



**HAL**  
open science

# Détection d'une source faible: modèles et méthodes statistiques. Application à la détection d'exoplanètes par imagerie directe.

Isabelle Smith

## ► To cite this version:

Isabelle Smith. Détection d'une source faible: modèles et méthodes statistiques. Application à la détection d'exoplanètes par imagerie directe.. Astrophysique [astro-ph]. Université Nice Sophia Antipolis, 2010. Français. NNT: . tel-00548905

**HAL Id: tel-00548905**

**<https://theses.hal.science/tel-00548905>**

Submitted on 20 Dec 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Laboratoire Fizeau – UMR 6525 CNRS / UNS / OCA  
UNIVERSITÉ DE NICE – SOPHIA ANTIPOLIS – UFR Sciences  
École Doctorale de Sciences Fondamentales et Appliquées

## THÈSE

pour obtenir le titre de  
**Docteur en Sciences**  
de l'UNIVERSITÉ de Nice-Sophia Antipolis  
Spécialité : *Sciences de l'Univers*

PRÉPARÉE ET SOUTENUE PUBLIQUEMENT LE 26 NOVEMBRE 2010 PAR :  
ISABELLE SMITH

**Détection d'une source faible :  
modèles et méthodes statistiques.**

**Application à la détection d'exoplanètes  
par imagerie directe.**

### Jury :

M. Jean-François CARDOSO	Directeur de recherche CNRS – Telecom/PariTech	Rapporteur
M. Jean-Yves TOURNERET	Professeur – Université de Toulouse	Rapporteur
M. Olivier MICHEL	Professeur – INPG	Président du jury
M. David MOUILLET	Astronome adjoint – Observatoire de Grenoble	Examineur
M. André FERRARI	Professeur – UNS	Directeur de thèse
M. Marcel CARBILLET	Maître de conférence – UNS	Co-Directeur de thèse



# Résumé / Summary

## Détection d'une source faible : modèles et méthodes statistiques. Application à la détection d'exoplanètes par imagerie directe.

Cette thèse contribue à la recherche de planètes extra-solaires à partir d'instruments au sol imageant une étoile et son environnement très proche. Le grand contraste lumineux et la proximité entre une potentielle exoplanète et son étoile parente rendent la détection de l'exoplanète extrêmement difficile. Une modélisation qualitative et probabiliste fine des données et l'utilisation de méthodes d'inférence adaptées permettent d'accroître *a posteriori* les performances des instruments. Cette thèse se focalise ainsi sur l'étape de traitement des données et sur un problème de méthodologie statistique plus général. Chaque étude est abordée sous des angles théoriques et appliqués.

La thèse décrit d'abord les données attendues pour le futur instrument SPHERE du *Very Large Telescope*, simulées à partir d'une modélisation physique détaillée. Un modèle probabiliste simple de ces données permet notamment de construire une procédure d'identification de candidats. Les performances des inférences sont aussi étudiées à partir d'un modèle décrivant de façon plus réaliste les bruits caractérisant les images (bruit de *speckle* corrélé, bruit de Poisson). On souligne la différence entre les probabilités de fausse alarme calculées à partir du modèle simple et à partir du modèle réaliste.

Le problème est ensuite traité dans le cadre bayésien. On introduit et étudie d'abord un outil original de test d'hypothèses : la distribution *a posteriori* du rapport de vraisemblance, notée PLR. Son étude théorique montre notamment que dans un cadre d'invariance standard le PLR est égal à une *p-value* fréquentiste. Par ailleurs, un modèle probabiliste des données est développé à partir du modèle initial et un modèle probabiliste de l'intensité de l'exoplanète est proposé. Ils sont finalement utilisés dans le PLR et le facteur de Bayes.

**Mots clés :** Traitement Statistique du Signal, Exoplanète, Détection, Estimation, Bayes, Imagerie directe, *Speckle*

## Detection of a weak source: statistical models and methods. Application to exoplanets detection using direct imaging.

This thesis is concerned by extra-solar planets research using ground-based instruments that image a star and its very closeby environment. Such a research is extremely difficult due to the high light contrast ratio and the proximity between a potential exoplanet and its parent star. A fine qualitative and probabilistic description of the data and relevant inferential methods increase *a posteriori* the instruments performances. Hence, this thesis focuses on signal processing methods and on a more general statistical methodological issue. All studies are performed theoretically and practically.

First, this thesis describes the data expected for the future SPHERE instrument of the Very Large Telescope, simulated from a detailed physical model. A simple probabilistic model of these data is used in particular to identify candidates. The inferences performances are also studied from a model that describes more realistically the noises that characterize the images (correlated speckle noise, Poisson noise). The difference between the false alarm probabilities computed from the simple model and from the realistic model is emphasized.

Then, the problem is adressed under the Bayesian paradigm. First, an original hypotheses test tool is introduced and studied: the posterior distribution of the likelihood ratio, denoted PLR. Its theoretical study especially shows that in a standard invariance framework the PLR is equal to a frequentist *p-value*. Further, a probabilistic model of the data is developed from the initial model and a probabilistic model of the light intensity of the exoplanet is proposed. They are finally used in the PLR and the Bayes factor.

**Keywords :** Statistical Signal Processing, Exoplanet, Detection, Estimation, Bayes, Direct Imaging, Speckle



# Remerciements

Cette thèse que j’espérais depuis une dizaine d’années ne devrait pas être signée que par moi. Comme toute réalisation dans la société et dans l’univers en général, elle est à considérer comme la partie apparente d’un iceberg dont les fondations reposent (Archimèdement parlant) bien plus profondément dans l’océan ou comme un simple accomplissement parmi toutes les curiosités et merveilles de tous types qui remplissent la nature. En tant que tâche rémunérée pleinement acceptée il était de ma responsabilité de la réaliser dans une mesure raisonnable et dans un cadre en partie pré-défini. Mais comme beaucoup des personnes qui s’orientent vers la recherche j’étais mue par une énergie émanant du monde plus ou moins proche de moi, la curiosité et la passion des Sciences en tant qu’objet et en tant que moyen de découverte construite et méthodique. Réaliser et achever cette thèse était ainsi particulièrement important pour moi et je souhaite en faire un *cadeau*, en tout cas original, pour tout ce monde, et en partie comme un résultat j’espère intéressant pour d’autres, au-delà du fait qu’il répond en partie à des problématiques de communauté(s) de recherche agréées.

En ce qui concerne les êtres humains –les seuls capables de *comprendre* des textes et les seuls malheureusement à qui j’adresserai mes remerciements pour ne choquer personne– je tiens déjà à remercier mes fondations. En premier lieu et instant vient ma famille, notamment pour m’avoir élevée dans la joie et fait grandir à l’abri du besoin mais aussi pour m’avoir éveillée aux curiosités et à la richesse de la Science. Vient ensuite tout le milieu dans lequel j’ai grandi, mes amis et connaissances diverses plus ou moins proches ainsi que tous les gens que je ne connais pas mais qui ont pourtant grandement contribué à ce que j’ai fait ou non de ma vie. Parmi les gens qui ont marqué directement et avec le plus d’insistance ma vie, un merci insondable à ma mère, mon père, Sophie, Christine, Koun, Vanessa, Mme Michel, Mélina, Claire, M. Grange, Thomas, Rima et Cedric, tous caractérisés par un très grand coeur, certains par leur sensibilité, certains (éventuellement les mêmes) par leur volonté et/ou leur exigence communicatives. Rencontrés plus tard pendant une année aux Etats-Unis et cotoyés moins longtemps, John, Sharon, Matt, Mike et Ryan ont aussi profondément marqué ma perception du fonctionnement de la société, des relations aux autres et aux choses *etc.* Au niveau scientifique, mais toujours antérieur à ma thèse à proprement parler, je tiens à remercier notamment par ordre chronologique mon père, Koun, l’ex-association Altaïr et ses membres dont Sophie, Clotilde, Igor, Raphaël et Guillaume pour leurs réalisations individuelles et collectives, M. Grange, Cedric, Emmanuel, Matt, Thomas Loredon et d’autres professeurs d’université, Alain, Keyan et Myriam.

Parmi les personnes que j’ai pu rencontrer pendant ma thèse, je remercie en premier lieu Andrea Ferrari, entre autres pour son efficacité et autres ingrédients nécessaires à la réalisation de projets, son expertise technique et la liberté qu’il m’a laissée. Je remercie mon second directeur de thèse Marcel Carbillet pour ses conseils et sa gentillesse. Merci aux rapporteurs pour le temps qu’ils ont passé à lire et discuter cette thèse et pour les améliorations qu’ils ont suggérées. Je remercie les examinateurs pour leur intérêt et le regard extérieur qu’ils ont apporté pendant la soutenance. Par ailleurs, j’adresse un merci particulier à Jean Arnaud, décédé alors qu’il dirigeait avec dévouement le laboratoire Fizeau.

Dans le cadre plus indirect de ma thèse, je tiens à remercier tous les amis doctorants, entre-soutiens fondamentaux à la réalisation d’une thèse notamment, et plus généralement d’une vie heureuse quand on a la chance de tomber sur des personnes exceptionnelles. Un infini merci en particulier à Alexis et Wassila avec qui j’ai tant partagé et sans qui je ne suis pas totalement sûre que j’aurais fini cette thèse. Sciences,



FIG. 1: Treifla Katfeuil et cousines de fortune.

questionnements sociétaux, collectifs, individuels, délires, sports extrêmes, besoins vitaux (remplissages gustatif et respiratoire en particulier), arts en tous genres et nombres, le tout avec beaucoup d'imagination, que demander de plus ? Dans le même ordre vracké d'idées, Eric, une source de renouvellement à peu près aussi inépuisable que les montagnes et les personnes généreuses que j'ai découvertes grâce à lui pendant quatre ans. Guillaume l'explosif, Guillaume le philosophe, inconfondables. Il est facile mais trop réducteur de réduire chacun à un qualificatif. Des pensées revigorantes pour Adrian, Alexandre, Amokrane, Bruno, Elsa, Emilie, Géraldine, Hamad, Hasan, Isabelle, Juan, Laurent, Luca, Mélanie, Nicolas, Slobodan, Tanya, Thierry, Vanessa ... Merci aux membres du laboratoire et en particulier à Fix, Eric-s, Mustafa, Cyrille, Gilbert, Valérie, Jocelyne, Gérard, David, Evelyne, Yan, Claude, Farrokh ... Et un remerciement élané pour toutes les personnes investies pour la collectivité *via* leurs responsabilités, leurs activités associatives, leur investissement en général. A Nice je pense *notamment* à Estelle Blanquet, Alain Puppo, Alexander Samuel, Adrien Catella, Cédric Jacob, Alexis Carlotti, Lionel Alberti et tant d'autres.

Un immense merci à tout et tous, sans oublier tous ceux qui agissent dans l'ombre et sans qui pourtant les choses seraient bien différentes.

Tout ceci ayant été relativement mal dit et trop sérieux peut-être<sup>1</sup>, un petit merci au Shah Dokteur puisque quand on ne sait pas où l'on va, il faut y aller et le plus vite possible (preuve et corrolaire : quand on ne sait pas ce que l'on dit, il faut le dire et le plus vite possible), aux exopingouins cométaires adeptes du golf marsouen, et à Treifla Katfeuil 'n family (voir la figure 1). Sans vous tous et associés la vie serait un peu rigid'ure et insensée.

<sup>1</sup>Comme le baragouinent si bien de grands sages jurassiques, no brain, no fear, XD forever.

# Introduction générale et notations

## Introduction générale

La recherche de planètes extra-solaires n'est devenue possible technologiquement qu'il y a à peine plus d'une dizaine d'années. La présence d'une étoile parente –telle que notre Soleil pour les planètes du Système Solaire– très intense et dans l'environnement très proche de l'exoplanète rend en effet cette tâche particulièrement difficile : avec les télescopes et instruments les plus actuels, exoplanète et étoile parente ne formaient jusqu'à récemment qu'un objet céleste indiscernable.

Les premières techniques de détection étaient indirectes, l'exoplanète étant détectée grâce à son effet sur l'étoile qu'elle accompagne et non par son effet direct (éclairage) sur l'observateur. L'émergence d'instruments capables de rendre la lumière provenant de l'exoplanète distincte de celle provenant de l'étoile est plus récente encore. Cette thèse s'attache particulièrement à l'instrument de recherche d'exoplanètes par imagerie directe SPHERE qui sera monté sur le VLT vers 2011.

Malgré la qualité des méthodes et matériels instrumentaux utilisés pour l'imagerie directe d'exoplanète, ceux-ci aboutissent, pour le type de données qui nous intéresse ici, à des images caractérisées par l'éventuelle présence d'une planète (minuscule tache d'intensité lumineuse négligeable) noyée dans un fond lumineux dû à la lumière diffusée et diffractée de l'étoile. Des méthodes de traitement des données sophistiquées sont nécessaires pour extraire une planète du bruit ambiant et quantifier ses caractéristiques avec soin.

A cette fin, deux ingrédients sont nécessaires : un modèle physique-statistique décrivant finement les processus aléatoires brouillant en permanence les données (atmosphère de la Terre traversée par la lumière de l'étoile, bruits de lecture de la caméra d'acquisition ...) et une méthode de construction d'inférences dont les performances sont en général bonnes (ex : estimation d'un paramètre par la valeur de paramètre qui rend les données les plus probables) et en particulier pour le modèle statistique supposé bien décrire les données.

Cette thèse s'attache essentiellement à développer ou utiliser des modèles statistiques des images attendues afin de construire des procédures de traitement de ces images et d'évaluer la qualité de ces procédures. Pour évaluer la qualité des méthodes de traitement, afin de choisir la plus performante et de constater si l'on devrait être capable en 2011 de détecter des exoplanètes avec SPHERE, des simulations physiques plus fines élaborées par le consortium chargé de la conception et de la construction de l'instrument SPHERE sont aussi utilisées pour tester les procédures de façon réaliste.

Par ailleurs, notons qu'une fois le modèle statistique des données choisi, la méthode pour construire un estimateur d'un paramètre numérique inconnu fait à peu près consensus dans la communauté des traiteurs de signaux, mais qu'en revanche la méthode pour construire un test d'hypothèses (ex : procédure qui répond à la question "Y a-t-il une exoplanète dans cette image?") est beaucoup plus débattue, notamment autour des notions de *p-value* et de probabilité de fausse alarme (PFA, probabilité pour une expérience donnée de répondre "oui" alors que la réponse était "non").

Cette thèse étudie ainsi une méthodologie générale de construction de test d'hypothèses. Son interprétation naturelle et ses propriétés sont en effet intéressantes tant quant à leur interprétation que leur utilisation pratique, mais la méthode avait encore été peu étudiée.



Plus précisément, cette thèse est structurée de la façon suivante :

- Le chapitre 1 introduit la problématique astrophysique traitée (partie destinée surtout aux non-astrophysiciens), puis les techniques et notions importantes pour construire des techniques de traitement du signal (partie destinée surtout aux non-traités de signaux), et liste finalement les grandes questions et les techniques les plus classiques relatives aux tests d’hypothèse (synthèse de notions et de notations intéressante pour tous).
- La partie I introduit tout d’abord le modèle physique fin utilisé par le consortium SPHERE pour simuler les données attendues pour le module de l’instrument SPHERE qui nous intéresse et présente une étude statistique qualitative des données ainsi simulées. Le chapitre 3 propose alors d’utiliser un modèle des données gaussien simple afin de construire une procédure de traitement des images. Une liste de candidats potentiels dans la recherche d’exoplanètes est alors extraite afin de pouvoir leur appliquer des techniques plus performantes mais plus gourmandes en temps de calcul, techniques qui seront l’objet des autres parties. Les propriétés statistiques de la procédure déduite du modèle gaussien sont étudiées, à la fois sous ce modèle simple et en la testant sur les simulations physiques fines. Un programme informatique, permettant en pratique d’appliquer la méthode mathématique, a été développé. Son utilisation est détaillée dans l’annexe A. Son application au jeu de données simulé finement indique les performances de la méthode, performances servant de point de comparaison aux méthodes développées dans la suite.
- La partie II est essentiellement destinée à calculer la probabilité que la procédure de détection précédente se trompe en l’absence de planète. Construire une procédure qui se trompe rarement est important mais le fait de le savoir l’est presque tout autant. En effet, sans pouvoir quantifier une PFA, notamment, on ne saurait quelle confiance attribuer à une annonce de détection d’exoplanète. Cependant, pour arriver à cette fin un modèle statistique détaillé est nécessaire. En l’occurrence, le modèle, développé en amont de cette thèse, repose sur une description fine des processus optiques déterministes et aléatoires se produisant à mesure que la lumière se déplace de l’étoile vers la caméra d’acquisition des données. Le modèle statistique étant particulièrement complexe, la méthode pour calculer la PFA associée l’est aussi. Les performances de la méthode d’estimation de la PFA, l’utilité du modèle statistique et la difficulté de son utilisation pratique sont étudiées et discutées.
- La partie III étudie une méthode de test d’hypothèses *bayésienne* naturelle et originale. Elle consiste à seuiliser une probabilité, sur l’ensemble des planètes qui pourraient être observées plutôt que sur l’ensemble de données qui pourraient être enregistrées, pour que la densité de probabilité des données observées soit plus faible en l’absence de planète qu’en présence d’une planète. On appelle cette probabilité le PLR. Après une introduction à la problématique et un état de l’art, des propriétés de base de ce concept sont dérivées. Par ailleurs, pour un modèle des données invariant et sous des conditions émergeant naturellement de cette condition (sujet introduit dans un chapitre dédié), il s’avère que le PLR est égal à une *p-value* fréquentiste, indiquant la correspondance qui peut exister entre probabilité sur un ensemble d’observations et probabilité sur un ensemble de planètes.
- Enfin, la partie IV développe un modèle bayésien qui propose notamment un modèle des données alternatif au modèle statistique détaillé de la partie II. Quelques tests –dont le PLR– sont construits à partir de ce modèle et discutés. Ils sont appliqués aux images simulées finement et leurs performances sur un grand nombre d’observations d’une étoile donnée sont étudiées à partir de données simulées plus simplement.
- Des conclusions et perspectives clôtureront finalement la thèse.

## Notations et définitions

### Fonctions usuelles des scalaires, vecteurs et matrices

#### • Scalaires (minuscules normales)

$a$	scalaire (réel ou complexe).
$\mathcal{Re}(a)$ ou $\mathcal{Im}(a)$	partie réelle ou respectivement imaginaire de $a$
$a^*$	complexe conjugué de $a$ ( $\mathcal{Re}(a^*) = \mathcal{Re}(a)$ et $\mathcal{Im}(a^*) = -\mathcal{Im}(a)$ )
$ a $	valeur absolue / module de $a$
$\lfloor a \rfloor$	partie entière de $a$

#### • Vecteurs (minuscules grasses)

$\mathbf{a}$	vecteur colonne
$a_\ell$	composante $\ell$ de $\mathbf{a}$
$\mathbf{a}^t$	vecteur ligne transposé de $\mathbf{a}$
$\mathbf{a}^H$	vecteur transposé Hermitien de $\mathbf{a}$ ( $\mathbf{a}^H = \mathbf{b}^t$ avec $b_\ell = a_\ell^*$ pour tout indice $\ell$ )
$\mathbf{1}_L$	vecteur de taille $L \times 1$ dont toutes les composantes sont égales à 1
$ \mathbf{a} $	vecteur contenant la valeur absolue de chaque composante de $\mathbf{a}$ ( $ \mathbf{a} _\ell =  a_\ell $ )
$\ \mathbf{a}\ $	norme $L^2$ de $\mathbf{a}$ ( $\ \mathbf{a}\  = \sqrt{\sum_\ell \ a_\ell\ ^2}$ )
$(a_1, \dots, a_L)$	concaténation en ligne des scalaires $a_\ell$ ( $\mathbf{a} = (a_1, \dots, a_L)^t$ est un vecteur colonne de taille $L \times 1$ )
$(\mathbf{a}(1)^t, \dots, \mathbf{a}(K)^t)$	concaténation en ligne des vecteurs lignes $\mathbf{a}(k)^t$ ( $\mathbf{a} = (\mathbf{a}(1)^t, \dots, \mathbf{a}(K)^t)^t$ est un vecteur de taille $KL \times 1$ si chaque $\mathbf{a}(k)$ est de taille $L \times 1$ )
$e^{\mathbf{a}}$	vecteur contenant les $e^{a_\ell}$ ( $e^{\mathbf{a}} = (e^{a_1}, \dots, e^{a_L})^t$ )

#### • Matrices (majuscules normales)

$A$	matrice
$A_{i,j}$	terme $(i, j)$ de $A$
$A^t$	matrice transposée de $A$
$A^H$	matrice transposée Hermitienne de $A$ ( $A^H = B^t$ avec $B_{i,j} = (A_{i,j})^*$ )
$I_L$	matrice identité de taille $L \times L$ ( $I_{\ell,\ell'} = 1$ si $\ell = \ell'$ et $I_{\ell,\ell'} = 0$ sinon)
$A^{-1}$	matrice inverse de $A$ (si $A$ est de taille $L \times L$ alors $A^{-1}A = AA^{-1} = I_L$ )
$D_{\mathbf{a}}$	matrice diagonale dont les termes diagonaux sont donnés par les composantes du vecteur $\mathbf{a}$
$Tr(A)$	trace de $A$ ( $Tr(A) = \sum_\ell A_{\ell,\ell}$ )
$ A $	déterminant de $A$

## Ensembles

$\mathbb{C}$	ensemble des nombres complexes
$\mathbb{R}$	ensemble des nombres réels
$\mathbb{R}_+$	ensemble des réels non négatifs
$\mathbb{R}_{+*}$	ensemble des réels positifs
$[a, b]$	ensemble des réels inclus entre $a$ et $b$
$\llbracket 1, L \rrbracket$	ensemble des entiers compris entre 1 et $L$
$\{\mathbf{x}(k), k \in \llbracket 1, K \rrbracket\}$	ensemble des $\mathbf{x}(k)$ pour $k$ variant dans $\llbracket 1, K \rrbracket$
/ ou :	“tel que” (ex. : $\{\ell \in \llbracket 1, L \rrbracket / x_\ell > \epsilon\}$ )
$I_{\mathcal{A}}(a)$	fonction indicatrice de l'ensemble $\mathcal{A}$ : $I_{\mathcal{A}}(a) = 1$ si $a \in \mathcal{A}$ et $I_{\mathcal{A}}(a) = 0$ sinon
$\mathcal{M}(\mathcal{A})$	mesure de l'ensemble (mesurable) $\mathcal{A}$ . Ex : mesure de probabilité $P(\mathcal{A}) = \Pr(\mathbf{x} \in \mathcal{A})$
$\text{Card}(\mathcal{A})$	cardinal (nombre d'éléments) de l'ensemble (dénombrable) $\mathcal{A}$

## Statistiques

### • Notions liée à une distribution de probabilité

$X$ ou $\mathbf{X}$	variable aléatoire scalaire ou respectivement vectorielle. Peu utilisé : une majuscule spécifique dans cette thèse en général que le terme désigne une matrice, pas une variable aléatoire.
$x$ ou $\mathbf{x}$	réalisation de $X$ ou respect. $\mathbf{X}$ . Désigne aussi par abus de langage $X$ ou $\mathbf{X}$ .
$\Pr(X \in [a, b])$ ou $\Pr(\mathbf{X} \in [a, b]^L)$	probabilité pour que $X$ (noté en général abusivement $x$ ) soit entre $a$ et $b$ ou respect. pour que chaque composante du vecteur $\mathbf{X}$ soit entre $a$ et $b$
$p(\mathbf{x})$ pour $p_{\mathbf{X}}(\mathbf{x})$	fonction de densité de probabilité (PDF) de la variable aléatoire continue $\mathbf{X}$ évaluée en $\mathbf{x}$ (on note par ex. $\Pr(\mathbf{x} \in [a, b]^L) = \int_{[a, b]^L} d\mathbf{x} p(\mathbf{x})$ )
$F_{\mathbf{X}}(\mathbf{x})$	fonction cumulative de densité (CDF) de la variable aléatoire continue $\mathbf{X}$ évaluée en $\mathbf{x}$ ( $F_{\mathbf{X}}(\mathbf{x}) = \Pr(\mathbf{X} \leq \mathbf{x})$ , inégalité composante par composante )
$h_{\mathbf{x}}(\mathbf{s})$ pour $h_{\mathbf{X}}(\mathbf{s})$	fonction génératrice de moments (MGF) de la variable aléatoire $\mathbf{X}$ évaluée en $\mathbf{s}$ ( si $\mathbf{x}$ est continue, $h_{\mathbf{x}}(\mathbf{s}) = \mathbb{E}[\exp(\mathbf{x}^t \mathbf{s})]$ , cf ligne suivante)
$\mathbb{E}[x]$ pour $\mathbb{E}[\mathbf{X}]$	espérance de $\mathbf{X}$ ( ex. : si $\mathbf{x}$ est continue, $\mathbb{E}[f(\mathbf{x})] = \int d\mathbf{x} f(\mathbf{x})p(\mathbf{x})$ )
$\text{var}[x]$ ou $\text{cov}[\mathbf{x}]$	variance de $X$ ou respectivement matrice de covariance du vecteur $\mathbf{X}$
$J$	matrice Jacobienne associée à un changement de variable bijectif (le Jacobien de $\mathbf{x} \rightarrow f(\mathbf{x})$ est donné par $J_{i,j} = \partial(f(\mathbf{x}))_i / \partial x_j$ )

## • Données et paramètres

En fréquentiste, une distribution de probabilité des données  $\mathbf{X}$  est simplement *paramétrée* par un ensemble de paramètres regroupés par exemple dans un vecteur  $\boldsymbol{\theta}$ . En bayésien en revanche une distribution *a posteriori* de  $\boldsymbol{\theta}$  n'est obtenue que si l'on explicite la dépendance en  $\boldsymbol{\theta}$  par un lien de *conditionnalité*. Une distribution conditionnelle n'est pas exactement une distribution de probabilité au sens strict mais nous ne nous soucierons pas des éventuels (rares) problèmes qui en découlent. Dans la suite, utilisant à la fois dans cette thèse les approches fréquentiste et bayésienne, nous confondrons paramétrisation et conditionnalité afin de coller chaque fois aux notations les plus standards.

Dans le même esprit, un compromis entre légèreté d'expression et absence d'ambiguïté nous amènera à préciser ou non la variable aléatoire considérée et le paramètre fixé. Nous n'utilisons donc pas de notations totalement unifiées dans cette thèse mais la description suivante devrait être suffisante.

$\mathcal{X}$	espace des données ( $\mathbf{x} \in \mathcal{X}$ )
$\Theta$	espace des paramètres ( $\boldsymbol{\theta} \in \Theta$ )
$p(\mathbf{x} \boldsymbol{\theta})$ ou $p_{\mathbf{X} \boldsymbol{\theta}}(\mathbf{x})$	PDF des données $\mathbf{X}$ évaluée en $\mathbf{x}$ pour une distribution de paramètre $\boldsymbol{\theta}$ . Aussi appelé ici "vraisemblance" quoique ce terme désigne en général $L(\boldsymbol{\theta}) \propto p(\mathbf{x} \boldsymbol{\theta})$ .
$F_{\mathbf{X} \boldsymbol{\theta}}(\mathbf{x})$	CDF de la variable aléatoire continue $\mathbf{X}$ évaluée en $\mathbf{x}$ pour une distribution de paramètre $\boldsymbol{\theta}$
$h_{\mathbf{x}}(\mathbf{s})$ ou $h_{\mathbf{x} \boldsymbol{\theta}}(\mathbf{s})$	MGF des données $\mathbf{X}$ évaluée en $\mathbf{s}$ (paramètre $\boldsymbol{\theta}$ en général implicite)
$E[\mathbf{x}]$ ou $E_{\boldsymbol{\theta}}[\mathbf{X}] \dots$	espérance de $\mathbf{X}$ (paramètre $\boldsymbol{\theta}$ parfois implicite)
$\pi(\boldsymbol{\theta})$	PDF du paramètre $\boldsymbol{\theta}$ (densité <i>a priori</i> d'un paramètre notée $\pi(\cdot)$ plutôt que $p(\cdot)$ )
$\pi(\boldsymbol{\theta} \mathbf{x})$	densité <i>a posteriori</i> de $\boldsymbol{\theta}$ (Bayes : $\pi(\boldsymbol{\theta} \mathbf{x}) = p(\mathbf{x} \boldsymbol{\theta})\pi(\boldsymbol{\theta}) (\int d\boldsymbol{\theta} p(\mathbf{x} \boldsymbol{\theta})\pi(\boldsymbol{\theta}))^{-1}$ )
$\hat{\boldsymbol{\theta}}(\mathbf{x})$	estimateur du paramètre $\boldsymbol{\theta}$ ( $\hat{\boldsymbol{\theta}}(\mathbf{x})$ : statistique <i>i.e.</i> fonction des données $\mathbf{x}$ seulement)
$\arg \max_{\mathbf{a}} f(\mathbf{a}, \mathbf{b})$	valeur de $\mathbf{a}$ qui maximise pour $\mathbf{b}$ fixé la fonction $\mathbf{a} \rightarrow f(\mathbf{a}, \mathbf{b})$
$\hat{\boldsymbol{\theta}}_{\text{ML}}(\mathbf{x})$	estimateur de <i>maximum de vraisemblance</i> de $\boldsymbol{\theta}$ : maximise $p(\mathbf{x} \boldsymbol{\theta})$ . $\hat{\boldsymbol{\theta}}_{\text{ML}}(\mathbf{x}) = \arg \max_{\boldsymbol{\theta}} p(\mathbf{x} \boldsymbol{\theta})$
$\hat{\boldsymbol{\theta}}_{\text{MAP}}(\mathbf{x})$	estimateur de <i>maximum a posteriori</i> de $\boldsymbol{\theta}$ : maximise $\pi(\boldsymbol{\theta} \mathbf{x})$
$\hat{\boldsymbol{\theta}}_{\text{MMSE}}(\mathbf{x})$	estimateur de <i>minimum d'erreur quadratique moyenne</i> de $\boldsymbol{\theta}$ : minimise la moyenne <i>a posteriori</i> de $\ \hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\ ^2$ (propriété : $\hat{\boldsymbol{\theta}}_{\text{MMSE}}(\mathbf{x}) = E[\boldsymbol{\theta} \mathbf{x}]$ )
$I(\boldsymbol{\theta})$	matrice d'information de Fisher associée au paramètre $\boldsymbol{\theta}$ (rarement utilisé ici, souvent égal à $I_{i,j}(\boldsymbol{\theta}) = -E_{\boldsymbol{\theta}}[\partial^2 \ln p(\mathbf{x} \boldsymbol{\theta}) / (\partial\theta_i \partial\theta_j)]$ )
$H_0$ ou $H_1$	hypothèse nulle ou respectivement hypothèse alternative
$S : \mathcal{A} \rightarrow \mathcal{B}$	fonction de domaine de départ $\mathcal{A}$ et de domaine d'arrivée $\mathcal{B}$ . Une fonction n'est pas notée en gras ( $S(\mathbf{a})$ ), sauf s'il s'agit explicitement d'un estimateur ( $\hat{\boldsymbol{\theta}}(\mathbf{x})$ ). L'utilisation d'une minuscule ou majuscule n'apporte pas d'information.

## • Distributions de probabilité standards

$\mathcal{P}(\mu)$	loi de Poisson de moyenne $\mu$
$\mathcal{N}(\mu, \sigma^2)$	loi normale monovariée de moyenne $\mu$ et de variance $\sigma^2$
$\mathcal{N}_L(\boldsymbol{\mu}, \Sigma)$	loi normale d'un vecteur aléatoire de taille $L$ de moyenne $\boldsymbol{\mu}$ et mat. de covariance $\Sigma$
$\mathcal{N}_L^c(\boldsymbol{\mu}, \Sigma)$	loi normale circulaire d'un vecteur complexe de taille $L$ de paramètres $\boldsymbol{\mu}, \Sigma$
$\mathcal{W}_L(\Sigma, n)$	loi de Wishart d'une matrice aléatoire de taille $L \times L$ de paramètres $\Sigma, n$
$\mathcal{W}_L^{-1}(\Sigma, n)$	loi inverse Wishart d'une matrice aléatoire de taille $L \times L$ de paramètres $\Sigma, n$ (si $S \sim \mathcal{W}_L(\Sigma, n)$ alors $S^{-1} \sim \mathcal{W}_L^{-1}(\Sigma^{-1}, n)$ )
$\mathcal{St}_L(\boldsymbol{\mu}, \Sigma, \nu)$	loi de Student d'une matrice aléatoire de taille $L \times L$ de paramètres $\boldsymbol{\mu}, \Sigma, \nu$
$\Gamma_L(\nu)$	fonction gamma multivariée de paramètre $\nu$ (il ne s'agit pas d'une distribution)

## Termes principaux de modélisation physique

Les données  $\mathbf{x}$  sont un ensemble d'images acquises à différents instants et éventuellement simultanément dans différentes bandes spectrales. La structure spatiale bi-dimensionnelle d'une image n'est en général pas utilisée pour la modélisation statistique, de telle sorte que l'on peut représenter sans difficulté une image comme un vecteur dont chaque composante  $\ell$  contient le nombre d'électrons enregistré par le pixel  $\ell$  du capteur CCD. L'ordonnement mathématique des pixels de la CCD se fait en général en concaténant les colonnes de la CCD les unes sous les autres.

### • Données : grandeurs mesurées

$K$	nombre de poses prises durant la durée d'observation
$L$	nombre de pixels d'un capteur CCD (1 capteur CCD par bande spectrale)
$M$	nombre de bandes spectrales (en général dans cette thèse $M = 1$ : cas $M > 1$ traité en second temps)
$x_\ell(k)$	pour $M = 1$ , nombre d'électrons lu par le pixel $\ell \in \llbracket 1, L \rrbracket$ à la pose $k \in \llbracket 1, K \rrbracket$
$\mathbf{x}(k)$	pour $M = 1$ , vecteur de taille $L \times 1$ contenant l'image acquise à l'instant $k$ (pour $M = 1$ ) $\mathbf{x}(k) = (x_1(k), \dots, x_L(k))^t$
$\mathbf{x}(k, m)$	vecteur de taille $L \times 1$ contenant l'image acquise à l'instant $k$ dans la bande spectrale $m$
$\mathbf{x}$	vecteur de taille $KLM \times 1$ contenant l'ensemble des images concaténées. Pour $M = 1$ par ex., $\mathbf{x} = (\mathbf{x}(1)^t, \dots, \mathbf{x}(K)^t)^t$

### • Composantes physiques : premier pas de modélisation

On ne considère ci-dessous que le cas mono-bande ( $M = 1$ ) :

$\alpha$	intensité (scalaire) de l'exoplanète. Noter qu' $\alpha$ désigne plutôt parfois le niveau de significativité (PFA) d'un test.
$r$	position de la planète sur la première image ( $k = 1$ )
$\mathbf{p}(k, r)$	vecteur de taille $L \times 1$ représentant le profil de la planète pour la pose $k$ si sa position initiale est $r$ . $r$ en général implicite car généralement supposé connu dans cette thèse.
$\alpha\mathbf{p}(k, r)$	vecteur de taille $L \times 1$ représentant la contribution totale d'une planète sur l'image $k$
$\mathbf{d}$ ou $\mathbf{i}$	vecteur de taille $L \times 1$ représentant la contribution (supposée statique mais non uniforme spatialement) de l'étoile, notations respectives des modèles 1 et 2

## Sigles

ADI	<i>Angular Differential Imaging</i>
AO	<i>Adaptive Optics</i>
BF	<i>Bayes Factor</i>
CCD	<i>Charge-Coupled Device</i>
CDF	<i>Cumulative Distribution Function</i>
DBI	<i>Dual-Band Imaging</i>
FBF	<i>Fractional Bayes Factor</i>
GLR	<i>Generalized Likelihood Ratio</i>
IRDIS	<i>Infra-Red Dual-beam Imaging and Spectroscopy</i>
LR	<i>Likelihood Ratio</i>
MCMC	<i>Monte Carlo Markov Chain</i>
MGF	<i>Moment Generating Function</i>
MOODS	<i>MOVing Objects Detection for SPHERE</i>
PD	<i>Probability of good Detection</i>
PDF	<i>Probability Density Function</i>
PFA	<i>Probability of False Alarm</i>
PLR	<i>Posterior cumulative distribution of the Likelihood Ratio</i>
PSF	<i>Point-Spread Function</i>
ROC	<i>Receiver Operating Characteristics</i>
SBI	<i>Single-Band Imaging</i>
SDI	<i>Spectral Differential Imaging</i>
SPHERE	<i>Spectro-Polarimetric High-contrast Exoplanet Research</i>
UMP(U)	<i>Uniformly Most Powerful (Unbiased)</i>
VLT	<i>Very Large Telescope</i>



# Table des matières

<b>Introduction générale et notations</b>	<b>v</b>
<b>Table des matières</b>	<b>1</b>
<b>1 Introduction à la détection d'exoplanètes et aux inférences statistiques</b>	<b>5</b>
1.1 Détection d'exoplanètes par imagerie directe et l'instrument SPHERE en particulier . . . . .	5
1.2 Développement de méthodes d'inférence statistiques pour le traitement du signal . . . . .	10
1.3 Introduction aux tests d'hypothèse fréquentistes et bayésiens . . . . .	17
<b>I Modèle gaussien simple pour construire des inférences applicables</b>	<b>31</b>
<b>2 Simulation physique fine des données SPHERE</b>	<b>35</b>
2.1 Simulation détaillée d'un jeu de données standard acquis par SPHERE . . . . .	35
2.2 Etude statistique qualitative des données simulées en l'absence de planète . . . . .	40
<b>3 Construction et analyse d'inférences à partir d'un modèle gaussien</b>	<b>55</b>
3.1 Modèle gaussien des données . . . . .	55
3.2 Estimation d'intensité et de position de la planète . . . . .	58
3.3 Consistance des estimateurs et estimation d'incertitude . . . . .	63
3.4 Détection initiale : obtention d'une liste de candidats . . . . .	68
3.5 Extension de la détection-estimation à l'imagerie simultanée dans plusieurs canaux spectraux	71
<b>4 Application aux données simulées finement et résultats</b>	<b>75</b>
4.1 Discussions plus détaillées de la validité des conditions . . . . .	75
4.2 Programme informatique <i>user-friendly</i> : MOODS sur Matlab . . . . .	78
4.3 Application de la méthode à des données (simulées) stationnaires et non-stationnaires . . . . .	81
4.4 Autres applications de la méthode . . . . .	86
<b>II Modèle fin d'imagerie <i>speckle</i> pour quantifier les inférences précédentes</b>	<b>91</b>
<b>5 Modèle fin des données et distribution statistique de l'estimateur linéaire d'intensité</b>	<b>95</b>
5.1 Description physique qualitative et modèle statistique des données (état de l'art) . . . . .	95
5.2 Distribution statistique fine de l'estimateur linéaire d'intensité de la section 3.2 . . . . .	102
5.3 Consistance de l'estimateur d'intensité . . . . .	104
<b>6 Probabilité de Fausse Alarme du détecteur de la section 3.4 sous modèle <i>speckle</i></b>	<b>105</b>
6.1 Approximations analytiques de la PFA . . . . .	105
6.2 Comparaison des méthodes d'approximation par simulation pour le modèle <i>speckle</i> . . . . .	107
6.3 Utilisation du modèle <i>speckle</i> en pratique . . . . .	112



<b>III</b>	<b>Méthode bayésienne de test d'hypothèse simple et originale : étude théorique</b>	<b>117</b>
<b>7</b>	<b>Introduction à la distribution <i>a posteriori</i> du rapport de vraisemblance (PLR)</b>	<b>121</b>
7.1	Approche de Neyman-Pearson . . . . .	121
7.2	Problématiques bayésiennes . . . . .	125
7.3	Etat de l'art sur la distribution cumulative <i>a posteriori</i> du rapport de vraisemblance (PLR)	128
<b>8</b>	<b>Quelques propriétés générales du PLR</b>	<b>133</b>
8.1	Etude de la loi <i>a posteriori</i> du rapport de vraisemblance . . . . .	133
8.2	Remarques simples concernant le PLR en tant que test binaire . . . . .	137
8.3	Implémentation et optimisation fréquentiste numériques du PLR . . . . .	141
<b>9</b>	<b>Introduction à l'invariance de groupe appliquée aux statistiques</b>	<b>143</b>
9.1	Motivation pour l'utilisation de l'invariance de groupe en statistiques . . . . .	143
9.2	<i>A priori</i> de Haar . . . . .	144
9.3	Famille de vraisemblances invariante sous un groupe de transformations . . . . .	152
<b>10</b>	<b>En cas invariant, équivalence entre le PLR et une <i>p-value</i> fréquentiste</b>	<b>155</b>
10.1	Théorème . . . . .	155
10.2	Conséquences de l'équivalence . . . . .	163
<b>IV</b>	<b>Modèle bayésien pour assouplir le modèle gaussien et application du PLR</b>	<b>167</b>
<b>11</b>	<b>Modèle bayésien hiérarchique</b>	<b>171</b>
11.1	Introduction . . . . .	171
11.2	Modèle des données . . . . .	173
11.3	Probabilité <i>a priori</i> de l'intensité de la planète dans les différents canaux spectraux . . . . .	175
11.4	Probabilité <i>a posteriori</i> de l'intensité et échantillonnage . . . . .	176
<b>12</b>	<b>Détection par le PLR pour les données SPHERE</b>	<b>181</b>
12.1	Jeux de données issus de la simulation physique fine . . . . .	181
12.2	Performances fréquentistes évaluées sur des données simulées plus simplement . . . . .	188
	<b>Conclusion et perspectives</b>	<b>195</b>
	<b>Publications et communications</b>	<b>199</b>
<b>V</b>	<b>Annexes</b>	<b>201</b>
<b>A</b>	<b>Algorithme MOODS</b>	<b>203</b>
A.1	Description textuelle de l'objectif scientifique de l'algorithme . . . . .	203
A.2	Motivations du choix de programmation, description plus détaillée et diagrammes d'activité	204
A.3	Listes des <i>inputs</i> demandés à l'utilisateur . . . . .	206
<b>B</b>	<b>Théorie bayésienne de la décision pour la construction d'inférences</b>	<b>211</b>
B.1	Théorie de la décision : généralités . . . . .	211
B.2	Estimation ponctuelle d'un paramètre . . . . .	212
B.3	Intervalle ou domaine de confiance d'un paramètre . . . . .	213
B.4	Test d'hypothèses . . . . .	215
<b>C</b>	<b>Figures et démonstrations supplémentaires</b>	<b>217</b>

C.1 Figures supplémentaires illustrant les données issues de la simulation physique fine . . . . . 217

C.2 Relation entre les PFA d'un test dans tout le champ et d'un test par pixel . . . . . 227

C.3 Images courte pose plus informatives qu'une longue pose . . . . . 230

C.4 Discontinuité du PLR en  $\zeta = \text{GLR}(\mathbf{x})$  . . . . . 231

**Bibliographie** . . . . . **235**



# Chapitre 1

## Introduction à la détection d'exoplanètes et aux inférences statistiques

### 1.1 Détection d'exoplanètes par imagerie directe et l'instrument SPHERE en particulier

#### Historique de la détection d'exoplanète

- Objets moins massifs que les étoiles

Dès l'Antiquité, les mouvements de Mercure, Vénus, Mars, Jupiter et Saturne relativement à la voûte céleste sont observés, ce qui vaut à ces objets le nom de “planète”, issu de l'expression grecque signifiant *astre errant*. L'instrumentation optique astronomique fait ses premiers pas en 1609 avec Galilée qui tourne vers le ciel les lunettes récemment inventées. En 1671, Newton réalise un premier type de télescope. Utilisant celui-ci, en 1781 Herschel identifie que l'objet maintenant nommé Uranus est une planète. En 1801, Piazzi découvre Cérès, la “planète manquante” prévue par Bode. Des ré-observations de cet objet l'année suivante –possibles grâce à la loi des moindres carrés de Gauss qu'il publiera en 1809– confirment que son orbite est plus circulaire que celle d'une comète. La découverte d'autres astéroïdes s'achève par le reclassement de Cérès du statut de *planète* à celui d'*astéroïde*. Enfin, prévue par les calculs de Le Verrier et Adams, Neptune est découverte en 1846, ce qui clôt la liste actuelle des huit planètes.

L'observation systématique du ciel permet à Tombaugh en 1930 de découvrir Pluton, qui est alors classée comme planète. Cependant, comme Cérès pour la ceinture d'astéroïdes, Pluton n'est qu'un membre assez quelconque d'une population maintenant nommée ceinture de Kuiper. Le premier de ces objets, en dehors de Pluton et de son plus gros satellite, est découvert en 1992 par Jewitt et Luu. Depuis, plus d'un millier de ces objets ont été détectés et identifiés, dont Eris en 2005 de masse légèrement supérieure à celle de Pluton (Jewitt [2010]). En 2006, l'Union Astronomique Internationale décide d'une nouvelle définition du mot planète et définit la classe des *planètes naines* dont Cérès et Pluton font dorénavant partie.

Parallèlement à la découverte de ces premiers objets transneptuniens, les premières naines brunes sont découvertes et identifiées à partir de 1995 (Basri [2000]). Cette classe d'objets constitue une classe tampon entre les étoiles et les exoplanètes, les étoiles étant durablement le siège de réactions nucléaires et les exoplanètes jamais au cours de leur vie. Le site de recensement de Gelino et al. [2010] fait état de plus de 700 naines brunes début 2010. En tant que classe adjacente, les naines brunes peuvent être difficilement discernables des exoplanètes.

Toujours dans la même période, les premières planètes extérieures au Système Solaire, appelées planètes extrasolaires ou exoplanètes, ont été découvertes et identifiées à partir de 1992 par Wolszczan and

Frail [1992] puis par Mayor and Queloz [1995]. Les exoplanètes détectées en 1992 sont en orbite autour d'un pulsar (étoile à neutrons particulière), qui constitue le stade tardif d'évolution d'une étoile massive : à ce stade l'étoile a soufflé ses couches externes et les éventuelles planètes l'accompagnant, empêchant les planètes détectées d'avoir la même histoire que les planètes du Système Solaire. Au contraire, l'exoplanète découverte en 1995 est en orbite autour d'une étoile dite de la séquence principale –comme le Soleil– en référence à la classification des étoiles dans le diagramme d'Hertzsprung-Russell. Le catalogue Schneider [2010] recense actuellement plus de 400 exoplanètes, dont 1% orbitant autour de pulsars. En 2009 a été détectée l'exoplanète la moins massive de celles connues actuellement, CoRoT-7b, de masse égale à 5 fois celle de la Terre (Queloz et al. [2009]).

### • Techniques de détection

En orbite autour d'une étoile, une exoplanète est très difficile à détecter de façon directe : le rapport de l'intensité lumineuse de l'étoile et de la planète est grand et la distance apparente qui les sépare petite, de telle sorte que la lumière de l'exoplanète est très difficilement discernable de celle de l'étoile. On mesure la distance apparente par l'angle entre l'étoile, la Terre et l'exoplanète, de même que n'importe quelle mesure de distance sur la sphère céleste. On s'intéresse typiquement à un champ (champ = zone du ciel) de quelques secondes d'arc au carré<sup>1</sup>. Avec un télescope de faible grossissement, une étoile devrait par exemple apparaître comme un point mais elle apparaît en fait comme une tache –turbulente dans le temps– dûe aux perturbations de la lumière de l'étoile lorsqu'elle traverse l'atmosphère de la Terre, et cette tache a une taille de l'ordre de la seconde d'arc. Un système Soleil-Terre vu à une distance de 10 parsecs (soit environ 33 années lumières, la distance que parcourt la lumière en 33 ans), *i.e.* dans l'environnement proche du Soleil, apparaît à l'observateur comme deux points lumineux distants de 0'1 et de rapport d'intensités lumineuses de  $5 \cdot 10^9$  en lumière visible et  $7 \cdot 10^6$  en infrarouge (Ollivier et al. [2009]).

Ainsi, sur les 500 exoplanètes actuellement détectées, moins de 3% ont été détectées par une méthode d'imagerie directe, et à partir de 2005 seulement (Chauvin et al. [2005]). Environ 95% des exoplanètes connues ont été détectées et confirmées par la combinaison des méthodes de spectroscopie Doppler, qui mesure l'éventuel mouvement de l'étoile induit par un autre corps, et de transit, qui mesure la variation d'intensité globale du système étoile + exoplanète en fonction du temps et permet de détecter des exoplanètes passant devant leur étoile et dont la période de rotation autour de celle-ci est suffisamment courte (quelques années). Ces méthodes ont récemment permis de détecter des exoplanètes de masse de l'ordre de 5 fois la masse de la Terre (exoplanètes les moins massives actuellement connues), voir Queloz et al. [2009], Charbonneau et al. [2009]. Les paramètres accessibles en premier à partir de ces techniques de détection indirecte sont la masse de l'exoplanète et ses paramètres orbitaux, à l'inclinaison de la ligne de visée par rapport au plan de l'orbite près.

Pour détecter la présence de vie sur une exoplanète, étudier son atmosphère est fondamental. Des composés chimiques plus ou moins complexes tels que l'ozone, l'eau ou des molécules organiques sont des éléments nécessaires à la vie sur Terre et leur détection constitue un premier indice dans la quête de la vie extra-terrestre. A cette fin, des méthodes spectroscopiques sont en cours de développement par méthode de transit (observation : Richardson et al. [2007]), par spectropolarimétrie (méthode : Stam [2008]) ou par imagerie directe (observation : Lafrenière et al. [2008]). L'imagerie directe permet de plus d'étudier directement l'interaction entre l'exoplanète et un éventuel disque de poussière orbitant autour de l'étoile.

Un panorama des méthodes de détection existantes et en développement est présenté sur la figure 1.1. Quelle que soit la méthode employée, des biais observationnels existent dans le recensement des différentes catégories d'exoplanètes et dans leur étude. Un des défis consiste à concevoir ou améliorer des techniques et instruments capables –conjointement avec les autres techniques– de détecter des planètes dont les caractéristiques physiques appartiennent à un espace des paramètres le plus grand possible.

<sup>1</sup>1 degré contient 60 minutes d'arc ( $1^\circ = 60'$ ) et 1 minute d'arc contient 60 secondes d'arc ( $1' = 60''$ ).

# Planet Detection Methods

Michael Perryman, Rep. Prog. Phys., 2000, 63, 1209 (updated 3 October 2007)

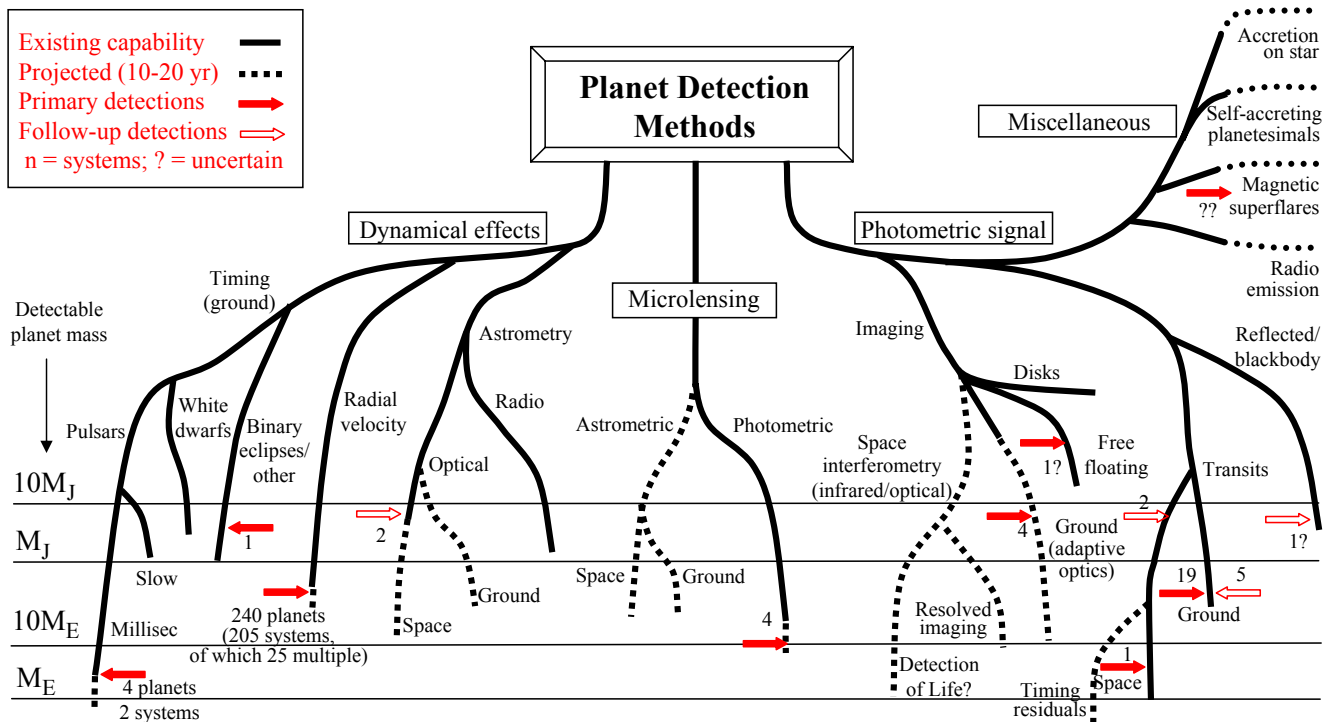


FIG. 1.1: Panorama des méthodes de détection d'exoplanètes, issu de Perryman [2007].

En particulier, afin de découvrir des planètes susceptibles d'héberger de la vie, il est *a priori* nécessaire de développer des méthodes capables de détecter des planètes peu massives et froides, ce qui correspond au cas le plus difficile quelle que soit la méthode de détection. Les planètes peu massives sont difficiles à détecter par la méthode de transit car une faible masse implique en général une faible taille et une faible vitesse de révolution (nous reviendrons sur le problème que cela pose), par la méthode de spectroscopie car l'effet gravitationnel de la planète sur l'étoile est alors faible et donc difficile à observer, et par les méthodes d'imagerie directe car une faible masse implique en général une faible luminosité.

En revanche, notons que les méthodes classiques de transit et de spectroscopie sont surtout capables de découvrir des exoplanète de faible période de rotation autour de l'étoile (période de révolution ou durée de l'année de cette planète), tandis que ce paramètre est presque totalement indifférent pour l'imagerie directe. En effet, la première classe de méthodes nécessite d'observer cette évolution temporelle tandis que la seconde requiert au contraire que pendant la durée des observations la planète n'ait pas bougé. Ainsi, puisque la détection d'exoplanètes n'a démarré qu'il y a une vingtaine d'années, la majorité des planètes détectées jusqu'ici par méthode de transit ou de spectroscopie ont une période de révolution inférieure à une année terrestre descendant jusqu'à des périodes de révolution de l'ordre de quelques jours, ce qui les rend assez différentes de la Terre qui constitue pourtant notre référence de *planète habitable*.

En ajoutant à ce type de considérations le fait que l'imagerie directe permet aussi une observation à la fois des planètes mais aussi de leur environnement (disque de poussière), on constate de façon simple que l'imagerie directe constitue une méthode instrumentale alternative certes difficile à mettre en œuvre mais aux potentialités de découverte importantes.

## Instrument SPHERE du Very Large Telescope au Chili

On peut distinguer les grandes familles de méthodes d'observation par imagerie directe selon qu'elle consiste d'un télescope suivie d'un coronographe ou qu'elle fait interférer la lumière de plusieurs télescopes, et selon qu'elle est située au sol ou dans l'espace. Nous nous attacherons ici à la méthodes d'imagerie directe par coronographie au sol. Ceci concerne notamment les deux grands projets instrumentaux GPI (*Gemini Planet Imager*, voir Macintosh et al. [2010]) et SPHERE (*Spectro-Polarimetric High-contrast Exoplanet Research*, voir Beuzit et al. [2006] et Wildi et al. [2009]) qui devraient générer leurs premiers jeux de données en 2011-2012. Nous nous focaliserons sur le futur instrument SPHERE qui sera monté sur un des télescopes de 8m de diamètre du VLT (*Very Large Telescope*, au Chili), tout en notant bien que les méthodes de traitement du signal développées dans cette thèse s'appliquent en fait de façon plus générale.

Plus spécifiquement, nous nous attacherons au module IRDIS (*Infra-Red Dual-beam Imaging and Spectroscopy*, décrit par Dohlen et al. [2008]) de SPHERE. Celui-ci a été récemment étudié dans la thèse de doctorat de Vigan [2009] dans la perspective de caractériser les potentialités de découverte d'exoplanètes de SPHERE. Vigan [2009] y présente ainsi une introduction complète à la fois aux enjeux astrophysiques, au concept et quantification instrumentaux et aux techniques standards de traitement du signal associés à ce module IRDIS. Ci-dessous nous résumons brièvement et simplement les concepts instrumentaux centraux pour cette thèse, partagés pour l'essentiel par tous les instruments permettant d'appliquer la technique de coronographie au sol.

Notons que dans la suite de la thèse, pour tester des méthodes de traitement de signal et étudier les potentialités de détection du module IRDIS, nous appliquerons nos techniques de traitement du signal sur des jeux de données simulés de façon réaliste. La section 2.1 présentera ainsi les spécifications quantitatives principales prévues pour l'instrument que nous étudions.

### • Correction de l'effet des perturbations de l'atmosphère terrestre par *optique adaptative*

Malgré ses avantages majeurs en termes de coût et de complexité de mise en œuvre, l'imagerie au sol pose un problème crucial par rapport à l'imagerie depuis l'espace : la lumière venant de l'étoile et de la planète traverse l'atmosphère de la Terre. A cause de ses fluctuations spatiales et temporelles de température et de pression, l'atmosphère terrestre perturbe notablement la trajectoire des rayons lumineux venant des objets célestes. Elle induit pour la même raison des effets de "trouble" tels que ceux que l'on observe parfois au dessus d'une route très chaude en été.

Ce "trouble" peut en partie être corrigé mécaniquement grâce à un *miroir déformable* : le rôle de ce miroir dont la forme change en permanence est de recevoir la lumière perturbée de l'objet céleste et de réfléchir cette lumière telle qu'on la verrait sans atmosphère, c'est-à-dire donnant une image bien nette de l'objet. Les mouvements infimes à induire sont provoqués par des actionneurs mécaniques qui répondent à des impulsions électriques commandées par des processeurs de calcul.

Les mouvements à imposer à chaque point  $P$  du miroir déformable sont calculés directement à partir de la direction locale de la lumière en  $P$ . Quelle que soit sa direction à l'instant  $t$ , elle doit repartir dans une direction fixée et bien définie pour chaque point  $P$ . Pour connaître les directions d'arrivée en chaque point du niveau du miroir déformable, il est nécessaire de les mesurer par un système optique recevant une partie de la lumière réfléchi par ce miroir. Ce dispositif capable de mesurer et calculer en partie la correction à appliquer au miroir déformable est appelé *analyseur de surface d'onde*<sup>2</sup>.

L'ensemble du système composé du miroir déformable et de l'analyseur de surface d'onde est appelé système d'*optique adaptative* (AO pour *Adaptive Optics*). Il a ainsi pour objectif de transformer un front d'onde "frappé" par l'atmosphère terrestre en un front d'onde plan. En France, le premier système d'AO dédié à l'astronomie a été mis au point à la fin des années 80. Diverses AO ont alors été progressivement

<sup>2</sup>Disons schématiquement que le front d'onde est la surface localement perpendiculaire en le point  $P$  au "rayon lumineux" passant par ce point  $P$ .

installées dans les plus grands télescopes du monde. En parallèle, de nouvelles générations d'AO ont vu le jour. En particulier, la recherche d'exoplanètes requiert des systèmes d'AO dits *extrêmes* dont les capacités de correction de la turbulence atmosphérique se rapprochent de la perfection. Ceux-ci sont ainsi toujours en cours de conception et réalisation. Plus de détails seront donnés dans la section 5.1 et des informations sur l'AO extrême de SPHERE (SAXO pour *Sphere Adaptive optics for eXoplanet Observation*) sont données par Wildi et al. [2009].

- **Affaiblissement de la lumière diffractée de l'étoile par coronographie**

Même dans le cas parfait sans atmosphère, le front d'onde d'une étoile n'est en réalité pas plan au niveau du miroir déformable. L'ouverture du télescope induit en effet un effet de *diffraction* qui renvoie la lumière de l'étoile dans diverses directions. On observerait ainsi sur une caméra d'acquisition munie de beaucoup de pixels que l'étoile est non un point mais une tache (la *tache d'Airy*) circulaire entourée d'anneaux concentriques de moins en moins intenses à mesure que leur distance au centre augmente. La taille de la tache et la distance entre les anneaux est d'autant plus grande que l'ouverture du télescope est petite.

A cause de l'effet de diffraction, sur la caméra d'acquisition même avec une AO parfaite on ne verrait donc pas un point mais une tache centrale et des anneaux. Ceci n'est pas gênant si l'on observe des étoiles un peu espacées les unes des autres, mais est en revanche une limitation fondamentale à la détection d'une exoplanète car celle-ci est très proche de son étoile parente, et est par conséquent noyée dans la tache de diffraction.

La coronographie est une méthode instrumentale destinée à "éteindre" l'étoile partout dans le champ, et donc en particulier là où se trouve de potentielles exoplanètes. Ceci est nécessaire car les exoplanètes sont très peu lumineuses par rapport à l'étoile (contraste de l'ordre du million comme nous l'avons précisé) et ne sont donc pas visibles lorsqu'elles sont superposées aux anneaux.

- **Acquisition simultanée d'images dans deux bandes spectrales très proches**

Malgré le développement d'un système d'AO et d'un coronographe performant, la détection d'exoplanètes dans les images reste difficile. Dans le module IRDIS de SPHERE, le mode d'observation DBI (*Double Band Imaging*) permettra d'observer simultanément, grâce à une lame séparant le faisceau en deux, dans deux domaines de longueurs d'onde différents. En l'occurrence, comme l'étoile et la planète sont physiquement des objets très différents – ayant en particulier des températures de surface très différentes – il est en général possible de les discriminer par leurs différences spectrales. Il s'avère ainsi qu'en infra-rouge, autour de la longueur d'onde de  $1.65\mu\text{m}$  pour les bandes H2 et H3 de référence de SPHERE, le spectre d'une étoile typique ne présente pas de bande d'absorption, tandis que la plupart des planètes d'intérêt présente une raie d'absorption dans une bande et pas dans l'autre. Ainsi, soustraire les images obtenues simultanément dans les deux bandes fait disparaître – en partie – l'étoile et non la planète. La différence entre les aberrations optiques dans une bande et dans l'autre (contribution de l'étoile) constitue la limitation centrale de la méthode simple de soustraction.

- **Effet de rotation de champ dû à la monture du VLT et à sa configuration pour l'instrument SPHERE**

La dernière caractéristique instrumentale fondamentale de SPHERE pour le problème de détection d'exoplanète est l'effet dit de *rotation de champ* : sur les images, la planète semble tourner circulairement autour de l'étoile à mesure que les heures passent. Il ne s'agit pas de la révolution de la planète autour de son étoile parente (on suppose que ce mouvement physique est négligeable en l'espace de quelques heures) mais d'un effet de projection dû à la monture du télescope et à sa configuration.

Afin de pouvoir tourner tout en étant posé au sol de façon rigide, un télescope doit être fixé à une *monture*. Il en existe de deux grands types : la monture alt-azimuthale qui permet de faire tourner le



télescope selon les axes horizontal et vertical et la monture équatoriale qui permet de le faire tourner selon l'axe parallèle à l'axe de rotation de la Terre (le seul axe qui tourne lorsque le télescope suit une étoile dans le ciel) et selon un axe orthogonal. Illustrons l'effet de rotation de champ par l'exemple d'une longue vue fixée à une monture alt-azimuthale. Si l'on fixe un capteur carré derrière la longue avec deux de ses côtés horizontaux, ceux-ci resteront horizontaux quelle que soit la direction de la longue vue. Si la longue vue suit une étoile dans le ciel, dans le référentiel du capteur le champ d'objets célestes situé autour de l'étoile semble alors tourner autour de l'étoile pointée. En particulier, si l'étoile est accompagnée d'une planète cette dernière semble décrire un cercle autour de l'étoile.

Le VLT est fixé à une monture alt-azimuthale et subit aussi un effet de rotation de champ même si celui-ci est légèrement différent (Avila and Wirenstrand [2005]). Grâce à un dérotateur de champ on se placera en particulier en mode dit *pupille stabilisée* dans lequel sur le capteur de la caméra la planète semble tourner autour de l'étoile. Plus spécifiquement, la planète semble se déplacer en cercle au sein d'aberrations optiques fixes dues à l'étoile. Cet effet de rotation est un effet de géométrie sphérique déterministe qui est connu dès que l'on connaît la position du VLT sur Terre, la position absolue de l'étoile visée et la date et l'heure de la nuit d'observation. On peut alors utiliser cet effet à nouveau par des différentielles d'images en soustrayant des images acquises à des moments suffisamment différents pour que la planète ait tourné, de sorte que la contribution de l'étoile se simplifie en grande partie mais pas celle de la planète.

## Techniques actuelles de traitement de données

En présentant les deux dernières caractéristiques de l'instrument SPHERE nous constatons que le développement de concepts instrumentaux va de pair avec un minimum de développement de technique de traitement des données. En l'occurrence, les idées de traitement consistant à appliquer des différences d'images afin de "réhausser le signal" ont déjà été testées et validées. La méthode de soustraction d'images de bande spectrale voisine (méthode dite SDI pour *Spectral Differential Imaging*), utilisée pour concevoir le mode DBI d'IRDIS, avait été proposé par Racine et al. [1999], et la méthode de soustraction d'images où la planète a changé d'angle (méthode dite ADI pour *Angular Differential Imaging*) avait été proposée par Marois et al. [2006]. Utilisées en général conjointement, ce sont là les deux techniques les plus standard de traitement pour des données du type de celles prévues pour notre utilisation de SPHERE.

Cependant, il n'est pas seulement nécessaire de "réhausser le signal", il convient aussi de finaliser la procédure de détection en parvenant au résultat binaire "il y a une planète" ou "il n'y a pas de planète" et d'associer un indicateur de confiance à attribuer au résultat. Pour cela, il est nécessaire de modéliser statistiquement les données. Certains modèles fins ont déjà été proposés comme nous le verrons par la suite. Dans un premier temps, voyons plus en détail en quoi consiste le développement de méthodes de traitement du signal à partir d'une modélisation statistique des données.

## 1.2 Développement de méthodes d'inférence statistiques pour le traitement du signal

Cette partie a essentiellement pour objectif de donner une vue d'ensemble simplifiée et rapide de la philosophie et de la procédure permettant d'utiliser la théorie des statistiques et probabilité comme outil pour développer des techniques de traitement du signal. Pour être simple et concrète, elle s'appuie tout du long sur l'exemple du problème astrophysique introduit dans le chapitre précédent et présenté dans Smith et al. [2008].

Pour un traitement plus exhaustif, on pourra commencer par Bickel and Doksum [2001] pour ce qui concerne les inférences statistiques, Billingsley [1995] pour un traitement plus rigoureux de la théorie des probabilités et de la mesure sous-jacente aux inférences statistiques ou Casella and Berger [2002] pour un traitement conjoint des inférences statistiques et de la théorie des probabilités. Petersen and Pedersen [2007] présentent succinctement des résultats de base du calcul matriciel.

### Modèle statistique : modèle probabiliste des données

Détecter une planète ou estimer son intensité et sa position à partir d'un jeu d'images sont des problèmes d'inférence statistique : le traiteur de signal dispose d'un ensemble d'observations stockées dans une variable numérique  $\mathbf{x}$  et d'informations sur la façon dont ces données ont été obtenues, et il souhaite obtenir des informations sur des paramètres sous-jacents à la formation de ces données. On appelle *modèle direct* le modèle qui associe aux paramètres d'intérêt les observations  $\mathbf{x}$ . En tant que modélisation physique de la nature, celui-ci a en général une composante déterministe et une composante aléatoire. Dans un premier temps, un *traiteur de signal* suppose ainsi que les données  $\mathbf{x}$  sont l'expression d'une variable aléatoire  $\mathbf{X}$  décrit par une distribution de probabilité.

Classiquement on définit cette distribution de probabilité par sa *densité*  $p(\cdot)$  dans le cas de données  $\mathbf{x}$  continues ( $\mathbf{x}$  continue signifie  $\mathbf{x}$  à valeur dans un espace continu,  $\mathcal{X} = \mathbb{R}^n$  par exemple) ou par sa *masse*  $p(\cdot)$  dans le cas de données  $\mathbf{x}$  discrètes ( $\mathbf{x}$  à valeurs dans un espace discret,  $\mathcal{X} = \{0, 2, 3, 5\}^n$  par exemple). Les probabilités pour que les données mesurées appartiennent à un certain ensemble  $S \subset \mathcal{X}$  de valeurs sont ainsi données par les relations

$$\Pr(\mathbf{X} \in S) = \int_S d\mathbf{x} p(\mathbf{x}) \quad \text{si } \mathbf{x} \text{ est continue}$$

$$\Pr(\mathbf{X} \in S) = \sum_{\mathbf{x} \in S} p(\mathbf{x}) \quad \text{si } \mathbf{x} \text{ est discret}$$

Dans la suite on se restreindra à  $\mathbf{x}$  continue et sauf mention contraire la mesure de base, notée  $d\mathbf{x}$  pour l'espace  $\mathcal{X}$ , sera alors la mesure de Lebesgue.

L'objet de l'étude statistique consiste à utiliser les données mesurées  $\mathbf{x}$  pour estimer tout ou partie de la distribution de probabilité  $p(\cdot)$ . Dans la suite, on supposera que la distribution de probabilité est connue à un ensemble de paramètres  $\boldsymbol{\theta} \in \Theta$  près : les inférences sont dites *paramétriques*. On définit ici la *vraisemblance* du paramètre  $\boldsymbol{\theta}$  comme la densité de probabilité d'effectivement observer les données  $\mathbf{x}$  pour ce paramètre  $\boldsymbol{\theta}$  et on la note ici en général sous la forme d'une densité de probabilité conditionnelle :  $p_{\boldsymbol{\theta}}(\mathbf{x}) = p(\mathbf{x}|\boldsymbol{\theta})$ .

**Exemple 1** Dans le cas de la détection d'exoplanètes à partir d'un jeu de  $K$  images par exemple, les données  $\mathbf{x} = \{\mathbf{x}(1), \dots, \mathbf{x}(K)\}$  se décomposent en  $K$  images constituées chacune de  $L$  pixels<sup>3</sup>. Le modèle direct sous-jacent le plus standard consiste à supposer que les images sont statistiquement indépendantes dans le temps, c'est-à-dire que la densité de probabilité conjointe des images  $\mathbf{x}(k)$  et  $\mathbf{x}(k')$  est égale au produit des densités de probabilité de  $\mathbf{x}(k)$  et de  $\mathbf{x}(k')$  pour  $k \neq k'$ , et que la densité de probabilité de l'image  $\mathbf{x}(k)$  est totalement définie à partir du paramètre  $\boldsymbol{\theta} = \{\mathbf{d}, \alpha, r, \mathbf{p}(k, r), \sigma\}$  (éventuellement supposé partiellement connu). On suppose ainsi qu'une image est la somme d'une combinaison déterministe et d'un bruit aléatoire  $\boldsymbol{\epsilon}(k)$  gaussien :

$$\mathbf{x}(k) = \mathbf{d} + \alpha \mathbf{p}(k, r) + \boldsymbol{\epsilon}(k) \quad \boldsymbol{\epsilon}(k)|\sigma \sim \mathcal{N}_L(0, \sigma^2 I_L)$$

$\mathbf{d}$  est l'image de l'étoile à travers le système optique (vecteur de taille  $L \times 1$ ),  $\mathbf{p}(k, r)$  est le profil de la planète pour la pose  $k$  pour une position initiale  $r$ ,  $\alpha$  est l'intensité scalaire de la planète et  $\boldsymbol{\epsilon}(k)$  est la réalisation d'un bruit aléatoire gaussien de moyenne nulle, indépendant spatialement et de variance  $\sigma^2$  égale pour tout pixel. Plus synthétiquement on écrit  $\mathbf{x}(k)|\boldsymbol{\theta} \sim \mathcal{N}_L(\mathbf{d} + \alpha \mathbf{p}(k, r), \sigma^2 I_L)$  i.e.

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{k=1}^K p(\mathbf{x}(k)|\boldsymbol{\theta}) \quad \text{avec } p(\mathbf{x}(k)|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{L}{2}} \sigma^L} \exp\left(-\frac{\|\mathbf{x}(k) - \mathbf{d} - \alpha \mathbf{p}(k, r)\|^2}{2\sigma^2}\right)$$

<sup>3</sup>Comme souvent en traitement d'image, les colonnes du détecteur (i.e. de chaque image) sont concaténées de telle sorte que  $\mathbf{x}(k)$  est représenté par un vecteur de taille  $L \times 1$ .  $\mathbf{x}$  peut alors en fait se construire sous la forme d'un vecteur de taille  $KL \times 1$ , voir la section de préambule.

On note  $\mathcal{X}$  le domaine d'appartenance des données  $\mathbf{x} \in \mathcal{X}$  et  $\Theta$  le domaine d'appartenance des paramètres  $\boldsymbol{\theta} \in \Theta$ , de telle sorte qu'un modèle paramétrique est totalement défini par  $\mathcal{P}_{\boldsymbol{\theta}} = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  où  $p(\cdot|\boldsymbol{\theta})$  est la distribution, supposée connue à  $\boldsymbol{\theta}$  près, de la variable aléatoire  $\mathbf{X}$ . Lorsque deux variables aléatoires  $\mathbf{X}$  et  $\mathbf{Y}$  sont en cours d'étude, on note  $p_{\mathbf{X}}(\cdot|\boldsymbol{\theta})$  et  $p_{\mathbf{Y}}(\cdot|\boldsymbol{\theta}')$  les distributions correspondant à chacune de ces variables aléatoires. Notons de plus qu'à ce stade il n'est pas nécessaire d'utiliser la notion de probabilité conditionnelle mais elle sera nécessaire plus loin donc nous utilisons d'emblée cette "notation".

## Deux grands types d'inférence : l'estimation ponctuelle et le test d'hypothèses

Dans un cadre paramétrique, une inférence consiste à appliquer une fonction qui aux données associe certaines "grandeurs d'intérêt" définies à partir de l'espace des paramètres  $\Theta$

- estimation ponctuelle : estimation d'une valeur  $\hat{\boldsymbol{\theta}}(\mathbf{x}) \in \Theta$
- estimation d'un intervalle de confiance  $C(\mathbf{x}) \subset \Theta$
- test d'hypothèses : réjection ou non d'une hypothèse telle que  $\boldsymbol{\theta} \in \Theta_0$  avec  $\Theta_0 \subset \Theta$
- sélection de modèle : choix d'une hypothèse parmi plusieurs

**Exemple 2** Dans l'exemple 1, diverses inférences peuvent être faites sur  $\boldsymbol{\theta}$ . En particulier, l'estimation de l'intensité et de la position de la planète consiste à estimer  $\alpha \in \mathbb{R}_+$  et  $\mathbf{r} \in [1, \sqrt{L}]^2$  par exemple, tandis que la détection de la planète consiste à décider si l'hypothèse nulle  $H_0 : \alpha = 0$  (pas de planète) est vraie ou fausse.

Il est fortement recommandé d'accompagner toute inférence d'une ou plusieurs grandeurs donnant une idée de la confiance à accorder à l'inférence. Ainsi, une "barre d'erreur" donne une idée de l'incertitude pour une estimation ponctuelle, une "probabilité d'erreur" donne une idée de la confiance à accorder à la décision pour un test d'hypothèses *etc.* Ce problème constitue aussi une inférence.

## Techniques de construction d'une procédure d'inférence

### • Construction de la procédure d'inférence

Un bagage classique de techniques est à la disposition du statisticien pour associer une procédure d'inférence à un problème d'inférence et un modèle statistique donnés (et éventuellement à d'autres ingrédients supplémentaires comme on le verra avec le paradigme bayésien). Voir Bickel and Doksum [2001], Casella and Berger [2002] pour une description plus exhaustive de telles méthodes.

Pour l'estimation ponctuelle, une méthode classique appelée **maximum de vraisemblance** (ML) consiste à estimer le paramètre  $\boldsymbol{\theta}$  par la valeur de  $\boldsymbol{\theta}$  qui maximise la probabilité (ou densité de probabilité) des données  $\mathbf{x}$ .

**Exemple 3** Dans notre exemple 2, si l'on souhaite **estimer** l'intensité  $\alpha$  et la position  $\mathbf{r}$  de la planète en supposant connus les profils  $\mathbf{p}(k, r)$  pour un  $r$  fixé, il n'est pas suffisant de dire que l'on choisit pour  $(\hat{\alpha}(\mathbf{x}), \hat{\mathbf{r}}(\mathbf{x}))$  l'estimateur ML car il s'agirait alors d'estimer  $\alpha$  et  $\mathbf{r}$  par les valeurs qui maximisent la vraisemblance  $p(\mathbf{x}|\alpha, r, \mathbf{d}, \sigma)$  alors que  $\mathbf{d}$  et  $\sigma$  sont inconnus. On appelle  $\mathbf{d}$  et  $\sigma$  les paramètres de nuisance par opposition aux paramètres d'intérêt  $\alpha$  et  $\mathbf{r}$ . C'est donc bien l'ensemble des paramètres qui sont estimés par méthode ML : on sélectionne **conjointement** l'ensemble du paramètre  $\boldsymbol{\theta}$  le plus vraisemblable, c'est-à-dire  $\boldsymbol{\theta}$  qui maximise la (densité de) probabilité des données  $\mathbf{x}$ . Si seules les valeurs  $(\hat{\alpha}(\mathbf{x}), \hat{\mathbf{r}}(\mathbf{x}))$  sont souhaitées on écrit simplement<sup>4</sup> :

$$(\hat{\alpha}_{ML}(\mathbf{x}), \hat{\mathbf{r}}_{ML}(\mathbf{x})) = \arg_{\alpha, \mathbf{r}} \max_{\alpha, r, \mathbf{d}, \sigma} p(\mathbf{x}|\alpha, r, \mathbf{d}, \sigma)$$

Avec le modèle partiellement linéaire de l'exemple 1, la solution  $(\hat{\alpha}_{ML}(\mathbf{x}), \hat{\mathbf{r}}_{ML}(\mathbf{x}))$  est une fonction des données en partie explicite, donnée dans Smith et al. [2008] et qui sera démontrée dans la section 3.2.

<sup>4</sup>Cf la section de préambule pour la notion d'arg maximum.

Pour tester une hypothèse, il est usuel de seuiller une statistique (*statistique* : fonction des données). On l'appelle "statistique de test" et on note en général  $\zeta$  le seuil du test.

**Exemple 4** Afin de **décider si une planète est présente ou non** (test de détection), un test simple à appliquer et intuitif consiste à décider qu'il y a une planète si l'intensité estimée de la planète est supérieure à un certain seuil :

$$\text{Rejeter } H_0 \text{ si } \hat{\alpha}_{ML}(\mathbf{x}) > \zeta \quad (1.1)$$

#### • Incertitude sur l'inférence

Une inférence seule est en général insuffisante : il convient d'obtenir des informations sur la qualité de l'inférence et donc la confiance qu'on peut attribuer à cette inférence. Ceci peut être fait

- *a priori* (*predata*), *i.e.* sans introduire de jeu de données observé. Il s'agit alors de considérer l'erreur moyenne d'une inférence (en intégrant sur toutes les données possibles, ou plus précisément sur un espace qui ne dépend pas des données effectivement mesurées), avant d'avoir des données, et suppose les paramètres vrais connus.
- *a posteriori* (*postdata*), *i.e.* à partir du jeu de données effectivement mesuré. Il s'agit alors d'estimer l'incertitude d'une inférence effectuée sur le jeu de données mesuré.<sup>5</sup>

**Estimation ponctuelle** Pour l'estimation ponctuelle, deux *erreurs predata* sont classiquement calculées : le **biais** (moyenne de la différence entre l'estimation et la valeur vraie) et l'**erreur quadratique moyenne** (EQM, moyenne de la différence au carré entre l'estimation et la valeur vraie). Ces deux grandeurs sont complémentaires car le biais indique si en moyenne l'estimateur indique la valeur vraie, en laissant une forte sous-estimation et une forte sur-estimation se compenser, tandis que dans l'EQM les grandes erreurs même de signes opposés contribuent à augmenter significativement l'erreur moyenne : il n'y a pas de compensation possible puisque l'on ne somme que des valeurs positives.

**Exemple 5** Ainsi, le biais et l'EQM de l'estimateur d'intensité sont définis par :

$$\begin{aligned} \text{biais} &= \mathbf{E}_{\theta}[\hat{\alpha}_{ML}(\mathbf{x})] - \alpha \\ \text{EQM} &= \mathbf{E}_{\theta}[(\hat{\alpha}_{ML}(\mathbf{x}) - \alpha)^2] = \text{var}_{\theta}[\hat{\alpha}_{ML}(\mathbf{x})] + \text{biais}^2 \end{aligned}$$

où  $\mathbf{E}_{\theta}$  signifie que la moyenne est calculée à partir de la distribution de  $\mathbf{x}$  de paramètre  $\theta$  supposé connu (voir la section de préambule). On rappelle que  $\theta$  représente ici  $(\alpha, \dots, \sigma)$ , voir l'exemple 1.

On souhaite d'un estimateur qu'en moyenne il estime correctement le paramètre estimé, *i.e.* que son biais soit nul et que son EQM soit faible (compromis entre les deux à choisir).

Si le biais d'un premier estimateur est une fonction affine du paramètre, on construit aisément à partir de cet estimateur un estimateur non biaisé. Dans ce cas, le biais est nul et l'EQM se réduit à la variance de l'estimateur. Cependant, même dans ce cas le paramètre ayant mené aux données est inconnu donc il est en principe impossible de calculer cette erreur quadratique afin d'avoir une idée en pratique de l'erreur faite lors de l'estimation. Il est cependant possible d'estimer, *postdata*, les erreurs *predata* précédentes en remplaçant le paramètre inconnu  $\theta$  par son estimateur.

**Exemple 6** On retranscrit alors classiquement le résultat de l'estimation d'intensité sous la forme d'une estimation nuancée par une barre d'erreur :

$$\begin{aligned} & \text{“ } \alpha = \hat{\alpha}_{ML}(\mathbf{x}) \pm \text{Err}(\mathbf{x}) \text{ ”} \\ & \text{avec } \text{Err}(\mathbf{x}) = \sqrt{\text{var}_{\hat{\theta}_{ML}(\mathbf{x})}[\hat{\alpha}_{ML}(\mathbf{x})]} \end{aligned}$$

<sup>5</sup>Notons qu'en *postdata* il est en plus possible de prédire l'erreur moyenne à attendre d'un prochain jeu de données. Nous ne considérerons cependant pas ici ce problème de prédiction.

où on rappelle que la variance est ici calculée en pratique en remplaçant, après avoir fait le calcul de variance,  $\theta$  par  $\hat{\theta}_{ML}(\mathbf{x})$ .

Idem pour une erreur sur la position : si  $r_1$  et  $r_2$  sont les numéros initiaux de ligne et colonne de la planète, une erreur serait par exemple donnée par la matrice de covariance sur l'estimation de  $(r_1, r_2)$ . La matrice de covariance recouvre la même idée que la variance mais pour un paramètre à plusieurs composantes.

**L'erreur moyenne : un critère de choix d'une inférence** On constate dès le problème d'estimation ponctuelle qu'étudier l'incertitude de plusieurs inférences permet d'une part d'estimer l'erreur faite sur une estimation donnée mais aussi en amont de choisir la procédure qui aura les erreurs les plus basses possibles (étant donné le compromis choisi par l'utilisateur). Selon les qualités considérées comme les plus importantes pour l'étude menée, on privilégiera ainsi une technique plutôt qu'une autre. Notons que si la seule propriété attendue de l'inférence est de minimiser en moyenne une erreur, il est plus généralement possible de se placer dans le cadre de la théorie de la décision et de déterminer conjointement la procédure et son erreur (voir l'annexe B). Mais en général, des propriétés pratiques de l'application de la procédure, telles qu'un temps de calcul réduit, sont aussi souhaitées.

**Test d'hypothèse** Le lien *predata* entre “choix de la procédure d'inférence” et “confiance à accorder à l'inférence” peut être partiellement calibré pour le test d'hypothèses : le seuil  $\zeta$  de l'exemple 1.1 est en général choisi en fixant la probabilité d'erreur de type I, *i.e.* la probabilité de rejeter  $H_0$  alors que les données  $\mathbf{x}$  ont été générées sous  $H_0$ . On l'appellera Probabilité de Fausse Alarme (PFA) dans la suite.

**Exemple 7** Dans l'exemple 3, la confiance a priori à accorder à la détection ou non détection de planète est typiquement indiquée par PFA (erreur de type I), la probabilité de détecter une planète alors qu'il n'y en a en réalité pas :

$$PFA(\zeta) = Pr(\hat{\alpha}_{ML}(\mathbf{x}) > \zeta | H_0) \quad (1.2)$$

Cet indicateur peut en fait être directement utilisé pour définir la procédure d'inférence : il suffit pour cela de fixer une probabilité d'erreur souhaitée  $p$  (typiquement  $p = 5\%$ , *i.e.* on s'autorise 5 fois sur 100 à détecter une planète alors qu'il n'y en a pas) et d'en déduire par relation inverse le seuil  $\zeta = PFA^{-1}(p)$  associé :

$$\text{Rejeter } H_0 \text{ si } \hat{\alpha}_{ML}(\mathbf{x}) > PFA^{-1}(p)$$

Notons que la PFA peut être choisie aussi petite qu'on le souhaite mais alors la probabilité d'erreur de type II pour une intensité de planète donnée peut être très proche de 1. La probabilité d'erreur de type II est la probabilité de ne pas rejeter  $H_0$  alors qu'en fait il y avait un signal d'intensité  $\alpha > 0$ . Ainsi, si dans l'exemple précédent on prend un seuil  $\zeta$  très grand, aucune planète ne sera détectée. Donc dans les cas où il n'y a pas de planète il n'y a effectivement pas d'erreur (PFA infime), mais dans les cas où une planète d'intensité  $\alpha$  donnée pas trop élevée est présente la procédure de détection continue à rarement détecter et fera donc souvent une erreur. Il peut donc effectivement être intéressant de fixer une certaine PFA telle que  $p = 5\%$ , puis de constater l'erreur de type II associée, ou de façon équivalente la Probabilité de bonne Détection (PD) définie comme 1 moins la probabilité d'erreur de type II.

**Exemple 8** Dans l'exemple 3, la probabilité de bien détecter une planète d'intensité  $\alpha$  est donnée par PD définie par

$$PD(\zeta, \alpha) = Pr(\hat{\alpha}_{ML}(\mathbf{x}) > \zeta | H_1, \alpha)$$

PD est simplement égal à 1 moins la probabilité de ne pas détecter une planète d'intensité  $\alpha$  alors qu'elle est présente (erreur de type II).

$\alpha$  est inconnue et à nouveau nous souhaiterions utiliser les données  $\mathbf{x}$  mesurées en pratique afin de donner une estimation d'erreur effectuée lors du test de détection. Une mesure *postdata* de probabilité d'erreur analogue à l'estimation d'erreur en estimation ponctuelle consisterait à injecter dans PD une estimation du paramètre vrai. Cependant, d'autres techniques bénéficient de propriétés plus attractives.

**Exemple 9** *Pour le test d'hypothèses, une mesure postdata de confiance à accorder à l'inférence apparaît naturellement dans le calcul de probabilité : plutôt que de donner que le résultat binaire “ $H_0$  rejetée ou non” et la valeur  $p$  choisie comme PFA ( $p = 5\%$  par ex.), il serait possible de donner (en plus) la valeur de la statistique seuillée  $\hat{\alpha}_{ML}(\mathbf{x})$  dont on peut voir si elle est très éloignée ou non du seuil  $\zeta$ , à condition de donner en plus le seuil  $\zeta$ . Cependant ce que l'on souhaite en général fixer est une PFA (pourcentage de planètes détectées alors qu'il n'y en avait pas), non  $\zeta$  qui n'a pas une interprétation aussi directe que la PFA. Comme précisé au dessus,  $\zeta$  était donc en général fixé par relation inverse de l'Eq. (1.2) :  $\zeta = PFA^{-1}(p)$  avec  $p = 5\%$  par ex. Donc plutôt que comparer  $\hat{\alpha}_{ML}(\mathbf{x})$  à  $\zeta$ , on préfère de même comparer “ $PFA(\hat{\alpha}_{ML}(\mathbf{x}))$ ” (PFA observée) à  $PFA(\zeta) = p$ . On appelle cette “PFA observée” la  $p$ -value. Plus précisément la  $p$ -value est ici définie comme la probabilité (sous  $H_0$ ) pour que l'estimation d'intensité  $\hat{\alpha}_{ML}(\mathbf{X})$  soit plus grande que l'estimation d'intensité estimée en pratique  $\hat{\alpha}_{ML}(\mathbf{x})$  :*

$$p_{val}(\hat{\alpha}_{ML}(\mathbf{x})) = Pr(\hat{\alpha}_{ML}(\mathbf{X}) > \hat{\alpha}_{ML}(\mathbf{x}) | H_0)$$

Comparer la  $p$ -value à  $p$  est de plus équivalent à comparer la statistique seuillée initiale  $\hat{\alpha}_{ML}(\mathbf{x})$  à  $\zeta$ .

Nous reverrons la  $p$ -value par la suite.

Pour les inférences *postdata*, présentons maintenant le paradigme bayésien dont l'objet est précisément de donner la distribution de probabilité “*a posteriori*”, *i.e.* connaissant les données  $\mathbf{x}$ , du paramètre d'intérêt.

## Introduction de connaissances sur les paramètres grâce au paradigme bayésien

### • Loi *a posteriori* de $\theta$

On appelle *fréquentistes* des inférences considérant toujours  $\theta$  comme un paramètre fixé, n'ayant donc pas de distribution de probabilité *a priori*. C'est le cas des inférences considérées jusqu'ici. Alternative-ment, des informations complémentaires peuvent être introduites sur le paramètre  $\theta$  via des contraintes déterministes et finalement le choix d'une distribution de probabilité *a priori*  $\pi(\theta)$ . Les données observées  $\mathbf{x}$  sont alors vues comme la variable fixée, connue, et  $\theta$  comme la variable inconnue et dont la connaissance est décrite par une distribution de probabilité. La distribution “*a posteriori*” de  $\theta$  conditionnelle aux données  $\mathbf{x}$  est donnée grâce au théorème de Bayes. Si  $\theta$  est une variable aléatoire continue, sa densité *a posteriori* est ainsi donnée par :

$$\pi(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} d\theta p(\mathbf{x}|\theta)\pi(\theta)} \quad \forall \theta \in \Theta \quad (1.3)$$

et l'intégrale est remplacée par une somme discrète si  $\theta$  est une variable aléatoire discrète. On appellera *bayésiennes* des inférences basées sur ces distributions *a posteriori*. Berger [1985] et Robert [2007] sont des ouvrages introductifs de référence pour l'approche bayésienne.

Deux grandes philosophies s'opposent quant à l'interprétation à attribuer à la distribution (*a priori* ou *a posteriori*) de  $\theta$ . Visiblement bien adaptée au problème astrophysique que nous traitons, celle que nous adoptons ici est une approche classique (adoptée notamment par Laplace, Bayes et Freedman), qui consiste schématiquement à supposer qu'il existe un paramètre sous-jacent  $\theta^*$  sur lequel on effectue des inférences grâce à un ensemble d'observations  $\mathbf{x}$ . La distribution *a priori* contient ainsi l'information que l'on souhaite prendre en compte concernant le paramètre d'intérêt  $\theta$  et la distribution *a posteriori* obtenue par l'Eq. (1.3) traduit notre connaissance *actualisée* du paramètre inconnu  $\theta$ .

- **Inférences bayésiennes classiques**

La distribution de probabilité *a posteriori* du paramètre  $\theta$  contient en général pour un traicteur de signaux toute l'information nécessaire sur  $\theta$  puisque cette loi de probabilité de  $\theta$  est conditionnelle à toute l'information dont il dispose. S'il attend plus synthétiquement des estimations ponctuelles associées "d'erreurs" estimées après acquisition des données  $\mathbf{x}$ , un estimateur ponctuel classique de  $\theta$  est sa moyenne (qui minimise l'erreur quadratique moyenne –moyenne sur  $\theta$  maintenant– et est donc notée MMSE) et "l'erreur" bayésienne de l'inférence est indiquée par la matrice de covariance de la distribution *a posteriori*. En particulier, si  $\theta$  est scalaire,

$$\hat{\theta}_{\text{MMSE}}(\mathbf{x}) = \mathbf{E}[\theta|\mathbf{x}]$$

$$E_B(\mathbf{x}) = \sqrt{\text{var}[\theta|\mathbf{x}]}$$

avec

$$\mathbf{E}[f(\theta)|\mathbf{x}] = \int_{\Theta} d\theta f(\theta) \pi(\theta|\mathbf{x})$$

$$\text{et } \text{var}[\theta|\mathbf{x}] = \mathbf{E}\left[\left(\theta - \mathbf{E}[\theta|\mathbf{x}]\right)^2 \middle| \mathbf{x}\right]$$

On peut alors, comme en fréquentiste, toujours indiquer une estimation de  $\theta$  sous la forme d'une barre d'erreur :

$$"\theta = \hat{\theta}_{\text{MMSE}}(\mathbf{x}) \pm E_B(\mathbf{x})"$$

Il est par ailleurs possible de calculer la probabilité *a posteriori* pour que le paramètre appartienne à un certain sous-espace<sup>6</sup>  $\Theta_0 \subset \Theta$  : la probabilité pour que  $\theta \in \Theta_0$  connaissant les données  $\mathbf{x}$  vaut simplement

$$\Pr(\theta \in \Theta_0|\mathbf{x}) = \int_{\Theta_0} d\theta \pi(\theta|\mathbf{x})$$

On peut ainsi définir l'événement  $\theta \in \Theta_0$  comme l'hypothèse nulle  $H_0$ . Dans ce cas, la probabilité précédente est un outil possible de test d'hypothèses car il se réécrit  $\Pr(\theta \in \Theta_0|\mathbf{x}) = \Pr(H_0|\mathbf{x})$ . Si on attend une information plus synthétique, du type *rejet* ou *acceptation d'hypothèse*, on peut par exemple seuiller cette probabilité<sup>7</sup>.

- **Probabilité *a priori*  $\pi(\theta)$**

L'outillage bayésien apparaît ainsi comme un outil naturel lorsque des informations sont effectivement disponibles. Et lorsque peu ou pas d'information est connue *a priori* sur le paramètre  $\theta$ , la notion "d'absence d'information" doit être précisée selon le cas étudié. Les lois *a priori non informatives* classiques sont décrites dans Robert [2007] par exemple.

Plus généralement, le passage d'une distribution de probabilité *a priori* de  $\theta$  (quelconque mais bien précisée) à une distribution de probabilité *a posteriori* de  $\theta$  en utilisant un ensemble d'observations  $\mathbf{x}$

<sup>6</sup>Notons que l'espace sur lequel est intégré la distribution *a posteriori* peut si on le souhaite dépendre de  $\mathbf{x}$ . En considérant  $\mathbf{x}$  comme une valeur fixée et non comme une variable aléatoire, il est même possible de calculer la probabilité *a priori* pour que  $\theta$  appartienne à un certain espace dépendant de  $\mathbf{x}$ .

<sup>7</sup>Notons cependant que lorsque  $\Theta_0 = \{\theta_0\}$  (comme dans notre cas de détection où l'intensité de la planète est nulle sous  $H_0 : \theta_0 = \mathbf{0}$ ) alors l'*a priori*  $\pi(\theta)$  requiert une composante discrète en  $\theta = \theta_0$  car sinon  $\Pr(\theta = \theta_0 | \mathbf{x}) = 0$ . Ce cas, problématique, sera détaillé plus loin. Plus généralement, l'objet d'une partie de cette thèse consiste à étudier une approche alternative à l'approche du test d'hypothèses consistant à comparer deux hypothèses en comparant leurs probabilités *a posteriori*.

qui traduit la vraisemblance vraie  $p(\mathbf{x}|\boldsymbol{\theta}^*)$  sous-jacente permet de quantifier mathématiquement l'actualisation de la distribution *a priori* par la vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta})$ .

Quoiqu'il en soit, il est important de toujours adopter une interprétation en considérant attentivement le cas étudié. Le sens à donner à la distribution *a priori* de  $\boldsymbol{\theta}$  dépend des informations conditionnelles que l'on souhaite inclure. Sous forme explicite ceci s'écrit  $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|I)$  où  $I$  dénote toute forme d'information connue à propos du problème, et  $\pi(\boldsymbol{\theta}|\mathbf{x}, I) \propto p(\mathbf{x}|\boldsymbol{\theta}, I)\pi(\boldsymbol{\theta}|I)$ .

**Exemple 10** *Dans l'exemple 1, il est assez naturel de supposer que pour un couple étoile-planète donné il existe une intensité  $\alpha_0$  sous-jacente "vraie", l'intensité de la planète. On peut alors par exemple supposer / décider que  $\pi(\alpha|I)$  représente la distribution supposée des intensités des objets orbitant autour d'une étoile. Dans le système solaire on constate avec les petits corps tels que les astéroïdes ou la ceinture de Kuiper que de telles distributions sont de forme hyperbolique :  $\pi(\alpha|I) \propto \alpha^{-\lambda}$  (voir par exemple Fuentes and Holman [2008]).*

*Autrement on peut souhaiter exprimer une forme de méconnaissance qui consiste à chercher un changement de variable sous lequel la forme de la loi a priori doit être la même. En l'occurrence, la valeur  $\alpha = 0$  a ici un sens physique et mathématique naturel : il n'y a pas d'objet, pas de lumière, et 0 est l'élément neutre de l'opération d'addition. Ainsi, travailler à partir de la grandeur  $\beta = \alpha + 3$  semble demander un effort inutile qui la rend inadaptée. Au contraire, lorsque l'on observe un processus dans le temps, le temps  $t = 0$  est choisi arbitrairement et on peut tout aussi bien travailler à partir de  $t$  que  $t' = t + 3$ .*

*Ce qui est dans notre cas arbitraire est l'échelle de l'intensité lumineuse  $\alpha$ . Supposons ainsi qu'une équipe mesure l'intensité en une unité qu'elle appelle Unit (qui est en fait égale au Kilowatt) et qu'une autre la mesure en Watt. Chacune des deux équipes se pose la question du choix de la distribution a priori  $\pi$ . Il semble qu'aucune des deux unités ne soit particulièrement plus adaptée que l'autre, donc si les équipes disposent de la même information et qu'elles essaient de prendre la décision "objectivement" on s'attend à ce que "dans leur référentiel respectif" elles choisissent la même loi de probabilité. Or comme les unités sont par ailleurs reliées par une opération bien définie ( $\beta = 1000\alpha$ ), les lois a priori de ces deux grandeurs sont nécessairement reliées par une loi de changement de variable standard. Ces deux considérations conduisent à contraindre la forme de la loi de probabilité comme nous le verrons plus tard.*

*L'absence d'un facteur d'échelle de référence implique finalement que la distribution doit être invariante par changement d'échelle. Autrement dit, on impose que  $\alpha$  ait la même probabilité d'appartenir à un intervalle numérique fixé pour un " $\alpha$ " exprimé en Watt ou un " $\alpha$ " exprimé en Kilowatt. La fonction qui vérifie cette contrainte est une densité (impropre) hyperbolique :  $\pi(\alpha|I) \propto \alpha^{-1}$ .*

Dans tous les cas, le choix de la distribution *a priori* dépend de l'information qualitative  $I$  que l'on souhaite incorporer quantitativement. Celle-ci n'apparaît pas exhaustivement dans la description du modèle direct qui mène des paramètres aux observations. Le choix de la distribution *a priori* est difficile mais explicite et correspond à la nécessité d'une description exhaustive du problème étudié. On est cependant souvent amené à choisir des hypothèses simplificatrices qui permettent de résoudre le problème en pratique sans rendre la solution invraisemblable ou de propriétés jugées non souhaitables.

### 1.3 Introduction aux tests d'hypothèse fréquentistes et bayésiens

Jusqu'ici quelques notions fondamentales ont été définies à partir de besoins apparaissant pour le problème de détection d'exoplanètes qui nous intéresse. La section qui suit formalise et généralise le traitement statistique du test d'hypothèses en l'abordant sous les angles fréquentiste et bayésien.



## Formulation d'une hypothèse à tester

### • Formulation générale

Le test d'une hypothèse consiste à décider à partir d'un jeu de données  $\mathbf{x}$  si l'hypothèse est correcte. En général, le test est effectué en comparant l'hypothèse testée à une hypothèse alternative, approche en général attribuée à Neyman and Pearson [1933]. Pour une famille paramétrique de distributions fixée et une hypothèse définie sur le paramètre  $\boldsymbol{\theta} \in \Theta$ , on définit en premier lieu la vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta})$ . Dans le cadre bayésien on fixe en plus l'*a priori*  $\pi(\boldsymbol{\theta})$ . Nous verrons plus tard comment il peut être possible de réexprimer un test d'hypothèses dans ce cas.

Dans le cas d'une famille paramétrique de distributions fixée, nous nous intéresserons exclusivement aux tests d'hypothèse qui s'expriment sous la forme

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad H_1 : \boldsymbol{\theta} \in \bar{\Theta}_0 \quad (1.4)$$

où  $\bar{\Theta}_0$  est l'espace complémentaire de  $\Theta_0$  dans  $\Theta$ .

Nous focalisant sur le problème de détection d'exoplanète, le test d'hypothèses central dans cette thèse est le test d'une hypothèse simple (le paramètre testé vaut une valeur fixée) versus une hypothèse composite (le paramètre est égal à n'importe quelle autre valeur) :

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \quad (1.5)$$

Pour un espace des paramètres  $\Theta$  convexe ces hypothèses divisent bien l'espace des paramètres en deux espaces complémentaires mais en même temps "l'hypothèse  $H_0$  est à la limite d'être incluse dans l'hypothèse  $H_1$ " dans le sens où  $\boldsymbol{\theta}_0$  (qui définit l'hypothèse  $H_0$ ) est adhérent à l'espace  $\bar{\Theta}_0$  (qui définit l'hypothèse  $H_1$ ).

**Exemple 11** *Pour un problème de détection de signal, l'hypothèse nulle est simple et consiste en " $H_0 : \boldsymbol{\alpha} = \mathbf{0}$ " (pas de signal) et l'hypothèse alternative est composite et consiste en " $H_1 : \boldsymbol{\alpha} \neq \mathbf{0}$ ". Pour la détection d'exoplanète, sous l'hypothèse nulle, i.e. l'hypothèse par défaut, il n'y a pas d'exoplanète dans les données enregistrées, et sous l'hypothèse alternative une exoplanète est présente mais son intensité est inconnue et peut prendre diverses valeurs : " $H_1 : \boldsymbol{\alpha} \in \mathbb{R}_{+*}^M$ " pour un signal d'intensité positive enregistré dans  $M$  canaux spectraux.*

### • Paramètres de nuisance

Notons que dans le cas d'une hypothèse simple *vs* une hypothèse composite, comme le test est effectué sur  $\boldsymbol{\theta}$  qui définit totalement la vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta})$  on suppose implicitement qu'il n'y a pas de paramètre de nuisance. L'étude menée ici gagne en clarté à ne pas considérer les paramètres de nuisance car pour les prendre en considération il faudrait définir l'hypothèse composite  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0, \boldsymbol{\eta} \in \Omega$  *vs* l'hypothèse composite  $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \boldsymbol{\eta} \in \Omega$  et donc quitter partiellement le cadre du test d'une hypothèse simple *vs* une hypothèse composite. On ne le quitterait que "partiellement" car la dimension de l'espace des paramètres sous  $H_0$  est inférieure à la dimension de l'espace des paramètres sous  $H_1$ .

En pratique il est toujours possible d'appliquer un test d'hypothèses simple *vs* composite même en présence de paramètres de nuisance : il suffit pour un fréquentiste ou un bayésien empirique d'estimer ces paramètres de nuisance par une fonction de  $\boldsymbol{\theta}$  (et de  $\mathbf{x}$ ), et pour un bayésien et quelques fréquentistes de les marginaliser à l'aide de la loi de Bayes :

$$p(\mathbf{x}|\boldsymbol{\theta}) = \int d\boldsymbol{\eta} p(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\eta})\pi(\boldsymbol{\eta}|\boldsymbol{\theta}) \quad (1.6)$$

$$\text{avec } \pi(\boldsymbol{\eta}|\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta}, \boldsymbol{\eta})}{\int d\boldsymbol{\eta} \pi(\boldsymbol{\theta}, \boldsymbol{\eta})}$$

Une comparaison générale entre le fait d'éliminer les paramètres de nuisance  $\boldsymbol{\eta}$  par estimation ou intégration est présentée par Berger et al. [1999].

## Problématiques traitées par un test d'hypothèses

### • Test binaire

**Définition 1** *Un test binaire d'hypothèses consiste à décider si une hypothèse  $H_0$  est correcte ou non à partir d'observations  $\mathbf{x}$ .*

Ainsi classiquement défini, un test d'hypothèses est un estimateur particulier : si on définit  $d = 0$  si  $H_0$  est vraie et  $d = 1$  si  $H_0$  est fautive, le test a pour but d'estimer le paramètre  $d$  en associant à l'espace des observations  $\mathcal{X}$  la statistique  $\hat{d}(\mathbf{x}) = 0$  s'il accepte l'hypothèse ou  $\hat{d}(\mathbf{x}) = 1$  s'il la rejette.

Plus généralement, en sélection de modèles plusieurs hypothèses sont définies. La sélection d'une hypothèse consiste alors en l'estimation d'une variable catégorique : l'hypothèse sous-jacente aux données  $\mathbf{x}$ . Un test d'hypothèses est un cas particulier de la sélection de modèle pour lequel le choix n'est laissé qu'entre deux hypothèses alternatives.

Les définitions et propriétés générales associées aux estimateurs ponctuels sont donc utilisables pour ces types d'inférence. Cependant, en général on ne parle d'*estimation* que pour l'estimation de variables numériques.

**Proposition 1** *En présence d'une hypothèse alternative, le test binaire d'hypothèses qui consiste à accepter ou rejeter l'hypothèse  $H_0$  est équivalent respectivement à rejeter ou accepter l'hypothèse alternative  $H_1$ . Pour une famille paramétrique  $\mathcal{P}_\theta$  de distributions, ces hypothèses sont définies sous la forme de l'Eq. (1.4).*

Dans l'exemple 2, l'hypothèse alternative est  $H_1 : \alpha > 0$  (une planète est présente mais son intensité est inconnue).

D'autre part, une procédure de test peut souvent être définie par les sous-espaces de l'espace des données  $\mathcal{X}$  d'acceptation et de réjection :  $\hat{d}(\mathbf{x}) = 0$  si et seulement si  $\mathbf{x} \in \mathcal{X}_0$  (**espace d'acceptation**) et  $\hat{d}(\mathbf{x}) = 1$  si et seulement si  $\mathbf{x} \in \mathcal{X}_1 = \bar{\mathcal{X}}_0$  (**espace de réjection**).

**Définition 2** *Un test binaire d'hypothèses est dit **déterministe** s'il associe avec une probabilité 1 chaque  $\mathbf{x}$  soit à l'acceptation soit à la réjection de l'hypothèse  $H_0$ , selon la valeur de  $\mathbf{x}$ .*

On montre que les tests déterministes ont certaines propriétés d'optimalité sur tout l'espace des tests, voir Lehmann and Romano [2005], Robert [2007]. Les estimateurs sont de plus en général restreints aux estimateurs déterministes. On se limitera de même dans cette thèse à l'étude des tests déterministes. Dès qu'un test  $T$  est déterministe on peut l'exprimer sous la forme

$$\text{Rejeter } H_0 \text{ si (et seulement si) } \mathbf{x} \in \mathcal{X}_T \tag{1.7}$$

où  $\mathcal{X}_T$  désigne ainsi l'espace de réjection  $\mathcal{X}_{(1)}$  associé au test  $T$ .

### • Performances moyennes et incertitude *postdata*

Le reste de ce chapitre et de cette thèse est sous-tendu par l'importante remarque suivante :

*En général lorsque l'on applique un test d'hypothèses on attend*

- *une décision binaire (réjection ou non de l'hypothèse  $H_0$  pour les données observées  $\mathbf{x}$ ),*
- *des informations *predata* (indépendantes d'éventuelles données observées  $\mathbf{x}$ ) concernant par exemple les performances moyennes, appelées caractéristiques opérationnelles, du test utilisé,*
- *des informations *postdata* (dépendantes des données observées  $\mathbf{x}$ ) concernant la confiance a posteriori à accorder à la décision.*

Les méthodes et outils pour satisfaire ces besoins soulèvent depuis toujours un large débat dans la communauté statistique, en particulier de part l'existence de deux paradigmes distincts : le paradigme fréquentiste (Neyman [1977]) et le paradigme bayésien (Jeffreys [1961, 1st ed. : 1939]). Une importante quantité de travail a été et est encore tournée vers ce problème, sous des approches philosophique, théorique et pratique. La théorie de la décision et la théorie de l'information sont deux approches qui permettent de définir et caractériser des procédures sous un angle systématique. Nous ne ferons pas ici de synthèse de tous ces travaux mais une introduction à la théorie bayésienne de la décision est donnée dans l'annexe B.

## Procédures fréquentistes classiques

Cette section généralise l'introduction au "test d'hypothèses" de la section 1.2.

### • Décision binaire : seuil du GLR ...

Comme dans l'exemple 3, la plupart des tests déterministes  $T$  de paramètre d'entrée  $\boldsymbol{\lambda}$  prennent la forme :

$$\text{Rejeter } H_0 \text{ si } S(\mathbf{x}) \leq \zeta \quad (1.8)$$

pour une certaine statistique  $S : \mathcal{X} \rightarrow \mathbb{R}$ . Ce test est équivalent au test Rejeter  $H_0$  si  $S'(\mathbf{x}) \geq \zeta'$  en remplaçant  $S'(\mathbf{x})$  par  $S(\mathbf{x})^{-1}$  et  $\zeta'$  par  $\zeta^{-1}$ . Dans ce cas, le paramètre d'entrée  $\boldsymbol{\lambda}$  qui caractérise complètement le test  $T$  est un simple scalaire :  $\lambda = \zeta$ , contrairement au test du "PLR" que nous étudierons par la suite dans lequel  $\boldsymbol{\lambda}$  a deux composantes ayant chacune une interprétation propre comme nous le discuterons.

Une méthode générale pour construire  $S(\mathbf{x})$  consiste par exemple à la définir comme le rapport de vraisemblance généralisé (GLR pour *Generalized Likelihood Ratio*) défini pour le test d'hypothèses général de l'Eq. (1.4) par

$$\text{GLR}(\mathbf{x}) = \frac{\max_{\boldsymbol{\theta} \in \Theta_0} p(\mathbf{x}|\boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \bar{\Theta}_0} p(\mathbf{x}|\boldsymbol{\theta})} = \frac{p(\mathbf{x}|\hat{\boldsymbol{\theta}}_{\text{ML},0})}{p(\mathbf{x}|\hat{\boldsymbol{\theta}}_{\text{ML},1})} \quad (1.9)$$

### • Caractéristiques *predata* d'un test : PFA

Dans les procédures fréquentistes classiques explicitant une hypothèse alternative, deux types d'erreur sont utilisés pour définir et/ou caractériser un test  $T$  paramétré par  $\boldsymbol{\lambda}$ . Ces erreurs sont dites respectivement de première espèce pour des données générées sous  $H_0$  (sous  $H_0$ , une erreur se produit si on estime que les données ne sont pas issues de  $H_0$ ) et de seconde espèce pour des données générées sous  $H_1$  (sous  $H_1$ , une erreur se produit si on estime que les données ne sont pas générées sous  $H_1$ ). Introduites dans la section 1.2, ces erreurs sont ainsi données par la PFA (pour *Probability of False Alarm*) et 1 moins PD (*Probability of good Detection*). De façon synthétique, pour un test d'hypothèses de la forme (1.5) on définit PFA et PD par

$$\text{PFA}(T, \boldsymbol{\lambda}) = \text{P}(T, \boldsymbol{\lambda}, \boldsymbol{\theta}_0) \quad (1.10)$$

$$\text{PD}(T, \boldsymbol{\lambda}, \boldsymbol{\theta}_1) = \text{P}(T, \boldsymbol{\lambda}, \boldsymbol{\theta}_1), \quad \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0 \quad (1.11)$$

avec

$$\text{P}(T, \boldsymbol{\lambda}, \boldsymbol{\theta}_i) = \text{Pr}(H_0 \text{ rejeté} | \boldsymbol{\theta}_i) = \int_{\mathcal{X}_{T,\boldsymbol{\lambda}}} d\mathbf{x} p(\mathbf{x}|\boldsymbol{\theta}_i) \quad (1.12)$$

où  $\mathcal{X}_{T,\boldsymbol{\lambda}} = \{\mathbf{x} : H_0 \text{ est rejeté par } T \text{ de paramètre } \boldsymbol{\lambda}\}$

La procédure de Neyman-Pearson consiste à fixer PFA à un certain niveau de confiance ( $p = 5\%$  par ex.) et à calculer un paramètre  $\boldsymbol{\lambda}$  associé. Ce paramètre  $\boldsymbol{\lambda}$  est en général unique comme ceci sera discuté plus tard.

Pour un test d'une hypothèse simple *vs* une hypothèse simple ( $\bar{\Theta}_0 = \{\theta_1\}$ ), le rapport de vraisemblance (LR pour *Likelihood Ratio*) est donné par

$$\text{LR}(\mathbf{x}) = \frac{p(\mathbf{x}|\theta_0)}{p(\mathbf{x}|\theta_1)} \quad (1.13)$$

Le test binaire associé consiste à utiliser  $S(\mathbf{x}) = \text{LR}(\mathbf{x})$  comme statistique de test dans l'Eq. (1.8). D'après le lemme de Neyman-Pearson (Lehmann and Romano [2005]), parmi tous les tests possibles ce test maximise PD pour une PFA fixée.

• **Evidence *postdata* d'une décision : *p-value***

Introduite à la fin de la section 1.2, la *p-value* est la mesure fréquentiste la plus classique de confiance à accorder à une hypothèse. Elle prend sa valeur non dans l'ensemble binaire  $\{0, 1\}$  mais dans tout l'intervalle réel  $[0, 1]$  et est plus informative comme on le verra. On considère souvent qu'elle mesure l'*évidence* de l'hypothèse  $H_0$ . C'est une sorte de "PFA observée", fonction des données observées  $\mathbf{x}$ . La *p-value* est définie de façon non-univoque comme la probabilité sous  $H_0$  d'observer des données "plus surprenantes" (surprenantes en se plaçant sous la perspective de  $H_0$ ) que les données observées  $\mathbf{x}$ .

La *p-value* admet une définition précise consensuelle lorsque  $H_0$  est une hypothèse simple :

**Définition 3** Pour une hypothèse  $H_0$  simple et pour un test d'hypothèses de la forme

$$\text{Rejeter } H_0 \text{ si } S(\mathbf{x}) \leq \zeta \quad (1.14)$$

la *p-value* est définie pour un jeu d'observations  $\mathbf{x}$  fixé par

$$p_{\text{val}}(S(\mathbf{x})) = \Pr(S(\mathbf{y}) \leq S(\mathbf{x})|\theta_0) = \int_{\{\mathbf{y}/S(\mathbf{y}) \leq S(\mathbf{x})\}} d\mathbf{y} p(\mathbf{y}|\theta_0) \quad (1.15)$$

Ceci se réécrit

$$p_{\text{val}}(S(\mathbf{x})) = F_{S|\theta_0}(S(\mathbf{x})) \quad (1.16)$$

où  $F_{S|\theta_0}$  est la distribution cumulative de la variable aléatoire  $S(\mathbf{X})$  sous l'hypothèse  $H_0$ .

**Théorème 1** Si  $S$  est une fonction continue, la *p-value* est égale pour un jeu d'observations  $\mathbf{x}$  fixé à la plus petite valeur de PFA telle que le test (1.14) rejette  $H_0$  pour le seuil  $\zeta$  associé à cette PFA :

$$p_{\text{val}}(S(\mathbf{x})) = \min\{p/S(\mathbf{x}) \leq \zeta \text{ avec } \zeta \text{ tel que } PFA(S, \zeta) = p\} \quad (1.17)$$

Preuve : comme  $F_{S|\theta_0}$  est une distribution cumulative elle est non décroissante. Comme de plus par hypothèse  $S$  est continue,  $F_{S|\theta_0}$  est strictement croissante. Ainsi,

$$S(\mathbf{x}) \leq \zeta \Leftrightarrow F_{S|\theta_0}(S(\mathbf{x})) \leq F_{S|\theta_0}(\zeta)$$

Par ailleurs, la PFA définie par les Eq. (1.10 et 1.12) s'exprime pour le test (1.14) simplement comme la distribution cumulative de  $S(\mathbf{X})$  évaluée en  $\zeta$  :

$$\text{PFA}(S, \zeta) = \int_{\{\mathbf{y}/S(\mathbf{y}) \leq \zeta\}} d\mathbf{y} p(\mathbf{y}|\theta_0) = F_{S|\theta_0}(\zeta)$$

Ainsi,

$$\begin{aligned} \min\{p/S(\mathbf{x}) \leq \zeta \text{ avec } \zeta / \text{PFA}(S, \zeta) = p\} &= \min\{p/F_{S|\theta_0}(S(\mathbf{x})) \leq F_{S|\theta_0}(\zeta) \text{ avec } \zeta / F_{S|\theta_0}(\zeta) = p\} \\ &= \min\{p/F_{S|\theta_0}(S(\mathbf{x})) \leq F_{S|\theta_0}(F_{S|\theta_0}^{-1}(p))\} \\ &= \min\{p/p_{\text{val}}(S(\mathbf{x})) \leq p\} \\ &= p_{\text{val}}(S(\mathbf{x})) \end{aligned}$$

Notons que le test (1.14) est équivalent au test

$$\text{Rejeter } H_0 \text{ si } \mathbf{x} \in \mathcal{X}_\zeta \quad (1.18)$$

si on en définit l'ensemble  $\mathcal{X}_\zeta$  par :

$$\mathcal{X}_\zeta = \{\mathbf{x}/S(\mathbf{x}) \leq \zeta\}$$

Réciproquement, si, pour un certain espace de réjection  $\mathcal{X}_\zeta$  paramétré par  $\zeta$ , il existe une fonction  $S$  telle que

$$\mathbf{x} \in \mathcal{X}_\zeta \Rightarrow S(\mathbf{x}) \leq \zeta \quad (1.19)$$

alors le test (1.18) est équivalent au test (1.14) de sorte que les propriétés écrites pour un test défini par  $(S, \zeta)$  sont alors équivalentes aux propriétés associées à un test défini par son espace de réjection  $\mathcal{X}_\zeta$ . La condition (1.19) est équivalente au fait que pour  $\zeta$  croissant l'espace de réjection  $\mathcal{X}_\zeta$  "s'accroît" au sens de l'inclusion d'ensembles.

La définition 3 et le théorème 1 conduisent ainsi à la définition de *p-value* de Lehmann and Romano [2005] dans le cas d'une hypothèse  $H_0$  simple : la *p-value* peut être définie pour tout test dont la région de réjection s'accroît lorsque le niveau de significativité augmente. La *p-value* est alors définie comme le niveau de confiance limite pour lequel le test rejette  $H_0$ .

Le théorème peut par ailleurs s'exprimer sous diverses formes.

Supposons que le test "initial" consiste à seuiller une statistique  $S(\mathbf{x})$  comme dans l'Eq. (1.8). L'approche de Neyman consiste à fixer un niveau de significativité du test, *i.e.* une PFA notée  $p$ , à en déduire le seuil  $\zeta$  à introduire dans le test, puis à rapporter pour les données  $\mathbf{x}$  observées le résultat binaire du test (1.8) et la valeur de PFA choisie (Neyman and Pearson [1933]) pour définir ce test. Alternativement, l'approche de Fisher consiste pour les données  $\mathbf{x}$  observées à calculer et rapporter la *p-value* (Eq. 1.15) de la statistique  $S(\mathbf{x})$  (Fisher [1970, 1st ed. : 1925]). L'opposition de ces approches est à peu près constamment soulignée (Berger [2003]), mais il est important de noter l'équivalence suivante qui ne transparaît pas clairement dans la littérature :

**Corrolaire 1** *Sous l'approche de Neyman consistant à fixer un niveau de significativité  $p$  afin d'appliquer un test binaire, pour une statistique  $S(\mathbf{x})$  réelle continue appliquer un test binaire consistant à seuiller  $S(\mathbf{x})$  ou à seuiller sa *p-value* est équivalent :*

$$\text{Rejeter } H_0 \text{ si } S(\mathbf{x}) \leq \zeta \quad \text{avec } \zeta \text{ choisi tel que } PFA(S, \zeta) = p \quad (1.20)$$

$$\Leftrightarrow \text{Rejeter } H_0 \text{ si } p_{val}(S(\mathbf{x})) \leq p \quad (1.21)$$

Ceci est une simple réinterprétation du théorème 1. L'approche de Neyman pour le test de l'équation (1.8) consiste en effet à fixer  $p$ , en déduire l'étalonnage  $\zeta = F_{S|\theta_0}^{-1}(p)$  et donc finalement appliquer le test suivant :

$$\text{Rejeter } H_0 \text{ si } S(\mathbf{x}) \leq F_{S|\theta_0}^{-1}(p)$$

Jusqu'ici nous avons étudié la *p-value* en tant qu'une statistique observée, *i.e.* pour un jeu de données  $\mathbf{x}$  fixé. Montrons deux résultats supplémentaires qui découlent directement du résultat précédent mais en considérant la *p-value* comme aléatoire. La *p-value* est une fonction des données observées  $\mathbf{x}$ . De même que les données observées sont la réalisation d'une variable aléatoire  $\mathbf{X}$ , en tant que fonction (déterministe) de  $\mathbf{x}$ , la *p-value* peut en effet aussi être vue comme une variable aléatoire.

**Corrolaire 2** *Le seuil  $p$  de la *p-value* utilisée comme statistique d'un test binaire (Eq. 1.21) donne directement le niveau de significativité de ce test :*

$$PFA(p_{val}, p) = p \quad (1.22)$$

Preuve : d'après l'expression (1.16) de la  $p$ -value,

$$\begin{aligned} \Pr(p_{\text{val}}(S(\mathbf{x})) \leq p \mid \boldsymbol{\theta}_0) &= \Pr(F_{S(\mathbf{y})|\boldsymbol{\theta}_0}(S(\mathbf{x})) \leq p \mid \boldsymbol{\theta}_0) \\ &= \Pr(S(\mathbf{x}) \leq F_{S(\mathbf{y})|\boldsymbol{\theta}_0}^{-1}(p) \mid \boldsymbol{\theta}_0) \\ &= F_{S(\mathbf{y})|\boldsymbol{\theta}_0}(F_{S(\mathbf{y})|\boldsymbol{\theta}_0}^{-1}(p)) \\ &= p \end{aligned}$$

On dit ainsi que la  $p$ -value est le niveau de significativité observé. Sous l'approche de Neyman, l'avantage d'utiliser directement la  $p$ -value comme statistique de test seuillée par  $p$  (test défini par l'Eq. 1.21), plutôt que le test initial donné par l'Eq. (1.14), est que l'expression de la  $p$ -value est bien connue, contrairement à  $F_{S|\boldsymbol{\theta}_0}$  qui aurait besoin d'être défini ou l'approche adoptée précisée.

En termes de distribution et non plus de caractéristique d'un test binaire, on a ainsi démontré que

**Corrolaire 3** *Si  $S(\mathbf{X})$  est une variable aléatoire réelle continue, sous l'hypothèse  $H_0$  la distribution de la  $p$ -value vue comme une variable aléatoire est uniforme sur l'intervalle  $[0, 1]$  :*

$$p_{\text{val}}(S(\mathbf{X})) \mid \boldsymbol{\theta}_0 \sim \mathcal{U}(0, 1) \quad (1.23)$$

Inverser ainsi une distribution cumulative constitue la méthode classique de génération d'une variable aléatoire sous une loi relativement quelconque à partir d'un échantillon distribué selon une loi uniforme.

D'autres informations concernant la  $p$ -value viendront dans la prochaine section où nous décrivons les procédures bayésiennes classiques.

## Procédures bayésiennes classiques pour une hypothèse $H_0$ non spécifiée

### • Notions bayésiennes standards

Le théorème de Bayes permet de donner un sens à la probabilité d'une hypothèse étant données les observations  $\mathbf{x}$ . Quel que soit l'espace sur lequel on définit les hypothèses (espace des paramètres, espace des vraisemblances), dans le cas simple où on suppose qu'il n'y a que deux hypothèses alternatives la probabilité *a posteriori* de l'hypothèse nulle étant donnée  $\mathbf{x}$  est donnée par :

$$\Pr(H_0 \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid H_0) \Pr(H_0)}{p(\mathbf{x} \mid H_0) \Pr(H_0) + p(\mathbf{x} \mid H_1) \Pr(H_1)} \quad (1.24)$$

avec  $\int_{\mathcal{X}} d\mathbf{x} p(\mathbf{x} \mid H_i) = 1$  et  $\Pr(H_0) + \Pr(H_1) = 1$ .

Cette probabilité est reliée directement au rapport de *odds a posteriori* (POR pour *Posterior Odds Ratio*)

$$\text{POR}(\mathbf{x}) = \frac{\Pr(H_0 \mid \mathbf{x})}{\Pr(H_1 \mid \mathbf{x})} = \frac{\Pr(H_0 \mid \mathbf{x})}{1 - \Pr(H_0 \mid \mathbf{x})} \quad (1.25)$$

lui-même relié au Facteur de Bayes, voir Kass and Raftery [1995] :

**Définition 4** *Le Facteur de Bayes (BF pour Bayes Factor) est donné par le rapport des vraisemblances (marginales) d'un couple d'hypothèses :*

$$BF(\mathbf{x}) = \frac{p(\mathbf{x} \mid H_0)}{p(\mathbf{x} \mid H_1)} \quad (1.26)$$

**Propriété 1** *Le facteur de Bayes est égal au rapport des rapport de odds a posteriori (POR) et a priori (pOR) :*

$$BF(\mathbf{x}) = \frac{\text{POR}(\mathbf{x})}{p\text{OR}}$$

où  $POR(\mathbf{x})$  est défini dans l'Eq. (1.25) et  $pOR$  par :

$$pOR = \frac{Pr(H_0)}{Pr(H_1)} \quad (1.27)$$

Nous verrons précisément dans la section 1.3 comment s'expriment ces tests si les hypothèses  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  et  $H_1 : \boldsymbol{\theta} \in \Theta_0$  sont lues au sens strict. Notons par avance que dans ce cas le BF vaut simplement

$$BF(\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)}{\int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \quad (1.28)$$

avec  $\pi$  un *a priori* continu sur  $\boldsymbol{\theta}$  décrivant sa loi sous  $H_1$ . Un problème important du BF est qu'il est alors défini à une constante multiplicative près si l'*a priori* est impropre, *i.e.* si  $\int d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) = \infty$ , ne définissant alors pas une distribution de probabilité. Cependant, en général même si l'*a priori* est impropre la distribution *a posteriori*  $\pi(\boldsymbol{\theta}|\mathbf{x})$  est propre et définit bien une distribution de probabilité : grâce aux données, la distribution *a priori* de  $\boldsymbol{\theta}$  initialement impropre est actualisée en une distribution *a posteriori* qui devient propre. Etendant cette idée, O'Hagan [1995], Berger and Pericchi [1996] proposent des facteurs de Bayes partiels (PBF pour *Partial Bayes Factor*) comme alternatives au BF lorsqu'un *a priori* impropre est utilisé. Schématiquement, ils supposent qu'il existe un "échantillon minimal d'entraînement"  $\mathbf{y}$  choisi à partir de l'ensemble des données  $\mathbf{x}$  tel que  $\pi(\boldsymbol{\theta}|\mathbf{y})$  est propre. Un BF (partiel) peut alors être défini de façon unique à partir du reste des données, utilisant  $\pi(\boldsymbol{\theta}|\mathbf{y})$  comme *a priori* propre. Parmi ceux-ci, le facteur de Bayes fractionnaire proposé par O'Hagan [1995] permet d'éviter d'avoir effectivement à manipuler le contenu des données  $\mathbf{x}$  et peut être appliqué aussi facilement à des données *i.i.d.* que non-*i.i.d.* : il pondère simplement la vraisemblance par rapport à l'*a priori* :

**Définition 5** *Le facteur de Bayes fractionnaire (ou FBF pour Fractional Bayes Factor) proposé par O'Hagan [1995] est défini dans le cas simple vs composite par*

$$FBF(\mathbf{x}, b) = \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)}{\int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \left( \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)^b}{\int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})^b \pi(\boldsymbol{\theta})} \right)^{-1} \quad (1.29)$$

Le FBF est aussi étudié par O'Hagan [1997] et comparé aux autres PBF, et plus spécifiquement à l'IBF (pour BF intrinsèque) proposé par Berger and Pericchi [1996].

- **Décision binaire : seuil de  $POR(\mathbf{x})$ ,  $BF(\mathbf{x})$  ...**

Chacune des statistiques  $Pr(H_0|\mathbf{x})$ ,  $POR(\mathbf{x})$ ,  $p(\mathbf{x}|H_0)$  et  $BF(\mathbf{x})$  ont un sens propre. Il est naturel de les seuiller pour obtenir un test d'hypothèses binaire, en attribuant une interprétation spécifique à leurs seuils respectifs. Préassigner une valeur de seuil de  $Pr(H_0|\mathbf{x})$  (ou du  $POR(\mathbf{x})$ , de  $p(\mathbf{x}|H_0)$  ou du  $BF(\mathbf{x})$ ) constitue une approche alternative à l'approche de Neyman consistant à fixer le seuil de la *p-value* par le "niveau de significativité du test" *i.e.* de sa PFA (voir le corrolaire 1). Notons par ailleurs que ces statistiques bayésiennes mènent à des tests équivalents dès lors que l'on établit une correspondance entre les seuils de chacun de ces tests. On vérifie par exemple aisément :

**Proposition 2** *Les tests suivants sont identiques :*

$$\begin{aligned} \text{Rejeter } H_0 \text{ si (et seulement si)} \quad & Pr(H_0|\mathbf{x}) < \zeta \\ \text{Rejeter } H_0 \text{ si (et seulement si)} \quad & BF(\mathbf{x}) < (\zeta^{-1} - 1)^{-1} pOR^{-1} \end{aligned} \quad (1.30)$$

• **Caractéristiques *predata* d'un test : PFA, risque intégré**

Des calculs fréquentistes (qui moyennent sur  $\mathcal{X}$  par la vraisemblance) peuvent aussi intéresser les bayésiens (Rubin [1984]). En plus de PFA et PD définis dans la section 1.3, le risque intégré –qui intègre une fonction de coût sur  $\mathcal{X}$  et  $\Theta$  par la vraisemblance et l'*a priori*– peut aussi être utilisé comme caractéristique *predata* d'un test (Robert and Casella [1994]). En tant que notion de théorie de la décision, il sera défini plus en détail dans l'annexe B.1. Notons simplement que dans le cas le plus simple le risque intégré est égal à la somme pondérée de PFA et PD.

• **Evidence *postdata* d'un test :  $\Pr(H_0|\mathbf{x})$  et retour sur les *p-values***

Alternativement aux *p-values* introduites initialement dans le paradigme fréquentiste, en bayésien c'est la probabilité *a posteriori* de  $H_0$  ( $\Pr(H_0|\mathbf{x})$ ) directement reliée au BF( $\mathbf{x}$ ) comme on vient de le voir) qui est considérée comme la mesure standard d'évidence de l'hypothèse  $H_0$ .

L'utilisation des *p-values* est depuis longtemps un sujet très controversé parmi les communautés fréquentistes aussi bien que bayésiennes. Elle a été impulsée par la pratique plutôt que sur des bases théoriques telles que celles de la théorie de la décision par exemple, et sa définition même est encore débattue dans certains cas. Un grand nombre de statisticiens tant des domaines appliqué que théorique ont cherché à relier cette notion intuitive à des notions statistiques un peu plus consensuelles comme le niveau de confiance d'un test ou la mesure d'évidence bayésienne. Schématiquement, les approches possibles d'étude sont les suivantes : des études de cas assez spécifiques (Meulepas [1998]), des études de cas plus générales (Tsui and Weerahandi [1989]), des analyses par la théorie de la décision (Hwang et al. [1992]) et/ou des arguments d'interprétation (Good [1983]). Les études fréquentistes et bayésiennes menées ont naturellement mené à la fois à des paradoxes (Lindley [1957]) et à de nouvelles formes de *p-value* (Meng [1994]).

Ainsi, les *p-values* sont encore très utilisées et étudiées (y compris sous une approche bayésienne, voir Meng [1994] pour la plus classique) mais  $\Pr(H_0|\mathbf{x})$  tient une place centrale dans le paradigme bayésien. De nombreuses études cherchent à concilier la *p-value* fréquentiste standard et  $\Pr(H_0|\mathbf{x})$ . Nous focalisant sur le cas du test d'une hypothèse simple *vs* une hypothèse composite *two-sided* (ex : hypothèse  $H_1 : \theta < \theta_1$  ou  $\theta > \theta_2$  avec  $\theta_1 \leq \theta_2$ ), voici quelques résultats qui indiquent la difficulté d'y parvenir :

- Le paradoxe de Lindley [1957] montre notamment que pour un *a priori* fixé et un  $\mathbf{x}$  qui s'ajuste de sorte que la *p-value* reste fixée,  $\Pr(H_0|\mathbf{x})$  tend vers 1 à mesure que la taille de l'échantillon  $\mathbf{x}$  augmente. Shafer [1982] discute ce sujet en rappelant que les hypothèses sous-jacentes d'un problème statistique sont cruciales dans le choix de la procédure d'inférence : selon le cas, cela peut être soit  $\Pr(H_0|\mathbf{x})$  –à travers le choix de l'*a priori*– soit la *p-value* qui est critiquable. Le paradoxe de Lindley sera présenté sous une autre forme dans le chapitre 7.
- Supposant que  $\Pr(H_0) = 1/2$ , Edwards et al. [1963] montrent dans certains cas que quel que soit l'*a priori*  $\pi$  sous  $H_1$ ,  $\Pr(H_0|\mathbf{x})$  est strictement supérieur à la *p-value*. Ce type d'étude est repris par Berger and Selke [1987] (et nombreuses références incluses) qui aboutit au même type de conclusion.
- Oh and DasGupta [1999] se préoccupent de plus du choix de  $\Pr(H_0)$  pour rendre la *p-value* égale à  $\Pr(H_0|\mathbf{x})$ .
- Sellke et al. [2001] essaient de concilier les *p-values* avec le BF en “calibrant” la *p-value* *i.e.* en définissant une fonction  $f$  de la *p-value* qui soit sur la même échelle que le BF. En particulier, leur étalonnage  $f$  mène à  $\text{BF}(\mathbf{x}) > f(p_{\text{val}}(\mathbf{x}))$  quel que soit le modèle et la *p-value* (pertinente) choisis, de sorte que même sous ce nouvel étalonnage Sellke et al. [2001] concluent que la *p-value* est biaisée vers un rejet trop fréquent de  $H_0$ , comme c'était déjà le cas sans nouvel étalonnage.
- Après de nombreuses études critiques de la *p-value* classique, Berger et al. [1997] cherchent finalement à concilier les procédures fréquentiste et bayésienne en utilisant à la fois une probabilité conditionnelle à une certaine statistique plutôt que la *p-value* pure et en même temps en introduisant une région de “non-décision” dans la procédure bayésienne.



Notons cependant que Casella and Berger [1987] indiquent qu'une telle irréconciliabilité entre  $\Pr(H_0|\mathbf{x})$  et  $p$ -value est beaucoup moins flagrante en cas *one-sided* (ie tel que sous  $H_1 : \theta > \theta_0$ ) que dans les cas *two-sided* référencés ici. Pour des familles d'*a priori* raisonnables la borne inférieure de  $\Pr(H_0|\mathbf{x})$  peut en effet alors coïncider avec la  $p$ -value.

## Application des procédures bayésiennes avec une hypothèse $H_0$ paramétrique

### • Définition du problème statistique

Nous supposerons partout dans cette thèse que tout test d'hypothèses est construit en définissant dans un premier temps les distributions sur  $\mathcal{X}$  et  $\Theta$  communes à toutes les hypothèses, puis l'hypothèse nulle paramétrique identique au cas fréquentiste :

1. un modèle paramétrique des données  $\mathcal{P}_\Theta = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  où  $p(\cdot|\boldsymbol{\theta})$  désigne la densité de probabilité de  $\mathbf{X}$
2. une densité *a priori*  $\pi(\cdot)$  sur  $\Theta$ , telle que  $\int_\Theta d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) = 1$  si l'*a priori* est propre ou telle que  $\int_\Theta d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) = +\infty$  si l'*a priori* est impropre.
3. une hypothèse nulle paramétrique de la forme  $H_0 : \boldsymbol{\theta} \in \Theta_0$ , comme dans la définition 1

Notons cependant tout de suite qu'avec ce choix, **pour une hypothèse simple  $\Theta_0 = \{\boldsymbol{\theta}_0\}$** , on réécrit en général de façon équivalente cette procédure. Si  $\boldsymbol{\theta}$  est une variable aléatoire continue sous  $H_1$  mais que  $\Theta_0 = \{\boldsymbol{\theta}_0\}$ , on supposait en fait d'emblée que l'*a priori*  $\pi$  était constitué d'un pic de Dirac en  $\boldsymbol{\theta}_0$ , affecté d'un coefficient  $\Pi_0$ . Dans ce cas, sous  $H_0$  la seule information contenue dans  $\pi$  est le scalaire  $\Pi_0$ , de sorte qu'il est alors plus commode de définir un *a priori* dit sous  $H_1$  qui est continu et normalisé non sur  $\Theta$  mais sur  $\bar{\Theta}_0$ . Dans ce cas, la procédure de définition de test d'hypothèses ci-dessus est équivalente à choisir dans l'ordre

1. un modèle paramétrique des données  $\mathcal{P}_\Theta = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  où  $p(\cdot|\boldsymbol{\theta})$  désigne la densité de probabilité de  $\mathbf{X}$
2. une hypothèse nulle paramétrique de la forme  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ , comme dans la définition 1 mais dans le cas simple
3. une densité *a priori* **continue  $\pi_1(\cdot)$  sur  $\bar{\Theta}_0$** , telle que  $\int_{\bar{\Theta}_0} d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) = 1$  si  $\int_{\bar{\Theta}_0} d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) < \infty$

L'indice 1 de  $\pi_1(\cdot)$  est alors parfois implicite. Cette approche est bien équivalente à l'approche définie ci-dessus dans la mesure où  $\pi_1(\cdot)$  est correctement défini à partir de  $\pi(\cdot)$ . Les détails seront donnés dans un instant.

### • Réexpression des notions bayésiennes précédentes pour $H_0 : \boldsymbol{\theta} \in \Theta_0$

En définissant une hypothèse paramétrique comme  $H_0 : \boldsymbol{\theta} \in \Theta_0$ , on définit implicitement via le signe “:” une fonction bijective entre l'espace des hypothèses et l'espace des paramètres. L'hypothèse  $H_1$  elle-même est alors définie sans ambiguïté par  $H_1 : \boldsymbol{\theta} \in \bar{\Theta}_0$ . Plus généralement, cette équivalence  $H_0 : \boldsymbol{\theta} \in \Theta_0$  permet de réexprimer n'importe laquelle des notions présentées dans la section 1.3 en remplaçant l'hypothèse  $H_i$  par son expression dans le domaine des paramètres  $\boldsymbol{\theta} \in \Theta_i$ .

Considérons pour exemple deux façons équivalentes de calculer  $\Pr(H_0|\mathbf{x})$  à partir de la vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta})$  et l'*a priori*  $\pi(\boldsymbol{\theta})$ .

**Réexpression directe de  $\Pr(H_0|\mathbf{x})$**  La probabilité *a posteriori* de l'hypothèse nulle s'exprime simplement à partir de la distribution *a posteriori* de  $\boldsymbol{\theta}$  selon

$$\Pr(H_0|\mathbf{x}) = \Pr(\boldsymbol{\theta} \in \Theta_0|\mathbf{x}) = \int_{\Theta_0} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}|\mathbf{x}) = \int_{\Theta_0} d\boldsymbol{\theta} \frac{p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} = \frac{\int_{\Theta_0} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \quad (1.31)$$

**Réexpression équivalente en passant par  $p(\mathbf{x}|H_0)$  et  $\Pr(H_0)$**  Toujours d'après notre construction du modèle statistique, la vraisemblance marginale de  $\mathbf{x}$  sous  $H_0$  et la probabilité *a priori* de  $H_0$  s'expriment comme

$$\begin{aligned} p(\mathbf{x}|H_0) &= \int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}, \boldsymbol{\theta}|H_0) = \int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta}, H_0)\pi(\boldsymbol{\theta}|H_0) = \int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\theta} \in \Theta_0) \\ &= \int_{\Theta_0} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta}) \frac{\pi(\boldsymbol{\theta})}{\Pr(\boldsymbol{\theta} \in \Theta_0)} = \frac{\int_{\Theta_0} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Theta_0} d\boldsymbol{\theta} \pi(\boldsymbol{\theta})} \end{aligned} \quad (1.32)$$

$$\Pr(H_0) = \Pr(\boldsymbol{\theta} \in \Theta_0) = \int_{\Theta_0} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) \quad (1.33)$$

**Définition d'un *a priori*  $\pi_0$  normalisé sur  $\Theta_0$**  Si on définit maintenant un *a priori*  $\pi_0(\cdot)$  sur  $H_0$  par

$$\begin{aligned} \int_{\Theta} \pi_0(\boldsymbol{\theta})d\boldsymbol{\theta} &= 1 \\ \pi_0(\boldsymbol{\theta}) &= \pi(\boldsymbol{\theta}) \Pi_0^{-1} \quad \forall \boldsymbol{\theta} \in \Theta_0 \\ \pi_0(\boldsymbol{\theta}) &= 0 \quad \text{sinon} \end{aligned} \quad (1.34)$$

alors les équations (1.32,1.33) se réécrivent

$$p(\mathbf{x}|H_0) = \int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta}) \quad (1.35)$$

$$\Pr(H_0) = \Pi_0 \quad (1.36)$$

• **Cas du test d'hypothèses simple versus composite :  $\Theta_0 = \{\boldsymbol{\theta}_0\}$**

Lorsque l'hypothèse nulle  $\boldsymbol{\theta} \in \Theta_0$  se résume à une hypothèse simple où  $\Theta_0 = \{\boldsymbol{\theta}_0\}$ , si l'on reste dans le cadre habituel précédent  $\boldsymbol{\theta}$  continue et donc  $\Pr(H_0) = 0$  alors

$$\Pr(H_0|\mathbf{x}) = \Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_0 | \mathbf{x}) = 0 \quad (1.37)$$

Afin de pouvoir utiliser les outils bayésiens classiques présentés ci-dessus, il est donc *nécessaire* de choisir une loi *a priori*  $\pi(\boldsymbol{\theta})$  *non régulière i.e.* non absolument continue, qui se décompose sur  $\Theta$  en une masse et une fonction continue pour  $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$  :

$$\pi(\boldsymbol{\theta}_0) = \Pi_0 \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \quad \text{avec } 0 < \Pi_0 < 1 \quad (1.38)$$

$$\pi(\boldsymbol{\theta}) = (1 - \Pi_0)\pi_1(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \quad (1.39)$$

de sorte que  $\pi_1$  est un *a priori* continu avec  $\int_{\Theta} d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) = 1$ . Le POR défini dans l'Eq. (1.25) se réécrit alors

$$\text{POR}(\mathbf{x}) = \frac{\Pi_0 p(\mathbf{x}|\boldsymbol{\theta}_0)}{\Pi_0 p(\mathbf{x}|\boldsymbol{\theta}_0) + (1 - \Pi_0) \int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta})} \quad (1.40)$$

Le choix du coefficient  $\Pi_0$  est problématique : les utilisateurs trouvent rarement de bonnes raisons de définir ce terme par une valeur plutôt qu'une autre, sauf éventuellement  $\Pi_0 = 1/2$  par souci de symétrie, symétrie vue du point de vue label des hypothèses  $i = 0$  et  $i = 1$ . L'utilisation du facteur de Bayes (BF, Eq. 1.26) permet d'éviter le problème tout en bénéficiant d'une interprétation spécifique : le BF est le rapport des vraisemblances marginales :

$$\text{BF}(\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)}{\int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta})} \quad (1.41)$$

Malgré tous ces efforts pour traiter le cas simple *vs* composite, ce test d'hypothèses continue à soulever des interrogations et difficultés telles que le paradoxe de Lindley [1957]. On a pu voir en particulier apparaître le problème posé par le cas simple *vs* composite en constatant qu'un *a priori* sur  $\Theta$  tout entier avait alors une composante discrète parmi une distribution ailleurs continue. Sans un tel *artifice*, pour un *a priori* continu partout nous aurions  $\Pr(H_0|\mathbf{x}) = 0$ . Berger and Selke [1987] considère néanmoins qu'une telle hypothèse n'a en réalité pas de sens et qu'une hypothèse  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  représente en fait plutôt la limite quand  $\zeta \rightarrow 0$  d'une hypothèse du type  $H_0 : d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \leq \zeta$  avec  $d$  une distance sur  $\Theta$ .

• **Réexpression des notions précédentes pour  $H_0 : p_{\boldsymbol{\theta}} \in \mathcal{P}_{\Theta_0}$**

En ajustement ou sélection de modèle,  $H_0$  est défini *via* le domaine d'appartenance de la distribution paramétrique de  $\mathbf{X}$ .

**Ajustement strict de modèle** Soit  $\mathcal{P}_{\Theta} = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  une famille paramétrique de densités de probabilité sur l'espace des données  $\mathcal{X}$ . Un problème de sélection de modèle paramétrique consiste à tester l'hypothèse

$$H'_0 : p(\cdot|\boldsymbol{\theta}) \in \mathcal{P}_{\Theta_0} \quad H'_1 : p(\cdot|\boldsymbol{\theta}) \in \mathcal{P}_{\bar{\Theta}_0} \quad (1.42)$$

où  $p(\cdot|\boldsymbol{\theta}) \in \mathcal{P}_{\Theta}$  est la densité inconnue de  $\mathbf{X}$ .

**Proposition 3** *Si on suppose que la densité inconnue  $p(\cdot|\boldsymbol{\theta})$  de  $\mathbf{X}$  appartient à  $\mathcal{P}_{\Theta} = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  et que deux fonctions  $p(\cdot|\boldsymbol{\theta}) \in \mathcal{P}_{\Theta}$  et  $p(\cdot|\boldsymbol{\theta}')$  sont égales si et seulement si  $\boldsymbol{\theta} = \boldsymbol{\theta}'$  alors les hypothèses*

$$\begin{aligned} H_0 &: \boldsymbol{\theta} \in \Theta_0 \\ H'_0 &: p(\cdot|\boldsymbol{\theta}) \in \mathcal{P}_{\Theta_0} \end{aligned}$$

sont équivalentes.

**Ajustement non strict de modèle** Jusqu'ici et partout ailleurs, une hypothèse paramétrique est définie comme l'appartenance stricte et la non-appartenance stricte du paramètre testé  $\boldsymbol{\theta}$  à un domaine  $\Theta_0 : H_0 : \boldsymbol{\theta} \in \Theta_0$ . Comme nous l'avons mentionné plus haut on peut alternativement introduire une distance  $d$  entre le paramètre testé et l'espace  $\Theta_0$ , définissant ainsi une hypothèse de la forme  $H_0(\zeta) : d(\boldsymbol{\theta}, \Theta'_0) \leq \zeta$ .

**Proposition 4** *Les hypothèses*

$$\begin{aligned} H_0 &: \boldsymbol{\theta} \in \Theta_0 \\ H_0(\zeta) &: d(\boldsymbol{\theta}, \Theta'_0) \leq \zeta \end{aligned}$$

sont équivalentes si  $\Theta'_0$  est fixé et que l'on définit  $\Theta_0$  par  $\Theta_0 = \{\boldsymbol{\theta} / d(\boldsymbol{\theta}, \Theta'_0) \leq \zeta\}$ .

En ajustement de modèle, la distance (ou la *divergence* qui désigne une distance non nécessairement symétrique), est définie comme la distance entre la distribution testée et la famille de distributions. Cette approche est intéressante car elle permet de faire appel à la théorie de l'information développée notamment par Kullback [1968]. Ce dernier définit en particulier la divergence de Kullback-Leibler, définie entre deux densités de distribution  $p$  et  $q$  par :

$$K(p, q) = \int_{\mathcal{X}} d\mathbf{x} p(\mathbf{x}) \log \left( \frac{p(\mathbf{x})}{q(\mathbf{x})} \right)$$

En utilisant ainsi une notion de *distance*, définir l'hypothèse sur l'espace  $\Theta$  ou sur l'espace des distributions  $\mathcal{P}_{\Theta}$  n'est alors plus nécessairement équivalent. Plus précisément, il n'est *a priori* pas toujours possible de trouver une "distance"  $d'$  à appliquer à l'espace des distributions telle que l'hypothèse

$H_0^\zeta : d(\boldsymbol{\theta}, \Theta_0) \leq \zeta$  soit équivalente à  $H'_0 : d'(p_{\boldsymbol{\theta}}, \mathcal{P}_{\Theta_0}) \leq \zeta'$ .

*Il convient alors de se demander selon l'étude menée si l'hypothèse doit être définie sur  $\Theta$  ou  $\mathcal{P}_\Theta$ .*

*Notons de plus que les distances (et divergences) classiques entre distributions ne font pas intervenir les données observées  $\boldsymbol{x}$  car les fonctions de  $\boldsymbol{x}$  et  $\boldsymbol{\theta}$  indiquant la divergence sont par exemple moyennées sur  $\mathcal{X}$ .*



## Première partie

# Modèle gaussien simple pour construire des inférences applicables



# Table des Matières

---

<b>2</b>	<b>Simulation physique fine des données SPHERE</b>	<b>35</b>
2.1	Simulation détaillée d'un jeu de données standard acquis par SPHERE . . . . .	35
	Définition du cas de simulation . . . . .	35
	Paramètres importants décrivant le jeu de données simulé . . . . .	36
2.2	Etude statistique qualitative des données simulées en l'absence de planète . . . . .	40
	Introduction . . . . .	40
	Description statistique des données . . . . .	43
	Discussion des analyses précédentes . . . . .	49
	Conclusion quant aux modèles statistiques étudiés dans cette thèse . . . . .	52
<b>3</b>	<b>Construction et analyse d'inférences à partir d'un modèle gaussien</b>	<b>55</b>
3.1	Modèle gaussien des données . . . . .	55
	Description qualitative d'une image . . . . .	55
	Discrimination entre signal (planète) et bruit (étoile) . . . . .	57
	Modèle statistique pour une bande spectrale . . . . .	57
3.2	Estimation d'intensité et de position de la planète . . . . .	58
	Estimation d'intensité $\hat{\alpha}(r)$ (et estimation <i>conjointe</i> du fond) . . . . .	59
	Estimation de la position initiale de la planète $\hat{r}_{ML}$ . . . . .	61
	Estimation d'un nombre de photons en pratique . . . . .	62
3.3	Consistance des estimateurs et estimation d'incertitude . . . . .	63
	Si $r$ est connu, $\hat{\alpha}_{ML} = \hat{\alpha}(r)$ est un estimateur non biaisé, efficace et consistant . . . . .	63
	Consistance de $\hat{\alpha}_{ML}$ si $r$ est inconnu . . . . .	66
	Estimation d'incertitude sur l'intensité par l'estimation de $\text{var}[\hat{\alpha}]$ . . . . .	67
3.4	Détection initiale : obtention d'une liste de candidats . . . . .	68
	Test de détection global dans tout le champ . . . . .	69
	Test de détection pour chaque position initiale $r$ . . . . .	70
3.5	Extension de la détection-estimation à l'imagerie simultanée dans plusieurs canaux spectraux	71
	Extension du modèle mono-canal . . . . .	71
	Estimateur d'intensité de la planète dans les différentes bandes . . . . .	72
	Test de détection . . . . .	73
<b>4</b>	<b>Application aux données simulées finement et résultats</b>	<b>75</b>
4.1	Discussions plus détaillées de la validité des conditions . . . . .	75
	Pré-estimation des profils de planète $\mathbf{p}(k, r)$ . . . . .	75
	Variation spatiale du terme de "normalisation" de $\hat{\alpha}(r)$ . . . . .	76
	Variation spatiale de la variance de bruit $\sigma^2$ . . . . .	76
	Si $r$ est connu, condition pour la consistance des estimateurs . . . . .	78



4.2	Programme informatique <i>user-friendly</i> : MOODS sur Matlab . . . . .	78
4.3	Application de la méthode à des données (simulées) stationnaires et non-stationnaires . .	81
	Données stationnaires . . . . .	81
	Données non-stationnaires . . . . .	84
4.4	Autres applications de la méthode . . . . .	86
	Application de la méthode sur des données de contrastes différents . . . . .	86
	Utilisation de deux bandes spectrales . . . . .	86
	Effet de la direction du vent . . . . .	88
	Comparaison de MOODS aux autres techniques standards . . . . .	88

---

## Chapitre 2

# Simulation physique fine des données SPHERE

### 2.1 Simulation détaillée d'un jeu de données standard acquis par SPHERE

Afin de tester la méthode il est nécessaire de disposer de jeux de données. Or l'instrument SPHERE introduit dans la section 1.1, et plus généralement les instruments disposant d'une nouvelle génération d'AO comme l'instrument GPI, n'est pas encore monté sur le VLT et ne délivre donc pas encore de vraies données. Il est donc nécessaire quoi qu'il en soit de faire appel à des simulations numériques détaillées.

La simulation physique *end-to-end* fine de l'instrument SPHERE fait l'objet d'un travail spécifique du consortium SPHERE. L'instrument SPHERE et le sous-instrument IRDIS en particulier sont simulés de façon détaillée par le *package* SPHERE (Carbillet et al. [2008]) du *problem-solving environment* CAOS (Carbillet et al. [2010]). Un ensemble de cas standards de simulation a été défini par le consortium afin notamment de permettre au groupe DRH (*Data Reduction and Handling*) de développer et tester leurs algorithmes sur de mêmes jeux de données.

En particulier, comme décrit par Smith et al. [2008], le cas de simulation détaillé ci-dessous, réalisé à partir de la version 3.0 du *package* SPHERE (Carbillet et al. [2008]), correspond au cas de simulation dit *non-stationnaire* dans la prochaine section. C'est aussi le cas traité partout ailleurs dans cette thèse, sauf le cas dit *stationnaire* de la prochaine section qui servira justement de point de comparaison.

#### Définition du cas de simulation

Le modèle physique utilisé pour les simulations a pour objectif de décrire de façon aussi réaliste que possible, mais avec un temps de simulation raisonnable, une scène astrophysique d'intérêt et la réponse de l'instrument SPHERE utilisé dans des conditions réalistes. Contrairement aux séries de simulations effectuées jusqu'à 2008, l'objectif ici est de définir un cas de simulation destiné aux études de traitement du signal et non plus aux études des sous-systèmes de SPHERE. La version 3.0 du *package* SPHERE tient compte en particulier de l'évolution temporelle des conditions atmosphériques et des aberrations instrumentales quasi-statiques dues au changement d'angle de hauteur visé par le télescope au cours de la nuit en suivant une étoile.

Comme introduit dans la section 1.1, le mode d'observation dit *pupille stabilisée* est choisi pour le dérotateur de champ et nous nous intéressons aux données issues du mode DBI du sous-instrument IRDIS de SPHERE. Ces deux hypothèses définissent le mode d'observation dit de *survey*. Parmi les deux coronographes possibles dans ce mode d'observation, dans la suite nous considérons le coronographe de Lyot apodisé et non le coronographe à quatre quadrants.

La modélisation physique utilisée dans la simulation inclut la variabilité des conditions atmosphériques (*seeing*, vitesse et direction du vent des différentes couches atmosphériques modélisées), la modélisation

des aberrations résiduelles post-AO intégrées dans le temps (images longues poses), les variations des aberrations instrumentales directement associées à la hauteur de l'étoile visée dans le ciel (résidus d'aberrations du correcteur de dispersion atmosphérique ou ADC, défauts optiques), les variations associées aux dérives de grande échelle temporelle (stabilité de pointage, stabilité de focalisation, etc) et les bruits de détecteur. La quantification de ces effets résulte notamment d'études expérimentales détaillées des sous-systèmes.

Les temps caractéristiques mis en jeu dans cette modélisation physique des données sont de différents ordres. Par exemple, alors que le temps de cohérence typique du front d'onde post-atmosphère est de l'ordre de  $10\text{ ms}$ , le temps caractéristique de stabilité des conditions de turbulence atmosphérique est mesurée en minutes. En ce qui concerne la vitesse de rotation de champ, le temps correspondant à une rotation de champ de  $\lambda/D$  s'étale de la minute à l'heure, dépendant de la direction de pointage et de la séparation entre l'étoile et la planète recherchée. Tandis que le temps d'intégration typique du détecteur (DIT) est de  $10\text{ s}$ , le temps d'observation total  $T$  est de typiquement quelques heures, le début de l'observation étant choisi de telle sorte que l'étoile passe au méridien au milieu de l'observation. Les aberrations instrumentales fonction de la hauteur de l'étoile pointée ont un temps de corrélation typique de l'ordre de quelques minutes.

La simulation est effectuée par le *package* SPHERE version 3.0 Carbillet et al. [2008] programmée dans le langage IDL (*Interactive Data Language*). Le *package* s'appuie dans un premier temps sur le *problem solving environment* CAOS pour simuler la propagation optique diffractive d'un front d'onde plan à travers l'atmosphère et l'ensemble de l'instrument, jusqu'au détecteur. Cette partie du code, dite *diffractive*, ne dépend pas des nombres de photons par seconde et par mètre carré atteignant le télescope issus respectivement de l'étoile et de la planète. L'étoile et la planète n'interfèrent pas car leurs fronts d'onde associés sont décohérents, de sorte que l'intensité de l'étoile se somme à celui de la planète au niveau du détecteur. Par ailleurs, l'information photométrique n'intervient qu'au niveau du détecteur où elle est nécessaire pour modéliser le bruit de Poisson. Ces derniers éléments sont donc pris en compte après la partie diffractive, dans la partie dite *photométrique* du code, programmée en pratique comme une bibliothèque classique de fonctions IDL. Comme la partie diffractive du code est très longue à s'exécuter, il est utile de partir de ses *outputs* afin de générer des jeux de données correspondant à différents cas astrophysiques photométriques (intensité de l'étoile, contraste entre la planète et l'étoile) et différents cas de rotation de champ.

## Paramètres importants décrivant le jeu de données simulé

Détaillons maintenant quels paramètres sont utilisés en pratique pour les données simulées traitées dans le reste de cette thèse notamment.

Nous nous intéressons à un système étoile-exoplanète dont la déclinaison est de  $-45\text{ deg}$  et dont l'angle horaire varie de  $-2\text{ hr}$  à  $+2\text{ hr}$ , c'est-à-dire pour une observation débutant 2 heures avant que le système ne traverse le méridien et se terminant 2 heures après que le système traverse le méridien. Le temps de pose choisi pour chaque longue pose est de  $100\text{ s}$ , (soit 10 DITs), induisant donc pour les 4 heures d'observation  $K = 144$  longues poses. Notons par ailleurs que 100 réalisations de fronts d'onde corrigés sont utilisées pour simuler une longue pose, dans laquelle les *speckles* atmosphériques sont donc assez bien moyennés.

Pour chacune de ces images longue pose, des paramètres optiques évoluent : l'erreur du front d'onde, les déplacements chromatiques associés à l'ADC, les dérives achromatiques lentes, etc. Les principales valeurs d'*input* pour le halo stellaire sont reportées dans le Tableau 2.1, tandis que la Figure 2.1 montre l'évolution de la turbulence en termes de *seeing* et de vitesse de vent, et l'évolution de l'erreur de front d'onde achromatique globale résultant de l'ensemble des instruments situés avant le coronographe.

La simulation ainsi définie résulte en le jeu de données illustré en Figure 2.2.

<b>Système étoile et exoplanète</b>	
intensité de l'étoile en "H2"	$2.80 \times 10^6 \gamma/\text{m}^2/\text{s}$
rapport d'intensité planète / étoile en "H2"	$8.3 \times 10^{-5}$
intensité de l'étoile en "H3"	$2.81 \times 10^6 \gamma/\text{m}^2/\text{s}$
rapport d'intensité planète / étoile en "H3"	$1.0 \times 10^{-5}$
séparations angulaires entre la planète et l'étoile	0".2, 0".5, 1" and 2"
déclinaison, angle horaire initial, angle zénithal (intervalle)	-45 deg , -2 hr, 20–32 deg
<b>Atmosphère et VLT</b>	
vitesse du vent des trois couches turbulentes (intervalle)	7–25 m/s
seeing à 500 nm (intervalle)	0".5–1"
échelle externe des turbulences $\mathcal{L}_0$	25 m
diamètre $D$	8 m
rapport d'obturation dû au miroir secondaire	0.14
point zéro	$2.1 \times 10^{10} e^-/\text{m}^2/\text{s}$
<b>Système d'AO "SAXO"</b>	
type d'analyseur de surface d'onde	Shack-Hartmann
nombre de micro-lentilles associées	40×40
magnitude de l'étoile de guidage	$m_V=8$
longueur d'onde centrale de l'analyseur	0.65 $\mu\text{m}$
temps d'intégration de l'analyseur	0.83 ms
retard de boucle, gain de boucle	1 ms, $\sim 0.47$
bruit de lecture (RON), courant d'obscurité	$0.5 e^-$ , $2 e^-/\text{s}$
efficacité du filtrage spatial (intervalle)	0.3–0.9
<b>Coronagraphe proche infra-rouge</b>	
bande de longueur d'ondes	"H2" : (1.56645 - 1.63922) $\mu\text{m}$
bande de longueur d'ondes supplémentaire (mode DBI)	"H3" : (1.61955 - 1.69478) $\mu\text{m}$
type de coronagraphe	Lyot apodisé
diamètres du masque de Lyot et du Lyot stop	$4\lambda/D$ et $0.96 D$
<b>Aberrations instrumentales</b>	
<i>jitter</i> instrumental	3 mas rms
miroirs M1, M2 et M3 du VLT	11.9 nm, 11.9 nm, 16.6 nm rms
erreur globale achromatique pré-coronographe (hors <i>tilt</i> )	29.2–30.1 nm rms
défocalisation chromatique pré-coronographe	0.8 nm rms
résidus de l'ADC residuals (intervalle)	11.5–12.0 nm rms
<i>tilt</i> achromatique sur le coronagraphe (intervalle)	0.07–0.28 nm rms
défocalisation chromatique post-coronographe	1.8 nm rms
erreur chromatique de front d'onde post-coronographe	7.5 nm rms
<b>Sous-instrument IRDIS</b>	
RON, bruit de <i>flat-field</i>	$10 e^-$ rms, $10^{-3}$ rms
échantionnage spatial	<i>Shannon</i> at 0.95 $\mu\text{m}$
DIT	10 s
temps de pose $\Delta t$	100 s
temps d'observation totale $T$	4 hr
vitesse de rotation de champ, mode pupille stabilisée	0.005–0.010 deg/s
transmission globale (VLT + SPHERE/IRDIS)	0.09

TAB. 2.1: Principales valeurs d'*input* pour la simulation.

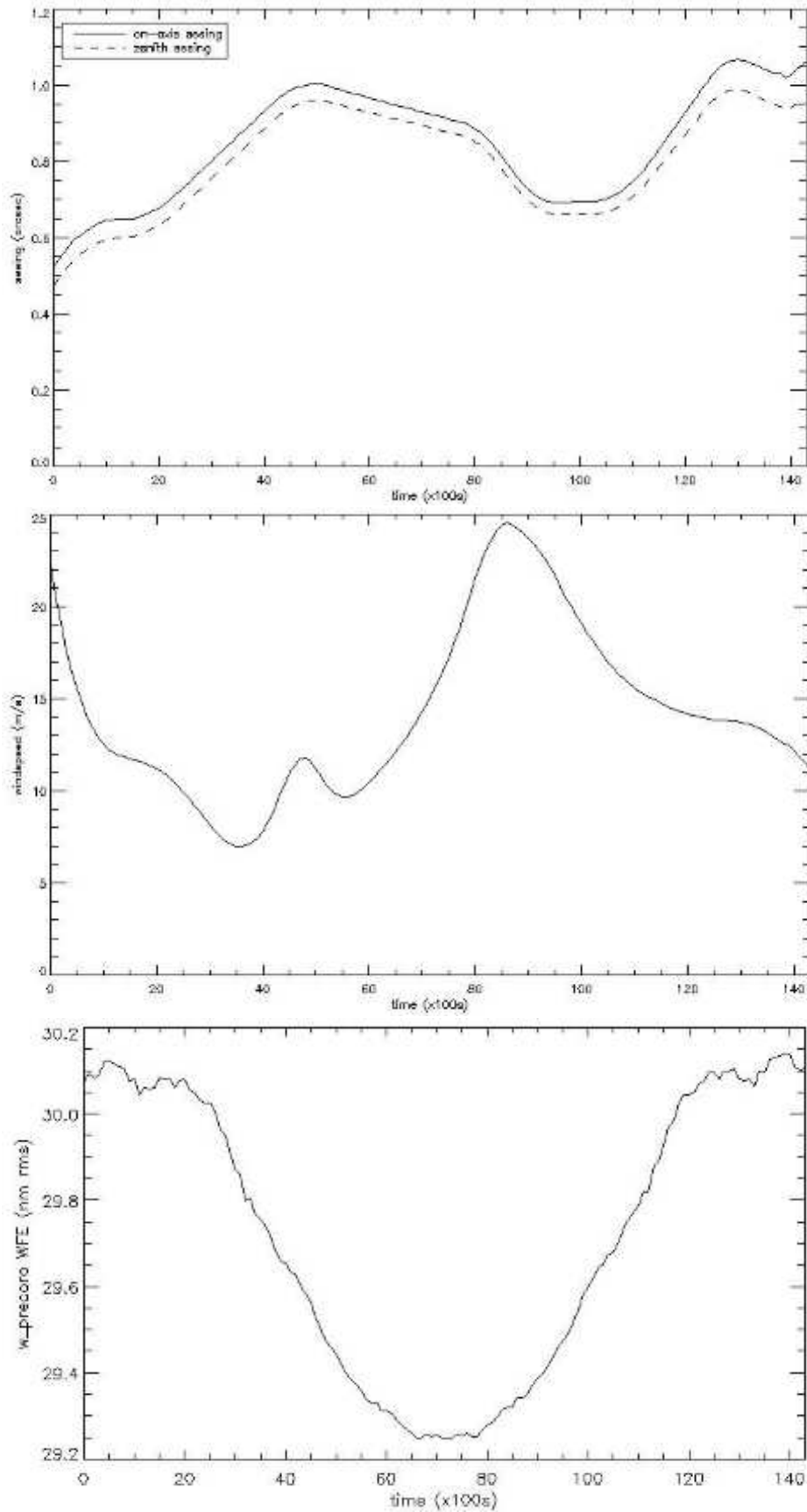


FIG. 2.1: En haut : Evolution temporelle du *seeing* pendant l'observation. Au milieu : évolution temporelle de la vitesse du vent des couches turbulentes de l'atmosphère. En bas : évolution temporelle de l'erreur globale achromatique du front d'onde due aux aberrations optiques pré-conographe (hors *tilt*).

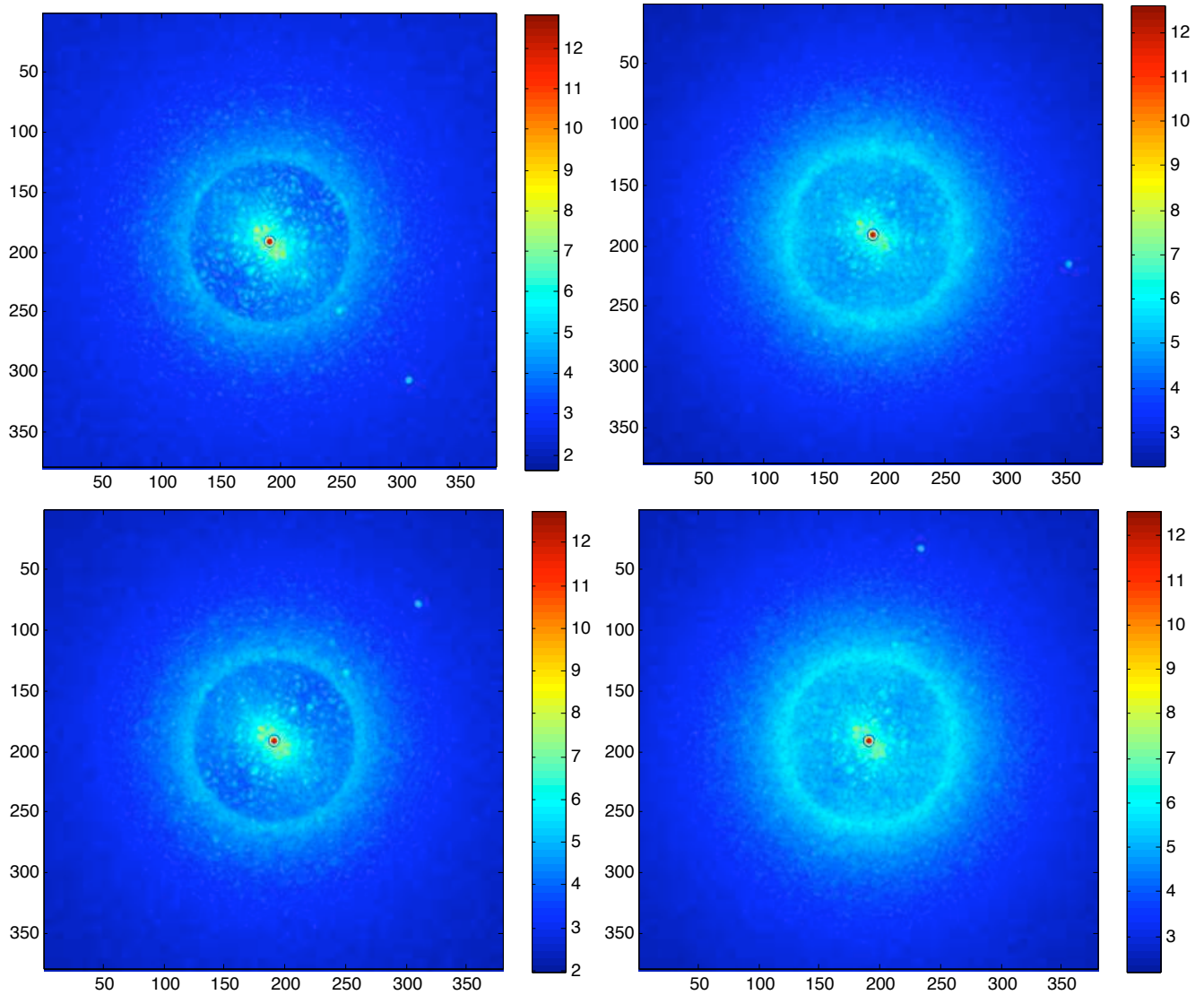


FIG. 2.2: Extrait d'un jeu de données typique constitué d'images longue pose sur lequel seront appliqués les algorithmes de traitement. Dans cette thèse, seuls la/les bandes spectrales et les intensités de l'étoile et de la planète peuvent changer par rapport à ce jeu de données. Celui-ci est simulé en utilisant le *package* SPHERE du *problem-solving environment* CAOS tel que détaillé dans la Section 2.1. Les poses  $k = 1$ ,  $k = 50$ ,  $k = 100$  et  $k = 144$  sont présentées dans le sens de la lecture, les images étant élevées à la puissance 0.2. Un pixel de détecteur correspond à  $0''.0122$  sur le ciel, c'est-à-dire environ  $0.3\lambda/D$ , de sorte que le champ de vue s'étend sur approximativement  $4''.5 \times 4''.5$ . Les quatre planètes simulées le long de la diagonale (de séparations  $0''.2$ ,  $0''.5$ ,  $1''$  et  $2''$ ) sont clairement visibles, sauf celle séparée de  $0''.2$  de l'étoile. Dans ce jeu de données les planètes sont simulées avec une intensité particulièrement élevée (contraste de  $10^4$  seulement entre l'étoile et la planète) afin de les faire apparaître mais l'enjeu de la thèse est de permettre la détection des planètes pour un contraste de  $10^6$  minimum.

## 2.2 Etude statistique qualitative des données simulées en l'absence de planète

### Introduction

- Description qualitative détaillée d'un unique jeu de données simulé

Afin de construire des inférences (estimation d'intensité de la planète, détecteur et probabilité de fausse alarme associée) performantes, il est nécessaire de construire un modèle probabiliste fin des données, celui-ci incluant des contraintes déterministes, comme dans notre cas où il est par exemple usuel de supposer la constance temporelle de la contribution de l'étoile. Nous nous intéressons ici uniquement à des données sans planète puisque la modélisation de la contribution de cette dernière dans un jeu de données est simple et que l'on suppose connue la réponse instrumentale d'une source hors axe. On désignera ainsi par *fond* l'intensité lumineuse mesurée due exclusivement à la lumière de l'étoile.

La sophistication du modèle physique utilisé pour les simulations décrites précédemment rend le jeu de données simulé réaliste, mais rend la simulation de nombreux jeux de données de caractéristiques statistiques différentes très longue à générer. Il est donc en pratique impossible à chercher à construire le modèle statistique uniquement à partir de données simulées et sans hypothèse construite par ailleurs. Quoiqu'il en soit, les données, qui sont ici constituées d'images mesurées durant quelques heures successives, ont comme nous allons le voir ci-dessous une structure statistique très complexe qui implique que les données doivent être étudiées en détail. Listons ainsi les principales caractéristiques des données, caractéristiques que nous observerons plus en détail par la suite.

- Malgré la haute performance de l'AO extrême prévue pour SPHERE, des variations temporelles réalistes des conditions atmosphériques rendent la distribution statistique des données *non stationnaires temporellement*.
- Cependant, une *corrélacion temporelle* importante des images peut être constatée. Ces corrélacions sont suffisamment importantes pour qu'il soit même souvent supposé que la moyenne du fond est constante dans le temps, éventuellement en ayant préalablement sélectionné des images pour lesquelles l'hypothèse est assez bien vérifiée.
- Il existe au moins *deux régimes spatiaux* bien distincts dans le champ de vue : la zone corrigée par l'AO et la zone non corrigée par l'AO. Pour la bande H2 considérée ici, l'AO arrête d'agir au delà d'une séparation de 0'81-0'85.
- Même au sein de chacun de ces deux régimes, les données sont *non stationnaires spatialement*. A l'extrême centre les moments empiriques sont variables radialement et orthoradialement. Dans le champ non corrigé par l'AO, l'intensité mesurée moyenne décroît radialement comme le halo atmosphérique standard dû à l'atmosphère terrestre.
- Malgré la haute performance de l'AO extrême et du coronographe, il existe de fortes *corrélacions spatiales*. En particulier, on observe partout dans le champ des corrélacions spatiales de petite échelle ( $\sim \lambda/D$  sachant que  $\lambda/D \approx 3.4$  pixels) formant des pics locaux d'intensité appelés *speckles*. Ceux-ci sont particulièrement importants car ils ont le même profil que le signal d'intérêt (la planète). Par ailleurs, dans le jeu de données simulé décrit précédemment, la zone la plus interne du champ (jusqu'à 0'5 environ) est corrélée spatialement sur une échelle de l'ordre de 10 à 20 pixels. Ceci peut s'expliquer en particulier par la présence de vents.
- Les considérations précédentes concernent essentiellement la moyenne et une forme qualitative de corrélacion des données. Cependant, la description des données par leurs premiers moments est insuffisante. La présence de *speckles* par exemple ne peut s'expliquer uniquement par une matrice de covariance. En effet, ceux-ci sont caractérisés par des profils de petite taille dont l'intensité globale est particulièrement élevée. Dans le cas le plus simple, pour un front d'onde cohérent, en un point donné et pour une amplitude complexe moyenne nulle, comme nous le reverrons, il est déjà bien connu que la présence de *speckles* reflète une distribution statistique exponentielle décroissante de l'intensité, bien modélisée théoriquement. Plus généralement il n'est donc pas suffisant *a priori*

d'injecter des moments dans une distribution gaussienne. L'étude de la *famille de distribution statistique* elle-même est donc importante, au moins dans un second temps.

L'objet de cette section est ainsi d'effectuer une étude qualitative détaillée du seul jeu de données simulé. Par qualitatif nous entendons qu'hormis la moyenne et l'écart type (ou valeur *rms* pour *root mean square*) empiriques aucun outil statistique quantitatif n'est utilisé. En particulier, deux grandeurs  $X$  et  $Y$  sont dites *corrélées* non pas si l'estimateur de  $\rho = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] / [\sqrt{\text{var}[X]}\sqrt{\text{var}[Y]}]$  est grand mais moins précisément si les valeurs prises par  $X$  et  $Y$  *semblent reliées*, et plus précisément que la structure des valeurs prises par  $X$  se retrouve dans celles prises par  $Y$ , mais pas nécessairement avec la même *dynamique*. Mathématiquement, on pourrait par exemple exprimer ceci par l'existence d'une fonction monotone  $f$  de  $Y$  telle que  $X$  et  $f(Y)$  sont corrélées au sens classique. Mais l'étude restera ici qualitative.

### • Division du champ d'intérêt en anneaux concentriques

Afin de mener cette étude, le champ d'intérêt sera divisé en anneaux concentriques de 10 pixels de large. La division en anneaux est récurrente pour l'étude des performances d'une AO et d'un coronographe car le profil de la réponse de l'instrument à une source ponctuelle est essentiellement à symétrie cylindrique. Celle-ci est aussi très naturelle en présence de rotation de champ induit car la planète suit alors un mouvement circulaire lui permettant de décrire un anneau. 10 pixels est un compromis entre la taille de la planète ( $\sim \lambda/D$  sachant que  $\lambda/D \approx 3.4$  pixels), la variation radiale de l'intensité moyenne (et plus généralement de la distribution statistique) et le nombre de pixels nécessaires à des statistiques significatives.

On désigne un anneau par la séparation angulaire du centre de l'anneau au centre de l'image, de sorte que l'anneau s'étend à 5 pixels de part et d'autre de sa séparation moyenne. On étudiera les anneaux centrés en des séparations multiples de 15 pixels. Un pixel correspondant à  $0''.0122$  sur le ciel, les séparations étudiées sont ainsi proches de  $0''.18$ ,  $0''.37$ ,  $0''.55$ ,  $0''.73$ ,  $0''.92$ ,  $1''.1$ , et ce jusqu'à  $2''.2$ . On peut observer la correspondance entre la séparation exprimée en pixels et celle exprimée en secondes d'arc sur la figure 2.3.

### • Introduction à la notion de *stationnarité*

Afin de construire un modèle probabiliste paramétrique des données qui soit effectivement utilisable en pratique et qui mène à des inférences performantes, il est nécessaire d'introduire une forme de *stationnarité* (spatiale et/ou temporelle) dans les données. Le cas limite de stationnarité en cas indépendant est le cas *indépendant et identiquement distribué (i.i.d.)*, qu'on entend ici à la fois spatialement et temporellement : dans le cas *i.i.d.* total, les intensités mesurées  $\{x_\ell(k), \ell \in \llbracket 1, L \rrbracket, k \in \llbracket 1, K \rrbracket\}$  sont indépendantes pour tout pixel  $\ell$  dans le champ et pour toutes les poses  $k$  et ont une distribution identique et donc en particulier une moyenne et une variance identique.

Le modèle gaussien simple proposé dans le premier chapitre s'écarte de cette hypothèse à travers la dépendance de la moyenne du fond en le pixel  $\ell$  : sous l'hypothèse nulle  $H_0$ , rappelons que ce modèle (3.1) s'écrivait pour un pixel  $\ell$  :

$$x_\ell(k) = d_\ell + \epsilon_\ell(k), \quad \ell = 1, \dots, L \quad k = 1, \dots, K \quad (2.1)$$

avec  $d_\ell$  le fond déterministe mais inconnu dû à l'étoile et  $\epsilon_\ell(k) \sim \mathcal{N}(0, \sigma^2)$  un bruit gaussien indépendant dans le temps et dans l'espace et de variance constante dans le temps et dans l'espace.

En conservant l'hypothèse de gaussiannité et d'indépendance spatiale et temporelle il existe deux façons d'introduire une non-stationnarité temporelle du fond : introduire une dépendance temporelle de la moyenne (" $d_\ell(k)$ ") ou bien l'introduire au niveau de la variance (" $\sigma^2(k)$ "). Concentrons-nous sur la première possibilité. En présence d'une planète, il n'est alors plus possible de distinguer la planète d'un fond variable de façon indéterminée à la fois en fonction du temps et de l'espace : un modèle où  $\alpha$  désigne



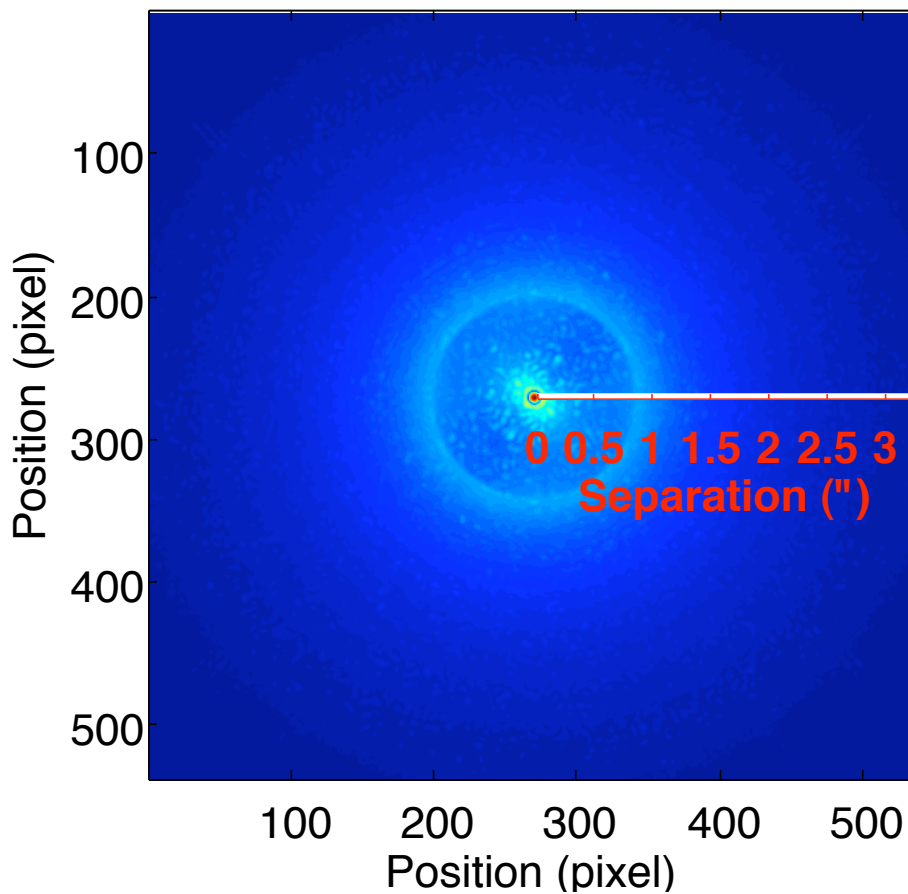


FIG. 2.3: Moyenne des images élevée à la puissance 0.2. La *séparation* désigne la distance entre le centre de l'image (position de l'étoile, axe optique) et le point de l'image considéré, *i.e.* la distance radiale. La première planète recherchée est située à  $0''.2$ , *i.e.* environ  $5\lambda/D$ .

la réponse de la planète pour un profil  $p_\ell(k, r)$  en  $(\ell, k)$  et où  $\delta_\ell(k)$  désigne la contribution de l'étoile en  $(\ell, k)$  est équivalent au modèle où les contributions de l'étoile et la planète sont données respectivement par  $\delta_\ell(k) + \alpha' p_\ell(k, r)$  et  $\alpha - \alpha'$ .

Il est néanmoins possible de paramétrer la variation temporelle de la moyenne du fond  $\delta_\ell(k)$  afin de rendre le problème identifiable à nouveau. La paramétrisation la plus simple consiste à supposer qu'il existe un *coefficient inconnu fonction du temps*  $\beta(k)$  constant pour tout pixel  $\ell$  :  $\delta_\ell(k) = \beta(k)d_\ell$ , de sorte que  $K$  paramètres sont ajoutés à l'ensemble de paramètres définis dans le modèle initial (2.1).  $\beta(k)$  n'est pas identifiable de  $d_\ell$  à un coefficient multiplicatif près, de sorte qu'il est nécessaire de fixer une contrainte telle que  $\|\mathbf{d}\| = 1$  par exemple. Cependant, en limitant la paramétrisation à l'expression  $\boldsymbol{\delta}(k) = \beta(k)\mathbf{d}$ , le problème est identifiable mais peut rester relativement trop peu contraint, rendant éventuellement l'estimation de  $\alpha$  moins bonne que dans le modèle initial (2.1). Quoiqu'il en soit, on peut montrer que le Hessien du critère à minimiser n'est pas défini positif sur le sous-espace défini par les paramètres  $\mathbf{d}$  et  $\beta$  de sorte que le critère à minimiser n'est alors plus convexe par rapport à l'ensemble des paramètres inconnus, contrairement au critère (3.4). L'espace des paramètres lui-même n'est pas convexe sous la contrainte  $\|\mathbf{d}\| = 1$ . Pour l'une ou l'autre de ces raisons la minimisation du critère associé à cette modélisation devient donc un problème compliqué et est probablement impossible à appliquer en pratique pour toutes les positions initiales  $r$  du champ d'intérêt.

La démarche contraire à laisser le paramètre  $\beta(k)$  totalement non contraint consisterait à le fixer. Ceci est éventuellement possible si la *structure temporelle* de la moyenne des données est accessible par

des mesures et étalonnages indépendants. Dans le cas qui nous intéresse, il est par exemple concevable que le *seeing* (surtout) et la vitesse des vents dans les couches atmosphériques les plus turbulentes (si des instruments le permettent) soient mesurables en fonction du temps et que l'étalonnage des aberrations optiques dûs au mouvement du télescope soit fiable. Ceux-ci ont été tous trois illustrés dans la figure 2.1, pour le cas de simulation qui a engendré le jeu de données que nous allons étudier ici et utiliser dans l'ensemble de la thèse. Cependant, même en présence de bonnes mesures parallèles il est encore nécessaire d'établir, par étalonnage ou modélisation physique, la correspondance causale ou statistique –mais supposée déterministe si l'on souhaite que  $\beta(k)$  soit fixé– entre ces mesures et le terme  $\beta(k)$ .

Plus généralement, nous avons vu à travers cet exemple de construction d'un modèle statistique qu'apporter de l'information parallèle pouvait éventuellement rendre les inférences d'intérêt plus performantes, dans la mesure en tout cas où cette information est convenablement prise en considération. Un des objectifs majeur de cette section est ainsi d'observer l'effet des variations temporelles du *seeing*, de la vitesse du vent (et éventuellement des aberrations optiques) illustrés en figure 2.1 sur les données simulées. En présence d'un seul jeu de simulations, il ne s'agit pas de dériver une relation systématique entre ces paramètres et une paramétrisation d'un modèle statistique mais de dégager une extension raisonnable de modèle statistique.

Notons cependant que l'étude statistique de l'évolution temporelle (ou spatiale) des paramètres d'une distribution à partir d'un jeu de données est délicate car cette évolution paramétrique est partiellement indiscernable de la variabilité naturelle des données. Grossièrement, comme pour le problème d'inférence qui nous intéresse, si la variation temporelle de la moyenne des données n'est pas notablement supérieure à l'écart type temporel des données il est difficile d'obtenir une estimation correcte de la variation déterministe de la moyenne.

## Description statistique des données

Au delà de chercher à déceler l'existence de structures sous-jacentes aux données  $\{x_\ell(k), \ell \in \llbracket 1, L \rrbracket, k \in \llbracket 1, K \rrbracket\}$ , l'analyse qui suit a pour objectif de vérifier *qualitativement* et sélectionner des hypothèses permettant de construire un modèle statistique pertinent et applicable en pratique. Nous vérifierons notamment que supposer en priorité la dépendance spatiale du fond et non sa dépendance temporelle n'est pas inacceptable.

### • Analyses de moyennes des données

Dans un premier temps, deux analyses dites de *moyennes* des données sont appliquées aux données. Par “moyenne” on entend que la structure (spatiale ou temporelle) est perdue.

**En un pixel donné** La première consiste à étudier la distribution empirique des données *pixel par pixel*, c'est-à-dire la distribution de  $\{x_\ell(k), k \in \llbracket 1, K \rrbracket\}$  pour chaque pixel  $\ell$ . Plus spécifiquement, la moyenne temporelle  $\bar{x}_\ell$  est calculée pour tout  $\ell$ , formant ainsi l'image moyenne  $\bar{\mathbf{x}} = K^{-1} \sum_k \mathbf{x}(k)$  habituelle (figure 2.4 en haut), ainsi que l'écart type empirique  $s_\ell = (K - 1)^{-1} \sqrt{\sum_k (x_\ell(k) - \bar{x}_\ell)^2}$  (figure 2.5 en haut). Des structures spatiales se dégagent de ces images “intégrées temporellement”.

Par ailleurs, ces deux fonctions de  $\ell$  sont chacune étudiée par anneau, en indiquant ainsi de la même façon, mais sur un graphe fonction de la seule séparation de l'anneau cette fois, la moyenne empirique et l'écart type indiqué sous forme de barre d'erreur. Ainsi, la moyenne empirique par anneau de l'image moyenne est la moyenne de l'ensemble des  $\{\bar{x}_\ell, \ell \in \mathcal{A}(\rho)\}$ , où  $\mathcal{A}(\rho)$  désigne l'ensemble des pixels appartenant à l'anneau de séparation moyenne  $\rho$ . De même, l'écart type de l'image moyenne est l'écart type des  $\{\bar{x}_\ell, \ell \in \mathcal{A}(\rho)\}$ . Ceux-ci sont représentés sur le graphe du bas de la figure 2.4. On procède de la même façon pour l'image des écarts types temporels, calculés on le rappelle pixel par pixel, ce qui conduit au graphe en bas de la figure 2.5.

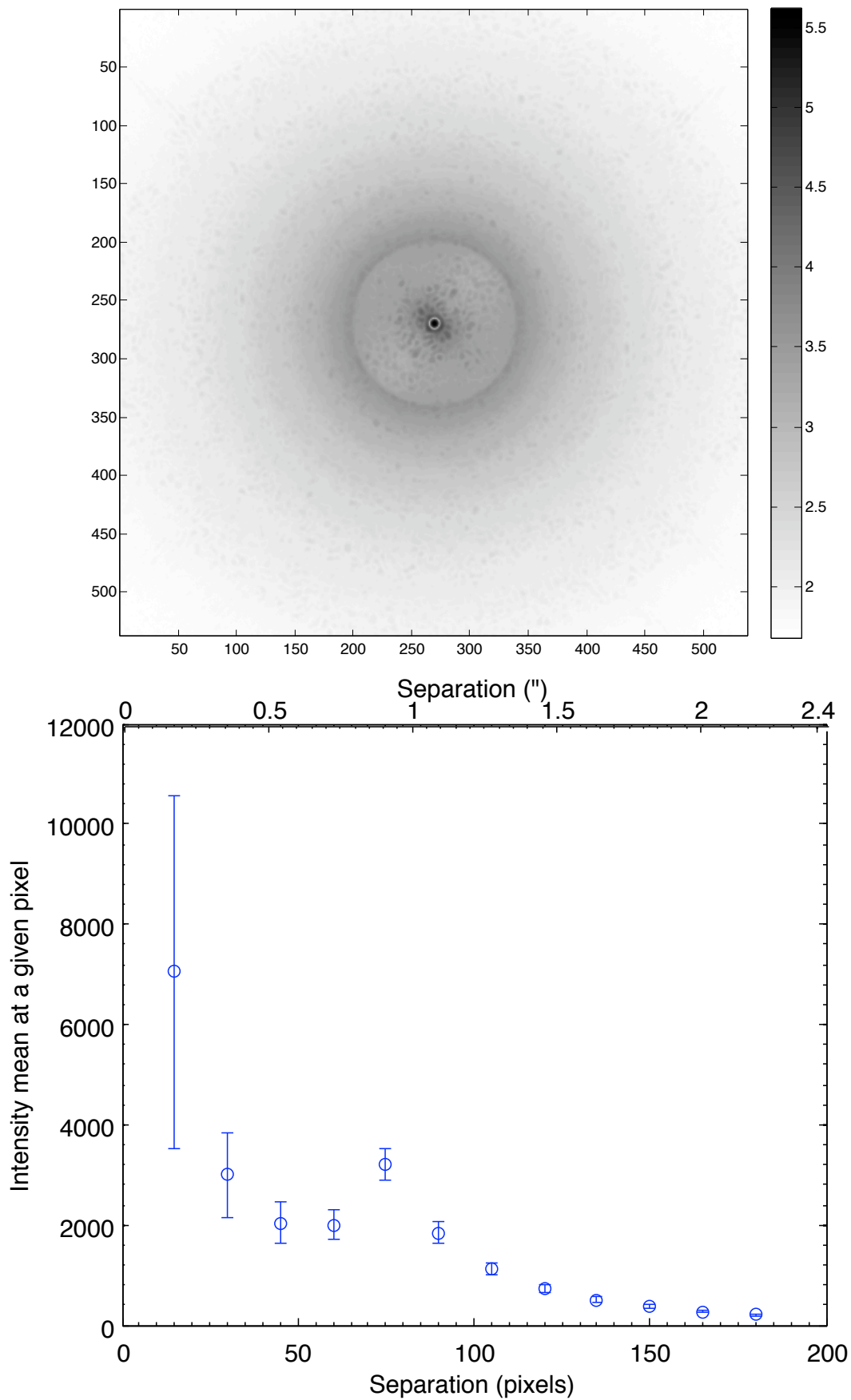


FIG. 2.4: Moyenne temporelle des données  $\{x_\ell(k), k \in \llbracket 1, K \rrbracket\}$ . En haut : représentée pour chaque position  $\ell$  en échelle logarithmique de base 10. En bas : représentée par sa moyenne (cercle) et son écart type (barre d'erreur) pour l'ensemble des pixels appartenant à un anneau.

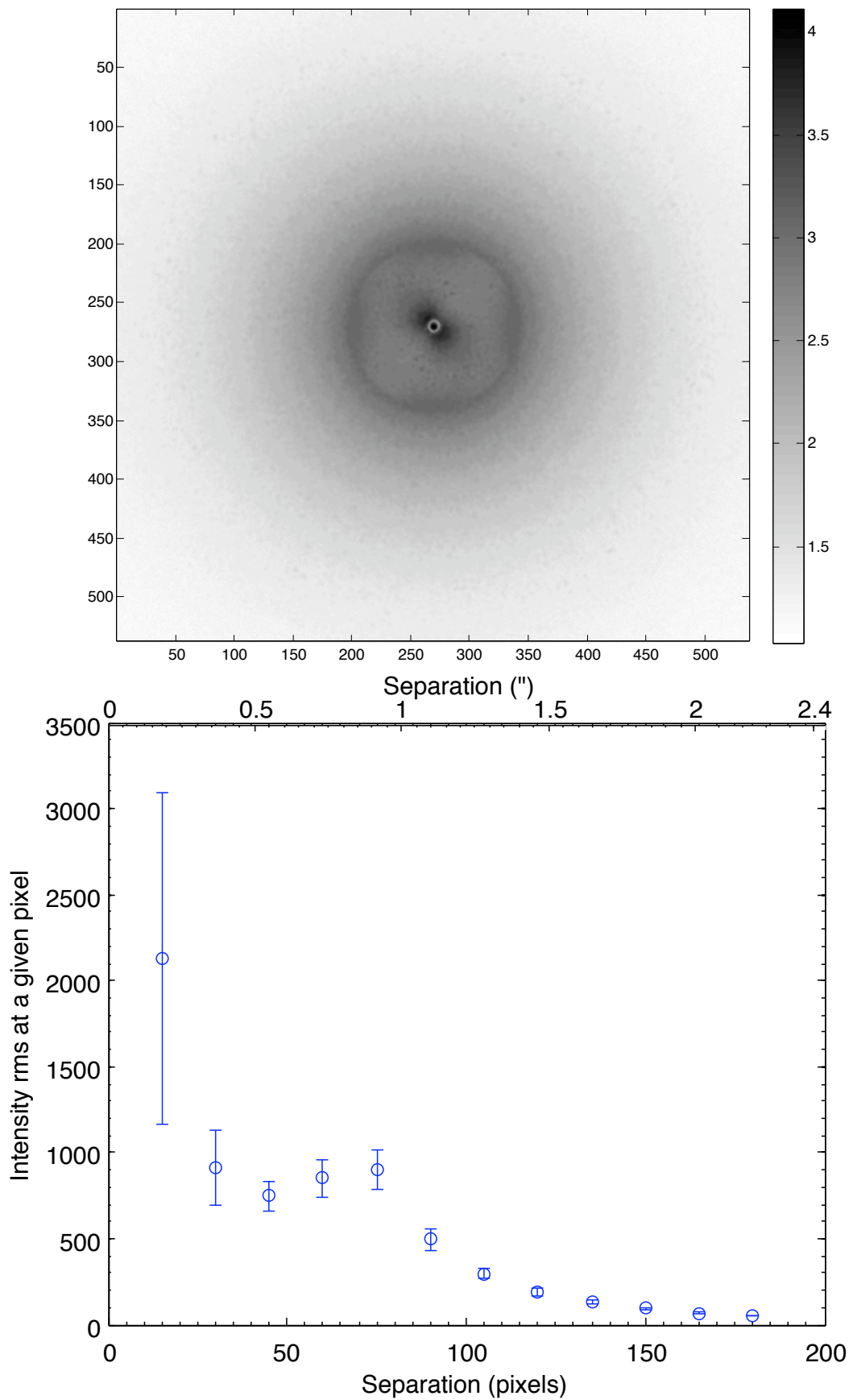


FIG. 2.5: Ecart type empirique temporel des données  $\{x_\ell(k), k \in \llbracket 1, K \rrbracket\}$ . En haut : représenté en échelle logarithmique de base 10. En bas : représenté par sa moyenne et son écart type par anneau.

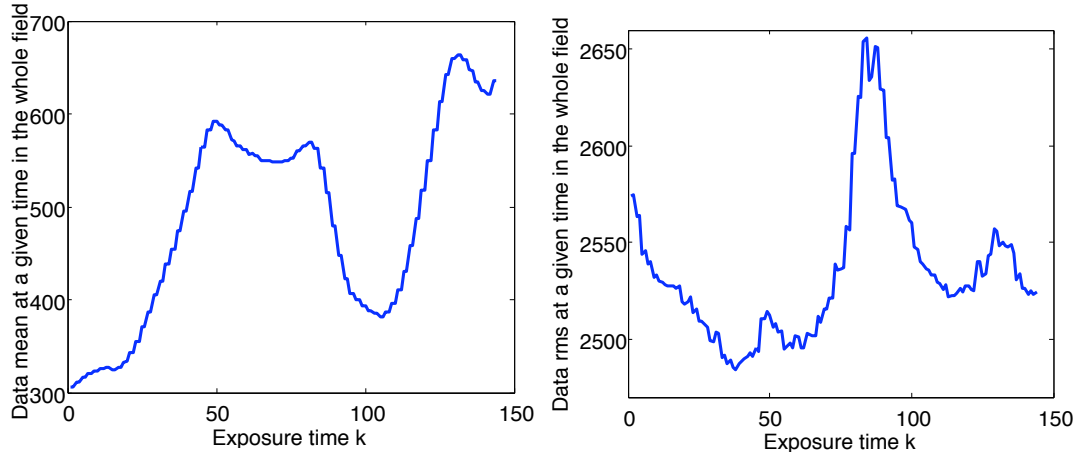


FIG. 2.6: Statistiques de la distribution pose par pose des données  $\{x_\ell(k), \ell \in \llbracket 1, L \rrbracket\}$  dans tout le champ. A gauche : moyenne empirique (nettement corrélée au *seeing* présenté sur la figure 2.1). A droite : écart type (nettement corrélé à la vitesse du vent présentée sur la figure 2.1).

**En un instant donné** On applique ensuite l’analyse *réciroque* de l’analyse précédente, intervertissant de façon totalement symétrique le rôle du temps et de l’espace, l’espace étant simplement limité à un anneau concentrique comme nous l’avons introduit plus haut. Dans un premier temps, l’étude concerne ainsi la distribution des données *pose par pose* au sein d’un anneau donné (figures 2.7 et 2.8 en haut). Cette fois ce ne sont donc plus des structures spatiales qui se dégagent mais des structures temporelles<sup>1</sup>. Les moyenne et écart type empiriques résultants sont alors de même que précédemment “moyennés”, mais cette fois temporellement et non plus spatialement (figures 2.7 et 2.8 en bas).

### • Analyse locale des données

A cause des structures apparues dans les analyses de moyennes (et bien présentes sur les données non intégrées), il n’est pas possible de considérer les données  $\{x_\ell(k), \ell \in \llbracket 1, L \rrbracket, k \in \llbracket 1, K \rrbracket\}$  comme étant *i.i.d.*. Il n’est ainsi pas possible *a priori* de modéliser les données comme étant *i.i.d.* de moyenne et écart type indiqués par les statistiques calculées jusqu’ici. L’objet de l’analyse dite *locale* des données est ainsi de décrire plus en détail, mais toujours qualitativement, chaque anneau.

En particulier, la structure spatiale et la distribution des données  $\{x_\ell(k), \ell \in \mathcal{A}(\rho)\}$  sont décrites pour  $k \in \{1, 50, 100, 144\}$ , c’est-à-dire que l’on étudie des poses séparées d’environ 1h20 qui correspondent à quatre régimes atmosphériques différents :

- $k = 1$  : très faible *seeing* et fort vent
- $k = 50$  : fort *seeing* et faible vent
- $k = 100$  : faible *seeing* et fort vent
- $k = 144$  : très fort *seeing* et faible vent

Les figures associées étant beaucoup plus nombreuses que celles de l’analyse de moyennes et ajoutant relativement peu d’information, elles sont reportées en annexe C.1. Chaque page de cette annexe est destinée à reprendre l’ensemble des informations jugées pertinentes concernant un anneau donné, et ce pour chaque anneau.

<sup>1</sup>La figure 2.6 intégrée dans tout le champ et non dans un anneau montre d’emblée les rôles majeurs joués par le *seeing* et la vitesse de vent dans la structure temporelle des données. Ces derniers sont indiqués sur la figure 2.1.

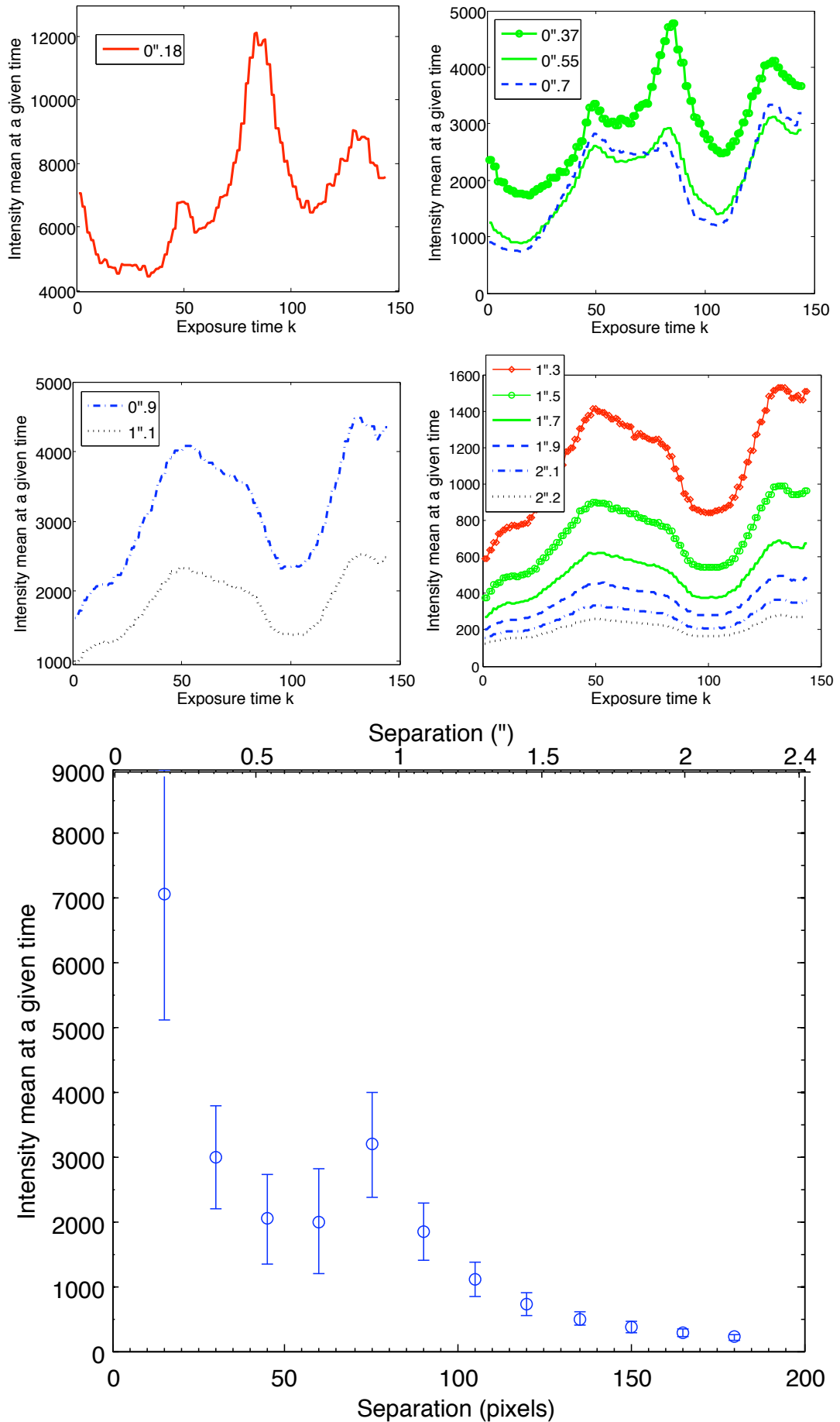


FIG. 2.7: Moyenne empirique de la distribution pose par pose des données  $\{x_\ell(k), \ell \in \mathcal{A}(\rho)\}$  pour chaque anneau  $A(\rho)$ ,  $\rho = 0''.18, 0''.37, \text{etc.}$  En haut : représentée pour chaque pose  $k$  (en haut : zone corrigée par AO, en bas : zone non corrigée par AO). En bas : représentée par sa moyenne et son écart type temporels.

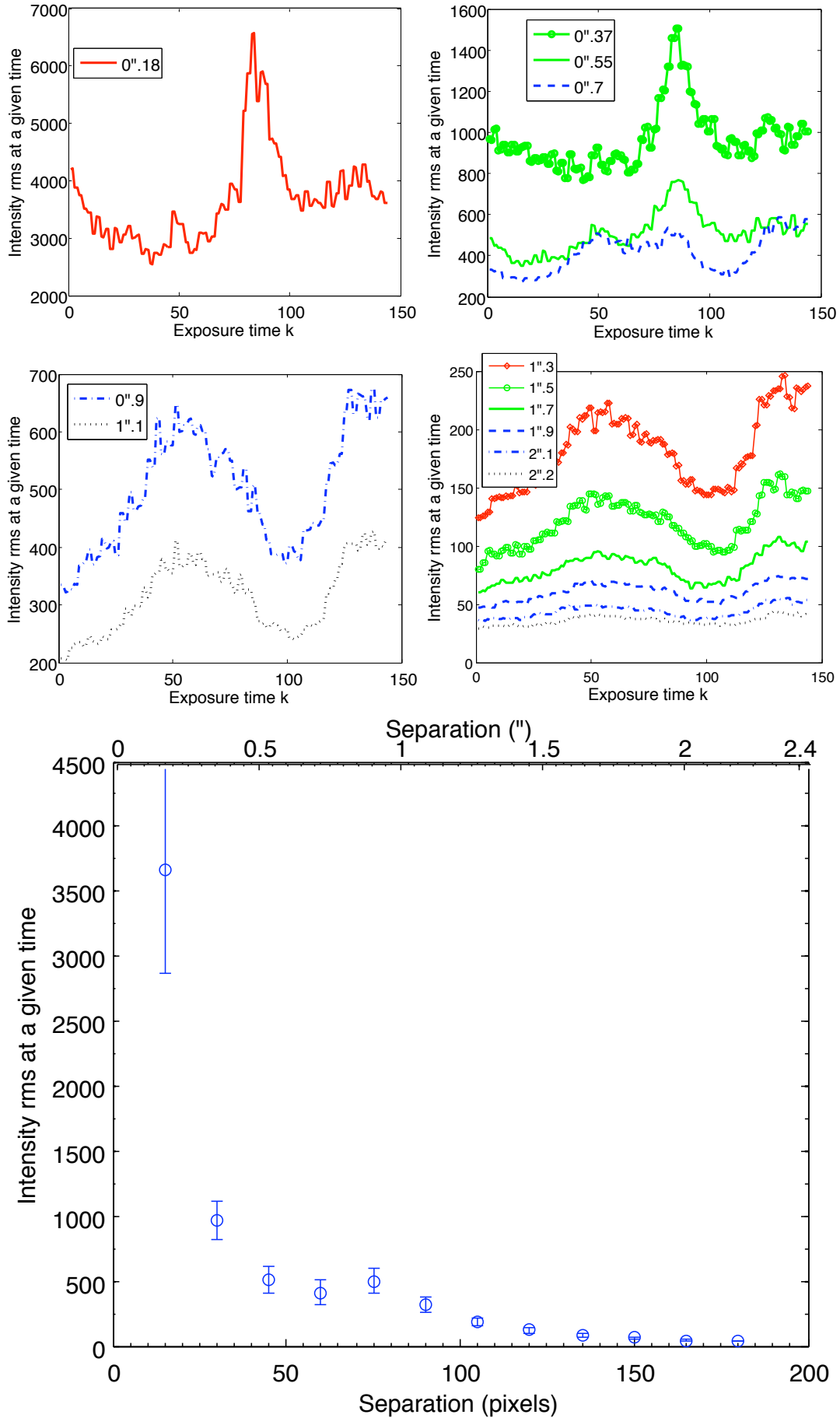


FIG. 2.8: *Ecart type* empirique de la distribution pose par pose des données  $\{x_\ell(k), \ell \in \mathcal{A}(\rho)\}$  pour chaque anneau  $A(\rho)$ ,  $\rho = 0''.18$ , etc. En haut : représenté pour chaque pose  $k$  (en haut : zone corrigée par AO, en bas : zone non corrigée par AO). En bas : représenté par sa moyenne et son écart type temporels.

## Discussion des analyses précédentes

### • Résumé

L'étude des figures présentées plus haut dans cette section pour l'analyse des moyennes des données et l'étude des figures présentées dans l'annexe C.1 pour l'analyse locale des données permet de discuter différents points fondamentaux dans la construction du modèle probabiliste des données. Dans un premier temps, nous allons décrire les structures spatiales et temporelles qui sont apparues ainsi que leurs éventuelles relations avec les paramètres physiques variables dans le temps, les paramètres atmosphériques en particulier. De la même façon, nous allons voir que supposer les données *i.i.d.* dans le temps plutôt qu'*i.i.d.* dans l'espace est important en pratique et n'est pas aberrant d'un point de vue statistique.

Ensuite, nous allons discuter la problématique plus délicate du choix de la famille de distribution. Compte-tenu des corrélations spatiales et temporelles fortes existant dans les données simulées, il est difficile de conclure quant à la famille de distributions à choisir pour modéliser les données. Le jeu de données est cependant suffisant pour conclure qu'il est préférable de choisir un modèle Gaussien plutôt que Poissonien car la variance empirique est de l'ordre de 100 fois plus importante que la moyenne dans la zone corrigée par AO. Il semble de plus assez clair que la distribution à queue lourde bien connue pour les *speckles* se retrouve bien dans le jeu de données.

### • Structures spatiales et temporelles

**Structure spatiale radiale** La limitation radiale à  $0''81 - 0''85$  de la zone de correction par l'AO se manifeste très clairement sur toutes les images individuelles et donc sur les figures moyennes. Dans le graphe de la moyenne empirique (2.4 ou 2.7) on observe une décroissance de l'intensité moyenne jusqu'à la séparation de  $0''92$  à laquelle la limite de correction est franchie. Après ce seuil, on observe la décroissance moyenne d'un halo atmosphérique en l'absence d'AO.

A l'intérieur de la zone corrigée par AO, à part l'anneau situé à  $0''18$ , à cause des fortes variations spatiales ou surtout temporelles (mais qui peuvent alors être réduites en classant les images en paquets *i.i.d.*) les moments des données ne s'excluent pas à plus de 1 ou  $2\sigma$ . Dans tous les anneaux on observe des différences d'un facteur 2 ou 3 dans les cartes d'écart type. L'anneau à  $0''73$  dont l'extrémité est à 3 pixels du début de la zone de non-corréction par AO a un profil de corrélation en croix. A part cet anneau, tous ceux de la zone de correction par AO ont des profils de corrélation radiale relativement proche, qui correspond en fait à la direction des vents.

La division de cette zone en anneaux n'est pas très clairement justifiée d'un point de vue statistique à cause de la force des effets asymétriques (vent). Cependant, elle est très pratique quoiqu'il en soit.

Au delà de  $0''92$ , un profil radial à symétrie cylindrique est très clairement visible. Il s'agit du halo atmosphérique usuel.

Ainsi, il est bien possible de diviser le champ d'intérêt en deux grands régimes spatiaux : la zone corrigée par AO dont le coeur et la partie externe ont des structures partiellement à symétrie cylindrique et partiellement à symétrie axiale (pour ce qui concerne les corrélations à grande échelle spatiale dûes au vent et aux aberrations à bas ordre) qui rendent une division en anneaux non totalement satisfaisante mais toujours acceptable, et la zone non corrigée par AO dont les propriétés statistiques sont bien à symétrie cylindrique.

**Structure temporelle** Les moyennes pose par pose présentées dans la figure 2.7 (avec une meilleure dynamique ces mêmes graphes sont présentés en annexe C.1) mises en vis-à-vis des graphes indiquant la variation temporelle du *seeing* et du vent (figure 2.1) indiquent clairement que la moyenne par anneau est corrélée à la vitesse du vent et/ou le *seeing* selon la séparation considérée. Plus précisément, pour ce jeu de données à  $0''18$  la moyenne est très bien reliée à la vitesse du vent (avec une petite contribution du *seeing* comme on peut le voir à la fin de l'observation), et à mesure que l'on s'écarte de l'axe optique la contribution relative du *seeing* à la moyenne augmente et celle du vent diminue. Après la limite de correction de l'AO, la moyenne est presque exclusivement, mais très clairement, indiquée par le *seeing*.



Ces conclusions qualitatives ne sont pas rejetées par une observation des profils temporels des pixels considérés individuellement.

Cependant, une distribution statistique n'est pas déterminée par sa seule moyenne. Les écarts type pose par pose sont présentés dans la figure 2.8 (et montrés dans l'annexe C.1 sous forme de variance empirique). Ceux-ci sont *plutôt* corrélés à la moyenne dans la zone corrigée par AO, et plus corrélés encore dans la zone non corrigée.

L'étude du second moment est moins concluante que celle du premier mais il semble acceptable de considérer que les régimes temporels définissant la distribution statistique sont définis à partir de l'évolution de paramètres physiques (*seeing* et vent), la dépendance entre les paramètres physiques et les moments de la distribution dépendant de la séparation angulaire de la position spatiale  $\ell$  considérée.

Notons que selon la force du vent et la contribution relative du *seeing*, la vitesse du vent et éventuellement les aberrations quasi-statiques, les conclusions pourraient varier. En particulier, les contributions quantitatives du *seeing* et de la vitesse du vent seules méritent d'être précisées.

### • Symétrie entre les dimensions spatiale et temporelle

Dans le modèle gaussien initial (3.1) nous avons supposé que le fond  $\mathbf{d}$  varie spatialement mais est constant en fonction du temps et que le bruit est indépendant spatialement et temporellement. Afin de rendre l'utilisation de ce modèle appropriée il peut être avantageux de diviser l'ensemble des images  $\{\mathbf{x}(k), k \in \llbracket 1, K \rrbracket\}$  en des sous-ensembles vérifiant chacun bien l'hypothèse d'images *i.i.d.*. Une extension naturelle de ce modèle encore non contraint spatialement est de tenir compte des corrélations spatiales tout en continuant à diviser le jeu de données afin de bien satisfaire l'hypothèse de stationnarité temporelle.

L'objet de ce paragraphe est d'indiquer des éléments qualitatifs décrivant la symétrie ou l'asymétrie entre les dimensions temporelle et spatiale. Une motivation sous-jacente est d'observer s'il serait préférable de construire un modèle *i.i.d.* temporellement, de moyenne variable spatialement et corrélé spatialement ou un modèle *i.i.d.* spatialement, de moyenne variable temporellement et corrélé temporellement, tout ceci *a priori* sans paramétrisation particulière du moins de la moyenne. Dans le premier cas, les poses  $k$  seraient à classer en des ensembles  $\mathcal{K}_i$  pour lesquels les images  $\mathbf{x}(k)$  ou  $\{x_\ell(k), \ell\}$  sont *i.i.d.* pour tout  $k \in \mathcal{K}_i$ , tandis que dans le second cas les pixels  $\ell$  du champ d'intérêt seraient à classer en des ensembles  $\mathcal{L}_i$  pour lesquels les données  $\{x_\ell(k), k\}$  sont *i.i.d.* pour tout  $\ell \in \mathcal{L}_i$ . Cependant, la variabilité inhérente au processus aléatoire de génération des données rend l'analyse délicate et nous vérifions simplement que l'une ou l'autre des approches n'est pas inacceptable.

- La *variabilité moyenne* temporelle peut être *quantifiée* par l'écart type empirique pixel par pixel, représenté en image puis synthétisé radialement sur la figure 2.5. La variabilité moyenne spatiale peut être quantifiée par l'écart type empirique pose par pose, représenté exhaustivement puis synthétisé sur l'ensemble des poses sur la figure 2.8. Même en tenant compte en partie de la diminution de variabilité qui découlerait d'une subdivision des données, on constate que : à 0''2 la variabilité temporelle est presque toujours ("toujours" référant à la probabilité associée à  $1\sigma$ ) plus élevée que la variabilité spatiale (la moyenne de la variabilité temporelle est presque 2 fois plus élevée que la moyenne de la variabilité spatiale), à 0''4 la variabilité temporelle est identique à la variabilité spatiale et au delà la variabilité temporelle est plus faible que la variabilité moyenne (facteur de l'ordre de 2 pour toutes les séparations). Dans la zone corrigée par AO, jusqu'à 0''55 inclus, les variabilités spatiale et temporelle pourraient être diminuées respectivement en divisant le champ d'intérêt en deux zones à symétrie axiale : la zone à fort flux (dans la direction du vent, qui ne change pas durant toute l'observation) et le reste, et en divisant la durée d'observation en la période de fort vent et le reste (voir les figures détaillées de l'annexe C.1). Par ailleurs, on peut aussi observer quantitativement les facteurs de variation absolue de l'intensité mesurée : ceux-ci sont de l'ordre de 3 jusqu'à 0''4, de l'ordre de 2 jusqu'à 1'', et de l'ordre de 1.5 ou 2 plus loin, le tout en temporel aussi bien qu'en spatial. Il ne semble pas possible à partir de ces éléments de conclure quant au choix de modélisation.

- La corrélation, en tant que dépendance de l'intensité mesurée pour un instant  $k$  et un autre instant  $k'$  pour la corrélation temporelle, est un élément important de choix de modèle. La *corrélation à petite échelle* est importante car tant spatialement que temporellement elle génère des profils proche de celui du signal recherché. Spatialement, à cause des *speckles* quasi-statiques et à cause de la distribution statistique à queue lourde décrivant les speckles, des profils identiques à la planète apparaissent et disparaissent sur les images. Temporellement, il faut un temps de l'ordre de quelques poses pour que la planète se soit suffisamment déplacée pour ne plus se superposer à elle-même. De même, si temporellement les corrélations atmosphériques et de correction AO font apparaître des temps de corrélation de cet ordre de grandeur il serait intéressant de les prendre en considération. Il serait possible que ceci existe temporellement mais il est certain qu'il existe spatialement, de sorte que le modèle *i.i.d.* temporellement et corrélé spatialement n'est du moins pas inacceptable.
- La *corrélation à grande échelle* est quant à elle importante non nécessairement pour être modélisée par une matrice de covariance (à moins de convenablement la paramétrer, mais cela induirait la création de trop de paramètres non contraints), mais surtout pour définir des zones *i.i.d.* (s'il existe une structure physique sous-jacente qui le permet).
- Afin de définir des zones (temporelles ou spatiales) *i.i.d.*, il est important de disposer de *mesures indépendantes* qui permettent de *classer statistiquement les données* sans utiliser ces dernières pour le choix de classification. En effet, dans la présente analyse nous pouvons constater que les problématiques sont trop intriquées pour être traitées à partir du seul jeu de données par ailleurs trop limité. En particulier, la variabilité reflète aussi bien à la fois la variation des moments que la nature aléatoire du processus. La corrélation entre les moments et des paramètres physiques sont des signes de l'interprétation à accorder à la variabilité mais il faut veiller à ne pas *surinterpréter* les données. Ainsi, l'observation de relations (causales ou statistiques) entre des paramètres physiques –mesurables par ailleurs– et des statistiques simples est très utile pour classer le jeu de données. Dans la première modélisation, le classement est effectué selon la composante temporelle, tandis que dans la seconde le classement est effectué selon la composante spatiale sur le plan focal de l'instrument SPHERE. Dans le second cas il est nécessaire de construire une relation entre des paramètres physiques atmosphériques et le plan focal. L'existence d'un tel lien est envisageable pour les paramètres physiques de direction<sup>2</sup> et de vitesse du vent, mais semble encore plus difficile à mettre en évidence qu'une relation systématique entre les valeurs de *seeing*, vitesse de vent et aberrations instrumentales d'un côté, et moyenne de distribution statistique de l'autre. La difficulté de diviser un ensemble de poses *vs* un ensemble de pixels doit par ailleurs être mesurée en fonction de la taille de ces zones, mise en perspective avec les tailles caractéristiques de corrélation à petite échelle.
- Comme nous l'avons mentionné, afin de tenir compte de corrélations (spatiales ou respectivement temporelles) il est nécessaire de disposer d'une *famille de distributions* adaptée. D'un point de vue physique, dans le premier cas, la corrélation spatiale est prise en compte par un modèle optique de formation d'une image à partir d'un front d'onde incident fixé. Dans le second, la corrélation temporelle doit être prise en compte par un modèle atmosphérique et associé à un modèle optique de formation d'image, le dernier modèle étant éventuellement déterministe.
- Il existe un point pour lequel les dimensions temporelle et spatiale jouent un rôle indiscutablement asymétrique : le *processus d'acquisition des données*. Un jeu de données minimal ("atome") est une image acquise à un instant  $k$ , c'est-à-dire un vecteur  $\mathbf{x}(k)$  constitué de  $L$  pixels. Acquérir plus de données consiste à acquérir un ensemble d'images  $\{\mathbf{x}(k), k \in \llbracket 1, K \rrbracket\}$  avec  $K$  croissant. Il est ainsi naturel de diviser l'ensemble des données en un ensemble de "jeux de données minimaux" et d'observer les propriétés des inférences lorsque le nombre d'éléments d'une classe donnée ou de l'ensemble des classes ( $K$ ) tend vers l'infini.

---

<sup>2</sup>Voir la figure C.1 de l'annexe C.1.

### • Famille de distribution

La présence de *structures* spatiales et temporelles a été montrée assez significativement. Réciproquement, il est donc délicat d'étudier à partir de ces mêmes données la famille de distributions (variabilité due à l'aléa du processus aléatoire pour un ensemble de paramètres fixés) sous-jacente à ces données. Néanmoins, dans le cadre de l'hypothèse de données identiquement distribuées dans le temps (éventuellement au sein d'un ensemble de poses donné), il est possible de regarder plus en détail les données correspondant à plusieurs poses ( $k \in \{1, 50, 100, 144\}$ ) fixées. Ceci a été réalisé pour chaque anneau, comme détaillé plus haut dans cette section. Ainsi, la moitié inférieure des figures présentées dans l'annexe C.1 présente dans un premier temps les données  $\{x_\ell(k), \ell \in \mathcal{A}(\rho)\}$  pour  $k \in \{1, 50, 100, 144\}$  pour les figures classées respectivement de gauche à droite, et dans un second temps les histogrammes associés, dans lesquels la structure spatiale est perdue mais au profit d'un classement des données en fonction de leurs valeurs. Dans le cadre d'un modèle où la moyenne temporelle des données est supposée dépendre de la position  $\ell$  sans contrainte, la moyenne peut dépendre de  $\ell$ , mais conditionnellement à la moyenne les données sont sensées être identiquement distribuées dans l'anneau.

- Pour les séparations supérieures à  $0''.5$ , les *régimes temporels* ont au minimum une légère influence sur la “forme” (paramètres hors position et échelle) des histogrammes. En particulier, en présence d'un faible vent et d'un très faible *seeing* ( $k = 1$ ), les histogrammes semblent *plutôt* avoir une queue à droite plus lourde que pour les autres régimes. A l'autre limite (fort vent et fort *seeing*), les histogrammes ont des ailes courtes et sont symétriques, y compris pour une séparation aussi faible que  $0''.5$ .
- Les régimes spatiaux jouent un rôle significatif. *Dans la zone corrigée* par AO, pour un *seeing* faible quelques *speckles* quasi-statiques de forte intensité forment une aile lourde à droite. En présence d'un *seeing* plus important ( $k = 50$  ou  $k = 144$ ), du flux apparaît de façon plutôt uniforme dans l'anneau, diminuant la contribution relative de ces *speckles* dans la distribution empirique d'intensité par anneau. Le vent quant à lui ( $k = 100$ ) apporte davantage de flux dans la direction dans laquelle il souffle. Dans tous les cas, les histogrammes ont tous une queue à droite plutôt lourde. Si les données présentes dans un anneau étaient considérées comme identiquement distribuées, les histogrammes ne pourraient nullement être représentés par des distributions Poissonniennes (voir les graphes de la seconde ligne des figures de l'annexe : la variance est de l'ordre de 100 fois égale à la moyenne). Comme on l'a dit, la moyenne peut cependant varier dans l'espace. Mais cette variation n'est pas suffisante pour expliquer les histogrammes à partir d'une distribution de Poisson de moyenne variable. Il est donc préférable de modéliser marginalement les données par une distribution gaussienne plutôt que par une distribution de Poisson car sa variance est non contrainte. Dans la mesure du possible l'idéal est certainement de modéliser les données par un modèle physique adapté, tel que le modèle *speckle* décrit dans la prochaine partie.
- *Dans la zone non corrigée* par AO, une queue à droite peut légèrement persister lorsque des *speckles* se distinguent effectivement sur l'image. Mais les histogrammes sont globalement relativement bien représentés par des distributions symétriques. La distribution de Poisson est toujours largement rejetée et une distribution gaussienne est ainsi préférable.

## Conclusion quant aux modèles statistiques étudiés dans cette thèse

### • Conclusion de l'analyse statistique

Une description détaillée du jeu de données simulé par le *package* SPHERE du *problem-solving environment* CAOS décrit dans le chapitre précédent a en premier lieu confirmé l'existence de deux grands *régimes spatiaux* : la zone corrigée par l'AO (zone de séparation inférieure à  $0''.83$ ) et la zone non-corrigée par l'AO. Les données mesurées dans la zone corrigée par l'AO ont un comportement variable à la fois en fonction de la séparation à l'étoile mais aussi en fonction de la position orthoradiale à cause d'effets

dont la symétrie n'est pas cylindrique tels que l'effet induit par le vent<sup>3</sup>. La zone non-correctée par l'AO a quant à elle un comportement à symétrie cylindrique bien connu.

Par ailleurs, comme nous l'avons discuté en détail il est naturel et assez bien justifié statistiquement de chercher à diviser le jeu de données global en des ensembles d'images  $\{\mathbf{x}(k), k \in \mathcal{K}_i\}$ , les images étant *i.i.d.* temporellement au sein d'un tel ensemble. Dans les parties suivantes, nous opterons pour un modèle *i.i.d.* temporellement, de moyenne variable spatialement et corrélé spatialement, en laissant la possibilité de classer les données en des ensembles  $\{\mathbf{x}(k), k \in \mathcal{K}_i\}$  afin de mieux vérifier l'hypothèse de distribution identique entre les poses. Vis-à-vis du modèle réciproque inversant les rôles du temps et de l'espace, ce modèle ne se justifie pas de façon indiscutable à partir des éléments statistiques discutés ici mais n'est du moins pas aberrant. Deux arguments transverses rendent ce modèle plus intéressant que sa réciproque : la très nette corrélation entre des moments des distribution empiriques et les valeurs de *seeing* et vitesse de vent (par ailleurs éventuellement mesurables pendant l'observation) et le rôle asymétrique joué par le temps et l'espace dans le processus d'acquisition des données, qui invite à considérer  $k$  comme la variable selon laquelle supposer le modèle *i.i.d.* en classant éventuellement les données pour bien vérifier cette condition.

En particulier, il a été observé qu'il existe en général une relation forte entre l'évolution temporelle de la moyenne spatiale (et en général aussi de la variation spatiale) des données et l'évolution temporelle de paramètres physiques (*seeing* et/ou vitesse du vent ici). Plus précisément, pour ce jeu de données à 0'2 la moyenne spatiale est très corrélée à la vitesse du vent (avec une petite contribution du *seeing* comme on peut le voir à la fin du temps d'observation) et à mesure que l'on s'écarte de l'axe optique la contribution relative du *seeing* augmente et celle du vent diminue. Finalement, après la limite de correction de l'AO, les statistiques spatiales sont presque exclusivement indiquée par le *seeing*. Cette relation forte entre la variation temporelle des caractéristiques des données et la variation temporelle de paramètres physiques éventuellement mesurables par ailleurs permet de pouvoir supposer qu'il serait éventuellement possible de classer les images  $\mathbf{x}(k)$  en images *i.i.d.* à l'aide de mesures physiques indépendantes. Le problème de division des données en ensembles *i.i.d.* n'est pas davantage investigué dans cette thèse : nous nous focalisons plutôt dans cette thèse sur l'étude d'un tel ensemble d'images *i.i.d.*

### • Choix des hypothèses statistiques utilisées dans cette thèse

Pour conclure cette analyse statistique qualitative et introduire le reste de cette thèse, le premier modèle que nous étudierons consistera à supposer les données *i.i.d.* temporellement, indépendantes spatialement et de moyenne variable spatialement. Ce modèle relativement simple, qui sera étudié dans le reste de cette partie, est nécessaire notamment pour construire une méthode de traitement applicable à un jeu de données tel que celui présenté ici et pour l'ensemble des positions du champ d'intérêt. Ce modèle permet de plus de construire une statistique linéaire des données pour chaque position du champ qui indique une estimation de l'intensité de la planète pour chaque position (initiale) que celle-ci pourrait occuper. La linéarité de cette statistique est nécessaire pour utiliser une modélisation physique plus fine des données et calculer une probabilité de fausse alarme fiable d'un test de détection (lorsque ce test consiste à seuiller cette statistique, c'est-à-dire à décider qu'une planète est présente si son intensité estimée est supérieure à un certain seuil à déterminer pour un certain choix de probabilité de fausse alarme).

Après avoir extrait un ensemble de positions pour lesquelles on suspecte la présence d'une planète il est possible d'appliquer une méthode de traitement plus évoluée et donc construite à partir d'une modélisation plus fine des données. Grâce à l'hypothèse d'images *i.i.d.* temporellement, une extension du modèle simple qui sera développé dans un premier temps consistera à prendre en compte les corrélations spatiales car celles-ci semblent fondamentales notamment parce que le signal a la même signature que les *speckles*. Ainsi, dans les parties suivantes nous étudierons les modèles *i.i.d.* temporellement et corrélé spa-

<sup>3</sup>Voir la seconde sous-section de l'annexe C.1 pour observer l'effet du vent sur une image.

tialement suivants : un modèle dit *modèle speckle* construit en amont de cette thèse par une généralisation multivariée du modèle de *speckles* en présence d'un fond d'amplitude complexe non nul, et un modèle Bayésien hiérarchique permettant d'éviter l'estimation d'une matrice de covariance.

## Chapitre 3

# Construction et analyse d'inférences à partir d'un modèle gaussien

### 3.1 Modèle gaussien des données

#### Description qualitative d'une image

Dans l'introduction nous avons décrit brièvement le trajet suivi par la lumière depuis l'étoile jusqu'à la caméra CCD de l'instrument SPHERE, modélisation physique reprise en grand détail par le *package* de simulation SPHERE. Synthétisons maintenant l'information de manière à mettre en évidence les hypothèses à utiliser pour construire le modèle statistique des *images mesurées* par la CCD située sur le plan focal de l'instrument SPHERE. Il s'agit ici de chercher les éléments discriminant la planète de l'étoile et du ciel. Nous omettons ici tout effet de *ghost* et réflexion interne de la lumière provenant par exemple d'autres étoiles. Nous supposons de plus que l'étoile est parfaitement alignée avec l'axe optique. Cela signifie en pratique qu'un prétraitement est supposé avoir réaligné les images, mais en laissant malheureusement éventuellement des variations d'intensité de l'étoile d'une pose à l'autre dues à la sensibilité du coronographe aux erreurs de tilt.

Schématiquement, la contribution de l'étoile sur une image peut être décomposée en une zone interne où l'Optique Adaptative (AO) corrige partiellement les effets de la turbulence atmosphérique qui est ainsi dominée par le profil résiduel coronographique, et une zone externe dominée par les turbulences atmosphériques. La zone corrigée est une zone circulaire dont le rayon angulaire sur l'image est donné par

$$r_{\text{correc}} = \frac{1}{2} \frac{\lambda}{d_{\text{act}}} \quad \text{avec} \quad d_{\text{act}} = \frac{D}{N_{\text{act}}}$$

où  $\lambda$  est la longueur d'onde d'acquisition,  $d_{\text{act}}$  la distance inter-actuateurs telle que vue par le miroir limitant (le miroir primaire) et  $N_{\text{act}}$  est le nombre d'actuateurs le long du diamètre du miroir déformable. Pour le VLT,  $D = 8m$  et avec l'AO extrême prévue pour SPHERE  $N_{\text{act}} = 40$ . Ainsi, en bande H2 où  $1.57\mu m \leq \lambda \leq 1.64\mu m$  la limite de correction de l'AO se situe à  $0''.81 \leq r_{\text{correc}} \leq 0''.85$ . Cette limitation est bien visible sur la figure 3.1 ; les pixels (225, 150) et (225, 300) font par exemple partie de cette limite circulaire. Ce profil serait à peu près à symétrie cylindrique en dehors d'effets asymétriques. Cependant, des éléments de petite, moyenne et grande échelles spatiales brisent la symétrie de ce profil d'intensité lumineuse : le vent, l'araignée due aux attaches du miroir secondaire, les aberrations optiques de tous ordres, *etc.*

En particulier, les *speckles* (aberrations de petite échelle appelées *tavelures* en français) compliquent considérablement notre problème car ils ont le même profil que le signal d'intérêt. Les *speckles* sont des taches de diffraction dont le profil est proche du profil d'Airy et dont la taille caractéristique est  $\lambda/D$ . On appellera par la suite ce profil plus simplement la PSF (pour *Point-Spread Function*) ou réponse

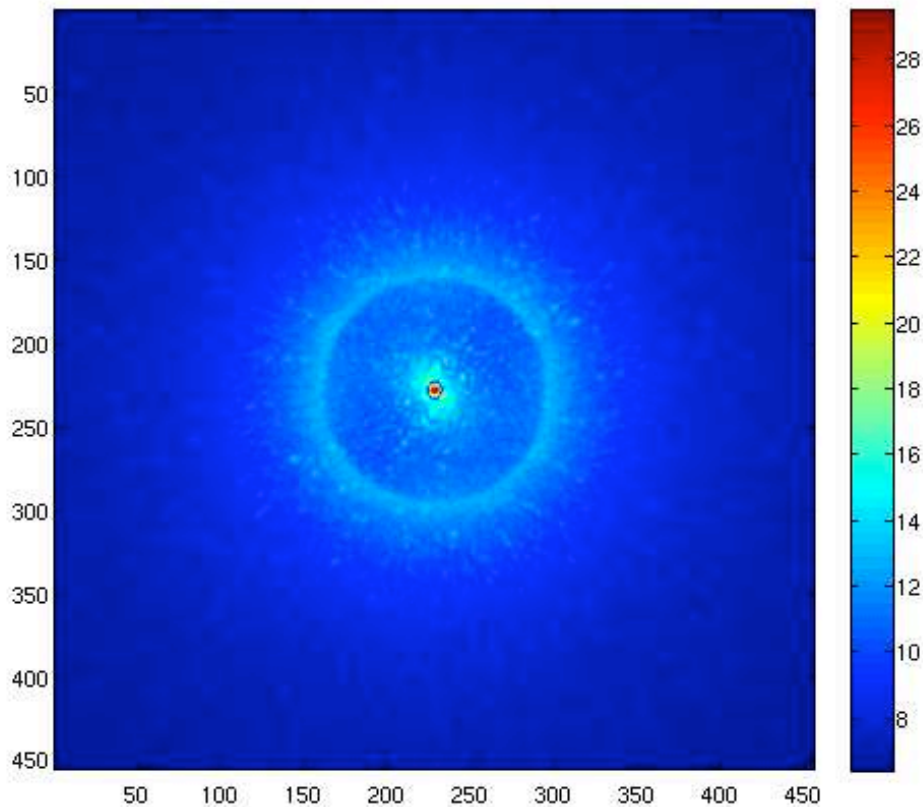


FIG. 3.1: Image obtenue à partir d'une simulation *end-to-end* du futur instrument VLT-SPHERE et représentée en échelle logarithmique (intensité de chaque pixel mise à la puissance 0.2) afin de faire apparaître les données de faible intensité. L'étoile est une source ponctuelle géométriquement située sur l'axe optique mais c'est elle qui illumine tout de même l'ensemble de la CCD. En effet, des effets de diffusion et diffraction de l'atmosphère et des optiques "renvoient de la lumière" autour de la position géométrique de l'objet. Cette lumière n'est pas diffractée "très loin" ni de façon "très puissante" mais la focale du télescope est grande de sorte que la résolution spatiale est grande (le télescope *zoom* beaucoup : haute résolution angulaire), et comme nous représentons le logarithme de l'image on observe ici des intensités à peu près invisibles à l'oeil (imagerie haute dynamique). La zone intérieure de relativement basse intensité correspond à la zone de correction d'AO. Cette intensité dite basse est pourtant déjà beaucoup plus élevée que les planètes que nous y cherchons et ne voyons par conséquent pas ici.

impulsionnelle du système<sup>1</sup>. Les *speckles* proviennent ainsi d'interférences constructives de la lumière émanant de l'étoile et dûes aux turbulences atmosphériques ou à des aberrations instrumentales. En terme de fluctuations aléatoires d'intensité lumineuse, les *speckles* ont un comportement bien connu (voir Goodman [2007]) qui sera l'objet du modèle statistique de la partie II. La planète quant à elle est une source ponctuelle d'intensité négligeable par rapport à celle de l'étoile, légèrement décalée spatialement de celle-ci, et dont la réponse instrumentale est bien approximée par une simple tache de diffraction et a donc le même profil qu'un *speckle*, appelé PSF.

En conclusion, il est important d'une part d'éviter toute hypothèse trop forte de symétrie cylindrique et plus généralement toute hypothèse sur la forme de la contribution de l'étoile, et d'autre part de trouver des critères déterministes de discrimination entre *speckle* et planète au niveau temporel et/ou spectral car ceux-ci ont le même profil sur une pose donnée. Comme nous le verrons, l'utilisation de modèles

<sup>1</sup>On ne peut pas parler à strictement parler d'une PSF avec une AO et un coronographe car leur réponse n'est pas identique à une translation près lorsque l'on déplace la source, mais en pratique la réponse est bien approximativement *shift-invariant* si la source est située un peu plus loin que l'axe optique.

statistiques fins basés sur les modèles d'imagerie *speckle* et de bruit Poissonien serait trop lourde pour traiter intégralement les gros jeux de données prévus pour l'instrument SPHERE.

### Discrimination entre signal (planète) et bruit (étoile)

Voici finalement des caractéristiques et hypothèses centrales que nous utilisons dans les modèles qui suivent et qui nous permettront de discriminer en partie le signal (planète) du bruit (étoile). Ces hypothèses et caractéristiques sont décrites sur une image donnée, sur deux images simultanées acquises dans deux bandes spectrales proches (H<sub>2</sub>-H<sub>3</sub> typiquement) et pour une séquence d'images successives.

- Sur une *seule image* : le signal a la même signature que les *speckles* (provenant des aberrations atmosphériques et instrumentales), *ie* un profil de diffraction proche d'un profil d'Airy.
- Sur deux *images simultanées*, acquises grâce au dispositif d'imagerie à *deux bandes spectrales* (DBI pour *Dual-Band Imaging*) :
  - L'intensité du signal est différente dans les deux filtres si la planète a une bande d'absorption dans l'un des deux filtres, tandis que le bruit est presque identique si les deux bandes sont suffisamment proches spectralement et qu'elles correspondent au continuum de l'étoile. Ceci constitue la principale motivation pour la création de ce dispositif instrumental et pour le choix des filtres.
  - Tous les profils de diffraction ont une taille qui augmente linéairement avec  $\lambda$  : la réponse dite *en axe* (centrée autour de l'axe optique), les aberrations *speckle* et le profil de la planète.
- Séquence d'*images successives dans le temps* pour un filtre donné (poses d'environ 10 s séparées de 0.001 s et éventuellement intégrées en des longues poses de 100 s) :
  - Pendant que le télescope suit l'étoile, sur le plan focal la planète tourne circulairement autour du centre de l'image. Ce phénomène dit de rotation de champ se produit car le VLT a une monture alt-azimutale (Avila and Wirenstrand [2005]).
  - Nous supposons que les images ne sont pas corrélées d'une pose à l'autre car les temps caractéristiques de corrélation haute fréquence de l'atmosphère et de l'optique adaptative sont de l'ordre de la milliseconde à la centaine de millisecondes, alors qu'une pose dure quelques secondes minimum.
  - Nous supposons que les conditions atmosphériques ne varient pas trop dans le temps et que les aberrations optiques restent fixes aussi. Afin de satisfaire ces hypothèses il est possible en pratique de définir plusieurs jeux de données correspondant à plusieurs parties de nuit correspondant à des données assez stationnaires.

Ces descriptions qualitatives déterministes doivent ensuite être incorporées dans un modèle statistique décrivant la nature aléatoire du signal mesuré.

### Modèle statistique pour une bande spectrale

Nous nous plaçons dans un premier temps dans le cas mono-canal. L'utilisation simultanée d'images dans deux canaux spectraux permise par le mode DBI de l'instrument IRDIS de SPHERE sera alors relativement immédiate et effectuée dans la section 3.5.

Afin de pouvoir construire des inférences applicables aux gros jeux de données prévus pour SPHERE, il est nécessaire de recourir à une phase d'initialisation simple de détection-estimation. Celle-ci est construite à partir d'un modèle statistique basique car la seule simulation de la rotation de la planète requiert un important temps de calcul pour tester l'ensemble des positions initiales.

On suppose qu'un jeu de données est constitué de  $K$  images successives statistiquement indépendantes. Chaque image acquise est représentée par la concaténation de ses colonnes en un vecteur  $\mathbf{x}(k)$  de taille  $L$ , où  $L$  est le nombre de pixels de la caméra (voir le tableau de notations). On suppose que l'image est alors la superposition de la contribution de l'étoile, de celle de la planète et d'un bruit gaussien non corrélé spatialement.

$$\mathbf{x}(k) = \mathbf{d} + \alpha \mathbf{p}(k, r) + \boldsymbol{\epsilon}(k), \quad k = 1, \dots, K \quad (3.1)$$



où :

- $k$  indique le numéro de la pose.
- $\mathbf{d}$  est un vecteur de taille  $L \times 1$  qui représente la contribution de l'étoile (halo atmosphérique, réponse du coronographe, aberrations optiques statiques et *speckles* statiques en particulier). Cette contribution est supposée déterministe, stationnaire dans le temps, non uniforme spatialement (et sans contrainte particulière) et inconnue.
- $\alpha$  est un scalaire représentant l'intensité déterministe inconnue de la potentielle planète.
- $\mathbf{p}(k, r)$  est la réponse déterministe de la planète (source ponctuelle) dans la pose  $k$ . Il s'agit à nouveau d'un vecteur de taille  $L \times 1$ , mais cette fois dont le seul paramètre inconnu est la position de la planète sur la première pose, position notée  $r$ . A cause de l'effet de rotation de champ et de la relative nécessité d'étalonnage instrumental de la PSF, le profil  $\mathbf{p}(k, r)$  n'a pas de forme analytique simple. Il ne peut pas non plus être obtenu simplement à partir d'un profil  $\mathbf{p}(k, r')$  pour une position initiale  $r'$  ne correspondant pas à la même séparation angulaire au centre de l'image que la position initiale  $r$ . Le terme  $\mathbf{p}(k, r)$  doit donc être en pratique en général simulé pour chaque position initiale  $r$  testée.
- $\boldsymbol{\epsilon}(k)$  est un vecteur de bruit aléatoire gaussien supposé indépendant dans le temps et dans l'espace et de variance  $\sigma^2$  uniforme dans le champ considéré :  $\boldsymbol{\epsilon}(k) \sim \mathcal{N}_L(\mathbf{0}, \sigma^2 I_L)$ , où  $\sigma^2$  est inconnu.

Le modèle (3.1) s'exprime de façon équivalente sous la forme de la fonction de densité de probabilité (PDF) des données  $\mathbf{x} = \{\mathbf{x}(1), \dots, \mathbf{x}(K)\}$  étant donnés les paramètres inconnus  $\alpha, r, \mathbf{d}, \sigma^2$  :

$$p(\mathbf{x}(1), \dots, \mathbf{x}(K) | \alpha, r, \mathbf{d}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{KL}{2}}} \exp\left(-\sum_{k=1}^K \frac{\|\mathbf{x}(k) - \mathbf{d} - \alpha\mathbf{p}(k, r)\|^2}{2\sigma^2}\right) \quad (3.2)$$

Ce modèle est pertinent dans les cas suivants :

- $H_0$  : hypothèse nulle, il n'y a pas de planète ( $\alpha = 0$ )
- $H_1$  : hypothèse alternative, il y a une planète ( $\alpha > 0$ )
- $H_1$  *bis* : il y a en fait plusieurs planètes mais les aires couvertes par ces planètes ne se superposent pas<sup>2</sup>. Le modèle reste alors le même : si une planète a été détectée par la méthode que nous allons présenter, il suffit alors par exemple d'appliquer un algorithme du type CLEAN en supprimant la contribution de la première planète détectée pour tenter d'en détecter une seconde en appliquant à nouveau la méthode de détection.

Pour traiter le cas où il y a deux planètes dont les aires couvertes se superposent, on peut étendre le modèle précédent en ajoutant le terme  $\alpha_2\mathbf{p}(k, r_2)$  à  $\mathbf{d} + \alpha\mathbf{p}(k, r)$  pour représenter la contribution de la seconde planète.

Ici nous traitons donc simplement le test d'hypothèses  $H_0$  vs  $H_1$ , où  $\alpha$  et  $r$  sont inconnus et recherchés.  $\mathbf{d}$  et  $\sigma^2$  sont deux autres paramètres inconnus mais leur valeur n'est pas d'un intérêt direct. On les appelle paramètres de nuisance comme vu dans l'introduction, section 1.2.

### 3.2 Estimation d'intensité et de position de la planète

Maximiser la vraisemblance (3.2) par rapport aux paramètres inconnus ( $\alpha, r, \mathbf{d}, \sigma^2$ ) est équivalent à minimiser l'opposé de la log-vraisemblance

$$\mathcal{L}_0(\alpha, r, \mathbf{d}, \sigma^2) = -\ln(p(\mathbf{x}(1), \dots, \mathbf{x}(K) | \alpha, r, \mathbf{d}, \sigma^2)) = \sum_{k=1}^K \frac{\|\mathbf{x}(k) - \mathbf{d} - \alpha\mathbf{p}(k, r)\|^2}{2\sigma^2} + \frac{KL}{2} \ln(2\pi\sigma^2) \quad (3.3)$$

<sup>2</sup>On définit simplement "l'aire couverte par une planète" comme le champ de la caméra qui a été vu au moins pour quelques poses par la planète. A strictement parler si la PSF a un support infini deux sources se superposeront toujours quelles que soient leurs positions initiales mais on considère que la PSF se limite par exemple à son premier anneau.

En fixant  $\sigma^2$ , nous voyons que la valeur de  $(\alpha, r, \mathbf{d})$  pour laquelle  $\mathcal{L}_0$  est minimum ne dépend pas de  $\sigma^2$ . Minimiser  $\mathcal{L}_0$  est donc équivalent à minimiser

$$\mathcal{L}(\alpha, r, \mathbf{d}) = \sum_{k=1}^K \|\mathbf{x}(k) - \mathbf{d} - \alpha \mathbf{p}(k, r)\|^2 \quad (3.4)$$

### Estimation d'intensité $\hat{\alpha}(r)$ (et estimation *conjointe* du fond)

Notons que pour une position initiale  $r$  donnée le modèle est linéaire par rapport au vecteur colonne  $\boldsymbol{\theta}$  des paramètres  $\mathbf{d}$  et  $\alpha$  :

$$\mathbf{x} = A(r)\boldsymbol{\theta} + \boldsymbol{\varepsilon} \quad (3.5)$$

où

- $\boldsymbol{\theta}$  est le vecteur de taille  $(L+1) \times 1$  des paramètres  $\boldsymbol{\theta} = (\mathbf{d}^t, \alpha)^t$ ,
- $\mathbf{x}$  représente la concaténation de taille  $KL \times 1$  des vecteurs  $\{\mathbf{x}(k), k = 1, \dots, K\}$ ,
- $\boldsymbol{\varepsilon}$  représente la concaténation de taille  $KL \times 1$  des vecteurs  $\{\boldsymbol{\varepsilon}(k), k = 1, \dots, K\}$ , de telle sorte que le bruit aléatoire est totalement spécifié par  $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_{KL})$ ,
- $A(r)$  est la concaténation de taille  $KL \times (L+1)$  des matrices  $A(k, r)$  définies par  $A(k, r) = (I_L, \mathbf{p}(k, r))$ .

Pour un  $r$  donné la fonction quadratique

$$\mathcal{L}(\boldsymbol{\theta}, r) = \|\mathbf{x} - A(r)\boldsymbol{\theta}\|^2 \quad (3.6)$$

a un minimum et son gradient y est nul. Le gradient de  $\mathcal{L}(\boldsymbol{\theta}, r)$  par rapport à  $\boldsymbol{\theta}$  est donné par :

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta}, r)}{\partial \boldsymbol{\theta}} = \frac{\partial (\mathbf{x} - A(r)\boldsymbol{\theta})^t (\mathbf{x} - A(r)\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -2A(r)^t (\mathbf{x} - A(r)\boldsymbol{\theta})$$

L'équation  $\frac{\partial \mathcal{L}(\boldsymbol{\theta}, r)}{\partial \boldsymbol{\theta}} = \mathbf{0}$  se réexprime donc  $A(r)^t \mathbf{x} = A(r)^t A(r)\boldsymbol{\theta}$  et sa solution donne finalement le résultat bien connu :

$$\hat{\boldsymbol{\theta}}(r) = (A(r)^t A(r))^{-1} A(r)^t \mathbf{x} \quad (3.7)$$

En utilisant le fait que  $A(r)$  est la concaténation verticale des  $A(k, r) = (I_L, \mathbf{p}(k, r))$  on a :

$$(A(r)^t A(r))^{-1} = \left( \sum_{k=1}^K A^t(k, r) A(k, r) \right)^{-1} = \frac{1}{K} \begin{pmatrix} I_L & \bar{\mathbf{p}}(r) \\ \bar{\mathbf{p}}(r)^t & \widetilde{\|\mathbf{p}\|^2}(r) \end{pmatrix}^{-1}$$

avec la définition des valeurs moyennes

$$\bar{\mathbf{p}}(r) = \frac{1}{K} \sum_{k=1}^K \mathbf{p}(k, r) \quad (3.8)$$

$$\widetilde{\|\mathbf{p}\|^2}(r) = \frac{1}{K} \sum_{k=1}^K \|\mathbf{p}(k, r)\|^2 \quad (3.9)$$

Or d'après le lemme d'inversion des matrices partitionnées (Golub and Loan [1989])

$$\begin{aligned} \text{si } A &= \begin{pmatrix} A_{(11)} & A_{(12)} \\ A_{(21)} & A_{(22)} \end{pmatrix} \\ \text{alors } A^{-1} &= \begin{pmatrix} (A_{(11)} - A_{(12)} A_{(22)}^{-1} A_{(21)})^{-1} & -(A_{(11)} - A_{(12)} A_{(22)}^{-1} A_{(21)})^{-1} A_{(12)} A_{(22)}^{-1} \\ -(A_{(22)} - A_{(21)} A_{(11)}^{-1} A_{(12)})^{-1} A_{(21)} A_{(11)}^{-1} & (A_{(22)} - A_{(21)} A_{(11)}^{-1} A_{(12)})^{-1} \end{pmatrix} \end{aligned}$$

Donc ici

$$(A(r)^t A(r))^{-1} = \frac{1}{K(\|\widetilde{\mathbf{p}}\|^2(r) - \|\bar{\mathbf{p}}(r)\|^2)} \begin{pmatrix} (\|\widetilde{\mathbf{p}}\|^2(r) - \|\bar{\mathbf{p}}(r)\|^2)I_L + \bar{\mathbf{p}}(r)\bar{\mathbf{p}}(r)^t & -\bar{\mathbf{p}}(r) \\ -\bar{\mathbf{p}}(r)^t & 1 \end{pmatrix} \quad (3.10)$$

où  $\bar{\mathbf{p}}(r)$  et  $\|\widetilde{\mathbf{p}}\|^2(r)$  sont définis par l'Eq. (3.8, 3.9). Notons que le dénominateur, que l'on va noter  $Kc_K(r)$  de sorte que  $c_K(r)$  soit un terme moyen, se réexprime à partir de la moyenne des normes au carré des résidus des profils  $\mathbf{p}(k, r)$  "nettoyés" de la moyenne  $\bar{\mathbf{p}}(r)$  :

$$c_K(r) = \|\widetilde{\mathbf{p}}\|^2(r) - \|\bar{\mathbf{p}}(r)\|^2 = \frac{1}{K} \sum_{k=1}^K \|\mathbf{p}(k, r) - \bar{\mathbf{p}}(r)\|^2 \geq 0 \quad (3.11)$$

car

$$\begin{aligned} \sum_{k=1}^K \|\mathbf{p}(k, r) - \bar{\mathbf{p}}(r)\|^2 &= \sum_k \|\mathbf{p}(k, r)\|^2 - 2 \sum_k \mathbf{p}(k, r)^t \bar{\mathbf{p}}(r) + \sum_k \|\bar{\mathbf{p}}(r)\|^2 \\ &= \sum_k \|\mathbf{p}(k, r)\|^2 - 2K\|\bar{\mathbf{p}}(r)\|^2 + K\|\bar{\mathbf{p}}(r)\|^2 \end{aligned}$$

L'expression de  $\hat{\boldsymbol{\theta}}(r)$  donnée par l'Eq. (3.7) est maintenant simplifiée. A  $r$  supposé connu on obtient l'estimateur de  $\alpha$  par la  $L + 1$  ème composante de  $\hat{\boldsymbol{\theta}}(r)$ . Quoique  $\hat{\alpha}(r)$  soit bien une fonction linéaire de  $\mathbf{x}$ , l'ensemble des  $\hat{\alpha}(r)$  (pour toutes les positions initiales  $r$  testées) n'est pas obtenue en appliquant simplement un filtre à  $\mathbf{x}$  car ce "filtre" dépend de  $r$  :

$$\hat{\alpha}(r) = \frac{1}{K} \sum_{k=1}^K \mathbf{w}(k, r)^t \mathbf{x}(k) \quad (3.12)$$

$$\text{avec } \mathbf{w}(k, r) = \frac{\mathbf{p}(k, r) - \bar{\mathbf{p}}(r)}{c_K(r)} \quad (3.13)$$

où  $\bar{\mathbf{p}}(r)$  et  $c_K(r)$  sont définis par les Eq. (3.8, 3.11). On peut aussi réexprimer  $\hat{\alpha}(r)$  par :

$$\hat{\alpha}(r) = \frac{\sum_k \mathbf{p}^t(k, r) \mathbf{x}(k) - K \bar{\mathbf{p}}^t(r) \bar{\mathbf{x}}}{\sum_k \mathbf{p}^t(k, r) \mathbf{p}(k, r) - K \bar{\mathbf{p}}^t(r) \bar{\mathbf{p}}(r)} \quad (3.14)$$

$$\text{avec } \bar{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}(k) \quad (3.15)$$

L'expression de l'estimation du fond  $\mathbf{d}$  dû exclusivement à l'étoile est donnée par les  $L$  premières composantes de  $\hat{\boldsymbol{\theta}}(r)$  mais s'exprime autrement très naturellement en fonction de  $\hat{\alpha}(r)$ .  $\hat{\mathbf{d}}(r)$  est égal à la moyenne de toutes les images dont on a retiré chaque fois la contribution de la planète :

$$\hat{\mathbf{d}}(r) = \bar{\mathbf{x}} - \hat{\alpha}(r) \bar{\mathbf{p}}(r) = \frac{1}{K} \left( \sum_k (\mathbf{x}(k) - \hat{\alpha}(r) \mathbf{p}(k, r)) \right) \quad (3.16)$$

La condition d'identifiabilité des composantes du paramètre  $\boldsymbol{\theta}$  est que le dénominateur  $Kc_K(r)$  de  $\hat{\boldsymbol{\theta}}(r)$  soit non nul. Réexprimons  $c_K(r)$  comme la moyenne de la moyenne des différences de toutes les paires de profils :

$$c_K(r) = K^{-2} \sum_k \sum_{k' < k} \|\mathbf{p}(k, r) - \mathbf{p}(k', r)\|^2 \quad (3.17)$$

En effet, d'après l'Eq. (3.11)

$$\begin{aligned}
c_K(r) &= K^{-1} \sum_k \|\mathbf{p}(k, r)\|^2 - K^{-2} \left\| \sum_k \mathbf{p}(k, r) \right\|^2 \\
&= K^{-2} \left( K \sum_k \|\mathbf{p}(k, r)\|^2 - \sum_k \mathbf{p}(k, r)^t \sum_{k'} \mathbf{p}(k', r) \right) = K^{-2} \left( \sum_{k, k'} (\|\mathbf{p}(k, r)\|^2 - \mathbf{p}(k, r)^t \mathbf{p}(k', r)) \right) \\
&= \frac{1}{2K^2} \left( \sum_{k, k'} (\|\mathbf{p}(k, r)\|^2 - 2\mathbf{p}(k, r)^t \mathbf{p}(k', r) + \|\mathbf{p}(k', r)\|^2) \right) = \frac{1}{2K^2} \sum_{k, k'} \|\mathbf{p}(k, r) - \mathbf{p}(k', r)\|^2
\end{aligned} \tag{3.18}$$

La planète est donc identifiable du fond dès que  $c_K(r) \neq 0$ , *i.e.* dès qu'il existe au moins deux poses dans lesquelles les profils de la planète ne sont pas confondus *i.e.* dès que

$$\exists (k, k') \in \llbracket 1, K \rrbracket^2 \quad / \quad k' \neq k \quad \text{et} \quad \mathbf{p}(k, r) \neq \mathbf{p}(k', r) \tag{3.19}$$

Ceci correspond en effet à notre modélisation : sans mouvement, la planète ne peut pas être distinguée du fond.

Notons au passage que si on procède sur le numérateur de l'estimateur  $\hat{\alpha}(r)$  de la même façon que pour obtenir l'Eq. (3.17) on obtient :

$$\hat{\alpha}(r) = \frac{\sum_k \sum_{k' < k} (\mathbf{p}(k, r) - \mathbf{p}(k', r))^t (\mathbf{x}(k) - \mathbf{x}(k'))}{K^2 c_K(r)} \tag{3.20}$$

L'estimation  $\hat{\alpha}$  est donc obtenue de façon équivalente par filtrage des différences de **tous** les couples d'images  $\{k, k' < k\}$ . La présente méthode est donc proche des méthodes différentielles angulaires classiques (Marois et al. [2006]), avec la contrainte ici que toutes les combinaisons doivent être utilisées. Sous notre hypothèse de départ que tous les  $\mathbf{x}(k)$  ont le même fond (Eq. 3.1), nous constaterons que cette contrainte est nécessaire pour obtenir l'estimateur d'intensité ( $\hat{\alpha}(r)$ ) de variance minimale.

Notons par le même genre de calcul que pour une position initiale  $r$  la zone des  $\mathbf{x}(k)$  qui intervient dans le calcul de l'estimation de  $\boldsymbol{\theta}$  est mathématiquement limitée aux pixels  $\ell$  qui voient au moins une fois la planète :

$$\{\ell \in \llbracket 1, L \rrbracket : \exists k \in \llbracket 1, K \rrbracket : p_\ell(k, r) \neq 0\} \tag{3.21}$$

On peut donc limiter le vecteur de données à cette zone spatiale pour l'estimation de  $\alpha$  et de  $\mathbf{d}$ .

### Estimation de la position initiale de la planète $\hat{r}_{\text{ML}}$

L'estimateur  $\hat{\boldsymbol{\theta}}(r)$  donnait jusqu'ici le minimum de  $\mathcal{L}(\boldsymbol{\theta}, r)$  (Eq. 3.6) pour chacune des valeurs possibles de  $r$ . L'estimateur du maximum de vraisemblance de  $r$  est alors donné par la valeur de  $r$  pour laquelle le minimum de  $\mathcal{L}(\hat{\boldsymbol{\theta}}(r), r)$  est atteint. Le minimum global de  $\mathcal{L}(\boldsymbol{\theta}, r)$  est donné par :

$$\mathcal{L}(\hat{\boldsymbol{\theta}}(r), r) = \|\mathbf{x} - A(r)\hat{\boldsymbol{\theta}}(r)\|^2 = \mathbf{x}^t \mathbf{x} - \mathbf{x}^t A(r) (A^t(r)A(r))^{-1} A^t(r) \mathbf{x}$$

Ce critère peut en fait se réexprimer très simplement en fonction de la statistique  $\hat{\alpha}(r)$  qui nous intéresse particulièrement. Repartant de l'expression (3.4) de  $\mathcal{L}(\boldsymbol{\theta}, r)$  et injectant l'expression (3.16) de  $\hat{\mathbf{d}}(r)$  qui faisait intervenir  $\hat{\alpha}(r)$  on obtient

$$\mathcal{L}(\hat{\alpha}(r), \hat{\mathbf{d}}(r), r) = \sum_{k=1}^K \|\mathbf{x}(k) - \hat{\mathbf{d}}(r) - \hat{\alpha}(r)\mathbf{p}(k, r)\|^2 = \sum_k \|\mathbf{a}(k) - \bar{\mathbf{a}}\|^2$$

avec  $\mathbf{a}(k) = \mathbf{x}(k) - \hat{\alpha}(r)\mathbf{p}(k, r)$  et  $\bar{\mathbf{a}}$  sa moyenne. De même que  $c_K(r)$  est donné à la fois par l'Eq.(3.17) et (3.18) on déduit

$$\mathcal{L}(\hat{\alpha}(r), \hat{\mathbf{d}}(r), r) = \frac{1}{K} \sum_k \|\mathbf{x}(k) - \bar{\mathbf{x}}\|^2 - \hat{\alpha}(r)^2 c_K(r) \quad (3.22)$$

où  $\bar{\mathbf{x}}$  et  $c_K(r)$  sont définis par les Eq. (3.15) et (3.11). La minimisation de ce critère par rapport à  $r$  s'exprime<sup>3</sup> alors très simplement à partir de la fonction  $r \rightarrow \hat{\alpha}(r)$  :

$$\hat{r}_{\text{ML}} = \arg \min_{r \in \llbracket 1, L \rrbracket} \mathcal{L}(\hat{\alpha}(r), \hat{\mathbf{d}}(r), r) = \arg \min_r \{\hat{\alpha}(r)^2 c_K(r)\}$$

L'estimateur du maximum de vraisemblance  $\hat{\alpha}_{\text{ML}}$  de  $\alpha$  est alors obtenu en injectant  $\hat{r}_{\text{ML}}$  dans  $\hat{\alpha}(r)$  puisque  $\hat{\alpha}(r)$  donné par l'Eq. (3.12) maximise la vraisemblance à  $r$  fixé.

$$\hat{\alpha}_{\text{ML}} = \hat{\alpha}(\hat{r}_{\text{ML}}) \quad (3.23)$$

Si  $\hat{\alpha}_{\text{ML}} > 0$ , d'après l'appendice de Smith et al. [2009],  $\hat{r}_{\text{ML}}$  se réexprime plus simplement comme

$$\hat{r}_{\text{ML}} = \arg \max_r \left\{ \hat{\alpha}(r) \sqrt{c_K(r)} \right\} \quad (3.24)$$

Si l'on peut négliger (localement) la variation de  $c_K(r)$  par rapport à  $r$ , on peut approximer l'estimateur de position et donc l'estimateur d'intensité par

$$\hat{r}_{\text{ML}} \approx \arg \max_r \hat{\alpha}(r) \quad (3.25)$$

$$\hat{\alpha}_{\text{ML}} \approx \max_r \hat{\alpha}(r) \quad (3.26)$$

Si  $c_K(r)$  varie très peu, l'estimation globale d'intensité de la planète  $\hat{\alpha}_{\text{ML}}$  est approximée par le maximum de  $\hat{\alpha}(r)$  sur toutes les positions  $r$  testées, et la position estimée de la planète  $\hat{r}_{\text{ML}}$  par l'endroit où ce maximum se trouve. Ceci rend les expressions de  $\hat{\alpha}_{\text{ML}}$  et  $\hat{r}_{\text{ML}}$  plus intuitives mais ne rend pas les calculs mathématiques plus simples en pratique. Cette approximation est discutée dans la section 4.1 dans le cas de SPHERE et s'avère en général bonne localement, sauf trop près de l'axe optique.

Notons que comme les profils  $\mathbf{p}(k, r)$  ne peuvent être exprimées comme une fonction explicite de  $r$ , les estimateurs qui sont fonctions des  $\mathbf{p}(k, r)$  n'ont pas de forme explicite non plus. Ainsi, l'estimation de  $r$  et donc de  $\alpha$  et  $\mathbf{d}$  doit passer par le calcul de  $\hat{\alpha}(r)$  –et de  $c_K(r)$  s'il varie en fonction de  $r$ – pour toutes les valeurs de  $r$ .

## Estimation d'un nombre de photons en pratique

L'objet de cette section n'est pas d'inverser le processus de Poisson (*i.e.* de changer de modèle statistique) mais simplement de ramener l'estimation précédente  $\hat{\alpha}_{\text{ML}}$  à une estimation d'intensité qui ait un sens physique et qui soit un peu plus proche de l'intensité lumineuse effectivement émise par la planète.

D'abord, l'estimation de  $\alpha$  doit être corrigée. La PSF a en général un support infini de sorte qu'en pratique lorsque l'on programme l'estimateur précédent on construit  $\mathbf{p}_{\text{instant}}$  en appliquant un masque binaire de support restreint à la PSF de support infini. Plus généralement, on peut étudier l'effet d'une erreur d'étalonnage de la PSF en observant la différence d'estimation entre une estimation faite à partir d'une "vraie" PSF et une estimation faite à partir de la "fausse" PSF utilisée en pratique par commodité.

<sup>3</sup>Il convient en principe de spécifier le domaine d'appartenance de  $r$ , la position initiale *possible* de la planète. En principe, en éludant ici toute considération sur l'échantillonnage,  $r$  peut représenter une position subpixelique que l'on décrirait par exemple en coordonnées cartésiennes comme  $\mathbf{r} \in [0, \sqrt{L}]^2$ . Pour fixer les idées, nous dirons dans la suite que nous testons les pixels du champ. En réalité, ce choix n'intervient nul part et en pratique nous testerons aussi des positions subpixeliques.

Notons  $\mathbf{q}(k, r)$  le “vrai” profil de la planète sur la pose  $k$  et  $\mathbf{p}(k, r)$  le profil pris comme le profil de la planète en pratique. Un masquage binaire de la PSF vraie s'exprime par exemple sous la forme :

$$\mathbf{p}(k, r) = D(k, r)\mathbf{q}(k, r)$$

où  $D(k, r)$  est le masque, c'est-à-dire une matrice diagonale de taille  $L \times L$  dont l'élément diagonal  $\ell$  est 0 si la vraie PSF est cachée en le pixel  $\ell$  et 1 si elle ne l'est pas.  $\hat{\alpha}(r)$  est l'estimation obtenue en pratique, comme décrit précédemment, à partir du modèle utilisant  $\mathbf{p}(k, r)$  :

$$\hat{\alpha}(r) = \frac{K \sum_k \mathbf{p}(k, r)^t \mathbf{x}(k) - \sum_k \mathbf{p}(k, r)^t \sum_k \mathbf{x}(k)}{K \sum_k \|\mathbf{p}(k, r)\|^2 - \|\sum_k \mathbf{p}(k, r)\|^2}$$

Le vrai modèle sous-jacent est cependant

$$\mathbf{x}(k) = \mathbf{d} + \alpha \mathbf{q}(k, r) + \boldsymbol{\epsilon}(k), \quad k = 1, \dots, K \quad (3.27)$$

On peut calculer le biais de l'estimation de  $\hat{\alpha}(r)$  étant donné ce vrai modèle.  $\mathbf{E}[\mathbf{x}(k)] = \mathbf{d} + \alpha \mathbf{q}(k, r)$  donc par linéarité de l'estimateur  $\hat{\alpha}(r)$  par rapport à  $\mathbf{x}$ ,

$$\mathbf{E}[\hat{\alpha}(r)] = \frac{K \sum_k \mathbf{p}(k, r)^t \mathbf{E}[\mathbf{x}(k)] - \sum_k \mathbf{p}(k, r)^t \sum_k \mathbf{E}[\mathbf{x}(k)]}{K \sum_k \|\mathbf{p}(k, r)\|^2 - \|\sum_k \mathbf{p}(k, r)\|^2} = \alpha \times \frac{K \sum_k \mathbf{p}(k, r)^t \mathbf{q}(k, r) - \sum_k \mathbf{p}(k, r)^t \sum_k \mathbf{q}(k, r)}{K \sum_k \|\mathbf{p}(k, r)\|^2 - \|\sum_k \mathbf{p}(k, r)\|^2}$$

On peut donc définir l'estimateur  $\hat{\alpha}'(r)$  corrigé du biais, *i.e.* tel que  $\mathbf{E}[\hat{\alpha}'(r)] = \alpha$ , à partir de l'estimation initiale  $\hat{\alpha}(r)$  :

$$\hat{\alpha}'(r) = \hat{\alpha}(r) \times \frac{K \sum_k \|\mathbf{p}(k, r)\|^2 - \|\sum_k \mathbf{p}(k, r)\|^2}{K \sum_k \mathbf{p}(k, r)^t \mathbf{q}(k, r) - \sum_k \mathbf{p}(k, r)^t \sum_k \mathbf{q}(k, r)} \quad (3.28)$$

On constate que les données extérieures à la zone vue par la planète masquée (Eq. 3.21) ne sont pas prises en compte dans cette correction du biais. L'information a en effet été définitivement perdue lorsque  $\hat{\alpha}(r)$  a été calculée. Ainsi, quoique le biais de  $\hat{\alpha}'(r)$  soit nul, on s'attend à ce que sa variance soit supérieure à celle d'une estimation faite à partir d'un profil non masqué.

Supposons maintenant que la norme 1 (somme simple de toutes les composantes d'un vecteur si celles-ci sont non négatives) de chaque profil  $\mathbf{p}(k, r)$  est identique quel que soit l'instant  $k$  (hypothèse en principe vérifiée) :

$$\sum_{\ell=1}^L q_{\ell}(k, r) = q \quad \forall k = 1, \dots, K \quad \forall r \quad (3.29)$$

Une première estimation du nombre de photons  $n$  reçus de la planète sur n'importe quelle image  $k$  est donc donnée par :

$$\hat{n}(r) = \hat{\alpha}'(r) \times q \quad (3.30)$$

En effet, estimer le nombre de photo-électrons  $n$  reçus de la planète sur n'importe quelle image  $k$  consiste à estimer  $\alpha q$ . Comme  $\hat{\alpha}'(r)$  est un estimateur non biaisé de  $\alpha$ ,  $\hat{n}(r)$  donné ci-dessus est un estimateur non biaisé de  $\alpha q$  :  $\mathbf{E}[\hat{n}(r)] = \alpha q$ .

Comme la correction  $\hat{\alpha}(r) \rightarrow \hat{\alpha}'(r)$  est linéaire nous pouvons cependant par simplicité continuer à utiliser  $\hat{\alpha}(r)$  pour l'étude de la consistance de l'estimation et la construction de la méthode de détection et réajuster *a posteriori* les résultats.

### 3.3 Consistance des estimateurs et estimation d'incertitude

Si  $r$  est connu,  $\hat{\alpha}_{\text{ML}} = \hat{\alpha}(r)$  est un estimateur non biaisé, efficace et consistant

- **Efficacité : borne de Rao-Cramer atteinte à  $K$  fixé (donc  $\hat{\alpha}$  meilleur estimateur)**

Si  $r$  était connu (ou estimé avec une erreur négligeable quel que soit  $\mathbf{x}$ ), comme exprimé par l'Eq. (3.5) le modèle serait gaussien et linéaire par rapport à  $\boldsymbol{\theta}$ . D'après l'Eq. (3.23) nous aurions alors directement  $\hat{\alpha}_{\text{ML}} = \hat{\alpha}(r)$  évalué en la valeur  $r$  vraie et plus généralement  $\hat{\boldsymbol{\theta}}_{\text{ML}} = \hat{\boldsymbol{\theta}}(r)$  tel que donné par l'Eq. (3.7).

Pour éviter toute confusion avec le cas général où  $r$  est inconnu, nous n'utilisons pas la notation  $\hat{\alpha}_{\text{ML}}$  et par simplicité nous gardons  $r$  implicite, de sorte que dans cette thèse dès que  $r$  sera supposé connu nous appellerons  $\hat{\alpha}$  l'estimateur  $\hat{\alpha}(r)$  évalué en la position initiale vraie.

Nous allons redémontrer le résultat classique selon lequel pour le modèle gaussien linéaire en  $\boldsymbol{\theta}$  de l'Eq. (3.5)  $\hat{\boldsymbol{\theta}}$  est non biaisé et sa covariance atteint la borne de Rao-Cramer, et donc son erreur est aussi petite que possible, voir Bickel and Doksum [2001]. En particulier,  $\hat{\boldsymbol{\theta}}$  estime en moyenne correctement  $\boldsymbol{\theta}$  et sa variance est uniformément (pour toute valeur de  $\alpha \in \mathbb{R}_+$ ) minimale parmi tous les estimateurs possible.

Les moments de  $\hat{\boldsymbol{\theta}}$  sont en effet simplement donnés par

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\theta}}] &= \mathbb{E}[(A^t A)^{-1} A^t \mathbf{x}] = (A^t A)^{-1} A^t A \boldsymbol{\theta} = \boldsymbol{\theta} \\ \text{cov}[\hat{\boldsymbol{\theta}}] &= \text{cov}[(A^t A)^{-1} A^t \mathbf{x}] = (A^t A)^{-1} A^t \text{cov}[\mathbf{x}] A (A^t A)^{-1} = \sigma^2 (A^t A)^{-1} \end{aligned}$$

où  $(A^t A)^{-1}$  est donné par l'Eq. (3.10).

D'autre part, comme  $\hat{\boldsymbol{\theta}}$  est non biaisé la borne de Rao-Cramer est donnée par l'inverse de la matrice d'information de Fisher. Celle-ci est bien définie ici car les trois hypothèses suivantes sont vérifiées :

- $\mathcal{X} = \{\mathbf{x} : p(\mathbf{x}|\boldsymbol{\theta}) > 0\}$  ne dépend pas de  $\boldsymbol{\theta}$  puisque  $\mathcal{X} = \mathbb{R}^{KL} \forall \boldsymbol{\theta}$ .  $\frac{\partial \log p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i}$  existe toujours et est non nul.
- $\hat{\boldsymbol{\theta}}$  est une statistique telle que  $\mathbb{E}[|\hat{\boldsymbol{\theta}}|_i] < \infty \forall \boldsymbol{\theta}$  pour chaque composante  $i$ . Et alors  $\frac{\partial}{\partial \theta_i} \left[ \int \hat{\boldsymbol{\theta}}(\mathbf{x}) p(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x} \right] = \int \hat{\boldsymbol{\theta}}(\mathbf{x}) \frac{\partial p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i} d\mathbf{x}$
- L'espace des paramètres est un ouvert de  $\mathbb{R}^{L+1}$ .

La composante  $(i, j)$  de la matrice d'information est alors définie par

$$I_{i,j}(\boldsymbol{\theta}) = \left( \mathbb{E} \left[ \frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i} \times \frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_j} \right] \right)$$

Comme  $p(\cdot|\boldsymbol{\theta})$  est différentiable deux fois et que l'intégrale double et la différenciation sous le signe intégral peuvent être interchangés,

$$I_{i,j}(\boldsymbol{\theta}) = -\mathbb{E} \left[ \frac{\partial^2 \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right]$$

Donc ici,

$$\begin{aligned} I_{i,j}(\boldsymbol{\theta}) &= -\mathbb{E} \left[ \frac{\partial}{\partial \theta_i} \left( \frac{\partial}{\partial \theta_j} \left( -\frac{(\mathbf{x} - A\boldsymbol{\theta})^t (\mathbf{x} - A\boldsymbol{\theta})}{2\sigma^2} \right) \right) \right] = \frac{1}{2\sigma^2} \mathbb{E}_{\boldsymbol{\theta}} \left[ \frac{\partial}{\partial \theta_i} \left( -2 \frac{\partial \boldsymbol{\theta}^t A^t \mathbf{x}}{\partial \theta_j} + \frac{\partial \boldsymbol{\theta}^t A^t A \boldsymbol{\theta}}{\partial \theta_j} \right) \right] \\ &= \frac{1}{2\sigma^2} \mathbb{E}_{\boldsymbol{\theta}} \left[ \frac{\partial}{\partial \theta_i} (-2(A^t \mathbf{x})_j + 2(A^t A \boldsymbol{\theta})_j) \right] = \frac{1}{\sigma^2} \mathbb{E}_{\boldsymbol{\theta}} [(A^t A)_{i,j}] = \frac{(A^t A)_{i,j}}{\sigma^2} \end{aligned}$$

et donc finalement

$$\mathbb{E}[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta} \tag{3.31}$$

$$\text{et } \text{cov}[\hat{\boldsymbol{\theta}}] = \sigma^2 (A^t A)^{-1} = I(\boldsymbol{\theta})^{-1} \quad \text{avec } (A^t A)^{-1} \text{ donné par l'Eq. (3.10)} \tag{3.32}$$

Le fait que la borne de Rao-Cramer soit atteinte implique qu'il n'est pas possible de trouver un estimateur  $\hat{\boldsymbol{\theta}}'$  non biaisé dont la matrice de covariance soit plus faible, c'est-à-dire telle que la matrice  $\text{cov}[\hat{\boldsymbol{\theta}}] - \text{cov}[\hat{\boldsymbol{\theta}}']$  soit définie positive. Un estimateur dont la matrice de covariance est aussi faible que la borne de Rao-Cramer est dit *efficace*.

La variance de  $\hat{\alpha}$  est donnée par la  $L+1$  ème composante de  $\text{cov}[\hat{\boldsymbol{\theta}}]$ , c'est-à-dire utilisant l'Eq. (3.10) :

$$\text{var}[\hat{\alpha}] = \frac{\sigma^2}{K_{CK}} \tag{3.33}$$

• **Erreur quadratique moyenne de l'estimateur tend vers 0 lorsque  $K \rightarrow \infty$**

Sous des conditions satisfaites dans notre cas en s'autorisant à observer plusieurs nuits, chaque terme de la matrice de covariance de l'estimateur  $\hat{\boldsymbol{\theta}}$  (donnée par l'Eq. (3.32) tend vers 0 lorsque le nombre d'images  $K$  "devient grand". Comme de plus le biais (erreur moyenne par rapport à la valeur vraie  $\boldsymbol{\theta}$ ) est nul, ceci implique que  $\hat{\boldsymbol{\theta}}$  converge en moyenne quadratique vers le paramètre vrai  $\boldsymbol{\theta}$ . Notons que la convergence de l'erreur quadratique moyenne vers 0 implique la convergence en probabilité vers le paramètre vrai, Lehmann and Casella [1998]. On dit que  $\hat{\alpha}$  est un estimateur *consistant* car, grossièrement, " $\hat{\alpha} \rightarrow \alpha$ " lorsque  $K \rightarrow \infty$ . Revenons maintenant dans les détails.

Chaque composante de  $\mathbf{p}(k)$  est bornée quel que soit  $k$  donc les termes du numérateur de  $(A^t A)^{-1}$  (donnée par l'Eq. 3.10) sont bornés donc chaque terme du numérateur de la matrice de covariance de  $\hat{\boldsymbol{\theta}}$  (Eq. 3.32) est borné. Ainsi, chaque terme de la matrice de covariance tend vers 0 si et seulement si  $Kc_K \rightarrow \infty$ . Ceci se produit par exemple dans les cas suivants :

- $p_{\min} = \min_{k=1, \infty} \|\mathbf{p}(k)\|^2 > 0$  et  $p_{\text{ov}} = \max_{k=1, \infty} \sum_{k' \neq k}^{\infty} \mathbf{p}(k)^t \mathbf{p}(k') < \infty$  existe. La seconde hypothèse signifie que si tous les  $\mathbf{p}(k)$  ont un support fini, la planète n'est pas autorisée à croiser un nombre infini de fois la même zone. Sous ces hypothèses,  $Kc_K \rightarrow \infty$  car, utilisant l'Eq. (3.18)  $Kc_K = \sum_k \|\mathbf{p}(k)\|^2 - K^{-1} \|\sum_k \mathbf{p}(k)\|^2$  et d'autre part

$$\sum_k \|\mathbf{p}(k)\|^2 - K^{-1} \|\sum_k \mathbf{p}(k)\|^2 > (K-1)p_{\min} - p_{\text{ov}} \quad (3.34)$$

- Il existe  $0 < a < 1$  et  $b > 0$  tels que  $\forall k, k', \|\mathbf{p}(k) - \mathbf{p}(k')\|^2 > b|k - k'|^{-a}$ . Cette condition suppose que les différences entre profils varient au moins comme la puissance de l'intervalle de temps (compté en nombre de poses). On utilise cette fois l'Eq. (3.17) qui donne  $Kc_K = K^{-1} \sum_k \sum_{k' < k} \|\mathbf{p}(k) - \mathbf{p}(k')\|^2$  et d'autre part sous cette condition on a :

$$\begin{aligned} K^{-1} \sum_k \sum_{k' < k} \|\mathbf{p}(k) - \mathbf{p}(k')\|^2 &> \frac{b}{K} \sum_k \sum_{k' < k} |k - k'|^{-a} = \frac{b}{K} \sum_{k=1}^K (K-k)k^{-a} = b \left( \frac{K^{1-a}}{a^2 - 3a + 2} + \zeta(a) + o(1) \right) \\ &> \frac{b}{a^2 - 3a + 2} K^{1-a} + O(1) \end{aligned}$$

où  $\zeta$  est la fonction Zeta de Riemann avec  $|\zeta(a)| < \infty$  lorsque  $0 < a < 1$ . Comme  $a^2 - 3a + 2 = (a-1)(a-2) > 0$  si  $a < 1$  et que  $K^{1-a}$  tend vers l'infini lorsque  $K \rightarrow \infty$ , sous cette condition on a bien  $Kc_K \rightarrow \infty$  lorsque le nombre de poses tend vers l'infini.

- Le dernier exemple de condition suffisante ne requiert pas que  $\|\mathbf{p}(k) - \mathbf{p}(k')\|^2 \neq 0$  pour tous les  $k, k'$  mais seulement pour une fraction d'entre eux. Définissons pour un  $c > 0$  donné

$$\mathcal{S}_K = \{(k, k') \in \llbracket 1, K \rrbracket^2, k > k' : \|\mathbf{p}(k) - \mathbf{p}(k')\|^2 > c\} \quad (3.35)$$

$\text{Card}(\mathcal{S}_K) = \sum_{k=1}^K f(k)$  avec  $f(k) = \text{Card}(\{k' \in \llbracket 1, k \rrbracket / \|\mathbf{p}(k) - \mathbf{p}(k')\|^2 > c\})$ . Sous l'hypothèse  $f(k) = \lfloor (\frac{k}{\lambda})^a \rfloor + 1$  où  $0 < a \leq 1$  et  $\lfloor x \rfloor$  représente la partie entière de  $x$ ,

$$\begin{aligned} K^{-1} \sum_k \sum_{k' < k} \|\mathbf{p}(k) - \mathbf{p}(k')\|^2 &\geq K^{-1} \sum_{(k, k') \in \mathcal{S}_K} \|\mathbf{p}(k) - \mathbf{p}(k')\|^2 > \frac{c}{K} \text{Card}(\mathcal{S}_K) \\ &> \frac{c}{K} \sum_{k=1}^K \left( \lfloor (\frac{k}{\lambda})^a \rfloor + 1 \right) \geq \frac{c}{K} \sum_{k=1}^K \left( \frac{k}{\lambda} \right)^a = \left( \frac{K}{\lambda} \right)^a \frac{c}{K} \sum_{k=1}^K \left( \frac{k}{K} \right)^a \\ &> \left( \frac{K}{\lambda} \right)^a c \left( \frac{1}{1+a} + \frac{1}{2K} + \frac{\zeta(-a)}{K^{a+1}} + o\left( \frac{1}{K^3} \right) \right) \\ &> \left( \frac{K}{\lambda} \right)^a \frac{c}{1+a} + o(K^a) \end{aligned}$$



Finalement, si  $r$  est connu et sous une des conditions listées ci-dessus, l'estimateur  $\hat{\alpha}$  converge en moyenne quadratique vers la vraie valeur, et est donc consistant. Dans le cas de la rotation de champ, présente notamment sur la monture du télescope qui alimente l'instrument SPHERE, la première condition n'est pas tout à fait valide si l'on considère qu'il faut plus d'une nuit pour avoir  $\hat{\alpha} \approx \alpha$ , mais la troisième est toujours valide, voir la discussion dans la section 4.1.

### Consistance de $\hat{\alpha}_{\text{ML}}$ si $r$ est inconnu

Le fait que  $r$  soit inconnu rend les résultats précédents incorrects en général, au moins au sens strict puisque  $\hat{\alpha}_{\text{ML}}$  prend alors une forme non linéaire, donnée au mieux par l'Eq. (3.26). En considérant  $\alpha, r, \mathbf{d}$  comme les paramètres d'intérêt, le modèle (3.1) lui-même n'est alors plus linéaire en ses paramètres. De plus, à cause du mouvement de la planète  $\mathbf{p}(k, r)$ , notre modèle de type "signal plus bruit" qui peut s'exprimer comme  $\mathbf{x}(k) = \mathbf{s}(k, \boldsymbol{\theta}, r) + \boldsymbol{\epsilon}(k)$ ,  $\boldsymbol{\epsilon}(k) \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_L)$  ne correspond pas au modèle standard *i.i.d.* (indépendant et identiquement distribué). Des propriétés asymptotiques de l'estimateur ML pour le cas indépendant mais non identiquement distribué ont été intensivement étudiées, comme par exemple par Philippou and Roussas [1975] et Hoadley [1971]. Cependant, grâce à la simplicité du modèle (3.1), des conditions suffisantes pour la consistance et la normalité asymptotique de  $(\hat{\mathbf{d}}_{\text{ML}}, \hat{\alpha}_{\text{ML}}, \hat{r}_{\text{ML}}, \widehat{\sigma}_{\text{ML}}^2)$  peuvent être directement obtenues à partir du résultat général obtenu par Sweeting [1980].

Suivant Sweeting [1980], l'hypothèse principale est l'existence d'une matrice déterministe  $B_K$  telle que  $B_K \rightarrow 0$  lorsque  $K \rightarrow \infty$  et telle que  $B_K^{-1} \mathcal{J}_K (B_K^{-1})^t$  converge en moyenne uniforme vers une matrice définie positive, où  $\mathcal{J}_K$  représente l'opposé de la matrice Hessienne de la log-vraisemblance (matrice des dérivées seconde par rapport à  $\boldsymbol{\theta}$ ).

Le calcul du Hessien de taille  $(L+3) \times (L+3)$  pour le modèle (3.1) est immédiat. Son expression est donnée par

$$\frac{\sigma^2}{K} \mathcal{J}_K = \begin{pmatrix} S_K^{0,0} & \mathbf{S}_K^{0t} & \alpha S_K^{0,1} - \sum_k \frac{\mathbf{p}^{(1)}(k,r)^t \boldsymbol{\epsilon}(k)}{K} & \sum_k \frac{\mathbf{p}(k,r)^t \boldsymbol{\epsilon}(k)}{K\sigma^2} \\ \mathbf{S}_K^0 & I_L & \alpha \mathbf{S}_K^{1t} & \sum_k \frac{\boldsymbol{\epsilon}(k)}{K\sigma^2} \\ \alpha S_K^{0,1} - \frac{1}{K} \sum_k \mathbf{p}^{(1)}(k,r)^t \boldsymbol{\epsilon}(k) & \alpha \mathbf{S}_K^1 & \alpha^2 S_K^{1,1} - \frac{\alpha}{K} \sum_k \boldsymbol{\epsilon}(k)^t \mathbf{p}^{(2)}(k,r) & \alpha \sum_k \frac{\mathbf{p}^{(1)}(k,r)^t \boldsymbol{\epsilon}(k)}{K\sigma^4} \\ \sum_k \frac{\mathbf{p}(k,r)^t \boldsymbol{\epsilon}(k)}{K\sigma^2} & \sum_k \frac{\boldsymbol{\epsilon}(k)}{K\sigma^2} & \alpha \sum_k \frac{\mathbf{p}^{(1)}(k,r)^t \boldsymbol{\epsilon}(k)}{K\sigma^4} & \sum_k \frac{\boldsymbol{\epsilon}(k)^t \boldsymbol{\epsilon}(k)}{K\sigma^4} - \frac{L}{2\sigma^2} \end{pmatrix} \quad (3.36)$$

supposant l'ordre  $(\alpha, \mathbf{d}, r, \sigma^2)$  et les notations suivantes :

$$\mathbf{S}_K^i = \frac{1}{K} \sum_{k=1}^K \mathbf{p}^{(i)}(k, r) \quad (3.37)$$

$$S_K^{i,j} = \frac{1}{K} \sum_{k=1}^K \mathbf{p}^{(i)}(k, r)^t \mathbf{p}^{(j)}(k, r) \quad (3.38)$$

où  $(i, j) \in \{(0, 0), (1, 0), (1, 1), (2, 0)\}$  et où  $\mathbf{p}^{(i)}(k, r)$  représente la  $i$ ème dérivée de  $\mathbf{p}(k, r)$  par rapport à  $r$ .

En choisissant  $A_K = \sqrt{K} I_{L+3}$  et en utilisant des résultats standards de convergence des moments empiriques d'échantillons gaussiens *i.i.d.*, on montre que la matrice  $A_K^{-1} \mathcal{J}_K (A_K^{-1})^t$  converge vers une matrice  $W$  si et seulement si les moyennes  $\mathbf{S}_K^i$  and  $S_K^{i,j}$  convergent vers certaines limites  $\mathbf{S}_\infty^i$  and  $S_\infty^{i,j}$ . Ceci constitue le premier ensemble de conditions pour obtenir la consistance de l'estimateur ML dans notre cas. Notons en particulier que si  $r$  est connue ces conditions se réduisent à la condition de la convergence de  $\mathbf{S}_K^0$  et  $S_K^{0,0}$ . Or comme  $Kc_K = K(S_K^{0,0} - \|\mathbf{S}_K^0\|^2)$ , cette condition implique bien la condition que nous attendions dans la section précédente.

Un second ensemble de conditions découle de la contrainte que  $W$  doit être définie positive. De telles conditions sont moins évidentes à trouver. Notons cependant par exemple que l'application du critère de Sylvestre à la limite de (3.37,3.38) requiert  $S_\infty^{0,0} > \|\mathbf{S}_\infty^0\|^2$ , qui revient à la condition d'identifiabilité énoncée dès la phase d'estimation par l'Eq. (3.19).

### Estimation d'incertitude sur l'intensité par l'estimation de $\text{var}[\hat{\alpha}]$

Si  $r$  était connu, que  $\sigma^2$  le soit ou non au moment de l'estimation, la variance de l'estimateur de  $\hat{\alpha}$  est donnée par l'Eq. (3.33). Cependant,  $r$  et  $\sigma^2$  sont inconnus donc pour obtenir une estimation d'incertitude de l'estimation il est classique d'estimer  $\text{var}[\hat{\alpha}]$  en remplaçant les valeurs vraies de  $r$  et  $\sigma^2$  par leur estimateur ML (voir la section 1.2). L'estimateur ML de  $r$  était donné par l'Eq. (3.24). Dérivons maintenant l'estimateur de  $\sigma^2$ .

#### • Estimation de $\sigma^2$ par ML

De la même façon que nous avons minimisé d'abord l'opposé de la log-vraisemblance  $\mathcal{L}_0(\alpha, r, \mathbf{d}, \sigma^2)$  par rapport à  $(\alpha, \mathbf{d})$  pour  $r$  et  $\sigma^2$  fixés (mais les minimiseurs ne dépendaient pas de  $\sigma^2$ ) puis  $\mathcal{L}_0(\hat{\alpha}(r), r, \hat{\mathbf{d}}(r))$  par rapport à  $r$ , nous pouvons maintenant chercher le maximiseur global de la vraisemblance en variant  $\sigma^2$ . Cherchons ainsi le minimiseur  $\hat{\sigma}_{\text{ML}}^2$  de  $\mathcal{L}_0(\hat{\alpha}_{\text{ML}}, \hat{r}_{\text{ML}}, \hat{\mathbf{d}}_{\text{ML}}, \sigma^2)$ . Notons que comme la transformation  $\sigma \rightarrow \sigma^2$  est bijective, l'estimateur du maximum de vraisemblance de  $\sigma$  est obtenu à partir de celui de  $\sigma^2$  par la transformation inverse  $(\hat{\sigma}_{\text{ML}})^2 = \hat{\sigma}_{\text{ML}}^2$ . On constate que la dérivée de  $\mathcal{L}_0$  par rapport à  $\sigma^2$  s'annule en le minimum de  $\mathcal{L}_0$ . Résolvant donc l'équation  $0 = \frac{\sum_k \|\mathbf{x}(k) - \hat{\mathbf{d}}_{\text{ML}} - \hat{\alpha}_{\text{ML}} \mathbf{p}(k, \hat{r}_{\text{ML}})\|^2}{2\sigma^4} - \frac{KL}{2\sigma^2}$  on obtient la forme classique

$$\hat{\sigma}_{\text{ML}}^2 = \sum_{k=1}^K \frac{\|\mathbf{x}(k) - \hat{\mathbf{d}}_{\text{ML}} - \hat{\alpha}_{\text{ML}} \mathbf{p}(k, \hat{r}_{\text{ML}})\|^2}{KL} = \frac{\mathcal{L}(\hat{\alpha}_{\text{ML}}, \hat{\mathbf{d}}_{\text{ML}}, \hat{r}_{\text{ML}})}{L} \quad (3.39)$$

On reconnaît en particulier le critère  $\mathcal{L}$  que nous avons minimisé.<sup>4</sup>

#### • Légères corrections de l'estimation de $\sigma^2$

Plus généralement, cet estimateur de  $\sigma^2$  se réexprime comme une fonction de  $\mathbf{x}$  et de la matrice  $A(\cdot)$  définie au paramètre de position initiale près :

$$\hat{\sigma}_{\text{ML}}^2 = \frac{\mathbf{x}^t \mathbf{x} - \mathbf{x}^t A(\hat{r}_{\text{ML}}) (A^t(\hat{r}_{\text{ML}}) A(\hat{r}_{\text{ML}}))^{-1} A^t(\hat{r}_{\text{ML}}) \mathbf{x}}{KL}$$

Cependant, l'estimateur  $\hat{r}_{\text{ML}}$  de  $r$  donné par l'Eq. (3.24) ou (3.25) ne peut être exprimé comme une fonction analytique simple de  $\mathbf{x}$ . Ainsi, de même que les moments de  $\hat{\boldsymbol{\theta}}_{\text{ML}}$  (estimateur qui est fonction de  $\hat{r}_{\text{ML}}$ ), l'espérance de  $\hat{\sigma}_{\text{ML}}^2$  n'est pas calculable analytiquement.

En revanche, à  $r$  connu, en conservant la ligne choisie dès l'expression de  $\hat{\boldsymbol{\theta}}$  où  $r$  était gardé implicite quand il était connu,

$$\mathbb{E}[\hat{\sigma}^2] = \frac{\mathbb{E}[\mathbf{x}^t \mathbf{x} - \mathbf{x}^t A(A^t A)^{-1} A^t \mathbf{x}]}{KL}$$

Or

$$\begin{aligned} \mathbb{E}[\mathbf{x}^t \mathbf{x}] &= \mathbb{E}[\text{Tr}(\mathbf{x} \mathbf{x}^t)] = \text{Tr}(\text{cov}[\mathbf{x}]) + \mathbb{E}[\mathbf{x}^t] \mathbb{E}[\mathbf{x}] = KL\sigma^2 + \mathbb{E}[\mathbf{x}^t] \mathbb{E}[\mathbf{x}] \\ \mathbb{E}[\mathbf{x}^t A(A^t A)^{-1} A^t \mathbf{x}] &= \sigma^2 \text{Tr}(A(A^t A)^{-1} A^t) + \mathbb{E}[\mathbf{x}^t] A(A^t A)^{-1} A^t \mathbb{E}[\mathbf{x}] \\ &= \sigma^2 \text{Tr}(A^t A(A^t A)^{-1}) + \boldsymbol{\theta}^t A^t A(A^t A)^{-1} A^t A \boldsymbol{\theta} \\ &= (L+1)\sigma^2 + \mathbb{E}[\mathbf{x}^t] \mathbb{E}[\mathbf{x}] \end{aligned}$$

<sup>4</sup>Utilisant sa réexpression donnée par l'Eq. (3.22) on obtient une expression utile si l'on ne souhaite jamais expliciter l'estimation de  $\mathbf{d}$  :

$$\hat{\sigma}_{\text{ML}}^2 = \frac{1}{L} \left( \frac{\sum_k \|\mathbf{x}(k) - \bar{\mathbf{x}}\|^2}{K} - (\hat{\alpha}(\hat{r}_{\text{ML}}))^2 c_K(\hat{r}_{\text{ML}}) \right) \quad (3.40)$$

avec  $\hat{\alpha}(\cdot)$  et  $c_K(\cdot)$  donnés par l'Eq. (3.12, 3.11).

donc

$$\mathbb{E}[\widehat{\sigma^2}] = \sigma^2 \left(1 - \frac{L+1}{KL}\right)$$

Quoique cette espérance soit différente de  $\sigma^2$ , comme elle reste une fonction linéaire de  $\sigma^2$  et que l'espérance est linéaire, il est aisé de définir un estimateur  $\widehat{\sigma^2}'$  non biaisé lorsque  $r$  est connu à partir de l'estimateur ML  $\widehat{\sigma^2}$  :

$$\widehat{\sigma^2}' = \sum_{k=1}^K \frac{\|\mathbf{x}(k) - \hat{\mathbf{d}} - \hat{\alpha}\mathbf{p}(k, r)\|^2}{L(K-1) - 1} \quad (3.41)$$

En pratique l'objectif ici est d'utiliser l'estimation de  $\sigma^2$  pour estimer la variance de l'estimateur d'intensité. Comme en pratique  $r$  est en fait inconnu, on peut estimer la variance du bruit par l'expression

$$\widehat{\sigma^2}' = \sum_{k=1}^K \frac{\|\mathbf{x}(k) - \hat{\mathbf{d}}_{\text{ML}} - \hat{\alpha}_{\text{ML}}\mathbf{p}(k, \hat{r}_{\text{ML}})\|^2}{L(K-1) - 1}$$

Notons cependant que le biais de  $\widehat{\sigma^2}$  était faible pour des valeurs raisonnablement grandes de  $K$  et  $L$ . Comme quoiqu'il en soit  $r$  est en fait inconnu et que le modèle statistique lui-même n'est pas très correct on n'utilise pas les résultats de cette page et on remplace  $r$  par  $\hat{r}_{\text{ML}}$  directement dans l'expression initiale (3.40).

En revanche, il est important, et relativement facile, dans notre cas de minimiser un peu l'erreur de modélisation qui consistait à supposer que  $\sigma^2$  était uniforme spatialement. Ce problème de non-uniformité spatiale de  $\sigma^2$  est discuté pour notre cas dans la section 4.1.

#### • Estimation finale d'incertitude sur le nombre de photons émis par la planète en une pose

On injecte finalement un estimateur de  $\sigma^2$  tel que décrit précédemment dans l'expression théorique de la variance de  $\hat{\alpha}$  :

$$\widehat{\text{var}}[\hat{\alpha}] = \frac{K\widehat{\sigma^2}'}{K \sum_k \|\mathbf{p}(k, \hat{r})\|^2 - \|\sum_k \mathbf{p}(k, \hat{r})\|^2} \quad (3.42)$$

Considérant à nouveau les éléments de la section 3.2 sur l'utilisation de  $\mathbf{p}(k, r)$  à la place d'un vrai profil  $\mathbf{q}(k, r)$  on écrit

$$\widehat{\text{var}}[\hat{\alpha}'] = \left( \frac{K \sum_k \|\mathbf{p}(k, \hat{r})\|^2 - \|\sum_k \mathbf{p}(k, \hat{r})\|^2}{K \sum_k \mathbf{p}(k, \hat{r})^t \mathbf{q}(k, \hat{r}) - \sum_k \mathbf{p}^t(k, \hat{r}) \sum_k \mathbf{q}(k, \hat{r})} \right)^2 \widehat{\text{var}}[\hat{\alpha}] \quad (3.43)$$

Finalement, une barre d'erreur sur le nombre de photons émis par la planète sur le temps d'une pose peut être estimée, avec  $q$  définie par l'Eq. (3.29), par :

$$\widehat{\sigma}(\hat{n}) = q \times \sqrt{\widehat{\text{var}}[\hat{\alpha}']}. \quad (3.44)$$

### 3.4 Détection initiale : obtention d'une liste de candidats

Nous avons terminé de construire la procédure à appliquer pour procéder à l'estimation de l'intensité et la position d'une potentielle planète. Il est néanmoins crucial de se demander si une planète est effectivement présente dans le jeu de données. La phase de détection correspond à un test d'hypothèses : il s'agit de décider entre une hypothèse nulle (aucune planète n'est présente) et son hypothèse alternative (une planète est présente). Une introduction au test d'hypothèses est donnée dans la section 1.3.

### Test de détection global dans tout le champ

Le test de détection que l'on souhaiterait en général *in fine* effectuer à partir d'un ensemble de données  $\mathbf{x}$  consiste à choisir entre les hypothèses  $H_0$  et  $H_1$  définies par

$$H_0 : \alpha = 0 \quad H_1 : \alpha > 0$$

Autrement dit, sous  $H_0$  (hypothèse par défaut) il n'y a pas de planète, et sous  $H_1$  il y a une planète d'intensité  $\alpha > 0$  inconnue et donc non spécifiée dans le test (test simple vs composite).

Comme introduit dans la section 1.3, une procédure de test classique consiste à seuiliser le rapport de vraisemblance généralisé (noté GLR pour *Generalized Likelihood Ratio*). Pour le modèle (3.2), celui-ci est défini par :

$$\begin{aligned} & \text{Rejeter } H_0 \text{ si } \text{GLR}(\mathbf{x}) \leq \zeta \\ \text{où } \text{GLR}(\mathbf{x}) &= \frac{\max_{\mathbf{d}, \sigma^2} p(\mathbf{x} | \alpha = 0, \mathbf{d}, \sigma^2)}{\max_{\alpha, r, \mathbf{d}, \sigma^2} p(\mathbf{x} | \alpha, r, \mathbf{d}, \sigma^2)} \end{aligned}$$

Les estimateurs ML ont été dérivés dans la section précédente sous  $H_1$ . Sous l'hypothèse nulle, par définition  $\alpha = 0$  et  $r$  n'est pas défini, de sorte que les expressions mathématiques des estimateurs sous  $H_0$  sont obtenues à partir des expressions des estimateurs sous  $H_1$  en remplaçant seulement  $\hat{\alpha}$  par 0.

D'après les expressions de la vraisemblance et de l'estimateur ML de la variance données respectivement par les Eq. (3.2) et (3.39), on constate après quelques calculs que

$$\text{GLR}(\mathbf{x}) = \left[ \frac{\widehat{\sigma^2}_{H_1}(\mathbf{x})}{\widehat{\sigma^2}_{H_0}(\mathbf{x})} \right]^{\frac{KL}{2}}$$

Le GLR est égal au rapport des puissances des “bruits résiduels” (estimations de variance) selon que l'on est sous une hypothèse ou une autre. D'après l'Eq. (3.22) il peut aussi se réexprimer à partir de la statistique  $\hat{\alpha}(\hat{r}_{\text{ML}})$  seulement :

$$\begin{aligned} \text{GLR}(\mathbf{x}) &= \left[ \left[ \frac{\sum_k \|\mathbf{x}(k) - \bar{\mathbf{x}}\|^2}{K} - [\hat{\alpha}(\hat{r}_{\text{ML}})]^2 c_K(\hat{r}_{\text{ML}}) \right] \left( \frac{\sum_k \|\mathbf{x}(k) - \bar{\mathbf{x}}\|^2}{K} \right)^{-1} \right]^{\frac{KL}{2}} \\ &= \left[ 1 - [\hat{\alpha}(\hat{r}_{\text{ML}})]^2 \frac{\sum_k \|\mathbf{p}(k, \hat{r}_{\text{ML}}) - \bar{\mathbf{p}}(\hat{r}_{\text{ML}})\|^2}{\sum_k \|\mathbf{x}(k) - \bar{\mathbf{x}}\|^2} \right]^{\frac{KL}{2}} \end{aligned}$$

avec  $\hat{\alpha}(\cdot)$ ,  $c_K(\cdot)$  et  $\hat{r}_{\text{ML}}$  donnés respectivement par les Eq. (3.12), (3.11) et (3.24). Le GLR peut ainsi être calculé sans expliciter l'estimation du fond dû à l'étoile  $\hat{\mathbf{d}}_{\text{ML}}$ , dont l'expression n'est pas de premier intérêt pour la détection d'exoplanète et l'estimation de ses paramètres.

L'interprétation de la statistique  $\text{GLR}(\mathbf{x})$  n'est pas suffisamment claire pour permettre de donner un sens direct utilisable en pratique au seuil  $\zeta$ . En particulier, on peut noter que  $\text{GLR}(\mathbf{x}) \leq 1$  quel que soit  $\mathbf{x}$ , de sorte qu'il faudrait choisir et donner un sens à un seuil  $\zeta$  nécessairement inférieur à 1, contrairement aux tests bayésiens standards (voir la section 1.3) pour lesquels la valeur 1 est une sorte de valeur “neutre”.

Souvent, et en particulier en l'absence d'une interprétation immédiate claire, le seuil  $\zeta$  est choisi en adoptant l'approche de Neyman-Pearson qui consiste à fixer un certain niveau de significativité, ou Probabilité de Fausse Alarme (PFA, typiquement au niveau de significativité de 5%), et à en déduire le seuil  $\zeta$  associé. Cette approche est décrite dans la section 1.3. La PFA associée au GLR est calculée à partir de la distribution statistique du GLR sous l'hypothèse  $H_0$ . L'expression du GLR vue comme une fonction de  $\mathbf{x}$  seulement est ici donnée comme une fonction en partie implicite *via*  $\hat{r}_{\text{ML}}$  notamment. Comme nous l'avons déjà mentionné, à cause du fait que les profils  $\mathbf{p}(k, r)$  ne peuvent s'exprimer comme une fonction simple de  $r$ , l'estimation de  $\hat{r}_{\text{ML}}$  ne peut être obtenue qu'en scannant tous les jeux de

profils  $\mathbf{p}(k, r)$  (“tous” : pour toutes les positions initiales possibles  $r$ ). Le GLR étant déjà long à calculer pour un jeu de données, sa distribution statistique obtenue en laissant  $\mathbf{x}$  décrire sa propre distribution (gaussienne) est simplement hors de portée.

Quoiqu'il en soit, comme nous le verrons plus loin (voir la section 6.3), le modèle gaussien étudié jusqu'ici est en général une description trop approximative pour calculer la PFA de n'importe quelle statistique de test. Cette distribution s'avère utile pour construire une procédure applicable sur de gros jeux de données et en particulier pour construire une fonction linéaire des données à appliquer en chaque pixel  $r$ , mais il est essentiel de recourir à une seconde phase de détection-estimation basée sur un modèle plus précis des données. Le modèle fin présenté en partie II aura pour objet de dériver un modèle fin des données.

La méthode de détection que nous nous apprêtons à construire sert donc de phase d'initialisation, ayant pour objet de sélectionner les meilleurs candidats pour la phase finale de détection. A l'issue de la phase d'initialisation, il est essentiel d'avoir non pas un résultat de détection ou non à l'échelle de tout le champ d'intérêt, mais un ensemble de candidats potentiels dans le champ, sur lesquels nous appliquerons dans un second temps une méthode de détection plus raffinée et donc plus gourmande en temps de calcul.

Il serait possible d'utiliser une procédure de détection globale pour obtenir finalement plusieurs candidats. Un algorithme itératif de type CLEAN consiste ainsi à appliquer une première fois l'étape de détection globale, de supprimer la contribution de l'objet estimé du jeu d'images, et de réitérer la procédure de détection autant de fois que souhaité. Cependant, ceci requiert de retraiter les données à chaque fois ce qui est extrêmement long lorsque le modèle direct n'est pas une simple convolution bruitée. Le modèle direct depuis  $\alpha$  n'est effectivement pas une convolution à cause de la dépendance en  $r$  de la forme du profil donc une telle approche itérative n'est pas envisageable dans notre cas si l'on souhaite disposer de trop de candidats dans la seconde phase de détection. L'étude des propriétés d'une telle procédure itérative de détection n'est de plus pas simple.

## Test de détection pour chaque position initiale $r$

### • PFA globale vs PFA par pixel

A la place d'associer à un jeu de données  $\mathbf{x}$  un résultat global de détection ou non détection, on lui associe pour *chaque* position initiale  $r$  testée le résultat “détection” ou “non détection”. Les PFA associées à un test de détection *global* et *par pixel* ne signifient pas la même chose. En particulier, il serait incohérent de souhaiter fixer l'une et l'autre à 5%. Empiriquement (calcul de “fréquence” qui tend asymptotiquement vers une probabilité), pour le premier type de test le nombre de fausses alarmes est calculé en comptant le nombre de jeux de données qui ont mené à une détection alors qu'il n'y avait pas de planète, alors qu'avec le second type de test, si on suppose que  $\hat{\alpha}(r)$  et  $\hat{\alpha}(r')$  sont décorrélés pour  $r \neq r'$ , alors le nombre de fausses alarmes est le nombre de positions initiales pour lesquelles une planète a été détectée alors qu'il n'y en avait pas à cette position (décompte poursuivi pour un grand nombre de jeux de données). Quantitativement, on constate à partir d'hypothèses simples que ces deux PFA sont en effet très distinctes l'une de l'autre, voir la section C.2.

### • Test du GLR

Une autre idée, plus simple à appliquer et à étudier, consiste à appliquer un test de détection en chaque position initiale “possible”. Dans ce cas, le test du GLR dont l'expression est donnée ci-dessus se réécrirait par exemple

$$\text{Rejeter } H_0 \text{ en } r \text{ si } \text{GLR}(\mathbf{x}, r) \leq \zeta(r)$$

$$\text{où } \text{GLR}(\mathbf{x}, r) = \left[ 1 - \hat{\alpha}(r)^2 \frac{\sum_k \|\mathbf{p}(k, r) - \bar{\mathbf{p}}(r)\|^2}{\sum_k \|\mathbf{x}(k) - \bar{\mathbf{x}}\|^2} \right]^{\frac{KL}{2}}$$

Si l'on ajoute de plus la contrainte que si  $\hat{\alpha}(r) < 0$  alors on redéfinit  $\hat{\alpha}(r) = 0$ , notons que ce test est équivalent à

$$\text{Rejeter } H_0 \text{ en } r \text{ si } S(\mathbf{x}, r) \geq \sqrt{1 - \zeta(r)^{\frac{2}{KL}}}$$

$$\text{avec } S(\mathbf{x}, r) = \frac{\hat{\alpha}(\mathbf{x}, r) \sqrt{Kc_K(r)}}{\sqrt{\sum_{k=1}^K \|\mathbf{x}(k) - \bar{\mathbf{x}}\|^2}}$$

Comme expliqué dans le paragraphe ci-dessus, ce test peut être appliqué en pratique pour rechercher des candidats sur lesquels appliquer la seconde phase de détection-estimation. Pour cela, on ordonne les  $\{S(\mathbf{x}, r), r \in \llbracket 1, L \rrbracket\}$  et on conserve comme positions initiales à réétudier celles correspondant aux plus grandes valeurs de  $S(\mathbf{x}, r)$ .

- **Seuil de l'estimation d'intensité (statistique linéaire)**

Le test du GLR proposé ci-dessus n'est cependant pas linéaire en les données  $\mathbf{x}$ . Son expression est trop compliquée pour que l'on soit *a priori* capable d'obtenir sa distribution statistique sous le modèle fin décrit dans la partie II.

Une autre procédure de test simple et intuitive consiste à détecter une planète en  $r$  si  $\hat{\alpha}(r)$ , l'estimation d'intensité de la planète située initialement en  $r$  (Eq. (3.12)) dépasse un seuil convenablement choisi :

$$\text{Rejeter } H_0 \text{ en } r \text{ si } \hat{\alpha}(\mathbf{x}, r) > \zeta(r) \quad (3.45)$$

$$\text{où } \hat{\alpha}(\mathbf{x}, r) \text{ est défini par l'Eq. (3.12)} \quad (3.46)$$

Nous avons maintenant à disposition une statistique linéaire des données  $\mathbf{x}$ . Le test peut être alors appliqué comme expliqué pour le GLR afin d'obtenir une liste de candidats.

### 3.5 Extension de la détection-estimation à l'imagerie simultanée dans plusieurs canaux spectraux

#### Extension du modèle mono-canal

Lorsque des images peuvent être acquises simultanément dans  $M$  canaux on peut étendre le modèle mono-bande développé dans la section 3.1 en imposant des contraintes sur la relation entre le fond dû à l'étoile dans les différents canaux (ex. : fonds égaux entre deux canaux à un facteur multiplicatif près *etc*).

En nous focalisant sur le cas DBI ( $M = 2$ ) de l'instrument IRDIS de SPHERE, on suppose que le fond dû à l'étoile est quasi-identique dans les deux bandes mais que l'intensité de la planète est en général différente. Les paires de filtres du mode DBI ont en effet été choisies de telle sorte que les deux filtres couvrent des bandes spectrales étroites et très proches spectralement, l'un des deux filtres correspondant à une raie d'absorption d'un élément atmosphérique classique pour une planète mais non pour une étoile. La paire de filtres considérée partout dans cette thèse est par exemple définie par la bande dite H2 ne laissant passer que la lumière de longueur d'onde  $1.57\mu m \leq \lambda_1 \leq 1.64\mu m$  et la bande dite H3 définie par  $1.62\mu m \leq \lambda_2 \leq 1.69\mu m$ , soit  $\boldsymbol{\lambda} \sim (1.60, 1.65)^t \mu m$ .

Nous modélisons ainsi, en étendant les notations du modèle (Eq. 3.1), une image acquise à l'instant  $k$  dans la bande  $m$ , avec  $M = 2$  en mode DBI, par

$$\mathbf{x}(k, m) = \mathbf{d} + \alpha_m \mathbf{p}(k, m, r) + \boldsymbol{\epsilon}(k, m), \quad k \in \llbracket 1, K \rrbracket \quad m \in \llbracket 1, M \rrbracket \quad (3.47)$$

en supposant que les images sont indépendantes entre les différents canaux. En l'absence de réfraction différentielle, l'optique géométrique indique que la position initiale  $r$  de la planète sur la CCD est identique dans les deux bandes. Les notations sont une simple extension des notations du modèle mono-bande :

- $\mathbf{d}$  est un vecteur de taille  $L \times 1$  qui représente la contribution de l'étoile, supposée identique dans toutes les bandes spectrales. Cette contribution est toujours supposée déterministe, stationnaire dans le temps, non uniforme spatialement (et sans contrainte particulière) et inconnue.
- $\alpha_m$  est un scalaire représentant l'intensité de la potentielle planète dans la bande  $m$ , qu'on concatène en  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)^t$ .
- $\mathbf{p}(k, m, r)$  est la réponse déterministe de la planète dans la pose  $k$  et la bande  $m$ . Il s'agit d'un vecteur de taille  $L \times 1$  dont le seul paramètre inconnu est la position initiale  $r$ .
- $\boldsymbol{\epsilon}(k, m)$  est un vecteur de bruit aléatoire gaussien supposé indépendant dans le temps, dans l'espace et entre les deux canaux spectraux et de variance  $\sigma^2$  uniforme dans le champ considéré :  $\boldsymbol{\epsilon}(k, m) \sim \mathcal{N}_L(\mathbf{0}, \sigma^2 I_L)$ , où  $\sigma^2$  est inconnu.

Revenons sur le modèle en tant que description physique des données. Les canaux  $m$  sont caractérisés par une longueur d'onde  $\lambda_m$  spécifique. Le profil de diffraction *on-axis* de l'étoile et la zone de correction d'optique adaptative sont ainsi légèrement plus larges –d'un même facteur radial– pour la longueur d'onde la plus grande. Cet effet est en général pris en compte dans les techniques actuelles dites SDI (*Spectral Differential Imaging*). Ici, comme dans ces techniques SDI, plutôt qu'introduire dans  $\mathbf{d}$  une paramétrisation un peu complexe à gérer en pratique dans la phase d'estimation, nous décidons de prétraiter toutes les images dans la bande 2 (par exemple) en changeant leur échelle radiale du facteur  $\lambda_1/\lambda_2$  pour que les fonds moyens dans les deux bandes coïncident *a priori* mieux. Nous appelons cette transformation d'image : *rescaling*.

Après *rescaling*, la position initiale  $r$  de la planète dans la bande 2 est transformée par le même changement d'échelle. Cependant, il n'est pas utile de définir explicitement une position initiale " $r_2$ " dans la bande 2 car la dépendance du profil en un certain canal  $m$  est déjà prise en compte lorsque l'on écrit  $\mathbf{p}(k, m, r)$ .  $r_2$  serait par ailleurs simplement une fonction déterministe et connue de  $r_1$ . Nous notons donc simplement  $r$  la position initiale de la planète pour  $m = 1$ .

Le modèle mathématique des images après *rescaling* s'écrit donc toujours comme le modèle (3.47) sauf que les profils  $\mathbf{p}(k, 2, r)$  doivent être légèrement transformés (*a priori* translatés radialement d'un facteur  $\lambda_1/\lambda_2$  uniquement).

Le traitement de la position initiale  $r$  pour  $M > 1$  est identique au cas mono-bande, n'induisant même pas de changement de notation. Nous supposons par conséquent dans la suite que  $r$  est connu et le gardons implicite.

### Estimateur d'intensité de la planète dans les différentes bandes

Le modèle (3.47) peut être réexprimé sous une forme vectorielle linéaire en concaténant convenablement vecteurs et matrices. On définit  $\mathbf{x}(k)$  comme le vecteur de taille  $LM \times 1$  constitué de l'ensemble des images obtenues simultanément dans les différentes bandes :  $\mathbf{x}(k)^t = (\mathbf{x}(k, 1)^t, \dots, \mathbf{x}(k, M)^t)$ . Les ensembles d'images sont alors indépendants et décrits par

$$\mathbf{x}(k) = A(k)\boldsymbol{\theta} + \boldsymbol{\epsilon}(k)$$

avec

- $\boldsymbol{\theta}$  le vecteur de paramètres de taille  $(L + M) \times 1$  défini par  $\boldsymbol{\theta}^t = (\mathbf{d}^t, \boldsymbol{\alpha}^t)$
- $A(k)$  la matrice de taille  $LM \times (L + M)$  définie par

$$A(k) = \begin{pmatrix} I_L & \mathbf{p}(k, 1) & \mathbf{0}_{L \times 1} & \dots & \mathbf{0}_{L \times 1} \\ I_L & \mathbf{0}_{L \times 1} & \mathbf{p}(k, 2) & \dots & \mathbf{0}_{L \times 1} \\ I_L & \mathbf{0}_{L \times 1} & \mathbf{0}_{L \times 1} & \dots & \mathbf{p}(k, M) \end{pmatrix}$$

Le maximum de vraisemblance du paramètre  $\boldsymbol{\theta}$  est obtenu par un calcul analogue au cas mono-bande développé dans la section 3.2. On rappelle que  $r$  est supposé connu et implicite ici. On obtient finalement l'estimateur de  $\boldsymbol{\alpha}$  à partir des  $M$  dernières composantes de  $\hat{\boldsymbol{\theta}}$ .

Pour le cas  $M = 2$  qui nous intéresse, on obtient finalement

$$\begin{aligned} \hat{\alpha}_1 = & \left( \sum_k \mathbf{p}(k, 1)^t \sum_k \mathbf{p}(k, 2) \right) \frac{2K \sum_k \mathbf{p}(k, 2)^t \mathbf{x}(k, 2) - \sum_k \mathbf{p}(k, 2)^t \sum_k (\mathbf{x}(k, 1) + \mathbf{x}(k, 2))}{\prod_{m=1}^2 (2K \sum_k \|\mathbf{p}(k, m)\|^2 - \|\sum_k \mathbf{p}(k, m)\|^2) - (\sum_k \mathbf{p}(k, 1)^t \sum_k \mathbf{p}(k, 2))^2} \\ & + \left( 2K \sum_k \|\mathbf{p}(k, 2)\|^2 - \|\sum_k \mathbf{p}(k, 2)\|^2 \right) \frac{2K \sum_k \mathbf{p}(k, 1)^t \mathbf{x}(k, 1) - \sum_k \mathbf{p}(k, 1)^t \sum_k (\mathbf{x}(k, 1) + \mathbf{x}(k, 2))}{\prod_m (2K \sum_k \|\mathbf{p}(k, m)\|^2 - \|\sum_k \mathbf{p}(k, m)\|^2) - (\sum_k \mathbf{p}(k, 1)^t \sum_k \mathbf{p}(k, 2))^2} \end{aligned} \quad (3.48)$$

L'expression de  $\hat{\alpha}_2$  est obtenue par symétrie entre les bandes 1 et 2. Cette forme développée de  $\boldsymbol{\alpha}$  peut être utile car elle évite de définir des vecteurs de plus grande taille et est totalement explicite. Cependant, elle fait intervenir tous les termes croisés entre les différents canaux, de sorte que pour  $M > 2$  la formule devient très longue. Repartons à partir de ce stade au cas pluri-canal  $M$  général.

On peut de même écrire  $\hat{\mathbf{d}}$  comme une fonction des données  $\mathbf{x}$  uniquement :  $\hat{\mathbf{d}}$  est égal aux  $L$  premières composantes de  $\hat{\boldsymbol{\theta}}$ . Mais comme en cas mono-bande, on préfère réécrire  $\hat{\mathbf{d}}$  sous une forme plus intuitive basée sur l'estimateur d'intensité de la planète dans les différents canaux  $\hat{\alpha}$  donné par l'Eq. (3.48).  $\hat{\mathbf{d}}$  est donné par la moyenne de toutes les images auxquelles on retire chaque fois la contribution de la planète :

$$\hat{\mathbf{d}} = \frac{\sum_k \sum_m (\mathbf{x}(k, m) - \hat{\alpha}_m \mathbf{p}(k, m))}{KM} \quad (3.49)$$

Réciproquement, l'estimation d'intensité de  $\alpha_m$  de l'Eq. (3.48) peut s'écrire très simplement à partir des seules données dans la bande  $m$  et de l'estimation du fond donnée par l'Eq. 3.49 :

$$\hat{\alpha}_m = \frac{\sum_k \mathbf{p}(k, m)^t (\mathbf{x}(k, m) - \hat{\mathbf{d}})}{\sum_k \|\mathbf{p}(k, m)\|^2} \quad (3.50)$$

A ce stade on peut rendre la position initiale  $r$  explicite à nouveau en l'injectant dans les profils  $\mathbf{p}(k, m, r)$  et en notant que les estimateurs  $\hat{\mathbf{d}}$  et  $\hat{\alpha}$  sont fonctions de  $r$ . Toujours comme dans le cas mono-bande de la section 3.2, on obtient ensuite l'estimateur de  $r$  en calculant une à une les vraisemblances  $p(\mathbf{x} | \hat{\mathbf{d}}, \hat{\alpha}, r, \sigma^2)$  pour toutes les positions  $r$  testées et en choisissant celle qui la maximise. Les estimateurs ML de  $\boldsymbol{\alpha}$  et  $\mathbf{d}$  sont alors obtenus en réinjectant  $\hat{r}_{\text{ML}}$  dans les estimateurs (3.48) et (3.49). Cependant, toujours par simplicité nous continuons de considérer  $r$  connu et le gardons implicite.

Finalement,  $\sigma^2$  est estimé en minimisant l'opposé de la log-vraisemblance par rapport à  $\sigma^2$  :

$$L_0(\hat{\mathbf{d}}, \hat{\alpha}, \sigma^2) = \frac{\sum_k \sum_m \|\mathbf{x}(k) - \hat{\mathbf{d}} - \hat{\alpha}_m \mathbf{p}(k, m)\|^2}{2\sigma^2} + \frac{KLM}{2} \ln(2\pi\sigma^2) \quad (3.51)$$

L'estimateur de  $\sigma^2$  est obtenu comme le paramètre qui annule la dérivée de  $L_0$  par rapport à  $\sigma^2$  et vaut simplement

$$\widehat{\sigma^2} = \frac{\sum_k \sum_m \|\mathbf{x}(k) - \hat{\mathbf{d}} - \hat{\alpha}_m \mathbf{p}(k, m)\|^2}{KLM} \quad (3.52)$$

## Test de détection

Comme précisé dans la section 3.4 on effectue le test pixel par pixel.  $r$  continue à être tenu implicite ici.



- **Test du GLR**

Injectant l'expression de  $\widehat{\sigma}_{H1}^2$  de l'Eq. (3.47) dans la vraisemblance définie par l'Eq. (3.52), on constate à nouveau que le GLR est simplement donné par

$$\text{GLR}(\mathbf{x}) = \left( \frac{\widehat{\sigma}_{H1}^2(\mathbf{x})}{\widehat{\sigma}_{H0}^2(\mathbf{x})} \right)^{\frac{KLM}{2}} \quad (3.53)$$

$$\text{où } \widehat{\sigma}_{Hi}^2(\mathbf{x}) = \frac{\sum_k \sum_m \|\mathbf{x}(k, m) - \widehat{\alpha}_{m,Hi} \mathbf{p}(k, m) - \widehat{\mathbf{d}}_{Hi}\|^2}{KLM}$$

avec  $\widehat{\alpha}_{H0} = \mathbf{0}$ ,  $\widehat{\alpha}_{m,H1}$  donné par l'Eq. (3.48) et  $\widehat{\mathbf{d}}_{Hi}$  donné par l'Eq. (3.49)

- **Seuil de l'estimation d'intensité dans une bande (statistique linéaire)**

Dans le cadre d'utilisation visé par le mode DBI de SPHERE, on s'attend à ce que l'intensité de la planète soit négligeable dans le canal (noté  $m = 2$ ) correspondant à la bande d'absorption de la planète. Cette hypothèse est centrale dans la technique différentielle SDI si les profils se chevauchent trop entre les deux canaux (ce qui est le cas) et correspond à considérer dans le modèle (3.47) que  $\alpha_2 \ll \alpha_1$  (pour des  $\mathbf{p}(k, m)$  normalisés de façon identiques entre  $m = 1$  et  $m = 2$ ). Il est alors naturel de conserver le test proposé en mono-canal mais appliqué uniquement au canal où l'on s'attend à recevoir du flux de la planète :

$$\text{Rejeter } H_0 \text{ si } \hat{\alpha}_1 > \zeta \quad (3.54)$$

Ceci permet en particulier de conserver une statistique de test linéaire et a du sens si le flux dans la seconde bande est effectivement faible. A l'extrême, pour  $\alpha_2 = 0$  ce test revient précisément au test mono-canal en présence d'un jeu d'images de *référence* dans lequel nulle contribution de planète n'est à rechercher.

- **Extension multivariée du seuil d'estimation d'intensité ?**

Plus généralement, l'extension directe de la procédure de détection mono-bande de la section 3.1 basée sur l'estimation d'intensité consiste à fixer  $M$  seuils  $\zeta_m$  et à rejeter  $H_0$  si et seulement si chacune des estimations d'intensité est supérieure à son seuil respectif :

$$\text{Rejeter } H_0 \text{ si } \hat{\alpha}(r) > \zeta \quad (3.55)$$

avec  $\zeta = (\zeta_1, \dots, \zeta_M)^t$ . La statistique de test est alors  $\alpha$ , toujours linéaire en les données. Cependant, pour  $M > 1$  le nombre de paramètres d'entrée du test (égale à  $M$ , la taille de  $\zeta$ ) est alors strictement supérieur à 1, ce qui empêche l'existence d'une paramétrisation bijective de ce paramètre d'entrée  $\zeta$  par la PFA. Il serait aussi possible de fixer tous les seuils à une unique valeur  $\zeta$  mais ceci serait soit trop soit pas assez conservatif, ou semblerait signifier que l'on aurait pu dès la modélisation écrire que tous les  $\alpha_m$  sont égaux. Il semble donc difficile de tirer profit de cette plus grande flexibilité, contrairement au test du PLR, paramétré par deux paramètres d'entrée, que l'on étudiera dans une prochaine partie. Cette procédure de test basée sur l'estimation d'intensité ne semble donc pas très adaptée si plus qu'une composante est supposée strictement positive.

## Chapitre 4

# Application aux données simulées finement et résultats

### 4.1 Discussions plus détaillées de la validité des conditions

#### Pré-estimation des profils de planète $\mathbf{p}(k, r)$

Dans le modèle (3.1) nous supposons les profils  $\mathbf{p}(k, r)$  connus à la position initiale  $r$  près. Nous avons expliqué dans la section 3.1 en quoi cette hypothèse était justifiée. Donnons maintenant la procédure pour construire les profils  $\mathbf{p}(k, r)$  concrètement.

La rotation de champ est un effet déterministe et connu (Avila and Wirenstrand [2005]) si nous supposons que le centre de la rotation est bien confondu avec l'axe optique *i.e.* le centre de l'image. Si la planète occupe une position initiale notée  $r$  nous connaissons alors la position de la planète à n'importe quel autre instant : elle décrit un cercle à une vitesse angulaire variable dans le temps, qui dépend en un mot de la position apparente de l'étoile. Les seules grandeurs à connaître (inputs de la fonction) sont donc les coordonnées célestes "angle horaire initial et déclinaison" de l'étoile observée, qui sont elles-mêmes directement fonction de la position absolue de l'étoile (ascension droite et déclinaison), de la position du lieu d'observation sur Terre et de la date et heure de début d'observation.

Connaissant la position de la planète au début de la pose  $k$ ,  $\mathbf{p}(k, r)$  est alors obtenu à partir de la PSF instantanée, notée  $\mathbf{p}_{\text{instant}}$ . Nous supposons que celle-ci peut être estimée à partir d'une phase d'étalonnage ou pris analytiquement comme un profil d'Airy par exemple.  $\mathbf{p}(k, r)$  est alors obtenu

- en positionnant  $\mathbf{p}_{\text{instant}}$  sur la position initiale  $r$  (nouvelle image de la même taille que la caméra CCD que l'on note  $\mathbf{p}_{\text{instant}}(r)$ )
- en appliquant à cette image une rotation dont le centre est confondu avec le centre de l'image et dont l'angle est donné par l'angle de rotation de champ subi du début de l'observation jusqu'au début de la pose  $k$
- en appliquant suffisamment de mini-rotations à l'image (chaque mini-rotation de l'image résultant en une *mini-pose*) pour simuler continûment le mouvement de la planète durant la pose  $k$  (on note  $n_k$  le nombre de mini-rotations)
- et enfin en ajoutant ces mini-poses les unes aux autres, ce qui résulte en un profil complet  $\mathbf{p}(k, r)$ .

Ainsi, on simule la réponse  $\mathbf{p}(k, r)$  de la planète sur la pose  $k$  par l'algorithme suivant :

$$\begin{aligned} \mathbf{p}(k, r) &= \sum_{i=0}^{n_k-1} \text{Rot}(\mathbf{p}_{\text{instant}}(r), \Delta\theta(i, k)) \\ \text{avec } \Delta\theta(i, k) &= \int_{t_0}^{t_0+k \times \Delta t + i \times \delta t} dt \dot{\theta}(t) \\ n_k &= \frac{\text{temps de pose}}{\delta t} \\ \text{où } \delta t \text{ est choisi tel que } \delta t &\leq \frac{\epsilon}{\rho_0 \dot{\theta}(t_0 + k \times \Delta t)} \end{aligned}$$

où

- $\dot{\theta}(t)$  est la vitesse angulaire de rotation de champ, connue entièrement dès que l'angle horaire initial et la déclinaison de l'étoile sont connus. Cette fonction est obtenue par un calcul de géométrie sphérique (Avila and Wirenstrand [2005]).
- $t_0$  est le temps initial d'observation (à cet instant la planète est en  $r$ )
- $\Delta t$  est égal au temps de pose plus au temps de lecture
- $\epsilon$  est la fraction de pixels entre les positions de la planète dans deux miniposes successives. Elle est fixée par l'utilisateur selon le degré de "continuité" du mouvement qu'il souhaite fixer.
- $\rho_0$  est la distance en pixels entre le centre de l'image et la planète

Notons que la fonction de rotation "Rot" peut ne consister qu'à effectuer une translation de  $\mathbf{p}_{\text{instant}}(r)$  si la PSF est à symétrie cylindrique.

### Variation spatiale du terme de "normalisation" de $\hat{\alpha}(r)$

Dans l'expression de l'estimateur de  $r$  par méthode ML, nous sommes arrivés à l'Eq. (3.24) donnée par :

$$\hat{r}_{\text{ML}} = \arg \max_r \left\{ \hat{\alpha}(r) \sqrt{c_K(r)} \right\}$$

Si l'on peut négliger la variation de  $c_K(r)$  par rapport à  $r$ , on peut approximer les estimations de position et d'intensité par

$$\begin{aligned} \hat{r}_{\text{ML}} &\approx \arg \max_r \hat{\alpha}(r) \\ \hat{\alpha}_{\text{ML}} &\approx \max_r \hat{\alpha}(r) \end{aligned}$$

Dans le cas spécifique discuté en chapitre 1,  $c_K(r)$  est représenté sur la figure 4.1. La séparation est la composante radiale (en unité –linéaire– d'angle sur le ciel) de la position  $r$ .  $c_K(r)$  ne varie pas en fonction de la composante orthoradiale de  $r$ . On constate que  $c_K(r)$  ne varie alors que de quelques pourcents dans tout le champ, hors zone trop proche de l'étoile.

Plus généralement, pour une étoile située typiquement à plus d'une vingtaine de degrés du zénith et pour une nuit d'observation de plus de 2h l'approximation n'est *a priori* pas mauvaise localement, mais il convient de bien la vérifier pour chaque cas car la vitesse de rotation de champ varie rapidement près du zénith et/ou près du méridien, et le terme  $c_K(r)$  varie quant à lui assez vite près du centre de l'image.

Rappelons quoiqu'il en soit que cette approximation rend le résultat simple et intuitif mais n'aide nullement à la réalisation des calculs.

### Variation spatiale de la variance de bruit $\sigma^2$

Le bruit considéré dans le modèle (3.1) est supposé gaussien. Ce type de bruit particulièrement simple et standard n'est pas très adapté à nos images très marquées par la contribution haute dynamique de

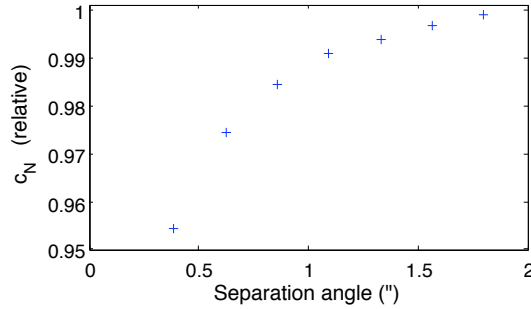


FIG. 4.1: Evolution du terme de normalisation ( $c_N$  dénote en fait  $c_K(r)$ ). Ce terme est fonction de  $r$  et plus spécifiquement fonction de la distance au centre de l'image car il n'y a pas de variation selon la composante orthoradiale de  $r$ .

l'étoile et des effets de diffraction et d'interférence se traduisant par l'existence de *speckles*. Cependant, comme nous l'avons dit, adopter un modèle gaussien était fondamental pour traiter de gros jeux de données. Vient alors la question des corrélations. Le modèle (3.1) suppose que le bruit est non corrélé temporellement, non corrélé spatialement et de variance  $\sigma^2$  identique pour tous les pixels de l'image. Nous avons justifié la première hypothèse de part les temps caractéristiques de corrélation de l'atmosphère et de l'AO. L'hypothèse de non-corrélation spatiale n'est pas justifiable d'un point de vue physique, mais à nouveau en termes de temps de calcul elle était difficilement évitable.

L'hypothèse d'une variance de bruit  $\sigma^2$  identique pour tous les pixels est quant à elle *a priori* assez bien vérifiée pour le bruit provenant du bruit de lecture de la CCD et du bruit de *flat field*. En principe elle ne se justifie cependant pas pour ce qui concerne le bruit de Poisson. Distribué selon une distribution de Poisson, le nombre de photo-électrons enregistrés par la CCD sur un pixel donné a une valeur moyenne donnée par le profil d'intensité lumineuse moyen de l'étoile et une variance égale à cette moyenne. Le profil lumineux de l'étoile étant à haute dynamique dans le champ d'intérêt, la variance varie donc de même de plusieurs ordres de grandeur dans le champ d'intérêt. Cependant, au premier ordre on peut considérer le profil moyen de l'étoile comme étant à symétrie cylindrique. Dans ce cas, la variance du bruit ne varie plus qu'en fonction de la séparation entre le pixel considéré et le centre de l'image, plus en fonction de l'angle orthoradial. Or, durant son mouvement dû à la rotation de champ, une planète reste toujours à la même distance de l'étoile. Ainsi, si elle était de taille nulle la planète verrait toujours des pixels caractérisés par une même variance de bruit. Cependant, le profil de la planète est celui d'une tache de diffraction, donc sa taille est égale à la distance caractéristique de variation spatiale de l'intensité de l'étoile. Ainsi,  $\sigma^2$  varie donc légèrement d'une extrémité à l'autre (dans la direction radiale) de la zone couverte par la planète lors de son mouvement mais l'hypothèse de constance de  $\sigma^2$  est dans un premier temps admissible.

Tenant compte de ces considérations, il faut cependant alors préciser un peu la phase d'estimation de  $\sigma^2$  de la section 3.3. D'après l'expression (3.14) de  $\hat{\alpha}(r)$ ,  $\hat{\alpha}(r)$  n'est fonction que des données incluses spatialement dans la zone vue par la planète au moins une fois (voir l'Eq. 3.21). Une délimitation des données n'est alors pas nécessaire à ce stade puisqu'une telle délimitation existe automatiquement de façon stricte dès que la PSF est masquée. En revanche, l'estimation de  $\sigma^2$ , supposé constant dans l'espace, "gagne" en principe à utiliser toutes les données disponibles. A ce stade il est donc important d'établir un compromis entre la nécessité de "moyenner" pour estimer  $\sigma^2$  et le fait qu'en pratique  $\sigma^2$  varie dans l'espace et que l'utilisation de données éloignées de la planète n'aide donc pas à l'estimation du  $\sigma^2$  entachant l'estimation de  $\alpha$ . En pratique,  $\sigma^2$  peut ainsi par exemple être estimé en limitant les données à l'anneau situé à la même séparation que la planète et à peine plus large que la planète. On sélectionne alors simplement cette zone de la CCD, on redéfinit de nouveaux vecteurs  $\mathbf{x}'(k)$  etc en sélectionnant les composantes associées de  $\mathbf{x}(k)$  etc et on applique le même calcul sur ces vecteurs de taille  $L' < L$ .

### Si $r$ est connu, condition pour la consistance des estimateurs

Dans la section 3.3 nous avons démontré que si  $r$  est connu et si l'on suppose que le modèle gaussien est valide, l'estimateur ML est strictement non biaisé et efficace de telle sorte qu'il est l'estimateur de plus faible erreur quadratique moyenne parmi tous les autres estimateurs non biaisés.

En revanche, cette erreur ne tend vers 0 quand le nombre de poses  $K$  tend vers l'infini que lorsque la variance (Eq. 3.33) tend vers 0 :

$$\text{var}[\hat{\alpha}] = \frac{\sigma^2}{Kc_K} \rightarrow 0$$

Ceci définit la condition nécessaire et suffisante  $Kc_K \rightarrow \infty$ , condition géométrique sur le déplacement des planètes dans l'image au cours d'une nuit et au cours de plusieurs nuits. Nous avons dérivé dans la section 3.3 trois conditions suffisantes plus simples à appréhender que la condition brute  $Kc_K \rightarrow \infty$ .

Dans notre cas particulier, on peut se satisfaire de la première condition (Eq. 3.34) si la vitesse de convergence est suffisamment rapide pour qu'une nuit d'observation "suffise" à la précision que nous attendons de l'estimation.

Sinon, cette seule condition est insuffisante car des nuits d'observation supplémentaires sont caractérisées par un fond supposé identique à la première nuit –par hypothèse– et un mouvement de la planète identique à la première nuit. En termes mathématiques, chaque nuit  $n$  est caractérisée par un ensemble de profils de planète  $\{\mathbf{p}(k), k \in S_n\}$  avec  $\{\mathbf{p}(k), k \in S_n\} = \{\mathbf{p}(k), k \in S_{n'}\}$  pour n'importe quelle nuit  $n' \neq n$ . La planète repassant toujours par les mêmes zones, la condition de l'Eq. (3.34) n'est pas valide.

En revanche, la condition de l'Eq. (3.35) est bien valide car chaque nuit une fraction des profils est bien dissociée, et cette fraction ne dépend pas de la nuit considérée. Cette fraction dépend du nombre de poses par nuit et de la vitesse de rotation de champ mais très grossièrement on constate que la condition est déjà valide pour  $a = 1$  et  $\lambda = 2$  par exemple. On peut localiser une zone du champ couverte par la planète et compter la fraction de temps passée par la planète sur cette zone pour le constater.

Finalement, si  $r$  était connu et que le modèle gaussien modélisait bien nos données astrophysiques, d'après les considérations ci-dessus  $\hat{\alpha}$  convergerait bien en erreur quadratique moyenne vers la vraie valeur et serait en particulier consistante.

## 4.2 Programme informatique *user-friendly* : MOODS sur Matlab

La méthode d'estimation et de détection proposée, étudiée et discutée dans le chapitre I a été programmée en pratique pour le consortium SPHERE. L'algorithme informatique, appelé MOODS pour *MOVing Objects Detection for SPHERE*, est décrit succinctement dans l'annexe A. Il est décrit de façon beaucoup plus détaillé dans le rapport technique Vedrenne et al. [2008] qui inclut la description des méthodes ANDROMEDA (Mugnier et al. [2007]) et MOODS, les deux techniques de traitement du signal développées dans le cadre du consortium.

La version initiale de MOODS programmée en Matlab a été conçue pour être simple et intuitive d'utilisation tout en se plaçant aussi près que possible du cadre d'une utilisation de l'algorithme sur une situation réelle qui requiert donc un grand nombre de paramètres d'entrée. Les paramètres d'entrée demandés à l'utilisateur sont listés en annexe A. Comme le temps de calcul pour tester quasiment tous les pixels du champ d'intérêt ( $0'2 - 2''$ ) est extrêmement long, il a été nécessaire de faire appel à autant d'optimisations que possibles qui rendent la structure interne du code un peu complexe.

MOODS est un algorithme de détection d'exoplanètes à appliquer à partir d'une série d'images IRDIS. Dans un premier temps, il estime conjointement les paramètres de la planète (position, intensité et incertitude sur l'intensité) et la contribution de l'étoile. Dans un second temps il effectue une détection partiellement automatique, présentant des graphes et images à l'utilisateur et lui demandant –vus ces résultats– de sélectionner un paramètre ou une méthode pour poursuivre la détection. Il offre aussi la

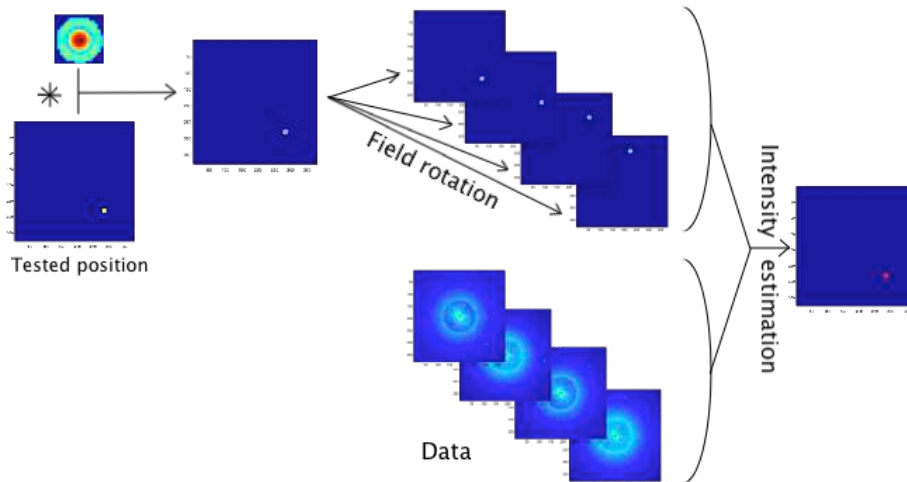


FIG. 4.2: Estimation de l’intensité d’une planète pour une position initiale  $r$  donnée, avec ici  $r \approx (280, 280)$ . Les *inputs* nécessaires sont :  $r$ , la PSF post-coronographe (dépendant éventuellement de  $r$  mais pas dans nos traitements), les coordonnées célestes de l’objet, le moment de la pose (date et heure) et les données  $\{\mathbf{x}(k)\}$ . L’*output* est l’estimation d’intensité  $\hat{\alpha}(r)$  telle que donnée par l’Eq. (3.12) qu’on localise simplement en  $r$  sur une carte de même taille que les images initiales et prise initialement égale à 0 partout.

possibilité d’utiliser la seconde bande spectrale comme référence où la planète est supposée être totalement inexistante. L’estimation conjointe du flux de la planète dans les deux bandes et le calcul du GLR décrits dans la section 3.5 ne sont pas décrits dans la version de MOODS présentée dans Vedrenne et al. [2008] car ils ont été développés et programmés par la suite.

La programmation de cette méthode mathématique est immédiate pour une position initiale  $r$  fixée. Pour estimer l’intensité qu’aurait une planète située initialement en un pixel  $r$  donné, on commence ainsi par simuler les profils  $\mathbf{p}(k, r)$  pour toutes les poses  $k$ , voir la section 4.1 et la figure 4.2. L’estimateur d’intensité  $\hat{\alpha}(r)$  donné par l’Eq. 3.12 est alors illustré sur la Figure 4.3.

Le problème survient lorsque l’on veut tester “toutes” (au sens d’un échantillonnage raisonnable proche de Shannon, *i.e.* proche de la résolution de l’instrument) les positions initiales malgré la taille des jeux de données. La programmation proposée repose sur le fait que la parallélisation n’est pas possible en Matlab et plus généralement difficile à programmer et implanter en pratique. A la place, pour simuler la rotation de champ qui est l’étape la plus longue de la phase d’estimation, il était essentiellement possible soit d’effectuer une rotation globale d’image, soit d’effectuer des translations subpixeliques de petites images sur une grande image. Il s’est avéré plus rapide d’adopter la méthode basée une rotation globale d’image et par des astuces associées, incluant la création de “boîtes”. La division du champ d’intérêt en boîtes (zones d’image définissant un ensemble d’indices utilisés au moment de la phase d’estimation finale) est illustrée sur la figure 4.4 et légèrement plus détaillée par Smith et al. [2008]. Les cartes d’estimation qui seront présentées figure 4.7 ont ainsi été obtenues en 42 à 45h avec un serveur HP Proliant DL585, avec les coefficients critiques  $dn\_rot = 100$  pixels (pas de simulation d’un haricot mais seulement d’un poi, différence d’estimation de l’ordre de quelques pourcents par rapport à  $dn\_rot = 0.5$ ),  $dtheta = 3$  pixels à  $2''$  et  $drho = 0.8$  pixels (paramètres définissant le nombre de grilles à tester et ainsi l’échantillonnage des positions initiales testées).

Notons cependant que la parallélisation du code est maintenant en partie développée sur le *cluster* de l’Observatoire de la Côte d’Azur par Mattei et al. [2010], ce qui devrait permettre notamment de pouvoir appliquer des méthodes d’inférences plus coûteuses en temps de calcul que celles présentées jusqu’ici.

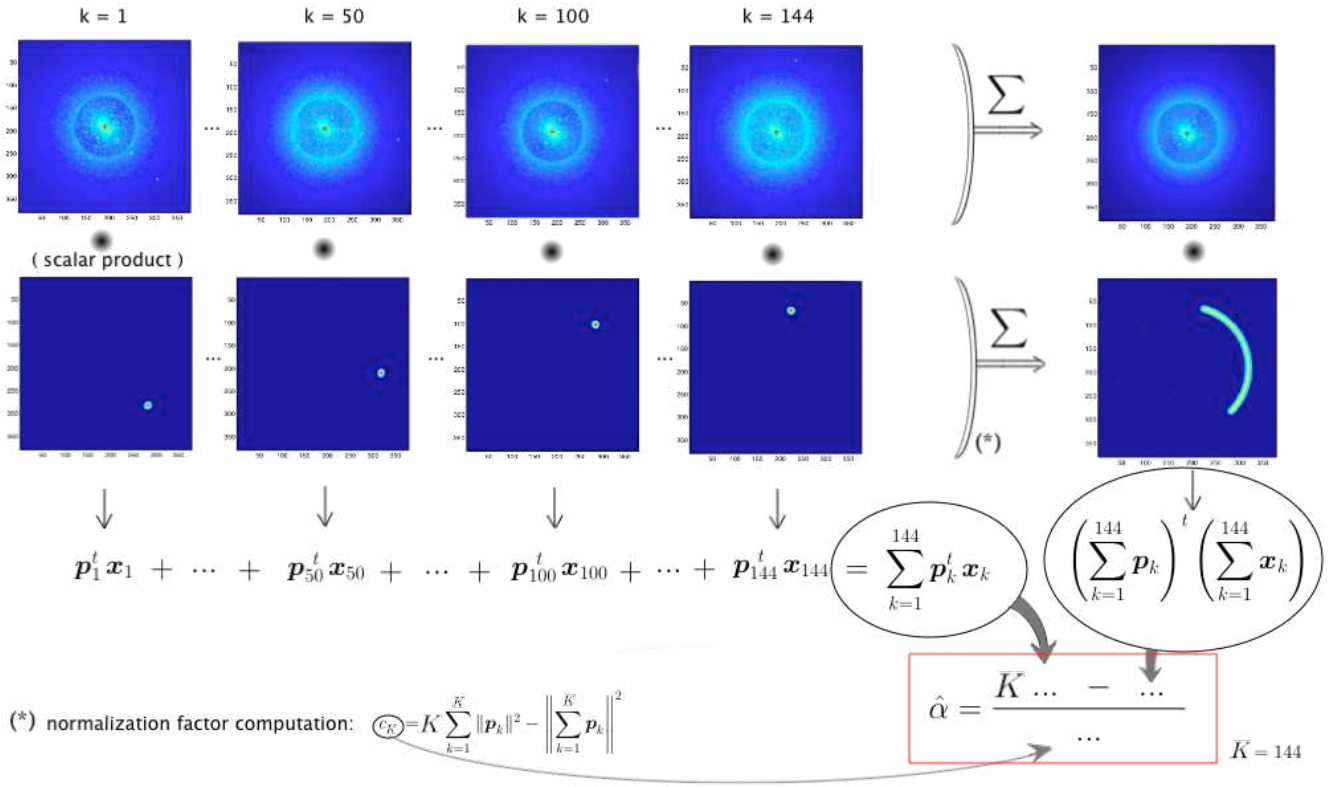


FIG. 4.3: Différents termes contribuant au calcul de l'estimation d'intensité  $\hat{\alpha}(r)$  donné par l'Eq. (3.12). Ici,  $r$  est implicite,  $\mathbf{x}_k$  désigne  $\mathbf{x}(k)$  et  $\mathbf{p}_k$  désigne  $\mathbf{p}(k, r)$ . Rappelons que le produit scalaire entre deux images (concaténées en des vecteurs pour simplifier les calculs mathématiques) consiste en pratique à multiplier les deux images pixel à pixel et à sommer ces  $L$  produits. Ainsi,  $\mathbf{p}(k, r)^t \mathbf{x}(k)$  est élevé s'il y a dans l'image  $\mathbf{x}(k)$  un pic d'intensité à l'endroit où le profil testé  $\mathbf{p}(k, r)$  est élevé.

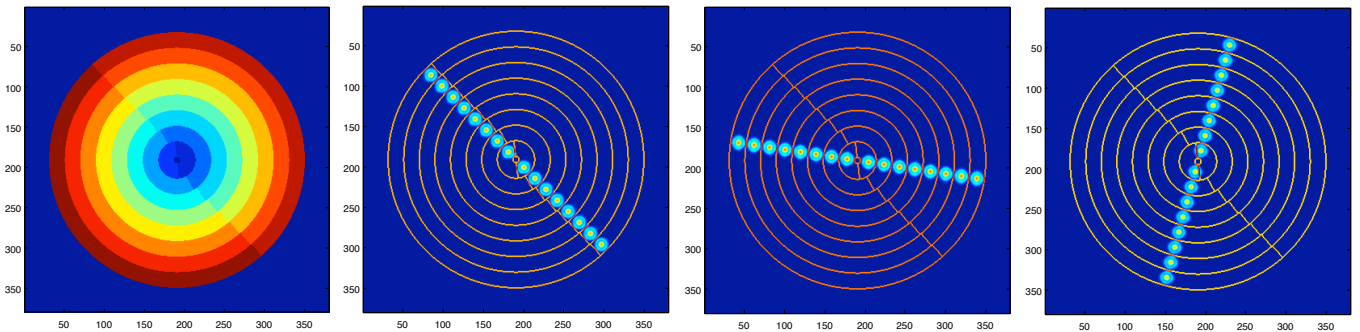


FIG. 4.4: Gauche : Grille constituée de 16 boîtes correspondant aux 16 positions initiales testées  $\{r_i\}_{i=1, \dots, 16}$  avec cette grille. Les autres figures présentent la projection dans la grille des haricots correspondants aux positions initiales testées ( $\mathbf{p}(k, r_i)$  élevé à la puissance 0.2). Les profils sont présentés pour  $k = 1$ ,  $k = 50$  et  $k = 144$ . Notons qu'à cause de la taille finie de la PSF, les planètes à l'intérieur remplissent davantage leurs boîtes au cours de l'observation que les planètes situées plus loin de l'étoile.

### 4.3 Application de la méthode à des données stationnaires et non-stationnaires

La stationnarité des données est une hypothèse forte qui constitue le point majeur de réflexion et d’ajustement de la plupart des techniques actuelles de traitement des données d’imagerie directe avec rotation de champ (Marois et al. [2006], Mugnier et al. [2007]). Les performances de l’algorithme MOODS sont comparées selon que les données sont simulées de façon stationnaire ou non-stationnaire.

Rappelons que la méthode mathématique sous-jacente est développée dans la section 3.2 en ce qui concerne l’estimation d’intensité et de position d’une hypothétique planète, dans la section 3.4 pour ce qui est de la procédure d’identification de candidats et dans la section 4.2 pour l’introduction du programme MOODS et l’illustration de la méthode d’estimation. Si l’on souhaite essayer de relier le seuil appliqué à  $\hat{\alpha}(r)$  à une PFA, il est naturel dans un premier temps d’exprimer le seuil en fonction de l’écart type de la statistique  $\hat{\alpha}(r)$ . Si la distribution de  $\hat{\alpha}(r)$  est gaussienne, pour un seuil  $\zeta(r) = z\sigma_{\hat{\alpha}(r)}$ , la PFA du test dépend seulement de  $z$ .

Notons cependant qu’ici l’écart type de  $\hat{\alpha}(r)$  n’est pas estimé comme indiqué dans la section 3.3 (complétée dans la section 4.1 pour le problème de variation spatiale de  $\sigma^2$ ) car le modèle gaussien décrit mal les données ; nous verrons que le modèle gaussien décrit en général mal la statistique  $\hat{\alpha}(r)$  elle-même malgré le fait qu’elle moyenne les images  $\mathbf{x}(k)$ .

À la place, un écart type empirique est calculé, ce qui correspond à la démarche standard notamment pour ce type de données. En l’occurrence,  $\hat{\sigma}_{\hat{\alpha}}(r)$  est calculé à partir de la carte des  $\hat{\alpha}(r)$  en tenant compte des considérations de la section 4.1. Pour une planète “détectée” en la position  $\hat{r}$ , on définit comme échantillon, supposé *i.i.d.* et décrit par la distribution de  $\hat{\alpha}(r)$  sous  $H_0$ , l’ensemble des  $\{\hat{\alpha}(r), r \in \mathcal{A}(\hat{r})\}$  où  $\mathcal{A}(\hat{r})$  est un anneau entourant  $\hat{r}$  mais dont les éléments sont suffisamment éloignés de  $\hat{r}$  pour que les  $\{\hat{\alpha}(r), r \in \mathcal{A}(\hat{r})\}$  soient bien décorrélés de  $\hat{\alpha}(\hat{r})$ .  $\hat{\sigma}_{\hat{\alpha}}(\hat{r})$  est ainsi calculé comme l’écart type empirique de  $\{\hat{\alpha}(r), r \in \mathcal{A}(\hat{r})\}$ .

À cause de la rotation de champ, la zone de corrélation des termes  $\hat{\alpha}(r)$  avec  $\hat{\alpha}(\hat{r})$  peut être grande. L’effet est facilement visualisable lorsque le rapport signal sur bruit est important, comme dans le cas stationnaire que nous verrons. Nous observerons qu’elle s’étend sur une fraction d’anneau dont la largeur est celle de la PSF et dont l’angle est égale au double de l’angle total de rotation de champ plus la taille angulaire de la PSF. En pratique, il peut donc ne rester qu’un faible échantillon  $\{\hat{\alpha}(r), r \in \mathcal{A}(\hat{r})\}$  qui peut être supposé décrit sous  $H_0$ .

L’effet déterministe de contamination d’une planète située en  $r_0$  sur la distribution d’une estimation  $\hat{\alpha}(r)$  en un autre pixel  $r$  peut être précisément quantifié lorsque  $r_0$ ,  $\alpha_0$  et la PSF sont connus. L’effet pourrait être pris en compte simplement par une méthode telle que CLEAN (implantée dans MOODS). Cependant, et quoiqu’il en soit, le modèle gaussien décrit en général mal la distribution empirique des  $\hat{\sigma}_{\hat{\alpha}}(r)$ , du moins pour les données non-stationnaires. Nous le constaterons en appliquant la méthode aux mêmes données mais en l’absence de planète.

En conclusion, pour les données simulées avec des planètes, les seuils seront donnés ici en fonction de  $\hat{\sigma}_{\hat{\alpha}}(r)$  uniquement à titre indicatif et ne sont pas reliés à une PFA calculée sous modèle gaussien puisque celui-ci décrit en général mal la distribution empirique des  $\hat{\sigma}_{\hat{\alpha}}(r)$ , comme nous le verrons avec le traitement des données sans planète.

#### Données stationnaires

- **Données simulées en présence de quatre planètes**

Dans un premier temps, afin de bien satisfaire l’hypothèse de stationnarité sous-jacente au modèle statistique des données présenté dans la section 3.2 et utilisé afin de construire la méthode d’inférence, nous appliquons la procédure à des données dites *stationnaires*. Toutes les images sont en effet caractérisées par un front d’onde incident sur la caméra identique pour les  $K = 144$  poses. Elles sont simulées en suivant le cas de simulation présenté dans la section 2.1, à ceci près qu’aucun paramètre physique (paramètres



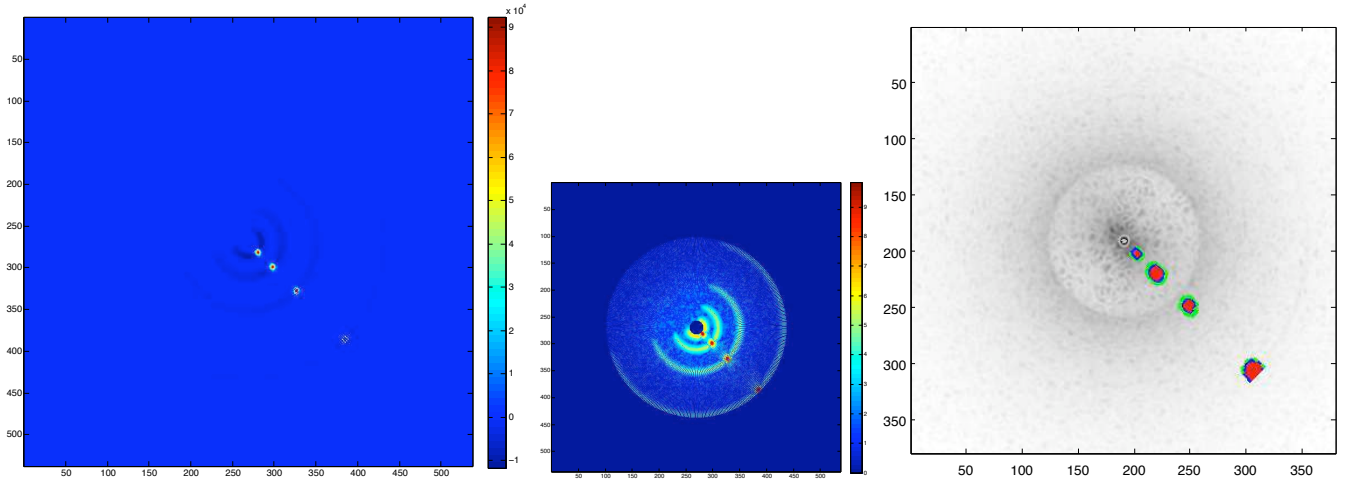


FIG. 4.5: Gauche : Carte des estimations d'intensité  $\hat{\alpha}(r)$  pour différentes positions initiales possibles  $r$  (en photo-électrons), pour le cas dit stationnaire. Centre : Carte identique, mais élevée à la puissance 0.2 afin de mettre en évidence les motifs de corrélation, et en particulier celui dû à la rotation de champ. Droite : Superposition de la somme des images  $\mathbf{x}(k)$  élevée à la puissance 0.2 et montrée en noir et blanc et de la carte de détection pour trois très grandes valeurs de seuil :  $\hat{\alpha}(r) > 50 \hat{\sigma}_{\hat{\alpha}}(r)$ ,  $\hat{\alpha}(r) > 200 \hat{\sigma}_{\hat{\alpha}}(r)$  et  $\hat{\alpha}(r) > 600 \hat{\sigma}_{\hat{\alpha}}(r)$ . Pour mieux visualiser la carte de détection, un opérateur de "dilatation" de quelques pixels de large est appliqué après le seuillage de  $\hat{\alpha}(r)$ .

atmosphériques et aberrations instrumentales) n'évolue avec le temps. Les  $K$  images correspondent alors simplement à différentes réalisations des bruits de détecteur (bruit de lecture, de courant d'obscurité, de *flat-field* et de bruit de Poisson) appliquées au champ d'intensité correspondant à la pose  $k = 1$  de la Figure 2.2.

L'image de gauche de la figure 4.5 présente la carte d'intensité des  $\hat{\alpha}(r)$  pour différentes positions initiales possibles  $r$ . L'image du centre présente la même carte mais élevée à la puissance 0.2. Dans l'image de droite, en arrière plan est indiquée la somme des images  $\mathbf{x}(k)$  et en premier plan la carte  $\hat{\alpha}(r)$  seuillée. Ainsi, chaque pixel coloré indique que  $\hat{\alpha}(r)$  est supérieur à une valeur correspondant à la couleur, pour trois valeurs de seuil et donc trois couleurs différentes. En l'occurrence, les valeurs de seuil, exprimées en fonction de  $\hat{\sigma}_{\hat{\alpha}}(r)$  décrit précédemment, sont très élevées, de sorte que décider que les zones effectivement supérieures au seuil abritent une planète constitue un test caractérisé par une très faible probabilité d'erreur empirique de détecter en l'absence de planète. Comme nous l'avons dit nous n'exprimons cependant pas la probabilité à partir d'une modélisation gaussienne de  $\hat{\alpha}(r)$  d'une part car celle-ci ne décrit pas toujours suffisamment bien la distribution de  $\hat{\alpha}(r)$  sous  $H_0$  pour en déduire une PFA crédible (voir paragraphe suivant où des données sans planète sont traitées) et d'autre part car le nombre d'estimations  $\hat{\alpha}(r)$  décorréliées sensées décrire  $H_0$  n'est ici pas vraiment suffisant pour la plupart des anneaux, comme on peut le voir sur la figure 4.5 du centre.

Le tableau 4.1 indique les valeurs maximales de  $\hat{\alpha}(r)$  par anneau et les valeurs vraies associées aux quatre planètes effectivement simulées. Les estimateurs d'intensité et de position sont très proches des vraies valeurs. Le tableau indique aussi le cardinal de  $\{\hat{\alpha}(r), r \in \mathcal{A}(\hat{r})\}$  et les valeurs de statistiques associées.

Finalement, comme les conditions de simulations vérifient assez bien les hypothèses de stationnarité et de gaussiannité utilisées pour construire l'estimateur d'intensité, les quatre planètes simulées sont correctement détectées même pour un seuil de  $600 \hat{\sigma}_{\hat{\alpha}}(r)$  et il se trouve que les premières fausses alarmes apparaissent pour un seuil de  $3 \hat{\sigma}_{\hat{\alpha}}(r)$ , et pour une séparation d'environ  $1''.5$ . De plus les erreurs des estimateurs d'intensité et de position sont faibles.

Anneau	0''12–0''37	0''37–0''61	0''61–0''85	0''85–1''1	1''1–1''3	1''3–1''6	1''6–1''8	1''8–2''
$\hat{\alpha}(r)$ max. [ $10^4 e^-$ ]	10.8	10.9	0.2	10.1	0.2	–	–	10.5
$\alpha$ simulé [ $10^4 e^-$ ]	10.3	10.3	–	10.3	–	–	–	10.3
$\hat{x}$ [px]	281	299	305	327	334	–	–	384
$\hat{y}$ [px]	280	298	304	327	332	–	–	384
$x = y$ simulé [px]	280	297	–	326	–	–	–	384
Séparation simulée	0''2	0''5	–	1''	–	–	–	2''
# de $\hat{\alpha}(r)$ sous $H_0$	760	1640	5610	3240	5210	7720	7630	2370
Moy. de $\hat{\alpha}(r)$ [ $e^-$ ]	-70	-14	-1	-4	-1	0	0	-1
$\hat{\sigma}_{\hat{\alpha}}$ [ $e^-$ ]	135	25	31	52	18	13	16	10

TAB. 4.1: Détection-estimation d'une planète par anneau à partir de la carte d'estimation d'intensité montrée sur la figure 4.5. “ $\hat{\alpha}(r)$  max” indique la valeur maximale de  $\hat{\alpha}(r)$  dans l'anneau complet si celle-ci est supérieure au seuil le plus bas de la figure 4.5, et la position associée est notée  $\hat{r} = (\hat{x}, \hat{y})$ . Les intensités et positions des quatre planètes simulées sont indiquées. Des statistiques des échantillons  $\{\hat{\alpha}(r), r \in \mathcal{A}(\hat{r})\}$  sont finalement indiquées pour chaque anneau, avec, comme décrit dans le texte,  $A(\hat{r})$  une fraction d'anneau excluant la zone autour de  $\hat{r}$ .

Une problématique importante consiste cependant à relier le seuil de détection à une PFA. Ceci nécessite de connaître la distribution de la statistique de test sous  $H_0$ . Pour cela on peut présupposer que la distribution appartient à une certaine famille (la famille gaussienne par exemple) et estimer ses paramètres à partir du jeu de données.

Or la planète simulée entache une grande fraction de l'anneau de sorte qu'en pratique il ne reste qu'un échantillon  $\{\hat{\alpha}(r), r \in \mathcal{A}(\hat{r})\}$  de petite taille pour décrire  $H_0$ . Cet effet peut éventuellement être partiellement pris en compte en pratique et n'est quoiqu'il en soit pas problématique pour un faible rapport signal sur bruit, mais ne doit pas être omis.

Ici, comme l'intensité de la planète est grande et que nous souhaitons éviter de compliquer la discussion, nous allons appliquer le même traitement des données au même jeu de données que précédemment, mais sans planète simulée, afin de disposer d'un échantillon plus grand de valeurs de  $\hat{\alpha}(r)$  sous  $H_0$  et d'observer si la PFA peut être estimée à partir d'une modélisation gaussienne de  $\hat{\alpha}(r)$ .

#### • Données simulées en l'absence de planète (cas $H_0$ )

Les traitements sont appliqués au même jeu de données que précédemment, mais sans planète simulée. Comme il n'y a plus de planète simulée, on s'intéresse uniquement à la distribution empirique de  $\hat{\alpha}(r)$  afin d'avoir une idée de la distribution sous-jacente de  $\hat{\alpha}(r)$  sous  $H_0$ .

En particulier, le tableau 4.2 donne des indicateurs relatifs à la gaussiannité de la distribution empirique de  $\hat{\alpha}(r)$  (sous  $H_0$ ). Il s'agit plus spécifiquement de voir si l'échantillon des  $\hat{\alpha}(r)$  par anneau peut s'accorder avec un modèle gaussien de moyenne et variance non prédéterminées. On aurait aussi pu fixer la moyenne à 0 mais le principal ici était de tester la forme de la distribution.

Les paramètres de *skewness* (coefficient de dissymétrie) et de *kurtosis* normalisé (coefficient d'aplatissement) valent 0 dans le cas gaussien. Dans les *p-values* suivantes, l'hypothèse nulle consiste à supposer que l'échantillon est issu d'une loi normale dont les moments sont les moments empiriques. La statistique de test, appelée  $\chi^2$  à cause de sa distribution dans le cas gaussien simple, est alors calculée à partir d'un histogramme constitué de 10 *bins* comme une moyenne d'écart entre la distribution observée et la distribution sous  $H_0$ . La *p-value* associée au test du  $\chi^2$  est alors égale à la probabilité d'observer une valeur de  $\chi^2$  plus extrême que celle effectivement observée. Typiquement, comme introduit dans la section 1.3, si  $p_{\text{val}} < 5\%$  on rejette  $H_0$ . La statistique de test de Kolmogorov-Smirnov est quant à elle calculée à partir de la distribution cumulative empirique de l'échantillon  $\hat{\alpha}(r)$  comme l'écart maximal entre la distribution

Anneau	0''12–0''37	0''37–0''61	0''61–0''85	0''85–1''1	1''1–1''3	1''3–1''6	1''6–1''8	1''8–2''
# de $\hat{\alpha}(r)$ sous $H_0$	2300	4950	6940	7630	7670	7720	7630	6770
Moy. de $\hat{\alpha}(r)$ [ $e^-$ ]	1	0	0	0	0	0	0	0
$\hat{\sigma}_{\hat{\alpha}}$ [ $e^-$ ]	38	22	17	20	15	11	10	9
Skewness	-0.05	-0.21	0.02	0.03	0.05	-0.10	-0.04	0.03
Kurtosis centré	-0.07	0.11	0.17	0.26	-0.10	0.21	0.15	- 0.01
$p_{\text{val}}$ du $\chi^2$ (%)	1.5	0.1	64.7	0.0	12.5	4.5	18.3	16.8
$p_{\text{val}}$ de KS (%)	34.7	0.1	9.4	1.2	24.1	3.7	6.0	43.2
$\hat{\alpha}(r)$ gaussien ?	?	non	oui	non	oui	$\sim$ non	oui	oui

TAB. 4.2: Indicateurs de gaussiannité de l'échantillon  $\{\hat{\alpha}(r), r \in \mathcal{A}\}$  pour chaque anneau  $\mathcal{A}$ . La carte des  $\hat{\alpha}(r)$  a ici été obtenue à partir du même jeu de données que celui ayant mené à la carte d'estimation de la figure 4.5, mais cette fois sans planète, afin d'étudier la statistique empirique de  $\hat{\alpha}(r)$  sous  $H_0$ . La réponse binaire quant à la gaussiannité de  $\hat{\alpha}(r)$  est "oui" si les  $p_{\text{val}}$  des tests du  $\chi^2$  et de Kolmogorov-Smirnov sont toutes deux supérieures à 5% et "non" si elles sont toutes deux inférieures.

sous  $H_0$  et la distribution observée. La  $p$ -value est alors calculée à partir d'une probabilité asymptotique (quand la taille de l'échantillon tend vers l'infini) de cet écart maximal. A nouveau, typiquement si  $p_{\text{val}} < 5\%$  on rejette  $H_0$ .

Les indicateurs du tableau 4.2 tendent plutôt à accepter l'hypothèse  $H_0$ , *i.e.* à considérer que les échantillons sont distribués selon une loi gaussienne, dans les anneaux où le fond est de faible intensité, et au contraire à rejeter  $H_0$  dans les anneaux où le fond est de forte intensité. Ceci est naturel puisque le bruit Poissonien contribue essentiellement pour les pixels de forte intensité.

## Données non-stationnaires

Les données dites *non-stationnaires* correspondent au cas décrit dans la section 2.1, c'est-à-dire le cas utilisé dans le reste de la thèse et illustré sur la figure 2.2. On applique la même procédure que dans le cas des données stationnaires décrit précédemment.

### • Données simulées en présence de quatre planètes

Les cartes d'estimation et de détection sont montrées sur la figure 4.6. Le tableau 4.3 indique les mêmes statistiques de  $\hat{\alpha}(r)$  que dans le cas stationnaire.

On peut noter que les estimations d'intensité  $\hat{\alpha}(r)$  au niveau des planètes simulées sont quasiment identiques au cas stationnaire pour les séparations de 1'' et 2'' mais diminuées de respectivement 30% et 5% pour les séparations de 0''2 et 0''5.

### • Données simulées en l'absence de planète (cas $H_0$ )

Comme pour le cas stationnaire, les traitements sont appliqués au même jeu de données que dans le cas non-stationnaire précédent mais sans planète simulée. On s'intéresse uniquement à la distribution empirique de  $\hat{\alpha}(r)$  afin d'avoir une idée de la distribution sous-jacente de  $\hat{\alpha}(r)$  sous  $H_0$ . En particulier, le tableau 4.4 donne des indicateurs relatifs à la gaussiannité de la distribution empirique de  $\hat{\alpha}(r)$  (sous  $H_0$ ).

L'hypothèse de gaussiannité ne semble plus adéquate pour modéliser  $\hat{\alpha}(r)$  sous  $H_0$ . Cependant, il faut noter la grande corrélation qui existe sur la carte des  $\hat{\alpha}(r)$ , de sorte que la non-stationnarité des données a rendu l'hypothèse d'échantillon *i.i.d.* non vérifiée, en particulier près de l'axe optique.

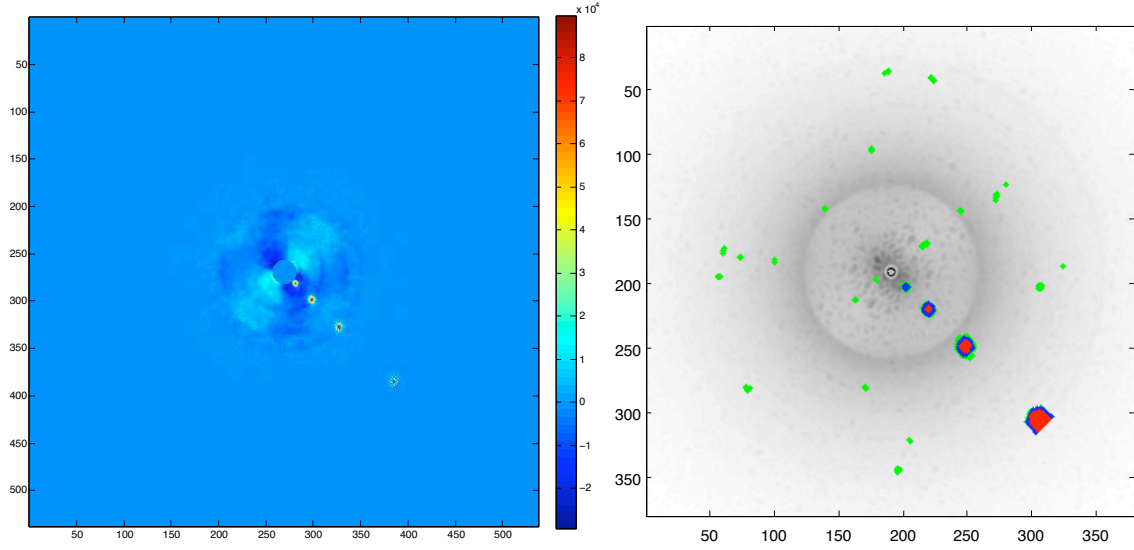


FIG. 4.6: Gauche : Carte des estimations d'intensité  $\hat{\alpha}(r)$  pour différentes positions initiales possibles  $r$  (en photo-électrons), pour le cas dit non-stationnaire, correspondant aux données illustrées dans la figure 2.2. Droite : Superposition de la somme des images  $\mathbf{x}(k)$  élevée à la puissance 0.2 et montrée en noir et blanc et de la carte de détection associées aux seuils  $\hat{\alpha}(r) > 3 \hat{\sigma}_{\hat{\alpha}}(r)$ ,  $\hat{\alpha}(r) > 6 \hat{\sigma}_{\hat{\alpha}}(r)$  et  $\hat{\alpha}(r) > 30 \hat{\sigma}_{\hat{\alpha}}(r)$ . Le même effet de dilatation est utilisé que précédemment.

Anneau	0''12–0''37	0''37–0''61	0''61–0''85	0''85–1''1	1''1–1''3	1''3–1''6	1''6–1''8	1''8–2''
$\hat{\alpha}(r)$ max. [ $10^4 e^-$ ]	8.2	10.3	0.8	10.0	0.2	–	–	10.5
$\alpha$ simulé [ $10^4 e^-$ ]	10.3	10.3	–	10.3	–	–	–	10.3
$\hat{x}$ [px]	282	299	302	327	335	–	–	384
$\hat{y}$ [px]	281	298	224	327	331	–	–	384
$x = y$ simulé [px]	280	297	–	326	–	–	–	384
Séparation simulée	0''2	0''5	–	1''	–	–	–	2''
# de $\hat{\alpha}(r)$ sous $H_0$	760	1640	5610	3240	5210	7720	7630	2370
Moy. de $\hat{\alpha}(r)$ [ $e^-$ ]	-6995	-2782	-450	51	-13	-2	0	15
$\hat{\sigma}_{\hat{\alpha}}$ [ $e^-$ ]	7910	2740	3440	959	411	235	139	100

TAB. 4.3: Détection-estimation d'une planète par anneau à partir de la carte d'estimation d'intensité montrée sur la figure 4.6. La signification des grandeurs indiquées est précisée dans le tableau 4.1.

Anneau	0''12–0''37	0''37–0''61	0''61–0''85	0''85–1''1	1''1–1''3	1''3–1''6	1''6–1''8	1''8–2''
# de $\hat{\alpha}(r)$ sous $H_0$	2300	4950	6940	7630	7670	7720	7630	6770
Moy. de $\hat{\alpha}(r)$ [ $e^-$ ]	76	3	-142	-16	-3	-2	0	1
$\hat{\sigma}_{\hat{\alpha}}$ [ $e^-$ ]	10590	3970	3170	1170	440	250	150	100
Skewness	-0.08	0.00	0.36	0.35	-0.06	-0.20	0.00	-0.03
Kurtosis centré	-0.77	-0.99	-0.83	2.69	-0.15	0.14	-0.09	0.17
$p_{\text{val}}$ du $\chi^2$ (%)	0.0	0.0	0.0	0.0	0.3	0.0	2.8	0.2
$p_{\text{val}}$ de KS (%)	0.0	0.0	0.0	0.0	11.1	0.0	12.0	12.9
$\hat{\alpha}(r)$ gaussien ?	non	non	non	non	?	non	?	?

TAB. 4.4: Indicateurs de gaussiannité de l'échantillon  $\{\hat{\alpha}(r), r \in \mathcal{A}\}$  pour chaque anneau  $\mathcal{A}$ . La carte des  $\hat{\alpha}(r)$  a ici été obtenue à partir du même jeu de données que celui ayant mené à la carte d'estimation de la figure 4.6, mais cette fois sans planète, afin d'étudier la statistique empirique de  $\hat{\alpha}(r)$  sous  $H_0$ . La signification des grandeurs indiquées est précisée dans le tableau 4.2 et le texte associé.

## 4.4 Autres applications de la méthode

### Application de la méthode sur des données de contrastes différents

A partir d'ici, les données simulées SPHERE seront systématiquement simulées selon le cas non-stationnaire présenté dans la section 2.1. On indique ci-dessous les cartes de  $\hat{\alpha}(r)$  obtenues pour des contrastes entre la planète et l'étoile variant de  $10^{-3}$  (cas le plus favorable) à  $10^{-6}$  (cas le plus défavorable, visé pour l'instrument SPHERE). Le contraste de  $10^{-4}$  correspond au cas trivial présenté précédemment à l'exception des paramètres photométriques (intensités de l'étoile et de la planète). Les planètes sont totalement invisibles pour un contraste de  $10^{-6}$ . La méthode appliquée jusqu'ici doit donc être notablement améliorée.

### Utilisation de deux bandes spectrales

La méthode de détection-estimation mono-canal a été étendue à l'utilisation de deux canaux spectraux dans la section 3.5 afin de tirer profit des caractéristiques du mode DBI de l'instrument IRDIS de SPHERE. Cette extension a été testée avec et sans *rescaling* sur quelques jeux de simulations physiques *end-to-end*. La partie diffractive de ces simulations de données (propagation de la lumière jusqu'à la CCD qui ne dépend pas de l'intensité lumineuse de l'étoile et la planète) a été effectuée grâce au *package* SPHERE (Carbillet et al. [2008]) du *problem-solving environment* CAOS (Carbillet et al. [2010]) tel que présenté dans la section 2.1. Les données ont alors été simulées (Vigan et al. [2010b], Vigan et al. [2010a]) en utilisant des valeurs photométriques d'intensité (de l'étoile et) de la planète dans chaque bande issues de modèles physiques d'atmosphères planétaires (Allard et al. [2007]).

La figure 4.8 présente des points de comparaison entre les méthodes d'estimation SBI (*Single Band Imaging*) et DBI. Sur les positions initiales des planètes simulées on applique pour la méthode SBI l'estimateur d'intensité de l'Eq. (3.12) et pour la méthode DBI l'estimateur d'intensité de l'Eq. (3.48), avec la correction photométrique nécessaire présentée dans la section 3.2.

On observe sur ces deux exemples de simulation que la méthode DBI apporte dans certains cas une amélioration par rapport à la méthode SBI, mais pas systématiquement ou pas de façon extrêmement significative. L'hypothèse d'égalité stricte entre les contributions de l'étoile dans les deux filtres est certainement trop forte. Il est possible d'assouplir cette contrainte en ajoutant des paramètres d'ajustement supplémentaires tels qu'un coefficient multiplicatif constant devant la contribution de l'étoile dans un des deux filtres. Nous choisirons plutôt dans la partie IV d'améliorer de façon globale la méthode en ajoutant notamment une corrélation (et pas une corrélation de 1 correspondant à une égalité stricte des fonds)

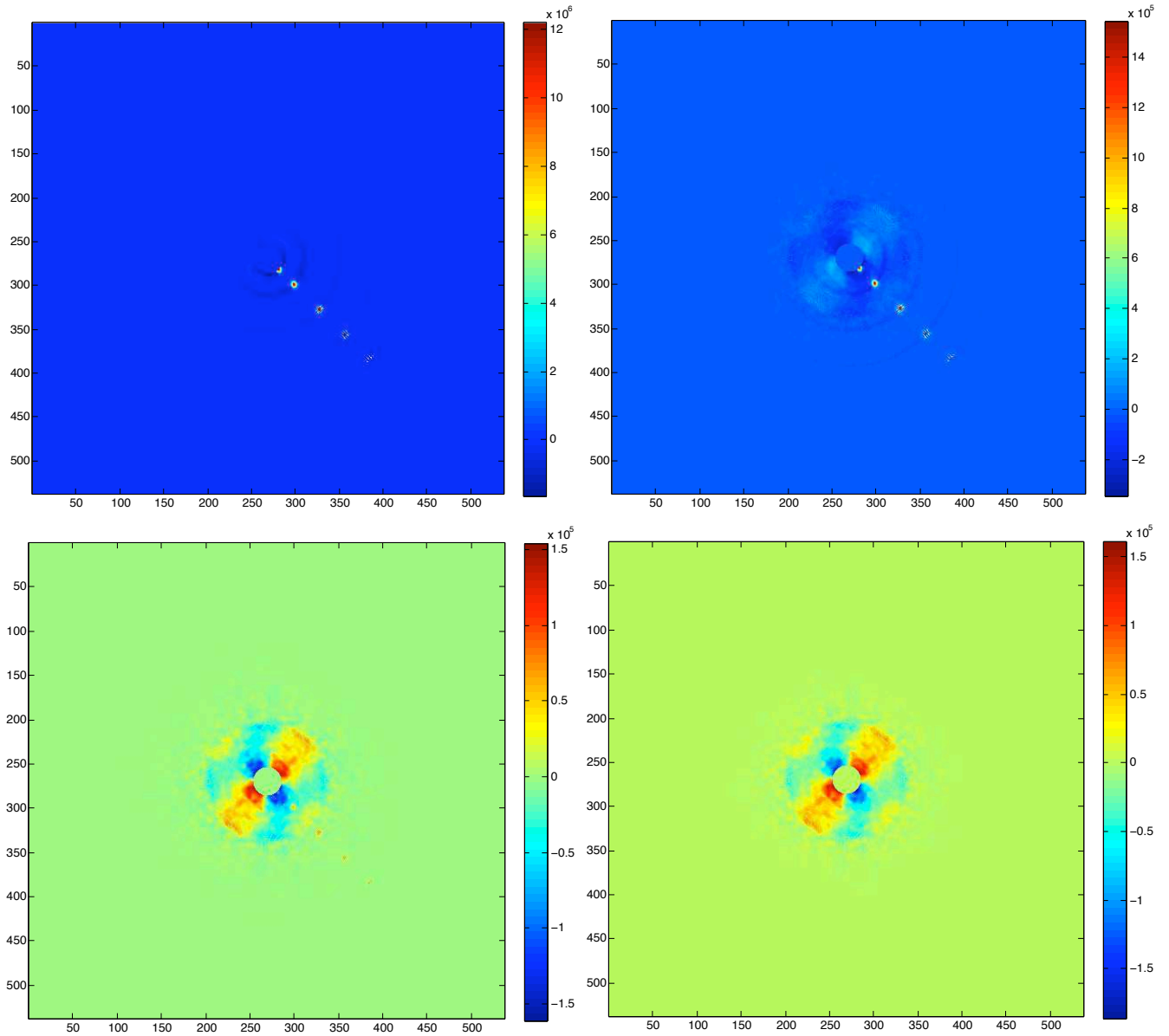


FIG. 4.7: Cartes d'estimation d'intensité de la planète  $\hat{\alpha}(r)$  ( $\hat{\alpha}(r)$  donné par l'Eq. (3.12)) pour toutes les positions initiales  $r$  possibles. Dans le sens de la lecture, contrastes respectifs entre la planète et l'étoile de  $10^{-3}$ ,  $10^{-4}$ ,  $10^{-5}$  et  $10^{-6}$ . Les cinq planètes simulées entre  $0''.2$  et  $2''$  sont clairement visibles pour les deux contrastes les plus favorables mais totalement invisibles pour le contraste le plus défavorable, pourtant visé par le consortium SPHERE.

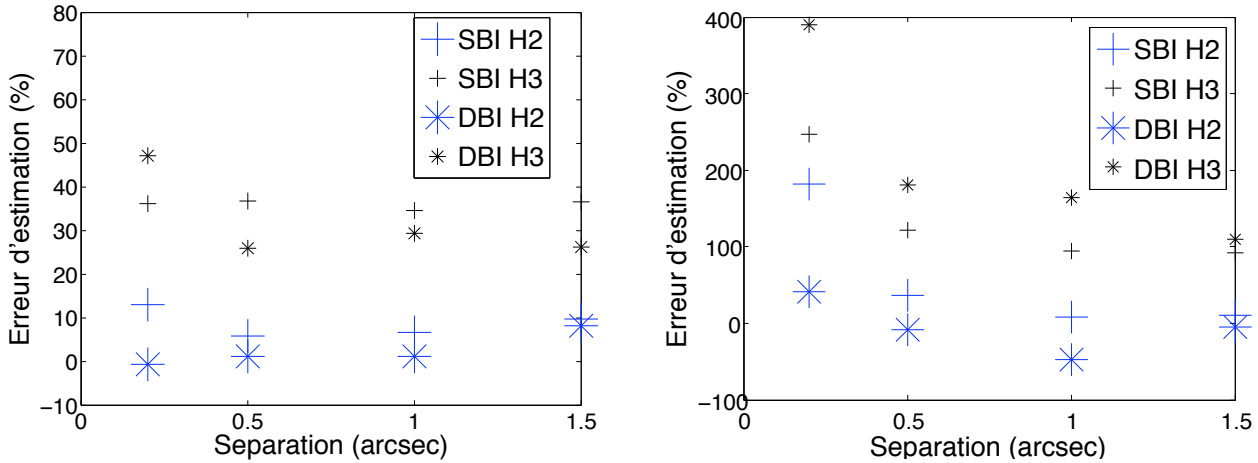


FIG. 4.8: Comparaison de la méthode d'estimation d'intensité de la planète utilisant séparément les images dans chacun des deux filtres (méthode SBI, section 3.2) et de la méthode utilisant conjointement les images dans chacun des deux filtres (méthode DBI, avec *rescaling*, section 3.5). Application de ces méthodes sur les simulations physiques *end-to-end* dédiées à SPHERE. A gauche : contraste de  $10^4$  entre l'étoile et la planète et flux quasiment identique de la planète dans les deux bandes (rapport d'intensité : 1.2). A droite : contraste de  $10^5$  entre l'étoile et la planète dans le filtre H2 et intensité de la planète 8.7 fois plus intense dans la bande H2 que dans la bande H3.

entre les deux filtres. Le test du GLR de l'Eq. (3.53) utilisant des images acquises dans deux canaux spectraux sera alors appliqué dans la section 12.1 à ces mêmes données simulées pour y être comparé à des détecteurs bayésiens.

### Effet de la direction du vent

La figure 4.9 présente la carte d'estimation  $\{\hat{\alpha}(r), r\}$  obtenue à partir d'un jeu de données simulé de la même façon que le jeu de données standard décrit dans le chapitre 2 mais pour une direction de vent horizontale et non plus diagonale<sup>1</sup>. Le jeu de données brut est illustré en annexe C.1, sur la dernière figure (figure C.1). L'objet de cette figure est uniquement de montrer qualitativement l'effet que peut avoir le vent sur la carte d'estimation  $\hat{\alpha}(r)$ .

### Comparaison de MOODS aux autres techniques standards

Les avantages fondamentaux des méthodes *différentielles* (*e.g.* Marois et al. [2006]) et de la méthode *conjointe* MOODS sont les suivants :

- Par construction les méthodes *différentielles* utilisent des paires d'images supposées avoir des fonds très similaires. Le fait que le fond évolue dans le temps est ainsi directement pris en considération de façon plus ou moins automatique, contrairement à MOODS où par défaut l'ensemble de la nuit ne définit qu'un ensemble de données  $\{\mathbf{x}(k), k \in \llbracket 1, K \rrbracket\}$  supposées avoir toute le même fond.
- La méthode *conjointe* MOODS estime conjointement le fond et la contribution de la planète dans les différentes bandes, de sorte que la planète peut partiellement se superposer dans son mouvement et de sorte que le flux de la planète peut être estimé dans les deux bandes spectrales, alors qu'une méthode différentielle estime uniquement une différence de flux.

Des extensions de ces deux méthodes devraient leur permettre de bénéficier chacune de tout ou partie des avantages de l'autre.

<sup>1</sup>On indique la direction dans le référentiel de la CCD.

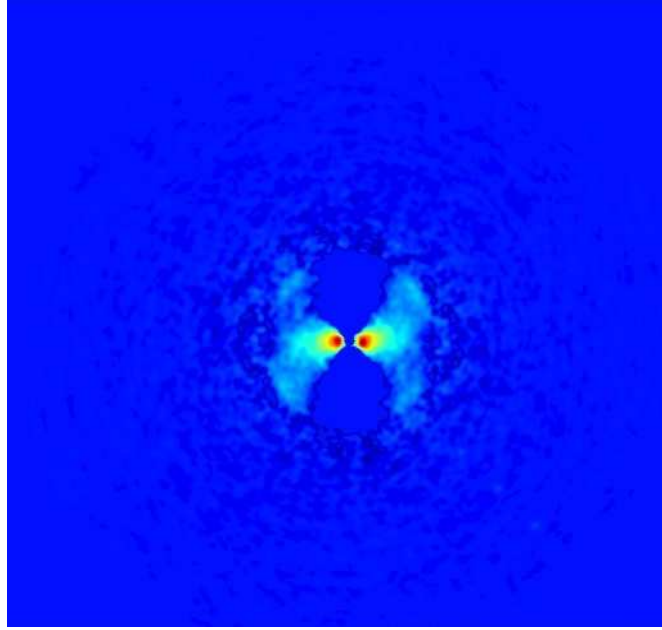


FIG. 4.9: Cartes d'estimation d'intensité de la planète  $\hat{\alpha}(r)$  ( $\hat{\alpha}(r)$  donné par l'Eq. (3.12)) pour toutes les positions initiales  $r$  possibles obtenue à partir d'un jeu de données pour lequel le vent a été simulé avec une direction horizontale.

Initialement, les méthodes différentielles considèrent uniquement des paires d'images comme ayant le même fond alors qu'il est plus performant de considérer que davantage d'images ont le même fond si tel est le cas (voir la section 3.3). Depuis, de nouvelles méthodes, toujours empiriques, permettent d'utiliser cette hypothèse implicite. Cependant, il ne leur est *a priori* pas encore possible d'estimer le flux de la planète dans l'autre bande. Rappelons que ce n'est pas forcément un problème pour SPHERE car par construction de l'instrument les paires de filtres du mode DBI sont choisies de telle sorte que pour les planètes d'intérêt, (très) peu de flux devrait être présent dans la seconde bande.

Une extension naturelle de la méthode MOODS consiste à l'appliquer à des sous-ensembles des données  $\{\mathbf{x}(k), k \in \llbracket 1, K \rrbracket\}$  pour lesquels les données sont à peu près stationnaires. Mathématiquement, un modèle reposant sur l'existence d'instantants pour lesquels le fond est identique s'écrit très simplement :

$$\mathbf{x}(k) = \mathbf{d}(j(k)) + \alpha \mathbf{p}(k, r) + \boldsymbol{\epsilon}(k) \quad k \in \llbracket 1, K \rrbracket, \quad j(k) \in \llbracket 1, J \rrbracket \quad (4.1)$$

où  $J \leq K$  désigne le nombre d'ensembles de poses supposées avoir le même fond. Mais une méthode basée sur un tel modèle se programme plus difficilement, même en supposant la fonction  $j(\cdot)$  connue pour éviter le problème combinatoire, si l'on souhaite éviter des temps de calcul démesurés.

Le problème rencontré en général pour toutes ces extensions plus ou moins empiriques de méthodes est la lourdeur de la programmation et/ou du calcul numérique.

Concrètement, comme on peut le voir dans la figure 4 de Vigan et al. [2010b] reproduite ici sur la figure 4.10 (sachant que la figure 4.8 traite les simulations correspondant aux contrastes de  $10^4$  et  $10^5$ ), les performances d'estimation de MOODS et des méthodes ADI et SDI+ADI sont assez similaires pour une utilisation typique de ces méthodes. Globalement, nous considérons qu'il n'y a pas de différence majeure entre toutes ces méthodes et que la différence de leur performance tiendra au temps accordé à leur ajustement, éventuellement au cas par cas, pour tenir compte de la variation de la contribution de l'étoile en fonction du temps.



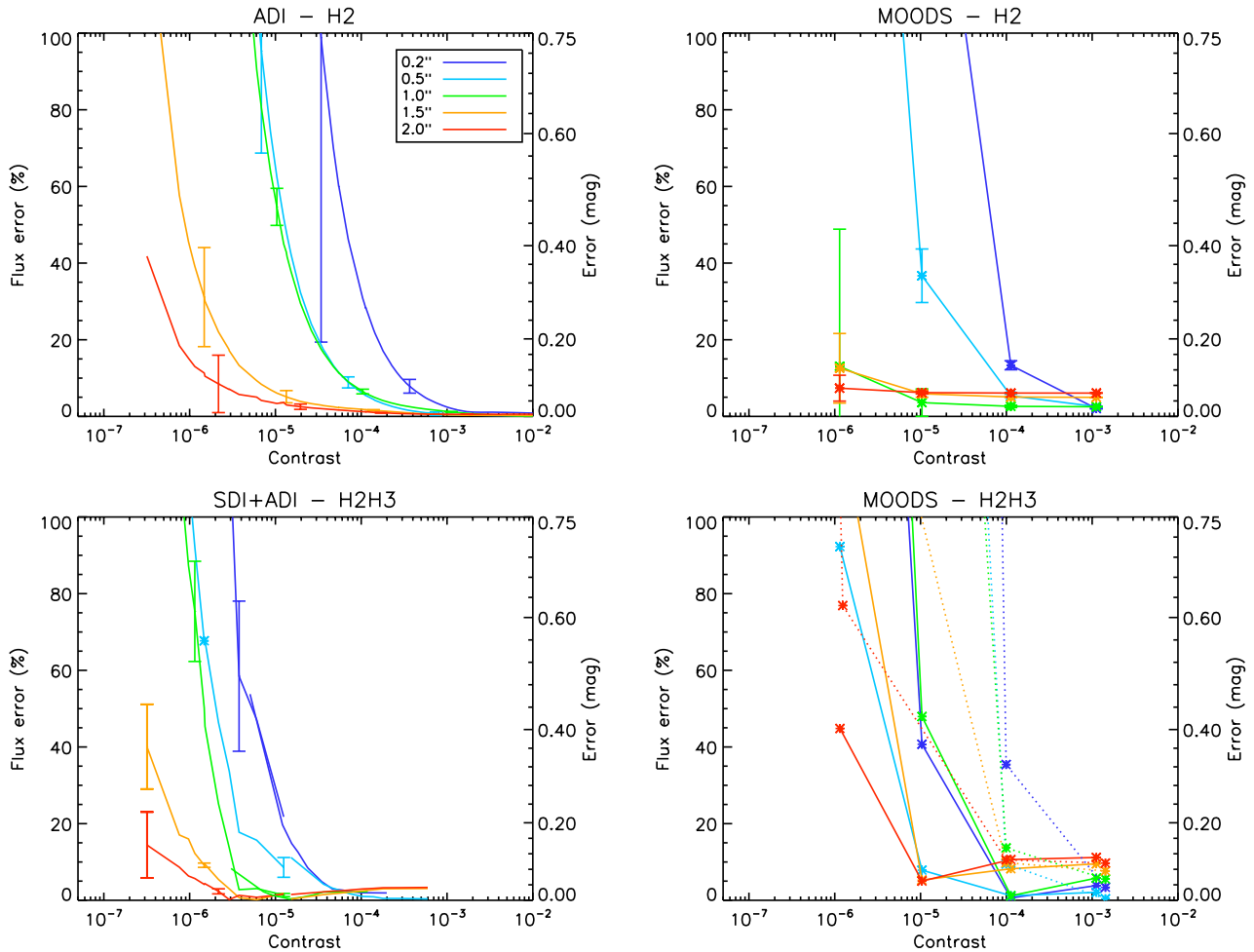


FIG. 4.10: Comparaison des erreurs d'estimation faites par la famille de méthodes ADI / ADI+SDI (gauche) et la famille de méthodes MOODS (droite). En haut, une seule bande est utilisée. En bas, les deux bandes sont utilisées conjointement. Les deux familles de méthode ont été appliquées sur un même ensemble de jeux de données sauf que les méthodes ADI / ADI+SDI ont traité beaucoup plus de cas de simulations (beaucoup plus de contrastes entre planète et étoile). La méthode ADI+SDI estime une différence de flux donc la figure en bas à gauche indique l'erreur de différence de flux entre les deux bandes, tandis que la figure en bas à droite indique en traits pleins l'erreur d'estimation de flux dans la bande où la planète a un fort flux et en pointillés l'erreur d'estimation de flux dans l'autre bande. Les méthodes ADI / ADI+SDI ont été appliquées sur 3 planètes à chaque fois (planètes simulées en trois positions orthoradiales différentes pour un jeu de données) de sorte que les barres d'erreur indiquent les 3 erreurs obtenues. Les barres d'erreur de MOODS indiquent l'estimation faite par MOODS de l'écart type de l'estimation d'intensité, comme développé dans la section 3.3 et 4.1. L'erreur n'a pas été estimée pour la méthode utilisant deux bandes spectrales. Figure issue de Vigan et al. [2010b].

## Deuxième partie

Modèle fin d'imagerie *speckle* pour  
quantifier les inférences précédentes



# Table des Matières

---

<b>5</b>	<b>Modèle fin des données et distribution statistique de l'estimateur linéaire d'intensité</b>	<b>95</b>
5.1	Description physique qualitative et modèle statistique des données (état de l'art) . . . . .	95
	Description physique qualitative . . . . .	95
	Modèle physique statistique des données (MGF) . . . . .	97
	Modèle statistique d'un jeu d'images en présence d'une exoplanète . . . . .	101
5.2	Distribution statistique fine de l'estimateur linéaire d'intensité de la section 3.2 . . . . .	102
	Motivation pour l'étude de l'estimateur linéaire d'intensité de la section 3.2 . . . . .	102
	Calcul de la MGF de $\hat{\alpha}(\mathbf{x})$ . . . . .	103
5.3	Consistance de l'estimateur d'intensité . . . . .	104
<b>6</b>	<b>Probabilité de Fausse Alarme du détecteur de la section 3.4 sous modèle <i>speckle</i></b>	<b>105</b>
6.1	Approximations analytiques de la PFA . . . . .	105
	Procédure de détection habituelle . . . . .	105
	Développements en série d'Edgeworth et de Gram-Charlier . . . . .	105
	Approximation du point selle ( <i>saddlepoint</i> ) . . . . .	106
6.2	Comparaison des méthodes d'approximation par simulation pour le modèle <i>speckle</i> . . . . .	107
	Cas de simulation . . . . .	107
	Evaluation des estimations de PFA et PD . . . . .	108
	Conclusion . . . . .	112
6.3	Utilisation du modèle <i>speckle</i> en pratique . . . . .	112
	Motivations pour ce modèle . . . . .	112
	Discussions de l'applicabilité du calcul de PFA à l'instrument SPHERE . . . . .	114

---



## Chapitre 5

# Modèle fin des données et distribution statistique de l'estimateur linéaire d'intensité

### 5.1 Description physique qualitative et modèle statistique des données (état de l'art)

#### Description physique qualitative

Le modèle gaussien développé dans la partie précédente était nécessaire pour appliquer un premier traitement rapide des données, extraire une liste de potentielles exoplanètes et dériver une statistique linéaire à étudier sous un modèle plus complexe. Une description plus fine de l'imagerie haute résolution au sol indique que la distribution gaussienne n'est pas appropriée, hormis éventuellement à très fort flux comme nous le verrons par la suite.

L'imagerie au sol est caractérisée par la présence de *speckles* turbulents (voir Goodman [2007] comme référence principale d'optique statistique) sur le plan focal de l'instrument. Ceux-ci sont dûs aux turbulences locales de l'atmosphère terrestre et à l'existence d'un orifice limitant, appelé diaphragme, au niveau de l'instrument de mesure. Les turbulences de l'atmosphère sont matérialisées par des fluctuations de température et de pression qui induisent directement des fluctuations spatiales et temporelles de l'indice de réfraction. Le trajet géométrique de la lumière émanant des objets célestes n'est donc pas totalement rectiligne, ce qui explique que nous percevions en général à l'oeil les étoiles comme des points scintillants : l'orifice limitant de l'oeil, la pupille, est de petite taille par rapport à la taille caractéristique des turbulences, donc schématiquement le fait que de la lumière entre ou non dans la pupille varie dans le temps.

En termes de propagation d'onde, les turbulences transforment le front d'onde plan émanant d'une étoile (source ponctuelle située infiniment loin de l'observateur) en un front d'onde perturbé. Rappelons que le front d'onde est défini comme la surface sur laquelle le champ électromagnétique a une phase constante, c'est-à-dire la surface à laquelle appartiennent "tous les photons qui ont été émis en même temps par l'étoile". La longueur caractéristique de cohérence spatiale définit une taille de zones localement planes du front d'onde à l'entrée d'un télescope (plan pupille). Chacune de ces zones "planes" constitue une source ponctuelle légèrement désaxée par rapport à l'étoile, dont la réponse par le télescope est schématiquement donnée par une réponse impulsionnelle hors axe (que nous appellerons PSF) légèrement décalée de la position géométrique de l'étoile. Pour un télescope sans occultation centrale et dont le miroir primaire –limitant, de diamètre  $D$ – est circulaire, ces petites taches ont ainsi typiquement le profil d'une tache d'Airy. Leur taille caractéristique à la longueur d'onde  $\lambda$  est ainsi  $\lambda/D$ , en taille angulaire car projetée sur le ciel.

En pointant un télescope modeste vers une étoile, ces taches lumineuses ne sont pas visibles sur des images *webcam* ou CCD car elles sont temporellement moyennées en un halo plus large autour de la position vraie de l'étoile<sup>1</sup> et car les pixels de l'instrument d'acquisition sont trop larges pour échantillonner ces taches lumineuses et donc distinguer ces dernières les unes des autres. En revanche, on peut imager ces *speckles* si l'on place sur le plan focal d'un télescope d'assez grand diamètre ( $D \sim 1m$  typiquement) et de grande focale ( $F \sim 20m$ ), un filtre très serré spectralement autour de  $\lambda \sim 1\mu m$  et une caméra dont la fréquence d'acquisition est suffisamment rapide ( $50ms$ , inférieur au temps caractéristique typique de corrélation des turbulences atmosphériques) et dont les pixels sont suffisamment petits (critère d'échantillonnage de Shannon : taille inférieure à  $\lambda/(2D)$  sachant que la taille angulaire d'un pixel est donné par  $p/F$  où  $p$  est la taille du pixel). Ce principe est à la base de la technique d'interférométrie des tavelures (*speckles* en anglais) développé par Labeyrie [1970].

Pour avoir des images de résolution plus élevée (*i.e.* meilleure) que celle induite par l'atmosphère de la Terre, il est possible d'appliquer des traitements numériques sur les mesures (opération dite de déconvolution lorsque la réponse impulsionnelle de l'atmosphère et instrument optique est uniforme dans le champ vu par la caméra), mais ces techniques sont insuffisantes si l'on souhaite obtenir des images de très haute résolution angulaire comme dans le cas de la recherche d'exoplanète. Il est alors nécessaire de corriger physiquement le front d'onde avant acquisition de l'intensité lumineuse par une caméra.

L'optique adaptative (AO) a précisément pour objet de corriger l'effet des turbulences atmosphériques sur la propagation de la lumière émanant des objets célestes. Un analyseur de surface d'onde récupérant une partie du flux mesure finement la forme du front d'onde au niveau du plan pupille (ex : analyseur de Shack-Hartmann fait d'une matrice de microlentilles, chaque microlentille permettant d'observer l'inclinaison d'une petite zone du front d'onde). Ces informations sont transmises à un miroir déformable qui compense alors en temps réel les retards ou avances locaux du front d'onde. Lorsqu'il parvient aux instruments finaux de mesure (coronographe, spectroscopie, caméra ...), le front d'onde est donc plan, à des erreurs inévitables de correction près.

Avec une correction parfaite par AO (comme si le télescope était placé en orbite au dessus de l'atmosphère), sans aberration optique et sans obstruction centrale, l'image d'une étoile serait alors par définition une tache d'Airy, profil de diffraction d'une ouverture circulaire. De profil cylindrique, celle-ci est caractérisée par un pic central lumineux et une alternance d'anneaux sombres (tombant à une intensité nulle) et d'anneaux de moins en moins lumineux. La séparation entre le centre et la première intensité nulle est de  $1.22\lambda/D$ , séparation qui définit le critère de Rayleigh ou critère de séparabilité théorique de points sources. La performance d'une optique adaptative et/ou la qualité du ciel vue de l'AO est en général quantifiée par le rapport de Strehl. Le rapport de Strehl est défini comme le rapport des intensités du centre de la réponse impulsionnelle dans le cas corrigé par AO et dans le cas d'une correction parfaite (maximum du profil d'Airy). Celui-ci dépend donc des conditions de turbulence.

Différents principes d'optique adaptative ont déjà été beaucoup étudiés et utilisés (ex actuels : NAOS pour le VLT et l'instrument NACO en particulier, ALTAIR pour le télescope Gemini et l'instrument NIRI, ou les NGS AO et LGS AO pour le Keck et la caméra NIRC-2). Ils ont déjà permis d'imager des systèmes binaires resserrés et d'assez fort contraste entre les deux objets (Chauvin et al. [2005] avec le VLT, Lafrenière et al. [2008] avec Gemini, Schneider [2010] en général). Cependant, ces performances sont encore insuffisantes pour l'observation directe de la plupart des exoplanètes d'intérêt. Plus généralement, l'imagerie à haute dynamique et haute résolution angulaire requiert une nouvelle génération d'AO appelée "AO extrême". Les futurs instruments destinés à l'imagerie directe d'exoplanète GPI (Macintosh et al. [2006]) et SPHERE (Wildi et al. [2007]) seront situés en aval de telles AO. Concrètement, pour le VLT les rapports de Strehl obtenus en pratique avec NAOS sont de l'ordre de 40-60% (voir Clenet et al. [2004]) et ceux visés pour la future AO extrême du VLT devraient atteindre régulièrement 80% (voir Fusco et al. [2006], Aller-Carpentier et al. [2008]).

Aux rapports de Strehl actuellement accessibles, le halo (image indiquant l'intensité lumineuse moyenne) dû aux aberrations atmosphériques aléatoires reste plus importants que le halo dû à la réponse déterministe

<sup>1</sup>La taille de ce halo est donnée par le *seeing* du ciel au moment des observations, typiquement 1".

de l'instrument *i.e.* à la PSF non normalisée (Racine et al. [1999]). En revanche, pour un rapport de Strehl de l'ordre de 90% la tendance s'inverse (Aime and Soummer [2004]). Cependant, la coronographie étant une méthode instrumentale requérant une description ondulatoire de la lumière, la description que nous avons donné ici en termes d'intensité moyenne de la lumière est insuffisante pour en déduire si un coronographe est vraiment utile pour l'instrument SPHERE qui sera installé derrière une AO extrême.

Comme nous le reverrons, comme la distribution de l'amplitude complexe en une position fixée peut être modélisée par une certaine loi gaussienne, que la composante déterministe de l'onde n'est plus négligeable par rapport à la composante aléatoire et que ces "deux contributions" sont cohérentes, la variance de l'intensité (même avant application du bruit de photon) peut s'exprimer comme une fonction non seulement du halo dû aux turbulences aléatoires (de moyenne nulle en amplitude complexe) mais aussi dû à la composante déterministe. Plus précisément, en une position fixée la moyenne et la variance de l'intensité s'expriment comme

$$i = i^{(c)} + i^{(s)} \quad (5.1)$$

$$\sigma_i^2 = \left(i^{(s)}\right)^2 + 2i^{(s)}i^{(c)} \quad (5.2)$$

où  $i^{(s)}$  est l'intensité moyenne (en une position du champ gardée implicite ici) dûe aux *speckles* et  $i^{(c)}$  l'intensité moyenne dûe à la partie déterministe de l'onde (profil type "Airy"). On constate ainsi que la variance de l'intensité en un point, dûe en premier lieu à la contribution aléatoire de l'onde *i.e.* aux *speckles*, est ainsi amplifiée par la contribution déterministe de l'onde,  $i^{(c)}$ , et donc en particulier au niveau des maxima des ailes de diffraction. Ces *speckles* sont appelés *pinned speckles* (Bloemhof [2004]). Plus l'intensité déterministe est amoindrie –par un coronographe– dans les zones d'intérêt, plus la variance de bruit  $y$  est affaiblie. Le bruit de Poisson accentue cet effet. Aime and Soummer [2004] quantifient plus en détail les régimes pour lesquels un coronographe apporte une véritable amélioration de l'image post-AO.

La modélisation adoptée par Aime and Soummer [2004] ne s'intéresse qu'à la densité de probabilité par pixel (densité marginale) de l'intensité lumineuse. Ferrari [2006] étudie un modèle joint de l'intensité lumineuse dans le champ, modèle issu d'une extension naturelle du modèle monovarié. Il quantifie ainsi la corrélation entre l'intensité d'un pixel et l'intensité des pixels voisins. Notons de plus que Aime and Soummer [2004] ne modélisent pas non plus les aberrations optiques post-AO et l'intégration temporelle de l'intensité le temps d'une pose. Soummer et al. [2007] introduisent dans ce modèle la présence d'aberrations statiques et quasi-statiques post-AO, considèrent le temps de vie de chaque composante et quantifient leur contribution dans la variance de l'intensité d'un pixel.

Notons que dans les modèles mentionnés ci-dessus le champ électromagnétique est toujours modélisé par un champ scalaire dans lequel chaque position est ainsi simplement caractérisée par une amplitude complexe et une phase. Quoiqu'il en soit, dans le cas de l'utilisation simple d'un laser (lumière cohérente et amplitude complexe moyenne nulle), on peut constater que la distribution statistique de l'intensité des *speckles* pour un pixel donné a *grosso modo* la même forme que l'onde soit polarisée ou non (Goodman [1985]). Notons aussi que les modèles supposent des images monochromatiques, hypothèse que l'on suppose suffisamment pertinente dans le cas de l'instrument IRDIS de SPHERE car les filtres sont "très étroits" ( $\Delta\lambda/\lambda \ll 1$ ).

## Modèle physique statistique des données (MGF)

Le modèle employé ici pour décrire le champ d'intensité dans le plan focal est le modèle multivarié de Ferrari [2006]. Le temps de corrélation et d'intégration temporelle d'intensité n'y est pas pris en compte explicitement mais peut être facilement intégré *a posteriori* dans le modèle. La lumière émise par la planète étant incohérente de celle de l'étoile, son intensité est simplement sommée à celle de l'étoile. Les données sont alors modélisées comme la réalisation de deux processus aléatoires supplémentaires : le bruit de photons et un bruit additif gaussien non corrélé décrivant notamment le bruit de lecture. Reprenons les étapes de ce modèle pas à pas.



• **Distribution de probabilité de l'intensité en un pixel du plan focal, principe des *speckles***

Goodman [1975] est à l'origine des calculs de statistiques en imagerie *speckle*. Plus généralement, Goodman [1985] ou Goodman [2007] constituent des références pour l'étude probabiliste de la formation d'images.

L'idée centrale dans la description statistique des *speckles* en régime dit "pleinement développé" (Goodman [1985]) et en présence d'un champ moyen non nul est de partir d'un plan pupille (ouverture du télescope par exemple) dans lequel le front d'onde s'exprime très généralement comme un champ complexe scalaire aléatoire de moyenne non nulle, cette moyenne correspondant à l'onde plane sans aberration. Ensuite, ce plan pupille est imagé par l'équivalent d'une lentille convergente, de sorte que si l'on suppose l'approximation de Fraunhofer valable, l'amplitude complexe de l'onde sur le plan focal est simplement donnée par la transformée de Fourier de l'amplitude dans le plan pupille. Si l'on note  $\psi_{\ell}$  le front d'onde dans le plan focal en la position  $\ell$  (position définie –uniquement le temps de cette section et la suivante– comme le vecteur des deux coordonnées cartésiennes) et  $\tilde{\psi}_{\mathbf{u}}$  le front d'onde dans le plan pupille en la position  $\mathbf{u}$ ,

$$\psi_{\ell} = \int d\mathbf{u} \tilde{\psi}_{\mathbf{u}} e^{-2i\pi\mathbf{u}^t\ell} \quad (5.3)$$

Schématiquement, l'idée est alors que si le champ aléatoire incident  $\tilde{\psi}$  est constitué de suffisamment de zones suffisamment décorrélatées et que le moyennage effectué entre toutes ces zones par les termes de phase  $e^{-2i\pi\mathbf{u}^t\ell}$  est assez important (ce qui n'est pas le cas au centre du plan focal, voir Soummer and Ferrari [2007]) et suffisamment *circulaire*, alors  $\psi_{\ell}$  a approximativement une distribution gaussienne complexe circulaire décentrée (définition en annexe de Ferrari [2006] ou dans Picinbono [1993]) quelle que soit la distribution de  $\tilde{\psi}$ . Ceci découle d'études asymptotiques de la transformée de Fourier présentées notamment par Brillinger [1981] et appliquées à l'imagerie par Goodman [1985].

On note ainsi

$$\psi_{\ell} \sim \mathcal{N}^c(\mu_{\ell}, \mathbf{E}[|\phi_{\ell}|^2]) \quad (5.4)$$

où on décompose

$$\psi = \boldsymbol{\mu} + \boldsymbol{\phi} \quad (5.5)$$

avec  $\boldsymbol{\mu}$  le fond déterministe ( $\boldsymbol{\mu} = \mathbf{E}[\boldsymbol{\psi}]$ ) et donc  $\boldsymbol{\phi}$  la contribution purement aléatoire de l'onde ( $\mathbf{E}[\boldsymbol{\phi}] = \mathbf{0}$ ) que l'on appelait jusqu'ici la contribution *speckle*.

Notons que si l'on ne spécifie pas "circulaire" le résultat découle naturellement du théorème central limite car un vecteur complexe est défini comme gaussien simplement si la concaténation de ses parties réelles et imaginaires a une distribution jointe gaussienne. La "circularité" demande une étude plus attentive de la distribution gaussienne complexe, voir Picinbono [1993].

L'intensité en ce point est donnée, conditionnellement à l'amplitude complexe dans tout le champ focal, par le module au carré de l'amplitude complexe en ce point :

$$i_{\ell} = |\psi_{\ell}|^2 = |\mu_{\ell}|^2 + |\phi_{\ell}|^2 + 2\mathcal{R}e(\mu_{\ell}^* \phi_{\ell}) \quad (5.6)$$

de telle sorte que ses deux premiers moments sont donnés par :

$$\mathbf{E}[i_{\ell}] = |\mu_{\ell}|^2 + \mathbf{E}[|\phi_{\ell}|^2] \quad (5.7)$$

$$\text{var}[i_{\ell}] = (\mathbf{E}[|\phi_{\ell}|^2])^2 + 2|\mu_{\ell}|^2 \mathbf{E}[|\phi_{\ell}|^2] \quad (5.8)$$

où on retrouve l'Eq. (5.1) avec  $i^{(s)} = \mathbf{E}[|\phi_{\ell}|^2]$  et  $i^{(c)} = |\mu_{\ell}|^2$ .

On constate bien que l'intensité moyenne est due en partie à la contribution déterministe de l'onde (réponse de l'instrument à une onde plane) et en partie à la variance du bruit (halo post-AO). La variation

statistique de l'intensité quant à elle est dûe d'une part à la variance du bruit mais aussi à la variance du bruit amplifiée par la réponse de l'instrument. Comme nous l'avons mentionné dans la section précédente, ce phénomène d'amplification de la variance des *speckles* sur la partie la plus brillante des anneaux est appelé phénomène de *pinned speckle*.

Au delà du terme de moyenne, la loi décrivant cette intensité est appelée loi de Rice modifiée, dérivée initialement par Goodman [1975]. Elle s'exprime sous la forme :

$$p(i_\ell) = \frac{1}{\mathbb{E}[|\phi_\ell|^2]} \exp\left(-\frac{i_\ell + |\mu_\ell|^2}{\mathbb{E}[|\phi_\ell|^2]}\right) I_0\left(\frac{2\sqrt{i_\ell|\mu_\ell|^2}}{\mathbb{E}[|\phi_\ell|^2]}\right)$$

où  $I_0$  est la fonction de Bessel modifiée d'ordre 0 du premier type. Elle peut être reliée à une loi du  $\chi^2$  décentrée. Pour  $\mu = \mathbf{0}$ , cette distribution est la loi exponentielle décroissante, décrivant la statistique d'un *speckle* "pur", *i.e.* non amplifié par un fond déterministe.

### • Applicabilité en présence d'optique adaptative et d'un coronographe

Ce résultat est bien connu en imagerie haute dynamique et haute résolution angulaire en astronomie. Il s'avère que ces distributions continuent à s'appliquer au niveau du plan focal post-AO (Canales and Cagigal [1999] et vérification expérimentale par Fitzgerald and Graham [2006] sur l'intensité) et au niveau du plan focal post-coronographe (calcul analytique et simulation de la distribution de l'amplitude complexe par Soummer et al. [2007]). Notons que le traitement spécifique du point central du plan focal, où les différents termes de phase de l'Eq. (5.4) valent 1, est traité par Soummer and Ferrari [2007] dans le cas d'un haut rapport de Strehl.

Une description physique détaillée de la propagation déterministe de la lumière et de sa composante aléatoire à travers un coronographe a été proposée notamment par Ferrari [2006] et Soummer et al. [2007] dans le cas d'un coronographe de Lyot ou de Roddier apodisé (Soummer et al. [2009]). En particulier, il s'avère que la famille du modèle statistique que nous adoptons pour décrire l'intensité dans le plan focal est identique qu'un coronographe soit utilisé ou non. Soummer et al. [2007] montrent ainsi qu'un coronographe parfait supprime la moyenne de l'amplitude complexe sans changer le type de distribution statistique. Après un calcul analytique de propagation de front d'onde à travers un coronographe, ils illustrent par simulation la distribution de l'amplitude complexe dans les quatre plans successifs du coronographe dans un cas de haut rapport de Strehl (*i.e.* avec une AO extrême). Pour un front d'onde incident strictement distribué selon une loi gaussienne complexe multivariée, Ferrari [2006] montre, par un développement de ce front sur la base des polynômes de Zernike, que le champ résultant dans le plan focal d'un coronographe est alors strictement gaussien complexe multivarié.

### • Distribution de probabilité jointe du champ d'intensité

Dans le modèle précédent (Eq. 5.4), seule la densité de probabilité en un point est dérivée. La corrélation des intensités entre pixels voisins ne peut donc être quantifiée sans généraliser le modèle monovarié. Ceci est nécessaire pour dériver la distribution de probabilité d'une statistique qui s'exprime comme une fonction de l'intensité mesurée simultanément en plusieurs pixels, telle que la statistique de test  $\hat{\alpha}$  construite dans la partie précédente et définie par l'Eq. (3.12).

Le modèle monovarié précédent (Eq. 5.4) est naturellement étendu au modèle joint suivant :

$$\psi \sim \mathcal{N}_L^c(\mu, \mathbb{E}[\phi\phi^H]) \quad (5.9)$$

où  $\mathcal{N}_L^c$  dénote la loi gaussienne complexe circulaire pour un vecteur aléatoire de taille  $L \times 1$  et  $\phi^H$  est la transposée Hermitienne de  $\phi$ .

Ferrari [2006] étudie alors la distribution jointe d'intensité, l'intensité étant donnée conditionnellement à  $\mu$  et  $\phi$  par l'équation déterministe (5.6) pour chaque position  $\ell$ . Dans un premier temps, il dérive les

moments du champ d'intensité grâce à leur développement en produits de cumulants (McCullagh [1987]). En particulier, généralisant en partie l'Eq. (5.7) et (5.8) :

$$\begin{aligned} \mathbf{E}[\mathbf{i}]_{\ell} &= |\mu_{\ell}|^2 + \mathbf{E}[|\phi_{\ell}|^2] \\ \text{cov}[\mathbf{i}]_{\ell, \ell'} &= |\mathbf{E}[\phi_{\ell}\phi_{\ell'}^*] + \mu_{\ell}\mu_{\ell'}^*|^2 - |\mu_{\ell}\mu_{\ell'}^*|^2 \end{aligned}$$

Plus généralement, la distribution du champ d'intensité est obtenue par Ferrari [2006] en considérant la matrice complexe  $\boldsymbol{\psi}\boldsymbol{\psi}^H$  dont la diagonale est égale au champ d'intensité  $\mathbf{i}$ . Notons  $\Sigma = \mathbf{E}[\boldsymbol{\phi}\boldsymbol{\phi}^H]$ . Si  $\boldsymbol{\psi} \sim \mathcal{N}_L^c(\boldsymbol{\mu}, \Sigma)$ , alors  $\boldsymbol{\psi}\boldsymbol{\psi}^H$  est distribuée selon la distribution de Wishart non centrée telle que discutée dans Tourneret et al. [2005] dans ce contexte ou plus généralement par Muirhead [1982] et Letac and Massam [2004]<sup>2</sup>. Cependant, par construction la matrice  $\boldsymbol{\psi}\boldsymbol{\psi}^H$  n'est pas inversible et sa densité de probabilité n'est donc pas définie par rapport à la mesure de Lebesgue. En revanche, la fonction génératrice des moments (MGF) existe et a une forme explicite simple donnée dans Ferrari [2006] :

$$h_{\boldsymbol{\psi}\boldsymbol{\psi}^H}(S) = \mathbf{E} \left[ e^{\text{Tr}(S\boldsymbol{\psi}\boldsymbol{\psi}^H)} \right] = \frac{e^{\text{Tr}[\boldsymbol{\mu}\boldsymbol{\mu}^H S(I_L - \Sigma S)^{-1}]} }{|I_L - \Sigma S|} \quad (5.10)$$

La distribution de  $\mathbf{i}$  est obtenue en marginalisant la distribution de  $\boldsymbol{\psi}\boldsymbol{\psi}^H$ . Or une densité marginale est obtenue par intégration de la densité jointe tandis qu'une MGF marginale est obtenue plus simplement par évaluation de la MGF jointe. En effet,

$$h_{\mathbf{i}}(\mathbf{s}) = \mathbf{E} \left[ e^{\sum_{\ell=1}^L s_{\ell} i_{\ell}} \right] = \mathbf{E} \left[ e^{\text{Tr}(D_{\mathbf{s}}\boldsymbol{\psi}\boldsymbol{\psi}^H)} \right] = h_{\boldsymbol{\psi}\boldsymbol{\psi}^H}(D_{\mathbf{s}})$$

avec  $D_{\mathbf{s}}$  la matrice diagonale dont les termes sont donnés par le vecteur  $\mathbf{s}$ . Incorporant l'Eq. (5.10) dans cette marginalisation on obtient :

$$h_{\mathbf{i}}(\mathbf{s}) = \frac{e^{\text{Tr}[\boldsymbol{\mu}\boldsymbol{\mu}^H D_{\mathbf{s}}(I_L - \Sigma D_{\mathbf{s}})^{-1}]} }{|I_L - \Sigma D_{\mathbf{s}}|} \quad (5.11)$$

Ferrari [2006] dérive finalement de cette distribution générale les distributions de variables aléatoires induites par  $\mathbf{i}$  dans les cas spécifiques  $L = 1$  et  $\boldsymbol{\mu} = \mathbf{0}$ . Notons que ce dernier cas peut être assoupli par l'approximation  $\text{Tr}[\boldsymbol{\mu}\boldsymbol{\mu}^H D_{\mathbf{s}}(I - \Sigma D_{\mathbf{s}})^{-1}] \ll 1$ , permettant ainsi d'expliquer la statistique d'une intensité dominée par la turbulence plutôt que par le fond déterministe (pas de *pinned speckle* : champ de *speckles* "purs").

Des études complémentaires de ce modèle et son utilisation pratique dans un cadre très proche du notre ont été proposées par Chatelain et al. [2006a] et Chatelain et al. [2007] qui prennent eux aussi en compte le bruit de photons qui suit.

### • Distribution de probabilité du champ de photo-électrons

Lorsque l'intensité lumineuse collectée par un pixel de CCD est convertie en électrons, un bruit dit de Poisson rend aléatoire le nombre d'électrons lus pour une intensité incidente donnée. Plus spécifiquement, si on note  $n_{\ell}$  le nombre d'électrons enregistrés par le pixel  $\ell$ ,

$$n_{\ell} | \mathbf{i} \sim \mathcal{P}(i_{\ell})$$

avec  $n_{\ell} | \mathbf{i}$  et  $n_{\ell'} | \mathbf{i}$  indépendants pour tous pixels  $\ell$  et  $\ell' \neq \ell$ .

La MGF multivariée du nombre de photo-électrons (vecteur d'entiers)  $\mathbf{n} = (n_1, \dots, n_L)^t$  est alors donnée à partir de celle de l'intensité (vecteur de réels non négatifs) grâce à Ferrari et al. [2004] par :

$$g_{\mathbf{n}}(\mathbf{z}) = \mathbf{E} \left[ \prod_{\ell=1}^L z_{\ell}^{n_{\ell}} \right] = h_{\mathbf{i}}(\mathbf{z} - \mathbf{1}) \quad (5.12)$$

<sup>2</sup>Si  $\boldsymbol{\mu} = \mathbf{0}$  la distribution de Wishart est dite centrée et la distribution induite du vecteur d'intensité est une distribution Gamma multivariée, voir Ferrari [2006], et pour son utilisation avec  $L = 2$  voir Chatelain et al. [2006b].

où  $h_i$  est donnée dans l'équation (5.11).

En plus des électrons dûs à la conversion des photons, des électrons peuvent apparaître du bruit de lecture et du courant d'obscurité. Un traitement classique des images CCD consiste à étalonner ces bruits additifs et soustraire leur moyenne des images d'intérêt. Ces images pré-traitées sont finalement modélisées par

$$\mathbf{x} = \mathbf{n} + \boldsymbol{\epsilon} \quad \text{avec } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_L)$$

et  $\mathbf{n}$  indépendant de  $\boldsymbol{\epsilon}$ . Ainsi,

$$h_{\mathbf{x}}(\mathbf{s}) = \mathbb{E} \left[ e^{\mathbf{s}^t \mathbf{x}} \right] = e^{\frac{\sigma^2 \|\mathbf{s}\|^2}{2}} g_{\mathbf{n}}(e^{\mathbf{s}}) \quad (5.13)$$

### Modèle statistique d'un jeu d'images en présence d'une exoplanète

La généralisation du modèle précédent à un ensemble de  $K$  images  $\{\mathbf{x}(k), k \in \llbracket 1, K \rrbracket\}$  en présence d'une exoplanète est immédiate.

À un instant donné  $t$ , comme les champs électromagnétiques de l'étoile et de la planète sont décorrélés, l'intensité instantanée  $\mathbf{i}^{(\delta)}(t)$  résultant des contributions de l'étoile et de la planète est égale à la somme de leurs intensités. D'après les hypothèses qualitatives de la partie I (détails dans la section 4.1), l'exoplanète a un profil connu et un mouvement apparent connu, à la position initiale  $r$  près. Conservant les notations précédentes de la partie I, si on note  $\mathbf{u}(t)$  le vecteur d'intensité instantanée due à l'étoile en l'instant  $t$ , on a alors :

$$\mathbf{i}^{(\delta)}(t) = \mathbf{u}(t) + \alpha \mathbf{p}(t, r) \quad (5.14)$$

$$\text{avec } u_\ell(t) = |\psi_\ell(t)|^2 \text{ et } \boldsymbol{\psi}(t) \sim \mathcal{N}_L^c(\boldsymbol{\mu}, \Sigma) \quad (5.15)$$

Cependant, classiquement la durée d'une pose est supérieure au temps de corrélation des bruits à haute fréquence temporelle (bruits post-AO). On modélise alors simplement l'intensité enregistrée le temps d'une pose  $k$  comme la somme de  $N_t$  vecteurs d'intensité instantanés indépendants. Comme nous le soulignerons, cette modélisation de l'intégration de l'intensité par une CCD est équivalente à supposer que ce sont les photo-électrons qui sont intégrés. Notons que dans le cas monovarié, Soummer et al. [2007] introduisent plus finement la prise en compte des corrélations temporelles en scindant la contribution aléatoire de l'onde en une contribution haute fréquence temporelle (résidus post-OA) et une contribution basse fréquence temporelle (speckles quasi-statiques). Ici comme dans le modèle précédent cependant nous supposons les *speckles* quasi-statiques *statiques*.

Si on note  $\delta_t$  le temps de corrélation des bruits de *speckle* post-AO et  $\Delta_t$  la durée d'une pose, on définit alors  $N_t$  par la partie entière de leur rapport :

$$N_t = \lfloor \Delta_t / \delta_t \rfloor \quad (5.16)$$

L'intensité enregistrée durant une pose est alors égale à

$$\mathbf{i}(k) = \sum_{t=1}^{N_t} \mathbf{i}^{(\delta)}(t)$$

et sa MGF à

$$h_{\mathbf{i}(k)}(\mathbf{s}) = \mathbb{E} \left[ e^{\mathbf{s}^t \sum_t \mathbf{i}^{(\delta)}(t)} \right] = \prod_t h_{\mathbf{i}^{(\delta)}(t)}(\mathbf{s}) \quad (5.17)$$

où, d'après l'Eq. (5.14) et comme la contribution de la planète est déterministe, la MGF instantanée  $h_{\mathbf{i}^{(\delta)}}$  est donnée par

$$h_{\mathbf{i}^{(\delta)}(t)}(\mathbf{s}) = \mathbb{E} \left[ e^{\mathbf{s}^t \mathbf{u}(t)} e^{\alpha \mathbf{s}^t \mathbf{p}(t, r)} \right] = e^{\alpha \mathbf{s}^t \mathbf{p}(t, r)} h_{\mathbf{u}(t)}(\mathbf{s})$$

où  $h_{\mathbf{u}(t)}(\mathbf{s})$  a été introduite dans l'Eq. (5.11). La distribution de  $\mathbf{u}(t)$  ne dépend pas de  $t$  puisque les moments du front d'onde émis par l'étoile ne dépendent pas du temps. On note ainsi la MGF plutôt  $h_{\mathbf{u}}(\mathbf{s})$ .

Finalement, partant de l'Eq. (5.17), la MGF de l'intensité d'une image se réexprime comme

$$h_{\mathbf{i}(k)}(\mathbf{s}) = e^{\alpha \mathbf{s}^t \mathbf{p}(k,r)} [h_{\mathbf{u}}(\mathbf{s})]^{N_t} \quad (5.18)$$

où  $h_{\mathbf{u}}$  est donnée par l'Eq. (5.11) et où on note comme auparavant  $\mathbf{p}(k,r)$  le profil de la planète intégré temporellement :  $\mathbf{p}(k,r) = \sum_{t=1}^{N_t} \mathbf{p}(t,r)$ .

Notons que la MGF  $(h_{\mathbf{u}}(\mathbf{s}))^{N_t}$  s'exprime aussi comme  $(h_{\mathbf{u}}(\mathbf{s}))^{N_t} = h_{\sum_{t=1}^{N_t} \mathbf{u}(t)}(\mathbf{s})$  avec  $\mathbf{u}(t)$  le vecteur donné par la diagonale de la matrice  $\boldsymbol{\psi}(t)\boldsymbol{\psi}(t)^H$  et avec  $\boldsymbol{\psi}(t) \sim \mathcal{N}_L^c(\boldsymbol{\mu}, \Sigma)$ . La distribution de  $\sum_{t=1}^{N_t} \boldsymbol{\psi}(t)\boldsymbol{\psi}(t)^H$  est alors la distribution de Wishart non centrée de moyenne  $N_t \boldsymbol{\mu} \boldsymbol{\mu}^H$  et de  $N_t$  degrés de liberté (Muirhead [1982]). Si  $L > N_t$ , cette dernière n'a pas de densité par rapport à la mesure de Lebesgue mais en possède une si  $L \leq N_t$ . Or en général, le nombre de pixels de l'image (et même uniquement de l'anneau étudié) est très supérieur aux nombres de poses instantanées décorréelées formant une pose. La distribution de  $\sum_t \boldsymbol{\psi}(t)\boldsymbol{\psi}(t)^H$  et donc de  $\mathbf{i}(k)$  n'a donc en général pas de PDF.

On peut de la même façon constater que contrairement à un modèle gaussien du vecteur d'intensités, passer du modèle *speckle* du vecteur d'intensités instantanées au modèle longue pose associé (où une longue pose est constituée de  $N_t$  poses instantanées décorréelées) ne revient pas à multiplier des paramètres de la distribution d'une pose instantanée par  $N_t$ . Il ne suffit par exemple pas de multiplier la contribution déterministe de la moyenne de l'intensité  $|\mu_\ell|^2$  par  $N_t$  car le déterminant du dénominateur de la MGF instantanée (5.11) est élevé à la puissance  $N_t$ .

Incorporant le bruit poissonnien et le bruit gaussien comme dans le cas instantané et sans planète, la MGF des données  $\mathbf{x}(k)$  mesurées en la pose  $k$  est finalement donnée par

$$h_{\mathbf{x}(k)}(\mathbf{s}) = e^{\frac{\sigma^2 \|\mathbf{s}\|^2}{2}} h_{\mathbf{i}(k)} [e^{\mathbf{s}} - \mathbf{1}] \quad (5.19)$$

avec  $h_{\mathbf{i}(k)}$  donné par l'équation (5.18).

## 5.2 Distribution statistique fine de l'estimateur linéaire d'intensité de la section 3.2

### Motivation pour l'étude de l'estimateur linéaire d'intensité de la section 3.2

Afin d'obtenir un estimateur utilisable en pratique sur les jeux de données attendus pour SPHERE, il était nécessaire de définir des statistiques d'inférence qui soient des fonctions relativement simples des données  $\mathbf{x}$ . Ces fonctions doivent être d'autant plus simples qu'il est important de pouvoir vérifier que leurs propriétés statistiques sont bonnes (estimation sans biais de la valeur estimée, erreur quadratique moyenne qui tend vers 0 lorsque le nombre de poses tend vers l'infini ...). Au delà de la phase d'estimation, il est primordial d'estimer à peu près correctement la probabilité d'erreur d'un test de détection, sans quoi on ne peut accorder aucune confiance au résultat d'une détection ou d'une non détection pour des rapports signal sur bruit particulièrement faibles.

Pour la phase d'estimation d'intensité de la planète, dans la partie précédente nous avons ainsi proposé dans la section 3.2 d'estimer l'intensité  $\alpha$  de la planète par la méthode de maximum de vraisemblance appliquée à un modèle gaussien simple des données  $\mathbf{x}$  (ML calculé sur l'ensemble des paramètres inconnus  $\alpha, r, \boldsymbol{\mu}, \sigma^2$  du modèle gaussien simple). L'estimateur global par ML donné par l'Eq. (3.23) est bien applicable en pratique mais il n'est en général pas possible d'étudier les propriétés de  $\hat{\alpha}_{\text{ML}}(\mathbf{x})$  (moyenne et variance) car l'expression de  $\hat{\alpha}_{\text{ML}}(\mathbf{x})$  fait appel, via  $\hat{r}_{\text{ML}}(\mathbf{x})$  donné par l'Eq. (3.24), à un *argmaximum* sur les positions initiales  $\{r \in \llbracket 1, L \rrbracket\}$  qui n'est pas simplifiable. Nous considérerons donc la position initiale  $r$  fixée et éventuellement connue, comme nous l'avons fait déjà dans la section 3.4 où nous avons

décidé d'appliquer le test de détection position initiale par position initiale (*i.e.* pixel par pixel). Ainsi,  $\hat{\alpha}_{\text{ML}}(\mathbf{x})$  se réduit à un estimateur pixel par pixel  $\hat{\alpha}(\mathbf{x}, r)$ , linéaire en les données  $\mathbf{x}$  et dont l'expression est donnée par l'Eq. (3.12) et que l'on rappelle ici :

$$\hat{\alpha}(\mathbf{x}, r) = \frac{1}{K} \sum_k \mathbf{w}(k, r)^t \mathbf{x}(k)$$

De même, pour la phase de détection la statistique simple utilisée pour l'étape de détection d'exoplanète est l'estimateur linéaire  $\hat{\alpha}(\mathbf{x}, r)$  défini par l'Eq. (3.12) dont les propriétés en tant qu'estimateur ont été étudiées notamment pour une position initiale supposée connue. Le test d'hypothèses basé sur  $\hat{\alpha}(\mathbf{x}, r)$  effectué sur chaque position initiale "possible"  $r$  a été décrit dans la section 3.4. Afin de relier le seuil  $\zeta$  à une probabilité de fausse alarme (PFA) proche de celle que l'on observerait empiriquement pour les données vraies, le modèle statistique décrivant la statistique  $\hat{\alpha}(\mathbf{x}, r)$  doit être précis et en particulier au niveau de la queue de la distribution puisque la PFA est égale à l'aire sous la queue de la densité de probabilité de  $\hat{\alpha}(\mathbf{x}, r)$ .

### Calcul de la MGF de $\hat{\alpha}(\mathbf{x})$

Une fois construite,  $\hat{\alpha}(\mathbf{x}, r)$  est une simple statistique, ou fonction des données  $\mathbf{x}$  paramétrée par  $r$ . La distribution de probabilité et les moments de  $\hat{\alpha}(\mathbf{x}, r)$  peuvent donc être étudiés selon que  $\mathbf{x}$  est décrit par un modèle statistique ou un autre. Sous le modèle gaussien de  $\mathbf{x}$  étudié dans la partie I,  $\hat{\alpha}(\mathbf{x}, r)$  suit une loi gaussienne de moyenne  $\alpha$  et de variance égale à la composante correspondante de la borne de Rao-Cramer, voir la section 3.3. Comme auparavant, gardons la dépendance de  $\hat{\alpha}(\mathbf{x}, r)$  en  $\mathbf{x}$  implicite.

Sous le modèle fin de *speckle* développé dans la section 5.1, la MGF d'une pose  $\mathbf{x}(k)$  est donnée par l'équation (5.19). Les poses étant indépendantes, la MGF de  $\hat{\alpha}$  est donnée sous  $H_1$  par

$$h_{\hat{\alpha}(r)}^{(1)}(z) = \mathbb{E}[e^{z\hat{\alpha}(r)}] = \prod_{k=1}^K h_{\mathbf{x}(k)} [zK^{-1}\mathbf{w}(k, r)] \quad (5.20)$$

En substituant la MGF des données grâce aux équations (5.19), (5.18) et (5.11) on obtient

$$\begin{aligned} h_{\hat{\alpha}(r)}^{(1)}(z) &= \prod_{k=1}^K e^{\frac{z^2\sigma^2\|\mathbf{w}(k,r)\|^2}{2K^2}} h_{\mathbf{i}(k)} [e^{zK^{-1}\mathbf{w}(k,r)} - \mathbf{1}] \\ &= e^{\frac{z^2\sigma^2\sum_k\|\mathbf{w}(k,r)\|^2}{2K^2}} \prod_k \exp \left[ \alpha \left( e^{zK^{-1}\mathbf{w}(k,r)} - \mathbf{1} \right)^t \mathbf{p}(k, r) \right] \left[ h_{\mathbf{u}} \left( e^{zK^{-1}\mathbf{w}(k,r)} - \mathbf{1} \right) \right]^{N_t} \end{aligned}$$

Or d'après l'équation (3.12),  $\sum_k \|\mathbf{w}(k, r)\|^2 = c_K^{-2} \sum_k \|\mathbf{p}(k, r) - \bar{\mathbf{p}}(r)\|^2$  et donc d'après l'équation (3.11),

$$\sum_k \|\mathbf{w}(k, r)\|^2 = \frac{K}{c_K}$$

Donc finalement,

$$\ln \left[ h_{\hat{\alpha}(r)}^{(1)}(z) \right] = \alpha \sum_k \left[ \exp \left( zK^{-1}\mathbf{w}(k, r) \right) - \mathbf{1} \right]^t \mathbf{p}(k, r) + \ln \left[ h_{\hat{\alpha}}^{(0)}(z) \right] \quad (5.21)$$

avec le logarithme de la MGF de  $\hat{\alpha}(r)$  donnée sous  $H_0$  par :

$$\begin{aligned} \ln \left[ h_{\hat{\alpha}(r)}^{(0)}(z) \right] &= \frac{z^2\sigma^2}{2Kc_K(r)} - N_t \sum_k \ln |I_L - \Sigma(D_{\exp(zK^{-1}\mathbf{w}(k,r))} - I_L)| \\ &+ N_t \text{Tr} \left[ \boldsymbol{\mu}\boldsymbol{\mu}^H \sum_k (D_{\exp(zK^{-1}\mathbf{w}(k,r))} - I_L)(I_L - \Sigma(D_{\exp(zK^{-1}\mathbf{w}(k,r))} - I_L))^{-1} \right] \quad (5.22) \end{aligned}$$

### 5.3 Consistance de l'estimateur d'intensité

On déduit directement les moments de  $\hat{\alpha}(\mathbf{x}, r)$  non par une intégrale de la PDF (impossible quoiqu'il en soit ici car la PDF des données n'existe pas pour des poses de durée raisonnable comme discuté précédemment) mais par une simple évaluation de dérivées de la MGF. Ces moments sont comme toujours calculés en supposant l'ensemble des paramètres du modèle connus.

A un nombre de poses  $K$  fixé, comme  $\sum_k \mathbf{w}(k, r) = \mathbf{0}$  quel que soit  $r$ , on constate en particulier que

$$E[\hat{\alpha}] = \alpha \quad (5.23)$$

$$\begin{aligned} \text{var}[\hat{\alpha}] &= \frac{\sigma^2}{K c_K(r)} + \alpha K^{-2} \sum_k \sum_\ell p_\ell(k, r) w_\ell(k, r)^2 + K^{-2} \sum_\ell (|\mu_\ell|^2 + \Sigma_{\ell, \ell}) \sum_k w_\ell(k, r)^2 \\ &\quad + K^{-2} \sum_{\substack{1 \leq \ell \leq L \\ 1 \leq \ell' \leq L}} (|\Sigma_{\ell, \ell'}|^2 + 2\text{Re}(\mu_\ell^* \mu_{\ell'} \Sigma_{\ell, \ell'})) \sum_k w_\ell(k, r) w_{\ell'}(k, r) \end{aligned} \quad (5.24)$$

et donc notamment que  $\hat{\alpha}(\mathbf{x}, r)$  est toujours non biaisé et que la variance de l'estimateur d'intensité de la section 3.2 est en fait bien plus grande que calculée sous le modèle gaussien par l'Eq. (3.33). La variance sous modèle gaussien ne constitue en effet que le premier terme de la variance plus réaliste calculée ci-dessus à partir du modèle *speckle*.

Pour une matrice de covariance  $\Sigma$  diagonale (front d'onde décorréolé spatialement), notons que cette variance peut s'exprimer sous la forme plus simple

$$\text{var}[\hat{\alpha}] = \frac{\sigma^2}{K c_K(r)} + \alpha K^{-2} \sum_{\ell, k} p_\ell(k, r) w_\ell(k, r)^2 + K^{-2} \sum_\ell (d_\ell(1 + 2\Sigma_{\ell, \ell}) + \Sigma_{\ell, \ell}(1 + \Sigma_{\ell, \ell})) \sum_k w_\ell(k, r)^2$$

en notant par analogie avec le modèle gaussien  $d_\ell = |\mu_\ell|^2$ .

Si  $\|\mathbf{p}(k, r)\| < \infty$  quels que soient  $k$  et  $r$  (ce qui est le cas en pratique), les second, troisième et quatrième termes de (5.24) contribuent au terme de variance comme  $O(K^{-1}c_K^{-2})$  pour  $K \rightarrow \infty$ . Pour que l'estimateur  $\hat{\alpha}(\mathbf{x}, r)$  soit consistant, ceci impose une condition légèrement plus restrictive sur le mouvement de la source que  $K^{-2}c_K^{-2} \rightarrow 0$  imposé avec le modèle gaussien.

Finalement, on constate que quoique le modèle gaussien simple utilisé en partie I pour construire  $\hat{\alpha}(\mathbf{x}, r)$  soit en général inadapté à des images acquises avec les instruments de seconde génération de détection d'exoplanète (AO extrême et coronographe), l'estimateur  $\hat{\alpha}(\mathbf{x}, r)$  qui en est dérivé a toujours certaines propriétés satisfaisantes sous le modèle fin de la section 5.1. Non seulement il est non biaisé mais il s'avère de plus que, sous des conditions à peine plus restrictives que celles dérivées et discutées dans le cas gaussien, la variance de  $\hat{\alpha}$  donnée dans l'équation (5.24) tend vers 0 lorsque le nombre de poses tend vers l'infini.

## Chapitre 6

# Probabilité de Fausse Alarme du détecteur de la section 3.4 sous modèle *speckle*

### 6.1 Approximations analytiques de la PFA

#### Procédure de détection habituelle

La procédure de détection consiste comme introduit en section 3.4 à seuiller la statistique linéaire  $\hat{\alpha}(\mathbf{x}, r)$  définie par l'Eq. (3.12) :

$$\text{Rejeter } H_0 \text{ en } r \text{ si } \hat{\alpha}(\mathbf{x}, r) > \zeta$$

Comme discuté alors, le test d'hypothèses basé sur  $\hat{\alpha}(\mathbf{x}, r)$  devrait donc être effectué en chaque position initiale "possible"  $r$ . L'enjeu de cette section est d'estimer la PFA (et la PD, supposant alors  $r$  connue) de ce test pour des seuils menant à des PFA raisonnablement faibles et pour des données  $\mathbf{x}$  décrites par la distribution *speckle* étudiée dans la présente partie. Rappelons que la PFA est simplement égal à 1 moins la distribution cumulative de  $\hat{\alpha}(\mathbf{x}, r)$ .

Le problème est beaucoup plus complexe que dans le cas gaussien où  $\hat{\alpha}(\mathbf{x}, r)$  était simplement décrit par une distribution gaussienne. Avec le modèle *speckle*, comme on l'a vu dans la section 5.2 en général la distribution des données n'a pas de densité par rapport à la mesure de Lebesgue sous le modèle *speckle* car le nombre d'intervalles de temps décorrélés dans une pose n'est pas suffisamment grand par rapport au nombre de pixels d'un anneau. Ici on considère le cas extrême où  $N_t = 1$ , ce qui sera justifié dans la section 6.3. Il a donc été nécessaire d'exprimer la distribution de  $\hat{\alpha}(\mathbf{x}, r)$  par sa MGF.

Afin de donner une expression analytique approximative de la PFA à partir de la MGF nous allons appliquer et comparer deux types de développements en série : d'une part les développements en série d'Edgeworth et de Gram-Charlier et d'autre part le développement autour d'un point selle d'une fonction analytique bien choisie (approximation dite du point selle ou *saddlepoint*).

#### Développements en série d'Edgeworth et de Gram-Charlier

Une technique standard pour approximer analytiquement la distribution cumulative d'une statistique consiste à développer en série la PDF de cette statistique autour d'une distribution gaussienne. Le  $i^{\text{ème}}$  terme de cette série requiert le calcul du cumulatif d'ordre  $i$  de la statistique. Celui-ci peut-être obtenu à partir de l'expansion de Taylor du logarithme de la MGF, donné par les équations (5.21) et (5.22) pour la statistique  $\hat{\alpha}(\mathbf{x}, r)$ .

Les développements en série dits d'Edgeworth et de Gram-Charlier sont des développements construits à partir de la série décrite précédemment, mais en ordonnant différemment les termes et en classant



différemment les ordres de développement, voir Helstrom [1995]. Leur premier ordre correspond à un ajustement gaussien de la distribution de  $\hat{\alpha}(\mathbf{x}, r)$  (moments estimés par les moments empiriques). On confond ces deux développements car ils sont strictement égaux jusqu'à leurs troisièmes termes de développement inclus.

Comme nous allons le constater, pour un développement d'ordre raisonnable ces expansions ne permettent pas d'approximer correctement la distribution cumulative pour une PFA raisonnablement faible.

### Approximation du point selle (*saddlepoint*)

Afin d'estimer convenablement la distribution cumulative d'une statistique réelle, il est nécessaire que la PDF estimée approxime bien la PDF vraie au niveau de l'aile de la PDF. Pourtant, les développements en série autour d'une distribution gaussienne estiment en général convenablement la distribution vraie autour de sa valeur moyenne mais non au niveau de ses ailes. Au contraire, l'approximation du point selle se focalise sur l'intégrand  $\phi(z)$  mis en jeu dans le calcul de distribution cumulative effectué à partir de la MGF.

Ainsi, l'approximation du point selle (ou *saddlepoint*), consiste à effectuer un développement de la distribution autour d'un point  $z_0 \in \mathbb{R}$  en lequel la dérivée complexe d'une fonction analytique  $\phi$  – qui s'exprime très simplement à partir de la MGF – s'annule. Cette technique est présentée de façon synthétique par Helstrom [1995] et plus détaillée par Jensen [1995].

Pour appliquer la technique du point selle, la transformée de Laplace  $\tilde{h}_i$  de la PDF de  $\hat{\alpha}(\mathbf{x}, r)$  sous l'hypothèse  $H_i$  est utilisée plutôt que sa MGF, ces deux fonctions s'exprimant simplement l'une en fonction de l'autre par  $\tilde{h}_i(z) = h_i(-z)$ . En utilisant uniquement le premier terme du développement du point selle, les approximations analytiques de PFA et PD sont alors respectivement données par :

$$\text{PFA}(\zeta, r) \approx -(2\pi\phi_0^{(2)}(z_0, r))^{-1/2} e^{\phi_0(z_0, r)} \quad (6.1)$$

$$\text{PD}(\zeta, r) \approx 1 - (2\pi\phi_1^{(2)}(z_1, r))^{-1/2} e^{\phi_1(z_1, r)} \quad (6.2)$$

$$\text{avec } \exp[\phi_i(z, r)] = \frac{\tilde{h}_i(z, r)e^{\zeta z}}{n(i)z}, \quad n(0) = -1, n(1) = 1 \quad (6.3)$$

$$\text{et } \phi_i^{(j)}(z_0, r) = \left. \frac{\partial^j \phi_i(z, r)}{\partial z^j} \right|_{z=z_0} \quad (6.4)$$

où  $h_i$  est donnée par les équations (5.21) et (5.22), et où les points selles  $z_0 < 0$  et  $z_1 > 0$  satisfont

$$\phi_i'(z_i, r) = 0, \quad z_0 < 0, \quad z_1 > 0 \quad (6.5)$$

Ces points sont appelés *point selle* car lorsque une fonction analytique sur  $\mathbb{C}$  atteint un maximum local selon l'axe des réels elle atteint en ce point un minimum local selon l'axe imaginaire.

Notons que

- comme mentionné par Helstrom [1995], les contraintes  $z_0 < 0$  et  $z_1 > 0$  sont des conséquences naturelles mais non-nécessaires de notre intérêt pour les cas où  $\text{E}[\hat{\alpha}(r)] < \zeta$  sous  $H_0$  (PFA faible) et  $\text{E}[\hat{\alpha}(r)] > \zeta$  sous  $H_1$  (PD élevée).
- le calcul des dérivées de premier et second ordres de  $\phi_0(z, r)$  et  $\phi_1(z, r)$  est effectué à partir des équations (5.22) et (5.21) en utilisant le fait que  $(\ln |A(z)|)' = -\text{Tr}(A(z)^{-1}A(z)')$  et que  $(A(z)^{-1})' = -A(z)^{-1}A(z)'A(z)^{-1}$ . Les expressions de ces dérivées sont longues, non difficiles à calculer et peu intéressantes et ne sont donc pas présentées ici.
- le calcul de PFA et PD à partir des équations (6.1) et (6.2) ne fait pas nécessairement intervenir la résolution de l'équation  $\phi_i'(z_i, r) = 0$ . Afin par exemple de tracer PFA ou PD comme une fonction de  $\zeta$ , un vaste ensemble de seuils doit être couvert quoiqu'il en soit. Or une réexpression des équations (6.3), (6.4) et (6.5) permet de paramétriser  $\zeta$  par  $z_i$  :

$$\zeta(z_i, r) = \frac{1}{z_i} - \left. \frac{d \ln(\tilde{h}_i(z, r))}{dz} \right|_{z=z_i} \quad (6.6)$$

Ainsi, sélectionner un intervalle de  $\zeta$  se résume à sélectionner un intervalle de  $z_i$ , à la main si les fonctions  $\zeta(z_i, \cdot)$ ,  $\text{PFA}(\zeta)$  etc sont suffisamment douces.

- si  $\sigma^2 = 0$ , la statistique de test  $\hat{\alpha}(\mathbf{x}, r)$  est une combinaison linéaire de variables aléatoires discrètes. Quoique les formes (6.1) et (6.2) soient pertinentes pour approximer les probabilités cumulatives de variables aléatoires continues, la validité de ces expressions pour une variable aléatoire discrète a été récemment discutée par Aldosari and Moura [2007].

## 6.2 Comparaison des méthodes d'approximation par simulation pour le modèle *speckle*

### Cas de simulation

L'approximation du point selle utilisée pour calculer PFA et PD pour les MGF (5.21) et (5.22) pour  $N_t = 1$  est validée en utilisant des données générées à partir du modèle *speckle* décrit dans la section 5.1. Pour estimer une probabilité cumulative vraie à partir d'intégrations de Monte-Carlo classiques, le nombre d'échantillons  $\hat{\alpha}(\mathbf{x}, r)$  doit être au minimum égal à un sur cette probabilité. Comme chaque échantillon résulte du traitement d'un jeu de données constitué de  $K \times L$  pixels, et comme la PFA d'intérêt sont typiquement aussi faibles que  $10^{-4}$ , de telles intégrations de Monte Carlo requièrent un grand temps de calcul. Les données sont par conséquent simulées avec une structure spatiale à une dimension, contrairement aux images astronomiques que nous considérons en pratique, et pour des valeurs faibles de  $K$  et  $L$ .

Quatre cas de simulation sont considérés. Un cas de simulation de référence est défini et noté "Data 1" sur les figures. Les paramètres définissant ce cas de simulation sont les suivants :

- $K = 10$  and  $L = 30$
- Les  $\psi(k)$  sont générés à partir d'un processus autorégressif gaussien circulaire de premier ordre. Sa matrice de covariance est définie par  $\Sigma_{i,j} = \rho^{|i-j|} e^{j\phi(j-i)}$  avec les paramètres  $\boldsymbol{\mu} = 0.08 \times \mathbf{1}$ ,  $\phi = \pi/4$  et  $\rho = 0.85$ . Une taille caractéristique de corrélation induite par  $\Sigma$  peut être définie à partir de la largeur à -3dB de  $|\Sigma_{i,j}|$  vue comme une fonction de  $|i - j|$ . Elle est égale à 4 pixels pour  $\rho = 0.85$ .
- La source d'intérêt est uniformément accélérée à partir d'une position initiale  $r = 1$ . Afin de nous rapprocher de notre problématique astrophysique, le profil de la source (planète) est choisi statique et de taille caractéristique égale à la taille caractéristique de corrélation induite par  $\Sigma$ .
- La variance du bruit additif gaussien est pris égal à  $\sigma^2 = 0.25$
- Les jeux de données sont simulés d'une part sous  $H_0$  ( $\alpha = 0$ ) et d'autre part  $H_1$  avec  $\alpha = 0.3$ .
- $10^6$  jeux de données indépendants sont simulés aussi bien pour  $H_0$  que pour  $H_1$ .

Le graphe de gauche de la figure 6.1 montre la contribution  $\alpha \mathbf{p}(k, r)$  de la source pour tous les instants  $k = 1, \dots, K$ . Le graphe de droite de la figure 6.1 montre une réalisation de  $\mathbf{x}(10)$  sous chacune des hypothèses. Sous  $H_1$ , la source peut être observée entre les pixels 20 et 25.

La statistique de détection  $\hat{\alpha}(\mathbf{x}, r)$  est ensuite calculée en la valeur vraie  $r$  à partir de son expression en équation (3.12). La figure 6.2 montre les histogrammes de  $\hat{\alpha}(\mathbf{x}, r)$  obtenus sous  $H_1$  et  $H_0$  pour les  $10^6$  jeux de données simulés.

Trois autres cas de simulation sont étudiés. Chacun diffère du cas de référence "Data 1" par deux paramètres maximum.

- "Data 2" : modification du nombre de temps de poses :  $K' = 3K$ . Afin de conserver le même faible nombre de pixels pixels  $L$ , on donne à la source un mouvement lent et de vitesse constante.
- "Data 3" : modification du paramètre de corrélation :  $\rho' = 0.95$ . Il induit une taille caractéristique de corrélation de 14 pixels.
- "Data 4" : nouvelle moyenne de fond :  $\boldsymbol{\mu}' = 100\boldsymbol{\mu}$ . On attend alors en moyenne une centaine de photo-électrons par pixel plutôt que quelques uns seulement. On choisit alors une nouvelle intensité de la source d'intérêt :  $\alpha' = 20\alpha$ .

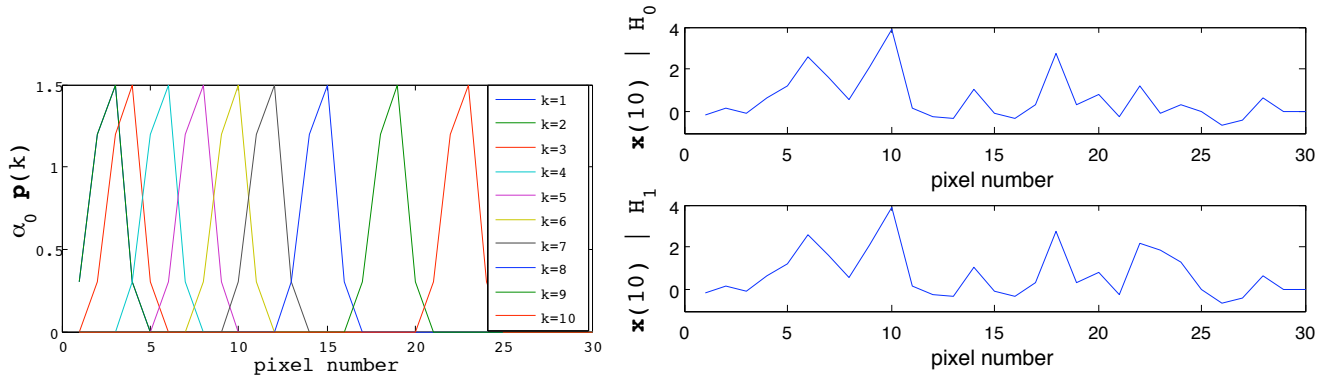


FIG. 6.1: Gauche : contribution  $\alpha_0 p(k, r)$  de la source pour  $k = 1, \dots, 10$  (noter que  $p(1, r) \approx p(2, r)$ ). Droite : réalisation de  $x(10)$  sous  $H_0$  (haut) et  $H_1$  (bas) pour le cas de référence, noté “Data 1”.

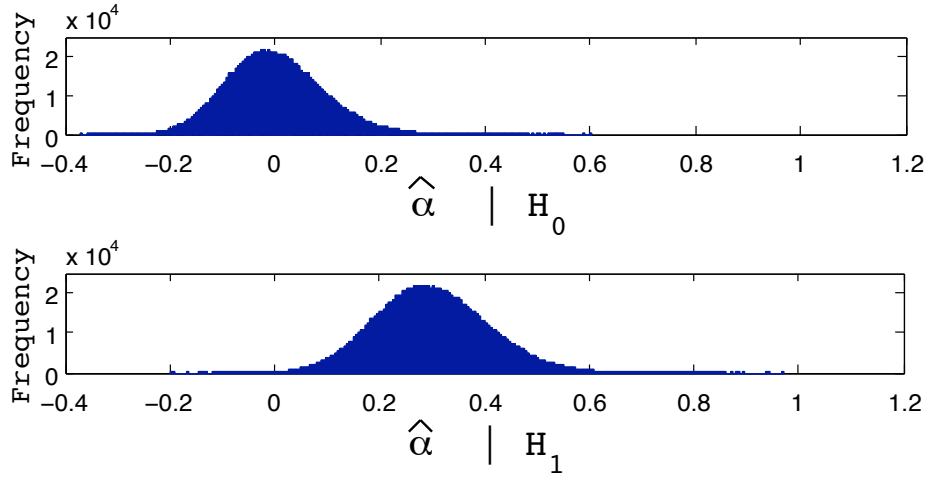


FIG. 6.2: Histogrammes de  $\hat{\alpha}(x, r)$  sous  $H_0$  (haut) et sous  $H_1$  (bas) pour le cas de référence, noté “Data 1”.

La figure 6.3 montre la courbe ROC associée à la procédure de test réintroduite dans la section 6.1 pour ces quatre cas de simulation. Chaque courbe est simplement estimée à partir des  $10^6$  échantillons de  $\hat{\alpha}(x, r)$  sous chacune des deux hypothèses.

### Evaluation des estimations de PFA et PD

La figure 6.4 montre PFA en fonction du seuil, pour la procédure de détection et selon les techniques d’approximation introduites dans la section 6.1. Plus précisément, les méthodes de calcul de PFA sont les suivantes :

- approximation du point selle par le premier ordre du développement
- distribution empirique de  $\hat{\alpha}(x, r)$  sous  $H_0$
- distribution gaussienne dont les moments sont estimés par les moments empiriques, qui sont égaux, à une précision relative de  $10^{-4}$ , aux moments vrais  $E[\hat{\alpha}] = \alpha$  et  $\text{var}[\hat{\alpha}]$  donné par l’équation (5.24).

Cet ajustement correspond au premier ordre des développements de Gram-Charlier et d’Edgeworth.

Dans tous les cas, pour une PFA allant de  $10^{-4}$  à  $10^{-1}$ , l’estimation de PFA par approximation de point selle ajuste parfaitement le graphe empirique, contrairement à l’approximation gaussienne. On

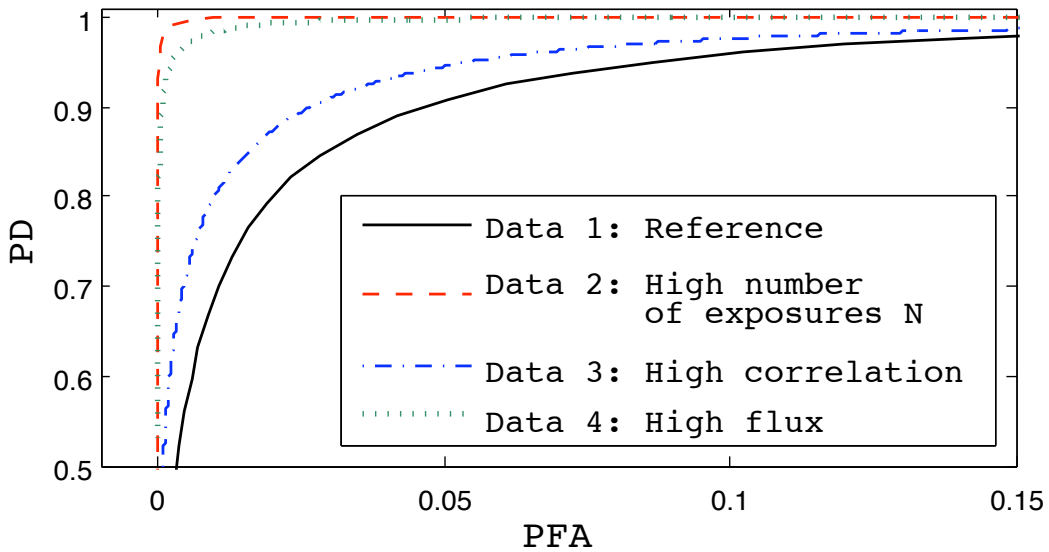


FIG. 6.3: Courbes ROC de la procédure de détection réintroduite dans la section 6.1 pour les quatre cas de simulation. Elles sont calculées à partir des distributions cumulatives, considérées ici comme les *vraies* PFA et PD dans l'analyse qui suit.

constate en fait que le comportement correct de l'approximation du point selle et le mauvais comportement de l'approximation gaussienne s'étendent pour des PFA plus faibles que  $10^{-10}$ . L'erreur relative de l'approximation gaussienne pour une PFA donnée est assez stable pour tous les cas étudiés. Ceci montre que l'approximation du point selle peut être considérée comme une estimation presque parfaite de la PFA si la PFA d'intérêt est plus faible que  $10^{-1}$  et que l'approximation gaussienne devrait dans la mesure du possible être évitée.

La figure 6.5 montre PD (pour la position initiale vraie) comme une fonction du seuil, calculée comme dans la figure 6.4, mais utilisant la seconde expression (6.2) afin de calculer l'approximation du point selle. L'approximation du point selle diverge de la PD *vraie* lorsque le seuil quitte la queue de la distribution de  $\hat{\alpha}$ , mais reste une meilleure approximation que l'approximation gaussienne pour  $PD < 0.6$ . Ce comportement qualitatif et les valeurs quantitatives sont assez stables pour les deux cas montrés, malgré le fait qu'ils représentent des cas de détection assez différents, comme ceci peut se voir sur la figure 6.3. L'approximation du point selle peut être considérée comme une estimation presque parfaite de PD pour  $PD > 0.8$  et devrait toujours être préférée à une approximation gaussienne pour  $PD > 0.6$ .

La figure 6.6 montre à nouveau PD comme une fonction du seuil, mais en utilisant aussi un développement autour de l'approximation gaussienne grâce aux développements de Gram-Charlier et d'Edgeworth. Cependant, l'utilisation de deux termes supplémentaires dans le développement de Gram-Charlier (le terme de base revenant à l'approximation gaussienne) ne met pas en évidence de différence significative. Le calcul des différents ordres du développement de Gram-Charlier fait intervenir le calcul de cumulants d'ordres trop élevés pour être calculés ici. Rappelons qu'à ce niveau de développement (trois termes) les développements de Gram-Charlier et d'Edgeworth sont identiques.

Finalement, la figure 6.7 montre que pour les cas où l'on peut s'attendre à une forte PD pour une PFA raisonnable ("Data 4", voir la figure 6.3), l'approximation du point selle reste une approximation très satisfaisante de PD. L'approximation gaussienne est de même qu'auparavant une mauvaise approximation. Ceci n'est pas très surprenant car l'intervalle de probabilité cumulative analysé ici est équivalent à celui analysé pour PFA. Notons aussi que pour un développement de Gram-Charlier d'ordre supérieur les estimations sont encore moins bonnes que l'approximation gaussienne.

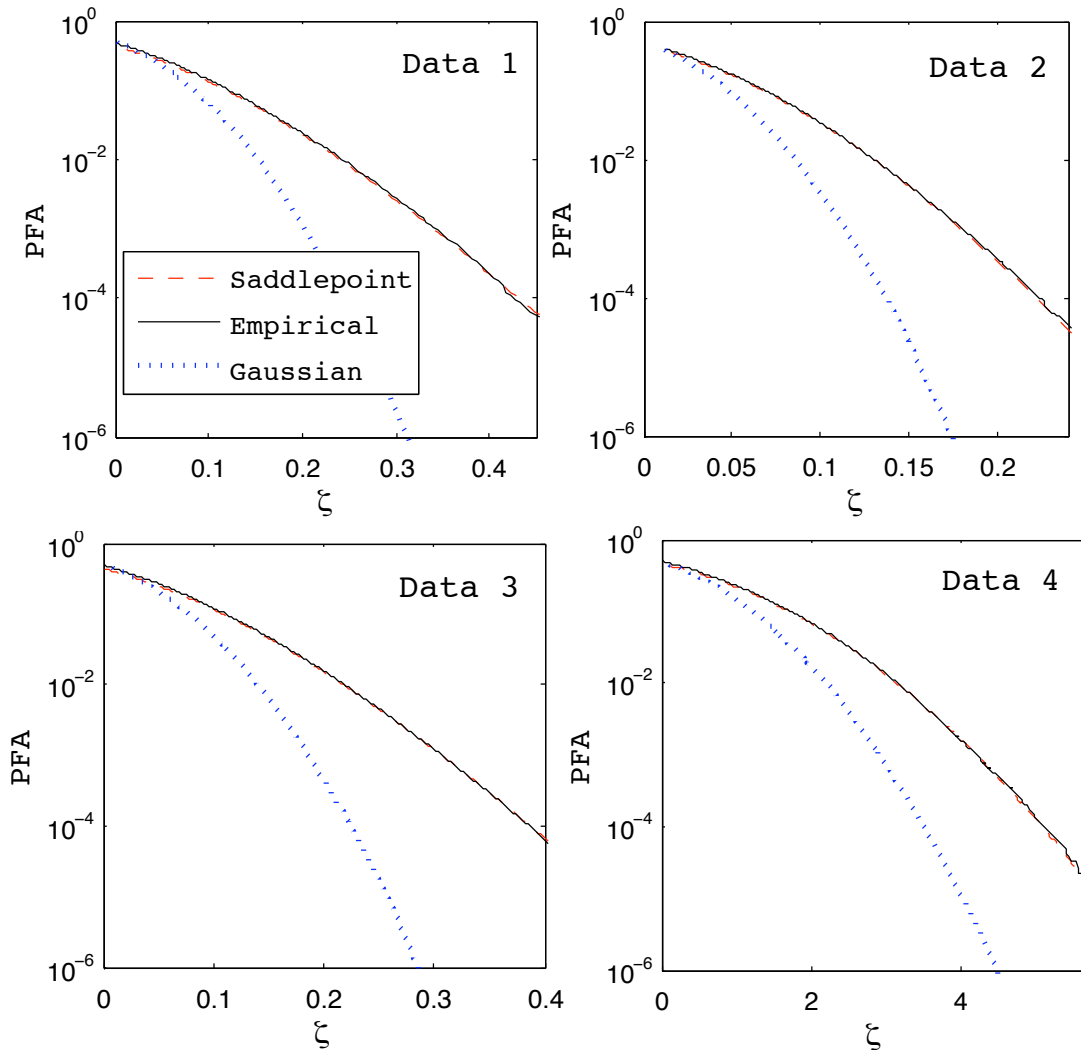


FIG. 6.4: PFA calculée à partir de l'approximations du point selle et de l'approximation gaussienne dont les moments sont estimés par les moments empiriques. Comparaison avec la PFA calculée à partir de la distribution empirique.

Autrement, les mauvaises performances de l'approximation gaussienne de la distribution empirique de  $\hat{\alpha}$  qui est observée dans les figures 6.4, 6.5, 6.6 et 6.7 peuvent être directement notées à partir des échantillons aléatoires de  $\hat{\alpha}$  montré sur la figure 6.2. L'application des tests de Kolmogorov-Smirnov et du Chi carré à la distribution empirique de  $\hat{\alpha}$  sous  $H_0$  et sous  $H_1$  rejète l'hypothèse d'une distribution gaussienne avec une  $p$ -value supérieure à 0.9999.

Par ailleurs, on peut noter que le *skewness*  $\gamma_1$  et le *kurtosis* centré  $\gamma_2$ , qui valent tous deux 0 pour une distribution gaussienne, valent ici  $\hat{\gamma}_1(H_0) \simeq 0.35$  sous  $H_0$ ,  $\hat{\gamma}_1(H_1) \simeq 0.24$  sous  $H_1$ ,  $\hat{\gamma}_2(H_0) \simeq 0.33$  et  $\hat{\gamma}_2(H_1) \simeq 0.18$ . Ceci quantifie le fait que la distribution de  $\hat{\alpha}$  est légèrement *right skewed*. On peut relier ces valeurs par exemple à celles correspondant à une distribution de Poisson. Cette comparaison est *a priori* relativement pertinente car  $\mathcal{P}(\lambda)$  a un paramètre de *skewness*  $\gamma_1(\lambda) = \lambda^{-1/2}$  et de *kurtosis*  $\gamma_2(\lambda) = \lambda^{-1}$  qui ont le même signe que  $\hat{\gamma}_1$  et  $\hat{\gamma}_2$ , et tendent vers ceux d'une distribution gaussienne lorsque  $\lambda \rightarrow \infty$ . On observe que :

- $\hat{\gamma}_1(H_0) > \gamma_1(10) \simeq 0.32$  et  $\hat{\gamma}_2(H_0) > \gamma_2(10) = 0.1$ . Ainsi, sous  $H_0$  la distribution empirique est (en termes de *skewness* et de *kurtosis*) "moins gaussienne" que  $\mathcal{P}(10)$ .
- $\hat{\gamma}_1(H_1) > \gamma_1(20) \simeq 0.22$ , et le *kurtosis* est même plus élevé que  $\gamma_2(10) > \gamma_2(20)$ . Ainsi, sous  $H_1$  la

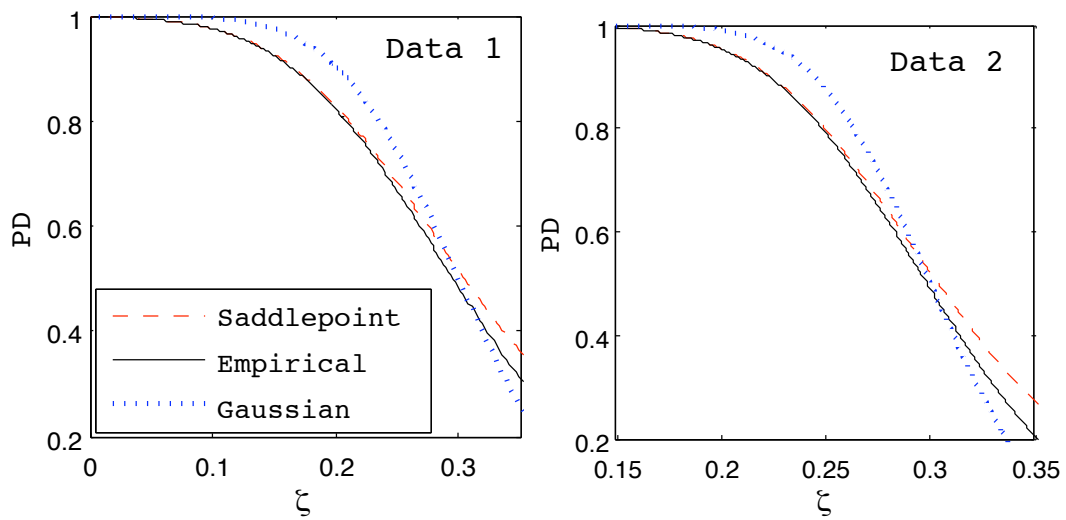


FIG. 6.5:  $PD(r)$  (pour la vraie position initiale  $r$ ) calculé par l'approximation du point selle et l'approximation gaussienne. Comparaison avec  $PD$  calculé à partir de la distribution empirique.

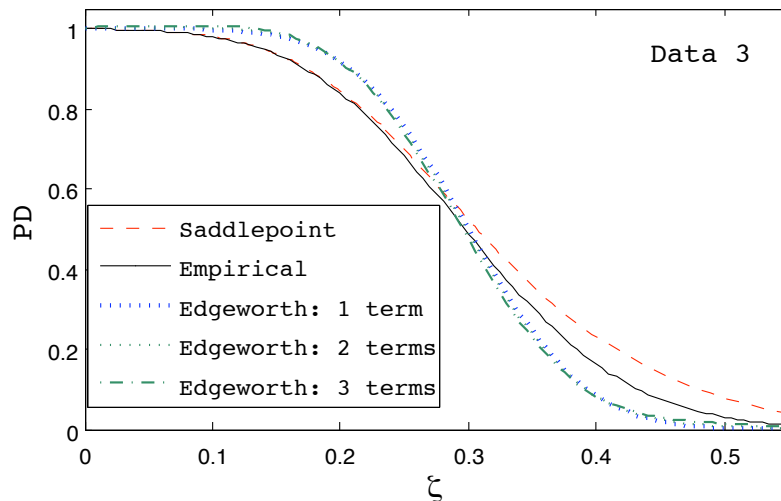


FIG. 6.6:  $PD(r)$  (pour la vraie position initiale  $r$ ) calculé par l'approximation du point selle et par le développement de Gram-Charlier tronqué après le premier terme (approximation gaussienne), après le second terme et après le troisième terme. Comparaison avec  $PD$  calculé à partir de la distribution empirique. Echelle linéaire de  $PD$ .

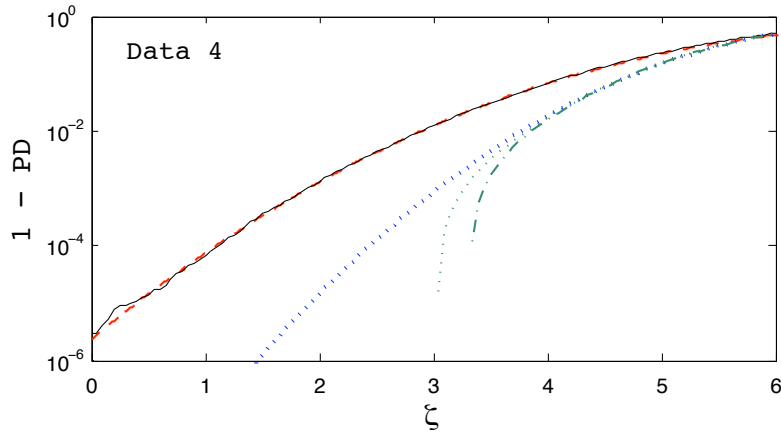


FIG. 6.7:  $PD(r)$  (pour la vraie position initiale  $r$ ) calculé par l'approximation du point selle et par le développement de Gram-Charlier tronqué après le premier terme (approximation gaussienne), après le second terme et après le troisième terme. Comparaison avec  $PD$  calculé à partir de la distribution empirique. Pour des valeurs extrêmes de  $PD$  :  $PD \approx 1$ .

distribution empirique est (dans le même sens que précédemment) beaucoup moins gaussien que  $\mathcal{P}(20)$ .

## Conclusion

L'objet de cette section était d'approximer analytiquement PFA et  $PD$  pour le modèle *speckle* décrit dans le chapitre précédent et pour la procédure de détection consistant à seuiller  $\hat{\alpha}(\mathbf{x}, r)$  dont l'expression a été définie dans la partie précédente. Pour cela, l'approximation du point selle pour une série tronquée au premier ordre de PFA et  $PD$  dont les expressions sont données respectivement par les équations (6.1) et (6.2) est appliquée aux MGF de  $\hat{\alpha}(\mathbf{x}, r)$  respectivement sous  $H_0$  et sous  $H_1$  dont les expressions sont données dans les équations (5.22) et (5.21). Cette approximation est comparée à des approximations classiques construites à partir d'un développement en série de Gram-Charlier ou d'Edgeworth.

Afin de mener cette étude, des simulations ont été effectuées à partir du modèle *speckle*. Les conditions et les différents cas de simulation (intensité de fond déterministe  $\mu$  différente, matrice de covariance de bruits post-AO  $\Sigma$  différente) ont été précisés au début de cette section. Ils aboutissent presque exactement aux mêmes conclusions, voir la figure 6.4 par exemple.

On constate pour le modèle *speckle* que l'approximation *saddlepoint* d'ordre 1 de la distribution cumulative de  $\hat{\alpha}$  ajuste parfaitement la distribution cumulative vraie de  $\hat{\alpha}$  pour la zone d'intérêt de PFA et pour une grande  $PD$ . En revanche, les développements d'Edgeworth et de Gram-Charlier, confondus entre eux pour un ordre de développement raisonnable, approximent très mal PFA et  $PD$ , comme ceci sera discuté sous un autre angle dans la prochaine section. On peut par ailleurs noter que l'expansion d'ordre 1 de la série de Gram-Charlier ou d'Edgeworth (approximation gaussienne) est moins mauvaise qu'une expansion faisant intervenir deux ordres supplémentaires de cette série, voir les figures 6.6 et 6.7).

## 6.3 Utilisation du modèle *speckle* en pratique

### Motivations pour ce modèle

Dans cette section, nous motivons et supposons donc ensuite que le modèle *speckle* de la section 5.1 pour le cas extrême  $N_t = 1$  décrit correctement les données et on s'interroge sur la perte accusée lorsque l'on suppose en fait le modèle gaussien comme correct.

- **Utiliser un temps d'observation moins long pour autant d'information**

Pour la plupart des modèles d'imagerie, comme celui de la partie I ou le modèle *speckle* de la section précédente, la somme de  $N_t$  images courte pose indépendantes est égale à l'image de temps de pose équivalent plus  $N_t - 1$  bruits de lecture gaussiens (voir App. C.3). On entend ici par l'égalité de ces deux images l'égalité de leurs distributions statistiques (même valeur en moyenne, même écart à la moyenne *etc*).

Ainsi, pour une durée totale d'observation donnée,  $KN_t$  images courte pose contiennent plus d'information que  $K$  images longue pose, au point que  $K$  images longue pose peuvent être approximativement déduites (au bruit de lecture près) des images courte pose. Utiliser de façon optimale des courtes poses plutôt que des longues poses permet ainsi de réduire le temps d'observation total pour une performance donnée.

Lorsque le temps de pose tend vers l'infini, pour le modèle *speckle* de la partie précédente par exemple, d'après le théorème central limite la distribution des données tend vers un modèle gaussien. Cependant, pour un temps de pose de quelques secondes le modèle gaussien ne modélise pas bien les données haute dynamique et haute résolution angulaire qui nous intéressent. Or pour pouvoir bénéficier du fait qu'une série de poses courtes contient plus d'information qu'une pose longue de durée correspondante, il convient de modéliser correctement la pose courte.

L'utilisation d'un modèle fin non asymptotique tel que 5.1 demande donc un effort supplémentaire de modélisation, de constructions d'inférences et de programmation, mais permet en principe de réduire le temps d'observation nécessaire. Ceci est d'autant plus vrai qu'en réduisant le temps d'observation on réduit du même coup les erreurs de modélisation liée à la variabilité des paramètres atmosphériques et instrumentaux supposés constants dans les modèles.

- **Erreur de plusieurs ordres de grandeur par un calcul de PFA par modèle gaussien**

Considérons maintenant le problème spécifique de qualité d'estimation de PFA. La PFA est en effet un des produits finaux les plus importants de la modélisation statistique des données pour un problème de détection. Rappelons que la PFA est la PFA associée au test partout discuté et utilisé dans cette thèse qui consiste à seuiller l'estimateur  $\hat{\alpha}(\mathbf{x}, r)$  habituel, toujours à une position initiale  $r$  fixée (section 3.4). Comme nous allons le discuter dans cette sous-section, supposer que la distribution de  $\hat{\alpha}(\mathbf{x}, r)$  est gaussienne peut mener à une très grosse erreur d'estimation de PFA pour des données en fait distribuées selon une distribution *speckle* de temps de pose égal au temps de corrélation haute fréquence. En effet, on constate sur la figure 6.4) que lorsque l'on croit fixer une PFA (par pixel puisque le test doit être appliqué pour chaque position possible  $r$ ) à  $10^{-6}$ , l'erreur d'estimation de PFA est ainsi de l'ordre de 1000 : 1000 fois plus de fausses alarmes que souhaitées sont enregistrées. Ce résultat est robuste à des changements des paramètres utilisés dans la simulation et menant à des performances illustrées dans la figure 6.3. Rappelons que l'on peut se faire une idée de la correspondance visuelle entre la PFA et l'allure de l'histogramme (*i.e.* de la densité) de  $\hat{\alpha}(\mathbf{x}, r)$  ayant mené à la PFA pour le cas "Data 1" grâce à l'histogramme en haut de la figure 6.2. Discutons plus en détail ce résultat.

Rappelons plus précisément que le résultat émane de simulations 1D effectuées à partir du modèle *speckle* décrit dans la section 5.1. 4 cas de simulations ont été considérés, avec  $N_t = 1$ ,  $K = 10$  ( $N_t$  et  $K$  sont faibles mais nous venons de mentionner que nous préférons économiser autant de temps d'observation que possible pour une performance donnée) et différentes paramétrisations du fond déterministe  $\boldsymbol{\mu}$  et de la covariance des bruits  $\Sigma$ . La PFA est alors calculée par méthode de Monte Carlo standard (génération aléatoire de jeux de données  $\{\mathbf{x}\}$ , calcul de  $\hat{\alpha}(\mathbf{x}, r)$  pour chaque  $\mathbf{x}$ , seuillage et comptage final du nombre de détections) sur chacun de ces 4 ensembles de simulations.

L'objet de l'étude était la comparaison de méthodes d'approximations (voir la section 6.2), mais comme l'ordre 0 des développements d'Edgeworth et de Gram-Charlier consiste en fait à approximer la PFA par une PFA sous modèle gaussien de la statistique considérée (ici  $\hat{\alpha}(\mathbf{x}, r)$ ), cet exemple permet



du même coup d’observer l’erreur d’approximation de la PFA lorsque l’on suppose que la distribution de  $\hat{\alpha}(\mathbf{x}, r)$  est gaussienne alors qu’elle est en fait donnée à partir du modèle *speckles* de la section 5.1.

L’erreur d’estimation de la PFA dépend de la valeur de PFA (gaussienne par exemple) considérée. Dans l’exemple, le test est appliqué “par pixel” (et en un seul pixel), non à tout le champ des positions possibles. D’après l’annexe C.2, pour des images de la taille de SPHERE ( $L \sim 10^4$ ) et pour une PFA globale  $PFA_{im} \sim 10^{-2}$  on conclut qu’il faudrait fixer une PFA de l’ordre de  $PFA_{pix} \sim 10^{-6}$ .

Utilisons alors la figure 6.4 comme on le ferait en pratique : on fixe  $PFA = 10^{-6}$  et on lit du graphe la valeur correspondante de seuil  $\zeta$  à injecter dans le test pour atteindre cette PFA. Si on suppose –à tort– que la distribution de  $\hat{\alpha}(\mathbf{x}, r)$  est gaussienne, la relation entre PFA et seuil est fautive. On lirait par exemple un seuil  $\zeta = 0.3$ , alors que pour un seuil  $\zeta = 0.3$  la PFA vraie est de l’ordre de  $10^{-3}$  au lieu des  $10^{-6}$  attendu. On ferait ainsi 1000 fois plus de fausses détections (“1000 fois plus de pixels qui s’allument à tort”) que ce à quoi on s’attendait en “fixant PFA à  $10^{-6}$ ”.

Pour des PFA plus faibles, le coeur de la distribution, bien approximée par une loi gaussienne, contribue encore moins à la valeur de la PFA, de sorte que l’approximation de la PFA devient de plus en plus mauvaise pour des PFA de plus en plus faibles. Ceci s’observe bien sur les simulations.

Notons de plus que même dans le cas d’un fond intense ( $\boldsymbol{\mu} = 100\boldsymbol{\mu}^{(0)}$ , et donc d’intensité  $d_\ell = |\mu_\ell|^2$  accrue d’un facteur  $10^4$ , où  $\boldsymbol{\mu}^{(0)}$  correspond au cas de simulation de référence “Data 1”), l’amélioration de la qualité d’estimation de la PFA par un modèle gaussien est peu visible sur la figure 6.4. Pourtant, l’approximation asymptotique de l’annexe de Smith et al. [2009] indique que pour une contribution déterministe du front d’onde qui devient très grande devant la contribution aléatoire (post-AO), la distribution des données standardisées tend vers une distribution gaussienne. Ce résultat de simulation semble indiquer soit que la non-standardisation rend le résultat asymptotique inapplicable soit que la vitesse de convergence est lente.

### • Optique adaptative extrême

Comme décrit par Aime and Soummer [2004] notamment, les rapports de Strehl qui devraient être régulièrement obtenus grâce à l’AO extrême utilisée pour SPHERE (voir Fusco et al. [2006], Allier-Carpentier et al. [2008]) feront apparaître des statistiques de bruit déterminées plus seulement essentiellement à partir des bruits résiduels post-AO mais aussi à partir du fond déterministe. Ceci exclut donc –pour une zone centrale plus ou moins large– une modélisation plus simple du champ d’intensité par une loi gamma multivariée, étudiée par exemple dans Chatelain et al. [2006b].

## Discussions de l’applicabilité du calcul de PFA à l’instrument SPHERE

L’approximation du *saddlepoint* évaluée analytiquement dans la section 6.1 est destinée à approximer la PFA du test décrit dans la section 3.4 sous le modèle *speckle* discuté dans la section 5.1. Dans les simulations introduites dans la section 6.2 il s’est avéré que cette approximation était excellente, et ce dès l’ordre 0 de l’expansion.

Cependant, la PFA était calculée en supposant les paramètres  $\boldsymbol{\mu}$  et  $\Sigma$  connus. Ceux-ci sont pourtant *a priori* inconnus en pratique. Dans le cas général ( $\boldsymbol{\mu} \neq \mathbf{0}$ ,  $\Sigma$  non diagonale ...) pour les estimer il est nécessaire à la fois de pouvoir utiliser un analyseur de surface d’onde destiné à la mesure indirecte de l’amplitude complexe dans le plan focal, et d’autre part de reprendre et développer des techniques d’estimation de  $\boldsymbol{\mu}$  et  $\Sigma$ .

Or, le problème d’estimation des aberrations dans le plan complexe est un problème récurrent en imagerie haute dynamique et à haute résolution angulaire et déjà très étudié et appliqué. Un dispositif de diversité de phase est ainsi prévu pour SPHERE pour la mesure des aberrations statiques  $\boldsymbol{\mu}$ . Un dispositif de diversité de phase classique mesure les intensités dans le plan focal et dans un plan légèrement défocalisé introduisant un déphasage connu, voir Gonsalves [1982]. En particulier, pour une distance de défocalisation faible, le terme de phase de la propagation de Fresnel peut être linéarisé, de sorte que l’amplitude complexe d’un point du plan défocalisé peut s’exprimer comme une combinaison linéaire (de

coefficients connus) des amplitudes complexes du plan focal. L'utilisation de tels dispositifs instrumentaux et la recherche de méthodes d'estimation liée au paramètre  $\boldsymbol{\mu}$  ont déjà été beaucoup étudiées dans la communauté de traitement du signal (Idier et al. [2005], Mugnier et al. [2009]).

Cependant, l'estimation d'une matrice de covariance est un problème à part entière, même lorsque celle-ci est à valeur dans  $\mathbb{R}^{L^2}$ . Pour une matrice de covariance d'un processus complexe, notons pour exemple que dans un problème d'estimation (conjointe) d'une matrice de covariance complexe pour un modèle quasiment identique au notre, Chatelain et al. [2007] supposent  $\boldsymbol{\mu}$  connu et utilisent des cliques et une paramétrisation de  $\Sigma$  qui permettent de réduire la dimensionnalité du problème. Le problème est déjà alors très compliqué. Une des grosses difficultés posée par l'estimation d'une matrice de covariance est son appartenance à l'espace des matrices définies positives. Nous n'aborderons par conséquent pas plus avant le problème d'estimation de  $\boldsymbol{\mu}$  et  $\Sigma$ . Il semblerait cependant très intéressant de poursuivre cette voie, dont une des applications directe est l'injection des estimées dans l'expression de l'approximation précise de la PFA présentée et discutée dans ce chapitre.



## Troisième partie

# Méthode bayésienne de test d'hypothèse simple et originale : étude théorique



# Table des Matières

---

<b>7</b>	<b>Introduction à la distribution <i>a posteriori</i> du rapport de vraisemblance (PLR)</b>	<b>121</b>
7.1	Approche de Neyman-Pearson . . . . .	121
	Hypothèse alternative : $\theta^*$ fixé mais inconnu, même sous l’approche bayésienne . . . . .	121
	Optimalité au sens de Neyman-Pearson : test UMP sur $\Theta_1$ . . . . .	122
	Optimalité du BF au sens de Neyman-Pearson ? . . . . .	123
	Développer un autre test utilisant l’ <i>a priori</i> $\pi_1$ . . . . .	125
7.2	Problématiques bayésiennes . . . . .	125
	Approche de Neyman-Pearson, théorie de la décision et approche bayésienne . . . . .	125
	Paradoxe de Lindley . . . . .	126
	Réconciliation des mesures d’évidence de $H_0$ fréquentiste ( <i>p-value</i> ) et bayésienne ( $\Pr(H_0 \mathbf{x})$ )	126
	Réconciliation des procédures de test fréquentiste et bayésienne . . . . .	127
	Nouvelles formes d’ <i>a priori</i> et de tests d’hypothèses bayésiens “objectifs” . . . . .	127
	Etude des <i>p-values</i> par la théorie de la décision et variables de divergence . . . . .	128
	Réconciliation <i>post-data</i> entre test d’hypothèses et intervalle de confiance . . . . .	128
	Nouvelles fonctions de coût pour les intervalles de confiance . . . . .	128
7.3	Etat de l’art sur la distribution cumulative <i>a posteriori</i> du rapport de vraisemblance (PLR)	128
	Définition du PLR . . . . .	128
	Suggestion de Dempster en 1974 (un résultat : $\text{PLR} = 1 - p_{\text{val}}$ en cas simple) . . . . .	129
	Travaux consécutifs d’Aitkin à partir de 1997 . . . . .	130
	Mesure d’évidence par le rapport de vraisemblance . . . . .	130
	<i>E-value</i> associée au FBST de Pereira et Stern à partir de 1999 . . . . .	131
<b>8</b>	<b>Quelques propriétés générales du PLR</b>	<b>133</b>
8.1	Etude de la loi <i>a posteriori</i> du rapport de vraisemblance . . . . .	133
	Expression analytique du PLR . . . . .	133
	Valeur de la borne inférieure du support du PLR . . . . .	134
	Discontinuité du PLR en le rapport de vraisemblance généralisé . . . . .	135
	Moments du PLR . . . . .	136
	$\text{PLR}(\mathbf{x}, \zeta)$ : une statistique synthétique et informative . . . . .	137
8.2	Remarques simples concernant le PLR en tant que test binaire . . . . .	137
	Quelques comparaisons pratiques entre le PLR et des tests usuels . . . . .	137
	Performances fréquentistes du PLR vu comme un détecteur . . . . .	138
8.3	Implémentation et optimisation fréquentiste numériques du PLR . . . . .	141
	Implémentation . . . . .	141
	Optimisation fréquentiste . . . . .	141

<b>9</b>	<b>Introduction à l'invariance de groupe appliquée aux statistiques</b>	<b>143</b>
9.1	Motivation pour l'utilisation de l'invariance de groupe en statistiques . . . . .	143
9.2	<i>A priori</i> de Haar . . . . .	144
	Mesures de Haar sur un groupe (topologique localement compact) . . . . .	144
	Mesures <i>a priori</i> de Haar induites sur $\Theta$ lorsque $\theta \rightarrow g\theta$ et $g \rightarrow g\theta$ sont bijectives . . . . .	146
	Propriétés des mesures <i>a priori</i> de Haar si le groupe est commutatif ou compact . . . . .	148
	Expression de la densité <i>a priori</i> de Haar dans un cas simple . . . . .	150
9.3	Famille de vraisemblances invariante sous un groupe de transformations . . . . .	152
<b>10</b>	<b>En cas invariant, équivalence entre le PLR et une <i>p-value</i> fréquentiste</b>	<b>155</b>
10.1	Théorème . . . . .	155
	Hypothèses simples . . . . .	155
	Hypothèses faisant appel à la notion d'orbites . . . . .	161
	Relation entre ce résultat et le théorème de Stein . . . . .	162
10.2	Conséquences de l'équivalence . . . . .	163
	Réconciliations entre fréquentiste et bayésien . . . . .	163
	Propriétés induites . . . . .	164
	Nouvelles <i>p-values</i> . . . . .	164

---

## Chapitre 7

# Introduction à la distribution *a posteriori* du rapport de vraisemblance (PLR)

### 7.1 Approche de Neyman-Pearson

**Hypothèse alternative :  $\theta^*$  fixé mais inconnu, même sous l'approche bayésienne**

Comme introduit dans la section 1.3, d'après les formulations standards de test d'hypothèses le test qui nous intéresse consiste à choisir parmi les hypothèses simple et composite suivante :

$$H_0 : \theta = \theta_0 \ ; \ H_1 : \theta \in \Theta_1 \tag{7.1}$$

avec  $\Theta_1$  et  $\theta_0 \notin \Theta_1$  préalablement fixés. Cette écriture de test d'hypothèses est la plus standard mais comme nous allons le voir peut être ambiguë si elle est lue sous un angle bayésien. A strictement parler, lorsque  $\theta$  est à valeurs discrètes, si on note  $n$  le nombre d'éléments de  $\Theta_1$  alors ce test d'hypothèses simple *vs* composite désigne le test d'hypothèses suivant :

$$H_0 : \theta = \theta_0 \ ; \ \left( H_1^{(1)} : \theta = \theta_1 \ \text{ou} \ H_1^{(2)} : \theta = \theta_2 \ \text{ou} \ \dots \ \text{ou} \ H_1^{(n)} : \theta = \theta_n \right)$$

où il s'agit alors de choisir entre l'hypothèse simple  $H_0$  et l'ensemble des hypothèses alternatives simples  $\{H_1^{(1)} \ \text{ou} \ H_1^{(2)} \ \text{ou} \ \dots \ \text{ou} \ H_1^{(n)}\}$ . En fréquentiste, le paramètre  $\theta$  sous-jacent au processus de génération des données décrit par la loi  $p(\mathbf{x}|\theta)$  a toujours une valeur fixée mais inconnue. Ainsi, implicitement le test précédent est équivalent à l'expression suivante, peu commune car elle nécessite des précisions textuelles :

$$H_0 : \theta = \theta_0 \ ; \ H_1' : \theta = \theta^* \tag{7.2}$$

avec  $\theta_0$  prédéterminé et  $\theta^* \in \Theta_1$  non prédéterminé. Par *prédéterminé* on entend “fixé préalablement par l'utilisateur du test” :  $\theta_0$  est fixé par le contexte, avec en détection  $\theta_0 = \mathbf{0}$ . Au contraire,  $\theta^*$  est déterministe, fixé, mais sa valeur précise n'est pas assignée, la seule contrainte étant  $\theta^* \neq \theta_0$  ; en détection  $\theta^*$  est “la” valeur du signal “sous-jacent à l'expérience” comme nous allons le discuter dans un instant.

Les expressions (7.2) et (7.1) peuvent être *considérées* comme équivalentes en fréquentiste mais pas en bayésien car en fréquentiste une hypothèse *indique* la problématique que l'on souhaite adresser alors que le paradigme bayésien permet de *mesurer* précisément l'hypothèse. En bayésien on associe en effet à  $\theta$  une mesure de probabilité *a priori*, de sorte que la mesure de probabilité *a posteriori* d'une hypothèse est bien définie, comme cela a été vu dans la section 1.3. En l'occurrence, pour les tests (7.2) et (7.1), la probabilité *a posteriori* des hypothèses alternatives définies en (7.2) et (7.1) sont respectivement données



par :

$$\Pr(H_1|\mathbf{x}) = \Pr(\boldsymbol{\theta} \in \Theta_1|\mathbf{x}) \neq 0 \quad (7.3)$$

$$\Pr(H_1'|\mathbf{x}) = \Pr(\boldsymbol{\theta} = \boldsymbol{\theta}^*|\mathbf{x}) = 0 \quad \text{si } \pi \text{ est continu sur } \Theta_1 \quad (7.4)$$

L'effet qui se produit ici pour  $H_1'$  a été décrit dans la section 1.3 pour  $H_0$  : lorsque l'on prend pour  $\Theta_0 \subset \Theta$  (définissant l'hypothèse nulle), l'espace  $\Theta_0 = \{\boldsymbol{\theta}_0\}$  de dimension inférieure à  $\Theta$ , un *a priori* continu sur  $\Theta$  implique  $\Pr(H_0|\mathbf{x}) = \Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_0|\mathbf{x}) = 0$ . Ceci constitue le problème central des tests d'hypothèses simple *vs* composite mis en évidence par exemple dans le paradoxe de Lindley introduit dans la section 1.3 et détaillé plus bas.

Pour résoudre le problème ou pour des raisons d'interprétation, certains considèrent qu'il n'existe pas de problème *réel* dans lequel on souhaite tester une hypothèse strictement simple, mais qu'éventuellement le cas de l'hypothèse simple peut être traité en tant que limite d'une hypothèse composite, voir Berger and Delampady [1987] par exemple. Ce n'est cependant pas notre point de vue car dès que l'on suppose qu'un paramètre *peut* prendre une valeur fixée sous-jacente (intensité d'une planète par exemple) alors on peut souhaiter d'une façon ou d'une autre tester le fait que le paramètre *soit égal* à cette valeur (ce qui n'arrive à strictement parler jamais pour un jeu de données de taille fini et doit donc être reformulé). Or sous l'hypothèse (7.2), le problème est encore plus important puisque la seule façon d'avoir les probabilités des hypothèses  $H_0$  et  $H_1$  non nulles est de prendre comme *a priori*s deux distributions de Dirac sous chacune des hypothèses.

Face au problème posé ici et selon le problème que l'on cherche à traiter, on peut essentiellement soit rejeter l'inférence "test d'hypothèses" dans son ensemble (approche adoptée par un certain nombre de bayésiens), soit rejeter l'expression (7.2) et choisir l'expression (7.1) (parce que cette dernière a une probabilité *a posteriori* qui a un sens clair et qui est non nulle ou parce que c'est effectivement l'hypothèse que l'on cherche à tester), soit chercher à reformuler le test d'hypothèses d'une façon qui réponde "mieux" à la problématique que l'on cherche à traiter.

En l'occurrence, nous estimons de façon générale qu'il est nécessaire que l'approche bayésienne soit en mesure d'effectuer un "test d'hypothèses" (incluant le problème de sélection d'un modèle) car c'est une question qui semble prendre du sens en pratique et que le fait d'inclure une incertitude ou une certaine mesure de probabilité sur  $\boldsymbol{\theta}$  nous semble en prendre aussi.

Par ailleurs, pour le test d'hypothèses qui nous intéresse (détection d'une exoplanète à côté d'une étoile donnée), comme nous l'avons écrit plus tôt il nous semble devoir traiter la problématique soulevée par (7.2). En effet, c'est la vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta}^*)$  qui décrit l'expérience reproductible associée au paramètre  $\boldsymbol{\theta}^*$  que l'on souhaite inférer. Ainsi, en imageant un nombre infini de fois l'étoile accompagnée de sa planète d'intensité  $\boldsymbol{\theta}^*$ , la proportion (fréquence) d'images  $\mathbf{x}$  tend vers la vraisemblance vraie  $p(\mathbf{x}|\boldsymbol{\theta}^*)$  dont on suppose toujours connaître la forme, au paramètre  $\boldsymbol{\theta}^*$  près. Mais sous une problématique *indiquée* par les hypothèses de (7.2), il nous faudra éventuellement chercher comment reformuler l'hypothèse pour que celle-ci soit pertinente en bayésien. Ainsi, nous verrons que si l'on souhaite exprimer notre outil de test à partir d'une forme de question similaire à (7.2), (7.1) etc, on définirait en fait une hypothèse  $H_0(\mathbf{x})$ .

Quoiqu'il en soit, (*comme les p-values*) il n'est pas nécessaire de définir une telle hypothèse pour que l'outil de test ait un sens. Le principal pour le moment est de bien comprendre que fondamentalement on ne souhaite pas lire directement dans notre cas (7.1) comme l'hypothèse à *mesurer* car la problématique est plutôt posée par (7.2). Notons cependant que jusqu'ici et par la suite nous utilisons tout de même l'expression (7.1) pour *indiquer* la problématique posée car cette notation est usuelle et recouvre souvent en pratique la même problématique que celle que nous adressons ici.

### Optimalité au sens de Neyman-Pearson : test UMP sur $\Theta_1$

Dans le test d'hypothèses simple *vs* composite bien connu réprécisé ci-dessus, l'approche de Neyman-Pearson introduite dans la section 1.3 et 1.2, consiste à choisir si possible le test qui maximise "PD" pour une PFA prédéterminée, fixée à la valeur notée  $\alpha$ .

Pour une valeur de  $\theta_1$  prédéterminée (test d'hypothèses simple *vs* simple classique), le lemme de Neyman-Pearson indique que

**Proposition 5** *Le test déterministe*

$$\text{Rejeter } H_0 \text{ si } LR(\mathbf{x}, \theta_1) = \frac{p(\mathbf{x}|\theta_0)}{p(\mathbf{x}|\theta_1)} \leq \zeta \quad (7.5)$$

$$\text{avec } \zeta \text{ tel que } PFA(LR, \zeta) = \alpha \quad (7.6)$$

maximise  $PD(T, \theta_1)$ , définie classiquement par l'Eq. (1.11), parmi tous les tests  $T$  de niveau  $\alpha$ .

Pour une valeur de  $\theta_1$  non prédéterminée mais fixée, comme dans notre problématique précisée dans la section précédente, sous l'approche de Neyman-Pearson on souhaiterait si possible utiliser un test  $T$  de niveau  $\alpha$  dont la probabilité de détection  $PD(T, \theta_1)$  soit plus élevée que  $PD(T', \theta_1)$  pour tous les autres tests  $T'$  (de niveau  $\alpha$ ) pour toutes les autres valeurs  $\theta_1$ . Un tel test  $T$  est dit uniformément plus performant (UMP), *ie* plus performant pour toutes les valeurs  $\theta_1$ .

Ainsi, dans le cas d'une hypothèse  $H_1 : \theta = \theta_1$  avec  $\theta_1 > \theta_0$ , on montre que pour une vraisemblance appartenant à la famille exponentielle il existe un test UMP, voir Lehmann and Romano [2005] par exemple. Cependant, il n'existe pas de test UMP pour tous les tests d'hypothèses ni pour toutes les vraisemblances. Par exemple, Pfanzagl [1968] montre, sous des conditions de régularité que nous présupposons ici partout, que pour ce test d'hypothèses ( $\theta_1 > \theta_0$ ), la seule vraisemblance pour laquelle il existe un test UMP parmi tous les tests de niveau  $\alpha$  (et quelle que soit la taille de l'échantillon, pour un jeu *i.i.d.* d'observations) est une vraisemblance appartenant à la famille exponentielle. De même, dans le cas très simple d'hypothèse simple *vs*  $H_1 : \theta \neq \theta_0$  (test dit *two-sided*), avec une vraisemblance normale de variance 1 on constate facilement<sup>1</sup> qu'il n'existe pas de test UMP.

La non-existence systématique d'un test UMP conduit ainsi parfois à restreindre l'ensemble des tests considérés dans l'optimisation. Une contrainte naturelle consiste par exemple à extraire de tout l'ensemble des tests les tests éventuellement excellents pour certaines valeurs de  $\theta_1$  mais de très mauvaise performance ailleurs. Plus spécifiquement, on appelle test UMPU un test  $T$  de niveau  $\alpha$  qui maximise  $PD(T', \theta_1)$  sur l'ensemble des tests  $T'$  non-biaisés *ie* dont  $PD(T', \theta_1)$  est supérieure ou égal à  $\alpha$  pour tout  $\theta_1$  et strictement supérieure à  $\alpha$  pour quelques  $\theta_1$ . Cette contrainte est assez naturelle puisque pour  $\alpha = 0.5$  un test pile ou face (ne dépendant donc pas des données  $\mathbf{x}$ ) vérifie  $PD(\text{pile ou face}, \theta_1) = \alpha$  quelle que soit la vraisemblance et notamment quel que soit  $\theta_1$ . Un test non-biaisé de niveau  $\alpha$  est donc un test qui est plus performant quel que soit  $\theta_1$  qu'un test aléatoire uniforme de niveau  $\alpha$ . Dans le cas *two-sided* considéré à la fin du dernier paragraphe, seuller  $|x|$  est par exemple un test UMPU.

## Optimalité du BF au sens de Neyman-Pearson ?

Nous nous apprêtons à intégrer au problème l'utilisation d'informations préalables concernant le paramètre testé  $\theta$ . Plus spécifiquement, suivant le paradigme bayésien introduit dans la section 1.2 nous souhaitons utiliser non seulement la vraisemblance  $p(\mathbf{x}|\theta)$  supposée modéliser nos données mais aussi une distribution de probabilité *a priori*  $\pi_1(\theta)$  sur l'intensité du paramètre  $\theta$  conditionnellement à l'hypothèse alternative  $H_1$ . Comme cela a aussi été brièvement introduit dans la section 1.2,  $\pi_1$  peut permettre d'injecter des informations *a priori* sur  $\theta$  effectivement connues avant l'expérience ou d'injecter la distribution qui apporte le moins d'*information* possible au problème (avec des traductions variées du concept d'information).

Pour le test d'hypothèses de l'Eq. (7.1), qui n'est cependant pas nécessairement celui qui nous intéresse comme nous l'avons discuté, le BF s'écrit :

$$\text{BF}(\mathbf{x}) = \frac{p(\mathbf{x}|H_0)}{p(\mathbf{x}|H_1)} = \frac{p(\mathbf{x}|\theta_0)}{\int_{\Theta} d\theta p(\mathbf{x}|\theta)\pi_1(\theta)}$$

<sup>1</sup>Pour cela on décompose l'étude de performance sur les deux zones définies dans le test *one-sided* précédent.

On peut souligner que

**Proposition 6** *Pour un niveau de PFA fixé, le BF maximise  $PD(T)$  parmi tous les tests  $T$  avec  $PD(T)$  définie comme la moyenne pondérée par  $\pi_1$  des  $PD(T, \boldsymbol{\theta})$  :*

$$PD(T) = \int d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) PD(T, \boldsymbol{\theta}) \quad (7.7)$$

où  $PD(T, \boldsymbol{\theta})$  est définie classiquement par l'Eq. (1.11).

Preuve : repartons de la forme générale du BF

$$BF(\mathbf{x}) = \frac{p(\mathbf{x}|H_0)}{p(\mathbf{x}|H_1)}$$

on constate que seuiller le BF consiste en le test de Neyman-Pearson pour le test d'hypothèses simple *vs* simple comparant la vraisemblance  $p(\mathbf{x}|H_0)$  à la vraisemblance  $p(\mathbf{x}|H_1)$ . Parmi tous les tests  $T$  de PFA fixée  $\alpha$ , si on note  $\mathcal{X}_1$  l'espace de réjection de  $H_0$  associé au test  $T$  et à  $\alpha$  (on se limite toujours dans les notations aux tests déterministes par simplicité), le BF maximise donc

$$PD(T) = \int_{\mathcal{X}_1} d\mathbf{x} p(\mathbf{x}|H_1)$$

Comme bien détaillé dans la section 1.3, sous le test d'hypothèses (7.1) entendu de façon strict,  $p(\mathbf{x}|H_1) = \int_{\Theta} d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta})$ . On obtient ainsi le résultat en intervertissant intégrale sur  $\mathbf{x}$  et intégrale sur  $\boldsymbol{\theta}$  :

$$PD(T) = \int_{\mathcal{X}_1} d\mathbf{x} \int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta}) = \int d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) \int_{\mathcal{X}_1} d\mathbf{x} p(\mathbf{x}|\boldsymbol{\theta}) = \int d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) PD(T, \boldsymbol{\theta})$$

En réalité, comme discuté plus haut et rediscuté ci-dessous, le test (7.1) ne représente pas précisément le test que nous souhaitons effectuer car  $p(\mathbf{x}|H_1) = \int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi_1(\boldsymbol{\theta})$  ne représente pas la vraisemblance de nos données. En revanche, la proposition précédente implique au moins bien pour notre problème que le BF est admissible (Lehmann and Romano [2005]) au sens de Neyman-Pearson pour notre problème.

**Proposition 7** *Le BF est admissible au sens de Neyman-Pearson pour le test (7.2).*

En effet, supposons au contraire que le BF est inadmissible et plaçons-nous à un niveau  $\alpha$  fixé. Par définition, "BF inadmissible" (au sens de Neyman-Pearson) signifie qu'il existe un autre test  $T$  tel que

$$\begin{aligned} \forall \boldsymbol{\theta} \in \Theta_1 \quad PD(T, \boldsymbol{\theta}) &\geq PD(BF, \boldsymbol{\theta}) \\ \text{et } \exists \boldsymbol{\theta} \in \Theta_1 \text{ tel que } PD(T, \boldsymbol{\theta}) &> PD(BF, \boldsymbol{\theta}) \end{aligned}$$

Dans ce cas on a directement et quel que soit  $\pi_1$  continu

$$\int d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) PD(T, \boldsymbol{\theta}) > \int d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}) PD(BF, \boldsymbol{\theta})$$

ce qui est impossible d'après la proposition 6.

## Développer un autre test utilisant l'*a priori* $\pi_1$

La proposition 6 est intéressante mais ne rend pas le BF optimal pour autant pour notre problème, l'optimalité au sens de Neyman-Pearson signifiant pour nous "UMP" sur  $\Theta_1$ . En effet, dans la vraisemblance intégrée  $p(\mathbf{x}|H_1)$  de la proposition 6,  $H_1$  est définie par l'Eq. (7.1) et n'est pas la vraisemblance ayant mené à nos données  $\mathbf{x}$ . Dans notre cas nous souhaitons en effet effectuer une inférence pour une étoile donnée, c'est-à-dire, s'il y en a une, pour une planète donnée. Il est alors facile et en général implicite d'interpréter la *fonction de vraisemblance de nos données* comme la limite quand le nombre de jeux de données tend vers l'infini d'une proportion d'images observées, celles-ci étant obtenues chaque fois pour la même expérience, supposée reproductible. La vraisemblance associée à cette définition est donc  $p(\mathbf{x}|\theta^*)$  avec  $\theta^*$  fixé mais inconnu et que l'on cherche à inférer. C'est pourquoi nous avons souligné qu'il pouvait être important de lever l'ambiguïté du test en notant plutôt le test d'hypothèses sous la forme de l'Eq. (7.2). Cependant, comme quoiqu'il en soit il ne convient pas de mesurer  $H'_1|\mathbf{x}$ , nous n'employons pas cette notation non plus. Cependant, c'est bien là l'idée sous-jacente de l'inférence à effectuer, de sorte qu'un test optimal au sens de Neyman-Pearson serait un test UMP sur  $\Theta_1$ .

La maximisation de la moyenne pondérée de  $PD(T, \theta)$  peut constituer un critère d'optimalité pour certains utilisateurs et constitue quoi qu'il en soit un atout du BF, impliquant même son admissibilité au sens de Neyman-Pearson, voir la proposition 7. Pourtant, ceci n'implique pas qu'il soit en général UMP. Ainsi, même sous l'approche de Neyman-Pearson fixant préalablement un niveau de PFA et cherchant si possible à maximiser PD, chercher de nouvelles structures de test éventuellement basées sur  $\pi$  se justifie car comme nous venons de le voir il n'existe pas toujours de test optimal au sens UMP ni même UMPU notamment.

En l'occurrence, le test que nous étudierons a certains intérêts vis-à-vis de l'approche de Neyman-Pearson. Il s'agit en fait d'une famille de tests paramétrée par deux *inputs*, qui ont chacun leur sens propre dans un cadre autre que celui de Neyman-Pearson. La famille de test est par conséquent optimisable au sens de Neyman-Pearson, contrairement aux tests usuels qui sont définis à partir d'un seul paramètre d'*input*, la valeur de cet *input* étant attribuée de façon unique par la valeur de PFA fixée. Par ailleurs, dans notre cas, nous constaterons que pour certaines valeurs de  $\theta_1$ ,  $PD(\text{BF}, \theta_1)$  peut être dépassé non seulement par le test optimal mais aussi par des paramétrisations naturelles du test, et ce pour toutes les valeurs de PFA d'intérêt.

## 7.2 Problématiques bayésiennes

### Approche de Neyman-Pearson, théorie de la décision et approche bayésienne

L'approche de Neyman-Pearson est très largement répandue mais ne constitue cependant pas l'unique approche possible de décision, même binaire. En théorie de la décision, on définit par exemple explicitement une fonction de coût, au même titre qu'une vraisemblance ou qu'un *a priori*. L'optimisation consiste alors à choisir la procédure qui minimise le risque résultant. On attribue alors une interprétation (et donc un seuil) au risque. Cette approche peut conduire à dériver un test optimal au sens de Neyman-Pearson, mais pour le problème le plus simple uniquement. Une introduction à la théorie bayésienne de la décision est donnée dans l'annexe B.

De même, les inférences bayésiennes classiques résultent de la théorie de la décision pour certaines fonctions de coût naturelles et sont par conséquent assorties de propriétés intéressantes, voir l'annexe B. On attribue cependant ici directement un sens au paradigme bayésien *via* le sens attribué à la mesure (en général une mesure de probabilité) sur  $\Theta$ , donnant ainsi un sens à la notion de probabilité *a posteriori* d'un paramètre.

La différence de point de vue peut apparaître par exemple lorsque l'on considère un sens possible de l'optimalité du BF au sens fréquentiste. Dans la "PD" maximisée par le BF ( $PD(T) = \int d\theta \pi_1(\theta) PD(T, \theta)$ ), l'*a priori*  $\pi_1$  joue en pratique le rôle de fonction de pondération de  $PD(T, \theta)$  et le fait qu'il s'agisse ou

non de la distribution de probabilité de  $\theta$  peut ne pas importer. Au contraire, le terme  $\Pr(\theta \in A|\mathbf{x})$  tient uniquement son sens de l'interprétation probabiliste accordée à  $\pi_1$ .

Ainsi, l'approche bayésienne classique appliquée au problème de test d'hypothèses consiste à sélectionner une certaine distribution *a priori* et à en dériver un test (classiquement, BF ou  $\Pr(H_0|\mathbf{x})$ , voir la section 1.3) et à prédéterminer le seuil non en fixant une PFA mais en interprétant directement le seuil de ces tests. Ceci peut être fait par chacun tout comme un choix de PFA. Sinon, des échelles standards telle que l'échelle dite de Jeffreys [1961, 1st ed. : 1939] permettent de décrire qualitativement des valeurs de BF.

Cependant, comme nous l'avons déjà vu le problème de test d'hypothèses vu sous l'angle bayésien amène encore des questions méthodologiques importantes auxquelles les méthodes classiques ne répondent pas de façon totalement satisfaisantes. Listons ici les problématiques auxquelles le test que nous allons étudier contribuent.

### Paradoxe de Lindley

Le paradoxe de Lindley [1957] est un des problèmes les plus apparents du BF et du POR. Il a été le sujet d'un grand nombre d'études et de discussions. Tsao [2006] est l'une des contributions les plus récentes à ce sujet. Reprenons l'exposé le plus bref et le plus simple de ce paradoxe.

Supposons que  $X \sim \mathcal{N}(\theta, 1)$  et que  $\pi_1$  représente un *a priori* diffus sur  $H_1$  à support convexe incluant 0 (test *two-sided*), en prenant par exemple  $\pi(\theta) \sim \mathcal{U}(-a/2, a/2)$ . Etudions alors le test d'hypothèses

$$H_0 : \theta = 0 \ ; \ H_1 : \theta \neq 0 \quad (7.8)$$

au sens direct de l'Eq. (7.1).

Pour un  $x$  fixé, la densité marginale de  $x$  sous  $H_1$  telle que définie dans le BF est donnée par

$$p(x|H_1) = \int d\theta \pi(\theta)p(x|\theta) = \frac{1}{a} \int_{-a/2}^{a/2} d\theta \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right) \quad (7.9)$$

Lorsque  $a \rightarrow \infty$ ,  $p(x|H_1) \rightarrow 0$  tandis que  $p(x|0)$  ne varie pas, de sorte que le BF accepte  $H_0$  quelle que soit la valeur de  $|x|$ , qui constitue pourtant le test UMPU comme nous l'avons dit plus haut.

Cet effet est conservé dans le POR, où on rappelle que le POR est le rapport des probabilités *a posteriori* des hypothèses (Eq. 1.25) de sorte que son application au test simple *vs* composite nécessite de définir un *a priori*  $\Pr(\theta = \theta_0) > 0$  même si sous  $H_1$   $\theta$  est continu, voir la section 1.3. Notons ainsi que la seule façon de compenser le côté "diffus" de l'*a priori* sous  $H_1$  ( $a \rightarrow \infty$ ) est de prendre  $\Pr(\theta = \theta_0) \rightarrow 0$ , c'est-à-dire de prendre  $\theta$  continu sur  $\mathbb{R}_+$  et non juste  $\mathbb{R}_{+*}$ .

Ainsi, même en excluant le test simple *vs* "simple mais inconnu" (7.2) qui nous intéresse, le BF et le POR se comportent encore mal sous le test simple *vs* composite de l'Eq. (7.1) : ils acceptent trop facilement  $H_0$ . Diverses interprétations de ce phénomène ont été proposées, remettant par exemple en question le sens d'une hypothèse simple ou d'un *a priori* diffus mais si l'on accepte le cas d'étude proposé, le BF et le POR ont un comportement difficilement acceptable.

Ce paradoxe est évité par des tests tels que le facteur de Bayes *a posteriori* de Aitkin [1991], le facteur de Bayes fractionnaire de O'Hagan [1995] ou le test de Bernardo [2002].

### Réconciliation des mesures d'évidence de $H_0$ fréquentiste (*p-value*) et bayésienne ( $\Pr(H_0|\mathbf{x})$ )

Pour étudier le paradoxe de Lindley on observe en général la *p-value* standard (notion introduite dans la section 1.3) en plus du BF et du POR abordés précédemment. En l'occurrence, dans le cas précédent la *p-value* est définie à partir du test UMPU consistant à seuiller  $|x|$ . Classiquement les études du paradoxe de Lindley consistent alors à choisir un cas d'étude et une classe d'*a priori* diffus, à donner les expressions du BF et/ou du POR et de la *p-value* et à discuter leur divergence.

Plus généralement, de nombreuses autres études abandonnent la contrainte d'*a priori* diffus et cherchent à concilier la *p-value* (en particulier lorsque celle-ci est définie de façon univoque) et  $\Pr(H_0|\mathbf{x})$ . En effet, ces deux statistiques reposent sur  $H_0$ , la première d'un point de vue fréquentiste et la seconde d'un point de vue bayésien. Rappelons en effet que la *p-value* est la probabilité sur  $\mathcal{X}$  sous  $H_0$  que les observations soient "plus extrêmes" que les observations obtenues en pratique, tandis que  $\Pr(H_0|\mathbf{x})$  est la mesure qui indique directement la probabilité de  $H_0$  pour des données  $\mathbf{x}$ . Ces deux quantités indiquent ainsi différemment si l'hypothèse  $H_0$  est *acceptable*, *évidente*, ou non.

Afin de concilier ces mesures d'évidence fréquentiste et bayésienne, pour différentes familles de vraisemblance et différents tests d'hypothèses (mais en restant en cas simple *vs* composite), l'*a priori* qui rend la *p-value* standard aussi proche que possible de  $\Pr(H_0|\mathbf{x})$  est recherchée parmi des familles d'*a priori*  $\pi_1$  plus ou moins générales, et/ou pour différents choix de  $\Pr(\boldsymbol{\theta} = \boldsymbol{\theta}_0)$ . Cette problématique a été discutée page 25. En un mot, en général la tendance est la même que dans le paradoxe de Lindley :

$$\Pr(H_0|\mathbf{x}) \geq p_{\text{val}}(\mathbf{x}) \quad (7.10)$$

En particulier, dans le cas *two-sided* il n'est pas possible (parmi toutes les familles d'*a priori* étudiées) de trouver un *a priori* qui permette de rendre ces deux quantités égales, voir Berger and Selke [1987] et références incluses.

Ceci soulève ainsi la question de savoir *quelles* mesures d'évidence bayésienne et/ou fréquentiste traitant du problème de test d'hypothèses peuvent se concilier de façon naturelle.

## Réconciliation des procédures de test fréquentiste et bayésienne

La réconciliation des *mesures d'évidence* fréquentiste (dite mesurée par la *p-value*) et bayésienne (dite mesurée par  $\Pr(H_0|\mathbf{x})$ ) n'est donc actuellement pas atteinte. Or connaître (et comprendre) les relations entre les approches fréquentiste et bayésienne est intéressante à divers égards : ces relations peuvent permettre de transposer les résultats obtenus sous une approche à l'autre approche, de pouvoir se rassurer en choisissant une approche sachant qu'en réalité elle est équivalente à l'autre, de finalement unifier les approches, *etc.*

Afin d'unifier les procédures de test fréquentiste et bayésienne, connaissant les problèmes classiques mentionnés ci-dessus il est donc nécessaire d'élargir le champ d'investigation. Une voie classique de rapprochement des concepts consiste à s'approprier par le paradigme bayésien des outils fréquentistes. L'outil le plus classique de cette famille est la *p-value* prédictive de Meng [1994], qui consiste à calculer la probabilité sur  $\mathcal{X}$  *via* des données prédites selon la distribution

$$p(\mathbf{x}^{\text{pred}}|\mathbf{x}) = \int d\boldsymbol{\theta} p(\mathbf{x}^{\text{pred}}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{x}) \quad (7.11)$$

Une autre consiste à modifier l'approche fréquentiste classique en effectuant les calculs de probabilité conditionnellement à la valeur de certaines statistiques, afin de se rapprocher du paradigme bayésien qui est lui totalement conditionné à  $\mathbf{x}$ . Pour le test d'hypothèses simple *vs* composite, Berger et al. [1997] effectuent un test binaire pour les deux procédures (bayésien avec une zone de non décision) puis unifient les notions de *probabilités d'erreur* vues sous les angles fréquentiste et bayésien en donnant d'un côté la PFA mais calculée conditionnellement à une certaine statistique  $S(\mathbf{x})$  ("PFA" =  $\Pr(\text{rejeter } H_0|S(\mathbf{x}), H_0)$ ) et de l'autre  $\Pr(H_0|\mathbf{x})$ .

## Nouvelles formes d'*a priori* et de tests d'hypothèses bayésiens "objectifs"

Par ailleurs, des recherches se poursuivent sur le mouvement bayésien dit "objectif" qui dérive des procédures bayésiennes à partir de la seule donnée de la vraisemblance et éventuellement de contraintes. En particulier, l'*a priori* dit de *référence* a été étudié dernièrement avec soin par Berger et al. [2009], tandis que Bernardo [2002] (élargissement par Bernardo [2010]) a construit, à partir de la théorie de la décision, de l'*a priori* de référence et de la divergence de Kullback-Leibler, un ensemble de procédures répondant aux inférences standards d'estimation ponctuelle, d'intervalle de confiance et de test d'hypothèses.

### Etude des *p-values* par la théorie de la décision et variables de divergence

Par ailleurs, comme souligné par Hwang et al. [1992] les études des *p-values* sous l'angle de la théorie de la décision sont encore très peu développées. Une problématique liée est la relative absence d'étude des variables dites de *divergence* (grandeur  $D(\mathbf{x}, \boldsymbol{\theta})$  servant à discriminer deux "situations") alors que celles-ci sont notamment à la base de la *p-value* prédictive de Meng [1994] et que cette dernière fait maintenant partie des outils standards en test d'hypothèses bayésien.

### Réconciliation *post-data* entre test d'hypothèses et intervalle de confiance

Il est connu que les probabilités *pre-data* (probabilité sur  $\mathcal{X}$  non conditionnelle à  $\mathbf{x}$ ) du rejet d'une hypothèse et d'un intervalle de confiance sont identiques. Il est aussi connu qu'une telle relation n'existe pas en *post-data* (probabilité sur  $\Theta$  conditionnelle à  $\mathbf{x}$ ). Goutis and Casella [1997] cherchent ainsi à les réconcilier, mais en partant toujours de  $\Pr(H_0|\mathbf{x})$ .

### Nouvelles fonctions de coût pour les intervalles de confiance

Comme souligné par Robert [2007] notamment, l'étude des fonctions de coût associées aux intervalles de confiance est encore relativement peu développée. Pourtant, ce type d'inférence constitue une méthode fondamentale pour associer une incertitude à une estimation ponctuelle. Cette problématique et des références associées sont introduites dans l'annexe B.3.

## 7.3 Etat de l'art sur la distribution cumulative *a posteriori* du rapport de vraisemblance (PLR)

### Définition du PLR

**Définition 6** La *distribution cumulative a posteriori du rapport de vraisemblance notée PLR* (PLR pour *Posterior cumulative distribution of the Likelihood Ratio*) est définie ici dans le cas simple  $\Theta_0 = \{\boldsymbol{\theta}_0\}$  vs composite seulement, et ici en l'absence de paramètre de nuisance<sup>2</sup>.

Le *rapport de vraisemblance LR* (LR pour *Likelihood Ratio*) est défini par

$$LR(\mathbf{x}, \boldsymbol{\theta}) = \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)}{p(\mathbf{x}|\boldsymbol{\theta})} \quad (7.12)$$

Sous l'approche bayésienne, sous l'hypothèse alternative  $H_1$ ,  $\boldsymbol{\theta}$  est une variable aléatoire décrite par une distribution *a posteriori* pour des données observées  $\mathbf{x}$ . Le rapport de vraisemblance  $LR(\mathbf{x}, \boldsymbol{\theta})$  est alors de même une variable aléatoire décrite par une distribution *a posteriori*. La distribution cumulative du LR est alors donnée par définition par

$$PLR(\mathbf{x}, \zeta) = \Pr(LR(\mathbf{x}, \boldsymbol{\theta}) \leq \zeta | \mathbf{x}) \quad (7.13)$$

$$= \int_{\{\boldsymbol{\theta}/p(\mathbf{x}|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}|\boldsymbol{\theta})\}} d\boldsymbol{\theta} \pi_1(\boldsymbol{\theta}|\mathbf{x}) \quad (7.14)$$

où PLR donne la probabilité *a posteriori* sous l'hypothèse alternative  $H_1$  pour que le rapport de vraisemblance soit inférieur à  $\zeta$  connaissant les données  $\mathbf{x}$ .

Le PLR est illustré sur la figure 12.6 p. 186 pour deux jeux de données  $\mathbf{x}$  issus de la simulation physique fine, l'un en présence d'une planète et l'autre sans.

<sup>2</sup>En pratique ils peuvent avoir été préalablement marginalisés.

A notre connaissance cet outil n'a été étudié en tant que tel que par Dempster [1974, 1997] et Aitkin [1997], Aitkin et al. [2005, 2009]<sup>3</sup>.

Notons cependant que nous avons découvert dernièrement la *surprise relative observée* (ORS) de Evans [1997] qui généralise le PLR en tenant compte d'une distribution sur  $\theta$  dite de référence qui n'est pas forcément égale à l'*a priori* et la *e-value* associée au *Full Bayesian Significance Test*, issu de Pereira and Stern [1999] et légèrement repris en particulier par Borges and Stern [2007], qui généralise le PLR en ajoutant –suivant l'ORS– une distribution de référence sur  $\theta$  et en traitant systématiquement le cas plus général où  $\Theta_0$  est de dimension inférieure à  $\Theta$  mais pas forcément de dimension nulle. Nous n'avons pas encore étudié complètement les travaux relatifs à l'ORS et au FBST mais il semble qu'en dehors des résultats listés dans la suite de cette section (qui présente l'état de l'art) les résultats obtenus dans cette thèse sont tous neufs.

### Suggestion de Dempster en 1974 (un résultat : $\text{PLR} = 1 - p_{\text{val}}$ en cas simple)

Dans le cadre du test d'hypothèses, A. Dempster a proposé en 1974 d'utiliser cette distribution cumulative, le PLR, comme une statistique à seuiller, voir Dempster [1974]. Dans un premier temps il reprend le test d'hypothèses simple *vs* simple ( $\{\theta_0\}$  *vs*  $\{\theta_1\}$ ). Dans ce cas, d'après le lemme de Neyman-Pearson (Lehmann and Romano [2005]), le test qui maximise la Probabilité de bonne Détection par rapport à la Probabilité de Fausse Alarme telles qu'introduites dans la section 1.2 est le test qui consiste à seuiller le rapport de vraisemblance :

$$\text{LR}(\mathbf{x}) = \frac{p(\mathbf{x}|\theta_0)}{p(\mathbf{x}|\theta_1)}$$

Rejeter  $H_0$  si  $\text{LR}(\mathbf{x}) < c$

Dempster s'intéresse alors au choix de la valeur de seuil  $c$ . La méthode de Neyman-Pearson consiste à fixer la PFA (ou erreur de type I) du test, définie par  $\alpha = \text{PFA}(c) = \Pr(\text{LR}(\mathbf{x}) < c | \theta_0)$ .  $c$  est alors obtenu en inversant la probabilité :  $c = \text{PFA}^{-1}(\alpha)$  avec une PFA typique de  $\alpha = 5\%$ . En bayésien il est usuel de fixer le POR (Eq. 1.25), *i.e.* la probabilité *a posteriori* limite pour l'hypothèse  $H_0$ , et ainsi d'en déduire  $c$ . Enfin, le test dit de vraisemblance opte pour une interprétation qu'il attribue à Fisher, dans lequel  $c$  est choisi directement à partir du rapport de vraisemblance exigé : “On rejette  $H_0$  si les données sont plus de 10 fois plus vraisemblables sous  $H_1$  que sous  $H_0$ .” qui induit  $c = 0.1$ . Pour le test d'hypothèses simple *vs* composite,  $\theta$  est inconnu sous l'hypothèse alternative. En tant que fonction de  $\theta$ ,  $\text{LR}(\mathbf{x}, \theta)$  est une variable aléatoire conditionnellement à  $\mathbf{x}$ . En bayésien il est alors possible de prendre la moyenne *a posteriori* de  $\text{LR}(\mathbf{x}, \theta)$  mais Dempster après une rapide analyse reposant sur des arguments théoriques “conceptuels” et sur l'interprétation d'exemples simples mais significatifs propose et opte pour l'utilisation du test que nous appellerons dans la suite “le test du PLR”, qui seuille la probabilité *a posteriori* que le rapport de vraisemblance soit inférieur à un seuil  $\zeta$ , seuil  $\zeta$  fixé selon la même interprétation que celle qu'il attribue à Fisher dans le cas simple *vs* simple :

$$\text{Rejeter } H_0 \text{ si } \text{PLR}(\mathbf{x}, \zeta) > p \tag{7.16}$$

<sup>3</sup>Ajoutons pour être tout à fait exhaustif la propriété suivante, remarquée par Newton and Raftery [1994] notamment, que pour un *a priori* propre (voir la section 8.1)

$$p(\mathbf{x}) = \left( \mathbb{E}[p(\mathbf{x}|\theta)^{-1} | \mathbf{x}] \right)^{-1} \tag{7.15}$$

Cette identité (identité de la moyenne harmonique) y est utilisée afin d'estimer une vraisemblance marginale par méthode de Monte Carlo à partir d'un échantillonnage de la loi *a posteriori*  $\pi(\theta|\mathbf{x})$ . Afin d'étudier les propriétés de convergence de cet estimateur et de proposer des estimateurs plus stables de la vraisemblance marginale, des études relatives aux moments *a posteriori* de fonctions de la vraisemblance  $p(\mathbf{x}|\theta)$  ont été menées, notamment par Raftery et al. [2007] et Calderhead and Girolami [2009]. Nous reprendrons des propriétés relatives aux moments du PLR mais nous attachons d'avantage à l'utilisation directe du PLR dans le cadre du test d'hypothèses et à l'étude générale de cette distribution.



La distribution cumulative PLR est donnée par l'équation (7.14). Citons ainsi Dempster [1974] (en adaptant ses notations) concernant sa préconisation quant au choix des valeurs de seuil  $\zeta$  et  $p$  : "The recommended approach is to reject  $H_0$  if the ratio  $p(\mathbf{x}|\boldsymbol{\theta}_1)/p(\mathbf{x}|\boldsymbol{\theta}_0)$  is reasonably sure to be large given knowledge of  $\mathbf{x}$ , where weight is put on values of the unknown  $\boldsymbol{\theta}_1$  in proportion to their plausibility given  $\mathbf{x}$ . For example, one might ask to be 0.90 sure that  $p(\mathbf{x}|\boldsymbol{\theta}_1)/p(\mathbf{x}|\boldsymbol{\theta}_0) \geq 100$ ."

### Travaux consécutifs d'Aitkin à partir de 1997

La distribution *a posteriori* du rapport de vraisemblance n'a ensuite été utilisée explicitement qu'en 1997 par M. Aitkin (Aitkin [1997]), afin de calibrer sur le cas normal (non-asymptotique puis en tant que limite asymptotique) la *p-value* standard, le facteur de Bayes *a posteriori* (Aitkin [1991]), le facteur de Bayes fractionnaire (FBF, O'Hagan [1995]) et le critère d'information de Akaike (AIC, Akaike [1973]) à partir du seuil  $\zeta$  du rapport de vraisemblance  $\text{LR}(\mathbf{x}, \boldsymbol{\theta})$  défini dans l'équation (7.12).

Le premier exemple qu'il considère est très proche de celui de Dempster : un cas normal scalaire *i.i.d.* simple de variance connue, avec un *a priori* uniforme sur la moyenne. Il constate, comme Dempster, que  $\text{PLR}(\mathbf{x}, 1)$  est simplement égal à 1 moins la *p-value* standard. Il retrouve ce même résultat sur un modèle quelconque mais développé asymptotiquement puis en prenant en compte  $\sigma$  comme paramètre de nuisance. Cet article a été l'occasion pour A. Dempster de republier son article de 1974 et de rédiger un commentaire supplémentaire (Dempster [1997]).

Aitkin et al. [2005] étendent l'étude de cette distribution en axant à nouveau une partie des cas qu'ils considèrent (exemple de modèle traité, approximations analytiques, prise en compte de paramètres de nuisance ...) de façon à démontrer la relation proche entre le PLR et la *p-value* standard. Dans les cas proches du cas gaussien avec un test sur la moyenne avec un *a priori* uniforme ils montrent à nouveau l'égalité  $\text{PLR}(\mathbf{x}, 1) = 1 - p_{\text{val}}(\mathbf{x})$ . Aitkin et al. [2009] utilisent finalement cet outil pour une application très spécifique qui inclut notamment un travail détaillé sur le modèle statistique.

Une des contributions de cette thèse est de démontrer que  $\text{PLR}(\mathbf{x}, 1) = 1 - p_{\text{val}}(\mathbf{x})$  est vrai pour toute famille de vraisemblances invariante en prenant l'*a priori* invariant associé et en choisissant comme statistique de test pour la *p-value* la vraisemblance pondérée par un terme traduisant la géométrie de la transformation de groupe. Le résultat est démontré si les espaces du problème (domaine d'appartenance d'une statistique exhaustive des données, espace des paramètres et espaces des transformations) sont isomorphes.

### Mesure d'évidence par le rapport de vraisemblance

Dempster [1974] propose l'utilisation du PLR en étendant l'utilisation du rapport de vraisemblance pour les tests de significativité en testant l'aire sous la queue d'une statistique. De nombreux autres auteurs ont travaillé sur le concept de vraisemblance : Fisher [1922], Birnbaum [1962], Dempster [1964], Fisher [1970, 1st ed. : 1925], Edwards [1972], Berger and Wolpert [1988], Hwang et al. [1992] ou Royall [1997] notamment. En particulier, Royall [1997] se focalise sur l'idée que la mesure d'une évidence est mesurée par le rapport de vraisemblance : il présente un paradigme basé sur la loi de la vraisemblance de Hacking (1965) qui affirme : "If hypothesis  $A$  implies that the probability that a random variable  $\mathbf{X}$  takes the value  $\mathbf{x}$  is  $p_A$ , while hypothesis  $B$  implies that the probability is  $p_B$ , then the observation  $\mathbf{X} = \mathbf{x}$  is evidence supporting  $A$  over  $B$  if  $p_A > p_B$ , and the likelihood ratio  $p_A/p_B$  measures the strength of that evidence.". Ceci suggère qu'il est possible d'attribuer une interprétation propre au seuil  $\zeta$  utilisé dans la définition 6 du PLR pour comparer une vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta}_1)$  possible sous l'hypothèse  $H_1$  à la vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta}_0)$  sous l'hypothèse  $H_0$ .

Dans leur article *Estimation of accuracy in testing* Hwang et al. [1992] étudient le problème de *test d'hypothèses* sous l'angle de la théorie de la décision (voir l'annexe B.4). L'inférence qu'ils souhaitent réaliser est "d'estimer la viabilité du domaine spécifié par  $H_0$ , c'est-à-dire d'estimer la fonction  $I_{\Theta_0}(\boldsymbol{\theta})$ " (la fonction indicatrice  $I_{\Theta_0}$  est définie dans le tableau de notations). Ils notent ainsi la décision  $d(\mathbf{x}) = \widehat{I_{\Theta_0}(\boldsymbol{\theta})}(\mathbf{x})$  et notent bien que " $d(\mathbf{x})$  ne mesure pas l'évidence dans un sens formel car ceci ne pourrait

être fait qu'au travers du rapport de vraisemblance" citant Birnbaum [1962] et Royall [1986], et ajoutant qu'ils font "la distinction importante de considérer  $d(\mathbf{x})$  comme une mesure de précision d'un test et non d'évidence".

La littérature concernant l'interprétation en termes de *mesure d'évidence* des méthodologies des tests de Neyman-Pearson, de Fisher, de la *p-value*, du facteur de Bayes, de la probabilité *a posteriori* de  $H_0$  etc est colossale et nous ne présenterons pas de synthèse ici. De nombreux ouvrages déjà référencés dans cette thèse permettent de s'introduire au sujet et d'y trouver un bon nombre de références aux articles et livres fondateurs.

### ***E-value* associée au FBST de Pereira et Stern à partir de 1999**

Nous reformulons ici des résultats qui ont en fait déjà été mis en évidence par Pereira and Stern [1999].

- **Invariance sous une transformation bijective de  $\mathbf{x}$**

Si l'on applique une transformation bijective  $S$  à  $\mathbf{x}$ , la vraisemblance de  $S(\mathbf{X})$  est égale à celle de  $\mathbf{X}$  multipliée par l'inverse du déterminant du Jacobien de  $S$ . Ce dernier se simplifie dans le rapport des vraisemblances seuilé ainsi que dans l'expression de la densité *a posteriori*. La valeur du PLR est donc invariante si l'on choisit de la calculer pour une fonction bijective de l'expression initiale des données  $\mathbf{x}$ .

- **Paradoxe de Lindley évité**

Le paradoxe de Lindley introduit dans la section 7.2 est évité ici puisqu'aucune densité marginale de  $\mathbf{x}$  n'est calculée.

Ceci se constate aisément sur l'exemple le plus simple considéré dans la section 7.2. Avec un *a priori* strictement uniforme sur  $\mathbb{R}$ , le PLR est parfaitement défini, contrairement au BF, et se réécrit comme

$$\text{PLR}(x, \zeta) = \Pr(\text{LR}(x, \theta) \leq \zeta | x) = \Pr((\theta - x)^2 \leq 2 \ln \zeta + x^2 | x) = \Pr(|X|^2 \leq 2 \ln \zeta + x^2 | x)$$

avec  $X \sim \mathcal{N}(0, 1)$

de sorte que l'espace d'intégration augmente avec  $|x|$  mais la distribution ne change pas, et donc  $\text{PLR}(x, \zeta)$  augmente bien avec  $|x|$ . En fait, pour  $\zeta = 1$  on vérifie même le résultat déjà connu de Dempster [1974]

$$\text{PLR}(x, 1) = \Pr(|X|^2 \leq x^2 | H_0) = 1 - p_{\text{val}}(|x|)$$

et que l'on généralisera dans le chapitre 10.



## Chapitre 8

# Quelques propriétés générales du PLR

Notons que la majorité des propriétés de la distribution *a posteriori* du rapport de vraisemblance (PLR) dérivées ci-dessous se réécrivent très simplement pour la distribution *a priori* du rapport de vraisemblance<sup>1</sup>. Cependant, nous nous focalisons dans cette thèse sur la distribution *a posteriori* du LR pour éviter toute confusion.

### 8.1 Etude de la loi *a posteriori* du rapport de vraisemblance

Les propriétés démontrées dans cette section sont en partie illustrées sur les figures 12.5 et 12.6 p. 186.

#### Expression analytique du PLR

L'expression analytique générale de la distribution *a posteriori* du LR (sous l'hypothèse alternative  $H_1$ , gardée implicite dans cette sous-section) est obtenue par définition à  $\mathbf{x}$  fixé. La dépendance de LR en  $\mathbf{x}$  n'intervient pas dans la transformation de variable effectuée sur l'espace  $\Theta$ , de sorte que  $\mathbf{x}$  est tenu implicite dans  $\text{LR}(\mathbf{x}, \boldsymbol{\theta})$  pour simplifier les notations. LR vu comme une fonction désigne ainsi la fonction  $\boldsymbol{\theta} \rightarrow \text{LR}(\mathbf{x}, \boldsymbol{\theta})$  et on note la *densité a posteriori* du LR évaluée en  $\zeta$  par  $p_{\text{LR}|\mathbf{x}}(\zeta)$  plutôt que  $p_{\text{LR}(\mathbf{x}, \boldsymbol{\theta})|\mathbf{x}}(\zeta)$  :

$$p_{\text{LR}|\mathbf{x}}(\zeta_0) = \left. \frac{d\text{PLR}(\mathbf{x}, \zeta)}{d\zeta} \right|_{\zeta=\zeta_0}$$

où le PLR est défini dans la définition 6. La densité *a posteriori* du LR est illustrée sur la figure 12.5 p. 186 et la distribution cumulative associée est illustrée sur la figure 12.6. Elles sont calculées pour deux jeux de données  $\mathbf{x}$  simulés finement, l'un en présence d'une planète l'autre sans.

Une technique classique pour obtenir la densité d'une fonction  $f(\mathbf{X}) \in \mathbb{R}^N$  à partir de la densité d'une variable aléatoire multivariée  $\mathbf{X} \in \mathbb{R}^M$  pour  $N \leq M$  consiste à définir une transformation  $\phi : \mathbb{R}^M \rightarrow \mathbb{R}^M$  qui à  $\mathbf{X}$  associe  $f(\mathbf{X})$  et autant de composantes de  $\mathbf{X}$  que nécessaire pour que  $\phi$  soit à valeur dans  $\mathbb{R}^M$ , à calculer la densité de cette nouvelle variable aléatoire par la règle de changement de variable classique faisant appel au Jacobien du changement de variable puis à marginaliser les composantes de  $\mathbf{X}$ .

Appliquer cette procédure pour obtenir la densité *a posteriori* de  $\text{LR}(\boldsymbol{\theta}) \in \mathbb{R}$  à partir de la densité *a posteriori* de  $\boldsymbol{\theta} \in \mathbb{R}^M$  consiste dans un premier temps à définir la transformation

$$\begin{aligned} \phi : \mathbb{R}^M &\rightarrow \mathbb{R}^M \\ \boldsymbol{\theta} &\rightarrow (\text{LR}, \check{\boldsymbol{\theta}}_1) \\ \text{où } \check{\boldsymbol{\theta}}_1 &= (\theta_2, \dots, \theta_M) \end{aligned}$$

---

<sup>1</sup>La distribution *a priori* d'une fonction de  $\mathbf{x}$  et  $\boldsymbol{\theta}$  a bien un sens si  $\mathbf{x}$  est considéré comme un paramètre fixé et non comme une variable aléatoire. En effet, la distribution *a posteriori* de  $\boldsymbol{\theta}$  conditionnelle au fait que la variable aléatoire  $\mathbf{X}$  est égale à une certaine valeur  $\mathbf{c}$  est définie quelle que soit la valeur  $\mathbf{c}$  et donc en particulier pour la valeur  $\mathbf{x}$  injectée dans la fonction  $\text{LR}(\mathbf{x}, \boldsymbol{\theta})$ .

Le déterminant de son Jacobien est  $|J| = |\partial \text{LR} / \partial \theta_1|$ . L'expression de la densité *a posteriori* de la variable aléatoire  $(\text{LR}, \check{\theta}_1)$  est alors obtenue grâce à la loi standard de transformation de variables :

$$p_{\text{LR}, \check{\theta}_1 | \mathbf{x}}(\zeta, \check{\mathbf{u}}_1) = \sum_{k=1}^{n(\zeta, \check{\theta}_1)} p_{\theta | \mathbf{x}}(\mathbf{u}^{(k)}) \left| \frac{\partial \text{LR}}{\partial \theta_1}(\mathbf{u}^{(k)}) \right|^{-1} \quad (8.1)$$

où  $p_{\theta | \mathbf{x}}(\mathbf{u})$  représente la densité *a posteriori* de  $\theta$  évaluée en  $\mathbf{u}$  (terme noté habituellement, en l'absence d'ambiguïté,  $\pi(\mathbf{u} | \mathbf{x})$ ), où les  $\mathbf{u}^{(k)}$  sont les solutions de  $\phi(\mathbf{u}^{(k)}) = (\zeta, \check{\mathbf{u}}_1)$  et  $n(\zeta, \check{\theta}_1)$  le nombre de solutions.

Pour  $M = 1$ , ceci donne directement la densité *a posteriori* du LR :

**Proposition 8** *Si  $\theta$  est réel et que la fonction  $\theta \rightarrow \text{LR}(\theta)$  est différentiable, la densité *a posteriori* du LR est donnée par*

$$p_{\text{LR} | \mathbf{x}}(\zeta) = \sum_{k=1}^{n(\zeta)} p_{\theta | \mathbf{x}}(u^{(k)}) \left| \frac{d \text{LR}}{d \theta}(u^{(k)}) \right|^{-1} \quad (8.2)$$

où les  $u^{(k)}$  sont les solutions de  $\phi(u^{(k)}) = \zeta$  et où la densité *a posteriori* de  $\theta$  est ici notée  $p_{\theta | \mathbf{x}}$  afin de distinguer la variable aléatoire de la valeur en laquelle est évaluée la densité.

Pour  $M > 1$ ,  $M - 1$  intégrations sont nécessaires pour marginaliser  $\check{\theta}_1$  :

$$\begin{aligned} p_{\text{LR} | \mathbf{x}}(\zeta) &= \int du_2 \dots \int du_M p_{\text{LR}, \check{\theta}_1 | \mathbf{x}}(\zeta, \check{\mathbf{u}}_1) \\ &= \int du_2 \dots \int du_M \sum_{k=1}^{n(\zeta, \check{\theta}_1)} p_{\theta | \mathbf{x}}(\mathbf{u}^{(k)}) \left| \frac{\partial \text{LR}}{\partial \theta_1}(\mathbf{u}^{(k)}) \right|^{-1} \end{aligned} \quad (8.3)$$

Ceci ne se simplifie pas en général.

Notons que la démarche utilisée ici consiste à utiliser la “loi de changement de variable standard” qui fait passer de la densité d’une variable aléatoire à la densité d’une autre variable aléatoire *de même taille* que la première, grâce au Jacobien usuel d’un changement de variable. Ici, si  $M > 1$  il est donc nécessaire de passer par une fonction intermédiaire,  $\phi : \theta \rightarrow (\text{LR}, \check{\theta}_1)$  par exemple, suivie d’une marginalisation sur  $\check{\theta}_1$ . L’utilisation d’une variable intermédiaire  $\check{\theta}_1$  brise temporairement la symétrie de la fonction  $\theta \rightarrow \text{LR}(\theta)$  où toutes les composantes de  $\theta$  jouent *a priori* le même rôle. L’utilisation de la mesure de Hausdorff par exemple permet d’éviter une telle brisure de symétrie en considérant directement la fonction  $\theta \rightarrow \text{LR}(\theta)$  (voir le théorème fort de la coaire, Federer [1969]), mais ceci fait appel à la notion d’espace rectifiable et au Jacobien associé qui ne font pas partie des outils de base en statistiques.

## Valeur de la borne inférieure du support du PLR

Dans l’approche fréquentiste, pour obtenir une statistique (fonction des données non fonction de paramètres inconnus), en général dans le rapport de vraisemblance le paramètre inconnu  $\theta$  est remplacé par la valeur de  $\theta$  qui maximise la vraisemblance. On appelle cette statistique, introduite dans la section 1.3, le rapport de vraisemblance généralisé<sup>2</sup> (GLR) :

$$\text{GLR}(\mathbf{x}) = \frac{p(\mathbf{x} | \theta_0)}{\max_{\theta \in \Theta_1} p(\mathbf{x} | \theta)} = \text{LR}(\mathbf{x}, \hat{\theta}_{\text{ML}}(\mathbf{x})) \quad (8.4)$$

<sup>2</sup>A strictement parler il faudrait remplacer le *maximum* de vraisemblance par la *borne supérieure* de vraisemblance car nous nous intéresserons au cas courant où l’espace des paramètres  $\Theta = \{\theta_0\} \cup \Theta_1$  est convexe, de sorte que  $\Theta_1$  est nécessairement ouvert. Or si la fonction  $\theta \rightarrow p(\mathbf{x} | \theta)$  est continue sur  $\Theta$  et si  $p(\mathbf{x} | \theta_0) \geq p(\mathbf{x} | \theta)$  quel que soit  $\theta \in \Theta$  alors  $\sup_{\theta \in \Theta_1} p(\mathbf{x} | \theta)$  est bien défini (et égal à  $p(\mathbf{x} | \theta_0)$ ) alors qu’à strictement parler  $\max_{\theta \in \Theta_1} p(\mathbf{x} | \theta)$  n’est pas défini. Un espace des paramètres  $\Theta$  est par exemple convexe pour le problème de détection de signal où  $H_0$  est définie par  $\theta_0 = \mathbf{0}$  et où sous  $H_1$  toutes les composantes de  $\theta$  sont réelles non nulles ou strictement positives. Nous ne rentrons pas ici dans ce niveau de précision.

Cette grandeur fréquentiste très classique joue aussi un rôle important dans la distribution *a posteriori* du LR :

**Proposition 9** *La borne inférieure du support du PLR est égal au GLR :*

$$\inf_{\zeta} \{\zeta : PLR(\mathbf{x}, \zeta) > 0\} = GLR(\mathbf{x}) \quad (8.5)$$

Ce résultat découle directement de la définition du GLR et se réécrit

$$\begin{aligned} PLR(\mathbf{x}, GLR(\mathbf{x})) &= 0 \\ PLR(\mathbf{x}, GLR(\mathbf{x}) + \epsilon) &> 0 \quad \forall \epsilon > 0 \end{aligned}$$

### Discontinuité du PLR en le rapport de vraisemblance généralisé

Par ailleurs, si la vraisemblance est “plate” en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})$ , le PLR est *discontinu* en le GLR. La vraisemblance doit être d’autant plus “plate” que  $M$ , la taille du vecteur  $\boldsymbol{\theta}$ , s’accroît. S’il existe plusieurs valeurs de  $\boldsymbol{\theta}$  qui maximisent la vraisemblance il suffit qu’il en existe une telle que la vraisemblance y soit “plate”. On notera  $\hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})$  cette valeur.

Plus précisément, on démontre dans l’annexe C.4 le résultat suivant :

**Théorème 2** *Sous une condition de régularité de la vraisemblance en  $\hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})$  telle que celle précisée ci-dessous, la fonction  $\zeta \rightarrow PLR(\mathbf{x}, \zeta)$  a une dérivée infinie quand  $\zeta \rightarrow GLR(\mathbf{x})^+$ . Une telle condition de régularité peut consister à supposer qu’il existe  $(\gamma_1, \dots, \gamma_M) \in \mathbb{R}_{+*}^M$ , tels que*

$$\sum_{m=1}^M \frac{1}{\gamma_m} < 1 \quad (8.6)$$

et  $(\beta_1, \dots, \beta_M) \in \mathbb{R}_{+*}^M$  tels que pour tout  $\boldsymbol{\theta}$  suffisamment proche de  $\hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})$

$$GLR(\mathbf{x}) \leq LR(\mathbf{x}, \boldsymbol{\theta}) \leq GLR(\mathbf{x}) + \sum_{m=1}^M \beta_m |\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})|_m^{\gamma_m} \quad (8.7)$$

On peut réécrire la condition de façon plus restrictive en imposant l’égalité des  $\gamma_m$  à une valeur  $\gamma$ , de sorte que la contrainte (C.16) se réécrit  $\gamma > M$ . En restreignant encore la condition en fixant  $\gamma = M + 1$ , on constate que le théorème s’applique sous la condition qu’il existe  $(\beta_1, \dots, \beta_M) \in \mathbb{R}_{+*}^M$  tels que pour tout  $\boldsymbol{\theta}$  suffisamment proche de  $\hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})$

$$GLR(\mathbf{x}) \leq LR(\mathbf{x}, \boldsymbol{\theta}) \leq GLR(\mathbf{x}) + \sum_{m=1}^M \beta_m |\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})|_m^{M+1}$$

Notons que comme pour la section précédente (section 8.1) on se place à  $\mathbf{x}$  fixé donc nous omettons à nouveau  $\mathbf{x}$  pour simplifier les notations. Le résultat énoncé par le théorème est évident en cas scalaire : si la vraisemblance est différentiable et que sa dérivée s’annule en  $\hat{\boldsymbol{\theta}}_{ML}$  alors d’après l’Eq. 8.2 la densité *a posteriori* du LR diverge en  $LR(\hat{\boldsymbol{\theta}}_{ML})$ . Le résultat est plus délicat à démontrer en cas multivarié. La condition consiste en une condition sur le développement limité de la fonction  $\boldsymbol{\theta} \rightarrow LR(\boldsymbol{\theta})$  en  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{ML}$ . La fonction doit être d’autant plus plate que le nombre d’intégrations impliqué dans la marginalisation explicitée dans l’équation (8.3) est élevé de sorte qu’en moyenne les ordres partiels  $\gamma_m$  doivent être d’autant plus grands que  $M$  est grand. Une démonstration ne faisant appel à aucune notion complexe est donnée dans l’annexe C.4.

Cette propriété supplémentaire permet par ailleurs d’estimer la valeur du GLR de façon assez précise comme on peut le voir sur la figure 12.6 p. 186. Échantillonner la distribution *a posteriori* sous une distribution *a priori* quelconque permet ainsi d’estimer la valeur fréquentiste du GLR.

## Moments du PLR

Le résultat suivant indique la relation forte existant entre le facteur de Bayes fractionnaire (FBF, introduit dans la section 1.3) et la distribution *a posteriori*<sup>3</sup> du LR.

**Théorème 3** *Si l'a priori  $\pi_1$  est propre, les moments fractionnaires a posteriori de LR sont égaux à FBF :*

$$\mathbb{E}_1[LR(\mathbf{x}, \boldsymbol{\theta})^b | \mathbf{x}] = \overline{FBF(\mathbf{x}, 1 - b)} \quad b \in \mathbb{R} \quad (8.8)$$

En particulier, pour  $b = 1$  on constate que le BF est égal à la moyenne a posteriori du LR :

$$\mathbb{E}_1[LR(\mathbf{x}, \boldsymbol{\theta}) | \mathbf{x}] = BF(\mathbf{x}) \quad (8.9)$$

Preuve :

$$\begin{aligned} \mathbb{E}_1[LR(\mathbf{x}, \boldsymbol{\theta})^b | \mathbf{x}] &= \int d\boldsymbol{\theta} \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)^b}{p(\mathbf{x}|\boldsymbol{\theta})^b} \pi_1(\boldsymbol{\theta} | \mathbf{x}) \\ &= p(\mathbf{x}|\boldsymbol{\theta}_0)^b \frac{\int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})^{1-b} \pi_1(\boldsymbol{\theta})}{p_1(\mathbf{x})} \\ &= BF(\mathbf{x}) \frac{\int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})^{1-b} \pi_1(\boldsymbol{\theta})}{p(\mathbf{x}|\boldsymbol{\theta}_0)^{1-b}} \\ &= BF(\mathbf{x}) \left( \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)^{1-b}}{\int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})^{1-b} \pi_1(\boldsymbol{\theta})} \right)^{-1} \\ &= \overline{FBF(\mathbf{x}, 1 - b)} \end{aligned}$$

Dans le cadre du test d'hypothèse, ce résultat –ainsi que d'autres relations du même type– a été remarqué à quelques reprises de façon indépendante<sup>4</sup>. Notons en particulier la contribution de Nicolae et al. [2008], l'un des auteurs de cet article étant le fondateur de la *p-value* prédictive (Meng [1994]).

Hors du cadre du test d'hypothèses, *i.e.* sans mention d'une valeur  $\boldsymbol{\theta}_0$  de référence, le résultat spécifique concernant le BF (Eq. 8.9) se réécrit

$$\mathbb{E}[p(\mathbf{x}|\boldsymbol{\theta})^{-1} | \mathbf{x}] = \left( \int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \right)^{-1} = p(\mathbf{x})^{-1}$$

Ce résultat permet d'estimer la vraisemblance marginale  $p(\mathbf{x})$  en échantillonnant la distribution *a posteriori*  $\pi(\boldsymbol{\theta} | \mathbf{x})$  et en calculant la moyenne harmonique de la vraisemblance non marginale. Ce résultat a été "découvert" de façon indépendante à quelques reprises, la découverte la plus connue étant semble-t-il celle de Newton and Raftery [1994]. Depuis, cette méthode d'estimation de la vraisemblance marginale par échantillonnage a été étudiée, utilisée et améliorée notamment par Chib [1995], Chib and Jeliazkhov [2001], Raftery et al. [2007] et Calderhead and Girolami [2009].

<sup>3</sup>Notons que contrairement aux autres propriétés dérivées dans cette section, cette propriété n'est pas conservée si l'on considère la distribution *a priori* du LR plutôt que sa distribution *a posteriori*. De façon équivalente, on note que cette relation n'est pas conservée si l'on inverse les hypothèses  $H_0$  et  $H_1$ , *i.e.* si l'on considère  $LR(\mathbf{x}, \boldsymbol{\theta})^{-1}$  plutôt que  $LR(\mathbf{x}, \boldsymbol{\theta})$ . En conservant les notations précédentes, le BF est alors directement obtenu à partir de la moyenne *a priori* du LR :

$$\mathbb{E}_1[LR(\mathbf{x}, \boldsymbol{\theta})^{-1}] = BF(\mathbf{x})^{-1}$$

et plus généralement on observe, par le même type de démonstration que celle du théorème qui suit, que :

$$\mathbb{E}_1[LR(\mathbf{x}, \boldsymbol{\theta})^{-b}] = \overline{FBF(\mathbf{x}, b)} BF(\mathbf{x})^{-1}$$

<sup>4</sup>par nous y compris

### PLR( $\mathbf{x}, \zeta$ ) : une statistique synthétique et informative

D'après les propriétés démontrées dans cette section, pour un jeu de données  $\mathbf{x}$ , les valeurs de la distribution cumulative *a posteriori* du LR, donnée par PLR( $\mathbf{x}, \zeta$ ), pour toutes les valeurs de  $\zeta \in \mathbb{R}_+$ , contiennent l'information nécessaire à la fois au calcul du GLR si  $\Theta$  est convexe et que la vraisemblance est continue sur  $\Theta$  et au calcul du BF si l'*a priori* est propre : le GLR est égale à la valeur de  $\zeta$  la plus faible telle que PLR( $\mathbf{x}, \zeta$ )  $> 0$  et le BF est égal à la moyenne *a posteriori* du LR, c'est-à-dire à la moyenne associée au PLR. Les valeurs de PLR( $\mathbf{x}, \zeta$ ) pour toutes les valeurs  $\zeta \in \mathbb{R}_+$  sont de plus intéressantes en tant que telles par une interprétation directe des valeurs des seuils  $\zeta$  et  $p$  comme ceci a été souligné dans la section 7.3.

## 8.2 Remarques simples concernant le PLR en tant que test binaire

Comme suggéré par Dempster [1974] et introduit dans l'équation (7.16), le PLR peut être utilisé afin de mesurer l'*évidence* de  $H_0$  mais aussi plus restrictivement, de même qu'une *p-value*, comme statistique de test binaire :

$$\text{Rejeter } H_0 \text{ si } \text{PLR}(\mathbf{x}, \zeta) > p \quad (8.10)$$

### Quelques comparaisons pratiques entre le PLR et des tests usuels

- **Equivalence entre PLR et GLR**

D'après la proposition 9, on constate que le test binaire associé au GLR, statistique de test dont l'expression est rappelée par l'équation (8.4), est équivalent à celui associé au PLR pour un seuil  $p = 0$  pour le PLR :

$$\text{PLR}(\mathbf{x}, \zeta) > 0 \Leftrightarrow \text{GLR}(\mathbf{x}) \leq \zeta \quad (8.11)$$

En effet, en considérant les événements complémentaires on constate que

$$\text{PLR}(\mathbf{x}, \zeta) = 0 \Leftrightarrow \Pr(\text{LR}(\mathbf{x}, \boldsymbol{\theta}) > \zeta | \mathbf{x}) = 1 \Leftrightarrow \text{LR}(\mathbf{x}, \boldsymbol{\theta}) > \zeta \forall \boldsymbol{\theta} \Leftrightarrow \text{GLR}(\mathbf{x}) > \zeta$$

- **Difficulté d'interprétation directe du seuil du GLR**

Le GLR mesure le rapport des vraisemblances les plus élevées sous chacune des deux hypothèses (ou plus exactement le rapport des bornes supérieures des vraisemblances, voir la note de bas de page associée à l'équation (8.4)). Or, si deux ensembles  $\Theta_a$  et  $\Theta_b$  de valeurs possibles pour  $\boldsymbol{\theta}$  sous une hypothèse mènent à la même valeur maximale de  $p(\mathbf{x}|\boldsymbol{\theta})$  alors le GLR résultant est identique que l'hypothèse soit définie par  $\Theta_a$  ou  $\Theta_b$ . Autrement dit, le GLR ne dépend de l'espace des paramètres sous une hypothèse qu'à travers la phase d'ordonnement des vraisemblances, qui est suivie par la sélection d'une *unique valeur* de  $\boldsymbol{\theta}$  : la valeur qui maximise la vraisemblance.

En particulier, pour un test d'hypothèses simple *vs* une hypothèse composite si l'espace des paramètres global  $\Theta$  est convexe et que la vraisemblance est continue sur  $\Theta$ ,  $\max_{\boldsymbol{\theta} \in \Theta_1} p(\mathbf{x}|\boldsymbol{\theta}) \geq p(\mathbf{x}|\boldsymbol{\theta}_0)$  ce qui implique directement

$$\text{GLR}(\mathbf{x}) \leq 1 \quad (8.12)$$

Pour un test binaire basé sur le GLR, le seuil  $\zeta$  doit donc *forcément être inférieur à 1*. En pratique, le choix de  $\zeta$  serait très difficilement effectué en attribuant une interprétation à la valeur du seuil  $\zeta$ . Un exemple des valeurs prises en pratique par le GLR est présenté dans la partie suivante, dans la sous-section p.184. Le seuil est ainsi, et/ou quoiqu'il en soit, systématiquement fixé par l'approche de Neyman-Pearson consistant à fixer une PFA et calculer le seuil  $\zeta$  associé.



Au contraire, dans les tests du BF et du PLR toutes les valeurs possible de  $\theta$  sont intégralement prises en compte à travers une intégration sur la loi *a priori*. Certaines valeurs de  $\theta$  rendront les données  $\mathbf{x}$  plus vraisemblables que sous  $\theta_0$ , d'autres les rendront moins vraisemblables. Pour le BF il s'agit alors de moyenniser la probabilité jointe de  $\mathbf{x}$  et  $\theta$  sur toutes les valeurs de  $\theta$  alors que pour le PLR il s'agit de calculer la probabilité *a posteriori* pour que les données  $\mathbf{x}$  soient plus vraisemblables sous un certain  $\theta \neq \theta_0$  que sous  $\theta_0$ . En particulier, il est possible de choisir  $\zeta = 1$  comme seuil dans les tests bayésiens, contrairement au GLR.

Par ailleurs, le fait que le GLR soit inférieur à 1 implique d'après la proposition 9 que  $\text{PLR}(\mathbf{x}, \zeta)$  est nécessairement strictement positif dès que  $\zeta > 1$ . Le théorème 2 indique de plus que la dérivée de  $\text{PLR}(\mathbf{x}, \zeta)$  tend vers l'infini lorsque  $\zeta \rightarrow \text{GLR}(\mathbf{x})^+$ , et de façon d'autant plus prononcée que la fonction  $\theta \rightarrow \text{LR}(\mathbf{x}, \theta)$  est "plate" en  $\hat{\theta}_{\text{ML}}(\mathbf{x})$ . A moins que la distribution *a priori* de  $\theta$  exclut de façon prononcée la valeur  $\hat{\theta}_{\text{ML}}(\mathbf{x})$  on peut par conséquent s'attendre à ce que la valeur  $\text{PLR}(\mathbf{x}, \zeta)$  soit *relativement élevée* pour  $\zeta > 1$ .

### • Equivalence entre PLR et BF

Si  $\mathbf{X}$  est une variable aléatoire continue, pour un jeu de données fixé  $\mathbf{x}$  si on fixe le seuil  $\alpha(\mathbf{x})$  du PLR à la probabilité *a posteriori* que le rapport de vraisemblance  $\text{LR}(\mathbf{x}, \theta)$  soit inférieur au rapport des vraisemblances marginales (BF), alors le test seuillant PLR par  $\alpha(\mathbf{x})$  est équivalent au test seuillant BF par  $\zeta$  :

$$\text{Soit } \alpha(\mathbf{x}) = \Pr\left(\text{LR}(\mathbf{x}, \theta) \leq \text{BF}(\mathbf{x}) \mid \mathbf{x}\right). \quad (8.13)$$

$$\text{PLR}(\mathbf{x}, \zeta) > \alpha(\mathbf{x}) \Leftrightarrow \text{BF}(\mathbf{x}) \leq \zeta \quad (8.14)$$

Ceci découle d'une propriété générale des distributions cumulatives continues. On a de ce fait la même équivalence en remplaçant  $\text{BF}(\mathbf{x})$  par  $\text{FBF}(\mathbf{x}, b)$  par exemple.

Cette équivalence permet d'observer au cas par cas comment le choix de la paramétrisation du PLR (ou respectivement du BF) est relié au choix de la paramétrisation du BF (ou respectivement du PLR).

### • PLR, BF et *a priori* impropre

Contrairement au facteur de Bayes, le PLR est défini de façon unique pour un *a priori* impropre, tant que la distribution *a posteriori* de  $\theta$  définit bien une probabilité<sup>5</sup>. Ceci permet de choisir les seuils  $\zeta$  et  $p$  quel que soit l'*a priori* choisi. Il faut cependant noter que pour un *a priori* impropre la moyenne *a posteriori* du PLR est infinie :

$$\mathbb{E}_1[\text{LR}(\mathbf{x}, \theta) \mid \mathbf{x}] = \int d\theta \frac{p(\mathbf{x} \mid \theta_0)}{p(\mathbf{x} \mid \theta)} \pi_1(\theta \mid \mathbf{x}) = p(\mathbf{x} \mid \theta_0) \frac{\int d\theta \pi_1(\theta)}{\int d\theta p(\mathbf{x} \mid \theta) \pi_1(\theta)} = +\infty \quad (8.15)$$

si  $\int d\theta \pi_1(\theta) = +\infty$  et  $\int d\theta p(\mathbf{x} \mid \theta) \pi_1(\theta) > 0$ .

## Performances fréquentistes du PLR vu comme un détecteur

On s'intéresse ici aux propriétés fréquentistes du PLR, et en particulier aux Caractéristiques Opérationnelles du Récepteur (ROC) qui représente la "probabilité de bonne détection PD" en fonction de la PFA.

### • Notations

Notons T un test tel que PLR, GLR ... et  $\lambda$  le(s) paramètre(s) le définissant :  $\lambda = (\zeta, p)$  pour le PLR par exemple. Notons de plus  $\mathcal{X}_{T, \lambda}$  l'espace de réjection associé à ce test et ce(s) paramètre(s). Notons  $P_i$

<sup>5</sup>La distribution *a posteriori* est bien définie si et seulement si  $0 < \int d\theta p(\mathbf{x} \mid \theta) \pi(\theta) < \infty$  i.e. si  $0 < \text{BF}(\mathbf{x}) < \infty$ .

une probabilité fréquentiste (conditionnelle à  $\boldsymbol{\theta} = \boldsymbol{\theta}_i$ ) de réjection de  $H_0 : P_0 = \text{PFA}$  et  $P_1 = \text{PD}(\boldsymbol{\theta}_1)$  pour la Probabilité de bonne Détection PD. Ceci s'écrit :

$$P_i(\mathbf{T}, \boldsymbol{\lambda}) = \int_{\mathcal{X}_{\mathbf{T}, \boldsymbol{\lambda}}} d\mathbf{x} p(\mathbf{x} | \boldsymbol{\theta}_i) \quad (8.16)$$

$$\text{avec } \mathcal{X}_{\mathbf{T}, \boldsymbol{\lambda}} = \{\mathbf{x} : H_0 \text{ est rejeté par } \mathbf{T} \text{ de paramètre } \boldsymbol{\lambda}\} \quad (8.17)$$

### • Famille de tests à deux paramètres

Comme nous l'avons vu, en général les tests sont paramétrés par un unique paramètre auquel on attribue une interprétation claire :  $\zeta$  (rapport de vraisemblances, rapport de probabilités *a posteriori*),  $\alpha$  (PFA) *etc.* Le PLR en tant que test binaire est défini quant à lui par *deux* paramètres tous deux interprétables :  $\zeta$  (rapport de vraisemblances) et  $p$  (probabilité *a posteriori*).

Pour étudier la courbe ROC, afin de tracer PD en fonction d'un unique paramètre (PFA), il convient de choisir un test à un paramètre issu de la famille de test à deux paramètres que constitue le PLR, *i.e.* de choisir une fonction  $\lambda : \mathbb{R}_{+*}^2 \rightarrow [0, 1]$  avec  $\lambda(\zeta, p) = \text{PFA}$ . Le PLR définit ainsi une famille de tests optimisable en le sens fréquentiste : l'ensemble des tests binaires  $\text{PLR}(\mathbf{x}, \zeta) > p$  définit un ensemble de courbes ROC et pas seulement une courbe ROC comme pour les tests usuels (à un paramètre). Nous verrons ci-dessous comment procéder numériquement.

### • Modèle invariant : $\text{PFA}(\text{PLR}, \zeta = 1, p) = 1 - p$

Dans un contexte de modèle invariant que nous étudierons plus loin, nous démontrerons sous des conditions très classiques le théorème 5 p.160 qui réexprime le PLR comme une intégrale fréquentiste (intégrale non plus sur  $\Theta$  mais sur  $\mathcal{X}$ ). Comme cela sera aussi démontré à la suite de ce théorème, le PLR peut alors être réexprimé en utilisant la fonction de densité cumulative  $F$  sous  $H_0$  d'une variable aléatoire  $T(\mathbf{X})$  où  $T$  est une certaine fonction continue déterministe de  $\mathbf{x}$  (statistique). Plus précisément, d'après la future Eq. (10.19) p. 161,

$$\text{PLR}(\mathbf{x}, \zeta) = 1 - F_{T(\mathbf{X})|\boldsymbol{\theta}_0}(\zeta^{-1}T(\mathbf{x})) \quad (8.18)$$

où l'expression de la fonction  $T(\cdot)$  n'importe pas ici<sup>6</sup>.

PFA et PD sont alors donnés par la définition générique  $P_i$  de l'Eq. (8.16).

$$P_i(\text{PLR}, \zeta, p) = \Pr(\text{PLR}(\mathbf{x}, \zeta) > p \mid \boldsymbol{\theta}_i)$$

et donc sous les conditions d'invariance mentionnées,

$$\begin{aligned} P_i(\text{PLR}, \zeta, p) &= \Pr\left(1 - F_{T(\mathbf{X})|\boldsymbol{\theta}_0}(\zeta^{-1}T(\mathbf{x})) > p \mid \boldsymbol{\theta}_i\right) \\ &= \Pr\left(F_{T(\mathbf{X})|\boldsymbol{\theta}_0}(\zeta^{-1}T(\mathbf{x})) < 1 - p \mid \boldsymbol{\theta}_i\right) \\ &= \Pr\left(\zeta^{-1}T(\mathbf{x}) < F_{T(\mathbf{X})|\boldsymbol{\theta}_0}^{-1}(1 - p) \mid \boldsymbol{\theta}_i\right) \end{aligned} \quad (8.19)$$

car  $\mathbf{x} \rightarrow T(\mathbf{x})$  est une fonction continue de telle sorte que la distribution cumulative de la variable aléatoire induite  $T(\mathbf{X})$  est strictement croissante de 0 à 1 exclus.

En particulier, pour  $\boldsymbol{\theta}_i = \boldsymbol{\theta}_0$  on peut réutiliser la distribution cumulative de  $T(\mathbf{X})$  sous  $H_0$  :

$$\text{PFA}(\text{PLR}, \zeta, p) = F_{T(\mathbf{X})|\boldsymbol{\theta}_0}\left(\zeta F_{T(\mathbf{X})|\boldsymbol{\theta}_0}^{-1}(1 - p)\right)$$

En particulier, pour  $\zeta = 1$  et sous  $H_0$  on constate que

<sup>6</sup>Juste pour exemple, supposons que la distribution de la variable aléatoire  $\mathbf{X}$  appartient à une famille de vraisemblances invariante par translation paramétrée par  $\boldsymbol{\theta}$ , c'est-à-dire que la distribution de la translation de  $\mathbf{X}$  appartient à cette même famille pour une valeur translatée de même de  $\boldsymbol{\theta}$ . Si on choisit comme *a priori* sur  $\Theta$  l'*a priori* uniforme sur  $\mathbb{R}^L$  et que l'on teste l'hypothèse  $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  alors d'après le théorème 5 p.160 le PLR est donné par l'équation (8.18) pour  $T(\mathbf{x}) = p(\mathbf{x}|\boldsymbol{\theta}_0)$ .

**Corrolaire 4** Dans les conditions d'invariance du théorème 5, le test binaire

$$\text{Rejeter } H_0 \text{ si } \text{PLR}(\mathbf{x}, \zeta = 1) > p \quad (8.20)$$

a une PFA donnée directement par le paramètre  $p$  :

$$\text{PFA}(\text{PLR}, 1, p) = 1 - p \quad (8.21)$$

Nous venons de redémontrer que la distribution d'une  $p$ -value sous  $H_0$  est uniforme de 0 à 1 car pour  $\zeta = 1$  on pouvait reconnaître dès l'équation (8.18) que le PLR évalué en  $\zeta = 1$  est égal à une  $p$ -value (voir le futur corrolaire dédié 7 p. 161).

### • Modèle invariant : une seule courbe ROC

On constate à partir de cette même dernière équation pour  $\zeta$  arbitraire, que contrairement au cas général, pour une PFA fixée, PD ne dépend que de PFA et non plus d'un second paramètre.

**Corrolaire 5** Dans les conditions d'invariance du théorème 5 p. 160, le test binaire

$$\text{Rejeter } H_0 \text{ si } \text{PLR}(\mathbf{x}, \zeta) > p \quad (8.22)$$

a une PD fixée pour une PFA fixée, quel que soit le choix de  $(\zeta, p)$  qui mène à cette PFA. Dans le contexte de modèle invariant considéré, le test du PLR n'a donc plus qu'une courbe ROC et non plus tout un ensemble de courbes ROC.

Preuve : Notons par  $\alpha$  la PFA associée aux paramètres  $(\zeta, p)$  : d'après l'Eq. (8.20),

$$\begin{aligned} \alpha &= F_{T(\mathbf{X})|\theta_0} \left( \zeta F_{T(\mathbf{X})|\theta_0}^{-1}(1-p) \right) \\ \text{i.e. } \zeta F_{T(\mathbf{X})|\theta_0}^{-1}(1-p) &= F_{T(\mathbf{X})|\theta_0}^{-1}(\alpha) \end{aligned}$$

Injectant cette grandeur dans l'expression de PD de l'Eq. (8.19) on obtient pour un  $\theta_1$  fixé quelconque

$$\text{PD}_{\theta_1}(\text{PLR}, \zeta, p) = \Pr \left( T(\mathbf{x}) < F_{T(\mathbf{X})|\theta_0}^{-1}(\alpha) \mid \theta_1 \right)$$

Le terme de droite dépend uniquement de  $\alpha$  (la PFA) et non plus d'une combinaison de  $\zeta$  et/ou  $p$ .

### • Remarques simples

Pour un seuil de probabilité  $p$  fixé, si  $\zeta' \leq \zeta$  alors l'espace de réjection de  $H_0$  défini par le PLR pour  $\zeta'$  est inclus dans l'espace de réjection du PLR pour  $\zeta$ . Ceci implique simplement que  $P_i(\text{PLR}, \zeta', p) < P_i(\text{PLR}, \zeta, p)$  si  $\zeta' \leq \zeta$ . Formellement on peut écrire la série d'implications menant à cette remarque ainsi : si  $\zeta' \leq \zeta$

$$\begin{aligned} \text{LR}(\mathbf{x}, \theta) \leq \zeta' &\Rightarrow \text{LR}(\mathbf{x}, \theta) \leq \zeta \\ \text{donc } \Pr(\text{LR}(\mathbf{x}, \theta) \leq \zeta' | \mathbf{x}) &\leq \Pr(\text{LR}(\mathbf{x}, \theta) \leq \zeta | \mathbf{x}) \\ \text{donc } \Pr(\text{LR}(\mathbf{x}, \theta) \leq \zeta' | \mathbf{x}) > p &\Rightarrow \Pr(\text{LR}(\mathbf{x}, \theta) \leq \zeta | \mathbf{x}) > p \\ \text{donc } \mathcal{X}_{\text{PLR}, \zeta', p} &\subset \mathcal{X}_{\text{PLR}, \zeta, p} \\ \text{donc } P_i(\text{PLR}, \zeta', p) &\leq P_i(\text{PLR}, \zeta, p) \quad \text{si } \zeta' \leq \zeta \end{aligned} \quad (8.23)$$

On obtient de la même façon d'autres inégalités liant les probabilités fréquentistes de certains tests :

$$P_i(\text{PLR}, \zeta, p'_0) \leq P_i(\text{PLR}, \zeta, p_0) \quad \text{si } p'_0 \geq p_0 \quad (8.24)$$

$$P_i(\text{GLR}, \zeta') \leq P_i(\text{GLR}, \zeta) \quad \text{si } \zeta' \leq \zeta \quad (8.25)$$

$$P_i(\text{GLR}, \zeta) \geq P_i(\text{PLR}, \zeta, p) \quad \forall \zeta, p \quad (8.26)$$

## 8.3 Implémentation et optimisation fréquentiste numériques du PLR

### Implémentation

L'implémentation de ce test peut paraître très compliquée. Cette complexité est grandement réduite en utilisant un algorithme MCMC (*Monte Carlo Markov Chain*) : il suffit d'échantillonner la loi *a posteriori* habituelle  $\pi(\boldsymbol{\theta}|\mathbf{x})$  comme ceci est illustré dans la sous-section de la page 179.

1. Échantillonner la loi *a posteriori*  $\pi(\boldsymbol{\theta}|\mathbf{x})$  en générant par un algorithme MCMC une chaîne de Markov de taille effective  $N_J$  :  $\{\boldsymbol{\theta}^{[j]} \sim \pi(\boldsymbol{\theta}|\mathbf{x}), j \in \llbracket 1, N_J \rrbracket\}$ .
2. Calculer la chaîne  $\{\text{LR}(\mathbf{x}, \boldsymbol{\theta}^{[j]}), j \in \llbracket 1, N_J \rrbracket\}$  en évaluant chaque élément par :

$$\text{LR}(\mathbf{x}, \boldsymbol{\theta}^{[j]}) = \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)}{p(\mathbf{x}|\boldsymbol{\theta}^{[j]})} \quad (8.27)$$

3. Estimer le PLR en tant que distribution cumulative empirique de la chaîne de LR, c'est-à-dire compter le nombre de  $\text{LR}(\mathbf{x}, \boldsymbol{\theta}^{[j]})$  inférieurs au seuil  $\zeta$  et diviser par  $N_J$  :

$$\text{PLR}(\mathbf{x}, \zeta) = \Pr(\text{LR}(\mathbf{x}, \boldsymbol{\theta}) \leq \zeta | \mathbf{x}) \approx \frac{\text{Card}\left(\left\{\text{LR}(\mathbf{x}, \boldsymbol{\theta}^{[j]}) \leq \zeta, j \in \llbracket 1, N_J \rrbracket\right\}\right)}{N_J} \quad (8.28)$$

4. Appliquer le test binaire en comparant pour  $(\zeta, p)$  fixés  $\text{PLR}(\mathbf{x}, \zeta)$  à  $p$ .
5. Si  $H_0$  est rejeté, utiliser la chaîne  $\{\boldsymbol{\theta}^{[j]}, j\}$  pour la phase d'estimation.

Le FBF peut aussi être calculé à partir de la chaîne de LR : comme nous l'avons mentionné dans la section 8.1, BF peut être estimé par une simple moyenne harmonique des LR. Une telle méthode d'estimation n'est cependant pas toujours satisfaisante. Pour plus de détails reprendre les références de la section 8.1.

Notons que dans le contexte du modèle invariant considéré dans le théorème 5 p.160, le théorème 5 exprime le PLR comme une intégrale sur  $\mathcal{X}$ . Ainsi, pour calculer le PLR en pratique il est alors possible d'échantillonner la vraisemblance sous  $\boldsymbol{\theta}_0$  plutôt que la loi *a posteriori*. La chaîne à constituer n'est alors plus celle des  $\text{LR}(\mathbf{x}, \boldsymbol{\theta}^{[j]})$  mais celle des  $T(\mathbf{x}^{[j]})$ .

### Optimisation fréquentiste

D'autre part, comme dans le cas général –et non dans le contexte du modèle invariant– le PLR est une famille de tests paramétrée par deux paramètres, comme nous l'avons mentionné dans la section 8.2 il est possible de sélectionner la paramétrisation  $\lambda(\zeta, p)$  qui optimise au sens de Neyman-Pearson le test  $(\text{PLR}, \zeta, p)$  au sein de toute la famille définie par  $\{(\zeta, p) \in [0, \infty) \times [0, 1]\}$ . La méthode consiste à :

1. calculer  $\text{PFA}(\zeta, p)$  et  $\text{PD}(\zeta, p) \forall (\zeta, p)$
2. fixer un  $\text{PFA}_0$  et obtenir  $\{(\zeta, p) : \text{PFA}(\zeta, p) = \text{PFA}_0\}$   
– choisir de cet ensemble  $(\zeta^*(\text{PFA}_0), p^*(\text{PFA}_0))$  qui maximise  $\text{PD}(\zeta, p)$ .

Cette procédure est appliquée numériquement en utilisant l'algorithme MCMC décrit ci-dessus et grâce à un réordonnement et une approximation asymptotique.  $N_I$  jeux de données sont simulés d'une part sous  $H_0$ , d'autre part sous  $H_1$  (pour  $\boldsymbol{\theta}_1 \in \Theta$  fixé par exemple). Pour chaque hypothèse, une matrice  $\mathbb{L}\mathbb{R}$  de taille  $N_I \times N_J$  est construite de sorte que la ligne  $i$  de  $\mathbb{L}\mathbb{R}$  contient la chaîne de Markov  $\{\text{LR}(\mathbf{x}, \boldsymbol{\theta}^{[j]}), j\}$  obtenue à partir du  $i^{\text{ème}}$  jeu de données. Chaque matrice est réordonnée en classant par ordre croissant chaque ligne afin d'obtenir la distribution cumulative pour un jeu de données fixé. Ensuite, chaque colonne est réordonnée par ordre croissant afin d'obtenir la distribution cumulative fréquentiste pour un seuil  $\zeta$  fixé. Pour un nombre  $N_I$  de jeux de données et une taille de chaîne  $N_J$  suffisamment grands, des valeurs approximatives de  $\text{PFA}(\zeta, p)$  et  $\text{PD}(\zeta, p)$  sont lues dans chaque matrice. Par exemple, avec  $\zeta(i, j) = \mathbb{L}\mathbb{R}_{i,j}$

$((i, j)$ <sup>ème</sup> composante de la matrice  $\mathbb{LR}$ ) et  $p(i, j) = j/N_J$ , la PFA est approximativement donnée par  $\text{PFA}(\zeta(i, j), p(i, j)) = i/N_I$ . Les paramètres approximativement optimaux peuvent alors être obtenus en lisant conjointement les deux matrices. Cette méthode est appliquée dans le prochain chapitre.

Une procédure analogue peut être appliquée au FBF lui aussi défini à partir de deux paramètres ( $\zeta$  et  $b$ ) : le FBF est calculé à partir de la chaîne de Markov  $\{\text{LR}(\mathbf{x}, \boldsymbol{\theta}^{[j]}), j \in \llbracket 1, N_J \rrbracket\}$  puis le test est optimisé par rapport à  $\zeta$  et  $b$ . Le sens d'une telle optimisation n'est cependant pas très clair.

## Chapitre 9

# Introduction à l'invariance de groupe appliquée aux statistiques

### 9.1 Motivation pour l'utilisation de l'invariance de groupe en statistiques

Comme dans l'exemple 10 p. 17, afin de définir la densité *a priori*  $\pi$  sur  $\Theta$ , il est parfois possible de formuler une contrainte d'invariance de la distribution *a priori* de  $\boldsymbol{\theta}$  sous l'application d'un endomorphisme inversible  $g : \Theta \rightarrow \Theta$  à  $\boldsymbol{\theta}$ . Comme précédemment, une telle contrainte s'écrit

$$\Pr(\boldsymbol{\theta} \in A) = \Pr(g(\boldsymbol{\theta}) \in A) \quad \forall A \subset \Theta \quad (9.1)$$

D'après la théorie des ensembles,

$$\Pr(g(\boldsymbol{\theta}) \in A) = \Pr(\boldsymbol{\theta} \in g^{-1}(A)) \quad (9.2)$$

où  $g^{-1}(A) = \{g^{-1}(\boldsymbol{\theta}), \boldsymbol{\theta} \in A\}$ . Pour alléger les notations nous omettrons les parenthèses et noterons par exemple  $g^{-1}A = \{g^{-1}\boldsymbol{\theta}, \boldsymbol{\theta} \in A\}$ . La contrainte d'invariance s'exprime donc

$$\Pr(\boldsymbol{\theta} \in A) = \Pr(\boldsymbol{\theta} \in g^{-1}A) \quad (9.3)$$

On se limite de plus ici par simplicité à une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue  $d\boldsymbol{\theta} : \Pr(\boldsymbol{\theta} \in A) = \int_A d\boldsymbol{\theta} \pi(\boldsymbol{\theta})$  (voir Halmos and Savage [1949] pour une introduction rapide aux propriétés de base des densités de mesure). Appliquant le changement de variable  $\boldsymbol{\theta}' = g\boldsymbol{\theta}$  au membre de droite et notant  $J_g$  son Jacobien (matrice des dérivées partielles), on obtient

$$\int_A d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) = \int_A d\boldsymbol{\theta}' \pi(g^{-1}\boldsymbol{\theta}') |J_g(\boldsymbol{\theta}')|^{-1}$$

Comme ceci est vrai  $\forall A \subset \Theta$  on obtient l'équation fonctionnelle

$$\pi(\boldsymbol{\theta}) = \pi(g^{-1}\boldsymbol{\theta}) |J_g(\boldsymbol{\theta})|^{-1} \quad \forall \boldsymbol{\theta} \in \Theta \quad (9.4)$$

Si la solution  $\pi(\cdot) \geq 0$  de cette équation existe, qu'elle est unique à une constante multiplicative près et qu'elle est  $L_1$  intégrable, *i.e.* telle que  $\int_{\Theta} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) < \infty$ , alors il est possible de fixer la constante de telle sorte que  $\Pr(\boldsymbol{\theta} \in \Theta) = 1$  et donc que  $\pi$  définisse bien une densité de probabilité. Cependant, pour certaines transformations usuelles la solution est bien définie à une constante multiplicative près mais  $\Pr(\boldsymbol{\theta} \in \Theta) = \infty$ , de telle sorte que l'Eq. (9.4) n'admet pas de solution qui soit une densité de probabilité.

**Exemple 12** Si  $g(\boldsymbol{\theta}) = \boldsymbol{\theta} + \mathbf{a} \quad \forall \boldsymbol{\theta} \in \Theta$  avec  $\Theta = \mathbb{R}^p$  et  $\mathbf{a} \in \Theta$ , l'Eq. (9.4) se réécrit  $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta} - \mathbf{a})$  dont l'ensemble de solutions  $\pi(\boldsymbol{\theta}) \geq 0$  pour tout  $\boldsymbol{\theta} \in \Theta$  est donné par

$$\pi(\boldsymbol{\theta}) = c \quad \forall \boldsymbol{\theta} \in \mathbb{R}^p$$

pour  $c \geq 0$  arbitraire. Excluant la solution  $\pi = 0$ ,  $\int_{\Theta} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) = \infty$  donc  $\pi$  n'est pas une densité de probabilité.

**Exemple 13** Dans l'exemple 10,  $g(\theta) = a\theta \quad \forall \theta \in \Theta$ , avec  $\Theta = \mathbb{R}_{+*}$  et  $a \in \Theta$ . Eq. (9.4) se réécrit alors  $\pi(\theta) = \pi\left(\frac{\theta}{a}\right)^{\frac{1}{a}}$  dont l'ensemble de solutions est donné par

$$\pi(\theta) = \frac{c}{\theta} \quad \forall \theta > 0$$

avec  $c > 0$  arbitraire. A nouveau, aucune de ces densités ne définit une densité de probabilité car  $\int_0^{+\infty} d\theta \theta^{-1} = \infty$ .

On constate dans ces exemples que les contraintes d'invariance les plus simples n'admettent pas de solution dans l'espace des mesures de probabilité. Les solutions dites *impropres* (i.e. telles que  $\int_{\Theta} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) = \infty$ ) sont pourtant intéressantes mais elles requièrent de reprendre plus en profondeur des arguments tels que présentés dans l'exemple 10 puisque ceux-ci mènent à des solutions qui sortent du cadre fixé initialement par l'utilisateur lorsqu'il imposait une égalité de *probabilités*.

En théorie de la mesure, les mesures vérifiant une contrainte d'invariance *du type* de l'Eq. (9.3) définissent la classe des mesures dites de Haar. Celles-ci sont cependant en fait définies sur l'espace des groupes et ne peuvent qu'ensuite être utilisées pour définir une mesure sur  $\Theta$ .

L'invariance de groupe appliquée aux mesures est un sujet très complexe et bien trop long pour être bien introduit dans cette thèse car il supposerait d'introduire aussi des bases de topologie et de théorie de la mesure. Supposant ces bases en topologie et en théorie de la mesure connues, des introductions à l'invariance de groupe appliquée aux statistiques sont données dans Berger [1985] pour une version simple se limitant à des sous-espaces de  $\mathbb{R}^n$ , Eaton [1989] pour un exposé à la fois clair, détaillé et rigoureux, ou Eaton [2007] pour une utilisation plus courante des Jacobiens que dans Eaton [1989].

Nous nous plaçons ici globalement dans le même cadre que ces ouvrages. Par simplicité et afin d'alléger la présentation, nous supposons implicite toute considération sur les  $\sigma$ -algèbres  $\mathcal{S}$  des sous-ensembles de  $\mathcal{X}$  définissant les espaces mesurables impliqués ici (notions de théorie de la mesure) ou sur la continuité d'espaces ou de transformations entre les différents espaces impliqués ici (notions de topologie). L'objectif est ici d'utiliser ces mathématiques, complexes mais inévitables, afin de démontrer certains résultats utilisables tels quels en traitement statistique du signal.

## 9.2 A priori de Haar

Nous introduisons ainsi ici le strict nécessaire de la théorie de la mesure de Haar qui permette d'introduire l'*a priori* dit *de Haar* (sur  $\Theta$ ) assez bien connu des statistiques appliquées. Voir Nachbin [1965] ou Eaton [1989] pour des introductions plus détaillées et rigoureuses mais qui ne s'étendent pas sur l'*a priori* induit sur  $\Theta$  par la mesure de Haar. Nous développons ensuite les propriétés de base de ce dernier, toujours dans une optique très appliquée en essayant cependant d'éviter de trop perdre en rigueur. A ce stade l'*a priori* de Haar est toujours défini comme le résultat d'une équation fonctionnelle. Leur expression explicite via un Jacobien est finalement donnée dans un cas relativement général mais non nécessaire à la démonstration qui suivra.

### Mesures de Haar sur un groupe (topologique localement compact)

Comme détaillé dans Nachbin [1965], la mesure de Haar est en premier lieu définie sur un groupe [topologique localement compact].

**Définition 7** On appelle **groupe**  $\mathcal{G}$  un ensemble muni d'une loi de composition  $\circ$  tel que

- $g_1, g_2 \in \mathcal{G}$  implique  $g_1 \circ g_2 \in \mathcal{G}$
- Il existe un élément neutre  $e \in \mathcal{G}$  tel que  $e \circ g = g \circ e = g$  pour  $g \in \mathcal{G}$
- Pour tout  $g \in \mathcal{G}$  il existe un unique élément  $g^{-1} \in \mathcal{G}$  tel que  $g \circ g^{-1} = g^{-1} \circ g = e$ .

Une composition est en général notée implicitement  $g_1 g_2$  plutôt que  $g_1 \circ g_2$ .

Illustrons les avertissements du précédent paragraphe : dans la suite  $\mathcal{G}$  ne sera en fait pas seulement un groupe puisqu'il doit aussi être mesurable et de propriétés convenables. Il s'agira en fait d'un groupe dit topologique localement compact, c'est-à-dire à la fois un groupe tel que défini ci-dessus et un espace topologique suffisamment "régulier" et tel que –notamment– la fonction  $(g_1, g_2) \rightarrow g_1 g_2$  est continue (voir Nachbin [1965] ou Eaton [1989]). Nous ne rentrons pas dans ces notions qui nécessiteraient des introductions bien trop longues et techniques. Voir Kechris [1995], Billingsley [1995] et Eaton [1989] pour respectivement : une introduction à la théorie des ensembles, son application à la mesure de probabilité et son application au sujet qui nous intéresse à savoir l'invariance de groupe appliquée aux statistiques. Une introduction détaillée à la mesure de Haar est donnée dans Nachbin [1965]. La plupart des définitions et propositions données ci-dessous sont des retranscriptions de Nachbin [1965] et/ou Eaton [1989]. Par souci de concision les propositions classiques sont données sans démonstration.

**Définition 8** On appelle **mesure invariante à gauche de Haar sur le groupe**  $\mathcal{G}$  une mesure de Radon<sup>1</sup>  $H^l$  ("l" pour left) telle que pour toute fonction  $f$  continue de support compact inclus dans  $\mathcal{G}$  (ensemble de fonctions noté  $K(\mathcal{G})$ )

$$\int_{\mathcal{G}} f(g) H^l(dg) = \int_{\mathcal{G}} f(g_0 g) H^l(dg) \quad \forall g_0 \in \mathcal{G} \quad (9.5)$$

**Définition 9** On appelle **mesure invariante à droite de Haar sur**  $\mathcal{G}$  une mesure  $H^r$  ("r" pour right) telle que

$$\int f(g) H^r(dg) = \int f(g g_0) H^r(dg) \quad \forall g_0 \in \mathcal{G}, \forall f \in K(\mathcal{G}) \quad (9.6)$$

L'absence de domaine d'intégration sous-entend toujours implicitement que celui-ci est le domaine tout entier.

Ces mesures invariantes tiennent leur nom de A. Haar, le premier à avoir établi pour de nombreux groupes topologiques l'existence de mesures invariantes.

**Proposition 10** Pour un groupe  $\mathcal{G}$  donné, la mesure invariante à gauche de Haar existe et est unique à une constante multiplicative près. Il en est de même pour la mesure invariante à droite de Haar.

**Définition 10** On appelle **module (droit) de**  $\mathcal{G}$  la fonction  $\Delta$  de  $\mathcal{G}$  à valeur dans  $\mathbb{R}_{+*}$  telle que si  $H^l$  est une mesure invariante à gauche de Haar

$$\int f(g g_0^{-1}) H^l(dg) = \Delta(g_0) \int f(g) H^l(dg) \quad \forall f \in K(\mathcal{G}) \quad (9.7)$$

D'après la propriété précédente d'unicité de la mesure invariante de Haar,

**Proposition 11** Le module  $\Delta$  existe et ne dépend pas du choix de  $H^l$ .

**Proposition 12** Le module  $\Delta$  est une fonction continue telle que

$$\Delta(g_1 g_2) = \Delta(g_1) \Delta(g_2) \quad \forall g_1, g_2 \in \mathcal{G}$$

ce qui implique aussi  $\Delta(e) = 1$  et  $\Delta(g^{-1}) = \Delta(g)^{-1}$ .

<sup>1</sup>On confondra mesure de Radon et intégrale car l'une définit l'autre de façon unique, voir Eaton [1989] par exemple.



**Proposition 13** Si  $H^l$  est une mesure invariante à gauche de Haar sur  $\mathcal{G}$  et  $\Delta$  le module de  $\mathcal{G}$  alors

$$\int f(g^{-1})H^l(dg) = \int f(g)\Delta(g)^{-1}H^l(dg) \quad \forall f \in K(\mathcal{G})$$

Le module permet de plus de relier mesures invariantes à gauche et à droite de Haar. La proposition précédente implique en particulier :

**Proposition 14** Si  $H^l$  est une mesure invariante à gauche de Haar et  $\Delta$  le module de  $\mathcal{G}$  alors la mesure définie par

$$H^r(dg) = \Delta(g)^{-1}H^l(dg) \quad (9.8)$$

est une mesure invariante à droite de Haar.

Autrement dit,

$$\int f(gg_0)\Delta(g)^{-1}H^l(dg) = \int f(gg_0)H^r(dg) = \int f(g)\Delta(g)^{-1}H^l(dg) \quad \forall g_0 \in \mathcal{G}, \forall f \in K(\mathcal{G})$$

Notons qu'en définissant ainsi  $H^r$ , la proposition 13 se réécrit

$$\int f(g^{-1})H^l(dg) = \int f(g)H^r(dg)$$

De même,

**Proposition 15** Si  $H^r$  est une mesure invariante à droite de Haar alors la mesure définie par

$$H^l(dg) = \Delta(g)H^r(dg) \quad (9.9)$$

est une mesure invariante à gauche de Haar.

### Mesures *a priori* de Haar induites sur $\Theta$ lorsque $\theta \rightarrow g\theta$ et $g \rightarrow g\theta$ sont bijectives

La notion de groupe doit maintenant être reliée à l'espace des paramètres  $\Theta$  qui nous intéresse. En statistiques, l'opération sous laquelle on a ou non "invariance" est une opération sur  $\Theta$  comme dans les exemples 12 et 13, et/ou sur  $\mathcal{X}$  comme on le verra bientôt.

Les éléments les plus naturels à attribuer à  $\mathcal{G}$  sont donc des transformations bijectives  $g : \Theta \rightarrow \Theta$ . Sur  $\mathcal{G}$ , l'opération naturelle du groupe est alors la composition de fonctions :  $(g_1g_2)(\theta) = g_1(g_2(\theta))$ . L'inverse de groupe est donc l'inverse au sens des fonctions et l'élément neutre  $e$  est la fonction identité.

**Définition 11** On appelle **groupe de transformations**  $\mathcal{G}$  un groupe tel que défini dans la définition 7 et dont les éléments sont des transformations bijectives.

On supposera de plus très simplement dans la suite que la transformation qui à  $g$  associe  $g(\theta)$ , noté  $g\theta$ , définit une bijection entre  $\mathcal{G}$  et  $\Theta$ . Autrement dit, on supposera dans la suite que les fonctions définies par

$$g : \Theta \rightarrow \Theta \quad (9.10)$$

$$\theta \rightarrow g\theta$$

$$\phi_\theta : \mathcal{G} \rightarrow \Theta \quad (9.11)$$

$$g \rightarrow g\theta$$

sont toutes bijectives, quel que soit  $g \in \mathcal{G}$  pour la première classe de fonctions et quel que soit  $\theta \in \Theta$  pour la seconde.<sup>2</sup>

<sup>2</sup>Nous verrons dans la section 10.1 que cette hypothèse peut être un peu assouplie mais au prix d'une plus grande technicité des hypothèses et de la démonstration. Elle est déjà suffisante pour de nombreux exemples importants.

**Exemple 14** Dans les exemples 12 et 13, si l'on souhaite imposer la contrainte d'invariance à toute transformation  $g$  on définit l'ensemble des transformations par

$$\mathcal{G} = \{g_{\mathbf{c}}, \mathbf{c} \in C\} \quad C = \Theta$$

avec dans le premier exemple  $\Theta = \mathbb{R}^n$  et  $g_{\mathbf{c}}(\boldsymbol{\theta}) = \boldsymbol{\theta} + \mathbf{c} \forall \boldsymbol{\theta} \in \Theta$  et dans le second  $\Theta = \mathbb{R}_{+*}$  et  $g_{\mathbf{c}}(\theta) = c\theta \forall \theta \in \Theta$ . Dans les deux cas  $\mathcal{G}$  définit bien un groupe de transformations tel que défini dans la définition 11.

La transformation  $\phi_{\boldsymbol{\theta}}$  définit une bijection entre  $\mathcal{G}$  et  $\Theta$  qui permet de définir simplement une mesure sur  $\Theta$  à partir d'une mesure sur  $\mathcal{G}$ . En particulier, les mesures de Haar sur  $\mathcal{G}$  définies plus haut induisent directement une mesure "a priori de Haar" sur  $\Theta$  définie en général de même à une constante multiplicative près. Villegas [1981] développe cette idée d'induire un a priori sur  $\Theta$  à partir d'une mesure de Haar et Halmos and Savage [1949] permettent de s'introduire pédagogiquement et succinctement au concept de transformation (mesurable) entre espaces mesurables.

**Définition 12** On appelle *mesure a priori à gauche de Haar* la mesure sur  $\Theta$  définie par  $\Pi_{\mathbf{a}}^l = H^l(\phi_{\mathbf{a}}^{-1})$  où  $H^l$  est une mesure invariante à gauche de Haar sur  $\mathcal{G}$  fixée et  $\phi_{\mathbf{a}}, \mathbf{a} \in \Theta$ , est l'isomorphisme tel que  $\phi_{\mathbf{a}}(g) = g\mathbf{a} \forall g \in \mathcal{G}$ .

**Définition 13** On appelle *mesure a priori à droite de Haar* la mesure sur  $\Theta$  définie par  $\Pi_{\mathbf{a}}^r = H^r(\phi_{\mathbf{a}}^{-1})$  où  $H^r$  est une mesure invariante à droite de Haar sur  $\mathcal{G}$  fixée.

Ceci se lit de la façon suivante : pour tout sous-ensemble mesurable  $A \subset \Theta$ ,

$$\begin{aligned} \Pi_{\mathbf{a}}^l(A) &= H^l(\phi_{\mathbf{a}}^{-1}A) \\ \text{où } \phi_{\mathbf{a}}^{-1}A &= \{\phi_{\mathbf{a}}^{-1}\boldsymbol{\theta}, \boldsymbol{\theta} \in A\} = \{g \in \mathcal{G} : \phi_{\mathbf{a}}g = \boldsymbol{\theta}, \boldsymbol{\theta} \in A\} = \{g \in \mathcal{G} : g\mathbf{a} = \boldsymbol{\theta}, \boldsymbol{\theta} \in A\} \end{aligned}$$

Si  $\Pi_{\mathbf{a}}^l$  est une mesure de probabilité (condition donnée dans la prochaine section), la mesure a priori sur  $\Theta$  s'écrit plus communément, :

$$\Pr(\boldsymbol{\theta} \in A) = \Pi_{\mathbf{a}}^l(A)$$

Notons une propriété intéressante de la mesure a priori à droite de Haar sur  $\Theta$  :

**Proposition 16** La mesure a priori à droite de Haar  $\Pi_{\mathbf{a}}^r$  ne dépend pas du choix de  $\mathbf{a}$ . Elle est donc unique pour une mesure  $H^r$  fixée et est noté  $\Pi^r$ .

Preuve :

$$\begin{aligned} \int f(\boldsymbol{\theta})\Pi_{\mathbf{a}}^r(d\boldsymbol{\theta}) &= \int f(\boldsymbol{\theta})H^r(\phi_{\mathbf{a}}^{-1}d\boldsymbol{\theta}) \quad f \in K(\Theta) \\ &= \int f(\phi_{\mathbf{a}}g)H^r(dg) \quad (\text{transformation de mesure : } \boldsymbol{\theta} = \phi_{\mathbf{a}}g) \\ &= \int f(\phi_{\mathbf{a}}(g \circ g_0))H^r(dg) \quad \forall g_0 \in \mathcal{G} \quad (\text{définition de } H^r \text{ par l'Eq. (9.8)}) \\ &= \int f(g(g_0(\mathbf{a})))H^r(dg) \quad (\text{définition de } \phi_{\mathbf{a}} \text{ et de l'action de } \mathcal{G}) \\ &= \int f(\phi_{\mathbf{b}}g)H^r(dg) \quad \text{avec } \mathbf{b} = g_0\mathbf{a} \\ &= \int f(\boldsymbol{\theta})\Pi_{\mathbf{b}}^r(d\boldsymbol{\theta}) \end{aligned}$$

Comme  $g_0 \rightarrow \mathbf{b}$  définit une bijection entre  $\mathcal{G}$  et  $\Theta$ ,  $\Pi^r$  ne dépend pas de  $\mathbf{a}$ . Villegas [1981] montre de plus que cette mesure a priori est la seule qui puisse être induite par la transformation  $\phi_{\mathbf{a}}$  sans dépendre du choix de  $\mathbf{a} \in \Theta$ .

On montre de la même façon :

**Proposition 17** *Pour une mesure à gauche de Haar  $H^l$  fixée, les mesures a priori à gauche de Haar de la classe  $\{\Pi_{\mathbf{a}}^l, \mathbf{a} \in \Theta\}$  sont égales à une constante multiplicative près.*

Preuve :

$$\begin{aligned}
\int f(\boldsymbol{\theta}) \Pi_{\mathbf{a}}^l(d\boldsymbol{\theta}) &= \int f(\boldsymbol{\theta}) H^l(\phi_{\mathbf{a}}^{-1} d\boldsymbol{\theta}) \quad f \in K(\Theta) \\
&= \int f(\phi_{\mathbf{a}} g) H^l(dg) \quad (\text{transformation de mesure : } \boldsymbol{\theta} = \phi_{\mathbf{a}} g) \\
&= \int \Delta(g) f(\phi_{\mathbf{a}} g) H^r(dg) \quad (\text{définition de } H^r \text{ par l'Eq. (9.8)}) \\
&= \int \Delta(g g_0) f(\phi_{\mathbf{a}}(g \circ g_0)) H^r(dg) \quad \forall g_0 \in \mathcal{G} \quad (\text{propriété de } H^r) \\
&= \int \Delta(g g_0) f(g(g_0(\mathbf{a}))) H^r(dg) \quad (\text{définition de } \phi_{\mathbf{a}} \text{ et de l'action de } \mathcal{G}) \\
&= \Delta(\phi_{\mathbf{a}}^{-1} \mathbf{b}) \int \Delta(g) f(\phi_{\mathbf{b}} g) H^r(dg) \quad \text{avec } \mathbf{b} = g_0 \mathbf{a} \quad (\text{et propriété de } \Delta) \\
&= k(\mathbf{a}, \mathbf{b}) \int f(\boldsymbol{\theta}) \Pi_{\mathbf{b}}^l(d\boldsymbol{\theta})
\end{aligned}$$

Comme  $g_0 \rightarrow \mathbf{b}$  définit une bijection entre  $\mathcal{G}$  et  $\Theta$ , toutes les mesures a priori à gauche de Haar  $\Pi_{\mathbf{a}}^l$  pour  $\mathbf{a} \in \Theta$  sont égales à une constante multiplicative près.

## Propriétés des mesures a priori de Haar si le groupe est commutatif ou compact

### • Groupe commutatif

D'après la définition des mesures invariantes à droite et à gauche de Haar on trouve immédiatement que

**Proposition 18** *Si le groupe  $\mathcal{G}$  est commutatif (i.e.  $g g_0 = g_0 g$  quels que soient  $g, g_0 \in \mathcal{G}$ ) alors la classe des mesures à gauche et la classe des mesures à droite de Haar sur  $\mathcal{G}$  sont confondues. Il en va de même pour les mesures a priori à gauche et à droite de Haar induites :  $\Pi^l(A) \propto \Pi^r(A) \quad \forall A \subset \Theta$ .*

C'est le cas en particulier pour les opérations simples d'addition dans  $\mathbb{R}^M$  (invariance par translation) et de multiplication dans  $\mathbb{R}$  (invariance par changement d'échelle). Ainsi, dans les exemples 10 et 18, l'utilisation de la contrainte de l'Eq. (9.1) est alors équivalente à l'utilisation d'une mesure a priori à droite de Haar.

### • Groupe compact

D'autre part, le fait que le groupe  $\mathcal{G}$  soit compact ou non a une conséquence majeure sur la mesure de Haar.

### Exemple 15

- $\mathbb{R}^n$  muni de la loi d'additivité usuelle est localement compact mais non compact.
- $\mathbb{R}_{+*}$  muni de la multiplication est localement compact mais non compact.
- $T_n^+$ , l'ensemble des matrices de taille  $n$  triangulaires inférieures définies positives muni de la multiplication matricielle, est localement compact mais non compact.
- $\mathcal{O}_n$ , l'ensemble des matrices orthogonales, est compact (et donc bien localement compact).

Le groupe est défini jusqu'ici comme un ensemble d'isomorphismes, pas un espace tel que ceux cités ici en exemples. En revanche, il est usuel qu'un tel espace constitue l'espace de paramétrisation du groupe

des isomorphismes. Le fait que  $\mathcal{G}$  soit compact ou non dépend alors des propriétés de la transformation qui relie le groupe des isomorphismes à l'espace de paramétrisation des isomorphismes.

Sinon on peut effectuer des hypothèses sur “l'action” du groupe : en théorie topologique des groupes (à partir de laquelle est définie la mesure de Haar) il est plus usuel de prendre pour  $\mathcal{G}$  un groupe standard tel que ceux cités ci-dessus, puis de définir l'action du groupe  $\mathcal{G}$  sur l'espace  $\Theta$  par l'intermédiaire –en général implicite– d'une fonction  $F$ .

**Définition 14** On dit que  $\mathcal{G}$  agit à gauche de  $\mathcal{X}$  s'il existe une fonction  $F$  de  $\mathcal{G} \times \mathcal{X}$  dans  $\mathcal{X}$  telle que

$$\begin{aligned} F(e, \mathbf{x}) &= \mathbf{x} & \forall \mathbf{x} \in \mathcal{X} \\ F(g_1 g_2, \mathbf{x}) &= F(g_1, F(g_2, \mathbf{x})) & \forall g_1, g_2 \in \mathcal{G}, \mathbf{x} \in \mathcal{X} \end{aligned}$$

$F$  définit l'action (à gauche) de  $\mathcal{G}$  sur  $\mathcal{X}$ .

En général on note plus simplement  $F(g, \mathbf{x}) = (g, \mathbf{x})$  puis  $(g, \mathbf{x}) = g\mathbf{x}$ , de telle sorte que  $g$  semble alors être une transformation comme dans l'approche précédente. Ainsi,  $\mathcal{G}$  représente en général dans la seconde approche l'espace de paramétrisation des endomorphismes de  $\mathcal{G}$  de la première approche. Nous utiliserons plutôt l'approche *action de groupe* afin de bien dissocier la loi de composition interne de son action sur  $\Theta$  (ou  $\mathcal{X}$ ). L'ensemble des matrices inversibles est par exemple un groupe sous la loi de produit matriciel :  $g_A g_B = AB$ . Mais son action sur des matrices de covariance doit souvent être définie comme  $g_A \mathbf{x} = A\mathbf{x}A^t$ ,  $\mathbf{x}$  étant symétrique définie positive. Ceci constitue bien une action au sens de la définition 14.

**Proposition 19** L'espace  $\mathcal{G}$  est de mesure finie sous toute mesure de Haar (i.e.  $H^l(\mathcal{G}) < \infty$  et  $H^r(\mathcal{G}) < \infty$ ) si et seulement si  $\mathcal{G}$  est compact.

Comme les mesures à gauche (ou à droite) de Haar sont reliées à un coefficient multiplicatif près et qu'une mesure à gauche de Haar est reliée à une mesure à droite de Haar par le module  $\Delta$ , il suffit que l'espace  $\mathcal{G}$  soit de mesure finie sous une des mesures pour qu'il soit de mesure finie sous toutes les mesures.

Si  $\mathcal{G}$  est compact on peut donc “normaliser” les mesures de Haar, i.e. choisir les deux mesures de Haar telles que  $\int H^l(dg) = 1$  et  $\int H^r(dg) = 1$ .

Il s'avère de plus que si  $\mathcal{G}$  est compact les mesures invariantes à gauche et à droite sont confondues (preuve dans la proposition 13 de Nachbin [1965] ou dans Eaton [1989]), de telle sorte que

**Proposition 20** Si (et seulement si) le groupe  $\mathcal{G}$  est compact, parmi les mesures invariantes à gauche de Haar il en existe une (unique) qui définit une mesure de probabilité. Il en va de même pour les mesures à droite de Haar. De plus, les mesures à gauche et à droite de Haar sont alors égales, définissant une unique mesure de probabilité  $H^r(A) = H^l(A) \forall A \subset \mathcal{G}$ .

Au contraire, lorsque  $\mathcal{G}$  est non compact  $H^r(\mathcal{G}) = \infty$  et donc  $H^r$  ne peut pas être normalisé et est donc fondamentalement défini à une constante multiplicative près. Comme  $\Pi^r$  est unique pour  $H^r$  fixé (proposition 16),  $\Pi^r$  est donc fondamentalement défini à une constante multiplicative près. D'après la proposition 17, quoique  $\Pi_a^l$  dépende en plus de  $\mathbf{a} \in \Theta$  il en est en fait aussi de même pour  $\Pi^l$ .

**Proposition 21** Si  $\mathcal{G}$  est non compact, les mesures a priori de Haar  $\Pi^r$  et  $\Pi^l$  sur  $\Theta$  ne peuvent être normalisées en des mesures de probabilité et sont donc définis à une constante multiplicative près.

Une loi a priori impropre ne définit pas une mesure de probabilité mais peut représenter la limite d'une famille de mesures de probabilité. D'autre part, la loi a posteriori peut bien être une probabilité même si la loi a priori n'en est pas une. Il suffit pour cela que  $\int p(\mathbf{x}|\boldsymbol{\theta})\Pi(d\boldsymbol{\theta})$  Une mesure a priori qui n'est pas une probabilité n'est donc pas nécessairement problématique en probabilité et statistiques.

**Définition 15** On dit qu'une mesure a priori est **propre** lorsqu'elle définit une mesure de probabilité, et qu'elle est **impropre** sinon.

**Proposition 22** *Les mesures de Haar sur  $\mathcal{G}$  induisent une mesure a priori sur  $\Theta$  qui est propre si et seulement si  $\mathcal{G}$  est compact. De plus, elle est alors unique.*

• **Mesures de Haar à gauche et à droite confondues**

Rappelons pour finir que même si les mesures de Haar ne peuvent être normalisées les classes des mesures invariantes à gauche et à droite peuvent être confondues. On a vu que cela était par exemple le cas si  $\mathcal{G}$  est commutatif. Utilisant finalement notamment la proposition 14 et 15 nous concluons sur ce point que

**Proposition 23** *Les classes des mesures à gauche et à droite de Haar sont égales si et seulement si le module  $\Delta$  de  $\mathcal{G}$ , défini dans la définition 10 p. 145, vaut 1. Cela se produit notamment si  $\mathcal{G}$  est commutatif ou compact.*

**Expression de la densité a priori de Haar dans un cas simple**

Comme dans l'exemple 14 p. 147 et comme nous l'avons mentionné précédemment, une transformation de groupe est en général une fonction dont la forme est connue mais dont un paramètre  $\mathbf{c}$  est inconnu, de telle sorte que le groupe est défini par un ensemble de fonctions de la forme :

$$\mathcal{G} = \{g_{\mathbf{c}}, \mathbf{c} \in C : g_{\mathbf{c}}(\mathbf{x}) = \mathbf{x} + \mathbf{c} \quad \forall \mathbf{x} \in \mathcal{X}\}$$

où  $\mathbf{x} + \mathbf{c}$  peut être remplacé par n'importe quelle fonction de  $\mathbf{x}$ ,  $\mathbf{c}$  et autres paramètres fixés. En théorie topologique des groupes il est usuel de confondre  $\mathcal{G}$  à son espace de paramétrisation  $C$  et de préciser l'action du groupe sur l'espace sur lequel il agit, *i.e.*  $\mathcal{G}$  lui-même,  $\mathcal{X}$  ou  $\Theta$ .

Dans cette sous-section, nous supposons comme Berger [1985] que  $\mathcal{G}$  est isomorphe (et autres propriétés topologiques nécessaires) à un sous-espace de  $\mathbb{R}^n$  de mesure de Lebesgue non nulle.

**Exemple 16** *Dans l'exemple 14, nous avons vu que dans les exemples 12 et 13,  $C = \Theta$  avec dans le premier cas  $\Theta = \mathbb{R}^n$  et  $g_{\mathbf{c}}(\boldsymbol{\theta}) = \boldsymbol{\theta} + \mathbf{c} \quad \forall \boldsymbol{\theta} \in \Theta$  et dans le second cas  $\Theta = \mathbb{R}_{+*}$  et  $g_{\mathbf{c}}(\theta) = c\theta \quad \forall \theta \in \Theta$ .*

Supposant qu'alors toute mesure de Radon (mesure de Haar incluse) est absolument continue par rapport à la mesure de Lebesgue<sup>3</sup>, on peut travailler de façon équivalente sur la densité de Haar (par rapport à la mesure de Lebesgue) afin de réécrire le problème sous une forme intégrale classique, en notant la mesure de Lebesgue  $\mathcal{L}(dg) = dg$ . Ainsi,

**Définition 16** *On appelle densité à gauche de Haar sur le groupe  $\mathcal{G}$  une densité  $h^l$  telle que*

$$\int dg f(g)h^l(g) = \int dg f(g_0g)h^l(g) \quad \forall g_0 \in \mathcal{G}, \forall f \in K(\mathcal{G}) \quad (9.12)$$

**Définition 17** *On appelle densité à droite de Haar sur le groupe  $\mathcal{G}$  une densité  $h^r$  telle que*

$$\int dg f(g)h^r(g) = \int dg f(gg_0)h^r(g) \quad \forall g_0 \in \mathcal{G}, \forall f \in K(\mathcal{G}) \quad (9.13)$$

Pour  $\mathcal{G} \subset \mathbb{R}^n$  il est souvent possible de résoudre l'équation fonctionnelle et ainsi d'obtenir une expression explicite des densités  $h^l$  et  $h^r$ .

Supposons que les transformations  $g \rightarrow g_0g$  (transformation à gauche) et  $g \rightarrow gg_0$  (transformation à droite) sont différentiables. Chacune admet alors un Jacobien, noté respectivement  $J_{g_0}^l$  et  $J_{g_0}^r$ . Restreignant  $f \in K(\mathcal{G})$  aux fonctions indicatrices  $f(g) = I_A(g) \quad \forall g \in \mathcal{G}$  avec  $A$  un sous-ensemble compact de  $\mathcal{G}$ ,

<sup>3</sup>A vérifier, mais cette section n'est quoiqu'il en soit pas utilisée dans le reste de cette thèse.

les contraintes s'écrivent, quel que soit  $g_0 \in \mathcal{G}$ ,

$$\int_A dg h^l(g) = \int_{g_0^{-1}A} dg h^l(g) \quad \forall A \subset \mathcal{G} \quad (9.14)$$

$$\int_A dg h^r(g) = \int_{Ag_0^{-1}} dg h^r(g) \quad \forall A \subset \mathcal{G} \quad (9.15)$$

avec  $g_0^{-1}A = \{g_0^{-1}g, g \in A\}$  et  $Ag_0^{-1} = \{gg_0^{-1}, g \in A\}$ .

Dans ce cas, effectuant sur les membres de droite des Eq. (9.14) et (9.15) les changements de variables respectifs  $g' = g_0g$  et  $g' = gg_0$  on obtient, un peu comme dans la section 9.1 sauf qu'ici l'application différenciée agit sur  $\mathcal{G}$  et non sur  $\Theta$ , les équations fonctionnelles

$$\begin{aligned} h^l(g) &= h^l(g_0^{-1}g) |J_{g_0}^l(g)|^{-1} \quad \forall g \in \mathcal{G}, g_0 \in \mathcal{G} \\ h^r(g) &= h^r(gg_0^{-1}) |J_{g_0}^r(g)|^{-1} \quad \forall g \in \mathcal{G}, g_0 \in \mathcal{G} \end{aligned}$$

En particulier, pour  $g = e$  et reprenant la notation “ $g$ ” pour noter  $g_0^{-1}$  (par exemple), on obtient

$$h^l(e) = h^l(g) |J_g^l(e)| \quad h^r(e) = h^r(g) |J_g^r(e)| \quad \forall g \in \mathcal{G}$$

Les densités à gauche et à droite de Haar sont chacune bien définies à un coefficient multiplicatif près (voir la proposition 10) donc nous avons obtenu l'ensemble des solutions.

**Proposition 24** *Les densités à gauche et à droite de Haar d'un groupe  $\mathcal{G}$  d'élément neutre  $e$  sont données respectivement par*

$$h^l(g) \propto \frac{1}{|J_g^l(e)|} \quad \forall g \in \mathcal{G} \quad (9.16)$$

$$h^r(g) \propto \frac{1}{|J_g^r(e)|} \quad \forall g \in \mathcal{G} \quad (9.17)$$

où  $J_{g_0}^l$  est le Jacobien de la transformation  $g \rightarrow g_0g$  et  $J_{g_0}^r$  est le Jacobien de la transformation  $g \rightarrow gg_0$ .

On obtient ensuite la densité *a priori* induite sur  $\Theta$  comme dans la section 9.2.

**Proposition 25** *Supposant que la transformation  $\phi_{\mathbf{a}}$  définie par l'Eq. (9.11) est bijective, les densités a priori de Haar sont données pour  $\mathbf{a} \in \Theta$  quelconque par*

$$\pi^l(\boldsymbol{\theta}) \propto \frac{1}{|J_{\phi_{\mathbf{a}}^{-1}(\boldsymbol{\theta})}^l(e)|} \quad \forall \boldsymbol{\theta} \in \Theta \quad (9.18)$$

$$\pi^r(\boldsymbol{\theta}) \propto \frac{1}{|J_{\phi_{\mathbf{a}}^{-1}(\boldsymbol{\theta})}^r(e)|} \quad \forall \boldsymbol{\theta} \in \Theta \quad (9.19)$$

**Exemple 17** *Pour la translation simple de l'exemple 16, en utilisant l'approche où  $g \in \mathcal{G}$  agit sur  $\Theta$  plutôt que celle où  $g$  est un endomorphisme, les notations et propositions dérivées ici s'écrivent :*

$$\begin{aligned} g_{\mathbf{c}} &= \mathbf{c} && \text{Notations : “}g\text{” pour rappeler élément de “}g\text{”roupe ; “}c\text{” seul pour raccourcir} \\ g_{\mathbf{c}} \circ g_{\mathbf{c}_0} &= \mathbf{c} + \mathbf{c}_0 && \text{composition du groupe : addition. } \mathcal{G} \text{ commutatif} \\ J_{g_0}^l(g_{\mathbf{c}}) &= 1 && \text{Jacobien de la transformation } g \rightarrow g_0g \\ F(g_{\mathbf{c}}, \boldsymbol{\theta}) &= g_{\mathbf{c}}\boldsymbol{\theta} = \mathbf{c} + \boldsymbol{\theta} && \Theta = \mathcal{G} ; \text{ action du groupe sur } \Theta : \text{ addition à nouveau} \\ \phi_{\mathbf{a}}(g_{\mathbf{c}}) &= g_{\mathbf{c}}\mathbf{a} = \mathbf{a} + \mathbf{c} && \mathbf{a} \in \Theta \\ \phi_{\mathbf{a}}^{-1}(\boldsymbol{\theta}) &= g_{\mathbf{b}} \Rightarrow \boldsymbol{\theta} = \phi_{\mathbf{a}}(g_{\mathbf{b}}) \Rightarrow \mathbf{b} = \boldsymbol{\theta} - \mathbf{a} \Rightarrow \phi_{\mathbf{a}}^{-1}(\boldsymbol{\theta}) = g_{\boldsymbol{\theta} - \mathbf{a}} \\ \pi^r(\boldsymbol{\theta}) &\propto \pi^l(\boldsymbol{\theta}) \propto \frac{1}{|J_{\phi_{\mathbf{a}}^{-1}(\boldsymbol{\theta})}^l(e)|} \propto \frac{1}{|J_{g_{\boldsymbol{\theta} - \mathbf{a}}}^l(e)|} \propto 1 \end{aligned}$$

Comme dans le cas général, si  $\mathcal{G}$  est non compact, les densités *a priori* de Haar  $\pi^r$  et  $\pi^l$  sur  $\Theta$  sont définies à une constante multiplicative près. Comme elles ne peuvent être normalisées en des mesures de probabilité on dit que ces densités *a priori* sont impropres.

Si  $\mathcal{G}$  est compact, la constante de proportionnalité des densités *a priori* données ci-dessus est obtenue par normalisation de la loi de probabilité (de telle sorte que la densité *a priori* est propre) et de plus on a alors  $\pi^l(\theta) = \pi^r(\theta) \forall \theta \in \Theta$  définissant une unique densité de probabilité.

### 9.3 Famille de vraisemblances invariante sous un groupe de transformations

Le choix du groupe de transformations  $\mathcal{G}$  sous lequel on souhaite imposer une invariance apparaît souvent naturellement à partir de la famille de vraisemblances elle-même si celle-ci est invariante sous un groupe de transformations et qu'il n'y a par ailleurs pas d'information *a priori* particulière.

Au début de l'étude d'un modèle statistique, on peut parfois constater l'existence d'une transformation qui appliquée aux observations  $\mathbf{X}$  mène dans la réalité et donc devrait mener à travers le modèle mathématique à une situation invariante si l'on ajuste convenablement les paramètres.

Initialement, on part donc d'un groupe  $\mathcal{G}$  qui agit sur un espace des données  $\mathcal{X}$ . Si on note  $P$  la mesure de probabilité sur  $\mathcal{X}$ , de telle sorte que  $\Pr(\mathbf{X} \in A) = P(A) \forall A \subset \mathcal{X}$  il est alors commode de définir l'action  $gP$  de  $\mathcal{G}$  sur l'espace des mesures de probabilité par

$$gP(A) = P(g^{-1}A) \quad \forall g \in \mathcal{G}, A \subset \mathcal{X}$$

On dit que l'action de  $\mathcal{G}$  sur  $\mathcal{X}$  induit une action de  $\mathcal{G}$  sur les mesures de probabilité. Voir Eaton [1989] pour plus de détails.

**Définition 18** *Un modèle statistique  $\mathcal{P} = \{P_1, P_2, \dots\}$  (ou famille de mesures de probabilité), est dit invariant sous un groupe de transformations  $\mathcal{G}$  si  $P \in \mathcal{P}$  implique  $gP \in \mathcal{P}$  quel que soit  $g \in \mathcal{G}$ .*

Dans un cadre paramétrique le modèle  $\mathcal{P}$  s'écrit  $\mathcal{P}_\Theta = \{P_\theta, \theta \in \Theta\}$ . L'invariance d'un modèle statistique implique alors qu'il existe  $\theta^* \in \Theta$  tel que  $gP_\theta = P_{\theta^*}$  : dire qu'une action sur  $\mathbf{X}$  laisse invariante certaines propriétés du problème "si l'on ajuste convenablement les paramètres" signifie que l'action de  $\mathcal{G}$  sur  $\mathcal{X}$  doit induire une action de  $\mathcal{G}$  sur  $\Theta$ . Il est alors commode d'induire une action de  $\mathcal{G}$  sur  $\Theta$  à partir de l'action de  $\mathcal{G}$  sur  $\mathbf{X}$  et l'invariance du modèle statistique. On peut par exemple définir pour un  $g \in \mathcal{G}$  fixé quelconque,

$$gP_\theta = P_{g\theta} \quad \forall \theta \in \Theta \tag{9.20}$$

On aurait aussi pu définir l'action induite de  $\mathcal{G}$  sur  $\Theta$  par exemple par  $gP_{g\theta} = P_\theta$  et les deux actions sont alors directement reliées par une relation inverse ( $g_2 = g_1^{-1}$ ).

Pour une famille de densités de probabilité par rapport à une mesure quelconque sur  $\mathcal{X}$  l'invariance s'exprime de la façon suivante :

**Définition 19** *La famille  $\mathcal{P}_\Theta = \{p(\cdot|\theta), \theta \in \Theta\}$  de densités par rapport à une mesure  $\mu$  quelconque sur  $\mathcal{X}$  est dite invariante sous un groupe  $\mathcal{G}$  de transformations si pour tout  $g \in \mathcal{G}$  il existe un unique  $\theta^* \in \Theta$  tel que si  $\mathbf{X}$  a pour densité  $p(\cdot|\theta) \in \mathcal{P}_\Theta$  alors  $\mathbf{Y} = g\mathbf{X}$  a pour densité  $p(\cdot|\theta^*) \in \mathcal{P}_\Theta$ .*

**Proposition 26** *Si la famille  $\mathcal{P}_\Theta = \{p(\cdot|\theta), \theta \in \Theta\}$  de densités par rapport à une mesure  $\mu$  quelconque sur  $\mathcal{X}$  est invariante sous  $\mathcal{G}$  alors le modèle statistique  $\mathcal{P}_\Theta = \{P_\theta, \theta \in \Theta : P_\theta(d\mathbf{x}) = p(\mathbf{x}|\theta)\mu(d\mathbf{x})\}$  est invariant sous  $\mathcal{G}$ .<sup>4</sup>*

<sup>4</sup>La réciproque est vraie "à un ensemble de mesure nulle près", voir Eaton [1989] pour un énoncé détaillé de la réciproque.

**Proposition 27** Soit  $\mathcal{P}_\Theta = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  une famille de densités par rapport à la mesure de Lebesgue. Plaçons-nous dans le cadre de la sous-section de la page 150, c'est-à-dire supposons que  $\mathcal{G}$  est isomorphe à un sous-espace de  $\mathbb{R}^n$  de mesure de Lebesgue non nulle. Si  $\mathcal{P}_\Theta$  est invariante sous  $\mathcal{G}$  et que  $gP_\theta = P_{g\theta}$  (avec  $P_\theta(d\mathbf{x}) = p(\cdot|\boldsymbol{\theta})d\mathbf{x}$ ) alors

$$p(\mathbf{x}|\boldsymbol{\theta}) = |J_g(\mathbf{x})|p(g\mathbf{x}|g\boldsymbol{\theta}) \quad \forall \mathbf{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta, g \in \mathcal{G} \quad (9.21)$$

où  $J_g(\mathbf{x})$  est le Jacobien –dont on suppose qu'il existe– de la transformation  $\mathbf{x} \rightarrow g\mathbf{x}$  appliqué en  $\mathbf{x}$ .

Notons que pour  $g \in \mathcal{G}$  fixé quelconque, si on note  $\boldsymbol{\theta}' = g\boldsymbol{\theta}$  alors  $\boldsymbol{\theta} = g^{-1}\boldsymbol{\theta}'$  et comme de plus  $\{g\boldsymbol{\theta}, \boldsymbol{\theta} \in \Theta\} = \Theta$ ,

$$gP_\theta = P_{g\theta} \quad \forall \boldsymbol{\theta} \in \Theta \quad \Leftrightarrow \quad gP_{g^{-1}\boldsymbol{\theta}} = P_\theta \quad \forall \boldsymbol{\theta} \in \Theta$$

de telle sorte que si  $gP_\theta = P_{g\theta}$  on a aussi

$$p(\mathbf{x}|g^{-1}\boldsymbol{\theta}) = |J_g(\mathbf{x})|p(g\mathbf{x}|\boldsymbol{\theta}) \quad \forall \mathbf{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta, g \in \mathcal{G}$$

**Exemple 18** La famille des lois normales à matrice de covariance connue est invariante par translation. Ceci s'écrit de la façon suivante : la famille de densités (par rapport à la mesure de Lebesgue)  $\mathcal{P}_\Theta = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  avec  $p(\cdot|\boldsymbol{\theta}) = \mathcal{N}_L(\boldsymbol{\theta}, \Sigma)$  et  $\Sigma$  connue est invariante sous le groupe des translations, où le groupe de translations est  $\mathcal{G} = \{g_{\mathbf{a}}, \mathbf{a} \in \mathbb{R}^L : g_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a} \quad \forall \mathbf{x} \in \mathbb{R}^L\}$ . Comme expliqué dans la section 9.2 et utilisé ci-dessus, on note plus simplement ce groupe de transformations comme le groupe usuel  $\mathcal{G} = \mathbb{R}^L$  avec l'action de  $\mathcal{G}$  sur  $\mathcal{X}$  donnée par  $g_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} + \mathbf{a}$ .

La famille est bien invariante car si  $\mathbf{X} \sim \mathcal{N}_L(\boldsymbol{\theta}, \Sigma)$ , alors comme la transformation  $g_{\mathbf{a}}$  est bijective la distribution de  $\mathbf{Y} = \mathbf{X} + \mathbf{a}$  est donnée par

$$p_{\mathbf{Y}}(\mathbf{y}|\boldsymbol{\theta}) = p_{\mathbf{X}}(\mathbf{y} - \mathbf{a}|\boldsymbol{\theta})|J_{g_{\mathbf{a}}}(\mathbf{y} - \mathbf{a})|^{-1} = \frac{1}{(2\pi)^{\frac{L}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(\mathbf{y} - \mathbf{a} - \boldsymbol{\theta})^t \Sigma^{-1}(\mathbf{y} - \mathbf{a} - \boldsymbol{\theta})}{2}\right) = p_{\mathbf{X}}(\mathbf{y}|\boldsymbol{\theta} + \mathbf{a})$$

où  $J_{g_{\mathbf{a}}}$  est le Jacobien de la transformation  $g_{\mathbf{a}}$  et vaut 1. On constate que la distribution de  $\mathbf{Y}$  est donnée par  $\mathbf{Y} \sim \mathcal{N}_L(\boldsymbol{\theta}^*, \Sigma)$  avec  $\boldsymbol{\theta}^* = \boldsymbol{\theta} + \mathbf{a} = g_{\mathbf{a}}\boldsymbol{\theta}$ . Les actions de  $\mathcal{G}$  sur  $\mathcal{X}$  et sur  $\Theta$  sont identiques. Ils (ou plus précisément leur espace de paramétrisation) sont aussi confondus avec  $\mathcal{X} = \mathbb{R}^L$  et  $\Theta = \mathbb{R}^L$ .

Ainsi, le groupe sous lequel la famille de vraisemblances est invariante induit le groupe sous lequel  $\boldsymbol{\theta}$  doit être modifié afin de laisser effectivement la mesure sur  $\mathcal{X}$  invariante. Comme présenté dans la section 9.2, on note  $\mathcal{G}$  le groupe “de base” et on spécifie seulement son action sur les différents espaces.

En l'absence d'autres informations *a priori* il est alors naturel et souvent recommandé de prendre une loi *a priori* sur  $\Theta$  induite par la mesure de Haar de ce groupe  $\mathcal{G}$ , et plus spécifiquement de la mesure à droite de Haar. Pour un modèle statistique invariant, Eaton and Sudderth [2002] montre par exemple que le seul *a priori* qui n'induit pas une inférence invariante qui soit inconsistente est l'*a priori* droit de Haar.

**Exemple 19** Pour la famille de vraisemblances invariante par translation de l'exemple 18, si le processus d'acquisition des données ne fait pas apparaître de choix d'origine spécifique (définition des mesures  $\mathbf{x}$  à une constante additive arbitraire près) il est naturel de définir comme loi *a priori* sur  $\boldsymbol{\theta}$  la loi correspondant au groupe de translations, sous lequel la famille de vraisemblances est invariante. Comme l'opération d'addition est commutative, l'*a priori* droit de Haar peut être obtenu simplement à partir de l'Eq. 9.18, voir l'exemple 17. Cet *a priori* est la loi uniforme  $\pi(\boldsymbol{\theta}) \propto 1$  et est impropre.

Dans le cadre de l'invariance d'un groupe topologique il n'est pas nécessaire que la transformation  $\mathbf{x} \rightarrow g\mathbf{x}$  soit différentiable. Ainsi, dans la proposition 27 on peut remplacer le Jacobien par le module d'une mesure sur  $\mathcal{X}$  dite relativement invariante sous l'action de  $g \in \mathcal{G}$ .



**Définition 20** Une mesure  $\mu$  sur  $\mathcal{X}$  est dite **relativement invariante de module**  $\chi$  sous le groupe  $\mathcal{G}$  si

$$\int f(\mathbf{x})\mu(d\mathbf{x}) = \chi(g) \int f(g\mathbf{x})\mu(d\mathbf{x}) \quad \forall f \in K(\mathcal{X}), g \in \mathcal{G} \quad (9.22)$$

Weil donne des conditions sur l'existence et l'unicité d'une telle mesure (voir Nachbin [1965] p. 138 par exemple) qui permet de relier ce module au module  $\Delta$  présenté plus haut dans le cadre des mesures sur  $\mathcal{G}$ . Nous considérerons ce problème plus tard car il fait appel à des notions pas encore présentées ici. Notons simplement que contrairement au Jacobien qui dépend de  $\mathbf{x}$  et est donc intégré,  $\chi(g)$  est ici à l'extérieur de l'intégrale.

Dans ce cadre, la définition 18 implique :

**Proposition 28** Soit  $\mathcal{P}_\Theta = \{p(\cdot|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  une famille de densités par rapport à une mesure  $\mu$  sur  $\mathcal{X}$  relativement invariante sous  $\mathcal{G}$ . Si  $\mathcal{P}_\Theta$  est invariante sous  $\mathcal{G}$  et que  $gP_\Theta = P_{g\boldsymbol{\theta}}$  (avec  $P_\Theta(d\mathbf{x}) = p(\cdot|\boldsymbol{\theta})\mu(d\mathbf{x})$ ) alors

$$p(\mathbf{x}|\boldsymbol{\theta}) = \chi(g)p(g\mathbf{x}|g\boldsymbol{\theta}) \quad \forall \mathbf{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta, g \in \mathcal{G} \quad (9.23)$$

où  $\chi$  est le module associé à la mesure relativement invariante  $\mu$ .

Nous avons brossé rapidement les définitions et résultats principaux de l'invariance de groupe (topologique localement compact) et son application aux statistiques via les notions de mesure induite et d'action d'un groupe sur les espaces  $\mathcal{X}$  et  $\Theta$ . Cette compilation sert de base aux résultats démontrés dans le chapitre suivant, avec le même niveau de rigueur et de détail que dans ce chapitre.

## Chapitre 10

# En cas invariant, équivalence entre le PLR et une *p-value* fréquentiste

Une introduction à l'application de l'invariance de groupe aux statistiques est présentée dans le chapitre précédent (chapitre 9). Elle présente spécifiquement les définitions et propriétés utilisées pour obtenir les résultats suivants.

Les termes indicés par 0 ne sont jamais des variables aléatoires : ce sont des grandeurs déterministes fixées. Ainsi, on utilise dans le déroulement du calcul  $\mathbf{x}_0$  représentant les données observées et  $\boldsymbol{\theta}_0$  la valeur de paramètre fixée dans la définition de l'hypothèse nulle. La conditionnalité à  $\mathbf{x}_0$  ou  $\boldsymbol{\theta}_0$  dans les expressions de probabilités  $\Pr(\cdot|\mathbf{x}_0)$  ou  $\Pr(\cdot|\boldsymbol{\theta}_0)$  indique que la distribution prise en compte est respectivement la distribution *a posteriori* basée sur la variable aléatoire  $\boldsymbol{\theta}|\mathbf{x}_0$  ou la distribution fréquentiste basée sur la variable aléatoire  $\mathbf{x}|\boldsymbol{\theta}_0$ .

### 10.1 Théorème

#### Hypothèses simples

- Hypothèse forte sur  $\mathcal{X}$

Soit  $\mathcal{P}_\Theta = \{P_\theta, \theta \in \Theta\}$  un modèle statistique sur  $\mathcal{X}$  invariant sous un groupe  $\mathcal{G}$  tel qu'introduit dans la section 9.2. Comme décrit dans la section 9.3, on définit l'action de  $\mathcal{G}$  sur  $\Theta$  à partir de l'action de  $\mathcal{G}$  sur  $\mathcal{X}$  et l'invariance du modèle statistique :

$$gP_{g\theta} = P_\theta \quad (10.1)$$

Ainsi, on suppose que tous les endomorphismes

$$\text{"}g\text{"} : \mathcal{X} \rightarrow \mathcal{X} \quad (10.2)$$

$$\mathbf{x} \rightarrow g\mathbf{x}$$

$$\text{"}g\text{"} : \Theta \rightarrow \Theta \quad (10.3)$$

$$\boldsymbol{\theta} \rightarrow g\boldsymbol{\theta}$$

avec  $g \in \mathcal{G}$  sont bijectifs<sup>1</sup>.

---

<sup>1</sup>A strictement parler on devrait appeler les endomorphismes autrement et dire qu'appliqué à  $\mathbf{x}$  (respectivement  $\boldsymbol{\theta}$ ) l'endomorphisme sur  $\mathcal{X}$  (resp.  $\Theta$ ) tel que défini ci-dessus est à égal  $g$  agissant sur  $\mathbf{x}$  (resp.  $\boldsymbol{\theta}$ ).

On suppose de plus que  $\mathcal{X}$ ,  $\Theta$  et  $\mathcal{G}$  sont isomorphes. En particulier, on suppose que les applications

$$\phi_{\mathbf{x}} : \mathcal{G} \rightarrow \mathcal{X} \quad (10.4)$$

$$g \rightarrow g\mathbf{x}$$

$$\phi_{\boldsymbol{\theta}} : \mathcal{G} \rightarrow \Theta \quad (10.5)$$

$$g \rightarrow g\boldsymbol{\theta}$$

sont bijectives quels que soient  $\mathbf{x} \in \mathcal{X}$  et  $\boldsymbol{\theta} \in \Theta$ . On exige ainsi que  $\mathcal{X}$  et  $\Theta$  soient isomorphes, une condition restrictive qui sera assouplie dans un second temps en remplaçant  $\mathbf{X}$  par une statistique exhaustive  $S(\mathbf{X})$ .

Soit  $H^r$  une mesure invariante à droite de Haar sur  $\mathcal{G}$ , normalisée en une mesure de probabilité si  $\mathcal{G}$  est compact, voir la section 9.2. Soit  $\Delta$  le module de Haar sur  $\mathcal{G}$  défini dans la définition 10 p. 145.

Soit  $H^l$  la mesure définie par

$$H^l(dg) = \Delta(g)H^r(dg) \quad (10.6)$$

$H^l$  est bien une mesure de Haar invariante à gauche d'après la proposition 15, p. 146.

On définit la mesure *a priori* sur  $\Theta$  comme dans la section 9.2 mais aussi une mesure sur  $\mathcal{X}$  : on note  $\mu$  et  $\Pi^r$  les mesures induites par  $H^r$  respectivement sur  $\mathcal{X}$  par  $\phi_{\mathbf{a}}$  et sur  $\Theta$  par  $\phi_{\mathbf{b}}$ , avec  $\mathbf{a} \in \mathcal{X}$  et  $\mathbf{b} \in \Theta$  avec

$$\mu(A) = H^r(\phi_{\mathbf{a}}^{-1}(A)) \quad \forall A \subset \mathcal{X} \quad (10.7)$$

$$\Pi^r(B) = H^r(\phi_{\mathbf{b}}^{-1}(B)) \quad \forall B \subset \Theta \quad (10.8)$$

Lorsque c'est possible, *i.e.* que  $\mathcal{G}$  est compact, la mesure *a priori*  $\Pi^r$  est bien normalisée en une mesure de probabilité car  $H^r$  a été choisie normalisée. Il n'est pas nécessaire d'introduire de coefficient de normalisation de  $\mu$  différent. Comme nous l'avons vu dans la section 9.2 ces mesures ne dépendent pas du choix de  $\mathbf{a} \in \mathcal{X}$  et  $\mathbf{b} \in \Theta$ .

Soit  $\mathcal{P}_{\Theta} = \{P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta\}$  une famille de mesures de probabilité sur  $\mathcal{X}$ . Supposons qu'elle est invariante sous  $\mathcal{G}$  avec  $gP_{\boldsymbol{\theta}} = P_{g\boldsymbol{\theta}}$ . Supposons de plus que chaque probabilité  $P_{\boldsymbol{\theta}}$  est absolument continue par rapport à  $\mu$  et notons  $p(\cdot|\boldsymbol{\theta})$  la densité associée :

$$P_{\boldsymbol{\theta}}(d\mathbf{x}) = p(\mathbf{x}|\boldsymbol{\theta})\mu(d\mathbf{x}) \quad (10.9)$$

Soit  $\Pi_{\mathbf{x}}^r$  la mesure *a posteriori* sur  $\Theta$  définie par

$$\Pi_{\mathbf{x}}^r(B) = \frac{\int_B p(\mathbf{x}|\boldsymbol{\theta})\Pi^r(d\boldsymbol{\theta})}{m(\mathbf{x})} \quad B \subset \Theta$$

$$\text{avec } m(\mathbf{x}) = \int p(\mathbf{x}|\boldsymbol{\theta})\Pi^r(d\boldsymbol{\theta})$$

On suppose que  $m(\mathbf{x}) < \infty$  de telle sorte que  $\Pi_{\mathbf{x}}^r$  définit bien une mesure de probabilité et on note  $\Pr(\cdot|\mathbf{x})$  la probabilité conditionnelle à  $\mathbf{x}$  :  $\Pr(\boldsymbol{\theta} \in B|\mathbf{x}) = \Pi_{\mathbf{x}}^r(B)$ .

**Lemme 1** *Les mesures  $\mu$  sur  $\mathcal{X}$  et  $\Pi^r$  sur  $\Theta$  induites ci-dessus par la mesure à droite de Haar  $H^r$  sur  $\mathcal{G}$  sont relativement invariantes de module  $\Delta^{-1}$ .*

Preuve : montrons le lemme pour  $\mu$ . La preuve pour  $\Pi^r$  est identique. D'après la définition 20 p. 154, une mesure sur  $\mathcal{X}$  est relativement invariante de module  $\chi$  sur  $\mathcal{G}$  si

$$\int f(\mathbf{x})\mu(d\mathbf{x}) = \chi(g) \int f(g\mathbf{x})\mu(d\mathbf{x})$$

Or

$$\begin{aligned}
\int f(g_0 \mathbf{x}) \mu(d\mathbf{x}) &= \int f(g_0 \mathbf{x}) H^r \phi_{\mathbf{a}}^{-1}(d\mathbf{x}) \quad (\text{Def. } \mu \text{ Eq. (10.7)}) \\
&= \int f(g_0 \phi_{\mathbf{a}} g) H^r(dg) \quad (\text{transformation de mesure : } g = \phi_{\mathbf{a}}^{-1} \mathbf{x}) \\
&= \int f(g_0 g \mathbf{a}) \Delta(g)^{-1} H^l(dg) \quad (\text{Def. de } \phi_{\mathbf{a}} \text{ et Def. de } H^l) \\
&= \Delta(g_0) \int f(g_0 g \mathbf{a}) \Delta(g_0 g)^{-1} H^l(dg) \quad (\text{Prop. 12 p. 145}) \\
&= \Delta(g_0) \int f(g \mathbf{a}) \Delta(g)^{-1} H^l(dg) \quad (H^l \text{ invariante à gauche}) \\
&= \Delta(g_0) \int f(\mathbf{x}) \mu(d\mathbf{x}) \quad (\text{calcul précédent réeffectué dans l'autre sens})
\end{aligned}$$

d'où le résultat.

**Corrolaire 6** *Un a priori de Haar induit comme ci-dessus (i.e. comme dans la section 9.2) à partir d'une mesure invariante à droite de Haar est relativement invariant.*

Notons d'après la proposition 23 p. 150 que cet *a priori* est invariant (définition 20 p. 154 avec  $\chi = 1$ ) si et seulement si  $\Delta = 1$ , c'est-à-dire si les classes des mesures à droite et à gauche de Haar sont confondues, ce qui se produit si  $\mathcal{G}$  est commutatif ou compact.

Dans le cadre fixé ci-dessus, le PLR (égal à une intégrale sur  $\Theta$ ) se réexprime comme une intégrale fréquentiste (intégrale sur  $\mathcal{X}$ ). On rappelle que  $\mathbf{x}$  et  $\boldsymbol{\theta}$  représentent dans ce chapitre des variables aléatoires ou des variables classiques et qu'on indice  $\mathbf{x}$  et  $\boldsymbol{\theta}$  par 0 lorsqu'elles prennent des valeurs fixées<sup>2</sup>.

**Théorème 4** *Soit  $\mathcal{P}_{\Theta} = \{p_{\mathcal{X}|\boldsymbol{\theta}}(\cdot), \boldsymbol{\theta} \in \Theta\}$  une famille de densités de probabilité par rapport à une mesure  $\mu$  sur  $\mathcal{X}$ . Supposons que*

1.  $\mathcal{P}_{\Theta}$  est invariante sous l'action du groupe  $\mathcal{G}$  sur  $\mathcal{X}$ .
2.  $\phi_{\mathbf{x}}$  et  $\phi_{\boldsymbol{\theta}}$  définis par les Eq. (10.4) et (10.5) sont bijectives quels que soit  $\boldsymbol{\theta} \in \Theta$  et  $\mathbf{x} \in \mathcal{X}$ .
3. La mesure a priori sur  $\Theta$  est la mesure induite par  $H^r$  via  $\phi_{\boldsymbol{\theta}}$ .
4. La mesure  $\mu$  sur  $\mathcal{X}$  est la mesure induite par  $H^r$  via  $\phi_{\mathbf{x}}$ .

Le PLR défini par l'Eq. (7.14) s'exprime alors, avec  $\mathbf{x}_0 \in \mathcal{X}$  (représentant les données observées),  $\boldsymbol{\theta}_0 \in \Theta$  et  $\zeta > 0$  et pour  $\mathbf{c} \in \mathcal{X}$  quelconque, comme l'intégrale fréquentiste :

$$PLR(\mathbf{x}_0, \zeta) = Pr\left( p(\mathbf{x}_0|\boldsymbol{\theta}_0) \Delta(\phi_{\mathbf{x}_0}^{-1} \mathbf{c}) \leq \zeta p(\mathbf{x}|\boldsymbol{\theta}_0) \Delta(\phi_{\mathbf{x}}^{-1} \mathbf{c}) \mid \boldsymbol{\theta}_0 \right) \quad (10.10)$$

avec  $\Delta$  le module du groupe  $\mathcal{G}$ , défini dans la définition 10 p. 145.

<sup>2</sup> $\mathbf{x}_0$  représente les données mesurées et  $\boldsymbol{\theta}_0$  la valeur du paramètre testé sous  $H_0$ . Ainsi, la conditionnalité à  $\mathbf{x}_0$  ou  $\boldsymbol{\theta}_0$  dans les expressions de probabilités  $Pr(\cdot|\mathbf{x}_0)$  ou  $Pr(\cdot|\boldsymbol{\theta}_0)$  indique que la distribution prise en compte est respectivement la distribution *a posteriori* basée sur la variable aléatoire  $\boldsymbol{\theta}|\mathbf{x}_0$  ou la distribution fréquentiste basée sur la variable aléatoire  $\mathbf{x}|\boldsymbol{\theta}_0$ .

Preuve :

$$\begin{aligned}
\text{PLR}(\mathbf{x}_0, \zeta) &= \Pr\left(p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|\boldsymbol{\theta}) \mid \mathbf{x}_0\right) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{\boldsymbol{\theta}: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|\boldsymbol{\theta})\}} p(\mathbf{x}_0|\boldsymbol{\theta}) \Pi^r(d\boldsymbol{\theta}) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{\boldsymbol{\theta}: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|\boldsymbol{\theta})\}} p(\mathbf{x}_0|\boldsymbol{\theta}) H^r(\phi_{\mathbf{b}}^{-1}(d\boldsymbol{\theta})) \quad (\text{Def. } \Pi^r \text{ Eq. (10.8)}) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{g: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|\phi_{\mathbf{b}}g)\}} p(\mathbf{x}_0|\phi_{\mathbf{b}}g) H^r(dg) \quad (g = \phi_{\mathbf{b}}^{-1}\boldsymbol{\theta}) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{g: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|g\mathbf{b})\}} p(\mathbf{x}_0|g\mathbf{b}) H^r(dg) \quad (\text{Def. } \phi_{\boldsymbol{\theta}} \text{ Eq. (10.5)}) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{g: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|g\mathbf{b})\}} p(\mathbf{x}_0|g\mathbf{b}) \Delta(g)^{-1} H^l(dg) \quad (\text{Def. } H^l \text{ Eq. (10.6)}) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{g: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|g^{-1}\mathbf{b})\}} p(\mathbf{x}_0|g^{-1}\mathbf{b}) H^l(dg) \quad (\text{Prop. 13, p. 146})
\end{aligned} \tag{10.11}$$

Or d'après le lemme 1,  $\mu$  est relativement invariant de module  $\Delta^{-1}$ . Comme par hypothèse la famille de vraisemblances est invariante, d'après la proposition 28 p. 154

$$p(\mathbf{x}|\boldsymbol{\theta}) = \Delta(g)^{-1} p(g\mathbf{x}|g\boldsymbol{\theta}) \quad \forall \mathbf{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta, g \in \mathcal{G}$$

ou de façon équivalente

$$p(\mathbf{x}|g^{-1}\boldsymbol{\theta}) = \Delta(g)^{-1} p(g\mathbf{x}|\boldsymbol{\theta}) \quad \forall \mathbf{x} \in \mathcal{X}, \boldsymbol{\theta} \in \Theta, g \in \mathcal{G}$$

Ainsi,

$$\begin{aligned}
\text{PLR}(\mathbf{x}_0, \zeta) &= \frac{1}{m(\mathbf{x}_0)} \int_{\{g: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(g\mathbf{x}_0|\mathbf{b})\Delta(g)^{-1}\}} \Delta(g)^{-1} p(g\mathbf{x}_0|\mathbf{b}) H^l(dg) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{g: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(g\mathbf{x}_0|\mathbf{b})\Delta(g)^{-1}\}} p(g\mathbf{x}_0|\mathbf{b}) H^r(dg) \quad (\text{Def. } H^l) \\
&= \frac{1}{m(\mathbf{x}_0)} \int_{\{g: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(g\mathbf{x}_0|\mathbf{b})\Delta(g)^{-1}\}} p(g\mathbf{x}_0|\mathbf{b}) \mu(\phi_{\mathbf{a}}(dg)) \quad (\text{Def. } \mu)
\end{aligned}$$

On constate que l'Eq. (10.11) ne dépend ni de  $\mathbf{a} \in \mathcal{X}$  ni de  $\mathbf{b} \in \Theta$ . Choisissons dès maintenant par commodité  $\mathbf{a} = \mathbf{x}_0$ . Dans ce cas on trouve simplement

$$\Pr\left(p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}_0|\boldsymbol{\theta}) \mid \mathbf{x}_0, \boldsymbol{\theta}_0\right) = \frac{1}{m(\mathbf{x}_0)} \int_{\{\mathbf{x}: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}|\mathbf{b})\Delta(\phi_{\mathbf{x}_0}^{-1}\mathbf{x})^{-1}\}} p(\mathbf{x}|\mathbf{b}) \mu(d\mathbf{x}) \quad (\mathbf{x} = \phi_{\mathbf{x}_0}g = g\mathbf{x}_0)$$

Par le même calcul on constate que

$$m(\mathbf{x}_0) = \int p(\mathbf{x}_0|\boldsymbol{\theta}) \Pi^r(d\boldsymbol{\theta}) = \int p(\mathbf{x}|\mathbf{b}) \mu(d\mathbf{x}) = 1$$

La densité marginale de  $\mathbf{X}$  ne dépend pas de  $\mathbf{x}$ , de même que le risque fréquentiste d'un estimateur invariant ne dépend pas de  $\boldsymbol{\theta}$ . Donc

$$\text{PLR}(\mathbf{x}_0, \zeta) = \int_{\{\mathbf{x}: p(\mathbf{x}_0|\boldsymbol{\theta}_0) \leq \zeta p(\mathbf{x}|\mathbf{b})\Delta(\phi_{\mathbf{x}_0}^{-1}\mathbf{x})^{-1}\}} p(\mathbf{x}|\mathbf{b}) \mu(d\mathbf{x})$$

Afin de se ramener à une forme plus proche d'une  $p$ -value, on termine le calcul en choisissant  $\mathbf{b} = \boldsymbol{\theta}_0$  et en notant que pour un  $\mathbf{c} \in \mathcal{X}$  quelconque

$$\Delta(\phi_{\mathbf{x}_0}^{-1}\mathbf{x}) = \frac{\Delta(\phi_{\mathbf{c}}^{-1}\mathbf{x})}{\Delta(\phi_{\mathbf{c}}^{-1}\mathbf{x}_0)} \quad (10.12)$$

En effet, si on note

$$\begin{aligned} g &= \phi_{\mathbf{x}_0}^{-1}\mathbf{x} \\ g_1 &= \phi_{\mathbf{c}}^{-1}\mathbf{x} \\ g_2 &= \phi_{\mathbf{c}}^{-1}\mathbf{x}_0 \end{aligned}$$

alors d'un côté  $g\mathbf{x}_0 = \mathbf{x}$  et de l'autre  $g_1(g_2^{-1}\mathbf{x}_0) = g_1\mathbf{c} = \mathbf{x}$  de telle sorte que

$$\begin{aligned} g\mathbf{x}_0 &= (g_1g_2^{-1})\mathbf{x}_0 \\ \text{donc } \phi_{\mathbf{x}_0}g &= \phi_{\mathbf{x}_0}(g_1g_2^{-1}) \\ \text{donc } g &= g_1g_2^{-1} \quad (\text{bijectivité de } \phi_{\mathbf{a}}) \\ \text{donc } \Delta(g) &= \frac{\Delta(g_1)}{\Delta(g_2)} \quad (\text{Prop. 12 p. 145}) \end{aligned}$$

Ainsi, finalement, pour  $\mathbf{c} \in \mathcal{X}$  quelconque

$$\text{PLR}(\mathbf{x}_0, \zeta) = \int_{\left\{ \mathbf{x}: \frac{p(\mathbf{x}_0|\boldsymbol{\theta}_0)}{\Delta(\phi_{\mathbf{c}}^{-1}\mathbf{x}_0)} \leq \zeta \frac{p(\mathbf{x}|\boldsymbol{\theta}_0)}{\Delta(\phi_{\mathbf{c}}^{-1}\mathbf{x})} \right\}} p(\mathbf{x}|\boldsymbol{\theta}_0)\mu(d\mathbf{x}) \quad (10.13)$$

Il est de plus intéressant de remarquer que

$$\begin{aligned} \phi_{\mathbf{a}}^{-1}\mathbf{b} &= (\phi_{\mathbf{b}}^{-1}\mathbf{a})^{-1} \\ \text{car } g = \phi_{\mathbf{a}}^{-1}\mathbf{b} &\Rightarrow g\mathbf{a} = \mathbf{b} \Rightarrow \mathbf{a} = g^{-1}\mathbf{b} \Rightarrow g^{-1} = \phi_{\mathbf{b}}^{-1}\mathbf{a} \end{aligned} \quad (10.14)$$

de telle sorte que de la même façon on a

$$\begin{aligned} \text{PLR}(\mathbf{x}_0, \zeta) &= \int_{\left\{ \mathbf{x}: p(\mathbf{x}_0|\boldsymbol{\theta}_0)\Delta(\phi_{\mathbf{x}_0}^{-1}\mathbf{c}) \leq \zeta p(\mathbf{x}|\boldsymbol{\theta}_0)\Delta(\phi_{\mathbf{x}}^{-1}\mathbf{c}) \right\}} p(\mathbf{x}|\boldsymbol{\theta}_0)\mu(d\mathbf{x}) \\ &= \Pr\left(p(\mathbf{x}_0|\boldsymbol{\theta}_0)\Delta(\phi_{\mathbf{x}_0}^{-1}\mathbf{c}) \leq \zeta p(\mathbf{x}|\boldsymbol{\theta}_0)\Delta(\phi_{\mathbf{x}}^{-1}\mathbf{c}) \mid \boldsymbol{\theta}_0\right) \end{aligned}$$

Fin de la démonstration du théorème 4.

#### • Invariance sur une statistique exhaustive $S(\mathbf{X})$ et non sur $\mathbf{X}$

Jusqu'ici nous avons supposé que  $\mathcal{G}$  était à la fois isomorphe à  $\mathcal{X}$  et à  $\Theta$  de telle sorte que  $\mathcal{X}$  et  $\Theta$  devaient être isomorphes. Une façon simple et usuelle de réduire la dimensionnalité de  $\mathcal{X}$  à celle de  $\Theta$  est de remplacer  $\mathbf{X}$  par une statistique exhaustive  $S(\mathbf{X})$ . Pour insister sur le fait que  $S(\mathbf{X})$  est une variable aléatoire, on devrait plutôt noter  $\mathbf{S} = S(\mathbf{X})$  la variable aléatoire induite par la fonction déterministe  $S$  appliquée à la variable aléatoire  $\mathbf{X}$ , mais nous choisissons plutôt de ne faire référence à  $S$  qu'en tant que fonction déterministe, appliquée notamment à la variable aléatoire  $\mathbf{X}$ .

**Définition 21** Soit  $\mathbf{X}$  une variable aléatoire dont la loi de probabilité est paramétrée par  $\boldsymbol{\theta}$ .  $S(\mathbf{X})$  est appelée **statistique exhaustive pour  $\boldsymbol{\theta}$**  si la loi de probabilité de  $\mathbf{X}$  conditionnelle à la variable aléatoire  $S(\mathbf{X})$  ne dépend pas de  $\boldsymbol{\theta}$ .

Le théorème 4 s'étend alors immédiatement en remplaçant dans les hypothèses  $\mathbf{X}$  par une statistique exhaustive  $S(\mathbf{X})$  et dans la statistique de la  $p$ -value la vraisemblance de  $\mathbf{X}$  par la vraisemblance marginale de  $S(\mathbf{X})$ . On remplace maintenant l'application définie par l'Eq. (10.4) par

$$\begin{aligned}\phi_{S(\mathbf{x})} : \mathcal{G} &\rightarrow S(\mathcal{X}) \\ g &\rightarrow g(S(\mathbf{x}))\end{aligned}\tag{10.15}$$

**Théorème 5** Soit  $\mathcal{P}_\Theta = \{p_{\mathbf{X}|\theta}(\cdot), \theta \in \Theta\}$  une famille de densités de probabilité par rapport à une mesure quelconque sur  $\mathcal{X}$ . Soit  $S(\mathbf{X})$  une statistique exhaustive de  $\mathbf{X}$ ,  $S(\mathcal{X})$  l'espace d'appartenance de  $S(\mathbf{X})$  et  $\mathcal{P}_{S,\Theta} = \{p_{S(\mathbf{X})|\theta}(\cdot), \theta \in \Theta\}$  la famille des densités marginales de  $S(\mathbf{X})$  par rapport à une mesure  $\mu$ . Supposons que

1.  $\mathcal{P}_{S,\Theta}$  est invariante sous l'action du groupe  $\mathcal{G}$  sur  $S(\mathcal{X})$ .
2.  $\phi_\theta$  et  $\phi_{S(\mathbf{x})}$  définis par les Eq. (10.5,10.15) sont bijectives quels que soit  $\theta \in \Theta$  et  $S(\mathbf{x}) \in S(\mathcal{X})$ .
3. La mesure a priori sur  $\Theta$  est la mesure induite par  $H^r$  via  $\phi_\theta$ .
4. La mesure  $\mu$  sur  $S(\mathcal{X})$  est la mesure induite par  $H^r$  via  $\phi_{S(\mathbf{x})}$ .

Le PLR défini par l'Eq. (7.14) s'exprime alors, avec  $\mathbf{x}_0 \in \mathcal{X}$ ,  $\theta_0 \in \Theta$  et  $\zeta > 0$  et pour  $\mathbf{c} \in \mathcal{X}$  quelconque, comme l'intégrale fréquentiste :

$$PLR(\mathbf{x}_0, \zeta) = Pr\left( p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}_0)) \Delta(\phi_{S(\mathbf{x}_0)}^{-1} \mathbf{c}) \leq \zeta p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x})) \Delta(\phi_{S(\mathbf{x})}^{-1} \mathbf{c}) \mid \theta_0 \right)\tag{10.16}$$

avec  $\Delta$  le module du groupe  $\mathcal{G}$ , défini dans la définition 10 p. 145.

Rappelons que nous avons fait apparaître un terme  $\mathbf{c} \in \mathcal{X}$  quelconque grâce à l'équation (10.12) et afin de rendre les termes à gauche et à droite de l'inégalité symétriques en  $\mathbf{x}$  et  $\mathbf{x}_0$ .

Preuve : Notons d'abord que l'expression du PLR ne dépend pas du choix de la mesure par rapport à laquelle est exprimée la densité de probabilité de  $\mathbf{X}$ . En effet, si la mesure de probabilité  $P_\theta(\cdot)$  est absolument continue par rapport à  $\lambda$  et  $\lambda'$  (non fonctions de  $\theta$ ) et que l'on note les densités correspondantes  $p_{\mathbf{X}|\theta}(\cdot)$  et  $p'_{\mathbf{X}|\theta}(\cdot)$ , i.e.  $\forall A \subset \mathcal{X}$

$$P_\theta(A) = \int_A p_{\mathbf{X}|\theta}(\mathbf{x}) \lambda(d\mathbf{x}) = \int_A p'_{\mathbf{X}|\theta}(\mathbf{x}) \lambda'(d\mathbf{x})$$

alors

$$\frac{P_\theta(d\mathbf{x})}{P_{\theta_0}(d\mathbf{x})} = \frac{p_{\mathbf{X}|\theta_0}(\mathbf{x}) \lambda(d\mathbf{x})}{p_{\mathbf{X}|\theta}(\mathbf{x}) \lambda(d\mathbf{x})} = \frac{p_{\mathbf{X}|\theta_0}(\mathbf{x})}{p_{\mathbf{X}|\theta}(\mathbf{x})}$$

et de même pour  $\lambda'$ , de telle sorte que

$$\frac{p_{\mathbf{X}|\theta_0}(\mathbf{x})}{p_{\mathbf{X}|\theta}(\mathbf{x})} = \frac{p'_{\mathbf{X}|\theta_0}(\mathbf{x})}{p'_{\mathbf{X}|\theta}(\mathbf{x})}$$

D'autre part,

$$\begin{aligned}PLR(\mathbf{x}_0, \zeta) &= Pr\left( (p_{\mathbf{X}|\theta_0}(\mathbf{x}_0) \leq \zeta p_{\mathbf{X}|\theta}(\mathbf{x}_0) \mid \mathbf{x}_0) \right) \\ &= Pr\left( (p_{\mathbf{X}|S(\mathbf{X})}(\mathbf{x}_0|S(\mathbf{x}_0)) p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}_0)) \leq \zeta p_{\mathbf{X}|S(\mathbf{X})}(\mathbf{x}_0|S(\mathbf{x}_0)) p_{S(\mathbf{X})|\theta}(S(\mathbf{x}_0)) \mid \mathbf{x}_0) \right)\end{aligned}$$

car comme  $S(\mathbf{x})$  est une fonction de  $\mathbf{x}$ ,  $p_{\mathbf{X}|\theta}(\mathbf{x}) = p_{\mathbf{X},S(\mathbf{X})|\theta}(\mathbf{x}, S(\mathbf{x}))$  et comme de plus  $S(\mathbf{X})$  est une statistique exhaustive de  $\mathbf{X}$ ,

$$\begin{aligned}p_{\mathbf{X}|\theta}(\mathbf{x}) &= p_{\mathbf{X}|S(\mathbf{X}),\theta}(\mathbf{x}|S(\mathbf{x})) p_{S(\mathbf{X})|\theta}(S(\mathbf{x})) \\ &= p_{\mathbf{X}|S(\mathbf{X})}(\mathbf{x}|S(\mathbf{x})) p_{S(\mathbf{X})|\theta}(S(\mathbf{x}))\end{aligned}\tag{10.17}$$

Simplifiant ainsi les densités ne dépendant pas de  $\theta$ ,

$$\begin{aligned} \text{PLR}(\mathbf{x}_0, \zeta) &= \Pr\left(p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}_0)) \leq \zeta p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}_0)) \mid S(\mathbf{x}_0)\right) \text{ si } p_{\mathbf{X}|S(\mathbf{X})}(\mathbf{x}_0|S(\mathbf{x}_0)) > 0 \\ &= \Pr\left(p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}_0))\Delta(\phi_{S(\mathbf{x}_0)}^{-1}\mathbf{c}) \leq \zeta p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}))\Delta(\phi_{S(\mathbf{x})}^{-1}\mathbf{c}) \mid \theta_0\right) \quad \forall \mathbf{c} \in S(\mathcal{X}) \text{ (Th. 4)} \end{aligned}$$

• **Relation entre le PLR et la  $p$ -value**

Voyons dans un premier temps comment réexprimer le PLR sous les conditions du théorème 5 en utilisant une distribution cumulative. Notons  $T(\mathbf{x})$  la statistique (fonction déterministe des données  $\mathbf{x}$ ) suivante :

$$T(\mathbf{x}) = p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}))\Delta(\phi_{S(\mathbf{x})}^{-1}\mathbf{c}) \quad (10.18)$$

Vu comme une variable aléatoire, le jeu de données  $\mathbf{x}$  induit la variable aléatoire  $T(\mathbf{X})$  de même qu'auparavant la statistique  $S(\mathbf{x})$  induisait la variable aléatoire  $S(\mathbf{X})$ . Notons  $F_{T(\mathbf{X})|\theta_0}$  la distribution cumulative de  $T(\mathbf{X})$  sous l'hypothèse nulle :

$$F_{T(\mathbf{X})|\theta_0}(\zeta) = \Pr(T(\mathbf{x}) \leq \zeta \mid \theta_0)$$

Partant du théorème 5, le PLR se réécrit alors

$$\text{PLR}(\mathbf{x}_0, \zeta) = 1 - F_{T(\mathbf{X})|\theta_0}(\zeta^{-1}T(\mathbf{x}_0)) \quad (10.19)$$

En particulier, pour un seuil  $\zeta = 1$ , on constate directement que le PLR est égal à 1 moins la  $p$ -value (voir la section d'introduction) associée à la statistique de test  $T(\mathbf{x})$ .

**Corrolaire 7** *Dans les conditions du théorème 5 et avec  $\mathbf{c} \in \mathcal{X}$  quelconque, le PLR de seuil 1 est égal à 1 moins la  $p$ -value d'une statistique de test proportionnelle à la vraisemblance marginale de la statistique exhaustive  $S(\mathbf{x})$  pondérée par le module de  $\mathcal{G}$  :*

$$\text{PLR}(\mathbf{x}, 1) = 1 - p_{\text{val}}\left(p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}))\Delta(\phi_{S(\mathbf{x})}^{-1}\mathbf{c})\right) \quad (10.20)$$

Le test fréquentiste correspondant au PLR est alors, pour un seuil  $\lambda > 0$  quelconque :

$$\text{Rejeter } H_0 \text{ si } p_{S(\mathbf{X})|\theta_0}(S(\mathbf{x}))\Delta(\phi_{S(\mathbf{x})}^{-1}\mathbf{c}) \leq \lambda \quad (10.21)$$

Rappelons à nouveau que nous avons fait apparaître un terme  $\mathbf{c} \in \mathcal{X}$  quelconque grâce à l'équation (10.12) et afin de rendre les termes à gauche et à droite de l'inégalité symétriques en  $\mathbf{x}$  et  $\mathbf{x}_0$ .

## Hypothèses faisant appel à la notion d'orbites

On peut en fait généraliser le théorème 5 p. 160 en faisant des hypothèses moins restrictives. Cependant, l'expression de telles hypothèses fait appel à des notions un peu plus techniques. Nous nous contentons ici de noter quelques généralisations possibles utilisées dans la littérature.

L'hypothèse de bijectivité de  $\phi_\theta$  défini par l'Eq. (10.5) p. 156 n'est pas nécessaire. Cette condition implique que  $\mathcal{G}$  agit transitivement sur  $\Theta$ , *i.e.*  $\{g\theta, g \in \mathcal{G}\} = \Theta \quad \forall \theta \in \Theta$ , et que  $\mathcal{G}$  agit librement, *i.e.*  $g\theta = \theta$  implique  $g = e$ . En général cette condition est légèrement étendue en la condition " $\mathcal{G}$  agit transitivement sur  $\Theta$  et il existe  $\theta_* \in \Theta$  tel que le sous-groupe  $\mathcal{G}_* = \{g \in \mathcal{G} : g\theta_* = \theta_*\}$  est compact", voir par exemple Zidek [1969] condition 3 ou Eaton and Sudderth [2002] condition 5.

En passant du théorème 4 p. 157 au théorème 5 p. 160, nous avons supprimé l'hypothèse de transitivité de  $\mathcal{G}$  sur  $\mathcal{X}$  en supposant que  $\mathcal{G}$  doit agir transitivement non sur  $\mathcal{X}$  mais sur un espace transformé  $S(\mathcal{X})$  de dimension plus faible que  $\mathcal{X}$ , où  $S(\mathbf{X})$  était alors une statistique exhaustive de  $\mathbf{X}$ . L'autre grande technique de réduction de dimensionnalité consiste à prendre pour  $S(\mathcal{X})$  une orbite de  $\mathcal{G}$  (par définition,  $\mathcal{G}$  agit transitivement sur ses orbites), et plus spécifiquement l'orbite correspondant aux données observées :  $O_{\mathbf{x}_0} = \{g\mathbf{x}_0, g \in \mathcal{G}\}$ . La notion de statistique exhaustive étant en général mieux connue que celle d'orbite, nous ne traitons pas cette seconde possibilité.



### Relation entre ce résultat et le théorème de Stein

Comme nous l'avons mentionné dans le chapitre 9 et au début de cette section, l'invariance de groupe est le cadre le plus classique d'équivalence entre les probabilités fréquentiste et bayésienne, c'est-à-dire qui répond par exemple à la question : *Quand a-t-on une égalité du type de l'égalité suivante ?*

$$\Pr(\boldsymbol{\theta} \in C(\mathbf{x})|\mathbf{x}) = \Pr(\boldsymbol{\theta} \in C(\mathbf{x})|\boldsymbol{\theta})$$

ie 
$$\int_{\{\boldsymbol{\theta} \in C(\mathbf{x})\}} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}|\mathbf{x}) = \int_{\{\mathbf{x}: \boldsymbol{\theta} \in C(\mathbf{x})\}} d\mathbf{x} p(\mathbf{x}|\boldsymbol{\theta})$$

Notons bien que normalement dans l'égalité précédente (qui constitue en fait le résultat du théorème de Stein que nous allons voir) le terme de gauche dépend de  $\mathbf{x}$  tandis que le terme de droite dépend de  $\boldsymbol{\theta}$ . Cette égalité-ci implique donc que ces intégrales ont une valeur fixée  $c$ .

Les théories de base sous-jacente à ce type d'unification sont la théorie pivotale de Fisher pour le côté "fréquentiste" et la mesure à droite de Haar pour le côté bayésien. La réconciliation de la théorie de Fisher et de l'approche bayésienne a alors démarré notamment avec Fraser [1961]. Stein [1965] esquisse ensuite de premières hypothèses impliquant ce qui s'appellera le théorème de Stein pour les domaines invariants. Le résultat de Stein, démontré sous différentes conditions depuis le résultat initial, énonce que :

**Théorème 6** *Si un domaine  $C(\mathbf{x}) \subset \Theta$  satisfait*

$$gC(\mathbf{x}) = C(g(\mathbf{x})) \tag{10.22}$$

avec  $gC(\mathbf{x}) = \{g\boldsymbol{\theta}, \boldsymbol{\theta} \in C(\mathbf{x})\}$ , alors sous certaines conditions d'invariance<sup>3</sup>,

$$\Pr(\boldsymbol{\theta} \in C(\mathbf{x})|\mathbf{x}) = c \quad \forall \mathbf{x} \in \mathcal{X} \quad (\text{probabilité bayésienne}) \tag{10.23}$$

$$\text{et } \Pr(\boldsymbol{\theta} \in C(\mathbf{x})|\boldsymbol{\theta}) = c \quad \forall \boldsymbol{\theta} \in \Theta \quad (\text{probabilité fréquentiste}) \tag{10.24}$$

D'autres études telles que celle de Zidek [1969] étendent ce type de résultats à l'ensemble des procédures invariantes. Un des stades d'unification le plus finalisé et général est atteint par exemple par Eaton and Sudderth [1999] qui montrent que sous des hypothèses assez standards en invariance ce qu'ils explicitent comme étant l'hypothèse centrale à la théorie pivotale de Fisher mène à une procédure identique à celle obtenue par la procédure bayésienne avec un *a priori* droit de Haar. Ils montrent aussi (et de façon plus générale dans Eaton and Sudderth [2002]) que pour qu'une inférence bayésienne invariante soit admissible (au sens où il n'existe pas une autre inférence invariante dont l'erreur quadratique moyenne soit plus faible pour tout  $\boldsymbol{\theta}$ ) il faut qu'elle soit induite par l'*a priori* droit de Haar sur  $\boldsymbol{\theta}$ .

Le théorème 4 dérivé dans ce chapitre ne découle pas du théorème de Stein car le domaine  $C(\mathbf{x}) \subset \Theta$  sur lequel nous intégrons n'est pas invariant. En effet, ceci ne serait vrai que si  $\boldsymbol{\theta}_0$  était invariant sous  $\mathcal{G}$ , ie  $g\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0$  quel que soit  $g \in \mathcal{G}$  (hypothèse  $H_0$  invariante sous  $\mathcal{G}$ ). Or, dans notre théorème nous supposons que  $\phi_{\boldsymbol{\theta}}$  est bijective quelle que soit  $\boldsymbol{\theta} \in \Theta$  ce qui implique que  $g\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0$  est équivalent à  $\bar{g} = e$  (fonction identité).

Ainsi, comme  $H_0$  n'est pas invariante sous  $\mathcal{G}$  le théorème de Stein ne s'applique pas, de sorte qu'il a été nécessaire de redériver l'égalité des probabilités bayésienne et fréquentiste dans un cas proche mais néanmoins différent. Notons à cet égard que l'ouvrage de référence Berger [1985] présente une démonstration simplifiée et très accessible du théorème de Stein mais que celle-ci confond par souci de simplicité élément de groupe et élément de  $\mathcal{X}$  ou  $\Theta$ , ce qui n'est pas trop gênant pour le théorème de Stein mais laisse trop d'ambiguïtés pour être réutilisée en pratique pour démontrer le théorème 4. Le problème posé par ces ambiguïtés transparait par exemple par le fait que le terme  $\Delta(\phi_{\mathbf{x}}^{-1}\mathbf{c})$  n'apparaît

<sup>3</sup>Le premier jeu de conditions date de Stein [1965]. Les hypothèses plus simples données par Chang and Villegas [1986] sont proches de celles utilisées dans la section 10.1.

pas dans le théorème de Stein mais apparaît dans le nouveau théorème.

Concernant le fait que  $H_0$  n'est pas invariante sous  $\mathcal{G}$ , notons que ceci suppose un cadre de test d'hypothèses différent du cas standard invariant où  $\Theta_0$  et  $\Theta_1$  sont invariants sous  $\mathcal{G}$  (voir Eaton [1989], Robert [2007]), mais cela n'a rien de choquant comme on peut le voir par exemple dans le test le plus simple et le plus classique discuté dans le paradoxe de Lindley (section 7.2 et section 7.3). L'invariance n'est utilisée ici que comme un outil pour démontrer une équivalence entre des concepts centraux dans un contexte plus large que celui de l'invariance, non pour dériver une procédure invariante.

Pour revenir à la problématique générale posée au début de cette sous-section, ici nous avons répondu à la question : *Quand a-t-on une égalité du type de l'égalité suivante ?*

$$\int_{C(\mathbf{x}, \boldsymbol{\theta}_0)} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}|\mathbf{x}) = \int_{C'(\mathbf{x}, \boldsymbol{\theta}_0)} d\mathbf{y} p(\mathbf{y}|\boldsymbol{\theta}_0)$$

ou plus précisément

$$\int_{\{\boldsymbol{\theta}: p(\mathbf{x}|\boldsymbol{\theta}_0) \leq p(\mathbf{x}|\boldsymbol{\theta})\}} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}|\mathbf{x}) = \int_{\{\mathbf{y}: p(\mathbf{x}|\boldsymbol{\theta}_0)f(\mathbf{x}) \leq p(\mathbf{y}|\boldsymbol{\theta}_0)f(\mathbf{y})\}} d\mathbf{y} p(\mathbf{y}|\boldsymbol{\theta}_0) \quad (10.25)$$

où ces intégrales sont ici (bien) fonctions de  $\mathbf{x}$  et de  $\boldsymbol{\theta}_0$  contrairement au théorème de Stein où les intégrales considérées prennent une valeur fixée.

## 10.2 Conséquences de l'équivalence

Dans le corollaire 7 nous avons démontré que, en quelques mots, si on note  $S(\mathbf{x})$  une statistique exhaustive des données  $\mathbf{x}$  de même taille que le paramètre  $\boldsymbol{\theta}$  et si la famille de vraisemblances de  $S(\mathbf{x})$  est invariante sous un groupe  $\mathcal{G}$  d'endomorphismes, alors prenant pour l'*a priori*  $\pi_1(\boldsymbol{\theta})$  l'*a priori* droit de Haar associé à  $\mathcal{G}$ , le PLR est égal à 1 moins une *p-value* basée sur la vraisemblance pondérée par un terme assez naturel de  $\mathcal{G}$  :  $\text{PLR}(\mathbf{x}, 1) = 1 - p_{\text{val}}(p(S(\mathbf{x})|\boldsymbol{\theta}_0)\Delta(\phi_{S(\mathbf{x})}^{-1}\mathbf{c}))$  quel que soit  $\mathbf{c} \in \mathcal{X}$ .

Ce résultat étend le résultat de Dempster [1974] et Aitkin [1997], contribue aux problématiques de réconciliation présentées dans la section 7.2 en partie introduite page 25 et contribue à des problématiques fréquentistes supplémentaires, toutes décrites ci-dessous.

### Réconciliations entre fréquentiste et bayésien

- **Réconciliation de mesures d'évidence fréquentiste (*p-value*) et bayésienne (PLR)**

Comme nous venons de le dire, ce résultat (corollaire 7) étend le résultat de Dempster [1974] et Aitkin [1997]. Comme précisé dans l'état de l'art de la section 7.3, ceux-ci ont démontré diverses variantes du fait que pour la loi normale de variance connue, avec un *a priori* uniforme le PLR est égal à 1 moins la *p-value* classique. Ici on démontre aussi que le PLR est égal à 1 moins une *p-value* fréquentiste. Le résultat démontré dans cette thèse implique bien le résultat de Dempster et Aitkin puisque la loi normale de variance connue est une famille de vraisemblances invariante par translation, l'*a priori* uniforme est bien l'*a priori* de Haar à droite associé au groupe des translations, et enfin la statistique  $|x|$  de la *p-value* du problème de Dempster et Aitkin est bien proportionnel à  $p(x|\boldsymbol{\theta})\Delta(\phi_{S(x)}^{-1}\mathbf{c})$  puisque pour le groupe des translations  $\Delta = 1$  (voir la proposition 23). Dans cette thèse le résultat de Dempster et Aitkin est ainsi étendu à toute famille invariante sous un groupe de transformations, en choisissant chaque fois l'*a priori* droit de Haar associé à ce groupe.

Le résultat contribue aussi à la problématique de réconciliation de mesures d'évidence fréquentiste (*p-value*) et bayésienne (PLR) présentée dans la section 7.2 et introduite en termes bibliographiques page 25. On constate que le problème de non conciliation des *mesures d'évidence* fréquentiste (*p-value*) et bayésienne ( $\text{Pr}(H_0|\mathbf{x})$ ) disparaît en partie si l'on remplace la forme stricte  $\text{Pr}(H_0|\mathbf{x})$  par une mesure du type " $\text{Pr}(H_0(\mathbf{x})|\mathbf{x})$ ". En particulier, le PLR est strictement égal à une *p-value* en cas invariant.

- **Réconciliation des approches bayésienne et fréquentiste pour le test d'hypothèses**

Plus généralement, comme la  $p$ -value peut servir à la fois de statistique de test et de mesure d'évidence pour  $H_0$  et que le PLR est supposé remplir la même fonction, constater qu'ils sont égaux dans certains cas permet de répondre à la problématique de réconciliation des procédures de test fréquentiste et bayésien.

En l'occurrence, ceci permet de définir une relation entre l'espace des  $a$  priori (côté bayésien) et *statistique de test* (côté fréquentiste), avec la correspondance  $a$  priori droit de Haar – module  $\Delta(\phi_{S(\mathbf{x})}^{-1}\mathbf{c})$  dans le cas d'une famille invariante avec un seuillage de la vraisemblance.

Prenant la perspective d'un *fréquentiste*, pour une famille de vraisemblances invariante, si l'on souhaiterait plutôt appliquer le test fréquentiste consistant à seuiller  $p(x|\theta)\Delta(\phi_{S(x)}^{-1}\mathbf{c})$ , on peut aussi décider d'utiliser, si l'on en a, de l'information sur  $\theta$  sans craindre d'aboutir à un résultat trop différent de celui obtenu par le test fréquentiste : on sait que l' $a$  priori droit de Haar aboutit à la  $p$ -value fréquentiste. En d'autres termes, l' $a$  priori droit de Haar constitue pour ce problème un  $a$  priori de "référence" pour un bayésien, un point de comparaison pour d'autres choix possibles d' $a$  priori.

## Propriétés induites

- **Propriétés de la  $p$ -value appliquées au PLR**

Le résultat nous a permis dans la section 8.2 de déduire des propriétés fréquentistes du PLR en cas invariant. L'une était induite par le fait que dans ce cas le PLR était une  $p$ -value, et l'autre repartait du théorème 5 à partir duquel nous avons observé que le PLR n'avait qu'une courbe ROC malgré le fait qu'en général, comme la structure du PLR est paramétrée par deux paramètres d'entrée  $\zeta$  et  $p$ , il a toute une gamme de performances selon la paramétrisation choisie.

- **Propriétés du PLR appliquées à la  $p$ -value**

Réciproquement, ceci nous informe que dans certains cas la  $p$ -value fréquentiste peut être égale à la probabilité  $a$  posteriori d'un domaine. En particulier, les idées de théorie de la décision s'appliquant au PLR peuvent donner des idées sur ce qui est alors vrai aussi d'une  $p$ -value.

De la même façon, il s'avère ici que la  $p$ -value est égale à une probabilité  $a$  posteriori sur l'hypothèse alternative  $H_1$ . Pourtant, comme la  $p$ -value est calculée à partir d'une probabilité sous  $H_0$  elle est régulièrement critiquée en tant qu'elle n'indique rien de l'hypothèse alternative. Nous montrons ici qu'une  $p$ -value peut être l'expression d'un test naturel incluant des informations raisonnables sur l'hypothèse alternative.

## Nouvelles $p$ -values

- **Nouvelle  $p$ -value fréquentiste ; transformation "géométrique" pour pondérer la vraisemblance**

D'autres formes de  $p$ -values fréquentistes sont toujours recherchées. Ainsi, Evans and Jang [2010] développent de nouvelles formes de  $p$ -value basées sur la densité d'une statistique de test éventuellement pondérée de sorte que cette  $p$ -value devienne invariante sous le changement de variable induite par cette statistique.

Dans le corrolaire 7, partir de l'expression générique du PLR et l'appliquer en cas invariant a fait apparaître une statistique de test donnée par la vraisemblance  $p(\mathbf{x}|\theta_0)$  pondérée par un terme naturel dans l'invariance de groupe, à savoir le module du groupe appliqué en la transformation qui à  $\mathbf{x}$  associe  $\mathbf{c} \in \mathcal{X}$  (quelconque, voir l'Eq. 10.12). Reste à savoir si un tel terme de pondération pourrait avoir un *équivalent* dans des problèmes plus généraux, de même que la théorie de l'information peut adopter des interprétations géométriques. L'utilisation de la géométrie pour les inférences statistiques est un problème standard, voir la synthèse de Kass [1989].

- **Nouvelle  $p$ -value bayésienne ; pas d'utilisation de données prédites**

En développant de nouvelles  $p$ -values, Meng [1994] et Evans and Jang [2010] indiquent, l'un pour le bayésien l'autre pour le fréquentiste, qu'il n'existe pas de variable de "divergence" qui fasse consensus. La nouveauté de la  $p$ -value prédictive de Meng [1994] tient à la structure de la  $p$ -value, qui utilise en effet la distribution de données prédites afin de se ramener à une intégrale notamment sur  $\mathcal{X}$  –comme pour la  $p$ -value traditionnelle– tout en incorporant l'information des données observées  $\mathbf{x}$ . Evans and Jang [2010] propose au contraire la variable de divergence à insérer dans la structure de  $p$ -value.

Ici, le PLR semble consister en la forme la plus directe d'utilisation bayésienne d'une variable de divergence (probabilité *a posteriori* pour que les variables soient différentes sous chacune des hypothèses) et emploie la forme la plus simple de variable de divergence (la vraisemblance). Pour autant, elle n'a été que très peu étudiée alors que le corollaire 7 confirme que le PLR (bayésien) est effectivement assimilable à une  $p$ -value (fréquentiste) car elle peut coïncider avec une  $p$ -value naturelle dans certains cas.



## Quatrième partie

# Modèle bayésien pour assouplir le modèle gaussien et application du PLR



# Table des Matières

---

<b>11</b>	<b>Modèle bayésien hiérarchique</b>	<b>171</b>
11.1	Introduction . . . . .	171
	Motivation pour une nouvelle modélisation . . . . .	171
	Modèle bayésien hiérarchique à partir du modèle gaussien simple . . . . .	172
11.2	Modèle des données . . . . .	173
	Vraisemblance de premier niveau . . . . .	173
	Vraisemblance marginale . . . . .	174
	Discussion du choix des hyperparamètres . . . . .	175
11.3	Probabilité <i>a priori</i> de l'intensité de la planète dans les différents canaux spectraux . . . . .	175
	<i>A priori</i> de premier niveau . . . . .	175
	<i>A priori</i> marginal . . . . .	176
	Discussion du choix de la distribution <i>a priori</i> et de ses hyperparamètres . . . . .	176
11.4	Probabilité <i>a posteriori</i> de l'intensité et échantillonnage . . . . .	176
	Méthode . . . . .	176
	Application sur un jeu de données simulé simplement . . . . .	177
<b>12</b>	<b>Détection par le PLR pour les données SPHERE</b>	<b>181</b>
12.1	Jeux de données issus de la simulation physique fine . . . . .	181
	Description des données simulées . . . . .	181
	Valeurs des hyperparamètres . . . . .	183
	Echantillonnage de la distribution <i>a posteriori</i> de $\alpha$ . . . . .	184
	Détection par le PLR et le facteur de Bayes . . . . .	184
	Données constituées de plus de poses . . . . .	187
12.2	Performances fréquentistes évaluées sur des données simulées plus simplement . . . . .	188
	Echantillonnage de $p(\mathbf{x} \alpha^*)$ et non de $p_1(\mathbf{x}) = \int d\alpha p(\mathbf{x} \alpha)\pi_1(\alpha)$ . . . . .	189
	Simulation d'un jeu de données selon la vraisemblance $p(\mathbf{x} \alpha^*)$ . . . . .	189
	Echantillonnage de la distribution <i>a posteriori</i> $\pi_1(\alpha \mathbf{x})$ . . . . .	190
	Relations entre les paramètres $(\zeta, p)$ et la PFA . . . . .	191
	Courbes ROC et paramétrisation optimale . . . . .	191
	<b>Conclusion et perspectives</b>	<b>195</b>
	<b>Publications et communications</b>	<b>199</b>

---





# Chapitre 11

## Modèle bayésien hiérarchique

Le premier objectif de ce chapitre est de construire un modèle des données *plus souple* que le modèle gaussien simple étudié dans la partie I. Le second objectif est d'employer le paradigme bayésien qui permet de définir la notion de probabilité *a posteriori* des paramètres d'intérêt à partir d'une distribution de probabilité *a priori* de ces paramètres.

*Pour des données affectées par la présence d'une planète, nous supposons dans cette partie que la position initiale  $r$  de la planète est connue et garderons ainsi  $r$  implicite.* Ainsi, à la place de noter une vraisemblance  $p(\mathbf{x}|\boldsymbol{\alpha}, r)$  on notera  $p(\mathbf{x}|\boldsymbol{\alpha})$ . Pour appliquer la méthode de détection en pratique, comme la position n'est pas connue, la démarche adoptée est la même que celle de la partie I : la procédure de détection pour une position  $r$  fixée est appliquée en chaque position  $r$  à tester<sup>1</sup>.

Comme discuté dans la section 7.1 le test d'hypothèses qui nous intéresse s'écrit sous la forme

$$H_0 : \boldsymbol{\alpha} = \mathbf{0} \quad H_1 : \boldsymbol{\alpha} \neq \mathbf{0} \quad (11.1)$$

où  $\boldsymbol{\alpha} \in \mathbb{R}_+^M$  représente l'intensité de la planète dans les  $M$  canaux spectraux. On rappelle (voir la section 7.1) qu'on entend en fait plutôt pour  $H_1$  le fait que  $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$  où  $\boldsymbol{\alpha}^* \neq \mathbf{0}$  est l'intensité vraie de la planète mais est inconnue. On note  $\pi_1(\boldsymbol{\alpha})$  sa densité de probabilité *a priori* conditionnellement au fait qu'une planète est présente à côté de l'étoile observée.

### 11.1 Introduction

#### Motivation pour une nouvelle modélisation

Nous avons constaté grâce à un modèle plus fin d'imagerie *speckle* étudié dans la partie II que le modèle gaussien développé dans la partie I a une queue trop fine. Le modèle gaussien suppose d'autre part que les intensités entre pixels voisins sont indépendantes conditionnellement à l'image moyenne. Le modèle fin d'images *speckle* prend bien en compte ces deux problématiques mais il est quant à lui difficilement utilisable en pratique car il suppose connu la moyenne du front d'onde et la matrice de covariance spatiale du front d'onde. Or l'estimation de ces moments complexes multivariés requiert l'existence et l'utilisation de dispositifs instrumentaux supplémentaires et le développement d'une méthode d'estimation spécifique à ces instruments incluant le développement d'une méthode d'estimation de matrice de covariance complexe. Enfin, la corrélation temporelle et la corrélation entre les différents canaux spectraux doivent aussi être prises en compte plus finement que dans le modèle gaussien simple du chapitre 3. On supposait en effet la moyenne de l'intensité lumineuse due à l'étoile identique avec une probabilité 1 dans toutes les images quel que soit l'instant et le canal spectral alors que les conditions atmosphériques varient dans le

---

<sup>1</sup>Dans la section 3.4 nous avons étudié un tel test d'hypothèses et l'avons appelé *test pixel par pixel*, mais le pixel sera implicite ici. De même que dans la partie I on indiquait la valeur de la statistique de test  $\hat{\alpha}(\mathbf{x}, r)$  pour tous les pixels  $r$ , on peut en pratique ici indiquer la valeur  $\text{PLR}(\mathbf{x}, \zeta, r)$  correspondant à la distribution *a posteriori* conditionnelle  $\pi(\boldsymbol{\alpha}|\mathbf{x}, r)$  pour une certaine valeur  $\zeta$  pour les pixels  $r$  testés.

temps et ont un effet important sur les données (voir la section 2.2 p. 40) et que les aberrations optiques dépendent de la longueur d'onde d'acquisition des images (voir la section 3.5 p. 71).

D'autre part nous avons mentionné que le paradigme bayésien permettait d'obtenir une distribution de probabilité du paramètre d'intérêt, ici l'intensité de l'exoplanète, étant données les images observées  $\mathbf{x}$ . Ceci a une interprétation intuitive bien fondée. L'utilisation de distributions *a priori* permet de plus de régulariser des problèmes dits mal posés au sens de Hadamard ou mal conditionnés : schématiquement, si la solution de l'inversion du modèle physique direct n'est pas unique, trop sensible aux données *etc*, la distribution *a priori* sélectionne une classe des solutions pertinentes. Enfin, dans le cas de la classification d'une éventuelle exoplanète détectée (schématiquement : planète gazeuse, tellurique, *etc* ...), il est possible grâce à la loi *a priori* sur l'intensité de la planète dans les différents canaux spectraux d'injecter le type de corrélation entre les différentes composantes spectrales que l'on peut attendre pour tel ou tel type d'exoplanète.

### Modèle bayésien hiérarchique à partir du modèle gaussien simple

Un modèle précis de la densité de probabilité de chaque image a été étudié dans la partie 2 dans le cas mono-bande. Il prend en compte un fond non-uniforme, un modèle général des *speckles* qui repose sur un modèle gaussien complexe corrélé du front d'onde arrivant sur la caméra CCD, le processus de Poisson et des bruits additifs gaussiens. Sous une hypothèse de fort flux, on a démontré que ce modèle tendait après renormalisation vers le modèle gaussien simple. Construire un modèle statistique plus fin que le modèle gaussien à partir du modèle gaussien n'est donc pas aberrant.

Une technique pour partir d'un modèle afin de construire un nouveau modèle est la modélisation hiérarchique. Pour cela il faut définir un modèle  $p(\mathbf{x}|\boldsymbol{\theta})$  dit de premier niveau et un *a priori*  $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$  où  $\boldsymbol{\lambda}$  sont appelés les hyperparamètres du modèle. Le modèle marginal est alors calculé par marginalisation du paramètre  $\boldsymbol{\theta}$  : d'après le théorème de Bayes,

$$p(\mathbf{x}|\boldsymbol{\lambda}) = \int d\boldsymbol{\theta} p(\mathbf{x}, \boldsymbol{\theta}|\boldsymbol{\lambda}) = \int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\lambda})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) = \int d\boldsymbol{\theta} p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda}) \quad (11.2)$$

L'intégration peut éventuellement être compliquée analytiquement et lourde numériquement. Une technique simple consiste à choisir l'*a priori*  $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$  conjugué au modèle  $p(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\lambda})$ .

**Définition 22** Une distribution *a priori* conjugué à une vraisemblance donnée est une distribution *a priori* telle que la distribution *a posteriori* associée appartient à la même famille de distributions que l'*a priori*.

L'utilisation d'un tel *a priori* supprime la nécessité d'effectuer l'intégration : en renversant la formule habituelle donnant la distribution *a posteriori* de  $\boldsymbol{\theta}$  on écrit

$$p(\mathbf{x}|\boldsymbol{\lambda}) = \frac{p(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\lambda})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})}{\pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\lambda})} \quad \forall \boldsymbol{\theta} \in \Theta \quad (11.3)$$

Par définition la densité de probabilité *a posteriori*  $\pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\lambda})$  a la même expression que la densité de probabilité *a priori* conjuguée  $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$ , aux paramètres près ceux-ci étant connaissables grâce au fait que  $\pi(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\lambda}) \propto p(\mathbf{x}|\boldsymbol{\theta}, \boldsymbol{\lambda})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$ . La loi marginale n'est donc plus égale à une intégrale mais au rapport de trois termes dont les expressions sont explicites.

La vraisemblance (de premier niveau) est ici la distribution gaussienne. La distribution *a priori* des paramètres de nuisance sera choisie mathématiquement pratique et physiquement recevable. L'*a priori* conjugué pour  $K = 1$  pour les paramètres de nuisance considérés est une distribution gaussienne – inverse Wishart. Cet *a priori* convient : son utilisation s'avère rester assez simple mathématiquement et physiquement il est satisfaisant car les hyperparamètres peuvent être choisis de telle sorte que l'*a priori* soit assez diffus, alourdissant ainsi la queue de la distribution gaussienne dont nous avons vu qu'elle était trop mince. Nous constaterons de plus *in fine* grâce à un très bon résultat de la détection à

partir du modèle bayésien que cette nouvelle vraisemblance semble mieux modéliser les données que la loi gaussienne simple.

La distribution *a priori* des paramètres d'intérêt sera aussi construite à partir d'un modèle hiérarchique à un niveau.

## 11.2 Modèle des données

### Vraisemblance de premier niveau

Les données  $\mathbf{x}$  sont constituées de  $K$  ensembles successifs de  $M$  images, où  $M$  est le nombre de canaux spectraux et chaque image est représentée par un vecteur de taille  $L \times 1$  noté  $\mathbf{x}(k, m)$  avec  $k \in \llbracket 1, K \rrbracket$  le numéro de la pose et  $m \in \llbracket 1, M \rrbracket$  le numéro du canal.

Afin d'augmenter les performances de détection, il est important de prendre en compte à un instant  $k$  donné les corrélations à la fois spatiales et spectrales. Les corrélations spatiales proviennent de divers phénomènes tels que les ordres faibles et élevés des aberrations quasi-statiques et les corrélations spectrales apparaissent en particulier pour des images enregistrées dans des canaux spectraux proches.

Afin de prendre en compte les deux types de corrélation sans supposer que les "fonds" dans les différents canaux sont trop simplement reliés, on définit un vecteur de taille  $LM \times 1$  par  $\mathbf{x}(k)^t = (\mathbf{x}(k, 1)^t, \dots, \mathbf{x}(k, M)^t)$ . Une extension directe du modèle mono-canal consiste à supposer que, conditionnellement à des paramètres constants, les ensembles d'images sont indépendants et décrits par

$$\mathbf{x}(k) | \mathbf{d}, \Sigma, \boldsymbol{\alpha}, A(k) \sim \mathcal{N}_{LM}(A(k)\boldsymbol{\alpha} + \mathbf{d}, \Sigma) \quad (11.4)$$

avec

- $\mathbf{d}^t = (\mathbf{d}(1)^t, \dots, \mathbf{d}(M)^t)$  où  $\mathbf{d}(m)$  est le vecteur de taille  $L \times 1$  ("vecteur-image") qui désigne le fond non uniforme du canal  $m$ . Dans le cas présent, il provient à la fois des résidus atmosphériques de l'étoile coronographiée et des *speckles* statiques liés aux aberrations optiques.
- $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)^t$  où  $\alpha_m$  est le scalaire dénotant l'intensité de la planète dans le canal  $m$ .
- $A(k)$  est la matrice de taille  $LM \times M$  définie par

$$A(k) = \begin{pmatrix} \mathbf{p}(k, 1) & \mathbf{0}_{L \times 1} & \dots & \mathbf{0}_{L \times 1} \\ \mathbf{0}_{L \times 1} & \mathbf{p}(k, 2) & \dots & \mathbf{0}_{L \times 1} \\ \mathbf{0}_{L \times 1} & \mathbf{0}_{L \times 1} & \dots & \mathbf{p}(k, M) \end{pmatrix}$$

où  $\mathbf{p}(k, m)$  est un vecteur-image contenant la réponse instrumentale supposée *connue* de la source à l'instant  $k$  dans le canal  $m$ . Dans le contexte astrophysique, on peut effectivement supposer que la forme du profil est connue (par étalonnage) et que le mouvement de la source est connu (mouvement de rotation de champ déterministe) à la position initiale près. Nous supposons cependant que la position à l'instant  $k = 1$  de l'éventuelle source est *connue*.

- $\Sigma$  est la matrice de covariance spatiale et spectrale de taille  $LM \times LM$ . Par exemple, avec le mode DBI de l'instrument IRDIS de SPHERE pour lequel  $M = 2$ , la bande diagonale principale de  $\Sigma$  (*i.e.* les termes indexés par  $\{(i, i), (i, i \pm 1), (i, i \pm 2), \dots\}$ ) décrit la corrélation spatiale, tandis que la bande hors-diagonale du milieu (d'indices  $\{(i, i + L), (i, i + L \pm 1), \dots\}$ ) décrit les corrélations entre un pixel d'une bande et le voisinage de ce pixel dans l'autre bande.

La vraisemblance se réécrit

$$p(\mathbf{x} | \mathbf{d}, \Sigma, \boldsymbol{\alpha}) = (2\pi)^{-KL} |\Sigma|^{-K/2} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} \sum_{k=1}^K (\mathbf{d} - \mathbf{x}(k) + A(k)\boldsymbol{\alpha})(\mathbf{d} - \mathbf{x}(k) + A(k)\boldsymbol{\alpha})^t \right) \quad (11.5)$$

où  $\text{etr}(B) = \exp(\text{Tr}(B))$ .

Notons que lorsque la source se déplace par rapport au fond, c'est-à-dire qu'il existe un instant  $k$  tel que  $\mathbf{p}(k, m) \neq \mathbf{p}(k', m)$  dans n'importe quelle bande, les  $\mathbf{x}(k)$  ne sont pas identiquement distribués. On note l'ensemble des données par  $\mathbf{x}^t = (\mathbf{x}(1)^t, \dots, \mathbf{x}(K)^t)$ .

### Vraisemblance marginale

Comme expliqué précédemment (section 11.1), il est possible de construire une nouvelle vraisemblance de  $\boldsymbol{\alpha}$  en marginalisant les paramètres de nuisance  $\mathbf{d}$  et  $\Sigma$  à partir de distributions *a priori* sur ces paramètres. La loi normale inverse-Wishart est conjuguée à la loi normale pour sa moyenne et matrice de covariance, voir Dickey et al. [1985]. Comme nous allons le voir cela reste en fait le cas pour  $K \neq 1$  malgré le mouvement de la source. Supposons ainsi que la loi de  $\mathbf{d}$  conditionnelle à  $\Sigma$  est une loi normale et que la loi de  $\Sigma$  est inverse Wishart (Anderson [2003]) :

$$\mathbf{d}|\Sigma, \mathbf{d}_0, \lambda_0 \sim \mathcal{N}_{LM}(\mathbf{d}_0, \lambda_0 \Sigma) \quad (11.6)$$

$$\Sigma|\Sigma_0, \nu_0 \sim \mathcal{W}_{LM}^{-1}(\Sigma_0, \nu_0) \quad (11.7)$$

où  $\Sigma_0$  est supposée symétrique définie positive,  $\lambda_0$  strictement positif et  $\nu_0 > LM$ . Notons que la distribution inverse Wishart vérifie  $E[\Sigma|\Sigma_0, \nu_0] = (\nu_0 - LM - 1)^{-1} \Sigma_0$ .

Avec ces *a priori* et sans planète ( $\boldsymbol{\alpha} = \mathbf{0}$ ), un résultat classique indique que la distribution marginale  $p(\mathbf{x}(k)|\boldsymbol{\alpha})$  est une loi de Student multivariée :

$$\mathbf{x}(k)|\boldsymbol{\alpha} = \mathbf{0} \sim \mathcal{St}_{LM} \left( \mathbf{d}_0, \frac{(\lambda_0 + 1)\Sigma_0}{\nu_0 + 1 - LM}, \nu_0 + 1 - LM \right) \quad (11.8)$$

Plus généralement, les Eq. (11.6,11.7) impliquent

$$\begin{aligned} p(\mathbf{d}, \Sigma|\mathbf{d}_0, \lambda_0, \nu_0) &= (2\pi)^{-LM/2} |\lambda_0 \Sigma|^{-1/2} \times \text{etr} \left( -\frac{1}{2\lambda_0} \Sigma^{-1} (\mathbf{d} - \mathbf{d}_0)(\mathbf{d} - \mathbf{d}_0)^t \right) \\ &\times \frac{|\Sigma_0|^{\nu_0/2}}{2^{\frac{\nu_0 LM}{2}} \Gamma_{LM}(\nu_0/2)} |\Sigma|^{-\frac{\nu_0 + LM + 1}{2}} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} \Sigma_0 \right) \end{aligned} \quad (11.9)$$

En combinant cette équation avec la vraisemblance (Eq. 11.5) on obtient :

$$p(\mathbf{x}, \mathbf{d}, \Sigma|\boldsymbol{\alpha}, \mathbf{d}_0, \lambda_0, \nu_0) = \frac{(2\pi)^{-\frac{LM(K+1)}{2}}}{2^{\frac{\nu_0 LM}{2}} \lambda_0^{\frac{LM}{2}} \Gamma_{LM}(\nu_0/2)} |\Sigma_0|^{\nu_0/2} \times |\Sigma|^{-\frac{\nu_0 + K + LM + 2}{2}} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} (\Sigma_0 + S(\mathbf{d}, \boldsymbol{\alpha}, \mathbf{x})) \right) \quad (11.10)$$

où  $S(\mathbf{d}, \boldsymbol{\alpha}, \mathbf{x})$  est défini par :

$$S(\mathbf{d}, \boldsymbol{\alpha}, \mathbf{x}) = \sum_k (\mathbf{d} - \mathbf{x}(k) + A(k)\boldsymbol{\alpha})(\mathbf{d} - \mathbf{x}(k) + A(k)\boldsymbol{\alpha})^t + \frac{1}{\lambda_0} (\mathbf{d} - \mathbf{d}_0)(\mathbf{d} - \mathbf{d}_0)^t$$

La matrice  $S(\mathbf{d}, \boldsymbol{\alpha}, \mathbf{x})$  peut s'exprimer comme une fonction quadratique de  $\mathbf{d}$  :

$$S(\mathbf{d}, \boldsymbol{\alpha}, \mathbf{x}) = (K + \lambda_0^{-1})(\mathbf{d} - h(\boldsymbol{\alpha}, \mathbf{x}))(\mathbf{d} - h(\boldsymbol{\alpha}, \mathbf{x}))^t + C(\mathbf{x}, \boldsymbol{\alpha}) \quad (11.11)$$

$$\text{avec } h(\boldsymbol{\alpha}, \mathbf{x}) = \frac{\lambda_0}{K\lambda_0 + 1} \sum_k (\mathbf{x}(k) - A(k)\boldsymbol{\alpha}) + \frac{1}{K\lambda_0 + 1} \mathbf{d}_0 \quad (11.12)$$

$$\text{et } C(\mathbf{x}, \boldsymbol{\alpha}) = \sum_k (\mathbf{x}(k) - A(k)\boldsymbol{\alpha})(\mathbf{x}(k) - A(k)\boldsymbol{\alpha})^t + \frac{1}{\lambda_0} \mathbf{d}_0 \mathbf{d}_0^t - (K + \lambda_0^{-1}) h(\boldsymbol{\alpha}, \mathbf{x}) h(\boldsymbol{\alpha}, \mathbf{x})^t \quad (11.13)$$

En substituant l'Eq. (11.11) dans l'Eq. (11.10) on constate que l'*a priori* de l'Eq. (11.9) est conjugué de l'Eq. (11.5). Une intégration de  $\mathbf{d}$  mène à :

$$p(\mathbf{x}, \Sigma|\boldsymbol{\alpha}) = \frac{(2\pi)^{-\frac{KLM}{2}}}{2^{\frac{\nu_0 LM}{2}} (K\lambda_0 + 1)^{\frac{LM}{2}} \Gamma_{LM}(\nu_0/2)} |\Sigma_0|^{\nu_0/2} |\Sigma|^{-\frac{\nu_0 + K + LM + 1}{2}} \text{etr} \left( -\frac{1}{2} \Sigma^{-1} (\Sigma_0 + C(\mathbf{x}, \boldsymbol{\alpha})) \right)$$

et une intégration de  $\Sigma$  donne finalement :

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \frac{\Gamma_{LM}(\frac{\nu_0 + K}{2})}{\Gamma_{LM}(\frac{\nu_0}{2})} \frac{|I_{LM} + \Sigma_0^{-1} C(\mathbf{x}, \boldsymbol{\alpha})|^{-\frac{\nu_0 + K}{2}}}{(\pi^K (K\lambda_0 + 1))^{LM/2} |\Sigma_0|^{K/2}} \quad (11.14)$$

## Discussion du choix des hyperparamètres

L'étude qui précède a mené à une vraisemblance  $p(\mathbf{x}|\boldsymbol{\alpha})$ , donnée dans l'Eq. (11.14), conditionnée aussi –mais de façon implicite car ce ne sont pas les paramètres testés– aux hyperparamètres  $\mathbf{d}_0, \lambda_0, \Sigma_0$  et  $\nu_0$ . Cette vraisemblance a ici été construite grâce au paradigme bayésien mais aurait aussi pu être définie telle quelle. Un des deux enjeux principaux de cette partie<sup>2</sup> consiste à valider la relative adéquation de la vraisemblance aux données simulées finement. Ceci est fait par l'application de tests de détection (fréquentiste ou bayésiens) construits à partir de cette vraisemblance.

Un traitement fréquentiste de la vraisemblance consisterait par exemple à estimer le paramètre  $\alpha$  et les hyperparamètres par la méthode de maximum de vraisemblance à partir du jeu de données  $\mathbf{x}$ . Une utilisation purement bayésienne de la vraisemblance consisterait par exemple à utiliser des informations *a priori* (mais conditionnées par l'expérience réalisée) pour déterminer les valeurs des hyperparamètres *a priori*. Pour la vraisemblance qui nous intéresse ici, les informations pourraient éventuellement être conditionnées au type spectral de l'étoile, à sa distance à la Terre, à des calibrations de la matrice de covariance spatiale, *etc.* L'approche bayésienne empirique est quant à elle présentée par Carlin and Louis [2000] notamment.

Fondamentalement, on peut noter que le choix des paramètres d'un modèle paramétrique constitue une des deux étapes dans la construction de la distribution de probabilité de  $\mathbf{x}$ , l'autre consistant à choisir le modèle paramétrique lui-même. Ainsi, le choix des hyperparamètres par une procédure systématique quantitative peut masquer le caractère qualitatif du choix du modèle paramétrique. Il est simplement nécessaire, pour être confiant en les résultats de l'inférence, que les résultats varient aussi peu que possible au sein de la zone d'*incertitude* entourant le choix joint du modèle paramétrique et des hyperparamètres.

Dans la suite, les ordres de grandeur des hyperparamètres  $\mathbf{d}_0$  et  $\Sigma_0$  sont parfois fixés à partir des données. Les autres hyperparamètres sont choisis relativement arbitrairement. Dans tous les cas, leurs valeurs sont systématiquement indiquées. Des études ultérieures seraient nécessaires mais comme nous l'avons dit l'enjeu initial consiste à valider la forme même du modèle, pour des valeurs d'hyperparamètres sélectionnées sans grande attention dans un premier temps.

## 11.3 Probabilité *a priori* de l'intensité de la planète dans les différents canaux spectraux

### *A priori* de premier niveau

La distribution *a priori* de  $\boldsymbol{\alpha}$  doit vérifier plusieurs contraintes importantes : la distribution *a priori* doit être propre si l'on souhaite pouvoir utiliser le facteur de Bayes, les intensités de la source dans les différentes bandes sont positives, elles sont plus ou moins corrélées entre elles et doivent pouvoir s'étendre sur plusieurs ordres de grandeur. Le dernier point est naturel en astronomie où l'intensité lumineuse des objets s'exprime en "magnitude", définie comme  $\text{mag} = -2.5 \log_{10}(F/F_0)$  avec  $F$  le flux électromagnétique de l'objet et  $F_0$  le flux d'un objet de référence. Les erreurs sur les intensités inférées en astronomie sont de même communément exprimées sur l'échelle des magnitudes (Wesley et al. [2008]). Spécifiquement ici, l'intensité du fond (et donc du rapport signal sur bruit à signal fixé) s'étale sur plusieurs ordres de grandeur (voir les échelles log de visualisation des profils coronographiques) et les intensités possibles des exoplanètes s'étalent elles-même sur plusieurs ordres de grandeur (Moutou et al. [2006]). On peut par exemple s'intéresser aux planètes de magnitude de l'ordre de  $15 \pm 5$ .

Il est par conséquent naturel de supposer que la distribution de  $(\ln \alpha_1, \dots, \ln \alpha_M)^t$  est conjointement gaussienne. En plus des propriétés mentionnées précédemment, l'avantage de cette distribution est qu'elle est décrite par peu de paramètres, reliés directement aux moments de la magnitude. La distribution du

<sup>2</sup>L'autre enjeu consiste à constater que pour des données distribuées selon cette vraisemblance le PLR n'est pas inadmissible par rapport au facteur de Bayes.

vecteur  $\alpha$  est appelée Log-Normale (Kotz et al. [2000]) :

$$p(\alpha|\mathbf{m}, B) = \frac{1}{|B|^{\frac{1}{2}}} \prod_{m=1}^M \frac{I_{\mathbb{R}_{+*}}(\alpha_m)}{\sqrt{2\pi\alpha_m}} \exp\left(-\frac{(\ln \alpha - \mathbf{m})^t B^{-1} (\ln \alpha - \mathbf{m})}{2}\right) \quad (11.15)$$

avec  $\mathbf{m} = E[\ln \alpha]$ ,  $B = \text{cov}[\ln \alpha]$  et  $I_{\mathbb{R}_{+*}}(\cdot)$  la fonction de Heaviside (égale à 1 sur  $\mathbb{R}_{+*}$  et nulle sinon).

### A priori marginal

Comme pour la vraisemblance, le modèle (11.15) est directement relié à une loi normale, à travers une loi log-normale cette fois et plus une moyenne dépendant du temps. Cependant, l'*a priori* normal-inverse Wishart est à nouveau conjugué pour les paramètres de nuisance. On suppose ainsi

$$\mathbf{m}|\mathbf{m}_0, c_0, B \sim \mathcal{N}_M(\mathbf{m}_0, c_0 B) \quad (11.16)$$

$$B|B_0, \omega_0 \sim \mathcal{W}_M^{-1}(B_0, \omega_0) \quad (11.17)$$

où  $B$  est supposée symétrique définie positive,  $c_0$  strictement positif et  $\omega_0 > M$ . On procède à un calcul similaire à celui effectué pour la vraisemblance. L'*a priori* marginal est alors finalement donné par :

$$\pi_1(\alpha) = \frac{(\pi(c_0 + 1))^{-\frac{M}{2}} \Gamma_M\left(\frac{\omega_0 + 1}{2}\right)}{\prod_{m=1}^M \alpha_m} \frac{\Gamma_M\left(\frac{\omega_0 + 1}{2}\right)}{\Gamma_M\left(\frac{\omega_0}{2}\right)} \frac{|B_0|^{\frac{\omega_0}{2}} \prod_{m=1}^M I_{\mathbb{R}_{+*}}(\alpha_m)}{\left|B_0 + \frac{(\ln \alpha - \mathbf{m}_0)(\ln \alpha - \mathbf{m}_0)^t}{c_0 + 1}\right|^{\frac{\omega_0 + 1}{2}}} \quad (11.18)$$

Malgré le fait que l'*a priori* (11.18) tend vers l'infini lorsque  $\alpha \rightarrow \mathbf{0}$ , on peut montrer que cet *a priori* est propre par des calculs standards sur des intégrales impropres. Notons que la distribution standard d'invariance d'échelle  $\prod_{m=1}^M \alpha_m^{-1} I_{\mathbb{R}_{+*}}(\alpha_m)$  tend aussi vers  $\infty$  lorsque  $\alpha \rightarrow \mathbf{0}$  mais est impropre, ce qui rend le facteur de Bayes inutilisable comme nous l'avons vu à maintes reprises.

## Discussion du choix de la distribution *a priori* et de ses hyperparamètres

Notons que si la vraisemblance revet un sens *relativement* clair (limite de la distribution empirique de données obtenues en réalisant un nombre infini de fois la même expérience dans des conditions identiques *et précisées*, c'est-à-dire par exemple en imageant un nombre infini de fois la même étoile dans les mêmes conditions atmosphériques par exemple), la distribution de probabilité *a priori* de l'intensité  $\alpha$  de la planète requiert elle absolument d'être explicitée car elle est construite conditionnellement à des informations ou à des souhaits qui peuvent être connus ou choisis assez différemment selon les observateurs.

Ainsi, l'expression "la probabilité pour que l'intensité de la planète soit supérieure à  $\zeta$  ..." n'a pas de sens à elle seule car elle contient toujours des présupposés (conditions  $I$  dans  $\pi(\alpha|I)$ ) tels que "*parmi toutes les planètes observables depuis la Terre*, la probabilité pour que l'intensité de la planète soit supérieure à  $\zeta$  ..." ou bien même "*parmi toutes les planètes de magnitude comprise entre 15 et 25*, la probabilité pour que l'intensité de la planète soit supérieure à  $\zeta$  ...".

## 11.4 Probabilité *a posteriori* de l'intensité et échantillonnage

### Méthode

On obtient la densité de probabilité *a posteriori* de l'intensité de la planète à partir des Eq. (11.14) et (11.18) et d'un choix raisonnable des hyperparamètres :

$$\pi_1(\alpha|\mathbf{x}) = \frac{p(\mathbf{x}|\alpha)\pi_1(\alpha)}{\int d\alpha p(\mathbf{x}|\alpha)p(\alpha)} \propto p(\mathbf{x}|\alpha)\pi_1(\alpha) \quad (11.19)$$

A partir d'un jeu de données  $\mathbf{x}$ , d'un choix des hyperparamètres et de l'expression analytique (11.19) on peut calculer quel que soit  $\boldsymbol{\alpha}$  la densité de probabilité  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$  (au facteur de normalisation  $p(\mathbf{x})$  près). Un tracé approximatif de  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$  est alors obtenu par exemple en échantillonnant uniformément  $\mathbb{R}_{+*}^M$ , en sélectionnant de façon déterministe un ensemble de valeurs de  $\boldsymbol{\alpha}$  également espacées.

Cette technique est cependant en général insuffisante pour calculer les moments *a posteriori* de  $\boldsymbol{\alpha}$  aussi bien que pour calculer la probabilité *a posteriori* de tout domaine  $C \subset \Theta$ . Pour ce faire il est préférable d'utiliser une méthode aléatoire de Monte Carlo pour échantillonner la distribution  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$ , voir Robert and Casella [1999]. Ce type d'algorithme basé sur la génération aléatoire de valeurs de  $\boldsymbol{\alpha}$  et de l'application de quelques règles (qui dépendent de la méthode considérée) résulte en un ensemble de valeurs  $\boldsymbol{\alpha}$  dont la distribution empirique doit tendre vers la distribution  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$  ou tout autre distribution visée. Parmi ces méthodes, la génération de chaînes de Markov par méthode de Monte Carlo (MCMC) est une méthode particulièrement répandue : elle consiste à obtenir la valeur numéro  $j + 1$  de l'ensemble des  $\boldsymbol{\alpha}$  en partant de la valeur numéro  $j$ . Avec les possibilités croissantes offertes par l'informatique, ces méthodes numériques prennent de plus en plus de place dans le paysage statistique, bayésien en particulier. Il existe déjà une littérature considérable sur le sujet, voir Robert and Casella [1999], Raftery et al. [2007] pour une introduction et une étude spécifique qui nous intéresse ici.

Dans la distribution *a posteriori* (11.19), le déterminant ne peut se réexprimer en un produit sur l'ensemble des composantes de  $\boldsymbol{\alpha}$ . Par conséquent, la méthode classique de Gibbs ne peut être utilisée ici. Nous utiliserons à la place "l'échantillonneur par tranche" (ou *slice sampling algorithm*) qui a des propriétés de convergence satisfaisante, est déjà programmé dans Matlab et est très facile d'utilisation.

### Application sur un jeu de données simulé simplement

Illustrons finalement l'obtention de la distribution de probabilité *a posteriori* de l'intensité  $\boldsymbol{\alpha}$  de la planète dans deux bandes spectrales à partir d'un jeu de données  $\mathbf{x}$ .

#### • Jeu de données simulé

Le jeu de données est ici simulé simplement en une dimension spatiale et pour deux canaux spectraux ( $M = 2$ ) et généré selon la vraisemblance gaussienne simple de l'équation (11.4) avec les paramètres  $L = 15$ ,  $K = 10$  et  $\boldsymbol{\alpha} = (20, 30)^t$ . De plus :

- Le profil de la source  $\mathbf{p}(k, m)$  est identique pour les deux canaux  $m \in \{1, 2\}$ . Il est de plus identique à tout instant  $k$ , en dehors d'un mouvement de translation caractérisé par une vitesse presque constante. Le profil a un support spatial de 3 pixels de large dont les valeurs sont  $(0.2, 0.5, 0.3)^t$ .
- Le fond est uniforme et identique dans les deux canaux, avec  $\mathbf{d} = 100\mathbf{1}_{LM}$ , *i.e.*  $d_i = 100 \forall i = 1, \dots, LM$ ,
- $\Sigma$  est une matrice de Toeplitz dont le premier vecteur colonne est donné par  $\mathbf{v} = 500\mathbf{u}$  avec  $u_1 = 1$ ,  $u_2 = 0.3$ ,  $u_3 = 0.05$ ,  $u_{L+1} = 0.6$ ,  $u_{L+2} = 0.2$ , toutes les autres composantes de  $\mathbf{u}$  étant nulles.

La figure 11.1 illustre le jeu de données  $\mathbf{x}$  ainsi simulé.

#### • Choix des hyperparamètres de la *priori*

La forme explicite complète de la distribution *a posteriori*  $\pi(\boldsymbol{\alpha}|\mathbf{x})$  donnée par l'Eq. (11.19) est obtenue en fixant les valeurs des hyperparamètres de la distribution *a priori* de l'Eq. (11.18) et de la vraisemblance de l'Eq. (11.14).

Les lignes de contour de  $\pi_1(\boldsymbol{\alpha})$  sont représentées sur la figure 11.2 pour  $M = 2$ ,  $\mathbf{m}_0 = (4, 4)^t$ ,  $c_0 = 0.99$ ,  $B_0 = (\omega_0 - 3)\mathbf{E}[B]$  où on fixe  $\mathbf{E}[B_{1,1}] = \mathbf{E}[B_{2,2}] = \sigma^2$  et  $\mathbf{E}[B_{1,2}] = \rho\sigma^2$  avec  $\sigma^2 = 10$  et  $\rho = 0.8$ . Le choix de  $\omega_0$  peut être fait à partir du choix des termes diagonaux de  $B$  :  $\text{var}(B_{i,i}) = \frac{2(B_0)_{i,i}^2}{(\omega_0 - 3)^2(\omega_0 - 5)^2}$ . On choisit  $\omega_0 = 4$ . Peu de corrélation est supposée entre  $\alpha_1$  et  $\alpha_2$  et une assez grande incertitude est supposée pour les différents paramètres. En pratique, le paramètres de corrélation peut être choisi de telle sorte qu'il s'ajuste à l'*a priori* de modèles physiques tels que ceux proposés par Moutou et al. [2006].



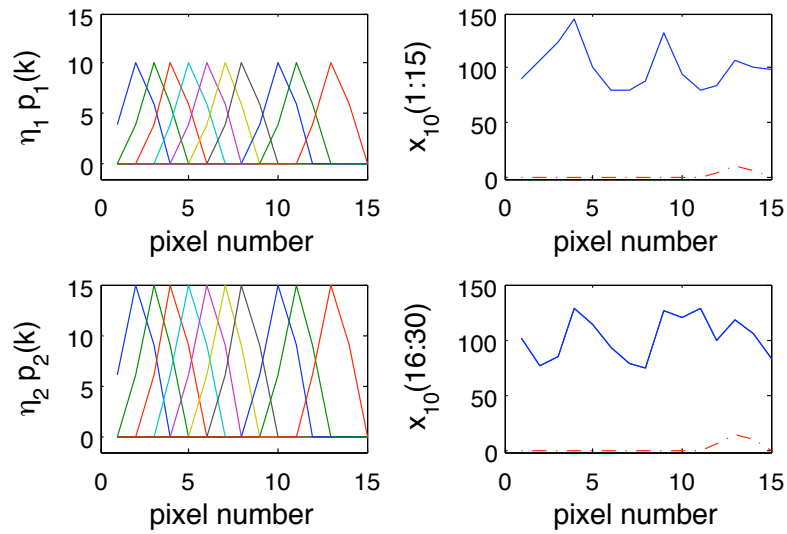


FIG. 11.1: Haut : canal  $m = 1$ . Bas : canal  $m = 2$ . Gauche : profils de la source  $\alpha_m \mathbf{p}(k, m)$  superposés pour toutes les poses  $k = 1, \dots, 10$ . Droite (ligne continue) : données  $\mathbf{x}(k)$  pour  $k = 10$  (première moitié du vecteur  $\mathbf{x}(10)$  pour  $m = 1$  et seconde moitié pour  $m = 2$ ). Droite (ligne en tirets) :  $\alpha_m \mathbf{p}(10, m)$ . Le terme “ $\eta$ ” sur la figure désigne en fait  $\alpha$ .

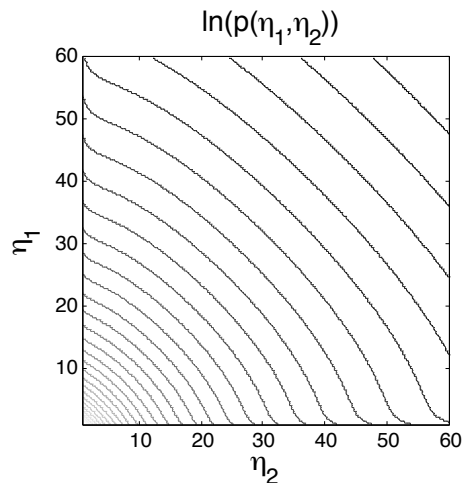


FIG. 11.2: Lignes de contour de la distribution *a priori*  $\pi_1(\boldsymbol{\alpha})$ , représentée en échelle log. Le terme “ $\eta$ ” sur la figure désigne  $\alpha$ .

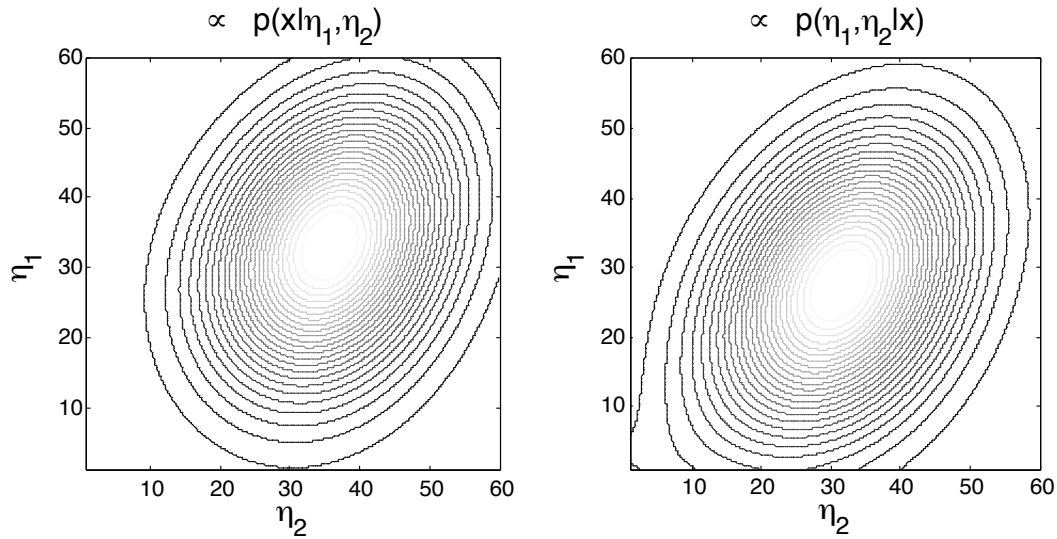


FIG. 11.3: Gauche : Lignes de contour de la vraisemblance (11.14). Droite : distribution *a posteriori* associée, pour la distribution *a priori* illustrée sur la figure 11.2. Le terme “ $\eta$ ” sur la figure désigne  $\alpha$ .

- **Choix des hyperparamètres de la vraisemblance**

Les lignes de contour de  $p(\mathbf{x}|\boldsymbol{\alpha})$  (Eq. 11.14) sont représentées sur l'image de gauche de la figure 11.3 pour  $\mathbf{d}_0 = 5001_{LM}$ ,  $\lambda_0 = 2$ ,  $\nu_0 = LM$  et  $\Sigma_0 = 4I_{LM}$  et pour les données  $\mathbf{x}$  illustrées dans la figure 11.1.

- **Echantillonnage de la loi *a posteriori***

La distribution *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$  (11.19) définie à partir de la distribution *a priori*  $\pi_1(\boldsymbol{\alpha})$  et la vraisemblance  $p(\mathbf{x}|\boldsymbol{\alpha})$  illustrées dans les figures précédentes a été échantillonnée selon la méthode MCMC décrite page 176. Le graphe de gauche de la figure 11.4 montre une partie de la chaîne de Markov pour chaque canal. Les histogrammes à droite de la figure 11.4 représentent les histogrammes (marginaux) associés à la chaîne de Markov. Les lignes de contour de la distribution *a posteriori* jointe sont représentées sur l'image de droite de la figure 11.3.

La loi *a posteriori* échantillonnée peut être utilisée à des fins d'estimation de l'intensité de la source dans les deux canaux. Rappelons que la valeur vraie était  $\boldsymbol{\alpha} = (20, 30)^t$ . Le maximum *a posteriori* pour les données  $\mathbf{x}$  est ici égal à  $\hat{\boldsymbol{\alpha}}_{MAP} = (27.5, 31.4)^t$ , la moyenne *a posteriori* est égale à  $\hat{\boldsymbol{\alpha}}_{PM} = (27.3, 31.1)^t$  et les déviations standards de l'intensité dans chaque canal sont égales à  $(12.2, 10.8)^t$ , ce qui est relativement large mais naturel étant donné la méconnaissance totale initiale de la valeur d'intensité et le profil relativement peu piqué de la vraisemblance.

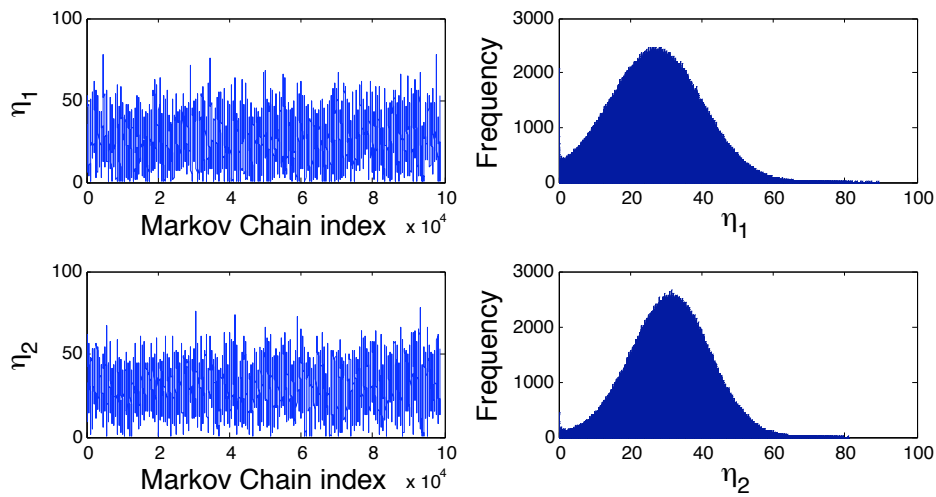


FIG. 11.4: Illustrations des échantillons  $\{(\alpha_1, \alpha_2)^{[j]}, j \in \llbracket 1, N_J \rrbracket\}$  de la loi *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\boldsymbol{x})$ . Haut : canal 1. Bas : canal 2, échantillons correspondants. Gauche : sous-ensemble de la chaîne de Markov constitué de  $10^5$  échantillons. Droite : histogrammes correspondants, mais pour la chaîne de Markov complète. Ces histogrammes sont les distributions empiriques servant à estimer les lois marginales  $\pi(\alpha_m|\boldsymbol{x})$ . Le terme “ $\eta$ ” sur la figure désigne  $\alpha$ .

## Chapitre 12

# Détection par le PLR pour les données SPHERE

### 12.1 Jeux de données issus de la simulation physique fine

#### Description des données simulées

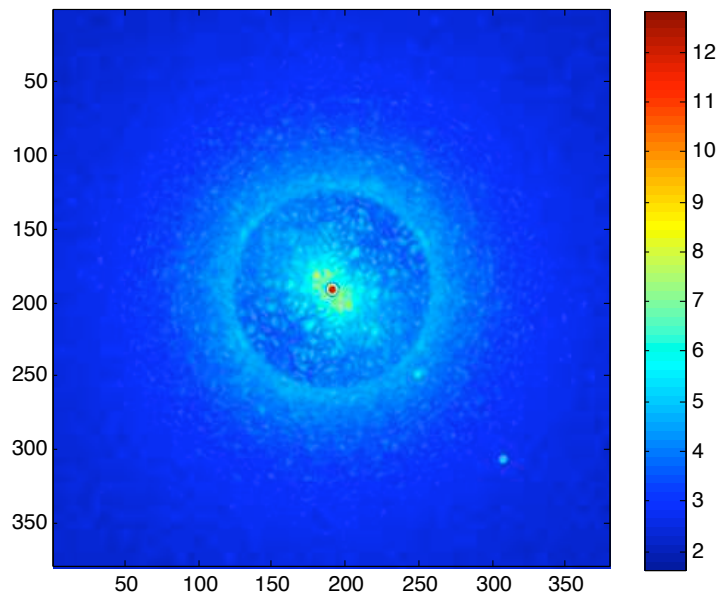


FIG. 12.1: Image obtenue à partir d'une simulation *end-to-end* du futur instrument VLT-SPHERE et élevée à la puissance 0.2. Quatre planètes sont simulées le long de la diagonale inférieure droite avec un contraste particulièrement bas de  $10^4$  : l'étoile est *seulement* 10000 fois plus lumineuse que les planètes. On considère dans cette section le même jeu de données mais pour un contraste de  $10^6$  et en se focalisant sur l'extrême centre de l'image.

Les tests de détection sont appliqués aux images issues de la simulation physique fine de l'instrument SPHERE décrite dans la section 2.1 p. 35. En l'absence de planète et pour un canal spectral, cet ensemble d'images est présenté dans le chapitre 2 et détaillé dans l'annexe C.1. En présence d'une planète lumineuse (contraste d'intensité de  $10^4$  entre l'étoile et la planète), l'ensemble d'images est illustré à nouveau sur la figure 12.1. Ces deux jeux de données respectivement sous  $H_0$  et sous  $H_1$  sont traités dans la section 4.3 p. 84 par la méthode issue du modèle gaussien simple. Rappelons de plus que le mode DBI du sous-instrument IRDIS de SPHERE permet l'acquisition simultanée d'images dans deux canaux

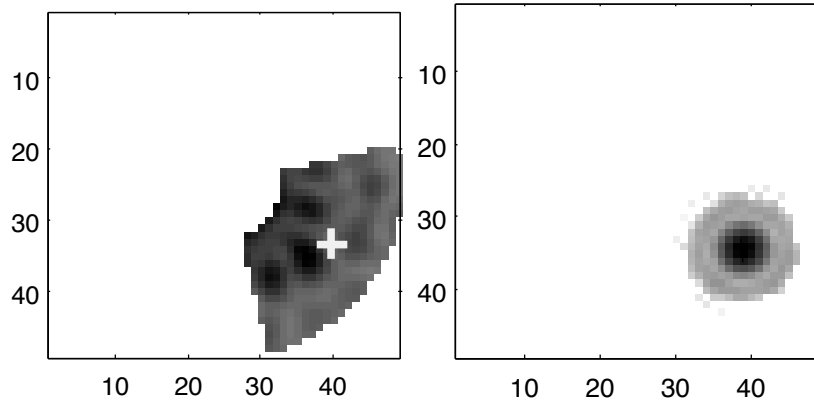


FIG. 12.2: Données simulées par le *package* de simulation SPHERE du *problem-solving environment* CAOS avec un contraste de  $10^6$  entre l'intensité de l'étoile et celle la planète et pour une séparation entre l'étoile et la planète de  $0''.2$ . Gauche : image  $\mathbf{x}(20, 2)$  élevée à la puissance 0.2. La croix indique la position de la planète. Droite : profil  $\mathbf{p}(20, 2)$  de la planète élevée à la puissance 0.2. Le profil de la planète est supposé connu, y compris sa position.

spectraux et est aussi simulé par le *package* de simulation SPHERE. Les données ainsi simulées, toujours pour un contraste de  $10^4$ , indiquent que l'introduction de plusieurs canaux spectraux dans le modèle gaussien simple, proposée dans la section 3.5 p. 71, améliore parfois mais pas systématiquement l'estimation d'intensité de la planète dans les deux canaux spectraux par rapport à des estimations indépendantes dans les deux canaux (voir figure 4.8 p. 88).

Dans la présente section, on traite le même jeu de données que discuté précédemment, et en utilisant les deux canaux spectraux, à la différence que les planètes sont maintenant remplacées par des planètes 100 fois moins lumineuses, que l'on ne traite que les  $K = 20$  ou  $K = 50$  premières images et non les  $K = 144$  images simulées (les conditions atmosphériques correspondantes sont indiquées sur la figure 2.1 p. 38) et que l'on s'intéresse exclusivement à la zone la plus proche de l'étoile, c'est-à-dire à la zone où les méthodes classiques de détection sont les moins performantes. Le contraste de  $10^6$  entre l'étoile et la planète correspond ici à une intensité de planète égale à  $\boldsymbol{\alpha}^* = (8, 0.5)^t \cdot 10^{-5}$  (vecteur indiquant l'intensité de la planète respectivement dans les canaux dit H2 ( $\lambda_1 \sim 1.60\mu m$ ) et H3 ( $\lambda_2 \sim 1.65\mu m$ )<sup>1</sup>.

Au sein de cet ensemble d'images simulées en présence de planètes sur la diagonale inférieure droite, on définit plusieurs *jeux de données*, un jeu de données étant caractérisé par une position<sup>2</sup>  $r$  et par le nombre d'images  $K$  prises en compte. Les 3 positions  $r$  testées dans cette section sont situées sur la diagonale inférieure droite. Ainsi, le jeu de données caractérisé par une séparation de  $0''.2$  correspond à des données sous  $H_1$  puisqu'une planète y est simulée, celui caractérisé par une séparation de  $0''.35$  satisfait l'hypothèse nulle  $H_0$  puisqu'aucune exoplanète n'y est simulée et que l'intensité des autres planètes y est négligeable, et à  $0''.5$  le jeu de données satisfait l'hypothèse  $H_1$  à nouveau.

<sup>1</sup>Le contraste est indiqué dans le canal spectral dans lequel l'intensité de la planète est la plus élevée (par choix des bandes spectrales dans le mode DBI de l'instrument SPHERE l'intensité de l'étoile varie quant à elle peu entre les deux canaux, voir la section 3.5 p. 71). Notons par ailleurs qu'en utilisant les images acquises dans le canal dans lequel l'intensité de la planète est la plus élevée et en utilisant la totalité des images, comme on le voit empiriquement en bas à droite de la figure 4.7 p. 87, la méthode gaussienne initiale de détection était totalement inefficace pour un contraste de  $10^6$ , et ce non seulement pour la zone la plus interne du champ mais même pour la zone la plus externe.

<sup>2</sup>On rappelle que le test est appliqué pour une position initiale de planète  $r$  fixée, fixant les réponses de la planète  $\mathbf{p}(k, r)$  pour tout  $k$ , de sorte qu'un jeu de données peut être défini en choisissant une position  $r$  plutôt qu'une autre.

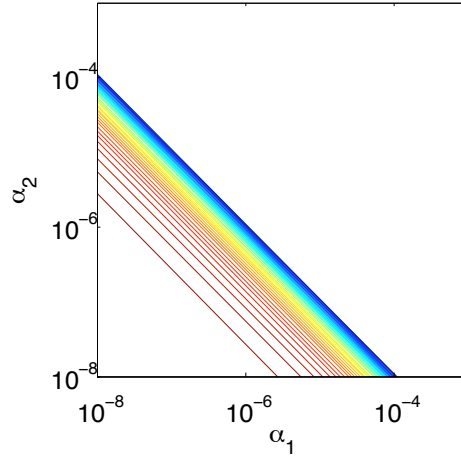


FIG. 12.3: Lignes de contour de l'*a priori*  $\pi_1(\boldsymbol{\alpha})$  utilisé pour le jeu de données simulé finement. Elles se confondent parfaitement avec celles de l'*a priori* invariant classique, si les composantes sont indépendantes,  $\pi_1(\boldsymbol{\alpha}) \propto \alpha_1^{-1} \alpha_2^{-1}$  pour  $\alpha_m > 0$ .

### Valeurs des hyperparamètres

La distribution *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$  (donnée par l'équation (11.19)) de l'intensité de la planète dans chaque bande spectrale est définie en explicitant non seulement les données  $\mathbf{x}$  utilisées mais aussi les hyperparamètres de modélisation.

Les hyperparamètres associés à la vraisemblance  $p(\mathbf{x}|\boldsymbol{\alpha})$  donnée par l'équation (11.14) p. 174 sont choisis de la façon suivante :

- La moyenne du fond  $\mathbf{d}_0$  est prise constante spatialement et spectralement avec un ordre de grandeur donné par les données. Ainsi,  $\mathbf{d}_0 = c\mathbf{1}_{LM}$  où le coefficient  $c$  est défini comme la médiane de l'ensemble des données, *i.e.* de tous les points  $\{x_l(k, m), l \in \llbracket 1, L \rrbracket, k \in \llbracket 1, K \rrbracket, m \in \llbracket 1, M \rrbracket\}$ .
- Le coefficient scalaire caractérisant la covariance de  $\mathbf{d}$  dans la construction du modèle hiérarchique est fixé à  $\lambda_0 = 2$ .
- Le coefficient  $\nu_0$  est fixé à  $\nu_0 = LM$ .
- La matrice de covariance  $\Sigma_0$  est prise proportionnelle à l'identité avec un ordre de grandeur donné par les données. Ainsi,  $\Sigma_0 = c'I_{LM}$  où le coefficient  $c'$  est défini comme la variance empirique de l'ensemble des données, *i.e.* de tous les points  $\{x_l(k, m), l \in \llbracket 1, L \rrbracket, k \in \llbracket 1, K \rrbracket, m \in \llbracket 1, M \rrbracket\}$ .

Les hyperparamètres associés à la distribution *a priori*  $\pi_1(\boldsymbol{\alpha})$  donnée par l'équation (11.18) p. 176 sont choisis de façon à ce que l'allure de la distribution *a priori* soit la même que l'*a priori* associé à l'invariance d'échelle sur chaque composante ( $\pi_1(\boldsymbol{\alpha}) \propto \prod_m \alpha_m^{-1}$  pour  $\alpha_m > 0$ ).

- La magnitude moyenne est choisie en fixant l'intensité moyenne de la planète à 1000 fois plus que ce qu'elle n'est en fait sous  $H_1$  :  $\mathbf{m}_0 = \ln(1000\boldsymbol{\alpha}^*)$ .
- Le coefficient scalaire caractérisant la covariance de  $\mathbf{m}$  dans la construction du modèle hiérarchique est fixé à  $c_0 = 0.7$ .
- Le coefficient  $\omega_0$  est fixé à  $\omega_0 = M + 5$ .
- Se souvenant que le nombre de canaux spectraux est égal à  $M = 2$ ,  $B_0$  est une matrice de Toeplitz dont le premier vecteur colonne est défini par  $4\sigma_B^2(1, \rho_B)^t$  avec  $\sigma_B^2 = 10$  et le terme de corrélation entre les deux canaux fixé à  $\rho_B = 0.5$ .

Les lignes de contour associées à cet *a priori* sont indiquées sur la figure 12.3 et coïncident avec celles associées à l'*a priori*  $\pi_1(\boldsymbol{\alpha}) \propto \prod_m \alpha_m^{-1}$ .

Une étude quantitative de la robustesse des résultats d'estimation et de détection et non-détection vis-à-vis d'un changement des valeurs d'hyperparamètres au sein d'un domaine *raisonnable* est nécessaire à terme mais n'est pas réalisée ici.

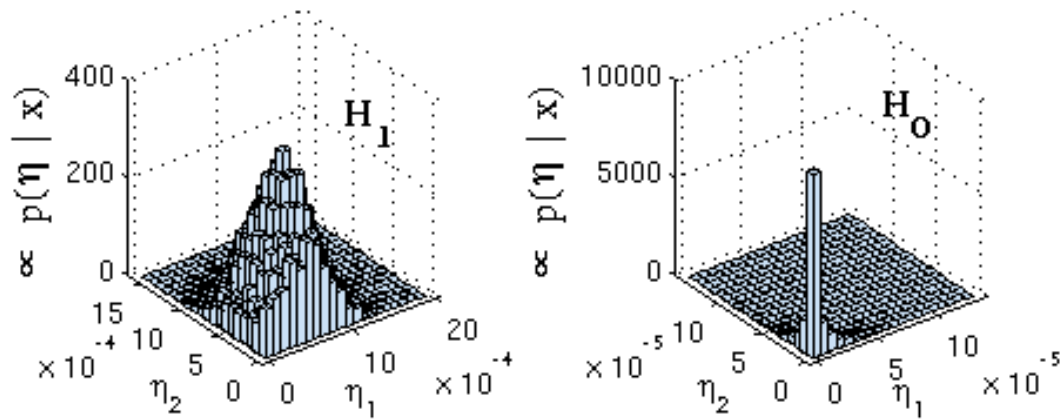


FIG. 12.4: Histogrammes des deux chaînes de Markov  $\alpha^{[j]} \sim \pi_1(\alpha|\mathbf{x})$  résultant des données avec planète ( $H_1$ , à gauche) et sans ( $H_0$ , à droite). A nouveau, sur le graphe  $\eta$  représente  $\alpha$ .

### Echantillonnage de la distribution *a posteriori* de $\alpha$

Dans un premier temps, uniquement deux jeux de données  $\mathbf{x}$  (un sous  $H_1$  et un sous  $H_0$ ) sont considérés parmi ceux présentés p. 181. Ils sont tous deux constitués des  $K = 20$  premières images de l'ensemble d'images, l'un au niveau de l'anneau de séparation moyenne 0''2 ( $H_1$ ), l'autre pour une séparation moyenne de 0''35 ( $H_0$ ).

Ainsi, les données sous  $H_1$  sont de taille  $(K, L, M) = (20, 425, 2)$ , soit 20 poses, 425 pixels et 2 bandes spectrales. Elles sont illustrées sur la figure 12.2.

Comme expliqué dans la section 11.4, la distribution *a posteriori*  $\pi_1(\alpha|\mathbf{x})$  donnée par l'équation (11.19) est échantillonnée par une méthode MCMC. Chaque chaîne de Markov est constituée de  $N_J = 2 \cdot 10^4$  échantillons  $\alpha^{[j]}$  minimum<sup>3</sup>. La figure 12.4 montre les histogrammes des chaînes de Markov résultant des deux jeux de données associés aux hypothèses  $H_0$  et  $H_1$ .

On calcule aisément les moments de ces distributions à partir des chaînes de Markov  $\{\alpha^{[j]}, j = 1, \dots, N_J\}$  représentées sur la figure 12.4. Utilisant la moyenne et l'écart type *a posteriori* on peut ainsi effectuer une estimation ponctuelle entachée d'une incertitude symétrique de l'intensité d'une potentielle planète pour chaque jeu de données. Sous  $H_1$ , on obtient ainsi  $\hat{\alpha} = (6.2 \pm 2.8, 4.6 \pm 2.6)^t \cdot 10^{-4}$  (la valeur vraie étant  $\alpha^* = (8, 0.5)^t \cdot 10^{-5}$ ) et sous  $H_0$  on obtient  $\hat{\alpha} = (8 \pm 13, 8 \pm 13)^t \cdot 10^{-6}$  (la valeur vraie étant  $\alpha_{H_0} = \mathbf{0}$ ). Notons qu'estimer un intervalle de confiance symétrique est peu pertinent pour les données sous  $H_0$  car comme on peut le voir sur l'histogramme de droite la distribution *a posteriori* de  $\alpha$  est hautement asymétrique. Quoiqu'il en soit l'estimation ne serait en pratique pas effectuée car comme on le verra aucune planète ne sera détectée sur ce jeu de données.

La discussion quantitative de ces estimées est difficile sans notamment une étude de leur robustesse vis-à-vis du choix des hyperparamètres. Quoiqu'il en soit, la phase d'estimation est considérée ici comme *moins importante* que la phase de détection car la phase de détection constitue la première phase de l'inférence de détection-estimation. Procédons donc maintenant à l'étape de détection effectuée grâce au PLR notamment.

### Détection par le PLR et le facteur de Bayes

On rappelle que pour la détection d'exoplanète, en supposant la position initiale  $r$  connue pour des données entachées de la présence d'une planète, un test d'hypothèses consiste à choisir entre les deux hypothèses suivantes :

$$H_0 : \alpha = \mathbf{0} \quad H_1 : \alpha \neq \mathbf{0} \quad (12.1)$$

<sup>3</sup>La vitesse de convergence de l'algorithme d'échantillonnage sera indiquée qualitativement sur la figure 12.7.

On rappelle (voir la section 7.1) qu'on entend en fait plutôt pour  $H_1$  le fait que  $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$  où  $\boldsymbol{\alpha}^* \neq \mathbf{0}$  est l'intensité vraie de la planète mais est inconnue. Sa méconnaissance est quantifiée avant l'obtention des données par la distribution  $\pi_1(\boldsymbol{\alpha})$  en l'occurrence donnée par l'équation (11.18).

Les statistiques bayésiennes de test utilisées ici pour l'étape de détection sont la distribution *a posteriori* du rapport de vraisemblance (PLR) étudiée dans la partie III p. 121 et le facteur de Bayes (BF) classique introduit dans la section 1.3 et en particulier dans l'équation (1.41) p. 27 et apparaissant dans certaines propriétés du PLR. Comme ceci est décrit dans la section 8.3 p. 141, ces deux statistiques peuvent être très facilement estimées à partir des échantillons *a posteriori*  $\{\boldsymbol{\alpha}^{[j]}, j \in \llbracket 1, N_J \rrbracket\}$ .

### • Détection par le PLR

Le PLR peut être utilisé comme statistique de test binaire :

$$\text{Rejeter } H_0 \text{ si } \text{PLR}(\boldsymbol{x}, \zeta) > p$$

Dans cette section, comme Dempster [1974] et Aitkin [1997] et comme le suggère Royall [1986] notamment, on effectue le test binaire en attribuant une interprétation directe des seuils  $\zeta$  et  $p$  (voir la section 7.3 p. 128) et en les fixant à un certain couple de valeurs  $(\zeta_0, p_0)$ . De la même façon, l'approche de Neyman consisterait à fixer une valeur de PFA, grandeur dont l'interprétation est naturelle aussi.

Il s'agit ainsi de rejeter l'hypothèse nulle  $H_0$  lorsque la vraisemblance  $p(\boldsymbol{x}|\boldsymbol{\alpha})$  est "souvent" (sous la distribution *a posteriori* de  $\boldsymbol{\alpha}$  sous  $H_1$ ) "beaucoup" plus élevée que la vraisemblance  $p(\boldsymbol{x}|\boldsymbol{\alpha} = \mathbf{0})$ . Ici, nous choisissons qu'une vraisemblance est "beaucoup" plus importante qu'une autre lorsqu'elle est 10 fois plus importante ( $\zeta_0 = 0.1$ ). Cependant, le fait que la vraisemblance  $p(\boldsymbol{x}|\boldsymbol{\alpha})$  soit plus de 10 fois plus grande que  $p(\boldsymbol{x}|\boldsymbol{\alpha} = \mathbf{0})$  ne peut pas se produire avec une probabilité 1 puisque l'intensité  $\boldsymbol{\alpha}$  de la planète n'est pas connue de façon certaine. On choisit en l'occurrence que l'événement qui nous intéresse ( $p(\boldsymbol{x}|\boldsymbol{\alpha}) \geq 10p(\boldsymbol{x}|\boldsymbol{\alpha}_0)$ ) se produit "souvent" si cela survient pour une probabilité *a posteriori* sous  $H_1$  supérieure à  $p_0 = 0.8$ .

On vérifie si le test binaire ainsi défini rejette l'hypothèse nulle ou non en lisant directement la figure 12.6 qui représente  $\text{PLR}(\boldsymbol{x}, \zeta)$  en fonction de  $\zeta$ , le graphe de gauche étant obtenu à partir des données  $\boldsymbol{x}$  avec planète et le graphe de droite à partir des données  $\boldsymbol{x}$  sans planète. On lit pour les données sous  $H_1$  que  $\text{PLR}(\boldsymbol{x}, 0.1) = 0.94$ , c'est-à-dire que la probabilité *a posteriori* sous  $H_1$  que la vraisemblance  $p(\boldsymbol{x}|\boldsymbol{\alpha})$  sous  $H_1$  soit plus de 10 fois supérieure à celle sous  $H_0$  est égale à 0.94. Cette valeur est élevée et en particulier supérieure à  $p_0 = 0.8$ , de sorte que le test rejette bien  $H_0$ . On constate pour les données sous  $H_0$  que  $\text{PLR}(\boldsymbol{x}, 0.1) = 0$ , c'est-à-dire qu'il y a une probabilité *a posteriori* nulle (à l'erreur d'échantillonnage près) sous  $H_1$  que la vraisemblance  $p(\boldsymbol{x}|\boldsymbol{\alpha})$  sous  $H_1$  soit plus de 10 fois supérieure à celle sous  $H_0$ .

Plus généralement, le PLR, sorte de *p-value* bayésienne (avec une valeur  $\zeta_0$  éventuellement différente de 1 pour le PLR considéré ici), indique les couples limites  $(\zeta_0, p_0)$  pour lesquels le test est à la limite de rejeter  $H_0$ . Ainsi, si par exemple pour une valeur choisie initialement  $(\zeta_0, p_0)$  le test rejette  $H_0$ , on peut lire si le test était loin ou non d'un test (autre couple  $(\zeta'_0, p'_0)$ ) qui aurait accepté  $H_0$ . En l'occurrence, pour chacun des deux jeux de données, le PLR *indique largement* la bonne hypothèse.

### • Détection par le facteur de Bayes

Le BF peut être utilisé comme statistique de test binaire :

$$\text{Rejeter } H_0 \text{ si } \text{BF}(\boldsymbol{x}) \leq \zeta$$

Comme pour le PLR, on peut souhaiter fixer le seuil  $\zeta$  à une valeur  $\zeta_0$  prédéfinie. L'échelle dite de Jeffreys associe ainsi, par convention, des termes qualitatifs à des valeurs prises par le BF (Kass and Raftery [1995]). La correspondance, à  $\boldsymbol{x}$  fixé, avec un test construit à partir du PLR, est indiquée dans la sous-section 8.2 p. 138. Un observateur utilisateur du BF pourrait par exemple souhaiter fixer  $\zeta_0 = 0.1$ .



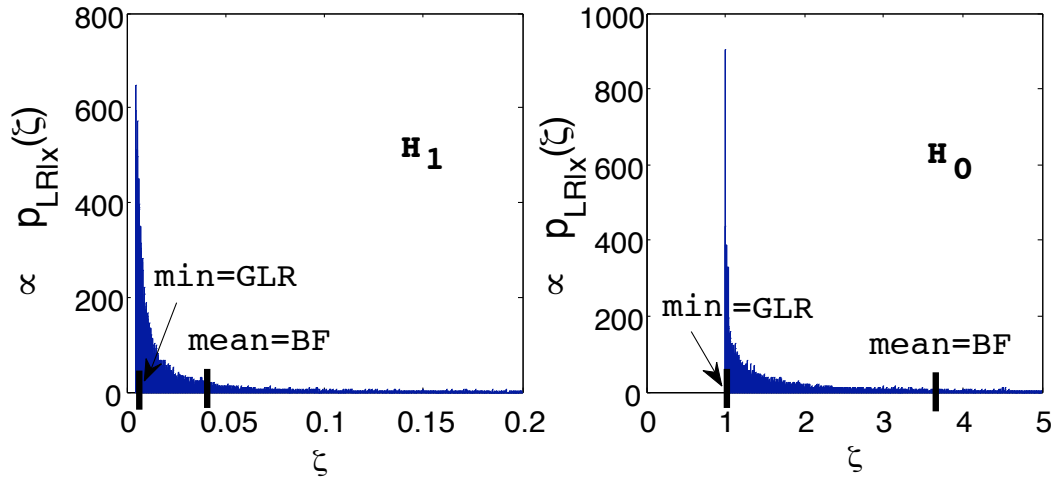


FIG. 12.5: Densités empiriques *a posteriori* de LR, tracées à partir des chaînes  $\alpha^{[j]}$  présentées sur la figure 12.4. Le GLR peut être directement lu comme le minimum du support de la distribution du LR (proposition 9 p. 135). Le BF peut être estimé qualitativement à partir de la figure car il correspond à la moyenne de la distribution (théorème 3 p. 136).

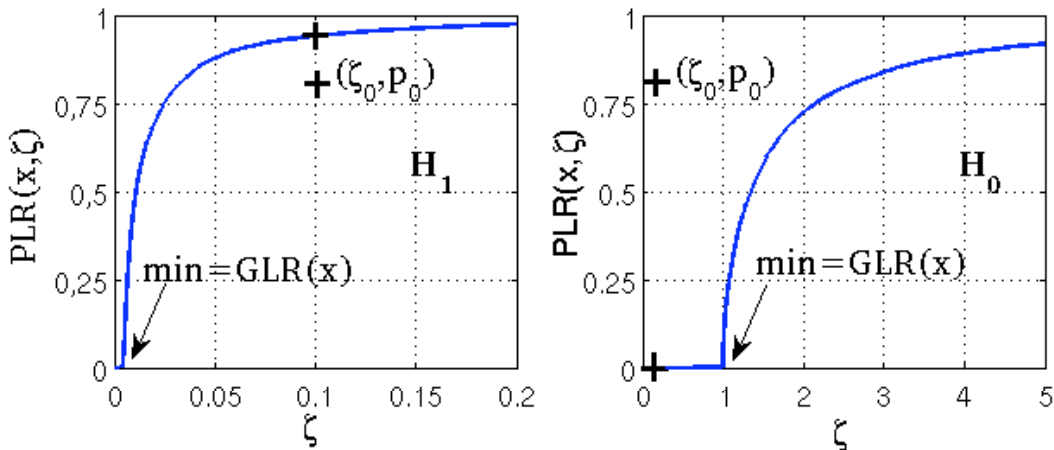


FIG. 12.6: Distributions cumulatives empiriques *a posteriori* de LR correspondant aux densités empiriques de LR montrées sur la figure 12.5. Elles sont tracées en pratique à partir des chaînes  $\alpha^{[j]}$  présentées sur la figure 12.4. On lit à gauche que  $H_0$  est rejeté par le test PLR pour  $\zeta_0 = 0.1$  et  $p_0 = 0.8$  car alors  $\text{PLR}(\mathbf{x}, \zeta_0) = \Pr(\text{LR}(\mathbf{x}, \alpha) < \zeta_0 | \mathbf{x}) > p_0$ .

Comme ceci est décrit dans la section 8.3 p. 141, le BF peut comme le PLR être très facilement (mais pas de façon optimale, voir la sous-section 8.1 p. 136) estimé à partir des échantillons *a posteriori*  $\{\alpha^{[j]}, j \in \llbracket 1, N_J \rrbracket\}$ . La valeur du BF pour chaque jeu de données est illustré sur la figure 12.5.

Comme pour le PLR, le BF choisit la bonne hypothèse pour chaque jeu de données :  $\text{BF}(\mathbf{x}) = 0.04 < \zeta_0$  pour les données sous  $H_1$  et  $\text{BF}(\mathbf{x}) = 3.7 > \zeta_0$  pour les données sous  $H_0$ . De plus, les valeurs de BF sont chaque fois assez éloignées du seuil fixé.

#### • Discussion concernant le GLR

Rappelons par ailleurs que le GLR (statistique fréquentiste) est indiquée par le PLR, en tant que première valeur de seuil  $\zeta_0$  pour laquelle  $\text{PLR}(\mathbf{x}, \zeta_0) > 0$ , voir la figure 12.5 ou 12.6. Cependant, comme discuté dans la sous-section 8.2 p. 137, il est difficile de fixer directement, par *interprétation*, une valeur

de seuil  $\zeta$  pour le test binaire basé sur  $\text{GLR}(\mathbf{x})$  car cette statistique est inférieure à 1 dans le cas qui nous intéresse. Pour utiliser le GLR il serait donc nécessaire de fixer le seuil  $\zeta$  à partir de l'approche de Neyman consistant à fixer une grandeur fréquentiste : la PFA. Calculer le GLR pour la vraisemblance considérée ici (Eq. 11.14) est déjà relativement difficile, mais calculer sa distribution fréquentiste l'est encore davantage. Notons en particulier qu'échantillonner l'espace des paramètres  $\Theta$  (selon la distribution *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$ ) est en général plus rapide qu'échantillonner l'espace des observations  $\mathcal{X}$  (selon la distribution fréquentiste  $p(\mathbf{x}|\boldsymbol{\alpha})$ ) qui se doit en effet d'être de dimension plus grande.

## Données constituées de plus de poses

Le PLR est appliqué à davantage de jeux de données. Ceux-ci ont été décrits dans la sous-section 12.1 p. 181. En particulier, on considère les données sous  $H_1$  pour lesquelles une planète est située respectivement à  $0''.2$  (cas précédent) et à  $0''.5$  du centre de l'image et les données sous  $H_0$  prises en étudiant un anneau dont les pixels  $\ell$  sont, en moyenne séparés du centre de  $0''.35$  (cas précédent). De plus, deux nombres de poses différents, débutant chaque fois à la pose  $k = 1$  selon l'indexation du jeu de 144 images<sup>4</sup>, sont pris en compte :  $K = 20$  (cas précédent) et  $K = 50$ .

Dans un premier temps, la figure 12.7 indique qualitativement la vitesse de convergence de l'échantillonnage de la loi *a posteriori* de  $\boldsymbol{\alpha}$ , vue du point de vue de l'allure de la distribution cumulative *a posteriori* du LR car il s'agit finalement de la grandeur qui nous intéresse. Ces graphes nous donnent en particulier une idée de l'incertitude associée à un nombre d'échantillons choisi relativement faible.

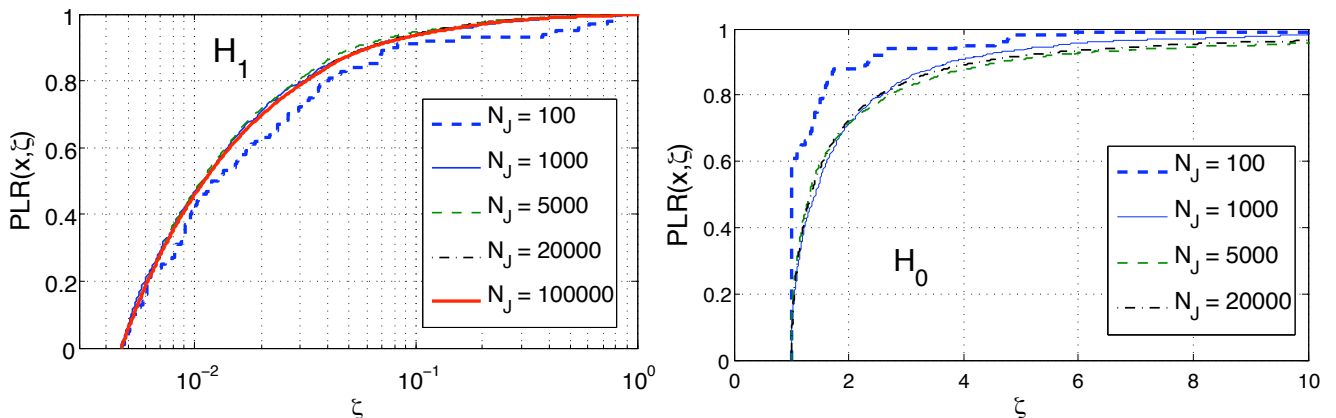


FIG. 12.7: Distributions cumulatives empiriques *a posteriori* de LR obtenues à partir de chaînes  $\{\boldsymbol{\alpha}^{[j]}, j \in \llbracket 1, N_J \rrbracket\}$  de différentes tailles  $N_J$  pour des données constituées de  $K = 20$  poses. A gauche : données de séparation moyenne égale  $0''.2$  ( $H_1$ ). A droite : données de séparation moyenne égale à  $0''.35$  ( $H_0$ ).

Finalement, la figure 12.8 montre le PLR pour les différents jeux de données décrits ci-dessus et pour un nombre d'échantillons égal à 1000. On observe des performances du PLR similaires pour la planète distante de  $0''.2$  et la planète distante de  $0''.5$  de l'axe optique. Pour un nombre de poses passant de  $K = 20$  à  $K = 50$ , on constate de plus une très nette amélioration des performances du PLR pour les données sous  $H_1$  et aucun changement significatif de performance pour les données sous  $H_0$ .

Ces constatations sont encourageantes mais ne peuvent être discutées sans appliquer le test du PLR sur davantage de jeux de données. Une étude systématique des performances du PLR requiert une programmation plus efficace de l'algorithme bayésien et/ou une parallélisation du code sur les différentes positions à tester.

<sup>4</sup>Ces poses correspondent à des valeurs de vent relativement faibles (figure 2.1 p. 38) et (ainsi) une valeur moyenne d'intensité environ 1.5 fois plus faible que pour les autres poses (figure 2.7 p. 2.7). Le vent est en moyenne plus faible pour un jeu de données constitué de 50 poses qu'un jeu de données constitué de 20 poses.

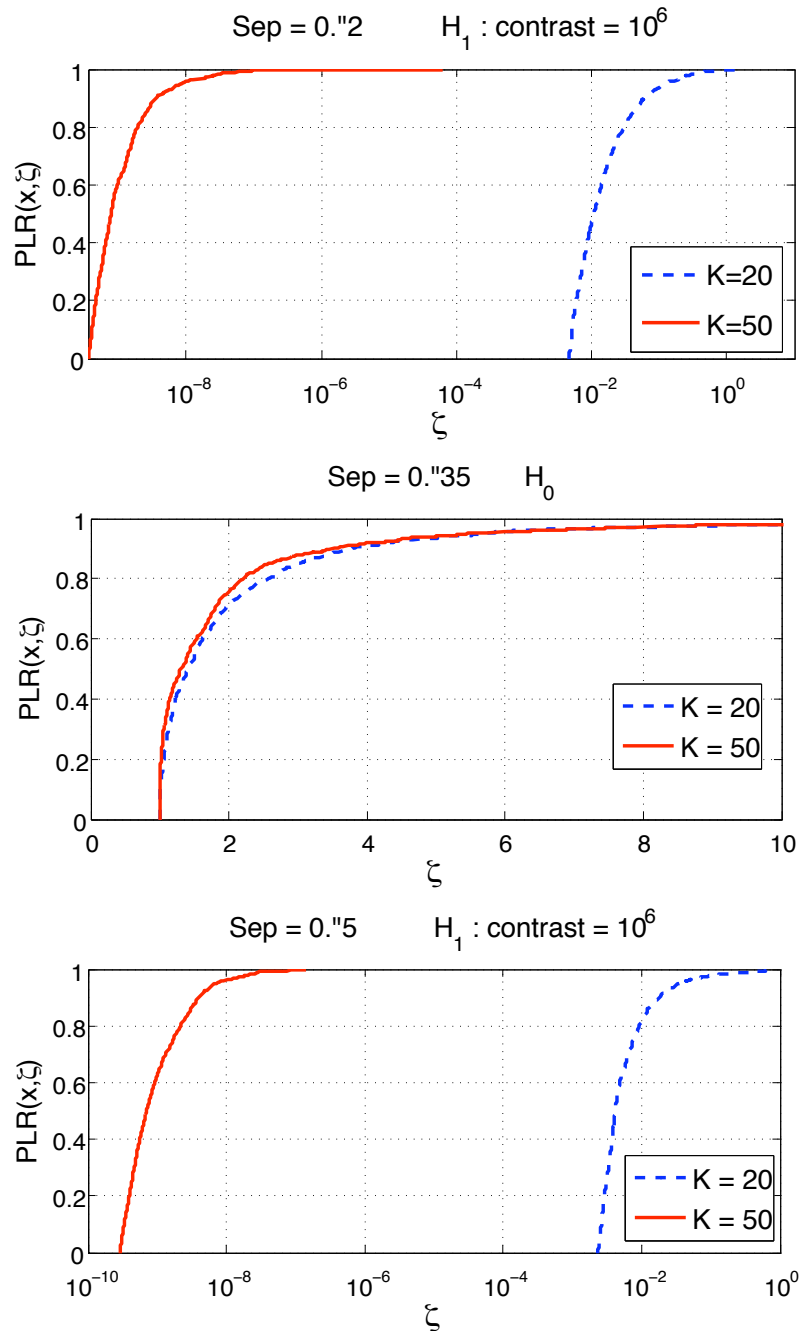


FIG. 12.8: PLR pour deux nombres de poses différents et pour différentes séparations, les séparations correspondant soit à des données sous  $H_1$  soit à des données sous  $H_0$ . En pointillés : PLR appliqué aux  $K = 20$  premières images. En trait continu : PLR appliqué aux  $K = 50$  premières images. En haut : séparation de  $0.''2$  ( $H_1$ ). Au milieu : séparation de  $0.''35$  ( $H_0$ ). En bas : séparation de  $0.''5$  ( $H_1$ ).

## 12.2 Performances fréquentistes évaluées sur des données simulées plus simplement

Étudions maintenant les performances fréquentistes du test binaire construit à partir du PLR. Par performances *fréquentistes* nous entendons que ce sont les performances qui seraient observées à partir d'un ensemble de jeux de données  $\{\mathbf{x}^{[i]}, i \in \llbracket 1, N_I \rrbracket\}$  qui serait obtenu en réalisant un nombre  $N_I$  infini

de fois une certaine expérience. En l'occurrence l'expérience à réaliser un nombre infini de fois est celle décrite de façon sous-jacente par la vraisemblance  $p(\mathbf{x}|\boldsymbol{\alpha}^*)$ , où  $\boldsymbol{\alpha}^* = \mathbf{0}$  sous  $H_0$  et  $\boldsymbol{\alpha}^* = \boldsymbol{\alpha}_1$  sous  $H_1$  avec  $\boldsymbol{\alpha}_1 \neq \mathbf{0}$  fixé. Pour un test paramétré par sa "probabilité de fausse alarme" PFA (*i.e.* son niveau dit de *significativité*), une courbe ROC représente ainsi la "probabilité de bonne détection" du test PD( $\boldsymbol{\alpha}_1$ ) en fonction de sa PFA, où PD est une probabilité calculée à partir de la distribution des données  $p(\mathbf{x}|\boldsymbol{\alpha}_1)$  sous  $H_1$ , pour une valeur  $\boldsymbol{\alpha}_1$  fixée, et PFA est une probabilité calculée à partir de la distribution des données  $p(\mathbf{x}|\mathbf{0})$  sous  $H_0$ .

Tout d'abord, nous allons constater que le PLR n'est pas inadmissible par rapport au BF, c'est-à-dire qu'il existe des valeurs  $\boldsymbol{\alpha}_1$  telles que  $\text{PD}(\text{PLR}, p, \boldsymbol{\alpha}_1) > \text{PD}(\text{BF}, p, \boldsymbol{\alpha}_1)$  où  $p$  désigne la valeur de PFA fixée, la valeur choisie pour  $p$  contraignant le(s) paramètre(s) définissant les tests binaires associés respectivement au PLR et au BF.

Par ailleurs (et cela précise l'assertion précédente), comme expliqué dans la section 8.2 p. 138, à travers ses paramètres d'entrée  $(\zeta_0, p_0)$  le PLR définit une famille de tests binaires. Dans le cas invariant discuté dans les chapitres précédents néanmoins, l'ensemble des courbes ROC se confond en une seule courbe (voir p. 140). Ici, nous ne sommes pas, du moins pas strictement, dans le cas invariant mais allons constater que les performances varient tout de même peu en fonction de la paramétrisation  $(\zeta_0, p_0)$  choisie.

Ces conclusions seront tirées de l'application des tests à un grand nombre de jeux de données, chacun étant simulé plus simplement que dans la section précédente, selon la vraisemblance marginale  $p(\mathbf{x}|\boldsymbol{\alpha})$  donnée dans l'équation (11.14) et utilisée dans la distribution *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$  de  $\boldsymbol{\alpha}$  sous l'hypothèse alternative.

### Echantillonnage de $p(\mathbf{x}|\boldsymbol{\alpha}^*)$ et non de $p_1(\mathbf{x}) = \int d\boldsymbol{\alpha} p(\mathbf{x}|\boldsymbol{\alpha})\pi_1(\boldsymbol{\alpha})$

Discutons tout d'abord le fait que pour des données sous  $H_1$  (présence d'une planète) on fixe "une certaine intensité  $\boldsymbol{\alpha}_1$ ". Le BF est optimal au sens de Neyman-Pearson pour la vraisemblance intégrée  $p_1(\mathbf{x}) = \int d\boldsymbol{\alpha} p(\mathbf{x}|\boldsymbol{\alpha})\pi_1(\boldsymbol{\alpha})$  (proposition 6 p. 124 qui découle directement du lemme de Neyman-Pearson). Considérer la vraisemblance  $p_1(\mathbf{x})$  revient à s'intéresser aux propriétés de détection lorsque l'on observe<sup>5</sup> un grand nombre de fois (intégration sur  $\mathcal{X}$ ) une étoile accompagnée d'une planète (caractérisée par son intensité  $\boldsymbol{\alpha}_1 \in \mathbb{R}_{+*}^M$ ), et que l'on étend les observations sur un grand nombre d'étoiles accompagnées d'une planète (intégration sur  $\Theta_1 = \mathbb{R}_{+*}^M$ ). Cependant, et comme discuté dans le reste de la section de la page 124, on s'intéresse ici aux propriétés qui émergent lorsque l'on observe un très grand nombre de fois une étoile donnée, caractérisée par une intensité  $\boldsymbol{\alpha}_1$  qui constitue l'objet d'inférence des tests que nous étudions ici.

### Simulation d'un jeu de données selon la vraisemblance $p(\mathbf{x}|\boldsymbol{\alpha}^*)$

Sur un jeu de données sous  $H_0$  et un jeu de données sous  $H_1$  nous avons constaté qu'à la fois le PLR et le BF choisissent, et avec certitude, la bonne hypothèse. Afin de confirmer cette tendance sur  $\mathcal{X}$ , il est nécessaire d'appliquer ces tests à plus de données. On utilise la courbe ROC pour comparer et optimiser numériquement les performances des familles de tests bayésiens. PFA et PD sont estimées par des intégrations de Monte Carlo classiques à partir d'un grand nombre  $N_I$  de jeux de données  $\mathbf{x}^{[i]}, i \in \llbracket 1, N_I \rrbracket$  générés respectivement sous  $H_0$  ou sous  $H_1$  (pour une intensité  $\boldsymbol{\alpha}_1$  fixée comme discuté juste au dessus). Comme nous l'avons déjà dit, il est cependant impossible de générer un grand nombre de données grâce au simulateur présenté en partie 1 basé sur un modèle physique fin et des tables d'étalonnage d'aberrations optiques fixées.

<sup>5</sup>Le détail de *la façon dont est menée une expérience sur une étoile donnée* concerne la relation entre le choix de la vraisemblance, hyperparamètres compris, et le choix des conditions d'expérimentation (conditions atmosphériques par exemple) et de même le détail de *la façon dont sont choisies les étoiles à observer* concerne la relation entre le choix de l'*a priori*, hyperparamètres compris, et le choix de l'expérience (type d'étoile à observer par exemple).

A la place, nous générons un grand nombre  $N_I$  de jeux de données sous la vraisemblance marginale donnée par l'Eq. (11.14) et avec une seule dimension spatiale. L'application des tests bayésiens requérant pour chaque  $\mathbf{x}^{[i]}$ ,  $i \in \llbracket 1, N_I \rrbracket$  la génération d'une chaîne de Markov  $\{\boldsymbol{\alpha}^{[j]}, j \in \llbracket 1, N_J \rrbracket\}_{\mathbf{x}^{[i]}}$  de grande taille  $N_J$ , chaque jeu de données  $\mathbf{x}^{[i]}$  doit être de taille restreinte.

En pratique un jeu de données  $\mathbf{x}^{[i]}$  est simulé à partir de la vraisemblance marginalisée (11.14) pour  $K = 5$  poses,  $L = 8$  pixels (disposés en ligne),  $M = 2$  canaux spectraux et pour les données sous  $H_1$  une intensité sous-jacente de planète donnée par  $\boldsymbol{\alpha}_1 = (0.5, 0.5)^t$ . Les valeurs d'hyperparamètres sont choisies de telle sorte que la détection ne soit ni trop simple (courbe ROC confondue avec la fonction Heaviside sur  $[0, 1]$ ) ni trop difficile (courbe ROC confondue avec la fonction  $f(x) = x$  sur  $[0, 1]$ ). Elles sont choisies égales à :

- $\mathbf{d}_0 = 10 \mathbf{1}_{LM}$
- $\lambda_0 = 2$
- $\Sigma_0$  est une matrice de Toeplitz dont le premier vecteur colonne est donné par  $\mathbf{v} = 5\mathbf{u}$  avec  $u_1 = 1$ ,  $u_2 = 0.3$ ,  $u_3 = 0.05$ ,  $u_{L+1} = 0.6$ ,  $u_{L+1} = 0.2$ , toutes les autres composantes de  $\mathbf{u}$  étant nulles
- $\nu_0 = 2M$

La génération d'un jeu de données  $\mathbf{x}$  selon cette vraisemblance est effectuée en échantillonnant les lois normales et inverse-Wishart (Eq. 11.6 et 11.7) pour simuler la marginalisation effectuée analytiquement page 174 et en injectant chaque couple de variable  $\mathbf{d}$  et  $\Sigma$  ainsi généré dans la vraisemblance gaussienne de niveau 1 (Eq. 11.5) utilisée finalement pour générer aléatoirement un jeu de données  $\mathbf{x}$ .

On échantillonne la loi inverse-Wishart grâce aux propriétés suivantes. Soit  $A$  une matrice aléatoire de taille  $K \times L$  telle que les  $K$  lignes sont indépendantes et soit  $A_{k*}$  la  $k^{\text{è}}$  ligne de  $A$ .

$$\begin{aligned} \text{Si } A_{k*} &\sim \mathcal{N}_L(0, \Sigma_0) \\ \text{alors } A^t A &\sim \mathcal{W}_L(\Sigma_0, K) \end{aligned}$$

De plus, si  $S \sim \mathcal{W}_L(\Sigma_0, K)$  alors  $S^{-1} \sim \mathcal{W}_L^{-1}(\Sigma_0^{-1}, K)$ . Pour échantillonner la loi inverse Wishart il suffit donc d'échantillonner la loi normale et d'appliquer les constructions indiquées.

Un jeu de données sous  $H_1$  est obtenu à partir d'un jeu de données sous  $H_0$  en ajoutant la contribution de la planète, dont l'intensité est fixée à  $\boldsymbol{\alpha}_1 = (0.5, 0.5)^t$  comme on l'a dit, et dont les profils  $\{\mathbf{p}(k, r), k \in \llbracket 1, K \rrbracket\}$  sont donnés par le profil  $\mathbf{p} = (0.2, 0.5, 0.3)^t$  translaté à partir de la première pose avec une vitesse de 1 pixel par pose et une accélération de 0.02 pixels par pose au carré.

### Echantillonnage de la distribution *a posteriori* $\pi_1(\boldsymbol{\alpha}|\mathbf{x})$

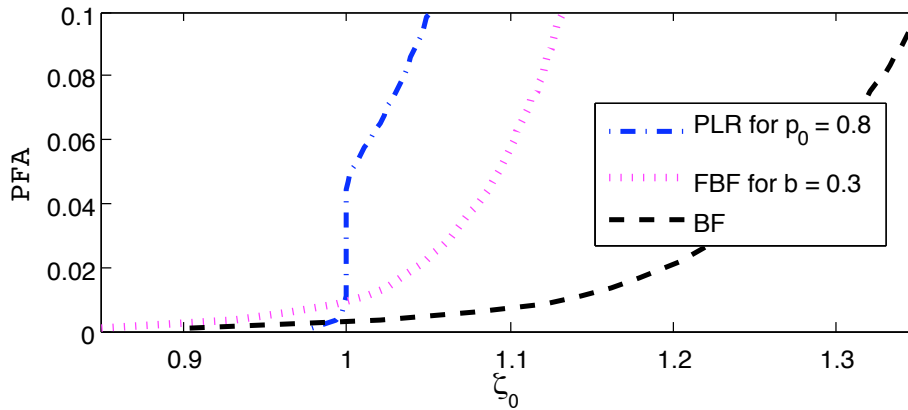
Pour appliquer les tests à un jeu de données  $\mathbf{x}^{[i]}$  ainsi simulé, il faut dans un premier temps définir les hyperparamètres nécessaires à la définition complète de la distribution *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\mathbf{x}^{[i]})$  de l'intensité d'une planète qui serait présente dans les données  $\mathbf{x}^{[i]}$ . Les hyperparamètres de la vraisemblance (11.14) sont évidemment choisis différents de ceux utilisés pour la génération des données mais choisis "raisonnables" :

- $\mathbf{d}_0 = 2 \mathbf{1}_{LM}$
- $\lambda_0 = 2$
- $\Sigma_0$  est prise proportionnelle à la matrice identité :  $\Sigma_0 = 10I_{LM}$
- $\nu_0 = 2M$

Les profils  $\mathbf{p}(k, r)$  sont supposés connus (et donc pris égaux aux profils décrits ci-dessus), y compris la position initiale  $r$  comme dans toute cette partie.

Les hyperparamètres de l'*a priori*  $\pi_1(\boldsymbol{\alpha})$  sont quant à eux les mêmes que ceux de la page 177 et l'*a priori* résultant était alors présenté sur la figure 11.2.

La distribution *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\mathbf{x}^{[i]})$  est finalement échantillonnée pour un jeu de données  $\mathbf{x}^{[i]}$  fixé par la même méthode MCMC qu'auparavant, décrite au début de la section 8.3 p. 141.

FIG. 12.9: PFA induite par différents tests paramétrés uniquement par le seuil  $\zeta_0$ .

### Relations entre les paramètres $(\zeta, p)$ et la PFA

L'échantillonnage de la distribution *a posteriori*  $\pi_1(\boldsymbol{\alpha}|\mathbf{x}^{[i]})$  est effectué pour deux grands ensembles de jeux de données :  $\{\mathbf{x}^{[i]}, i \in \llbracket 1, N_I \rrbracket\}_{H_0}$  et  $\{\mathbf{x}^{[i]}, i \in \llbracket 1, N_I \rrbracket\}_{H_1, \alpha_1}$ . Sous  $H_i$  tous les jeux de données sont générés de la même façon, et les statistiques de test  $\text{PLR}(\mathbf{x}^{[i]}, \zeta)$  et du facteur de Bayes fractionnaire (définition 5 p. 24)  $\text{FBF}(\mathbf{x}^{[i]}, b)$  sont évidemment calculées de la même façon que les données aient été générées sous  $H_0$  ou sous  $H_1$ .

Afin d'appliquer le test binaire consistant à seuiller  $\text{PLR}(\mathbf{x}, \zeta)$  par  $p$  et  $\text{FBF}(\mathbf{x}, b)$  en adoptant l'approche de Neyman, c'est-à-dire en fixant une PFA plutôt que  $\zeta$  et/ou  $p$ , un étalonnage de ces paramètres serait nécessaire pour traiter un jeu de données  $\mathbf{x}$  fixé. Considérant dans un premier temps l'ensemble de jeux de données générés sous  $H_0$ , en faisant varier les valeurs de  $\zeta$  et/ou  $p$  et en comptant le nombre de fausses alarmes (estimation de PFA) associés à ces tests, on peut observer ces étalonnages<sup>6</sup>.

La figure 12.9 indique la relation entre la valeur de seuil  $\zeta_0$  choisie pour différents tests et la PFA induite. Les tests considérés sont le test consistant à seuiller  $\text{PLR}(\mathbf{x}, \zeta)$  par  $p_0 = 0.8$ , le test consistant à seuiller  $\text{FBF}(\mathbf{x}, b)$  par  $\zeta$  pour  $b = 0.3$  et le test consistant à seuiller  $\text{BF}(\mathbf{x})$  (égal à  $\text{FBF}(\mathbf{x}, 0)$ ) par  $\zeta$ . La figure 12.9 montre ainsi la valeur de seuil  $\zeta_0$  à choisir d'après l'approche de Neyman-Pearson, pour la vraisemblance et l'*a priori* considérés, pour atteindre une PFA fixée dans les tests du PLR lorsque l'autre paramètre est fixé. Notons la discontinuité de la valeur de PFA associée au PLR autour de la valeur  $\zeta = 1$  pour un seuil  $p$  fixé. La figure 12.10 indique quant à elle la PFA associée au test consistant à seuiller  $\text{PLR}(\mathbf{x}, \zeta)$  par  $p$  pour  $\zeta = 1$ .

### Courbes ROC et paramétrisation optimale

Comme introduit dans la section 7.1 page 121 et précisé page 189, on ne considère par pour PD la moyenne de la probabilité de bonne détection sur tous les  $\alpha_1$  possibles et décrits par la loi *a priori* mais bien  $\text{PD}(\alpha_1)$  pour une valeur  $\alpha_1$  fixée. Comme nous le verrons la valeur de  $\alpha_1$  choisie mène à une détection ni trop simple ni trop difficile.

#### • Paramétrisation optimale du PLR

L'étude de la relation entre PD et les paramètres des tests pourrait aussi être effectuée grâce aux données générées sous  $H_1$ , et toutes ces études pour encore bien d'autres paramétrisations ( $(\zeta, p)$  pour le PLR par exemple) des tests. Comme ceci est expliqué dans la seconde partie de la section 8.3 p. 141, une optimisation, au sens de Neyman-Pearson, des deux familles de tests associées respectivement au PLR

<sup>6</sup>Noter que l'indice 0 des seuils n'est pas très utile ici mais est utile lorsque  $\text{PLR}(\mathbf{x}, \zeta)$  est tracé en fonction de  $\zeta$  par exemple.

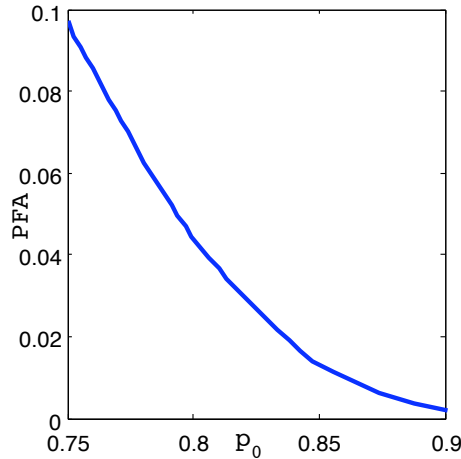


FIG. 12.10: PFA induite par le test consistant à seuiller  $\text{PLR}(\mathbf{x}, \zeta)$  par  $p_0$  pour  $\zeta = 1$ . L'intervalle de PFA considéré est le même que dans la figure précédente.

et au FBF est ainsi en principe possible puisque ces statistiques définissent des tests binaires paramétrés par deux paramètres de test. Pour le PLR les deux paramètres  $\zeta$  et  $p$  sont laissés libres ( $\lambda = (\zeta_0, p_0)$ ) et pour le FBF les deux paramètres  $b$  et  $\zeta$  sont laissés libres ( $\lambda = (b, \zeta_0)$ ). Pour la famille associée au PLR, pour chaque PFA fixée on détermine ainsi la paramétrisation  $\lambda(\text{PFA})$  qui maximise PD par rapport à toutes les autres paramétrisations possibles.

L'optimisation numérique proposée dans la section 8.3 pour le PLR et le FBF est ainsi appliquée aux deux ensembles de données générés sous  $H_0$  et sous  $H_1$  (ensemble de données caractérisé par l'intensité  $\alpha_1$ ). La figure 12.12 illustre le fait que la paramétrisation optimale du PLR est très proche de  $\{\zeta = 1, p \in [0.1, 0.4]\}$  sur une gamme de PFA classique.

La paramétrisation optimale au sens de Neyman-Pearson du FBF est indiquée sur la figure 12.12.

### • Courbes ROC associées au PLR, au FBF et au GLR construit dans la partie 1 de cette thèse

La figure 12.13 illustre maintenant les courbes ROC obtenues pour les tests consistant à seuiller la distribution *a posteriori* du rapport de vraisemblance  $\text{PLR}(\mathbf{x}, \zeta)$ , le facteur de Bayes fractionnaire  $\text{FBF}(\mathbf{x}, b)$  et le rapport de vraisemblance généralisé “ $\text{GLR}_2(\mathbf{x})$ ” construit à partir du modèle gaussien simple de la première partie de cette thèse et dont l'expression est donnée dans la sous-section de la page 73. Les courbes ROC associées aux paramétrisations optimales du PLR et du FBF discutées dans le paragraphe précédent sont aussi représentées.

On constate que :

- Le facteur de Bayes classique ( $\text{FBF}(\mathbf{x}, 0)$ ) est dépassé uniformément sur la gamme de PFA considérée par le PLR et le FBF paramétrés de façon standard (en fixant respectivement  $p = 0.8$ ,  $\zeta = 1$  (et  $b = 0.3$  dont l'interprétation n'est pas discutée ici)). Ceci implique en particulier que le PLR n'est pas inadmissible par rapport au BF, c'est-à-dire qu'il existe des valeurs  $\alpha_1$  telles que  $\text{PD}(\text{PLR}, \text{PFA}, \alpha_1) > \text{PD}(\text{BF}, \text{PFA}, \alpha_1)$ , et ce en l'occurrence pour toute une gamme de PFA. Pour  $\text{PFA} = 0.1$ , les performances du PLR dépassent celles du facteur de Bayes de 15%.
- Les tests à paramètres fixés (PLR pour  $p = 0.8$  ou  $\zeta = 1$  et FBF pour  $b = 0.3$ ) ont ici des performances très proches des performances optimales. Ceci permet de pouvoir fixer les paramètres  $\zeta$ ,  $p$  et/ou  $b$  (par l'interprétation qu'on leur attribue, plutôt que fixer une PFA) sans ce soucier de l'optimalité fréquentiste de ce choix puisque toutes les paramétrisations sont à peu près optimales.
- Pour la simulation effectuée pour obtenir ces courbes ROC, le test consistant à seuiller le  $\text{GLR}_2$  (statistique développée dans la première partie de cette thèse, où  $\Sigma$  était supposée à tort proportion-

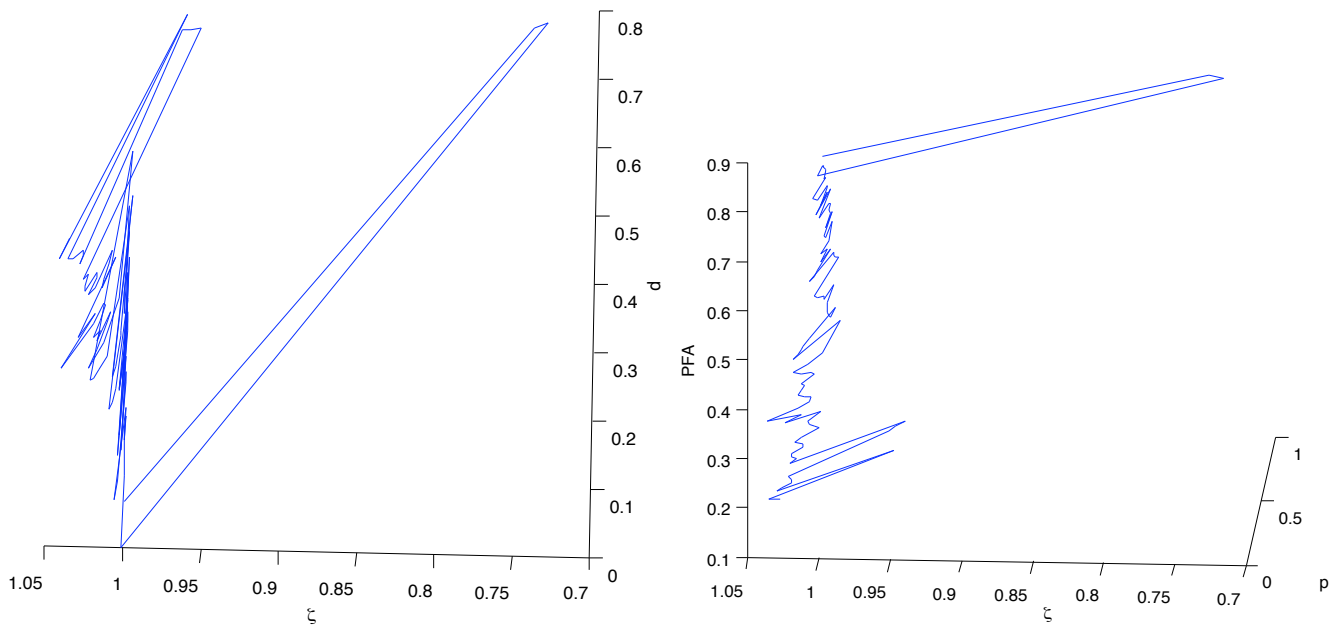


FIG. 12.11: Gauche : Paramétrisation  $\{(\zeta, p), \text{PFA} \in [0, 1]\}$  optimale au sens de Neyman-Pearson (PD maximisé pour PFA fixé) de la famille de tests  $\{\text{PLR}(\mathbf{x}, \zeta) > p, \zeta \in \mathbb{R}_+$  et  $p \in [0, 1]\}$ . Droite : La PFA résultant de la paramétrisation choisie est indiquée en fonction de la paramétrisation optimale. L'axe vertical indique la valeur de PFA associée à chaque couple  $(\zeta, p)$  optimal, l'axe horizontal indique  $\zeta$  et l'axe en profondeur indique le seuil  $p$ .

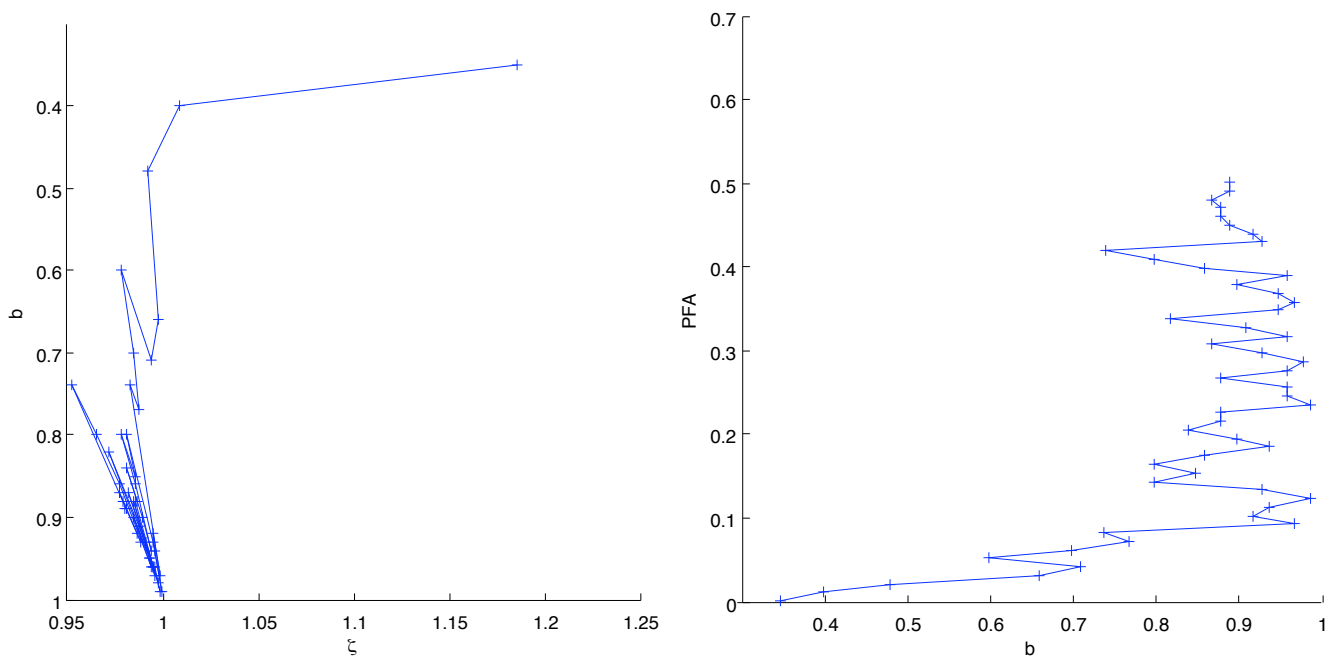


FIG. 12.12: Gauche : Paramétrisation  $\{(\zeta, p), \text{PFA} \in [0, 1]\}$  optimale au sens de Neyman-Pearson (PD maximisé pour PFA fixé) de la famille de tests  $\{\text{FBF}(\mathbf{x}, b) > \zeta, b \in [0, 1]$  et  $\zeta \in \mathbb{R}_+\}$ . Droite : La PFA résultant de la paramétrisation choisie est indiquée en fonction de la valeur de la valeur  $b$  optimale.



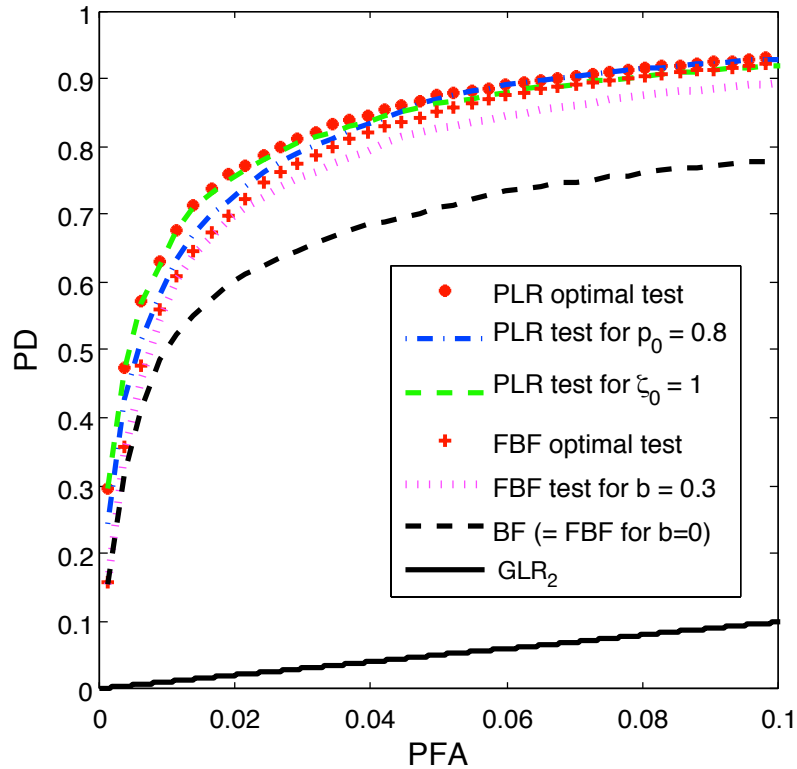


FIG. 12.13: Courbes ROC du PLR, du (F)BF et du GLR<sub>2</sub>. PD est calculée pour une valeur  $\alpha_1$  fixée.

nelle à l'identité) est extrêmement peu performant : ce test revient ici, pour PFA=0.5, à détecter ou non selon l'issue d'un jeu de *pile ou face*. Il faut cependant bien noter que ce GLR n'est pas calculé à partir de la vraisemblance marginale  $p(\mathbf{x}|\boldsymbol{\alpha})$  donnée par l'équation (11.14) p. 174 utilisée pour générer ici les données.

# Conclusion et perspectives

## Conclusion

La recherche de planètes extra-solaires est l'objectif ambitieux d'un grand nombre de méthodes et projets instrumentaux actuels. Parmi ceux-ci, des instruments au sol tels que le futur instrument SPHERE qui sera monté sur le VLT en 2011 vise à détecter une exoplanète grâce à la lumière qu'elle émet. Discerner la lumière de l'exoplanète de celle de l'étoile parente est particulièrement difficile car l'exoplanète a en infra-rouge (gamme de longueurs d'onde la plus favorable) une intensité lumineuse typiquement un million de fois plus faible que celle de l'étoile, tout en étant séparée de celle-ci d'un angle de  $0''.2$ , c'est-à-dire du même angle que deux points séparés de 1 millimètre et observés à 1 kilomètre de distance. De plus, des effets tels que les turbulences de l'atmosphère de la Terre, la diffraction de la lumière par le télescope et les fluctuations statistiques naturelles de l'intensité lumineuse mesurée en présence d'un front d'onde cohérent détériorent considérablement les images, noyant l'exoplanète dans la lumière émise par l'étoile. Malgré la performance des techniques d'optique adaptative et de coronagraphie existantes et en cours de développement, des techniques de traitement des images performantes et informatives sont nécessaires à la détection d'exoplanètes à partir d'images coronographiques acquises depuis le sol terrestre.

Cette thèse s'est ainsi attachée au traitement d'images acquises grâce au mode DBI du module d'imagerie IRDIS de l'instrument SPHERE. Ces images seront notamment caractérisées par la présence d'un signal (l'exoplanète) d'intensité très faible et se déplaçant de façon connue sur un fond (la lumière résiduelle de l'étoile) presque, mais pas suffisamment, constant dans le temps et dans les deux canaux spectraux d'acquisition. Pour détecter et estimer l'intensité et la position d'une exoplanète à partir d'un ensemble d'images acquises simultanément dans une, deux ou plusieurs bandes spectrales, nous avons développé des modèles statistiques de ces images afin de construire des inférences applicables ou de calculer précisément des incertitudes, développé un programme informatique permettant d'appliquer de façon réaliste ces inférences à un jeu de données, et contribué à l'étude théorique d'une méthode générale et informative bayésienne de test d'hypothèses, utile pour le problème de détection notamment.

Les quatre parties de cette thèse nous permettent de dresser les conclusions suivantes :

- Il émerge de l'ensemble de la thèse que pour appliquer une méthode de détection-estimation performante à l'ensemble d'images  $\mathbf{x}$  traité, il est nécessaire de procéder à une étape d'initialisation consistant à partir des données  $\mathbf{x}$  à identifier dans l'environnement proche de l'étoile les zones susceptibles d'abriter une planète. Des méthodes plus performantes et donc plus coûteuses en temps de calcul pourront alors être appliquées à ces candidats.
- La méthode mathématique sous-jacente à l'étape d'initialisation de détection-estimation est développée dans la partie I. Elle repose sur l'hypothèse de stationnarité temporelle du fond et sur un modèle des données gaussien non-corrélé spatialement. L'étape d'initialisation est implantée dans le programme informatique *user-friendly* MOODS, utilisable en principe pour des données réelles telles que celles qui seront acquises avec SPHERE et auxquelles des prétraitements standards auront été appliquées. MOODS a alors été appliqué à des données simulées selon un modèle physique pas-à-pas fin dédié à l'instrument SPHERE. Les résultats sont qualitativement comparables à ceux des méthodes classiques actuelles dédiées au traitement de telles données.

- Cependant, un dispositif d'imagerie employant une optique adaptative extrême et un coronographe, tel que c'est le cas pour l'instrument SPHERE, est caractérisé par l'existence d'un front d'onde cohérent, suffisamment échantillonné par la caméra d'acquisition pour voir apparaître des *speckles* (petites taches de même profil que la planète) et des *speckles* pinnés (*speckles* dont l'intensité est réhaussée lorsqu'ils sont situés sur une zone de forte amplitude complexe moyenne). Ainsi, pour des poses courtes, plus informatives et donc préférables à des poses longues si elles sont correctement modélisées, et sous des conditions atmosphériques stables, les images sont plus rigoureusement modélisées par un modèle statistique dit de *speckle* développé en amont de cette thèse et bien plus complexe que le modèle gaussien utilisé précédemment. Nous avons ainsi estimé dans la partie II à partir d'une méthode d'estimation, ici très performante, de probabilité d'erreur appliquée au modèle détaillé des *speckles*, que la probabilité d'erreur telle que calculée à partir du modèle gaussien simple induit typiquement 1000 fois plus de fausses alarmes (planètes détectées à tort) que ce ne serait le cas en réalité. Utiliser le modèle *speckle* pour calculer cette probabilité d'erreur en pratique serait donc important, mais l'utilisation du modèle *speckle* suppose l'estimation du front d'onde moyen dans le plan de la caméra d'acquisition ainsi que l'estimation de la matrice de covariance spatiale dans ce même plan, estimations qui n'ont pas été réalisées ici.
- Par *probabilité d'erreur* nous entendions précédemment la “probabilité” que l'on constaterait en appliquant la méthode de détection sur l'ensemble des jeux de données possibles, où ces jeux de données décriraient une loi de probabilité (la vraisemblance) faisant tous intervenir les mêmes paramètres  $\theta$  sous-jacents (probabilité fréquentiste). Cependant, ces paramètres sous-jacents, l'intensité de la planète en particulier, sont en général inconnus. Dans le paradigme bayésien, il est possible d'utiliser une description des valeurs que les paramètres sous-jacents  $\theta$  peuvent prendre par une distribution de probabilité dite *a priori* de ces paramètres. Dans ce cadre, il est aussi possible de définir une certaine *probabilité d'erreur* comme la “probabilité” sur  $\theta$  que l'on constaterait en appliquant une méthode de détection supposant  $\theta$  connu. La probabilité est alors calculée non sur les valeurs que pourrait prendre  $x$  mais sur les valeurs que pourrait prendre  $\theta$  (probabilité bayésienne). Dans notre cas, cela signifie par exemple calculer une probabilité en *comptant* sur l'ensemble des planètes sur lesquelles on aurait pu pointer le télescope. Ce calcul peut en particulier être effectué *a posteriori*, c'est-à-dire conditionnellement à la seule variable effectivement connue, à savoir le jeu de données observé  $x$ . En particulier, nous avons étudié dans la partie III les propriétés générales d'un outil de test bayésien simple et original. Celles-ci s'avèrent utiles en pratique et/ou assez significatives. Il permet par ailleurs, et contrairement aux outils classiques bayésiens de test d'hypothèses, d'établir dans certains cas une équivalence entre probabilité sur l'espace des observations (probabilité fréquentiste) et l'espace des paramètres sous-jacents (probabilité bayésienne).
- Par ailleurs, le paradigme bayésien nous a permis de construire dans la partie IV un modèle des données plus flexible que le modèle gaussien initial et plus facilement applicable que le modèle dit de *speckle*. Associé à une distribution *a priori* de l'intensité de la planète dans les différents canaux spectraux du même type que celle observée en pratique pour la distribution des petits corps (astéroïdes ...), ce modèle s'avère très adapté aux quelques jeux de données simulés finement sur lesquels la méthode de détection a été appliquée. En effet, en utilisant la méthode de détection discutée précédemment ou la méthode bayésienne standard de détection, la détection d'une planète séparée de  $0''.2$  de l'étoile et d'intensité un million de fois plus faible que l'étoile est *évidente* en utilisant 20 poses, et réciproquement la méthode de détection indique avec une très forte probabilité qu'il n'y a pas de planète dans un jeu de données effectivement simulé sans planète. Les performances d'une détection par la méthode discutée dans la partie III et appliquée à ce modèle bayésien semblent ainsi très supérieures à celles de la méthode de détection construite à partir du modèle gaussien. Cependant, des études de la *robustesse* de ces résultats à des changements des valeurs d'hyperparamètres et à l'application de la méthode de détection à davantage de jeux de données simulés finement sont encore nécessaires.

## Perspectives

Les perspectives de cette thèse émergent naturellement des limites apparaissant dans la conclusion.

- Nous avons vu qu'utiliser des poses courtes plutôt que des poses longues permet en principe d'atteindre une performance équivalente pour un temps d'observation total plus faible. Ceci est vrai d'une part car une pose longue est quasiment équivalente, en terme statistique, à une somme de poses courtes, et d'autre part car pour une durée d'observation totale plus courte la condition de stationnarité du fond (*i.e.* des conditions atmosphériques) est mieux vérifiée que pour une durée d'observation totale plus longue. Cependant, ce gain de performance n'est vrai qu'à condition que les images de poses courtes soient correctement modélisées, c'est-à-dire selon le modèle *speckle* utilisé dans la partie II et supposant la stationnarité temporelle du fond dans le jeu d'images. Or afin d'utiliser le modèle *speckle*, il est nécessaire d'une part d'estimer la moyenne et la matrice de covariance spatiale du front d'onde dans le plan de la caméra et d'autre part de classer l'ensemble des images obtenues lors d'une nuit d'observation en des séries d'images vérifiant chacune bien l'hypothèse de stationnarité.

Pour le problème d'estimation des moments du front d'onde, notons d'une part qu'un dispositif de diversité de phase, qui mesure l'intensité dans un plan éloigné du plan focal d'une distance connue, est déjà régulièrement utilisé en astronomie pour estimer les aberrations dans le plan focal (Mugnier et al. [2009]) et pourrait peut-être être suffisant pour l'estimation de la matrice de covariance, et d'autre part l'estimation d'une matrice de covariance complexe est un problème régulièrement rencontré dans l'étude des radars.

Concernant le problème de classification des images en des séries stationnaires, nous avons constaté dans le chapitre 2 et l'annexe associée que les deux premiers moments empiriques de la distribution des données simulées finement sont en général très nettement corrélés soit –dans la zone centrale du champ– avec l'intensité moyenne pondérée du vent dans les couches de l'atmosphère de la Terre, soit –dans la zone à la limite de la correction par optique adaptative et la zone non corrigée– avec le *seeing*. Dans la mesure où le vent ne change pas de direction pendant la nuit, il suffirait donc *a priori* et si ces constatations sont renforcées par d'autres études du même type, de mesurer ces grandeurs physiques caractérisant l'état des turbulences de l'atmosphère de la Terre afin de classer les données en des séries relativement stationnaires sans avoir besoin d'utiliser les données à traiter elles-mêmes. Le *seeing* est systématiquement mesuré en temps réel sur un site d'observation astronomique, et la mesure de l'intensité des vents dans les couches les plus turbulentes pourrait être envisageable à la fréquence temporelle nécessaire si des instruments de mesure, tels que SCIDAR, sont utilisés à cet effet.

- Le modèle bayésien développé dans la partie IV est applicable et s'est avéré très performant sur quelques jeux de données simulés finement et sur des simulations moins fidèles des futures données à attendre de l'instrument SPHERE. Cependant, un plus grand nombre d'applications de la méthode, en particulier sur les données simulées finement, sont nécessaires afin de tester la robustesse des résultats vis-à-vis d'un changement des valeurs des hyperparamètres définissant la distribution probabiliste des données et la distribution probabiliste de l'intensité de la planète et vis-à-vis du jeu de données considéré. En particulier, ici les valeurs d'hyperparamètres ont été choisies *très simplement* mais il est assez clair qu'en principe leurs valeurs devraient être prises différentes selon la stationnarité et l'état des conditions atmosphériques caractérisant le jeu de données. L'objectif initial d'études ultérieures pourrait être d'étudier d'autres jeux de données mais en conservant les valeurs d'hyperparamètres *simples* utilisées dans cette thèse afin de voir s'il est nécessaire, et si oui comment, de classer les images en séries stationnaires.

Pour cela, et quoiqu'il en soit, il est nécessaire d'améliorer la programmation actuelle de la méthode car elle est en l'état inapplicable à de gros jeux de données et/ou à l'étude de robustesse envisagée dans le paragraphe précédent. Trois éléments sont particulièrement limitatifs actuellement : l'étude séquentielle des différentes positions initiales possibles de la planète, le calcul du déterminant de la vraisemblance donnée dans l'équation (11.14) p. 174 et l'échantillonnage par la méthode de

*slice sampling* effectué actuellement par une fonction programmée pour le logiciel Matlab. La parallélisation du traitement des différentes positions initiales sur le *cluster* de la Côte d’Azur est en cours (voir Mattei et al. [2010]) mais les deux autres problèmes doivent encore être résolus.

En ce qui concerne l’application de la méthode développée dans la partie III, il s’avère ici d’après des données simulées simplement que les paramétrisations standards (de même que fixer un niveau de probabilité d’erreur à 5% est standard) sont toutes proches de la paramétrisation optimale, en terme fréquentiste, de cette famille de tests. Ceci est vrai en particulier de la paramétrisation  $\zeta = 1$  qui semble par ailleurs jouer un rôle particulier. Si cette constatation se confirme analytiquement et/ou par simulation, fixer le paramètre de seuil  $\zeta$  à 1 et lire et seuiller la valeur de la statistique de test induite peut constituer une bonne procédure de détection, analogue à la procédure effectuée avec une *p-value* fréquentiste, mais ici en version bayésienne.

- Plus généralement, la méthode de test d’hypothèses développée dans la partie III peut continuer à être étudiée de façon théorique sous de nombreux angles ainsi que d’un point de vue plus fondamental. Fondamentalement, une première étape pourrait consister à dériver des propriétés d’optimalité liées à cette méthode, par une approche bayésienne du lemme de Neyman-Pearson en particulier. Une seconde étape pourrait alors consister à étendre la méthode pour le test d’une hypothèse composite *vs* une hypothèse composite ainsi que les propriétés d’optimalité associées<sup>7</sup>. Une troisième étape pourrait consister à se réinterroger sur la pertinence de comparer une hypothèse à une autre plutôt que tester une hypothèse sans utiliser d’alternative, comme le fait une *p-value* fréquentiste. Par ailleurs, l’effet de la valeur spécifique  $\zeta = 1$  mériterait d’être approfondi. Enfin, les études d’une méthode élargissant celle étudiée dans cette thèse (méthode définie pour un test d’une hypothèse simple *vs* une hypothèse composite) requièreraient d’être comparées aux études menées sur le problème de sélection de modèle, déjà très discuté.

---

<sup>7</sup>Ces deux premières étapes ont en fait déjà été réalisées mais trop tard pour être incluses dans cette thèse.

# Publications et communications

## Publications de revue

I. Smith, A. Ferrari et M. Carbillet. Detection of a moving source in speckle noise. Application to exoplanet detection. Dans *IEEE Transactions on Signal Processing*, 2009.

A. Vigan, C. Moutou, M. Langlois, F. Allard, A. Boccaletti, M. Carbillet, D. Mouillet et I. Smith. Photometric characterization of exoplanets using angular and spectral differential imaging. Dans *Monthly Notices of the Royal Astronomical Society*, 2010.

## Publications de conférence internationale à comité de lecture

I. Smith et A. Ferrari. Detection and performance analysis for a moving point source in speckle noise, application to exoplanet detection by direct imaging. Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) Proceedings*, 2008.

I. Smith et A. Ferrari. Detection from a multi-channel sensor using a hierarchical Bayesian model. Dans *ICASSP Proceedings*, 2009.

A. Ferrari, C. Richard, I. Smith et C. Theys. Transductive kernel matrix learning with hierarchic Bayesian model, application to hyperspectral images. Dans *IEEE International Geoscience & Remote Sensing Symposium (IGARSS) Proceedings*, 2010.

I. Smith et A. Ferrari. Bayesian detection with the posterior distribution of the likelihood ratio. Dans *European Signal Processing Conference (EUSIPCO) Proceedings*, 2010.

## Publications de conférence nationale à comité de lecture

I. Smith et A. Ferrari. Détection d'exoplanètes par imagerie directe, une approche bayésienne. Dans *Groupe d'Etudes du Traitement du Signal et des Images (GRETSI)*, 2009.

## Publications de conférence internationale sans comité de lecture

M. Carbillet, A. Boccaletti, C. Thalmann, T. Fusco, A. Vigan, I. Smith, D. Mouillet, K. Dohlen, P. Bendjoya et A. Ferrari. The Software Package SPHERE : a numerical tool for end-to-end simulations of the VLT instrument SPHERE. Dans *Adaptive Optics System, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Proceedings*, 2008.

I. Smith, M. Carbillet, A. Ferrari, D. Mouillet, A. Boccaletti et K. Dohlen. Simulation of moving exoplanets detection using the VLT instrument SPHERE. Dans *Adaptive Optics System, SPIE Proceedings*, 2008.

D. Mattei, I. Smith, A. Ferrari et M. Carbillet. Parallelization of exoplanets detection algorithms based on field rotation ; example of the MOODS algorithm for SPHERE. Dans *In the spirit of Lyot*, 2010.

A. Vigan, C. Moutou, M. Langlois, D. Mouillet, K. Dohlen, A. Boccaletti, M. Carbillet, I. Smith, A. Ferrari, L. Mugnier et C. Thalmann. Comparison of methods for detection and characterization of exoplanets with SPHERE/IRDIS. Dans *SPIE Proceedings*, 2010.

I. Smith et A. Ferrari. The posterior distribution of the likelihood ratio as a measure of evidence. Dans *Maxent Proceedings*, 2010.

## Publications de conférence nationale sans comité de lecture

I. Smith, A. Ferrari, M. Carbillet, A. Boccaletti, T. Fusco, K. Dohlen, D. Mouillet et M. Langlois. An exoplanet detection algorithm for field-rotated coronagraphic images, and preliminary results for SPHERE. Dans *Journées SF2A*, 2007.

## Rapports techniques

N. Vedrenne, I. Smith, A. Cornia, L. Mugnier et A. Ferrari. Specification of signal extraction for IRDIS DBI. Dans *VLT reports*, 2008.

I. Smith, F. Delarue et A. Ferrari. Posterior distribution of the LR at the GLR. Rapport technique, 2009.

## Cinquième partie

### Annexes





# Annexe A

## Algorithme MOODS

### A.1 Description textuelle de l'objectif scientifique de l'algorithme

La méthode mathématique utilisée pour la phase d'estimation de MOODS est présentée dans le chapitre précédent, et plus brièvement dans Smith et al. [2008] et Smith et al. [2009]. Rappelons que celle-ci est très basique et repose principalement sur

- un modèle des données gaussien indépendant temporellement et spatialement
- l'hypothèse que le fond (lumière due à l'étoile) est spatialement non-uniforme mais invariant dans le temps
- le fait qu'une planète se déplace sur ce "fond" à cause de la rotation de champ due à la monture alt-azimuthale du VLT, de sorte que son mouvement est connu à la position initiale  $r$  près
- l'hypothèse que la planète a un profil connu à un coefficient près : l'intensité de la planète  $\alpha$

Avant d'appliquer MOODS, les images brutes doivent être prétraitées afin de corriger des effets de courants d'obscurité, de "*flat field*", de décentrage de l'étoile. Elles peuvent aussi éventuellement être prétraitées de façon à injecter dans MOODS des données qui vérifient mieux l'hypothèse de constance dans le temps du fond.

L'algorithme MOODS est divisé en deux parties successives :

1. Estimation de l'intensité de l'exoplanète pour un ensemble de positions initiales possibles
2. Détection éventuelle d'exoplanètes et estimation finale de leurs positions, intensités et incertitudes sur les estimations d'intensité

Pour l'étape 1, tous les inputs que l'observateur doit entrer sont inclus dans un fichier de paramètres, appelé `call_estim.m`. C'est aussi la seule fonction qui a besoin d'être exécutée par l'observateur. Les paramètres d'entrée sont tous listés ci-dessous. En principe, il suffirait d'appliquer le calcul de  $\hat{\alpha}(r)$  pour chacune des positions initiales une par une (positions à définir à partir de paramètre plus *user-friendly* tels que la zone d'intérêt *etc*), mais en pratique la simulation des  $\mathbf{p}(k, r)$  est extrêmement long et des optimisations de temps de calcul étaient nécessaires.

Pour l'étape 2, la même structure est utilisée : le seul fichier `call_detec.m` intègre les paramètres d'entrée demandés à l'observateur et est le seul fichier que l'observateur doit exécuter. Cependant, à cause du manque intrinsèque de robustesse de l'étape de détection vis-à-vis de la distribution statistique précise des données et donc vis-à-vis des données  $\mathbf{x}$  traitées en pratique, l'étape 2 requiert des aller-retours entre le programme et l'observateur au moment de l'exécution du programme. Ainsi, des graphes sont montrés successivement à l'observateur et lui laisse choisir -selon ce qu'il observe- d'adopter telle ou telle stratégie pour continuer à dérouler la procédure de détection.

## A.2 Motivations du choix de programmation, description plus détaillée et diagrammes d'activité

Voici une rapide description textuelle de la programmation actuelle de MOODS :

1. Estimation initiale d'intensité : Tester chaque pixel -ou chaque position subpixelique d'un certain ensemble- est très gourmand en temps de calcul. Afin de réduire le temps nécessaire à l'algorithme, autant de positions initiales que possibles devraient être testées "en même temps". A cette fin, utilisant le temps total d'observation, la taille d'une PSF (préalablement masquée par MOODS) et les *inner* et *outer working angles* (qui délimite l'anneau autour de l'étoile où l'on veut chercher une planète), une grille polaire est créée, où une case (appelée *boîte*) de la grille correspond à une position initiale testée. Une PSF est translatée sur chacune de ces positions initiales. Cette étape revient à une convolution de la PSF masquée avec une image binaire en 0 et 1 avec 1 sur chaque position initiale testée. Ensuite, la rotation de champ, connue de façon déterministe, voir la section 4.1, est appliquée sur cette image pour simuler une pose, de telle sorte que la PSF instantanée (*pois*) est transformée en un petit *haricot*. Pour chaque pose, cette image est utilisée avec les données afin de démarrer le calcul d'estimation de  $\hat{\alpha}(r)$  pour chacun des  $r$  de la grille. Boîte par boîte, l'intensité et l'incertitude sont finalement estimée une fois que toutes les poses  $k$  ont été traitées. Cette phase finale d'estimation est réitérée pour toutes les boîtes (toutes les positions initiales  $r$ ) de la grille. Finalement, le processus complet recommence pour toutes les grilles requises pour le *scan* de toutes les positions initiales à étudier.
2. Détection et estimation finale d'intensité. La relation entre le seuil et une réelle PFA repose sur une modélisation statistique précise des données : contrairement à l'étape d'estimation, l'étape de détection est très sensible à l'aile de la distribution de la statistique à seuiller. Elle requiert donc davantage d'étude et de temps de calcul que ceux permis avec le modèle gaussien. Cependant, il est déjà important de visualiser et extraire des candidats potentiels, traités par la suite par un algorithme plus performant, non inclus dans MOODS. L'observateur a ainsi la possibilité de choisir entre deux types de seuil : un seuil théorique reposant directement sur la variance de l'estimation calculée dans l'étape 1 et un seuil empirique reposant sur les estimations d'intensité dans un anneau donné considérées comme non affectées par la présence d'une planète. Une fois que la détection est faite, les estimées d'intensité et d'incertitude sont corrigées de l'utilisation d'un masque sur la PSF dans l'étape 1. Finalement, de nouvelles images "nettoyées" sont créées en supprimant la contribution des planètes détectées des images initiales, comme dans l'approche CLEAN développée initialement par Högbom [1974] et largement répandue depuis. Ce jeu de données "nettoyées" peut être retraiter depuis l'étape 1 afin de détecter d'éventuelles autres planètes à une même séparation de l'étoile.

Les figures A.1 et A.2 décrivent schématiquement la structure de l'algorithme MOODS et les relations entre les différentes fonctions. Chacune de ces fonctions est décrite plus en détail dans le document joint en Annexe.

Les noms entre crochets en bas des boîtes jaunes et des cadres indiquent des noms de fonctions (extension .m implicite). Chaque boîte jaune indique la tâche accomplie par la fonction inscrite en bas de la boîte et chaque flèche représente l'information sous-jacente à la méthode statistique. Il s'agit d'un diagramme d'aide à la compréhension du principe de l'algorithme, pas un diagramme de "flux" car les flèches ne représentent pas les paramètres d'inputs et d'outputs programmés en pratique. Il comment les différentes grandes étapes sont réparties entre les différentes fonctions et comment elles s'articulent dans le principe.

Toutes les fonctions utilisées dans l'algorithme MOODS sont présentes dans les diagrammes et listées à nouveau ici :

- Etape 1 : `call_estim`, `estim_main`, `estim_psf`, `boxes_and_grids`, `load_data`, `load_ref`, `create_boxes` et `all_estim`

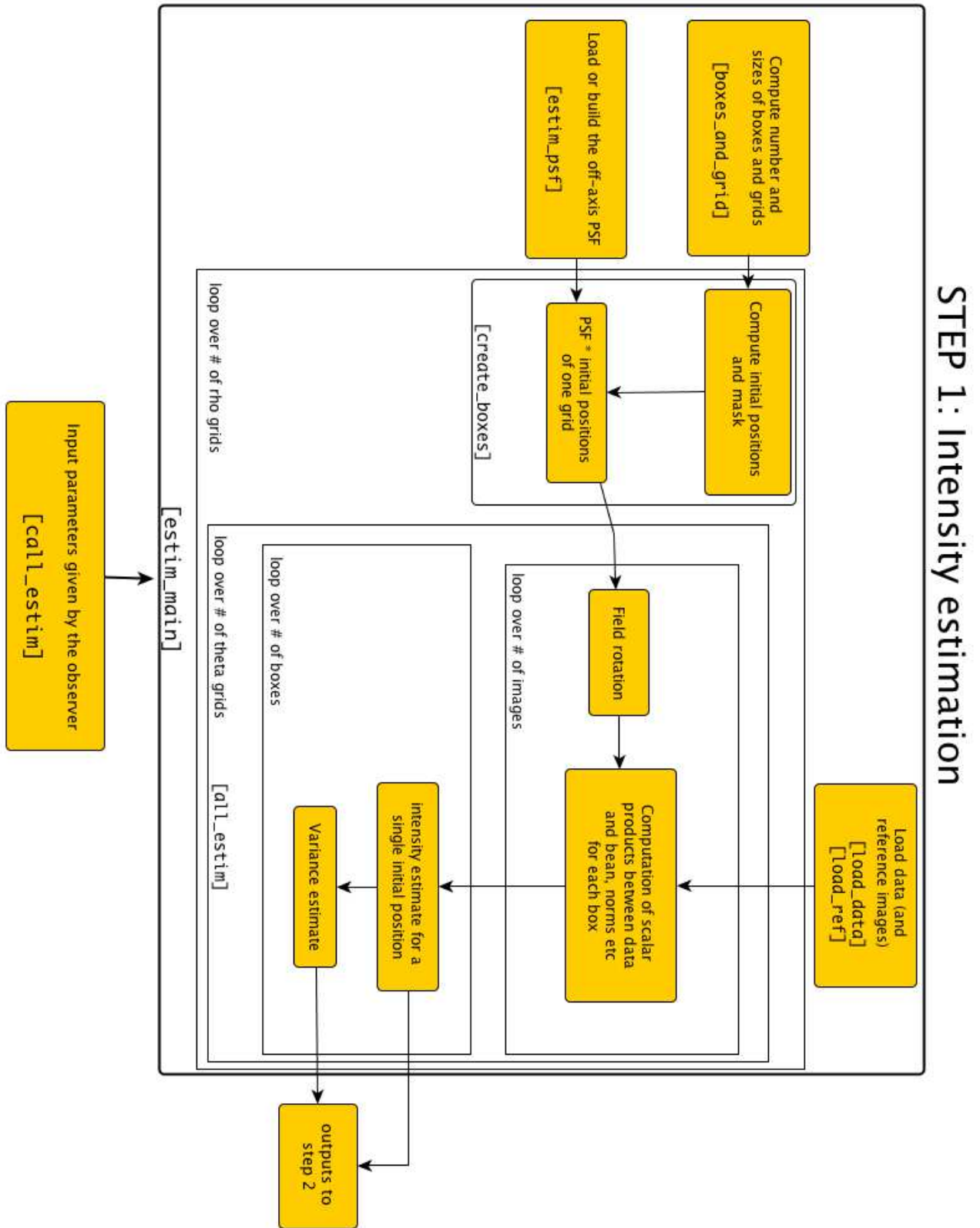


FIG. A.1: Diagramme d'activité de l'étape initiale d'estimation de MOODS.

## STEP 2: Detection

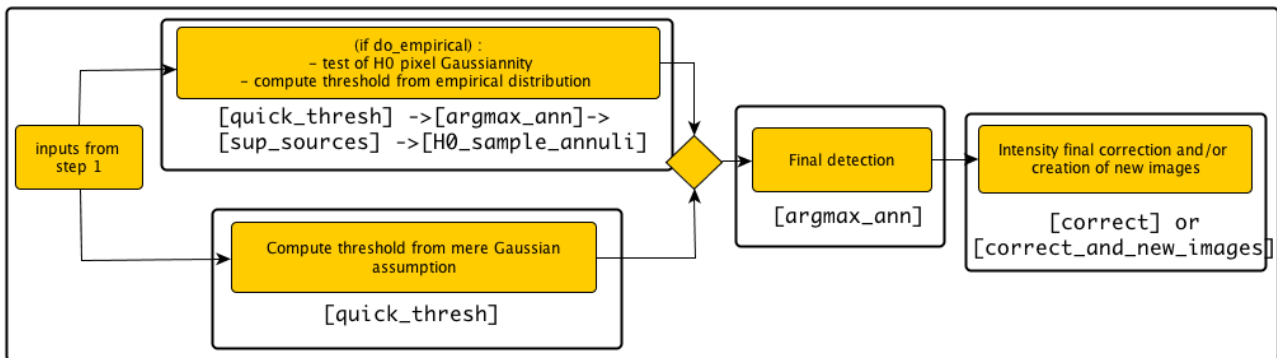


FIG. A.2: Diagramme d'activité de l'étape de détection et d'estimation finale de MOODS.

- Etape 2 : call\_detec, detec\_main, quick\_thresh, argmax\_ann, sup\_sources, H0\_sample\_annuli, correct et correct\_and\_new\_images

Ces fonctions sont rapidement décrites dans le document mis en annexe. Sont indiqués pour chaque fonction :

- la liste des inputs
- la liste des outputs
- les prérequis “physiques” à l'utilisation de la fonction (ex : “la PSF doit avoir été préalablement estimée et stockée dans un fichier” ...)
- un pseudo-code *synthétisant* ce qui est fait dans la fonction

Elles ont été initialement programmées en Matlab et sont disponibles sur demande. D'autres sous-routines sont appelées par cet ensemble de fonctions principales. Celles-ci sont seulement rapidement décrites à la fin du document mis en annexe, soit parce qu'elles sont usuelles (et peuvent même facilement être remplacées par de meilleurs programmes) soit parce qu'elles sont courtes (moins de 30 lignes).

### A.3 Listes des *inputs* demandés à l'utilisateur

On précise pour chaque input dans quelle fonction il est utilisé pour la première fois : cela permet de donner une idée de l'utilisation de cet input.

#### Etape 1 : Estimation

L'ensemble des inputs demandés à l'utilisateur pour l'étape d'estimation sont des variables à définir dans la seule fonction call\_estim.

- Inputs utilisés pour la première fois dans estim\_psf.m
  - do\_estim\_psf : exécute estim\_psf (do\_estim\_psf=1) ou recharge le résultat de la dernière exécution de estim\_psf (do\_estim\_psf=0)  
Par défaut : do\_estim\_psf = 1
  - how\_estim\_psf : indique comment la PSF (réponse de la planète) est estimée
    - si how\_estim\_psf = 'creation' elle est produite à partir d'une formule analytique
    - si how\_estim\_psf = 'chargeSP' elle est chargée à partir d'images d'étalonnage (“SP” signifie “SPHERE”)
  - Par défaut : how\_estim\_psf = 'chargeSP' (mais dépend des données)

- `dir_data` : adresse du répertoire qui contient l'ensemble des données à traiter et éventuellement l'estimée de la PSF  
Par défaut : `'.././../rot+phot_2008/jeu_070608.T2/'` (mais dépend des données)
  - `f` : numéro du filtre IRDIS utilisé ( $f=1$  or  $f=2$ )  
Par défaut :  $f = 1$  (mais dépend des données)
  - `mask_mode` : un masque binaire est appliqué sur la PSF pour réduire sa taille. La nouvelle PSF est appelée *pois*.
    - si `mask_mode = 'intensity'` le critère de la taille du masque est la fraction d'intensité de la PSF conservée dans le masque
    - si `mask_mode = 'radiusmax'` le critère de la taille du masque est directement le rayon souhaité pour le masque
 Par défaut : `mask_mode = 'intensity'`
  - `val` :
    - si `mask_mode = 'intensity'` `val` = fraction d'intensité de la PSF conservée dans le masque
    - si `mask_mode = 'radiusmax'` `val` = rayon du masque
 Par défaut : `val = 0.8`
  - `dr` :
    - si `mask_mode = 'intensity'` `dr` = pas radial pour le calcul du masque correspondant à l'intensité `val` (en pixels)
    - si `mask_mode = 'radiusmax'` `dr` n'est pas utilisé
 Par défaut : `dr = 0.5`
  - `numerical_psf_norm` : coefficient numérique de normalisation, choisi de telle sorte que la PSF prenne des valeurs commensurables avec les données  
Par défaut : `numerical_psf_norm = 100` (inutile en général)
2. Inputs utilisés pour la première fois dans `boxes_and_grids.m`
- `do_boxes_and_grids` : exécute `boxes_and_grids` ou recharge le résultat de la dernière exécution de `boxes_and_grids` (comme `do_estim_psf`, le premier input listé)  
Par défaut : `do_boxes_and_grids = 1`
  - `dt` : pas de temps utilisé pour approximer l'angle total de rotation de champ pour le calcul de la taille orthoradiale d'une boîte (en secondes)  
Par défaut : `dt = 0.1`
  - `tintegration` : durée total d'observation total time integration (en secondes)  
Par défaut : `tintegration = 14400` (mais dépend des données)
  - `dec` : déclinaison de l'étoile (en degrés)  
Par défaut : `dec = -45` (mais dépend des données)
  - `hahr_init` : angle horaire initial de l'étoile (lorsque la première pose démarre, en heures)  
Par défaut : `hahr_init=-2.` (mais dépend des données)
  - `size_pix` : taille d'un pixel en secondes d'arc sur le ciel  
Par défaut : `size_pix = 0.0122`
  - `[rho0_input_arcs,theta_test]` : position de la position initiale de référence testée [arcsec, rad]. `rho0_input` définit la séparation minimale testée, *i.e.* l'*inner working angle*.  
Par défaut : `[rho0_input_arcs,theta_test] = [0.2,  $\pi/4$ ]` (le second paramètre n'est pas important en général)
  - `rho_max_arcs` : *outer working angle* (en secondes d'arc)  
Par défaut : `rho_max_arcs = 1` (peut être étendu à 2 en général)
  - `opt_grid` :
    - `opt_grid = 'Ngrid'` si le nombre de grilles est directement défini par l'utilisateur
    - `opt_grid = 'resol'` si le nombre de grilles est défini à partir du nombre de pixels séparant deux positions initiales testées consécutives
 Par défaut : `opt_grid = 'Ngrid'`

- val\_rho :
    - si opt\_grid = 'resol' val\_rho = nombre de pixels souhaités le long de l'axe radial entre deux positions initiales testées consécutives
    - si opt\_grid = 'Ngrid' val\_rho = Nrho\_grid souhaité
    - Par défaut : val\_rho = 3
  - val\_theta :
    - si opt\_grid = 'resol' val\_theta = nombre MAXIMUM de pixels souhaités le long de la coordonnée orthoradiale entre deux positions initiales testées consécutives (maximum car ce nombre varie avec rho et on choisit de considérer l'*outer working angle* pour définir val\_theta)
    - si opt\_grid = 'Ngrid' val\_theta = Ntheta\_grid souhaité
    - Par défaut : val\_theta = 5
  - r\_secur : nombre de pixels ajoutés à la taille de la boîte pour éviter des erreurs numériques
  - Par défaut : r\_secur = 1
3. Inputs utilisés pour la première fois dans load\_data.m et load\_ref.m
- do\_load\_data : exécute load\_data ou non (comme do\_estim\_psf, do\_boxes\_and\_grids ...)
  - Par défaut : do\_load\_data = 1
  - Nexpos : nombre de poses  $K$
  - Par défaut : Nexpos = 144 (mais dépend des données)
  - suppr :
    - si suppr = 'suppr' la taille des images est réduite et conserve un champ carré s'arrêtant à rho\_max
    - si suppr = 'nosup' la taille des images n'est pas réduite
    - Par défaut : suppr = 'suppr'
  - utilisation\_ref = 0 si aucune image de référence n'est utilisée ; utilisation\_ref = 1 sinon
  - Par défaut : utilisation\_ref = 0
  - do\_load\_ref : exécute load\_ref (do\_load\_ref=1) ou non (comme do\_estim\_psf, do\_boxes\_and\_grids ...)
  - Par défaut : do\_load\_ref=0
  - Nexpos\_ref : nombre d'images de référence (images ne contenant pas d'exoplanète mais ayant à peu près le même fond que les données traitées)
  - Par défaut : Nexpos\_ref = 40 (inutile ici puisque utilisation\_ref = 0)
4. Inputs utilisés pour la première fois dans map\_pol.m
- do\_map\_pol : exécute map\_pol ou non (comme do\_estim\_psf, do\_boxes\_and\_grids ...)
  - Par défaut : do\_map\_pol = 1
5. Inputs utilisés pour la première fois dans create\_boxes.m
- m : facteur de suréchantillonnage utilisé dans l'interpolation pour la translation du *pois* sur la position initiale testée
  - Par défaut : m = 5
6. Inputs utilisés pour la première fois dans all\_estim.m
- do\_create\_bean : crée les haricots ( $\{\mathbf{p}(k, r)\}_{k=1, \dots, K}$ ) ou recharge les anciens
  - Par défaut : do\_create\_bean = 1
  - tpose\_unit : durée d'une pose (en secondes)
  - Par défaut : tpose\_unit = 100 (dépend des données)
  - tlecture : temps de transfert d'une pose qui vient s'ajouter à la durée d'une pose (en secondes)
  - Par défaut : tlecture = 0 (dépend des données)
  - dn\_rot : distance autorisée entre deux pois situés à rho\_max (pour produire un haricot  $\mathbf{p}(k, r)$  assez continu pour une pose  $k$  donnée, voir la section 4.1) (en pixels)
  - Par défaut : dn\_rot = 0.7

## Etape 2 : Détection et estimation finale

L'ensemble des inputs demandés à l'utilisateur pour l'étape d'estimation sont des variables à définir dans la seule fonction `call_detec`.

1. Pour charger les inputs et outputs de l'étape 1 d'estimation :
  - `file_param` : adresse du répertoire qui contient les inputs de l'étape 1  
Par défaut : `file_param = '../step1_estimation/results/param_070608_T2_f1'` (dépend de la structure des répertoires)
  - `file_result` : adresse du répertoire qui contient les inputs de l'étape 2  
Par défaut : `file_result = '../step1_estimation/results/res_070608_T2_f1'` (dépend de la structure des répertoires)

Les inputs listés ci-dessous sont spécifiques à l'étape de détection.
2. Input fondamental pour l'ensemble de l'étape de détection :
  - `do_empirical` : essai de calcul empirique de seuil (`do_empirical = 1`) ou non (`do_empirical = 0`)  
Par défaut : `do_empirical = 1`
3. Inputs utilisés pour la première fois dans `quick_thresh.m`
  - `do_quick_thresh1` : exécute `quick_thresh` (`do_quick_thresh1=1`) ou recharge le résultat d'une exécution précédente (`do_quick_thresh1=0`)  
Par défaut : `do_quick_thresh1 = 1`
  - `r_resol` : taille du nouveau pixel (pour une résolution plus grande que la taille des pixels de la CCD) exprimée comme une fonction de l'ancienne. Une résolution plus fine (`r_resol > 1`) évite une convolution artificielle d'une carte sous-jacente réelle  $\hat{\alpha}(r)$  et permet donc de plus fines détections, mais accepte en contrepartie des détections dues à des erreurs numériques non compensées par un moyennage local.  
Par défaut : `r_resol = 0.5`
  - `z_thresh` : vecteur contenant les valeurs  $z$  du seuil (si la variable aléatoire  $\hat{\alpha}(r)$  était gaussienne,  $z=3$  correspondrait à  $PFA = 0.001$ )  
Par défaut : `z_thresh = [10 20 30 40 50 60]`
  - `rad_dil_init` : rayon du disque utilisé pour remplir les trous laissés par l'échantillonnage des positions initiales  $r$  utilisées pour le calcul de  $\hat{\alpha}(r)$  (effet de visualisation)  
Par défaut : `rad_dil_init = 2`
4. Inputs utilisés pour la première fois dans `argmax_ann.m`
  - `do_source_search1`  
Par défaut : `do_source_search1 = 1`
  - `drho_case`  
Par défaut : `drho_case = 15`
  - `d_max`  
Par défaut : `d_max = 20`
5. Inputs utilisés pour la première fois dans `sup_sources.m`
  - `do_sup_sources`  
Par défaut : `do_sup_sources = 1`
  - `delta_p`  
Par défaut : `delta_p = d_max`
6. Inputs utilisés pour la première fois dans `H0_sample_annuli.m`
  - `do_stats_H0`  
Par défaut : `do_stats_H0 = 1`
  - `nbin_hist`  
Par défaut : `nbin_hist = 300`
  - `nbin_ks`  
Par défaut : `nbin_ks = 300`



- drho\_study\_H0  
Par défaut : drho\_study\_H0 = 30
- 7. Inputs utilisés pour la première fois dans the second use of quick\_thresh.m
  - do\_final\_thresh  
Par défaut : do\_final\_thresh = 1
  - pfa  
Par défaut : pfa=[ $10^{-3}$ ,  $10^{-5}$ ,  $10^{-7}$ ,  $10^{-9}$ ]
- 8. Inputs utilisés pour la première fois dans the third use of quick\_thresh.m (si do\_empirical true)
  - z\_empir  
Par défaut : z\_empir = [1 3 5]
- 9. Inputs utilisés pour la première fois dans the second use of argmax\_ann.m
  - do\_final\_estim  
Par défaut : do\_final\_estim = 1
- 10. Inputs utilisés pour la première fois dans correct.m ou correct\_and\_new\_images.m
  - do\_correct  
Par défaut : do\_correct = 1
  - do\_new\_images  
Par défaut : do\_new\_images = 0

## Annexe B

# Théorie bayésienne de la décision pour la construction d'inférences

Pour une introduction à la théorie de la décision bayésienne dans un cadre proche du notre, voir Berger [1985]. Pour une introduction synthétique et informative à la théorie de la décision appliquée au test d'hypothèses, voir Hwang et al. [1992].

### B.1 Théorie de la décision : généralités

En théorie de la décision, une inférence (décision statistique  $d(\mathbf{x})$  à valeur dans un espace de décision  $\mathcal{D}$ ) est obtenue par la minimisation d'un *risque* moyen. Le risque moyen d'une décision est la moyenne d'un coût  $L(f(\boldsymbol{\theta}), d(\mathbf{x})) \in \mathbb{R}_+$  où  $d(\mathbf{x})$  fait une inférence sur  $f(\boldsymbol{\theta})$ . Typiquement, pour les  $\mathbf{x}$  et  $\boldsymbol{\theta}$  tels que  $d(\mathbf{x})$  infère parfaitement  $f(\boldsymbol{\theta})$  le coût  $L(f(\boldsymbol{\theta}), d(\mathbf{x}))$  est pris nul, et le coût est choisi ensuite d'autant plus grand que  $d(\mathbf{x})$  infère mal  $f(\boldsymbol{\theta})$ . Parmigiani and Inoue [2009] présente une synthèse introductive de l'état de l'art concernant la théorie de la décision. Berger [1985], Robert [2007] permettent de s'introduire au sujet à travers une présentation du problème de l'inférence bayésienne.

Pour une estimation ponctuelle de  $f(\boldsymbol{\theta}) = \boldsymbol{\theta}$  par exemple, on note l'inférence  $d(\mathbf{x}) = \hat{\boldsymbol{\theta}}(\mathbf{x})$  à valeur dans  $\mathcal{D} = \Theta$  et on choisit classiquement comme fonction de coût le coût quadratique défini comme l'erreur quadratique de  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  :

$$L_2(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) = \|\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\|^2 \quad (\text{B.1})$$

Selon ce qui est souhaité, la fonction de coût peut être moyennée sur l'ensemble  $\mathcal{X}$  des observations possibles et/ou sur l'ensemble  $\Theta$  des paramètres possibles.

**Définition 23** On définit les risques fréquentiste  $R(\boldsymbol{\theta}, d)$ , bayésien  $\rho(\pi, d(\mathbf{x}))$  et totalement intégré  $r(\pi, d)$  par

$$R(\boldsymbol{\theta}, d) = \int_{\mathcal{X}} d\mathbf{x} L(\boldsymbol{\theta}, d(\mathbf{x}))p(\mathbf{x}|\boldsymbol{\theta}) \quad (\text{B.2})$$

$$\rho(\pi, d(\mathbf{x})) = \int_{\Theta} d\boldsymbol{\theta} L(\boldsymbol{\theta}, d(\mathbf{x}))\pi(\boldsymbol{\theta}|\mathbf{x}) \quad (\text{B.3})$$

$$r(\pi, d) = \int_{\mathcal{X}} d\mathbf{x} \int_{\Theta} d\boldsymbol{\theta} L(\boldsymbol{\theta}, d(\mathbf{x}))p(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (\text{B.4})$$

Pour définir le risque comme une simple moyenne de la fonction de coût et de telle sorte qu'il puisse être minimisé, il était bien nécessaire de définir la fonction de coût sur un sous-espace de  $\mathbb{R}$  afin d'avoir une relation d'ordre sur l'espace du risque.

**Définition 24** On appelle décision de Bayes une statistique  $d_B(\mathbf{x})$  qui minimise le risque bayésien (B.3) par rapport à  $d(\mathbf{x})$  où  $\mathbf{x}$  sont les données observées.

$$d_B(\mathbf{x}) = \arg \min_{d(\mathbf{x}) \in \mathcal{D}} \rho(\pi, d(\mathbf{x}))$$

Ainsi définie quel que soit  $\mathbf{x} \in \mathcal{X}$ , la fonction  $d_B(\cdot)$  minimise aussi le risque intégré (B.4). Sous l'approche de la théorie de la décision, on obtient donc une décision  $d(\mathbf{x})$  lorsque la vraisemblance  $p(\mathbf{x}|\boldsymbol{\theta})$ , la fonction de coût  $L(\boldsymbol{\theta}, d(\mathbf{x})) \in \mathbb{R}_+$  et éventuellement l'a priori  $\pi(\boldsymbol{\theta})$  sont fixés.

Notons que les risques sont les moyennes d'une *erreur* et indiquent donc naturellement une précision *moyenne* de l'inférence  $d_B$ . Grâce à la théorie de la décision, construction d'une inférence et estimation d'erreur (puisque c'est justement la grandeur minimisée) de cette inférence forment bien un problème joint. Rukhin [1988] montre même que sous une certaine classe de fonctions de coût jointes sur l'inférence  $d$  et sa *précision*, la décision de Bayes revient bien à opter pour la décision  $d_B(\mathbf{x})$  et à estimer l'erreur par son risque comme nous l'avons fait ci-dessus par exemple. Plus généralement, toute fonction de coût jointe vérifiant cette propriété est appelée une *scoring rule* propre Schervish [1989].

## B.2 Estimation ponctuelle d'un paramètre

L'estimation ponctuelle consiste à définir et utiliser une décision  $d(\mathbf{x})$  à valeur dans le même espace que le paramètre à estimer, typiquement  $d(\mathbf{x}) = \hat{\boldsymbol{\theta}}(\mathbf{x})$  avec  $\mathcal{D} = \Theta$ . Faire une inférence consiste à construire une fonction déterministe des données, donc estimer  $f(\boldsymbol{\theta})$  peut revenir à estimer une fonction  $g(\boldsymbol{\theta}, \mathbf{x})$ . Par souci de clarté cependant nous nous limitons dans cette section à l'estimation classique de  $f(\boldsymbol{\theta}) = \boldsymbol{\theta}$ .

Comme on l'a dit, une fonction de coût classique pour l'estimation ponctuelle est le coût quadratique (B.1). Elle a les intéressantes propriétés suivantes :

**Théorème 7** L'estimateur bayésien de  $\boldsymbol{\theta}$  associé au coût quadratique (B.1) est donné par la moyenne a posteriori de  $\boldsymbol{\theta}$  :

$$\hat{\boldsymbol{\theta}}_{MMSE}(\mathbf{x}) = \mathbb{E}[\boldsymbol{\theta}|\mathbf{x}] \quad (\text{B.5})$$

On le note  $\hat{\boldsymbol{\theta}}_{MMSE}(\mathbf{x})$  pour *Minimum Mean Square Error*.

**Proposition 29** Le risque bayésien associé à l'estimateur MMSE est simplement la somme des variances marginales a posteriori de  $\boldsymbol{\theta}$  :

$$\rho(\pi, \hat{\boldsymbol{\theta}}_{MMSE}(\mathbf{x})) = \mathbb{E}[\|\boldsymbol{\theta} - \mathbb{E}[\boldsymbol{\theta}|\mathbf{x}]\|^2|\mathbf{x}] = \sum_m \text{var}[\theta_m|\mathbf{x}] \quad (\text{B.6})$$

En particulier, si  $\theta$  est scalaire la racine du risque associé au coût quadratique est ainsi appelée en général déviation standard a posteriori, et comme elle a la dimension de  $\theta$  elle donne directement une mesure d'*incertitude* sur  $\theta$  :

$$\begin{aligned} \text{Err}_B(\mathbf{x}) &= \sqrt{\rho(\pi, d_B(\mathbf{x})|\mathbf{x})} = \text{std}[\theta|\mathbf{x}], \quad \theta \in \mathbb{R} \\ \theta &= \hat{\theta}_{MMSE}(\mathbf{x}) \pm \text{Err}_B(\mathbf{x}) \end{aligned} \quad (\text{B.7})$$

Plus généralement, si  $\boldsymbol{\theta}$  est un vecteur, on remplace l'intervalle par un domaine et celui-ci n'est donc pas défini uniquement à partir du risque réel  $\rho$ . Il est donné classiquement par l'ellipse centrée en  $\hat{\boldsymbol{\theta}}_{MMSE}(\mathbf{x})$  et dont la forme est définie par la matrice de covariance de  $\boldsymbol{\theta}$  :

$$\text{cov}[\boldsymbol{\theta}|\mathbf{x}] = \mathbb{E}[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{MMSE}(\mathbf{x}))(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{MMSE}(\mathbf{x}))^t|\mathbf{x}] \quad (\text{B.8})$$

Une autre fonction de coût usuel est le coût en norme  $L_1$  :

$$L_1(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) = \|\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}\|_1 = \sum_m |\hat{\theta}_m(\mathbf{x}) - \theta_m| \quad (\text{B.9})$$

On montre que

**Proposition 30** *Tout estimateur bayésien associé à la fonction de coût (B.9) est donné par  $(\hat{\theta}_{B,1})_m = \text{med}[\theta_m|\mathbf{x}]$  où  $\text{med}[\theta_m|\mathbf{x}]$  est une médiane marginale a posteriori de  $\theta_m$ .*

**Proposition 31** *Le risque bayésien associé à l'estimateur "médiane a posteriori" est égal à la somme des erreurs absolues marginales calculées par rapport à la médiane a posteriori :*

$$\rho(\pi, \hat{\theta}_{B,1}(\mathbf{x})) = \sum_m \mathbb{E}[|\theta_m - \text{med}[\theta_m|\mathbf{x}]||\mathbf{x}] \quad (\text{B.10})$$

**Exemple 20** *Soit  $X_1, \dots, X_n$  des variables aléatoires scalaires indépendantes conditionnellement à  $\theta$  et identiquement distribuées selon  $p(x_i|\theta) \propto \exp(-\frac{(x_i-\theta)^2}{2\sigma^2})$  avec  $\sigma$  connu. Supposons un a priori gaussien<sup>1</sup>  $\pi(\theta) \propto \exp(-\frac{(\theta-\theta_0)^2}{2\sigma^2})$  avec  $\theta_0$  connu. Afin d'estimer  $\theta$ , que l'on choisisse la fonction de coût (B.1) ou (B.9), la minimisation de (B.3) donne  $\hat{\theta}_B(\mathbf{x}) = \frac{\sum_{i=1}^n x_i + \theta_0}{n+1}$  :  $\theta$  est estimée par la moyenne pondérée de la moyenne des données et de la moyenne a priori de  $\theta$ .*

Une dernière fonction de coût usuelle est associée à la fonction de coût binaire

$$L_{0-1}(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{X})) = 1 - I_{\{\hat{\boldsymbol{\theta}}(\mathbf{X})\}}(\boldsymbol{\theta}) \quad (\text{B.11})$$

où  $I$  est la fonction indicatrice définie par

$$\begin{aligned} I_A(\boldsymbol{\theta}) &= 1 && \text{si } \boldsymbol{\theta} \in A \\ &= 0 && \text{sinon} \end{aligned} \quad (\text{B.12})$$

**Proposition 32** *Si  $\theta$  est réel, l'estimateur bayésien associé à la fonction de coût (B.11) est donné par  $\hat{\theta}_{MAP}(\mathbf{x}) = \arg \max_{\theta \in \mathbb{R}} (\pi(\theta|\mathbf{x}))$ .*

En revanche si  $\boldsymbol{\theta}$  est multivarié le risque bayésien associé à la fonction de coût (B.11) ne dépend pas de  $\mathbf{x}$  donc ce coût convient d'être ajusté. On peut par exemple définir le coût

$$L_{0-1}^\zeta(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}(\mathbf{x})) = 1 - I_{\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{x})\| < \zeta\}}(\boldsymbol{\theta}) \quad (\text{B.13})$$

**Proposition 33** *L'estimateur bayésien associé à la fonction de coût (B.13) tend, lorsque  $\zeta$  tend vers 0, vers  $\hat{\boldsymbol{\theta}}_{MAP}(\mathbf{x}) = \arg \max_{\boldsymbol{\theta} \in \Theta} (\pi(\boldsymbol{\theta}|\mathbf{x}))$ .*

## B.3 Intervalle ou domaine de confiance d'un paramètre

En estimation ponctuelle bayésienne d'un paramètre, il est standard d'écrire comme dans l'Eq. (B.8) que la covariance du paramètre donne une idée de l'espace d'appartenance du paramètre vrai  $\boldsymbol{\theta}$  sous-jacent aux données  $\mathbf{x}$ . De façon plus restrictive, si  $\boldsymbol{\theta}$  est réel le risque peut être utilisé comme dans l'Eq. (B.7) comme estimateur de l'erreur de  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  : le risque associé à l'estimation ponctuelle  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  permet comme en cas multivarié de définir un intervalle centré "d'incertitude"  $C(\mathbf{x})$  :  $C(\mathbf{x}) = [\hat{\boldsymbol{\theta}}_{\text{MMSE}}(\mathbf{x}) - \text{Err}_B(\mathbf{x}), \hat{\boldsymbol{\theta}}_{\text{MMSE}}(\mathbf{x}) + \text{Err}_B(\mathbf{x})]$ .

Cependant, ceci ne peut être fait que dans un cas très limitatif et même alors cet espace n'est pas à strictement parler un espace de confiance car la théorie reposait sur le coût associé à une estimation ponctuelle. En fait, un intervalle de confiance est défini à partir de la probabilité d'appartenance de  $\boldsymbol{\theta}$  à un domaine  $C(\mathbf{x})$ , la probabilité étant calculée conditionnellement à  $\mathbf{x}$  en bayésien (et conditionnellement à  $\boldsymbol{\theta}$  en fréquentiste). En particulier, si la densité a posteriori du paramètre à estimer est multimodale ou dissymétrique cette mesure de précision comme "erreur" n'est pas adaptée.

<sup>1</sup>Cet a priori est dit *conjugué* avec la vraisemblance : la loi a posteriori est alors de la même famille que la loi a priori.

Ainsi, l'estimation d'un intervalle de confiance (ou région de confiance si  $\Theta$  a plus d'une dimension ou que le domaine n'est pas convexe) constitue une inférence à proprement parler. Une synthèse récente des travaux sur l'estimation d'un espace de confiance dans le cadre de la théorie de la décision est faite dans Rice et al. [2008] et plus anciennement dans Casella et al. [1993].

Pour ce type d'inférence, la décision  $d(\mathbf{x})$  appartient à un espace  $\mathcal{D}$  de plus grande dimension que la dimension de l'espace du paramètre à inférer ( $\Theta$  typiquement). Pour  $\Theta = \mathbb{R}$  par exemple,  $\mathcal{D}$  regroupe toutes les unions d'intervalles de  $\mathbb{R}$ . Les fonctions de coût utilisées en estimation ponctuelle ne sont donc plus utilisables en tant que telles pour ce problème.

Deux critères apparaissent comme les critères fondamentaux pour définir une estimation de domaine : le fait que la valeur vraie  $\theta$  appartienne bien au domaine (notion de *biais*) et que le domaine ne soit pas trop étendu (notion de *précision*). Si le biais est considéré comme une condition sine qua non, elle peut constituer une contrainte et la précision un paramètre à minimiser. Nous verrons ceci plus bas. Sinon en restant dans le cadre strict décrit ci-dessus il est possible de considérer les deux paramètres conjointement dans une fonction de coût.

La fonction de coût la plus classique est une simple somme pondérée de ces deux contributions :

$$L(\theta, C(\mathbf{x})) = 1 - I_{C(\mathbf{x})}(\theta) + \zeta \text{vol}(C(\mathbf{x})) \quad (\text{B.14})$$

où  $\lambda > 0$ ,  $I$  est la fonction indicatrice définie par l'Eq. (B.12) et "vol" est le volume sous la mesure de Lebesgue :  $\text{vol}(C(\mathbf{x})) = \int_{C(\mathbf{x})} d\theta$ . On constate que ce coût se rapproche fortement du coût menant à l'estimateur MAP dans le cas de l'estimation ponctuelle.

Ainsi, si  $\theta$  est réel, la décision bayésienne associée à ce coût est directement donnée par une région HPD :

**Proposition 34** *Si  $\mathcal{D}$  est l'ensemble des unions des sous-espaces mesurables de  $\Theta$  alors sous la fonction de coût (B.14) la décision de Bayes  $C_B(\mathbf{x})$  est une région donnée par :*

$$C_B(\mathbf{x}) = \{\theta : \pi(\theta|\mathbf{x}) \geq \zeta\} \quad (\text{B.15})$$

**Définition 25** *Une région de la forme de l'Eq. (B.15) est appelée région **HPD** (pour **H**ighest **P**osterior **D**ensity).*

Cependant cette fonction de coût peut mener à des situations paradoxales. Casella et al. [1993] se focalisent par exemple sur le cas standard normal de moyenne et variance inconnues et observent que la décision bayésienne associée à ce coût peut être uniformément dominée (risque bayésien inférieur pour tout  $\mathbf{x}$ ) par une statistique non désirable intuitivement. Ils proposent alors pour résoudre ce paradoxe la fonction de coût alternative  $L(\theta, C(\mathbf{x})) = I_{\bar{C}(\mathbf{x})}(\theta) + \zeta S(\text{vol}(C(\mathbf{x})))$  où  $S$  est une fonction non linéaire et non décroissante.

Comme Berger [1985] on peut aussi définir le coût

$$L(\theta, C(\mathbf{x})) = I_{\bar{C}(\mathbf{x})}(\theta) + \zeta T(C(\mathbf{x})) \quad (\text{B.16})$$

avec comme mesure de taille d'un domaine  $T(A) = \int_A d\theta t(\theta)$ . En particulier, on peut souhaiter utiliser la mesure *a priori*. Si la densité *a priori* est toujours dominée par la mesure de Lebesgue on écrit simplement  $t(\cdot) = \pi(\cdot)$  et dans ce cas on montre comme précédemment que

**Théorème 8** *Si  $\mathcal{D}$  est l'ensemble des unions de sous-espaces mesurables de  $\Theta$  alors sous la fonction de coût (B.16) avec  $T(A) = \int_A d\theta \pi(\theta)$  la décision de Bayes  $C_B(\mathbf{x})$  est donnée par :*

$$C_B(\mathbf{x}) = \{\theta : \pi(\theta|\mathbf{x}) \geq \zeta \pi(\theta)\} = \{\theta : p(\mathbf{x}|\theta) \geq \zeta p(\mathbf{x})\} \quad (\text{B.17})$$

Le risque bayésien associé est

$$\rho(\pi, C_B(\mathbf{x})) = \Pr(p(\mathbf{x}|\boldsymbol{\theta}) < \zeta p(\mathbf{x})|\mathbf{x}) + \zeta \int_{C_B(\mathbf{x})} d\boldsymbol{\theta} \pi(\boldsymbol{\theta}) \quad (\text{B.18})$$

Une autre estimation classique (voir Berger [1985], Robert [2007] par exemple) de précision d'une estimation de région de confiance est simplement donnée par l'estimateur sous le coût quadratique (Eq. B.1) de  $I_{C(\mathbf{x})}(\boldsymbol{\theta})$ , soit :

$$\Pr(\boldsymbol{\theta} \in C_B(\mathbf{x})|\mathbf{x}) = \Pr(p(\mathbf{x}|\boldsymbol{\theta}) \geq \zeta p(\mathbf{x})|\mathbf{x}) \quad (\text{B.19})$$

A partir de maintenant et dans les deux prochaines sous-sections, nous ajoutons une composante au problème : l'existence d'une référence ( $\boldsymbol{\theta}_0$  ou espace  $\Theta_0$ ) à laquelle on souhaite *comparer* d'une façon ou d'une autre le paramètre  $\boldsymbol{\theta}$  inconnu. Cette différence peut s'exprimer via le choix de la fonction à inférer, la forme de la fonction de coût, voire la procédure bayésienne complète (en définissant deux risques par exemple, mais nous ne traiterons pas cette possibilité ici).

## B.4 Test d'hypothèses

### Hypothèse $H_0$ générale

La théorie de la décision mène aux statistiques de la section 1.3 lorsque l'on cherche à *estimer*  $i \in \{0, 1\}$ , l'indice de l'hypothèse  $H_i$ , sous les fonctions de coût standards vues dans la section B.2. L'inférence *test d'hypothèses* consiste alors en l'*estimation* d'une variable catégorique  $H_i$  ou d'une variable numérique ayant seulement deux valeurs possible : 0 ou 1. Cependant, on peut autoriser l'estimateur numérique à appartenir à l'intervalle  $\mathcal{D} = [0, 1]$  de telle sorte que les deux fonctions de coût présentées dans la section B.2 sont utilisables. Elles s'écrivent ici sous la forme

$$L_k(i, \hat{i}(\mathbf{x})) = |\hat{i}(\mathbf{x}) - i|^k \quad (\text{B.20})$$

avec  $k = 1, 2$ . D'après les définitions et théorèmes donnés dans la section B.2 on déduit simplement les estimateurs bayésiens résultants, ainsi que leurs précisions respectives données par leurs risques bayésiens  $\mathbb{E}[L_k(\hat{i}_{B,k}(\mathbf{x}), i)|\mathbf{x}] = \mathbb{E}[|\hat{i}(\mathbf{x}) - i|^k|\mathbf{x}]$ .

**Théorème 9** *Sous la fonction de coût binaire (Eq. B.11) aussi bien que sous la fonction de coût linéaire (Eq. B.20 avec  $k = 1$ ), l'estimateur de Bayes est donné par la décision binaire*

$$\begin{aligned} \hat{i}_{B,1}(\mathbf{x}) &= 0 & \text{si } \Pr(H_0|\mathbf{x}) > \frac{1}{2} \\ &= 1 & \text{sinon} \end{aligned} \quad (\text{B.21})$$

La précision a posteriori de cet estimateur est donnée pour les deux fonctions de coût par :

$$\begin{aligned} \rho_1(\pi, \hat{i}_{B,1}(\mathbf{x})|\mathbf{x}) &= \Pr(H_1|\mathbf{x}) & \text{si } \Pr(H_0|\mathbf{x}) > \frac{1}{2} \\ &= \Pr(H_0|\mathbf{x}) & \text{sinon} \end{aligned} \quad (\text{B.22})$$

La décision bayésienne est équivalente au test binaire donné par l'Eq. (1.30) pour  $\zeta = 1/2$ . Comme décrit par Robert [2007] on obtiendrait un coefficient  $\zeta \neq 1/2$  en pondérant la fonction de coût différemment selon que l'erreur concerne  $H_0$  ou  $H_1$ .

**Théorème 10** *Sous la fonction de coût quadratique (Eq. B.20 avec  $k = 2$ ), l'estimateur de Bayes est donné par*

$$\hat{i}_{B,2}(\mathbf{x}) = \Pr(H_1|\mathbf{x}) \quad (\text{B.23})$$

La précision a posteriori de cet estimateur est donnée par :

$$\rho_2(\pi, \hat{i}_{B,2}(\mathbf{x})|\mathbf{x}) = \Pr(H_0|\mathbf{x})\Pr(H_1|\mathbf{x})$$

Quelle que soit la fonction de coût, on constate que la valeur de  $\Pr(H_0|\mathbf{x})$  (et pas seulement le fait qu'elle soit inférieure ou supérieure à  $1/2$ ) joue un rôle central dans le test d'hypothèses : pour les fonctions de coût binaire et linéaire,  $\Pr(H_0|\mathbf{x})$  (ou  $1 - \Pr(H_0|\mathbf{x})$ ) indique la précision de la décision binaire, et pour la fonction de coût quadratique  $1 - \Pr(H_0|\mathbf{x})$  indique directement l'estimation de  $i$ .

Finalement, nous constatons simplement que la théorie de la décision converge vers l'utilisation de  $\Pr(H_0|\mathbf{x})$  en lui donnant des interprétations assez simples et tout à fait cohérentes avec l'interprétation bayésienne standard de la *probabilité a posteriori de l'hypothèse  $H_0$* .

### Test d'hypothèses paramétriques

L'application des résultats généraux de théorie de la décision en test d'hypothèses à une hypothèse de la forme

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \tag{B.24}$$

découle de la description faite dans la section 1.3.

Comme décrit dans Berger [1985], Hwang et al. [1992], Robert [2007], la théorie de la décision s'exprime dans ce cas non plus par l'estimation de  $i \in \{0, 1\}$  mais de façon équivalente par l'estimation de la fonction indicatrice  $I_{\Theta_0}(\boldsymbol{\theta})$  définie dans l'Eq. (B.12) :

$$I_{\Theta_0}(\boldsymbol{\theta}) = 1 - i \tag{B.25}$$

Avec les fonctions de coût binaire, linéaire et quadratique, les conclusions générales de la section B.4 s'appliquent bien entendu en faisant les changements de notations que l'on vient de préciser.

### Ajustement de modèle et théorie de l'information

L'application des résultats généraux de théorie de la décision en test d'hypothèses à une hypothèse de la forme

$$H_0 : p_{\boldsymbol{\theta}} \in \mathcal{P}_{\Theta_0} \tag{B.26}$$

découle de la description faite dans la section 1.3.

# Annexe C

## Figures et démonstrations supplémentaires

### C.1 Figures supplémentaires illustrant les données issues de la simulation physique fine

L'ensemble des figures de cette annexe a pour objet de décrire qualitativement et/ou statistiquement les données simulées par l'outil de simulation physique fine du *package* SPHERE du *problem-solving environment* CAOS, décrit dans la section 2.1. Cette analyse statistique est introduite dans la section 2.2.

#### Analyse statistique des données de référence, pour chaque anneau de 0'18 à 2'2

Les figures utilisées dans l'analyse dite *locale* de la section 2.2 sont présentées dans cette section. On appelle l'analyse *locale* car le champ d'intérêt est divisé en anneaux concentriques dont on présente ici individuellement les statistiques. Les anneaux sont caractérisés par une largeur identique de 10 pixels et des séparations moyennes données par  $\rho \in \{0'18, 0'37, 0'55, 0'73, 0'92, 1'1, 1'3, 1'5, 2'2\}$ . On peut visualiser sur la figure 2.3 où sont situés ces anneaux par rapport aux structures spatiales globales de l'image.

Pour chaque séparation  $\rho$  les données étudiées sont ainsi  $\{x_\ell(k), k \in \llbracket 1, K \rrbracket, \ell \in \mathcal{A}(\rho)\}$  avec  $\mathcal{A}(\rho)$  l'ensemble des pixels appartenant à l'anneau de séparation moyenne  $\rho$ . Une page présente les statistiques d'un anneau.

- Les deux images du haut de la page présentent la moyenne temporelle (à gauche) et l'écart type temporel (à droite) des données présentes dans l'anneau étudié (analyse pixel par pixel). Ces figures sont des détails des images des figures 2.4 et 2.5 respectivement.
- Les graphes de la seconde ligne de figures de la page présentent réciproquement la moyenne spatiale (à gauche) et la variance empirique spatiale (à droite) des données étudiées (analyse pose par pose). Ces figures sont des détails des figures 2.7 et 2.8 respectivement, mais plutôt sous forme de variance que d'écart type pour la seconde figure.
- Les 4 images de la troisième ligne de figures présentent les données  $\{x_\ell(k), \ell \in \mathcal{A}(\rho)\}$  pour  $k \in \{1, 50, 100, 144\}$ , qui correspondent à quatre régimes atmosphériques différents comme décrit rapidement dans la section 2.2.
- Les 4 histogrammes finaux présentent les distributions associées aux 4 images.

L'information contenue dans l'ensemble des figures présentées ici est discutée dans la section 2.2.



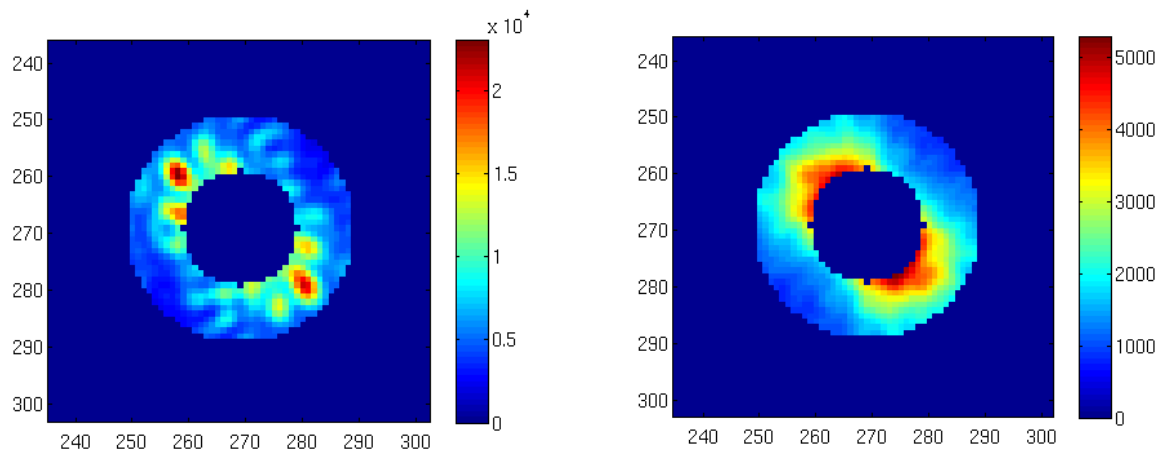


FIG. 1 – Séparation moyenne : 15 pixels, ie 0''18

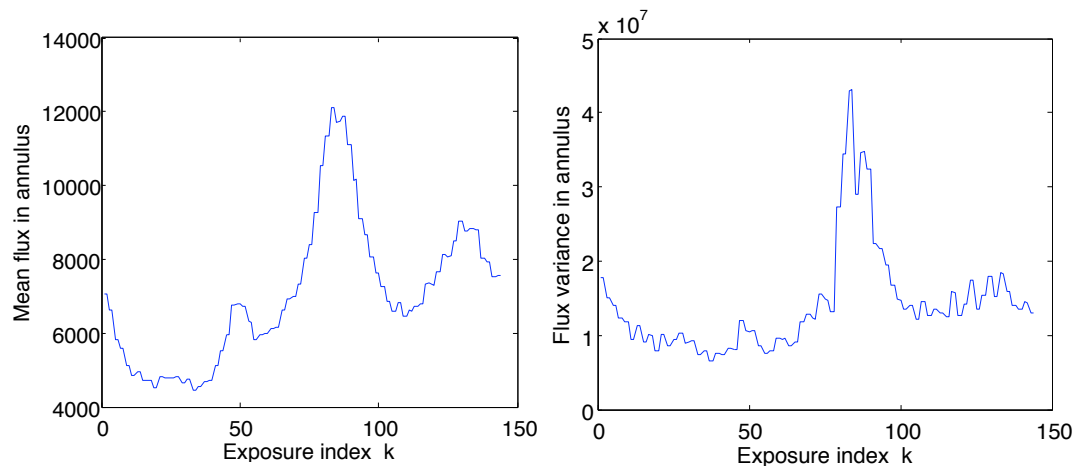


FIG. 2 – Séparation moyenne : 15 pixels, ie 0''18

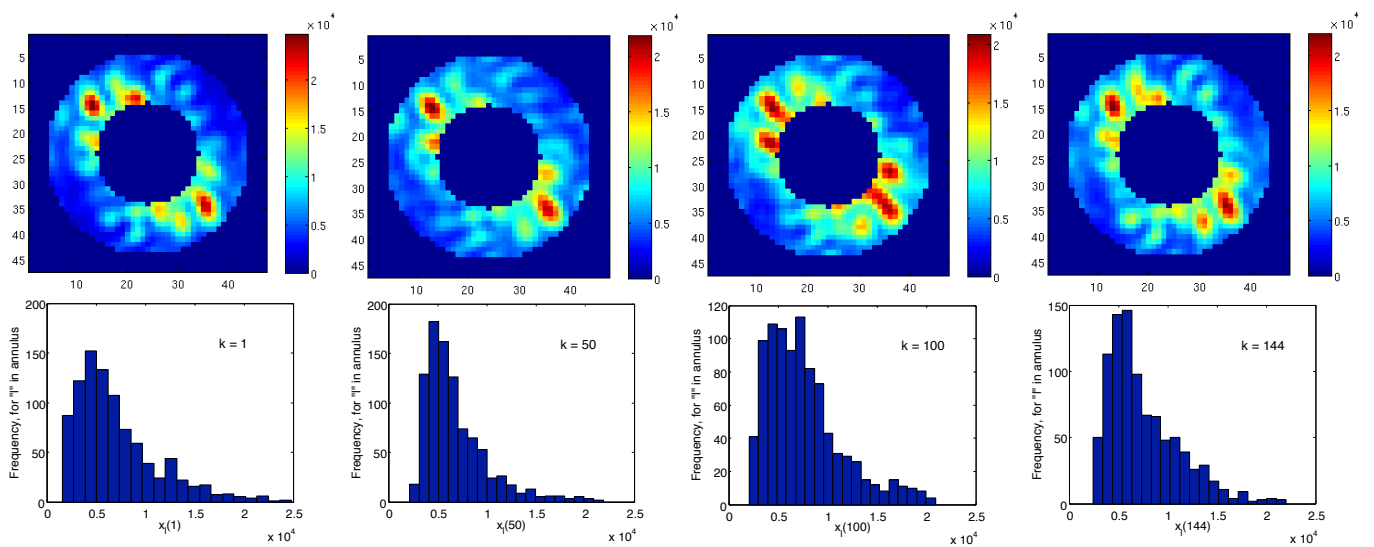


FIG. 3 – Séparation moyenne : 15 pixels, ie 0''18

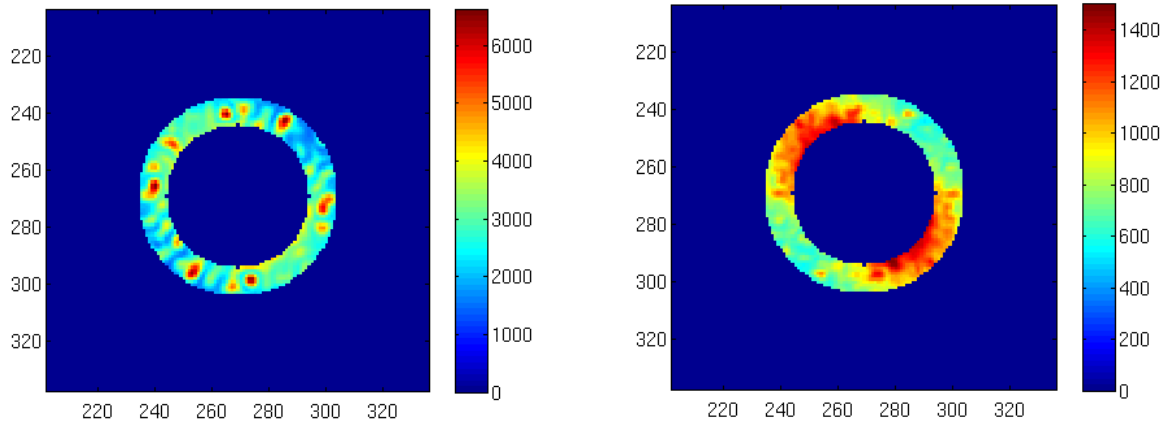


FIG. 4 – Séparation moyenne : 30 pixels, ie 0'37

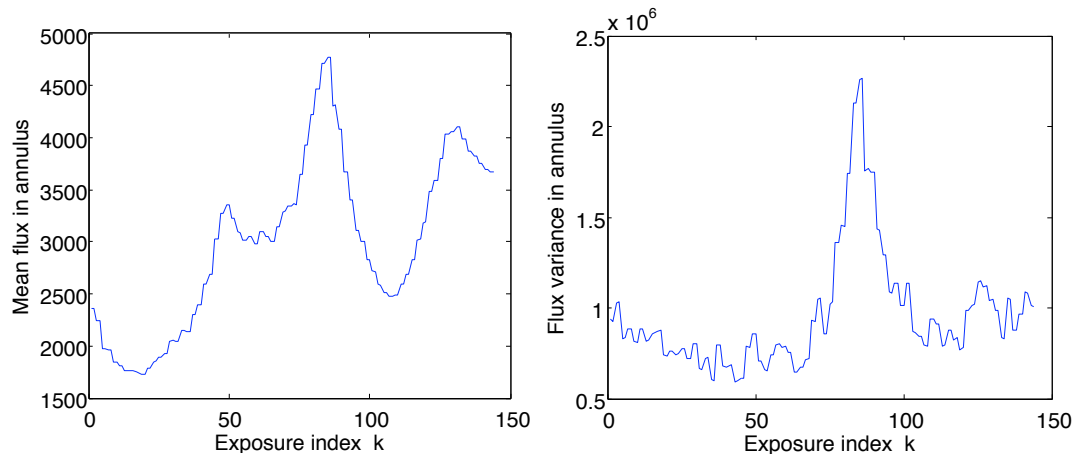


FIG. 5 – Séparation moyenne : 30 pixels, ie 0'37

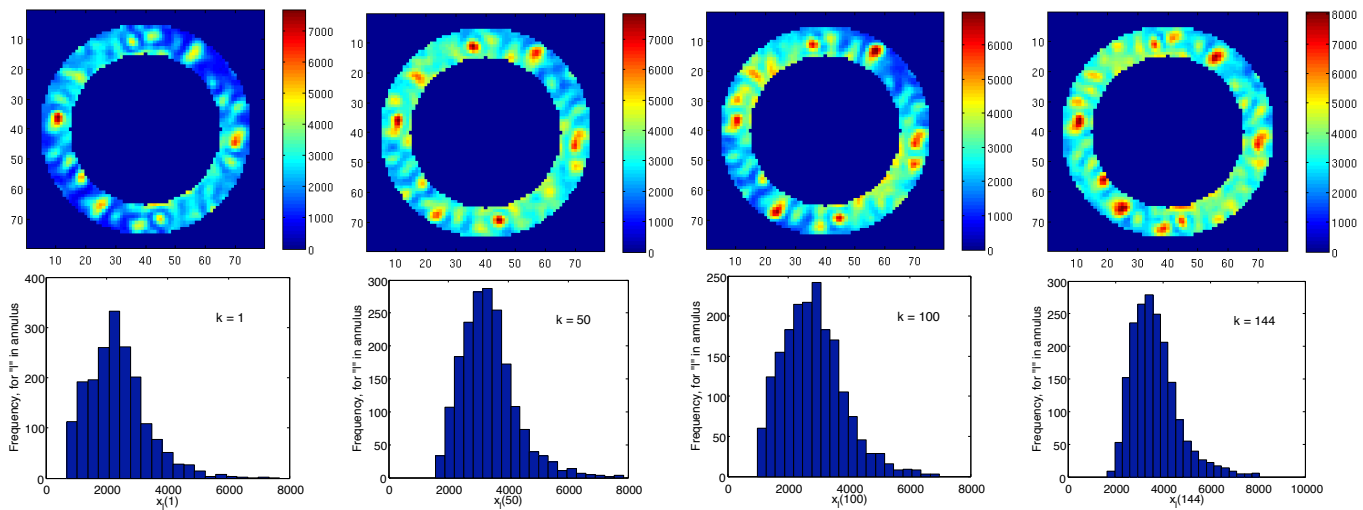


FIG. 6 – Séparation moyenne : 30 pixels, ie 0'37

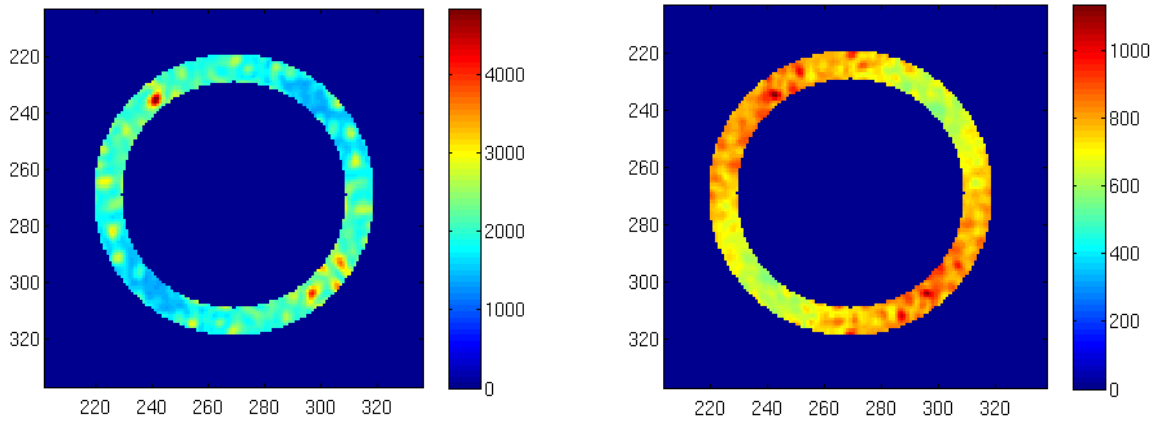


FIG. 7 – Séparation moyenne : 45 pixels, ie 0''55

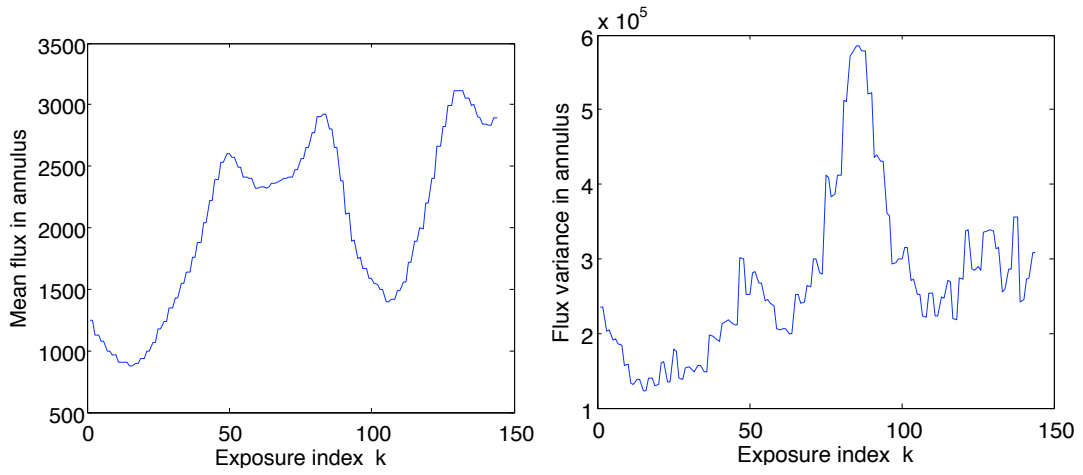


FIG. 8 – Séparation moyenne : 45 pixels, ie 0''55

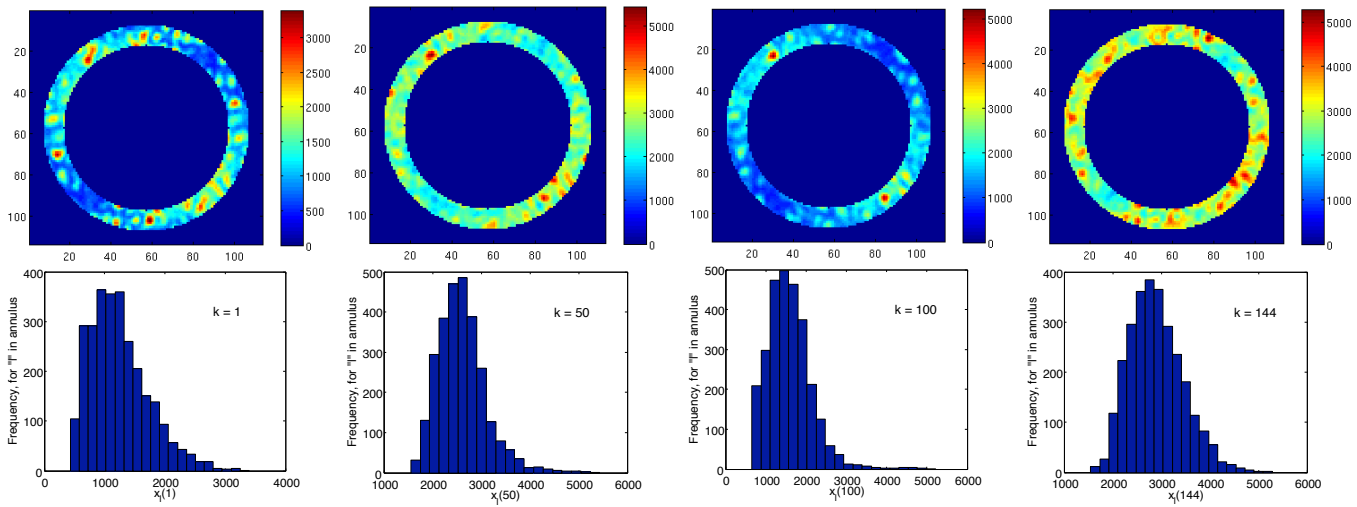


FIG. 9 – Séparation moyenne : 45 pixels, ie 0''55

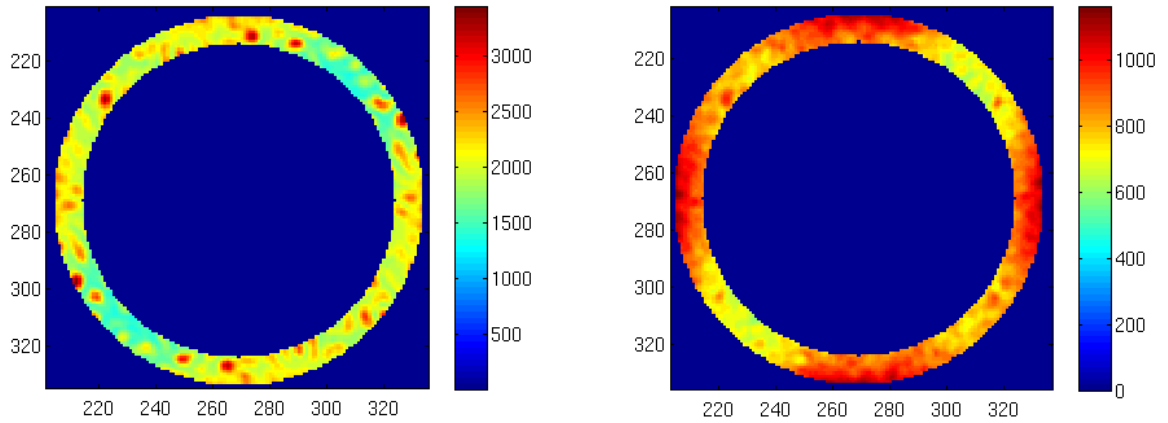


FIG. 10 – Séparation moyenne : 60 pixels, ie 0''.73

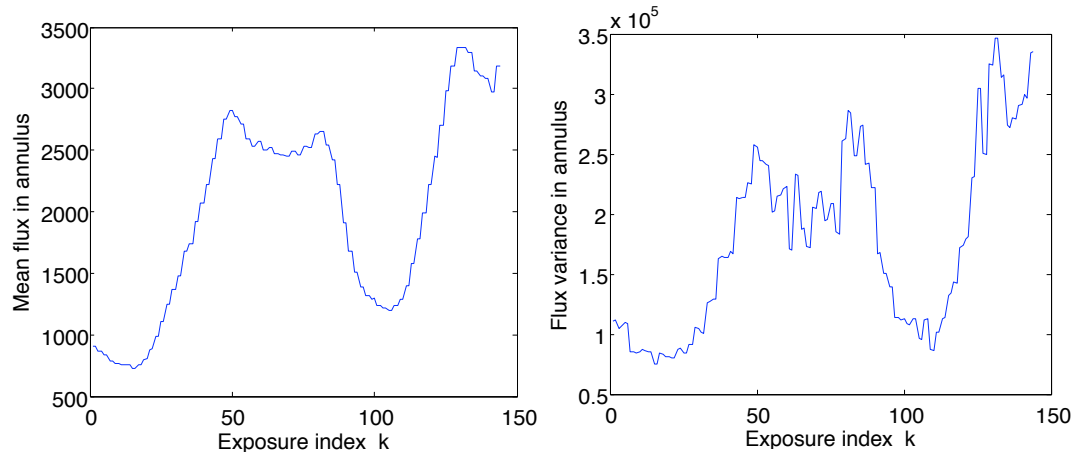


FIG. 11 – Séparation moyenne : 60 pixels, ie 0''.73

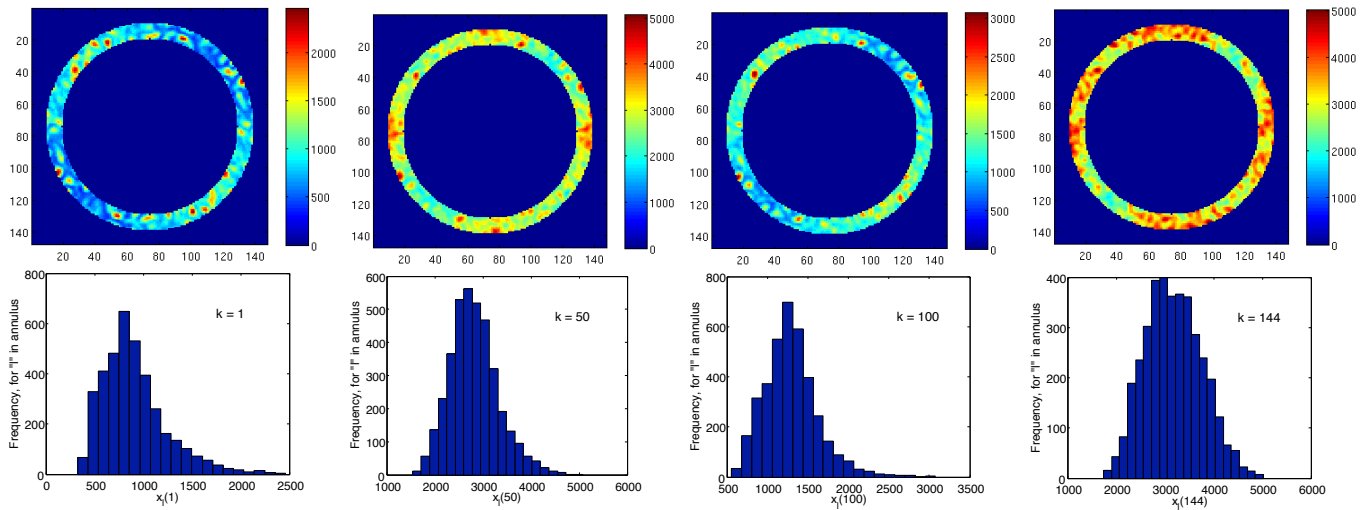


FIG. 12 – Séparation moyenne : 60 pixels, ie 0''.73

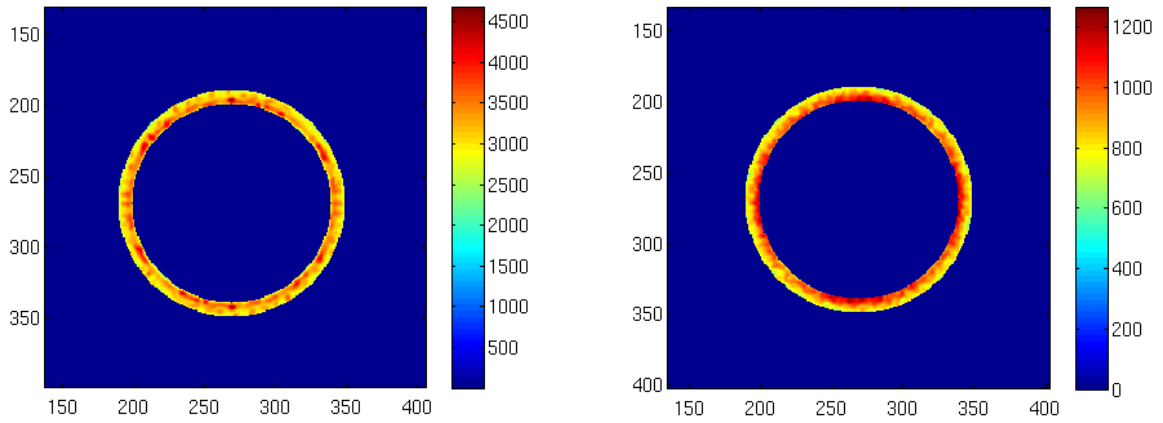


FIG. 13 – Séparation moyenne : 75 pixels, ie 0'92

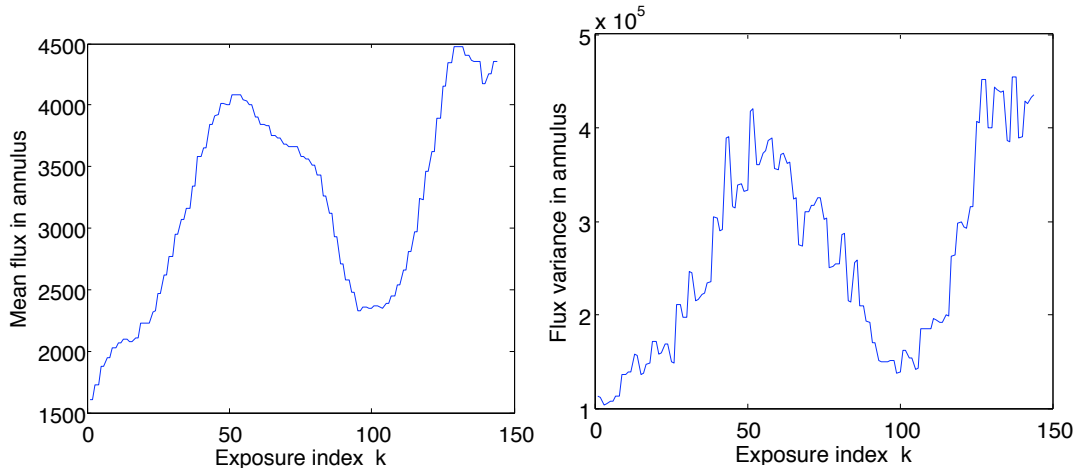


FIG. 14 – Séparation moyenne : 75 pixels, ie 0'92

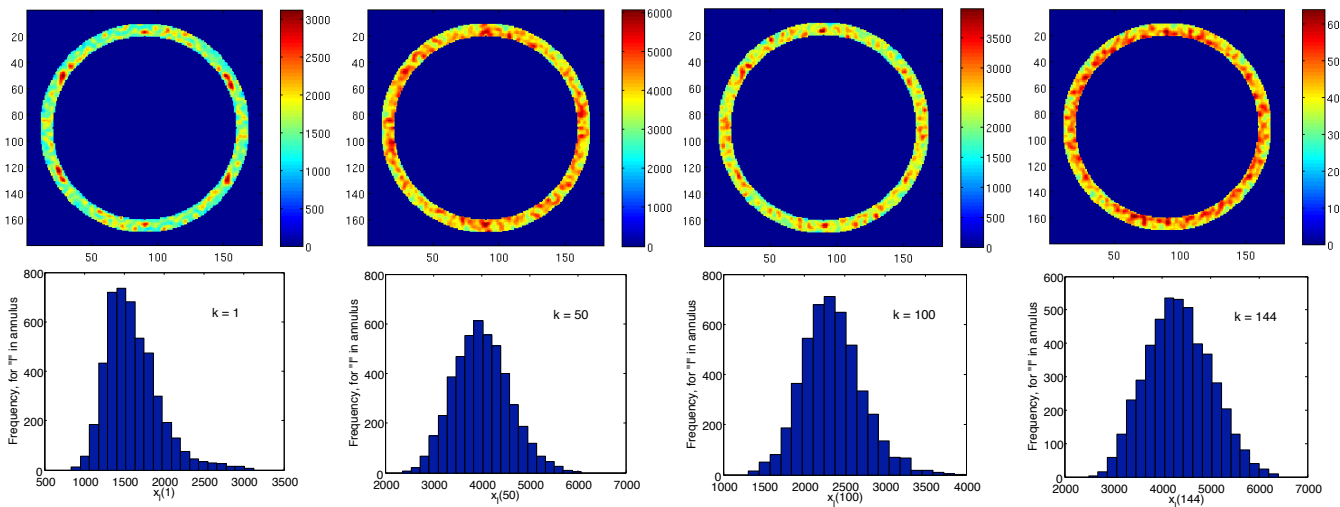


FIG. 15 – Séparation moyenne : 75 pixels, ie 0'92

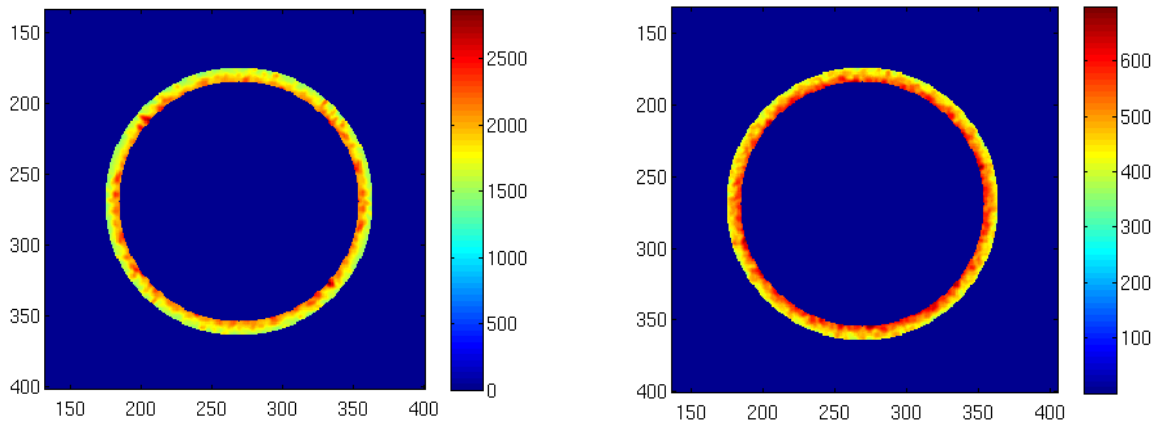


FIG. 1 – Séparation moyenne : 90 pixels, ie 1''

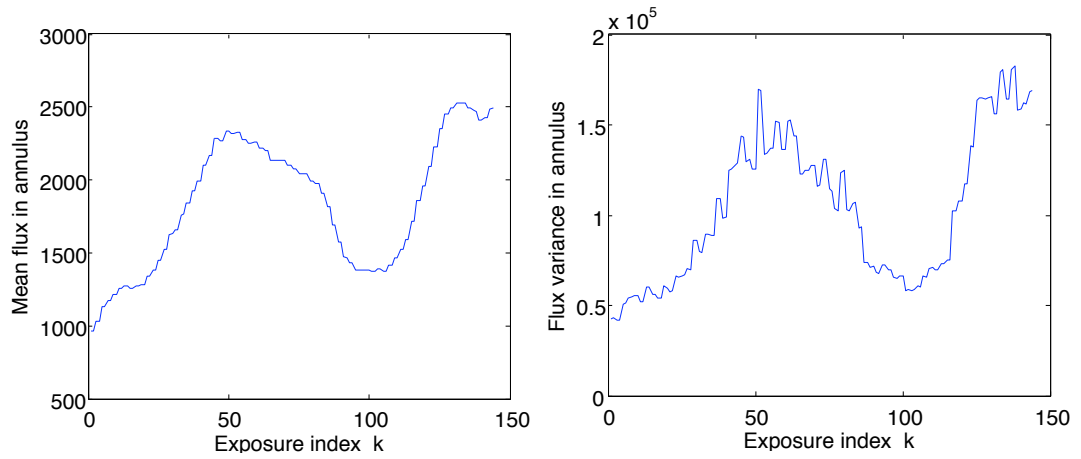


FIG. 2 – Séparation moyenne : 90 pixels, ie 1''

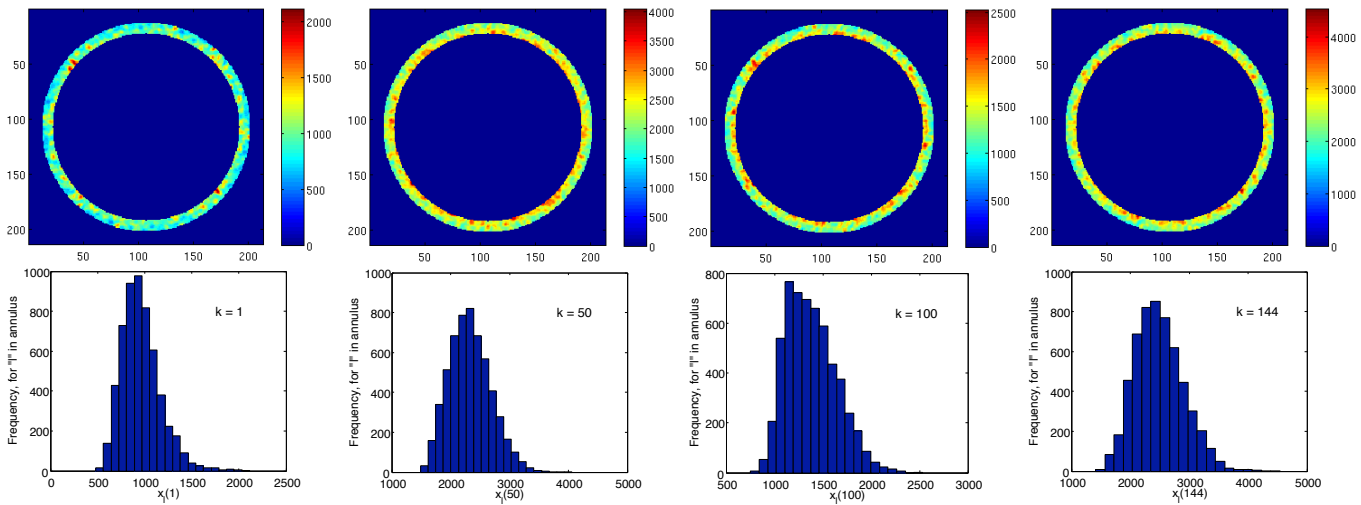


FIG. 3 – Séparation moyenne : 90 pixels, ie 1''

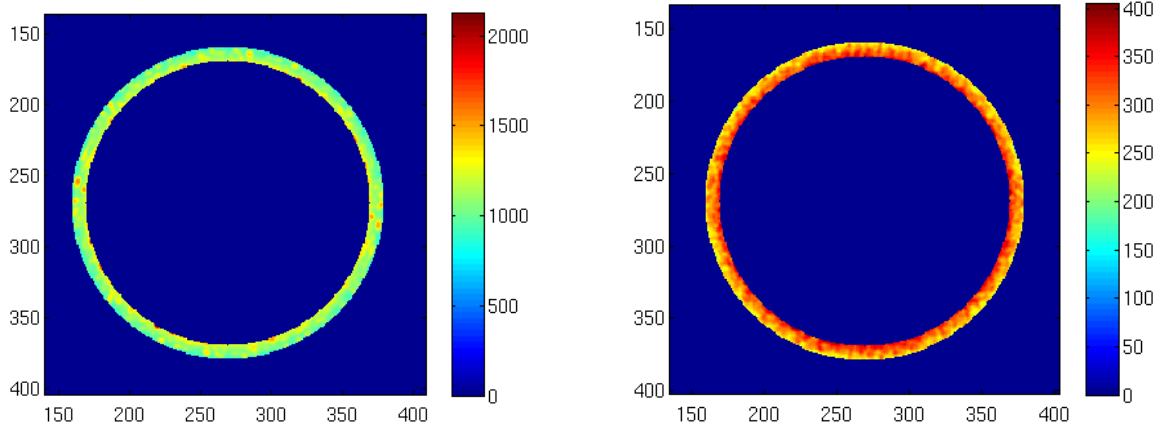


FIG. 4 – Séparation moyenne : 105 pixels, ie 1''3

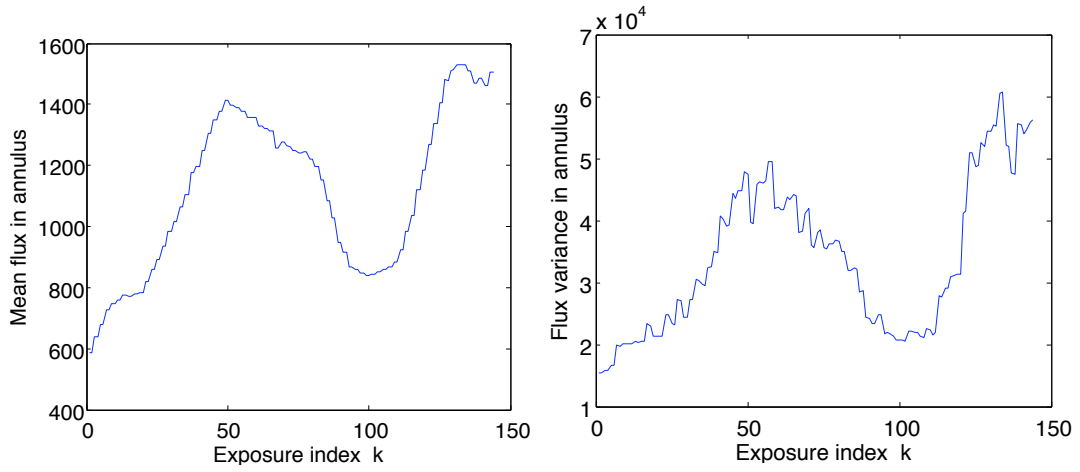


FIG. 5 – Séparation moyenne : 105 pixels, ie 1''3

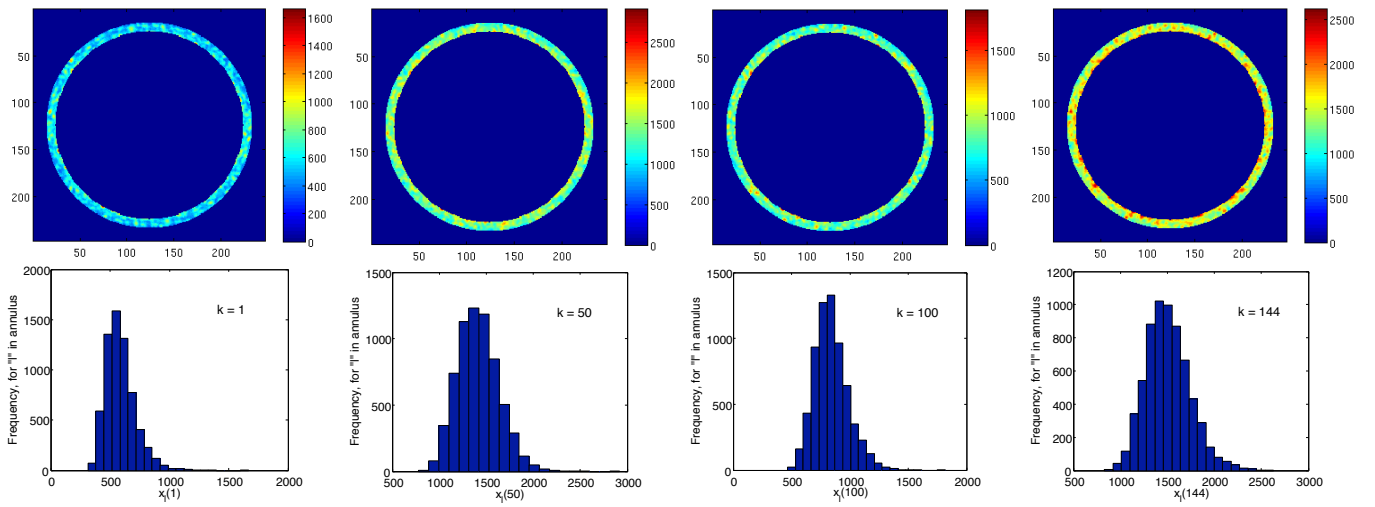


FIG. 6 – Séparation moyenne : 105 pixels, ie 1''3

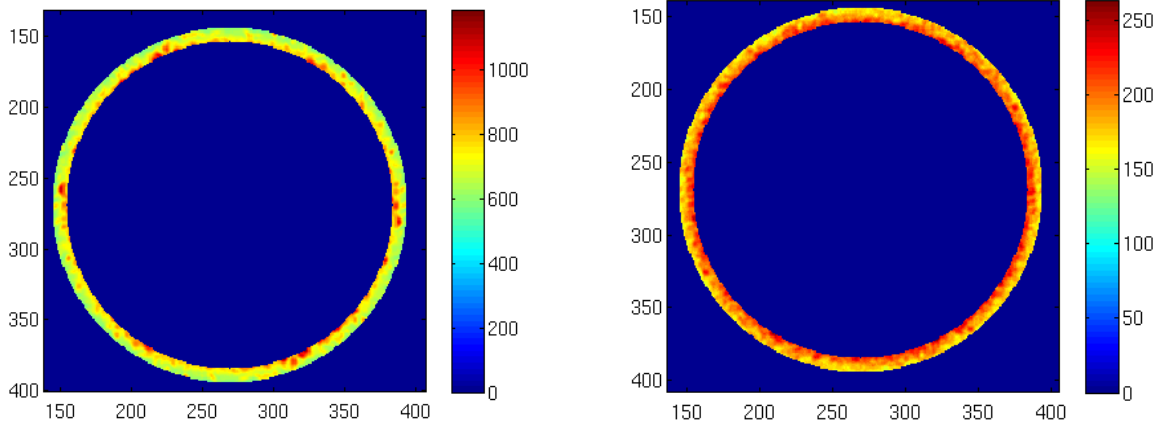


FIG. 7 – Séparation moyenne : 120 pixels, ie 1''/5

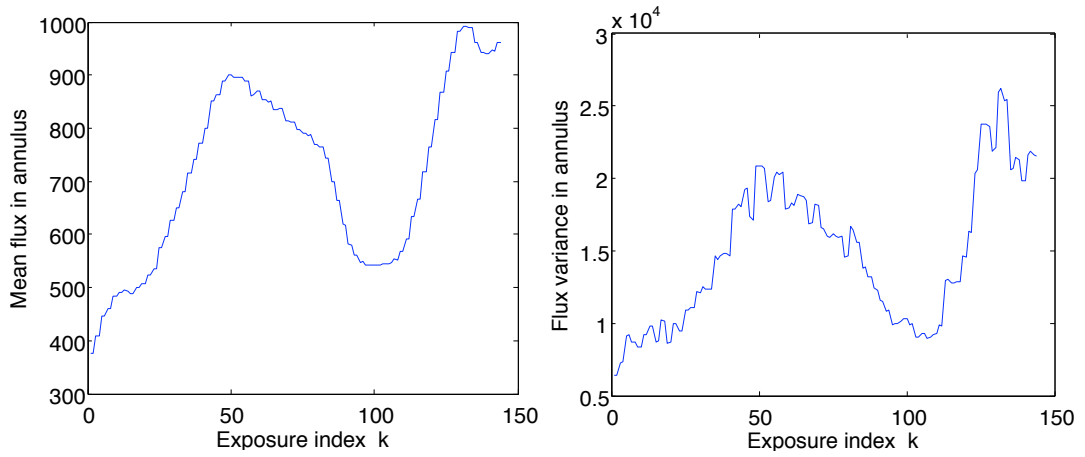


FIG. 8 – Séparation moyenne : 120 pixels, ie 1''/5

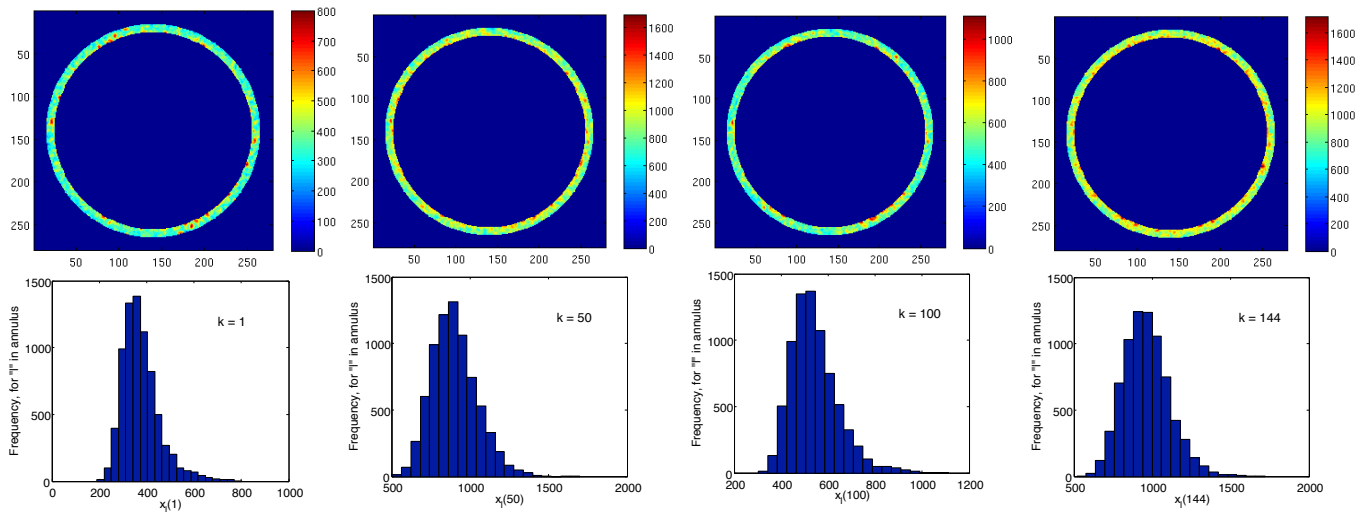


FIG. 9 – Séparation moyenne : 120 pixels, ie 1''/5



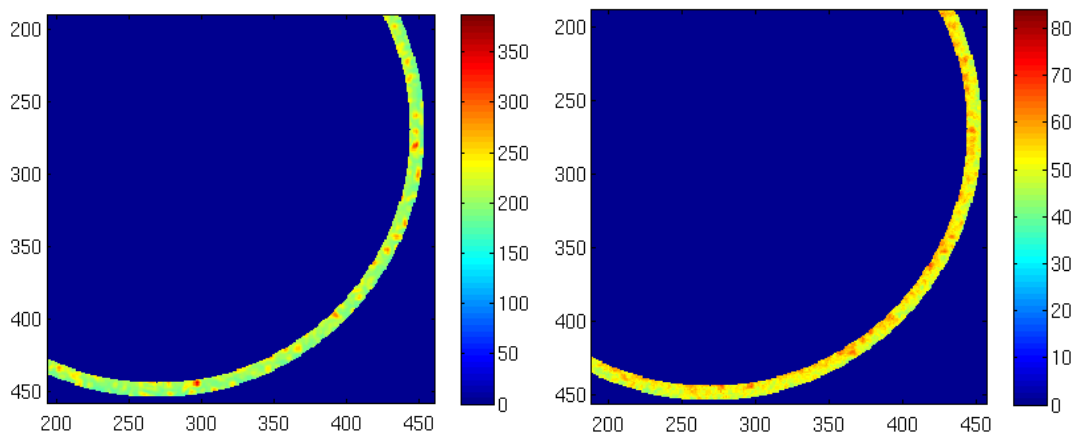


FIG. 10 – Séparation moyenne : 180 pixels, ie 2'/2

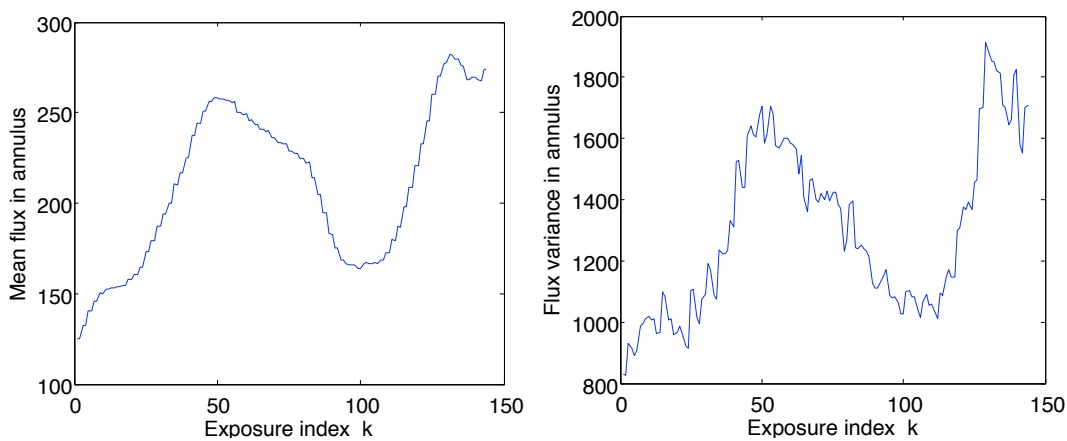


FIG. 11 – Séparation moyenne : 180 pixels, ie 2'/2

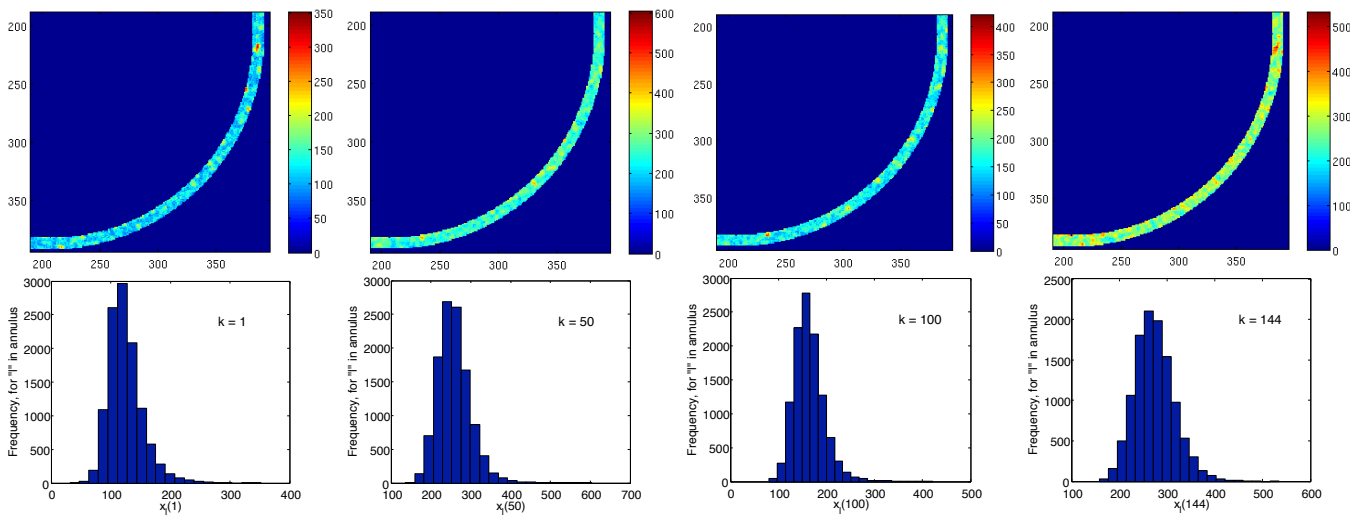


FIG. 12 – Séparation moyenne : 180 pixels, ie 2'/2

## Simulation de données pour une direction de vent différente des données de référence

La simulation physique fine décrite dans la section 2.1 est ici réappliquée mais pour une direction de vent non plus diagonale (direction vue dans le plan focal où est situé la CCD) mais horizontale. Les données résultantes sont illustrées dans la figure C.1 pour 4 poses distinctes. Qualitativement, à l'extrême centre de l'image on observe ainsi une intensité forte dans une bande horizontale alors que dans le jeu de données de référence illustré dans la figure 2.2 de la section 2.1 un tel profil était visible mais orienté selon la diagonale. Ceci confirme le constat de la section 2.2 selon lequel le vent a un effet majeur sur les statistiques des données au centre de l'image et confirme l'idée que la direction du vent pourrait *éventuellement* servir de paramètre physique permettant de définir une zone spatiale *i.i.d.* spatialement.

## C.2 Relation entre les PFA d'un test dans tout le champ et d'un test par pixel

Dans le modèle initial des données (3.2) la position initiale de la planète  $r$  est un paramètre inconnu. Si l'on sait qu'une planète est présente quelque part dans l'image,  $r$  peut ainsi être estimée, ce qui motive l'étude concernant les propriétés de l'estimateur par vraisemblance de  $r$  de la section 3.3.

Dans la phase de détection, la définition du test d'hypothèses de la section 3.4

$$H_0 : \alpha = 0 \quad H_1 : \alpha > 0 \quad (\text{C.1})$$

est aussi considérée dans un premier temps le test d'hypothèses relié directement au modèle (3.2), pour lequel  $r$  est ainsi indéterminé sous  $H_0$  et bien déterminé (mais inconnu) sous  $H_1$ . Un test consiste alors à seuiliser le GLR ou plus simplement  $\hat{\alpha}_{\text{ML}}$ , où les maxima de vraisemblance sont aussi calculés sur  $r$ . "Rejeter  $H_0$ " signifie alors implicitement "Décider qu'il n'y a de planète nulle part". Dans la suite nous considérerons par exemple, pour être concrets, la statistique de test  $\hat{\alpha}_{\text{ML}} \approx \max_r \hat{\alpha}(r)$  :

$$\text{Rejeter } H_0 \text{ si } \max_r \hat{\alpha}(r) > \zeta \quad (\text{C.2})$$

Cependant, comme il est trop compliqué de calculer la PFA de ce test, voir la section 3.4, nous avons réduit le test d'hypothèses à un test sur chaque pixel  $r$  :

$$(\text{pour } r \text{ fixé}) \quad H_0 : \alpha = 0 \quad H_1 : \alpha > 0 \quad (\text{C.3})$$

"Rejeter  $H_0$ " signifie alors implicitement "Décider qu'il n'y a pas de planète pour la position initiale  $r$ ". On considère par exemple le test

$$\text{Rejeter } H_0 \text{ (en } r) \text{ si } \hat{\alpha}(r) > \zeta \quad (\text{C.4})$$

Un jeu d'images est alors associé à une carte binaire de détection, associant à chaque position initiale  $r$  testée la valeur 1 si une planète a été détectée, 0 sinon.

$H_0$  au sens de (C.1) est donc équivalent à  $H_0(r) \forall r \in \llbracket 1, L \rrbracket$  au sens de (C.3) et  $H_1$  au sens de (C.1) est équivalent à  $\exists ! r \in \llbracket 1, L \rrbracket : H_1(r)$  au sens de (C.3) (et  $H_0(r)$  pour les autres  $r$ ).

Il est ainsi important de noter que la PFA n'a pas le même sens selon que le test est défini sur toute la zone de positions initiales possibles (PFA notée  $\text{PFA}_{im}$ ) ou sur chaque position initiale (PFA notée  $\text{PFA}_{pix}(r)$ ). On les définit respectivement par

$$\text{PFA}_{im} = \Pr(\max_r \hat{\alpha}(r) > \zeta | H_0) \quad (\text{C.5})$$

$$\text{PFA}_{pix}(r) = \Pr(\hat{\alpha}(r) > \zeta | H_0(r)) \quad (\text{C.6})$$

En tant que différentes fonctions d'un même  $\mathbf{x}$ , les statistiques  $\hat{\alpha}(r)$  pour les différents  $r$  sont des variables aléatoires *a priori* corrélées et non identiquement distribuées. Nous choisissons de tester tous

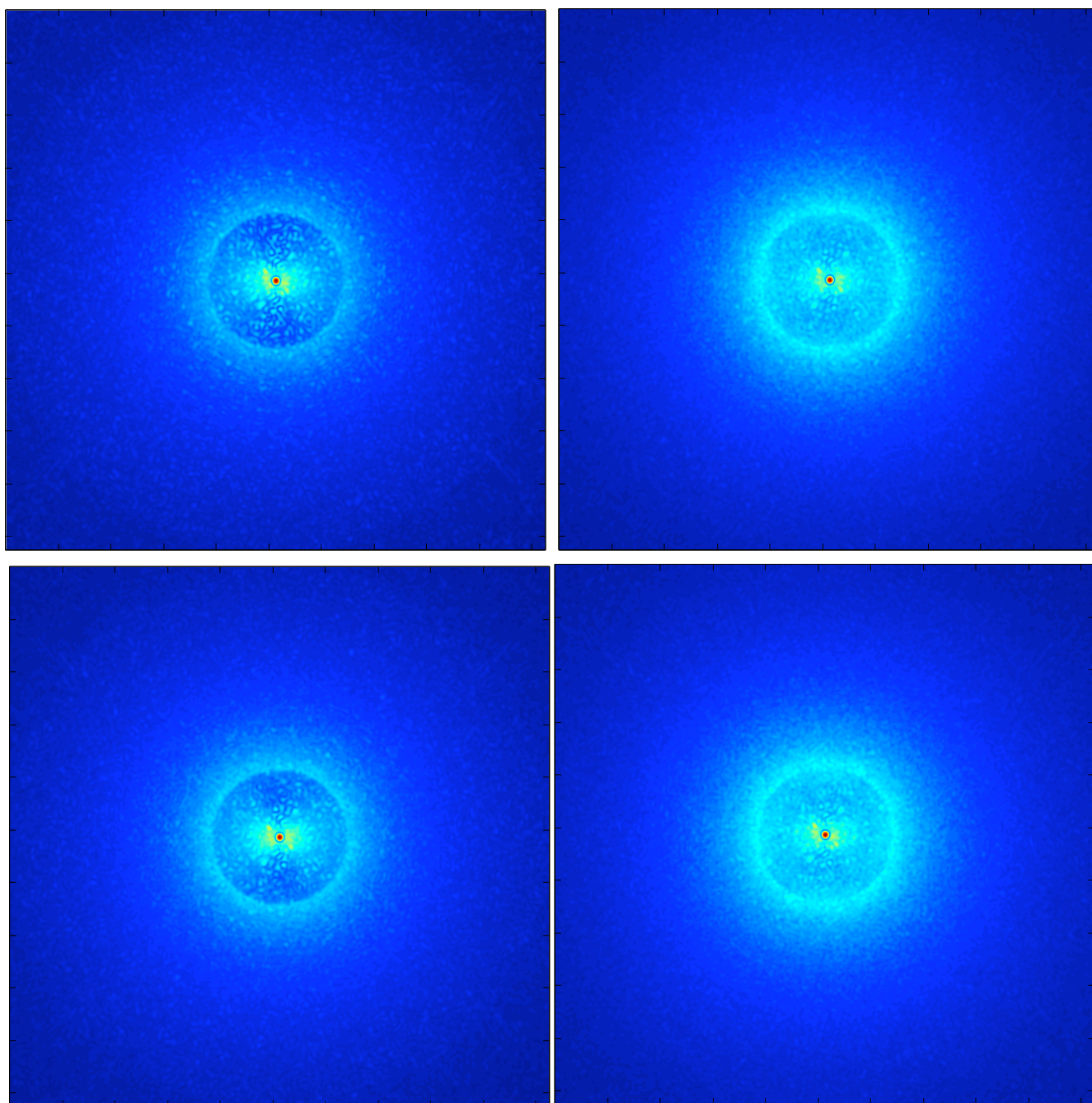


FIG. C.1: Illustration d'un jeu de données issu de la simulation physique fine décrite dans la section 2.1 mais ici réappliquée pour une direction de vent horizontale et non plus diagonale comme pour les données de référence illustrées dans la figure 2.2. Les poses  $k = 1$ ,  $k = 50$ ,  $k = 100$  et  $k = 144$  sont présentées dans le sens de la lecture, les images étant élevées à la puissance 0.2.

les pixels  $r \in \llbracket 1, L \rrbracket$  pour être concret.  $\text{PFA}_{im}$  est alors égale à 1 moins la distribution cumulative du maximum de l'échantillon  $\hat{\alpha}(1), \dots, \hat{\alpha}(L)$  alors que  $\text{PFA}_{pix}(r)$  est égale à 1 moins la distribution cumulative marginale de  $\hat{\alpha}(r)$ . Pour le modèle (3.2), comme  $\hat{\alpha}(r)$  est une fonction linéaire de  $x$  on a directement  $\hat{\alpha}(r) \sim \mathcal{N}(\mu(r), \Sigma(r))$  en général.

Dans le cas général, on peut noter que comme pour tout  $r \in \llbracket 1, L \rrbracket$  " $\hat{\alpha}(r) > \zeta$ "  $\Rightarrow$  " $\max_r \hat{\alpha}(r) > \zeta$ ", alors pour tout  $r \in \llbracket 1, L \rrbracket$   $\Pr(\hat{\alpha}(r) > \zeta | H_i) \leq \Pr(\max_r \hat{\alpha}(r) > \zeta | H_i)$  et donc

$$\text{PFA}_{im} \geq \max_r \text{PFA}_{pix}(r) \quad (\text{C.7})$$

$$\text{PD}_{im} \geq \max_r \text{PD}_{pix}(r) \quad (\text{C.8})$$

Ces relations peuvent être considérablement précisées selon que l'on étudie  $H_0$  ( $\text{PFA}_{im}$ ) ou  $H_1$  ( $\text{PD}_{im}$ ).

Etudions d'abord le cas sous  $H_0$ . Supposons que les  $\hat{\alpha}(r)$  sont i.i.d. Cette hypothèse est assez bien vérifiée dans notre cas, en tout cas hors zone totalement centrale, puisque  $\mathbb{E}[\hat{\alpha}] = \alpha = 0$  et  $\text{var}[\hat{\alpha}] = K\sigma^2 c_K^{-1}(r)$  et que  $c_K^{-1}(r)$  varie peu dans le champ, voir la section 4.1. Ainsi, prendre le maximum de PFA sur toutes les positions initiales  $r$  est identique à considérer la PFA de n'importe quel  $r$  :  $\text{PFA}_{pix}(r) = \text{PFA}_{pix}$  constant. On écrit alors en reprenant le calcul classique de distribution de statistique d'ordre :

$$\Pr(\max_r \hat{\alpha}(r) \leq \zeta) = \Pr(\hat{\alpha}(1) \leq \zeta \cap \hat{\alpha}(2) \leq \zeta \cap \dots \cap \hat{\alpha}(L) \leq \zeta) = \Pr(\hat{\alpha}(r) \leq \zeta)^L \quad \text{quel que soit } r$$

qui donne directement

$$\text{PFA}_{im} = 1 - (1 - \text{PFA}_{pix})^L \quad (\text{C.9})$$

Pour  $\text{PFA}_{pix} \ll 1$ ,

$$(1 - \text{PFA}_{pix})^L = 1 - L \times \text{PFA}_{pix} + \frac{L(L-1) \times \text{PFA}_{pix}^2}{2!} - \dots$$

et donc pour  $L \gg 1$  (mais  $L \times \text{PFA}_{pix} \ll 1$ ),

$$(1 - \text{PFA}_{pix})^L = 1 - L \times \text{PFA}_{pix} + \frac{L^2 \times \text{PFA}_{pix}^2}{2!} - \dots = \exp(-L \times \text{PFA}_{pix})$$

En tronquant le développement au 1er ordre on a simplement

$$\text{PFA}_{im} \approx L \times \text{PFA}_{pix} \quad (\text{C.10})$$

Rappelons que dans le cas identiquement distribué de ce calcul,  $\text{PFA}_{pix} = \max_r \text{PFA}_{pix}(r)$  ce qui rend l'inégalité (C.7) particulièrement faible.

Ainsi, si l'on souhaite par exemple fixer  $\text{PFA}_{im} = 5\%$ , c'est-à-dire détecter 5 planètes dans 100 jeux de données sans planète, si le nombre de pixels est  $L = 10^3$  il faut fixer  $\text{PFA}_{pix} = 5.10^{-5}$ , c'est-à-dire détecter 5 planètes dans 100 000 positions initiales sans planète. En réalité cette relation n'est pas complètement exacte car  $\hat{\alpha}(r)$  et  $\hat{\alpha}(r')$  sont en général corrélés pour des positions initiales proches, mais l'idée est qu'il faut prendre garde à ne pas choisir systématiquement " $\text{PFA} = 5\%$ " sans réfléchir au sens donné à cette PFA et en particulier qu'entre une détection pour un champ complet d'intérêt et une détection par position initiale les PFA ne sont pas du tout égales.

Sous  $H_1$  en revanche, il existe un unique  $r_0 \in \llbracket 1, L \rrbracket$  tel que  $H_1(r_0)$  au sens (C.3). L'hypothèse valide pour les autres positions initiales  $r$  est  $H_0(r)$ . Si l'estimateur de  $r_0$  par méthode ML est excellent (*i.e.*  $\mathbb{E}[\hat{r}_{\text{ML}}] = r_0$  et  $\text{var}[\hat{r}_{\text{ML}}] \approx 0$  à l'échelle des positions initiales testées), on peut approximer  $\hat{r}_{\text{ML}}$  par la valeur déterministe  $r_0$  de sorte que

$$\text{PD}_{im} \approx \Pr(\hat{\alpha}(r_0) > \zeta | H_1, r_0) \approx \text{PD}_{pix}(r_0) \quad (\text{C.11})$$

avec naturellement  $\text{PD}_{pix}(r_0) \approx \max_r \text{PD}_{pix}(r)$ .

L'Eq. (C.9) ou (C.10) et l'Eq. (C.11) précisent grandement les inégalités très peu contraignantes des Eqs. (C.7) et (C.8). Pour un nombre suffisant de pixels par rapport à la zone de corrélation d'une statistique de test, on peut ainsi garder à l'esprit les inégalités suivantes :

$$\text{PFA}_{im} \gg \max_r \text{PFA}_{pix}(r) \quad (\text{C.12})$$

$$\text{PD}_{im} \gtrsim \max_r \text{PD}_{pix}(r) \quad (\text{C.13})$$

Nous utiliserons d'autre part l'approximation linéaire de l'Eq. (C.10) afin de fixer un ordre de grandeur de  $\text{PFA}_{pix}$  après avoir fixé un  $\text{PFA}_{im}$ , plus intuitif, typiquement à une valeur de l'ordre de 5% dans la phase finale de détection.

### C.3 Images courte pose plus informatives qu'une longue pose

Il existe deux grandes façons d'obtenir une image correspondant à l'intégration d'une séquence temporelle. Soit on laisse le capteur intégrer "physiquement" pendant le temps souhaité, soit on acquiert plusieurs images successives que l'on somme "numériquement" ensuite. Avoir à disposition plus d'images semble intuitivement indiquer que l'on dispose d'une plus grande information. Cette idée est ici exacte (au bruit additif gaussien près) mais cela dépend en général des processus aléatoires et déterministes mis en jeu dans la comparaison de ces approches. On dit ici qu'une variable aléatoire est plus informative qu'une autre dans un sens très restrictif :  $X$  est plus informative que  $Y$  si la distribution d'une statistique de  $X$  est égale la distribution de  $Y$ .

Supposons ici correct le modèle *speckle* de la section 5.1. Plus généralement, supposons que l'intensité résulte d'une intégration dans le temps d'intensités indépendantes et que les données résultent ensuite d'une Poissonisation de cette intensité et de l'ajout de bruits gaussiens indépendants.

Soit  $i(t)$  l'intensité enregistrée pendant le temps  $\delta_t$  sur la CCD en la mini-pose  $t \in \llbracket 1, N_t \rrbracket$ . On choisit par simplicité  $\delta_t$  tel que deux poses  $i(t)$  et  $i(t')$ ,  $t' \neq t$  sont (approximativement) indépendantes. Le fait que le modèle soit monovarié ou multivarié ne change rien au problème donc on suppose que l'on étudie le pixel  $\ell$ , gardé implicite (on note par exemple  $i(t)$  à la place de  $i_\ell(t)$ ).

La première approche consiste à enregistrer une unique pose longue, en laissant ainsi le temps passer avant de lire les électrons de la CCD. On ne considère pas ici les effets de corrélation temporelle partiels. La modélisation physique de la conversion de l'intensité en photo-électrons fait appel à des processus quantiques de physique des solides complexes. Supposons qu'elle peut approximativement s'exprimer par une intégration déterministe qui a lieu soit avant le processus de Poissonisation soit après. Conditionnellement aux  $\{i(t), t \in \llbracket 1, N_t \rrbracket\}$ , la distribution du nombre total de photo-électrons le temps de la longue pose est donnée pour les deux modélisations physiques par :

$$\begin{aligned} \text{somme avant Poissonisation : } n|i &\sim \mathcal{P}(i) & \text{avec } i &= \sum_{t=1}^{N_t} i(t) \\ \text{somme après Poissonisation : } n &= \sum_{t=1}^{N_t} n(t) & \text{avec } n(t)|i(t) &\sim \mathcal{P}(i(t)) \end{aligned}$$

avec  $n_\ell|i$  et  $n_{\ell'}|i$  indépendants pour tous pixels  $\ell$  et  $\ell' \neq \ell$ .

Les MGF respectives de  $n$  sous les deux approches sont alors données à partir de celle de l'intensité instantanée  $i(t)$  par :

$$\begin{aligned} \text{somme avant Poissonisation : } g_n(z) &= h_i(z-1) = \mathbb{E}[e^{(z-1)\sum_t i(t)}] = \prod_t h_{i(t)}(z-1) \\ \text{somme après Poissonisation : } g_n(z) &= \mathbb{E}[z^{\sum_t n(t)}] = \prod_t \mathbb{E}[z^{n(t)}] = \prod_t h_{i(t)}(z-1) \end{aligned}$$

Statistiquement, les modèles “intensité intégrée avant Poissonisation” et “intensité Poissonisée au fur et à mesure” ne sont donc pas discernables.

Si la distribution d’intensité  $h_{i(t)}$  est stationnaire, la MGF du nombre d’électrons enregistrés lors d’une longue pose est donc donnée par

$$g_n(z) = (h_{i(t)}(z-1))^{N_t} \quad (\text{C.14})$$

On suppose alors que les données sont entachées d’un bruit additif gaussien indépendant du nombre de photo-électrons. On suppose de plus schématiquement qu’une partie de ce bruit gaussien ne s’accumule pas avec le temps (bruit de *lecture*) et qu’une autre partie s’accumule avec le temps (courant d’obscurité).

$$x^{(1)} = n + \epsilon_1 + \sum_{t=1}^{N_t} \epsilon_2(t) \quad \text{avec } \epsilon_1 \sim \mathcal{N}(0, \sigma_1^2), \quad \epsilon_2(t) \sim \mathcal{N}(0, \sigma_2^2)$$

La MGF d’une image enregistrée  $x^{(1)}$  acquise par longue pose est donc donnée par :

$$h_{x^{(1)}}(s) = \mathbb{E} \left[ e^{sx^{(1)}} \right] = e^{\frac{(\sigma_1^2 + N_t \sigma_2^2) s^2}{2}} g_n(e^s) = e^{\frac{\sigma_1^2 s^2}{2}} \left( e^{\frac{\sigma_2^2 s^2}{2}} h_{i(t)}(e^s - 1) \right)^{N_t}$$

La seconde approche consiste à obtenir une image  $x^{(2)}$  en sommant  $N_t$  images courte pose  $x^{(2)}(t)$ . Nous avons supposé la durée  $\delta_t$  d’une mini-pose telle que  $x^{(2)}(t)$  est indépendant de  $x^{(2)}(t')$  pour  $t \neq t'$ . On obtient alors suivant la même modélisation que précédemment :

$$h_{x^{(2)}}(s) = \mathbb{E} \left[ e^{s \sum_t x^{(2)}(t)} \right] = \left( \mathbb{E} \left[ e^{s x^{(2)}(t)} \right] \right)^{N_t} = \left( e^{\frac{(\sigma_1^2 + \sigma_2^2) s^2}{2}} g_{n(t)}(e^s) \right)^{N_t} = \left( e^{\frac{(\sigma_1^2 + \sigma_2^2) s^2}{2}} h_{i(t)}(e^s - 1) \right)^{N_t}$$

Ainsi on constate que les distributions statistiques d’une image obtenue par longue pose et d’une image obtenue par sommation d’images courte pose ont la même distribution, à un terme près dû au bruit additif gaussien à ajouter au moment de chaque pose (bruit de lecture typiquement) :

$$h_{x^{(2)}}(s) = e^{\frac{(N_t - 1) \sigma_1^2 s^2}{2}} h_{x^{(1)}}(s)$$

La somme des images courte pose est donc égale (en distribution) à l’image longue pose plus un bruit équivalent à  $(N_t - 1)$  bruits de lecture.

$$x^{(2)} = x^{(1)} + \epsilon \quad \text{avec } \epsilon \sim \mathcal{N}(0, (N_t - 1) \sigma_1^2) \quad (\text{C.15})$$

A ce bruit près, les images courte pose permettent d’obtenir l’image longue pose, de sorte que les images courte pose sont plus informatives que l’image longue pose. Mais il convient d’utiliser convenablement cette information, en n’utilisant en particulier pas d’approximation asymptotique de distribution qui supposerait le temps de pose très long (approximation qui mène par le théorème central limite à une distribution gaussienne en général).

## C.4 Discontinuité du PLR en $\zeta = \text{GLR}(\mathbf{x})$

Nous donnons ici une démonstration pour  $M > 1$  du théorème 2 de la section 8.1 selon lequel :

*S’il existe  $(\gamma_1, \dots, \gamma_M) \in \mathbb{R}_{+*}^M$ , tels que*

$$\sum_{m=1}^M \frac{1}{\gamma_m} < 1 \quad (\text{C.16})$$

et  $(\beta_1, \dots, \beta_M) \in \mathbb{R}_{+*}^M$  tels que pour tout  $\boldsymbol{\theta}$  suffisamment proche de  $\hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})$

$$GLR(\mathbf{x}) \leq LR(\mathbf{x}, \boldsymbol{\theta}) \leq GLR(\mathbf{x}) + \sum_{m=1}^M \beta_m |\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{ML}(\mathbf{x})|_m^{\gamma_m} \quad (\text{C.17})$$

alors la fonction  $\zeta \rightarrow PLR(\mathbf{x}, \zeta)$  a une dérivée infinie quand  $\zeta \rightarrow GLR(\mathbf{x})^+$ .

On procède ici par une récurrence simple sur les dimensions de  $\boldsymbol{\theta}$ . Appelons par simplicité

$$\begin{aligned} g_0 &= GLR(\mathbf{x}) \\ \hat{\boldsymbol{\theta}} &= \hat{\boldsymbol{\theta}}_{ML}(\mathbf{x}) \\ \boldsymbol{\delta} &= \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \quad \forall \boldsymbol{\theta} \in \Theta \\ T : \boldsymbol{\theta} &\rightarrow \boldsymbol{\delta} \\ g(\boldsymbol{\delta}) &= g_0 + \sum_{m=1}^M \beta_m |\delta_m|^{\gamma_m} \quad \forall \boldsymbol{\delta} \in T^{-1}(\Theta) \end{aligned}$$

où  $\beta_m > 0$  pour tout  $m \in \llbracket 1, M \rrbracket$  et les  $\gamma_m$  vérifient l'Eq. (C.16). L'existence du développement de l'Eq. (C.17) s'écrit alors

$$g_0 < LR(\boldsymbol{\delta}) \leq g(\boldsymbol{\delta}) \quad \forall \boldsymbol{\delta} \in T^{-1}(\Theta)$$

Comme dans la section 8.1, la densité de  $(g, \check{\boldsymbol{\delta}}_1)$  est donnée par

$$p_{g, \check{\boldsymbol{\delta}}_1}(\zeta, \check{\mathbf{u}}_1) = \sum_{k=1}^{n(\zeta, \check{\mathbf{u}}_1)} p_{\boldsymbol{\delta}}(\mathbf{u}^{(k)}) \left| \frac{\partial g}{\partial \delta_1}(\mathbf{u}^{(k)}) \right|^{-1}$$

où les  $\mathbf{u}^{(k)}$  sont les solutions de  $\phi(\mathbf{u}^{(k)}) = (\zeta, \check{\mathbf{u}}_1)$  où  $\phi : \boldsymbol{\delta} \rightarrow (g, \check{\boldsymbol{\delta}}_1)$ . Notant  $\phi_1$  une restriction de  $\phi$  telle que  $\phi_1$  est bijective,

$$p_{g, \check{\boldsymbol{\delta}}_1}(\zeta, \check{\mathbf{u}}_1) \geq p_{\boldsymbol{\delta}}(\phi_1^{-1}(\zeta, \check{\mathbf{u}}_1)) \left| \frac{\partial g}{\partial \delta_1}(\phi_1^{-1}(\zeta, \check{\mathbf{u}}_1)) \right|^{-1}$$

Or

$$p_{\boldsymbol{\delta}}(\phi_1^{-1}(\zeta, \check{\mathbf{u}}_1)) > A > 0$$

De plus,

$$\begin{aligned} \left| \frac{\partial g}{\partial \delta_1}(\boldsymbol{\delta}) \right| &= \gamma_1 \delta_1^{\gamma_1 - 1} \quad \forall \boldsymbol{\delta} \in \mathbb{R}_{+*}^M \\ \text{qui implique} \quad \left| \frac{\partial g}{\partial \delta_1}(\phi_1^{-1}(\zeta, \check{\mathbf{u}}_1)) \right| &= \gamma_1 \left( \zeta - g_0 - \sum_{m=2}^M u_m^{\gamma_m} \right)^{\frac{\gamma_1 - 1}{\gamma_1}} \\ \text{Or } u_2 > 0 \Rightarrow \zeta - g_0 - \sum_{m=2}^M u_m^{\gamma_m} &< \zeta - g_0 - \sum_{m=3}^M u_m^{\gamma_m} \\ \text{et comme } \gamma_1 > 1, \left| \frac{\partial g}{\partial \delta_1}(\phi_1^{-1}(\zeta, \check{\mathbf{u}}_1)) \right| &< \gamma_1 \left( \zeta - g_0 - \sum_{m=3}^M u_m^{\gamma_m} \right)^{\frac{\gamma_1 - 1}{\gamma_1}} \end{aligned} \quad (\text{C.18})$$

Ainsi,

$$p_{g, \check{\boldsymbol{\delta}}_1}(\zeta, \check{\mathbf{u}}_1) > \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=3}^M u_m^{\gamma_m} \right)^{\frac{1 - \gamma_1}{\gamma_1}}$$

Eq. (C.18) est définie pour  $u_2 > 0$  de telle sorte que

$$0 < \zeta - g_0 - \sum_{m=2}^M u_m^{\gamma_m}$$

ie  $u_2 < u_2^{max}$  avec

$$u_2^{max} = \left( \zeta - g_0 - \sum_{m=3}^M u_m^{\gamma_m} \right)^{1/\alpha_2}$$

Donc

$$\begin{aligned} p_{g,\delta_3,\dots,\delta_M}(\zeta, u_3, \dots, u_M) &= \int_0^{u_2^{max}} du_2 p_{g,\delta_2,\dots,\delta_M}(\zeta, u_2, \dots, u_M) \\ &> \int_0^{u_2^{max}} du_2 \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=3}^M u_m^{\gamma_m} \right)^{\frac{1-\gamma_1}{\gamma_1}} = u_2^{max} \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=3}^M u_m^{\gamma_m} \right)^{\frac{1-\gamma_1}{\gamma_1}} \\ &> \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=3}^M u_m^{\gamma_m} \right)^{-1 + \frac{1}{\gamma_1} + \frac{1}{\alpha_2}} \end{aligned}$$

Supposons que pour  $1 < i < M$

$$p_{g,\delta_i,\delta_{i+1},\dots,\delta_M}(\zeta, u_i, u_{i+1}, \dots, u_M) > \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=i+1}^M u_m^{\gamma_m} \right)^{-1 + \sum_{m=1}^{i-1} \frac{1}{\gamma_m}} \quad (C.19)$$

$$\text{with } 0 < u_i < u_i^{max} \text{ with } u_i^{max} = \left( \zeta - g_0 - \sum_{m=i+1}^M u_m^{\alpha_m} \right)^{1/\alpha_i} \quad (C.20)$$

Alors,

$$\begin{aligned} p_{g,\delta_{i+1},\dots,\delta_M}(\zeta, u_{i+1}, \dots, u_M) &= \int_0^{u_i^{max}} du_i p_{g,\delta_i,\dots,\delta_M}(\zeta, u_i, \dots, u_M) \\ &> \int_0^{u_i^{max}} du_i \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=i+1}^M u_m^{\gamma_m} \right)^{-1 + \sum_{m=1}^{i-1} \frac{1}{\gamma_m}} \\ &> u_i^{max} \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=i+1}^M u_m^{\gamma_m} \right)^{-1 + \sum_{m=1}^{i-1} \frac{1}{\gamma_m}} \\ &> \frac{A}{\gamma_1} \left( \zeta - g_0 - \sum_{m=i+1}^M u_m^{\gamma_m} \right)^{-1 + \sum_{m=1}^i \frac{1}{\gamma_m}} \end{aligned}$$

Ceci est défini pour  $u_{i+1} > 0$  de telle sorte que

$$0 < \zeta - g_0 - \sum_{m=i+1}^M u_m^{\gamma_m}$$

ie  $u_{i+1} < u_{i+1}^{max}$  with

$$u_{i+1}^{max} = \left( \zeta - g_0 - \sum_{m=i+2}^M u_m^{\gamma_m} \right)^{1/\alpha_{i+1}}$$



Les hypothèses (C.19) et (C.20) sont vérifiées pour  $i = 2$  et impliquent les mêmes équations pour  $i + 1$ . Elles sont donc vraies pour tout  $2 \leq i \leq M$ . Alors,

$$p_{g, \delta_M}(\zeta, u_M) > \frac{A}{\gamma_1} (\zeta - g_0)^{-1 + \sum_{m=1}^{M-1} \frac{1}{\gamma_m}}$$

avec  $0 < u_M < (\zeta - g_0)^{1/\gamma_M}$

donc

$$p_g(\zeta) > \frac{A}{\gamma_1} (\zeta - g_0)^{-1 + \sum_{m=1}^M \frac{1}{\gamma_m}} \quad \text{pour } \zeta > g_0 \quad (\text{C.21})$$

Finalement,

$$\begin{aligned} \Pr(g(\boldsymbol{\theta}) \in (g_0, g_0 + \epsilon]) &= \int_{g_0}^{g_0 + \epsilon} d\zeta p_g(\zeta) \\ &> \int_{g_0}^{g_0 + \epsilon} d\zeta \frac{A}{\gamma_1} (\zeta - g_0)^{-1 + \sum_{m=1}^M \frac{1}{\gamma_m}} \end{aligned}$$

or  $\zeta < g_0 + \epsilon \Rightarrow (\zeta - g_0)^{-1 + \sum_{m=1}^M \frac{1}{\gamma_m}} > \epsilon^{-1 + \sum_{m=1}^M \frac{1}{\gamma_m}}$  ssi  $-1 + \sum_{m=1}^M \frac{1}{\gamma_m} < 0$

$$\begin{aligned} \text{donc } \frac{1}{\epsilon} \Pr(g(\boldsymbol{\theta}) \in (g_0, g_0 + \epsilon]) &> \frac{1}{\epsilon} \int_{g_0}^{g_0 + \epsilon} d\zeta \frac{A}{\gamma_1} \epsilon^{-1 + \sum_{m=1}^M \frac{1}{\gamma_m}} \\ &> \frac{A}{\gamma_1} \frac{\epsilon^{-1 + \sum_{m=1}^M \frac{1}{\gamma_m}}}{\epsilon} \\ &> \frac{A}{\alpha_1} \epsilon^{-1 + \sum_{m=1}^M \frac{1}{\gamma_m}} \end{aligned}$$

Ainsi,

$$\frac{1}{\epsilon} \Pr(g(\boldsymbol{\theta}) \in (g_0, g_0 + \epsilon]) \rightarrow \infty \quad \text{quand } \epsilon \rightarrow 0$$

mais comme  $g_0 \leq \text{LR}(\boldsymbol{\theta}) \leq g(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta}$  assez proche de  $\hat{\boldsymbol{\theta}}$

$$\text{alors } \frac{1}{\epsilon} \Pr(\text{LR}(\boldsymbol{\theta}) \in (g_0, g_0 + \epsilon]) \geq \frac{1}{\epsilon} \Pr(g(\boldsymbol{\theta}) \in (g_0, g_0 + \epsilon])$$

$$\text{et finalement } \frac{1}{\epsilon} \Pr(\text{LR}(\boldsymbol{\theta}) \in (g_0, g_0 + \epsilon]) \rightarrow \infty \quad \text{quand } \epsilon \rightarrow 0$$

# Bibliographie

- Claude Aime and Rémi Soummer. The effect of a coronagraph on the statistics of adaptive optics pinned speckles. In *EAS Publications Series*. EDP Sciences, 2004.
- Murray Aitkin. Posterior Bayes factors. *Journal of the Royal Statistical Society. Series B*, 1991.
- Murray Aitkin. The calibration of p-values, posterior Bayes factors and the AIC from the posterior distribution of the likelihood. *Statistics and Computing*, 1997.
- Murray Aitkin, Richard J. Boys, and Tom Chadwick. Bayesian point null hypothesis testing via the posterior likelihood ratio. *Statistics and Computing*, 2005.
- Murray Aitkin, Charles C. Liu, and Tom Chadwick. Bayesian model comparison and model averaging for small-area estimation. *Annals of Applied Statistics*, 2009.
- Hiroto Akaike. Information theory and the extension of the maximum likelihood principle. In *Proc. 2nd Int. Symp. Information Theory*, 1973.
- S. A. Aldosari and J.M.F Moura. Detection in sensor networks : The saddlepoint approximation. *IEEE Transactions on Signal Processing*, 55(1) :327–340, 2007.
- F. Allard, N. F. Allard, et al. K-H2 quasi-molecular absorption detected in the T-dwarf  $\epsilon$  Indi Ba. *Astronomy & Astrophysics*, 2007.
- Emmanuel Aller-Carpentier, Markus Kasper, Patrice Martinez, et al. High order test bench for extreme adaptive optics system optimization. In *SPIE Proceedings*, 2008.
- Theodore Anderson. *An introduction to multivariate statistical analysis*. Wiley, 3rd edition, 2003.
- G. Avila and K. Wirestrand. Field and pupil rotations for the VLT 8m unit telescopes. Technical report, European Southern Observatory, VLT report, 2005.
- Gibor Basri. Observations of brown dwarfs. *Annual Review of Astronomy and Astrophysics*, 2000.
- James Berger and Thomas Selke. Testing a point null hypothesis : the irreconcilability of P values and evidence (with discussion). *Journal of the American Statistical Association*, 1987.
- James Berger, B. Boukai, and Y. Wang. Unified frequentist and Bayesian testing of a precise hypothesis. *Statistical Science*, 1997.
- James O. Berger. Could Fisher, Jeffreys and Neyman have agreed on testing? (with discussion). *Statistical Science*, 2003.
- James O. Berger. *Statistical decision theory and Bayesian analysis*. Springer-Verlag, 2nd edition, 1985.
- James O. Berger and Mohan Delampady. Testing precise hypotheses (with discussion). *Statistical Science*, 1987.

- James O. Berger and Luis Pericchi. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 1996.
- James O. Berger and R. Wolpert. *The Likelihood principle*. Haywood, 2nd edition, 1988.
- James O. Berger, Liseo, and Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 1999.
- James O. Berger, José Bernardo, and Dongchu Sun. The formal definition of reference priors. *Annals of Statistics*, 2009.
- José Bernardo. Bayesian hypothesis testing : a reference approach. *International Statistical Review*, 2002.
- José Bernardo. Integrated objective Bayesian estimation and hypothesis testing. In *Bayesian Statistics 9*, 2010.
- J.-L. Beuzit, Markus Feldt, et al. SPHERE : a ‘Planet Finder’ instrument for the VLT. *Messenger*, 2006.
- Peter J. Bickel and Kjell A. Doksum. *Mathematical Statistics, basic ideas and selected topics*, volume 1. Prentice Hall, 2001.
- Patrick Billingsley. *Probability and measure*. John Wiley & sons, 3rd edition, 1995.
- Allan Birnbaum. On the foundation of statistical inference *Journal of the American Statistical Association*, 1962.
- E. Bloemhof. Anomalous intensity of pinned speckles at high adaptive correction. *Optics Letters*, 2004.
- Wagner Borges and Julio Stern. The rules of logic composition for the Bayesian epistemic e-values. *Logic journal of the IGPL*, 2007.
- David Brillinger. *Times series : data analysis and theory*. New York : Mc Graw-Hill, 1981.
- Ben Calderhead and Mark Girolami. Estimating Bayes factors via thermodynamic integration and population MCMC. *Computational Statistics & Data Analysis*, 2009.
- Vidal Canales and Manuel Cagigal. Photon statistics in partially compensated wave fronts. *J. Opt. Soc. Am. A*, 1999.
- M. Carbillet, A. Boccaletti, Christian Thalmann, Thierry Fusco, Arthur Vigan, Isabelle Smith, David Mouillet, Kjetil Dohlen, Philippe Bendjoya, and André Ferrari. The software package SPHERE : a numerical tool for end-to-end simulations of the VLT instrument SPHERE. In N. Hubin et al., editors, *Adaptive Optics Systems*. SPIE, 2008.
- M. Carbillet, G. Desiderà, E. Augier, A. La Camera, A. Riccardi, A. Boccaletti, L. Jolissaint, and D. Ab Kadir. The CAOS problem-solving environment : recent developments. In B. Ellerbroek, M. Hart, N. Hubin, and P. Wizinowich, editors, *Adaptive Optics Systems II*, volume 7736. SPIE Proc., 2010.
- Bradley Carlin and Thomas Louis. *Bayes and empirical Bayes methods for data*. Chapman and Hall / CRC Press, 2000.
- George Casella and Roger L. Berger. *Statistical inference*. Duxbury, 2nd edition, 2002.
- George Casella and Roger L. Berger. Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *Journal of the American Statistical Association*, 1987.
- George Casella, Jiunn Hwang, and Christian Robert. A paradox in decision-theoretic interval estimation. *Statistica Sinica*, 1993.

- Ted Chang and Cesareo Villegas. On a theorem of Stein relating Bayesian and classical inferences in group models. *The Canadian Journal of Statistics*, 1986.
- David Charbonneau et al. A super-Earth transiting a nearby low-mass star. *Nature Letters*, 2009.
- Florent Chatelain, André Ferrari, and Jean-Yves Tourneret. Parameter estimation for multivariate mixed Poisson distributions. In *ICASSP*, 2006a.
- Florent Chatelain, Jean-Yves Tourneret, Jordi Inglada, and André Ferrari. Parameter estimation for multivariate Gamma distributions. In *EUSIPCO*, 2006b.
- Florent Chatelain, André Ferrari, and Jean-Yves Tourneret. Détection d'exoplanètes à l'aide de méthodes des moments. In *GRETSI*, 2007.
- G. Chauvin et al. Confirmation of a giant planet companion to the brown dwarf 2MASSWJ1207334. *Astronomy & Astrophysics*, 2005.
- Siddhartha Chib. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 1995.
- Siddhartha Chib and I. Jeliaskhov. Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, 2001.
- Yann Clenet, Markus Kasper, Nancy Ageorges, et al. NACO performance : status after 2 years of operation. In *SPIE Proceedings*, 2004.
- Arthur P. Dempster. On the difficulties inherent in Fisher's fiducial argument. *Journal of the American Statistical Association*, 1964.
- Arthur P. Dempster. The direct use of likelihood for significance testing. In *Proceedings of Conference on Foundational Questions in Statistical Inference*, 1974.
- Arthur P. Dempster. Commentary on the paper by Murray Aitkin, and on discussion by Mervyn Stone. *Statistics and Computing*, 1997.
- J. M. Dickey, D. V. Lindley, and S. J. Press. Bayesian estimation of the dispersion matrix of a multivariate normal distribution. *Communications in Statistics - Theory and Methods*, 14 :1019–1034, 1985.
- Kjetil Dohlen, Maud Langlois, Michel Saisse, et al. The Infra Red Dual Imaging and Spectrograph for SPHERE : design and performance. In *Proc. SPIE*, 2008.
- Morris Eaton. *Multivariate statistics*. Institute of Mathematical Statistics, 2007.
- Morris Eaton. *Group invariance applications in Statistics*. Regional Conf. Series in Prob. and Stat., 1989.
- Morris Eaton and William Sudderth. Group invariant inference and right Haar measure. *Journal of Statistical planning and inference*, 2002.
- Morris Eaton and William Sudderth. Consistency and strong inconsistency of group-invariant predictive inferences. *Bernoulli*, 1999.
- A. W. Edwards. *Likelihood*. Cambridge University Press, 1972.
- W. Edwards, H. Lindman, and L. Savage. Bayesian statistical inference for psychological research. *Psychological Review*, 1963.
- Michael Evans. Bayesian inference procedures derived via the concept of relative surprise. *Communications in Statistics*, 1997.

- Michael Evans and Gun Jang. Invariant p-values for model checking. *Annals of Statistics*, 2010.
- Herbert Federer. *Geometric measure theory*. Springer-Verlag, 1969.
- André Ferrari. Modeling for direct imaging of exoplanets data processing. In *EAS Publications Series*. EDP Sciences, 2006.
- André Ferrari, Gérard Letac, and Jean-Yves Tournet. Multivariate mixed Poisson distributions. In *EUSIPCO*, 2004.
- R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A.*, 1922.
- Ronald Fisher. *Statistical methods for research workers*. Oliver and Boyd, 14th edition, 1970, 1st ed. : 1925.
- Michael Fitzgerald and James Graham. Speckle statistics in adaptively corrected images. *The Astrophysical Journal*, 2006.
- D. A. S. Fraser. The fiducial method and invariance. *Biometrika*, 1961.
- Cesar Fuentes and Matthew Holman. A Subaru archival search for faint trans-Neptunian objects. *Astronomical Journal*, 2008.
- Thierry Fusco, Cyril Petit, Gérard Rousset, et al. Design of the extreme AO system for SPHERE, the planet finder instrument of the VLT. In *SPIE Proceedings*, 2006.
- Chris Gelino, Davy Kirkpatrick, and Adam Burgasser. Brown dwarfs catalog, 2010. URL [browndwarfs.org](http://browndwarfs.org).
- Gene H Golub and Charles F Van Loan. *Matrix computations*. Johns Hopkins series in the mathematical sciences. Johns Hopkins University Press, Baltimore, Md., 2nd ed. edition, 1989. ISBN 0-8018-3772-3.
- R. Gonsalves. Phase retrieval and diversity in adaptive optics. *Optical Engineering*, 1982.
- I. J. Good. *Good thinking : the foundations of probability and its applications*. University of Minnesota Press, 1983.
- J. Goodman. *Statistical Optics*. Wiley, 1985.
- Joseph Goodman. Statistical properties of laser speckle patterns. In *Topics in Applied Physics*. Springer Berlin / Heidelberg, 1975.
- J.W. Goodman. *Speckle Phenomena in Optics*. Roberts and Company Publishers, 2007.
- Constantinos Goutis and George Casella. Relationships between post-data accuracy measures. *Annals of the Institute of Statistical Mathematics*, 1997.
- Paul Halmos and Leonard Savage. Applications of the Radon-Nikodym theorem to the theory of sufficient statistics. *Annals of Mathematical Statistics*, 1949.
- Carl W Helstrom. *Elements of signal detection and estimation*. PTR Prentice Hall, Englewood Cliffs, N.J., 1995. ISBN 013808940X.
- Bruce Hoadley. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *Annals of Mathematical Statistics*, 1971.

- Jan Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy & Astrophysics Suppl.*, 1974.
- Jiunn Hwang, George Casella, Christian Robert, Martin Wells, and Roger Farrell. Estimation of accuracy in testing. *Annals of Statistics*, 1992.
- Jérôme Idier, Laurent Mugnier, and A. Blanc. Statistical behavior of joint least square estimation in the phase diversity context. *IEEE Trans. on Image Processing*, 2005.
- Harold Jeffreys. *Theory of probability*. Oxford University Press, 3rd edition, 1961, 1st ed. : 1939.
- Jens L. Jensen. *Saddlepoint Approximations*. Oxford University Press, 1995.
- David Jewitt. Kuiper belt, 2010. URL <http://www2.ess.ucla.edu/~jewitt/kb.html>.
- Robert Kass. The geometry of asymptotic inference. *Statistical Science*, 1989.
- Robert Kass and Adrian Raftery. Bayes factors. *Journal of the American Statistical Association*, 1995.
- Alexander Kechris. *Classical descriptive set theory*. Springer-Verlag, 1995.
- Samuel Kotz, N. Balakrishnan, and Norman L. Johnson. *Continuous Multivariate Distributions, Volume 1, Models and Applications, 2nd Edition*. Wiley-Interscience, 2000. ISBN 0-471-18387-3.
- Solomon Kullback. *Information theory and statistics*. Dover, 2nd edition, 1968.
- Antoine Labeyrie. Attainment of diffraction-limited resolution in large telescopes by Fourier-analyzing speckle patterns in star images. *Astronomy & Astrophysics*, 1970.
- David Lafrenière, Ray Jayawardhana, and Marten van Kerkwijk. Direct imaging and spectroscopy of a planetary mass candidate companion to a young solar analog. *The Astrophysical Journal Letter*, 2008.
- Erich L. Lehmann and George Casella. *Theory of point estimation*. Springer, 2nd edition, 1998.
- Erich L. Lehmann and Joseph P. Romano. *Testing statistical hypotheses*. Springer, 3rd edition, 2005.
- Guy Letac and H. Massam. A tutorial on noncentral Wishart distributions. Technical Report Available on request, Lab. Stat. Prob., 2004.
- Dennis Lindley. A statistical paradox. *Biometrika*, 1957.
- B. Macintosh, J. Graham, D. Palmer, R. Doyon, et al. The Gemini Planet Imager. In B. Ellerbroek and D. Bonaccini Calia, editors, *Advances in Adaptive Optics II*. SPIE, 2006.
- B. Macintosh, D. Palmer, et al. The Gemini planet imager : status and integration. In *SPIE Proceedings*, volume 7736, 2010.
- C. Marois, D. Lafrenière, R. Doyon, B. Macintosh, and D. Nadeau. Angular differential imaging : a powerful high-contrast imaging technique. *The Astrophysical Journal*, 641 :556–564, April 2006.
- D. Mattei, I. Smith, Andrea Ferrari, and Marcel Carillet. Parallelization of exoplanets detection algorithms based on field rotation ; example of the MOODS algorithm for SPHERE. In *In the Spirit of Lyot Proceedings*, 2010.
- Michel Mayor and Didier Queloz. A jupiter-mass companion to a solar-type star. *Nature*, 1995.
- Peter McCullagh. *Tensor Methods in Statistics*. Chapman and Hall, 1987.
- Xiao-Li Meng. Posterior predictive p-values. *Annals of Statistics*, 1994.

- E. Meulepas. A two-tailed p-value for Fisher's exact test. *Biometrical Journal*, 1998.
- C. Moutou et al. Science case for VLT-Planet Finder. In C. Aime and F. Vakili, editors, *IAU Colloq. 200 : Direct Imaging of Exoplanets : Science and Techniques*, 2006.
- Laurent Mugnier, A. Blanc, and Jérôme Idier. La diversité de phase : principe, applications, estimateurs. In *GdR ISiS*, 2009.
- L.M. Mugnier, J.-F. Sauvage, T. Fusco, and G. Rousset. Multi-channel planet detection algorithm for angular differential imaging. In B. Ellerbroek, editor, *Optical Society of America, Adaptive Optics : Analysis and Methods*, 2007.
- Robb J Muirhead. *Aspects of multivariate statistical theory*. Wiley series in probability and mathematical statistics. Wiley, New York, 1982. ISBN 0471094420.
- Leopoldo Nachbin. *The Haar integral*. Van Nostrand, 1965.
- Michael Newton and Adrian Raftery. Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society Series B*, 1994.
- Jerzy Neyman. Frequentist probability and frequentist statistics. *Synthese*, 1977.
- Jerzy Neyman and Egon Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London, Series A*, 1933.
- Dan Nicolae, Xiao-Li Meng, and Augustine Kong. Quantifying the fraction of missing information for hypothesis testing in statistical and genetic studies. *Statistical Science*, 2008.
- Hyun Oh and Anirban DasGupta. Comparison of the p-value and posterior probability. *Journal of Statistical planning and inference*, 1999.
- Anthony O'Hagan. Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society*, 1995.
- Anthony O'Hagan. Properties of intrinsic and fractional Bayes factors. *Test*, 1997.
- Marc Ollivier, Thérèse Encrenaz, Françoise Roques, Franck Selsis, and Fabienne Casoli. *Planetary systems : detection, formation and habitability of extrasolar planets*. Springer-Verlag, 2009.
- Giovanni Parmigiani and Lurdes Inoue. *Decision theory : principles and approaches*. Wiley, 2009.
- Carlos Pereira and Julio Stern. Evidence and credibility : full Bayesian significance test for precise hypotheses. *Entropy*, 1999.
- Michael Perryman. Planet detection methods, 2007. URL <http://www.rssd.esa.int/SA-general/Projects/Staff/perryman/planet-figur%e.pdf>.
- Kaare Petersen and Michael Pedersen. *The Matrix Cookbook*. matrixcookbook.com, 2007.
- Johann Pfanzagl. A characterization of the one parameter exponential family by existence of uniformly most powerful tests. *Sankhya : the Indian Journal of Statistics*, 1968.
- A.N. Philippou and G.G. Roussas. Asymptotic normality of the maximum likelihood estimate in the independent not identically distributed case. *Annals of the Institute of Statistical Mathematics*, 27(1) : 45–55, 1975.
- Bernard Picinbono. *Random Signals and Systems*. Pearson US Imports & PHIPES, 1993. ISBN 013015900X.

- Didier Queloz et al. The CoRoT-7 planetary system : two orbiting Super-Earths. *Astronomy & Astrophysics*, 2009.
- René Racine, Gordon Walker, et al. Speckle noise and the detection of faint companions. *Publ. of the Astron. Soc. of the Pacific*, 1999.
- Adrian E. Raftery, Michael A. Newton, Jaya M. Satagopan, and Pavel N. Krivitsky. Estimating the integrated likelihood via posterior simulation using the harmonic mean identity. *Bayesian statistics*, 2007.
- Kenneth Rice, Thomas Lumley, and Adam Szpiro. Trading bias for precision : decision theory for intervals and sets. *UW Biostatistics Working Paper Series*, 2008.
- L. J. Richardson et al. A spectrum of an extrasolar planet. *Nature Letters*, 2007.
- Christian Robert and George Casella. Distance weighted losses for testing and confidence set evaluation. *Test*, 1994.
- Christian P. Robert. *The Bayesian choice*. Springer, 2nd edition, 2007.
- Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, 1999.
- Richard Royall. The effect of sample size on the meaning of significance tests. *American Statistician*, 1986.
- Richard Royall. *Statistical evidence : a likelihood paradigm*. Chapman and Hall / CRC Press, 1997.
- Donald Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*, 1984.
- Andrew Rukhin. Loss functions for loss estimation. *Annals of Statistics*, 1988.
- Mark Schervish. A general method for comparing probability assessors. *Annals of Statistics*, 1989.
- J. Schneider. Extrasolar planets encyclopaedia, 2010. URL <http://exoplanet.eu/>.
- Sellke, Bayarri, and James O. Berger. Calibration of p-values for testing precise null hypotheses. *American Statistician*, 2001.
- Glenn Shafer. Lindley's paradox (with discussion). *Journal of the American Statistical Association*, 1982.
- I. Smith, A. Ferrari, and M. Carbillet. Detection of a moving source in speckle noise. Application to exoplanet detection. *IEEE Transactions on Signal Processing*, 57(3), 2009.
- Isabelle Smith, Marcel Carbillet, André Ferrari, David Mouillet, Anthony Boccaletti, and Kjetil Dohlen. Simulation of moving exoplanets detection using the VLT instrument SPHERE / IRDIS. In Norbert Hubin et al., editors, *Adaptive Optics Systems*, volume 7015. SPIE Proceedings, 2008.
- R. Soummer, A. Ferrari, C. Aime, and L. Jolissaint. Speckle noise and dynamic range in coronagraphic images. *The Astrophysical Journal*, 669 :642–656, nov 2007.
- Rémi Soummer and André Ferrari. The Strehl ratio in adaptive optics images : statistics and estimation. *Astrophysical Journal Letters*, 2007.
- Rémi Soummer et al. Apodized pupil Lyot coronagraphs for arbitrary apertures. ii. theoretical properties and application to extremely large telescopes. *Astrophysical Journal*, 2009.



- D. M. Stam. Spectropolarimetric signatures of Earth-like extrasolar planets. *Astronomy & Astrophysics*, 2008.
- Charles Stein. Approximation of improper prior measures by prior probability measures. In *Bernoulli, Bayes, Laplace Festschrift*. Springer-Verlag, 1965.
- Sweeting. Uniform asymptotic normality of the maximum likelihood estimator (and correction). *Annals of Statistics*, 1980.
- Jean-Yves Tournet, André Ferrari, and Gérard Letac. The noncentral Wishart distribution : properties and applications to speckle imaging. In *IEEE-SP Workshop Stat. Signal Processing*, 2005.
- C. Andy Tsao. A note on Lindley's paradox. *Test*, 2006.
- Tsui and Weerahandi. Generalized p-values in significance testing hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, 1989.
- Nicolas Vedrenne, Isabelle Smith, Alberto Cornia, Laurent Mugnier, and André Ferrari. Specification of signal extraction for IRDIS DBI. Technical report, VLT report, 2008.
- Arthur Vigan. *Détection et caractérisation des exoplanètes par imagerie directe avec IRDIS*. PhD thesis, Université de Provence (Aix-Marseille I), 2009.
- Arthur Vigan, Claire Moutou, Maud Langlois, France Allard, Anthony Boccaletti, Marcel Carbillet, David Mouillet, and Isabelle Smith. Photometric characterization of exoplanets using angular and spectral differential imaging. *Mon. Not. R. Astron. Soc.*, 2010a.
- Arthur Vigan, Claire Moutou, Maud Langlois, David Mouillet, Kjetil Dohlen, Anthony Boccaletti, Marcel Carbillet, Isabelle Smith, André Ferrari, Laurent Mugnier, and Christian Thalmann. Comparison of methods for detection and characterization of exoplanets with SPHERE/IRDIS. In *SPIE Proceedings*, 2010b.
- C. Villegas. Inner statistical inference II. *Annals of Statistics*, 1981.
- F. Wesley, J.J. Kavelaars, et al. The Kuiper belt luminosity function from  $m(R)=21$  to 26. *Icarus*, 195 (2) :827–843, 2008.
- F. Wildi, J.-L. Beuzit, M. Feldt, et al. The SPHERE exoplanet imager : Status report at PDR. In R. Tyson and M. Lloyd-Hart, editors, *SPIE Proceedings*, volume 6691. SPIE, September 2007.
- F. Wildi, J.-L. Beuzit, et al. SPHERE : the VLT planet imager in the post FDR phase. *SPIE*, 7440, 2009.
- Aleksander Wolszczan and Dale Frail. A planetary system around the millisecond pulsar PSR1257+2. *Nature*, 1992.
- James Zidek. A representation of Bayesian invariant procedures in terms of Haar measure. *Annals of the Institute of Statistical Mathematics*, 1969.