



HAL
open science

Analyse statistique de processus de poisson non homogènes. Traitement statistique d'un multidétecteur de particules

Jean-Pierre Lacombe

► **To cite this version:**

Jean-Pierre Lacombe. Analyse statistique de processus de poisson non homogènes. Traitement statistique d'un multidétecteur de particules. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1985. Français. NNT: . tel-00318875

HAL Id: tel-00318875

<https://theses.hal.science/tel-00318875>

Submitted on 5 Sep 2008

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.

THESE

présentée à

l' Université Scientifique et Médicale de Grenoble

pour obtenir le grade de
DOCTEUR DE 3ème CYCLE
«Mathématiques appliquées»

par

Jean-Pierre LACOMBE



- ANALYSE STATISTIQUE DE PROCESSUS DE POISSON
NON HOMOGENES

- TRAITEMENT STATISTIQUE D'UN MULTIDECTEUR
DE PARTICULES.



Thèse soutenue le 19 décembre 1985 devant la commission d'examen.

A. LE BRETON	}	Président
B. VAN CUTSEM		Examineurs
A. ANTONIADIS		
A. FILHOL		
G. GREGOIRE		

UNIVERSITE SCIENTIFIQUE ET MEDICALE DE GRENOBLE

Année universitaire 1982-1983

Président de l'Université : M. TANCHE

MEMBRES DU CORPS ENSEIGNANT DE L'U.S.M.G.

(RANG A)

SAUF ENSEIGNANTS EN MEDECINE ET PHARMACIE

PROFESSEURS DE 1^{ère} CLASSE

ARNAUD Paul	Chimie organique
ARVIEU Robert	Physique nucléaire I.S.N.
AUBERT Guy	Physique C.N.R.S.
AYANT Yves	Physique approfondie
BARBIER Marie-Jeanne	Electrochimie
BARBIER Jean-Claude	Physique expérimentale C.N.R.S. (labo de magnétisme)
BARJON Robert	Physique nucléaire I.S.N.
BARNOUD Fernand	Biosynthèse de la cellulose-Biologie
BARRA Jean-René	Statistiques - Mathématiques appliquées
BELORISKY Elie	Physique
BENZAKEN Claude (M.)	Mathématiques pures
BERNARD Alain	Mathématiques pures
BERTRANDIAS Françoise	Mathématiques pures
BERTRANDIAS Jean-Paul	Mathématiques pures
BILLET Jean	Géographie
BONNIER Jean-Marie	Chimie générale
BOUCHEZ Robert	Physique nucléaire I.S.N.
BRAVARD Yves	Géographie
CARLIER Georges	Biologie végétale
CAUQUIS Georges	Chimie organique
CHIBON Pierre	Biologie animale
COLIN DE VERDIERE Yves	Mathématiques pures
CRABBE Pierre (détaché)	C.E.R.M.O.
CYROT Michel	Physique du solide
DAUMAS Max	Géographie
DEBELMAS Jacques	Géologie générale
DEGRANGE Charles	Zoologie
DELOBEL Claude (M.)	M.I.A.G. Mathématiques appliquées
DEPORTES Charles	Chimie minérale
DESRE Pierre	Electrochimie
DOLIQUE Jean-Michel	Physique des plasmas
DUCROS Pierre	Cristallographie
FONTAINE Jean-Marc	Mathématiques pures
GAGNAIRE Didier	Chimie physique

.../...

GASTINEL Noël	Analyse numérique - Mathématiques appliquées
GERBER Robert	Mathématiques pures
GERMAIN Jean-Pierre	Mécanique
GIRAUD Pierre	Géologie
IDELMAN Simon	Physiologie animale
JANIN Bernard	Géographie
JOLY Jean-René	Mathématiques pures
JULLIEN Pierre	Mathématiques appliquées
KAHANE André (détaché DAFCO)	Physique
KAHANE Josette	Physique
KOSZUL Jean-Louis	Mathématiques pures
KRAKOWIAK Sacha	Mathématiques appliquées
KUPTA Yvon	Mathématiques pures
LACAZE Albert	Thermodynamique
LAJZEROWICZ Jeannine	Physique
LAJZEROWICZ Joseph	Physique
LAURENT Pierre	Mathématiques appliquées
DE LEIRIS Joël	Biologie
LLIBOUTRY Louis	Géophysique
LOISEAUX Jean-Marie	Sciences nucléaires I.S.N.
LOUP Jean	Géographie
MACHE Régis	Physiologie végétale
MAYNARD Roger	Physique du solide
MICHEL Robert	Minéralogie et pétrographie (géologie)
MOZIERES Philippe	Spectrométrie - Physique
OMONT Alain	Astrophysique
OZENDA Paul	Botanique (biologie végétale)
PAYAN Jean-Jacques (détaché)	Mathématiques pures
PEBAY PEYROULA Jean-Claude	Physique
PERRIAUX Jacques	Géologie
PERRIER Guy	Géophysique
PIERRARD Jean-Marie	Mécanique
RASSAT André	Chimie systématique
RENARD Michel	Thermodynamique
RICHARD Lucien	Biologie végétale
RINAUDO Marguerite	Chimie CERMAV
SENGEL Philippe	Biologie animale
SERGERAERT Francis	Mathématiques pures
SOUTIF Michel	Physique
VAILLANT François	Zoologie
VALENTIN Jacques	Physique nucléaire I.S.N.
VAN CUTSEN Bernard	Mathématiques appliquées
VAUQUOIS Bernard	Mathématiques appliquées
VIALON Pierre	Géologie

PROFESSEURS DE 2^{ème} CLASSE

ADIBA Michel	Mathématiques pures
ARMAND Gilbert	Géographie

AURIAULT Jean-Louis	Mécanique
BEGUIN Claude (M.)	Chimie organique
BOEHLER Jean-Paul	Mécanique
BOITET Christian	Mathématiques appliquées
BORNAREL Jean	Physique
BRUN Gilbert	Biologie
CASTAING Bernard	Physique
CHARDON Michel	Géographie
COHENADDAD Jean-Pierre	Physique
DENEUVILLE Alain	Physique
DEPASSEL Roger	Mécanique des fluides
DOUCE Roland	Physiologie végétale
DUFRESNOY Alain	Mathématiques pures
GASPARD François	Physique
GAUTRON René	Chimie
GIDON Maurice	Géologie
GIGNOUX Claude (M.)	Sciences nucléaires I.S.N.
GUITTON Jacques	Chimie
HACQUES Gérard	Mathématiques appliquées
HERBIN Jacky	Géographie
HICTER Pierre	Chimie
JOSELEAU Jean-Paul	Biochimie
KERCKOVE Claude (M.)	Géologie
LE BRETON Alain	Mathématiques appliquées
LONGEQUEUE Nicole	Sciences nucléaires I.S.N.
LUCAS Robert	Physiques
LUNA Domingo	Mathématiques pures
MASCLE Georges	Géologie
NEMOZ Alain	Thermodynamique (CNRS - CRTBT)
OUDET Bruno	Mathématiques appliquées
PELMONT Jean	Biochimie
PERRIN Claude (M.)	Sciences nucléaires I.S.N.
PFISTER Jean-Claude (détaché)	Physique du solide
PIBOULE Michel	Géologie
PIERRE Jean-Louis	Chimie organique
RAYNAUD Hervé	Mathématiques appliquées
ROBERT Gilles	Mathématiques pures
ROBERT Jean-Bernard	Chimie physique
ROSSI André	Physiologie végétale
SAKAROVITCH Michel	Mathématiques appliquées
SARROT REYNAUD Jean	Géologie
SAXOD Raymond	Biologie animale
SOUTIF Jeanne	Physique
SCHOOL Pierre-Claude	Mathématiques appliquées
STUTZ Pierre	Mécanique
SUBRA Robert	Chimie
VIDAL Michel	Chimie organique
VIVIAN Robert	Géographie



Je tiens à exprimer ma reconnaissance à Monsieur LE BRETON qui a bien voulu me faire l'honneur de présider le jury.

Je remercie vivement Messieurs VAN CUTSEM, GREGOIRE et FILHOL qui ont accepté de juger ce travail et de participer au jury.

J'ai plaisir à reconnaître l'aide de Monsieur ANTONIADIS qui a dirigé cette thèse et dont les enseignements m'ont été précieux.

Melle MARTIN a assuré avec autant de soin que de rapidité la dactylographie du manuscrit ; je la remercie très sincèrement.



PREMIERE PARTIE

*ANALYSE STATISTIQUE DE PROCESSUS DE POISSON
SPATIAUX ET NON HOMOGENES*



TABLE DES MATIERES

INTRODUCTION

CHAPITRE I

- I.1 - Notations et définitions.
- I.2 - Définition de la version canonique du processus.

CHAPITRE II

- II.1 - Dérivée de Radon-Nikodym d'une loi de Poisson par rapport à une autre.
- II.2 - Définition d'un test d'hypothèses simples.

CHAPITRE III

- III.1 - Quelques valeurs caractéristiques de la statistique $\psi(\log(f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)})) \circ X$.
- III.2 - Loi asymptotique de la variable $\psi(\log(f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)}))$.

CHAPITRE IV

- IV.1 - Lemme
- IV.2 - Consistance du test
- IV.3 - Exemples

CHAPITRE V

- V.1 - Famille de mesures monotone
- V.2 - Famille de mesures exponentielle



INTRODUCTION

Le travail que nous présentons ici s'inscrit dans le cadre de l'étude statistique des processus de Poisson non homogènes. Il traite plus particulièrement du problème de tests d'hypothèses sur les intensités de tels processus.

Exposons tout d'abord qu'elle était notre motivation en commençant ce travail.

De nombreux auteurs ont établi la convergence de processus ponctuels vers une limite poissonnienne, et, ce faisant, nombre d'entre eux ont suggéré que certains processus ponctuels peuvent être représentés de manière satisfaisante par les processus de Poisson. On trouve de tels propos dans les articles de Goldman [9], Breiman [2] ...

Plus prosaïquement, le problème pratique auquel nous nous trouvons confrontés, et que nous exposons dans la partie appliquée de ce travail, consiste à traiter un enregistrement d'observations indépendantes effectué sur un domaine borné de \mathbb{R}^2 , observations qui semblent pouvoir être assimilées en première approximation à des réalisations de processus de Poisson.

Nous nous sommes alors tout naturellement proposés de mettre en place une procédure visant à tester les intensités des processus de Poisson non homogènes.

Dans [3], Mark Brown définit un tel test, mais se restreint aux processus à valeurs dans l'espace des mesures σ finies sur \mathbb{R} . Ce travail est une généralisation de l'étude effectuée par Brown aux processus de Poisson à valeurs dans l'ensemble des mesures de Radon, positives sur un espace localement compact à base dénombrable.

Décrivons maintenant comment nous allons exposer nos résultats.

Dans un premier chapitre, nous rappelons la définition de tels processus, et présentons les outils mathématiques nécessaires à cette étude ; c'est ainsi que nous serons amenés à traiter des fonctionnelles de Laplace et des versions canoniques associées aux processus de Poisson non homogènes. Cette partie probabiliste doit beaucoup aux travaux de J. Neveu [14] et de K. Krickeberg [12], auxquels nous renvoyons les lecteurs peu familiarisés avec la théorie des processus ponctuels.

Dans un deuxième chapitre, nous nous sommes attachés à définir un test de l'hypothèse H_0 "l'intensité du processus est la mesure λ_{θ_0} " contre l'hypothèse H_1 "l'intensité du processus est la mesure λ_{θ_1} ". Pour ce faire, nous recherchons des conditions nécessaires et suffisantes à l'existence de la dérivée de Radon - Nikodym d'une loi de Poisson relativement à une autre. Dès lors que nous sommes en mesure de donner l'expression de cette dérivée, le lemme de Neyman-Pearson permet de solutionner le problème de test d'hypothèses.

Nous consacrons le troisième chapitre de ce travail à l'étude du test de Neyman-Pearson. Nous donnons dans un premier temps des conditions assurant la normalité asymptotique de la statistique définissant le test, ceci afin de remédier dans une certaine mesure à l'impossibilité dans

laquelle nous nous trouvons de déterminer avec exactitude la loi de cette variable aléatoire. Nous abordons ensuite l'étude de la puissance du test, et montrons en particulier que, sous certaines hypothèses, ce test est consistant.

Enfin, l'objet du dernier chapitre est l'étude des processus de Poisson non homogènes ayant pour intensités des éléments d'une famille de mesures monotone ou exponentielle. Nous prouvons notamment que l'on peut définir dans ce cas précis un test d'hypothèses multiples à partir du test de Neyman-Pearson, et énonçons de nouvelles conditions pour lesquelles la statistique définissant le test est asymptotiquement normale, ainsi que d'autres assurant la consistance du test.



CHAPITRE 1

I.1 NOTATIONS ET DEFINITIONS :

Soit T un espace vectoriel localement compact, et à base dénombrable ; T est séparable, et peut être muni d'une distance d pour laquelle il est polonais. La tribu engendrée par les boules est alors la tribu borélienne, et on peut choisir la distance d de manière à ce que les boules soient relativement compactes.

Soient \mathcal{C} la tribu borélienne de T , et \mathfrak{B} l'anneau des boréliens bornés. Désignons par $\mathcal{M}(T, \mathcal{C})$ l'espace des mesures de Radon positives sur (T, \mathcal{C}) , c'est à dire l'ensemble des mesures positives sur \mathcal{C} qui sont finies sur \mathfrak{B} , et introduisons sur cet espace la tribu $\mathcal{B}(\mathcal{M})$, tribu engendrée par la famille des applications qui, à une mesure μ de $\mathcal{M}(T, \mathcal{C})$, associe la valeur $\mu(A)$ dans \mathbb{R}^+ , lorsque A parcourt \mathcal{C} .

On appelle processus ponctuel sur T défini sur l'espace probabilisé (Ω, \mathcal{A}, P) tout élément aléatoire X , défini sur (Ω, \mathcal{A}, P) , à valeurs dans $(\mathcal{M}(T, \mathcal{C}), \mathcal{B}(\mathcal{M}))$, tel que pour tout $\omega \in \Omega$, $X(\omega)$ est une mesure ponctuelle sur (T, \mathcal{C}) .

Les réalisations d'un tel processus sont, pour tout $\omega \in \Omega$, de la forme $X(\omega) = \sum_{j=1, n(\omega)} \delta_{t_j(\omega)}$, où $n(\omega)$ est un nombre aléatoire de points de T , où $(t_1(\omega), \dots, t_{n(\omega)}(\omega))$ est une suite de points de T , et où l'on note δ_t la mesure de Dirac au point t .

Définissons, pour tout borélien A appartenant à \mathcal{C} , la variable aléatoire N_A par la relation suivante :

$$\forall \omega \in \Omega, N_A(\omega) = \langle X(\omega), \mathbf{1}_A \rangle = \int_A d(\sum_{j=1, n(\omega)} \delta_{t_j(\omega)})$$

N_A ainsi définie est la variable de comptage du nombre de points $t_j(\omega)$ sur le borélien A .

Etant donné une mesure λ appartenant à $\mathcal{M}(T, \mathcal{C})$, un processus ponctuel X est dit processus de Poisson d'intensité λ si :

i) Pour tout borélien A appartenant à \mathcal{C} , la variable aléatoire N_A suit une loi de Poisson de paramètre $\lambda(A)$. (Si $\lambda(A) = \infty$, alors $N_A = \infty$ ps)

ii) Les variables aléatoires $N_{A_1} \dots N_{A_k}$, associées à toute suite finie de boréliens $A_1 \dots A_k$ deux à deux disjoints de \mathcal{C} , sont indépendantes.

La loi sur $(\mathcal{M}(T, \mathcal{C}), \mathcal{B}(\mathcal{M}))$ d'un tel processus étant univoquement déterminée par les propriétés i) et ii) (cf Neveu [14]) nous posons :

Definition 1.1.1 :

Pour toute mesure λ appartenant à $\mathcal{M}(T, \mathcal{E})$, on appelle loi de Poisson d'intensité λ sur $(\mathcal{M}(T, \mathcal{E}), \mathcal{G}(\mathcal{M}))$, la loi de probabilité P_λ d'un processus de Poisson X d'intensité λ sur T .

Définition 1.1.2 :

On appelle fonctionnelle de Laplace d'une loi P_λ sur $(\mathcal{M}(T, \mathcal{E}), \mathcal{G}(\mathcal{M}))$ l'application φ définie par $\varphi(f) = \int_{\mathcal{M}(T, \mathcal{E})} e^{-\langle \mu, f \rangle} P_\lambda(d\mu)$, où $\langle \mu, f \rangle = \int f d\mu$, f parcourant le cône des fonctions boréliennes positives sur T .

Les problèmes que nous nous proposons d'étudier sont définis sur la structure statistique $(\mathcal{M}(T, \mathcal{E}), \mathcal{G}(\mathcal{M}), P_{\lambda_\theta}; \theta \in \Theta)$, où $(\lambda_\theta, \theta \in \Theta)$ est une famille de mesures σ finies sur \mathcal{E} , dominée par une mesure ν σ finie, et paramétrée par Θ .

Cela suppose en particulier que toute mesure λ_θ est absolument continue par rapport à ν . Il existe alors une fonction f_θ , unique à une ν équivalence près, telle que, pour tout borélien A appartenant à \mathcal{E} , on ait $\lambda_\theta(A) = \int_A f_\theta d\nu$. De plus, f_θ est ν_{pp} finie puisque ν est σ finie. (cf [15]).

Pour tout $r > 0$, on notera (T_r, \mathcal{E}_r) l'espace mesurable trace de (T, \mathcal{E}) avec la boule centrée de rayon r , et $P_{\lambda_\theta}(r)$ la loi d'un processus de Poisson sur (T_r, \mathcal{E}_r) d'intensité la restriction de λ_θ sur \mathcal{E}_r . On désignera par $\lambda_\theta^{(r)}$ cette intensité, et par $f_\theta^{(r)}$ un représentant de la dérivée de

Radon-Nikodym de $\lambda_\theta^{(r)}$ relativement à ν .

Avant de procéder à l'étude de tests d'hypothèses, on se propose d'envisager le problème de discrimination de deux processus de Poisson. Pour ce faire, nous allons définir une version canonique du processus sur T_r pour laquelle les calculs seront plus aisés.

I.2. DEFINITIONS DE LA VERSION CANONIQUE DU PROCESSUS X SUR

T_r :

Soient T_r^n l'espace produit des points $t = (t_1, \dots, t_n)$, où $t_i \in T_r$, $\mathcal{C}_r^{\otimes n}$ la tribu produit des tribus \mathcal{C}_r , et $\lambda_\theta^{\otimes n}$ la mesure produit des mesures λ_θ sur $(T_r^n, \mathcal{C}_r^{\otimes n})$.

Prenons pour Ω' l'ensemble des suites finies ordonnées d'éléments de T_r , et notons ω'_0 la suite vide ; pour tout $n \geq 0$, T_r^n est inclus dans Ω' , et $T_r^0 = \{\omega'_0\}$.

Définissons \mathcal{A}' de la manière suivante :

$$\mathcal{A}' = \{ F \subset \Omega' ; F \cap T_r^n \in \mathcal{C}_r^{\otimes n}, n \geq 0 \}$$

\mathcal{A}' est une tribu de parties de Ω' . Donnons nous alors sur l'espace probabillisable (Ω', \mathcal{A}') la probabilité $Q\lambda_\theta$ telle que :

$$\forall F \in \mathcal{A}', Q\lambda_\theta(F) = e^{-\lambda\theta(T_r)} \sum_{n \geq 0} (\lambda_\theta^{\otimes n}(F \cap T_r^n) / n!)$$

$$\text{avec } \lambda^{\otimes 0} = \delta_{\omega'_0}$$

A un élément ω' de Ω' différent de ω'_0 est associée une suite finie

$t_1(\omega'), \dots, t_n(\omega')(\omega')$ de points de T_r .

Soit alors l'application Y , de (Ω', \mathcal{A}') dans $\mathcal{M}(T_r, \mathcal{E}_r)$ définie par :

$$\forall \omega' \in \Omega', Y(\omega') = \sum_{j=1, n(\omega')} \delta_{t_j}(\omega')$$

$$Y(\omega'_0) = 0$$

Montrons que l'application Y est $\mathcal{A}' - \mathcal{E}(\mathcal{M})$ mesurable, ou, ce qui est équivalent ([15], p253), que les applications qui, à un élément ω' de Ω' , associe $\langle Y(\omega'), \mathbb{1}_A \rangle$ sont mesurables sur Ω' pour tout borélien A de \mathcal{E}_r .

$$\text{On a } \langle Y(\omega'), \mathbb{1}_A \rangle = \int_A d(\sum_{j=1, n(\omega')} \delta_{t_j}(\omega')) = N_A(\omega').$$

Soient alors deux entiers quelconques k et n , avec $n > 0$;

$$\text{Si } k > n, \text{ alors } N_A^{-1}((k)) \cap T_r^n = \emptyset$$

Si $k \leq n$, $N_A^{-1}((k)) \cap T_r^n$ est la réunion finie d'ensembles produits dont k facteurs sont égaux à A , les $n-k$ autres étant égaux à A^c .

Dans tous les cas, $N_A^{-1}((k)) \cap T_r^n \in \mathcal{E}_r^{\otimes n}$ et, par définition, $N_A^{-1}((k)) \in \mathcal{A}'$. Cela permet alors de conclure à la mesurabilité de Y .

Proposition 1.2.1 :

La loi image par Y de Q_{λ_θ} sur $\mathcal{M}(T_r, \mathcal{E}_r)$ est la même que celle de P par X , lorsque X est un processus d'intensité $\lambda_\theta^{(r)}$.

Démonstration :

Calculons l'expression de la fonctionnelle de Laplace de Y en une fonction f intégrable sur T_r .

$$\varphi_Y(f) = E(e^{-\langle Y, f \rangle}) = \int_{\Omega} e^{-\langle Y(\omega'), f \rangle} dQ_{\lambda_{\theta}}(\omega')$$

De par la définition même de Y , le dernier membre de cette égalité s'écrit encore :

$$\begin{aligned} & \sum_{n \geq 0} (e^{-\lambda_{\theta}(r)(T_r)} / n!) \int_{T_r^n} e^{-\sum_{j=1, n} f(t_j)} d\lambda_{\theta}^{\otimes n}(t_1, \dots, t_n) \\ &= e^{-\lambda_{\theta}(r)(T_r)} \sum_{n \geq 0} \left[\left(\int_{T_r} e^{-f(t)} d\lambda_{\theta}^{(r)}(t) \right) / n! \right]^n \\ &= \exp \int_{T_r} (\exp(-f) - 1) d\lambda_{\theta}^{(r)} \end{aligned}$$

La fonctionnelle de Laplace de Y en f est donnée par

$$\varphi_Y(f) = \exp \int_{T_r} (\exp(-f) - 1) d\lambda_{\theta}^{(r)}$$

Par ailleurs, on sait ([14], p 260) qu'un processus dont la fonctionnelle de Laplace est de cette forme est un processus de Poisson d'intensité $\lambda_{\theta}(r)$; la loi image de $Q_{\lambda_{\theta}}$ par Y sur $\mathcal{M}(T_r, \mathcal{E}_r)$ est donc $P_{\lambda_{\theta}}^{(r)}$.

CHAPITRE 2

DERIVEE DE RADON-NIKODYM D'UNE LOI DE POISSON PAR RAPPORT A UNE AUTRE - TEST DE NEYMAN-PEARSON.

Nous allons dans ce chapitre définir sur l'espace T_r la dérivée de Radon-Nikodym d'une loi de Poisson relativement à une autre. Une fois cette dérivée établie, il nous sera possible, à partir du lemme de Neyman-Pearson, de construire un test de l'hypothèse H_0 "l'intensité de X est $\lambda_{\theta_0}^{(r)}$ " contre l'hypothèse H_1 "l'intensité de X est $\lambda_{\theta_1}^{(r)}$ ".

II.1. DERIVEE DE RADON-NIKODYM D'UNE LOI DE POISSON PAR RAPPORT A UNE AUTRE.

Soient λ_{θ_0} et λ_{θ_1} deux mesures de $\mathcal{M}(T, \mathcal{E})$, de restrictions $\lambda_{\theta_0}^{(r)}$ et $\lambda_{\theta_1}^{(r)}$ sur \mathcal{E}_r .

Théorème II.1.1 : (KRICKEBERG)

Une condition nécessaire et suffisante pour que les lois de Poisson $P_{\lambda_{\theta_1}^{(r)}}$ et $P_{\lambda_{\theta_0}^{(r)}}$ soient équivalentes sur $(\mathcal{M}(T_r, \mathcal{E}_r), \mathcal{B}(\mathcal{M}))$ est que les

mesures λ_{θ_0} et λ_{θ_1} le soient sur (T_r, \mathcal{E}_r) ; sous cette hypothèse, la dérivée de Radon-Nikodym de $P_{\lambda_{\theta_1}}^{(r)}$ relativement à $P_{\lambda_{\theta_0}}^{(r)}$ est donnée par :

$$\begin{aligned} & (dP_{\lambda_{\theta_1}}^{(r)}/dP_{\lambda_{\theta_0}}^{(r)})(X(\omega)) \\ & = \exp\left(-\int_{T_r} (f_{\lambda}^{(r)} - 1) d\lambda_{\theta_0}^{(r)} + \Psi(\log f_{\lambda}^{(r)}) \circ X(\omega)\right) \end{aligned} \quad (1)$$

où $\Psi(f) \circ X(\omega) = \sum_{j=1, n(\omega)} f(t_j(\omega))$, et où $f_{\lambda}^{(r)} = [d\lambda_{\theta_1}^{(r)}/d\lambda_{\theta_0}^{(r)}]$

Démonstration :

Donnons tout d'abord la preuve de la première partie de ce théorème.

Une condition nécessaire et suffisante pour que les lois de Poisson $P_{\lambda_{\theta_0}}^{(r)}$ et $P_{\lambda_{\theta_1}}^{(r)}$ soient équivalentes est que les probabilités $Q\lambda_{\theta_0}$ et $Q\lambda_{\theta_1}$ le soient sur (Ω', \mathcal{A}') . De par la définition même de ces probabilités, il faut et il suffit, pour obtenir le résultat escompté, que les mesures $\lambda_{\theta_0}^{\otimes n}$ et $\lambda_{\theta_1}^{\otimes n}$ soient équivalentes sur $(T_r^n, \mathcal{E}_r^{\otimes n})$, ou encore que $\lambda_{\theta_0}^{(r)}$ et $\lambda_{\theta_1}^{(r)}$ le soient sur (T_r, \mathcal{E}_r) .

Supposons maintenant que la mesure λ_{θ_1} soit dominée par λ_{θ_0} , et établissons que, sous cette hypothèse, la dérivée de Radon-Nikodym de $P_{\lambda_{\theta_1}}^{(r)}$ relativement à $P_{\lambda_{\theta_0}}^{(r)}$ est donnée par l'expression (1).

Nous allons pour ce faire calculer la dérivée de Radon-Nikodym des lois associées à la version canonique du processus X , et nous nous ramènerons ensuite à ce processus.

Soit F un élément quelconque de \mathcal{A}' ; par définition de la probabilité $Q\lambda_{\theta_1}$, on a :

$$Q\lambda_{\theta_1}(F) = e^{-\lambda_{\theta_1}(r)(T_r)} \sum_{n \geq 0} (1/n!) \int_{F \cap T_r} \lambda_{\theta_1}^{(r) \otimes n} d\lambda_{\theta_1}^{(r) \otimes n}$$

La mesure $\lambda_{\theta_1}^{\otimes n}$ est dominée par $\lambda_{\theta_0}^{\otimes n}$, et on peut écrire encore $Q\lambda_{\theta_1}(F)$ comme :

$$\exp \left[- \int_{T_r} f_{\lambda}^{(r)} d\lambda_{\theta_0}^{(r)} \right] \sum_{n \geq 0} (1/n!) \int_{F \cap T_r} f_{\lambda}^{(r) \otimes n} d\lambda_{\theta_0}^{(r) \otimes n}$$

expression que l'on peut identifier à :

$$\exp \left[- \int_{T_r} (f_{\lambda}^{(r)} - 1) d\lambda_{\theta_0}^{(r)} \right] \int_F \prod_{j=1, n(\omega')} f_{\lambda}^{(r)}(t_j(\omega')) d\lambda_{\theta_0}(\omega').$$

Nous avons établi l'égalité suivante :

$$\int_F dQ\lambda_{\theta_1} =$$

$$\exp \left(- \int_{T_r} (f_{\lambda}^{(r)} - 1) d\lambda_{\theta_0}^{(r)} \right) \int_F \prod_{j=1, n(\omega')} f_{\lambda}^{(r)}(t_j(\omega')) dQ\lambda_{\theta_0}(\omega') \quad (2)$$

Puisque λ_{θ_1} est dominée par λ_{θ_0} , on a :

$$\lambda_{\theta_0} \{ x \in T_r, f_{\lambda}(x) = 0 \} = 0$$

L'égalité (2) permet alors d'établir la dérivée de Radon-Nikodym de $Q\lambda_{\theta_1}$ relativement à $Q\lambda_{\theta_0}$:

$$(dQ\lambda_{\theta_1}/dQ\lambda_{\theta_0})(\omega') = \exp \left(- \int_{T_r} (f_{\lambda}^{(r)} - 1) d\lambda_{\theta_0}^{(r)} \right) \exp(\Psi(\log f_{\lambda}^{(r)}) \circ \gamma(\omega'))$$

$$\text{où } \Psi(f) \circ \gamma(\omega') = \sum_{j=1, n(\omega')} f(t_j(\omega'))$$

En passant au processus X , il vient :

$$(dP_{\lambda_{\theta_1}}^{(r)}/dP_{\lambda_{\theta_0}}^{(r)})(X(\omega)) =$$

$$\exp \left(- \int_{T_r} (f_{\lambda}^{(r)} - 1) d\lambda_{\theta_0}^{(r)} \right) \exp(\Psi(\log f_{\lambda}^{(r)}) \circ X(\omega))$$

Nous allons dans le second paragraphe de ce chapitre définir, à partir du lemme de Neyman-Pearson, un test de l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_0}^{(r)}$ contre

l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_1}^{(r)}$.

II.2. DEFINITION D'UN TEST DE L'HYPOTHESE $\lambda_{\theta}^{(r)} = \lambda_{\theta_0}^{(r)}$ CONTRE L'HYPOTHESE $\lambda_{\theta}^{(r)} = \lambda_{\theta_1}^{(r)}$:

Soient λ_{θ_0} et λ_{θ_1} deux mesures de $\mathcal{M}(T, \mathcal{E})$, dominées par une mesure ν appartenant elle aussi à $\mathcal{M}(T, \mathcal{E})$; notons f_{θ_0} et f_{θ_1} les représentants des dérivées de Radon-Nikodym des mesures λ_{θ_0} et λ_{θ_1} par rapport à ν .

Les mesures λ_{θ_0} et λ_{θ_1} étant par hypothèse dominées par ν , le théorème II.1.1 permet d'affirmer que les lois de Poisson $P_{\lambda_{\theta_0}^{(r)}}$ et $P_{\lambda_{\theta_1}^{(r)}}$ sont dominées par $P_{\nu}^{(r)}$ sur $(\mathcal{M}(T_r, \mathcal{E}_r), \mathfrak{B}(\mathcal{M}))$.

Le lemme de Neyman-Pearson permet alors d'établir le résultat suivant :

Proposition II.2.1 :

Soit $\lambda_{\theta}^{(r)}$ l'intensité d'un processus de Poisson X , défini sur $(\Omega, \mathfrak{A}, P)$, à valeurs dans $(\mathcal{M}(T_r, \mathcal{E}_r), \mathfrak{B}(\mathcal{M}))$.

Un test de niveau α de l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_0}^{(r)}$ contre l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_1}^{(r)}$ est donné par la région critique :

$$R_C = \{ \omega, \Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X(\omega) > k \}, \text{ avec}$$

$$P_{\lambda_{\theta_0}^{(r)}}(R_C) = \alpha.$$

Démonstration :

Les lois de Poisson $P_{\lambda_{\theta_0}^{(r)}}$ et $P_{\lambda_{\theta_1}^{(r)}}$ sont dominées par $P_{\nu}^{(r)}$; leurs

dérivées de Radon-Nikodym sont données par :

$$\Phi_{\lambda_{\theta_0}}(X(\omega)) = \left(\frac{dP_{\lambda_{\theta_0}}(r)}{dP_{\nu}}(r) \right) (X(\omega)) = \\ \exp\left(-\int_{T_r} (f_{\theta_0}(r) - 1) d\nu + \Psi(\log f_{\theta_0}(r)) \circ X(\omega)\right)$$

$$\Phi_{\lambda_{\theta_1}}(X(\omega)) = \left(\frac{dP_{\lambda_{\theta_1}}(r)}{dP_{\nu}}(r) \right) (X(\omega)) = \\ \exp\left(-\int_{T_r} (f_{\theta_1}(r) - 1) d\nu + \Psi(\log f_{\theta_1}(r)) \circ X(\omega)\right)$$

En conséquence, le rapport de vraisemblance s'écrit :

$$\Phi_{\lambda_{\theta_1}}(X(\omega)) / \Phi_{\lambda_{\theta_0}}(X(\omega)) = \\ \exp\left(-\int_{T_r} (f_{\theta_1}(r) - f_{\theta_0}(r)) d\nu + \Psi(\log f_{\theta_1}(r) / f_{\theta_0}(r)) \circ X(\omega)\right)$$

Le lemme de Neyman-Pearson permet alors de définir la région critique du test comme étant :

$$R_c = \{\omega, \exp\left(-\int_{T_r} (f_{\theta_1}(r) - f_{\theta_0}(r)) d\nu + \Psi(\log f_{\theta_1}(r) / f_{\theta_0}(r)) \circ X(\omega)\right) > k'\}$$

où la constante k' est fixée par le niveau α du test.

Posons $k = \log(k' \exp(\int_{T_r} (f_{\theta_1}(r) - f_{\theta_0}(r)) d\nu))$; il vient :

$$R_c = \{\omega, \Psi(\log f_{\theta_1}(r) / f_{\theta_0}(r)) \circ X(\omega) > k\}$$

Avant de nous pencher plus avant sur les propriétés de ce test, ce qui fera l'objet d'un prochain chapitre, nous allons étudier certaines caractéristiques de la statistique $\Psi(\log f_{\theta_1}(r) / f_{\theta_0}(r)) \circ X(\omega)$.



CHAPITRE 3

ETUDE DE LA STATISTIQUE DEFINISSANT LE TEST DE NEYMAN-PEARSON

La loi de la variable aléatoire réelle $\Psi(\log (f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X$ est difficile à établir d'une manière générale ; c'est une des raisons qui font que nous nous attacherons, dans le second paragraphe de ce chapitre, à définir des conditions pour lesquelles cette variable est asymptotiquement normale. Nous allons consacrer un premier paragraphe à calculer certaines caractéristiques de la statistique $\Psi(\log (f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X$.

III.1. QUELQUES CARACTERISTIQUES DE LA VARIABLE

$\Psi(\log (f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X :$

Désignons par H_0 l'hypothèse selon laquelle l'intensité du processus de Poisson est $\lambda_{\theta_0}^{(r)}$, et par H_1 la contre hypothèse.

Calculons dans un premier temps la probabilité qu'a la variable aléatoire $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X$ de prendre la valeur $-\infty$ sous l'hypothèse H_0

$$P[\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X(\omega) = -\infty] =$$

$$P[\sum_{j=1, n(\omega)} \log(f_{\theta_1}^{(r)}(t_j(\omega)) / f_{\theta_0}^{(r)}(t_j(\omega))) = -\infty]$$

Puisque la mesure λ_{θ_0} est finie sur \mathcal{T}_r , la sommation s'effectue sur un nombre fini de termes, et $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X(\omega)$ prend la valeur $-\infty$ si et seulement si il existe un ω de Ω pour lequel $f_{\theta_1}^{(r)}(t_j(\omega)) = 0$.

Soit $A = \{x \in T_r, f_{\theta_1}^{(r)}(x) = 0\}$; on a :

$$P[\sum_{j=1, n(\omega)} \log(f_{\theta_1}^{(r)}(t_j(\omega)) / f_{\theta_0}^{(r)}(t_j(\omega))) = -\infty] =$$

$$= P_{\lambda_{\theta_0}^{(r)}}(\langle X(\omega), \mathbf{1}_A \rangle \neq 0)$$

$$= 1 - P_{\lambda_{\theta_0}^{(r)}}(N_A = 0) = 1 - \exp(-\lambda_{\theta_0}^{(r)}(A))$$

En particulier, si la mesure λ_{θ_0} est dominée par λ_{θ_1} , la statistique $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X$ ne prend la valeur $-\infty$ qu'avec une probabilité nulle $P_{\lambda_{\theta_0}^{(r)}}$.

Calculons de la même manière, sous l'hypothèse H_0 , la quantité

$$P[\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X = \infty].$$

$$\text{On a } P[\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X = \infty] = P_{\lambda_{\theta_0}^{(r)}}(\langle X(\omega), \mathbf{1}_{A'} \rangle \neq 0), \text{ où}$$

l'on désigne par A' l'ensemble suivant :

$$A' = \{x \in T_r, f_{\theta_0}^{(r)}(x) = 0\}.$$

L'ensemble A' étant naturellement de mesure $\lambda_{\theta_0}^{(r)}$ nulle, on a :

$$P[\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X = \infty] = 1 - \exp(-\lambda_{\theta_0}^{(r)}(A)) = 0$$

Nous pouvons alors formuler la remarque suivante :

Remarque III.1.1 :

Si la mesure $\lambda_{\theta_0}^{(r)}$ est dominée par $\lambda_{\theta_1}^{(r)}$, la statistique $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X$ est finie $P_{\lambda_{\theta_0}^{(r)}}$ p.p..

Soit $B = \{x \in T_r, f_{\theta_0}^{(r)}(x) = 0\}$

Un raisonnement en tous points analogue à celui que nous venons d'effectuer permet d'établir, sous l'hypothèse H_1 , les égalités :

$$P[\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X = \infty] = 1 - \exp(-\lambda_{\theta_1}^{(r)}(B))$$

$$P[\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X = -\infty] = 0$$

Il en résulte une nouvelle remarque :

Remarque III.1.2 :

Si la mesure $\lambda_{\theta_1}^{(r)}$ est dominée par $\lambda_{\theta_0}^{(r)}$, la statistique $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X$ est finie $P_{\lambda_{\theta_1}^{(r)}}$ p.p..

Un outil important pour la suite de ce travail est la transformée de Laplace de la statistique $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \circ X$, lorsque la loi de X est

$P_{\lambda_{\theta_0}^{(r)}}$. Nous allons donc dans ce paragraphe faire une place à ce calcul, ou, plus exactement, à l'étude de la transformée de Laplace de la variable $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \mathbb{1}_{A^c} \circ X$, nous restreignant ainsi à l'ensemble sur

lequel cette variable est finie $P_{\lambda_{\theta_0}^{(r)}}$ presque partout.

Pour toute valeur positive de u , on a :

$$E(\exp(-u \Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) \mathbb{1}_{A^c}) \circ X) =$$

$$\int_{\Omega} \exp(-\langle X(\omega), u \log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \mathbb{1}_{A^c} \rangle) dP(\omega).$$

En passant à la version canonique du processus X , cette expression devient, sous l'hypothèse H_0 :

$$\int_{\Omega'} \exp(-\langle Y(\omega'), u \log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \mathbb{1}_{A^c} \rangle) dQ_{\lambda_{\theta_0}}(\omega').$$

On reconnaît sous cette forme la fonctionnelle de Laplace de Y , calculée en la fonction $f(x) = u \log(f_{\theta_1}^{(r)}(x) / f_{\theta_0}^{(r)}(x)) \mathbb{1}_{A^c}(x)$. Celle-ci est bien définie, puisque $\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})$ est finie sur A^c , donc intégrable sur $T_r \cap A^c$.

La valeur de la fonctionnelle de Laplace de Y , calculée en cette fonction f , est donnée (cf. p10) par :

$$\begin{aligned} \varphi_Y(f) &= \exp(-\lambda_{\theta_0}(T_r)) \exp \int_{T_r} \exp(-u \log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \mathbb{1}_{A^c}) d\lambda_{\theta_0} \\ &= \exp(-\lambda_{\theta_0}(T_r)) \exp \int_{T_r} (\mathbb{1}_A + \exp(-u \log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \mathbb{1}_{A^c})) d\lambda_{\theta_0} \\ &= \exp \int_{A^c} (\exp(-u \log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) - 1) d\lambda_{\theta_0} \end{aligned}$$

Remarque III.1.3 :

La transformée de Laplace de la variable $\Psi(\log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \mathbb{1}_{A^c}) \circ X$ calculée en u est donnée, sous l'hypothèse H_0 par

$$\varphi(u) = \exp \int_{A^c} (\exp(-u \log(f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)})) - 1) d\lambda_{\theta_0}$$

III.2. LOI ASYMPTOTIQUE DE LA VARIABLE $\Psi(\log(f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)})) \circ X$:

Nous avons défini au chapitre II le test de Neyman-Pearson de l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_0}^{(r)}$ contre $\lambda_{\theta}^{(r)} = \lambda_{\theta_1}^{(r)}$ à partir de la statistique $\Psi(\log(f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)})) \circ X$. L'objet de ce paragraphe est d'établir des conditions pour lesquelles la loi de cette variable est asymptotiquement normale.

On désigne par :

$$A_c = \{x \in T, |f_{\theta_1}(x)/f_{\theta_0}(x) - 1| \leq c\}$$

$$S_{rM} = \{x \in T_r, |\log(f_{\theta_1}^{(r)}(x)/f_{\theta_0}^{(r)}(x))| \leq M\}$$

$$a_{rM} = \int_{S_{rM}} \log(f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)}) d\lambda_{\theta_0}^{(r)}$$

$$b_{rM}^2 = \int_{S_{rM}} (\log(f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)}))^2 f_{\theta_0} dv$$

Théorème III.2.1 :

Si la mesure λ_{θ_0} est dominée par λ_{θ_1} , s'il existe une valeur de M pour laquelle $\lambda_{\theta_0}(S_{rM}^c)$ est finie, et si l'une au moins des conditions i) ou ii) ci-après est vérifiée, alors la variable aléatoire réelle $(\Psi(\log(f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)})) \circ X - a_{rM}) / b_{rM}$ converge en loi, lorsque r tend vers l'infini, vers une variable dont la loi est normale centrée réduite.

$$i) \forall c > 0, \int_{A_c} (f_{\theta_1}/f_{\theta_0} - 1)^2 f_{\theta_0} dv = \infty$$

$$ii) \exists c > 0 \text{ tel que } \lambda_{\theta_0}(\{x, |\log(f_{\theta_1}(x)/f_{\theta_0}(x))| > c\}) = \infty$$

Démonstration

Nous avons établi la preuve de ce théorème selon la démarche suivante :

Nous démontrons dans un premier temps que la variable $\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}^c}) \circ X$ est finie P p.p., puis que b_{rM} tend vers l'infini lorsque r tend vers l'infini.

Cela permet d'obtenir ce premier résultat :

Montrer la convergence en loi de $(\Psi(\log(f_{\theta_1}/f_{\theta_0}) \circ X - a_{rM})/b_{rM}$ vers une variable aléatoire U suivant une loi normale centrée réduite revient à établir la convergence de la variable $(\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}}) \circ X - a_{rM})/b_{rM}$ vers U .

Arguant ensuite d'un résultat classique de la théorie des transformées de Laplace (cf. [7], p 408), nous prouverons, pour établir la convergence de ces variables, celle de leurs transformées de Laplace.

Montrons que $|\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}^c}) \circ X|$ est finie P p.p..

$$|\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}^c}) \circ X| =$$

$$|\sum_{j=1, n(\omega)} \log(f_{\theta_1}(t_j(\omega))/f_{\theta_0}(t_j(\omega))) \mathbb{1}_{S_{rM}^c}(t_j(\omega))|$$

Puisque $\lambda_{\theta_0}(S_{rM}^c)$ est finie, la sommation s'effectue sur un nombre fini de termes. Alors,

$$\{\omega, |\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}^c}) \circ X| = \infty\}$$

$$= \{\omega, f_{\theta_1}(t_j(\omega)) = 0\} \cup \{\omega, f_{\theta_0}(t_j(\omega)) = 0\}$$

En conséquence,

$$P(|\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}^c}) \circ X| = \infty)$$

$$\leq P \lambda_{\theta_0}(\langle \mathbb{1}_{\{f_{\theta_1}=0\}}, X(\omega) \rangle) + P \lambda_{\theta_0}(\langle \mathbb{1}_{\{f_{\theta_0}=0\}}, X(\omega) \rangle)$$

λ_{θ_0} étant dominée par λ_{θ_1} , $P \lambda_{\theta_0}(\langle \mathbb{1}_{\{f_{\theta_1}=0\}}, X(\omega) \rangle) = 0$. De plus,

$$P_{\lambda_{\theta_0}} (\langle \mathbb{1}_{(f_{\theta_0}=0)}, X(\omega) \rangle) = 1 - \exp(-\lambda_{\theta_0} (f_{\theta_0} = 0)) = 0$$

La variable $|\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}}) \circ X|$ est donc finie $P_{\lambda_{\theta_0}}$ p.p.

Etablissons maintenant que, si la condition i) du théorème est vérifiée, b_{rM} tend vers l'infini lorsque r tend vers l'infini.

Rappelons la condition i) :

$$\forall c > 0, \int_{A_c} (f_{\theta_1}/f_{\theta_0} - 1)^2 f_{\theta_0} dv = \infty$$

Soit c_0 une valeur de c pour laquelle $M = \max(\log(c_0+1), |\log(1-c_0)|)$;

Alors,

$$\forall x \text{ tel que } |f_{\theta_1}(x)/f_{\theta_0}(x) - 1| \leq c_0,$$

$$(f_{\theta_1}(x)/f_{\theta_0}(x) \geq 1) \Rightarrow (f_{\theta_1}(x)/f_{\theta_0}(x) \leq c_0 + 1) \Rightarrow \log(f_{\theta_1}(x)/f_{\theta_0}(x)) \leq M$$

$$(f_{\theta_1}(x)/f_{\theta_0}(x) < 1) \Rightarrow (f_{\theta_1}(x)/f_{\theta_0}(x) \geq 1 - c_0)$$

$$\Rightarrow |\log(f_{\theta_1}(x)/f_{\theta_0}(x))| \leq |\log(1 - c_0)|$$

Dans tous les cas, l'expression $|\log(f_{\theta_1}(x)/f_{\theta_0}(x))|$ est majorée par M ; on a ainsi établi l'inclusion :

$$\{x, |f_{\theta_1}(x)/f_{\theta_0}(x) - 1| \leq c_0\} \subset \{x, |\log f_{\theta_1}(x)/f_{\theta_0}(x)| \leq M\} = S_M$$

La condition i) entraîne donc l'assertion suivante :

$$\int_{S_M} (f_{\theta_1}/f_{\theta_0} - 1)^2 f_{\theta_0} dv = \infty$$

Puisque, de plus, $(f_{\theta_1}/f_{\theta_0} - 1)$ et $\log(f_{\theta_1}/f_{\theta_0})$ sont du même ordre lorsque $f_{\theta_1}/f_{\theta_0}$ est proche de 1 ; il vient :

$$\lim_{r \rightarrow \infty} b_{rM}^2 = \int_{S_M} \log(f_{\theta_1}/f_{\theta_0})^2 f_{\theta_0} dv = \infty$$

Puisque d'une part b_{rM} tend vers l'infini lorsque r tend vers l'infini, et

que d'autre part la variable $\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}}) \circ X$ est finie P p.p, $\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}}) \circ X / b_{rM}$ est un $o_p(1)$.

D'après la remarque que nous avons établie en préliminaire à cette démonstration, il nous suffit, pour démontrer le théorème, de nous assurer de la convergence de la transformée de Laplace de $(\Psi(\log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}}) \circ X - a_{rM}) / b_{rM}$, notée $\Phi_r(u)$, vers $\exp(u^2/2)$.

Un calcul identique à celui effectué au paragraphe III.1 nous amène à écrire :

$$\Phi_r(u) = E_{Q_{\lambda_{\theta_0}}}(\exp(-\langle Y, (u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0}) \mathbb{1}_{S_{rM}} \rangle)) e^{ua_{rM}/b_{rM}}$$

On reconnaît dans le second membre de cette égalité la valeur de la fonctionnelle de Laplace de Y en la fonction $f(x) = (u/b_{rM}) \log(f_{\theta_1}(x)/f_{\theta_0}(x)) \mathbb{1}_{S_{rM}}(x)$, ce à une constante multiplicative près.

Cette expression est bien définie puisque $\log(f_{\theta_1}(x)/f_{\theta_0}(x))$ est majoré par M sur S_{rM} , donc λ_{θ_0} intégrable sur T_r .

Une démarche là encore similaire à celle du paragraphe III.1 nous conduit à l'égalité :

$$\varphi_r(u) = \exp\left(\int_{S_{rM}} (e^{(-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0})} - 1) d\lambda_{\theta_0}\right) \exp(ua_{rM}/b_{rM}) \quad (1)$$

En substituant dans (1) a_{rM} par sa valeur, il vient :

$$\begin{aligned} \varphi_r(u) &= \exp\left(\int_{S_{rM}} (e^{(-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0})} - 1 + u/b_{rM} \log(f_{\theta_1}/f_{\theta_0})) d\lambda_{\theta_0}\right) \\ &= \exp(u^2/2) \exp\left(\int_{S_{rM}} \sum_{j \geq 3} [((-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0}))^j / j!] d\lambda_{\theta_0}\right) \end{aligned}$$

Il s'en suit que la convergence de $\varphi_r(u)$ vers $\exp(u^2/2)$ est assurée si l'on prouve que la quantité $\exp\left(\int_{S_{rM}} \sum_{j \geq 3} [((-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0}))^j / j!] d\lambda_{\theta_0}\right)$ tend vers 1 lorsque r tend vers l'infini

Puisque $|e^v - 1| \leq e^{|v|} - 1$, nous allons montrer que

$$\begin{aligned} \lim_{r \rightarrow \infty} \left| \int_{S_{rM}} \sum_{j \geq 3} [((-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0}))^j / j!] \right| &= 0. \text{ Or,} \\ \left| \int_{S_{rM}} \sum_{j \geq 3} [((-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0}))^j / j!] d\lambda_{\theta_0} \right| &= \\ \left| \sum_{j \geq 3} \int_{S_{rM}} [((-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0}))^j / j!] d\lambda_{\theta_0} \right| &\leq \\ \sum_{j \geq 3} \int_{S_{rM}} |(-u/b_{rM}) \log(f_{\theta_1}/f_{\theta_0})|^j d\lambda_{\theta_0} / j! &\quad (2) \end{aligned}$$

L'expression (2) est majorée par :

$$\sum_{j \geq 3} [u^j M^{j-2} / j!] \left[\int_{S_{rM}} (\log(f_{\theta_1}/f_{\theta_0}))^2 d\lambda_{\theta_0} \right] \left[\int_{S_{rM}} (\log(f_{\theta_1}/f_{\theta_0}))^2 d\lambda_{\theta_0} \right]^{-j/2} \quad (3)$$

De plus, (3) peut encore être majorée par :

$$\left[\int_{S_{rM}} (\log(f_{\theta_1}/f_{\theta_0}))^2 d\lambda_{\theta_0} \right]^{-1/2} \sum_{j \geq 3} [u^j M^{j-2} / j!]$$

Comme b_{rM} tend vers l'infini lorsque r tend vers l'infini, on a :

$$\lim_{r \rightarrow \infty} \left[\int_{S_{rM}} (\log(f_{\theta_1}/f_{\theta_0}))^2 d\lambda_{\theta_0} \right]^{-1/2} \sum_{j \geq 3} [u^j M^{j-2} / j!] = 0$$

Cela achève la démonstration du théorème lorsque la condition i) est vérifiée.

Il reste à nous assurer de la convergence en loi de la variable $(\Psi(\log(f_{\theta_1}/f_{\theta_0})) - a_{rM}) / b_{rM}$ sous l'hypothèse ii) du théorème. La démonstration en est pratiquement identique, la seule modification qu'il y ait lieu d'apporter tenant à ce fait :

Pour prouver que b_{rM} tend vers l'infini lorsque r tend vers l'infini, nous avons utilisé la condition i). Nous devons établir que c'est encore le cas lorsque la condition ii) est satisfaite.

Les hypothèses sont maintenant les suivantes :

$$\exists c > 0, \lambda_{\theta_0} (x, |\log (f_{\theta_1}^{(x)} / f_{\theta_0}^{(x)})| > c) = \infty$$

$$\exists M > 0, \lambda_{\theta_0} (x, |\log (f_{\theta_1}^{(x)} / f_{\theta_0}^{(x)})| > M) < \infty$$

De ces deux assertions se déduit une troisième :

$$\exists c > 0, \exists M > 0, \lambda_{\theta_0} (x, c < |\log (f_{\theta_1}^{(x)} / f_{\theta_0}^{(x)})| < M) = \infty$$

Alors

$$\begin{aligned} \lim_{r \rightarrow \infty} \int \mathbf{1}_{]c, M[} \cap T_r (|\log (f_{\theta_1} / f_{\theta_0})|) d\lambda_{\theta_0} \\ = \lambda_{\theta_0} (c < |\log (f_{\theta_1}^{(x)} / f_{\theta_0}^{(x)})| < M) = \infty \end{aligned}$$

Comme $b_{rM}^2 = \int_{S_{rM}} (\log (f_{\theta_1} / f_{\theta_0}))^2 d\lambda_{\theta_0} \geq c^2 \int \mathbf{1}_{]c, M[} \cap T_r (|\log (f_{\theta_1} / f_{\theta_0})|) d\lambda_{\theta_0}$

b_{rM} tend vers l'infini lorsque r tend vers l'infini.

CHAPITRE IV

ETUDE DE LA CONSISTANCE DU TEST DE NEYMAN-PEARSON

Nous allons maintenant aborder le problème de la consistance du test de l'hypothèse $\lambda = \lambda_{\theta_0}$ contre l'hypothèse $\lambda = \lambda_{\theta_1}$ tel que nous l'avons défini au second chapitre de notre étude.

Nous établirons tout d'abord l'équivalence entre la consistance de ce test et le fait que les lois de Poisson $P_{\lambda_{\theta_0}}$ et $P_{\lambda_{\theta_1}}$ soient étrangères l'une par rapport à l'autre.

Nous donnerons ensuite une condition nécessaire et suffisante, dérivée de celle énoncée par M Brown dans [4] pour que les lois $P_{\lambda_{\theta_0}}$ et $P_{\lambda_{\theta_1}}$ soient étrangères, donc assurant la consistance du test.

Enfin, nous étudierons un exemple de mesures pour lesquelles le test de Neyman-Pearson est effectivement consistant.

IV.1. Lemme :

On se propose ici d'établir l'équivalence de la consistance du test de Neyman-Pearson lorsque r tend vers l'infini avec l'orthogonalité des deux mesures de Poisson définissant l'hypothèse nulle et la contre-hypothèse.

Soient donc $P_{\lambda_{\theta_0}}$ et $P_{\lambda_{\theta_1}}$ deux mesures de Poisson sur $(\mathcal{M}(T, \mathcal{C}), \mathfrak{B}(\mathcal{M}))$, d'intensités respectives λ_{θ_0} et λ_{θ_1} sur (T, \mathcal{C}) . Désignons par $(\Sigma_r)_{r>0}$ la famille des sous tribus de \mathcal{C} engendrées par \mathcal{C}_r . Ainsi, la famille $(\Sigma_r)_{r>0}$ est une famille filtrante croissante de \mathcal{C} , engendrant \mathcal{C} puisque T est polonais, et, par construction, les restrictions de $P_{\lambda_{\theta_i}}$ ($i = 0, 1$) à Σ_r coïncident avec $P_{\lambda_{\theta_i}}^{(r)}$ sur (T_r, \mathcal{C}_r) .

Pour établir la consistance du test de Neyman-Pearson, nous aurons besoin du lemme fondamental suivant ([5]) :

Lemme IV.1.1.

« Soient (Ω, Σ) un espace probabilisable et $(\Sigma_\alpha)_{\alpha \in I}$ une famille croissante de sous tribus de Σ , engendrant Σ , où I est un ensemble filtrant croissant quelconque. Désignons par P et Q deux mesures de probabilité sur (Ω, Σ) , et par P_α et Q_α leurs restrictions à Σ_α . Introduisons les décompositions de Lebesgue $Q = f P + R$ et $Q_\alpha = f_\alpha P_\alpha + R_\alpha$; on a alors :

- a) la famille $(f_\alpha)_{\alpha \in I}$ est une surmartingale positive
- b) f_α tend vers f en P -mesure »

On peut dès lors énoncer

Proposition IV.1.2.

Une condition nécessaire et suffisante pour que le test de Neyman-Pearson de $\lambda_\theta = \lambda_{\theta_0}$ contre $\lambda_\theta = \lambda_{\theta_1}$ soit consistant quand r tend vers l'infini est que les mesures $P_{\lambda_{\theta_0}}$ et $P_{\lambda_{\theta_1}}$ soient orthogonales.

Démonstration

Considérons les décompositions de Lebesgue :

$$P_{\lambda_{\theta_1}} = f P_{\lambda_{\theta_0}} + R$$

$$P_{\lambda_{\theta_1}^r} = f_r P_{\lambda_{\theta_0}^r} + R_r$$

D'après le lemme, f_r tend vers f en $P_{\lambda_{\theta_0}}$ mesure. Or, l'orthogonalité de $P_{\lambda_{\theta_0}}$ et $P_{\lambda_{\theta_1}}$ équivaut à la nullité $P_{\lambda_{\theta_0}}$ p.s. de f , et donc f_r converge vers 0 en $P_{\lambda_{\theta_0}}$ mesure. De plus, et pour tout r , la fonction mesurable f_r coïncide, à un ensemble de $P_{\lambda_{\theta_0}}$ mesure nulle près, avec la statistique du rapport des vraisemblances définie au paragraphe II.1.

Il s'en suit que le logarithme du rapport des vraisemblances $\Psi(\log f^{(r)}(X)) - \int_{T_r} (f^{(r)} - 1) d\lambda_{\theta_0}^{(r)}$ tend vers $-\infty$ en $P_{\lambda_{\theta_0}}$ mesure.

Un raisonnement identique permet d'établir la convergence de cette statistique vers $+\infty$ en $P_{\lambda_{\theta_1}}$ mesure ; ainsi, pour tout seuil $\alpha > 0$, le test de Neyman-Pearson de $\lambda = \lambda_{\theta_0}$ contre $\lambda = \lambda_{\theta_1}$, ou de $\lambda = \lambda_{\theta_1}$ contre $\lambda = \lambda_{\theta_0}$ a une puissance approchant 1 lorsque r tend vers l'infini.

Réciproquement, si pour tout seuil $\alpha > 0$, le test de Neyman-Pearson est

consistant, il s'en suit que la statistique $\Psi(\log f^{(r)}(x)) - \int_{T_r} (f^{(r)} - 1) d\lambda_{\theta_0}^{(r)}$ diverge, et donc que $P_{\lambda_{\theta_0}}$ et $P_{\lambda_{\theta_1}}$ sont orthogonales.

IV.2. CONSISTANCE DU TEST :

Théorème IV.2.1 :

Une condition nécessaire et suffisante pour que le test de l'hypothèse $\lambda_{\theta} = \lambda_{\theta_0}$ contre l'hypothèse $\lambda_{\theta} = \lambda_{\theta_1}$ soit consistant est que l'une au moins des assertions suivantes soit vérifiée :

$$a) \exists m > 0 \text{ tel que } \int_{B_m} |f_{\theta_0} - f_{\theta_1}| d\nu = \infty$$

$$\text{où } B_m = \{x, |f_{\theta_0}(x) / f_{\theta_1}(x) - 1| > m\}$$

$$b) \forall m > 0 \int_{B_m^c} (f_{\theta_0} / f_{\theta_1} - 1)^2 f_{\theta_1} d\nu = \infty$$

Démonstration

Pour faire la preuve de ce théorème, il suffit d'après le lemme IV.1.1. de montrer que l'une ou l'autre des conditions a) ou b) entraîne que les lois de Poisson $P_{\lambda_{\theta_0}}$ et $P_{\lambda_{\theta_1}}$ sont étrangères.

Cela se déduit tout simplement d'un théorème que Brown énonce dans [4], et que nous allons rappeler.

« Soient P_{μ} et P_{ν} deux lois de Poisson définies sur un espace mesurable $(\mathcal{X}, \mathcal{G})$ dont les intensités μ et ν sont des mesures σ finies. Soient $\mu = \lambda + w$ la décomposition de Lebesgue de μ relativement à ν , et f un

représentant de la dérivée de Radon-Nikodym de λ par rapport à ν .

Les lois P_μ et P_ν sont étrangères si et seulement si l'une des conditions suivantes est vérifiée :

$$i) w(\mathcal{X}) = \infty$$

$$ii) \exists m > 0 \text{ tel que } \int_{|f-1|>m} |f-1| d\nu = \infty$$

$$iii) \forall m > 0, \int_{|f-1|\leq m} (f-1)^2 d\nu = \infty. \quad \gg$$

Considérons la décomposition de Lebesgue de λ_{θ_0} par rapport à λ_{θ_1} en $\lambda_{\theta_0} = f\lambda_{\theta_1} + w$, où w est une mesure étrangère à λ_{θ_1} . Désignons par B le support de w .

On a, bien sûr, $w(B^c) = 0$, mais également :

$$\begin{aligned} w(B^c) &= \lambda_{\theta_0}(B^c) - \int_{B^c} f d\lambda_{\theta_1} \\ &= \int_{B^c} (f_{\theta_0} - f f_{\theta_1}) d\nu = \int_{B^c} (f_{\theta_0}/f_{\theta_1} - f) d\lambda_{\theta_1} \end{aligned}$$

Cela entraîne que $f = f_{\theta_0}/f_{\theta_1}$ sur l'ensemble B^c , donc que $f = f_{\theta_0}/f_{\theta_1}$ λ_{θ_1} pp puisque $\lambda_{\theta_1}(B) = 0$

Remplaçons f par $f_{\theta_0}/f_{\theta_1}$ dans les conditions ii) et iii) du théorème de Brown ; il vient :

$$\exists m > 0 \text{ tel que } \int_{B_m^c} |f_{\theta_0}/f_{\theta_1} - 1| d\lambda_{\theta_1} = \infty$$

$$\forall m > 0 \int_{B_m^c} (f_{\theta_0}/f_{\theta_1} - 1)^2 d\lambda_{\theta_1} = \infty$$

On retrouve ainsi les conditions que nous avons énoncées au théorème IV.2.1.

IV. 3. EXEMPLE :

Prenons pour T l'espace \mathbb{R}^2 , et pour \mathcal{C} la tribu des boréliens de \mathbb{R}^2 , $\mathfrak{B}_{\mathbb{R}^2}$.

Soient ν , λ_{θ_0} et λ_{θ_1} des mesures de $\mathcal{M}(\mathbb{R}^2, \mathfrak{B}_{\mathbb{R}^2})$ définies de la manière suivante :

ν est la somme de la mesure de Lebesgue sur \mathbb{R}^2 et de la mesure de comptage sur \mathbb{N}^2

λ_{θ_0} et λ_{θ_1} sont deux mesures dominées par ν , et leurs dérivées de Radon-Nikodym par rapport à cette mesure, notées f_{θ_0} et f_{θ_1} , sont données par :

$$\forall (x,y) \in \mathbb{R}^2,$$

$$f_{\theta_0}(x,y) = \mathbb{1}_{\mathbb{R}^2 \setminus \mathbb{N}^2}(x,y)$$

$$f_{\theta_1}(x,y) = e^{x+y}$$

On montre le résultat suivant :

La statistique définissant la région critique du test de l'hypothèse

$\lambda_{\theta_0}^{(r)} = \lambda_{\theta_0}^{(r)}$ contre $\lambda_{\theta_1}^{(r)} = \lambda_{\theta_1}^{(r)}$ est donnée, $P_{\lambda_{\theta_0}}$ pp, par :

$$\Psi(\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \circ X(\omega) =$$

$$\sum_{j=1, n(\omega)} (x_j(\omega) + y_j(\omega)) \mathbb{1}_{(\mathbb{R}^2 \setminus \mathbb{N}^2) \cap T_r}(x_j(\omega), y_j(\omega)).$$

De plus, ce test est consistant lorsque r tend vers l'infini.

Preuve

Nous avons défini dans la proposition II.2.1 la région critique de ce test comme étant :

$$R_C = \{\omega, \Psi(\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \circ X(\omega) > k\}$$

$$\log f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)}(x,y) = +\infty \mathbf{1}_{\mathbb{N}^2 \cap T_r}(x,y) + (x+y) \mathbf{1}_{(\mathbb{R}^2 \setminus \mathbb{N}^2) \cap T_r}(x,y).$$

Alors,

$$\Psi(\log f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)}) \circ X(\omega) =$$

$$\sum_{j=1, n(\omega)} (\infty \mathbf{1}_{\mathbb{N}^2 \cap T_r}(x_j(\omega), y_j(\omega)) + (x_j(\omega) + y_j(\omega)) \mathbf{1}_{(\mathbb{R}^2 \setminus \mathbb{N}^2) \cap T_r}(x_j(\omega), y_j(\omega)))$$

Nous avons choisi la mesure λ_{θ_0} de façon à ce que l'on ait $\lambda_{\theta_0}(\mathbb{N}^2) = 0$; cela entraîne l'égalité :

$$\Psi(\log f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)}) \circ X(\omega) =$$

$$\sum_{j=1, n(\omega)} (x_j(\omega) + y_j(\omega)) \mathbf{1}_{(\mathbb{R}^2 \setminus \mathbb{N}^2) \cap T_r}(x_j(\omega), y_j(\omega)) P_{\lambda_{\theta_0}}$$

Etablissons maintenant que le test est consistant. Pour ce faire, montrons que la condition i) du théorème IV.2.1. est vérifiée.

Rappelons cette condition :

$$\exists m > 0, \int_{B_m} |f_{\theta_0} - f_{\theta_1}| \, d\nu = \infty$$

$$\text{où } B_m = \{x, |f_{\theta_0}(x) / f_{\theta_1}(x) - 1| > m\}$$

Pour tout couple $(x,y) \in \mathbb{R}^2$, on a :

$$|f_{\theta_0}(x,y) - f_{\theta_1}(x,y)| = | \mathbf{1}_{\mathbb{R}^2 \setminus \mathbb{N}^2}(x,y) - e^{x+y} |$$

$$|f_{\theta_0}(x,y) / f_{\theta_1}(x,y) - 1| = | e^{-(x+y)} \mathbf{1}_{\mathbb{R}^2 \setminus \mathbb{N}^2}(x,y) - 1 |$$

Prenons pour m la valeur $1/2$.

$$\forall (x,y) \in \mathbb{N}^2, |f_{\theta_0}(x,y) / f_{\theta_1}(x,y) - 1| > m$$

Cela entraîne l'inégalité :

$$\int_{\mathbb{N}^2} | \mathbf{1}_{\mathbb{R}^2 \setminus \mathbb{N}^2}(x,y) - e^{x+y} | \, d\nu \leq \int_{B_m} |f_{\theta_0} - f_{\theta_1}| \, d\nu$$

D'autre part,

$$\int_{\mathbb{N}^2} |f_{\theta_0}(x,y) - e^{x+y}| d\nu = \int_{\mathbb{N}^2} e^{x+y} d\nu = \sum_{x_i \in \mathbb{N}} \sum_{y_j \in \mathbb{N}} e^{x_i+y_j} = \infty$$

On a ainsi exhibé une valeur de m pour laquelle l'intégrale $\int_{B_m} |f_{\theta_0} - f_{\theta_1}| d\nu$ est infinie, et le test est consistant.

Le théorème III.2.1. ne permet pas de conclure à la normalité asymptotique de la statistique du test de Neyman-Pearson, car il n'est pas possible de trouver une valeur de M pour laquelle $\lambda_{\theta_0}(S_{MC})$ est finie.

CHAPITRE V

TEST DE NEYMAN - PEARSON, ET PROPRIETES DE CE TEST POUR DES PROCESSUS DEFINIS A PARTIR DE FAMILLES DE MESURES PARTICULIERES

Nous nous restreignons dans ce chapitre au cas où $T = \mathbb{R}^S$, et où \mathcal{G} est la tribu des boréliens de \mathbb{R}^S .

Dans le premier paragraphe, nous réétudions le test de Neyman-Pearson pour des processus de Poisson dont les intensités appartiennent à une famille de mesures monotones.

Nous consacrons le second paragraphe à l'étude des propriétés de ce test lorsque les processus ont pour intensités des éléments d'une famille de mesures exponentielles.

V.1. FAMILLE DE MESURES MONOTONES :

L'objet de ce paragraphe est d'établir un test d'hypothèses multiples à partir du test d'hypothèses simples que nous avons mis en place au chapitre II.

Avant d'envisager ce problème, commençons par expliciter ce que nous

entendons par famille de mesures monotone.

Définissons tout d'abord la notion d'inégalité vectorielle telle que nous la concevons :

$$\forall \theta \in \mathbb{R}^S, \forall \theta_0 \in \mathbb{R}^S,$$

$$\theta \leq \theta_0 \Leftrightarrow \forall i \in (1 \dots s), \theta^{(i)} \leq \theta_0^{(i)}$$

$$\theta < \theta_0 \Leftrightarrow (\theta \leq \theta_0) \text{ et } (\exists j \in (1 \dots s), \theta^{(j)} < \theta_0^{(j)}).$$

Soit $(\lambda_\theta, \theta \in I \subset \mathbb{R}^S)$ une famille de mesures de $\mathcal{M}(\mathbb{R}^S, \mathfrak{B}_{\mathbb{R}^S})$, dominée par une mesure ν σ finie.

Définition V.1.1 :

La famille de mesures $(\lambda_\theta, \theta \in I)$ est dite monotone (croissante) si les dérivées de Radon-Nikodym de λ_θ par rapport à ν vérifient la condition :

$$\forall \theta \in I, \forall \theta' \in I, \theta < \theta' \Rightarrow f_\theta(x) < f_{\theta'}(x).$$

Nous pouvons alors énoncer la proposition suivante :

Proposition V.1.2 :

Soient θ_0 et θ_1 deux valeurs du paramètre θ telles que $\theta_0 < \theta_1$.

Le test de niveau α de l'hypothèse $\lambda_\theta^{(r)} = \lambda_{\theta_0}^{(r)}$ est encore de niveau α comme test de $\lambda_\theta^{(r)} \in [\lambda_{\theta'}^{(r)}, \theta' \leq \theta_0]$ contre $\lambda_\theta^{(r)} \in [\lambda_{\theta''}^{(r)}, \theta'' > \theta_0]$, et sa fonction puissance est monotone non décroissante.

Démonstration :

La preuve de cette démonstration demande le résultat préliminaire suivant :

Lemme V.1.3 :

Soit k une fonction non négative définie sur T_r ; alors, pour tout $a > 0$,

$$P_{\lambda_{\theta_0}^{(r)}}[X(\omega), \Psi(k) \circ X(\omega) > a] \leq P_{\lambda_{\theta_1}^{(r)}}[X(\omega), \Psi(k) \circ X(\omega) > a]$$

Démonstration du lemme :

Soit $(A_{j,n})_{j=0 \dots 4^n-1, n \in \mathbb{N}^*}$ une suite de sous ensembles de T_r tels que :

$$A_{j,n} = \{x \in T_r, k(x) \in [(j/2^n), ((j+1)/2^n)]\}$$

Notons $(C_n)_{n \in \mathbb{N}^*}$ la suite de variables aléatoires définies, pour toute valeur ω de Ω , et pour tout $n \in \mathbb{N}^*$ par :

$$C_n(\omega) = \langle X(\omega), \sum_{j=0, 4^n-1} (j/2^n) \mathbf{1}_{A_{j,n}} \rangle$$

Soit λ_{θ_i} , $i = 0$ ou 1 , l'intensité du processus de Poisson X ; les variables $\langle X(\omega), \mathbf{1}_{A_{j,n}} \rangle$ suivent des lois de Poisson de paramètre $\lambda_{\theta_i}^{(r)}(A_{j,n})$.

La famille de mesures $(\lambda_{\theta}, \theta \in I)$ étant monotone, on a :

$$\forall x \in T_r, f_{\theta_0}^{(r)}(x) \leq f_{\theta_1}^{(r)}(x).$$

Alors, $\forall n \in \mathbb{N}^*, \forall j \in \{0, \dots, 4^n-1\}$,

$$\int_{A_{j,n}} f_{\theta_0}^{(r)}(x) dv \leq \int_{A_{j,n}} f_{\theta_1}^{(r)}(x) dv$$

et $\lambda_{\theta_0}^{(r)}(A_{j,n}) \leq \lambda_{\theta_1}^{(r)}(A_{j,n})$.

Soit alors $((M_{j,n})_{j=0 \dots 4^n-1})_{n \in \mathbb{N}^*}$ une suite de variables aléatoires suivant des lois de Poisson de paramètre $(\lambda_{\theta_1}^{(r)}(A_{j,n}) - \lambda_{\theta_0}^{(r)}(A_{j,n}))$, et indépendantes, pour toute valeur de n , des variables C_n .

On a choisi la suite $(M_{j,n})$ de manière à ce que l'on ait :

$$\forall a \in \mathbb{R}^+, P_{\lambda_{\theta_0}^{(r)}}[C_n + \sum_{j=0,4^{n-1}} (j/2^n) M_{j,n} > a] = P_{\lambda_{\theta_1}^{(r)}}[C_n > a]$$

Puisque $(\sum_{j=0,4^{n-1}} (j/2^n) M_{j,n})_n$ est à valeurs positives ou nulles, il vient :

$$\forall a \in \mathbb{R}^+, P_{\lambda_{\theta_0}^{(r)}}[C_n > k] \leq P_{\lambda_{\theta_1}^{(r)}}[C_n > k] \quad (1)$$

De plus, la suite de fonctions étagées $(\sum_{j=0,4^{n-1}} (j/2^n) \mathbb{1}_{[j/2^n, (j+1)/2^n[}(k(x)))_n$ converge en tout point de T_r vers $k(x)$;

La suite de variables aléatoires $(C_n)_n$ converge donc presque partout vers $\Psi(k) \circ X$, et l'on arrive, par passage à la limite dans l'inégalité (1), au résultat :

$$P_{\lambda_{\theta_0}^{(r)}}[\Psi(k) \circ X > a] \leq P_{\lambda_{\theta_1}^{(r)}}[\Psi(k) \circ X > a]$$

Etablissons maintenant la preuve de la proposition .

Le test de niveau α de l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_0}^{(r)}$ contre l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_1}^{(r)}$ est défini par la région critique :

$$R_C = \{\omega, \Psi(\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \circ X(\omega) > \lambda_{\alpha}\} = \alpha$$

La fonction $\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}$ est à valeurs positives puisque la famille de mesures $(\lambda_{\theta}, \theta \in I)$ est monotone. Le lemme V.1.3., appliqué à cette fonction, nous donne le résultat suivant :

L'application qui, à une valeur θ du paramètre, associe la probabilité $P_{\lambda_{\theta}^{(r)}}[\Psi(\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \circ X > \lambda_{\alpha}]$, est croissante.

En particulier, on a :

$$- \forall \theta \leq \theta_0, P_{\lambda_{\theta}^{(r)}}[\Psi(\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \circ X > \lambda_{\alpha}] \leq \alpha.$$

Le test est encore de niveau α comme test de $\lambda_{\theta}^{(r)} \in \{ \lambda_{\theta_0}^{(r)}, \theta \leq \theta_0 \}$
 contre $\lambda_{\theta}^{(r)} \in \{ \lambda_{\theta_1}^{(r)}, \theta > \theta_0 \}$

$$- \forall \theta \geq \theta_1, P_{\lambda_{\theta}^{(r)}} [\Psi(\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \circ X > \lambda_{\alpha}] \geq$$

$$P_{\lambda_{\theta_1}^{(r)}} [\Psi(\log f_{\theta_1}^{(r)} / f_{\theta_0}^{(r)}) \circ X > \lambda_{\alpha}]$$

la fonction puissance est non décroissante, et l'on peut ajouter une remarque :

Remarque V.1.4 :

Si les lois de Poisson $P_{\lambda_{\theta_0}^{(r)}}$ et $P_{\lambda_{\theta_1}^{(r)}}$ sont étrangères, la puissance tend vers 1 pour toute valeur du paramètre $\theta \geq \theta_1$, et tout test de NEYMAN-PEARSON de l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_0}^{(r)}$ contre l'hypothèse $\lambda_{\theta}^{(r)} = \lambda_{\theta_1}^{(r)}$, où $\theta \geq \theta_1$, est consistant lorsque r tend vers l'infini.

Passons maintenant à l'étude des processus de Poisson ayant pour intensités des éléments d'une famille de mesures exponentielles.

V.2. STRUCTURE EXPONENTIELLE :

Nous abordons ce paragraphe en rappelant ce que l'on entend par structure exponentielle.

Définition V.2.1 :

On dit que la structure statistique $(\mathbb{R}^S, \mathfrak{B}_{\mathbb{R}^S}, \lambda_{\theta}, \theta \in I \subset \mathbb{R}^S)$ est exponentielle si $f_{\theta}(x)$ est de la forme

$$f_{\theta}(x) = c(\theta) \exp \left(\sum_{j=1, S} p_j(\theta) k_j(x) \right) h(x).$$

Nous pouvons dès lors faire la proposition suivante :

Proposition V.2.2. :

La structure statistique $(\mathcal{M}(T_r, \mathcal{E}_r), \mathfrak{B}(\mathcal{M}), P_{\lambda_\theta}^{(r)}, \theta \in I)$ admet la statistique exhaustive $(\Psi(k_1) \circ X, \dots, \Psi(k_s) \circ X)$; de plus, cette statistique est complète dès que l'image par p de I contient un ouvert de \mathbb{R}^S .

Démonstration :

La famille de mesures $(\lambda_\theta, \theta \in I)$ est dominée par une mesure ν σ finie. Le théorème II.1.1 permet alors d'affirmer que la dérivée de Radon-Nikodym de $P_{\lambda_\theta}^{(r)}$ par rapport à $P_\nu^{(r)}$ existe, et est donnée par :

$$(dP_{\lambda_\theta}^{(r)}/dP_\nu^{(r)})(X(\omega)) = \exp \left[- \int_{T_r} (f_\theta^{(r)} - 1) d\nu + \Psi(\log f_\theta^{(r)}) \circ X(\omega) \right]$$

En remplaçant dans cette égalité $f_\theta^{(r)}$ par sa valeur, il vient :

$$(dP_{\lambda_\theta}^{(r)}/dP_\nu^{(r)}) = \exp \left[- \int_{T_r} (c(\theta) \exp(\sum_{j=1,s} p_j(\theta) k_j(x)) h(x) - 1) \nu(dx) + \right. \\ \left. \Psi(\log(c(\theta) h(x)) \sum_{j=1,s} p_j(\theta) k_j(x)) \circ X(\omega) = \right. \\ \left. c(\theta) \exp(\sum_{j=1,s} p_j(\theta) \Psi(k_j) \circ X(\omega)) h(x) \right]$$

La structure statistique $(\mathcal{M}(T_r, \mathcal{E}_r), \mathfrak{B}(\mathcal{M}), P_{\lambda_\theta}^{(r)}, \theta \in I)$ est donc exponentielle.

Un théorème que J.R. BARRA énonce dans [1] permet alors de conclure à l'exhaustivité de la variable

$$T = (\Psi(k_1) \circ X \dots \Psi(k_s) \circ X)$$

De plus, si l'on s'est assuré de ce que l'ensemble $(p(\theta), \theta \in I)$ contienne un ouvert de \mathbb{R}^S , on montre (cf [20]) que la statistique T est complète.

Lorsque la structure statistique $(\mathbb{R}^S, \mathfrak{B}_{\mathbb{R}^S}, \lambda_\theta, \theta \in I)$ est exponentielle, on peut définir des conditions, plus simples que celles du théorème IV.2.1, impliquant la consistance du test de $\lambda_\theta = \lambda_{\theta_0}$ contre $\lambda_\theta = \lambda_{\theta_1}$. Il est également possible de donner de nouvelles conditions sous lesquelles la statistique du test est asymptotiquement normale. C'est à cette étude que nous consacrons la seconde partie de ce paragraphe.

Soit $(\mathbb{R}^S, \mathfrak{B}_{\mathbb{R}^S}, \lambda_\theta, \theta \in I)$ une structure statistique de type exponentielle, dominée par une mesure ν σ finie. Désignons par f_θ un représentant de la dérivée de Radon-Nikodym de λ_θ par rapport à ν , dont l'expression est donnée par $f_\theta(x) = \exp \langle p(\theta), k(x) \rangle$, où k est une application définie sur \mathbb{R}^S , à valeurs dans $(\mathbb{R}^+)^S$ et où p , définie de \mathbb{R}^S dans \mathbb{R}^S est bijective, et croissante au sens où elle vérifie la condition :

$$\forall \theta_1 \in I, \forall \theta_2 \in I, \theta_1 < \theta_2 \Rightarrow p(\theta_1) < p(\theta_2)$$

Soient θ_0 et θ_1 deux valeurs du paramètre telles que $\theta_0 > \theta_1$.

Proposition V.2.3.

Si l'une des conditions suivantes c_1 ou c_2 est vérifiée, le test de l'hypothèse $\lambda_\theta = \lambda_{\theta_0}$ contre $\lambda_\theta = \lambda_{\theta_1}$ est consistant

$$c_1 \quad \exists c \in \mathbb{R}^{+*}, \lambda_{\theta_1}(x, \forall j, k_j(x) > c) = \infty$$

$$c_2 \quad \forall a \in \mathbb{R}^{+*}, \forall i \in \{1 \dots s\}, \int k_i^2(x) \mathbb{1}(x, \forall j, k_j(x) \leq a) d\lambda_{\theta_1} = \infty$$

Démonstration :

Nous allons montrer que les hypothèses c_1 et c_2 impliquent les

conditions de consistance que nous avons établies lors du théorème IV.2.1.

$$\text{Soit } A_c = \{x \in \mathbb{R}^S, \forall j, k_j(x) > c\}$$

Supposons la condition c_1 vérifiée, ie $\lambda_{\theta_1}(A_c) = \infty$. Puisque l'application p est croissante, et qu'on a supposé θ_0 supérieur à θ_1 , il existe un indice j_0 au moins tel que $p_{j_0}(\theta_0) > p_{j_0}(\theta_1)$.

Posons $\alpha = \exp((p_{j_0}(\theta_0) - p_{j_0}(\theta_1))c) - 1$; on a :

$$\forall x \in A_c, \exp(p(\theta_0) - p(\theta_1), k(x)) - 1 > \alpha$$

Cela entraîne l'inclusion $A_c \subset B_\alpha$, où B_α est défini par :

$$B_\alpha = \{x \in \mathbb{R}^S, \exp(p(\theta_0) - p(\theta_1), k(x)) - 1 > \alpha\}$$

Puisque $\lambda_{\theta_1}(A_c) = \infty$, l'ensemble B_α est de mesure λ_{θ_1} infinie, et il vient :

$$\begin{aligned} \int \mathbb{1}_{\{(\|f_{\theta_0}/f_{\theta_1} - 1\|) > \alpha\}} d\lambda_{\theta_1} &= \int_{B_\alpha} (\exp(p(\theta_0) - p(\theta_1), k(x)) - 1) d\lambda_{\theta_1} \\ &\geq \alpha \int_{B_\alpha} d\lambda_{\theta_1} = \infty \end{aligned}$$

On a exhibé une valeur $\alpha \in \mathbb{R}^{+*}$ pour laquelle $\int \mathbb{1}_{\{(\|f_{\theta_0}/f_{\theta_1} - 1\|) > \alpha\}} d\lambda_{\theta_1}$ est infinie, ce qui montre que la condition c_1 de cette proposition entraîne la condition 1) du théorème IV.2.1, et donc est une condition suffisante pour la consistance du test.

Supposons maintenant que l'hypothèse c_2 est vérifiée. Puisque l'application p est bijective et croissante, il existe, pour toute valeur positive de c , un réel a positif tel que :

$$a = \text{Log}(c+1) / \left(\sum_{j=1, S} (p_j(\theta_0) - p_j(\theta_1)) \right)$$

Désignons alors par :

$$C_a = \{x \in \mathbb{R}^S, \forall j, k_j(x) \leq a\}$$

$$D_c = \{x \in \mathbb{R}^S, \exp \langle p(\theta_0) - p(\theta_1), k(x) \rangle - 1 \leq c\}$$

Etablissons l'inclusion $C_a \subset D_c$:

$$\forall x \in C_a, \exp \langle p(\theta_0) - p(\theta_1), k(x) \rangle - 1 \leq \exp \left(a \sum_{j=1, S} (p_j(\theta_0) - p_j(\theta_1)) \right) - 1$$

De par le choix que nous avons fait de c , on a :

$$c = \exp \left(a \sum_{j=1, S} (p_j(\theta_0) - p_j(\theta_1)) \right) - 1, \text{ et } C_a \subset D_c.$$

Attachons nous maintenant à retrouver la seconde condition du théorème IV.2.1.

$$\begin{aligned} \int (r_{\theta_0}/r_{\theta_1} - 1)^2 \mathbb{1}(|r_{\theta_0}/r_{\theta_1} - 1| \leq c) d\lambda_{\theta_1} &= \int_{D_c} (\exp \langle p(\theta_0) - p(\theta_1), k(x) \rangle - 1)^2 d\lambda_{\theta_1} \\ &\geq \int_{D_c} \langle p(\theta_0) - p(\theta_1), k(x) \rangle^2 d\lambda_{\theta_1} \end{aligned}$$

Soit j_0 un indice tel que $p_{j_0}(\theta_0) > p_{j_0}(\theta_1)$; alors,

$$\int_{D_c} \langle p(\theta_0) - p(\theta_1), k(x) \rangle^2 d\lambda_{\theta_1} \geq (p_{j_0}(\theta_0) - p_{j_0}(\theta_1))^2 \int_{D_c} k^2_{j_0}(x) d\lambda_{\theta_1}$$

Puisque $D_c \subset C_a$, il vient :

$$(p_{j_0}(\theta_0) - p_{j_0}(\theta_1))^2 \int_{D_c} k^2_{j_0}(x) d\lambda_{\theta_1} \geq (p_{j_0}(\theta_0) - p_{j_0}(\theta_1))^2 \int_{C_a} k^2_{j_0}(x) d\lambda_{\theta_1}$$

Comme, par hypothèse, l'intégrale $\int_{C_a} k^2_{j_0}(x) d\lambda_{\theta_1}$ est infinie pour toute valeur de a , on a :

$$\forall c > 0, \int (r_{\theta_0}/r_{\theta_1} - 1)^2 \mathbb{1}(|r_{\theta_0}/r_{\theta_1} - 1| \leq c) d\lambda_{\theta_1} = \infty$$

On a ainsi établi que la condition c_2 entraîne la consistance du test.

Nous nous proposons maintenant de définir des conditions plus simples que celles énoncées au théorème III.2.1 pour lesquelles la variable $(\Psi(\log r_{\theta_1}^{(r)}/r_{\theta_0}^{(r)}) - a_{rM}) / b_{rM}$ est asymptotiquement normale.

Afin de clarifier la démonstration de la proposition que nous allons établir, nous supposons ici que θ_0 et θ_1 sont deux valeurs du paramètre θ telles que $\theta_0 < \theta_1$.

Désignons par :

$$S_{rM} = \{x \in T_r, \|k(x)\| \leq M\}$$

$$a_{rM} = \int_{S_{rM}} \langle p(\theta_0) - p(\theta_1), k(x) \rangle d\lambda_{\theta_0}$$

$$b_{rM}^2 = \int_{S_{rM}} \langle p(\theta_0) - p(\theta_1), k(x) \rangle^2 d\lambda_{\theta_0}$$

Proposition V.2.4. :

S'il existe une valeur de M pour laquelle $\lambda_{\theta_0}(S_{rM}^c)$ est finie, et si l'une au moins des conditions i) ou ii) suivante est vérifiée, alors la variable aléatoire $(\Psi(\log f_{\theta_1}^{(r)}/f_{\theta_0}^{(r)}) \text{ ou } -a_{rM} / b_{rM})$ converge en loi, lorsque r tend vers l'infini, vers une variable de loi normale centrée réduite.

$$i) \forall a \in \mathbb{R}^{+*}, \forall i \in (1 \dots s), \int k_i^2(x) \mathbb{1}(x, \forall j, k_j(x) \leq a) d\lambda_{\theta_0} = \infty$$

$$ii) \exists c \in \mathbb{R}^{+*}, \lambda_{\theta_0}(x, \forall j \in (1 \dots s), k_j(x) > c) = \infty$$

Démonstration :

Il nous faut montrer que ces hypothèses entraînent les conditions de normalité asymptotique du théorème III.2.1.

Supposons donc connue l'existence d'une valeur de M pour laquelle $\lambda_{\theta_0}(S_{rM}^c)$ est finie, et prouvons que nous pouvons exhiber un M_0 tel que :

$$\lambda_{\theta_0}(x, \|\log f_{\theta_1}(x)/f_{\theta_0}(x)\| > M_0) < \infty$$

Posons $M_0 = M \|p(\theta_1) - p(\theta_0)\|$; pour tout x vérifiant l'inégalité $\|\log f_{\theta_1}(x)/f_{\theta_0}(x)\| > M_0$, ou, ce qui est équivalent, $|\langle p(\theta_1) - p(\theta_0), k(x) \rangle| >$

M_0 , on a :

$$\|p(\theta_1) - p(\theta_0)\| \|k(x)\| \geq |\langle p(\theta_1) - p(\theta_0), k(x) \rangle| > M \|p(\theta_1) - p(\theta_0)\|$$

Puisque l'application p est bijective, cela entraîne $\|k(x)\| > M$, et donc l'inclusion :

$$\{x, \|\log f_{\theta_1}(x)/f_{\theta_0}(x)\| > M_0\} \subset \{x, \|k(x)\| > M\}$$

En particulier,

$$(\lambda_{\theta_0} \{x, \|k(x)\| > M\} < \infty) \Rightarrow (\lambda_{\theta_0} \{x, \|\log f_{\theta_1}(x)/f_{\theta_0}(x)\| > M_0\} < \infty)$$

Il nous reste à établir que les hypothèses i) et ii) de cette proposition impliquent les conditions i) et ii) du théorème III.2.1.

$$i) (\forall a \in \mathbb{R}^{+*}, \forall l \in (1 \dots s), \int k_l^2(x) \mathbb{1}_{\{x, \forall j, k_j(x) \leq a\}} d\lambda_{\theta_0} = \infty)$$

$$\Rightarrow (\forall c > 0, \int (f_{\theta_1}/f_{\theta_0} - 1)^2 \mathbb{1}_{\{|f_{\theta_1}/f_{\theta_0} - 1| \leq c\}} f_{\theta_0} dv = \infty)$$

Un résultat similaire à celui-ci a été établi au cours de la démonstration de la proposition V.2.3. ; nous ne reviendrons donc pas sur la preuve de cette assertion.

$$ii) (\exists c \in \mathbb{R}^{+*}, \lambda_{\theta_0} \{x, \forall j \in (1 \dots s), k_j(x) \geq c\} = \infty)$$

$$\Rightarrow (\exists c' > 0, \lambda_{\theta_0} \{x, \|\log f_{\theta_1}(x)/f_{\theta_0}(x)\| \geq c'\} = \infty).$$

Là encore, nous allons pour prouver ce résultat montrer l'inclusion de l'ensemble $A_c = \{x, \forall j \in (1 \dots s), k_j(x) \geq c\}$ dans l'ensemble $B_{c'} = \{x, \|\log f_{\theta_1}(x)/f_{\theta_0}(x)\| \geq c'\}$

Posons $c' = c \sum_{j=1, s} (p_j(\theta_1) - p_j(\theta_0))$; quelque soit $x \in A_c$, on a :

$$\|\log f_{\theta_1}(x)/f_{\theta_0}(x)\| = \langle p(\theta_1) - p(\theta_0), k(x) \rangle = \sum_{j=1, s} (p_j(\theta_1) - p_j(\theta_0)) k_j(x)$$

x appartenant à A_c , le dernier membre de cette égalité peut être minoré

par c' , donc $A_c \subset B_{c'}$, et $\lambda_{\theta_0}(B_{c'}) = \infty$.

DEUXIEME PARTIE

TRAITEMENT STATISTIQUE

D'UN MULTIDECTEUR DE PARTICULES

TABLE DES MATIERES

INTRODUCTION - AVERTISSEMENT

CHAPITRE I: Réponse du multidétecteur à un flux homogène de particules.

1.1 - Quelques définitions

1.2 - Etude de l'homogénéité dans l'espace du détecteur

1.3 - Etude de la stabilité dans le temps du multidétecteur

CHAPITRE II: Mode de calcul de l'intensité intégrée d'une diffraction de BRAGG

11.1 - Position du problème et description du traitement envisagé

11.2 - Calcul de l'intensité intégrée d'une diffraction - cas général

11.3 - Calcul de l'intensité intégrée d'une diffraction pour une valeur élevée du bruit de fond

ANNEXE 1 Enregistrement d'une diffraction sur un multidétecteur de particules

ANNEXE 2 Simulation d'un enregistrement

ANNEXE 3 Traitement d'une diffraction ; cas général

ANNEXE 4 Traitement d'une diffraction ; cas d'un bruit de fond élevé



INTRODUCTION

Rappelons quelques notions élémentaires de cristallographie qui permettent de mieux comprendre le fondement de ce travail.

Les cristaux possèdent cette propriété singulière de renvoyer un faisceau incident de rayons X ou de neutrons dans un ensemble discret de directions privilégiées. Appliquée aux cristaux moléculaires, l'étude de cette propriété constitue la principale méthode de détermination de structures moléculaires.

La figure de diffraction permet d'une part d'identifier le cristal si un spectre a été obtenu antérieurement, et d'autre part, ce qui est beaucoup plus important, de reconnaître la symétrie du cristal, les dimensions de sa maille élémentaire, et de déterminer la localisation des atomes dans cette maille.

Sans entrer dans des considérations d'ordre physique (cf [22] pour une introduction à ce sujet), disons que l'interprétation des phénomènes de diffraction peut être scindée en deux parties distinctes :

a) analyse des directions de diffraction afin de connaître les caractéristiques géométriques du réseau cristallin (dimensions et symétrie de la maille élémentaire)

b) analyse des intensités diffractées pour déterminer la position des atomes dans la maille.

Nous allons nous intéresser plus particulièrement au problème de l'extraction des intensités diffractées à partir des diagrammes mesurés sur un détecteur à localisation bidimensionnelle (multidétecteur). Plus précisément, nous traitons le cas de données ayant des caractéristiques du type de celles enregistrées sur le détecteur multifils à gaz (64 anodes et 16 cathodes) du diffractomètre à neutrons pour monocristaux D 19 A, de l'Institut Laue Langevin à Grenoble (cf. [8]).

Lorsqu'un plan cristallin est en position de diffraction, on enregistre dans l'espace bidimensionnel du multidétecteur un signal que l'on appelle tache de BRAGG. Son intensité, ses dimensions, sa forme (généralement proche de celle d'une gaussienne bidimensionnelle) sont fonction de l'instrument et de l'échantillon. A cette tache, ainsi que sur toute la surface de l'appareil, se superpose un bruit de fond. Un enregistrement se résume donc pour nous en la donnée de 1024 mesures effectuées sur les 1024 cellules de l'appareil (Annexe 1).

La détermination de la position des atomes dans la maille du cristal étudié fait appel au calcul des intensités diffractées, c'est à dire aux comptages enregistrés par les cellules du multidétecteur touchées par la tache de BRAGG, une fois éliminée la contribution du bruit de fond à ces mesures.

Nous allons effectuer le travail selon la démarche suivante :

Dans un premier chapitre, nous étudions les caractéristiques de réponse

du multidétecteur lorsqu'on soumet celui-ci à un flux homogène de particules. Nous nous attachons en particulier à mettre en place des procédures visant à tester l'homogénéité et la stationnarité du processus de comptage dans ce cas précis.

L'objet du second chapitre est le calcul effectif des intensités diffractées. Nous supposons dans cette partie que le bruit de fond est homogène sur la surface entière de l'appareil, et sous cette hypothèse, établissons deux méthodes originales de traitement, selon que le bruit de fond moyen auquel est soumis l'appareil est faible ou au contraire élevé.

Avertissement :

Compte tenu de la nature du phénomène physique, tout se passe comme si l'on observait dans l'espace bidimensionnel H du multidétecteur de particules les réalisations d'un processus de Poisson de loi $P_{\lambda_{\theta}}$.

Désignons par A un borélien de \mathbb{R}^2 , et par N_A la variable de comptage du nombre de particules relevées sur A ; on a :

$$P(N_A = k) = \exp(-\lambda_{\theta}(A \cap H)) (\lambda_{\theta}(A \cap H))^k / k!$$

Puisque l'on a supposé que le bruit de fond est homogène, il vient de plus :

$$\lambda_{\theta}(A \cap H) = \int_{D_0} \mathbf{1}_{(A \cap H)} d\nu + \lambda_{\alpha}(A \cap H) \quad \text{où } \nu \text{ désigne la mesure de LEBESGUE, et } D_0 \text{ le sous ensemble de } H \text{ ne percevant pas la diffraction.}$$

Soient $\theta_j = (D_j, \alpha)$ diverses valeurs du paramètre θ associées aux différents sous ensembles D_j de H .

Déterminer l'intensité d'une diffraction équivaut à localiser le sous ensemble D_0 qui n'est pas touché par la tache de BRAGG, et peut encore se concevoir comme tester les hypothèses $\lambda_\theta = \lambda_{\theta_1}$ contre $\lambda_\theta = \lambda_{\theta_j}$.

Nous sommes alors bien évidemment tentés d'appliquer ici la théorie que nous avons développée dans la partie précédente de notre étude. Pour ce faire, il nous faut estimer la densité f de la mesure λ_θ , estimation qui pose un problème insoluble ! Explicitons en la teneur.

M. CURIONI [6], puis M. VAN BELLINGHEN [21], ont développé diverses méthodes permettant d'estimer la densité de l'intensité d'un processus : méthodes de l'histogramme, de la fenêtre mobile, du noyau ...

Afin d'illustrer le problème auquel nous nous trouvons confrontés, nous allons étudier l'une de ces procédures, celle dite de l'histogramme par exemple.

Soit p_1, \dots, p_n un n échantillon d'un processus de Poisson observé sur H , dont nous effectuons par ailleurs une partition en

$$H = [0, p[\times [0, q[= \bigcup_{r_1=1, k(n)} \bigcup_{r_2=1, l(n)} v_{r_1, r_2}, \text{ où}$$

$$v_{r_1, r_2} = [(r_1-1)p/k(n), r_1 p/k(n)[\times [(r_2-1)q/l(n), r_2 q/l(n)[$$

Désignons par $v_{r_1, r_2}^{(j)}$ le nombre de particules détectées sur le sous ensemble v_{r_1, r_2} au cours d'une réalisation p_j , et notons

$$\hat{f}_{nk(n)l(n)}(x) = \sum_{j=1, n} v_{r_1, r_2}^{(j)} (k(n) l(n) / npq)$$

Proposition :

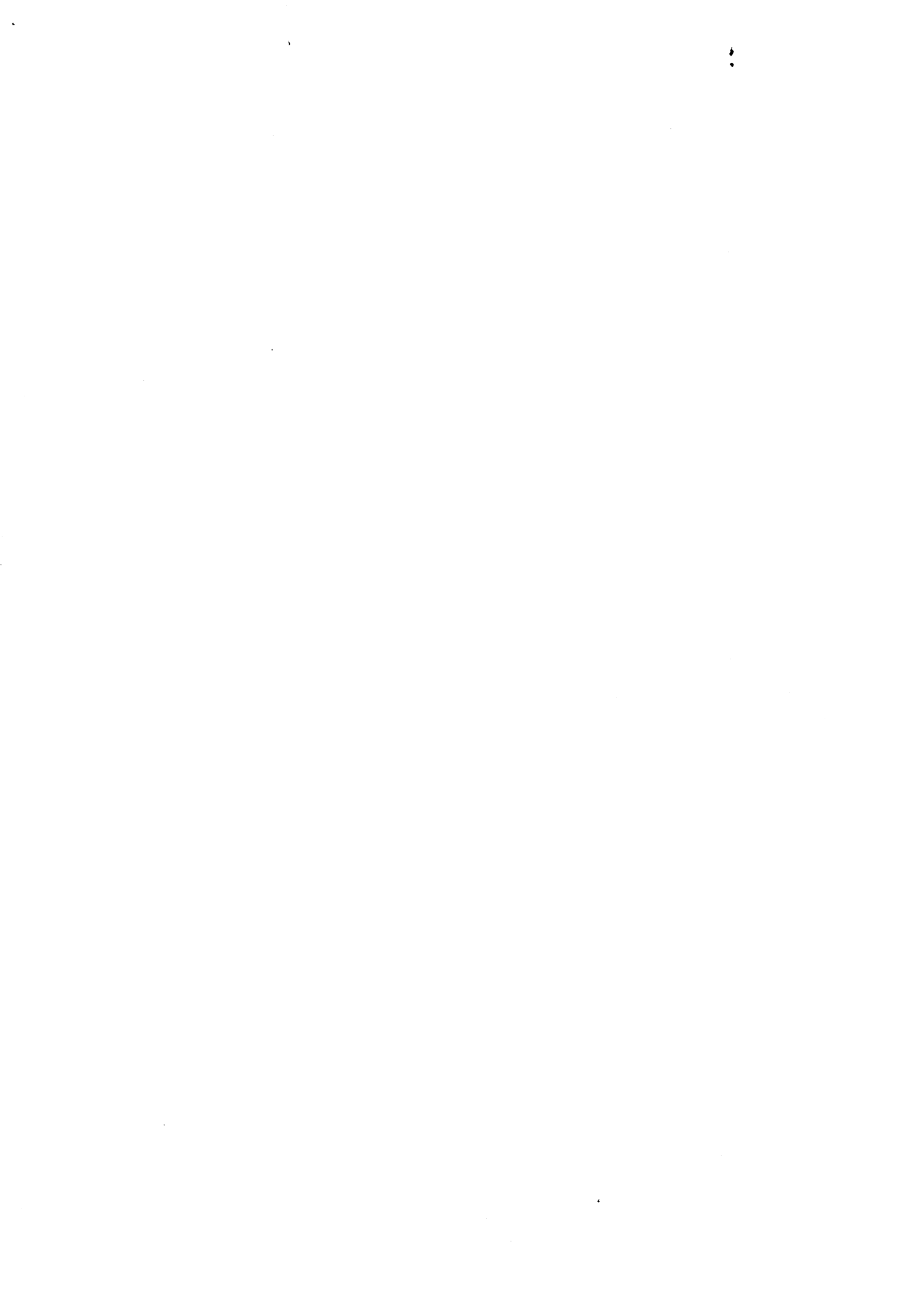
$\hat{f}_{nk(n)l(n)}(x)$ converge uniformément vers f si et seulement si les conditions suivantes sont vérifiées :

i) $k(n) l(n)$ tend vers l'infini lorsque $n \rightarrow \infty$

ii) $(k(n) l(n)/pq) = O(n/\log n)$

Une condition nécessaire de convergence est donc que $k(n) l(n)$ tende vers l'infini, ce qui s'avère physiquement irréalisable puisqu'on est limité, dans le choix de la partition de H , par la dimension élémentaire des cellules.

A notre connaissance, on rencontre le même problème dans les autres méthodes, et la solution consistant à effectuer des tests d'hypothèses sur les intensités du processus afin de déterminer l'intensité diffractée d'une tache de BRAGG ne peut, de ce fait, être envisagée.



CHAPITRE I

REPONSE DU MULTIDECTEUR A UN FLUX HOMOGENE DE PARTICULES

I.1. QUELQUES DEFINITIONS :

Lorsque l'on soumet le multidétecteur à un flux homogène de particules, on enregistre une série de 1024 observations dont la nature est telle qu'elle nous permet de supposer que les comptages effectués sur chaque cellule se font selon un processus de Poisson.

Désignons par $n_i(t, t+T)$ le comptage enregistré sur la cellule i du multidétecteur pendant la durée d'acquisition $[t, t+T]$. Nous modélisons $n_i(t, t+T)$ comme la réalisation de $N_i(t+T) - N_i(t)$, où l'on désigne par $(N_i(t), t \geq 0)$ un processus de Poisson d'intensité f_i ; cela équivaut encore à considérer chaque observation $n_i(t, t+T)$ comme la réalisation d'une variable aléatoire suivant une loi de Poisson de paramètre $\int_{t, t+T} f_i(u) du$.

D'une manière générale, les variables $(N_i(t+T) - N_i(t))$ et $(N_j(s+h) - N_j(s))$ sont indépendantes dès que $[t, t+T] \cap [s, s+h] = \emptyset$; nous supposons de plus

que les comptages effectués sur deux cellules i et j pendant une même durée d'acquisition sont les réalisations de variables indépendantes, c'est à dire que $(N_i(t+T) - N_i(t))$ et $(N_j(t+T) - N_j(t))$ sont indépendantes dès que i diffère de j .

Le multidétecteur de particules est dit homogène dans l'espace si chaque observation $n_i(t, t+T)$ ne dépend que de la période pendant laquelle on enregistre cette valeur, et non du positionnement de la cellule i sur l'appareil ; cela revient à énoncer :

Définition 1.1.1.

Le multidétecteur est homogène dans l'espace si chaque comptage $n_i(t, t+T)$ est la réalisation d'une variable suivant une loi de Poisson de paramètre $\int_{t, t+T} f(u) du$.

Nous allons dans ce chapitre mettre en place des procédures permettant de nous assurer de l'homogénéité dans l'espace du détecteur en réponse à un flux homogène de particules.

Un point également important de cette partie est la stabilité dans le temps du multidétecteur, dont voici une définition :

Définition 1.1.2 :

Le détecteur est stable en valeur relative si, pour une durée d'observation $h < T$, l'intensité $f_i(u)$ est constante sur $[t, t+h]$, donnée par $f_i(u) = \gamma_i(t)$.

Le détecteur est dit stable temporellement si $\gamma_i(t)$ est une fonction constante en t .

Il y a donc équivalence entre la stabilité dans le temps du

multidécteur et la stationnarité du processus de comptage lié à chaque cellule.

Enfin, l'appareil est à la fois homogène dans l'espace et stable dans le temps si chaque comptage $n_i[t, t+T]$ est la réalisation d'une variable suivant une loi de Poisson de paramètre γT .

1.2 ETUDE DE L'HOMOGENEITE DANS L'ESPACE DU DETECTEUR :

L'objet de ce paragraphe est de présenter deux méthodes permettant de tester l'hypothèse d'homogénéité dans l'espace du multidécteur en réponse à un flux homogène de particules.

Désignons par $(n_i(t, t+T))_{i=1, \dots, 1024}$ les observations effectuées pendant une période d'acquisition $[t, t+T]$. Le détecteur est homogène dans l'espace si chaque mesure $n_i(t, t+T)$ est la réalisation d'une variable aléatoire suivant une loi de Poisson de paramètre $\int_{t, t+T} f(u) du$, que nous notons, pour plus de commodité, $\lambda(t, t+T)$.

En conséquence, le multidécteur est dit homogène si les observations $(n_i(t, t+T))_{i=1 \dots 1024}$ constituent un 1024 échantillon d'une loi de Poisson de paramètre $\lambda(t, t+T)$ inconnu. Nous pouvons dès lors envisager comme test d'homogénéité dans l'espace un test d'adéquation de l'échantillon considéré à une loi de Poisson $P(\lambda(t, t+T))$.

1.2.1. Le Chi 2 d'adéquation comme test d'homogénéité:

On construit l'histogramme de l'échantillon $(n_i(t, t+T))_{i=1 \dots 1024}$ à partir des effectifs k_1, k_2, \dots, k_r de différentes classes. On cherche à savoir si l'échantillon considéré provient du modèle théorique de Poisson $P(\lambda(t, t+T))$

qui affecterait les probabilités p_1, \dots, p_r aux classes indiquées. Ces probabilités à priori sont inconnues, mais on les estime par les quantités $\hat{p}_1, \dots, \hat{p}_r$ calculées en prenant pour $\lambda(t, t+T)$ l'estimateur du maximum de vraisemblance $\hat{\lambda}(t, t+T) = 1/1024 \sum_{i=1, 1024} n_i(t, t+T)$.

Proposition :

Sous l'hypothèse d'homogénéité, $\sum_{j=1, r} (k_j - 1024\hat{p}_j)^2 / 1024 \hat{p}_j$ est la réalisation d'une variable aléatoire suivant la même loi que la variable $T_R = Z_1^2 + Z_2^2 + \dots + Z_{r-2}^2 + \mu Z_{r-1}^2$ où : $0 \leq \mu \leq 1$

$$\begin{array}{c} \Gamma \quad Z_1 \quad \Gamma \\ | \quad \quad | \quad \sim N_{r-1} (0, \mathbb{I}_{r-1}) \\ \text{L } Z_{r-1} \text{ J} \end{array}$$

Utiliser la loi du Chi2 à $r-2$ degrés de liberté pour la variable aléatoire T_R augmente le risque de première espèce, et prendre comme loi une loi du chi2 à $r-1$ degrés de liberté le diminue.

En première approximation, nous considérons que la variable T_R suit une loi du Chi2 à $r-1$ degrés de liberté.

Nous rejetons au niveau α l'hypothèse d'homogénéité dans l'espace si

$$\sum_{j=1, r} (k_j - 1024\hat{p}_j)^2 / 1024 \hat{p}_j > \lambda_{1-\alpha}$$

où λ_q désigne le quantile d'ordre q d'une loi du chi2 à $r-1$ degrés de liberté.

1.2.2 Test d'homogénéité basé sur les propriétés de l'indice de dispersion.

Une autre manière d'envisager ce problème consiste à mesurer le caractère agrégatif du processus étudié ; c'est ce à quoi nous allons nous consacrer dans la seconde partie de ce paragraphe.

Soit X un processus de Poisson homogène sur \mathbb{R}^2 , d'intensité λ . Désignons par D un borélien de \mathbb{R}^2 , et par $N(D)$ la variable de comptage du nombre d'occurrences sur D . L'événement $N(D) = n$ a lieu avec la probabilité $\exp(-\lambda(D)) \lambda(D)^n/n!$.

Conditionnellement à cet événement, les n points sont indépendamment et identiquement distribués suivant la loi uniforme sur D .

On dit qu'un processus est agrégé (respectivement régulier) si la probabilité de trouver un ou des atomes du processus dans un voisinage de x , sachant que la réalisation observée charge x est plus grande (respectivement plus petite) que pour les processus de Poisson homogènes ayant le même moment d'ordre 1.

Un procédé pour caractériser l'agrégation ou la régularité d'un processus, très utilisé par les écologistes (voir MEAD [13], PIELOU [16] ...) est le suivant :

Soient n_i les observations effectuées sur le domaine D . On calcule leur moyenne m , et leur variance V , puis le rapport $I = (n-1) V/m$, où n désigne le nombre d'observations sur D .

Dans le cas d'un processus homogène, le rapport $I/(n-1)$ est proche de 1. Une valeur très inférieure à 1 indique un processus régulier, une valeur très supérieure à 1 un processus agrégé. De plus, on a le résultat suivant (cf PIELOU [16]) :

Proposition

Si le processus de Poisson est homogène, la quantité I est en première

approximation la réalisation d'un variable aléatoire suivant une loi du Chi2 à n-1 degrés de liberté

Il s'agit pour nous de déterminer si le multidétecteur est homogène dans l'espace, auquel cas le processus de Poisson, défini dans l'espace bidimensionnel des cellules et lié aux comptages effectués sur un ensemble de capteurs, est homogène, ou s'il ne l'est pas, et dans ce cas le processus se trouve être agrégé ou encore régulier sur tout ou partie de l'appareil.

Désignons par $\bar{n}(t,t+T)$ et $v(t,t+T)$ les moyennes et variances des observations $(n_i(t,t+T))_{i=1,\dots,1024}$.

$$\bar{n}(t,t+T) = 1/1024 \sum_{i=1,1024} n_i(t,t+T)$$

$$v(t,t+T) = 1/1023 \sum_{i=1,1024} (n_i(t,t+T) - \bar{n}(t,t+T))^2$$

Corollaire :

Sous l'hypothèse d'homogénéité dans l'espace, la quantité $I = 1023 [V(t,t+T)/\bar{n}(t,t+T)]$ est, en première approximation, la réalisation d'une variable suivant une loi du Chi2 à 1023 degrés de liberté.

On rejettera donc l'hypothèse d'homogénéité si $I > \lambda_{1-\alpha}$, où l'on désigne par λ_q le quantile d'ordre q d'une loi du Chi2 à 1023 degrés de liberté.

Remarque

1) Dans la pratique, on utilise l'approximation de la loi du Chi2 par une gaussienne, selon la formule de FISHER. Nous considérons donc que $(2I)^{1/2} - (2(1023)-1)^{1/2}$ est distribuée comme une loi normale centrée réduite.

ii) Bien qu'attrayants par leur aspect élémentaire, ces tests ont le défaut de ne pas avoir une puissance élevée, et, notamment pour le second, une alternative beaucoup trop large.

1.3. ETUDE DE LA STABILITE DANS LE TEMPS DU DETECTEUR :

Nous supposons dans ce paragraphe que le multidétecteur est stable en valeur relative.

Nous allons dans une première partie définir un test permettant de juger de la stabilité dans le temps pour une cellule donnée du détecteur.

Dans une deuxième partie, nous montrons que, sous l'hypothèse supplémentaire d'homogénéité dans l'espace, nous pouvons étendre notre procédure de test à un domaine quelconque du détecteur. Enfin, nous consacrons la troisième partie de ce paragraphe à l'étude de la puissance des tests.

1.3.1 Stabilité dans le temps d'une cellule :

Soit $(n_i(t + (k-1)h_1, t+k h_1))_{k=1 \dots n}$ un enregistrement effectué sur une cellule quelconque du multidétecteur pendant les durées d'acquisition successives $(t + (k-1)h_1, t+k h_1)_{k=1 \dots n}$. Nous avons vu en introduction à ce chapitre que, sous l'hypothèse de stabilité relative du détecteur, nous modélisons $n_i(t + (k-1)h_1, t+kh_1)$ comme réalisation d'une variable aléatoire $N_i(t+ kh_1) - N_i(t+(k-1)h_1)$ suivant une loi de Poisson de paramètre $\gamma_i(t + kh_1)h_1$.

Sous l'hypothèse H_0 de stabilité dans le temps, les observations $(n_i(t + (k-1)h_1, t+kh_1))_{k=1 \dots n}$ constituent alors un échantillon d'une variable aléatoire suivant une loi de Poisson de paramètre $\gamma_i(t) h_1$.

La démarche que nous allons suivre est de définir une transformation de la variable aléatoire $N_i(t+kh_1) - N_i(t+(k-1)h_1)$ de manière à nous ramener à un problème de statistique sur échantillon gaussien. La justification de cette transformation repose sur le théorème suivant, énoncé par WILKS [23].

Théorème :

<< Soit (x_1, \dots, x_n) un n échantillon, réalisation d'une variable aléatoire de moyenne μ et de variance σ^2 finies. Soit g une fonction dérivable dans un voisinage de μ , et telle que la dérivée de g en μ soit non nulle.

Alors, la variable aléatoire $g(\bar{X})$ suit asymptotiquement, lorsque la taille n de l'échantillon tend vers l'infini, une loi normale, de moyenne $g(\mu)$, et de variance $(\sigma g'(\mu))^2/n$ >>.

Ainsi, sous l'hypothèse H_0 , le n échantillon dont nous disposons est la réalisation d'une loi de Poisson de moyenne $\gamma_i(t) h_1$ finie, et si l'on prend pour g l'application définie de \mathbb{R} dans \mathbb{R} par $g(x) = x^{1/2}$, le théorème de WILKS permet d'établir le corollaire suivant :

Corollaire :

Sous l'hypothèse H_0 de stabilité dans le temps du détecteur, la variable aléatoire

$$X^{1/2} = \left(\sum_{k=1,n} N_i(t+kh_1) - N_i(t+(k-1)h_1) \right)^{1/2},$$

suit asymptotiquement une loi normale, de moyenne $(n \gamma_i(t) h_1)^{1/2}$, et de variance $1/4$.

Considérons maintenant un second enregistrement effectué sur la même cellule i de l'appareil, pendant les intervalles de temps

$(s+(k-1)h_2, s+kh_2)_{k=1\dots n}$, que l'on note $(n_i(s+(k-1)h_2, s+kh_2))_{k=1\dots n}$. Sous l'hypothèse de stabilité dans le temps, ces observations constituent encore un n échantillon d'une variable poissonnienne, de paramètre $\gamma_i(s)h_2$; et, sous l'hypothèse H_0 , la variable

$$Y^{1/2} = (\sum_{k=1,n} N_i(s+kh_2) - N_i(s+(k-1)h_2))^{1/2}$$

suit asymptotiquement une loi normale, de moyenne $(n \gamma_i(s) h_2)^{1/2}$, et de variance $1/4$.

Posons $d = h_1 / h_2$

Nous avons énoncé comme définition de la stabilité dans le temps le critère $\gamma_i(t) = \text{cste}$; l'hypothèse H_0 se traduit donc par $\gamma_i(t) = \gamma_i(s)$, la contre-hypothèse à tester étant $\gamma_i(s) < \gamma_i(t)$ car en cas de non stabilité, il y a en général atténuation de l'intensité dans le cas des détecteurs à gaz. En effet mis à part les pannes, le défaut le plus classique est celui d'une diminution lente de la réponse globale ou locale (dégradation ou fuite de gaz de détection...)

Cette remarque nous permet d'énoncer la proposition suivante et son corollaire.

Proposition :

Sous l'hypothèse de stabilité dans le temps du multidétecteur de particules, la variable $X^{1/2} - (dY)^{1/2}$ suit asymptotiquement une loi normale, de moyenne nulle, et de variance $(1+d) / 4$.

Corollaire :

Le test de l'hypothèse $\gamma_i(s) = \gamma_i(t)$ contre $\gamma_i(s) < \gamma_i(t)$, défini par la région critique

$$R_c = \{ \omega, [2((X(\omega) + 3/8))^{1/2} - (d(Y(\omega) + 3/8))^{1/2}] / (1+d)^{1/2} \geq z_{1-\alpha} \},$$

où z_q désigne le quantile d'ordre q de la loi normale centrée réduite, est de niveau approximativement égal à $1 - \alpha$.

Nous reviendrons, dans la suite de cette étude, sur l'introduction du terme $3/8$ dans la définition de cette région critique, ainsi que sur la puissance de ce test. Nous nous proposons maintenant de définir un test de stabilité dans le temps effectué non plus cellule par cellule, mais pour une région D du multidétecteur de particules.

1.3.2. Stabilité dans le temps du multidétecteur :

Nous faisons dans ce paragraphe deux hypothèses :

- i) le multidétecteur est relativement stable dans le temps
- ii) l'appareil est également homogène dans l'espace.

Désignons par D un domaine quelconque du détecteur comprenant un nombre n_D de cellules.

Notons $n_D(t, t+h_1)$ le comptage effectué sur le domaine D pendant une durée d'acquisition $[t, t+h_1]$; on a :

$$n_D(t, t+h_1) = \sum_{i=1, 1024} n_i(t, t+h_1) \mathbf{1}_D(i).$$

Sous l'hypothèse faite à priori que le détecteur est homogène dans l'espace, nous avons établi déjà que les observations $(n_i(t, t+h_1))_{i=1, \dots, 1024}$ constituent un 1024 échantillon d'une variable de Poisson $N_i(t+h_1) - N_i(t)$, de moyenne $\int_{t, t+h_1} \gamma(u) du$; si, de plus, on suppose que l'appareil est relativement stable dans le temps, $N_i(t+h_1) - N_i(t)$ suit une loi de Poisson

de paramètre $\gamma(t) h_1$.

Le théorème de WILKS que nous avons rappelé précédemment permet alors d'établir :

Corollaire :

La variable aléatoire $X^{1/2} = (\sum_{i=1,1024} (N_i(t+h_1) - N_i(t)) \mathbb{1}_D(i))^{1/2}$ suit asymptotiquement une loi normale, de moyenne $(n_D \gamma(t) h_1)^{1/2}$, et de variance $1/4$.

Soit $n_D(s, s+h_2)$ un second enregistrement effectué sur le même domaine D pendant une nouvelle période d'acquisition $[s, s+h_2]$. La variable aléatoire

$Y^{1/2} = (\sum_{i=1,1024} (N_i(s+h_2) - N_i(s)) \mathbb{1}_D(i))^{1/2}$ suit asymptotiquement une loi normale, de moyenne $(n_D \gamma(s) h_2)^{1/2}$, et de variance $1/4$.

Posons $d = h_1/h_2$, $\gamma_D(t) = n_D \gamma(t)$, et $\gamma_D(s) = n_D \gamma(s)$.

L'hypothèse de stabilité dans le temps se traduit cette fois par $\gamma_D(t) = \gamma_D(s)$, la contre-hypothèse étant $\gamma_D(s) < \gamma_D(t)$, et l'on a :

Proposition :

Sous l'hypothèse de stabilité dans le temps du multidétecteur, la variable aléatoire $X^{1/2} - (dY)^{1/2}$ suit asymptotiquement une loi normale, centrée, et de variance $(1+d)/4$.

Comme corollaire à cette proposition, il vient :

Corollaire :

Le test de l'hypothèse $\gamma_D(t) = \gamma_D(s)$ contre $\gamma_D(t) > \gamma_D(s)$, défini par la région critique :

$$R_c = \{ \omega, [2((X(\omega) + 3/8))^{1/2} - (d(Y(\omega) + 3/8))^{1/2}] / (1+d)^{1/2} \geq z_{1-\alpha} \},$$

où z_q désigne le quantile d'ordre q de la loi normale centrée réduite, est approximativement UMPB de niveau α .

On introduit là encore un terme additif dans la définition de la région critique, dont on peut dire en simplifiant qu'il a pour effet de diminuer la variance de la statistique du test. C'est un point sur lequel nous allons revenir dans l'étude que nous développons à présent.

1.3.3. Quelques remarques générales quant à la puissance des tests de stabilité :

Supposons que pour une cellule donnée, ou plus généralement pour une région D quelconque du multidétecteur, on ait effectué deux enregistrements $n_D(t, t+h_1)$ et $n_D(s, s+h_2)$ pour des durées d'acquisition distinctes.

Sous l'hypothèse à priori d'une stationnarité segmentée, donc si l'on suppose que le multidétecteur est relativement stable dans le temps, les observations $n_D(t, t+h_1)$ et $n_D(s, s+h_2)$ sont modélisées comme réalisations de deux variables X et Y suivant des lois de Poisson de moyenne respective $\gamma_D(t) h_1$ et $\gamma_D(s) h_2$.

Nous avons établi alors que l'hypothèse H_0 à tester se traduit par $\gamma_D(s) = \gamma_D(t)$, la contre-hypothèse étant $\gamma_D(s) < \gamma_D(t)$.

Nous allons dans ce paragraphe comparer les procédures existant pour effectuer un tel test ; nous énonçons également une formule permettant de

déterminer la durée d'acquisition nécessaire à l'obtention d'une puissance donnée pour le test que nous avons défini.

Il existe un test uniformément le plus puissant parmi les tests sans biais pour tester l'égalité $\gamma_D(s) = \gamma_D(t)$, énoncé par LEHMANN dans ([12], p141). Ce test est basé sur la loi binomiale, et est un test conditionnel. Etant défini à partir d'une loi discrète, il offre le désavantage d'avoir un seuil de signification réel inférieur en général au seuil théorique désiré.

Dans le cas où les paramètres h_1 et h_2 sont égaux, un test approximatif de l'hypothèse H_0 a été proposé par SICHEL [19], qui s'avère supérieur au test exact car son niveau de signification réel est plus proche du seuil théorique que celui du test exact. Néanmoins, aucune comparaison de puissance n'est faite dans cet article.

SHIUE et BAIN [18] ont généralisé ce test au cas où h_1 diffère de h_2 , et en ont étudié la puissance

Posons $d = h_1 / h_2$, et $\rho = \gamma_D(t) / \gamma_D(s)$.

Le test de l'hypothèse $\gamma_D(s) = \gamma_D(t)$ contre $\gamma_D(s) < \gamma_D(t)$ établi par SHIUE et BAIN est défini par la donnée de la région critique :

$$Rc = \{\omega, Z_1(\omega) = [(X(\omega) - d Y(\omega)) / (d X(\omega) + Y(\omega))]^{1/2} \geq z_{1-\alpha}\},$$

où z_q est le quantile d'ordre q de la loi normale centrée réduite.

La puissance exacte de ce test se calcule aisément ; de plus, en utilisant l'approximation gaussienne pour la loi de Z_1 , les auteurs montrent que pour des valeurs fixées de d et de ρ , le choix approprié du paramètre $h_2 \gamma_D(s)$ pour atteindre une puissance p est donné par :

$$h_2 \gamma_D(s) = (1 + \rho/d) \left(\frac{(1+d\rho)}{(d+\rho)} \right)^{1/2} z_{1-\alpha} + z_p)^2 / (\rho-1)^2 \quad (1)$$

Ainsi, si l'on dispose d'une estimation de la vraie valeur du paramètre $\gamma_D(s)$, les temps d'enregistrement h_1 et h_2 peuvent être déterminés pour des valeurs données de ρ , d , α , et p .

Bien que le niveau de signification et la puissance du test ainsi défini soient approximativement corrects, le vrai niveau est un peu supérieur au théorique pour une valeur petite de d et un peu inférieur pour une grande valeur de d , avec pour conséquence de donner des valeurs de h_2 $\gamma_D(s)$ inférieures ou supérieures, selon l'ordre de grandeur de d , aux valeurs effectivement nécessaires à l'obtention d'une puissance donnée.

Ceci peut s'expliquer d'une part par le fait que la convergence de la variable Z_1 vers la loi normale est lente, et d'autre part par le fait que la variance de cette variable est estimée de manière récursive, ce qui conduit à une approximation grossière du paramètre $h_2 \gamma_D(s)$.

Le test de l'hypothèse $\gamma_D(s) = \gamma_D(t)$ contre $\gamma_D(s) < \gamma_D(t)$ que nous avons mis en place repose sur la donnée de la région critique :

$$R_c = \{\omega, Z_2(\omega) = [2((X(\omega) + 3/8)^{1/2} - (d(Y(\omega) + 3/8))^{1/2})] / (1+d)^{1/2} \geq z_{1-\alpha}\},$$

où z_q désigne le quantile d'ordre q de la loi normale centrée réduite.

Il est connu que la transformation de stabilisation de la variance, utilisée aux paragraphes 1.3.1 et 1.3.2 accélère la convergence en loi vers une loi normale centrée réduite : il s'en suit que la variable Z_2 converge plus rapidement que Z_1 vers la loi gaussienne lorsque $\lambda = \min(h_2 \gamma_D(s), h_1 \gamma_D(s))$ tend vers l'infini.

L'introduction du terme $3/8$ dans la définition de notre test est justifiée par la remarque suivante :

Remarque :

L'espérance de Z_2 , que la transformation contienne ou non le terme $3/8$ est donnée, à l'ordre $O(\lambda^{-1/2})$, par

$$E(Z_2) = 2(e^{1/2} - 1) (h_2 \gamma_D(s) d)^{1/2} / (1+d)^{1/2}.$$

Par contre, la variance diminue par l'effet de ce terme, et passe de l'ordre $1 + O(\lambda^{-1})$ à $1 + O(\lambda^{-2})$.

Etablissons la preuve de ce résultat.

Soit Z_a la variable définie par :

$$Z_a(\omega) = 2((X(\omega) + a)^{1/2} - (d(Y(\omega) + a))^{1/2}) / (1+d)^{1/2}.$$

Un rapide calcul montre que l'espérance de la variable $(X + a)^{1/2}$ est donnée, à l'ordre $O((h_1 \gamma_D(t))^{-5/2})$ par :

$$E((X + a)^{1/2}) = (h_1 \gamma_D(t))^{1/2} + [(4a-1) / 8(h_1 \gamma_D(t))^{1/2}] - [(16a^2-24a+7)/128(h_1 \gamma_D(t))^{3/2}]$$

A l'ordre $O(\lambda^{-1/2})$, on a donc $E((X + a)^{1/2}) = (h_1 \gamma_D(t))^{1/2}$.

Un calcul identique permet d'établir l'égalité

$$E((Y + a)^{1/2}) = (h_2 \gamma_D(s))^{1/2} + O(\lambda^{-1/2})$$

il vient alors :

$$\begin{aligned} E(Z_a) &= 2((h_1 \gamma_D(t))^{1/2} - (h_2 \gamma_D(s))^{1/2}) / (1+d)^{1/2} + O(\lambda^{-1/2}) \\ &= 2(h_2 \gamma_D(s) d)^{1/2} (e^{1/2} - 1) / (1+d)^{1/2} + O(\lambda^{-1/2}) \end{aligned}$$

L'espérance de la variable Z_a ne dépend donc pas, à l'ordre $O(\lambda^{-1/2})$, de la constante a .

Développons maintenant le calcul de la quantité $\text{Var}(Z_a)$; on a bien sûr :

$$\text{Var}(Z_a) = 4(\text{Var}((X+a)^{1/2}) + d \text{Var}((Y+a)^{1/2})) / (1+d),$$

et $\text{Var}((X + a)^{1/2}) = E(X+a) - (E[(X+a)^{1/2}])^2$

A l'ordre $O((h_1 \gamma_D(t))^{-2})$, $(E[(X+a)^{1/2}])^2$ s'écrit :

$$(E((X+a)^{1/2}))^2 = h_1 \gamma_D(t) + a - 1/4 - (6-16a)/(64 h_1 \gamma_D(t))$$

D'où l'égalité

$$\text{Var}(X+a) = 1/4 + (6-16a)/(64 h_1 \gamma_D(t)) + O((h_1 \gamma_D(t))^{-2})$$

De la même manière, il vient :

$$\text{Var}(Y+a) = 1/4 + (6-16a)/(64 h_2 \gamma_D(s)) + O((h_2 \gamma_D(s))^{-2})$$

La variance de la variable Z_a s'écrit alors, à l'ordre $O(\lambda^{-2})$, comme :

$$\text{Var}(Z_a) = 1 + (6-16a) \left[\frac{1}{16 h_1 \gamma_D(t)} + \frac{d}{16 h_2 \gamma_D(s)} \right] / (1+d) + O(\lambda^{-2})$$

Il résulte de cette égalité que, pour un choix de $a = 3/8$, la variance de Z_a passe de l'ordre $1 + O(\lambda^{-1})$ à $1 + O(\lambda^{-2})$.

Calculons maintenant la puissance exacte de ce dernier test :

$$p = P \left[2((X + 3/8)^{1/2} - (d(Y+3/8))^{1/2}) \geq z_{1-\alpha} (1+d)^{1/2} \right]$$

En conditionnant par rapport à Y , cette égalité devient :

$$p = \sum_{k=0, \infty} P \left[(X + 3/8)^{1/2} \geq 1/2 z_{1-\alpha} (1+d)^{1/2} + (d(k+3/8))^{1/2} / Y=k \right] P[Y=k]$$

Soit encore, avec

$$l_0 = \text{partie entière} \left[(0,5 z_{1-\alpha} (1+d)^{1/2} + (d(k+3/8))^{1/2})^2 - 3/8 \right]$$

$$p = \sum_{k=0, \infty} \sum_{l=l_0+1, \infty} \exp(-\gamma_D(s) h_2 - \gamma_D(t) h_1) (\gamma_D(s) h_2)^k (\gamma_D(t) h_1)^l / k! l!$$

En utilisant la loi asymptotique comme loi de Z_2 , une approximation de cette puissance est donnée par

$$p = 1 - \Phi \left(z_{1-\alpha} - 2(p^{1/2}-1) (h_2 \gamma_D(s) d)^{1/2} / (1+d)^{1/2} \right),$$

où Φ désigne la fonction de répartition d'une loi normale centrée réduite.

Le choix approprié du paramètre $h_2 \gamma_D(s)$ pour atteindre la puissance p est alors défini par l'égalité

$$h_2 \gamma_D(s) = (z_{1-\alpha} + z_p)^2 (1+d) / 4d (p^{1/2}-1)^2 \quad (2)$$

Pour des valeurs données de α , p , d et ρ , nous avons calculé :

- le niveau de signification et la puissance exacts du test de SHIUE et BAIN (notés $\alpha(A)$ et $p(A)$)
- le niveau de signification et la puissance exacts du test que nous avons mis en place (notés $\alpha(B)$ et $p(B)$)

Dans le tableau ci-joint ne figurent que les valeurs calculées pour $\alpha = 0.05$ et $p = 0.9$, mais les résultats obtenus pour d'autres valeurs des paramètres α et p sont très similaires. Nous donnons également dans ce tableau les estimations du paramètre $h_2 \gamma_D(s)$ établies par l'équation (1), puis par l'équation (2).

D'une manière générale, le niveau de signification et la puissance de notre test sont très proches des valeurs requises pour α et p , plus que ne le sont le seuil et la puissance du test défini par SHIUE et BAIN. Les durées d'enregistrement que nous préconisons pour atteindre la puissance p sont donc également plus proches de la réalité que celles qui découlent de la formule de SHIUE et BAIN.

d	p	$\alpha(A)$	P(A)	$h_2 \gamma_{D(s)}(1)$	$h_2 \gamma_{D(s)}(2)$	P(B)	$\alpha(B)$
0.1	4	0.063	0.915	22.92	23.56	0.903	0.047
	3	0.059	0.910	42.69	43.96	0.901	0.047
	2	0.057	0.907	133.9	137.31	0.902	0.048
0.5	4	0.059	0.922	6.89	6.42	0.902	0.055
	3	0.054	0.915	12.49	11.99	0.902	0.051
	2	0.052	0.908	37.90	37.45	0.900	0.049
1	4	0.052	0.922	4.76	4.28	0.896	0.051
	3	0.049	0.916	8.57	7.99	0.898	0.050
	2	0.049	0.908	25.70	24.97	0.900	0.049
2	4	0.042	0.920	3.62	3.21	0.901	0.049
	3	0.041	0.913	6.51	5.99	0.896	0.053
	2	0.047	0.905	19.48	18.72	0.898	0.050
4	4	0.038	0.912	3.01	2.68	0.896	0.048
	3	0.038	0.908	5.43	4.99	0.896	0.050
	2	0.044	0.902	16.30	15.60	0.898	0.049

**Comparaison du niveau de signification et de la puissance
des tests étudiés**

CHAPITRE II

MODE DE CALCUL DE L'INTENSITE INTEGREE **D'UNE DIFFRACTION DE BRAGG**

II.1. POSITION DU PROBLEME ET DESCRIPTION DU TRAITEMENT **ENVISAGE**

II.1.1. Position du problème et définitions :

Lorsqu'un plan cristallin est en position de diffraction, on enregistre sur une partie du multidétecteur de particules un signal que l'on a coutume de nommer tache de BRAGG.

Précisons que l'appareil considéré est ainsi conçu que, sauf recouvrements de pics voisins, il n'y a qu'une seule réflexion qui soit active. Nous faisons l'hypothèse qu'il n'y a pas recouvrements, et donc que nous n'observons qu'une seule tache de BRAGG par enregistrement.

Sur l'ensemble des capteurs vient se superposer un bruit de fond que l'on considère en première approximation comme un flux homogène de particules.

Désignons par $(n_i(t,t+T))_{i=1\dots 1024}$ un enregistrement effectué lorsqu'un noeud de BRAGG est en position de diffraction. Nous pouvons décomposer

chaque mesure $n_i(t, t+T)$ en

$$n_i(t, t+T) = n_i^B(t, t+T) + n_i^S(t, t+T)$$

où $n_i^B(t, t+T)$ est la contribution du bruit de fond au comptage $n_i(t, t+T)$, et

où $n_i^S(t, t+T)$ est l'intensité de la diffraction perçue par la cellule i . (avec éventuellement $n_i^S(t, t+T) = 0$ si cette cellule n'est pas touchée par la tache de BRAGG)

Définition :

On appelle intensité intégrée de la diffraction la quantité

$$I = \sum_{i=1, \dots, 1024} n_i^S(t, t+T)$$

On désigne par rapport signal/bruit le rapport

$$S/B = \sum_{i=1, \dots, 1024} n_i^S(t, t+T) / \sum_{i=1, \dots, 1024} n_i^B(t, t+T)$$

Décrivons maintenant en quelques mots le modèle mathématique qui nous a paru raisonnable pour représenter ce problème physique.

Sous l'hypothèse a priori que le multidétecteur est homogène dans l'espace lorsqu'il est soumis à un flux homogène de particules, les comptages $n_i^B(t, t+T)_{i=1, \dots, 1024}$ constituent un 1024 échantillon d'une loi de Poisson de paramètre $\lambda(t, t+T)$.

Compte tenu de la nature du phénomène, nous modélisons là encore $n_i(t, t+T)$ comme réalisation d'une loi de Poisson de moyenne $\lambda(t, t+T) + s_i(t, t+T)$, avec éventuellement $s_i(t, t+T) = 0$ pour des comptages ne percevant pas la diffraction.

II.1.2. Mode de calcul adopté pour extraire l'intensité intégrée d'une diffraction :

Avant d'exposer la méthode que nous avons retenue pour effectuer le

calcul de l'intensité intégrée d'une tache de BRAGG, explicitons les raisons de notre choix, et étudions pour ce faire les limites de la procédure employée dans [8].

a) Mode de calcul usité actuellement :

Le multidétecteur de particules étant soumis à un bruit de fond dont on suppose dans cet article qu'il est homogène, on enregistre un 1024 échantillon d'une loi de Poisson de paramètre μ , et l'on a :

$$\forall x \in \mathbb{N}, p(x) = P\{n_i(t, t+T) = x\} = e^{-\mu} \mu^x / x!$$

$$y(x) = L_N(x! p(x)) = -\mu + x L_N(\mu) \quad (1)$$

Désignons par $f(x)$ la fréquence empirique de x , et par $\hat{y}(x)$ la valeur $\hat{y}(x) = L_N(x! f(x))$; soit $y = ax + b$ la droite d'ajustement aux moindres carrés des couples $(x, \hat{y}(x))$.

Les auteurs retiennent trois estimateurs possibles du paramètre μ :

- L'estimateur du maximum de vraisemblance,

$$\hat{\mu}_c = 1/1024 \sum_{i=1,1024} n_i(t, t+T)$$

- L'estimateur calculé à partir de l'ordonnée à l'origine de la droite définie par l'équation (1),

$$\hat{\mu}_a = -L_N(N(0)/1024)$$

où $N(0)$ désigne le nombre de cellules sur lesquelles on a enregistré un comptage nul.

- Enfin, un dernier estimateur établi à partir de la droite d'ajustement aux moindres carrés,

$$\hat{\mu}_b = \exp(a).$$

Lorsqu'une tache de BRAGG se superpose au bruit de fond, il n'est plus

envisageable d'utiliser $\hat{\mu}_c$ comme estimateur de μ ; on retient alors l'un ou l'autre des estimateurs $\hat{\mu}_A$ ou $\hat{\mu}_b$ selon le critère

$$\hat{\mu} = \begin{cases} \hat{\mu}_a & \text{si } (N(0) / 1024) > k \\ \hat{\mu}_b & \text{sinon} \end{cases}$$

où k est un seuil fixé par les utilisateurs.

Cela signifie donc que l'estimation du bruit moyen est effectuée à partir des seuls comptages nuls dès que $N(0)$ est suffisamment grand.

Une fois le bruit de fond moyen estimé par $\hat{\mu}$, le calcul de l'intensité intégrée est fait suivant la formule :

$$\hat{I} = \sum_{i=1, 1024} n_i(t, t+T) - 1024 \hat{\mu}$$

Nous pouvons formuler les remarques suivantes :

i) Quel que soit l'estimateur retenu pour le paramètre μ , on lui préférerait l'estimateur du maximum de vraisemblance calculé à partir des cellules sur lesquelles $n_i^S(t, t+T) = 0$.

ii) Nonobstant la médiocre qualité de l'estimateur $\hat{\mu}$, il est à noter que la variance de \hat{I} croît avec le nombre de cellules sur lesquelles s'effectue la sommation (elle porte ici sur les 1024 cellules du détecteur). La diffraction n'occupant d'une manière générale que trente pour cent du multidétecteur, il est désavantageux d'utiliser dans le calcul de \hat{I} un nombre de cellules inutilement élevé.

Il convient en conclusion, pour optimiser ce calcul, de définir la frontière tache de BRAGG-bruit de fond du mieux possible. C'est ce à quoi nous nous sommes attachés dans notre étude.

b) descriptif de la procédure que nous envisageons :

Exposons brièvement qu'elle est la démarche que nous avons suivie.

Nous cherchons en tout premier lieu à localiser le domaine D du détecteur occupé par la tache de BRAGG. Une fois ce domaine déterminé, et puisque les comptages $n_j^S(t, t+T)$ sont nuls sur les cellules de D^C , nous estimons le bruit de fond moyen $\lambda(t, t+T)$ par son estimateur du maximum de vraisemblance sur D^C :

$$\hat{\lambda}(t, t+T) = \sum_{k=1, 1024} n_k(t, t+T) \mathbf{1}_{D^C}(k) / \sum_{k=1, 1024} \mathbf{1}_{D^C}(k)$$

Enfin nous estimons l'intensité intégrée de la diffraction par la quantité

$$\hat{I} = \sum_{k=1, 1024} n_k(t, t+T) \mathbf{1}_D(k) - \hat{\lambda}(t, t+T) \sum_{k=1, 1024} \mathbf{1}_D(k)$$

L'estimateur de I ainsi défini est de variance inférieure à celui utilisé dans [8].

Les procédures que nous allons proposer pour estimer l'intensité intégrée d'une tache de BRAGG ont donc pour principe de base de localiser l'ensemble des cellules occupées par cette tache, localisation qu'il est apparue très difficile à déterminer avec précision. En effet, si l'on sait que la restriction à la partie D^C du processus de comptage est homogène, et donc que les observations effectuées sur ce domaine constituent un échantillon d'une même loi, on ne sait rien à priori des comptages effectués sur D ; pour cette raison, des tests d'hypothèses sont délicats à mettre en place, et nous avons choisi de procéder de la manière suivante :

Nous nous efforçons dans un premier temps de détecter le "noyau de la tache de BRAGG", partie de l'appareil sur laquelle le processus est en général agrégé en ce sens que, si l'une des cellules enregistre un comptage élevé, les cellules voisines enregistrent également un nombre élevé de particules.

Ensuite, et compte tenu de l'unicité de la tache de BRAGG, nous procédons à une phase d'agglomération autour du noyau, qui, en quelques mots, consiste à réexaminer l'appartenance à la partie D du détecteur des cellules voisines des cellules du noyau.

La complexité du problème traité impose l'emploi d'algorithmes d'analyse des données, fondés sur des tests successifs d'hypothèses multiples dont il est difficile de contrôler les seuils, et qui perdent tout caractère d'optimalité.

Nous décrivons deux méthodes que nous avons développées plus précisément : Dans la première de ces méthodes, nous ne faisons aucune hypothèse à priori quant à la valeur moyenne du bruit de fond : à contrario, nous envisageons dans la seconde le cas où l'intensité du bruit de fond est élevée, et donc où l'on peut se ramener à un problème sur échantillons gaussiens.

II.2. CALCUL DE L'INTENSITE INTEGREE D'UNE DIFFRACTION - CAS GENERAL :

II.2.1. Extration du noyau du signal :

En présence d'une diffraction de BRAGG, le multidétecteur comprend, comme on l'a souligné au paragraphe II.1.2, deux parties qu'il nous faut distinguer.

- Un domaine D^C soumis, au seul bruit de fond, et sur lequel le processus de Poisson lié aux comptages est homogène.
- Un domaine D recouvert par la tache de BRAGG et dont on sait qu'il contient un sous ensemble au moins sur lequel le processus est agrégé.

Résoudre le problème de classification d'une cellule i du détecteur dans la partie D^C ou dans la partie "noyau du signal" revient donc approximativement à répondre à la question suivante :

La cellule i appartient-elle à une partie sur laquelle le processus est homogène ou, au contraire, agrégé ?

Nous avons résolu déjà, dans le premier chapitre de cette étude appliquée, un problème du même ordre, utilisant pour ce faire les

propriétés de l'indice de dispersion. Nous allons user dans cette partie d'une procédure quelque peu similaire à celle que nous avons alors mise en place.

Soit C_1, C_2, \dots, C_N une partition du multidétecteur en 128 rectangles contigus et égaux comprenant chacun 8 cellules. Désignons par I_j l'indice de dispersion calculé à partir des comptages $(n_{ij}(t, t+T)_{i=1 \dots 8})$ effectués sur la plage C_j , et qui, rappelons le, est donné par :

$$I_j = \sum_{i=1,8} [(n_{ij}(t, t+T) - \bar{n}_j(t, t+T))^2] / \bar{n}_j(t, t+T),$$

où
$$\bar{n}_j(t, t+T) = 1/8 \sum_{i=1,8} n_{ij}(t, t+T).$$

Sous l'hypothèse d'homogénéité du processus de Poisson sur la plage C_j , I_j est en première approximation la réalisation d'une variable aléatoire suivant une loi du Chi2 à 7 degrés de liberté.

Cela nous permet de proposer le

Test :

On rejette l'hypothèse $C_j \subset D^C$ au profit de $C_j \cap D \neq \emptyset$ si

$$I_j > z_{1-\alpha_1},$$

où z_q désigne le quantile d'ordre q d'une loi du Chi2 à 7 degrés de liberté.

De plus, une valeur élevée de I_j indique que le processus est agrégé sur C_j ; on procède alors de la manière suivante :

Soit α_2 un niveau choisi arbitrairement très faible ; nous affectons chaque cellule de la plage C_j

- au sous ensemble D^C si $I_j \leq z_{1-\alpha_1}$

- au sous ensemble D si $I_j > z_{1-\alpha_2}$

Remarques

a) On doit tenir compte du fait que, sur certaines parties de D, le processus peut être homogène. Cela peut en particulier être le cas lorsque le rapport signal-bruit est très faible. Le test que nous effectuons conduit alors à ne pas classer ces plages dans l'ensemble D, voire à les affecter à l'ensemble D^C (dans ce dernier cas, cela ne modifie pas outre mesure la valeur estimée du bruit de fond moyen, puisque l'intensité de la diffraction est alors en général très faible).

b) nous nous donnons une valeur très élevée de $z_{1-\alpha_2}$, afin de minimiser le risque qui consiste à classer à tort une cellule dans D. En contre-partie, il peut se faire qu'on ne parvienne pas, à ce niveau α_2 , à détecter une plage de D ; dans ce cas, on retient comme noyau de la diffraction la plage enregistrant les comptages les plus élevés.

Développons maintenant la phase d'agglomération

II.2.2. Agglomération des cellules voisines du noyau de la tache de BRAGG :

Nous allons nous ramener à un problème de tests sur deux échantillons poissonniens. La procédure que nous établirons alors repose sur la proposition suivante, énoncée par WILKS dans [23].

Proposition :

< < Soient $(x_{11}, \dots, x_{1n_1})$ et $(x_{21}, \dots, x_{2n_2})$ des échantillons indépendants de variables aléatoires suivant des lois de Poisson de paramètres μ_1 et μ_2 inconnus. Désignons par \bar{x}_1 la moyenne du premier

échantillon, par \bar{x}_2 la moyenne du second, et notons \bar{x} la quantité

$$\bar{x} = (n_1 \bar{x}_1 + n_2 \bar{x}_2) / (n_1 + n_2)$$

Soit à tester l'hypothèse $H_0 : \mu_1 = \mu_2$ contre $\mu_1 < \mu_2$. On montre que le rapport de vraisemblance λ est donné par l'égalité :

$$\lambda = \bar{x}^{(n_1+n_2)} / \bar{x}_1^{n_1} \bar{x}_2^{n_2}$$

De plus, la distribution asymptotique de $-2 \log \lambda$ est, sous l'hypothèse H_0 , une loi du Chi2 à un degré de liberté $> >$.

Soit C_k une plage du multidétecteur non affectée encore à la partie "Tache de BRAGG", mais voisine d'une plage que nous savons être, à ce niveau de notre étude, dans l'ensemble D .

Désignons par $(n_{ik}(t,t+T))_{i=1...8}$ les comptages enregistrés par les cellules de C_k .

Soient $n_{D^c C_1}(t,t+T), \dots, n_{D^c C_{n_1}}(t,t+T)$ les observations effectuées sur les n_1 cellules de l'ensemble $D^c \cap C_k^c$; ces observations constituent un n_1 échantillon d'une loi de Poisson de paramètre $\lambda(t,t+T)$.

a) si, au seuil α_1 , nous avons accepté l'hypothèse $C_k \subset D^c$, le processus de Poisson lié aux comptages sur C_k est homogène, et les observations $(n_{ik}(t,t+T))_{i=1...8}$ constituent un deuxième échantillon, de taille 8, d'une loi de Poisson de paramètre $\lambda(t,t+T) + s(t,t+T)$, et indépendant du premier.

De plus, et puisque le processus est homogène sur C_k , on a soit $C_k \subset D$, soit $C_k \subset D^c$, et tester l'hypothèse $s(t,t+T) = 0$ contre $s(t,t+T) > 0$ revient à tester $C_k \subset D^c$ contre $C_k \subset D$.

Nous nous appuyons sur la proposition de WILKS pour proposer :

Test 1

On rejette l'hypothèse $C_k \subset D^c$ si

$$-2 L_n((\bar{n}^{(n_1+8)\bar{n}}) / (\bar{n}_{D^c}^{n_1} \bar{n}_{D^c} \bar{n}_k^{8\bar{n}_k})) > z_{1-\alpha_3},$$

où z_q est le quantile d'ordre q d'une loi du Chi2 à un degré de liberté, et où :

$$\bar{n}_k = 1/8 \sum_{i=1,8} n_{ik}(t, t+T)$$

$$\bar{n}_{D^c} = 1/n_1 \sum_{i=1, n_1} n_{D^c i}(t, t+T)$$

$$\bar{n} = (8 \bar{n}_k + n_1 \bar{n}_{D^c}) / (n_1 + 8)$$

b) Si, au seuil α_1 , nous avons rejeté l'hypothèse $C_k \subset D^c$, et puisque le processus n'est pas agrégé sur C_k (sinon, on aurait $C_k \subset D$), nous admettons que C_k peut être scindé en deux sous groupes homogènes.

Soient $(n'_{ik}(t, t+T))_{i=1 \dots 8}$ les comptages ordonnés effectués sur C_k . Nous regroupons ces valeurs en deux sous groupes homogènes $((n'_{1k}(t, t+T), \dots, (n'_{pk}(t, t+T))$ et $((n'_{p+1k}(t, t+T), \dots, (n'_{8k}(t, t+T))$ selon l'indice p qui rend minimum la quantité :

$$p = \text{Min}_{1 \leq s \leq 8} ((s-1)(v_s / \bar{n}_s) + (8-s-1)(v_{8-s} / \bar{n}_{8-s})),$$

où \bar{n}_s et v_s désignent les moyennes et variances des s premières observations, \bar{n}_{8-s} et v_{8-s} les moyennes et variances des comptages résiduels.

La valeur p est donc choisie de manière à rendre minimum la somme des indices de dispersion de chaque sous groupe.

Désignons par C_k^0 et C_k^1 respectivement le premier et le second des sous ensembles ainsi générés.

Le processus de comptage restreint à chacun des ensembles C_k^i , pour $i = 0$ ou 1 , est homogène, et nous avons soit $C_k^i \subset D^C$, soit $C_k^i \subset D$.

Cela nous permet de suggérer

Test 2 :

On rejette l'hypothèse $C_k^i \subset D^C$ au profit de l'hypothèse $C_k^i \subset D$ si

$$-2 L_n \left[\frac{(\bar{n}_i (n_1 + i(8-p) + (1-i)p))}{(\bar{n}_{D^C} n_1 \bar{n}_{D^C} \bar{n}_{ki} ((8-p) + (1-i)p) \bar{n}_{ki})} \right] > z_{1-\alpha_3},$$

où \bar{n}_{ki} désigne la moyenne de l'un ou l'autre des sous ensembles C_k^i , et où \bar{n}_i est donné par :

$$\bar{n}_i = (n_1 \bar{n}_{D^C} + (i(8-p) + (1-i)p) \bar{n}_{ki}) / (n_1 + i(8-p) + (1-i)p)$$

Traitement effectif :

Décrivons maintenant de manière plus pragmatique notre démarche.

On affecte dans un premier temps chaque plage C_j du multidétecteur à la partie D ou à la partie D^C de l'appareil, selon le critère :

$$C_j \subset D^C \text{ si } I_j < z_{1-\alpha_1} \text{ pour } \alpha_1 = 0.3$$

$$C_j \subset D \text{ si } I_j > z_{1-\alpha_2} \text{ pour } \alpha_2 = 5 \cdot 10^{-4}$$

(on se désintéresse momentanément des plages pour lesquelles $I_j \in [z_{1-\alpha_1}, z_{1-\alpha_2}]$).

On établit alors à partir des n_1 cellules classées dans D^C une première estimation du bruit de fond moyen \bar{n}_{D^C} ; ensuite, et pour chaque plage C_k , voisine d'une plage affectée à D , non classée elle même dans D , on procède à la phase d'agglomération.

Dans le cas où l'on affecte la plage C_k à la partie D , on réitère la phase

d'agglomération pour les plages voisines de C_k qui ne sont pas classées encore dans D.

Dans le cas contraire, on établit une nouvelle estimation du paramètre \bar{n}_{D^C} , cette fois à partir des $(n_1 + 8)$ cellules de l'ensemble D^C .

Enfin, une fois cette procédure terminée, toutes les plages qui n'ont pas été agglomérées au noyau de la diffraction sont classées dans la partie D^C , et l'on effectue alors le calcul de l'intensité intégrée de la tache de BRAGG.

Le programme correspondant au traitement d'une diffraction figure en annexe 3 ; y figurent également certains résultats marquants, ainsi qu'un exemple de traitement.

11.3. CALCUL DE L'INTENSITE INTEGREE D'UNE DIFFRACTION POUR UNE VALEUR ELEVEE DU BRUIT DE FOND :

Nous supposons dans cette partie que le multidétecteur est stable dans le temps, c'est à dire que le processus de comptage est stationnaire.

La démarche que nous allons suivre est de définir une transformation similaire à celle exposée au chapitre I de manière à nous ramener à un problème sur échantillons gaussiens.

Rappelons la manière dont est effectuée une telle transformation.

11.3.1. Transformation de l'échantillon :

Soit $(n_i(t, t+n h_1))_{i=1 \dots 1024}$ un enregistrement effectué pendant une durée d'acquisition $(t, t+n h_1)$ en présence d'une tache de BRAGG. Chaque comptage $n_i(t, t+n h_1)$ se décompose en

$$n_i(t, t+n h_1) = \sum_{k=1, n} n_i(t+(k-1) h_1, t+k h_1)$$

Sous l'hypothèse de stabilité dans le temps du multidétecteur, les observations $n_i(t+(k-1)h_1, t + kh_1)_{k=1\dots n}$ constituent un n échantillon d'une loi de Poisson de paramètre $\lambda(t, t+h_1) + s_i(t, t+h_1)$.

Nous savons dès lors que les valeurs $n_i(t, t+nh_1)^{1/2}$ sont les réalisations de lois approximativement normales, de moyenne $[\lambda(t, t+h_1) + s_i(t, t+h_1)]^{1/2}$, et de variance $1/4$, avec $s_i(t, t+h_1) = 0$ si la cellule i se trouve appartenir à l'ensemble D^C , ensemble sur lequel la contribution de la diffraction aux comptages $n_i(t, t+nh_1)$ est nulle.

11.3.2 Extraction du noyau d'une diffraction :

Nous allons mettre en place pour effectuer une première classification des cellules une procédure de type séquentiel, basée sur un test d'égalité en moyenne d'échantillons gaussiens.

En voici le fondement.

Supposons qu'à un niveau donné de notre étude, nous ayons affecté k cellules à la partie D^C du détecteur, partie ne percevant que le seul bruit de fond.

Les valeurs $(n_i(t, t+nh_1)^{1/2})_{i=1\dots k}$ calculées sur ces k cellules constituent un k échantillon d'une loi approximativement normale, de moyenne $(n\lambda(t, t+h_1))^{1/2}$, et de variance $1/4$.

Désignons par \bar{n}_k la moyenne de ces valeurs :

$$\bar{n}_k = 1/k \sum_{i=1, k} n_i(t, t+nh_1)^{1/2}$$

\bar{n}_k est la réalisation d'une variable \bar{N}_k dont la loi est approximativement normale, de moyenne $(n\lambda(t, t+h_1))^{1/2}$, et de variance $1/4K$.

Considérons maintenant un ensemble de p valeurs $(n_i(t, t+nh_1))^{1/2}_{i=k+1 \dots k+p}$, calculées sur une partie C_p du multidétecteur comprenant p cellules, et dont nous voulons tester l'inclusion dans D^C

Posons $\bar{n}_{k,k+p} = 1/p \sum_{i=k+1, k+p} (n_i(t, t+nh_1))^{1/2}$.

La loi de la variable $\bar{N}_{k,k+p}$, dont $\bar{n}_{k,k+p}$ est une réalisation, est :

- sous l'hypothèse H_0 où $C_p \subset D^C$ une loi approximativement normale, de moyenne $(n\lambda(t, t+h_1))^{1/2}$, et de variance $1/4p$.

- sous l'hypothèse H_1 ($C_p \cap D \neq \emptyset$) une loi approximativement normale, de variance $1/4p$ et de moyenne m_0 donnée par :

$$m_0 = 1/p \sum_{i=k+1, k+p} (n(\lambda(t, t+h_1) + s_i(t, t+h_1)))^{1/2},$$

les $s_i(t, t+h_1)$ étant non tous nuls.

Proposition

La loi de la variable $\bar{N}_{k,k+p} - \bar{N}_k$ est une loi approximativement normale, de variance $1/4k + 1/4p$ et de moyenne :

- nulle sous l'hypothèse H_0

- $m_1 > 0$ sous l'hypothèse H_1 , avec $m_1 = m_0 - (n\lambda(t, t+h_1))^{1/2}$.

En corollaire, nous pouvons énoncer le résultat suivant :

Corollaire :

Le test de l'hypothèse $C_p \subset D^C$ contre $C_p \cap D \neq \emptyset$, défini par la région critique

$$R_C = \{\omega, \bar{N}_{k,k+p}(\omega) - \bar{N}_k(\omega) > c_{1-\alpha}\}$$

où $c_q (1/4k + 1/4p)^{-1/2}$ est le quantile d'ordre q d'une loi normale centrée

réduite est approximativement de niveau α .

En conséquence, nous affectons l'ensemble des cellules de C_p

- à la partie D^C du détecteur si $\bar{n}_{k,k+p} - \bar{n}_k < c_{1-\alpha_1}$
- à la partie D si $\bar{n}_{k,k+p} - \bar{n}_k > c_{1-\alpha_2}$

où l'on choisit une valeur arbitrairement très faible pour α_2 .

Traitement effectif :

D'un point de vue beaucoup plus pragmatique, voici la manière dont nous procédons :

Nous effectuons une partition du multidétecteur en 128 plages disjointes C_1, \dots, C_N comprenant chacune 8 cellules.

En tout premier lieu, il nous faut effectuer une première estimation du paramètre \bar{n}_k ; dans l'absolu, nous devrions pour ce faire extraire de l'appareil un nombre donné de plages qui aient une probabilité élevée de ne pas être touchées par la diffraction, ce dont on pourrait s'assurer en retenant les plages pour lesquelles l'adéquation à une loi normale de moyenne $(n\lambda(t, t+h_1))^{1/2}$ et de variance $1/4$ est la meilleure. Afin de ne pas compliquer la procédure, nous supposons que la tache de BRAGG est centrée, et estimons la quantité \bar{n}_k sur 4 plages situées à la périphérie du détecteur.

Nous comparons ensuite successivement et pour chaque plage C_p les moyennes de deux échantillons gaussiens, l'un de taille n_1 correspondant aux cellules déjà affectées à la partie D^C lorsque l'on teste la plage C_p , l'autre de taille 8 constitué des enregistrements effectués sur C_p

On classe ensuite l'ensemble des cellules de la plage C_p .

- dans la partie D^C si l'on accepte H_0 au niveau $\alpha_1 = 0.025$
- dans la partie D si l'on n'accepte pas H_0 au niveau $\alpha_2 = 0.001$.

S'il s'explique par notre désir de ne commettre l'erreur consistant à classer une plage de D^C dans la partie occupée par la diffraction qu'avec une très faible probabilité, un tel choix du paramètre α_2 appelle toutefois deux remarques :

i) Nous prenons le risque de classer à tort des plages dans la partie D^C de l'appareil ; la phase d'agglomération est donc nécessaire pour extraire le signal dans son entité.

ii) Il peut également se faire qu'à ce niveau, on ne soit pas à même de détecter une plage susceptible d'appartenir au noyau de la diffraction ; nous résolvons ce problème en retenant alors comme noyau la plage sur laquelle les comptages sont les plus élevés.

Avant d'explicitier la manière dont nous avons conçu la phase d'agglomération, étudions plus précisément la procédure que nous venons de mettre en place.

Désignons par C_p une plage sur laquelle on enregistre effectivement la diffraction de BRAGG, et efforçons nous de répondre à cette question :

" De quel ordre de grandeur doit être le rapport signal/bruit sur cette plage pour qu'au niveau α_2 , le test nous permette de classer C_p dans la partie D du multidétecteur avec une probabilité donnée, 0.75 par exemple."

Désignons par $(S/B)_{C_p}$ la valeur du rapport signal/bruit sur la seule plage C_p . En d'autres termes, nous désirons déterminer la quantité $(S/B)_{C_p}$

de manière à ce que :

$$P [\bar{N}_{k,k+p} - \bar{N}_k > C_{1-\alpha_2} / C_p \leq D] = 0.75.$$

Développons le calcul qui nous permet de solutionner ce problème :

$$\begin{aligned} (P [\bar{N}_{k,k+p} - \bar{N}_k > C_{1-\alpha_2} / C_p \leq D] = 0.75) \\ = (P [(\bar{N}_{k,k+p} - \bar{N}_k - m_1) / (1/4k + 1/4p)^{1/2} \\ > (C_{1-\alpha_2} - m_1) / (1/4k + 1/4p)^{1/2}] = 0.75) \end{aligned}$$

En désignant par Φ la fonction de répartition de la loi normale centrée réduite, on a :

$$\Phi ((C_{1-\alpha_2} - m_1) / (1/4k + 1/4p)^{1/2}) = 0.25 \quad (1)$$

De plus,

$$\begin{aligned} m_1 &= 1/p \sum_{i=k+1, k+p} (n(\lambda(t, t+h_1) + s_i(t, t+h_1)))^{1/2} - (n\lambda(t, t+h_1))^{1/2} \\ &= 1/p \sum_{i=k+1, k+p} (n\lambda(t, t+h_1))^{1/2} [(1 + (ns_i(t, t+h_1)) / (n\lambda(t, t+h_1)))^{1/2} - 1] \end{aligned}$$

égalité qui, à l'ordre $O((n\lambda(t, t+h_1))^{-3/2})$, s'écrit encore comme

$$m_1 = 1/p \sum_{i=k+1, k+p} (n\lambda(t, t+h_1))^{1/2} (ns_i(t, t+h_1)) / (2n\lambda(t, t+h_1))$$

De sorte que l'on a, à cet ordre :

$$m_1 = (n\lambda(t, t+h_1))^{1/2} (S/B)_{C_p} / 2$$

En substituant m_1 par cette valeur dans l'égalité (1) il vient

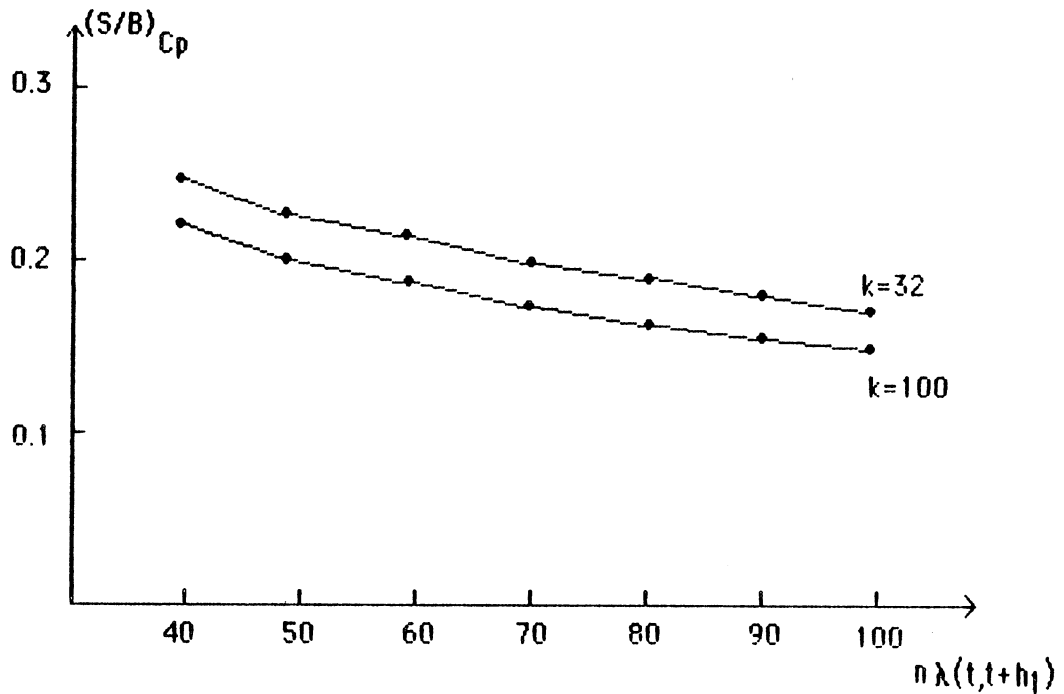
$$\Phi ((2C_{1-\alpha_2} - (n\lambda(t, t+h_1))^{1/2} (S/B)_{C_p}) / 2(1/4k + 1/4p)^{1/2}) = 0.25$$

soit encore, tout calcul effectué :

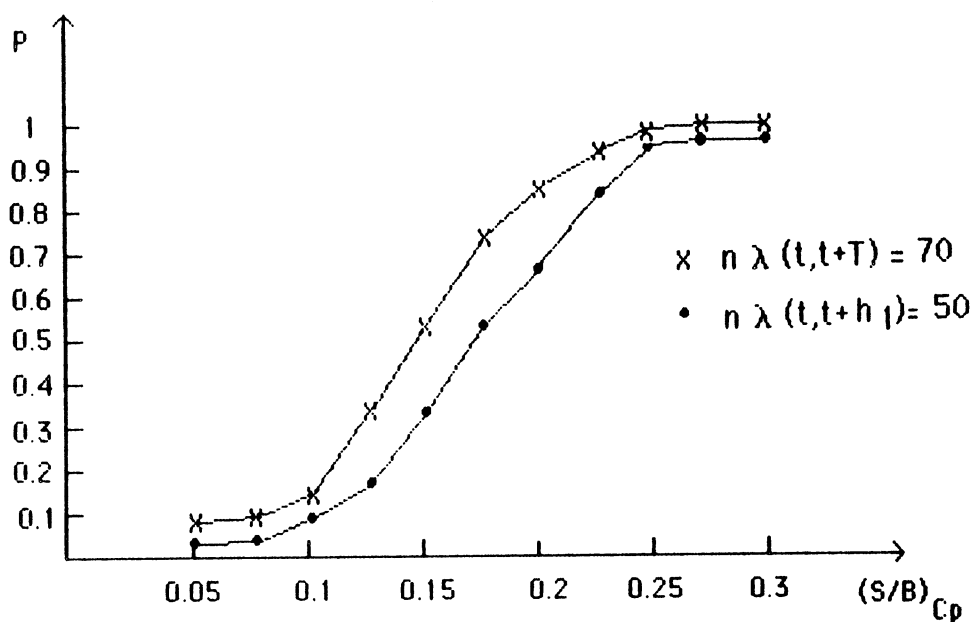
$$(S/B)_{C_p} = (3.7645 \times 2 \times (1/4k + 1/4p)^{1/2}) / (n\lambda(t, t+h_1))^{1/2}$$

La figure ci-dessous illustre l'évolution du rapport signal-bruit nécessaire à l'obtention d'une puissance $p = 0.75$ en fonction du bruit

moyen d'une part ($n \lambda(t, t+h_1)$) et du nombre k de cellules affectées à la "partie bruit" d'autre part ($32 \leq k \leq 1016$).



Il est également intéressant de constater comment évolue la puissance du test en fonction du rapport signal-bruit pour un bruit moyen et un nombre k de cellules donné ; c'est ce que nous avons représenté dans une deuxième figure.



Voyons maintenant comment nous envisageons la phase d'agglomération.

11.3.3. Agglomération des cellules voisines du noyau d'une diffraction :

Une procédure similaire à celle que nous avons développée lorsque le bruit de fond est de faible intensité est là encore envisageable. Celle-ci étant compliquée et exigeant un temps de calcul élevé, nous avons préféré mettre en place une procédure d'agglomération plus simple, mais nécessitant l'hypothèse supplémentaire que les comptages enregistrés sur la partie D du détecteur le sont de manière homogène. Nous justifions cette hypothèse par le fait que, puisque le bruit de fond moyen est élevé, on a

$$s_j(t, t+h_1) \ll \lambda(t, t+h_1)$$

Soient $n_1(t, t+h_1), \dots, n_k(t, t+h_1)$ les comptages effectués sur k cellules que nous savons appartenir à la partie D^C du détecteur. Ces observations constituent un k échantillon d'une loi de Poisson de paramètre $n\lambda(t, t+h_1)$, loi que l'on peut approximer pour une valeur élevée de $n\lambda(t, t+h_1)$ par une loi normale $N(m_0, \sigma_0^2)$.

Désignons par $n_{k+1}(t, t+nh_1), \dots, n_{k+p}(t, t+nh_1)$ les enregistrements relatifs aux cellules affectées à la partie D du multidétecteur. Sous l'hypothèse, établie à priori, que les comptages sont homogènes, ces valeurs constituent un p échantillon d'une variable dont la loi est, en première approximation, normale, de moyenne m_1 et de variance σ_1^2 .

Soit $n'_i(t, t+nh_1)$ un comptage effectué sur une cellule frontière, c'est à dire une cellule non affectée à D, mais voisine d'un capteur classé dans D.

La méthode de classement que nous avons adoptée est décrite par J.M. ROMEDER dans [17] ; en voici le principe :

On affecte la cellule i au groupe pour lequel la probabilité d'observer $n'_i(t, t+nh_1)$ est la plus forte, donc au groupe pour lequel la probabilité a posteriori est la plus grande.

La région d'affectation à la partie D^C ne percevant pas la diffraction est donnée par

$$R_B = (x, 1/\sigma_0 \sqrt{2\pi} \exp(-(x-m_0)^2 / 2\sigma_0^2) > 1/\sigma_1 \sqrt{2\pi} \exp(-(x-m_1)^2 / 2\sigma_1^2))$$

Le programme relatif au traitement mis en place dans cette partie, ainsi que certains résultats et exemples figurent en annexe 4.

ANNEXE 1

**Représentation d'un enregistrement
obtenu sur l'instrument D 19A [8]**

RUN DATECAR

FIRST MINOR

10939

LAST MINOR=10939

D19 10939 20-DEC-80 15:21:20 RATE DE CAM

UTILISATEUR:THOMAS LOCAL CONTACT:THOMAS

COMMENT:PHTHALOCYANINE TEMP AMBIENTE , DEC 80. B AXIS OFFSET.

MONITEUR= 179. TEMPS= 4000. ANGLE= 31.45 PHI=-128.72 CHI= 31.45 OMEGA= 31.38 GAMMA= 62.76 FSI= 0.00

MATRICE UR: 0.072203 0.022822 0.007874 0.031754 0.022872 0.058422 0.011356 -0.209069 0.007251

LAMBDA= 1.578864 TEMPERATURE= 0.000 CHAMP MAGNETIQUE= 0.000 H= -5.00 K= 0.00 L= -6.00

ONE SCAN ANGLE DEFART= 30.55 PAS= 0.05 DOMAINE= 1.65 PRESET= 4000.

(T = 40 sec) X ↑ Y

Cathodes

	K 0	K 1	K 2	K 3	K 4	K 5	K 6	K 7	K 8	K 9	K 10	K 11	K 12	K 13	K 14	K 15
A 0	0.	0.	2.	2.	1.	4.	1.	0.	1.	0.	1.	0.	1.	0.	0.	0.
A 1	2.	1.	1.	1.	1.	1.	1.	1.	1.	2.	1.	0.	2.	1.	1.	2.
A 2	2.	1.	2.	0.	2.	0.	2.	1.	1.	0.	0.	0.	2.	1.	1.	2.
A 3	4.	0.	0.	0.	1.	0.	2.	1.	0.	0.	0.	1.	0.	1.	3.	2.
A 4	1.	1.	0.	0.	1.	1.	0.	0.	0.	0.	0.	1.	0.	2.	0.	0.
A 5	0.	2.	0.	0.	0.	1.	1.	2.	1.	0.	0.	3.	0.	1.	0.	0.
A 6	0.	0.	1.	0.	0.	1.	0.	0.	0.	0.	0.	3.	2.	0.	0.	0.
A 7	1.	1.	0.	0.	0.	0.	0.	0.	1.	2.	0.	0.	1.	1.	0.	0.
A 8	0.	0.	2.	0.	4.	0.	2.	2.	0.	0.	1.	0.	0.	0.	0.	0.
A 9	0.	0.	1.	1.	2.	0.	0.	0.	0.	0.	0.	0.	3.	1.	2.	0.
A 10	0.	0.	0.	0.	1.	0.	0.	0.	0.	1.	0.	0.	2.	0.	2.	1.
A 11	1.	1.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	1.	0.
A 12	1.	2.	1.	0.	0.	1.	0.	1.	0.	0.	2.	0.	0.	0.	0.	1.
A 13	0.	1.	0.	2.	1.	0.	0.	0.	1.	0.	1.	0.	1.	0.	0.	0.
A 14	1.	0.	2.	1.	0.	0.	0.	1.	0.	0.	0.	0.	0.	0.	0.	0.
A 15	1.	0.	1.	1.	0.	2.	0.	0.	0.	0.	0.	0.	0.	1.	0.	0.
A 16	0.	0.	1.	0.	2.	0.	0.	0.	1.	0.	2.	1.	1.	0.	2.	3.
A 17	1.	0.	0.	1.	0.	0.	1.	0.	0.	3.	1.	0.	0.	0.	1.	0.
A 18	0.	2.	1.	0.	1.	3.	0.	0.	0.	0.	1.	1.	0.	0.	1.	0.
A 19	3.	1.	0.	0.	1.	0.	1.	3.	2.	1.	0.	0.	2.	1.	2.	1.
A 20	0.	2.	1.	0.	1.	0.	0.	0.	2.	0.	0.	0.	2.	2.	1.	1.
A 21	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	4.	2.
A 22	3.	0.	1.	0.	0.	0.	0.	0.	0.	0.	2.	0.	0.	0.	0.	0.
A 23	0.	2.	0.	2.	1.	3.	0.	3.	3.	0.	1.	1.	0.	1.	1.	1.
A 24	3.	4.	0.	0.	2.	0.	1.	1.	1.	1.	1.	0.	0.	0.	0.	1.
A 25	0.	2.	4.	1.	2.	2.	1.	0.	1.	3.	0.	1.	0.	0.	0.	1.
A 26	0.	3.	2.	0.	1.	0.	0.	1.	1.	0.	1.	0.	0.	0.	1.	0.
A 27	0.	1.	2.	0.	0.	3.	1.	4.	4.	2.	1.	1.	3.	0.	0.	0.
A 28	1.	1.	2.	3.	7.	5.	11.	10.	11.	9.	3.	4.	1.	2.	0.	1.
A 29	0.	0.	0.	1.	33.	100.	149.	115.	117.	90.	76.	15.	4.	0.	2.	0.
A 30	1.	0.	2.	7.	67.	236.	380.	350.	320.	291.	189.	37.	5.	2.	1.	0.
A 31	4.	1.	1.	8.	45.	219.	532.	583.	334.	297.	229.	40.	8.	2.	2.	0.
A 32	4.	4.	4.	9.	100.	376.	556.	641.	326.	330.	358.	54.	20.	5.	0.	1.
A 33	2.	2.	1.	7.	106.	230.	206.	284.	218.	163.	144.	33.	9.	1.	1.	1.
A 34	2.	0.	1.	0.	18.	27.	29.	49.	35.	25.	10.	3.	0.	1.	0.	3.
A 35	2.	0.	2.	0.	0.	5.	4.	2.	1.	0.	2.	7.	1.	0.	2.	0.
A 36	0.	3.	1.	0.	1.	5.	2.	1.	3.	0.	0.	2.	1.	0.	0.	1.
A 37	0.	0.	1.	3.	1.	0.	2.	0.	1.	1.	1.	1.	0.	0.	0.	0.
A 38	1.	1.	0.	1.	2.	0.	0.	0.	1.	2.	3.	1.	0.	0.	0.	0.
A 39	0.	1.	1.	3.	1.	0.	1.	2.	1.	0.	0.	0.	0.	1.	1.	1.
A 40	1.	1.	2.	0.	3.	1.	0.	1.	0.	0.	1.	4.	0.	0.	0.	0.
A 41	1.	1.	1.	1.	1.	1.	1.	1.	0.	0.	0.	0.	2.	0.	1.	0.
A 42	2.	0.	1.	2.	1.	1.	1.	0.	0.	0.	0.	1.	1.	0.	0.	1.
A 43	3.	0.	0.	0.	0.	0.	0.	1.	2.	2.	0.	0.	0.	0.	0.	1.
A 44	1.	2.	0.	3.	1.	0.	1.	0.	0.	0.	0.	1.	0.	2.	1.	2.
A 45	0.	1.	2.	1.	1.	0.	0.	0.	3.	0.	1.	1.	0.	1.	1.	0.
A 46	0.	1.	0.	1.	0.	0.	1.	0.	0.	1.	0.	1.	0.	0.	0.	0.
A 47	0.	1.	1.	0.	1.	2.	1.	0.	0.	0.	0.	2.	1.	0.	0.	0.
A 48	0.	1.	1.	0.	0.	1.	0.	1.	2.	0.	0.	0.	1.	0.	0.	0.
A 49	0.	1.	1.	2.	0.	0.	0.	0.	0.	2.	1.	0.	0.	0.	0.	0.
A 50	1.	2.	1.	0.	1.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
A 51	1.	1.	0.	2.	1.	0.	1.	0.	0.	0.	1.	0.	0.	0.	1.	0.
A 52	0.	2.	1.	0.	1.	0.	0.	0.	1.	0.	0.	1.	0.	0.	1.	0.
A 53	1.	1.	0.	0.	1.	0.	0.	0.	0.	0.	0.	1.	0.	1.	0.	0.
A 54	0.	0.	1.	1.	4.	0.	0.	1.	1.	0.	0.	0.	1.	0.	0.	2.
A 55	1.	1.	0.	0.	0.	1.	0.	1.	2.	1.	1.	0.	0.	0.	0.	0.
A 56	1.	0.	1.	0.	0.	0.	0.	0.	0.	0.	3.	0.	0.	0.	0.	1.
A 57	0.	1.	2.	2.	0.	0.	2.	1.	0.	0.	0.	0.	0.	0.	1.	0.
A 58	1.	1.	1.	0.	0.	2.	2.	0.	0.	0.	0.	1.	0.	0.	0.	2.
A 59	1.	2.	0.	1.	0.	2.	0.	0.	0.	0.	1.	0.	1.	1.	0.	1.
A 60	0.	0.	0.	0.	0.	0.	2.	0.	1.	2.	0.	2.	0.	0.	0.	0.
A 61	0.	0.	1.	0.	1.	0.	0.	0.	0.	0.	0.	0.	0.	0.	2.	0.
A 62	0.	2.	0.	1.	0.	0.	2.	0.	2.	0.	0.	0.	0.	0.	0.	1.
A 63	3.	1.	4.	1.	0.	0.	0.	1.	0.	0.	0.	0.	0.	1.	0.	0.

DATECAR STOP

ANNEXE 2

Simulation d'un enregistrement

Nous allons expliquer ici comment nous simulons un enregistrement, et présenter quelques exemples.

Nous assimilons le multidétecteur de particules à un borélien D de \mathbb{R}^2 , $D = [-h, h[\times]-l, l[$, union disjointe de 1024 pavés C_{ij} correspondant aux dimensions des 1024 cellules de l'appareil :

$$D = \bigcup_{i=1}^{16} \bigcup_{j=1}^{64} [-h + (i-1)c_1, -h + i c_1 [\times]-l + (j-1)c_2, -l + j c_2 [$$

Désignons par U un vecteur gaussien, centré et de matrice de variance-covariance Λ donnée par :

$$\Lambda = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

Les comptages effectués par le multidétecteur se présentent sous la forme d'une matrice de taille 16×64 ; pour simuler un enregistrement, nous générons donc une matrice N , de cette taille, dont chaque élément n_{ij} est la somme des réalisations n'_{ij} et n''_{ij} de deux lois de Poisson, où :

- n'_{ij} est la réalisation d'une loi de Poisson de paramètre k_1 , et représente la contribution du bruit de fond au comptage effectué.
- n''_{ij} est la réalisation d'une loi de Poisson de moyenne $k_2 P[U \in C_{ij}]$, et figure la valeur du signal enregistré sur la cellule correspondante.

En conséquence, nous générons successivement :

- un 1024 échantillon d'une loi de Poisson de paramètre k_1 ; ce paramètre correspond au bruit de fond moyen théorique, et la moyenne de l'échantillon au bruit de fond moyen effectivement généré.
- 1024 valeurs de lois de Poisson de moyenne $k_2 P[U \in C_{ij}]$, dont la somme représente l'intensité intégrée de la diffraction simulée.

Remarques :

i) Nous disposons d'un générateur de nombres aléatoires selon la loi uniforme sur $[0,1[$; nous générons une loi de Poisson de moyenne λ de la manière suivante :

à chaque tirage d'une valeur x de $[0,1[$, on associe l'entier n , tel que :

$$n = \max\{i \in \mathbb{N}, \sum_{j=0,i} \exp(-\lambda) (\lambda^j / j!) \leq x\}$$

ii) nous approximos la quantité $P[U \in C_{ij}]$ par :

$$P[U \in C_{ij}] \simeq$$

$$(1/2\pi \sigma_1 \sigma_2) c_1 c_2 \exp [-((h+ic_1-c_1/2)^2 / 2\sigma_1^2) - ((-1+jc_2-c_2/2)^2 / 2\sigma_2^2)]$$

Quelques mots concernant le programme.

Le programme est conçu selon un mode conversationnel, l'utilisateur ayant à répondre aux questions posées concernant la valeur des paramètres

k_1, k_2, σ_1 et σ_2 .

Une fois le traitement effectué, il y a création d'un fichier contenant :

- la valeur des paramètres $k_1, k_2, \sigma_1, \sigma_2$
- la valeur du bruit de fond moyen généré, ainsi que celle de l'intensité intégrée de la diffraction simulée
- la simulation d'un enregistrement.


```
program simulation;
```

```
{SIMULATION d'un enregistrement sur le multidetecteur}
```

```
const
```

```
pas1=4.92;pas2=2.54; {pas1 et pas2 sont les dimensions d'une cellule }
```

```
pi=3.14159;
```

```
type
```

```
matrix=array[1..16,1..64] of real;
```

```
var
```

```
f:text;
```

```
ma,mb:matrix;
```

```
w,p,q,mul:integer;
```

```
tot,bruit,landa,sig1,sig2:real;
```

```
external function hasard:real;
```

```
function poisson(x:real;land:real):integer;
```

```
{on genere une loi de poisson de parametre land a partir de la loi uniforme sur [0,1[ }
```

```
var
```

```
a,b:real;
```

```
m:integer;
```

```
begin
```

```
a:=exp(-land);
```

```
b:=a;
```

```
m:=0;
```

```
while a<x do
```

```
begin
```

```
m:=m+1;
```

```
b:=b*land/m;
```

```
a:=a+b {fonction de repartition F(m)}
```

```
end;
```

```
poisson:=m {m=max(u, F(u)<x}
```

```
end;
```

```
procedure lecture; {acquisition des donnees}
```

```
begin
```

```
writeln('VALEURS DU SIGNAL GAUSSIEN CENTRE ');
```

```
writeln('entrer l ecart type de la premiere marginale ');
```

```
readln(sig1);
```

```
writeln('entrer l ecart type de la seconde marginale ');
```

```
readln(sig2);
```

```
writeln('entrer la constante multiplicative du gaussien');
```

```
readln(mul);
```

```
writeln;
```

```
writeln('entrer la valeur moyenne du bruit ');
```

```
readln(landa)
```

```
end;
```

```
procedure ecriture; {impression de l'enregistrement simule}
```

```
var
```

```
i,j:integer;
```

```
begin
```

```
writeln(f,' CARACTERISTIQUE d un ENREGISTREMENT ');
```

```
writeln(f);
```

```
writeln(f,' ecart type de la premiere marginale ',sig1:7:2);
```

```
writeln(f,' ecart type de la seconde marginale ',sig2:7:2);
```

```
writeln(f,' valeur de la constante multiplicative ',mul:8);
```

```
writeln(f, ' valeur du bruit de fond moyen demande ',landa:7:3);
writeln(f);
writeln(f,'Valeurs simulees ');
writeln(f,' bruit de fond moyen effectivement genere ',bruit/1024:7:2);
writeln(f,' intensite integree de la diffraction simulee ',tot-bruit:7:2);
writeln(f);
writeln(f,' Simulation d un enregistrement ');
writeln(f);
for i:=1 to 64 do
begin
for j:=1 to 16 do write(f,round(ma[j,i]):3,' ');
writeln(f)
end
end;

procedure gauss(var mc:matrix);
{on calcule dans cette procedure la quantite k2*P[u appartient a Ci]}
var
i,j:integer;
x1,x2,r,s:real;
begin
for i:=1 to 16 do
begin
for j:=1 to 64 do mc[i,j]:=0
end;
x1:=15*pas1/2;
x2:=-63*pas2/2;
for i:=1 to 8 do
begin
r:=exp(-sqr(x1-(i-1)*pas1)/(2*sqr(sig1)))/(sqrt(2*pi)*sig1);
for j:=1 to 32 do
begin
s:=exp(-sqr(x2+(j-1)*pas2)/(2*sqr(sig2)))/(sqrt(2*pi)*sig2);
mc[i,j]:=abs(s*r*mul)
end
end;
for i:=1 to 8 do
begin
for j:=1 to 32 do
begin
mc[i+8,j]:=mc[8-(i-1),j];
mc[i,j+32]:=mc[i,32-(j-1)];
mc[i+8,j+32]:=mc[8-(i-1),32-(j-1)]
end
end
end;

begin{program}
{on genere dans un premier temps une loi de poisson de parametre k1, corresponda
nt au seul bruit de fond '}
assign(f,'b:result.sim');
rewrite(f);
bruit:=0;
tot:=0;
lecture;
for p:=1 to 16 do
begin
for q:=1 to 64 do
begin
ma[p,q]:=poisson(hasard,landa);
```

```
bruit:=bruit+ma[p,q]
end
end;
writeln('bruit effectivement genere ',bruit/1024:6:2);
{ajout de la partie signal}
gauss(mb);
for p:=1 to 16 do
begin
for q:=1 to 64 do
begin
ma[p,q]:=ma[p,q]+poisson(hazard,mb[p,q]);
tot:=tot+ma[p,q]
end
end;
écriture;
close(f,w)
end.
```

CARACTERISTIQUE d un ENREGISTREMENT

ecart type de la premiere marginale 7.00
 ecart type de la seconde marginale 8.00
 valeur de la constante multiplicative 10000
 valeur du bruit de fond moyen demande 3.000

Valeurs simulees

bruit de fond moyen effectivement genere 2.95
 intensite integree de la diffraction simulee 832.00

Simulation d un enregistrement

1	2	0	4	4	1	1	0	3	3	1	2	1	0	4	5
5	3	3	2	1	1	5	2	5	3	5	1	6	2	0	3
4	4	3	5	4	2	3	5	1	1	3	2	2	7	2	4
5	1	6	5	5	4	2	6	2	4	0	0	2	2	2	2
4	0	0	4	6	5	3	2	3	3	3	2	3	4	3	1
6	1	4	1	4	4	7	3	0	6	2	2	1	1	2	0
4	3	4	3	3	3	4	1	4	1	2	3	2	5	1	3
3	0	2	5	4	3	2	3	4	3	4	5	1	3	2	2
2	6	2	4	6	4	3	2	2	2	4	2	2	2	1	2
6	2	3	8	4	3	4	3	2	3	3	5	4	3	5	5
2	4	5	2	1	5	5	2	3	4	1	1	1	3	4	4
0	8	3	1	5	3	1	4	3	2	3	6	5	5	1	6
3	2	6	3	0	1	1	3	0	5	5	1	3	2	2	1
5	3	3	5	2	4	2	3	2	3	5	5	5	1	4	3
3	0	2	6	0	5	1	4	5	2	1	3	2	2	5	1
2	2	1	0	3	2	3	4	3	2	2	5	5	5	3	8
5	3	1	1	0	2	2	3	3	4	2	3	3	1	1	4
1	3	3	3	3	4	1	5	3	2	1	2	1	5	2	6
4	2	3	2	2	4	1	4	3	2	5	5	4	7	2	2
2	3	4	3	4	2	3	0	4	2	2	2	1	1	0	1
1	4	4	3	3	0	3	4	2	3	6	0	2	0	4	3
0	4	2	2	4	2	1	4	4	3	2	3	4	2	0	6
3	0	2	4	6	1	7	4	0	3	1	2	2	2	1	2
2	2	0	5	1	1	5	4	0	4	3	2	6	4	8	3
5	2	2	2	0	5	5	2	6	7	4	4	1	3	2	3
6	4	0	1	1	3	6	4	8	5	3	1	4	4	5	3
2	3	3	2	4	3	5	9	9	9	5	2	1	2	4	3
3	2	5	4	1	10	4	12	11	7	7	0	3	5	6	3
1	6	4	3	7	11	15	18	18	10	8	0	2	3	5	3
0	4	4	5	2	12	12	23	23	24	13	5	6	4	1	5
8	2	2	4	8	8	23	29	28	16	11	1	1	4	2	5
5	4	2	5	5	12	23	27	33	20	11	6	2	2	2	2
4	2	4	1	3	8	15	25	35	23	7	5	4	4	2	4
7	2	5	7	4	10	19	26	21	7	13	4	2	7	2	5
3	6	2	2	5	6	20	21	25	10	13	6	2	4	3	2
2	3	2	3	7	2	10	21	14	16	10	3	3	4	3	2
2	4	2	0	6	6	6	14	10	6	1	4	3	5	1	4
3	1	7	7	4	2	9	8	6	6	1	1	3	2	3	1
4	2	2	3	2	3	3	3	4	3	1	1	4	6	4	2
3	6	1	3	2	3	2	4	2	5	5	3	3	2	0	4
3	3	6	1	2	5	7	1	6	0	3	3	1	2	3	2
1	4	7	6	4	6	1	4	9	5	5	2	3	2	3	1
3	5	4	1	2	5	2	3	3	6	4	5	3	3	2	4
3	5	6	6	4	2	5	3	3	1	7	0	3	3	2	3
4	1	5	0	1	0	3	6	1	3	4	4	4	6	8	2
7	4	4	2	2	2	3	4	1	1	2	1	4	4	3	1
6	2	6	4	3	1	3	2	2	5	2	5	6	3	2	1
1	4	2	6	2	3	1	3	1	2	3	2	5	4	2	2
6	5	3	3	4	5	1	3	4	1	2	0	3	1	1	5
7	3	3	4	3	3	2	1	3	3	4	2	2	1	1	2
2	4	0	1	3	2	1	1	3	2	4	1	2	3	6	1
3	3	5	5	4	0	3	2	3	3	4	5	1	4	3	6
2	4	1	5	3	2	2	8	6	2	3	1	2	2	8	4
1	4	2	4	5	2	6	3	1	4	3	1	4	1	2	3
1	2	3	5	2	3	2	0	5	3	5	4	5	2	6	2
0	2	1	5	3	2	4	4	3	5	2	1	3	3	1	4
3	3	3	2	4	7	4	3	2	4	1	2	6	3	1	3
2	2	2	3	2	3	4	1	3	1	5	5	3	0	2	3
3	1	4	2	4	5	1	3	1	5	3	4	4	3	3	1
3	1	4	2	4	5	1	1	5	3	3	4	1	4	1	4
4	2	2	1	5	3	3	3	5	5	2	2	0	2	8	5
1	2	2	4	0	5	2	4	7	3	8	1	2	5	2	3
5	2	3	6	1	1	4	3	6	2	4	1	2	3	3	6
6	4	4	3	2	4	0	3	3	2	1	1	1	3	2	0

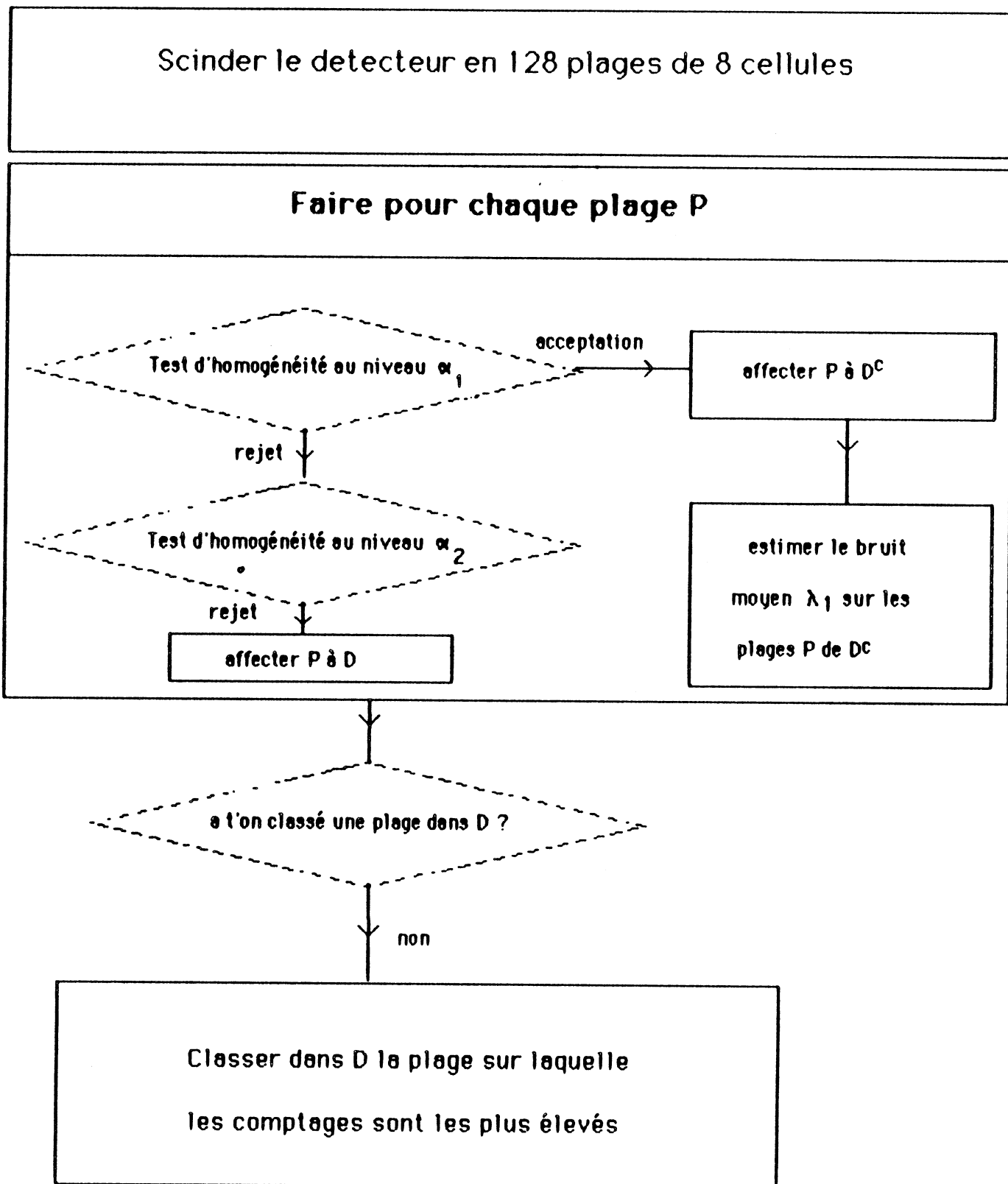


ANNEXE 3

Le programme que nous avons conçu pour estimer l'intensité intégrée d'une diffraction comporte 3 blocs principaux qui sont, outre le calcul proprement dit de l'intensité estimée, l'extraction du noyau du signal d'une part, l'agglomération des plages frontières d'autre part. Pour la clarté de cette étude, nous présentons ici deux organigrammes correspondant à ces deux dernières phases.

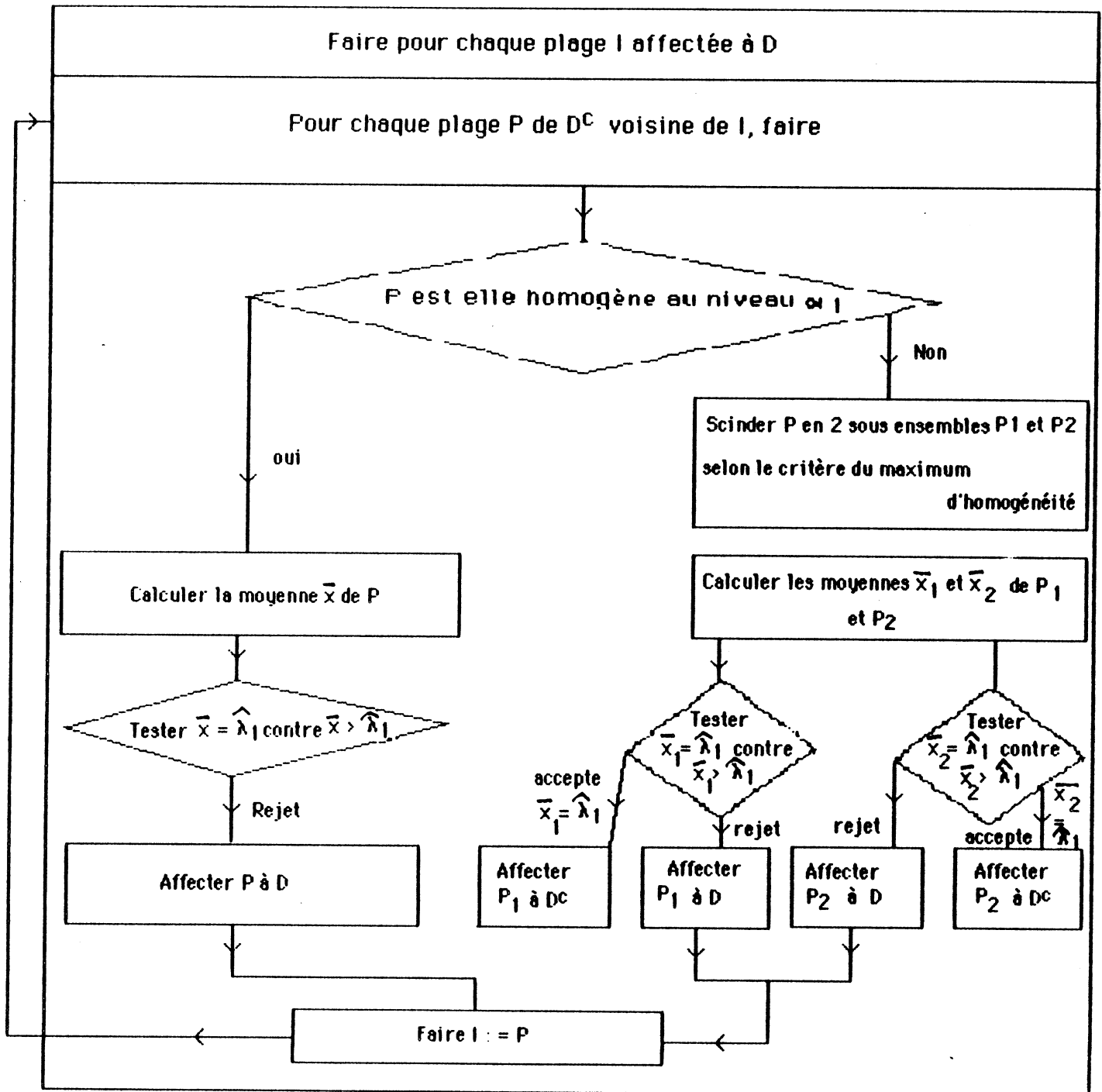
La lecture de ces organigrammes rendra plus claire le programme, assez long, qui figure en deuxième partie de cette annexe.

Enfin, nous consacrons une troisième partie à l'énoncé de quelques résultats et aux commentaires qu'ils nécessitent.



Organigramme de la phase " détection du noyau de la diffraction ; Première estimation du bruit"

Detection du noyau - Première estimation du bruit



Organigramme de la phase "Agglomération des plages frontières"


```
program traitement;
```

```
{Calcul de l'intensite integree d'une diffraction _ cas general}
```

```
const s1=8.38;s2=2.17;  
type  
mat=array[1..8] of real;  
matrix=array[1..16,1..64,1..2] of real;  
var  
f,g:text;  
bo:boolean;  
a,b,c,nb:integer;  
land,chi7:real;  
sign:real;  
x:matrix;  
y:mat;  
p:array[1..8,1..16,1..2] of integer;
```

```
procedure lecture;          {acquisition des donnees}
```

```
var  
i,j:integer;  
begin  
for i:=1 to 13 do readln(f);  
for j:=1 to 64 do  
begin  
for i:=1 to 16 do  
begin  
read(f,x[i,j,1]);  
x[i,j,2]:=0  
end  
end;  
chi7:=26  
end;
```

```
procedure calcul;          {calcul de l'intensite integree et impression des resultat  
s}
```

```
var  
i,j:integer;  
begin  
land:=0;sign:=0;nb:=0;  
for j:=1 to 64 do  
begin  
for i:=1 to 16 do  
begin  
if (p[trunc((i+1)/2),trunc((j+3)/4),1]=2) then x[i,j,2]:=1;  
if x[i,j,2]=1 then  
begin  
sign:=sign+x[i,j,1];  
write(g,round(x[i,j,1]):3,' ')  
end;  
if x[i,j,2]=0 then  
begin  
nb:=nb+1;  
land:=land+x[i,j,1];  
write(g,'*** ')  
end  
end;  
end;  
writelng)
```

```
end;
land:=land/nb;
sign:=sign-(1024-nb)*land;
writeln(g);
writeln(g,'valeur du bruit ',land:8:2);
writeln(g,'valeur du signal integre ',sign:9:2)
end;
```

```
procedure homogene(z1,z2:integer;seuil:real;var w:integer);
```

```
{detection du noyau de la diffraction }
```

```
var i,j:integer;
mean,va:real;
begin
mean:=0;
va:=0;
w:=0;
for i:=2*z1-1 to 2*z1 do
begin
for j:=4*z2-3 to 4*z2 do
begin
mean:=mean+x[i,j,1];
va:=va+sqr(x[i,j,1])
end
end;
mean:=mean/8;
va:=va/8-sqr(mean);
if (7*va/mean>s1) then w:=1;
if (7*va/mean>seuil) then w:=2
end;
```

```
procedure maxi(var a1:integer;var a2:integer);
```

```
{recherche de la plage enregistrant la plus forte moyenne }
```

```
var moy,mtes:real;
i,j,k,l:integer;
begin
moy:=0;mtes:=0;
for i:=1 to 8 do
begin
for j:=1 to 16 do
begin
moy:=0;
for k:=2*i-1 to 2*i do
begin
for l:=4*j-3 to 4*j do moy:=moy+x[k,l,1]
end;
if moy>mtes then
begin
a1:=i;
a2:=j;
mtes:=moy
end
end
end
end;
```

```
procedure testpl(k,l:integer;var m:integer;var n:integer;var x1:matrix);
{test d'agglomeration des plages frontieres }

var
i,j,q,ind,t1,t2:integer;
y:mat;
m01,m02,m1,m2,m3,v2,v3,rap:real;

procedure tri(k1,l1:integer;var y1:mat);
{tri d'une plage pour en ordonner les comptages }
var
i,j,r:integer;
trans:real;
begin
r:=1;
for i:=2*k1-1 to 2*k1 do
begin
for j:=4*l1-3 to 4*l1 do
begin
y1[r]:=x1[i,j,1];
r:=r+1
end
end;
i:=1;
while i<=7 do
begin
for j:=i to 8 do
begin
if y[i]>y[j] then
begin
trans:=y[j];
y[j]:=y[i];
y[i]:=trans
end
end;
i:=i+1
end
end;

procedure teste(nbre:integer;moy:real;var resul:integer);
{test d'appartenance d'un nombre nbre de comptages a la tache de Bragg}
begin
resul:=0;
if (-2*((nb*land+nbre*moy)*ln((nb*land+nbre*moy)/(nb+nbre))-nb*land*ln(land)-nbre*moy*ln(moy))>s2) then resul:=2 ;
end;

procedure agglo(deb,fin:integer;var x2:matrix);
{agglomeration de comptages a la diffraction }
var
i,j,z:integer;
begin
for z:=deb to fin do
begin
i:=2*k-1;
j:=4*l-3;
while ((x2[i,j,1]<>y[z]) or (x2[i,j,2]=1)) do
```

```
begin
if j=4*1 then
begin
i:=i+1;
j:=j-3
end
else j:=j+1
end;
x2[i,j,2]:=1
end
end;

begin          {testp1}
n:=1;
if m=0 then   {cas d'une plage homogene}
begin
m1:=0;
for i:=2*k-1 to 2*k do
begin
for j:=4*1-3 to 4*1 do m1:=m1+x1[i,j,1]
end;
m1:=m1/8;
teste(8,m1,m)
end
else          {cas d'une plage heterogene }
begin
tri(k,1,y);
rap:=1000;
m01:=0;m02:=0;t1:=0;t2:=0;
q:=2;
while q<7 do
begin
m2:=0;v2:=0;
for i:=1 to q do
begin
m2:=m2+y[i];
v2:=v2+sqr(y[i])
end;
m3:=0;v3:=0;
for i:=q+1 to 8 do
begin
m3:=m3+y[i];
v3:=v3+sqr(y[i])
end;
m2:=m2/q;
v2:=v2/q-sqr(m2);
m3:=m3/(8-q);
v3:=v3/(8-q)-sqr(m3);
if (((q-1)*v2/m2+(8-q-1)*v3/m3) <rap ) then
begin
rap:=(q-1)*v2/m2+(8-q-1)*v3/m3;
ind:=q;
m01:=m2;
m02:=m3
end;
q:=q+1
end;
if m01>land then teste(ind,m01,t1);
if t1=2 then agгло(1,ind,x1);
if m02>land then teste((8-ind),m02,t2);
```

```
if t2=2 then agglo(ind+1,8,x1);
if ((t1=2) and (t2=2)) then m:=2
end;
if m=2 then
begin
if ((k>1) and (p[k-1,1,1]<2) and (p[k-1,1,2]<1)) then testpl(k-1,1,p[k-1,1,1],p[
k-1,1,2],x1);
if ((k<8) and (p[k+1,1,1]<2) and (p[k+1,1,2]<1)) then testpl(k+1,1,p[k+1,1,1],p[
k+1,1,2],x1);
if ((1>1) and (p[k,1-1,1]<2) and (p[k,1-1,2]<1)) then testpl(k,1-1,p[k,1-1,1],p[
k,1-1,2],x1);
if ((1<16) and (p[k,1+1,1]<2) and (p[k,1+1,2]<1)) then testpl(k,1+1,p[k,1+1,1],p[
[k,1+1,2],x1)
end
end;
```

```
begin      {program}
assign(f,'b:result.sim');
assign(g,'b:result.hom');
reset(f);
rewrite(g);
lecture;
bo:=false;
for a:=1 to 8 do
begin
for b:=1 to 16 do
begin
homogene(a,b,chi7,p[a,b,1]);
if p[a,b,1]=2 then bo:=true
end
end;
land:=0;
nb:=0;
for a:=1 to 16 do
begin
for b:=1 to 64 do
begin
if (p[trunc((a+1)/2),trunc((b+3)/4),1]=2) then x[a,b,2]:=1;
if (p[trunc((a+1)/2),trunc((b+3)/4),1]=0) then
begin
land:=land+x[a,b,1];
nb:=nb+1
end
end
end;
land:=land/nb;
if bo=false then
begin
maxi(a,b);
p[a,b,1]:=2
end;
for a:=1 to 8 do
begin
for b:=1 to 16 do
begin
if p[a,b,1]=2 then
begin
```

```
if ((a>1) and (p[a-1,b,1]<2) and (p[a-1,b,2]<>1)) then testpl(a-1,b,p[a-1,b,1],p
[a-1,b,2],x);
if ((a<8) and (p[a+1,b,1]<2) and (p[a+1,b,2]<>1)) then testpl(a+1,b,p[a+1,b,1],p
[a+1,b,2],x);
if ((b>1) and (p[a,b-1,1]<2) and (p[a,b-1,2]<>1)) then testpl(a,b-1,p[a,b-1,1],p
[a,b-1,2],x);
if ((b<16) and (p[a,b+1,1]<2) and (p[a,b+1,2]<>1)) then testpl(a,b+1,p[a,b+1,1],
p[a,b+1,2],x)
end
end
end;
calcul;
close(g,c)
end.
```

Les tableaux 3.1 et 3.2 illustrent l'évolution de la qualité de l'estimation de l'intensité intégrée d'une diffraction en fonction de la valeur moyenne du bruit de fond.

Nous avons simulé pour différentes valeurs du bruit deux séries de 25 diffractions, d'intensité intégrée moyenne $s_g = 471$ pour la première (tableau 3.1) et $s_g = 2000$ pour la seconde 5, (Tableau 3.2).

Dans le premier cas, l'intensité intégrée de la diffraction est faible, mais la tache de Bragg n'occupe qu'un nombre restreint de cellules, et le rapport signal/bruit sur la partie effectivement touchée par la diffraction est élevé.

La diffraction simulée dans le second exemple a une intensité intégrée élevée, mais occupe un domaine D important ; le rapport signal/bruit sur le seul domaine D est assez faible.

Les résultats sont à lire de la manière suivante :

La première colonne donne la valeur moyenne du bruit de fond, et la seconde la valeur du rapport signal/ bruit. La troisième colonne indique le biais moyen du ratio intensité estimé/intensité intégrée, notée $|1 - \hat{s}/s_g|$; enfin, la quatrième colonne donne la variance des quantités $(1 - \hat{s}/s_g)$.

Si les résultats sont dans les deux cas de figure assez encourageants, il apparaît une différence notable à la lecture de ces tableaux : Le tableau 1 montre qu'il nous faut descendre jusqu'à un rapport signal/ bruit de l'ordre de 4% pour atteindre un biais moyen de 0.1, biais dépassé dans le tableau 2 alors que le rapport signal/bruit est encore de 15% !

La différence entre ces résultats s'explique par la nature même des diffractions simulées :

Les tableaux 3.3 et 3.4 montrent l'évolution de la qualité de l'estimation en fonction de l'intensité intégrée de la tache. Pour ce faire, nous fixons la

valeur moyenne du bruit de fond ainsi que, dans la mesure du possible, la forme de la tache (peu étendue dans le tableau 3.3 et de moyenne étendue dans le tableau 3.4), et générons deux séries de 25 enregistrements pour différentes valeurs de l'intensité intégrée.

Dans les deux cas, le biais est inférieur à 0.1 dès que l'on atteint un rapport signal/bruit de l'ordre de 6%, et majoré par 0.05 pour des rapports signal/bruit supérieurs à 12%.

Il est intéressant, pour apprécier ces résultats d'observer la nature de tels enregistrements ; on se référera aux figures 3.5 et 3.6 qui représentent un enregistrement dont le rapport signal/bruit est d'environ 6%, et son traitement.

Tableau 3.1 ($\sigma_1 = 3, \sigma_2 = 4, k_2 = 6000$)

λ	s_g / b	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
4	0.115	0.049	0.001
6	0.076	0.071	0.0019
8	0.057	0.069	0.0038
10	0.046	0.089	0.0039
12	0.038	0.096	0.0042
14	0.033	0.121	0.0080
16	0.029	0.140	0.0081

Tableau 3.2 ($\sigma_1 = 10, \sigma_2 = 12, k_2 = 25000$)

λ	s_g / b	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
4	0.488	0.066	0.0008
6	0.326	0.057	0.0012
8	0.244	0.077	0.0018
10	0.195	0.079	0.0015
12	0.163	0.065	0.0017
14	0.140	0.105	0.0076
16	0.122	0.123	0.0073

Tableau 3.3

k_2	s_g / b	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
1000	0.026	0.21	0.021
1500	0.039	0.19	0.020
2000	0.051	0.12	0.005
2500	0.065	0.07	0.0047
3000	0.079	0.07	0.0024
3500	0.093	0.07	0.0025
4000	0.104	0.06	0.0013
4500	0.117	0.04	0.0013
5000	0.131	0.04	0.0012
5500	0.142	0.03	0.0007
6000	0.158	0.03	0.0006
6500	0.169	0.03	0.0005
7000	0.183	0.03	0.0006
7500	0.197	0.02	0.0004

Tableau 3.4

k_2	s_g / b	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
1000	0.027	0.36	0.04
1500	0.039	0.24	0.021
2000	0.053	0.14	0.015
2500	0.066	0.08	0.0035
3000	0.078	0.08	0.0034
3500	0.091	0.08	0.003
4000	0.103	0.07	0.002
4500	0.117	0.06	0.002
5000	0.136	0.05	0.0009
5500	0.145	0.05	0.001
6000	0.159	0.05	0.001
6500	0.169	0.04	0.0009
7000	0.183	0.04	0.0007
7500	0.198	0.03	0.0005

CARACTERISTIQUE d un ENREGISTREMENT

ecart type de la premiere marginale 7.00
 ecart type de la seconde marginale 8.00
 valeur de la constante multiplicative 2500
 valeur du bruit de fond moyen demande 3.000

Valeurs simulees

bruit de fond moyen effectivement genere 3.02
 intensite integree de la diffraction simulee 201.00

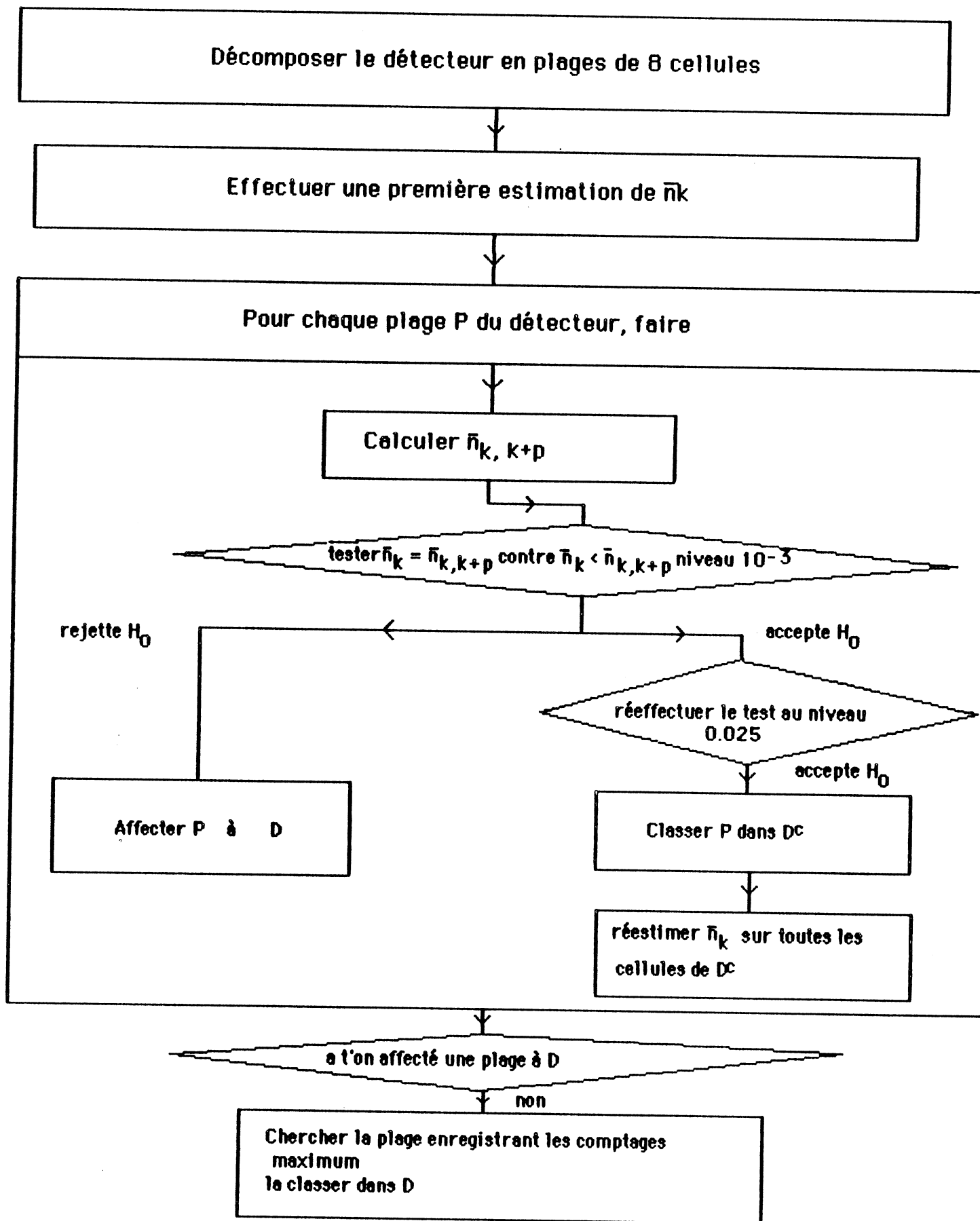
Simulation d un enregistrement

3	2	2	3	5	3	3	4	2	3	1	4	5	3	4	3
3	1	1	3	3	2	2	3	8	3	2	3	6	2	1	3
5	2	6	3	4	3	2	3	0	3	3	1	2	3	2	3
3	7	5	2	1	3	1	2	1	3	6	5	0	5	2	5
3	1	3	0	2	3	5	2	4	3	3	4	2	2	4	4
1	0	4	2	3	6	3	1	2	5	2	1	5	2	3	2
5	1	4	5	4	6	3	2	6	4	4	3	2	2	6	3
3	3	3	0	0	2	3	3	5	4	4	3	1	0	1	3
6	2	3	5	4	4	0	3	4	1	1	2	0	2	0	3
4	4	8	5	5	4	2	5	5	3	2	2	0	2	3	3
4	4	0	1	4	2	2	2	3	2	4	4	2	1	1	8
2	4	0	1	2	3	2	6	1	3	2	2	1	4	3	2
3	0	4	3	3	3	2	4	3	4	4	4	5	3	2	2
4	3	4	3	3	3	4	4	3	4	2	5	3	0	4	6
1	4	5	3	3	4	3	5	2	2	4	2	4	2	4	1
1	2	5	4	2	3	4	4	2	3	5	4	1	1	3	1
4	2	1	2	1	1	1	4	2	2	4	2	0	0	3	3
6	3	4	6	3	3	4	4	4	3	3	7	6	3	2	1
5	2	5	3	3	3	3	4	0	3	6	3	1	3	3	3
6	1	3	2	2	2	1	2	2	1	0	7	2	2	4	2
7	2	1	2	2	3	1	6	5	5	6	5	2	3	3	4
1	0	5	1	1	5	2	2	3	1	2	4	3	2	2	4
5	6	3	3	2	2	1	7	6	1	4	2	6	5	1	2
2	4	1	6	3	4	2	8	2	4	0	7	5	2	2	5
2	1	3	4	2	2	1	3	2	2	3	2	0	2	3	3
9	3	3	2	2	7	4	5	7	2	3	6	1	5	4	5
3	3	4	4	3	3	3	2	2	6	4	0	3	1	2	1
6	5	3	2	2	7	6	4	6	3	7	4	5	3	1	2
2	1	4	2	2	4	4	4	4	3	4	3	1	1	3	2
4	2	0	2	1	3	6	5	6	7	2	4	3	4	4	4
1	2	4	4	4	4	12	13	6	6	1	5	3	4	1	5
6	3	7	1	4	6	4	6	8	6	11	4	4	4	4	6
4	4	4	8	3	7	7	9	10	11	8	4	3	7	7	6
4	2	1	5	5	5	5	16	8	6	5	1	5	2	4	2
1	4	4	3	1	5	4	14	6	6	3	3	2	5	3	1
2	4	4	2	5	1	2	7	3	6	5	1	1	1	7	5
3	2	3	3	0	3	7	3	7	6	4	4	6	6	2	4
1	2	2	1	3	5	2	2	5	3	3	5	2	5	2	1
2	5	4	7	2	2	2	2	2	3	3	2	3	3	3	3
4	3	2	0	3	3	6	4	4	7	1	2	3	5	1	1
3	0	6	1	2	3	6	6	3	2	0	4	2	3	5	4
3	0	2	4	4	4	2	4	4	3	4	6	3	5	2	4
3	7	3	4	2	1	3	7	7	3	3	3	5	4	3	2
3	2	5	3	5	5	2	1	2	2	6	4	2	3	4	3
3	1	4	2	1	5	6	4	3	3	1	6	3	2	5	0
5	4	5	3	5	4	1	5	3	3	3	5	3	4	3	2
4	2	4	1	5	5	1	1	1	1	3	3	5	3	1	0
4	2	2	0	0	5	4	1	0	4	5	6	4	4	7	2
6	2	3	4	2	3	4	2	1	4	4	6	4	3	2	5
4	2	4	5	4	1	4	0	3	3	0	2	4	4	4	2
1	4	3	4	0	3	2	3	5	2	2	3	4	3	4	4
5	1	3	1	2	5	2	2	4	2	4	2	6	2	3	3
2	2	5	2	7	7	1	1	1	1	1	1	1	4	0	4
3	3	1	7	4	1	4	6	3	2	0	4	1	4	5	4
5	2	2	2	7	4	7	2	1	2	0	4	2	5	7	3
4	1	5	3	2	0	3	4	2	2	5	1	2	6	1	4
1	4	5	0	2	9	1	3	1	3	6	1	2	5	4	3
2	1	3	5	6	4	4	1	1	3	4	2	3	3	1	3
1	4	1	2	2	2	3	3	4	4	4	3	3	3	5	4
4	5	2	6	0	3	7	4	1	2	4	3	3	3	1	1
3	4	0	4	1	2	2	2	4	3	3	5	4	5	1	2

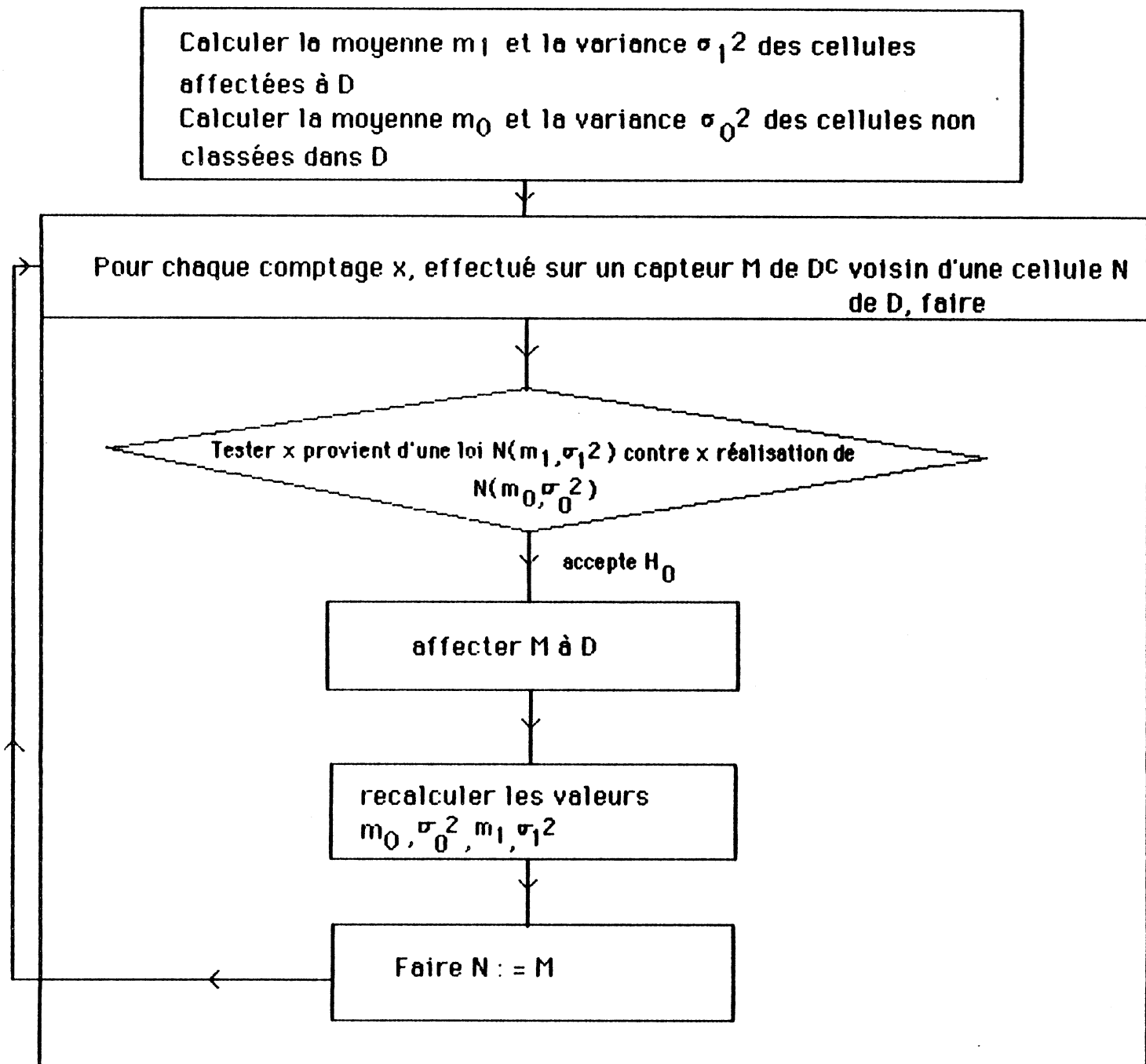
Figure 3.5.

ANNEXE 4

Avant de lister le programme relatif au calcul de l'intensité intégrée d'une tache de BRAGG perturbée par un bruit de fond élevé, nous en explicitons la conception par le biais de deux organigrammes, qui figurent l'un la phase de détection du noyau de la diffraction, l'autre la phase d'agglomération des cellules autour du dit noyau. Dans une dernière partie, nous illustrons ce programme en présentant quelques résultats, ainsi que quelques exemples de traitement.



Organigramme de la phase "Détection du noyau"



Organigramme de la phase d'agglomération autour du noyau


```
program traitement;
```

```
  {CALCUL de l'intensite integree d'une diffraction}  
  {cas d'un bruit de fond moyen eleve}
```

```
type  
matrix=array[1..16,1..64,1..2] of real;  
var  
f,g:text;  
ma:matrix;  
u,v,p,q,nb:integer;  
va0,va1,tot,bruit,signal,br:real;  
bo:boolean;
```

```
procedure lecture; {acquisition des donnees}
```

```
var  
i,j:integer;  
begin  
for i:=1 to 13 do readln(f);  
for j:=1 to 64 do  
begin  
for i:=1 to 16 do  
begin  
read(f,ma[i,j,1]);  
ma[i,j,2]:=0  
end;  
readln(f)  
end  
end;
```

```
procedure initb; {premiere estimation de sqrt(n*landa(t,t+h1))}
```

```
var  
i,j:integer;  
begin  
br:=0;  
for i:=1 to 2 do  
begin  
for j:=1 to 4 do br:=br+sqrt(ma[i,j,1])+sqrt(ma[14+i,j,1])+sqrt(ma[i,60+j,1])+sqrt(ma[14+i,60+j,1])  
end;  
br:=br/32;  
nb:=32  
end;
```

```
procedure maxi(var a1:integer;var a2:integer);
```

```
var  
moy,mtes:real;  
i,j,k,l:integer;  
begin  
moy:=0;  
mtes:=0;  
for i:=1 to 8 do  
begin  
for j:=1 to 16 do  
begin  
moy:=0;  
for k:=2*i-1 to 2*i do  
begin  
for l:=4*j-3 to 4*j do moy:=moy+ma[k,l,1]  
end;  
end;
```

```
if moy>mtes then
begin
a1:=i;
a2:=j;
mtes:=moy
end
end
end
end;
```

```
procedure teste(k,l:integer;var moy:real;var nbre:integer;var m2:matrix;var bo1:
boolean);
```

```
{detection du noyau de la diffraction }
```

```
var
i,j:integer;
ms,land,seuil:real;
begin
ms:=0;
seuil:=3.1;{cela correspond a un niveau alpha egal a 0.001}
for i:=k to k+1 do
begin
for j:=1 to 1+3 do ms:=ms+sqrt(m2[i,j,1])
end;
ms:=ms/8;
land:=sqrt(1/(4*nbre)+1/32)*seuil;
if ((ms-moy)>land) then
begin
bo1:=true;
for i:=k to k+1 do
begin
for j:=1 to 1+3 do m2[i,j,2]:=1
end
end
else
begin
if ((ms-moy)<sqrt(1/(4*nbre)+1/32)*1.96) then
begin
nbre:=nbre+8;
moy:=(nbre*moy+8*ms)/(nbre+8)
end
end
end;
```

```
procedure tesaggl(p1,q1:integer; var m0:real;var v0:real;var na:integer;var m1:r
eal;var v1:real;var h:matrix);
```

```
{agglomeration des points frontieres }
```

```
begin
if (exp(-sqrt(h[p1,q1,1]-m0)/(2*v0))/(sqrt(v0)) < exp(-sqrt(h[p1,q1,1]-m1)/(2*v1))
/sqrt(v1)) then
begin
h[p1,q1,2]:=1;
v0:=((v0+sqrt(m0))*na-sqrt(h[p1,q1,1]))/(na-1);
m0:=(na*m0-h[p1,q1,1])/(na-1);
v0:=v0-sqrt(m0);
v1:=((v1+sqrt(m1))*(1024-na)+sqrt(h[p1,q1,1]))/(1024-na+1);
m1:=((1024-na)*m1+h[p1,q1,1])/(1024-na+1);
```

```
v1:=v1-sqr(m1);
na:=na-1;
if h[p1+1,q1,2]=0 then tesaggl(p1+1,q1,m0,v0,na,m1,v1,h);
if h[p1-1,q1,2]=0 then tesaggl(p1-1,q1,m0,v0,na,m1,v1,h);
if h[p1,q1-1,2]=0 then tesaggl(p1,q1-1,m0,v0,na,m1,v1,h);
if h[p1,q1+1,2]=0 then tesaggl(p1,q1+1,m0,v0,na,m1,v1,h)
end
end;
```

```
procedure ecriture; {impression des resultats }
var
i,j:integer;
begin
writelng,'TRAITEMENT d une TACHE de BRAGG ';
writelng;
for i:=1 to 64 do
begin
for j:=1 to 16 do
begin
if ma[j,i,2]=1 then write(g,round(ma[j,i,1]):3,' ')
else write(g,'***',' ')
end;
write(g,' ');
writelng;
end;
writelng;
writelng,'Valeur estimee du bruit de fond moyen ',bruit:7:2);
writelng,'nombre de cellules affectees a la partie ne percevant pas la diffract
ion ',nb:3);
writelng,' valeur estimee de l intensite integree de la diffraction simulee ',s
ignal-bruit*(1024-nb):8:2)
end;
```

```
begin {program}
assign(f,'b:donnees');
assign(g,'b:result');
reset(f);
rewrite(g);
lecture;
bruit:=0;
signal:=0;
initb;
bo:=false;
p:=1;
while (p<=15) do
begin
q:=1;
if ((p=15) or (p=1)) then q:=5;
while (q<=61) do
begin
teste(p,q,br,nb,ma,bo);
q:=q+4;
if (((p=1) or (p=15)) and (q=61)) then q:=62
end;
p:=p+2
end;
if bo=false then
begin
maxi(p,q);
ma[p,q,2]:=1
```

```
end;
  bruit:=0;
  va0:=0;val:=0;
  signal:=0;
  nb:=0;
  for q:=1 to 64 do
  begin
  for p:=1 to 16 do
  begin
  if ma[p,q,2]=0 then
  begin
  nb:=nb+1;
  bruit:=bruit+ma[p,q,1];
  va0:=va0+sqr(ma[p,q,1])
  end;
  if ma[p,q,2]=1 then
  begin
  signal:=signal+ma[p,q,1];
  val:=val+sqr(ma[p,q,1])
  end
  end;
  end;
  bruit:=bruit/nb;
  va0:=va0/nb-sqr(bruit);
  signal:=signal/(1024-nb);
  val:=val/(1024-nb)-sqr(signal);
  for p:=2 to 15 do
  begin
  for q:=2 to 63 do
  begin
  if ((ma[p,q,2]=0) and ((ma[p+1,q,2]=1) or (ma[p-1,q,2]=1) or (ma[p,q-1,2]=1) or
  (ma[p,q+1,2]=1))) then
  begin
  tesaggl(p,q,bruit,va0,nb,signal,val,ma)
  end
  end
  end;
  signal:=0;
  for q:=1 to 64 do
  begin
  for p:=1 to 16 do
  begin
  if ma[p,q,2]=1 then
  begin
  signal:=signal+ma[p,q,1]
  end
  end
  end;
  writeln('BRUIT ESTIME ',bruit:7:2,' sur ',nb:3,' CELLULES ');
  writeln('VALEUR ESTIMEE du SIGNAL INTEGRE ',signal-bruit*(1024-nb):8:2);
  ecrisure;
  close(g,u)
  end.
```

Les tableaux IV.1 et IV.2 illustrent la bonne qualité générale de l'estimation de l'intensité intégrée d'une diffraction par cette seconde méthode.

Le premier de ces tableaux correspond à la détection d'une tache d'intensité intégrée peu élevée, et de faible étendue. Le rapport s_g/b n'évolue que très faiblement, et le biais moyen des ratios \hat{s}/s_g varie très peu. Dans tous les cas, il reste inférieur à 6% alors que le rapport signal/bruit atteint 1%, ce qui est très raisonnable.

Le second tableau figure la détection d'une tache d'intensité intégrée élevée, mais plus étendue que la précédente ; là encore le biais du ratio \hat{s}/s_g varie peu, et, s'il est en règle générale plus fort que dans le tableau IV.1, il reste dans des proportions somme toutes raisonnables.

Nous illustrons dans les tableaux IV.3 et IV.4 l'évolution de la qualité de l'estimation en fonction de l'intensité intégrée de la tache de BRAGG. Nous procédons pour ce faire de la même manière que dans l'étude faite en annexe 3 ; le tableau IV.3 correspond aux simulations d'une tache de faible amplitude, l'autre aux simulations d'une diffraction d'ampleur moyenne (Cela revient à énoncer le fait que le rapport signal/bruit restreint à la seule partie du détecteur touchée par la tache de BRAGG est plus faible dans le second cas que dans le premier).

Il ressort de ces résultats que l'on atteint un biais inférieur à 0.1 pour un rapport signal/bruit de 1% pour la première simulation et de l'ordre de 1.2% pour la seconde, ce qui est encore satisfaisant (cf figures 4.5 et 4.6). Ces résultats ainsi que ceux qui figurent en annexe 3 sont naturellement à moduler en fonction de la forme de la tache de BRAGG simulée : c'est ainsi que d'une manière générale, l'estimation de l'intensité intégrée d'une diffraction de faible amplitude sera meilleure que celle d'une tache de taille élevée. D'une manière générale, les procédures que nous proposons ont tendance à introduire une troncature importante des pics étalés.

Tableau IV.1 ($\sigma_1 = 4, \sigma_2 = 5, k_2 = 10000$)

λ	s_g / b %	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
30	2.6	0.045	0.0018
40	1.9	0.048	0.0017
50	1.6	0.053	0.0026
60	1.3	0.046	0.0025
70	1.1	0.055	0.0025
80	1	0.060	0.0024

Tableau IV.2 ($\sigma_1 = 8, \sigma_2 = 10, k_2 = 29000$)

λ	s_g / b %	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
30	7.6	0.059	0.0011
40	5.7	0.072	0.0017
50	4.5	0.067	0.0019
60	3.8	0.063	0.0018
70	3.2	0.076	0.0021
80	2.8	0.076	0.0019

Tableau IV.3

k_2	s_g / b (%)	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
5000	0.78	0.187	0.018
6000	0.93	0.111	0.0069
7000	1.09	0.082	0.005
8000	1.25	0.044	0.0011
9000	1.40	0.060	0.0015
10000	1.56	0.055	0.0015
11000	1.72	0.045	0.0011
12000	1.87	0.045	0.0009

k_2	s_g / b (%)	$ 1 - \hat{s}/s_g $	$\sigma^2(1 - \hat{s}/s_g)$
5000	0.78	0.20	0.029
6000	0.94	0.13	0.009
7000	1.09	0.12	0.006
8000	1.25	0.07	0.002
9000	1.40	0.09	0.003
10000	1.56	0.09	0.003
11000	1.72	0.08	0.003
12000	1.87	0.08	0.002

Tableau IV.4

CARACTERISTIQUE d un ENREGISTREMENT

ecart type de la premiere marginale 7.00
 ecart type de la seconde marginale 8.00
 valeur de la constante multiplicative 10000
 valeur du bruit de fond moyen demande 50.000

Valeurs simulees

bruit de fond moyen effectivement genere 49.77
 intensite integree de la diffraction simulee 827.00

Simulation d un enregistrement

39	51	50	51	51	55	50	47	45	54	53	44	52	47	52	46
47	42	60	49	53	47	56	50	44	51	54	53	52	51	41	47
47	61	56	45	55	50	47	51	52	51	61	52	52	46	56	48
52	39	52	62	51	46	50	51	44	55	42	47	45	35	49	54
49	58	54	38	48	38	49	47	46	50	48	55	50	50	44	41
51	38	43	46	49	59	47	38	54	54	47	47	54	56	50	51
45	51	38	57	49	47	46	62	53	53	36	44	41	47	45	49
45	65	45	57	39	66	53	41	49	48	42	59	41	40	49	62
38	63	40	48	48	47	67	42	53	46	53	37	42	39	50	47
39	52	54	52	53	55	36	47	43	64	58	43	48	51	48	64
49	59	46	39	41	40	45	49	47	44	44	61	52	50	47	49
47	49	42	51	39	57	45	46	50	40	47	46	53	60	40	51
35	42	44	45	52	48	59	43	51	46	55	49	57	62	41	47
60	48	47	49	61	59	46	47	35	56	55	57	43	46	61	52
56	46	47	53	53	47	39	43	42	55	49	57	44	41	53	52
66	54	36	53	39	44	44	51	45	41	44	53	47	50	43	50
60	56	41	52	51	54	77	42	62	52	47	59	51	45	64	64
41	47	56	49	65	50	46	36	59	46	50	56	54	46	62	50
40	48	57	66	50	53	38	46	45	50	53	55	60	44	44	57
50	49	59	52	43	48	50	45	56	51	62	58	52	37	53	42
43	44	56	52	55	47	60	59	62	46	45	45	48	47	41	64
52	58	62	55	54	55	40	43	46	38	46	48	50	61	55	51
50	50	40	46	51	44	51	60	45	48	37	47	58	32	52	58
53	51	49	56	69	48	51	46	52	46	57	42	47	40	53	55
52	45	53	52	50	49	38	61	64	47	57	45	40	59	60	62
46	48	65	61	54	50	43	54	47	45	49	43	47	49	56	51
41	52	53	40	46	58	55	65	51	56	44	56	43	54	57	54
57	52	54	62	64	39	54	62	65	56	44	35	47	54	49	49
59	44	45	50	71	52	62	56	77	61	52	42	48	51	38	53
49	47	47	59	55	65	53	68	74	76	59	46	62	39	51	62
62	37	64	46	46	48	72	76	63	74	57	49	40	51	59	45
49	42	49	50	66	59	77	89	75	82	61	47	60	52	48	64
42	50	51	50	49	48	71	80	79	62	45	47	44	44	58	49
50	45	48	52	39	49	61	65	74	66	51	53	43	49	59	51
53	40	58	58	48	51	77	81	58	57	57	53	36	51	46	53
44	39	48	46	39	53	79	57	73	57	51	55	51	43	59	61
62	53	50	59	44	44	48	53	55	55	45	53	52	47	65	56
50	57	40	41	50	51	50	54	59	53	51	49	55	55	52	56
55	46	41	62	46	38	49	44	61	50	47	39	52	52	46	54
41	52	43	54	56	52	28	54	56	61	36	48	35	49	40	53
49	42	46	47	46	52	44	48	66	59	40	48	48	55	43	41
41	47	54	49	56	50	67	42	51	69	42	44	60	48	51	47
66	44	43	34	46	62	58	42	57	52	47	48	51	56	48	56
47	38	59	40	53	56	47	49	39	51	56	49	50	45	44	53
50	50	58	43	50	40	62	45	64	48	50	57	40	43	56	47
40	34	54	48	51	50	51	59	53	50	54	46	51	56	47	66
56	43	61	50	60	56	56	68	48	43	40	47	52	54	57	49
55	46	49	54	49	58	44	50	55	53	48	38	46	60	61	37
47	62	46	49	48	47	49	51	54	39	48	45	43	49	56	51
50	50	56	48	55	52	55	47	43	50	52	41	56	50	54	41
47	50	51	40	55	42	50	57	40	41	45	54	44	52	50	43
60	58	49	53	52	63	52	44	51	68	48	50	51	53	63	51
40	49	44	64	42	51	48	50	52	55	48	62	53	43	53	39
58	51	46	45	56	57	51	55	49	47	47	31	50	43	62	47
45	44	53	50	51	53	46	55	57	47	49	58	46	61	46	56
38	52	50	54	49	62	53	56	37	57	47	53	40	43	55	53
50	42	49	44	43	46	42	56	42	47	45	49	46	33	43	48
51	45	60	57	45	52	47	39	37	48	49	45	56	49	49	45
45	37	53	41	46	55	58	48	52	50	55	46	48	40	53	50
49	49	50	58	53	51	46	40	73	53	52	52	39	45	52	49
60	59	46	52	38	50	54	49	44	47	52	56	49	71	47	45
56	44	53	55	54	48	39	49	54	47	62	54	39	51	66	43
39	45	51	48	55	43	45	37	55	44	49	50	50	57	46	43
44	46	49	52	57	61	59	52	52	43	58	51	42	50	49	60

Figure 4.5.

BIBLIOGRAPHIE

- [1] BARRA J.R. : Notions fondamentales de statistiques mathématiques. 1971, DUNOD. p.170.
- [2] BREIMAN L. : The Poisson tendency in traffic distributions, 1963, AMS, 34, p. 308-11
- [3] BROWN M. : Statistical analysis of non homogeneous Poisson processes, 1972, WILEY, p. 67-88.
- [4] BROWN M. : Discrimination of Poisson processes, 1971, AMS, 42, p. 773-776.
- [5] CHATTERJI S.D.; MANDREKAR V. : Equivalence and singularity of Gaussian measures and applications, 1978, Proba analysis and related topics.
- [6] CURIONI M. : Estimation de la densité des processus de Poisson non homogènes, 1977, Thèse de 3^{eme} cycle.
- [7] FELLER W. : An introduction to probability theory and its applications, 1966, WILEY.

- [8] FILHOL A. ; THOMAS M. : Traitement des données de D 19 A équipé d'un multidétecteur bidimensionnel "oeil de mouche", 1981, Rapport interne ILL, (81 TH 08 T).
- [9] GOLDMAN J.R. : Stochastic point processes : limit theorems, 1968, AMS, 39, p. 771-779.
- [10] GREGOIRE G. : Quelques outils en vue de la modélisation des processus ponctuels et des processus de clustering, 1980, Thèse 3^{eme} cycle, p. 5.
- [11] KRICKEBERG K. : Processus ponctuels, 1981, Lecture Notes in maths.
- [12] LEHMAN E. L. : Testing statistical hypotheses, 1959, WILEY, p. 141.
- [13] MEAD R. : A test for spatial pattern at several scales using data from a grid of contiguous quadrats, 1974, Biometrics.
- [14] NEVEU J. : Processus ponctuels, 1977, Lecture Notes in maths.
- [15] NEVEU J. : Bases mathématiques du calcul des probabilités, 1964, Masson.
- [16] PIELOU E.C. : An introduction to mathematical ecology, 1969, WILEY, p. 91.
- [17] ROMEDER J.M. : Méthodes et programmes d'analyse discriminante, 1973, DUNOD.

- [18] SHIUE W.K.; BAIN L.J. : Experiment size and power comparisons for two-samples Poisson tests, 1982, JRSS applied stat,31, p. 130-134.
- [19] SICHEL H.S. : On a significance test for two Poisson variables, 1973, applied stat, 22, p. 50-58.
- [20] SOLER J.L. : Notion de liberté en statistique mathématique, 1970, Thèse 3^{eme} cycle, p. 29.
- [21] VAN BELLINGHEN M. : Etude de la densité moyenne d'une mesure aléatoire, 1978, thèse 3^{eme} cycle.
- [22] VAN MEERSHE, FENEAU-DUPONT : Introduction à la cristallographie et à la chimie structurale, 1977, OYEZ.
- [23] WILKS S.S. : Mathematical statistics, 1962, WILEY.



DERNIERE PAGE D'UNE THESE

3È CYCLE, DOCTEUR INGÉNIEUR OU UNIVERSITÉ

Vu les dispositions de l'arrêté du 16 avril 1974,

Vu les rapports de M. A. ANTONIADIS.....

M. A. LE BRETON.....

M. Jean-Pierre LACOMBE..... est autorisé
à présenter une thèse en vue de l'obtention du grade de DOCTEUR 3^e cycle en
Mathématiques Appliquées.....

Grenoble, le 4 11 1985

Le Président de l'Université Scientifique
et Médicale



P/ M. TANCHE
φ
D. TANCHE
Vice-Président



Résumé de thèse

La première partie de cette thèse est consacrée à l'étude statistique des processus de Poisson non homogènes et spatiaux . Nous y définissons un test de type Neyman-Pearson concernant la mesure intensité de ces processus . Nous énonçons des conditions pour lesquelles la consistance du test est assurée , et d'autres entraînant la normalité asymptotique de la statistique de test .

Dans la seconde partie de ce travail , nous étudions certaines techniques de traitement statistique de champs poissonniens et leurs applications à l'étude d'un multidétecteur de particules . Nous proposons en particulier des tests de qualité de l'appareillage ainsi que des méthodes d'extraction du signal .

Mots clés

Processus de Poisson - Tests d'hypothèses - Normalité asymptotique - Multidétecteur bidimensionnel .

