



HAL
open science

Modélisation stochastique du vieillissement en temps discret

Cyril Bracquemond

► **To cite this version:**

Cyril Bracquemond. Modélisation stochastique du vieillissement en temps discret. Modélisation et simulation. Institut National Polytechnique de Grenoble - INPG, 2001. Français. NNT: . tel-00004670

HAL Id: tel-00004670

<https://theses.hal.science/tel-00004670>

Submitted on 16 Feb 2004

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.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

THÈSE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité: MATHÉMATIQUES APPLIQUÉES

préparée au Laboratoire de Modélisation et Calcul

dans le cadre de l'école doctorale **Mathématiques, Sciences et Technologies de l'Information**

présentée et soutenue publiquement

par

Cyril Bracquemond

le 26 octobre 2001

Titre :

**Modélisation stochastique
du vieillissement en temps discret**

Directeur de thèse :

Olivier Gaudoin

JURY

M. Soler	, Président
M. Nikulin	, Rapporteur
M. Linnios	, Rapporteur
M. Gaudoin	, Directeur de thèse
M. Chevalier	, Co-encadrant
M. Ligeron	, Examineur

Remerciements

Mes premiers remerciements vont à Jean-Louis Soler, professeur à l'INP Grenoble, qui m'a fait l'honneur de présider le jury.

Je tiens à exprimer ma profonde gratitude envers Mikhail Nikulin, professeur à l'Université Bordeaux II, pour sa lecture très attentive de mon manuscrit et notamment pour ses nombreux commentaires et références concernant le chapitre traitant des tests d'adéquation.

Je souhaite exprimer mes vifs remerciements à Nikolaos Limnios, professeur à l'Université Technologique de Compiègne, qui a accepté de rapporter ma thèse. Ses remarques appropriées et les multiples perspectives de recherche qu'il entrevoit témoignent de son intérêt pour ce domaine et également de sa volonté d'explorer ce sujet quelque peu délaissé.

Je tiens à remercier plus chaleureusement Olivier Gaudoin, maître de conférences à l'INP Grenoble, d'avoir encadré cette thèse industrielle avec autant de zèle. Je lui suis reconnaissant de m'avoir transmis tout autant des compétences statistiques que des méthodes de travail. Sans nul doute que sans tout le temps qu'Olivier m'a consacré, cette thèse n'aurait pas eu la même valeur, aussi bien au niveau mathématique qu'au niveau de l'applicabilité.

Je tiens à remercier Daniel Perez, responsable du service Systèmes et Ingénierie de Services (SIS) de Schneider Electric, de m'avoir accueilli au sein de son équipe depuis mon stage de DESS. Ses compétences statistiques alliées à son excellente connaissance des produits de Schneider Electric ont permis à cette thèse d'avoir de nombreuses possibilités d'applications pratiques en interne. Applications dont sera chargé Michel Bonneton, que je remercie sincèrement de s'être investi si rapidement pour reprendre mes travaux de recherche.

La recherche de données réelles (mais *discrètes*) fut laborieuse et je serai éternellement reconnaissant à Claude Grelier, de Schneider Electric, et André Lannoy, d'Electricité de France, de m'avoir fourni leurs données de défaillance. Leurs avis pratiques d'experts sur les résultats d'analyse de ces données et leur intérêt pour les modèles de fiabilité en temps discret auront été une source de motivation pour explorer tous les domaines de la fiabilité en temps discret.

J'ai été sincèrement ravi de travailler avec les membres de l'équipe SIS dans une ambiance toujours très contractée, où les échanges furent très fructueux. Je ne sais pas si j'ai correctement répondu à leurs interrogations mathématiques, mais ces trois ans furent pour moi riches d'enseignements sur le plan professionnel et personnel (en électronique, électromécanique, syndicalisme, motocyclisme, coinche, etc). Je remercie également tous mes amis et les thésards du LMC, avec une mention spéciale pour Jean-François qui m'a également encouragé et supporté, même jusqu'à des heures tardives.

Je remercie Séverine qui a dû trouver que le temps, durant ces trois ans de thèse, fut tout sauf discret.

Enfin, je remercie toute ma famille, et plus particulièrement mes parents et mes grands-parents qui m'ont toujours soutenu et permis de mener à ma guise mes études, beaucoup moins brillantes que celles de Denis Poisson, également pithivérien.

Table des matières

1	Introduction	11
1.1	Contexte de l'étude	11
1.2	Présentation des appareils électromécaniques	11
1.2.1	Les différentes fonctions du disjoncteur	12
1.2.2	Les différentes types de défaillance du disjoncteur	13
1.3	Les essais chez Schneider Electric	13
1.3.1	Les essais en phase de développement	14
1.3.2	Les essais en phase de lancement	14
1.3.3	Les essais en phase d'exploitation	14
1.4	Analyse du retour d'expérience	15
1.5	Objectifs de la thèse	15
1.6	Plan de la thèse	16
2	Les grandeurs de la fiabilité des systèmes non réparables en temps discret	19
2.1	Rappels des notions de fiabilité en temps continu	19
2.2	Définition des grandeurs caractéristiques de la fiabilité en temps discret	20
2.3	Quelques résultats	22
2.3.1	Sur le taux de défaillance	22
2.3.2	Sur le MTTF	23
2.3.3	Sur la durée de vie résiduelle moyenne	24
2.4	Correspondances entre les différentes grandeurs de la fiabilité en temps continu et en temps discret	25
2.5	Notions de vieillissement	25
2.5.1	Notions de vieillissement en temps continu	25
2.5.2	Notions de vieillissement en temps discret	28
2.6	Simulation de Monte-Carlo des modèles de durée de vie en temps discret	36
2.6.1	Méthode utilisant le taux de défaillance	36
2.6.2	Méthode d'inversion	36
2.6.3	Utilisation de modèles en temps continu	37

3	Modèles de durée de vie en temps discret	38
3.1	Bibliographie chronologique	38
3.2	Lois discrètes usuelles	39
3.2.1	La loi géométrique	39
3.2.2	La loi binomiale négative décalée	40
3.2.3	La loi de Poisson décalée	41
3.3	Les modèles en temps discrets issus du temps continu	42
3.3.1	Modèles en temps discret aux caractéristiques identiques à un modèle en temps continu	42
3.3.2	Modèles discrets dérivés du temps continu par discrétisation de l'échelle du temps	48
3.4	Modèles basés sur l'urne de Pólya	53
3.4.1	La loi de Pólya inverse IFR	54
3.4.2	La loi de Pólya inverse DFR	56
3.4.3	Les lois de Salvia et Bollinger	58
3.4.4	Les lois de Salvia et Bollinger généralisées	59
3.4.5	La loi d' Eggenberger-Pólya	59
3.5	Conclusion	61
4	Vers une nouvelle définition du taux de défaillance en temps discret	62
4.1	Quelques problèmes avec la définition du taux de défaillance usuel	62
4.2	Une nouvelle définition du taux de défaillance	64
4.3	Relation avec les autres grandeurs de la fiabilité	66
4.3.1	Relation avec le taux de défaillance usuel	66
4.3.2	Cas des modèles issus du temps continu	67
4.3.3	Relation avec les autres grandeurs de la fiabilité	68
4.4	Applications aux modèles de fiabilité en temps discret	68
4.4.1	Loi géométrique	68
4.4.2	Lois de Weibull	69
4.5	Conclusion	69
5	Estimation non paramétrique	71
5.1	Notations	71
5.2	La fiabilité empirique	72
5.3	Le taux de défaillance empirique	79
5.3.1	Définition	79
5.3.2	Loi exacte du taux de défaillance empirique	80
5.3.3	Loi asymptotique du taux de défaillance empirique	83
5.3.4	Relation entre R et λ	89

5.4	Le nouveau taux de défaillance empirique	90
5.4.1	Définition	90
5.4.2	Loi asymptotique du nouveau taux de défaillance empirique	91
5.5	La durée de vie résiduelle moyenne empirique	93
5.6	Conclusion	95
6	Estimation paramétrique	96
6.1	La loi géométrique	96
6.1.1	Estimation ponctuelle	97
6.1.2	Estimation par intervalle de confiance	103
6.1.3	Comparaison discret/continu	106
6.2	La loi de Weibull de type I	108
6.3	La loi de Weibull de type III	111
6.4	La loi de Pólya inverse IFR	112
6.5	La loi d'Éggenberger-Pólya	113
6.6	La loi logistique tronquée discrétisée	115
6.7	La loi S	116
6.8	Conclusion	116
7	Tests d'adéquation	118
7.1	Tests graphiques d'adéquation	118
7.1.1	La loi géométrique	119
7.1.2	La loi de Weibull de type I	122
7.1.3	La loi de Weibull de type III	125
7.1.4	La loi de Pólya inverse IFR	126
7.1.5	La loi d'Éggenberger-Pólya	126
7.1.6	La loi S	127
7.2	Tests statistiques d'adéquation : Introduction et notations	127
7.2.1	Test du χ^2	128
7.2.2	Tests basés sur la fonction de répartition empirique	130
7.2.3	Tests basés sur la fonction génératrice empirique	133
7.2.4	Tests lisses de Neyman	135
7.3	Tests basés sur la transformation de Smirnov généralisée	136
7.3.1	Test d'adéquation à une loi complètement spécifiée	136
7.3.2	Test d'adéquation à une famille paramétrée de lois	137
7.4	Application à la loi géométrique	138
7.4.1	Test de Kocherlakota et Kocherlakota	138
7.4.2	Test de Rueda, Pérez-Abreu et O'Reilly	139
7.4.3	Test de Baringhaus et Henze	140

7.4.4	Test lisse de Neyman	141
7.4.5	Tests basés sur la transformation de Smirnov généralisée	142
7.4.6	Comparaison des tests	144
7.5	Application à la loi de Weibull de type I	149
7.5.1	Tests basés sur la transformation de Smirnov généralisée	150
7.5.2	Comparaison des tests	151
7.6	Conclusion	153
8	Procédure d'analyse du retour d'expérience en temps discret	155
8.1	Procédure d'analyse	156
8.1.1	Notations	156
8.1.2	Estimation non paramétrique	157
8.1.3	Test graphique d'adéquation à la famille des lois géométriques	158
8.1.4	Test statistique d'adéquation à la famille des lois géométriques	158
8.1.5	Estimation des fonctions usuelles de la fiabilité pour la loi géométrique	159
8.1.6	Test graphique d'adéquation à la famille des lois de Weibull de type I	160
8.1.7	Test statistique d'adéquation à la famille des lois de Weibull de type I	161
8.1.8	Estimation des fonctions usuelles de la fiabilité pour la loi de Weibull de type I	161
8.1.9	Autres modèles de fiabilité	162
8.1.10	Récapitulatif	162
8.2	Application pratique : données non censurées	163
8.3	Application pratique : données censurées	167
8.3.1	Estimation non paramétrique	167
8.3.2	Tests d'adéquation graphiques	172
8.3.3	Estimation paramétrique	173
8.3.4	Conclusion	174
9	Conclusion et perspectives	175
	Bibliographie	178

Notations :

I_{cc}	:	Courant de court-circuit.
CEI	:	Commission Electrotechnique Internationale.
$ANSI$:	American National Standard Institute.
$\mathbb{R}^k, \mathbb{N}^k$:	Ensemble des k-uplets réels, d'entiers.
$\mathbb{P}, \mathbb{E}, \text{Var}, \text{Cov}$:	Probabilité, espérance, variance et covariance.
$\Gamma(\cdot)$:	Fonction Gamma.
ϕ, z_α	:	Fonction de répartition, quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite.
$\mathcal{F}_{(\alpha; \nu_1, \nu_2)}$:	Quantile d'ordre α de la loi de Fischer-Snedecor à (ν_1, ν_2) degrés de liberté.
v.a.	:	Variable aléatoire.
v.a.i.i.d.	:	Variations aléatoires indépendantes, identiquement distribuées.
b.o.n.	:	Base orthonormée.
$[x]$:	Partie entière d'un réel x .
$\mathbb{I}_E(\cdot)$:	Fonction indicatrice d'un ensemble E .
$\mathcal{M}_K(\mathbb{R})$:	Ensemble des matrices carrées de taille K à coefficients réels.
I_n	:	Matrice identité de $\mathcal{M}_n(\mathbb{R})$.
A^t, A^{-1} , pour $A \in \mathcal{M}_K(\mathbb{R})$:	Matrice transposée de A , inverse de A .
$\xrightarrow{\mathbb{P}}$, $\xrightarrow{p.s.}$, $\xrightarrow{\mathcal{L}}$:	Convergence en probabilité, presque sûre, en loi.

$$(a)_{(k)} = \begin{cases} \prod_{i=1}^k (a + i - 1) & \text{si } k \geq 1, \\ 1 & \text{si } k = 0. \end{cases} \quad : \text{ Notation de Pochhammer.}$$

$${}_1F_1(a, c; z) = \sum_{k=0}^{+\infty} \frac{(a)_{(k)}}{(c)_{(k)} k!} z^k \quad : \text{ Série de Kummer.}$$

$${}_2F_1(a, b, c, z) = \sum_{k=0}^{+\infty} \frac{(a)_{(k)} (b)_{(k)}}{(c)_{(k)} k!} z^k \quad : \text{ Série hypergéométrique de Gauss.}$$

$$\sum_{i=1}^0 a_i = 0, \quad \prod_{i=1}^0 a_i = 1 \quad : \text{ Convention, pour toute suite } \{a_i\}.$$

1.1 Contexte de l'étude

Pour répondre aux enjeux actuels de compétitivité et de sûreté de fonctionnement dans le domaine de la distribution électrique et du contrôle commande, Schneider Electric est engagé depuis longtemps dans le développement des méthodes et outils de la sûreté de fonctionnement. L'utilisation de la plupart des appareils fabriqués par Schneider Electric n'est pas continue car ceux-ci fonctionnent seulement lorsqu'ils sont sollicités. Un travail de recherche sur la durée de vie des systèmes en temps discret a alors été initié sous la forme d'une thèse. Ce travail de recherche permettra à Schneider Electric de :

- favoriser la maîtrise de la sûreté des produits,
- contrôler les coûts associés,
- aider au diagnostic pour la maintenance préventive.

1.2 Présentation des appareils électromécaniques

Le travail de modélisation de la thèse s'applique sur les appareils électromécaniques fabriqués par Schneider Electric, mais ce travail aurait pu également porter sur d'autres types d'appareils tels que des automates.

Avec ses quatre marques internationales - Merlin Gerin, Modicon, Square D et Telemecanique - Schneider Electric fabrique les appareils électromécaniques suivants :

- le sectionneur
- l'interrupteur
- le disjoncteur
- le fusible

Nous nous intéresserons plus particulièrement au disjoncteur qui est un produit central dans l'offre du catalogue Merlin Gerin et pour lequel le Retour d'Expérience est le plus conséquent. De plus, le disjoncteur est l'appareil qui a l'impact le plus important pour la sécurité et la disponibilité du réseau qu'il protège.

Le disjoncteur est un appareil qui assure la commande et la protection d'un réseau. Le disjoncteur est capable d'établir, de supporter et d'interrompre les courants de service ainsi que les courants de court-circuit.

1.2.1 Les différentes fonctions du disjoncteur

a) Supporter les contraintes du réseau

Cette fonction est essentielle car le disjoncteur est toujours sous tension. Pendant toute sa durée de vie, l'appareil (fermé ou ouvert) doit supporter une tension entre phases et entre phase et masse.

Le niveau d'isolement fixe la tenue diélectrique des matériels aux surtensions d'origine interne et aux ondes de choc de foudre.

Les surtensions d'origine interne accompagnent toute modification intervenant dans un circuit : ouverture ou fermeture d'un circuit, claquage ou contournement d'un isolant,...

Elles sont simulées en laboratoire par la tenue à la fréquence industrielle pendant une minute.

Les surtensions d'origine externe ou atmosphérique se produisent lorsque la foudre tombe sur la ligne ou à sa proximité.

Les ondes de tension qui en résultent sont simulées en laboratoire et sont appelées ondes de choc.

b) Établir un courant

Le disjoncteur peut :

- établir le courant dans une charge (moteur, transformateur, condensateur,...),
- fermer sur courant de court-circuit,
- fermer une ligne à vide.

c) Conduire le courant

Le circuit principal doit supporter sans dommage :

- le courant thermique = courant de court-circuit pendant 1 ou 3 secondes,
- le courant électrodynamique = $2,5I_{cc}$ (normes CEI) ou $2,7I_{cc}$ (normes ANSI)
- le courant de charge permanent.

Un disjoncteur étant la plupart du temps en position "fermé", le courant de charge doit circuler sans emballement thermique pendant toute la durée de vie de l'appareil.

Les échauffements dépendent de l'échange thermique avec l'air ambiant. Les normes définissent les limites maximales d'échauffement.

d) Couper le courant

Le disjoncteur doit couper les courants de charge et de court-circuit.

1.2.2 Les différents types de défaillance du disjoncteur

Le tableau de données (fictives ici pour des raisons évidentes de confidentialité) ci-dessous provient d'une fiche de prélèvement d'un produit de Schneider Electric. Le tableau contient, période par période, pour chaque type d'appareil, le nombre de sollicitations effectuées, soit jusqu'à la défaillance de l'appareil, soit jusqu'à un seuil maximal de sollicitations (fixé fictivement à 15). Lorsqu'il y a défaillance de l'appareil, on précise la cause de la défaillance.

Période	Résultat	Remarques
<i>P1</i>	14	rupture d'axe Ø7
<i>P2</i>	15	OK
<i>P3</i>	15	OK
<i>P4</i>	12	rupture de bielle
<i>P5</i>	13	réarmement incomplet
<i>P6</i>	15	OK
<i>P7</i>	9	casse ressort cliquet ouverture
<i>P8</i>	15	OK
<i>P9</i>	13	casse manivelle
<i>P10</i>	12	afficheur HS sur déclencheur
<i>P11</i>	10	réarmement incomplet
<i>P12</i>	15	OK

On peut noter que la durée de vie des appareils est le nombre de sollicitations jusqu'à l'apparition de la première défaillance. Cet exemple montre clairement que les appareils considérés sont en fait des systèmes, relativement complexes, dont l'origine des défaillances est multiple. Dans un souci de simplification, nous ne distinguerons pas les causes et modes de défaillances, mais nous considérerons que l'appareil est défaillant lorsqu'il ne peut assurer une des fonctions décrites à la section 1.2.1.

1.3 Les essais chez Schneider Electric

Dans le processus de création de produits nouveaux, trois phases apparaissent : le développement, le lancement (industriel et commercial) et l'exploitation. Chacune de ces phases possède un plan de qualification pour passer à l'étape suivante. Tous les essais décrits ci-dessous consistent à solliciter les appareils jusqu'à l'apparition de la première défaillance.

1.3.1 Les essais en phase de développement

Lors de la création d'un nouveau produit, les concepteurs doivent qualifier successivement les sous-ensembles, puis l'ensemble du prototype au regard des normes (CEI 60947-1, 60947-2 et 60947-3 selon le type d'appareil électromécanique développé).

1.3.2 Les essais en phase de lancement

Les caractéristiques concernant les disjoncteurs sont définies par deux documents de référence :

- La CEI (Commission Electrotechnique Internationale) publication du Technical Committee 56,
- l'ANSI (American National Standard Institute) C37-06.

Chaque contrat avec un client fait généralement référence à un de ces deux documents. Les recommandations CEI et l'ANSI C37-06 définissent les caractéristiques, assignées obligatoires et particulières d'un disjoncteur.

Parmi les caractéristiques assignées obligatoires figure une séquence de manoeuvres assignée (voir paragraphe 4.104 CEI 56). Un essai (dit de type) de fonctionnement mécanique à la température de l'air ambiant permet de vérifier cette caractéristique obligatoire pour le disjoncteur. La norme donne le choix parmi trois séquences de manoeuvres (lent, rapide 1, et rapide 2) qui consistent en des cycles de fermeture-ouverture avec des temps inter-manoeuvres imposés sous un courant de court-circuit.

Des essais, dits de routine, ont également pour but de vérifier les caractéristiques du matériel, ses dispositifs de commande et ses équipements auxiliaires. Ces essais sont réalisés par Schneider Electric et font l'objet d'un rapport d'essai paraphé par le département contrôle Qualité de l'usine. Par exemple, l'endurance mécanique demandée par la norme CEI 56 est de 2000 manoeuvres pour le domaine de la Moyenne Tension, dans des conditions normales de fonctionnement.

1.3.3 Les essais en phase d'exploitation

Après la phase de lancement d'un nouveau produit, il est nécessaire de contrôler la qualité de fabrication du produit afin de vérifier les objectifs de fiabilité. Le service Qualité de l'usine de production prévoit donc de prélever un certain nombre d'appareils avec une certaine fréquence.

Il est important de noter que toutes les normes précédemment citées ne prennent pas en compte le vieillissement à la sollicitation des appareils. C'est à dire que, sachant que l'appareil a correctement fonctionné jusqu'à la $k^{\text{ème}}$ sollicitation, la probabilité de défaillance

à la $k + 1^{\text{ème}}$ sollicitation est supposée être la même quel que soit k . Cette hypothèse est considérée comme peu réaliste par de nombreux spécialistes (concepteurs, experts, etc) de Schneider Electric. On peut supposer que les normes actuelles ne prennent pas en compte le vieillissement à la sollicitation en raison de la simplicité des calculs sous cette hypothèse, et surtout par un manque évident de travaux de recherches dans le domaine de la fiabilité en temps discret. À partir de ce constat et la volonté de Schneider Electric d'approfondir la connaissance de ses appareils fonctionnant à la sollicitation, un travail de recherche a été initié.

1.4 Analyse du retour d'expérience

La procédure actuellement suivie par Schneider Electric dans l'étude des données issues des essais d'endurance mécanique provient des normes CEI 60605-4, 60605-6 et 61649. La démarche suivante décrit brièvement la procédure d'analyse du Retour d'Expérience :

1. Tester si le taux de défaillance est constant, c'est à dire faire un test d'adéquation à la loi exponentielle (voir CEI 60605-6).
2. Si la loi exponentielle n'est pas rejetée, on donne une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse exponentielle (voir CEI 60605-4).
3. Si la loi exponentielle est rejetée, on fait un test d'adéquation à la loi de Weibull. Si la loi de Weibull n'est pas rejetée, on donne une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse de Weibull (voir CEI 61649).
4. Si la loi de Weibull est rejetée, une étude spécifique est nécessaire.

1.5 Objectifs de la thèse

Dans la plupart des études de fiabilité, on considère que le temps est continu et on utilise les modèles usuels de fiabilité en temps continu, comme la loi exponentielle ou la loi de Weibull. Or dans certaines situations, le temps calendaire n'est pas la meilleure façon de décrire la durée de vie d'un système. C'est le cas des appareils électromécaniques décrits ci-dessus pour lesquels le bon fonctionnement s'exprime en termes de nombre de sollicitations effectuées jusqu'à défaillance. On peut également citer le cas où les données observées s'expriment en nombre de cycles, nombre de milliers de sollicitations jusqu'à défaillance, ou nombre de mois de bon fonctionnement. La caractéristique commune de toutes ces situations est que le temps est discret ou assimilable à un temps discret.

Le premier objectif de la thèse concerne la modélisation de durées de vie en temps discret. La modélisation supposera les appareils non réparables, hypothèse justifiée par la nature des

données disponibles. En effet, la collecte des défaillances des appareils en service chez les clients étant faible et souvent imprécise, surtout après l'expiration de la période de garantie, les seules données disponibles sont celles des essais d'endurance mécanique effectués en interne chez Schneider Electric (voir la section 1.3). Dans un souci de réalisme en vue d'applications industrielles, la modélisation considèrera des lois de durées de vie qui permettront de mettre en évidence un vieillissement à la sollicitation.

Le second objectif de la thèse sera de donner une procédure d'analyse en temps discret du Retour d'Expérience calquée sur celle utilisée actuellement en temps continu (écrite à la section 1.4) :

1. Tester si le taux de défaillance est constant, c'est à dire faire un test d'adéquation à la loi géométrique.
2. Si la loi géométrique n'est pas rejetée, donner une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse géométrique.
3. Si la loi géométrique est rejetée, existe-t'il une loi de Weibull en temps discret? Faire alors un test d'adéquation à la loi de Weibull discrète. Si la loi de Weibull discrète n'est pas rejetée, donner une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse de Weibull discrète.
4. Si la loi de Weibull discrète est rejetée, quels sont les autres modèles de fiabilité en temps discret disponibles?

Les normes ne prenant pas encore en compte le vieillissement à la sollicitation, les différents calculs de fiabilité effectués dans le cadre de cette thèse ne peuvent pas être utilisés pour la justification aux organismes d'homologation.

1.6 Plan de la thèse

Le plan du présent manuscrit est le suivant :

- Le deuxième chapitre commence par donner la définition des grandeurs de fiabilité en temps discret. Nous mettons en avant les différences avec les définitions du temps continu, notamment concernant le taux de défaillance. Nous nous intéressons ensuite aux notions de vieillissement en temps discret. Nous donnons des résultats concernant les taux de défaillance, croissant et croissant en moyenne, et les systèmes meilleurs nouveaux qu'utilisés, qui viennent compléter les travaux de Shaked, Shanthikumar et Valdez-Torres [84] sur le sujet.
- Le troisième chapitre est un état de l'art des modèles de fiabilité en temps discret. Notre présentation regroupe les modèles en trois familles : les lois discrètes usuelles, les

modèles présentant une analogie avec un modèle de fiabilité en temps continu, et enfin les modèles issus du schéma d'urne de Pólya.

- Dans le chapitre 4, nous mettons en évidence un certain nombre de différences importantes entre les notions de fiabilité usuelles en temps continu et leurs équivalents en temps discret. Nous montrons que ces différences sont dues à la définition du taux de défaillance en temps discret. Nous proposons une nouvelle définition qui résout les problèmes rencontrés. Avec cette nouvelle définition, les notions de fiabilité en temps discret sont nettement plus proches des notions correspondantes en temps continu qu'avec la définition usuelle.
- Avec le chapitre 5, nous abordons la partie d'inférence statistique. Ce chapitre est consacré à l'estimation non paramétrique des grandeurs de la fiabilité. On rappelle la définition de l'estimateur empirique de la fiabilité et ses propriétés. On s'intéresse plus particulièrement au taux de défaillance empirique dont on donne la définition ainsi que ses propriétés en tant qu'estimateur (à taille d'échantillon fixée et asymptotiques). La définition de l'estimateur empirique du nouveau taux de défaillance est également introduite et une étude asymptotique de ses propriétés est menée. Concernant la durée de vie résiduelle moyenne, seule l'expression de l'estimateur empirique est donnée, la complexité de son expression ne permettant pas d'étudier ses propriétés.
- Le chapitre 6 porte sur l'étude statistique d'une partie des modèles de fiabilité présentés au chapitre 3. Pour la loi géométrique, nous donnons les estimateurs ponctuels du paramètre p et de la fiabilité, ainsi que trois méthodes de construction d'intervalles de confiance pour p . Nous présentons une comparaison de l'erreur commise sur l'estimation de la fiabilité lors du traitement de données discrètes à l'aide de la loi géométrique ou de la loi exponentielle. Les autres lois discrètes ne possédant pas d'estimateurs explicites de leurs paramètres, nous présentons les propriétés d'estimation obtenues à l'aide de simulations de Monte-Carlo.
- Les tests d'adéquation aux lois discrètes sont traités au chapitre 7. Nous commençons par présenter les tests d'adéquation graphiques, puis nous passons en revue les principaux tests statistiques issus de notre étude bibliographique. Nous utilisons la transformation de Smirnov généralisée pour construire un test d'adéquation aux lois discrètes définies à l'aide de la partie entière de v.a. de lois continues. Une étude de puissance des tests d'adéquation aux lois géométrique et Weibull de type I clôt le chapitre.
- Le chapitre 8 propose une procédure d'analyse du Retour d'Expérience en temps discret calquée sur celle du temps continu utilisée chez Schneider Electric. À titre applicatif, nous traitons complètement un jeu de données issues du Retour d'Expérience. Enfin, nous abordons le cas des données censurées pour lesquelles une étude approfondie reste

à faire.

Les grandeurs de la fiabilité des systèmes non réparables en temps discret

Ce chapitre présente le cadre général de l'étude. Les grandeurs de base de la fiabilité des systèmes non réparables en temps discret sont définies : fiabilité, taux de défaillance, taux de défaillance cumulé, durée de vie résiduelle moyenne, MTTF. Après avoir présenté quelques propriétés de ces grandeurs, on s'intéresse aux notions de vieillissement en temps discret : taux de défaillance croissant (IFR) et croissant en moyenne (IFRA), système meilleur nouveau qu'utilisé (NBU). On constate que, contrairement à ce qui se passe en temps continu, ces notions peuvent avoir plusieurs définitions qui en sont pas équivalentes. Enfin, on montre la complétude de la classe des lois IFRA en temps discret.

2.1 Rappels des notions de fiabilité en temps continu

On donne à titre de rappel, les définitions en temps continu des principales grandeurs de la fiabilité. Soit T une variable aléatoire continue à valeurs dans \mathbb{R}^+ , représentant la durée de vie en temps continu d'un système, de sa mise en service jusqu'à sa défaillance.

1. La fonction de répartition de T donne la probabilité que le système défaille avant l'instant t :

$$\forall t \geq 0, F(t) = \mathbb{P}(T \leq t)$$

2. La densité de T est définie, si elle existe, comme la dérivée de $F(t)$ par rapport à t :

$$\forall t \geq 0, f(t) = dF(t)/dt$$

3. La fiabilité donne la probabilité que le système soit toujours en fonctionnement à l'instant t :

$$\forall t \geq 0, R(t) = 1 - F(t)$$

4. La durée de vie résiduelle à l'instant t est le temps qu'il reste à vivre au système sachant qu'il a fonctionné jusqu'à cet instant. La fiabilité résiduelle à l'instant t est :

$$\forall s \geq 0, R(s|t) = \mathbb{P}(T > t + s | T > t) = \frac{R(t+s)}{R(t)}$$

5. La durée de vie résiduelle moyenne est la durée moyenne qu'il reste à vivre au système sachant qu'il a survécu jusqu'à l'instant t :

$$\forall t \geq 0, m(t) = \mathbb{E}(T - t | T > t) = \int_0^{\infty} R(x|t) dx$$

6. Le temps moyen jusqu'à défaillance (Mean Time To Failure) est (sous réserve que l'intégrale converge) :

$$MTTF = \mathbb{E}(T) = \int_0^{\infty} x f(x) dx$$

7. Le taux de défaillance est défini par :

$$\forall t \geq 0, \lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{f(t)}{R(t)}$$

8. Le taux de défaillance cumulé est défini par :

$$\forall t \geq 0, \Lambda(t) = \int_0^t \lambda(x) dx = -\ln R(t)$$

2.2 Définition des grandeurs caractéristiques de la fiabilité en temps discret

On considère que la durée de vie d'un système en temps discret est le nombre K de sollicitations nécessaires pour que le système défaille, ou, de manière équivalente, K est le rang de la première sollicitation pour laquelle le système ne fonctionne pas.

Remarques :

On notera que Salvia et Bollinger [80] et d'autres auteurs utilisent une variable aléatoire discrète à valeurs dans \mathbb{N} . Compte tenu de la définition ci-dessus, on considèrera que K est à valeurs dans \mathbb{N}^* . Cela nécessitera de modifier les résultats obtenus pour les variables à valeurs

dans \mathbb{N} . De plus, contrairement à de nombreuses références anglo-saxonnes, nous n'utilisons pas la fonction de survie, mais plutôt la fonction de fiabilité, notamment pour définir le taux de défaillance en temps discret.

On donne ci-dessous les définitions des principales grandeurs de la fiabilité en temps discret, analogues à celles présentées pour le temps continu en 2.1.

Définition 1

1. La probabilité que le système défaille à la $k^{\text{ème}}$ sollicitation est :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \mathbb{P}(K = k)$$

2. La fonction de répartition, qui traduit la probabilité que le système défaille entre la première et la $k^{\text{ème}}$ sollicitation est :

$$\forall k \in \mathbb{N}^*, \quad F(k) = \mathbb{P}(K \leq k) = \sum_{i=1}^k p(i)$$

3. La fiabilité, qui traduit la probabilité que le système fonctionne encore à la $k^{\text{ème}}$ sollicitation est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = \mathbb{P}(K > k) = 1 - F(k) = 1 - \sum_{i=1}^k p(i)$$

4. La durée de vie résiduelle à l'instant k est le temps qu'il reste à vivre au système sachant qu'il fonctionne encore à l'instant k . La fiabilité résiduelle à l'instant k est :

$$\forall i \in \mathbb{N}^*, \quad R(i|k) = \mathbb{P}(K > k + i | K > k) = \frac{R(k + i)}{R(k)}$$

5. La durée de vie résiduelle moyenne est la durée moyenne qu'il reste à vivre au système sachant qu'il a survécu à k sollicitations :

$$\forall k \in \mathbb{N}^*, \quad m(k) = \mathbb{E}(K - k | K > k)$$

6. Le temps moyen jusqu'à défaillance (Mean Time To Failure) est (sous réserve que la série converge) :

$$MTTF = \mathbb{E}(K) = m(0) = \sum_{i=1}^{\infty} ip(i)$$

7. Le taux de défaillance est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = \mathbb{P}(K = k | K \geq k) = \frac{\mathbb{P}(K = k)}{\mathbb{P}(K \geq k)} = \frac{p(k)}{R(k-1)}$$

Le taux de défaillance donne la probabilité conditionnelle de défaillance du système à l'instant k , sachant qu'il a fonctionné jusqu'à l'instant $k - 1$.

8. Le taux de défaillance cumulé est :

$$\forall k \in \mathbb{N}^*, \quad \Lambda(k) = \sum_{i=1}^k \lambda(i).$$

La première mention d'un taux de défaillance (ou taux de hasard) en temps discret apparaît dans l'article de Barlow-Marshall-Proschan [8]. La définition proposée est $\lambda(k) = \frac{p(k)}{\sum_{i=k}^{+\infty} p(i)}$, $k \geq 0$, ce qui revient, mais sans l'écrire, à $\lambda(k) = \mathbb{P}(K = k | K \geq k)$. Les auteurs notent que dans le cas discret, $\lambda(k) \leq 1$. La notion de taux de défaillance en temps discret apparaît ensuite dans les travaux de Barlow et Proschan [9], Cox [20], Kaufmann-Grouchko-Cruon [46], Kalbfleisch-Prentice [44], Landberg-Leon-Lynch-Proschan [49] et Lawless [50] mais n'est pas plus développée. La notion de fiabilité en temps discret est évoquée dans les livres de Kalbfleisch-Prentice [44] et Lawless [50]. La mise en place plus précise des concepts de base est due à Salvia-Bollinger [80].

2.3 Quelques résultats

2.3.1 Sur le taux de défaillance

Il est important de noter que, contrairement au cas continu, le taux de défaillance en temps discret est une probabilité conditionnelle, donc est compris entre 0 et 1. Comme dans le cas continu, toutes les grandeurs de la fiabilité peuvent s'exprimer à l'aide du taux de défaillance.

$$\lambda(k) = \mathbb{P}(K = k | K \geq k) = \frac{p(k)}{R(k-1)} = \frac{R(k-1) - R(k)}{R(k-1)} = 1 - \frac{R(k)}{R(k-1)} \quad (2.1)$$

On obtient alors la formule de récurrence :

$$R(k) = R(k-1)(1 - \lambda(k))$$

On obtient ainsi :

$$\forall k \in \mathbb{N}^* \quad R(k) = \prod_{i=1}^k (1 - \lambda(i)) \quad (2.2)$$

$$\text{puis, } \forall k \in \mathbb{N}^* \quad p(k) = \lambda(k) \prod_{i=1}^{k-1} (1 - \lambda(i)). \quad (2.3)$$

Comme il est possible de définir la probabilité élémentaire de la durée de vie K à partir du taux de défaillance, il est important de trouver des conditions nécessaires et suffisantes pour qu'une fonction λ soit un taux de défaillance. Une condition incomplète a été donnée par Salvia-Bollinger dans [80]. Dans [84], Shaked, Shanthikumar et Valdes-Torres ont corrigé celle-ci pour obtenir finalement la condition nécessaire et suffisante suivante :

Théorème 1

Une suite $\{\lambda(k)\}_{k \geq 1}$ est le taux de défaillance d'une variable aléatoire à valeurs dans \mathbb{N}^* , si et seulement si,

(a) $\exists m \in \mathbb{N}^*, \forall i < m, \lambda(i) < 1$ et $\lambda(m) = 1$.

ou

(b) $\forall k \in \mathbb{N}^*, \lambda(k) \in [0, 1[$ et $\sum_{k=1}^{\infty} \lambda(k) = +\infty$

Le support de la loi correspondante est $\{1, \dots, m\}$ dans le cas (a) et \mathbb{N}^* dans le cas (b).

La démonstration de ce théorème se base sur le fait que $\forall m \in \mathbb{N}^*$, et quelle que soit la suite $\{\lambda(k)\}_{k \geq 1}$ à valeurs dans $[0, 1]$, on a :

$$\sum_{k=1}^{m-1} p(k) = \sum_{k=1}^{m-1} \lambda(k) \prod_{i=1}^{k-1} (1 - \lambda(i)) = 1 - \prod_{i=1}^{m-1} (1 - \lambda(i))$$

et d'après Knopp [47] (pp 219-221) :

$$\prod_{i=1}^{+\infty} (1 - \lambda(i)) = 0 \Leftrightarrow \begin{cases} \exists k \in \mathbb{N}^* / \lambda(k) = 1 \\ \text{ou} \\ \forall k \in \mathbb{N}^*, \lambda(k) < 1 \text{ et } \sum_{k=1}^{+\infty} \lambda(k) = +\infty \end{cases}$$

2.3.2 Sur le MTTF

On peut exprimer le MTTF en fonction de la fiabilité :

$$\begin{aligned} MTTF &= \mathbb{E}(K) = \sum_{k=1}^{\infty} k \mathbb{P}(K = k) \\ &= \sum_{k=1}^{\infty} k (R(k-1) - R(k)) = \sum_{k=0}^{\infty} (k+1)R(k) - \sum_{k=1}^{\infty} kR(k) \\ \text{D'où } MTTF &= \sum_{k=0}^{+\infty} R(k) \end{aligned} \tag{2.4}$$

En utilisant la relation (2.2) dans (2.4), on obtient l'expression du MTTF en fonction du taux de défaillance :

$$MTTF = \sum_{k=0}^{+\infty} \prod_{j=1}^k (1 - \lambda(j)) = 1 + \sum_{k=1}^{+\infty} \prod_{j=1}^k (1 - \lambda(j)) \tag{2.5}$$

Les conditions d'existence du MTTF sont données par Padgett et Spurrer [66] dans le théorème suivant :

Théorème 2

Le MTTF existe si et seulement si il existe $\varepsilon > 0$ et $m \in \mathbb{N}$ tels que pour tout $k > m$, $\frac{k\lambda(k)}{1 - \lambda(k)} > 1 + \varepsilon$.

D'après Salvia [79], on a les résultats suivants :

1. Si le taux de défaillance est croissant, on a $MTTF \leq \frac{1}{\lambda(1)}$
2. Si le taux de défaillance est décroissant, on a $MTTF \geq \frac{1}{\lambda(1)}$
3. Si le taux de défaillance est constant, on a $MTTF = \frac{1}{\lambda}$

2.3.3 Sur la durée de vie résiduelle moyenne

On peut exprimer la durée de vie résiduelle moyenne en fonction de la fiabilité :

$$\begin{aligned}
 m(k) &= \mathbb{E}(K - k | K > k) = \sum_{i=1}^{\infty} i \mathbb{P}(K - k = i | K > k) \\
 &= \sum_{i=1}^{\infty} i \frac{\mathbb{P}(K - k = i \cap K > k)}{\mathbb{P}(K > k)} = \sum_{i=1}^{\infty} i \frac{\mathbb{P}(K = k + i)}{\mathbb{P}(K > k)} \\
 &= \sum_{i=1}^{\infty} i \frac{R(k + i - 1) - R(k + i)}{R(k)} = \sum_{i=1}^{\infty} i \frac{R(k + i - 1)}{R(k)} - \sum_{i=1}^{\infty} i \frac{R(k + i)}{R(k)} \\
 &= \sum_{j=0}^{\infty} (j + 1) \frac{R(k + j)}{R(k)} - \sum_{j=0}^{\infty} j \frac{R(k + j)}{R(k)} = \sum_{j=0}^{+\infty} R(k + j) \\
 \text{D'où } m(k) &= \frac{1}{R(k)} \sum_{j=0}^{+\infty} R(k + j) \tag{2.6}
 \end{aligned}$$

À l'aide de (2.2) et (2.6), on peut aussi exprimer la durée de vie résiduelle moyenne en fonction du taux de défaillance :

$$\begin{aligned}
 m(k) &= \frac{1}{R(k)} \sum_{j=0}^{+\infty} R(k + j) = 1 + \sum_{j=1}^{+\infty} \frac{\prod_{i=1}^{k+j} (1 - \lambda(i))}{\prod_{i=1}^k (1 - \lambda(i))} \\
 &= 1 + \sum_{j=1}^{+\infty} \prod_{i=k+1}^{k+j} (1 - \lambda(i)) = 1 + \sum_{j=k+1}^{+\infty} \prod_{i=k+1}^j (1 - \lambda(i)) \\
 \text{alors } m(k) - 1 &= 1 - \lambda(k + 1) + \sum_{j=k+2}^{+\infty} \prod_{i=k+1}^j (1 - \lambda(i))
 \end{aligned}$$

$$\text{donc } \frac{m(k) - 1}{1 - \lambda(k+1)} = 1 + \sum_{j=k+2}^{+\infty} \prod_{i=k+2}^j (1 - \lambda(i)) = m(k+1)$$

d'où le résultat :

$$\forall k \in \mathbb{N}^*, \lambda(k) = 1 - \frac{m(k-1) - 1}{m(k)} \text{ et, si } \lambda(k) \neq 1, m(k) = \frac{m(k-1) - 1}{1 - \lambda(k)} \quad (2.7)$$

On en déduit que :

Théorème 3

Si $0 < \lim_{k \rightarrow +\infty} \lambda(k) < 1$ alors, $\lim_{k \rightarrow +\infty} m(k) = \frac{1}{\lim_{k \rightarrow +\infty} \lambda(k)}$.

2.4 Correspondances entre les différentes grandeurs de la fiabilité en temps continu et en temps discret

Les tableaux 2.1 et 2.2 de la page suivante récapitulent les correspondances entre les différentes grandeurs de la fiabilité en temps continu et en temps discret.

Des similitudes claires apparaissent entre les deux cas, mais également des différences sensibles, notamment pour le taux de défaillance.

2.5 Notions de vieillissement

Le vieillissement d'un système peut être caractérisé de différentes façons, que la durée de vie du système soit discrète ou continue. Dans cette partie, nous étudions les différentes notions de vieillissement en temps discret sur la base des travaux de Shaked, Shanthikumar et Valdez-Torres [84], en les complétant.

2.5.1 Notions de vieillissement en temps continu

a) Taux de défaillance croissant (IFR)

La variable aléatoire T à valeurs dans \mathbb{R}^+ , est à taux de défaillance croissant (Increasing Failure Rate) si et seulement si, de manière équivalente :

- IFR1 : $\lambda(t)$ est une fonction croissante.
- IFR2 : $\forall s \geq 0, R(s|t)$ est une fonction décroissante en t .
- IFR3 : $\forall s_1, s_2, t_1, t_2 \in \mathbb{R}^+$ tels que $s_1 < s_2$ et $t_1 < t_2$, on a $\left| \begin{array}{cc} R(s_1 - t_1) & R(s_1 - t_2) \\ R(s_2 - t_1) & R(s_2 - t_2) \end{array} \right| \geq 0$
- IFR4 : $\ln R(t)$ est une fonction concave.

	$R(t)$	$f(t)$	$\lambda(t)$	$m(t)$
$R(t)$		$\int_t^\infty f(x)dx$	$\exp \left[- \int_0^t \lambda(x)dx \right]$	$\frac{m(0)}{m(t)} \cdot \exp \left\{ - \int_0^t \frac{1}{m(x)} dx \right\}$
$f(t)$	$-\frac{dR(t)}{dt}$		$\lambda(t) \exp \left[- \int_0^t \lambda(x)dx \right]$	$(1 + m'(t)) \frac{m(0)}{m(t)^2} \exp \left[- \int_0^t \frac{dx}{m(x)} \right]$
$\lambda(t)$	$-\frac{1}{R(t)} \frac{dR(t)}{dt}$	$\frac{f(t)}{\int_t^\infty f(x)dx}$		$\frac{1 + m'(t)}{m(t)}$
$m(t)$	$\frac{\int_t^\infty R(x)dx}{R(t)}$	$\frac{\int_t^\infty x f(x)dx}{\int_t^\infty f(x)dx} - t$	$\frac{\int_t^\infty \exp \left[- \int_0^x \lambda(u)du \right] dx}{\exp \left[- \int_0^t \lambda(x)dx \right]}$	

TAB. 2.1 – Grandeurs de la fiabilité en temps continu

	$R(k)$	$p(k)$	$\lambda(k)$	$m(k)$
$R(k)$		$\sum_{j=k+1}^{+\infty} p(j)$	$\prod_{j=1}^k (1 - \lambda(j))$	$\prod_{j=1}^k \frac{m(j-1) - 1}{m(j)}$
$p(k)$	$R(k-1) - R(k)$		$\lambda(k) \prod_{j=1}^{k-1} (1 - \lambda(j))$	$\left(1 - \frac{m(k-1) - 1}{m(k)}\right) \prod_{j=1}^{k-1} \frac{m(j-1) - 1}{m(j)}$
$\lambda(k)$	$\frac{R(k-1) - R(k)}{R(k-1)}$	$\frac{p(k)}{\sum_{j=k}^{+\infty} p(j)}$		$1 - \frac{m(k-1) - 1}{m(k)}$
$m(k)$	$\frac{1}{R(k)} \sum_{j=0}^{+\infty} R(j+k)$	$\frac{\sum_{j=k+1}^{\infty} jp(j)}{\sum_{j=k+1}^{\infty} p(j)} - k$	$1 + \sum_{j=k+1}^{+\infty} \prod_{i=k+1}^j (1 - \lambda(i))$	

TAB. 2.2 – Grandeurs de la fiabilité en temps discret

La notion IFR exprime le fait que le système étudié vieillit au cours de son utilisation. La notion IFR3 est également appelée PF_2 (fréquence de Pólya d'ordre 2). L'équivalence entre les quatre notions ci-dessus a été démontrée par Barlow et Proschan dans [9].

b) Taux de défaillance croissant en moyenne (IFRA)

La variable aléatoire T est à taux de défaillance croissant en moyenne (Increasing Failure Rate in Average) si et seulement si, de manière équivalente :

- IFRA1 : $[R(t)]^{1/t}$ est une fonction décroissante.
- IFRA2 : $\frac{\Lambda(t)}{t}$ est une fonction croissante.

La notion IFRA a été introduite en temps continu par Barlow et Proschan [10] afin de définir la plus petite classe de lois de durée de vie d'un système cohérent constitué de composants à taux de défaillance croissant. Par conséquent, si l'on considère un système constitué de composants indépendants ayant chacun un taux de défaillance croissant, alors le système est à taux de défaillance croissant en moyenne. On notera également que tout système constitué de composants IFRA est lui-même IFRA. Cette propriété que nous appellerons dans la suite, la complétude de la classe IFRA, n'est pas valable pour la notion IFR.

c) Système meilleur nouveau qu'utilisé (NBU)

La variable aléatoire T appartient à la famille NBU (New Better Than Used) si et seulement si, de manière équivalente :

- NBU1 : $R(s|t) \leq R(s) \quad \forall s \geq 0, \forall t \geq 0$.
- NBU2 : $\Lambda(t+s) \geq \Lambda(t) + \Lambda(s) \quad \forall t \geq 0, \forall s \geq 0$.

La notion NBU s'explique facilement par l'intermédiaire de la fiabilité résiduelle : la fiabilité d'un système ayant fonctionné jusqu'à un instant t est plus petite que celle d'un système neuf à cet instant t .

d) Durée de vie résiduelle moyenne décroissante (DMRL)

La variable aléatoire T appartient à la famille DMRL (Decreasing Mean Residual Life) si et seulement si :

- DMRL : $m(t)$ est une fonction décroissante.

En temps continu, on a (voir [10]) :

$$\begin{array}{ccc}
 IFR & \Rightarrow & IFRA \Rightarrow NBU \\
 \Downarrow & & \\
 DMRL & &
 \end{array}
 \tag{2.8}$$

De façon similaire, les mêmes notions sont valables pour le rajeunissement qui s'exprime par un taux de défaillance décroissant (DFR), un taux de défaillance décroissant en moyenne (DFRA), un système dit moins bon nouveau qu'utilisé (NWU) et une durée de vie résiduelle moyenne croissante (IMRL).

2.5.2 Notions de vieillissement en temps discret

a) Taux de défaillance croissant (IFR)

Nous donnons l'expression en temps discret des quatre notions de vieillissement décrites précédemment. Soit K une variable aléatoire discrète à valeurs dans \mathbb{N}^* .

Les définitions en temps discret correspondant aux définitions équivalentes à IFR en temps continu sont :

- IFR1: $\{\lambda(k)\}_{k \geq 1}$ est une suite croissante.
- IFR2: $\forall i \in \mathbb{N}^*, \{R(i|k)\}_{k \geq 1}$ est une suite décroissante en k .
- IFR3: $\forall j_1, j_2, k_1, k_2 \in \mathbb{N}^*$ tels que $j_1 < j_2$ et $k_1 < k_2$, on a $\begin{vmatrix} R(j_1 - k_1) & R(j_1 - k_2) \\ R(j_2 - k_1) & R(j_2 - k_2) \end{vmatrix} \geq 0$
- IFR4: $\{\ln R(k)\}_{k \geq 1}$ est une suite concave.

Comme dans le cas continu, la notion IFR3 est appelée PF_2 par Barlow-Marshall-Proshan [8], qui avancent, sans démonstration, la propriété suivante :

Propriété 1

Les notions IFR1, IFR2, IFR3 et IFR4 sont équivalentes.

Preuve :

Nous montrerons d'abord que $\text{IFR1} \Rightarrow \text{IFR2} \Rightarrow \text{IFR3} \Rightarrow \text{IFR1}$, puis l'équivalence $\text{IFR4} \Leftrightarrow \text{IFR1}$.

- $\text{IFR1} \Rightarrow \text{IFR2}$

$$\begin{aligned} R(i|k) &= \frac{R(k+i)}{R(k)} = \frac{R(k+i)}{R(k+i-1)} \cdot \frac{R(k+i-1)}{R(k+i-2)} \cdots \frac{R(k+1)}{R(k)} \\ &= \prod_{j=1}^i \frac{R(k+j)}{R(k+j-1)} = \prod_{j=1}^i (1 - \lambda(k+j)) \end{aligned}$$

Donc si $\{\lambda(k)\}_{k \geq 1}$ est une suite croissante, $\{R(i|k)\}_{k \geq 1}$ est une suite décroissante en k .

- $\text{IFR2} \Rightarrow \text{IFR3}$

La notion IFR2 s'écrit $\forall i, \forall k, l > k \Rightarrow R(i|k) \geq R(i|l)$

C'est à dire $\frac{R(k+i)}{R(k)} \geq \frac{R(l+i)}{R(l)}$ ou $R(k+i)R(l) \geq R(k)R(l+i)$.

Soient $(j_1, j_2, k_1, k_2) \in \mathbb{N}^{*4}$ tels que $j_1 < j_2$ et $k_1 < k_2$. On pose $i = j_2 - j_1, k = j_1 - k_2$

et $l = j_1 - k_1$. $k_1 < k_2$ implique $k < l$, on peut donc appliquer le résultat précédent :

$$R(j_1 - k_1)R(j_2 - k_2) = R(l)R(k + i) \geq R(k)R(l + i) = R(j_1 - k_2)R(j_2 - k_1)$$

Ce qui prouve la propriété IFR3.

• IFR3 \Rightarrow IFR1

Pour $k \in \mathbb{N}^*$, posons $k_1 = 1, k_2 = 2, j_1 = k + 1$ et $j_2 = k + 2$. On a bien $j_1 < j_2$ et $k_1 < k_2$.

$$\text{Donc } R(j_1 - k_1)R(j_2 - k_2) = R(k)R(k) \geq R(j_1 - k_2)R(j_2 - k_1) = R(k - 1)R(k + 1)$$

$$\text{D'où } \forall k \in \mathbb{N}^*, \frac{R(k)}{R(k - 1)} \geq \frac{R(k + 1)}{R(k)}$$

$$\text{ou } \forall k \in \mathbb{N}^*, 1 - \frac{R(k)}{R(k - 1)} \leq 1 - \frac{R(k + 1)}{R(k)}$$

$$\text{donc } \forall k \in \mathbb{N}^*, \lambda(k) \leq \lambda(k + 1).$$

Ce qui prouve que $\{\lambda(k)\}_{k \geq 1}$ est une suite croissante.

Pour montrer que IFR1 \Leftrightarrow IFR4, il faut rappeler que $\{\varphi(k)\}_{k \geq 1}$ est une suite concave si et seulement si :

$$\forall (k_1, k_2, k_3) \in \mathbb{N}^{*3}, k_1 < k_2 < k_3 \Rightarrow \frac{\varphi(k_2) - \varphi(k_1)}{k_2 - k_1} \geq \frac{\varphi(k_3) - \varphi(k_1)}{k_3 - k_1} \geq \frac{\varphi(k_3) - \varphi(k_2)}{k_3 - k_2}$$

En posant $k_1 = k, k_2 = k + i$ et $k_3 = k + i + j$, on montre facilement que cette définition est équivalente à :

$$\forall (k, i, j) \in \mathbb{N}^{*3}, \frac{1}{j} [\varphi(k + i + j) - \varphi(k + i)] \leq \frac{1}{i} [\varphi(k + i) - \varphi(k)]$$

Pour $\varphi = \ln R$, on obtient :

$$\forall (k, i, j) \in \mathbb{N}^{*3}, \frac{1}{j} \ln \frac{R(k + i + j)}{R(k + i)} \leq \frac{1}{i} \ln \frac{R(k + i)}{R(k)}$$

• IFR4 \Rightarrow IFR1

En posant $l = k + 1, i = j = 1$, la définition précédente devient :

$$\ln \frac{R(l + 1)}{R(l)} \leq \ln \frac{R(l)}{R(l - 1)}.$$

$$\text{D'où } \forall l \in \mathbb{N}^*, \lambda(l) = 1 - \frac{R(l)}{R(l - 1)} \leq 1 - \frac{R(l + 1)}{R(l)} = \lambda(l + 1).$$

Ce qui prouve que $\{\lambda(k)\}_{k \geq 1}$ est une suite croissante.

• IFR1 \Rightarrow IFR4

La concavité de $\{\ln R(k)\}_{k \geq 1}$ s'écrit aussi :

$$\forall (k, i, j) \in \mathbb{N}^{*3}, \frac{1}{j} \sum_{l=1}^j (1 - \lambda(k + i + l)) \leq \frac{1}{i} \sum_{l=1}^i (1 - \lambda(k + l)).$$

Si $\{\lambda(k)\}_{k \geq 1}$ est une suite croissante, on a :

- $\forall l \in \{1, \dots, i\}, \lambda(k + l) \leq \lambda(k + i)$
- $\forall l \in \{1, \dots, j\}, \lambda(k + i + l) \geq \lambda(k + i)$

Donc on en déduit :

- $\frac{1}{i} \sum_{l=1}^i (1 - \lambda(k + l)) \geq 1 - \lambda(k + i)$
- $\frac{1}{j} \sum_{l=1}^j (1 - \lambda(k + i + l)) \leq 1 - \lambda(k + i)$

Ce qui prouve que $\{\ln R(k)\}_{k \geq 1}$ est concave. \square

En temps continu, les notions de log-concavité et log-convexité de la fiabilité permettent très facilement de déterminer le sens de variation du taux de défaillance (voir Barlow-Proschan [10]). Gupta *et al.* [32] ont proposé une démarche analogue en temps discret. Leur résultat n'est pas exact pour les lois à support borné, aussi nous ne l'énonçons que pour les lois dont le support est \mathbb{N}^* .

Théorème 4

On suppose que $\forall k \in \mathbb{N}^*, p(k) \neq 0$. Alors :

- La loi est log-concave ssi la suite $\left\{ \frac{p(k+1)}{p(k)} \right\}_{k \geq 1}$ est décroissante. Alors le taux de défaillance est croissant.
- La loi est log-convexe ssi la suite $\left\{ \frac{p(k+1)}{p(k)} \right\}_{k \geq 1}$ est croissante. Alors le taux de défaillance est décroissant.
- Si la suite $\left\{ \frac{p(k+1)}{p(k)} \right\}_{k \geq 1}$ est constante, alors le taux de défaillance est constant, donc la loi est géométrique.

Ce théorème permettra de déterminer facilement le sens de variation de certains taux de défaillance dont l'expression est complexe.

b) Taux de défaillance croissant en moyenne (IFRA)

Les définitions en temps discret correspondant aux définitions équivalentes à IFRA en temps continu sont :

- IFRA1: $\left\{ R(k)^{\frac{1}{k}} \right\}_{k \geq 1}$ est une suite décroissante.
- IFRA2: $\left\{ \frac{\Lambda(k)}{k} \right\}_{k \geq 1}$ est une suite croissante.

Dans [84], Shaked *et al.* affirment, sans prouver l'implication, que :

Propriété 2

$IFRA1 \Rightarrow IFRA2$ et $IFRA2 \not\Rightarrow IFRA1$.

Preuve :

- IFRA1 \Rightarrow IFRA2

D'une part,

$$\begin{aligned} \left\{ \frac{\Lambda(k)}{k} \right\}_{k \geq 1} \text{ croissante} &\Leftrightarrow \forall k \in \mathbb{N}^*, \frac{1}{k} \sum_{j=1}^k \lambda(j) \leq \frac{1}{k+1} \sum_{j=1}^{k+1} \lambda(j) \\ &\Leftrightarrow \frac{1}{k} \sum_{j=1}^k \lambda(j) \leq \frac{1}{k+1} \sum_{j=1}^k \lambda(j) + \frac{1}{k+1} \lambda(k+1) \\ &\Leftrightarrow \left[\frac{1}{k} - \frac{1}{k+1} \right] \sum_{j=1}^k \lambda(j) \leq \frac{1}{k+1} \lambda(k+1) \\ &\Leftrightarrow \frac{1}{k(k+1)} \sum_{j=1}^k \lambda(j) \leq \frac{1}{k+1} \lambda(k+1) \end{aligned}$$

Donc $\left\{ \frac{\Lambda(k)}{k} \right\}_{k \geq 1}$ croissante $\Leftrightarrow \forall k \in \mathbb{N}^*, \frac{1}{k} \sum_{j=1}^k \lambda(j) \leq \lambda(k+1)$.

D'autre part, $\left\{ R(k)^{\frac{1}{k}} \right\}_{k \geq 1}$ décroissante

$$\begin{aligned} &\Leftrightarrow \forall k \in \mathbb{N}^*, R(k)^{\frac{1}{k}} \geq R(k+1)^{\frac{1}{k+1}} \\ &\Leftrightarrow \left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k}} \geq \left[\prod_{i=1}^{k+1} (1 - \lambda(i)) \right]^{\frac{1}{k+1}} = \left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k+1}} (1 - \lambda(k+1))^{\frac{1}{k+1}} \\ &\Leftrightarrow \left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k} - \frac{1}{k+1}} \geq (1 - \lambda(k+1))^{\frac{1}{k+1}} \end{aligned}$$

$$\begin{aligned} &\Leftrightarrow \left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k(k+1)}} \geq (1 - \lambda(k+1))^{\frac{1}{k+1}} \\ &\Leftrightarrow \left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k}} \geq 1 - \lambda(k+1) \\ &\Leftrightarrow \lambda(k+1) \geq 1 - \left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k}} \end{aligned}$$

On utilise l'inégalité des moyennes arithmétique et géométrique suivante :

$$\text{Si les } a_i \text{ sont des réels positifs, alors } \forall k \in \mathbb{N}^*, \quad \left[\prod_{i=1}^k a_i \right]^{\frac{1}{k}} \leq \frac{1}{k} \sum_{i=1}^k a_i$$

En posant $a_i = 1 - \lambda(i)$ dans cette inégalité, on obtient :

$$\left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k}} \leq \frac{1}{k} \sum_{i=1}^k (1 - \lambda(i))$$

La décroissance de $\left\{ R(k)^{\frac{1}{k}} \right\}_{k \geq 1}$ entraîne donc :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k+1) \geq 1 - \left[\prod_{i=1}^k (1 - \lambda(i)) \right]^{\frac{1}{k}} \geq 1 - \frac{1}{k} \sum_{i=1}^k (1 - \lambda(i)) = \frac{1}{k} \sum_{i=1}^k \lambda(i)$$

Ainsi $\left\{ \frac{1}{k} \Lambda(k) \right\}_{k \geq 1}$ est croissante.

• IFRA2 $\not\Rightarrow$ IFRA1

Pour démontrer que la réciproque est fautive, Shaked *et al.* proposent l'exemple de la v.a. K dont le taux de défaillance est défini par $\lambda(1) = 0.1$, $\lambda(2) = 0.9$ et $\lambda(k) = 0.6$ pour $k \geq 3$. Alors K est bien IFRA2 mais $R(2)^{1/2} < R(3)^{1/3}$ et donc la suite $\{R(k)^{1/k}\}_{k \geq 1}$ n'est pas décroissante. \square

Remarque :

En temps continu, $\Lambda(t) = -\ln R(t)$, ce qui prouve immédiatement l'équivalence des deux notions IFRA. Mais en temps discret, $\Lambda(k) \neq -\ln R(k)$, ce qui fait que les notions IFRA1 et IFRA2 ne sont pas équivalentes. Lawless [50] a noté ce fait et en a simplement conclu que le terme "taux de défaillance cumulé" n'est pas approprié en temps discret pour $H(k) = -\ln R(k)$.

L'implication du théorème suivant a été énoncée sans démonstration par Shaked *et al.* :

Théorème 5

$IFR \Rightarrow IFRA1$ et $IFRA1 \not\Rightarrow IFR$.

Preuve :

• IFR \Rightarrow IFRA1

Par analogie avec le temps continu, nous utilisons la notion de suite étoilée pour démontrer cette implication. En temps continu, une fonction $\varphi(x)$ définie sur \mathbb{R}^+ est étoilée si et seulement si $\forall x \in \mathbb{R}^+, \forall \alpha \in [0, 1], \varphi(\alpha x) \leq \alpha \varphi(x)$. Cela est équivalent à dire que $\frac{\varphi(x)}{x}$ est une fonction croissante. De plus, toute fonction convexe qui passe par l'origine est étoilée (Barlow-Proschan [10]).

Nous présentons l'équivalent de ces définitions en temps discret.

Définition 2

Une suite $\{\varphi(k)\}_{k \geq 1}$ est étoilée si et seulement si, de façon équivalente :

- $\forall (j, k) \in \mathbb{N}^{*2}, j \leq k \Rightarrow \varphi(j) \leq \frac{j}{k} \varphi(k)$
- $\left\{ \frac{\varphi(k)}{k} \right\}_{k \geq 1}$ est une suite croissante.

L'équivalence entre les deux définitions de la notion étoilée est immédiate si l'on réécrit la première définition sous la forme : $j \leq k \Rightarrow \frac{\varphi(j)}{j} \leq \frac{\varphi(k)}{k}$.

Lemme 1

Toute suite convexe passant par l'origine est étoilée.

Preuve :

On déduit d'un élément de la preuve de la propriété 1 que $\{\varphi(k)\}_{k \geq 1}$ est une suite convexe si et seulement si :

$$\forall (u, v, w) \in \mathbb{N}^{*3}, (u + v)\varphi(w + u) - v\varphi(w) - u\varphi(w + u + v) \leq 0$$

φ passant par l'origine, on a $\varphi(w) = 0$ pour $w = 0$.

$$\Rightarrow \forall (u, v) \in \mathbb{N}^{*2}, (u + v)\varphi(u) - u\varphi(u + v) \leq 0$$

Posons $j = u$ et $k = u + v$, on obtient :

$$\forall (j, k) \in \mathbb{N}^{*2}, j < k \Rightarrow k\varphi(j) - j\varphi(k) \leq 0, \text{ d'où } \varphi(j) \leq \frac{j}{k} \varphi(k)$$

Ce qui prouve que $\{\varphi(k)\}_{k \geq 1}$ est étoilée. \square

La démonstration de IFR \Rightarrow IFRA est alors simple puisque :

$$\begin{aligned} \text{K est IFR} &\Rightarrow \{\ln R(k)\}_{k \geq 1} \text{ est concave} \\ &\Rightarrow \{-\ln R(k)\}_{k \geq 1} \text{ est une suite convexe passant par l'origine} \\ &\Rightarrow \{-\ln R(k)\}_{k \geq 1} \text{ est étoilée} \\ &\Rightarrow \left\{ -\frac{\ln R(k)}{k} \right\}_{k \geq 1} \text{ est croissante} \\ &\Rightarrow \left\{ R(k)^{\frac{1}{k}} \right\}_{k \geq 1} \text{ est décroissante} \\ &\Rightarrow \text{K est IFRA1} \end{aligned}$$

• IFRA1 $\not\Rightarrow$ IFR

Soit la suite $\{\lambda(k)\}_{k \geq 1}$ définie par :

$$\lambda(1) = 0.1, \lambda(2) = 0.4, \forall k \geq 3 \quad \lambda(k) = 0.3.$$

On a $R(1) = 0.9$, $R(2) = 0.54$, et $\forall k \geq 3 \quad R(k) = 0.54 * 0.7^{k-2}$.

On montre que $\frac{R(k+1)^{1/(k+1)}}{R(k)^{1/k}} = \left[\frac{49}{54} \right]^{\frac{1}{k(k+1)}} \leq 1$.

Donc la suite $\{R(k)^{\frac{1}{k}}\}_{k \geq 1}$ est décroissante et la loi est IFRA1 alors qu'elle n'est pas IFR. \square

Théorème 6 *Complétude de la classe IFRA1*

On considère un système cohérent constitué de n composants indépendants dont les durées de vie sont à valeurs dans \mathbb{N}^ . Alors si les lois des durées de vie des composants sont IFRA1, la loi de la durée de vie du système est IFRA1.*

Preuve :

La démonstration s'inspire de celle de Barlow et Proschan [10] (page 85) dans le cas continu. Soient $R_1(k), \dots, R_n(k)$ les fiabilités des n composants d'un système cohérent. Soit $R(k)$ la fiabilité du système, et soit h la fonction de fiabilité du système :

$$\forall k \in \mathbb{N}^*, R(k) = h(R_1(k), \dots, R_n(k))$$

La loi de la durée de vie du système sera IFRA1 si et seulement si :

$$\forall (j, k) \in \mathbb{N}^{*2}, j \leq k \Rightarrow R(j) \geq R(k)^{\frac{j}{k}}$$

Si tous les composants sont de loi IFRA1, alors :

$$\forall (j, k) \in \mathbb{N}^{*2}, \forall i \in \{1, \dots, n\}, j \leq k \Rightarrow R_i(j) \geq R_i(k)^{\frac{j}{k}}$$

Comme la fonction de fiabilité d'un système cohérent est monotone, on a :

$$R(j) = h(R_1(j), \dots, R_n(j)) \geq h(R_1(k)^{\frac{j}{k}}, \dots, R_n(k)^{\frac{j}{k}})$$

Or pour toute fonction de fiabilité d'un système cohérent, on a (voir page 85 de [10]):

$$\forall \alpha \in]0, 1[, \quad h(p_1^\alpha, \dots, p_n^\alpha) \geq h^\alpha(p_1, \dots, p_n)$$

Cette propriété est valable quel que soit le type de loi des durées de vie des composants, discrète ou continue. Donc

$$\forall (j, k) \in \mathbb{N}^{*2}, R(j) \geq h(R_1(k)^{\frac{j}{k}}, \dots, R_n(k)^{\frac{j}{k}}) \geq h^{\frac{j}{k}}(R_1(k), \dots, R_n(k)) = R(k)^{\frac{j}{k}}$$

Ce qui prouve que la loi de la durée de vie du système est IFRA1. \square

c) Système meilleur nouveau qu'utilisé (NBU)

Les définitions en temps discret correspondant aux définitions équivalentes à NBU en temps continu sont :

- NBU1 : $\forall k \in \mathbb{N}^*, \forall j \in \mathbb{N}^*, R(j|k) \leq R(j)$,
- NBU2 : $\forall k \in \mathbb{N}^*, \forall j \in \mathbb{N}^*, \Lambda(k + j) \geq \Lambda(k) + \Lambda(j)$.

Propriété 3

Dans le cas NBU, aucune des deux formulations NBU1 et NBU2 n'implique l'autre.

Preuve :

Les deux contre-exemples ci-dessous ont été proposés par Shaked *et al.* [84]. Soit K une v.a. de taux de défaillance défini par $\lambda(1) = 0.5, \lambda(2) = 0.8, \lambda(3) = 0.8, \lambda(4) = 0.5, \lambda(5) = 0.9, \lambda(6) = 0.6$, et $\lambda(k) = 0.99$ pour $k \geq 7$.

On montre alors que : $\forall k \in \mathbb{N}^*, \forall j \in \mathbb{N}^*, \prod_{i=1}^k (1 - \lambda(i)) \geq \prod_{i=j+1}^{j+k} (1 - \lambda(i))$ et donc K est

NBU1. Mais $\sum_{i=1}^3 \lambda(i) > \sum_{i=4}^6 \lambda(i)$ et donc K n'est pas NBU2.

Si on définit le taux de défaillance de la v.a. K par $\lambda(1) = 0.4, \lambda(2) = 0.6, \lambda(3) = 0.5, \lambda(4) = 0.5$, et $\lambda(k) = 0.99$ pour $k \geq 5$, on montre que la v.a. K est NBU2, mais n'est pas NBU1 car $\prod_{i=1}^2 (1 - \lambda(i)) < \prod_{i=3}^4 (1 - \lambda(i))$. □

Les implications IFRA1 \Rightarrow NBU1 et IFRA2 \Rightarrow NBU2 sont immédiates. Par ailleurs, Tang-Yu-Chew [90] ont montré que IFR \Rightarrow DMRL et Ebrahimi [26] a montré que DMRL $\not\Rightarrow$ IFR. L'analogie en discret de (2.8) est donc finalement :

$$\begin{array}{ccc}
 IFR & \Rightarrow & IFRA1 \Rightarrow NBU1 \\
 \Downarrow & & \Downarrow \\
 DMRL & & IFRA2 \Rightarrow NBU2
 \end{array} \tag{2.9}$$

Remarque 2 :

Grâce à la formule (2.7), Ebrahimi [26] a donné une caractérisation simple des lois DMRL ou IMRL. Dans le même ordre d'idée, Langberg *et al.* [49] ont caractérisé la classe des lois DFR en temps discret.

Remarque 3 :

Des résultats analogues pour des taux de défaillance “en baignoire” ou “en baignoire inversée” en temps continu et discret ont été montrés par Guess-Park [31], Tang-Yu-Chew [90], Ghai-Mi [29] et Mi [56].

2.6 Simulation de Monte-Carlo des modèles de durée de vie en temps discret

2.6.1 Méthode utilisant le taux de défaillance

Soit un modèle de fiabilité en temps discret défini par son taux de défaillance $\lambda(k)$, $k \in \mathbb{N}^*$. On souhaite simuler des réalisations de K , la variable aléatoire ayant ce taux de défaillance. On sait que :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(K = k) = \lambda(k) \prod_{i=1}^{k-1} (1 - \lambda(i))$$

Soit $\{U_i\}_{i \geq 1}$ une suite de v.a. indépendantes de lois de Bernoulli de paramètres respectifs $\lambda(i)$. On pose $X = \min\{i \geq 1; U_i = 1\}$. X est une v.a. à valeurs dans \mathbb{N}^* .

$$\begin{aligned} \forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) &= \mathbb{P}(\text{les } k - 1 \text{ premiers } U_i \text{ sont nuls et le } k^{\text{ème}} \text{ est égal à } 1) \\ &= \left(\prod_{i=1}^{k-1} \mathbb{P}(U_i = 0) \right) \mathbb{P}(U_k = 1) \\ &= \left[\prod_{i=1}^{k-1} (1 - \lambda(i)) \right] \lambda(k) \end{aligned}$$

Donc X a la loi cherchée.

On peut donc simuler très facilement n'importe quelle loi de fiabilité en temps discret en utilisant l'algorithme suivant :

```

i ← 1
TantQue (random > λ(i)) faire
    i ← i+1
finTantQue
X ← i

```

2.6.2 Méthode d'inversion

L'algorithme de simulation par inversion de la fonction de répartition est l'algorithme suivant :

```

i ← 1
choix ← random
TantQue (choix > F(i)) faire
    i ← i+1
finTantQue
X ← i

```

Dans la première méthode, on effectue X appels de *random*, alors que dans la seconde, on n'en fait qu'un seul. C'est la raison pour laquelle la méthode de simulation par inversion est plus rapide en général que la première, sauf lorsque $\lambda(i)$ est plus simple à calculer que $F(i)$.

2.6.3 Utilisation de modèles en temps continu

Dans le chapitre 3, nous recenserons des modèles de durée de vie en temps discret issus de modèles usuels en temps continu. Si T est une variable aléatoire continue, alors on obtient une variable aléatoire à valeurs dans \mathbb{N}^* en posant $K = [T] + 1$. Pour simuler ces modèles discrets, il suffit de simuler T par la méthode d'inversion, d'en prendre la partie entière et d'ajouter 1. Lorsque cette méthode est utilisable, elle est beaucoup plus rapide que les deux précédentes.

Modèles de durée de vie en temps discret

Ce chapitre présente une revue des différents modèles de durée de vie en temps discret existant. Nous proposons une classification en trois familles. Une première famille concerne les lois discrètes usuelles qui n'ont pas été spécifiquement proposées dans un contexte de fiabilité. Une deuxième famille regroupe les modèles en temps discret issus de modèles de fiabilité usuels en temps continu, avec des analogies pouvant porter sur différents critères (même expression du taux de défaillance, de la fiabilité, etc). La troisième famille englobe les lois de probabilité issues du schéma d'urne de Pólya.

Pour chaque modèle, nous donnons l'expression des grandeurs de base p , R , λ , ainsi que, lorsque c'est possible, du MTTF, et nous donnons un graphe du taux de défaillance. Les caractéristiques des modèles, leurs avantages et leurs défauts sont précisés.

3.1 Bibliographie chronologique

Nakagawa et Osaki [58] sont les premiers en 1975 à proposer un modèle de durée de vie en temps discret. Le modèle est défini pour être l'équivalent en temps discret de la loi de Weibull en temps continu. Il faudra attendre 1984 pour que des travaux sur la loi de Weibull en temps discret soient poursuivis.

La notion de persistance en climatologie peut être comprise comme le complémentaire à 1 du taux de défaillance. Il est donc possible d'adapter pour la fiabilité des modèles d'étude des séquences climatologiques, en particulier le modèle proposé par Le Breton et Martin [52] en 1979 d'après les travaux d'EGgenberger-Pólya [28].

En 1982, Salvia et Bollinger [80] illustrent leurs résultats fondamentaux sur le taux de défaillance en temps discret, par des exemples simples. Stein et Dattero [88] définissent, en 1984, une nouvelle loi de Weibull discrète, dite de type II. En 1985, Padgett et Spurrier [66] reprennent le travail de Salvia et Bollinger et généralisent leurs modèles en ajoutant un second paramètre afin de les rendre plus "flexibles". Ces modèles existent en fait dans la littérature (voir [41] et [42]) sous le nom de modèles inverses de Pólya. Les auteurs donnent une troisième forme discrète de la loi de Weibull. En 1983, Xekalaki [94], s'inspirant d'un

article d'Irwin [39], s'intéresse à caractériser les lois à taux de défaillance inverse linéaire, dont la loi de Waring qui s'avère être une autre appellation de la loi de Salvia-Bollinger généralisée à taux de défaillance décroissant. Un travail similaire concernant les lois à durée de vie résiduelle moyenne linéaire a été effectué par Roy et Gupta en 1999 [77].

En 1992, Kaaniche et Kanoun [43], et Soler [86], en 1996, proposent des modèles de croissance de fiabilité des logiciels dont on peut extraire de chacun un modèle de fiabilité : la loi hyper-exponentielle discrète pour les premiers et la loi "S" pour Soler. En 1997, Clarotti, Lannoy et Procaccia [16] effectuent une modélisation bayésienne du vieillissement à la sollicitation, qui utilise un modèle inverse de Pólya. Enfin, en 1997, Gupta-Gupta-Tripathi [32] proposent des familles étendues de lois à taux de défaillance croissant, comme les familles de Katz ou Kemp.

Remarque :

Par rapport aux modèles originaux tels qu'ils sont donnés dans la littérature, il a souvent été nécessaire d'opérer un décalage d'une unité sur les valeurs prises par la variable discrète afin de travailler sur \mathbb{N}^* .

3.2 Lois discrètes usuelles

Dans cette section, nous nous contenterons de citer seulement les lois discrètes usuelles qui, ayant un support sur \mathbb{N} ou \mathbb{N}^* , ont été considérées par les premiers auteurs travaillant sur le sujet (Barlow-Marshall-Proschan [8], Kaufmann-Grouchko-Cruon [46],...) comme de possibles modèles de fiabilité : les lois géométrique, binomiale négative et Poisson.

3.2.1 La loi géométrique

Si la v.a. K suit la loi géométrique de paramètre p , alors K représente le nombre d'essais nécessaires pour obtenir la réalisation d'un événement de probabilité p dans une suite d'expériences identiques et indépendantes. La loi géométrique $\mathcal{G}(p)$ est définie par :

$$\forall k \in \mathbb{N}^*, \quad p(k) = p(1-p)^{k-1}, \quad p \in]0, 1[$$

La fiabilité est alors :

$$\forall k \in \mathbb{N}^*, \quad R(k) = (1-p)^k$$

Le temps moyen jusqu'à défaillance est :

$$MTTF = \frac{1}{p}$$

Le taux de défaillance d'une loi géométrique est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = p$$

Le taux de défaillance d'une loi géométrique est constant et égal à la valeur du paramètre p , ce qui signifie une absence d'usure pour le phénomène étudié. Cette propriété d'absence de vieillissement s'écrit :

$$\forall (i, k) \in \mathbb{N}^{*2}, \quad \mathbb{P}(K > i + k | K > i) = \mathbb{P}(K > k)$$

Cette caractérisation (et d'autres telle que la conservation de la loi par le minimum, etc) fait de la loi géométrique l'analogie en temps discret de la loi exponentielle.

3.2.2 La loi binomiale négative décalée

Habituellement, la loi binomiale négative représente la loi du nombre d'essais nécessaires pour observer la $r^{\text{ème}}$ occurrence d'un événement de probabilité p dans une suite d'expériences identiques et indépendantes. C'est la raison pour laquelle la loi binomiale négative est à valeurs dans $r, r + 1, \dots, +\infty$. Il suffit de poser $K = X - r + 1$ où la variable aléatoire X suit la loi binomiale négative $\mathcal{BN}(p, r)$ pour nous ramener sur \mathbb{N}^* .

L'expression des probabilités élémentaires de la loi binomiale négative décalée $\mathcal{BN}_d(p, r)$ est :

$$\forall k \in \mathbb{N}^*, \quad p(k) = C_{k+r-2}^{r-1} p^r (1-p)^{k-1}, \quad r \in \mathbb{N}^*, \quad p \in]0, 1[$$

La fiabilité est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = 1 - \sum_{i=1}^{k-1} C_{i+r-2}^{r-1} p^r (1-p)^{i-1}$$

Le temps moyen jusqu'à défaillance est :

$$MTTF = \frac{r(1-p)}{p} + 1.$$

Le taux de défaillance est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = \frac{C_{k+r-2}^{r-1} p^r (1-p)^{k-1}}{R(k-1)}$$

$\frac{p(k+1)}{p(k)} = (1 + \frac{r-1}{k})(1-p)$ est une fonction décroissante de k . Donc, en vertu du théorème 4 sur le rapport de probabilités successives, la loi binomiale négative décalée est log-concave et son taux de défaillance est croissant. Pour $r = 1$, on retrouve l'expression de la loi géométrique et le taux de défaillance est alors constant.

Remarque :

La loi binomiale négative est la loi de la somme de r variables aléatoires indépendantes de même loi géométrique. La loi géométrique étant l'équivalent en temps discret de la loi exponentielle, on peut considérer que la loi binomiale négative est l'analogie en discret de la loi Gamma, qui est la loi de la somme de variables aléatoires indépendantes et de même loi exponentielle.

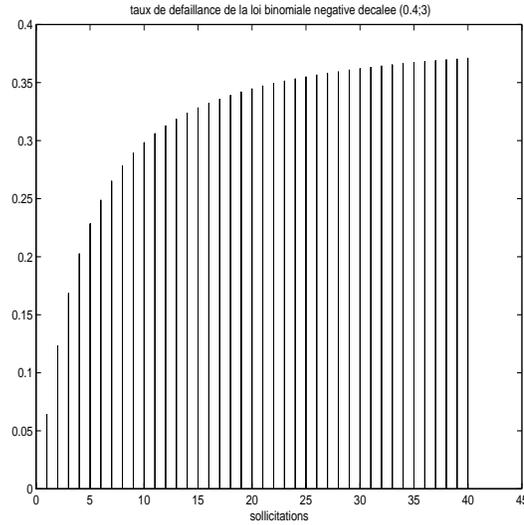


FIG. 3.1 – Taux de défaillance croissant de la loi binomiale négative décalée

3.2.3 La loi de Poisson décalée

Habituellement, dans le domaine de la fiabilité, on utilise la loi de Poisson pour modéliser le nombre de défaillances par unité de temps sur un parc d'appareils similaires fiables. Pour notre domaine d'application, on posera $K = X + 1$ où la variable aléatoire X suit la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre λ .

L'expression des probabilités élémentaires de la loi de Poisson décalée $\mathcal{P}_d(\lambda)$ est :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!}, \quad \lambda \in \mathbb{R}^{+*}$$

Le temps moyen jusqu'à défaillance est :

$$\forall k \in \mathbb{N}^*, \quad MTTF = 1 + \lambda$$

La fiabilité est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = 1 - \sum_{i=1}^{k-1} \frac{e^{-\lambda} \lambda^{i-1}}{(i-1)!}$$

Le taux de défaillance est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)! \left(1 - \sum_{i=1}^{k-1} \frac{e^{-\lambda} \lambda^{i-1}}{(i-1)!}\right)}$$

$\frac{p(k+1)}{p(k)} = \frac{\lambda}{k}$ est une fonction décroissante de k , donc la loi de Poisson décalée est log-concave et son taux de défaillance est croissant.

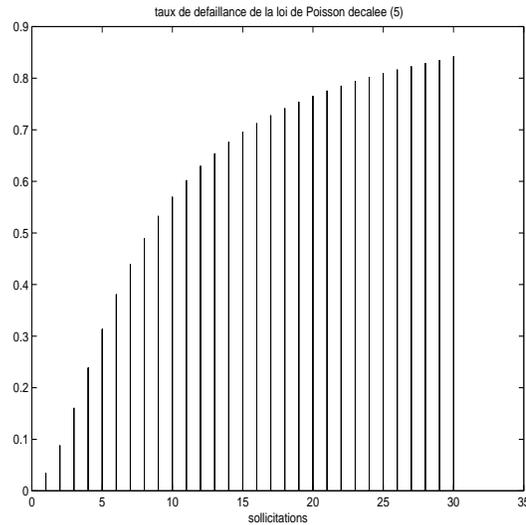


FIG. 3.2 – Taux de défaillance de la loi de Poisson décalée

3.3 Les modèles en temps discrets issus du temps continu

Les modèles en temps discret dérivés du temps continu peuvent être classés en deux familles. La première concerne les modèles construits sur une similitude d'expression entre temps continu et temps discret pour une certaine grandeur de la fiabilité. La seconde considère une durée de vie en temps discret comme la partie entière d'une durée de vie en temps continu.

3.3.1 Modèles en temps discret aux caractéristiques identiques à un modèle en temps continu

a) La loi géométrique

La loi géométrique est l'analogie en temps discret de la loi exponentielle car les deux lois possèdent les mêmes propriétés de caractérisation : taux de défaillance constant, conservation de la loi par le minimum, etc. Les caractéristiques de la loi géométrique sont données à la section 3.2.1.

b) La loi de Weibull discrète de type I

La loi de Weibull discrète, dite de type I, a été définie par Nakagawa et Osaki [58] dans le but de définir l'analogie en temps discret de la loi de Weibull continue. C'est la première loi de probabilité définie spécifiquement pour être un modèle de fiabilité en temps discret. Le modèle est basé sur la similitude d'expression de la fiabilité entre la loi continue et la loi discrète. Si T est de loi de Weibull $\mathcal{W}(\eta, \beta)$ en temps continu, $R(t) = e^{-\left(\frac{t}{\eta}\right)^\beta}$. Une expression similaire pour la fiabilité en temps discret est $R(k) = e^{-\left(\frac{k}{\eta}\right)^\beta}$ qui peut s'écrire $R(k) = q^{k^\beta}$ où

$\beta \in \mathbb{R}^{+*}$ et $q \in]0, 1[$.

La loi de Weibull de type I est donc définie par :

$$\forall k \in \mathbb{N}^*, \quad p(k) = q^{(k-1)\beta} - q^{k\beta}$$

Cette loi est également citée en exemple par Kalbfleisch et Prentice [44].

Le taux de défaillance de cette loi est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = 1 - q^{k\beta - (k-1)\beta}$$

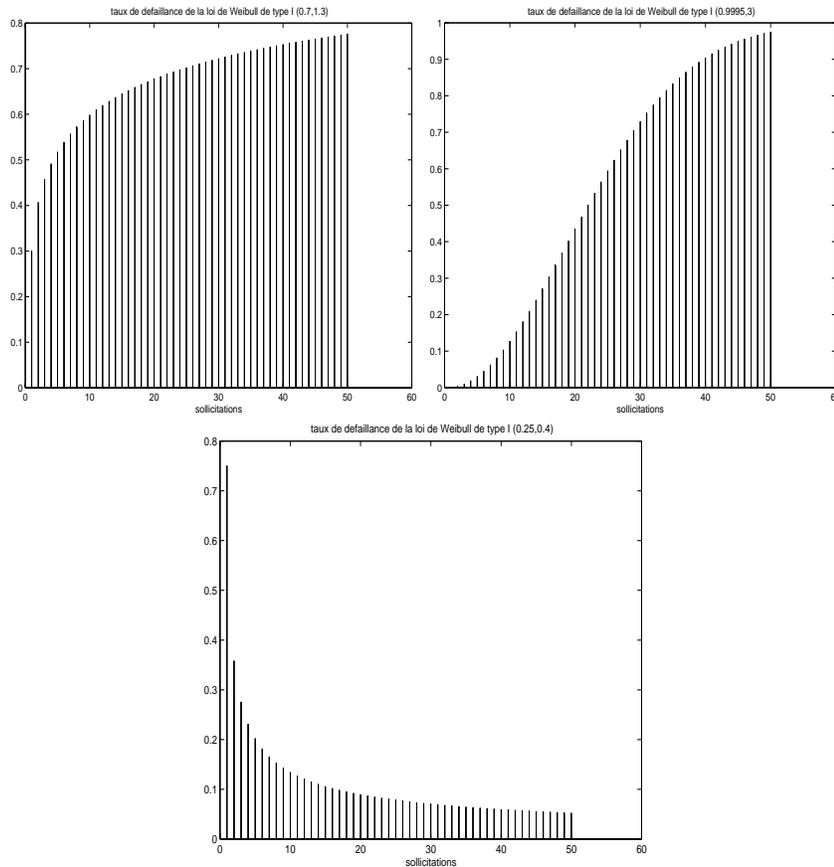


FIG. 3.3 – Taux de défaillance de la loi de Weibull de type I avec $\beta = 1.3, 3$ et 0.4

Comme pour la loi de Weibull en temps continu, le sens de variation du taux de défaillance possède la propriété de dépendre de la valeur de β de la façon suivante :

1. Pour $\beta = 1$, le modèle se réduit à la loi géométrique.
2. Pour $0 < \beta < 1$, la loi est DFR.
3. Pour $\beta > 1$, la loi est IFR.

On a $MTTF = \sum_{k=0}^{+\infty} q^{k\beta}$. Il n'existe pas d'expression plus simple du MTTF.

c) La loi de Weibull discrète de type II

Stein et Dattero [88] ont donné une nouvelle forme de la loi de Weibull discrète, dite de type II, dont la construction repose sur la conservation de l'expression du taux de défaillance entre le temps discret et le temps continu.

Si T est de loi de Weibull $\mathcal{W}(\eta, \beta)$ en temps continu, $\lambda(t) = (\frac{t}{\eta})^{\beta-1}$. Une expression similaire pour le taux de défaillance en temps discret est $\lambda(k) = (\frac{k}{\eta})^{\beta-1} = ck^{\beta-1}$ avec $\beta \in \mathbb{R}^{+*}$ et $c \in]0, 1]$. Mais comme $\lambda(k)$ doit être inférieur à 1, il est nécessaire d'introduire l'entier m , borne supérieure du support de la loi :

$$m = \begin{cases} \lfloor c^{-\frac{1}{\beta-1}} \rfloor, & \text{si } \beta > 1 \\ +\infty, & \text{si } \beta \leq 1 \end{cases}$$

La nouvelle loi de Weibull discrète de type II est alors définie par son taux de défaillance :

$$\lambda(k) = \begin{cases} ck^{\beta-1} & \text{si } k \in \{1, 2, \dots, m\} \\ 0 & \text{sinon} \end{cases}$$

Il s'avère que cette définition de la loi de Weibull II ne vérifie pas le théorème 1 pour toutes les valeurs de c et donc ne définit pas une loi de probabilité. Ce modèle discret ne définit une loi de probabilité que pour les valeurs de c telles que $c^{-\frac{1}{\beta-1}}$ soit entier. Nous allons donc redéfinir le taux de défaillance de la loi de Weibull II avec cette hypothèse sur c :

$$\lambda(k) = \begin{cases} (\frac{k}{m})^{\beta-1} & \text{si } k \in \{1, 2, \dots, m\} \\ 0 & \text{sinon} \end{cases}$$

Par la formule (2.3), on peut donner l'expression des probabilités élémentaires :

$$p(k) = \begin{cases} (\frac{k}{m})^{\beta-1} \prod_{i=1}^{k-1} \left(1 - (\frac{i}{m})^{\beta-1}\right) & \text{si } k \leq m \\ 0 & \text{si } k > m \end{cases}$$

L'expression de la fiabilité est donnée par :

$$\forall k \in \mathbb{N}^*, \quad R(k) = \prod_{i=1}^{\inf(k, m)} \left(1 - (\frac{i}{m})^{\beta-1}\right)$$

Remarque :

D'un point de vue pratique, un modèle à support borné est peu vraisemblable : on n'est jamais certain qu'un système tombera en panne en moins de m sollicitations. On aura donc tendance à ne pas utiliser ce modèle.

d) La loi de Weibull de type III

Dans [66], Padgett et Spurrier donnent une autre forme d'une loi de probabilité analogue à la loi de Weibull dans le sens où le taux de défaillance peut être croissant, constant ou décroissant, suivant la valeur d'un paramètre :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = 1 - e^{-ck^\beta}, \quad c > 0, -\infty < \beta < +\infty$$

β est le paramètre de forme, et $c = -\ln(1 - p(1))$. Par la formule (2.3), on peut donner l'expression des probabilités élémentaires :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \left(1 - e^{-ck^\beta}\right) e^{-c \sum_{i=1}^{k-1} i^\beta}$$

L'expression de la fiabilité est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = e^{-c \sum_{j=1}^k j^\beta}$$

L'expression du taux de défaillance possède l'avantage d'être un peu plus simple que celle de la loi de Weibull discrète de type I. Son sens de variation dépend de la valeur de β :

1. Pour $\beta = 0$, le modèle se réduit à la loi géométrique.
2. Pour $\beta > 0$, la loi est IFR.
3. Pour $\beta < 0$, la loi est DFR.

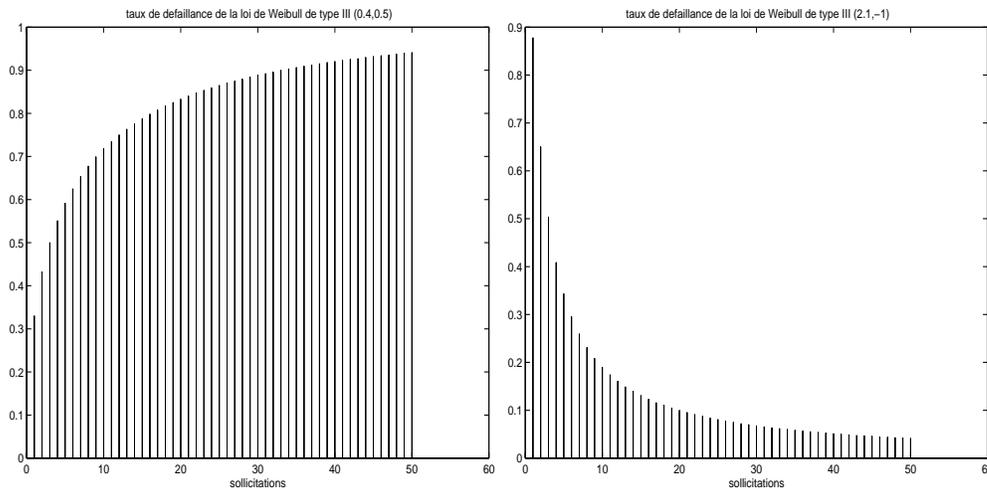


FIG. 3.4 – Taux de défaillance de la loi de Weibull de type III

On peut montrer que le *MTTF* existe si $\beta > -1$, ou $\beta = -1$ et $c > 1$, mais il n'a pas d'expression explicite simple.

e) La loi hyperexponentielle discrète

Kaaniche-Kanoun [43] ont introduit un modèle de croissance de fiabilité des logiciels en temps discret, équivalent au modèle hyperexponentiel en temps continu. La loi hyperexponentielle peut modéliser la durée de vie d'un système non réparable en temps discret. Son taux de défaillance est :

$$\lambda(k) = \frac{\theta p_{sup}(1 - p_{sup})^{k-1} + \bar{\theta} p_{inf}(1 - p_{inf})^{k-1}}{\theta(1 - p_{sup})^{k-1} + \bar{\theta}(1 - p_{inf})^{k-1}}$$

avec $\theta \in [0, 1]$, $\bar{\theta} = 1 - \theta$ et $p_{inf} \leq p_{sup}$. Il est décroissant, mais on peut facilement construire un modèle analogue à taux de défaillance croissant. Cette loi possède trois paramètres et est de ce fait plus complexe que les autres lois présentées ici.

f) La loi "S"

Soler [86] a proposé une loi appelée loi S , pour la durée de vie en temps continu de systèmes soumis à des stress. On propose ici son analogue en temps discret. On considère un appareil susceptible de subir, à chaque sollicitation, un stress avec la probabilité p , et de ne pas le subir avec la probabilité $q = 1 - p$. L'instant de la première défaillance K , est tel que le taux de défaillance, conditionnellement à la suite des stress est, à la $k^{\text{ème}}$ sollicitation, une fonction croissante du nombre de stress N_k subis jusque-là. On exprime ce fait par la formule :

$$\forall k \in \mathbb{N}^*, \quad \lambda_N(k) = \mathbb{P}(K = k | K \geq k, \{N_j\}_{j \geq 1}) = 1 - \pi^{N_k}, \quad \text{avec } \pi \in]0, 1].$$

La fiabilité, conditionnellement à la suite des stress, est :

$$\begin{aligned} R(k) &= P(K > k) = \mathbb{E}[P(K > k | \{N_j\}_{j \geq 1})] \\ &= \mathbb{E} \left[\prod_{i=1}^k (1 - \mathbb{P}(K = i | K \geq i, \{N_j\}_{j \geq 1})) \right] \\ &= \mathbb{E} \left[\prod_{i=1}^k \pi^{N_i} \right] \end{aligned}$$

Les v.a. N_i n'étant pas indépendantes, on définit alors les v.a. U_k par :

$\forall k \geq 1, \quad U_k = N_k - N_{k-1}$. On a $N_k = \sum_{i=1}^k U_i$, avec les v.a. U_i indépendantes et de loi de

Bernoulli $\mathcal{B}(p)$ caractérisant l'occurrence d'un stress à l'instant i .

$$\begin{aligned} \text{Alors } \sum_{i=1}^k N_i &= N_k - N_{k-1} + 2(N_{k-1} - N_{k-2}) + \cdots + (k-1)(N_2 - N_1) + kN_1 \\ &= \sum_{i=1}^k (k-i+1)U_i \\ \text{D'où } R(k) &= \mathbb{E} \left[\prod_{i=1}^k \pi^{N_i} \right] = \mathbb{E} \left[\pi^{\sum_{i=1}^k N_i} \right] = \mathbb{E} \left[\pi^{\sum_{i=1}^k (k-i+1)U_i} \right] = \prod_{i=1}^k \mathbb{E} \left[\pi^{(k-i+1)U_i} \right] \\ &= \prod_{i=1}^k G(\pi^{k-i+1}) \end{aligned}$$

où G est la fonction génératrice de la loi $\mathcal{B}(p)$ définie par $G(u) = pu + 1 - p$.

Finalement, la fiabilité du système est donnée par :

$$R(k) = \prod_{i=1}^k (q + p\pi^{k-i+1}) = \prod_{i=1}^k (q + p\pi^i)$$

On définit ainsi les probabilités élémentaires de la loi S :

$$\forall k \in \mathbb{N}^*, \quad p(k) = p(1 - \pi^k) \prod_{i=1}^{k-1} (q + p\pi^i)$$

Le taux de défaillance de cette loi discrète est croissant :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = p(1 - \pi^k)$$

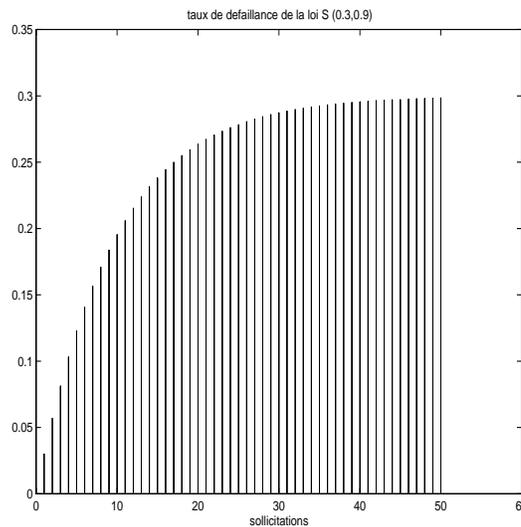


FIG. 3.5 – Taux de défaillance de la loi S

En particulier, si on considère qu'un stress survient à chaque sollicitation, $p = 1$, on obtient un taux de défaillance très simple : $\lambda(k) = 1 - \pi^k$. C'est un cas particulier du taux de

défaillance de la loi de Weibull de type III avec $\beta = 1$ et $c = -\ln \pi$.

Le MTTF de la loi S n'a pas d'expression explicite simple.

3.3.2 Modèles discrets dérivés du temps continu par discrétisation de l'échelle du temps

Soit T une variable aléatoire réelle positive représentant la durée de vie d'un système en temps continu. On pose $K = \lfloor T \rfloor + 1$ (où $\lfloor \cdot \rfloor$ désigne la partie entière). K est une variable aléatoire discrète à valeurs dans \mathbb{N}^* . On notera par $\lambda_K, F_K, R_K, \Lambda_K$ et $\lambda_T, F_T, R_T, \Lambda_T$ les grandeurs de la fiabilité associées respectivement aux v.a. K et T .

Il nous faut exprimer les probabilités élémentaires de la loi de K à l'aide de la fonction de répartition de T :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(K = k) = \mathbb{P}(k - 1 \leq T < k) = F_T(k) - F_T(k - 1) \quad (3.1)$$

De plus :

$$F_K(k) = \mathbb{P}(K \leq k) = \mathbb{P}(\lfloor T \rfloor + 1 \leq k) = \mathbb{P}(T < k) = F_T(k)$$

D'où :

$$R_K(k) = R_T(k) \quad (\text{ou } F_K(k) = F_T(k)).$$

Le taux de défaillance de la v.a. K s'écrit à l'aide de la fiabilité de T de la façon suivante :

$$\begin{aligned} \forall k \in \mathbb{N}^*, \quad \lambda_K(k) &= 1 - \frac{R_T(k)}{R_T(k-1)} = 1 - \frac{e^{-\int_0^k \lambda_T(u) du}}{e^{-\int_0^{k-1} \lambda_T(u) du}} \\ &= 1 - e^{-\int_{k-1}^k \lambda_T(u) du} = 1 - e^{-[\Lambda_T(k) - \Lambda_T(k-1)]} \end{aligned} \quad (3.2)$$

On peut remarquer que λ_T et λ_K possèdent le même sens de variation.

Un encadrement des moments d'ordre 1 et 2 de la variable discrète K est possible, à partir de la variable continue T dont elle est issue. En effet, à partir de la définition de K , on a l'inégalité suivante : $T < K \leq T + 1$ *p.s.* . On obtient les inégalités suivantes pour la moyenne et la variance de T :

$$\begin{cases} \mathbb{E}(T) < \mathbb{E}(K) \leq \mathbb{E}(T) + 1 \\ \text{Var}(T) - 2\mathbb{E}(T) - 1 < \text{Var}(K) \leq \text{Var}(T) + 2\mathbb{E}(T) + 1 \end{cases}$$

On constate ainsi que, pour de grandes valeurs de T , le passage à la partie entière ne change presque pas la moyenne, mais en revanche, la variance en temps discret peut être très éloignée de la variance en temps continu.

a) La loi exponentielle

On suppose que la variable aléatoire T suit la loi exponentielle $\mathcal{E}(\lambda)$ de fonction de répartition $F_T(t) = 1 - e^{-\lambda t}$. À l'aide de l'équation (3.1), on obtient :

$$\forall k \in \mathbb{N}^*, \quad p(k) = (e^{-\lambda})^{k-1}(1 - e^{-\lambda})$$

On reconnaît l'expression de la loi géométrique de paramètre $(1 - e^{-\lambda})$, dont le taux de défaillance constant est $\lambda(k) = 1 - e^{-\lambda}$.

Remarque :

Le taux de défaillance de la loi géométrique n'est donc pas égal au taux de défaillance de la loi exponentielle correspondante.

La fiabilité a pour expression :

$$\forall k \in \mathbb{N}^*, \quad R(k) = e^{-\lambda k}$$

b) La loi de Weibull

Si T est de loi de Weibull $\mathcal{W}(\eta, \beta)$, on a : $R_K(k) = e^{-(\frac{k}{\eta})^\beta} = q^{k^\beta}$ avec $q = e^{-\frac{1}{\eta^\beta}}$. On en déduit l'expression du taux de défaillance $\lambda(k) = 1 - q^{k^\beta - (k-1)^\beta}$. On retrouve donc la loi de Weibull discrète de type I.

c) La loi de Pareto

Une variable aléatoire réelle T suit une loi de Pareto $P(c, d)$, si sa fonction de répartition est $F_T(t) = 1 - \left(\frac{d}{t+d}\right)^c$ pour $t > 0, c \geq 0, d \geq 0$. À l'aide de l'équation (3.1), on obtient :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \left(\frac{d}{k-1+d}\right)^c - \left(\frac{d}{k+d}\right)^c$$

Cette expression n'est pas celle d'une loi connue, sa moyenne et sa variance ne peuvent être calculées explicitement.

La fiabilité de la loi de Pareto discrétisée est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = \left(\frac{d}{k+d}\right)^c$$

L'expression de son taux de défaillance est calculée à partir de la formule (3.2) et on trouve :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = 1 - \left(\frac{k-1+d}{k+d}\right)^c$$

Le taux de défaillance d'une loi de Pareto discrétisée est donc décroissant puisque $c \geq 0$.

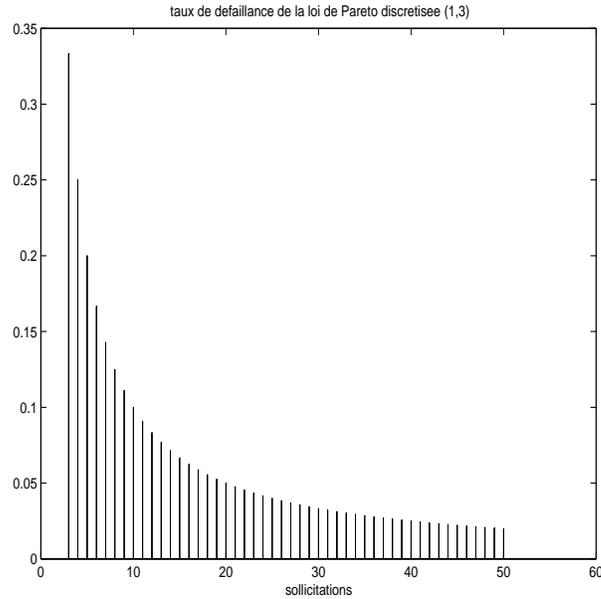


FIG. 3.6 – Taux de défaillance de la loi de Pareto discrétisée

d) La loi logistique tronquée

On suppose que la variable aléatoire T suit la loi logistique $\text{Log}(c, d)$ tronquée sur \mathbb{R}^+ (avec $d > 0$ et $c \in \mathbb{R}$) dont la fonction de répartition est $F_T(t) = \frac{1 - e^{-\frac{t}{d}}}{1 + e^{-\frac{t-c}{d}}}$. On trouve alors :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \frac{e^{-\frac{k-1}{d}} (1 - e^{-\frac{1}{d}}) (1 + e^{\frac{c}{d}})}{(1 + e^{-\frac{k-c}{d}}) (1 + e^{-\frac{k-1-c}{d}})}$$

Une fois de plus, la moyenne et la variance ne peuvent être données explicitement à cause de la complexité des probabilités élémentaires.

L'expression de la fiabilité est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = \frac{e^{-\frac{k-c}{d}} + e^{-\frac{k}{d}}}{1 + e^{-\frac{k-c}{d}}}$$

Le taux de défaillance de cette loi est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = \frac{1 - e^{-1/d}}{1 + e^{-\frac{k-c}{d}}}$$

Le taux de défaillance de la loi logistique discrétisée est donc croissant.

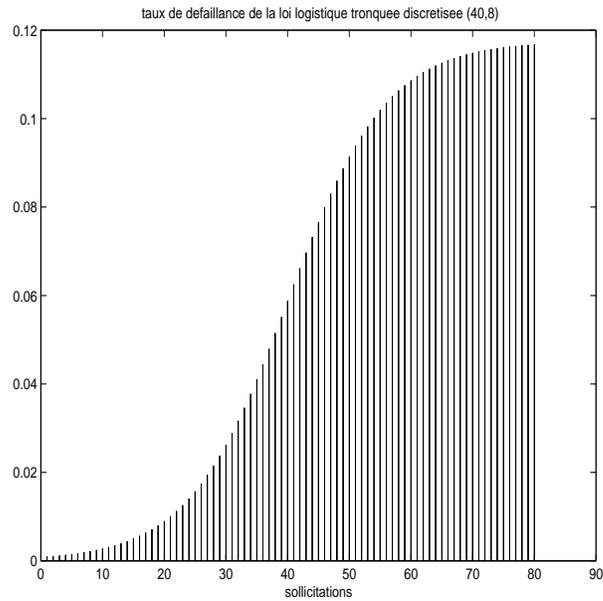


FIG. 3.7 – Taux de défaillance croissant de la loi logistique tronquée discrétisée

e) La loi log-logistique

On suppose que la variable aléatoire T suit la loi log-logistique $\mathcal{L}og - logistic(c, d)$ (avec $c > 0, d > 0$) de fonction de répartition $F_T(t) = \frac{t^d}{t^d + c}$.

On trouve alors par l'équation (3.1):

$$\forall k \in \mathbb{N}^*, \quad p(k) = \frac{c (k^d - (k-1)^d)}{(k^d + c) ((k-1)^d + c)}$$

L'expression de la fiabilité est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = \frac{c}{k^d + c}$$

Le taux de défaillance de cette loi est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = \frac{k^d - (k-1)^d}{k^d + c}$$

Suivant les valeurs des paramètres, le taux de défaillance est soit décroissant, soit croissant puis décroissant.

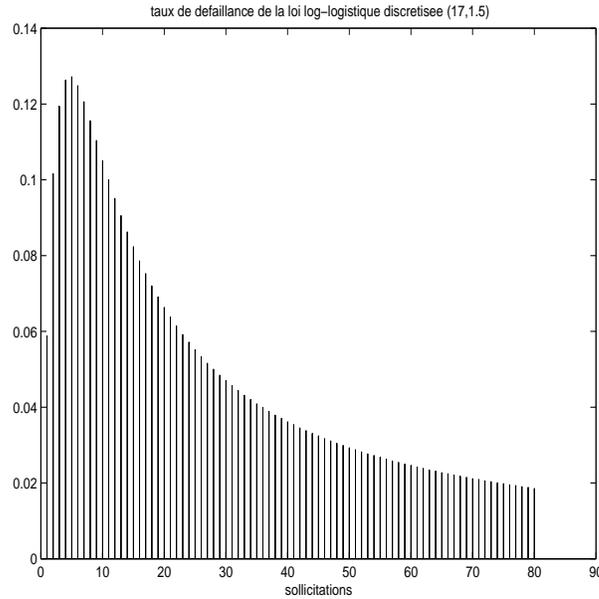


FIG. 3.8 – Taux de défaillance de la loi log-logistique discrétisée

f) Les modèles mixtes

Les experts industriels interrogés décrivent souvent la vie opérationnelle des dispositifs comme la succession de deux périodes. Pendant la première période, et jusqu'à un instant critique (noté τ), le matériel ne vieillit pas; durant la seconde période, i.e. au-delà du seuil critique, les effets du vieillissement sont plus notables. Du point de vue probabiliste, on peut donc supposer que :

- le taux de défaillance est constant jusqu'à l'instant critique τ ;
- à partir de τ , le taux de défaillance est croissant.

Si on considère qu'à partir de τ , le taux de défaillance croît comme celui d'une loi de Weibull, on obtient la loi exponentielle-Weibull introduite par Zacks [97]. Sa fonction de répartition est :

$$\forall t \geq 0, \quad F_T(t) = 1 - e^{-\lambda t - [\lambda(t-\tau)^+]^\beta}$$

où $Y^+ = \max(0, Y)$, $\lambda \in \mathbb{R}^{+*}$ est un paramètre d'échelle, $\beta \in \mathbb{R}^+$ est un paramètre de forme et $\tau \in \mathbb{R}^+$ est le paramètre de changement de tendance.

Pour la construction de la loi exponentielle-Weibull discrétisée, qu'on appellera géométrique-Weibull, le paramètre τ sera à valeurs dans \mathbb{N}^* et correspond à un seuil critique de sollicitations. L'expression de la probabilité élémentaire de cette loi est :

$$\forall k \in \mathbb{N}^*, \quad p(k) = e^{-\lambda(k-1) - [\lambda(k-\tau-1)^+]^\beta} - e^{-\lambda k - [\lambda(k-\tau)^+]^\beta}$$

L'espérance et la variance ne sont pas calculables explicitement. La fiabilité a pour expression :

$$\forall k \in \mathbb{N}^*, \quad R(k) = e^{-\lambda k - [\lambda(k-\tau)^+]^\beta}$$

Le taux de défaillance de cette loi est :

$$\begin{cases} \lambda(k) = 1 - e^{-\lambda + \lambda^\beta [(k-\tau-1)^\beta - (k-\tau)^\beta]} & \text{si } k \geq \tau + 1 \\ \lambda(k) = 1 - e^{-\lambda} & \text{si } k \leq \tau \end{cases}$$

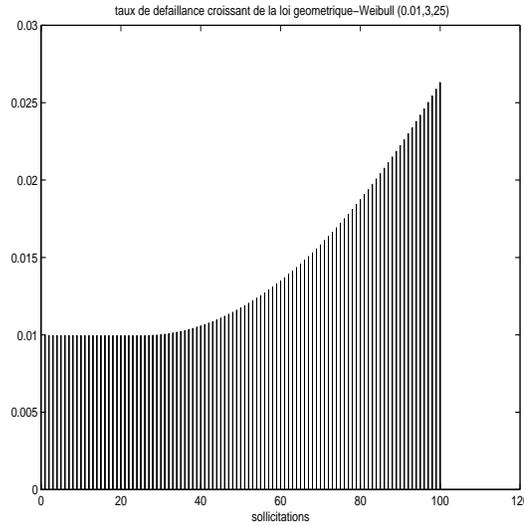


FIG. 3.9 – Taux de défaillance de la loi géométrique-Weibull

3.4 Modèles basés sur l'urne de Pólya

Comme on peut le voir dans Johnson-Kotz-Kemp [42], de nombreuses lois de probabilités discrètes ont été construites à partir de représentations de tirages dans des urnes. Parmi celles-ci, certaines sont particulièrement adaptées à un contexte de fiabilité.

Le schéma d'urne proposé par Eggenberger-Pólya est le suivant. Une urne contient B boules blanches et R boules rouges. À chaque tirage d'une boule, on adopte une certaine politique de remise. Les lois de Pólya sont les lois du nombre de boules rouges obtenues en N tirages. Les lois de Pólya inverses (voir p 192 de [41]) sont les lois du nombre de tirages à effectuer pour obtenir un nombre spécifié r de boules rouges. Il y a autant de lois différentes que de politiques de remise possibles. Par exemple, si on se contente de remettre la boule tirée dans l'urne, la loi de Pólya obtenue est la loi binomiale et la loi de Pólya inverse correspondante est la loi binomiale négative (loi géométrique pour $r = 1$).

Dans notre contexte, on considèrera que le tirage d'une boule blanche correspond à un bon fonctionnement à la sollicitation et le tirage d'une boule rouge à une défaillance. Par conséquent, la durée de vie du système considéré sera de loi de Pólya inverse avec $r = 1$.

Si la politique de remise consiste à remettre dans l'urne la boule (blanche) tirée et à rajouter Δ boules rouges, on voit que l'on augmente le risque de défaillance. La loi de durée de vie correspondante aura donc un taux de défaillance croissant. Inversement, si on remet

dans l'urne la boule (blanche) tirée ainsi que Δ autres boules blanches, la loi de durée de vie correspondante aura un taux de défaillance décroissant.

Dans ce cas, le taux de défaillance au $k^{\text{ème}}$ tirage peut être compris comme la probabilité de tirer (pour la première fois) une boule rouge au $k^{\text{ème}}$ tirage, sachant qu'aucune boule rouge n'a été tirée au cours des $k - 1$ tirages précédents.

Dans cette section, nous décrivons les lois de Pólya inverses IFR et DFR présentées ci-dessus. Ensuite, nous présenterons les lois de Salvia et Bollinger et leurs généralisations qui sont en fait des lois de Pólya inverses. Enfin, nous présenterons le modèle d'EGgenberger-Pólya, issu d'études climatologiques.

3.4.1 La loi de Pólya inverse IFR

Si on remet la boule blanche tirée et que l'on rajoute Δ boules rouges, on voit que le taux de défaillance est donné par :

$$\lambda(k) = \frac{R + (k - 1)\Delta}{R + B + (k - 1)\Delta}$$

En posant $\theta = \frac{R}{R + B}$ et $\delta = \frac{\Delta}{R + B}$, on obtient

$$\lambda(k) = \frac{\theta + (k - 1)\delta}{1 + (k - 1)\delta} = 1 - \frac{1 - \theta}{1 + (k - 1)\delta}, \quad 0 < \theta < 1, \quad 0 < \delta \quad (3.3)$$

Le paramètre $\theta \in]0, 1[$ est la probabilité d'échec au premier tirage. c'est à dire dans notre contexte, la probabilité de défaillance à la première sollicitation. $\delta > 0$ représente l'importance du phénomène de vieillissement. Pour $\delta = 0$, on retrouve la loi géométrique.

Ce modèle a été utilisé par Clarotti-Lannoy-Procaccia [16] pour modéliser le vieillissement à la sollicitation des diesels dans les centrales nucléaires françaises.

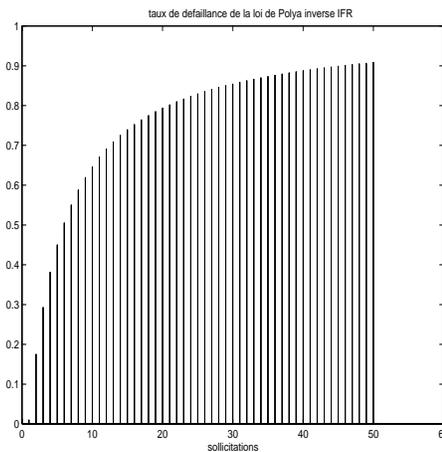


FIG. 3.10 – Taux de défaillance de la loi de Pólya inverse IFR

Par la formule (2.3), on peut donner l'expression de la probabilité élémentaire qui correspond à ce taux de défaillance :

$$p(k) = \frac{(1-\theta)^{k-1}[\theta + (k-1)\delta]}{\prod_{i=1}^k [1 + (i-1)\delta]} = \frac{(1-\theta)^{k-1}[\theta + (k-1)\delta]}{\delta^k \left(\frac{1}{\delta}\right)_{(k)}}$$

en utilisant la notation de Pochhammer : $(a)_{(k)} = \begin{cases} \prod_{i=1}^k (a+i-1) & \text{si } k \geq 1, \\ 1 & \text{si } k = 0. \end{cases}$

L'expression de la fiabilité est :

$$R(k) = \frac{(1-\theta)^k}{\delta^k \left(\frac{1}{\delta}\right)_{(k)}}$$

Propriété 4

$$MTTF = \frac{(1-\delta)\delta^{\frac{1}{\delta}-2}}{(1-\theta)^{\frac{1-\delta}{\delta}}} e^{\frac{1-\theta}{\delta}} \gamma\left(\frac{1-\delta}{\delta}, \frac{1-\theta}{\delta}\right)$$

où $\gamma(b, x) = \int_0^x t^{b-1} e^{-t} dt$ est la fonction Gamma incomplète.

Preuve :

$$\begin{aligned} MTTF &= \sum_{k=0}^{+\infty} R(k) = \sum_{k=0}^{+\infty} \frac{(1-\theta)^k}{\delta^k \left(\frac{1}{\delta}\right)_{(k)}} \\ &= \sum_{k=0}^{+\infty} \frac{k!}{\left(\frac{1}{\delta}\right)_{(k)} k!} \left(\frac{1-\theta}{\delta}\right)^k = \sum_{k=0}^{+\infty} \frac{(1)_{(k)}}{\left(\frac{1}{\delta}\right)_{(k)} k!} \left(\frac{1-\theta}{\delta}\right)^k \\ &= {}_1F_1\left(1, \frac{1}{\delta}; \frac{1-\theta}{\delta}\right) \end{aligned}$$

où ${}_1F_1(a, c; z) = \sum_{k=0}^{+\infty} \frac{(a)_{(k)}}{(c)_{(k)} k!} z^k$ est une série de Kummer (voir [42]).

Dans [42], Johnson-Kotz-Kemp donnent la représentation sous forme intégrale de la série de Kummer :

$${}_1F_1(a, c; z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 u^{(a-1)} (1-u)^{c-a-1} e^{zu} du$$

En utilisant cette formule, pour $a = 1$, on obtient :

$${}_1F_1(1, c, z) = \frac{c-1}{z^{c-1}} e^z \int_0^z t^{c-2} e^{-t} dt = \frac{c-1}{z^{c-1}} e^z \gamma(c-1, z)$$

Il est alors facile d'en déduire que :

$$MTTF = \frac{(1 - \delta)\delta^{\frac{1}{\delta}-2}}{(1 - \theta)^{\frac{1-\delta}{\delta}}} e^{\frac{1-\theta}{\delta}} \gamma\left(\frac{1 - \delta}{\delta}, \frac{1 - \theta}{\delta}\right). \quad \square$$

Remarque :

Nous avons détaillé tout le calcul du MTTF car il semble que ce résultat soit nouveau, malgré l'ancienneté du schéma d'urne de Pólya (1923). Nous ne chercherons pas l'expression de la variance de cette loi en raison de la complexité du résultat obtenu pour le MTTF.

3.4.2 La loi de Pólya inverse DFR

Si on remet la boule blanche tirée et que l'on rajoute Δ boules blanches, on voit que le taux de défaillance est :

$$\lambda(k) = \frac{R}{R + B + (k - 1)\Delta} = \frac{\theta}{1 + (k - 1)\delta}, \quad 0 < \theta < 1, \quad 0 < \delta.$$

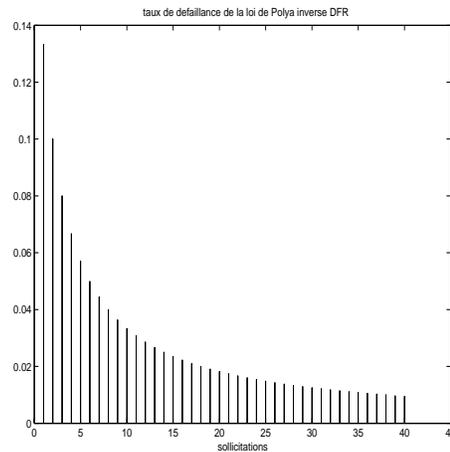


FIG. 3.11 – Taux de défaillance de la loi de Pólya inverse DFR

Par la formule (2.3), on peut donner l'expression de la probabilité élémentaire qui correspond à ce taux de défaillance :

$$p(k) = \frac{\theta \prod_{i=1}^{k-1} [1 - \theta + (i - 1)\delta]}{\prod_{i=1}^k [1 + (i - 1)\delta]} = \frac{\theta}{1 - \theta - \delta} \frac{\left(\frac{1-\theta-\delta}{\delta}\right)_{(k)}}{\left(\frac{1}{\delta}\right)_{(k)}}$$

L'expression de la fiabilité est :

$$R(k) = \frac{\prod_{i=1}^k [1 - \theta + (i - 1)\delta]}{\prod_{i=1}^k [1 + (i - 1)\delta]} = \frac{\left(\frac{1-\theta}{\delta}\right)_{(k)}}{\left(\frac{1}{\delta}\right)_{(k)}}$$

Le MTTF est fini pour $\theta > \delta$ et vaut :

$$MTTF = \sum_{k=0}^{+\infty} \frac{\left(\frac{1-\theta}{\delta}\right)_{(k)}}{\left(\frac{1}{\delta}\right)_{(k)}} = {}_2F_1\left(\frac{1-\theta}{\delta}, 1, \frac{1}{\delta}, 1\right) = \frac{1-\delta}{\theta-\delta}$$

où ${}_2F_1(a, b, c, z) = \sum_{k=0}^{+\infty} \frac{(a)_{(k)}(b)_{(k)} z^k}{(c)_{(k)} k!} = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{+\infty} \frac{\Gamma(a+k)\Gamma(b+k)}{\Gamma(c+k)} \frac{z^k}{k!}$ est une série hypergéométrique de Gauss.

Dans la perspective d'une estimation des paramètres de la loi par la méthode des moments, il est intéressant de donner le moment d'ordre 2 :

$$\mathbb{E}(K^2) = 2\theta {}_2F_1\left(\frac{1-\theta}{\delta}, 3, \frac{1}{\delta} + 1, 1\right) - MTTF = \frac{(1-\delta)(2-\theta-2\delta)}{(\theta-\delta)(\theta-2\delta)}$$

La loi de Pólya inverse DFR est aussi appelée loi de Waring et a été étudiée par Irwin [39]. Son taux de défaillance est dit "inverse linéaire", c'est à dire qu'il est de la forme $\lambda(k) = \frac{1}{a+bk}$. Le cas étudié ici correspond à $b > 0$. Xekalaki [94] a également considéré le cas $b < 0$. Dans ce cas, la loi est à support borné $\{1, \dots, m\}$ et son taux de défaillance est de la forme $\lambda(k) = \frac{1}{1+c(m-k)}$, $k \in \{1, \dots, m\}$, $c > 0$. Dans le cas particulier où $\frac{1}{c}$ est entier, les probabilités élémentaires peuvent s'écrire :

$$p(k) = \frac{C_{m+\frac{1}{c}-k-1}^{\frac{1}{c}-1}}{C_{m+\frac{1}{c}-1}^{\frac{1}{c}}}$$

et la loi de K est la loi hypergéométrique négative décalée.

Par ailleurs, il est facile de constater (voir Roy-Gupta [77]) qu'un taux de défaillance inverse linéaire est équivalent à une durée de vie résiduelle moyenne linéaire. Dans le cas de la loi de Pólya inverse DFR, on obtient, à condition que $\theta > \delta$:

$$m(k) = \frac{\delta(k-1) + 1}{\theta - \delta}$$

On retrouve comme cas particulier $MTTF = m(0) = \frac{1-\delta}{\theta-\delta}$.

Enfin, Gupta-Gupta-Tripathi [32] ont retrouvé la loi de Waring comme cas particulier de la famille de Katz étendue, définie par un rapport de probabilités successives de la forme :

$$\frac{p(k+1)}{p(k)} = \frac{\alpha + \beta k}{\gamma + k}, \text{ pour } \alpha = \frac{1-\theta-\delta}{\delta}, \beta = 1 \text{ et } \gamma = \frac{1}{\delta}$$

De la même façon pour la loi de Pólya inverse IFR, on a :

$$\frac{p(k+1)}{p(k)} = \frac{1-\theta}{\delta} \frac{\frac{\theta}{\delta} + k}{\left(\frac{\theta}{\delta} - 1 + k\right) \left(\frac{1}{\delta} + k\right)}$$

qui est un cas particulier de la famille généralisée de Kemp (voir p 86 de [42]).

Le fait d'avoir classé les lois de Pólya inverses dans des familles de lois plus larges n'apporte aucune information intéressante supplémentaire sur ces modèles.

3.4.3 Les lois de Salvia et Bollinger

En 1982, Salvia et Bollinger [80] ont proposé deux lois très simples à un seul paramètre. Nous les mentionnons ici car leurs généralisations sont des lois de Pólya inverses.

a) Taux de défaillance croissant (IFR)

Le taux de défaillance est $\lambda(k) = 1 - c/k$, avec $0 < c \leq 1$ et $k \in \mathbb{N}^*$. À l'aide de la formule (2.3), on a :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \left(1 - \frac{c}{k}\right) \prod_{i=1}^{k-1} \frac{c}{i} = (k-c) \frac{c^{k-1}}{k!}$$

L'expression de la fiabilité est donnée par :

$$\forall k \in \mathbb{N}^*, \quad R(k) = \frac{c^k}{k!}$$

L'expression du MTTF est donnée par :

$$MTTF = {}_1F_1(1, 1, c) = e^c$$

Remarque :

Cette loi n'a que peu d'intérêt en pratique car son espérance est comprise entre 1 et e . Les valeurs prises par K seront donc très souvent égales à 1 ou 2 et très rarement supérieures.

b) Taux de défaillance décroissant (DFR)

Le taux de défaillance est $\lambda(k) = c/k$, avec $0 < c \leq 1$ et $k \in \mathbb{N}^*$. À l'aide de la formule (2.3), on a :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \frac{c}{k} \prod_{i=1}^{k-1} \left(1 - \frac{c}{i}\right) = \frac{c}{k!} (1-c)_{(k-1)}$$

L'expression de la fiabilité est donnée par :

$$\forall k \in \mathbb{N}^*, \quad R(k) = \prod_{i=1}^k \left(\frac{i-c}{i}\right) = \frac{(1-c)_{(k)}}{k!}$$

Le MTTF n'existe pas car la condition $\theta > \delta = 1$ n'est pas vérifiée.

3.4.4 Les lois de Salvia et Bollinger généralisées

Les modèles précédents de Salvia et Bollinger ne sont pas suffisamment élaborés pour modéliser un grand nombre de situations. Padgett et Spurrier [66] ont proposé de généraliser ces modèles en ajoutant un second paramètre α . Quand $\alpha = 1$, les modèles se réduisent à ceux de Salvia et Bollinger. Quand $\alpha = 0$, le modèle est la loi géométrique.

a) Taux de défaillance croissant

Le taux de défaillance est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = 1 - \frac{c}{(k-1)\alpha + 1} \quad \alpha \geq 0, \quad 0 \leq c \leq 1.$$

b) Taux de défaillance décroissant

Le taux de défaillance est :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = \frac{c}{(k-1)\alpha + 1}, \quad \alpha \geq 0, \quad 0 \leq c \leq 1.$$

On voit que ces lois sont exactement des lois de Pólya inverses, ce qui n'avait pas été remarqué par Padgett et Spurrier.

3.4.5 La loi d'Engenberger-Pólya

Le Breton et Martin [52] se sont intéressés à la construction d'un modèle pour l'étude des séquences climatologiques. Leur modèle peut être utilisé en fiabilité car il a été choisi pour ses propriétés concernant la persistance, notion complémentaire à 1 au taux de défaillance. Le modèle est basé sur une loi de probabilité obtenue par Engenberger et Pólya comme limite de lois de Pólya quand on fait tendre le nombre de tirages vers l'infini (voir [41], pp 190-191).

La probabilité élémentaire de défaillance d'un appareil à la $k^{\text{ème}}$ sollicitation est :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \frac{1}{(1+d)^{h/d}} \cdot \frac{(h/d)_{(k-1)}}{(k-1)!} \cdot \left(\frac{d}{d+1} \right)^{k-1}$$

L'expression du taux de défaillance est obtenue par récurrence :

$$\lambda(k+1) = \frac{p(k+1)}{\sum_{j=k}^{+\infty} p(j) - p(k)}$$

$$\text{Or } p(k+1) = \frac{h + (k-1)d}{k(1+d)} p(k)$$

D'où

$$\lambda(k+1) = \frac{h + (k-1)d}{k(1+d)} \frac{\lambda(k)}{1 - \lambda(k)}$$

La résolution de cette relation de récurrence donne :

$$\forall k \in \mathbb{N}^*, \quad \frac{1}{\lambda(k)} = \sum_{j=0}^{+\infty} \frac{(h/d + k - 1)_{(j)}}{(k)_{(j)}} \left(\frac{d}{1+d} \right)^j = {}_2F_1 \left(\frac{h}{d} + k - 1, 1, k, \frac{d}{1+d} \right)$$

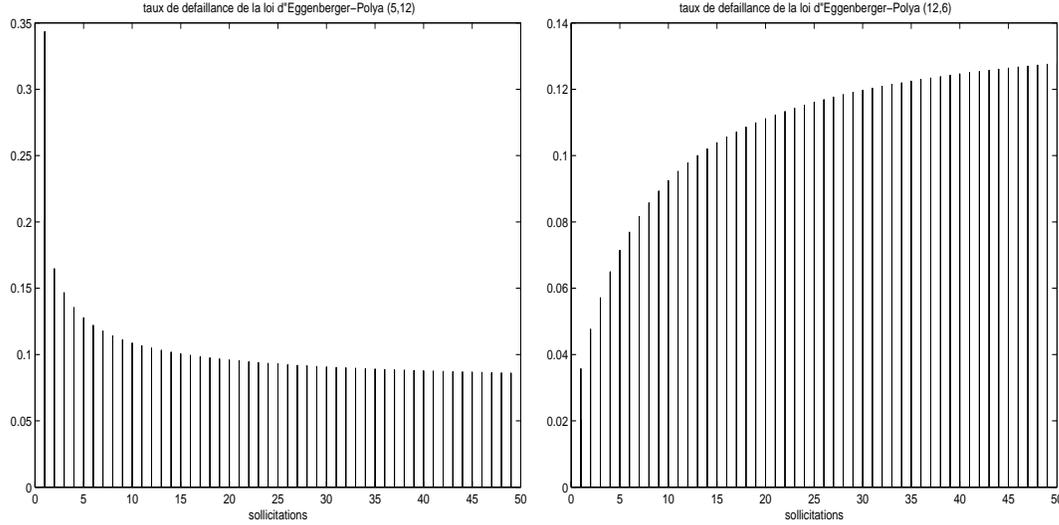


FIG. 3.12 – Taux de défaillance de la loi d'Eggenberger-Pólya

L'expression du taux de défaillance est donc très complexe, mais on peut déterminer facilement son sens de variation grâce au théorème 4 sur les rapports des probabilités élémentaires successives :

$$\frac{p(k+1)}{p(k)} = \frac{d}{1+d} + \frac{h-d}{1+d} \cdot \frac{1}{k}$$

1. Si $h = d$ alors le taux de défaillance est constant et la loi est géométrique de paramètre $\frac{1}{1+d}$,
2. Si $h < d$, la loi est log-convexe et donc DFR,
3. Si $h > d$, la loi est log-concave et donc IFR.

Dans les trois cas $\lim_{k \rightarrow \infty} \lambda(k) = \frac{1}{1+d}$.

À l'aide de la fonction génératrice de la loi, on obtient les expressions de l'espérance et la variance de la loi :

$$\begin{cases} \mathbb{E}(K) &= h + 1 \\ \text{Var}(K) &= h(d + 1) \end{cases}$$

Le paramètre h est le nombre moyen de sollicitations jusqu'à défaillance moins une. Le paramètre d n'a pas de signification tangible.

3.5 Conclusion

Ce chapitre révèle l'existence de nombreux modèles de durée de vie de systèmes non réparables en temps discret. Cependant, exception faite de la loi géométrique, ces modèles sont très peu connus dans le domaine de la fiabilité, et donc peu utilisés en pratique dans le milieu industriel. De cet état de l'art des modèles de fiabilité en temps discret, nous retiendrons pour leurs propriétés les modèles géométrique, Weibull de type I et d'Engenberger-Pólya. La loi géométrique est simple d'utilisation en raison de son unique paramètre, qui possède de plus une interprétation simple en terme de taux de défaillance. La loi de Weibull de type I est flexible puisque son taux de défaillance peut être DFR, CFR ou IFR selon la valeur de ses paramètres. Elle est l'analogue en temps discret de la loi de Weibull continue qui est très largement utilisée, à tel point que le paramètre de forme β est devenu une mesure du vieillissement dans la pratique. La loi d'Engenberger-Pólya est également flexible et ses deux premiers moments possèdent une expression simple qui facilitera l'estimation des paramètres. Son paramètre h s'interprète directement à l'aide du MTTF. Nous verrons au chapitre 6 que l'estimation paramétrique confirme cette sélection de modèles.

Nous ne retenons pas la loi de Weibull de type III en raison de son analogie avec celle de type I. Les lois basées sur le schéma d'urne de Pólya supposent que l'on sache a priori si le phénomène observé engendre un vieillissement ou un rajeunissement pour choisir le modèle adéquat, ce qui les rend moins flexibles que les lois de Weibull I et d'Engenberger-Pólya.

4 Vers une nouvelle définition du taux de défaillance en temps discret

Dans ce chapitre, nous mettons en évidence un certain nombre de différences importantes entre les notions de fiabilité usuelles en temps continu et leurs équivalents en temps discret. Nous montrons que ces différences sont dues à la définition du taux de défaillance en temps discret. Nous proposons une nouvelle définition qui résout les problèmes rencontrés. Avec cette nouvelle définition, les notions de fiabilité en temps discret sont nettement plus proches des notions correspondantes en temps continu qu'avec la définition usuelle. Enfin, nous donnons l'expression de r pour quelques modèles de fiabilité en temps discret.

4.1 Quelques problèmes avec la définition du taux de défaillance usuel

On rappelle la définition du taux de défaillance en temps discret :

$$\lambda(k) = \mathbb{P}(K = k | K \geq k) = \frac{\mathbb{P}(K = k)}{\mathbb{P}(K \geq k)} \quad (4.1)$$

et sa définition en temps continu :

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} \mathbb{P}(t < T \leq t + dt | T > t) \quad (4.2)$$

Le taux de défaillance en temps discret comporte de par sa définition, des contraintes et des différences importantes par rapport à sa définition en temps continu :

1. En temps discret, le taux de défaillance est défini comme une probabilité conditionnelle, alors que ce n'est pas le cas en temps continu. Ceci peut être une source de confusion entre le taux de défaillance et les probabilités élémentaires, d'autant plus que, même en temps continu, le taux de défaillance est souvent interprété à tort comme une probabilité.

2. En temps discret, le taux de défaillance est borné (par 1). Or la plupart des modèles de fiabilité en temps continu n'ont pas de taux de défaillance borné (lois de Weibull, gamma, log-normale, ...). Donc ces modèles ne peuvent pas avoir d'analogues directs en temps discret.
3. Puisqu'il est borné, le taux de défaillance en temps discret ne peut pas prendre n'importe quelle forme. Par exemple, il ne peut pas croître linéairement ni exponentiellement. Plus largement, il ne peut être convexe. Par conséquent, un problème d'interprétation se pose car, en temps continu, un taux de défaillance convexe est usuellement interprété comme caractéristique d'une accélération du vieillissement en fin de vie des systèmes (fin de la courbe en "baignoire").

En temps continu, le modèle le plus fréquemment adopté pour ce cas de figure est la loi de Weibull avec $\beta > 2$. En temps discret, il ne peut y avoir de modèle équivalent. La figure 4.1 montre que le taux de défaillance de la loi de Weibull discrète de type I avec $\beta > 2$ commence par être convexe, mais devient rapidement concave :

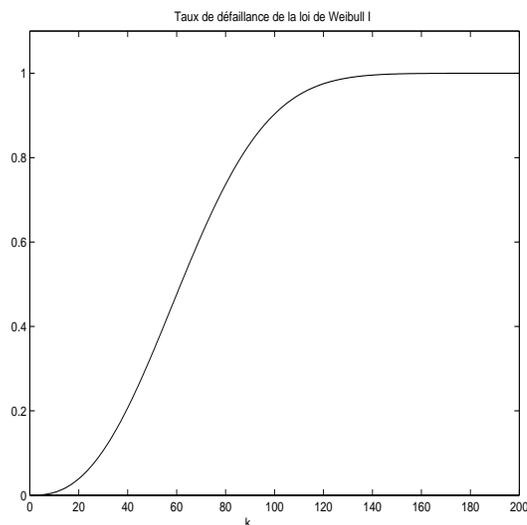


FIG. 4.1 – Taux de défaillance de la loi de Weibull de type I de paramètre $\beta = 3.5$

4. Pour un système série, le taux de défaillance en temps discret n'est pas additif. En effet, considérons un système constitué de n composants indépendants en série de taux de défaillance et fiabilité respectifs $\lambda_i, R_i, i = 1, \dots, n$. La fiabilité du système est usuellement :

$$R(k) = \prod_{i=1}^n R_i(k)$$

Le taux de défaillance du système est :

$$\begin{aligned}\lambda(k) &= \frac{R(k-1) - R(k)}{R(k-1)} = \frac{\prod_{i=1}^n R_i(k-1) - \prod_{i=1}^n R_i(k)}{\prod_{i=1}^n R_i(k-1)} = 1 - \prod_{i=1}^n \frac{R_i(k)}{R_i(k-1)} \\ &= 1 - \prod_{i=1}^n (1 - \lambda_i(k)) \neq \sum_{i=1}^n \lambda_i(k)\end{aligned}$$

Par conséquent, le taux de défaillance d'un système de n composants indépendants en série n'est pas égal à la somme des taux de défaillance des composants. Il est ennuyeux qu'une des propriétés les plus connues et utilisées du taux de défaillance en temps continu ne soit pas valable en temps discret.

5. Comme on l'a déjà remarqué au chapitre 2, la relation usuelle entre la fiabilité et le taux de défaillance cumulé en temps continu, $\Lambda(t) = -\ln R(t)$, n'est pas valable en temps discret : $\Lambda(k) \neq H(k) = -\ln R(k)$.
6. Comme on l'a vu au chapitre 2, il existe en temps discret deux définitions non équivalentes de la notion de vieillissement IFRA :

$$\begin{aligned}\text{IFRA1: } & \left\{ R(k)^{\frac{1}{k}} \right\}_{k \geq 1} \text{ est une suite décroissante,} \\ \text{IFRA2: } & \left\{ \frac{\Lambda(k)}{k} \right\}_{k \geq 1} \text{ est une suite croissante.}\end{aligned}$$

La non équivalence entre les deux notions est liée au fait que $H(k) \neq \Lambda(k)$, donc le point 6 est en fait une conséquence du point 5. Le problème est similaire pour la notion de vieillissement NBU qui possède également deux définitions en temps discret.

En conclusion, plusieurs propriétés de base des taux de défaillance et taux de défaillance cumulé en temps continu ne sont plus vraies en temps discret. Il nous est donc venu à l'esprit que la définition (4.1) du taux de défaillance en temps discret, proposée par Barlow-Marshall-Proschan [8] pourrait ne pas être appropriée, et que l'on pourrait définir autrement ce taux de défaillance de façon à éviter les problèmes rencontrés.

4.2 Une nouvelle définition du taux de défaillance

Le problème énoncé au point 5 ci-dessus peut être résolu si le taux de défaillance usuel λ est remplacé par une nouvelle définition du taux de défaillance, notée r , telle que $-\ln R(k) = \sum_{i=1}^k r(i)$, c'est à dire :

$$r(k) = -\ln R(k) + \ln R(k-1) = -\ln \frac{R(k)}{R(k-1)}$$

Une nouvelle définition du taux de défaillance en temps discret est donc :

Définition 3

Pour une variable aléatoire à valeurs dans \mathbb{N}^* définie par sa fonction de fiabilité $R(k)$, on définit le nouveau taux de défaillance $r(k)$ par :

$$\forall k \in \mathbb{N}^*, \quad r(k) = \ln \frac{R(k-1)}{R(k)} \quad (4.3)$$

On peut comprendre $\lambda(k) = \frac{p(k)}{R(k-1)}$ comme l'analogie en temps discret de la relation $\lambda(t) = \frac{f(t)}{R(t)}$ en temps continu, tandis que $r(k) = -\ln \frac{R(k)}{R(k-1)}$ serait l'analogie de $\lambda(t) = -\frac{d}{dt} \ln R(t)$.

On vérifie alors que cette nouvelle définition résout les problèmes énoncés à la section 4.1 :

1. Contrairement au taux de défaillance usuel en temps discret, le nouveau taux de défaillance r n'est pas une probabilité conditionnelle et donc aucune confusion n'est possible avec les probabilités élémentaires.
2. r n'est pas borné contrairement à λ .
3. r peut être convexe comme le montre le graphe suivant de r (et λ) pour la loi de Weibull de type I avec $\beta > 2$:

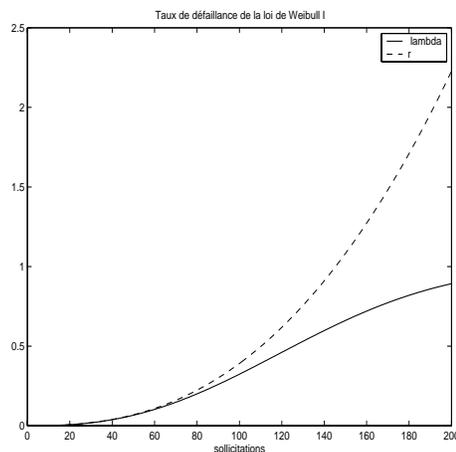


FIG. 4.2 – Comparaison des taux de défaillance en temps discret

4. r vérifie la propriété d'additivité pour un système série. En effet, avec les notations de la section 4.1, le nouveau taux de défaillance d'un système constitué de n composants

indépendants en série est :

$$\begin{aligned} r(k) &= \ln \frac{R(k-1)}{R(k)} = \ln \frac{\prod_{i=1}^n R_i(k-1)}{\prod_{i=1}^n R_i(k)} = \ln \prod_{i=1}^n \frac{R_i(k-1)}{R_i(k)} \\ &= \sum_{i=1}^n \ln \frac{R_i(k-1)}{R_i(k)} = \sum_{i=1}^n r_i(k) \end{aligned}$$

r est donc bien égal à la somme des nouveaux taux de défaillance des composants.

5. Par définition de r , la relation entre la fiabilité et le taux de défaillance cumulé est valable en temps discret puisque $H(k) = -\ln R(k) = \sum_{i=1}^k r(i)$.

Ainsi, l'appellation de taux de défaillance cumulé pour H n'est plus inappropriée comme c'était le cas pour le taux de défaillance usuel.

6. Nous proposons les définitions suivantes de la notion IFRA en temps discret correspondant aux définitions équivalentes en temps continu :

- IFRA1 : $\left\{ R(k)^{\frac{1}{k}} \right\}_{k \geq 1}$ est une suite décroissante.
- IFRA2 : $\left\{ \frac{H(k)}{k} \right\}_{k \geq 1}$ est une suite croissante.

En remplaçant Λ par H dans la définition de la notion IFRA2, nous obtenons ainsi l'équivalence entre les notions IFRA1 et IFRA2 en temps discret puisque $H(k) = -\ln R(k)$.

Nous proposons également les définitions suivantes de la notion NBU en temps discret correspondant aux définitions équivalentes en temps continu :

- NBU1 : $\forall k \in \mathbb{N}^*, \forall j \in \mathbb{N}^*, R(j|k) \leq R(j)$,
- NBU2 : $\forall k \in \mathbb{N}^*, \forall j \in \mathbb{N}^*, H(k+j) \geq H(k) + H(j)$.

Comme pour le cas IFRA, la relation $H(k) = -\ln R(k)$ permet de montrer l'équivalence en temps discret entre les notions NBU1 et NBU2.

4.3 Relation avec les autres grandeurs de la fiabilité

4.3.1 Relation avec le taux de défaillance usuel

La propriété suivante donne la relation entre les taux de défaillance $r(k)$ et $\lambda(k)$:

Propriété 5

$$r(k) = -\ln[1 - \lambda(k)] \quad \text{ou} \quad \lambda(k) = 1 - e^{-r(k)} \quad (4.4)$$

Preuve :

Par définition du nouveau taux de défaillance :

$$\begin{aligned} r(k) &= -\ln \frac{R(k)}{R(k-1)} \\ &= -\ln \left[1 - \frac{R(k-1) - R(k)}{R(k-1)} \right] \\ &= -\ln[1 - \lambda(k)]. \quad \square \end{aligned}$$

La relation (4.4) entraîne en particulier que r et λ ont même sens de variation. Ainsi, on peut rajouter aux quatre définitions de la notion de vieillissement IFR (voir 2.5.2) une nouvelle définition équivalente :

- IFR5 : $\{r(k)\}_{k \geq 1}$ est une suite croissante.

La relation (4.4) permet également d'écrire :

$$r(k) = -\ln[1 - \lambda(k)] = \lambda(k) + o(\lambda(k))$$

C'est à dire que $r(k)$ et $\lambda(k)$ sont équivalents quand $\lambda(k)$ est proche de 0 (ce qui est souvent le cas pour des appareils fiables), puis les taux de défaillance s'éloignent pour les grandes valeurs de k comme on peut le constater sur le graphe 4.2.

Dans [80], Salvia et Bollinger ont proposé des approximations des grandeurs de la fiabilité à partir d'hypothèses effectuées sur le taux de défaillance. Pour les petites valeurs du taux de défaillance $\lambda(k)$, on utilise l'approximation $1 - \lambda(k) \simeq \exp(-\lambda(k))$. Les formules (2.2) et (2.3) deviennent :

$$R(k) \simeq \exp[-\Lambda(k)] \text{ et } p(k) \simeq \lambda(k)\exp[-\Lambda(k-1)]$$

On voit que la démarche des auteurs revient en fait à remplacer Λ par H . Dans ce cas les approximations deviennent des égalités.

4.3.2 Cas des modèles issus du temps continu

À la section 3.3.2, nous nous sommes intéressés à la construction de la v.a. entière K à partir de la v.a. réelle positive T ($K = \lfloor T \rfloor + 1$) et ainsi à la définition de λ_K à partir de λ_T . Il est alors logique d'introduire l'analogie pour le nouveau taux de défaillance, c'est à dire r_K .

On rappelle la relation valable pour le taux de défaillance usuel :

$$\lambda_K(k) = 1 - e^{\int_{k-1}^k \lambda_T(u) du} = 1 - e^{[\Lambda_T(k) - \Lambda_T(k-1)]}$$

En utilisant la relation (4.4), on obtient :

$$r_K(k) = \int_{k-1}^k \lambda_T(u) du = \Lambda_T(k) - \Lambda_T(k-1)$$

4.3.3 Relation avec les autres grandeurs de la fiabilité

Propriété 6

On a les relations suivantes entre les grandeurs de la fiabilité et le nouveau taux de défaillance :

$$\begin{aligned}
 p(k) &= [1 - e^{-r(k)}] \exp \left\{ - \sum_{i=1}^{k-1} r(i) \right\} \\
 R(k) &= \exp \left\{ - \sum_{i=1}^k r(i) \right\} \\
 H(k) &= -\ln R(k) = \sum_{i=1}^k r(i) \\
 m(k) &= 1 + \sum_{j=k+1}^{\infty} \exp \left\{ - \sum_{i=k+1}^j r(i) \right\}
 \end{aligned}$$

4.4 Applications aux modèles de fiabilité en temps discret

Nous donnons les expressions du nouveau taux de défaillance pour les lois géométrique et Weibull de type I car ce sont les lois les plus faciles à utiliser en temps discret. Nous étudions également la loi de Weibull de type III pour laquelle nous avons obtenu un résultat intéressant puisque le nouveau taux de défaillance de cette loi est en fait égal au taux de défaillance usuel de la loi de Weibull de type II.

4.4.1 Loi géométrique

On rappelle que pour la loi géométrique $\mathcal{G}(p)$, on a : $\forall k \in \mathbb{N}^*$, $\lambda(k) = p$. À partir de l'expression de (4.4), l'expression du nouveau taux de défaillance pour la loi géométrique $\mathcal{G}(p)$ est :

$$r(k) = -\ln(1 - p)$$

On peut remarquer que le nouveau taux de défaillance de la loi géométrique ne conserve pas la propriété du taux de défaillance usuel qui est d'être égal au paramètre de la loi géométrique.

Si on considère maintenant la loi géométrique comme la discrétisation de la loi exponentielle $\mathcal{E}(\lambda)$, on a vu au 3.3.2 que $\lambda_K(k) = 1 - e^{-\lambda}$. Donc le nouveau taux de défaillance est :

$$r(k) = -\ln [1 - (1 - e^{-\lambda})] = \lambda = \lambda_T(k), \text{ alors que } \lambda_K(k) \neq \lambda_T(k)$$

Donc si on discrétise une loi exponentielle, le nouveau taux de défaillance r de la loi géométrique obtenu est égal au taux de défaillance de la loi exponentielle initiale, ce qui n'était pas le cas du taux de défaillance usuel λ .

4.4.2 Lois de Weibull

Pour la loi de Weibull de type I, $\lambda(k) = 1 - q^{k^\beta - (k-1)^\beta}$, donc

$$r(k) = -\ln \left[\frac{q^{(k-1)^\beta}}{q^{k^\beta}} \right] = [(k-1)^\beta - k^\beta] \ln q \quad (4.5)$$

Si on considère la loi de Weibull de type I comme la discrétisation de la loi de Weibull $\mathcal{W}(\eta, \beta)$, on a vu au 3.3.2 que $\lambda_K(k) = 1 - e^{-\left(\frac{k}{\eta}\right)^\beta + \left(\frac{k-1}{\eta}\right)^\beta} = 1 - q^{k^\beta - (k-1)^\beta}$ avec $q = e^{-\frac{1}{\eta^\beta}}$.

Remarque :

Bien que le taux de défaillance défini par (4.5) puisse croître linéairement pour $\beta = 2$ comme dans le cas continu, son expression ne coïncide pas avec celle du taux de défaillance de la loi de Weibull en temps continu :

Soit T de loi de Weibull $\mathcal{W}(1, 2)$. Alors $\lambda_T(t) = 2t$, $\lambda_K(k) = 1 - e^{-(2k-1)}$ et $r_K(k) = k^2 - (k-1)^2 = 2k - 1$.

Soit T de loi de Weibull $\mathcal{W}(1, 3)$. Alors $\lambda_T(t) = 3t^2$, $\lambda_K(k) = 1 - e^{-(3k^2-3k+1)}$ et $r_K(k) = k^3 - (k-1)^3 = 3k^2 - 3k + 1 = \lambda_T(k) - 3k + 1$.

On constate donc qu'il y a nettement plus de similitude entre λ_T et r_K qu'entre λ_T et λ_K .

La loi de Weibull de type III est définie par $\lambda(k) = 1 - e^{-ck^\beta}$. On a donc immédiatement $r(k) = ck^\beta$. On constate donc que l'expression du nouveau taux de défaillance pour la loi de Weibull de type III est, d'une part la même que celle du taux de défaillance usuel pour la loi de Weibull de type II (mis à part le support), et d'autre part la même que celle du taux de défaillance de la loi de Weibull en temps continu.

Par conséquent, si jusqu'à maintenant on a considéré que l'équivalent en temps discret de la loi de Weibull continue était la loi de Weibull de type I, cette propriété nous amène à considérer que cet équivalent pourrait être la loi de Weibull de type III.

4.5 Conclusion

L'introduction d'une nouvelle définition du taux de défaillance en temps discret a permis de remédier aux différences énoncées en début de ce chapitre entre le temps discret et le temps continu. La nouvelle définition a été introduite à des fins purement mathématiques et son inconvénient majeur est qu'elle n'a pas d'interprétation probabiliste simple comme λ . En effet, $\lambda(k)$ s'interprète facilement comme la probabilité conditionnelle de défaillance d'un système à la $k^{\text{ème}}$ sollicitation sachant qu'il a correctement fonctionné au cours des $k-1$ premières sollicitations. L'introduction de la nouvelle définition a néanmoins comme conséquence de renforcer les similitudes entre la fiabilité en temps discret et la fiabilité en temps continu. Ce point est important pour le développement en pratique de l'utilisation de

la fiabilité dans un contexte discret car ce développement ne saurait se faire sans analogie directe avec la fiabilité en temps continu.

Estimation non paramétrique

L'estimation non paramétrique permet l'analyse et l'interprétation de données sans utiliser d'hypothèses sur le phénomène étudié, l'utilisation d'un modèle inadéquat pouvant conduire à des résultats erronés. Par exemple, dans le domaine de la fiabilité en temps discret, l'utilisation de la loi géométrique est inappropriée pour l'étude de données qui traduisent un vieillissement. L'analyse non paramétrique est généralement utilisée comme une étape intermédiaire avant une analyse paramétrique qui permet d'utiliser des procédures de statistique inférentielle plus précises et plus poussées. Encore faut-il vérifier que les hypothèses que requiert l'utilisation du modèle soient bien validées pour les données étudiées.

Nous nous intéressons aux estimateurs non paramétriques des grandeurs usuelles de la fiabilité (fonction de répartition, taux de défaillance et durée de vie résiduelle moyenne) pour des données discrètes et non censurées. Nous rappellerons les résultats d'estimation, ponctuelle, par intervalle et bande de confiance, de la fiabilité, qui nous serviront ensuite pour étudier le taux de défaillance (usuel) empirique. Nous donnerons également les expressions des estimateurs empiriques des autres grandeurs de la fiabilité (le nouveau taux de défaillance et la durée de vie résiduelle moyenne), leur complexité ne nous ayant pas permis une étude complète comme pour la fiabilité et le taux de défaillance empiriques.

5.1 Notations

Soit k_1, \dots, k_n des observations qu'on supposera être des réalisations de variables aléatoires K_1, \dots, K_n indépendantes et de même loi à valeurs dans \mathbb{N}^* , représentant par exemple des dates de défaillance. On notera par $k_{(i)}$ la $i^{\text{ème}}$ observation de l'échantillon ordonné dans l'ordre croissant. Si des répétitions existent, on notera par $k_{(1)}^* < k_{(2)}^* < \dots < k_{(l)}^*$, $l \leq n$, les instants de défaillance distincts. Soient d_j le nombre de défaillances à la date $k_{(j)}^*$ et r_j le nombre cumulé de défaillances entre l'instant initial et $k_{(j)}^*$:

$$r_j = \max\{i; k_{(i)} = k_{(j)}^*\} = \sum_{k=1}^j d_k$$

On notera par n_j le nombre d'appareils encore en fonctionnement à l'instant $k_{(j)}^*$: $n_j = n - r_j$, avec $n_0 = n$.

Exemple :

Nous utilisons un exemple cité par Xie et Goth [95] portant sur des données issues d'un contrôle de qualité. Cet exemple sera utilisé tout au long de ce chapitre.

La variable aléatoire discrète décrit, dans un processus industriel, le nombre d'appareils inspectés avant la découverte d'un appareil défectueux.

L'échantillon comporte $n = 28$ données au total dont $l = 14$ valeurs distinctes :

k_j	13	1	1	17	5	3	2	3	2	2	29	14	1	1	5	3	2
$k_{(j)}$	1	1	1	1	1	1	2	2	2	2	3	3	3	4	4	4	5
k_j	4	18	11	1	1	14	26	9	4	7	4						
$k_{(j)}$	5	7	9	11	13	14	14	17	18	26	29						

Avec les notations introduites ci-dessus, on obtient :

$k_{(j)}^*$	1	2	3	4	5	7	9	11	13	14	17	18	26	29
d_j	6	4	3	3	2	1	1	1	1	2	1	1	1	1
r_j	6	10	13	16	18	19	20	21	22	24	25	26	27	28
n_j	22	18	15	12	10	9	8	7	6	5	3	2	1	0

5.2 La fiabilité empirique

On définit les fréquences empiriques $\mathbb{P}_n(k)$, pour l'échantillon k_1, \dots, k_n par :

$$\begin{aligned} \forall k \in \mathbb{N}^*, \quad \mathbb{P}_n(k) &= \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{K_i=k\}} \\ &= \frac{\text{nombre d'observations} = k}{\text{nombre total d'observations}} \end{aligned}$$

La fonction de répartition empirique de l'échantillon k_1, \dots, k_n est définie par :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{F}_n(k) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{K_i \leq k\}}$$

Dans un contexte de fiabilité, il est plus intéressant de considérer directement la fiabilité empirique, définie par :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{R}_n(k) = 1 - \mathbb{F}_n(k) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{K_i > k\}}$$

L'estimateur $\mathbb{R}_n(k)$ correspond au pourcentage d'observations strictement supérieures à k , ou le pourcentage d'appareils encore en fonctionnement à l'instant k . La fiabilité empirique

est une suite décroissante, valant 1 avant la plus petite observation, diminuant de d_j/n en chaque $k_{(j)}^*$ et valant 0 pour la plus grande des observations. Elle est entièrement déterminée par ses valeurs en chaque point de l'échantillon :

$$\forall j \in \{1, \dots, l\}, \quad \mathbb{R}_n(k_{(j)}^*) = 1 - \frac{r_j}{n} = \frac{n_j}{n}$$

Exemple :

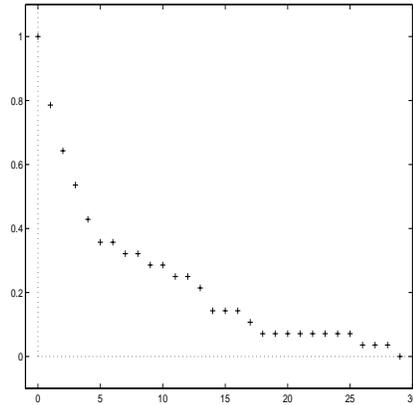


FIG. 5.1 – Graphe de la fiabilité empirique pour les données du contrôle de qualité

Le graphe met en évidence une décroissance rapide de la fiabilité empirique qui s'explique par le nombre élevé de faibles valeurs dans l'échantillon.

Nous rappelons maintenant les propriétés de base de la fiabilité empirique qui seront utilisées dans la suite du document. En fait, les propriétés de la fiabilité empirique sont directement transposées de celles de la fonction de répartition empirique, données par exemple par Billingsley [12] et Serfling [83]. Ces propriétés sont valables pour des observations continues comme pour des observations discrètes.

Propriété 7

$$\forall k \in \mathbb{N}^* \quad , \quad n\mathbb{R}_n(k) \text{ suit la loi binomiale } \mathcal{B}(n, R(k)), \quad (5.1)$$

$$\forall k \in \mathbb{N}^* \quad , \quad \mathbb{E}(\mathbb{R}_n(k)) = R(k) \text{ et } \text{Var}(\mathbb{R}_n(k)) = \frac{R(k)(1 - R(k))}{n}, \quad (5.2)$$

$$\forall (i, j) \in \mathbb{N}^{*2} \quad , \quad \text{Cov}(\mathbb{R}_n(i), \mathbb{R}_n(j)) = \frac{1}{n}[R(\max(i, j)) - R(i)R(j)]. \quad (5.3)$$

La formule (5.3) est obtenue à partir du résultat connu pour la fonction de répartition : $\text{Cov}(\mathbb{F}_n(i), \mathbb{F}_n(j)) = \frac{1}{n}[F(\min(i, j)) - F(i)F(j)]$ et en montrant que :

$$\begin{aligned} F(\min(i, j)) - F(i)F(j) &= 1 - R(\min(i, j)) - (1 - R(i))(1 - R(j)) \\ &= -R(\min(i, j)) + R(i) + R(j) - R(i)R(j) \\ &= R(\max(i, j)) - R(i)R(j). \end{aligned}$$

Les propriétés de convergence de la fiabilité empirique sont données par :

Propriété 8

$$\forall k \in \mathbb{N}^* \quad , \quad \mathbb{R}_n(k) \xrightarrow{p.s.} R(k) \text{ quand } n \rightarrow +\infty, \quad (5.4)$$

$$\forall k \in \mathbb{N}^* \quad , \quad \sqrt{n}(\mathbb{R}_n(k) - R(k)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, R(k)(1 - R(k))), \quad (5.5)$$

$$\sup_{k \in \mathbb{N}^*} |\mathbb{R}_n(k) - R(k)| \xrightarrow{p.s.} 0, \quad (5.6)$$

$\forall m \in \mathbb{N}^*$, et pour toute famille $x_1 < x_2 < \dots < x_m$ de m entiers classés par ordre croissant, le vecteur aléatoire $(\sqrt{n}(\mathbb{R}_n(x_i) - R(x_i)), 1 \leq i \leq m)^t$ converge en distribution quand $n \rightarrow \infty$ vers un vecteur aléatoire gaussien de \mathbb{R}^m , centré et de matrice de variance Γ définie par :

$$\begin{cases} \Gamma(x_i, x_i) = R(x_i)[1 - R(x_i)] \text{ pour } 1 \leq i \leq m; \\ \Gamma(x_i, x_j) = R(\max(x_i, x_j)) - R(x_i)R(x_j) \text{ si } i \neq j. \end{cases} \quad (5.7)$$

La formule (5.4) énonce la convergence simple de la fiabilité empirique vers la vraie fiabilité. Le résultat (5.6), démontré par Glivenko (1933) pour les lois continues et par Cantelli (1933) pour des lois quelconques, énonce que la convergence est uniforme. La formule (5.5) résulte de l'approximation de la loi binomiale par la loi normale. La formule (5.7) assure notamment la convergence des combinaisons linéaires des $\sqrt{n}(\mathbb{R}_n(x_i) - R(x_i))$ vers une variable normale.

Grâce aux propriétés données ci-dessus, on obtient la propriété suivante :

Propriété 9

$\forall k \in \mathbb{N}^*$, $\mathbb{R}_n(k)$ est un estimateur sans biais, convergent uniformément et presque sûrement de $R(k)$.

Les résultats sur les intervalles de confiance pour le paramètre p de la loi binomiale $\mathcal{B}(n, p)$ permettent de construire un intervalle de confiance exact $[R_{inf}(k), R_{sup}(k)]$, de seuil $\alpha \in]0, 1[$ pour $R(k)$:

$$R_{inf}(k) = \left\{ 1 + \frac{(n - n\mathbb{R}_n(k) + 1)\mathcal{F}(1 - \alpha/2; 2n - 2n\mathbb{R}_n(k) + 2, 2n\mathbb{R}_n(k))}{n\mathbb{R}_n(k)} \right\}^{-1}$$

$$R_{sup}(k) = \left\{ 1 + \frac{n - n\mathbb{R}_n(k)}{(n\mathbb{R}_n(k) + 1)\mathcal{F}(1 - \alpha/2; 2n\mathbb{R}_n(k) + 2, 2n - 2n\mathbb{R}_n(k))} \right\}^{-1}$$

où $\mathcal{F}_{(p; \nu_1, \nu_2)}$ est le p quantile de la loi de Fischer-Snedecor à (ν_1, ν_2) degrés de liberté.

Il est également possible de construire un intervalle de confiance asymptotique basé sur l'approximation normale de la loi binomiale. En utilisant la formule (5.5) de la propriété 8,

on peut écrire que

$$\sqrt{n} \frac{\mathbb{R}_n(k) - R(k)}{\sqrt{R(k)(1-R(k))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

On utilise le théorème de Slutsky (p. 68 de [18]) :

Théorème 7

Soient $\{X(k)\}_{k \geq 1}$ une suite de v.a. qui converge en loi et $\{Y(k)\}_{k \geq 1}$ une suite de v.a. qui converge en probabilité vers une constante c . Alors, si g est une fonction continue, les suites $\{g(X(k), Y(k))\}_{k \geq 1}$ et $\{g(X(k), c)\}_{k \geq 1}$ ont même loi limite.

Le théorème de Slutsky et la convergence presque sûre de $\mathbb{R}_n(k)$ vers $R(k)$ assurent alors que :

$$Z_n(k) = \sqrt{n} \frac{\mathbb{R}_n(k) - R(k)}{\sqrt{\mathbb{R}_n(k)(1 - \mathbb{R}_n(k))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (5.8)$$

On obtient alors un intervalle de confiance asymptotique $[R'_{inf}(k), R'_{sup}(k)]$, de seuil $\alpha \in]0, 1[$ pour $R(k)$ donné par :

$$\begin{aligned} R'_{inf}(k) &= \mathbb{R}_n(k) - z_\alpha \sqrt{\frac{\mathbb{R}_n(k)(1 - \mathbb{R}_n(k))}{n}} \\ R'_{sup}(k) &= \mathbb{R}_n(k) + z_\alpha \sqrt{\frac{\mathbb{R}_n(k)(1 - \mathbb{R}_n(k))}{n}} \end{aligned}$$

Pour les échantillons de petite taille, il est possible que $R'_{inf}(k) < 0$ ou $R'_{sup}(k) > 1$. Pour éviter ce problème, une autre approximation normale peut être obtenue avec la transformation du logit ([54]) :

$$\text{logit}(p) = \ln \frac{p}{1-p}, \text{ pour } p \in]0, 1[.$$

On utilise le résultat de convergence suivant :

$$\sqrt{n} (\text{logit}(\mathbb{R}_n(k)) - \text{logit}(R(k))) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{R(k)(1-R(k))}\right)$$

qui résulte de la relation (5.5) et de l'utilisation de la propriété suivante (méthode delta) : si $\sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$ alors $\sqrt{n}(\varphi(X_n) - \varphi(\mu)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \varphi'^2(\mu))$.

La convergence presque sûre de $\mathbb{R}_n(k)$ vers $R(k)$ et le théorème de Slutsky permettent d'écrire que :

$$\sqrt{n \mathbb{R}_n(k)(1 - \mathbb{R}_n(k))} (\text{logit}(\mathbb{R}_n(k)) - \text{logit}(R(k))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Un intervalle de confiance asymptotique $[R''_{inf}(k), R''_{sup}(k)]$, de seuil $\alpha \in]0, 1[$ pour $R(k)$ est alors donné par :

$$\begin{aligned} R''_{inf}(k) &= \frac{\mathbb{R}_n(k)}{\mathbb{R}_n(k) + (1 - \mathbb{R}_n(k))w_\alpha(k)} \\ R''_{sup}(k) &= \frac{\mathbb{R}_n(k)}{\mathbb{R}_n(k) + (1 - \mathbb{R}_n(k))/w_\alpha(k)} \end{aligned}$$

$$\text{où } w_\alpha(k) = \exp \left\{ \frac{z_\alpha}{\sqrt{n\mathbb{R}_n(k)(1 - \mathbb{R}_n(k))}} \right\}.$$

Exemple :

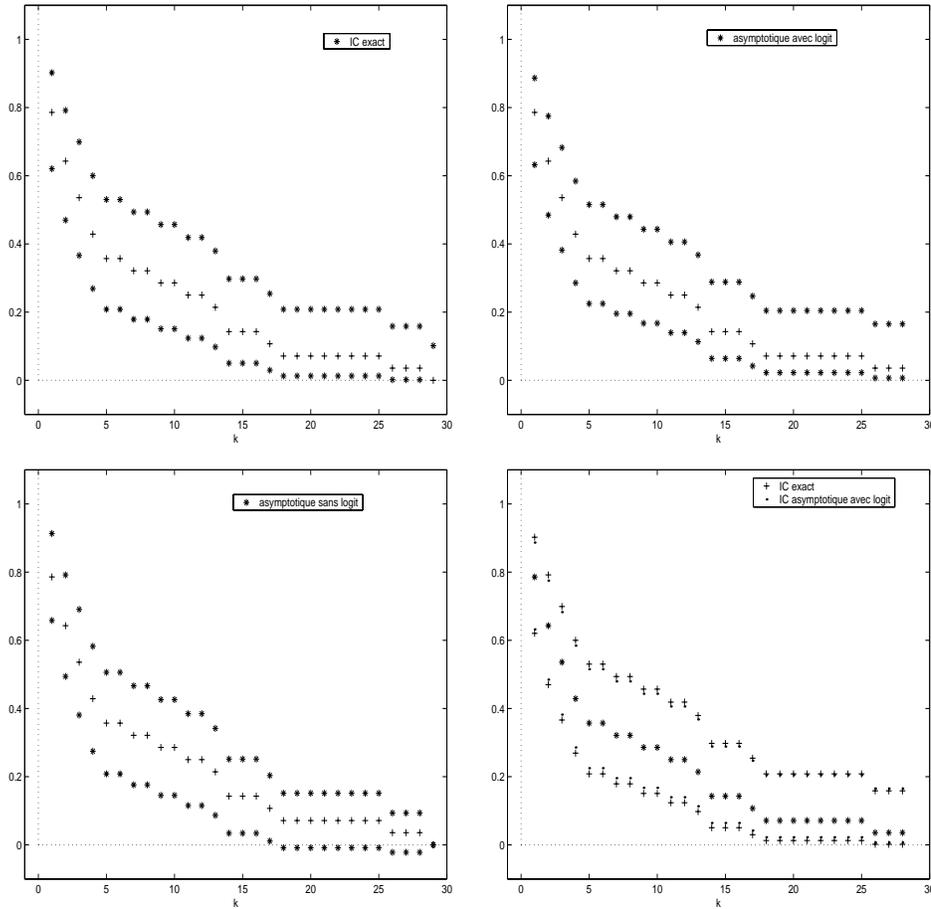


FIG. 5.2 – Intervalles de confiance exact et asymptotiques, avec ou sans logit, de seuil $\alpha = 0.1$ pour la fiabilité empirique avec les données du contrôle de qualité

On peut se rendre compte que les trois intervalles de confiance diffèrent peu pour les petites valeurs de k . L'approximation normale sans logit peut conduire à donner des bornes de l'intervalle de confiance de la fiabilité qui soient en dehors de l'intervalle $[0, 1]$. On remarquera que l'intervalle de confiance avec logit possède l'inconvénient d'être toujours plus large que les autres intervalles de confiance sans logit. La comparaison des intervalles de confiance exact et avec le logit montre qu'ils diffèrent très peu, quelle que soit la valeur de k . Les intervalles de confiance exacts sont légèrement plus larges que les intervalles de confiance asymptotiques avec logit.

L'intervalle de confiance permet d'obtenir de l'information sur R en une valeur ponctuelle k . Cependant, il est parfois nécessaire de quantifier l'incertitude sur l'estimation de R pour

tout un intervalle de valeurs $[a, b]$.

Définition 4

Une bande de confiance de seuil α ($\alpha \in [0, 1]$) pour la fonction R sur l'intervalle $[a, b]$ est définie par les fonctions $L(k)$ et $H(k)$ données par :

$$\mathbb{P}(L(k) \leq R(k) \leq H(k), \quad \forall k \in [a, b]) = 1 - \alpha$$

Par définition, à un instant donné, une bande de confiance est plus large qu'un intervalle de confiance.

La construction d'une bande de confiance asymptotique pour R repose sur la loi asymptotique de la variable aléatoire

$$Z_{a,b}^n = \max_{\{a \leq k \leq b\}} Z_n(k)$$

où $Z_n(k)$ est définie par (5.8). Les variables aléatoires $Z_n(k)$ sont asymptotiquement de loi normale centrée réduite, mais elles ne sont pas indépendantes comme le rappelle la formule (5.7) de la propriété 8. Il s'agit alors de calculer la loi du maximum des composantes d'un vecteur gaussien dont on connaît la matrice de covariance. Plus simplement, on peut se contenter d'utiliser une bande de confiance asymptotique construite à partir de la statistique de Kolmogorov-Smirnov pour une fonction de répartition continue. Si on dispose d'un échantillon d'une loi continue de fiabilité $R(t)$ et que l'on note $\mathbb{R}_n(t)$ la fiabilité empirique, on a :

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\sqrt{n} \sup_{t \in \mathbb{R}^+} |\mathbb{R}_n(t) - R(t)| \geq c \right) = 2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 c^2}$$

Une bande de confiance asymptotique de Kolmogorov-Smirnov $[R_{inf}(t), R_{sup}(t)]$, de seuil $\alpha \in]0, 1[$ pour R est alors donnée par :

$$\begin{aligned} R_{inf}(t) &= \mathbb{R}_n(t) - \frac{e_{1-\alpha}}{\sqrt{n}} \\ R_{sup}(t) &= \mathbb{R}_n(t) + \frac{e_{1-\alpha}}{\sqrt{n}} \end{aligned}$$

où $e_{1-\alpha}$ est tel que $2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 e_{1-\alpha}^2} = \alpha$.

Dans [57], Nair donne une approximation de la série $2 \sum_{k=1}^{\infty} (-1)^{k+1} e^{-2k^2 x^2}$ par $2e^{-2x^2}$. Les expressions des bornes de la bande de confiance asymptotique de Kolmogorov-Smirnov deviennent plus simplement :

$$\begin{aligned} R_{inf}(k) &= \mathbb{R}_n(k) - \sqrt{-\frac{1}{2n} \ln \frac{\alpha}{2}} \\ R_{sup}(k) &= \mathbb{R}_n(k) + \sqrt{-\frac{1}{2n} \ln \frac{\alpha}{2}} \end{aligned}$$

La bande de Kolmogorov-Smirnov est valable pour $0 \leq t < t_{(n)}$. C'est la bande de confiance la plus répandue pour les variables aléatoires continues et elle se révèle conservatrice (la probabilité qu'elle contienne R est supérieure à $1 - \alpha$) (voir Bickel et Krieger [11]) lorsque la loi est discrète.

Exemple :

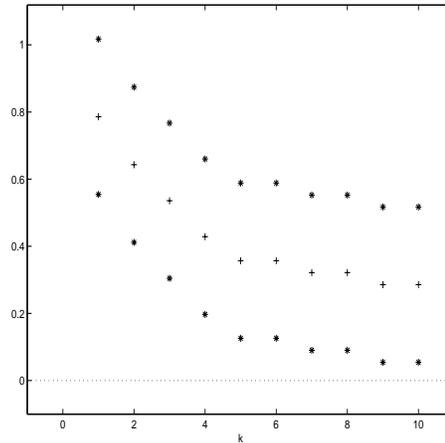


FIG. 5.3 – Bande de confiance de Kolmogorov-Smirnov sur $[1, 10]$, de seuil $\alpha = 0.1$ pour les données du contrôle de qualité

Nous avons volontairement limité le graphe de la bande de confiance pour des valeurs de k comprises entre 1 et 10, à titre de comparaison avec les autres graphes des intervalles et des bandes de confiance de ce chapitre. Mais il s'avère que la bande de confiance de Kolmogorov-Smirnov est très large et coupe l'axe des abscisses à partir de $k = 13$.

Nair présente différentes méthodes pour construire des bandes de confiance asymptotiques pour la fiabilité d'une loi continue. Ces méthodes de construction reposent sur le résultat suivant qui n'est plus valable pour une variable discrète : si T est une variable aléatoire de fonction de répartition F continue sur \mathbb{R} , alors $F(T)$ suit une loi uniforme sur $[0, 1]$.

Dans [57], Nair donne en temps continu la construction de la bande de confiance asymptotique $[R'_{inf}(t), R'_{sup}(t)]$, $\forall t \in [a, b]$ avec $k_{(1)} < a < b < k_{(n)}$, de seuil $\alpha \in]0, 1[$ pour R :

$$R'_{inf}(t) = \mathbb{R}_n(t) - q_\alpha(a, b) \sqrt{\frac{\mathbb{R}_n(t)(1 - \mathbb{R}_n(t))}{n}}$$

$$R'_{sup}(t) = \mathbb{R}_n(t) + q_\alpha(a, b) \sqrt{\frac{\mathbb{R}_n(t)(1 - \mathbb{R}_n(t))}{n}}$$

avec $q_\alpha(a, b)$ tel que $A(q_\alpha(a, b)) = \alpha/2$ où $A(t) = te^{-t^2/2} \log \left[\frac{(1 - \mathbb{R}_n(a))\mathbb{R}_n(b)}{\mathbb{R}_n(a)(1 - \mathbb{R}_n(b))} \right] / \sqrt{8\pi}$.

Exemple :

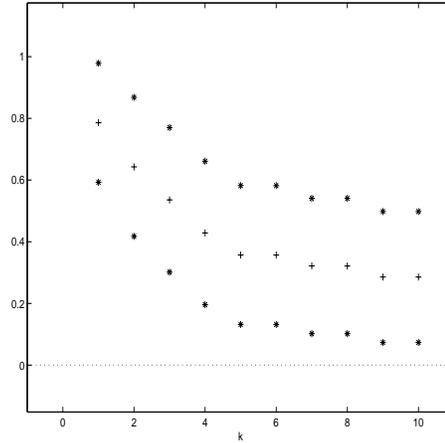


FIG. 5.4 – Bande de confiance de Nair sur $[1, 10]$, de seuil $\alpha = 0.1$ pour les données du contrôle de qualité

Bien que cette bande de confiance ne soit a priori pas valable pour une variable discrète, on peut constater sur notre exemple que cette bande de confiance diffère très peu de celle de Kolmogorov-Smirnov.

5.3 Le taux de défaillance empirique

5.3.1 Définition

On rappelle la définition du taux de défaillance :

$$\lambda(k) = \frac{\mathbb{P}(K = k)}{\mathbb{P}(K \geq k)} = \frac{R(k-1) - R(k)}{R(k-1)}$$

On propose alors de définir le taux de défaillance empirique en remplaçant la fiabilité par son estimateur empirique dans (4.1). La définition du taux de défaillance empirique est alors :

$$\begin{aligned} \lambda_n(k) &= \frac{\mathbb{P}_n(k)}{\mathbb{R}_n(k-1)} = \frac{\mathbb{R}_n(k-1) - \mathbb{R}_n(k)}{\mathbb{R}_n(k-1)}, \quad \forall k \leq k_{(n)} \\ &= \frac{\text{nombre d'observations} = k}{\text{nombre d'observations} \geq k}. \end{aligned} \quad (5.9)$$

Les particularités du taux de défaillance empirique sont les suivantes :

1. λ_n n'est pas défini pour $k > k_{(n)}$, puisque le dénominateur doit être non nul,
2. si k n'est pas égal à une observation, alors $\lambda_n(k) = 0$,
3. $\lambda_n(k_{(j)}^*) = \frac{d_j}{n_{j-1}}$, $j = 1, \dots, l$,
4. $\lambda_n(k_{(l)}^*) = 1$.

Remarques :

Il est important de noter que l'estimateur empirique du taux de défaillance que nous introduisons n'a pas d'équivalent en temps continu à la différence des estimateurs empiriques de R et F . En effet, en temps discret, il n'existe qu'un seul estimateur empirique très simple de $\mathbb{P}(K = k)$, tandis qu'il existe en temps continu plusieurs estimateurs empiriques, plus complexes, de la densité.

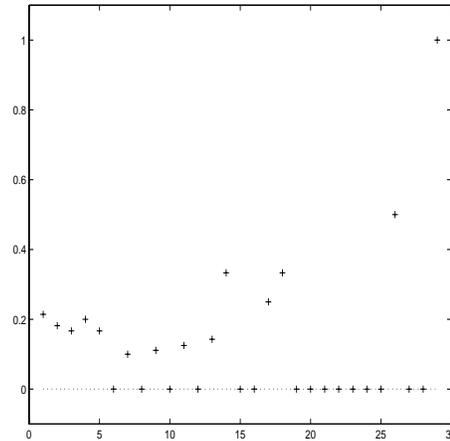
Exemple :

FIG. 5.5 – Taux de défaillance empirique pour les données du contrôle de qualité

Pour apprécier objectivement le sens de variation du taux de défaillance empirique, il faut prendre en compte également les valeurs nulles de celui-ci, ce qui rend l'interprétation graphique d'autant plus difficile. La figure 5.5 illustre bien la difficulté à se prononcer sur le sens de variation du taux de défaillance. On peut remarquer que la variance de l'estimateur empirique $\lambda_n(k)$ semble être une fonction croissante de k .

Nous allons étudier la loi de probabilité de cet estimateur empirique, tout d'abord à taille d'échantillon fixée puis asymptotiquement. Nous pourrions en déduire ainsi les propriétés usuelles (biais, variance et théorème limite), et construire un intervalle et une bande de confiance pour le taux de défaillance.

5.3.2 Loi exacte du taux de défaillance empirique

La définition du taux de défaillance empirique donnée par (5.9) est de la forme $\frac{i}{m}$ où $0 \leq i \leq m \leq n$, $m > 0$. Un problème apparaît car le taux de défaillance empirique en k est défini par une fraction qui peut ne pas être réduite ($\frac{ij}{mj} = \frac{i}{m}$, $\forall j$). Pour étudier la loi de probabilité du taux de défaillance empirique, il faut alors considérer $\frac{i}{m}$ sous sa forme

irréductible et envisager tous les rapports $\frac{ij}{mj}$, avec la contrainte $mj \leq n$, donc $j \leq \frac{n}{m}$.

Propriété 10

$\forall k \leq k_{(n)}$, la loi de probabilité du taux de défaillance empirique en k est définie par :

- $\mathbb{P}(\lambda_n(k) = 0) = (1 - p(k))^n$
- Pour $1 \leq i \leq m \leq n$, avec i et m premiers entre eux :

$$\mathbb{P}(\lambda_n(k) = \frac{i}{m}) = [F(k-1)]^n \sum_{j=1}^{\lfloor \frac{n}{m} \rfloor} \frac{n!}{(n-jm)!(ji)!(jm-ji)!} \left[\frac{R(k)^{m-i} p(k)^i}{F(k-1)^m} \right]^j$$

Preuve :

$$\begin{aligned} \mathbb{P}(\lambda_n(k) = 0) &= \mathbb{P}(\text{aucune observation n'est égale à } k) \\ &= (1 - p(k))^n \end{aligned}$$

Pour un k donné, on peut partager les n observations en trois groupes :

- les observations inférieures ou égales à $k-1$ (au nombre de $n\mathbb{F}_n(k-1)$),
- les observations égales à k (au nombre de $n\mathbb{P}_n(k)$),
- les observations strictement supérieures à k (au nombre de $n\mathbb{R}_n(k)$).

Le vecteur $(n\mathbb{F}_n(k-1), n\mathbb{P}_n(k), n\mathbb{R}_n(k))$ suit la loi multinomiale $\mathcal{M}(n; F(k-1), p(k), R(k))$.

Par définition $\lambda_n(k) = \frac{\mathbb{P}_n(k)}{\mathbb{R}_n(k-1)} = \frac{n\mathbb{P}_n(k)}{n - n\mathbb{F}_n(k-1)}$, d'où, pour $1 \leq i \leq m \leq n$, avec i et m premiers entre eux :

$$\begin{aligned} \mathbb{P}(\lambda_n(k) = \frac{i}{m}) &= \sum_{j=1}^{\lfloor \frac{n}{m} \rfloor} \mathbb{P}(n\mathbb{P}_n(k) = ji \cap n\mathbb{F}_n(k-1) = n - jm) \\ &= \sum_{j=1}^{\lfloor \frac{n}{m} \rfloor} \mathbb{P}(n\mathbb{F}_n(k-1) = n - jm, n\mathbb{P}_n(k) = ji, \mathbb{R}_n(k) = jm - ji) \\ &= \sum_{j=1}^{\lfloor \frac{n}{m} \rfloor} \frac{n!}{(n-jm)!(ji)!(jm-ji)!} F(k-1)^{n-jm} p(k)^{ji} R(k)^{jm-ji} \quad \square \end{aligned}$$

À partir de la loi du taux de défaillance empirique donnée par la propriété 10, on peut donner ses deux premiers moments :

Propriété 11

$$\begin{aligned} \mathbb{E}(\lambda_n(k)) &= \lambda(k) [1 - F(k-1)^n] \\ \mathbb{E}(\lambda_n^2(k)) &= F(k-1)^n p(k) \sum_{m=1}^n \frac{1}{m} C_n^m \frac{1}{F(k-1)^m} \left\{ p(k)(m-1) [R(k) + p(k)]^{m-2} + [R(k) + p(k)]^{m-1} \right\} \end{aligned}$$

Preuve :

$$\mathbb{E}(\lambda_n(k)) = \sum_{m=1}^n \sum_{i=1}^m \frac{i}{m} \mathbb{P}(n\mathbb{P}_n(k) = i \cap n\mathbb{F}_n(k-1) = n-m)$$

En utilisant les résultats de la propriété 10 sur la loi de probabilité de l'estimateur empirique du taux de défaillance, on obtient :

$$\begin{aligned} \mathbb{E}(\lambda_n(k)) &= \sum_{m=1}^n \sum_{i=1}^m \frac{i}{m} \frac{n!}{(n-m)!i!(m-i)!} \frac{p(k)^i R(k)^{m-i}}{F(k-1)^m} F(k-1)^n \\ &= F(k-1)^n \sum_{m=1}^n \frac{1}{m} \frac{n!}{(n-m)!(m-1)!} \frac{1}{F(k-1)^m} \sum_{i=1}^m \frac{i(m-1)!}{i!(m-i)!} R(k)^{m-i} p(k)^i \\ &= F(k-1)^n \sum_{m=1}^n C_n^m \frac{1}{F(k-1)^m} \sum_{i=0}^{m-1} C_{m-1}^i R(k)^{m-i+1} p(k)^{i+1} \\ &= F(k-1)^n \sum_{m=1}^n C_n^m \frac{1}{F(k-1)^m} p(k) [R(k) + p(k)]^{m-1} \\ &= F(k-1)^n p(k) \sum_{m=1}^n C_n^m \frac{[1 - F(k-1)]^{m-1}}{F(k-1)^m} \\ &= F(k-1)^n p(k) \frac{1}{1 - F(k-1)} \sum_{m=1}^n C_n^m \left[\frac{1 - F(k-1)}{F(k-1)} \right]^m \\ &= \frac{F(k-1)^n p(k)}{1 - F(k-1)} \left[\left(1 + \frac{1 - F(k-1)}{F(k-1)} \right)^n - 1 \right] \\ &= \frac{F(k-1)^n p(k)}{1 - F(k-1)} \left[\frac{1}{F(k-1)^n} - 1 \right] \\ &= \lambda(k) [1 - F(k-1)^n] \end{aligned}$$

$$\begin{aligned} \mathbb{E}(\lambda_n^2(k)) &= \sum_{m=1}^n \sum_{i=1}^m \left(\frac{i}{m} \right)^2 \frac{n!}{(n-m)!i!(m-i)!} \frac{R(k)^{m-i} p(k)^i}{F(k-1)^m} F(k-1)^n \\ &= F(k-1)^n \sum_{m=1}^n \frac{n!}{m^2(n-m)!} \frac{1}{F(k-1)^m} \sum_{i=1}^m \frac{i^2}{i!(m-i)!} R(k)^{m-i} p(k)^i \\ &= F(k-1)^n \sum_{m=1}^n \frac{n!}{m^2(n-m)!(m-1)!} \frac{1}{F(k-1)^m} \sum_{i=1}^m \frac{i(m-1)!}{(i-1)!(m-i)!} R(k)^{m-i} p(k)^i \\ &= F(k-1)^n \sum_{m=1}^n \frac{n!}{m^2(n-m)!(m-1)!} \frac{1}{F(k-1)^m} \sum_{i=1}^m \frac{i(m-1)!}{(i-1)!(m-i)!} R(k)^{m-i} p(k)^i \\ &= F(k-1)^n \sum_{m=1}^n \frac{1}{m} C_n^m \frac{1}{F(k-1)^m} \sum_{i=1}^m i C_{m-1}^{i-1} R(k)^{m-i} p(k)^i \\ &= F(k-1)^n p(k) \sum_{m=1}^n \frac{1}{m} C_n^m \frac{1}{F(k-1)^m} \sum_{j=0}^{m-1} (j+1) C_{m-1}^j R(k)^{m-j-1} p(k)^j \\ &= F(k-1)^n p(k) \sum_{m=1}^n \frac{1}{m} C_n^m \frac{1}{F(k-1)^m} \left\{ p(k)(m-1) [R(k) + p(k)]^{m-2} + [R(k) + p(k)]^{m-1} \right\} \square \end{aligned}$$

Remarques :

1. $\mathbb{E}(\lambda_n(k)) \leq \lambda(k)$. On aura donc tendance à sous-estimer le taux de défaillance. Cela est probablement dû au fait que $\mathbb{P}(\lambda_n(k) = 0) \neq 0$. On peut envisager alors de corriger l'estimation en multipliant $\lambda_n(k)$ par une quantité supérieure à 1.
2. Le biais de $\lambda_n(k)$ est d'autant plus fort que $F(k-1)$ est grand, donc que k est grand. En particulier, $\lambda_n(k_{(l)}^*) = 1$ traduit bien l'existence d'un biais important. Par conséquent, l'estimation de $\lambda_n(k)$ n'est exploitable, pour de petites tailles d'échantillon, que pour les faibles valeurs de k . Ainsi, pour juger du sens de variation de λ au vu du graphe de λ_n , il vaut mieux ne pas utiliser les plus grandes observations.

Propriété 12

$\forall k \in \mathbb{N}^*$, $\lambda_n(k)$ est un estimateur biaisé et asymptotiquement sans biais de $\lambda(k)$.

Preuve :

Ce résultat est une conséquence directe de la propriété 11. \square

5.3.3 Loi asymptotique du taux de défaillance empirique

Nous allons maintenant étudier les propriétés asymptotiques du taux de défaillance empirique.

Propriété 13

$$\forall k \in \mathbb{N}^*, \quad \lambda_n(k) \xrightarrow{p.s.} \lambda(k)$$

Preuve :

D'après la loi des grands nombres, on a $\mathbb{P}_n(k) \xrightarrow{p.s.} p(k)$. On sait également que $\mathbb{R}_n(k) \xrightarrow{p.s.} R(k)$, d'où $\lambda_n(k) \xrightarrow{p.s.} \lambda(k)$. \square

Par analogie avec les résultats obtenus pour la fiabilité, nous introduisons la notation suivante :

Définition 5

Soit \mathbb{T}_n le processus empirique associé au taux de défaillance, défini par :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{T}_n(k) = \sqrt{n}(\lambda_n(k) - \lambda(k)).$$

Théorème 8

Pour $k \in \mathbb{N}^*$ fixé, on a $\mathbb{T}_n(k) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\lambda(k)(1 - \lambda(k))}{R(k-1)}\right)$

Preuve :

Fixons k un entier non nul. Le processus empirique associé au taux de défaillance peut s'écrire :

$$\begin{aligned}
\mathbb{T}_n(k) &= \sqrt{n} \left(\frac{\mathbb{R}_n(k-1) - \mathbb{R}_n(k)}{\mathbb{R}_n(k-1)} - \frac{R(k-1) - R(k)}{R(k-1)} \right) \\
&= \sqrt{n} \left(-\frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)} - \frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)} - \frac{R(k-1) - R(k)}{R(k-1)} \right) \\
&= \sqrt{n} \left(-\frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)} + \frac{R(k)\mathbb{R}_n(k-1) - R(k)R(k-1)}{\mathbb{R}_n(k-1)(R(k-1))} \right) \\
&= \sqrt{n} \left(-\frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)} + \frac{R(k)}{R(k-1)} \frac{\mathbb{R}_n(k-1) - R(k-1)}{\mathbb{R}_n(k-1)} \right) \\
&= \sqrt{n} \left(-\frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)} + (1 - \lambda(k)) \frac{\mathbb{R}_n(k-1) - R(k-1)}{\mathbb{R}_n(k-1)} \right).
\end{aligned}$$

En utilisant les résultats de convergence de la fiabilité empirique rappelés par la propriété 8 et le théorème de Slutsky, on obtient la convergence en loi du vecteur

$$\begin{pmatrix} \sqrt{n}(1 - \lambda(k)) \frac{\mathbb{R}_n(k-1) - R(k-1)}{\mathbb{R}_n(k-1)} \\ \sqrt{n} \frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)} \end{pmatrix}$$

vers le vecteur gaussien centré de matrice de covariance

$$\frac{1}{R(k-1)^2} \begin{pmatrix} (1 - \lambda(k))^2 R(k-1)(1 - R(k-1)) & (1 - \lambda(k))R(k)(1 - R(k-1)) \\ (1 - \lambda(k))R(k)(1 - R(k-1)) & R(k)(1 - R(k)) \end{pmatrix}.$$

La suite $\{\mathbb{T}_n(k)\}_{n \geq 1}$ converge donc vers une variable aléatoire $\mathbb{T}(k)$ de loi normale en tant que combinaison linéaire de composantes d'un vecteur gaussien.

Il est clair que la variable $\mathbb{T}(k)$ est centrée. D'après ce qui précède, la variance $\Pi(k)$ de la variable $\mathbb{T}(k)$ est donnée par :

$$\Pi(k) = \frac{R(k)(1 - R(k))}{R(k-1)^2} + \frac{(1 - \lambda(k))^2 R(k-1)(1 - R(k-1))}{R(k-1)^2} - 2 \frac{(1 - \lambda(k))R(k)(1 - R(k-1))}{R(k-1)^2}$$

Après simplifications, on obtient :

$$\Pi(k) = \frac{R(k)}{R(k-1)^3} [R(k-1) - R(k)] = \frac{\lambda(k)(1 - \lambda(k))}{R(k-1)} \quad \square$$

Pour n suffisamment grand, le théorème 8 permet d'admettre que la loi de $\mathbb{T}_n(k)$, en tout $k \in \mathbb{N}^*$, est approximativement la loi normale $\mathcal{N}(0, \Pi(k))$. On peut ainsi construire un intervalle de confiance asymptotique de seuil α pour $\lambda(k)$:

$$\left[\lambda_n(k) - z_\alpha \sqrt{\frac{\Pi(k)}{n}}, \lambda_n(k) + z_\alpha \sqrt{\frac{\Pi(k)}{n}} \right]$$

Mais cette dernière formule n'est en fait utilisable que si on connaît la variance asymptotique $\Pi(k)$. Or, celle-ci dépend de la fonction de répartition et est donc inconnue. Il est donc nécessaire d'estimer la fonction de répartition à l'aide la fonction de répartition empirique. Soit $\tilde{\Pi}_n(k)$ la variance estimée :

$$\tilde{\Pi}_n(k) = \frac{\mathbb{R}_n(k)}{\mathbb{R}_n(k-1)^3} [\mathbb{R}_n(k-1) - \mathbb{R}_n(k)]$$

La convergence presque sûre de $\mathbb{R}_n(k)$ vers $R(k)$ assure que $\tilde{\Pi}_n(k)$ converge presque sûrement vers $\Pi(k)$ quand $n \rightarrow +\infty$.

Le théorème de Slutsky nous permet d'écrire :

$$Z_n(k) = \sqrt{n} \frac{(\lambda_n(k) - \lambda(k))}{\sqrt{\tilde{\Pi}_n(k)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (5.10)$$

La construction d'un intervalle de confiance asymptotique $[\lambda_{inf}(k), \lambda_{sup}(k)]$, de seuil α pour $\lambda(k)$ est donnée par :

$$\begin{aligned} \lambda_{inf}(k) &= \lambda_n(k) - z_\alpha \sqrt{\frac{\tilde{\Pi}_n(k)}{n}} \\ \lambda_{sup}(k) &= \lambda_n(k) + z_\alpha \sqrt{\frac{\tilde{\Pi}_n(k)}{n}} \end{aligned}$$

Comme pour la fonction de fiabilité, la transformation du logit se révèle utile pour construire un intervalle de confiance pour le taux de défaillance, avec la certitude d'avoir des bornes comprises dans $[0, 1]$. La construction de l'intervalle de confiance asymptotique utilisant la transformation du logit, repose sur le résultat de convergence suivant :

$$\sqrt{n} (\text{logit}(\lambda_n(k)) - \text{logit}(\lambda(k))) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{\lambda(k)R(k)}\right)$$

La convergence presque sûre de $\lambda_n(k)$ vers $\lambda(k)$ et le théorème de Slutsky permettent d'écrire que :

$$Z_n^{\text{logit}}(k) = \sqrt{n \lambda_n(k) \mathbb{R}_n(k)} (\text{logit}(\lambda_n(k)) - \text{logit}(\lambda(k))) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (5.11)$$

On obtient alors un intervalle de confiance asymptotique $[\lambda'_{inf}(k), \lambda'_{sup}(k)]$, de seuil $\alpha \in]0, 1[$ pour $\lambda(k)$ donné par :

$$\begin{aligned} \lambda'_{inf}(k) &= \frac{\lambda_n(k)}{\lambda_n(k) + (1 - \lambda_n(k)) \exp\left[\frac{z_\alpha}{\sqrt{\lambda_n(k) \mathbb{R}_n(k)}}\right]} \\ \lambda'_{sup}(k) &= \frac{\lambda_n(k)}{\lambda_n(k) + (1 - \lambda_n(k)) \exp\left[-\frac{z_\alpha}{\sqrt{\lambda_n(k) \mathbb{R}_n(k)}}\right]} \end{aligned}$$

Exemple :

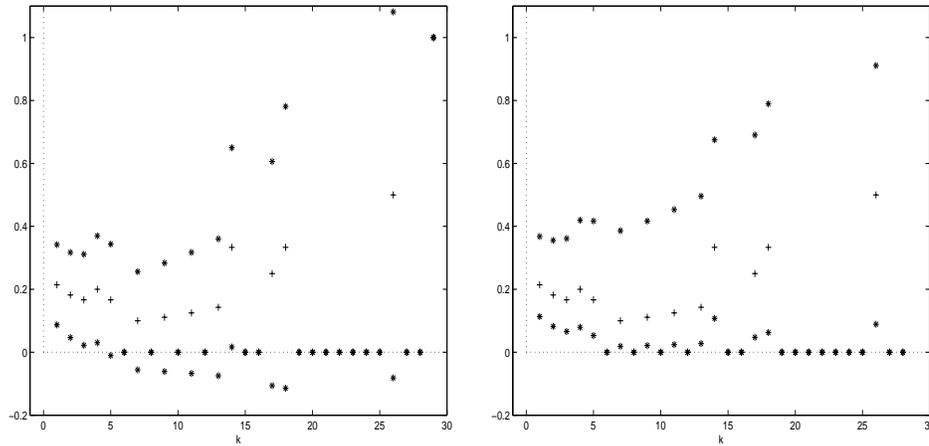


FIG. 5.6 – Intervalles de confiance asymptotiques du taux de défaillance, sans ou avec le logit, au seuil $\alpha = 0.1$ pour les données du contrôle de qualité

On peut constater que la largeur des intervalles de confiance augmente rapidement à partir de $k \geq 15$. De plus, la borne inférieure peut prendre des valeurs négatives pour l'intervalle de confiance sans logit, en raison de l'approximation et de la variance importante du processus empirique. Comme pour la fonction de fiabilité, l'intervalle de confiance avec logit n'est pas symétrique. Contrairement à ce que l'on a observé pour la fiabilité, les intervalles de confiance avec logit semblent moins larges que les intervalles de confiance sans logit.

Nous allons maintenant nous intéresser à la loi des vecteurs fini-dimensionnels $(\mathbb{T}_n(x_j) : 1 \leq j \leq m)^t$ pour la construction de bandes de confiance.

Théorème 9

Pour tout entier $m \geq 1$ et toute suite (x_1, \dots, x_m) de m nombres entiers classés par ordre croissant, le vecteur aléatoire

$$(\mathbb{T}_n(x_j) : 1 \leq j \leq m)^t$$

converge en loi, quand $n \rightarrow +\infty$, vers un vecteur aléatoire gaussien \mathbb{T}^t de \mathbb{R}^m , centré, de matrice de variance Π définie par :

$$\begin{cases} \Pi(j, k) = 0 \text{ pour } j \neq k \\ \Pi(k, k) = \frac{R(k)}{[R(k-1)]^3} [R(k-1) - R(k)] = \Pi(k) \end{cases}$$

Preuve :

Le vecteur $(\mathbb{T}_n(x_j) : 1 \leq j \leq m)^t$ peut s'écrire

$$\left(\frac{R(x_j)}{R(x_j-1)} \sqrt{n} \frac{\mathbb{R}_n(x_j-1) - R(x_j-1)}{\mathbb{R}_n(x_j-1)} - \sqrt{n} \frac{\mathbb{R}_n(x_j) - R(x_j)}{\mathbb{R}_n(x_j-1)} : 1 \leq j \leq m \right)^t.$$

Le théorème de Slutsky permet d'écrire que ce vecteur a même limite en loi que le vecteur

$$\left(\frac{R(x_j)}{R^2(x_j - 1)} \sqrt{n}(\mathbb{R}_n(x_j - 1) - R(x_j - 1)) - \frac{1}{R(x_j - 1)} \sqrt{n}(\mathbb{R}_n(x_j) - R(x_j)) : 1 \leq j \leq m \right)^t$$

qui est une transformation linéaire du vecteur $(\sqrt{n}(\mathbb{R}_n(x_j) - R(x_j)) : 1 \leq j \leq m)^t$. D'après (5.7), celui-ci est asymptotiquement gaussien, donc $(\mathbb{T}_n(x_j) : 1 \leq j \leq m)^t$ est asymptotiquement gaussien. L'espérance limite est clairement nulle, il reste à calculer la matrice de covariance limite Π .

Soient $(j, k) \in \mathbb{N}^{*2}$, et $\Pi_n(k, j) = Cov(\mathbb{T}_n(k), \mathbb{T}_n(j))$.

$$\begin{aligned} \Pi_n(k, j) &= Cov \left(\sqrt{n} \frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)}, \sqrt{n} \frac{\mathbb{R}_n(j) - R(j)}{\mathbb{R}_n(j-1)} \right) \\ &- (1 - \lambda(j)) Cov \left(\sqrt{n} \frac{\mathbb{R}_n(k) - R(k)}{\mathbb{R}_n(k-1)}, \sqrt{n} \frac{\mathbb{R}_n(j-1) - R(j-1)}{\mathbb{R}_n(j-1)} \right) \\ &- (1 - \lambda(k)) Cov \left(\sqrt{n} \frac{\mathbb{R}_n(k-1) - R(k-1)}{\mathbb{R}_n(k-1)}, \sqrt{n} \frac{\mathbb{R}_n(j) - R(j)}{\mathbb{R}_n(j-1)} \right) \\ &+ (1 - \lambda(j))(1 - \lambda(k)) Cov \left(\sqrt{n} \frac{\mathbb{R}_n(k-1) - R(k-1)}{\mathbb{R}_n(k-1)}, \sqrt{n} \frac{\mathbb{R}_n(j-1) - R(j-1)}{\mathbb{R}_n(j-1)} \right). \end{aligned}$$

La formule (5.7) de la propriété 8 permet d'écrire que pour tout $(i, j) \in \mathbb{N}^{*2}$, on a :

$$\lim_{n \rightarrow \infty} Cov(\sqrt{n}(\mathbb{R}_n(i) - R(i)), \sqrt{n}(\mathbb{R}_n(j) - R(j))) = R(\max(i, j)) - R(i)R(j).$$

On en déduit que :

$$\begin{aligned} \Pi(k, j) &= \frac{1}{R(k-1)^2 R(j-1)^2} \{ \\ &[R(\max(k, j)) - R(k)R(j)]R(k-1)R(j-1) \\ &- [R(\max(k-1, j)) - R(k-1)R(j)]R(k)R(j-1) \\ &- [R(\max(k, j-1)) - R(k)R(j-1)]R(j)R(k-1) \\ &+ [R(\max(k-1, j-1)) - R(k-1)R(j-1)]R(j)R(k) \}. \end{aligned}$$

Pour $k = j$, on trouve après simplifications : $\Pi(k, k) = \frac{R(k)}{R(k-1)^3} [R(k-1) - R(k)]$.

Pour $k < j$, le calcul aboutit à $\Pi(k, j) = 0$. C'est à dire que pour $j \neq k$, $\lambda_n(j)$ et $\lambda_n(k)$ sont asymptotiquement indépendants. \square

La construction d'une bande de confiance pour $\lambda(k)$ repose sur la loi de la variable aléatoire

$$Z_{a,b}^n = \max_{\{a \leq k \leq b\}} Z_n(k)$$

On sait d'après le théorème 9 que le vecteur $(\mathbb{T}_n(k) : a \leq k \leq b)^t$ converge en loi vers un vecteur de variables normales indépendantes et centrées. En utilisant le théorème de Slutsky et la convergence presque sûre de $\tilde{\Pi}_n(k)$ vers $\Pi(k)$, on a la convergence en loi du vecteur

$(Z_n(k) : a \leq k \leq b)^t$ vers un vecteur de variables indépendantes normales centrées et réduites. La suite $\{Z_{a,b}^n\}_{n \geq 1}$ converge vers la v.a. $Z_{a,b}$. D'après ce qui précède, la loi de $Z_{a,b}$ est celle du maximum de $b - a + 1$ variables normales centrées réduites indépendantes. On construit la bande de confiance asymptotique de λ sur $[a, b]$, de seuil α en explicitant le quantile correspondant de la variable aléatoire $|Z_{a,b}|$:

$$\begin{aligned}
 P(|Z_{a,b}| < q_{1-\alpha}) = 1 - \alpha &\iff P(-q_{1-\alpha} < Z_{a,b} < q_{1-\alpha}) = 1 - \alpha \\
 &\iff F_{Z_{a,b}}(q_{1-\alpha}) - F_{Z_{a,b}}(-q_{1-\alpha}) = 1 - \alpha \\
 &\iff \phi^{b-a+1}(q_{1-\alpha}) - \phi^{b-a+1}(-q_{1-\alpha}) = 1 - \alpha \\
 &\iff \phi^{b-a+1}(q_{1-\alpha}) - [1 - \phi(q_{1-\alpha})]^{b-a+1} = 1 - \alpha \quad (5.12)
 \end{aligned}$$

où ϕ est la fonction de répartition de la loi normale centrée réduite.

Finalement, la bande de confiance asymptotique $[\lambda_{inf}''(k), \lambda_{sup}''(k)]$, $\forall k \in [a, b]$, de seuil $\alpha \in]0, 1[$ pour λ est donnée par :

$$\begin{aligned}
 \lambda_{inf}''(k) &= \lambda_n(k) - q_{1-\alpha} \sqrt{\frac{\tilde{\Pi}_n(k)}{n}}, \forall k \in [a, b] \\
 \lambda_{sup}''(k) &= \lambda_n(k) + q_{1-\alpha} \sqrt{\frac{\tilde{\Pi}_n(k)}{n}}, \forall k \in [a, b]
 \end{aligned}$$

avec $q_{1-\alpha}$ tel que $\phi^{b-a+1}(q_{1-\alpha}) - [1 - \phi(q_{1-\alpha})]^{b-a+1} = 1 - \alpha$.

Exemple :

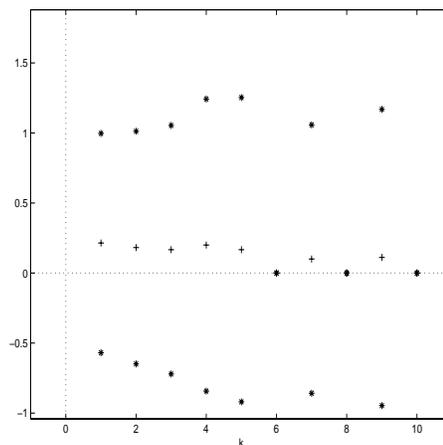


FIG. 5.7 – Bande de confiance asymptotique au seuil $\alpha = 0.1$ du taux de défaillance pour les données du contrôle de qualité

À l'instar des intervalles de confiance, la variance $\Pi(k)$ du processus empirique associé au taux de défaillance pouvant prendre de grandes valeurs, il est possible que les bornes de

la bande de confiance soient en dehors de l'intervalle $[0, 1]$. On peut alors remédier à cela en utilisant la transformation du logit et en procédant comme ci-dessus.

Soit $Z_{a,b,n}^{logit} = \max_{\{a \leq k \leq b\}} Z_n^{logit}(k)$, avec $Z_n^{logit}(k)$ définie par (5.11). Le théorème 9 assure que le vecteur $(\sqrt{n}(\text{logit}(\lambda_n(k)) - \text{logit}(\lambda(k))) : a \leq k \leq b)^t$ converge en loi vers un vecteur de variables normales centrées et indépendantes. Mais concernant le vecteur $(Z_n^{logit}(k) : a \leq k \leq b)^t$, la présence du produit $\lambda_n(k)\mathbb{R}_n(k)$ implique que l'on ne peut utiliser le théorème de Slutsky et ainsi vérifier l'indépendance asymptotique entre les composantes. Par conséquent, on ne peut pas obtenir facilement une bande de confiance avec cette méthode.

Exemple :

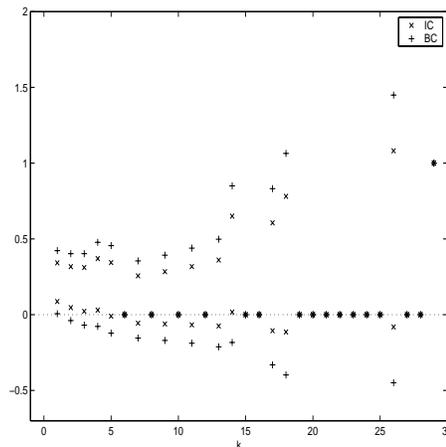


FIG. 5.8 – Comparaison de la bande et de l'intervalle de confiance asymptotiques, au seuil $\alpha = 0.1$ du taux de défaillance pour les données du contrôle de qualité

Comme on pouvait s'y attendre, la bande de confiance sur λ pour tout l'intervalle $[1, 29]$ est plus large que l'ensemble des intervalles de confiance pour chaque $\lambda(k)$, avec k variant de 1 à 29.

5.3.4 Relation entre R et λ

L'expression de la fiabilité en fonction du taux de défaillance est la suivante :

$$R(k) = \prod_{i=1}^k (1 - \lambda(i))$$

$$\text{On a } \mathbb{R}_n(k) = \prod_{i=1}^k \frac{\mathbb{R}_n(i)}{\mathbb{R}_n(i-1)} = \prod_{i=1}^k (1 - \lambda_n(i)) = \prod_{j:k_{(j)}^* \leq k} (1 - \lambda_n(j)) = \prod_{j:k_{(j)}^* \leq k} \left(1 - \frac{d_j}{n_{j-1}}\right)$$

Donc en temps discret, l'estimateur empirique de la fiabilité est directement sous une forme produit de type Kaplan-Meier.

5.4 Le nouveau taux de défaillance empirique

5.4.1 Définition

La définition du nouveau taux de défaillance donnée au chapitre 4 est :

$$\forall k \in \mathbb{N}^*, \quad r(k) = \ln \frac{R(k-1)}{R(k)} = -\ln(1 - \lambda(k))$$

On définit logiquement l'estimateur empirique du nouveau taux de défaillance en remplaçant R par son estimateur empirique \mathbb{R}_n :

$$\forall k < k_{(n)}, \quad r_n(k) = \ln \frac{\mathbb{R}_n(k-1)}{\mathbb{R}_n(k)} = -\ln(1 - \lambda_n(k)) \quad (5.13)$$

Les caractéristiques du nouveau taux de défaillance sont les suivantes :

1. si k n'est pas égal à une observation, alors $r_n(k) = 0$ (comme pour $\lambda_n(k)$),
2. r_n n'est pas défini pour $k = k_{(n)}$, a contrario de λ_n ($\lambda_n(k_{(n)}) = 1$),
3. r_n n'est pas borné par 1 comme l'est λ_n ,
4. $r_n(k_{(j)}^*) = \ln\left(1 - \frac{d_j}{n_{j-1}}\right) = -\ln \frac{n_j}{n_{j-1}}$

L'estimateur empirique du nouveau taux de défaillance donne une estimation supérieure à celle de l'estimateur empirique du taux de défaillance usuel. Cela s'explique à l'aide du résultat suivant appliqué à (5.13) : $\forall x, x < -\ln(1 - x)$.

Exemple :

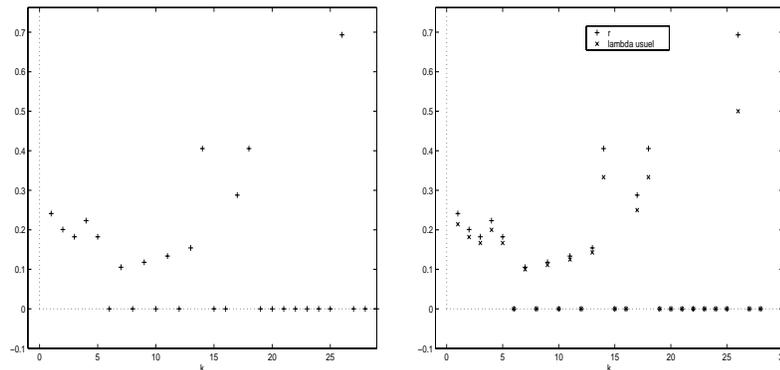


FIG. 5.9 – Nouveau taux de défaillance empirique pour les données du contrôle de qualité

On peut voir comme prévu que les deux taux de défaillance empiriques sont équivalents pour les petites valeurs de k , mais que des différences notables apparaissent pour les grandes valeurs de k .

La loi exacte de $r_n(k)$ est trop complexe à calculer en raison de la présence du logarithme. On ne peut donc obtenir le biais de ce nouveau taux de défaillance empirique.

5.4.2 Loi asymptotique du nouveau taux de défaillance empirique

La relation (4.4) entraîne la validité des résultats asymptotiques obtenus sur le taux de défaillance empirique usuel pour le nouveau taux de défaillance empirique, c'est à dire :

Propriété 14

$$\forall k \in \mathbb{N}^*, \quad r_n(k) \xrightarrow{p.s.} r(k).$$

Comme pour le taux de défaillance usuel, on peut s'intéresser au comportement asymptotique du processus empirique $\sqrt{n}(r_n(k) - r(k))$:

Propriété 15

$$\forall k < k_{(n)}, \quad \sqrt{n}(r_n(k) - r(k)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{R(k)} - \frac{1}{R(k-1)}\right)$$

Preuve :

D'après la relation (4.4) qui est de la forme $r_n(k) = \phi(\lambda(k))$, et le résultat de convergence asymptotique du processus empirique associé à λ (théorème 8), l'application de la méthode delta conduit à :

$$\forall k < k_{(n)}, \quad \sqrt{n}(r_n(k) - r(k)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\lambda(k)(1 - \lambda(k))}{R(k-1)} \phi'^2(\lambda(k))\right)$$

où $\phi'(u) = \frac{1}{1-u}$.

La variance du processus empirique se simplifie de la façon suivante :

$$\begin{aligned} \frac{\lambda(k)(1 - \lambda(k))}{R(k-1)} \phi'^2(\lambda(k)) &= \frac{\lambda(k)(1 - \lambda(k))}{R(k-1)} \frac{1}{(1 - \lambda(k))^2} \\ &= \frac{\lambda(k)}{(1 - \lambda(k))R(k-1)} \\ &= \frac{\lambda(k)}{R(k)} \\ &= \frac{1}{R(k)} - \frac{1}{R(k-1)}. \quad \square \end{aligned}$$

On déduit directement de la propriété 15, la propriété suivante :

Propriété 16

$\forall k \in \mathbb{N}^*$, $r_n(k)$ est un estimateur asymptotiquement sans biais de $r(k)$.

Nous allons nous intéresser maintenant à construire un intervalle de confiance pour le nouveau taux de défaillance en appliquant la même démarche utilisée pour la fiabilité et le taux de défaillance usuel.

La convergence presque sûre de $\mathbb{R}_n(k)$ vers $R(k)$ et l'utilisation du théorème de Slutsky entraînent alors :

$$Z_n(k) = \sqrt{n} \frac{(r_n(k) - r(k))}{\sqrt{\frac{1}{\mathbb{R}_n(k)} - \frac{1}{\mathbb{R}_n(k-1)}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (5.14)$$

Un intervalle de confiance asymptotique $[r_{inf}(k), r_{sup}(k)]$, de seuil $\alpha \in]0, 1[$ pour $r(k)$ est donc donné par :

$$\begin{aligned} r_{inf}(k) &= r_n(k) - z_\alpha \sqrt{\frac{1}{n} \left[\frac{1}{\mathbb{R}_n(k)} - \frac{1}{\mathbb{R}_n(k-1)} \right]} \\ r_{sup}(k) &= r_n(k) + z_\alpha \sqrt{\frac{1}{n} \left[\frac{1}{\mathbb{R}_n(k)} - \frac{1}{\mathbb{R}_n(k-1)} \right]} \end{aligned}$$

L'utilisation d'une transformation du nouveau taux de défaillance r peut se révéler d'autant plus utile que l'on sait que l'estimateur empirique du nouveau taux de défaillance donne une estimation supérieure à celle du taux de défaillance usuelle. La transformation du logit n'est pas la plus judicieuse à utiliser car $r(k)$ est à valeurs sur \mathbb{R}^+ , et non pas sur l'intervalle $[0, 1]$ comme le sont R et λ . Il faut plutôt utiliser une simple transformation logarithmique.

L'intervalle de confiance asymptotique utilisant la transformation logarithmique repose sur le résultat suivant :

$$\sqrt{n} (\log(r_n(k)) - \log(r(k))) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \left(\frac{1}{R(k)} - \frac{1}{R(k-1)} \right) \frac{1}{r^2(k)} \right)$$

Les convergences presque sûres de $\mathbb{R}_n(k)$ et $r_n(k)$ respectivement vers $R(k)$ et $r(k)$ et le théorème de Slutsky permettent d'écrire que :

$$Z_n^{log}(k) = \sqrt{n} \frac{(\log(r_n(k)) - \log(r(k)))}{\frac{1}{r_n(k)} \sqrt{\frac{1}{\mathbb{R}_n(k)} - \frac{1}{\mathbb{R}_n(k-1)}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (5.15)$$

Un intervalle de confiance asymptotique, utilisant la transformation logarithmique, $[r'_{inf}(k), r'_{sup}(k)]$, de seuil $\alpha \in]0, 1[$ pour $r(k)$ est alors donné par :

$$\begin{aligned} r'_{inf}(k) &= r_n(k) \exp \left(-\frac{z_\alpha}{r_n(k)} \sqrt{\frac{1}{n} \left[\frac{1}{\mathbb{R}_n(k)} - \frac{1}{\mathbb{R}_n(k-1)} \right]} \right) \\ r'_{sup}(k) &= r_n(k) \exp \left(\frac{z_\alpha}{r_n(k)} \sqrt{\frac{1}{n} \left[\frac{1}{\mathbb{R}_n(k)} - \frac{1}{\mathbb{R}_n(k-1)} \right]} \right) \end{aligned}$$

Exemple :

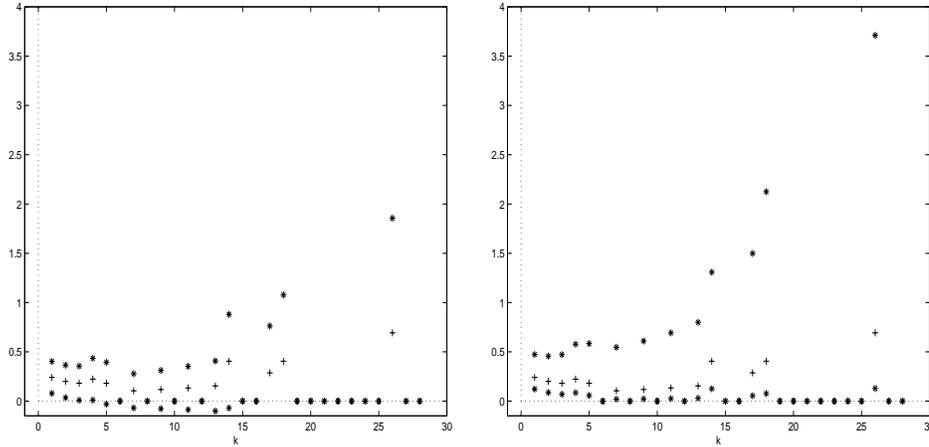


FIG. 5.10 – Intervalles de confiance asymptotiques, sans ou avec la transformation \log , au seuil $\alpha = 0.1$, du taux de défaillance pour les données du contrôle de qualité

L'intervalle de confiance asymptotique utilisant la transformation logarithmique renvoie bien une borne inférieure pour r positive ou nulle, contrairement à l'intervalle de confiance sans transformation. En revanche, l'intervalle de confiance avec la transformation logarithmique est plus grand, ce qui se traduit surtout par une borne supérieure plus élevée que sans l'utilisation du \log , ce phénomène étant croissant avec k .

Nous ne présentons pas de construction de bandes de confiance pour le nouveau taux de défaillance car les composantes du vecteur $(Z_n(k) : a \leq k \leq b)^t$ (avec $Z_n(k)$ définie par (5.14)) ne sont probablement pas indépendantes.

5.5 La durée de vie résiduelle moyenne empirique

Comme le taux de défaillance empirique, la durée de vie résiduelle moyenne empirique ne semble pas avoir été étudiée auparavant. On rappelle que la durée de vie résiduelle moyenne est définie par $m(k) = \mathbb{E}(K - k | K > k)$. Il semble logique de proposer la définition suivante pour la durée de vie résiduelle moyenne empirique :

$$\begin{aligned} m_n(k) &= \frac{\sum_{i=1}^n (K_i - k) \mathbb{I}_{\{K_i > k\}}}{\sum_{i=1}^n \mathbb{I}_{\{K_i > k\}}} = \frac{\sum_{i=1}^n K_i \mathbb{I}_{\{K_i > k\}}}{\mathbb{R}_n(k)} - k, \quad \forall k < k_{(n)}^* \\ &= \frac{\text{Somme des observations} > k}{\text{Nombre d'observation(s)} > k} - k \end{aligned}$$

Les caractéristiques de cet estimateur empirique sont les suivantes :

1. $m_n(0) = \overline{K}_n$, ce qui est cohérent avec $m(0) = \mathbb{E}(K)$,
2. $m_n(k_{(n)}^* - 1) = 1$, donc m_n possède une allure générale au moins en partie décroissante, commençant à \overline{K}_n pour finir à 1.
3. si k n'est pas égal à une observation, alors $m_n(k) = m_n(k - 1) - 1$ (car $\lambda_n(k) = 0$ dans (2.7)),
4. $m_n(k)$ n'est pas définie pour $k \geq k_{(n)}^*$ pour éviter la forme indéterminée 0/0.
5. si $k = k_{(j)}^*$, alors $m_n(k_{(j)}^*) = \frac{n_j - 1}{n_j} [m_n(k_{(j)}^*) - 1]$.

La durée de vie résiduelle moyenne empirique est donc une succession de segments de droites de pente -1 avec des sauts en chaque $k_{(j)}^*$.

Exemple :

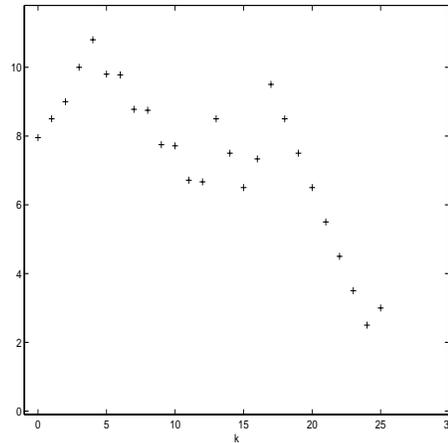


FIG. 5.11 – *Durée de vie résiduelle moyenne empirique pour les données du contrôle de qualité*

On rappelle la relation entre le taux de défaillance et la durée de vie résiduelle moyenne :

$$\text{Si } \lambda(k) \neq 1, \quad m(k) = \frac{m(k-1) - 1}{1 - \lambda(k)}$$

On peut vérifier que cette relation est également vérifiée par les estimateurs empiriques de m et λ :

Propriété 17

$$\forall k \in \mathbb{N}^*, \quad m_n(k) = \frac{m_n(k-1) - 1}{1 - \lambda_n(k)}$$

Preuve :

$$\begin{aligned}
\frac{m_n(k-1) - 1}{1 - \lambda_n(k)} &= \left(\frac{\sum_{i=1}^n K_i \mathbb{I}_{\{K_i > k-1\}}}{\mathbb{R}_n(k-1)} - (k-1) - 1 \right) \frac{\mathbb{R}_n(k-1)}{\mathbb{R}_n(k)} \\
&= \frac{\sum_{i=1}^n K_i \mathbb{I}_{\{K_i \geq k\}} - k \mathbb{R}_n(k-1)}{\mathbb{R}_n(k)} \\
&= \frac{\sum_{i=1}^n K_i \mathbb{I}_{\{K_i > k\}} - k \mathbb{R}_n(k) + \sum_{i=1}^n K_i \mathbb{I}_{\{K_i = k\}} - k \sum_{i=1}^n K_i \mathbb{I}_{\{K_i = k\}}}{\mathbb{R}_n(k)} \\
&= m_n(k) + \frac{\sum_{i=1}^n K_i \mathbb{I}_{\{K_i = k\}} - k \sum_{i=1}^n \mathbb{I}_{\{K_i = k\}}}{\mathbb{R}_n(k)}
\end{aligned}$$

En remarquant que $\sum_{i=1}^n K_i \mathbb{I}_{\{K_i = k\}} = k \sum_{i=1}^n \mathbb{I}_{\{K_i = k\}}$, on obtient le résultat cherché. \square

La durée de vie résiduelle moyenne empirique semble être un indicateur de vieillissement plus intéressant que le taux de défaillance empirique dans la mesure où on évite le problème des zéros et où le graphe obtenu est plus lisse que le précédent. Cependant, le fait que m_n varie de \overline{K}_n à 1 reste un handicap pour l'interprétation du vieillissement.

5.6 Conclusion

L'estimateur empirique du taux de défaillance introduit dans ce chapitre ne possède malheureusement pas les propriétés de l'estimateur empirique de la fiabilité. En effet, c'est un estimateur biaisé et surtout qui s'annule en dehors des observations. L'information qu'il apporte sur le phénomène observé est par conséquent assez faible, surtout quand on a peu de données, ce qui est le cas des matériels de Schneider Electric. En revanche, dans des contextes où les données sont plus nombreuses, comme l'exemple climatologique traité par Le Breton et Martin [52], le taux de défaillance empirique peut être très utile pour apprécier le vieillissement (ou la persistance).

Dans ce chapitre, nous nous intéressons à l'estimation des paramètres de la plupart des modèles de fiabilité présentés au chapitre 3. Une section importante de ce chapitre est consacrée à la loi géométrique. Cette loi est le point de départ de notre procédure d'analyse du retour d'expérience, et plus généralement de toute analyse de données représentant des durées de vie puisqu'elle suppose l'absence de vieillissement dans le phénomène observé. Pour chaque modèle, nous donnons les propriétés (biais, variance et coefficient de variation) des estimateurs ponctuels des paramètres obtenus à l'aide des méthodes usuelles d'estimation paramétrique que sont la méthode du maximum de vraisemblance (MV) et la méthode des moments (MM). Pour la loi géométrique, nous avons également donné des méthodes de construction d'un intervalle de confiance sur le paramètre.

Dans la plupart des cas, les estimateurs ne possèdent pas de forme explicite et requièrent une résolution numérique. Les propriétés des estimateurs sont alors obtenues à l'aide de simulations de Monte-Carlo : pour une taille d'échantillon n et une valeur des paramètres de la loi considérée fixés, nous avons simulé $N = 2000$ échantillons pour lesquels nous avons estimé à chaque fois les paramètres. Nous donnons les résultats sous forme de tableaux et de graphes pour les biais, variances et coefficients de variation empiriques des estimateurs. Les valeurs des paramètres ont été choisies de façon à obtenir des réalisations de la v.a. simulée de l'ordre de 10^k , pour $k \in \{1, 2, 3, 4\}$. Nous avons également choisi de simuler des échantillons de tailles $n \in \{5, 10, 20, 40, 60, 100, 200\}$ pour couvrir un grand nombre de cas pratiques.

6.1 La loi géométrique

On rappelle l'expression des probabilités élémentaires de la loi géométrique $\mathcal{G}(p)$:

$$p(k) = p(1-p)^{k-1}, \quad k \in \mathbb{N}^*$$

qui a pour espérance $\mathbb{E}(K) = \frac{1}{p}$ et pour variance $\text{Var}(K) = \frac{1-p}{p^2}$.

6.1.1 Estimation ponctuelle

Soit K_1, \dots, K_n un échantillon issu de la loi géométrique $\mathcal{G}(p)$. On note $S = \sum_{i=1}^n K_i$, notation qui sera utilisée dans toute cette section. La variable aléatoire S est de loi binomiale négative $\mathcal{BN}(n, p)$ dont les probabilités élémentaires sont données par :

$$\mathbb{P}(S = k) = C_{k-1}^{n-1} p^n (1-p)^{k-n} \quad k = n, n+1, \dots$$

La v.a S a pour espérance $\mathbb{E}(S) = \frac{n}{p}$ et pour variance $\mathbb{V}\text{ar}(S) = \frac{n(1-p)}{p^2}$.

Propriété 18

L'estimateur du MV de p est :

$$\hat{p}_n = \frac{n}{\sum_{i=1}^n K_i} = 1/\bar{K}_n$$

Il est biaisé et son biais est donné par :

$$\mathbb{E}(\hat{p}_n) = p^n {}_2F_1(n, n, n+1, 1-p)$$

où ${}_2F_1$ est la fonction hypergéométrique de Gauss.

Preuve :

La fonction de vraisemblance est :

$$\mathcal{L}(p; k_1, \dots, k_n) = \prod_{i=1}^n p(1-p)^{k_i-1} = p^n (1-p)^{\sum_{i=1}^n k_i - n}$$

Donc l'EMV est : $\hat{p}_n = \frac{n}{\sum_{i=1}^n K_i}$.

Le biais de l'estimateur \hat{p}_n se calcule alors de la façon suivante :

$$\begin{aligned} \mathbb{E}(\hat{p}_n) &= \mathbb{E}\left(\frac{n}{S}\right) = \sum_{s=n}^{\infty} \frac{n}{s} \mathbb{P}(S = s) = \sum_{s=n}^{\infty} \frac{n}{s} C_{s-1}^{n-1} p^n (1-p)^{s-n} \\ &= \sum_{i=0}^{\infty} \frac{n}{n+i} C_{n+i-1}^{n-1} p^n (1-p)^i = p^n \sum_{i=0}^{\infty} \frac{n(n+i-1)!}{(n+i)(n-1)!i!} (1-p)^i \\ &= p^n \frac{n}{(n-1)!} \sum_{i=0}^{\infty} \frac{\Gamma(n+i)\Gamma(n+i)}{\Gamma(n+i+1)} \frac{(1-p)^i}{i!} \\ &= p^n \frac{\Gamma(n+1)}{\Gamma(n)\Gamma(n)} \sum_{i=0}^{\infty} \frac{\Gamma(i+n)\Gamma(i+n)}{\Gamma(i+n+1)} \frac{(1-p)^i}{i!} = p^n \sum_{i=0}^{\infty} \frac{(n)_{(i)}(n)_{(i)}}{(n+1)_{(i)}} \frac{(1-p)^i}{i!} \\ &= p^n {}_2F_1(n, n, n+1, 1-p) \quad \square \end{aligned}$$

Propriété 19

L'estimateur sans biais de variance minimale de p est :

$$\tilde{p}_n = \frac{n-1}{\sum_{i=1}^n K_i - 1}$$

Preuve :

La statistique S est exhaustive et complète. On sait que $\mathbb{I}_{\{K_1=1\}}$ est un estimateur sans biais de p puisque :

$$\mathbb{E}(\mathbb{I}_{\{K_1=1\}}) = \mathbb{P}(K_1 = 1) = p$$

Donc l'ESBVM de p est obtenu par projection de l'estimateur sans biais $\mathbb{I}_{\{K_1=1\}}$ sur la statistique exhaustive et complète S , c'est à dire $\mathbb{E}[\mathbb{I}_{\{K_1=1\}}|S]$.

$$\begin{aligned} \mathbb{E}[\mathbb{I}_{\{K_1=1\}}|S = s] &= \mathbb{P}(K_1 = 1|S = s) = \frac{\mathbb{P}\left(K_1 = 1 \cap \sum_{i=1}^n K_i = s\right)}{\mathbb{P}(S = s)} \\ &= \frac{\mathbb{P}(K_1 = 1) \mathbb{P}\left(\sum_{i=2}^n K_i = s - 1\right)}{\mathbb{P}(S = s)} = \frac{p C_{s-2}^{n-2} p^{n-1} (1-p)^{s-1-(n-1)}}{C_{s-1}^{n-1} p^n (1-p)^{s-n}} \\ &= \frac{n-1}{s-1}. \quad \square \end{aligned}$$

Propriété 20

La variance de l'estimateur sans biais \tilde{p}_n est :

$$\text{Var}(\tilde{p}_n) = p^n {}_2F_1(n-1, n-1, n, 1-p) - p^2$$

Preuve :

$$\begin{aligned} \mathbb{E}(\tilde{p}_n^2) &= (n-1)^2 \mathbb{E}\left[\frac{1}{(\sum_{i=1}^n K_i - 1)^2}\right] \\ &= (n-1)^2 \sum_{s=n}^{+\infty} \frac{1}{(s-1)^2} C_{s-1}^{n-1} p^n (1-p)^{s-n} \\ &= (n-1)^2 p^n \sum_{k=0}^{+\infty} \frac{1}{(k+n-1)^2} C_{k+n-1}^{n-1} (1-p)^k \\ &= (n-1)^2 p^n \sum_{k=0}^{+\infty} \frac{1}{(k+n-1)^2} \frac{(k+n-1)!}{(n-1)!k!} (1-p)^k \\ &= p^n \sum_{k=0}^{+\infty} (n-1) \frac{1}{(k+n-1)} \frac{(n+k-2)!}{(n-2)!} \frac{(1-p)^k}{k!} \\ &= p^n \sum_{k=0}^{+\infty} \frac{\Gamma(n)}{\Gamma(n-1)} \frac{\Gamma(n+k-1)}{\Gamma(n+k)} \frac{\Gamma(n+k-1)}{\Gamma(n-1)} \frac{(1-p)^k}{k!} \\ &= p^n \frac{\Gamma(n)}{\Gamma(n-1)\Gamma(n-1)} \sum_{k=0}^{+\infty} \frac{\Gamma(n+k-1)\Gamma(n+k-1)}{\Gamma(n+k)} \frac{(1-p)^k}{k!} \\ &= p^n {}_2F_1(n-1, n-1, n, 1-p). \quad \square \end{aligned}$$

La variance de \tilde{p}_n est une fonction croissante puis décroissante en p comme le montre la figure FIG. 6.1.

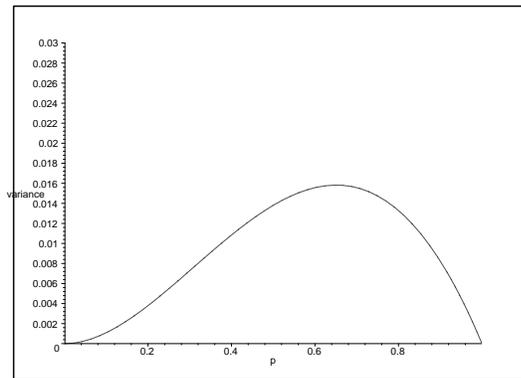


FIG. 6.1 – variance théorique de \tilde{p}_n pour un échantillon de taille 10

Nous allons maintenant étudier le coefficient de variation de \tilde{p}_n qui mesure la dispersion relative de \tilde{p}_n , contrairement à la variance qui est une mesure de dispersion absolue. Le coefficient de variation, noté CV , est défini comme le rapport de l'écart-type de la variable aléatoire considérée sur l'espérance de cette v.a.. Ainsi, le coefficient de variation est une grandeur sans dimension, ce qui permet de comparer les dispersions de v.a. d'ordres de grandeur très différents.

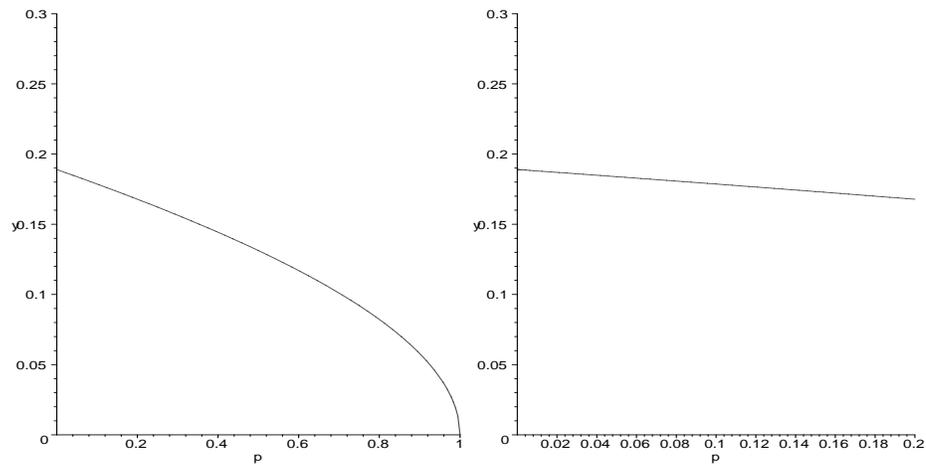
Pour l'estimateur \tilde{p}_n , le coefficient de variation est donné par :

$$CV(\tilde{p}_n) = \frac{\sqrt{\text{Var}(\tilde{p}_n)}}{\mathbb{E}(\tilde{p}_n)} = \sqrt{p^{n-2} {}_2F_1(n-1, n-1, n, 1-p) - 1}$$

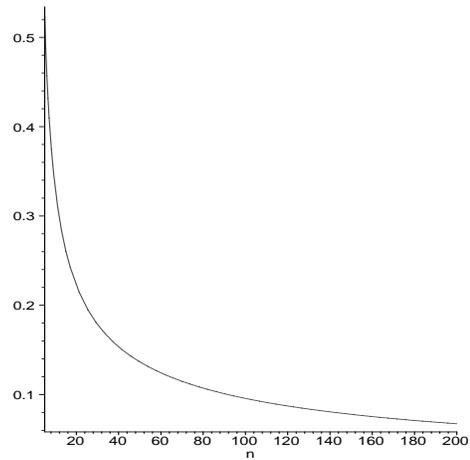
Un estimateur empirique du coefficient de variation, noté CV_n , peut être défini à partir d'un échantillon en remplaçant la moyenne et l'écart-type de la v.a. par leurs estimateurs empiriques respectifs. On considère généralement, et ce critère sera valable pour tout ce chapitre, qu'une estimation de paramètre est satisfaisante lorsque CV ou $CV_n \leq 15\%$ (voir [5]).

On notera que nous nous intéressons plus particulièrement aux petites valeurs de p ($p \leq 0.1$) pour des raisons applicatives.

À taille d'échantillon fixée, le coefficient de variation théorique de \tilde{p}_n est une fonction décroissante de p (voir le premier graphe de la figure FIG. 6.2), donc on estime d'autant mieux le paramètre p qu'il est proche de 1. Mais pour les petites valeurs de p , celles qui nous concernent, la décroissance de \tilde{p}_n est très faible comme le montre le second graphe de la figure FIG. 6.2.

FIG. 6.2 – coefficient de variation théorique de \tilde{p}_n pour $n = 30$

La figure FIG. 6.3 du coefficient de variation théorique de \tilde{p}_n montre que l'estimation de p est satisfaisante à partir d'échantillons de taille $n = 40$.

FIG. 6.3 – $CV(\tilde{p}_n)$ pour $p = 0.1$

Les résultats empiriques pour $CV_n(\tilde{p}_n)$, obtenus par simulation de Monte-Carlo dans les tableaux TAB. 6.1, 6.2 et 6.3 confirment ceux énoncés ci-dessus pour $CV(\tilde{p}_n)$.

valeur du paramètre p	0.1					
taille de l'échantillon	moyenne		variance		cv	
	empirique	théorique	empirique	théorique	empirique	théorique
5	9.95 -02	0.1	2.60 -03	2.75 -03	0.51	0.52
10	9.97 -02	0.1	1.11 -03	1.09 -03	0.33	0.33
20	1.00 -01	0.1	5.24 -04	4.94 -04	0.23	0.22
40	9.95 -02	0.1	2.43 -04	2.36 -04	0.16	0.15
60	1.00 -01	0.1	1.61 -04	1.55 -04	0.13	0.12
100	1.00 -01	0.1	8.98 -05	9.16 -05	0.10	0.10
200	9.99 -02	0.1	4.59 -05	4.54 -05	0.07	0.07

TAB. 6.1 – Résultats d'estimation du paramètre de la loi géométrique

valeur du paramètre p	0.01					
taille de l'échantillon	moyenne		variance		cv	
	empirique	théorique	empirique	théorique	empirique	théorique
5	1.02 -02	1.00 -02	3.86 -05	3.27 -05	0.61	0.57
10	1.00 -02	1.00 -02	1.25 -05	1.23 -05	0.35	0.35
20	9.97 -03	1.00 -02	5.41 -06	5.49 -06	0.23	0.23
40	1.00 -02	1.00 -02	2.76 -06	2.60 -06	0.17	0.16
60	9.99 -03	1.00 -02	1.64 -06	1.71 -06	0.13	0.13
100	9.97 -03	1.00 -02	1.01 -06	1.01 -06	0.10	0.10
200	1.00 -02	1.00 -02	4.85 -07	5.00 -07	0.07	0.07

TAB. 6.2 – Résultats d'estimation du paramètre de la loi géométrique

valeur du paramètre p	0.001					
taille de l'échantillon	moyenne		variance		cv	
	empirique	théorique	empirique	théorique	empirique	théorique
5	1.00 -03	1.00 -03	3.10 -07	3.33 -07	0.56	0.58
10	9.90 -04	1.00 -03	1.40 -07	1.25 -07	0.38	0.35
20	1.00 -03	1.00 -03	5.30 -08	5.55 -08	0.23	0.23
40	1.00 -03	1.00 -03	2.60 -08	2.63 -08	0.16	0.16
60	1.00 -03	1.00 -03	1.70 -08	1.72 -08	0.13	0.13
100	1.00 -03	1.00 -03	1.10 -08	1.02 -08	0.11	0.10
200	1.00 -03	1.00 -03	5.09 -09	5.05 -09	0.07	0.07

TAB. 6.3 – Résultats d'estimation du paramètre de la loi géométrique

Nous traitons également l'estimation ponctuelle pour la fiabilité de la loi géométrique dans cette section car c'est la seule loi pour laquelle on possède des résultats explicites.

La fiabilité de la loi géométrique est définie par $R(k) = (1 - p)^k$, $\forall k \in \mathbb{N}^*$. L'estimateur de maximum de vraisemblance de $R(k)$ est :

$$\widehat{R}(k) = (1 - \hat{p}_n)^k = \left(1 - \frac{n}{S}\right)^k$$

Seules les simulations permettront d'évaluer le biais de cet estimateur, car son expression est trop complexe pour le calculer formellement.

Propriété 21

L'estimateur sans biais de variance minimale $\widetilde{R}(k)$ de $R(k)$ est :

$$\widetilde{R}(k) = \begin{cases} \prod_{i=1}^{n-1} \left(1 - \frac{k}{S-i}\right) = \frac{C_{S-k-1}^{n-1}}{C_{S-1}^{n-1}} & \text{si } k \leq S - n \\ 0 & \text{si } k > S - n \end{cases}$$

où $S = \sum_{i=1}^n K_i$.

Preuve :

On sait que $\mathbb{I}_{\{K_1 > k\}}$ est un estimateur sans biais de $R(k)$, car $\mathbb{E}(\mathbb{I}_{\{K_1 > k\}}) = \mathbb{P}(K > k) = R(k)$. En projetant l'estimateur sans biais $\mathbb{I}_{\{K_1 > k\}}$ sur la statistique exhaustive et complète S , on obtient l'expression de l'ESBVM de $R(k)$ donné par $\widetilde{R}(k) = \mathbb{E} \left[\mathbb{I}_{\{K_1 > k\}} \mid \sum_{i=1}^n K_i \right]$.

Par définition, on a $\forall i \in \{1, \dots, n\}$, $K_i \geq 1$. Donc si $K_1 > k$, alors $\sum_{i=1}^n K_i > n - 1 + k$. Par conséquent, il faut que $s > n - 1 + k$, c'est à dire que $k \leq s - n$. Sous cette hypothèse, on a :

$$\begin{aligned} \mathbb{E} [\mathbb{I}_{\{K_1 > k\}} \mid S = s] &= \mathbb{P}(K_1 > k \mid S = s) = \frac{\mathbb{P} \left(K_1 > k \cap \sum_{i=1}^n K_i = s \right)}{\mathbb{P}(S = s)} \\ &= \frac{1}{C_{s-1}^{n-1} p^n (1-p)^{s-n}} \sum_{j=k+1}^{\infty} \mathbb{P} \left(K_1 = j \cap \sum_{i=2}^n K_i = s - j \right) \end{aligned}$$

La probabilité $\mathbb{P} \left(K_1 = j \cap \sum_{i=2}^n K_i = s - j \right)$ est nulle si $s - j < n - 1$, c'est à dire si $j > s - n + 1$, et elle est donc non nulle pour $j \leq s - n + 1$.

$$\begin{aligned}
\mathbb{E} [\mathbb{I}_{\{K_1 > k\}} | S = s] &= \frac{1}{C_{s-1}^{n-1} p^n (1-p)^{s-n}} \sum_{j=k+1}^{s-n+1} \mathbb{P}(K_1 = j) \mathbb{P} \left(\sum_{i=2}^n K_i = s-j \right) \\
&= \frac{1}{C_{s-1}^{n-1} p^n (1-p)^{s-n}} \sum_{j=k+1}^{s-n+1} p(1-p)^{j-1} C_{s-j-1}^{n-2} p^{n-1} (1-p)^{s-j-n+1} \\
&= \frac{1}{C_{s-1}^{n-1}} \sum_{j=k+1}^{s-n+1} C_{s-j-1}^{n-2} = \frac{1}{C_{s-1}^{n-1}} \sum_{i=0}^{s-n+k} C_{s-i-k-2}^{n-2} \\
&= \frac{1}{C_{s-1}^{n-1}} \sum_{j=0}^{s-n+k} C_{s-[s-n-k-j]-k-2}^{n-2} = \frac{1}{C_{s-1}^{n-1}} \sum_{j=0}^{s-n+k} C_{j+n-2}^{n-2}
\end{aligned}$$

$$\text{Or } \sum_{j=0}^m C_{j+k}^k = C_{m+k+1}^{k+1}$$

$$\begin{aligned}
\text{Donc } \mathbb{E} [\mathbb{I}_{\{K_1 > k\}} | S = s] &= \frac{C_{s-k-1}^{n-1}}{C_{s-1}^{n-1}} = \frac{(s-k-1)!(s-n)!}{(s-k-n)!(s-1)!} = \prod_{i=1}^{n-1} \frac{s-k-i}{s-i} \\
&= \prod_{i=1}^{n-1} \left(1 - \frac{k}{s-i} \right). \quad \square
\end{aligned}$$

6.1.2 Estimation par intervalle de confiance

a) Première méthode

Dans [17], Clemans donne un intervalle de confiance exact pour p . La construction de l'intervalle de confiance repose sur la variable aléatoire S de loi binomiale négative $\mathcal{BN}(n, p)$, dont la fonction de répartition est :

$$F_S(s|p) = \mathbb{P}(S \leq s) = \sum_{j=n}^s C_{j-1}^{n-1} p^n (1-p)^{j-n} \quad s = n, n+1, \dots$$

$F_S(s|p)$ est une fonction continue et croissante de p , quel que soit s . On a un intervalle de confiance $[p_1, p_2]$ de seuil α en trouvant (p_1, p_2) tel que $F_S(s|p_1) = \frac{\alpha}{2}$ et $F_S(s-1|p_2) = 1 - \frac{\alpha}{2}$, c'est à dire :

$$\begin{aligned}
\sum_{k=n}^s C_{k-1}^{n-1} p_1^n (1-p_1)^{k-n} &= \frac{\alpha}{2} \\
\sum_{k=n}^{s-1} C_{k-1}^{n-1} p_2^n (1-p_2)^{k-n} &= 1 - \frac{\alpha}{2}
\end{aligned}$$

b) Deuxième méthode : approximation normale

Par le théorème Central Limite (TCL), pour n assez grand, S est approximativement de loi normale $\mathcal{N}\left(\frac{n}{p}, \frac{n(1-p)}{p^2}\right)$. On a alors :

$$\mathbb{P}\left(\left|\frac{pS - n}{\sqrt{n(1-p)}}\right| \leq z_\alpha\right) \approx 1 - \alpha$$

$$\begin{aligned} \left|\frac{pS - n}{\sqrt{n(1-p)}}\right| \leq z_\alpha &\iff (pS - n)^2 \leq z_\alpha^2 n(1-p) \\ &\iff S^2 p^2 + (z_\alpha^2 - 2S)np + n(n - z_\alpha^2) \leq 0 \end{aligned}$$

Ce trinôme en p est toujours positif, sauf entre ses racines p'_1 et p'_2 ($p'_1 < p'_2$) qui sont les bornes de l'intervalle de confiance de seuil α :

$$\begin{aligned} p'_1 &= \frac{(2S - z_\alpha^2)n - \sqrt{nz_\alpha^2[nz_\alpha^2 + 4S(S - n)]}}{2S^2} \\ p'_2 &= \frac{(2S - z_\alpha^2)n + \sqrt{nz_\alpha^2[nz_\alpha^2 + 4S(S - n)]}}{2S^2} \end{aligned}$$

c) Troisième méthode : approche bayésienne

Les probabilités élémentaires de la loi géométrique $\mathcal{G}(p)$ sont données par :

$$\forall k \in \mathbb{N}^*, \quad f(k|p) = p(1-p)^{k-1}$$

La famille de lois a priori conjuguée à la loi géométrique est la famille des lois Beta de première espèce $Beta(a, b)$ (voir par exemple [76]) dont les probabilités élémentaires sont données par :

$$\forall k \in \mathbb{N}^*, \quad f(p) = \frac{p^{a-1}(1-p)^{b-1}}{\beta(a, b)} \mathbb{I}_{[0,1]}(p)$$

Le calcul de la densité a posteriori par la formule de Bayes donne :

$$\forall k \in \mathbb{N}^*, \quad f(p|k) = \frac{p^a(1-p)^{b-k-2}}{\beta(a+1, b+k-1)} \mathbb{I}_{[0,1]}(p)$$

qui est la densité de la loi Beta $\beta(a+1, b+k-1)$. Dans le cas d'un échantillon K_1, \dots, K_n issu de la loi géométrique $\mathcal{G}(p)$, la loi a posteriori du paramètre p est la loi Beta $\beta(a+n, b+s-n)$. On choisit de prendre $a = b = 1$ qui correspond à considérer une loi a priori de type non-informatif puisque la loi a priori Beta $\beta(1, 1)$ est égale à la loi uniforme sur $[0, 1]$. Mais on pourrait éventuellement prendre d'autres valeurs si on avait des informations a priori sur la valeur de p , par exemple quand on sait que le système est très fiable.

L'intervalle de confiance $[p''_1, p''_2]$ de seuil α sur le paramètre p s'obtient de la façon suivante :

$$\begin{aligned} p''_1 &= Beta_{(\alpha/2, n+1, S-n+1)} \\ p''_2 &= Beta_{(1-\alpha/2, n+1, S-n+1)} \end{aligned} \tag{6.1}$$

où $Beta_{(\alpha, \nu_1, \nu_2)}$ est le quantile d'ordre α de la loi Beta $\beta(\nu_1, \nu_2)$. Pour une utilisation pratique, on peut exprimer (6.1) à l'aide des quantiles de la loi de Fischer-Snedecor :

$$p_1'' = \frac{(n+1)\mathcal{F}_{(\alpha/2, 2(n+1), 2(S-n+1))}}{S-n+1 + (n+1)\mathcal{F}_{(1-\alpha/2, n+1, S-n+1)}}$$

$$p_2'' = \frac{(n+1)\mathcal{F}_{(1-\alpha/2, n+1, S-n+1)}}{S-n+1 + (n+1)\mathcal{F}_{(1-\alpha/2, 2(n+1), 2(S-n+1))}}$$

d) Comparaison

Des simulations ont été menées pour comparer la précision et la rapidité des trois méthodes précédentes de construction d'intervalle de confiance sur le paramètre de la loi géométrique. Pour une valeur du paramètre $p \in \{0.3, 0.1, 0.01\}$, nous avons simulé $N = 2000$ échantillons de loi géométrique $\mathcal{G}(p)$, de taille $n \in \{5, 10, 20, 40, 60, 100\}$. À partir de chaque échantillon simulé, nous avons construit un intervalle de confiance bilatéral de seuil $\alpha \in \{0.1, 0.05\}$ sur le paramètre p par les trois méthodes présentées à la section 6.1.2. Les tableaux TAB. 6.4 et 6.5 donnent le pourcentage des N intervalles de confiance construits qui ne contiennent pas la valeur p , respectivement pour un seuil de $\alpha = 0.1$ et $\alpha = 0.05$. Nous ne présentons pas les résultats concernant l'intervalle de confiance exact pour $p = 0.01$ car les calculs à effectuer sont extrêmement longs.

n \ p	Méthode exacte		Méthode bayésienne			Méthode avec TCL		
	0.3	0.1	0.3	0.1	0.01	0.3	0.1	0.01
5	7.0	9.0	10.1	10.5	9.0	8.7	8.0	7.0
10	9.9	10.8	11.3	10.3	8.9	9.7	9.6	9.1
20	7.9	9.5	9.3	9.3	10.6	9.9	10.3	8.8
40	8.4	8.9	9.9	9.5	9.8	9.5	10.1	9.1
60	9.0	8.5	10.2	10.3	11.0	9.5	9.5	10.1
100	9.7	9.5	9.9	10.3	9.5	9.1	9.0	9.8

TAB. 6.4 – Comparaison des constructions d'intervalle de confiance pour $\alpha = 0.1$

On peut constater dans les tableaux TAB. 6.4 et 6.5, que c'est pour la méthode exacte que les seuils empiriques sont les plus éloignés des seuils théoriques. Cela est sûrement dû à la résolution numérique des deux équations dont les bornes de l'intervalle sont solution. C'est également la raison pour laquelle cette méthode est la plus lente des trois étudiées. Dans l'ensemble, les seuils empiriques pour la méthode bayésienne sont les plus proches des seuils théoriques. La méthode utilisant le TCL donne de très bons résultats dès $n \geq 10$.

En conclusion, on recommande d'utiliser plutôt la méthode bayésienne, quelle que soit la taille de l'échantillon étudié.

n \ p	Méthode exacte		Méthode bayésienne			Méthode avec TCL		
	0.3	0.1	0.3	0.1	0.01	0.3	0.1	0.01
5	3.6	3.4	5.3	5.1	5.4	4.6	4.3	4.1
10	4.8	4.9	5.6	5.6	4.9	4.7	5.3	5.5
20	3.8	4.2	5.6	5.3	4.1	4.9	5.5	4.5
40	3.9	4.6	5.0	4.6	4.6	5.8	4.7	4.7
60	4.7	3.0	4.6	5.1	5.2	4.3	5.1	4.8
100	3.9	4.3	4.6	5.2	5.1	4.1	4.5	4.4

TAB. 6.5 – Comparaison des constructions d'intervalle de confiance pour $\alpha = 0.05$

6.1.3 Comparaison discret/continu

A priori, il est logique d'utiliser un modèle de fiabilité en temps discret quand le phénomène étudié est de nature discrète. Cependant, compte-tenu des difficultés liées à l'analyse statistique de ces modèles et à la méconnaissance des modèles de fiabilité en temps discret, il peut être tentant d'utiliser des modèles en temps continu, qui sont bien connus depuis longtemps. Il est donc intéressant d'évaluer l'erreur induite par cette approximation et de déterminer dans quels cas elle peut être acceptable. Remarquons tout d'abord qu'aucun modèle en temps continu ne permettra d'évaluer correctement la probabilité de défaillance à la première sollicitation, qui est une des grandeurs clés du problème. Mais les deux types de modèles peuvent fournir une estimation du MTTF et de la fiabilité.

Nous allons nous placer sous l'hypothèse d'absence de vieillissement. Soient K_1, \dots, K_n les durées de vie (en temps discret) de n systèmes identiques et indépendants. Le bon modèle en temps discret est la loi géométrique. L'approximation en temps continu est la loi exponentielle. Les deux modèles fournissent la même estimation du MTTF, à savoir la moyenne empirique. Mais leurs estimations de la fiabilité sont différentes. En effet, si les K_i , $i = 1, \dots, n$, sont supposés de loi géométrique, alors l'ESBVM de $R(k) = (1 - p)^k$ est:

$$\tilde{R}_{\mathcal{G}}(k) = \prod_{i=1}^{n-1} \left(1 - \frac{k}{S - i}\right) \mathbb{I}_{\{k < S - n\}}$$

Si les K_i , $i = 1, \dots, n$, sont supposés de loi exponentielle, alors l'ESBVM de $R(k) = e^{-\lambda k}$ est:

$$\tilde{R}_{\mathcal{E}}(k) = \left(1 - \frac{k}{S}\right)^{n-1} \mathbb{I}_{\{k < S\}}$$

À partir d'un échantillon issu de la loi géométrique, on souhaite connaître l'erreur commise quand on estime la fiabilité $R(k)$ par $\tilde{R}_{\mathcal{G}}(k)$ ou $\tilde{R}_{\mathcal{E}}(k)$. On sait que $\tilde{R}_{\mathcal{G}}(k)$ est sans biais et de variance minimale. Le biais de $\tilde{R}_{\mathcal{E}}(k)$ est trop complexe pour être calculé explicitement, mais il peut être estimé à l'aide de simulations de Monte-Carlo. Des échantillons de loi géométrique ont été simulés avec différentes valeurs de p ($p \in \{0.5, 0.1, 0.0001\}$) et des tailles

d'échantillons variant de $n = 5$ à $n = 200$. Les résultats des simulations montrent que l'erreur commise sur l'estimation de $R(k)$ par $\tilde{R}_{\mathcal{E}}(k)$ est quasiment indépendante de n . Nous avons alors fixé arbitrairement la taille de l'échantillon à 30; les figures FIG. 6.4 et 6.5 donnent les graphes des fonctions $R(k)$, $\tilde{R}_{\mathcal{G}}(k)$ et $\tilde{R}_{\mathcal{E}}(k)$ pour des échantillons de lois géométriques de paramètres respectifs 0.5, 0.1 et 0.0001. On constate que, quel que soit p , $\tilde{R}_{\mathcal{G}}(k)$ est un excellent estimateur de $R(k)$, ce qui est normal. Quand p est grand, l'erreur commise en estimant $R(k)$ par $\tilde{R}_{\mathcal{E}}(k)$ est importante, mais elle diminue avec p pour devenir négligeable quand $p = 0.0001$. Par conséquent, cela signifie que, en l'absence de vieillissement, l'erreur commise en utilisant un modèle en temps continu à la place d'un modèle en temps discret est significative pour des systèmes très peu fiables (p grand) et négligeable pour des systèmes très fiables (p petit). Autrement dit, pour des données de l'ordre de grandeur de la centaine, il est acceptable d'utiliser un modèle continu pour traiter des données discrètes. Il reste à déterminer si cette conclusion est encore valable en présence de vieillissement.

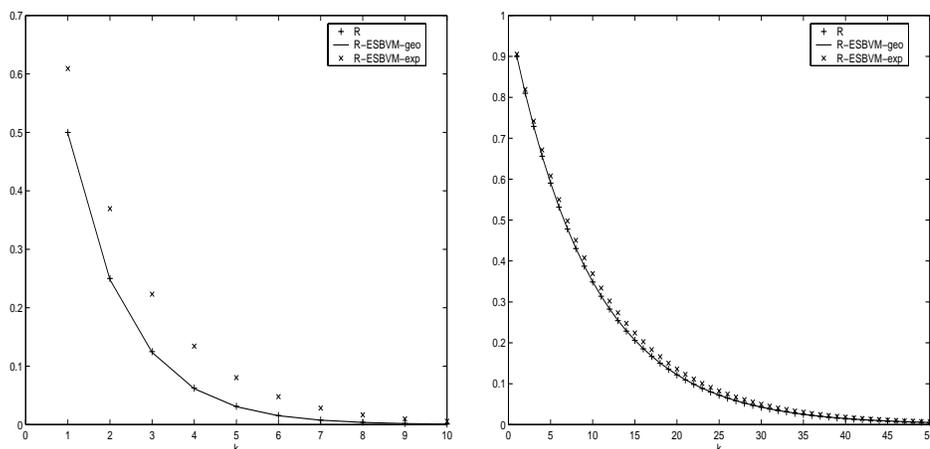


FIG. 6.4 – Graphes de $R(k)$, $\tilde{R}_{\mathcal{G}}(k)$ et $\tilde{R}_{\mathcal{E}}(k)$ pour $p = 0.5$ et $p = 0.1$

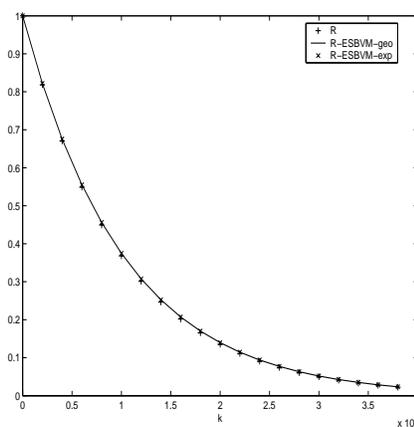


FIG. 6.5 – Graphes de $R(k)$, $\tilde{R}_{\mathcal{G}}(k)$ et $\tilde{R}_{\mathcal{E}}(k)$ pour $p = 0.0001$

6.2 La loi de Weibull de type I

On rappelle l'expression des probabilités élémentaires de la loi de $\mathcal{W}_I(q, \beta)$:

$$\forall k \in \mathbb{N}^*. \quad p(k) = q^{(k-1)^\beta} - q^{k^\beta}$$

Dans les simulations, le choix des valeurs de β (vitesse d'usure) correspond au désir d'avoir une loi IFR ($\beta > 1$) avec un vieillissement plus ou moins important ($\beta \in \{1.5, 2, 3\}$). Quant au choix des valeurs de q , il répond au désir d'obtenir des données de l'ordre de la dizaine, de la centaine, du millier et de la dizaine de milliers ($1 - q \in \{0.1, \dots, 10^{-6}\}$). On rappelle que les échantillons simulés sont de taille $n \in \{5, 10, 20, 40, 60, 100, 200\}$.

Dans [2], Ali Khan-Khalique-Abouammoh s'intéressent à l'estimation des paramètres de la loi de Weibull de type I. Les auteurs présentent une méthode simple, dite *méthode des proportions*, qu'ils comparent à une méthode des moments approchés.

La méthode des proportions pour la loi de Weibull de type I $\mathcal{W}_I(q, \beta)$ se base sur le fait que :

$$\mathbb{P}(K = 1) = 1 - q$$

Donc une estimation de q est le pourcentage d'observations de l'échantillon strictement supérieures à 1 :

$$\hat{q} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{K_i > 1\}} = \mathbb{R}_n(1)$$

De même, on a :

$$\begin{aligned} \mathbb{P}(K = 2) &= p(2) = q - q^{2^\beta} \\ \text{d'où } \beta &= \frac{1}{\ln 2} \{ \ln \ln (q - p(2)) - \ln \ln q \} \end{aligned}$$

On peut alors estimer β par :

$$\hat{\beta} = \frac{1}{\ln 2} \{ \ln \ln [\mathbb{R}_n(1) - (\mathbb{R}_n(1) - \mathbb{R}_n(2))] - \ln \ln \mathbb{R}_n(1) \} = \frac{1}{\ln 2} \ln \frac{\ln \mathbb{R}_n(2)}{\ln \mathbb{R}_n(1)}$$

On peut de la même façon utiliser deux entiers quelconques à la place de 1 et 2.

La méthode des moments n'est pas utilisable pour la loi de Weibull de type I car les expressions des deux premiers moments de la loi ne s'expriment que sous la forme d'une somme infinie. Les auteurs ont eu l'idée d'approcher les moments par des sommes partielles. Les estimateurs des moments approchés minimisent alors la somme des écarts quadratiques entre les moments empiriques et les sommes partielles.

La méthode des proportions nécessite un grand nombre d'observations ex-æquos, et plus précisément d'observations égales à 1 et 2 afin que les estimations empiriques de $p(1)$ et $p(2)$ soient non-nulles et proches des vraies valeurs. Cela n'est pas très souvent le cas dans le domaine d'application de Schneider Electric. D'autre part, cette méthode utilise seulement une petite partie des données pour estimer les paramètres, et on peut regretter de ne pas utiliser toute l'information contenue dans les observations, qui plus est dans le contexte industriel où les données sont souvent peu nombreuses. Enfin, la méthode des proportions est très sensible aux fluctuations de l'échantillon. Pour toutes ces raisons, nous n'avons pas utilisé cette méthode d'estimation, mais la méthode du maximum de vraisemblance.

Il s'avère que pour des données de l'ordre du millier ($1 - q \leq 0.0001$), l'estimation du paramètre q , obtenue par maximisation de la log-vraisemblance par la fonction `fmins` de MATLAB, se trouve hors de l'intervalle $]0,1[$. Nous avons alors décidé d'implémenter directement la méthode de Newton-Raphson. On obtient de bons résultats lorsque les valeurs d'initialisation des paramètres pour l'algorithme itératif de Newton-Raphson sont "proches" de la valeur finale des estimateurs du MV. Les valeurs initiales sont fournies par une méthode d'estimation graphique décrite à la section 7.1.2. Nous nous sommes aperçus que le problème de maximisation de la log-vraisemblance était mal conditionné lorsque la loi est paramétrée avec les paramètres q et β . En effet, plus les valeurs prises par la v.a. K sont grandes, plus le paramètre q est proche de 1, et par conséquent son estimation peut être supérieure à 1. En reparamétrant la loi de Weibull de type I comme la loi de Weibull continue, avec les paramètres η et β où $q = e^{-\left(\frac{1}{\eta}\right)^\beta}$, on ne rencontre plus ce problème car η est à valeurs dans \mathbb{R}^+ . Nous conserverons néanmoins, autant que possible, la paramétrisation avec q et β de la loi de Weibull de type I, afin d'éviter toute confusion avec la loi de Weibull continue.

Les résultats des simulations indiquent que l'estimation de q est légèrement biaisée (voir la figure FIG. 6.6) : q est sur-estimé pour $q < 0.9$ et sous-estimé pour $q \geq 0.9$. De plus, d'autres simulations non fournies ici semblent montrer que le biais de \hat{q} est indépendant de la valeur de β .

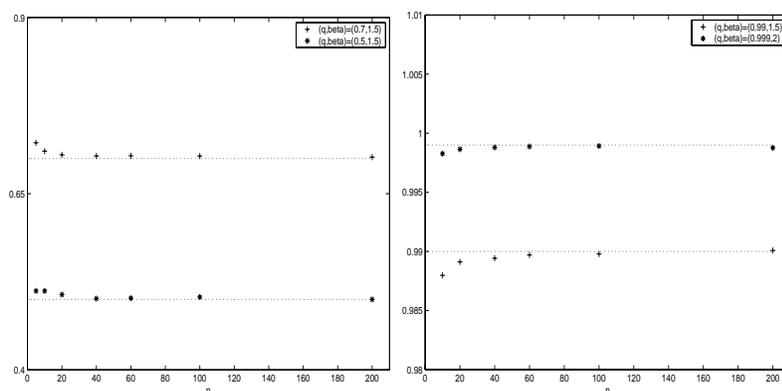


FIG. 6.6 – *Biais empirique de \hat{q}*

L'estimation de q est très satisfaisante, et cela dès que $n = 5$ pour $q \geq 0.9$, tandis que pour $q < 0.9$, l'estimation est satisfaisante à partir de $n = 20$ en général (voir la figure FIG. 6.7).

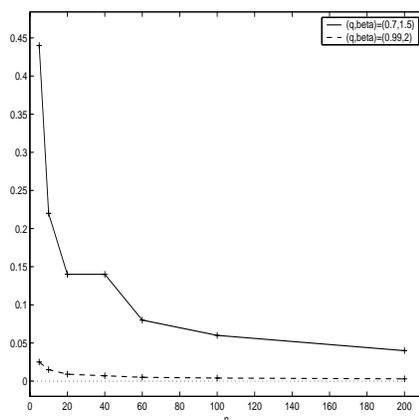


FIG. 6.7 – CV empirique de \hat{q}

L'estimation de β semble être légèrement biaisée comme l'indique le premier graphe de la figure FIG. 6.8. La convergence de $\hat{\beta}$ vers β est un peu plus lente que pour celle de \hat{q} vers q . En effet, l'estimation de β est satisfaisante pour $n \geq 40$ comme le montre le second graphe de la figure FIG. 6.8. À partir des nombreuses simulations effectuées, il apparaît que la qualité d'estimation de β semble indépendante de q , donc de l'ordre de grandeur des données. Ceci est normal, puisque β est un paramètre de forme.

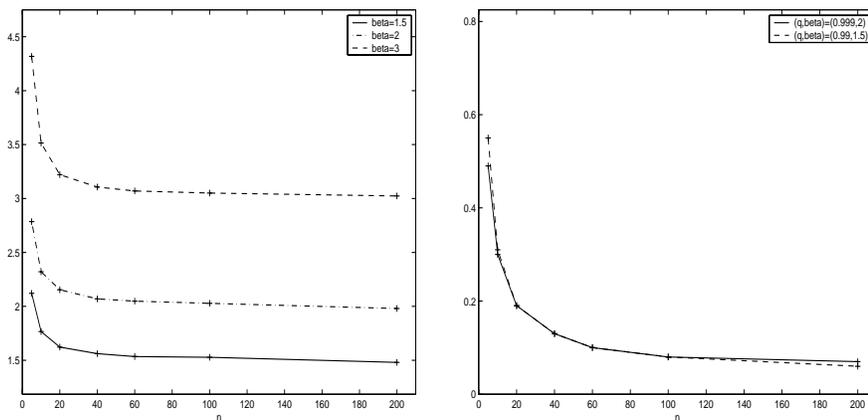


FIG. 6.8 – Biais et CV empiriques de $\hat{\beta}$

6.3 La loi de Weibull de type III

On rappelle l'expression des probabilité élémentaires :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \left(1 - e^{-ck^\beta}\right) e^{-c \sum_{i=1}^{k-1} i^\beta}.$$

La complexité de cette expression augmente considérablement les temps de calculs de la vraisemblance par rapport à la loi de Weibull de type I. Exceptés les cas où l'algorithme de maximisation de la vraisemblance ne converge pas à l'intérieur des intervalles de validité des paramètres, les simulations indiquent que l'estimation de c est peu biaisée (voir figure FIG. 6.9). La qualité d'estimation semble être décroissante avec c , et même pour $c = 0.1$, valeur la plus élevée parmi celles simulées, nos résultats montrent que l'estimation n'est pas satisfaisante même pour $n = 200$.

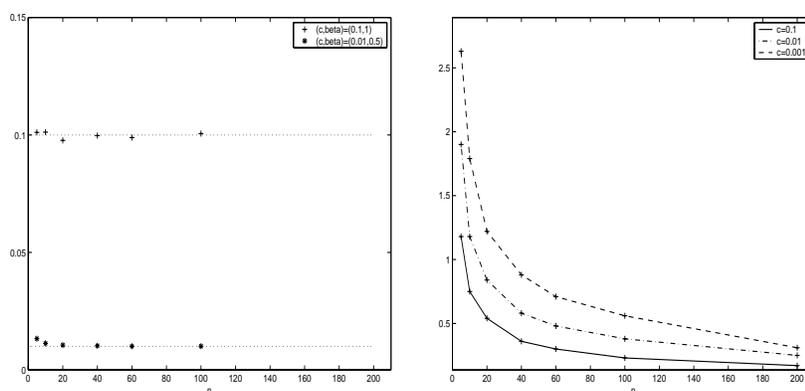


FIG. 6.9 – Biais et CV empiriques de \hat{c}

L'estimation du paramètre β semble biaisée (voir le premier graphe de FIG. 6.10). L'estimation de β n'est satisfaisante que pour des échantillons de taille supérieure ou égale à 150 comme le montre le second graphe de la figure FIG. 6.10. Les résultats des simulations semblent indiquer également que la qualité d'estimation est indépendante de la valeur de c .

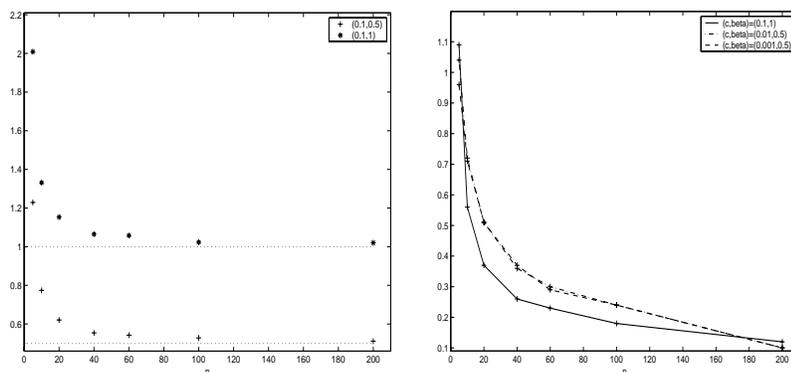


FIG. 6.10 – Biais et CV empiriques de $\hat{\beta}$

Nous avons eu des problèmes d'estimation pour les valeurs de c proches de zéro (i.e. pour les grandes valeurs de la v.a. K). On retrouve alors le même problème que pour la loi de Weibull de type I avec une estimation qui peut devenir négative pour le paramètre c . On peut alors penser à une reparamétrisation, qui reste à définir, qui permettrait d'avoir des paramètres à valeurs dans des intervalles plus grands. Il apparaît également un problème pour le paramètre β dont l'estimation pour les échantillons de taille 200 est parfois plus biaisée que pour les échantillons de taille inférieure.

6.4 La loi de Pólya inverse IFR

La méthode des moments n'est pas utilisable pour la loi de Pólya inverse IFR. En effet, l'expression du premier moment de la loi (voir la section 3.4.1) est trop complexe et ne permet pas de trouver explicitement les expressions des estimateurs qui résolvent les équations des moments. En revanche, les expressions des moments pour la loi de Pólya inverse DFR sont relativement simples et permettent ainsi l'utilisation de la méthode des moments pour estimer ses paramètres.

Pour la loi de Pólya inverse IFR, des problèmes apparaissent dans l'algorithme d'optimisation de la vraisemblance qui renvoie des estimations hors des intervalles de validité pour les échantillons de taille inférieure à 40. C'est la raison pour laquelle $n \geq 40$ sur les graphes des figures FIG. 6.11 et 6.12. Nos résultats se basent donc sur un faible nombre de simulations en raison de ces problèmes. Au vu de la figure FIG. 6.11, il semble que l'estimation de θ soit faiblement biaisée et satisfaisante à partir de $n = 40$. Les graphes de la figure FIG. 6.12 indiquent que l'estimation de δ semble fortement biaisée, et que son estimation n'est même pas satisfaisante pour $n = 200$.

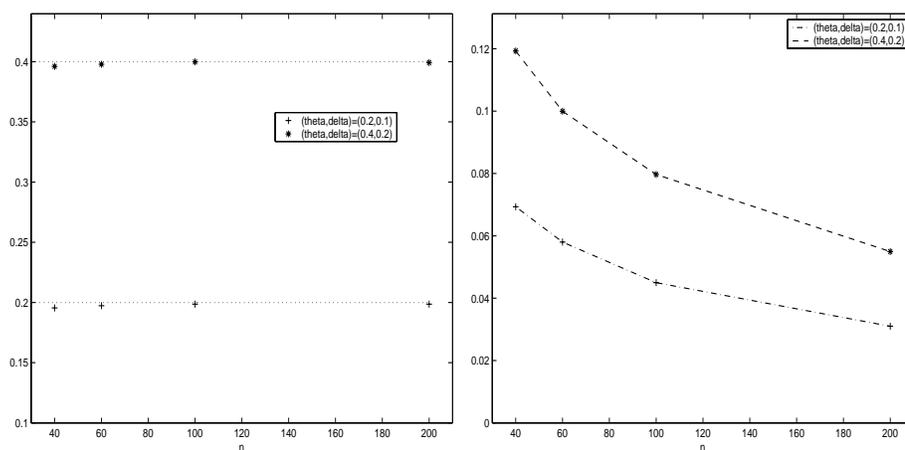
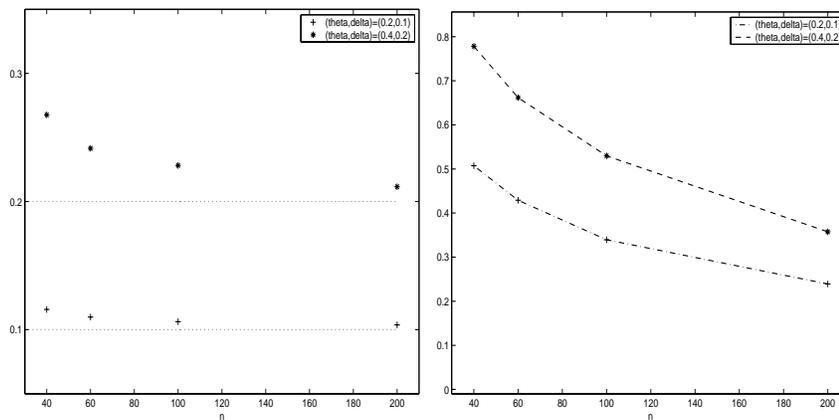


FIG. 6.11 – Biais et CV empiriques de $\hat{\theta}$

Clarotti-Lannoy-Proccacia [16] ont eu recours à la statistique bayésienne pour estimer les

FIG. 6.12 – Biais et CV empiriques de $\hat{\delta}$

paramètres θ et δ . Le loi de Pólya inverse IFR n'appartenant pas à la famille exponentielle, il n'existe pas de famille de loi a priori conjuguée à ce modèle. Les auteurs ont dû développer des techniques d'intégration numérique pour obtenir la loi a posteriori des paramètres.

6.5 La loi d'Éggenberger-Pólya

Dans leur article, Le Breton et Martin [52] utilisent deux méthodes d'estimation : une méthode graphique, que l'on verra en 7.1.5, et la méthode des moments explicitée ci-dessous.

Soit K une variable aléatoire de loi d'Éggenberger-Pólya de paramètres h et d . Alors les probabilités élémentaires sont données par :

$$\forall k \in \mathbb{N}^*, \quad p(k) = \frac{1}{(1+d)^{h/d}} \cdot \frac{(h/d)_{k-1}}{(k-1)!} \cdot \left(\frac{d}{d+1}\right)^{k-1}$$

La moyenne et la variance de la loi sont données par :

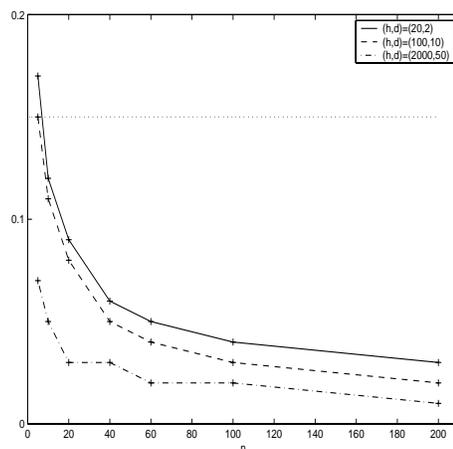
$$\begin{cases} \mathbb{E}(K) &= h + 1 \\ \text{Var}(K) &= h(d + 1) \end{cases}$$

La méthode des moments donne explicitement des estimateurs des paramètres :

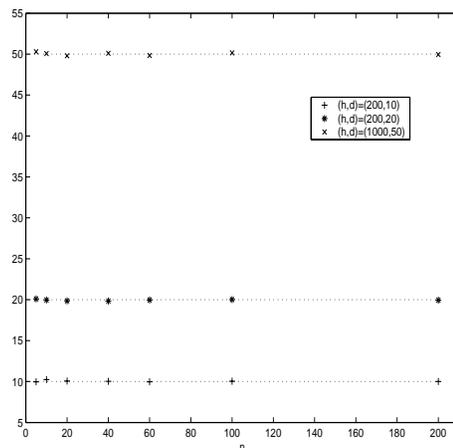
$$\begin{cases} \hat{h} &= \bar{K}_n - 1 \\ \hat{d} &= \frac{S_{n-1}^2}{\bar{K}_n - 1} - 1. \end{cases}$$

où S_n^2 est la variance empirique de l'échantillon.

\hat{h} est un estimateur sans biais de h , de variance $\frac{h(d+1)}{n}$. Le coefficient de variation théorique de l'estimateur \hat{h} est : $CV(\hat{h}) = \sqrt{\frac{d+1}{hn}}$. Les simulations confirment effectivement l'absence de biais dans l'estimation de h . La figure FIG. 6.13 illustre l'allure de $CV_n(\hat{h})$ et on constate que l'estimation est satisfaisante dès que $n \geq 10$.

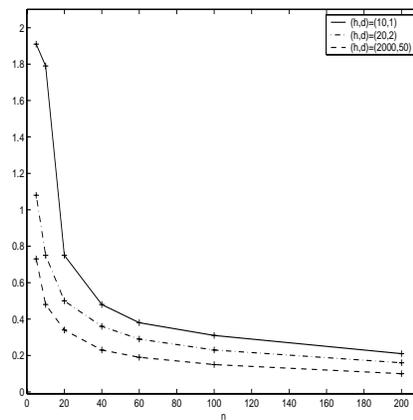
FIG. 6.13 – CV empirique de \hat{h} (par la méthode des moments)

L'estimation du paramètre d semble faiblement biaisée, et cela quel que soit l'ordre de grandeur du paramètre. La figure 6.14 donne quelques résultats tirés des simulations :

FIG. 6.14 – Biais empirique de \hat{d} (par la méthode des moments)

L'estimation n'est correcte que pour des échantillons de taille comprise entre 100 et 200 comme le montre la figure FIG. 6.15. La différence de qualité d'estimation entre h et d est due au fait que h est une moyenne et d est lié à la variance.

La méthode du MV n'est pas facilement utilisable pour l'estimation des paramètres de la loi en raison de la complexité des probabilités élémentaires. Les résultats d'estimation sont néanmoins aussi satisfaisants que pour la méthode des moments. La seule différence avec la méthode des moments concerne le temps de calcul qui est beaucoup plus long en raison de la résolution numérique des équations de MV.

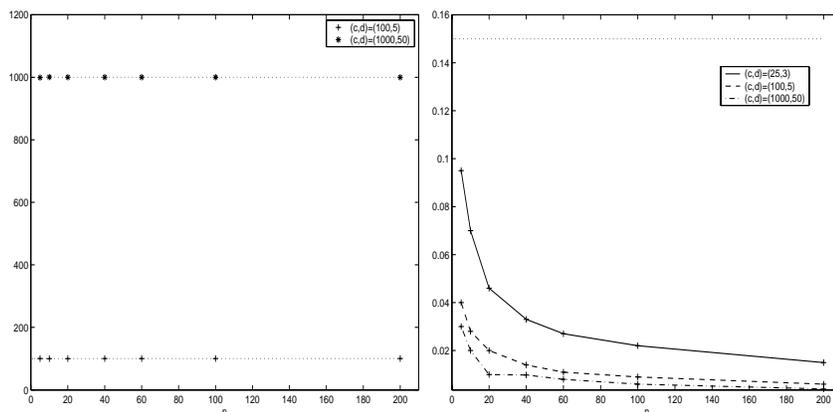
FIG. 6.15 – CV empirique de \hat{d} (par la méthode des moments)

6.6 La loi logistique tronquée discrétisée

On rappelle l'expression des probabilités élémentaires de la loi logistique tronquée discrétisée $\mathcal{L}og_d(c, d)$:

$$\forall k \in \mathbb{N}^*, \quad p(k) = \frac{e^{-\frac{k-1}{d}} (1 - e^{-\frac{1}{d}}) (1 + e^{\frac{c}{d}})}{(1 + e^{-\frac{k-c}{d}}) (1 + e^{-\frac{k-1-c}{d}})}$$

Les simulations montrent que l'estimation du paramètre c est très satisfaisante et semble faiblement biaisée, quel que soit l'ordre de grandeur de c et d comme le montre la figure FIG. 6.16 :

FIG. 6.16 – Biais et CV empiriques de \hat{c}

Pour le paramètre d , le coefficient de variation empirique dépend fortement de la valeur des deux paramètres c et d . En effet, pour les valeurs simulées $(c, d) = (1000, 25)$, l'estimation est satisfaisante à partir de $n = 40$, tandis que pour $(c, d) = (1000, 100)$, il faut que $n \geq 200$. La figure FIG. 6.17 met en évidence ce point.

L'estimation de d semble biaisée, comme nous le montre clairement le premier graphe de la

figure FIG. 6.17 ci-dessous. Bizarrement, le biais ne semble pas diminuer quand la taille de l'échantillon augmente :

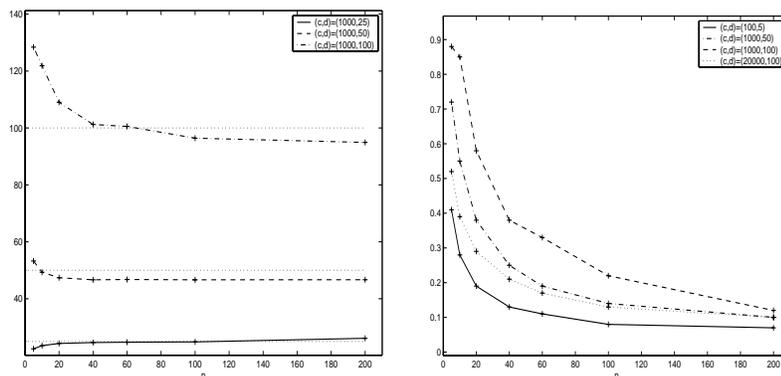


FIG. 6.17 – Biais et CV empiriques de \hat{d}

6.7 La loi S

Un problème numérique apparaît dans l'estimation du paramètre π de la loi $S(\pi, p)$. Il s'avère que pour des échantillons de petite taille, l'algorithme de maximisation renvoie systématiquement une estimation de π négative. Il faut alors avoir recours à des méthodes de maximisation de la vraisemblance avec contraintes, ce que fait la fonction `fmincon` de MATLAB. Malheureusement, l'optimisation de la vraisemblance par cette fonction `fmincon` ne trouve pas de meilleure solution que les valeurs initiales des paramètres fournies par la méthode graphique pourtant peu précise (voir la section 7.1.6).

6.8 Conclusion

La sélection des lois géométrique, Weibull de type I et d'Éggenberger-Pólya, basée sur des critères a priori (voir en fin du chapitre 3), est confirmée par ce chapitre d'estimation paramétrique. Les paramètres de la loi de Weibull de type I et celui de la loi géométrique sont très correctement estimés, même pour des échantillons de petite taille. Pour la loi d'Éggenberger-Pólya, si l'estimation du paramètre h est très satisfaisante, on notera que ce n'est pas le cas pour le paramètre d , même pour des échantillons de grande taille. À ces modèles, il faut ajouter la loi logistique tronquée discrétisée dont l'estimation des paramètres s'avère très satisfaisante, même pour les petits échantillons. La loi logistique tronquée discrétisée n'est pas aussi flexible que les lois de Weibull de type I et d'Éggenberger-Pólya puisqu'elle ne peut pas modéliser un rajeunissement. Mais en pratique, on observe beaucoup plus souvent des phénomènes de vieillissement, comme les applications pratiques chez Schneider Electric le prouvent.

Pour les lois de Pólya inverse IFR, S et Weibull de type III, des problèmes numériques doivent être résolus pour estimer les paramètres de ces lois lorsque ceux-ci sont proches des bornes de leur intervalle de validité.

Pour pouvoir effectuer une analyse statistique pertinente de données de retour d'expérience à l'aide des procédures paramétriques présentées au chapitre précédent, il est indispensable de pouvoir choisir des modèles adaptés à chaque jeu de données étudié.

Les tests d'adéquation, qu'ils soient graphiques ou statistiques, sont des outils mathématiques qui ne permettent pas de choisir le "meilleur" modèle, mais plutôt de rejeter un modèle inadéquat. La littérature dans ce domaine est très abondante en temps continu (voir par exemple [23]) contrairement au cas discret où la discontinuité des processus empiriques étudiés semble être la cause principale du faible nombre de travaux de recherche dans ce domaine.

Nous commençons par présenter des tests d'adéquation graphiques. Puis, nous présentons plusieurs types de tests d'adéquation statistiques : les tests du χ^2 , les tests basés sur la fonction de répartition empirique, la fonction génératrice empirique, les tests lisses et enfin les tests basés sur la transformation de Smirnov généralisée.

La plupart des tests d'adéquation aux lois discrètes ont été appliqués essentiellement à la loi de Poisson, et pratiquement jamais à la loi géométrique. Nous présentons ici une application détaillée de ces techniques à la loi géométrique et, dans une moindre mesure, à la loi de Weibull de type I. Les puissances des tests étudiés sont comparées à l'aide de simulations.

7.1 Tests graphiques d'adéquation

Le principe des tests graphiques d'adéquation d'un modèle à des observations repose sur l'existence d'une transformation fonctionnelle d'une des grandeurs de la fiabilité (R , F , λ , m , ...) du modèle qui admette une représentation graphique linéaire. La grandeur de fiabilité étant inconnue, on l'estime à l'aide d'un estimateur non paramétrique, et on vérifie l'adéquation des données au modèle en comparant l'estimateur empirique à une droite sur un papier à échelles fonctionnelles. Le graphe ainsi obtenu est appelé graphe de probabilité.

Remarque :

Pour certains modèles de fiabilité en temps discret, les expressions des grandeurs de fiabilité sont trop complexes pour trouver une transformation fonctionnelle qui admette une

représentation graphique linéaire, à moins que certains paramètres soient supposés connus. C'est le cas des lois logistique tronquée discrétisée et S .

7.1.1 La loi géométrique

L'expression de la fiabilité pour la loi géométrique $\mathcal{G}(p)$ est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = (1 - p)^k$$

En prenant le logarithme, on obtient :

$$\forall k \in \mathbb{N}^*, \quad \ln R(k) = k \ln(1 - p)$$

On trace le nuage de points $\left(k_{(j)}^*, \ln \mathbb{R}_n(k_{(j)}^*)\right)_{j=1, \dots, l} = \left(k_{(j)}^*, \ln\left(1 - \frac{r_j}{n}\right)\right)_{j=1, \dots, l}$ avec les notations du chapitre 5. Si les points du nuage sont approximativement alignés selon une droite de pente négative et passant par l'origine, on admettra que la loi géométrique est un modèle acceptable pour ces données. Dans ce cas, la pente de la droite fournit un estimateur de $\ln(1 - p)$ donc de p . Pour la loi exponentielle $\mathcal{E}(\lambda)$, on a :

$$\forall t \in \mathbb{R}^+, \quad \ln R(t) = -\lambda t$$

Donc le graphe de probabilité de la loi exponentielle pour des observations t_1, \dots, t_n est le nuage de points $\left(t_{(i)}, \ln \mathbb{R}_n(t_{(i)})\right)_{i=1, \dots, n} = \left(t_{(i)}, \ln\left(1 - \frac{i}{n}\right)\right)_{i=1, \dots, n}$.

Si dans le cas de la loi géométrique, il n'y a pas d'ex-æquos, le nuage de points est $\left(k_{(i)}, \ln\left(1 - \frac{i}{n}\right)\right)_{i=1, \dots, n}$. Donc les graphes de probabilités pour les lois géométrique et exponentielle sont identiques quand il n'y a pas d'ex-æquos. On peut se demander ce qui se passe si on teste graphiquement l'adéquation à la loi géométrique en traçant le graphe de probabilité de la loi exponentielle, même quand il y a des ex-æquos. Par exemple, pour les données de Xie-Goth décrites au chapitre 5, on obtient :

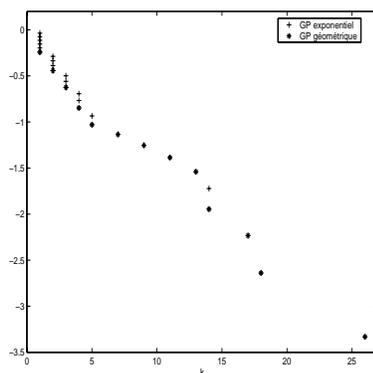


FIG. 7.1 – Graphes de probabilité des lois exponentielle et géométrique

Le fait d'avoir plusieurs ordonnées pour la même abscisse est perturbant pour juger de la linéarité du graphe de probabilité. De plus, si on estime le paramètre de la loi par régression linéaire sur ce nuage de points, on obtiendra des résultats différents dans les deux cas.

Par exemple, les tableaux TAB. 7.1 et 7.2 donnent les résultats d'estimation de ce paramètre à partir de $N = 2000$ simulations pour des échantillons de loi géométrique.

On constate que les biais des deux estimateurs sont comparables, mais que la variance de l'estimateur obtenu pour le graphe de la loi exponentielle est inférieure à celle obtenue pour le graphe de la loi géométrique. C'est probablement dû au fait que, à cause des ex-æquos, le graphe de probabilité de la loi géométrique possède moins de points que celui de la loi exponentielle. Donc, paradoxalement, il est peut être plus efficace d'utiliser le graphe de probabilité de la loi exponentielle que celui de la loi géométrique.

$p = 0.3$	Graphe de probabilité de la loi exponentielle			Graphe de probabilité de la loi géométrique		
	moyenne	variance	cv	moyenne	variance	cv
5	0.3840	2.2 -02	0.39	0.3428	2.8 -02	0.49
10	0.3491	1.1 -02	0.31	0.3455	1.5 -02	0.35
20	0.3318	5.5 -03	0.22	0.3309	7.8 -03	0.27
40	0.3196	2.5 -03	0.16	0.3208	4.1 -03	0.20
60	0.3142	1.7 -03	0.13	0.3162	3.2 -03	0.18
100	0.3101	1.1 -03	0.11	0.3122	2.2 -03	0.15
200	0.3064	5.8 -04	0.08	0.3086	1.5 -03	0.13

TAB. 7.1 – Estimation graphique du paramètre de la loi géométrique

$p = 0.1$	Graphe de probabilité de la loi exponentielle			Graphe de probabilité de la loi géométrique		
	moyenne	variance	cv	moyenne	variance	cv
5	0.1565	9.7 -03	0.63	0.1597	1.1 -02	0.67
10	0.1245	2.4 -03	0.39	0.1251	2.7 -03	0.42
20	0.1124	8.3 -04	0.26	0.1131	9.5 -04	0.27
40	0.1077	4.1 -04	0.19	0.1086	5.1 -04	0.21
60	0.1058	2.7 -04	0.16	0.1068	3.7 -04	0.18
100	0.1045	1.7 -04	0.12	0.1056	2.6 -04	0.15
200	0.1023	7.4 -05	0.08	0.1032	1.4 -04	0.12

TAB. 7.2 – Estimation graphique du paramètre de la loi géométrique

Nous nous sommes également intéressés à la puissance du test graphique d'adéquation à la loi géométrique. Nous avons simulé d'autres lois que la loi géométrique, puis nous avons tracé pour chaque échantillon simulé, le graphe de probabilité de la loi géométrique. On peut constater sur la figure FIG. 7.2, que les points des nuages ne semblent raisonnablement alignés que pour l'échantillon provenant effectivement de la loi géométrique, ce qui montre l'efficacité de la méthode. Remarquons cependant que, dans ces exemples, les valeurs des paramètres

des lois simulées ont été choisies de façon à ce que les lois soient nettement éloignées de la loi géométrique. Pour d'autres valeurs des paramètres (par exemple pour une loi de Weibull de type I avec β proche de 1), on n'obtient évidemment pas d'aussi bons résultats. Globalement, cette méthode est quand même performante pour détecter rapidement la présence ou pas d'un vieillissement.

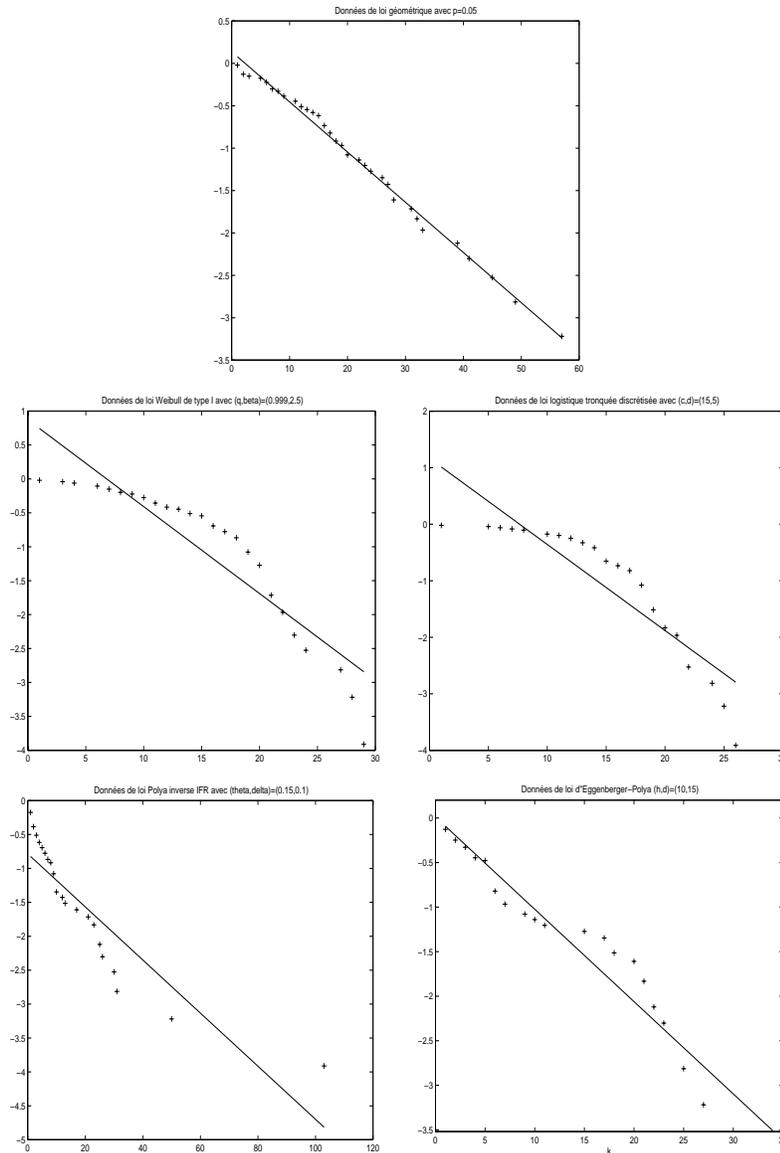


FIG. 7.2 – Graphes de probabilité de la loi géométrique pour des échantillons simulés de lois géométrique, Weibull de type I, logistique tronquée discrétisée, Pólya inverse IFR et Eggenberger-Pólya

7.1.2 La loi de Weibull de type I

L'expression de la fiabilité pour la loi de Weibull de type I $\mathcal{W}_I(q, \beta)$ est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = q^{k^\beta}, \quad \beta > 0, 1 > q > 0.$$

Après transformation log-logarithmique, on obtient :

$$\ln[-\ln R(k)] = \beta \ln k + \ln[-\ln q]$$

On trace le nuage de points $\left(\ln k_{(j)}^*, \ln[-\ln \mathbb{R}_n(k_{(j)}^*)] \right)_{j=1, \dots, l} = \left(\ln k_{(j)}^*, \ln[-\ln(1 - \frac{r_j}{n})] \right)_{j=1, \dots, l}$. Si les points du nuage sont approximativement alignés selon une droite de pente positive, on admettra que la loi de Weibull de type I est un modèle acceptable pour ces données. Dans ce cas, la pente et l'ordonnée à l'origine de la droite fournissent des estimateurs de β et q .

Pour la loi de Weibull continue $\mathcal{W}(\eta, \beta)$, on a :

$$\forall t \in \mathbb{R}^+, \quad R(t) = e^{-\left(\frac{t}{\eta}\right)^\beta}$$

Après transformation log-logarithmique, on obtient :

$$\ln[-\ln R(t)] = \beta \ln t - \beta \ln \eta$$

Le graphe de probabilité de la loi de Weibull pour des observations t_1, \dots, t_n est alors donné par $(\ln t_{(i)}, \ln[-\ln(1 - \frac{i}{n})])_{i=1, \dots, n}$.

Dans le cas où il n'y a pas d'ex-æquos dans les observations, les graphes de probabilité pour les lois de Weibull continue et Weibull discrète de type I sont identiques. Comme précédemment pour les lois exponentielles et géométrique, nous avons voulu juger l'effet, sur l'estimation des paramètres q et β , de l'utilisation du graphe de probabilité de la loi de Weibull continue pour des échantillons simulés de loi de Weibull discrète de type I. Les tableaux TAB. 7.3 et 7.5 donnent les résultats des simulations, que l'on peut comparer avec ceux des tableaux TAB. 7.4 et 7.6 donnant les résultats d'estimation graphique avec l'utilisation du graphe de probabilité de la loi de Weibull discrète de type I.

On peut remarquer que, comme dans le cas géométrique, la variance des estimateurs est meilleure pour le graphe de probabilité de la loi de Weibull continue. En revanche, le biais est nettement meilleur pour le graphe de probabilité de la loi de Weibull discrète de type I. On préférera alors utiliser le graphe de probabilité de la loi de Weibull discrète de type I.

$q = 0.9$ et $\beta = 1.5$	Graphe de probabilité de la loi de Weibull continue					
	q			β		
taille de l'échantillon	moyenne	variance	cv	moyenne	variance	cv
5	0.8766	8.4 -03	0.10	1.823	0.9804	0.54
10	0.9100	2.6 -03	0.06	1.705	0.2654	0.30
20	0.9291	8.8 -04	0.03	1.729	0.0827	0.17
40	0.9407	4.1 -04	0.02	1.788	0.0406	0.11
60	0.9445	2.6 -04	0.02	1.809	0.0268	0.09
100	0.9482	1.4 -04	0.01	1.832	0.0170	0.07
200	0.9503	6.3 -05	0.01	1.843	0.0076	0.05

TAB. 7.3 – Estimation graphique des paramètres de la loi de Weibull de type I

$q = 0.9$ et $\beta = 1.5$	Graphe de probabilité de la loi de Weibull discrète de type I					
	q			β		
taille de l'échantillon	moyenne	variance	cv	moyenne	variance	cv
5	0.8377	2.4 -02	0.19	1.749	1.2140	0.63
10	0.8780	6.1 -03	0.09	1.580	0.3039	0.35
20	0.8949	2.6 -03	0.06	1.561	0.0959	0.20
40	0.9010	1.5 -03	0.04	1.562	0.0571	0.15
60	0.9011	1.0 -03	0.04	1.543	0.0353	0.12
100	0.9018	5.7 -04	0.03	1.530	0.0195	0.09
200	0.9005	2.6 -04	0.02	1.513	0.0076	0.06

TAB. 7.4 – Estimation graphique des paramètres de la loi de Weibull de type I

$q = 0.9$ et $\beta = 2$	Graphe de probabilité de la loi de Weibull continue					
	q			β		
taille de l'échantillon	moyenne	variance	cv	moyenne	variance	cv
5	0.8847	8.3 -03	0.10	2.212	0.9132	0.43
10	0.9144	2.6 -03	0.06	2.191	0.4366	0.30
20	0.9358	9.1 -04	0.03	2.254	0.1554	0.17
40	0.9462	3.8 -04	0.02	2.325	0.0711	0.11
60	0.9500	2.4 -04	0.02	2.358	0.0511	0.10
100	0.9531	1.4 -04	0.01	2.384	0.0310	0.07
200	0.9558	6.2 -05	0.01	2.414	0.0150	0.05

TAB. 7.5 – Estimation graphique des paramètres de la loi de Weibull de type I

$q = 0.9$ et $\beta = 2$	Graphe de probabilité de la loi de Weibull discrète de type I					
	q			β		
taille de l'échantillon	moyenne	variance	cv	moyenne	variance	cv
5	0.7553	6.7 -02	0.34	1.907	1.1530	0.56
10	0.8712	7.8 -03	0.10	2.030	0.5222	0.36
20	0.8941	2.9 -03	0.06	2.058	0.1801	0.21
40	0.8999	1.7 -03	0.05	2.069	0.1120	0.16
60	0.9003	1.1 -03	0.04	2.049	0.0748	0.13
100	0.8997	7.0 -04	0.03	2.022	0.0402	0.10
200	0.8997	3.3 -04	0.02	2.010	0.0176	0.07

TAB. 7.6 – Estimation graphique des paramètres de la loi de Weibull de type I

Les tests d'adéquation graphiques à la loi de Weibull de type I effectués sur des échantillons simulés de lois de Pólya inverse, d'Engenberger-Pólya et de loi logistique tronquée discrétisée donnent malheureusement des représentations graphiques approximativement linéaires comme on peut le constater sur la figure FIG. 7.3. Le graphe de probabilité ne permet pas de reconnaître que des données issues de ces autres modèles ne proviennent pas de la loi de Weibull de type I.

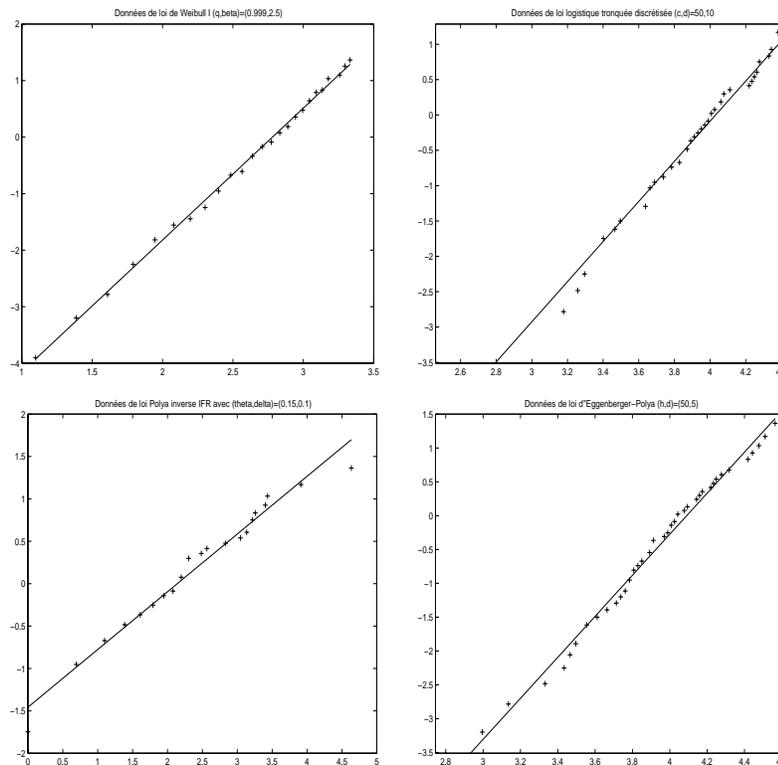


FIG. 7.3 – Graphes de probabilité de la loi de Weibull de type I pour des échantillons simulés de lois de Weibull I, logistique tronquée discrétisée, Pólya inverse IFR et Engenberger-Pólya

7.1.3 La loi de Weibull de type III

L'expression de la fiabilité de la loi de Weibull de type III $\mathcal{W}_{III}(c, \beta)$ est :

$$\forall k \in \mathbb{N}^*, \quad R(k) = e^{-c \sum_{j=1}^k j^\beta}$$

Aucune transformation de cette expression ne permet d'obtenir une représentation graphique linéaire, alors on a essayé d'utiliser plutôt le taux de défaillance qui est donné par :

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = 1 - e^{-ck^\beta}$$

Après transformation log-logarithmique, on obtient :

$$\ln[-\ln(1 - \lambda(k))] = \beta \ln k + \ln c$$

On trace le nuage de points $\left(\ln k_{(j)}^*, \ln[-\ln(1 - \lambda_n(k_{(j)}^*))] \right)_{j=1, \dots, l} = \left(\ln k_{(j)}^*, \ln[-\ln \frac{n_j}{n_j-1}] \right)_{j=1, \dots, l}$. Si les points du nuage sont approximativement alignés selon une droite de pente positive et d'ordonnée à l'origine négative, on admettra que la loi de Weibull de type III est un modèle acceptable pour ces données.

Le graphe de probabilité de la loi de Weibull de type III est basé sur le taux de défaillance empirique et non pas sur la fiabilité empirique comme c'est le cas pour les lois géométrique et Weibull de type I. On a vu au chapitre 5 que le taux de défaillance empirique était un moins bon estimateur que ne l'est la fiabilité empirique, et c'est sans doute la raison pour laquelle les tests d'adéquation graphique sont moins performants pour ce modèle comme on peut le constater sur la figure FIG. 7.4 : bien que les données soient simulées selon le modèle de Weibull de type III, les points du graphe de probabilité sont loin d'être approximativement alignés.

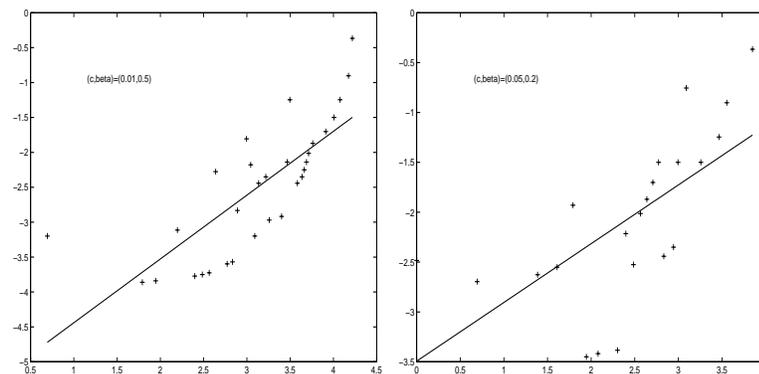


FIG. 7.4 – Graphes de probabilité de deux échantillons de taille 50 de loi de Weibull de type III

Par conséquent, on déconseillera la méthode du graphe de probabilité pour la loi de Weibull de type III.

7.1.4 La loi de Pólya inverse IFR

L'expression du taux de défaillance de la loi de de Pólya inverse IFR est :

$$\lambda(k) = \frac{\theta + (k-1)\delta}{1 + (k-1)\delta}, \quad 0 < \theta < 1, \quad 0 < \delta$$

Après quelques manipulations, on obtient :

$$\frac{1}{1 - \lambda(k)} = \frac{(k-1)\delta + 1}{1 - \theta}$$

On trace le nuage de points $\left(k_{(j)}^*, \frac{1}{1 - \lambda_n(k_{(j)}^*)}\right)_{j=1, \dots, l}$. Si les points du nuage sont approximativement alignés selon une droite de pente positive et d'ordonnée à l'origine dans $[0, 1]$, on admettra que la loi de Pólya inverse IFR est un modèle acceptable pour ces données.

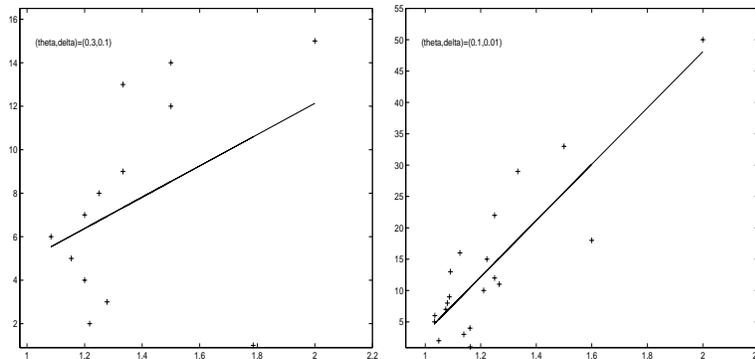


FIG. 7.5 – Graphes de probabilité de deux échantillons de taille 50 de loi Pólya inverse IFR

Comme on peut le constater sur la figure FIG. 7.5, les tests graphiques donnent de très mauvais résultats, principalement dus aux piètres performances de l'estimateur empirique du taux de défaillance.

Comme pour la loi de Weibull de type III, on déconseille cette méthode pour la loi de Pólya inverse IFR.

7.1.5 La loi d'Engenberger-Pólya

Les expressions des grandeurs R , λ et m pour la loi d'Engenberger-Pólya sont trop complexes pour pouvoir trouver une transformation fonctionnelle qui permette une représentation graphique linéaire. En 3.4.5, on a calculé le rapport de deux probabilités successives de la loi d'Engenberger-Pólya $EP(h, d)$:

$$\frac{p(k+1)}{p(k)} = \frac{d}{1+d} + \frac{h-d}{1+d} \frac{1}{k}, \quad \forall k \in \mathbb{N}^*$$

On estime les probabilités $p(k)$ par les fréquences empiriques $\mathbb{P}_n(k)$ correspondantes de l'échantillon observé. Avec les notations du chapitre 5, on a :

$$\frac{\mathbb{P}_n(k_{(j)}^* + 1)}{\mathbb{P}_n(k_{(j)}^*)} = \begin{cases} \frac{d_{j+1}}{d_j} & \text{si } k_{(j+1)}^* = k_{(j)}^* \\ 0 & \text{sinon} \end{cases}$$

On trace alors le nuage de points $\left(\frac{1}{k_{(j)}^*}, \frac{d_{j+1}}{d_j}\right)_{j=1, \dots, l}$, si les points du nuage sont approximativement alignés, on admettra que la loi d'Engenberger-Pólya est un modèle acceptable pour ces données. Cette méthode, introduite par Rao [72], a été utilisée par Le Breton et Martin en climatologie [52].

Remarque :

Cette méthode graphique présente un inconvénient similaire à la méthode d'estimation des proportions car elle nécessite que l'échantillon contienne de nombreux ex-æquos : si l'échantillon ne contient pas d'ex-æquo, les rapports de probabilités empiriques valent tous 1 ou 0. Les autres méthodes graphiques présentées ici ne nécessitent pas nécessairement d'avoir des ex-æquos dans l'échantillon étudié.

7.1.6 La loi S

On rappelle l'expression du taux de défaillance de la loi S $S(\pi, p)$:

$$\forall k \in \mathbb{N}^*, \quad \lambda(k) = p(1 - \pi^k)$$

Une estimation graphique des paramètres à partir d'une expression de fiabilité de cette loi n'est possible qu'en fixant un des paramètres. Le paramètre p représentant la probabilité que l'appareil subisse un stress lors d'une sollicitation, on pourra par exemple le fixer à 1 dans certains cas.

On obtient alors :

$$\forall k \in \mathbb{N}^*, \quad \ln(1 - \lambda(k)) = k \ln \pi$$

On trace le nuage de points $\left(k_{(j)}^*, \ln(1 - \lambda_n(k_{(j)}^*))\right)_{j=1, \dots, l}$. Si les points du nuage sont approximativement alignés selon une droite de pente négative et passant par l'origine, on admettra que la loi S avec $p = 1$ est un modèle acceptable pour ces données.

7.2 Tests statistiques d'adéquation : Introduction et notations

Soit K, K_1, \dots, K_n des variables aléatoires indépendantes et identiquement distribuées, à valeurs dans \mathbb{N}^* . La fonction de répartition F de K déterminant de façon unique la loi

associée, nous les assimilons dans le reste du chapitre.

Soit $\mathcal{F}_\Theta = \{F(\cdot; \theta) : \theta \in \Theta\}$ une famille paramétrée de lois discrètes. Soit $d = \dim \theta$.

Un test d'adéquation d'un échantillon K_1, \dots, K_n à une loi discrète complètement spécifiée $F(\cdot; \theta_0) = F_0(\cdot)$ (i.e. $\theta = \theta_0$ connu), consiste à tester :

$$\mathcal{H}_0 : F = F_0 \text{ contre } \mathcal{H}_1 : F \neq F_0.$$

Un test d'adéquation d'un échantillon K_1, \dots, K_n à une famille paramétrée de lois, consiste à tester :

$$\mathcal{H}_0 : F \in \mathcal{F}_\Theta \text{ contre } \mathcal{H}_1 : F \notin \mathcal{F}_\Theta.$$

7.2.1 Test du χ^2

Le test du χ^2 est sûrement le test le plus répandu pour tester l'adéquation d'une loi, continue ou discrète, à des données. Dans notre contexte industriel, le manque de données en nombre suffisant ne permet pas souvent l'utilisation de ce test. C'est la raison pour laquelle nous nous limitons à une présentation succincte des tests du χ^2 de Pearson et de Nikulin-Rao-Robson qui ne reflète pas l'importance de ce test en général (voir par exemple Greenwood et Nikulin [30]).

a) Test du χ^2 de Pearson

Le principe du test du χ^2 de Pearson ne diffère pas selon que la loi à tester soit continue ou discrète.

On répartit les observations en r classes. On notera par N_1, \dots, N_r et $p_1(\theta), \dots, p_r(\theta)$, respectivement les effectifs et les probabilités de chacune des classes sous \mathcal{H}_0 . On a donc $\mathbb{E}(N_i) = np_i(\theta)$ sous l'hypothèse \mathcal{H}_0 . On considère alors la statistique τ_n^2 définie par :

$$\tau_n^2 = \sum_{j=1}^r \frac{(N_j - np_j(\theta))^2}{np_j(\theta)}$$

Dans le cas d'un test d'adéquation à une loi discrète complètement spécifiée, i.e. $\theta = \theta_0$ connu et les $p_i(\theta_0)$ également connus pour $i = 1, \dots, r$, la statistique τ_n^2 suit asymptotiquement une loi du χ^2 à $r - 1$ degrés de liberté sous l'hypothèse \mathcal{H}_0 . Pour le cas du test d'adéquation à une famille paramétrée de lois, il faut alors estimer le vecteur θ des paramètres inconnus. Si on note $\hat{\theta}_n$ l'estimateur du MV et $\hat{p}_i = p_i(\hat{\theta}_n)$, $i = 1, \dots, r$, la probabilité estimée d'appartenir à la classe i , la statistique de test devient alors :

$$\hat{\tau}_n^2 = \sum_{j=1}^r \frac{(N_j - n\hat{p}_j)^2}{n\hat{p}_j}$$

Si les estimations du MV des paramètres inconnus sont basées sur les données groupées dans les classes, alors le nombre de degrés de liberté du χ^2 devient $r - 1 - d$. En revanche, si l'estimation du MV est effectuée sur les données brutes, la loi asymptotique de $\hat{\tau}_n^2$ est comprise

entre une loi χ_{r-1}^2 et χ_{r-1-d}^2 .

En pratique, le choix du nombre r de classes est souvent donné par $r \simeq 2n^{\frac{2}{5}}$. On choisit ensuite les bornes des classes de façon à avoir des classes le plus équiprobables possibles. On admet que l'effectif pour chaque classe doit être supérieur ou égal à 5 pour que l'on puisse considérer que la loi de τ_n^2 est approximativement de loi χ_{r-1}^2 . On remarquera que les classes doivent être construites a priori, c'est à dire indépendamment des observations. Mais lorsque la loi supposée des observations n'est pas connue, il en va généralement de même de l'ordre de grandeur des observations et donc de la construction des classes. Lorsque la construction des classes dépend des observations, la loi limite de $\widehat{\tau}_n^2$ sous \mathcal{H}_0 est une loi du χ^2 à laquelle s'ajoute une combinaison linéaire de lois χ_1^2 avec des coefficients dépendant de la vraie valeur des paramètres. On peut remarquer dans ce cas, que si les paramètres estimés sont des paramètres de forme ou d'échelle, la loi asymptotique de $\widehat{\tau}_n^2$ ne dépend pas de ces paramètres.

b) Test du χ^2 avec la statistique modifiée de Nikulin-Rao-Robson

On utilise les notations du paragraphe précédent, auxquelles on ajoute la matrice d'information de Fisher $J(\theta)$, et la matrice $B(\theta)$ dont les coefficients sont $p_i(\theta) - \frac{1}{2} \frac{\partial p_i(\theta)}{\partial \theta_i}$.

La statistique de Nikulin-Rao-Robson est donnée par :

$$\widehat{NRR}_n = V(\widehat{\theta}_n)^t Q(\widehat{\theta}_n) V(\widehat{\theta}_n),$$

où

- $\widehat{\theta}_n$ est l'estimateur du MV de θ sur les données brutes,
- $V(\theta)$ est le vecteur de taille r qui a pour composantes $[N_i - np_i(\theta)] / (np_i(\theta))^{\frac{1}{2}}$,
- $Q(\theta) = I_r + B(\theta)[J(\theta) - B(\theta)^t B(\theta)]^{-1} B(\theta)^t$,
- I_r est la matrice identité d'ordre r .

En 1973, Nikulin ([62], [63], [64] et [65]) a étudié cette statistique pour une loi continue quelconque et Rao-Robson [73] l'ont proposé en 1974 pour la famille des lois exponentielles. Nikulin a démontré que la statistique \widehat{NRR}_n , suivait asymptotiquement sous \mathcal{H}_0 la loi du χ^2 à $r - 1 - d$ degrés de liberté. On notera que ce résultat est valable quel que soit le choix des bornes des classes. En pratique, on veillera à utiliser ce test avec des classes le plus équiprobables possible. C'est là que se trouve la différence essentielle entre les lois discrètes et les lois continues : dans le cas des lois discrètes, la présence d'ex-æquos rend souvent impossible l'obtention de classes équiprobables.

En pratique, pour appliquer efficacement les tests du χ^2 de Pearson et de Nikulin-Rao-Robson, il est nécessaire d'avoir au moins 50 observations. Comme nous le précisons ci-

dessus, nous leur préférerons des tests plus puissants et qui ne nécessitent pas un nombre trop important de données.

7.2.2 Tests basés sur la fonction de répartition empirique

a) Rappels : en temps continu

Le test de Kolmogorov-Smirnov (KS) est très largement utilisé pour tester l'hypothèse qu'un échantillon est issu d'une loi continue spécifiée ou d'une famille de lois. Soit X_1, \dots, X_n un échantillon de taille n de variables aléatoires continues. On souhaite tester :

$$\mathcal{H}_0 : F = F_0 \text{ contre } \mathcal{H}_1 : F \neq F_0$$

où F est la fonction de répartition inconnue de l'échantillon, et $F_0(\cdot) = F(\cdot; \theta_0)$ est une fonction de répartition complètement spécifiée (i.e. θ_0 connu).

Soit \mathbb{F}_n la fonction de répartition empirique des observations. Si l'on souhaite tester l'hypothèse \mathcal{H}_0 contre l'alternative \mathcal{H}_1 , la statistique du test de KS est donnée par :

$$K_n = \sqrt{n} \sup_x |\mathbb{F}_n(x) - F_0(x)|.$$

Sous \mathcal{H}_0 , la statistique K_n est asymptotiquement libre, c'est à dire que sa loi ne dépend pas de F_0 . Les quantiles de cette loi ont été tabulés pour différentes valeurs du niveau α , $\alpha \in]0, 1[$. Dans le cas où le vecteur des paramètres θ est inconnu et doit être estimé par un estimateur $\hat{\theta}_n$, la loi de la statistique \hat{K}_n (où $\hat{\theta}_n$ remplace θ_0) dépend, sous \mathcal{H}_0 , de la famille de lois testée. D'Agostino et Stephens [23] donnent les quantiles des statistiques pour différentes familles de lois, qui permettent de faire des tests asymptotiques de niveau α .

Le test de Kolmogorov-Smirnov porte sur la mesure du maximum de la différence entre $\mathbb{F}_n(x)$ et $F(x)$. Il existe une seconde classe de mesures d'écart donnée par la statistique quadratique suivante :

$$Q_n^2 = n \int_{-\infty}^{+\infty} (\mathbb{F}_n(x) - F(x))^2 \psi(x) dF(x)$$

où la fonction $\psi(x)$ est une fonction pondératrice. Lorsque $\psi(x) \equiv 1$, la statistique est appelée statistique de Cramér-von Mises, notée W_n^2 , et lorsque $\psi(x) = 1/\{F(x)(1 - F(x))\}$, la statistique est appelée statistique d'Anderson-Darling, notée A_n^2 .

Comme pour le test de Kolmogorov-Smirnov, les lois limites des statistiques W_n^2 et A_n^2 sous \mathcal{H}_0 sont libres et des tables de leurs quantiles permettent ainsi de construire des tests asymptotiques de niveau α voulu. Lorsque le vecteur des paramètres θ est inconnu et doit être estimé par un estimateur $\hat{\theta}_n$, les lois limites des statistiques \hat{A}_n^2 , \hat{W}_n^2 (où $\hat{\theta}_n$ remplace θ_0) dépendent, sous \mathcal{H}_0 , de la famille de lois testée, mais pas de θ , si les paramètres sont de position et d'échelle (voir [25] et [24]).

b) Test d'adéquation à des lois discrètes

Une première batterie de résultats en temps discret concerne le test d'adéquation de Kolmogorov-Smirnov à une loi entièrement spécifiée. La statistique de test est :

$$KS_n = \sqrt{n} \max_k |\mathbb{F}_n(k) - F(k, \theta_0)| \quad (7.1)$$

Schmid [81] a donné la loi limite de KS_n sous \mathcal{H}_0 quand F_0 admet un nombre fini de discontinuités. Pour une loi quelconque, Conover a donné une procédure pour calculer approximativement les quantiles de la loi de KS_n sous \mathcal{H}_0 , pour n fini ($n \leq 30$). Pettitt-Stephens [69] ont donné une table de quantiles exacts pour cette loi, en se plaçant dans le cas de données issues de groupement de données continues en classes. Ces auteurs, ainsi que Horn [38] ont comparé sur des exemples le test de Kolmogorov-Smirnov et le test du χ^2 et ont conclu que le test de Kolmogorov-Smirnov était le plus puissant. En particulier, il est plus performant que le χ^2 quand les données présentent un ordre naturel, ce qui est le cas des données de fiabilité. Enfin, Wood-Altavela [3] ont donné la loi limite de KS_n sous \mathcal{H}_0 pour n'importe quelle loi discrète.

Il ressort de tous ces travaux que la principale différence entre le cas continu et le cas discret réside dans le fait que la loi de KS_n dépend de F_0 pour les lois discrètes, alors qu'elle n'en dépendait pas pour les lois continues.

Choulakian, Lockart et Stephens [15] se sont intéressés aux tests de Cramér-von Mises dans le cas de l'adéquation à une loi entièrement spécifiée et à support fini. Spinelli-Stephens [87] ont généralisé ce travail au cas d'un support infini et aux tests d'adéquation à une famille de lois. Les statistiques de Cramér-von Mises et d'Anderson-Darling pour un échantillon d'une loi discrète sont :

$$W_n^2 = n \sum_{k=1}^{\infty} [\mathbb{F}_n(k) - F(k; \theta)]^2 p(k; \theta) \quad (7.2)$$

$$A_n^2 = n \sum_{k=1}^{\infty} \frac{[\mathbb{F}_n(k) - F(k; \theta)]^2 p(k; \theta)}{F(k; \theta)(1 - F(k; \theta))} \quad (7.3)$$

Lorsque les paramètres de F doivent être estimés, les expressions des statistiques (7.1), (7.2) et (7.3) deviennent alors :

$$\widehat{KS}_n = \sqrt{n} \max_{k \in \mathbb{N}^*} |\mathbb{F}_n(k) - F(k, \hat{\theta}_n)| \quad (7.4)$$

$$\widehat{W}_n^2 = n \sum_{k=1}^{\infty} [\mathbb{F}_n(k) - F(k; \hat{\theta}_n)]^2 p(k; \hat{\theta}_n) \quad (7.5)$$

$$\widehat{A}_n^2 = n \sum_{k=1}^{\infty} \frac{[\mathbb{F}_n(k) - F(k; \hat{\theta}_n)]^2 p(k; \hat{\theta}_n)}{F(k; \hat{\theta}_n)(1 - F(k; \hat{\theta}_n))} \quad (7.6)$$

Pour les calculs pratiques, le maximum et les séries précédentes doivent être finies et une troncature est nécessaire. Pour la statistique de Kolmogorov-Smirnov, il est clair que l'on

peut se limiter au maximum sur les entiers inférieurs à la plus grande des observations $k_{(n)}$, puisque $\forall k \geq k_{(n)}, \mathbb{F}_n(k) = 1$. La formule (7.4) devient alors :

$$\widehat{KS}_n = \sqrt{n} \max_{k \leq k_{(n)}} |\mathbb{F}_n(k) - F(k; \hat{\theta}_n)|$$

Pour le calcul des séries, Spinelli et Stephens proposent une autre procédure pour calculer les bornes de sommation :

- la borne inférieure M_i est choisie telle que $\forall j < M_i, \mathbb{P}_n(j) = 0$ et $p(j; \hat{\theta}_n) < 10^{-3}/n$,
- la borne supérieure M_s est choisie telle que $\forall j \geq M_s, \mathbb{P}_n(j) = 0$ et $p(j; \hat{\theta}_n) < 10^{-3}/n$.

Dans [34], Henze définit seulement une borne supérieure pour le calcul des statistiques de Cramér-von Mises :

$$M_s = \min \{k \geq k_{(n)}; (1 - F(k; \hat{\theta}_n))^3 \leq 10^{-4}/n\}$$

Les formules (7.5) et (7.6) deviennent alors :

$$\begin{aligned} \widehat{W}_n^2 &= n \sum_{k=M_i}^{M_s} [\mathbb{F}_n(k) - F(k; \hat{\theta}_n)]^2 p(k; \hat{\theta}_n) \\ \widehat{A}_n^2 &= n \sum_{k=M_i}^{M_s} \frac{[\mathbb{F}_n(k) - F(k; \hat{\theta}_n)]^2 p(k; \hat{\theta}_n)}{F(k; \hat{\theta}_n)(1 - F(k; \hat{\theta}_n))} \end{aligned}$$

Les auteurs avancent que ces troncatures ne modifient pas significativement la valeur des statistiques.

Henze propose également d'utiliser la statistique de test suivante :

$$\widehat{W}_n^{2*} = n \sum_{k=1}^{\infty} [\mathbb{F}_n(k) - F(k; \hat{\theta}_n)]^2 \mathbb{P}_n(k)$$

Le calcul pratique de la statistique \widehat{W}_n^{2*} se réduit à une somme finie puisque les fréquences empiriques $\mathbb{P}_n(k)$ sont nulles en dehors des observations.

Pour effectuer des tests d'adéquation à l'aide de ces statistiques, les démarches de Choulakian-Lockart-Stephens, Spinelli-Stephens et Henze sont bien différentes. Les premiers auteurs écrivent la loi asymptotique des statistiques de test comme la loi d'une somme pondérée de v.a. du χ^2 indépendantes. Les poids sont les valeurs propres d'une certaine matrice. Après les avoir déterminées, on peut déterminer ces lois sous une hypothèse \mathcal{H}_0 donnée. Henze, lui, a établi la convergence en loi du processus empirique estimé $\sqrt{n}(\mathbb{F}_n(\cdot) - F(\cdot; \hat{\theta}_n))$ vers une suite gaussienne centrée avec une certaine fonction de covariance. Les statistiques de test sont des fonctionnelles de ce processus, ce qui permet de déterminer leur loi asymptotique. Dans les deux cas, le principal problème est que les lois des statistiques de test sous \mathcal{H}_0 dépendent de la vraie valeur de θ , ce qui n'était pas le cas pour beaucoup de lois continues.

Les tests basés sur la statistique de test $\widehat{T}_n \in \{\widehat{KS}_n, \widehat{W}_n^2, \widehat{A}_n^2, \widehat{W}_n^{2*}\}$ consistent à rejeter l'hypothèse \mathcal{H}_0 au seuil α si \widehat{T}_n est supérieure à une valeur critique c_α . La valeur critique c_α dépend de la taille de l'échantillon n , du niveau de signification α , et aussi de la vraie valeur inconnue θ . Par conséquent, il faut estimer c_α . Pour cela, on utilise la méthode du bootstrap paramétrique décrite ci-après.

À partir de l'échantillon K_1, \dots, K_n , on estime le paramètre inconnu θ par l'estimateur $\widehat{\theta}_n$. Ensuite, on simule $N = 500$ échantillons de taille n de loi F de paramètre $\widehat{\theta}_n$. Pour chaque échantillon simulé, on calcule la statistique de test. La valeur critique du test, \widehat{c}_α , est le quantile empirique d'ordre $1 - \alpha$ des N valeurs obtenues de \widehat{T}_n . Henze [34] a montré la validité de cette procédure, à savoir que sous \mathcal{H}_0 , $\lim_{\substack{n \rightarrow +\infty \\ N \rightarrow +\infty}} \mathbb{P}(\widehat{T}_n > \widehat{c}_\alpha) = \alpha$.

On remarquera que l'utilisation de la méthode du bootstrap paramétrique complique considérablement l'application de ces tests par rapport aux tests analogues en temps continu.

7.2.3 Tests basés sur la fonction génératrice empirique

Définition 6

La fonction génératrice, notée $\varphi(\cdot; \theta)$, de la loi de K , est définie par

$$\varphi(t; \theta) = \mathbb{E}(t^K), \text{ avec } |t| < 1.$$

Définition 7

À partir de l'échantillon K_1, \dots, K_n , on définit la fonction génératrice empirique, notée φ_n , par $\varphi_n(t) = \frac{1}{n} \sum_{i=1}^n t^{K_i}$, avec $|t| < 1$.

a) Test de Kocherlakota et Kocherlakota

Kocherlakota et Kocherlakota [48] ont eu l'idée de construire un test d'adéquation basé sur la fonction génératrice puisque cette fonction caractérise de façon unique une loi. Pour t_0 fixé, les auteurs ont démontré la convergence suivante :

$$\frac{\varphi_n(t_0) - \varphi(t_0; \theta)}{\sigma_{(t_0; \theta)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad (7.7)$$

avec $\sigma_{(t_0; \theta)}^2 = \frac{1}{n} (\varphi(t_0^2; \theta) - \varphi^2(t_0; \theta))$.

Lorsque le vecteur θ des paramètres est estimé par l'estimateur du maximum de vraisemblance $\widehat{\theta}_n$, les auteurs ont montré, sous les conditions usuelles de régularité d'estimation, que :

$$\widehat{K}K_n = \frac{\varphi_n(t_0) - \varphi(t_0; \widehat{\theta}_n)}{\sigma_{(t_0; \widehat{\theta}_n)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

$$\text{où } \sigma_{(t_0; \hat{\theta}_n)}^2 = \frac{1}{n} \left(\varphi(t_0^2; \hat{\theta}_n) - \varphi^2(t_0; \hat{\theta}_n) \right) - \sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} \frac{\partial \varphi(t_0; \theta)}{\partial \theta_i} \frac{\partial \varphi(t_0; \theta)}{\partial \theta_j}$$

et σ_{ij} désigne le terme d'ordre $i \times j$ de l'inverse de la matrice d'information de Fisher $J(\theta)$.

L'hypothèse \mathcal{H}_0 est rejetée au seuil α si $|\widehat{K\bar{K}}_n| > z_\alpha$. Au vu des simulations menées dans le cas de l'adéquation à la loi de Poisson, les auteurs préconisent d'utiliser des petites valeurs pour t_0 , de l'ordre de 0.1. Mais on ne sait justifier ce choix autrement que de façon empirique et on ne sait pas s'il peut être étendu à d'autres lois discrètes, notamment à la loi géométrique.

b) Test de Rueda, Pérez-Abreu et O'Reilly

L'inconvénient majeur du test de Kocherlakota et Kocherlakota réside dans sa dépendance à la valeur arbitraire de t_0 . Pour remédier à ce problème, Rueda, Pérez-Abreu, et O'Reilly [78] proposent une statistique de test similaire à celle de Cramér-von Mises qui permet de prendre en compte toutes les valeurs de t .

Soit $\{\xi_n(t; \theta)\}_{t \geq 0}$ le processus empirique associé à la fonction génératrice :

$$\forall t \in \mathbb{R}^+, \quad \xi_n(t; \theta) = \sqrt{n} (\varphi(t; \theta) - \varphi_n(t))$$

On note $\hat{\theta}_n$ l'estimateur du MV de θ . Les auteurs ont montré dans le cas où $\dim \theta = 1$, que le processus $\{\xi_n(t; \hat{\theta}_n)\}_{t \geq 0}$ converge vers un processus gaussien centré de fonction de covariance $C_\theta(s, t)$ donnée par

$$C_\theta(s, t) = \varphi(st; \theta) - \varphi(s; \theta)\varphi(t; \theta) + (J(\theta) - 2) \frac{\partial \varphi(s; \theta)}{\partial \theta} \frac{\partial \varphi(t; \theta)}{\partial \theta}$$

La statistique du test est :

$$\widehat{RPO}_n = \int_0^1 \xi_n^2(t; \hat{\theta}_n) dt$$

Sous \mathcal{H}_0 , la loi de la statistique \widehat{RPO}_n dépend de la vraie valeur de θ , donc le calcul de la valeur critique du test de niveau α est effectué par la méthode de bootstrap paramétrique décrite pour les tests basés sur la fonction de répartition empirique.

c) Test de Baringhaus et Henze

En 1992, Baringhaus et Henze [7] ont eu l'idée de construire un nouveau test d'adéquation pour la loi de Poisson portant sur la fonction génératrice empirique. Les auteurs considèrent la fonction génératrice comme solution d'une équation différentielle.

Nous écrirons l'équation différentielle dont la fonction génératrice de la loi est solution sous la forme générale :

$$\psi(t; \varphi; \varphi'; \theta) = 0$$

Pour un test d'adéquation à une famille paramétrée de lois, il suffit alors de construire la statistique \widehat{BH}_n suivante qui est une mesure de l'écart entre la fonction génératrice empirique et la fonction génératrice estimée sous \mathcal{H}_0 :

$$\widehat{BH}_n = n \int_0^1 [\psi(t; \varphi_n; \varphi'_n; \hat{\theta}_n)]^2 dt$$

où φ'_n est la dérivée de la fonction génératrice empirique.

L'hypothèse nulle est rejetée pour les grandes valeurs de \widehat{BH}_n , la valeur critique étant déterminée par une méthode de bootstrap paramétrique.

Remarque :

Les trois tests basés sur la fonction génératrice n'ont été appliqués que pour la loi de Poisson, ainsi que les lois de Pascal-Poisson, Poisson bivariée et Neyman de type A bivariée pour le test de Kocherlakota et Kocherlakota.

7.2.4 Tests lisses de Neyman

Ces tests sont qualifiés de "lisses" (traduction française de "smooth") car Neyman [59] les a construits pour détecter les densités de probabilité qui diffèrent faiblement de celle spécifiée par l'hypothèse nulle : par un décalage de la moyenne, de la variance, des coefficients d'asymétrie ou d'aplatissement. Les tests lisses de Neyman s'appliquent à des lois continues ou discrètes. Les principaux résultats énoncés sur les tests lisses de Neyman proviennent du livre de Rayner et Best [75].

On définit la loi de probabilité $g_m(\cdot; \beta; \theta)$ par :

$$g_m(k; \beta; \theta) = C(\beta, \theta) \exp \left\{ \sum_{j=1}^m \beta_j h_j(k; \theta) \right\} p(k; \theta).$$

où $\{h_j(k; \theta)\}_{j=1, \dots, m}$ est une base orthonormée complète par rapport à la mesure de probabilité $p(k; \theta)$, c'est à dire $\forall i, j : \mathbb{E}_0[h_i(K; \theta_0)h_j(K; \theta_0)] = \delta_{ij}$, \mathbb{E}_0 désignant l'espérance par rapport à la loi de probabilité $p(k; \theta)$. Par convention, $h_0(k; \theta) \equiv 1$, afin que $\mathbb{E}_0[h_j(K)] = \delta_{j0} = 0$. $C(\beta, \theta)$ est une constante de normalisation.

Le principe des tests lisses de Neyman est de plonger la loi définie par $p(k; \theta)$ dans la loi définie par $g_m(\cdot; \theta; \beta)$, puis à tester l'hypothèse suivante $\mathcal{H}_0 : \beta = (\beta_1, \dots, \beta_m)^t = 0$ qui revient à faire un test d'adéquation à la loi définie par $p(\cdot; \theta)$. Le test du score est utilisé pour tester l'hypothèse \mathcal{H}_0 .

On définit alors la matrice $M(\theta)$ par :

$$M(\theta) = I_m - Cov_0 \left(h_j, \frac{\partial \ln p}{\partial \theta_i} \right)_{\substack{i=1, \dots, d \\ j=1, \dots, m}} \left\{ Cov_0 \left(\frac{\partial \ln p}{\partial \theta_i}, \frac{\partial \ln p}{\partial \theta_j} \right) \right\}_{\substack{i=1, \dots, d \\ j=1, \dots, d}}^{-1} Cov_0 \left(\frac{\partial \ln p}{\partial \theta_i}, h_j \right)_{\substack{i=1, \dots, d \\ j=1, \dots, m}}.$$

où $Cov_0(\cdot)$ désigne la covariance sous \mathcal{H}_0 lorsque l'espérance est prise par rapport à $p(k, \theta)$.

Dans le cas d'un test d'adéquation à une loi complètement spécifiée $F(\cdot; \theta_0)$ (θ_0 connu), la matrice $M(\theta_0)$ se réduit à la matrice identité I_m . La statistique du test du score pour tester l'hypothèse \mathcal{H}_0 est alors donnée par :

$$S_m = U(\theta_0)^t U(\theta_0) = \sum_{j=1}^m U_j(\theta_0)^2$$

où les composantes U_j du vecteur du score U sont définies par $U_j(\theta_0) = \sum_{i=1}^n h_j(k_i; \theta_0) / \sqrt{n}$.

Sous \mathcal{H}_0 , la statistique S_m suit asymptotiquement la loi du χ_m^2 .

Dans le cas où θ est inconnu, on note $\hat{\theta}_n$ l'estimateur du Maximum de Vraisemblance de θ . On note alors $\widehat{M} = M(\hat{\theta}_n)$ et $\widehat{U}_j = \sum_{i=1}^n h_j(k_i; \hat{\theta}_n) / \sqrt{n}$.

Sous réserve que la matrice \widehat{M} ne soit pas singulière, la statistique du test du score est donnée par :

$$\widehat{S}_m = \widehat{U}^t \widehat{M}^{-1} \widehat{U}$$

Sous \mathcal{H}_0 , la statistique \widehat{S}_m est asymptotiquement de loi χ_m^2 . Lorsque la matrice \widehat{M} est singulière, une solution consiste à retirer une ou plusieurs composantes orthonormales h_j et à calculer la statistique du score sans ces composantes. La loi asymptotique de la statistique ainsi modifiée est toujours de loi du χ^2 , le nombre de degrés de liberté étant égal au nombre de composantes.

7.3 Tests basés sur la transformation de Smirnov généralisée

En 1992, Nikulin [61] propose d'utiliser la transformation de Smirnov généralisée pour effectuer un test d'adéquation d'un échantillon à une loi, discrète ou continue. En 1997, Hocine [36] reprend cette idée pour construire des tests d'adéquation à quelques lois discrètes usuelles complètement spécifiées.

7.3.1 Test d'adéquation à une loi complètement spécifiée

La transformation de Smirnov généralisée est donnée par (page 44 de Greenwood et Nikulin [30]) :

Propriété 22 Transformation de Smirnov généralisée

Soit X une v.a. réelle de fonction de répartition F_0 . On note $F_0^-(x) = \mathbb{P}_0(X < x)$. Soient X et U deux v.a. indépendantes, respectivement de loi F_0 et $\mathcal{U}([0, 1])$. Alors la variable aléatoire $T = F_0^-(X) + [F_0(X) - F_0^-(X)]U$ est de loi $\mathcal{U}([0, 1])$.

En remarquant que pour une variable discrète K à valeurs dans \mathbb{N}^* on a : $\forall k \in \mathbb{N}^*, F_0(k) - F_0^-(k) = \mathbb{P}_0(K = k)$, le résultat de la propriété 22 peut être formulé de la façon suivante :

Propriété 23

Soit K une v.a. à valeurs dans \mathbb{N}^* , de fonction de répartition F_0 . Soit U une v.a. de loi $\mathcal{U}([0, 1])$, indépendante de K . Alors la variable aléatoire $T = F_0(K - 1) + \mathbb{P}_0(K)U$ est de loi $\mathcal{U}([0, 1])$.

La propriété suivante étend le résultat de la propriété 23, valable pour une variable aléatoire discrète K , à tout un échantillon K_1, \dots, K_n :

Propriété 24

Soit K_1, \dots, K_n n v.a. indépendantes à valeurs dans \mathbb{N}^* , de même fonction de répartition F_0 . Soient U_1, \dots, U_n n v.a. indépendantes et de même loi $\mathcal{U}([0, 1])$, indépendantes de K_1, \dots, K_n . Alors les variables aléatoires

$$T_i = F_0(K_i - 1) + \mathbb{P}_0(K_i)U_i \quad (7.8)$$

sont indépendantes et de même loi $\mathcal{U}([0, 1])$.

Pour tester l'adéquation d'un échantillon K_1, \dots, K_n à la loi discrète de fonction de répartition F_0 , la procédure est la suivante :

- transformer l'échantillon K_1, \dots, K_n en l'échantillon T_1, \dots, T_n à l'aide de la formule (7.8),
- tester l'adéquation de l'échantillon T_1, \dots, T_n à la loi $\mathcal{U}([0, 1])$.

7.3.2 Test d'adéquation à une famille paramétrée de lois

On notera par $\hat{\theta}_n = \hat{\theta}(K_1, \dots, K_n)$ un estimateur satisfaisant les hypothèses de régularité usuelles. On s'intéresse maintenant au résultat de la propriété 24 où θ est remplacé par $\hat{\theta}_n$:

Propriété 25

Soit K_1, \dots, K_n et U_1, \dots, U_n deux échantillons indépendants respectivement de loi $F(\cdot; \theta)$ et $\mathcal{U}([0, 1])$, avec $F(k; \theta)$ continue en k et θ . Soit $\hat{\theta}_n$ un estimateur convergent en probabilité vers θ . Pour $i \in \{1, \dots, n\}$, on pose :

$$\hat{T}_i = F(K_i - 1; \hat{\theta}_n) + \mathbb{P}(K_i; \hat{\theta}_n)U_i \quad (7.9)$$

Alors $\forall i \in \{1, \dots, n\}$, $\hat{T}_i \xrightarrow{\mathcal{L}} \mathcal{U}([0, 1])$.

Preuve :

La démonstration repose sur le théorème de Slutsky énoncé au chapitre 5 (théorème 7), avec les notations suivantes, pour i fixé :

- $X_n = (K_i, U_i)$,

- $Y_n = \hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$,
- $g(X_n, Y_n) = F(K_i - 1; \hat{\theta}_n) + \mathbb{P}(K_i; \hat{\theta}_n)U_i = \hat{T}_i$,
- $g(X_n, \theta) = F(K_i - 1; \theta) + \mathbb{P}(K_i; \theta)U_i = T_i$.

Le théorème de Slutsky permet d'affirmer que \hat{T}_i et T_i ont même loi limite, l'utilisation de la propriété 23 donnant le résultat cherché puisque T_i est de loi $\mathcal{U}([0, 1])$. \square

On est donc tenté de tester l'adéquation de K_1, \dots, K_n à la famille de lois \mathcal{F}_Θ en testant l'adéquation de $\hat{T}_1, \dots, \hat{T}_n$ à la loi uniforme sur $[0, 1]$, en utilisant par exemple la statistique d'Anderson-Darling. Mais il se passe le même phénomène que pour les tests basés sur la fonction de répartition empirique pour des lois continues : l'estimation de θ par $\hat{\theta}_n$ modifie la loi asymptotique de la statistique de test par rapport au cas où θ est connu. Par conséquent, on ne connaît pas a priori la loi asymptotique de la statistique de test sous \mathcal{H}_0 . Une étude spécifique de chaque famille de lois est alors nécessaire, comme l'ont fait D'Agostino et Stephens pour quelques lois continues usuelles dans [23].

7.4 Application à la loi géométrique

Il est important de préciser qu'à l'exception du test lisse de Neyman et du test du χ^2 , aucun des tests présentés à la section précédente n'a été appliqué à la loi géométrique.

7.4.1 Test de Kocherlakota et Kocherlakota

Pour la loi géométrique $\mathcal{G}(p)$, la fonction génératrice est donnée par :

$$\varphi(t; p) = \frac{pt}{1 - (1-p)t}, \quad t \in [0, 1]$$

Propriété 26

Le terme $\sigma_{(t_0; \hat{p}_n)}^2$ de la statistique $\widehat{K}\widehat{K}_n$ du test de Kocherlakota et Kocherlakota pour l'adéquation à la famille des lois géométriques est donné par :

$$\sigma_{(t_0; \hat{p}_n)}^2 = \frac{\hat{p}_n t_0^2 (1 - \hat{p}_n)^2 (1 - t_0)^4}{n [1 - (1 - \hat{p}_n) t_0^2] [1 - (1 - \hat{p}_n) t_0]^4}$$

Preuve :

Puisque $\dim \theta = 1$, l'expression de $\sigma_{(t_0; p)}^2$ est :

$$\begin{aligned} \sigma_{(t_0; p)}^2 &= \frac{1}{n} (\varphi(t_0^2; p) - \varphi^2(t_0; p)) - J^{-1}(p) \left(\frac{\partial \varphi(t_0; p)}{\partial p} \right)^2 \\ &= \frac{1}{n} \left(\frac{pt_0^2}{1 - (1-p)t_0^2} - \frac{p^2 t_0^2}{[1 - (1-p)t_0]^2} \right) - J^{-1}(p) \left(\frac{\partial \varphi(t_0; p)}{\partial p} \right)^2 \end{aligned}$$

Le calcul de la dérivée partielle donne :

$$\frac{\partial}{\partial p} \varphi(t; p) = \frac{t(1-t)}{[1 - (1-p)t]^2}$$

L'information de Fisher est : $J(p) = \frac{n}{p^2(1-p)}$.

On obtient alors :

$$\sigma_{(t_0;p)}^2 = \frac{1}{n} \left(\frac{pt_0^2}{1 - (1-p)t_0^2} - \frac{p^2 t_0^2}{[1 - (1-p)t_0]^2} \right) - \frac{p^2(1-p)t_0^2(1-t_0)^2}{n[1 - (1-p)t_0]^4}$$

Le résultat final est obtenu en réduisant les fractions ci-dessus au même dénominateur et en remplaçant p par \hat{p}_n . \square

Finalement, la statistique de test de Kocherlakota et Kocherlakota est :

$$\widehat{KK}_n = \sqrt{n} \frac{[1 - (1 - \hat{p}_n)t_0^2][1 - (1 - \hat{p}_n)t_0]^3}{t_0(1-t_0)^2(1-\hat{p}_n)\sqrt{\hat{p}_n}} \left\{ \frac{1}{n} [1 - (1 - \hat{p}_n)t_0] \sum_{i=1}^n t_0^{K_i} - \hat{p}_n t_0 \right\}$$

Remarque :

Au vu des simulations effectuées, il s'avère que le choix de $t_0 = 0.5$ semble donner de meilleurs résultats que la valeur $t_0 = 0.125$ suggérée par Kocherlakota et Kocherlakota dans le cas du test d'adéquation à la loi de Poisson.

7.4.2 Test de Rueda, Pérez-Abreu et O'Reilly

Propriété 27

La statistique du test de Rueda et al. pour tester l'adéquation à la famille des lois géométriques est donnée par :

$$\begin{aligned} RPO_n(p) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{K_i + K_j + 1} \\ &+ 2p \sum_{i=1}^n \left[\left(\sum_{j=1}^{K_i+1} \frac{1}{(1-p)^j (K_i + 2 - j)} \right) + \frac{\ln p}{(1-p)^{K_i+2}} \right] \\ &+ \frac{np(p+1)}{(1-p)^2} + \frac{2np^2 \ln p}{(1-p)^3} \end{aligned}$$

Preuve :

$$\begin{aligned} RPO(p) &= \int_0^1 n (\varphi_n(t) - \varphi(t;p))^2 dt \\ &= \int_0^1 n \left(\frac{1}{n} \sum_{i=1}^n t^{K_i} - \frac{pt}{1 - (1-p)t} \right)^2 dt \\ &= \int_0^1 n \left[\frac{1}{n^2} \left(\sum_{i=1}^n t^{K_i} \right)^2 - \frac{2}{n} \frac{pt}{1 - (1-p)t} \sum_{i=1}^n t^{K_i} + \frac{p^2 t^2}{[1 - (1-p)t]^2} \right] dt \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{K_i + K_j + 1} - 2 \sum_{i=1}^n \int_0^1 \frac{pt^{K_i+1}}{1 - (1-p)t} dt + n \int_0^1 \frac{p^2 t^2}{[1 - (1-p)t]^2} dt \end{aligned}$$

$$\text{Soit } A = 2 \sum_{i=1}^n \int_0^1 \frac{pt^{K_i+1}}{1 - (1-p)t} dt \text{ et } B = n \int_0^1 \frac{p^2 t^2}{[1 - (1-p)t]^2} dt.$$

Après calculs, on trouve :

$$A = 2p \sum_{i=1}^n \left[\left(\sum_{j=1}^{K_i+1} \frac{1}{(1-p)^j (K_i - j + 2)} \right) + \frac{\ln p}{(1-p)^{K_i+2}} \right]$$

$$B = \frac{np(p+1)}{(1-p)^2} + \frac{2np^2 \ln p}{(1-p)^3}. \quad \square$$

On obtient la valeur critique du test de niveau α par une méthode de bootstrap paramétrique.

7.4.3 Test de Baringhaus et Henze

Propriété 28

La fonction génératrice de la loi géométrique $\mathcal{G}(p)$ est l'unique fonction génératrice solution de l'équation différentielle suivante :

$$y'(t; p) - \frac{y^2(t; p)}{pt^2} = 0 \quad (E)$$

Preuve :

L'équation différentielle est une équation dite de Bernoulli. On pose $z(t) = 1/y(t)$. L'équation (E) devient :

$$z'(t) + \frac{1}{pt^2} = 0 \quad (E')$$

Les solutions de (E') sont de la forme $z_c(t; p) = \frac{1}{pt} + c$, avec $c \in \mathbb{R}$. Les solutions de (E) sont alors de la forme $y_c(t; p) = \frac{pt}{1 + cpt}$. Pour que y soit une fonction génératrice, on doit avoir $y_c(1; p) = 1$, ce qui implique que $c_0 = \frac{p-1}{p}$. On obtient $y_{c_0}(t; p) = \frac{pt}{1 - (1-p)t}$ qui est l'expression de la fonction génératrice de la loi géométrique de paramètre p .

Propriété 29

La statistique du test de Baringhaus et Henze pour la loi géométrique est :

$$BH_n(p) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{K_i K_j}{K_i + K_j - 1} - \frac{2}{n^2 p} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{K_i}{K_i + K_j + K_k - 2}$$

$$+ \frac{1}{p^2 n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \frac{1}{K_i + K_j + K_k + K_l - 3}.$$

Preuve :

$$\begin{aligned}
BH_n(p) &= n \int_0^1 \left\{ \varphi'_n(t) - \frac{\varphi_n^2(t)}{pt^2} \right\}^2 dt \\
&= n \int_0^1 \left\{ \frac{1}{n} \sum_{i=1}^n K_i t^{K_i-1} - \frac{1}{pt^2} \left(\sum_{i=1}^n t^{K_i} \right)^2 \right\}^2 dt \\
&= \frac{1}{n} \int_0^1 \left(\sum_{i=1}^n K_i t^{K_i-1} \right)^2 dt - \frac{2}{p} \int_0^1 \frac{1}{t^2} \left(\sum_{i=1}^n K_i t^{K_i-1} \right) \left(\sum_{i=1}^n t^{K_i} \right)^2 dt \\
&\quad + \frac{n}{p^2} \int_0^1 \frac{1}{t^4} \left(\sum_{i=1}^n t^{K_i} \right)^4 dt
\end{aligned}$$

On calcule séparément les termes de la statistique de test :

$$\begin{aligned}
\int_0^1 \left(\sum_{i=1}^n K_i t^{K_i-1} \right)^2 dt &= \sum_{i=1}^n \sum_{j=1}^n \frac{K_i K_j}{K_i + K_j - 1} \\
\int_0^1 \frac{1}{t^2} \left(\sum_{i=1}^n K_i t^{K_i-1} \right) \left(\sum_{i=1}^n t^{K_i} \right)^2 dt &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{K_i}{K_i + K_j + K_k - 2} \\
\int_0^1 \frac{1}{t^4} \left(\sum_{i=1}^n t^{K_i} \right)^4 dt &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \frac{1}{K_i + K_j + K_k + K_l - 3}
\end{aligned}$$

On obtient finalement la statistique de test :

$$\begin{aligned}
BH_n(p) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \frac{K_i K_j}{K_i + K_j - 1} - \frac{2}{p} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \frac{K_i}{K_i + K_j + K_k - 2} \\
&\quad + \frac{n}{p^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \frac{1}{K_i + K_j + K_k + K_l - 3}. \quad \square
\end{aligned}$$

La valeur critique du test se calcule par une méthode du bootstrap paramétrique. On peut constater que le calcul de la statistique de test n'est pas simple et nécessite donc un temps de calcul prohibitif. De plus, ce test utilise la méthode du bootstrap qui demande de répéter un grand nombre de fois le calcul de la statistique et rend donc cette méthode difficilement applicable en pratique.

7.4.4 Test lisse de Neyman

Rayner et Best [75] ont appliqué les tests lisses de Neyman à différentes lois discrètes et particulièrement à la loi géométrique. Des études de puissance ont été menées pour comparer la puissance du test lisse avec le test de Vit [91], et choisir également le nombre de composantes m du test lisse (voir page 97 de [75]). m sera choisi égal à 4 et la statistique du test d'ordre $m = 4$ est alors :

$$S_4 = \frac{1}{n} \sum_{j=2}^5 \left(\sum_{i=1}^n h_j(k_i; p) \right)^2$$

où les h_j sont un cas particulier des polynômes de Meixner (1934) :

$$\begin{aligned}
h_1(k;p) &= \frac{pk - 1}{\sqrt{1-p}} \\
h_2(k;p) &= \frac{k^2 p^2 + p(p-4)k + 2}{2(1-p)} \\
h_3(k;p) &= \frac{k^3 p^3 - k^2(9p^2 - 3p^3)}{6\sqrt{1-p}(1-p)} + \frac{k(2p^3 - 9p^2 + 18p) - 6}{6\sqrt{1-p}(1-p)} \\
h_4(k;p) &= \frac{k^4 p^4 + k^3(6p^4 - 16p^3) + k^2(72p^2 - 48p^3 + 11p^4)}{24(1-p)^2} \\
&\quad + \frac{k(6p^4 - 32p^3 + 72p^2 - 96p) + 24}{24(1-p)^2} \\
h_5(k;p) &= \frac{k^5 p^5 + k^4(10p^5 - 25p^4) + k^3(35p^5 - 150p^4 + 200p^3)}{120\sqrt{1-p}(1-p)^2} \\
&\quad + \frac{k^2(50p^5 - 275p^4 + 600p^3 - 600p^2)}{120\sqrt{1-p}(1-p)^2} + \frac{k(24p^5 - 150p^4 + 400p^3 - 600p^2 + 600p) - 120}{120\sqrt{1-p}(1-p)^2}
\end{aligned}$$

Sous l'hypothèse \mathcal{H}_0 , la statistique \hat{S}_4 suit asymptotiquement la loi χ_4^2 lorsque le paramètre de la loi géométrique est estimé par le maximum de vraisemblance $\hat{p}_n = 1/\bar{K}_n$. Pour les échantillons de petite taille, Rayner et Best proposent une correction portant sur le quantile de niveau $\alpha = 0.05$ de la loi χ_4^2 (égal à 9.488) qui se calcule ainsi :

$9.488 \times \{1 + 3.643/n - 2.314/\sqrt{n} - 0.447/\sqrt{n(1-\hat{p}_n)}\}$, en remplaçant $n(1-p)$ par n lorsque $n(1-\hat{p}_n) < 1$.

7.4.5 Tests basés sur la transformation de Smirnov généralisée

On a vu à la section 7.3.2 que les tests d'adéquation à une famille de lois basés sur la transformation de Smirnov généralisée nécessitent une étude spécifique.

Nous utilisons dans cette section le fait que si $E \rightsquigarrow \mathcal{E}(\lambda)$, alors la v.a. K à valeurs dans \mathbb{N}^* définie par $K = \lfloor E \rfloor + 1$ est de loi $\mathcal{G}(1 - e^{-\lambda})$. Inversement on peut penser à créer, à partir d'un échantillon de loi géométrique, un échantillon de loi exponentielle.

Pour créer un tel échantillon, on utilise la transformation de Smirnov généralisée: soit $K \rightsquigarrow \mathcal{G}(p)$ et $U \rightsquigarrow \mathcal{U}([0, 1])$ indépendante de K . Alors la propriété 23 établit que :

$$T = 1 - (1 - pU)(1 - p)^{K-1} \rightsquigarrow \mathcal{U}([0, 1])$$

Or si $V \rightsquigarrow \mathcal{U}([0, 1])$, alors $-\lambda \ln(1 - V) \rightsquigarrow \mathcal{E}(\lambda)$,

donc

$$E = K - 1 + \frac{\ln(1 - pU)}{\ln(1 - p)} \rightsquigarrow \mathcal{E}(-\ln(1 - p)).$$

On retrouve bien $K = \lfloor E \rfloor + 1$.

Ainsi, on testera l'adéquation d'un échantillon K_1, \dots, K_n à la loi $\mathcal{G}(p)$, p connu, en testant l'adéquation des $E_i = K_i - 1 + \frac{\ln(1 - pU_i)}{\ln(1 - p)}$ à la loi $\mathcal{E}(-\ln(1 - p))$.

Dans le cas où le paramètre p est inconnu et doit être estimé, l'idée naturelle est de tester l'adéquation des

$$\widehat{E}_i = K_i - 1 + \frac{\ln(1 - \widehat{p}_n U_i)}{\ln(1 - \widehat{p}_n)}, \quad \text{où } \widehat{p}_n = \frac{n}{\sum_{i=1}^n K_i} \quad (7.10)$$

à la famille des lois exponentielles, par exemple par le test d'Anderson-Darling.

Pour valider ce test, il faut vérifier que la loi asymptotique de la statistique de test ainsi définie est la même que celle de la statistique d'Anderson-Darling pour tester l'adéquation d'un échantillon à la loi exponentielle.

Faute d'avoir pu prouver ce résultat, nous avons comparé à l'aide de simulations de Monte Carlo les quantiles empiriques de la loi de la statistique d'Anderson-Darling obtenue à partir d'un échantillon géométrique par la transformation de Smirnov généralisée, aux quantiles théoriques de la loi asymptotique de la statistique correspondante pour le test d'adéquation à la loi exponentielle. La figure FIG. 7.6 montre que ces quantiles diffèrent très peu, ce qui nous laisse penser que le test proposé est effectivement applicable.

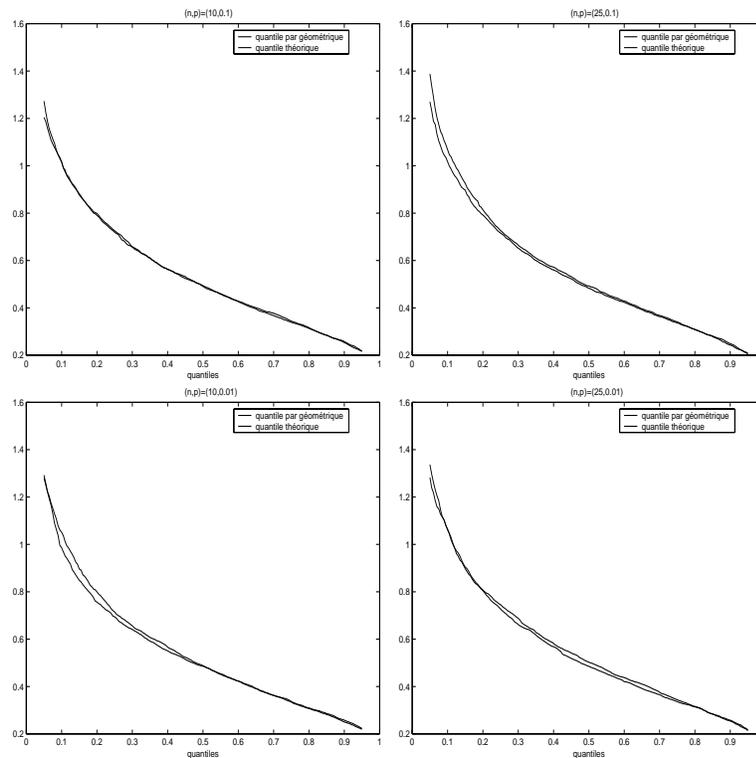


FIG. 7.6 – Comparaison des quantiles pour des échantillons de taille 10 et 25 des lois géométriques de paramètres $p = 0.1$ et $p = 0.01$

La procédure pour tester l'adéquation de l'échantillon K_1, \dots, K_n à la famille des lois géométriques est donc la suivante. On commence par transformer l'échantillon K_1, \dots, K_n en un échantillon $\widehat{E}_1, \dots, \widehat{E}_n$ par la formule (7.10). Puis on teste l'adéquation de l'échantillon

$\widehat{E}_1, \dots, \widehat{E}_n$ à la famille des lois exponentielles. On présume que n'importe quel test conviendra. Nous avons utilisé le test d'Anderson-Darling.

Il est clair que la même procédure est applicable à d'autres lois discrètes, par exemple à la loi de Weibull de type I qui est issue de la loi de Weibull continue par discrétisation.

7.4.6 Comparaison des tests

Dans cette section nous nous intéressons à la comparaison des tests suivants dans le cas de l'adéquation à la famille des lois géométriques :

- \widehat{S}_4 : test lisse de Neyman,
- \widehat{A}_n^2 : test d'Anderson-Darling de Henze pour une loi discrète,
- \widehat{W}_n^2 : test de Cramér-von Mises de Henze pour une loi discrète,
- \widehat{A}_{TSG}^2 : test d'Anderson-Darling utilisant la transformation de Smirnov généralisée,
- \widehat{KK}_n : test de Kocherlakota et Kocherlakota,
- \widehat{RPO}_n : test de Rueda, Pérez-Abreu et O'Reilly,
- \widehat{BH}_n : test de Baringhaus et Henze.

Nous avons également mis en place le test de Nikulin-Rao-Robson, mais les résultats sont tellement mauvais que nous avons choisi de ne pas le présenter. Nous précisons que pour les tests d'Anderson-Darling et Cramér-von Mises, nous avons utilisé le critère d'arrêt de Henze. Pour le test de Kocherlakota et Kocherlakota, nous avons choisi $t_0 = 0.5$.

Les résultats s'obtiennent très vite pour les tests \widehat{S}_4 , \widehat{A}_{TSG}^2 et \widehat{KK}_n . En revanche, l'utilisation du bootstrap fait que les simulations sont très longues pour les tests \widehat{A}_n^2 , \widehat{W}_n^2 , \widehat{RPO}_n et \widehat{BH}_n . Nous avons fait appel à la puissance de calcul de la grappe de PC du *i-cluster*, projet commun à l'INRIA et Hewlett-Packard (<http://icluster.imag.fr/>). Malgré cela, certaines simulations excessivement longues n'ont pu être effectuées et sont signalées par un signe '?' dans les tableaux de résultats.

Notre premier travail a consisté à vérifier le niveau de ces tests par des simulations de Monte-Carlo. Pour une valeur $p \in \{0.5, 0.3, 0.1, 0.01, 0.001, 0.0001\}$ fixée, nous avons simulé un échantillon de loi géométrique $\mathcal{G}(p)$ de taille $n \in \{5, 10, 20, 40, 60, 80, 100, 200\}$ fixée, auquel nous avons appliqué tous les tests d'adéquation cités ci-dessus au seuil $\alpha = 5\%$. Cette procédure a été répliquée $N = 2000$ fois pour chaque couple (p, n) , et le tableau TAB. 7.7 donne pour chaque test le pourcentage de rejets de l'hypothèse géométrique parmi les N simulations.

Les tests basés sur la fonction de répartition empirique, \widehat{W}_n^2 , \widehat{A}_n^2 et \widehat{A}_{TSG}^2 , se comportent tous les trois de façon identique et satisfaisante, et ce dès $n = 5$. Les seuils empiriques pour le test lisse sont très voisins du seuil théorique, avec pour $n \geq 10$, un seuil empirique plus

	\widehat{S}_4	\widehat{W}_n^2	\widehat{A}_n^2	\widehat{A}_{TSG}^2	$\widehat{K}\widehat{K}_n$	\widehat{RPO}_n	\widehat{BH}_n	\widehat{S}_4	\widehat{W}_n^2	\widehat{A}_n^2	\widehat{A}_{TSG}^2	$\widehat{K}\widehat{K}_n$	\widehat{RPO}_n	\widehat{BH}_n
p	$n = 5$							$n = 10$						
0.0001	4.1	4.7	4.7	5.7	4.2	6.5	6.8	6.0	5.2	5.6	5.1	6.9	3.7	4.3
0.001	4.0	4.7	4.8	5.8	6.2	5.8	5.9	6.0	5.2	5.0	5.2	6.3	3.8	5.6
0.01	4.1	4.7	4.8	5.9	4.2	5.4	5.8	6.0	4.6	4.9	4.8	6.7	4.7	5.3
0.1	4.2	4.1	4.4	5.9	4.6	5.7	5.3	6.5	4.4	4.5	5.5	6.3	5.7	5.5
0.3	5.3	4.0	4.1	6.0	4.3	3.9	3.1	6.0	4.6	4.0	5.6	7.4	5.0	4.0
0.5	4.8	5.7	2.5	6.2	4.6	1.3	5.7	6.0	5.4	3.6	5.2	6.3	3.1	3.6
p	$n = 20$							$n = 40$						
0.0001	5.7	5.7	5.2	5.6	4.3	5.1	5.6	5.8	4.5	5.0	5.8	4.6	5.3	4.2
0.001	5.8	5.6	5.2	5.6	4.5	4.4	6.5	5.8	4.4	4.9	5.8	5.4	5.1	3.4
0.01	5.7	5.5	5.1	5.5	4.5	4.4	5.0	5.9	4.5	5.5	5.8	4.6	4.4	4.6
0.1	6.1	4.8	4.5	5.7	4.3	5.0	4.5	6.1	4.4	5.2	6.0	5.0	4.8	5.0
0.3	5.7	5.0	5.0	5.3	5.0	4.9	4.9	6.4	4.7	5.1	5.5	5.3	6.0	6.0
0.5	5.7	5.7	4.6	5.2	3.5	5.1	5.0	5.5	5.6	4.5	5.4	4.7	5.0	4.7
p	$n = 60$							$n = 80$						
0.0001	6.1	4.7	5.2	4.7	4.7	5.8	5.9	5.5	5.5	5.2	5.1	4.0	5.4	5.8
0.001	6.4	4.1	4.7	4.6	3.9	6.0	6.0	5.5	5.0	4.8	5.1	4.1	5.2	5.6
0.01	6.4	5.1	4.4	4.7	3.9	6.7	6.7	5.5	4.8	4.7	5.2	4.0	4.9	5.3
0.1	6.2	5.3	4.7	4.6	5.5	5.6	5.8	5.4	5.5	4.8	4.5	5.1	5.3	5.0
0.3	6.6	5.6	4.8	4.4	5.3	5.4	4.6	6.2	5.6	4.8	4.4	5.4	5.0	6.0
0.5	6.5	5.1	4.7	4.3	5.0	4.2	4.0	5.5	5.0	4.9	4.7	4.1	5.4	4.6
p	$n = 100$							$n = 200$						
0.0001	5.0	5.5	4.9	4.8	3.8	-	-	5.2	5.0	4.3	5.4	4.5	-	-
0.001	5.0	5.2	5.2	4.8	3.9	-	-	5.3	5.0	6.0	5.5	5.4	-	-
0.01	5.0	5.9	6.2	4.6	4.0	-	-	5.2	6.3	6.5	5.4	5.0	-	-
0.1	5.3	6.0	5.8	4.5	4.4	-	-	5.3	5.9	6.3	5.2	4.6	-	-
0.3	5.5	5.9	5.0	5.3	4.8	-	-	5.0	5.4	5.9	5.4	5.2	-	-
0.5	5.5	5.3	5.1	5.4	5.0	-	-	4.9	5.0	5.6	4.7	4.7	-	-

TAB. 7.7 – Seuil empirique pour les tests à la loi géométrique de niveau $\alpha = 0.05$

proche de 6% que de 5%. Pour les tests \widehat{KK}_n , \widehat{RPO}_n et \widehat{BH}_n , les seuils empiriques sont un peu plus éloignés du seuil théorique, mais cela reste satisfaisant.

Les valeurs les plus éloignées du seuil théorique sont obtenues pour n petit et $p = 0.5$. Dans ce cas, l'échantillon est à moitié constitué de 1, ce qui a peu de chances d'arriver en fiabilité. Globalement, tous les tests proposés peuvent donc être utilisés pour tester l'adéquation d'un échantillon, même de petite taille, à la loi géométrique.

Une étude de puissance de ces tests a été également menée à l'aide de simulations. Les données ont été simulées selon la loi de Weibull discrète de type I $\mathcal{W}_I(q, \beta)$, car cette loi peut être aussi bien DFR que IFR et qu'elle se réduit à la loi géométrique lorsque $\beta = 1$. Le contexte industriel et l'expérience acquise nous incitent à choisir surtout des valeurs de β largement supérieures à 1, mais nous avons également choisi des valeurs de β proches de 1 pour tester la puissance des tests à discerner deux lois proches. Finalement on a $\beta \in \{0.8, 1.2, 2\}$. De même, différentes valeurs de $q \in \{0.5, 0.7, 0.9, 0.99, 0.999, 0.9999\}$ ont été choisies afin d'obtenir des observations de l'ordre de l'unité jusqu'à la dizaine de milliers. La taille de l'échantillon varie de façon à couvrir les applications observées en pratique : $n \in \{10, 20, 40, 60, 80, 100, 200\}$, et le seuil des tests est $\alpha = 5\%$. Les tableaux TAB. 7.8, 7.9 et 7.10 présentent pour chaque test et pour chaque triplet (q, β, n) , les pourcentages de rejets de l'hypothèse géométrique pour $N = 2000$ simulations.

Pour analyser les résultats de l'étude de puissance, nous classons les données en trois ordres de grandeur. Les données seront dites : d'un très grand ordre de grandeur lorsqu'elles sont de l'ordre du millier à la dizaine de milliers ($q \in \{0.999, 0.9999\}$), d'un grand ordre de grandeur pour des données de l'ordre de la dizaine à la centaine ($q \in \{0.9, 0.99\}$) et d'un petit ordre de grandeur pour des données inférieures à l'ordre de la dizaine ($q < 0.9$).

Pour $\beta = 0.8$, le test de Baringhaus et Henze est le plus puissant, quel que soit l'ordre de grandeur des données, devançant le test de Rueda *al.* La puissance du test lisse est très satisfaisante, et cela pour tous les ordres de grandeur des données. Il semble que pour tous les tests, à l'exception de \widehat{A}_n^2 , \widehat{W}_n^2 et \widehat{KK}_n , la puissance soit une fonction croissante de q . Pour des données d'un très grand ordre de grandeur, les tests \widehat{A}_n^2 , \widehat{W}_n^2 et \widehat{KK}_n ont une puissance très faible, quel que soit n . On évitera donc de les utiliser pour ce type de données. Pour les tests basés sur la fonction de répartition empirique, cela vient peut être du fait que ce sont des tests spécifiques aux lois discrètes alors que pour cet ordre de grandeur, les données sont proches de celles d'une loi continue. C'est probablement pour cette raison que le test \widehat{A}_{TSG}^2 a une très bonne puissance, puisque par l'intermédiaire de la transformation de Smirnov généralisée, on effectue un test d'adéquation à un échantillon de loi continue. Inversement on peut noter que pour des données d'un petit ordre de grandeur et pour des tailles d'échantillons importantes, la puissance du test \widehat{A}_{TSG}^2 est en retrait par rapport aux autres tests. Cela peut s'expliquer par le fait que dans ces conditions, le caractère fortement discret des observations

(i.e. la présence de nombreux ex-æquos) est perdu par la transformation en temps continu sur laquelle repose le test.

	\hat{S}_4	\widehat{W}_n^2	\hat{A}_n^2	\hat{A}_{TSG}^2	$\widehat{K}\widehat{K}_n$	\widehat{RPO}_n	$\widehat{B}\widehat{H}_n$	\hat{S}_4	\widehat{W}_n^2	\hat{A}_n^2	\hat{A}_{TSG}^2	$\widehat{K}\widehat{K}_n$	\widehat{RPO}_n	$\widehat{B}\widehat{H}_n$
q	n = 10							n = 20						
0.9999	18.6	1.6	1.4	18.5	0.6	23.5	28.2	27.1	0.8	0.8	27.5	1.1	25.6	32.8
0.999	18.6	1.5	1.7	18.3	3.0	21.9	25.0	27.1	0.8	0.8	27.5	5.9	22.3	33.0
0.99	18.6	1.5	2.4	17.8	12.8	19.8	21.6	27.2	4.7	6.6	27.6	21.8	28.7	30.7
0.9	18.6	9.5	12.9	16.4	15.5	17.4	20.5	26.6	23.9	27.5	26.1	23.4	27.8	32.5
0.7	17.3	10.1	10.2	14.1	12.0	12.7	16.6	24.7	17.7	17.3	21.2	19.3	20.5	25.7
0.5	15.6	8.0	5.9	10.4	9.0	9.4	13.4	21.5	12.2	11.7	14.8	15.4	17.2	20.7
q	n = 40							n = 60						
0.9999	39.4	0.3	0.3	45.0	1.5	49.2	52.5	49.8	0.0	0.0	57.4	2.0	65.0	68.5
0.999	39.4	0.4	0.4	45.0	9.6	48.4	52.4	49.8	15.2	1.0	57.4	12.5	64.2	69.5
0.99	39.5	38.0	34.7	44.9	30.7	45.1	50.8	49.7	60.6	57.0	57.5	31.5	60.7	65.3
0.9	39.1	39.9	43.5	42.7	37.4	45.0	50.3	49.6	54.4	55.6	50.5	50.1	56.3	61.1
0.7	36.3	32.2	31.5	34.0	34.9	35.9	41.2	46.0	47.5	44.0	45.3	48.3	50.6	56.2
0.5	30.4	23.0	23.3	22.2	27.6	25.9	29.3	38.8	35.5	33.8	29.3	38.8	39.2	40.8
q	n = 80							n = 100						
0.9999	58.2	0.1	0.1	69.1	2.6	79.0	83.1	67.2	0.2	0.0	77.8	3.3	80.2	90.4
0.999	58.2	49.3	7.3	69.1	14.2	77.5	80.9	67.2	6.9	26.5	77.7	17.8	78.5	88.2
0.99	58.2	70.3	72.3	69.3	38.0	74.3	79.8	67.3	76.9	83.3	77.5	41.4	74.1	85.6
0.9	58.0	63.5	66.8	68.1	66.6	78.1	77.4	66.5	71.6	76.2	76.5	70.1	85.3	88.4
0.7	53.8	58.2	55.7	56.3	55.4	58.5	65.0	61.3	66.6	66.5	64.8	69.4	71.1	69.3
0.5	45.6	44.3	43.4	38.7	17.1	50.2	52.3	51.6	52.3	52.8	44.4	59.2	64.8	51.6
q	n = 200													
0.9999	92.2	1.5	85.4	73.2	6.3	97.1	98.1							
0.999	91.5	96.0	93.5	90.8	31.8	97.0	98.3							
0.99	91.2	96.2	98.1	96.8	60.8	96.4	97.5							
0.9	90.1	95.4	97.0	96.6	94.0	96.2	97.3							
0.7	86.9	92.5	92.7	97.5	94.1	91.2	91.5							
0.5	77.9	82.8	82.5	96.6	86.2	85.2	87.1							

TAB. 7.8 – Puissance empirique des tests d'adéquation à la géométrie pour $\beta = 0.8$

Comme pour $\beta = 0.8$, les résultats des simulations pour $\beta = 1.2$ indiquent que la puissance des tests est globalement une fonction croissante de q . Pour des données d'un grand ordre de grandeur et des échantillons de grande taille, le meilleur test est celui de $\widehat{B}\widehat{H}_n$. Puis par ordre décroissant de puissance, on a le test de \hat{A}_{TSG}^2 , les tests \widehat{W}_n^2 et \hat{A}_n^2 qui ont une puissance comparable, et enfin le test \hat{S}_4 . Le problème de la très faible puissance des tests \widehat{W}_n^2 et \hat{A}_n^2 pour $q > 0.999$, rencontré pour $\beta = 0.8$, n'est plus valable. Seuls les tests $\widehat{K}\widehat{K}_n$ et \widehat{RPO}_n possèdent toujours ce problème pour des données d'un grand ordre de grandeur. Pour des données d'un petit ordre de grandeur et des petits échantillons ($n \leq 40$), le test de \hat{A}_n^2 est le plus puissant, tandis que pour des échantillons plus grands, les tests de $\widehat{K}\widehat{K}_n$, \widehat{RPO}_n et $\widehat{B}\widehat{H}_n$ donnent les meilleurs résultats. On notera que pour $\beta = 1.2$, la supériorité des tests basés sur la fonction génératrice semble moins évidente que pour le cas précédent avec $\beta = 0.8$.

	\widehat{S}_4	\widehat{W}_n^2	\widehat{A}_n^2	\widehat{A}_{TSG}^2	$\widehat{K}K_n$	\widehat{RPO}_n	\widehat{BH}_n	\widehat{S}_4	\widehat{W}_n^2	\widehat{A}_n^2	\widehat{A}_{TSG}^2	$\widehat{K}K_n$	\widehat{RPO}_n	\widehat{BH}_n
q	n = 10							n = 20						
0.9999	7.0	11.8	12.3	6.3	0.5	1.4	1.8	8.7	18.3	17.9	11.1	0.8	2.3	6.9
0.999	7.0	12.7	12.2	6.3	1.9	2.1	2.0	8.7	17.4	17.4	11.0	2.5	1.6	7.7
0.99	7.1	9.9	9.2	6.4	0.9	3.0	2.3	8.9	10.8	10.4	11.0	0.4	8.5	6.6
0.9	7.8	6.4	7.9	6.5	6.7	8.1	2.4	9.1	13.6	14.5	11.0	11.9	14.9	7.2
0.7	10.4	9.1	8.1	6.4	7.1	8.2	4.0	9.2	14.3	14.0	9.9	12.4	12.1	7.3
0.5	9.3	9.6	6.2	6.0	3.0	4.4	3.8	9.1	13.2	11.9	7.1	6.7	9.0	7.5
q	n = 40							n = 60						
0.9999	12.8	24.5	24.0	22.5	1.0	0.5	17.5	17.6	30.0	32.7	33.5	1.3	2.1	29.5
0.999	12.8	18.0	19.0	22.6	1.3	8.0	23.0	17.6	23.0	24.9	33.5	0.7	18.0	33.0
0.99	12.9	17.2	14.1	22.6	18.0	33.0	23.5	17.6	25.2	26.7	33.1	29.8	30.2	29.5
0.9	13.0	23.1	23.4	22.1	22.0	24.1	16.2	17.2	29.5	34.5	31.9	32.9	35.7	26.8
0.7	12.1	21.2	20.4	18.5	22.1	23.2	18.0	14.7	25.8	30.2	23.7	32.0	31.9	25.3
0.5	9.4	18.1	15.9	10.3	13.6	15.2	14.2	10.7	22.2	22.5	14.2	20.7	21.2	24.5
q	n = 80							n = 100						
0.9999	24.4	40.1	36.6	43.5	1.0	7.2	42.9	29.8	45.9	39.5	52.2	1.2	15.2	65.8
0.999	24.3	34.0	31.5	43.5	0.3	28.0	46.6	29.8	43.2	38.2	52.2	0.5	38.6	58.7
0.99	24.4	37.6	39.0	43.4	12.3	39.4	47.2	29.9	48.4	41.8	52.1	17.5	48.2	56.5
0.9	23.6	42.9	44.5	39.9	42.2	46.5	41.1	29.1	50.9	51.5	47.8	52.0	55.8	51.4
0.7	20.4	36.0	37.6	28.9	40.2	34.2	34.5	23.6	44.8	44.4	37.0	48.4	49.6	39.4
0.5	12.8	30.5	26.8	16.0	26.9	32.0	28.7	15.7	36.7	33.8	21.0	34.3	34.5	32.0
q	n = 200													
0.9999	63.7	79.7	75.2	41.5	2.5	-	-							
0.999	62.0	80.5	78.4	67.8	0.1	-	-							
0.99	61.4	83.0	86.5	82.0	45.3	-	-							
0.9	60.3	81.5	83.2	81.9	82.5	-	-							
0.7	47.2	67.5	70.5	84.1	77.3	-	-							
0.5	31.9	57.6	57.3	84.1	61.3	-	-							

TAB. 7.9 – Puissance des tests d'adéquation à la loi géométrique pour $\beta = 1.2$

Les résultats des simulations pour $\beta = 2$ indiquent que tous les tests, à l'exception de \widehat{KK}_n , ont une puissance élevée lorsque la taille d'échantillon est supérieure ou égale à 20, et rejettent donc fortement l'hypothèse géométrique. Les tests \widehat{S}_4 , \widehat{W}_n^2 , \widehat{A}_n^2 , \widehat{A}_{TSG}^2 et \widehat{RPO}_n ont des puissances similaires pour $q \leq 0.99$. Le test \widehat{BH}_n apparaît un peu moins puissant, surtout pour les petits échantillons. Le problème du test \widehat{KK}_n s'est encore accentué au point que, selon la valeur de q et n , le test rejette ou accepte presque toujours l'hypothèse géométrique.

	\widehat{S}_4	\widehat{W}_n^2	\widehat{A}_n^2	\widehat{A}_{TSG}^2	\widehat{KK}_n	\widehat{RPO}_n	\widehat{BH}_n	\widehat{S}_4	\widehat{W}_n^2	\widehat{A}_n^2	\widehat{A}_{TSG}^2	\widehat{KK}_n	\widehat{RPO}_n	\widehat{BH}_n
q	n = 5							n = 10						
0.9999	15.2	33.8	19.1	20.4	0.1	0.0	0.0	57.6	60.4	54.4	54.7	0.0	31.7	23.5
0.999	16.6	32.3	19.5	21.3	0.0	0.7	0.1	58.1	55.6	55.0	54.6	0.0	31.3	39.3
0.99	19.0	28.5	23.7	19.1	0.3	27.2	0.1	54.8	51.7	57.5	52.9	24.2	62.2	24.2
0.9	32.4	28.1	27.8	16.7	21.8	26.7	1.6	56.7	58.2	57.9	46.0	54.9	58.8	36.2
q	n = 20							n = 40						
0.9999	89.8	88.3	92.7	92.0	0.0	85.7	75.2	99.9	99.9	100	100	0.0	98.3	99.9
0.999	90.6	87.0	92.5	92.0	0.0	85.6	76.0	99.6	99.9	100	99.9	76.2	99.6	99.9
0.99	89.3	89.1	91.9	91.3	80.9	93.7	87.6	99.9	100	99.9	99.9	99.5	100	100
0.9	86.8	89.7	89.9	83.1	91.5	92.6	86.0	99.0	99.8	100	98.8	99.8	99.9	99.9
q	n = 60							n = 80						
0.9999	100	100	100	100	0.0	100	100	100	100	100	100	100	100	100
0.999	100	100	100	100	95.2	100	100	100	100	100	100	100	100	100
0.99	100	100	100	100	100	100	100	100	100	100	100	100	100	100
0.9	100	100	100	99.9	100	100	100	100	100	100	100	100	100	100

TAB. 7.10 – Puissance des tests d'adéquation à la loi géométrique $\beta = 2$

Conclusion :

À partir des résultats obtenus, il semble que le test de Baringhaus et Henze soit le plus puissant lorsque $\beta < 1$, et que ce soit le test de Rueda *et al.* lorsque $\beta > 1$. Il apparaît que pour les valeurs de β proches de 1, la puissance des tests dépend fortement de l'ordre de grandeur des données. Ce phénomène s'amenuisant rapidement avec la croissance de β , puisque pour $\beta = 2$, tous les tests ont une puissance comparable et excellente.

On pourra conseiller l'utilisation du test d'Anderson-Darling utilisant la transformation de Smirnov généralisée dont les résultats de puissance sont satisfaisants dans presque tous les cas. Le test de Rueda *et al.* peut être également conseillé puisqu'il est même plus puissant que le test \widehat{A}_{TSG}^2 dans les deux cas suivants : $\beta < 1$, $\beta > 1$ avec q petit.

7.5 Application à la loi de Weibull de type I

La fonction génératrice de la loi de Weibull ne possédant pas de forme explicite simple, les tests d'adéquation de Kocherlakota-Kocherlakota, Baringhaus-Henze et Rueda *et al.* ne sont pas applicables. Pour le test lisse de Neyman, le calcul de la b.o.n pour la loi de Weibull de type I n'est pas possible en raison de la forme des moments sous forme de séries.

Les seuls tests disponibles pour la loi de Weibull de type I sont donc les tests basés sur la fonction de répartition empirique et les tests basés sur la transformation de Smirnov généralisée.

7.5.1 Tests basés sur la transformation de Smirnov généralisée

On utilise la même procédure que pour les tests d'adéquation à la loi géométrique basés sur la transformation de Smirnov généralisée. En remarquant que si W est de loi de Weibull continue $\mathcal{W}(\eta, \beta)$, alors la variable $K = [W] + 1$ est de loi de Weibull discrète de type I $\mathcal{W}_I(e^{-\frac{1}{n^\beta}}, \beta)$, on va créer un échantillon de loi de Weibull continue à partir d'un échantillon de loi de Weibull discrète de type I.

Soient $K \rightsquigarrow \mathcal{W}_I(q, \beta)$, et $U \rightsquigarrow \mathcal{U}([0, 1])$, indépendante de K . La propriété 23 établit alors que

$$T = 1 - q^{(K-1)^\beta} + \left(q^{(K-1)^\beta} - q^{K^\beta} \right) U \rightsquigarrow \mathcal{U}([0, 1])$$

On testera ainsi l'adéquation de l'échantillon K_1, \dots, K_n à la loi de Weibull de type I $\mathcal{W}_I(q, \beta)$, avec q et β connus, en testant l'adéquation de l'échantillon des $T_i = 1 - q^{(K_i-1)^\beta} + \left(q^{(K_i-1)^\beta} - q^{K_i^\beta} \right) U_i$, $i = 1, \dots, n$ à la loi $\mathcal{U}([0, 1])$.

D'autre part si $V \rightsquigarrow \mathcal{U}([0, 1])$, alors $\eta [-\ln(1 - V)]^{\frac{1}{\beta}} \rightsquigarrow \mathcal{W}(\eta, \beta)$, donc

$$W = \left[(K - 1)^\beta + \frac{1}{\ln(q)} \ln \left[1 - (1 - q^{K^\beta - (K-1)^\beta}) U \right] \right]^{\frac{1}{\beta}} \rightsquigarrow \mathcal{W} \left((-\ln q)^{-\frac{1}{\beta}}, \beta \right)$$

On note par \hat{q}_n et $\hat{\beta}_n$ les estimateurs du maximum de vraisemblance respectivement de q et β . On testera l'adéquation de l'échantillon K_1, \dots, K_n à la loi de Weibull de type I $\mathcal{W}_I(q, \beta)$, avec q et β inconnus, en testant l'adéquation de l'échantillon des

$$\widehat{W}_i = \left[(K_i - 1)^{\hat{\beta}_n} + \frac{1}{\ln \hat{q}_n} \ln \left[1 - (1 - \hat{q}_n^{K_i^{\hat{\beta}_n} - (K_i-1)^{\hat{\beta}_n}}) U_i \right] \right]^{\frac{1}{\hat{\beta}_n}}, \quad i = 1, \dots, n, \text{ à la loi de Weibull.}$$

Pour cela on pourra utiliser un test basé sur la fonction de répartition empirique tel que le test d'Anderson-Darling, ou le test de Mann-Scheuer-Fertig [53], décrits ci-dessous.

Le test d'Anderson Darling à la loi de Weibull (voir [23]) utilise la propriété suivante : si la v.a. $W \rightsquigarrow \mathcal{W}(\eta, \beta)$, alors la v.a. $Y = -\ln W$ est de loi de Gumbel $G(\frac{1}{\beta}, -\ln \eta)$. Il suffit alors d'utiliser la procédure classique des tests basés sur la fonction de répartition empirique telle que Stephens et D'Agostino la décrivent page 149 de [23] pour la loi de Gumbel.

Mann, Scheuer et Fertig [53] ont proposé un test d'adéquation à la loi de Weibull pour des données censurées de type II dont nous présentons seulement la version pour des données non censurées par soucis de simplicité. On pose $Y_i = \ln \widehat{W}_i$.

On définit alors

$$\forall i = 1, \dots, n, \quad l_i = \frac{Y_{(i+1)} - Y_{(i)}}{\ln \ln \left[\frac{4n+1}{4(n-i-1)+3} \right] - \ln \ln \left[\frac{4n+1}{4(n-i)+3} \right]}$$

La statistique de test pour tester l'hypothèse que l'échantillon $\widehat{W}_1, \dots, \widehat{W}_n$ est issu d'une loi de Weibull est :

$$M = \frac{\lfloor \frac{n}{2} \rfloor \sum_{i=\lfloor \frac{n}{2} \rfloor + 1}^{n-1} l_i}{\lfloor \frac{n-1}{2} \rfloor \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} l_i}$$

Sous l'hypothèse nulle, la statistique M est approximativement de loi de Fisher $\mathcal{F}_{(2\lfloor \frac{n-1}{2} \rfloor, 2\lfloor \frac{n}{2} \rfloor)}$, approximation valable pour $n \geq 20$.

7.5.2 Comparaison des tests

L'étude du seuil a été menée seulement avec les tests d'adéquation de Mann-Scheuer-Fertig et Anderson-Darling, utilisant tous les deux la transformation de Smirnov généralisée. Par manque de temps, les simulations pour les tests basés sur la fonction de répartition empirique seront effectuées plus tard. Les valeurs des paramètres utilisés pour les simulations sont : $q \in \{0.5, 0.7, 0.9, 0.99, 0.999\}$ et $\beta \in \{0.8, 1.2, 1.5, 2.0\}$. Le tableau TAB. 7.11 donne les pourcentages de rejets des tests au seuil $\alpha = 5\%$ pour $N = 2000$ répliques.

Pour $q > 0.7$, les seuils empiriques du test d'Anderson-Darling sont légèrement supérieurs à 5% lorsque $n \leq 40$, quel que soit β . Pour $q \leq 0.7$, les quantiles empiriques sont inférieurs au seuil théorique, quels que soient β et n . Pour le test de Mann-Scheuer-Fertig, les quantiles empiriques sont toujours nettement inférieurs à 5%, quel que soit le triplet (q, β, n) . Par conséquent, il vaudra sans doute mieux ne pas utiliser ce test.

De même que pour l'étude du seuil des tests d'adéquation, l'étude de puissance ne concerne que les deux tests utilisant la transformation de Smirnov généralisée. Pour les simulations, nous avons choisi les lois d'EGGENBERGER-PÓLYA, logistique tronquée discrétisée et Weibull de type III, car à l'instar de la loi de Weibull de type I, elles peuvent modéliser le vieillissement. Les valeurs des paramètres de ces lois ont été choisies de façon à obtenir des données de l'ordre de grandeur allant de la dizaine jusqu'à la dizaine de milliers. Le seuil des tests est fixé à 10%.

Le test de Mann-Scheuer-Fertig est le moins puissant des deux tests, avec de plus un comportement très inégal de sa puissance, dépendant des valeurs des paramètres de la loi de Weibull de type III simulée et de la taille de l'échantillon. Il semble que la puissance du test dépende fortement du paramètre β . Ce point pourra être vérifié par des simulations supplémentaires avec $\beta > 1$. La puissance du test d'Anderson-Darling se comporte de façon analogue, en étant toutefois tout le temps supérieure.

	n = 20		n = 40		n = 60		n = 100		n = 200	
$\beta = 0.8$										
q	\widehat{A}_{TSG}^2	\widehat{MSF}_{TSG}^2								
0.999	6.3	3.2	5.8	3.6	4.8	2.8	5.2	2.9	5.2	2.9
0.99	6.2	3.0	5.8	3.4	4.8	3.0	5.0	3.0	5.2	2.9
0.9	6.0	3.1	5.8	3.4	4.7	2.8	5.2	2.2	5.6	3.2
0.7	3.9	2.4	3.4	2.6	3.8	2.6	4.0	2.6	4.4	3.0
0.5	3.0	1.4	4.1	2.2	4.0	2.1	5.4	3.1	6.5	4.8
$\beta = 1.2$										
0.999	6.3	3.2	5.7	3.5	4.8	2.8	5.1	2.9	5.1	2.7
0.99	6.3	3.1	5.9	3.4	4.8	3.0	5.2	3.1	5.6	3.0
0.9	5.7	3.1	5.9	3.7	4.9	3.1	5.1	2.4	5.5	3.0
0.7	4.4	2.2	3.7	2.6	3.8	2.7	4.6	2.1	4.7	2.5
0.5	3.0	2.0	3.3	2.1	4.0	2.0	4.1	2.3	4.3	2.3
$\beta = 1.5$										
0.999	6.3	3.3	5.8	3.4	4.7	3.0	5.1	3.0	5.2	2.8
0.99	6.3	3.3	5.6	3.7	4.6	2.9	5.3	3.0	5.8	2.6
0.9	5.3	3.2	5.2	3.4	4.9	3.2	4.9	2.2	5.6	3.0
0.7	4.2	2.4	3.9	2.2	3.8	2.2	4.2	2.0	4.5	2.6
0.5	3.3	1.9	3.1	1.7	3.5	1.4	3.7	1.9	3.5	2.1
$\beta = 2$										
0.999	6.2	3.4	5.5	3.9	4.9	2.8	5.5	3.0	5.2	3.2
0.99	5.9	2.9	5.8	3.9	4.8	3.0	5.2	3.0	5.8	2.9
0.9	5.1	3.2	5.9	2.8	5.4	2.8	5.7	2.2	5.1	2.7
0.7	4.8	1.9	5.1	2.3	4.8	2.7	4.5	2.8	5.0	2.8
0.5	3.3	1.9	2.7	1.5	3.3	1.7	3.7	2.0	3.7	1.5

TAB. 7.11 – Seuil empirique pour les tests au niveau $\alpha = 0.05$

	n = 20		n = 40		n = 60		n = 100		n = 200	
(c, β)	\widehat{A}_{TSG}^2	\widehat{MSF}_{TSG}^2								
(0.1,1.0)	6.0	3.5	5.9	3.3	6.0	3.1	7.4	3.8	9.6	4.9
(0.1,0.5)	14.0	7.0	22.6	11.8	36.0	16.8	59.8	28.5	94.1	54.2
(0.05,0.5)	14.0	7.5	22.2	12.8	34.2	19.7	57.9	32.6	93.1	60.8
(0.01,0.5)	11.8	6.0	15.8	9.0	24.0	13.6	39.7	21.3	78.1	43.0
(0.001,0.5)	7.9	3.9	9.5	5.5	13.0	5.9	19.0	9.9	41.2	18.9
(0.0001,0.5)	6.7	3.2	6.7	3.8	7.4	3.8	9.2	5.2	17.4	8.3
(0.0001,1.0)	100	77.8	100	78.6	100	82.3	100	86.5	100	93.5

TAB. 7.12 – Puissance des tests à la loi de Weibull I pour des simulations de la loi de Weibull de type III

	n = 20		n = 40		n = 60		n = 100		n = 200	
(h, d)	\widehat{A}_{TSG}^2	\widehat{MSF}_{TSG}^2								
(4,2)	28.1	6.10	57.9	6.8	86.5	7.7	100	6.2	100	7.4
(20,5)	15.9	10.5	21.8	15.0	28.6	18.6	40.6	28.5	67.8	45.7
(1000,50)	24.6	19.0	39.40	34.0	53.5	48.6	77.9	68.1	97.1	94.7
(5000,1500)	14.2	11.3	20.2	14.5	27.6	22.1	39.2	31.7	64.5	52.3

TAB. 7.13 – Puissance empirique des tests à la loi de Weibull I pour des simulations de la loi d'EGgenberger-Pólya

Le test d'Anderson-Darling est semble-t'il plus puissant que le test de Mann-Scheuer-Fertig. La dépendance de la puissance à l'ordre de grandeur des données est très nette pour le test \widehat{A}_{TSG}^2 .

	n = 20		n = 40		n = 60		n = 100		n = 200	
(c,d)	\widehat{A}_{TSG}^2	\widehat{MSF}_{TSG}^2								
(4,2)	38.1	15.1	77.2	21.3	95.9	28.8	100	39.4	100	63.4
(20,10)	20.6	16.3	37.0	25.3	53.8	31.3	76.8	46.4	97.6	70.7
(100,50)	20.0	15.6	34.9	24.6	53.1	30.8	71.4	44.9	95.5	71.6
(1000,500)	20.9	17.1	36.3	24.3	32.0	32.0	73.3	45.6	95.6	72.3
(10000,5000)	21.2	17.2	35.7	22.3	51.2	33.3	73.3	46.2	95.8	70.5
(25000,7500)	23.0	21.7	42.8	31.3	59.7	36.6	80.9	44.1	97.9	60.5

TAB. 7.14 – Puissance empirique des tests à la loi de Weibull I pour des simulations de la loi logistique tronquée discrétisée

Dans le cas de la simulation de la loi logistique tronquée discrétisée, le test d'Anderson-darling est encore le plus puissant des deux tests. La puissance des tests est croissante avec la taille de l'échantillon. L'ordre de grandeur des données ne semble pas jouer un rôle important pour les deux tests, excepté pour le test \widehat{A}_{TSG}^2 avec des données d'un petit ordre de grandeur.

Conclusion

Le test d'Anderson-Darling est plus puissant que le test de Mann-Scheuer-Fertig pour toutes les lois simulées et les tailles d'échantillon considérées. Il semble que la loi de Weibull de type III soit pour les deux tests, la loi la plus difficile à différencier de la loi de Weibull de type I. Pour toutes les lois simulées, la puissance du test d'Anderson-Darling est excellente pour des données d'un petit ordre de grandeur.

7.6 Conclusion

On a vu à la section 7.1 que seul le graphe de probabilité de la loi géométrique est vraiment un test graphique efficace. Pour les autres lois discrètes, la transformation fonctionnelle ou les qualités de l'estimateur empirique utilisé, sont la cause de l'impossibilité à conclure efficacement sur le rejet du modèle étudié. On ne conseillera finalement l'utilisation du graphe

de probabilité que pour la loi géométrique, et l'utilisation des tests statistiques d'adéquation pour toutes les lois. Pour la loi géométrique, on conseillera indifféremment l'utilisation du test d'Anderson-Darling utilisant la transformation de Smirnov généralisée ou du test de Rueda *et al.* pour leur puissance générale au vu de la grande variété de données testées. Pour tester l'adéquation à la loi de Weibull de type I, l'étude de puissance doit être complétée par les tests basés sur la fonction de répartition empirique et d'autres valeurs des paramètres des lois simulées.

Procédure d'analyse du retour d'expérience en temps discret

Comme on l'a précisé au chapitre 1, la procédure actuelle d'analyse de données du Retour d'EXpérience (REX) en temps continu chez Schneider Electric est la suivante :

1. Tester si le taux de défaillance est constant, c'est à dire faire un test d'adéquation à la loi exponentielle (voir CEI 60605-6).
2. Si la loi exponentielle n'est pas rejetée, on donne une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse exponentielle (voir CEI 60605-4).
3. Si la loi exponentielle est rejetée, on fait un test d'adéquation à la loi de Weibull. Si la loi de Weibull n'est pas rejetée, on donne une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse de Weibull (voir CEI 61649).
4. Si la loi de Weibull est rejetée, une étude spécifique est nécessaire.

Un des objectifs annoncés de la thèse était de proposer son analogue en temps discret, qui a priori devrait être constitué des étapes suivantes :

1. Tester si le taux de défaillance est constant, c'est à dire faire un test d'adéquation à la loi géométrique.
2. Si la loi géométrique n'est pas rejetée, donner une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse géométrique.
3. Si la loi géométrique est rejetée, existe-t'il une loi de Weibull en temps discret? Faire alors un test d'adéquation à la loi de Weibull discrète. Si la loi de Weibull discrète n'est pas rejetée, donner une estimation ponctuelle et un intervalle de confiance unilatéral à 60% du taux de défaillance et de la fiabilité sous l'hypothèse de Weibull discrète.

4. Si la loi de Weibull discrète est rejetée, quels sont les autres modèles de fiabilité en temps discret disponibles?

Dans ce chapitre, nous décrivons la procédure d'analyse du REX effectivement adoptée, qui diffère légèrement de celle décrite ci-dessus. L'analyse présentée exploite les résultats des chapitres précédents et est directement utilisable par un praticien n'ayant pas lu les chapitres précédents.

Un logiciel de Fiabilité et d'Exploitation pour des Données Discrètes de Retour d'Expérience (*FEDDRE*) a été développé sous MATLAB. Dans sa version actuelle, le logiciel FEDDRE permet :

- la simulation des principaux modèles de fiabilité en temps discret,
- la description statistique des données, réelles ou simulées,
- l'estimation non paramétrique (ponctuelle) des principales grandeurs de la fiabilité,
- la construction de tests d'adéquation graphiques,
- l'estimation des paramètres des modèles de fiabilité et la réalisation de prévisions.

Ce logiciel possède une interface graphique conviviale, réalisée entièrement sous MATLAB, qui facilite son utilisation par le personnel de Schneider Electric familiarisé avec les méthodes statistiques.

8.1 Procédure d'analyse

On suppose que l'on dispose de données k_1, \dots, k_n (non censurées) représentant les durées de vie en temps discret de n systèmes identiques et indépendants. Nous proposons de commencer par une étude non paramétrique des données, puis seulement ensuite de tester l'adéquation des observations à la famille des lois géométriques comme le veut la procédure d'analyse proposée.

8.1.1 Notations

On notera la statistique d'ordre par : $k_{(1)} \leq \dots \leq k_{(n)}$. Si des répétitions existent, on notera par $k_{(1)}^* < k_{(2)}^* < \dots < k_{(l)}^*$, $l \leq n$, les observations sans répétitions. Soient d_j le nombre de défaillances à la date $k_{(j)}^*$ et r_j le nombre cumulé de défaillances entre l'instant initial et $k_{(j)}^*$:

$$r_j = \sum_{k=1}^j d_k$$

On notera par n_j le nombre d'appareils encore en fonctionnement à l'instant $k_{(j)}^*$: $n_j = n - r_j$, avec $n_0 = n$.

8.1.2 Estimation non paramétrique

1) Estimation du taux de défaillance

Le taux de défaillance empirique est défini par :

$$\forall k \leq k_{(l)}, \quad \lambda_n(k) = \frac{\text{nombre d'observations} = k}{\text{nombre d'observations} \geq k}.$$

2) Estimation de la fiabilité

La fiabilité empirique est définie par :

$$\forall k \in \mathbb{N}^*, \quad \mathbb{R}_n(k) = \frac{\text{nombre d'observations} > k}{\text{nombre total d'observations}}.$$

3) Estimation par intervalle de confiance du taux de défaillance

Pour le taux de défaillance, on utilise un intervalle de confiance non paramétrique avec la transformation du logit afin d'être sûr d'avoir des bornes comprises entre 0 et 1. La borne supérieure d'un intervalle de confiance asymptotique unilatéral $[0, \lambda_{sup}(k)]$, de seuil $\alpha \in]0, 1[$ pour $\lambda(k)$ est donnée par :

$$\lambda_{sup}(k) = \frac{\lambda_n(k)}{\lambda_n(k) + (1 - \lambda_n(k)) \exp \left[-\frac{z_{2\alpha}}{\sqrt{\lambda_n(k) \mathbb{R}_n(k)}} \right]}$$

où $z_{2\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi normale centrée réduite.

4) Estimation par intervalle de confiance de la fiabilité

Pour la fiabilité, on utilise un intervalle de confiance non paramétrique exact. Un intervalle de confiance unilatéral exact de seuil $\alpha \in]0, 1[$ pour $R(k)$ est donné par $[R_{inf}(k), 1]$, avec :

$$R_{inf}(k) = \left\{ 1 + \frac{(n - n\mathbb{R}_n(k) + 1) \mathcal{F}(1 - \alpha; 2n - 2n\mathbb{R}_n(k) + 2, 2n\mathbb{R}_n(k))}{n\mathbb{R}_n(k)} \right\}^{-1}$$

où $\mathcal{F}_{(p; \nu_1, \nu_2)}$ est le quantile d'ordre p de la loi de Fischer-Snedecor à (ν_1, ν_2) degrés de liberté.

5) Estimation par bande de confiance du taux de défaillance

Une bande de confiance asymptotique $[\lambda''_{inf}(k), \lambda''_{sup}(k)]$, $\forall k \in [a, b]$, de seuil $\alpha \in]0, 1[$ pour λ est donnée par :

$$\begin{aligned} \lambda''_{inf}(k) &= \lambda_n(k) - q_{1-\alpha} \sqrt{\frac{\tilde{\Pi}_n(k)}{n}} \\ \lambda''_{sup}(k) &= \lambda_n(k) + q_{1-\alpha} \sqrt{\frac{\tilde{\Pi}_n(k)}{n}} \end{aligned}$$

avec

- $\tilde{\Pi}_n(k) = \frac{\mathbb{R}_n(k)}{\mathbb{R}_n(k-1)^3} [\mathbb{R}_n(k-1) - \mathbb{R}_n(k)],$
- $q_{1-\alpha}$ tel que $\phi^{b-a+1}(q_{1-\alpha}) - [1 - \phi(q_{1-\alpha})]^{b-a+1} = 1 - \alpha,$
- ϕ est la fonction de répartition de la loi normale centrée réduite.

6) Estimation par bande de confiance de la fiabilité

Une bande de confiance asymptotique de Kolmogorov-Smirnov $[R_{inf}(k), R_{sup}(k)], \forall k \in [1, k_{(l)}[$, de seuil $\alpha \in]0, 1[$ pour R est donnée par :

$$\begin{aligned} R_{inf}(k) &= \mathbb{R}_n(k) - \sqrt{-\frac{1}{2n} \ln \frac{\alpha}{2}} \\ R_{sup}(k) &= \mathbb{R}_n(k) + \sqrt{-\frac{1}{2n} \ln \frac{\alpha}{2}} \end{aligned}$$

8.1.3 Test graphique d'adéquation à la famille des lois géométriques

Avant de faire un test statistique d'adéquation à la famille des lois géométriques, on effectue un test graphique.

Le graphe de probabilité de la loi géométrique est le nuage des points $\left(k_{(j)}^*, \ln\left(1 - \frac{r_j}{n}\right)\right)_{j=1, \dots, l}$. Si les points du nuage sont approximativement alignés selon une droite de pente négative et passant par l'origine, on admettra que la loi géométrique est un modèle acceptable pour ces données. Dans ce cas, la pente de la droite fournit un estimateur de $\ln(1-p)$ donc de p .

8.1.4 Test statistique d'adéquation à la famille des lois géométriques

L'étude de puissance des tests d'adéquation menée au chapitre 7 recommande le test d'Anderson-Darling utilisant la transformation de Smirnov-généralisée. Le paramètre p de la loi géométrique étant inconnu, on commence par l'estimer.

La procédure pour tester l'adéquation de l'échantillon K_1, \dots, K_n à la loi géométrique $\mathcal{G}(p)$, avec p inconnu, est :

1. Calculer $\hat{p}_n = \frac{n}{\sum_{i=1}^n K_i},$
2. simuler n variables aléatoires indépendantes de loi uniforme sur $[0, 1]: U_1, \dots, U_n,$
3. calculer $\hat{T}_i = K_i - 1 + \frac{\ln(1 - \hat{p}_n U_i)}{\ln(1 - \hat{p}_n)},$ pour $i = 1, \dots, n,$
4. calculer, à partir de $\hat{T}_1, \dots, \hat{T}_n,$ la statistique \hat{A}_{TSG}^2 du test d'adéquation d'Anderson-Darling à la famille exponentielle :
 - Ordonner les données : $\hat{T}_{(1)}, \dots, \hat{T}_{(n)},$
 - calculer $\hat{\lambda}_n = \frac{n}{\sum_{i=1}^n \hat{T}_i},$
 - calculer $Z_{(i)} = 1 - \exp[-\hat{\lambda}_n \hat{T}_{(i)}],$ pour $i = 1, \dots, n,$

– calculer la statistique de test :

$$\widehat{A}_{TSG}^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1) \ln Z_{(i)} + (2n+1-2i) \ln(1-Z_{(i)})]$$

Pour les échantillons de petite taille, d'Agostino et Stephens [23] conseillent de corriger cette statistique en la multipliant par $(1 + \frac{0.6}{n})$. Si $\widehat{A}_{TSG}^2 > c_\alpha$, où c_α est tabulé dans [23], on décide de rejeter au seuil α l'hypothèse que les données sont issues de la loi géométrique. La procédure d'analyse se poursuit alors à la section 8.1.6. Sinon, aucune preuve ne permet de rejeter l'hypothèse d'une loi géométrique pour les données et l'étape suivante est la section 8.1.5.

8.1.5 Estimation des fonctions usuelles de la fiabilité pour la loi géométrique

Pour la loi géométrique, les grandeurs usuelles de la fiabilité s'expriment simplement en fonction du paramètre de la loi, nous commençons par donner l'expression de son estimateur.

1) Estimation du paramètre p

L'estimateur sans biais et de variance minimale (ESBVM) du paramètre p est donné par :

$$\tilde{p}_n = \frac{n-1}{\sum_{i=1}^n K_i - 1}$$

2) Estimation du taux de défaillance

Le taux de défaillance de la loi géométrique est constant et égal au paramètre p . L'estimateur du taux de défaillance est donc égal à \tilde{p}_n .

3) Estimation du temps moyen de bon fonctionnement (MTTF)

Le MTTF de la loi géométrique est égal à l'inverse du paramètre de la loi. L'ESBVM du MTTF est : $\widehat{MTTF} = \frac{1}{\tilde{p}_n} = \frac{1}{n} \sum_{i=1}^n K_i$.

Remarque :

La durée de vie résiduelle moyenne de la loi géométrique étant constante (et donc égale au MTTF), on obtient l'ESBVM, \hat{m} , de la durée de vie résiduelle moyenne à partir de celui du MTTF donné ci-dessus : $\hat{m} = \frac{1}{\tilde{p}_n}$.

4) Estimation de la fiabilité

Pour tout $k \in \mathbb{N}^*$, l'estimateur sans biais de variance minimale $\tilde{R}(k)$ de $R(k)$ est :

$$\tilde{R}(k) = \begin{cases} \prod_{i=1}^{n-1} \left(1 - \frac{k}{S-i}\right) = \frac{C_{S-k-1}^{n-1}}{C_{S-1}^{n-1}} & \text{si } k \leq S-n \\ 0 & \text{si } k > S-n \end{cases}$$

où $S = \sum_{i=1}^n K_i$.

5) Estimation par intervalle de confiance du paramètre p

Les conclusions de la section 6.1.2 consacrée aux intervalles de confiance pour p recommandent l'utilisation de l'intervalle de confiance basé sur la méthode bayésienne.

L'intervalle de confiance bilatéral $[p_{inf}, p_{sup}]$, de seuil α , $\alpha \in]0, 1[$, pour le paramètre p est donné par :

$$p_{inf} = \frac{(n+1)\mathcal{F}_{(\alpha/2; 2(n+1), 2(S-n+1))}}{S-n+1 + (n+1)\mathcal{F}_{(1-\alpha/2; n+1, S-n+1)}}$$

$$p_{sup} = \frac{(n+1)\mathcal{F}_{(1-\alpha/2; n+1, S-n+1)}}{S-n+1 + (n+1)\mathcal{F}_{(1-\alpha/2; 2(n+1), 2(S-n+1))}}$$

où $\mathcal{F}_{(p; \nu_1, \nu_2)}$ désigne le quantile d'ordre p de la loi de Fischer-Snedecor à (ν_1, ν_2) degrés de liberté, et $S = \sum_{i=1}^n K_i$.

Pour la construction d'un intervalle de confiance unilatéral $[0, p'_{sup}]$, de seuil α ($\alpha \in]0, 1[$), la borne supérieure est donnée par :

$$p'_{sup} = \frac{(n+1)\mathcal{F}_{(1-\alpha; n+1, S-n+1)}}{S-n+1 + (n+1)\mathcal{F}_{(1-\alpha; 2(n+1), 2(S-n+1))}}$$

6) Estimation par intervalle de confiance de la fiabilité

L'expression de la fiabilité pour la loi géométrique $\mathcal{G}(p)$ est :

$$R(k) = (1-p)^k$$

La fiabilité étant une fonction décroissante de p , l'intervalle de confiance de la fiabilité s'obtient directement à partir de l'intervalle de confiance pour p .

L'intervalle de confiance bilatéral (respectivement unilatéral) de seuil α pour la fiabilité en k est donc donné par $[(1-p_{sup})^k, (1-p_{inf})^k]$ (respectivement $[(1-p'_{sup})^k, 1]$).

8.1.6 Test graphique d'adéquation à la famille des lois de Weibull de type I

Le graphe de probabilité de la loi de Weibull de type I est donné par :

$\left(\ln k_{(j)}^*, \ln[-\ln(1 - \frac{r_j}{n})] \right)_{j=1, \dots, l}$. Si les points du nuage sont approximativement alignés selon une droite d'équation $y = ak + b$, avec $a > 0$, on admettra que la loi de Weibull de type I est un modèle acceptable pour ces données. Dans ce cas, le coefficient a de la droite fournit un estimateur de β , et $\exp[-\exp(b)]$ un estimateur de q .

8.1.7 Test statistique d'adéquation à la famille des lois de Weibull de type I

On teste l'adéquation de l'échantillon K_1, \dots, K_n à la loi de Weibull de type I $\mathcal{W}_I(q, \beta)$, avec q et β inconnus, par le test d'Anderson-Darling utilisant la transformation de Smirnov généralisée. On commence alors par estimer les paramètres de la loi.

1. Calculer les estimateurs du MV \hat{q}_n et $\hat{\beta}_n$ selon la procédure du 8.1.8.1,
2. simuler n variables aléatoires indépendantes de loi uniforme sur $[0, 1]$: U_1, \dots, U_n ,
3. calculer $\hat{T}_i = \left[(K_i - 1)^{\hat{\beta}_n} + \frac{1}{\ln \hat{q}_n} \ln \left[1 - (1 - \hat{q}_n^{K_i^{\hat{\beta}_n} - (K_i - 1)^{\hat{\beta}_n}}) U_i \right] \right]^{\frac{1}{\hat{\beta}_n}}$, pour $i = 1, \dots, n$,
4. calculer, à partir de $\hat{T}_1, \dots, \hat{T}_n$, la statistique \hat{A}_{TSG}^2 du test d'adéquation d'Anderson-Darling à la famille des lois de Weibull :
 - Transformer l'échantillon des $\hat{T}_1, \dots, \hat{T}_n$ en l'échantillon Y_1, \dots, Y_n avec $Y_i = -\ln \hat{T}_i$,
 - ordonner l'échantillon : $Y_{(1)}, \dots, Y_{(n)}$,
 - à partir de $Y_{(1)}, \dots, Y_{(n)}$, calculer les estimateurs du MV $\hat{\mu}$ et $\hat{\nu}$ de la loi des valeurs extrêmes de fonction de répartition $F(x; \mu, \nu) = \exp \left[-\exp \left(-\frac{x-\mu}{\nu} \right) \right]$,
 - calculer $Z_{(i)} = F(Y_{(i)}; \hat{\mu}, \hat{\nu})$, pour $i = 1, \dots, n$,
 - calculer la statistique de test :

$$\hat{A}_{TSG}^2 = -n - \frac{1}{n} \sum_{i=1}^n \left[(2i - 1) \ln Z_{(i)} + (2n + 1 - 2i) \ln(1 - Z_{(i)}) \right]$$

Pour les échantillons de petite taille, d'Agostino et Stephens [23] conseillent de corriger cette statistique en la multipliant par $(1 + \frac{0.2}{\sqrt{n}})$. Si $\hat{A}_{TSG}^2 > c_\alpha$, avec c_α tabulé dans [23], on décide de rejeter au seuil α l'hypothèse que les données sont issues de la loi de Weibull de type I, l'analyse se poursuit alors à la section 8.1.9. Sinon, aucune preuve ne permet de rejeter l'hypothèse d'une loi de Weibull pour les données, l'étape suivante est alors la section 8.1.8.

8.1.8 Estimation des fonctions usuelles de la fiabilité pour la loi de Weibull de type I

1) Estimation des paramètres

L'estimation par la méthode du Maximum de Vraisemblance des paramètres q et β est obtenue numériquement en maximisant la fonction de vraisemblance \mathcal{L} donnée ci-dessous :

$$\mathcal{L}_n(q, \beta) = \prod_{i=1}^n \left(q^{(k_i - 1)\beta} - q^{k_i^\beta} \right)$$

Soient \hat{q}_n et $\hat{\beta}_n$ les estimateurs du maximum de vraisemblance.

2) Estimation du taux de défaillance

L'expression du taux de défaillance estimé de la loi de Weibull de type I est :

$$\hat{\lambda}(k) = 1 - \hat{q}_n^{k\hat{\beta}_n - (k-1)\hat{\beta}_n}$$

où \hat{q}_n et $\hat{\beta}_n$ sont les estimateurs du maximum de vraisemblance de q et β .

3) Estimation de la fiabilité

L'expression de la fiabilité estimée est :

$$\hat{R}(k) = \hat{q}_n^{k\hat{\beta}_n}$$

L'état actuel des recherches sur la loi de Weibull de type I ne permet pas de construire d'intervalle de confiance pour le taux de défaillance et la fiabilité comme la procédure d'analyse le prévoyait en début de thèse. Néanmoins, il est possible de construire des intervalles de confiance non paramétriques pour λ et R , comme nous les avons présentés au chapitre 5.

8.1.9 Autres modèles de fiabilité

Les applications pratiques sur des données discrètes ont montré que les lois d'Engenberger-Pòlya et logistique tronquée discrétisée peuvent être des modèles envisageables dans le domaine d'application de Schneider Electric. Malheureusement, il n'existe pas encore de test d'adéquation à ces modèles. Néanmoins l'estimation paramétrique est possible.

8.1.10 Récapitulatif

La procédure d'analyse de données du REX en temps discret souhaitée en début de thèse n'est pas applicable complètement à cause de l'absence d'intervalles de confiance pour le taux de défaillance et la fiabilité dans le cas de la loi de Weibull de type I. Néanmoins, il est possible de construire des intervalles de confiance non paramétriques pour λ et R , comme nous les avons présentés au chapitre 5. Même si un intervalle de confiance non paramétrique est moins précis que ne l'est un intervalle de confiance paramétrique, il est utilisable dans tous les cas, même quand aucun modèle n'a été retenu. Le diagramme de la figure FIG. 8.1 illustre la procédure d'analyse effective à la fin de thèse.

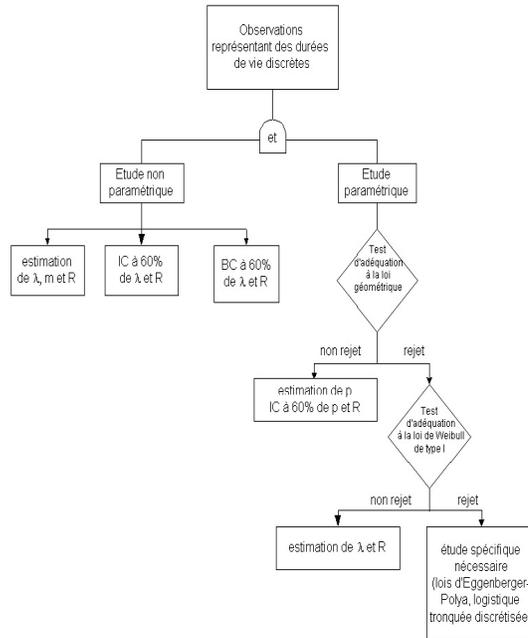


FIG. 8.1 – Procédure effective d'analyse du REX en temps discret

8.2 Application pratique : données non censurées

Nous présentons un jeu de $n = 23$ données représentant des nombres de sollicitations jusqu'à défaillance, provenant d'essais effectués par Schneider Electric :

12 15 15 15 17 18 18 19 20 20 22 22 23 23 24 25 25 25 29 31
32 32

Pour cet exemple, nous avons choisi d'évaluer les grandeurs de la fiabilité à la 20^{ème} sollicitation pour des raisons liées au contexte applicatif de Schneider Electric.

L'analyse des données commence par la construction du graphe de probabilité de la loi géométrique :

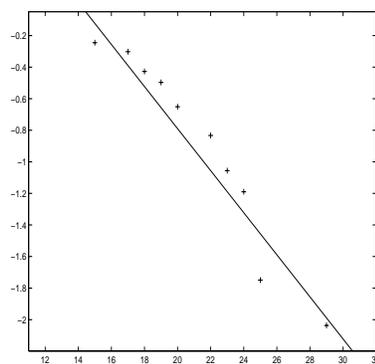


FIG. 8.2 – Graphe de probabilité de la loi géométrique

Il apparaît sur le graphe de probabilité que les points du nuage ne sont pas raisonnablement alignés : le modèle géométrique ne semble donc pas acceptable pour nos données.

La procédure du test d'adéquation à la loi géométrique d'Anderson-Darling utilisant la transformation de Smirnov généralisée appliquée aux données donne le résultat suivant pour la statistique de test corrigée : $\widehat{A}_{TSG}^2 = 5.93$. La valeur critique au seuil de 5% étant $c_{0.05} = 1.321$, le test rejette l'hypothèse géométrique au seuil de 5%. En notant que $c_{0.0025} = 2.534$, on en déduit que le risque de rejeter à tort l'hypothèse géométrique est inférieur à 0.25% et la loi géométrique est donc largement rejetée.

La procédure indique dans le cas du rejet de la loi géométrique, de tester l'adéquation à la loi de Weibull de type I. Nous commençons alors naturellement par construire le graphe de probabilité pour cette loi, donné à la figure FIG. 8.3.

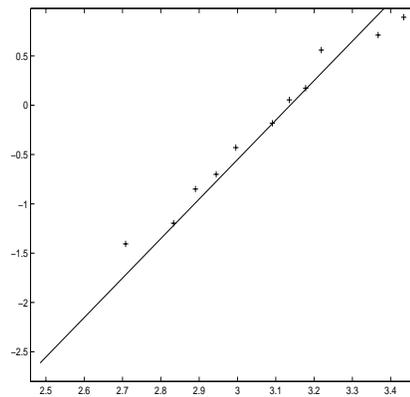


FIG. 8.3 – Graphe de probabilité de la loi de Weibull de type I

Il est difficile de se prononcer sur l'alignement des points du nuage. Il est donc nécessaire d'effectuer un test statistique pour prendre une décision sur le rejet ou non du modèle de Weibull de type I. À titre indicatif, la régression linéaire sur le nuage de points nous donne l'équation de droite suivante : $y = 4.00k - 12.56$. Les estimateurs graphiques des paramètres sont $\hat{q} = 0.99999650$ et $\hat{\beta} = 4.00$.

Le test d'adéquation à la loi de Weibull de type I d'Anderson-Darling utilisant la transformation de Smirnov généralisée appliqué à ces données donne les résultats suivants :

1. Les estimateurs du MV de la loi de Weibull de type I sont : $\hat{q}_n = 0.999997$ et $\hat{\beta}_n = 4.08$,
2. la statistique de test corrigée est $\widehat{A}_{TSG}^2 = 0.414$. La valeur critique est $c_{0.05} = 0.757$. Le test ne rejette pas la loi de Weibull de type I au seuil de 5%. La loi de Weibull de type I est donc un modèle acceptable pour représenter la loi des durées de vie des appareils étudiés.

Remarque :

On constate que l'estimation graphique des paramètres est très proche des estimateurs du MV.

L'estimation paramétrique du taux de défaillance à la 20^{ème} sollicitation est alors : $\hat{\lambda}(20) = 9.7 \cdot 10^{-2}$. L'estimation non paramétrique ponctuelle est $\lambda_n(20) = 0.14$. L'intervalle de confiance non paramétrique unilatéral à 60% pour $\lambda(20)$ est $[0, 0.17]$. L'estimation paramétrique de la fiabilité à la 20^{ème} sollicitation est $\hat{R}(20) = 0.58$. L'estimation non paramétrique ponctuelle de la fiabilité est $R_n(20) = 0.52$. L'intervalle de confiance non paramétrique unilatéral à 60% pour $R(20)$ est $[0.47; 1]$. Les estimations, paramétrique et non paramétrique, de R sont relativement proches l'une de l'autre. Tandis que pour le taux de défaillance, l'estimation non paramétrique est légèrement supérieure à l'estimation paramétrique, donc plus pessimiste. Ceci étant probablement dû au fait que l'estimateur empirique du taux de défaillance est nul en dehors des observations.

Les graphes du taux de défaillance et de la fiabilité estimés à l'aide de la loi de Weibull de type I sont donnés par la figure FIG. 8.4.

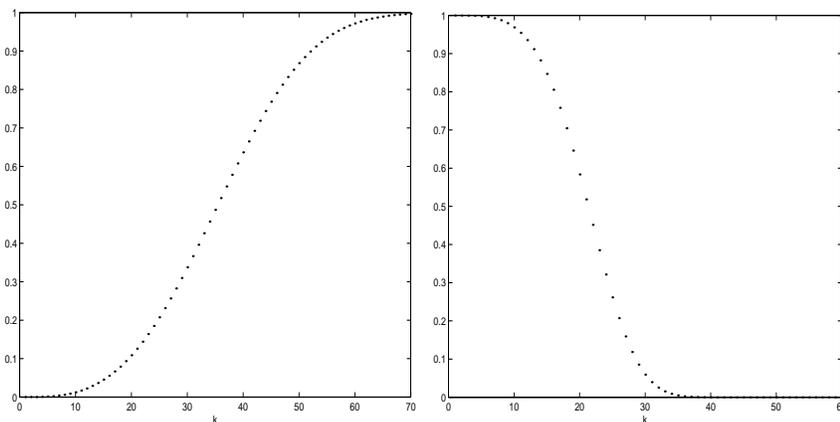


FIG. 8.4 – Taux de défaillance et fiabilité estimés de la loi de Weibull de type I

La figure FIG. 8.5 donne les représentations graphiques de l'intervalle et de la bande de confiance non paramétriques du taux de défaillance.

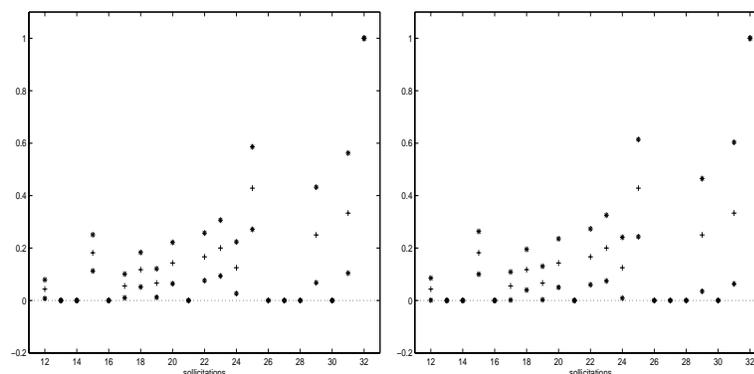


FIG. 8.5 – Intervalle et bande de confiance non paramétriques bilatéraux à 60% du taux de défaillance

La figure FIG. 8.6 donne les représentations graphiques de l'intervalle et de la bande de confiance non paramétriques de la fiabilité.

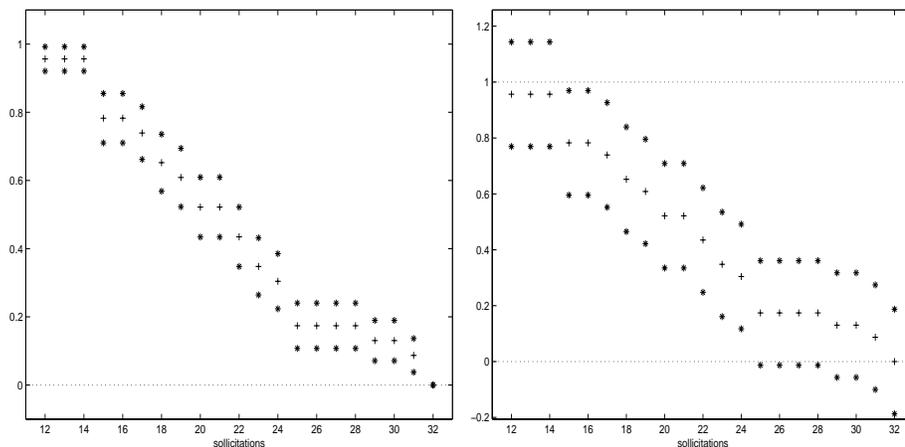


FIG. 8.6 – Intervalle et bande de confiance non paramétriques bilatéraux à 60% de la fiabilité

Les bandes de confiance de λ et R sont logiquement plus larges que les intervalles de confiance correspondants.

À titre de comparaison, nous donnons les résultats d'estimation paramétrique pour la loi géométrique (même si celle-ci a été rejetée par les tests d'adéquation), car cette loi modélise l'absence de vieillissement, hypothèse souvent faite dans les études de fiabilité. L'estimateur (ESBVM) de p est $\hat{p}_n = 4.4 \cdot 10^{-2}$. Un intervalle de confiance unilatéral à 60% pour p est $[0, 5.0 \cdot 10^{-2}]$.

La figure FIG. 8.7 donne les graphes des estimations paramétriques et non paramétriques du taux de défaillance et de la fiabilité.

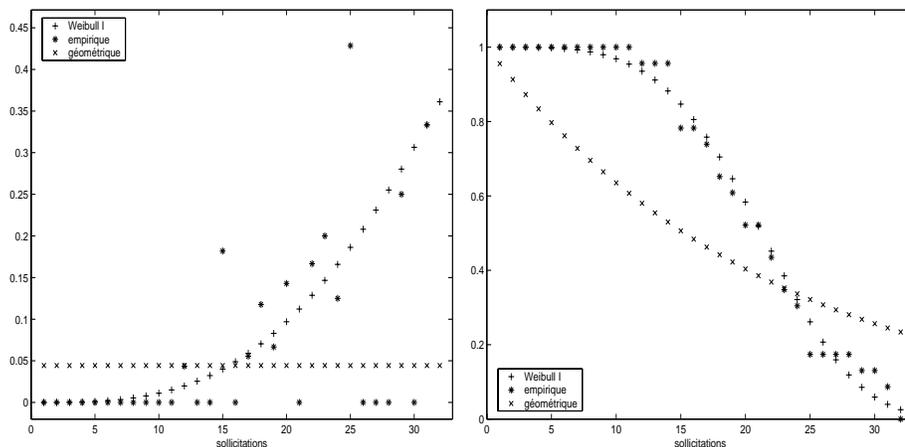


FIG. 8.7 – Comparaison des estimations du taux de défaillance et de la fiabilité

Ces graphes confirment clairement que la loi de Weibull de type I est nettement plus appropriée que la loi géométrique pour ces données.

8.3 Application pratique : données censurées

Dans la pratique, les données issues du REX comportent souvent des censures en plus des dates de défaillances. Aussi, dans cette section, nous présentons un second exemple d'application au REX de Schneider Electric qui se compose de données multicensurées à droite. Comme dans cette thèse, nous n'avons pas étudié le cas de données censurées, nous ne donnons que des résultats de base, issus pour la plupart de résultats usuels en temps continu. L'étude des données censurées constitue un thème de recherches futures.

Les nombres de sollicitations jusqu'à défaillance sont les suivants :

2900 6767 9886 13000 15000 16300 18771 21032 21120 21160 21421 22340 22644 22821
23000 24961 25660 25777 27085 27215 27240 27343 27473 28114 28161 28926 29500

Les nombres de sollicitations jusqu'à censure :

20000 20000 20000 20000 20000 20000 30000 30000 30000

En raison de l'ordre de grandeur des observations et de l'absence d'ex-æquos, il ne semble pas nécessaire d'utiliser un modèle de fiabilité en temps discret pour modéliser la durée de vie des appareils, mais plus simplement un modèle continu usuel. Le but de cet exemple est de montrer, dans un cas extrême de validité de la démarche en temps discret, l'intérêt du cadre discret et également ses limites par rapport au temps continu. Nous expliciterons les calculs de fiabilité en $k = 20000$.

Ces données comportent des censures et à notre connaissance, il n'existe pas de test d'adéquation aux lois discrètes pour ce genre de données. Par conséquent nous ne pouvons tester l'adéquation à la loi géométrique comme le veut la procédure d'analyse.

Nous commençons par présenter les techniques d'estimation non paramétrique pour des données multicensurées à droite. Nous les appliquerons à la construction de graphes de probabilité, puis nous aborderons l'estimation paramétrique.

8.3.1 Estimation non paramétrique

Nous introduisons les notations suivantes, valables pour des données multicensurées à droite de type I. Supposons qu'à la fin d'un test portant sur n appareils identiques, on observe k_1, \dots, k_r réalisations indépendantes d'une variable aléatoire discrète K représentant des durées de bon fonctionnement et t_1, \dots, t_{n-r} , $n - r$ instants de censure. Le problème consiste à estimer, à partir des observations $k_1, \dots, k_r, t_1, \dots, t_{n-r}$, une certaine grandeur choisie de fiabilité de la loi de K . Dans un souci de simplicité, nous reprenons les notations d_j et n_j introduites pour les données non censurées au chapitre 5, avec un changement pour n_j qui représente maintenant le nombre d'appareils encore en test (et non plus encore en fonctionnement) à l'instant $k_{(j)}^*$. La différence tient au fait qu'un appareil censuré peut très bien être en état de bon fonctionnement alors qu'il n'est plus en test. On pose $n_0 = n$.

$k_{(j)}^* (t_{(i)}^*)$	2900	6767	9886	13000	15000	16300	18771	(20000)	21032
d_j	1	1	1	1	1	1	1		1
n_{j-1}	36	35	34	33	32	31	30		23

$k_{(j)}^* (t_{(i)}^*)$	21120	21160	21421	22340	22644	22821	23000	24961	25660	25777
d_j	1	1	1	1	1	1	1	1	1	1
n_{j-1}	22	21	20	19	18	17	16	15	14	13

$k_{(j)}^* (t_{(i)}^*)$	27085	27215	27240	27343	27473	28114	28161	28296	29500	(30000)
d_j	1	1	1	1	1	1	1	1	1	
n_{j-1}	12	11	10	9	8	7	6	5	4	

a) La fiabilité empirique

1) Estimation ponctuelle

On a vu (voir la section 5.3.4) dans le cas de données discrètes non censurées, que la relation entre la fiabilité et le taux de défaillance conduisait à l'estimateur empirique de la fiabilité suivant :

$$\mathbb{R}_n(k) = \prod_{j:k_{(j)}^* \leq k} \frac{n_{j-1} - d_j}{n_{j-1}}$$

C'est exactement l'expression de l'estimateur de la fiabilité pour des données censurées proposé par Kaplan-Meier [45]. Cet estimateur est valable en temps continu comme en temps discret. $\{\mathbb{R}_n(k)\}_{k \geq 1}$ est une suite décroissante, égale à 1 pour $k = 1$, diminuant à chaque $k_{(j)}^*$ et dont la valeur ne change pas aux instants de censure. L'effet de la censure intervient dans le calcul des n_j et par conséquent dans les sauts de $\mathbb{R}_n(k)$. L'estimateur de Kaplan-Meier de la fiabilité est bien défini pour $k \leq \max(k_{(r)}, t_{(n-r)})$. Si la plus grande donnée correspond à une défaillance, alors l'estimateur vaut zéro au delà. En revanche, si la plus grande observation est un instant de censure, alors l'estimateur de Kaplan-Meier de la fiabilité ne s'annulera jamais. C'est la raison pour laquelle on définit seulement l'estimateur de Kaplan-Meier pour $k \leq \max(k_{(r)}, t_{(n-r)})$.

Exemple :

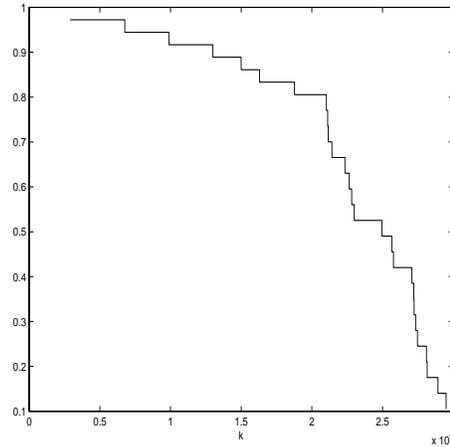


FIG. 8.8 – *Estimateur empirique de Kaplan-Meier de R pour les données multicensurées*

On trouve $\mathbb{R}_n(20000) = 0.806$.

Remarque :

Pour des données non censurées, cette estimation empirique correspond à l'estimation usuelle de la fiabilité par $\mathbb{R}_n(k_{(j)}^*) = \frac{n_j}{n}$ car $n_{j-1} - d_j = n_j$.

L'estimateur $\mathbb{R}_n(k)$ de Kaplan-Meier est un estimateur consistant de $R(k)$ sous certaines conditions générales, de variance asymptotique estimée donnée par :

$$\hat{\sigma}_{\mathbb{R}_n}^2(k) = \mathbb{R}_n^2(k) \sum_{j:k_{(j)}^* \leq k} \frac{d_j}{n_{j-1}(n_{j-1} - d_j)} \quad (8.1)$$

La formule (8.1) est connue sous le nom de "formule de Greenwood". Cox et Oakes montrent dans [21] que la formule de Greenwood est également valable dans le cas discret.

2) Estimation par intervalle de confiance

Un intervalle de confiance exact n'est plus calculable comme dans le cas de données non censurées. En revanche, il est possible de construire un intervalle de confiance asymptotique utilisant l'estimateur de Kaplan-Meier, qui repose sur le fait que la variable

$$Z_{\mathbb{R}_n} = \frac{\mathbb{R}_n(k) - R(k)}{\hat{\sigma}_{\mathbb{R}_n}(k)}$$

est asymptotiquement de loi normale centrée réduite. Un intervalle de confiance asymptotique de seuil α pour $R(k)$ est $[R_{inf}(k), R_{sup}(k)]$, donné par :

$$R_{inf}(k) = \mathbb{R}_n(k) - z_\alpha \hat{\sigma}_{\mathbb{R}_n}(k)$$

$$R_{sup}(k) = \mathbb{R}_n(k) + z_\alpha \hat{\sigma}_{\mathbb{R}_n}(k)$$

Comme nous l'avons vu dans le cas des données non censurées, une meilleure approximation peut être obtenue en utilisant la transformation du logit. La construction de l'intervalle de confiance asymptotique repose sur le fait que la variable aléatoire

$$Z_{logit} = \frac{\text{logit}(\mathbb{R}_n(k)) - \text{logit}(R(k))}{\widehat{\sigma}_{logit}(k)}$$

avec $\widehat{\sigma}_{logit}(k) = \frac{\widehat{\sigma}_{\mathbb{R}_n}(k)}{\mathbb{R}_n(k)(1 - \mathbb{R}_n(k))}$, est asymptotiquement de loi normale centrée réduite. L'intervalle de confiance asymptotique de seuil α pour $R(k)$ est $[R'_{inf}(k), R'_{sup}(k)]$, donné par :

$$R'_{inf}(k) = \frac{\mathbb{R}_n(k)}{\mathbb{R}_n(k) + (1 - \mathbb{R}_n(k))w_\alpha(k)}$$

$$R'_{sup}(k) = \frac{\mathbb{R}_n(k)}{\mathbb{R}_n(k) + (1 - \mathbb{R}_n(k))/w_\alpha(k)}$$

où $w_\alpha(k) = \exp\{z_\alpha \widehat{\sigma}_{logit}(k)\}$.

Exemple :

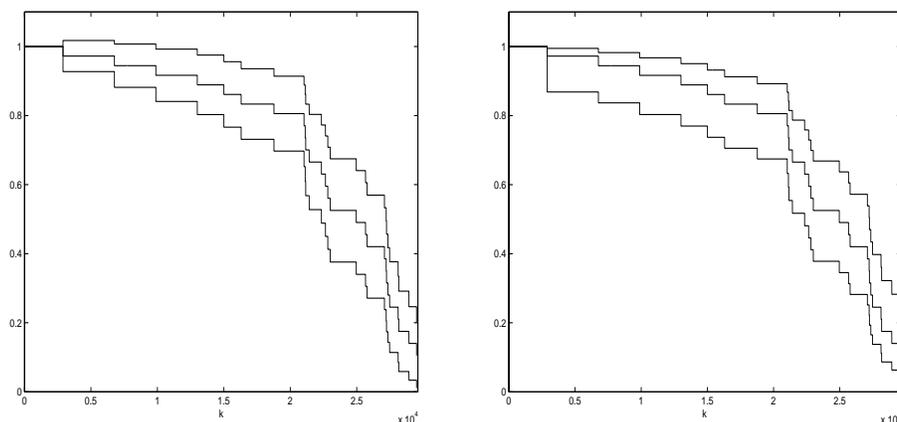


FIG. 8.9 – Intervalles de confiance asymptotiques au seuil $\alpha = 0.1$, sans et avec le logit, de R pour les données multicensurées

On constate donc que les résultats sont très proches. On peut constater sur la figure FIG. 8.9 que l'intervalle de confiance avec logit est un peu plus large, ce que l'on avait déjà observé pour des données non censurées. L'intérêt de la transformation logit intervient seulement lorsque les bornes de l'intervalle de confiance risquent d'être en dehors de l'intervalle $[0, 1]$. Les intervalles de confiance asymptotiques unilatéraux pour $R(20000)$, sans et avec logit, au seuil de $\alpha = 0.40$ sont respectivement $[0.786, 1]$ et $[0.787, 1]$.

b) Le taux de défaillance empirique

La définition du taux de défaillance empirique pour des données censurées est la même que celle utilisée au chapitre 5 pour des données non censurées rappelées ci-dessous :

$$\begin{aligned}\lambda_n(k) &= \frac{\mathbb{P}_n(k)}{\mathbb{R}_n(k-1)} = \frac{\mathbb{R}_n(k-1) - \mathbb{R}_n(k)}{\mathbb{R}_n(k-1)}, \quad \forall k \leq \max(k_{(r)}, t_{(n-r)}) \\ &= \frac{\text{nombre de défaillances} = k}{\text{nombre d'observations} \geq k}.\end{aligned}$$

Les particularités du taux de défaillance empirique sont les suivantes :

1. λ_n n'est pas défini pour $k > \max(k_{(r)}, t_{(n-r)})$, puisque le dénominateur n'est pas défini,
2. si k n'est pas égal à une défaillance, alors $\lambda_n(k) = 0$,
3. $\lambda_n(k_{(j)}^*) = \frac{d_j}{n_{j-1}}$, $j = 1, \dots, r$.

Nous avons vu que cet estimateur du taux de défaillance est utilisé par Kaplan et Meier pour estimer la fonction de fiabilité.

Exemple :

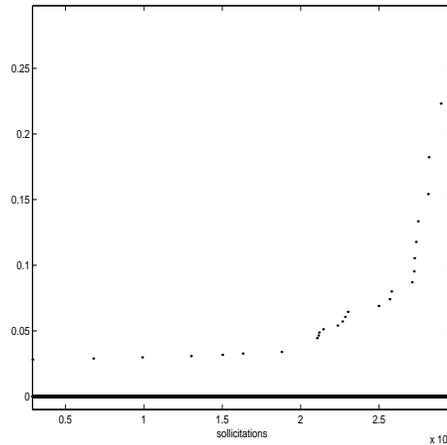


FIG. 8.10 – Taux de défaillance empirique pour les données multicensurées

Aucune défaillance n'intervenant à la 20000^{ème} sollicitation, l'estimateur empirique de $\lambda(20000)$ est donc nul. L'ordre de grandeur élevé et l'absence d'ex-æquo parmi les observations rendent cet estimateur peu utile car presque toujours nul. Ce point est également perturbateur dans l'appréciation du sens de variation du taux de défaillance.

c) La durée de vie moyenne résiduelle empirique

À partir de l'estimateur de Kaplan-Meier de la fiabilité, il est possible d'utiliser la relation entre la fiabilité et la durée de vie résiduelle moyenne pour construire un estimateur empirique,

noté m_n de la durée de vie résiduelle moyenne pour des données multicensurées :

$$m(k) = \frac{1}{R(k)} \sum_{j=k}^{\infty} R(j)$$

$$\Rightarrow m_n(k) = \frac{1}{\mathbb{R}_n(k)} \sum_{j=k}^{\max(k_{(r)}, t_{(n-r)})} \mathbb{R}_n(j)$$

La sommation dans l'expression de $\hat{m}(k)$ est limitée par la plus grande observation car l'estimateur de Kaplan-Meier n'est pas défini au delà.

Il n'est pas possible d'utiliser la relation de récurrence entre la durée de vie résiduelle moyenne et le taux de défaillance pour obtenir un estimateur empirique de m , car on ne connaît pas la valeur de $m_n(0)$, c'est à dire d'un estimateur empirique de la moyenne pour des données censurées.

Exemple :

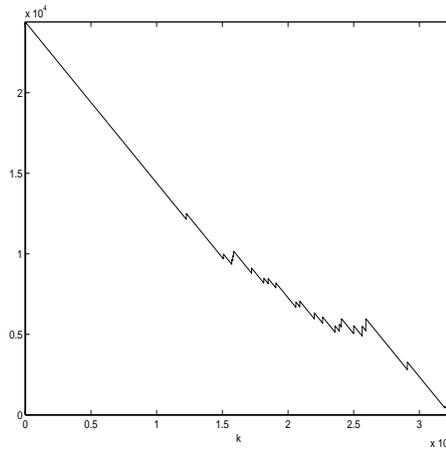


FIG. 8.11 – *Durée de vie résiduelle moyenne empirique pour les données multicensurées*

On obtient $m_n(20000) = 5717.3$ sollicitations.

8.3.2 Tests d'adéquation graphiques

Tous les tests graphiques d'adéquation (voir la section 7.1) construits pour les modèles de fiabilité en temps discret restent valables pour des données censurées. Pour chaque loi, la transformation fonctionnelle de la grandeur de fiabilité étant toujours valable, il suffit de remplacer la grandeur de fiabilité par l'estimateur non paramétrique pour des données multicensurées et de tracer seulement les points correspondants aux défaillances.

Exemple :

Le nuage de points du graphe de probabilité de la loi géométrique est le suivant :

$$\left(k_{(j)}^*, \ln \mathbb{R}_n(k_{(j)}^*) \right)_{j=1, \dots, r}.$$

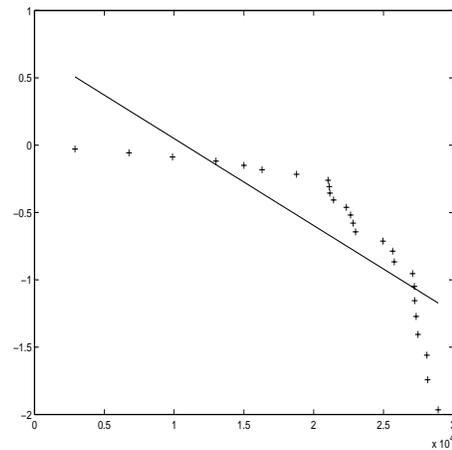


FIG. 8.12 – Graphe de probabilité de la loi géométrique pour les données multicensurées

La figure FIG. 8.12 indique clairement que la loi géométrique n'est pas un modèle acceptable pour les données étudiées.

Le graphe de probabilité de la loi de Weibull de type I est : $\left(\ln k_{(j)}^*, \ln \left[-\ln \left(1 - \frac{r_j}{n} \right) \right] \right)_{j=1, \dots, l}$.

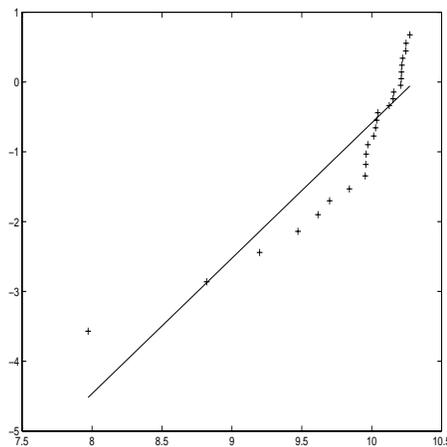


FIG. 8.13 – Graphe de probabilité de la loi de Weibull de type I pour les données multicensurées

La figure FIG. 8.13 indique clairement que la loi de Weibull de type I n'est pas un modèle acceptable pour les données étudiées.

Pour les autres modèles de fiabilité en temps discret, les résultats médiocres obtenus dans le cas non censuré pour le test graphique ou l'impossibilité de construire un test graphique, aboutissent à l'impossibilité de choisir un modèle pour ces données multicensurées.

8.3.3 Estimation paramétrique

Pour des données censurées, la méthode des moments n'est plus valable comme cela pouvait être le cas pour des données non censurées, seule l'estimation par la méthode du

Maximum de Vraisemblance est disponible. L'estimation par la méthode du Maximum de Vraisemblance du vecteur θ des paramètres est obtenue en maximisant la fonction de vraisemblance \mathcal{L} donnée ci-dessous :

$$\mathcal{L}(\theta) = \prod_{i=1}^r p(k_i) \prod_{j=1}^{n-r} R(t_j)$$

Remarque :

La loi géométrique est la seule loi discrète qui possède une forme explicite pour l'estimateur du MV de son paramètre :

$$\hat{p}_n = \frac{r}{\sum_{i=1}^r k_i + \sum_{j=1}^{n-r} t_j}$$

Pour toutes les autres lois discrètes, une résolution numérique de l'équation de vraisemblance est nécessaire pour obtenir les estimateurs du MV.

Exemple :

L'estimation du paramètre de la loi géométrique donne : $\hat{p}_n = 3.3 \cdot 10^{-5}$. Pour la loi de Weibull de type I, on obtient : $\hat{\eta}_n = 25970.4$ et $\hat{\beta}_n = 3.7$. La paramétrisation de la loi de Weibull de type I avec q au lieu de η n'est pas utilisable en raison de l'ordre de grandeur élevé des observations qui conduit à une estimation de q égale à 1.

Finalement, pour ce jeu de données particulier, seule l'analyse non paramétrique permet d'obtenir des résultats exploitables.

8.3.4 Conclusion

L'analyse non paramétrique de données discrètes censurées est presque aussi développée que pour des données non censurées. À notre connaissance, il n'existe pas de test statistique d'adéquation à des données discrètes censurées, les seuls tests disponibles sont les tests graphiques. A priori, comme dans le cas non censuré, le seul test graphique à conseiller est celui de la loi géométrique, les possibilités de tests sont alors extrêmement réduites par rapport au cas de données non censurées. Ce jeu de données a également mis en évidence les limites de l'analyse paramétrique de données censurées en temps discret lorsque l'ordre de grandeur des observations traitées est élevé.

Conclusion et perspectives

Ce travail de recherche constitue une première approche d'un domaine de recherche peu souvent étudié, celui de la fiabilité des systèmes non réparables en temps discret. Nous présentons un éventail de concepts et de méthodes qui se veut aussi complet et synthétique que possible, depuis une réflexion sur les concepts de base jusqu'à l'application aux données réelles, en passant par la description de modèles et de méthodes statistiques permettant leur utilisation. Nous espérons ainsi favoriser la diffusion des concepts de fiabilité en temps discret.

Cette étude a permis de montrer que les différences importantes entre les notions de fiabilité en temps continu et en temps discret proviennent principalement de la définition usuelle du taux de défaillance en temps discret. La nouvelle définition que nous avons proposé permet de résoudre ces problèmes. Il ressort de notre état de l'art un grand nombre de modèles de fiabilité, et plus précisément de modèles pouvant modéliser le vieillissement. Notre répartition des modèles en trois familles facilite leur classification et peut contribuer à éviter l'emploi de noms différents pour un même modèle, comme c'est le cas pour les modèles de Salvia-Bollinger généralisés qui sont en fait des lois de Pólya inverses. Nous avons démontré la complétude de la classe IFRA1, résultat classique en temps continu, qui, étonnamment, n'avait pas été démontré en temps discret.

La définition d'un estimateur empirique du taux de défaillance en temps discret est également novatrice. Son étude, à taille d'échantillon fixée et asymptotique, a permis de mettre en évidence ses qualités et défauts d'estimateur et aussi ses particularités par rapport à l'estimateur empirique de la fiabilité. On pourra penser à modifier la définition de cet estimateur empirique pour éviter le fait qu'il puisse s'annuler un grand nombre de fois pour des données d'un ordre de grandeur élevé, ce qui perturbe toute interprétation et estimation graphique.

Le test d'adéquation à la famille géométrique d'Anderson-Darling utilisant la transformation de Smirnov généralisée semble présenter le meilleur compromis entre puissance et robustesse (par rapport à l'ordre de grandeur des données et la taille d'échantillon). Seule la loi de Weibull de type I a été utilisée pour les simulations de l'étude de puissance. Il reste à poursuivre ce travail avec d'autres alternatives. Tous les tests utilisés pour tester l'hypothèse

géométrique n'ont pu être appliqués à la loi de Weibull de type I. Il serait intéressant d'avoir accès à de fortes puissances de calcul pour poursuivre cette étude. Les temps de calculs prohibitifs (pour les études de seuil et de puissance) étant dus à la méthode du bootstrap, on pourrait chercher à construire des tests n'utilisant pas cette méthode, tel que le test lisse, pour lequel la difficulté repose sur la construction de la base orthonormée rendue complexe à cause de l'expression des moments de la loi de Weibull de type I. Seuls les tests d'adéquation aux lois géométrique et Weibull de type I sont actuellement possibles, un thème de recherche futur pourrait être d'étudier la possibilité de tester les autres lois discrètes, notamment celles pouvant modéliser un vieillissement telles que les lois d'Engenberger-Pólya et logistique tronquée discrétisée.

Le premier objectif fixé en début de thèse concernant la modélisation des durées de vie en temps discret est atteint. Le second objectif visait à donner une procédure d'analyse du REX en temps discret. Cet objectif est en grande partie atteint puisque seules manquent les intervalles de confiance du taux de défaillance et de la fiabilité dans le cas où le modèle retenu est la loi de Weibull de type I. La construction de ces intervalles de confiance n'a pu être abordée dans cette thèse par manque de temps. L'intérêt de ce thème de recherche est pratique, mais également théorique au vu de la littérature abondante concernant la loi de Weibull en temps continu.

Du point de vue industriel, cette thèse a permis de faire connaître les concepts de la fiabilité en temps discret aux ingénieurs du service de sûreté de fonctionnement de Schneider Electric. Les concepteurs de produits nouveaux ont été également intéressés d'approfondir la connaissance en terme de fiabilité de leurs produits, ne serait-ce que pour leur permettre de valider les objectifs de fiabilité des normes, mais aussi pour évaluer la performance à la sollicitation des nouveaux appareils. Les services technico-commerciaux ont pu s'aider de ces travaux pour juger plus précisément du vieillissement à la sollicitation et ainsi estimer les risques associés à la vente de services en rapport avec la durée de vie des appareils (contrats de maintenance, d'assurances,...). Le logiciel FEDDRE, développé sous MATLAB par Schneider Electric, intègre la plupart des fonctionnalités nécessaires à l'analyse de données discrètes présentées dans cette thèse.

Cette thèse constitue pour Schneider Electric une première étape de la démarche engagée pour approfondir la connaissance de ses produits. La prise en compte du vieillissement à la sollicitation est indubitablement un progrès par rapport aux normes actuelles, mais la vie opérationnelle des appareils est encore plus complexe. En effet, au vieillissement à la sollicitation des appareils, il faut ajouter une composante continue due au vieillissement dans le temps. Par ailleurs, les conditions d'exploitation peuvent être relativement éloignées des conditions nominales spécifiées par le constructeur et influencer fortement sur la durée de vie de l'appareil selon le stress subi (température, humidité, tension,...). La présence de stress n'apparaît plus seulement dans l'utilisation opérationnelle des appareils, mais également dans les essais en interne où le niveau élevé du stress permet de diminuer la durée de ceux-ci. À

notre connaissance, il n'existe pas de travaux de recherche sur la prise en compte simultanée du vieillissement à la sollicitation et du vieillissement temporel. De même, il n'existe pas non plus de modèle de vieillissement à la sollicitation prenant en compte le stress subi par l'appareil, alors que pour des appareils à durée de vie continue subissant des stress, la recherche, autant en industrie que dans les laboratoires de recherche est très développée.

L'obtention de données peut s'avérer être la plus grande difficulté et conditionne en réalité le degré de complexité de la modélisation. C'est un peu le cas du sujet de cette thèse puisque les seules données disponibles étaient des nombres de sollicitation jusqu'à l'apparition de la première défaillance. Nous avons alors considéré que les appareils étaient non réparables. En réalité, les appareils sont réparables, mais le Retour d'Expérience n'est pas encore suffisamment important pour pouvoir utiliser des modèles de fiabilité pour des systèmes réparables. Pour compenser le manque de données, on peut avoir recours aux techniques bayésiennes qui prennent en compte l'information a priori disponible. Le coût des essais conduit aussi à avoir peu de données disponibles pour un même type d'appareil étudié (souvent moins de 10 observations).

Le coût des essais est souvent la cause de la censure des durées de vie des appareils en test, censure qui est quasiment omniprésente dans le Retour d'Expérience. Les données à traiter sont donc de plus en plus souvent censurées et nécessitent des techniques d'analyse statistiques spécifiques. En temps discret, toute la partie statistique traitée dans cette thèse pour des données non censurées, reste (en grande partie) à être étudiée dans le cas censuré. Toutefois, les estimateurs non paramétriques des principales grandeurs de fiabilité existent dans le cas censuré. On regrette alors que la statistique non paramétrique soit peu utilisée dans la pratique industrielle alors qu'elle permet souvent de donner simplement une réponse valable au problème posé, même pour des échantillons de taille faible.

Bibliographie

- [1] ABRAMOVITZ, M. ET STEGUN, I.A. *Handbook of Mathematical Functions*. National Bureau of Standards, 9th edition, 1970.
- [2] ALI KHAN M.S., KHALIQUE A. ET ABOUAMMOH A.M. *On estimating parameters in a discrete Weibull distribution*. IEEE Transactions on Reliability, Vol. 38, No. 3, pp 348–350, August 1989.
- [3] ALTAVELA, M.M. ET WOOD C.L. *Large sample results for Kolmogorov-Smirnov statistics for discrete distributions*. Biometrika, Vol. 65, pp 235-239, 1978.
- [4] BACHA, M., CELEUX G. EDWIGE I. LANNOY A. ET VASSEUR D.
Estimation de modèles de durées de vie fortement censurées. Collection de la Direction des Études et Recherches d'Électricité de France, éditions Eyrolles, 1998.
- [5] BAILLARGEON G. *Méthodes statistiques de l'ingénieur*. Volume 1, Trois-Rivières, Les Éditions SMG, 1990.
- [6] BAIN L.J. ET ENGELHARDT M. *Statistical Analysis of Reliability and Life-testing Models, Theory and Methods, Second Edition*. Marcel Dekker, 1991.
- [7] BARINGHAUS, L. ET HENZE, N. *A goodness of fit test for the Poisson distribution based on the empirical generating function*. Statistics and Probability letters, Vol. 13, pp 269-274, 1992.
- [8] BARLOW R.E. MARSHALL A.W. ET PROSCHAN F. *Properties of probability distributions with monotone hazard rate*. Annals of Statistics, Vol. 34, pp 375-389, 1963.
- [9] BARLOW R.E. ET PROSCHAN F. *Mathematical Theory of Reliability*. John Wiley and Sons, Inc, New York, 1965.
- [10] BARLOW R.E. ET PROSCHAN F. *Statistical Theory of Reliability and Life Testing: Probability Models*. Second edition, Silver Spring, MD: To begin with, 1975.
- [11] BICKEL P.J. ET KRIEGER A.M. *Confidence Bands for a Distribution Function Using the Bootstrap*. Journal of the American Statistical Association, Vol. 53, No. 405, pp 95–100, 1989.
- [12] BILLINGSLEY P. *Convergence of probability measures*. John Wiley and Sons, Inc, New York, 1968.

- [13] BRACQUEMOND C. *Modélisation du vieillissement à la sollicitation de matériels électromécaniques*. Mémoire de DEA, Laboratoire de Modélisation et Calcul, Grenoble, 1998.
- [14] BRESLOW N.E. ET CROWLEY J. *A large sample study of the life table and product limit estimates under random censorship*. The Annals of Statistics, No. 2, pp 434–453, 1974.
- [15] CHOULAKIAN, V. LOCKART, R.A. ET STEPHENS, M.A. *Cramer-Von-Mises tests for discrete distributions*. Canadian Journal of Statistics, Vol. 22, pp 125-137, 1994.
- [16] CLAROTTI A., LANNOY A. ET PROCACCIA H. *Probability risk analysis of ageing components which fail on demand-a bayesian model: Application to maintenance optimization of diesel engine linings. Ageing of Materials and Methods for the Assesment of Lifetimes of Engineering Plant*. In Penny, editor, Balkema, Rotterdam, pp 85-93, 1997.
- [17] CLEMANS K.G. *Confidence limits in the case of the geometric distribution*. Biometrika, Vol. 46, pp 260-264, 1959.
- [18] COCOZZA-THIVENT C. *Processus stochastiques et fiabilité des systèmes*. Mathematiques and Applications (Paris). 28. Berlin : Springer.
- [19] CONOVER, W.J. *A Kolmogorov Goodness-of-fit Test for Discontinuous Distributions*. Journal of the American Statistical Association, Vol. 67, No. 339, p 591-596, 1972.
- [20] COX D.R. *Regression models and life tables (with discussion)*. Journal of the Royal Statistical Society, Serie B, 34, pp 187-220, 1972
- [21] COX D.R. ET OAKES D. *Analysis of survival data*. London : Chapman and Hall, 1984.
- [22] COX, D.R. ET HINKLEY, D.V. *Theoretical statistics*. London : Chapman and Hall, 1974.
- [23] D'AGOSTINO, R.B. ET STEPHENS, M.A. *Goodness of Fit Techniques*. Marcel Dekker, INC, 1986.
- [24] DAVID F.N. ET JOHNSON N. L. *The probability integral transformation when parameters are estimated from the sample*. Biometrika, Vol. 35, pp 182-190, 1948.
- [25] DURBIN J. *Weak convergence of the sample distribution function when parameters are estimated*. Annals of Statistics, Vol. 1, No. 2, pp 279–290, 1973.
- [26] EBRAHIMI N. *Classes of discrete decreasing and increasing mean-residual-life distributions*. IEEE Transactions on Relability, Vol. 35, No. 4, pp 403–405, October 1986.
- [27] EFRON B. *The two sample problem with censored data*. Proceedings Fifth Berkeley Symposium, 4, pp 831–853, 1967.
- [28] EGGENBERGER, F. ET POLYA, G. *Über die Statistik verketter Vorgänge*. Zeitschrift für angewandte Mathematik und Mechanik, 3, pp 279-289, 1923.
- [29] GHAI G.L. ET MI J. *Mean Residual Life and Its Association With Failure Rate*. IEEE Transactions on Reliability, Vol. 48, No. 3, pp 262–266, September 1999.
- [30] GREENWOOD, P.E. ET NIKULIN, M.S. *A guide to Chi-Squared Testing*. Wiley Series in Probability and Statistics, 1996. .

- [31] GUESS F.M. ET PARK D.H. *Modeling discrete bathtub and upside-down bathtub mean residual-life functions*. IEEE Transactions on Reliability, Vol. 37, No.5, pp 545-549, 1988.
- [32] GUPTA P.L., GUPTA R.C. ET TRIPATHI R.C. *On the monotonic properties of the discrete failure rates*. Journal of Statistical Planning and Inference, No. 65, pp 255-268, 1997.
- [33] GUTTMAN I. *A note on a series solution of a problem in estimation*. Biometrika, Vol. 45, pp 565-567, 1958.
- [34] HENZE, N. *Empirical distribution function goodness of fit tests for discrete models*. Canadian Journal of Statistics, Vol. 24, No. 1, pp 81-93, 1996.
- [35] HERD G.R. *Estimation of Reliability from Incomplete Data*. Proceedings of the 6th National Symposium on Reliability and Quality Control, IEEE, 345 East 47 St., New York, NY 10017, pp 202-217, 1960.
- [36] HOCINE M.E. *Tests basés sur les processus empiriques pour les lois discontinues*. Actes des XXIXèmes Journées de la Statistique, Carcassonne, pp 454-455, 1997.
- [37] HOEYLAND A. ET RAUSAND M. *System reliability theory. Models and statistical methods*. Wiley Series in Probability and Mathematical Statistics, 1994.
- [38] HORN, S.D. *Goodness-of-Fit for discrete data*. Biometrics, Vol. 33, pp 237-248, 1977.
- [39] IRWIN, J.O. *The generalized Waring distribution*. Journal of the Royal Statistical Society, Series B, No. 138, pp 18-31 (partie I); pp 204-227 (partie II); pp 374-384 (partie III), 1975.
- [40] JOHNSON L.G. *The statistical treatment of fatigue experiments*. Elsevier, New York, 1964.
- [41] JOHNSON N.L. ET KOTZ S. *Urn models and their application: an approach to modern discrete probability theory*. Wiley series in probability and mathematical statistics, 1977.
- [42] JOHNSON N.L., KOTZ S. ET KEMP, A. W. *Univariate discrete distributions*. Wiley series in probability and mathematical statistics, second edition, 1992.
- [43] KAA NICHE M. ET KANOUN K. *The discrete time hyperexponential model for software reliability growth evaluation*. Proceedings 3rd International Symposium on Software Reliability Engineering, North Carolina, March 1992.
- [44] KALBFLEISCH J.D. ET PRENTICE R.L. *The statistical analysis of failure time data*. Wiley Series in Probability and Mathematical Statistics, 1980.
- [45] KAPLAN E.L. ET MEIER P. *Nonparametric estimation from incomplete observations*. Journal of the American Statistical Association, 53, pp 457-481, 1958.
- [46] KAUFMANN A., GROUCHKO D. ET CRUON R. *Modèles mathématiques pour l'étude de la fiabilité des systèmes*. Masson, 1974.
- [47] KNOPP K. *Infinite Sequences and Series*. Dover, New York, 1956.
- [48] KOCHERLAKOTA, S. ET KOCHERLAKOTA, K. *Goodness-of-fit tests for discrete distributions*. Communications in Statistics, Theory Methods, Vol. 15, pp 815-829, 1986.

- [49] LANDBERG N.A., LEON R.V., LYNCH J. ET PROSCHAN F. *Extreme points of the class of discrete decreasing failure rate life distributions*. Mathematics of Operations Research, Vol. 5, No. 1, pp 35–42, Février 1980.
- [50] LAWLESS J.F. *Statistical models and methods for lifetime data*. John Wiley and Sons, Inc, New York, 1982.
- [51] LEE C.C. *Chi-squared tests for and against an order restriction on multinomial parameters*. Journal of the American Statistical Association, 82, pp 611–618, 1987.
- [52] LE BRETON A. ET MARTIN S. *Un modèle pour l'étude des séquences climatologiques*. Séminaire de l'équipe S.M.S, Université Grenoble I, 1979.
- [53] MANN, N. R., SCHEUER, E. M. ET FERTIG, K. W. *A new goodness-of-fit test for the two-parameter Weibull or extreme-value distribution with unknown parameters*. Communications in Statistics, 2, pp 383-400, 1973.
- [54] MEEKER W.Q. ET ESCOBAR L.A. *Statistical Methods for Reliability Data*. Wiley Series in Probability and Mathematical Statistics, 1998.
- [55] MEIER P. *Estimation of a distribution function from incomplete observations*. Perspectives in Probability and Statistics, J. Gani, Ed. Sheffield, ENGLAND, Applied Probability Trust, 1975.
- [56] MI J. *Discrete Bathtub Failure Rate and Upside-Down Bathtub Mean Residual Life*. Naval Research Logistics, Vol. 40, pp 361–371, 1993.
- [57] NAIR V. N. *Confidence Bands For Survival Functions With Censored Data: A comparative Study*. Technometrics, Vol. 26, No 3, August 1984.
- [58] NAKAGAWA T. ET OSAKI S. *The discrete Weibull distribution*. IEEE Transactions on Reliability, Vol. 24, No. 5, pp 300–301, December 1975.
- [59] NEYMAN, J. *“Smooth” test for goodness of fit*. Skand. Aktuarie Tidskr., Vol. 20, pp 150–199, 1937.
- [60] NEYMAN, J. ET PEARSON E.S. *On the use and interpretation of certain test criteria for purposes of statistical inference*. Biometrika, Vol. 20, pp 175–240 et 263–294, 1928.
- [61] NIKULIN, M.S. *Gihman statistic and goodness-of-fit tests for grouped data*. C.R. Math. Rep. Acad. Sci. Canada, Vol. 14, No. 4, pp 151–156, 1992.
- [62] NIKULIN, M.S. *Chi-square test for normality*. International Vilnius Conference on Probability Theory and Mathematical Statistics, Vol. 2, pp 119-122, 1973.
- [63] NIKULIN, M.S. *Chi-square test for continuous distributions with shift and scale parameters*. Theory of Probability and its Applications, Vol. 18, pp 559–568, 1973.
- [64] NIKULIN, M.S. *On a chi-square test for continuous distributions*. Theory of Probability and its Applications, Vol. 18, pp 638–639, 1973.
- [65] NIKULIN, M.S. *On a quantile test*. Theory of Probability and its Applications, Vol. 19, pp 410–413, 1974.

- [66] PADGETT W.J. ET SPURRIER J.D. *Discrete failure models*. IEEE Transactions on Reliability, Vol. 34, No. 3, pp 253–256, August 1985.
- [67] PATEL J.K., KAPADIA C.H. ET OWEN D.B. *Handbook of Statistical Distributions*, volume 20 Statistics: textbooks and Monographs. Marcel Dekker, Inc, 1976.
- [68] PETERSON A.V. *Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions*. Journal of the American Statistical Association, 72, pp 854–858, 1977.
- [69] PETTITT, A.N. ET STEPHENS, M.A. *The Kolmogorov-Smirnov goodness-of-fit statistics with discrete and grouped data*. Technometrics, Vol. 19, pp 205-210, 1977.
- [70] RANGLES R.H. ET WOLFE D.A., *Introduction to the Theory of Nonparametrics Statistics*. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, 1979.
- [71] RAO, C.R. *Tests of significance in multivariate analysis*. Biometrika, Vol. 35, pp 58–79, 1948.
- [72] RAO G.V. *A test for the fitting of some discrete distributions*. Publication I.S.U.P., Vol. XX, Fascicule 3-4, p 121-128, 1971 (paru en 1974).
- [73] RAO, K.C. ET ROBSON D.S. *A chi-square statistic for goodness-of-fit tests within the exponential family*. Communications in Statistics, Theory Methods, Vol. 3, pp 1139-1153, 1974.
- [74] RAYNER, J.C.W. ET BEST, D.J. *Goodness-of-fit for grouped data using components of Pearson's χ^2* . Computational Statistics and Data Analysis, Vol. 5, pp 53-57, 1987.
- [75] RAYNER, J.C.W. ET BEST, D.J. *Smooth Tests of Goodness of Fit*. Oxford University Press, 1989.
- [76] ROBERT C. *L'analyse statistique bayésienne*. Collection Economica, 1992.
- [77] ROY D. ET GUPTA R.P. *Characterizations and model selections through reliability measures in the discrete case*. Statistics & Probability Letters, Vol. 43, pp 197–206, 1999.
- [78] RUEDA, R., PÉREZ-ABREU, V. ET O'REILLY, F. *Goodness-of-fit test for the Poisson distribution based on the probability generating function*. Communications in Statistics, Theory Methods, Vol. 20, pp 3093-3110, 1991.
- [79] SALVIA A.A. *Some results on Discrete Mean Residual Life*. IEEE Transactions on Reliability, Vol. 35, No. 3, pp 359–361, September 1996.
- [80] SALVIA A.A. ET BOLLINGER R.C. *On discrete hazard functions*. IEEE Transactions on Reliability, Vol. 31, No. 5, pp 458–459, December 1982.
- [81] SCHMID P. *On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions* Annals of Mathematical Statistics, Vol. 29, pp 1011–1027. 1958.
- [82] SCHOENBERG I.J. *On the zeros of the generating functions of multiply positive sequences and functions* Annals of Mathematical Statistics, Vol. 62, pp 447-471. 1955.

- [83] SERFLING R.J. *Approximations theorems of mathematical statistics*. Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, 1980.
- [84] SHAKED M., SHANTHIKUMAR J.G. ET VALDEZ-TORRES J.B. *Discrete hazard rate functions*. Computers and Operations Research, Vol. 22, No. 4, pp 391–402, 1995.
- [85] SHORACK G.R. ET J.A. WELLNER J.A. *Empirical processes with applications in statistics*. Wiley Series in Probability and Mathematical Statistics, 1986.
- [86] SOLER J.L. *Croissance de fiabilité des versions d'un logiciel*. Revue de Statistique Appliquée, Vol. XLIV, No. 1, pp 5–20, 1996.
- [87] SPINELLI, J.J. ET STEPHENS, M.A. *Cramer-Von-Mises tests of fit for the Poisson distribution*. Canadian Journal of Statistics, Vol. 25, pp 257-268, 1997.
- [88] STEIN W.E. ET DATTERO R. *A new discrete Weibull distribution*. IEEE Transactions on Reliability, Vol. 33, No. 2, pp 196–197, June 1984.
- [89] STURGES . *The choice of a class interval*. Journal of the American Statistical Association, Vol. 21, pp 65–66, 1926.
- [90] TANG L.C., LU Y., CHEW E.P. *Mean Residual Life of Lifetime Distributions*. IEEE Transactions on Reliability, Vol. 48, No. 1, pp 73–78, March 1999.
- [91] VIT P. *Testing for homogeneity; the geometric distribution*. Biometrika, Vol. 61, pp 565–568, 1974.
- [92] WALD, A. *Tests of statistical hypotheses concerning several parameters when the number of observations is large*. Transactions of the American Mathematical Society, Vol. 54, pp 426–482, 1943.
- [93] XEKALAKI E. *Chance mechanisms for the univariate generalized Waring distribution and related characterizations*. Statistical distributions in scientific work, Vol. 4, Proc. NATO Adv. Study Inst., Trieste/Italy 1980, pp 157-171 (paru en 1981).
- [94] XEKALAKI E. *Hazard functions and life distributions in discrete time*. Communications in Statistics, Theory Methods, Vol. 12, No. 21, pp 2503–2509 , 1983.
- [95] XIE M. ET GOTH T.N. *SPC of a nearzero defect process subject to random shock*. Quality and Reliability Engineering International, Vol. 9, pp 89-93, 1993.
- [96] XIE M., GAUDOIN O. ET BRACQUEMOND C. *Redefining Failure Rate Function for Discrete Distributions*. Rapport de Recherche RR 1029,LMC/IMAG, Mai 2000.
- [97] ZACKS S. *Estimating the shift to wear-out of systems having exponential-Weibull life*. Operations Research, No. 32, pp 741–749, 1984.
- [98] NORME CEI 60605-4 : *Essai de fiabilité des équipements. Quatrième partie: Méthode de calcul des estimateurs ponctuels et des limites de confiance résultant d'essais de détermination de la fiabilité d'équipements*. Janvier 1986.
- [99] NORME CEI 60605-6 : *Essais de fiabilité des équipements - Partie 6: Tests de validité des hypothèses du taux de défaillance constant ou de l'intensité de défaillance constante*. Avril 1997.

- [100] NORME CEI 61649 : *Procédures pour le tests d'adéquation, les intervalles de confiance et les limites inférieures de confiance pour les données suivant la distribution de Weibull*. Mai 1997.
- [101] NORME CEI 60947-1 : *Appareillage à basse tension - Partie 1 : Règles générales*. Avril 1997.
- [102] NORME CEI 60947-2 : *Appareillage à basse tension - Partie 2 : Disjoncteurs*. Décembre 1995.
- [103] NORME CEI 60947-3 : *Appareillage basse tension - Partie 3 : Interrupteurs, sectionneurs, interrupteurs-sectionneurs et combinés-fusibles*. Commission Electronique Internationale, Juillet 1999.
- [104] NORME ANSI C37-06 : *AC High-Voltage Circuit Breakers Rated on a Symmetrical Current Basis Preferred Ratings and Related Required Capabilities*. American National Standard Institute, 1997.

Résumé

Dans les études de fiabilité, il est possible que les durées de vie s'expriment par des valeurs entières. C'est le cas par exemple pour des appareils fonctionnant à la sollicitation. Le but de cette thèse est de revisiter les concepts classiques de la fiabilité des systèmes non réparables, quand on suppose que le temps est discret. Les grandeurs usuelles de la fiabilité en temps discret sont définies, ainsi que les principales notions de vieillissement. Nous passons ensuite en revue les principaux modèles de fiabilité que nous avons classés en trois familles à la suite de notre étude bibliographique. Après avoir mis en évidence qu'un certain nombre de différences entre les notions de fiabilité usuelles en temps continu et leurs équivalents en temps discret sont dues à la définition du taux de défaillance en temps discret, nous proposons une nouvelle définition qui résout les problèmes rencontrés. L'estimation non paramétrique des fonctions pouvant caractériser le vieillissement est traitée et nous donnons des résultats d'estimation ponctuelle, par intervalle et bande de confiance pour le taux de défaillance (usuel). L'estimation paramétrique des modèles de fiabilité est également abordée, avec une section importante consacrée à la loi géométrique dont les estimateurs sont explicites, contrairement à la plupart des autres lois discrètes. Un état de l'art des tests d'adéquation statistiques aux lois discrètes est présenté. Nous proposons un test basé sur la transformation de Smirnov généralisée dont la puissance est étudiée pour tester l'adéquation aux lois géométrique et Weibull de type I.

Mots clés : Fiabilité, temps discret, systèmes non réparables, modèles de durées de vie, vieillissement, taux de défaillance, test d'adéquation, transformation de Smirnov généralisée.

Stochastic modelling of ageing in discrete time

Abstract

In this thesis, we propose a complete review of non repairable systems reliability concepts and models when systems lifetime are discrete random variables. First, we give the basic concepts of reliability in discrete time, including ageing notions. We show that the differences between continuous and discrete reliability ageing concepts are due to the definition of discrete failure rate. We propose a new definition of failure rate in discrete time which solves the problems. A bibliographical study of discrete reliability models highlights a possible classification of them into three categories. A complete statistical analysis is presented, from a non parametric study of reliability functions involved in the characterization of ageing, to parameters estimation and goodness-of-fit tests for discrete models. We propose a new goodness-of-fit test based on the generalized Smirnov transformation whose power is studied in the case of fitting the geometric and type I discrete Weibull distributions.

Keywords : Reliability, discrete time, ageing, failure rate, goodness-of-fit test.