

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITÉ MOULOUD MAMMARI, TIZI-OUZOU



FACULTÉ DES SCIENCES
DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du
DIPLÔME DE MAGISTER (ÉCOLE DOCTORALE)
EN MATHÉMATIQUES

Option : Statistique

Présenté par : **Fahima Bouzekria**

Titre

**Modélisation bayésienne des points de changements de l'historique des
pannes dans les portes du train**

Devant le jury d'examen composé de :

Fazia BEDOUHENE	Professeur	UMMTO	Présidente
Mohamed BOUALEM	Maître de Conférences A	UAMB	Rapporteur
Karima NOUALI	Maître de Conférences A	UMMTO	Examinatrice
Karim ABBAS	Maître de Conférences A	UAMB	Examinateur
Louiza BOURAINE	Maître de Conférences B	UAMB	Invitée

Année universitaire 2013/2014

Résumé

Les défaillances des systèmes réparables sont souvent décrites au moyen de processus de Poisson non homogène, identifiées par leur fonction d'intensité et la valeur de leur moyenne. L'intervention sur les systèmes sont susceptibles de modifier leur fiabilité, et les variations des fonctions d'intensité et la valeur de la moyenne sont donc amenés. Nous considérons différents scénarios dans lesquels les interventions prennent place et proposent des modèles décrivant chacun d'eux. L'analyse bayésienne, est illustrée avec des applications de simulation sur les données réelles, en s'appuyant sur les méthodes de Monte Carlo par Chaîne de Markov (MCMC).

Mots-Clés : Inference Bayésien, processus de la lois de puissance, méthodes MCMC, points de changements.

Abstract

Failures in repairable systems are often described by means of non-homogeneous Poisson processes, identified by their intensity and mean value functions. Intervention on the systems are likely to modify their reliability, and changes in intensities and mean value functions are therefore induced. We consider different scenarios in which interventions take places and propose models describing each of them. Bayesian analyses, relying on Markov-chain Monte Carlo methods, are illustrated along with applications to simulated and real, widely-known, data.

Key words : Bayesian inference, power law process, MCMC methods, change points.

Remerciements

Tout d'abord, je tiens à remercier le Bon Dieu, sans lui ce mémoire n'aurait pu exister.

Je voudrais en premier lieu exprimer toute ma gratitude au Docteur BOUALEM Mohamed (Université A. Mira Bejaia), pour sa disponibilité, son soutien, ses encouragements, au cours de cette année, mais aussi pour avoir fait preuve, à plusieurs reprises, de sa confiance en mon travail de recherche.

Je remercie vivement, Professeur BEDOUHENE Fazia (Université Mouloud Mammeri Tizi -Ouzou) pour l'honneur qu'elle me fait en acceptant d'être présidente du jury.

Mes remerciements chaleureux s'adressent également à Docteur ABBAS Karim (Université A. Mira Bejaia), Docteur NOUALI Karima (Université Mouloud Mammeri Tizi -Ouzou), Docteur Bouraine Louiza (Université A. Mira Bejaia) pour avoir accepté d'examiner ce travail.

Finalement, j'aimerais remercier mon mari YESSINE, et ma grande famille, en particulier ma mère, mon père, et tous ceux qui m'ont encouragé pendant ces années d'études.

Table des matières

Introduction générale	4
1 Analyse statistique bayésienne	8
1.1 Introduction	8
1.2 Approche statistique classique	10
1.3 Approche bayésienne	11
1.4 Lien entre les deux approches	11
1.5 Notations	12
1.6 Fondements de la statistique bayésienne	13
1.6.1 Le paradigme bayésien	14
1.6.2 Distributions a priori	16
1.7 Inférence Bayésienne	20
1.7.1 Prédiction	20
1.7.2 L'estimation ponctuelle	21
1.7.3 Estimation paramétrique Bayésienne	29
1.7.4 Analyse Bayésienne non paramétrique	31
1.7.5 Estimateur de Bayes de la fonction de répartition	35
1.7.6 Tests et régions de confiance	35
1.8 Les méthodes de simulation de Monte Carlo par Chaîne de Markov(MCMC)	39
1.8.1 Les chaînes de Markov	39
1.8.2 Méthodes de Monte Carlo par chaîne de Markov	42

2	Processus de Poisson	55
2.1	Introduction	55
2.2	Définitions et description du processus de Poisson	56
2.3	Caractérisation d'un processus de Poisson par ses temps d'arrivée	58
2.4	Propriétés supplémentaires	59
2.4.1	Décomposition, superposition	59
2.5	Processus de Poisson et loi binomiale	59
2.6	Processus de Poisson et loi uniforme	60
2.7	Processus de Poisson composés	60
2.8	Processus de Poisson homogène(HPP)	62
2.9	Processus de Poisson non-homogène(NHPP)	62
2.9.1	Propriétés de processus de Poisson non homogène	65
2.9.2	Classes des PPNH	66
3	Fiabilité	71
3.0.3	Rappel sur la théorie de la fiabilité	72
3.1	Fiabilité des systèmes réparables :	74
3.1.1	Systèmes réparables	74
3.2	Points de changement dans les NHPP	77
3.2.1	Défaillances dans les métros : Fiabilité de la porte du train	78
3.2.2	Les données	80
3.2.3	Analyse exploratoire des données	81
3.2.4	Modèle pour l'apparition de défaillances	86
3.2.5	Estimation bayésienne non paramétrique	87
3.3	Application	90
3.3.1	Estimation bayésienne des paramètres	90
3.3.2	Estimation bayésienne des paramètres par les méthodes MCMC	91
3.3.3	Discussion des résultats	92

Introduction générale

L'objectif principal de la statistique est de faire, à partir d'observations d'un phénomène aléatoire, une inférence au sujet de la loi générant ces observations en vue d'analyser le phénomène ou de prévoir un événement futur.

Pour réduire la complexité du phénomène étudié, deux approches statistiques sont utilisées dans la littérature : modélisation non paramétrique et modélisation paramétrique. Dans la modélisation non paramétrique, on considère souvent que l'inférence statistique doit prendre en considération la complexité autant que possible et donc on cherche à estimer la distribution de l'ensemble du phénomène, mettant en œuvre l'estimation des fonctionnelles (densité, droite de régression, etc.). Par opposition, le modèle paramétrique est utilisé lorsque la famille de départ est paramétrée par un vecteur θ de dimension finie ; dans ce cas sa résolution se résume à l'identification de ce paramètre. D'une manière générale, une fois le modèle est construit, on cherche à établir une inférence sur θ , c'est-à-dire à utiliser les observations x afin d'évaluer θ , en vue d'une décision liée à ce paramètre.

Notons que la démarche statistique est fondée sur une démarche d'inversion qui est de relier des "*effets*" (observations) aux "*causes*" (paramètres). Pour réaliser ce lien *effets-causes* plusieurs approches sont proposées dans la littérature on cite : l'approche fréquentiste et l'approche bayésienne. Pour certains auteurs, ces deux approches (fréquentiste ou probabiliste, bayésienne) s'opposent alors que d'autres voient en elles une certaine complémentarité. Cette opposition réside en fait dans la démarche même de ces deux approches.

Une autre technique d'une grande importance numérique dans l'analyse statistique est l'utilisation des méthodes de Monte Carlo par Chaîne de Markov (MCMC). Ces méthodes ont ouvert la voie à une analyse statistique correcte des modèles complexes. Elles sont introduites par Metropolis et al. (1953) [40] pour les calculs numérique, puis généralisées par Hastings (1970) [56] à la simulation statistique. Il faut attendre le début des années 90 et l'apparition de l'article de Gelfand et Smith [4], pour que leurs utilisation se développe en statistique appliquée [25].

Les méthodes MCMC sont en généralement suffisantes pour simuler une densité complexe et de dimension élevée. Elles consistent à générer des échantillons distribués asymptotiquement suivant la loi d'intérêt (cible). Elles se divisent en deux catégories : l'algorithme de Metropolis-Hastings (MH) et de l'échantillonneur de Gibbs. Ce dernier est un cas particulier de l'algorithme de Metropolis-Hastings. Cet algorithme sert à échantillonner des valeurs d'une fonction multidimensionnelle avec un nombre fixe de variables. La densité conjointe n'est pas nécessairement connue; toutefois, toutes les distributions conditionnelles doivent être connues. C'est dans cette optique que cette méthode semble être une bonne méthode pour faire l'échantillonnage d'une distribution a posteriori dans l'approche bayésienne.

La théorie de la fiabilité a pour objectif d'étudier l'aptitude de dispositifs techniques (machines, équipements,...), à accomplir une fonction requise, dans des conditions données, durant un temps donné. Actuellement, c'est une discipline à part entière. Prévoir la fiabilité d'un système est essentielle pour des problèmes de sécurité (systèmes de freinage, systèmes nucléaires, systèmes informatiques...), la quasi-impossibilité de réparer certains matériels (satellites), les problèmes économiques (coûts des défaillances, gestion du personnel de maintenance, maintenance des stocks des pièces de rechange),... se qui rend la connaissance de la fiabilité des systèmes utilisés indispensable.

Les défaillances se produisent généralement de façon aléatoire, il est logique de faire appel aux calculs probabilistes pour l'étude d'un problème de fiabilité. Ainsi, nous définissons la fiabilité d'un dispositif comme étant sa probabilité de fonctionner correctement pendant une durée donnée, ou, ce qui revient au même, la probabilité qu'aucune défaillance ne se produise pendant cette durée.

Les systèmes réparables sont ces systèmes qui en cas de panne, peuvent être réparés, par exemple en remplaçant un composant, et les ramener à un fonctionnement normal. Dans certains cas, la fiabilité d'un système, après une réparation, revient au même état qu'avant la panne. En revanche, les réparations "parfaites" apportent la fiabilité à l'état auquel était le système au début de l'opération. Les échecs des anciens systèmes réparables sont souvent décrits au moyen de Processus de Poisson Non Homogènes (NHPP), alors que les processus de renouvellement décrivent généralement ces derniers systèmes.

Dans la littérature, un bon nombre de travaux traitent la fiabilité des systèmes. A titre d'exemple, on cite les travaux de Argiento et al. (2003) [25] qui examinent des données réelles représentant les défaillances du système d'ouverture de portes de 40 trains enregistrés entre le 6 Avril 1990 et le 31 Décembre 1998. Par ailleurs, Cavallo (1999)[17] et Cavallo & Ruggeri (2001)[18] ont proposé de modéliser ces dernières données par différents processus : processus de Poisson homogène, NHPP à un seul paramètre, le processus aux lois de puissance et le processus logarithmique de Poisson non homogène.

L'objectif de notre travail est l'étude de l'historique des pannes des portes des trains. Nous allons utiliser les méthodes MCMC pour estimer les paramètres du processus de Poisson non homogène modélisant l'historique des pannes en prenant en considération le point de changement de ce processus.

Le mémoire présenté comprend une introduction, deux chapitres et une conclusion. Le premier chapitre a un caractère introductif. Dans ce chapitre, on décrit en détail l'analyse

statistique bayésienne et les méthodes de simulation de Monte Carlo par Chaîne de Markov basée sur l'algorithme Metropolis-Hastings et l'échantillonneur de Gibbs. Dans le second chapitre, en premier lieu, on a synthétisé quelques résultats concernant les processus de Poisson (homogènes et non homogènes) et la fiabilité des systèmes réparables. En deuxième lieu, on a analysé les données de défaillance récoltées sur 40 rames de métro, qui ont été livrés à une entreprise de transport entre Novembre 1989 et Mars 1991 et tous ont été mis en service à partir du 6 Avril 1990 au 20 Juillet 1992. Dans ce sens vous avez donné une modélisation bayésienne des points de changements de l'historique des pannes dans les portes du train avec une estimation bayésienne des paramètres d'un modèle de processus de Poisson non-homogène.

Ce travail se termine par une conclusion générale et quelques perspectives.

Chapitre 1

Analyse statistique bayésienne

1.1 Introduction

Avant de mettre en place les éléments nécessaires à la construction d'une machine inférentielle bayésienne, nous considérons tout d'abord quels sont les points essentiels définissant la science statistique. Il nous apparaît qu'une définition concise est de mener, grâce à l'observation d'un phénomène aléatoire, une inférence soit donc une démarche déductive logique sur la distribution de probabilité à l'origine de ce phénomène, pour ainsi fournir une analyse (ou une description) d'un phénomène passé, ou bien une prévision d'un phénomène à venir (et de même nature). Il va de soi que les étapes nécessaires au recueil des données comme la construction de plans de sondage ou d'expérience fait aussi partie du domaine de la statistique et que l'approche bayésienne peut également apporter un éclairage nouveau sur ces opérations.

L'approche statistique est par essence formelle (ou mathématiquement structurée) parce qu'elle repose sur une formalisation poussée de la réalité objective. En particulier, nous insistons ici sur l'interprétation décisionnelle de l'inférence statistique parce que, tout d'abord, les analyses et prédictions mentionnées ci-dessus sont la plupart du temps motivées par un seul objectif (comme la construction d'un portefeuille boursier ou la validation d'un

contrôle de qualité) ayant des conséquences (quantitativement et souvent monétairement) mesurables (résultats financiers, taux de retour des pièces défectives) et que d'autre part, ce degré supplémentaire de formalisme permet en retour la construction d'une machine automatique d'inférence bayésienne. Notons par ailleurs, que la statistique doit être considérée comme l'interprétation du phénomène observé, plutôt que comme son explication. En effet, l'inférence statistique est précédé d'une modélisation probabiliste et celle-ci implique nécessairement une étape de formalisation réductrice : sans cette base probabiliste, aucune conclusion utile ne pourrait être obtenue. On pourra peut-être regretter cette apposition d'un modèle probabiliste sur un phénomène inexpliqué, comme il est possible que le phénomène observé soit entièrement déterministe ou tout du moins sans rapport direct avec le modèle pré-supposé. Cependant, cette critique de la modélisation probabiliste n'a guère de consistance si nous considérons la statistique sous l'angle de l'interprétation, évoquée ci-dessus. Ces modèles probabilistes formels permettent en effet d'incorporer simultanément les informations disponibles sur le phénomène (facteurs déterminants, fréquence, amplitude, etc.) et les incertitudes inhérentes à ces informations. Ils autorisent donc un discours qualitatif sur le problème en fournissant, à travers la théorie des probabilités, un véritable calcul de l'incertain qui permet de dépasser le stade descriptif des modèles déterministes.

Évidemment la modélisation probabiliste n'a de sens pour l'analyse que si elle fournit une représentation suffisamment proche du phénomène observé. Dans de nombreux cas, la formalisation statistique est bien réductrice au sens où elle n'est qu'une approximation de la réalité, perdant une partie de la richesse de cette réalité mais gagnant en efficacité. Face à ce possible volant de réduction dans la complexité du phénomène observé, deux approches statistiques s'opposent. La première approche suppose que l'inférence statistique doit prendre en compte cette complexité autant que possible et elle cherche donc à estimer la distribution sous-jacente du phénomène sous des hypothèses minimales, en ayant recours en général à l'estimation fonctionnelle (densité, fonction de régression, etc.). Cette approche est dite non paramétrique. Par opposition, l'approche paramétrique représente la distribution des observations par une fonction de densité $f(x|\theta)$, où seul le paramètre θ

(de dimension finie) est inconnu. Les deux approches ont leurs avantages respectifs et, bien que dans cet chapitre, nous ne considérons que l'approche paramétrique pour des raisons pratiques, il existe également des résolutions bayésiennes du problème de l'inférence non paramétrique (Dey et al. 1998 [16]).

1.2 Approche statistique classique

Dans l'approche fréquentiste, les données observées sont considérées comme des observations de variables aléatoires. Elles servent alors à faire porter l'inférence sur les paramètres θ ayant dirigé leur mécanisme de génération. Autrement dit, l'information provenant des données observées est l'unique source d'information.

Dans cette approche, l'inversion est flagrante dans la notion de vraisemblance. En effet, on écrit

$$l(\theta|x) = f(x|\theta).$$

en considérant $l(\theta|x)$ comme une fonction de θ , on la normalise (quand cela est possible) pour en faire une fonction de densité sur Θ et on l'utilise dans l'estimation de θ . Par exemple, en estimation, on cherche la valeur $\hat{\theta}_{MV}$ qui maximise $l(\theta|x)$, c'est l'estimation au sens du maximum de vraisemblance :

$$\hat{\theta}_{MV} = \arg_{\theta \in \Theta} \max \{l(\theta|x)\}.$$

On utilise $l(\theta|x)$ comme si elle était une fonction de densité de probabilité conditionnelle aux observations x . Cette inversion est purement formelle alors que, dans l'approche bayésienne, comme nous le verrons un peu plus tard, cette inversion se fait d'une manière plus satisfaisante par la formule de Bayes.

1.3 Approche bayésienne

Dans la section précédente, nous avons considéré le cas où le paramètre inconnu θ est non aléatoire. Dans l'approche bayésienne, l'idée de base consiste à traiter le paramètre inconnu θ comme une variable aléatoire admettant une densité de probabilité $\pi(\theta)$ qui s'appelle densité a priori.

L'objectif est donc d'utiliser cette information supplémentaire. Sachant que l'information contenue dans les observations x est contenue dans $\pi(x|\theta)$ et l'information a priori sur θ dans $\pi(\theta)$, on peut utiliser la formule de Bayes pour combiner ces deux types d'informations en définissant la densité a posteriori par :

$$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\int \pi(x|\theta)\pi(\theta)d\theta},$$

qui contiendra toutes les informations sur θ .

On remarque que l'inversion de cause à effet est beaucoup plus naturelle. Elle se fait d'une manière cohérente, car l'état de connaissance a priori sur θ traduite par la densité a priori $\pi(\theta)$ est transformé, après les observations x , en état de connaissance a posteriori par la densité a posteriori $\pi(\theta|x)$.

On remarque que la densité a posteriori peut s'écrire aussi comme suite :

$$\pi(\theta|x) \propto l(\theta|x)\pi(\theta).$$

1.4 Lien entre les deux approches

L'aspect d'inversion de la statistique, il est tentant de considérer, sous réserve de l'intégrabilité, la fonction de vraisemblance $l(\theta|x)$ comme une loi de probabilité sur θ dont l'estimateur du maximum de vraisemblance serait le mode. En effet, celle-ci est équivalente

à la densité a posteriori $\pi(\theta|x)$ lorsqu'on choisit une loi a priori uniforme définie sur $[0,1]$ pour $\pi(\theta)$. Laplace [46] considérait que l'absence d'information a priori justifiait le choix de la loi uniforme. Cette position est défendable lorsqu'il s'agit d'un paramètre de position, conduit à des paradoxes et à des contradictions qui montrent clairement la nécessité de la théorie bayésienne, plus élaborée incluant les notions des lois a priori non informatives, des lois conjuguées et des lois a priori de références (voir Robert 1992[12]).

L'approche à privilégier pour l'inférence bayésienne est celle passant par la loi a posteriori. En effet, travaillant conditionnellement aux observations, cette approche suit d'une manière cohérente l'idée d'inversion des causes aux effets, tout en restant fidèle au principe de vraisemblance. En fait, la loi a posteriori représente l'actualisation de l'information a priori, $\pi(\theta)$, au vu de l'information contenue dans les observations x , au travers de la vraisemblance $l(\theta|x)$. Disposant ainsi d'une distribution de probabilité sur θ , le champ de l'inférence est beaucoup plus vaste que dans le cadre classique qui se contentait de $l(\theta|x)$. Nous avons calculé par exemple la moyenne, la médiane ou le mode.

A ce stade, on peut dire que l'approche statistique fréquentiste où toute l'inférence est basée sur la vraisemblance est un cas particulier de l'approche bayésienne avec une densité a priori uniforme définie sur $[0,1]$. En effet, si $\pi(\theta) = c$, la densité a posteriori $\pi(\theta|x) \propto l(\theta|x)$ et, d'après le principe de vraisemblance, toute inférence tirée de ces deux approches sera équivalente. Cependant, seule l'approche bayésienne permet d'introduire des informations complémentaires sur θ sous forme d'une densité a priori non uniforme.

Une fois la densité a posteriori calculée, il reste encore à savoir l'utiliser correctement.

1.5 Notations

Le formalisme fondamental d'une approche statistique est de supposer que les observations x_1, \dots, x_n , sur lesquelles l'analyse statistique se fonde, proviennent d'une loi de proba-

bilité paramétrée, ce qui se traduit par les hypothèses selon lesquelles x_1 a une distribution de densité $f_1(x|\theta_1)$ sur un espace mesurable comme \mathbb{R}^p et $x_i (2 \leq i \leq n)$ a, conditionnellement aux observations x_1, \dots, x_{i-1} , une distribution de densité $f_i(x_i|\theta_i, x_1, \dots, x_{i-1})$ sur le même espace mesurable. Dans ce cas, le paramètre θ_i est inconnu (et constitue un objet d'inférence), mais la fonction générique f_i est connue. Ce modèle peut être réécrit plus succinctement par

$$x \sim f(x|\theta') = f_1(x_1|\theta_1) \prod_{i=2}^n f_i(x_i|\theta_i, x_1, \dots, x_{i-1}), \quad (1.1)$$

où x est le vecteur des observations, $x = (x_1, \dots, x_n)^t$, et θ' l'ensemble des paramètres, $\theta' = (\theta_1, \dots, \theta_n)$, les composants étant éventuellement tous égaux. Cette représentation est unificatrice dans le sens où elle recouvre les cas d'une observation isolée, d'observations dépendantes, ou d'observations distribuées de façon indépendante et identiquement distribuées (iid), les x_1, \dots, x_n étant tous de même loi, de densité $f(x_1|\theta)$. Dans le dernier cas, $\theta' = \theta$ et

$$f(x|\theta') = \prod_{i=1}^n f(x_i|\theta). \quad (1.2)$$

Les densités $f(x|\theta)$ peuvent ainsi correspondre à des densités binomiales, de Poisson, normales ou Gamma, pour citer quelques exemples standard. Une simplification de notation adoptée dans la suite est que les densités des variables aléatoires continues (comme les variables normales) et discrètes (comme les variables de Poisson) sont représentées par les mêmes symboles, la mesure de référence étant fournie naturellement par le contexte. De plus, nous écrirons " x est distribué selon f " ou " $x \sim f$ " au lieu de " x est une observation de la distribution de densité f " par souci de concision.

1.6 Fondements de la statistique bayésienne

Le message que nous voulons communiquer dans ce rapide survol de la statistique bayésienne est on ne peut plus simple : il est possible, sans expertise préalable, de réaliser

une analyse bayésienne de tout problème statistique. En particulier, nous insistons sur le fait que ce qui est souvent considéré comme les deux difficultés majeures de l'approche bayésienne, à savoir le choix de la loi à priori et le calcul des procédures bayésiennes, qui peuvent être surmontées en suivant des formules simples.

1.6.1 Le paradigme bayésien

Étant donné un modèle paramétrique d'observation $x \sim f(x|\theta)$, ou $\theta \in \Theta$, un espace de dimension finie, l'analyse statistique bayésienne vise à exploiter le plus efficacement possible l'information apportée par x sur le paramètre θ , pour ensuite construire des procédures d'inférence sur θ . Bien que x ne soit qu'une réalisation (aléatoire) d'une loi gouvernée par θ , elle apporte une actualisation aux informations préalablement recueillies par l'expérimentateur. Pour des raisons diverses dont certaines apparaîtront dans le prochain paragraphe, l'information fournie par l'observation x est contenue dans la densité $f(x|\theta)$, que l'on représente classiquement sous la forme inversée de vraisemblance :

$$l(\theta'|x) = f(x|\theta'), \quad (1.3)$$

pour traduire qu'il s'agit d'une fonction de θ' , qui est inconnu, dépendant de la valeur observée x . L'inversion des rôles de x et de θ' par rapport à la modélisation probabiliste reflète le but premier de la statistique qui est de reconstruire (avec un certain degré de précision) le paramètre θ' au vu de la réalisation aléatoire x . C'est donc pourquoi elle est naturellement liée au théorème de Bayes qui formalise l'inversion des conditionnements dans les probabilités : Si A et E sont des événements tels que $p(E) \neq 0$, alors $p(E|A)$ et $p(A|E)$ alors ils sont reliés par :

$$\begin{aligned} P(A|E) &= \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)} \\ &= \frac{P(E|A)P(A)}{P(E)}. \end{aligned} \quad (1.4)$$

Une version continue de ce résultat permet d'inverser les densités conditionnelles, à savoir :

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}. \quad (1.5)$$

Le lien entre ces propriétés probabilistes et l'inférence bayésienne est que, dans le paradigme bayésien, le paramètre inconnu θ' n'est plus considéré comme inconnu et déterministe, mais comme une variable aléatoire. On considère ainsi que l'incertitude sur le paramètre θ' d'un modèle peut être décrite par une distribution de probabilité π sur θ' , appelée distribution a priori (par opposition à la distribution a posteriori qui inclut l'information contenue dans l'observation x), ce qui revient à supposer que θ est distribué suivant $\pi(\theta')$, $\theta \sim \pi(\theta')$, "avant" que x ne soit généré suivant $f(x|\theta')$, le conditionnement implicite dans cette notation prenant alors tout son sens. Sans vouloir nous engager dans un débat philosophique sur la nature du hasard, notons que le rôle central de la distribution à priori dans l'analyse statistique bayésienne ne réside pas dans le fait que le paramètre d'intérêt θ' puisse (ou ne puisse pas) être perçu comme étant distribué selon π , ou même comme étant une variable aléatoire, mais plutôt dans la démonstration que l'utilisation d'une distribution à priori et de l'appareillage probabiliste qui l'accompagne est la manière la plus efficace (au sens de nombreux critères) de résumer l'information disponible (ou le manque d'information) sur ce paramètre ainsi que l'incertitude résiduelle. Un point plus technique est que le seul moyen de construire une approche mathématiquement justifiée opérant conditionnellement aux observations, tout en restant dans un schéma probabiliste, est d'introduire une distribution correspondante pour les paramètres.

D'un point de vue pratique, le choix de la loi a priori est souvent perçu comme une difficulté. D'un point de vue pratique, le choix de la loi a priori est souvent perçu comme une difficulté majeure de l'approche bayésienne en ce que l'interprétation de l'information a priori disponible est rarement assez précise pour conduire à la détermination d'une seule

et unique loi. D'autre part, il est aisé de constater sur des exemples classiques que des choix contrastés de lois a priori conduisent à des inférences divergentes. Il existe néanmoins des lois calibrées en fonction de la distribution des observations, dites lois conjuguées, et des lois à faible contenu informatif, dites lois non-informatives, qui permettent d'évaluer l'influence d'une loi a priori donnée.

1.6.2 Distributions a priori

Le choix de la loi a priori représente une étape cruciale dans l'analyse statistique bayésienne, puisqu'elle influence directement le reste de l'inférence. L'information a priori, quand elle est disponible, n'est en général pas formulée en termes précis ou même statistiques : par exemple l'information a priori " θ appartient à l'intervalle $[-1, 1]$ avec 95% de chance" ne détermine en rien la forme de la loi a priori π qui peut être choisie indifféremment gaussienne, de Cauchy, uniforme, etc.

Lois conjuguées

Une famille F de lois de probabilité sur Θ est dite conjuguée pour la vraisemblance $f(x|\theta)$ si pour toute loi a priori $\pi \in F$, la loi a posteriori $\pi(\cdot|x)$ appartient également à F .

Les familles de lois conjuguées sont souvent considérées en premier lieu pour choisir la loi a priori car elles permettent d'effectuer les estimations sans avoir recours à des techniques complexes d'approximation numérique : si la famille conjuguée est une famille paramétrée $F = F_{\alpha \in A}$, la loi a posteriori appartenant à $F_{\alpha \in A}$, son calcul se résume à une mise à jour du paramètre $\alpha \in A$. En particulier l'accès aux quantités telles que l'espérance, la variance ou les intervalles de crédibilité a posteriori devient presque immédiat. Lorsqu'une information a priori est disponible, elle sert généralement à choisir le paramètre $\alpha \in A$ pour fixer le choix de la loi a priori au sein de la famille conjuguée. Cette facilité à mener l'estimation a toutefois un prix, puisque le choix de la loi a priori demeure restreint à la famille $F_{\alpha \in A}$

considérée.

Lois hiérarchiques

D'une manière très générale, dès que le choix de la loi a priori est restreint à une famille paramétrée $F_{\alpha \in A}$, que cette famille soit ou non conjuguée pour le modèle, il est nécessaire de choisir le paramètre $\alpha \in A$. Dans le cas où l'information a priori ne permet pas de fixer α , l'approche hiérarchique modélise le manque d'information sur α à l'aide d'une distribution a priori sur ce nouveau paramètre du modèle (alors appelé hyper-paramètre). La loi a priori hiérarchique ainsi construit s'exprime alors souvent sous la forme :

$$\pi(\theta, \alpha) = \pi(\theta|\alpha)\pi(\alpha). \quad (1.6)$$

Le principe de l'approche hiérarchique peut aussi s'étendre à lui-même dont la loi a priori peut dépendre d'un nouvel hyper-paramètre, etc. L'utilisation de lois a priori hiérarchiques conduit à des estimateurs plus robustes, comme illustré dans Congdon 2003 [40], au sens où l'inférence menée est moins sensible au choix des paramètres fixés par l'utilisateur.

Lois non informatives

En l'absence d'information a priori, le choix de la loi a priori s'effectue parmi les lois a priori dites non informatives puisqu'elles minimisent, en un certain sens, l'influence de la loi a priori sur la loi a posteriori. Nous ne présentons ici que quelques-unes des possibilités listées dans Kass and Wasserman (1996). Notons qu'il peut arriver que le choix de π considéré comme loi a priori ne définisse qu'une mesure positive et non une loi de probabilité sur Θ , i.e. $\pi(\theta) > 0$ pour tout $\theta \in \Theta$ et on a :

$$\int \pi(\theta) d\theta < \infty. \quad (1.7)$$

Le cadre bayésien s'étend toutefois à de tels choix de lois a priori, dites impropres, dès lors que la loi

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}, \quad (1.8)$$

est bien définie, c'est-à-dire dès que

$$\int f(x|\theta)\pi(\theta)d\theta < +\infty. \quad (1.9)$$

1. Loi de Laplace

La loi a priori de Laplace (1774) correspond au choix

$$\pi(\theta) \propto \mathbf{1}_{\Theta}(\theta). \quad (1.10)$$

En fonction de l'ensemble Θ des paramètres, nous retrouvons alternativement une loi uniforme Θ ou une loi impropre. Le choix de la loi de Laplace peut sembler naturel car aucune valeur de paramètre n'est a priori favorisée par rapport à une autre mais cette loi n'est pas invariante par reparamétrisation. En effet, si la reparamétrisation $\eta = g(\theta)$ est considérée (g étant une bijection) et si la loi de Laplace est choisie comme loi a priori sur θ , alors par changement de variable, on obtient :

$$\pi(\theta) \propto 1 \implies \tilde{\pi}(\eta) \propto \left| \frac{d}{d\eta} g^{-1}(\eta) \right|, \quad (1.11)$$

où $\tilde{\pi}$ désigne la loi a priori sur $\eta = g(\theta)$ correspondante. Bien qu'aucune information a priori ne soit disponible sur η , puisqu'aucune information a priori n'est disponible

sur θ , le choix de loi a priori sur η n'est donc plus (en général) la loi de Laplace, et le choix de la loi a priori semble donc dépendre de la formulation même du problème.

2. Loi de Jeffreys

Jeffreys (1946) propose une loi a priori qui répond à la demande d'invariance par reparamétrisation. L'approche repose sur l'information de Fisher du modèle supposé régulier (voir Lehmann et Casella, 1998) définie pour $\theta \in \Theta \subset \mathbb{R}^d$ comme la matrice $I(\theta)$ dont les coefficients sont donnés pour $1 \leq i, j \leq d$ par :

$$I_{ij}(\theta) = E_{\theta} \left[\left(\frac{\partial \log f(x|\theta)}{\partial \theta_i} \right) \left(\frac{\partial \log f(x|\theta)}{\partial \theta_j} \right) \right], \quad (1.12)$$

La loi non informative de Jeffreys est définie par :

$$\pi^* \propto \det^{\frac{1}{2}} I(\theta), \quad (1.13)$$

Clarke and Barron (1990) démontrent, dans le cas d'observations indépendantes et identiquement distribuées (i.i.d.), que ce choix minimise l'influence de la loi a priori sur la loi a posteriori au sens où elle maximise la divergence de Kullback-Leibler entre ces deux lois.

3. Loïs de référence

Bernardo (1979) propose de construire des lois dite de référence : dans cette approche, les coordonnées sont regroupées par blocs sur lesquels un ordre est fixé et la loi a priori de référence est construite de manière conditionnelle. Le choix du nombre de blocs de coordonnées ainsi que leurs compositions et l'ordre qui leur est associé influencent

donc la construction de la loi a priori de référence. Par exemple, pour $\theta = (\theta_1, \theta_2)$, où θ_1 désigne le paramètre d'intérêt et θ_2 le paramètre de nuisance, alors la loi a priori de référence est calculée en définissant d'abord $\pi(\theta_2|\theta_1)$ comme la loi de Jeffreys associée à $f(x|\theta)$ conditionnellement à θ_1 , puis $\pi(\theta_1)$ comme la loi de Jeffreys associée à :

$$\tilde{f}(x|\theta_1) = \int f(x|\theta_1, \theta_2)\pi(\theta_2|\theta_1)d\theta_2, \quad (1.14)$$

Les lois de référence constituent une généralisation de la loi de Jeffreys au sens où elles demeurent invariantes par reparamétrisation au sein de chaque bloc de coordonnées.

1.7 Inférence Bayésienne

Nous considérerons que cette loi a posteriori est disponible et nous présenterons dans quelques idées générales comment conduire une inférence sur θ (estimation, tests et prévision) en utilisant cette distribution a posteriori, c'est-à-dire une inférence au sens Bayésien.

1.7.1 Prédiction

Le contexte du problème de la prédiction est le suivant : les observations X sont identiquement distribuées selon P_θ , absolument continue par rapport à une mesure dominante μ et donc qu'il existe une fonction de densité conditionnelle $f(\cdot|\theta)$. Par ailleurs on suppose que θ suit une loi a priori π .

Il s'agit alors à partir de n tirages X_1, \dots, X_n de déterminer le plus précisément possible ce que pourrait être le tirage suivant X_{n+1} .

Dans l'approche fréquentiste, on calcule $f(X_{n+1}|X_1, \dots, X_n, \hat{\theta})$. Comme le révèle la notation $\hat{\theta}$, on ne connaît pas exactement θ . On doit donc l'estimer dans un premier temps et de ce fait, on utilise deux fois les données : une fois pour l'estimation du paramètre et une nouvelle fois pour la prédiction dans la fonction f . En général, ceci amène à sous-estimer les intervalles de confiance.

La stratégie du paradigme bayésien, désormais bien comprise par la lectrice et peut-être un peu assimilé par le lecteur, consiste à intégrer la prévision suivant une loi a priori sur θ et ce, afin d'avoir la meilleure prédiction compte tenu à la fois de notre savoir et de notre ignorance sur le paramètre. La loi prédictive s'écrit ainsi :

$$f^\pi(X_{n+1}|X_1, \dots, X_n) = \int_{\theta} f(X_{n+1}|X_1, \dots, X_n, \theta)\pi(\theta|X_1, \dots, X_n)d\theta, \quad (1.15)$$

Dans le cas des tirages indépendants et identiquement distribués, ceci devient :

$$f^\pi(X_{n+1}|X_1, \dots, X_n) = \int_{\theta} f(X_{n+1}|\theta)\pi(\theta|X_1, \dots, X_n)d\theta, \quad (1.16)$$

1.7.2 L'estimation ponctuelle

Avant de parler de l'estimation, un passage sur la théorie de la décision est très utile, car comme nous allons le voir, déterminer un estimateur de Bayes revient à déterminer une règle de décision.

Décision bayésienne

Le problème statistique étant quelle décision prendre sans expérimenter à l'aide de la loi a priori ? Si on choisit l'action a , la perte est $L(\theta, a)$ et le risque moyen est $\int_{\Theta} L(\theta, a)\pi(d\theta)$ et correspond à la stratégie déterministe S identifiée à la décision déterministe a .

Définition 1.7.1. On appelle risque bayésien pour la loi a priori π la quantité

$$\inf_A \int_{\Theta} L(\theta, \delta) \pi(d\theta), \quad (1.17)$$

avec A ce t'un espace de décision.

Si cette borne est atteinte pour une action $\delta \in A$; on dit que δ est la règle de décision bayésienne.

Plus généralement, une règle de décision δ_0 est une règle de décision bayésienne par rapport à la probabilité à priori π et la classe des règles de décision D , si

$$\int_{\Theta} R_{\delta_0}(\theta) \pi(d\theta) = \inf_{\delta \in D} \int_{\Theta} R_{\delta}(\theta) \pi(d\theta), \quad (1.18)$$

avec $R_{\delta}(\theta) = \int_A L(\theta, \delta') \pi_{\theta}^{\delta'}(d\delta')$, où δ' est la stratégie déterministe définie par la règle de décision δ .

Les règles de décision bayésiennes possèdent des propriétés intéressantes, en particulier elles sont admissibles alors que ce n'est pas toujours le cas d'une fonction de décision sans biais.

Définition 1.7.2. (Risque fréquentiste)

Pour une fonction de perte donnée $l(\theta, \delta)$, la fonction de risque associée est donnée par :

$$\begin{aligned} R(\delta, \theta) &= E_{\theta}[l(\theta, \delta(x))] \\ &= \int_X l(\theta, \delta(x)) f(x|\theta) d\mu(x). \end{aligned}$$

C'est une fonction de θ et ne définit pas un ordre total sur D et ne permet donc pas de comparer toutes décisions et estimateurs. Il n'existe donc pas de meilleur estimateur dans un sens absolu. Ainsi, l'approche fréquentiste restreint l'espace d'estimation en préférant la classe des estimateurs sans biais dans laquelle il existe des estimateurs de risque uniformément minimal; l'école Bayésienne ne perd pas en définissant un risque a posteriori. L'idée est d'intégrer sur l'espace des paramètres pour pallier à cette difficulté.

Définition 1.7.3. (Le risque a posteriori)

Pour un Bayésien, θ est une variable aléatoire de distribution a priori $\pi(\theta)$, et après que les données seront disponibles, la distribution pertinente de θ sera donnée par la distribution a posteriori $\pi(\theta|x)$ et le risque pertinent sera le risque a posteriori ou bien le risque Bayésien, voir que :

$$\begin{aligned} \rho(\pi, \delta|x) &= E^\pi(l(\theta, \delta(x))|x) \\ &= \int_{\Theta} l(\theta, \delta(x)) \pi(\theta|x) d\theta. \end{aligned}$$

Ainsi, le problème change selon les données; ceci dû à la non existence d'un ordre total sur les estimateurs.

Définition 1.7.4. (Le risque intégré)

Pour une fonction de perte donnée, le risque intégré est défini par :

$$\begin{aligned} r(\pi, \delta) &= E(R(\theta, \delta)|x) \\ &= \int_{\Theta} R(\theta, \delta)\pi(\theta)d\theta. \end{aligned}$$

Une fois que la loi a posteriori sur le paramètre est disponible, le problème de l'estimation Bayésienne ponctuelle peut être exprimé comme un problème de décision.

Proposition 1.7.1. ([45]) *Si la loi a priori π est telle que $\pi(\Delta) > 0; \forall \Delta \subseteq \Theta$, alors la règle de Bayes relative à π est admissible.*

Proof. Supposons que la règle de Bayes δ_0 n'est pas admissible. Il existe alors une règle de décision δ' meilleure que δ_0 au sens que $R_{\delta'}(\theta) \leq R_{\delta_0}(\theta), \forall \theta \in \Theta$ et $R_{\delta'}(\theta) < R_{\delta_0}(\theta)$, il s'ensuit que les risques bayésiens relatifs aux règles de décision δ et δ_0 satisfont l'équation suivante :

$$R_{\delta'}^{\pi}(\theta) < R_{\delta_0}^{\pi}(\theta), \quad (1.19)$$

ce qui contredit que δ_0 est une règle de décision bayésienne.

Fonctions de coût usuelles

Coût quadratique

Définition 1.7.5. La fonction de coût quadratique est la fonction définie par :

$$L(\theta, \delta(x)) = (\theta - \delta(x))^2. \quad (1.20)$$

Une variante de cette fonction de coût est une fonction de coût quadratique pondérée de la forme :

$$L(\theta, \delta(x)) = w(\theta)(\theta - \delta(x))^2. \quad (1.21)$$

Proposition 1.7.2. ([45]) : Sous l'hypothèse d'un coût quadratique, l'estimateur de Bayes $\delta^\pi(x)$ de θ associé à la loi à priori π est la moyenne a posteriori de θ :

$$\delta^\pi(x) = E^{\pi(\cdot|x)}(\theta) = \int_{\theta \in \Theta} L(\theta, \delta(x)) \pi(\theta|x) d\theta. \quad (1.22)$$

Preuve : Par définition, l'estimateur de Bayes minimise le coût a posteriori i.e. :

$$\rho(\pi, \delta) = E^{\pi(\cdot|x)}[L(\theta, \delta(x))]. \quad (1.23)$$

Sous l'hypothèse d'un coût quadratique, on a :

$$\rho(\pi, \delta) = E^{\pi(\cdot|x)}[(\theta - \delta(x))^2] = E^{\pi(\cdot|x)}(\theta)^2 - 2\delta(x)E^{\pi(\cdot|x)}(\theta) + \delta(x)^2. \quad (1.24)$$

Il s'agit d'un polynôme du second degré en $\delta(x)$. Il sera minimum en $E^{\pi(\cdot|x)}(\theta)$.

Coût absolu

Définition 1.7.6. La fonction de coût absolue est la fonction définie par :

$$L(\theta, \delta(x)) = \begin{cases} k_2(\theta - \delta(x)), & \text{si } \theta > \delta(x), \\ k_1(\delta(x) - \theta), & \text{sinon.} \end{cases} \quad (1.25)$$

Proposition 1.7.3. ([25]) : Un estimateur de Bayes associé à π et au coût absolue, est une fractile d'ordre $k_2/(k_1 + k_2)$ de $\pi(\theta/x)$.

Preuve :

$$\begin{aligned}
E^{\pi(\cdot|x)}[L(\theta, \delta(x))] &= \int_{\theta \in \Theta} L(\xi, \delta(x)) \pi(\xi|x) d\xi \\
&= \int_{\delta(x)}^{+\infty} k_2(\xi - \delta(x)) \pi(\xi|x) d\xi + \int_{-\infty}^{\delta(x)} k_1(\xi - \delta(x)) \pi(\xi|x) d\xi.
\end{aligned} \tag{1.26}$$

On remarque que : $\pi(\xi|x)d\xi = dF(\xi|x) = -d(1 - F(\xi|x))$ et on écrit donc :

$$\begin{aligned}
E^{\pi(\cdot|x)}[L(\theta, \delta(x))] &= [k_2(\xi - \delta(x))(1 - F(\xi|x))]_{\delta(x)}^{+\infty} + \int_{\delta(x)}^{+\infty} k_2 P^{\delta(\cdot|x)}(\theta > \xi) d\xi \\
&\quad + k_1(\xi - \delta(x))F(\xi|x) + \int_{-\infty}^{\delta(x)} k_1 P^{\delta(\cdot|x)}(\theta < \xi) d\xi \\
&\quad \int_{\delta(x)}^{+\infty} k_2 P^{\delta(\cdot|x)}(\theta > \xi) d\xi + \int_{-\infty}^{\delta(x)} k_1 P^{\delta(\cdot|x)}(\theta < \xi) d\xi.
\end{aligned} \tag{1.27}$$

On dérive par rapport à $\delta(x)$, on obtient :

$$-k_2 P^{\delta(\cdot|x)}(\theta > \delta(x)) + k_1 P^{\delta(\cdot|x)}(\theta < \xi) = 0. \tag{1.28}$$

$$\iff -k_2(1 - P^{\delta(\cdot|x)}(\theta > \delta(x))) + k_1 P^{\delta(\cdot|x)}(\theta < \xi) = 0. \tag{1.29}$$

$$\iff (k_1 + k_2) P^{\delta(\cdot|x)}(\theta < \delta(x)) - k_2 = 0, \tag{1.30}$$

d'où

$$P^{\delta(\cdot|x)}(\theta < \delta(x)) = \frac{k_2}{k_1 + k_2}, \tag{1.31}$$

et le coût est donc maximisé pour $\tilde{\theta} = \delta(x)$ tel que $P^{\delta(\cdot|x)}(\theta < \delta(x)) = k_2/(k_1 + k_2)$.

Coût 0-1

Définition 1.7.7. On appelle coût 0 – 1, l'application L définie par :

$$L(\theta, \delta(x)) = \begin{cases} 0, & \text{si la décision est bonne;} \\ 1, & \text{sinon.} \end{cases} \quad (1.32)$$

On retrouve en utilisant cette fonction de coût, les résultats de la théorie des tests d'hypothèses. Un problème de test est un problème de choix (de prise de décision) entre $H_0 : "$ $\theta \in \Theta_0$ " contre $H_1 : "$ $\theta \in \Theta_1$ ".

On définit donc la décision de la manière suivante :

$\delta(X) = 1$: on accepte H_0 ;

$\delta(X) = 0$: on rejette H_0 .

On a un espace d'actions de la forme : $A = \{0, 1\}$.

Soit W la région de rejet i.e. le sous-ensemble de \mathcal{X} qui conduit à rejeter H_0 . On peut construire une fonction de coût de la manière suivante : supposons $\theta \in \Theta_0$,

si $X \in W$, on prend la décision de rejeter i.e., $\delta(X) = 0$, mais la décision n'est pas bonne, on va pénaliser et $L(\theta, \delta(x)) = 1$.

si $X \notin W$, on ne rejette pas, on prend la décision $\delta(X) = 1$, la décision est bonne $L(\theta, \delta(X)) = 0$.

Le coût s'écrit sous la forme :

$$L(\theta, \delta(x)) = \begin{cases} 1 - \delta(x) & \text{si } \theta \in \Theta_0; \\ \delta(x) & \text{sinon.} \end{cases} \quad (1.33)$$

Ce qu'on peut écrire : $L(\theta, \delta(x)) = \mathbf{1}_{(x \in W)}$ et on calcule la fonction de risque :

$$R(\theta, \delta) = E[L(\theta, \delta(X))] = \int_{\mathcal{X}} L(\theta, \delta(x)) dp_{\theta}(x) = P_{\theta}(x \in W), \theta \in \Theta_0. \quad (1.34)$$

On retrouve le risque de première espèce.

Définition 1.7.8. (Estimateur de Bayes)

On appelle estimateur de Bayes associé à un coût L et à une distribution a priori π , toute décision δ^{π} qui minimise le risque de Bayes $r(\pi, \delta)$.

On a :

$$\delta^{\pi}(x) = \text{Argmin}_{\delta} r(\pi, \delta). \quad (1.35)$$

Intervalle de crédibilité bayésien

L'intervalle de crédibilité est un reflet réel de la confiance que l'on peut avoir sur la valeur du paramètre concerné. Il met en évidence le point de vue original de l'approche bayésienne qui ne considère pas que le paramètre est une quantité fixe inconnue mais une variable aléatoire dont nous avons une appréciation plus ou moins exacte.

Par définition, un intervalle de crédibilité au niveau α est un intervalle tel que la probabilité du paramètre de lui appartenir selon la distribution a posteriori est de $(1 - \alpha)$.

TAB. 1.1 – Lois a priori conjuguées pour quelques familles exponentielles

Loi de x	Loi conjuguée	Moyenne a posteriori
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\frac{\mu\sigma^2 + \tau^2 x}{\tau^2 + \sigma^2}$
Poisson $p(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + x}{\beta + 1}$
Gamma $\mathcal{G}(\nu, \theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + \nu}{\beta + x}$
Binomiale $B(n, \theta)$	Bêta $Be(\alpha, \beta)$	$\frac{\alpha + x}{\alpha + \beta + n}$
Binomiale Négative $Neg(m, \theta)$	Bêta $Be(\alpha, \beta)$	$\frac{\alpha + n}{\alpha + \beta + x + n}$
Multinomiale $M_k(\theta_1, \dots, \theta_k)$	Dirichlet $D(\alpha_1, \dots, \alpha_k)$	$\frac{\alpha_i + x_i}{\sum_j \alpha_j + n}$
Normale $\mathcal{N}(\mu, 1/\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\frac{\alpha + 1}{\beta + (\mu - x)^2}$

Il y a plusieurs manière de le construire ; par exemple, on peut prendre celui donné par les $\alpha/2$ et $(1 - \alpha/2)$ quantiles.

Si on s'intéresse à plusieurs paramètres simultanément, on parlera de région de crédibilité.

Le tableau ci-dessous présente quelques estimateur de Bayes de paramètre θ sous Coût quadratique pour les lois a priori conjuguées des familles exponentielles usuelles.

Quelques estimateurs de Bays usuels :

1.7.3 Estimation paramétrique Bayésienne

Estimateur MMSE (posterior minimum mean square estimate)

L'estimateur MMSE est défini par la moyenne de la densité à posteriori considérée. Étant donné un vecteur de paramètres θ et un vecteur d'observation y on a :

$$\hat{\theta}_{MMSE} = \int_{\Theta} p(\theta|y) d\theta. \quad (1.36)$$

Estimateur MAP(maximum a posteriori)

L'estimateur MAP est déterminé par le mode de la densité à posteriori considérée :

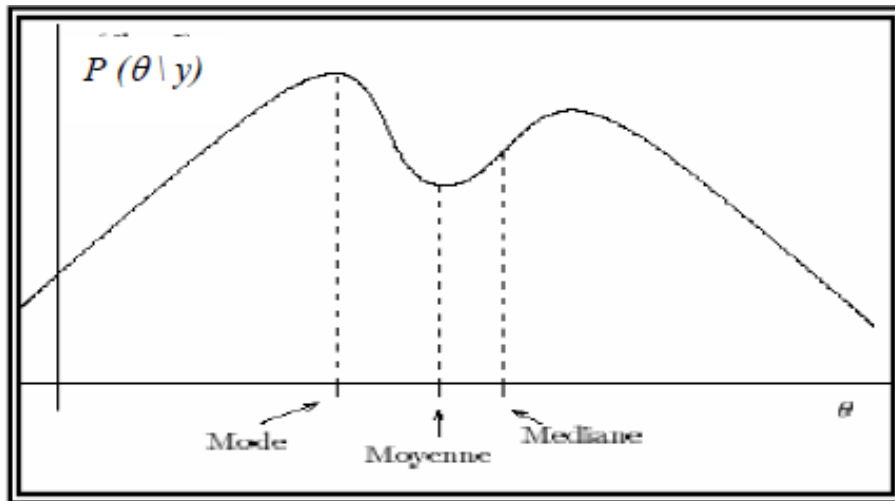


FIG. 1.1 – Estimateur paramétrique Bayésien

Pour résoudre un tel problème, nous proposons une méthode de simulation des densités de probabilité telle que la méthode Monte Carlo par Chaîne de Markov, basée sur la génération de variables aléatoires distribuées suivant une loi π à simuler.

Un estimateur de référence de θ fondé sur $\pi(\theta|x)$ est un estimateur de maximum a posteriori (MAP), défini comme le mode a posteriori par :

$$\delta_{MAP} = \underset{\theta \in \Theta}{\operatorname{argmax}} \pi(\theta|x). \quad (1.37)$$

Cette estimateur est associé au coût, cette fonction indique 1 quand $\theta \neq \delta$ ce qui donne à cette estimateur un signification globale, il s'agit d'optimiser un seul critère. La densité jointe $\pi(\theta|x)$, dans le cas continue puisque pour tout $\delta \in \Theta$

$$\int_{\Theta} 1_{\theta \neq \delta} \pi(\delta/x) d\delta = 1. \quad (1.38)$$

La fonction de coût associé peut être remplacée par une suite de coût $L_{\zeta}(d, \delta) = 1_{\|\delta-d\|>\zeta}$, donc l'estimateur MAP est alors la limite des estimateur de Bayes à L_{ζ} . quand ζ tend vers 0.

Le MAP peut être s'exprimer comme un estimateur du maximum de vraisemblance lorsque la taille d'échantillon tend vers l'infini.

Le grand avantage de cet estimateur est qu'il ne dépend pas d'une fonction de perte, et est utile pour les approches théoriques, sauf dans des procédures des tests où l'estiamteur MAP s'interprète comme l'estimateur associé à la fonction de perte 0 - 1 comme nous allons le constater dans la prochaine section.

1.7.4 Analyse Bayésienne non paramétrique

Dans le cadre paramétrique, le modèle statistique classique $(\mathcal{X}, \mathcal{A}, p_{\theta}, \theta \in \Theta)$, où \mathcal{X} est l'espace des observations, \mathcal{A} est la tribu associée, p_{θ} la loi de l'observation, Θ l'espace des paramètres ($\Theta \subset \mathbb{R}^s$), est transformé en un modèle statistique bayésien par la donnée d'une loi a priori $\pi(\cdot)$ sur θ .

Dans le cadre non paramétrique, l'objet sur lequel on infère est non plus θ mais p_{θ} la distribution de l'observation ou d'une manière plus générale, une fonctionnelle $g(P_{\theta})$. Ce peut être la fonction de répartition, la densité, la fonction de survie (ou de fiabilité), le taux de survie (ou de défaillance), l'intensité d'un processus, etc.

L'approche bayésienne de l'estimation d'une fonctionnelle nécessite donc la construction d'une loi a priori sur P_θ , la mesure probabilité associée à l'observation. Elle est maintenant considérée comme une mesure aléatoire.

Soit M l'ensemble des mesures de Radon sur \mathbb{R} i.e. l'ensemble des mesures μ telle que, pour tout $A \in \mathcal{A}$, $\mu(A) < \infty$.

On peut donner la définition suivante d'une mesure aléatoire :

Définition 1.7.9. Une mesure aléatoire sur \mathbb{R} est une application mesurable de $(\mathcal{X}, \mathcal{A})$ dans (\mathbb{M}, M) .

Soit η une mesure aléatoire, $\forall A \in \mathcal{A}$, $\eta(A)$ est une variable aléatoire.

Le processus associé à cette mesure aléatoire η sera caractérisée par la donnée de la densité jointe de $(\eta(A_1), \eta(A_2), \dots, \eta(A_{m+1}))$ où A_1, A_2, \dots, A_{m+1} est une partition de \mathcal{X} .

η est un processus stochastique indexé par des éléments de \mathcal{A} .

Exemple. Le processus de Poisson est un exemple simple de processus répondant aux conditions énoncées ci-dessus. Considérant, une mesure de comptage η sur \mathbb{R}_+ et supposant que pour tout A , $\eta(A)$ est une v.a. discrète qui suit une loi de Poisson de paramètre $\mu(A)$ ou μ est une mesure de Radon. On dira que η est une mesure aléatoire de Poisson si :

$$Pr(\eta(A) = k) = \frac{\mu(A)^k}{k!} \exp -\mu(A), \quad k \in \mathbb{N}.$$

On définira le processus de Poisson d'intensité μ associée a cette mesure η par :

Définition 1.7.10. Soit μ une mesure finie non nulle sur \mathbb{R}^+ .

On dit que la mesure η est un processus de Poisson de paramètre μ , et on note $\eta \in \mathbf{p}(\mu)$ si, pour toute partition mesurable $(A_1, A_2, \dots, A_{m+1})$ de \mathbb{R}^+ , la loi de $(\eta(A_1), \eta(A_2), \dots, \eta(A_m))$ est un produit de lois de Poisson de paramètre respectif $\mu(A_i)$, $i = 1, \dots, m + 1$.

Le processus fondateur de l'approche bayésienne des problèmes non paramétriques est le processus de Dirichlet. Rappelons tout d'abord, la définition et les propriétés de la loi de Dirichlet :

Définition 1.7.11. Un vecteur aléatoire (U_1, U_2, \dots, U_m) à valeurs dans le sous-espace de \mathbb{R}^m :

$$\Delta_m = \left\{ (u_1, u_2, \dots, u_m); u_i \geq 0, i = 1, 2, \dots, m; \sum_{i=1}^m u_i \leq 1 \right\},$$

permet d'exploiter la structure récursive d'un modèle pour l'estimer, plutôt que d'être forcé d'estimer tous les paramètres simultanément.

et suit une loi de Dirichlet de paramètres $(\alpha_1, \alpha_2, \dots, \alpha_m; \alpha_{m+1})$ si la densité de sa loi s'écrit :

$$\frac{\Gamma(\alpha_1, \alpha_2, \dots, \alpha_{m+1})}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{m+1})} u_1^{\alpha_1-1} \dots u_m^{\alpha_m-1} (1 - u_1 - \dots - u_m)^{\alpha_{m+1}-1}. \quad (1.39)$$

Remarquons que cette distribution est une manière de généralisation de la loi Beta. Pour $m = 1$, on retrouve la loi Beta de paramètres $(\alpha_1; \alpha_2)$.

On définit alors le processus de Dirichlet comme suit :

Définition 1.7.12. Soit α une mesure finie non nulle sur \mathbb{R}^+ . On dit que la mesure P est un processus de Dirichlet de paramètre α , et on note $P \in \mathcal{D}(\alpha)$, si, pour toute

partition mesurable $(A_1, A_2, \dots, A_{m+1})$ de \mathbb{R}_+ , la loi de $(P(A_1), P(A_2), \dots, P(A_m))$ est une loi de Dirichlet de paramètres $(\alpha(A_1), \alpha(A_2), \dots, \alpha(A_m); \alpha(A_{m+1}))$.

Proposition 1.7.4. ([1]) Soient $(Y_1, Y_2, \dots, Y_{m+1})$, $m+1$ variables aléatoires indépendantes de loi Gamma de paramètres respectifs $(\alpha_i, 1)$, $i = 1, \dots, m+1$. On considère les variables :

$$X_i = \frac{Y_i}{Y_1 + \dots + Y_{m+1}}, \quad i = 1, \dots, m.$$

Alors (X_1, \dots, X_m) suit une loi de Dirichlet de paramètres $(\alpha_1, \alpha_2, \dots, \alpha_m; \alpha_{m+1})$.

Ce résultat permet d'établir :

Propriété 1.7.1. ([1]) Si $Y_l \sim \mathcal{G}(\alpha_l, 1)$, alors pour tous nombres entiers $i < j$, $\sum_{l=i}^j Y_l \sim \mathcal{G}(\sum_{l=i}^j \alpha_l, 1)$.

A partir de ces résultats, on peut montrer que la loi marginale de chaque X_j est une loi Beta de paramètres $(\alpha_j, \alpha - \alpha_j)$ avec $\alpha = \sum_{j=1}^m \alpha_j$. On en déduit l'espérance mathématique de chacune des composantes du vecteur aléatoire (X_1, \dots, X_m) : $E(X_j) = \alpha_j / \alpha, j = 1, \dots, m$. Ces résultats se transposent au cas du processus de Dirichlet et permettent de construire un estimateur de Bayes de la fonction de répartition.

Proposition 1.7.5. ([1]) Si (X_1, \dots, X_m) suit une loi de Dirichlet de paramètres $(\alpha_1, \dots, \alpha_m; \alpha_{m+1})$ alors la loi marginale de $(X_{i_1}, X_{i_2}, \dots, X_{i_q}), q < m$, suit une loi de Dirichlet de paramètres $(\alpha_{i_1}, \dots, \alpha_{i_q}, \sum_{j \neq i_l, l=1, \dots, q} \alpha_j)$. Cette proposition s'obtient en considérant la propriété précédente et en appliquant la propriété d'additivité de la loi Gamma.

1.7.5 Estimateur de Bayes de la fonction de répartition

Soit (X_1, \dots, X_n) un n -échantillon. On note $(X_{(1)}, \dots, X_{(m)})$, les m observations ordonnées et distinctes. Considérons la partition de \mathbb{R} formée des intervalles dont les bornes sont les $X_{(j)}, j = 1, \dots, m$ avec $X_{(0)} = -\infty$ et $X_{(m+1)} = +\infty$. On pose $I_j = [X_{(j-1)}, X_{(j)}[, j = 1, \dots, m+1$ Considérons maintenant que l'observation consiste en un vecteur $(\varepsilon_X(I_1), \dots, \varepsilon_X(I_m))$ où $\varepsilon_X(I_j) = \sum_{i=1}^n \mathbf{1}_{(X_i \in I_j)}$.

Ce vecteur suit une loi multinomiale de paramètre $(P(I_1), \dots, P(I_m); P(I_{m+1}))$:

$$f(X|P) = \prod_{j=1}^{m+1} P(I_j)^{\varepsilon_X(I_j)}.$$

ε_X est un processus multinomiale de mesure-paramètre P . Une loi a priori naturelle (loi conjuguée) pour une loi multinomiale est la loi de Dirichlet. On considérera donc comme loi a priori sur le processus P , un processus de Dirichlet de mesure-paramètre α :

$$\pi(P) = \prod_{j=1}^{m+1} P(I_j)^{\alpha(I_j)-1}.$$

Et la loi a posteriori sera donnée par :

$$\pi(P|X) \propto f(X|P)\pi(P) = \prod_{j=1}^{m+1} P(I_j)^{(\varepsilon+\alpha)(I_j)-1}.$$

j La loi a posteriori de P est une loi de Dirichlet ou $P|_X$ est un processus de Dirichlet de paramètre-mesure $\varepsilon + \alpha$.

1.7.6 Tests et régions de confiance

Région de confiance

Définition 1.7.13. Région α -crédible

Une région C de Θ est dite α -crédible si et seulement si $P^\pi(\theta \in C|X) > 1 - \alpha$.

Notons que le paradigme bayésien permet une nouvelle fois de s'affranchir d'un inconvénient de l'approche fréquentiste. En effet, au sens fréquentiste, une région de confiance C est définie par :

$$\forall \theta; P_{\theta}(\theta \in C) \geq 1 - \alpha, \quad (1.40)$$

et correspond à l'interprétation suivante. En refaisant l'expérience un grand nombre de fois, la probabilité que θ soit dans C est plus grande que $1 - \alpha$. Une région de confiance n'a donc de sens que pour un très grand nombre d'expériences tandis que la définition bayésienne exprime que la probabilité que θ soit dans C au vue des celles déjà réalisées est plus grande que $1 - \alpha$. Il n'y a donc pas besoin ici d'avoir recours à un nombre infini d'expériences pour définir une région α -crédible, seule compte l'expérience effectivement réalisée.

Il y a une infinité de régions α -crédibles, il est donc logique de s'intéresser à R la région qui a le volume minimal. Le volume étant défini par $vol(C) = \int_C d\nu(\theta)$, si $\pi(\theta|X)$ est absolument continue par rapport à une mesure de référence ν .

Définition 1.7.14. Région HPD (Highest Posteriori Density)

C_{α}^{π} est une région HPD si et seulement si $C_{\alpha}^{\pi} = \{\theta | \pi(\theta|X) \geq h_{\alpha}\}$ où h_{α} est défini par $h_{\alpha} = \sup\{h, P^{\pi}(\theta, \pi(\theta|X) \geq h|X) \geq 1 - \alpha\}$.

C_{α}^{π} est parmi les régions qui ont une probabilité supérieure à $1 - \alpha$ de contenir θ (et qui sont donc α -crédibles) et sur lesquelles la densité a posteriori ne descend pas sous un certain niveau (restant au dessus de la valeur la plus élevée possible).

Tests

D'un point de vue statistique, un test soit au sens Bayésien ou au sens classique peut être considéré comme une des deux approches suivantes : Soit comme un procédé statistique, c'est-à-dire une fonction définie sur l'espace des observations à valeurs dans un espace à deux points appelées "accepter" et "rejeter" une hypothèse. Dans ce cas, on peut envisager un problème de test comme un problème de décision avec deux actions possibles. Sinon, il peut être considéré comme une façon pour le statisticien de gérer ses doutes relatifs à son modèle statistique .

Facteur de Bayes

Supposons que nous avons deux hypothèses :

$$H_0 : \theta \in \Theta_0,$$

$$H_1 : \theta \in \Theta_1,$$

et nous devons choisir une parmi les deux dans un concept Bayésien. Pour ce faire, il suffit de comparer les probabilités a posteriori des deux hypothèses suivantes :

$$p_0 = P(H_0|x) = P(\theta \in \Theta_0|x);$$

$$p_1 = P(H_1|x) = P(\theta \in \Theta_1|x);$$

la règle de Bayes consiste à choisir l'hypothèse de plus grande probabilité a posteriori.

Supposons aussi que nous disposons des probabilités a priori des hypothèses :

$$\pi_0 = P(\theta \in \Theta_0) = P(H_0),$$

$$\pi_1 = P(\theta \in \Theta_1) = P(H_1) = 1 - \pi_0.$$

Le odds a priori (the odds prior) est le rapport des probabilités a priori de H_0 relativement à H_1

$$\frac{\pi_0}{\pi_1},$$

Ce rapport égal à 1 signifie que les hypothèses sont les mêmes avant d'observer les données. De même nous pouvons définir le odds a posteriori (the odds posterior) comme

$$\frac{p_0}{p_1},$$

Le facteur de Bayes en faveur de H_0 relativement à H_1 est le rapport des deux odds :

$$B_F = \frac{\text{odds a posteriori}}{\text{odds a priori}} = \frac{p_0 \pi_1}{\pi_1 \pi_0}. \quad (1.41)$$

Les facteurs de Bayes sont très flexibles pour la comparaison des hypothèses multiples et des modèles.

Interprétation du facteur du Bayes

Comme la règle de Bayes consiste à choisir l'hypothèse dont le modèle a la plus grande probabilité a posteriori. Le facteur de Bayes peut être interprété comme suit :

$B_F \geq 1$	nous acceptons H_0
$10^{-1/2} \leq B_F < 1$	la certitude que H_0 est fautive est minimale
$10^{-1} \leq B_F < 10^{-1/2}$	cette certitude est substantielle
$10^{-2} \leq B_F < 10^{-1}$	la certitude est forte
$B_F < 10^{-2}$	la certitude est décisive et nous devons rejeter H_0

Le facteur de Bayes peut être aussi utilisé pour comparer deux modèles

$$\begin{aligned}
M_0 &= \{f_0(x|\theta_0), g_0(\theta_0)\}, \\
M_1 &= \{f_1(x|\theta_1), g_1(\theta_1)\},
\end{aligned}$$

le facteur de Bayes dans ce cas est défini comme suit :

$$\begin{aligned}
B_F &= \frac{P(X|M_0)}{P(X|M_1)} \\
&= \frac{\int_{\Theta_0} f_0(x|\theta_0) g_0(\theta_0) d\theta_0}{\int_{\Theta_1} f_1(x|\theta_1) g_1(\theta_1) d\theta_1}.
\end{aligned}$$

Le odds a posteriori est défini de la même façon que la précédente, à savoir :

$$\frac{P(M_0|X)}{P(M_1|X)} = \frac{P(M_0)}{P(M_1)} B_F.$$

Dans la pratique, on prend souvent $P(M_0) = P(M_1) = 1/2$, c'est-à-dire que le odds a posteriori sera égal au facteur de Bayes.

1.8 Les méthodes de simulation de Monte Carlo par Chaîne de Markov(MCMC)

1.8.1 Les chaînes de Markov

Une chaîne de Markov est une suite de variables aléatoires $(X_i, i \in \mathbb{N})$ qui permet de modéliser l'évolution dynamique d'un système aléatoire : X_i représente l'état du système à l'instant i . La propriété fondamentale des chaînes de Markov, dite propriété de Markov, est

que son évolution future ne dépend du passé qu'au travers de sa valeur actuelle. Autrement dit, conditionnellement à $X_i, (X_0, \dots, X_i)$ et $(X_{i+k}, k \in \mathbb{N})$ sont indépendants. Les applications des chaînes de Markov sont très nombreuses (réseaux, génétique des populations, mathématiques financières, gestion de stock, algorithmes stochastiques d'optimisation, simulation, . . .).

Définition 1.8.1. Une chaîne de Markov est une collection de variables aléatoires $(X_i, i \in \mathbb{N})$. L'évolution de cette chaîne sur un espace est régie par le noyau de transition qui est un mécanisme d'écrivant le mouvement de la probabilité d'un état à un autre basant sur l'état actuel, et qui correspond à la distribution conditionnelle de X_{i+1} sachant tout le passé :

$$P(x, A) = P(X_{i+1} \in A | X_i = x, X_j, j < i), x \in \Omega, A \subset \Omega \quad (1.42)$$

$$= P(X_{i+1} \in A | X_i = x). \quad (1.43)$$

Irréductibilité, récurrence, périodicité

Définition Soit $X = (X_i, i \in \mathbb{N})$ une chaîne de Markov de matrice de transition P . On dit que x est un état absorbant de la chaîne X si $P(x, x) = 1$.

En particulier, si la chaîne de Markov atteint un de ses points absorbants, elle ne peut plus s'en échapper.

Définition 1.8.2. On dit qu'une chaîne de Markov, ou sa matrice de transition, est irréductible si pour tous $x, y \in \Omega$, la probabilité partant de x d'atteindre y est strictement positive, autrement dit : si pour tous $x, y \in \Omega$, il existe $n = n_{x,y} \geq 1$ (dépendant a priori de x et y) tel que $P^n(x, y) > 0$.

La condition $P^n(x, y) > 0$ est équivalente à l'existence de $n > 1, x_0 = x, x_1, \dots, x_n = y$ tels que :

$$P(X_1 = x_1, \dots, X_n = x_n | X_0 = x_0) = \prod_{k=1}^n P(x_{k-1}, x_k) > 0. \quad (1.44)$$

Définition 1.8.3. Une chaîne de Markov irréductible est récurrente si l'espérance du nombre de visites qu'elle accorde à chaque état est infini :

$$\forall(x, y), E(x \rightarrow y) = \sum_{n=1}^{\infty} P^n(x, y) = \infty. \quad (1.45)$$

Dans le cas où l'espace d'états est fini toute chaîne irréductible est récurrente, en effet, le nombre d'états étant fini, il existe donc au moins un état qui est visité infiniment souvent qu'on itère la chaîne à l'infini. Cet état étant connecté à tous les autres, chacun des états étant visité infiniment souvent. La question de la récurrence de la chaîne ne se pose donc réellement que lorsque l'espace d'états est infini.

Une chaîne récurrente est positive si la fréquence de visites de tout sous-ensemble A à partir d'un état de départ θ est strictement positive, et nous avons alors un candidat à une distribution invariante de la chaîne.

Définition 1.8.4. Soit X une chaîne de Markov de matrice de transition P . La période d'un état $x \in \Omega$ est le PGCD de $n \in \mathbb{N}^*$; $P^n(x, x) > 0$. Un état est dit apériodique si sa période est 1, sinon il est dit périodique. Une chaîne de Markov est dite apériodique si tous ses états sont apériodiques.

Lorsque la chaîne de Markov vérifie toutes ces propriétés, c'est-à-dire qu'elle est apériodique, irréductible et récurrente positive, elle sera dite chaîne ergodique.

1.8.2 Méthodes de Monte Carlo par chaîne de Markov

La méthode MCMC tire son nom de l'idée que, pour produire des approximations acceptables d'intégrales et d'autres fonctions dépendant d'une loi d'intérêt, il suffit de générer une chaîne de Markov $(\theta(t))_t$ de loi limite la loi d'intérêt.

Le but des méthodes MCMC est de simuler selon π et son idée de base est de construire une chaîne de Markov ergodique d'une loi stationnaire π .

On suppose ici que l'on cherche à simuler une distribution π sur un ensemble de dimension élevé et qu'il est difficile, voire impossible, de générer cette distribution directement. Les algorithmes MCMC (méthodes de Monte Carlo par chaîne de Markov) sont des méthodes permettant d'effectuer une telle simulation, via la simulation d'une chaîne de Markov dont la mesure invariante est π . Rappelons que si P est la matrice des probabilités de transition d'une chaîne de Markov sur l'espace d'état S et à temps discret, une distribution de probabilité π sur E est dite mesure invariante de la chaîne si, vue comme un vecteur ligne, on a $\pi = \pi p$. On dit aussi que π est une distribution stationnaire de la chaîne.

Deux algorithmes MCMC sont présentés dans la section suivante : l'algorithme de Metropolis Hastings et l'échantillonnage de Gibbs. Ces méthodes permettent de construire un noyau de transition en fonction d'une distribution d'intérêt donnée telle que les séquences d'échantillons simulés qui forment une chaîne de Markov convergente vers la distribution stationnaire voulue.

Les algorithmes de Metropolis et de Gibbs sont deux méthodes populaires que nous allons décrire pour échantillonner une distribution π donnée. L'algorithme de Metropolis, développé en 1953 sur des problèmes de transport de particules, est une technique générale permettant de construire une classe entière de matrices de probabilité de transition P

(correspondant à une chaîne de Markov $X = (X_i), i \geq 0$) ayant toutes la mesure invariante p désirée, définies sur le même espace S que π . L'algorithme est le suivant :

Algorithme de Metropolis-Hastings

La technique de Historiquement, Metropolis-Hastings est la première des méthodes MCMC, elle a été développée par Metropolis et al. (1953)[40], au départ pour la physique particulière, et généralisée par Hastings (1970)[56] dans un cadre plus statistique. Elle est fondée sur la construction d'une distribution de proposition J qui génère un candidat θ^* , et sur la base de la probabilité d'acceptation; nous acceptons ou rejetons ce candidat mais conservant la valeur précédemment simulée en cas de rejet comme nous le verrons prochainement. L'algorithme de Metropolis-Hastings est une généralisation de l'algorithme d'acceptation-rejet, l'idée est parvenue de fait que, les réalisations successives de θ_t dans l'algorithme d'acceptation-rejet sont indépendantes ce qui implique que la décision de considérer une réalisation issue de la loi à simuler comme réalisation candidate de la loi d'intérêt ne peut être prise que sur les données présentes, ce qui nécessite une décision plus nuancée qui peut s'appuyer sur l'utilisation des informations antérieurs sur la chaîne $(\theta^{(t)})_{t \geq 0}$ qui doit avoir alors une structure de mémoire, la structure la plus simple et la mieux placée que l'on puisse envisager pour répondre à ce problème est celle de chaînes de Markov.

Pour une distribution a posteriori donnée $\pi(\theta|x)$, la procédure itérative de Metropolis-Hastings, génère à partir d'une valeur θ_i , la valeur suivante θ_{i+1} sur la base d'un algorithme en deux temps :

1. D'abord, on choisit une valeur candidate θ^* tirée aléatoirement d'une distribution de probabilité $J(\theta^*|\theta_i)$ éventuellement dépendante de θ_i . Cette loi est dite loi de proposition, mais aussi appelée "fonction de saut" parce qu'elle permet à la chaîne de bouger dans Θ à partir d'un point donné.

2. Ensuite le candidat θ^* est accepté avec une probabilité :

$$\alpha(\theta_i, \theta^*) = \min\left(1, \frac{\pi(\theta^*|x)J(\theta^*|\theta^{(i)})}{\pi(\theta^{(i)}|x)J(\theta^{(i)}|\theta^*)}\right), \quad (1.46)$$

”Accepter” le candidat signifie le choisir comme valeur suivante de la chaîne : $\theta_{i+1} = \theta^*$. Si le candidat est refusé, alors la chaîne ne bouge pas de θ_i : $\theta_{i+1} = \theta_i$. En pratique, si le rapport de la formule (1.46) est supérieur à 1, on accepte le candidat. S’il est inférieur à 1, on tire une valeur u d’une loi uniforme $U(0, 1)$ et on définit :

$$\theta_{i+1} = \theta^* \quad \text{si} \quad u \leq \alpha(\theta_i, \theta^*), \quad (1.47)$$

$$\theta_{i+1} = \theta_i \quad \text{si} \quad u > \alpha(\theta_i, \theta^*). \quad (1.48)$$

Notons que $\theta^{(0)}$ doit avoir une probabilité positive

$$P(\theta^{(0)}|x) > 0. \quad (1.49)$$

Lemme 1.8.1. ([44]) *Lorsque le support de $J(\cdot|\theta)$ contienne le support de $\pi(\cdot|x)$, i.e., $\text{supp}\pi(\cdot|x) \subset \text{supp}J(\cdot|\theta)$. La chaîne de Markov $(\theta^{(i)})_i$ produite par l’algorithme de Metropolis-Hastings est irréductible.*

L’irréductibilité de la chaîne découle de la condition sur le support de J , qui n’est cependant pas nécessaire pour assurer la validité de l’algorithme.

Échantillonnage de Gibbs

L’échantillonneur de Gibbs est la technique MCMC la plus simple. Sa popularité date de l’application de Geman et Geman (1984)[48]. L’expression ”échantillonneur de Gibbs” vient de l’utilisation que Geman et Geman ont fait de la distribution Gibbs pour modéliser les images satellites, mais son applicabilité est beaucoup plus générale.

Écrivons le vecteur de paramètres par $\theta = (\theta_1, \theta_2, \dots, \theta_J)'$, où θ_j peut être un élément ou un sous-ensemble de θ . Si le modèle est assez complexe, la distribution marginale $p(\theta_j|x)$ sera non standard. Par contre, la distribution conditionnelle $p(\theta_j|\theta_{-j}, x)$ sera souvent standard, où θ_{-j} représente les paramètres du modèle autres que θ_j . Supposons qu'il est possible de simuler des tirages artificiels $\theta_j \sim p(\theta_j|\theta_{-j}, x)$. Si tel est le cas, on peut générer une séquence aléatoire selon l'algorithme suivant :

$$\theta_1^i \sim p(\theta_1|\theta_2^{i-1}, \theta_3^{i-1}, \dots, \theta_J^{i-1}, x),$$

$$\theta_2^i \sim p(\theta_2|\theta_1^{i-1}, \theta_3^{i-1}, \dots, \theta_J^{i-1}, x),$$

$$\theta_3^i \sim p(\theta_3|\theta_1^{i-1}, \theta_2^{i-1}, \dots, \theta_J^{i-1}, x),$$

$$\theta_j^i \sim p(\theta_j|\theta_1^{i-1}, \theta_2^{i-1}, \dots, \theta_{j-1}^{i-1}, x).$$

Notons que cet algorithme décrit une chaîne markovienne de premier ordre, puisque la distribution conditionnelle d'un tirage dépend de la réalisation précédente. De plus, si la densité $p(\theta_j|\theta_{-j}, x)$ est positive pour toutes les valeurs possibles de θ_{-j} , la chaîne est aussi ergodique. Dans ce cas, la chaîne markovienne converge vers sa distribution stable $p(\cdot|x)$. Si la séquence converge avant l'itération n , la statistique $N^{-1} \sum_{i=n+1}^{n+N} g(\theta^i)$ converge presque sûrement vers $E[g(\theta)|x]$.

En principe, les valeurs générées à chaque itération peuvent être utilisées pour estimer $E[g(\theta)|x]$. Mais en pratique, le taux de convergence peut être plutôt lent si la corrélation entre θ_i et θ_{i-1} est trop élevée. Puisque la même valeur de θ_j est retenue pour $J - 1$ itérations, la corrélation entre les valeurs de θ simulées par l'algorithme précédent sera grande. Dans les applications de l'échantillonneur de Gibbs, on utilise des valeurs de θ générées à chaque J itérations - un tour - pour que chaque tirage soit différent.

Difficultés des méthodes MCMC

En contrepartie d'une relative facilité d'implémentation, les méthodes MCMC peuvent présenter des difficultés importantes.

La convergence - c'est-à-dire l'atteinte de l'équilibre de la chaîne de Markov - peut être très lente, surtout lorsque $\pi(\theta|x)$ est difficile à approximer par une densité de transition facilement simulable. ; par ailleurs, une faible maniabilité de l'expression de $\pi(\theta|x)$ rend cette convergence parfois difficile à prouver.

Par ailleurs, même si convergence il y a, détecter le “temps de chauffe” est loin d'être évident. Une importante littérature est ainsi consacrée à la recherche de critères d'arrêt des méthodes MCMC. Voir ainsi les livres de Robert (1998)[13] et Robert et Casella (2004, Chap. 12)[11].

Évaluation de la convergence

Dans une application des techniques MCMC - soit l'échantillonnage de Gibbs ou l'algorithme Metropolis Hastings, la procédure est toujours :

1. Choisir une valeur de départ θ^0 .
2. Simuler n réalisations de la chaîne markovienne avec n assez grande pour atténuer l'effet du choix de θ^0 .
3. Conserver les N valeurs suivantes. Cet échantillon est utilisé pour calculer la statistique $\bar{g} \equiv N^{-1} \sum_{i=n+1}^{n+N} g(\theta^i)$ pour estimer $E[g(\theta)|x]$.

Puisque la chaîne markovienne est ergodique, la valeur de départ n'a aucune importance si n est suffisamment grand. De la même façon, si N est suffisamment grand, la précision de l'estimateur \bar{g} peut atteindre n'importe quel niveau désiré. .

Exemple. Détermination de la valeur centennale du niveau journalier de la mer au Port Pirie (Australie)[28]

La théorie des valeurs extrême est du grand intérêt puisque, une protection qui fonctionne pour des évènements extrême pare aussi des évènements de moindre ampleur. La théorie des valeurs extrêmes considère un n -échantillon iid et s'interroge sur la distribution de la plus grande ou la plus petite valeur de cet échantillon lorsque sa taille tend vers l'infini. Les inondations, avalanches, séismes,...ect, sont des évènements d'autant plus redoutés que leur intensité est grande, on sait bien qu'une valeur extrême est une intensité qui a heureusement peu de chance d'être observée.

1- Position du problème :

Notre étude est basée sur le niveau de mer à Port Pirie (Australie), ces données couvrent la période 1923-1987, pour ces données nous ne disposons que des maxima annuels et nous calibrons un modèle GEV. Soit la série des maxima de nos données Z_1, \dots, Z_k , un processus stochastique à temps discret est divisé en k blocs. La fonction de répartition de l'échantillon des maxima appelée loi généralisée des valeurs extrêmes est donnée par :

$$G(z/\mu, \sigma, \xi) = \exp\left(-\left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-1/\xi}\right), \quad (1.50)$$

avec

$$1 + \xi \frac{z - \mu}{\sigma} > 0, \quad (1.51)$$

et

$$\begin{cases} \mu \in R_+, & \text{un paramètre de centrage;} \\ \sigma \in R, & \text{un paramètre d'échantillon;} \\ \xi \in R, & \text{un paramètre de forme.} \end{cases} \quad (1.52)$$

La fonction de densité de probabilité du modèle GEV s'obtient en différenciant la fonction de répartition en z :

$$g(z/\mu, \sigma, \xi) = \frac{1}{\sigma} \left(1 + \xi \frac{z - \mu}{\sigma}\right)^{-1/\xi - 1} G(z/\mu, \sigma, \xi). \quad (1.53)$$

Le reparamétrage suivant facilite la programmation informatique :
on prend

$$\rho_0 = \sigma^{-1} > 0, \beta_0 = -\rho_0 \xi, \quad (1.54)$$

Le modèle GEV devient :

$$P(z \leq \beta_0, \mu, \rho_0) = \exp[-(1 - \beta_0(z - \mu))^{\rho_0/\beta_0}], \quad (1.55)$$

le modèle GEV est donc caractérisé par le paramètre $\theta = (\beta_0, \mu, \rho_0)$ et la densité de probabilité conjointe s'écrit :

$$g(z/\theta) = \rho_0 (1 - \beta_0(z - \mu))^{\rho_0/\beta_0 - 1} G(z/\theta), \quad (1.56)$$

où $\rho_0 \in R_0^+$, $\beta_0 \in R_0$, $\mu \in R$ et $\beta_0(Z - \mu) < 1$. L'hypothèse iid entraîne la vraisemblance d'un k échantillons de maxima :

$$g(z_1, \dots, z_k/\theta) = \rho_0^k \prod_{i=1}^k \{(1 - \beta_0(z_i - \mu))^{\rho_0/\beta_0 - 1} G(z_i/\theta)\}. \quad (1.57)$$

Pour la priori : on postulera l'indépendance des composantes du vecteur θ : $\beta_0 \perp \mu \perp \rho_0$ et un prior non informatif simple a la forme suivante :

$$\pi(\theta) = \pi(\beta_0)\pi(\mu)\pi(\rho_0) \propto \frac{1}{\rho_0}. \quad (1.58)$$

Le prior arbituel est une distribution gamma dont les paramètres tendent vers zéro.

$$\pi(\rho_0/a, b) \propto \rho_0^{a-1} \exp(-b\rho_0) \xrightarrow{a, b \rightarrow 0} \frac{1}{\rho_0}. \quad (1.59)$$

l'application de la règle de bays donne le posteriori non normalisé :

$$g(\theta/z_1, \dots, z_k) \propto \rho_0^{k-1} \prod_{i=1}^k \{(1 - \beta_0(z_i - \mu))^{\rho_0/\beta_0 - 1} G(z_i/\theta)\}. \quad (1.60)$$

La normalisation par calcul intégral n'est pas possible et aucun conditionnelle n'est standard, l'inférence peut se faire via un algorithme de Metropolis Hasting.

2- Algorithme MH séquentiel appliqué au modèle GEV :

Il sera commande de poser :

$$f(\beta_0, \mu, \sigma_0) = \rho_0^{k-1} \prod \{(1 - \beta_0(z_i - \mu))^{\rho_0/\beta_0 - 1} G(Z_i/\theta)\}, \quad (1.61)$$

puisque $\rho_0 > 0$, l'algorithme de Metropolis-Hasting est mettre en oeuvre avec le changement de paramètres suivant :

$$\phi = \ln \rho_0 \Rightarrow \rho_0 = \exp^\phi, \quad (1.62)$$

la transformation logarithmique donne :

$$\ln f(\beta_0, \mu, \sigma_0) = (k-1)\phi + \left(\frac{e^\phi}{\beta_0} - 1\right) \sum_{i=1}^k \ln[(1 - \beta_0(z_i - \mu))] - \sum_{i=1}^k [1 - \beta_0(z_i - \mu)]^{e^\phi/\beta_0}. \quad (1.63)$$

où :

$$\beta_0(z_i - \mu) < 1, i = 1, \dots, k \quad (1.64)$$

- L'algorithme :

Soit une marche aléatoire réalisée dans R^3 à partir d'un point initial $\theta^0 = (\beta_0^0, \mu^0, \phi^0)$ pour loi instrumentale, nous avons choisi le produit de trois normales unidimensionnelles indépendantes :

$$\beta_0 \sim \mathcal{N}(\beta_0^{i-1}, v_{\beta_0}), \quad (1.65)$$

$$\mu \sim \mathcal{N}(\mu^{i-1}, v_\mu), \quad (1.66)$$

et

$$\phi \sim \mathcal{N}(\phi^{i-1}, v_\phi), \quad (1.67)$$

où les variances instrumentales, v_{β_0} , v_{μ} , v_{ϕ} règlent la force des sauts respectifs. Soit :

$$\theta^{i-1} = (\beta_0^{i-1}, \mu^{i-1}, \phi^{i-1}), \quad (1.68)$$

la valeur de triplet à l'itération $(i - 1)$, à l'itération i : on réalise les trois séquences suivantes :

1.

$$(\beta_0^{i-1}, \mu_0^{i-1}, \phi) \longrightarrow (\beta_0^{i-1}, \mu^{i-1}, \phi^i). \quad (1.69)$$

2. Sous la condition

$$\beta_0(z_t - \mu^{i-1}) < 1, \quad (\beta_0, \mu^{i-1}, \phi^i) \longrightarrow (\beta_0^i, \mu^{i-1}, \phi^i). \quad (1.70)$$

3. Sous la condition

$$\beta_0^i(z_t - \mu) < 1, \quad (\beta_0^i, \mu, \phi^i) \longrightarrow (\beta^i, \mu^i, \phi^i) = \theta^i. \quad (1.71)$$

La rapidité de la convergence dépend du choix du point initial et du choix des variances instrumentales. Notons enfin que disposant des paramètres ϕ et β_0 , il est facile de retourner aux paramètres initiaux σ et ξ .

$$\sigma = \exp^{-\phi}, \quad \xi = -\exp^{-\phi} \beta_0. \quad (1.72)$$

Il est judicieux de choisir un point initial qui favorise une convergence rapide de la marche aléatoire. Pour chaque paramètre, la force des sauts aléatoires est réglée par la variance de la loi normale. Certains auteurs proposent de juger la performance d'un algorithme MH en regardant le taux d'acceptation des candidats, avec une loi instrumentale unidimensionnelle, ce taux devrait être de l'ordre 0,44.

- Les résultats obtenus :

La figure montre le profil des maximum annuel. La variabilité du signal semble stationnaire et il est donc raisonnable de postuler que les maxima sont iid.

Le tableau suivant donne l'estimation des paramètres du modèle GEV :

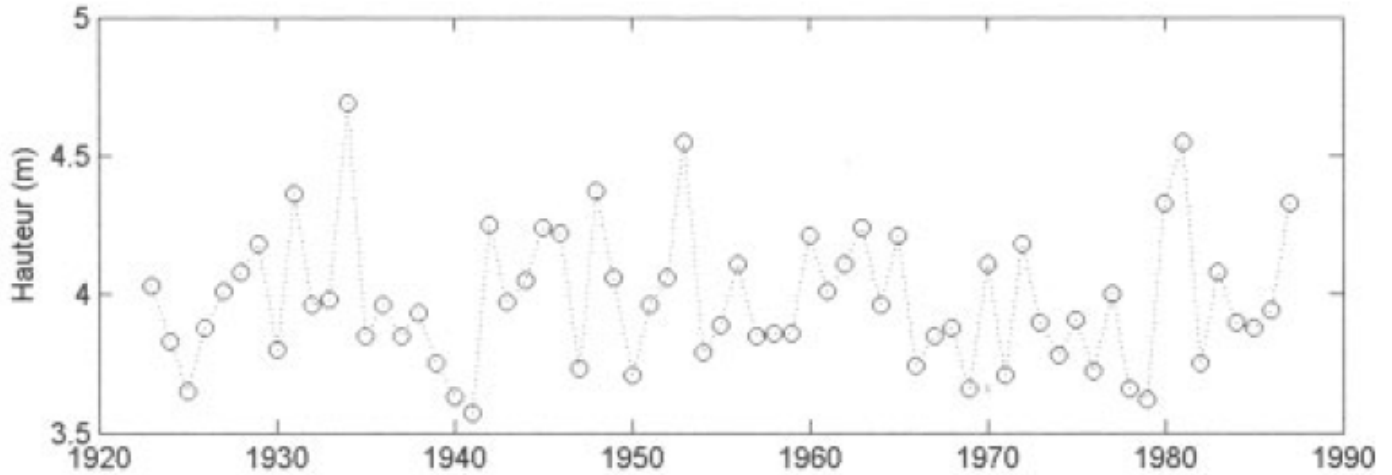


FIG. 1.2 – Maxima annuels des niveaux de retour

$\hat{\sigma}$	$\hat{\xi}$	$\hat{\mu}$	$\hat{Z}_{0,01}$
0,2	-0,05	3,9	4,7

Les valeurs initiaux pour le réglage de la marche aléatoire et le taux d'acceptation sont données dans le tableau suivant :

	$\phi = \ln \rho$	β	μ
valeur initial	1	-0,5	4,5
variance	$(0,23)^2$	$(1,5)^2$	$(0,06)^2$
taux	0,44	0,44	0,45

Après une marche aléatoire de $N = 2000$, le tableau suivant donne les intervalles de crédibilité à 90% en écartant les 500 premières valeurs.

Q	ρ	ξ	μ	$Z_{0,01}$
5	0,17	-0,16	3,83	4,5
50	0,20	-0,03	3,87	4,7
95	0,24	0,15	3,91	5,1

Remarquons que ξ n'est pas significativement différent de zéro ($\xi \rightarrow 0$), et que l'estimation de la valeur centennale du niveau journalier de la mer $\simeq 4,7$, est la médiane de sa distribution marginal a posteriori.

La figure suivante montre le profil des chaînes de Markov :

La figure ci-dessous montre les lois marginales a posteriori de chacun des paramètres.

A l'exception de la conclusion sur ξ . L'application aux données du Port Pirie confirme le résultat de l'estimation et que la distribution suit une loi de Weibul.

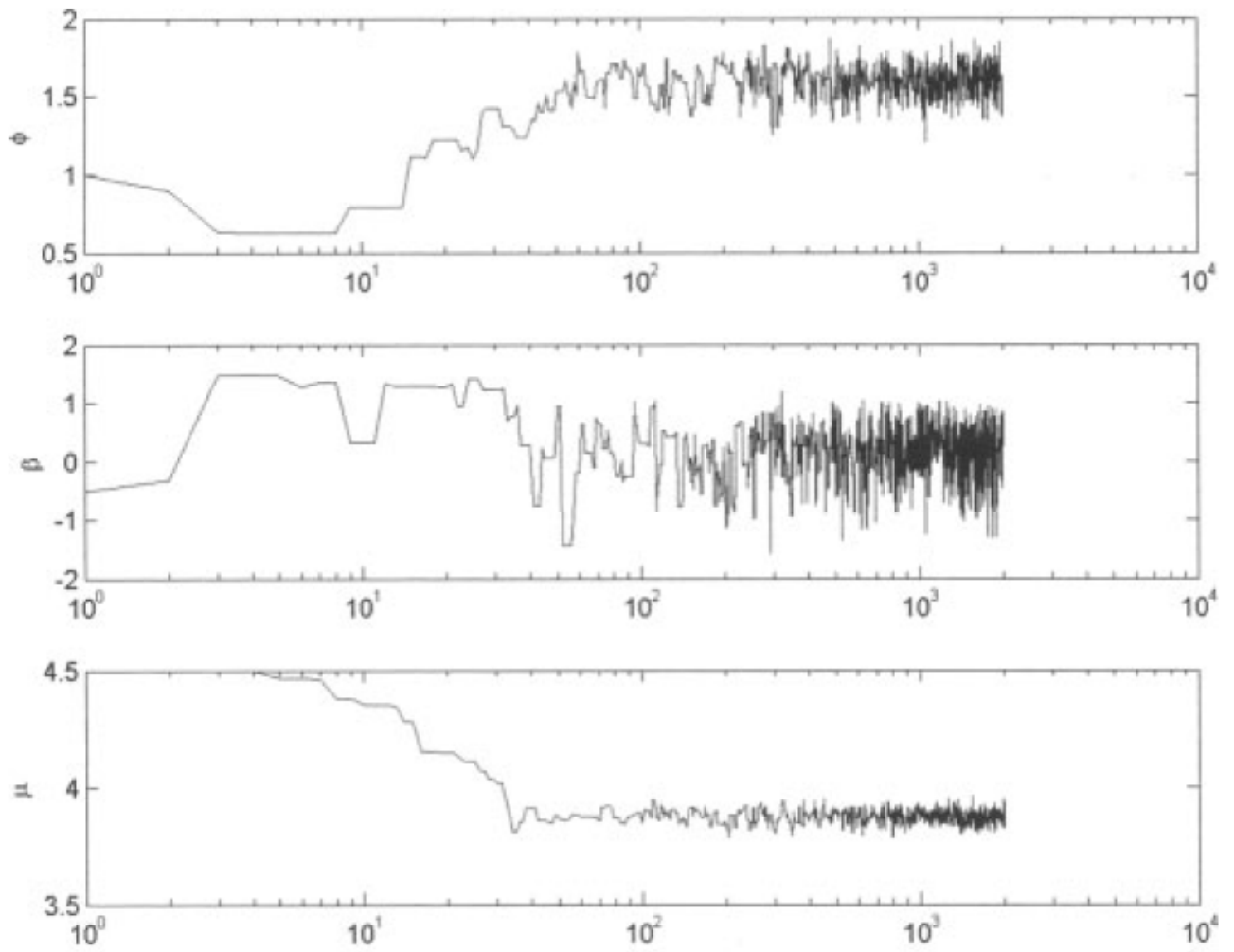


FIG. 1.3 – Port Pirie : Profils des marches aléatoires

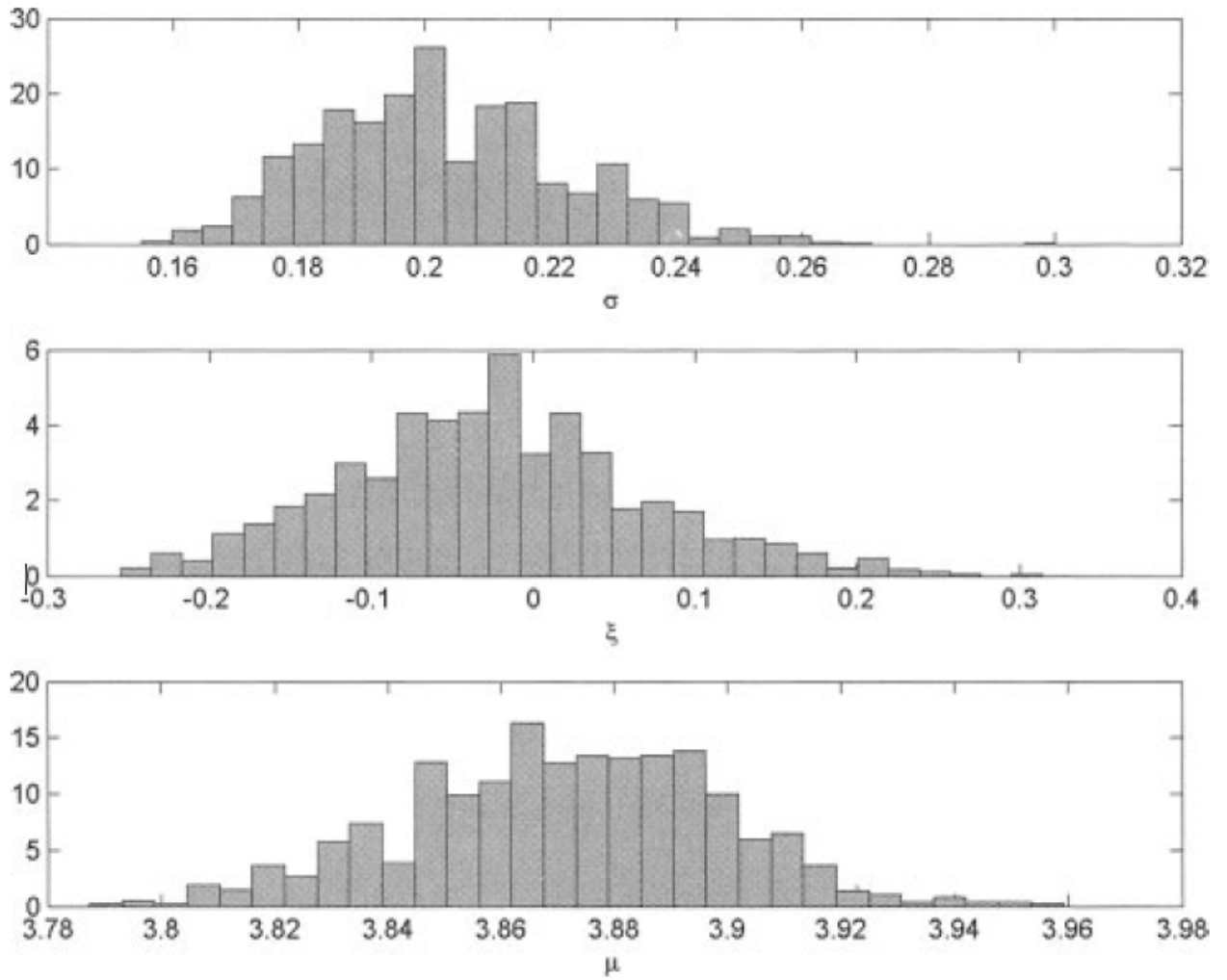


FIG. 1.4 – Port Pirie : Loi marginales a posteriori des paramètres du modèle GEV

Chapitre 2

Processus de Poisson

2.1 Introduction

De nombreux phénomènes aléatoires se manifestent par des “arrivées” survenant une par une à des instants aléatoires successifs.

Exemples :

- Arrivées d’appels à une centrale téléphonique ;
- Impacts de micrométéorites sur un satellite ;
- Passage de véhicules à un péage d’autoroute ;
- Arrivées de clients à un guichet, occurrence d’accidents dans une ville, pannes de machines dans une usine... .

De tels phénomènes peuvent se définir par la famille $(A_n)_{n \in \mathbb{N}}$. des temps d’arrivées qui sont des variables aléatoires. Mais on peut aussi le faire à partir du processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ ou par la famille $(T_n)_{n \in \mathbb{N}}$. des intervalles de temps entre deux arrivées.

N_t est le nombre d’événements apparus jusqu’à l’instant t .

$N_{t+u} - N_u$ est le nombre d'événements apparus entre u et $u + t$.

L'espace des états du processus $(N_t)_{t \in \mathbb{R}_+}$ est $E = \mathbb{N}$ et l'espace des temps est $T = \mathbb{R}_+$. Le processus qui modélise convenablement les exemples cités ci-dessus est le processus de Poisson. On conviendra que $N_0 = 0$.

On note :

- * A_n l'instant de réalisation du $n^{\text{ème}}$ événement ;
- * T_n la durée séparant le $(n - 1)^{\text{ème}}$ événement du $n^{\text{ème}}$ événement pour $n \geq 2$ et $T_1 = A_1$.

On a :

- * $A_n = T_1 + T_2 + \dots + T_n$ pour tout $n \in \mathbb{N}^*$,
- * $T_1 = A_1$ et $T_n = A_n - A_{n-1}$ pour tout $n \geq 2$.

Ainsi, la connaissance de la famille $(A_n)_{n \in \mathbb{N}^*}$ équivaut à celle de la famille $(T_n)_{n \in \mathbb{N}^*}$.

D'autre part, $A_n \leq t$ signifie que le $n^{\text{ème}}$ événement a eu lieu à l'instant t ou avant, c'est-à-dire qu'à l'instant t , au moins n événements ont eu lieu, c'est-à-dire que $N_t \geq n$.

Ainsi

$$F_{A_n}(t) = P([A_n \leq t]) = P([N_t \geq n]) = 1 - \sum_{k=0}^{n-1} P([N_t = k]),$$
$$P([N_t = n]) = P([N_t \geq n]) - P([N_t \geq n + 1]) = P([A_n \leq t]) - P([A_{n+1} \leq t]).$$

Par conséquent, la connaissance de $(N_t)_{t \in \mathbb{R}_+}$ équivaut à celle de la famille $(A_n)_{n \in \mathbb{N}^*}$.

2.2 Définitions et description du processus de Poisson

Soit $(N_t)_{t \in \mathbb{R}_+}$ un processus aléatoire à valeurs dans \mathbb{N} tel que $N_0 = 0$.

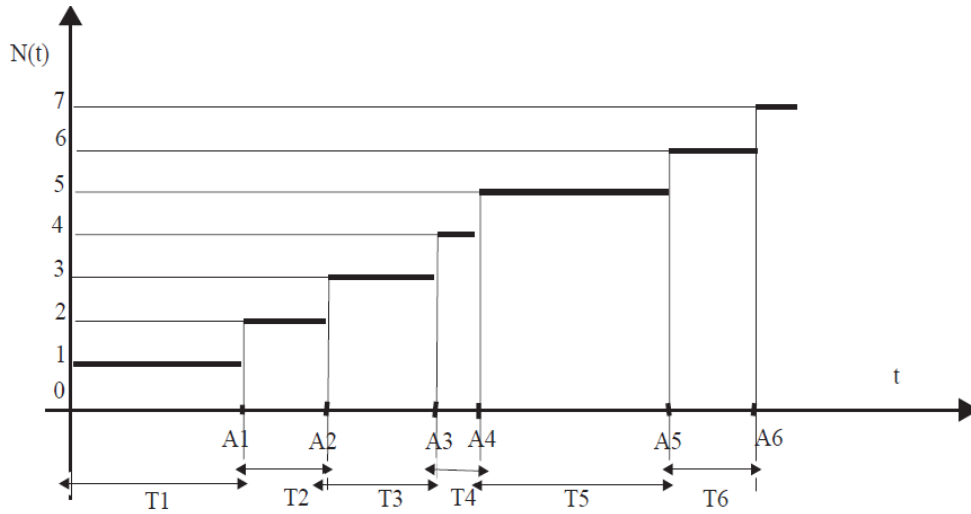


FIG. 2.1 – Processus de renouvellement.

Définition 2.2.1. Le processus $(N_t)_{t \in \mathbb{R}_+}$ est appelé processus de comptage si c'est un processus croissant, c'est-à-dire si pour tout $s \leq t$, alors $N_s \leq N_t$. La variable aléatoire $N_t - N_s$ est alors appelée accroissement du processus sur $]s, t]$.

Définition 2.2.2. Un processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ est appelé processus à accroissements indépendants si pour tout $n \in \mathbb{N}$ et pour tous t_1, \dots, t_n tels que $t_1 < t_2 \dots < t_n$, les accroissements $N_{t_1} - N_0, N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ sont des variables aléatoires indépendantes.

Les processus vérifiant cette hypothèse sont assez nombreux : il semble en effet assez naturel que les arrivées se produisant dans des intervalles disjoints ne soient pas liées entre elles.

Définition 2.2.3. Un processus est dit homogène (ou stationnaire dans le temps), si pour tout s et pour tout t , l'accroissement $N_{t+s} - N_s$ suit la même loi que N_t .

Remarque 2.2.1. Cette propriété est semblable à l'axiome d'homogénéité pour les chaînes

de Markov : seule la durée écoulée entre deux instants (et non pas les deux instants eux-mêmes) compte pour déterminer la loi de l'accroissement.

Définition 2.2.4. Un processus à accroissement indépendant stationnaire $(N_t)_{t \in \mathbb{R}_+}$ est dit à événements rares si $\lim_{h \rightarrow 0_+} P([N_h > 0]) = 0$ et si $\lim_{h \rightarrow 0_+} \frac{P([N_h > 1])}{P([N_h = 1])} = 0$.

Remarque 2.2.2. La deuxième propriété traduit l'improbabilité d'arrivées simultanées. Si on pose $f(t) = P([N_t = 0])$ (et donc $f(0) = 1$), la première hypothèse traduit la continuité de f en 0, car $P([N_h > 0]) = 1 - f(h)$ et donc $\lim_{h \rightarrow 0_+} f(h) = f(0)$.

Définition 2.2.5. Un processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ tel que $N_0 = 0$ est un processus de Poisson si

- $(N_t)_{t \in \mathbb{R}_+}$ est stationnaire,
- $(N_t)_{t \in \mathbb{R}_+}$ est un processus à accroissement indépendant,
- $(N_t)_{t \in \mathbb{R}_+}$ est un processus à événements rares.

Le nom donné au processus de Poisson s'explique par ce qui suit :

Propriété 2.2.1. ([11]) Un processus de comptage $(N_t)_{t \in \mathbb{R}_+}$ tel que $N_0 = 0$ est un processus de Poisson si et seulement si :

- $(N_t)_{t \in \mathbb{R}_+}$ est stationnaire,
- $(N_t)_{t \in \mathbb{R}_+}$ est un processus à accroissements indépendants,
- il existe $\lambda > 0$ tel que, pour tout $t \geq 0$, la variable aléatoire N_t suit la loi de Poisson de paramètre λt .

2.3 Caractérisation d'un processus de Poisson par ses temps d'arrivée

Soit A_n l'instant de la $n^{\text{ième}}$ arrivée : $A_n = \inf\{t \geq 0; N_t = n\}$ et T_n le $n^{\text{ième}}$ temps d'attente pour $n \in \mathbb{N}^*$: $T_n = A_n - A_{n-1}$ (en convenant $A_0 = 0$).

On a $A_n = \sum_{i=1}^n T_i$ et $N_t = \max\{n \geq 0; A_n \leq t\}$.

Théorème 2.3.1. ([15]) $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson de paramètre λ si et seulement si les variables aléatoires T_n sont indépendantes de même loi exponentielle de paramètre λ (de densité $f_{T_n}(t) = \lambda e^{-\lambda t} 1_{]0, +\infty[}(t)$).

Propriété 2.3.1. ([15]) Si $(N_t)_{t \in \mathbb{R}_+}$ est un processus de Poisson de paramètre λ , le temps aléatoire U qui sépare un instant θ du prochain événement et le temps aléatoire V qui sépare θ du dernier événement suivent la loi exponentielle de paramètre λ .

2.4 Propriétés supplémentaires

2.4.1 Décomposition, superposition

On peut décomposer en deux un processus de Poisson suivant un certain critère.

Exemple

Compteur à fonctionnement aléatoire : Des particules arrivant vers un compteur forment un processus de Poisson de paramètre $\lambda = 6$ part/mn. Chaque particule a la probabilité $2/3$ d'être enregistrée par le compteur. Les particules enregistrées forment un processus de Poisson de paramètre $\lambda = 6 \times \frac{2}{3} = 4$ part/mn.

2.5 Processus de Poisson et loi binomiale

Propriété 2.5.1. ([54]) Pour $s \leq t$, la loi conditionnelle de N_s sachant $[N_t = n]$ est la loi binomiale $\mathcal{B}(n, \frac{s}{t})$.

Preuve :

$$P^{[N_t=n]}([N_s = k]) = \frac{P([N_s = k] \cap [N_t = n])}{P([N_t = n])}$$

$$\begin{aligned}
&= \frac{P([N_s = k] \cap [N_t - N_s = n - k])}{P([N_t = n])} \\
&= \frac{P([N_s = k])P([N_t - N_s = n - k])}{P([N_t = n])} \\
&= \frac{P([N_s = k])P([N_t - N_s = n - k])}{P([N_t = n])} \\
&= \frac{e^{-\lambda s} \frac{(\lambda s)^k}{k!} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^{n-k}}{(n-k)!}}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} \\
&= C_n^k \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}.
\end{aligned}$$

2.6 Processus de Poisson et loi uniforme

La loi de (A_1, A_2, \dots, A_n) sachant $[N_t = n]$ est celle de $(U_1^*, U_2^*, \dots, U_n^*)$; où U_i^* est la $i^{\text{ème}}$ plus petite valeur parmi les U_1, U_2, \dots, U_n , qui sont des variables aléatoires indépendantes de même loi uniforme sur $[0, t]$ (en particulier $U_1^* = \min(U_1, \dots, U_n)$ et $U_n^* = \max(U_1, \dots, U_n)$). Cette loi a pour densité f^* définie par :

$$f^*(s_1, s_2, \dots, s_n) = \begin{cases} \frac{n!}{t^n} & \text{si } 0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq t; \\ 0 & \text{sinon.} \end{cases}$$

Ainsi, pour simuler les n premiers temps d'arrivée d'un processus de Poisson sachant que $[N_t = n]$, il suffit de tirer n nombres au hasard entre 0 et 1 (touche "random" de la calculatrice), de les ranger par ordre croissant, puis de les multiplier par t en faisant éventuellement les conversions en format horaire adéquat (heures, minutes, seconde).

2.7 Processus de Poisson composés

Pour un processus de Poisson, la condition C3 entraîne l'impossibilité d'une réalisation simultanée de deux événements ou plus. Si on supprime cette condition, les événements peuvent se produire par "grappes", l'effectif de la grappe étant lui-même aléatoire.

Exemples

- * Arrivées d'avions dans un aéroport : chaque avion transporte un certain nombre de passagers ;
- * Traffic routier : chaque accident engendre un certain nombre de blessés...

Les instants d'occurrence des grappes d'événements sont les A_n d'un processus de Poisson de paramètre λ .

A chaque instant A_n est associée une variable aléatoire Y_n désignant le nombre d'événements se produisant à l'instant A_n .

On suppose les Y_n indépendantes et de même loi.

Z_t est le nombre d'événements apparus jusqu'à l'instant t .

On a la relation fondamentale suivante :

$$Z_t = Y_1 + Y_2 + \dots + Y_{N_t}.$$

Cette relation permet, en particulier, de déterminer la fonction génératrice de Z_t et donc son espérance et sa variance.

Propriété 2.7.1. ([15]) $G_{Z_t}(s) = \exp(\lambda t(G_Y(s) - 1))$; $\mathbb{E}(Z_t) = \lambda t \mathbb{E}(Y)$ et $\text{var}(Z_t) = \lambda t \mathbb{E}(Y^2)$.

Exemple. Le nombre d'accidents par jour dans une ville suit un processus de Poisson de paramètre 2 et le nombre de personnes impliquées suit la loi géométrique de paramètre 1/2 . Quelle est la moyenne et la variance du nombre de personnes accidentées durant une semaine ?

$\mathbb{E}(Y) = 2$ et $var(Y) = 2 = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2$ d'où $\mathbb{E}(Y^2) = 6$ et, pour $t = 7$, $\mathbb{E}(Z_t) = 2 \times 7 \times 2 = 28$ et $var(Z_t) = 2 \times 7 \times 6 = 84$.

Remarque 2.7.1. On a $\mathbb{E}(N_t) = \lambda t$, donc, par exemple, dans le cas de l'avion, le nombre moyen de passagers arrivés jusqu'à l'instant t est égal au nombre moyen $\mathbb{E}(N_t)$ d'avions arrivés jusqu'à t multiplié par le nombre moyen de passagers par avion $\mathbb{E}(Y)$, ce qui paraît évident.

2.8 Processus de Poisson homogène(HPP)

Le processus de Poisson homogène est un processus tel que l'intensité est une constante. Donc on a alors :

$$\lambda \in \mathbb{R}_+, P\{N(t + \tau) - N(t) = k\} = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!}, k = 0, 1, \dots$$

Propriété 2.8.1. ([15]) Le temps du $n^{\text{ème}}$ événement à partir d'un système modélisé par une HPP a une distribution gamma avec le paramètre $\alpha = n$ et $\beta = 1/\lambda$.

Propriété 2.8.2. ([15]) La somme de deux processus de Poisson indépendants de paramètres respectifs $\lambda_1(t)$ et $\lambda_2(t)$ est un processus de Poisson de paramètre $\lambda_1(t) + \lambda_2(t)$.

2.9 Processus de Poisson non-homogène(NHPP)

Un processus de comptage $\{N(t), t \geq 0\}$ est appelé un processus de Poisson non homogène, avec la fonction d'intensité $\lambda(t)$ et $t \geq 0$, s'il satisfait les propriétés suivantes :

- (a) $N(0) = 0$,
- (b) le processus $\{N(t)\}$ a accroissements indépendants,
- (c) $p(N(t, t + h) \geq 2) = 0(h)$,
- (d) $p(N(t, t + h) = 1) = \lambda(t)(h) + 0(h)$.

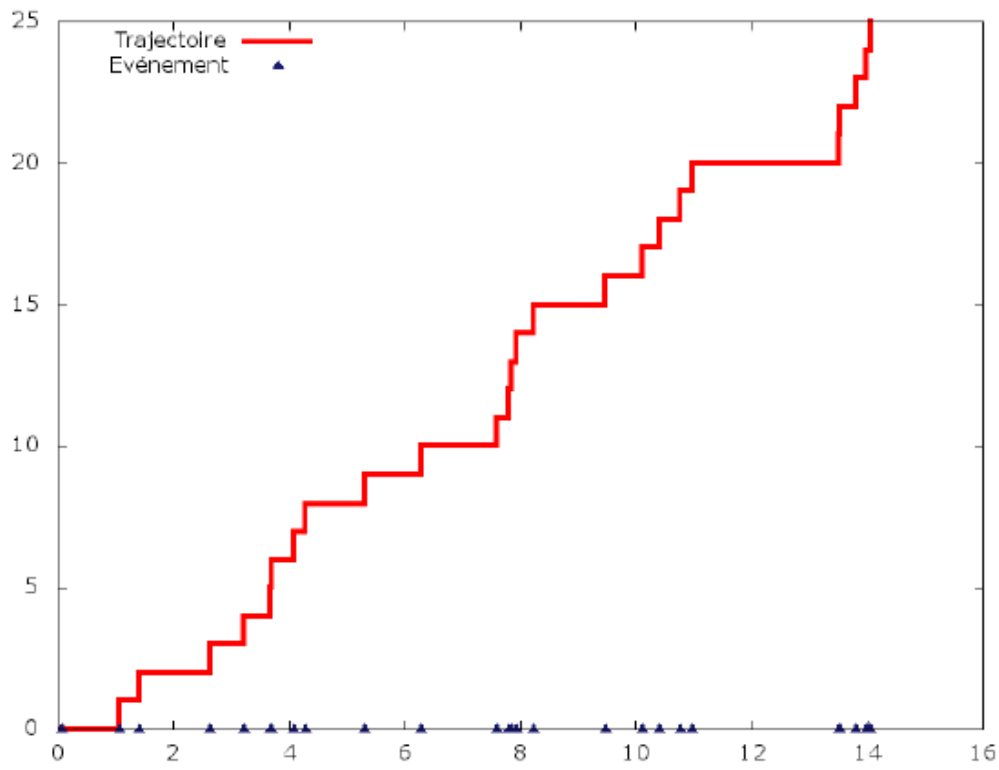


FIG. 2.2 – Simulation d'un processus de Poisson homogène avec $\lambda = 1.8$

Lorsque l'intensité de $\lambda(t)$ est une constante dans le temps, alors on obtient un processus de Poisson homogène (PPH). Si λ est le taux d'intensité d'une PPH et en plus il représente les temps inter-arrivées entre deux événements qui sont iid RV, alors l'PPH est aussi un processus de renouvellement, contrairement à l'autre PPNH dont les temps inter-arrivées ne sont pas iid RV.

La fonction de l'intensité d'un NHPP peut être interprétée comme :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{pr(N(t, t + \Delta t) \geq 1)}{\Delta t}.$$

La valeur moyenne de la fonction du PPNH est définie comme la valeur non décroissante et non négative

$$M(t) = E(N(s, t)) \quad 0 \leq s < t,$$

avec

$$M(t) = E(N_t) \quad t \geq 0.$$

En supposant que $M(t)$ est différentiable, alors $\mu(t) = \frac{dM(t)}{dt}$ est le taux d'occurrence des défaillances (ROCOF) pour le PPNH. La propriété (3) dans la Définition 1 implique que $\mu(t) = \lambda(t)$ de telle sorte que $M(y, s) = \int_Y^s M(t) dt$.

Le PPNH est défini de fait par :

$$P(N(y, t) = k) = \frac{[M(y, s)]^k}{k!} \exp\{-M(y, s)\},$$

pour tout entier k .

il existe deux protocoles d'échantillonnage différentes qui fournissent des données : (i) défaillance tronquée et (ii) Temps tronqué (time truncated case).

Les données sont dites de défaillances tronquées lorsque le test s'arrête après un nombre prédéterminé d'échecs. Supposons que l'on observe jusqu'à n événements qui se produisent (n fixé), si l'on observe les temps de défaillance chronologiques $t_1 < t_2 < \dots < t_n$ est le temps de la $n^{\text{ème}}$ défaillance. On remarque, dans ce cas, le nombre de défaillances est fixé mais le temps, où le test s'arrête, est aléatoire.

Les données sont dites à temps tronquées lorsque le test s'arrête à un temps t prédéterminé. Nous observons une série de temps de défaillances $t_1 < \dots < t_n \leq t$. Dans ce cas, le moment où le test s'arrête est fixé mais le nombre de défaillances n est aléatoire.

Supposons que le système est observé dans un intervalle $(0, y]$ et les défaillances sont enregistrées à des instants $0 < t_1 < \dots < t_n \leq y$. La limite supérieure y pourrait être un point final prédéterminé, ou le moment de la $n^{\text{ème}}$ (prédéterminée) défaillance. Dans le premier cas, il s'agit d'un temps d'une expérience tronquée, tandis que le second il s'agit d'une défaillance tronquée. Finalement, dans les deux expériences la fonction de l'intensité est donnée par : $\lambda(t) = \lambda(t; \theta)$,

et la fonction de vraisemblance :

$$L(\theta, t) = \prod_i^n \lambda(t_i; \theta) \exp\left\{-\int_0^y \lambda(t; \theta) dt\right\}. \quad (2.1)$$

Si on considère un PPH avec un taux λ (taux d'intensité), alors la vraisemblance devient $\lambda^n \exp\{-\lambda y\}$.

2.9.1 Propriétés de processus de Poisson non homogène

Propriété 2.9.1. ([22]) La somme de deux processus de Poisson non homogène indépendants de paramètres respectifs $\lambda_1(t)$ et $\lambda_2(t)$ est un processus de Poisson de paramètre $\lambda_1(t) + \lambda_2(t)$.

Propriété 2.9.2. ([22]) La distribution conjointe, définie aux instants t_1, t_2, \dots, t_n , de PPNH d'intensité $\lambda(t)$ est donnée par :

$$f(t_1, t_2, \dots, t_n) = \prod_i^n \lambda(t_i) \exp\left\{-\int_0^y \lambda(t) dt\right\}, \quad (2.2)$$

où y est le temps d'arrêt : Pour le cas d'événements tronqués on a $y = t_n$, et $y = t$ pour le cas de temps tronqué. Dans un cadre bayésien, les deux expériences conduisent aux mêmes estimations, alors que pour l'approche fréquentiste, certaines différences sont possibles.

Propriété 2.9.3. ([22]) Soient les événements qui suivent processus de Poisson de taux λ , et qui sont indépendants de tout ce qui s'est produit avant (un événement qui se produit à l'instant t est compté avec la probabilité $p(t)$). Le processus des événements comptés constitue un processus de Poisson non homogène avec une intensité $\lambda(t) = \lambda p(t)$.

Pour la simulation, on a considéré le cas suivant :

$$\lambda(t) = \lambda_0 + \alpha e^{-\beta t}.$$

2.9.2 Classes des PPNH

Une classe générale des PPNH est décrite par la fonction d'intensité $\lambda(t, \alpha, \beta) = \alpha g(t, \beta)$, avec $\alpha, \beta > 0$. Cette classe contient des procédés bien connus, tels que le Musa-Okumoto, le Cox-Lewis, les processus de la Loi de Puissance et les processus log-linéaires.

Le processus de Musa-Okumoto

Ce processus a une fonction d'intensité

$$\lambda(t, \alpha, \beta) = \alpha / (t + \beta),$$

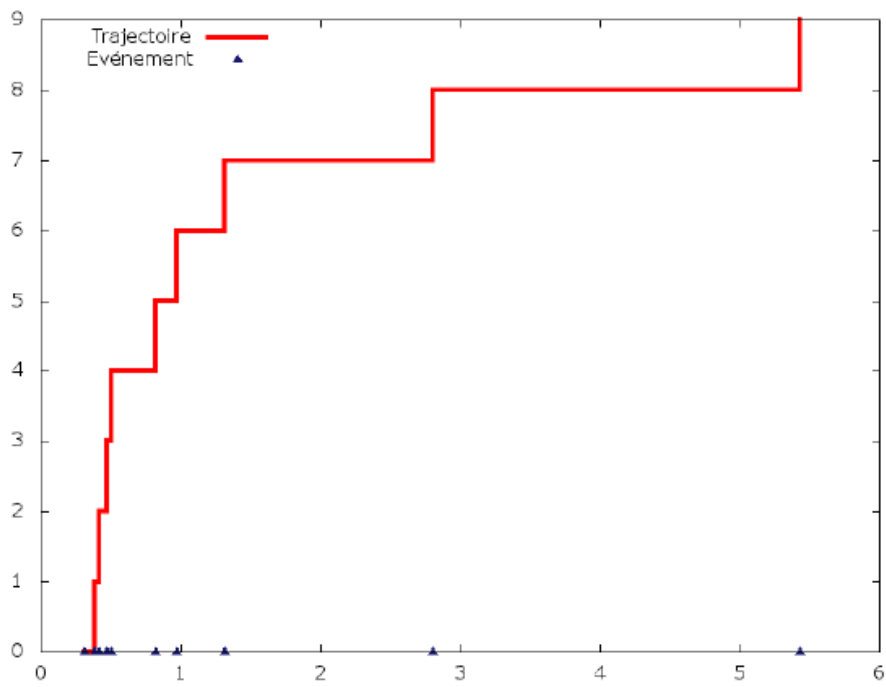


FIG. 2.3 – Simulation d'un processus de Poisson non homogène avec $\lambda_0 = 0.1, \alpha = 4.5, \beta = 0.7$

et sa fonction des accroissements finis est donnée par :

$$\begin{aligned}
 M(t, \alpha, \beta) &= \int x(\alpha/x + \beta)dx \\
 &= \alpha \int (x/x + \beta)dx \\
 &= \alpha \log(t + \beta).
 \end{aligned}$$

Le processus de Cox-Lewis

Le processus de Cox-Lewis a une fonction $\lambda(t, \alpha, \beta) = \alpha \exp\{\beta t\}$ et sa fonction des accroissements finis $M(t, \alpha, \beta)$ est donnée par :

$$\begin{aligned}
 M(t, \alpha, \beta) &= \int x(\alpha \exp\{\beta x\})dx \\
 &= \alpha \int x(\exp\{\beta x\})dx \\
 &= \alpha([t/\beta \exp\{\beta t\}] - \int \frac{1}{\beta} \exp\{\beta x\}dx) \\
 &= \frac{\alpha}{\beta} t \exp\{\beta t\} - \frac{\alpha}{\beta} \exp\{\beta t\}.
 \end{aligned}$$

Donc

$$M(t, \alpha, \beta) = (\alpha/\beta)[\exp\{\beta t\} - 1].$$

Processus de Loi de Puissance

La fonction d'intensité de la Loi de Puissance a la forme $\lambda(t, \alpha, \beta) = \alpha\beta t^{\beta-1}$ et sa fonction des accroissements finis est donnée par :

$$M(t, \alpha, \beta) = \int x(\alpha\beta x^{\beta-1})dx.$$

Donc

$$M(t, \alpha, \beta) = \alpha t^\beta.$$

Un traitement unifié de cette classe de point de vue bayésien est possible. Considérons un PPNH avec une fonction d'intensité $\lambda(t)$ et supposons que nous observons le système à un instant y et à n (nombre des défaillances survenant au moment $t_1 < \dots < t_n$), alors la fonction de vraisemblance est donnée par

$$f(t_1, t_2, \dots, t_n) = \prod_i^n \lambda(t_i) \exp\left\{-\int_0^y \lambda(t) dt\right\}, \quad (2.3)$$

où $I = (t_1, t_2, \dots, t_n)$.

Dans le cas d'un processus de la loi de puissance avec une intensité $\lambda(t) = \alpha\beta t^{\beta-1}$, la fonction de vraisemblance est définie par :

$$\begin{aligned} L(\alpha, \beta, T) &= \prod_i^n \alpha\beta t_i^{\beta-1} \exp\left\{-\int_0^y \alpha\beta t_i^{\beta-1} dt\right\} \\ &= \alpha^n \beta^n \prod_i^n t_i^{\beta-1} \exp\left\{-\int_0^y \alpha\beta t_i^{\beta-1} dt\right\} \\ &= \alpha^n \beta^n \prod_i^n t_i^{\beta-1} \exp\{-\alpha y^\beta\}. \end{aligned} \quad (2.4)$$

Une possible distribution a priori pour α et β est donnée par

- $\alpha \sim \Gamma(a, b)$,
- $\beta \sim Uniform(0, 2)$.

Nous supposons que α et β sont indépendants.

L'application du théorème de Bayes, nous donne :

$$\begin{aligned} \pi(\alpha, \beta \setminus I) &\propto L(\alpha, \beta, T) \pi(\alpha) \pi(\beta), \\ \pi(\alpha, \beta \setminus I) &\propto \alpha \beta \prod_i^n t_i^{\beta-1} \exp\{-\alpha y^\beta\} \alpha^{a-1} \exp^{-b\alpha} I_{(0,2)(\beta)}, \\ \pi(\alpha, \beta \setminus I) &\propto \alpha^{n+a-1} \beta^n \exp\{-\alpha y^\beta + b\} \exp^{\beta \sum_{i=1}^n \log t_i} I_{(0,2)(\beta)}. \end{aligned} \quad (2.5)$$

Processus log-linéaire

Un processus log-linéaire est un PPNH avec une fonction d'intensité

$$\lambda(t) = \exp(\gamma_0 + \gamma_1 t),$$

où les paramètres γ_0 et γ_1 sont définis sur \mathbb{R} . Pour un processus log-linéaire, le nombre moyen de défaillances à l'instant t est :

$$\Lambda(t) = \exp(\gamma_0)[\exp(\gamma_1 t) - 1]/\gamma_1.$$

Chapitre 3

Fiabilité

La théorie de la fiabilité a pour objectif d'étudier l'aptitude de dispositifs techniques (machines, équipements,...), à accomplir une fonction requise, dans des conditions données, durant un temps donné. Actuellement, c'est une discipline à part entière. Prévoir la fiabilité d'un système est essentielle pour des problèmes de sécurité (systèmes de freinage, systèmes nucléaires, systèmes informatiques...), la quasi-impossibilité de réparer certains matériels (satellites), les problèmes économiques (coûts des défaillances, gestion du personnel de maintenance, maintenance des stocks des pièces de rechange),... ce qui rend la connaissance de la fiabilité des systèmes utilisés indispensable.

Les défaillances se produisent généralement de façon aléatoire, il est logique de faire appel aux calculs probabilistes pour l'étude d'un problème de fiabilité. Ainsi, nous définissons la fiabilité d'un dispositif comme étant sa probabilité de fonctionner correctement pendant une durée donnée, ou, ce qui revient au même, la probabilité qu'aucune défaillance ne se produise pendant cette durée.

3.0.3 Rappel sur la théorie de la fiabilité

Fonction de répartition

La fonction de répartition de T est la probabilité que le système tombe en panne avant l'instant t

$$\forall t \geq 0, F(t) = \Pr(T \leq t),$$

où : T est la durée de vie d'un système.

Fonction de fiabilité

La fonction de fiabilité de système à l'instant t est la probabilité pour que ce système fonctionne sur tout l'intervalle $[0, t]$ et on a :

$$\begin{aligned} R(t) &= 1 - F(t) \\ &= \Pr(X > t). \end{aligned}$$

Fonction de structure

La fonction de structure traduit les relations fonctionnelles entre les composants du système et son état de panne / fonctionnement.

Soit $X = (X_1, X_2, \dots, X_n) \in \{0, 1\}^n$ le vecteur décrivant conjointement les états des composants. On définit une fonction $\Phi(X)$ décrivant l'état du système à valeurs dans $\{0, 1\}$ par :

$$\Phi(X) = \begin{cases} 1 & \text{si le système est en bon état.} \\ 0 & \text{si le système est en panne.} \end{cases}$$

où :

$$X_i = \begin{cases} 1 & \text{si le composant } i \text{ est en bon état.} \\ 0 & \text{si le composant } i \text{ est en panne.} \end{cases}$$

Définition 3.0.1. La variable aléatoire T , généralement absolument continue, représente la durée de vie d'un dispositif (ou la durée de bon fonctionnement jusqu'à sa première panne).

- Fiabilité à l'instant t : $R(t) = P([T > t])$.
- Fonction de répartition de T : $F(t) = P([T \geq t])$.
- Densité de défaillance : f , densité de la loi de T .
- Taux de défaillance instantané : $\lambda(t) = \lim_{h \rightarrow 0} P^{[T > t]}([t < T < t + h])$.

Proposition 3.0.1. ([9]) *L'une de ces données suffit à caractériser la loi de T .*

Preuve : On a

$$\begin{aligned} R(t) &= 1 - F(t), \\ F(t) &= \int_0^t f(u) du, \\ R(t) &= \int_t^{+\infty} f(u) du, \\ f(t) &= F'(t) = -R'(t). \end{aligned}$$

De plus, $\lambda(t) = \frac{F'(t)}{1-F(t)} = -\frac{R'(t)}{R(t)}$ et $R(t) = \exp \left[\int_0^t \lambda(u) du \right]$.

Lois utilisées

- Loi exponentielle : $f(t) = \lambda e^{-\lambda t} \mathbf{1}_{]0, +\infty[}(t)$, $\tau = \mathbb{E}(T) = \frac{1}{\lambda}$, $\lambda(t) = \lambda$ (taux de défaillance constant, caractérise les dispositifs sans usure).
- Loi de Weibull : $f(t) = \lambda \beta (t - t_0)^{\beta-1} e^{-\lambda(t-t_0)^\beta}$, $\mathbb{E}(T) = \Gamma(1 + \beta^{-1}(\lambda(t - t_0))^\beta)^{-1}$.
- Loi gamma : $\gamma(\lambda, a)$, $f(t) = \frac{\lambda^a}{\Gamma(a)} t^{a-1} e^{-\lambda t} \mathbf{1}_{]0, +\infty[}(t)$, $\mathbb{E}(T) = \frac{a}{\lambda}$.
- Loi normale : $\mathcal{N}(m, \sigma^2)$, $f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-m)^2}{2\sigma^2}}$, $\mathbb{E}(T) = m$.
- Loi log-normale : $f(t) = \frac{1}{t\sigma\sqrt{2\pi}} e^{-\frac{(\ln t - m)^2}{2\sigma^2}} \mathbf{1}_{]0, +\infty[}(t)$, $\mathbb{E}(T) = e^{m + \frac{\sigma^2}{2}}$.

3.1 Fiabilité des systèmes réparables :

Dans cette section, nous allons nous concentrer sur les processus de Poisson non homogènes en illustrant quelques-unes de ses propriétés et de ses applications. En particulier, nous donnons un aperçu sur les liens existants en fiabilité, avec une illustration rapide des notions mathématiques.

3.1.1 Systèmes réparables

Les systèmes sujets à des pannes peuvent être divisés en deux catégories principales : Les systèmes irréparables et les systèmes réparables. Lorsqu'une pompe hydraulique dans une centrale électrique tombe en panne, elle sera remplacée (la pompe hydraulique est un système non réparable), tandis que la centrale électrique est un système réparable.

Une fois qu'un système subit une panne, différentes stratégies de réparation ont des influences sur sa fiabilité, généralement définie comme la probabilité d'aucune défaillance dans des intervalles de temps. En réparation "parfaite", le système est ramené à sa fiabilité initiale, alors qu'une réparation minimale instantanée d'un petit composant restaure la même fiabilité du système juste avant la panne. L'ancienne stratégie correspond à un état communément appelé "bon comme neuf" ou "même comme neuf", tandis que la seconde correspond à la "mauvais-comme-vieux" ou "de même que -vieux" état. Les défaillances du système dans les deux stratégies sont généralement modélisées par des processus de renouvellement et NHPP. respectivement.

La fiabilité peut être constante au fil du temps, comme il arrive dans le processus de Poisson homogène (HPP), le seul NHPP qui est un processus de renouvellement aussi. Les systèmes sont soumis à la décroissance ou la croissance de la fiabilité (avec constance comme un cas particulier des deux); dans leur durée de vie, ils peuvent vivre l'une ou l'autre ou les alterner à certains points de changement. La détection séquentielle de problèmes dans le logiciel, sans introduction de nouveaux, implique une croissance de fiabilité partout

dans la phase de test. En revanche, il existe des systèmes soumis à de nombreux échecs précoces, puis une diminution des échecs est suivie d'une longue période de défaillances rares et par une dernière période avec un nombre croissant d'échecs. Après la phase initiale de "fiabilisation par rodage" suivie d'une croissance de fiabilité, ces systèmes connaissent une fiabilité constante ("vie utile") suivie d'une phase finale de décroissance de fiabilité, le terme "baignoire" est utilisé pour décrire ce comportement en raison de la forme de la fonction d'intensité de la NHPP correspondant.

Les deux stratégies de réparation illustrées avant sont deux stratégies opposées et extrêmes. Dans la pratique, les réparations augmentent habituellement la fiabilité du système par rapport à celle qu'elle avait juste avant la panne, mais elles ne le ramènent pas à sa valeur initiale. En outre, les réparations ne sont pas, en général, instantanées et l'indisponibilité doit être modélisée ainsi.

Enfin, il convient de mentionner que les systèmes complexes peuvent être divisés en composants et leur fiabilité pourrait être le résultat de la superposition de la fiabilité de leurs composants.

Analyse statistique de la simple NHPP

Considérons un NHPP avec une fonction d'intensité $\lambda(t; \theta)$. Supposons que nous observons le système à un temps y . Soit n le nombre d'échecs qu'a eu lieu aux instants $t_1 < t_2 \dots < t_n$; puis la fonction de vraisemblance est donnée par :

$$L(\theta; t) = \prod_{i=1}^n \lambda(t_i) \exp\left(-\int_0^y \lambda(t) dt\right), \quad (3.1)$$

où $t = (t_1, \dots, t_n)$.

Au moins deux expériences sont possibles : L'observation du système à un instant donné, ou jusqu'à ce que la $n^{\text{ème}}$ défaillance se produit. Le premier cas est appelé troncature de

temps, tandis que le second est appelé troncature d'échec, avec $y = t_n$. Exceptionnellement différent, les deux expériences conduisent aux mêmes estimations dans un cadre bayésien, alors que certaines différences sont possibles selon une approche fréquentiste, telle que dans l'estimation du maximum de vraisemblance des paramètres du processus de la lois de puissance.

Mesures de fiabilité

L'estimation des paramètres n'est pas, sans doute, l'objectif le plus important d'une analyse statistique de la fiabilité. L'intérêt majeur d'une étude de fiabilité statistique repose sur la possibilité de prévoir les futures défaillances, que ce soit sur le système sous observation ou un nouveau système. Un exemple de la quantité d'intérêt est la fiabilité du système, qui est défini comme $R(y, s) = P\{N(y, s) = 0\}$ pour le système observé jusqu'à temps y ou $R(s) = P\{N(s) = 0\}$ pour un nouveau système. Pour un PLP, la fiabilité du système devient $R(y, s) = \exp\{-Ms^\beta + My^\beta\}$ et $R(s) = \exp\{-Ms^\beta\}$, respectivement. Une quantité connexe est la prédiction du nombre de défaillances dans un certain intervalle de temps.

Une autre quantité d'intérêt est le nombre prévu de défaillances dans des intervalles futurs, soit $E[N(y, s)]$ pour le même système ou $E[N(s)]$ pour un nouveau système. Pour un PLP, le nombre attendu de défaillances dans des intervalles futurs est donné par $E[N(y; s)] = M[s^\beta - y^\beta]$ et $E[N(s)] = Ms^\beta$, respectivement.

Enfin, l'estimation de la fonction d'intensité à y est importante. NHPP sont parfois utilisés pour décrire les défaillances dans les tests d'un prototype. Un nouveau produit est testé avant d'être commercialisé. La réalisation de fiabilité satisfaisant, le coût excessif des essais supplémentaires, le risque d'obsolescence du produit sont quelques causes qui impliquent l'interruption de l'essai au moment y . Une fois que le produit est commercialisé, aucun autre essai n'est autorisé pour l'améliorer et sa future fiabilité est la même qu'il avait

eu au temps y .

3.2 Points de changement dans les NHPP

Cette partie, décrit un système dont le comportement change à un instant donné. On considère deux types différents de modèles de point de changement. Dans le premier, on considère des modèles permettant des changements dans le niveau de fiabilité après chaque échec (comme le système est réparé et mis en fonctionnement), par exemple : la fiabilité du logiciel. Dans le second, on considère un modèle permettant des changements aux points aléatoires dans le temps, en raison de panne d'un composant sans provoquer l'échec du système ou en raison d'interventions par équipe de maintenance à points inconnus dans le temps.

Fabrizio Ruggeri et Siva Sivaganesan (2005)[21] considèrent les points de changement dans le processus de la lois de puissance juste après chaque échec (à l'instant t_i^+, s), modifiant la valeur de β . Les variations de M pourraient être semblables mais encombrante. On note par β_i la valeur du paramètre à l'instant $t_i^+, i = 1, \dots, n$, juste après un échec, identifiant le processus $(t_i, t_{i+1}]$, où β_0 est la valeur du paramètre sur $(t_0, t_1]$. Ici, nous prenons $t_0 = 0$ et $t_{n+1} = y$, c'est-à-dire les extrémités de l'intervalle d'observation.

Fabrizio Ruggeri et Siva Sivaganesan (2005)[21] considèrent à la fois un modèle hiérarchique et un autre dynamique. Dans le premier modèle, ils suppose de fait que (ϕ, σ^2) et le β_i, s sont iid avec un $\mathcal{LN}(\phi, \sigma^2); i = 0, \dots, n$ de distribution log-normale. Dans le dernier modèle, ϕ et σ^2 suivent respectivement, une distribution a priori normal $\mathcal{N}(\mu, \tau^2)$ et une distribution gamma inverse a priori $\mathcal{IG}(\rho, \gamma)$.

La vraisemblance devient

$$M^n \prod_{i=1}^n \beta_{i-1} t_i^{\beta_{i-1}-1} \exp -M \sum_{i=1}^{n+1} (t_i^{\beta_{i-1}-1} - t_{i-1}^{\beta_{i-1}}). \quad (3.2)$$

En supposant que la distribution a priori de M est une Gamma, de ce fait, on obtient les distributions conditionnelles a posteriori de M , σ^2 et ϕ qui suivent respectivement une distribution Gamma, une distribution Gamma inverse et une distribution normale. A l'exception de la constante de normalisation, les distributions conditionnelles des β_i , s sont connues. C'est une situation typique dans laquelle les distributions conditionnelles complètes peuvent être utilisées pour simuler la distribution conjointe a posteriori par l'algorithme Metropolis-Hastings et l'échantillonneur de Gibbs.

Dans [21], les auteurs considèrent un modèle dynamique, pour le paramètre β_{i-1} qui se modifier aux instants t_i^+ , $i = 1, \dots, n$, défini par :

$$\log \beta_i = \log(a) + \log(\beta_{i-1}) + \epsilon_i,$$

où a une constante positive et ϵ_i une variable aléatoire iid issue d'une loi normale d'une moyenne de 0 et de variance σ^2 (bruit blanc). Des choix appropriés des probabilités a priori conduisent à la simulation de la distribution conjointe a posteriori par l'échantillonneur de Gibbs avec des pas de Metropolis.

3.2.1 Défaillances dans les métros : Fiabilité de la porte du train

L'exemple suivant présente, une estimation des paramètres du modèle de fiabilité, modélisé par un processus de loi de puissance, basée sur les méthodes de simulation MCMC.

Nous considérons que les données de défaillance de 40 rames de métro, qui ont été livrés à une entreprise européenne de transport entre Novembre 1989 et mars 1991 et tous ont été mis en service à partir du 20 Mars 1990 au 20 Juillet 1992. Surveillance de défaillance

terminée le 31 Décembre 1998. Quand une panne a eu lieu, à la fois la lecture de l'odomètre et la date de l'échec ont été enregistrés, avec le code du composant défectueux.

La recherche initiale visait à analyser la fiabilité de la porte avant l'expiration de la garantie. On examine tout les échecs, sans faire de distinction entre les différents modes de défaillance. Dans cette étude, nous réalisons d'abord une analyse exploratoire des données afin d'identifier les principaux modes de défaillance parmi les sept modes classifiés. Dans notre cas, on se limite à deux types de défaillances : défaillances dues aux commandes d'ouverture des portes et de pièces mécaniques.

Les portes, comme d'autres composants de train, doivent être construites suivant certaines exigences techniques, en assurant un niveau minimum de fiabilité, c'est-à-dire une contrainte sur le nombre maximum d'échecs attendus dans des intervalles prescrits (en jours) et d'espace (en kilomètres).

Le but est de trouver des modèles capables de décrire l'historique des échecs et de prédire le nombre de défaillances dans des intervalles de temps futurs.

Les portes de train peuvent être considérées comme des systèmes réparables complexes, soumis à des réparations partielles en cas des échecs. En effet, les portes en général ne sont pas remplacées, quand une panne surviennent. L'intervention de réparation s'effectue sur le composant défectueux : soit de le réparer ou de le remplacer. Par conséquent, les interventions après les échecs n'ont pas d'incidence significative sur le comportement et la fiabilité de la porte. C'est la situation typique décrite dans la fiabilité " aussi mauvais que vieux ", puisque les réparations sont partielles alors ce qui ramène la fiabilité du système à son état juste avant la panne.

Nous avons considéré un modèle hiérarchique afin de relier les kilomètres et les jours ; ici nous considérons une fonction d'intensité bidimensionnelle pour le processus de Poisson.

Les tracés de kilomètres par rapport aux jours pour chaque échec montrent une relation plutôt linéaire entre eux et cela se reflète dans la fonction d'intensité bidimensionnelle. Dans le même document, les données combinées sur tous les modes de défaillance ont montré une périodicité de 12 mois ; ici nous introduisons une composante périodique dans la fonction d'intensité ainsi, basée sur les résultats de périodogrammes. En outre, la concavité de la courbe de nombre cumulé de défaillances par rapport au temps induit de considérer un élément de base décroissant avec le temps en fonction de l'intensité.

Les distributions a posteriori des paramètres des modèles divertis ont été obtenus via des simulations MCMC et ils ont été utilisés principalement pour faire les prévisions sur le nombre de défaillances futures.

3.2.2 Les données

Dans notre projet de consultation et de motivation, nous avons analysé les données de défaillance de 40 rames de métro, qui ont été livrés à une entreprise de transport entre Novembre 1989 et Mars 1991 et tous ont été mis en service à partir du 6 Avril 1990 au 20 Juillet 1992. La surveillance des pannes terminée le 31 Décembre 1998. Quand une panne a eu lieu, à la fois la lecture de l'odomètre et la date de l'échec ont été enregistrées. Les données ont été recueillies dans le but d'analyser les composants d'une rame de métro étaient les plus influentes sur le coût du cycle de vie de l'ensemble des trains. Les quatre composantes les plus pertinentes ont été identifiées : portes, roues de moteur et deux convertisseurs. Roues de moteur sont des composants mécaniques dont les échecs sont principalement dues à deux composants, alors que les convertisseurs sont constitués d'un grand nombre de composants, principalement électroniques, sous réserve des échecs. Nous avons décidé de mettre l'accent sur les échecs des portes d'ouverture, car ils sont les composants qui ne sont plus soumis à des échecs. La prévention de ces échecs et la réduction de leur coût sont importantes pour la société. Par conséquent, il est important que les portes sont fiables au moins comme indiqué par le fabricant dans le contrat signé avec la compagnie de transport. Les écarts

dans la fiabilité des composants sont pris en charge par une période de garantie dont des défaillances structurelles des composants sont facturés au fabricant qui doit soit payer une indemnité à la compagnie de transport ou de régler le problème par l'amélioration des interventions. Par conséquent, la compagnie de transport est intéressé à vérifier la fiabilité des composants avant leur garantie arrive à échéance et le présent document vise à fournir un outil pour décrire les échecs et faire des prévisions.

3.2.3 Analyse exploratoire des données

Dans la Figure 1, nous montrons le nombre cumulé de défaillances de tous les trains en fonction du temps (exprimée en jours) ; les astérisques indiquent quand chaque train a été mis en service. Un comportement transitoire a lieu pendant les 500 premiers jours, quand 35 trains sont mettre en service, après cette période, le nombre cumulé de défaillances, qui est maintenant calculé sur un nombre sensiblement constant de trains, prend peu à peu une concave façonner.

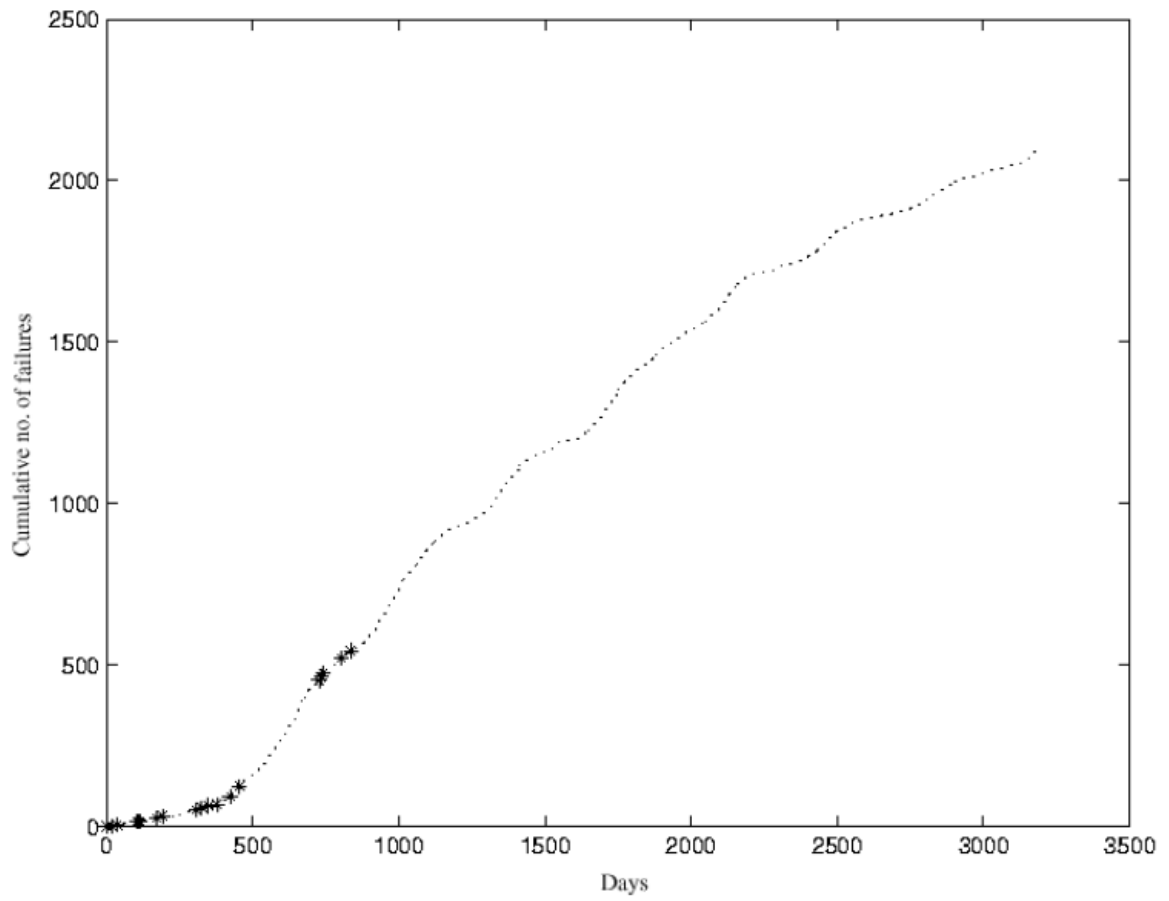


FIG. 3.1 – Nombre cumulés de défaillances par rapport aux jours pour tous les trains.

Oscillations périodiques peuvent également être remarqué dans la courbe, comme l'a confirmé par la Figure. 2, qui montre une tendance saisonnière dans l'apparition de défaillances.

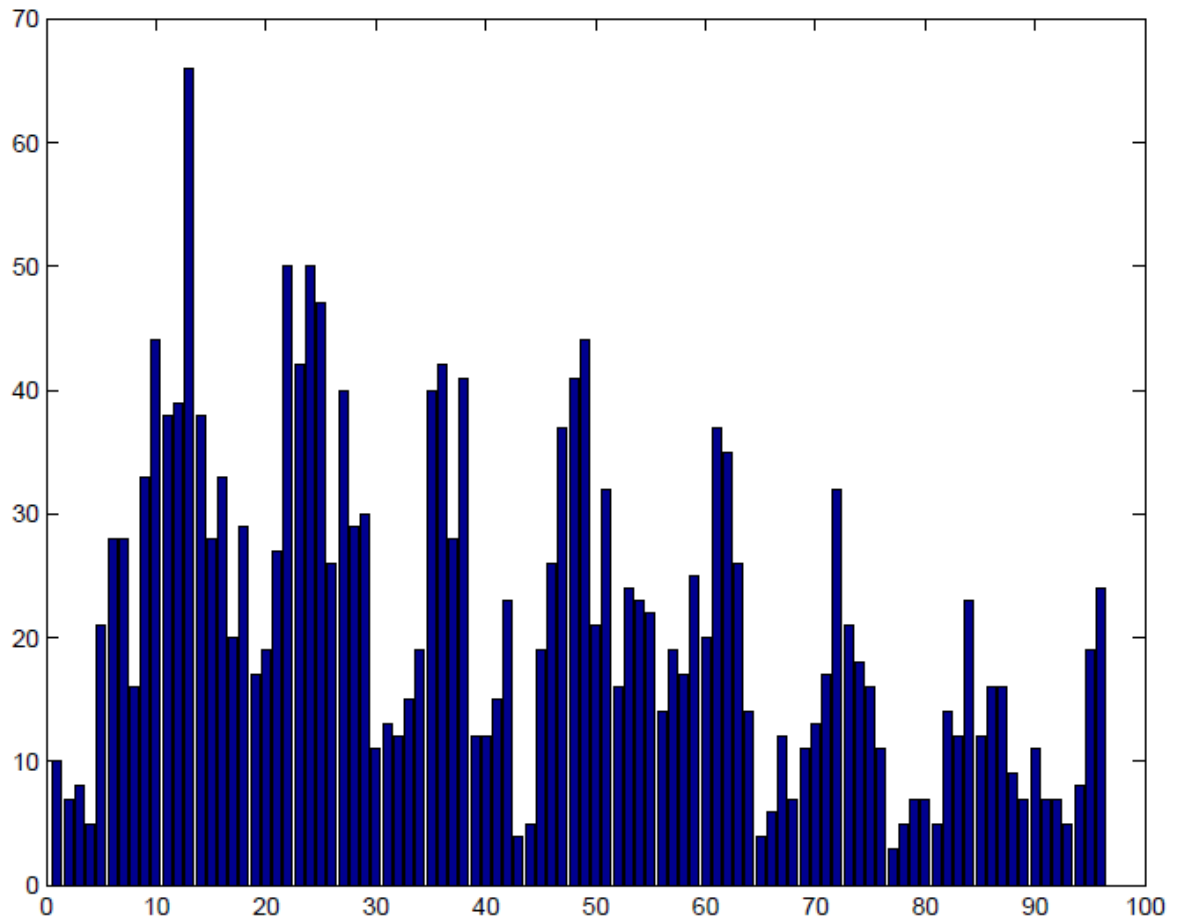


FIG. 3.2 – Pas Mensuel d'échecs pour les 40 trains à partir Janvier 1991.

Ces résultats nous ont conduit à envisager le développement en série de Fourier de la série temporelle du nombre mensuel de défaillances. Depuis l'ensemble de la période a 104 mois et nous sommes intéressés à identifier les cycles saisonniers possibles, les huit premiers mois sont jetés afin que les données sont considérées comme plus exactement huit ans. L'harmonique de la j de l'expansion de Fourier, $j = 1, \dots, 48$, est alors donnée par

$$\left\{ a_j \cos\left(\frac{2\pi j}{96}i\right) + b_j \sin\left(\frac{2\pi j}{96}i\right) \right\}_i^{96},$$

où a_j et b_j sont les coefficients de Fourier. La période de l'harmonique $j^{\text{ième}}$ est de $96/j$, et de composants saisonniers sont identifiés par des valeurs entières de j . Par exemple, lorsque $j = 8$, la période est de 12, correspondant à un cycle d'un an. Le périodogramme $\{a_j^2 + b_j^2\}_j$ est tracée sur la Figure 3. Ici, la seule fréquence qui est nettement prédominante est celle correspondant à la huitième harmonique (correspondant à une période de 12 mois), l'exclusion de la première harmonique qui représente la tendance.

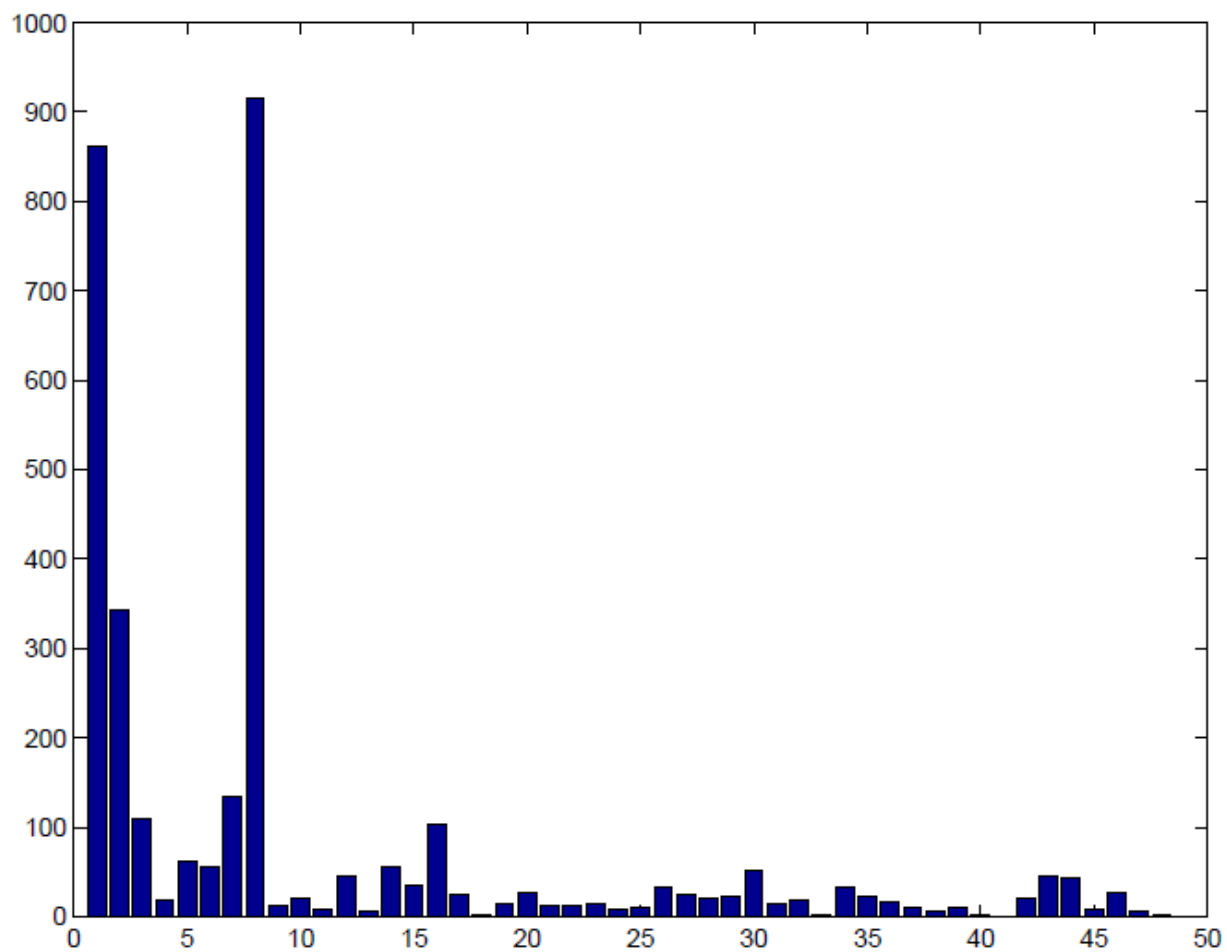


FIG. 3.3 – Périodogramme de la série temporelle du nombre mensuel d'échecs de 1991 à 1998

En ce qui concerne la courbe du nombre cumulé de défaillances en km pour tous les trains, une représentation plus lisible exigerait d'avoir un nombre constant d'trains. Ainsi, nous écartons les huit trains qui ont couru moins de 500 000 km . Maintenant 507 291 km , est le minimum de la dernière lecture des odomètres de reste dès 32 trains.

3.2.4 Modèle pour l'apparition de défaillances

Nous introduisons un modèle pour les pannes du système d'ouverture de porte d'un seul train, qui est composé de trois wagons. Les réparations du système considéré dans son ensemble consistent la substitution de petites pièces. Ensuite, nous sommes dans une situation de réparation minimale et un modèle de processus de Poisson semble approprié.

La distance parcourue est évidemment le meilleur moyen de mesurer le vieillissement d'un train, mais le temps doit aussi être considéré pour tenir compte de la saisonnalité dans les données. Parce que la distance parcourue est essentiellement une fonction du temps, laissant t désigne le temps écoulé en jours et $g(t)$ la distance parcourue au temps t , nous pouvons intégrer les deux échelles de mesure à l'intérieur d'une fonction d'intensité qui dépend de t seulement :

$$\lambda(t; \theta) = \mu(g(t); \theta_1) s(t; \theta_2), t \geq 0; \theta = (\theta_1, \theta_2). \quad (3.3)$$

Ici, $\mu(\cdot; \theta_1)$) représente la fonction de l'intensité de base du procédé, et

$$s(t; \theta_2) = \exp\{\rho \cos(2\pi\omega t + \varphi)\}, \theta_2 = \rho \geq 0.$$

Où seule l'amplitude ρ est inconnue. La fréquence est $\omega = 1/365$ (one-year cycles), et la phase φ dépend de la date de départ de chaque train, et peut être considéré comme $2\pi\omega t^*$, où t^* est le nombre de jours à partir du 1er Janvier de l'année au cours de laquelle le train a commencé son service depuis la date de départ lui-même. Les plus hauts sommets de la série chronologique des comptes mensuels observés se produisent en fait autour de Janvier.

Les analyses préliminaires nous amènent à renoncer aux processus de Poisson qui sont couramment utilisés dans la fiabilité (par exemple, le processus de la loi de puissance), donc nous considérons

$$\mu(k; \theta_1) = b_0 \frac{\log(1 + b_1 k)}{1 + b_1 k}, \theta_1 = (b_0, b_1), b_0, b_1 \geq 0, \quad (3.4)$$

une fonction qui augmente la hauteur de son point maximum à $(e - 1)/b_1$. Puis passons à zéro lorsque $k \rightarrow \infty$. Cela représente une situation où les premiers échecs ont lieu, et, peut-être après certaine réparation d'amélioration, la fiabilité améliore modérément. En outre, la fonction de valeur moyenne correspondante, $b_0 \log^2(1 + b_1 t)/(2b_1)$, a une forme qui ressemble au nombre cumulé des échecs sous enquête.

Pour un train contrôlé dans un intervalle de temps $[0, T]$ qui a échoué à des moments $(t_1, \dots, t_n) = t$, et des lectures de l'odomètre $(k_1, \dots, k_n) = k$, la fonction de vraisemblance est alors :

$$L(\theta, k, t) = \prod_{i=1}^n \mu(g(t_i), \theta_1) s(t_i, \theta_2) \exp\left[-\int_0^T \mu(g(t), \theta_1) s(t, \theta_2) dt\right]. \quad (3.5)$$

3.2.5 Estimation bayésienne non paramétrique

L'estimation de régression paramétrique de la relation entre le temps et les kilomètres terme ne s'est pas avérée satisfaisante, puisque les changements imprévisibles apparaissent dans le modèle d'utilisation de plusieurs trains. Par conséquent, la modélisation g comme un processus stochastique non négatif et non décroissant inaperçu semble approprié. Si g suit un processus de gamma avec (positif) paramètres a et b , puis il a accroissements indépendants et $g(t)$ suit une densité gamma avec la forme à l'échelle et b . L'hypothèse d'incrément indépendant est raisonnable, parce qu'un train a un horaire quotidien fixe qu'il devrait suivre. L'hypothèse de distribution implique que la lecture prévue de l'odomètre au moment t est at/b , avec une déviation standard \sqrt{at}/b . Cela prend en compte la linéarité

essentielle de la relation entre le temps et les kilomètres, et l'incertitude qui augmente à mesure que l'on regarde plus loin dans l'avenir.

Le modèle peut être résumé par la structure hiérarchique suivante :

$$g(t) \sim G(at, b),$$

$$\theta \sim \pi(\theta),$$

$$[t|g, \theta] = NHPP\mu(g(t); \theta_1)s(t; \theta_2),$$

$$[k|t, g] = \prod_{i=1}^n \delta_{g(t_i)}(\cdot).$$

où $\pi(\theta)$ est une densité a priori spécifiée sur θ , les crochets sur le stand de gauche pour "distribution de probabilité, et $\delta_{g(t_i)}(\cdot)$ (\cdot) est la masse point $g(t_i)$. NHPP. signifie "processus de Poisson non homogène" avec sa fonction d'intensité entre accolades.

Pour θ nous prenons des distributions a priori indépendantes pour chacun de ses éléments. Si la probabilité a priori sur b_0 est impropre et ceux sur b_1 et ρ ont un support borné supérieur, alors il est facile de montrer que la conjointe a posteriori est propre. Nous aimerions utiliser les probabilités a priori impropres partout, mais il n'est pas évident que, dans ce cas, la conjointe a posteriori est toujours bonne. Une estimation pilote des paramètres sera nécessaire pour déterminer les limites supérieures raisonnables pour les probabilités a priori de b_1 et ρ .

Pour l'estimation bayésienne du modèle, Fabrizio Ruggeri et Siva Sivaganesan [14] ont mis en place un algorithme de simulation, composante par composante, qui est semblable à une chaîne de Markov Monte Carlo (MCMC). Nous disons "similaire" parce que, contrairement à des situations où la fonction de la valeur moyenne elle-même est une réalisation d'un processus de Gamma, la distribution a posteriori de g n'est pas celle d'un processus

de Gamma, de sorte que nous ne pouvons simuler directement. L'algorithme est alors le suivant :

1. Mise à jour b_0 selon sa densité conditionnelle totale qui est gamma :

$$G(n + 1, \int_0^T \frac{\log(1 + b_1 g(t))}{1 + b_1 g(t)} s(t; \rho) dt).$$

2. Mise à jour b_1 et ρ séparément avec une étape de Metropolis chacun, où à la fois la proposition et les densités a priori sont uniformes sur une gamme appropriée. Le taux d'acceptation est alors un rapport de vraisemblances.
3. Mise à jour g avec une étape d'acceptation/rejet. Un nouvel g est proposé à partir de la distribution de g conditionnel à savoir que les valeurs de g aux points en t sont celles en k . En négligeant les informations que les défaillances ont été effectivement observées à l'instant t , cette répartition se réduit à celle d'un processus de gamma conditionnés à traverser les points $(t_i, k_i)_i$. Ensuite, il peut être montré que dans chaque intervalle $(t_{i-1}, t_i), i = 1, \dots, n(t_0 = 0)$, la trajectoire de la courbe g est celle d'une fonction de distribution cumulative tirée d'un processus de Dirichlet multiplié par $g(t_i) - g(t_{i-1})$ ci-dessus et décalée par $g(t_{i-1})$. Le processus de Dirichlet possède un paramètre $a(t_{i+1} - t_i)U_i(\cdot)$, où $U_i(\cdot)$ Désigne la distribution uniforme sur (t_{i-1}, t_i) . Dans l'intervalle (t_n, T) de la courbe g est sans contrainte, il est donc simplement d'un processus de gamma de paramètres a et b décalée de $g(t_n)$.

Après cette proposition, une étape d'acceptation est effectuée en utilisant le rapport de vraisemblance en tant que probabilité d'acceptation (qui n'est pas la probabilité d'acceptation requis pour l'équilibre).

Maintenant, nous allons présenter une application sur l'estimation Bayésienne des paramètres d'un modèle de processus de poisson non-homogène en utilisant les méthodes MCMC présentées dans le premier Chapitre.

3.3 Application

3.3.1 Estimation bayésienne des paramètres

Nous avons estimé les paramètres d'un modèle de processus de la Loi de puissance abordée dans le deuxième Chapitre de ce mémoire avec une intensité $\lambda(t, \alpha, \beta) = \alpha\beta t^{\beta-1}$, donc la fonction de vraisemblance est donnée par :

$$L(\alpha, \beta|T) = (\beta/\alpha)^n \left(\prod_i^n t_i/\alpha \right)^{\beta-1} \exp\{-(y/\alpha)^\beta\}, \quad (3.6)$$

où $y_n = t$ pour la défaillance tronquée et $y = t$ pour le temps tronqué.

On considère une analyse Bayésienne pour ce modèle basée sur une loi a priori non informative définie comme suit :

$$\pi(\alpha, \beta) \propto (\alpha\beta^\gamma)^{-1}, \quad (3.7)$$

pour $\gamma = 0$ et $\gamma = 1$ qui correspondent respectivement aux cas de dépendance et d'indépendance.

La loi a priori de α, β dans l'équation (2.7) se transforme en utilisant le Théorème de Bayes sous la forme suivante :

$$\pi(\alpha, \beta|t) = C_\gamma(t) \beta^{n-\gamma} \left(\prod_{i=1}^n t_i \right)^\beta e^{-(y/\alpha)^\beta} / \alpha^{n\beta+1}, \alpha > 0, \beta > 0, \quad (3.8)$$

où C est une constante donnée par :

$$C_\gamma(t) = \left(\ln \prod_{i=1}^n y/t_i \right)^{n-\gamma} / (\Gamma(n)\Gamma(n-\gamma)), \text{ pour } \gamma = 0, 1.$$

Pour cela, on obtient les estimateurs Bayésiens qui sont les suivants :

$$\begin{aligned} \hat{\beta} &= n / \sum_{i=1}^n \ln y/t_i, \\ \hat{\alpha} &= yn^{-1/\hat{\beta}}. \end{aligned}$$

3.3.2 Estimation bayésienne des paramètres par les méthodes MCMC

En remplace la loi a priori proposée dans l'équation (2.12) par :

$$\pi(\alpha) \sim \Gamma(a, b).$$

$$\pi(\beta) \sim \text{Uniform}(0, 2).$$

Dans ce cas la loi a priori est donnée :

$$\pi(\alpha, \beta) = \alpha^{a-1} e^{-b\alpha} \mathbf{I}_{(0,2)\beta},$$

et la loi a posteriori des paramètres α et β est :

$$\pi(\alpha, \beta|T) \propto \alpha^{n+a-1} \beta^n e^{-\alpha y^\beta + b} e^{\beta \sum_{i=1}^n \log t_i} \mathbf{I}_{(0,2)\beta}. \quad (3.9)$$

L'échantillonneur de Gibbs considère que la loi cible est le produit des densités complètes. et pour l'appliquer, il faut tout d'abord déterminer la distribution conditionnelle de chaque paramètre. Dans notre cas, les distributions conditionnelles complètes de α et β sont faciles à extraire. Cependant, on obtient la distribution complète de α en illuminant tous les termes dans la formule (2.14) qui ne dépendent pas de α , et on refait la même chose pour β , d'où

$$P(\alpha|T) \propto \alpha^{n+a-1} e^{-\alpha y^\beta + b}. \quad (3.10)$$

$$P(\beta|T) \propto \beta^n e^{\beta \sum_{i=1}^n \log t_i} \mathbf{I}_{(0,2)\beta}. \quad (3.11)$$

On constate que la formule (2.15) n'est autre qu'une distribution Gamma ($\Gamma(n + a, b + y^\beta)$). L'échantillonneur de Gibbs, dans ce cas, est plus adéquat pour simuler α . Concernant la formule (2.16), il serait utile d'introduire l'algorithme de Metropolis-Hastings pour simuler β .

TAB. 3.1 – Valeurs simulées de α et β

t	1000	1100	1200	1300	1500	1600	1800	2000
$\hat{\alpha}(10^{-4})$	2.86	2.80	2.93	2.96	2.83	2.78	2.79	2.89
$\hat{\beta}$	2.97	2.66	2.46	2.26	2.06	1.90	1.64	1.48

L'implémentation de l'algorithme de Metropolis-Hastings nécessite la mise en place de la distribution de proposition. Pour notre cas nous avons choisi la distribution $\Gamma(n+1, ns^2)$ comme une distribution de proposition pour générer un candidat θ .

Pour le traitement des données disponibles, nous avons implémenté les deux algorithmes précédents sous le logiciel **R**. Afin d'estimer α et β ainsi que la valeur de leur fonction de vraisemblance en fonction de point de changement t , nous avons réalisé 10 000 simulations sur nos données pour $a = 5$, $b = 0.43$ et différentes valeurs t .

Les résultats numériques obtenus sont résumés dans la Table 3.1.

3.3.3 Discussion des résultats

À partir de la Table 3.1 :

- ♣ On constate que les estimateurs $\hat{\beta}$ et $\hat{\alpha}$ dépendent de choix de point de changement.
- ♣ Bien que les vraies valeurs de α et β sont respectivement 2.90×10^{-4} et 2.51. De ce fait, on conclut que le point de changement $t = 1200$ jours est meilleur de fait que les estimateurs fournis en ce point sont les plus proches aux valeurs exactes.
- ♣ En guise de conclusion, cela se traduit par l'utilité et l'importance d'un choix adéquat du point ou des points de changement dans l'analyse des données modélisées par les processus non homogènes.

Conclusion générale

Le présent travail, se focalise sur l'étude des données réelles représentant les pannes de portes du train sur la ligne de métro d'une grande ville européenne. Après la modélisation de ces données par un Processus Poissonien Non Homogène (PPNH) nous proposons d'utiliser la technique MCMC afin d'estimer les paramètres de ce processus ainsi que ses points de changements.

Le premier chapitre a un caractère introductif. Dans ce chapitre, on décrit en détail l'analyse statistique bayésienne et les méthodes de simulation de Monte Carlo par Chaîne de Markov. Dans le second chapitre, on a synthétisé quelques résultats concernant les processus de Poisson (homogènes et non homogènes) et la fiabilité des systèmes réparables. De plus, on a analysé les données de défaillance collectées sur 40 rames de métro, qui ont été livrés à une entreprise de transport entre Novembre 1989 et Mars 1991 et tous ont été mis en service à partir du 6 Avril 1990 au 20 Juillet 1992 dans ce sens une modélisation bayésienne des points de changements de l'historique des pannes dans les portes du train a été faite. Finalement, une application numérique est réalisée sur les données en question. Pour cela, les résultats obtenus montre que l'estimation des paramètres d'un modèle non homogène nécessitent un choix adéquat du point (ou des points) de changement afin d'obtenir des bonnes estimations.

Comme perspectives, on cite :

- Une étude comparative entre les techniques classiques et bayésienne est souhaitable.
- Une extension de ce travail paraît nécessaire en considérant d'autres lois à priori.

- Analyser les techniques d'estimation des points de changements (théorique et application).
- Application de ces différentes techniques sur d'autres échantillons réels avec différentes tailles (petite, moyenne, grande).

Bibliographie

- [1] **A. Kottas**. A nonparametric Bayesian approach to inference for nonhomogeneous Poisson processes, University of California, 1156 High Street, Santa Cruz, CA 95064, USA, NSF grant DMS-0505085, August 23, 2006.
- [2] **A. Zellner**. An Introduction to Bayesian Inference in Econometrics, 448 pages. Wiley 1996.
- [3] **A.F.M. Smith, G.O. Robert**. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. Journal of the Royal Statistical Society, vol(55), 3-23, 1993.
- [4] **A.E. Gelfand and A.F.M. Smith**. Sampling-based approaches to calculating marginal densities. of the American Statistical Association, vol(85(410)), 398-410, June 1990.
- [5] **A. Turkmann, M. A.** Bayesian analysis of an autoregressive process with exponential white noise. Statistics, vol(4), 601-608, 1990.
- [6] **A.E. Raftery, V.E. Akman** . Bayesian analysis of a Poisson process with a change-point. Biometrika, vol(73), 85-89, 1986.
- [7] **A.F.M. Smith**. A Bayesian approach to inference about a change-point in a sequence of random variables. Biometrika. vol(62),407-416, 1975.
- [8] **B. YCART**, Notion de Fiabilité et Files d'Attente, Centre de publication universitaire, Tunis. 132 pages 2004.
- [9] **C. Mazzali, F. Ruggeri**. Bayesian analysis of failure count data from repairable systems. Technical Report. CNR - IAMI, Milano, vol(19), 1998.

- [10] **C. Coccozza.** Processus stochastique et fiabilité des systeme, 436 pages, 1997.
- [11] **C. Robert, G. Cassela.** Monte Carlo statistical methods. Springer-Verlag, New York, 1999.
- [12] **C.Robert.** L'analyse stalislique bayésienne. Statistique mathématique et probabilité. Economica, 1992.
- [13] **C. Robert, G. Cassela.** Monte Carlo Statistical Methods. Springer-Verlag, New York, second édition, 289 pages, 2004.
- [14] **D. Foata, A. Fuchs.** Processus stochastiques : processus de Poisson, Chaines de Marcov et Martingales, Paris, 236 pages, 2004.
- [15] **D. Dey, P. Müller, D. Sinha.** Practical Nonparametrics and Semiparametrics in Bayesian Statistical Inference, vol 133 dans Lecture Notes in Statistics. Springer-Verlag, New York. 1998.
- [16] **D. Cavallo.** Nuovi modelli bayesiani nell'analisi dell'affidabilità dei sistemi riparabili. B.Sc. Thesis, University of Milano, Milano, 1999.
- [17] **D. Cavallo, F. Ruggeri.** Bayesian models for failures in a gas network. In E. Zio, M. Demichela, and N. Piccinini, editors, Safety and Reliability - ESREL 2001, pages 1963-1970, 2001.
- [18] **E. Parent, J. Bernier,** Le raisonnement Bayésien : Modélisation et Inférence. Springer-Verlag France, Paris. 364 pages, 2007.
- [19] **F. Ruggeri, S. Sivaganesan.** On modeling change points in nonhomogeneousPoisson processes. Tentatively accepted by Stat. Inference Stoch. Process, vol(8), 311–329, 2005.
- [20] **F. Ruggeri, A. Pievatolo.** Bayesian modelling of train doors reliability, In Handbook of Applied Bayesian Analysis (A. O'Hagan and M. West Eds.), Oxford Press, Part II, Chapter 10. Oxford University Press. 271-294, 2010.
- [21] **F. Rugger, A. Pievatolo.** Bayesian reliability analysis of complex repairable systems. Applied Stochastic Models in Business and Industry, vol(20), 253-264, 2004.

- [22] **F. Ruggeri**. On the Reliability of Repairable Systems : Methods and Applications. Progress in Industrial Mathematics at ECMI 2004, A. Di Bucchianico, R.M.M. Mattheij, and M.A. Peletier, Eds, vol(8), 535-553, 2005.
- [23] **F. Ruggeri, A. Pievatolo, R. Argiento** Bayesian analysis and prediction of failures in underground trains . Quality Reliability Engineering International, vol(19), 327-336, 2003.
- [24] **J.O. Berger**. Statistical Decision Theory and Bayesian Analysis, Second Edition, Springer-Verlag, . 617 pages, 1985.
- [25] **J. Besag**. Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society : Series B (Methodological), 36, 192-236. 1974.
- [26] **J.K. Ghosh, M. Delampady, T. Samanta**. An Introduction to Bayesian Analysis Theory and Methods. 366 pages, Springer 2007.
- [27] **J.J. Boreux, E. Parent, J. Bernier**. Pratique du calcul Bayésien. Springer-Verlag France Paris, 1 online resource (xxiv, 333 pages), 2010.
- [28] **J. Albert, S. Chib**. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. Journal of Business and Economic Statistics, vol(11), 1-15, 1993.
- [29] **J. Bernardo**. Reference posterior distributions for Bayesian inference (with discussion). J. Royal Statist. Soc. Series B, vol(41), 113-147, 1979.
- [30] **L.H. Crow**. Confidence interval procedures for the Weibull process with applications to reliability growth, Technometrics, vol(24), 67-72, 1982.
- [31] **L. Kuo, T.Y. Yang**. Bayesian computation for nonhomogeneous Poisson processes in software reliability, J. Amer. Statist. Assoc. vol(91), 763-773, 1996.
- [32] **L.DeRobertis, J.A. Hartigan**. Bayesian inference using intervals of measure, Ann. Statist, vol(1), 235-244, 1981.

- [33] **M. Ibazizen, H. Fellag.** Bayesian estimation of an AR(1) process with exponential white noise. *Statistics : A journal of Theoretical and Applied Statistics*, vol 37(5), 365-372, 2003.
- [34] **M. Guida, R. Calabria, and G. Pulcini.** Bayes inference for a nonhomogeneous Poisson process with power intensity law. *IEEE Transactions on Reliability*, vol(38), 603-609, 1989.
- [35] **M. Tanner, W. Wong.** The calculation of posterior distributions by data augmentation. *J. American Statist. Assoc.*, vol(82), 528-550, 1987.
- [36] **N. Metropolis, A.W. Rosenbluth, M.N.Rosenbluth, A.H. Teller, E. Teller.** Equation of state calculations by fast computins machines. *Chemical Physics*, vol(21(6)), 1087-1092, june 1953.
- [37] **P. O'connor,** Testing for reliability, *Quality and reliability engineering*, vol(19), 73-84, 2003.
- [38] **P. Sander, R.Badoux,** *Bayesian Methods in Reliability*, Kluwer Academic Publisher, 1991.
- [39] **P. Robert, Christian.** *Le choix Bayésien : Principes et pratique.* Springer-Verlag France, Paris, vol(1), 638 pages, 2006.
- [40] **P. Laplace.** *Essai Philosophique sur les Probabilités.* Epistémé. Christian Bourgeois, Paris, 1795, reprinted in 1986.
- [41] **R. Kass, A. Raftery.** Bayes factor and model uncertainty. *J. American Statist. Assoc.*, vol(90), 773-795, 1995.
- [42] **S.E. Rigdon, A.P. Basu .** The Power Law process : a model for the reliability of repairable systems. *Journal of Quality Technology*, vol(10), 251-260, 1989.
- [43] **S. Bar-Lev, I. Lavi, B. Reiser.** Bayesian inference for the Power Law process. *Ann. Inst. Statist. Math*, vol(44), 623-639, 1992.
- [44] **S.E. Rigdon, Basu A.P.** The Power Law process : a model for the reliability of repairable systems. *Journal of Quality Technology*, vol(10), 251-260, 1989.

- [45] **S. Chib, E. Greenberg** . Understanding the Metropolis-Hastings Algorithm. Journal of the American Statistical Association, vol(49), 327-335, 1995.
- [46] **T.Y. Yang, L. Kuo**. Bayesian Computation for the Superposition of Nonhomogeneous poisson processes. vol(27), 547-556, 1999.
- [47] **W.I. Bryc**. Applied Probability and Stochastic Processes, Lecture Notes for 577-578 Class, March 21, 1996.
- [48] **W.A. Massey, A. Geraldine** . Parker and Ward Whitt, Estimating the Parameters of a Nonhomogeneous Poisson Process with Linear Rate, AT and T Bell Laboratories Murray Hill, NJ 07974-0636 November 16, 1993.
- [49] **W.K. Hastings**. Monte Carlo Sampling methods using Markov chains and their application, Biometrika, vol(57), 97-109, April 1970.
- [50] **Y. Goegebeur**, The Poisson and the nonhomogeneous Poisson process, ST505 - Statistical Simulation, December 3, 2007.
- [51] **Z. Chen**. Empirical Bayes Analysis on the Power Law Process with Natural Conjugate Priors, Journal of Data Science. vol(8), 139-149, 2010.