École Doctorale Carnot-Pasteur

Thèse de doctorat

Discipline : Mathématiques

présentée par

Alexis Flesch

Bandes de confiance par vraisemblance empirique : δ-méthode fonctionnelle et applications aux processus des événements récurrents

---

dirigée par Jean-Yves Dauxois et Davit Varron

Rapporteurs : Laurent Bordes et Ingrid Van Keilegom

Soutenue le 12 juillet 2012 davant le jury composé de :

M. Philippe Berthet	Examinateur
M. Laurent Bordes	Rapporteur
Mme Fabienne Comte	Examinatrice
M. Jean-Yves Dauxois	Directeur
Mme Agathe Guilloux	Examinatrice
Mme Katy Paroux	Examinatrice
M. Davit Varron	Directeur

Table des matières

Remerciements

Après presque quatre ans de thèse pour un total de huit ans passés à l'Université de Franche-Comté, une page se tourne et ma vie d'étudiant s'achève. Mais quelle belle page et que de belles rencontres pendant mes études !

Il convient avant tout de remercier chaudement mes deux directeurs de thèse, Jean-Yves Dauxois et Davit Varron.

Jean-Yves, tout d'abord, pour avoir eu le courage (la folie) d'accepter de m'encadrer. Pour avoir su trouver des plages horaires dans son planning surchargé, pour m'écouter et répondre à mes questions. Mais aussi pour m'avoir permis de travailler avec Davit.

Davit. Tant d'heures passées dans son bureau, et pourtant, sa porte est toujours restée ouverte (principalement parce qu'il ne fermait pas son bureau à clef). Merci à lui pour sa patience sans limite et pour les heures qu'il a passées à relire mes calculs.

Mais, avant d'être des chercheurs, ce sont aussi des personnes désintéressées qui m'ont encouragé dans mes choix de carrière là où d'autres m'auraient poussé dans une direction qui n'était pas la mienne. Encore merci à eux pour leur sympathie et leur bonne humeur.

Deux autres personnes ont joué un rôle primordial dans ma thèse, il s'agit bien évidemment de mes rapporteurs.

Merci à Laurent Bordes et à Ingrid Van Keilegom d'avoir accepté la lourde tâche de relire mon manuscrit. Moi qui ai besoin de plusieurs semaines pour déchiffrer quelques pages d'un texte scientifique, je n'ose imaginer la pénibilité de la relecture complète d'une thèse.

Je me dois aussi de remercier tous les membres du Jury pour avoir fait le déplacement jusqu'à Besançon en plein mois de juillet. Merci à Philippe Berthet, que j'avais déjà eu la chance de rencontrer lors du colloque des jeunes statisticiens à Aussois. Merci à Fabienne Comte et à Agathe Guilloux. Enfin, merci à Katy Paroux, qui avant de faire partie de mon jury m'a enseigné la statistique lors de mon année de préparation à l'agrégation et m'a donné le goût pour cette matière. Merci à elle pour la qualité de son enseignement, son investissement et sa bonne humeur.

Lors de mes études à l'Université de Franche-Comté, j'ai eu la chance de faire de très belles rencontres, autant au niveau de mes enseignants que des étudiants ainsi que des chercheurs et du personnel du laboratoire de

mathématiques.

Je tiens à remercier tout particulièrement Gilles Lancien, à qui je dois d'avoir (encore) mon agrégation. Cet enseignant hors pair (et souvent hors jeu) a toujours eu la patience de répondre à mes questions avec une pédagogie sans pareil.

Merci aussi à Stéphane Chrétien. S'il ne m'avait pas initié à l'optimisation (non) convexe pendant mon Master en me proposant d'encadrer les TDs de recherche opérationnelle avec lui, beaucoup de questions dans ma thèse seraient restées en suspens. Merci à lui pour sa bonne humeur permanente et ses mauvais jeux de mots [1].

Merci à toute l'équipe de la prépa agreg de Besançon, sans laquelle je n'aurais pas mon poste aujourd'hui : Jean-Robert Béliard, Stéphane Chrétien, Jean-Yves Dauxois, Yves Dutrilleux, Florence Lancien, Gilles Lancien, Christian Lemerdy, Éric Ricard, et j'en oublie sans doute.

Merci aussi à toute l'équipe de Probabilités et Statistique : Landy, Célestin, Youri, Stéphane, Jean-Yves, Davit, Julien, Francial, Juan-Pablo.

Merci à Richard et à Romain, qui sont toujours disponibles pour dépanner. Merci à Catherine Pagani et à Catherine Vuillemenot dont la gentillesse et l'efficacité ne sont plus à démontrer.

Entre Karine, Émilie et Céline, la vie n'a pas toujours été simple au bureau. Car oui, elles sont toutes les trois adorables, mais aussi un peu folles. Que de bons souvenirs dans ce bureau 401 à leurs côtés ! Que de bons moments passés autour d'un café/thé et de petits gâteaux maison faits par les uns et les autres. Je n'aurais pu rêver meilleurs acolytes que ces trois-là. Mais ce ne sont pas les seules à avoir animé l'équipe des doctorants. Il y a bien sûr Guillaume, compagnon de promotion pendant 8 ans ! Mais aussi Guixiang, grâce à qui je balbutie quelques mots de chinois. Stéphane Viguié, un peu déconnecté, mais le sourire aux lèvres et la canette à la bouche. Bien sûr, n'oublions pas Anthony qui a participé activement à la bonne ambiance au laboratoire, ainsi que Mathilde et Vanessa. Merci enfin à mes aînés, Dimitri, Florent et Olivier qui m'ont montré la voie.

Merci à tous mes amis qui m'ont aidé à « travailler mes statistiques » depuis des années : Jérémie, Nicolas, Aurélien, Marine, Martine et Tartine. Mais aussi David, Édouard, et j'en oublie... Merci aussi à Damien, Jean et Maxime pour les soirées PES (Probabilités Et Statistique), FIFA (Fun In Functional Analysis) et coinche.

Merci à toute l'équipe de l'AC2000 avec laquelle j'ai essuyé plus d'une défaite, mais toujours dans la bonne humeur. Merci en particulier à Olivier Braun et à JC pour leurs mauvais jeux de mots [2].

Merci à Thomas, pour qui j'ai une pensée émue, ainsi qu'à toute sa famille.

Enfin, merci à mes parents qui m'ont permis de poursuivre des études longues et qui ont même eu le courage de relire mon manuscrit à l'assaut

1. à côté desquels les miens passent pour du Raymond Devos.
2. à côté desquels ceux de Stéphane passent pour du Raymond Devos.

des fautes d'orthographe et de grammaire (non sans succès !). Merci enfin à mon frère, Étienne.

Introduction

En 1975, Thomas et Grunkemeier ont posé les bases de ce qu'on allait plus tard appeler la vraisemblance empirique. Dans un problème de durées de vie avec censure aléatoire indépendante, ils ont trouvé une méthode permettant d'obtenir des régions de confiance pour des variables aléatoires réelles i.i.d. dans L^2 sans avoir à estimer de variance. Depuis, beaucoup de résultats ont été publiés sur le sujet et on dénombre beaucoup d'applications en Statistique. Les principales références sur le sujet sont le livre d'Owen (1990) ainsi que l'article de Hall & La Scala (1990).

La construction de régions de confiance par vraisemblance empirique est coûteuse en temps, en particulier lorsque les données sont dans des espaces de grande dimension. De plus, pour démontrer des résultats de convergence sur le rapport de vraisemblance empirique, nous serons amené[3] à utiliser le théorème de la limite centrale. Pourquoi alors ne pas s'arrêter à un tel résultat de normalité asymptotique ?

Parmi les avantages de la vraisemblance empirique par rapport au théorème de la limite centrale, citons la forme des régions de confiance : celle-ci est déterminée par la géométrie des données, alors que les régions obtenues par le théorème de la limite centrale ont une forme ellipsoïdale qui ne rend pas compte de la non-symétrie éventuelle de la distribution sous-jacente.

Il existe d'autres raisons pour lesquelles la vraisemblance empirique est populaire. Hall & La Scala (1990) étudient en particulier les avantages de la vraisemblance empirique par rapport au bootstrap et montrent que les régions obtenues par vraisemblance empirique admettent une correction de Bartlett. Cependant, nous n'étudierons pas ces aspects dans notre travail.

Plusieurs centaines d'articles ont déjà été publiés sur la vraisemblance empirique et il serait trop long ici d'en faire une revue complète. Cependant, les applications dans les problèmes d'événements récurrents n'ont pas encore été explorées en profondeur, bien que quelques articles aient déjà vu le jour. Citons entre autre Zhao & Hsu (2005) qui ont construit des bandes de confiance pour le paramètre de régression dans un modèle additif avec durées de vie censurées. Citons encore Adimari (1997) et Ren (2001) qui ont utilisé des méthodes de vraisemblance empirique pour étudier la moyenne dans un problème de durées de vie censurées. Enfin, citons

3. Bien que ce manuscrit comporte des fautes grammaticales avec une probabilité non nulle, ceci n'en est pas une, le *nous de modestie* s'accordant de manière sylleptique.

Harari-Kermadec (2011) qui a étudié les chaînes de Markov récurrentes positives par vraisemblance empirique. Bien sûr, cette bibliographie est loin d'être exhaustive. Cependant, on ne dénombre pas plus de quelques dizaines d'articles sur le sujet.

Tout au long de cette thèse, nous nous sommes appliqué à étudier les problèmes en lien avec les durées de vie ou les processus d'événements récurrents. Que ce soit en biostatistiques, en sciences sociales, ou encore en fiabilité, les applications sont nombreuses. Nous nous sommes particulièrement intéressé aux événements récurrents, aux risques concurrents, en présence de covariables, d'un événement terminal ainsi que d'une censure aléatoire indépendante. Nous avons proposé divers modèles issus de la littérature et avons construit des bandes de confiance par vraisemblance empirique pour les processus d'intérêt dans chacun des modèles.

La première partie de cette thèse est un rappel des résultats classiques qui nous seront nécessaires pour la suite.

L'étude des processus étant un enjeu majeur de ce travail, le premier chapitre donne le cadre d'étude dans lequel nous nous plaçons, ainsi que les « bonnes définitions » de la convergence en loi. Nous nous sommes en particulier attaché à faire le lien entre la métrique de Skorokhod et l'approche de Hoffman-Jørgensen, lien que nous n'avons pas trouvé présenté de manière claire dans la littérature. Les résultats sur la métrique de Skorokhod sont empruntés à Billingsley (1968), et certaines démonstrations ont été simplifiées et adaptées à notre cadre d'étude.

Dans le deuxième chapitre, nous présentons la théorie des processus empiriques au sens de Pollard (1990) et reprenons le cheminement permettant de démontrer des résultats de type « loi des grands nombres uniforme » ou « théorème de la limite centrale fonctionnel ». Nous donnons les éléments clefs des démonstrations et faisons le lien avec la théorie des classes de Vapnik et Chervonenkis.

Le troisième chapitre est consacré à la vraisemblance empirique. Nous y énonçons le théorème d'Owen ainsi qu'un théorème démontré par Hjort, McKeague et Van Keilegom (2004). Nous donnons une formule plus générale en supprimant l'hypothèse sur le rang de la matrice de variance-covariance. Bien qu'il ne s'agisse que d'un jeu d'écriture, nous verrons au chapitre 11 en quoi cela peut rendre service dans la pratique.

Dans le quatrième chapitre, nous énonçons quelques résultats classiques des problèmes de durées de vie, en particulier les propriétés des estimateurs de Nelson-Aalen et de Kaplan-Meier. Ce chapitre est presque entièrement issu des résultats présentés dans le livre de Andersen *et al.* (1993).

La deuxième partie de ce travail, découpée en plusieurs chapitres, reprend les principaux résultats démontrés au cours de cette thèse.

Le chapitre 5 est consacré à l'étude de l'enveloppe convexe de processus à valeurs vectorielles. Nous y présentons des résultats préliminaires que nous utilisons dans tous les chapitres suivants.

Dans le chapitre 6, nous construisons des bandes de confiance pour la fonction de risque cumulé par vraisemblance empirique dans diverses situations. Nous nous intéressons en particulier aux événements de types multiples en présence d'une censure aléatoire indépendante ainsi qu'aux problèmes multi-états.

Le chapitre 7 se concentre sur l'estimation de fonctions moyennes pour des événements récurrents avec risques concurrents, événement terminal et censure aléatoire à droite. Il reprend le modèle présenté dans Dauxois & Sencey (2009) ainsi qu'un de leurs résultats. Nous y conduisons des simulations pour vérifier la validité de notre modèle sur des échantillons de taille finie et appliquons notre résultat à un jeu de données d'infections nosocomiales.

Dans le chapitre 8, nous reprenons le modèle présenté dans Cai & Schaubel (2004b) et obtenons un résultat de convergence dans un contexte de durées de vie avec événements récurrents, risques concurrents, covariables et censure aléatoire à droite indépendante.

Le chapitre 9 est consacré au problème de la construction de bandes de confiance par vraisemblance empirique pour des fonctionnelles différentiables. Il est principalement basé sur les travaux de Bertail (2006) et donne une généralisation de son résultat adaptée à l'étude des processus. Nous y expliquons sur des exemples simples comment s'applique notre résultat en pratique.

Le chapitre 10, tout comme le chapitre 6, donne une méthode pour construire des bandes de confiance par vraisemblance empirique pour les fonctions de risque cumulé. Nous choisissons alors une autre approche en appliquant les résultats du chapitre 9 plutôt que les propriétés de l'estimateur de Nelson-Aalen.

Dans le chapitre 11, nous appliquons le résultat énoncé au chapitre 9 pour étudier les problèmes multi-états et expliquons comment construire des bandes de confiance par vraisemblance empirique pour la matrice de transition d'un processus de Markov non homogène à temps continu soumis à une censure aléatoire à droite indépendante.

La troisième partie de cette thèse est quant à elle consacrée aux méthodes employées pour la simulation du processus de comptage présenté au chapitre 7 et nous y donnons une version simplifiée des algorithmes utilisés. Nous y présentons aussi succinctement le module de vraisemblance empirique écrit pour l'occasion en Python [4].

4. Python est un langage de programmation interprété possédant de nombreuses extensions destinées aux applications numériques. Sa syntaxe claire et sa rapidité en font un des langages les plus utilisés à ce jour.

Première partie

Outils Mathématiques

CHAPITRE 1

Convergence en distribution

1.1 Introduction

Dans tout notre travail, nous nous sommes concentré sur l'étude de processus et nous avons démontré des résultats de convergence, notamment en distribution.

Il existe plusieurs définitions de convergence en loi pour des processus. Les deux plus connues sont celle au sens de Skorokhod et celle au sens de Hoffmann-Jørgensen, c'est pourquoi on trouve dans la littérature des théorèmes de convergence pour l'une ou l'autre définition. Comme nous allons utiliser des résultats provenant des deux théories, il nous semble nécessaire de les rappeler ici et de montrer qu'elles se rejoignent dans le cadre qui nous intéresse.

Dans ce chapitre nous allons étudier les problèmes de mesurabilité inhérents à la convergence en loi pour des variables aléatoires indexées par un paramètre de temps. Commençons par définir formellement ces objets.

Définition 1.1 (processus)**.** Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et T un espace quelconque. On dira que f est un *processus* (sur $(\Omega, \mathcal{A}, \mathbb{P})$) indexé par T si f est une application de $\Omega \times T$ dans $\mathbb{R}$ telle que $f(\cdot, t)$ est borélienne pour tout t de T.

Si $X_1, \dots, X_n$ sont des variables aléatoires réelles, alors, la *fonction de répartition empirique* F_n définie ci-dessous est bien un processus au sens de la définition précédente :

$$F_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{X_i \leqslant t\}}.$$

Dans ce cas, $T = \mathbb{R}$.

Lorsque l'on est en présence d'un processus X, il peut être intéressant d'étudier la fonction $\sup_{t \in T} X(t)$, qui n'est pas nécessairement mesurable. Cependant, elle le sera dès que le processus sera séparable au sens de la définition ci-dessous.

Définition 1.2 (processus séparable)**.** Soit (T, ρ) un espace topologique séparable et f un processus indexé par T. On dira que f est *séparable* si il existe un ensemble négligeable $N \subset \Omega$ et un espace séparable $T_0 \subset T$ tel que pour tout ω dans $\Omega \backslash N$ et pour tout $t \in T_0$, il existe une suite $(t_n)_n$ dans T_0 telle que $t_n \to t$ et $f(\omega, t_n) \to f(\omega, t)$.

Dans ce qui suit, sauf mention contraire, on s'intéressera à des processus indexés par $[0, \tau]$ où τ est une constante strictement positive fixée.

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de processus aléatoires, et soit X un processus aléatoire. On pourrait se contenter de définir la convergence en loi de la suite $(X_n)_{n\in\mathbb{N}^\star}$ vers X par la convergence des marges fini-dimensionnelles. Cela signifie que pour tout entier p, pour tout $(t_1, \dots, t_p) \in [0, \tau]^p$ et pour toute fonction f continue bornée de $\mathbb{R}^p$ dans $\mathbb{R}$, on a :

$$\mathbb{E}\, f(X_n(t_1), \dots, X_n(t_p)) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\, f(X(t_1), \dots, X(t_p)).$$

Cependant, cette définition n'est pas satisfaisante car elle est trop faible. On aimerait en effet pouvoir dire, par exemple, que si $(X_n)_{n\in\mathbb{N}^\star}$ converge en loi vers X, alors, pour tout réel M, on a :

$$\mathbb{P}(\|X_n\| \leqslant M) \underset{n\to+\infty}{\longrightarrow} \mathbb{P}(\|X\| \leqslant M). \tag{1.1}$$

Ce type de convergence est essentiel en statistique pour construire des bandes de confiance uniformes par exemples. La convergence des marges fini-dimensionnelles n'est cependant pas une condition suffisante pour obtenir ce résultat.

Il nous faut alors trouver une autre définition de la convergence en loi. De manière naïve, on pourrait voir les processus aléatoires X_n comme des variables aléatoires à valeurs dans $(\mathcal{D}([0, \tau]), \|\cdot\|_\infty)$, l'ensemble des fonctions càdlàg sur $[0, \tau]$ et utiliser la définition usuelle de la convergence en loi, à savoir :

$$\forall f \in \mathscr{C}_b(\mathcal{D}[0, \tau], \mathbb{R}), \quad \mathbb{E}\, f(X_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\, f(X).$$

Pour que cette définition ait un sens, il faudrait que les $f(X_n)$ soient mesurables. Se pose alors d'emblée la question de la topologie dont on munit $\mathcal{D}([0, \tau])$. Si on choisit par exemple d'utiliser la norme infinie, alors, même des processus très simples peuvent ne pas être mesurables comme le montre l'exemple ci-dessous.

Exemple 1.3. Soit $\Omega = [0, 1]$ muni de la tribu borélienne et de la mesure de Lebesgue. On définit sur Ω une variable aléatoire de loi $\mathcal{U}([0, 1])$ en posant :

$$\forall \omega \in \Omega, \quad U(\omega) = \omega.$$

Soit maintenant G le processus aléatoire défini pour tout $(\omega, t) \in \Omega \times [0, 1]$ par $G(\omega, t) = \mathbb{1}_{\{U(\omega)\leqslant t\}}$. Définissons encore pour tout a dans $[0, 1]$ l'application

$$\begin{array}{rccc} f_a & : & [0,1] & \to & \mathbb{R} \\ & & t & \mapsto & \mathbb{1}_{\{a\leqslant t\}}, \end{array}$$

ainsi que l'ensemble

$$F_a := \big\{\omega \in \Omega,\ \|G(\omega,\cdot) - f_a\|_\infty < 1/2\big\}.$$

Alors :

$$\begin{aligned}\omega \in F_a &\Leftrightarrow \forall t \in [0,1],\ |G(\omega,t) - f_a(t)| < 1/2\\ &\Leftrightarrow \forall t \in [0,1],\ \mathbb{1}_{\{\omega \leqslant t\}} = \mathbb{1}_{\{a \leqslant t\}}\\ &\Leftrightarrow \omega = a.\end{aligned}$$

On en déduit donc que $F_a = \{a\}$. Posons maintenant :

$$O_a = \big\{f \in \mathcal{D}([0,1])\ :\ \|f - f_a\|_\infty < 1/2\big\}.$$

Alors, O_a est un ouvert de $(\mathcal{D}([0,1]), \|\cdot\|_\infty)$. De plus, $G^{-1}(O_a) = F_a$. Soit maintenant A une partie de $[0,1]$ qui ne soit pas un borélien. Alors, on peut écrire :

$$A = \bigcup_{a\in A}\{a\} = \bigcup_{a\in A} F_a = \bigcup_{a\in A} G^{-1}(O_a) = G^{-1}\left(\bigcup_{a\in A} O_a\right).$$

Or, $\cup O_a$ est un ouvert de $\mathcal{D}([0,1])$ et donc l'application G ne peut pas être borélienne.

Pour contourner ces problèmes de mesurabilité, il existe deux méthodes principales : celle qui utilise la topologie de Skorokhod ainsi que celle proposée par Hoffmann-Jørgensen.

La première consiste à construire sur $\mathcal{D}([0,\tau])$ une distance plus faible que celle induite par la norme uniforme.

La seconde, quant à elle, permet de contourner les problèmes de mesurabilité en définissant l'espérance de fonctions non mesurables et en conservant sur $\mathcal{D}([0,\tau])$ la norme infinie. L'avantage de cette deuxième méthode est qu'elle n'impose que la mesurabilité du processus limite X.

1.2 La métrique de Skorokhod sur $D([0,\tau])$

1.2.1 Définition et premières propriétés

L'approche de Skorokhod consiste à construire sur $\mathcal{D}([0,\tau])$ une distance d_S plus faible que la norme infinie, qui rend cet espace séparable et pour laquelle le processus empirique est mesurable. Cette distance est cependant suffisamment forte pour que des propriétés comme (1.1) soient conservées.

La définition de la convergence en loi sera alors la définition « classique » avec les fonctions test et sera transportée par les applications continues.

Avec cette nouvelle métrique, la séparabilité de $(\mathcal{D}([0,\tau]), d_S)$ assure que la tribu engendrée par les boules ouvertes coïncide avec la tribu engendrée par les ouverts. Les fonctions continues de $(\mathcal{D}([0,\tau]), d_S)$ dans $\mathbb{R}$ seront donc mesurables.

Intuitivement, la métrique de Skorokhod permet de dire que deux fonctions sont proches si on peut passer de l'une à l'autre en se permettant une petite déformation dans l'espace et dans le temps. *A contrario*, la norme uniforme sur $\mathcal{D}([0,\tau])$ ne permet qu'une déformation dans l'espace.

Pour illustrer ceci, considérons une suite de fonctions $(f_n)_{n\in\mathbb{N}^\star}$ où les f_n sont définies pour tout n par (on pourra se reporter à la figure 1.1) :

$$\begin{array}{rccl} f_n & : & [0,1] & \to & \mathbb{R} \\ & & t & \mapsto & \begin{cases} 1-1/n & \text{si } t < 1/2+1/n, \\ 2+1/n & \text{sinon.} \end{cases} \end{array}$$

Posons encore :

$$\begin{array}{rccl} f & : & [0,1] & \to & \mathbb{R} \\ & & t & \mapsto & \begin{cases} 1 & \text{si } t < 1/2, \\ 2 & \text{sinon.} \end{cases} \end{array}$$

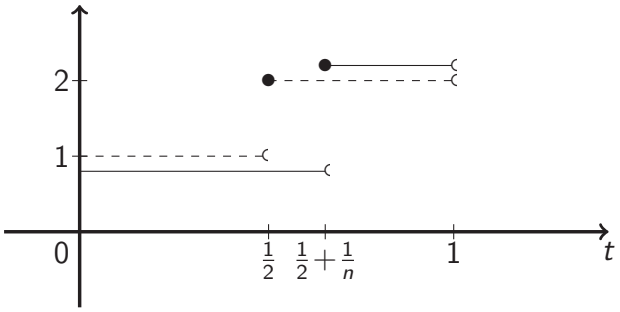


FIGURE 1.1 – Représentation de f_n (en trait plein) et de f (en pointillés).

Alors, $\|f_n - f\|_\infty \geqslant |f_n(1/2) - f(1/2)| \geqslant 1$. Ainsi, il n'y a pas convergence de la suite $(f_n)_{n\in\mathbb{N}^\star}$ vers f pour la norme infinie. Cependant, il y a convergence pour la métrique de Skorokhod que nous allons définir ici.

Définition 1.4 (distance de Skorokhod)**.** Soit Λ l'ensemble des fonctions bijectives continues et strictement croissantes de $[0,\tau]$ dans $[0,\tau]$. Soit Id la fonction identité sur $[0,\tau]$. On définit la distance de Skorokhod entre deux fonctions f et g de $\mathcal{D}([0,\tau])$ par :

$$d_S(f,g) = \inf_{\lambda\in\Lambda} \max\left(\|\lambda - \mathrm{Id}\|_\infty, \|f - g\circ\lambda\|_\infty\right).$$

Remarque 1.5. En prenant $\lambda := \mathrm{Id}$, on voit que $d_S(f,g) \leqslant \|f-g\|_\infty$. Ainsi, la topologie engendrée par d_S est plus faible que la topologie issue de la norme infinie. De plus, nous allons démontrer que $(\mathcal{D}([0,\tau]), d_S)$ est un espace séparable : la tribu borélienne sur $(\mathcal{D}([0,\tau]), d_S)$ est donc la même que celle engendrée par les boules ouvertes.

Énonçons maintenant un résultat élémentaire qui nous sera utile dans la suite pour faire le lien entre l'approche de Skorokhod et celle de Hoffmann-Jørgensen.

Proposition 1.6. *Soit $(f_n)_{n\in\mathbb{N}^\star}$ une suite d'éléments de $\mathcal{D}([0,\tau])$ et soit f une fonction continue sur $[0,\tau]$. Alors, lorsque $n \to +\infty$:*

$$f_n \xrightarrow{\|\cdot\|_\infty} f \quad \Longleftrightarrow \quad f_n \xrightarrow{d_S} f.$$

Démonstration. Soit $(f_n)_{n\in\mathbb{N}^\star}$ une suite convergente pour la distance de Skorokhod. Notons f sa limite ; il s'agit de montrer qu'il y a convergence en norme infinie (la réciproque étant vraie d'après la remarque précédente). Par convergence de la suite $(f_n)_{n\in\mathbb{N}^\star}$, il existe une suite $(\lambda_n)_{n\in\mathbb{N}^\star}$ d'éléments de Λ telle que $\|\lambda_n - \mathrm{Id}\| \to 0$ et $\|f(\lambda_n) - f_n\| \to 0$. Or :

$$\|f - f_n\| \leqslant \|f - f(\lambda_n)\| + \|f(\lambda_n) - f_n\|. \tag{1.2}$$

Le deuxième terme dans (1.2) converge vers 0 par définition de la suite $(\lambda_n)_{n\in\mathbb{N}^\star}$. Quant au premier terme, par uniforme continuité de f et puisque la suite $(\lambda_n)_{n\in\mathbb{N}^\star}$ converge uniformément vers Id, il tend lui aussi vers 0. □

Remarque 1.7. En reprenant la démonstration ci-dessus, on démontre facilement que si $\mathcal{C}_f$ désigne l'ensemble des points de continuité de f et que f_n converge vers f pour d_S, alors, $|f_n(s) - f(s)| \to 0$ uniformément en $s \in \mathcal{C}_f$.

L'approche de Skorokhod consiste alors à définir la convergence en loi d'une suite de processus aléatoires $(X_n)_{n\in\mathbb{N}^\star}$ définis de $[0,\tau]$ dans $\mathbb{R}$ par la condition :

$$\forall f \in \mathscr{C}_b(\mathcal{D}([0,\tau]),\mathbb{R}), \quad \mathbb{E}\, f(X_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\, f(X),$$

où on a muni $\mathcal{D}([0,\tau])$ de la distance de Skorokhod. Le problème ici est que *a priori*, la convergence en loi n'est transportée que pour les applications Skorokhod-mesurables.

1.2.2 Séparabilité et mesurabilité

Dans son livre, Billingsley (1968) démontre que les processus (au sens de la définition 1.1) sont mesurables en tant que fonctions de $(\Omega,\mathcal{A},\mathbb{P})$ dans $(D([0,\tau]), d_S)$. Pour ce faire, il commence par énoncer le lemme suivant (cf. Billingsley (1968) page 122) :

Lemme 1.8. *Soit $f \in \mathcal{D}([0,\tau])$ et $\varepsilon > 0$. Alors, il existe une subdivision $0 = t_0 < t_1 < \ldots < t_n = \tau$ de l'intervalle $[0,\tau]$ telle que :*

$$\forall i \in [\![1,n]\!], \quad \sup_{(s,t)\in[t_{i-1},t_i[^2} |f(s) - f(t)| < \varepsilon.$$

De ce lemme, on déduit en particulier que l'ensemble des points de discontinuité de f est au plus dénombrable et que $\|f\|_\infty < \infty$. Billingsley (1968) montre ensuite le lemme ci-dessous.

Lemme 1.9. *Soit $t \in [0,\tau]$. Alors, l'application $\pi_t : f \mapsto f(t)$ est borélienne en tant qu'application de $(\mathcal{D}([0,\tau]), d_S)$ dans $(\mathbb{R},\mathcal{B}_\mathbb{R})$* [1].

1. $\mathcal{B}_\mathbb{R}$ désigne la tribu borélienne sur $\mathbb{R}$.

Démonstration. Pour $\varepsilon > 0$ fixé, on définit la fonction h_ε de $\mathcal{D}([0,\tau])$ dans $\mathbb{R}$ par :

$$h_\varepsilon(f) = \frac{1}{\varepsilon}\int_t^{t+\varepsilon} f(s)\mathrm{d}s.$$

Montrons que les fonctions h_ε sont continues pour la métrique de Skorokhod. Soit $(f_n)_{n\in\mathbb{N}^\star}$ une suite de $\mathcal{D}([0,\tau])$ convergeant vers f pour d_S. Alors, d'après la remarque 1.7, on sait que f_n converge uniformément vers f sauf éventuellement sur un sous-ensemble de $[0,1]$ de mesure nulle (correspondant aux points de discontinuité de f). Donc, d'après le théorème de convergence dominée :

$$\lim_{n\to+\infty} h_\varepsilon(f_n) = h_\varepsilon(f).$$

De plus, par continuité à droite des éléments de $\mathcal{D}([0,\tau])$:

$$\forall f \in \mathcal{D}([0,\tau])\ ,\quad h_{1/n}(f) \underset{n\to+\infty}{\longrightarrow} f(t) = \pi_t(f).$$

Ainsi, π_t est mesurable en tant que limite de fonctions mesurables. □

Proposition 1.10. *L'espace de Skorokhod $(\mathcal{D}([0,\tau]), d_S)$ est séparable. De plus, une partie dénombrable dense D est donnée par l'ensemble des fonctions de la forme :*

$$\sum_{i=1}^{n} r_i \mathbb{1}_{[q_i,q_i'[},\quad q_i < q_i',\ (r_i,q_i,q_i') \in \mathbb{Q}\times\left(\mathbb{Q}\cap[0,\tau]\right)^2, n\in\mathbb{N}^\star.$$

Démonstration. Nous proposons ici une démonstration sensiblement différente de celle proposée par Billingsley (1968). Soit $f \in \mathcal{D}([0,\tau])$ et $\varepsilon > 0$. Soit $t_0 < \ldots < t_n$ une subdivision de l'intervalle $[0,\tau]$ donnée par le lemme 1.8 appliqué avec $\varepsilon/2$. Posons :

$$f_\varepsilon = \sum_{i=1}^{n} f(t_{i-1})\mathbb{1}_{[t_{i-1},t_i[}.$$

Alors, $d_S(f,f_\varepsilon) \leqslant \varepsilon/2$. De plus, si les t_i et les $f(t_i)$ sont dans $\mathbb{Q}$, alors $f_\varepsilon \in D$. Si ce n'est pas le cas, alors on peut choisir une subdivision de rationnels $0 = q_0 < q_1 < \ldots < q_n = \tau$ telle que pour tout $i \in [\![1,n]\!]$ on ait $|q_i - t_i| < \varepsilon$. Soit alors λ_ε la bijection croissante affine par morceaux de $[0,\tau]$ dans lui-même définie par son graphe (cf. figure 1.2). Il existe des rationnels $r_1,\ldots,r_n$ tels que pour tout $i \in [\![1,n]\!]$ on ait : $|f(t_i) - r_i| < \varepsilon/2$. Soit maintenant g_ε la fonction définie par :

$$g_\varepsilon = \sum_{i=1}^{n} r_{i-1}\mathbb{1}_{[q_{i-1},q_i[},$$

de sorte que

$$g_\varepsilon \circ \lambda_\varepsilon^{-1} = \sum_{i=1}^{n} r_{i-1}\mathbb{1}_{[t_{i-1},t_i[}.$$

Alors, en prenant $\lambda = \lambda_\varepsilon^{-1}$ dans la définition de la métrique de Skorokhod, on remarque que $d_S(f,g_\varepsilon) \leqslant \varepsilon$ puisque $\|f_\varepsilon - g_\varepsilon\circ\lambda_\varepsilon^{-1}\|_\infty \leqslant \max_i |f(t_i) - r_i| \leqslant \varepsilon/2$ et que $\|\lambda_\varepsilon^{-1} - \mathrm{Id}\|_\infty \leqslant \varepsilon$. □

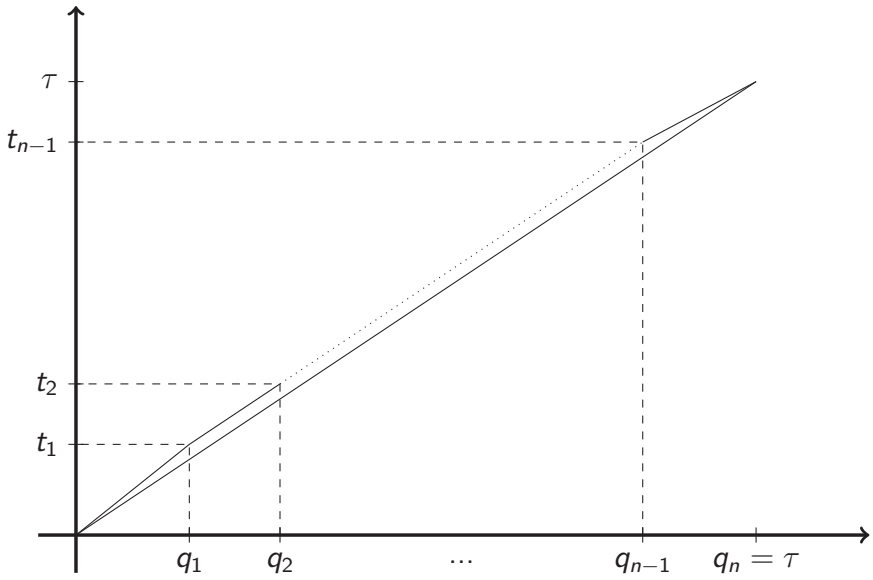


FIGURE 1.2 – Graphe de la fonction λ_ε et de la fonction identité.

Remarque 1.11. La démonstration précédente est encore valable en remplaçant $\mathbb{Q}$ par une partie dense de $[0, \tau]$ contenant 0 et τ.

Théorème 1.12

Une fonction X d'un ensemble $(\Omega, \mathcal{A})$ dans $(\mathcal{D}([0,\tau]), d_S)$ est mesurable pour la tribu borélienne si et seulement si pour tout $t \in \mathbb{Q} \cap [0,\tau]$, $X(t)$ est mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}, \mathcal{B}_\mathbb{R})$.

Nous proposons ici encore une démonstration légèrement différente de celle présentée par Billingsley (1968). De même que pour la proposition 1.10, on peut remplacer $\mathbb{Q}$ par une partie dense de $[0, \tau]$ contenant 0 et τ.

Notation 1.13. Soit $f \in \mathcal{D}([0,\tau])$ et $(q_k)_{k\in\mathbb{N}^\star}$ une énumération de $\mathbb{Q} \cap [0,\tau]$. On notera :

$$\psi_n(f) = \sum_{i=1}^{n} f(q_{(i-1)}) \mathbb{1}_{[q_{(i)}, q_{(i-1)}[},$$

où les $(q_{(i)})_i$ sont les $(q_i)_i$ réordonnés.

Lemme 1.14. *Soit $f \in \mathcal{D}([0,\tau])$. Alors, $\psi_n(f) \xrightarrow{d_S} f$ lorsque $n \to +\infty$.*

Démonstration. Soit $f \in \mathcal{D}([0,\tau])$ et $\varepsilon > 0$. On sait (cf. démonstration de la proposition 1.10) qu'il existe des rationnels $r_1, \dots, r_p$ tels que $d_S(f, g_\varepsilon) \leqslant \varepsilon$ où g_ε est la fonction donnée par :

$$g_\varepsilon = \sum_{i=1}^{p} f(r_{(i-1)}) \mathbb{1}_{[r_{(i-1)}, r_{(i)}[}.$$

On a aussi construit une fonction λ_ε telle que $f \circ \lambda_\varepsilon^{-1}$ oscille peu au sens où :

$$\forall i \in [\![1, p]\!] , \quad \sup_{(s,t)\in[r_{(i-1)},r_{(i)}[^2} |f \circ \lambda_\varepsilon^{-1}(s) - f \circ \lambda_\varepsilon^{-1}(t)| < \varepsilon.$$

Donc, si on affine la subdivision $0 = r_{(0)} < r_{(1)} < ... < r_{(p)} = 1$ en rajoutant un point r_{p+1}, alors, la fonction

$$h_\varepsilon = \sum_{i=1}^{p+1} f(r_{(i)}) \mathbb{1}_{[r_{(i-1)},r_{(i)}[}$$

vérifie $d_S(f, h_\varepsilon) \leqslant d_S(f, g_\varepsilon) \leqslant \varepsilon$. Or, comme la suite $(q_k)_{k\in\mathbb{N}^\star}$ est une énumération de $\mathbb{Q}\cap[0, \tau]$, à partir d'un certain rang n_0 les $(r_i)_i$ sont dans $\{q_1, ... , q_{n_0}\}$ et donc :

$$\forall n \geqslant n_0, \quad d_S(\psi_n(f), f) \leqslant \varepsilon.$$

□

Démonstration du théorème 1.12. Par mesurabilité des applications π_t, il est évident que si X est mesurable, alors $X(t)$ est mesurable pour tout t. Il suffit donc de montrer que l'application identité est mesurable en tant qu'application de $(\mathcal{D}([0, \tau]), \sigma(\pi_t, t \in \mathbb{Q}))$ dans $(\mathcal{D}([0, \tau]), \mathcal{B}_{d_S})$, où $\mathcal{B}_{d_S}$ désigne la tribu borélienne engendrée par les ouverts de d_S. Or, il est clair que les ψ_n sont cylindres-mesurables[2]. Donc, l'application identité est mesurable en tant que limite de fonctions mesurables. □

1.3 L'approche de Hoffmann-Jørgensen

Pour pallier les problèmes de mesurabilité, Hoffmann-Jørgensen a introduit le concept d'*espérance extérieure* qui s'applique à des fonctions non nécessairement mesurables. Il a ensuite défini la convergence en loi à l'aide des fonctions continues bornées sur $\mathcal{D}([0, \tau])$ muni de la norme infinie. Cette approche est encore valable en remplaçant $\mathcal{D}([0, \tau])$ par un espace métrique quelconque $(\mathcal{X}, d)$. Cependant, elle impose la mesurabilité du processus limite. De plus, sa définition dépend *a priori* de la tribu définie sur l'espace de départ.

Dans cette section, nous supposerons que $(\Omega, \mathcal{A}, \mathbb{P})$ est un espace probabilisé sur lequel sont définies une suite de fonctions $(X_n)_{n\in\mathbb{N}^\star}$ ainsi qu'une autre fonction X, toutes à valeurs dans un espace pseudo-métrique $(\mathcal{X}, d)$. Nous supposerons que X est mesurable lorsqu'on équipe $(\mathcal{X}, d)$ de la tribu borélienne.

2. Les cylindres sont les $\pi^{-1}_{t_1,...,t_k}(H)$ où $\pi_{t_1,...,t_k}(f) := (f(t_1), ... , f(t_k))$, où H est un borélien de $\mathbb{R}^k$ et où $(t_1, ... , t_k) \in \mathbb{Q}^k$.

1.3.1 Espérance extérieure et enveloppe mesurable

Définition 1.15 (Espérance extérieure)**.** L'espérance extérieure d'une fonction H définie de Ω dans $\mathbb{R}$ est le réel :

$$\mathbb{E}^\star H := \inf\{\mathbb{E}\, h\ ,\ \ H \leqslant h \ \text{ et } \ h \text{ intégrable}\}.$$

Nous pouvons maintenant donner la définition de convergence en loi au sens de Hoffmann-Jørgensen.

Définition 1.16 (Convergence en distribution)**.** On dit que la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge en distribution vers la loi de probabilité P et on écrit $X_n \rightsquigarrow P$ si :

$$\forall f \in \mathscr{C}_b((\mathcal{X}, d), \mathbb{R})\ ,\ \ \mathbb{E}^\star f(X_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}_P f.$$

Si X est une variable aléatoire définie de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $\mathcal{X}$, alors, on peut définir la convergence en distribution de la suite $(X_n)_{n\in\mathbb{N}^\star}$ vers X en remplaçant $\mathbb{E}_P f$ par $\mathbb{E}\, f(X)$ dans la définition précédente.

Cette définition permet de transporter la convergence en loi en composant par des applications continues.

Théorème 1.17

Supposons que $X_n \rightsquigarrow X$ et que $g \in \mathscr{C}(\mathcal{X}, \mathcal{X}')$, alors, $g(X_n) \rightsquigarrow g(X)$.

Pour pouvoir faire le lien entre les deux approches, nous allons voir plus en détails les concepts d'espérance extérieure, de *probabilité extérieure*, ainsi que les modes de convergence qu'ils induisent et les liens qui les unissent.

On peut définir la probabilité extérieure d'un événement non mesurable d'une manière similaire à celle employée pour définir l'espérance extérieure.

Définition 1.18 (Probabilité extérieure)**.** Soit $B \subset \Omega$ un ensemble non nécessairement mesurable. La *probabilité extérieure* de B est le réel :

$$\mathbb{P}^\star(B) := \inf\{\mathbb{P}(A),\ A \supset B \ \text{ et } \ A \in \mathcal{A}\}.$$

On remarque que, comme pour l'espérance extérieure, la probabilité extérieure dépend de la tribu de départ. De plus, l'infimum dans ces définitions est toujours atteint comme le montre le lemme suivant.

Lemme 1.19 (enveloppe mesurable)**.** *Soit $T\colon \Omega \to \bar{\mathbb{R}}$ une fonction quelconque. Alors, il existe une fonction mesurable $T^\star\colon \Omega \to \bar{\mathbb{R}}$ telle que :*

i. $T^\star \geqslant T$;

ii. Pour toute fonction $U\colon \Omega \to \bar{\mathbb{R}}$ mesurable, si $U \geqslant T$ presque sûrement, alors $U \geqslant T^\star$ presque sûrement.

De plus, une telle fonction $T^\star$ vérifie $\mathbb{E}^\star T = \mathbb{E}T^\star$ dès que $\mathbb{E}T^\star$ existe, ce qui est le cas si $\mathbb{E}^\star T < \infty$.

On peut montrer que si $T^\star$ et $T'^\star$ sont deux fonctions enveloppes d'une même fonction T, alors elles sont égales partout sauf éventuellement sur un ensemble de probabilité nulle. Par abus de notation, on écrira souvent $T^\star$ un représentant quelconque de la classe d'équivalence.

Définition 1.20. Soit $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ un espace probabilisé et ϕ une application de $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ dans $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que ϕ est *parfaite* si pour toute fonction bornée $T : \Omega \to \mathbb{R}$, $(T \circ \phi)^\star = T^\star \circ \phi$.

1.3.2 Convergence presque sûre et en probabilité

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de fonctions définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans un espace métrique $(\mathcal{X}, d)$, et soit X une fonction de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(\mathcal{X}, d)$.

Définition 1.21. On dit que :

1. $(X_n)_{n\in\mathbb{N}^\star}$ converge en probabilité extérieure vers X si $d(X_n, X)^\star \to 0$ en probabilité. Cela signifie que :
$$\forall \varepsilon > 0, \quad \mathbb{P}(d(X_n, X)^\star > \varepsilon) = \mathbb{P}^\star(d(X_n, X) > \varepsilon) \to 0.$$
On note $X_n \xrightarrow{\mathbb{P}^\star} X$.
2. $(X_n)_{n\in\mathbb{N}^\star}$ converge presque sûrement extérieurement vers X si il existe une version de $d(X_n, X)^\star$ qui converge presque sûrement vers 0. On note $X_n \xrightarrow{\text{p.s.}^\star} X$.
3. $(X_n)_{n\in\mathbb{N}^\star}$ converge presque sûrement vers X si $\mathbb{P}^\star(\{X_n \not\to X\}) = 0$. On note $X_n \xrightarrow{\text{p.s.}} X$.

Le concept de convergence presque sûre extérieure doit être examiné soigneusement. En effet, on peut construire (voir van der Vaart & Wellner (1996)) une suite convergente pour tout ω de Ω qui ne soit pas presque sûrement extérieurement convergente. Mais, la définition donnée ici de convergence presque sûre extérieure permet de retrouver les résultats classiques liant les différents types de convergences (extérieures). De plus, on peut démontrer la convergence presque sûre extérieure à partir de la convergence presque sûre sous certaines hypothèses (dans leur livre, van der Vaart & Wellner (1996) donnent des conditions nécessaires et suffisantes pour passer de l'une à l'autre).

Théorème 1.22

Supposons que X est mesurable par rapport à la tribu borélienne sur $(\mathcal{X}, d)$ et que $\mathbb{P}_X$ est séparable [3]. Si $(X_n)_{n\in\mathbb{N}^\star}$ converge presque sûrement vers X et que les X_n ainsi que X sont mesurables par rapport à la tribu engendrée par les boules, alors, il y a convergence presque sûre extérieure de la suite $(X_n)_{n\in\mathbb{N}^\star}$ vers X.

3. Signifie qu'il existe un ensemble mesurable séparable de probabilité 1 pour $\mathbb{P}_X$.

Lemme 1.23. *Supposons que X est mesurable par rapport à la tribu borélienne sur $(\mathcal{X}, d)$. Alors, la convergence presque sûre extérieure de $(X_n)_{n\in\mathbb{N}^\star}$ vers X entraîne la convergence en probabilité extérieure de $(X_n)_{n\in\mathbb{N}^\star}$ vers X.*

Lemme 1.24. *Supposons que X est mesurable par rapport à la tribu borélienne sur $(\mathcal{X}, d)$. Alors, la convergence en probabilité extérieure de la suite $(X_n)_{n\in\mathbb{N}^\star}$ vers X entraîne la convergence en loi de $(X_n)_{n\in\mathbb{N}^\star}$ vers X.*

Théorème 1.25 (Représentation presque sûre)

Supposons que $X_n \rightsquigarrow X$, où P_X est borélienne séparable. Alors, il existe un espace probabilisé $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ et des applications ϕ_n et ϕ mesurables parfaites de $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ dans $(\Omega, \mathcal{A})$ telles que les applications $\widetilde{X}_n := X_n \circ \phi_n$ et $\widetilde{X} := X \circ \phi$ vérifient :

1. $\widetilde{X}_n \xrightarrow{\text{p.s.}} \widetilde{X}$.
2. Les deux suites ont la même loi au sens où pour toute fonction bornée $f : \mathcal{X} \to \mathbb{R}$ et pour tout n : $\mathbb{E}^\star f(X_n) = \widetilde{\mathbb{E}}^\star f(\widetilde{X}_n)$. De même, $\mathbb{E}^\star f(X) = \widetilde{\mathbb{E}}^\star f(\widetilde{X})$.

1.4 Lien entre les deux méthodes

1.4.1 De Hoffmann-Jørgensen à Skorokhod

Supposons que $(X_n)_{n\in\mathbb{N}^\star}$ est une suite de variables aléatoires allant d'un ensemble $(\Omega, \mathcal{A}, \mathbb{P})$ dans $(\mathcal{D}([0,\tau]), d_S)$ muni de la tribu engendrée par les boréliens, où d_S est la distance de Skorokhod présentée à la section 1.2. Supposons de plus que la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge en loi vers une variable aléatoire X au sens de Hoffmann-Jørgensen, c'est-à-dire :

$$\forall f \in \mathscr{C}_b((\mathcal{D}([0,\tau]), \|\cdot\|_\infty), \mathbb{R}) \ , \quad \mathbb{E}^\star f(X_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\, f(X).$$

Alors, comme $\mathscr{C}_b((\mathcal{D}([0,\tau]), d_S), \mathbb{R}) \subset \mathscr{C}_b((\mathcal{D}([0,\tau]), \|\cdot\|_\infty), \mathbb{R})$, on a aussi :

$$\forall f \in \mathscr{C}_b((\mathcal{D}([0,\tau]), d_S), \mathbb{R}) \ , \quad \mathbb{E}^\star f(X_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\, f(X).$$

Mais, par construction de la métrique de Skorokhod, l'espérance extérieure ci-dessus est en fait une espérance. Ainsi, la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge aussi en distribution vers X au sens de Skorokhod.

1.4.2 De Skorokhod à Hoffmann-Jørgensen

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires définies sur un ensemble $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $(\mathcal{D}([0,\tau]), d_S)$ muni de la tribu engendrée par les boréliens, où d_S est la distance de Skorokhod présentée à la section 1.2.

Supposons que la suite $(X_n)_{n\in\mathbb{N}^\star}$ converge en loi vers une variable aléatoire X à trajectoires presque sûrement continues et séparable. Cette convergence est à comprendre au sens de Skorokhod :

$$\forall f \in \mathscr{C}_b((\mathcal{D}([0,\tau]), d_S), \mathbb{R}), \quad \mathbb{E}\, f(X_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\, f(X).$$

Nous allons montrer que la convergence en loi a encore lieu au sens de Hoffmann-Jørgensen en prenant $(\mathcal{D}([0,\tau]), \|\cdot\|_\infty)$ muni de la tribu borélienne comme espace d'arrivée. Pour ce faire, remarquons que nous pouvons appliquer le théorème de représentation presque sûr de Hoffmann-Jørgensen (théorème 1.25) à la suite $(X_n)_{n\in\mathbb{N}^\star}$. En effet, les X_n, en tant que variables aléatoires de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $(\mathcal{D}([0,\tau]), d_S)$ convergent en loi au sens usuel et donc aussi au sens de Hoffmann-Jørgensen. Il existe donc des applications ϕ_n et ϕ parfaites ainsi qu'un espace probabilisé $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{\mathbb{P}})$ tels que si on pose $\widetilde{X}_n := X_n \circ \phi_n$ et $\widetilde{X} := X \circ \phi_n$ alors :

1. $\widetilde{X}_n \to \widetilde{X}$ presque sûrement (pour d_S) ;
2. $\forall f \in \mathscr{C}_b((\mathcal{D}([0,\tau]), d_S), \mathbb{R}),\ \mathbb{E}\, f(X_n) = \widetilde{\mathbb{E}}\, f(\widetilde{X}_n)$.

Remarquons que par mesurabilité des applications en jeu, les espérances extérieures du théorème de représentation presque sûr deviennent des espérances. De plus, le processus $\widetilde{X}$ étant presque sûrement à trajectoires continues, la convergence presque sûre pour d_S implique (d'après la proposition 1.6) la convergence presque sûre pour la norme infinie :

$$\|\widetilde{X}_n - \widetilde{X}\|_\infty \to 0 \quad \text{presque sûrement.}$$

De plus, on sait que les $\widetilde{X}_n$ ainsi que $\widetilde{X}$ sont mesurables par rapport à la tribu engendrée par les boules. Donc, d'après le théorème 1.22, la convergence presque sûre entraîne la convergence presque sûre extérieure qui implique elle-même la convergence en loi (au sens de Hoffmann-Jørgensen : cf lemmes 1.23 et 1.24). Ainsi :

$$\forall f \in \mathscr{C}_b((\mathcal{D}([0,\tau]), \|\cdot\|_\infty), \mathbb{R}), \quad \widetilde{\mathbb{E}}\, f(\widetilde{X}_n) \underset{n\to+\infty}{\longrightarrow} \widetilde{\mathbb{E}}\, f(\widetilde{X}). \tag{1.3}$$

Pour conclure, il suffit de montrer que l'équation (1.3) est encore vraie « sans les tildes ». Soit alors $f \in \mathscr{C}_b((\mathcal{D}([0,\tau]), \|\cdot\|_\infty), \mathbb{R})$ et $n \in \mathbb{N}^\star$. Comme ϕ_n est parfaite :

$$\widetilde{\mathbb{E}}^\star f(\widetilde{X}_n) = \int f(X_n \circ \phi_n)^\star \mathrm{d}\widetilde{\mathbb{P}} = \int (f(X_n) \circ \phi_n)^\star \mathrm{d}\widetilde{\mathbb{P}} = \int f(X_n)^\star \circ \phi \,\mathrm{d}\widetilde{\mathbb{P}} = \mathbb{E}^\star f(X_n).$$

1.5 L'approche de Pollard

Dans son livre, Pollard (1990) définit la convergence en loi d'une manière similaire à celle de Hoffmann-Jørgensen, mais en utilisant des fonctions test bornées uniformément continues. Cela lui permet de simplifier quelques démonstration et il montre que les définitions se rejoignent dans les cas « importants », c'est-à-dire dès lors que la loi limite est borélienne et concentrée sur un espace séparable.

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de processus indexés par un espace pseudo-métrique $(\mathcal{X}, d)$. Si on note $\mathcal{U}_b(\mathcal{X})$ l'ensemble des fonctions uniformément continues bornées de $(\mathcal{X}, d)$ dans $\mathbb{R}$, alors, la convergence en loi au sens de Pollard de la suite $(X_n)_{n\in\mathbb{N}^\star}$ vers une fonction borélienne de $(\Omega, \mathcal{A})$ dans $\mathcal{D}(\mathcal{X})$ est définie par :

$$\forall f \in \mathcal{U}_b((\mathcal{X}, d), \mathbb{R}) \ , \quad \mathbb{E}^\star f(X_n) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}\, f(X).$$

Pour démontrer son théorème de la limite centrale fonctionnel, Pollard (1990) utilise une caractérisation de la convergence en loi pour les processus (cf. théorème 1.26). On retrouve cette même caractérisation dans van der Vaart & Wellner (1996) qui décrit l'approche de Hoffmann-Jørgensen, ce qui montre que les deux définitions se rejoignent dès lors que la loi limite est borélienne séparable.

Nous avons expliqué dans l'introduction de ce chapitre que la convergence des marges fini-dimensionnelles n'était pas suffisante pour que la définition de la convergence en loi soit satisfaisante. Rappelons que cela signifie que pour tout k, pour tout $t_1, \dots, t_k$ dans T, il existe une mesure de probabilité borélienne $P(\cdot|t_1, \dots, t_k)$ sur $\mathbb{R}^k$ telle que :

$$(X_n(\cdot, t_1), \dots, X_n(\cdot, t_k)) \xrightarrow{\mathcal{L}} P(\cdot|t_1, \dots, t_k). \tag{1.4}$$

Il existe un lien entre la convergence des marges fini-dimensionnelles et la définition de la convergence en loi au sens de Hoffmann-Jørgensen (ou au sens de Pollard) comme le montre le théorème ci-dessous que l'on retrouve dans Pollard (1990) ainsi que dans van der Vaart & Wellner (1996).

Théorème 1.26

Soit $(X_n(\cdot))_{n\in\mathbb{N}^\star}$ une suite de processus indexés par un espace pseudo-métrique précompact (T, ρ). Supposons que :

(i) les marges fini-dimensionnelles convergent comme dans (1.4) ;

(ii) Pour tout $\varepsilon > 0$ et pour tout $\eta > 0$, il existe $\delta > 0$ tel que :

$$\overline{\lim_n}\ \mathbb{P}^\star \left(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \eta \right) < \varepsilon. \tag{1.5}$$

Alors, il existe une mesure de probabilité borélienne P dont les projections fini-dimensionnelles sont les $P(\cdot|t_1, \dots, t_k)$ telle que X_n converge en loi vers P. Réciproquement, si X_n converge vers une mesure de probabilité borélienne P séparable, alors, il existe une semi-métrique ρ rendant T précompact et telle que (i) et (ii) sont satisfaites.

On dira qu'une suite de processus $(X_n)_{n\in\mathbb{N}^\star}$ est *asymptotiquement équicontinue* si elle vérifie (1.5). Dans son livre, Billingsley (1968) montre un résultat similaire pour la convergence en loi au sens de Skorokhod. Notons

que la difficulté n'est pas la même suivant l'approche choisie. En effet, avec la convergence en loi au sens de Hoffmann-Jørgensen, le point (i) est évident grâce au théorème de transport alors que le point (ii) pose une difficulté. Au contraire, la convergence en loi au sens de Skorokhod permet de montrer simplement le point (ii) en utilisant le fait que $\mathcal{D}([0,\tau])$ est un espace polonais [4] alors que le point (i) pose problème puisque les projections sont mesurables (cf lemme 1.9) mais pas nécessairement continues : on ne peut donc plus utiliser le théorème de transport.

1.6 Discussion

Nous avons vu que l'approche de Hoffmann-Jørgensen était valable pour des processus indexés par un espace pseudo-métrique $(\mathcal{X}, d)$: elle s'adapte donc très bien à l'étude du processus empirique. Par contre, on a construit la métrique de Skorokhod sur $\mathcal{D}([0,\tau])$ et elle n'est donc pas *a priori* adaptée pour l'étude du processus empirique. Pollard (1984) définit une métrique de Skorokhod sur $\mathcal{D}([0,+\infty[)$ et on peut l'étendre de manière canonique à $\mathcal{D}(\mathbb{R})$.

Il n'existe pas, à notre connaissance, de généralisation adéquate de la métrique de Skorokhod pour des processus indexés par des espaces plus complexes que $\mathbb{R}^K$. Cependant, nous n'utiliserons dans notre travail que des résultats portant sur la convergence de processus dans l'espace $\mathcal{D}([0,\tau])$ ou $\mathcal{D}([0,\tau])^K$.

1.7 Extension au cas vectoriel

Nous aurons dans la suite affaire à des processus à valeurs vectorielles. La définition de la convergence en loi au sens de Hoffmann-Jørgensen donnée précédemment est encore valide pour des processus à valeurs dans $\mathbb{R}^K$, où K est un entier naturel non nul [5]. De plus, il existe une caractérisation à l'aide de l'équicontinuité asymptotique et de la convergence des marges fini-dimensionnelles (cf chapitre 1.8 de van der Vaart & Wellner (1996)).

Soit $(X_n(\cdot))_{n\in\mathbb{N}^*}$ une suite de processus à valeurs dans $\mathbb{R}^K$. Notons X_n^k les coordonnées du vecteur X_n. Dire alors que les marges fini-dimensionnelles de la suite de processus convergent signifie que pour tout p, pour tout $t_1, \dots, t_p$ dans T, il existe une mesure de probabilité borélienne $P(\cdot|t_1,\dots,t_p)$ sur $\mathbb{R}^p$ telle que :

$$(X_n^1(\cdot,t_1),\dots,X_n^K(\cdot,t_1),\dots,X_n^1(\cdot,t_p),\dots,X_n^K(\cdot,t_p)) \xrightarrow{\mathcal{L}} P(\cdot|t_1,\dots,t_p). \qquad (1.6)$$

4. Un espace polonais est un espace métrique séparable (E,d) tel qu'il existe une distance d' équivalente à d et rendant E complet.

5. Un processus f à valeurs dans $\mathbb{R}^K$ indexé par T est une application de $\Omega \times T$ dans $\mathbb{R}^K$ telle que $f(\cdot,t)$ est borélienne pour tout t de T.

Avec ces nouvelles définitions, on peut maintenant énoncer un résultat similaire au théorème 1.26 qui lie la convergence en distribution à la convergence des marges fini-dimensionnelles.

Théorème 1.27

Soit $(X_n(\cdot))_{n\in\mathbb{N}^\star}$ une suite de processus à valeurs dans $\mathbb{R}^K$ indexés par un espace pseudo-métrique précompact (T,ρ). Supposons que :

(i) les marges fini-dimensionnelles convergent comme dans (1.6) ;

(ii) Pour tout $\varepsilon > 0$ et pour tout $\eta > 0$, il existe $\delta > 0$ tel que :

$$\overline{\lim_n}\ \mathbb{P}^\star\left(\sup_{\rho(s,t)<\delta}\max_{1\leqslant k\leqslant K}|X_n^k(s)-X_n^k(t)|>\eta\right)<\varepsilon.$$

Alors, il existe une mesure de probabilité borélienne P dont les projections fini-dimensionnelles sont les $P(\cdot|t_1,\dots,t_p)$ telle que $(X_n)_{n\in\mathbb{N}^\star}$ converge en loi vers P. Réciproquement, si $(X_n)_{n\in\mathbb{N}^\star}$ converge vers une mesure de probabilité borélienne P séparable, alors, il existe une semi-métrique ρ rendant T précompact et telle que (i) et (ii) sont satisfaites.

CHAPITRE 2

Processus Empiriques

2.1 Introduction

Dans ce chapitre, nous reprenons les résultats présentés dans Pollard (1990) qui seront fondamentaux dans la suite de notre travail. Nous allons revoir le cheminement, classique en théorie des processus empiriques, qui mène à des résultats de type loi des grands nombres uniforme et théorème de la limite centrale fonctionnel. Nous donnerons les éléments clefs des preuves ainsi que des démonstrations simplifiées et adaptées aux cas auxquels nous aurons affaire dans la suite. Nous ferons aussi un lien avec la théorie de Vapnik et Chervonenkis.

Dans tout le chapitre, on considérera une suite de processus $(f_i)_{i\in\mathbb{N}^\star}$ définis sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ et indexés par un ensemble T. On notera :

$$S_n(\omega, t) := \sum_{i=1}^{n} f_i(\omega, t),$$
$$M_n(t) := \mathbb{E}\, S_n(\cdot, t).$$

Lorsque T est fini, on connaît grâce à la loi des grands nombres des conditions suffisantes simples pour lesquelles $\frac{1}{n}S_n$ converge presque sûrement et dans $L^1(\Omega)$. De même, on connaît des conditions suffisantes simples pour la convergence en loi de $n^{-1/2}(S_n - M_n)$. La question ici est de savoir si ces convergences sont encore satisfaites au sens suivant :

$$\sup_{t\in T} \frac{1}{n} |S_n(\cdot, t) - M_n(t)| \overset{\mathbb{P},\,\text{p.s.}}{\longrightarrow} 0 \qquad \text{lorsque } n \to +\infty, \tag{2.1}$$
$$n^{-1/2}(S_n - M_n) \rightsquigarrow G \qquad \text{lorsque } n \to +\infty.$$

Pour pouvoir parler de convergence presque sûre ou en probabilité dans (2.1), il faut s'assurer de la mesurabilité du sup. Nous supposerons dans toute la suite qu'il est mesurable, ce qui sera le cas dès que les processus en

jeu seront séparables (cf. définition 1.2). Le but des sections suivantes est de borner la variable aléatoire $\sup_{t\in T} |S_n(\cdot, t) - M_n(t)|$. Pour ce faire, Pollard (1990) procède en plusieurs étapes que nous allons reprendre ici de manière succincte.

2.2 Symétrisation et conditionnement

La première étape, que l'on retrouve aussi dans van der Vaart & Wellner (1996) pour l'étude des classes de Vapnik–Chervonenkis, consiste à ajouter de l'aléa au problème. Plutôt que d'étudier directement $S_n(\omega, t)$, on introduit une suite $(\sigma_i)_{i\in\mathbb{N}^\star}$ de variables aléatoires de Rademacher symétriques [1] mutuellement indépendantes et indépendantes des f_i, et on définit :

$$S_n^\circ(\omega, t) := \sum_{i=1}^{n} \sigma_i f_i(\omega, t).$$

Cette *symétrisation* du processus S_n permet de conditionner par les f_i et d'utiliser les propriétés de sous-gaussianité des processus de Rademacher [2]. Ceci servira à transposer le problème à celui de l'étude des propriétés géométriques des ensembles $\mathcal{F}_{n\omega} \subset \mathbb{R}^n$ définis par :

$$\mathcal{F}_{n\omega} := \Big\{ \big(f_1(\omega, t), \dots, f_n(\omega, t)\big),\ t \in T \Big\}.$$

En effet, Pollard (1990) montre le théorème suivant :

Théorème 2.1 (Pollard)

Notons $\sigma = (\sigma_1, \dots, \sigma_n)^t$ et $f = (f_1, \dots, f_n)^t$. Alors, pour toute fonction convexe croissante ϕ, on a :

$$\mathbb{E}\,\phi\left(\sup_{t\in T} |S_n(\cdot, t) - M_n(t)|\right) \leqslant \mathbb{E}_\omega\,\mathbb{E}_\sigma\,\phi\left(2 \sup_{f\in\mathcal{F}_{n\omega}} |\langle \sigma, f\rangle|\right).$$

La démonstration de ce théorème fait appel à l'inégalité de Jensen ainsi qu'aux propriétés de $(\sigma_1, \dots, \sigma_n)$, et en particulier au fait que la loi de ce n-uplet est invariante par permutation.

Nous verrons par la suite qu'en utilisant à nouveau les propriétés des variables de Rademacher, il nous suffira de donner des conditions sur les $\mathcal{F}_{n\omega}$ uniformément en n et en ω pour trouver les résultats souhaités. Ce qui sera important sera alors la « taille » des $\mathcal{F}_{n\omega}$ ou plutôt à quel point ils « occupent l'espace » en un sens à définir.

1. Signifie que $\mathbb{P}(\sigma_i = 1) = \mathbb{P}(\sigma_i = -1) = 1/2$.

2. Un *processus de Rademacher* est un processus de la forme $X(a) := \sum_{i=1}^n a_i\sigma_i$, indexé par une partie A de $\mathbb{R}^n$.

Remarquons que si on intègre le coefficient 2 dans la fonction ϕ, le problème devient alors celui de la recherche de bornes pour :

$$\mathbb{E}_\sigma \, \phi \left(\sup_{f \in \mathcal{F}_{n\omega}} |\langle \sigma, f \rangle| \right).$$

2.3 Chaînage

Avant d'entrer dans le vif du sujet, il est nécessaire de faire quelques rappels sur les normes d'Orlicz.

Définition 2.2. Soit ϕ une fonction convexe strictement croissante telle que $\phi(0) \in [0, 1[$. La ϕ-norme d'Orlicz d'une variable aléatoire Z est définie par :

$$\|Z\|_\phi = \inf \left\{ C > 0 \ : \ \mathbb{E}\, \phi(|Z|/C) \leqslant 1 \right\},$$

où l'infimum peut être éventuellement égal à $+\infty$.

On démontre que l'application qui à une variable aléatoire Z associe $\|Z\|_\phi$ est une semi-norme en tant que jauge d'un convexe non vide. On peut remarquer que les normes d'Orlicz associées aux fonctions $x \mapsto x^p$ correspondent aux normes $\|\cdot\|_p$ usuelles.

Dans tout ce chapitre, on notera ψ la fonction convexe croissante définie pour tout x de $\mathbb{R}$ par $\psi(x) = e^{x^2}/5$. Dire que la ψ-norme d'Orlicz d'une variable aléatoire Z est finie signifie en particulier que Z est sous-gaussienne. Plus précisément, ses queues de distribution décroissent au moins aussi vite que $5e^{-t^2/C}$, où $C = \|Z\|_\psi$. En effet, en utilisant l'inégalité de Markov, on a :

$$\begin{aligned} \mathbb{P}(|Z| \geqslant t) &= \mathbb{P}(e^{Z^2/C^2} \geqslant e^{t^2/C^2}) \\ &\leqslant \frac{\mathbb{E}\, e^{Z^2/C^2}}{e^{t^2/C^2}} \\ &\leqslant 5\, e^{-t^2/C^2}. \end{aligned}$$

De plus, on montre (cf. van der Vaart & Wellner (1996)) le résultat suivant qui nous sera utile dans la suite.

Proposition 2.3. *Soit $p \geqslant 1$. Il existe une constante C_p telle que :*

$$\|\cdot\|_p \leqslant C_p \|\cdot\|_\psi.$$

Lemme 2.4 (Pollard)**.** *Soient $Z_1, \ldots, Z_m$ des variables aléatoires. Alors :*

$$\| \max_{i \leqslant m} |Z_i| \|_\psi \leqslant \sqrt{2 + \ln m} \max_{i \leqslant m} \|Z_i\|_\psi.$$

La forte croissance de la fonction ψ permet d'obtenir une inégalité bien plus fine que celle obtenue par inégalité triangulaire, à savoir :

$$\| \max_{i \leqslant m} |Z_i| \|_\psi \leqslant m \max_{i \leqslant m} \|Z_i\|_\psi.$$

Ainsi, si l'espace des temps T était fini, on pourrait appliquer directement le lemme 2.4 pour obtenir une borne sur $\|\sup_{\mathcal{F}_{n\omega}} |\langle \sigma, f \rangle|\|_\psi$. Ce ne sera évidemment pas le cas en général. Une manière de contourner le problème est d'utiliser une technique dite de *chaînage*. Considérons alors un espace pseudo-métrique séparable (T, d) et introduisons deux nouvelles définitions.

Définition 2.5 (nombre de saturation)**.** Le *nombre de saturation* (ou *packing number* en anglais) $D(\varepsilon, T_0)$ d'un sous-espace T_0 de T est le plus grand entier m pour lequel il existe m points de T_0 qui soient ε-écartés (c'est-à-dire qu'il existe $t_1, \dots, t_m$ dans T_0 tels que pour tout $i \neq j$, $d(t_i, t_j) > \varepsilon$).

Définition 2.6 (nombre de recouvrement)**.** Le *nombre de recouvrement* (ou *covering number* en anglais) $N(\varepsilon, T_0)$ d'un sous-espace T_0 de T est le plus petit nombre de boules de rayon ε nécessaires pour recouvrir T_0 (les centres des boules n'étant pas nécessairement des points de T_0).

Les concepts de nombre de saturation et de nombre de recouvrement sont très proches puisque :

$$N(\varepsilon, T_0) \leqslant D(\varepsilon, T_0) \leqslant N(\varepsilon/2, T_0).$$

On les retrouve encore dans l'approche de Vapnik–Chervonenkis.

Supposons maintenant que nous sommes en présence d'un processus Z indexé par (T, d) et ayant des trajectoires continues, au sens où pour tout ω, les fonctions $Z(\omega, \cdot)$ sont continues de T dans $\mathbb{R}$. Dans notre cas, le processus Z est en fait $Z(\sigma, f) = \langle \sigma, f \rangle$, et Z est clairement une fonction continue en f à σ fixé, par continuité à droite du produit scalaire (f et σ étant des éléments de $\mathbb{R}^n$).

Lemme 2.7 (Pollard)**.** *Supposons que Z est à trajectoires continues et que :*

$$\forall (s, t) \in T^2, \quad \|Z(s) - Z(t)\|_\psi \leqslant d(s, t).$$

Soit $t_0 \in T$ et $\delta := \sup_{t \in T} d(t, t_0)$. Alors :

$$\|\sup_{t \in T} |Z(t)|\|_\psi \leqslant \|Z(t_0)\|_\psi + \sum_{i=0}^{+\infty} \frac{\delta}{2^i} \sqrt{2 + \ln D(\delta/2^{i+1}, T)}.$$

Démonstration. Ce résultat étant essentiel pour la suite, reprenons ici les grandes lignes de la démonstration présentée dans Pollard (1990). L'idée est d'utiliser la continuité des trajectoires du processus et de l'approcher sur une grille de plus en plus fine. Notons $\delta_i = \delta/2^i$, et supposons que le membre de droite dans l'inégalité est fini, sans quoi le résultat est trivial.
On construit d'abord par récurrence une suite croissante de sous-ensembles finis $\{t_0\} =: T_0 \subseteq T_1 \subseteq \dots$ vérifiant :

1. Pour tout t dans T : $\min_{t^\star \in T_i} d(t, t^\star) \leqslant \delta_i$.
2. Les points de T_i sont δ_i-écartés.

Par définition du nombre de saturation, on sait que $\sharp T_i \leqslant D(\delta_i, T) =: m_i$. En utilisant l'inégalité triangulaire ainsi que le lemme 2.4 on montre que :

$$\| \max_{t\in T_{k+1}} |Z(t)| \|_\psi \leqslant \| \max_{t\in T_k} |Z(t)| \|_\psi + \delta_k \sqrt{2 + \ln m_{k+1}}.$$

On applique ensuite cette inégalité k fois. Apparaît alors le terme $\|Z(t_0)\|_\psi$ plus une somme. Puis, en faisant tendre k vers $+\infty$, on obtient un ensemble T_∞ dénombrable et dense dans T. On démontre alors le résultat souhaité en utilisant le théorème de convergence monotone ainsi que la continuité des trajectoires, puisque :

$$\| \max_{t\in T_{k+1}} |Z(t)| \|_\psi \nearrow \| \sup_{t\in T_\infty} |Z(t)| \|_\psi = \| \sup_{t\in T} |Z(t)| \|_\psi.$$

□

Pollard (1990) montre ensuite le lemme 2.8 ci-dessous. Il permet de faire disparaître les variables de Rademacher précédemment introduites pour un coût moindre (un coefficient 2) et de travailler dorénavant avec la norme euclidienne des éléments de $\mathcal{F}$.

Lemme 2.8 (Pollard)**.** *Pour tout f dans $\mathbb{R}^n$, on a : $\|\langle \sigma, f\rangle\|_\psi \leqslant 2\|f\|_2$. Par conséquent, pour tout $(f, g) \in \mathbb{R}^n \times \mathbb{R}^n$:*

$$\|\langle \sigma, f - g\rangle\|_\psi \leqslant \|f - g\|_2.$$

Ce résultat montre en particulier qu'un processus de Rademacher est sous-gaussien[3]. Les lemmes 2.7 et 2.8 permettent maintenant de borner le $\sup_{f\in\mathcal{F}} |\langle \sigma, f\rangle|$. Pollard (1990) utilise encore une inégalité élémentaire ainsi qu'une comparaison série/intégrale (cf. figure 2.1) pour obtenir l'inégalité énoncée au théorème 2.9.

Théorème 2.9 (Pollard)

Pour tout sous-espace $\mathcal{F}$ de $\mathbb{R}^n$,

$$\| \sup_{f\in\mathcal{F}} |\langle \sigma, f\rangle| \|_\psi \leqslant 9 \int_0^\delta \sqrt{\ln D(x, \mathcal{F})}\, dx, \qquad \text{où} \quad \delta = \sup_{f\in\mathcal{F}} \|f\|_2.$$

2.4 Inégalité maximale

On peut maintenant regrouper les résultats vus aux sections précédentes, en prenant soin cette fois-ci de ne pas omettre la dépendance en ω. Notons $L_n(\sigma, \omega) = \sup_{t\in T} |\sum_{i=1}^n \sigma_i f_i(\omega, t)|$. Rappelons que :

$$\|L_n(\sigma, \omega)\|_\psi = \| \sup_{f\in\mathcal{F}_{n\omega}} |\langle \sigma, f\rangle| \|_\psi \leqslant J_n(\omega) := 9 \int_0^{\delta_n(\omega)} \sqrt{\ln D(x, \mathcal{F}_{n\omega})}\, dx,$$

3. Un processus $X(\cdot)$ est dit sous-gaussien si $\mathbb{P}(|X(s) - X(t)| \geqslant x) \leqslant C \exp\left(\frac{-x^2}{2\|s-t\|_2^2}\right)$.

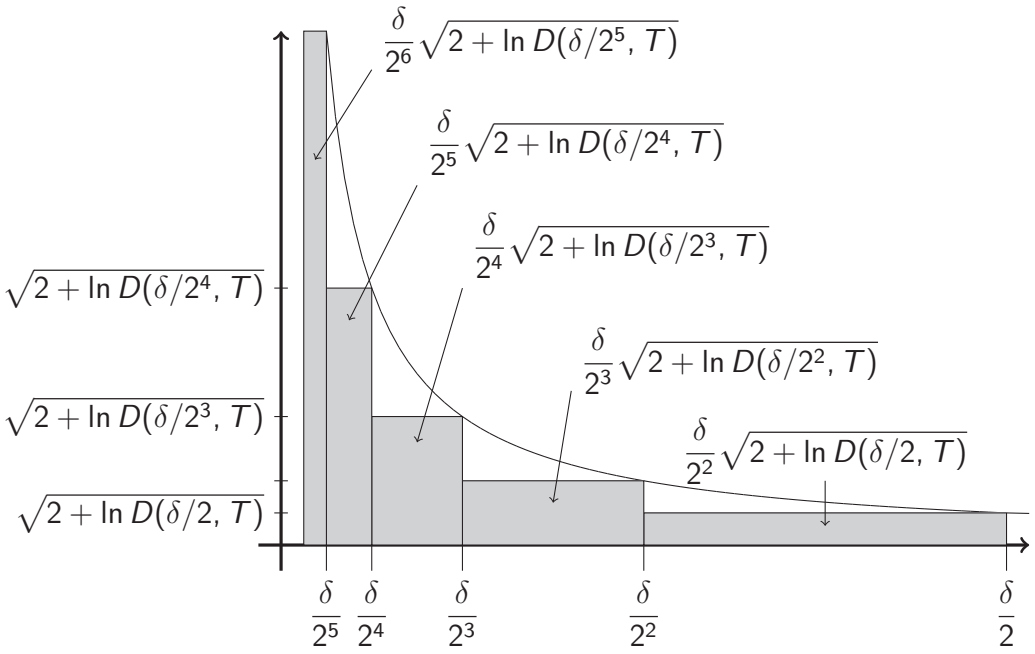


FIGURE 2.1 – Comparaison série/intégrale.

où $\delta_n(\omega) = \sup_{f\in\mathcal{F}_{n\omega}} \|f\|_2$. Donc, par définition de la ψ-norme, on en déduit que, presque sûrement :

$$\mathbb{E}_\sigma \exp\left(\left(\frac{L_n(\omega,\sigma)}{J_n(\omega)}\right)^2\right) \leqslant 5.$$

On remarque que le comportement de $J_n(\omega)$ va jouer un rôle important dans la suite. Rappelons que les classes $\mathcal{F}_{n\omega}$ s'écrivent :

$$\mathcal{F}_{n\omega} = \Big\{ \big(f_1(\omega,t),\dots,f_n(\omega,t)\big),\ t\in T \Big\}.$$

On notera $F_n := \sup_{t\in T} f_n(\cdot,t)$ et on dira que les F_n sont des *enveloppes*[4] des f_n. Supposons maintenant qu'il existe une constante C et des fonctions $\lambda_n\colon [0,1]\to\mathbb{R}$ telles que :

1. Les F_n sont presque sûrement bornées par C ;
2. $D\big(x\big\|\big(F_1(\omega),\dots,F_n(\omega)\big)^t\big\|_2, \mathcal{F}_{n\omega}\big) \leqslant \lambda_n(x)$ presque sûrement ;
3. $\Lambda_n(t) := \int_0^t \sqrt{\ln \lambda_n(x)}\,dx$ est bien définie et est finie pour tout $t\in[0,1]$.

Alors, après un simple changement de variable, on obtient :

$$\begin{aligned} J_n(\omega) &= 9\int_0^{\delta_n(\omega)} \sqrt{\ln D(x,\mathcal{F}_{n\omega})}\,dx \\ &\leqslant 9\,\big\|\big(F_1(\omega),\dots,F_n(\omega)\big)^t\big\|_2\,\Lambda_n\left(\frac{\delta_n(\omega)}{\big\|\big(F_1(\omega),\dots,F_n(\omega)\big)^t\big\|_2}\right) \qquad (2.2)\\ &\leqslant \sqrt{n}\,C\Lambda_n(1). \qquad (2.3)\end{aligned}$$

4. De manière générale, on dira que F_n est une enveloppe de f_n si $\sup_{t\in T}|f_n(\omega,t)| \leqslant F_n(\omega)$.

Souvenons-nous maintenant de la proposition 2.3 et appliquons-la à L_n avec $p = 1$. On obtient :

$$\mathbb{E}_\sigma |L_n(\sigma, \omega)| \leqslant C_1 J_n(\omega).$$

Ainsi :

$$\mathbb{E} \sup_{t \in T} |S_n(\omega, t) - M_n(t)| \leqslant 2C_1 \, \mathbb{E} \, J_n.$$

D'où, en réinjectant dans les équations (2.2) et (2.3) :

$$\mathbb{E} \sup_{t \in T} |S_n(\omega, t) - M_n(t)| \leqslant \sqrt{n} \, C_2 \, \mathbb{E} \left(\Lambda_n \left(\frac{\delta_n}{\left\| \left(F_1(\omega), \dots, F_n(\omega) \right)^t \right\|_2} \right) \right) \tag{2.4}$$

$$\leqslant C_3 \sqrt{n}. \tag{2.5}$$

Ces équations nous serviront plus tard pour expliquer le théorème de la limite centrale fonctionnel ainsi que la loi des grands nombres uniforme.

Définition 2.10 (paramètre d'échelle)**.** On appellera paramètre d'échelle tout vecteur de $\mathbb{R}^n$ dont toutes les coordonnées sont positives.

Notation 2.11. Si $\alpha = (\alpha_1, \dots, \alpha_n)$ est un paramètre d'échelle de $\mathbb{R}^n$ et que $f \in \mathbb{R}^n$, on notera $\alpha \odot f$ le vecteur dont les coordonnées sont les $\alpha_i f_i$.

Pollard (1990) donne alors la définition suivante de *processus manageables* pour lesquels il montrera une loi des grands nombres uniforme ainsi qu'un théorème central limite fonctionnel. On reconnaît dans cette définition des éléments qui nous ont permis d'obtenir les inégalité (2.4) et (2.5).

Définition 2.12 (processus *manageable*)**.** Une suite $(f_n)_{n \in \mathbb{N}^\star}$ de processus est dite *manageable* (par rapport aux enveloppes F_n) si il existe une fonction λ telle que :

1. $\int_0^1 \sqrt{\ln \lambda(x)} \, \mathrm{d}x < \infty$,
2. pour tout $x \in]0, 1]$, pour tout ω, pour tout n et pour tout paramètre d'échelle α, on ait :

$$D\left(x \left\| \alpha \odot \left(F_1(\omega), \dots, F_n(\omega) \right) \right\|_2, \alpha \odot \mathcal{F}_{n\omega}\right) \leqslant \lambda(x).$$

Définition 2.13 (processus euclidiens)**.** Si dans la définition précédente de processus *manageable* la fonction λ peut s'écrire $\lambda(x) = Ax^{-W}$, où A et W sont des constantes strictement positives, on parlera alors de *processus euclidiens*.

2.5 Pseudo-dimension et classes euclidiennes

On a vu à la section précédente que les processus euclidiens avaient de bonnes propriétés (cf. équations (2.4) et (2.5) par exemple). Il nous faut maintenant développer un outil qui nous permette de montrer simplement qu'une classe est euclidienne. L'approche de Pollard consiste à projeter les classes $\mathcal{F} \subset \mathbb{R}^n$ sur $\mathbb{R}^k$ ($k \leqslant n$) et à regarder « à quel point » elles occupent l'espace, en un sens à définir.

Définition 2.14 (orthant)**.** Pour tout $t \in \mathbb{R}^k$ et pour tout ensemble d'indices $J \subset [\![1, k]\!]$, on définit le *J-ième orthant autour de t* comme étant l'ensemble des éléments $x \in \mathbb{R}^k$ tels que :

$$\begin{aligned} x_i > t_i \quad & \text{si } i \in J, \\ x_i < t_i \quad & \text{si } i \in J^C. \end{aligned}$$

Définition 2.15. On dira qu'un sous-ensemble A de $\mathbb{R}^k$ *occupe* le J-ième orthant autour de t s'il contient au moins un élément dans cet orthant. On dira que t est *cerné* par A si les 2^k orthants autour de t sont tous occupés par A.

Définition 2.16 (projection propre)**.** On appelle *projection propre* toute application de $\mathbb{R}^n$ dans $\mathbb{R}^k$ ($k \leqslant n$) qui peut s'écrire $p(x_1, \dots, x_n) = (x_{i(1)}, \dots, x_{i(k)})$, où les $i(j)$ sont des entiers distincts compris entre 1 et n.

Exemple 2.17. L'application définie ci-dessous est une projection propre :

$$\begin{array}{rccc} p & : & \mathbb{R}^4 & \to & \mathbb{R}^2 \\ & & (x, y, z, t)^t & \mapsto & (z, x)^t. \end{array}$$

Définition 2.18 (pseudo-dimension)**.** On dit que $\mathcal{F} \subset \mathbb{R}^n$ est de *pseudo-dimension au plus k* si pour tout point $u \in \mathbb{R}^{k+1}$ et pour toute projection propre $p : \mathbb{R}^n \to \mathbb{R}^{k+1}$, u n'est pas cerné par $p(\mathcal{F})$.

Remarque 2.19. Comme le remarque Pollard (1990), si $\mathcal{F}$ et $\mathcal{G}$ sont de pseudo-dimension finie, alors $\mathcal{F} \wedge \mathcal{G}$, $\mathcal{F} \vee \mathcal{G}$ et $\mathcal{F} \cup \mathcal{G}$ sont aussi de pseudo-dimension finie[5]. Cependant, on ne peut rien dire *a priori* sur la pseudo-dimension de $\mathcal{F} + \mathcal{G}$ ou de $\mathcal{F}\mathcal{G}$. Énonçons ici un autre résultat de stabilité que nous n'avons pas trouvé dans la littérature.

Théorème 2.20

Soient $f_1, \dots, f_n$ des fonctions d'un ensemble T dans $\mathbb{R}$. Définissons :

$$\begin{aligned} \mathcal{F} &:= \Big\{ \big(f_1(t), \dots, f_n(t)\big),\ t \in T \Big\}, \\ \mathcal{G} &:= \Big\{ \big(\mathbb{1}_{\{f_1(t)>0\}}, \dots, \mathbb{1}_{\{f_n(t)>0\}}\big),\ t \in T \Big\}. \end{aligned}$$

Si $\mathcal{F}$ est de pseudo-dimension k, alors il en est de même pour $\mathcal{G}$.

Démonstration. Soit p une projection propre à valeurs dans $\mathbb{R}^{k+1}$. Il existe un orthant autour de l'origine qui n'est pas occupé par $p(\mathcal{F})$. Donc :

$$p(\mathcal{G}) \neq \big\{0; 1\big\}^{k+1}. \tag{2.6}$$

On en déduit en particulier que $\mathcal{G}$ est de pseudo-dimension au plus k. En effet, soit $u \in \mathbb{R}^{k+1}$, il s'agit de montrer que tous les orthants autour de u ne sont pas occupés par $p(\mathcal{G})$. On distingue alors trois cas (cf. figure 2.2).

5. On note $x \wedge y = \min(x, y)$ et $x \vee y = \max(x, y)$, et on étend canoniquement ces définitions à des ensembles.

- Si u est en-dehors du carré $[0,1]^{k+1}$, alors, le résultat est évident.
- Si u est situé sur la frontière du carré $[0,1]^{k+1}$, alors, le résultat est encore évident puisque nous avions utilisé des inégalités strictes dans la définition des orthants autour de u.
- Si u est situé à l'intérieur du carré $[0,1]^2$, alors, dire qu'il existe un orthant autour de u qui n'est pas occupé, signifie qu'un des coins du carré $[0,1]^{k+1}$ n'est pas dans $p(\mathcal{G})$ ce qui est le cas d'après (2.6).

□

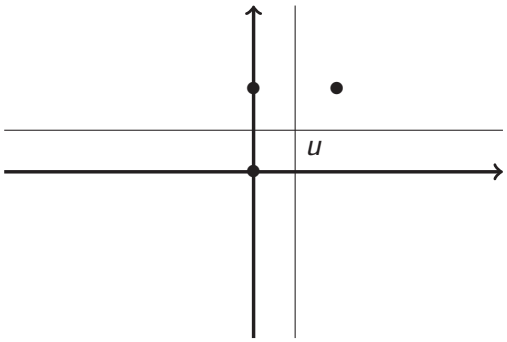


FIGURE 2.2 – Illustration du théorème 2.20 pour $k = 1$.

Remarque 2.21. Il n'y a pas équivalence dans le théorème 2.20 : $\mathcal{G}$ peut être de pseudo-dimension au plus k sans que $\mathcal{F}$ ne le soit. En effet, la classe $\mathcal{G}$ décrit uniquement le comportement de la classe $\mathcal{F}$ autour de l'origine.

Il est très simple de se représenter ce que signifie avoir une pseudo-dimension k pour une classe $\mathcal{F}$. Pourtant, comme le montre Pollard (1990), cela a des conséquences *a priori* difficiles à montrer sur le nombre de recouvrement de la classe $\mathcal{F}$. Notons que le résultat ci-dessous s'appuie sur un lemme combinatoire (lui-même basé sur les résultats de Sauer (1972)). Il est similaire à celui utilisé dans l'approche de Vapnik et Chervonenkis qui consiste en l'étude des classes éponymes et que nous détaillerons à la section 2.11.

Théorème 2.22 (Pollard)

Soit $\mathcal{F} \subset \mathbb{R}^n$ bornée admettant pour enveloppe[6] F. Si $\mathcal{F}$ est de pseudo-dimension au plus k, alors il existe une constante A dépendant uniquement de k telle que pour tout $\varepsilon \in]0,1]$ et pour tout paramètre d'échelle α, on ait :

$$D(\varepsilon \|\alpha \odot F\|_2, \alpha \odot \mathcal{F}) \leqslant A\, \varepsilon^{-4k}.$$

Par analogie avec les processus euclidiens, on parlera alors de *classes euclidiennes* (relativement à une enveloppe).

6. Une enveloppe $F = (F_1, \dots, F_n)^t$ d'un ensemble $\mathcal{F}$ est un vecteur tel que pour tout $f = (f_1, \dots, f_n)^t$ de F et pour tout $i \in [\![1, n]\!]$, $|f_i| \leqslant F_i$.

En particulier, si $(f_n)_{n\in\mathbb{N}^\star}$ est une suite de processus indexés par un ensemble T et vérifiant que :

1. les classes $\mathcal{F}_{n\omega} := \big\{(f_1(\omega,t),\dots,f_n(\omega,t)),\ t\in T\big\}$ sont de pseudo-dimension k indépendante de n et de ω ;
2. les processus f_n admettent une même enveloppe intégrable F ;

alors, la suite $(f_n)_{n\in\mathbb{N}^\star}$ est euclidienne.

Le théorème ci-dessous, dû à Bilias *et al.* (1997), appliqué de concert avec les propriétés de stabilité que nous énoncerons plus loin, nous permettra de prouver que certaines classes que nous allons rencontrer sont euclidiennes. Ce résultat est illustré à la figure 2.3.

Théorème 2.23

Soient $f_1,\dots,f_n$ des fonctions croissantes d'un ensemble T dans $\mathbb{R}$. Alors, la classe $\mathcal{F} := \big\{(f_1(t),\dots,f_n(t)),\ t\in T\big\}$ est de pseudo-dimension au plus 1.

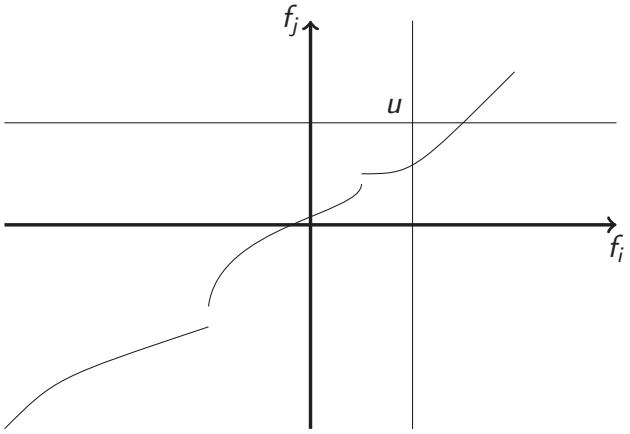


FIGURE 2.3 – Illustration du théorème 2.23.

2.6 Stabilité

Pour montrer qu'une classe est euclidienne, on sera souvent amené à la décomposer en classes plus simples pour lesquelles le résultat est évident, puis on utilisera des propriétés de stabilité. On utilise cette même approche lorsque l'on étudie les classes de Vapnik–Chervonenkis.

Dans leur article, Pakes & Pollard (1989) énoncent des propriétés de stabilité pour les classes euclidiennes. Rappelons ici un de leurs résultats que nous utiliserons dans la suite de notre étude.

Proposition 2.24. *Soient $\mathcal{F}$ et $\mathcal{G}$ deux classes euclidiennes admettant pour enveloppes respectives F et G. Alors :*

i. $\mathcal{F} + \mathcal{G}$ *est euclidienne pour l'enveloppe* $F + G$.

ii. $\mathcal{F}\mathcal{G}$ *est euclidienne pour l'enveloppe* FG.

iii. $\mathcal{F} \wedge \mathcal{G}$ *et* $\mathcal{F} \vee \mathcal{G}$ *sont euclidiennes pour l'enveloppe* $F \vee G$.

iv. Pour tout $M > 0$, *la classe* $\{\alpha f,\ f \in \mathcal{F},\ \alpha \in \mathbb{R},\ |\alpha| \leqslant M\}$ *est euclidienne pour l'enveloppe* MF.

2.7 Fonctions à variations bornées

Nous avons vu au théorème 2.23 qu'il était possible de construire des classes euclidiennes à partir de fonctions croissantes. Nous allons ici rappeler quelques résultats classiques sur les fonctions à variations bornées et donner un résultat similaire au théorème 2.23. On pourra se reporter au livre de Titchmarsh (1986) pour plus de détails sur le sujet.

Notation 2.25. Soit $f : [a, b] \to \mathbb{R}$ une application et $\sigma = (a = t_0 < t_1 < \ldots < t_n = b)$ une subdivision de l'intervalle $[a, b]$. La variation totale de f sur σ est :

$$V(f, \sigma) := \sum_{i=0}^{n-1} |f(t_{i+1}) - f(t_i)|.$$

Définition 2.26 (fonction à variations bornées)**.** On dit que $f : [a, b] \to \mathbb{R}$ est à *variations bornées* si :

$$V_a^b(f) := \sup_{\sigma} V(f, \sigma) < +\infty,$$

où le sup est pris sur l'ensemble des subdivisions de l'intervalle $[a, b]$.

Soit f une fonction à variations bornées sur $[a, b]$. Posons pour tout $x \in [a, b]$:

$$g(x) := V_a^x(f) \qquad \text{et} \qquad h(x) := g(x) - f(x).$$

Alors, on montre que g et h sont deux fonctions croissantes et que :

$$V_a^b(h) \leqslant 2V_a^b(g) = 2V_a^b(f).$$

Ainsi, toute fonction à variations bornées est différence de deux fonctions croissantes bornées et est bornée.

De plus, on montre facilement que le produit de deux fonctions à variations bornées est encore à variations bornées. Enfin, on a le résultat de stabilité par composition ci-dessous.

Proposition 2.27. *Si $f : \mathbb{R} \to \mathbb{R}$ est de classe $\mathscr{C}^1$ et que $g : [a, b] \to \mathbb{R}$ est à variations bornées, alors, $f \circ g$ est à variations bornées.*

Énonçons maintenant un résultat que nous utiliserons en particulier au chapitre 8 pour construire des classes euclidiennes et démontrer des résultats de convergence uniforme.

Théorème 2.28

Soient $f_1, \dots, f_n$ des fonctions de $[a, b]$ dans $\mathbb{R}$ et $M \in \mathbb{R}$. Supposons que pour tout $i \in [\![1, n]\!]$, $V_a^b(f_i) \leqslant M$. Posons :

$$\mathcal{F}_n := \Big\{ \big(f_1(t), \dots, f_n(t)\big),\ t \in [a, b] \Big\}.$$

Alors, $\mathcal{F}_n$ est euclidienne.

Démonstration. Comme les fonctions f_i sont à variations bornées, on peut les écrire comme différence de fonctions croissantes g_i et h_i dont les variations totales sont bornées par $2M$. Posons alors :

$$\mathcal{G}_n := \Big\{ \big(g_1(t), \dots, g_n(t)\big),\ t \in [a, b] \Big\},$$
$$\mathcal{H}_n := \Big\{ \big(h_1(t), \dots, h_n(t)\big),\ t \in [a, b] \Big\}.$$

Alors, $\mathcal{G}_n$ et $\mathcal{H}_n$ sont de pseudo-dimension au plus 1 d'après le théorème 2.23 et bornées par $2M$: elles sont donc euclidiennes. Ainsi, d'après la proposition 2.24, $\mathcal{F}_n$ est euclidienne. □

2.8 Loi des grands nombres uniforme

Théorème 2.29 (loi des grands nombres uniforme (Pollard))

Soit $(f_n)_{n\in\mathbb{N}^\star}$ une suite de processus aléatoires indépendants. Si les classes $\mathcal{F}_{n\omega} := \big\{(f_1(\omega, t), \dots, f_n(\omega, t)),\ t \in T\big\}$ sont euclidiennes et que les f_n admettent une même enveloppe F de carré intégrable, alors, lorsque $n \to +\infty$:

$$\frac{1}{n} \sup_{t\in T} |S_n(\cdot, t) - M_n(t)| \longrightarrow 0 \quad \text{p.s. et dans } L^1.$$

Remarque 2.30. Si on suppose de plus que les processus sont identiquement distribués, alors, on peut montrer la convergence dans L^1 en supposant uniquement l'existence d'un moment d'ordre 1 pour F en utilisant un argument de martingale inversée.

Remarque 2.31. Dans les conditions du théorème 2.29, si F n'est pas de carré intégrable mais simplement intégrable, alors, on peut démontrer la convergence en probabilité.

Démonstration. La démonstration de ce résultat est disponible dans Pollard (1990). Nous proposons ici une preuve élémentaire d'une version plus faible que nous utiliserons souvent par la suite. Supposons que $(f_n)_{n\in\mathbb{N}^\star}$ est euclidienne et que les enveloppes F_n sont bornées par une constante. En prenant $\alpha = 1$ dans la définition d'une suite de processus euclidiens, on remarque que les conditions présentées page 32 sont satisfaites. On obtient alors grâce à (2.5) page 33 que :

$$\mathbb{E}\sup_{t\in T}|S_n(\omega,t)-M_n(t)| \leqslant C_3\sqrt{n}.$$

L'inégalité de Markov permet alors de conclure à la convergence en probabilité. □

On peut grâce à ce résultat retrouver le théorème de Glivenko-Cantelli. Supposons en effet que $(Z_i)_{1\leqslant i\leqslant n}$ soit une suite de variables aléatoires i.i.d., et considérons les processus f_i définis par :

$$f_i(\omega,t) := \mathbb{1}_{]-\infty,t]}(Z_i(\omega)).$$

À ω fixé, les fonctions $f_i(\omega,\cdot)$ sont croissantes sur $\mathbb{R}$. Ainsi, d'après le théorème 2.23, les $\mathcal{F}_{n\omega}$ sont de pseudo-dimension au plus 1. De plus, les f_i sont bornés par 1 et donc $G \equiv 1$ est une enveloppe intégrable. On peut alors appliquer le théorème 2.29. Notons alors F_n la fonction de répartition empirique associée aux Z_i et F la fonction de répartition commune des Z_i. On a alors par définition :

$$\begin{aligned} S_n(\cdot,t) &= \sum_{i=1}^{n}\mathbb{1}_{]-\infty,t]}(Z_i) = nF_n(t),\\ M_n(t) &= n\,\mathbb{E}\,F_n(t) = nF(t). \end{aligned}$$

2.9 Théorème de la limite centrale fonctionnel

Avant d'énoncer le théorème de la limite centrale fonctionnel de Pollard (1990), il nous faut introduire quelques notations. Précisons aussi que nous nous plaçons dans le cadre de la convergence en loi au sens de Pollard.

Soit $(f_n)_{n\in\mathbb{N}^\star}$ une suite de processus indépendants paramétrés par un ensemble T. Notons :

$$\begin{aligned} X_n(\omega,t) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}f_i(\omega,t)-\mathbb{E}f_i(\cdot,t)\right),\\ \rho_n(s,t) &= \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\mathbb{E}|f_i(\cdot,s)-f_i(\cdot,t)|^2\right)^{1/2}. \end{aligned}$$

Définissons, lorsqu'elle existe, la fonction ρ comme étant la limite de la suite ρ_n au sens où :

$$\forall(s,t)\in T^2,\quad \rho(s,t)=\lim_{n\to+\infty}\rho_n(s,t).$$

Théorème 2.32 (de la limite centrale fonctionnel (Pollard))

Supposons que :

(i) les f_n sont *manageables* ;

(ii) pour tout $(s,t) \in T^2$, la limite $H(s,t) := \lim_n \mathbb{E}\, X_n(s)X_n(t)$ est bien définie ;

(iii) $\overline{\lim_n}\, \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, F_i^2 < \infty$;

(iv) pour tout $\varepsilon > 0$, $\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, \left(F_i^2\, \mathbb{1}_{\{F_i > \varepsilon\sqrt{n}\}}\right) \to 0$ lorsque $n \to +\infty$;

(v) la limite $\rho(\cdot,\cdot)$ est bien définie et pour toutes suites $(s_n)_n$ et $(t_n)_n$ telles que $\rho(s_n,t_n) \underset{n\to\infty}{\longrightarrow} 0$, on a $\rho_n(s_n,t_n) \underset{n\to\infty}{\longrightarrow} 0$.

Alors, $(X_n)_n$ converge en loi vers un processus gaussien centré dont la fonction de variance-covariance est donnée par H.

Remarque 2.33. Lorsque nous sommes en présence de processus identiquement distribués, on a $\rho_n(s,t) = \rho(s,t)$ et la fonction ρ est alors toujours bien définie. Le point (v) du théorème de la limite centrale fonctionnel ci-dessus est alors automatiquement vérifié. De plus, si on suppose que les f_n sont uniformément bornés (en n et en t) par une même enveloppe F de carré intégrable, alors, les conditions (iii) et (iv) sont aussi vérifiées.

Heuristique de la démonstration dans un cas simple. On se reportera à Pollard (1990) pour une démonstration détaillée de ce résultat. Plaçons-nous ici encore une fois dans un cadre plus simple. Supposons les processus f_i identiquement distribués et uniformément bornés par une constante. Alors, on a vu que (cf. équation (2.4) page 33) :

$$\mathbb{E}\left(\frac{1}{\sqrt{n}} \sup_{t\in T} |S_n(\omega,t) - M_n(t)|\right) \leqslant C_2\, \mathbb{E}\left(\Lambda_n\left(\frac{\delta_n}{\|F_n\|_2}\right)\right). \tag{2.7}$$

De plus, comme les f_i sont identiquement distribués, la fonction ρ s'écrit :

$$\rho(s,t) = \left(\mathbb{E}\,|f(\cdot,s) - f(\cdot,t)|^2\right)^{1/2}.$$

Pour montrer la convergence en distribution, on utilise alors la caractérisation vue page 23 au théorème 1.26. On démontre la convergence des marges fini-dimensionnelles grâce au théorème central limite classique. Quant à l'équicontinuité asymptotique, elle fait intervenir la semi-métrique définie par ρ via les classes :

$$\mathcal{F}_{r\omega n} := \left\{ \big(f_1(\omega,s) - f_1(\omega,t), \ldots, f_n(\omega,s) - f_n(\omega,t)\big),\ \rho(s,t) < r \right\}.$$

En utilisant les propriétés de stabilité des classes *manageables*, on montre une inégalité similaire à (2.7) valable pour les éléments de $\mathcal{F}_{r\omega n}$. Le sup devient alors un sup sur l'ensemble des couples (s,t) tels que $\rho(s,t) < r$ et il s'agit de contrôler le diamètre $\delta_{r\omega n}$ des classes $\mathcal{F}_{r\omega n}$. Or, par définition :

$$\delta_{r\omega n} = \sup_{\rho(s,t)<r} \left(\frac{1}{n} \sum_{i=1}^{n} \big(f_i(\omega,s) - f_i(\omega,t) \big)^2 \right)^{1/2}.$$

Mais, un argument de type Glivenko-Cantelli permet de dire que cette quantité se comporte asymptotiquement comme $(\mathbb{E}\,|f_i(\cdot,s) - f_i(\cdot,t)|^2)^{1/2} = \rho(s,t)$. Ainsi, si s et t sont suffisamment proches pour la semi-métrique ρ, alors, le diamètre des $\mathcal{F}_{n\omega r}$ peut être rendu arbitrairement petit et l'inégalité de Markov assure alors l'équicontinuité asymptotique recherchée. □

Tout comme pour la loi des grands nombres uniforme, on peut retrouver ici le théorème de Donsker. Soit en effet $(Z_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires i.i.d. et posons encore :

$$f_i(\omega,t) := \mathbb{1}_{]-\infty,t]}(Z_i(\omega)).$$

Alors, les $f_i(\omega,\cdot)$ sont des fonctions croissantes et bornées par 1. Toutes les conditions du théorème sont donc réunies, sauf peut-être la deuxième. Notons F_n la fonction de répartition empirique des Z_i et F la fonction de répartition commune des Z_i. Alors, un rapide calcul montre que :

$$\begin{aligned}
\mathbb{E}\,X_n(s)X_n(t) &= \frac{1}{n}\,\mathbb{E}\left(\left(\sum_{i=1}^{n} \Big(\mathbb{1}_{]-\infty,s]}(Z_i) - F(s) \Big) \right) \left(\sum_{i=1}^{n} \Big(\mathbb{1}_{]-\infty,t]}(Z_i) - F(t) \Big) \right) \right) \\
&= \frac{1}{n} \Big(nF(s\wedge t) - nF(s)F(t) \Big) \\
&= F(s\wedge t) - F(s)F(t),
\end{aligned}$$

puisque $\mathbb{1}_{]-\infty,t]}(Z_i)\mathbb{1}_{]-\infty,s]}(Z_i) = \mathbb{1}_{]-\infty,s\wedge t]}(Z_i)$.

2.10 Théorème de la limite centrale fonctionnel multivarié

Dans le chapitre 8, nous allons utiliser un résultat de type théorème de la limite centrale pour des processus à valeurs vectorielles. Voyons ici un corollaire immédiat du théorème présenté à la section précédente. Supposons donc être en présence de K suites de processus indépendants $(f_i^k(\omega,t))$, et notons $f_i = (f_i^1, \dots, f_i^K)^t$. Notons encore pour tout $k \in [\![1,K]\!]$:

$$X_n^k(\omega,t) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} f_i^k(\omega,t) - \mathbb{E}\,f_i^k(\cdot,t) \right).$$

Si on peut appliquer le théorème de la limite centrale aux X_n^k (à k fixé), alors, chacune des suites $(X_n^k)_{n\in\mathbb{N}^\star}$ est asymptotiquement ρ_k-équicontinue, où les

ρ_k sont les semi-métriques définies à la section précédente. Rappelons que cela signifie que pour tout $\varepsilon > 0$ et pour tout $\eta > 0$, il existe $\delta > 0$ tel que :

$$\overline{\lim_n}\, \mathbb{P}^\star \left(\sup_{\rho_k(s,t)<\delta} |X_n^k(s) - X_n^k(t)| > \eta \right) < \varepsilon.$$

Soit alors ρ la semi-métrique définie par :

$$\rho(s,t) := \max_{1\leqslant k\leqslant K} \rho_k(s,t).$$

On en déduit alors que la suite de processus $(X_n = (X_n^1, \dots, X_n^K)^t)_{n\in\mathbb{N}^\star}$ est asymptotiquement ρ-équicontinue, au sens où pour tout $\varepsilon > 0$ et pour tout $\eta > 0$, il existe $\delta > 0$ tel que :

$$\overline{\lim_n}\, \mathbb{P}^\star \left(\sup_{\rho(s,t)<\delta} \max_{1\leqslant k\leqslant K} |X_n^k(s) - X_n^k(t)| > \eta \right) < \varepsilon.$$

De plus, comme K est fini, l'espace (T,ρ) est encore pré-compact. Ainsi, pour démontrer la convergence en loi de la suite de processus $(X_n)_{n\in\mathbb{N}^\star}$ à valeurs dans $\mathbb{R}^K$, il suffit d'appliquer le théorème de la limite centrale à ses coordonnées (pour démontrer l'équicontinuité asymptotique) puis de démontrer la convergence de ses marges fini-dimensionnelles, en utilisant par exemple le théorème de la limite centrale multivarié classique.

2.11 Les classes de Vapnik-Chervonenkis

2.11.1 Introduction

Comme on l'a vu précédemment, l'approche de Pollard s'adapte particulièrement bien à l'étude des processus aléatoires indexés par un paramètre de temps. En théorie des processus empiriques, il existe une autre approche : celle de Vapnik et Chervonenkis. Considérons des variables aléatoires $X_1, \dots, X_n$ à valeurs dans un espace métrique $(\mathcal{X}, d)$ que l'on munit de sa tribu borélienne $\mathscr{B}$. La loi de probabilité empirique associée est la mesure aléatoire :

$$\mathbb{P}_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}.$$

Soit maintenant un ensemble $C \in \mathscr{B}$, on peut définir :

$$\mathbb{P}_n(C) := \int \mathbb{1}_C \mathrm{d}\mathbb{P}_n.$$

De même, si f est une application mesurable de $\mathcal{X}$ dans $\mathbb{R}$, on peut définir :

$$\mathbb{P}_n(f) := \int f \mathrm{d}\mathbb{P}_n.$$

Soit alors $\mathcal{F}$ une collection de fonctions mesurables de $\mathcal{X}$ dans $\mathbb{R}$. La loi de probabilité empirique peut être vue comme une fonction aléatoire à valeurs dans l'ensemble des fonctions de $\mathcal{F}$ dans $\mathbb{R}$ via l'identification :

$$\begin{array}{rccc} \mathbb{P}_n & : & \mathcal{F} & \to & \mathbb{R} \\ & & f & \mapsto & \mathbb{P}_n(f) = \int f \mathrm{d}P_n. \end{array}$$

Définissons maintenant $\ell_\infty(\mathcal{F})$ comme l'ensemble des applications ψ de $\mathcal{F}$ dans $\mathbb{R}$ qui sont bornées pour la norme définie par :

$$\|\psi\|_\infty := \sup_{f \in \mathcal{F}} |\psi(f)|.$$

L'approche de Vapnik-Chervonenkis consiste à chercher des conditions suffisantes sur $\mathcal{F}$ pour obtenir des résultats de type loi des grands nombres uniforme et théorème de la limite centrale, en travaillant dans $\ell_\infty(\mathcal{F})$. Nous allons présenter très brièvement quelques résultats sur ce sujet, et nous mettrons en évidence quelles sont les difficultés rencontrées si on ne fait pas l'identification entre les mesures sur $\mathcal{X}$ et $\ell_\infty(\mathcal{F})$. On se reportera au livre de van der Vaart & Wellner (1996) pour plus de détails.

2.11.2 Définitions et résultats classiques

Tout comme dans l'approche de Pollard, ce sont des conditions simples sur la géométrie des classes $\mathcal{F}$ qui vont permettre de démontrer les résultats de convergence.

Soit $\mathcal{C}$ une collection de sous-ensembles d'un ensemble $\mathcal{X}$ et soit $x_1, \dots, x_n$ n points de $\mathcal{X}$. On dira que $\mathcal{C}$ *attrape* un sous-ensemble de $\{x_1, \dots, x_n\}$ si ce sous-ensemble peut s'écrire $C \cap \{x_1, \dots, x_n\}$ pour un certain $C \in \mathcal{C}$. On dira que $\mathcal{C}$ *pulvérise* $\{x_1, \dots, x_n\}$ s'il attrape toute partie de $\{x_1, \dots, x_n\}$. On a alors la définition ci-dessous.

Définition 2.34. Soit $\mathcal{C}$ une collection d'ensembles et soit $A \subset \mathcal{X}$ un ensemble fini. On dit que $\mathcal{C}$ *pulvérise* A si :

$$\mathcal{C} \cap A := \{C \cap A,\ C \in \mathcal{C}\} = \mathcal{P}(A).$$

Définition 2.35. On dit que $\mathcal{C}$ est de dimension de Vapnik au plus k si tout ensemble A de cardinal k n'est pas pulvérisé par $\mathcal{C}$. La *dimension de Vapnik* de $\mathcal{C}$ est l'infimum de ces k (on pose par convention $\inf \varnothing = +\infty$). On parlera de *classes de Vapnik-Chervonenkis* lorsque cet infimum est fini.

Exemple 2.36. Soit $\mathcal{C} := \{]-\infty, a],\ a \in \mathbb{R}\}$, alors $\mathcal{C}$ est de dimension 2. Il est clair que tout singleton $\{x\}$ est pulvérisé par $\mathcal{C}$: il suffit de prendre $C_1 :=]-\infty, x]$ et $C_2 =]-\infty, x-1]$, et ainsi $\mathcal{C} \cap \{x\} = \{\{x\}; \varnothing\}$. Considérons maintenant 2 points x et y et supposons $x < y$ pour fixer les idées. Alors, $\{y\} \notin \mathcal{C} \cap \{x; y\}$. En effet, si $a \in \mathbb{R}$ est tel que $C :=]-\infty, a]$ attrape y, cela signifie que $a \geqslant y$ et donc que $a \geqslant x$. Ainsi, C attrape aussi x. Ce résultat élémentaire est illustré à la figure 2.4.

FIGURE 2.4 – Illustration de l'exemple 2.36 : il est impossible pour $\mathcal{C}$ d'attrapper y sans attrapper x.

Exemple 2.37. Soit $\mathcal{C} := \{\,[a,b],\ (a,b) \in \mathbb{R}^2\}$, alors $\mathcal{C}$ est de dimension 3. Il est clair que tout singleton $\{x\}$ est pulvérisé par $\mathcal{C}$. De même, toute partie de $\mathbb{R}$ à deux éléments est pulvérisée par $\mathcal{C}$. Considérons maintenant 3 points x, y et z et supposons $x < y < z$ pour fixer les idées. Alors, $\{x;z\} \notin \mathcal{C} \cap \{x;y;z\}$. En effet, si a et b sont deux points de $\mathbb{R}$ tels que $C := [a,b]$ attrape x et z, alors $a \leqslant x$ et $b \geqslant z$ et donc $y \in [a,b]$. Ainsi, C attrape aussi y. Nous avons illustré ce résultat à la figure 2.5.

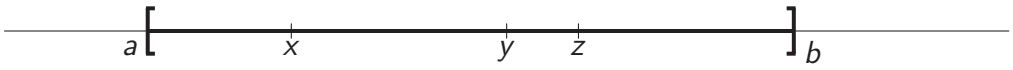


FIGURE 2.5 – Illustration de l'exemple 2.37 : il est impossible pour $\mathcal{C}$ d'attraper x et z sans attraper y.

Définition 2.38. Soit $\mathcal{F}$ une classe de fonctions de $\mathcal{X}$ dans $\mathbb{R}$. On dira que $\mathcal{F}$ est une VC-classe de dimension (au plus) k si l'ensemble des hypographes[7] des fonctions de $\mathcal{F}$ forme une VC-classe (de $\mathcal{X} \times \mathbb{R}$) de dimension (au plus) k.

Remarquons que cette définition est cohérente avec la définition 2.35 puisqu'on peut montrer que si $\mathcal{C}$ est une classe de Vapnik au sens de la définition 2.35, alors les fonctions $\mathbb{1}_C$ forment une classe de Vapnik au sens de la définition ci-dessus.

Dans leur livre, van der Vaart & Wellner (1996) montrent qu'une VC-classe de dimension finie admettant une enveloppe F vérifie une condition d'entropie (tout comme les classes euclidiennes de Pollard). Soit Q une mesure de probabilité, définissons :

$$\|f\|_{p,Q} = \left(\int |f|^p \mathrm{d}Q\right)^{1/p}.$$

Supposons que $\mathcal{F}$ est une VC-classe de dimension k admettant une enveloppe F. Alors, pour toute mesure de probabilité Q telle que $\|F\|_{2,Q} > 0$ on a :

$$N(\varepsilon\|F\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}) \leqslant A\,\varepsilon^{-W}, \tag{2.8}$$

où A et W sont des constantes dépendant uniquement de la dimension k. En particulier, la condition d'uniforme entropie suivante est satisfaite :

$$\int_0^{+\infty} \sup_Q \sqrt{\ln N(\varepsilon\|F\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q})} < +\infty, \tag{2.9}$$

7. L'hypographe de f est l'ensemble $\big\{(x,t) \in \mathcal{X} \times \mathbb{R} \ :\ t < f(x)\big\}$.

où le supremum est pris sur l'ensemble des mesures de probabilité discrètes Q telles que $\|F\|_{2,Q} > 0$. Tout comme dans l'approche de Pollard, c'est cette condition d'entropie qui va permettre de démontrer des résultats de convergence. Remarquons de plus que van der Vaart & Wellner (1996) énoncent des résultats de stabilité par réunion finie, intersection, passage au complémentaire ou encore image réciproque pour les classes de Vapnik-Chervonenkis.

Dans le cadre particulier des processus séparables, on peut montrer que l'espace probabilisé de départ n'intervient pas dans le théorème 2.32 de la limite centrale de Pollard. Ce n'est pas le cas lorsqu'on s'intéresse à des processus empiriques plus complexes. C'est pourquoi van der Vaart & Wellner (1996) se fixent un espace de départ et démontrent leurs résultats dans ce cadre. Dans ce qui suit, les variables aléatoires $X_1, X_2, \ldots$ seront donc définies sur l'espace produit $(\Omega^\infty, \mathcal{A}^\infty)$ [8]. Ainsi, $\mathbb{P}_n$ sera considéré comme un élément aléatoire [9] à valeurs dans $(\ell_\infty(\mathcal{F}), \|\cdot\|_\infty)$ muni de la tribu borélienne.

Notation 2.39. Soit P une mesure signée sur $(\mathcal{X}, d)$. On notera $\mathcal{B}(\mathcal{F}, P)$ l'ensemble des fonctions de $\ell_\infty(\mathcal{F})$ qui sont uniformément continues pour la norme $\|\cdot\|_{2,P}$.

Théorème 2.40

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires i.i.d. de loi commune P_0 et soit $\mathcal{F}$ une VC-classe de dimension finie. Supposons que $\mathcal{F}$ admette une enveloppe F vérifiant $\|F\|_{2,P_0} < +\infty$. Alors, il existe un processus gaussien centré séparable G tel que :

$$\sqrt{n}(\mathbb{P}_n - P_0) \rightsquigarrow G.$$

De plus, $\mathbb{P}(G \in \mathcal{B}(\mathcal{F}, P_0)) = 1$.

En appliquant le théorème de la limite centrale classique, on remarque que le processus limite G vérifie :

$$\forall f \in \mathcal{F} \ : \quad G(f) \sim \mathcal{N}\left(0, \int f^2 \mathrm{d}P - \left(\int f \mathrm{d}P\right)^2\right),$$

$$\forall (f,g) \in \mathcal{F}^2 \ : \quad \mathbb{C}\mathrm{ov}(f,g) = \int fg \mathrm{d}P - \int f \mathrm{d}P \int g \mathrm{d}P.$$

De plus, van der Vaart & Wellner (1996) montrent que ces deux équations déterminent entièrement le processus G. Enfin, on peut montrer (cf. van der Vaart & Wellner (1996)) que si on remplace la condition $\|F\|_{2,P_0} < +\infty$ par $\|F\|_{1,P_0} < +\infty$ dans le théorème précédent, alors, presque sûrement :

$$\|\mathbb{P}_n - P_0\|_{\mathcal{F}} \longrightarrow 0 \quad \text{lorsque } n \to +\infty.$$

8. Cet espace est le produit tensoriel infini des $(\Omega, \sigma(X_i))$.

9. On appelle élément aléatoire une fonction sur laquelle on ne fait pas d'hypothèse de mesurabilité.

2.11.3 VC-classes et classes euclidiennes

Il existe un lien fort entre les classes de Vapnik et les classes euclidiennes. Étant donné la similarité entre les théories de Pollard (1990) et de van der Vaart & Wellner (1996) (qui démontrent les deux des conditions d'entropie), le résultat ci-dessous n'a rien de surprenant.

Théorème 2.41 (Pakes & Pollard (1989))

Soit $\mathcal{C}$ une VC-classe de dimension k et $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires définies sur un même espace. Posons :

$$\mathcal{F}_{n\omega} := \left\{ \left(\mathbb{1}_C(X_1(\omega)), \dots, \mathbb{1}_C(X_n(\omega)) \right),\ C \in \mathcal{C} \right\}.$$

Alors, $\mathcal{F}_{n\omega}$ est de pseudo-dimension au plus $k-1$.

Démonstration. Soient $i_1, \dots, i_k$ des indices distincts compris entre 1 et n. Pour montrer le résultat, on peut reprendre le raisonnement de la démonstration du théorème 2.20. Il suffit alors de montrer que :

$$\left\{ \left(\mathbb{1}_C(X_{i_1}), \dots, \mathbb{1}_C(X_{i_k}) \right),\ C \in \mathcal{C} \right\} \neq \left\{0;1\right\}^k.$$

Autrement dit, il suffit de montrer que $\mathcal{C}$ n'attrape pas une des parties de l'ensemble $\{X_{i_1}, \dots, X_{i_k}\}$. Or, $\mathcal{C}$ étant de dimension de Vapnik k, elle ne peut pas pulvériser l'ensemble $\{X_{i_1}, \dots, X_{i_k}\}$. □

2.11.4 Convergence de mesures signées aléatoires

Résultat

On a vu que l'approche de Vapnik-Chervonenkis consistait à considérer les mesures sur $(\mathcal{X}, d)$ comme des éléments de $\ell_\infty(\mathcal{F})$. Cependant, on pourrait décider de travailler directement sur l'ensemble $\mathcal{P}$ des mesures signées sur $\mathcal{F}$, que l'on munirait de la semi-norme :

$$\|Q\|_{\mathcal{F}} := \sup_{f\in\mathcal{F}} \left| \int f \mathrm{d}Q \right|.$$

En ce qui concerne la loi des grands nombres uniforme, cela ne changerait rien puisque, lorsque $n \to +\infty$:

$$\|\mathbb{P}_n - P_0\|_{\mathcal{F}} \to 0 \iff \sup_{f\in\mathcal{F}} |\mathbb{P}_n(f) - P_0(f)| \to 0 \iff \|P_n - P_0\|_\infty \to 0.$$

Cependant, des problèmes se posent lorsqu'on s'intéresse à la convergence en loi, comme nous allons le montrer ici. Plaçons-nous dans le cadre du théorème de la limite centrale 2.40 et notons $\mathbb{G}_n$ le processus empirique renormalisé considéré comme un élément aléatoire à valeurs dans $\ell_\infty(\mathcal{F})$:

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_0) \colon (\Omega, \mathcal{A}, \mathbb{P}) \to \ell_\infty(\mathcal{F}).$$

Munissons tous les espaces en jeu de leur tribu borélienne associée et notons encore Q_n l'élément aléatoire $\sqrt{n}(\mathbb{P}_n - P_0)$ à valeur dans l'ensemble des mesures signées sur $(\mathcal{X}, d)$. Nous allons montrer que la convergence en loi de $(\mathbb{G}_n)_{n\in\mathbb{N}^\star}$ implique la convergence en loi de $(Q_n)_{n\in\mathbb{N}^\star}$. Nous allons donc montrer qu'il existe une mesure borélienne [10] μ telle que :

$$\forall \phi \in \mathscr{C}_b((\mathcal{P}, \|\cdot\|_{\mathcal{F}}), \mathbb{R}) \ : \quad \mathbb{E}^\star(\phi(Q_n)) \to \int \phi \, \mathrm{d}\mu. \tag{2.10}$$

À cet effet, définissons le quotient :

$$\widetilde{\mathcal{P}} := \mathcal{P}/\{Q \ : \ \|Q\|_{\mathcal{F}} = 0\},$$

ainsi que l'application

$$\begin{array}{rccc} \theta & : & \widetilde{\mathcal{P}} & \to & (\mathcal{F} \to \mathbb{R}) \\ & & \bar{P} & \mapsto & (f \mapsto \int f \mathrm{d}P), \end{array}$$

où P est un représentant quelconque de la classe $\bar{P}$. Cette application est bien définie puisque :

$$\bar{P} = \bar{Q} \iff \|P - Q\|_{\mathcal{F}} = 0 \iff \forall f \in \mathcal{F} \ : \ \int f \mathrm{d}P = \int f \mathrm{d}Q.$$

De plus, par construction, $\theta(\bar{Q}_n) = \mathbb{G}_n$. Enfin :

$$\|\theta(P)\| = \sup_{f\in\mathcal{F}} \left| \int f \mathrm{d}P \right| = \|P\|_{\mathcal{F}}.$$

On en déduit que θ est une isométrie linéaire surjective de $\widetilde{\mathcal{P}}$ dans son image. Notons θ^{-1} sa réciproque qui est une application continue. Appliquons maintenant l'axiome du choix à la famille $(\bar{P})_{\bar{P}\in\widetilde{\mathcal{P}}}$ pour construire une fonction permettant de remonter de $\widetilde{\mathcal{P}}$ vers $\mathcal{P}$. Il existe une fonction « choix » Γ :

$$\begin{array}{rccc} \Gamma & : & (\widetilde{\mathcal{P}}, \|\cdot\|_{\mathcal{F}}) & \to & (\mathcal{P}, \|\cdot\|_{\mathcal{F}}) \\ & & \bar{P} & \mapsto & \Gamma(\bar{P}), \end{array}$$

telle que pour tout $\bar{P}$, $\Gamma(\bar{P}) \in \bar{P}$. Remarquons alors que la fonction Γ est continue. En effet :

$$\begin{aligned} \|\bar{P} - \bar{Q}\| < \varepsilon \iff \forall (P, Q) \in \bar{P} \times \bar{Q} \ : \ \sup_{f\in\mathcal{F}} \left| \int f \mathrm{d}P - \int f \mathrm{d}Q \right| < \varepsilon \\ \implies |\Gamma(\bar{P}) - \Gamma(\bar{Q})| < \varepsilon. \end{aligned}$$

On déduit de ce qui précède que Q_n est une transformation continue de $\mathbb{G}_n$:

$$Q_n = (\Gamma \circ \theta^{-1})(\mathbb{G}_n).$$

10. Cette mesure borélienne étant définie sur la fermeture de $\mathcal{P}$.

Pour conclure, il faut utiliser un résultat un peu plus fin que le transport de la convergence en loi par les applications continues. En effet, $\Gamma \circ \theta^{-1}$ n'est pas défini *a priori* en G, le processus limite de $\mathbb{G}_n$. Cependant, le théorème de Prohorov (cf van der Vaart & Wellner (1996) page 22) permet de démontrer qu'il existe une sous-suite de $(Q_n)_{n\in\mathbb{N}^\star}$ qui converge en distribution et on peut alors simplement montrer que toute la suite $(Q_n)_{n\in\mathbb{N}^\star}$ converge en distribution vers une mesure borélienne μ, au sens où l'assertion (2.10) est satisfaite.

Étude de la limite

Lorsqu'on utilise l'approche de van der Vaart & Wellner (1996), la loi limite de $(\mathbb{G}_n)_{n\in\mathbb{N}^\star}$ est presque sûrement un élément de $\mathcal{B}(\mathcal{F}, P_0)$. Cela montre que l'espace $\ell_\infty(\mathcal{F})$ est suffisamment riche pour étudier ladite suite. Cependant, si on ne considère pas les G_n comme des éléments de $\ell_\infty(\mathcal{F})$ mais comme des mesures aléatoires sur $(\mathcal{X}, d)$, alors, la limite n'est pas nécessairement une mesure signée, comme nous allons le voir ici. En fait, c'est simplement une application additive définie sur l'ensemble des boréliens de $(\mathcal{X}, d)$. De ce fait, elle est beaucoup plus difficile à appréhender qu'un élément de $\mathcal{B}(\mathcal{F}, P_0)$, ce qui justifie l'identification faite par van der Vaart & Wellner (1996).

Soit $(X_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires réelles indépendantes de loi $P_0 \sim \mathcal{U}([0,1])$. Définissons $\mathcal{F}$ comme étant l'ensemble des fonctions f_t, $t \in [0,1]$, où :

$$\begin{array}{rccc} f_t & : & [0,1] & \to & \mathbb{R} \\ & & x & \mapsto & \mathbb{1}_{[0,t]}(x). \end{array}$$

Alors, d'après le théorème 2.40, on sait qu'il y a convergence en distribution de $\sqrt{n}(\mathbb{P}_n - P_0)$ vers un processus $G(\cdot) \in \mathcal{B}(\mathcal{F}, P_0)$ vérifiant que pour tout $(s,t) \in [0,1]^2$:

$$\begin{aligned} G(f_t) &\sim \mathcal{N}(0, t(1-t)), \\ \mathbb{C}\mathrm{ov}(G(f_s), G(f_t)) &= s \wedge t - st. \end{aligned}$$

Ainsi, G peut être identifié à un pont brownien B sur $[0,1]$ en posant :

$$\forall t \in [0,1] \ : \quad B(s) := G(f_s).$$

Supposons alors que le processus limite $G(\cdot)$ puisse être représenté par une mesure signée Q. On aurait alors :

$$\forall t \in [0,1] \ : \quad Q([0,t]) = \int f_t \, \mathrm{d}Q = Z(f_t) = B(t).$$

Or, les variations positives et négatives du pont brownien ne sont pas bornées, ce qui montre que Q ne peut pas définir une mesure signée.

Remarquons que nous pouvons démontrer le résultat sans avoir à utiliser les propriétés du pont brownien. En effet, plaçons-nous dans le même cadre que précédemment et considérons les fonctions de Haar $(f_n)_{n\in\mathbb{N}^\star}$ (cf. figure 2.6) définies par :

$$\forall t \in [0,1[: \qquad f_n(t) := \sum_{k=0}^{2^n-1} (-1)^k \mathbb{1}_{[k2^{-n},(k+1)2^{-n}[}(t).$$

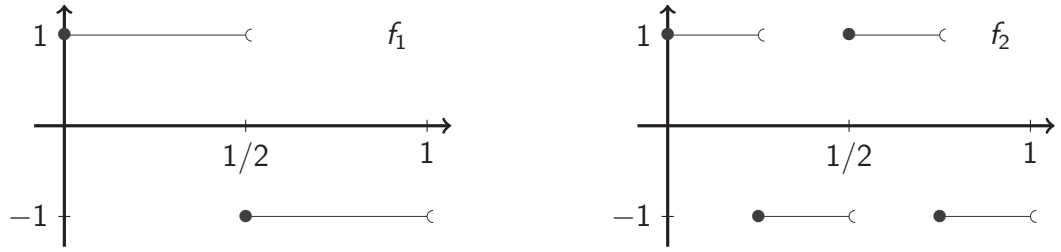


FIGURE 2.6 – Les fonctions de Haar.

On montre facilement (en utilisant le théorème 2.40) que $(G(f_n))_{n\in\mathbb{N}^\star}$ est une suite de variables aléatoires mutuellement indépendantes et de loi $\mathcal{N}(0,1)$. Supposons alors que G soit représenté par une mesure Q. Notons Q^- et Q^+ ses parties positives et négatives données par la décomposition de Jordan. On a, pour tout $n \in \mathbb{N}^\star$:

$$\begin{aligned} |Q(f_n)| &= \left| \sum_{k=0}^{2^n-1} (-1)^k Q([k2^{-n},(k+1)2^{-n}[) \right| \\ &\leqslant \sum_{k=0}^{2^n-1} (Q^+ + Q^-)([k2^{-n},(k+1)2^{-n}[) \\ &= Q^+([0,1]) + Q^-([0,1]). \end{aligned}$$

Mais, presque sûrement :

$$\sup_{n\in\mathbb{N}^\star} |Q(f_n)| = +\infty.$$

De plus :

$$Q^+([0,1]) - Q^-([0,1]) = Q([0,1]) \sim \mathcal{N}\big(0, P_0([0,1])(1-P_0([0,1]))\big) = \mathcal{N}(0,0).$$

On en déduit que, presque sûrement, $Q^+([0,1]) = Q^-([0,1]) = +\infty$ ce qui est incompatible avec le fait que Q est une mesure signée.

CHAPITRE 3

Vraisemblance Empirique

3.1 Le théorème d'Owen

Étant données n variables aléatoires réelles i.i.d. $X_1, \dots, X_n$ de fonction de répartition F_0, la vraisemblance non paramétrique d'une fonction de répartition F est donnée par :

$$L(F) = \prod_{i=1}^{n} \big(F(X_i) - F(X_i^-)\big).$$

On peut montrer en utilisant une inégalité de convexité que la fonction de répartition maximisant L est la fonction de répartition empirique F_n associée aux X_i. De manière équivalente, F_n est l'unique fonction de répartition maximisant le rapport de vraisemblance :

$$L' \colon F \mapsto \frac{L(F)}{L(F_n)}.$$

De plus, on sait grâce aux théorèmes de Donsker et de Glivenko-Cantelli que dans le cas de variables aléatoires i.i.d., F_n a de « bonnes propriétés ». L'idée de la vraisemblance empirique est d'étudier le maximum de cette fonction L' en se restreignant à certaines fonctions de répartition. Avant d'entrer dans les détails, remarquons que dans l'optique de maximiser la fonction L', on peut se contenter de travailler avec des lois de probabilité discrètes à support dans $\{X_1, \dots, X_n\}$. Ainsi, si F est une fonction de répartition et que l'on note :

$$\forall i \in [\![1, n]\!] \ : \quad p_i := F(X_i) - F(X_i^-),$$

alors, le problème est devenu celui de maximiser la fonction

$$L' \colon p \in \mathbb{S}_{n-1} \mapsto \prod_{i=1}^{n} np_i \in [0, 1],$$

où on a noté $\mathbb{S}_{n-1}$ le simplexe unité de $\mathbb{R}^n$. Supposons maintenant que les X_i sont de carré intégrable et notons μ_0 leur espérance commune. Que se passe-t-il si on cherche le maximum du rapport de vraisemblance sur l'ensemble des fonctions de répartition admettant un moment d'ordre 1 égale à μ_0 ? Autrement dit, que peut-on dire de la quantité :

$$\mathrm{EL}_n(\mu_0) := \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i X_i = \mu_0 \ \text{ et } \ p \in \mathbb{S}_{n-1}\right\},$$

où on a posé par convention $\max \varnothing = 0$? C'est la question que se sont posée Thomas & Grunkemeier (1975) dans dans un problème de survie avec censure à droite. Ils ont alors remarqué que $\mathrm{EL}_n(\mu_0)$ suivait asymptotiquement une loi du χ^2 dès que les X_i avait une variance non nulle. Plus tard, Owen (1990) énoncera le résultat ci-dessous.

Théorème 3.1 (Owen)

Soit $(X_i)_{i\in\mathbb{N}^\star}$ une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^p$ admettant un moment d'ordre 2. Notons μ_0 l'espérance commune des X_i et k le rang de la matrice de variance-covariance des X_i. Alors :

$$-2\ln \mathrm{EL}_n(\mu_0) \xrightarrow{\mathcal{L}} \chi^2_k, \quad \text{lorsque } n \to +\infty.$$

La convention $\max \varnothing = 0$ n'a pas de conséquence puisque comme le montre Owen (1990), la probabilité pour que μ_0 soit dans l'enveloppe convexe des $(X_i)_{1\leqslant i\leqslant n}$ tend vers 1 quand n tend vers l'infini.

3.2 Construction de régions de confiance

Le théorème d'Owen énoncé à la section précédente permet d'obtenir des régions de confiance asymptotiques pour la moyenne commune μ_0 d'un échantillon $X_1, \dots, X_n$ de variables aléatoires de L^2 à valeurs dans $\mathbb{R}^p$. En effet, si $\alpha \in [0,1[$, que k désigne le rang de la matrice de variance-covariance commune des X_i et et que v_α est le quantile d'ordre $1-\alpha$ de la loi du χ^2_q, alors :

$$\mathbb{P}(-2\ln \mathrm{EL}_n(\mu_0) \leqslant v_\alpha) \longrightarrow 1-\alpha, \quad \text{lorsque } n \to +\infty.$$

Ainsi, on en déduit une région de confiance asymptotique :

$$R_{n,1-\alpha} := \left\{\mu \in \mathbb{R}^p \ :\ -2\ln \mathrm{EL}_n(\mu) \leqslant v_\alpha\right\}. \tag{3.1}$$

Remarquons que l'on peut réécrire $R_{n,1-\alpha}$ comme :

$$R_{n,1-\alpha} = \left\{\sum_{i=1}^n p_i X_i,\ -2\sum_{i=1}^n \ln(np_i) \leqslant v_\alpha\right\}.$$

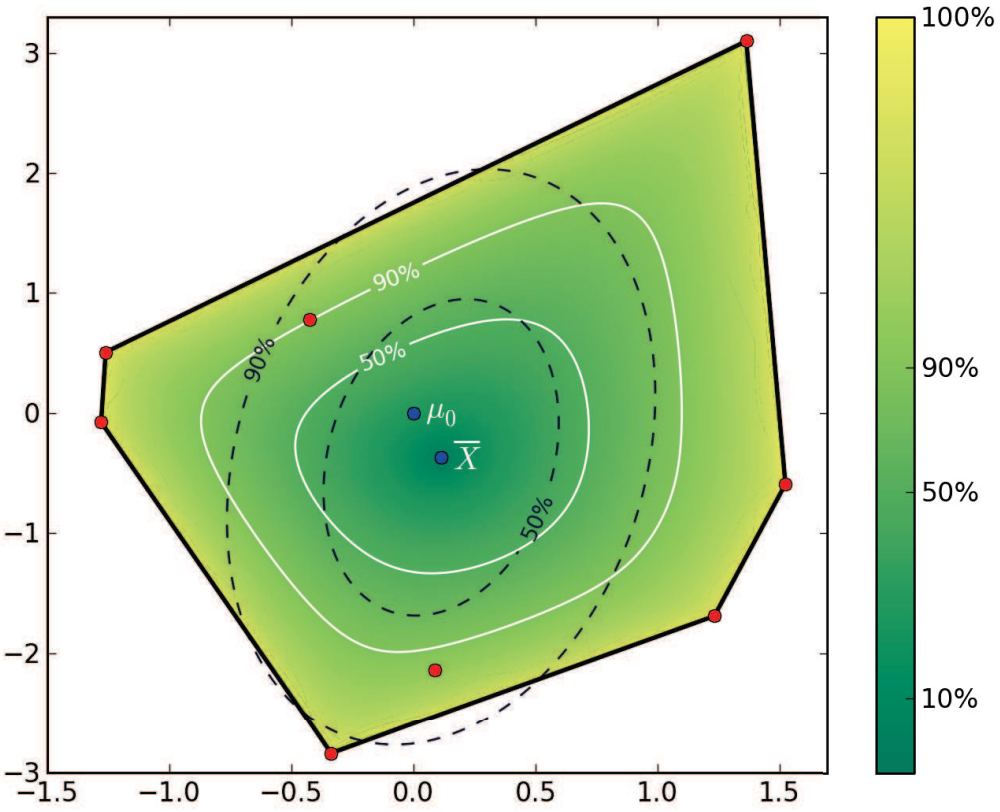


FIGURE 3.1 – Régions de confiance par vraisemblance empirique et par théorème de la limite centrale.

Cependant, ce deuxième point de vue a un inconvénient majeur en pratique puisqu'il est très long numériquement de construire l'ensemble des vecteurs de probabilité dont le produit des composantes dépasse un certain seuil (lorsque n est grand). Au contraire, en utilisant la définition (3.1), on voit qu'il suffit de faire varier μ dans l'enveloppe convexe des $(X_i)_{1\leqslant i\leqslant n}$ et de ne garder que les μ vérifiant $-2\ln \mathrm{EL}_n(\mu) \leqslant v_\alpha$ pour déterminer $R_{n,1-\alpha}$.

3.3 Illustration

Pour illustrer le théorème d'Owen, on a généré 8 réalisations d'une loi $\mathcal{N}(O,\Sigma)$ où $\Sigma = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$, et on a tracé sur un même graphique les régions de confiance obtenues par vraisemblance empirique et par le théorème de la limite centrale. Sur la figure 3.1, on a représenté en rouge les 8 réalisations et en bleu $\bar{X}$ et $\mu_0 = (0,0)^t$. Le dégradé de couleurs correspond au niveau des régions de confiance obtenues par vraisemblance empirique. Les ellipses en pointillés correspondent quant à elles aux régions de confiance (de niveau 50% et 90%) obtenues grâce au théorème de la limite centrale. On peut remarquer que les régions obtenues par vraisemblance empirique captent la géométrie des données.

3.4 Une généralisation du théorème d'Owen

3.4.1 Le théorème de Hjort *et al.* (2004)

Dans tout ce qui suit, on considère une suite $(X_n)_{n\in\mathbb{N}^\star}$ de variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^p$ et de loi inconnue P_0. Le paramètre que l'on cherche à estimer est $\mu_0 = \mu_0(P_0) \in \mathbb{R}^p$. On supposera qu'il existe une fonction estimante $m = m(X_i, \mu, \widehat{h})$ à valeurs dans $\mathbb{R}^d$ muni de la norme du sup (notée $\|\cdot\|$) et d'espérance nulle, où $\widehat{h}$ est un paramètre de nuisance de valeur réelle $h_0 = h_0(P_0)$. Pour bien comprendre à quoi correspond cette fonction m, plaçons-nous par exemple dans le cadre du théorème d'Owen. Il n'y a alors pas de paramètre de nuisance $\widehat{h}$ et la fonction m est donnée par :

$$m(X_i, \mu) = X_i - \mu.$$

Nous verrons dans les applications des exemples où intervient un « paramètre de nuisance ». Cela correspondra en particulier à des cas où on ne peut pas calculer la fonction m mais seulement une estimation $\widehat{m}$ de m. Par exemple, si m dépend de la variance (inconnue) σ^2 des X_i, on pourra considérer la fonction $\widehat{m}$ où on aura remplacé σ^2 par la variance empirique des X_i dans l'expression de m. On notera alors indifféremment $\widehat{m}(X_i, \mu)$ ou $m(X_i, \mu, \widehat{h})$.

Si A est une matrice, on notera $A \succ 0$ pour dire que A est définie positive. Si x est un vecteur, on écrira $x^{\otimes 2}$ la matrice xx^t. Et, par analogie avec le théorème d'Owen, on considérera les quantités suivantes :

$$\mathrm{EL}_n(\mu, \widehat{h}) := \max\left\{ \prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i m(X_i, \mu, \widehat{h}) = 0 \text{ et } p \in \mathbb{S}_{n-1} \right\},$$

$$M_n(\mu, \widehat{h}) := \frac{1}{n}\sum_{i=1}^n m(X_i, \mu, \widehat{h}),$$

$$S_n(\mu, \widehat{h}) := \frac{1}{n}\sum_{i=1}^n m(X_i, \mu, \widehat{h})^{\otimes 2}.$$

Théorème 3.2 (Hjort *et al.* (2004))

Supposons que :

(A0) $\mathbb{P}\left(\mathrm{EL}_n(\mu_0, \widehat{h}) = 0\right) \longrightarrow 0$, lorsque $n \to +\infty$;

(A1) $\sqrt{n} \cdot M_n(\mu_0, \widehat{h}) \xrightarrow{\mathcal{L}} U$, lorsque $n \to +\infty$;

(A2) $S_n(\mu_0, \widehat{h}) \xrightarrow{P} V$, où $V \succ 0$, lorsque $n \to +\infty$;

(A3) $\max\limits_{1\leqslant i\leqslant n} \|m(X_i, \mu_0, \widehat{h})\| = o_p(n^{1/2})$.

Alors :

$$-2\ln \mathrm{EL}_n(\mu_0, \widehat{h}) \xrightarrow{\mathcal{L}} U^t V^{-1} U.$$

Remarque 3.3. On peut remplacer la loi limite dans le théorème par une somme pondérée de χ^2 indépendantes comme l'ont remarqué Qin & Jing (2001). En effet, supposons que $Z_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id_p)$ et que $A \in M_p(\mathbb{R})$ est une matrice définie positive admettant $r_1, \dots, r_p$ pour valeurs propres. Alors :

$$Z_n^t A Z_n \xrightarrow{\mathcal{L}} r_1\chi^2_{1,1} + \dots + r_p\chi^2_{p,1}.$$

Pour le démontrer, diagonalisons A dans une base orthonormale de vecteurs propres. Écrivons alors $A = P^t D P$ avec $D = \text{diag}\,(r_1, \dots, r_p)$, et posons $\widetilde{Z}_n = P^t Z_n$. Comme la matrice P est orthonormale, le vecteur $\widetilde{Z}_n$ converge encore en loi vers une loi $\mathcal{N}(0, Id_p)$. D'où :

$$Z_n^t A Z_n = (PZ_n)^t D (PZ_n) \xrightarrow{\mathcal{L}} r_1\chi^2_{1,1} + \dots + r_p\chi^2_{p,1},$$

puisque $Z_n^t A Z_n = r_1\widetilde{Z}^2_{n,1} + \dots + r_p\widetilde{Z}^2_{p,1}$.

3.4.2 Démonstration du théorème de Hjort *et al.* (2004)

Problème dual

Étonnamment, dans les papiers présentant des résultats de vraisemblance empirique, on utilise des arguments de dualité forte sans prendre le temps de vérifier leur légitimité. Nous allons donner ici des conditions suffisantes pour lesquelles il y a dualité forte, et nous verrons que dans le cadre du théorème de Hjort *et al.* (2004), ces conditions sont satisfaites. Le problème que l'on considère dans cette section est :

$$\max_{p \in \mathfrak{X}} f(p),$$

soumis aux contraintes :

$$g_1(p) = 0, \dots, g_m(p) = 0.$$

Pour simplifier l'écriture, on notera g le vecteur $(g_1, \dots, g_m)^t$, et C l'ensemble des points admissibles. On suppose de plus que toutes les fonctions du problème sont continues et que $\mathfrak{X}$ est une partie de $\mathbb{R}^n$.

Définition 3.4 (Fonction de Lagrange)**.** La fonction de Lagrange associée au problème est :

$$\begin{array}{rccc} L & : & \mathfrak{X} \times \mathbb{R}^m & \to & \mathbb{R} \\ & & (p, \lambda) & \mapsto & f(p) + \langle \lambda, g(p) \rangle \end{array}.$$

Remarque 3.5. Si on pose :

$$\begin{array}{rccc} \eta & : & \mathbb{R}^n & \to & \mathbb{R} \\ & & p & \mapsto & \inf\limits_{\lambda \in \mathbb{R}^n} L(p, \lambda) = \begin{cases} -\infty & \text{si } \exists j : g_j(p) \neq 0, \\ f(p) & \text{sinon,} \end{cases} \end{array}$$

alors,

$$\sup_{p \in C} f(p) = \sup_{p \in \mathfrak{X}} \eta(p) = \sup_{p \in \mathfrak{X}} \inf_{\lambda \in \mathbb{R}^n} L(p, \lambda).$$

La dualité lagrangienne consiste à regarder le problème où le supremum et l'infimum sont pris dans l'ordre inverse.

Définition 3.6 (Fonction duale)**.** La fonction duale du problème de départ est :

$$\begin{array}{rccl} \psi & : & \mathbb{R}^m & \to & \mathbb{R} \\ & & \lambda & \mapsto & \sup\limits_{p\in\mathfrak{X}} L(p,\lambda) \end{array}$$

Notation 3.7. Pour tout $\lambda \in \mathbb{R}^m$, on pose :

$$\mathfrak{X}_\lambda := \left\{ p \in \mathbb{R}^n : p \text{ maximise } L(\cdot\,,\lambda) \right\}.$$

Théorème 3.8 (dualité forte)

Si λ^* minimise ψ et que ψ est différentiable en λ^*, alors, pour toute solution p^* du problème primal, $\psi(\lambda^*) = f(p^*)$. Si de plus $\mathfrak{X}_{\lambda^*} \neq \emptyset$, alors, pour tout $p_\lambda \in \mathfrak{X}_{\lambda^*}$, p_λ est solution du problème primal.

Maintenant que le théorème de dualité forte a été énoncé, revenons au problème de départ. Quitte à passer au logarithme, on peut l'écrire sous la forme :

$$\max_{p\in\mathfrak{X}} \sum_{i=1}^n \ln(np_i), \quad s.c. \quad \begin{cases} \sum_{i=1}^n p_i - 1 = 0, \\ \sum_{i=1}^n p_i m(X_i, \mu_0, \widehat{h}) = 0, \end{cases}$$

où $\mathfrak{X} = (\mathbb{R}_{+*})^n$. Pour alléger les notations, nous écrirons $\widehat{m}_i = m(X_i, \mu_0, \widehat{h})$. Remarquons que d'après l'hypothèse (A0), avec probabilité tendant vers 1, 0 est dans l'enveloppe convexe des $\widehat{m}_i$ et l'ensemble C des points admissibles (celui sur lequel on maximise la fonction) est non vide. La fonction de Lagrange associée au problème est alors :

$$L(p,u,v) = \sum_{i=1}^n \ln(np_i) - u\left(\sum_{i=1}^n p_i - 1\right) - v^t\left(\sum_{i=1}^n p_i\widehat{m}_i\right).$$

Le problème étant séparable au sens où maximiser $L(p,u,v)$ en p revient à maximiser pour tout i les fonctions φ_i définies ci-dessous, les calculs sont très simples :

$$\varphi_i(p_i,u,v) := \ln(np_i) - u(p_i - 1) - p_i v^t\widehat{m}_i.$$

Mais, les φ_i étant concaves en p_i, on cherche à annuler leurs dérivées et on trouve :

$$\frac{\partial\varphi_i}{\partial p_i} = \frac{1}{p_i} - u - v^t\widehat{m}_i = 0 \Leftrightarrow p_i = \frac{1}{u + v^t\widehat{m}_i}.$$

On en déduit donc l'expression de la fonction duale ψ :

$$\begin{aligned} \psi(u,v) &= \sup_{p\in\mathfrak{X}} L(p,u,v) \\ &= \sum_{i=1}^n \ln\left(\frac{n}{u + v^t\widehat{m}_i}\right) - u\left(\sum_{i=1}^n \frac{1}{u + v^t\widehat{m}_i} - 1\right) - v\left(\sum_{i=1}^n \frac{\widehat{m}_i}{u + v^t\widehat{m}_i}\right) \\ &= \sum_{i=1}^n \ln\left(\frac{n}{u + v^t\widehat{m}_i}\right) - \sum_{i=1}^n \left(\frac{1}{u + v^t\widehat{m}_i}\left(u - \frac{u^2}{n} - \frac{u}{n}v^t\widehat{m}_i + v^t\widehat{m}_i\right)\right). \end{aligned}$$

D'où, après simplification :

$$\psi(u, v) = \sum_{i=1}^{n} \ln\left(\frac{n}{u + v^t \widehat{m}_i}\right) + n - u.$$

Pour utiliser le théorème de dualité forte on montre que la fonction ψ est différentiable à l'optimum. Remarquons que $O := \{(u, v) : \psi(u, v) < \infty\}$ est un ouvert. De plus, ψ est clairement différentiable sur O. Il y a donc dualité forte. Avant de chercher où ψ atteint son minimum, on remarque que pour tout $p \in C$:

$$0 = \sum_{i=1}^{n} p_i \frac{\partial \varphi_i}{\partial p_i} = n - u \sum_{i=1}^{n} p_i - v^t \sum_{i=1}^{n} p_i \widehat{m}_i = n - u.$$

On en déduit que $u = n$ et on peut alors réécrire la fonction duale :

$$\psi(u, v) = \psi(v) = \sum_{i=1}^{n} \ln\left(\frac{1}{1 + v^t \widehat{m}_i / n}\right).$$

Posons alors $\lambda = v/n$ pour simplifier les notations. La fonction duale étant convexe, on trouve son minimum en annulant son gradient et on remarque qu'il est atteint en λ tel que les p_i résolvant le problème primal exprimés en fonction de λ sont donnés par :

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^t \widehat{m}_i} \quad \text{et} \quad \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{m}_i}{1 + \lambda^t \widehat{m}_i} = 0.$$

Étude asymptotique de $\|\lambda\|$

Nous allons dorénavant reprendre la preuve dérivée de celle de Owen (1990) et présentée par Hjort *et al.* (2004) dans leur pré-print. On retrouve des preuves de ce type dans beaucoup de papier traitant de vraisemblance empirique. Dans la version finale de leur papier, Hjort *et al.* (2004) présentent une démonstration alternative intéressante en travaillant directement sur la fonction duale.

Dans tout ce qui suit, on munira $M_d(\mathbb{R})$ de la norme subordonnée à la norme du sup dont on a équipé $\mathbb{R}^d$: on la noterons encore $\|\cdot\|$.

D'après les calculs menés à la section précédente, on peut exprimer $\mathrm{EL}_n(\mu_0, \widehat{h})$ à l'aide du paramètre λ et on a :

$$\mathrm{EL}_n(\mu_0, \widehat{h}) = \prod_{i=1}^{n} (1 + \lambda^t \widehat{m}_i)^{-1},$$

où λ est tel que

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{m}_i}{1 + \lambda^t \widehat{m}_i} = 0.$$

Notons $\lambda = \|\lambda\| u$. Nous allons montrer que $\|\lambda\| = O_p(n^{-1/2})$. Pour ce faire, on remarque que :

$$\begin{aligned} 0 = u^t \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{m}_i}{1 + \lambda^t \widehat{m}_i} &= u^t \frac{1}{n} \sum_{i=1}^{n} \widehat{m}_i \left(1 - \frac{\lambda^t \widehat{m}_i}{1 + \lambda^t \widehat{m}_i} \right) \\ &= u^t \frac{1}{n} \sum_{i=1}^{n} \widehat{m}_i - u^t \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{m}_i \lambda^t \widehat{m}_i}{1 + \lambda^t \widehat{m}_i} \\ &= u^t M_n(\mu_0, \widehat{h}) - \|\lambda\| u^t \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{m}_i \widehat{m}_i^t}{1 + \lambda^t \widehat{m}_i} u. \end{aligned}$$

Ainsi :

$$\begin{aligned} \|\lambda\| u^t S_n(\mu_0, \widehat{h}) u &\leqslant \|\lambda\| u^t \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{m}_i \widehat{m}_i^t}{1 + \lambda^t \widehat{m}_i} u (1 + \max_{1 \leqslant i \leqslant n} \lambda^t \widehat{m}_i) \\ &\leqslant \|\lambda\| u^t \frac{1}{n} \sum_{i=1}^{n} \frac{\widehat{m}_i \widehat{m}_i^t}{1 + \lambda^t \widehat{m}_i} u \left(1 + \|\lambda\| \max_{1 \leqslant i \leqslant n} \|\widehat{m}_i\| \right) \\ &= u^t M_n(\mu_0, \widehat{h}) \left(1 + \|\lambda\| \max_{1 \leqslant i \leqslant n} \|\widehat{m}_i\| \right). \end{aligned}$$

D'où :

$$\|\lambda\| \left(u^t S_n(\mu_0, \widehat{h}) u - \max_{1 \leqslant i \leqslant n} \|\widehat{m}_i\| u^t M_n(\mu_0, \widehat{h}) \right) \leqslant u^t M_n(\mu_0, \widehat{h}).$$

Soient maintenant $\sigma_1 \leqslant \cdots \leqslant \sigma_p$ les valeurs propres de V. On sait d'après l'hypothèse (A2) que $\sigma_1 > 0$. Donc, presque sûrement :

$$\underline{\lim}_n \, u^t S_n(\mu_0, \widehat{h}) u \geqslant \sigma_1 > 0 \qquad \text{et} \qquad \overline{\lim}_n \, u^t S_n(\mu_0, \widehat{h}) u \leqslant \sigma_p.$$

D'autre part, d'après (A3) et (A1) :

$$\left\| \max_{1 \leqslant i \leqslant n} \|\widehat{m}_i\| u^t M_n(\mu_0, \widehat{h}) \right\| = o_p(n^{1/2}) O_p(n^{-1/2}) = o_p(1).$$

Ainsi :

$$\|\lambda\| \left(u^t S_n(\mu_0, \widehat{h}) u - o_p(1) \right) = O_p(n^{-1/2}).$$

Finalement :

$$\|\lambda\| = O_p(n^{-1/2}).$$

Étude asymptotique du rapport de vraisemblance empirique

Maintenant que l'on a étudié le comportement asymptotique du multiplicateur de Lagrange λ, on va pouvoir, comme dans la démonstration usuelle

du théorème de la limite centrale, effectuer un développement limité pour obtenir une convergence en loi. On a :

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{m}_i}{1+\lambda^t \widehat{m}_i} \\
&= \frac{1}{n}\sum_{i=1}^{n} \widehat{m}_i \left(1 - \lambda^t \widehat{m}_i + \frac{(\lambda^t \widehat{m}_i)^2}{1+\lambda^t \widehat{m}_i}\right) \\
&= M_n(\mu_0, \widehat{h}) - S_n(\mu_0, \widehat{h})\lambda + \frac{1}{n}\sum_{i=1}^{n} \frac{\widehat{m}_i (\lambda^t \widehat{m}_i)^2}{1+\lambda^t \widehat{m}_i} \\
&=: M_n(\mu_0, \widehat{h}) - S_n(\mu_0, \widehat{h})\lambda + \alpha_n.
\end{aligned}$$

Or, on sait que $|1+\lambda^t \widehat{m}_i| = O_p(1)$, donc, d'après (A2), (A3) et le résultat sur $\|\lambda\|$:

$$\begin{aligned}
\|\alpha_n\| &\leqslant \|\lambda\|^2 \left\| \frac{1}{n}\sum_{i=1}^{n} \widehat{m}_i^{\otimes 2} \right\| \max_{1\leqslant i\leqslant n} \|\widehat{m}_i\| \\
&= O_p(n^{-1}) O_p(1) o_p(n^{1/2}) \\
&= o_p(n^{1/2}).
\end{aligned}$$

On peut donc écrire :

$$\lambda = S_n(\mu_0, \widehat{h})^{-1} M_n(\mu_0, \widehat{h}) + \beta_n,$$

avec $\beta_n = o_p(n^{-1/2})$. Par ailleurs, on a vu que $\|\lambda\| = O_p(n^{-1/2})$, et par l'hypothèse (A3), on sait que $\max_i \|\widehat{m}_i\| = o_p(n^{1/2})$. Ainsi, $\max_i |\lambda^t \widehat{m}_i| = o_p(1)$. On a alors le développement limité :

$$\ln(1+\lambda^t \widehat{m}_i) = \lambda^t \widehat{m}_i - \frac{1}{2}(\lambda^t \widehat{m}_i)^2 + \frac{1}{3}\frac{(\lambda^t \widehat{m}_i)^3}{(1+\xi_{n,i})^3}, \quad \text{où } |\xi_i| \leqslant \lambda^t \widehat{m}_i.$$

Remarquons encore que :

$$\left\| \sum_{i=1}^{n} \frac{(\lambda^t \widehat{m}_i)^3}{(1+\xi_{n,i})^3} \right\| = o_p(1).$$

Pour s'en convaincre, on regardera la majoration de α_n un peu plus haut. De plus :

$$\beta_n^t S_n(\mu_0, \widehat{h})\beta_n = o_p(n^{-1/2}) O_p(1) o_p(n^{-1/2}) = o_p(n^{-1}).$$

Finalement, en écrivant $M = M_n(\mu_0, \widehat{h})$, $\beta = \beta_n$ et $S = S_n(\mu_0, \widehat{h})$ pour simplifier les notations, on obtient :

$$\begin{aligned}
-2\ln \mathrm{EL}_n(\mu_0, \widehat{h}) &= 2\sum_{i=1}^{n} \ln(1 + \lambda^t \widehat{m}_i) \\
&= 2\sum_{i=1}^{n} \lambda^t \widehat{m}_i - \sum_{i=1}^{n} (\lambda^t \widehat{m}_i)^2 + \frac{2}{3}\sum_{i=1}^{n} \frac{(\lambda^t \widehat{m}_i)^3}{(1+\xi_{n,i})^3} \\
&= 2n\lambda^t M - n\lambda^t S\lambda + o_p(1) \\
&= 2nM^t S^{-1}M + 2n\beta^t M - nM^t S^{-1}M - 2nM^t\beta - n\beta^t S\beta + o_p(1) \\
&= nM_n(\mu_0, \widehat{h})^t S_n(\mu_0, \widehat{h})^{-1} M_n(\mu_0, \widehat{h}) + o_p(1) \\
&\xrightarrow{\mathcal{L}} U^t V^{-1} U,
\end{aligned}$$

où la convergence en loi est obtenue en utilisant les hypothèses (A1) et (A2) ainsi que le lemme de Slutsky. □

Le théorème d'Owen peut être vu rétrospectivement comme une conséquence du théorème 3.2. Il s'agit comme nous l'avons déjà remarqué du cas où il n'y a pas de paramètre $\widehat{h}$ de perturbation et où la fonction m est donnée par $m(X_i, \mu, \widehat{h}) = X_i - \mu$. Quitte à se restreindre à $\mathrm{Im}(X_1)$ on peut supposer que V_0, la matrice de variance-covariance des X_i, est définie positive. On a grâce au théorème de la limite centrale et à la loi des grands nombres :

$$\sqrt{n}\, M_n(\mu_0) = \sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0) = \sqrt{n}\left(\bar{X} - \mu_0\right) \xrightarrow{\mathcal{L}} U, \text{ où } U \sim \mathcal{N}_p(0, V_0),$$

$$S_n(\mu_0) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)(X_i - \mu_0)^t \xrightarrow{P} V_0.$$

Pour démontrer la condition (A0), Owen (1990) montre que tous les demi-espaces autour de μ sont occupés asymptotiquement par les X_i (on se reportera à Owen (1990) page 217 pour plus de détails). Nous présenterons des résultats similaires par la suite, c'est pourquoi nous ne donnons pas plus de détails ici. La condition (A3) du théorème est quant à elle vérifiée grâce au lemme ci-dessous, que l'on trouve dans Owen (1990).

Lemme 3.9. *Soit $(Z_n)_{n\in\mathbb{N}^\star}$ une suite de variables aléatoires i.i.d de L^2. Alors :*

$$\max_{1\leqslant i\leqslant n} |Z_i| = o_p(n^{1/2}).$$

Démonstration. On a :

$$\sum_{n=1}^{\infty} \mathbb{P}(Z_n^2 > n) = \sum_{n=1}^{\infty}\sum_{i=n}^{\infty} \mathbb{P}(Z_n^2 \in]i, i+1]) = \sum_{i=1}^{\infty}\sum_{n=1}^{i} \mathbb{P}(Z_n^2 \in]i, i+1]).$$

Donc :

$$\sum_{n=1}^{\infty} \mathbb{P}(Z_n^2 > n) = \sum_{i=1}^{\infty} i\mathbb{P}(Z_1^2 \in]i, i+1]) = \sum_{i=1}^{\infty} \mathbb{E}(i\mathbb{1}_{\{Z_1^2 \in]i, i+1]\}}).$$

Finalement :

$$\sum_{n=1}^{\infty} \mathbb{P}(Z_n^2 > n) \leqslant \sum_{i=1}^{\infty} \mathbb{E}(Z_1^2 \mathbb{1}_{\{Z_1^2 \in]i,i+1]\}}) \leqslant \mathbb{E}(Z_1^2) < \infty.$$

Donc, d'après le lemme de Borel-Cantelli, avec probabilité 1, l'événement $\left\{|Z_n| > n^{1/2}\right\}$ n'est réalisé que pour un nombre fini de n. Il n'y a donc qu'un nombre fini de n tels que $\max_{1\leqslant i\leqslant n} |Z_i| > n^{1/2}$. D'où, presque sûrement :

$$\overline{\lim_n} \max_{1\leqslant i\leqslant n} |Z_i| n^{-1/2} \leqslant 1.$$

De même, pour tout $A > 0$, presque sûrement :

$$\overline{\lim_n} \max_{1\leqslant i\leqslant n} |Z_i| n^{-1/2} \leqslant A.$$

Donc :

$$\mathbb{P}\left(\bigcap_{A\in\mathbb{Q}_+} \left\{\overline{\lim_n} \max_{1\leqslant i\leqslant n} |Z_i| n^{-1/2} \leqslant A\right\}\right) = 1.$$

□

Les conditions (A0), (A1), (A2) et (A3) étant vérifiées, il y a convergence en loi de $\mathrm{EL}_n(\mu_0)$ vers $U^t V_0^{-1} U$. Or, $U \sim \mathcal{N}_p(0, V_0)$. On peut donc écrire $U = V_0^{1/2} N$ avec $N \sim \mathcal{N}_p(0, Id)$. Ainsi :

$$U^t V_0^{-1} U = N^t (V_0^{1/2})^t V_0^{-1} V_0^{1/2} N = N^t N \sim \chi_q^2.$$

3.4.3 Généralisation

Dans cette section, nous donnons une généralisation du théorème de Hjort *et al.* (2004) où nous ne faisons plus d'hypothèse sur le rang de la matrice de variance-covariance. Il s'agit en fait d'un simple jeu d'écriture sur la loi limite. Bien que cela puisse paraître inutile de prime abord, nous verrons que dans le cas de processus complexes, cela peut simplifier grandement les calculs.

Définition 3.10. Soit $A \in M_n(\mathbb{R})$ une matrice symétrique positive. Notons $\lambda_1, \ldots, \lambda_p$ ses valeurs propres non nulles. Soit P une matrice orthonormale telle que :

$$A = P \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & 0_{p,n-p} \\ & & \lambda_p & \\ & 0_{n-p,p} & & 0_{n-p,n-p} \end{pmatrix} P^t.$$

On appelle inverse généralisée de A et on note $A^\dagger$ la matrice :

$$A^\dagger = P \begin{pmatrix} 1/\lambda_1 & & & \\ & \ddots & & 0_{p,n-p} \\ & & 1/\lambda_p & \\ & 0_{n-p,p} & & 0_{n-p,n-p} \end{pmatrix} P^t.$$

Reprenons les notations de la section 3.4.1. Supposons que la matrice V de variance-covariance des $\widehat{m}_i := m(X_i, \mu_0, \widehat{h})$ admette k valeurs propres non nulles et que toutes les autres hypothèses du théorème 3.2 soient satisfaites. Soit P une matrice de passage comme dans la définition précédente. Notons :

$$Y_i := P^t \widehat{m}_i \qquad \text{et} \qquad \mathbb{V}\mathrm{ar}\, Y_i = P^t\, \mathbb{V}\mathrm{ar}(\widehat{m}_i)\, P =: \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix},$$

où D est une matrice diagonale de taille $k \times k$ et de rang plein. Rappelons que nous étudions la variable aléatoire :

$$\mathrm{EL}_n(\mu_0, \widehat{h}) = \max \left\{ \prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i \widehat{m}_i = 0 \text{ et } p \in \mathbb{S}_{n-1} \right\}.$$

Soit maintenant $Z_i := (Y_{i,1}, \dots, Y_{i,k})^t \in \mathbb{R}^k$ le vecteur contenant les k premières composantes du vecteur Y. On a :

$$\sum_{i=1}^n p_i \widehat{m}_i = 0 \Longleftrightarrow \sum_{i=1}^n p_i Y_i = 0 \Longleftrightarrow \sum_{i=1}^n p_i Z_i = 0.$$

Or, comme les Z_i ont une matrice de variance-covariance de rang plein (D), on peut leur appliquer le théorème 3.2. En effet, toutes les hypothèses sont clairement vérifiées et on trouve :

$$\begin{aligned} -2 \ln \mathrm{EL}_n(\mu_0, \widehat{h}) &\rightsquigarrow \Big((P^t U)_1, \dots, (P^t U)_k \Big) D^{-1} \begin{pmatrix} (P^t U)_1 \\ \vdots \\ (P^t U)_k \end{pmatrix} \\ &= (P^t U)^t \begin{pmatrix} D^{-1} & 0 \\ 0 & 0 \end{pmatrix} P^t U \\ &= U^t V^\dagger U. \end{aligned}$$

Ainsi, on peut enlever l'hypothèse sur le rang de la matrice V dans le théorème 3.2 en remplaçant V^{-1} par $V^\dagger$ dans la loi limite. Remarquons que si V est inversible, alors $V^\dagger = V^{-1}$ et on retrouve bien la même formule.

3.5 Extension aux processus

On peut trouver dans le pre-print de l'article de Hjort *et al.* (2004) une version de leur théorème pour des processus. Ce sera ce résultat de vraisemblance empirique que nous utiliserons dans la suite. Les notations sont très similaires à celles présentées à la section 3.4.1 à ceci près qu'il y a un paramètre de temps t supplémentaire.

Soit $(X_n(\cdot))_{n\in\mathbb{N}}$ une suite de processus aléatoires indexés par un espace pseudo-métrique (T, ρ). On suppose que les processus Z_i sont i.i.d. et que les $X_i(t)$ sont à valeurs dans $\mathbb{R}^q$ et de loi inconnue P_0. Le paramètre que l'on cherche à estimer est $\mu_0(t) = \mu_0(P_0, t)$. On l'estime grâce à une fonction

$m(X, \mu, \widehat{h}, t)$, où $\widehat{h}(t)$ est un paramètre de nuisance de valeur réelle $h_0(t) = h_0(P_0, t)$. On note :

$$\mathrm{EL}_n(\mu, \widehat{h}, t) := \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i m(X_i, \mu, \widehat{h}, t) = 0 \text{ et } p \in \mathbb{S}_{n-1}\right\},$$

$$M_n(\mu, \widehat{h}, t) := \frac{1}{n}\sum_{i=1}^n m(X_i, \mu, \widehat{h}, t),$$

$$S_n(\mu, \widehat{h}, t) := \frac{1}{n}\sum_{i=1}^n m(X_i, \mu, \widehat{h}, t)^{\otimes 2}.$$

Théorème 3.11 (Hjort *et al.* (2004))

Supposons qu'il existe deux constantes δ et M strictement positives ainsi qu'une matrice $V(\cdot)$ dont les valeurs propres sont uniformément bornées dans l'intervalle $[\delta, M]$ sur T. Supposons de plus que :

(A0*) $\mathbb{P}\left(\exists t \in T :\ \mathrm{EL}_n(\mu_0, \widehat{h}, t) = 0\right) \longrightarrow 0$, lorsque $n \to +\infty$;

(A1*) $\sqrt{n} \cdot M_n(\mu_0, \widehat{h}, \cdot) \rightsquigarrow U(\cdot)$, lorsque $n \to +\infty$;

(A2*) $\sup_{t\in T} |S_n(\mu_0, \widehat{h}, t) - V(t)| \xrightarrow{\mathbb{P}^\star} 0$, lorsque $n \to +\infty$;

(A3*) $\max_{1\leqslant i\leqslant n} \sup_{t\in T} \|m(X_i, \mu_0, \widehat{h}, t)\| = o_p(n^{1/2})$.

Alors :

$$-2\ln \mathrm{EL}_n(\mu_0, \widehat{h}, \cdot) \rightsquigarrow U^t(\cdot)V^{-1}(\cdot)U(\cdot).$$

Démonstration. La preuve de ce théorème est omise dans le pré-print de Hjort *et al.* (2004), c'est pourquoi nous présentons ici une idée de la démonstration. Pour montrer la convergence en distribution, on utilise la caractérisation donnée dans le théorème 1.26. La convergence des marges finidimensionnelles est une conséquence du théorème 3.2. De plus, en reprenant la démonstration du théorème 3.2, on remarque qu'il suffit de montrer l'équicontinuité asymptotique de :

$$\sqrt{n}M_n(\mu_0, \widehat{h}, \cdot)S_n(\mu_0, \widehat{h}, \cdot)\sqrt{n}M_n(\mu_0, \widehat{h}, \cdot),$$

En effet, comme les valeurs propres de V sont uniformément bornées sur T, le reste dans le développement asymptotique de $-2\ln \mathrm{EL}_n(\mu_0, \widehat{h}, \cdot)$ est encore un $o_p(1)$. Or, $S_n(\mu_0, \widehat{h}, \cdot)$ converge vers V en probabilité extérieure dans $(\mathcal{D}(T), \|\cdot\|_\infty)$ et $\sqrt{n}M_n(\mu_0, \widehat{h}, \cdot)$ converge en loi vers U. Il suffit donc d'appliquer le lemme de Slutsky (cf. page 32 de van der Vaart & Wellner (1996)) ainsi que le transport de la convergence en loi par composition avec des applications continues pour démontrer le résultat. □

Remarque 3.12. Tout comme pour le théorème 3.2, on peut affaiblir l'hypothèse sur la matrice de variance-covariance et supposer que les valeurs propres non nulles de $V(t)$ sont uniformément bornées sur T dans l'intervalle $[\delta, M]$. En effet, d'après ce qui a été vu à la section 3.4.3, on a encore convergence des marges fini-dimensionnelles (en remplaçant V^{-1} par $V^{\dagger}$ dans la loi limite). De plus, le reste dans le développement asymptotique de $-2\ln \mathrm{EL}_n(\mu_0, \widehat{h}, \cdot)$ est encore un $o_p(1)$ puisque seules les valeurs propres non nulles de V interviennent et qu'elles sont uniformément bornées sur T.

3.6 Bootstrap

Au chapitre 7, nous avons effectué des simulations pour déterminer l'efficacité de nos résultats sur des échantillons de taille finie. Ces résultats faisant appel au théorème 3.11, nous avons décidé d'utiliser la technique du bootstrap pour déterminer un quantile du supremum de la loi limite. Nous allons dans cette section expliquer succinctement pourquoi même dans un cas simple des problèmes peuvent se poser.

Plaçons-nous alors dans le cadre du théorème d'Owen. Nous sommes donc en présence d'une suite de variables aléatoires i.i.d. $(X_i)_{i\in\mathbb{N}^\star}$ dans $L^2(\Omega)$ et nous cherchons une région de confiance pour l'espérance commune des X_i. Plutôt que d'utiliser la loi limite (connue) du rapport de vraisemblance empirique, nous allons l'estimer par bootstrap. Nous noterons n la taille de l'échantillon.

On tire donc $(X_1^\star, \dots, X_n^\star)$ uniformément dans $\{X_1, \dots, X_n\}$ et on calcule le rapport de vraisemblance empirique en $\bar{X}$:

$$\mathrm{EL}_n^\star(\bar{X}) = \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i X_i^\star = \bar{X} \text{ et } p \in \mathbb{S}_{n-1}\right\}$$

On a vu que, asymptotiquement, l'espérance commune des X_i se trouve dans l'enveloppe convexe de $\{X_1, \dots, X_n\}$. Cependant, il se peut, comme on l'a illustré à la figure 3.2, que $\bar{X}$ ne soit pas dans l'enveloppe convexe des $X_i^\star$, en particulier pour des tailles d'échantillon petites. Dans ce cas, on avait posé par convention :

$$\mathrm{EL}_n^\star(\bar{X}) = \max \varnothing = 0,$$

Ceci va donc induire un biais et on aura tendance à sous-évaluer le quantile de la loi limite de EL_n, donc à sur-évaluer celui de $-2\ln \mathrm{EL}_n$.

Comment peut-on alors contourner ce problème ? Le plus simple est sans doute de ne pas tenir compte de cette itération bootstrap. Cependant, comme nous le verrons plus tard (cf. section 7.4.3), ce n'est pas toujours aussi simple, en particulier lorsqu'on est en présence de processus.

Remarque 3.13. Imaginons que l'on connaisse la vraie espérance μ_0 des X_i et que l'on veuille vérifier numériquement que le bootstrap fonctionne bien. A-t-on intérêt à calculer $\mathrm{EL}_n^\star(\mu_0)$? La réponse est non et la raison est simple : l'espérance des $X_i^\star$ est $\bar{X}$ et non μ_0. Ainsi, $\mathrm{EL}_n^\star(\mu_0)$ aura tendance à être strictement inférieur à $\mathrm{EL}_n^\star(\bar{X})$ en moyenne.

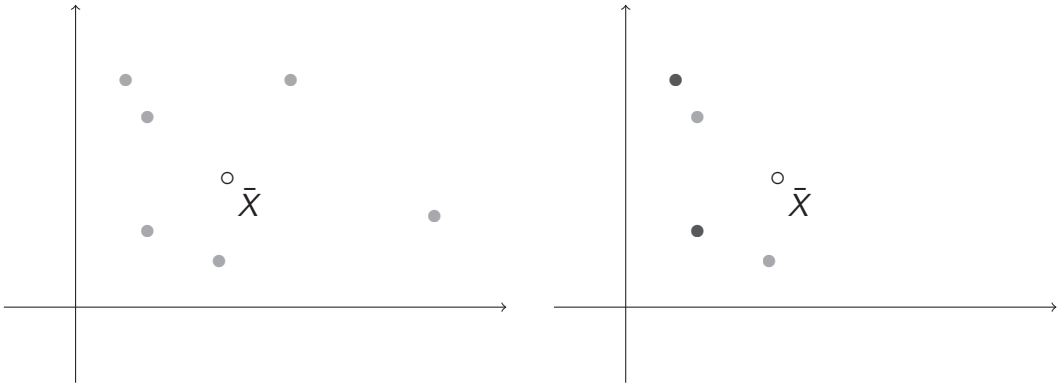


FIGURE 3.2 – Vraisemblance empirique et bootstrap. À gauche, les X_i, à droite les $X_i^\star$. Les points foncés représentent des doublons.

Remarque 3.14. Rappelons que la loi limite dans le théorème 3.2 est de la forme $U^t V^{-1} U$. Dans leur article, Hjort *et al.* (2004) proposent d'estimer U par bootstrap et V par :

$$\widehat{V} := \frac{1}{n} \sum_{i=1}^{n} m(X_i, \widehat{\mu}, \widehat{h})^{\otimes 2},$$

où $\widehat{\mu}$ est un estimateur consistant de μ_0. Ils donnent des conditions suffisantes pour lesquelles cette approximation fonctionne. Il est possible que leur méthode s'étende aux problématiques impliquant des processus (et en particulier au théorème 3.11). Cependant, nous ne nous sommes pas penché sur la question et avons bootstrappé directement le processus limite dans nos simulations.

CHAPITRE 4

Processus de comptage

4.1 Introduction

Les processus de comptage peuvent être utilisés pour décrire un grand nombre de phénomènes. En biostatistiques, ils peuvent modéliser la contraction d'infections nosocomiales par un patient au cours du temps ; en fiabilité, on peut les utiliser pour modéliser les pannes récurrentes d'un système. En outre, les propriétés de sous-martingales inhérentes aux processus de comptage en font des objets de choix pour le statisticien, en particulier grâce à des résultats de type théorème de la limite centrale pour les martingales. Ainsi, il n'est pas étonnant que la théorie des martingales soit le fondement d'ouvrages traitant de processus de comptage comme le livre de Fleming & Harrington (1991) ou encore celui de Andersen *et al.* (1993).

Dans ce chapitre, nous donnons des résultats théoriques sur les processus de comptage, en particulier certains concernant les estimateurs de Nelson-Aalen et de Kaplan-Meier qui nous seront utiles aux chapitres 6 et 7.

4.2 Exemple

Soit X une durée de vie (i.e. une variable aléatoire positive) de fonction de répartition F. Si X est absolument continue de densité f, alors, on peut définir la fonction de risque instantané α par :

$$\begin{array}{rccc} \alpha & : & [0, \tau_F] & \to & \overline{\mathbb{R}^+} \\ & & t & \mapsto & \dfrac{f(t)}{S(t)}, \end{array}$$

où S est la fonction de survie de X définie par $S(t) = 1 - F(t) = \mathbb{P}(X > t)$, et où $\tau_F = \sup\{s \ : \ F(s) < 1\}$. Le nom *risque instantané* s'explique par la formule :

$$\alpha(t) = \lim_{u \to 0^+} \frac{1}{u} \mathbb{P}\left(X \in [t, t+u[\,|X \geqslant t\right).$$

Le *risque cumulé* est quant à lui :

$$\begin{array}{rccc} A & : & [0,\tau_F] & \to & \overline{\mathbb{R}^+} \\ & & t & \mapsto & \displaystyle\int_0^t \alpha(s)ds = -\ln S(t). \end{array}$$

Dans leur livre, Andersen *et al.* (1993) présentent un exemple introductif qui permet de mettre en exergue certains des objets clefs qui apparaissent dans les démonstrations. Il permet en outre de présenter une heuristique de la décomposition de Doob-Meyer pour les processus de comptage que nous introduirons formellement à la section 4.3. Reprenons ici cet exemple. Supposons alors être en présence de données censurées. Au lieu d'observer directement des réalisations de variables aléatoires $X_1, \dots, X_n$, nous observons des couples $(\widetilde{X}_i, \delta_i)$, $i = 1, \dots, n$, où les δ_i sont des indicateurs de censure :

$$\begin{aligned} X_i &= \widetilde{X}_i \quad \text{si } \delta_i = 1, \\ X_i &> \widetilde{X}_i \quad \text{si } \delta_i = 0. \end{aligned}$$

Notons $N(\cdot)$ le processus défini sur $[0,\tau]$ par :

$$N(t) := \sum_{i=1}^{n} \mathbb{1}_{\{\widetilde{X}_i \leqslant t, \delta_i = 1\}}.$$

Rappelons qu'une *filtration* $\mathscr{F} = (\mathscr{F}_t)_t$ est une suite croissante de tribus continue à droite [1]. Rappelons aussi que, par définition, le processus $N(\cdot)$ est dit *adapté* à la filtration $\mathscr{F}$ si $N(t)$ est $\mathscr{F}_t$-mesurable pour tout $t \in [0,\tau]$. Dans ce qui suit, nous considérerons la filtration *naturelle* associée au problème [2].
Dans ce cadre, si on suppose que les censures sont indépendantes des X_i, alors, de manière heuristique :

$$\lim_{u\to 0^+} \frac{1}{u} \mathbb{P}\left(\widetilde{X}_i \in [t, t+u[, \delta_i = 1 \,\middle|\, \mathcal{F}_{t^-}\right) = \begin{cases} \alpha(t) & \text{si } \widetilde{X}_i \geqslant t, \\ 0 & \text{si } \widetilde{X}_i < t. \end{cases}$$

Ainsi, si on pose $Y(t) = \sum_{i=1}^n \mathbb{1}_{\{\widetilde{X}_i \geqslant t\}}$, alors :

$$\begin{aligned} \lim_{u\to 0^+} \frac{1}{u} \mathbb{E}\left(\sum_{i=1}^{n} \mathbb{1}_{\{\widetilde{X}_i \in [t,t+u[, \delta_i = 1\}} \,\middle|\, \mathcal{F}_{t^-}\right) &= \left(\sum_{i=1}^{n} \mathbb{1}_{\{\widetilde{X}_i \geqslant t\}}\right) \alpha(t) \\ &= Y(t)\alpha(t). \end{aligned}$$

Notons maintenant $\mathrm{d}N(t)$ l'incrément $N((t+\mathrm{d}t)^-) - N(t^-)$ et λ la fonction $Y\alpha$. Alors, on peut résumer ce qui vient d'être vu par :

$$\mathbb{E}(\mathrm{d}N(t)|\mathcal{F}_{t^-}) = \lambda(t)\mathrm{d}t.$$

1. La continuité à droite signifie que pour tout s, $\mathscr{F}_s = \bigcap_{t>s} \mathscr{F}_t$.
2. La filtration naturelle est celle définie pour tout t par $\mathscr{F}_t := \sigma(\{N(s), s \leqslant t\})$.

Ainsi, l'application M définie ci-dessous est une martingale :

$$M(t) := N(t) - \Lambda(t) := N(t) - \int_0^t \lambda(s)\mathrm{d}s.$$

En effet, $\lambda(t)$ est $\mathcal{F}_{t^-}$-mesurable et donc :

$$\begin{aligned}\mathbb{E}(\mathrm{d}M(t)|\mathcal{F}_{t^-}) &= \mathbb{E}(\mathrm{d}N(t) - \mathrm{d}\Lambda(t)|\mathcal{F}_{t^-}) \\ &= \mathbb{E}(\mathrm{d}N(t) - \lambda(t)\mathrm{d}t|\mathcal{F}_{t^-}) \\ &= \mathbb{E}(\mathrm{d}N(t)|\mathcal{F}_{t^-}) - \lambda(t)\mathrm{d}t \\ &= 0.\end{aligned}$$

Nous voyons ici que dans un contexte de problème de durées de vie, il sera possible de construire de nouveaux objets et de profiter de leurs propriétés de martingales.

4.3 Intensité d'un processus de comptage

Un processus de comptage multivarié peut être vu comme un processus aléatoire qui compte le nombre d'occurrences de certains types d'événements distincts au cours du temps. Dans toute la suite, on étudiera des processus indexés par $[0, \tau]$, où τ est une constante fixée, et on supposera l'existence d'une filtration adaptée $(\mathcal{F}_t)_{t\in[0,\tau]}$.

Définition 4.1. On appelle *processus de comptage multivarié* tout vecteur $N = (N_1, \dots, N_k)$ de processus adaptés càdlàg prenant tous la valeur 0 au temps 0, dont les composantes sont croissantes, en escalier, admettant des sauts de taille 1 et tel que deux composantes ne sautent pas en même temps (*i.e.* $\mathbb{P}(\exists t \in [0,\tau], \exists i \neq j \ : \ N_i(t) - N_i(t^-) = N_j(t) - N_j(t^-) = 1) = 0$).

Rappelons maintenant un résultat essentiel en théorie des martingales, à savoir le théorème de décomposition de Doob-Meyer, que nous appliquerons à nos processus de comptage.

Théorème 4.2 (Décomposition de Doob-Meyer)

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ une base stochastique et $N(\cdot)$ une $(\mathcal{F}_t)_t$-sous-martingale positive. Alors, il existe une $(\mathcal{F}_t)_t$-martingale càdlàg $M(\cdot)$ et un processus prévisible croissant $\Lambda(\cdot)$ tels que :

i) $\forall t \in [0,\tau],\ \mathbb{E}|\Lambda(t)| < \infty$;

ii) $M(t) = N(t) - \Lambda(t)$;

Si on impose de plus $A_0 = 0$, alors cette décomposition est unique.

Définition 4.3 (compensateur)**.** Le processus prévisible $\Lambda(\cdot)$ dans la décomposition de Doob-Meyer est appelé *compensateur* de $N(\cdot)$.

Remarque 4.4. Soit $N(\cdot)$ un processus de comptage, il est donc croissant. Ainsi, pour tout $t \geqslant s$ on a $N(t) \geqslant N(s)$, et donc :

$$\mathbb{E}(N(t)|\mathscr{F}_s) \geqslant \mathbb{E}(N(s)|\mathscr{F}_s) = N(s).$$

Ainsi, les processus de comptage sont des sous-martingales positives et on peut leur appliquer la décomposition de Doob-Meyer.

Définition 4.5. Soit $N(\cdot)$ un processus de comptage et $\Lambda(\cdot)$ son compensateur. On dira que le processus prévisible $\lambda(\cdot)$ est une intensité de $N(\cdot)$ si :

$$\forall t \in [0, \tau] \ : \quad \Lambda(t) = \int_0^t \lambda(s)\mathrm{d}s.$$

4.4 Deux cas particuliers

Dans notre travail, nous nous sommes intéressés à plusieurs modèles, notamment ceux impliquant plusieurs types d'événements ou encore des problèmes multi-états. Ces modèles que nous allons détailler ici partagent une même particularité, celle de posséder une intensité dite *multiplicative*. Cela signifie que leur intensité $\lambda(\cdot)$ peut s'écrire sous la forme :

$$\lambda(t) = \alpha(t)Y(t),$$

où α est déterministe et où Y est prévisible. Dans ce cas, on dira que α est la fonction de risque instantané associée au processus de comptage.

Dans toute la suite on supposera que les processus de comptage sous-jacents sont presque sûrement bornés par une constante B. Cette hypothèse nous permettra de construire des classes euclidiennes bornées et d'appliquer la loi des grands nombres uniforme.

4.4.1 Événements de types multiples

Dans cette partie, on suppose que l'on observe des individus $i = 1, \dots, n$ pouvant expérimenter plusieurs types d'événements $h = 1, \dots, K$. On suppose de plus que l'observation de ces individus est soumise à une censure aléatoire à droite indépendante C. On pourra penser par exemple à des patients d'un hôpital pouvant contracter différent types d'infections nosocomiales, comme ce sera le cas au chapitre 7. Ainsi, C pourra correspondre à une fenêtre d'observation.

Notons alors $N^\star = (N_h^\star; h = 1, \dots, K)$ le processus de comptage correspondant, où h indexe les différents types d'événements. Notons encore $N_h(t) = N_h^\star(t \wedge C)$. Alors, les variables observées sont des réalisations i.i.d. $(N_{ih}(\cdot), \mathbb{1}_{\{C_i \geqslant \cdot\}})_{1 \leqslant i \leqslant n}$ de $(N_h(\cdot), \mathbb{1}_{\{C \geqslant \cdot\}})$, pour $h = 1, \dots, K$. Dans ces conditions, Andersen *et al.* (1993) montrent que les processus $N_h(\cdot)$ admettent une intensité multiplicative de la forme $\lambda_h = \alpha_h Y$, où $Y(t)$ représente le nombre d'individus « à risque » à l'instant t :

$$Y(t) := \sum_{i=1}^{n} \mathbb{1}_{\{C_i \geqslant t\}},$$

et où les $\alpha_h(t)$ sont les fonctions de risque instantané données par

$$\alpha_h(t) = \lim_{u \to 0^+} \frac{1}{u} \mathbb{P}(N_h(t+u) - N_h(t) = 1 | N(s), s < t).$$

Bien que Y ne dépendent pas de h, nous noterons $Y_h = Y$ à la section 4.5 pour que les notations soient cohérentes.

4.4.2 Processus de Markov non homogènes

Dans cette section, on suppose être en présence d'individus indexés par $i = 1, \dots, n$, ces individus pouvant passer d'un état $h = 1, \dots, k$ à un autre. On s'intéresse aux processus $N_{hj}(\cdot)$, $h \neq j$, comptant les transitions d'un état h à un état j et on suppose que les observations sont soumises à une censure aléatoire à droite C, c'est-à-dire qu'on n'observe plus l'état de la chaîne après un temps C aléatoire indépendant. On pourra penser par exemple à des individus pouvant être dans les états `sain`, `malade` ou `décédé`, ce dernier état étant absorbant. La censure aléatoire à droite pourrait être dans ce cas une fenêtre d'observation correspondant à la durée de l'étude. On modélise alors cette situation par un processus de Markov non homogène continu en temps et à espace d'états fini.

Notons alors $X_i(t)$ le processus décrivant l'état de l'individu i au temps t. Notons aussi $Y_h(t)$ le nombre d'individus « à risque » (c'est-à-dire dans l'état h) au temps t^- et α_{hj} les fonctions de risque instantané de passage de l'état h à l'état j. Formellement, posons (pour $h \neq j$) :

$$\alpha_{hj}(t) = \lim_{u \to 0^+} \frac{1}{u} \mathbb{P}(X_i(t+u) = j | X_i(t) = h),$$
$$Y_h(t) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i(t^-) = h\}}.$$

On peut montrer que, dans ce cadre, les processus de comptage en jeu ont une intensité multiplicative de la forme $\lambda_{hj}(t) = \alpha_{hj}(t) Y_h(t)$ (cf. Andersen *et al.* (1993)). Pour obtenir des résultats de convergence, nous ferons l'hypothèse dans la suite que la chaîne $X(\cdot)$ est définie sur un intervalle de temps $[0, \tau]$ tel que :

$$\forall t \in [0, \tau], \ \forall h \in [\![1, k]\!] \ : \quad \mathbb{P}(X(t^-) = h) > 0.$$

4.5 Résultats

4.5.1 Estimateur de Nelson-Aalen

Dans cette section, nous nous plaçons dans le cadre d'événements de types multiples présenté à la section 4.4.1 ou dans le cadre de processus de Markov non homogènes présenté à la section 4.4.2. Le but est ici de trouver

un estimateur de la fonction :

$$\begin{array}{rccc} A_h & : & [0,\tau] & \to & \mathbb{R} \\ & & t & \mapsto & \displaystyle\int_0^t \alpha_h(s)\,\mathrm{d}s, \end{array}$$

où τ est un réel positif fixé tel que $A_h(\tau) < +\infty$. Dans le cas de chaînes de Markov non homogènes, il peut s'agir d'estimer le risque cumulé pour la transition d'un état h à un autre état. On pourra aussi s'intéresser à la transition d'un état h à un état h' : les calculs sont les mêmes mais nous n'écrirons pas l'indice h' pour simplifier les notations. Pour trouver un estimateur de A_h, l'idée est d'utiliser la propriété de martingale de :

$$M_h(t) = N_h(t) - \int_0^t \alpha_h(s) Y_h(s)\,\mathrm{d}s.$$

Ainsi, si on écrit symboliquement $\mathrm{d}N_h(t) = \alpha_h(t)Y_h(t)\mathrm{d}t + \mathrm{d}M_h(t)$ et que l'on considère $M_h(t)$ comme étant un bruit de moyenne nulle, alors, un estimateur naturel de $A_h(t)$ est l'estimateur de Nelson-Aalen donné par la formule :

$$\widehat{A}_h(t) := \int_0^t Y_h(s)^{-1} J_h(s)\,\mathrm{d}N_{\bullet h}(s),$$

où on a posé par convention $0/0 = 0$ et où on a noté

$$J_h(s) := \mathbb{1}_{\{Y(s)>0\}} \quad \text{et} \quad N_{\bullet h}(s) := \sum_{i=1}^n N_{ih}(s).$$

Comme le montrent Andersen *et al.* (1993), cet estimateur est consistant et converge en loi après normalisation. Les conditions suffisantes pour la convergence des $\widehat{A}_h$ énoncées dans Andersen *et al.* (1993) sont très générales et ils démontrent que dans les cas qui nous intéressent ici (risques concurrents et chaînes de Markov) elles sont vérifiées. Notons alors $A = (A_1, \dots, A_k)^t$ et $\widehat{A} = (\widehat{A}_1, \dots, \widehat{A}_k)^t$ et énonçons leur résultat.

Théorème 4.6 (Andersen *et al.* (1993))

Il existe des martingales gaussiennes indépendantes $G_1, \dots, G_n$ continues telles que, lorsque $n \to +\infty$:

$$\sup_{t\in[0,\tau]} |\widehat{A}_h(t) - A(t)| \xrightarrow{\mathbb{P}} 0,$$
$$\sqrt{n}(\widehat{A}(\cdot) - A(\cdot)) \rightsquigarrow G(\cdot) = (G_1(\cdot), \dots, G_k(\cdot))^t.$$

Remarque 4.7. Le résultat de convergence en loi est donné au sens de Skorokhod dans le livre de Andersen *et al.* (1993), mais comme nous l'avons vu au chapitre 1, la convergence au sens de Hoffmann-Jørgensen est équivalente dans ce cas.

4.5.2 Estimateur de Kaplan-Meier

Dans le cadre de durées de vies censurées, il peut être intéressant d'estimer la fonction de survie de la variable de durée de vie. Une solution est apportée par Kaplan et Meier (cf. Kaplan & Meier (1958) ou Andersen *et al.* (1993) pour une étude détaillée). En gardant les mêmes notations que dans l'exemple introductif, l'estimateur de Kaplan-Meier est donné par la formule :

$$\widehat{S}(t) = \prod_{s \leqslant t} \left(1 - \frac{\Delta N(s)}{Y(s)}\right).$$

On peut remarquer que dans le cas de données non censurées, cet estimateur se ramène à $1 - F_n$ où F_n est la fonction de répartition empirique.

Dans leur livre, Andersen *et al.* (1993) montrent (cf. page 261) que dans les situations auxquelles nous nous intéressons, cet estimateur est uniformément consistant. Ils obtiennent aussi un résultat de convergence en distribution que nous ne mentionnerons pas ici car nous ne nous en servirons pas dans la suite.

Théorème 4.8

Soit τ un réel tel que $S(\tau) > 0$. Alors, dans les conditions décrites précédemment, lorsque $n \to +\infty$:

$$\sup_{t \in [0,\tau]} |\widehat{S}(t) - S(t)| \xrightarrow{\mathbb{P}} 0.$$

Deuxième partie

Résultats

CHAPITRE 5

Résultats préliminaires

Dans les chapitres suivants, nous appliquerons le théorème de Hjort *et al.* (2004) dans diverses situations. Pour vérifier la condition (A0*), nous utiliserons les résultats présentés ici.

5.1 Processus à valeurs réelles

Soit $(m_i(\cdot))_i$ une suite de processus aléatoires indépendants et identiquement distribués à valeurs dans $\mathbb{R}$, définis sur un intervalle de temps $[\tau_1, \tau]$. Soit $(\widehat{m}_i(\cdot))_i$ une autre suite de processus non nécessairement i.i.d. telle que, lorsque $n \to +\infty$:

$$\max_{1 \leqslant i \leqslant n} \|\widehat{m}_i - m_i\|_\infty \xrightarrow{\mathbb{P}} 0.$$

Supposons que les m_i sont centrés et que :

$$0 < \inf_{t \in [\tau_1, \tau]} \mathbb{E}\,|m_i(t)| \leqslant \sup_{t \in [\tau_1, \tau]} \mathbb{V}\mathrm{ar}\, m_i(t) < \infty.$$

Définissons encore :

$$\mathcal{F}_{n\omega} := \Big\{ \big(m_1(t)(\omega), \ldots m_n(t)(\omega)\big),\ t \in [\tau_1, \tau] \Big\},$$

et supposons que les $\mathcal{F}_{n\omega}$ sont euclidiennes et que les m_i admettent une même enveloppe F de carré intégrable.

Lemme 5.1. *Dans les conditions précédentes, lorsque $n \to +\infty$:*

$$\mathbb{P}\left(\inf_{t \in [\tau_1, \tau]} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\widehat{m}_i(t) > 0\}} > 0 \right) \longrightarrow 1, \tag{5.1}$$

$$\mathbb{P}\left(\inf_{t \in [\tau_1, \tau]} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\widehat{m}_i(t) < 0\}} > 0 \right) \longrightarrow 1. \tag{5.2}$$

En particulier, la probabilité pour qu'il existe $t \in [\tau_1, \tau]$ tel que 0 ne soit pas dans l'enveloppe convexe des $\widehat{m}_i(t)$ tend vers 0 lorsque n tend vers $+\infty$.

Avant de donner une démonstration de ce résultat, énonçons un lemme technique que nous utiliserons par la suite.

Lemme 5.2. *Soit z une variable aléatoire centrée admettant un moment d'ordre 2 non nul et soit $\eta = \mathbb{E}\,|z|/4$. Alors :*

$$\mathbb{P}(z \geqslant \eta) \geqslant \frac{\eta^2}{\mathbb{E}\,z^2}.$$

Démonstration du lemme 5.2. Commençons par remarquer que comme z est centré :

$$\eta = \frac{1}{4}\,\mathbb{E}\,|z| = \frac{1}{2}\,\mathbb{E}\,(z\mathbb{1}_{\{z\geqslant 0\}}).$$

Appliquons alors l'inégalité de Cauchy-Schwarz :

$$\begin{aligned}
\mathbb{P}(z \geqslant \eta) &\geqslant \frac{1}{\mathbb{E}(z^2)}\mathbb{E}^2(z\mathbb{1}_{\{z\geqslant\eta\}}) \\
&= \frac{1}{\mathbb{E}(z^2)}\big(\mathbb{E}(z\mathbb{1}_{z\geqslant 0\}}) - \big|\mathbb{E}(z\mathbb{1}_{\{z\geqslant 0\}}) - \mathbb{E}(z\mathbb{1}_{\{z\geqslant\eta\}})\big|\big)^2 \\
&\geqslant \frac{1}{\mathbb{E}(z^2)}(2\eta - \eta)^2.
\end{aligned}$$

□

Démonstration du lemme 5.1. Il est clair que l'on peut démontrer (5.2) à partir de (5.1) en considérant les variables $-\widehat{m}_i$ à la place des $\widehat{m}_i$. Nous nous contenterons donc de démontrer (5.1). Notons :

$$\bar{m} := \frac{1}{n}\sum_{i=1}^{n} m_i.$$

On a alors, par inégalité triangulaire :

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{\widehat{m}_i(t)>0\}} &\geqslant \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}_{\{m_i(t)>\eta/2\}} - \mathbb{1}_{\{\|\widehat{m}_i - m_i\|_\infty > \eta/2\}}\right) \\
&\geqslant \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}_{\{m_i(t)-\bar{m}(t)>\eta,\ |\bar{m}(t)|\leqslant\eta/2\}} - \mathbb{1}_{\{\|\widehat{m}_i - m_i\|_\infty > \eta/2\}}\right) \\
&\geqslant \frac{1}{n}\sum_{i=1}^{n}\left(\mathbb{1}_{\{m_i(t)-\bar{m}(t)>\eta\}} - \mathbb{1}_{\{|\bar{m}|>\eta/2\}} - \mathbb{1}_{\{\|\widehat{m}_i - m_i\|_\infty > \eta/2\}}\right) \quad (5.3)
\end{aligned}$$

On sait par hypothèse que la troisième indicatrice dans (5.3) tend vers 0 en probabilité lorsque $n \to +\infty$. D'après la loi des grands nombres uniforme, il en est de même pour la deuxième indicatrice. Il suffit donc de montrer que :

$$\mathbb{P}\left(\inf_{t\in[\tau_1,\tau]}\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{m_i(t)-\bar{m}(t)>\eta\}} > 0\right) \longrightarrow 1, \quad \text{lorsque } n \to +\infty.$$

Si les $\left(\mathbb{1}_{\{m_i>\eta/2\}}\right)_i$ formaient une classe euclidienne, on pourrait encore effectuer une inégalité triangulaire pour faire disparaître $\bar{m}$ puis utiliser la loi des grands nombres uniformes et enfin appliquer le lemme 5.2.

Bien que ce sera le cas dans nos applications, nous pouvons démontrer le résultat sans avoir à faire cette hypothèse. Pour ce faire, nous allons appliquer le lemme 5.2 à une probabilité empirique. Définissons alors pour tout $n \in \mathbb{N}^\star$ les fonctions z_n par :

$$\forall i \in [\![1, n]\!] \ : \quad z_n(i) := m_i(\cdot) - \bar{m}(\cdot).$$

Munissons $[\![1, n]\!]$ de la tribu engendrée par ses parties et de la probabilité uniforme $\widehat{\mathbb{P}}$. Alors :

$$\widehat{\mathbb{E}}(z_n) = \frac{1}{n}\sum_{i=1}^{n} z(i) = 0,$$
$$\widehat{\mathbb{E}}(z_n^2) = \frac{1}{n}\sum_{i=1}^{n} z(i)^2 = \frac{1}{n}\sum_{i=1}^{n} m_i(\cdot)^2 - \bar{m}(\cdot)^2.$$

Ainsi, les z_n sont centrés et d'après la loi des grands nombres uniforme [1], la probabilité pour que leurs moments d'ordre 2 soient non nuls tend vers 1 lorsque n tend vers $+\infty$. Posons alors :

$$\eta_n(\cdot) := \frac{1}{4}\,\widehat{\mathbb{E}}|z_n| \geqslant \frac{1}{4n}\sum_{i=1}^{n} |m_i(\cdot)| - \frac{1}{4}\,|\bar{m}(\cdot)|.$$

Donc, en appliquant encore une fois la loi des grands nombres uniforme, on en déduit que :

$$\mathbb{P}\left(\inf_{t\in[\tau_1,\tau]} \eta_n(t) > 0\right) \longrightarrow 1, \quad \text{lorsque } n \to +\infty,$$

car les m_i sont centrés. Finalement, en appliquant le lemme 5.2 on trouve :

$$\widehat{\mathbb{P}}(z_n \geqslant \eta) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{m_i(\cdot)-\bar{m}(\cdot)>\eta\}} \geqslant \frac{\eta_n(\cdot)^2}{\widehat{\mathbb{E}}\left(z_n(\cdot)^2\right)}.$$

Or, $\eta_n(t)$ est asymptotiquement de l'ordre de $\mathbb{E}|m_i(t)|$ et $\widehat{\mathbb{E}}\left(z_n^2\right)$ est asymptotiquement de l'ordre de $\mathbb{V}\mathrm{ar}\, m_i(t)$, d'où le résultat.

□

5.2 Processus à valeurs vectorielles

Supposons maintenant être en présence d'une suite de processus $(m_i(\cdot))_i$ i.i.d. à valeurs dans $\mathbb{R}^K$ et définis sur un intervalle de temps $[\tau_1, \tau]$. Soit

1. Voir remarque 2.31 qui suit la loi des grands nombres uniforme.

$(\widehat{m}_i(\cdot))_i$ une autre suite de processus non nécessairement i.i.d. telle que, lorsque $n \to +\infty$:

$$\max_{1 \leqslant i \leqslant n} \|\widehat{m}_i - m_i\|_\infty \xrightarrow{\mathbb{P}} 0,$$

où la norme infinie est définie par

$$\|z\|_\infty := \max_{1 \leqslant k \leqslant K} \sup_{t \in [\tau_1, \tau]} |z_k(t)|.$$

Notons $V(t)$, pour $t \in [\tau_1, \tau]$ fixé, la matrice de variance-covariance commune des $m_i(t)$. Supposons que les m_i sont centrés. Notons Θ la sphère unité de $\mathbb{R}^K$ et supposons que :

$$0 < \inf_{t \in [\tau_1, \tau]} \inf_{\theta \in \Theta} \mathbb{E}\left|\langle m_i(t), \theta \rangle\right| \leqslant \sup_{t \in [\tau_1, \tau]} \sup_{\theta \in \Theta} \left|\langle V(t)\theta, \theta \rangle\right| < +\infty.$$

Définissons encore pour tout $k \in [\![1, K]\!]$:

$$\mathcal{F}_{n\omega}^k := \left\{ \big(m_{1k}(t)(\omega), \dots m_{nk}(t)(\omega)\big),\ t \in [\tau_1, \tau] \right\},$$

et supposons que les $\mathcal{F}_{n\omega}^k$ sont euclidiennes et que les m_{ik} admettent une même enveloppe F de carré intégrable.

Lemme 5.3. *Alors, lorsque $n \to +\infty$:*

$$\mathbb{P}\left(\inf_{t \in [\tau_1, \tau]} \inf_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\langle \widehat{m}_i(t), \theta \rangle > 0\}} > 0 \right) \longrightarrow 1. \tag{5.4}$$

En particulier, la probabilité pour que tous les demi-espaces autour de 0 *soient occupés par les $\widehat{m}_i$ tend vers* 1.

Ce lemme montre que la probabilité pour qu'il existe $t \in [\tau_1, \tau]$ tel que 0 ne soit pas dans l'enveloppe convexe des $\widehat{m}_i(t)$ tend vers 0 lorsque n tend vers $+\infty$.

Démonstration. Nous allons appliquer les résultats de la section précédente aux processus réels définis par :

$$\forall (t, \theta) \in [\tau_1, \tau] \times \Theta \ : \quad \begin{cases} T_i(t, \theta) & := \langle m_i(t), \theta \rangle, \\ \widehat{T}_i(t, \theta) & := \langle \widehat{m}_i(t), \theta \rangle. \end{cases}$$

Par hypothèse, $\mathbb{V}\mathrm{ar}\, T_i(t, \theta) = \langle V(t)\theta, \theta \rangle$ est uniformément bornée dans un intervalle $[\delta, M]$ sur $[\tau_1, \tau] \times \Theta$. C'est aussi le cas de $\mathbb{E}\,|T_i(t, \theta)|$. De plus, d'après l'inégalité de Cauchy-Schwarz (cas d'égalité) :

$$\sup_{\theta \in \Theta} |\langle m_i(\cdot) - \widehat{m}_i(\cdot), \theta \rangle| = \|m_i(\cdot) - \widehat{m}_i(\cdot)\|_2,$$

où $\|\cdot\|_2$ désigne la norme euclidienne. On en déduit que, lorsque $n \to +\infty$:

$$\sup_{t \in [\tau_1, \tau]} \sup_{\theta \in \Theta} |T_i(t, \theta) - \widehat{T}_i(t, \theta)| \xrightarrow{\mathbb{P}} 0.$$

Remarquons aussi que les T_i admettent une même enveloppe de carré intégrable d'après l'inégalité de Cauchy-Schwarz et les hypothèses faites sur les m_i.

Pour conclure, il suffit de montrer que les classes ci-dessous sont euclidiennes :

$$\mathcal{G}_{n\omega} := \Big\{ \big(T_1(t,\theta)(\omega), \dots T_n(t,\theta)(\omega) \big),\ (t,\theta) \in [\tau_1, \tau] \times \Theta \Big\}.$$

Or, si on pose

$$\mathcal{H}^k_{n\omega} := \Big\{ \big(\alpha m_{1k}(t)(\omega), \dots \alpha m_{nk}(t)(\omega) \big),\ (\alpha, t) \in [-1,1] \times [\tau_1, \tau] \Big\},$$

alors, d'après la proposition 2.24, les $\mathcal{H}^k_{n\omega}$ sont euclidiennes. De plus :

$$\mathcal{G}_{n\omega} \subset \sum_{k=1}^{K} \mathcal{H}^k_{n\omega},$$

d'où le résultat par stabilité. □

5.3 Processus à valeurs vectorielles - première généralisation

Reprenons les notations introduites à la section 5.2, mais supposons cette fois que V admet des valeurs propres nulles. L'idée ici est de dire que si les valeurs propres non nulles de V se comportent bien (i.e. comme à la section précédente) et que les $\widehat{m}_i$ sont égaux aux m_i sur les sous-espaces correspondant aux valeurs propres nulles, alors, la condition (A0)* est encore satisfaite.

Pour $t \in [\tau_1, \tau]$ fixé, on a $\mathbb{R}^K = \mathrm{Im}V(t) \oplus \mathrm{Ker}V(t)$. On décompose alors les vecteurs m_i et $\widehat{m}_i$ comme suit :

$$\begin{aligned} m_i(t) &= a_i(t) + b_i(t), \\ \widehat{m}_i(t) &= \widehat{a}_i(t) + \widehat{b}_i(t), \end{aligned}$$

où les a_i et les $\widehat{a}_i$ sont des éléments de $\mathrm{Im}V$ et où les b_i et les $\widehat{b}_i$ sont des éléments de $\mathrm{Ker}V$. On supposera que :

$$\begin{aligned} \forall t \in [\tau_1, \tau] \ : \quad & \widehat{b}_i(t) = b_i(t), \\ & \max_{1 \leqslant i \leqslant n} \|\widehat{m}_i - m_i\|_\infty \xrightarrow{\mathbb{P}} 0. \end{aligned}$$

On supposera de plus que :

$$0 < \inf_{t \in [\tau_1,\tau]} \inf_{\theta \in \Theta \cap \mathrm{Im}V(t)} \mathbb{E} \left| \langle m_i(t), \theta \rangle \right| \leqslant \sup_{t \in [\tau_1,\tau]} \sup_{\theta \in \Theta \cap \mathrm{Im}V(t)} \left| \langle V(t)\theta, \theta \rangle \right| < +\infty.$$

Alors, lorsque $n \to +\infty$:

$$\mathbb{P}\left(\inf_{t\in[\tau_1,\tau]}\ \inf_{\theta\in\Theta\cap \mathrm{Im}V(t)}\ \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{\langle\widehat{m}_i(t),\theta\rangle>0\}} > 0\right) \longrightarrow 1. \tag{5.5}$$

En particulier, cela montre que :

$$\mathbb{P}\Big(\exists t\in[\tau_1,\tau]\ :\ 0\notin \mathrm{Conv}\big(\widehat{a}_1(t),\dots,\widehat{a}_n(t)\big)\Big) \longrightarrow 0, \quad \text{lorsque } n\to+\infty.$$

Remarquons que comme les $\widehat{b}_i(t)$ sont égaux aux $b_i(t)$, cela implique qu'ils sont nuls (les m_i étant centrés). On a donc aussi :

$$\mathbb{P}\Big(\exists t\in[\tau_1,\tau]\ :\ 0\notin \mathrm{Conv}\big(\widehat{b}_1(t),\dots,\widehat{b}_n(t)\big)\Big) = 0.$$

Ainsi, la probabilité pour qu'il existe $t\in[\tau_1,\tau]$ tel que 0 ne soit pas dans l'enveloppe convexe des $\widehat{m}_i(t)$ tend vers 0 lorsque n tend vers $+\infty$.

Démonstration. Pour démontrer (5.5), il suffit, comme à la section précédente, de se ramener à des processus réels. Pour ce faire, on étudie les processus T_i et $\widehat{T}_i$ définis par :

$$\forall(t,\theta)\in[\tau_1,\tau]\times\Theta\cap\mathrm{Im}V(t)\ :\quad \begin{cases} T_i(t,\theta) &:= \langle m_i(t),\theta\rangle,\\ \widehat{T}_i(t,\theta) &:= \langle \widehat{m}_i(t),\theta\rangle. \end{cases}$$

Il est clair que ces processus vérifient toutes les propriétés suffisantes présentées à la section 5.1 d'après les calculs effectués à la section précédente. Les calculs sont en effet les mêmes et les classes $\mathcal{G}_{n\omega}$ correspondantes sont plus petites que celles de la section précédente ce qui démontre qu'elle sont euclidiennes. □

5.4 Processus à valeurs vectorielles - deuxième généralisation

Plaçons-nous ici dans le même cadre qu'à la section précédente et reprenons les mêmes notations. Rappelons qu'on avait imposé $\widehat{m}_i$ d'être égaux aux m_i sur $\mathrm{Ker}V$ pour pouvoir conclure que :

$$\mathbb{P}\Big(\exists t\in[\tau_1,\tau]\ :\ 0\notin \mathrm{Conv}\big(\widehat{b}_1(t),\dots,\widehat{b}_n(t)\big)\Big) = 0. \tag{5.6}$$

Il est clair que (5.6) n'est pas vérifiée en général (prendre par exemple $b_i = 0$ et $\widehat{b}_i = 1/n$). Cependant, on peut modifier les $\widehat{m}_i$ à moindre coût pour obtenir (5.6) comme nous allons le voir ici.

5.4.1 Idée en dimension 1

Pour fixer les idées, nous allons dans un premier temps supposer que $\mathrm{Ker}V$ est de dimension 1 : les $\widehat{b}_i$ et les b_i seront donc des processus à valeurs vectorielles. Soit alors $(R_{i,n})_{1\leqslant i\leqslant n, n\in\mathbb{N}^\star}$ une suite i.i.d. de variables aléatoires de Rademacher symétriques [2]. Définissons :

$$\alpha_n := \max_{1\leqslant i\leqslant n} \|\widehat{b}_i\|_\infty.$$

Notons encore :

$$\forall t\in[\tau_1,\tau] \ : \quad \widetilde{b}_i(t) := \widehat{b}_i(t) + \alpha_n R_{i,n},$$

de sorte que le signe de $\widetilde{b}_i(t)$ ne dépende que de $R_{i,n}$. Alors, il est clair que :

$$0\in\mathrm{Conv}\big(R_{i,n},\ 1\leqslant i\leqslant n\big) \implies \forall t\in[\tau_1,\tau] \ : \ 0\in\mathrm{Conv}\big(\widetilde{b}_i(t),\ 1\leqslant i\leqslant n\big).$$

On en déduit que, lorsque $n\to+\infty$:

$$\mathbb{P}\Big(\exists t\in[\tau_1,\tau] \ : \ 0\notin\mathrm{Conv}\big(\widetilde{b}_1(t),\dots,\widetilde{b}_n(t)\big)\Big) \leqslant \left(\frac{1}{2}\right)^{n-1} \xrightarrow{\mathbb{P}} 0.$$

De plus, par construction, les $\widetilde{b}_i$ sont proches des $\widehat{b}_i$ au sens où :

$$\max_{1\leqslant i\leqslant n} \|\widetilde{b}_i - \widehat{b}_i\|_\infty = \max_{1\leqslant i\leqslant n} |\alpha_n| = \max_{1\leqslant i\leqslant n} \|\widehat{b}_i\|_\infty \xrightarrow{\mathbb{P}} 0, \quad \text{lorsque } n\to+\infty.$$

Ainsi, on pourrait être tenté de remplacer les $\widehat{b}_i$ par les $\widetilde{b}_i$ pour que la condition (5.6) soit vérifiée. De plus, cela n'aurait pas *a priori* de conséquence sur la convergence en loi du rapport de vraisemblance empirique puisque les éléments du noyau de V disparaissent dans la loi limite.

Cependant, il se peut qu'on ne connaisse pas $\mathrm{Ker}V$, et il devient alors impossible de faire la substitution proposée ici. Dans ce cas, nous proposons de perturber les $\widehat{m}_i$ non plus uniquement sur $\mathrm{Ker}V$ mais sur $\mathbb{R}^K$ tout entier.

5.4.2 Perturbation en dimension quelconque

Résultat

Dans la continuité de ce qui a été vu à la section précédente, considérerons une suite de variables aléatoires i.i.d. $(U_{i,n})_{1\leqslant i\leqslant n, n\in\mathbb{N}^\star}$ de loi uniforme sur la sphère unité de $\mathbb{R}^K$. Définissons alors :

$$\alpha_n := \max_{1\leqslant i\leqslant n} \sup_{t\in[\tau_1,\tau]} \|\widehat{m}_i - m_i\|_2,$$

où $\|\cdot\|_2$ désigne la norme euclidienne sur $\mathbb{R}^K$. Notons encore :

$$\forall t\in[\tau_1,\tau] \ : \quad \widetilde{m}_i(t) := \widehat{m}_i(t) + 2\alpha_n U_{i,n}.$$

2. Rappelons que cela signifie que $\mathbb{P}(R_{i,n}=1)=\mathbb{P}(R_{i,n}=-1)=1/2$.

Décomposons alors $\widetilde{m}_i$ sur $\mathrm{Im}\, V \oplus \mathrm{Ker}\, V$ et écrivons pour tout $t \in [\tau_1, \tau]$:

$$\widetilde{m}_i(t) := \widetilde{a}_i(t) + \widetilde{b}_i(t).$$

Nous allons montrer que si les $\widehat{m}_i$ vérifient les mêmes hypothèses que celles énoncées à la section 5.3 sauf éventuellement « $\widehat{b}_i = b_i$ », alors la condition (A0*) est vérifiée par les $\widetilde{m}_i$ et la déformation effectuée ne perturbe pas la loi limite du rapport de vraisemblance empirique.

Il est clair que si à $t \in [\tau_1, \tau]$ fixé $\mathrm{Ker}\, V$ est de dimension 0, alors il n'y a rien à démontrer. Nous supposerons donc que la dimension du noyau de V est strictement positive dans la suite.

Soit $\theta \in \Theta \cap \mathrm{Ker}\, V$ et soit p la dimension de $\mathrm{Ker}\, V$. Alors, il existe une constante c_p strictement positive telle que :

$$\mathbb{P}\left(\langle U_{i,n}, \theta\rangle > \frac{1}{2}\right) = c_p.$$

De plus, pour tout $t \in [\tau_1, \tau]$, d'après l'inégalité de Cauchy-Schwarz :

$$\left|\langle \widehat{b}_i(t), \theta\rangle\right| \leqslant \|\widehat{b}_i(t)\|_2 \leqslant \|\widehat{m}_i(t)\|_2.$$

On en déduit que, pour tout $t \in [\tau_1, \tau]$:

$$\langle U_{i,n}, \theta\rangle > \frac{1}{2} \implies \langle \widetilde{b}_i(t), \theta\rangle \geqslant \langle \widehat{b}_i(t), \theta\rangle + \alpha_n \geqslant -\alpha_n + \alpha_n = 0.$$

D'où :

$$\forall t \in [\tau_1, \tau],\ \forall \theta \in \Theta\ :\quad \mathbb{P}\big(\langle \widetilde{b}_i(t), \theta\rangle > 0\big) \geqslant c_p.$$

Ainsi :

$$\mathbb{P}\Big(\exists t \in [\tau_1, \tau]\ :\ 0 \notin \mathrm{Conv}\big(\widetilde{b}_1(t), \dots, \widetilde{b}_n(t)\big)\Big) \longrightarrow 0, \quad \text{lorsque } n \to +\infty.$$

Remarquons enfin que :

$$\max_{1\leqslant i\leqslant n} \|\widetilde{m}_i - m_i\|_\infty \xrightarrow{P} 0, \quad \text{lorsque } n \to +\infty.$$

Ainsi, les $\widetilde{m}_i$ vérifient la condition (A0*) du théorème 3.11 d'après les calculs menés à la section 5.3. De plus, ils vérifient les autres conditions dès que les $\widehat{m}_i$ les vérifient. En effet, rappelons que :

(A1*) $\sqrt{n} \cdot M_n(\cdot) \rightsquigarrow G(\cdot)$, lorsque $n \to +\infty$;

(A2*) $\sup_{t\in[\tau_1,\tau]} \|S_n(t) - V(t)\| \xrightarrow{\mathbb{P}^\star} 0$, lorsque $n \to +\infty$;

(A3*) $\max_{1\leqslant i\leqslant n} \sup_{t\in[\tau_1,\tau]} \|\widehat{m}(t)\| = o_p(n^{1/2})$.

On a alors :

$$\sqrt{n}\,\frac{1}{n}\sum_{i=1}^n \widetilde{m}_i = \sqrt{n}\,\frac{1}{n}\sum_{i=1}^n \widehat{m}_i + \left(\max_{1\leqslant i\leqslant n} \|\widehat{m}_i - m\|_\infty\right) \sqrt{n}\,\frac{1}{n}\sum_{i=1}^n U_{i,n},$$

où le deuxième terme tend en probabilité vers 0 d'après le théorème de la limite centrale et le lemme de Slutsky. Pour ce qui est de (A2*), on a :

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{m}_i^{\otimes 2} = \frac{1}{n}\sum_{i=1}^{n}\widehat{m}_i^{\otimes 2} + \left(\max_{1\leqslant i\leqslant n}\|\widehat{m}_i - m\|_\infty\right)^2 \frac{1}{n}\sum_{i=1}^{n} U_{i,n}U_{i,n}^t + \left(\max_{1\leqslant i\leqslant n}\|\widehat{m}_i - m\|_\infty\right)\frac{1}{n}\sum_{i=1}^{n}\left(U_{i,n}\widehat{m}_i^t + \widehat{m}_i U_{i,n}^t\right), \quad (5.7)$$

et on montre facilement que les deux derniers termes tendent vers 0 en probabilité (en norme infinie). Enfin, pour ce qui est de (A3*), on remarque que :

$$\|\widetilde{m}_i\|_\infty \leqslant 2\|\widehat{m}_i\|_\infty.$$

Conclusion

Supposons que les $\widehat{m}_i$ vérifient les hypothèses (A1*), (A2*) et (A3*) du théorème 3.11. Notons Θ la sphère unité de $\mathbb{R}^k$ et supposons de plus que :

$$\max_{1\leqslant i\leqslant n}\|\widehat{m}_i - m_i\|_\infty \xrightarrow{P} 0, \quad \text{lorsque } n \to +\infty,$$

et que

$$0 < \inf_{t\in[\tau_1,\tau]}\ \inf_{\theta\in\Theta\cap \mathrm{Im}\, V(t)} \mathbb{E}\left|\langle m_i(t),\theta\rangle\right| \leqslant \sup_{t\in[\tau_1,\tau]}\ \sup_{\theta\in\Theta\cap \mathrm{Im}\, V(t)} \left|\langle V(t)\theta,\theta\rangle\right| < +\infty.$$

Enfin, définissons :

$$\mathcal{F}_{n\omega} := \left\{\big(m_1(t)(\omega), \dots m_n(t)(\omega)\big),\ t\in[\tau_1,\tau]\right\},$$

et supposons que les $\mathcal{F}_{n\omega}$ sont euclidiennes et que les m_i admettent une même enveloppe F de carré intégrable. Alors, lorsque $n \to +\infty$:

$$-2\ln\widetilde{\mathrm{EL}_n}(\cdot) \rightsquigarrow G(\cdot)^t V(\cdot)^\dagger G(\cdot),$$

où on a posé pour tout $t \in [\tau_1,\tau]$

$$\widetilde{\mathrm{EL}_n}(t) := \max\left\{\prod_{i=1}^{n} np_i,\ \sum_{i=1}^{n} p_i\widetilde{m}_i(t) \text{ et } p\in\mathbb{S}_{n-1}\right\}.$$

CHAPITRE 6

Bande de confiance pour les fonctions de risque cumulé

6.1 Introduction

Nous avons vu au chapitre 4 que les fonctions de risque cumulé apparaissaient naturellement dans l'étude des processus de comptage à intensité multiplicative. De plus, elles peuvent dans certains cas contenir suffisamment d'informations pour retrouver des grandeurs d'intérêt. Nous avons vu par exemple à la section 4.2 que l'on pouvait retrouver la fonction de survie à partir de la fonction de risque cumulé dans un cadre de durées de vie censurées. Dans l'étude de processus de Markov, il est encore possible d'exprimer la matrice de transition à l'aide des fonctions de risque cumulé comme nous le verrons au chapitre 11.

Ainsi, il peut être intéressant de chercher des estimateurs pour ces fonctions de risque cumulé. L'estimateur le plus connu est sans doute celui de Nelson-Aalen, proposé par Aalen (1978) et généralisant celui proposé indépendamment par Nelson (1972) et Altshuler (1970). Il est étudié en détails dans Andersen *et al.* (1993) et nous avons donné certaines de ses propriétés au chapitre 4.

Dans ce chapitre, nous allons donner une méthode pour construire des bandes de confiance par vraisemblance empirique pour les fonctions de risque cumulé. Nous nous placerons alors soit dans le cadre d'événements multiples décrits à la section 4.4.1, soit dans le cadre de processus de Markov non homogènes décrits à la section 4.4.2. Nous appliquerons des techniques de vraisemblance empirique en pondérant l'estimateur de Nelson-Aalen et en utilisant ses propriétés asymptotiques.

Nous reviendrons au chapitre 10 sur les fonctions de risque cumulé et proposerons alors une deuxième approche pour la construction de bandes de confiance par vraisemblance empirique. Cette deuxième approche, plus directe, ne sera plus basée sur les propriétés asymptotiques de l'estimateur de Nelson-Aalen.

6.2 Modèle

Les notations utilisées dans ce chapitre sont celles introduites au chapitre 4, en particulier celles des sections 4.4.1 et 4.4.2. Nous étudions donc conjointement les problèmes avec événements de types multiples et les processus de Markov non homogènes. Nous donnons des conditions suffisantes pour lesquelles on peut construire des bandes de confiance pour les fonctions de risque cumulé par vraisemblance empirique.

Dans tout ce qui suit, nous supposerons l'existence d'un réel τ tel que $A_h(\tau) < \infty$. Nous noterons :

$$\begin{array}{rccc} \bar{Y}_h & : & [0,\tau] & \to & [0,1] \\ & & t & \mapsto & \dfrac{1}{n} Y_h(t). \end{array}$$

La fonction Y_h correspondra donc au nombre moyen d'individus à risque. Dans ces conditions, on peut montrer un résultat de type Glivenko-Cantelli :

$$\sup_{t\in[0,\tau]} |\bar{Y}_h(t) - \mathbb{E}\bar{Y}_h(t)| \xrightarrow{\mathbb{P}} 0, \quad \text{lorsque } n \to +\infty. \tag{6.1}$$

Ce résultat peut se démontrer simplement à l'aide du théorème 2.29 (la loi des grands nombres uniforme) pour les événements de types multiples. En ce qui concerne les processus de Markov non homogènes, on se reportera à Andersen *et al.* (1993) pages 296 et 318 pour une démonstration.

Dans tout ce qui suit, nous supposerons qu'il existe un réel τ_1 tel que, uniformément en n :

$$\inf_{t\in[\tau_1,\tau]} \mathbb{E}\bar{Y}_h(t) > 0. \tag{6.2}$$

Cette hypothèse utilisée conjointement avec (6.1) implique en particulier que :

$$\mathbb{P}\left(\inf_{t\in[\tau_1,\tau]} \bar{Y}_h(t)\,\mathbb{E}\bar{Y}_h(t) > 0\right) \underset{n\to+\infty}{\longrightarrow} 1. \tag{6.3}$$

Intéressons-nous maintenant à la fonction de risque cumulée. Que ce soit dans le cadre d'événements de types multiples présentés à la section 4.4.1 ou dans le cadre de processus de Markov non homogènes présentées à la section 4.4.2, on a vu que l'estimateur de Nelson-Aalen de la fonction de risque cumulé s'écrivait :

$$\widehat{A}_h(t) := \int_0^t Y_h(s)^{-1} J_h(s)\, \mathrm{d}N_{\bullet h}(s).$$

Pour pouvoir appliquer le théorème de Hjort *et al.* (2004), l'idée est de faire apparaître une moyenne empirique dans l'expression de $\widehat{A}_h(\cdot)$. À cet effet, posons :

$$\widehat{m}_{ih}(A_h, t) := \int_0^t \frac{1}{\bar{Y}_h(s)} \mathrm{d}N_{ih}(s) - A_h(t),$$

$$\mathrm{EL}_n(A_h, t) := \max\left\{ \prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i \widehat{m}_{ih}(A_h, t) = 0 \text{ et } p \in \mathbb{S}_{n-1} \right\},$$

où les $\widehat{m}_{ih}$ ont été définis de sorte que :

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{m}_{ih} = 0 \iff \widehat{A}_h = A_h.$$

Notons encore :

$$m_{ih}(A_h, t) := \int_0^t \frac{1}{\mathbb{E}\bar{Y}_h(s)} \mathrm{d}N_{ih}(s) - A_h(t).$$

On peut montrer que les $m_{ih}(A_h, \cdot)$ sont centrés. Pour s'en convaincre, on pourra se référer à la décomposition de Doob-Meyer et consulter Andersen *et al.* (1993). De plus, pour tout $t \in [\tau_1, \tau]$:

$$\begin{aligned}|m_{ih}(A_h, t) - \widehat{m}_{ih}(A_h, t)| &= \left| \int_0^t \frac{1}{\bar{Y}_h(s)\mathbb{E}\bar{Y}_h(s)} \left(\mathbb{E}\bar{Y}_h(s) - \bar{Y}_h(s)\right) \mathrm{d}N_{i,h}(s) \right| \\ &\leqslant \frac{1}{\inf\limits_{t\in[\tau_1,\tau]} \bar{Y}_h(t)\,\mathbb{E}\bar{Y}_h(t)} \cdot \left\|\mathbb{E}\bar{Y}_h - \bar{Y}_h\right\|_\infty \cdot B.\end{aligned}$$

On en déduit d'après (6.1) et (6.3) que, lorsque $n \to +\infty$:

$$\max_{1\leqslant i\leqslant n} \|m_{ih}(A_h, \cdot) - \widehat{m}_{ih}(A_h, \cdot)\|_\infty \xrightarrow{\mathbb{P}} 0. \tag{6.4}$$

6.3 Résultat

Théorème 6.1

Supposons que :

$$0 < \inf_{t\in[\tau_1,\tau]} \mathbb{E}\left|m_i(t)\right| \leqslant \sup_{t\in[\tau_1,\tau]} \left|V(t)\right| < +\infty.$$

Alors :

$$-2\ln \mathrm{EL}_n(A_h, \cdot) \rightsquigarrow G_h^2(\cdot)V_h^{-1}(\cdot), \quad \text{lorsque } n \to +\infty,$$

où G_h est le processus limite donné au théorème 4.6.

Pour pouvoir appliquer le théorème de Hjort *et al.* (2004), il nous faut vérifier les hypothèses (A0*) à (A3*). Rappelons que

$$M_n(A_h, t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{m}_{ih}(A_h, t),$$

$$S_n(A_h, t) = \frac{1}{n}\sum_{i=1}^{n}\widehat{m}_{ih}(A_h, t)^2,$$

et que :

(A0*) : $\mathbb{P}(\exists t \in [\tau_1, \tau] \,:\, \mathrm{EL}_n(A_h, t) = 0) \longrightarrow 0$, lorsque $n \to \infty$;

(A1*) : $\sqrt{n}\, M_n(A_h, \cdot) \rightsquigarrow G(\cdot)$, lorsque $n \to \infty$;

(A2*) : $\sup_{t\in[\tau_1,\tau]} |S_n(A_h, t) - V_h(t)| \xrightarrow{\mathbb{P}} 0$, lorsque $n \to \infty$;

(A3*) : $\max_{1\leqslant i\leqslant n} \|\widehat{m}_{ih}(A_h, \cdot)\|_\infty = o_p(n^{1/2})$.

Démonstration de (A1*). Le point (A1*) est en fait une conséquence directe de la convergence en loi de l'estimateur de Nelson-Aalen puisque :

$$\sqrt{n}\, M_n(A_h, \cdot) = \sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} \widehat{m}_i(A_h, \cdot) = \sqrt{n}\,(\widehat{A}_h(\cdot) - A_h(\cdot)).$$

□

Démonstration de (A2*). Réécrivons S_n sous la forme :

$$S_n(A_h, t) = \frac{1}{n} \sum_{i=1}^{n} \left(\Upsilon_{ih}(t)^2 + 2\Upsilon_{ih}(t) m_{ih}(A_h, t) + m_{ih}(A_h, t)^2\right),$$

où :

$$\Upsilon_{ih}(t) := \int_0^t \left(\frac{1}{\bar{Y}_h(s)} - \frac{1}{\mathbb{E}\bar{Y}_h(s)}\right) \mathrm{d}N_{ih}(s).$$

• **Première étape.** Montrons que : $\|\frac{1}{n}\sum_{i=1}^n m_{ih}(A_h, \cdot)^2 - V_h(\cdot)\|_\infty \xrightarrow{\mathbb{P}} 0$ lorsque $n \to \infty$.
Posons $f_{ih}(\omega, t) := m_{ih}(A_h, t)^2$ et $\mathcal{F}^h_{n\omega} := \left\{\left(f_{1,h}(\omega, t), \dots, f_{n,h}(\omega, t)\right), t \in [\tau_1, \tau]\right\}$. Écrivons encore :

$$g_{i,h}(\omega, t) = \int_0^t \frac{1}{\mathbb{E}\bar{Y}_h(s)} \mathrm{d}N_{ih}(s, \omega),$$

et

$$\mathcal{G}^h_{n\omega} = \Big\{(g_{1,h}(\omega, t), \dots, g_{n,h}(\omega, t)), t \in [\tau_1, \tau]\Big\}.$$

Alors, presque sûrement :

$$\begin{aligned} |g_{i,h}(\omega, t)| &= \left|\int_0^t \frac{1}{\mathbb{E}\bar{Y}_h(s)} \mathrm{d}N_{ih}(s, \omega)\right| \\ &\leqslant \frac{B}{\inf_{t\in[\tau_1,\tau]} \mathbb{E}\bar{Y}_h(t)}. \end{aligned}$$

Ainsi, la classe $\mathcal{G}^h_{n\omega}$ est euclidienne. En effet, les g_{ih} étant croissantes, la classe est de pseudo-dimension 1 (voir théorème 2.23). Comme elle est bornée, elle est alors euclidienne d'après le théorème 2.22. Les mêmes arguments s'appliquent encore à la classe $\mathcal{H}^h_n = \{(A_h(t), \dots, A_h(t)), t \in [\tau_1, \tau]\}$ puisque $A_h(\cdot)$ est borné sur $[\tau_1, \tau]$ par hypothèse. Remarquons maintenant

que $\mathcal{F}^h_{n\omega} \subset (\mathcal{G}^h_{n\omega} - \mathcal{H}^h_n)^2$. Comme les classes euclidiennes sont stables par addition et multiplication, $\mathcal{F}^h_{n\omega}$ est euclidienne. Donc, comme les $m_{ih}(t)$ sont centrés, on peut appliquer le théorème 2.29 et on en déduit que :

$$\left\| \frac{1}{n} \sum_{i=1}^{n} m_{ih}(A_h, \cdot)^2 - V_h(\cdot) \right\|_\infty = \sup_{t \in [\tau_1, \tau]} \left| \frac{1}{n} \sum_{i=1}^{n} f_{i,h}(t, \omega) - \mathbb{E}\, f_{i,h}(t, \cdot) \right| \xrightarrow[n \to +\infty]{\mathbb{P}} 0.$$

• **Deuxième étape.** Montrons que : $\| \frac{1}{n} \sum_{i=1}^{n} \Upsilon^2_{ih} \|_\infty \xrightarrow{\mathbb{P}} 0$, lorsque $n \to +\infty$.
Par définition, on a :

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \Upsilon^2_{ih} \right\|_\infty \leqslant \max_{1 \leqslant i \leqslant n} \| m_{ih}(A_h, \cdot) - \widehat{m}_{ih}(A_h, \cdot) \|^2_\infty \xrightarrow{\mathbb{P}} 0.$$

D'où le résultat en utilisant (6.4).
• **Troisième étape.** Montrons que : $\| \frac{2}{n} \sum_{i=1}^{n} \Upsilon_{i,h}(\cdot) m_{ih}(A_h, \cdot) \|_\infty \xrightarrow{\mathbb{P}} 0$, lorsque $n \to +\infty$.
On montre comme à l'étape précédente que $\| \frac{1}{n} \sum_{i=1}^{n} \Upsilon_{ih} \|_\infty \xrightarrow{\mathbb{P}} 0$. On conclut alors en remarquant que :

$$\max_{1 \leqslant i \leqslant n} \| m_{ih}(A_h, \cdot) \|_\infty \leqslant \frac{B}{\inf\limits_{t \in [\tau_1, \tau]} \mathbb{E} \bar{Y}_h(t)} + A_h(\tau).$$

□

Démonstration de (A3*). Remarquons que :

$$\begin{aligned} \| \widehat{m}_{ih}(A_h, \cdot) \|_\infty &= \left\| \int_0^\cdot \frac{1}{Y_h(s)} \mathrm{d}N_{ih}(s) - A_h(\cdot) \right\|_\infty \\ &\leqslant \frac{B}{\inf\limits_{t \in [\tau_1, \tau]} \bar{Y}_h(t)} + A_h(\tau). \end{aligned}$$

Ainsi, la condition (A3*) est satisfaite puisque :

$$\mathbb{P}\left(\inf_{t \in [\tau_1, \tau]} \bar{Y}_h(t) > 0 \right) \to 1, \quad \text{lorsque } n \to +\infty.$$

□

Démonstration de (A0*). Pour démontrer (A0*) on applique le résultat présenté à la section 5.3. On sait que les $m_{ih}(\widehat{A}_h, \cdot)$ sont centrés. De plus, ils forment une classe euclidienne bornée d'après la première étape de la démonstration de (A2*). D'où le résultat d'après (6.4) et les hypothèses faites dans l'énoncé du théorème. □

CHAPITRE 7

Estimation de la fonction moyenne

7.1 Introduction

Nous disposons d'un jeu de données collectées dans une unité de soin intensif d'un hôpital français sur une période de dix ans. Ce jeu de données a déjà été étudié par Dauxois & Sencey (2009) et nous utiliserons dans la suite un de leurs résultats. Pour chaque patient, on connaît sa date d'entrée à l'hôpital ainsi que sa date de sortie (ou, si il est décédé la date de son décès). On sait aussi si il a contracté des infections nosocomiales, et, le cas échéant, de quelles infections il s'agit.

Nous sommes donc en présence d'événements récurrents (une même infection nosocomiale peut être contractée plusieurs fois) et de risques concurrents (un patient peut contracter plusieurs types d'infections). De plus, lorsqu'un patient quitte le service, cela peut être interprété comme un événement terminal.

Nous nous sommes donc attaché à l'étude d'événements récurrents en présence de risques concurrents, censure à droite et événement terminal (bien qu'il n'y ait pas de censure dans le jeu de données). Dans ce cadre, nous avons souhaité construire des bandes de confiance pour les fonctions moyennes qui représentent le nombre moyen d'événements d'un certain type au cours du temps.

Les données concernent un total de 7867 patients. Nous nous sommes concentré sur les trois infections les plus fréquentes, à savoir les pneumonies, les septicémies et les infections urinaires. À titre d'exemple, 373 patients ont contracté un total de 463 pneumonies.

Les modèles d'événements récurrents sont utiles dans divers domaines : en sciences humaines pour l'étude de périodes de chômage, en fiabilité pour l'étude de pannes récurrentes ou encore en biostatistiques comme c'est notre cas ici.

Beaucoup d'outils statistiques ont été mis à contributions ces dernières décennies pour l'étude des événements récurrents. Citons entre autres les modèles markoviens, les martingales ou encore les processus de Poisson.

Andersen *et al.* (1993) dans leur livre font une synthèse d'une partie de la recherche en utilisant une approche martingale. Lawless (1995) et Lawless & Nadeau (1995) se sont intéressés à la fonction moyenne. Wang & Wells (1998), Lin *et al.* (1999), Lin & Ying (2001), Cai & Schaubel (2004a) et plus récemment Du (2009) se sont intéressés à la distribution des temps intersauts. Plus récemment, Cook & Lawless (2007) ont écrit un livre sur les événements récurrents.

Il se peut, dans certaines situations, qu'un événement terminal interrompe l'observation du processus de comptage. C'est par exemple le cas dans notre étude lorsqu'un patient décède. Il apparaît alors naturel de ne pas faire d'hypothèse d'indépendance entre l'événement terminal et le processus de comptage. Il se peut aussi que l'observation du processus de comptage soit interrompue par une censure à droite indépendante de l'état du patient. C'est le cas par exemple lorsqu'un patient sort de l'étude ou simplement lorsque l'étude arrive à son terme. C'est pourquoi dans notre modèle nous inclurons une censure aléatoire à droite indépendante ainsi qu'un événement terminal (qui lui ne sera pas supposé indépendant). Parmi d'autres, citons les papiers de Cook *et al.* (2009) et de Cai & Schaubel (2004a) qui traitent d'événements récurrents avec censure (et en présence d'un événement terminal pour le premier des deux articles).

Enfin, il se peut que plusieurs types d'événements, que l'on qualifiera alors de concurrents, entrent en jeu. C'est le cas dans notre étude puisque plusieurs types d'infections peuvent toucher un même patient. C'est aussi le cas en fiabilité si on étudie par exemple les pannes sur une voiture : plusieurs mécanismes différents peuvent être mis en cause. De même, dans le cadre de l'étude des périodes de chômage, plusieurs facteurs peuvent être à l'origine de la perte d'un travail. Il est donc naturel d'étudier les événements récurrents avec risques concurrents comme l'ont fait Dauxois & Sencey (2009) dans leur article.

7.2 Notations

Par souci de simplicité, nous supposerons que seulement deux types d'événements sont observés, la généralisation à un nombre quelconque étant évidente. Pour $j = 1, 2$, nous noterons $N_j^\star(t)$ le nombre total d'événements de type j ayant eu lieu jusqu'à l'instant t. Nous supposerons que les processus de comptage $N_j^\star$ sont presque sûrement bornés par une constante B. Nous supposerons de plus la présence d'un événement terminal D *a priori* dépendant des $N_j^\star$, et après lequel les processus de comptage sont constants. Enfin, nous supposerons que les observations des processus sont soumises à une censure aléatoire à droite C indépendante des $N_j^\star$ et de D.

Nous noterons $X = D \wedge C$ et $\delta = \mathbb{1}_{\{X \leqslant C\}}$. Cette variable aléatoire indique si c'est l'événement terminal qui a stoppé l'observation des processus de comptage ou la censure. Nous noterons encore $Y(t) = \mathbb{1}_{\{X \geqslant t\}}$ et $N_j(t) = N_j^\star(t \wedge C)$ pour $t \geqslant 0$ et $j = 1, 2$. Alors, les observations sont des répliques i.i.d. $(N_{i,j}(t), X_i, \delta_i)$ de $(N_j(t), X, \delta)$, pour $t \in [0, \tau]$ où τ est une constante

fixée vérifiant $\mathbb{P}(C > \tau)\mathbb{P}(X > \tau) > 0$.

Pour construire des bandes de confiance simultanées pour les fonctions moyennes $t \mapsto \mu_j(t) := \mathbb{E}(N_j^\star(t))$, nous utilisons les estimateurs $\widehat{\mu}_j(t)$ proposés par Dauxois & Sencey (2009) :

$$\widehat{\mu}_j(t) = \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{\widehat{S}(u^-)}{\bar{Y}(u)}\mathrm{d}N_{i,j}(u),$$

où $\widehat{S}$ est l'estimateur de Kaplan-Meier de S, la fonction de survie de D (on se reportera à la section 4.5.2 pour plus de détails). Comme μ s'écrit comme une somme, l'idée est de la pondérer pour utiliser un résultat de vraisemblance empirique. Nous écrirons donc pour tout t dans $[0, \tau]$:

$$\begin{aligned}
\widehat{m}_{i,j}(N_{i,j}, \mu_j, t) &:= \int_0^t \frac{\widehat{S}(u^-)}{\bar{Y}(u)}\mathrm{d}N_{i,j}(u) - \mu_j(t),\\
m_{i,j}(N_{i,j}, \mu_j, t) &:= \int_0^t \frac{S(u^-)}{\mathbb{E}Y(u)}\mathrm{d}N_{i,j}(u) - \mu_j(t),\\
\mathrm{EL}_n(\mu_j, t) &:= \max\left\{\prod_{i=1}^{n} np_i,\ \sum_{i=1}^{n} p_i\widehat{m}_{i,j}(N_{i,j}, \mu_j, t) = 0 \text{ et } p \in \mathbb{S}_{n-1}\right\}.
\end{aligned}$$

Enfin nous définirons :

$$\begin{aligned}
M_{n,j}(\mu_j, t) &:= \frac{1}{n}\sum_{i=1}^{n} m_{i,j}(N_{i,j}, \mu_j, t),\\
S_{n,j}(\mu_j, t) &:= \frac{1}{n}\sum_{i=1}^{n} m_{i,j}(N_{i,j}, \mu_j, t)^2.
\end{aligned}$$

Lorsque cela ne prêtera pas à confusion, nous noterons $m_{i,j}(\cdot)$ et $\widehat{m}_{i,j}(\cdot)$ les processus $m_{i,j}(N_{i,j}, \mu_j, \cdot)$ et $m_{i,j}(N_{i,j}, \mu_j, \cdot)$.

7.3 Résultats

7.3.1 Convergence des marginales

Pour démontrer le théorème principal de cette section nous utiliserons le résultat ci-dessous qui est une conséquence du Théorème 2 de Dauxois & Sencey (2009) :

Théorème 7.1

Soit $j \in [\![1, 2]\!]$, alors, lorsque $n \to \infty$:

$$\sqrt{n}\,(\widehat{\mu}_j(\cdot) - \mu_j(\cdot)) \rightsquigarrow G_j(\cdot),$$

où $G_j(\cdot)$ est un processus gaussien centré.

Nous pouvons maintenant énoncer le résultat principal de ce chapitre : la convergence en distribution du rapport de vraisemblance empirique sous la condition de variance habituelle.

Théorème 7.2

Soit $V_j(t) := \mathbb{V}\mathrm{ar}\, m_{i,j}(t)$. Supposons qu'il existe un réel τ_1 tel que :

$$0 < \inf_{t\in[\tau_1,\tau]} \mathbb{E}\,|m_{i,j}(t)| \leqslant \sup_{t\in[\tau_1,\tau]} V_j(t) < +\infty.$$

Alors :

$$-2\ln \mathrm{EL}_n(\mu_j,\cdot) \rightsquigarrow G_j^2(\cdot)V_j^{-1}(\cdot), \quad \text{lorsque } n\to+\infty.$$

Nous allons appliquer le théorème de Hjort *et al.* (2004) pour démontrer ce résultat. Rappelons qu'il est suffisant de montrer que les conditions (A0*) à (A3*) sont vérifiées, où :

(A0)* : $\mathbb{P}(\exists t\in[\tau_1,\tau] \,:\, \mathrm{EL}_n(\mu_j(t),t)=0) \longrightarrow 0$, lorsque $n\to\infty$;

(A1)* : $\sqrt{n}\,M_{n,j}(\mu_j,\cdot) \rightsquigarrow G_j(\cdot)$, lorsque $n\to\infty$;

(A2)* : $\sup_{t\in[\tau_1,\tau]} |S_{n,j}(\mu_j,t) - V_j(t)| \xrightarrow{\mathbb{P}} 0$, lorsque $n\to\infty$;

(A3)* : $\max_{1\leqslant i\leqslant n} |m_{i,j}(N_i,\mu_j,\cdot)|_\infty = o_p(n^{1/2})$,

où $\|\cdot\|_\infty$ désigne la norme infinie sur $\mathcal{D}([\tau_1,\tau])$.

Démonstration de (A1*). D'après le théorème 7.1, on a :

$$\sqrt{n}\,M_{n,j}(\mu_j,\cdot) = \sqrt{n}\left(\widehat{\mu}_j(\cdot) - \mu(\cdot)\right) \rightsquigarrow G_j(\cdot), \quad \text{lorsque } n\to+\infty.$$

□

Démonstration de (A2*). On peut réécrire $S_{n,j}$ comme suit :

$$S_{n,j}(\mu_j,t) = \frac{1}{n}\sum_{i=1}^{n}\left(A_{i,j}(t)^2 + 2A_{i,j}(t)m_{i,j}(t) + m_{i,j}(t)^2\right),$$

où

$$A_{i,j}(t) := \int_0^t \left(\frac{\widehat{S}(u^-)}{\bar{Y}(u)} - \frac{S(u^-)}{\mathbb{E}Y(u)}\right)\mathrm{d}N_{i,j}(u).$$

• **Première étape** Montrons que : $\|\frac{1}{n}\sum_{i=1}^n m_{i,j}^2 - V_j\|_\infty \xrightarrow{\mathbb{P}} 0$ lorsque $n\to\infty$. Écrivons $f_{i,j}(\omega,t) := m_{i,j}(t)(\omega)^2$ et :

$$\mathcal{F}_{n\omega}^j := \Big\{\big(f_{1,j}(\omega,t),\ldots,f_{n,j}(\omega,t)\big),\ t\in[\tau_1,\tau]\Big\}.$$

Posons encore :

$$g_{i,j}(\omega, t) := \int_0^t \frac{S(u^-)}{\mathbb{E}\, Y_1(u)} \mathrm{d}N_{i,j}(u, \omega),$$

et

$$\mathcal{G}^j_{n\omega} := \Big\{ \big(g_{1,j}(\omega, t), \dots, g_{n,j}(\omega, t)\big), t \in [\tau_1, \tau] \Big\}.$$

Alors, presque sûrement :

$$\begin{aligned} |g_{i,j}(\omega, t)| &= \left| \int_0^t \frac{1}{\mathbb{P}(C \geqslant u)} \mathrm{d}N_{i,j}(u, \omega) \right| \\ &\leqslant \frac{1}{\mathbb{P}(C \geqslant \tau)} \int_0^\tau \mathrm{d}N_{i,j}(u, \omega) \\ &\leqslant \frac{B}{\mathbb{P}(C \geqslant \tau)}. \end{aligned}$$

Ainsi, la classe $\mathcal{G}^j_{n\omega}$ est euclidienne. En effet, les $g_{i,j}$ étant croissantes, la classe est de pseudo-dimension 1 (voir théorème 2.23). Comme elle est bornée, elle est alors euclidienne. Les mêmes arguments s'appliquent encore à la classe $\mathcal{H}^j_n = \{(\mu_j(t), \dots, \mu_j(t)), t \in [\tau_1, \tau]\}$ puisque $|\mu_j(t)| \leqslant B$. Remarquons maintenant que $\mathcal{F}^j_{n\omega} \subset (\mathcal{G}^j_{n\omega} - \mathcal{H}^j_n)^2$. Comme les classes euclidiennes admettant une enveloppe sont stables par addition et multiplication, $\mathcal{F}^j_{n\omega}$ est euclidienne. Donc, comme les $m_{i,j}(t)$ sont centrés, on en déduit d'après la loi des grands nombres uniforme que :

$$\sup_{t \in [\tau_1, \tau]} \left| \frac{1}{n} \sum_{i=1}^n f_{i,j}(t, \omega) - \mathbb{E}\, f_{i,j}(t, \cdot) \right| \xrightarrow{\mathbb{P}} 0, \quad \text{lorsque } n \to +\infty.$$

• **Deuxième étape.** Montrons que : $\| \frac{1}{n} \sum_{i=1}^n A^2_{i,j} \|_\infty \xrightarrow{\mathbb{P}} 0$, lorsque $n \to +\infty$. On peut réécrire $A_{i,j}(t)$ comme :

$$\int_0^t \frac{1}{\bar{Y}(u)\mathbb{E}\, Y(u)} \Big(\widehat{S}(u^-)(\mathbb{E}Y(u) - \bar{Y}(u)) - \bar{Y}(u)(S(u^-) - \widehat{S}(u^-)) \Big) \mathrm{d}N_{i,j}(u).$$

Or, par hypothèse sur τ, on a :

$$\inf_{t \in [\tau_1, \tau]} \mathbb{E}Y(t) = \mathbb{P}(D \geqslant \tau)\mathbb{P}(C \geqslant \tau) > 0.$$

De plus :

$$\bar{Y}(u) = \frac{1}{n} \sum_{i=1}^n I(X_i \geqslant u) \geqslant \bar{Y}(\tau),$$

où le dernier terme converge presque sûrement vers $\mathbb{P}(D \geqslant \tau)\mathbb{P}(C \geqslant \tau) > 0$. Mais, presque sûrement :

$$\left\| \frac{1}{n} \sum_{i=1}^n A^2_{i,j} \right\|_\infty \leqslant \left(\frac{\|\mathbb{E}Y(\cdot) - \bar{Y}(\cdot)\|_\infty + \|S(\cdot) - \widehat{S}(\cdot)\|_\infty}{\bar{Y}(\tau)\mathbb{E}Y(\tau)} \right)^2 B^2,$$

car les $N_{i,j}$ sont bornés par B. Cette inégalité conjointement à la convergence de $\|\mathbb{E}Y(\cdot) - \bar{Y}(\cdot)\|_\infty$ ainsi que celle de $\|S(\cdot) - \widehat{S}(\cdot)\|_\infty$ (cf. section 4.5.2) démontre la convergence recherchée pour la deuxième étape.

• **Troisième étape.** Montrons que : $\|\frac{2}{n}\sum_{i=1}^n A_{i,j} m_{i,j}\|_\infty \xrightarrow{\mathbb{P}} 0$, lorsque $n \to +\infty$.

Les mêmes arguments que ceux utilisés à la deuxième étape permettent de montrer que $\|\frac{1}{n}\sum_{i=1}^n A_{i,j}\|_\infty \xrightarrow{\mathbb{P}} 0$. On conclut en remarquant que :

$$\max_{1\leqslant i\leqslant n} \|m_{i,j}\|_\infty \leqslant \max_{1\leqslant i\leqslant n} \left(\int_0^\tau \frac{S(u^-)}{\mathbb{E}Y(u)} \mathrm{d}N_{i,j}(u) + \mu_j(\tau) \right) \leqslant \frac{B}{\mathbb{E}Y(\tau)} + B.$$

□

Démonstration de (A3*). Remarquons d'abord que :

$$\begin{aligned} \|m_{i,j}(N_i, \mu_j, \cdot)\|_\infty &= \sup_{t\in[\tau_1,\tau]} \left| \int_0^t \frac{\widehat{S}(u^-)}{\bar{Y}(u)} \mathrm{d}N_{i,j}(u) - \mu_j(t) \right| \\ &\leqslant \int_0^\tau \frac{1}{\bar{Y}(u)} \mathrm{d}N_{i,j}(u) + \mu_j(\tau) \\ &\leqslant \frac{B}{\bar{Y}(\tau)} + B. \end{aligned}$$

Mais, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(n^{-1/2} \cdot \frac{B}{\bar{Y}(\tau)} > \varepsilon \right) = \mathbb{P}\left(\sqrt{n}(\bar{Y}(\tau) - \mathbb{E}Y(\tau)) < \frac{B}{\varepsilon} - \sqrt{n} \cdot \mathbb{E}Y(\tau) \right) \underset{n\to\infty}{\longrightarrow} 0.$$

Ainsi, $\|m_{i,j}(N_i, \mu_j, \cdot)\|_\infty = o_p(n^{1/2})$. □

Démonstration de (A0*). Il est suffisant de montrer que les conditions présentées à la section 5.1 sont satisfaites. Or, pour tout t dans $[\tau_1, \tau]$, presque sûrement :

$$|\widehat{m}_{i,j}(t) - m_{i,j}(t)| \leqslant B \left| \frac{\widehat{S}(u^-)}{\bar{Y}(u)} - \frac{S(u^-)}{\mathbb{E}Y_1(u)} \right|.$$

Mais, on sait d'après le théorème de Glivenko-Cantelli, que $\|\mathbb{E}Y(\cdot) - \bar{Y}(\cdot)\|_\infty$ tend en probabilité vers 0 lorsque n tend vers $+\infty$. De plus, l'estimateur de Kaplan-Meier étant uniformément consistant, $\|S(\cdot) - \widehat{S}(\cdot)\|_\infty$ converge aussi en probabilité vers 0 lorsque n tend vers $+\infty$ (voir section 4.5.2). On en déduit que :

$$\max_{1\leqslant i\leqslant n} \|\widehat{m}_{i,j} - m_{i,j}\|_\infty \xrightarrow{\mathbb{P}} 0.$$

Rappelons que les $m_{i,j}$ sont centrés et qu'ils sont euclidiens et bornés d'après ce qui a été vu à la première étape de la démonstration de (A2*). Ainsi, toutes les conditions sont satisfaites et (A0*) est démontrée. □

7.3.2 Convergence jointe

Dans cette section nous énonçons un corollaire du théorème 7.2 dans lequel nous considérons 2 événements. Nous démontrons alors une convergence du rapport de vraisemblance empirique pour les deux fonctions μ_1 et μ_2. Définissons :

$$\mathrm{EL}_n(\mu, t) = \max\left\{\prod_{i=1}^{n} np_i,\ \sum_{i=1}^{n} p_i m_i(N_i, \mu, t) = 0 \text{ et } p \in \mathbb{S}_{n-1}\right\},$$

où

$$m_i(N_i, \mu, t) = (m_{i,j}(N_{i,j}, \mu_j, t))_{j=1,2}.$$

Comme précédemment, lorsqu'il n'y aura pas de confusion possible, nous écrirons :

$$m_i(t) = (m_{i,j}(N_{i,j}, \mu_j, t))_{j=1,2} = \left(\int_0^t \frac{S(u^-)}{\mathbb{E}Y(u)} \mathrm{d}N_{i,j}(u) - \mu_j(t)\right)_{j=1,2},$$
$$\widehat{m}_i(t) = (m_{i,j}(N_{i,j}, \mu_j, t))_{j=1,2} = \left(\int_0^t \frac{\widehat{S}(u^-)}{\bar{Y}(u)} \mathrm{d}N_{i,j}(u) - \mu_j(t)\right)_{j=1,2}.$$

Avant d'énoncer le corollaire du théorème 7.2, rappelons le résultat suivant démontré par Dauxois & Sencey (2009) :

Théorème 7.3

Lorsque $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \widehat{\mu}_1 - \mu_1 \\ \widehat{\mu}_2 - \mu_2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} =: G.$$

Corollaire 7.4. *Soit $V(\cdot)$ la matrice de variance-covariance commune des $m_i(\cdot)$ et soit Θ la sphère unité de $\mathbb{R}^2$. Supposons qu'il existe un réel τ_1 tel que :*

$$0 < \inf_{t\in[\tau_1,\tau]} \inf_{\theta\in\Theta} \mathbb{E}\left|\langle m_i(t), \theta\rangle\right| \leqslant \sup_{t\in[\tau_1,\tau]} \sup_{\theta\in\Theta} \left|\langle V(t)\theta, \theta\rangle\right| < +\infty.$$

Alors :

$$-2\ln \mathrm{EL}_n(\mu, \cdot) \rightsquigarrow G(\cdot)^t V^{-1}(\cdot) G(\cdot), \quad \text{lorsque } n \to +\infty.$$

Posons :

$$M_n(\mu, t) := \frac{1}{n}\sum_{i=1}^{n} m_i(N_i, \mu, t) \quad \text{et} \quad S_n(\mu, t) := \frac{1}{n}\sum_{i=1}^{n} m_i(N_i, \mu, t)^{\otimes 2}.$$

Il est clair que les hypothèses (A1*) et (A3*) sont vérifiées : il suffit de l'écrire et d'utiliser le théorème 7.3. L'hypothèse (A0*) est quant à elle vérifiée car toutes les hypothèses de la section 5.2 sont réunies.

Démonstration de (A2). Remarquons que :

$$S_n(N, \mu, t)) = \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} m_{i,1}^2(t) & m_{i,1}(t)m_{i,2}(t) \\ m_{i,2}(t)m_{i,1}(t) & m_{i,2}(t)^2 \end{pmatrix}.$$

On a déjà montré la convergence des termes diagonaux de cette matrice dans la preuve du théorème 7.2. En utilisant des arguments similaire, on peut montrer que, lorsque $n \to +\infty$:

$$\sup_{t\in[\tau_1,\tau]} \left| \frac{1}{n} \sum_{i=1}^{n} m_{i,1}(t)m_{i,2}(t) - \mathbb{C}\mathrm{ov}(m_1(t), m_2(t)) \right| \xrightarrow{\mathbb{P}} 0.$$

□

7.4 Simulations

7.4.1 Modélisation

Dans cette partie nous considérerons des événements récurrents d'un seul type. On note N le processus de comptage associé. Pour la simulation de N, nous utilisons le même cadre que celui présenté dans Dauxois & Sencey (2009) (on se reportera au chapitre 12 pour les propriétés de la loi bivariée exponentielle).

Soient $T_1 < T_2 < \ldots$ les instants de sauts de N. Le vecteur (T_1, D) sera tiré suivant une loi $\mathrm{BVE}(\lambda_1, \lambda_D, \lambda_{12})$. Les autres instants de sauts de N seront simulés de sorte que $(T_i, T_{i+1} - T_i)$ suive une loi $\mathrm{BVE}(\lambda_1, \lambda_D, \lambda_{12})$. Le choix des paramètres a été fait afin que $\mathrm{corr}(T_i, T_{i+1} - T_i) = \mathrm{corr}(D, T_1) = \rho$. On en déduit alors que :

$$\lambda_1 = \lambda_2 = \frac{1-\rho}{1+\rho} \qquad \text{et} \qquad \lambda_{12} = \frac{2\rho}{1+\rho}.$$

Les valeurs de ρ que l'on a retenues pour les simulations sont $\rho = 0$ et $\rho = 0.25$. La censure C a été tirée suivant une loi uniforme sur $[0, c]$, c étant une constante déterminée de manière heuristique (de même que λ_D) de sorte que l'on observe 1 ou 2 événements par individu en moyenne. Enfin, trois tailles d'échantillon ont été choisies : $n = 30, 100, 500$.

Pour approcher la loi limite du rapport de vraisemblance empirique nous avons utilisé une technique de bootstrap que nous discutons à la section 7.4.3.

Nous avons alors construit des régions de confiance d'ordre $1 - \alpha$ pour $\alpha = 1\%, 5\%$ et 10% et avons testé l'appartenance du vrai paramètre à ces régions. Après 10000 itérations, nous avons compté le pourcentage d'erreurs de première espèce.

7.4.2 Résultats des simulations

Nous avons résumé dans le tableau ci-dessous les résultats des simulations présentées à la section précédente.

		Nombre moyen d'événements : 1				Nombre moyen d'événements : 2			
		$\rho = 0$		$\rho = 0.25$		$\rho = 0$		$\rho = 0.25$	
		0%	30%	0%	30%	0%	30%	0%	30%
$n = 30$	α_1	1.40	1.16	2.12	1.91	0.96	1.09	1.48	1.00
	α_5	6.13	5.78	6.92	6.18	4.99	5.18	5.95	5.38
	α_{10}	11.46	11.59	12.18	11.50	10.28	10.92	11.12	10.66
$n = 100$	α_1	1.23	0.93	1.12	1.00	0.94	0.90	1.00	1.01
	α_5	5.31	5.05	5.31	5.34	4.84	4.99	4.70	4.76
	α_{10}	10.66	10.47	10.06	10.70	9.98	10.37	9.95	10.55
$n = 500$	α_1	0.99	0.88	1.11	1.10	0.94	1.09	1.14	0.91
	α_5	5.31	5.18	5.67	5.55	5.04	5.33	4.89	5.09
	α_{10}	10.60	10.45	11.21	11.07	10.01	10.77	9.95	10.06

On peut remarquer que les valeurs obtenues sont en moyenne supérieures aux valeurs théoriques, ce qui est sans doute dû (au moins en partie) à l'utilisation du bootstrap pour estimer la loi limite comme nous l'expliquons à la section 7.4.3.

De plus, à de très rares exceptions près, le niveau empirique se rapproche du niveau théorique lorsqu'on passe de 1 à 2 pour le nombre moyen d'événements.

Cependant, cette propriété n'est pas toujours vérifiée pour le passage de $n = 100$ à $n = 500$, ce qui peut être dû à un nombre insuffisant d'itérations comme nous l'expliquons par la suite.

7.4.3 Discussion sur les simulations

Bootstrap

On ne connaît pas *a priori* la loi limite du rapport de vraisemblance empirique. C'est pourquoi nous avons choisi d'utiliser la méthode du bootstrap pour l'approcher.

Soit $Z = (N_i, X_i, \delta_i)_{1 \leqslant i \leqslant n}$ l'échantillon simulé. On tire $Z^\star := (Z_1^\star, \dots Z_n^\star)$ uniformément dans $\{Z_1, \dots, Z_n\}$, puis on calcule le rapport de vraisemblance empirique des $Z^\star$ en $\widehat{\mu}_j$:

$$\mathrm{EL}_n^\star(\widehat{\mu}_j, t) = \max \left\{ \prod_{i=1}^n np_i, \ \sum_{i=1}^n p_i m_{i,j}(N_i^\star, \widehat{\mu}_j, t) = 0 \text{ et } p \in \mathbb{S}_{n-1} \right\}.$$

On définit ensuite $S^\star$ comme étant le supremum de $\mathrm{EL}_n^\star$ sur $[\tau_1, \tau]$. On répète alors 1000 fois cette procédure et on obtient ainsi une approximation de la fonction de répartition du sup de la loi limite.

Comme on l'a déjà dit dans la partie 3.6, calculer le rapport de vraisemblance empirique d'un échantillon bootstrap peut poser problème en particulier lorsque la taille de l'échantillon est petite. C'est encore plus vrai lorsque l'on considère des processus.

Pour illustrer le problème, reportons-nous à la figure 7.1 et imaginons être en présence de trois réalisations d'un processus de comptage. Il est possible que lors d'une itération bootstrap, le processus dessiné en pointillés (celui qui saute en premier) ne soit pas retenu. Dans ce cas, on ne pourra pas calculer le rapport de vraisemblance empirique en $t = \tau_1$ (on l'avait en fait par convention défini comme étant nul). Mais, pour des valeurs plus grandes de t, on sera en mesure de calculer le rapport de vraisemblance empirique. Dans nos simulations, nous avons donc décidé de ne pas écarter cette itération bootstrap. Rappelons-nous que, *in fine*, c'est la loi du sup du rapport de vraisemblance empirique qui nous intéresse. Ainsi, si nous gardons la convention $\max \varnothing = 0$, en passant au logarithme nous allons nous retrouver en présence de valeurs infinies et nous perdrons toute l'information donnée par les valeurs de t où on sait calculer le rapport de vraisemblance empirique. C'est pourquoi nous avons décidé dans nos algorithmes de poser $\max \varnothing = 1$

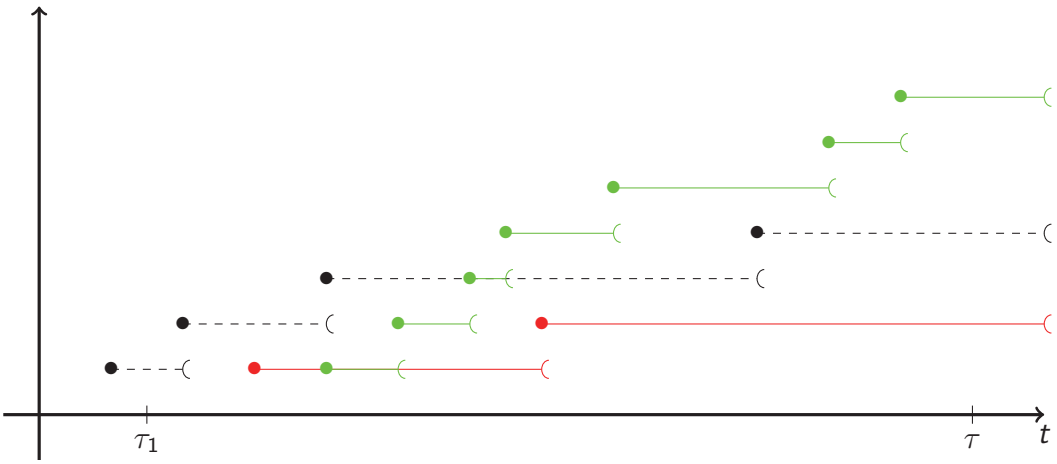


FIGURE 7.1 – Vraisemblance empirique et bootstrap

Cette méthode induit donc un biais vers le bas puisque le sup du rapport de vraisemblance empirique pouvait *a priori* être atteint là où on n'a pas pu le calculer. Ainsi, lorsque l'on calculera le risque de première espèce, on aura un biais vers le haut et les valeurs de α_0 auront tendance à être sur-estimées.

Temps de Calcul

Les calculs ont été effectués au Mésocentre de Calcul de Franche-Comté. Nous avons disposé en permanence de 30 cœurs pour paralléliser les calculs. Cependant, remplir le tableau présenté à la section 7.4.2 aura pris quatre mois. Pourquoi ?

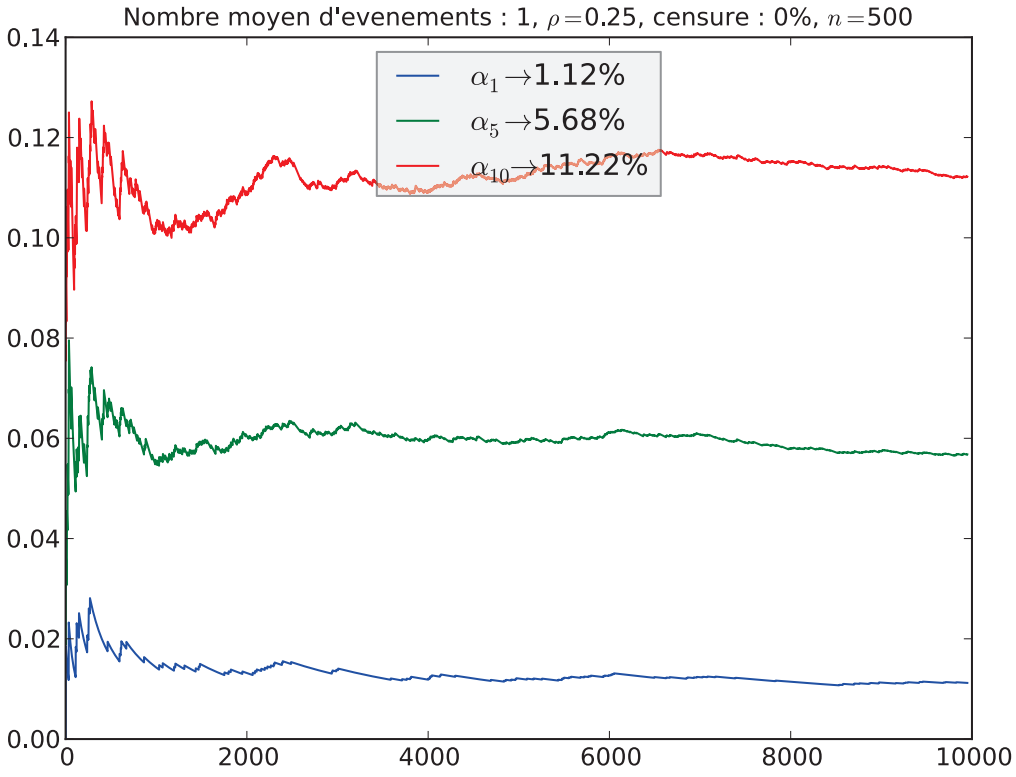
Étant donnés $X_1, \dots, X_n$ réels, calculer le rapport de vraisemblance empirique de $X_1, \dots, X_n$ en un point μ_0 est très rapide. Il s'agit en effet d'un problème d'optimisation convexe et il se résout donc numériquement en une

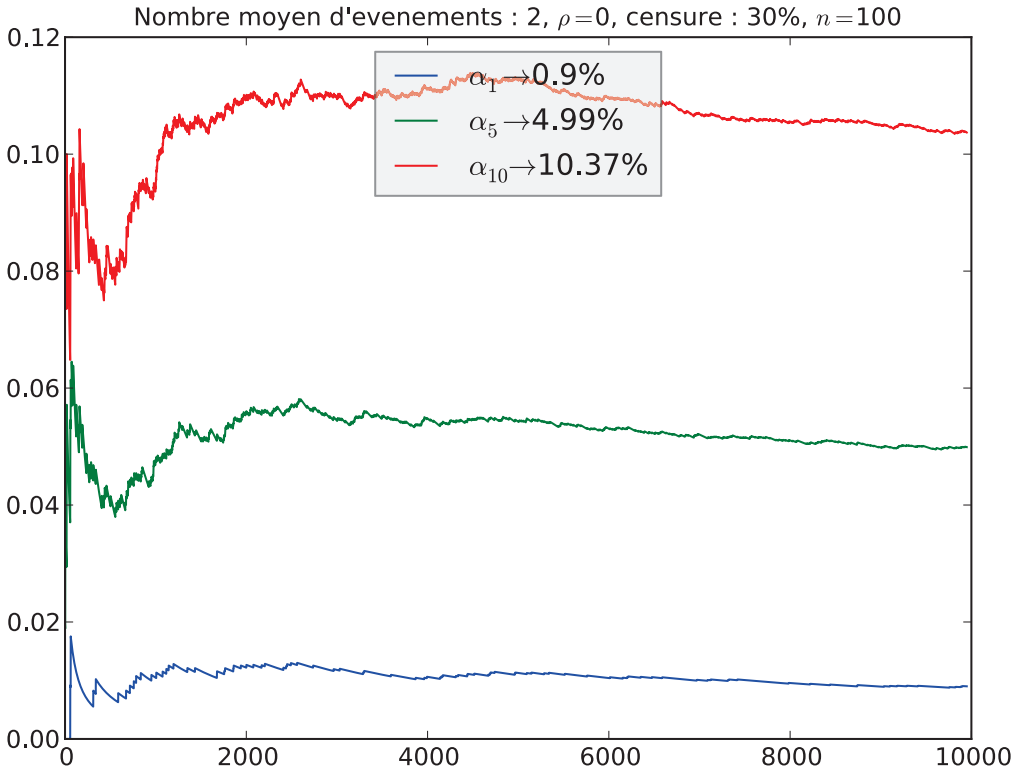
fraction de seconde. Ce n'est cependant pas « instantané ». Dans nos simulations, nous avons discrétisé l'intervalle de temps $[\tau_1, \tau]$: cela multiplie déjà le nombre de problèmes d'optimisations convexes à résoudre. Ensuite, pour chacune des trajectoires simulées (qui sont au nombre de 10000 dans chacune des 24 cases du tableau), on a effectué 1000 itérations bootstrap (sur l'intervalle de temps discrétisé) ce qui représente un très grand nombre de problèmes d'optimisation convexe (environ $3,6 \cdot 10^{11}$) !

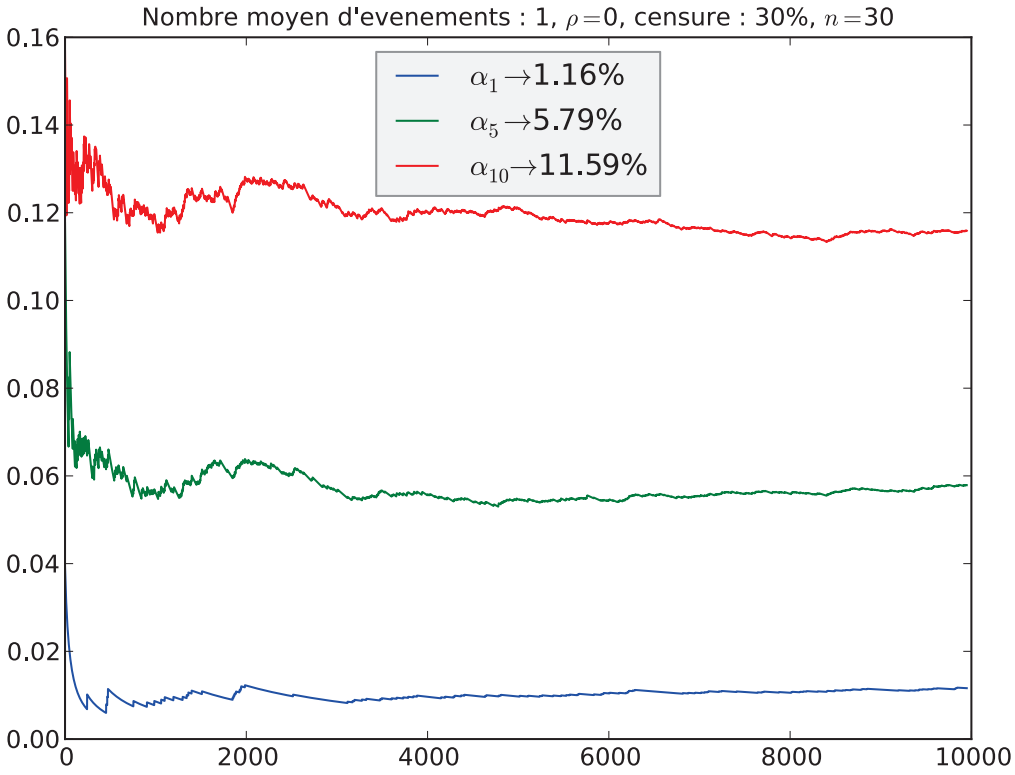
Dans ces conditions, on peut se demander si l'approche « vraisemblance empirique » est bien adaptée au problème. En fait, comme on l'a dit, remplir le tableau est très long. Cependant, le temps de calcul demandé à la construction d'une bande de confiance pour un jeu de données est quant à lui tout à fait raisonnable : de l'ordre d'une dizaine de minutes.

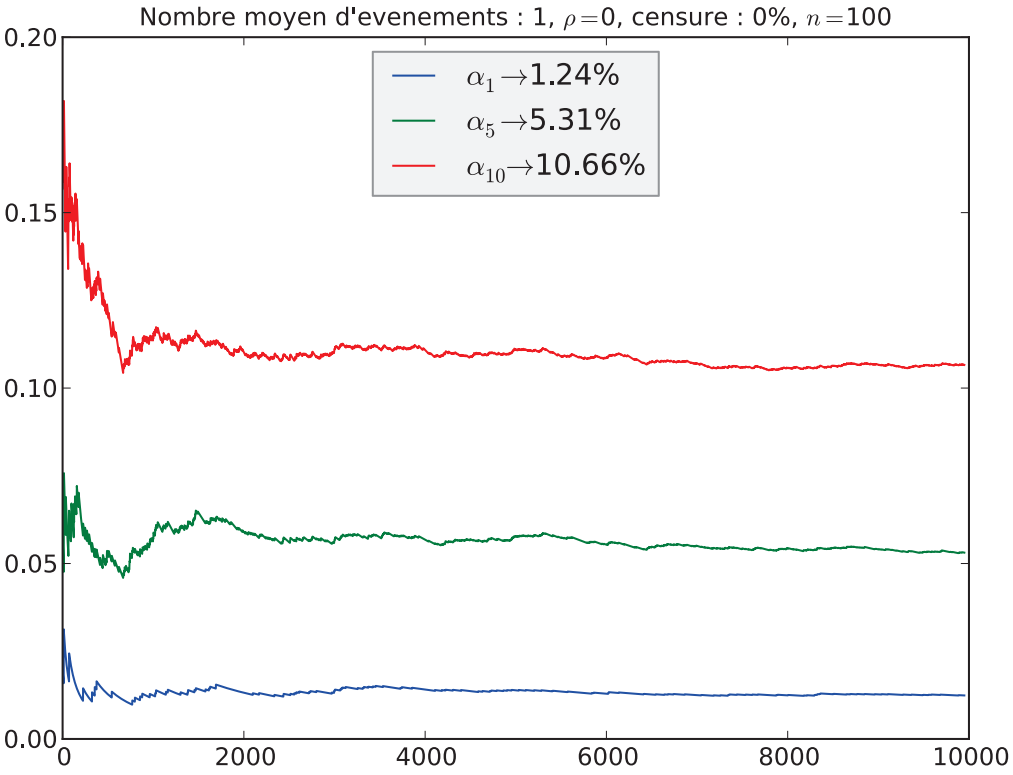
Biais

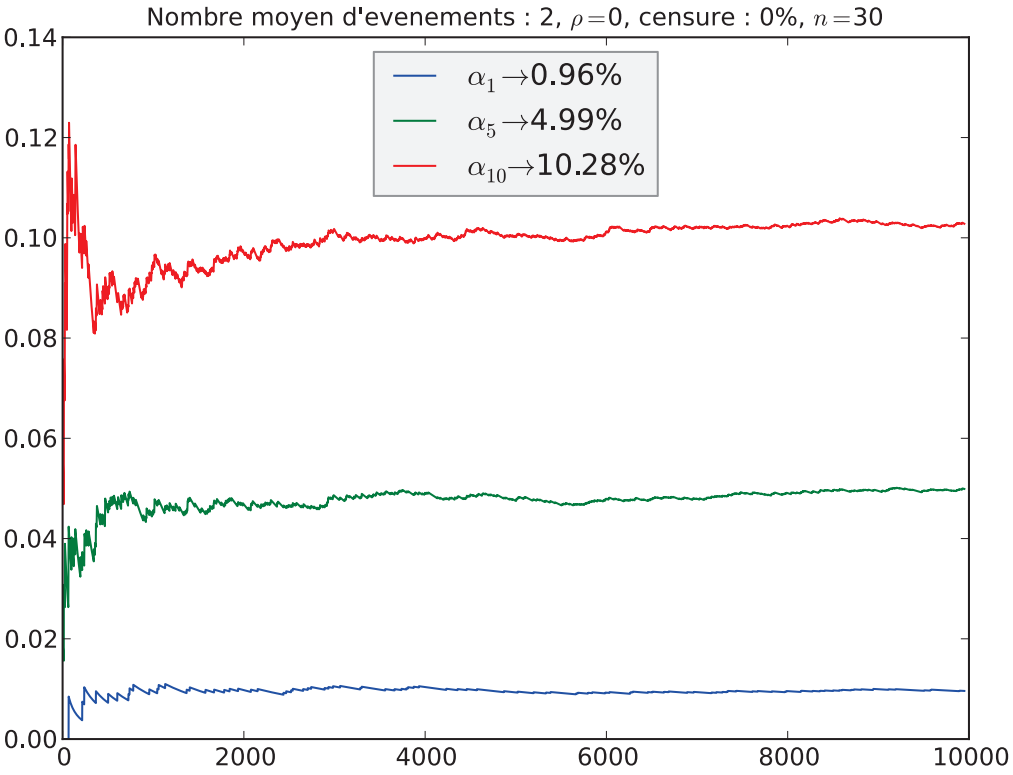
Comme on l'a vu à la section 7.4.3, l'approximation de la loi limite par bootstrap induit un biais vers le haut, en particulier pour les échantillons de petite taille. De plus, les temps de calculs étant très grands, nous nous sommes limité à 10000 itérations Monte-Carlo. Dans certains cas, il aurait été préférable de mener quelques itérations supplémentaires. Sur les graphiques présentés dans la suite, nous avons tracé l'évolution des approximations de α_1, α_5 et α_{10} en fonction du nombre d'itérations pour mettre en évidence ce phénomène. Remarquons que pour certaines simulations, le nombre d'itérations semble suffisant.




Nombre moyen d'evenements : 2, $\rho=0$, censure : 30%, $n=100$
$\alpha_1 \rightarrow$0.9%
$\alpha_5 \rightarrow$4.99%
$\alpha_{10} \rightarrow$10.37%
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2000
4000
6000
8000
10000


Nombre moyen d'evenements : 1, $\rho=0$, censure : 30%, $n=30$
$\alpha_1 \rightarrow$1.16%
$\alpha_5 \rightarrow$5.79%
$\alpha_{10} \rightarrow$11.59%
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2000
4000
6000
8000
10000


Nombre moyen d'evenements : 1, $\rho=0$, censure : 0%, $n=100$
$\alpha_1 \rightarrow$1.24%
$\alpha_5 \rightarrow$5.31%
$\alpha_{10} \rightarrow$10.66%
0.20
0.15
0.10
0.05
0.00
0
2000
4000
6000
8000
10000


Nombre moyen d'evenements : 2, $\rho=0$, censure : 0%, $n=30$
$\alpha_1 \rightarrow$0.96%
$\alpha_5 \rightarrow$4.99%
$\alpha_{10} \rightarrow$10.28%
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
2000
4000
6000
8000
10000

7.5 Application des résultats au jeu de données

Dans cette section nous allons appliquer les résultats vus précédemment pour construire des bandes de confiance pour le nombre moyen d'un certain type d'infection nosocomiale. Nous nous concentrerons uniquement sur les trois infections les plus fréquemment rencontrées dans notre jeu de données, à savoir les pneumonies, les septicémies ainsi que les infections urinaires. Dans un premier temps, nous construirons pour chaque type d'infection une bande de confiance d'ordre 95% pour le nombre moyen d'infections contractées au cours du temps. Ensuite, nous construirons un tube de confiance où nous considérerons deux types d'infections simultanément.

7.5.1 Présentation du jeu de données

Les données, collectées dans une unité de soins intensifs dans un hôpital français sur une période de dix ans, concernent $n = 7867$ patients. Certaines pathologies n'apparaissent qu'une seule fois et nous avons donc choisi de ne regarder que les trois types d'infections les plus fréquentes. Les pneumonies par exemple sont au nombre de 463 et ont touché un total de 373 patients. Nous sommes donc bien en présence d'événements récurrents. De plus, un même patient peut contracter plusieurs infections différentes, c'est pourquoi le cadre des événements concurrents semble adapté à cette étude.

En reprenant les notations introduites à la section 7.1, la dernière observation ayant lieu au jour 380, on a ici $\max_{1 \leqslant i \leqslant n} X_i = 380$.

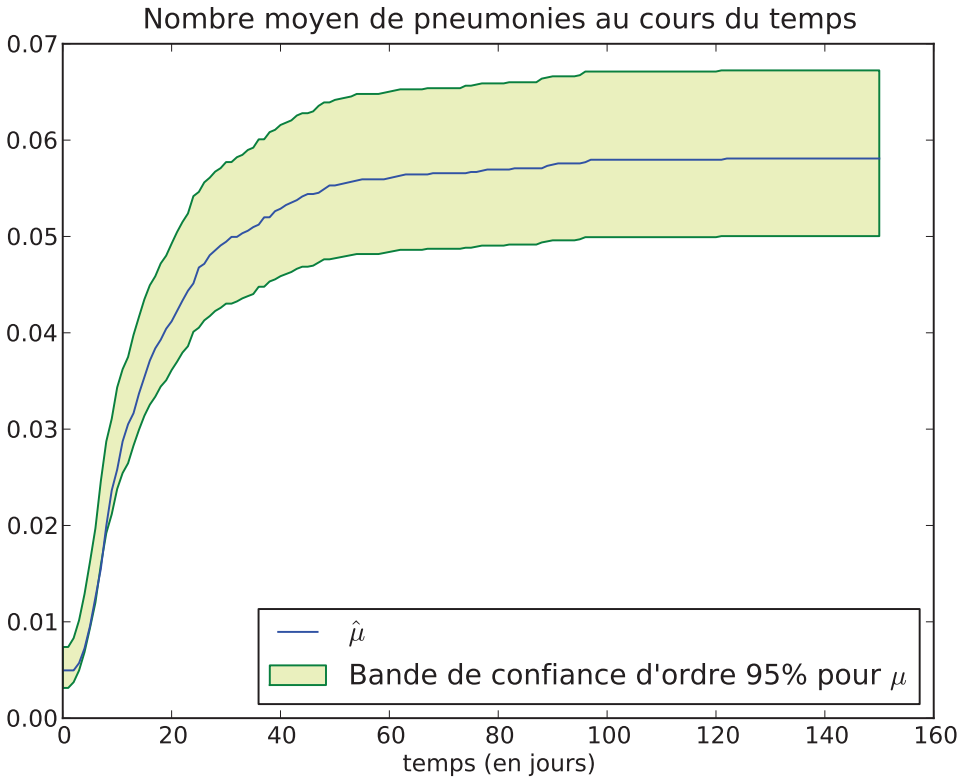
7.5.2 Bandes de confiance

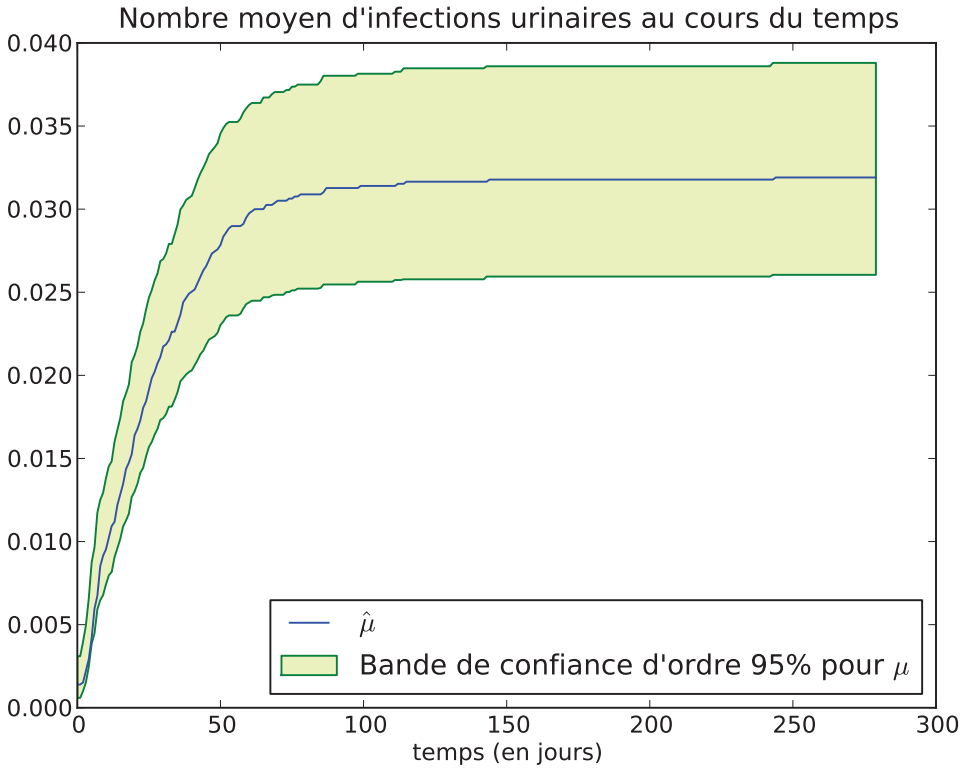
La loi limite a été approchée par bootstrap comme pour les simulations. Pour déterminer une bande de confiance, on utilise la méthode décrite au chapitre 14. On peut remarquer sur les figures ci-dessous plusieurs choses.

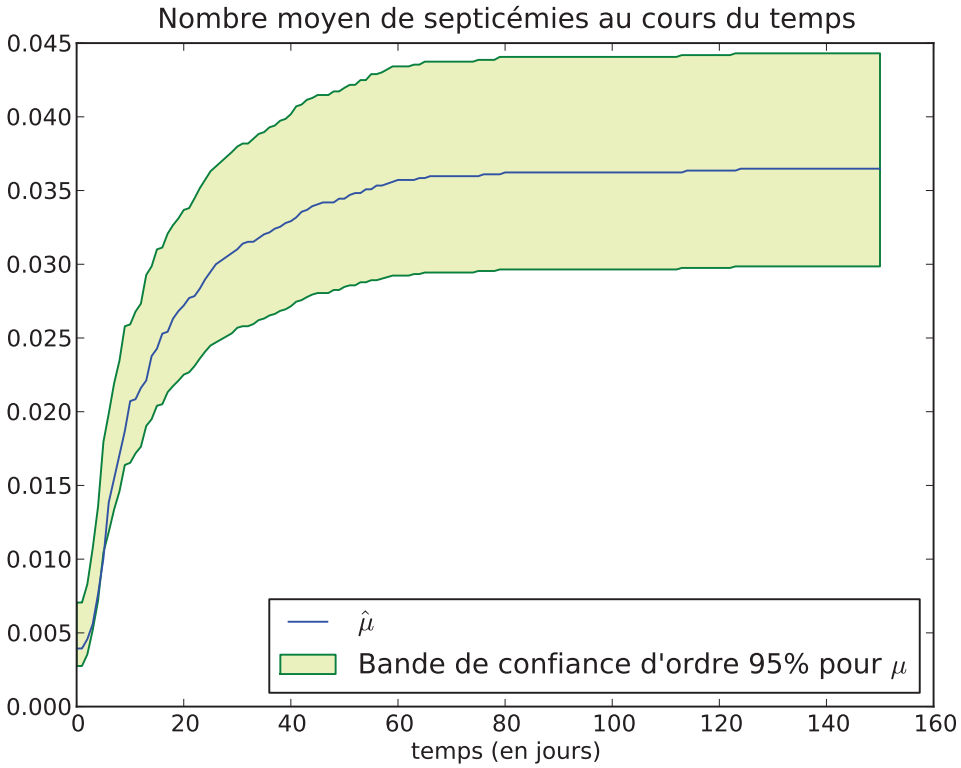
Premièrement, la forme de la bande de confiance ne change plus après le dernier saut observé : il n'y a par exemple eu aucun patient qui a contracté une pneumonie après le 125-ième jour.

Deuxièmement, la bande de confiance n'est pas centrée autour de $\widehat{\mu}$: cela est dû à l'utilisation de la vraisemblance empirique.

Enfin, comme on l'a déjà fait remarquer à la section 3.3, les régions de confiance obtenues par vraisemblance empirique capturent la géométrie des données.


Nombre moyen de pneumonies au cours du temps
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
80
100
120
140
160
temps (en jours)
$\hat{\mu}$
Bande de confiance d'ordre 95% pour μ

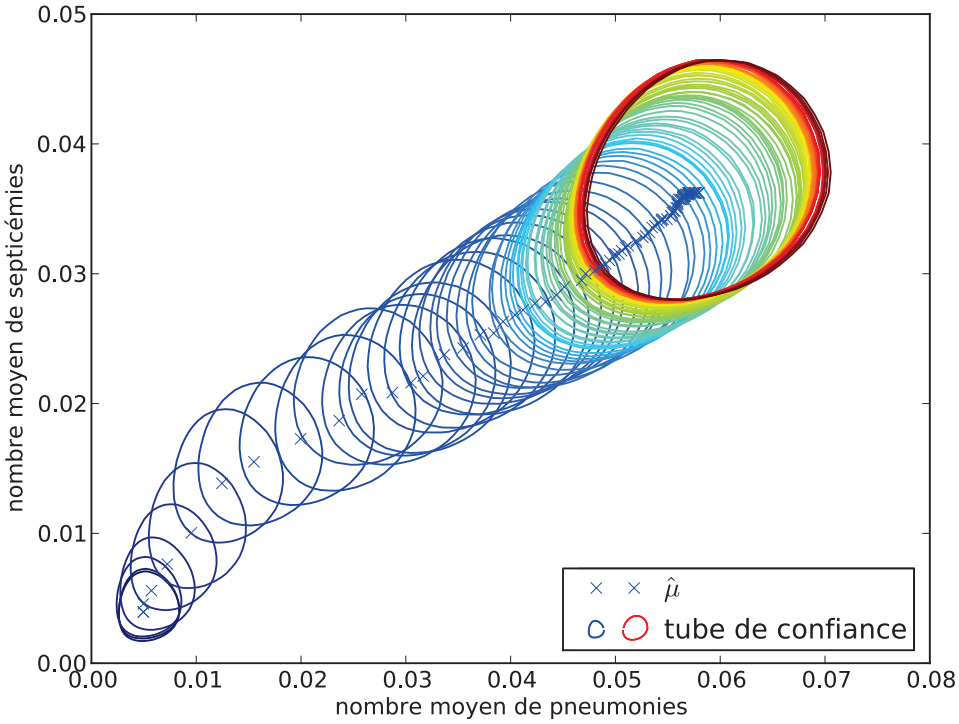

Nombre moyen d'infections urinaires au cours du temps
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0
50
100
150
200
250
300
temps (en jours)
$\hat{\mu}$
Bande de confiance d'ordre 95% pour μ

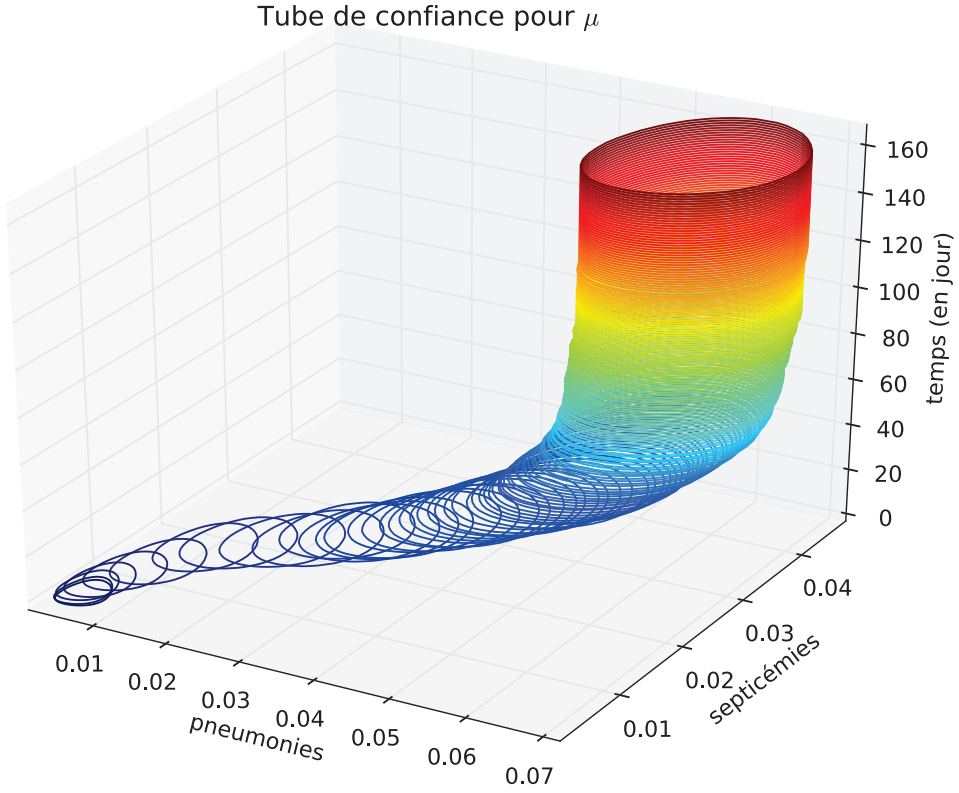


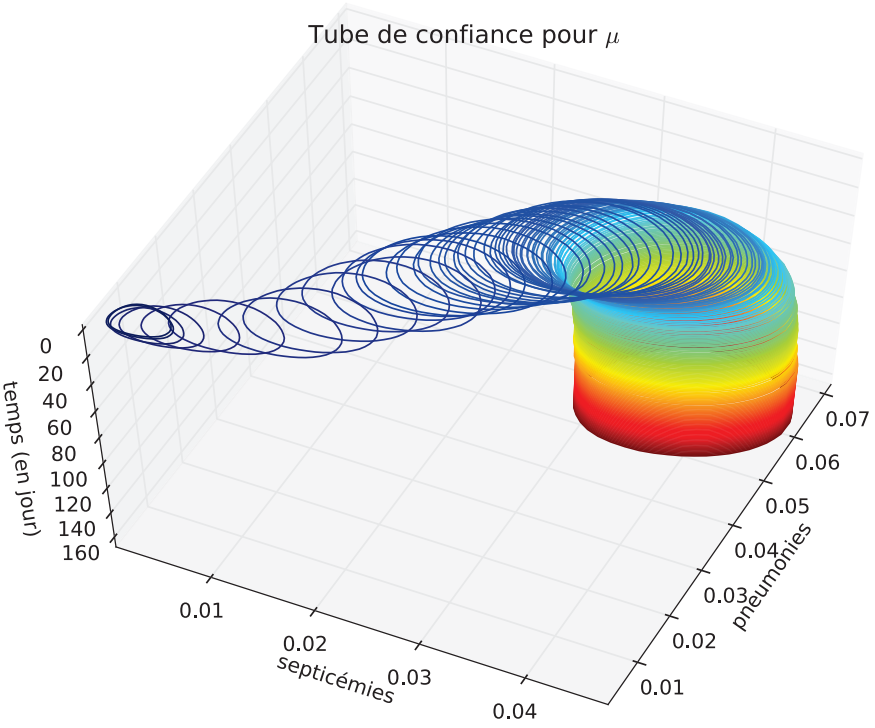
7.5.3 Tube de confiance

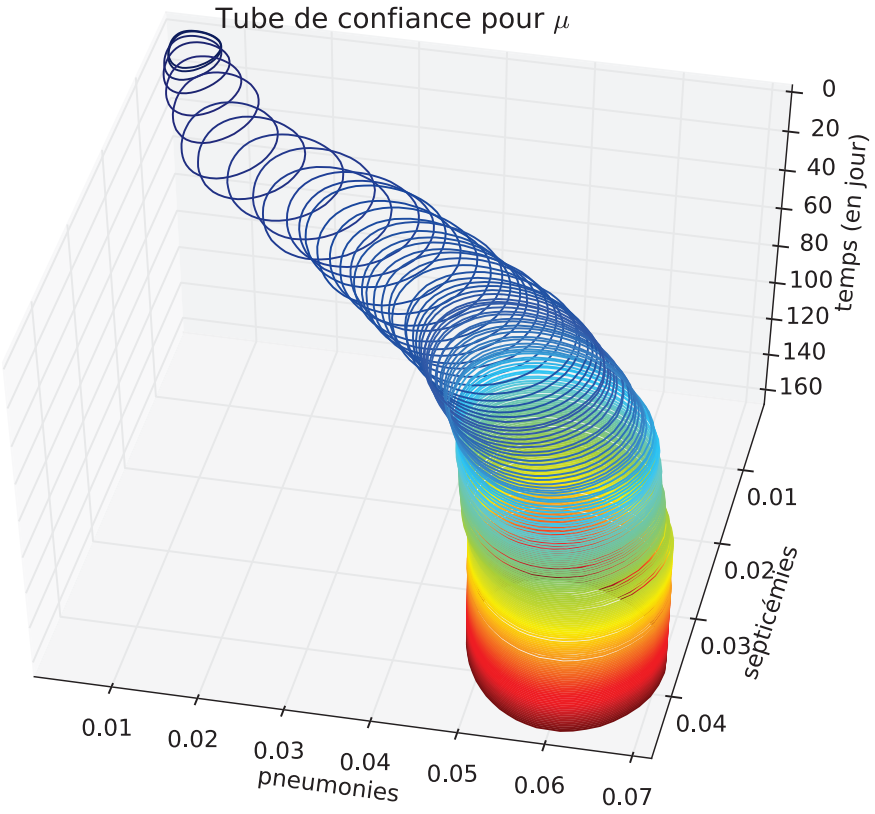
Les résultats présentés à la section 7.3.2 permettent la construction de régions de confiance en dimension supérieure. Nous avons donc tracé un tube de confiance pour le nombre moyen de pneumonies et de septicémies au cours du temps. Par souci de lisibilité il a été « aplati » et dessiné sous plusieurs angles. Une fois de plus, on peut remarquer que le tube n'est pas symétrique comme cela aurait été le cas si on avait fait appel à un théorème central limite.

Le tube de confiance présenté ici est certes difficile à interpréter, mais nous l'avons construit pour montrer sa faisabilité. De plus, on retrouve les résultats de la section précédente en le projetant sur le plan adéquat. Ajoutons enfin que les algorithmes écrits en dimension 1 ne demandent que très peu de modification pour obtenir ce résultat et qu'il aurait alors été dommage de s'en passer.


0.05
0.04
0.03
0.02
0.01
0.00
nombre moyen de septicémies
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
nombre moyen de pneumonies
$\hat{\mu}$
tube de confiance


Tube de confiance pour μ
160
140
120
100
80
60
40
20
0
temps (en jour)
0.01
0.02
0.03
0.04
0.05
0.06
0.07
pneumonies
0.01
0.02
0.03
0.04
septicémies


Tube de confiance pour μ
temps (en jour)
0
20
40
60
80
100
120
140
160
septicémies
0.01
0.02
0.03
0.04
pneumonies
0.01
0.02
0.03
0.04
0.05
0.06
0.07


Tube de confiance pour μ
pneumonies
0.01
0.02
0.03
0.04
0.05
0.06
0.07
septicémies
0.01
0.02
0.03
0.04
temps (en jour)
0
20
40
60
80
100
120
140
160

CHAPITRE 8

Covariables

8.1 Introduction

Dans certaines situations, l'intensité d'un processus de comptage peut dépendre de covariables. En fiabilité par exemple, on peut imaginer que les conditions d'exploitation d'un système ont une influence sur la récurrence des pannes. En sociologie, les périodes de chômage d'un individu peuvent être fonction de son âge, de son sexe, de son niveau d'études ou encore de sa situation géographique. En biostatistiques, dans le cadre de l'étude d'infections nosocomiales, on peut imaginer que beaucoup de données entrent en jeu, notamment l'âge, le sexe, les antécédents médicaux ou encore certaines données biologiques.

Il existe plusieurs modèles permettant de prendre en compte un vecteur de covariables et dans lesquels on s'intéresse à un vecteur de régression faisant état de l'influence de chacune de ces covariables sur l'intensité des processus de comptage en jeu. Citons entre autres le modèle de Cox ou encore le modèle additif de Aalen. On pourra se reporter à Andersen *et al.* (1993) pour une étude détaillée de ces modèles dans laquelle ils donnent des estimateurs consistants et asymptotiquement normaux des paramètres de régression.

Nous avons quant à nous voulu construire une région de confiance asymptotique pour un vecteur de régression par vraisemblance empirique dans un cadre d'événements récurrents. Nous avons choisi de reprendre le modèle présenté dans l'article de Cai & Schaubel (2004b) : il est adapté à l'étude des événements récurrents avec risques concurrents, censure aléatoire à droite et covariables. Précisons que ce modèle n'inclut pas d'événement terminal contrairement à celui du chapitre 7. Dans leur article, Cai & Schaubel (2004b) considèrent un vecteur de covariables dépendant du temps et un vecteur de régression constant. Ils construisent un estimateur des paramètres de régression et montrent sa consistance et sa normalité asymptotique.

Dans ce chapitre, nous ajouterons un paramètre de temps au vecteur de régression et le considérerons donc comme une fonction $\boldsymbol{\beta}_0 \colon [0,\tau] \to \mathbb{R}^p$, où $[0,\tau]$ est un intervalle de temps. C'est dans ce cadre que nous construirons une région de confiance asymptotique pour la fonction $\boldsymbol{\beta}_0$ par vraisemblance empirique.

Dans toute la suite, on supposera être en présence d'individus $i = 1, \dots, n$ pouvant expérimenter K types d'événements sur un intervalle de temps $[0,\tau]$. On notera $N^\star_{ik}(t)$, $k = 1, \dots, K$ les processus de comptage correspondant. On supposera que les observations sont soumises à des censures aléatoires à droite indépendantes (des processus de comptage et des covariables) C_{ik} vérifiant $\mathbb{P}(C_{ik} > \tau) > 0$. On notera alors $Y_{ik}(s) := \mathbb{1}_{\{C_{ik} \geqslant s\}}$ et :

$$N_{ik}(t) := \int_0^t Y_{ik}(s)\, \mathrm{d}N^\star_{ik}(s) = N^\star_{ik}(t \wedge C_{ik}).$$

On notera encore $\mathbf{Z}_{ik}(t) = (\mathbf{Z}_{ik1}(t), \dots, \mathbf{Z}_{ikp}(t))^t \in \mathbb{R}^p$ un vecteur de covariables aléatoires i.i.d. pour tout t dans $[0,\tau]$.

Tout comme dans l'exemple introductif présenté à la section 4.2, on notera $\mathrm{d}N(t)$ l'incrément $N((t+dt)^-) - N(t^-)$. On supposera qu'il existe une fonction $\boldsymbol{\beta}_0 \colon [0,\tau] \to \mathbb{R}^p$ telle que :

$$\mathbb{E}(\mathrm{d}N^\star_{ik}(t) | \mathbf{Z}_{ik}(t)) = g(\boldsymbol{\beta}_0(t)^t \mathbf{Z}_{ik}(t)) \mathrm{d}\mu_{0k}(t), \tag{8.1}$$

où $\mu_{0k}(t) = \int_0^t \mathrm{d}\mu_{0k}(s)$ est une fonction inconnue et où g est une fonction positive, connue et de classe $\mathscr{C}^2$. On pourra par exemple pour g penser à $g(x) = \mathrm{e}^x$ en s'inspirant du modèle de Cox ou encore à $g(x) = 1 + x$ en s'inspirant du modèle additif de Aalen. On se reportera à Andersen *et al.* (1993) pour plus de détails sur ces modèles.

Dans la suite, on supposera que $\boldsymbol{\beta}_0$ ainsi que le vecteur de covariables sont à variations bornées sur $[0,\tau]$, au sens où il existe une constante c_Z telle que, presque sûrement :

$$\forall (i,k,l) \in [\![1,n]\!] \times [\![1,K]\!] \times [\![1,p]\!] \ : \quad V_0^\tau(\mathbf{Z}_{ikl}) < c_Z.$$

On supposera aussi que les N_{ik} sont uniformément bornés par une constante B et que :

$$\exists \delta > 0 \ / \ \forall s \in [0,\tau] \ : \ g(\boldsymbol{\beta}_0(s)^t \mathbf{Z}_{ik}(s)) > \delta.$$

Enfin, nous munirons $M_p(\mathbb{R})$ et $\mathbb{R}^p$ des normes définies par :

$$\forall x = (x_1, \dots, x_p)^t \in \mathbb{R}^p, \quad \|x\| := \max_{1 \leqslant i \leqslant p} |x_i|,$$
$$\forall A = (a_{ij})_{ij} \in M_p(\mathbb{R}), \qquad \|A\| := \max_{1 \leqslant i,j \leqslant p} |a_{i,j}|.$$

8.2 Résultats préliminaires

Comme le remarquent Cai & Schaubel (2004b) dans leur article, l'équation (8.1) suggère d'estimer $\boldsymbol{\beta}_0$ au travers de :

$$\sum_{i=1}^n \int_0^\tau \left(\mathrm{d}N_{ik}(s) - Y_{ik}(s) g(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s)) \mathrm{d}\mu_{0k}(s) \right) = 0. \tag{8.2}$$

Or, dans cette équation apparaissent les paramètres inconnus μ_{0k}. L'idée de Cai & Schaubel (2004b), empruntée à Liang & Zeger (1986), est de proposer une autre équation pour estimer $\boldsymbol{\beta}_0$:

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \mathbf{Z}_{ik}(s)\frac{g'(\boldsymbol{\beta}(s)^t\mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}(s)^t\mathbf{Z}_{ik}(s))}\big(\mathrm{d}N_{ik}(s) - Y_{ik}(s)g(\boldsymbol{\beta}(s)^t\mathbf{Z}_{ik}(s))\mathrm{d}\mu_{0k}(s)\big) = \mathbf{0}. \tag{8.3}$$

Pour faire disparaître la fonction inconnue $\mathrm{d}\mu_{0k}$ dans (8.3), on l'estime via (8.2) par :

$$\mathrm{d}\mu_{0k}(s) = \frac{1}{n}\sum_{i=1}^{n}\frac{n}{\sum_{i=1}^{n} Y_{ik}(s)g(\boldsymbol{\beta}_0(s)^t\mathbf{Z}_{ik}(s))}\mathrm{d}N_{ik}(s). \tag{8.4}$$

En combinant alors (8.4) et (8.3) on obtient une nouvelle équation pour estimer $\boldsymbol{\beta}_0$ qui ne fait plus intervenir les μ_{0k}, à savoir :

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\int_0^\tau \left(\mathbf{Z}_{ik}(s)\cdot\frac{g'(\boldsymbol{\beta}(s)^t\mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}(s)^t\mathbf{Z}_{ik}(s))} - \mathbf{E}_k(s,\boldsymbol{\beta})\right)\mathrm{d}N_{ik}(s) = \mathbf{0}, \tag{8.5}$$

où $\mathbf{E}_k$ est la fonction définie par :

$$\mathbf{E}_k(s,\boldsymbol{\beta}) = \frac{\sum_{i=1}^{n} Y_{ik}(s)\mathbf{Z}_{ik}(s)g'(\boldsymbol{\beta}(s)^t\mathbf{Z}_{ik}(s))}{\sum_{i=1}^{n} Y_{ik}(s)g(\boldsymbol{\beta}(s)^t\mathbf{Z}_{ik}(s))}.$$

C'est cette équation (8.5) faisant intervenir une somme de variables aléatoires « presque i.i.d. » qui motive l'utilisation de la vraisemblance empirique. En effet, comme nous le verrons à la section 8.3, l'idée est alors de maximiser le rapport de vraisemblance empirique sous la contrainte que la version pondérée du membre de gauche dans (8.5) est égale à $\mathbf{0}$.

Avant de rentrer dans les détails, il nous faut démontrer quelques résultats préliminaires. Notons :

$$\mathbf{e}_k(s,\boldsymbol{\beta}_0) = \frac{\mathbb{E}(Y_{ik}(s)\mathbf{Z}_{ik}(s)g'(\boldsymbol{\beta}_0(s)^t\mathbf{Z}_{ik}(s)))}{\mathbb{E}(Y_{ik}(s)g(\boldsymbol{\beta}_0(s)^t\mathbf{Z}_{ik}(s))}.$$

Alors, d'après la loi des grands nombres, on a pour tout $s \in [0,\tau]$:

$$\mathbf{e}_k(s,\boldsymbol{\beta}) = \lim_{n\to\infty}\mathbf{E}_k(s,\boldsymbol{\beta}) \quad \text{p.s..}$$

Lemme 8.1. *Il existe un processus gaussien centré U^k multivarié de dimension p tel que :*

$$\sqrt{n}\big(\mathbf{E}_k(\cdot,\boldsymbol{\beta}_0) - \mathbf{e}_k(\cdot,\boldsymbol{\beta}_0)\big) \rightsquigarrow U^k(\cdot), \quad \text{lorsque } n\to+\infty.$$

Démonstration. Pour démontrer ce lemme, nous allons utiliser le théorème de la limite centrale fonctionnel vectoriel ainsi que la δ-méthode fonctionnelle. À cet effet, posons pour tout i :

$$f_{ik}(\omega,s) := \Big(Y_{ik}(s)\mathbf{Z}_{ik}(s)g'(\boldsymbol{\beta}_0(s)^t\mathbf{Z}_{ik}(s))\Big)(\omega).$$

Notons encore pour tout $l \in [\![1, p]\!]$:

$$f_{ikl}(\cdot, s) := Y_{ik}(s)\mathbf{Z}_{ikl}(s)g'(\beta_0(s)^t\mathbf{Z}_{ik}(s)).$$

Par hypothèse, les $\mathbf{Z}_{ikl}$ sont à variations bornées uniformément en ω et en i. On a aussi supposé que β_0 était à variations bornées et que la fonction g était de classe $\mathscr{C}^2$. Ainsi, à ω fixé, les $f_{ikl}(\omega, \cdot)$ sont à variations bornées par une constante M indépendante de i et de ω (on se reportera à la section 2.7 pour plus de détails sur les fonctions à variations bornées). Posons maintenant pour tout $l \in [\![1, p]\!]$:

$$\mathcal{F}_{n\omega}^{kl} := \left\{ \big(f_{1kl}(\omega, s), , \dots, f_{nkl}(\omega, s)\big),\ s \in [0, \tau] \right\}.$$

Alors, d'après le théorème 2.28, les classes $\mathcal{F}_{n\omega}^{kl}$ sont euclidiennes. De plus, les $f_{ikl}(\omega, t)$ sont uniformément bornés (en i, ω et t) par $2M$. On peut donc appliquer le théorème de la limite centrale fonctionnel puisque les processus en jeu sont i.i.d. et dans $L^2(\Omega)$ (cf. théorème 2.32 et remarque 2.33). Ainsi, il existe des processus gaussiens centrés $U_1^{k1}, \dots, U_1^{kp}$ tels que pour tout $l \in [\![1, p]\!]$:

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_{ik}\mathbf{Z}_{ikl}g'(\beta_0^t\mathbf{Z}_{ikl}) - \mathbb{E}\left(Y_{ik}\mathbf{Z}_{ikl}g'(\beta_0^t\mathbf{Z}_{ikl})\right)\right) \rightsquigarrow U_1^{kl}.$$

On démontre de même qu'il existe des processus U_2^k tels que :

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} Y_{ik}g(\beta_0^t\mathbf{Z}_{ik}) - \mathbb{E}\left(Y_{ik}g(\beta_0^t\mathbf{Z}_{ik})\right)\right) \rightsquigarrow U_2^k.$$

Notons alors pour k fixé W_i le processus à valeurs dans $\mathbb{R}^{p+1}$ dont les coordonnées sont les $Y_{ik}\mathbf{Z}_{ikl}g'(\beta_0^t\mathbf{Z}_{ikl})$ ainsi que les $Y_{ik}g(\beta_0^t\mathbf{Z}_{ik})$. D'après ce qui précède, la suite de terme général $n^{-1/2}\sum_{i=1}^{n}(W_i(\cdot) - \mathbb{E}W_i(\cdot))$ est asymptotiquement équicontinue (cf section 2.10). Pour montrer qu'elle converge en distribution, il suffit donc de montrer que ses marges fini-dimensionnelles convergent. Or, comme les W_i sont i.i.d. et bornés dans $\mathbb{R}^{p+1}$, le théorème de la limite centrale multivarié classique nous assure que c'est le cas.

Soit maintenant ψ la fonction définie par :

$$\begin{array}{rccc} \psi & : & \mathbb{R}^p \times \mathbb{R}^\star & \to & \mathbb{R}^p \\ & & (x, y) & \mapsto & y^{-1}x. \end{array}$$

Alors, la fonction ψ est différentiable. On sait aussi par hypothèse que pour tout $s \in [0, \tau]$, la fonction $g(\beta_0(s)^t\mathbf{Z}_{ik}(s))$ est non nulle. On peut donc appliquer la δ-méthode fonctionnelle (voir par exemple van der Vaart & Wellner (1996) page 372). D'où :

$$\sqrt{n}\big(\mathbf{e}_k(\cdot, \beta_0) - \mathbf{E}_k(\cdot, \beta_0)\big) \rightsquigarrow \mathrm{d}\psi(\theta).(U_1^k, U_2^k) =: U^k,$$

où $\theta := \left(\mathbb{E}\left(Y_{ik}\mathbf{Z}_{ik}g'(\beta_0^t\mathbf{Z}_{ik})\right), \mathbb{E}\left(Y_{ik}g(\beta_0^t\mathbf{Z}_{ik})\right)\right)^t$. □

Le lemme ci-dessous, présenté dans l'annexe de l'article de Lin *et al.* (2000), nous sera utile dans la section 8.3 pour démontrer que l'hypothèse (A1) du théorème de Hjort *et al.* (2004) est satisfaite.

Lemme 8.2 (Lin *et al.* (2000))**.** *Soient* $(f_n)_{n\in\mathbb{N}}$ *et* $(g_n)_{n\in\mathbb{N}}$ *deux suites de fonctions définies sur* $[0,\tau]$*. Supposons que :*

(i) les g_n *sont monotones et bornées ;*

(ii) il existe une fonction f *continue sur* $[0,\tau]$ *telle que :*

$$\sup_{t\in[0,\tau]} |f_n(t) - f(t)| \longrightarrow 0 \quad \text{lorsque } n \to +\infty;$$

(iii) il existe une fonction g *bornée sur* $[0,\tau]$ *telle que :*

$$\sup_{t\in[0,\tau]} |g_n(t) - g(t)| \longrightarrow 0 \quad \text{lorsque } n \to +\infty.$$

Alors :

$$\sup_{t\in[0,\tau]} \left| \int_0^t f_n(s)\mathrm{d}g_n(s) - \int_0^t f(s)\mathrm{d}g(s) \right| \longrightarrow 0 \quad \text{lorsque } n \to +\infty.$$

8.3 Vraisemblance empirique

Comme nous l'avons vu à la section 8.2, l'équation (8.5) suggère de considérer les quantités :

$$\widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}) := \sum_{k=1}^{K} \int_0^\tau \left(\mathbf{Z}_{ik}(s) \cdot \frac{g'(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))} - \mathbf{E}_k(s, \boldsymbol{\beta}) \right) \mathrm{d}N_{ik}(s),$$

et

$$\mathrm{EL}_n(\boldsymbol{\beta}) := \max \left\{ \prod_{i=1}^{n} np_i \ , \quad \sum_{i=1}^{n} p_i \widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}) = 0, \ p \in \mathbb{S}_{n-1} \right\}.$$

Notons encore :

$$m_i(\mathbf{Z}_i, \boldsymbol{\beta}) := \sum_{k=1}^{K} \int_0^\tau \left(\mathbf{Z}_{ik}(s) \cdot \frac{g'(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))} - \mathbf{e}_k(s, \boldsymbol{\beta}) \right) \mathrm{d}N_{ik}(s),$$

$$\Gamma_{in}(\boldsymbol{\beta}) := \sum_{k=1}^{K} \int_0^\tau \big(\mathbf{e}_k(s, \boldsymbol{\beta}) - \mathbf{E}_k(s, \boldsymbol{\beta}) \big) \mathrm{d}N_{ik}(s).$$

Enfin, définissons les quantités désormais classiques du théorème de Hjort *et al.* (2004) :

$$M_n(\boldsymbol{\beta}) := \frac{1}{n} \sum_{i=1}^{n} \widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}) \quad \text{et} \quad S_n(\boldsymbol{\beta}) := \frac{1}{n} \sum_{i=1}^{n} \widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta})^{\otimes 2}.$$

Théorème 8.3

Supposons que $V := \mathbb{V}\mathrm{ar}\, m(\mathbf{Z}_i, \boldsymbol{\beta}_0) \succ 0$. Supposons de plus que :

$$\inf_{\theta \in \Theta} \mathbb{E}\,|\langle m(\mathbf{Z}_i, \boldsymbol{\beta}_0), \theta\rangle| > 0,$$

où Θ désigne la sphère unité de $\mathbb{R}^k$. Alors, il existe un vecteur aléatoire centré G tel que :

$$-2 \ln \mathrm{EL}_n(\boldsymbol{\beta}) \rightsquigarrow G^t V^{-1} G \quad \text{lorsque } n \to +\infty.$$

Pour pouvoir appliquer le théorème de Hjort *et al.* (2004), nous allons montrer que les hypothèses (A0) à (A3) sont satisfaites. Rappelons que :

(A0) : $\mathbb{P}(\mathrm{EL}_n(\boldsymbol{\beta}_0) = 0) \longrightarrow 0$, lorsque $n \to \infty$;

(A1) : $\sqrt{n}\, M_n(\boldsymbol{\beta}_0) \rightsquigarrow G$, lorsque $n \to \infty$;

(A2) : $S_n(\boldsymbol{\beta}_0) \xrightarrow{\mathbb{P}} V$, lorsque $n \to \infty$;

(A3) : $\max_{1 \leqslant i \leqslant n} \|\widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}_0)\| = o_p(n^{1/2})$.

Démonstration de (A1). Écrivons, en omettant le paramètre de temps pour alléger les notations :

$$\begin{aligned}
\sqrt{n} \cdot M_n(\boldsymbol{\beta}_0) &= \sum_{k=1}^{K} \sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left(\mathbf{Z}_{ik} \cdot \frac{g'(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})}{g(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})} - \mathbf{e}_k(\cdot, \boldsymbol{\beta}_0) \right) \mathrm{d}N_{ik} \\
&\quad + \sum_{k=1}^{K} \int_0^\tau \sqrt{n}\big(\mathbf{e}_k(\cdot, \boldsymbol{\beta}_0) - \mathbf{E}_k(\cdot, \boldsymbol{\beta}_0)\big) \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{d}N_{ik} \right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(\mathbf{Z}_i, \boldsymbol{\beta}_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_{in}(\boldsymbol{\beta}_0).
\end{aligned}$$

Première étape. Convergence du premier terme. Par indépendance des censures C_{ik} avec les covariables, on a :

$$\mathbf{e}_k(s, \boldsymbol{\beta}_0) = \frac{\mathbb{E}(\mathbf{Z}_{ik}(s) g'(\boldsymbol{\beta}_0(s)^t \mathbf{Z}_{ik}(s)))}{\mathbb{E}(g(\boldsymbol{\beta}_0(s)^t \mathbf{Z}_{ik}(s))}.$$

Donc, d'après le théorème de Fubini :

$$\begin{aligned}
&\int_0^\tau \mathbb{E}\left(\mathbb{E}\left(\left(\mathbf{Z}_{ik} \frac{g'(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})}{g(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})} - e_k(\cdot, \boldsymbol{\beta}) \right) \mathrm{d}N_{ik} \,\middle|\, \mathbf{Z}_{ik} \right) \right) \\
&\quad = \int_0^\tau \mathbb{E}\left(\left(\mathbf{Z}_{ik} \frac{g'(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})}{g(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})} - e_k(\cdot, \boldsymbol{\beta}) \right) \mathbb{E}\left(\mathrm{d}N_{ik} \,|\, \mathbf{Z}_{ik} \right) \right) \\
&\quad = \int_0^\tau \mathbb{E}\left(\left(\mathbf{Z}_{ik} \frac{g'(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})}{g(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})} - e_k(\cdot, \boldsymbol{\beta}) \right) \mathbb{P}(C_{ik} \geqslant \cdot) g(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik}) \mathrm{d}\mu_{0k} \right) \\
&\quad = \int_0^\tau \left(\mathbb{E}(\mathbf{Z}_{ik} g'(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik})) - \mathbb{E}\left(\mathbf{Z}_{ik} g'(\boldsymbol{\beta}_0^t \mathbf{Z}_{ik}) \right) \right) \mathbb{P}(C_{ik} \geqslant \cdot) \mathrm{d}\mu_{0k} \\
&\quad = 0.
\end{aligned}$$

La convergence en loi est alors une conséquence du théorème de la limite centrale multivarié.

Deuxième étape. Convergence du deuxième terme. D'après la loi uniforme des grands nombres (cf théorème 2.29 page 38), comme les N_{ik} sont presque sûrement bornés, on a, presque sûrement :

$$\sup_{t\in[0,\tau]} \left| \frac{1}{n} \sum_{i=1}^{n} N_{ik}(t) - \mu_k(t) \right| \longrightarrow 0 \quad \text{lorsque } n \to +\infty,$$

où $\mu_k(\cdot) = \mathbb{E}\, N_{ik}(\cdot)$. Donc, d'après le lemme 8.1 :

$$\begin{pmatrix} \sqrt{n}(\mathbf{e}_k - \mathbf{E}_k) \\ \frac{1}{n}\sum_{i=1}^{n} N_{ik} \end{pmatrix} \rightsquigarrow \begin{pmatrix} U \\ \mu_k \end{pmatrix}.$$

Ainsi, d'après le théorème de représentation presque sûre (cf page 21), il existe deux suites $(f_n)_{n\in\mathbb{N}^\star}$ et $(g_n)_{n\in\mathbb{N}^\star}$ de même loi que $\sqrt{n}(\mathbf{e}_k - \mathbf{E}_k)$ et que $\frac{1}{n}\sum_{i=1}^{n} N_{ik}$ et qui convergent presque sûrement. Soit Ω_0 un ensemble de probabilité 1 sur lequel il y a convergence partout. Alors, pour tout ω de Ω_0 :

(i) les $g_n(\omega, \cdot)$ sont croissantes et bornées sur $[0, \tau]$;

(ii) $U(\omega, \cdot)$ est une fonction continue sur $[0, \tau]$ et :

$$\sup_{t\in[0,\tau]} |f_n(\omega, t) - U(\omega, t)| \longrightarrow 0 \quad \text{lorsque } n \to +\infty;$$

(iii) $\mu_k(\cdot) = \mathbb{E}\, N_{ik}(\cdot)$ est bornée et :

$$\sup_{t\in[0,\tau} |g_n(\omega, t) - \mu_k(t)| \longrightarrow 0 \quad \text{lorsque } n \to +\infty.$$

Donc, d'après le lemme 8.2, presque sûrement :

$$\sup_{t\in[0,\tau]} \left| \int_0^t f_n(s)\mathrm{d}g_n(s) - \int_0^t U(s)\mathrm{d}\mu(s) \right| \longrightarrow 0 \quad \text{lorsque } n \to +\infty.$$

Mais, les processus en jeu sont mesurables par rapport à la tribu engendrée par les boules (cf. van der Vaart & Wellner (1996) page 46) et il y a donc aussi convergence en loi. Finalement :

$$\int_0^{\cdot} \sqrt{n}\big(\mathbf{e}_k(s, \beta_0) - \mathbf{E}_k(s, \beta_0)\big) \left(\frac{1}{n} \sum_{i=1}^{n} \mathrm{d}N_{ik}(s) \right) \rightsquigarrow \int_0^{\cdot} U(s)\mathrm{d}\mu(s).$$

Troisième étape. D'après ce qui précède, les deux suites de processus de termes général respectifs $n^{-1/2}\sum_{i=1}^{n} m(\mathbf{Z}_i, \beta_0)$ et $n^{-1/2}\sum_{i=1}^{n} \Gamma_{in}(\beta_0)$ sont asymptotiquement équicontinues. Cela signifie pour la première qu'il existe une semi-métrique ρ_1 telle que pour tout $\varepsilon > 0$ et pour tout $\eta > 0$, il existe $\delta > 0$ tel que :

$$\overline{\lim_n}\ \mathbb{P}^\star \left(\sup_{\rho_1(s,t)<\delta} |n^{-1/2}\big(m_i(\mathbf{Z}_i, \beta_0)(s) - m_i(\mathbf{Z}_i, \beta_0)(s)\big)| > \eta \right) < \varepsilon.$$

De même, il existe une semi-métrique ρ_2 pour laquelle la deuxième suite est asymptotiquement équicontinue. Ainsi, on peut construire une semi-métrique ρ (définie par exemple par $\rho(s,t) = \max(\rho_1(s,t), \rho_2(s,t))$ pour laquelle la somme des deux processus est asymptotiquement équicontinue.

Donc, pour montrer que $\sqrt{n}M_n(\beta_0)$ converge en loi, il suffit de montrer que ses marges fini-dimensionnelles convergent. Or, c'est le cas d'après le théorème de la limite centrale multivarié usuel.

□

Démonstration de (A2). En gardant les mêmes notations qu'au point (A1), on a :

$$\begin{aligned} S_n(\beta_0) = \frac{1}{n}\sum_{i=1}^{n} m_i(\mathbf{Z}_i,\beta_0)^{\otimes 2} + \frac{1}{n}\sum_{i=1}^{n} m_i(\mathbf{Z}_i,\beta_0)\Gamma_{in}(\beta_0)^t \\ + \frac{1}{n}\sum_{i=1}^{n} \Gamma_{in}(\beta_0) m_i(\mathbf{Z}_i,\beta_0)^t + \frac{1}{n}\sum_{i=1}^{n} \Gamma_{in}(\beta_0)^{\otimes 2}. \end{aligned} \tag{8.6}$$

En utilisant le lemme 8.1 et le fait que les N_{ik} sont presque sûrement bornés (uniformément en i et en k) on montre facilement que :

$$\frac{1}{n}\sum_{i=1}^{n} \Gamma_{in}(\beta_0)^{\otimes 2} \xrightarrow{\mathbb{P}} 0, \quad \text{lorsque } n \to +\infty.$$

Pour ce faire, il suffit d'appliquer l'inégalité élémentaire ci-dessous

$$\|xy^t\| = \max_{i,j} |x_i y_j| \leqslant \|x\|\,\|y\|, \tag{8.7}$$

et de remarquer que, presque sûrement :

$$\|\Gamma_{in}(\beta_0)\| \leqslant \sup_{t\in[\tau_1,\tau]} \|\mathbf{e}_k(s,\beta_0) - \mathbf{E}_k(s,\beta_0)\| B. \tag{8.8}$$

De plus, on a d'après la loi des grands nombres :

$$\frac{1}{n}\sum_{i=1}^{n} m_i(\mathbf{Z}_i,\beta_0)^{\otimes 2} \to \mathbb{V}\mathrm{ar}(m_i(\mathbf{Z}_i,\beta_0)), \quad \text{presque sûrement.}$$

Finalement, on démontre que les termes croisés dans (8.6) convergent en probabilité vers 0 en utilisant l'inégalité (8.7) ainsi que l'inégalité de Cauchy-Schwarz. Par exemple, pour le premier terme on obtient la majoration suivante :

$$\begin{aligned} \left\| \frac{1}{n}\sum_{i=1}^{n} m_i(\mathbf{Z}_i,\beta_0)\Gamma_{in}(\beta_0)^t \right\| &\leqslant \frac{1}{n}\sum_{i=1}^{n} \|\Gamma_{in}(\beta_0)\|\,\|m_i(\mathbf{Z}_i,\beta_0)\| \\ &\leqslant \left(\frac{1}{n}\sum_{i=1}^{n} \|\Gamma_{in}(\beta_0)\|^2\right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n} \|m_i(\mathbf{Z}_i,\beta_0)\|^2\right)^{1/2}. \end{aligned}$$

□

Démonstration de (A0). Nous avons démontré au point (A2) (cf. équation (8.8)) que :

$$\max_{1\leqslant i\leqslant n} \|m_i(\mathbf{Z}_i, \boldsymbol{\beta}_0) - \widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}_0)\| \xrightarrow{\mathbb{P}} 0, \quad \text{lorsque } n \to +\infty.$$

Donc, toutes les conditions présentées à la section 5.2 sont réunies. □

Démonstration de (A3). On remarque que :

$$\|\widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}_0)\| \leqslant \sup_{s\in[0,\tau]} \sum_{k=1}^{K} \left\| \mathbf{Z}_{ik}(s) \frac{g'(\boldsymbol{\beta}_0(s)^t \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}_0(s)^t \mathbf{Z}_{ik}(s))} - \mathbf{E}_k(s, \boldsymbol{\beta}_0) \right\| B.$$

Mais, par hypothèse, les fonctions apparaissant dans cette équation sont toutes à variations bornées. De plus, les $g(\boldsymbol{\beta}_0(s)^t \mathbf{Z}_{ik}(s))$ sont presque sûrement minorées (uniformément en i). On en déduit donc que les $\|\widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}_0)\|_2$ sont bornés par une constante indépendante de i ce qui achève la démonstration de (A3). □

8.4 Extension à des processus

Dans notre modèle, nous avons considéré $\boldsymbol{\beta}_0$ comme étant une fonction dépendant du temps. Le théorème présenté à la section précédente ne permet pas *a priori* de récupérer toute l'information sur $\boldsymbol{\beta}_0$ puisque nous considérons des intégrales d'une fonction de $\boldsymbol{\beta}_0$ sur $[0, \tau]$. Ainsi, nous ne pouvons espérer que décrire le comportement moyen de $\boldsymbol{\beta}_0$ sur $[0, \tau]$. C'est pourquoi dans cette section nous étudions le rapport de vraisemblance empirique en faisant varier la borne supérieure dans l'intégrale, dans le but d'obtenir des régions de confiance « localisantes ». À cet effet, posons :

$$\widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}, t) := \sum_{k=1}^{K} \int_0^t \left(\mathbf{Z}_{ik}(s) \cdot \frac{g'(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))} - \mathbf{E}_k(s, \boldsymbol{\beta}) \right) \mathrm{d}N_{ik}(s),$$

$$\mathrm{EL}_n(\boldsymbol{\beta}, t) := \max \left\{ \prod_{i=1}^{n} np_i \ , \ \sum_{i=1}^{n} p_i \widehat{m}(\mathbf{Z}_i, \boldsymbol{\beta}, t) = 0, \ p \in \mathbb{S}_{n-1} \right\}.$$

Notons encore :

$$m_i(\mathbf{Z}_i, \boldsymbol{\beta}, t) := \sum_{k=1}^{K} \int_0^t \left(\mathbf{Z}_{ik}(s) \cdot \frac{g'(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}(s)^t \mathbf{Z}_{ik}(s))} - \mathbf{e}_k(s, \boldsymbol{\beta}) \right) \mathrm{d}N_{ik}(s),$$

$$\Gamma_{in}(\boldsymbol{\beta}, t) := \sum_{k=1}^{K} \int_0^t \big(\mathbf{e}_k(s, \boldsymbol{\beta}) - \mathbf{E}_k(s, \boldsymbol{\beta}) \big) \mathrm{d}N_{ik}(s).$$

Théorème 8.4

Soit Θ la sphère unité de $\mathbb{R}^K$ et soit $V(\cdot) := \mathbb{V}\mathrm{ar}\, m_i(\mathbf{Z}_i, \boldsymbol{\beta}_0, \cdot)$. Supposons que :

$$0 < \inf_{t\in[\tau_1,\tau]} \inf_{\theta\in\Theta} \mathbb{E}\left|\langle m_i(t), \theta\rangle\right| \leqslant \sup_{t\in[\tau_1,\tau]} \sup_{\theta\in\Theta} \left|\langle V(t)\theta, \theta\rangle\right| < +\infty.$$

Alors, il existe un processus centré G tel que :

$$-2\ln \mathrm{EL}_n(\boldsymbol{\beta}_0, \cdot) \rightsquigarrow G(\cdot)^t V^{-1}(\cdot) G(\cdot), \quad \text{lorsque } n \to +\infty.$$

Un corollaire immédiat de ce théorème est que si u_α désigne le quantile d'ordre $1-\alpha$ du $\sup_{t\in[\tau_1,\tau]} G(t)^t V(t)^{-1} G(t)$, alors, une région de confiance asymptotique d'ordre $1-\alpha$ pour $\boldsymbol{\beta}_0$ est :

$$R_\alpha = \left\{ \boldsymbol{\beta} \colon [\tau_1, \tau] \to \mathbb{R}^p \ : \ \sup_{t\in[\tau_1,\tau]} -2\ln \mathrm{EL}_n(\boldsymbol{\beta}, t) \leqslant u_\alpha \right\}.$$

Pour démontrer le théorème 8.4, nous allons vérifier que les hypothèses (A0)* à (A3)* du théorème de Hjort *et al.* (2004) sont vérifiées. Rappelons que :

(A0*) $\mathbb{P}\left(\exists t \in [\tau_1, \tau] : \ \mathrm{EL}_n(\boldsymbol{\beta}_0, t) = 0\right) \longrightarrow 0$, lorsque $n \to +\infty$;

(A1*) $\sqrt{n} \cdot M_n(\boldsymbol{\beta}_0, \cdot) \rightsquigarrow G(\cdot)$, lorsque $n \to +\infty$;

(A2*) $\sup_{t\in[\tau_1,\tau]} |S_n(\boldsymbol{\beta}_0, t) - V(t)| \xrightarrow{\mathbb{P}} 0$, lorsque $n \to +\infty$;

(A3*) $\max_{1\leqslant i\leqslant n} \sup_{t\in[\tau_1,\tau]} \|\widehat{m}_i(\mathbf{Z}_i, \boldsymbol{\beta}_0, t)\| = o_p(n^{1/2})$.

Démonstration de (A1*). Écrivons :

$$\sqrt{n} \cdot M_n(\boldsymbol{\beta}_0, \cdot) = \frac{1}{\sqrt{n}} \sum_{i=1}^n m_i(\mathbf{Z}_i, \boldsymbol{\beta}_0, \cdot) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \Gamma_{in}(\boldsymbol{\beta}_0, \cdot). \tag{8.9}$$

On sait que l'intégrande dans l'expression de $m_i(\mathbf{Z}_i, \boldsymbol{\beta}_0, \cdot)$ est à variations bornées : il est donc borné. De plus, les N_{ik} sont des processus de comptage bornés par B. Donc, les $m_i(\mathbf{Z}_i, \boldsymbol{\beta}_0, \cdot)$ sont à variations bornées. Posons :

$$\mathcal{F}_{n\omega}^k := \left\{ \big(m_{1k}(\mathbf{Z}_i, \boldsymbol{\beta}_0, t)(\omega), \ldots, m_{nk}(\mathbf{Z}_i, \boldsymbol{\beta}_0, t)(\omega)\big), \ t \in [\tau_1, \tau] \right\}.$$

Alors, d'après le théorème 2.28, les $\mathcal{F}_{n\omega}^k$ sont euclidiennes. De plus, on a vu page 116 que les $m_i(\mathbf{Z}_i, \boldsymbol{\beta}_0, \cdot)$ étaient centrés. On peut donc appliquer le théorème 2.32 de la limite centrale fonctionnel d'après la remarque 2.33 pour conclure que le premier terme dans (8.9) converge en distribution.

On a par ailleurs déjà montré page 117 la convergence en distribution du deuxième terme dans (8.9), ce qui démontre l'équicontinuité asymptotique de $\sqrt{n}\, M_n(\boldsymbol{\beta}_0, \cdot)$.

Enfin, la convergence des marges fini-dimensionnelles est une conséquence du théorème de la limite centrale multivarié classique. □

Démonstration de (A2*). On écrit encore :

$$\begin{aligned} S_n(\boldsymbol{\beta}_0,\cdot) = \frac{1}{n}\sum_{i=1}^{n} m_i(\mathbf{Z}_i,\boldsymbol{\beta}_0,\cdot)^{\otimes 2} + \frac{1}{n}\sum_{i=1}^{n} m_i(\mathbf{Z}_i,\boldsymbol{\beta}_0,\cdot)\Gamma_{in}(\boldsymbol{\beta}_0,\cdot)^t \\ + \frac{1}{n}\sum_{i=1}^{n} \Gamma_{in}(\boldsymbol{\beta}_0,\cdot)m_i(\mathbf{Z}_i,\boldsymbol{\beta}_0,\cdot)^t + \frac{1}{n}\sum_{i=1}^{n} \Gamma_{in}(\boldsymbol{\beta}_0,\cdot)^{\otimes 2}. \end{aligned} \tag{8.10}$$

Alors, en reprenant le raisonnement de l'étape (A1*), on peut appliquer la loi des grands nombres uniforme au premier terme dans (8.10) par stabilité des classes euclidiennes par multiplication.

On peut reprendre les arguments présentés page 118 pour démontrer que le dernier terme converge en probabilité vers 0 (uniformément en $t \in [\tau_1, \tau]$.

Quant aux termes croisés, on leur applique la même méthode que celle décrite page 118 pour démontrer qu'ils convergent eux aussi en probabilité vers 0 (uniformément en $t \in [\tau_1, \tau]$). □

Démonstration de (A3*). On a déjà démontré au point (A3) que les $\widehat{m}(\mathbf{Z}_i, \boldsymbol{\beta}_0)$ étaient uniformément bornés en i et en t ce qui démontre (A3*). □

Démonstration de (A0*). Il suffit de montrer que toutes les conditions présentées à la section 5.2 sont réunies. Or, on a déjà montré que les $\mathcal{F}^k_{n\omega}$ étaient euclidiennes. Il reste donc uniquement à vérifier que :

$$\max_{1\leqslant i\leqslant n} \|\widehat{m}_i(\mathbf{Z}_i,\boldsymbol{\beta}_0,\cdot) - m_i(\mathbf{Z}_i,\boldsymbol{\beta}_0,\cdot)\|_\infty \xrightarrow{\mathbb{P}} 0, \quad \text{lorsque } n \to +\infty.$$

Les arguments présentés au point (A0) page 119 permettent de conclure. □

8.5 Discussion

Comme nous l'avons dit précédemment, on peut donner une région de confiance asymptotique pour la fonction $\boldsymbol{\beta}_0$ donnée par :

$$R_\alpha = \left\{ \boldsymbol{\beta}\colon [\tau_1,\tau] \to \mathbb{R}^p \ : \ \sup_{t\in[\tau_1,\tau]} -2\ln \mathrm{EL}_n(\boldsymbol{\beta}, t) \leqslant u_\alpha \right\}.$$

Pour construire numériquement une telle région, on peut discrétiser l'intervalle $[\tau_1, \tau]$ puis chercher les fonctions $\boldsymbol{\beta}_0$ constantes par morceaux sur ces intervalles qui vérifient $-2\ln \mathrm{EL}_n(\boldsymbol{\beta}, t) \leqslant u_\alpha$. Admettons alors avoir choisi N_1 points dans l'intervalle $[\tau_1, \tau]$. Comme $\boldsymbol{\beta}$ est à valeurs dans $\mathbb{R}^p$, il nous faut aussi discrétiser un sous-ensemble borné bien choisi de $\mathbb{R}^p$: on obtiendra alors de l'ordre de N_2^p points. Pour déterminer une approximation de R_α, nous aurons donc au final de l'ordre de $N_1 N_2^p$ problèmes d'optimisation convexe à résoudre. Si N_1 et N_2 sont de l'ordre de 100 et que $p = 5$, cela représente 10^{12} problèmes d'optimisation convexe, ce qui est conséquent. En effet, supposons qu'un problème d'optimisation convexe soit réalisé en $0,1$ seconde, cela représente alors 105 années de calcul sur 30 cœurs en parallèle.

Ajoutons à cela que, *a priori*, la région de confiance R_α n'a aucune raison d'être convexe. Même si la fonction g est simple (par exemple si g est la fonction exponentielle), on ne peut pas simplifier l'expression de R_α.

CHAPITRE 9

Différentiabilité

9.1 Introduction

Supposons être en présence de variables aléatoires i.i.d. $X_1, \dots, X_n$ à valeurs dans un espace métrique $(\mathcal{X}, d)$ et de loi $P_0 \in \mathcal{P}$, où $\mathcal{P}$ est un ensemble de lois de probabilités. Soit T une fonction de $\mathcal{P}$ à valeurs dans un Banach $(\mathcal{B}, \|\cdot\|)$. Le problème que l'on se pose dans ce chapitre est celui de la construction d'une région de confiance pour $T(P_0)$ par vraisemblance empirique. Posons :

$$\mathcal{P}_{n-1} := \left\{ \widetilde{\mathbb{P}}_n := \sum_{i=1}^{n} p_i \delta_{X_i},\ p \in \mathbb{S}_{n-1} \right\}.$$

On peut alors s'intéresser à la variable aléatoire suivante :

$$\mathrm{EL}_n(T, P) := \max \left\{ \prod_{i=1}^{n} n p_i,\ T(\widetilde{\mathbb{P}}_n) = T(P_0) \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \right\},$$

où, pour alléger les notations, on a omis de préciser que le $\widetilde{\mathbb{P}}_n$ correspondait au vecteur $(p_1, \dots, p_n)$. Dans le cas où T est une application linéaire de rang fini, on peut réécrire EL_n comme :

$$\mathrm{EL}_n(T, P) = \max \left\{ \prod_{i=1}^{n} n p_i,\ \sum_{i=1}^{n} p_i T(\delta_{X_i}) = T(P_0) \text{ et } p \in \mathbb{S}_{n-1} \right\}.$$

On s'est alors ramené à un problème de vraisemblance empirique « classique » que l'on peut traiter grâce au théorème de Hjort *et al.* (2004). Cependant, lorsque T n'est pas une application linéaire, la démonstration du théorème de Hjort *et al.* (2004) ne fonctionne plus en général. En effet, pour prouver ce résultat, on a utilisé la séparabilité de la fonction de Lagrange qui est une conséquence de la linéarité de T.

Pour se ramener à une application linéaire, l'idée est alors d'étudier (si elle existe) la différentielle de T. Cette approche est déjà présente dans Owen (1990) lorsque $T = T(\mu)$ est une application différentiable de la moyenne μ. On la retrouve dans un cadre plus général dans l'article de Bertail (2006) que nous allons détailler et généraliser ici.

Dans tout ce qui suit, nous supposerons que $\mathcal{P}$ est un ensemble convexe de mesures signées contenant les mesures de Dirac et P_0. Nous noterons $\mathcal{F}$ une classe de fonctions de $\mathcal{X}$ dans $\mathbb{R}$ et munirons $\mathcal{P}$ de la semi-norme $\|\cdot\|_{\mathcal{F}}$ définie par :

$$\|P - Q\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \left| \int f \mathrm{d}P - \int f \mathrm{d}Q \right|.$$

Nous considérerons $\mathcal{P}$ comme un sous-ensemble de $\ell_\infty(\mathcal{F})$ grâce à l'identification présentée à la section 2.11 :

$$\forall P \in \mathcal{P}, \forall f \in \mathcal{F} \ : \quad P(f) = \int f \mathrm{d}P.$$

La fonctionnelle T sera donc considérée comme une application dont l'espace de départ est l'ensemble des trajectoires $\left\{ f \mapsto \int f \mathrm{d}P,\ P \in \mathcal{P} \right\}$. Nous noterons encore $T(P)$ les images de T pour alléger les notations. Définissons :

$$\overline{\mathcal{P}_{n,1-\alpha}} := \left\{ \widetilde{\mathbb{P}}_n = \sum_{i=1}^n p_{i,n}\delta_{X_i},\ \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \text{ et } \prod_{i=1}^n np_{i,n} \geqslant u_\alpha \right\},$$

où u_α est fixé et correspondra plus tard à $\exp(-\chi^2_{1-\alpha}(q)/2)$. Nous supposerons de plus que $\mathcal{F}$ est une classe de fonctions admettant pour enveloppe H vérifiant :

$$0 < \int H^2 \mathrm{d}P_0 < \infty. \tag{9.1}$$

Nous supposerons aussi la condition d'uniforme entropie suivante :

$$\int_0^{+\infty} \sup_{Q\in\mathcal{D}} \sqrt{\ln(\mathcal{N}(\varepsilon\|H\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}))}\mathrm{d}\varepsilon < \infty, \tag{9.2}$$

où $\|f\|_{2,Q} = (\mathbb{E}_Q(f^2))^{1/2}$ et où $\mathcal{D}$ désigne l'ensemble des mesures de probabilité discrètes telles que $0 < \int H^2 \mathrm{d}Q$.

Sous ces conditions, on sait (voir section 2.11 page 42) que $\sqrt{n}(\mathbb{P}_n - P_0)$ converge en distribution (dans $\ell_\infty(\mathcal{F})$) vers un processus gaussien G. Dans son article, Bertail (2006) remarque que les $\widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}}$ sont proches de $\mathbb{P}_n$ et démontre qu'il y a encore convergence du processus empirique pondéré renormalisé vers G, et ce, de manière uniforme sur $\overline{\mathcal{P}_{n,1-\alpha}}$. Plus précisément, rappelons ici son résultat.

Lemme 9.1 (Bertail (2006))**.** *Soit $\alpha \in [0,1[$. Alors, il existe des constantes strictement positives $a(\alpha) < 1 < b(\alpha)$ telles que :*

$$\forall \widetilde{\mathbb{P}}_n = \sum_{i=1}^n p_{i,n}\delta_{X_i} \in \overline{\mathcal{P}_{n,1-\alpha}} \ : \quad \frac{a(\alpha)}{n} \leqslant p_{i,n} \leqslant \frac{b(\alpha)}{n}.$$

De plus, si $\alpha \in]0,1[$ *et que* $\mathcal{F}$ *vérifie (9.1) et (9.2), alors :*

$$\left(\sum_{i=1}^{n} p_{i,n}^2\right)^{-1/2} \left(\widetilde{\mathbb{P}}_n - P_0\right) \rightsquigarrow G, \tag{9.3}$$

et ce, uniformément en $\widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}}$.

Remarque 9.2. Nous n'avons pas défini formellement la convergence en loi uniforme au chapitre 1. Notons $\mathbb{S}_{n-1}^{1-\alpha}$ les éléments du simplexe unité de $\mathbb{R}^n$ dont le produit des coordonnées dépasse u_α. Notons encore $\widetilde{\mathbb{G}}_{n,p}$ le processus empirique pondéré correspondant pour tout p de $\mathbb{S}_{n-1}^{1-\alpha}$ (défini par (9.3)). Alors, le deuxième point du lemme précédent signifie que, lorsque $n \to +\infty$:

$$\forall \phi \in \mathscr{C}_b(\ell_\infty(\mathcal{F}), \mathbb{R}) \ : \quad \sup_{p \in \mathbb{S}_{n-1}^{1-\alpha}} \left| \mathbb{E}^\star \phi(\widetilde{\mathbb{G}}_{n,p}) - \mathbb{E}\, \phi(G) \right| \longrightarrow 0.$$

Ce lemme, particulièrement important pour la suite de ce chapitre, nous permettra de négliger des quantités de la forme $\|\widetilde{\mathbb{P}}_n - P_0\|_{\mathcal{F}}$ qui seront alors des $O_P(n^{-1/2})$. Nous pourrons aussi négliger les quantités de la forme $T(\widetilde{\mathbb{P}}_n) - T(P_0)$ lorsque T sera différentiable.

9.2 Idées et heuristique de preuve

Comme nous l'avons vu à la section précédente, on ne peut pas en général appliquer directement le théorème de Hjort *et al.* (2004) dans le cas où la fonctionnelle T n'est pas linéaire. Supposons alors que T est différentiable en P_0 au sens usuel et posons :

$$\mathrm{EL}_n^L(T, P) := \max \left\{ \prod_{i=1}^{n} n p_i, \ \mathrm{d}T(P_0).(\widetilde{\mathbb{P}}_n - P) = 0 \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \right\}.$$

Si, par exemple, T était à valeurs dans $\mathcal{D}([0,\tau])$, alors, sous les conditions du théorème de Hjort *et al.* (2004) on pourrait montrer la convergence en distribution de $\mathrm{EL}_n^L(T, P_0)$. Ainsi, une région de confiance asymptotique pour $T(P_0)$ serait donnée par :

$$R_{n,1-\alpha} := \left\{ T(P_0) + \mathrm{d}T(P_0).(\widetilde{\mathbb{P}}_n - P_0), \ \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}} \right\}.$$

Comme on ne connaît pas P_0, l'idée est d'écrire un développement de Taylor de T au voisinage de P_0 :

$$T(\widetilde{\mathbb{P}}_n) = T(P_0) + \mathrm{d}T(P_0).(\widetilde{\mathbb{P}}_n - P_0) + R_n(\widetilde{\mathbb{P}}_n, P_0).$$

On s'attend alors à ce que $R_n(\widetilde{\mathbb{P}}_n, P_0) = O_P(n^{-1/2})$ en vertu du lemme 9.1. Ainsi peut-on proposer une autre région de confiance asymptotique pour $T(P_0)$:

$$R_{n,1-\alpha}^T := \left\{ T(\widetilde{\mathbb{P}}_n), \ \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}} \right\}.$$

Cependant, plusieurs problèmes se posent. Premièrement, $R^T_{n,1-\alpha}$ n'a aucune raison d'être convexe. Par continuité de T, on peut seulement conclure à la connexité et il peut être difficile d'interpréter le résultat obtenu. Pour l'illustrer, nous avons simulé cinq réalisations d'une loi normale centrée réduite et avons tracé sur la figure 9.1 la région $R^{T_{ev}}_{n,1-\alpha}$ pour la fonctionnelle T_{ev} définie par :

$$T_{ev}(P) = \left(\int X \mathrm{d}P, \int X^2 \mathrm{d}P - \left(\int X \mathrm{d}P \right)^2 \right)^t .$$

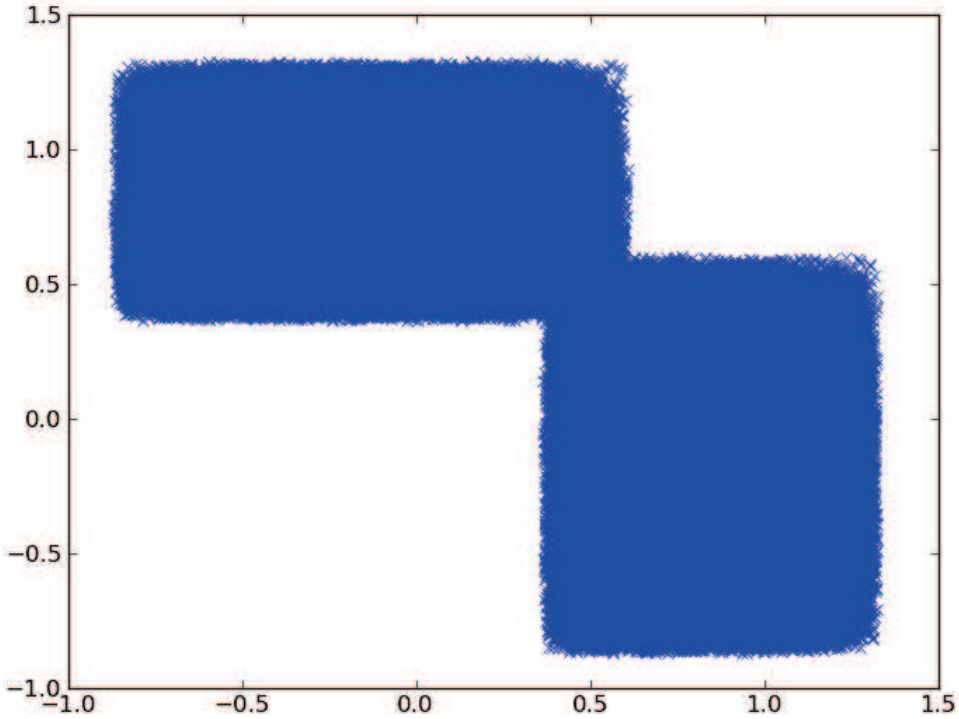


FIGURE 9.1 – Région de confiance $R^{T_{ev}}_{n,1-\alpha}$ pour la moyenne et la variance.

Bien sûr, comme $R^T_{n,1-\alpha}$ et $R_{n,1-\alpha}$ sont proches à un $O_P(n^{-1/2})$ près, la région $R^T_{n,1-\alpha}$ est « asymptotiquement convexe » (car $R_{n,1-\alpha}$ l'est en tant que transformation linéaire d'un ensemble convexe). De plus, si $\mathrm{d}T$ était continue en P_0, alors, on pourrait s'intéresser à la région de confiance asymptotique définie par :

$$R^C_{n,1-\alpha} = \left\{ T(\mathbb{P}_n) + \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - \mathbb{P}_n),\ \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}} \right\},$$

qui n'est autre que $R_{n,1-\alpha}$ où P_0 a été substituée par $\mathbb{P}_n$. Cette région est quant à elle convexe. Nous l'avons aussi tracée avec les mêmes données à la figure 9.2. Remarquons que le vrai paramètre (ici $(0,1)$) n'est pas dans la région de confiance, mais cela est sans doute dû au fait que notre échantillon est très petit ($n = 5$).

Cela nous amène à l'autre problème posé par ces régions de confiance asymptotiques. Si nous avons utilisé un échantillon si petit, c'est parce qu'il est très long, numériquement, de décrire l'ensemble $\overline{\mathcal{P}_{n,1-\alpha}}$. Rappelons qu'il

faut à cet effet trouver tous les vecteurs de probabilité de $\mathbb{R}^n$ dont le produit des composantes dépasse un certain seuil. Ainsi, pour le décrire, il faut effectuer $n-1$ boucles imbriquées sur une grille suffisamment fine ce qui prend énormément de temps !

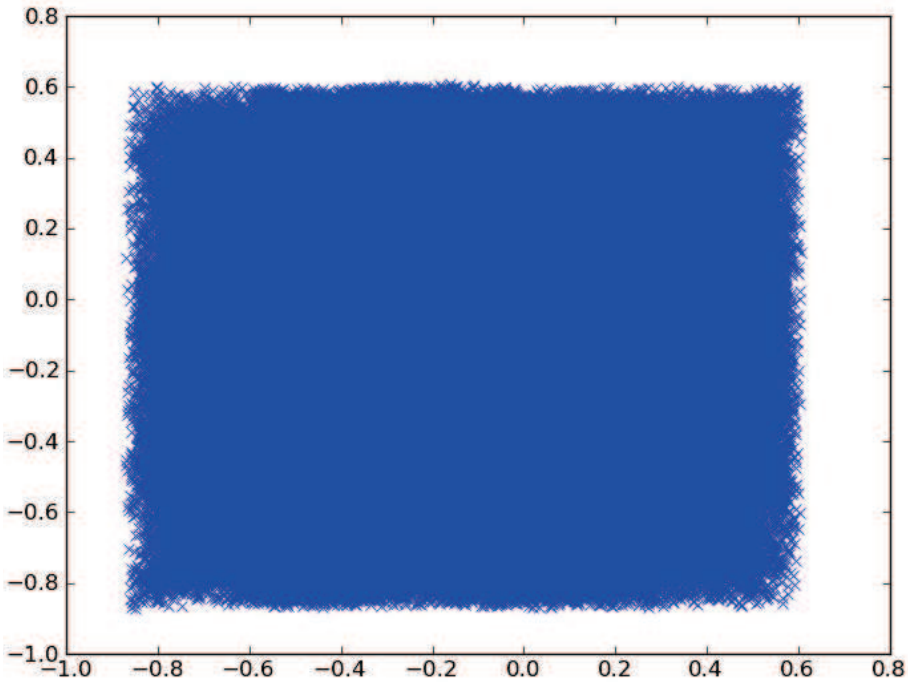


FIGURE 9.2 – Région de confiance $R^C_{n,1-\alpha}$ pour la moyenne et la variance.

Pour palier ce problème, nous proposons ici de considérer le rapport de vraisemblance empirique suivant :

$$\mathrm{EL}^{LL}_n(T,P) := \max\left\{\prod_{i=1}^n np_i,\ \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0 \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1}\right\},$$

ainsi que la région

$$R^L_{n,1-\alpha} := \left\{T(\mathbb{P}_n) + \mathrm{d}T(\mathbb{P}_n).(P-\mathbb{P}_n),\ \mathrm{EL}^{LL}_n(T,P) \geqslant u_\alpha\right\}.$$

Comme précédemment, sous certaines conditions de continuité, on montrera alors que $R_{n,1-\alpha}$ et $R^L_{n,1-\alpha}$ sont proches à un $O_P(n^{-1/2})$ près. Cette méthode a l'avantage de donner une région convexe. De plus, elle est numériquement viable comme nous le verrons par la suite, et ce, quelle que soit la taille de l'échantillon. Comme précédemment, nous avons tracé $R^L_{n,1-\alpha}$ pour T_{ev} avec les mêmes données (cf. courbe bleue sur la figure figure 9.3).

Avant de conclure cette section, il nous faut remarquer que le problème de la construction d'une région de confiance par vraisemblance empirique pour la moyenne et la variance est en fait très simple à résoudre avec le théorème d'Owen (ce qui n'aurait pas été le cas si on avait considéré la variance seule). En effet, il suffit de définir :

$$\mathrm{EL}^L_n(\mu,\sigma^2) = \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i\begin{pmatrix} X_i \\ X_i^2\end{pmatrix} = \begin{pmatrix}\mu \\ \sigma^2+\mu^2\end{pmatrix} \text{ et } p\in\mathbb{S}_{n-1}\right\}.$$

Cependant, la région obtenue avec cette méthode (en rouge sur la figure 9.3) se trouve être la même que $R^L_{n,1-\alpha}$ (en bleu sur la même figure) à un changement de repère affine près. De plus, les régions sont asymptotiquement équivalentes, à un $O_P(n^{-1/2})$ près.

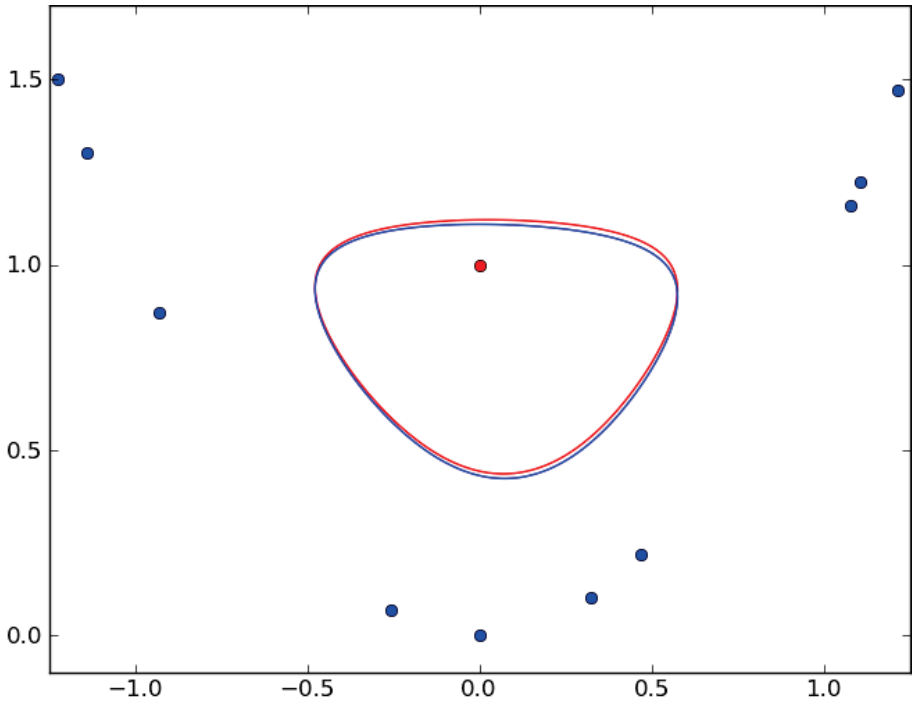


FIGURE 9.3 – Comparaison entre $R^L_{n,1-\alpha}$ pour la moyenne et la variance et la région obtenue directement avec le théorème d'Owen.

9.3 Hadamard-différentiabilité

Nous avons présenté à la section précédente les idées permettant de travailler avec des fonctionnelles différentiables. Cependant, la différentiabilité usuelle (au sens de Fréchet) n'est pas vérifiée dans beaucoup de situations. C'est pourquoi nous allons considérer des fonctions *Hadamard-différentiables*.

Définition 9.3 (Hadamard-différentiabilité)**.** Soit A et B deux espaces vectoriels munis de semi-normes. On dira que $T : A_T \subset A \to B$ est Hadamard-différentiable en a tangentiellement à $A_0 \subset A$ si il existe une application linéaire continue, notée $\mathrm{d}T(a)$ et définie sur A_0, telle que pour toute suite $z_n \to z \in A_0$ et pour toute suite $t_n \to 0$ vérifiant $a + t_n z_n \in A_T$, on ait :

$$\frac{1}{t_n}\Big(T(a + t_n z_n) - T(a)\Big) - \mathrm{d}T(a).z \longrightarrow 0 \quad \text{lorsque } n \to +\infty.$$

Dans la suite, on s'intéressera à des fonctions T qui seront Hadamard-différentiables en $P_0 \in \ell_\infty(\mathcal{F})$, le vrai paramètre, tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$ défini à la section 2.11 et qui vérifie que $\mathbb{P}(G \in \mathcal{B}(\mathcal{F}, P_0)) = 1$, où G est le processus limite de $\sqrt{n}(\mathbb{P}_n - P_0)$.

La notion de Hadamard-différentiabilité est stable par composition, ce qui simplifiera grandement les calculs que nous effectuerons par la suite.

Lemme 9.4. *Soit $T : E_T \subset E \to F$ une fonction Hadamard-différentiable en x_0 tangentiellement à $E_0 \subset E$ et $S : F_S \subset F \to G$ Hadamard-différentiable en $T(x_0)$ tangentiellement à $\mathrm{d}T(E_0)$. Alors, $S \circ T : E_T \to G$ est Hadamard-différentiable en x_0 tangentiellement à E_0 et :*

$$\mathrm{d}(S \circ T)(x_0) = \mathrm{d}S(T(x_0)) \circ \mathrm{d}T(x_0).$$

Rappelons que nous avions défini dans la section précédente les quantités suivantes :

$$\mathrm{EL}_n^L(T, P, \cdot) = \max \left\{ \prod_{i=1}^n np_i,\ \mathrm{d}T(P_0).(\widetilde{\mathbb{P}}_n - P) = 0 \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \right\},$$
$$\mathrm{EL}_n^{LL}(T, P, \cdot) = \max \left\{ \prod_{i=1}^n np_i,\ \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0 \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \right\}.$$

Or, si T est Hadamard-différentiable en P_0 (ou en $\mathbb{P}_n$ pour EL_n^{LL}) tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$, on ne peut considérer que des quantités du type $\mathrm{d}T(P_0).Q$ (ou $\mathrm{d}T(\mathbb{P}_n).Q$ pour EL_n^{LL}) avec $Q \in \mathcal{B}(\mathcal{F}, P_0)$. Ajoutons que, *a priori*, les $(\widetilde{\mathbb{P}}_n - P_0)$ ne sont pas des éléments de $\mathcal{B}(\mathcal{F}, P_0)$. Donc, pour que ces quantités aient un sens, nous supposerons que T est Gâteaux-différentiable en P_0 suivant les directions δ_x, $x \in \mathcal{X}$. Nous poserons alors pour tout $P \in \mathcal{B}(\mathcal{F}, P_0)$:

$$\mathrm{d}T(P_0).(\widetilde{\mathbb{P}}_n - P) = \mathrm{d}T(P_0).\widetilde{\mathbb{P}}_n - \mathrm{d}T(P_0).P,$$

où la première différentielle sera celle au sens de Gâteaux et la deuxième celle au sens de Hadamard. Remarquons que si T est différentiable en un point P au sens de Hadamard et au sens de Gâteaux, alors, les deux différentielles sont les mêmes ce qui justifie l'utilisation de la même notation. Ajoutons que comme $P_0 \in \mathcal{B}(\mathcal{F}, P_0)$, nous pourrons considérer les quantités $\mathrm{d}T(P_0).(\widetilde{\mathbb{P}}_n - P_0)$.

Définition 9.5 (Gâteaux-différentiabilité)**.** Une fonction $T : A_T \subset A \to B$ est Gâteaux-différentiable en a_0 suivant la direction $a \in A$ si il existe une application linéaire continue de A dans B, notée $\mathrm{d}T(a_0)$, telle que :

$$\frac{1}{t}\Big(T(a_0 + ta) - T(a_0)\Big) - \mathrm{d}T(a_0).a \longrightarrow 0, \quad \text{lorsque } t \to 0.$$

9.4 Résultats

9.4.1 Vraisemblance empirique

Dans la continuité de ce qui a été présenté aux sections précédentes, nous considérerons une suite de variables aléatoires i.i.d. $(X_n)_{n \in \mathbb{N}^\star}$ à valeurs

dans un espace métrique $(\mathcal{X}, d)$. Nous noterons $\mathcal{F}$ une classe de fonctions de $\mathcal{X}$ dans $\mathbb{R}$ et $\ell_\infty(\mathcal{F})$ l'ensemble des fonctions de $\mathcal{F}$ dans $\mathbb{R}$ qui sont bornées pour la norme infinie définie par :

$$\forall \psi \in \ell_\infty(\mathcal{F}) \ : \quad \|\psi\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\psi(f)|.$$

Nous noterons $\mathcal{P}$ un ensemble de mesures signées sur $(\mathcal{X}, d)$ muni de la tribu borélienne. Nous supposerons que $\mathcal{P}$ contient P_0, $\mathbb{P}_n$ ainsi que les mesures $P_0 + u\delta_x$ et $\mathbb{P}_n + u\delta_x$ pour u suffisamment petit et $x \in \mathcal{X}$. Les éléments de $\mathcal{P}$ induisant des éléments de $\ell_\infty(\mathcal{F})$, nous identifierons $\mathcal{P}$ à un sous-ensemble de $\ell_\infty(\mathcal{F})$ et nous considérerons une fonctionnelle :

$$T : \mathcal{P} \subset \ell_\infty(\mathcal{F}) \longrightarrow \mathcal{D}([0, \tau])^k,$$

où $k \in \mathbb{N}^\star$. Nous utiliserons les mêmes normes sur $\mathbb{R}^k$ et $M_k(\mathbb{R})$ que celles définies au chapitre 8, à savoir celles définies par :

$$\forall x = (x_1, \dots, x_k)^t \in \mathbb{R}^k, \quad \|x\| := \sup_{1 \leqslant i \leqslant k} |x_i|,$$
$$\forall A = (a_{ij})_{ij} \in M_k(\mathbb{R}), \qquad \|A\| := \sup_{1 \leqslant i,j \leqslant k} |a_{i,j}|.$$

Nous supposerons de plus que :

(i) les conditions (9.1) et (9.2) sont satisfaites ;

(ii) T est Hadamard-différentiable en P_0 tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$;

(iii) T est Gâteaux-différentiable en P_0 suivant les δ_x, $x \in \mathcal{X}$;

(iv) $\mathrm{d}T : \mathrm{Vect}(\{\delta_x, x \in \mathcal{X}\} \cup \mathcal{B}(\mathcal{F}, P_0)) \to \mathcal{D}([0, \tau])^k$ est continue en P_0 pour la norme définie par :

$$|||\mathrm{d}T(\psi)|||_{\mathcal{F}} := \sup_{\|\psi\|_{\mathcal{F}} \leqslant 1} \sup_{t \in [0,\tau]} \|(\mathrm{d}T(P).\psi)(t)\|.$$

Nous écrirons encore pour tout $f = (f_1, \dots, f_k) \in \mathcal{D}([0, \tau])^k$:

$$\|f\|_\infty = \sup_{t \in [0,\tau]} \|f(t)\| = \sup_{t \in [0,\tau]} \max_{1 \leqslant i \leqslant k} |f_i(t)|.$$

Définissons maintenant pour tout $P \in \mathcal{B}(\mathcal{F}, P_0)$:

$$\mathrm{EL}_n^{LL}(T, P, \cdot) := \max \left\{ \prod_{i=1}^n np_i, \ \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0 \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \right\}.$$

À partir d'un certain rang, cette quantité a un sens puisque $\mathrm{d}T$ est continue en P_0, que $\mathbb{P}_n \to P_0$ lorsque $n \to +\infty$ et que T est différentiable en P_0. Nous allons maintenant appliquer le théorème de Hjort *et al.* (2004) pour démontrer la convergence en distribution de EL_n^{LL}. À cet effet, définissons :

$$\widehat{m}_i(X_i, T, P, \cdot) := \mathrm{d}T(\mathbb{P}_n).(\delta_{X_i} - P),$$
$$m_i(X_i, T, P, \cdot) := \mathrm{d}T(P_0).(\delta_{X_i} - P).$$

Théorème 9.6

Soit $V(\cdot) := \mathbb{V}\mathrm{ar}\, m_i(X_i, T, P, \cdot)$. Supposons que les m_i sont séparables et qu'il existe un réel τ_1 tel que, lorsque $n \to +\infty$:

$$\text{(A0}^\star\text{)} \qquad \mathbb{P}\left(\exists t \in [\tau_1, \tau] : \ \mathrm{EL}_n^{LL}(T, P_0, t) = 0\right) \longrightarrow 0;$$

$$\text{(B2}^\star\text{)} \qquad \sup_{t\in[\tau_1,\tau]} \left| \frac{1}{n}\sum_{i=1}^{n} m_i(X_i, T, P_0, t)^{\otimes 2} - V(t) \right| \xrightarrow{\mathbb{P}^\star} 0.$$

Supposons de plus qu'il existe deux réels strictement positifs δ et M tels que les valeurs propres non nulles de V sont uniformément bornées dans l'intervalle $[\delta, M]$ sur $[\tau_1, \tau]$. Alors, il existe un processus gaussien centré $G(\cdot)$ sur $[\tau_1, \tau]$ tel que :

$$-2 \ln \mathrm{EL}_n^{LL}(T, P_0, \cdot) \rightsquigarrow G(\cdot) V^\dagger(\cdot) G(\cdot), \quad \text{lorsque } n \to +\infty.$$

Rappelons que :

(A0*) $\mathbb{P}\left(\exists t \in [\tau_1, \tau] : \ \mathrm{EL}_n^{LL}(T, P_0, t) = 0\right) \longrightarrow 0$, lorsque $n \to +\infty$;

(A1*) $\sqrt{n} \cdot M_n(T, P_0, \cdot) \rightsquigarrow G(\cdot)$, lorsque $n \to +\infty$;

(A2*) $\sup_{t\in[\tau_1,\tau]} \|S_n(T, P_0, t) - V(t)\| \xrightarrow{\mathbb{P}^\star} 0$, lorsque $n \to +\infty$;

(A3*) $\max_{1\leqslant i\leqslant n} \sup_{t\in[\tau_1,\tau]} \|\widehat{m}(X_i, T, P_0, t)\| = o_p(n^{1/2})$.

Démonstration de (A1*). On sait qu'il existe un processus gaussien centré Z tel que :

$$\sqrt{n}(\mathbb{P}_n - P_0) \rightsquigarrow Z,$$

où Z est presque sûrement un élément de $\mathcal{B}(\mathcal{F}, P_0)$. Donc, via un transport par $\mathrm{d}T(P_0)$ qui est une application continue et par linéarité, on a aussi :

$$\sqrt{n}\left(\mathrm{d}T(P_0).\mathbb{P}_n - \mathrm{d}T(P_0).P_0\right) \rightsquigarrow \mathrm{d}T(P_0).Z.$$

De plus :

$$\begin{aligned}\sqrt{n} M_n(T, P_0, \cdot) &= \sqrt{n}\, \mathrm{d}T(\mathbb{P}_n).(\mathbb{P}_n - P_0) \\ &= \left(\mathrm{d}T(\mathbb{P}_n) - \mathrm{d}T(P_0)\right).\left(\sqrt{n}(\mathbb{P}_n - P_0)\right) + \mathrm{d}T(P_0).\left(\sqrt{n}(\mathbb{P}_n - P_0)\right).\end{aligned}$$

Le deuxième terme dans l'équation ci-dessus converge en distribution vers $G := \mathrm{d}T(P_0).Z$. Nous allons montrer que le premier terme converge vers 0 en probabilité extérieure. Soit $\varepsilon > 0$ fixé. Comme $(\sqrt{n}(\mathbb{P}_n - P_0))_{n\in\mathbb{N}^\star}$ converge en distribution vers Z qui est séparable, la suite est asymptotiquement ten-

due [1]. Donc, il existe M_ε ainsi que $n_1 \in \mathbb{N}^\star$ tels que :

$$\forall n \geqslant n_1 \ : \quad \mathbb{P}^\star\Big(\|\sqrt{n}(\mathbb{P}_n - P_0)\|_{\mathcal{F}} \geqslant M_\varepsilon\Big) \leqslant \frac{\varepsilon}{2}.$$

De plus, $\mathbb{P}_n \to P_0$ en probabilité dans $\ell_\infty(\mathcal{F})$. Donc, par continuité de $\mathrm{d}T$ en P_0, il existe $n_2 \in \mathbb{N}^\star$ tel que :

$$\forall n \geqslant n_2 \ : \quad \mathbb{P}^\star\Big(\|\mathrm{d}T(\mathbb{P}_n) - \mathrm{d}T(P_0)\|_{\mathcal{F}} \geqslant \frac{\varepsilon}{M_\varepsilon}\Big) \leqslant \frac{\varepsilon}{2}.$$

Or :

$$\big\|\big(\mathrm{d}T(\mathbb{P}_n) - \mathrm{d}T(P_0)\big).\big(\sqrt{n}(\mathbb{P}_n - P_0)\big)\big\|_\infty \leqslant \|\mathrm{d}T(\mathbb{P}_n) - \mathrm{d}T(P_0)\|_{\mathcal{F}} \cdot \|\sqrt{n}(\mathbb{P}_n - P_0)\|_{\mathcal{F}}$$

Donc, en découpant le membre de droite dans l'équation ci-dessus suivant si $\|\sqrt{n}(\mathbb{P}_n - P_0)\|_{\mathcal{F}}$ est plus petit ou plus grand que M_ε, on montre que pour tout $n \geqslant \max(n_1, n_2)$:

$$\mathbb{P}^\star\Big(\big\|\big(\mathrm{d}T(\mathbb{P}_n) - \mathrm{d}T(P_0)\big).\big(\sqrt{n}(\mathbb{P}_n - P_0)\big)\big\|_\infty \geqslant \varepsilon\Big) \leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

Démonstration de (A2*). Rappelons que :

$$S_n(T, P_0, \cdot) = \frac{1}{n}\sum_{i=1}^n \widehat{m}_i(X_i, T, P_0, \cdot)^{\otimes 2}.$$

On peut donc réécrire S_n de cette manière :

$$\begin{aligned} S_n(T, P_0, \cdot) = &\frac{1}{n}\sum_{i=1}^n \big(\widehat{m}_i(X_i, T, P_0, \cdot) - m_i(X_i, T, P_0, \cdot)\big)^{\otimes 2} \\ &+ \frac{1}{n}\sum_{i=1}^n \big(\widehat{m}_i(X_i, T, P_0, \cdot) - m_i(X_i, T, P_0)\big) m_i(X_i, T, P_0, \cdot)^t \\ &+ \frac{1}{n}\sum_{i=1}^n m_i(X_i, T, P_0, \cdot)\big(\widehat{m}_i(X_i, T, P_0, \cdot) - m_i(X_i, T, P_0)\big)^t \\ &+ \frac{1}{n}\sum_{i=1}^n m_i(X_i, T, P_0, \cdot)^{\otimes 2}. \end{aligned} \tag{9.4}$$

Première étape : convergence du premier terme. Rappelons que par définition, le premier terme dans (9.4) s'écrit :

$$\frac{1}{n}\sum_{i=1}^n \Big(\big(\mathrm{d}T(P_0) - \mathrm{d}T(\mathbb{P}_n)\big).(\delta_{X_i} - P_0)\Big)^{\otimes 2}.$$

1. $(X_n)_{n\in\mathbb{N}^\star}$ est asymptotiquement tendue si pour tout $\varepsilon > 0$, il existe un compact K tel que pour tout $\delta > 0$, $\overline{\lim}\, \mathbb{P}^\star(X_n \notin K^\delta) \leqslant \varepsilon$, où $K^\delta := \{y : d(y, K) < \delta\}$. On se reportera à van der Vaart & Wellner (1996) page 21 pour une démonstration de ce résultat. Notons que ce résultat est essentiel pour montrer l'équicontinuité asymptotique d'une suite convergente.

Or :

$$\|\delta_{X_i} - P_0\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \left| f(X_i) - \int f \mathrm{d}P_0 \right| \leqslant H(X_i) + \int H \mathrm{d}P_0.$$

De plus, par continuité de $\mathrm{d}T$ en P_0, il existe $n_0 \in \mathbb{N}^\star$ tel que :

$$\forall n \geqslant n_0 \ : \quad \|\mathrm{d}T(\mathbb{P}_n) - \mathrm{d}T(P_0)\|_{\mathcal{F}} \leqslant \varepsilon.$$

Ainsi, pour tout $n \geqslant n_0$, presque sûrement [2] :

$$\frac{1}{n}\sum_{i=1}^{n} \left\| \Big(\big(\mathrm{d}T(P_0) - \mathrm{d}T(\mathbb{P}_n)\big).(\delta_{X_i} - P_0) \Big)^{\otimes 2} \right\|_\infty \leqslant \varepsilon \frac{1}{n}\sum_{i=1}^{n} \left(H(X_i) + \int H \mathrm{d}P_0 \right)^2,$$

d'où le résultat en utilisant la loi des grands nombres.

Deuxième étape : convergence des termes croisés. Nous allons reprendre ici les arguments présentés page 118 pour démontrer la convergence en probabilité vers 0 des deux termes du milieu dans (9.4). Il s'agit simplement d'appliquer l'inégalité (8.7) puis l'inégalité de Cauchy-Schwarz. Écrivons alors $m_i := m_i(X_i, T, P_0, \cdot)$ et $\widehat{m}_i := \widehat{m}_i(X_i, T, P_0, \cdot)$ pour simplifier les notations et considérons par exemple le premier des deux termes croisés. On a :

$$\begin{aligned} \left\| \frac{1}{n}\sum_{i=1}^{n} (\widehat{m}_i - m_i) m_i^t \right\|_\infty &\leqslant \frac{1}{n}\sum_{i=1}^{n} \|\widehat{m}_i - m_i\|_\infty \|m_i\|_\infty \\ &\leqslant \left(\frac{1}{n}\sum_{i=1}^{n} \|\widehat{m}_i - m_i\|_\infty^2 \right)^{1/2} \left(\frac{1}{n}\sum_{i=1}^{n} \|m_i\|_\infty^2 \right)^{1/2}. \end{aligned} \tag{9.5}$$

On obtient la même majoration pour le deuxième terme. De plus, on a montré à l'étape précédente que la première somme dans (9.5) convergeait vers 0 en probabilité (uniformément en t). Quant au deuxième, il est par hypothèse borné en probabilité extérieure, et ce, uniformément en t.

Troisième étape : convergence du troisième terme. On sait, par hypothèse sur les m_i, que :

$$\sup_{t\in[\tau_1,\tau]} \left\| \frac{1}{n}\sum_{i=1}^{n} m_i(X_i, T, P_0, t)^{\otimes 2} - V(t) \right\| \xrightarrow{\mathbb{P}} 0.$$

□

Démonstration de (A3*). Remarquons que pour tout $t \in [\tau_1, \tau]$:

$$\|\widehat{m}_i(X_i, T, P_0, t)\| \leqslant \|\widehat{m}_i(X_i, T, P_0, t) - m_i(X_i, T, P_0, t)\| + \|m_i(X_i, T, P_0, t)\|. \tag{9.6}$$

Or, les m_i étant des processus séparables, les $\sup_{t\in[\tau_1,\tau]} |m_i(t)|$ sont des variables aléatoires, et par hypothèse elles sont i.i.d. et dans L^2. On peut donc

2. Cela résulte de l'inégalité élémentaire (8.7) donnée page 118 ainsi que des propriétés de la norme subordonnée.

leur appliquer le lemme 3.9 pour montrer que le deuxième terme dans (9.6) est un $o_p(n^{1/2})$. Le premier terme est quant à lui presque sûrement inférieur à :

$$\|\mathrm{d}T(\mathbb{P}_n) - \mathrm{d}T(P_0)\|_{\mathcal{F}} \left(H(X_i) + \int H \mathrm{d}P_0 \right),$$

qui est aussi un $o_p(n^{1/2})$ par continuité de $\mathrm{d}T$ en P_0 et en vertu du lemme 3.9. □

Remarque 9.7. Dans la pratique, on pourra utiliser les résultats présentés au chapitre 5 pour démontrer (A0*). Notons que certaines des conditions suffisantes qui y sont énoncées sont automatiquement vérifiées. En effet :

$$\begin{aligned} \|\mathbb{E}\, \widehat{m}_i(X_i, T, P_0, \cdot) - \mathbb{E}\, m_i(X_i, T, P_0, \cdot)\|_\infty &= \|\mathbb{E}\, \mathrm{d}T(\mathbb{P}_n).\delta_{X_i} - \mathbb{E}\, \mathrm{d}T(P_0).P_0\|_\infty \\ &= \|\mathrm{d}T(\mathbb{P}_n).P_0 - \mathrm{d}T(P_0).P_0\|_\infty, \end{aligned}$$

qui tend vers 0 lorsque $n \to +\infty$ par continuité de la différentielle en T_0. De plus, par linéarité de la différentielle, on a aussi $\mathbb{E}\, m_i(X_i, T, P_0) = 0$.

Remarque 9.8. Dans les applications, pour montrer que l'application T est Hadamard-différentiable en P_0 tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$, nous considérerons un élément $\psi \in \mathcal{B}(\mathcal{F}, P_0)$, une suite $(\psi_n)_{n\in\mathbb{N}^\star}$ d'éléments de $\ell_\infty(\mathcal{F})$ convergeant vers ψ, ainsi qu'une suite $(t_n)_{n\in\mathbb{N}^\star}$ de réels positifs tendant vers 0. Nous supposerons de plus que pour tout $n \in \mathbb{N}^\star$, $P_0 + t_n\psi_n \in \mathcal{P}$. Ainsi, par stabilité par combinaisons linéaires de $\mathcal{P}$, nous pourrons identifier les ψ_n à des mesures signées.

9.4.2 Construction de régions de confiance

Résultats

D'après le théorème 9.6, une région de confiance asymptotique pour $\mathrm{d}T(\mathbb{P}_n).P_0$ est donnée par :

$$\begin{aligned} S_{n,1-\alpha} &= \left\{ \mathrm{d}T(\mathbb{P}_n).\widetilde{\mathbb{P}}_n,\ \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}} \right\} \\ &= \left\{ \mathrm{d}T(\mathbb{P}_n).P,\ \sup_{t\in[\tau_1,\tau]} \mathrm{EL}_n^{LL}(T, P, t) \geqslant u_\alpha \right\}. \end{aligned}$$

On en déduit qu'une région de confiance asymptotique pour $T(P_0)$ est donnée par :

$$R_{n,1-\alpha} = \left\{ T(P_0) + \mathrm{d}T(\mathbb{P}_n).(P - P_0),\ \sup_{t\in[\tau_1,\tau]} \mathrm{EL}_n^{LL}(T, P, t) \geqslant u_\alpha \right\}.$$

Cependant, il est clair qu'on ne peut pas construire $R_{n,1-\alpha}$. Nous allons donc montrer que les régions ci-dessous sont encore des régions de confiance

asymptotiques pour $T(P_0)$:

$$R^C_{n,1-\alpha} = \left\{ T(\mathbb{P}_n) + \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - \mathbb{P}_n),\ \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}} \right\},$$
$$R^L_{n,1-\alpha} = \left\{ T(\mathbb{P}_n) + \mathrm{d}T(\mathbb{P}_n).(P - \mathbb{P}_n),\ \sup_{t\in[\tau_1,\tau]} \mathrm{EL}^{LL}_n(T, P, t) \geqslant u_\alpha \right\}.$$

Rappelons que :

$$\overline{\mathcal{P}_{n,1-\alpha}} := \left\{ \widetilde{\mathbb{P}}_n = \sum_{i=1}^n p_{i,n}\delta_{X_i},\ \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \text{ et } \prod_{i=1}^n np_{i,n} \geqslant u_\alpha \right\},$$

où u_α est tel que

$$\mathbb{P}\left(\sup_{t\in[\tau_1,\tau]} G(t)V^{-1}(t)G(t) \leqslant -2\ln u_\alpha \right) = 1-\alpha,$$

et où G et V sont donnés au théorème 9.6.

Pour comparer les différentes régions, nous allons évaluer la distance qui les sépare. Pour ce faire, on définira la distance $d(A, B)$ entre deux ensembles A et B d'un espace métrique (E, d) par :

$$d(A, B) := \sup_{a\in A} d(a, B) = \sup_{a\in A}\ \inf_{b\in B} d(a, b).$$

Pour démontrer que les régions $R^C_{n,1-\alpha}$ et $R^L_{n,1-\alpha}$ sont des régions de confiance asymptotiques, on montre qu'elles sont asymptotiquement équivalentes à $R_{n,1-\alpha}$ au sens où leur distance à $R_{n,1-\alpha}$ est un $O_p(n^{-1/2})$. Comme les calculs sont très similaires dans les deux cas, et comme la région $R^C_{n,1-\alpha}$ n'est pas exploitable en pratique, nous nous contenterons ici de démontrer le résultat pour $R^L_{n,1-\alpha}$. Commençons par remarquer que :

$$\mathrm{EL}^{LL}_n(T, P) \geqslant u_\alpha \iff \exists\, \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}}\ :\ \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0.$$

Ainsi :

$$\theta \in R^L_{n,1-\alpha} \iff \begin{cases} \exists\, \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}}\ :\ \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0, \\ \theta = T(\mathbb{P}_n) + \mathrm{d}T(\mathbb{P}_n).(P - \mathbb{P}_n). \end{cases}$$

En injectant la première équation dans la deuxième, on peut donc aussi écrire :

$$\theta \in R^L_{n,1-\alpha} \iff \begin{cases} \exists\, \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}}\ :\ \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0, \\ \theta = T(\mathbb{P}_n) + \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - \mathbb{P}_n). \end{cases}$$

Soit alors $\theta \in R^L_{n,1-\alpha}$. On déduit de ce qui précède que :

$$d(\theta, R_{n,1-\alpha}) \leqslant \underbrace{T(P_0) - T(\mathbb{P}_n)}_{=O_p(n^{-1/2})} + \mathrm{d}T(P_0).(\underbrace{\widetilde{\mathbb{P}}_n - P_0}_{=O_p(n^{-1/2})}) - \mathrm{d}T(\mathbb{P}_n).(\underbrace{\widetilde{\mathbb{P}}_n - \mathbb{P}_n}_{=O_p(n^{-1/2})}),$$

où les $O_p(n^{-1/2})$ sont tous uniformes en $\widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}}$. Donc, puisque $\mathrm{d}T$ est continue au voisinage de P_0, $d(R^L_{n,1-\alpha}, R_{n,1-\alpha}) = O_P(n^{-1/2})$. Ainsi, les deux régions sont asymptotiquement équivalentes.

Remarque 9.9. On peut reprendre les arguments présentés dans Bertail (2006) pour démontrer que la région de confiance définie ci-dessous est asymptotiquement équivalente à $R_{n,1-\alpha}$ à un $O_p(n^{-1/2})$ près :

$$R^T_{n,1-\alpha} = \left\{ T(\widetilde{\mathbb{P}}_n),\ \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}} \right\}.$$

Cependant, nous n'exploiterons pas ce résultat dans la suite, car comme nous l'avons expliqué à la section 9.2, la construction de $R^T_{n,1-\alpha}$ n'est pas numériquement viable.

Application

Nous avons donné plusieurs choix de régions de confiance asymptotiques à la section précédente. En pratique, nous nous intéresserons uniquement à $R^L_{n,1-\alpha}$. Rappelons que :

$$\mathrm{EL}^{LL}_n(T, P, \cdot) = \max \left\{ \prod_{i=1}^n np_i,\ \mathrm{d}T(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0 \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \right\},$$

$$R^L_{n,1-\alpha} = \left\{ T(\mathbb{P}_n) + \mathrm{d}T(\mathbb{P}_n).(P - \mathbb{P}_n),\ \mathrm{EL}^{LL}_n(T, P) \geqslant u_\alpha \right\}.$$

Ainsi, à $t \in [\tau_1, \tau]$ fixé, on peut réécrire $\mathrm{EL}^{LL}_n(T, P, t)$ en fonction d'un paramètre $\mu(t) \in \mathrm{Conv}\left(\left(\mathrm{d}T(\mathbb{P}_n).\delta_{X_i}\right)(t), 1 \leqslant i \leqslant n\right) \subset \mathbb{R}^k$:

$$\mathrm{EL}^{LL}_n(\mu, t) = \max \left\{ \prod_{i=1}^n np_i,\ \mathrm{d}T(\mathbb{P}_n).\widetilde{\mathbb{P}}_n(t) = \mu(t) \text{ et } \widetilde{\mathbb{P}}_n \in \mathcal{P}_{n-1} \right\}.$$

On s'est donc ramené à un problème « simple » à résoudre, du moins en petite dimension. En effet, si $k = 1$, alors on peut déterminer une région de confiance par dichotomies successives (en faisant varier t dans l'intervalle $[\tau_1, \tau]$). Si $k = 2$, alors l'enveloppe convexe des $\left(\mathrm{d}T(\mathbb{P}_n).\delta_{X_i}\right)(t)$ dans laquelle doit varier le paramètre μ est de dimension 2 et on peut encore, en passant en coordonnées polaires, déterminer par dichotomies successives la région de confiance. On se reportera à la partie III pour plus de détails.

9.5 Quelques exemples élémentaires

Nous allons voir dans les deux exemples qui suivent comment appliquer le théorème 9.6 dans des situations simples où les processus en jeu sont constants.

9.5.1 Variance

Supposons que l'on observe n réalisations de variables aléatoires i.i.d. $X_1, \ldots, X_n$ à valeurs réelles et que l'on souhaite construire une région de

confiance pour la variance des X_i. Posons $\mathcal{F} := \{f_0, f_1\}$ où $f_0 : x \mapsto x$ et où $f_1 : x \mapsto x^2$. Définissons :

$$\begin{array}{rccc} T_v & : & \mathcal{P} & \to & \mathbb{R} \\ & & P & \mapsto & P(f_1) - P(f_0)^2, \end{array}$$

où $\mathcal{P}$ désigne l'ensemble des mesures signées sur $\mathbb{R}$ admettant une variance. Alors, T_v est différentiable au sens usuel puisque pour tout $(P, Q) \in \mathcal{P}^2$:

$$T_v(P+Q) - T_v(P) = \int X^2 \mathrm{d}Q - 2\int X\mathrm{d}P \int X\mathrm{d}Q - \left(\int X\mathrm{d}Q\right)^2.$$

Or, par définition :

$$\|Q\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} \left|\int f\mathrm{d}Q\right| \geqslant \int X\mathrm{d}Q.$$

On en déduit que T_v est différentiable sur $\mathcal{P}$ et que pour tout $(P, Q) \in \mathcal{P}^2$:

$$\begin{aligned} \mathrm{d}T_v(P).Q &= \int X^2\mathrm{d}Q - 2\int X\mathrm{d}P\int X\mathrm{d}Q \\ &= Q(f_1) - 2P(f_0)Q(f_0). \end{aligned}$$

Il reste à démontrer que T_v est Hadamard-différentiable en P_0 tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$. Pour ce faire, considérons une suite $t_n \to 0$ et une suite $\psi_n \to \psi \in \mathcal{B}(\mathcal{F}, P_0)$ telles que pour tout n, $P_0 + t_n\psi_n \in \mathcal{P}$. On a :

$$\begin{aligned} T_v(P_0 + t_n\psi_n) - T_v(P_0) &= (P_0 + t_n\psi_n)(f_1) - \big((P_0 + t_n\psi_n)(f_0)\big)^2 - T_v(P_0) \\ &= t_n\psi_n(f_1) - 2t_nP_0(f_0)\psi_n(f_0) - t_n^2\psi_n(f_0)^2. \end{aligned}$$

Le choix de la classe $\mathcal{F}$ nous permet donc de conclure que T_v est Hadamard-différentiable en P_0 tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$ et que sa différentielle est donnée par :

$$\forall \psi \in \mathcal{B}(\mathcal{F}, P_0) \ : \quad \mathrm{d}T_v(P_0).\psi = \psi(f_1) - 2P(f_0)\psi(f_0).$$

On remarque que les formules sont les mêmes pour les différentielles au sens de Gâteaux ou au sens de Hadamard. De plus, il est clair que l'on peut reprendre les calculs précédents en remplaçant P_0 par n'importe quel élément de $\mathcal{P}$. On remarque ainsi que $\mathrm{d}T_v$ est continue en P_0. Remarquons maintenant que :

$$\begin{aligned} \widehat{m}_i(X_i, T, P_0) &= \mathrm{d}T_v(\mathbb{P}_n).(\delta_{X_i} - P_0) \\ &= (\delta_{X_i} - P_0)(f_1) - 2P_0(f_0)(\delta_{X_i} - P_0)(f_0) \\ &= X_i^2 - \sigma^2 - \mu^2 - 2\bar{X}X_i + 2\bar{X}\mu, \end{aligned}$$

où on a noté σ^2 et μ la variance et l'espérance de P_0, respectivement. On en déduit que si les X_i admettent un moment d'ordre 4, alors la condition

(B2)* du théorème 9.6 est vérifiée. Quant à la condition (A0*), elle est clairement vérifiée. Rappelons maintenant que la version linéarisée du rapport de vraisemblance empirique s'écrit :

$$\mathrm{EL}_n^{LL}(P, T_v) = \max\left\{\prod_{i=1}^n np_i,\ \mathrm{d}T_v(\mathbb{P}_n)(\widetilde{\mathbb{P}}_n - P) = 0 \text{ et } \widetilde{\mathbb{P}}_n \in \overline{\mathcal{P}_{n,1-\alpha}}\right\}.$$

On peut donc la réécrire en fonction du paramètre réel $a = a(P) = \mathrm{d}T_v(\mathbb{P}_n).P$:

$$\mathrm{EL}_n^{LL}(a) = \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i\left(X_i^2 - 2\bar{X}X_i\right) = a \text{ et } p \in \mathbb{S}_{n-1}\right\}.$$

Ainsi, une région de confiance asymptotique est donnée par :

$$\begin{aligned} R_{1-\alpha}^L &= \left\{T_v(\mathbb{P}_n) + \mathrm{d}T_v(\mathbb{P}_n).(P - \mathbb{P}_n),\ \mathrm{EL}_n^{LL}(a) \geqslant u_\alpha\right\} \\ &= \left\{a + (\bar{X})^2,\ \mathrm{EL}_n^{LL}(a) \geqslant u_\alpha\right\}. \end{aligned}$$

9.5.2 Espérance et variance

Supposons que l'on observe n réalisations de variables aléatoires i.i.d. $X_1, \ldots, X_n$ et que l'on veuille construire une région de confiance simultanée pour l'espérance et la variance des X_i. On définit la même classe $\mathcal{F}$ et le même ensemble $\mathcal{P}$ qu'à la section précédente. On définit aussi la fonctionnelle T_{ev} par :

$$\begin{array}{rccc} T_{ev} & : & \mathcal{P} & \to & \mathbb{R}^2 \\ & & P & \mapsto & \left(P(f_0), P(f_1) - P(f_0)^2\right)^t. \end{array}$$

En reprenant les résultats présentés à la section précédente, on démontre facilement que T_{ev} est différentiable et que sa différentielle est continue (en prenant la même classe $\mathcal{F}$. On montre aussi que toutes les hypothèses du théorème 9.6 sont vérifiées dès que les X_i admettent un moment d'ordre 4. De plus, un simple calcul montre que EL_n^{LL} ne dépend de P qu'au travers de son espérance et de sa variance. On obtient alors aisément la formule ci-dessous pour la version linéarisée du rapport de vraisemblance empirique :

$$\mathrm{EL}_n^{LL}(\mu, \sigma^2) = \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i\begin{pmatrix} X_i \\ X_i^2 \end{pmatrix} = \begin{pmatrix} \mu \\ \sigma^2 + \mu^2 \end{pmatrix} \text{ et } p \in \mathbb{S}_{n-1}\right\}.$$

On en déduit une région de confiance asymptotique pour la moyenne et la variance qui est :

$$R_{n,1-\alpha}^L = \left\{\begin{pmatrix} \mu \\ \sigma^2 + \mu^2 + 2\bar{X}\mu - 3(\bar{X})^2 \end{pmatrix},\ \mathrm{EL}_n^{LL}(\mu, \sigma^2) \geqslant u_\alpha\right\}.$$

Comme nous l'avions remarqué à la section 9.2, cette région est une transformation affine de celle obtenue directement avec le théorème d'Owen. En effet :

$$R_{n,1-\alpha}^L = \left\{\begin{pmatrix} 1 & 0 \\ 2\bar{X} & 1 \end{pmatrix}\begin{pmatrix} \mu \\ \sigma^2 + \mu^2 \end{pmatrix} + \begin{pmatrix} 0 \\ -3(\bar{X})^2 \end{pmatrix},\ \mathrm{EL}_n^{LL}(\mu, \sigma^2) \geqslant u_\alpha\right\}.$$

CHAPITRE 10

Bande de confiance pour les fonctions de risque cumulé - deuxième approche

10.1 Introduction

Nous avons présenté les fonctions de risque cumulé au chapitre 4. Nous avons ensuite construit des bandes de confiance par vraisemblance empirique pour ces fonctions au chapitre 6. Pour ce faire, nous nous sommes basé sur l'estimateur de Nelson-Aalen et avons mis à profit ses propriétés asymptotiques.

Dans ce chapitre, nous allons voir comment utiliser les résultats du chapitre 9 pour déterminer des bandes de confiance par vraisemblance empirique pour la fonction de risque cumulé dans le cadre de durées de vie censurées. Cette approche directe ne nécessite aucune connaissance sur l'estimateur de Nelson-Aalen et fait appel uniquement aux résultats du chapitre 9.

10.2 Durées de vie non censurées

10.2.1 Rappels

On a vu que si P était une loi de probabilité absolument continue sur $\mathbb{R}^+$, alors la fonction de risque cumulé associée à P était définie par :

$$A_P(t) = -\ln P(]t, +\infty[). \tag{10.1}$$

Dans le cas où P n'était pas absolument continue, on avait posé :

$$A_P(t) = \int_{]0,t]} \frac{\mathrm{d}P(]-\infty, s])}{P([s, +\infty[)}, \tag{10.2}$$

cette dernière définition est équivalente à la première dès que P est absolument continue (on se reportera au chapitre 4 pour plus de détails). De plus,

si Q est une mesure signée sur $\mathbb{R}^+$, alors, on peut encore définir le risque cumulé avec (10.2) lorsque cette quantité a un sens et dans le cas où Q est absolument continue on a la formule :

$$A_Q(t) = -\ln Q(]t, +\infty[) + \ln Q(]0, +\infty[), \tag{10.3}$$

si ces quantités sont bien définies.

Supposons que l'on observe n réalisations de variables aléatoires réelles i.i.d. $X_1, \dots, X_n$ de loi absolument continue sur $\mathbb{R}^+$. Soit τ un réel positif tel que $\mathbb{P}(X_i > \tau) > 0$. On cherche à construire une région de confiance pour $A_{P_0} \in \mathcal{D}([0, \tau])$. Dans ce qui suit, nous noterons $\mathcal{P}$ l'ensemble des mesures signées pour lesquelles on peut définir la fonction de risque cumulé via (10.2). Il est clair que P_0 est un élément de $\mathcal{P}$ et que pour n suffisamment grand, $\mathbb{P}_n$ est aussi un élément de $\mathcal{P}$ par définition de τ. De plus, pour tout u positif, $P_0 + u\delta_x$ est dans $\mathcal{P}$ et donc $\mathcal{P}$ vérifie les conditions du théorème 9.6. Définissons maintenant $\mathcal{F}$ comme étant l'ensemble des fonctions f_t, $t \in [0, \tau]$ où :

$$\begin{array}{rccc} f_t & : & \mathbb{R}^+ & \to & \mathbb{R} \\ & & x & \mapsto & \mathbb{1}_{[0,t]}(x). \end{array}$$

Notons encore $f_\infty \equiv 1$ et définissons la fonction :

$$\begin{array}{rccc} T_A & : & \mathcal{P} \subset \ell_\infty(\mathcal{F}) & \to & \mathcal{D}([0, \tau]) \\ & & P & \mapsto & \displaystyle\int_{]0,\cdot]} \frac{\mathrm{d}P(f_s)}{P(f_\infty) - \lim\limits_{n\to+\infty} P(f_{s-1/n})}, \end{array}$$

de sorte que $T_A(P) = A_P$. Remarquons que la limite dans la définition de T_A existe par définition d'une loi de probabilité. Il est clair que la classe $\mathcal{F}$ vérifie les conditions du théorème 9.6 puisqu'elle est de dimension de Vapnik inférieure à 2 et qu'elle admet f_∞ comme enveloppe de carré intégrable. On se reportera à la section 2.11 pour plus de détails.

10.2.2 Hadamard-différentiabilité

Montrons que T_A est Hadamard-différentiable en P_0 tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$. Pour ce faire, considérons deux suites $(t_n)_{n\in\mathbb{N}^\star}$ et $(Q_n)_{n\in\mathbb{N}^\star}$ telles que $t_n \to 0$, $Q_n \to \psi \in \mathcal{B}(\mathcal{F}, P_0)$ et telles que pour tout n, $P_0 + t_n Q_n \in \mathcal{P}$. Avant de nous lancer dans les calculs, remarquons que ψ vérifie :

$$\forall s \in]0, \tau] \ : \quad \lim_{n\to\infty} \psi(f_{s-1/n}) = \psi(f_s).$$

En effet, ψ étant $\|\cdot\|_{2,P_0}$-uniformément continue, on sait que :

$$\forall \varepsilon > 0, \exists \eta > 0 \ : \quad \|f_s - f_t\|_{2,P_0} < \eta \Rightarrow |\psi(f_s) - \psi(f_t)| < \varepsilon.$$

Mais :

$$\|f_{s-1/n} - f_s\|^2_{2,P_0} = \int |f_{s-1/n} - f_s|^2 \mathrm{d}P_0 = P_0([s - 1/n, s]).$$

D'où le résultat par absolue continuité de P_0. Calculons maintenant l'accroissement de T_A :

$$T_A(P_0 + t_n Q_n)(t) = \int_{]0,t]} \frac{\mathrm{d}P_0([0,s]) + t_n \,\mathrm{d}Q_n([0,s])}{P_0([s,+\infty[) + t_n Q_n([s,+\infty[)}.$$

Pour différentier T_A, nous allons effectuer un développement limité. Il existe une fonction ε tendant vers 0 et 0 telle que si on note

$$R_n(t) := \int_{]0,t]} t_n \frac{Q_n([s,+\infty[)}{P_0([s,+\infty[)^2}\, \varepsilon\left(t_n \frac{Q_n([s,+\infty[)}{P_0([s,+\infty[)^2} \right) \mathrm{d}P_0([0,s]),$$

alors :

$$\begin{aligned}
&\int_{]0,t]} \frac{\mathrm{d}P_0([0,s])}{P_0([s,+\infty[) + t_n Q_n([s,+\infty[)} \\
&\quad = \int_{]0,t]} \left(\frac{1}{P_0([s,+\infty[)} - t_n \frac{Q_n([s,+\infty[)}{P_0([s,+\infty[)^2} \right) \mathrm{d}P_0([0,s]) + R_n(t) \\
&\quad = T_A(P_0)(t) - t_n \int_{]0,t]} \frac{Q_n([s,+\infty[)}{P_0([s,+\infty[)^2} \mathrm{d}P_0([0,s]) + R_n(t).
\end{aligned}$$

Or :

$$\sup_{t\in[0,\tau]} R_n(t) \leqslant t_n \frac{1}{P_0([\tau,+\infty[)^2} \sup_{t\in[0,\tau]} \left| Q_n([s,+\infty[) \varepsilon\left(t_n \frac{Q_n([s,+\infty[)}{P_0([s,+\infty[)^2} \right) \right|. \quad (10.4)$$

Mais, par définition de la norme sur $\ell_\infty(\mathcal{F})$, comme $Q_n \to \psi$, on a :

$$\sup_{t\in[0,\tau]} \left| Q_n([0,t]) - \psi(f_t) \right| \longrightarrow 0, \quad \text{lorsque } n \to +\infty.$$

De plus, ψ est bornée. On en déduit que $Q_n([\cdot,+\infty[)$ est uniformément borné et on peut alors passer à la limite dans (10.4). D'où :

$$\lim_{n\to+\infty} \sup_{t\in[0,\tau]} \frac{1}{t_n} R_n(t) \longrightarrow 0, \quad \text{lorsque } n \to +\infty.$$

On montre de même qu'il existe une fonction $R_n'(t)$ telle que :

$$t_n \int_{]0,t]} \frac{\mathrm{d}Q_n(]-\infty,s])}{P_0([s,+\infty[) + t_n Q_n([s,+\infty[)} = t_n \int_{]0,t]} \frac{\mathrm{d}Q_n([0,s])}{P_0([s,+\infty[)} + R_n'(t). \quad (10.5)$$

Cependant, on ne peut pas, *a priori*, passer à la limite dans (10.5) car la fonction ψ n'est pas nécessairement à variations bornées. On effectue donc une intégration par parties [1], on trouve que la partie linéaire du membre de droite dans (10.5) est égale à :

$$\frac{Q_n([0,t])}{P_0(]t,+\infty[)} - \frac{Q_n(\{0\})}{P_0(]0,+\infty[)} + \int_{]0,t]} \frac{Q_n([0,s])\mathrm{d}P_0([s,+\infty[)}{P_0([s,+\infty[)P_0(]s,+\infty[)}$$

1. Si F et G sont deux fonctions à variations bornées, on a la formule (cf. Shorack & Wellner (2009)) :

$$F(b^+)G(b^+) - F(a^+)G(a^+) = \int_{]a,b]} U(s^-)\mathrm{d}V(s) + \int_{]a,b]} V(s^+)\mathrm{d}U(s).$$

On peut aussi effectuer une intégration par parties pour montrer que $R_n'(t)$ converge vers 0 uniformément en t lorsque n tend vers $+\infty$. Finalement, on en déduit que T_A est Hadamard-différentiable en P_0 tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$ et que pour tout $\psi \in \mathcal{B}(\mathcal{F}, P_0)$ on a :

$$\begin{aligned}(\mathrm{d}T_A(P_0).\psi)(t) = &- \int_{]0,t]} \frac{\psi(f_\infty) - \psi(f_s)}{P_0([s,+\infty[)^2}\mathrm{d}P_0([0,s]) + \frac{\psi(f_t)}{P_0(]t,+\infty[)} \\ &- \frac{\psi(f_0)}{P_0(]0,+\infty[)} + \int_{]0,t]} \frac{\psi(f_s)}{P_0([s,+\infty[)P_0(]s,+\infty[)}\mathrm{d}P_0([s,+\infty[)\end{aligned} \tag{10.6}$$

On peut utiliser l'absolue continuité de P_0 pour simplifier l'expression ci-dessus. On trouve alors, pour tout $t \in [0,\tau]$:

$$\big(\mathrm{d}T_A(P_0).\psi\big)(t) = \frac{-1}{P_0([t,+\infty[)}\Big(\psi(f_\infty) - \psi(f_t)\Big) + \psi(f_\infty) - \psi(f_0).$$

10.2.3 Gâteaux-différentiabilité

Montrons maintenant que T_A est Gâteaux-différentiable en P_0 dans les directions δ_x. Soit alors $x \in \mathbb{R}^+$ et δ_x la mesure de Dirac associée. Soit encore $u \in \mathbb{R}^+$, on a :

$$T_A(P_0 + u\delta_x)(t) = \int_{]0,t]} \frac{\mathrm{d}P_0([0,s]) + \mathrm{d}u\delta_x([0,s])}{P_0([s,+\infty[) + u\delta_x([s,+\infty[)}$$

Comme précédemment, nous allons effectuer un développement limité. Il existe une fonction ε tendant vers 0 en 0 telle que si on note

$$R_u(t) := \int_{]0,t]} u\frac{\delta_x([s,+\infty[)}{P_0([s,+\infty[)^2}\,\varepsilon\left(u\frac{\delta_x([s,+\infty[)}{P_0([s,+\infty[)^2}\right)\mathrm{d}P_0([0,s]),$$

alors :

$$\begin{aligned}&\int_{]0,t]} \frac{\mathrm{d}P_0([0,s])}{P_0([s,+\infty[) + u\delta_x([s,+\infty[)} \\ &\quad= \int_{]0,t]} \left(\frac{1}{P_0([s,+\infty[)} - u\frac{\delta_x([s,+\infty[)}{P_0([s,+\infty[)^2}\right)\mathrm{d}P_0([0,s]) + R_u(t) \\ &\quad= T_A(P_0)(t) - u\int_{]0,t]} \frac{\delta_x([s,+\infty[)}{P_0([s,+\infty[)^2}\mathrm{d}P_0([0,s]) + R_u(t) \\ &\quad= T_A(P_0)(t) - u\left(\frac{1}{P_0([x\wedge t,+\infty[)} - \frac{1}{P_0([0,+\infty[)}\right) + R_u(t).\end{aligned}$$

On montre sans difficulté que le reste tend uniformément (en $t \in [0,\tau]$) vers 0 lorsque u tend vers 0. De même :

$$\begin{aligned}u\int_{]0,t]} \frac{\mathrm{d}\delta_x([0,s])}{P_0([s,+\infty[) + u\delta_x([s,+\infty[)} &= u\int_{]0,t]} \frac{\mathrm{d}\delta_x([0,s])}{P_0([s,+\infty[)} + o(u) \\ &= u\,\frac{1}{P_0([x,+\infty[)}\mathbb{1}_{\{x\leqslant t\}} + o(u).\end{aligned}$$

On en déduit que T_A est Gâteaux-différentiable suivant les directions δ_x, $x \in \mathbb{R}^+$, et que :

$$\mathrm{d}T_A(P_0).\delta_x \equiv \frac{1}{P_0([x \wedge \cdot, +\infty[)} - \frac{1}{P_0([0, +\infty[)} + \frac{1}{P_0([x, +\infty[)} \mathbb{1}_{\{x \leqslant \cdot\}}$$

Puis, en découpant suivant si $x \leqslant t$ ou $x > t$, on trouve :

$$\big(\mathrm{d}T_A(P_0).\delta_x\big)(t) = -\frac{1}{P_0([t, +\infty[)} \delta_x([t, +\infty[) + \delta_x([0, +\infty[).$$

De plus, comme les X_i sont absolument continus, on peut se restreindre sans perte de généralité aux δ_x, $x \in \mathbb{R}_+^\star$. On retrouve alors la même formule que pour la Hadamard-différentiabilité :

$$\big(\mathrm{d}T_A(P_0).\delta_x\big)(t) = -\frac{1}{P_0([t, +\infty[)} \Big(\delta_x(f_\infty) - \lim_{n\to\infty} \delta_x(f_{t-1/n})\Big) + \delta_x(f_\infty) - \delta_x(\{0\}).$$

10.2.4 Continuité de la différentielle

Soit $P \in \mathcal{P}$ quelconque. On peut reprendre les calculs présentés aux sections précédentes pour démontrer que T_A est Hadamard-différentiable en P tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$ et Gâteaux-différentiable suivant les δ_x pour $x \in \mathbb{R}_+^\star$ quelconque. On retrouve alors les mêmes formules pour l'expression de la différentielle, bien qu'elles ne se simplifient plus. Il reste à montrer que :

$$\|\mathrm{d}T_A(P_0) - \mathrm{d}T_A(P)\|_{\mathcal{F}} \to 0 \quad \text{lorsque } P \to P_0.$$

Or, si $\psi \in \mathrm{Vect}\left(\{\delta_x, x \in \mathbb{R}^+\} \cup \mathcal{B}(\mathcal{F}, P_0)\right)$ est de norme 1, dans l'expression de $(\mathrm{d}T_A(P_0) - \mathrm{d}T_A(P)).\psi$ apparaissent des quantités de la forme :

$$\frac{1}{P_0([t, +\infty[)} - \frac{1}{P([t, +\infty[)},$$

où encore :

$$\int_{]0,t]} \left(\frac{\mathrm{d}P_0([0, s])}{P_0([s, +\infty[)^2} - \frac{\mathrm{d}P([0, s])}{P([s, +\infty[)^2}\right),$$

qui convergent toutes vers 0 lorsque P tend vers P_0. En effet, par définition de la classe $\mathcal{F}$:

$$\|P - P_0\|_{\mathcal{F}} = \sup_{t\in[0,\tau]} \Big|P([0, t]) - P_0([0, t])\Big|.$$

10.2.5 Dernières vérifications

Vérifions maintenant la condition (B2*) du théorème 9.6. Pour ce faire, définissons :

$$\mathcal{F}_{n\omega} := \Big\{\big(m_i(X_i, T, P_0, t)(\omega)\big)_{1\leqslant i\leqslant n},\ t \in [0, \tau]\Big\}.$$

D'après le calculs menés précédemment, on a :

$$m_i(X_i, T, P_0, t) = -\frac{\mathbb{1}_{\{X_i \geqslant t\}}}{P_0([t, +\infty[)} + 1. \tag{10.7}$$

On en déduit facilement que $\mathcal{F}_{n\omega}$ est euclidienne et bornée par :

$$|m_i(X_i, T, P_0, t)| \leqslant \frac{1}{P_0([\tau, +\infty[)} + 1.$$

Ainsi, par stabilité par produit des classes euclidiennes, la condition (B2*) est vérifiée.

Il reste maintenant à vérifier la condition (A0*). D'après la remarque 9.7, comme les $\mathcal{F}_{n\omega}$ sont euclidiennes, il suffit de vérifier que :

$$0 < \inf_{t\in[\tau_1,\tau]} \mathbb{E}\,|m_i(t)| \leqslant \sup_{t\in[\tau_1,\tau]} \mathbb{V}\mathrm{ar}\, m_i(t) < \infty,$$

où τ_1 est un réel à déterminer. Or, les m_i étant bornés (uniformément en t), leur variance est bornée. De plus, d'après (10.7), il est clair que si τ_1 est un réel vérifiant $P_0([\tau_1, +\infty[) < 1$, alors la condition recherchée est vérifiée.

10.2.6 Conclusion

D'après ce qui précède, toutes les conditions du théorème 9.6 sont satisfaites. On en déduit que EL_n^{LL} converge en distribution. De plus, on a :

$$\mathrm{d}T_A(\mathbb{P}_n).(\widetilde{\mathbb{P}}_n - P) = 0 \Longleftrightarrow \forall t \in [0, \tau] \ : \ \widetilde{\mathbb{P}}_n(]t, +\infty]) = P(]t, +\infty])$$

Si on note $S = S(P)$ la fonction de survie associée à P, alors, on peut réécrire EL_n^{LL} en fonction de S :

$$\mathrm{EL}_n^{LL}(S, t) = \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i \mathbb{1}_{]t,+\infty]}(X_i) = S(t) \text{ et } p \in \mathbb{S}_{n-1}\right\}.$$

Ainsi, la région de confiance $R_{n,1-\alpha}^L$ est-elle égale à :

$$R_{n,1-\alpha}^L = \left\{-\ln \mathbb{P}_n(]\cdot, +\infty]) - \frac{S}{\mathbb{P}_n(]\cdot, +\infty])} - 1,\ \sup_{t\in[0,\tau]} \mathrm{EL}_n^{LL}(S, t) \geqslant u_\alpha\right\}.$$

Définissons :

$$R_{n,1-\alpha}^d = \left\{S,\ \sup_{t\in[0,\tau]} \mathrm{EL}_n^{LL}(S, t) \geqslant u_\alpha\right\}.$$

Il est clair que si nous arrivons à construire $R_{n,1-\alpha}^d$ numériquement, alors, nous pourrons construire $R_{n,1-\alpha}^L$ par une simple transformation affine. Or, on peut construire $R_{n,1-\alpha}^d$ très simplement en procédant comme suit :

1. On fixe $t \in [0, \tau]$;
2. Pour a variant entre 0 et 1, on évalue $\mathrm{EL}_n^{LL}(a, t)$;

3. Si la valeur obtenue est supérieure ou égale à u_α, alors on ajoute le point (t, a) au tube de confiance pour S_0.

En fait, par convexité à t fixé de la région de confiance obtenue pour $S_0(t)$, on procédera par dichotomies pour déterminer les bornes de l'intervalle correspondant, en utilisant le fait que la fonction de survie empirique est dans $R^d_{n,1-\alpha}$.

Remarque 10.1. On a utilisé le théorème 9.6 pour déterminer une région de confiance pour la fonction de risque cumulé. Cependant, on aurait aussi pu déterminer une région de confiance R_n pour la fonction de survie puis considérer $-\ln R_n$. Pour ce faire, une approche naturelle aurait été de définir la fonction $\mathrm{EL}_n(S, \cdot)$ sur $[\tau_1, \tau]$ par :

$$\mathrm{EL}_n(S, t) := \max\left\{\prod_{i=1}^n np_i,\ \sum_{i=1}^n p_i \mathbb{1}_{]t,+\infty]}(X_i) = S(t) \text{ et } p \in \mathbb{S}_{n-1}\right\}.$$

On retrouve ici la même fonction que EL_n^{LL}. Cependant, la région de confiance qui en résulte pour la fonction de risque cumulé est sensiblement différente puisqu'on trouve cette fois :

$$R^A_{n,1-\alpha} = -\ln R^d_{n,1-\alpha} = \left\{-\ln S,\ \sup_{t\in[0,\tau]} \mathrm{EL}_n^{LL}(S, t) \geqslant u_\alpha\right\}.$$

Ici, la convergence en distribution de $\mathrm{EL}_n(S_0, \cdot)$ est une conséquence du théorème 9.6, mais elle peut être vue aussi comme une conséquence du théorème de Hjort *et al.* (2004).

10.3 Durées de vie censurées

Supposons être en présence de données censurées. On observe alors les $X_i = (\widetilde{X}_i \wedge C_i, d_i)$, où les $\widetilde{X}_i$ sont des durées de vie, les C_i des censures et $d_i = \mathbb{1}_{\{\widetilde{X}_i \leqslant C_i\}}$ (on ne notera pas les indicatrices de censure avec des δ comme dans les chapitres précédents pour éviter toute confusion avec les masses de Dirac). On supposera que les censures sont indépendantes des durées de vie et que les lois des $\widetilde{X}_i$ et des C_i sont absolument continues. La construction présentée à la section précédente ne fonctionne alors plus. En effet, on ne peut plus exprimer la fonctionnelle T de la même manière puisqu'on ne connaît pas la loi des $\widetilde{X}_i$. Cependant, la loi de $(\widetilde{X} \wedge C, \delta)$ contient suffisamment d'information pour déterminer la fonction de survie de $\widetilde{X}$, ou, de manière équivalente, sa fonction de risque cumulé. En effet, un simple calcul montre que, grâce à l'hypothèse d'indépendance entre $\widetilde{X}$ et C, on a :

$$\begin{aligned}\alpha(t) &= \lim_{h\to 0^+} \frac{1}{h}\,\mathbb{P}(\widetilde{X} \in [t, t+h[\,|\,\widetilde{X} \geqslant t)\\ &= \lim_{h\to 0^+} \frac{1}{h}\,\mathbb{P}(\widetilde{X} \wedge C \in [t, t+h[, d = 1\,|\,\widetilde{X} \wedge C \geqslant t). \qquad (10.8)\end{aligned}$$

On peut donc retrouver la fonction de risque cumulé A de $\widetilde{X}$ à partir de la loi du couple $(\widetilde{X} \wedge C, d)$ puisque :

$$A(t) = \int_0^t \alpha(s)\mathrm{d}s.$$

Cependant, si $\widetilde{X}$ n'est pas absolument continue, alors, il faut définir A autrement. Mais, on montre facilement qu'on peut définir A par :

$$A(t) := \int_{]0,t]} \frac{\mathrm{d}F(s)}{G(s)}, \tag{10.9}$$

où

$$F(s) := \mathbb{P}(\widetilde{X} \wedge C \leqslant s, d = 1) \quad \text{et} \quad G(s) := \mathbb{P}(\widetilde{X} \wedge C \geqslant s),$$

et que les deux définitions se rejoignent dès que $\widetilde{X}$ et C sont absolument continues.

Fixons alors τ positif tel que $\mathbb{P}(\widetilde{X} \wedge C > \tau) > 0$ et cherchons à construire une bande de confiance pour A, la fonction de risque cumulé de $\widetilde{X}$ que l'on considère comme un élément de $\mathcal{D}([0,\tau])$. Les X_i étant à valeurs dans $\mathbb{R}^+ \times \{0;1\}$, nous nous intéresserons à l'ensemble $\mathcal{P}$ des mesures signées P sur $\mathbb{R}^+ \times \{0;1\}$ pour lesquelles on peut définir la fonction de risque cumulé par (10.9), c'est-à-dire celles pour lesquelles la fonction $A_P \in \mathcal{D}([0,\tau])$ définie ci-dessous a un sens :

$$A_P(\cdot) := \int_{]0,\cdot]} \frac{\mathrm{d}P([0,s] \times \{1\})}{P([s,+\infty[\times\{0;1\})}.$$

Il est clair que $\mathcal{F}$ vérifie les hypothèses du théorème 9.6 par définition de τ. Notons maintenant $\mathcal{F}$ l'ensemble des fonctions $f_{t,L}$, où $t \in [0,\tau]$, $L \subset \{0;1\}$ et où :

$$\begin{array}{rccc} f_{t,L} & : & \mathbb{R} \times \{0;1\} & \to & \mathbb{R} \\ & & (z,d) & \mapsto & \mathbb{1}_{\{z \leqslant t, d \in L\}}. \end{array}$$

Pour alléger les notations nous écrirons $f_{t,0} = f_{t,\{0\}}$, $f_{t,1} = f_{t,\{1\}}$ et $f_t = f_{t,\{0;1\}}$. Nous écrirons encore f_∞ la fonction constante égale à 1. Il est clair que la classe $\mathcal{F}$ est une VC-classe de dimension finie et qu'elle vérifie toutes les hypothèses du théorème. Définissons alors la fonctionnelle T_A par :

$$\begin{array}{rccc} T_A & : & \mathcal{P} \subset \ell_\infty(\mathcal{F}) & \to & \mathcal{D}([0,\tau]) \\ & & P & \mapsto & \displaystyle\int_{]0,\cdot]} \frac{\mathrm{d}P(f_{s,1})}{P(f_\infty) - \lim\limits_{n\to\infty} P(f_{s-1/n})}. \end{array}$$

10.3.1 Différentiabilité

La fonctionnelle T_A étant définie de manière similaire à celle présentée à la section 10.2, les calculs de différentiabilité ne changent pratiquement pas, c'est pourquoi nous nous contenterons ici d'énoncer les résultats.

On montre que T_A est Hadamard-différentiable au voisinage de P_0 tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$ et qu'elle est Gâteaux-différentiable au voisinage de P_0 suivant les directions $\delta_{(z,d)}$, $(z, d) \in \mathbb{R}^+ \times \{0; 1\}$. On trouve alors la formule ci-dessous, valable pour tout $\psi \in \mathcal{B}(\mathcal{F}, P_0)$:

$$(\mathrm{d}T_A(P_0).\psi)(t) = -\int_{]0,t]} \frac{\psi(f_\infty) - \psi(f_s)}{P_0([s, +\infty[\times\{0;1\})^2} \mathrm{d}P_0([0, s] \times \{1\}) \tag{10.10}$$

$$+ \int_{]0,t]} \frac{\psi(f_{s,1})\, \mathrm{d}P_0([s, +\infty[\times\{0;1\})}{P_0([s, +\infty[\times\{0;1\}) P_0(]s, +\infty[) \times \{0;1\})}$$

$$+ \frac{\psi(f_{t,1})}{P_0(]t, +\infty[\times\{0;1\})} - \frac{\psi(f_{0,1})}{P_0(]0, +\infty[\times\{0;1\})} \tag{10.11}$$

Cette formule peut paraître compliquée. Cependant, pour appliquer le théorème 9.6, nous devrons étudier $\mathrm{d}T_A(P_0).(\delta_{X_i} - P_0)$. Or, évaluée en une mesure signée Q, l'expression de $\mathrm{d}T_A(P_0)$ se simplifie en intégrant par parties et on trouve :

$$\mathrm{d}T_A(P_0).Q \equiv -\int_{]0,\cdot]} \frac{Q([s, +\infty[\times\{0;1\})}{P_0([s, +\infty[\times\{0;1\})^2} \mathrm{d}P_0(]-\infty, s] \times \{1\})$$

$$+ \int_{]0,\cdot]} \frac{\mathrm{d}Q(]-\infty, s] \times \{1\})}{P_0([s, +\infty[\times\{0;1\})}.$$

Concernant la Gâteaux-différentiabilité, on trouve la même formule, comme dans le cadre de durées de vie non censurées. De plus, les arguments présentés à la section 10.2.4 sont encore valables ici : $\mathrm{d}T_A$ est donc continue au voisinage de P_0.

10.3.2 Dernières vérifications

Vérifions maintenant la condition (B2*) du théorème 9.6. Pour ce faire, remarquons que :

$$m_i(X_i, T, P_0, t) = \int_{]0,t]} \frac{\mathrm{d}\mathbb{1}_{\{\widetilde{X}_i \leqslant s, \delta_i = 1\}}}{P_0([\widetilde{X}_i, +\infty[\times\{0;1\})} - \int_0^t \frac{\mathbb{1}_{\{\widetilde{X}_i \geqslant s\}} \mathrm{d}P_0(]-\infty, s] \times \{1\})}{P_0([s, +\infty[\times\{0;1\})^2}$$

$$= \frac{\mathbb{1}_{\{\widetilde{X}_i \leqslant t, \delta_i = 1\}}}{P_0([\widetilde{X}_i, +\infty[\times\{0;1\})} - \int_0^{t\wedge\widetilde{X}_i} \frac{\mathrm{d}P_0(]-\infty, s] \times \{1\})}{P_0([s, +\infty[\times\{0;1\})^2}.$$

Ainsi, la loi des grands nombres uniforme appliquée aux classes ci-dessous démontre la convergence en probabilité de $\frac{1}{n}\sum_{i=1}^n m_i^2$ vers une fonction V :

$$\mathcal{F}_{n\omega} := \left\{ \left(\frac{\mathbb{1}_{\{0 < \widetilde{X}_i(\omega) \leqslant t, \delta_i = 1\}}}{P_0([\widetilde{X}_i(\omega), +\infty[\times\{0;1\})} \right)_{1 \leqslant i \leqslant n}, \ t \in [0, \tau] \right\},$$

$$\mathcal{G}_{n\omega} := \left\{ \left(\int_0^{t\wedge\widetilde{X}_i(\omega)} \frac{\mathrm{d}P_0([s, +\infty[\times\{1\})}{P_0([s, +\infty[\times\{0;1\})^2} \right)_{1 \leqslant i \leqslant n}, \ t \in [0, \tau] \right\}.$$

En effet, ces classes sont euclidiennes car les fonctions en jeu sont croissantes et elles sont bornées par hypothèse sur P_0. Soit $\tau_1 \in]0, \tau[$ tel que $\mathbb{P}_0([0, \tau_1] \times \{1\}) > 0$. D'après la remarque 9.7, il suffit, pour montrer (A0*), de vérifier que :

$$0 < \inf_{t\in[\tau_1,\tau]} \mathbb{E}\,|m_i(t)| \leqslant \sup_{t\in[\tau_1,\tau]} \mathbb{V}\mathrm{ar}\, m_i(t) < \infty.$$

Comme les m_i sont centrés, il suffit de montrer que

$$\inf_{t\in[\tau_1,\tau]} \mathbb{P}(m_i(X_i, T, P_0, t) \neq 0) > 0,$$

et que :

$$\sup_{t\in[\tau_1,\tau]} m_i(X, T, P_0, t) < C^{ste}.$$

La deuxième propriété est évidente car les m_i sont bornés (uniformément en t). Pour vérifier la première, considérons $\omega \in \Omega$ tel que $\widetilde{X}_i(\omega) > \tau$. On a, pour tout $t \in [\tau_1, \tau]$:

$$\begin{aligned} m_i(X_i(\omega), T, P_0, t) &= \int_0^t \frac{\mathrm{d}P_0(]-\infty, s] \times \{1\})}{P_0([s, +\infty[\times\{0;1\})} \\ &\geqslant \int_0^{\tau_1} \mathrm{d}P_0(]-\infty, s] \times \{1\}) \\ &= P_0(]-\infty, \tau_1] \times \{1\}) > 0. \end{aligned}$$

Mais, $\mathbb{P}(\widetilde{X}_i > \tau) > 0$, ce qui démontre le résultat.

CHAPITRE 11

Problèmes multi-états

11.1 Avant-propos

Dans ce chapitre, nous nous intéressons aux problèmes multi-états décrits par des processus de Markov non homogènes à temps continu. Ce modèle peut être utilisé pour décrire beaucoup de situations réelles. En fiabilité, par exemple, on pourrait considérer qu'un objet peut être dans deux états différents : en marche ou en panne. Dans le cas d'infections nosocomiales, un patient peut être dans plusieurs états correspondants aux différentes maladies, à l'état « décédé », ou encore à l'état « sain ». Il se peut que les observations soient soumises à une censure aléatoire à droite, correspondant à une fenêtre de temps d'observation. Dans ces conditions, il peut être intéressant de déterminer des bandes de confiance pour les matrices de transitions de telles chaînes. On trouvera dans Andersen *et al.* (1993) une réponse à ce problème via une approche de type martingale. Nous avons quant à nous voulu explorer cette question du point de vue de la vraisemblance empirique. Pour ce faire, nous avons utilisé le théorème 9.6. Avant de décrire précisément notre modèle, il nous faut introduire un nouvel objet, appelé produit intégral, et qui est au produit ce que l'intégrale est à la somme.

Définition 11.1 (Produit intégral)**.** Soit $X\colon [0,\tau] \to M_k(\mathbb{R})$ une fonction à valeurs matricielles dont les composantes sont des fonctions càdlàg à variations bornées. On appelle produit intégral de X et on écrit :

$$Y = \prod(\mathrm{Id} + \mathrm{d}X),$$

la fonction matricielle définie pour tout $t \in [0,\tau]$ par :

$$Y(t) = \prod_{s\in[0,t]} (\mathrm{Id} + \mathrm{d}X(s)) = \lim_{\max|t_i - t_{i-1}|\to 0} \prod (\mathrm{Id} + X(t_i) - X(t_{i-1})), \qquad (11.1)$$

où $0 = t_0 < t_1 < \ldots < t_n = t$ est une subdivision de $[0,t]$ et où le produit se lit de gauche à droite. Le terme $X(0)$ dans la définition doit être remplacé

par $X(0^-) := 0$ car 0 fait partie de l'intervalle d'intégration $[0, \tau]$. Tout comme pour l'intégration de fonctions à variations bornées, on peut définir le produit intégral en excluant une des bornes (ou les deux) de l'intervalle d'intégration.

On démontre que le produit intégral est bien défini de la même manière que l'on démontre que l'intégrale de Stieltjes l'est. Il existe une relation de Chasles qui est donnée par :

$$\prod_{[0,t]}(\mathrm{Id} + \mathrm{d}X) = \prod_{[0,s[}(\mathrm{Id} + \mathrm{d}X) \prod_{[s,t]}(\mathrm{Id} + \mathrm{d}X)$$

Énonçons maintenant un résultat de différentiabilité qui nous sera utile pour la suite. On en trouvera une démonstration dans Fleming (1978).

Proposition 11.2. *Soit $E_M^{k^2} \subset (\mathcal{D}([0,\tau]))^{k^2}$ l'ensemble des fonctions à valeurs dans $M_k(\mathbb{R})$ et dont les composantes sont à variations bornées par M. Soit encore :*

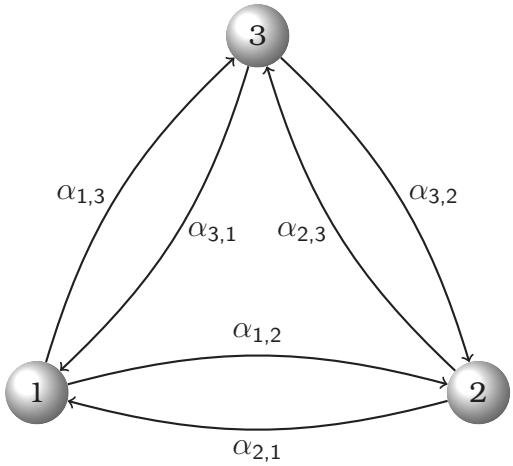
$$\begin{array}{rccc} \phi & : & E_M^{k^2} & \to & (\mathcal{D}([0,\tau]))^{k^2} \\ & & X & \mapsto & \prod(\mathrm{Id} + \mathrm{d}X). \end{array}$$

Alors, ϕ est Hadamard-différentiable tangentiellement à l'ensemble des fonctions de $(\mathcal{D}([0,\tau]))^{k^2}$ à variations bornées et sa différentielle est donnée par la formule :

$$\mathrm{d}\phi(X).H(t) = \int_0^t \prod_{[0,s[}(\mathrm{Id} + \mathrm{d}X)\mathrm{d}H(s) \prod_{]s,t]}(\mathrm{Id} + \mathrm{d}X).$$

11.2 Modèle

Considérons un processus de Markov non homogène à espace d'états finis $[\![1, k]\!]$ et à temps continu. Notons-le $X(\cdot)$. Notons encore $(P_{hj}(s,t))_{hj}$ sa matrice de transition et $(\alpha_{hj}(\cdot))_{h \neq j}$ les intensités correspondantes.



Rappelons que les α_{hj}, $h \neq j$, sont donnés par la formule :

$$\alpha_{hj}(t) = \lim_{u \to 0^+} \frac{1}{u} \mathbb{P}(X(t+h) = j \mid X(t) = h). \tag{11.2}$$

On sait que la loi de la chaîne est entièrement déterminée par les $(\alpha_{hj}(\cdot))$. En effet, notons $\alpha_{hh} := -\sum_{j\neq h}\alpha_{hj}$ et :

$$A_{hj}(t) := \int_0^t \alpha_{hj}(s)\mathrm{d}s.$$

Alors, on a la formule (cf Andersen *et al.* (1993) page 287) :

$$P(s,t) = \prod_{]s,t]}(\mathrm{Id} + \mathrm{d}A(u)). \tag{11.3}$$

Pour comprendre cette formule, on peut remarquer que si Z est une durée de vie de loi absolument continue et de fonction de risque cumulé A, alors, en passant au logarithme dans (11.1), on a :

$$\mathbb{P}(Z \geqslant t) = \exp(-A(t)) = \prod_{[0,t[}(1 - \mathrm{d}A)$$

La différence de signe entre cette équation et (11.3) vient du fait que les termes diagonaux sont dans notre cas négatifs.
Nous nous intéresserons donc au processus càdlàg sur $[0,\tau]$ défini par :

$$G(t) = \prod_{]0,t]}(\mathrm{Id} + \mathrm{d}A(u)).$$

On a déjà étudié la fonction de risque cumulé à la section 10.3. Nous allons généraliser ici quelques uns des résultats déjà démontrés. D'après ce qui précède, on sait que la loi de la chaîne est entièrement déterminée par les α_{hj}. Notons alors pour tout $h \neq j$ Z_{hj} une variable aléatoire de fonction de risque instantané α_{hj}. Une telle variable aléatoire existe car sa loi est entièrement déterminée par :

$$\mathbb{P}(Z_{hj} \geqslant t) = \exp\left(-\int_0^t \alpha_{hj}(s)\mathrm{d}s\right).$$

Si la chaîne se trouve dans un état h et qu'elle change d'état, on n'observera pas directement les Z_{hj} mais seulement le minimum de ces variables, ainsi que l'état dans lequel la chaîne se trouve après avoir quitté l'état h. Autrement dit, on observera les variables :

$$W_h := \min_{j\neq h} Z_{hj} \qquad \text{et} \qquad d_h := \operatorname*{argmin}_{j\neq h} Z_{hj}.$$

De plus, on peut retrouver les α_{hj} à partir de ces variables. En effet, en utilisant la propriété de Markov, on a :

$$\begin{aligned}
\alpha_{hj}(t) &= \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\mathbb{P}(X(t+\varepsilon) = j \mid X(t) = h) \\
&= \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\mathbb{P}(X(t+\varepsilon) = j \mid \forall s \leqslant t, X(t) = h) \\
&= \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\mathbb{P}(W_h \in [t, t+\varepsilon[,\ d_h = j \mid W_h \geqslant t).
\end{aligned}$$

Nous supposerons que la chaîne est soumise à une censure aléatoire à droite C indépendante après laquelle on n'observe plus l'état de la chaîne. Une telle censure peut être comprise comme une fenêtre d'observation, et on ajoutera à la chaîne un état absorbant 0 correspondant à la censure. Comme dans le cas de durées de vies censurées, on peut retrouver la loi de la chaîne à partir de la loi des $(W_h \wedge C, d_h)$. On peut donc identifier la loi de la chaîne à la loi de ces variables et travailler sur l'ensemble des mesures signées sur $(\mathbb{R}^+ \times [\![0,k]\!])^k$. Nous noterons alors $\mathcal{P}$ l'ensemble de ces mesures signées P pour lesquelles la fonction $T_A(P)$ est bien définie, où :

$$T_A(P) = \left(\int_0^{\cdot} \frac{\mathrm{d}P_h([0,s] \times \{j\})}{P_h([s,+\infty[\times [\![0,k]\!])} \right)_{h \neq j}, \tag{11.4}$$

et où on a noté P_h la h-ième marginale de P. Dans tout ce qui suit, nous supposerons qu'il existe $\tau > 0$ tel que les variables W_h et d_h définissant la loi de la chaîne vérifient :

$$\forall h \neq j \ : \quad \mathbb{P}(W_h > \tau, d_h = j) > 0. \tag{11.5}$$

Nous noterons alors $f_{t,h,j}$ les fonctions définies pour tout $t \in [0,\tau]$ et pour tout $(h,j) \in [\![0,k]\!]$ par :

$$\begin{array}{rccc} f_{t,h,j} & : & (\mathbb{R}^+ \times [\![0,k]\!])^k & \to & \mathbb{R} \\ & & (z_1, d_1, \dots, z_k, d_k) & \mapsto & \mathbb{1}_{\{z_h \leqslant t, d_h = j\}}. \end{array}$$

Nous noterons encore $f_{t,h}$ les fonctions définies pour tout $t \in [0,\tau]$ et pour tout $h \in [\![0,k]\!]$ par :

$$\begin{array}{rccc} f_{t,h} & : & (\mathbb{R}^+ \times [\![0,k]\!])^k & \to & \mathbb{R} \\ & & (z_1, d_1, \dots, z_k, d_k) & \mapsto & \mathbb{1}_{\{z_h \leqslant t\}}. \end{array}$$

Finalement, posons $f_\infty \equiv 1$. Nous noterons alors T_A la fonctionnelle :

$$\begin{array}{rccc} T_A & : & \mathcal{P} \subset \ell_\infty(\mathcal{F}) & \to & \mathcal{D}([0,\tau])^{k(k-1)} \\ & & P & \mapsto & \left(\displaystyle\int_0^{\cdot} \frac{\mathrm{d}P(f_{s,h,j})}{P(f_\infty) - \lim\limits_{n\to\infty} P(f_{h,s-1/n})} \right)_{h \neq j}, \end{array}$$

qui coïncide avec la fonction donnée à l'équation (11.4). On notera encore $T_A(P)$ la matrice dont les éléments diagonaux sont donnés par :

$$\big(T_A(P)\big)_{hh} = -\sum_{j \neq h} \big(T_A(P)\big)_{hj}.$$

Soit maintenant ϕ la fonction donnée à la proposition 11.2. Alors, la fonctionnelle T à laquelle nous nous intéressons est :

$$\begin{array}{rccc} T & : & \mathcal{P} & \to & \mathcal{D}([0,\tau])^{k^2} \\ & & P & \mapsto & \phi \circ T_A(P). \end{array}$$

On connaît la différentielle de ϕ par la proposition 11.2. On connaît aussi la différentielle de T_A d'après la section 10.3. En effet, on peut calculer la différentielle coordonnée par coordonnée, et les calculs sont les mêmes que ceux de la section 10.3 à ceci près que les d_h sont à valeurs dans $[\![0, k]\!]$ au lieu de $\{0; 1\}$. L'expression de la différentielle est alors donnée par une formule similaire, donnée pour tout $P \in \mathcal{P}$ et pour tout ψ de $\mathcal{B}(\mathcal{F}, P_0)$ par :

$$\begin{aligned}(\mathrm{d}T_A(P).\psi)(t) = &- \int_{]0,t]} \frac{\psi(f_\infty) - \psi(f_{s,h})}{P([s, +\infty[\times[\![0, k]\!])^2} \mathrm{d}P([0, s] \times \{j\}) \\ &+ \int_{]0,t]} \frac{\psi(f_{s,h,j})\, \mathrm{d}P([s, +\infty[\times[\![0, k]\!])}{P([s, +\infty[\times[\![0, k]\!])P(]s, +\infty[) \times [\![0, k]\!])} \\ &+ \frac{\psi(f_{t,h,j})}{P(]t, +\infty[\times[\![0, k]\!])} - \frac{\psi(f_{0,h,j})}{P(]0, +\infty[\times[\![0, k]\!])}\end{aligned}$$

De plus, pour toute mesure signée Q, on a la formule (où $h \neq j$) :

$$\begin{aligned}\big(\mathrm{d}T_A(P).Q\big)_{hj}(t) = &- \int_{]0,t]} \frac{Q_h([s, +\infty[\times[\![0, k]\!])}{P_h([s, +\infty[\times[\![0, k]\!])^2} \mathrm{d}P_h([0, s] \times \{j\}) \\ &+ \int_{]0,t]} \frac{\mathrm{d}Q_h([0, s] \times \{j\})}{P_h([s, +\infty[\times[\![0, k]\!])},\end{aligned}$$

cette expression étant valable pour la Hadamard-différentiabilité tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$ ou pour la Gâteaux-différentiabilité suivant les Dirac. Par composition, on en déduit que T est Gâteaux-différentiable suivant les Dirac et Hadamard-différentiable tangentiellement à $\mathcal{B}(\mathcal{F}, P_0)$ et que sa différentielle et donnée pour toute mesure signée Q par :

$$(\mathrm{d}T(P).Q)(t) = \int_0^t \prod_{[0,s[} (\mathrm{Id} + \mathrm{d}T_A(P)) \big(\mathrm{d}T_A(P).Q\big)(\mathrm{d}s) \prod_{]s,t]} (\mathrm{Id} + \mathrm{d}T_A(P)).$$

Pour pouvoir appliquer le théorème 9.6, il nous faut vérifier des conditions sur $\mathrm{d}T(P_0).(\delta_{X_i} - P_0)$. Cependant, cet objet doit être considéré comme une fonction à valeurs vectorielles et non à valeurs matricielles. Nous noterons donc par la suite m_i le vecteur colonne dont les composantes sont celles de $\mathrm{d}T(P_0).(\delta_{X_i} - P_0)$ réordonnés par lignes et $\widehat{m}_i$ celui correspondant à $\mathrm{d}T(\mathbb{P}_n).(\delta_{X_i} - P_0)$. D'après l'équation (11.3), l'expression de $\mathrm{d}T$ se simplifie lorsqu'on l'évalue en P_0 et on trouve :

$$\begin{aligned}(\mathrm{d}T(P_0).(\delta_{X_i} - P_0))(t) &= \int_0^t P(0, s^-) \big(\mathrm{d}T_A(P_0).(\delta_{X_i} - P_0)\big)(\mathrm{d}s) P(s, t) \\ &= \int_0^t P(0, s^-) \big(\mathrm{d}T_A(P_0).\delta_{X_i}\big)(\mathrm{d}s) P(s, t). \qquad (11.6)\end{aligned}$$

Notons B la matrice $\mathrm{d}T_A(P_0).\delta_{X_i}$ et b_{hj} son terme général. Notons de même p_{hj} le terme général de P. Alors, le terme général de (11.6) est donné par :

$$a_{hj}(t) = \sum_{l=1}^{k} \sum_{m=1}^{k} \int_0^t p_{hl}(0, s^-) p_{mj}(s, t)\, \mathrm{d}b_{ml}(s). \qquad (11.7)$$

Supposons maintenant que X_i corresponde à l'observation d'une transition d'un état h_0 à un état j_0. Alors, presque tous les termes dans (11.7) sont nuls et $a_{hj}(t)$ se simplifie en :

$$a_{hj}(t) = \int_0^t \left(p_{hj_0}(0, s^-) - p_{hh_0}(0, s^-)\right) p_{h_0,j}(s, t)\, \mathrm{d}b_{h_0,j_0}(s). \tag{11.8}$$

On a déjà rencontré des quantités de ce type à la section 10.3 lorsqu'on a étudié la fonction de risque cumulé dans le cadre de durées de vie censurées. On avait alors montré que les b_{ml} pouvaient s'écrire comme différence de deux fonctions croissantes bornées. Il en est donc de même pour les a_{hj} car les fonctions p_{hj} sont majorées par 1. En particulier :

$$\forall (h, j) \in [\![1, k]\!]^2 \ : \quad |a_{hj}(t)| \leqslant 2 \sup_{t\in[0,\tau]} |(\mathrm{d}T_A(P_0).\delta_{X_i})(t)| \leqslant \mathrm{C}^{\mathrm{ste}} < +\infty. \tag{11.9}$$

Alors, par stabilité par combinaison linéaire finie, les classes ci-dessous sont euclidiennes bornées :

$$\mathcal{F}_{n\omega} := \left\{ \big(m_1(\omega, t), \dots, m_{k^2}(\omega, t)\big),\ t \in [\tau_1, \tau] \right\}.$$

On en déduit, par stabilité par produit des classes euclidiennes, qu'il existe une fonction V à valeurs matricielles telle que :

$$\sup_{t\in[0,\tau]} \left| \frac{1}{n} \sum_{i=1}^n m_i^{\otimes 2}(t) - V(t) \right| \xrightarrow{\mathbb{P}} 0. \tag{11.10}$$

De plus, d'après (11.9), on a :

$$\sup_{t\in[0,\tau]} \sup_{\theta\in\Theta} \langle V(t)\theta, \theta\rangle < +\infty,$$

où on a noté Θ la sphère unité de $\mathbb{R}^{k^2}$. Cependant, il se peut que $V(\cdot)$ admette des valeurs propres nulles (cf. équation (11.8)). Dans ce cas, on peut perturber les $\widehat{m}_i$ comme on l'a expliqué à la section 5.4.2 et considérer le rapport de vraisemblance empirique ci-dessous, défini pour tout $t \in [0, \tau]$:

$$\widetilde{\mathrm{EL}_n^{LL}}(t) := \max \left\{ \prod_{i=1}^n n p_i,\ \sum_{i=1}^{k^2} p_i \widetilde{m}_i(t) \text{ et } p \in \mathbb{S}_{k^2-1} \right\}.$$

Théorème 11.3

Soit Θ la sphère unité de $\mathbb{R}^{k^2}$. Supposons qu'il existe un réel τ_1 tel que :

$$0 < \inf_{t\in[\tau_1,\tau]} \inf_{\theta\in\Theta\cap \mathrm{Im}\, V(t)} \mathbb{E}\left|\langle m_i(t), \theta\rangle\right| \leqslant \sup_{t\in[\tau_1,\tau]} \sup_{\theta\in\Theta\cap \mathrm{Im}\, V(t)} \left|\langle V(t)\theta, \theta\rangle\right| < +\infty.$$

Alors, il existe un processus gaussien centré G tel que :

$$-2 \ln \widetilde{\mathrm{EL}_n^{LL}}(\cdot) \rightsquigarrow G(\cdot) V^\dagger(\cdot) G(\cdot), \quad \text{lorsque } n \to +\infty.$$

Démonstration. Il suffit de démontrer que toutes les conditions du théorème 3.11 sont remplies. On sait d'après ce qui a été vu à la section 5.4.2 qu'il suffit de vérifier les conditions (A1*) à (A3*) pour les $\widehat{m}_i$ puisque la remarque 9.7 nous assure que :

$$\max_{1\leqslant i\leqslant n} \|\widehat{m}_i - m_i\|_\infty \xrightarrow{\mathbb{P}} 0, \quad \text{lorsque } n \to +\infty.$$

De plus, la démonstration du théorème 9.6 implique que les conditions (A1*) à (A3*) sont remplies puisque (B2*) est vérifiée d'après (11.10) et que les conditions sur les valeurs propres de V sont présentes dans les hypothèses. □

Troisième partie

Méthodologie et algorithmes

CHAPITRE 12

Loi Exponentielle Bivariée

Pour simuler le processus de comptage dans nos algorithmes, nous avons choisi d'utiliser des lois exponentielles bivariées. Nous donnons dans ce chapitre quelques propriétés élémentaires de ces lois qui nous seront utiles par la suite.

Définition 12.1. Soit $(\lambda_1, \lambda_2, \lambda_{12}) \in \mathbb{R}_+^3$. On dit que (X, Y) suit une loi bivariée exponentielle de paramètres $(\lambda_1, \lambda_2, \lambda_{12})$ s'il existe $U \sim \mathcal{E}(\lambda_1)$, $V \sim \mathcal{E}(\lambda_2)$ et $W \sim \mathcal{E}(\lambda_{12})$ mutuellement indépendantes telles que $X = U \wedge W$ et $Y = V \wedge W$. On notera $(X, Y) \sim \mathsf{BVE}(\lambda_1, \lambda_2, \lambda_{12})$

Lemme 12.2. *Si* $(X, Y) \sim \mathsf{BVE}(\lambda_1, \lambda_2, \lambda_{12})$, *alors, pour tout* $(x, y) \in \mathbb{R}_+^2$, *on a :*

$$\mathbb{P}(X > x, Y > y) = \exp(-\lambda_1 x - \lambda_2 y - \lambda_{12} \max(x, y)).$$

Démonstration. Il suffit d'écrire :

$$\begin{aligned}\mathbb{P}(X > x, Y > y) &= \mathbb{P}(U > x, W > x, V > y, W > y) \\ &= \mathbb{P}(U > x)\mathbb{P}(V > y)\mathbb{P}(W > \max(x, y)).\end{aligned}$$

□

Remarque 12.3. On peut remplacer tout ou partie des inégalités du lemme 12.2 par des inégalités larges.

Rappel 12.4. On rappelle que si Z et T sont deux variables aléatoires indépendantes de lois exponentielles de paramètres respectifs λ et μ, alors $Z \wedge T$ suit une loi exponentielle de paramètre $\lambda + \mu$.

Propriété 12.5. *Si* $(X, Y) \sim \mathsf{BVE}(\lambda_1, \lambda_2, \lambda_{12})$, *alors, pour tout* $(x, y) \in \mathbb{R}_+^2$, *on a :*

$$\mathbb{P}(Y > y | X = x) = \begin{cases} \exp(-\lambda_2 y) & \text{si } y < x, \\ \frac{\lambda_1}{\lambda_1 + \lambda_{12}} \exp(\lambda_{12} x) \exp(-(\lambda_2 + \lambda_{12}) y) & \text{si } y > x. \end{cases}$$

Démonstration. Supposons que $y > x$. Soit $h > 0$ suffisamment petit pour que $y > x + h$. On a alors grâce au lemme 12.2 :

$$\begin{aligned}\mathbb{P}(Y > y | X \in [x, x+h]) &= \frac{\mathbb{P}(Y > y, X \geqslant x) - \mathbb{P}(Y > y, X > x+h)}{\mathbb{P}(U \wedge W \in [x, x+h])} \\ &= \frac{\exp(-\lambda_1 x - \lambda_2 y - \lambda_{12} y)(1 - \exp(-\lambda_1 h))}{\exp(-(\lambda_1+\lambda_{12})x)(1 - \exp(-(\lambda_1+\lambda_{12})h))} \\ &= \frac{1 - \exp(-\lambda_1 h)}{1 - \exp(-(\lambda_1+\lambda_{12})h)} \exp(\lambda_{12} x - (\lambda_2 + \lambda_{12}) y).\end{aligned}$$

d'où le résultat en passant à la limite lorsque h tend vers 0. Supposons maintenant que $y < x$. Soit $h > 0$, on obtient alors, toujours grâce au lemme 12.2 :

$$\begin{aligned}\mathbb{P}(Y > y | X \in [x, x+h]) &= \frac{\mathbb{P}(Y > y, X \geqslant x) - \mathbb{P}(Y > y, X > x+h)}{\mathbb{P}(U \wedge W \in [x, x+h])} \\ &= \frac{\exp(-\lambda_1 x - \lambda_2 y - \lambda_{12} x)(1 - \exp(-\lambda_{12} h))}{\exp(-(\lambda_1+\lambda_{12})x)(1 - \exp(-\lambda_{12} h))} \\ &= \exp(-\lambda_2 y).\end{aligned}$$

□

Corollaire 12.6. *Soit $(X, Y) \sim \mathrm{BVE}(\lambda_1, \lambda_2, \lambda_{12})$ et $x > 0$. Alors, une densité de la partie continue de la loi conditionnelle de Y sachant $[X = x]$ est donnée par :*

$$f_Y^{X=x}(y) = \lambda_2 \mathrm{e}^{-\lambda_2 y} \mathbb{1}_{\{y<x\}} + \frac{\lambda_1(\lambda_2+\lambda_{12})}{\lambda_1+\lambda_{12}} \mathrm{e}^{\lambda_{12} x} \mathrm{e}^{-(\lambda_2+\lambda_{12})y} \mathbb{1}_{\{y>x\}}.$$

De plus :

$$\mathbb{P}(Y = x | X = x) = \frac{\lambda_{12}}{\lambda_1+\lambda_{12}} \mathrm{e}^{-\lambda_2 x}.$$

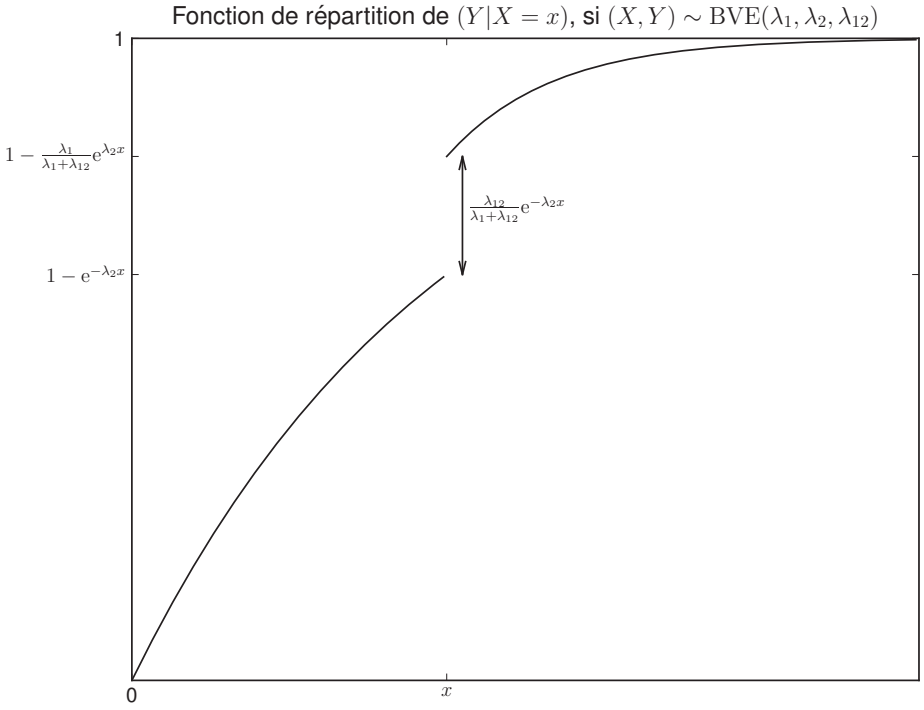
Propriété 12.7. *Soit $(X, Y) \sim \mathrm{BVE}(\lambda_1, \lambda_2, \lambda_{12})$. Alors :*

$$\begin{aligned}\mathbb{P}(Y = X) &= \frac{\lambda_{12}}{\lambda_1+\lambda_2+\lambda_{12}}, \\ \mathbb{P}(Y > X) &= \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_{12}}.\end{aligned}$$

Démonstration. En utilisant le corollaire 12.6, on trouve :

$$\begin{aligned}\mathbb{P}(Y = X) &= \int \mathbb{P}(Y = X | X = x) \mathrm{d}\mathbb{P}_X(x) \\ &= \int_0^{+\infty} \frac{\lambda_{12}}{\lambda_1+\lambda_{12}} \mathrm{e}^{-\lambda_2 x} (\lambda_1+\lambda_{12}) \mathrm{e}^{-(\lambda_1+\lambda_{12})x} \mathrm{d}x \\ &= \frac{\lambda_{12}}{\lambda_1+\lambda_2+\lambda_{12}}.\end{aligned}$$

L'autre égalité se déduit de cette première propriété ainsi que de la propriété 12.5. □



Propriété 12.8. *Soit* $(X, Y) \sim \mathrm{BVE}(\lambda_1, \lambda_2, \lambda_{12})$. *Alors, le coefficient de corrélation de* (X, Y) *est donné par :*

$$\rho(X, Y) = \frac{\lambda_{12}}{\lambda_1 + \lambda_2 + \lambda_{12}}.$$

Lemme 12.9. *Soit* $Z \sim \mathcal{E}(\lambda)$ *et* $w > 0$. *Alors :*

$$\mathbb{E}(Z \wedge w) = \frac{1}{\lambda}\left(1 - e^{-\lambda w}\right).$$

Démonstration du lemme 12.9. Il suffit de remarquer que :

$$\mathbb{P}(Z \wedge w \leqslant t) = \begin{cases} 1 - e^{-\lambda t} & \text{si } t < w, \\ 1 & \text{sinon,} \end{cases}$$

puis de calculer l'espérance grâce à une intégration par parties. □

Démonstration de la propriété 12.8. Soit $w > 0$, alors, en gardant les mêmes notations que précédemment :

$$\begin{aligned} \mathbb{E}(XY|W = w) &= \mathbb{E}(U \wedge w \cdot V \wedge w|W = w) \\ &= \frac{1}{\lambda_1\lambda_2}\left(1 - e^{-\lambda_1 w} - e^{-\lambda_2 w} - e^{-(\lambda_1+\lambda_2)w}\right). \end{aligned}$$

Or, si β est un réel quelconque, on a :

$$\int e^{-\beta w} d\mathbb{P}_W(w) = \frac{\lambda_{12}}{\beta + \lambda_{12}} \int_0^{+\infty} (\beta + \lambda_{12}) e^{-(\beta+\lambda_{12})w} dw = \frac{\lambda_{12}}{\beta + \lambda_{12}}.$$

Ainsi :

$$\mathbb{E}(XY) = \frac{1}{\lambda_1\lambda_2}\left(1 - \frac{\lambda_{12}}{\lambda_1+\lambda_{12}} - \frac{\lambda_{12}}{\lambda_2+\lambda_{12}} + \frac{\lambda_{12}}{\lambda_1+\lambda_2+\lambda_{12}}\right).$$

Le résultat voulu s'obtient alors après quelques calculs élémentaires. □

CHAPITRE 13

Algorithmes

Tous les algorithmes présentés ici ont été écrits en **Python**. Ils reposent sur la rapidité de `numpy` et de `cvxopt` ainsi que sur la simplicité de **Python**.

13.1 Le module `emplik`

13.1.1 Présentation

Nous avons écrit un module **Python** `emplik`, qui est l'équivalent du paquet éponyme sous **R**. Il utilise le module `cvxopt` d'optimisation convexe écrit par L. Vandenberghe et permet de résoudre le problème d'optimisation ci-dessous :

$$\max \prod_{i=1}^{n} np_i \ , \quad \text{avec} \begin{cases} \sum_{i=1}^{n} p_i X_i = n\mu, \\ \forall i \in [\![1, n]\!] : p_i \geqslant 0, \\ \sum_{i=1}^{n} p_i = n, \end{cases}$$

où X, n et μ sont les données du problème. Le contenu du module `emplik` est pratiquement déjà écrit dans la documentation de `cvxopt`.

```
#!/usr/bin/python
#-*- coding: utf-8 -*-

Resout le probleme d'optimisation suivant :

    max_p \prod p_i  ,    sous les contraintes :
    \sum p_i = n
    \sum p_i X_i = n*mu

Commande a lancer :
    acent(contraintes(X, len(X), mu))
```

```
from cvxopt import solvers, matrix, spdiag, log
solvers.options["maxiters"] = 15
solvers.options["show_progress"] = False

def acent(A, b):
    m, n = A.size
    def F(x = None, z = None):
        if x is None: return 0, matrix(1.0, (n, 1))
        if min(x) <= 0.0: return None
        f = -sum(log(x))
        Df = -(x ** -1).T
        if z is None: return f, Df
        H = spdiag(z[0] * x ** -2)
        return f, Df, H
    return solvers.cp(F, A = A, b = b)

def contraintes(X, n, mu):
    A = matrix(1., (1, n))
    A = matrix([A, X])
    b = matrix([float(n), n * mu])
    return A, b

def contraintes2d(Mnms1, Mnms2, mu1, mu2, n):
    A = matrix(1., (1, n))
    A = matrix([A, Mnms1, Mnms2])
    b = matrix([float(n), n * mu1, n * mu2])
    return A, b
```

13.1.2 Réduction de la dimension

Il peut être intéressant dans le cas où beaucoup d'observations ont la même valeur de réduire la dimension du problème d'optimisation rencontré en vraisemblance empirique. Considérons le problème d'optimisation suivant :

$$\min_{p\in\mathbb{R}^n} -\sum_{i=1}^{n} \ln(p_i) \ , \quad \text{avec} \ \begin{cases} \sum_{i=1}^n p_i X_i = n\mu, \\ \sum_{i=1}^n p_i = n. \end{cases} \tag{13.1}$$

On suppose qu'il y a des ex-æquo dans les données. Notons alors $(Y_1, \dots, Y_m)$ les valeurs distinctes prises par les X_i, et, pour $j \in [\![1, m]\!]$, notons n_j le cardinal de l'ensemble $\{i \in [\![1, n]\!] : X_i = Y_j\}$. La valeur Y_j est alors observée n_j fois. On montre facilement que le problème (13.1) est équivalent au problème ci-dessous :

$$\min_{q\in\mathbb{R}^m} -\sum_{j=1}^{m} n_j \ln(q_j) \ , \quad \text{avec} \ \begin{cases} \sum_{j=1}^m n_j q_j Y_j = n\mu, \\ \sum_{j=1}^m n_j q_j = n. \end{cases}$$

13.2 Le module `escalier`

Les processus considérés dans nos résultats sont des processus en escalier càdlàg. Il a donc fallu trouver une manière simple et efficace pour travailler avec de telles fonctions. Nous avons choisi de représenter ces fonctions sous forme de tableaux. Prenons par exemple la fonction f définie ci-dessous :

$$f : \mathbb{R} \longrightarrow \mathbb{R}$$
$$x \mapsto \begin{cases} -1 & \text{si } x < -2, \\ 2 & \text{si } -2 \leqslant x < 1, \\ -3 & \text{si } 1 \leqslant x < 3, \\ -5 & \text{sinon.} \end{cases}$$

Cette fonction sera représentée par le tableau :

$$\begin{bmatrix} -2 & 1 & 3 & +\infty \\ -1 & 2 & -3 & -5 \end{bmatrix}.$$

Le module `escalier` permet de réaliser des opérations élémentaires sur ces fonctions : somme, moyenne, troncature, supremum, etc ...

```
#!/usr/bin/python
#-*- coding: utf-8 -*-
```

```
Module pour travailler avec les fonctions en escalier
```

```
3 from __future__ import division
4 import numpy as np

6 def moins_f(f):
7     return np.array([f[0], -f[1]])

9 def fDeX(f, x):
10    return f[1][f[0] > x][0]

12 def un_sur_f(f):
13    return np.array([f[0], np.true_divide(1, f[1])])

15 def f_plus_g(f, g):
16    h0 = np.unique1d(np.append(f[0], g[0]))  # trie supprime les doublons
17    h1 = np.array([])
18    for x in np.append(-np.inf, h0[: -1]):
19        h1 = np.append(h1, fDeX(f, x) + fDeX(g, x))
20    return np.array([h0, h1])
```

```
22 def f_moins_g(f, g):
23     return f_plus_g(f, moins_f(g))

25 def sup_f(f):
26     return np.max(f[1])

28 def f_moyenne(liste, moy0 = False):
29     |Calcule la moyenne de n fonctions
30     n = len(liste)
31     if  not moy0.any():
32         moy0 = np.array([])
33         for f in liste:
34             moy0 = np.append(moy0, f[0])
35         moy0 = np.unique1d(moy0)
36     moy1 = np.array([])
37     for x in np.append(-np.inf, moy0[: -1]):
38         y = 0
39         for f in liste:
40             y+ = fDeX(f, x)
41         moy1 = np.append(moy1, y)
42     return np.array([moy0, moy1/n])

44 def tronque(f, a, b):
45     i1 = f[0] > a
46     i2 = f[0] < b
47     i = i1 * i2
48     f0 = np.append(np.append(a, f[0][i]), b)
49     f1 = np.append(np.append(fDeX(f, a), f[1][i]), fDeX(f, b))
50     return np.array([f0, f1])
```

13.3 Simulation du processus de comptage

Soient $T_1 < ... < T_n$ les instants de saut du processus. On suppose que (T_1, D) suit une loi exponentielle bivariée de paramètres $(\lambda_1, \lambda_D, \lambda_{12})$. On suppose aussi que les couples $(T_i, T_{i+1} - T_i)$ suivent la loi BVE $(\lambda_1, \lambda_2, \lambda_{12})$. Pour simuler le processus de comptage, il faut donc pouvoir simuler une loi bivariée exponentielle ainsi qu'une loi bivariée exponentielle conditionnelle. Les propriétés vues au chapitre 12 permettent de le faire. En effet, en utilisant la méthode d'inversion ainsi que la propriété 12.5, tout se passe sans heurt.

```
#!/usr/bin/python
#-*- coding: utf-8 -*-
```

```
1 from __future__ import division
2 from escalier import *
3 import numpy as np
4 import emplik
5 from cvxopt import matrix
6 import matplotlib.pyplot as plt

8 def bve(l1, l2, l12):
9     u = np.random.exponential(1/l1)
10    v = np.random.exponential(1/l2)
11    if l12 == 0:
12        return u, v
13    else:
14        w = np.random.exponential(1/l12)
15        return min(u, w), min(v, w)

17 def bve_cond(x, l1, l2, l12):
18    u = np.random.random()
19    if u <= 1 - np.exp(-l2 * x):
20        return -np.log(u)/l2
21    elif 1 - l1/(l1 + l12) * np.exp(-l2 * x) < u:
22        return -(np.log((u) * (l1 + l12)/l1) - l12 * x)/(l2 + l12)
23    else:
24        return x

26 def censure(c, D, noCensure = False):
27    if noCensure:
28        return D + 1
29    else:
30        return c * np.random.random()

32 def traj(l1, l2, l12, lD, tau, c):
33    |Simule une trajectoire du processus a sauts
      |que l'on arrete a D\wedge C
35    l = l12 * (1 + lD/l1)/2.
36    T1, D = bve(l1, lD, l)
37    C = censure(c, D)
38    if C < D:
39        X = C
40        delta = 0
41    else:
42        X = D
43        delta = 1
44    T = np.array([T1])
45    while True:
46        last = T[-1]
```

```
47          if last > X ∨ last > tau:
48              return T[: −1], delta, X
49          T = np.append(T, last + bve_cond(last, I1, I2, I12))
```

13.4 Calcul des divers estimateurs

Étant donné qu'on estime la loi limite par bootstrap, il faut tenir compte des ex-æquo dans nos algorithmes.

```
#!/usr/bin/python
#-*- coding: utf-8 -*-
```

```
Calcul des divers estimateurs
```

```
 3 from __future__ import division
 4 from escalier import *
 5 import numpy as np

 7 def f_hChapeau(X, delta, n, censure):
 8     | Calcule hChapeau
 9     if not censure:
10         return np.array([[np.inf], [1]])
11     indices = np.argsort(X)
12     X2 = X[indices]
13     delta2 = delta[indices]
14     sauts_unique = np.unique(X)
15     Z = []
16     nbMorts = []
17     for i in sauts_unique:
18         nbMorts.append(np.sum(delta2[X2 == i]))
19         Z.append(sum(X == i))
20     nbRisque = np.insert((n − np.cumsum(Z))[: −1], 0, n)
21     nbMorts = np.array(nbMorts)
22     Hc0 = np.append(sauts_unique, np.inf)
23     Y1 = np.append(nbRisque, 0)/n
24     S1 = np.insert(np.cumprod(1 − nbMorts/nbRisque), 0, 1)
25     return np.array([Hc0, S1/Y1])

28 def f_mij2(X, Ni, delta, Hc, n, tau, censure):
```

```
29      mij2 = m_ij + mu_j
30      M2 = [fDeX(Hc, tk) for tk in Ni[0][: -1]]
31      M2.insert(0, 0)
32      M2 = np.cumsum(M2)
33      M = np.array([Ni[0], M2])
34      return tronque(M, 0.5, tau)

37 def f_M(X, N, delta, n, tau, censure):
38      M == vecteur des mij2
39      Hc = f_hChapeau(X, delta, n, censure)
40      return [f_mij2(X, Ni, delta, Hc, n, tau, censure) for Ni in N]

43 def f_mujchap(M):
44      Calcule le fameux \hat{\mu}_j
45      return f_moyenne(M, moy0 = np.arange(0.5, tau, .05))
```

13.5 Programme principal

Dans l'algorithme principal, B désigne le nombre d'itérations bootstrap et n la taille de l'échantillon. On désigne par censure un booléen qui indique si il y a une censure et par c le réel tel que $C \sim \mathcal{U}[0, c]$. Les paramètres étant donnés, l'algorithme simule un processus de comptage et calcule $\widehat{\mu}$. Il estime ensuite par bootstrap les quantiles d'ordre 90%, 95% et 99% de la loi limite. Enfin, il teste grâce à la fonction `is_in_tube` l'appartenance de la fonction $\widehat{\mu}$ au tube de confiance. Cet algorithme est appelé 10000 fois mais pour en simplifier la lecture, on n'a laissé qu'une itération. De même, les garde-fous dans la dichotomie ont été enlevés, d'autant plus qu'ils se sont révélés inutiles en pratique.

#!/usr/bin/python
#-*- coding: utf-8 -*-

Algorithme principal

```
3 from __future__ import division
4 import numpy as np
5 import emplik

7 def new_moins_2_log(X, n, t):
```

```
 8     | Calcule la vraisemblance empirique au point t
 9     A, b = emplik.contraintes(X, n, t)
10     if max(X) == min(X):
11         return np.inf
12     else:
13         try:
14             res = emplik.acent(A, b)
15         except :
16             return np.inf
17         if res['status'] ≠ "optimal":
18             return np.inf
19         else:
20             return 2 * res['primal objective']

23 def f_Zetoile(N, X, delta, n):
24     | Cree un echantillon de Z* en tirant
       |    uniformement dans {Z_1,...,Z_n}
26     indicesEtoile = np.random.randint(n, size = n)
27     N_e = [N[i] for i in indicesEtoile]
28     X_e = X[indicesEtoile]
29     delta_e = delta[indicesEtoile]
30     return N_e, X_e, delta_e

33 def un_bootstrap(N, X, delta, tau, n, mujchap):
34     | Effectue un bootstrap a partir d'une trajectoire
35     N_e, X_e, delta_e = f_Zetoile(N, X, delta, n)
36     M_e = f_M(X_e, N_e, delta_e, n, tau, censure)
37     return moins_2_log(M_e, n, tau, mujchap)

40 def une_simulation(l1, l2, l12, lD, tau, c, n, B, censure):
41     Veboot = np.array([])
42     N, X, delta = f_N(l1, l2, l12, lD, tau, c, n, censure)
43     newtau = np.max(X)   # Inutile de faire les calculs au-dela !
44     M = f_M(X, N, delta, n, newtau, censure)
45     mujchap = f_mujchap(M)
46     for i in range(B):
47         b = un_bootstrap(N, X, delta, newtau, n, mujchap)
48         Veboot = np.append(Veboot, sup_f(b))
49     ecdf = np.sort(Veboot)
50     u_alpha = np.array([ecdf[int(B * .99)]])
51     u_alpha = np.append(u_alpha, ecdf[int(B * .95)])
52     u_alpha = np.append(u_alpha, ecdf[int(B * .9)])
53     res = []
54     for u in u_alpha:
55         res.append(is_in_tube(N, X, delta, M, mujchap, u, newtau, n))
56     np.savetxt("result.txt", np.array(res))
```

```
59 def is_in_tube(N, X, delta, M, mujchap, u_alpha, tau, n):
60     Retourne vrai si muj est dans le tube, faux sinon
61     x = np.arange(0.5, tau - .0000001, .05)
62     for xi in x:
63         mujchapxi = fDeX(mujchap, xi)
64         muxi = fDeX(mu, xi)
65         Mnms = np.array([fDeX(mij2, xi) for mij2 in M])
66         AA = matrix(Mnms).trans()
67         ab = np.array([0, mujchapxi])
68         y = new_moins_2_log(AA, n, 0)
69         while abs(y - u_alpha) > .05:
70             c = np.mean(ab)
71             y = new_moins_2_log(AA, n, c)
72             if y < u_alpha:
73                 ab[1] = c
74             else:
75                 ab[0] = c
76         if muxi < c:
77             return False
78         ab = np.array([mujchapxi, 800])
79         y = new_moins_2_log(AA, n, 800)
80         while abs(y - u_alpha) > .05:
81             c = np.mean(ab)
82             y = new_moins_2_log(AA, n, c)
83             if y > u_alpha:
84                 ab[1] = c
85             else:
86                 ab[0] = c
87         if muxi > c:
88             return False
89     return True
```

CHAPITRE 14

Bande de confiance

La construction d'une bande de confiance se fait comme dans l'algorithme nommé `is_in_tube`. À un jour k fixé, on part de $\widehat{\mu}_j(k)$ et on effectue deux dichotomies : une vers le bas et une vers le haut. Comme on sait que le rapport de vraisemblance empirique est positif, on peut effectuer la première dichotomie sur $[0, \widehat{\mu}_j(k)]$. Par contre, il n'y a pas *a priori* de borne sup, on a donc choisi empiriquement l'intervalle $[\widehat{\mu}_j(k), 5]$.

Encore une fois, pour simplifier la lecture de l'algorithme, on a enlevé certains paramètres puisque la fonction originale prenait par exemple en argument l'infection nosocomiale concernée.

Les paramètres de la fonction sont les données (N, X, δ), la taille de l'échantillon (n), l'intervalle sur lequel on doit tracer la bande de confiance ($[t_1, t_2]$) ainsi que le quantile u_α estimé au préalable par bootstrap.

```
#!/usr/bin/python
#-*- coding: utf-8 -*-
```

```
Bande de confiance
```

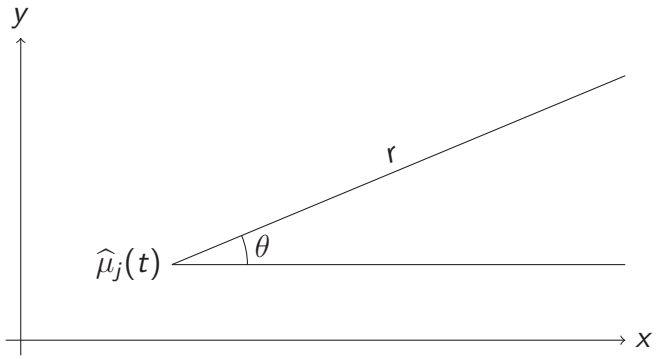
```
3 from __future__ import division
4 import numpy as np

6 def une_simulation(N, X, delta, n, t1, t2, u_alpha):
7     M = f_M(N, X, delta, n, t1, t2)
8     mujchap = f_mujchap(M, t1, t2)
9     binf = []
10    bsup = []
11    x = np.arange(t1, t2)
12    for t in x:
13        mujchapt = fDeX(mujchap, t)
14        ab = np.array([0, mujchapt])
```

```
15          y = moins_2_log(M, n, t, mujchapt)
16          while abs(y − u_alpha) > .1:
17              c = np.mean(ab)
18              y = moins_2_log(M, n, t, c)
19              if y < u_alpha:
20                  ab[1] = c
21              else:
22                  ab[0] = c
23          binf.append(c)
24          ab = np.array([mujchapt, 5])
25          y = moins_2_log(M, n, t, mujchapt)
26          while abs(y − u_alpha) > .1:
27              c = np.mean(ab)
28              y = moins_2_log(M, n, t, c)
29              if y > u_alpha:
30                  ab[1] = c
31              else:
32                  ab[0] = c
33          bsup.append(c)
34      binf = np.array([x, binf])
35      bsup = np.array([x, bsup])
36      np.savetxt("binf.txt", binf)
37      np.savetxt("bsup.txt", bsup)
```

CHAPITRE 15

Tube de confiance

La fonction ci-dessous prend en argument deux processus de comptage N_1 et N_2, le vecteur des X_i et des δ_i correspondant, le quantile calculé par bootstrap et le jour t auquel on souhaite calculer la région de confiance. Elle détermine ensuite pour un angle θ variant entre 0 et 2π l'intervalle de confiance dans la direction θ par dichotomie. Comme on sait que $\widehat{\mu}_j$ est dans la région de confiance, on le prend comme point de départ et on prend un rayon r suffisamment grand (ici on a choisi $r = 100$).




```
1  def une_simulation(N1, N2, X, delta, n, t1, t2, u_alpha, t):
2      M1, M2 = f_M(N1, N2, X, delta, n, t1, t2)
3      mujchap1, mujchap2 = f_mujchap(M1, M2, t1, t2)
4      Theta = np.arange(0, 2 * np.pi, 2 * np.pi/50)
5      r = []
6      for theta in Theta:
7          mujchap1_en_t = fDeX(mujchap1, t)
8          mujchap2_en_t = fDeX(mujchap2, t)
9          ab = np.array([0., 100])
10         y = moins_2_log(M1, M2, n, mujchap1_en_t, mujchap2_en_t, t)
11         while abs(y - u_alpha) > .1:
12             c = np.mean(ab)
```

```
13              y = moins_2_log(M1, M2, n, mujchap1_en_t + c * np.cos(theta),
                    mujchap2_en_t + c * np.sin(theta), t)
15              if y > u_alpha:
16                  ab[1] = c
17              else:
18                  ab[0] = c
19          r.append(c)
20      return np.array(r)
```

Conclusion et perspectives

Dans son livre, Owen (1990) pose les bases de la théorie de la vraisemblance empirique. Il explique comment obtenir des régions de confiance pour la moyenne d'un échantillon i.i.d. dans L^2. Il montre aussi comment, à partir d'une équation estimant un paramètre μ (de la forme $f(\mu) = 0$), on peut appliquer les méthodes de vraisemblance empirique pour déterminer des régions de confiance pour μ. Enfin, il propose une sorte de δ-méthode pour la construction de régions de confiances pour un paramètre s'exprimant comme une fonction différentiable de la moyenne. Dans ce travail de thèse, nous avons utilisé ces trois aspects dans le cadre d'événements récurrents, mais d'un point de vue des processus.

Nous avons dans une première partie énoncé des résultats nécessaires à notre étude et avons donné des preuves succinctes, parfois nouvelles, en espérant rendre ces chapitres auto-suffisants.

Dans une deuxième partie, nous avons énoncé les résultats principaux que nous avons pu démontrer, et qui sont tous basés sur une version plus générale du théorème d'Owen et adaptée à l'étude de processus : on la trouvera dans le papier de Hjort *et al.* (2004).

Dans les chapitres 6 et 7, nous avons construit des bandes de confiance par vraisemblance empirique à partir d'un processus moyen. Cette étude est à rapprocher du théorème d'Owen « classique ». Elle nous a permis d'étudier la fonction de risque cumulé dans deux cadres : celui de l'étude de risques concurrents avec censure aléatoire à droite et événement terminal, ainsi que celui de l'étude de processus de Markov non homogènes à temps continu et soumis à une censure aléatoire à droite. Cette étude nous a aussi permis de construire des bandes de confiance pour les fonctions moyennes dans le cadre d'événements récurrents avec risques concurrents, événement terminal et censure aléatoire à droite. Nous avons de plus appliqué ce résultat à un jeu de données traitant d'infections nosocomiales dans une unité de soins intensifs et avons réalisé des simulations sur des échantillons de taille finie pour juger de la qualité de notre modèle.

Dans le chapitre 8, nous nous sommes intéressé à la prise en compte de covariables pour la modélisation des événements récurrents. Nous avons alors cherché à construire une bande de confiance pour des paramètres de régression fonctionnels dans un cadre d'événements récurrents avec risques concurrents, covariables et censure aléatoire à droite. Nous avons alors ob-

tenu un résultat de convergence et une région de confiance. Cependant, ce résultat *semble* inexploitable dans la pratique.

Nous avons par la suite voulu construire des bandes de confiance par vraisemblance empirique pour la matrice de transition d'un processus de Markov non homogène à temps continu et soumis à une censure aléatoire à droite. Pour ce faire, nous avons dû énoncer un nouveau résultat de type δ-méthode fonctionnelle adaptée à l'étude de processus par vraisemblance empirique. Nous avons alors mis à contribution la théorie des classes de Vapnik et Chervonenkis pour contrôler le comportement du processus empirique pondéré qui apparaît naturellement dans ce cadre. Nous avons donné des régions de confiance constructibles numériquement, ce qui n'existait pas encore dans la littérature à notre connaissance. Enfin, nous avons pu appliquer ce résultat aux processus de Markov.

Ce travail de thèse permet de voir comment, dans diverses situations, on peut appliquer la vraisemblance empirique à des processus, en particulier aux processus des événements récurrents. Des questions restent cependant en suspend.

Par exemple, les résultats du chapitre 7 sont-ils encore vrais si on ne suppose plus les processus de comptage presque sûrement bornés ? Est-il possible, quitte à rajouter des conditions « peu contraignantes », d'exploiter le résultat trouvé au chapitre 8 pour déterminer une bande de confiance pour le vecteur de covariables ? Peut-on donner des conditions suffisantes « simples » pour montrer que les conditions du théorème présenté au chapitre 11 traitant des problèmes multi-états sont satisfaites ?

De plus, il paraît intéressant de voir comment appliquer à des problèmes de fiabilité le modèle traitant des processus de Markov et présenté au chapitre 11. En effet, les processus markoviens voir semi-markoviens sont très utilisés en fiabilité. Il serait sans doute intéressant d'utiliser les techniques de vraisemblance empirique pour construire des bandes de confiance pour les grandeurs d'intérêt.

Toujours concernant le résultat sur les processus de Markov, on pourrait s'essayer à conduire des simulations sur des échantillons de taille finie, et appliquer notre résultat au jeu de données d'infections nosocomiales. Nous avons en effet obtenu des résultats pour le nombre moyen d'infections nosocomiales contractées au cours du temps au chapitre 7. Or, il aurait aussi été intéressant de décrire l'état d'un patient au cours du temps par un processus de Markov et de construire une bande de confiance pour sa matrice de transition. Notons cependant que le nombre d'états dans la chaîne conditionne grandement les temps de calcul. Ainsi, construire des régions de confiance sur un jeu de données en estimant la loi limite par bootstrap reste raisonnable lorsqu'il n'y a que trois ou quatre états, mais vérifier la qualité des régions de confiance par Monte-Carlo s'avère pratiquement impossible : le temps de calcul pour remplir un tableau similaire à celui présenté au chapitre 7 serait de plusieurs dizaines d'années.

Au chapitre 9, nous avons expliqué comment construire des régions de confiance par vraisemblance empirique pour des fonctionnelles différen-

tiables. Nous avons considéré un ensemble de trajectoires comme espace de départ et l'ensemble des fonctions càdlàg comme espace d'arrivée. Nous avons donc utilisé la théorie des classes de Vapnik et Chervonenkis pour mener à bien notre démonstration. Cependant, plutôt que de travailler sur $\ell_\infty(\mathcal{F})$, nous aurions pu utiliser la théorie des processus empiriques de Pollard (1990) qui paraît plus naturelle étant donné l'espace d'arrivée de notre fonctionnelle. Il pourrait donc être intéressant de voir comment mettre en œuvre une telle approche.

Enfin, on pourrait s'intéresser aux propriétés théoriques du bootstrap de suprema de processus, en particulier pour le carré d'un processus gaussien. En effet, nous avons effectué au chapitre 7 un bootstrap pour estimer les quantiles d'une loi limite sans avoir vérifié la légitimité théorique de cette approche, bien qu'elle n'ait pas été contredite par nos simulations.

Quatrième partie

Annexes

Index des notations

$x \wedge y$: le minimum entre x et y ;

$x \vee y$: le maximum entre x et y ;

x^t : le vecteur transposé de x ;

$x \odot y$: le vecteur $(x_1 y_1, \dots, x_n y_n)$;

$x^{\otimes 2}$: la matrice xx^t ;

$A \succ 0$: signifie que la matrice A est définie positive ;

$A \succeq 0$: signifie que la matrice A est positive ;

$\|\cdot\|_{p,Q}$: la semi-norme définie par $\|f\|_{p,Q}^p = \int |f|^p \mathrm{d}Q$;

$\mathcal{B}_{\mathbb{R}}$: la tribu borélienne sur $\mathbb{R}$;

$\mathcal{B}(\mathcal{F}, Q)$: l'ensemble des fonctions bornées de $\ell_\infty(\mathcal{F})$ qui sont uniformément continues pour $\|\cdot\|_{2,Q}$:

F_n : la fonction de répartition empirique associée à $(X_1, \dots, X_n)$;

$\tilde{F}_n$: une version pondérée de F_n, i.e. de la forme $\sum_{i=1}^n p_i \mathbb{1}_{]-\infty,\cdot]}(X_i)$;

$\mathbb{P}_n$: la probabilité empirique associée à $(X_1, \dots, X_n)$;

$\widetilde{\mathbb{P}}_n$: une version pondérée de P_n, i.e. de la forme $\sum_{i=1}^n p_i \delta_{X_i}$;

$\mathbb{S}_{n-1}$: le simplexe unité de $\mathbb{R}^n$;

$V(f, \sigma)$: les variations de la fonction f sur la subdivision σ ;

$V_a^b(f)$: la variation totale de la fonction f sur l'intervalle $[a, b]$;

$X_n \rightsquigarrow X$: signifie que X_n converge en distribution vers X (cf. chapitre 1).

$X_n \xrightarrow{\mathbb{P}} X$: signifie que X_n converge en probabilité vers X ;

Index

Bibliographie

Aalen, O. (1978). Nonparametric inference for a family of counting processes. Ann. Statist. **6**, 701–726.

Adimari, G. (1997). Empirical likelihood type confidence intervals under random censorship. Ann. Inst. Statist. Math. **49**, 447–466.

Altshuler, B. (1970). Theory for the measurement of competing risks in animal experiments. Mathematical Biosciences **6**, 1–11.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1993). Statistical models based on counting processes. Springer Series in Statistics. Springer-Verlag, New York.

Bertail, P. (2006). Empirical likelihood in some semiparametric models. Bernoulli **12**, 299–331.

Bilias, Y., Gu, M. & Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. Ann. Statist. **25**, 662–682.

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Cai, J. & Schaubel, D. E. (2004a). Analysis of recurrent event data. In Advances in survival analysis, vol. 23 of Handbook of Statist. Elsevier, Amsterdam, pp. 603–623.

Cai, J. & Schaubel, D. E. (2004b). Marginal means/rates models for multiple type recurrent event data. Lifetime Data Anal. **10**, 121–138.

Cook, R. & Lawless, J. (2007). The statistical analysis of recurrent events. Springer.

Cook, R. J., Lawless, J. F., Lakhal-Chaieb, L. & Lee, K.-A. (2009). Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination : application to skeletal complications in cancer metastatic to bone. J. Amer. Statist. Assoc. **104**, 60–75.

Dauxois, J.-Y. & Sencey, S. (2009). Non-parametric tests for recurrent events under competing risks. Scand. J. Stat. **36**, 649–670.

Du, P. (2009). Nonparametric modeling of the gap time in recurrent event data. Lifetime Data Anal. **15**, 256–277.

Fleming, T. R. (1978). Asymptotic distribution results in competing risks estimation. Ann. Statist. **6**, 1071–1079.

Fleming, T. R. & Harrington, D. P. (1991). Counting processes and survival analysis. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York.

Hall, P. & La Scala, B. (1990). Methodology and algorithms of empirical likelihood. International Statistical Review/Revue Internationale de Statistique , 109–127.

Hand, D. J. (2008). The statistical analysis of recurrent events by richard j. cook, jerald f. lawless. International Statistical Review **76**, 142–143.

Harari-Kermadec, H. (2011). Regenerative block empirical likelihood for markov chains. Journal of Nonparametric Statistics **23**, 781–802.

Hjort, N. L., McKeague, I. W. & Van Keilegom, I. (2004). Extending the scope of empirical likelihood (pre-print). Ann. Statist. **37**, 1079–1111.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. **53**, 457–481.

Lawless, J. F. (1995). The analysis of recurrent events for multiple subjects. Journal of the Royal Statistical Society. Series C (Applied Statistics) **44**, pp. 487–498.

Lawless, J. F. & Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. Technometrics **37**, 158–168.

Liang, K. & Zeger, S. (1986). Longitudinal data analysis using generalized linear models. Biometrika **73**, 13–22.

Lin, D. Y., Sun, W. & Ying, Z. (1999). Nonparametric estimation of the gap time distributions for serial events with censored data. Biometrika **86**, 59–70.

Lin, D. Y., Wei, L. J., Yang, I. & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. Journal of the Royal Statistical Society. Series B (Statistical Methodology) **62**, pp. 711–730.

Lin, D. Y. & Ying, Z. (2001). Nonparametric tests for the gap time distributions of serial events based on censored data. Biometrics **57**, 369–375.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. Technometrics , 945–966.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. Annals of Statistics **18**, 90–120.

Pakes, A. & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. Econometrica **57**, 1027–1057.

Pollard, D. (1984). Convergence of stochastic processes. Springer.

Pollard, D. (1990). Empirical processes : theory and applications. NSF-CBMS Regional Conference Series in Probability and Statistics, 2. Institute of Mathematical Statistics, Hayward, CA.

Qin, G. & Jing, B.-Y. (2001). Empirical likelihood for censored linear regression. Scand. J. Statist. **28**, 661–673.

Ren, J.-J. (2001). Weighted empirical likelihood ratio confidence intervals for the mean with censored data. Ann. Inst. Statist. Math. **53**, 498–516.

Sauer, N. (1972). On the density of families of sets. Journal of Combinatorial Theory, Series A **13**, 145 – 147.

Shorack, G. & Wellner, J. (2009). Empirical processes with applications to statistics, vol. 59. Society for Industrial Mathematics.

Thomas, D. R. & Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. J. Amer. Statist. Assoc. **70**, 865–871.

Titchmarsh, E. (1986). The theory of functions, (reprinted and corrected version of the original 1939 2nd edition). Oxford University Press, New York.

van der Vaart, A. W. & Wellner, J. A. (1996). Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Wang, W. & Wells, M. T. (1998). Nonparametric estimation of successive duration times under dependent censoring. Biometrika **85**, 561–572.

Zhao, Y. & Hsu, Y.-S. (2005). Semiparametric analysis for additive risk model via empirical likelihood. Communications in Statistics - Simulation and Computation **34**, 135–143.

Résumé

Disposant d'un jeu de données sur des infections nosocomiales, nous utilisons des techniques de vraisemblance empirique pour construire des bandes de confiance pour certaines quantité d'intérêt. Cette étude nous amène à renforcer les outils déjà existants afin qu'ils s'adaptent à notre cadre.

Nous présentons dans une première partie les outils mathématiques issus de la littérature que nous utilisons dans ce travail de thèse. Nous les appliquons ensuite à diverses situations et donnons de nouvelles démonstrations lorsque cela est nécessaire. Nous conduisons aussi des simulations et obtenons des résultats concrets concernant notre jeu de données. Enfin, nous détaillons les algorithmes utilisés.

Mots-clés : vraisemblance empirique, processus empiriques, processus des événements récurrents, risques concurrents, censure, événement terminal, covariables, problèmes multi-états, classes de Vapnik-Chervonenkis, classes euclidiennes, δ-méthode.

Classification AMS : 60G51, 62N01, 62N02, 62N05, 62G08.

Abstract

The starting point of this thesis is a data set of nosocomial infections in an intensive care unit of a French hostipal. We focused our attention on building confidence bands for some parameters of interest using empirical likelihood techniques. In order to do so, we had to adapt and develop some already existing methods so that they fit our setup.

We begin by giving a state of the art of the different theories we use. We then apply them to different setups and demonstrate new results when needed. Finally, we conduct simulations and describe our algorithms.

Keywords: empirical likelihood, empirical processes, recurrent events, competing risks, censoring, terminal event, covariates, multi-states problems, Vapnik-Chervonenkis classes, euclidean classes, δ-method.

AMS Classification: 60G51, 62N01, 62N02, 62N05, 62G08.