Estimation des paramètres pour des modèles adaptés aux séries de records

Anis Hoayek

THÈSE

Pour obtenir le grade de

Docteur

Délivré par **UNIVERSITE de MONTPELLIER**

Préparée au sein de l'école doctorale

Information, Structures, Systèmes (I2S)

Et de l'unité de recherche

Institut Montpelliérain Alexander Grothendieck (IMAG)

Spécialité : **Biostatistique**

Présentée par **Anis HOAYEK**

Estimation des paramètres pour des modèles adaptés aux séries de records.

Soutenue le 25/11/2016 devant le jury composé de

M. Clément DOMBRY, Professeur, Université de Franche Comté	Rapporteur
M. Joseph NGATCHOU-WANDJI, Professeur, Université de Lorraine	Rapporteur
M. Jean-Noël BACRO, Professeur, Université de Montpellier	Examinateur
Mme. Zainab ASSAGHIR, Maitre de conférences, Université Libanaise	Examinateur
M. Gilles DUCHARME, Professeur, Université de Montpellier	Directeur
M. Hassan ZEINEDDINE, Professeur, Université Libanaise	Co-Directeur
M. Zaher KHRAIBANI, Maitre de conférences, Université Libanaise	Co-Encadrant

Avec toi, Marcellin, Pèlerin d'espérance

Pour de nouveaux matins, nous sommes la semence.

Avec toi, Marcellin, notre Saint d'espérance,

Nous bâtirons demain le monde de l'enfance.

Remerciements

Je remercie avec gratitude tous ceux qui m'ont aidé à réaliser cette thèse de doctorat :

- Dieu pour ses grâces ;
- Mon directeur de thèse professeur Gilles DUCHARME : je ne trouve pas les mots appropriés pour vous exprimer ma profonde gratitude et respect. Vous m'avez inspiré à devenir un chercheur indépendant et m'avez aidé à atteindre la puissance du raisonnement critique. Vous m'avez aussi montré ce qu'un scientifique brillant et travailleur sérieux peut accomplir ;
- Mon co-directeur professeur Hassan ZEINEDDINE et mon co-encadrant docteur Zaher KHRAIBANI : je vous remercie pour le temps que vous m'avez consacré et pour me proposer des commentaires afin d'améliorer mon travail ;
- Le CNRS Libanais et l'Université de Montpellier pour l'opportunité donnée et surtout docteur Charles TABET qui a cru en moi ;
- Les rapporteurs pour leur analyse critique et les examinateurs pour avoir accepté d'être membres du jury ;
- Mes collègues au laboratoire pour leur gentillesse et leur amitié pendant ces trois années de thèse ;
- Monsieur Jean ABI AKEL pour qui je dois ma réussite du fait qu'il m'a transmis la passion des mathématiques ;
- Mes parents pour leur encouragement continu et leur patience ;
- Et finalement ma femme Nicole.

Enfin, je tiens à vous ré exprimer mes sincères remerciements et mes sentiments les plus respectueux.

Résumé

Dans une série chronologique $\{X_t, t \geq 1\}$, une observation X_j est un record au temps j si sa valeur est supérieure à toutes les valeurs précédentes, c'est-à-dire $X_j > \max\{X_1, ..., X_{j-1}\}$. Dans cette thèse, nous considérons les records supérieurs mais les records inférieurs peuvent être définis de manière similaire, par exemple en multipliant les séries chronologiques par « -1 ». Suivant l'augmentation de t, considérons la suite des valeurs de records $\{R_n, n \geq 1\}$ et la suite des indices d'occurrence des records $\{L_n, n \geq 1\}$. Ici L_n est l'indice d'occurrence du $n^{ième}$ record de valeur $R_n = X_{L_n}$.

L'intérêt des records augmente quand ils sont les seuls valeurs accessibles dans une série chronologique donnée. Parce que les records font partie de la culture populaire, ils sont généralement conservés dans des lieux facilement accessibles, par exemple le livre « *Guinness World Records* » pour des événements sportifs et similaires. Les records sont également rencontrés dans la recherche sur le climat (Wergen et Krug, 2010), en épidémiologie (Khraibani *et al.*, 2015) et en finance (Wergen, 2014). Dans de telles applications, les données disponibles sont les couples $\{(R_n, L_n), n = 1, ..., N_T\}$, où T désigne le temps présent et N_T est le nombre de records dans $\{X_t, t = 1, ..., T\}$. Cette information limitée, a un coût : de nombreuses questions sur le comportement de la série chronologique restent sans réponses à moins d'information externe est disponible. Mais certaines autres questions importantes, telles que la prédiction de la valeur de record suivant R_{N_T+1} et son indice L_{N_T+1}, sont accessibles.

Les propriétés stochastiques des suites de valeurs de records ont été largement étudiés dans le cas où les X_t sont des variables aléatoires (*va*) indépendantes et identiquement distribuées (*iid*) (Arnold *et al.*, 1998 ; Nevzorov, 2001). Il se trouve que beaucoup de ces propriétés sont universelles, c'est-à-dire elles tiennent pour n'importe quelle loi de probabilité commune des X_t. Cela a permis une bonne compréhension du comportement des records. En

particulier, ils ont tendance à devenir plus séparés dans le temps quand t ou n augmente (Arnold *et al.*, 1998 page 26). Cependant, ce n'est pas ce que l'on observe dans de nombreux ensembles de données réelles. Par exemple, les progrès technologiques font que les records sportifs se produisent plus souvent que prévu par le cas *iid*. Ceci a conduit à l'élaboration de modèles plus complexes pour fournir une meilleure prédiction.

Le modèle, peut-être le plus simple mais en tout cas le plus populaire, pour une série de records issus d'observations indépendantes mais non identiquement distribuées est le modèle à dérive linéaire (LDM) :

$$X_t = Y_t + \theta t,$$

où $\{Y_t, t \geq 1\}$ est une suite de *va iid* et $\theta > 0$ est un paramètre à déterminer. Ce modèle a été étudié par de nombreux auteurs (Ballerini et Resnick, 1985 ; Borovkov, 1999 ; Franke *et al.*, 2010) et trouvé en accord avec certains types de données où l'hypothèse *iid* ne tient pas. Cependant, dans des situations pratiques, l'utilisation du LDM nécessite la détermination de θ et cela amène le problème dans le domaine des statistiques.

Il existe une similitude entre les records et le traitement de données censurées en analyse de survie. En particulier, toutes les valeurs de X_t entre R_n et R_{n+1}, et au-delà de R_{N_T} peuvent être considérées comme des observations censurées par le dernier record observé. Pour mettre en évidence cette similitude, considérons la suite $\{\delta_t, t \geq 1\}$ des indicatrices de records : $\delta_t = 1$ si t se produit dans la suite $\{L_n, n \geq 1\}$. Il existe une relation d'équivalence entre $\{\delta_t, t = 1, ..., T\}$ et $\{L_n, n = 1, ..., N_T\}$ de sorte que $\{(R_n, L_n), n = 1, ..., N_T\}$ contient les mêmes informations que $\{(W_t, \delta_t), t = 1, ..., T\}$, où $W_t = X_t$ si $\delta_t = 1$ et est censurée par le dernier record si $\delta_t = 0$.

Ceci a conduit Smith (1988) à proposer l'application des méthodes basées sur le principe du maximum de vraisemblance standard en analyse de survie à la suite (W_t, δ_t). Carlin et Gelfand (1993) donnent l'expression de la vraisemblance en termes de (R_n, L_n). Les procédures statistiques associées exigent la connaissance de la distribution jointe des Y_t. En outre, les propriétés des estimateurs résultants sont largement inconnus. Feuerverger et Hall (1996) ont suggéré l'estimation de θ par la méthode des moindres carrés. Leur méthode, basée uniquement sur la suite des valeurs de records $\{R_n, n = 1, ..., N_T\}$, est indépendante de la distribution sous-jacente (distribution des Y_t) et le comportement de leur estimateur peut être approximé asymptotiquement et par des techniques de bootstrap. Mais, la qualité de l'estimateur se dégrade avec

la diminution de θ. Cela est facile à comprendre : lorsque θ est petit, X_t a une faible probabilité d'être un record et dans des échantillons modérés, le nombre de données sur lesquelles ils appliquent leur approche est probablement faible. En outre, à notre connaissance, il n'y a pas de tests d'ajustement pour le LDM.

Un autre modèle populaire est le modèle Yang-Nevzorov :

$$X_t \sim F(\cdot)^{\rho_t},$$

où les ρ_t $(t \geq 1)$ sont des constantes réelles ≥ 1 et $F(\cdot)$ est une fonction de répartition. Ce modèle a été initialement proposé par Yang (1975) avec $\rho_t = \gamma^t$ et généralisé par Nevzorov (1990) à des suites $\{\rho_t\}$ générales. Ce dernier auteur a aussi beaucoup développé les propriétés stochastiques du modèle. Ce modèle est intéressant car il a la structure d'un modèle à risque proportionnel en analyse de survie, lequel a montré son utilité afin de modéliser de nombreux jeux de données. Cependant, à notre connaissance, l'inférence statistique pour le modèle Yang-Nevzorov a été peu développé.

Le but de ce travail est d'introduire certains estimateurs des paramètres θ et γ des modèles LDM et Yang respectivement et d'en tirer leurs propriétés statistiques. Il est montré que le mécanisme de censure est informatif pour θ et γ. Cela justifie le travail sur l'utilité des estimateurs qui peuvent être obtenus à partir de $\{\delta_t, t = 1, ..., T\}$. Nous donnons quelques propriétés exactes et asymptotiques de ces estimateurs. Il se trouve que dans le modèle de Yang, le comportement des différents estimateurs est indépendant de la distribution sous-jacente. Notons que nos estimateurs peuvent être utilisés même lorsque les valeurs exactes de records sont elles-mêmes indisponibles ou de mauvaise qualité et les seules indicatrices δ_t sont disponibles ou fiables. En outre, il est montré que des tests d'ajustement du modèle de Yang peuvent aussi être dérivés de ces δ_t. Ces tests ont même des capacités diagnostiques qui peuvent aider à suggérer des corrections au modèle.

Dans le Chapitre 1, nous présentons ce que sont les records et pourquoi leur étude est d'une certaine importance en pratique. Ensuite on discute de leur comportement dans le cas classique (*iid*) et on présente quelques arguments qui renforcent l'idée que le passage au delà du contexte de records classiques est parfois indispensable.

Dans le Chapitre 2, nous considérons le cas d'un modèle LDM. Notre but est de développer le comportement stochastique de ce modèle.

Nous appliquons les résultats du Chapitre 2 à l'estimation du paramètre θ d'un modèle LDM dans le cas d'une distribution sous-jacente de Gumbel de

paramètres connus et étudions le comportement asymptotique des différents estimateurs dans le Chapitre 3.

Le Chapitre 4 est consacré à explorer le modèle de Yang-Nevzorov et sa relation avec le modèle LDM du Chapitre 2. De plus, nous estimons le paramètre de puissance γ d'un modèle de Yang-Nevzorov dans le cas d'une distribution sous-jacente quelconque de paramètres connus.

Toujours dans le contexte d'un modèle de Yang, dans le Chapitre 5 nous étudions le comportement stochastique du temps inter-records et nous donnons sa loi asymptotique, indépendamment de la loi des *va* sous-jacentes. Puis, en se basant sur ces résultats, combinés aux estimateurs obtenus au chapitre précédent, nous introduisons plusieurs tests d'adéquation d'un modèle de Yang. On montre que ces tests peuvent aider à suggérer des corrections au modèle. Enfin, nous appliquons nos résultats théoriques à des données analysées précédemment par Yang (1975).

Dans le Chapitre 6, nous passons à l'utilisation de la totalité des données disponibles (valeurs et indices/indicatrices de records) afin de calculer, par plusieurs méthodes, des estimateurs des paramètres des modèles LDM et Yang-Nevzorov dans des cas où la distribution sous-jacente n'est pas nécessairement Gumbel et/ou de paramètres connus. De plus, nous introduisons des tests statistiques qui nous aident à vérifier la conformité du choix de la distribution sous-jacente et à choisir entre un modèle LDM et de Yang.

Enfin, le Chapitre 7 présente une conclusion générale de notre travail et quelques perspectives pour les travaux futurs.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Les records, une introduction

On trouve des données sous forme de records dans de nombreuses disciplines utilisatrices de la statistique : changement climatique (Wergen et Krug, 2010) et (Wergen, 2013), risque d'émergence d'une pathologie (Khraibani *et al.*, 2015), finance (Wergen *et al.*, 2011), sports (Yang, 1975), etc. L'étude de records a commencé avec (Chandler, 1952) et a connu de nombreux développements dans les travaux de Arnold, Nevzorov et leurs collaborateurs durant les années 1980 et 1990. Les premiers résultats ont été obtenus dans le cas dit « classique » où les variables aléatoires (*va*) sont indépendantes et identiquement distribuées (*iid*). Par la suite et reconnaissant que ce cas n'ajustait pas correctement plusieurs jeux de données, les chercheurs ont tenté d'aller au delà du contexte des records classiques et se sont penchés sur certains cas où les observations sont indépendantes mais pas identiquement distribuées.

Dans ce chapitre, nous présentons d'abord ce que sont les records et pourquoi leur étude est d'une certaine importance en pratique. Ensuite, dans le double but de donner un aperçu de ce domaine de recherche et d'introduire quelques résultats théoriques utilisés par la suite, on discute de leur comportement dans le cas classique. Le lecteur intéressé pourra consulter les ouvrages d'Arnold *et al.* (1998) et de Nevzorov (2001) pour une présentation plus complète des principaux résultats théoriques concernant les records dans ce cadre. On présente ensuite quelques arguments qui renforcent l'idée que le passage au delà du contexte de records classiques est parfois indispensable.

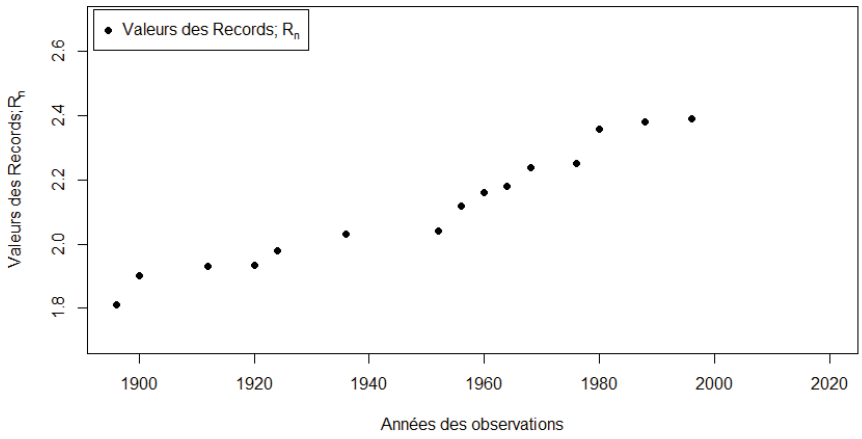

FIGURE 1.1 – Les records du gagnant de la médaille d'or en saut en hauteur homme (Mètres) aux différents jeux olympiques depuis 1896.

1.1 Définition d'un record

Simplement dit, un record est un résultat dans une chaîne donnée d'événements qui dépasse tout ce qui a été rencontré auparavant. Par conséquent, un nouveau record est toujours quelque chose de remarquable et qui attire l'attention, indépendamment de fait que le record est associé ou non à une bonne nouvelle. Un exemple prototype apparait à la Figure 1.1 qui présente les records olympiques du saut en hauteur homme (Mètres) (pour une référence voir le site internet : `http://www.olympic.org`). Dans cette figure, et en fonction des années des différents jeux olympiques, chaque point noir présente la hauteur du saut de l'athlète qui a remporté la médaille d'or aux jeux olympiques en battant le record existant. Ces valeurs de records sont notées par la suite R_n. La série de points représente évidemment une tendance strictement croissante.

Les records font partie de la culture populaire et le livre *Guinness de records* répertorie de nombreuses séries de records observées au fil des années dans la plupart des champs de l'activité humaine. Il est amusant de signaler que ce livre détient lui-même le record du livre le plus vendu au monde. Sur le plan statistique, et le livre Guinness en témoigne, les records sont souvent

beaucoup plus facilement accessibles que les données à partir desquelles ils sont construits. Par exemple, la Figure 1.2, obtenue après un peu plus de recherche que la précédente, donne les hauteurs du saut maximal des athlètes qui ont obtenu la médaille d'or aux différents jeux olympiques : les points noirs correspondent aux records de la figure précédente et les points rouges sont les sauts de ceux qui n'ont pas battu le record existant. Ces valeurs (rouges et noires) sont notées par la suite X_t.

On peut déjà sentir que les records sont conceptuellement proches des valeurs extrêmes ; il n'est donc pas surprenant que plusieurs structures apparaissant dans la théorie des valeurs extrêmes se retrouvent également dans la théorie de records. Cette impression est renforcée en notant que les valeurs de chacun des points rouges ou noirs sont elles-mêmes les maximums des meilleurs essais de tous les participants à un certain jeu olympique, que l'on a représentés par les points en bleus dans la Figure 1.3. Déjà, l'obtention de ces points bleus est une entreprise beaucoup plus compliquée. En outre, les participants qui ont donné ces points bleus sont eux-mêmes « générateurs de valeurs extrêmes » car un pays n'envoie aux olympiques que ses meilleurs athlètes. Les hauteurs de leurs sauts sont à leur tour des valeurs extrêmes parmi ceux des autres athlètes d'un même pays, au fil des différentes compétitions qui ont eu lieu pour la sélection finale. Ainsi on peut déplorer que la simple observation de records corresponde à une immense quantité de données perdue dont la partie émergente est la série des R_n, mais l'entièreté de ces données est extrêmement difficile à observer. C'est pour cette raison que dans de nombreuses situations, la série de records est tout ce dont dispose le statisticien pour étudier le phénomène qui l'intéresse et apporter des réponses aux questions qui lui sont posées (e.g. quelle hauteur maximale peut-on atteindre avec la technologie existante et en quelle année cette hauteur maximale est atteinte). Enfin, comme la théorie de records est l'étude des valeurs extrêmes de valeurs extrêmes, il n'est pas déraisonnable de voir les records comme un filtre sur une série chronologique de données suivant la loi de Gumbel, Weibull ou Fréchet qui sont les 3 lois possibles de valeurs extrêmes d'un grand nombre de données.

Signalons que récemment la théorie de records se diversifie dans plusieurs sens et selon différentes pistes de recherche : par exemple Nevzorov et Stepanov (2014) l'utilisent pour créer un test détectant les valeurs aberrantes dans un jeu de données.

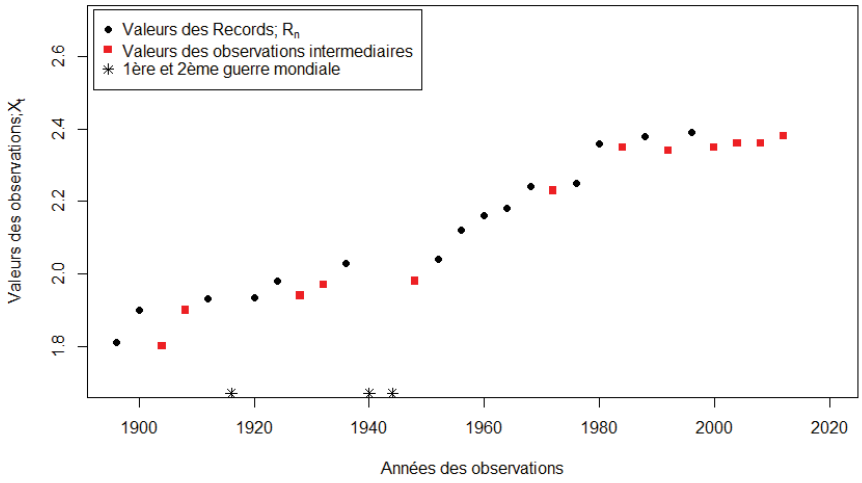

FIGURE 1.2 – Hauteurs du saut maximal des athlètes qui ont obtenu la médaille d'or aux différents jeux olympiques depuis 1896.

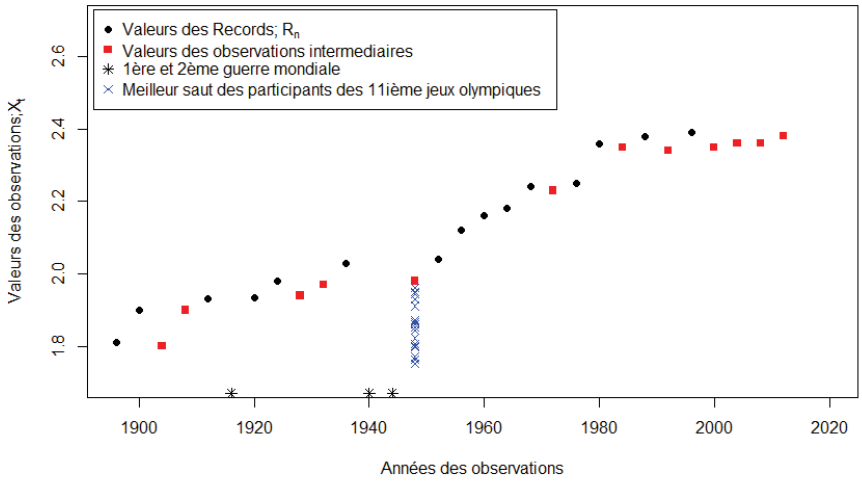

FIGURE 1.3 – Hauteurs du saut maximal des athlètes qui ont obtenu la médaille d'or aux différents jeux olympiques depuis 1896.

1.2 Contexte général et notations

Dans la suite de ce travail, $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé et X une variable aléatoire réelle définie sur Ω de fonction de répartition $F_X(\cdot)$ parfois notée $F(\cdot)$ quand il n'y a pas de risque de confusion et de densité $f_X(\cdot)\,(= f(\cdot))$ par rapport à la mesure de Lebesgue. On suppose que $(\Omega, \mathcal{F}, \mathbb{P})$ est suffisamment riche pour supporter une suite infinie $\{X_t, t \geq 1\}$ de copies indépendantes de X ; elles sont donc des *va iid.* Une observation X_t est appelé un « Record » si sa valeur est supérieure à celle de toutes les observations l'ayant précédées :

$$X_t > \max\{X_1, \ldots, X_{t-1}\},$$

et dans ce cas, on la note R_n. Notons que $R_1 = X_1$, que l'on appelle le record trivial. Notons aussi qu'on considère ici des records supérieurs. On pourrait aussi considérer des records inférieurs, où X_t est appelée un record si $X_t < \min\{X_1, \ldots, X_{t-1}\}$. Dans la mesure où pour se ramener à l'un ou l'autre cas il suffit de multiplier les *va* par « -1 », il n'y a pas de perte de généralité à considérer les records supérieurs. C'est ce que nous ferons dans la suite de ce travail.

On définit aussi δ_t l'indicatrice d'un record (avec $\delta_1 = 1$) :

$$\delta_t = \begin{cases} 1 & \text{si } X_t > \max(X_1, \ldots, X_{t-1}) \\ 0 & \text{sinon} \end{cases},$$

et L_n l'indice du $n^{ième}$ record :

$$\begin{aligned} L_1 &= 1, \\ L_n &= \inf\left\{t > L_{n-1} : X_t > X_{L_{n-1}}\right\}. \end{aligned}$$

Avec ces notations, $R_n = X_{L_n}$. Enfin, soit N_T le nombre total de records parmi $X_1, \ldots, X_T$ où T désigne le temps présent ; on pose $N_1 = 1$ et par la suite :

$$N_T = \sum_{t=1}^{T} \delta_t.$$

Dans la suite, on suppose disposer des données dans les suites $\{R_n, n \geq 1\}$ et $\{L_n, n \geq 1\}$ et/ou $\{\delta_t, t \geq 1\}$. La suite $\{N_T, T \geq 1\}$ se calcule à partir de l'une ou l'autre de ces deux dernières suites.

Le résultat suivant montre que pour un T fixe, l'information contenue dans $\{L_n, 1 \leq n \leq N_T\}$ est équivalente à celle dans $\{\delta_t, 1 \leq 1 \leq T\}$.

Proposition 1.1. *Pour un T fixe et un nombre de records $N_T = m$ donné, à partir de la suite des indices de records $\{L_n, 1 \leq n \leq m\}$ on peut déterminer la configuration complète de la suite des indicatrices de records $\{\delta_t, 1 \leq t \leq T\}$, et réciproquement.*

Démonstration. ($\Longrightarrow$) Si $\{L_1 = l_1 = 1, L_2 = l_2, \ldots, L_m = l_m\}$, c'est donc que

$$\{\delta_1 = \delta_{l_1} = 1, \delta_2 = 0, \ldots, \delta_{l_2-1} = 0, \delta_{l_2} = 1, \delta_{l_2+1} = 0, \ldots, \\ \delta_{l_m-1} = 0, \delta_{l_m} = 1, \delta_{l_m+1} = 0, \ldots, \delta_T = 0\}.$$

($\Longleftarrow$) Réciproquement, si

$$\{\delta_1 = \delta_{l_1} = 1, \delta_2 = 0, \ldots, \delta_{l_2-1} = 0, \delta_{l_2} = 1, \delta_{l_2+1} = 0, \ldots, \\ \delta_{l_m-1} = 0, \delta_{l_m} = 1, \delta_{l_m+1} = 0, \ldots, \delta_T = 0\},$$

alors les indices t des $\delta_t = 1$ sont les valeurs des indices de records $\{L_1 = l_1 = 1, L_2 = l_2, \ldots, L_m = l_m\}$. □

Signalons aussi que dans certaines applications, il est commode de travailler avec les records à battre à l'instant t

$$\begin{aligned} W_t &= \text{Record à battre à l'instant } t, \\ &= \max(X_1, \ldots, X_t). \end{aligned}$$

Ainsi, pour l'analyse statistique et probabiliste de records, on suppose qu'il est donné d'observer la suite de couples $\{(R_n, L_n); n = 1, ..., N_T\}$ ou, de façon équivalente, la suite $\{(W_t, \delta_t), 1 \leq t \leq T\}$.

De toute façon, à cause de cette information restreinte extraite uniquement des couples $\{(R_n, L_n); n = 1, ..., N_T\}$ ou $\{(W_t, \delta_t), 1 \leq t \leq T\}$, de nombreuses questions sur le comportement de la série chronologique des X_t doivent rester sans réponses e.g. $\mathbb{P}[X_{t+1} \leq x \mid X_t = y]$. Cependant, certaines questions importantes, telles que la prédiction de la valeur et de l'instant du prochain record, R_{N_T+1} et L_{N_T+1}, sont accessibles.

Enfin, profitons de l'occasion pour signaler la similitude entre les records et le traitement de données censurées en analyse de survie. En particulier, toutes les valeurs de X_t entre R_n et R_{n+1} peuvent être considérées comme des observations censurées par R_n. En outre, on remarque dans les chapitres suivants, que ce mécanisme de censure peut contenir des informations sur les paramètres inconnus d'un modèle de records, information qui est donc exploitable.

1.3 Cas *iid*

Les propriétés stochastiques des séries de records ont été largement étudiés dans le cas où les X_t sont indépendantes et identiquement distribuées (Arnold *et al.*, 1998 ; Nevzorov, 2001). Il se trouve que beaucoup de ces propriétés sont universelles, c'est-à-dire elles tiennent pour une distribution quelconque de X_t. Cela a permis un progrès dans la compréhension globale du comportement stochastique de records.

Considérons le cas où les X_t sont *iid* de densité $f(\cdot)$ et de fonction de répartition $F(\cdot)$. Nous présentons un certain nombre de résultats connus afin de donner un aperçu de quelques faits importants en théorie de records. On note $\mathbb{P}[\delta_t = 1]$ par P_t, le taux de record à l'instant t. C'est la probabilité que la $t^{ième}$ observation, X_t, soit un record.

Théorème 1.2. *(Nevzorov, 2001 ; page 58), Pour tout $T \geq 1$, les va δ_1,...,δ_T et $M_T = \max(X_1, \ldots, X_T)$ sont mutuellement indépendantes avec*

$$\delta_t \backsim \textit{Bernoulli}\left(\frac{1}{t}\right),$$

de sorte que :

$$\mathbb{P}[\delta_t = 1] = \frac{1}{t}.$$

Démonstration.

$$\begin{aligned}
\mathbb{P}[\delta_t = 1] &= \mathbb{P}[\max\{X_1, \ldots, X_{t-1}\} < X_t], \\
&= \int_{\mathbb{R}} \mathrm{d}\mathbb{P}(\max\{X_1, \ldots, X_{t-1}\} < X_t, X_t = x), \\
&= \int_{\mathbb{R}} (F_X(x))^{t-1}\, \mathrm{d}F_X(x), \\
&= \int_0^1 u^{t-1} \mathrm{d}u, \\
&= \frac{1}{t}.
\end{aligned}$$

Les *va* δ_t prennent seulement deux valeurs. Ainsi, pour montrer l'indépendance il suffit de prouver, Nevzorov (2001), que pour tout x ; $T = 1, 2, 3, \ldots$;

$N_T = 1, \ldots, T$ et

$$L_1 = l_1 = 1 < L_2 = l_2 < \ldots < L_{N_T} = l_{N_T} \leq T, \text{ indices de records,}$$

on a :

$$\mathbb{P}\left[\delta_{l_1} = 1, \ldots, \delta_{l_{N_T}} = 1, M_T < x\right] = \mathbb{P}\left[\delta_{l_1} = 1\right] \times \ldots \times \mathbb{P}\left[\delta_{l_{N_T}} = 1\right] \times \mathbb{P}\left[M_T < x\right],$$

avec $\mathbb{P}\left[\delta_{l_1} = 1\right] = 1$ (car la première observation est toujours un record, c'est le record trivial). Considérons les cas où $N_T = 2$ et $N_T = 3$:

Pour $N_T = 2$:

$$\begin{aligned}
\mathbb{P}\left[\delta_{l_1} = 1, \delta_{l_2} = 1, M_T < x\right] &= \mathbb{P}\left[\max\left(X_1, \ldots, X_{l_2-1}\right) < X_{l_2} < x; \right. \\
&\quad \left. \max\left(X_{l_2+1}, \ldots, X_T\right) < x\right], \\
&= \left(F\left(x\right)\right)^{T-l_2} \int_{-\infty}^{x} \left(F_X\left(u\right)\right)^{l_2-1} dF_X(u), \\
&= \frac{F_X^T(x)}{l_2}, \\
&= 1 \times \frac{1}{l_2} \times F_X^T(x), \\
&= \mathbb{P}\left[\delta_{l_1} = 1\right] \times \mathbb{P}\left[\delta_{l_2} = 1\right] \times \mathbb{P}\left[M_T < x\right].
\end{aligned}$$

Pour $N_T = 3$:

$$\begin{aligned}
\mathbb{P}\left[\delta_{l_1} = 1, \delta_{l_2} = 1, \delta_{l_3} = 1, M_T < x\right] &= \mathbb{P}\left[\max\left(X_1, \ldots, X_{l_2-1}\right) < X_{l_2} < x; \right. \\
&\quad \max\left(X_{l_2}, \ldots, X_{l_3-1}\right) < X_{l_3} < x; \\
&\quad \left. \max\left(X_{l_3+1}, \ldots, X_T\right) < x\right], \\
&= \left(F_X\left(x\right)\right)^{T-l_3} \int_{-\infty}^{x} \int_{u}^{x} \left(F_X\left(u\right)\right)^{l_2-1} \\
&\quad \times \left(F_X\left(v\right)\right)^{l_3-l_2-1} dF\left(v\right) dF\left(u\right), \\
&= \frac{F_X^T\left(x\right)}{l_2 l_3}, \\
&= 1 \times \frac{1}{l_2} \times \frac{1}{l_3} \times F_X^T(x), \\
&= \mathbb{P}\left[\delta_{l_1} = 1\right] \times \mathbb{P}\left[\delta_{l_2} = 1\right] \times \mathbb{P}\left[\delta_{l_3} = 1\right] \\
&\quad \times \mathbb{P}\left[M_T < x\right].
\end{aligned}$$

Le cas $N_T > 3$ se prouve de façon analogue. □

Théorème 1.3. *Arnold* et al. *(1998), La suite* $\{R_n, n \geq 1\}$ *des valeurs de records forme une chaîne de Markov de probabilité de transition :*

$$f_{R_n|R_{n-1}}(r_n|r_{n-1}) = \frac{f(r_n)}{1 - F(r_{n-1})},$$

et d'espace d'état $E = \{R_1 = X_1 = r_1 < R_2 = r_2 < \cdots < R_n = r_n < \cdots\}$.

Théorème 1.4. *Shorrock (1972), La distribution exacte de* N_T *est donnée par :*

$$\mathbb{P}[N_T = m] = \frac{\mathcal{S}(T, m)}{T!}, 0 \leq m \leq T,$$

où $\mathcal{S}(T, m)$ *est le nombre de Stirling de première espèce (voir Charalambides et Singh (1988)), définie par :*

$$\prod_{t=0}^{T-1}(s - t) = \sum_{m=0}^{T} \mathcal{S}(T, m)\, s^m.$$

Remarque 1.5. En écrivant $N_T = \delta_1 + \cdots + \delta_T$ nous obtenons l'espérance et la variance de N_T :

$$\begin{aligned} \mathbb{E}(N_T) &= \sum_{t=1}^{T} \frac{1}{t}, \\ &= \ln(T) + \varpi, \end{aligned}$$

où ϖ est la constante d'Euler-Mascheroni : $\varpi = \lim_{T\to\infty} \left(\sum_{t=1}^{T} \frac{1}{t} - \ln(T)\right) \simeq 0.57722$. Et

$$\begin{aligned} \mathbb{V}(N_T) &= \sum_{t=1}^{T} \left(\frac{1}{t}\right)\left(1 - \frac{1}{t}\right), \\ &= \sum_{t=1}^{T} \frac{1}{t} - \sum_{t=1}^{T} \frac{1}{t^2}. \end{aligned}$$

Proposition 1.6. *Arnold* et al. *(1998), La distribution jointe des indices de records est :*

$$\begin{aligned}
\mathbb{P}\left[L_1 = 1, L_2 = l_2, \ldots, L_{N_T} = l_{N_T}\right] &= \mathbb{P}\left[\delta_1 = 1, \delta_2 = 0, \ldots, \delta_{l_2-1} = 0,\right. \\
&\quad \left.\delta_{l_2} = 1, \ldots, \delta_{l_{N_T}} = 1\right], \\
&= \mathbb{P}\left[\delta_1 = 1\right] \times \mathbb{P}\left[\delta_2 = 0\right] \times \cdots \\
&\quad \times \mathbb{P}\left[\delta_{l_{N_T}} = 1\right], \\
&= 1 \times \left(1 - \frac{1}{2}\right) \times \cdots \times \frac{1}{l_{N_T}}, \\
&= \frac{1}{l_2 - 1} \times \cdots \times \frac{1}{l_{N_T} - 1} \times \frac{1}{l_{N_T}}.
\end{aligned}$$

1.4 Cas non *iid*

Arnold *et al.*(1998, page 26) montrent que, dans le cas *iid*, lorsque n augmente, les records ont tendance à devenir plus espacés dans le temps. Cependant, dans de nombreux jeux de données réelles, ce phénomène ne se produit pas et les raisons pour cela sont nombreuses. Par exemple, les progrès technologiques peuvent rendre l'apparition de records plus fréquente que prévue par le modèle *iid*. De plus, l'intérêt pour une activité sportive où les records sont imbattables diminue autant chez les spectateurs et les praticiens. Ceci a conduit à l'élaboration de modèles plus complets qui peuvent fournir une meilleure prédiction.

La Figure 1.4 présente les records olympiques du saut en hauteur homme (Mètres) contre les records extraits d'une série chronologique de variables aléatoires *iid*. De cette figure, on remarque qu'il n'est pas ici raisonnable de considérer le modèle *iid*, car le nombre de records croit rapidement. De plus, dans le cas *iid* les records se concentrent parmi les premières observations. Ceci montre l'intérêt de modèles qui dépassent l'hypothèse *iid*. Dans le Chapitre 2, nous présentons plusieurs autres arguments qui renforcent l'idée que le passage au delà d'un modèle *iid* est parfois indispensable.

Les modèles de records les plus populaires au-delà du cas *iid* sont :

1. Le modèle à dérive linéaire (LDM).
2. Le modèle de Yang-Nevzorov (Modèle F^{ρ}).

Dans ces deux modèles, les X_t sont indépendantes mais pas identiquement distribuées.

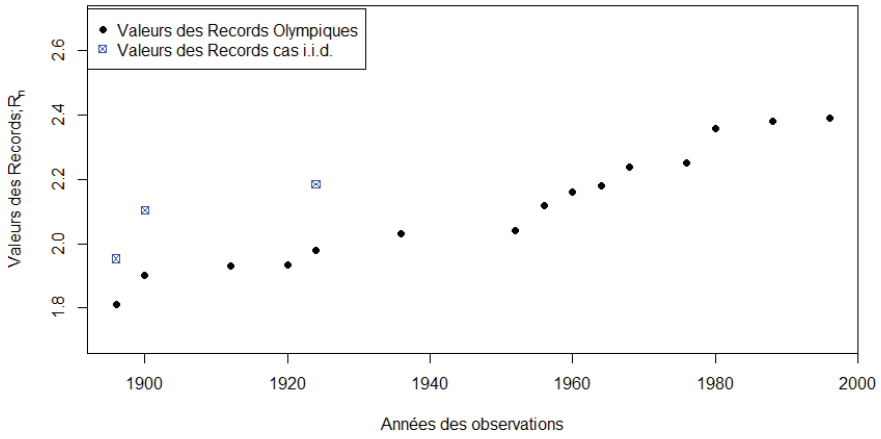

FIGURE 1.4 – Records olympiques du saut en hauteur homme (Mètres) vs Records cas *iid.*

1.5 Conclusion

Dans ce chapitre, nous avons donné une définition simple d'un record et nous avons fait une revue de quelques résultats de la théorie de records dans le cas où les variables aléatoires X_t sont indépendantes et identiquement distribuées. Dans le chapitre suivant, nous considérons le cas d'un modèle LDM. Notre but est de développer le comportement stochastique de ce modèle.

Chapitre 2

Modèle LDM : Quelques résultats sur le comportement stochastique

Le modèle, peut-être le plus simple mais en tout cas le plus populaire, pour une série de records issus d'observations indépendantes mais non identiquement distribuées est le modèle à dérive linéaire (LDM) introduit par Ballerini et Resnick (1985). Dans ce chapitre nous développons le comportement stochastique d'un modèle LDM et nous présentons certaines propriétés des indicatrices de records $\{\delta_t, 1 \leq t \leq T\}$ où T désigne le temps présent. De plus, en se basant sur ces derniers nous déterminons la distribution du nombre de records N_T et celle des indices de records $\{L_n, n \geq 1\}$.

Dans un modèle LDM, la $t^{ième}$ observation de la série chronologique est de la forme :

$$X_t = Y_t + \theta \times t, \tag{2.0.1}$$

où θ est une constante strictement positive appelée « drift » , $1 \leq t \leq T$ et $\{Y_t, 1 \leq t \leq T\}$ est une suite de variables aléatoires *iid* de fonction de répartition $F(\cdot)$. Ainsi, les *va* $\{X_t, 1 \leq t \leq T\}$ sont indépendantes mais non identiquement distribuées de fonctions de répartition $\{F_t(\cdot), 1 \leq t \leq T\}$ et de densités $\{f_t(\cdot), 1 \leq t \leq T\}$ où

$$\begin{aligned} f_t(x) &= f(x - \theta \times t), \\ F_t(x) &= F(x - \theta \times t). \end{aligned}$$

Avec ce modèle, le taux de record est :

$$\begin{aligned} P_t(\theta) &= \int f_t(x) \left(\prod_{k=1}^{t-1} F_k(x) \right) \mathrm{d}x, \\ &= \int f(x - \theta \times t) \left(\prod_{k=1}^{t-1} F(x - \theta \times k) \right) \mathrm{d}x, \\ &= \int f(y) \left(\prod_{k=1}^{t-1} F(y + \theta \times k) \right) \mathrm{d}y. \end{aligned}$$

2.1 Résultats préliminaires

Dans le cadre d'un modèle LDM, Ballerini et Resnick (1985) montrent que pour un $\theta > 0$ et si $\mathbb{E}(Y_+) < \infty$ ($y_+ = \max(0, y)$), on peut définir un taux de record asymptotique :

$$P(\theta) = \lim_{t \to \infty} P_t(\theta).$$

Cette quantité limite $P(\theta)$ doit, en général, être évaluée numériquement, mais une expression explicite est disponible dans certains cas.

Exemple 2.1. Si les Y_t suivent une loi de Gumbel de paramètres connus, que sans perte de généralité et afin de simplifier la présentation nous prendrons comme étant $\mu = 0$ (paramètre de position) et $\beta = 1$ (paramètre d'échelle), notée $G(0, 1)$ de sorte que $F(y) = \exp(-\exp(-y))$, on a alors la relation :

$$\begin{aligned} F(y + a) &= \exp\left(-e^{-y-a}\right), \\ &= \exp\left(-e^{-y}\right)^{e^{-a}}, \\ &= F(y)^{e^{-a}}. \end{aligned}$$

Donc,

$$\begin{aligned} P_t(\theta) &= \int f(y) \left(\prod_{k=1}^{t-1} F(y + \theta \times k) \right) \mathrm{d}y, \\ &= \int f(y) \left(\prod_{k=1}^{t-1} F(y)^{e^{-\theta k}} \right) \mathrm{d}y, \end{aligned}$$

$$\begin{aligned} P_t(\theta) &= \int_0^1 u^{\sum_{k=1}^{t-1}\tau^k}\,\mathrm{d}u, \text{ avec } \tau = e^{-\theta}, \\ &= \frac{1}{\sum_{k=0}^{t-1}\tau^k}, \\ &= \frac{1-e^{-\theta}}{1-e^{-\theta t}}. \end{aligned}$$

Ainsi,

$$\lim_{t\to\infty} P_t(\theta) = P(\theta) = 1 - e^{-\theta}.$$

Remarque 2.2. Nous verrons au Chapitre 6 que si les Y_t suivent une loi de Gumbel de paramètres inconnus, notée $G(\mu,\beta)$, alors le taux de record dépendra uniquement du drift θ et du paramètre d'échelle β et aura comme expression $P_t = \frac{1-e^{-\frac{\theta}{\beta}}}{1-e^{-\frac{\theta}{\beta}t}}$. Ainsi les paramètres θ et β ne sont pas identifiables. Il est donc raisonnable de considérer le cas $\beta = 1$ pour simplifier. D'autre part, l'importance de considérer le cas où la distribution sous-jacente suit la loi de Gumbel vient du fait que :

1. Dans un modèle LDM, Borovkov (1999) montre que la distribution $G(\mu,\beta)$ est l'unique loi caractérisée par l'indépendance des δ_t.
2. Gumbel est le domaine d'attraction de nombreuses densités populaires (Normale, Exponentielle, ...).

Théorème 2.3. *(Ballerini et Resnick(1985-1987)) Dans le contexte d'un modèle LDM, si $F(\cdot)$ est continue, $\theta > 0$, et $\mathbb{E}[Y_+] < \infty$ alors*

$$\mathbb{E}\left[\frac{N_T}{T}\right] \underset{T}{\longrightarrow} P(\theta),$$

et

$$\frac{N_T}{T} \longrightarrow P(\theta), \textit{ presque sûrement.}$$

De plus, si $\mathbb{E}[Y_+]^2 < \infty$, il existe un $0 < \sigma^2 < \infty$ tel que

$$\sqrt{T}\left(\frac{N_T}{T} - P(\theta)\right) \longrightarrow N\left(0,\sigma^2\right).$$

Dans le cas d'une distribution sous-jacente de loi de Gumbel, $\sigma^2 = P(\theta) - P^2(\theta)$ (voir aussi le Chapitre 3).

2.2 Distribution de N_T

Le nombre de records N_T est lié aux indicatrices de records par la relation :

$$N_T = \sum_{t=1}^{T} \delta_t, \tag{2.2.1}$$

avec δ_t l'indicatrice de record dont on rappelle la définition :

$$\delta_t = \begin{cases} 1 & \text{si } X_t > \max(X_1, \ldots, X_{t-1}) \\ 0 & \text{sinon} \end{cases}.$$

Proposition 2.4. *Si dans le modèle LDM, les va $\{Y_t, t \geq 1\}$ suivent la loi $G(0,1)$, alors N_T a pour fonction de masse :*

$$\mathbb{P}_\theta [N_T = m] = \frac{\exp(-T\theta)\, \mathcal{S}(T, m \mid \vec{u}(\theta))}{\prod_{t=1}^{T} u_t(\theta)}, \tag{2.2.2}$$

où le vecteur $\vec{u}(\theta) = (u_0(\theta), \ldots, u_T(\theta))$ est défini par

$$u_t(\theta) = \frac{e^{-\theta}\left(1 - e^{-t\theta}\right)}{(1 - e^{-\theta})}, \text{ avec } u_0(\theta) = 0, \tag{2.2.3}$$

et $\mathcal{S}(T, m \mid \vec{u}(\theta))$ est la généralisation du nombre de Stirling de 1er espèce, définie par la comparaison des deux polynômes suivants :

$$\prod_{t=0}^{T-1} (s + u_t(\theta)) = \sum_{m=0}^{T} \mathcal{S}(T, m \mid \vec{u}(\theta))\, s^m. \tag{2.2.4}$$

Démonstration. Borovkov (1999) montre que dans un modèle LDM, la distribution de Gumbel est l'unique loi caractérisée par l'indépendance des indicatrices de records. Ainsi, la fonction génératrice de N_T est :

$$\begin{aligned} G_{N_T}(s) &= \mathbb{E}\left(s^{N_T}\right), \\ &= \mathbb{E}\left(s^{\sum_{t=1}^{T} \delta_t}\right), \\ &= \prod_{t=1}^{T} G_{\delta_t}(s). \end{aligned} \tag{2.2.5}$$

D'après l'exemple 2.1, le taux de record $\mathbb{P}[\delta_t = 1]$, pour une distribution de $G(0,1)$, est donné par :

$$P_t(\theta) = \frac{1 - e^{-\theta}}{1 - e^{-t\theta}}, \tag{2.2.6}$$

de sorte que,

$$\begin{aligned} G_{\delta_t}(s) &= \mathbb{E}\left(s^{\delta_t}\right), \\ &= s^0 \mathbb{P}[\delta_t = 0] + s\mathbb{P}[\delta_t = 1], \\ &= (1 - P_t(\theta)) + sP_t(\theta), \\ &= \frac{1 - e^{-\theta}}{1 - e^{-t\theta}} \left[\frac{1 - e^{-t\theta}}{1 - e^{-\theta}} - 1 + s\right]. \end{aligned} \tag{2.2.7}$$

En remplaçant (2.2.7) dans (2.2.5) on obtient :

$$\begin{aligned} G_{N_T}(s) &= \left(\prod_{t=1}^{T} \frac{1 - e^{-\theta}}{1 - e^{-t\theta}}\right) \prod_{t=1}^{T} \left(s + \frac{e^{-\theta} - e^{-t\theta}}{1 - e^{-\theta}}\right), \\ &= \left(\prod_{t=1}^{T} \frac{1 - e^{-\theta}}{1 - e^{-t\theta}}\right) \prod_{t=0}^{T-1} \left(s + \frac{e^{-\theta} - e^{-(t+1)\theta}}{1 - e^{-\theta}}\right), \\ &= \left(\prod_{t=1}^{T} \frac{1 - e^{-\theta}}{1 - e^{-t\theta}}\right) \prod_{t=0}^{T-1} \left(s + \frac{e^{-\theta}\left(1 - e^{-t\theta}\right)}{1 - e^{-\theta}}\right). \end{aligned}$$

Définissant les $u_t(\theta)$ comme en (2.2.3), on obtient de (2.2.4)

$$\begin{aligned} G_{N_T}(s) &= \left(\prod_{t=1}^{T} \frac{1 - e^{-\theta}}{1 - e^{-t\theta}}\right) \prod_{t=0}^{T-1} (s + u_t(\theta)), \\ &= \left(\prod_{t=1}^{T} \frac{1 - e^{-\theta}}{1 - e^{-t\theta}}\right) \sum_{m=0}^{T} \mathcal{S}(T, m \mid \vec{u}(\theta)) s^m. \end{aligned} \tag{2.2.8}$$

Or, $\mathbb{P}_\theta[N_T = 0] = 0$ de sorte que

$$G_{N_T}(s) = \sum_{m \geq 1} \mathbb{P}_\theta[N_T = m] s^m. \tag{2.2.9}$$

Par comparaison de (2.2.8) et (2.2.9) :

$$\begin{aligned}\mathbb{P}_\theta[N_T = m] &= \mathcal{S}(T, m \mid \vec{u}(\theta)) \prod_{t=1}^{T} \frac{1-e^{-\theta}}{1-e^{-t\theta}}, \\ &= \mathcal{S}(T, m \mid \vec{u}(\theta)) \prod_{t=1}^{T} \frac{e^{-\theta}}{u_t(\theta)}, \\ &= \frac{\mathcal{S}(T, m \mid \vec{u}(\theta))\, e^{-T\theta}}{\prod_{t=1}^{T} u_t(\theta)}.\end{aligned}$$

□

Remarque 2.5. Pour calculer explicitement le nombre de Stirling généralisé de première espèce, on peut utiliser la formule de récurrence suivante, (Khraibani (2008), page 51) :

$$\mathcal{S}(T+1, m \mid \vec{u}(\theta)) = \mathcal{S}(T, m-1 \mid \vec{u}(\theta)) + u_T \mathcal{S}(T, m \mid \vec{u}(\theta)),$$

pour $T \geq 1$ et $0 \leq m \leq T$.

2.3 Moments de N_T

Depuis (2.2.1) on a :

$$\begin{aligned}\mathbb{E}[N_T] &= \sum_{t=1}^{T} \mathbb{E}[\delta_t], \\ &= \sum_{t=1}^{T} P_t(\theta),\end{aligned}$$

et dans le cas de la distribution $G(0,1)$,

$$\mathbb{E}[N_T] = \sum_{t=1}^{T} \frac{1-e^{-\theta}}{1-e^{-t\theta}}.$$

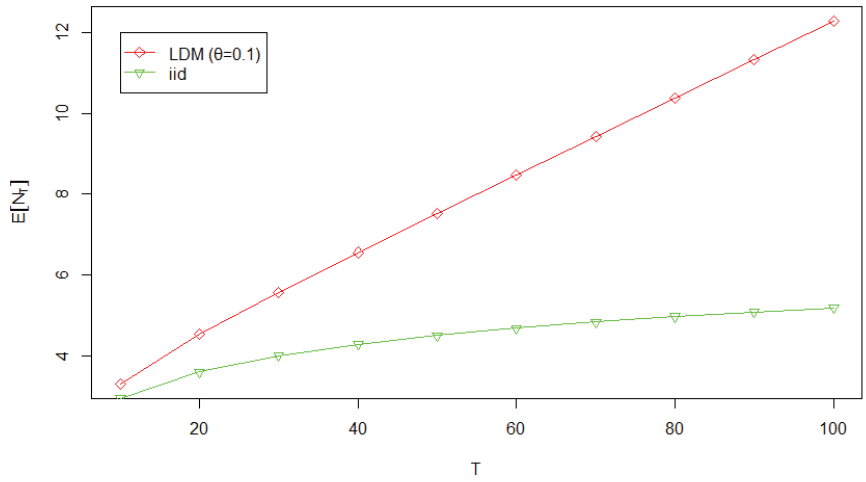

FIGURE 2.1 – Espérance de N_T dans le cas de la loi de Gumbel : cas *iid* vs modèle LDM.

Utilisant l'indépendance des indicatrices de records pour la loi de Gumbel, nous calculons l'expression de la variance de N_T :

$$\mathbb{V}[N_T] = \mathbb{E}[N_T] - \sum_{t=1}^{T} \left(\frac{1 - e^{-\theta}}{1 - e^{-t\theta}} \right)^2 .$$

La Figure 2.1, montre que la moyenne du nombre de records est plus grande dans le cas du LDM (ici avec une distribution sous-jacente $G(0,1)$ et un drift $\theta = 0.1$) que dans le cas *iid* et la différence augmente avec T. La Figure 2.2, compare la fonction de masse de N_T dans le cas LDM, calculée suivant (2.2.2) (distribution sous-jacente $G(0,1)$ et différentes valeurs de θ) avec le cas *iid*. On remarque que dans le cas LDM le mode de la distribution est toujours plus grand que celui du cas *iid* et la différence augmente de concert avec le drift θ. De plus, la queue de droite de la fonction de masse de N_T est toujours plus épaisse dans le cas LDM.

Ceci montre, comme on l'a pressenti au Chapitre 1, que le passage du cas *iid* à un modèle LDM est une façon d'obtenir une souplesse accrue pour ajuster correctement une série de données de records.

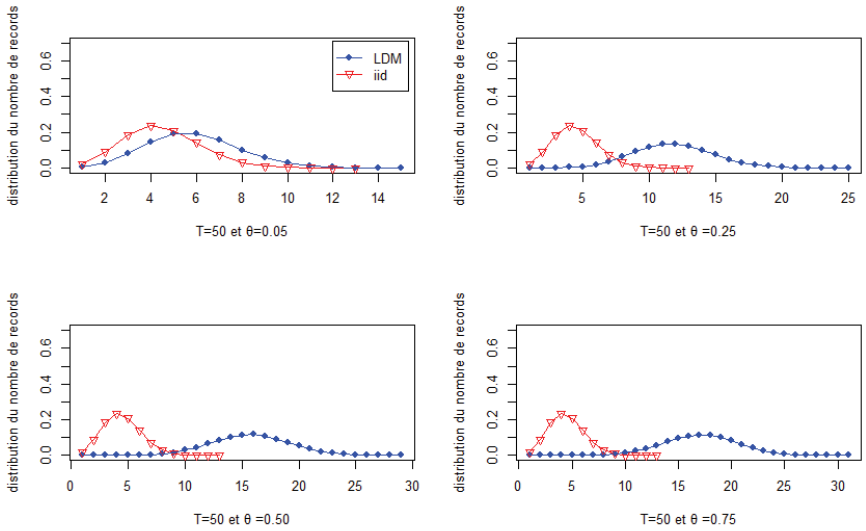

FIGURE 2.2 – Fonction de masse de N_T dans un modèle *iid* et dans un modèle LDM. Cas de la loi de Gumbel.

2.4 Valeurs et indices de records

On rappelle que les suites $\{R_n : n \geq 1\}$ et $\{L_n : n \geq 1\}$ représentent respectivement les valeurs et les indices de records. Plus précisément, dans un modèle LDM avec un drift θ :

$$\begin{aligned} L_1 &= 1\,(\text{Record trivial}). \\ L_n &= \inf\left\{t > L_{n-1} : X_t > X_{L_{n-1}}\right\}, \\ &= \inf\left\{t > L_{n-1} : Y_t + \theta t > Y_{L_{n-1}} + \theta L_{n-1}\right\}, \\ &= \inf\left\{t > L_{n-1} : Y_t > Y_{L_{n-1}} + \theta\left(L_{n-1} - t\right)\right\}, \end{aligned}$$

et

$$\begin{aligned} R_1 &= X_1. \\ R_n &= X_{L_n}, \\ &= Y_{L_n} + \theta L_n. \end{aligned}$$

Remarque 2.6. Si une observation Y_t est un record dans la suite $\{Y_t, t \geq 1\}$, alors la valeur correspondante à cette observation dans le modèle LDM, $X_t = Y_t + \theta t$ est aussi un record dans la suite $\{X_t, t \geq 1\}$. Mais, la réciproque n'est pas vraie.

De cette remarque on peut tirer le corollaire suivant :

Corollaire 2.7. *Dans un modèle LDM et pour une distribution sous-jacente quelconque on a les inégalités suivantes*

$$L_n \leq L_n^{\theta=0},$$

et

$$N_T \geq N_T^{\theta=0},$$

où $L_n^{\theta=0}$ et $N_T^{\theta=0}$ désignent respectivement l'indice du $n^{ième}$ record et le nombre de records dans la suite $\{Y_t, 1 \leq t \leq T\}$.

Remarque 2.8. Ceci montre le fait évoqué dans la Figure 2.1. Par contre, les deux quantités R_n (cas LDM) et R_n (cas *iid*) ne sont pas aussi facilement comparable.

2.5 Distribution des L_n

Considérons la suite des indices de records $L_1 = 1 < L_2 < \ldots < L_m$, pour un nombre de records $N_T = m$ donné dans la suite $\{X_t,\ 1 \leq t \leq T\}$.

Dans la suite, on considère que les valeurs $l_2, \ldots, l_m$ des indices de records sont des entiers telles que $1 < l_2 < \ldots < l_m$. Pour alléger les expressions, on escamote le fait que si les l_j ne sont pas des entiers satisfaisant les contraintes alors les probabilités qu'on rencontre sont nulles.

Proposition 2.9. *Dans un modèle LDM, si les va $\{Y_t, t \geq 1\}$ suivent la loi $G(0,1)$, alors la distribution jointe des indices de records est :*

$$\mathbb{P}\left[L_1 = 1, L_2 = l_2, \ldots, L_m = l_m\right] = \frac{\exp\left(-\left(l_m - m\right)\theta\right)}{1 - \exp\left(-l_m\theta\right)} \prod_{n=2}^{m} \left(\frac{1 - \exp\left(-\theta\right)}{1 - \exp\left(-\left(l_n - 1\right)\theta\right)} \right).$$

Démonstration. L'indépendance des δ_t dans le cas de la loi de Gumbel implique que :

$$\begin{aligned}\mathbb{P}\left[L_1 = 1, L_2 = l_2, \ldots, L_m = l_m\right] &= \mathbb{P}\left[\delta_1 = 1, \delta_2 = 0, \ldots, \delta_{l_2} = 1, \ldots,\right. \\ &\quad \left.\delta_{l_m - 1} = 0, \delta_{l_m} = 1\right],\end{aligned}$$

$$\begin{aligned}\mathbb{P}\left[L_1 = 1, L_2 = l_2, \ldots, L_m = l_m\right] &= 1 \times \left(1 - \frac{1 - \exp(-\theta)}{1 - \exp(-2\theta)}\right) \times \cdots \times \left(\frac{1 - \exp(-\theta)}{1 - \exp(-l_2\theta)}\right) \times \cdots \\ &\quad \times \left(1 - \frac{1 - \exp(-\theta)}{1 - \exp(-(l_m - 1)\theta)}\right) \times \left(\frac{1 - \exp(-\theta)}{1 - \exp(-l_m\theta)}\right), \\ &= (1 - \exp(-\theta))^m \exp(-(l_m - m)\theta) \\ &\quad \times \frac{1}{1 - \exp(-l_m\theta)} \left[\prod_{n=2}^{m} \left(\frac{1}{1 - \exp(-(l_n - 1)\theta)}\right)\right], \\ &= \frac{\exp(-(l_m - m)\theta)}{1 - \exp(-l_m\theta)} \prod_{n=2}^{m} \left(\frac{1 - \exp(-\theta)}{1 - \exp(-(l_n - 1)\theta)}\right).\end{aligned}$$

□

Corollaire 2.10. *Dans un modèle LDM, si les va* $\{Y_t, t \geq 1\}$ *suivent la loi* $G(0,1)$*, alors la distribution marginale d'un indice de records est :*

$$\mathbb{P}\left[L_n = l_n\right] = \frac{1 - \exp(-\theta)}{1 - \exp(-l_n\theta)} \frac{\exp(-(l_n - 1)\theta)\, \mathcal{S}\left(l_n - 1, n - 1 \mid \overrightarrow{u}(\theta)\right)}{\prod_{t=1}^{l_n - 1} u_t(\theta)}. \tag{2.5.1}$$

Démonstration.

$$\begin{aligned}\mathbb{P}\left[L_n = l_n\right] &= \mathbb{P}\left[N_{l_n} = n, N_{l_n - 1} = n - 1\right], \\ &= \mathbb{P}\left[\delta_{l_n} = 1, N_{l_n - 1} = n - 1\right].\end{aligned}$$

Dans le cas de la loi de Gumbel les évènements $\{\delta_{l_n} = 1\}$ et $\{N_{l_n - 1} = n - 1\}$ sont indépendants, de sorte qu'en utilisant la Proposition 2.4 on obtient (2.5.1). □

Proposition 2.11. *Dans un modèle LDM, si les va* $\{Y_t, t \geq 1\}$ *suivent la loi* $G(0,1)$*, la suite des indices de records* $\{L_n, n \geq 1\}$ *forme une chaîne de Markov de probabilité de transition :*

$$\mathbb{P}\left[{}^{L_n=l_n}\!/_{L_{n-1}=l_{n-1}}\right] = \frac{(1-\exp(-\theta))(1-\exp(-\theta l_{n-1}))\exp(-(l_n-l_{n-1}-1)\theta)}{(1-\exp(-(l_n-1)\theta))(1-\exp(-\theta l_n))},$$

et d'espace d'état $E = \{L_1 = 1 < L_2 = l_2 < \cdots < L_n = l_n < \cdots\}$.

Démonstration. Considérons la probabilité :

$$\begin{aligned}
\mathbb{P}\left[{}^{L_n=l_n}\!/_{L_{n-1}=l_{n-1},\ldots,L_2=l_2,L_1=1}\right] &= \frac{\mathbb{P}[L_n=l_n, L_{n-1}=l_{n-1},\ldots,L_2=l_2,L_1=1]}{\mathbb{P}[L_{n-1}=l_{n-1},\ldots,L_2=l_2,L_1=1]}, \\
&= \frac{\frac{\exp(-(l_n-n)\theta)}{1-\exp(-l_n\theta)}\prod_{j=2}^{n}\left(\frac{1-\exp(-\theta)}{1-\exp(-(l_j-1)\theta)}\right)}{\frac{\exp(-(l_{n-1}-(n-1))\theta)}{1-\exp(-l_{n-1}\theta)}\prod_{j=2}^{n-1}\left(\frac{1-\exp(-\theta)}{1-\exp(-(l_j-1)\theta)}\right)}, \\
&= \frac{(1-\exp(-\theta))(1-\exp(-l_{n-1}\theta))}{(1-\exp(-(l_n-1)\theta))} \\
&\quad \times \frac{\exp(-(l_n-l_{n-1}-1)\theta)}{(1-\exp(-l_n\theta))}.
\end{aligned}$$

Or,

$$\begin{aligned}
\mathbb{P}\left[{}^{L_n=l_n}\!/_{L_{n-1}=l_{n-1}}\right] &= \mathbb{P}\left[\delta_{l_{n-1}+1}=0, \delta_{l_{n-1}+2}=0,\ldots,\delta_{l_n-1}=0,\delta_{l_n}=1\right], \\
&= \left(1-\frac{1-\exp(-\theta)}{1-\exp(-(l_{n-1}+1)\theta)}\right)\times\cdots \\
&\quad \times\left(1-\frac{1-\exp(-\theta)}{1-\exp(-(l_n-1)\theta)}\right)\times\left(\frac{1-\exp(-\theta)}{1-\exp(-l_n\theta)}\right), \\
&= \frac{(1-\exp(-\theta))(1-\exp(-l_{n-1}\theta))}{(1-\exp(-(l_n-1)\theta))} \\
&\quad \times \frac{\exp(-(l_n-l_{n-1}-1)\theta)}{(1-\exp(-l_n\theta))}.
\end{aligned}$$

Par suite, $\mathbb{P}\left[{}^{L_n=l_n}\!/_{L_{n-1}=l_{n-1},\ldots,L_2=l_2,L_1=1}\right] = \mathbb{P}\left[{}^{L_n=l_n}\!/_{L_{n-1}=l_{n-1}}\right]$. Donc, toute l'information utile pour la prédiction du futur est contenue dans l'état présent du processus $\{L_n, n \geq 1\}$. □

Chapitre 3

Modèle LDM : Estimation du paramètre θ de dérive

Dans ce chapitre, notre but est d'estimer le paramètre de dérive θ, qu'on appelle aussi le « drift », d'un modèle LDM dans le cas d'une distribution sous-jacente de Gumbel de paramètres connus. Sans perte de généralité et afin de simplifier la présentation nous prendrons $\mu = 0$ (paramètre de position) et $\beta = 1$ (paramètre d'échelle), notée $G(0, 1)$. Nous utilisons plusieurs méthodes d'estimation, principalement : la méthode des moments et celle du maximum de vraisemblance. Nos méthodes d'estimation ponctuelle et par intervalle de confiance sont basées sur les distributions de probabilité du nombre N_T, des indicatrices δ_t et des indices L_n de records. Nous étudions le comportement asymptotique des différents estimateurs. Puis, nous validons ces résultats théoriques par des simulations numériques sous R en comparant la qualité de chaque estimateur selon plusieurs critères : le biais, l'écart-type et la probabilité de couverture de l'intervalle de confiance qui en découle.

3.1 En utilisant N_T : Estimation par maximum de vraisemblance (EMV)

Nous estimons ponctuellement et par intervalle de confiance le drift θ d'un modèle LDM dans le cas d'une distribution sous-jacente de loi $G(0, 1)$, en se basant sur le principe du maximum de vraisemblance (EMV) et en utilisant la distribution de probabilité du nombre de records N_T. Nous supposons que la longueur T de la série temporelles est fixée.

$N_T = m$	1	2	3	4	5	...	9	10
$\hat{\theta}_1$	$\simeq 0$	$\simeq 0$	$\simeq 0$	0.256	0.489	...	2.184	∞

TABLE 3.1 – Valeurs de $\hat{\theta}_1$ pour toutes les valeurs du N_T, quand $T = 10$

Ayant observé $N_T = m$, notre premier travail consiste à trouver $\hat{\theta}_1$ qui maximise par rapport à θ la fonction de masse de N_T donnée en (2.2.2) et que l'on rappelle ici :

$$\mathbb{P}_\theta [N_T = m] = \frac{\exp(-T\theta)\, \mathcal{S}(T, m \mid \vec{u}(\theta))}{\prod_{t=1}^{T} u_t(\theta)}. \tag{3.1.1}$$

A titre d'exemple, appliquons cette méthode d'estimation à un modèle LDM où $T = 10$. Pour chaque valeur possible de m allant de 1 à T, on calcule numériquement le θ qui maximise la fonction de masse (3.1.1) de N_T. Ceci donne la Table 3.1 suivante qui donne la valeur de $\hat{\theta}_1$, pour chaque valeur m possible.

Cette table montre qu'on ne peut pas dire grand chose sur le comportement de cet estimateur. En effet, pour certaines valeurs de m, $\hat{\theta}_1 = \infty$. Donc, on ne peut même pas calculer les moments de l'estimateur.

Mais, en adaptant un théorème bien connu (voir Plackett (1974), page 41), on peut néanmoins obtenir un intervalle de confiance de niveau exact pour la valeur θ du drift.

Théorème 3.1. *Ayant observé $N_T = m$, un intervalle de confiance $(\theta_{\mathcal{L}}, \theta_{\mathcal{U}})$ de niveau $1-\alpha$ pour le paramètre θ d'un modèle LDM est obtenu en résolvant :*

$$\mathbb{P}_{\theta_{\mathcal{U}}} [N_T \leq m] \;=\; \frac{\alpha}{2}, \tag{3.1.2}$$

et

$$\mathbb{P}_{\theta_{\mathcal{L}}} [N_T \geq m] = \frac{\alpha}{2}. \tag{3.1.3}$$

A titre d'exemple de l'application du Théorème 3.1, considérons un modèle LDM avec $T = 100$, $\alpha = 5\%$ et supposons qu'on a observé $m = 44$. L'application de (3.1.2) et (3.1.3) donne les valeurs : $\theta_{\mathcal{L}} = 0.398$ et $\theta_{\mathcal{U}} = 0.766$. Notons qu'en principe on peut faire ce travail a priori pour toutes les valeurs

de T, α et m et créer une table à trois entrées (ou un programme informatique) à laquelle on peut se référer en toutes circonstances. Mais comme la fonction de masse (3.1.1) de N_T est compliquée à manipuler, nous n'irons donc pas plus loin dans l'étude du comportement de cet estimateur.

3.2 En utilisant N_T : Estimation par une variante simple de la méthode des moments

Ici notre but est d'obtenir un estimateur ponctuel du drift θ en appliquant une variante de la méthode des moments à la distribution du nombre de records N_T. On espère pouvoir étudier plus finement cet estimateur que ce qui avait été possible de faire avec l'estimateur du maximum de vraisemblance de la section précédente. On rappelle que :

$$N_T = \sum_{t=1}^{T} \delta_t,$$

où les δ_t sont indépendants et de loi $Bernoulli\left(P_t\left(\theta\right)\right)$.

Ballerini et Resnick (1985), montrent que le taux de record asymptotique existe :

$$\begin{aligned} \lim_{t\to\infty} P_t\left(\theta\right) &= P\left(\theta\right), \\ &= 1 - e^{-\theta}. \end{aligned}$$

Par ailleurs, d'après le Théorème 2.3,

$$\mathbb{E}\left[\frac{N_T}{T}\right] \underset{T}{\longrightarrow} P\left(\theta\right),$$

et

$$\frac{N_T}{T} \longrightarrow P\left(\theta\right) \text{ presque sûrement.}$$

Ainsi, on peut définir un deuxième estimateur $\hat{\theta}_2$ du drift par l'expression :

$$\begin{aligned} \hat{\theta}_2 &= P^{-1}\left(\frac{N_T}{T}\right), \\ &= -\log\left(1 - \frac{N_T}{T}\right). \end{aligned}$$

Cette méthode d'estimation est basée sur le fait que le moment $\mathbb{E}\left[\frac{N_T}{T}\right]$ est approximé par le taux de record asymptotique $P(\theta)$ afin de rendre $\hat{\theta}_2$ plus facile à obtenir (car le calcul direct de la fonction réciproque de $\mathbb{E}\left[\frac{N_T}{T}\right]$ est plus compliqué). Ce qui fait qu'on peut voir cette méthode d'estimation comme une variante de la méthode des moments. Une version plus complexe de cet estimateur est présentée à la Section 3.2.2.

3.2.1 Loi de $\hat{\theta}_2$

Supposons que la suite $\{Y_t, t \geq 1\}$ est étendue à une suite doublement infinie $\{Y_t, -\infty < t < +\infty\}$. Nous pouvons définir pour $-\infty < t < +\infty$:

$$\begin{aligned} M_{t-1} &= \max_{1 \leq i < \infty} (Y_{t-i} - \theta i). \\ \delta_t^* &= \begin{cases} 1 & \text{si } Y_t > M_{t-1} \\ 0 & \text{sinon} \end{cases}. \end{aligned}$$

et

$$N_T^* = \sum_{t=1}^{T} \delta_t^*.$$

Or,

$$\begin{aligned} \delta_t^* = 1 &\iff Y_t > M_{t-1}, \\ &\iff Y_t > \max_{1 \leq i < \infty} (Y_{t-i} - \theta i), \\ &\iff (X_t - \theta t) > \max_{1 \leq i < \infty} (X_{t-i} - \theta (t-i) - \theta i), \\ &\iff X_t > \max_{1 \leq i < \infty} (X_{t-i}), \\ &\iff X_t \text{ est un record dans la suite doublement infinie} \\ &\quad \{X_t, -\infty < t < +\infty\}. \end{aligned}$$

De plus, définissons :

$$\begin{aligned} r_h &= \mathbb{E}\left(\delta_t^* \delta_{t+h}^*\right), \text{ avec } -\infty < t < +\infty \text{ et } h \geqslant 0, \\ &= \mathbb{P}\left[X_t > \max_{k<t} X_k \text{ et } X_{t+h} > \max_{k<t+h} X_k\right], \end{aligned}$$

$$\begin{aligned} r_h &= \int_{\mathbb{R}} \left(f_t(x) \left(\prod_{k=-\infty}^{t-1} F_k(x) \right) \times \left(\int_x^{+\infty} \left(f_{t+h}(z) \left(\prod_{k=t+1}^{(t+h)-1} F_k(z) \right) \right) \mathrm{d}z \right) \right) \mathrm{d}x, \\ &= \int_{\mathbb{R}} \left(f(x-\theta t) \left(\prod_{k=-\infty}^{t-1} F(x-\theta k) \right) \right. \\ &\quad \left. \times \left(\int_x^{+\infty} \left(f(z-\theta(t+h)) \left(\prod_{k=t+1}^{(t+h)-1} F(z-\theta k) \right) \right) \mathrm{d}z \right) \right) \mathrm{d}x. \end{aligned}$$

Posons,

$$G_h(x) = \prod_{k=1}^{h-1} F(x+\theta k),$$

et

$$G_\infty(x) = \prod_{k=1}^{\infty} F(x+\theta k). \tag{3.2.1}$$

L'existence de la limite (3.2.1) est assurée par Ballerini et Resnick (1985). En effectuant le changement de variable, $w = z - \theta(t+h)$, on obtient :

$$\begin{aligned} r_h &= \int_{\mathbb{R}} \left(f(x-\theta t) \left(\prod_{k=-\infty}^{t-1} F(x-\theta k) \right) \right. \\ &\quad \left. \times \left(\int_{x-\theta(t+h)}^{+\infty} \left(f(w) \left(\prod_{k=t+1}^{(t+h)-1} F(w+\theta(t+h-k)) \right) \right) \mathrm{d}w \right) \right) \mathrm{d}x, \\ &= \int_{\mathbb{R}} \left(f(x-\theta t) \left(\prod_{k=-\infty}^{t-1} F(x-\theta k) \right) \right. \\ &\quad \left. \times \left(\int_{x-\theta(t+h)}^{+\infty} \left(f(w) \left(\prod_{k=1}^{h-1} F(w+\theta k) \right) \right) \mathrm{d}w \right) \right) \mathrm{d}x, \\ &= \int_{\mathbb{R}} \left(f(x-\theta t) \left(\prod_{k=-\infty}^{t-1} F(x-\theta k) \right) \right. \\ &\quad \left. \times \left(\int_{x-\theta(t+h)}^{+\infty} f(w) G_h(w) \mathrm{d}w \right) \right) \mathrm{d}x. \end{aligned}$$

En effectuant un deuxième changement de variable, $y = x - \theta t$, on obtient :

$$\begin{aligned} r_h &= \int_{\mathbb{R}} \left(f(y) \left(\prod_{k=-\infty}^{t-1} F(y + \theta(t-k)) \right) \right. \\ &\quad \left. \times \left(\int_{y-\theta h}^{+\infty} f(w) G_h(w) \, \mathrm{d}w \right) \right) \mathrm{d}y, \\ &= \int_{\mathbb{R}} \left(f(y) \left(\prod_{k=1}^{\infty} F(y + \theta k) \right) \right. \\ &\quad \left. \times \left(\int_{y-\theta h}^{+\infty} f(w) G_h(w) \, \mathrm{d}w \right) \right) \mathrm{d}y, \\ &= \int_{\mathbb{R}} \left(f(y) G_\infty(y) \left(\int_{y-\theta h}^{+\infty} f(w) G_h(w) \, \mathrm{d}w \right) \right) \mathrm{d}y. \end{aligned} \tag{3.2.2}$$

Comme on a supposé que les Y_j sont de loi $G(0,1)$, et en posant $\tau = e^{-\theta}$, on a :

$$\begin{aligned} G_h(w) &= \prod_{k=1}^{h-1} F(w + \theta k), \\ &= \prod_{k=1}^{h-1} e^{-e^{-(w+\theta k)}}, \\ &= \prod_{k=1}^{h-1} F(w)^{\tau^k}, \\ &= F(w)^{\sum_{k=1}^{h-1} \tau^k}. \end{aligned}$$

De même,

$$\begin{aligned} G_\infty(y) &= F(y)^{\sum_{k=1}^{\infty} \tau^k}, \\ &= F(y)^{\frac{\tau}{1-\tau}}. \end{aligned} \tag{3.2.3}$$

Par suite

$$\int_{y-\theta h}^{+\infty} f(w) G_h(w) \, \mathrm{d}w = \int_{y-\theta h}^{+\infty} f(w) F(w)^{\sum_{k=1}^{h-1} \tau^k} \, \mathrm{d}w,$$

$$\begin{aligned}\int_{y-\theta h}^{+\infty} f(w)\, G_h(w)\, \mathrm{d}w &= \int_{F(y-\theta h)}^{1} u^{\sum_{k=1}^{h-1}\tau^k}\mathrm{d}u, \\ &= \frac{1-F(y)^{\sum_{k=0}^{h-1}\tau^{k-h}}}{\sum_{k=0}^{h-1}\tau^k}, \\ &= \frac{1-\tau}{1-\tau^h}\left(1-F(y)^{\frac{\tau^h-1}{(\tau-1)\tau^h}}\right). \qquad (3.2.4)\end{aligned}$$

On remplace (3.2.4) et (3.2.3) dans (3.2.2) pour obtenir

$$\begin{aligned} r_h &= \frac{1-\tau}{1-\tau^h}\left[\int_{\mathbb{R}} f(y)\, F(y)^{\frac{\tau}{1-\tau}}\,\mathrm{d}y - \int_{\mathbb{R}} f(y)\, F(y)^{\frac{\tau^{h+1}-\tau^h+1}{(1-\tau)\tau^h}}\,\mathrm{d}y\right], \\ &= \frac{1-\tau}{1-\tau^h}\left[(1-\tau) - \int_0^1 u^{\frac{\tau^{h+1}-\tau^h+1}{(1-\tau)\tau^h}}\,\mathrm{d}u\right], \\ &= \frac{1-\tau}{1-\tau^h}\left[(1-\tau) - (1-\tau)\,\tau^h\right], \\ &= (1-\tau)^2, \\ &= P^2(\theta). \end{aligned}$$

Maintenant on revisite le Théorème 2.3 du Chapitre précédent afin de préciser la valeur exacte de σ^2 :

Théorème 3.2. *Ballerini et Resnick* (1985, 1987), *Si* $F(\cdot)$ *est continue,* $\theta > 0$ *et* $\mathbb{E}[Y_+]^2 < \infty$ $(y_+ = \max(0, y))$ *:*

$$\sqrt{T}\left(\frac{N_T}{T} - P(\theta)\right) \longrightarrow N\left(0, \sigma^2\right),$$

avec

$$\sigma^2 = P(\theta) - P^2(\theta) + 2\sum_{h=1}^{\infty}\left(r_h - P^2(\theta)\right).$$

En appliquant ce théorème à notre cas où $F = G(0,1)$, on trouve

$$\sigma^2 = P(\theta) - P^2(\theta).$$

Or on a définit $\hat{\theta}_2 = P^{-1}\left(\frac{N_T}{T}\right)$. En utilisant la méthode delta, (Oehlert, 1992), on trouve que :

$$\sqrt{T}\left(P^{-1}\left(\frac{N_T}{T}\right) - P^{-1}\left(P\left(\theta\right)\right)\right) \xrightarrow{\mathcal{L}} N\left(0, \lambda\left(\theta\right)\right),$$

où $\xrightarrow{\mathcal{L}}$ dénote la convergence en loi et $\lambda\left(\theta\right) = \sigma^2\left[\frac{\mathrm{d}P^{-1}(\theta)}{\mathrm{d}\theta}\Big|_{\theta=P(\theta)}\right]^2 = \left(\frac{1-e^{-\theta}}{e^{-\theta}}\right)$.
Par suite,

$$\sqrt{T}\left(\hat{\theta}_2 - \theta\right) \xrightarrow{\mathcal{L}} N\left(0, \lambda\left(\theta\right)\right),$$

Ainsi

$$\frac{\sqrt{T}\left(\hat{\theta}_2 - \theta\right)}{\sqrt{\lambda\left(\theta\right)}} \xrightarrow{\mathcal{L}} N\left(0, 1\right). \tag{3.2.5}$$

Pour un niveau de confiance asymptotique $1-\alpha$, on peut alors obtenir un intervalle de confiance pour θ en s'appuyant sur le fait que pour tout $\theta > 0$:

$$\mathbb{P}_\theta\left[\hat{\theta}_2 - \frac{\sqrt{\lambda\left(\theta\right)}}{\sqrt{T}} z_{1-\alpha/2} \leq \theta \leq \hat{\theta}_2 + \frac{\sqrt{\lambda\left(\theta\right)}}{\sqrt{T}} z_{1-\alpha/2}\right] \longrightarrow 1-\alpha, \tag{3.2.6}$$

où $z_{1-\alpha/2}$ = quantile d'ordre $1-\alpha/2$ d'une $N\left(0,1\right)$. Or dans cette expression θ est inconnu et (3.2.6) est inutilisable. Cependant, d'après Slutsky (1925), on a aussi pour tout $\theta > 0$:

$$\mathbb{P}_\theta\left[\hat{\theta}_2 - \frac{\sqrt{\lambda\left(\hat{\theta}_2\right)}}{\sqrt{T}} z_{1-\alpha/2} \leq \theta \leq \hat{\theta}_2 + \frac{\sqrt{\lambda\left(\hat{\theta}_2\right)}}{\sqrt{T}} z_{1-\alpha/2}\right] \longrightarrow 1-\alpha.$$

Ainsi, un intervalle de confiance de niveau asymptotique $1-\alpha$ pour θ et basé sur N_T est donné par :

$$\left[\hat{\theta}_2 - \frac{\sqrt{\lambda\left(\hat{\theta}_2\right)}}{\sqrt{T}} z_{1-\alpha/2}, \hat{\theta}_2 + \frac{\sqrt{\lambda\left(\hat{\theta}_2\right)}}{\sqrt{T}} z_{1-\alpha/2}\right]. \tag{3.2.7}$$

3.2.2 $\hat{\theta}_2$ amélioré

Dans le contexte de la Section 3.2, nous cherchons à améliorer la qualité de $\hat{\theta}_2$. Notons que :

$$\begin{aligned}\mathbb{E}\left[\frac{N_T}{T}\right] &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\delta_t\right], \\ &= \frac{1}{T}\sum_{t=1}^{T}\frac{1-e^{-\theta}}{1-e^{-\theta t}}, \\ &= g_T\left(\theta\right).\end{aligned}$$

L'estimateur de la section précédente utilise le fait, prouvé dans Ballerini et Resnick (1985), que

$$g_T\left(\theta\right) \longrightarrow g\left(\theta\right) = P\left(\theta\right) = 1 - e^{-\theta},$$

et notre estimateur $\hat{\theta}_2$ du drift est défini par :

$$\begin{aligned}\hat{\theta}_2 &= g^{-1}\left(\frac{N_T}{T}\right), \\ &= -\log\left(1 - \frac{N_T}{T}\right).\end{aligned}$$

Comme déjà signalé à la Section 3.2, la version classique de la méthode des moments amènerait à définir $\hat{\theta}_2^M = g_T^{-1}\left(N_T/T\right)$. Mais comme le calcul direct de la fonction réciproque de g_T est compliqué, on va utiliser une approche alternative qui consiste à corriger le biais de $\hat{\theta}_2$ en utilisant le développement de Taylor de g^{-1} au voisinage de $g\left(\theta\right)$:

$$\begin{aligned}g^{-1}\left(\frac{N_T}{T}\right) &= g^{-1}\left(g\left(\theta\right)\right) + \left(\frac{N_T}{T} - g\left(\theta\right)\right) \times \left[\left.\frac{\mathrm{d}g^{-1}\left(\theta\right)}{\mathrm{d}\theta}\right|_{\theta=g(\theta)}\right] \\ &\quad + \frac{1}{2}\left(\frac{N_T}{T} - g\left(\theta\right)\right)^2 \times \left[\left.\frac{\mathrm{d}^2 g^{-1}\left(\theta\right)}{\mathrm{d}\theta^2}\right|_{\theta=g(\theta)}\right] + o\left(\left(\frac{N_T}{T} - g\left(\theta\right)\right)^2\right),\end{aligned}$$

$$\begin{aligned}\hat{\theta}_2 &= \theta + \left(\frac{N_T}{T} - g\left(\theta\right)\right) \times \frac{1}{1-g\left(\theta\right)} + \frac{1}{2}\left(\frac{N_T}{T} - g\left(\theta\right)\right)^2 \times \frac{1}{\left(1-g\left(\theta\right)\right)^2} \\ &\quad + o\left(\left(\frac{N_T}{T} - g\left(\theta\right)\right)^2\right).\end{aligned}$$

Ainsi,

$$\mathbb{E}\left[\hat{\theta}_2\right] \simeq \theta + \frac{g_T(\theta) - g(\theta)}{1 - g(\theta)} + \frac{1}{2\left(1 - g(\theta)\right)^2}\mathbb{E}\left[\left(\frac{N_T}{T} - g(\theta)\right)^2\right]. \quad (3.2.8)$$

Or,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{N_T}{T} - g(\theta)\right)^2\right] &= \mathbb{E}\left[\left(\frac{N_T}{T}\right)^2 + g^2(\theta) - 2\frac{N_T}{T}g(\theta)\right], \\
&= \mathbb{E}\left[\left(\frac{N_T}{T}\right)^2\right] + g^2(\theta) - 2g_T(\theta)\, g(\theta), \\
&= \mathbb{V}\left[\frac{N_T}{T}\right] + g_T^2(\theta) + g^2(\theta) - 2g_T(\theta)\, g(\theta), \\
&= \mathbb{V}\left[\frac{N_T}{T}\right] + \left(g_T(\theta) - g(\theta)\right)^2.
\end{aligned}$$

En rappelant que les δ_t sont indépendants dans le cas de la loi de Gumbel on a,

$$\begin{aligned}
\mathbb{V}\left[\frac{N_T}{T}\right] &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{V}\left[\delta_t\right], \\
&= \frac{1}{T^2}\sum_{t=1}^{T}P_t(\theta)\left(1 - P_t(\theta)\right),
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}\left[\frac{N_T}{T}\right] &= \frac{1}{T^2}\left(\mathbb{E}\left[N_T\right] - \sum_{t=1}^{T}P_t^2(\theta)\right), \\
&= \frac{1}{T}g_T(\theta) - \frac{1}{T^2}\sum_{t=1}^{T}P_t^2(\theta).
\end{aligned}$$

Alors,

$$\mathbb{E}\left[\left(\frac{N_T}{T} - g(\theta)\right)^2\right] = \frac{1}{T}g_T(\theta) - \frac{1}{T^2}\sum_{t=1}^{T}P_t^2(\theta) + \left(g_T(\theta) - g(\theta)\right)^2. \quad (3.2.9)$$

Par suite, en remplaçant (3.2.9) dans (3.2.8) on obtient,

$$\mathbb{E}\left[\hat{\theta}_2\right] \simeq \theta + G_T(\theta),$$

où,

$$G_T(\theta) = \frac{g_T(\theta) - g(\theta)}{1 - g(\theta)} + \frac{1}{2(1-g(\theta))^2}\left[\frac{1}{T}g_T(\theta) - \frac{1}{T^2}\sum_{t=1}^{T} P_t^2(\theta) + (g_T(\theta) - g(\theta))^2\right]. \tag{3.2.10}$$

Donc,

$$\mathbb{E}\left[\hat{\theta}_2\right] - \theta \simeq G_T(\theta).$$

Par ailleurs,

$$\begin{aligned}
\sqrt{T}(g_T(\theta) - g(\theta)) &= \sqrt{T}\left(\mathbb{E}\left[\frac{N_T}{T}\right] - P(\theta)\right), \\
&= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} P_t(\theta) - P(\theta)\right), \\
&= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}(P_t(\theta) - P(\theta))\right), \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(P_t(\theta) - P(\theta)).
\end{aligned}$$

Or,

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(P_t(\theta) - P(\theta)) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{1-e^{-\theta}}{1-e^{-t\theta}} - \left(1-e^{-\theta}\right)\right), \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{e^{-t\theta}\left(1-e^{-\theta}\right)}{(1-e^{-t\theta})}, \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sqrt{t}\frac{e^{-t\theta}\left(1-e^{-\theta}\right)}{(1-e^{-t\theta})\sqrt{t}}.
\end{aligned}$$

La suite $b_t = \sqrt{t}; t \geqslant 1$ est une suite croissante de réels positifs divergeant vers l'infini. Considérons la série de terme général positif

$$U_t = \frac{e^{-t\theta}\left(1-e^{-\theta}\right)}{\left(1-e^{-t\theta}\right)\sqrt{t}}, t \geq 1.$$

On a :

$$\begin{aligned}
\frac{U_{t+1}}{U_t} &= \frac{e^{-(t+1)\theta}\left(1-e^{-\theta}\right)}{\left(1-e^{-(t+1)\theta}\right)\sqrt{t+1}} \times \frac{\left(1-e^{-t\theta}\right)\sqrt{t}}{e^{-t\theta}\left(1-e^{-\theta}\right)}, \\
&= \frac{e^{-\theta}e^{-t\theta}\left(1-e^{-t\theta}\right)\sqrt{t}}{e^{-t\theta}\left(1-e^{-(t+1)\theta}\right)\sqrt{t}\sqrt{1+\frac{1}{t}}}, \\
&\longrightarrow e^{-\theta} < 1, \text{ car } \theta > 0.
\end{aligned}$$

Ainsi, par le critère de convergence de D'Alembert, la série $\sum_{t\geq 1} U_t$ est convergente. Par suite, en appliquant le Lemme de Kronecker, Feller (1971, page 239) :

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(P_t\left(\theta\right) - P\left(\theta\right)\right) &= \frac{1}{b_T}\sum_{t=1}^{T} b_t U_t, \\
&\longrightarrow 0.
\end{aligned}$$

Ainsi,

$$\begin{aligned}
\sqrt{T}\left(g_T\left(\theta\right) - g\left(\theta\right)\right) &= o\left(1\right), \\
\left(g_T\left(\theta\right) - g\left(\theta\right)\right) &= o\left(T^{-1/2}\right), \\
\left(g_T\left(\theta\right) - g\left(\theta\right)\right)^2 &= o\left(T^{-1}\right).
\end{aligned}$$

Alors, on peut corriger le biais de $\hat{\theta}_2$ à l'ordre T^{-1} en considérant le nouvel estimateur :

$$\begin{aligned}
\hat{\theta}_2^* &= \hat{\theta}_2 - G_T\left(\hat{\theta}_2\right), \\
&= H_T\left(\hat{\theta}_2\right),
\end{aligned}$$

avec $H_T\left(\theta\right) = \theta - G_T\left(\theta\right)$. Ainsi, d'après la méthode delta appliqué à (3.2.5) on a :

$$\sqrt{T}\left(H_T\left(\hat{\theta}_2\right) - H_T(\theta)\right) \xrightarrow{\mathcal{L}} N(0, \vartheta(\theta)).$$

où $\vartheta(\theta) = \lambda(\theta)\left(\frac{\mathrm{d}H_T(\theta)}{\mathrm{d}\theta}\right)^2$ et

$$\begin{aligned}
\frac{\mathrm{d}H_T(\theta)}{\mathrm{d}\theta} &= 1 - \frac{\mathrm{d}G_T(\theta)}{\mathrm{d}\theta}, \\
&= 1 - \frac{\frac{\mathrm{d}P(\theta)}{\mathrm{d}\theta}}{(1-P(\theta))^2}\left[\frac{1}{T}\sum_{t=1}^{T} P_t(\theta) - P(\theta)\right] \\
&\quad + \frac{1}{(1-P(\theta))}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{d}P_t(\theta)}{\mathrm{d}\theta} - \frac{\mathrm{d}P(\theta)}{\mathrm{d}\theta}\right] \\
&\quad + \frac{\frac{\mathrm{d}P(\theta)}{\mathrm{d}\theta}}{(1-P(\theta))^3}\left[\frac{1}{T^2}\sum_{t=1}^{T} P_t(\theta)(1-P_t(\theta)) + \left(\frac{1}{T}\sum_{t=1}^{T} P_t(\theta) - P(\theta)\right)^2\right] \\
&\quad + \frac{1}{2(1-P(\theta))^2}\left[\frac{1}{T^2}\sum_{t=1}^{T}\frac{\mathrm{d}P_t(\theta)}{\mathrm{d}\theta} - \frac{1}{T^2}\sum_{t=1}^{T}\frac{\mathrm{d}P_t^2(\theta)}{\mathrm{d}\theta}\right. \\
&\quad \left. + 2\left(\frac{1}{T}\sum_{t=1}^{T} P_t(\theta) - P(\theta)\right)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{d}P_t(\theta)}{\mathrm{d}\theta} - \frac{\mathrm{d}P(\theta)}{\mathrm{d}\theta}\right)\right],
\end{aligned}$$

avec, $P_t(\theta) = \frac{1-e^{-\theta}}{1-e^{-t\theta}}$ et $P(\theta) = 1 - e^{-\theta}$. Par suite,

$$\sqrt{T}\left(\hat{\theta}_2^* - H_T(\theta)\right) \xrightarrow{\mathcal{L}} N(0, \vartheta(\theta)). \tag{3.2.11}$$

De plus, afin de monter que $H_T(\theta) \longrightarrow \theta$, il suffit de prouver que $G_T(\theta) \longrightarrow 0$. Or, en appliquant le fait que $g_T(\theta) \longrightarrow P(\theta)$ sur l'équation (3.2.10) il nous reste à montrer que $\frac{1}{T^2}\sum_{t=1}^{T} P_t^2(\theta) \longrightarrow 0$. Mais, $P_t^2(\theta) \leq 1$ alors,

$$0 \leq \frac{1}{T^2}\sum_{t=1}^{T} P_t^2(\theta) \leq \frac{1}{T} \longrightarrow 0.$$

Et,

$$\frac{1}{T^2}\sum_{t=1}^{T} P_t^2(\theta) \longrightarrow 0.$$

Par conséquent, $H_T(\theta) \longrightarrow \theta$ et

$$\sqrt{T}\left(\hat{\theta}_2^* - \theta\right) \xrightarrow{\mathcal{L}} N\left(0, \vartheta\left(\theta\right)\right). \tag{3.2.12}$$

Ainsi, un intervalle de confiance de niveau asymptotique $1-\alpha$ pour θ est donné par :

$$\left[\hat{\theta}_2^* - \frac{\sqrt{\vartheta\left(\hat{\theta}_2^*\right)}}{\sqrt{T}} z_{1-\alpha/2}, \hat{\theta}_2^* + \frac{\sqrt{\vartheta\left(\hat{\theta}_2^*\right)}}{\sqrt{T}} z_{1-\alpha/2}\right]. \tag{3.2.13}$$

3.3 En utilisant $\{\delta_t, t \geq 1\}$: Estimation par maximum de vraisemblance

3.3.1 Non exhaustivité de N_T

Le passage de méthodes basées sur N_T aux indicatrices de records est justifié par le théorème suivant :

Théorème 3.3. *Dans un modèle LDM de distribution sous-jacente $G(0,1)$, la statistique $N_T = \sum_{t=1}^{T} \delta_t$, n'est pas exhaustive pour θ.*

Démonstration.

$$\begin{aligned}
\mathbb{P}\left[\delta_1=d_1,\ldots,\delta_T=d_T/N_T=m\right] &= \frac{\mathbb{P}\left[\delta_1 = d_1, \ldots, \delta_T = d_T; N_T = m\right]}{\mathbb{P}_\theta\left[N_T = m\right]}. \\
\text{L'équation (3.1.1) implique que} &= \begin{cases} 0 & \text{si } \sum_{t=1}^{T} d_t \neq m \\ \frac{\prod_{t=2}^{T} P_t^{d_t}(\theta)(1-P_t(\theta))^{1-d_t}}{\frac{\exp(-T\theta)\mathcal{S}(T,m|\vec{u}(\theta))}{\prod_{t=1}^{T} u_t(\theta)}} & \text{sinon} \end{cases}, \\
\text{et d'après (2.2.6)} &= \begin{cases} 0 & \text{si } \sum_{t=1}^{T} d_t \neq m \\ \frac{\prod_{t=2}^{T} P_t^{d_t}(\theta)(1-P_t(\theta))^{1-d_t}}{\mathcal{S}(T,m|\vec{u}(\theta))\prod_{t=1}^{T} P_t(\theta)} & \text{sinon} \end{cases}, \\
&= \begin{cases} 0 & \text{si } \sum_{t=1}^{T} d_t \neq m \\ \frac{1}{\mathcal{S}(T,m|\vec{u}(\theta))}\prod_{t=2}^{T}\left(\frac{1-P_t(\theta)}{P_t(\theta)}\right)^{1-d_t} & \text{sinon} \end{cases}.
\end{aligned}$$

Puisque $P_t(\theta)$ est fonction de θ, la probabilité conditionnelle dépend de θ et N_T n'est pas exhaustive. □

Alors, le passage de N_T à l'utilisation des δ_t donne en principe plus d'informations sur le paramètre θ. Ainsi, en se basant sur les indicatrices de records on peut espérer améliorer la qualité des estimateurs.

3.3.2 Calcul de $\hat{\theta}_3$

Notre but est d'estimer le drift θ dans le cas d'une distribution sous-jacente $G(0,1)$, en se basant sur la méthode du maximum de vraisemblance et en utilisant la distribution de probabilité des indicatrices de records $\{\delta_t, 1 \leq t \leq T\}$. Les δ_t sont indépendants et suivent la loi de Bernoulli de paramètre $Q_t(\tau) = \frac{1-\tau}{1-\tau^t}$, en utilisant pour simplifier la reparamétrisation $\tau = \exp(-\theta)$, $0 < \tau < 1$. Notre travail consiste alors à trouver le τ qui maximise l'expression :

$$\begin{aligned} L(\tau) &= \mathbb{P}[\delta_1, \ldots, \delta_T; \tau], \\ &= \prod_{t=2}^{T} Q_t(\tau)^{\delta_t} (1 - Q_t(\tau))^{1-\delta_t}. \end{aligned} \tag{3.3.1}$$

Pour ce faire, une approche consiste à dériver $\log L(\tau)$ par rapport à τ puis à calculer les racines de cette dérivée :

$$\begin{aligned} \log L(\tau) &= \sum_{t=2}^{T} [\delta_t \log(Q_t(\tau)) + (1-\delta_t) \log(1 - Q_t(\tau))], \\ &= \sum_{t=2}^{T} \left[\delta_t \log(1-\tau) - \delta_t \log\left(\tau\left(1-\tau^{t-1}\right)\right) + \log\left(\tau\left(1-\tau^{t-1}\right)\right) \right. \\ &\quad \left. - \log\left(1-\tau^t\right)\right]. \end{aligned}$$

Or $\sum_{t=2}^{T} \delta_t = N_T - 1$. Par suite,

$$\log L(\tau) = N_T \log(1-\tau) + (T - N_T) \log(\tau) - \log\left(1-\tau^T\right) - \sum_{t=2}^{T} \delta_t \log\left(1-\tau^{t-1}\right).$$

Ainsi, il faut trouver la valeur $\hat{\tau}$ de τ telle que :

$$\left(\frac{\mathrm{d} \log L(\tau)}{\mathrm{d}\tau}\right)_{\tau=\hat{\tau}} = 0. \tag{3.3.2}$$

Or,

$$\begin{aligned}\frac{\mathrm{d}\log L(\tau)}{\mathrm{d}\tau} &= \frac{-N_T}{1-\tau}+\frac{T-N_T}{\tau}+\frac{T\tau^{T-1}}{1-\tau^T}+\sum_{t=2}^{T}\delta_t\frac{(t-1)\tau^{t-2}}{1-\tau^{t-1}},\\ &= \frac{T\left(1-\frac{N_T}{T}-\tau\right)}{\tau(1-\tau)}+\frac{T\tau^{T-1}}{1-\tau^T}+\sum_{t=2}^{T}\delta_t\frac{(t-1)\tau^{t-2}}{1-\tau^{t-1}}. \quad (3.3.3)\end{aligned}$$

Une première difficulté est de trouver les racines de cette dérivée. Pour la surmonter, on a recours à une méthode numérique en prenant $\tau = \exp\left(-\hat{\theta}_2\right)$ comme point de départ de l'algorithme. Ce qui du coup justifie le travail accompli aux sections précédentes. Une seconde difficulté concerne le comportement de cet estimateur. En effet, les δ_t sont indépendants mais pas identiquement distribués de sorte que les théorèmes standards concernant le comportement des estimateurs de vraisemblance maximale ne s'appliquent pas. Cependant, en appliquant à notre contexte un théorème de Leroy *et al.* (2016) sur le comportement asymptotique des EMV dans le cas où les observations sont indépendantes mais non identiquement distribuées, on montre que $\hat{\tau}$ est consistant et que

$$\frac{\hat{\tau}-\tau}{\sqrt{I_T^{-1}(\tau)}} \xrightarrow{\mathcal{L}} N(0,1), \quad (3.3.4)$$

où $I_T(\tau)$ dénote l'information de Fisher (Pour la vérification des conditions du théorème de Leroy *et al.* (2016) voir la Section 3.7) :

$$I_T(\tau) = -\mathbb{E}\left[\frac{\mathrm{d}^2\log L(\tau)}{\mathrm{d}\tau^2}\right].$$

Or,

$$\begin{aligned}\frac{\mathrm{d}^2\log L(\tau)}{\mathrm{d}\tau^2} &= \frac{-N_T}{(1-\tau)^2}+\frac{N_T-T}{\tau^2}+\frac{T\tau^{T-2}\left(T+\tau^T-1\right)}{\left(1-\tau^T\right)^2}\\ &\quad+\sum_{t=2}^{T}\delta_t\frac{(t-1)\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{\left(1-\tau^{t-1}\right)^2}.\end{aligned}$$

Par suite,

$$\begin{aligned} I_T(\tau) &= \frac{1}{(1-\tau)^2}\sum_{t=1}^{T} Q_t(\tau) + \frac{1}{\tau^2}\left(T - \sum_{t=1}^{T} Q_t(\tau)\right) - \frac{T\tau^{T-2}\left(T+\tau^T-1\right)}{\left(1-\tau^T\right)^2} \\ &\quad - \sum_{t=2}^{T} \frac{(t-1)\,\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{\left(1-\tau^{t-1}\right)^2} Q_t(\tau). \end{aligned} \tag{3.3.5}$$

Notons que lorsque $N_T = T$, le maximum de la fonction de vraisemblance (3.3.1) n'est pas atteint et dans ce cas $\hat{\tau} = 0$ et $\hat{\theta}_3 = +\infty$. Mais cet événement a une probabilité qui tend vers zéro lorsque T tend vers l'infini.

Étudions maintenant la convergence du rapport $\frac{I_T(\tau)}{T}$. D'une part, si T est suffisamment grand et en utilisant le fait que $\frac{1}{T}\sum_{t=1}^{T} Q_t(\tau)$ converge vers $Q(\tau) = 1-\tau$ (Théorème 2.3) :

$$\begin{aligned} \frac{1}{T}\left[\frac{1}{(1-\tau)^2}\sum_{t=1}^{T} Q_t(\tau) + \frac{1}{\tau^2}\left(T - \sum_{t=1}^{T} Q_t(\tau)\right)\right] &\longrightarrow \frac{1-\tau}{(1-\tau)^2} + \frac{1}{\tau^2}\left(1-(1-\tau)\right), \\ &= \frac{1}{(1-\tau)\,\tau}, \end{aligned} \tag{3.3.6}$$

et

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T}\frac{T\tau^{T-2}\left(T+\tau^T-1\right)}{\left(1-\tau^T\right)^2} &= \lim_{T\to\infty} \frac{T\tau^T\left(1+{}^{\tau^T}\!/_{T} - {}^{1}\!/_{T}\right)}{\tau^2\left(1-\tau^T\right)^2}, \\ &= \lim_{T\to\infty} \frac{Te^{T\log(\tau)}\left(1+\frac{e^{T\log(\tau)}}{T} - {}^{1}\!/_{T}\right)}{\tau^2\left(1-\tau^T\right)^2}, \\ &= 0 \text{ car l'exponentielle l'emporte sur la puissance et } 0<\tau<1. \end{aligned} \tag{3.3.7}$$

D'autre part posons

$$\begin{aligned} H_T(\tau) &= \frac{1}{T}\sum_{t=2}^{T} \frac{(t-1)\,\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{\left(1-\tau^{t-1}\right)^2} Q_t(\tau), \\ &= \frac{1}{T}\sum_{t=2}^{T} t\frac{(t-1)\,\tau^{t-3}\left(t-2+\tau^{t-1}\right)(1-\tau)}{t\left(1-\tau^{t-1}\right)^2\left(1-\tau^t\right)}. \end{aligned}$$

La suite $b_t = t$; $t \geqslant 1$ est une suite croissante de réels positifs divergeant vers l'infini. Considérons la série de terme général positif

$$U_t = \frac{(t-1)\,\tau^{t-3}\,(t-2+\tau^{t-1})\,(1-\tau)}{t\,(1-\tau^{t-1})^2\,(1-\tau^t)},\ t \geq 2.$$

On a :

$$\begin{aligned}
\frac{U_{t+1}}{U_t} &= \frac{t\tau^{t-2}\,(t-1+\tau^t)\,(1-\tau)}{(t+1)\,(1-\tau^t)^2\,(1-\tau^{t+1})} \times \frac{t\,(1-\tau^{t-1})\,(1-\tau^t)}{(t-1)\,\tau^{t-3}\,(t-2+\tau^{t-1})\,(1-\tau)}, \\
&= \frac{t^3\tau\,(1-1/t+\tau^t/t)\,(1-\tau^{t-1})^2}{(t^2-1)\,(1-\tau^t)\,(1-\tau^{t+1})\,(t-2+\tau^{t-1})}, \\
&= \frac{t^3\tau\,(1-1/t+\tau^t/t)\,(1-\tau^{t-1})^2}{t^3\,(1-1/t^2)\,(1-\tau^t)\,(1-\tau^{t+1})\,(1-2/t+\tau^{t-1}/t)}, \\
&\longrightarrow \tau < 1.
\end{aligned}$$

Ainsi, d'après le critère de convergence de D'Alembert, la série $\sum_{t\geq 1} U_t$ est convergente. Par suite, en appliquant le Lemme de Kronecker :

$$\begin{aligned}
H_T(\tau) &= \frac{1}{b_T}\sum_{t=2}^{T} b_t U_t, \\
&\longrightarrow 0. \qquad (3.3.8)
\end{aligned}$$

Ainsi, d'après les équations (3.3.6), (3.3.7) et (3.3.8) :

$$\frac{I_T(\tau)}{T} \longrightarrow I(\tau) = \frac{1}{(1-\tau)\,\tau}.$$

Par suite, en se basant sur l'Équation (3.3.4) on a

$$\sqrt{T}\,(\hat{\tau}-\tau) \xrightarrow{\mathcal{L}} N\left(0, I^{-1}(\tau)\right).$$

Or, le drift θ de notre modèle LDM est donné par $\theta = -\log(\tau) = h(\tau)$. Donc, d'après la méthode delta on obtient :

$$\frac{\sqrt{T}\left(\hat{\theta}_3-\theta\right)}{\sqrt{(h'(\tau))^2\,I^{-1}(\tau)}} = \frac{\sqrt{T}\left(\hat{\theta}_3-\theta\right)}{\frac{1}{\tau}\sqrt{I^{-1}(\tau)}} \xrightarrow{\mathcal{L}} N(0,1),$$

et, pour un niveau de confiance asymptotique $1-\alpha$, un intervalle de confiance asymptotique de θ est donné par :

$$\left[\hat{\theta}_3 - \frac{\sqrt{I^{-1}(\hat{\tau})}}{\hat{\tau}\sqrt{T}} z_{1-\alpha/2}, \hat{\theta}_3 + \frac{\sqrt{I^{-1}(\hat{\tau})}}{\hat{\tau}\sqrt{T}} z_{1-\alpha/2}\right]. \tag{3.3.9}$$

Il ne reste plus qu'à réexprimer l'information de Fisher (3.3.5) dans la paramétrisation originale (c'est à dire en θ), en appliquant la méthode delta à l'équation (3.3.4) on obtient :

$$I_T^{-1}(\theta) = I_T^{-1}(\tau) \times \left(\frac{\mathrm{d}h(\tau)}{\mathrm{d}\tau}\right)^2.$$

Ainsi

$$I_T(\theta) = \frac{I_T(\tau)}{\left(\frac{\mathrm{d}h(\tau)}{\mathrm{d}\tau}\right)^2},$$

$$\begin{aligned} I_T(\theta) &= \tau^2 I_T(\tau). \\ &= \frac{e^{-2\theta}}{(1-e^{-\theta})^2}\sum_{t=1}^{T} P_t(\theta) + T - \sum_{t=1}^{T} P_t(\theta) - \frac{Te^{-T\theta}\left(T + e^{-T\theta} - 1\right)}{(1-e^{-T\theta})^2} \\ &\quad - \sum_{t=2}^{T} \frac{(t-1)\, e^{-(t-1)\theta}\left(t - 2 + e^{-(t-1)\theta}\right)}{\left(1-e^{-(t-1)\theta}\right)^2} P_t(\theta). \end{aligned}$$

et

$$\frac{\left(\hat{\theta}_3 - \theta\right)}{\sqrt{I_T^{-1}(\theta)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{3.3.10}$$

Alors, pour un niveau de confiance asymptotique $1-\alpha$, un intervalle de confiance pour θ est donné par :

$$\left[\hat{\theta}_3 - \sqrt{I_T^{-1}\left(\hat{\theta}_3\right)} z_{1-\alpha/2}, \hat{\theta}_3 + \sqrt{I_T^{-1}\left(\hat{\theta}_3\right)} z_{1-\alpha/2}\right]. \tag{3.3.11}$$

3.4 En utilisant $\{L_n, n \geq 1\}$: Estimation par maximum de vraisemblance

Dans le même contexte d'un modèle LDM avec une loi sous-jacente $G(0,1)$, calculons un quatrième estimateur du drift, en se basant sur le principe du maximum de vraisemblance et en utilisant la distribution de probabilité des indices de records $\{L_n, 1 \leq n \leq N_T = m\}$.

D'après la Proposition 2.11, la suite $\{L_n, 1 \leq n \leq N_T\}$ forme une chaîne de Markov de probabilité de transition :

$$\mathbb{P}\left[{}^{L_n=k}/_{L_{n-1}=j}\right] = \frac{(1-\tau)\left(1-\tau^{j}\right)\tau^{k-j-1}}{\left(1-\tau^{k-1}\right)\left(1-\tau^{k}\right)},$$

et d'espace d'état $E = \{L_1 = 1 < L_2 = l_2 < \cdots < L_{N_T} = l_{N_T} \leq T\}$. Notons que dans l'expression de la probabilité de transition on a réutilisé la reparamétrisation $\tau = e^{-\theta}$.

Le travail consiste alors à trouver τ qui maximise la vraisemblance de cette chaîne de Markov. On a

$$\begin{aligned}
T(\tau) &= \mathbb{P}\left[L_1 = l_1, \ldots, L_{N_T} = l_{N_T}; \tau\right], \\
&= \prod_{n=1}^{N_T-1} \mathbb{P}\left[{}^{L_{n+1}=l_{n+1}}/_{L_n=l_n}\right] \mathbb{P}\left[L_1 = l_1\right],\ l_1 = 1 \text{ (record trivial)}, \\
&= \frac{(1-\tau)\left(1-\tau^{l_1}\right)\tau^{l_2-l_1-1}}{\left(1-\tau^{l_2-1}\right)\left(1-\tau^{l_2}\right)} \times \cdots \times \frac{(1-\tau)\left(1-\tau^{l_{N_T-1}}\right)\tau^{l_{N_T}-l_{N_T-1}-1}}{\left(1-\tau^{l_{N_T}-1}\right)\left(1-\tau^{l_{N_T}}\right)}, \\
&= \frac{(1-\tau)^{N_T}\,\tau^{l_{N_T}-N_T}}{\left(1-\tau^{l_{N_T}}\right)\prod_{n=2}^{N_T}\left(1-\tau^{l_n-1}\right)}.
\end{aligned}$$

Pour trouver $\hat{\hat{\tau}}$ qui dénote notre estimateur par la méthode du maximum de vraisemblance, la méthode classique est de dériver $\log T(\tau)$ par rapport à τ, puis à calculer les racines de cette dérivée. On a

$$\begin{aligned}
\log T(\tau) &= N_T \log(1-\tau) + (l_{N_T} - N_T)\log(\tau) - \log\left(1-\tau^{l_{N_T}}\right) \\
&\quad - \sum_{n=2}^{N_T} \log\left(1-\tau^{l_n-1}\right).
\end{aligned}$$

Ensuite il faut trouver la valeur $\hat{\hat{\tau}}$ de τ telle que,

$$\left(\frac{\mathrm{d}\log T(\tau)}{\mathrm{d}\tau}\right)_{\tau=\hat{\hat{\tau}}} = 0.$$

Or,

$$\frac{\mathrm{d}\log T(\tau)}{\mathrm{d}\tau} = \frac{-N_T}{1-\tau} + \frac{l_{N_T}-N_T}{\tau} + \frac{l_{N_T}\tau^{l_{N_T}-1}}{1-\tau^{l_{N_T}}} + \sum_{n=2}^{N_T}\frac{(l_n-1)\,\tau^{l_n-2}}{1-\tau^{l_n-1}}, \tag{3.4.1}$$

et pour calculer les racines de cette fonction, on a recours à une méthode numérique.

Nous n'irons donc pas plus loin dans l'étude du comportement de cet estimateur, d'autant qu'il fait intervenir des observations formant une chaîne de Markov, donc ni indépendantes ni identiquement distribuées. De plus, en se basant sur des simulations numériques, on remarque à la prochaine section que ce nouvel estimateur, noté $\hat{\theta}_4$, est d'une qualité moins bonne que celui obtenu par la méthode précédente (EMV basée sur les indicatrices de records). Notamment le biais est plus important, ainsi que l'écart-type, si θ est petit.

3.5 Simulations numériques

Pour évaluer le comportement de nos estimateurs, en particulier l'utilité des approximations asymptotiques des équations (3.2.5), (3.2.12) et (3.3.10), des simulations ont été effectuées. Nous nous concentrons sur le biais, l'approximation asymptotique des écarts-types et la probabilité de couverture des intervalles de confiance. Pour ce faire, 5000 séries chronologiques ont été générées selon un modèle LDM pour différentes valeurs de θ avec $T = 100$ observations et une distribution sous-jacente $G(0,1)$. Pour chacune de ces séries, la suite des indicatrices de records $\{\delta_t, 1 \leq t \leq T\}$ a été extraite et les estimateurs $\hat{\theta}_2$, $\hat{\theta}_2^*$, $\hat{\theta}_3$ et $\hat{\theta}_4$ ont été calculés. De ces 5000 séries, les caractéristiques mentionnées précédemment ont été empiriquement estimées.

La Table 3.2 donne les biais des estimateurs. Les estimateurs $\hat{\theta}_2$ et $\hat{\theta}_4$ ont un grand biais et dans tous les cas $\hat{\theta}_2^*$ et $\hat{\theta}_3$ doivent être préférés, avec une légère préférence pour $\hat{\theta}_3$ lorsque θ est petit. La Table 3.3 présente les

	Biais			
θ	$\hat{\theta}_2$	$\hat{\theta}_2^*$	$\hat{\theta}_3$	$\hat{\theta}_4$
0.05	0.04	0.01	$\simeq 0.00$	0.04
0.10	0.03	0.01	$\simeq 0.00$	0.02
0.15	0.03	$\simeq 0.00$	$\simeq 0.00$	0.02
0.20	0.02	$\simeq 0.00$	$\simeq 0.00$	0.01
0.25	0.02	$\simeq 0.00$	$\simeq 0.00$	0.01

TABLE 3.2 – Biais empiriques de $\hat{\theta}_2$, $\hat{\theta}_2^*$, $\hat{\theta}_3$ et $\hat{\theta}_4$ pour différentes valeurs du drift θ avec $T = 100$ observations (Loi de Gumbel)

écarts-types. Les colonnes 2, 3 et 4 du tableau donnent les écarts-types empiriques (à partir des 5000 séries) de $\hat{\theta}_2$, $\hat{\theta}_2^*$ et $\hat{\theta}_4$. Les approximations asymptotiques des équations (3.2.5) et (3.2.12) apparaissent dans les colonnes 5 et 6. Les deux dernières colonnes montrent le même pour $\hat{\theta}_3$ et son approximation $\sqrt{I_T^{-1}(\theta)}$ de l'équation (3.3.10). Dans l'ensemble, les approximations des équations (3.2.5) et (3.2.12) sous-estiment les valeurs des écarts-types. L'équation (3.3.10) donne des résultats plus précis. Pour ces raisons on peut laisser tomber les estimateurs $\hat{\theta}_2$ et $\hat{\theta}_4$ et garder $\hat{\theta}_3$. Aussi, comme $\hat{\theta}_2^*$ est presque sans biais et a un écart-type très proche de l'expression $\sqrt{I_T^{-1}(\theta)}$, on peut aussi garder ce dernier estimateur avec $\sqrt{I_T^{-1}(\theta)}$ comme approximation de l'écart-type.

Enfin, pour vérifier l'exactitude de l'approximation asymptotique des équations (3.2.12) et (3.3.10), les probabilités de couverture des intervalles de confiance (3.2.13) et (3.3.11) (pour les probabilités de couverture basées sur $\hat{\theta}_2^*$, nous considérons $\sqrt{I_T^{-1}(\theta)}$ comme approximation de l'écart-type), pour différents niveaux de confiance $1-\alpha$, sont présentées dans la Table 3.4. En terme de probabilités de couverture l'estimateur $\hat{\theta}_3$ doit être préféré à $\hat{\theta}_2^*$. Dans les 3 dernières colonnes de la Table (3.4) les niveaux de confiance réels sont proches de ceux visés, sauf pour les petites valeurs de θ. Ceci est provoqué par l'asymétrie de la distribution de $\hat{\theta}_3$ qui vient de la contrainte $\theta > 0$. Sinon, l'utilisation de $\hat{\theta}_3$ devrait conduire à une inférence correcte.

θ	$\sqrt{\widehat{\mathbb{V}}\left(\hat{\theta}_2\right)}$	$\sqrt{\widehat{\mathbb{V}}\left(\hat{\theta}_2^*\right)}$	$\sqrt{\widehat{\mathbb{V}}\left(\hat{\theta}_4\right)}$	$\sqrt{\frac{\lambda(\theta)}{T}}$	$\sqrt{\frac{\vartheta(\theta)}{T}}$	$\sqrt{\widehat{\mathbb{V}}\left(\hat{\theta}_3\right)}$	$\sqrt{I_T^{-1}(\theta)}$
0.05	0.028	0.031	0.321	0.023	0.027	0.030	0.030
0.10	0.035	0.038	0.039	0.032	0.035	0.037	0.037
0.15	0.043	0.044	0.046	0.040	0.042	0.043	0.044
0.20	0.049	0.050	0.051	0.047	0.049	0.050	0.050
0.25	0.055	0.056	0.058	0.053	0.055	0.056	0.056

TABLE 3.3 – Écarts-types de $\hat{\theta}_2$, $\hat{\theta}_2^*$, $\hat{\theta}_3$ et $\hat{\theta}_4$ pour différentes valeurs du drift θ avec $T = 100$ observations (Loi de Gumbel). Les colonnes 2, 3 et 4 donnent les écarts-types empiriques de $\hat{\theta}_2$, $\hat{\theta}_2^*$ et $\hat{\theta}_4$. Les approximations asymptotiques des équations (3.2.5) et (3.2.12) apparaîssent dans les colonnes 5 et 6. Les deux dernières colonnes montrent le même pour $\hat{\theta}_3$ et son approximation $\sqrt{I_T^{-1}(\theta)}$ de l'équation (3.3.10).

	$\hat{\theta}_2^*$			$\hat{\theta}_3$		
	$1-\alpha$			$1-\alpha$		
θ	90%	95%	99%	90%	95%	99%
0.05	92.8%	98.5%	99.9%	89.5%	98.9%	99.7%
0.10	86.5%	93.2%	97.5%	89.2%	93.6%	97.7%
0.15	89.1%	94.1%	97.9%	89.5%	94.0%	98.0%
0.20	89.5%	94.9%	97.8%	89.2%	94.2%	98.1%
0.25	86.7%	92.8%	98.2%	89.5%	94.5%	98.2%

TABLE 3.4 – Probabilités de couverture des intervalles de confiance (3.2.13) et (3.3.11) pour différentes valeurs du drift θ et du niveau de confiance $1-\alpha$ avec $T = 100$ observations (Loi de Gumbel)

3.6 Conclusion

Nous avons présenté plusieurs méthodes d'estimation du drift d'un modèle LDM avec une distribution sous-jacente de Gumbel de paramètres connus, en se basant sur le nombre N_T, les indicatrices δ_t et les indices L_n de records respectivement et en utilisant la méthode des moments ou le principe du maximum de vraisemblance.

Le meilleur estimateur sur les plans considérés (biais, écart-type et probabilité de couverture) a été obtenu en utilisant les indicatrices de records et en se basant sur le principe du maximum de vraisemblance. On a aussi donné la preuve de son comportement asymptotique normal.

Une restriction importante du présent travail est cette limitation à la loi de Gumbel. Son avantage ici est que les δ_t sont indépendants sous cette loi, ce qui a permis l'étude asymptotique de $\hat{\theta}_2, \hat{\theta}_2^*$ et $\hat{\theta}_3$, et la loi de Gumbel est la seule qui possède cette propriété. Cette loi est certes une loi importante, mais ce n'est pas la seule possible dans un contexte de records et cette restriction, causée par des raisons techniques, laisse un goût d'inachevé. On serait heureux de pouvoir la lever. Nous n'y sommes pas arrivé directement. Cependant, en emboîtant le présent modèle LDM dans un autre modèle, le modèle de Yang-Nevzorov, il s'est avéré possible de retrouver des δ_t indépendants pour toutes les lois possibles et donc de généraliser notre travail à cette famille plus large de modèles.

Dans le chapitre suivant, nous considérons le cas du modèle de Yang-Nevzorov avec une distribution sous-jacente quelconque. Notre but est de calculer des estimateurs du paramètre de ce modèle et d'étudier leur comportement asymptotique en suivant de très près le scénario du présent chapitre.

3.7 Vérification des conditions de Leroy *et al.* (2016)

Dans la Section 3.3 on a obtenu un estimateur du drift θ d'un modèle LDM de distribution sous-jacente de Gumbel $G(0,1)$ en se basant sur le principe du maximum de vraisemblance et en utilisant la loi des $\{\delta_t, 1 \leq t \leq T\}$ qui forment une suite de variables aléatoires indépendantes, chacune distribuée

suivant la loi de Bernoulli de fonction de masse :

$$f_t(d,\tau) = \begin{cases} Q_t(\tau) = \frac{1-\tau}{1-\tau^t} & \text{si } d = 1 \\ 1 - Q_t(\tau) = \frac{\tau-\tau^t}{1-\tau^t} & \text{si } d = 0 \end{cases},$$

avec, $\tau = \exp(-\theta)$, $0 < \tau < 1$.

Le théorème dont l'application permet de confirmer que notre estimateur $\hat{\tau}$ est asymptotiquement sans biais et de loi normale est dû à Leroy, Dauxois et Tubert-Bitter (2016). Nous l'exprimons en utilisant la notation du présent chapitre.

Théorème 3.4. *Si les Hypothèses 1-10 (ci-dessous) sont satisfaites, l'estimateur du maximum de vraisemblance $\hat{\tau}$ est un estimateur convergent de τ^0 (la valeur réelle de τ), et pour tout $\varepsilon > 0$, $\mathbb{P}[|\hat{\tau} - \tau^0| > \varepsilon] \longrightarrow 0$. Si en plus l'Hypothèse 11 (ci-dessous) est satisfaite, la variable aléatoire $Z = \sqrt{T}(\hat{\tau} - \tau^0)$ converge asymptotiquement vers la loi normale d'espérance 0 et de variance $[I(\tau^0)]^{-1}$, avec $I(\tau^0) = -\lim_T \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left(\frac{d^2 \log f_t(\delta_t,\tau)}{d\tau^2}\Big|_{\tau^0} \right)$.*

Dans la présente section, notre but est de montrer que nous sommes dans un cas où les hypothèses suivantes sont vérifiées :

Hypothèse 1. L'estimateur du maximum de vraisemblance $\hat{\tau}$ est solution de l'équation (3.3.2).

Démonstration. Par définition notre estimateur est la solution de l'équation (3.3.2). □

Hypothèse 2. L'équation (3.3.2) a une racine unique.

Démonstration. On va montrer, que si T est suffisamment grand $\frac{1}{T}\frac{\mathrm{d}^2 \log L(\tau)}{\mathrm{d}\tau^2} < 0$, presque sûrement sur $0 < \tau < 1$, c'est-à-dire la fonction de vraisemblance est presque sûrement concave sur son domaine. Or,

$$\begin{aligned} \frac{1}{T}\frac{\mathrm{d}^2 \log L(\tau)}{\mathrm{d}\tau^2} &= \frac{1}{T}\left[\frac{-N_T}{(1-\tau)^2} + \frac{N_T - T}{\tau^2} + \frac{T\tau^{T-2}\left(T + \tau^T - 1\right)}{\left(1-\tau^T\right)^2} \right. \\ &\quad \left. + \sum_{t=2}^{T} \delta_t \frac{(t-1)\tau^{t-3}(t - 2 + \tau^{t-1})}{(1-\tau^{t-1})^2} \right]. \end{aligned}$$

Ballerini et Resnick (1985) montrent que dans un modèle LDM avec une distribution sous-jacente de Gumbel, $\frac{N_T}{T} \longrightarrow (1-\tau)$ presque sûrement. Par suite, presque sûrement :

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\left[\frac{-N_T}{(1-\tau)^2}+\frac{1}{\tau^2}\left(N_T-T\right)\right] &= \frac{-(1-\tau)}{(1-\tau)^2}+\frac{1}{\tau^2}\left((1-\tau)-1\right),\\ &= \frac{-1}{(1-\tau)\tau}<0, \qquad (3.7.1)\end{aligned}$$

et,

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\frac{T\tau^{T-2}\left(T+\tau^T-1\right)}{\left(1-\tau^T\right)^2} &= \lim_{T\to\infty}\frac{T\tau^T\left(1+{}^{\tau^T}\!/_T-{}^1\!/_T\right)}{\tau^2\left(1-\tau^T\right)^2},\\ &= \lim_{T\to\infty}\frac{Te^{T\log(\tau)}\left(1+\frac{e^{T\log(\tau)}}{T}-{}^1\!/_T\right)}{\tau^2\left(1-\tau^T\right)^2},\\ &= 0. \qquad (3.7.2)\end{aligned}$$

D'autre part,

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T}\delta_t\frac{(t-1)\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{\left(1-\tau^{t-1}\right)^2} &\leq \frac{1}{T}\sum_{t=2}^{T}\frac{(t-1)\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{\left(1-\tau^{t-1}\right)^2},\\ &= \frac{1}{T}\sum_{t=2}^{T}t\frac{(t-1)\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{t\left(1-\tau^{t-1}\right)^2}. \qquad (3.7.3)\end{aligned}$$

La suite $b_t = t;\ t \geqslant 1$ est une suite croissante de réels positifs divergeant vers l'infini. Considérons la série de terme général positif

$$U_t = \frac{(t-1)\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{t\left(1-\tau^{t-1}\right)^2};\ t \geq 2.$$

On a :

$$\frac{U_{t+1}}{U_t} = \frac{t\tau^{t-2}\left(t-1+\tau^t\right)}{(t+1)\left(1-\tau^t\right)^2}\times\frac{t\left(1-\tau^{t-1}\right)^2}{(t-1)\tau^{t-3}\left(t-2+\tau^{t-1}\right)},$$

$$\begin{aligned}\frac{U_{t+1}}{U_t} &= \frac{t^3\tau\left(1-{}^1\!/_t+{}^{\tau^t}\!/_t\right)\left(1-\tau^{t-1}\right)^2}{\left(t^2-1\right)\left(1-\tau^t\right)^2\left(t-2+\tau^{t-1}\right)},\\ &= \frac{t^3\tau\left(1-{}^1\!/_t+{}^{\tau^t}\!/_t\right)\left(1-\tau^{t-1}\right)^2}{t^3\left(1-{}^1\!/_{t^2}\right)\left(1-\tau^t\right)^2\left(1-{}^2\!/_t+{}^{\tau^{t-1}}\!/_t\right)},\\ &\longrightarrow \tau<1.\end{aligned}$$

Ainsi, d'après le critère de convergence de D'Alembert la série $\sum_{t\geq 1} U_t$ est convergente. Par suite, en appliquant le lemme de Kronecker :

$$\begin{aligned}\frac{1}{T}\sum_{t=2}^{T} t\frac{(t-1)\,\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{t\left(1-\tau^{t-1}\right)^2} &= \frac{1}{b_T}\sum_{t=2}^{T} b_t U_t,\\ &\longrightarrow 0.\end{aligned}$$

Alors, en se basant sur l'Équation (3.7.3) et d'après le critère de convergence de comparaison :

$$\frac{1}{T}\sum_{t=2}^{T}\delta_t\frac{(t-1)\,\tau^{t-3}\left(t-2+\tau^{t-1}\right)}{\left(1-\tau^{t-1}\right)^2}\longrightarrow 0. \tag{3.7.4}$$

Donc, d'après les équations (3.7.1), (3.7.2) et (3.7.4) on a presque sûrement

$$\frac{1}{T}\frac{\mathrm{d}^2\log L(\tau)}{\mathrm{d}\tau^2}\longrightarrow\frac{-1}{(1-\tau)\,\tau}<0.$$

□

Hypothèse 3. Les dérivées :

$$\frac{\mathrm{d}\log f_t(\cdot,\tau)}{\mathrm{d}\tau},\ \frac{\mathrm{d}^2\log f_t(\cdot,\tau)}{\mathrm{d}\tau^2}\text{ et }\frac{\mathrm{d}^3\log f_t(\cdot,\tau)}{\mathrm{d}\tau^3},$$

existent pour presque toutes valeurs de δ_t (qui sont, on le rappelle, 0 et 1).

Démonstration.

$$\log f_t(d,\tau) = \begin{cases}\log(1-\tau)-\log\left(1-\tau^t\right) & \text{si } d=1\\ \log(\tau-\tau^t)-\log\left(1-\tau^t\right) & \text{si } d=0\end{cases}.$$

$$\frac{\mathrm{d}\log f_t(d,\tau)}{\mathrm{d}\tau} = \begin{cases} \frac{-1}{1-\tau}+\frac{t\tau^{t-1}}{1-\tau^t} & \text{si } d=1 \\ \frac{1-t\tau^{t-1}}{\tau-\tau^t}+\frac{t\tau^{t-1}}{1-\tau^t} & \text{si } d=0 \end{cases}.$$

$$\frac{\mathrm{d}^2\log f_t(d,\tau)}{\mathrm{d}\tau^2} = \begin{cases} \frac{-1}{(1-\tau)^2}+\frac{t(t-1)\tau^{t-2}+t\tau^{2t-2}}{(1-\tau^t)^2} & \text{si } d=1 \\ \frac{t(1-t)\tau^{t-1}-t\tau^{2t-2}}{(\tau-\tau^t)^2}+\frac{t(t-1)\tau^{t-2}+t\tau^{2t-2}}{(1-\tau^t)^2} & \text{si } d=0 \end{cases}.$$

$$\frac{\mathrm{d}^3\log f_t(d,\tau)}{\mathrm{d}\tau^3} = \begin{cases} \frac{-2}{(1-\tau)^3}+\frac{t(t-1)(t-2)\tau^{t-3}}{(1-\tau^t)^3} & \\ +\frac{t(t-1)(t+4)\tau^{2t-3}+2t\tau^{3t-3}}{(1-\tau^t)^3} & \text{si } d=1 \\ \frac{t\left(2t-t^2-3\right)\tau^{t-1}+t(t-1)(-t-3)\tau^{2t-2}-2t\tau^{3t-3}}{(\tau-\tau^t)^3} & \\ +\frac{t(t-1)(t-2)\tau^{t-3}+t(t-1)(t+4)\tau^{2t-3}+2t\tau^{3t-3}}{(1-\tau^t)^3} & \text{si } d=0 \end{cases}.$$

□

Hypothèse 4. La dérivée $\frac{\mathrm{d}f_t(d,\tau)}{\mathrm{d}\tau}$ est une fonction intégrable sur le support S_t de δ_t et :

$$\int_{S_t}\frac{\mathrm{d}f_t(d,\tau)}{\mathrm{d}\tau}\,\mathrm{d}d = \frac{\mathrm{d}}{\mathrm{d}\tau}\int_{S_t} f_t(d,\tau)\,\mathrm{d}d. \tag{3.7.5}$$

Hypothèse 5. La dérivée $\frac{\mathrm{d}^2 f_t(d,\tau)}{\mathrm{d}\tau^2}$ est une fonction intégrable sur le support S_t de δ_t et :

$$\int_{S_t}\frac{\mathrm{d}^2 f_t(d,\tau)}{\mathrm{d}\tau^2}\,\mathrm{d}d = \frac{\mathrm{d}}{\mathrm{d}\tau}\int_{S_t}\frac{\mathrm{d}f_t(d,\tau)}{\mathrm{d}\tau}\,\mathrm{d}d. \tag{3.7.6}$$

Démonstration. Les Hypothèses 4 et 5 sont vraies car les fonctions à intégrer sont discrètes. □

Hypothèse 6.

$$\frac{1}{T}\sum_{t=2}^{T}\frac{\mathrm{d}\log f_t(\delta_t,\tau)}{\mathrm{d}\tau}\xrightarrow[T\to\infty]{\mathcal{P}}\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}\log f_t(\delta_t,\tau)}{\mathrm{d}\tau}\right)=0,$$

où $\xrightarrow{\mathcal{P}}$ dénote la convergence en probabilité.

Hypothèse 7.

$$\mathbb{E}\left(\frac{\mathrm{d}^2\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^2}\right)\text{ et }\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^2\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^2}\right),$$

existent, et

$$\frac{1}{T}\sum_{t=2}^{T}\frac{\mathrm{d}^2 \log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^2} \xrightarrow{\mathcal{P}} \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^2 \log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^2}\right).$$

Hypothèse 8.

$$\mathbb{E}\left(\frac{\mathrm{d}^3 \log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^3}\right) \text{ et } \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^3 \log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^3}\right),$$

existent, et

$$\frac{1}{T}\sum_{t=2}^{T}\frac{\mathrm{d}^3 \log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^3} \xrightarrow[T\to\infty]{\mathcal{P}} \lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^3 \log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^3}\right).$$

Démonstration. La loi faible des grands nombres (Feller (1968) page 253), donne une condition suffisante pour les convergences en probabilité des Hypothèses 6-8.

$$\begin{aligned}
\mathbb{E}\left(\frac{\mathrm{d}\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau}\right) &= \frac{\mathrm{d}\log f_t\left(0,\tau\right)}{\mathrm{d}\tau}f_t\left(0,\tau\right)+\frac{\mathrm{d}\log f_t\left(1,\tau\right)}{\mathrm{d}\tau}f_t\left(1,\tau\right),\\
&= \left(\frac{1-t\tau^{t-1}}{\tau-\tau^t}+\frac{t\tau^{t-1}}{1-\tau^t}\right)\left(\frac{\tau-\tau^t}{1-\tau^t}\right)\\
&\quad +\left(\frac{-1}{1-\tau}+\frac{t\tau^{t-1}}{1-\tau^t}\right)\left(\frac{1-\tau}{1-\tau^t}\right),\\
&= \frac{1-t\tau^{t-1}}{1-\tau^t}-\frac{1}{1-\tau^t}+\frac{t\tau^{t-1}}{1-\tau^t},\\
&= 0.
\end{aligned}$$

Par suite, l'hypothèse 6 est vérifiée.

$$\begin{aligned}
\mathbb{E}\left(\frac{\mathrm{d}^2\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^2}\right) &= \frac{\mathrm{d}^2\log f_t\left(0,\tau\right)}{\mathrm{d}\tau^2}f_t\left(0,\tau\right)+\frac{\mathrm{d}^2\log f_t\left(1,\tau\right)}{\mathrm{d}\tau^2}f_t\left(1,\tau\right),\\
&= \frac{-1}{\left(1-\tau\right)\left(1-\tau^t\right)}+\frac{t\left(1-t\right)\tau^{t-1}-t\tau^{2t-2}-1}{\left(\tau-\tau^t\right)\left(1-\tau^t\right)}\\
&\quad +\frac{t\left(t-1\right)\tau^{t-2}+t\tau^{2t-2}}{\left(1-\tau^t\right)^2}.
\end{aligned}$$

Il faut étudier $\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^2\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^2}\right)$. Or,

$$\lim_{t\to\infty}\frac{-1}{(1-\tau)(1-\tau^t)}=\frac{-1}{(1-\tau)},$$

et

$$\begin{aligned}\lim_{t\to\infty}\frac{t(1-t)\tau^{t-1}-t\tau^{2t-2}-1}{(\tau-\tau^t)(1-\tau^t)} &= \lim_{t\to\infty}\frac{t(1-t)e^{(t-1)\log\tau}-te^{(2t-2)\log\tau}-1}{(\tau-\tau^t)(1-\tau^t)},\\ &= -\frac{1}{\tau}\text{ car l'exponentielle l'emporte sur la puissance et } 0<\tau<1,\end{aligned}$$

de plus,

$$\lim_{t\to\infty}\frac{t(t-1)\tau^{t-2}+t\tau^{2t-2}}{(1-\tau^t)^2}=0.$$

Alors,

$$\begin{aligned}\lim_{t\to\infty}\mathbb{E}\left(\frac{\mathrm{d}^2\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^2}\right) &= \frac{-1}{(1-\tau)}-\frac{1}{\tau},\\ &= -\frac{1}{\tau(1-\tau)}.\end{aligned}$$

Ainsi, d'après le lemme de Cesàro, Hardy (1949, page 100) :

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^2\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^2}\right) = -\frac{1}{\tau(1-\tau)}.$$

Par suite, l'hypothèse 7 est vérifiée.

De même pour l'hypothèse 8 on trouve que

$$\begin{aligned}\mathbb{E}\left(\frac{\mathrm{d}^3\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^3}\right) &= \frac{-2}{(1-\tau)^2(1-\tau^t)}\\ &\quad+\frac{t\left(2t-t^2-3\right)\tau^{t-1}+t(t-1)(-t-3)\tau^{2t-2}-2t\tau^{3t-3}}{(\tau-\tau^t)^2(1-\tau^t)}\\ &\quad+\frac{t(t-1)(t-2)\tau^{t-3}+t(t-1)(t+4)\tau^{2t-3}+2t\tau^{3t-3}}{(1-\tau^t)^3},\end{aligned}$$

et

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^3\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^3}\right)=\frac{-2}{(1-\tau)^2}.$$

□

Hypothèse 9. Il existe M tel que pour tout τ dans un voisinage de τ^0 :

$$\left|\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^3\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^3}\right)\right| < M.$$

Démonstration. On a déjà montré que :

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^3\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^3}\right) = \frac{-2}{\left(1-\tau\right)^2}.$$

Par suite, pour un certain $\zeta > 0$ assez petit tel que $(\tau^0 - \zeta) > 0$ et $(\tau^0 + \zeta) < 1$ il suffit de prendre le voisinage $V\left(\tau^0\right) =]\tau^0 - \zeta, \tau^0 + \zeta[$ de τ^0 et de poser :

$$M = \frac{2}{\left(1-\left(\tau^0+\zeta\right)\right)^2}.$$

Ainsi, pour tout τ dans $V\left(\tau^0\right)$ on obtient :

$$\begin{aligned}\left|\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^3\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^3}\right)\right| &= \frac{2}{\left(1-\tau\right)^2},\\ &< M.\end{aligned}$$

□

Hypothèse 10. La matrice $I\left(\tau^0\right)$ est définie positive.

Démonstration. Dans notre cas $I\left(\tau^0\right) = -\lim_T \frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\left.\frac{\mathrm{d}^2\log f_t(\delta_t,\tau)}{\mathrm{d}\tau^2}\right|_{\tau^0}\right)$. Or pour tout τ

$$-\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left(\frac{\mathrm{d}^2\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau^2}\right) = \frac{1}{\tau\left(1-\tau\right)} > 0.$$

Par suite :

$$\left[I\left(\tau^0\right)\right]^{-1} = \tau^0\left(1-\tau^0\right) > 0.$$

□

Hypothèse 11. Pour tout $\varepsilon > 0$,

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{T}\mathbb{E}\left[\left(\frac{\mathrm{d}\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau}\Big|_{\tau^0}\right)^2 \mathrm{Ind}\left\{\left(\left(\frac{\mathrm{d}\log f_t\left(\delta_t,\tau\right)}{\mathrm{d}\tau}\Big|_{\tau^0}\right)^2\right)^{1/2} > \varepsilon\sqrt{T}\right\}\right] = 0. \tag{3.7.7}$$

Avec $\mathrm{Ind}\left\{A\right\}$ est l'indicatrice de l'ensemble A.

Démonstration. Pour un certain $\varepsilon > 0$ fixé, posons pour $t \geq 2$,

$$\begin{aligned} B_t &= \frac{\mathrm{d}\log f_t\left(d,\tau\right)}{\mathrm{d}\tau}, \\ &= \begin{cases} b_{1t} = \frac{-1}{1-\tau} + \frac{t\tau^{t-1}}{1-\tau^t} & \text{si } d = 1 \\ b_{2t} = \frac{1-t\tau^{t-1}}{\tau-\tau^t} + \frac{t\tau^{t-1}}{1-\tau^t} & \text{si } d = 0 \end{cases}. \end{aligned}$$

Rappelons que,

$$\delta_t = \begin{cases} 1 \text{ avec une probabilité } Q_{1t} = \frac{1-\tau}{1-\tau^t} \\ 0 \text{ avec une probabilité } Q_{2t} = \frac{\tau-\tau^t}{1-\tau^t} \end{cases}.$$

D'autre part, dans l'Hypothèse 6 on a montré que :

$$\mathbb{E}\left[B_t\right] = 0.$$

Ainsi, (3.7.7) s'écrit :

$$\lim_{T\longrightarrow\infty}\frac{1}{T}\sum_{t=2}^{T}\int_{|b_t|>\varepsilon\sqrt{T}} b_t^2 \mathrm{d}F_{B_t}\left(b_t\right) = 0.$$

Or,

$$\begin{aligned} \int_{|b_t|>\varepsilon\sqrt{T}} b_t^2 \mathrm{d}F_{B_t}\left(b_t\right) &= \sum_{\substack{k=1 \\ |b_{kt}|>\varepsilon\sqrt{T}}}^{2} b_{kt}^2 Q_{kt}, \\ &\leq \mathrm{Ind}\left\{|b_{1t}| > \varepsilon\sqrt{t}\right\} \times b_{1t}^2 Q_{1t} \\ &\quad + \mathrm{Ind}\left\{|b_{2t}| > \varepsilon\sqrt{t}\right\} \times b_{2t}^2 Q_{2t}, \end{aligned}$$

et

$$\begin{aligned} b_{1t}^2 Q_{1t} &= \left(\frac{-1}{1-\tau} + \frac{t\tau^{t-1}}{1-\tau^t}\right)^2 \frac{1-\tau}{1-\tau^t}, \\ &= \frac{1}{1-\tau} - \frac{2t\tau^{t-1}}{(1-\tau^t)^2} + \frac{(t\tau^{t-1})^2(1-\tau)}{(1-\tau^t)^3}, \\ &\longrightarrow \frac{1}{1-\tau} \text{ car l'exponentielle l'emporte sur la puissance et } 0<\tau<1. \end{aligned}$$

De même on montre que :

$$b_{2t}^2 Q_{2t} \longrightarrow \frac{1}{\tau}.$$

Ainsi, $b_{1t}^2 Q_{1t}$ et $b_{2t}^2 Q_{2t}$ admettent des limites finies. Et comme les indicatrices tendent vers 0 (leurs limites sont des indicatrices sur l'ensemble vide). Alors,

$$\lim_{t\longrightarrow\infty} \int_{|b_t|>\varepsilon\sqrt{T}} b_t^2 \mathrm{d}F_{B_t}(b_t) = 0.$$

Donc, d'après le lemme de Cesàro :

$$\lim_{T\longrightarrow\infty} \frac{1}{T}\sum_{t=2}^{T} \int_{|b_t|>\varepsilon\sqrt{T}} b_t^2 \mathrm{d}F_{B_t}(b_t) = 0.$$

□

Chapitre 4

Modèle de Yang-Nevzorov.

Dans ce chapitre, notre but est d'explorer le modèle de Yang-Nevzorov et sa relation avec le modèle LDM du Chapitre 2. Ce modèle est intéressant car il a la structure d'un modèle à risque proportionnel en analyse de survie, lequel a montré son utilité afin de modéliser de nombreux jeux de données. Nous estimons le paramètre de puissance γ d'un modèle de Yang-Nevzorov dans le cas d'une distribution sous-jacente quelconque de paramètres connus. Nous utilisons, comme dans le chapitre précédent, plusieurs méthodes d'estimations basées sur la suite des indicatrices de records δ_t et le nombre de records N_T. Nous étudions le comportement asymptotique de nos estimateurs. Enfin, nous validons nos résultats théoriques par des simulations numériques sous R en analysant la qualité de l'estimateur selon plusieurs critères : le biais, l'écart-type et la probabilité de couverture de l'intervalle de confiance qui en découle.

Avant de développer les différentes méthodes d'estimation, nous présentons quelques résultats préliminaires qui sont utilisés afin de calculer les estimateurs.

4.1 Résultats préliminaires

Dans le modèle introduit par Yang (1975), un nombre entier fixe ρ_t de *va iid* Y de fonction de répartition $F(\cdot)$ est généré et disponible simultanément au temps t, desquelles est extrait $X_t = \max\{\text{de ces } \rho_t Y\}$. Ainsi, nous considérons la suite $\{X_t, t \geq 1\}$ de *va* indépendantes mais non identiquement distribuées avec

$$F_{X_t}(x) = F_t(x) = \{F(x)\}^{\rho_t}, \ \rho_t > 0. \tag{4.1.1}$$

Il nous est donné d'observer seulement les records R_n de cette suite ainsi que les L_n.

En raison de la propriété *iid* des Y, la probabilité qu'il y ait un record parmi les ρ_t *va* nouvellement générées, est donnée par

$$P_t = \mathbb{P}[\delta_t = 1] = \frac{\rho_t}{S_t}, \ t \geq 1, \tag{4.1.2}$$

où $S_t = \sum_{k=1}^{t} \rho_k$.

Nevzorov (1990) montre que de façon plus générale (4.1.2) tient si (4.1.1) est vrai avec $\rho_t > 0$ réel. De plus, il montre (Nevzorov 2001, page 114) que l'indépendance des indicatrices de records $\{\delta_t, t \geq 1\}$ reste valable pour n'importe quelle distribution sous-jacente, contrairement au modèle LDM. Donc, la suite des δ_t est un processus de Bernoulli avec probabilité de succès P_t. Dans le cas où on peut définir un taux de record asymptotique $P = \lim_{t\to\infty} P_t$, on dit qu'on a un modèle de Nevzorov de type 1. Notons que si $\rho_t = 1 \ \forall t \geq 1$, le modèle de Yang-Nevzorov n'est autre que le modèle classique du Chapitre 1 où les X_t sont *iid*.

Nous rappelons que la distribution du nombre de records N_T est liée aux indicatrices de records par la relation suivante :

$$N_T = \sum_{t=1}^{T} \delta_t.$$

Ainsi, les moments de N_T sont :

$$\mathbb{E}[N_T] \ = \ \sum_{t=1}^{T} P_t, \tag{4.1.3}$$

et

$$\mathbb{V}[N_T] = \mathbb{E}[N_T] - \sum_{t=1}^{T} P_t^2. \tag{4.1.4}$$

Théorème 4.1. *Ballerini et Resnick (1987), dans le contexte d'un modèle de Yang-Nevzorov, s'il existe une constante* P $(0 < P \leq 1)$ *telle que*

$$\lim_{T \longrightarrow \infty} T^{-1/2} \sum_{t=1}^{T} (P_t - P) = 0,$$

alors,

$$\frac{N_T}{T} \longrightarrow P \text{ presque sûrement,}$$

et

$$\sqrt{T}\left(\frac{N_T}{T} - P\right) \longrightarrow N\left(0, P - P^2\right).$$

4.2 Modèle de Yang-Nevzorov vs Modèle LDM

Proposition 4.2. *Un modèle de Yang-Nevzorov, ayant une distribution sous-jacente* $Y \sim G(0,1)$ *de fonction de répartition* $F_Y(y) = \exp(-\exp(-y))$ *et* $\rho_t = \exp(\theta t)$ *est équivalent à un modèle LDM avec la même distribution sous-jacente et un drift* θ*.*

Démonstration. Pour un modèle de Yang-Nevzorov de distribution sous-jacente $G(0,1)$:

$$\begin{aligned} F_{X_t}(x) &= \{F(x)\}^{\rho_t}, \\ &= [\exp(-\exp(-x))]^{\rho_t}, \\ &= \exp(-\exp(-x) \times \rho_t), \\ &= \exp(-\exp(-x) \times \exp(\log(\rho_t))), \\ &= \exp(-\exp(-(x - \log(\rho_t)))), \\ &= F_Y(x - \log(\rho_t)). \end{aligned}$$

Ainsi,

$$X_t = Y_t + \log(\rho_t). \tag{4.2.1}$$

Si $\rho_t = \exp(\theta t)$, où θ est une constante positive, on retombe sur le modèle LDM de distribution sous-jacente de Gumbel :

$$X_t = Y_t + \theta t.$$

□

Dans un modèle LDM, $\mathbb{V}[X_t]$ est constante. Ce n'est pas nécessairement le cas dans un modèle de Yang-Nevzorov.

Proposition 4.3. *Dans un modèle de Yang-Nevzorov la variance des X_t dépend de t.*

Démonstration. Considérons un modèle de Yang-Nevzorov où les $Y \sim U(0,1)$. Pour $x \in (0,1)$

$$\begin{aligned} F_t(x) &= \{F(x)\}^{\rho_t}, \\ &= x^{\rho_t}, \end{aligned}$$

et

$$f_t(x) = \rho_t x^{\rho_t - 1}.$$

Ainsi,

$$\begin{aligned} \mathbb{E}[X_t] &= \int_0^1 \rho_t x^{\rho_t} \, \mathrm{d}x, \\ &= \frac{\rho_t}{\rho_t + 1}, \end{aligned}$$

et,

$$\begin{aligned} \mathbb{E}\left[X_t^2\right] &= \int_0^1 \rho_t x^{\rho_t + 1} \, \mathrm{d}x, \\ &= \frac{\rho_t}{\rho_t + 2}. \end{aligned}$$

Par suite,

$$\begin{aligned} \mathbb{V}[X_t] &= \frac{\rho_t}{\rho_t + 2} - \left(\frac{\rho_t}{\rho_t + 1}\right)^2, \\ &= \frac{\rho_t}{(\rho_t + 2)(\rho_t + 1)^2}, \end{aligned}$$

qui est fonction de t. □

4.3 Modèle à croissance exponentielle - Modèle de Yang

Dans cette section nous choisissons la forme paramétrique suivante

$$\rho_t(\gamma) = \gamma^t, \text{ avec } \gamma > 1,$$

appelée modèle de Yang (1975) et qui est antérieur au modèle de Nevzorov (1990). $\rho_t(\gamma)$ représente une croissance exponentielle du nombre des *va* disponibles. Ainsi,

$$\begin{aligned} S_t(\gamma) &= \sum_{t=1}^{T} \gamma^t, \\ &= \gamma \frac{\gamma^t - 1}{\gamma - 1}. \end{aligned}$$

Par suite,

$$\begin{aligned} P_t(\gamma) &= \frac{\rho_t(\gamma)}{S_t(\gamma)}, \\ &= \frac{\gamma^t(\gamma - 1)}{\gamma(\gamma^t - 1)}. \end{aligned}$$

Ainsi, le modèle de Yang est un modèle de Nevzorov de type 1 car :

$$\begin{aligned} \lim_{t\to\infty} P_t(\gamma) &= \frac{\gamma - 1}{\gamma}, \\ &= P(\gamma). \end{aligned}$$

De plus,

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (P_t(\gamma) - P(\gamma)) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\frac{\gamma^t(\gamma - 1)}{\gamma(\gamma^t - 1)} - \frac{\gamma - 1}{\gamma} \right), \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{(\gamma - 1)}{\gamma(\gamma^t - 1)}, \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sqrt{t} \frac{(\gamma - 1)}{\gamma(\gamma^t - 1)\sqrt{t}}. \end{aligned}$$

La suite $b_t = \sqrt{t}$; $t \geqslant 1$ est une suite croissante de réels positifs divergeant vers l'infini. Considérons la série de terme général positif

$$U_t = \frac{(\gamma - 1)}{\gamma\left(\gamma^t - 1\right)\sqrt{t}},\ t \geq 1.$$

On a :

$$\begin{aligned}
\frac{U_{t+1}}{U_t} &= \frac{(\gamma - 1)}{\gamma\left(\gamma^{t+1} - 1\right)\sqrt{t+1}} \times \frac{\gamma\left(\gamma^t - 1\right)\sqrt{t}}{(\gamma - 1)}, \\
&= \frac{\left(1 - \frac{1}{\gamma^t}\right)\gamma^t\sqrt{t}}{\left(\gamma - \frac{1}{\gamma^t}\right)\gamma^t\sqrt{t}\sqrt{1 + \frac{1}{t}}}, \\
\longrightarrow\ & \frac{1}{\gamma} < 1.
\end{aligned}$$

Ainsi, par le critère de convergence de D'Alembert, la série $\sum_{t\geq 1} U_t$ est convergente. Alors, en appliquant le Lemme de Kronecker :

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(P_t\left(\gamma\right) - P\left(\gamma\right)\right) &= \frac{1}{b_T}\sum_{t=1}^{T} b_t U_t, \\
&\longrightarrow\ 0.
\end{aligned}$$

Ainsi, dans un modèle de Yang le Théorème 4.1 est applicable avec $P = \frac{\gamma - 1}{\gamma}$.

Pour illustrer un jeu de données provenant d'un modèle de Yang, la Figure 4.1 présente une trajectoire simulée avec $T = 50$, une distribution sous-jacente de loi de Weibull $W(\mu, \beta, \sigma)$ avec un paramètre de position $\mu = 0$, paramètre d'échelle $\beta = 1$ et paramètre de forme $\sigma = 1$ et $\gamma = 1.25$ et 1 respectivement ($\gamma = 1$ représente le cas *iid*). On remarque que, comme dans le modèle LDM, le nombre de records dans un modèle de Yang augmente plus rapidement que le cas *iid* où les records se concentrent parmi les premières observations. Donc, dans certain cas le passage à un modèle de Yang (γ>1) permet de mieux coller à des données.

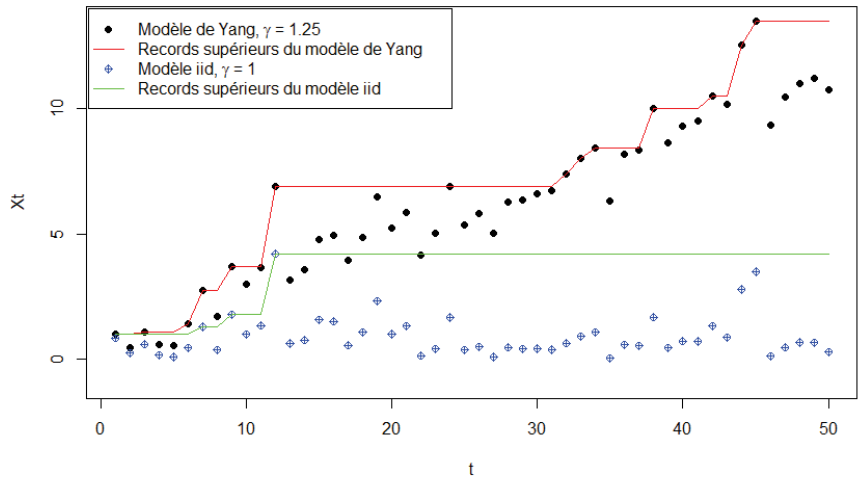

FIGURE 4.1 – Trajectoire simulée d'un modèle à croissance exponentielle où $\gamma = 1.25, 1$ et $T = 50$ avec une distribution sous-jacente de Weibull

4.4 Estimation de γ dans un modèle de Yang :

4.4.1 En utilisant N_T : Estimation par maximum de vraisemblance (EMV)

Nous estimons ponctuellement et par intervalle de confiance le paramètre γ d'un modèle de Yang, en se basant sur le principe du maximum de vraisemblance (EMV) et en utilisant la distribution de probabilité du nombre de records N_T. Nous supposons que la longueur T de la série temporelles est fixée.

Ayant observé $N_T = m$, notre premier travail consiste à trouver $\hat{\gamma}_1$ qui maximise par rapport à γ la fonction de masse de N_T donnée en Khraibani (2008, page 48) :

$$\mathbb{P}_\gamma [N_T = m] = \frac{\mathcal{S}(T, m \mid \vec{u}(\gamma))}{\gamma^T \prod_{t=1}^{T} u_t(\gamma)}, \tag{4.4.1}$$

où le vecteur $\vec{u}(\gamma) = (u_0(\gamma), \ldots, u_T(\gamma))$ est défini par

$N_T = m$	1	2	3	4	5	...	9	10
$\hat{\gamma}_1$	$\simeq 1$	$\simeq 1$	$\simeq 1$	1.291	1.631	...	8.885	∞

TABLE 4.1 – Valeurs de $\hat{\gamma}_1$ pour toutes les valeurs du N_T, quand $T = 10$

$$u_t(\gamma) = \frac{1-\gamma^t}{\gamma^t(1-\gamma)}, \text{ avec } u_0(\gamma) = 0,$$

et $\mathcal{S}(T, m \mid \vec{u}(\gamma))$ est la généralisation du nombre de Stirling de 1^{er} espèce, définie par la comparaison des deux polynômes suivants :

$$\prod_{t=0}^{T-1}(s + u_t(\gamma)) = \sum_{m=0}^{T}\mathcal{S}(T, m \mid \vec{u}(\gamma))\, s^m.$$

A titre d'exemple, appliquons cette méthode d'estimation à un modèle de Yang où $T = 10$. Pour chaque valeur possible de m allant de 1 à T, on calcule numériquement le γ qui maximise la fonction de masse (4.4.1) de N_T. Ceci donne la Table 4.1 suivante qui donne la valeur de $\hat{\gamma}_1$, pour chaque valeur m possible.

Cette table montre qu'on ne peut pas dire grand chose sur le comportement de cet estimateur. En effet, pour certaines valeurs de m, $\hat{\gamma}_1 = \infty$. Donc, on ne peut même pas calculer les moments de l'estimateur.

Mais, en adaptant le Théorème 3.1, on peut néanmoins obtenir un intervalle de confiance $(\gamma_{\mathcal{L}}, \gamma_{\mathcal{U}})$ de niveau exact pour γ. A titre d'exemple, considérons un modèle de Yang avec $T = 100$, $\alpha = 5\%$ et supposons qu'on a observé $m = 44$. L'application du théorème 3.1 donne les valeurs : $\gamma_{\mathcal{L}} = 1.489$ et $\gamma_{\mathcal{U}} = 2.151$. Notons qu'en principe on peut faire ce travail a priori pour toutes les valeurs de T, α et m et créer une table à trois entrées (ou un programme informatique) à laquelle on peut se référer en toutes circonstances. Mais comme la fonction de masse (4.4.1) de N_T est compliquée à manipuler, nous n'irons donc pas plus loin dans l'étude du comportement de cet estimateur.

4.4.2 En utilisant N_T : Estimation par la méthode des moments

Ici notre but est d'obtenir un estimateur ponctuel du paramètre γ du modèle de Yang en se basant sur la distribution du nombre de records N_T et

en appliquant une variante de la méthode des moments. On rappelle que :

$$N_T = \sum_{t=1}^{T} \delta_t,$$

où les δ_t sont indépendants et de loi $Bernoulli\left(P_t\left(\gamma\right)\right)$.

À la Section 4.3 on a montré que le taux de record asymptotique existe :

$$\lim_{t\to\infty} P_t\left(\gamma\right) \;=\; 1 - \frac{1}{\gamma}.$$

Par ailleurs, d'après le Théorème 4.1,

$$\frac{N_T}{T} \longrightarrow P\left(\gamma\right) \text{ presque sûrement,}$$

et

$$\sqrt{T}\left(\frac{N_T}{T} - P\left(\gamma\right)\right) \longrightarrow N\left(0, P\left(\gamma\right) - P^2\left(\gamma\right)\right).$$

Ainsi on peut définir un second estimateur $\hat{\gamma}_2$ de γ par l'expression :

$$\begin{aligned} \hat{\gamma}_2 &= P^{-1}\left(\frac{N_T}{T}\right), \\ &= \frac{1}{1 - \frac{N_T}{T}}. \end{aligned}$$

En utilisant la méthode delta on trouve que :

$$\sqrt{T}\left(P^{-1}\left(\frac{N_T}{T}\right) - P^{-1}\left(P\left(\gamma\right)\right)\right) \xrightarrow{\mathcal{L}} N\left(0, \lambda\left(\gamma\right)\right).$$

où $\lambda\left(\gamma\right)$=$\left(P\left(\gamma\right) - P^2\left(\gamma\right)\right)\left[\left.\frac{\mathrm{d}P^{-1}(\gamma)}{\mathrm{d}\gamma}\right|_{\gamma=P(\gamma)}\right]^2$=$\gamma^2\left(\gamma - 1\right)$. Alors,

$$\sqrt{T}\left(\hat{\gamma}_2 - \gamma\right) \xrightarrow{\mathcal{L}} N\left(0, \lambda\left(\gamma\right)\right). \tag{4.4.2}$$

Ainsi

$$\frac{\sqrt{T}\left(\hat{\gamma}_2 - \gamma\right)}{\sqrt{\lambda\left(\gamma\right)}} \xrightarrow{\mathcal{L}} N\left(0, 1\right), \tag{4.4.3}$$

et un intervalle de confiance de niveau asymptotique $1-\alpha$ pour γ et basé sur N_T est donné par :

$$\left[\hat{\gamma}_2 - \frac{\sqrt{\lambda\left(\hat{\gamma}_2\right)}}{\sqrt{T}} z_{1-\alpha/2}, \hat{\gamma}_2 + \frac{\sqrt{\lambda\left(\hat{\gamma}_2\right)}}{\sqrt{T}} z_{1-\alpha/2}\right]. \tag{4.4.4}$$

4.4.3 $\hat{\gamma}_2$ amélioré

Dans le contexte de la Sous-Section 4.4.2, nous cherchons à améliorer la qualité de $\hat{\gamma}_2$, surtout au niveau du biais. Notons que :

$$\begin{aligned} \mathbb{E}\left[\frac{N_T}{T}\right] &= \frac{1}{T}\sum_{t=1}^{T} \frac{\gamma^t(\gamma-1)}{\gamma\left(\gamma^t-1\right)}, \\ &= h_T(\gamma). \end{aligned}$$

D'après le Théorème 4.1

$$h_T(\gamma) \longrightarrow h(\gamma) = P(\gamma) = 1 - \frac{1}{\gamma},$$

et notre estimateur $\hat{\gamma}_2$ est défini par :

$$\begin{aligned} \hat{\gamma}_2 &= h^{-1}\left(\frac{N_T}{T}\right), \\ &= \frac{1}{1-\frac{N_T}{T}}. \end{aligned}$$

La version classique de la méthode des moments amènerait à définir $\hat{\gamma}_2^M = h_T^{-1}(N_T/T)$. Mais comme le calcul direct de la fonction réciproque de h_T est compliqué, on va utiliser une approche alternative qui consiste à corriger le biais de $\hat{\gamma}_2$ en utilisant le développement de Taylor de h^{-1} au voisinage de $h(\gamma)$ qui est une approximation de $\mathbb{E}\left[\frac{N_T}{T}\right]$. Ce qui fait qu'on peut voir cette méthode d'estimation comme une variante de la méthode des moments :

$$\begin{aligned} h^{-1}\left(\frac{N_T}{T}\right) &= h^{-1}(h(\gamma)) + \left(\frac{N_T}{T} - h(\gamma)\right) \times \left[\left.\frac{\mathrm{d}h^{-1}(\gamma)}{\mathrm{d}\gamma}\right|_{\gamma=h(\gamma)}\right] \\ &\quad + \frac{1}{2}\left(\frac{N_T}{T} - h(\gamma)\right)^2 \times \left[\left.\frac{\mathrm{d}^2 h^{-1}(\gamma)}{\mathrm{d}\gamma^2}\right|_{\gamma=h(\gamma)}\right] + o\left(\left(\frac{N_T}{T} - h(\gamma)\right)^2\right), \end{aligned}$$

$$\begin{aligned}\hat{\gamma}_2 &= \gamma + \left(\frac{N_T}{T} - h(\gamma)\right) \times \frac{1}{(1-h(\gamma))^2} + \left(\frac{N_T}{T} - h(\gamma)\right)^2 \times \frac{1}{(1-h(\gamma))^3} \\ &\quad + o\left(\left(\frac{N_T}{T} - h(\gamma)\right)^2\right).\end{aligned}$$

Ainsi,

$$\mathbb{E}[\hat{\gamma}_2] \simeq \gamma + \frac{h_T(\gamma) - h(\gamma)}{(1-h(\gamma))^2} + \frac{1}{(1-h(\gamma))^3}\mathbb{E}\left[\left(\frac{N_T}{T} - h(\gamma)\right)^2\right]. \quad (4.4.5)$$

Or,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{N_T}{T} - h(\gamma)\right)^2\right] &= \mathbb{E}\left[\left(\frac{N_T}{T}\right)^2 + h^2(\gamma) - 2\frac{N_T}{T}h(\gamma)\right], \\ &= \mathbb{E}\left[\left(\frac{N_T}{T}\right)^2\right] + h^2(\gamma) - 2h_T(\gamma)h(\gamma), \\ &= \mathbb{V}\left[\frac{N_T}{T}\right] + h_T^2(\gamma) + h^2(\gamma) - 2h_T(\gamma)h(\gamma), \\ &= \mathbb{V}\left[\frac{N_T}{T}\right] + (h_T(\gamma) - h(\gamma))^2.\end{aligned}$$

Aussi,

$$\begin{aligned}\mathbb{V}\left[\frac{N_T}{T}\right] &= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{V}[\delta_t], \\ &= \frac{1}{T^2}\sum_{t=1}^{T}P_t(\gamma)(1-P_t(\gamma)),\end{aligned}$$

$$\begin{aligned}\mathbb{V}\left[\frac{N_T}{T}\right] &= \frac{1}{T^2}\left(\mathbb{E}[N_T] - \sum_{t=1}^{T}P_t^2(\gamma)\right), \\ &= \frac{1}{T}h_T(\gamma) - \frac{1}{T^2}\sum_{t=1}^{T}P_t^2(\gamma).\end{aligned}$$

Alors,

$$\mathbb{E}\left[\left(\frac{N_T}{T}-h\left(\gamma\right)\right)^2\right]=\frac{1}{T}h_T\left(\gamma\right)-\frac{1}{T^2}\sum_{t=1}^{T}P_t^2\left(\gamma\right)+\left(h_T\left(\gamma\right)-h\left(\gamma\right)\right)^2. \quad (4.4.6)$$

Par suite, en remplaçant (4.4.6) dans (4.4.5) on obtient,

$$\mathbb{E}\left[\hat{\gamma}_2\right] \simeq \gamma+H_T\left(\gamma\right),$$

où,

$$H_T\left(\gamma\right)=\frac{h_T\left(\gamma\right)-h\left(\gamma\right)}{\left(1-h\left(\gamma\right)\right)^2}+\frac{1}{\left(1-h\left(\gamma\right)\right)^3}\left[\frac{1}{T}h_T\left(\gamma\right)-\frac{1}{T^2}\sum_{t=1}^{T}P_t^2\left(\gamma\right)+\left(h_T\left(\gamma\right)-h\left(\gamma\right)\right)^2\right]. \quad (4.4.7)$$

Donc,

$$\mathbb{E}\left[\hat{\gamma}_2\right]-\gamma \simeq H_T\left(\gamma\right).$$

Par ailleurs,

$$\begin{aligned}\sqrt{T}\left(h_T\left(\gamma\right)-h\left(\gamma\right)\right) &= \sqrt{T}\left(\mathbb{E}\left[\frac{N_T}{T}\right]-P\left(\gamma\right)\right),\\ &= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}P_t\left(\gamma\right)-P\left(\gamma\right)\right),\\ &= \sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}\left(P_t\left(\gamma\right)-P\left(\gamma\right)\right)\right),\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(P_t\left(\gamma\right)-P\left(\gamma\right)\right).\end{aligned}$$

Or dans la Section 4.3 nous avons montré que

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(P_t\left(\gamma\right)-P\left(\gamma\right)\right)\longrightarrow 0.$$

Ainsi,

$$\begin{aligned}\sqrt{T}\left(h_T(\gamma)-h(\gamma)\right) &= o(1),\\ \left(h_T(\gamma)-h(\gamma)\right) &= o\left(T^{-1/2}\right),\\ \left(h_T(\gamma)-h(\gamma)\right)^2 &= o\left(T^{-1}\right).\end{aligned}$$

Alors, on peut corriger le biais de $\hat{\gamma}_2$ à l'ordre T^{-1} en considérant un nouvel estimateur :

$$\begin{aligned}\hat{\gamma}_2^* &= \hat{\gamma}_2 - H_T(\hat{\gamma}_2),\\ &= K_T(\widehat{\gamma}_2),\end{aligned}$$

avec $K_T(\gamma)=\gamma-H_T(\gamma)$. Ainsi, d'après la méthode delta appliqué à (4.4.2) on a :

$$\sqrt{T}\left(K_T(\hat{\gamma}_2)-K_T(\gamma)\right)\xrightarrow{\mathcal{L}}N(0,\vartheta(\gamma)),$$

où $\vartheta(\gamma)=\lambda(\gamma)\left(\frac{\mathrm{d}K_T(\gamma)}{\mathrm{d}\gamma}\right)^2$ et

$$\begin{aligned}\frac{\mathrm{d}K_T(\gamma)}{\mathrm{d}\gamma} &= 1-\frac{\mathrm{d}H_T(\gamma)}{\mathrm{d}\gamma},\\ &= 1-\frac{2\frac{\mathrm{d}P(\gamma)}{\mathrm{d}\gamma}}{(1-P(\gamma))^3}\left[\frac{1}{T}\sum_{t=1}^{T}P_t(\gamma)-P(\gamma)\right]\\ &\quad+\frac{1}{(1-P(\gamma))^2}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{d}P_t(\gamma)}{\mathrm{d}\gamma}-\frac{\mathrm{d}P(\gamma)}{\mathrm{d}\gamma}\right]\\ &\quad+\frac{3\frac{\mathrm{d}P(\gamma)}{\mathrm{d}\gamma}}{(1-P(\gamma))^4}\left[\frac{1}{T^2}\sum_{t=1}^{T}P_t(\gamma)(1-P_t(\gamma))+\left(\frac{1}{T}\sum_{t=1}^{T}P_t(\gamma)-P(\gamma)\right)^2\right]\\ &\quad+\frac{1}{(1-P(\gamma))^3}\left[\frac{1}{T^2}\sum_{t=1}^{T}\frac{\mathrm{d}P_t(\gamma)}{\mathrm{d}\gamma}-\frac{1}{T^2}\sum_{t=1}^{T}\frac{\mathrm{d}P_t^2(\gamma)}{\mathrm{d}\gamma}\right.\\ &\quad\left.+2\left(\frac{1}{T}\sum_{t=1}^{T}P_t(\gamma)-P(\gamma)\right)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{d}P_t(\gamma)}{\mathrm{d}\gamma}-\frac{\mathrm{d}P(\gamma)}{\mathrm{d}\gamma}\right)\right],\end{aligned}$$

avec, $P_t(\gamma)=\frac{\gamma^t-\gamma^{t-1}}{\gamma^t-1}$ et $P(\gamma)=1-\frac{1}{\gamma}$. Donc,

$$\sqrt{T}\left(\hat{\gamma}_2^* - K_T(\gamma)\right) \xrightarrow{\mathcal{L}} N(0, \vartheta(\gamma)). \tag{4.4.8}$$

De plus, afin de monter que $K_T(\gamma) \longrightarrow \gamma$, il suffit de prouver que $H_T(\gamma) \longrightarrow 0$. Or, en appliquant le fait que $\mathbb{E}\left[\frac{N_T}{T}\right] \longrightarrow P(\gamma)$ sur l'équation (4.4.7) il nous reste à montrer que $\frac{1}{T^2}\sum_{t=1}^{T} P_t^2(\gamma) \longrightarrow 0$. Mais, $P_t^2(\gamma) \leq 1$ alors,

$$0 \leq \frac{1}{T^2}\sum_{t=1}^{T} P_t^2(\gamma) \leq \frac{1}{T} \longrightarrow 0.$$

Par suite,

$$\frac{1}{T^2}\sum_{t=1}^{T} P_t^2(\gamma) \longrightarrow 0.$$

Par conséquent, $K_T(\gamma) \longrightarrow \gamma$ et

$$\sqrt{T}\left(\hat{\gamma}_2^* - \gamma\right) \xrightarrow{\mathcal{L}} N(0, \vartheta(\gamma)). \tag{4.4.9}$$

Ainsi, un intervalle de confiance de niveau asymptotique $1-\alpha$ pour γ est donné par :

$$\left[\hat{\gamma}_2^* - \frac{\sqrt{\vartheta\left(\hat{\gamma}_2^*\right)}}{\sqrt{T}} z_{1-\alpha/2}, \hat{\gamma}_2^* + \frac{\sqrt{\vartheta\left(\hat{\gamma}_2^*\right)}}{\sqrt{T}} z_{1-\alpha/2}\right]. \tag{4.4.10}$$

4.4.4 En utilisant $\{\delta_t, t \geq 1\}$: Estimation par maximum de vraisemblance

Ici, notre but est d'estimer le paramètre γ d'un modèle de Yang, en se basant sur le principe du maximum de vraisemblance et en utilisant la distribution de probabilité des indicatrices de records $\{\delta_t, 1 \leq t \leq T\}$. Les δ_t sont indépendants et suivent la loi de Bernoulli de paramètre $P_t(\gamma) = \frac{\gamma^t(\gamma-1)}{\gamma(\gamma^t-1)}$. Reparamétrisons $\upsilon = \frac{1}{\gamma}$, $0 < \upsilon < 1$ et $Q_t(\upsilon) = \frac{1-\upsilon}{1-\upsilon^t}$. Notre travail consiste alors à trouver le υ qui maximise l'expression :

$$\begin{aligned} L(\upsilon) &= \mathbb{P}[\delta_1, \ldots, \delta_T; \upsilon], \\ &= \prod_{t=2}^{T} Q_t(\upsilon)^{\delta_t} \left(1 - Q_t(\upsilon)\right)^{1-\delta_t}. \end{aligned} \tag{4.4.11}$$

Pour ce faire, une approche consiste à dériver $\log L(\upsilon)$ par rapport à υ puis à calculer les racines de cette dérivée :

$$\log L(\upsilon) = N_T \log(1-\upsilon) + (T - N_T)\log(\upsilon) - \log\left(1-\upsilon^T\right) - \sum_{t=2}^{T} \delta_t \log\left(1-\upsilon^{t-1}\right), \tag{4.4.12}$$

avec $\sum_{t=2}^{T} \delta_t = N_T - 1$. Ainsi, il faut trouver la valeur $\hat{\upsilon}$ de υ, laquelle dénote notre estimateur par la méthode du maximum de vraisemblance, telle que :

$$\left(\frac{\mathrm{d}\log L(\upsilon)}{\mathrm{d}\upsilon}\right)_{\upsilon=\hat{\upsilon}} = 0.$$

Or,

$$\frac{\mathrm{d}\log L(\upsilon)}{\mathrm{d}\upsilon} = \frac{-N_T}{1-\upsilon} + \frac{T-N_T}{\upsilon} + \frac{T\upsilon^{T-1}}{1-\upsilon^T} + \sum_{t=2}^{T} \delta_t \frac{(t-1)\,\upsilon^{t-2}}{1-\upsilon^{t-1}}. \tag{4.4.13}$$

Les δ_t sont indépendants mais pas identiquement distribués de sorte que les théorèmes standards concernant le comportement des estimateurs de vraisemblance maximale ne s'appliquent pas. Cependant, en appliquant à notre contexte un théorème de Leroy *et al.* (2016) sur le comportement asymptotique des EMV dans le cas où les observations sont indépendantes mais non identiquement distribuées, on montre (même démarche que dans la Sous-Section 3.3.2 du Chapitre 3 car l'équation (4.4.13) est la même que (3.3.3) à un changement de variable près) que $\hat{\upsilon}$ est consistant et que

$$\frac{\hat{\upsilon}-\upsilon}{\sqrt{I_T^{-1}(\upsilon)}} \xrightarrow{\mathcal{L}} N(0,1), \tag{4.4.14}$$

où $I_T(\upsilon)$ dénote l'information de Fisher (Pour la vérification des conditions du Théorème de Leroy *et al.* (2016) voir la Section 3.7 du Chapitre 3) :

$$I_T(\upsilon) = -\mathbb{E}\left[\frac{\mathrm{d}^2\log L(\upsilon)}{\mathrm{d}\upsilon^2}\right].$$

Or,

$$\begin{aligned}\frac{\mathrm{d}^2 \log L(\upsilon)}{\mathrm{d}\upsilon^2} &= \frac{-N_T}{(1-\upsilon)^2} + \frac{N_T - T}{\upsilon^2} + \frac{T\upsilon^{T-2}\left(T+\upsilon^T-1\right)}{\left(1-\upsilon^T\right)^2} \\ &\quad + \sum_{t=2}^{T} \delta_t \frac{(t-1)\,\upsilon^{t-3}\left(t-2+\upsilon^{t-1}\right)}{\left(1-\upsilon^{t-1}\right)^2}.\end{aligned}$$

Ainsi,

$$\begin{aligned}I_T(\upsilon) &= \frac{1}{(1-\upsilon)^2}\sum_{t=1}^{T} Q_t(\upsilon) + \frac{1}{\upsilon^2}\left(T - \sum_{t=1}^{T} Q_t(\upsilon)\right) - \frac{T\upsilon^{T-2}\left(T+\upsilon^T-1\right)}{\left(1-\upsilon^T\right)^2} \\ &\quad - \sum_{t=2}^{T} \frac{(t-1)\,\upsilon^{t-3}\left(t-2+\upsilon^{t-1}\right)}{\left(1-\upsilon^{t-1}\right)^2} Q_t(\upsilon). \qquad (4.4.15)\end{aligned}$$

Notons que lorsque $N_T = T$, le maximum de la fonction de vraisemblance (4.4.11) n'est pas atteint et dans ce cas $\hat{\upsilon} = 0$ et $\hat{\gamma} = +\infty$. Mais cet événement a une probabilité qui tend vers zéro lorsque T tend vers l'infini.

Par ailleurs, suivant la même preuve qu'à la Sous-Section 3.3.2 :

$$\frac{I_T(\upsilon)}{T} \xrightarrow[T]{} I(\upsilon) = \frac{1}{(1-\upsilon)\,\upsilon}.$$

Par suite, en se basant sur l'équation (4.4.14) on a

$$\sqrt{T}\left(\hat{\upsilon} - \upsilon\right) \xrightarrow{\mathcal{L}} N\left(0, I^{-1}(\upsilon)\right).$$

Maintenant le paramètre γ de notre modèle de Yang est donné par $\gamma = \frac{1}{\upsilon}$. Donc, d'après la méthode delta on obtient :

$$\frac{\sqrt{T}\left(\hat{\gamma}_3 - \gamma\right)}{\sqrt{\left(\frac{\mathrm{d}}{\mathrm{d}\upsilon}\left(\frac{1}{\upsilon}\right)\right)^2 I^{-1}(\upsilon)}} = \frac{\sqrt{T}\left(\hat{\gamma}_3 - \gamma\right)}{\frac{1}{\upsilon^2}\sqrt{I^{-1}(\upsilon)}} \xrightarrow{\mathcal{L}} N(0,1),$$

et, pour un niveau de confiance asymptotique $1-\alpha$, un intervalle de confiance asymptotique de γ est donné par :

$$\left[\hat{\gamma}_3 - \frac{\sqrt{I^{-1}(\widehat{\upsilon})}}{\widehat{\upsilon}^2\sqrt{T}} z_{1-\alpha/2}, \hat{\gamma}_3 + \frac{\sqrt{I^{-1}(\widehat{\upsilon})}}{\widehat{\upsilon}^2\sqrt{T}} z_{1-\alpha/2}\right]. \qquad (4.4.16)$$

Il ne reste plus qu'à réexprimer l'information de Fisher (4.4.15) dans la paramétrisation originale (c'est à dire en γ), en appliquant la méthode delta sur l'équation (4.4.14) on obtient :

$$I_T^{-1}(\gamma) = I_T^{-1}(\upsilon) \times \left(\frac{\mathrm{d}}{\mathrm{d}\upsilon}\left(\frac{1}{\upsilon}\right)\right)^2.$$

Ainsi

$$\begin{aligned} I_T(\gamma) &= \frac{I_T(\upsilon)}{\left(\frac{\mathrm{d}}{\mathrm{d}\upsilon}\left(\frac{1}{\upsilon}\right)\right)^2}, \\ &= \upsilon^4 I_T(\upsilon). \\ &= \frac{1}{\gamma^2(\gamma-1)^2}\sum_{t=1}^{T} P_t(\gamma) + \frac{1}{\gamma^2}\left(T - \sum_{t=1}^{T} P_t(\gamma)\right) - \frac{T\left(1+\gamma^T(T-1)\right)}{\gamma^2\left(\gamma^T-1\right)^2} \\ &\quad - \sum_{t=2}^{T} \frac{(t-1)\left(1+(t-2)\gamma^{t-1}\right)}{\gamma^2\left(\gamma^{t-1}-1\right)^2} P_t(\gamma). \end{aligned}$$

et

$$\frac{\hat{\gamma}_3 - \gamma}{\sqrt{I_T^{-1}(\gamma)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{4.4.17}$$

Alors, pour un niveau de confiance $1-\alpha$, un intervalle de confiance de γ est donné par :

$$\left[\hat{\gamma}_3 - \sqrt{I_T^{-1}(\hat{\gamma}_3)}z_{1-\alpha/2}, \hat{\gamma}_3 + \sqrt{I_T^{-1}(\hat{\gamma}_3)}z_{1-\alpha/2}\right]. \tag{4.4.18}$$

4.5 Simulations numériques

Afin d'étudier le comportement de nos estimateurs, en particulier l'utilité des approximations asymptotiques des équations (4.4.3), (4.4.9) et (4.4.17), des simulations ont été effectuées. Nous nous concentrons sur le biais, l'approximation asymptotique des écarts-types et la probabilité de couverture des intervalles de confiance. Pour ce faire, 5000 séries chronologiques ont été générées selon un modèle de Yang pour différentes valeurs de γ et T et une distribution sous-jacente de loi de Weibull $W(\mu, \beta, \sigma)$ avec $\mu = 0$, $\beta = 1$ et $\sigma = 1$, ces derniers choix n'ayant aucune incidence car les estimateurs sont

indépendants de la loi de la distribution sous-jacente. Pour chacune de ces séries, la suite des indicatrices de records $\{\delta_t, 1 \leq t \leq T\}$ a été extraite et les estimateurs $\hat{\gamma}_2$, $\hat{\gamma}_2^*$ et $\hat{\gamma}_3$ ont été calculés. De ces 5000 séries, les caractéristiques mentionnées précédemment ont été empiriquement estimées.

La Table 4.2 donne les biais empiriques des estimateurs. L'estimateur $\hat{\gamma}_2$ a un grand biais et dans tous les cas $\hat{\gamma}_2^*$ et $\hat{\gamma}_3$ doivent être préférés, avec une légère préférence pour $\hat{\gamma}_2^*$ lorsque γ est grand. La Table 4.3 présente les écarts-types pour $T = 25$ et $T = 50$ observations respectivement. Dans les Sous-Tables 4.3a et 4.3b les colonnes 2 et 3 donnent les écarts-types empiriques (à partir de 5000 séries) de $\hat{\gamma}_2$ et $\hat{\gamma}_2^*$. L'approximation asymptotique de l'équation (4.4.3) apparait dans la quatrième colonne et celle de l'équation (4.4.9) dans la cinquième. Les deux dernières colonnes montrent le même pour $\hat{\gamma}_3$ et son approximation $\sqrt{I_T^{-1}(\gamma)}$ de l'équation (4.4.17). Dans l'ensemble, l'approximation des équations (4.4.3) et (4.4.9) sous-estiment les valeurs exactes des écarts types. L'équation (4.4.17) donne des résultats beaucoup plus précis. Pour ces raisons, il convient de laisser tomber l'estimateur $\hat{\gamma}_2$ et nous gardons $\hat{\gamma}_3$. Aussi, comme $\hat{\gamma}_2^*$ est presque sans biais et a un écart-type très proche de l'expression $\sqrt{I_T^{-1}(\gamma)}$, nous gardons ce dernier estimateur avec $\sqrt{I_T^{-1}(\gamma)}$ comme approximation de l'écart-type.

Enfin, pour vérifier l'exactitude de l'approximation asymptotique des équations (4.4.9) et (4.4.17), les probabilités de couverture des intervalles de confiance (4.4.10) et (4.4.18) (pour les probabilités de couverture basées sur $\hat{\gamma}_2^*$, nous considérons $\sqrt{I_T^{-1}(\gamma)}$ comme approximation de l'écart-type), pour différents niveaux de confiance $1 - \alpha$ et de T, sont présentées respectivement dans les Tables 4.4 et 4.5. En terme de probabilités de couverture l'estimateur $\hat{\gamma}_3$ doit être préféré à $\hat{\gamma}_2^*$. Dans la Table (4.5) les niveaux de confiance réels sont proches de ceux visés, sauf pour les petites valeurs de γ. Ceci est provoqué par l'asymétrie de la distribution de $\hat{\gamma}_3$ qui vient de la contrainte $\gamma > 1$. Sinon, l'utilisation de $\hat{\gamma}_3$ devrait conduire à une inférence correcte pour n'importe quelle distribution sous-jacente.

	Biais					
	$T = 25$			$T = 50$		
γ	$\hat{\gamma}_2$	$\hat{\gamma}_2^*$	$\hat{\gamma}_3$	$\hat{\gamma}_2$	$\hat{\gamma}_2^*$	$\hat{\gamma}_3$
1.05	0.17	0.02	0.02	0.08	0.01	$\simeq 0.00$
1.10	0.16	0.01	0.01	0.07	$\simeq 0.00$	$\simeq 0.00$
1.15	0.15	0.01	0.01	0.07	$\simeq 0.00$	$\simeq 0.00$
1.25	0.15	$\simeq 0.00$	$\simeq 0.00$	0.07	$\simeq 0.00$	$\simeq 0.00$
1.50	0.17	- 0.01	0.03	0.08	$\simeq 0.00$	0.01
1.75	0.19	- 0.02	0.06	0.09	$\simeq 0.00$	0.03

TABLE 4.2 – Biais empiriques de $\hat{\gamma}_2$, $\hat{\gamma}_2^*$ et $\hat{\gamma}_3$ pour différentes valeurs de γ et T

γ	$\sqrt{\mathbb{V}(\hat{\gamma}_2)}$	$\sqrt{\mathbb{V}(\hat{\gamma}_2^*)}$	$\sqrt{\frac{\lambda(\gamma)}{T}}$	$\sqrt{\frac{\vartheta(\gamma)}{T}}$	$\sqrt{\mathbb{V}(\hat{\gamma}_3)}$	$\sqrt{I_T^{-1}(\gamma)}$
1.05	0.10	0.10	0.05	0.06	0.09	0.12
1.10	0.12	0.12	0.07	0.08	0.11	0.12
1.15	0.14	0.13	0.09	0.10	0.13	0.13
1.25	0.17	0.16	0.13	0.13	0.17	0.16
1.50	0.28	0.25	0.21	0.20	0.27	0.24
1.75	0.40	0.34	0.30	0.27	0.40	0.33

(a) $T = 25$ observations

γ	$\sqrt{\mathbb{V}(\hat{\gamma}_2)}$	$\sqrt{\mathbb{V}(\hat{\gamma}_2^*)}$	$\sqrt{\frac{\lambda(\gamma)}{T}}$	$\sqrt{\frac{\vartheta(\gamma)}{T}}$	$\sqrt{\mathbb{V}(\hat{\gamma}_3)}$	$\sqrt{I_T^{-1}(\gamma)}$
1.05	0.05	0.06	0.03	0.04	0.05	0.06
1.10	0.06	0.07	0.05	0.05	0.07	0.07
1.15	0.08	0.08	0.06	0.07	0.08	0.08
1.25	0.10	0.10	0.09	0.09	0.10	0.10
1.50	0.17	0.16	0.15	0.14	0.17	0.16
1.75	0.24	0.23	0.21	0.20	0.24	0.22

(b) $T = 50$ observations

TABLE 4.3 – Écarts-types de $\hat{\gamma}_2$, $\hat{\gamma}_2^*$ et $\hat{\gamma}_3$ pour différentes valeurs de γ et T. Les colonnes 2 et 3 donnent les écarts-types empiriques de $\hat{\gamma}_2$, $\hat{\gamma}_2^*$. L'approximation asymptotique des équations (4.4.3) et (4.4.9) apparaissent dans les colonnes 4 et 5 respectivement. Les deux dernières colonnes montrent le même pour $\hat{\gamma}_3$ et son approximation $\sqrt{I_T^{-1}(\gamma)}$ de l'équation (4.4.17).

	Probabilités de couverture			
	$T = 25$		$T = 50$	
γ	90%	95%	90%	95%
1.05	98.5%	99.9%	98.2%	99.2%
1.10	98.8%	99.9%	98.1%	99.8%
1.15	99.8%	99.9%	91.5%	92.9%
1.25	91.5%	99.9%	87.3%	94.4%
1.50	81.8%	91.8%	89.7%	93.8%
1.75	81.3%	90.5%	86.5%	92.1%

TABLE 4.4 – Probabilités de couverture de l'intervalle de confiance (4.4.10) pour différentes valeurs de γ, T et du niveau de confiance $1 - \alpha$

	Probabilités de couverture			
	$T = 25$		$T = 50$	
γ	90%	95%	90%	95%
1.05	96.7%	97.7%	96.6%	98.0%
1.10	97.1%	98.0%	89.2%	99.3%
1.15	97.6%	98.3%	89.6%	93.7%
1.25	89.1%	92.9%	88.7%	93.2%
1.50	88.1%	91.9%	89.3%	93.8%
1.75	89.0%	92.0%	90.3%	93.8%

TABLE 4.5 – Probabilités de couverture de l'intervalle de confiance (4.4.18) pour différentes valeurs de γ, T et du niveau de confiance $1 - \alpha$

4.6 Conclusion

Nous avons développé quelques aspects du comportement stochastique du modèle de Yang-Nevzorov, qui est, avec le modèle LDM du chapitre précédent, l'un des modèles les plus populaires pour des séries de records dans le cas non *iid*. De plus, en se basant sur ces résultats stochastiques, nous avons introduit plusieurs méthodes d'estimation du paramètre γ d'un modèle de Yang avec une distribution sous-jacente quelconque, en utilisant la méthode des moments ou le principe du maximum de vraisemblance. Il importe de signaler que la contrainte d'une distribution sous-jacente de loi de Gumbel du modèle LDM peut être levée si on considère plutôt un modèle de Yang, lequel est par ailleurs attrayant parce qu'il croise le modèle LDM (cas de la Gumbel) et qu'il est de même structure que le modèle à hasard proportionnel, très populaire en analyse de survie. À l'instar du modèle LDM, ce modèle de Yang devrait donner un ajustement correct à des nombreux jeux de données sans présenter les contraintes du LDM. D'autres avantages du modèle de Yang sont signalés au chapitre suivant.

Ainsi, dans le chapitre suivant, notre but est de construire plusieurs tests d'adéquation pour un modèle de Yang, basés sur le comportement asymptotique de la distribution du temps inter-records.

Chapitre 5

Temps inter-records et tests d'adéquation

Toujours dans le contexte d'un modèle de Yang, nous étudions dans ce chapitre le comportement stochastique du temps inter-records et nous donnons sa loi asymptotique, indépendamment de la loi des *va* sous-jacentes. Puis, en se basant sur ces résultats, combinés aux estimateurs obtenus au chapitre précédent, nous introduisons plusieurs tests d'adéquation d'un modèle de Yang. On remarque que ces tests peuvent aider à suggérer des corrections au modèle. Enfin, nous appliquons nos résultats théoriques à des données analysées précédemment par Yang (1975) présentant les records de la course de 200 mètres dans les Jeux olympiques.

5.1 Distribution du temps inter-records

On rappelle que dans un modèle de Yang, la suite $\{X_t, t \geq 1\}$ de *va* indépendantes mais non identiquement distribuées est telle que

$$F_{X_t}(x) = \{F(x)\}^{\rho_t},$$

où $\rho_t = \gamma^t$ et $\gamma > 1$. Notons :

$$S(t) \;=\; \sum_{k=1}^{t} \rho_t = \gamma + \gamma^2 + \cdots + \gamma^t,$$

et,

$$\triangle_{L_n} = L_{n+1} - L_n,\ n \geq 1,$$

le temps entre le $(n+1)^{ième}$ et le $n^{ième}$ records, sachant que $L_1 = 1$ (record trivial).

En s'inspirant des travaux de Yang (1975) réalisés dans le cadre où les paramètres ρ_t sont discrétisés en prenant $\rho_t = \left\lfloor \gamma^t + \frac{1}{2} \right\rfloor$, $\lfloor x \rfloor$ désignant la partie entière d'un réel x, nous présentons les Théorèmes 5.1 et 5.2 et donnons leurs preuves dans le cas plus général $\rho_t = \gamma^t$. Notez que, le Théorème 5.1 est démontré dans le contexte plus large d'un modèle de Nevzorov, c'est-à-dire pour un $\rho_t > 0$ quelconque.

Comme à la Section 2.5, dans la suite de ce chapitre, on considère que les valeurs des indices de records $l_2, \ldots, l_n, \ldots$ sont des entiers telles que $1 < l_2 < \ldots < l_n < \ldots$. Pour alléger les expressions, on escamote le fait que si les l_j ne sont pas des entiers satisfaisant les contraintes alors les probabilités qu'on rencontre sont nulles.

Théorème 5.1. *Dans le cadre d'un modèle de Nevzorov*

$$\mathbb{P}\left[\Delta_{L_1} > j\right] = \frac{\rho_1}{S\left(j+1\right)},$$

et,

$$\mathbb{P}\left[\Delta_{L_n} > j\right] = \sum_{l_2=2}^{\infty} \sum_{l_3=l_2+1}^{\infty} \cdots \sum_{l_n=l_{n-1}+1}^{\infty} \frac{\rho_{l_1} \times \cdots \times \rho_{l_n}}{S\left(l_2-1\right) \times \cdots \times S\left(l_n-1\right) \times S\left(l_n+j\right)}. \tag{5.1.1}$$

Démonstration. La distribution de Δ_{L_n} ne dépend pas de la loi de la variable aléatoire sous-jacente car elle dépend uniquement des rangs de la suite des *va* $X_1, \ldots, X_T$ (T désigne le temps présent). Sans perte de généralité, et afin de simplifier le calcul, nous considérons le cas où la distribution sous-jacente possède une loi exponentielle de fonction de masse $f\left(x\right) = e^{-x}$, $x > 0$. Ainsi,

$$F_{X_1}\left(x\right) = \left(1 - e^{-x}\right)^{\rho_1} \text{ et } f_{X_1}\left(x\right) = \rho_1 \left(1 - e^{-x}\right)^{\rho_1 - 1} e^{-x}.$$

Par suite,

$$\begin{aligned} \mathbb{P}\left[L_2 > j\right] &= \int_0^{\infty} F_{X_2}\left(x\right) \cdots F_{X_j}\left(x\right) f_{X_1}\left(x\right) \mathrm{d}x, \\ &= \int_0^{\infty} \left(1 - e^{-x}\right)^{\rho_2 + \cdots + \rho_j} f_{X_1}\left(x\right) \mathrm{d}x. \end{aligned}$$

Ainsi,

$$\begin{aligned}\mathbb{P}\left[\Delta_{L_1} > j\right] &= \mathbb{P}\left[L_2 - L_1 > j\right], \\ &= \mathbb{P}\left[L_2 > j+1\right],\end{aligned}$$

$$\begin{aligned}\mathbb{P}\left[\Delta_{L_1} > j\right] &= \int_0^\infty \left(1-e^{-x}\right)^{\rho_2+\cdots+\rho_{j+1}} \rho_1 \left(1-e^{-x}\right)^{\rho_1-1} e^{-x}\,\mathrm{d}x, \\ &= \int_0^\infty \left(1-e^{-x}\right)^{S(j+1)-1} \rho_1 e^{-x}\,\mathrm{d}x, \\ &= \frac{\rho_1}{S\left(j+1\right)}.\end{aligned}$$

D'autre part,

$$\begin{aligned}\mathbb{P}\left[X_{L_2} \leq x, L_2 = l_2\right] &= \mathbb{P}\left[X_{L_2} \leq x \mid L_2 = l_2\right] \times \mathbb{P}\left[L_2 = l_2\right], \\ &= \mathbb{P}\left[X_1 \leq x, X_2 \leq x, \ldots, X_{l_2} \leq x\right] \\ &\quad \times \mathbb{P}\left[X_2 \leq X_1, X_3 \leq X_1, \ldots, X_{l_2-1} \leq X_1, X_{l_2} > X_1\right],\end{aligned}$$

$$\begin{aligned}\mathbb{P}\left[X_{L_2} \leq x, L_2 = l_2\right] &= \left(1-e^{-x}\right)^{\rho_1+\cdots+\rho_{l_2}} \int_0^\infty \Big\{\left(1-e^{-x}\right)^{\rho_2+\cdots+\rho_{l_2-1}} \\ &\quad \times \left(1-\left(1-e^{-x}\right)^{\rho_{l_2}}\right) \rho_1 \left(1-e^{-x}\right)^{\rho_1-1} e^{-x}\Big\}\,\mathrm{d}x, \\ &= \left(1-e^{-x}\right)^{S(l_2)} \rho_1 \left[\frac{1}{S\left(l_2-1\right)} - \frac{1}{S\left(l_2\right)}\right], \\ &= \frac{\rho_1 \rho_{l_2}}{S\left(l_2-1\right) S\left(l_2\right)} \left(1-e^{-x}\right)^{S(l_2)}. \qquad (5.1.2)\end{aligned}$$

Par conséquent,

$$\begin{aligned}\mathbb{P}\left[\Delta_{L_2} > j\right] &= \mathbb{P}\left[L_3 - L_2 > j\right], \\ &= \sum_{l_2=2}^\infty \int_0^\infty \prod_{t=l_2+1}^{l_2+j} \mathbb{P}\left[X_t \leq x\right] \mathrm{d}\mathbb{P}\left(X_{L_2} \leq x, L_2 = l_2\right), \\ &= \sum_{l_2=2}^\infty \frac{\rho_1 \rho_{l_2}}{S\left(l_2-1\right)} \int_0^\infty \left(1-e^{-x}\right)^{\rho_{l_2+1}+\cdots+\rho_{l_2+j}} \left(1-e^{-x}\right)^{S(l_2)-1} e^{-x}\,\mathrm{d}x, \\ &= \sum_{l_2=2}^\infty \frac{\rho_1 \rho_{l_2}}{S\left(l_2-1\right)} \int_0^\infty \left(1-e^{-x}\right)^{S(l_2+j)-1} e^{-x}\,\mathrm{d}x,\end{aligned}$$

$$\mathbb{P}\left[\Delta_{L_2} > j\right] = \sum_{l_2=2}^{\infty} \frac{\rho_1 \rho_{l_2}}{S\left(l_2-1\right) S\left(l_2+j\right)}.$$

De même,

$$\begin{aligned}
\mathbb{P}\left[X_{L_n} \leq x, L_2 = l_2, \ldots, L_n = l_n\right] &= \mathbb{P}\left[X_{l_{n-1}} < X_{l_n} \leq x, X_{l_n-1} \leq X_{l_{n-1}},\right. \\
&\quad X_{l_{n-1}+1} \leq X_{l_{n-1}}, X_{l_{n-1}} > X_{l_{n-2}}, \ldots, \\
&\quad \left. X_{l_2} > X_1, X_{l_2-1} \leq X_1, \ldots, X_2 \leq X_1\right], \\
&= \int_0^x f_{X_{l_n}}\left(x_n\right) \mathrm{d}x_n \\
&\quad \times \int_0^{x_n} \prod_{t=l_{n-1}+1}^{l_n-1} \mathbb{P}\left[X_t \leq x_{n-1}\right] f_{X_{l_{n-1}}}\left(x_{n-1}\right) \mathrm{d}x_{n-1} \\
&\quad \times \int_0^{x_{n-1}} \prod_{t=l_{n-2}+1}^{l_{n-1}-1} \mathbb{P}\left[X_t \leq x_{n-2}\right] f_{X_{l_{n-2}}}\left(x_{n-2}\right) \mathrm{d}x_{n-2} \times \cdots \\
&\quad \times \int_0^{x_3} \prod_{t=l_2+1}^{l_3-1} \mathbb{P}\left[X_t \leq x_2\right] f_{X_{l_2}}\left(x_2\right) \mathrm{d}x_2 \\
&\quad \times \int_0^{x_1} \prod_{t=2}^{l_2-1} \mathbb{P}\left[X_t \leq x\right] f_{X_1}\left(x\right) \mathrm{d}x.
\end{aligned}$$

Après avoir calculé les intégrales et en suivant la même méthode aboutissant à (5.1.2), on obtient :

$$\mathbb{P}\left[X_{L_n} \leq x, L_2 = l_2, \ldots, L_n = l_n\right] = \frac{\rho_{l_1} \rho_{l_2} \cdots \rho_{l_n}}{S\left(l_2-1\right) S\left(l_3-1\right) \ldots S\left(l_n-1\right) S\left(l_n\right)} \left(1-e^{-x}\right)^{S(l_n)}.$$

Ainsi,

$$\begin{aligned}
\mathbb{P}\left[\Delta_{L_n} > j\right] &= \mathbb{P}\left[L_{n+1} - L_n > j\right], \\
&= \sum_{l_n=n}^{\infty} \sum_{l_{n-1}=n-1}^{l_n-1} \cdots \sum_{l_2=2}^{l_3-1} \int_0^{\infty} \prod_{t=l_n+1}^{l_n+j} \mathbb{P}\left[X_t \leq x\right] \mathrm{d}P\left[X_{L_n} \leq x, L_2 = l_2, \ldots, L_n = l_n\right], \\
&= \sum_{l_n=n}^{\infty} \sum_{l_{n-1}=n-1}^{l_n-1} \cdots \sum_{l_2=2}^{l_3-1} \frac{\rho_{l_1} \times \cdots \times \rho_{l_n}}{S\left(l_2-1\right) \times \cdots \times S\left(l_n-1\right) \times S\left(l_n+j\right)}.
\end{aligned}$$

Maintenant, si on permute les indices de sommation on obtient l'équation (5.1.1). □

En s'appuyant sur le Théorème 5.1 on obtient la fonction de masse de Δ_{L_n} :

$$\begin{aligned}
\mathbb{P}\left[\Delta_{L_n}=j\right] &= \mathbb{P}\left[L_{n+1}-L_n=j\right], \\
&= \mathbb{P}\left[\Delta_{L_n}>j-1\right]-\mathbb{P}\left[\Delta_{L_n}>j\right], \\
&= \sum_{l_2=2}^{\infty}\sum_{l_3=l_2+1}^{\infty}\cdots\sum_{l_n=l_{n-1}+1}^{\infty}\frac{\rho_{l_1}\times\cdots\times\rho_{l_n}}{S\left(l_2-1\right)\times\cdots\times S\left(l_n-1\right)\times S\left(l_n+j-1\right)} \\
&\quad -\sum_{l_2=2}^{\infty}\sum_{l_3=l_2+1}^{\infty}\cdots\sum_{l_n=l_{n-1}+1}^{\infty}\frac{\rho_{l_1}\times\cdots\times\rho_{l_n}}{S\left(l_2-1\right)\times\cdots\times S\left(l_n-1\right)\times S\left(l_n+j\right)}, \\
&= \sum_{l_2=2}^{\infty}\sum_{l_3=l_2+1}^{\infty}\cdots\sum_{l_n=l_{n-1}+1}^{\infty}\left\{\frac{\rho_{l_1}\times\cdots\times\rho_{l_n}}{S\left(l_2-1\right)\times\cdots\times S\left(l_n-1\right)}\right. \\
&\quad \left.\times\left[\frac{1}{S\left(l_n+j-1\right)}-\frac{1}{S\left(l_n+j\right)}\right]\right\}.
\end{aligned}$$

Afin d'appliquer les résultats du Théorème 5.1 au modèle de Yang et pour simplifier les notations définissons :

$$\begin{aligned}
\Lambda\left(t\right) &= 1+\gamma+\cdots+\gamma^{t-1}=\frac{1-\gamma^t}{1-\gamma}, \\
p\left(n,0\right) &= 1,\, n\geq 1, \\
p\left(1,j\right) &= \frac{1}{\Lambda\left(j+1\right)},\, j\geq 1, \\
p\left(n,j\right) &= \sum_{l_2=2}^{\infty}\sum_{l_3=l_2+1}^{\infty}\cdots\sum_{l_n=l_{n-1}+1}^{\infty}\left\{\prod_{i=2}^{n}\frac{\gamma^{l_i-1}}{\Lambda\left(l_i-1\right)}\right\}\frac{1}{\Lambda\left(l_n+j\right)}. \quad (5.1.3)
\end{aligned}$$

Ainsi, dans le cadre d'un modèle de Yang :

$$\begin{aligned}
\mathbb{P}\left[\Delta_{L_n}>0\right] &= p\left(n,0\right)=1,\, n\geq 1, \\
\mathbb{P}\left[\Delta_{L_n}>j\right] &= p\left(n,j\right),\, n\geq 1 \,\text{et}\, j\geq 1, \\
\mathbb{P}\left[\Delta_{L_n}=j\right] &= p\left(n,j-1\right)-p\left(n,j\right),\, n\geq 1 \,\text{et}\, j\geq 1.
\end{aligned}$$

Le Théorème suivant est une spécialisation au cas du modèle de Yang.

Théorème 5.2. *Dans le cadre d'un modèle de Yang nous avons la formule récursive suivante :*

$$p\left(n,j\right)=\sum_{i=0}^{j}\frac{p\left(n-1,i\right)}{\Lambda\left(j+1\right)}, \quad (5.1.4)$$

De plus, $\lim_{n\longrightarrow\infty} p(n,j)$ *existe et converge vers* $q_j = \frac{1}{\gamma^j}$, $j \geq 0$.

Démonstration. Considérons la dernière somme de l'équation (5.1.3) :

$$\begin{aligned}
\sum_{l_n=l_{n-1}+1}^{\infty} \frac{\gamma^{l_n-1}}{\Lambda(l_n-1)\Lambda(l_n+j)} &= (1-\gamma)^2 \sum_{l_n=l_{n-1}+1}^{\infty} \frac{\gamma^{l_n-1}}{(1-\gamma^{l_n-1})(1-\gamma^{l_n+j})}, \\
&= (1-\gamma)^2 \sum_{l_n=l_{n-1}+1}^{\infty} \frac{\gamma^{l_n-1}}{\gamma^{l_n-1}-\gamma^{l_n+j}} \left[\frac{1}{1-\gamma^{l_n-1}} - \frac{1}{1-\gamma^{l_n+j}}\right], \\
&= (1-\gamma)^2 \sum_{l_n=l_{n-1}+1}^{\infty} \frac{1}{1-\gamma^{j+1}} \left[\frac{1}{1-\gamma^{l_n-1}} - \frac{1}{1-\gamma^{l_n+j}}\right], \\
&= \frac{(1-\gamma)^2}{1-\gamma^{j+1}} \sum_{l_n=l_{n-1}+1}^{\infty} \left[\frac{1}{1-\gamma^{l_n-1}} - \frac{1}{1-\gamma^{l_n+j}}\right].
\end{aligned}$$

On remarque qu'à partir de $l_n = l_{n-1} + j + 2$, la série devient télescopique. Ainsi, il reste un nombre fini de termes :

$$\begin{aligned}
\sum_{l_n=l_{n-1}+1}^{\infty} \frac{\gamma^{l_n-1}}{\Lambda(l_n-1)\Lambda(l_n+j)} &= \frac{(1-\gamma)^2}{1-\gamma^{j+1}} \sum_{l_n=l_{n-1}+1}^{l_{n-1}+j+1} \frac{1}{1-\gamma^{l_n-1}}, \\
&= \frac{(1-\gamma)^2}{1-\gamma^{j+1}} \sum_{i=0}^{j} \frac{1}{1-\gamma^{l_{n-1}+i}}, \\
&= \frac{1}{\Lambda(j+1)} \sum_{i=0}^{j} \frac{1}{\Lambda(l_{n-1}+i)}. \qquad (5.1.5)
\end{aligned}$$

On remplace (5.1.5) dans (5.1.3) :

$$p(n,j) = \sum_{l_2=2}^{\infty} \sum_{l_3=l_2+1}^{\infty} \cdots \sum_{l_{n-1}=l_{n-2}+1}^{\infty} \left\{\prod_{i=2}^{n-1} \frac{\gamma^{l_i-1}}{\Lambda(l_i-1)}\right\} \frac{1}{\Lambda(j+1)} \sum_{i=0}^{j} \frac{1}{\Lambda(l_{n-1}+i)},$$

$$\begin{aligned}
p(n,j) &= \frac{1}{\Lambda(j+1)} \sum_{i=0}^{j} \left[\sum_{l_2=2}^{\infty} \sum_{l_3=l_2+1}^{\infty} \cdots \sum_{l_{n-1}=l_{n-2}+1}^{\infty} \left\{\prod_{i=2}^{n-1} \frac{\gamma^{l_i-1}}{\Lambda(l_i-1)}\right\} \frac{1}{\Lambda(l_{n-1}+i)}\right], \\
&= \sum_{i=0}^{j} \frac{p(n-1,i)}{\Lambda(j+1)}.
\end{aligned}$$

Utilisant (5.1.4), on obtient :

$$\begin{aligned} p(2,j) &= \frac{1}{\Lambda(j+1)} \sum_{i=0}^{j} p(1,i), \\ &= p(1,j) \sum_{i=0}^{j} \frac{1}{\Lambda(i+1)}, \\ &= p(1,j) \left[1 + \frac{1}{\Lambda(2)} + \cdots + \frac{1}{\Lambda(j+1)} \right], \\ &\geq p(1,j). \end{aligned}$$

De plus,

$$\begin{aligned} p(n+1,j) - p(n,j) &= \sum_{i=0}^{j} \frac{p(n,i)}{\Lambda(j+1)} - \sum_{i=0}^{j} \frac{p(n-1,i)}{\Lambda(j+1)}, \\ &= \frac{1}{\Lambda(j+1)} \sum_{i=0}^{j} \{p(n,i) - p(n-1,i)\}, \\ &= \frac{1}{\Lambda(j+1)} \sum_{i=1}^{j} \{p(n,i) - p(n-1,i)\}. \end{aligned}$$

En poursuivant le calcul d'une façon récursive et en utilisant le fait que $p(2,j) \geq p(1,j)\ \forall j \geq 0$, on montre que $p(n+1,j) \geq p(n,j), \forall j \geq 0$. Ainsi, comme la suite $\{p(n,j)\}_{n\geq 1}$ est croissante et majorée, elle est donc convergente. Notons $\lim_{n\longrightarrow\infty} p(n,j) = q_j,\ j \geq 0$.

Pour $j = 0$,

$$\lim_{n\longrightarrow\infty} p(n,0) = 1.$$

Ainsi, $q_0 = 1 = \frac{1}{\gamma^0}$. En raisonnant par récurrence, et pour un $j \geq 1$, supposons que $q_i = \frac{1}{\gamma^i}\ \forall 0 \leq i \leq j-1$ et montrons que $q_j = \frac{1}{\gamma^j}$.

D'après l'équation (5.1.4), et en passant à la limite, on obtient :

$$q_j = \frac{1}{\Lambda(j+1)} \sum_{i=0}^{j} q_i,$$

$$\begin{aligned}\left(1-\frac{1}{\Lambda(j+1)}\right)q_j &= \frac{1}{\Lambda(j+1)}\sum_{i=0}^{j}\frac{1}{\gamma^i},\\ \left(1-\frac{1-\gamma}{1-\gamma^{j+1}}\right)q_j &= \frac{1-\gamma}{1-\gamma^{j+1}}\frac{1-\frac{1}{\gamma^j}}{1-\frac{1}{\gamma}},\\ q_j &= \frac{1}{\gamma^j}.\end{aligned}$$

□

Ainsi, d'après le Théorème 5.2, la loi de la *va* Δ_{L_n} est asymptotiquement géométrique de paramètre $\left(1-\frac{1}{\gamma}\right)$ et de fonction de masse :

$$\begin{aligned}p_j(\gamma) &= \lim_{n\longrightarrow\infty}\mathbb{P}\left[\Delta_{L_n}=j\right],\\ &= \left(1-\frac{1}{\gamma}\right)\left(\frac{1}{\gamma}\right)^{j-1},\\ &= \frac{\gamma-1}{\gamma^j},\ j\geq 1. \qquad (5.1.6)\end{aligned}$$

5.2 Tests d'adéquation

5.2.1 Test pour le modèle *iid*

Avant d'ajuster un modèle de Yang ou LDM à un ensemble de données, il faut d'abord évaluer la nécessité de ce niveau supplémentaire de complexité. Pour ce faire, une approche consiste à tester l'hypothèse nulle H_0 : les données proviennent d'une suite de variables aléatoires *iid*. Nous pouvons fonder un tel test d'adéquation sur la suite des indicatrices de records $\{\delta_t, 1\leq t\leq T\}$ en utilisant un théorème d'Arnold *et al.* (1998, page 25) qui montrent que, sous cette hypothèse nulle

$$\mathcal{N}_T=\frac{N_T-\log(T)}{\sqrt{\log(T)}}\xrightarrow{\mathcal{L}}N(0,1).$$

Comme les modèles de Yang et LDM impliquent une augmentation du nombre de records, on rejette H_0 si

$$\mathcal{N}_T>z_{1-\alpha},$$

où $z_{1-\alpha}$ = quantile d'ordre $1-\alpha$ d'une $N(0,1)$ et α est le risque d'erreur de première espèce (la valeur du risque α doit être fixée a priori).

5.2.2 Tests d'adéquation pour le modèle de Yang

Dans le contexte d'un modèle de Yang, d'après le Théorème 5.2, la loi du temps inter-records $\{\Delta_{L_n}, n \geq 1\}$ est asymptotiquement géométrique de paramètre $\left(1-\frac{1}{\gamma}\right)$ et de fonction de masse :

$$p_j(\gamma) = \lim_{n\longrightarrow\infty} \mathbb{P}\left[\Delta_{L_n} = j\right] = \frac{\gamma-1}{\gamma^j},\ j \geq 1. \tag{5.2.1}$$

Dans la suite de cette section, l'adéquation à un modèle de Yang est évaluée en vérifiant si les valeurs observées des temps inter-records, après une période d'échauffement permettant à l'effet asymptotique de s'installer, sont en accord avec la distribution géométrique (5.2.1), où le paramètre γ doit être estimé. Notons aussi que dans la suite de cette section N_T dénote le nombre de records après la période d'échauffement et T le temps présent.

5.2.2.1 Test χ^2 de Pearson pour le modèle de Yang

Une première approche, consiste à construire un test d'adéquation basé sur le χ^2 de Pearson (1900). Conditionnons sur l'événement $N_T = m$. Fixons $K > 1$. Le problème de choisir K en pratique est très complexe et ne sera pas discuté ici. Le lecteur intéressé peut consulter Rayner et Rayner (2001) . Sur la base d'une partition (de sous ensembles disjoints) $\Pi_1 \bigcup ... \bigcup \Pi_K$ de l'ensemble $\{1, 2, \ldots, \infty\}$, n_k, $1 \leq k \leq K$ dénote le nombre de Δ_{L_n} qui tombent dans Π_k avec $n_1 + \ldots + n_K = m-1$ (car pour $N_T = m$ records il correspond $m-1$ inter-records). De plus, notons $\pi_k(\gamma) = \sum_{j\in\Pi_k} p_j(\gamma)$, $1 \leq k \leq K$. La statistique de χ^2 de Pearson est calculée comme étant

$$\chi(\gamma) = \sum_{k=1}^{K} \frac{\left(n_k - (m-1)\pi_k(\gamma)\right)^2}{(m-1)\pi_k(\gamma)}. \tag{5.2.2}$$

La valeur de cette statistique est comparée à $x^2_{K-1,1-\alpha}$ le quantile d'ordre $(1-\alpha)$ de la loi khi-deux à $K-1$ degrés de liberté, notée χ^2_{K-1}. Quand $\chi(\gamma) > x^2_{K-1,1-\alpha}$, le test rejette, au niveau asymptotique α, l'hypothèse H_0. Ainsi, les valeurs observées de Δ_{L_n} ne sont pas en accord avec la distribution géométrique et le modèle n'est pas de Yang.

Cependant, la statistique (5.2.2) est inutilisable car le paramètre γ est inconnu. On doit donc l'estimer. Pour ce faire, on calcule la valeur $\widetilde{\gamma}$ qui minimise $\chi(\gamma)$. La statistique utilisable est

$$\chi(\widetilde{\gamma}) = \arg\min_{\gamma} \chi(\gamma). \tag{5.2.3}$$

Selon des résultats classiques (Bishop *et al.* (2007, page 348)) si les données proviennent bien d'une loi géométrique, et en conditionnant sur la suite N_T, la statistique $\chi(\widetilde{\gamma}) \xrightarrow{\mathcal{L}} \chi^2_{K-2}$. En pratique donc la valeur de $\chi(\widetilde{\gamma})$ est comparée à $x^2_{K-2,1-\alpha}$ le quantile d'ordre $(1-\alpha)$ de χ^2_{K-2}. Si $\chi(\widetilde{\gamma}) > x^2_{K-2,1-\alpha}$, le test rejette, au niveau asymptotique α, l'hypothèse nulle que les Δ_{L_n} suivent une loi géométrique, ce qui jette le doute sur le modèle de Yang.

Notons que lorsque le test rejette, l'analyse des composantes $\frac{(n_k-(m-1)\pi_k(\widetilde{\gamma}))^2}{(m-1)\pi_k(\widetilde{\gamma})}$ peut aider à identifier les Π_k où les données ne collent pas avec le modèle de Yang. Ceci peut aider à proposer des améliorations.

5.2.2.2 Test lisse d'adéquation pour le modèle de Yang

Un deuxième test d'adéquation pour le modèle de Yang est basé sur le test lisse développé par Rayner et Best (1989). Nous conditionnons aussi sur $N_T = m$ et fixons $K > 1$. Afin de tester l'hypothèse H_0 =« la distribution asymptotique du temps inter-records est géométrique de paramètre γ inconnu », la statistique du test (Rayner et Best (1989) page 96) est donnée comme étant

$$\hat{S}_K = \sum_{r=2}^{K+1} \hat{V}_r^2, \tag{5.2.4}$$

avec,

$$\hat{V}_r = \frac{1}{\sqrt{(m-1)(r!)^2\left(\left(\overline{\Delta}-1\right)^2+\left(\overline{\Delta}-1\right)\right)^r}} \sum_{n=1}^{m-1} h_r\left(\Delta_{L_n}-1, \overline{\Delta}-1\right),$$

où $\overline{\Delta} = \frac{\sum_{n=1}^{m-1}\Delta_{L_n}}{m-1}$ et les $h_r(x,a)$ sont les polynômes de Meixner (1934). Ces cinq premiers polynômes sont donnés par :

$$\begin{aligned} h_1(x,a) &= (x-a), \\ h_2(x,a) &= x(x-1) - 4ax + 2a^2, \end{aligned}$$

$$
\begin{aligned}
h_3(x,a) &= x(x-1)(x-2) - 9ax(x-1) + 18xa^2 - 6a^3, \\
h_4(x,a) &= 24C_x^4 - 96aC_x^3 + 144a^2C_x^2 - 96xa^3 + 24a^4, \\
h_5(x,a) &= 120C_x^5 - 600aC_x^4 + 1200a^2C_x^3 - 1200a^3C_x^2 + 600a^4x - 120a^5,
\end{aligned}
$$

avec,

$$
C_x^k = \begin{cases} \frac{x!}{k!(x-k)!} & \text{si } x \geq k \\ 0 & \text{si } x < k \end{cases}.
$$

Si $\alpha = 5\%$ et $K = 4$, tel que recommandé par Rayner et Best (1989), et sous le modèle de Yang avec γ inconnu, la statistique de test (5.2.4) est comparée au quantile corrigée d'une χ^2_4

$$
\widehat{x^2_{4,0.95}} = 9.488 \times \left(1 + \frac{3.643}{m-1} - \frac{2.314}{\sqrt{m-1}} - \frac{0.447}{\sqrt{(m-1)\frac{\overline{\Delta}-1}{\overline{\Delta}}}} \right).
$$

Quand $\hat{S}_4 > \widehat{x^2_{4,0.95}}$, le test rejette, au niveau asymptotique $\alpha = 5\%$, l'hypothèse H_0. Ainsi les valeurs observées de Δ_{L_n} ne sont pas en accord avec la distribution géométrique ce qui jette le doute sur le modèle de Yang.

Ici aussi, lorsque le test rejette H_0, des informations diagnostiques peuvent être obtenues à partir des composantes $\hat{V}_r^2$ afin de suggérer des améliorations, mais celles-ci doivent être normalisées. Pour plus de détails, voir Henze et Klar (1996).

5.3 Application

Yang (1975) a considéré les records olympiques de la course de 200 mètres (Hommes) de 1900 jusqu'à 1972 avec $T = 16$ observations. Depuis ce temps, plusieurs records ont été observés. La table 5.1 donne les records de 1900 jusqu'à 2012 avec $T = 26$ observations.

Yang, afin d'ajuster ses observations à son modèle, avait besoin d'une estimation de γ. Pour cela, il a utilisé un argument externe basé sur le fait que durant le $20^{ième}$ siècle, la population a presque doublé tous les 36 ans, ce qui l'a conduit à estimer γ par $\hat{\gamma} = 1.08$. Pour vérifier la validité de son estimation et de son modèle, il a utilisé le résultat du Théorème 5.2 sur la distribution géométrique asymptotique (5.2.1) du temps inter-records $\Delta_{L_n} = L_{n+1} - L_n$. Il a pris une période d'échauffement de 2 records et a donc considéré les 5 derniers temps inter-records.

Année du record	Valeur du record (secondes)	Année du record	Valeur du record (secondes)
1900	22.20	1964	20.30
1904	21.60	1968	19.83
1932	21.20	1984	19.80
1936	20.70	1988	19.75
1956	20.60	1996	19.32
1960	20.50	2008	19.30

TABLE 5.1 – Records olympiques de la course de 200 mètres (Hommes) (`http://www.olympic.org`)

j	1	2	3	4+
Données de Yang (1900-1972)	4	0	1	0
$p_j(1.08)$	0.074	0.069	0.064	0.793
Totalité des données (1900-2012)	5	1	2	1
$p_j(1.77)$	0.435	0.246	0.139	0.180

TABLE 5.2 – $p_j\,(\hat{\gamma})$ et les fréquences observées du temps inter-records à partir des données de Yang et des données complètes respectivement.

Les lignes 2 et 3 de la Table 5.2 donnent les fréquences observées (au cours des 5 derniers temps inter-records) pour les données de Yang (1900-1972), ainsi que les premiers p_j (1.08). En utilisant l'estimation de Yang, $\hat{\gamma} = 1.08$, nous remarquons que la loi asymptotique géométrique du temps inter-records a comme moyenne $\frac{\hat{\gamma}}{\hat{\gamma}-1}$ =13.5 jeux entre deux records. Yang remarque que cette moyenne est beaucoup plus élevée que le temps d'attente moyen réel, qui est de 1.4 jeux, et explique cette écart flagrant en suggérant que l'augmentation rapide du nombre de records olympiques est due à des raisons autres que la croissance de la population. Il conclut aussi que γ devrait être plus élevé.

Nous essayons d'étayer les conclusions de Yang en utilisant la série complète de records (1900-2012).

Tout d'abord, nous vérifions la nécessité d'utiliser un modèle plus complexe que le modèle *iid* en appliquant le test de la Sous-Section 5.2.1. On obtient $\mathcal{N}_T = 4.84$ (p-value $= 6.49 \times 10^{-7}$). Donc, au niveau asymptotique $\alpha = 5\%$, l'hypothèse d'un modèle *iid* doit être rejetée.

Ensuite, pour vérifier si le modèle de Yang est en accord avec ces données de records, nous appliquons le test de khi-deux et le test lisse des Sous-Sections 5.2.2.1 et 5.2.2.2 respectivement. Nous utilisons la même période d'échauffement que Yang, soit les 9 derniers temps inter-records. Les fréquences observées de ces 9 inter-records apparaissent à la ligne 4 de la Table 5.2. Afin d'appliquer le test de khi-deux nous considérons la partition $\Pi_1 = \{1\}$, $\Pi_2 = \{2\}$, $\Pi_3 = \{3\}$ et $\Pi_4 = \{4+\}$ ($K = 4$). La statistique (5.2.3) prend la valeur de $1.438 < x^2_{2,0.95} = 5.992$ (p-value $= 0.487$) tandis que la statistique du test lisse (5.2.4) vaut $0.766 < \widehat{x^2_{4,0.95}} = 3.949$. Ainsi, au niveau asymptotique $\alpha = 5\%$, les données ne remettent pas en cause le modèle de Yang.

Enfin, en utilisant la méthode d'estimation du paramètre γ d'un modèle de Yang basée sur le principe du maximum de vraisemblance de la Sous-Section 4.4.4 du chapitre précédent et en se basant sur la totalité des données disponibles (les records olympiques de 1900 jusqu'à 2012) nous calculons un nouveau estimateur de γ, $\hat{\gamma}_3 = 1.77$, avec un intervalle de confiance de niveau asymptotique 95% donné par $(1.1213, 2.4215)$. En utilisant cette valeur $\hat{\gamma}_3 = 1.77$ et en appliquant l'approximation asymptotique du Théorème 5.2 nous obtenons un temps moyen inter-record estimé de 2.3 jeux (= 9.2 années), ce qui est proche du temps d'attente moyen réel de 1.89 jeux (basé sur les 9 derniers temps inter-records des données complètes (1900-2012)). Ainsi, en

se basant sur les probabilités de la ligne 5 de la Table 5.2, et sachant que le dernier record de la course de 200 mètres a eu lieu en 2008 (le record n'a pas été battu en 2012), on peut prévoir qu'il aura un nouveau record olympique de cette course durant les jeux de Rio 2016 avec une probabilité de $\frac{p_2(1.77)}{(1-p_1(1.77))} = 43.5\%$.

5.4 Conclusion

Dans ce chapitre, nous avons montré que dans le contexte d'un modèle de Yang, le comportement asymptotique de la distribution du temps inter-records $\Delta_{L_n} = L_{n+1} - L_n$ est géométrique. Puis, en se basant sur cette propriété, nous avons développé deux tests d'adéquation pour le modèle de Yang. De plus, nous avons construit un test statistique, basé sur le nombre de records N_T, qui évalue la nécessité d'un passage à un modèle non *iid* afin de modéliser certaines séries de records. Enfin, nous avons appliqué nos résultats théoriques à des données analysées précédemment par Yang (1975) présentant les records de la course de 200 mètres dans les Jeux olympiques. Les différents tests statistiques ont montré que le passage au delà d'un modèle *iid* est nécessaire et que l'hypothèse d'un modèle de Yang n'est pas déraisonnable pour ce type de données. De plus, en appliquant la méthode d'estimation du chapitre précédent afin d'estimer le paramètre γ du modèle de Yang, nous avons obtenu un estimateur qui colle mieux que celui proposé par Yang (1975) avec les données réelles.

Chapitre 6

LDM et Yang-Nevzorov : Estimation basée sur (R_n, L_n)

Dans ce chapitre, en premier lieu, nous estimons le drift θ et les paramètres de la distribution sous-jacente, dans le cas d'une loi de Gumbel de paramètres inconnus μ (paramètre de position) et β (paramètre d'échelle), d'un modèle LDM, en se basant uniquement sur les indicatrices de records. Puis, nous passons à l'utilisation de la totalité des données disponibles (valeurs et indices/indicatrices de records) afin de calculer, par plusieurs méthodes, des estimateurs des paramètres des modèles LDM et Yang-Nevzorov dans des cas où la distribution sous-jacente n'est pas nécessairement Gumbel. De plus, nous évaluons les différentes méthodes d'estimation par des simulations numériques sous R et Matlab en comparant la qualité de chaque estimateur sur plusieurs niveaux : biais, écart-type et probabilité de couverture. Enfin, nous introduisons des tests statistiques qui nous aident à vérifier la conformité du choix de la distribution sous-jacente et à choisir entre un modèle LDM et de Yang. Aussi nous évaluons les différents tests statistiques par des simulations numériques sous Matlab en calculant les quantiles empiriques sous H_0 et les variations de la puissance du test pour différentes valeurs de θ et γ.

On peut citer les articles de Carlin et Gelfand (1993), Feuerverger et Hall (1996) et Smith (1988) pour d'autres méthodes d'estimation des paramètres des modèles de records.

6.1 Estimation basée uniquement sur les δ_t

Au Chapitre 3, nous avons considéré le cas d'un modèle LDM de distribution sous-jacente de loi de Gumbel avec paramètre connu et, sans perte de généralités, nous avons considéré la loi $G(0,1)$. Ici nous nous placons dans le cadre d'un modèle LDM avec une distribution sous-jacente de loi de Gumbel de paramètres inconnus. La $t^{ième}$ observation de la série temporelle est de la forme :

$$X_t = Y_t + \theta t, \tag{6.1.1}$$

avec θ le « drift » et $\{Y_t, t \geq 1\}$ une suite infinie de variables aléatoires iid qui suivent la loi $G(\mu, \beta)$ de fonction de répartition $F(y) = \exp\left(-\exp\left(-\frac{y-\mu}{\beta}\right)\right)$.

Évidemment, il ne nous est pas donné d'observer la suite X_t, seulement ses records $R_n = X_{L_n}$ où L_n est l'indice du $n^{ième}$ record.

Notre but est d'étendre les résultats du Chapitre 3 et de tenter d'obtenir des estimateurs des paramètres θ, μ et β en utilisant le principe du maximum de vraisemblance basé sur la distribution des indicatrices de records $\{\delta_t, t \geq 1\}$, avec lesquels on avait obtenu des résultats utiles au Chapitre 3.

Si la densité de Y_t est $f_t(y) = \frac{1}{\beta}\exp\left(-\frac{y-\mu}{\beta}\right)\exp\left(-\exp\left(-\frac{y-\mu}{\beta}\right)\right)$ (la loi $G(\mu, \beta)$), d'après l'exemple 2.1, le taux de record est donné par :

$$P_t = \int f(y)\left(\prod_{k=1}^{t-1}(F(y))^{\exp\left(-\frac{\theta k}{\beta}\right)}\right)\mathrm{d}y.$$

En posant $\nu = \exp\left(-\frac{\theta}{\beta}\right)$ et $u = F(y)$, on obtient :

$$\begin{aligned} Q_t(\nu) &= \int_0^1 \left(\prod_{k=1}^{t-1} u^{\nu^k}\right)\mathrm{d}u, \\ &= \int_0^1 u^{\frac{\nu-\nu^t}{1-\nu}}\,\mathrm{d}u, \\ &= \frac{1-\nu}{1-\nu^t}. \end{aligned}$$

Ceci montre qu'en utilisant uniquement les indicatrices de records comme variable de censure, on peut extraire des informations sur le drift θ et le

paramètre d'échelle β uniquement, via le terme ν. Or comme ν est fonction de $\frac{\theta}{\beta}$, ces deux paramètres ne sont pas identifiables. En fait il est logique que ça ne le soit pas car $X_t = Y_t + \theta t \iff \frac{X_t}{\beta} = \frac{Y_t}{\beta} + \frac{\theta}{\beta} t$. Alors, on peut conclure que notre censure est informative pour le quotient $\frac{\theta}{\beta}$ mais non-informative sur μ, β et θ pris individuellement.

En utilisant la propriété d'indépendance des indicatrices de records dans un modèle LDM de distribution sous-jacente de loi de Gumbel (Borovkov, 1999), on peut alors construire une fonction de vraisemblance basée sur les δ_t. En appliquant le principe du maximum de vraisemblance, le travail consiste à trouver ν qui maximise l'expression suivante :

$$\begin{aligned} L(\nu) &= \mathbb{P}[\delta_1, \ldots, \delta_T; \nu], \\ &= \prod_{t=2}^{T} \mathbb{P}[\delta_t], \ \delta_1 = 1 \,(\text{record trivial}), \\ &= \prod_{t=2}^{T} Q_t(\nu)^{\delta_t} (1 - Q_t(\nu))^{1-\delta_t}, \\ &= \prod_{t=2}^{T} \frac{(1-\nu)^{\delta_t} (\nu - \nu^t)^{1-\delta_t}}{(1-\nu^t)}. \end{aligned}$$

Étant donné l'aspect « régulier » de cette vraisemblance, la méthode classique est de dériver $\log L(\nu)$ par rapport à ν, puis de calculer les racines de cette dérivée.

$$\begin{aligned} \log L(\nu) &= \sum_{t=2}^{T} \left[\delta_t \log(Q_t(\nu)) + (1-\delta_t) \log(1 - Q_t(\nu))\right], \\ &= \sum_{t=2}^{T} \left[\delta_t \log(1-\nu) - \delta_t \log\left(\nu\left(1-\nu^{t-1}\right)\right) + \log\left(\nu\left(1-\nu^{t-1}\right)\right)\right. \\ &\quad \left. - \log\left(1-\nu^t\right)\right], \\ &= N_T \log(1-\nu) + (T - N_T) \log(\nu) - \log\left(1-\nu^T\right) \\ &\quad - \sum_{t=2}^{T} \delta_t \log\left(1-\nu^{t-1}\right), \end{aligned} \tag{6.1.2}$$

où $\sum_{t=2}^{T} \delta_t = N_T - 1$, avec $N_T =$ le nombre de records parmi les T valeurs de δ_t.

Ainsi, il faut trouver l'estimateur $\hat{\nu}$ de ν tel que,

$$\begin{aligned}\left(\frac{\mathrm{d}\log L(\nu)}{\mathrm{d}\nu}\right)_{\nu=\hat{\nu}} &= \left(\frac{-N_T}{1-\nu} + \frac{T-N_T}{\nu} + \frac{T\nu^{T-1}}{1-\nu^T} + \sum_{t=2}^{T}\delta_t \frac{(t-1)\,\nu^{t-2}}{1-\nu^{t-1}}\right)_{\nu=\hat{\nu}}, \\ &= 0. \end{aligned} \tag{6.1.3}$$

On remarque que c'est la même équation que celle de la Section 3.3 portant sur l'estimation de $\tau = e^{-\theta}$. Ainsi, en appliquant de nouveau un théorème de Leroy *et al.* (2016) sur le comportement asymptotique des EMV dans le cas où les observations sont indépendantes mais non identiquement distribuées, on montre que $\hat{\nu}$ existe et

$$\frac{\hat{\nu}-\nu}{\sqrt{I_T^{-1}(\nu)}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.1.4}$$

où $I_T(\nu)$ dénote l'information de Fisher :

$$\begin{aligned} I_T(\nu) &= -\mathbb{E}\left[\frac{\mathrm{d}^2\log(L(\nu))}{\mathrm{d}\nu^2}\right], \\ &= \frac{1}{(1-\nu)^2}\sum_{t=1}^{T} Q_t(\nu) + \frac{1}{\nu^2}\left(T - \sum_{t=1}^{T} Q_t(\nu)\right) - \frac{T\nu^{T-2}\left(T+\nu^T-1\right)}{\left(1-\nu^T\right)^2} \\ &\quad - \sum_{t=2}^{T}\frac{(t-1)\,\nu^{t-3}\left(t-2+\nu^{t-1}\right)}{\left(1-\nu^{t-1}\right)^2} Q_t(\nu). \end{aligned} \tag{6.1.5}$$

De plus, on montre aussi que

$$\frac{I_T(\nu)}{T} \longrightarrow I(\nu) = \frac{1}{(1-\nu)\,\nu}.$$

Par suite, en se basant sur l'Équation (6.1.4) on a

$$\sqrt{T}\,(\hat{\nu}-\nu) \xrightarrow{\mathcal{L}} N\left(0, I^{-1}(\nu)\right).$$

Or, le quotient $\frac{\theta}{\beta}$ du modèle LDM est donné par $\frac{\theta}{\beta} = -\log(\nu)$. Ainsi, d'après la méthode delta on obtient :

$$\frac{\sqrt{T}\left(\widehat{\left(\frac{\theta}{\beta}\right)}-\frac{\theta}{\beta}\right)}{\frac{1}{\nu}\sqrt{I^{-1}(\nu)}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.1.6}$$

et, pour un niveau de confiance asymptotique $1-\alpha$, un intervalle de confiance asymptotique de $\frac{\theta}{\beta}$ est donné par :

$$\left[\widehat{\left(\frac{\theta}{\beta}\right)}-\frac{\sqrt{I^{-1}(\hat{\nu})}}{\hat{\nu}\sqrt{T}}z_{1-\alpha/2}, \widehat{\left(\frac{\theta}{\beta}\right)}+\frac{\sqrt{I^{-1}(\hat{\nu})}}{\hat{\nu}\sqrt{T}}z_{1-\alpha/2}\right]. \tag{6.1.7}$$

Il ne reste plus qu'à réexprimer l'information de Fisher (6.1.5) dans la paramétrisation originale (c'est à dire en $\frac{\theta}{\beta}$), en appliquant la méthode delta sur l'équation (6.1.4) et en utilisant encore une fois la relation $\frac{\theta}{\beta} = -\log(\nu)$. On obtient :

$$I_T^{-1}\left(\frac{\theta}{\beta}\right) = I_T^{-1}(\nu) \times \left(\frac{\mathrm{d}}{\mathrm{d}\nu}(-\log(\nu))\right)^2.$$

Ainsi

$$\begin{aligned} I_T\left(\frac{\theta}{\beta}\right) &= \frac{I_T(\nu)}{\left(\frac{\mathrm{d}}{\mathrm{d}\nu}(-\log(\nu))\right)^2}, \\ &= \nu^2 I_T(\nu), \\ &= \frac{e^{-\frac{\theta}{\beta}}}{\left(1-e^{-\frac{\theta}{\beta}}\right)^2}\sum_{t=1}^{T} P_t\left(\frac{\theta}{\beta}\right) - \frac{T^2 e^{-T\frac{\theta}{\beta}}}{\left(1-e^{-T\frac{\theta}{\beta}}\right)^2} \\ &\quad -\sum_{t=2}^{T}\frac{(t-1)^2 e^{-(t-1)\frac{\theta}{\beta}}}{\left(1-e^{-(t-1)\frac{\theta}{\beta}}\right)^2}P_t\left(\frac{\theta}{\beta}\right). \end{aligned}$$

et

$$\frac{\left(\widehat{\left(\frac{\theta}{\beta}\right)}-\frac{\theta}{\beta}\right)}{\sqrt{I_T^{-1}\left(\frac{\theta}{\beta}\right)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{6.1.8}$$

En conclusion, en se basant uniquement sur les indicatrices de records on peut estimer le quotient $\frac{\theta}{\beta}$, mais on ne peut pas estimer les autres paramètres

μ, β, θ d'un modèle LDM avec distribution sous-jacente $G(\mu, \beta)$ de paramètres inconnus.

Donc, pour les modèles LDM où on a plusieurs paramètres à estimer, il est indispensable d'utiliser plus d'informations, à savoir ici les couples des données disponibles $\{(R_n, L_n), n \geq 1\}$ (valeurs et indices de records), pour tenter d'obtenir des estimateurs des paramètres.

6.2 Estimation basée sur les couples (R_n, L_n)

6.2.1 Estimation de θ du modèle LDM

Notre but est maintenant de revisiter les résultats du Chapitre 3 et d'estimer le drift θ d'un modèle LDM de la forme $X_t = Y_t + \theta t$ dans le cas d'une distribution sous-jacente $G(0,1)$, en se basant sur le principe du maximum de vraisemblance et en utilisant maintenant les couples $\{(R_n, L_n), n \geq 1\}$.

Pour une suite de T observations $X = (X_1, X_2, \ldots, X_T)$, on suppose que les données disponibles sont les indices de records $L_1 = 1, L_2, \ldots, L_m$ et les valeurs de records $R_1 = X_1, R_2 = X_{L_2}, \ldots, R_m = X_{L_m}$, avec $m = N_T$ nombre de records parmi les T observations. Par conséquent, nous savons aussi qu'il n'y a pas de records parmi les X inobservables.

Remarquons que la situation ressemble beaucoup à un contexte où il y a censure. Cette similitude a été exploitée par Smith (1988) pour obtenir de façon ad hoc un estimateur de θ. Cependant ici, comme on l'a vu, la censure (les δ_t) est informative de sorte que l'approche de Smith (1988) n'exploite pas toute l'information contenue dans les données.

6.2.1.1 Vraisemblance de Carlin et Gelfand (1993)

Selon Carlin et Gelfand (1993) la fonction de vraisemblance associée à la totalité des informations disponibles est donnée par

$$\begin{aligned} L(\theta) &= L\left(L_1 = 1, X_1 = x_1, L_2 = l_2, X_{L_2} = x_{l_2}, \ldots, \right. \\ &\quad \left. L_{N_T} = l_{N_T}, X_{L_{N_T}} = x_{l_{N_T}}; \theta\right), \qquad (6.2.1) \\ &= f(x_{l_1})\,\mathbb{P}[l_2 \mid x_{l_1}] \times f(x_{l_2} \mid x_{l_1}, l_2) \times \ldots \times \mathbb{P}\left[l_{N_T} \mid x_{l_1}, l_2, \ldots, x_{l_{N_T-1}}\right] \\ &\quad \times f\left(x_{l_{N_T}} \mid x_{l_1}, l_2, \ldots, l_{N_T}\right) \times \mathbb{P}\left[\text{Pas de records après } l_{N_T} \mid x_{l_1}, l_2, \ldots, x_{l_{N_T}}\right]. \\ &\qquad (6.2.2) \end{aligned}$$

On définit les événements :

$$\begin{aligned} A_i &= \left\{X_{l_i+1} \leq x_{l_i}, \ldots, X_{l_{i+1}-1} \leq x_{l_i}\right\}, \\ &= \text{Pas de records entre deux records consécutifs, } l_i \text{ et } l_{i+1},\ i = 1, \ldots, N_T - 1, \end{aligned}$$

et

$$\begin{aligned} A_{N_T} &= \left\{X_{l_{N_T}+1} \leq x_{l_{N_T}}, \ldots, X_T \leq x_{l_{N_T}}\right\}, \\ &= \text{Pas de records après le record } l_{N_T}. \end{aligned}$$

De plus, notons,

$$U = \left(X_{l_1}, \ldots, X_{l_{N_T}}\right) \text{ de densité } f(u;\theta) \text{ et } V = X \setminus U.$$

Ainsi,

$$\begin{aligned} f(x;\theta) &= f(x_1, x_2, \ldots, x_T;\theta), \\ &= f(u, v;\theta), \\ &= f(u;\theta)\, f(v \mid u;\theta), \end{aligned}$$

où $f(v \mid u;\theta)$ est la densité conditionnelle de $v \mid u$. Posons,

$$B = \bigcap_{i=1}^{N_T} A_i \,(\text{Les observations non records}).$$

Alors, l'expression (6.2.2) s'écrit :

$$\begin{aligned} \int_B f(u, v;\theta)\, \mathrm{d}v &= f(u;\theta) \int_B f(v \mid u;\theta)\, \mathrm{d}v, \\ &= f(u;\theta)\, \mathbb{P}[B \mid u;\theta], \\ &= f\left(x_{l_1}, x_{l_2}, \ldots, x_{l_{N_T}}\right) \times \mathbb{P}\left[\bigcap_{i=1}^{N_T} A_i \mid x_{l_j}, j = 1, \ldots, N_T;\theta\right], \end{aligned}$$

$$\begin{aligned} \int_B f(u, v;\theta)\, \mathrm{d}v &= f(x_{l_1}) \times \left\{\prod_{i=2}^{N_T} f\left(x_{l_i} \mid x_{l_j}, j = 1, \ldots, i-1\right)\right\} \\ &\quad \times \mathbb{P}\left[\bigcap_{i=1}^{N_T} A_i \mid x_{l_j}, j = 1, \ldots, N_T\right]. \end{aligned} \tag{6.2.3}$$

Arnold *et al.* (1998, page 28) montrent que, dans le cas *iid,* la suite des valeurs de records forme une chaîne de Markov. Ainsi, il n'est pas déraisonnable de poser comme première approximation que les valeurs de records forment une chaîne de Markov. Et comme les X_t sont indépendantes, (6.2.3) s'écrit,

$$L(\theta) = f(x_{l_1}) \left\{ \prod_{i=2}^{N_T} f\left(x_{l_i} \mid x_{l_{i-1}}\right) \right\} \left\{ \prod_{i=1}^{N_T-1} \mathbb{P}\left[A_i \mid x_{l_i}, x_{l_{i+1}}\right] \right\} \mathbb{P}\left[A_{N_T} \mid x_{l_{N_T}}\right]. \tag{6.2.4}$$

Afin de maximiser l'expression (6.2.4) et trouver un estimateur $\hat{\theta}$ de θ, il faut déterminer les probabilités conditionnelles $\left\{f\left(x_{l_i} \mid x_{l_{i-1}}\right)\right\}_{2 \leq i \leq N_T}$. Par suite, il faut définir la loi jointe $f_{X_{l_i}, X_{l_{i-1}}}\left(x_{l_i}, x_{l_{i-1}}\right)$ de deux records consécutifs.

Comme nous sommes dans le cas d'un modèle LDM avec une distribution sous-jacente de loi de Gumbel, une première approche, qui ne semble pas déraisonnable à première vue, est de supposer que la distribution jointe suit une loi bi-Gumbel.

Gumbel et Mustafi (1967) ont présenté deux telles lois qu'ils appellent des bi-Gumbel. Si X et Y sont deux variables aléatoires qui suivent la loi $G(0,1)$ de fonction de répartition $F(x) = e^{-e^{-x}}$, alors la fonction de répartition jointe de la première formulation est :

$$F_{X,Y}(x, y, a) = \exp\left[-\left(e^{-x} + e^{-y}\right) + a\left(e^{x} + e^{y}\right)^{-1}\right], \tag{6.2.5}$$

et celle de la seconde est

$$F_{X,Y}(x, y, b) = \exp\left[-\left(e^{-bx} + e^{-by}\right)^{1/b}\right]. \tag{6.2.6}$$

Dans les deux cas $0 \leq a < 1$ et $b \geq 1$ sont des paramètres de dépendance entre X et Y. Selon notre choix de la formulation (6.2.5) ou (6.2.6) on dit que nous sommes dans le cas d'un $a - modèle$ ou $b - modèle$.

Ainsi, la densité jointe est :

- Cas $a - modèle$,

$$\begin{aligned} f_{X,Y}(x, y, a) &= F_{X,Y}(x, y, a)\, e^{-(x+y)} \left[1 - a\left(e^{2x} + e^{2y}\right)\left(e^{x} + e^{y}\right)^{-2}\right. \\ &\quad \left. + 2ae^{2(x+y)}\left(e^{x} + e^{y}\right)^{-3} + a^2 e^{2(x+y)}\left(e^{x} + e^{y}\right)^{-4}\right]. \end{aligned} \tag{6.2.7}$$

- Cas $b - modèle$,

$$\begin{aligned} f_{X,Y}(x,y,b) &= F_{X,Y}(x,y,b) \times e^{-b(x+y)} \times \left(e^{-bx}+e^{-by}\right)^{-2+\frac{1}{b}} \\ &\quad \times \left[b-1+\left(e^{-bx}+e^{-by}\right)^{\frac{1}{b}}\right]. \end{aligned} \tag{6.2.8}$$

Cependant l'utilisation des Équations (6.2.7) ou (6.2.8) afin de calculer l'expression (6.2.4) et obtenir un estimateur du drift θ en se basant sur le principe du maximum de vraisemblance présente plusieurs inconvénients :

1. Nous devons supposer que dans un modèle LDM de distribution sous-jacente de loi de Gumbel, la distribution jointe des valeurs de records consécutifs suit la loi bi-Gumbel, ce qui n'est pas nécessairement vrai, surtout que, dans le cas *iid,* Arnold *et al.* (1998) et Nevzorov (2001) montrent que la loi des valeurs de records est différente que la loi des observations sous-jacentes (Théorème de dualité).
2. D'après Borovkov (1999), dans un modèle LDM la suite des valeurs de records $\{R_n, n \geq 1\}$ n'est pas nécessairement une chaîne de Markov.

Donc, en se basant sur les points ci dessus, il est plus avantageux de chercher un estimateur du drift en travaillant directement sur la maximisation de l'expression (6.2.1) et d'éviter l'expression (6.2.4) où nous sommes obligés de poser des hypothèses, plus ou moins exactes, sur la distribution jointe et la propriété Markovienne des valeurs de records, afin d'expliciter la fonction de vraisemblance à maximiser.

6.2.1.2 Nouvelle méthode d'estimation

Théorème 6.1. *Borovkov (1999), Dans un modèle LDM* $(\theta > 0)$ *la suite des couples* $\{(R_n, L_n), n \geq 1\}$ *forme une chaîne de Markov homogène de probabilité de transition*

$$\mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}\right] = f\left(x_{l_{n+1}} - \theta l_{n+1}\right) \prod_{t=l_n+1}^{l_{n+1}-1} F\left(x_{l_n} - \theta t\right), \tag{6.2.9}$$

avec $F(\cdot)$*, la fonction de répartition des* Y*, et un espace d'état*

$$\begin{aligned} E &= \{L_1 = l_1 = 1 < L_2 = l_2 < \cdots < L_{N_T} = l_{N_T} \leq T \text{ et} \\ &\quad R_1 = x_{l_1} < R_2 = x_{l_2} < \cdots < R_{N_T} = x_{l_{N_T}}\}. \end{aligned}$$

En combinant le résultat du Théorème 6.1 avec l'Équation (6.2.1), on peut conclure que notre travail consiste à trouver le θ qui maximise l'expression suivante :

$$\begin{aligned}
L(\theta) &= L\left(L_1 = l_1, X_1 = x_{l_1}, L_2 = l_2, X_{L_2} = x_{l_2}, \ldots, L_{N_T} = l_{N_T}, X_{L_{N_T}} = x_{l_{N_T}}; \theta\right), \\
&= \mathbb{P}\left[L_1 = l_1, X_1 = x_{l_1}\right] \times \mathbb{P}\left[L_2 = l_2, X_{L_2} = x_{l_2} \mid L_1 = l_1, X_1 = x_{l_1}\right] \times \ldots \\
&\quad \times \mathbb{P}\Big[L_{N_T} = l_{N_T}, X_{L_{N_T}} = x_{l_{N_T}} \mid L_{N_T-1} = l_{N_T-1}, \\
&\quad X_{L_{N_T-1}} = x_{l_{N_T-1}}\Big], \, l_1 = 1 \, (\text{record trivial}), \\
&= \mathbb{P}\left[X_1 = x_{l_1}\right] \times f\left(x_{l_2} - \theta l_2\right) \times \left(\prod_{t=l_1+1}^{l_2-1} F\left(x_{l_1} - \theta t\right)\right) \times \ldots \\
&\quad \times f\left(x_{l_{N_T}} - \theta l_{N_T}\right) \times \left(\prod_{t=l_{N_T-1}+1}^{l_{N_T}-1} F\left(x_{l_{N_T-1}} - \theta t\right)\right).
\end{aligned} \tag{6.2.10}$$

En spécialisant cette expression au cas d'une distribution sous-jacente $G(0,1)$,

$$\begin{aligned}
L(\theta) &= e^{-\left(x_{l_1} - \theta\right)} e^{-e^{-\left(x_{l_1} - \theta\right)}} e^{-\left(x_{l_2} - \theta l_2\right)} e^{-e^{-\left(x_{l_2} - \theta l_2\right)}} \left(\prod_{t=l_1+1}^{l_2-1} e^{-e^{-\left(x_{l_1} - \theta t\right)}}\right) \times \ldots \\
&\quad \times e^{-\left(x_{l_{N_T}} - \theta l_{N_T}\right)} e^{-e^{-\left(x_{l_{N_T}} - \theta l_{N_T}\right)}} \left(\prod_{t=l_{N_T-1}+1}^{l_{N_T}-1} e^{-e^{-\left(x_{l_{N_T-1}} - \theta t\right)}}\right), \\
&= \exp\left(-\sum_{k=1}^{N_T} \left(x_{l_k} - \theta l_k\right)\right) \exp\left(-\sum_{k=1}^{N_T} \left(e^{-\left(x_{l_k} - \theta l_k\right)}\right)\right) \left[\exp\left(-\sum_{t=l_1+1}^{l_2-1} e^{-\left(x_{l_1} - \theta t\right)}\right) \times \ldots \right. \\
&\quad \left. \times \exp\left(-\sum_{t=l_{N_T-1}+1}^{l_{N_T}-1} e^{-\left(x_{l_{N_T-1}} - \theta t\right)}\right)\right], \\
&= \exp\left(-\sum_{k=1}^{N_T} \left(x_{l_k} - \theta l_k\right)\right) \times \exp\left(-\sum_{k=1}^{N_T} \left(e^{-\left(x_{l_k} - \theta l_k\right)}\right)\right) \\
&\quad \times \exp\left(-\sum_{i=1}^{N_T-1} \sum_{t=l_i+1}^{l_{i+1}-1} e^{-\left(x_{l_i} - \theta t\right)}\right).
\end{aligned} \tag{6.2.11}$$

Comme cette vraisemblance est « régulière », l'approche classique est de dériver $\log L(\theta)$ par rapport à θ puis calculer les racines de cette dérivée.

$$\log L(\theta) = -\left[\sum_{k=1}^{N_T}(x_{l_k} - \theta l_k) + \sum_{k=1}^{N_T}\left(e^{-\left(x_{l_k}-\theta l_k\right)}\right) + \sum_{i=1}^{N_T-1}\sum_{t=l_i+1}^{l_{i+1}-1} e^{-\left(x_{l_i}-\theta t\right)}\right].$$

Ainsi l'estimateur $\hat{\theta}_5$ de θ est tel que,

$$\left(\frac{\mathrm{d}\log L(\theta)}{\mathrm{d}\theta}\right)_{\theta=\hat{\theta}_5} = 0.$$

Or,

$$\frac{\mathrm{d}\log L(\theta)}{\mathrm{d}\theta} = -\left[-\sum_{k=1}^{N_T} l_k + \sum_{k=1}^{N_T} l_k e^{-\left(x_{l_k}-\theta l_k\right)} + \sum_{i=1}^{N_T-1}\sum_{t=l_i+1}^{l_{i+1}-1} t e^{-\left(x_{l_i}-\theta t\right)}\right],$$

et, de plus,

$$\frac{\mathrm{d}^2\log L(\theta)}{\mathrm{d}\theta^2} = -\left[\sum_{k=1}^{N_T} l_k^2 e^{-\left(x_{l_k}-\theta l_k\right)} + \sum_{i=1}^{N_T-1}\sum_{t=l_i+1}^{l_{i+1}-1} t^2 e^{-\left(x_{l_i}-\theta t\right)}\right].$$

Par suite, $\frac{\mathrm{d}^2\log L(\theta)}{\mathrm{d}\theta^2} < 0$, ce qui permet de conclure que la log-vraisemblance est concave et admet un unique maximum.

Remarque 6.2. En utilisant la présente méthode d'estimation on peut facilement aller au delà d'une distribution sous-jacente de loi de Gumbel. En effet, en se basant sur l'Équation (6.2.10) on peut considérer une distribution sous-jacente quelconque de fonction de répartition $F(\cdot)$ et de densité $f(\cdot)$, et calculer nos estimateurs en se basant sur le principe du maximum de vraisemblance comme dans le cas d'une loi de Gumbel. Ainsi, l'utilisation de la totalité des données disponibles nous a aidé à dépasser la contrainte d'une distribution sous-jacente de loi de Gumbel.

Afin d'étudier empiriquement la qualité de l'estimateur obtenu par cette nouvelle méthode d'estimation, 5000 séries chronologiques selon un modèle LDM pour différentes valeurs de θ avec $T = 100$ observations et une distribution sous-jacente $G(0,1)$ ont été générées. Pour chacune de ces séries, la suite des indices et des valeurs de records ont été extraites et l'estimateur

θ	Biais	$\sqrt{\hat{\mathbb{V}}\left(\hat{\theta}_5\right)}$	$\overline{\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\theta)}{\mathrm{d}\theta^2}\right)^{-1}}}$
0.05	0.01	0.0410	0.0334
0.10	$\simeq 0.00$	0.0074	0.0072
0.15	$\simeq 0.00$	0.0054	0.0054
0.20	$\simeq 0.00$	0.0045	0.0045
0.25	$\simeq 0.00$	0.0039	0.0040

TABLE 6.1 – Biais et écarts-types de $\hat{\theta}_5$, pour différentes valeurs du drift θ

$\hat{\theta}_5$ a été calculé. De ces 5000 séries, les caractéristiques habituelles ont été empiriquement estimées.

La Table 6.1 donne le biais de l'estimateur dans la deuxième colonne. La troisième colonne du tableau donne l'écart-type empirique (à partir de 5000 séries) de $\hat{\theta}_5$. Enfin, la dernière colonne donne la moyenne empirique (sur les 5000 séries) de $\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\theta)}{\mathrm{d}\theta^2}\right)^{-1}}$.

On remarque que l'estimateur $\hat{\theta}_5$ de θ est de bonne qualité, tant sur le plan du biais et de l'écart-type. D'autre part, en comparant notre estimateur avec $\hat{\theta}_3$, obtenu par la méthode EMV basée uniquement sur les indicatrices de records du Chapitre 3, Section 3.3 (voir les Tables 3.2 et 3.3 pages 60, 61), on peut conclure que les deux estimateurs sont presque sans biais, avec une petite préférence pour $\hat{\theta}_3$ pour les petites valeurs de θ, mais l'estimateur obtenu en se basant sur les couples des données (R_n, L_n) est moins variable, surtout pour les grandes valeurs de θ. Notons que, l'estimateur $\hat{\theta}_3$ obtenu au Chapitre 3 est aussi utile pour donner un point de départ de l'algorithme numérique menant à $\hat{\theta}_5$. Enfin, signalons que $\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\theta)}{\mathrm{d}\theta^2}\right)^{-1}}$ est une bonne approximation de l'écart-type exact de $\hat{\theta}_5$. Par suite il n'est pas déraisonnable de conjecturer que

$$Z = \frac{\hat{\theta}_5 - \theta}{\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\theta)}{\mathrm{d}\theta^2}\right)^{-1}}} \sim N(0,1).$$

	$1-\alpha$		
θ	90%	95%	99%
0.05	79.58%	85.32%	93.10%
0.10	83.72%	89.60%	95.84%
0.15	86.70%	92.04%	97.38%
0.20	87.12%	92.44%	97.70%
0.25	87.94%	93.00%	98.38%

TABLE 6.2 – Probabilité de couverture de l'intervalle de confiance (6.2.12) pour différentes valeurs du drift θ et du degré de confiance $1-\alpha$ avec une distribution sous-jacente $G(0,1)$

On pourrait ainsi obtenir un intervalle de confiance de θ de niveau asymptotique $1-\alpha$:

$$\mathbb{P}\left[\hat{\theta}_5 - \sqrt{-\left(\left.\frac{\mathrm{d}^2 \log L(\theta)}{\mathrm{d}\theta^2}\right|_{\theta=\hat{\theta}_5}\right)^{-1}} \times z_{1-\alpha/2} \leq \theta \leq \hat{\theta}_5 + \sqrt{-\left(\left.\frac{\mathrm{d}^2 \log L(\theta)}{\mathrm{d}\theta^2}\right|_{\theta=\hat{\theta}_5}\right)^{-1}} \times z_{1-\alpha/2}\right] \sim 1-\alpha. \quad (6.2.12)$$

Pour valider cette conjecture, la Table 6.2 présente les probabilités de couverture de l'intervalle de confiance (6.2.12), pour différents niveaux de confiance $1-\alpha$. Les niveaux de confiance réels sont proches de ceux visés, sauf pour les petites valeurs de θ. Mais globalement les niveaux de confiance réels de l'intervalle de confiance (3.3.11) basé sur $\hat{\theta}_3$ collent mieux avec ceux visés (voir la Table 3.4 page 61).

Ainsi, l'utilisation des valeurs de records, en plus des indices de records, donne plus d'informations sur notre paramètre à estimer, ce qui aide à obtenir un estimateur de bonne qualité, surtout pour les grandes valeurs de θ, et indépendamment de la distribution sous-jacente.

Nous pouvons aussi appliquer cette nouvelle méthode d'estimation pour estimer le drift d'un modèle LDM de distribution sous-jacente autre que

Gumbel, mais en restant sur une fonction de vraisemblance construite à partir d'une distribution de Gumbel (6.2.11). Ici en fait c'est en utilisant une pseudo vraisemblance.

Comme application numérique, la Table 6.3 donne le biais empirique et l'écart-type approximé par simulation $\sqrt{\hat{\mathbb{V}}\left(\hat{\theta}_5\right)}$ de $\hat{\theta}_5$ de plus de la moyenne, du maximum et du minimum du rapport $\frac{N_T}{T}$ pour différentes valeurs de θ et pour des distributions sous-jacentes de $Weibull\left(\mu = \varpi - \frac{\pi}{\sqrt{6}}, \beta = \frac{\pi}{\sqrt{6}}, \sigma = 1\right)$ et $N\left(\varpi, \frac{\pi}{\sqrt{6}}\right)$ respectivement (μ =paramètre de position, β =paramètre d'échelle ($\beta > 0$), σ =paramètre de forme ($\sigma > 0$) et $\varpi = 0.5772$ la constante d'Euler-Mascheroni), avec une fonction de vraisemblance construite suivant la densité de $G(0,1)$, pour différentes valeurs de T. Le choix des paramètres des distributions sous-jacentes est basé sur la remarque que la distribution $G(0,1)$ a la même espérance et le même écart-type que $Weibull\left(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\right)$ ou $N\left(\varpi, \frac{\pi}{\sqrt{6}}\right)$.

Encore une fois, on remarque que l'estimateur $\hat{\theta}_5$ du drift θ est de bonne qualité indépendamment de la distribution sous-jacente, surtout pour $T = 200$ et pour les grandes valeurs de θ. De plus, aussi indépendamment de la distribution sous-jacente, la qualité des estimateurs s'améliore avec l'augmentation de la moyenne du nombre de records, surtout au niveau de l'écart-type.

6.2.2 Estimation de γ du modèle de Yang-Nevzorov

Dans cette sous-section notre but est de revisiter les résultats du Chapitre 4 et d'estimer γ d'un modèle de Yang-Nevzorov de la forme $X_t \sim F(\cdot)^{\gamma^t}$, en se basant sur le principe du maximum de vraisemblance et en utilisant aussi les couples $\{(R_n, L_n), n \geq 1\}$. Pour une suite de T observations $X = (X_1, X_2, \ldots, X_T)$, on rappelle que les données disponibles sont les indices de records $L_1 = 1, L_2, \ldots, L_m$ et les valeurs de records $R_1 = X_1, R_2 = X_{L_2}, \ldots, R_m = X_{L_m}$, avec $m = N_T$ nombre de records parmi les T observations.

Dans le Théorème 6.1 Borovkov (1999) montre que, dans le cas d'un modèle LDM, la suite des couples $\{(R_n, L_n), n \geq 1\}$ forme une chaîne de

θ	Biais	$\sqrt{\hat{\mathbb{V}}\left(\hat{\theta}_5\right)}$	$\overline{\left(\frac{N_T}{T}\right)}$	$\max(N_T)$	$\min(N_T)$
0.05	0.02	0.054	0.0766	19	1
0.10	0.01	0.026	0.1087	22	2
0.15	$\simeq 0.00$	0.0074	0.1443	29	4
0.20	$\simeq 0.00$	0.0057	0.1793	34	6
0.25	$\simeq 0.00$	0.0046	0.2162	38	8

(a) Cas de $Weibull\left(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\right)$, $T = 100$ observations

θ	Biais	$\sqrt{\hat{\mathbb{V}}\left(\hat{\theta}_5\right)}$	$\overline{\left(\frac{N_T}{T}\right)}$	$\max(N_T)$	$\min(N_T)$
0.05	$\simeq 0.00$	0.0158	0.1153	26	2
0.10	$\simeq 0.00$	0.0031	0.1864	33	4
0.15	$\simeq 0.00$	0.0023	0.2625	48	10
0.20	$\simeq 0.00$	0.0018	0.3348	56	15
0.25	$\simeq 0.00$	0.0016	0.4114	65	21

(b) Cas de $Weibull\left(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\right)$, $T = 200$ observations

θ	Biais	$\sqrt{\hat{\mathbb{V}}\left(\hat{\theta}_5\right)}$	$\overline{\left(\frac{N_T}{T}\right)}$	$\max(N_T)$	$\min(N_T)$
0.05	$\simeq 0.00$	0.0056	0.1058	20	3
0.10	$\simeq 0.00$	0.0042	0.1594	30	6
0.15	$\simeq 0.00$	0.0036	0.2068	34	9
0.20	$\simeq 0.00$	0.0034	0.2482	39	14
0.25	$\simeq 0.00$	0.0032	0.2870	45	17

(c) Cas $N\left(\varpi, \frac{\pi}{\sqrt{6}}\right)$, $T = 100$ observations

θ	Biais	$\sqrt{\hat{\mathbb{V}}\left(\hat{\theta}_5\right)}$	$\overline{\left(\frac{N_T}{T}\right)}$	$\max(N_T)$	$\min(N_T)$
0.05	$\simeq 0.00$	0.0018	0.1854	30	7
0.10	$\simeq 0.00$	0.0014	0.2978	49	17
0.15	$\simeq 0.00$	0.0013	0.3937	58	25
0.20	$\simeq 0.00$	0.0012	0.4825	66	29
0.25	$\simeq 0.00$	0.0012	0.5611	75	40

(d) Cas $N\left(\varpi, \frac{\pi}{\sqrt{6}}\right)$, $T = 200$ observations

TABLE 6.3 – Biais et écarts-types de $\hat{\theta}_5$, pour différentes valeurs de θ et T d'un modèle LDM de distribution sous-jacente autre que Gumbel, mais en restant sur une fonction de vraisemblance construite à partir d'une distribution de Gumbel.

Markov. Le Théorème suivant est une généralisation du résultat obtenu par Borovkov (1999) au cas d'un modèle de Yang-Nevzorov.

Théorème 6.3. *Dans un modèle de Yang-Nevzorov la suite des couples* $\{(R_n, L_n), n \geq 1\}$ *forme une chaîne de Markov homogène de probabilité de transition*

$$\begin{aligned}\mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}\right] &= \rho_{l_{n+1}} \times f\left(x_{l_{n+1}}\right) \\ &\quad \times F\left(x_{l_{n+1}}\right)^{\rho_{l_{n+1}}-1} \times F\left(x_{l_n}\right)^{\sum_{t=l_n+1}^{l_{n+1}-1}\rho_t},\end{aligned} \tag{6.2.13}$$

avec $F(\cdot)$*, la fonction de répartition de la va sous-jacente, et un espace d'état*

$$\begin{aligned}E &= \{L_1 = l_1 = 1 < L_2 = l_2 < \cdots < L_{N_T} = l_{N_T} \leq T \text{ et} \\ &\quad R_1 = x_{l_1} < R_2 = x_{l_2} < \cdots < R_{N_T} = x_{l_{N_T}}\}\end{aligned}$$

Démonstration.

$$\mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}, \ldots, L_1 = l_1, R_1 = x_{l_1}\right] =$$

$$\mathbb{P}\left[X_{l_n+1} < x_{l_n}, \ldots, X_{l_{n+1}-1} < x_{l_n}, X_{l_{n+1}} = x_{l_{n+1}}\right] = f_{X_{l_{n+1}}}\left(x_{l_{n+1}}\right) \times \prod_{t=l_n+1}^{l_{n+1}-1} F_{X_t}\left(x_{l_n}\right).$$

Or,

$$\begin{aligned}\mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}\right] &= \mathbb{P}\left[X_{l_n+1} < x_{l_n}, \ldots,\right. \\ &\quad \left. X_{l_{n+1}-1} < x_{l_n}, X_{l_{n+1}} = x_{l_{n+1}}\right], \\ &= f_{X_{l_{n+1}}}\left(x_{l_{n+1}}\right) \times \prod_{t=l_n+1}^{l_{n+1}-1} F_{X_t}\left(x_{l_n}\right).\end{aligned}$$

Par suite,

$$\begin{aligned}&\mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}, \ldots, L_1 = l_1, R_1 = x_{l_1}\right] = \\ &\qquad \mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}\right].\end{aligned}$$

Donc, toute l'information utile pour la prédiction du futur est contenue dans l'état présent du processus $\{(R_n, L_n), n \geq 1\}$, et la probabilité de transition s'écrit :

$$\begin{aligned}\mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}\right] &= \rho_{l_{n+1}} \times f\left(x_{l_{n+1}}\right) \times F\left(x_{l_{n+1}}\right)^{\rho_{l_{n+1}}-1} \\ &\quad \times \prod_{t=l_n+1}^{l_{n+1}-1} F\left(x_{l_n}\right)^{\rho_t}, \\ &= \rho_{l_{n+1}} \times f\left(x_{l_{n+1}}\right) \times F\left(x_{l_{n+1}}\right)^{\rho_{l_{n+1}}-1} \\ &\quad \times F\left(x_{l_n}\right)^{\sum_{t=l_{n+1}}^{l_{n+1}-1} \rho_t}.\end{aligned}$$

□

En appliquant la probabilité de transition du Théorème 6.3 au cas d'un modèle de Yang, c'est à dire $\rho_t = \gamma^t$, $\gamma > 1$, on obtient :

$$\begin{aligned}\mathbb{P}\left[L_{n+1} = l_{n+1}, R_{n+1} = x_{l_{n+1}} \mid L_n = l_n, R_n = x_{l_n}\right] &= \gamma^{l_{n+1}} \times f\left(x_{l_{n+1}}\right) \times F\left(x_{l_{n+1}}\right)^{\gamma^{l_{n+1}}-1} \\ &\quad \times F\left(x_{l_n}\right)^{\sum_{t=l_{n+1}}^{l_{n+1}-1} \gamma^t}, \\ &= \gamma^{l_{n+1}} \times f\left(x_{l_{n+1}}\right) \times F\left(x_{l_{n+1}}\right)^{\gamma^{l_{n+1}}-1} \\ &\quad \times F\left(x_{l_n}\right)^{\frac{\gamma^{l_n+1}-\gamma^{l_{n+1}}}{1-\gamma}}. \qquad (6.2.14)\end{aligned}$$

En combinant le résultat du Théorème 6.3 avec la vraisemblance de Carlin et Gelfand (6.2.1), on peut conclure que notre travail consiste à trouver le γ qui maximise l'expression suivante :

$$\begin{aligned}L(\gamma) &= L\left(L_1 = l_1, X_1 = x_{l_1}, L_2 = l_2, X_{L_2} = x_{l_2}, \ldots, L_{N_T} = l_{N_T}, X_{L_{N_T}} = x_{l_{N_T}}; \gamma\right), \\ &= \mathbb{P}\left[L_1 = l_1, X_1 = x_{l_1}\right] \times \mathbb{P}\left[L_2 = l_2, X_{L_2} = x_{l_2} \mid L_1 = l_1, X_1 = x_{l_1}\right] \times \ldots \\ &\quad \times \mathbb{P}\Big[L_{N_T} = l_{N_T}, X_{L_{N_T}} = x_{l_{N_T}} \mid L_{N_T-1} = l_{N_T-1}, \\ &\quad X_{L_{N_T-1}} = x_{l_{N_T-1}}\Big], \; l_1 = 1 \,(\text{record trivial}), \\ &= \gamma^{l_1} f\left(x_{l_1}\right) F\left(x_{l_1}\right)^{\gamma^{l_1}-1} \times \gamma^{l_2} f\left(x_{l_2}\right) F\left(x_{l_2}\right)^{\gamma^{l_2}-1} \times F\left(x_{l_1}\right)^{\frac{\gamma^{l_1+1}-\gamma^{l_2}}{1-\gamma}} \times \ldots \\ &\quad \times \gamma^{l_{N_T}} f\left(x_{l_{N_T}}\right) F\left(x_{l_{N_T}}\right)^{\gamma^{l_{N_T}}-1} \times F\left(x_{l_{N_T-1}}\right)^{\frac{\gamma^{l_{N_T-1}+1}-\gamma^{l_{N_T}}}{1-\gamma}}, \\ &= \gamma^{\sum_{n=1}^{N_T} l_n} \prod_{n=1}^{N_T} \left(f\left(x_{l_n}\right) F\left(x_{l_n}\right)^{\gamma^{l_n}-1}\right) \prod_{n=1}^{N_T-1} \left(F\left(x_{l_n}\right)^{\frac{\gamma^{l_n+1}-\gamma^{l_{n+1}}}{1-\gamma}}\right). \qquad (6.2.15)\end{aligned}$$

Comme cette vraisemblance est « régulière », l'approche classique est de dériver $\log L(\gamma)$ par rapport à γ puis calculer les racines de cette dérivée.

$$\begin{aligned}
\log L(\gamma) &= \log(\gamma)\sum_{n=1}^{N_T} l_n + \sum_{n=1}^{N_T}\log f(x_{l_n}) + \sum_{n=1}^{N_T}\left(\gamma^{l_n}-1\right)\log F(x_{l_n}) \\
&\quad + \sum_{n=1}^{N_T-1}\left(\frac{\gamma^{l_n+1}-\gamma^{l_{n+1}}}{1-\gamma}\right)\log F(x_{l_n}), \\
&= \log(\gamma)\sum_{n=1}^{N_T} l_n + \sum_{n=1}^{N_T}\log f(x_{l_n}) + \sum_{n=1}^{N_T-1}\left(\frac{\gamma^{l_n}-\gamma^{l_{n+1}}}{1-\gamma}-1\right)\log F(x_{l_n}) \\
&\quad + \left(\gamma^{l_{N_T}}-1\right)\log F\left(x_{l_{N_T}}\right). \qquad (6.2.16)
\end{aligned}$$

Ainsi l'estimateur $\hat{\gamma}_4$ de γ est tel que,

$$\left(\frac{\mathrm{d}\log L(\gamma)}{\mathrm{d}\gamma}\right)_{\gamma=\hat{\gamma}_4} = 0.$$

Or,

$$\begin{aligned}
\frac{\mathrm{d}\log L(\gamma)}{\mathrm{d}\gamma} &= \frac{1}{\gamma}\sum_{n=1}^{N_T} l_n + l_{N_T}\gamma^{l_{N_T}-1}\log F\left(x_{l_{N_T}}\right) \\
&\quad + \sum_{n=1}^{N_T-1}\left(\frac{\left(l_n+(1-l_n)\gamma\right)\gamma^{l_n-1}-\left(l_{n+1}+(1-l_{n+1})\gamma\right)\gamma^{l_{n+1}-1}}{(1-\gamma)^2}\right)\log F(x_{l_n}),
\end{aligned}$$

et, de plus,

$$\begin{aligned}
\frac{\mathrm{d}^2\log L(\gamma)}{\mathrm{d}\gamma^2} &= -\frac{1}{\gamma^2}\sum_{n=1}^{N_T} l_n + l_{N_T}\times(l_{N_T}-1)\times\gamma^{l_{N_T}-2}\times\log F\left(x_{l_{N_T}}\right) \\
&\quad + \sum_{n=1}^{N_T-1}\left(\frac{\left((l_n-1)l_n - 2l_n\gamma(l_n-2)+(l_n-2)(l_n-1)\gamma^2\right)\gamma^{l_n-2}}{(1-\gamma)^3}\right. \\
&\quad \left. -\frac{\left((l_{n+1}-1)l_{n+1} - 2l_{n+1}\gamma(l_{n+1}-2)+(l_{n+1}-2)(l_{n+1}-1)\gamma^2\right)\gamma^{l_{n+1}-2}}{(1-\gamma)^3}\right) \\
&\quad \times\log F(x_{l_n}).
\end{aligned}$$

γ	Biais	$\sqrt{\hat{\mathbb{V}}(\hat{\gamma}_4)}$	$\sqrt{\left(-\left(\frac{\mathrm{d}^2 \log L(\gamma)}{\mathrm{d}\gamma^2}\right)\right)^{-1}}$
1.05	0.01	0.0484	0.0347
1.10	$\simeq 0.00$	0.0091	0.0092
1.15	$\simeq 0.00$	0.0067	0.0065
1.20	$\simeq 0.00$	0.0058	0.0058
1.25	$\simeq 0.00$	0.0053	0.0054

TABLE 6.4 – Biais et écarts-types de $\hat{\gamma}_4$, pour différentes valeurs de γ

Afin d'étudier empiriquement la qualité de l'estimateur obtenu par cette nouvelle méthode d'estimation, 5000 séries chronologiques selon un modèle de Yang pour différentes valeurs de γ avec $T = 100$ observations et une distributions sous-jacente $Weibull\left(\mu = \varpi - \frac{\pi}{\sqrt{6}}, \beta = \frac{\pi}{\sqrt{6}}, \sigma = 1\right)$ ont été générées ($\varpi = 0.5772$ la constante d'Euler-Mascheroni). Pour chacune de ces séries, la suite des indices et des valeurs de records ont été extraites et l'estimateur $\hat{\gamma}_4$ a été calculé. De ces 5000 séries, les caractéristiques habituelles ont été empiriquement estimées.

La Table 6.4 donne le biais de l'estimateur dans la deuxième colonne. La troisième colonne du tableau donne l'écart-type empirique (à partir de 5000 séries) de $\hat{\gamma}_4$. Enfin, la dernière colonne donne la moyenne empirique (sur les 5000 séries) de $\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\gamma)}{\mathrm{d}\gamma^2}\right)^{-1}}$.

On remarque que l'estimateur $\hat{\gamma}_4$ de γ est de bonne qualité, tant sur le plan du biais et de l'écart-type. Notons aussi que, l'estimateur $\hat{\gamma}_3$ obtenu au Chapitre 4 est aussi utile pour donner un point de départ de l'algorithme numérique menant à $\hat{\gamma}_4$. Enfin, signalons que $\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\gamma)}{\mathrm{d}\gamma^2}\right)^{-1}}$ est une bonne approximation de l'écart-type exact de $\hat{\gamma}_4$. Par suite il n'est pas déraisonnable de conjecturer que

$$\frac{\hat{\gamma}_4 - \gamma}{\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\gamma)}{\mathrm{d}\gamma^2}\right)^{-1}}} \sim N(0,1).$$

	$1-\alpha$		
γ	90%	95%	99%
1.05	78.5%	85.0%	93.1%
1.10	83.5%	89.4%	95.7%
1.15	85.9%	91.7%	97.0%
1.20	86.3%	92.0%	97.5%
1.25	87.9%	92.9%	97.9%

TABLE 6.5 – Probabilité de couverture de l'intervalle de confiance (6.2.17) pour différentes valeurs de γ et du degré de confiance $1-\alpha$ avec une distributions sous-jacente de Weibull

On pourrait ainsi obtenir un intervalle de confiance de γ de niveau asymptotique $1-\alpha$:

$$\mathbb{P}\left[\hat{\gamma}_4 - \sqrt{-\left(\frac{\mathrm{d}^2 \log L(\gamma)}{\mathrm{d}\gamma^2}\Big|_{\gamma=\hat{\gamma}_4}\right)^{-1}} \times z_{1-\alpha/2} \leq \gamma \right.$$
$$\left. \leq \hat{\gamma}_4 + \sqrt{-\left(\frac{\mathrm{d}^2 \log L(\gamma)}{\mathrm{d}\gamma^2}\Big|_{\gamma=\hat{\gamma}_4}\right)^{-1}} \times z_{1-\alpha/2}\right] \sim 1-\alpha. \quad (6.2.17)$$

Pour tenter un peu de valider cette conjecture, la Table 6.5 présente les probabilités de couverture de l'intervalle de confiance (6.2.17), pour différents niveaux de confiance $1-\alpha$. Les niveaux de confiance réels sont proches de ceux visés, sauf pour les petites valeurs de γ. Mais globalement les niveaux de confiance réels de l'intervalle de confiance (4.4.18) basé sur $\hat{\gamma}_3$ collent mieux avec ceux visés.

Ainsi, l'utilisation des valeurs de records en plus des indices de records, donne d'informations sur notre paramètre à estimer, ce qui nous aide à obtenir un estimateur de bonne qualité, surtout pour les grandes valeurs de γ, et indépendamment de la distribution sous-jacente.

Remarque 6.4. Jusqu'à présent on n'a pas encore la théorie permettant de donner les preuves des conjectures sur le comportement asymptotique de nos

estimateurs. En effet, les (R_n, L_n) forment une chaîne de Markov de sorte que les théorèmes standards concernant le comportement des estimateurs de vraisemblance maximale ne s'appliquent pas directement.

6.3 Estimation de θ et des paramètres de la distribution sous-jacente dans un modèle LDM

Dans cette section, notre but est d'estimer les paramètres d'un modèle LDM dans le cas d'une distribution sous-jacente de loi $G(\mu, \beta)$ de paramètres inconnus, c'est-à-dire de fonction de répartition

$$F(x) = \exp\left(-\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)\right),$$

et de densité

$$f(x) = \frac{1}{\beta}\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)\exp\left(-\exp\left(-\left(\frac{x-\mu}{\beta}\right)\right)\right),$$

en se basant sur la méthode combinant le résultat du Théorème 6.1 avec l'Équation (6.2.1).

Ainsi, avec une distribution sous-jacente $G(\mu, \beta)$, la fonction de vraisemblance (6.2.10) s'écrit :

$$\begin{aligned} L(\theta, \mu, \beta) &= \mathbb{P}[X_{L_1} = x_{l_1}] f(x_{l_2} - \theta l_2) \left(\prod_{t=l_1+1}^{l_2-1} F(x_{l_1} - \theta t)\right) \times \dots \\ &\quad \times f\left(x_{l_{N_T}} - \theta l_{N_T}\right) \left(\prod_{t=l_{N_T-1}+1}^{l_{N_T}-1} F\left(x_{l_{N_T-1}} - \theta t\right)\right), \end{aligned}$$

$$
\begin{aligned}
L(\theta,\mu,\beta) &= \frac{1}{\beta} e^{-\frac{\left(x_{l_1}-\theta\right)-\mu}{\beta}} e^{-e^{-\frac{\left(x_{l_1}-\theta\right)-\mu}{\beta}}} \frac{1}{\beta} e^{-\frac{\left(x_{l_2}-\theta l_2\right)-\mu}{\beta}} e^{-e^{-\frac{\left(x_{l_2}-\theta l_2\right)-\mu}{\beta}}} \\
&\quad \times \left(\prod_{t=l_1+1}^{l_2-1} e^{-e^{-\frac{\left(x_{l_1}-\theta t\right)-\mu}{\beta}}} \right) \times \cdots \times \frac{1}{\beta} e^{-\frac{\left(x_{l_{N_T}}-\theta l_{N_T}\right)-\mu}{\beta}} e^{-e^{-\frac{\left(x_{l_{N_T}}-\theta l_{N_T}\right)-\mu}{\beta}}} \\
&\quad \times \left(\prod_{t=l_{N_T-1}+1}^{l_{N_T}-1} e^{-e^{-\frac{\left(x_{l_{N_T-1}}-\theta t\right)-\mu}{\beta}}} \right), \\
&= \frac{1}{\beta^{N_T}} \exp\left(-\sum_{k=1}^{N_T} \frac{(x_{l_k}-\theta l_k)-\mu}{\beta} \right) \exp\left(-\sum_{k=1}^{N_T} \left(e^{-\frac{\left(x_{l_k}-\theta l_k\right)-\mu}{\beta}} \right) \right) \\
&\quad \times \left[\exp\left(-\sum_{t=l_1+1}^{l_2-1} e^{-\frac{\left(x_{l_1}-\theta t\right)-\mu}{\beta}} \right) \times \ldots \times \exp\left(-\sum_{t=l_{N_T-1}+1}^{l_{N_T}-1} e^{-\frac{\left(x_{l_{N_T-1}}-\theta t\right)-\mu}{\beta}} \right) \right], \\
&= \frac{1}{\beta^{N_T}} \exp\left(-\sum_{k=1}^{N_T} \frac{(x_{l_k}-\theta l_k)-\mu}{\beta} \right) \exp\left(-\sum_{k=1}^{N_T} \left(e^{-\frac{\left(x_{l_k}-\theta l_k\right)-\mu}{\beta}} \right) \right) \\
&\quad \times \exp\left(-\sum_{i=1}^{N_T-1} \sum_{t=l_i+1}^{l_{i+1}-1} e^{-\frac{\left(x_{l_i}-\theta t\right)-\mu}{\beta}} \right). \qquad (6.3.1)
\end{aligned}
$$

Notre travail consiste donc à trouver les θ, μ et β qui maximisent l'expression (6.3.1). Comme cette vraisemblance est « régulière », la méthode classique est de dériver $\log L(\theta,\mu,\beta)$ par rapport à θ, μ et β respectivement, puis calculer les racines de ces dérivées.

$$
\begin{aligned}
\log L(\theta,\mu,\beta) &= -\left[N_T \log\beta + \sum_{k=1}^{N_T} \frac{(x_{l_k}-\theta l_k)-\mu}{\beta} + \sum_{k=1}^{N_T} \left(e^{-\frac{\left(x_{l_k}-\theta l_k\right)-\mu}{\beta}} \right) \right. \\
&\quad \left. + \sum_{i=1}^{N_T-1} \sum_{t=l_i+1}^{l_{i+1}-1} e^{-\frac{\left(x_{l_i}-\theta t\right)-\mu}{\beta}} \right].
\end{aligned}
$$

Ainsi, il faut trouver les estimateurs $\hat{\theta}$, $\hat{\mu}$ et $\hat{\beta}$ de θ, μ et β respectivement tel que,

$$
\left(\frac{\partial \log L(\theta,\mu,\beta)}{\partial\theta} \right)_{\substack{\theta=\hat{\theta} \\ \mu=\hat{\mu} \\ \beta=\hat{\beta}}} = \left(\frac{\partial \log L(\theta,\mu,\beta)}{\partial\mu} \right)_{\substack{\theta=\hat{\theta} \\ \mu=\hat{\mu} \\ \beta=\hat{\beta}}} = \left(\frac{\partial \log L(\theta,\mu,\beta)}{\partial\beta} \right)_{\substack{\theta=\hat{\theta} \\ \mu=\hat{\mu} \\ \beta=\hat{\beta}}} = 0.
$$

Où,

$$\begin{aligned}
\frac{\partial \log L(\theta,\mu,\beta)}{\partial \theta} &= \frac{1}{\beta}\sum_{k=1}^{N_T} l_k - \frac{1}{\beta}\sum_{k=1}^{N_T} l_k e^{-\frac{\left(x_{l_k}-\theta l_k\right)-\mu}{\beta}} - \frac{1}{\beta}\sum_{i=1}^{N_T-1}\sum_{t=l_i+1}^{l_{i+1}-1} t e^{-\frac{\left(x_{l_i}-\theta t\right)-\mu}{\beta}}.\\
\frac{\partial \log L(\theta,\mu,\beta)}{\partial \mu} &= \frac{N_T}{\beta} - \frac{1}{\beta}\sum_{k=1}^{N_T} e^{-\frac{\left(x_{l_k}-\theta l_k\right)-\mu}{\beta}} - \frac{1}{\beta}\sum_{i=1}^{N_T-1}\sum_{t=l_i+1}^{l_{i+1}-1} e^{-\frac{\left(x_{l_i}-\theta t\right)-\mu}{\beta}}.\\
\frac{\partial \log L(\theta,\mu,\beta)}{\partial \beta} &= -\frac{N_T}{\beta} + \frac{1}{\beta^2}\sum_{k=1}^{N_T}\left[\left(x_{l_k}-\theta l_k\right)-\mu\right] - \frac{1}{\beta^2}\sum_{k=1}^{N_T}\left[\left(x_{l_k}-\theta l_k\right)-\mu\right] e^{-\frac{\left(x_{l_k}-\theta l_k\right)-\mu}{\beta}}\\
&\quad -\frac{1}{\beta^2}\sum_{i=1}^{N_T-1}\sum_{t=l_i+1}^{l_{i+1}-1}\left[\left(x_{l_i}-\theta t\right)-\mu\right] e^{-\frac{\left(x_{l_i}-\theta t\right)-\mu}{\beta}}.
\end{aligned}$$

Ce qui doit se faire numériquement.

Maintenant, afin d'étudier la qualité des estimateurs obtenus par cette méthode d'estimation, 5000 séries chronologiques selon un modèle LDM pour différentes valeurs de T, $\theta = 0.25$ et une distribution sous-jacente d'une loi de $G(\mu = 0, \beta = 1)$ ont été générées. La Table 6.6 donne les biais empiriques et les écarts-types exacts approximés par simulation $\sqrt{\hat{\mathbb{V}}\left(\hat{\theta}\right)}$, $\sqrt{\hat{\mathbb{V}}\left(\hat{\mu}\right)}$ et $\sqrt{\hat{\mathbb{V}}\left(\hat{\beta}\right)}$ de nos estimateurs.

On remarque que les estimateurs des paramètres du modèle LDM semblent de bonne qualité, sur le plan du biais et de l'écart-type (sauf l'estimateur du paramètre de localisation μ qui s'améliore cependant avec l'augmentation du nombre des observations). Donc, l'utilisation des valeurs de records en plus que les indices de records, nous donne plus d'informations afin d'estimer les paramètres de la loi sous-jacente de Gumbel, ce qui était impossible en utilisant uniquement les indicatrices ou les indices de records.

6.4 Au delà de Gumbel

Nous explorons brièvement l'utilisation de la méthode précédente pour tenter d'estimer les paramètres de la distribution sous-jacente d'un modèle

	θ	μ	β
Biais	$\simeq 0.00$	-0.06	$\simeq 0.00$
Écart type	0.0086	0.5502	0.1708

(a) $T = 100$ observations

	θ	μ	β
Biais	$\simeq 0.00$	-0.05	$\simeq 0.00$
Écart-type	0.0096	0.3854	0.1331

(b) $T = 200$ observations

	θ	μ	β
Biais	$\simeq 0.00$	-0.05	$\simeq 0.00$
Écart-type	0.0103	0.2950	0.1233

(c) $T = 300$ observations

TABLE 6.6 – Biais et écarts-types de $\hat{\theta}$, $\hat{\mu}$ et $\hat{\beta}$ pour différentes valeurs du nombre d'observations T

LDM de distribution sous-jacente autre que la loi de Gumbel, mais en restant sur une fonction de vraisemblance construite à partir d'une distribution de loi de Gumbel de paramètres inconnus (Équation (6.3.1)). Il s'agit ici d'un problème dit de « misspecification » où l'utilisateur a correctement choisi le modèle LDM mais incorrectement choisi la loi sous-jacente. L'intérêt ici est de voir si l'on peut néanmoins retrouver certains des paramètres inconnus.

Nous considérons le problème d'estimation des paramètres d'un modèle LDM dans le cas d'une distribution sous-jacente de loi de $Weibull\,(\mu, \beta, \sigma)$, en se basant sur le principe du maximum de vraisemblance, où la fonction de vraisemblance est construite à l'aide de la densité de la loi $G\,(\mu, \beta)$.

Pour étudier le comportement des estimateurs de la section précédente, 5000 séries chronologiques selon un modèle LDM pour différentes valeurs de T, $\theta = 0.25$ et une distribution sous-jacente de loi de $Weibull\left(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\right)$ ont été générées. Puis, on a injecté les données générées dans la fonction de vraisemblance (6.3.1) construite à partir de la densité de la loi $G\,(\mu, \beta)$. Enfin, on maximise notre fonction de vraisemblance afin d'obtenir des estimateurs de θ, μ et β. La Table 6.7 donne les biais et les écarts-types approximés par simulations des estimateurs pour $T = 100$, 200 et 500 observations respectivement.

	θ	μ	β
Biais	$\simeq 0.00$	0.41	-0.18
Écart-type	0.0091	0.5986	0.2268

(a) $T = 100$ observations

	θ	μ	β
Biais	$\simeq 0.00$	0.41	-0.16
Écart-type	0.0031	0.4175	0.1699

(b) $T = 200$ observations

	θ	μ	β
Biais	$\simeq 0.00$	0.42	-0.15
Écart-type	0.0008	0.2445	0.1028

(c) $T = 500$ observations

TABLE 6.7 – Biais et écarts-types de $\hat{\theta}$, $\hat{\mu}$ et $\hat{\beta}$ pour différentes valeurs du nombre d'observations T, une distribution sous-jacente de loi de $Weibull\left(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\right)$ et $\theta = 0.25$ en utilisant le principe du maximum de vraisemblance appliqué sur l'Équation (6.3.1) construite à partir d'une distribution de loi $G(\mu, \beta)$.

Ainsi, même avec une mauvaise vraisemblance (données Weibull, modèle supposé Gumbel) le drift θ est estimable et l'estimateur semble de bonne qualité sur le plan du biais et de l'écart-type. Par contre, le paramètre de position μ et le paramètre d'échelle β ne sont pas estimables dans le sens que la méthode d'estimation ne semble pas donner un estimateur convergent : même si l'écart-type diminue, il ne semble pas que le biais disparaisse.

6.5 Tests statistiques

Dans cette section, nous introduisons des tests statistiques basés sur les vraisemblances (6.2.10) et (6.2.15) obtenues en utilisant la suite des couples $\{(R_n, L_n), n \geq 1\}$. En premier lieu, notre but est de vérifier la conformité du choix de la distribution sous-jacente dans le cas des modèles LDM et Yang respectivement. En second lieu, nous présentons un test qui nous aide à choisir entre un modèle LDM et de Yang. De plus, nous évaluons les différents tests statistiques par des simulations numériques sous Matlab en calculant les quantiles empiriques sous H_0 et les variations de la puissance du test pour différentes valeurs de θ et γ.

6.5.1 Cas LDM

Dans le cadre d'un modèle LDM de drift $\theta > 0$, notre but est de tester l'hypothèse $H_0 =$ « la distribution sous-jacente suit la loi A avec un drift θ inconnu » contre $H_1 =$ « la distribution sous-jacente suit la loi B avec un drift θ inconnu » en utilisant un test du rapport de vraisemblances maximales. Pour faciliter le lien avec les sections précédentes on considère par la suite que la loi A est la loi $G(\mu_0, \beta_0)$ et la loi B est la $Weibull(\mu_1, \beta_1, \sigma_1)$, mais toute autre paire de loi sous H_0 et H_1 aurait pu être choisie.

Pour ce faire, on maximise la vraisemblance (6.2.10), basée sur les (R_n, L_n) où on rappelle que $R_n = X_{L_n}$, sous H_0 (c'est à dire calculer $\max_\theta L^{Gumbel}(\theta)$) et sous H_1 (c'est à dire calculer $\max_\theta L^{Weibull}(\theta)$) et de faire le quotient des deux. La statistique du test qui va servir à tester l'hypothèse est la variable aléatoire :

$$K_T^{LDM} = 2\log\left(\frac{\max_\theta L^{Weibull}(\theta)}{\max_\theta L^{Gumbel}(\theta)}\right),$$

où T désigne le temps présent. Cette statistique est appelée la déviance.

La valeur de cette statistique est comparée au quantile empirique d'ordre $(1-\alpha)$ de K_T^{LDM}, notée $K_{T,1-\alpha}^{LDM}$. Quand $K_T^{LDM} > K_{T,1-\alpha}^{LDM}$, le test rejette, au niveau asymptotique α, l'hypothèse H_0. Ainsi, la distribution sous-jacente du modèle LDM ne serait pas en accord avec la loi de Gumbel.

Cette loi K_T^{LDM} étant inconnue, afin d'approcher ses quantiles sous H_0, 5000 séries chronologiques selon un modèle LDM pour différentes valeurs de θ et de T avec une distribution sous-jacente $G(0,1)$ ont été générées. On injecte chacune de ces séries dans les deux vraisemblances $L^{Gumbel}(\theta)$ et $L^{Weibull}(\theta)$ en considérant une loi de $Weibull(\mu{=}\varpi - \frac{\pi}{\sqrt{6}},\beta{=}\frac{\pi}{\sqrt{6}},\sigma{=}1)$. On rappelle que le choix des paramètres de la distribution sous-jacente vient du fait que la distribution $G(0,1)$ a la même espérance et le même écart-type que $Weibull\Big(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\Big)$, où $\varpi = 0.5772$ la constante d'Euler-Mascheroni. Ensuite, on maximise les deux fonctions de vraisemblances par rapport à θ. Finalement, on calcule la statistique K_T^{LDM} afin d'obtenir empiriquement les quantiles de la loi de K_T^{LDM} sous H_0. D'autre part, pour calculer la puissance du test, 5000 séries chronologiques selon un modèle LDM pour différentes valeurs de θ et de T avec une distribution sous-jacente de $Weibull\left(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\right)$ ont été générées. On injecte chacune de ces séries dans les deux vraisemblances $L^{Gumbel}(\theta)$ et $L^{Weibull}(\theta)$ et on maximise les deux fonctions de vraisemblances par rapport à θ afin de calculer K_T^{LDM}. En comptant le nombre de fois que $K_T^{LDM} > K_{T,1-\alpha}^{LDM}$ (rejet de H_0), on obtient la puissance du test. La Table 6.8 représente les quantiles empiriques, sous H_0, d'ordre $(1-\alpha)$ de K_T^{LDM} pour différentes valeurs de θ et de T avec $\alpha = 5\%$. De plus, la Figure 6.1 représente les variations de la puissance du test pour différentes valeurs de T. On remarque bien évidemment que la puissance du test s'améliore avec l'augmentation de θ et de T.

6.5.2 Cas de Yang

Maintenant, en suivant la même méthode qu'à la sous-section précédente mais dans le cas d'un modèle de Yang de paramètre $\gamma > 1$, nous avons considéré le problème de tester l'hypothèse $H_0 =$ « la distribution sous-jacente suit la loi de $G(\mu_0, \sigma_0)$ avec un paramètre γ inconnu » contre $H_1 =$ « la distribution sous-jacente suit la loi de $Weibull(\mu_1, \sigma_1, \lambda_1)$ avec un paramètre γ inconnu » en utilisant aussi un test basé sur la déviance. Encore là, le choix des densités sous H_0, H_1 aurait pu être différent sans que cela change

	$K^{LDM}_{T,95\%}$			
θ	$T = 25$	$T = 50$	$T = 100$	$T = 200$
0.05	2.30	2.12	1.96	2.21
0.10	2.25	2.09	2.07	2.26
0.15	2.25	2.10	2.31	2.74
0.20	2.27	2.10	2.40	2.72
0.25	2.42	2.52	2.71	2.93
0.50	3.59	4.14	4.37	3.96

TABLE 6.8 – Quantiles empiriques, sous H_0, d'ordre $(1-\alpha)$ de K^{LDM}_T pour différentes valeurs de θ et de T avec $\alpha = 5\%$

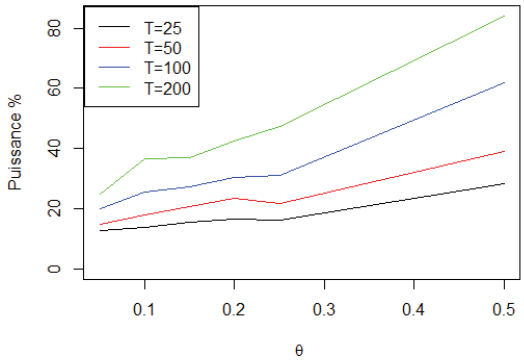

FIGURE 6.1 – Variations de la puissance du test pour différentes valeurs de T (cas LDM)

la stratégie de test. Ainsi, il suffit de maximiser la vraisemblance (6.2.15), basée sur les (R_n, L_n) où on rappelle que $R_n = X_{L_n}$, sous H_0 (c'est à dire calculer $\max_{\gamma} L^{Gumbel}(\gamma)$) et sous H_1 (c'est à dire calculer $\max_{\gamma} L^{Weibull}(\gamma)$) et de faire le quotient des deux. La statistique du test qui va servir à tester l'hypothèse est la variable aléatoire :

$$K_T^{Yang} = 2\log\left(\frac{\max_{\gamma} L^{Weibull}(\gamma)}{\max_{\gamma} L^{Gumbel}(\gamma)}\right).$$

La valeur de cette statistique est comparée au quantile empirique d'ordre $(1-\alpha)$ de K_T^{Yang}, notée $K_{T,1-\alpha}^{Yang}$. Quand $K_T^{Yang} > K_{T,1-\alpha}^{Yang}$, le test rejette, au niveau asymptotique α, l'hypothèse H_0. Ainsi, la distribution sous-jacente du modèle de Yang ne serait pas en accord avec la loi de Gumbel.

Pour étudier numériquement le comportement de cette stratégie de test, on applique la procédure de la section précédente mais en générant des séries chronologiques selon un modèle de Yang. La Table 6.9 représente les quantiles empiriques, sous H_0, d'ordre $(1-\alpha)$ de K_T^{Yang} pour différentes valeurs de γ et de T avec $\alpha = 5\%$. Le « NA » dans la Table 6.9 signifie que la réponse n'est pas disponible. En fait, dans un modèle de Yang et pour des grandes valeurs de T et de γ la série des observations simulée $\{X_t, t \geq 1\}$ prend des valeurs infinies ce qui nous ne permet pas d'aller plus loin dans l'analyse des simulations. Notons que normalement le modèle de Yang avec une distribution sous-jacente de Gumbel coïncide avec le LDM avec cette même loi sous-jacente. Ainsi, en comparant les quantiles sous H_0 obtenus ici avec ceux de la section précédente, on remarque qu'asymptotiquement les deux séries de quantiles se rapprochent. De plus, la Figure 6.2 représente les variations de la puissance du test pour différentes valeurs de T. On remarque aussi évidemment que la puissance du test s'améliore avec l'augmentation de γ et de T.

6.5.3 LDM vs Yang

Enfin, la bonne pratique statistique commande que l'on dispose d'un moyen permettant de discriminer entre les modèles LDM et de Yang. Ici, notre but est de développer une stratégie de test de H_0 = « Le modèle est LDM avec une distribution sous-jacente de loi A avec un drift θ inconnu » contre H_1 = « Le modèle est de Yang avec une distribution sous-jacente de

	$K^{Yang}_{T,95\%}$			
γ	$T = 25$	$T = 50$	$T = 100$	$T = 200$
1.05	2.01	1.99	1.95	2.34
1.10	2.01	2.01	2.16	2.61
1.15	2.01	2.19	2.49	2.60
1.20	2.01	2.37	2.57	*NA*
1.25	2.17	2.25	2.39	*NA*
1.50	2.17	2.41	*NA*	*NA*

TABLE 6.9 – Quantiles empiriques, sous H_0, d'ordre $(1-\alpha)$ de K_T^{Yang} pour différentes valeurs de γ et de T avec $\alpha = 5\%$

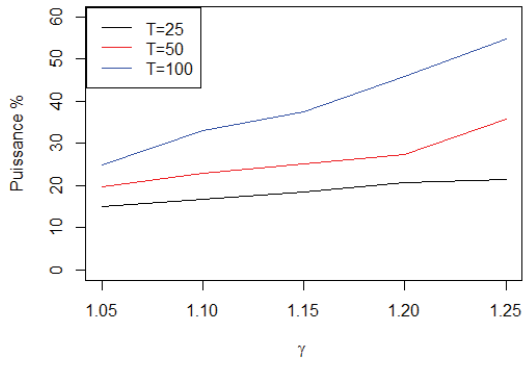

FIGURE 6.2 – Variations de la puissance du test pour différentes valeurs de T (cas Yang)

loi A avec un paramètre γ inconnu » en utilisant un test basé sur la déviance. Ici nous prendrons la loi A comme étant la $Weibull\,(\mu_0, \beta_0, \sigma_0)$ pour faciliter le lien avec les sections précédentes, mais toute autre loi aurait été possible. Il suffit alors de maximiser la vraisemblance (6.2.10) sous H_0 (c'est à dire calculer $\max_{\theta} L^{Weibull}(\theta)$) et de maximiser la vraisemblance (6.2.15) sous H_1 (c'est à dire calculer $\max_{\gamma} L^{Weibull}(\gamma)$) et de faire le quotient des deux. La statistique du test qui va servir à tester l'hypothèse est la variable aléatoire :

$$K_T = 2\log\left(\frac{\max_{\gamma} L^{Weibull}(\gamma))}{\max_{\theta} L^{Weibull}(\theta)}\right).$$

La valeur de cette statistique est comparée au quantile empirique d'ordre $(1-\alpha)$ de K_T, notée $K_{T,1-\alpha}$. Quand $K_T > K_{T,1-\alpha}$, le test rejette, au niveau asymptotique α, l'hypothèse H_0. Ainsi, le modèle adapté à notre série de records n'est pas en accord avec le modèle LDM. Évidemment, par une simple inversion du quotient dans l'expression de K_T, on peut tester l'hypothèse H_0^* d'un modèle de Yang contre H_1^* : le bon modèle est le LDM.

Il reste à déterminer la loi, exacte ou asymptotique, de la statistique K_T sous H_0. L'une ou l'autre étant à ce stade inaccessible, on procède par simulation de Monte-Carlo. On a généré 5000 séries chronologiques selon un modèle LDM pour différentes valeurs de θ et de T avec une distribution sous-jacente de $Weibull\left(\mu = \varpi - \frac{\pi}{\sqrt{6}}, \beta = \frac{\pi}{\sqrt{6}}, \sigma = 1\right)$. On injecte chacune de ces séries dans les deux vraisemblances $L^{Weibull}(\theta)$ et $L^{Weibull}(\gamma)$ que l'on maximise par rapport à θ et γ respectivement. Finalement, on calcule la statistique K_T afin d'obtenir empiriquement les quantiles de la loi de K_T sous H_0. D'autre part, pour calculer la puissance du test, 5000 séries chronologiques selon un modèle de Yang pour différentes valeurs de γ et de T avec une distribution sous-jacente de $Weibull\left(\varpi - \frac{\pi}{\sqrt{6}}, \frac{\pi}{\sqrt{6}}, 1\right)$ ont été générées. On injecte chacune de ces séries dans les deux vraisemblances $L^{Weibull}(\theta)$ et $L^{Weibull}(\gamma)$ et on maximise les deux fonctions de vraisemblances par rapport à θ et γ respectivement afin de calculer la statistique K_T. En comptant le nombre de fois que $K_T > K_{T,1-\alpha}$ (rejet de H_0), on obtient la puissance du test. La Table 6.8 représente les quantiles empiriques, sous H_0, d'ordre $(1-\alpha)$ de K_T pour différentes valeurs de θ et de T avec $\alpha = 5\%$. De plus, la Figure 6.1 représente les variations de la puissance du test pour différentes valeurs de T. On remarque encore une fois que la puissance du test s'améliore avec

	$K_{T,95\%}$			
θ	$T = 25$	$T = 50$	$T = 100$	$T = 200$
0.05	0.90	0.69	0.51	0.46
0.10	0.94	0.81	0.69	0.84
0.15	0.99	0.92	1.05	1.37
0.20	1.05	1.08	1.29	1.65
0.25	1.10	1.10	1.37	*NA*
0.50	1.16	0.89	*NA*	*NA*

TABLE 6.10 – Quantiles empiriques, sous H_0, d'ordre $(1-\alpha)$ de K_T pour différentes valeurs de θ et de T avec $\alpha = 5\%$

l'augmentation de θ et de T.

Remarque 6.5. L'article de Ducharme et Frichot (2003) présente le comportement asymptotique des différentes statistiques de tests des Sections 6.5.1, 6.5.2 et 6.5.3 dans le cas standard où les données sont *iid*. Les lois limites de versions centrés réduites des statistiques de tests plus haut dans ce cas sont des $N(0,1)$. Il semble assez compliqué d'obtenir ces comportements asymptotiques dans le présent cas de records.

6.6 Conclusion

Nous avons présenté plusieurs méthodes d'estimation des paramètres d'un modèle LDM avec une distribution sous-jacente de paramètres inconnus, en utilisant le principe du maximum de vraisemblance.

D'abord, en se basant uniquement sur les indicatrices de records, on a vu qu'on n'avait pas assez d'informations pour bien estimer tous les paramètres d'un modèle LDM. Par suite, le passage à l'utilisation des couples des données disponibles $\{(R_n, L_n), n \geq 1\}$ devient indispensable.

Ensuite, nous avons supposé que les données disponibles sont les indices et les valeurs de records. Nous avons présenté une nouvelle méthode d'estimation des paramètres des modèles LDM et Yang basée sur la propriété Markovienne de la suite des couples (R_n, L_n) et sur les travaux de Carlin et Gelfand (1993). En utilisant cette nouvelle méthode d'estimation, on a obtenu des estimateurs de bonne qualité, sur le plan du biais et de l'écart-type. De plus, l'utilisation de la totalité des données disponibles nous a aidé à dépasser

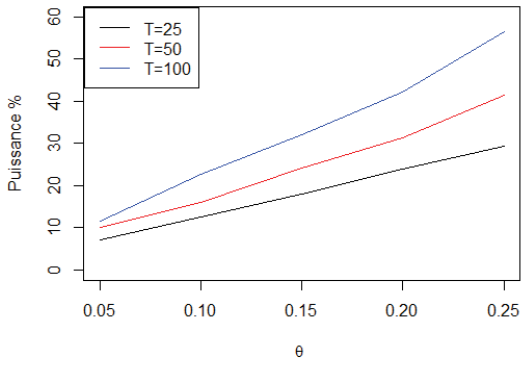

FIGURE 6.3 – Variations de la puissance du test pour différentes valeurs de T (cas LDM vs Yang)

la contrainte d'une distribution sous-jacente de Gumbel dans un modèle LDM. Nous remarquons aussi que indépendamment de la distribution sous-jacente, la qualité des estimateurs s'améliore avec l'augmentation du nombre de records, surtout au niveau de l'écart-type. D'autre part, nous avons aussi utilisé cette vraisemblance plus complète pour estimer les paramètres d'un modèle LDM de distribution sous-jacente autre que Gumbel, mais en restant sur une fonction de vraisemblance construite à partir d'une distribution de Gumbel et nous avons étudié empiriquement leur comportement.

Donc l'utilisation des valeurs de records en plus que les indices de records, nous donne plus d'informations sur les paramètres à estimer, ce qui implique l'obtention des estimateurs de meilleure qualité.

Enfin, nous avons introduit des tests statistiques basés sur les vraisemblances (6.2.10) et (6.2.15) obtenues en utilisant la suite des couples $\{(R_n, L_n), n \geq 1\}$. En premier lieu, nous avons vérifié la conformité du choix de la distribution sous-jacente dans le cas des modèles LDM et Yang respectivement. En second lieu, nous avons présenté un test qui nous aide à choisir entre un modèle LDM et de Yang. D'après les simulations numériques sous Matlab on a remarqué que la puissance des tests s'améliore avec l'augmentation de θ, γ et T.

Chapitre 7

Conclusion et perspectives

7.1 Conclusion

Dans cette thèse nous avons présenté ce que sont les records et pourquoi leur étude est d'une certaine importance en pratique. De plus, nous avons développé le comportement stochastique des différents modèles adaptés aux séries de records en se concentrant sur le cas d'une série de records issus d'observations indépendantes mais non identiquement distribuées. Dans ce contexte, les modèles de records les plus populaires sont le modèle à dérive linéaire (LDM) et le modèle de Yang-Nevzorov.

Sachant que dans une série de records, les données disponibles sont les valeurs $\{R_n, n \geq 1\}$, indices $\{L_n, n \geq 1\}$ et indicatrices $\{\delta_t, t \geq 1\}$ de records, nous avons présenté plusieurs méthodes d'estimation du drift θ d'un modèle LDM avec une distribution sous-jacente de Gumbel de paramètres connus et du paramètre γ d'un modèle de Yang avec une distribution sous-jacente quelconque aussi de paramètres connus , en se basant uniquement sur le nombre N_T, les indicatrices δ_t et les indices L_n de records respectivement et en utilisant la méthode des moments ou le principe du maximum de vraisemblance. Le meilleur estimateur sur les plans considérés (biais, écart-type et probabilité de couverture) a été obtenu en utilisant les indicatrices de records et en se basant sur le principe du maximum de vraisemblance. On a aussi donné la preuve du comportement asymptotique normal des différents estimateurs obtenus.

L'importance du modèle de Yang-Nevzorov vient du fait que les δ_t sont indépendants pour toutes les distributions sous-jacentes possibles et que le

modèle de Yang coïncide avec le LDM dans le cas d'une distribution sous-jacente de Gumbel. De plus, ce modèle est intéressant car il a la structure d'un modèle à risque proportionnel en analyse de survie, lequel a montré son utilité afin de modéliser de nombreux jeux de données.

En outre, et toujours dans le contexte d'un modèle de Yang, nous avons montré que le comportement asymptotique de la distribution du temps inter-records $\Delta_{L_n} = L_{n+1} - L_n$ est géométrique. Puis, en se basant sur cette propriété, nous avons développé deux tests d'adéquation pour le modèle de Yang. D'autre part, nous avons construit un test statistique, basé sur le nombre de records N_T, qui évalue la nécessité d'un passage à un modèle non *iid* afin de modéliser certaines séries de records.

Nous avons appliqué nos résultats théoriques à des données analysées précédemment par Yang (1975) présentant les records de la course de 200 mètres dans les Jeux olympiques. Les différents tests statistiques ont montré que le passage au delà d'un modèle *iid* est nécessaire et que l'hypothèse d'un modèle de Yang n'est pas déraisonnable pour ce type de données. Et en appliquant la méthode d'estimation basée sur les indicatrices de records et le principe du maximum de vraisemblance afin d'estimer le paramètre γ du modèle de Yang, nous avons obtenu un estimateur qui colle mieux que celui proposé par Yang (1975) avec les données réelles.

Dans le dernier chapitre, nous avons montré qu'en se basant uniquement sur les indicatrices de records, on n'a pas assez d'informations pour bien estimer tous les paramètres d'un modèle LDM. Par suite, le passage à l'utilisation des couples des données disponibles $\{(R_n, L_n), n \geq 1\}$ devient indispensable. Ensuite, Nous avons présenté une nouvelle méthode d'estimation des paramètres des modèles LDM et Yang basée sur la propriété Markovienne de la suite des couples (R_n, L_n) et sur les travaux de Carlin et Gelfand (1993). En utilisant cette nouvelle méthode d'estimation, on a obtenu des estimateurs de bonne qualité, sur le plan du biais et de l'écart-type. De plus, l'utilisation de la totalité des données disponibles nous a aidé à dépasser la contrainte d'une distribution sous-jacente de Gumbel dans un modèle LDM. D'autre part, nous avons aussi utilisé cette vraisemblance plus complète pour estimer les paramètres d'un modèle LDM de distribution sous-jacente autre que Gumbel, mais en restant sur une fonction de vraisemblance construite à partir d'une distribution de Gumbel et nous avons étudié empiriquement leur comportement.

Enfin, nous avons introduit des tests statistiques basés sur les fonctions de vraisemblances obtenues en utilisant la suite des couples $\{(R_n, L_n), n \geq 1\}$. En

premier lieu, nous avons vérifié la conformité du choix de la distribution sous-jacente dans le cas des modèles LDM et Yang respectivement. En second lieu, nous avons présenté un test qui nous aide à choisir entre un modèle LDM et de Yang. On remarque bien évidemment que la puissance du test s'améliore avec l'augmentation de θ, γ et T.

7.2 Perspectives

Les résultats des simulations du Chapitre 6 montrent qu'il n'est pas déraisonnable de conjecturer que le comportement asymptotique des estimateurs $\hat{\theta}_5$ et $\hat{\gamma}_4$, obtenus en utilisant le principe du maximum de vraisemblance et la propriété Markovienne des couples (R_n, L_n), est normal de variance l'inverse de la dérivée seconde des fonctions log-vraisemblances. C'est à dire

$$\frac{\hat{\theta}_5 - \theta}{\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\theta)}{\mathrm{d}\theta^2}\right)^{-1}}} \sim N(0,1),$$

et

$$\frac{\hat{\gamma}_4 - \gamma}{\sqrt{-\left(\frac{\mathrm{d}^2 \log L(\gamma)}{\mathrm{d}\gamma^2}\right)^{-1}}} \sim N(0,1).$$

Jusqu'à présent on n'a pas encore la théorie permettant de donner les preuves des conjectures, surtout que les (R_n, L_n) forment une chaîne de Markov de sorte que les théorèmes standards concernant le comportement des estimateurs de vraisemblance maximale ne s'appliquent pas directement. Ainsi, une première perspective de cette thèse est d'essayer de développer une preuve théorique du comportement asymptotique de ces deux estimateurs. Pour débuter ce travail nous suggérons l'utilisation de la méthode de « one step estimator » basée sur les différentes fonctions de vraisemblances du Chapitre 6 et en prenant $\hat{\theta}_3$ et $\hat{\gamma}_3$ comme point de départ de l'algorithme itérative.

Une deuxième perspective, est aussi d'étudier le comportement asymptotique des différents estimateurs obtenus dans le contexte du problème de « misspecification » de la Section 6.4. Jusqu'à présent on n'a pas des idées claires sur la théorie permettant d'estimer les écarts-types de ces estimateurs.

De plus, une troisième perspective est d'étudier le comportement asymptotique exact des différentes statistiques de tests de la Section 6.5 en se basant sur les travaux de Ducharme et Frichot (2003) qui font l'analyse dans le cas standard où les données sont *iid.* Aussi, pour suppléer à la théorie manquante sur le comportement asymptotique des statistiques de tests, une perspective possible serait la méthode du bootstrap. Mais encore là, la preuve que cette approche marche dans le présent contexte non-standard pourrait être délicate et doit faire l'objet de travaux futurs.

Enfin, nous cherchons à appliquer nos résultats théoriques à des données records dans plusieurs domaines, surtout pour des phénomènes naturels, afin de calculer la probabilité d'un prochain record.

Productions scientifiques

- Juillet 2016 : Article soumis au journal «Extremes», intitulé : « Distribution-Free Inference in Record Series ».
- Juin 2016 : J'ai assisté à la conférence « Les 48e Journées de Statistique (JdS2016) » tenue à Montpellier, et j'ai présenté un exposé intitulé : « Estimation des paramètres pour des modèles adaptés aux séries de records ».
- Mai 2015 : J'ai assisté à la conférence « Lebanese International Conference on Mathematics and Applications LICMA'2015 » tenue à Beyrouth, et j'ai présenté un exposé intitulé : « Records in classical model and in the presence of a linear drift ».

Bibliographie

Barry C Arnold, Narayanaswamy Balakrishnan et Haikady Navada Nagaraja : *Records*. John Wiley & Sons, New York, 1998.

Rocco Ballerini et Sidney Resnick : Records from improving populations. *Journal of Applied Probability*, pages 487–502, 1985.

Rocco Ballerini et Sidney I Resnick : Embedding sequences of successive maxima in extremal processes, with applications. *Journal of Applied Probability*, pages 827–837, 1987.

Yvonne M Bishop, Stephen E Fienberg et Paul W Holland : *Discrete multivariate analysis : theory and practice*. Springer Science & Business Media, New York, 2007.

K Borovkov : On records and related processes for sequences with trends. *Journal of Applied Probability*, 36:668–681, 1999.

Bradley P Carlin et Alan E Gelfand : Parametric likelihood inference for record breaking problems. *Biometrika*, 80:507–515, 1993.

KN Chandler : The distribution and frequency of record values. *Journal of the Royal Statistical Society. Series B*, pages 220–228, 1952.

A Charalambides, Ch et Jagbir Singh : Review of the stirling numbers, their generalizations and statistical applications. *Communications in Statistics-Theory and Methods*, 17:2507–2532, 1988.

Gilles R Ducharme et Benoît Frichot : Quasi most powerful invariant goodness-of-fit tests. *Scandinavian journal of statistics*, 30:399–414, 2003.

William Feller : *An Introduction to Probability Theory and Its Applications. Volume I.* John Wiley & Sons London-New York-Sydney-Toronto, 1968.

William Feller : *An Introduction to Probability Theory and Its Applications. Volume II. second edition.* Wiley, New York, 1971.

Andrey Feuerverger et Peter Hall : On distribution-free inference for record-value data with trend. *The Annals of Statistics*, pages 2655–2678, 1996.

Jasper Franke, Gregor Wergen et Joachim Krug : Records and sequences of records from random variables with a linear trend. *Journal of Statistical Mechanics : Theory and Experiment*, 2010:10013, 2010.

Emil Julius Gumbel et Chandan K Mustafi : Some analytical properties of bivariate extremal distributions. *Journal of the American Statistical Association*, 62(318):569–588, 1967.

Godfrey Harold Hardy : *Divergent series.* Oxford University Press, London, 1949.

Norbert Henze et Bernhard Klar : Properly rescaled components of smooth tests of fit are diagnostic. *Australian Journal of Statistics*, 38:61–74, 1996.

Z Khraibani : *Risque d'emergence d'une pathologie dans une population : Evaluation a l'aide d'une approche par processus extreme*. These sous la direction de Christine Jacob- Paris 11, Paris, 2008.

Zaher Khraibani, Christine Jacob, Christian Ducrot, Myriam Charras-Garrido et Carole Sala : A non parametric exact test based on the number of records for an early detection of emerging events : Illustration in epidemiology. *Communications in Statistics-Theory and Methods*, 44:726–749, 2015.

Fanny Leroy, Jean-Yves Dauxois et Pascale Tubert-Bitter : On the parametric maximum likelihood estimator for independent but non-identically distributed observations with application to truncated data. *Journal of Statistical Theory and Applications*, 15:96–107, 2016.

Joseph Meixner : Orthogonale polynomsysteme mit einer besonderen gestalt der erzeugenden funktion. *Journal of the London Mathematical Society*, 1:6–13, 1934.

Valery B Nevzorov : *Records : mathematical theory*. American Mathematical Society, Rhode Island, 2001.

VB Nevzorov : Records for nonidentically distributed random variables. *Proceedings of the Fifth Vilnius Conference*, 2:227–233, 1990.

VB Nevzorov et A Stepanov : Records with confirmation. *Statistics & Probability Letters*, 95:39–47, 2014.

Gary W Oehlert : A note on the delta method. *The American Statistician*, 46:27–29, 1992.

Karl Pearson : X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50:157–175, 1900.

R. L. Plackett : *The Analysis of Categorical Data*. Griffin, London, 1974.

Glen D Rayner et John CW Rayner : Power of the neyman smooth tests for the uniform distribution. *Advances in Decision Sciences*, 5(3):181–191, 2001.

JCW Rayner et DJ Best : *Smooth tests of goodness of fit*. Oxford University Press, New York, 1989.

RW Shorrock : On record values and record times. *Journal of Applied Probability*, pages 316–326, 1972.

E Slutsky : About stochastic asymptotes and limits. *Metron*, 5:3–89, 1925.

Richard L Smith : Forecasting records by maximum likelihood. *Journal of the American Statistical Association*, 83:331–338, 1988.

G Wergen : Records in stochastic processes, theory and applications. *Journal of Physics A : Mathematical and Theoretical*, 46:223001, 2013.

Gregor Wergen : Modeling record-breaking stock prices. *Physica A : Statistical Mechanics and its Applications*, 396:114–133, 2014.

Gregor Wergen, Miro Bogner et Joachim Krug : Record statistics for biased random walks, with an application to financial data. *Physical Review E*, 83:051109, 2011.

Gregor Wergen et Joachim Krug : Record-breaking temperatures reveal a warming climate. *Europhysics Letters*, 92:30008, 2010.

Mark CK Yang : On the distribution of the inter-record times in an increasing population. *Journal of Applied Probability*, pages 148–154, 1975.