

Théorie des valeurs extrêmes

M2 Data Science

Enseignant : Gilles Stupfler

Email : `gilles.stupfler@univ-angers.fr`

Département de Mathématiques, bureau I115

Printemps 2022/2023

Bibliographie et références complémentaires

- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, John Wiley & Sons, Chichester.
- Billingsley, P. (1995). *Probability and measure (third edition)*, John Wiley and Sons.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.
- Dekkers, A.L.M., Einmahl, J.H.J., de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Annals of Statistics* **17**(4): 1833–1855.
- Diebolt, J., Guillou, A., Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *Journal of Statistical Planning and Inference* **137**(3): 841–857.
- Drees, H., Ferreira, A., de Haan, L. (2004). On maximum likelihood estimation of the extreme value index, *Annals of Applied Probability* **14**(3): 1179–1201.
- Embrechts, P., Klüppelberg, C., Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.
- Grimshaw, S.D. (1993). Computing maximum likelihood estimates for the Generalized Pareto distribution, *Technometrics* **35**(2): 185–191.
- de Haan, L., Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer, New York.¹
- Hall, P. (1982). On estimating the endpoint of a distribution, *Annals of Statistics* **10**(2): 556–568.
- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution, *Annals of Statistics* **3**(5): 1163–1174.
- Hosking, J.R.M., Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics* **29**(3): 339–349.
- Pickands, J. (1975). Statistical inference using extreme order statistics, *Annals of Statistics* **3**(1): 119–131.
- Resnick, S.I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

¹Erratum partiel disponible à <https://personal.eur.nl/1dehaan/EVTbook.correction.pdf>

Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations, *Journal of the American Statistical Association* **73**(364): 812–815.

Zhou, C. (2009). Existence and consistency of the maximum likelihood estimator for the extreme value index, *Journal of Multivariate Analysis* **100**: 794–815.

Zhou, C. (2010). The extent of the maximum likelihood estimator for the extreme value index, *Journal of Multivariate Analysis* **101**: 971–983.

Table des matières

1	Modèles de valeurs extrêmes	1
1.1	Motivation	1
1.1.1	Exemple 1 : données d'assurance incendie	2
1.1.2	Exemple 2 : données de température	3
1.2	Vers des modèles de valeurs extrêmes : convergence du maximum et des excès	3
2	Estimation dans les modèles à queue lourde	9
2.1	Estimateur de quantile intermédiaire	10
2.2	Estimateur de Hill	11
2.3	Estimateur de Weissman d'un quantile extrême	15
2.4	Application sur données réelles	17
3	Estimation hors des modèles à queue lourde	20
3.1	Approche paramétrique	20
3.2	Application sur données réelles	24

Chapitre 1

Modèles de valeurs extrêmes

1.1 Motivation

Bien que la majorité des applications de la statistique vise à estimer et interpréter des quantités qu'on peut qualifier de « centrales » (une moyenne, une médiane, des quartiles...), certaines applications requièrent au contraire de se concentrer sur des paramètres ou indicateurs « extrêmes » de la distribution sous-jacente au phénomène étudié. Ces applications constituent la motivation de la théorie moderne des valeurs extrêmes.

La première telle application historiquement connue remonte à la fin des années 1920, avec le travail de Leonard Tippett dans le domaine de l'industrie textile. Tippett, dont la recherche portait sur la résistance des fils de coton, avait compris que la résistance d'un fil était directement liée à la résistance de ses fibres les plus faibles, ce qui l'a naturellement conduit à étudier spécifiquement la probabilité qu'une fibre de coton soit « faible ». Une autre application d'importance majeure, proposée par le gouvernement des Pays-Bas au cours des années 1950, est l'estimation de la hauteur des digues abritant les zones habitées sous le niveau de la mer afin de garantir leur protection contre des fortes tempêtes provoquant des inondations. Le gouvernement néerlandais proposait d'utiliser cette estimation afin de construire des digues qui ne pourraient être submergées qu'une fois tous les 10 000 ans en moyenne. La difficulté ici, comme dans tout problème d'étude de valeurs extrêmes, est la quantité de données disponibles : les archives contiennent des données sur quelques centaines d'années au maximum, ce qui semble bien insuffisant pour estimer la probabilité d'un événement se produisant en moyenne une fois tous les 10 000 ans !

La question de la compréhension des valeurs extrêmes d'un phénomène aléatoire se pose aussi en gestion de risque financier et actuariel. En assurance, la directive Solvabilité II de

l'Union Européenne impose qu'une compagnie d'assurance (ayant ses activités au sein de l'Union) calcule et provisionne une quantité de fonds suffisante pour couvrir ses obligations sur les 12 mois à venir avec probabilité 0.995. Dans ce type de cadre réglementaire, il est donc crucial pour une compagnie d'assurance d'analyser avec précision les distributions des indemnités qu'elle verse, et tout particulièrement le comportement probabiliste des « grandes » indemnités, afin d'éviter non seulement une sous-estimation du risque qui la mettrait en difficulté vis-à-vis du régulateur, mais aussi une sur-estimation de ce risque qui pourrait l'inciter à se montrer trop prudente et donc non compétitive.

1.1.1 Exemple 1 : données d'assurance incendie

On présente ici un exemple dans lequel les méthodes de théorie de valeurs extrêmes prennent tout leur sens. On utilise les données d'assurance incendie de l'entreprise Copenhagen Re, collectées entre 1980 et 1990 et ajustées (en raison de l'inflation) à leur niveau équivalent en 1985. Ces données sont disponibles dans plusieurs packages R, parmi lesquels `evir`¹, en appelant `data(danish)`. Le jeu de données, de taille $n = 2\,167$, est représenté sur la première ligne de la Figure 1.1. Cette représentation n'est que partielle : sur les 2 167 observations, 2 058 (soit 95%) sont inférieures à 10, mais 15 d'entre elles sont supérieures à 30, avec une observation maximale égale à 263.25. Ce comportement (une grande majorité de petites observations et une petite fraction d'observations d'un tout autre ordre de grandeur) est typique d'un jeu de données dit à queue lourde.

L'histogramme en haut à gauche de la Figure 1.1 suggère à première vue une distribution unimodale à décroissance rapide, peut-être exponentielle. Une famille de modèles raisonnable semble être la famille des lois de Weibull, dont la densité est donnée par :

$$\forall x > 0, f(x|\beta, \lambda) = \beta\lambda(\lambda x)^{\beta-1} \exp(-(\lambda x)^\beta), \text{ où } \beta, \lambda > 0.$$

On utilise la méthode du maximum de vraisemblance pour estimer β et λ , à l'aide de la fonction `fitdist` du package `fitdistrplus`. La densité correspondant au maximum de vraisemblance est superposée à l'histogramme des données sur le graphique en bas à gauche de la Figure 1.1, et on a représenté un diagramme quantile-quantile, ou *QQ-plot*, comparant les quantiles de cette distribution estimée avec les quantiles empiriques des données sur le dernier graphique, en bas à droite de la Figure 1.1.

Il est clair, d'après ces deux derniers graphiques, que si la loi de Weibull estimée donne une idée globalement raisonnable du comportement de la majorité des données, elle représente

¹Un des « grands » packages de statistique des valeurs extrêmes, avec, entre autres, `evd` et `extRemes`.

par contre très mal ses extrêmes. Le QQ-plot indique que les grandes observations de l'échantillon sont beaucoup plus grandes que celles qui seraient typiquement générées par cette loi de Weibull. En d'autres termes, ce jeu de données présente une queue plus lourde qu'une distribution classique à décroissance exponentielle. L'erreur qui serait faite en utilisant la loi de Weibull pour décrire les extrêmes de cet échantillon est potentiellement très importante : par exemple, au niveau 99%, le quantile de la loi de Weibull estimée est environ 15.88, alors que le quantile empirique des données est environ 25.07. Une prédiction des valeurs extrêmes en utilisant la loi de Weibull serait donc trop optimiste, ce qui peut s'avérer très préjudiciable en assurance.

1.1.2 Exemple 2 : données de température

Les applications de la théorie des valeurs extrêmes ne concernent pas seulement les données présentant une queue lourde, dont on verra qu'elles sont nécessairement non bornées. Pour illustrer cela, on utilise des données du KNMI, l'institut de météorologie des Pays-Bas, sur la température journalière moyenne enregistrée à l'aéroport d'Amsterdam-Schiphol. Ces données sont disponibles sur le site Web du KNMI² (station 240). Le jeu de données complet, de taille 25 202, est représenté à l'aide d'une courbe de température et d'un histogramme sur la première ligne de la Figure 1.2. On a également représenté cette même courbe et l'histogramme associé pour l'année 2019.

D'après ces deux histogrammes, il semble que la distribution des données possède deux modes ; il paraît raisonnable de proposer un modèle constitué d'un mélange de deux lois normales. Pour la prédiction de températures extrêmes (très élevées ou très basses), ce modèle n'est pourtant pas satisfaisant, ne serait-ce que parce qu'il ne peut pas donner une estimation des températures maximale et minimale physiquement possibles, puisque la loi normale n'est pas bornée !

1.2 Vers des modèles de valeurs extrêmes : convergence du maximum et des excès

Pour trouver des modèles appropriés à la description des grandes valeurs d'un échantillon, on va chercher à comprendre directement le comportement probabiliste de ces grandes valeurs. On se donne un échantillon (X_1, \dots, X_n) de variables aléatoires indépendantes

²Voir <https://www.knmi.nl/nederland-nu/klimatologie/daggegevens>

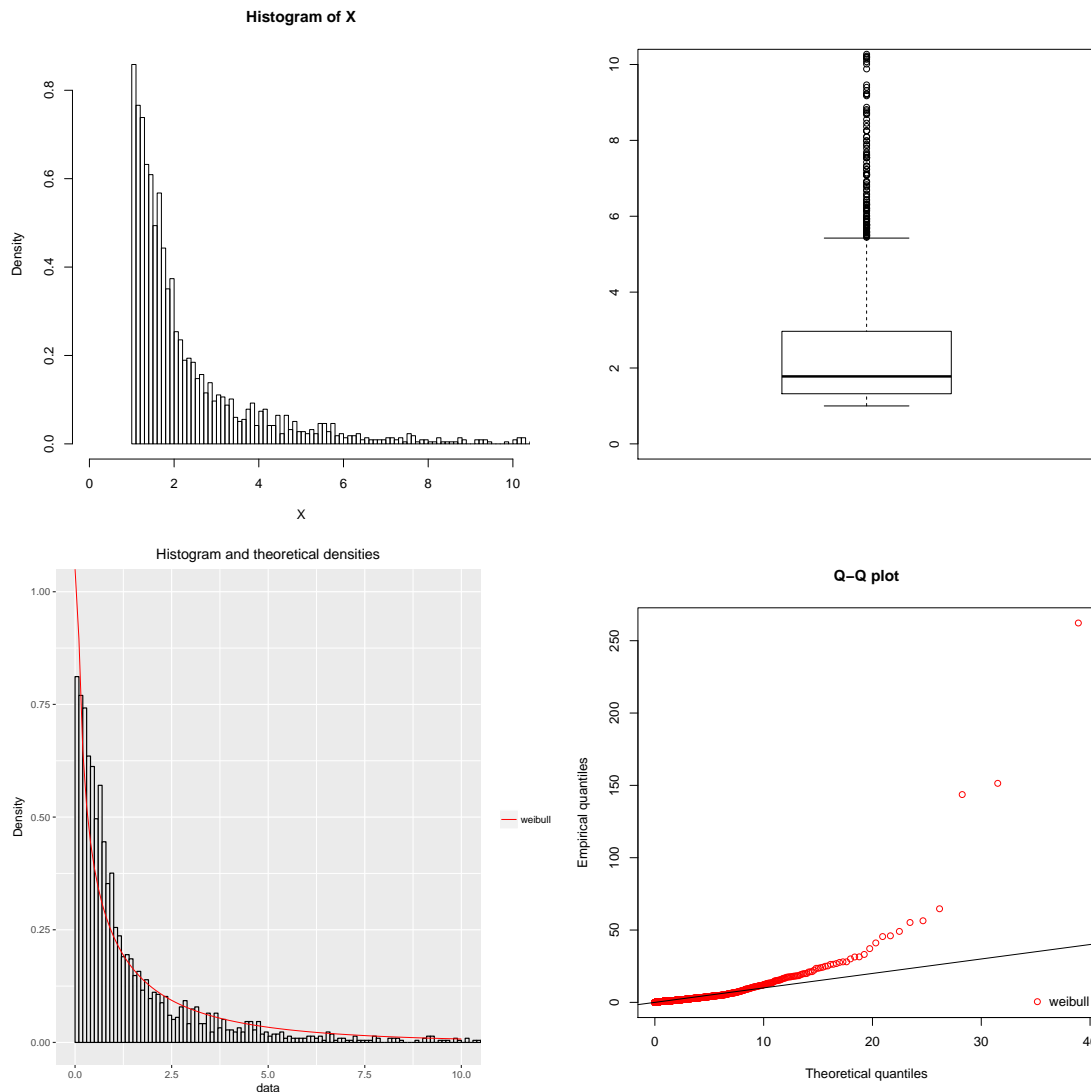


Figure 1.1 : Données d'assurance incendie. De gauche à droite et de haut en bas : histogramme des données (entre 1 et 10), boîte à moustaches des données (entre 1 et 10), histogramme et densité de Weibull estimée par maximum de vraisemblance (appliqué aux données auxquelles on a retranché 1), et diagramme quantile-quantile des données contre les quantiles de la loi de Weibull estimée.

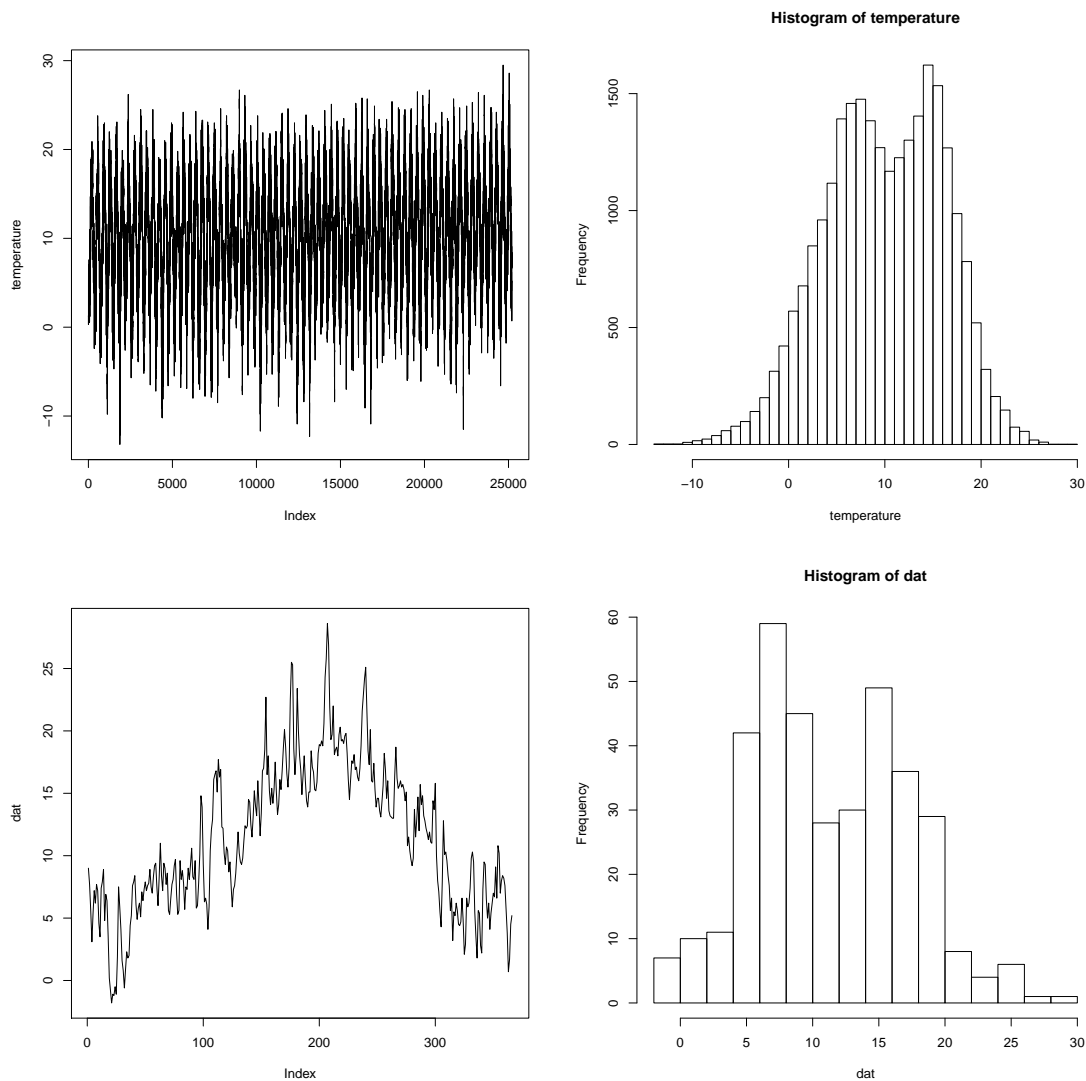


Figure 1.2 : Données sur la température enregistrée à Amsterdam-Schiphol. De gauche à droite et de haut en bas : courbe de température moyenne journalière, histogramme de la température moyenne journalière, courbe de température moyenne journalière en 2019 et histogramme de la température moyenne journalière en 2019.

ayant la même loi qu'une variable aléatoire X de fonction de répartition F , et de fonction quantile associée $q : \alpha \mapsto \inf\{t \in \mathbb{R} \mid F(t) \geq \alpha\}$, pour $\alpha \in (0, 1)$.

Parmi les grandes valeurs de l'échantillon (X_1, \dots, X_n) , la plus évidente est la plus grande d'entre elles, c'est-à-dire la variable aléatoire

$$M_n = \max_{1 \leq i \leq n} X_i.$$

La recherche de la classe de lois limites possibles pour M_n est un problème intéressant mais difficile, dont la solution est donnée par le théorème de Fisher-Tippett-Gnedenko.

Théorème 1.2.1 (Fisher-Tippett (1928), Gnedenko (1943)). *Supposons qu'il existe deux suites réelles (a_n) , strictement positive, et (b_n) , telles que la suite de variables aléatoires $(M_n - b_n)/a_n$ converge en loi, vers une loi non constante. Alors il existe $a > 0$, $b \in \mathbb{R}$ et $\gamma \in \mathbb{R}$ tels que la loi limite a pour fonction de répartition $G_\gamma(ax + b)$, où*

$$G_\gamma(x) = \exp\left(-(1 + \gamma x)^{-1/\gamma}\right), \text{ pour tout } x \text{ tel que } 1 + \gamma x > 0.$$

Lorsque $\gamma = 0$, la fonction G_0 se lit $G_0(x) = \exp(-e^{-x})$ pour tout $x \in \mathbb{R}$.

La famille de lois ayant pour fonction de répartition G_γ est appelée *famille GEV* (pour *Generalized Extreme Value*). Le Théorème 1.2.1 est important car il apporte une famille de référence pour modéliser le comportement d'un maximum. En pratique, il s'applique en construisant des blocs de données disjoints dont on prend les maxima : en partant de la série (X_1, \dots, X_n) , on peut construire les blocs (X_1, \dots, X_r) , (X_{r+1}, \dots, X_{2r}) , ... Si r est assez grand, alors les maxima correspondants $M_{r,1}, M_{r,2}, \dots$ sont indépendants et vérifient

$$M_{r,i} \stackrel{d}{\approx} b_r + a_r \text{GEV}(\gamma) = m + \sigma \text{GEV}(\gamma)$$

où $m = b_r$ et $\sigma = a_r$ sont des paramètres de position et d'échelle. Le paramètre γ est appelé *indice des valeurs extrêmes*, ou paramètre de forme, de la loi GEV. Il caractérise le comportement des extrêmes de cette distribution, au sens où il contrôle le comportement de $G_\gamma(x)$ pour de grandes valeurs de x . On peut donc, par exemple, appliquer le Théorème 1.2.1 à la modélisation des maxima mensuels, saisonniers, ou annuels d'une température ou de niveaux de pluie (si on ignore le problème de dépendance entre les données, qui est une question difficile).

Il y a cependant de nombreuses applications dans lesquelles prendre des maxima n'est pas une opération naturelle. En assurance, par exemple, il est intéressant de comprendre tous les sinistres « atypiques » ayant occasionné des pertes élevées pour une compagnie, alors

que se concentrer sur le sinistre de valeur maximale sur une période n'aurait qu'un intérêt très limité. Ceci demande d'analyser plus globalement le comportement de la loi de X pour X « grand », ce qui fait l'objet du théorème de Pickands-Balkema-de Haan.

Théorème 1.2.2 (Pickands-Balkema-de Haan). *Supposons que F a pour point terminal $\theta = q(1)$. Les deux propositions suivantes sont équivalentes :*

(i) *Il existe des suites réelles (a_n) , strictement positive, et (b_n) telles que $(M_n - b_n)/a_n$ converge en loi, vers la loi de fonction de répartition G_γ .*

(ii) *Il existe une fonction σ , strictement positive, telle que*

$$\lim_{t \uparrow \theta} \frac{\mathbb{P}((X - t)/\sigma(t) > x)}{\mathbb{P}(X > t)} = (1 + \gamma x)^{-1/\gamma} \text{ pour tout } x \text{ tel que } 1 + \gamma x > 0.$$

Pour $\gamma = 0$, la limite se lit e^{-x} .

Ce résultat entraîne la convergence

$$\lim_{t \uparrow \theta} \mathbb{P}\left(\frac{X - t}{\sigma(t)} > x \mid X > t\right) = (1 + \gamma x)^{-1/\gamma} \text{ pour tout } x > 0 \text{ tel que } 1 + \gamma x > 0.$$

Celle-ci s'interprète de la façon suivante : pour t proche de θ , la loi des excès $X - t$, conditionnellement à $X > t$, est approximativement une *loi GP* (pour *Generalized Pareto*), avec pour fonction de répartition :

$$H_{\sigma, \gamma}(x) = 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-1/\gamma}, \text{ pour tout } x > 0 \text{ avec } 1 + \frac{\gamma x}{\sigma} > 0.$$

Le paramètre $\sigma = \sigma(t) > 0$ est un paramètre d'échelle. Le Théorème 1.2.1 et le Théorème 1.2.2 nous donnent donc deux manières de modéliser les extrêmes d'un échantillon, dont la pertinence dépend de l'application considérée.

- Le modèle GEV est adapté à la modélisation de maxima, le plus souvent périodiquement observés. Son utilisation requiert un choix de fréquence d'enregistrement des données (la taille des blocs).
- Le modèle GP est adapté à la modélisation des grandes valeurs d'un échantillon dans leur globalité. Son utilisation requiert un choix de seuil au-delà duquel les observations sont considérées comme extrêmes.

Ces deux modèles paramétriques ne sont cependant pas faciles à utiliser tels quels en pratique. Il faut choisir la taille du bloc (pour la loi GEV) ou le seuil (pour la loi GP),

mais aussi et surtout implémenter l'algorithme du maximum de vraisemblance dans chacun de ces cas, ce qui s'avère difficile pour de multiples raisons.

Une analyse plus fine de la convergence vers la loi GEV permet enfin de montrer que la propriété de valeurs extrêmes se traduit en une propriété sur les grands quantiles de X . C'est l'argument qui permettra justement d'estimer des quantiles extrêmes.

Théorème 1.2.3. *Supposons que F a pour point terminal θ . On note $U(t) = q(1 - t^{-1})$ (pour $t > 1$) la fonction quantile de queue associée à F . Pour un réel γ fixé, il y a équivalence entre :*

(i) *Il existe des suites réelles (a_n) , strictement positive, et (b_n) telles que $(M_n - b_n)/a_n$ converge en loi, vers la loi GEV de fonction de répartition G_γ .*

(ii) *Il existe une fonction a , strictement positive, telle que*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma} \quad \text{pour tout } x > 0.$$

Pour $\gamma = 0$, cette limite se lit $\log x$.

Lorsque $\gamma \neq 0$, on peut réécrire (ii) plus simplement :

1. *Pour $\gamma > 0$ (lois à queue lourde) : $\theta = \infty$ et $a(t) = \gamma U(t)$, ce qui donne*

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma.$$

2. *Pour $\gamma < 0$ (lois à queue courte) : $\theta < \infty$ et $a(t) = -\gamma(\theta - U(t))$, ce qui donne*

$$\lim_{t \rightarrow \infty} \frac{\theta - U(tx)}{\theta - U(t)} = x^\gamma.$$

Il découle de ce résultat que la propriété de valeurs extrêmes se réécrit

$$q(1 - 1/(tx)) \approx q(1 - 1/t) + a(t) \frac{x^\gamma - 1}{\gamma}$$

pour t grand et $x \geq 1$. En d'autres termes, on peut estimer un quantile de niveau arbitrairement grand si l'on dispose d'un seul grand quantile $q(1 - 1/t)$, du paramètre d'échelle correspondant $a(t)$, et du paramètre de forme γ . La question de l'estimation du paramètre γ est donc primordiale en théorie des valeurs extrêmes.

Chapitre 2

Estimation dans les modèles à queue lourde

Le problème de base de la théorie des valeurs extrêmes est qu'à taille d'échantillon finie, on ne peut pas estimer un quantile de niveau quelconque non-paramétriquement. Par exemple, dans le problème de l'estimation des digues aux Pays-Bas, avec 100 ans de données climatiques, on peut certainement estimer la force d'une tempête se produisant une fois tous les 10 ans en moyenne, car une telle tempête est un évènement statistiquement observable plusieurs fois sur 100 ans. Par contre, on ne pourra pas raisonnablement estimer la force d'une tempête se produisant une fois tous les 100 ans en moyenne, car on aura observé ce phénomène une ou peut-être deux fois (voire pas du tout).

Ceci se traduit mathématiquement de la manière suivante. Etant donné un échantillon d'observations X_1, \dots, X_n indépendantes de même loi, la fonction quantile empirique associée $\hat{q}_n(\alpha) := X_{[n\alpha],n}$ (où on a réordonné l'échantillon X_1, \dots, X_n en ses statistiques d'ordre $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$) est certes un estimateur convergent de tout quantile $q(\alpha)$, quand $n \rightarrow \infty$, mais la quantité $X_{n,n}$ n'est pas un bon estimateur du quantile extrême $q(1 - 1/n)$. La question de la détermination des ordres $\alpha = \alpha_n \rightarrow 1$, qui sont extrêmes mais tels que $\hat{q}_n(\alpha_n)$ est un estimateur convergent de $q(\alpha_n)$, est une question importante qu'on abordera dans ce chapitre.

Pour pouvoir obtenir un bon estimateur d'un quantile trop extrême pour être estimé correctement par la fonction de quantile empirique, on doit donc utiliser des informations supplémentaires sur la distribution sous-jacente. Dans le cas d'une loi à queue lourde, sur

lequel on se concentrera dans ce chapitre, on a l'approximation dite de Pareto

$$\frac{U((1-\beta)^{-1})}{U((1-\alpha)^{-1})} \approx \left(\frac{1-\beta}{1-\alpha}\right)^{-\gamma} \quad \text{et donc} \quad q(\beta) \approx \left(\frac{1-\beta}{1-\alpha}\right)^{-\gamma} q(\alpha) \quad \text{pour } \alpha, \beta \uparrow 1. \quad (2.1)$$

Ce calcul est au coeur de la construction de l'estimateur de quantile extrême extrapolé de Weissman (1978), dont l'étude est l'objectif central de ce chapitre.

2.1 Estimateur de quantile intermédiaire

On se donne, dans tout le reste de ce chapitre, un échantillon (X_1, \dots, X_n) de variables aléatoires indépendantes ayant la même loi à queue lourde. On note $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ les statistiques d'ordre de l'échantillon (X_1, \dots, X_n) , et on définit la fonction quantile empirique associée $\hat{q}_n(\alpha) := X_{[n\alpha],n}$. On note également F , q et U les fonctions de répartition, quantile, et quantile de queue des X_i , respectivement (pour la fonction quantile de queue, voir le Théorème 1.2.3.)

On va tout d'abord montrer dans quelle mesure on peut estimer un quantile extrême $q(\alpha) = q(\alpha_n)$, pour $\alpha_n \uparrow 1$, à l'aide de la fonction quantile empirique classique. On a vu plus haut que $X_{n,n}$ n'est pas un estimateur convergent de $q(1 - 1/n)$. Une justification intuitive est la suivante : si la fonction de répartition F des X_i est continue,

$$1 - \hat{F}_n(q(1 - 1/n)) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i > q(1 - 1/n)\}} \approx 1 - F(q(1 - 1/n)) = 1/n$$

par la loi des grands nombres. En d'autres termes, il n'y a en moyenne qu'une observation au sein de l'échantillon qui dépasse $q(1 - 1/n)$, et on ne peut donc pas estimer avec une précision correcte ce niveau de quantile de façon purement non-paramétrique. De manière générale, on ne peut pas non plus estimer $q(1 - k/n)$, avec k un entier fixé, de cette manière. [C'est un défaut classique des méthodes purement non-paramétriques : elles ne peuvent pas extrapoler au-delà de ce qu'elles observent « fréquemment ».] La solution est d'autoriser un niveau de quantile qui est extrême, mais dépassé par un nombre croissant de données $k = k_n \rightarrow \infty$ avec $k/n \rightarrow 0$, auquel cas $\hat{q}_n(1 - k/n) = X_{n-k,n}$ sera bien un estimateur convergent et asymptotiquement normal de $q(1 - k/n)$. Le quantile $q(1 - k/n)$ est alors dit *intermédiaire*, car il est « extrême, mais pas trop ».

Pour comprendre la convergence de $\hat{q}_n(1 - k/n) = X_{n-k,n}$ vers $q(1 - k/n)$ à travers un résultat de normalité asymptotique, on doit contrôler le terme d'erreur dans l'approximation

de Pareto, donc connaître la vitesse de la convergence

$$\lim_{t \rightarrow \infty} \left(\frac{U(tx)}{U(t)} - x^\gamma \right) = 0.$$

C'est l'objet de la condition ci-dessous, dite *condition de second ordre*.

Condition $\mathcal{C}_2(\gamma, \rho, A)$: Il existe $\gamma > 0$, $\rho \leq 0$ et une fonction A , qui tend vers 0 à l'infini et est de signe constant, tels que

$$\lim_{t \rightarrow \infty} \frac{1}{A(t)} \left(\frac{U(tx)}{U(t)} - x^\gamma \right) = x^\gamma \frac{x^\rho - 1}{\rho} \quad \text{pour tout } x > 0.$$

Pour $\rho = 0$ le membre de droite se lit $x^\gamma \log x$.

La constante ρ , appelé *paramètre de second ordre*, a, comme l'indice des valeurs extrêmes γ , une interprétation sur le comportement de la distribution des X_i . On peut montrer qu'en général $|A(t)| = t^\rho \mathcal{L}(t)$, où \mathcal{L} est une fonction à *variation lente* (essentiellement, $\mathcal{L}'(t)$ se comporte comme $1/t$ au voisinage de l'infini). Autrement dit, lorsque ρ est proche de 0, la convergence des extrêmes de la loi des X_i vers ceux de la loi de Pareto d'indice γ se fait lentement. Toute distribution pour laquelle le paramètre de second ordre ρ est proche de 0 doit donc être considérée comme un cas difficile dans le cadre extrême.

Théorème 2.1.1. *Supposons que la condition $\mathcal{C}_2(\gamma, \rho, A)$ soit vérifiée. Soit $k = k_n$ une suite vérifiant $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$. Si de plus $\sqrt{k}A(n/k) = O(1)$, alors*

$$\sqrt{k} \left(\frac{X_{n-k,n}}{q(1-k/n)} - 1 \right) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

La condition supplémentaire $\sqrt{k}A(n/k) = O(1)$ est une condition de biais classique qui dit essentiellement que le biais dû à l'approximation de Pareto dans la condition $\mathcal{C}_2(\gamma, \rho, A)$ est négligeable dans l'estimation. Cette condition contraint k à ne pas être « trop grand ».

Maintenant qu'on dispose d'un estimateur convergent du quantile intermédiaire $q(1-k/n)$, on va chercher à extrapoler ce quantile à des niveaux arbitrairement extrêmes. Pour cela, il nous faut un estimateur de l'indice γ .

2.2 Estimateur de Hill

L'estimation de l'indice γ diffère de celle du quantile intermédiaire au sens où il est un paramètre d'une distribution, qui est la loi de Pareto approchant la loi des X_i au voisinage de l'infini. On peut donc chercher un argument d'estimation paramétrique pour l'estimation de γ .

Supposons que les X_i suivent exactement une loi de Pareto, c'est-à-dire que $F(x) = 1 - x^{-1/\gamma}$ pour tout $x > 1$. La densité associée est $f(x|\gamma) = \gamma^{-1}x^{-1/\gamma-1}$ pour tout $x > 1$, d'où l'estimateur du maximum de vraisemblance

$$\tilde{\gamma}_n = \frac{1}{n} \sum_{i=1}^n \log X_i.$$

Comme dans tout modèle régulier, l'estimateur du maximum de vraisemblance est convergent, asymptotiquement normal, optimal en termes de variance, et a pour variance asymptotique l'inverse de l'information de Fisher, qui est ici (pour une seule observation) $I(\gamma) = 1/\gamma^2$. On en déduit que

$$\sqrt{n}(\tilde{\gamma}_n - \gamma) \xrightarrow{d} \mathcal{N}(0, \gamma^2).$$

Revenons au cas général d'une loi à queue lourde. On réécrit l'hypothèse comme

$$\mathbb{P}(X/t > x \mid X > t) = \mathbb{P}(X > tx \mid X > t) = \frac{1 - F(tx)}{1 - F(t)} \rightarrow x^{-1/\gamma} \text{ pour tout } x > 1$$

lorsque $t \rightarrow \infty$. Pour t grand et sachant $X > t$, la variable X/t suit donc approximativement une loi de Pareto d'indice γ . Pour utiliser cette approximation en pratique, on se restreint à de grandes observations $X_{n-k,n}, X_{n-k+1,n}, \dots, X_{n,n}$, pour $k = k_n$ vérifiant $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$. On sait que $X_{n-k,n} \xrightarrow{\mathbb{P}} \infty$, donc pour n grand, les variables aléatoires $X_{n-i+1,n}/X_{n-k,n}$, $1 \leq i \leq k$, forment (approximativement !) un échantillon de la loi de Pareto de paramètre γ . On en déduit alors l'estimateur de Hill (1975) :

$$\hat{\gamma}_k = \frac{1}{k} \sum_{i=1}^k \log \frac{X_{n-i+1,n}}{X_{n-k,n}} = \frac{1}{k} \sum_{i=1}^k \log X_{n-i+1,n} - \log X_{n-k,n}.$$

La normalité asymptotique de l'estimateur de Hill est donnée dans le théorème ci-dessous.

Théorème 2.2.1. *Supposons que la condition $\mathcal{C}_2(\gamma, \rho, A)$ soit vérifiée. Soit $k = k_n$ une suite vérifiant $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$. Si de plus $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$, alors*

$$\sqrt{k}(\hat{\gamma}_k - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

L'estimateur de Hill est, comme l'estimateur du maximum de vraisemblance dans le modèle purement Pareto, asymptotiquement normal et de variance asymptotique γ^2 . Il y a deux différences principales :

- La vitesse de convergence de l'estimateur de Hill est \sqrt{k} , comparée à \sqrt{n} pour l'estimateur du maximum de vraisemblance. Il n'y a en effet que k observations de l'échantillon prises en compte dans l'estimateur de Hill.

- Il y a une composante de biais dans l'estimateur de Hill, contrôlée par la fonction A apparaissant dans la condition $\mathcal{C}_2(\gamma, \rho, A)$. Cette composante de biais est présente car l'estimateur de Hill ne maximise qu'une quasi-vraisemblance, puisque les $X_{n-i+1,n}/X_{n-k,n}$ ne forment pas un échantillon indépendant et identiquement distribué de la loi de Pareto.

Pour obtenir une bonne estimation de γ , un bon choix de k est donc crucial. Prendre k trop petit donnera un estimateur avec une grande variance puisque basé sur un petit nombre d'observations, alors que prendre k trop grand donnera un estimateur biaisé puisque prenant en compte des observations trop loin de la queue de distribution et donc non représentatives des extrêmes des X_i . En ce sens, le choix de k est un problème analogue à celui du choix du paramètre de lissage pour l'estimation d'une fonction de densité ou d'une moyenne conditionnelle.

Pour illustrer la difficulté de ce choix, on a représenté sur la Figure 2.1 quatre exemples de courbes $k \mapsto \hat{\gamma}_k$, pour des jeux de données simulées de taille $n = 500$ avec $\gamma = 1/2$, dans les cas suivants :

- La loi de Pareto, de fonction de répartition $F(x) = 1 - x^{-1/\gamma}$ (pour $x > 1$),
- La loi de Fréchet, de fonction de répartition $F(x) = e^{-x^{-1/\gamma}}$ (pour $x > 0$),
- La loi de Burr, de fonction de répartition $F(x) = 1 - (1 + x^{-\rho/\gamma})^{1/\rho}$ (pour $x > 0$), où on prend $\rho = -1$,
- La loi de Burr encore, cette fois-ci avec $\rho = -1/2$.

Pour la loi de Pareto, le biais est nul, et l'estimateur de Hill, après de fortes oscillations pour k petit, se stabilise autour de la vraie valeur. Pour les lois de Fréchet et Burr, on observe que l'estimateur de Hill se stabilise pour k « intermédiaire » (ni trop petit, ni trop grand) puis s'éloigne de la vraie valeur $\gamma = 1/2$. Ce dernier comportement est typique, et en pratique, un bon choix de k se ramène souvent à la détection d'une plage de stabilité suffisamment longue parmi les petites valeurs de k , dans laquelle on prend une estimation $\hat{\gamma}_k$. On constate aussi que le biais est plus fort lorsque ρ est proche de 0 (comparer les échelles verticales des deux graphiques au bas de la Figure 2.1). Sur certains exemples, la plage de stabilité peut également être plus courte et difficile à détecter dans un cas où ρ est proche de 0, mais ce n'est pas nécessairement le cas.

Pour vérifier que k a été correctement choisi et donc que l'estimation de γ est raisonnable, on peut se servir d'une autre interprétation de γ . On montre que, lorsque Y_1, \dots, Y_n ont

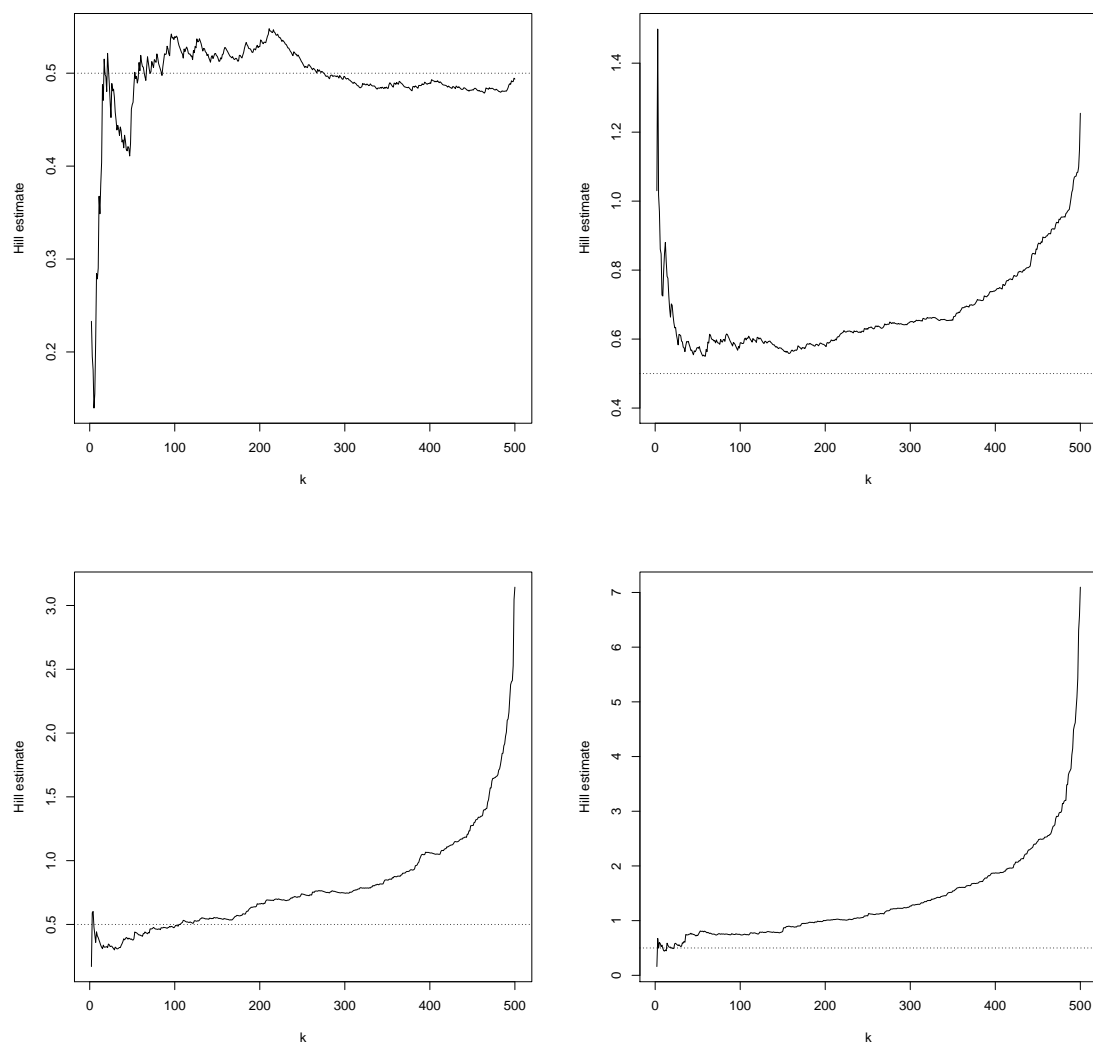


Figure 2.1 : Courbe de l'estimateur de Hill sur un jeu de données de taille $n = 500$, pour $\gamma = 1/2$. De gauche à droite et de haut en bas : loi de Pareto, loi de Fréchet, loi de Burr avec $\rho = -1$, loi de Burr avec $\rho = -1/2$. Horizontalement et en pointillés : droite $\gamma = 1/2$.

une loi de Pareto d'indice γ ,

$$\left(i \log \frac{Y_{n-i+1,n}}{Y_{n-i,n}} \right)_{1 \leq i \leq k} \stackrel{d}{=} (\gamma i (E_{n-i+1,n} - E_{n-i,n}))_{1 \leq i \leq k} \stackrel{d}{=} (\gamma E_i)_{1 \leq i \leq k}$$

où les E_i sont indépendantes et suivent une loi exponentielle de paramètre 1. Ceci suggère que pour un bon choix de k , les quantités $i \log(X_{n-i+1,n}/X_{n-i,n})$, pour $1 \leq i \leq k$, devraient être « approximativement » indépendantes et de même loi exponentielle de moyenne γ . On peut vérifier cela avec un QQ-plot construit comme suit :

- Construire les variables E_1^*, \dots, E_k^* définies par

$$E_i^* = i \log \frac{X_{n-i+1,n}}{X_{n-i,n}}.$$

- Construire le QQ-plot exponentiel

$$\left\{ \left(\hat{\gamma}_k \log \left(\frac{k+1}{k+1-i} \right), E_{i,k}^* \right), 1 \leq i \leq k \right\}.$$

Ce QQ-plot compare les quantiles théoriques de la loi exponentielle de moyenne γ aux niveaux $1/(k+1), \dots, k/(k+1)$, avec les quantiles empiriques des E_i^* .

- Si k et $\hat{\gamma}_k$ ont été bien choisis (et que l'hypothèse de queue lourde est raisonnable), ce QQ-plot doit être approximativement aligné sur la droite d'équation $y = x$.

On donne deux exemples de cette procédure sur la Figure 2.2. Dans les deux exemples, l'estimation de γ est globalement satisfaisante, même si le cas avec ρ proche de 0 est plus difficile, comme on devait s'y attendre.

2.3 Estimateur de Weissman d'un quantile extrême

On peut maintenant construire un estimateur de quantile extrême à tout ordre $\beta = \beta_n \uparrow 1$. Dans le cas $n(1 - \beta_n) \rightarrow \infty$, on peut utiliser l'estimateur de quantile intermédiaire de la Section 2.1. On se concentre donc sur le cas de l'estimation d'un quantile $q(\beta_n)$, où $n(1 - \beta_n) \rightarrow c < \infty$. On rappelle la relation d'extrapolation (2.1) :

$$q(\beta) \approx \left(\frac{1-\beta}{1-\alpha} \right)^{-\gamma} q(\alpha) \text{ pour } \alpha, \beta \uparrow 1.$$

Pour obtenir un estimateur du quantile extrême $q(\beta) = q(\beta_n)$, on utilise cette approximation pour extrapoler un estimateur de quantile intermédiaire $q(\alpha) = q(\alpha_n)$, avec

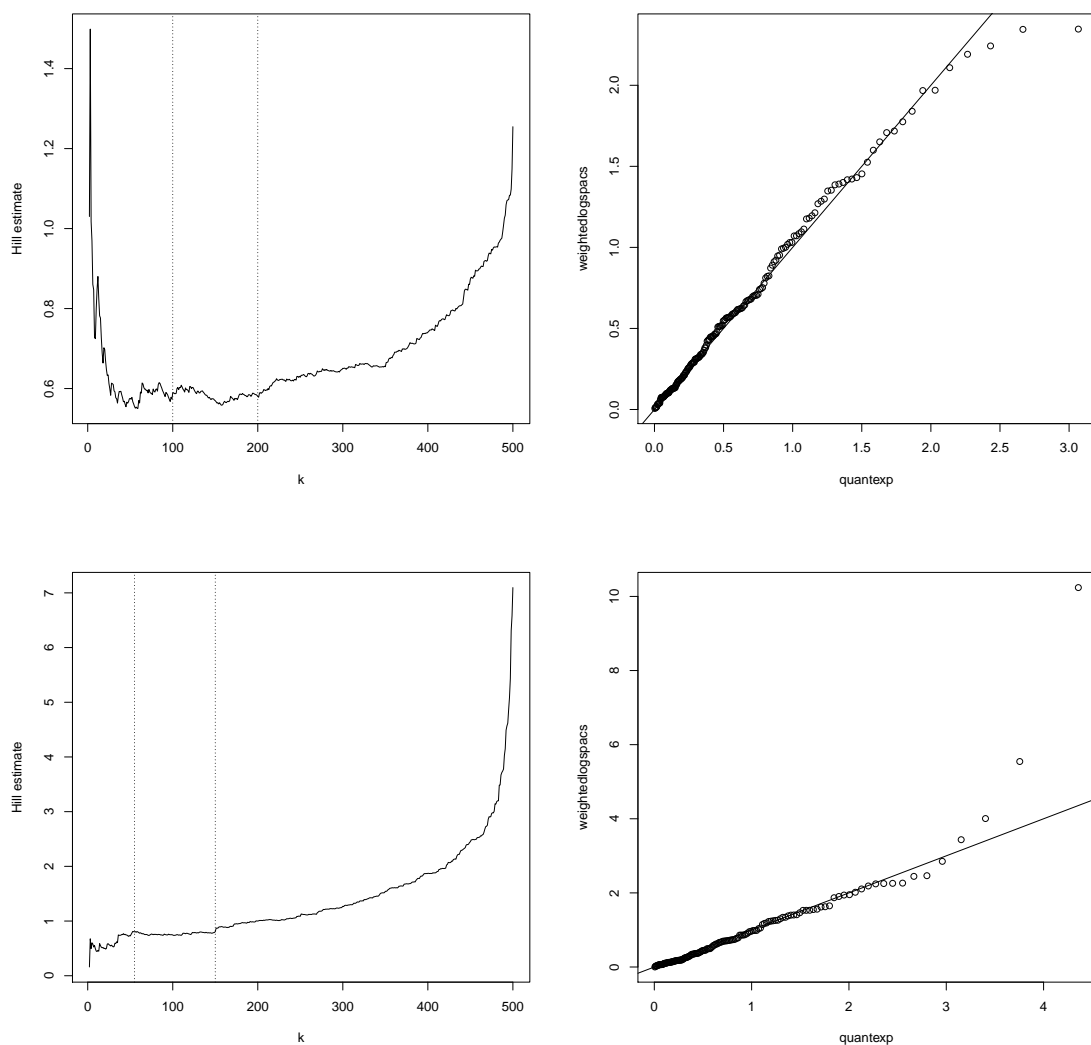


Figure 2.2 : Courbe de l'estimateur de Hill sur un jeu de données de taille $n = 500$, pour $\gamma = 1/2$, et diagnostic par QQ-plot exponentiel. En haut : loi de Fréchet, en bas : loi de Burr avec $\rho = -1/2$. A gauche : estimateur de Hill, avec plage de stabilité, à droite : QQ-plot exponentiel, où $\hat{\gamma}_k$ est calculé en prenant la dernière valeur sur la plage de stabilité.

$n(1 - \alpha_n) \rightarrow \infty$, en remplaçant γ par l'estimateur de Hill $\hat{\gamma}_k$. On prend le plus souvent $\alpha_n = 1 - k/n$, ce qui donne l'estimateur extrapolé de Weissman

$$\hat{q}_n^*(\beta_n|k) := \left(\frac{n(1 - \beta_n)}{k}\right)^{-\hat{\gamma}_k} \hat{q}_n(1 - k/n) = \left(\frac{n(1 - \beta_n)}{k}\right)^{-\hat{\gamma}_k} X_{n-k,n}.$$

La normalité asymptotique de l'estimateur de Weissman est en fait essentiellement celle de l'estimateur de Hill, comme le montre le résultat ci-dessous.

Théorème 2.3.1. *Supposons que la condition $\mathcal{C}_2(\gamma, \rho, A)$ soit vérifiée avec $\rho < 0$. Soit $k = k_n$ une suite vérifiant $k \rightarrow \infty$ et $k/n \rightarrow 0$, et supposons que $\beta_n \uparrow 1$ vérifie $n(1 - \beta_n) \rightarrow c < \infty$ quand $n \rightarrow \infty$. Si de plus $\sqrt{k}/\log(k/[n(1 - \beta_n)]) \rightarrow \infty$ et $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$, alors*

$$\frac{\sqrt{k}}{\log(k/[n(1 - \beta_n)])} \left(\frac{\hat{q}_n^*(\beta_n|k)}{q(\beta_n)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1 - \rho}, \gamma^2\right).$$

La vitesse de convergence de l'estimateur de Weissman est presque celle de l'estimateur de Hill, à un facteur logarithmique près. Ce terme est le prix à payer pour l'utilisation de la relation d'extrapolation suggérée par l'approximation de Pareto, et ce prix augmente à mesure que l'on cherche à extrapoler plus loin dans la queue de distribution. En pratique, la construction d'intervalles de confiance se fait plutôt en utilisant la version du Théorème 2.3.1 sur l'échelle logarithmique :

$$\frac{\sqrt{k}}{\log(k/[n(1 - \beta_n)])} \log \frac{\hat{q}_n^*(\beta_n|k)}{q(\beta_n)} \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1 - \rho}, \gamma^2\right).$$

L'intervalle de confiance asymptotique au niveau $1 - \alpha$ s'en déduisant (en négligeant la composante de biais) est

$$\hat{I}_{n,\alpha}(\beta_n) = \left[\hat{q}_n^*(\beta_n|k) \exp\left(\pm \frac{\log(k/[n(1 - \beta_n)])}{\sqrt{k}} \hat{\gamma}_k \times z_{1-\alpha/2}\right) \right]$$

où z_α est le quantile de niveau α de la loi normale centrée réduite.

2.4 Application sur données réelles

On reprend les données d'assurance incendie de l'entreprise Copenhagen Re qu'on a utilisées à la Section 1.1.1. On a représenté en haut à droite de la Figure 2.3 l'estimateur de Hill pour ce jeu de données de taille $n = 2167$, accompagné d'intervalles de confiance calculés grâce au Théorème 2.2.1 (en négligeant le terme de biais). L'estimateur a une plage de stabilité assez longue et semble osciller autour de $\hat{\gamma} = 0.7$. Il faut cependant veiller à choisir

une valeur de k relativement petite, afin d'assurer que l'approximation de Pareto, utilisée pour le calcul de l'estimateur de Weissman, est raisonnable. Le choix $k = 550$ paraît sensé ici. Pour cette valeur de k , le QQ-plot exponentiel décrit à la fin de la Section 2.2 est représenté en bas à gauche de la Figure 2.3. On constate que l'approximation de Pareto est effectivement crédible sur ces k plus grandes valeurs de l'échantillon. Enfin, en bas à droite de la Figure 2.3, on a représenté l'estimateur de Weissman, accompagné d'intervalles de confiance calculés grâce au Théorème 2.3.1 (sur l'échelle logarithmique, au lieu de l'échelle relative). Avec $\beta = 1 - 1/n = 0.9995385$, et $k = 550$, on trouve un estimateur de Weissman de $q(\beta)$ égal à $\hat{q}^*(\beta) = 251.3$. Cette valeur est très largement supérieure au quantile empirique des données au niveau β , qui est $\hat{q}(\beta) = 152.5$. Ceci montre tout l'intérêt de l'extrapolation, puisque dans cet exemple, estimer le quantile extrême de façon non-paramétrique induit une forte sous-estimation du risque.

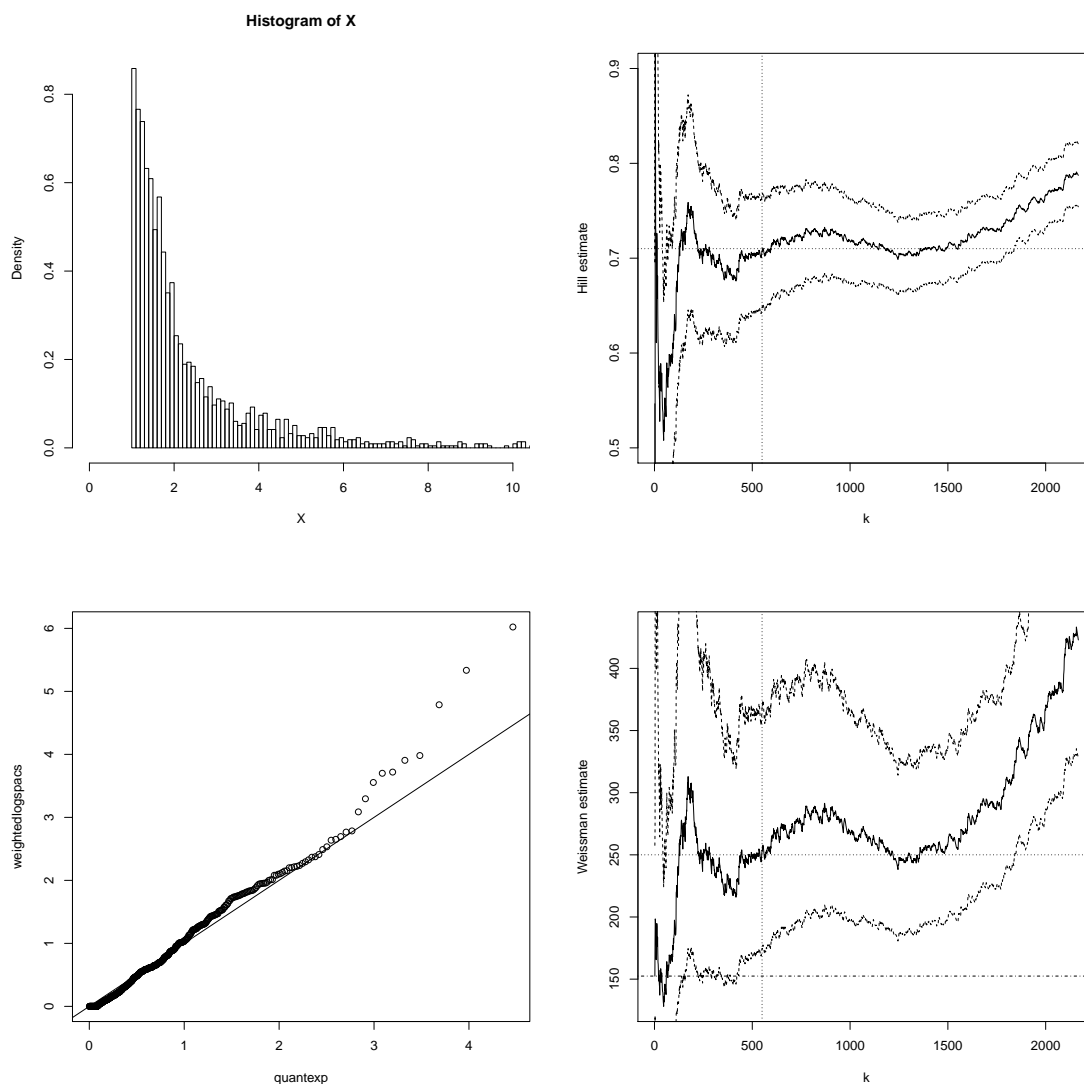


Figure 2.3 : Jeu de données d'assurance incendie, de taille $n = 2167$. De gauche à droite et de haut en bas : histogramme des données (entre 1 et 10), estimateur de Hill, QQ-plot exponentiel décrit à la fin de la Section 2.2, et estimateur de Weissman au niveau $\beta = 1 - 1/n = 0.9995385$. Les courbes en pointillés représentent des intervalles de confiance asymptotiques au niveau de confiance 95%.

Chapitre 3

Estimation hors des modèles à queue lourde

Il existe de nombreux cas dans lesquels les modèles à queue lourde ne s'appliquent pas. Lorsqu'on souhaite estimer des niveaux de température, par exemple, il est raisonnable d'utiliser un modèle borné (à la surface de la Terre en tout cas et dans les conditions actuelles). Dans certaines modélisations en assurance, notamment celles qui ne tiennent pas compte d'événements rares (vol, assurance automobile...) les modèles à queue légère représentent bien le type de remboursement qu'on observe. Dans ces cas, on utilise une approximation des extrêmes de la loi sous-jacente par une loi GP, comportant un paramètre de position, d'échelle et le paramètre de forme γ . Cette approximation est très flexible, car elle s'applique à tout modèle, qu'il soit à queue lourde, courte ou légère :

$$U(tx) \approx U(t) + a(t) \frac{x^\gamma - 1}{\gamma} \quad \text{pour tout } x > 0 \text{ et } t \rightarrow \infty.$$

On interprète ici $U(t) = q(1 - 1/t)$ comme le paramètre de position, estimé avec un quantile empirique intermédiaire, et $a(t)$ joue le rôle de paramètre d'échelle. On propose ici une approche paramétrique pour l'estimation, par maximum de vraisemblance.

3.1 Approche paramétrique

On suppose disposer d'un échantillon (X_1, \dots, X_n) de variables aléatoires indépendantes ayant la même loi et satisfaisant la propriété de valeurs extrêmes. On note $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ les statistiques d'ordre de l'échantillon (X_1, \dots, X_n) , ainsi que F , q et U les

fonctions de répartition, quantile, et quantile de queue des X_i , respectivement. On note θ le point terminal des X_i .

Estimateurs des paramètres γ et σ On sait, par le Théorème 1.2.2, que pour t grand, on peut approximativement modéliser l'excès $X - t$ par une loi GP, ayant pour fonction de répartition

$$H_{\sigma,\gamma}(x) = 1 - \left(1 + \frac{\gamma x}{\sigma}\right)^{-1/\gamma} \text{ pour tout } x > 0 \text{ avec } 1 + \frac{\gamma x}{\sigma} > 0.$$

Le paramètre $\sigma = \sigma(t) > 0$ est un paramètre d'échelle. La densité de probabilité correspondante est

$$h(x|\sigma, \gamma) = \frac{1}{\sigma} \left(1 + \frac{\gamma x}{\sigma}\right)^{-1/\gamma-1} \text{ pour tout } x > 0 \text{ avec } 1 + \frac{\gamma x}{\sigma} > 0.$$

En pratique, on peut utiliser cette approximation en fixant $t = X_{n-k,n}$, où $k = k_n$ vérifie $k \rightarrow \infty$ et $k/n \rightarrow 0$, et en décidant de calibrer une loi GP pour décrire la loi des variables aléatoires $(Y_i = Y_{i,k} := X_{n-i+1,n} - X_{n-k,n})_{1 \leq i \leq k}$. En traitant ces variables aléatoires comme formant un échantillon indépendant de loi GP, la vraisemblance correspondante est

$$L(Y_1, \dots, Y_k|\sigma, \gamma) = \prod_{i=1}^k h(Y_i|\sigma, \gamma) = \sigma^{-k} \prod_{i=1}^k \left(1 + \frac{\gamma Y_i}{\sigma}\right)^{-1/\gamma-1} \mathbb{1}_{\{1+\gamma Y_i/\sigma > 0\}}.$$

[Les Y_i sont strictement positives par définition, donc on omet cette contrainte dans la vraisemblance.] Maximiser cette (quasi-)vraisemblance est difficile car le modèle GP n'est pas régulier, puisque le support des variables Y_i dépend des paramètres σ et γ . On montre en fait que cette vraisemblance n'est jamais bornée si l'on optimise sur le domaine complet $(\sigma, \gamma) \in]0, \infty[\times \mathbb{R}$. La restriction au domaine $(\sigma, \gamma) \in]0, \infty[\times]-1/2, \infty[$, auquel on se limite dans ce qui suit, permet d'obtenir des estimateurs du maximum de vraisemblance qui sont convergents et asymptotiquement gaussiens.

Pour écrire les équations définissant les estimateurs du maximum de vraisemblance, il est intéressant au préalable de reparamétriser le problème en posant $\tau = \gamma/\sigma$ et en optimisant en (τ, γ) . La vraisemblance s'écrit

$$L(Y_1, \dots, Y_k|\tau, \gamma) = \left(\frac{\tau}{\gamma}\right)^k \prod_{i=1}^k (1 + \tau Y_i)^{-1/\gamma-1} \mathbb{1}_{\{1+\tau Y_i > 0\}}.$$

Les équations définissant les estimateurs $\hat{\tau}_k^{\text{ML}}$ et $\hat{\gamma}_k^{\text{ML}}$ sont alors

$$\frac{1}{\hat{\tau}_k^{\text{ML}}} = \left(\frac{1}{\hat{\gamma}_k^{\text{ML}}} + 1\right) \times \frac{1}{k} \sum_{i=1}^k \frac{Y_i}{1 + \hat{\tau}_k^{\text{ML}} Y_i} \text{ et } \hat{\gamma}_k^{\text{ML}} = \frac{1}{k} \sum_{i=1}^k \log(1 + \hat{\tau}_k^{\text{ML}} Y_i).$$

[Il est à noter que ces équations n'ont pas de sens si $\gamma \leq -1/2$ car la vraisemblance n'est alors pas différentiable.] Ces estimateurs ne sont pas explicites et doivent être calculés numériquement. Ils ont cependant l'avantage de donner automatiquement un estimateur de σ en plus d'un estimateur de γ , en posant $\hat{\sigma}_k^{\text{ML}} = \hat{\gamma}_k^{\text{ML}} / \hat{\tau}_k^{\text{ML}}$.

Leur convergence est un problème difficile. On pourra consulter Drees *et al.* (2004) et Zhou (2009) pour la preuve du résultat suivant, formulé ici sous la condition $\mathcal{H}_2(\gamma, a, \rho, A)$ comme dans le Théorème 3.4.2 dans de Haan et Ferreira (2006) :

Condition $\mathcal{H}_2(\gamma, a, \rho, A)$: Il existe $\gamma \in \mathbb{R}$, une fonction strictement positive a , $\rho \leq 0$ et une fonction A , qui tend vers 0 à l'infini et est de signe constant, tels que

$$\lim_{t \rightarrow \infty} \frac{1}{A(t)} \left(\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma} \right) = \frac{1}{\rho} \left(\frac{x^{\gamma+\rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right) \quad \text{pour tout } x > 0.$$

Pour γ ou $\rho = 0$ le membre de droite se lit comme sa limite quand $\gamma \rightarrow 0$ ou $\rho \uparrow 0$.

On pourra également consulter Zhou (2010) pour une extension au cas $\gamma \in]-1, -1/2]$.

Théorème 3.1.1. *Supposons que la condition $\mathcal{H}_2(\gamma, a, \rho, A)$ soit vérifiée avec $\gamma > -1/2$. Soit $k = k_n$ une suite vérifiant $k \rightarrow \infty$ et $k/n \rightarrow 0$ quand $n \rightarrow \infty$. Si de plus $\sqrt{k}A(n/k) \rightarrow \lambda \in \mathbb{R}$, alors les équations du maximum de vraisemblance ont une solution $(\hat{\gamma}_k^{\text{ML}}, \hat{\sigma}_k^{\text{ML}})$ qui vérifie*

$$\sqrt{k} \left(\hat{\gamma}_k^{\text{ML}} - \gamma, \frac{\hat{\sigma}_k^{\text{ML}}}{\sigma_k} - 1 \right) \xrightarrow{d} \mathcal{N}(\lambda b^{\text{ML}}(\gamma, \rho), \Sigma^{\text{ML}}(\gamma)).$$

La matrice de covariance asymptotique est

$$\Sigma^{\text{ML}}(\gamma) = \begin{pmatrix} (1 + \gamma)^2 & -(1 + \gamma) \\ -(1 + \gamma) & 1 + (1 + \gamma)^2 \end{pmatrix}.$$

[Le terme de biais a une expression explicite, mais plus complexe.]

Les estimateurs du maximum de vraisemblance n'étant pas explicites, on doit utiliser un algorithme pour les calculer numériquement. Il existe plusieurs possibilités pour cela, dont la routine `fevd`, du package R `extRemes`. Un des inconvénients de l'utilisation de ces estimateurs est leur temps de calcul parfois élevé, en plus de leur limitation aux cas pour lesquels $\gamma > -1/2$. La réduction du problème à celui de la recherche d'une racine d'une fonction d'une seule variable a été étudiée par Grimshaw (1993) et reprise en partie par Zhou (2009, 2010) dans son travail théorique sur la convergence des estimateurs.

On peut également utiliser l'approximation paramétrique pour construire des estimateurs des moments. Ce type d'approche est fondée sur le calcul de quelques moments bien choisis

de la loi GP, qui sont ensuite combinés pour en déduire des estimateurs de σ et γ . On peut, par exemple, utiliser les relations

$$\mathbb{E}(Y) = \frac{\sigma}{1-\gamma} \quad \text{et} \quad \text{Var}(Y) = \frac{\sigma^2}{(1-\gamma)^2(1-2\gamma)},$$

qui sont valides si Y suit une loi GP de paramètres σ et γ avec $\gamma < 1/2$, desquelles on déduit

$$\gamma = \frac{1}{2} \left(1 - \frac{[\mathbb{E}(Y)]^2}{\text{Var}(Y)} \right) \quad \text{et} \quad \sigma = \frac{\mathbb{E}(Y)}{2} \left(1 + \frac{[\mathbb{E}(Y)]^2}{\text{Var}(Y)} \right).$$

Il est alors très simple de construire des estimateurs basés sur les moments empiriques des variables aléatoires $(Y_i := X_{n-i+1,n} - X_{n-k,n})_{1 \leq i \leq k}$, pour $k = k_n$ vérifiant $k \rightarrow \infty$ et $k/n \rightarrow 0$: on pose

$$\bar{m}_k := \frac{1}{k} \sum_{i=1}^k X_{n-i+1,n} - X_{n-k,n} \quad \text{et} \quad \bar{s}_k^2 := \frac{1}{k} \sum_{i=1}^k (X_{n-i+1,n} - X_{n-k,n} - \bar{m}_k)^2$$

et on définit

$$\hat{\gamma}_k^M = \frac{1}{2} \left(1 - \frac{\bar{m}_k^2}{\bar{s}_k^2} \right) \quad \text{et} \quad \hat{\sigma}_k^M = \frac{\bar{m}_k}{2} \left(1 + \frac{\bar{m}_k^2}{\bar{s}_k^2} \right).$$

Ces estimateurs sont très faciles à calculer mais ont tendance à être instables, car ils sont particulièrement sensibles aux grandes valeurs de l'échantillon. On a tendance à leur préférer leurs versions construites en utilisant des moments pondérés de la forme $M_{p,r,s}(Y) = \mathbb{E}(Y^p[F(Y)]^r[1-F(Y)]^s)$ pour de bons choix de p , r et s . C'est l'approche utilisée par Hosking et Wallis (1987) et plus tard Diebolt *et al.* (2007). Les estimateurs qui en résultent peuvent être calculés pour une grande gamme de valeurs de γ , sans condition de signe, et constituent par exemple d'excellentes valeurs de départ pour une estimation par maximum de vraisemblance par un algorithme itératif de type Newton-Raphson.

Estimateur de quantile extrême On utilise la formule d'extrapolation, qui demande l'estimation du paramètre de position, d'échelle et de forme. Si l'on suit l'approche par maximum de vraisemblance, elle a l'avantage d'être automatique puisque les estimateurs de $q(1-k/n)$, σ_k et γ ont été calculés simultanément. On peut donc directement calculer l'estimateur extrapolé

$$\hat{q}_n^{*,\text{ML}}(\beta_n|k) := X_{n-k,n} + \frac{\hat{\sigma}_k^{\text{ML}}}{\hat{\gamma}_k^{\text{ML}}} \left(\left[\frac{n(1-\beta_n)}{k} \right]^{-\hat{\gamma}_k^{\text{ML}}} - 1 \right).$$

Sous la condition de second ordre et sous une condition de biais, on montre que

$$\sqrt{k} \frac{\hat{q}_n^{*,\text{ML}}(\beta_n|k) - q(\beta_n)}{\hat{\sigma}_k^{\text{ML}} \psi_{\hat{\gamma}_k^{\text{ML}}}(k/[n(1-\beta_n)])}, \quad \text{avec} \quad \psi_\gamma(t) = \int_1^t s^{\gamma-1} \log s \, ds,$$

converge en loi vers une loi normale, ayant pour variance

$$V^{\text{ML}}(\gamma) = \begin{cases} (1 + \gamma)^2 & \text{si } \gamma \geq 0, \\ 1 + 4\gamma + 5\gamma^2 + 2\gamma^3 + 2\gamma^4 & \text{si } \gamma < 0. \end{cases}$$

Un intervalle de confiance asymptotique gaussien s'en déduit immédiatement.

3.2 Application sur données réelles

On reprend les données de température à l'aéroport d'Amsterdam-Schiphol qu'on a utilisées à la Section 1.1.2. Pour éliminer tout problème de saisonnalité, on retravaille les données pour ne retenir que les températures moyennes journalières enregistrées au mois d'août, de 1950 à 2019. Le jeu de données qui en résulte, de taille $n = 2\,139$, est représenté en haut à gauche de la Figure 3.1. Une tendance à la hausse est visible, qu'on estime par un modèle linéaire avec l'estimateur des moindres carrés, et qu'on représente sur cette même figure. Le jeu de données auquel on a retiré la tendance est représenté en haut à droite de la Figure 3.1. On estime ensuite l'indice des valeurs extrêmes γ à l'aide de l'estimateur du maximum de vraisemblance $\hat{\gamma}_k^{\text{ML}}$; ce dernier est calculé à l'aide de la fonction `fevd` du package `extRemes`. On prend $k = 800$, qui correspond à des estimateurs $\hat{\sigma}_k^{\text{ML}} = 3.07$ et $\hat{\gamma}_k^{\text{ML}} = -0.24$. Un des avantages de l'utilisation de la routine `fevd` est qu'elle permet d'obtenir immédiatement des diagnostics de la qualité de l'approximation GP au moyen d'un simple `plot` : on en donne un exemple sur ce jeu de données avec la Figure 3.2. On constate que la qualité de l'approximation GP est globalement satisfaisante. Enfin, on conclut en estimant un quantile extrême de la température au moyen de la formule

$$\hat{q}_n^{*,\text{ML}}(\beta_n|k) := X_{n-k,n} + \frac{\hat{\sigma}_k^{\text{ML}}}{\hat{\gamma}_k^{\text{ML}}} \left(\left[\frac{n(1-\beta_n)}{k} \right]^{-\hat{\gamma}_k^{\text{ML}}} - 1 \right).$$

Avec $\beta_n = 1 - 1/n = 0.9995325$, on trouve $\hat{q}_n^{*,\text{ML}}(\beta_n|k) = 26.7$. Cette estimation est représentée en bas à droite de la Figure 3.1, superposée aux données. On constate que l'estimation du quantile extrême extrapole bien au-delà des valeurs couvertes par l'échantillon. Il n'est cependant pas raisonnable d'utiliser ici d'intervalle de confiance asymptotique suggéré par les calculs de la Section 3.1 en raison de la nature des données. [Pourquoi ?]

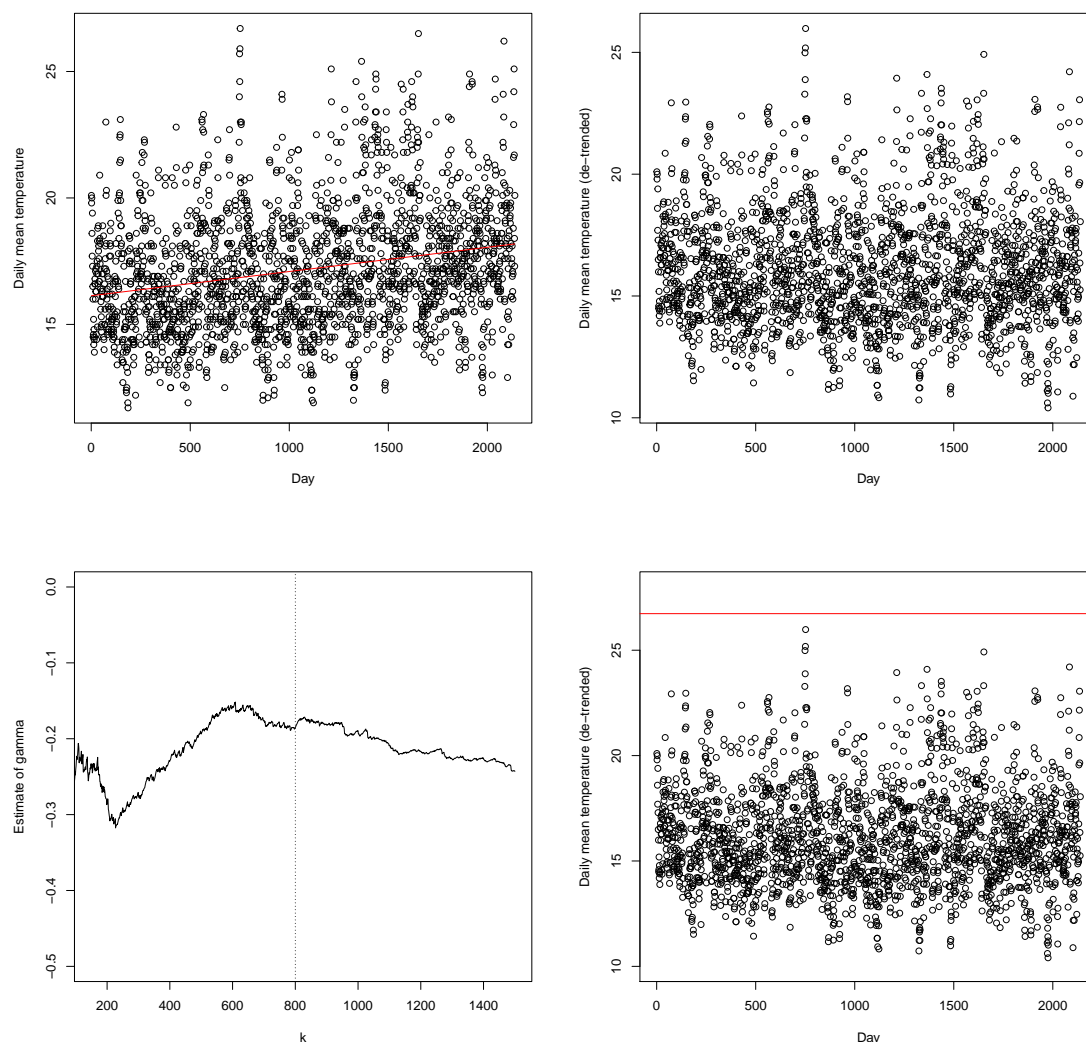


Figure 3.1 : Données sur la température enregistrée à Amsterdam-Schiphol. En haut à gauche : données brutes sur la température moyenne journalière (avec tendance temporelle en rouge). En haut à droite : données auxquelles on a retiré la tendance. En bas à gauche : estimateur du maximum de vraisemblance $\hat{\gamma}_k^{\text{ML}}$. En bas à droite : données auxquelles on a retiré la tendance avec estimateur du quantile extrême au niveau $\beta = 1 - 1/n = 0.9995325$.

```
fevd(x = X, threshold = thresh, type = "GP")
```

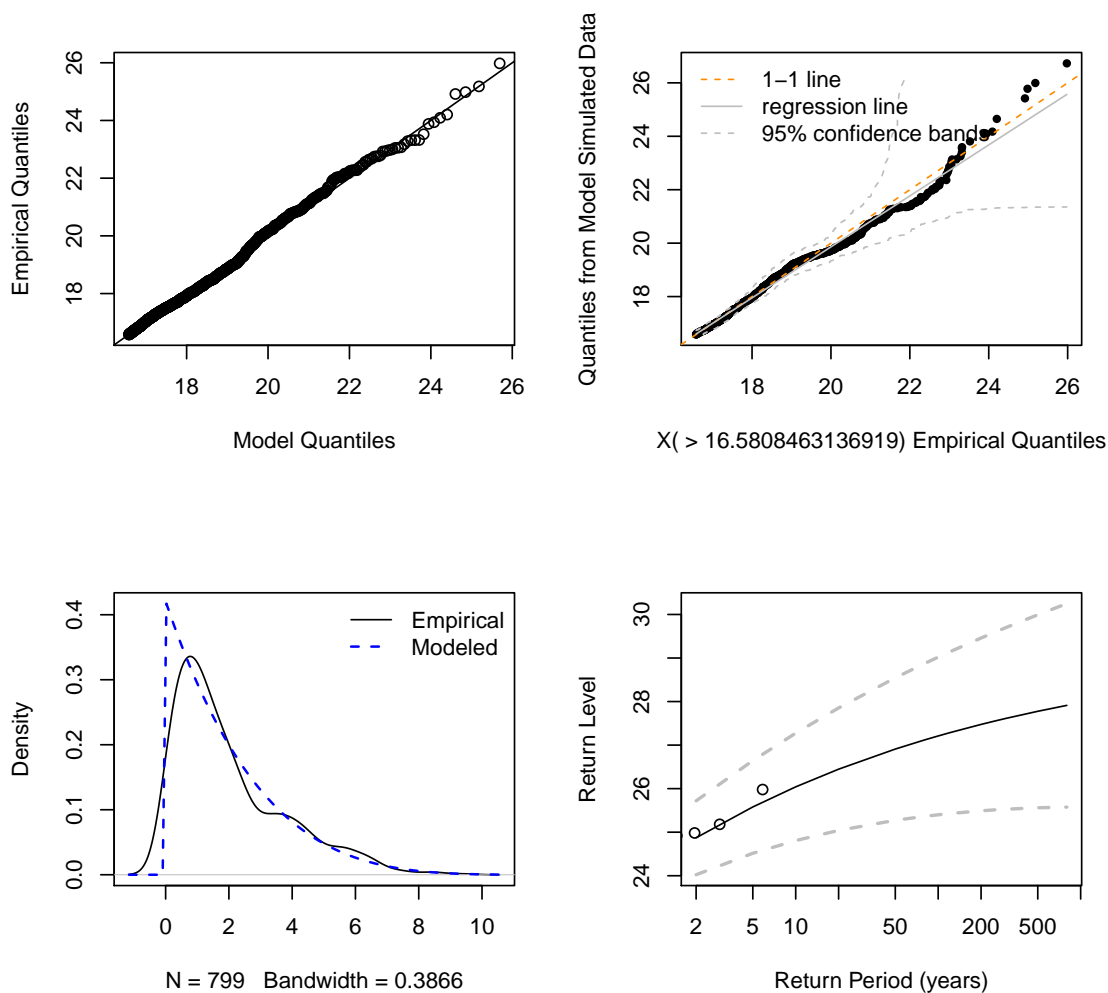


Figure 3.2 : Données sur la température enregistrée à Amsterdam-Schiphol. En haut à gauche : QQ-plot basé sur la loi GP estimée pour les k plus grandes valeurs de l'échantillon. En haut à droite : QQ-plot empirique comparant un jeu de données simulées de la loi GP estimée et les k plus grandes valeurs de l'échantillon. En bas à gauche : comparaison entre un estimateur non-paramétrique de la densité des observations et la densité estimée de la loi GP. En bas à droite : estimation de niveaux de retours extrêmes. Les courbes en pointillés représentent des intervalles de confiance (ou de prédiction) asymptotiques au niveau de confiance 95%.