Correction des exercices du Chapitre 03

Correction de l'exercice 01

Méthode d'estimation du maximum de vraisemblance

La fonction log-vraisemblance de la loi GPD est donnée par :

$$l\left((Y_1, Y_2, \ldots, Y_{N_u}), (\xi, \sigma(u))\right) = -N_u \log(\sigma(u)) - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^{N_u} \log\left(1 + \xi \frac{Y_i}{\sigma(u)}\right)$$

avec $1 + \xi \frac{Y_i}{\sigma(u)} > 0$ pour $i = 1, 2, \ldots, N_u$.

En posant $\eta = \frac{\xi}{\sigma(u)}$, l'équation de vraisemblance

$$\frac{\partial\left((Y_1, Y_2, \ldots, Y_{N_u}), (\xi, \sigma(u))\right)}{\partial(\xi, \sigma(u))} = 0$$

nous conduit au système suivant :

$$\begin{cases} \hat{\xi} = \frac{1}{N_u} \sum_{i=1}^{N_u} \log(1 + \eta Y_i) = \hat{\xi}(\eta) \\ \frac{1}{\eta} = \frac{1}{N_u}\left(1 + \frac{1}{\hat{\xi}}\right) \sum_{i=1}^{N_u} \frac{Y_i}{1 + \eta Y_i} \end{cases}$$

Donc l'estimateur du maximum de vraisemblance de (ξ, η) est $(\hat{\xi} = \hat{\xi}(\hat{\eta}), \hat{\eta})$ où $\hat{\eta}$ est solution de l'équation

$$\frac{1}{\eta} = \frac{1}{N_u}\left(1 + \frac{1}{\hat{\xi}}\right) \sum_{i=1}^{N_u} \frac{Y_i}{1 + \eta Y_i}$$

Il est difficile d'avoir l'expression explicite de la solution de cette dernière équation, on fait alors appel à des méthodes de résolution numérique telles que l'algorithme de Newton-Raphson . Davison et Smith le font remarquer que l'estimateur du maximum de vraisemblance est cependant peu utilisé en pratique car il pose des problèmes numériques et il est peu performant sur des échantillons de petite taille ($n < 500$). Dans ce cas passons donc à l'estimation de ces paramètres par les deux autres méthodes.

fonction densité est donné pour $\xi \neq 0$ par :

$$g_{\xi,\mu,\sigma}(x) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}$$

avec $\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right] \geq 0$

Méthode de maximum de vraisemblance

Cette méthode souvent utilisée en statistique, offre des résultats asymptotiquement efficaces. Nous obtenons alors des estimateurs convergeant sous certaines conditions vers les vraies valeurs des paramètres. Considérons un échantillon $(X_1, X_2, ..., X_n)$ supposé indépendant identiquement distribué de densité de probabilité g_θ, où $\theta = (\xi, \mu, \sigma)$. L'expression de la vraisemblance est donc donnée par :

$$\begin{aligned} L\left((X_1, X_2, ..., X_n); \theta\right) &= \prod_{i=1}^{n} g_\theta(x_i) \\ &= \prod_{i=1}^{n} \frac{1}{\sigma}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp\{-\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\} \end{aligned}$$

Alors la fonction log-vraisemblance est donnée par

$$l\left((X_1, X_2, ..., X_n); \theta\right) = -n\log\sigma - \left(\frac{1+\xi}{\xi}\right)\sum_{i=1}^{n}\log\left(1+\xi\frac{x_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\left(1+\xi\frac{x_i-\mu}{\sigma}\right)^{-\frac{1}{\xi}}$$

Pour $\xi = 0$ on a

$$l\left((X_1, X_2, ..., X_n); \theta\right) = -n\log\sigma - \sum_{i=1}^{n}\exp\left(-\frac{x_i-\mu}{\sigma}\right) - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)$$

L'estimateur $\hat{\theta}$ est donc donné par le système d'équations suivant :

$$\begin{cases} \dfrac{\partial l((X_1, X_2, ..., X_n); \hat{\theta})}{\partial\theta} = 0, \\ \dfrac{\partial^2\left((X_1, X_2, \ldots, X_n); \hat{\theta}\right)}{\partial\theta^2} < 0 \end{cases}$$

Ce qui nous ramène au système d'équations suivant :

$$\begin{cases} \dfrac{\partial l\left((X_1, X_2, \ldots, X_n); \hat{\theta}\right)}{\partial\mu} = n + \sum_{i=1}^{n}\dfrac{x_i-\mu}{\sigma}\left[\exp(-\dfrac{x_i-\mu}{\sigma} - 1)\right] = 0 \\ \dfrac{\partial l\left((X_1, X_2, \ldots, X_n); \hat{\theta}\right)}{\partial\sigma} = n - \sum_{i=1}^{n}\exp\left(-\dfrac{x_i-\mu}{\sigma}\right) \qquad = 0 \end{cases} \quad (1)$$

Cette méthode s'avère assez complexe du fait que la résolution du système (1) est assez difficile et n'admet pas de solution explicite. Dans ce cas on fait appel à des méthodes d'optimisations numériques telles que l'algorithme de Newton-Raphson . L'estimateur du maximum de vraisemblance conduit à un estimateur non biaisé, asymptotiquement normal. De plus, la variance asymptotique est égale à la borne de Fréchet – Darmois – Cramer – Rao.

Correction de l'exercice 02

1) La densité de la v.a.r. générique dans ce modèle de la loi uniforme est :

$$f_\theta(x) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x).$$

La vraisemblance de l'échantillon $(x_1, \ldots, x_n)$ est alors :

$$\mathcal{L}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_\theta(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} \mathbb{1}_{[0,\theta]}(x_i)$$

$$= \frac{1}{\theta^n} \mathbb{1}_{[0,+\infty[}(\inf_{i=1,\ldots,n} x_i) \mathbb{1}_{[0,\theta]}(\sup_{i=1,\ldots,n} x_i) = \frac{1}{\theta^n} \mathbb{1}_{[0,+\infty[}(x_{(1)}) \mathbb{1}_{]-\infty,\theta]}(x_{(n)}).$$

2) La fonction $\theta \mapsto \mathcal{L}(x_1, \ldots, x_n; \theta)$ est nulle sur l'intervalle $]-\infty, x_{(n)}[$ et coïncide avec la fonction $1/\theta^n$ sur $[x_{(n)}, +\infty[$. Cette fonction n'est pas continue en $x_{(n)}$ (et donc pas dérivable). Ainsi elle n'est pas convexe sur $\mathbb{R}$. On ne peut donc appliquer le raisonnement habituel (recherche du zéro de la dérivée première).

Mais il apparait clairement que le maximum de la vraisemblance est atteint en $\theta = x_{(n)}$ puisque avant (strictement) ce point la vraisemblance est nulle, qu'en ce point elle prend la valeur

$$\mathcal{L}(x_1, \ldots, x_n; x_{(n)}) = \frac{1}{\left(x_{(n)}\right)^n}$$

et qu'après elle est décroissante. Ainsi l'estimateur du maximum de vraisemblance est

$$\hat{\theta}_n = X_{(n)}.$$

Correction de l'exercice 03

1. La densité est :

$f(x;\theta) = \theta a^{\theta} x^{-(\theta+1)} I_{[a,+\infty[}(x)$ qui ne peut être mis sous la forme de la classe exponentielle en raison de $I_{[a,+\infty[}(x)$.

2. L'estimateur des moments pour a est solution de $\frac{a\theta}{\theta-1} = \overline{X}$ si $\theta > 1$ soit $\widehat{a}^M = \frac{\theta-1}{\theta}\overline{X}$.

3.

$$\prod_{i=1}^{n} f(x_i;a) = \theta^n a^{n\theta} \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)} \prod_{i=1}^{n} I_{[a,+\infty[}(x_i),$$

or $\prod_{i=1}^{n} I_{[a,+\infty[}(x_i) = \prod_{i=1}^{n} I_{[a,+\infty[}(x_{(1)})$ où $x_{(1)} = \min\{x_1, \cdots, x_n\}$. Donc :

$$\prod_{i=1}^{n} f(x_i;a) = a^{n\theta} \prod_{i=1}^{n} I_{[a,+\infty[}(x_{(1)}) \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)} \theta^n$$

$X_{(1)}$ est exhaustive. Elle est minimale du fait qu'elle est de dimension 1.

4. La fonction de répartition de $X_{(1)}$ est $1-[1-F(x;a)]^n = 1-\left(\frac{a}{x}\right)^{n\theta}$ si $x \geq a$. Donc $X_{(1)}$ suit une loi de Pareto de paramètre de seuil a et de paramètre de forme $n\theta$, d'où $E\left(X_{(1)}\right) = \frac{n\theta}{n\theta-1}a$, $E\left(\frac{n\theta-1}{n\theta}X_{(1)}\right) = a$ et $(1-\frac{1}{n\theta})X_{(1)}$ est un estimateur sans biais pour a.

Correction de l'exercice 04

a) Estimateur de Hill

Cette méthode est seulement applicable dans le cas où $\xi > 0$, autrement dit dans le cas des distributions de type Fréchet.

Définition 9. Soit $X_1, X_2, \ldots, X_n$ un échantillon de variables aléatoires i.i.d de fonction de répartition $F \in \mathrm{MDA}(G_\xi)$ pour $\xi > 0$.
L'estimateur de Hill est défini par :

$$\xi_{(k,n)}^{Hill} = \frac{1}{k}\sum_{i=1}^{k} \log X_{(n-i+1,n)} - \log X_{(n-k,n)}, \quad \text{pour} \quad k \in \{1, 2, \ldots, n-1\}$$

Construction de l'estimateur de Hill

On peut construire l'estimateur de Hill par plusieurs approches, parmi lesquels on peut citer la méthode du maximum de vraisemblance. On considère une suite de variables aléatoires (i.i.d) de loi de Pareto de paramètre $\alpha > 0$, de fonction de répartition donnée par

$$F(x) = 1 - x^{-\alpha} \quad \text{pour} \quad x \geq 1$$

alors

$$f(x) = \alpha x^{-\alpha-1} \quad \text{si} \quad x \geq 1$$

On veut trouver l'estimateur du maximum de vraisemblance (EMV) de α. Pour cela, on donne $L(.)$ la fonction de vraisemblance et les dérivées du logarithme de la fonction vraisemblance par rapport à α

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; \alpha) &= \prod_{i=1}^{n} f(x_i) \\ &= n\alpha \left(\prod_{i=1}^{n} x_i^{-1-\alpha}\right) \end{aligned}$$

$$\frac{\partial \log L(x_1, x_2, \ldots, x_n, \alpha)}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^{n} \log x_i$$

$$\frac{\partial^2 L(x_1, x_2, \ldots, x_n, \alpha)}{\partial \alpha^2} = -\frac{n}{\alpha^2} < 0$$

L'estimateur du maximum de vraisemblance de $\frac{1}{\alpha}$ est donc donnée par la statistique :

$$\hat{T} = \frac{1}{n}\sum_{i=1}^{n} \log X_i = \frac{1}{n}\sum_{i=1}^{n} \log X_{(i,n)}$$

Une généralisation concerne

$$F(x) = Cx^{-\alpha}, \quad x > 0$$

Si on pose $C = u^{\alpha}$,avec $x \geq u > 0$, on obtient donc l'EMV de α,

$$\hat{\alpha} = \left(\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{X_{(i,n)}}{u}\right)\right)^{-1} = \left(\frac{1}{n}\sum_{i=1}^{n} \log X_{(i,n)} - \log u\right)^{-1}$$

Souvent, on n'a pas l'expression de la fonction de répartition mais dans le domaine maximal d'attraction de Fréchet, on sait que F suit la loi de Pareto à partir d'un certain seuil connu u.
Soit $K = card\left\{i = 1, 2, \ldots, n : X_{(i,n)} > u\right\}$.
Conditionnellement, à l'événement $\{K = k\}$, les estimateurs du maximum de vraisemblance des paramètres α et C associés à un échantillon de loi commune définie par :

$$f_{(X_{(1,n)}, X_{(2,n)}, \ldots, X_{(k,n)})}(x_1, x_2, \ldots, x_n) = \frac{n!}{(n-k)!}\left(1 - Cx^{-\alpha}\right)^{n-k} C^k \alpha^k \prod_{i=1}^{n} x_i^{-(\alpha+1)}$$

où $f_{(X_{(1,n)}, X_{(2,n)}, \ldots, X_{(n,n)})}(.)$ est la loi conjointe du vecteur $(X_{(1,n)}, X_{(2,n)}, \ldots, X_{(n,n)})$ définie pour $x_1 > x_2 > \ldots > x_k > u$.

Après calcul, les estimateurs $\hat{\alpha}$ et $\hat{C}$ respectivement de α et C sont :

$$\hat{\xi}^{Hill}_{(k,n)} = \frac{1}{n}\sum_{i=1}^{n} \log \frac{X_{(i,n)}}{X_{k,n}} = \frac{1}{n}\sum_{i=1}^{n} \log X_{(i,n)} - \log X_{(k,n)}$$

$$\hat{C}_{(k,n)} = \frac{k}{n} X_{(k,n)}^{\hat{\xi}^{Hill}_{(k,n)}}$$

b) Estimateur de Pickands

L'avantage de l'estimateur de Pickands est qu'il est valable quel que soit le domaine d'attraction de la distribution et par conséquent, du domaine de définition de l'indice des valeurs extrêmes.

Définition 10. Soit $(X_i, i = 1, 2, \ldots, n)$ une suite de variables aléatoires indépendantes de loi F appartenant à l'un des domaines d'attractions. Soit une suite $(k(n))_{(n\geq 1)}$ d'entiers avec $1 \leq k(n) \leq n$, on définit l'estimateur de Pickands par :

$$\hat{\xi}^P_{(k(n),n)} = \frac{1}{\log 2} \log \frac{(X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)})}{(X_{(n-2k(n)+1)} - X_{(n-4k(n)+1,n)})} \quad , \quad \xi \in \mathbb{R}$$

Construction de l'estimateur de Pickands

Soit $U(x) = \left(\frac{1}{1-F(x)}\right)^{-1}$ et posons $\varphi_t(x) = \int_1^x u^{t-1}du$, $x > 0$ et $t \in \mathbb{R}$.
De Haan montre que la relation (1.4) est vérifiée si et seulement si

$$\lim_{t\to\infty} = \frac{U(tx) - U(t)}{U(ty) - U(t)} = \frac{\varphi_\xi(x)}{\varphi_\xi(y)} \quad \text{pour tout} \quad x, y > 0 \quad \text{et} \quad y \neq 1 \tag{2.8}$$

En remplaçant maintenant dans (2.8) U par $\hat{U}_n = \left[\frac{1}{1-F_n}\right]^{-1}$ (où F_n est la fonction de répartition empirique), t par $\frac{n}{2k(n)}$, x par $\frac{1}{2}$ et y par 2, on a pour n suffisamment grand

$$2^{-\xi}\left(\frac{X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)} - X_{(n-4k(n)+1,n)}}\right) = 1 \tag{2.9}$$

L'estimateur de Pickands $\hat{\xi}^P_{(k(n),n)}$ est alors la solution de l'équation (2.9).
Il vient

$$\log\left[2^{-\hat{\xi}^P_{(k(n),n)}}\left(\frac{X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)} - X_{(n-4k(n)+1,n)}}\right)\right] = 0$$

$$\hat{\xi}^P_{(k(n),n)} \log 2 = \log\left(\frac{X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)} - X_{(n-4k(n)+1,n)}}\right)$$

D'où on a

$$\hat{\xi}^P_{(k(n),n)} = \frac{1}{\log 2}\log\left(\frac{X_{(n-k(n)+1,n)} - X_{(n-2k(n)+1,n)}}{X_{(n-2k(n)+1,n)} - X_{(n-4k(n)+1,n)}}\right)$$

Une explication à cela est que $\hat{\xi}^P_{(k(n),n)}$, qui utilise l'information apportée par des distances entre deux statistiques d'ordres n'utilise pas le maximum de l'échantillon $X_{(n,n)}$, ce qui implique une perte d'information sur la queue de distribution.

Correction de l'exercice 05

1) Sur $[0, \infty[$, $\bar{F}^*(\cdot)$ est évidemment décroissante puisque $F(\cdot)$ est une fonction de répartition (et donc une fonction croissante). De plus, $\bar{F}^*(x) = x^{1/\gamma}L(x)$ qui est à variations régulières d'indice $1/\gamma$.

2) En résolvant l'équation $\bar{F}^*(x) = \alpha$ c'est-à-dire l'équation $F(x_F - 1/x) = 1/\alpha$ on a bien que $(\bar{F}^*)^{\leftarrow}(\alpha) = (x_F - F^{\leftarrow}(1-\alpha))^{-1}$. D'après le rappel précédent, on sait qu'il existe une fonction à variations lentes $\ell^*(\cdot)$ telle que

$$(\bar{F}^*)^{\leftarrow}(1/x) = x^{-\gamma}\ell^*(x) = (x_F - F^{\leftarrow}(1-\alpha))^{-1}.$$

En posant $\alpha = 1/x$, on en déduit le résultat annoncé avec $\ell(\cdot) = 1/\ell^*(\cdot)$.

3) On sait que

$$\{X_1, \ldots, X_n\} \stackrel{\mathcal{L}}{=} \{F^{\leftarrow}(U_1), \ldots, F^{\leftarrow}(U_n)\} = \{x_F - (1-U_i)^{-\gamma}\ell((1-U_i)^{-1}),\ i = 1, \ldots, n\}.$$

En rangeant par ordre décroissant, il vient

$$\{X_{n-i+1,n},\ i = 1, \ldots, n\} \stackrel{\mathcal{L}}{=} \{x_F - (1-U_{i,n})^{-\gamma}\ell((1-U_{i,n})^{-1}),\ i = 1, \ldots, n\}.$$

4) La question 3) implique que

$$\left\{\frac{x_F - X_{n-i+1,n}}{x_F - X_{n-k,n}},\ i = 1, \ldots, k\right\} \stackrel{\mathcal{L}}{=} \left\{\frac{U_{i,n}^{-\gamma}}{U_{k+1,n}^{-\gamma}}\frac{\ell(U_{i,n}^{-1})}{\ell(U_{k+1,n}^{-1})},\ i = 1, \ldots, k\right\}.$$

En prenant le logarithme et en sommant sur $i = 1, \ldots, k$, on obtient le résultat.

5) La représentation de Rényi implique que

$$\left\{\frac{U_{i,n}^{-1}}{U_{k+1,n}^{-1}},\ i = 1, \ldots, k\right\} \stackrel{\mathcal{L}}{=} \left\{\frac{T_{k+1}}{T_i},\ i = 1, \ldots, k\right\} \stackrel{\mathcal{L}}{=} \{U_{i,k}^{-1},\ i = 1, \ldots, k\}.$$

En sommant, on a donc

$$\sum_{i=1}^{k} \log\frac{U_{i,n}^{-1}}{U_{k+1,n}^{-1}} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{k} -\log(U_{i,k}) \stackrel{\mathcal{L}}{=} \sum_{i=1}^{k} -\log(1 - U_{k-i+1,k}) = \sum_{i=1}^{k} -\log(1 - U_i).$$

Comme $-\log(1 - \cdot)$ est l'inverse de la fonction de répartition d'une loi exponentielle standard, on trouve le résultat annoncé.

6) C'est direct en utilisant la loi faible des grands nombres.
Pour $x > 0$, on a

$$\begin{aligned}\mathbb{P}(n^{-\gamma}(x_F - X_{n,n}) \leq x) &= \mathbb{P}(X_{n,n} \geq x_F - n^{\gamma}x) = 1 - F^n(x_F - n^{\gamma}x) \\ &= 1 - \exp\left\{n \log\left[1 - \frac{x^{-1/\gamma}}{n} L\left(\frac{n^{-\gamma}}{x}\right)\right]\right\}.\end{aligned}$$

Comme

$$\frac{x^{-1/\gamma}}{n} L\left(\frac{n^{-\gamma}}{x}\right) \sim \frac{1}{n} c x^{-1/\gamma} \to 0,$$

on en déduit que

$$\mathbb{P}(n^{-\gamma}(x_F - X_{n,n}) \leq x) \to 1 - \exp(-c x^{-1/\gamma}),$$

pour $x > 0$. On déduit aisément de ce résultat que $X_{n,n} \xrightarrow{\mathrm{P}} x_F$.

On propose d'estimer γ par

$$\hat{\gamma}_n := \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{X_{n,n} - X_{n-i+1,n}}{X_{n,n} - X_{n-k,n}}\right).$$

Cet estimateur est connu dans la littérature sous le nom de *negative Hill estimator.*