Estimation et Identification statistiques
Contrôle Année 2020-2021
Correction

Les exercices sont indépendants.

Exercice 1 *Estimation fréquentiste*

1. On a :
$$E[Z_m] = \frac{1}{n}\sum_{i=1}^{n} E[Z_i] = \frac{1}{n}\sum_{i=1}^{n} \frac{a\theta}{a-1} = \frac{a\theta}{a-1}. \tag{1}$$
Ainsi $E[KZ_m] = KE[Z_m] = K\frac{a\theta}{a-1}$, et l'estimateur KZ_m est donc sans biais *si et seulement si* $K = \frac{a-1}{a}$. Dans ce cas, on obtient :
$$Var(KZ_m) = K^2 Var(Z_m) = \frac{(a-1)^2}{a^2}\frac{1}{n^2}\sum_{i=1}^{n} Var(Z_i) = \frac{(a-1)^2}{a^2}\frac{1}{n}\frac{a\theta^2}{(a-1)^2(a-2)} = \frac{\theta^2}{na(a-2)} \tag{2}$$
Ainsi : $\lim\limits_{n\to+\infty} Var(KZ_m) = 0$.
KZ_m est donc un estimateur sans biais à variance asymptotiquement nulle. D'après une propriété vue en cours, on peut en déduire qu'il s'agit d'un estimateur convergent.
2. Calcul de la fonction de répartition G de $\tilde{Z}$ en fonction de celle de Z :
$$\begin{aligned} G(x) &= P(\tilde{Z} < x) = 1 - P(\tilde{Z} > x) = 1 - P(min(Z_1, ..., Z_n) > x) = 1 - P(Z_1 > x, ..., Z_n > x) \\ &= 1 - \prod_{i=1}^{n} P(Z_i > x) = 1 - \prod_{i=1}^{n} \big(1 - P(Z_i < x)\big) = 1 - \big(1 - F(x)\big)^n \end{aligned} \tag{3}$$
En remplaçant F par son expression, on trouve : $G(x) = 1 - \left(\frac{\theta}{x}\right)^{an}$, pour $x \geq \theta$.
Ainsi, la fonction densité de probabilité s'écrit pour $x \geq \theta$:
$$g(x) = G'(x) = \frac{an\theta^{an}}{x^{an+1}}. \tag{4}$$
On reconnaît ici la densité de probabilité d'une loi de Pareto de paramètres (θ, na), et on en déduit donc $E[\tilde{Z}] = \frac{na\theta}{na-1}$, et $Var(\tilde{Z}) = \frac{na\theta^2}{(na-1)^2(na-2)}$.
Alternativement, on peut également déterminer espérance et variance par les calculs habituels. L'espérance de l'estimateur $\tilde{Z}$ s'écrit alors :
$$E[\tilde{Z}] = \int_{\theta}^{+\infty} xg(x)dx = an\theta^{an}\int_{\theta}^{+\infty} \frac{1}{x^{an}}dx = an\theta^{an}\frac{1}{(an-1)\theta^{an-1}} = \frac{an\theta}{an-1}. \tag{5}$$
Son moment d'ordre deux s'écrit :
$$E[\tilde{Z}^2] = \int_{\theta}^{+\infty} x^2 g(x)dx = an\theta^{an}\int_{\theta}^{+\infty} \frac{1}{x^{an-1}}dx = an\theta^{an}\frac{1}{(an-2)\theta^{an-2}} = \frac{an\theta^2}{an-2}. \tag{6}$$
D'où l'on déduit la variance :
$$Var(\tilde{Z}) = E[\tilde{Z}^2] - E[\tilde{Z}]^2 = \frac{an\theta^2}{an-2} - \frac{a^2n^2\theta^2}{(an-1)^2} = \frac{an}{(an-2)(an-1)^2}\theta^2 \tag{7}$$

L'estimateur $K'\tilde{Z}$ a pour espérance $E[K'\tilde{Z}] = K'E[\tilde{Z}] = K'\frac{an\theta}{an-1}$. Ainsi, cet estimateur est sans biais *si et seulement si* $K' = \frac{an-1}{an}$. Dans ce cas, sa variance sera : $Var(K'\tilde{Z}) = K'^2 Var(\tilde{Z}) = \frac{(an-1)^2}{(an)^2}\frac{an}{(an-2)(an-1)^2}\theta^2 = \frac{\theta^2}{an(an-2)}$. Ainsi : $\lim\limits_{n\to+\infty} Var(K'\tilde{Z}) = 0$.

$K'\tilde{Z}$ est donc un estimateur sans biais à variance asymptotiquement nulle. D'après une propriété vue en cours, on peut en déduire qu'il s'agit d'un estimateur convergent.

3. On a :

$$Var(KZ_m) - Var(K'\tilde{Z}) = \frac{\theta^2}{na(a-2)} - \frac{\theta^2}{na(na-2)} = \frac{(n-1)a}{na(na-2)(a-2)}\theta^2. \tag{8}$$

Comme $a > 2$, on en déduit que $Var(KZ_m) > Var(K'\tilde{Z})$ pour $n > 1$, et $Var(KZ_m) = Var(K'\tilde{Z})$ pour $n = 1$. Ainsi, l'estimateur $K'\tilde{Z}$ est plus efficace que l'estimateur KZ_m.

4. Les Z_i étant tous indépendants, la vraisemblance s'écrit :

$$L(a; z_1, ..., z_n) = \prod_{i=1}^{n} f(z_i) = \frac{a^n\theta^{na}}{\left(\prod\limits_{i=1}^{n} z_i\right)^{a+1}} \tag{9}$$

Ainsi, la log-vraisemblance s'écrit :

$$\ln L(a; z_1, ..., z_n) = n\ln(a) + na\ln(\theta) - (a+1)\sum_{i=1}^{n}\ln(z_i) \tag{10}$$

Le calcul des dérivées par rapport à a donne :

$$\begin{cases} \frac{\partial \ln L(a;z_1,...,z_n)}{\partial a} & = \frac{n}{a} + n\ln(\theta) - \sum\limits_{i=1}^{n}\ln(z_i) \\ \frac{\partial^2 \ln L(a;z_1,...,z_n)}{\partial a^2} & = -\frac{n}{a^2} \end{cases} \tag{11}$$

Ainsi, la dérivée première s'annule pour une unique valeur $a = \frac{n}{-n\ln(\theta)+\sum\limits_{i=1}^{n}\ln(z_i)} = \frac{n}{\sum\limits_{i=1}^{n}\ln(z_i/\theta)}$, et pour cette valeur de a, la dérivée seconde est strictement négative. On en déduit l'existence de l'estimateur de a au sens du maximum de vraisemblance donné par $T(Z) = \frac{n}{\sum\limits_{i=1}^{n}\ln(Z_i/\theta)}$.

5. En utilisant les deux résultats proposées, on en déduit que $T(Z) = \frac{n}{X}$ où $X \sim \Gamma(n, a)$. Ainsi, en appliquant le théorème de transfert, on en déduit :

$$\begin{aligned} & E[T(Z)] = nE[\tfrac{1}{X}] = n\int\limits_{\mathbb{R}} \tfrac{1}{x} f_X(x)dx \\ & = n\int\limits_{0}^{+\infty} \tfrac{1}{x}\tfrac{a^n}{(n-1)!}x^{n-1}e^{-ax}dx \\ & = \tfrac{na^2}{(n-1)(n-2)}\int\limits_{0}^{+\infty} x\tfrac{a^{(n-2)}}{((n-2)-1)!}x^{(n-2)-1}e^{-ax}dx, \end{aligned} \tag{12}$$

où l'on reconnait que l'intégrale est l'espérance d'une loi Gamma de paramètres $(n-2, a)$. Ainsi : $E[T(z)] = \frac{na^2}{(n-1)(n-2)}\frac{n-2}{a} = \frac{n}{n-1}a$. Et donc $K''T(z)$ est un estimateur sans biais de a si et seulement si $K'' = \frac{n-1}{n}$.

Exercice 2 *Estimation bayésienne* (4 points)

1. A partir de $z = \theta + b$, on en déduit $z - \theta = b$ dont $z - \theta$ a la loi de b soit une densité de probabilité égale à $f(z-\theta) = \frac{z-\theta}{2}$ différente de 0 si $z - \theta \in [0,\ 2]$, Donc $f_{z|\theta} = \frac{z-\theta}{2}$ avec $\theta \in [z-2,\ z]$. $f_\theta = 2\theta$ par hypothèse . Par la formule de Bayes, $f_{\theta|z} = \frac{f_{z|\theta}f_\theta}{f_z}$ soit $f_{\theta|z} = \frac{\frac{z-\theta}{2}2\theta}{f_z} = \frac{(z-\theta)\theta}{f_z}$ avec $\theta \in [\max(z-2, 0),\ \min(z, 1)]$

2. — Le risque moyen en valeur absolue s'écrit : $r_{MVA} = \underset{\theta}{\mathrm{E}}|\theta - \hat{\theta}|$ ce qui s'écrit $r_{MVA} = \underset{\theta|Z}{\mathrm{E}}|\theta - \hat{\theta}|$ soit $\int_z^{+\infty} |\theta - \hat{\theta}| f(\theta|Z) d\theta$.

   Pour minimiser ce risque, on cherche $\hat{\theta}$ tel que $\frac{\partial}{\partial \hat{\theta}} \int_{\max(z-2,0)}^{\min(z,1)} |\theta - \hat{\theta}|(z-\theta)\theta d\theta = 0$, soit $\frac{\partial}{\partial \hat{\theta}} \int_{0.5}^{\hat{\theta}} |\theta - \hat{\theta}|(z-\theta)\theta d\theta + \frac{\partial}{\partial \hat{\theta}} \int_{\hat{\theta}}^{1} |\theta - \hat{\theta}|(z-\theta)\theta d\theta = 0$ qui se réduit à $-\frac{\partial}{\partial \hat{\theta}} \int_{0.5}^{\hat{\theta}} (\theta - \hat{\theta})(z-\theta)\theta d\theta + \frac{\partial}{\partial \hat{\theta}} \int_{\hat{\theta}}^{1} (\theta - \hat{\theta})(z-\theta)\theta d\theta = 0$ soit $\hat{\theta}$ doit satisfaire $(\frac{2.5\theta^2}{2} - \frac{\theta^3}{3})2 - (\frac{2.5 \times 0.5^2}{2} - \frac{0.5^3}{3}) = (\frac{2.5^2}{2} - \frac{2.5^3}{3})$, soit $\hat{\theta} \approx 0.76$.

   — Le risque quadratique s'écrit : $r_{MVA} = \underset{\theta}{\mathrm{E}}(\theta - \hat{\theta})^2$. Par le même raisonnement que précédemment, $\hat{\theta}$ minimisant ce risque doit satisfaire $\frac{\partial}{\partial \hat{\theta}} \int_{0.5}^{1} (\theta - \hat{\theta})^2 (z-\theta)\theta d\theta = 0$ soit $2\int_{0.5}^{1} (\theta - \hat{\theta})(z-\theta)\theta d\theta = 0$ soit $\hat{\theta} \int_{0.5}^{+1} (z-\theta)\theta d\theta = \int_{0.5}^{1} \theta(z-\theta)\theta d\theta$ ou encore $\hat{\theta} = \frac{1}{0.6458} \int_{0.5}^{1} \theta(2.5-\theta)\theta d\theta$. On retrouve le résultat classique sur l'estimateur bayésien du risque quadratique, à savoir que $\hat{\theta}$ est l'espérance a posteriori, qui vaut ici 0.79.

   — On retrouve également le résultat classique de correspondance entre l'estimateur du risque moyen quadratique et l'estimateur de la moyenne a posteriori.

   — Pour le maximum de vraisemblance a posteriori, l'estimateur doit maximiser $f(\theta|z)$ donc maximiser $(2.5-\theta)\theta$ avec $\theta \in [0.5\ 1]$ donc $\hat{\theta} = 1$

3. Comparer ces résultats. Quel est selon vous le meilleur estimateur ? Les deux premiers estimateurs correspondent à une valeur intérieure à l'intervalle donc on peut considérer qu'on a acquis de l'information avec le risque. Le maximum de vraisemblance a posteriori est à une des bornes de l'intervalle, donc moins informatif

Correction alternative

1. On a :
$$f_{Z|\Theta}(z|\theta) = \frac{z-\theta}{2} \mathbf{1}_{[0,2]}(z-\theta) = \frac{z-\theta}{2} \mathbf{1}_{[z-2,z]}(\theta). \tag{13}$$
   Ainsi, la loi de Bayes nous donne :
$$f_{\Theta|Z}(\theta|z) = \frac{f_{Z|\Theta}(z|\theta) f_\Theta(\theta)}{f_Z(z)} = \frac{(z-\theta)\theta}{f_Z(z)} \mathbf{1}_{[z-2,z]}(\theta) \mathbf{1}_{[0,1]}(\theta) = \frac{(z-\theta)\theta}{f_Z(z)} \mathbf{1}_{[max(z-2,0),min(z,1)]}(\theta) \tag{14}$$

2. Dans cette question, nous avons $z = 2.5$. Ainsi :
$$f_{\Theta|Z}(\theta|z) = \frac{(2.5-\theta)\theta}{f_Z(2.5)} \mathbf{1}_{[0.5,1]}(\theta) \tag{15}$$
   — L'estimateur au sens du risque quadratique moyen s'écrit :
$$\hat{\theta}^{RQM} = \arg\min_T E_\Theta[(\theta - T)^2 | Z = z] = \arg\min_T \int_{-\infty}^{+\infty} (\theta - T)^2 f_{\Theta|Z}(\theta|z) d\theta = \arg\min_T \int_{0.5}^{1} (\theta - T)^2 \frac{(2.5-\theta)\theta}{f_Z(2.5)} d\theta \tag{16}$$
   — De même, l'estimateur au sens du risque en valeur absolue a posteriori s'écrit :
$$\hat{\theta}^{RQM} = \arg\min_T E_\Theta[|\theta - T| | Z = z] = \arg\min_T \int_{-\infty}^{+\infty} |\theta - T| f_{\Theta|Z}(\theta|z) d\theta = \arg\min_T \int_{0.5}^{1} |\theta - T| \frac{(2.5-\theta)\theta}{f_Z(2.5)} d\theta \tag{17}$$
   — Enfin, l'estimateur au sens de la moyenne a posteriori s'écrit :
$$\hat{\theta}^{RQM} = E_\Theta[\theta | Z = z] = \int_{-\infty}^{+\infty} \theta f_{\Theta|Z}(\theta|z) d\theta = \int_{0.5}^{1} \theta \frac{(2.5-\theta)\theta}{f_Z(2.5)} d\theta \tag{18}$$
   Cet estimateur est égal à l'estimateur du risque quadratique moyen (cf. cours).

— L'estimateur du maximum de vraisemblance a posteriori s'écrit :

$$\hat{\theta}^{MAP} = \arg\max_{\theta}\{f_{\Theta|Z}(\theta|z)\}. = \arg\max_{\theta}\{\frac{(2.5-\theta)\theta}{f_Z(2.5)}\mathbf{1}_{[0.5,1]}(\theta)\} = \arg\max_{\theta\in[0.5,1]}\{\frac{(2.5-\theta)\theta}{f_Z(2.5)}\} \quad (19)$$

En calculant les dérivées successives de $f_{\Theta|Z}(\theta|z)$ par rapport à θ, on obtient :

$$\begin{cases} \frac{\partial f_{\Theta|Z}(\theta|z)}{\partial\theta} & = \frac{z-2\theta}{f(z)} \\ \frac{\partial^2 f_{\Theta|Z}(\theta|z)}{\partial\theta^2} & = \frac{-2}{f(z)} \end{cases} \quad (20)$$

La valeur de θ annulant la dérivée première est $\theta = z/2 = 1.25$.
On déduit des résultat précédents que la fonction $f_{\Theta|Z}(\theta|z)$ est nulle sur l'intervalle $]-\infty;0.5[$, croissante, continue, et positive sur l'intervalle $[0.5;1]$, et nulle sur l'intervalle $]1;+\infty[$. Ainsi, le maximum de vraisemblance est atteint pour la valeur : $\hat{\theta}^{MAP} = 1$.

Exercice 3 Un système est modélisé sous la forme $z_n = \theta u_n + b_n$, pour $1 \leq n \leq N$. θ doit être estimé à partir de N mesures, réalisations de v.a.rélles indépendantes $z_1, \ldots, z_N$. Les $b_n, 1 \leq n \leq N$ forment une séquence de v.a. i.i.d. suivant une loi gaussienne de moyenne nulle et de variance σ^2. La séquence $u_n, 1 \leq n \leq N$ est supposée connue (de façon exacte).

1. Pour tout $n \in \{1, ..., N\}$, on a : $Z_n \sim \mathcal{N}(\theta u_n, \sigma^2)$. D'où : $Z \sim \mathcal{N}(\theta U, \sigma^2 I_N)$. Ainsi :

$$f(Z|\theta) = \frac{1}{\sqrt{2\pi}^N\sigma^N}\exp\left(-\frac{1}{2\sigma^2}(Z-\theta U)^T(Z-\theta U)\right). \quad (21)$$

2. Compte-tenu de l'information de l'énoncé, on fait l'*a priori* suivant sur θ :

$$\theta \sim \mathcal{U}([1;5]). \quad (22)$$

Ainsi, $f(\theta) = \frac{1}{4}\mathbf{1}_{[1;5]}(\theta)$.

3. Par la loi de Bayes, on trouve :

$$f(\theta|Z) = \frac{f(\theta)f(Z|\theta)}{f(Z)} = \frac{1}{f(Z)}\frac{1}{4\sqrt{2\pi}^N\sigma^N}\exp\left(-\frac{1}{2\sigma^2}(Z-\theta U)^T(Z-\theta U)\right). \quad (23)$$

4. Allures des densités.
5. La log-vraisemblance s'écrit :

$$\ln(L(\theta|Z)) = \ln(f(Z|\theta) = \ln\left(\frac{1}{f(Z)}\frac{1}{4\sqrt{2\pi}^N\sigma^N}\right) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(z_n-\theta u_n)^2 \quad (24)$$

Ainsi :

$$\begin{cases} \frac{\partial\ln(L(\theta|Z))}{\partial\theta} & = \frac{1}{\sigma^2}\sum\limits_{n=1}^{N}(z_n-\theta u_n)u_n \\ \frac{\partial^2\ln(L(\theta|Z))}{\partial\theta^2} & = -\frac{1}{\sigma^2}\sum\limits_{n=1}^{N}u_n^2 \end{cases} \quad (25)$$

L'équation $\frac{\partial\ln(L(\theta|Z))}{\partial\theta} = 0$ donne une unique solution comme extremum : $\theta = \frac{\sum\limits_{n=1}^{N} z_n u_n}{\sum\limits_{n=1}^{N} u_n^2}$. De plus, sous l'hypothèse qu'au moins un des u_n soit non nul, $\frac{\partial^2\ln(L(\theta|Z))}{\partial\theta^2} < 0$, ce qui indique que l'extremum calculé est un maximum.
Finalement, on en déduit l'existence de l'estimateur du maximum de vraisemblance :

$$\hat{\theta}_N^{EMV} = \frac{\sum\limits_{n=1}^{N} Z_n u_n}{\sum\limits_{n=1}^{N} u_n^2}. \quad (26)$$

La recherche de l'estimateur du maximum *a posteriori* revient à calculer :

$$\hat{\theta}_N^{MAP} = \arg\max_\theta f(\theta|Z). \tag{27}$$

Les calculs sont identiques à ceux de l'estimateur du maximum de vraisemblance et conduisent au même résultat : les deux estimateurs sont égaux.

Exercice 4 *Sensibilité aux données aberrantes* (6 points)

1. L'estimé au sens du risque quadratique minimal correspond à $r_{MVA} = \underset{\theta}{\mathrm{E}}(\theta - \hat{\theta})^2$. Dans le cas considéré, on cherche $\hat{\theta}$ qui minimise $\sum(z_i - \hat{\theta})^2$.
2. L'estimé au sens du risque absolu minimal correspond à $r_{MVA} = \underset{\theta}{\mathrm{E}}|\theta - \hat{\theta}|$. Cet estimé correspond à la médiane de θ , lorsque N est impair, et les z classés par ordre croissant, cet estimé est égal à $z_{\frac{N+1}{2}}$, lorsque N est pair, cet estimé peut être défini comme $(z_{\frac{N}{2}} + z_{\frac{N}{2}+1})/2$.
3. En introduisant une valeur z aberrante, l'expression $\sum(z_i - \hat{\theta})^2$ va être fortement modifiée (de l'ordre de l'écart quadratique en la valeur normale et la valeur aberrante). La valeur de l'estimé qui minimise ce critère va donc être déplacée par rapport à la valeur estimée initiale de l'ordre de cet écart. Par contre lorsque on introduit une valeur aberrante et qu'on utilise l'estimateur de l'écart en valeur absolue, la médiane sera soit inchangée si on conserve le même nombre de données de part et d'autre de la médiane précédente ou passer de la valeur courante $z_{\frac{N+1}{2}}$ à la valeur suivante ou précédente, ce qui ne représente pas une variation proportionnelle à l'écart entre la valeur normale et la valeur aberrante. Donc pour les deux estimateurs, c'est l'estimateur en valeur absolue qui sera le moins affecté par la donnée aberrante donc il est plus robuste.