ENSAE

ECOLE NATIONALE DE LA STATISTIQUE ET DE L'ADMINISTRATION ECONOMIQUE

Cursus Intégré
2004-2005

Rappels de statistique mathématique
Corrigé des travaux dirigés n°3

Guillaume Lacôte
Bureau **E03**
✉ Guillaume.Lacote@ensae.fr
☞ http://ensae.no-ip.com/SE222/

Corrigé de l'exercice 1

☞ Q1 On se donne Y variable aléatoire sur $\mathcal{M}_{N,1}(\mathbb{R})$ et X sur $\mathcal{M}_{N,K}(\mathbb{R})$.
La vraisemblance (conditionnelle) du modèle normal s'écrit

$$f(Y|X,b,\theta) = \frac{1}{(2\pi)^{\frac{N}{2}}\sqrt{|S(\theta)|}} \exp\left(-\frac{1}{2}(Y-Xb)'\cdot S(\theta)^{-1}\cdot(Y-Xb)\right)$$

Note : il s'agit bien d'un nombre réel car $(Y-Xb)\in\mathcal{M}_{N,1}(\mathbb{R})$.
La log-vraisemblance conditionnelle s'écrit donc

$$\ln L(Y|X,b,\theta) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|S(\theta)| - \frac{1}{2}(Y-Xb)'\cdot S(\theta)^{-1}\cdot(Y-Xb)$$

☞ Q2 Le vecteur score $\mathcal{S}(b,\theta)\in\mathcal{M}_{K+1,1}(\mathbb{R})$ se décompose entre les composantes suivant b, s
$\mathcal{S}_b(b)\in\mathcal{M}_{K,1}(\mathbb{R})$, et celle suivant θ, soit $\mathcal{S}_\theta(\theta)\in\mathbb{R}$, que nous calculons alternativement.

– Composantes suivant b :

$$\begin{aligned}
\mathcal{S}_b(b) &= \frac{\partial \ln L(Y|X,\cdot,\theta)}{\partial b}(b)\\
&= -\frac{1}{2}\left(\frac{\partial}{\partial b}\left(b\mapsto (Y-Xb)'\cdot S(\theta)^{-1}\cdot(Y-Xb)\right)\right)(b)\\
&= -\frac{1}{2}\left(\frac{\partial}{\partial b}\left(b\mapsto Y'S(\theta)^{-1}Y - b'X'S(\theta)^{-1}Y - Y'S(\theta)^{-1}Xb + b'X'S(\theta)^{-1}Xb\right)\right)(
\end{aligned}$$

Rappelons que pour $A\in\mathcal{M}_{n,p}(\mathbb{R})$ on a $\frac{\partial}{\partial b}(b\mapsto b'A) = A$ et $\frac{\partial}{\partial b}(b\mapsto A'b) = A'$, donc pou
symétrique

$$\frac{\partial}{\partial b}(b\mapsto b'\Sigma b)(b) = 2\Sigma b$$

Donc

$$\begin{aligned}
\mathcal{S}_b(b) &= -\frac{1}{2}\left(0 - X'S(\theta)^{-1}Y - X'S(\theta)^{-1}Y + 2(X'S(\theta)^{-1}X)b\right)\\
&= X'\cdot S(\theta)^{-1}\cdot(Y-Xb)
\end{aligned}$$

– Composante suivant θ :

$$\begin{aligned}
\mathcal{S}_\theta(\theta) &= \frac{\partial \ln L(Y|X,b,\cdot)}{\partial\theta}(\theta)\\
&= -\frac{1}{2}\left(\frac{\partial}{\partial\theta}\left(\theta\mapsto \ln|S(\theta)| + (Y-Xb)'\cdot S(\theta)^{-1}\cdot(Y-Xb)\right)\right)(\theta)
\end{aligned}$$

Or

$$\left(\frac{\partial}{\partial\theta}\left(\theta\mapsto\ln|S(\theta)|\right)\right)(\theta) = Tr\left(S(\theta)^{-1}\frac{\partial S}{\partial\theta}(\theta)\right)$$

et

$$\left(\frac{\partial}{\partial\theta}\left(\theta \mapsto S(\theta)^{-1}\right)\right)(\theta) = -S(\theta)^{-1}\cdot\frac{\partial S}{\partial\theta}(\theta)\cdot S(\theta)^{-1}$$

La première assertion découle en effet directement du théorème 2 du chapitre 3 de *"Matrix Differential Calculus"*, Jan R. Magnus et Heinz Neudecker, qui stipule que si F : $\left(\mathcal{M}_{n,p}(\mathbb{R}) \to \mathcal{S}_k^+(\mathbb{R})\right)$ est k fois différentiable, alors $\ln|F|$ l'est et admet pour différentielle $d\ln|F| = Tr\left(F^{-1}\right)dF$ (il suffit de poser $n = p = 1$ et $F = S$). Elle peut même être redémontrée directement, en notant λ_i^A la i-ième valeur propre d'une matrice A (sans ordre de multiplicité) :

$$\begin{aligned}
\left(\frac{\partial}{\partial\theta}\left(\theta \mapsto \ln|S(\theta)|\right)\right)(\theta) &= \left(\frac{\partial}{\partial\theta}\left(\theta \mapsto \ln\left(\Pi_{i=1}^n \lambda_i^{S(\theta)}\right)\right)\right)(\theta)\\
&= \sum_{i=1}^n \left(\frac{\partial}{\partial\theta}\left(\theta \mapsto \left(\ln\lambda_i^{S(\theta)}\right)\right)\right)(\theta)\\
&= \sum_{i=1}^n \left(\frac{\partial\lambda_i^{S(\theta)}}{\partial\theta}(\theta)\right)\left(\lambda_i^{S(\theta)}\right)^{-1}\\
&= \sum_{i=1}^n \left(\lambda_i^{\frac{\partial S(\theta)}{\partial\theta}(\theta)}\right)\left(\lambda_i^{S(\theta)^{-1}}\right)\\
&= Tr\left(\left(\frac{\partial S(\theta)}{\partial\theta}(\theta)\right)\times\left(S(\theta)^{-1}\right)\right)
\end{aligned}$$

Une preuve de la seconde assertion est également proposée dans le théorème 3 de ce même ouvrage, ou peut aussi être obtenue en développant l'égalité

$$\frac{\partial}{\partial\theta}\left(S(\theta)\times S(\theta)^{-1}\right) = \frac{\partial I_N}{\partial\theta} = (0)$$

Il en ressort que

$$\begin{aligned}
\mathcal{S}_\theta(\theta) &= -\frac{1}{2}Tr\left(S(\theta)^{-1}\frac{\partial S}{\partial\theta}(\theta)\right) + \frac{1}{2}\underbrace{(Y-Xb)'\left(S(\theta)^{-1}\cdot\frac{\partial S}{\partial\theta}(\theta)\cdot S(\theta)^{-1}\right)(Y-Xb)}_{\in\mathbb{R}}\\
&= -\frac{1}{2}Tr\left(\left(S(\theta)^{-1}\frac{\partial S}{\partial\theta}(\theta)S(\theta)^{-1}\right)\left(S(\theta)-(Y-Xb)'(Y-Xb)\right)\right)
\end{aligned}$$

Reste à montrer que le score est centré :

– On a d'un part $\mathbb{E}(Y|X,b) = Xb$,
  en conséquence de quoi $\mathbb{E}(Y - Xb|X,b) = (0)$
  de sorte que $\mathbb{E}(\mathcal{S}_b(b)|X,b) = (0)$: les composantes selon b sont centrées.
– Et d'autre part $\mathbb{E}_\theta\left((Y-Xb)(Y-Xb)'\right) = S(\theta)$,
  de sorte que $\mathbb{E}(\mathcal{S}_\theta(\theta)|X,b) = 0$.

En conséquence, $\mathbb{E}_{b,\theta}(\mathcal{S}(b,\theta)|X,b,\theta) = (0)$: le vecteur score est centré.

☞ Q3 Le calcul de la matrice d'information de Fisher est encore plus convivial, car c'est un prétexte au calcul des dérivées secondes de la log-vraisemblance (le calcul alternatif de la variance du score calculé est laissé en exercice).

– Dérivée seconde selon (b,b) :

$$\frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial b\partial b'}(\theta,b) = -X'\cdot S(\theta)^{-1}\cdot X$$

Donc, comme $S(0) = I_N$,

$$\mathbb{E}_{b,\theta=0}\left(\frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial b\partial b'}(\theta,b)\right) = -X'X$$

– Dérivée seconde selon (b,θ) :

$$\frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial b\partial\theta}(\theta,b) = -X'\cdot S(\theta)^{-1}\frac{\partial S}{\partial\theta}(\theta)S(\theta)^{-1}\cdot(Y-Xb)$$

Donc, comme $\mathbb{E}_{b,\theta}(Y-Xb) = 0$,

$$\mathbb{E}_{b,\theta=0}\left(\frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial b\partial\theta}(\theta,b)\right) = (0)$$

– Dérivée seconde selon (θ,b) :

Loins de nous lancer dans la dérivation selon b de $\frac{\partial\ln L(Y|X,\cdot,\cdot)}{\partial\theta}$, constatons astucieusement c
$(b,\theta)\mapsto\mathcal{S}(b,\theta)$ est de classe C^2 sur un voisinage de 0 : elle est en effet la composée d'u
fonction de classe C^∞, à savoir
$\left(\begin{array}{ccc}\mathbb{R}\times\mathcal{M}_{N,N}(\mathbb{R}) & \to & \mathbb{R}\\ (x,A) & \mapsto & -\frac{N}{2}\ln 2 - \frac{1}{2}x - \frac{1}{2}(Y-Xb)'\cdot A\cdot(Y-Xb)\end{array}\right)$, et des deux foncti
$(M\mapsto\ln|M|)$ et $(M\mapsto M^{-1})$ qui sont de classe C^∞ sur une partie dense d'un voisina
de 0. Cauchy nous assure alors de l'égalité des dérivées croisées pour θ voisin de
$\forall(b,\theta)$, $\frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial b\partial\theta}(\theta,b) = \frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial\theta\partial b}(\theta,b)$, ce qui permet de conclure que

$$\mathbb{E}_{b,\theta=0}\left(\frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial\theta\partial b}(\theta,b)\right) = (0)$$

– Dérivée seconde selon (θ,θ) :

$$\begin{aligned}
&\mathbb{E}_{b,\theta}\left(\frac{\partial^2\ln L(Y|X,\cdot,\cdot)}{\partial\theta^2}(\theta,b)\right)\\
&= \mathbb{E}_{b,\theta}\left(\frac{\partial}{\partial\theta}\left(\theta\mapsto -\frac{1}{2}Tr\left(\underbrace{\left(S(\theta)^{-1}\frac{\partial S}{\partial\theta}(\theta)S(\theta)^{-1}\right)}_{\psi(\theta)}\underbrace{\left(S(\theta)-(Y-Xb)'(Y-Xb)\right)}_{h(\theta)}\right)\right)\right)\\
&= -\frac{1}{2}\mathbb{E}_{b,\theta}\left(Tr\left(\frac{\partial\psi}{\partial\theta}(\theta)h(\theta) + \psi(\theta)\frac{\partial h}{\partial\theta}(\theta)\right)\right)\\
&= -\frac{1}{2}Tr\left(\mathbb{E}_{b,\theta}\left(\frac{\partial\psi}{\partial\theta}(\theta)h(\theta)\right)\right) - \frac{1}{2}\mathbb{E}_{b,\theta}\left(Tr\left(\psi(\theta)\frac{\partial h}{\partial\theta}(\theta)\right)\right)\\
&= -\frac{1}{2}Tr\left(\mathbb{E}_{b,\theta}\left(\frac{\partial\psi}{\partial\theta}(\theta)\underbrace{\mathbb{E}_b(h(\theta)|\theta)}_{=0}\right)\right) - \frac{1}{2}\mathbb{E}_{b,\theta}\left(Tr\left(\psi(\theta)\frac{\partial S}{\partial\theta}(\theta)\right)\right)\\
&= -\frac{1}{2}Tr\left(\psi(\theta)\frac{\partial S}{\partial\theta}(\theta)\right)
\end{aligned}$$

Donc
$$\mathbb{E}_{b,\theta}\left(-\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial\theta^2}(\theta,b)\right)=\frac{1}{2}Tr\left(S(\theta)^{-1}\ \frac{\partial S}{\partial\theta}(\theta)\ S(\theta)^{-1}\ \frac{\partial S}{\partial\theta}(\theta)\right)$$
et par suite, sachant que $S(0)=I_N$ et en notant $A=\frac{\partial S}{\partial\theta}(0)$
$$\mathbb{E}_{b,\theta=0}\left(-\frac{\partial^2 \ln L(Y|X,\cdot,\cdot)}{\partial\theta^2}(\theta,b)\right)=\frac{1}{2}Tr\left(A^2\right)$$

On a donc finalement :
$$\boxed{I(b,0)=\left(\begin{array}{c|c} & 0\\ X'X & \dots\\ & 0\\ \hline 0\dots0 & \frac{1}{2}Tr\left(A^2\right)\end{array}\right)}$$

☞ Q4 On a $S(\theta)=\left(\theta^{|i-j|}\right)_{(i,j)\in[\![1,N]\!]^2}$.

Or $\frac{\partial}{\partial\theta}\left(\theta\mapsto\theta^k\right)(\theta)=\begin{cases}0 & \text{si } k=0\\ 1 & \text{si } k=1\\ \theta^{k-1} & \text{sinon}\end{cases}$ En particulier $A=\frac{\partial S}{\partial\theta}(0)=\left(\mathbb{1}_{|i-j|=1}\right)_{(i,j)\in[\![1,N]\!]^2}$.

En conséquence
$$A^2=\begin{pmatrix}1&\star&\star&\dots&\star&\star\\ \star&2&\star&\dots&\star&\star\\ \star&\star&2&\dots&\star&\star\\ \vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\ \star&\star&\star&\dots&2&\star\\ \star&\star&\star&\dots&\star&1\end{pmatrix}$$

et donc $Tr\left(A^2\right)=2(N-1)$.
Ainsi,
$$\boxed{I_{b,\theta}(b,0)=\begin{pmatrix}X'X & 0\\ 0 & N-1\end{pmatrix}}$$

⋆
⋆ ⋆

Corrigé de l'exercice 2

☞ Q1 On a
$$f_X(x,p)=p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\mathbb{1}_{x\in\{0,1\}}$$
d'où on tire que
$$\frac{\partial \ln f_X}{\partial p}(x,p)=\frac{\sum_{i=1}^n x_i}{p}-\frac{n-\sum_{i=1}^n x_i}{1-p}$$
en conséquence de quoi un estimateur $\widehat{p}$ de p vérifie
$$(1-\widehat{p})\sum_{i=1}^n x_i=\widehat{p}(n-\sum_{i=1}^n x_i)$$
soit $\widehat{p}=\frac{1}{n}\sum_{i=1}^n x_i=\overline{X}$ (la réciproque étant immédiate).
On a par ailleurs
$$\frac{\partial^2 \ln f_X}{\partial p^2}(x,\theta)=-\frac{1}{p^2}\sum_{i=1}^n x_i-\frac{n-\sum_{i=1}^n x_i}{(1-p)^2}$$
d'où $I_n(\theta)=\frac{n}{p(1-p)}$ et donc d'après le Théorème Central Limite il vient
$$\sqrt{n}(\widehat{p}-p)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}\left(0,p(1-p)\right)$$

☞ Q2 Le paramètre à estimer est ici (m,σ^2) (et pas (m,σ) : on s'abstiendra d'écrire σ^4 mais plu $(\sigma^2)^2$). On a
$$f_X(x,m,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-m)^2}$$
et donc
$$\frac{\partial \ln f_X}{\partial m}(x,m,\sigma^2)=\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-m)$$
Par ailleurs,
$$\frac{\partial \ln f_X}{\partial(\sigma^2)}(x,m,\sigma^2)=-\frac{n}{2\sigma^2}+\frac{1}{2\left(\sigma^2\right)^2}\sum_{i=1}^n(x_i-m)^2$$
L'estimateur du maximum de vraisemblance $\begin{pmatrix}\widehat{m}\\ \widehat{\sigma^2}\end{pmatrix}$ de $\begin{pmatrix}m\\ \sigma^2\end{pmatrix}$ vérifie alors pour tout x
$$\frac{\partial \ln f_X}{\partial\begin{pmatrix}m\\ \sigma^2\end{pmatrix}}\left(\begin{pmatrix}\widehat{m}(x)\\ \widehat{\sigma^2}(x)\end{pmatrix}\right)=\begin{pmatrix}0\\ 0\end{pmatrix}$$
ce dont on tire
$$\begin{cases}\widehat{m}(x)=x\\ \frac{1}{\widehat{\sigma^2}(x)}\sum_{i=1}^n\left(x_i-\widehat{m}(x)\right)=0\end{cases}$$
soit encore
$$\begin{pmatrix}\widehat{m}\\ \widehat{\sigma^2}\end{pmatrix}:x\mapsto\begin{pmatrix}\overline{x}\\ \frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})^2\end{pmatrix}$$

On a enfin

$$\begin{aligned}
\frac{\partial^2 \ln f_X}{\partial m^2}(x,m,\sigma^2) &= -\frac{n}{\sigma^2} \\
\frac{\partial^2 \ln f_X}{\partial m \partial(\sigma^2)}(x,m,\sigma^2) &= \left(-\frac{1}{(\sigma^2)^2}\right)\sum_{i=1}^{n}(x_i - m) \\
\frac{\partial^2 \ln f_X}{\partial(\sigma^2)^2}(x,m,\sigma^2) &= \frac{n}{2\,(\sigma^2)^2} + \frac{1}{2}\sum_{i=1}^{n}(x_i - m)^2\left(-\frac{2}{(\sigma^2)^3}\right)
\end{aligned}$$

De $\mathbb{E}\left(X_i - m\right) = 0$ et $\mathbb{E}\left((X_i - m)^2\right) = \sigma^2$ on tire finalement que

$$\begin{aligned}
\mathbb{E}\left(\frac{\partial^2 \ln f_X}{\partial(\sigma^2)^2}(X;m,\sigma^2)\right) &= \frac{n}{2\,(\sigma^2)^2} - \frac{1}{2}\frac{2}{(\sigma^2)^3}\left(\sigma^2\right) \\
&= -\frac{n}{2\,(\sigma^2)^2}
\end{aligned}$$

de sorte que

$$I_n(m,\sigma^2) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{pmatrix}$$

et donc en définitive

$$\sqrt{n}\begin{pmatrix} \widehat{m} - m \\ \widehat{\sigma^2} - \sigma^2 \end{pmatrix} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2(\sigma^2)^2 \end{pmatrix}\right)$$

Les variables aléatoires (asymptotiquement normales) $\widehat{m}$ et $\widehat{\sigma^2}$ sont de covariance asymptotique nulle, donc[1] sont asymptotiquement indépendantes.

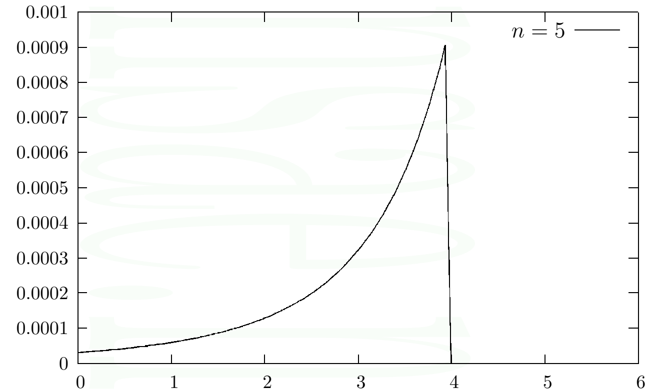
☞ Q3 La vraisemblance s'écrit :

$$L_X(x;a,b) = \frac{1}{(b-a)^n}\mathbb{1}_{\min x_i \geq a}\mathbb{1}_{\max x_i \leq b}$$

et n'est pas dérivable en a et b (disconstinuités en $a = \min x_i$ et $b = \max x_i$). Les conditions de régularité habituelles ne sont pas vérifiées, et le point $(\widehat{a}(x), \widehat{b}(x))$ qui maximise la vraisemblance ne vérifie pas les conditions du premier ordre.

Il peut néanmoins être déterminé directement : en effet pour $x \in \mathbb{R}^n$ et $b \in \mathbb{R}$ fixés, $\begin{pmatrix}]-\infty, b[& \to & \mathbb{R}^+ \\ a & \mapsto & \frac{1}{(b-a)^n}\mathbb{1}_{\min x_i \geq a}\mathbb{1}_{\max x_i \leq b} \end{pmatrix}$ est identiquement nulle si $\max_i x_i > b$, et sinon admet pour graphe

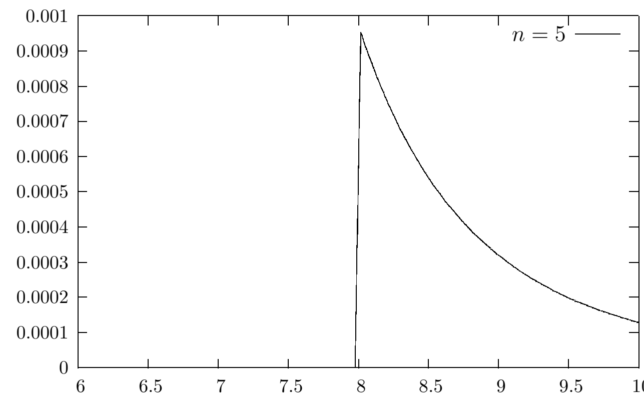
[1] Attention, propriété propre à la loi normale



Graphe de $a \mapsto L_X(x;a,b)$ lorsque $b = 8 > \max_i x_i$ et $\min_i x_i = 4$

Elle est donc dans tous les cas maximale en $\widehat{a}(x) = \min_i x_i$ (qui ne dépend pas de b).

De même pour $x \in \mathbb{R}^n$ et $a \in \mathbb{R}$ fixés, $\begin{pmatrix}]a, +\infty[& \to & \mathbb{R}^+ \\ b & \mapsto & \frac{1}{(b-a)^n}\mathbb{1}_{\min x_i \geq a}\mathbb{1}_{\max x_i \leq b} \end{pmatrix}$ est identiq
ment nulle si $\min_i x_i < a$, et sinon admet pour graphe



Graphe de $b \mapsto L_X(x;a,b)$ lorsque $a = 4 \leq \min_i x_i$ et $\max_i x_i = 8$

Elle est donc dans tous les cas maximale en $\widehat{b}(x) = \max_i x_i$ (qui ne dépend pas de a).

Par conséquent l'estimateur du maximum de vraisemblance de (a,b) est $x \mapsto (\min_i x_i, \max_i$
Cependant cette statistique ne vérifie pas les propriétés habituelles de l'e.m.v., à commencer
la normalité asymptotique. La loi limite du couple $(\widehat{a}, \widehat{b})$ doit donc être retrouvée 'à la ma

Cherchons tout d'abord pour $\alpha > 0$ la loi de $n^\alpha(\widehat{a} - a)$:

$$\begin{aligned}
\mathbb{P}(n^\alpha(\widehat{a}(X) - a) > t) &= \mathbb{P}(\min X_i > a + \frac{t}{n^\alpha}) \\
&= \mathbb{P}(X_1 > a + \frac{t}{n^\alpha})^n \\
&= \left(1 - \frac{t}{n^\alpha}(b-a)\right)^n \\
&= e^{n \ln\left(1 - \frac{t}{n^\alpha}(b-a)\right)} \\
&\xrightarrow[n\to+\infty]{} \begin{cases} 0 & \text{si } \alpha < 1 \\ e^{-t/(b-a)} & \text{si } \alpha = 1 \\ +\infty & \text{si } \alpha > 1 \end{cases}
\end{aligned}$$

En particulier, $\mathbb{P}(n(\widehat{a}(X) - a) > t) \xrightarrow[n\to+\infty]{} e^{-t/(b-a)}$ et donc la fonction de survie de $n(\widehat{a} - a) > t$ tend quand $n \to +\infty$ vers la fonction de survie d'une loi exponentielle $\mathcal{E}\left(\frac{1}{b-a}\right)$. Rappelons que la convergence de la fonction de répartition en tout point t de continuité (et donc de la loi de survie) entraîne la convergence en loi. On montre de la même façon (attention au sens de l'inégalité) que

$$n(b - \widehat{b}(X)) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}(1/(b-a))$$

Calculons enfin la loi jointe de $n\begin{pmatrix} \widehat{a} - a \\ b - \widehat{b} \end{pmatrix}$: pour ce faire calculons de même

$$\begin{aligned}
\mathbb{P}\left(n(\widehat{a}(X) - a) > t \wedge n(b - \widehat{b}(X)) > u\right) &= \mathbb{P}\left((\min_i X_i) > \frac{t}{n} + a \ \wedge\ (\max_i X_i) < b - \frac{u}{n}\right) \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} f_{(\min_i X_i, \max_i X_i)}(y,z)\ \mathbb{1}_{y > \frac{t}{b} + a \ \wedge\ z < b - \frac{u}{n}}\ dydz
\end{aligned}$$

La loi du couple $(\min_i X_i, \max_i X_i)$ s'obtient alors à partir de celle du n-uplet ordonné $(X_{(1)}, \ldots, X_{(n)})$, à savoir

$$l(y_1, \ldots, y_n) = n! \mathbb{1}_{y_1 < \ldots < y_n} f(y_1) \ldots f(y_n)$$

(Voir TD 2, exercice 1) en intégrant successivement (d'après le théorème de projection) suivant y_{n-1}, puis $y_{n-2}, \ldots$, et ainsi de suite jusqu'à intégrer selon y_2. Il vient en tout état de cause

$$f_{(\min_i X_i, \max_i X_i)}(y,z) = \frac{n!}{(n-2)!} f(y) \left(F_X(z) - F_X(y)\right)^{n-2} f(z) \mathbb{1}_{y<z}$$

En conséquence,

$$\begin{aligned}
&\mathbb{P}\left(n(\widehat{a}(X) - a) > t \ \wedge\ n(b - \widehat{b}(X)) > u\right) \\
=\ & \int_{a+\frac{t}{n}}^{b-\frac{u}{n}} \left(\int_y^{b-\frac{u}{n}} \frac{n!}{(n-2)!} \left(\frac{1}{b-a}\mathbb{1}_{[a,b]}(y)\right) \left(\frac{z-a}{b-a} - \frac{y-a}{b-a}\right)^{n-2} \left(\frac{1}{b-a}\mathbb{1}_{[a,b]}(z)\right) dz \right. \\
& \vdots \\
=\ & \left(1 - \frac{1}{n}\frac{y+z}{b-a}\right)^n \\
\xrightarrow[n\to+\infty]{}\ & e^{-\frac{y+z}{b-a}} \\
=\ & e^{-\frac{y}{b-a}} \times e^{-\frac{z}{b-a}}
\end{aligned}$$

qui est la loi jointe de deux exponentielles **indépendantes**.

<u>*Autre méthode (beaucoup plus simple) :*</u>

On peut directement déterminer la loi jointe de $n\begin{pmatrix} \widehat{a}(X) - a \\ b - \widehat{b}(X) \end{pmatrix}$:

$$\begin{aligned}
\mathbb{P}\left(n(\widehat{a}(X) - a) > t \wedge n(b - \widehat{b}(X)) < u\right) &= \mathbb{P}\left((\min_i X_i) > \frac{t}{n} + a \ \wedge\ (\max_i X_i) < b - \right. \\
&= \mathbb{P}\left(\forall i \in [\![1,n]\!],\ X_i \in \left]\frac{t}{n} + a\, ,\, b - \frac{u}{n}\right[\right) \\
&= \mathbb{P}\left(X_1 \in \left]\frac{t}{n} + a\, ,\, b - \frac{u}{n}\right[\right)^n \\
&= \left(\int_{\frac{t}{n}+a}^{b-\frac{u}{n}} \frac{1}{b-a} dx\right)^n \\
&= \left(\int_{\frac{t}{n}+a}^{b-\frac{u}{n}} \frac{1}{b-a} dx\right)^n \\
&= \left(1 - \frac{1}{n}\frac{t+u}{b-a}\right)^n \\
&\xrightarrow[n\to+\infty]{} e^{-\frac{t+u}{b-a}} \\
&= e^{-\frac{t}{b-a}} \times e^{-\frac{u}{b-a}}
\end{aligned}$$

se dont on déduit tout d'abord que les $\widehat{a}(X) - a$ et $b - \widehat{b}(X)$ sont indépendantes, et en ou qu'elles suivent une même loi exponentielle de paramètre $\frac{1}{b-a}$, ce qui s'écrit simplement

$$\boxed{n\begin{pmatrix} \widehat{a}(X) - a \\ b - \widehat{b}(X) \end{pmatrix} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{E}\left(\tfrac{1}{b-a}\right)^{\otimes 2}}$$

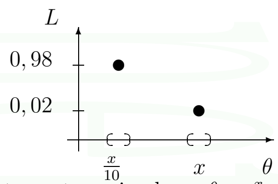
⋆
⋆ ⋆

Corrigé de l'exercice 3

☞ Q1 La vraisemblance s'écrit

$$L(x;\theta) = 0.98 \cdot \mathbb{1}_{10\theta}(x) \ + \ 0.02 \cdot \mathbb{1}_{\theta}(x)$$

Le graphe de $\theta \mapsto L(x;\theta)$ est alors, à x fixé



La vraisemblance est donc manifestement maximale en $\theta = \frac{x}{10}$.

Donc $\widehat{\theta} = \frac{x}{10}$ et $\mathbb{P}(\widehat{\theta}(X) = \theta) = \mathbb{P}\left(\frac{X}{10} = \theta\right) = \mathbb{P}(X = 10\theta) = \boxed{0.98}$

☞ Q2 La vraisemblance s'écrit cette fois

$$L(x;\theta) = 0.02 \cdot \mathbb{1}_{\theta}(x) + \sum_{i=1}^{980} 0.001 \cdot \mathbb{1}_{a_i\theta}(x)$$

et est maximale en $\theta = x$.

En conséquence, $\mathbb{P}(\widetilde{\theta} = \theta) = \mathbb{P}(X = \theta) = 0.02$ On a même $\mathbb{P}(\widetilde{\theta} \geq 10\theta) = \mathbb{P}(X \geq 10\theta) = 1 - \mathbb{P}(X = \theta) = 0.98$: on est presque certain de se tromper d'un facteur 10 !

Ce résultat est paradoxal, puisqu'il incite à penser qu'une même méthode d'estimation (en l'occurence, la maximisation de la vraisemblance) peut conduire, face à deux problèmes quasi-identiques, à un estimation satisfaisante dans un cas mais très décevante dans l'autre.

Plus inquiétant encore (sauf bien-sûr pour l'étudiant de l'ENSAE, qui est vif d'esprit et averti), supposons que les a_i soient inconnus et tirés indépendamment selon une même loi aléatoire $\mathcal{L}$, d'espérance 10.5 et (pour simplifier) à support borné. Alors l'e.m.v. reprendrait à nouveau une valeur beaucoup plus satisfaisante ($\widehat{\theta} = \frac{x}{10.5}$) : en d'autres termes, l'information supplémentaire "connaissance des a_i" se traduit par une estimation beaucoup moins bonne que si on ne disposait pas de cette information !

L'explication de ce paradoxe tient à la discontinuité de la vraisemblance, et à ce que l'estimateur par maximum de vraisemblance n'est justifié qu'asymptotiquement (lorsque $n \to +\infty$) ; cet exemple donne en fait un cas extrême d'écart entre la loi à distance finie et la loi asymptotique (ici $n = 1$, mais il faut noter qu'on peut aisément construire d'autres paradoxes de ce type pour un nombre quelconque d'observations n).

Cet exemple illustre le danger des méthodes dont la justification repose uniquement sur un comportement asymptotique, alors que le nombre de données dont on dispose est toujours fini en pratique.

⋆
⋆ ⋆