



Cursus Intégré  
2004-2005

Rappels de statistique mathématique  
*Réponses question par question des travaux  
dirigés n° 3*

Guillaume Lacôte  
Bureau **E03**

✉ [Guillaume.Lacote@ensae.fr](mailto:Guillaume.Lacote@ensae.fr)

👉 <http://ensae.no-ip.com/SE222/>

<b>Exercice corrigé 1</b>
---------------------------

Soit  $Y$  un vecteur aléatoire de taille  $N$  et  $X$  une matrice aléatoire à  $N$  lignes et  $K$  colonnes. Soit  $\theta \rightarrow S(\theta)$  une application de classe  $C^1$  définie sur un voisinage de 0 dans  $\mathbb{R}$ , à valeurs dans l'ensemble des matrices symétriques définies positives de taille  $N$ . On suppose que  $S(0) = I_N$  et on note :  $A = \frac{\partial S}{\partial \theta}(0)$ .

On rappelle par ailleurs que

$$\begin{aligned} \frac{\partial \ln(|S|)}{\partial \theta} &= \text{Tr} \left( S^{-1} S' \right) \\ \frac{\partial S^{-1}}{\partial \theta} &= -S^{-1} S' S^{-1} \end{aligned}$$

On considère le modèle linéaire (conditionnel) gaussien suivant :

$$\mathbb{P}_Y^X = \mathcal{N}(Xb, S(\theta))$$

où  $b \in \mathbb{R}^K$  et  $\theta \in \mathbb{R}$  sont les paramètres inconnus.

☞ Q1 

Ecrire la vraisemblance du modèle.
------------------------------------

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 

Ecrire le vecteur score (de taille $K + 1$ ) et vérifier qu'il est d'espérance nulle.
---

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

– Composantes suivant  $b$  :

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

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 pour  $\Sigma$  symétrique

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

Donc

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

– Composante suivant  $\theta$  :

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

Or

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

et

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

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  $F : (\mathcal{M}_{n,p}(\mathbb{R}) \rightarrow \mathcal{S}_k^+(\mathbb{R}))$  est  $k$  fois différentiable, alors  $\ln |F|$  l’est et admet pour différentielle  $d \ln |F| = \text{Tr}(F^{-1})dF$  (il suffit de poser  $n = p = 1$  et  $F = S$ ). Elle peut même être redémontrée directement, en notant  $\lambda_i^A$  la  $i$ -ième valeur propre d’une matrice  $A$  (sans ordre de multiplicité) :

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

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

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

Il en ressort que

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

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  $\mathbb{E}_\theta((Y - Xb)(Y - Xb)') = S(\theta)$ ,  
 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

Calculer la matrice d'information du modèle en  $(b, \theta = 0)$ .  
 Le résultat s'exprime très simplement en fonction de  $X$  et  $A$ .

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  $(b, \theta)$  :

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

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  $(\theta, b)$  :

Loins de nous lancer dans la dérivation selon  $b$  de  $\frac{\partial \ln L(Y|X, \cdot, \cdot)}{\partial \theta}$ , constatons astucieusement que  $(b, \theta) \mapsto \mathcal{S}(b, \theta)$  est de classe  $C^2$  sur un voisinage de 0 : elle est en effet la composée d'une fonction de classe  $C^\infty$ , à savoir

$$\left( \begin{array}{ccc} \mathbb{R} \times \mathcal{M}_{N,N}(\mathbb{R}) & \rightarrow & \mathbb{R} \\ (x, A) & \mapsto & -\frac{N}{2} \ln 2 - \frac{1}{2}x - \frac{1}{2}(Y - Xb)' \cdot A \cdot (Y - Xb) \end{array} \right), \text{ et des deux fonctions}$$

$(M \mapsto \ln |M|)$  et  $(M \mapsto M^{-1})$  qui sont de classe  $C^\infty$  sur une partie dense d'un voisinage de 0. Cauchy nous assure alors de l'égalité des dérivées croisées pour  $\theta$  voisin de 0 :  $\forall(b, \theta), \frac{\partial^2 \ln L(Y|X, \cdot, \cdot)}{\partial b \partial \theta}(\theta, b) = \frac{\partial^2 \ln L(Y|X, \cdot, \cdot)}{\partial \theta \partial b}(\theta, b)$ , 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  $(\theta, \theta)$  :

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

Donc

$$\mathbb{E}_{b, \theta} \left( -\frac{\partial^2 \ln L(Y|X, \cdot, \cdot)}{\partial \theta^2}(\theta, b) \right) = \frac{1}{2} \text{Tr} \left( S(\theta)^{-1} \frac{\partial S}{\partial \theta}(\theta) S(\theta)^{-1} \frac{\partial S}{\partial \theta}(\theta) \right)$$

et par suite, sachant que  $S(0) = I_N$  et en notant  $A = \frac{\partial S}{\partial \theta}(0)$

$$\mathbb{E}_{b, \theta=0} \left( -\frac{\partial^2 \ln L(Y|X, \cdot, \cdot)}{\partial \theta^2}(\theta, b) \right) = \frac{1}{2} \text{Tr} (A^2)$$

On a donc finalement :

$$I(b, 0) = \left( \begin{array}{c|c} X'X & 0 \\ \hline & \dots \\ & 0 \\ \hline 0 \dots 0 & \frac{1}{2} \text{Tr} (A^2) \end{array} \right)$$

Dans cette question, on suppose que  $S(\theta)$  est la matrice de terme général :

$$S(\theta)_{i,j} = (\theta^{|i-j|})_{1 \leq i,j \leq N}$$

Q4

Calculer  $I_{b,\theta}(b, 0)$ .

On a  $S(\theta) = (\theta^{|i-j|})_{(i,j) \in \llbracket 1,N \rrbracket^2}$ .

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

En conséquence

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

et donc  $Tr(A^2) = 2(N - 1)$ .

Ainsi,

$$I_{b,\theta}(b, 0) = \begin{pmatrix} X'X & 0 \\ 0 & N - 1 \end{pmatrix}$$

### Exercice corrigé 2

Calculer l'estimateur du maximum de vraisemblance (e.m.v.)  $\hat{p}$  de  $p$  dans le modèle

Q1

$$X_i \underset{iid}{\sim} \mathcal{B}(1, p)$$

et calculer la loi limite de  $\sqrt{n}(\hat{p} - p)$ .

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  $\hat{p}$  de  $p$  vérifie

$$(1 - \hat{p}) \sum_{i=1}^n x_i = \hat{p}(n - \sum_{i=1}^n x_i)$$

soit  $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}$  (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

$$\sqrt{n}(\hat{p} - p) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, p(1-p))$$

Calculer l'e.m.v.  $(\hat{m}, \hat{\sigma}^2)$  de  $(m, \sigma^2)$  dans le modèle

$$X_i \underset{iid}{\sim} \mathcal{N}(m, \sigma^2)$$

→ Q2

et donner la loi limite du vecteur

$$\sqrt{n} \begin{pmatrix} \hat{m} - m \\ \hat{\sigma}^2 - \sigma^2 \end{pmatrix}$$

Le paramètre à estimer est ici  $(m, \sigma^2)$  (et pas  $(m, \sigma)$  : on s'abstiendra d'écrire  $\sigma^4$  mais plutôt  $(\sigma^2)^2$ ). 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,

$$\frac{\partial \ln f_X}{\partial (\sigma^2)}(x, m, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - m)^2$$

L'estimateur du maximum de vraisemblance  $\begin{pmatrix} \hat{m} \\ \hat{\sigma}^2 \end{pmatrix}$  de  $\begin{pmatrix} m \\ \sigma^2 \end{pmatrix}$  vérifie alors pour tout  $x$

$$\frac{\partial \ln f_X}{\partial \begin{pmatrix} m \\ \sigma^2 \end{pmatrix}} \left( \begin{pmatrix} \hat{m}(x) \\ \hat{\sigma}^2(x) \end{pmatrix} \right) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

ce dont on tire

$$\begin{cases} \hat{m}(x) = x \\ \frac{1}{\hat{\sigma}^2(x)} \sum_{i=1}^n (x_i - \hat{m}(x)) = 0 \end{cases}$$

soit encore

$$\begin{pmatrix} \hat{m} \\ \hat{\sigma}^2 \end{pmatrix} : x \mapsto \begin{pmatrix} \bar{x} \\ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{pmatrix}$$

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

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

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

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

Les variables aléatoires (asymptotiquement normales)  $\hat{m}$  et  $\hat{\sigma}^2$  sont de covariance asymptotique nulle, donc<sup>1</sup> sont asymptotiquement indépendantes.

Calculer l'e.m.v.  $(\hat{a}, \hat{b})$  de  $(a, b)$  dans le modèle

$$X_i \underset{iid}{\rightsquigarrow} \mathcal{U}([a, b]) \quad \text{loi uniforme sur } [a, b]$$

☞ Q3

Donner la loi limite du vecteur

$$n \begin{pmatrix} \hat{a} - a \\ b - \hat{b} \end{pmatrix}$$

La vraisemblance s'écrit :

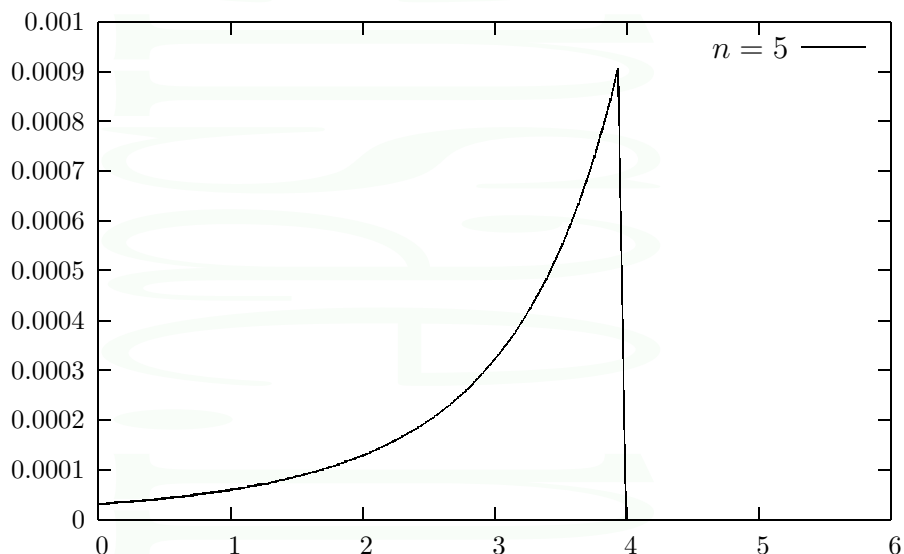
$$L_X(x; a, b) = \frac{1}{(b-a)^n} \mathbb{1}_{\min x_i \geq a} \mathbb{1}_{\max x_i \leq b}$$

et n'est pas dérivable en  $a$  et  $b$  (discontinuités en  $a = \min x_i$  et  $b = \max x_i$ ). Les conditions de régularité habituelles ne sont pas vérifiées, et le point  $(\hat{a}(x), \hat{b}(x))$  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,  $\begin{pmatrix} ] -\infty, b[ \rightarrow \mathbb{R}^+ \\ a \mapsto \frac{1}{(b-a)^n} \mathbb{1}_{\min x_i \geq a} \mathbb{1}_{\max x_i \leq b} \end{pmatrix}$  est identiquement nulle si  $\max_i x_i > b$ , et sinon admet pour graphe

<sup>1</sup>Attention, propriété propre à la loi normale

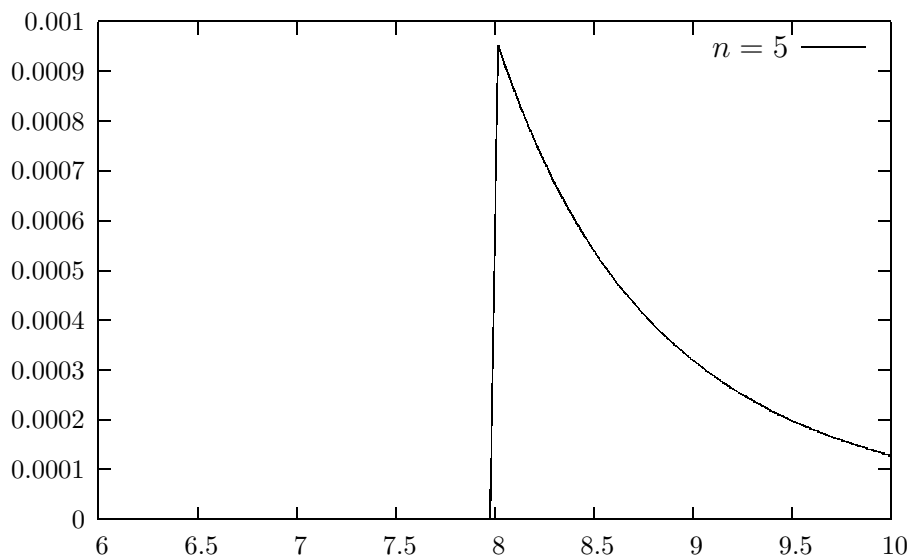




Graphes de  $a \mapsto L_X(x; a, b)$  lorsque  $b = 8 > \max_i x_i$  et  $\min_i x_i = 4$

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

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



Graphes 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  $\hat{b}(x) = \max_i x_i$  (qui ne dépend pas de  $a$ ).

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

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

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

En particulier,  $\mathbb{P}(n(\hat{a}(X) - a) > t) \xrightarrow{n \rightarrow +\infty} e^{-t/(b-a)}$  et donc la fonction de survie de  $n(\hat{a} - a) > t$  tend quand  $n \rightarrow +\infty$  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

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

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

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

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

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

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

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

En conséquence,

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

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

Autre méthode (beaucoup plus simple) :

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

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

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

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

## Exercice corrigé 3

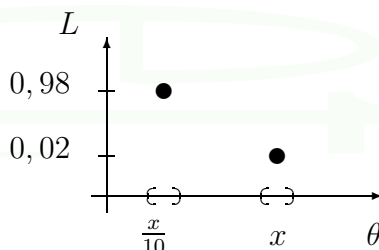
Exemple tiré de *Basu D. (1988) Stastical Information and Likelihood, Springer-Verlag, N. Y.*  
 Dans une urne contenant 1000 tickets, 20 sont marqués  $\theta$  et 980 sont marqués  $10\theta$ .

⇒ Q1 Donner l'estimateur du maximum de vraisemblance  $\hat{\theta}$  de  $\theta$  lorsque l'on tire un unique ticket de valeur  $X$ , et montrer que  $P(\hat{\theta} = \theta) = 0.98$ .

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

⇒ Q2 On renumérote les tickets marqués  $10\theta$  par  $a_i\theta$  ( $1 \leq i \leq 980$ ) où les  $a_i$  sont des réels connus, deux-à-deux distincts, et compris dans l'intervalle  $[10, 10.1]$ . Donner le nouvel estimateur du maximum de vraisemblance  $\tilde{\theta}$  et montrer que  $P(\tilde{\theta} = \theta) = 0.02$ . Ce résultat vous semble-t-il paradoxal ?

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,  $\mathbb{P}(\tilde{\theta} = \theta) = \mathbb{P}(X = \theta) = 0.02$  On a même  $\mathbb{P}(\tilde{\theta} \geq 10\theta) = \mathbb{P}(X \geq 10\theta) = 1 - \mathbb{P}(X = \theta) = 0.98$  : 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'occurrence, la maximisation de la vraisemblance) peut conduire, face à deux problèmes quasi-identiques, à une 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 ( $\hat{\theta} = \frac{x}{10.5}$ ) : 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  $n \rightarrow +\infty$ ) ; 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.