Cursus Intégré
2004-2005

Rappels de statistique mathématique
Corrigé des travaux dirigés n°4

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

Corrigé de l'exercice 1

☞ Q1 On obtient : $\widehat{\alpha} = \min x_i$, $\widehat{\theta} = \overline{x} - \min x_i$.

☞ Q2 On a :

$$\mathbb{P}(n(\widehat{\alpha} - \alpha) \geq t) = \mathbb{P}(\min x_i \geq \alpha + t/n) = \left(\exp^{-t/n\theta}\right)^n = \exp(-t/\theta)$$

et donc $n(\widehat{\alpha} - \alpha) \rightsquigarrow \mathcal{E}\left(\frac{1}{\theta}\right)$.

☞ Q3 On ne peut pas appliquer le théorème de la normalité asymptotique à $\widehat{\theta}$, car $\widehat{\theta}$ n'est que l'u des des deux composantes de l'estimateur du maximum de vraisemblance $(\widehat{\alpha}, \widehat{\theta})$ dont les co posantes ne sont pas indépendantes, et qui dans ce cas particulier ne vérifie pas les conditi de régularité habituelles (notamment dérivabilité de la log-vraisemblance).

Il faut retrouver la loi limite d'une autre façon. Ici :

$$\sqrt{n}(\widehat{\theta} - \theta) = \sqrt{n}\,(\overline{x} - (\theta + \alpha)) + \sqrt{n}(\alpha - \widehat{\alpha})$$

Comme $\mathbb{E}\,(X) = \theta + \alpha$, $\mathbb{V}\,(X) = \theta^2$, il vient d'après le théorème central limite :

$$\sqrt{n}\,(\overline{x} - (\theta + \alpha)) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \theta^2)$$

De plus,

$$\sqrt{n}(\alpha - \widehat{\alpha}) = \frac{-1}{\sqrt{n}} n(\widehat{\alpha} - \alpha) \xrightarrow{P} 0$$

puisque $n(\widehat{\alpha} - \alpha) \rightsquigarrow \mathcal{E}\left(\frac{1}{\theta}\right)$. Rappelons que :

$$X_n \xrightarrow[n\to+\infty]{\mathcal{L}} X, Y_n \xrightarrow{P} a : X_n + Y_n \xrightarrow[n\to+\infty]{\mathcal{L}} X + a$$

si a est une constante. Donc :

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0, \theta^2)$$

☞ Q4 La densité de l'échantillon ordonné $Y = (X_{(1)}, ..., X_{(n)})$ s'écrit

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

En d'autres termes,

$$\begin{aligned} f_{X_{(1)},\dots,X_{(n)}(y_1,\dots,y_n)} &= n! \mathbb{1}_{y_1<\dots<y_n} \frac{1}{\theta^n} \left(\Pi_{i=1}^n e^{-\frac{y_i - \alpha}{\theta}}\right) \mathbb{1}_{min_i y_i \geq \alpha} \\ &= \frac{n!}{\theta^n} \mathbb{1}_{\alpha \leq y_1 < \dots < y_n} e^{-\sum_{i=1}^n \frac{y_i - \alpha}{\theta}} \end{aligned}$$

Soit alors

$$\phi : \begin{pmatrix} \mathbb{R}^n & \rightarrow & \mathbb{R}^n \\ (y_1, \dots, y_n) & \mapsto & (ny_1, (n-1)(y_2 - y_1), \dots, (y_n - y_{n-1})) \end{pmatrix}$$

Alors ϕ est de clasee C^∞ et en outre

$$f_{\phi(Y)}(z_1,\ldots,z_n) = \left|Jac\phi^{-1}\right| f_Y(\phi^{-1}(z_1,\ldots,z_n))$$

(ceci se montre en effectuant le changement de variable $z = \phi(y)$ dans l'intégrale $\int_A f_Y(y)$, pour toute partie mesurable A de $\mathbb{R}^n$).
Or $\phi(Y)_1 = nY_1$, donc

$$\begin{aligned} \frac{\partial\phi(\cdot)_1}{\partial Y_1} &= n \\ \frac{\partial\phi(\cdot)_1}{\partial Y_i} &= 0, \quad \forall i \geq 2 \end{aligned}$$

puis $\phi(Y)_2 = (n-1)(Y_2 - Y_1)$, et donc

$$\begin{aligned} \frac{\partial\phi(\cdot)_2}{\partial Y_1} &= -(n-1) \\ \frac{\partial\phi(\cdot)_2}{\partial Y_2} &= (n-1), \quad \forall i \geq 2 \\ \frac{\partial\phi(\cdot)_2}{\partial Y_i} &= 0, \quad \forall i \geq 3 \end{aligned}$$

de sorte qu'en définitive

$$Jac(\phi) = \begin{pmatrix} n & 0 & 0 & \ldots & 0 & 0 \\ -(n-1) & (n-1) & 0 & \ldots & 0 & 0 \\ 0 & -(n-2) & (n-2) & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \ldots & \vdots & \vdots \\ 0 & 0 & \ldots & \ldots & -1 & 1 \end{pmatrix}$$

En conséquence, $|Jac(\phi^{-1})| = |Jac(\phi)^{-1}| = |Jac(\phi)|^{-1} = \frac{1}{n!}$
D'autre part, $\mathbb{1}_{y_1 \geq \alpha} = \mathbb{1}_{\phi(y)_1 \geq n\alpha}$, puis pour $i \geq 2$, $\mathbb{1}_{y_i < y_{i+1}} = \mathbb{1}_{\phi(y)_i > 0}$.
Enfin, une récurrence immédiate montre que $\forall i \in [\![1,n]\!]$, $\phi^{-1}(z)_i = \sum_{j=1}^{i} \frac{z_j}{n-j+1}$, de sorte qu'en définitive

$$\boxed{f_Z(z_1,...,z_n) = \theta^{-n} \exp^{-(\sum z_i - n\alpha)/\theta} \mathbb{1}_{z_1 \geq n\alpha} \mathbb{1}_{z_2 > 0} \ldots \mathbb{1}_{z_n > 0}}$$

Autrement dit, les Z_i sont indépendants, avec $Z_1 \leadsto n\alpha + \mathcal{E}\left(\frac{1}{\theta}\right)$, et $Z_i \leadsto \mathcal{E}\left(\frac{1}{\theta}\right)$ pour $i \geq 2$. Finalement, constatant que $\widehat{\alpha} = \frac{1}{n} Z_1$ et $\widehat{\theta} = \frac{1}{n}\sum_{i=2}^n Z_i$, on conclut que $\widehat{\alpha}$ et $\widehat{\theta}$ sont indépendants.

⋆
⋆ ⋆

Corrigé de l'exercice 2

☞ Q1 On a par définition de la probabilité conditionnelle

$$\forall i \in [\![1,n]\!],\ \mathbb{P}(A|H_i) = \frac{\mathbb{P}(A \cap H_i)}{\mathbb{P}(H_i)}$$

et donc

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) \\ &= \mathbb{P}(A \cap (\cup_{i=1}^n H_i)) \quad \text{car } (H_1,\ldots,H_n) \text{ est complet} \\ &= \sum_{i=1}^n \mathbb{P}(A \cap H_i) \quad \text{car } (H_1,\ldots,H_n) \text{ est disjoint} \\ &= \sum_{i=1}^n \mathbb{P}(A|H_i)\,\mathbb{P}(H_i) \quad \text{car chaque } H_i \text{ est non-vide} \end{aligned}$$

(cette formule est dite *des probabilités totales*).
Par ailleurs on a

$$\mathbb{P}(A|H_i)\,\mathbb{P}(H_i) = \mathbb{P}(A \cap H_i) = \mathbb{P}(H_i|A)\,\mathbb{P}(A)$$

et donc

$$\begin{aligned} \mathbb{P}(H_i|A) &= \frac{\mathbb{P}(H_i)\,\mathbb{P}(A|H_i)}{\mathbb{P}(A)} \\ &= \boxed{\frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\sum_{i=1}^n \mathbb{P}(H_i)\mathbb{P}(A|H_i)}} \end{aligned}$$

Cette relation exprime la probabilité d'un événement H_i, une fois connue la réalisation de
en fonction de sa probabilité a priori.

☞ Q2 Dans le cas où $\Theta = \{\theta_i \,/\, i \in \mathbb{N}\}$ est dénombrable il suffit d'appliquer le résultat précéden
$A = \{(x_1,\ldots,x_n)\}$ et $\forall i \in \mathbb{N}$, $H_i = \{\theta_i\}$.
Dans le cas général le calcul est similaire : la formule des probabilités totales s'écrit pour to
$\mathcal{A} \subset \mathbb{R}^n$ mesurable

$$\mathbb{P}(X \in \mathcal{A}) = \int_\Theta \mathbb{P}(X \in \mathcal{A}|\theta = \theta_0)\,\mathbb{P}(\theta = \theta_0)\,d\theta_0$$

et par suite pour toute $\mathcal{B} \subset \Theta$ mesurable, à supposer que $\mathbb{P}(X \in \mathcal{A}) \neq 0$

$$\mathbb{P}(\theta \in \mathcal{B}|X \in \mathcal{A}) = \frac{\mathbb{P}(\theta \in \mathcal{B})\,\mathbb{P}(X \in \mathcal{A}|\theta \in \mathcal{B})}{\mathbb{P}(X \in \mathcal{A})}$$

et donc

$$\boxed{\pi_{\cdot|x}(\theta|X = x) = \frac{f_{\mathcal{L}_\theta^{\otimes n}}(x_1,\ldots,x_n)\,\pi_0(\theta)}{\int_\Theta f_{\mathcal{L}_\theta^{\otimes n}}(x_1,\ldots,x_n)\,\pi_0(\theta)d\theta}}$$

☞ Q3 Attention aux notations : $\mathbb{E}_\theta\left(\dots\widehat{\theta}(X)\dots\theta\dots\right)$ désigne l'espérance sur la variable X, espérance qui dépend de θ, et non pas une espérance sur la variable θ. En outre l'énoncé suppose ici que Θ est de dimension 1 (sans quoi il faudrait noter $\mathbb{E}_\theta\left\|\widehat{\theta}(x)-\theta\right\|^2$).

Soit $\mathcal{X}=\mathbb{R}^n$.

Première méthode :

On cherche l'élément $T:\mathcal{X}\to\Theta$ (qui est une *fonction*) qui minimise

$$\begin{aligned} R_{\pi_0}(T) &= \int_\Theta\int_{\mathcal{X}}(T(x)-\theta)^2 f_{\mathcal{L}_\theta}(x)\pi_0(\theta)dxd\theta \\ &= \int_{\mathcal{X}}\left(\int_\Theta (T(x)-\theta)^2 f_{\mathcal{L}_\theta}(x)\pi_0(\theta)d\theta\right)dx \quad \text{d'après Fubini} \end{aligned}$$

Or la fonction

$$\phi:\left(\begin{array}{ccc} \mathcal{X}^2 & \to & \Theta \\ (x,t) & \mapsto & \int_\Theta (t-\theta)^2 f_{\mathcal{L}_\theta}(x)\pi_0(\theta)d\theta \end{array}\right)$$

est positive, donc il suffit de minimiser point-par-point :

$$\widehat{\theta}=argmin_{T:\mathcal{X}\to\Theta}\int_{\mathcal{X}}\phi\left(T(x)\right)dx=\left(\begin{array}{ccc} \mathcal{X} & \to & \Theta \\ x & \mapsto & argmin_{t\in\Theta}\,\phi(x,t) \end{array}\right)$$

Soit donc $x\in\mathcal{X}$ **fixé**, et calculons le nombre $\widehat{\theta}(x)=argmin_{t\in\Theta}\,\phi(x,t)$. On a

$$\begin{aligned} \phi(x,t) &= \int_\Theta (t-\theta)^2 f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta \\ &= \underbrace{\left(\int_\Theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta\right)}_{a>0}t^2+\underbrace{\left(-2\int_\Theta \theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta\right)}_{b}t+\underbrace{\left(\int_\Theta \theta^2 f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta\right)}_{c} \end{aligned}$$

Donc $\phi(x,t)=at^2+bt+c$ est une parabole en t (à x fixé), et est donc extrêmale en $-\frac{b}{2a}$; comme $a>0$ elle est *minimale* en $-\frac{b}{2a}$ et donc

$$\begin{aligned} argmin_{t\in\Theta}\,\phi(x,t) &= -\frac{b}{2a} \\ &= \frac{\int_\Theta \theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta}{\int_\Theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta} \\ &= \int_\Theta \theta\frac{f_{\mathcal{L}_\theta}(x)\phi_0(\theta)}{\int_\Theta f_{\mathcal{L}_\theta}(x)\phi_0(\theta)d\theta}d\theta \end{aligned}$$

Par conséquent,

$$\boxed{\forall x\in\mathcal{X},\ \widehat{\theta}(x)=\int_\Theta \theta\pi_{\cdot|x}(\theta)d\theta}$$

☞ Q4 Dans le cas d'une loi a priori uniforme on a simplement

$$\begin{aligned} \widehat{\theta}(x) &= \int_{[a,b]}\theta\pi_{\cdot|x}(\theta)d\theta \\ &= \int_{[a,b]}\theta\frac{f_{\mathcal{L}_\theta}(x_1,\dots,x_n)\frac{1}{b-a}}{\int_\Theta f_{\mathcal{L}_\theta}(x_1,\dots,x_n)\frac{1}{b-a}d\theta}d\theta \\ &= \frac{\int_{[a,b]}\theta f_{\mathcal{L}_\theta}(x)d\theta}{\int_{[a,b]} f_{\mathcal{L}_\theta}(x)d\theta} \end{aligned}$$

Dans le cas où $\mathcal{L}_\theta=\mathcal{E}(\theta)$ la densité de l'échantillon indépendant $(X_1,\dots,X_n)$ est

$$\begin{aligned} f_{\mathcal{E}(\theta)^{\otimes n}}\left(X_1=x_1,\dots,X_n=x_n\right) &= \Pi_{i=1}^n\left(\theta e^{-\theta x_i}\mathbb{1}_{x_i\geq 0}\right) \\ &= \theta^n e^{-\theta(x_1+\cdots+x_n)}\mathbb{1}_{\min_i x_i\geq 0} \end{aligned}$$

et donc pour $(x_1,\dots,x_n)\in(\mathbb{R}^+)^n$ il vient

$$\begin{aligned} \widehat{\theta}(x) &= \frac{\int_{[a,b]}\theta^{n+1}e^{-\theta(x_1+\cdots+x_n)}d\theta}{\int_{[a,b]}\theta^n e^{-\theta(x_1+\cdots+x_n)}d\theta} \\ &= \frac{\left[\theta^{n+1}\frac{e^{-\theta\sum_{i=1}^n x_i}}{-\sum_{i=1}^n x_i}\right]_a^b-\int_{[a,b]}(n+1)\theta^n\frac{e^{-\theta\sum_{i=1}^n x_i}}{-\sum_{i=1}^n x_i}d\theta}{\int_{[a,b]}\theta^n e^{-\theta(x_1+\cdots+x_n)}d\theta} \quad \text{par parties} \\ &= \frac{1}{\sum_{i=1}^n x_i}\left(n+1-\frac{b^{n+1}e^{-b\sum_{i=1}^n x_i}-a^{n+1}e^{-a\sum_{i=1}^n x_i}}{\int_{[a,b]}\theta^n e^{-\theta(x_1+\cdots+x_n)}d\theta}\right) \end{aligned}$$

Ainsi $\boxed{\widehat{\theta}(x)=\frac{1}{\sum_{i=1}^n x_i}\left(n+1-\frac{\overline{\theta}^{n+1}e^{-\overline{\theta}\sum_{i=1}^n x_i}}{\int_{[0,\overline{\theta}]}\theta^n e^{-\theta(x_1+\cdots+x_n)}d\theta}\right)}$ lorsque $\Theta=[0,\overline{\theta}]$ par exemple.

L'estimateur du maximum de vraisemblance de θ, par comparaison, est $\widehat{\theta}^{emv}(x)=\frac{n}{\sum_{i=1}^n x_i}$.

⋆
⋆ ⋆

Corrigé de l'exercice 3

☞ Q1 Laplace considère en fait de ne pas détenir d'information sur p autre que celle apportée par observations : en d'autres termes, il n'a pas d' "a priori" sur p, ce qu'il modélise par une lo priori uniforme. Cela revient intuitivement à donner une "probabilité égale" à toutes les vale possibles de p. Mais cette intuition est en fait trompeuse : si on change la paramétrisat du modèle (en remplaçant p par $q=p^2$ par exemple), on se rend compte alors que la lo priori sur q n'est plus uniforme ! Cette reparamétrisation ne nous a pourtant apporté aucune information supplémentaire. Ce paradoxe illustre la difficulté de bien choisir une loi a priori, surtout dans un cadre non-informatif (c'est à dire lorsqu'aucune information n'est disponible a priori). Cependant, nous verrons en 4) qu'il est assez simple de se restreindre au moins à un choix limité de lois a priori raisonnables.

☞ Q2 La densité de la loi a posteriori $\pi(\theta|x_1,...,x_n)$ du paramètre θ s'écrit sous la forme (voir exercice 2) :

$$\pi(\theta|x_1,...,x_n) = \frac{\pi(\theta)\mathbb{P}(x_1,...,x_n|\theta)}{\int_\Theta \pi(\theta)\mathbb{P}(x_1,...,x_n|\theta)d\theta}$$

Il vient

$$\begin{aligned}\pi(p|N_g = n_g) &= \frac{\pi(p)\mathbb{P}(N_g = n_g|p)}{\int_0^1 \pi(p)\mathbb{P}(N_g = n_g|p)dp} \\ &= \frac{p^{n_g}(1-p)^{n-n_g}}{\int_0^1 p^{n_g}(1-p)^{n-n_g}dp}\end{aligned}$$

Par conséquent

$$\begin{aligned}\mathbb{P}(p > \frac{1}{2}|N_g = n_g) &= \int_{[0,1]} \frac{\pi(p)\mathbb{P}(N_g = n_g|p)}{\int_0^1 \pi(p)\mathbb{P}(N_g = n_g|p)dp} \mathbb{1}_{p>\frac{1}{2}} dp \\ &= \frac{\int_{\frac{1}{2}}^1 p^{n_g}(1-p)^{n-n_g}dp}{\int_0^1 p^{n_g}(1-p)^{n-n_g}dp}\end{aligned}$$

☞ Q3 Notons π_p la densité de la loi a posteriori : π_p est la densité de la loi $\mathcal{B}(n_g+1, n-n_g+1)$. Son espérance et sa variance sont donc :

$$\begin{aligned}\mathbb{E}_{\pi_p}[p] &= \frac{n_g+1}{n+2} \\ \mathbb{V}_{\pi_p}[p] &= \frac{(n_g+1)(n-n_g+1)}{(n+2)^2(n+3)}\end{aligned}$$

Or $n_g = \sum_{i=1}^n g_i$ où $g_i = \mathbb{1}_{\text{l'enfant } i \text{ est un garçon}} \underset{iid}{\rightsquigarrow} \mathcal{B}(1, p_0)$.

Donc en vertu de la loi forte des grands nombres $\frac{n_g}{n} = \frac{1}{n}\sum_{i=1}^n g_i \xrightarrow[n\to+\infty]{p.s.} p_0$ et donc

$$\begin{aligned}\mathbb{E}_{\pi_p}[p] &\xrightarrow[n\to+\infty]{p.s.} p_0 \\ \mathbb{V}_{\pi_p}[p] &\underset{n\to+\infty}{\overset{p.s.}{\rightsquigarrow}} \frac{p_0(1-p_0)}{n}\end{aligned}$$

En d'autres termes, lorsque $n \to +\infty$ la loi a posteriori se "rétrécie" (sa variance tend vers 0) et se concentre autour de la vraie valeur p_0. Ce résultat est intuitif : plus on dispose d'observations, plus faible est l'incertidude sur la valeur du paramètre p.

☞ Q4 Si on change la loi a priori uniforme (qui correspond d'ailleurs à une loi $\mathcal{B}(1,1)$) pour une lo priori plus générale $\mathcal{B}(\alpha,\beta)$, on obtient une loi a posteriori :

$$\begin{aligned}\pi(p|N_g &= n_g) \propto p^{\alpha-1}(1-p)^{\beta-1}p^{n_g}(1-p)^{n-n_g}1_{[0,1]}(p) \\ &\propto p^{\alpha+n_g-1}(1-p)^{\beta+n-n_g-1}1_{[0,1]}(p)\end{aligned}$$

(où $\propto$ signifie "proportionnel à")

La loi a posteriori est donc une loi $\mathcal{B}(n_g+\alpha, n-n_g+\beta)$. On vérifie alors facilement que $\mathbb{E}_{\pi_p}$ et $\mathbb{V}_{\pi_p}[p]$ conservent le même comportement asymptotique que dans le cas précédent. En ef plus grand est le nombre d'observations, plus le poids de l'a priori est faible, et ce poids devi nul asymptotiquement.

Cependant, la Statistique Bayésienne ne repose pas sur des justifications asymptotiques. distance finie (c'est à dire pour un nombre fini d'observations), une loi a priori mal cho peut sérieusement déformer la loi a posteriori, et mener à une inférence complètement érron Ainsi, au vu du probème posé (situer p dans l'un des deux intervalles $[0,\frac{1}{2}]$ ou $[\frac{1}{2},1]$), on aucune raison de favoriser *a priori* l'une de ces deux régions, seules les observations pouv nous permettre de les départager. Il est donc naturel de se restreindre au moins aux lois a pri symétriques ($\alpha = \beta$). Parmi celles-ci, on peut aussi écarter les lois de très faible variance telles que α prend de fortes valeurs, ex : $\mathcal{B}(10,10)$), qui privilégient inutilement un voisin trop restreint du point $\frac{1}{2}$.

Finalement, la loi a priori uniforme proposée par Laplace semble donc un choix raisonna parmi d'autres, au moins pour le problème posé. Il est en fait possible de proposer une plus satisfaisante (dite loi a priori de Jeffreys, ici $\mathcal{B}(1/2,1/2)$) qui permette de s'affranchir paradoxe de la reparamétrisation présenté en 1).

⋆
⋆ ⋆