Probabilités et statistiques

Corrigé de l'examen du 10 mai 2012

Problème 1

1. La première marginale est donnée par
$$f_1(x) = \int_0^1 f(x,y)\,\mathrm{d}y = C\int_0^x \frac{1}{\sqrt{xy}}\,\mathrm{d}y 1_{]0,1[}(x) = C\frac{2\sqrt{y}}{\sqrt{x}}\bigg|_{y=0}^{y=x} 1_{]0,1[}(x) = 2C1_{]0,1[}(x)\ .$$
C'est la densité de la loi uniforme sur $]0,1[$, à condition que $C = 1/2$. La seconde marginale est donnée par
$$f_2(y) = \int_0^1 f(x,y)\,\mathrm{d}x = \left(\frac{1}{\sqrt{y}} - 1\right)1_{]0,1[}(y)\ .$$

2. On trouve $\mathbb{E}(X) = 1/2$, $\mathbb{E}(Y) = 1/6$ et
$$\mathbb{E}(XY) = \int_{\mathbb{R}^2} xyf(x,y)\,\mathrm{d}x\,\mathrm{d}y = \frac{1}{2}\int_0^1\int_0^x \sqrt{xy}\,\mathrm{d}y\,\mathrm{d}x = \frac{1}{2}\int_0^1 \frac{2}{3}x^2\,\mathrm{d}x = \frac{1}{9}$$
d'où $\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 1/36$.

Problème 2

1. Y admet la densité $f_Y(y) = \lambda\,\mathrm{e}^{-\lambda y}\,1_{\{y>0\}}$, donc
$$L_Y(u) = \int_0^\infty \mathrm{e}^{-uy}\,\lambda\,\mathrm{e}^{-\lambda y}\,\mathrm{d}y = \frac{\lambda}{\lambda+u}\ .$$

2. (a) $f(x,y) = f_X(x)\lambda\,\mathrm{e}^{-\lambda y}\,1_{\{y>0\}}$.
   (b) On a
$$\mathbb{P}\left\{Y > X\right\} = \int_0^\infty\int_x^\infty f(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_0^\infty f_X(x)\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \mathbb{E}\left(\mathrm{e}^{-\lambda X}\right)\ .$$

3. Le résultat précédent montre que
$$\mathbb{P}\left\{Y > X_1 + \dots + X_n\right\} = \mathbb{E}\left(\mathrm{e}^{-\lambda(X_1+\dots+X_n)}\right)\ ,$$
et l'indépendance implique
$$\mathbb{E}\left(\mathrm{e}^{-\lambda(X_1+\dots+X_n)}\right) = \mathbb{E}\left(\mathrm{e}^{-\lambda X_1}\right)\dots\mathbb{E}\left(\mathrm{e}^{-\lambda X_n}\right)\ .$$
Or le résultat précédent montre aussi que $\mathbb{E}(\mathrm{e}^{-\lambda X_k}) = \mathbb{P}\{Y > X_k\}$ pour chaque k.

4. (a) Les événements $\{M = X_1\},\dots,\{M = X_n\}$ ont tous la même probabilité, et comme les X_i ont une densité, leurs intersections ont une probabilité nulle. De plus comme $M \in \{X_1,\dots,X_n\}$, leur réunion est Ω. Par la loi des probabilités totales,
$$\begin{aligned}\mathbb{P}\left\{2M > X_1 + \dots + X_n\right\} &= \sum_{k=1}^n \mathbb{P}\left\{M = X_k, 2M > X_1 + \dots + X_n\right\}\\ &= n\mathbb{P}\left\{M = X_1, X_1 > X_2 + \dots + X_n\right\}\\ &= n\mathbb{P}\left\{X_1 > X_2 + \dots + X_n\right\}\ .\end{aligned}$$

(b) i. Par les points 1. et 2.(b), $\mathbb{P}\{X_1 > X_2\} = L_X(\lambda) = \lambda/(2\lambda) = 1/2$.

ii. Par les points 4.(a) et 3.,

$$\mathbb{P}\big\{2M > X_1 + \dots + X_n\big\} = n\mathbb{P}\big\{X_1 > X_2\big\}^{n-1} = \frac{n}{2^{n-1}} .$$

Problème 3

1. X_1 admet la densité $f_X(x) = \lambda\,\mathrm{e}^{-\lambda x}\,1_{\{x>0\}}$. Comme $X_1 = g(Y)$ où $g(y) = \log y$, le théorème de changement de variable montre que Y admet la densité

$$f_Y(y) = f_X(\log y)|g'(y)| = \frac{\lambda}{y^{1+\lambda}} 1_{\{y>1\}}$$

(loi dite de Pareto).

2. S_n suit une loi Gamma de paramètres n, λ.

3. On a

$$\mathbb{E}\left(\mathrm{e}^{uX_1}\right) = \int_0^\infty \mathrm{e}^{ux}\,\lambda\,\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{\lambda}{\lambda - u}$$

pour tout $u < \lambda$. Par conséquent,

$$f_n(u) = \mathbb{E}\left(\mathrm{e}^{uS_n}\right) = \mathbb{E}\left(\mathrm{e}^{uX_1}\right)\dots\mathbb{E}\left(\mathrm{e}^{uX_n}\right) = \left(\frac{\lambda}{\lambda - u}\right)^n .$$

4. On trouve

$$f_n'(u) = \frac{n\lambda^n}{(\lambda - u)^{n+1}} \qquad \text{puis} \qquad f_n^{(k)}(u) = \frac{n(n+1)\dots(n+k-1)\lambda^n}{(\lambda-u)^{n+k}}$$

pour tout $k \geqslant 1$.

5. Comme $\sum_{k\geqslant 0} \frac{1}{k!} u^k \mathbb{E}\left(S_n^k\right) = \mathbb{E}\left(\mathrm{e}^{uS_n}\right) = \sum_{k\geqslant 0} \frac{1}{k!} f_n^{(k)}(0) u^k$ pour tout $u < \lambda$, on a

$$\mathbb{E}\left(S_n^k\right) = f_n^{(k)}(0) = \frac{n(n+1)\dots(n+k-1)}{\lambda^k} = \frac{1}{\lambda^k}\frac{\Gamma(n+k)}{\Gamma(n)} .$$

Problème 4

1. Comme montré au cours, l'estimateur du maximum de vraisemblance est égal à la moyenne empirique, $\hat\theta = \frac{1}{n}\sum_{i=1}^n X_i$. Il est donc non biaisé et consistant.

2. L'intervalle de confiance est donné par

$$I = \left[\hat\theta - \sqrt{\frac{\hat\theta(1-\hat\theta)}{n}} c_\beta, \hat\theta + \sqrt{\frac{\hat\theta(1-\hat\theta)}{n}} c_\beta\right]$$

avec $c_\beta = \Phi^{-1}(\frac{1+\beta}{2}) \simeq 1.96$ pour $\beta = 0.95$.

3. (a) On a $\hat\theta = 0.48$ et $n = 1000$, ce qui donne $I = [0.449, 0.511]$.

(b) Pour que I ait une largeur de 0.02, il faut que $n = \hat\theta(1-\hat\theta)c_\beta^2(0.01)^{-2} \simeq 9589$.

(c) On prend comme région de rejet $R = \{\hat\theta < \theta_0\}$ avec θ_0 tel que $\sup_{\theta>1/2}\mathbb{P}_\theta(R) = \mathbb{P}_{1/2}(R) = 0.05$.
Le TCL fournit $\theta_0 \simeq \frac{1}{2} - 1.96 \cdot \sqrt{0.25/1000} \simeq 0.469$.
On rejette H_0 s'il y a moins de 469 opinions A.