Probabilités et statistiques

Corrigé de l'examen du 4 mai 2011

Problème 1

1. La fonction f est manifestement non-négative. De plus on a
$$\int_{\mathbb{R}^2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^\infty \int_0^y \mathrm{d}x\,\mathrm{e}^{-y}\,\mathrm{d}y = \int_0^\infty y\,\mathrm{e}^{-y}\,\mathrm{d}y = -y\,\mathrm{e}^{-y}\Big|_0^\infty + \int_0^\infty \mathrm{e}^{-y}\,\mathrm{d}y = 1 ,$$
ce qui montre que f est bien une densité.

2. Les densités marginales valent respectivement
$$\begin{aligned} f_1(x) &= \int_{\mathbb{R}} f(x,y)\,\mathrm{d}y = \int_x^\infty \mathrm{e}^{-y}\,\mathrm{d}y\,1_{\{x>0\}} = \mathrm{e}^{-x}\,1_{\{x>0\}} , \\ f_2(y) &= \int_{\mathbb{R}} f(x,y)\,\mathrm{d}x = \int_0^y \mathrm{e}^{-y}\,\mathrm{d}x\,1_{\{y>0\}} = y\,\mathrm{e}^{-y}\,1_{\{y>0\}} . \end{aligned}$$
X suit donc une loi exponentielle de paramètre 1, alors que Y suit une loi Gamma de paramètres $(1,1)$.

3. Comme $f_1(x)f_2(y) = y\,\mathrm{e}^{-(x+y)}\,1_{\{x>0,y>0\}} \neq f(x,y)$, les variables X et Y ne sont pas indépendantes.

4. Pour $0 \leqslant z \leqslant 1$, et $Y > 0$, la condition $X/Y \leqslant z$ équivaut à $X \leqslant zY$. Il suit
$$\mathbb{P}\left\{\frac{X}{Y} \leqslant z, Y \leqslant t\right\} = \int_0^t \int_0^{zy} \mathrm{e}^{-y}\,\mathrm{d}x\,\mathrm{d}y = \int_0^t zy\,\mathrm{e}^{-y}\,\mathrm{d}y = z\big[1 + (t-1)\,\mathrm{e}^{-t}\big] .$$
On en déduit
$$\mathbb{P}\left\{\frac{X}{Y} \leqslant z\right\} = \lim_{t\to\infty} \mathbb{P}\left\{\frac{X}{Y} \leqslant z, Y \leqslant t\right\} = z$$
pour $z \in [0,1]$, donc que X/Y suit la loi uniforme sur $[0,1]$.

5. Les variables X/Y et Y sont indépendantes, puisque nous venons d'établir que leur fonction de répartition conjointe se factorise en produit d'une fonction de z et d'une fonction de t.

Problème 2

1. La densité du couple (X,Y) est $f_{X,Y}(x,y) = \lambda^2\,\mathrm{e}^{-\lambda(x+y)}\,1_{\{x>0,y>0\}}$.

2. L'application $\Phi : (x,y) \mapsto (u,z) = (x, y/x)$ est un difféomorphisme de $\mathbb{R}_+^* \times \mathbb{R}_+^*$ vers $\mathbb{R}_+^* \times \mathbb{R}_+^*$, admettant la réciproque $\Phi^{-1} : (u,z) \mapsto (x,y) = (u, zu)$. Le jacobien de Φ est donné par
$$\frac{\partial(u,z)}{\partial(x,y)} = \begin{vmatrix} 1 & 0 \\ -\frac{y}{x^2} & \frac{1}{x} \end{vmatrix} = \frac{1}{x} = \frac{1}{u} .$$
La formule de changement de variable donne alors la densité du couple (U,Z)
$$f_{U,Z}(u,z) = \lambda^2 u\,\mathrm{e}^{-\lambda(1+z)u}\,1_{\{u>0,z>0\}} .$$

3. Comme $(x,y) = (u, zu)$, la densité (marginale) de Z est donnée par
$$f_Z(z) = \int_0^\infty f_{U,Z}(u,z)\,\mathrm{d}u = \frac{\lambda}{1+z}\int_0^\infty \mathrm{e}^{-\lambda(1+z)u}\,\mathrm{d}u\,1_{\{z>0\}} = \frac{1}{(1+z)^2}1_{\{z>0\}} \;.$$
4. Comme $Z > 0$, on a
$$\mathbb{E}\left(|(1+Z)^\alpha|\right) = \mathbb{E}\left((1+Z)^\alpha\right) = \int_{\mathbb{R}} (1+Z)^\alpha f_Z(z)\,\mathrm{d}z = \int_0^\infty \frac{1}{(1+z)^{2-\alpha}}\,\mathrm{d}z \;.$$
Le critère de Riemann montre que cette intégrale existe pour $\alpha < 1$ et vaut dans ce cas
$$\int_0^\infty \frac{1}{(1+z)^{2-\alpha}}\,\mathrm{d}z = \frac{1}{1-\alpha} \;.$$
5. On sait que $\mathbb{E}(X) = 1/\lambda$, et on calcule, en intégrant d'abord sur x puis sur z,
$$\mathbb{E}\left(X(1+Z)^\alpha\right) = \int_0^\infty\int_0^\infty \lambda^2 x^2 (1+z)^\alpha\,\mathrm{e}^{-\lambda(1+z)x}\,\mathrm{d}x\,\mathrm{d}z = \frac{2}{1-\alpha}$$
pour autant que $\alpha < 1$. Il suit que
$$\operatorname{cov}(X, (1+Z)^\alpha) = \frac{2-1/\lambda}{1-\alpha} \;.$$

Problème 3

1. L'indépendance des X_i implique que
$$\mathbb{E}\left(z^{S_n}\right) = \mathbb{E}\left(z^{\sum_{i=1}^n X_i}\right) = \mathbb{E}\left(\prod_{i=1}^n z^{X_i}\right) = \prod_{i=1}^n \mathbb{E}\left(z^{X_i}\right) = G_X(z)^n \;.$$
2. La loi des probabilités totales permet d'écrire
$$\mathbb{P}\{S_N = k\} = \sum_{n\geqslant 0} \mathbb{P}\{S_N = k | N = n\}\mathbb{P}\{N = n\} = \sum_{n\geqslant 0} \mathbb{P}\{S_n = k\}\mathbb{P}\{N = n\} \;.$$
Il suit que
$$\begin{aligned} G_S(z) &= \mathbb{E}\left(z^{S_N}\right) \\ &= \sum_{k\geqslant 0} z^k \mathbb{P}\{S_N = k\} \\ &= \sum_{k\geqslant 0} z^k \sum_{n\geqslant 0} \mathbb{P}\{S_n = k\}\mathbb{P}\{N = n\} \\ &= \sum_{n\geqslant 0} \mathbb{P}\{N = n\} \underbrace{\sum_{k\geqslant 0} z^k \mathbb{P}\{S_n = k\}}_{=\mathbb{E}\left(z^{S_n}\right) = G_X(z)^n} \\ &= \sum_{n\geqslant 0} G_X(z)^n \mathbb{P}\{N = n\} = G_N(G_X(z)) \;. \end{aligned}$$
3. Les fonctions génératrices de N et des X_i sont données respectivement par
$$G_N(z) = \mathrm{e}^{\lambda(z-1)} \;, \qquad G_X(z) = 1 - p + pz \;.$$
Le résultat précédent montre que
$$G_S(z) = G_N(1-p+pz) = \mathrm{e}^{\lambda(-p+pz)} = \mathrm{e}^{p\lambda(z-1)} \;.$$
Par conséquent, S_N suit une loi de Poisson de paramètre $p\lambda$.

Problème 4

1. La fonction $f_{\mu,\sigma}$ est non-négative, et le changement de variable $y = \log x$ montre que son intégrale sur $\mathbb{R}$ vaut

$$\int_0^\infty \frac{1}{\sqrt{2\pi}\,\sigma x}\,\mathrm{e}^{-(\log x-\mu)^2/(2\sigma^2)}\,\mathrm{d}x = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\,\sigma}\,\mathrm{e}^{-(y-\mu)^2/(2\sigma^2)}\,\mathrm{d}y = 1\;,$$

puisque l'intégrant dans la dernière expression est la densité d'une loi gaussienne.

2. En utilisant à nouveau le changement de variable $y = \log x$, on obtient

$$\begin{aligned}\mathbb{E}\left(\log X\right) &= \int_{-\infty}^\infty y\frac{1}{\sqrt{2\pi}\,\sigma}\,\mathrm{e}^{-(y-\mu)^2/(2\sigma^2)}\,\mathrm{d}y = \mu\;,\\ \mathbb{E}\left((\log X)^2\right) &= \int_{-\infty}^\infty y^2\frac{1}{\sqrt{2\pi}\,\sigma}\,\mathrm{e}^{-(y-\mu)^2/(2\sigma^2)}\,\mathrm{d}y = \sigma^2+\mu^2\;.\end{aligned}$$

En effet, on reconnaît les deux premiers moments d'une loi gaussienne d'espérance μ et de variance σ^2. (On peut également se ramener au cas standard par le changement de variable $z = (y-\mu)/\sigma$).

3. La log-vraisemblance est donnée par

$$\log L_x = -n\log(\sqrt{2\pi}\sigma) - \sum_{i=1}^n x_i - \sum_{i=1}^n \frac{(\log x_i-\mu)^2}{2\sigma^2}\;.$$

Les dérivées partielles par rapport aux paramètres sont données par

$$\begin{aligned}\frac{\partial}{\partial\mu}\log L_x &= \frac{1}{\sigma^2}\sum_{i=1}^n(\log x_i-\mu)\;,\\ \frac{\partial}{\partial\sigma}\log L_x &= -\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^n(\log x_i-\mu)^2\;.\end{aligned}$$

Elles s'annulent les deux à la fois en

$$\hat\mu = \frac{1}{n}\sum_{i=1}^n\log x_i\;, \qquad \hat\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^n(\log x_i-\hat\mu)^2}\;.$$

On vérifie que c'est bien un maximum, donc $(\hat\mu,\hat\sigma)$ est l'estimateur du maximum de vraisemblance de (μ,σ).

4. Le résultat de la question 2. montre que $\mathbb{E}\left(\hat\mu\right) = \mu$, donc que l'estimateur $\hat\mu$ est sans biais.

5. On a

$$\begin{aligned}\mathbb{E}\left(\hat\mu^2\right) &= \frac{1}{n^2}\sum_{i=1}^n\mathbb{E}\left((\log X_i)^2\right) + \frac{1}{n^2}\sum_{i\neq j,i,j=1}^n\mathbb{E}\left(\log X_i\right)\mathbb{E}\left(\log X_j\right)\\ &= \frac{n(\sigma^2+\mu^2)}{n^2} + \frac{n(n-1)\mu^2}{n^2} = \frac{\sigma^2}{n}+\mu^2\;.\end{aligned}$$

Le risque quadratique moyen vaut donc

$$\mathbb{E}\left((\hat\mu-\mu)^2\right) = \mathbb{E}\left(\hat\mu^2\right) - \mu^2 = \frac{\sigma^2}{n}\;.$$

L'estimateur est donc consistant.