Probabilités et statistiques

Corrigé de l'examen du 19 mai 2010

Problème 1

1. L'intégrale de la densité de la loi gamma valant 1, on a

$$\int_0^\infty t^{n-1}\,\mathrm{e}^{-\theta t}\,\mathrm{d}t = \frac{(n-1)!}{\theta^n}\;.$$

2. On se donne $Y_1 \sim \gamma(n,\theta)$ et $Y_2 \sim \gamma(1,\theta)$, indépendantes.

   (a) La densité de Y_1+Y_2 est donnée par la convolution

$$\begin{aligned} f_{Y_1+Y_2}(t) &= \int_{-\infty}^{\infty} f_{Y_1}(s) f_{Y_2}(t-s)\,\mathrm{d}s \\ &= \int_{-\infty}^{\infty} \frac{\theta^n}{(n-1)!} s^{n-1}\,\mathrm{e}^{-\theta s}\,\theta\,\mathrm{e}^{-\theta(t-s)}\,1_{\{0<s<t\}}\,\mathrm{d}s \\ &= \frac{\theta^{n+1}}{(n-1)!}\,\mathrm{e}^{-\theta t}\int_0^t s^{n-1}\,\mathrm{d}s\,1_{\{t>0\}} \\ &= \frac{\theta^{n+1}}{n!} t^n\,\mathrm{e}^{-\theta t}\,1_{\{t>0\}}\;. \end{aligned}$$

   Y_1+Y_2 suit donc une loi $\gamma(n+1,\theta)$.

   (b) Par récurrence sur N, la somme de N variables i.i.d. de loi $\gamma(1,\theta)$ suit une loi $\gamma(N,\theta)$. En effet, c'est trivialement vrai pour $N=1$, et le résultat précédent montre que si c'est vrai pour $N=n$, c'est encore vrai pour $N=n+1$.

3. Si X admet la densité $f_X(t) = \theta t^{\theta-1} 1_{\{0<t<1\}}$, on a

$$\mathbb{E}(X) = \int_0^1 t\,\theta t^{\theta-1}\,\mathrm{d}t = \frac{\theta}{\theta+1}$$

et

$$\mathbb{E}(X^2) = \int_0^1 t^2\,\theta t^{\theta-1}\,\mathrm{d}t = \frac{\theta}{\theta+2}\;.$$

On en déduit que la variance de X vaut

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{\theta}{(\theta+1)^2(\theta+2)}\;.$$

4. Soit $Z = \log(1/X) = -\log(X)$. Alors $X = g(Z)$ où $g(t) = \mathrm{e}^{-t}$ est strictement décroissante. La formule de changement de variable montre que la densité de Z est donnée par

$$f_Z(t) = f_X(g(t))|g'(t)| = \theta(\mathrm{e}^{-t})^{\theta-1} 1_{\{0<\mathrm{e}^{-\theta t}<1\}}\,\mathrm{e}^{-t} = \theta\,\mathrm{e}^{-\theta t}\,1_{\{t>0\}}\;.$$

Par conséquent, Z suit une loi $\gamma(1,\theta)$, c-à-d une loi exponentielle de paramètre θ. En appliquant le point 1., on trouve $\mathbb{E}(Z) = 1/\theta$.

5. La fonction de vraisemblance est donnée par

$$L_x(\theta) = \theta^n x_1^{\theta-1} \dots x_n^{\theta-1} 1_{\{0<x_i<1, i=1,\dots,n\}} \ .$$

Pour $x \in]0,1[^n$, la log-vraisemblance vaut donc

$$\log L_x(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^n \log x_i \ .$$

Comme

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log L_x(\theta) = \frac{n}{\theta} + \sum_{i=1}^n \log x_i \qquad \text{et} \qquad \frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log L_x(\theta) = -\frac{n}{\theta^2} < 0 \ ,$$

la vraisemblance admet un unique maximum en $\hat{\theta}$, où $\hat{\theta}^{-1} = -\frac{1}{n} \sum_{i=1}^n \log x_i$. L'estimateur du maximum de vraisemblance est donc donné par

$$W_n = \frac{n}{-\sum_{i=1}^n \log X_i} = \frac{n}{\sum_{i=1}^n \log \frac{1}{X_i}} \ .$$

6. Soit $S_n = \sum_{i=1}^n \log(1/X_i) = \sum_{i=1}^n Z_i$. Les $Z_i = \log(1/X_i)$ sont i.i.d. et suivent une loi $\gamma(1, \theta)$ en vertu du point 4. On sait donc par le point 2. que S_n suit une loi $\gamma(n, \theta)$. Ceci permet de calculer (en utilisant la question 1.)

$$\mathbb{E}\left(\frac{1}{S_n}\right) = \int_0^\infty \frac{1}{t} \frac{\theta^n}{(n-1)!} t^{n-1} \, \mathrm{e}^{-\theta t} \, \mathrm{d}t = \frac{\theta}{n-1} \ .$$

Comme $W_n = n/S_n$, on en déduit $\mathbb{E}(W_n) = \frac{n}{n-1}\theta$, donc le biais de W_n vaut

$$\mathbb{E}(W_n) - \theta = \frac{n}{n-1}\theta - \theta = \frac{\theta}{n-1} \ .$$

L'estimateur W_n est donc biaisé, mais asymptotiquement sans biais.

7. Par la loi faible des grands nombres, S_n/n converge en probabilité vers $\mathbb{E}(Z) = 1/\theta$. La fonction $x \mapsto 1/x$ étant continue en $1/\theta > 0$, W_n converge en probabilité vers θ. L'estimateur W_n est donc consistant.

   Le risque quadratique de W_n est donné par

$$\mathbb{E}\left[(W_n - \theta)^2\right] = \mathbb{E}(W_n^2) - 2\theta \mathbb{E}(W_n) + \theta^2 \ .$$

   Comme

$$\mathbb{E}\left(\frac{1}{S_n^2}\right) = \int_0^\infty \frac{1}{t^2} \frac{\theta^n}{(n-1)!} t^{n-1} \, \mathrm{e}^{-\theta t} \, \mathrm{d}t = \frac{\theta^2}{(n-1)(n-2)} \ ,$$

   on obtient

$$\mathbb{E}(W_n^2) = \frac{n^2}{(n-1)(n-2)}\theta^2 \ ,$$

   et finalement

$$\mathbb{E}\left[(W_n - \theta)^2\right] = \frac{n+2}{(n-1)(n-2)}\theta^2 \ .$$

   Le risque quadratique converge donc vers 0 lorsque $n \to \infty$, c-à-d que W_n converge vers θ en moyenne quadratique.

Problème 2

1. Soit Z_1 une variable aléatoire réelle suivant une loi exponentielle de paramètre $\lambda > 0$.

   (a) La densité de Z_1 est $f_{Z_1}(t) = \lambda \, e^{-\lambda t} \, 1_{\{t>0\}}$.

   (b) On pose $N_1(t) = 1_{\{Z_1 \leqslant t\}}$. L'image de $N_1(t)$ est l'ensemble $\{0, 1\}$. On a

$$\mathbb{P}\{N_1(t) = 0\} = \mathbb{P}\{Z_1 > t\} = \int_t^\infty \lambda \, e^{-\lambda t} \, dt = e^{-\lambda t} \; ,$$

   et par conséquent $\mathbb{P}\{N_1(t) = 1\} = 1 - e^{-\lambda t}$.

2. Soit Z_2 une variable aléatoire réelle indépendante de Z_1, et de même loi que Z_1.

   (a) La densité conjointe de (Z_1, Z_2) est donnée par

$$f_{Z_1,Z_2}(z_1, z_2) = f_{Z_1}(z_1) f_{Z_2}(z_2) = \lambda^2 \, e^{-\lambda(z_1+z_2)} \, 1_{\{z_1>0, z_2>0\}} \; .$$

   (b) On pose $X_1 = Z_1$ et $X_2 = Z_1 + Z_2$. On a $(Z_1, Z_2) = g(X_1, X_2)$ avec $g(x_1, x_2) = (x_1, x_2 - x_1)$. L'application linéaire g est un difféomorphisme de Jacobien

$$\left| \det \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \right| = 1 \; ,$$

   tel que $g^{-1}\{(z_1, z_2) \colon z_1 > 0, z_2 > 0\} = \{(x_1, x_2) \colon 0 < x_1 < x_2\}$. Par la formule de changement de variables, la densité conjointe de (X_1, X_2) est donnée par

$$f_{X_1,X_2}(x_1, x_2) = f_{Z_1,Z_2}(x_1, x_2 - x_1) = \lambda^2 \, e^{-\lambda x_2} \, 1_{\{0<x_1<x_2\}} \; .$$

   (c) Les densités marginales de X_1 et X_2 sont données respectivement par

$$\int_{-\infty}^\infty f_{X_1,X_2}(x_1, x_2) \, dx_2 = \lambda^2 \int_{x_1}^\infty e^{-\lambda x_2} \, dx_2 \, 1_{\{x_1>0\}} = \lambda \, e^{-\lambda x_1} \, 1_{\{x_1>0\}} \; ,$$

$$\int_{-\infty}^\infty f_{X_1,X_2}(x_1, x_2) \, dx_1 = \lambda^2 \, e^{-\lambda x_2} \int_0^{x_2} dx_1 \, 1_{\{x_2>0\}} = \lambda^2 x_2 \, e^{-\lambda x_2} \, 1_{\{x_2>0\}} \; .$$

   On aurait pu les obtenir sans faire de calculs, puisque $X_1 = Z_1$ suit une loi exponentielle $\gamma(1, \lambda)$ et $X_2 = Z_1 + Z_2$ suit une loi $\gamma(2, \lambda)$.

   (d) On pose $N_2(t) = 1_{\{X_1 \leqslant t\}} + 1_{\{X_2 \leqslant t\}}$. L'image de $N_2(t)$ est l'ensemble $\{0, 1, 2\}$ et

$$\begin{aligned}\mathbb{P}\{N_2(t) = 0\} &= \mathbb{P}\{X_1 > t, X_2 > t\} \\ &= \lambda^2 \int_t^\infty \underbrace{\int_t^\infty e^{-\lambda x_2} \, 1_{\{x_2>x_1>0\}} \, dx_2}_{= \int_{x_1}^\infty e^{-\lambda x_2} \, dx_2 = \frac{1}{\lambda} e^{-\lambda x_1}} \, dx_1 = e^{-\lambda t} \; .\end{aligned}$$

   On aurait aussi pu remarquer que $\mathbb{P}\{X_1 > t, X_2 > t\} = \mathbb{P}\{X_1 > t\} = e^{-\lambda t}$ puisque $X_2 > X_1$. Pour la même raison, on a $\mathbb{P}\{X_1 > t, X_2 \leqslant t\} = 0$, donc

$$\begin{aligned}\mathbb{P}\{N_2(t) = 1\} &= \mathbb{P}\{X_1 \leqslant t, X_2 > t\} + \mathbb{P}\{X_1 > t, X_2 \leqslant t\} \\ &= \lambda^2 \int_0^t \underbrace{\int_t^\infty e^{-\lambda x_2} \, 1_{\{x_2>x_1>0\}} \, dx_2}_{= \int_t^\infty e^{-\lambda x_2} \, dx_2 = \frac{1}{\lambda} e^{-\lambda t}} \, dx_1 + 0 = \lambda t \, e^{-\lambda t} \; .\end{aligned}$$

   Finalement, $\mathbb{P}\{N_2(t) = 2\} = 1 - (1 + \lambda t) \, e^{-\lambda t}$.

3. Soient $Z_1, \dots, Z_n$ des variables i.i.d. de loi exponentielle de paramètre λ. Pour $k = 1, \dots n$, on pose

$$X_k = \sum_{i=1}^{k} Z_i \;, \qquad N_k(t) = \sum_{i=1}^{k} 1_{\{X_i \leqslant t\}} \;.$$

(a) La loi conjointe de $(Z_1, \dots, Z_n)$ est donnée par

$$f_{Z_1,\dots,Z_n}(z_1, \dots, z_n) = \lambda^n \,\mathrm{e}^{-\lambda(z_1+\dots+z_n)}\, 1_{\{z_1>0,\dots,z_n>0\}} \;.$$

Le difféomorphisme permettant d'exprimer Z en fonction de X s 'écrit

$$g(x_1, \dots, x_n) = (x_1, x_2 - x_1, \dots, x_n - x_{n-1}) \;.$$

Sa matrice jacobienne est triangulaire, avec des 1 sur la diagonale, donc son Jacobien vaut 1. Le théorème de changement de variable montre que la densité conjointe de $(X_1, \dots, X_n)$ vaut

$$\begin{aligned} f_{X_1,\dots,X_n}(x_1, \dots, x_n) &= f_{Z_1,\dots,Z_n}(x_1, x_2 - x_1, \dots, x_n - x_{n-1}) \\ &= \lambda^n \,\mathrm{e}^{-\lambda x_n}\, 1_{\{0<x_1<\dots<x_n\}} \;. \end{aligned}$$

On a donc nécessairement $\mathbb{P}\{X_1 < X_2 < \dots < X_n\} = 1$, puisque la densité est nulle en dehors de cet ensemble.

(b) La densité conjointe de $(X_1, \dots, X_{n-1})$ est donnée par

$$\int_{x_{n-1}}^{\infty} \lambda^n \,\mathrm{e}^{-\lambda x_n}\, 1_{\{0<x_1<\dots<x_{n-1}\}} \,\mathrm{d}x_n = \lambda^{n-1} \,\mathrm{e}^{-\lambda x_{n-1}}\, 1_{\{0<x_1<\dots<x_{n-1}\}} \;.$$

Par récurrence, on trouve pour la densité conjointe de $(X_1, \dots, X_k)$

$$f_{X_1,\dots,X_k}(x_1, \dots, x_k) = \lambda^k \,\mathrm{e}^{-\lambda x_k}\, 1_{\{0<x_1<\dots<x_k\}} \;, \qquad k = 1, \dots, n \;.$$

(c) La densité conjointe de $(X_2, \dots, X_k)$ est donnée par

$$\int_{0}^{x_2} \lambda^k \,\mathrm{e}^{-\lambda x_k}\, 1_{\{x_2<\dots<x_k\}} \,\mathrm{d}x_1 = \lambda^k x_2 \,\mathrm{e}^{-\lambda x_k}\, 1_{\{x_2<\dots<x_k\}} \;.$$

Par récurrence, on trouve pour la densité conjointe de $(X_l, \dots, X_k)$

$$f_{X_l,\dots,X_k}(x_l, \dots, x_k) = \frac{\lambda^k}{(l-1)!} x_l^{l-1} \,\mathrm{e}^{-\lambda x_k}\, 1_{\{x_l<\dots<x_k\}} \;, \qquad l = 1, \dots, k \;.$$

(d) Comme $X_1 < X_2 < \dots < X_n$, on a $\{X_{k+1} \leqslant t\} \subset \{X_k \leqslant t\}$, donc $1_{\{X_{k+1} \leqslant t\}} \leqslant 1_{\{X_k \leqslant t\}}$ pour tout $k \leqslant n-1$. L'événement $N_n(t) = k$ ne peut donc se réaliser que si les k premiers X_i sont inférieurs ou égaux à t, alors que les autres X_i sont supérieurs à t. En d'autres termes, on a $\mathbb{P}\{N_n(t) = k\} = \mathbb{P}\{X_k \leqslant t, X_{k+1} > t\}$ pour $k = 1, \dots, n-1$.

(e) Il suit que

$$\begin{aligned} \mathbb{P}\{N_n(t) = k\} &= \int_0^t \int_t^\infty f_{X_k,X_{k+1}}(x_k, x_{k+1}) \,\mathrm{d}x_{k+1} \,\mathrm{d}x_k \\ &= \frac{\lambda^{k+1}}{(k-1)!} \int_0^t x_k^{k-1} \int_t^\infty \mathrm{e}^{-\lambda x_{k+1}} \,\mathrm{d}x_{k+1} \,\mathrm{d}x_k = \frac{(\lambda t)^k}{k!} \,\mathrm{e}^{-\lambda t} \end{aligned}$$

pour $k = 1, \dots, n-1$, alors que $\mathbb{P}\{N_n(t) = n\}$ se déduit du fait que la somme des $\mathbb{P}\{N_n(t) = k\}$ vaut 1. Faisant tendre n vers l'infini, on obtient que la loi de la limite est une loi de Poisson de paramètre λt.