Examen Final Statistique Inférentielle

Exercice 01 :

Partie I : En 1897, l'économiste suisse Vilfredo Pareto (1848-1923), professeur d'économie politique à l'Université de Lausanne, eut l'idée de modéliser la loi du revenu en postulant que le nombre relatif de personnes dont le revenu dépasse une valeur x est inversement proportionnel à une puissance de x. La définition suivante fut adoptée : une variable aléatoire X absolument continue suit une distribution de Pareto avec les paramètres α et θ si sa densité est donnée par

$$f(x;\alpha,\theta) = kx^{-\alpha}\mathbf{1}_{x\geq\theta}.$$

où $\theta > 0$ et $\alpha > 1$.

(1) Démontrer que la constante de normalisation $k = (\alpha-1)\,\theta^{\alpha-1}$.

(2) Ecrivez la fonction de répartition F_X de X.

(3) Calculez l'espérance $E(X)$. Précisez la condition d'existence.

(4) Calculez la variance $V(X)$. Précisez la condition d'existence.

(5) Calculez le moment d'ordre s $E(X^s)$ $(s \in \mathbb{N}^*)$. Précisez la condition d'existence.

(6) Soit Y une variable aléatoire définie par $Y = (\alpha-1)\ln\dfrac{X}{\theta}$.

(a) Donner le support de la variable aléatoire Y.

(b) Démontrer que $Y \sim Exp(1)$.

Partie II : Soient les variables aléatoires $X_1, ..., X_n$ i.i.d. de loi de Pareto de paramètres α et θ.

(1) Ecrire la fonction de vraisemblance de la loi de Pareto.

(2) Supposons que θ est connu.

(a) Donnez une statistique exhaustive pour α.

(b) Cette statistique est-elle minimale pour α.

(c) Proposez un estimateur $\hat{\alpha}$ de α par la méthode du maximum de vraisemblance.

Exercice 02 : Soit $X = (X_1, X_2, X_3)$ un vecteur gaussien centré de matrice de covariance

$$\Gamma_X = \begin{pmatrix} 3 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(1) Montrer que X_1 et X_3 sont indépendantes, et que X_2 et X_3 sont également indépendantes.

(2) Ecrire la densité de probabilité du vecteur $X = (X_1, X_2, X_3)$.

(3) Considérons le vecteur $Y = (X_1, X_2)$.

(a) Démontrer que Y est un vecteur gaussien.

(b) Calculer la densité de probabilité du vecteur $Y = (X_1, X_2)$.

(c) En déduire la moyenne $E(Y)$ et la matrice de covariance Γ_Y du vecteur Y.

(4) Montrer que pour tout $a \in \mathbb{R}$ le vecteur $(X_2, aX_1 + X_2)$ est un vecteur gaussien.

(5) Déterminer la valeur de a telle que $cov(X_2, aX_1 + X_2) = 0$.

Correction de l'examen

Ex01 (13pts)

1) On a : $$\int_{-\infty}^{+\infty} f(x;\alpha,\theta)\,dx = 1 \iff \int_{-\infty}^{+\infty} k x^{-\alpha} \mathbb{1}_{x\geq\theta}\,dx = 1$$

$$\iff k\int_{\theta}^{+\infty} x^{-\alpha}\,dx = 1 \iff k\left[\frac{1}{1-\alpha}x^{1-\alpha}\right]_{\theta}^{+\infty} = 1$$ (1pt)

$$\iff k\left[\frac{(+\infty)^{1-\alpha}}{1-\alpha} - \frac{1}{1-\alpha}\theta^{1-\alpha}\right] = 1 \iff k = (\alpha-1)\theta^{\alpha-1} \Rightarrow f(x;\alpha,\theta) = \frac{(\alpha-1)}{\theta}\left(\frac{\theta}{x}\right)^{\alpha}\mathbb{1}_{x\geq\theta}$$

2) $$F_X(x) = \int_{-\infty}^{x} f(t;\alpha,\theta)\,dt = \int_{-\infty}^{x}(\alpha-1)\theta^{\alpha-1}x^{-\alpha}\mathbb{1}_{x\geq\theta}\,dt$$

$$= \frac{\alpha-1}{\theta^{1-\alpha}}\int_{\theta}^{x} t^{-\alpha}\,dt = \frac{\alpha-1}{\theta^{1-\alpha}}\left[\frac{t^{1-\alpha}}{1-\alpha}\right]_{\theta}^{x} = \frac{\alpha-1}{\theta^{1-1}}\left[\frac{x^{1-\alpha}}{1-\alpha} - \frac{\theta^{1-\alpha}}{1-\alpha}\right]$$

$$F_X(x) = \begin{cases} 1 - \left(\frac{\theta}{x}\right)^{\alpha-1} & \forall x \geq 0. \\ 0 & \text{sinon (si } x < 0) \end{cases}$$ (1pt)

3) $$E(X) = \int_{0}^{+\infty} x f(x;\alpha,\theta)\,dx = \int_{\theta}^{+\infty}(\alpha-1)\theta^{\alpha-1}x^{1-\alpha}\,dx = (\alpha-1)\theta^{\alpha-1}\left[\frac{x^{2-\alpha}}{2-\alpha}\right]_{\theta}^{+\infty}$$

$$= (\alpha-1)\theta^{\alpha-1}\frac{\theta^{2-\alpha}}{\alpha-2} \quad \text{pour } 2-\alpha<0 \iff \alpha>2.$$

$$= \frac{\alpha-1}{\alpha-2}\theta \quad . \alpha>2.$$ (1pt)

4) Calculons d'abord

$$E(X^2) = \int_{0}^{+\infty} x^2 f(x;\alpha,\theta)\,dx = \int_{\theta}^{+\infty}(\alpha-1)\theta^{\alpha-1}x^{2-\alpha}\,dx$$

$$= (\alpha-1)\,\theta^{\alpha-1}\left[\frac{x^{3-\alpha}}{3-\alpha}\right]_{\theta}^{+\infty} = (\alpha-1)\,\theta^{\alpha-1}\,\frac{\theta^{3-\alpha}}{\alpha-3} \quad \text{si } \alpha>3$$

$$= \frac{\alpha-1}{\alpha-3}\,\theta^2 \quad \text{si } \alpha>3$$

$$\text{Var}(X) = E(X^2) - E^2(X) = \frac{\alpha-1}{\alpha-3}\,\theta^2 - \frac{(\alpha-1)^2}{(\alpha-2)^2}\,\theta^2$$

$$= \frac{(\alpha-1)(\alpha-2)^2 - (\alpha-1)^2(\alpha-3)}{(\alpha-2)^2(\alpha-3)}\,\theta^2$$

$$= \frac{(\alpha-1)\left[(\alpha-2)^2 - (\alpha-1)(\alpha-3)\right]}{(\alpha-2)^2(\alpha-3)}\,\theta^2$$

$$= \frac{\alpha-1}{(\alpha-2)^2(\alpha-3)}\,\theta^2 \quad \text{pour } \alpha>3 .$$ (1pt)

5) $$E(X^S) = \int_0^{+\infty} x^S f(x;\alpha,\theta)\,dx = \int_{\theta}^{+\infty} (\alpha-1)\,\theta^{\alpha-1}\,x^{S-\alpha}\,dx$$

$$= (\alpha-1)\,\theta^{\alpha-1}\left[\frac{x^{S-\alpha+1}}{S-\alpha+1}\right]_{\theta}^{+\infty} \quad \text{si } \alpha>S+1$$

$$= (\alpha-1)\,\theta^{\alpha-1}\,\frac{\theta^{S-\alpha+1}}{\alpha-S-1} = \frac{\alpha-1}{\alpha-S-1}\,\theta^S \quad \text{pour } \alpha>S+1 .$$ (1pt)

6) a) On a $\quad Y = (\alpha-1)\ln\left(\frac{X}{\theta}\right) \Leftrightarrow X = \theta\, e^{\frac{Y}{\alpha-1}}$

et si $X \geq \theta \Rightarrow \frac{X}{\theta} \geq 1 \Rightarrow \ln\left(\frac{X}{\theta}\right) \geq \ln(1) = 0$

$\Rightarrow (\alpha-1)\ln\left(\frac{X}{\theta}\right) \geq 0 .$

$\Rightarrow Y \geq 0 \quad$ Support de Y est $D_Y = \mathbb{R}_*^+ .$ (1pt)

6) On Pose $Y = g(X) = (\alpha-1) \ln\left(\frac{X}{\theta}\right)$.

$h(Y) = \theta\, e^{y/\alpha-1}$

et $J \equiv g'(X) = \frac{\alpha-1}{X}$.

$$f_Y(y) = f_X(h(y)) \left[g'(h(y))\right]^{-1}$$

$$= \frac{(\alpha-1)\theta^{\alpha-1}}{\left(\theta e^{y/(\alpha-1)}\right)^{\alpha}} \frac{\theta e^{y/\alpha-1}}{\alpha-1} = e^{-y}$$

Donc $Y \sim Exp(1)$. (1pt)

7) a) $\mathcal{L}(\alpha,\theta;x) = \prod_{i=1}^{n} f_{X_i}(x_i;\alpha,\theta)$

$$= \prod_{i=1}^{n} (\alpha-1)\, \theta^{\alpha-1}\, x_i^{-\alpha}\, 1_{x_i \geq \theta}$$

$$= (\alpha-1)^n\, \theta^{n(\alpha-1)} \prod_{i=1}^{n} x_i^{-\alpha} \prod_{i=1}^{n} 1_{x_i \geq \theta}.$$ (2pts)

b) i) D'après le théorème de Factorisation de F-N On a,

$$L(\alpha;x) = g(t(x);\alpha)\, h(x)$$

telles que $t(x) = \prod_{i=1}^{n} x_i^{-\alpha}$ et $g(y,\alpha) = (\alpha-1)^n\, \theta^{n(\alpha-1)}\, y$

$$h(x) = \prod_{i=1}^{n} 1_{x_i \geq \theta}$$

ou

$\mathcal{L}(\alpha;x) = g(t(x);\alpha)\, h(x)$ telles que

$t(x) = \prod_{i=1}^{n} x_i^{\alpha}$, $g(t(x),\alpha) = (\alpha-1)^n\, \theta^{n\alpha}\, t(x)$

$$h(x) = \frac{\prod_{i=1}^{n} 1_{x_i \geq \theta}}{\theta^n} = \frac{1_{\max_i\{x_i\} \geq \theta}}{\theta^n}.$$ (1pt)

iii) $\ell(\alpha;x) = \ln\left[\mathcal{L}(\alpha;x)\right] = \ln\left[(\alpha-1)^n \theta^{n(\alpha-1)} \left(\prod_{i=1}^n x_i\right)^{-\alpha} 1_{x_i \geq \theta}\right]$

$$= n\ln(\alpha-1) + n(\alpha-1)\ln\theta - \alpha\sum_{i=1}^n \ln(X_i) + \sum_{i=1}^n 1_{x_i \geq \theta}$$

$$\frac{\partial}{\partial\alpha}\ell(\alpha;x) = \frac{n}{\alpha-1} + n\ln\theta - \sum_{i=1}^n \ln(X_i)$$

$$= \frac{n}{\alpha-1} + \sum_{i=1}^n \ln\theta - \sum_{i=1}^n \ln(X_i)$$

$$= \frac{n}{\alpha-1} + \sum_{i=1}^n \left(\frac{X_i}{\theta}\right)$$

$$\frac{\partial}{\partial\alpha}\ell(\alpha;x) = 0 \Leftrightarrow \frac{n}{\alpha-1} + \sum_{i=1}^n \frac{X_i}{\theta} = 0 \Leftrightarrow \hat{\alpha} = 1 + \frac{n}{\sum_{i=1}^n \ln\left(\frac{X_i}{\theta}\right)}.$$

(2pts)

iv) On calcul $\frac{f(x;\alpha)}{f(y;\alpha)}$. On a :

$$\frac{f(x;\alpha)}{f(y;\alpha)} = \frac{(\alpha-1)^n \theta^{n(\alpha-1)} \left(\prod_{i=1}^n x_i\right)^{-\alpha} \prod_{i=1}^n 1_{x_i \geq \theta}}{(\alpha-1)^n \theta^{n(\alpha-1)} \left(\prod_{i=1}^n y_i\right)^{-\alpha} \prod_{i=1}^n 1_{y_i \geq \theta}}$$

$$= \left(\frac{\prod_{i=1}^n x_i}{\prod_{i=1}^n y_i}\right)^{-\alpha} \frac{\prod_{i=1}^n 1_{x_i \geq \theta}}{\prod_{i=1}^n 1_{y_i \geq \theta}}$$ ne dépend pas de α ssi

$$\frac{\prod_{i=1}^n x_i}{\prod_{i=1}^n y_i} = 1 \Leftrightarrow \prod_{i=1}^n x_i = \prod_{i=1}^n y_i \Leftrightarrow t(x) = t(y)$$

donc la statistique t est minimale pour α.

(1pt)

Exo 2 : (7pts)

1) La densité de probabilité de la loi gaussienne est

$$f(x)=f(x_1,x_2,x_3)=\frac{1}{(2\pi)^{3/2}\sqrt{\det\Gamma}}\exp\left[\frac{-1}{2}(x-M)^t\,\Gamma^{-1}(x-M)\right]$$

telles que M est la moyenne donc $E(X)=M=0$ (vecteur centré)

et $\Gamma=\begin{pmatrix}3&1&1\\1&2&0\\0&0&1\end{pmatrix}\Rightarrow \det\Gamma=5$, donc

$$f(x)=\frac{1}{2\sqrt{10\pi^3}}\exp\left[\frac{-1}{2}\begin{pmatrix}x_1&x_2&x_3\end{pmatrix}\begin{pmatrix}3&1&1\\1&2&0\\0&0&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}\right]$$

$$=\frac{2}{2\sqrt{10\pi 3}}\exp\left\{\frac{-1}{2}\left[3x_1^2+2x_1x_2+2x_2^2+x_3^2\right]\right\}$$ (1pt)

2) D'après la matrice de covariance Γ, on obtient :

$$\Gamma=\begin{pmatrix}Var(X_1)&Cov(X_1,X_2)&Cov(X_1,X_3)\\Cov(X_2,X_1)&Var(X_2)&Cov(X_2,X_3)\\Cov(X_3,X_1)&Cov(X_3,X_2)&Var(X_3)\end{pmatrix}=\begin{pmatrix}3&1&0\\1&2&0\\0&0&1\end{pmatrix}\Leftrightarrow$$

$Cov(X_1,X_3)=Cov(X_2,X_3)=0\Leftrightarrow$ les variables X_1, X_3 et X_2, X_3 sont indépendantes

$\Leftrightarrow X_3$ et le vecteur (X_1,X_2) sont indépendantes.

3)a) On a $X=(X_1,X_2,X_3)$ est gaussien $\Rightarrow aX_1+bX_2+cX_3$ est gaussien $\Rightarrow$ si $c=0$ aX_1+bX_2 est gaussien (1pt)

b) On a :

$$f_{(X_1,X_2)}(x_1,x_2)=\int_{-\infty}^{+\infty}f_{(X_1,X_2,X_3)}(x_1,x_2,x_3)\,dx_3=\int_{-\infty}^{+\infty}f_X(x)\,dx_3$$

$$=\frac{1}{2\sqrt{10\pi^3}}\int_{-\infty}^{+\infty}\exp\left[\frac{-1}{2}\left(3x_1^2+2x_1x_2+2x_2^2+x_3^2\right)\right]dx_3$$

$$=\frac{1}{2\sqrt{10\pi 3}}\exp\left[\frac{-1}{2}\left(3x_1^2+2x_1x_2+2x_2^2\right)\right]\int_{-\infty}^{+\infty}e^{-\frac{1}{2}x_3^2}\,dx_3$$

$$=\frac{1}{2\pi\sqrt{5}}\exp\left[-\frac{1}{2}\left(3x_1^2+2x_1x_2+2x_2^2\right)\right].$$ (1pt)

3) On sait que

$$f_{(X_1,X_2)}(x_1,x_2) = \frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}\sqrt{1-\rho^2}} \operatorname{Exp}\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-E(X_1))^2}{\sigma_{X_1}^2} - 2\rho\frac{(x_1-E(X_1))(x_2-E(X_2))}{\sigma_{X_1}\sigma_{X_2}} + \frac{(x_2-E(X_2))^2}{\sigma_{X_2}^2}\right]\right\}$$

$$= \frac{1}{2\pi\sqrt{5}} \operatorname{Exp}\left[\frac{-1}{2}\left(3x_1^2 + 2x_1x_2 + 2x_2^2\right)\right]$$

Par Identification, on obtient:

$E(X_1) = E(X_2) = 0$.

$(1-\rho^2)\sigma_{X_1}^2 = 1/3$

$\frac{-2\rho}{(1-\rho^2)\sigma_X\sigma_Y} = 2$

$(1-\rho^2)\sigma_{X_2}^2 = 1/2$

$\Leftrightarrow \rho = \frac{-1}{\sqrt{6}}$ et $\sigma_X^2 = \frac{2}{5}$, $\sigma_Y^2 = \frac{3}{\sqrt{5}^2}$, $Cov(X,Y) = \rho\sigma_X\sigma_Y = \frac{-1}{5}$

donc $\Gamma = \begin{pmatrix} 2/5 & -\frac{1}{5} \\ -\frac{1}{5} & \frac{3}{5} \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$. (1pt)

4) On sait que (X_1, X_2, X_3) est gaussien $\Leftrightarrow \forall \alpha, \beta, \gamma \in \mathbb{R}$

$\alpha X_1 + \beta X_2 + \gamma X_3$ est gaussienne.

On prend $\alpha = a$, $\beta = 2$, $\gamma = 0$. donc $aX_1 + 2X_2$ est gaussien, Alors le vecteur

$(X_2, X_2 + aX_1)$ est gaussien. (1pt)

5) Comme $(X_2, X_2 + aX_1)$ est gaussien $\Leftrightarrow$ X_2 et $X_2 + aX_1$ sont indépendantes

$\Leftrightarrow Cov(X_2, X_2 + aX_1) = 0 \Leftrightarrow$ Donc, d'après la relation $Cov(X+Y, Z) = Cov(X,Z) + Cov(Y,Z)$

On écrit

$$Cov(X_2, X_2 + aX_1) = Cov(X_2, X_2) + Cov(X_2, aX_1)$$

$$= Cov(X_2, X_2) + a\,Cov(X_1, X_2) = Var(X_2) + a\,Cov(X_1, X_2)$$

(1pt)