Exercice 1 — Chose promise chose due — 3 points

1. Donnez un exemple de phénomène pouvant être modélisé par une loi de Pareto ?
2. Donnez le code en python permettant de visualiser un nuage de point de deux variables x et y représentant respectivement le département de naissance d'une personne et sa catégorie socio professionnelle ?
3. Lorsque l'on déclare qu'une carte n'est pas un territoire, l'OPM est-ce la carte ou le territoire ?
4. Soient X et Y deux variable aléatoires quantitatives. Montrez que, si $\mathbb{E}(Y) = 0$

$$cov(X, Y) = \mathbb{E}(XY)$$

5. Que mesure le risque d'un estimateur ?
6. A quoi sert l'information de Fischer ?

Exercice 2 — Corcovado — 3 points

Soient X et Y deux variables aléatoires indépendantes dont les lois sont données par

x	1	2	3
$P(X = x)$	1/3	1/3	1/3

et

y	0	2	4
$P(Y = y)$	1/4	1/2	1/4

1. Calculez la covariance entre X et Y.

Exercice 3 — La famille exponentielle normale et naturelle — 4 points

Soit X une variable aléatoire réelle suivant une loi normale de paramètres μ et σ^2 tous deux inconnus, et donc de densité :

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

1. Montrez que $f_X(x)$ peut s'écrire

$$f_X(x) = \sqrt{\frac{\theta_2}{\pi}} exp^{\left(-\theta_2 x^2 + \theta_1 x - \frac{\theta_1^2}{4\theta_2}\right)}$$

avec

$$\theta_1 = \frac{\mu}{\sigma^2} \qquad \text{et } \theta_2 = \frac{1}{2\sigma^2}$$

2. Montrez que les estimateurs du maximum de vraisemblance de θ_1 et θ_2 s'écrivent

$$\begin{cases} \widehat{\theta}_1 = \dfrac{\dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i}{\dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i^2 - \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i\right)^2} \\ \widehat{\theta}_2 = \dfrac{1}{2}\dfrac{1}{\dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i^2 - \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i\right)^2} \end{cases}$$

Exercice 4 Se lever dès potron minet c'est l'assurance de paraitre tôt 10 points

Soit $(X_1, \ldots, X_n)$ un échantillon i.i.d. dont la densité de la loi parente s'écrit, pour deux paramètres $\theta > 0$ et $a > 0$ inconnus,

$$f(x) = \begin{cases} \theta\, a^{\theta}\, x^{-(\theta+1)} & \text{si } x \geq a \\ 0 & \text{sinon} \end{cases} .$$

1. Calculer la log vraisemblance de l'échantillon.
2. En déduire l'estimateur maximum de vraisemblance de θ connaissant a.
3. Quelle est la borne de Cramer-Rao du paramètre θ ?
4. Quelle est la distribution asymptotique de l'estimateur du maximum de vraisemblance de θ connaissant a ?
5. Construire un estimateur efficace d'une fonction de θ connaissant a et donnez sa variance.
6. Donner l'estimateur maximum de vraisemblance de a connaissant θ.

Formulaire

— fréquences $\widehat{f_i} = \frac{n_i}{n}$ où n est le nombre total d'observations et n_i le nombre d'observation de la modalité i

— moyenne : $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \hat{f}_i x_i$

— variance empirique : $\widehat{\sigma}_x^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$

— espérance d'une variable aléatoire discrète : $\mathbb{E}(X) = \displaystyle\sum_i x_i \mathbb{P}(x_i)$.

— espérance d'une variable aléatoire continue de densité $f(x)$: $\mathbb{E}(X) = \displaystyle\int x f(x) dx$.

— variance d'une variable aléatoire X : $Var(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big)$.

— covariance de deux variables aléatoires : $C(X,Y) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\Big)$

— médiane : $\mathbb{P}(X < M) = 0,5$

— mode : $\underset{x\in\Omega}{Argmax}\{\mathbb{P}(x)\}$

— fractiles à l'ordre p, $\forall p \in [0,1], \widehat{\Phi}_p$ telle que $\widehat{\mathbb{P}}(X \leq \widehat{\Phi}_p) = p$

— les quartiles :
  — $\widehat{\Phi}_{\frac{1}{4}} = \widehat{Q}_1,$ telle que $\widehat{F}(\widehat{Q}_1) = \frac{1}{4}$,
  — $\widehat{\Phi}_{\frac{1}{2}} = \widehat{Q}_2 = \widehat{M},$ telle que $\widehat{F}(\widehat{M}) = \frac{1}{2}$,
  — $\widehat{\Phi}_{\frac{3}{4}} = \widehat{Q}_3,$ telle que $\widehat{F}(\widehat{Q}_2) = \frac{3}{4}$.

— covariance de deux échantillons : $c_{xy} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$

— corrélation : $\mathrm{cor}(x,y) = \dfrac{c_{xy}}{\sqrt{\widehat{\sigma}_x^2\widehat{\sigma}_y^2}}$

— probabilité conditionnelle : $\mathbb{P}(X = x_i | Y = y_j) = \dfrac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)}$

— densité conditionnelle : $f_{X|Y}(x|y) = \dfrac{f_{(X,Y)}(x,y)}{f_Y(y)}$

— espérance conditionnelle (v.a. discrète) : $\mathbb{E}[Y|X = a] = \displaystyle\sum_{i=1}^{n} y_i \mathbb{P}(Y = y_i | X = a)$

— La loi des grands nombres : $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i \quad \underset{n\to\infty}{\longrightarrow} \quad \mathbb{E}(X)$

— Le théorème central limite : $\sqrt{n}\Big(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i - \mathbb{E}(X)\Big) \quad \underset{n\to\infty}{\overset{(d)}{\longrightarrow}} \quad \mathcal{N}(0,\sigma^2)$

Estimateurs :

— Le risque : $R_{\widehat{\theta}}(\theta) = \mathbb{E}\big((\widehat{\theta}(X_1,\ldots,X_n) - \theta)^2\big)$
  — Le biais d'un estimateur : $\mathbb{E}(\widehat{\theta}) - \theta$
  — La variance d'un estimateur : $\mathbb{E}\Big(\big(\widehat{\theta} - \mathbb{E}(\widehat{\theta})\big)^2\Big)$

— Soit X une variable aléatoire de distribution $f_\theta(x)$ et $(X_1,\ldots,X_n)$ une famille de n variables aléatoires i.i.d. ayant pour loi parente la loi de X dépendant d'un paramètre $\theta \in \mathbb{R}$.
  — Vraisemblance :

$$L(\theta, X_1, \ldots, X_n) = \prod_{i=1}^{n} f_\theta(X_i)$$

Log vraisemblance :

$$\ell(\theta, X_1, \ldots, X_n) = log\, L(\theta, X_1, \ldots, X_n) = \sum_{i=1}^{n} log\, f_\theta(X_i)$$

Le score :

$$s(\theta, X_1, \ldots, X_n) = \frac{\partial \ell(\theta, X_1, \ldots, X_n)}{\partial \theta}$$

— L'information de Fisher : $I_n(\theta) = Var\Big(\dfrac{\partial \ell(\theta, X_1, \ldots, X_n)}{\partial \theta}\Big) = -\mathbb{E}\left(\dfrac{\partial^2 \ell(\theta, X_1, \ldots, X_n)}{\partial \theta^2}\right)$

— Pour $\theta \in \mathbb{R}^p$, L'information de Fisher est la matrice $p \times p$

$$I_n(\theta) = \mathbb{E}\big(\nabla \ell(\theta, X_1, \ldots, X_n) \nabla \ell(\theta, X_1, \ldots, X_n)^\top\big)$$

— Borne de Cramer Rao

  — pour un estimateur sans biais de θ

$$BCR(\theta) = \frac{1}{I_n(\theta)} \;\leq\; \mathrm{Var}\big(\widehat{\theta}(X_1, \ldots, X_n)\big)$$

  — Pour un estimateur sans biais d'une fonction de θ : $\mathbb{E}(\widehat{u}(\theta)) = u(\theta)$

$$BCR(\theta) = \frac{u'(\theta)}{I_n(\theta)} \;\leq\; \mathrm{Var}\big(\widehat{u}(\theta(X_1, \ldots, X_n)\big)$$

  — pour un estimateur biaisé de biais B

$$BCR(\theta) = \frac{1 + B'(\theta)}{I_n(\theta)} \;\leq\; \mathrm{Var}\big(\widehat{\theta}(X_1, \ldots, X_n)\big)$$

Variables aléatoires et lois

— Soit $Y \sim \mathcal{N}(0, 1)$ une variable aléatoire normale centrée réduite.

— Soit $Y_1, Y_2, ..., Y_n$ un échantillon i.i.d. de taille n de loi parente $\mathcal{N}(0, 1)$.

— On appelle loi du χ^2 à n degrés de libertés la loi de la variable aléatoire $Z_n = \sum_{i=1}^n Y_i^2$. L'espérance et la variance de cette loi sont $\mathbb{E}(Z_n) = n$ et $Var(Z_n) = 2n$

— On appelle loi de student à n degrés de libertés la loi de la variable aléatoire T_n

$$T_n = \frac{N}{\sqrt{\frac{X_n}{n}}} \qquad \begin{array}{l} N \sim \mathcal{N}(0, 1) \\[1em] X_n \sim \chi_n^2 \end{array}$$