PARTIEL 2019 - 2020

(3 HEURES – DOCUMENTS NON AUTORISÉS)

Exercice 1 (Modèle Poisson-Gamma, 6pts)

On considère des observations $X_1, \dots, X_n$ i.i.d de loi de Poisson $\mathcal{P}(\lambda)$ d'intensité $\lambda > 0$ dont on rappelle la densité $f_\lambda(x) = \lambda^x e^{-\lambda}/x!$ par rapport à la mesure de comptage sur $\mathbb{N}$.

1. (0.5pt) Calculez $\mathbb{E}(X_1)$ et $\mathrm{Var}(X_1)$

2. (0.5pt) Montrez que l'estimateur $\widehat{\lambda}_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais et calculez son risque quadratique. Donnez la loi limite de $\sqrt{n}(\widehat{\lambda}_n - \lambda)$.

3. (1,5pts) Construisez deux intervalles de confiance pour λ asymptotiquement de niveau $1-\alpha$, où $\alpha \in (0,1)$, l'un en utilisant le lemme de Slutsky et l'autre en utilisant la Δ-méthode.

4. (1pt) Construisez un test d'hypothèse $H_0 : \lambda = \lambda_0$ contre $H_1 : \lambda \neq \lambda_0$, où $\lambda_0 > 0$ est fixé, asymptotiquement de niveau $\alpha \in (0,1)$. Calculez la p-valeur de ce test.

5. (1pt) On rappelle que la loi Gamma de paramètres de forme $a > 0$ et d'intensité $\tau > 0$, notée $\mathrm{Gamma}(a,\tau)$, a pour densité $g_{a,\tau}(x) = \frac{\tau^a}{\Gamma(a)} x^{a-1} e^{-\tau x} \mathbf{1}_{x \geq 0}$ contre la densité de Lebesgue sur $\mathbb{R}$. Si G est une variable aléatoire de loi $\mathrm{Gamma}(a,\tau)$, calculez $\mathbb{E}(G)$ et $\mathrm{Var}(G)$.

6. (1pt) On se place maintenant dans un cadre bayesien. On suppose que $X_1, \dots, X_n$ sont i.i.d et de loi conditionnelle à $\boldsymbol{\lambda} = \lambda$ de loi $\mathcal{P}(\lambda)$ et que $\boldsymbol{\lambda}$ suit la loi a priori $\mathrm{Gamma}(a,\tau)$. On introduit $X = (X_1, \dots, X_n)$ et $x = (x_1, \dots, x_n) \in \mathbb{N}^n$. Montrez que la loi a posteriori de $\boldsymbol{\lambda}$ sachant $X = x$ est $\mathrm{Gamma}(\sum_{i=1}^n x_i + a, n + \tau)$. Que dit-on alors concernant les lois Poisson et Gamma ?

7. (1,5pts) Montrez que l'estimateur bayesien de λ pour la perte quadratique est égal à $\widehat{\lambda}_n^B = (\sum_{i=1}^n X_i + a)(n+\tau)$. Calculez son risque quadratique et le risque bayesien.

Exercice 2 (ANOVA, 4pts)

On considère des variables aléatoires $X_{i,j} \sim \mathcal{N}(m_i, \sigma^2)$ pour $i = 1, \dots, k$ et $j = 1, \dots, n_i$ toutes indépendantes. Cela revient à observer k échantillons gaussiens de tailles respectives $n_1, n_2, \dots, n_k$, notés $X_{i,\bullet} = [X_{i,1} \cdots X_{i,n_i}]^\top \in \mathbb{R}^{n_i}$. On suppose que les paramètres $m = [m_1 \cdots m_k]^\top \in \mathbb{R}^k$ et $\sigma^2 > 0$ sont inconnus. On veut construire un test d'hypothèses

$$H_0 : m_1 = m_2 = \cdots = m_k \quad \text{contre} \quad H_1 : \exists i \neq i' \text{ tels que } m_i \neq m_{i'},$$

autrement dit, on veut tester si les moyennes des k échantillons sont égales.

1. (1pt) On considère le vecteur aléatoire $X = (X_{i,j})_{1 \leq i \leq k, 1 \leq j \leq n_i} \in \mathbb{R}^n$ avec $n = \sum_{i=1}^k n_i$. Montrez que $\mu = \mathbb{E}(X)$ appartient à un sous-espace vectoriel E de dimension k. Calculez $X_E = \mathrm{proj}_E(X)$ la projection orthogonale de X dans E.

2. (1pt) Montrez que l'hypothèse nulle s'écrit $H_0 : \mu \in F$, où F est un sous-espace vectoriel de E de dimension 1. Calculez X_F.

3. (2pts) Construisez alors un test de Fisher pour les hypothèses décrites au dessus.

Exercice 3. (Sur le range d'un échantillon, 3pts)

Soient $X_1, \ldots, X_n$ des variables aléatoires réelles i.i.d. On pose

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad \text{range}(X_1, \ldots, X_n) = \max_{1\leq i\leq n} X_i - \min_{1\leq i\leq n} X_i.$$

1. (1,5pt) On suppose que les $X_1, \ldots, X_n$ sont de loi $\mathcal{N}(m, \sigma^2)$. Montrez que $\overline{X}_n$ et range$(X_1, \ldots, X_n)$ sont indépendants.

2. (1,5pt) On suppose que les $X_1, \ldots, X_n$ sont de loi exponentielle $\mathcal{E}(\lambda)$ d'intensité $\lambda > 0$. Déterminez la loi de range$(X_1, \ldots, X_n)$.

Exercice 4. (Loi de Student, 4pts)

Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$ indépendante d'une suite $(Y_n)_{n \geq 1}$ de variables aléatoires où Y_n suit la loi $\chi^2(n)$ du Chi-deux à n degrés de liberté. On rappelle que la loi $\chi^2(n)$ correspond à une loi Gamma$(n/2, 1/2)$ dont la densité est rappelée dans l'Exercice 1. On pose $T_n = X/\sqrt{Y_n/n}$.

1. (1,5pts) Montrez que la densité t_n de T_n est donnée par

$$t_n(x) = \frac{1}{\sqrt{\pi}}\frac{\Gamma((n+1)/2)}{\sqrt{n}\,\Gamma(n/2)}\frac{1}{(1+x^2/n)^{(n+1)/2}}.$$

2. (1,5pts) Montrez que

$$\lim_{n\to\infty} t_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

pour tout $x \in \mathbb{R}$. On pourra tout d'abord montrer que $\Gamma(a+1/2)/\Gamma(a) \sim \sqrt{a}$ quand $a \to \infty$ en remarquant que $\Gamma(a+1/2)/\Gamma(a) = \mathbb{E}[\sqrt{U}]$ où U suit une loi $\Gamma(a, 1)$.

3. (1pts) En déduire que T_n converge en loi vers $\mathcal{N}(0, 1)$.

Exercice 5. (Modèle linéaire hétéroscédastique, 3pts)

On observe un vecteur aléatoire $Y = [Y_1 \cdots Y_n]^\top \in \mathbb{R}^n$ de loi $\mathcal{N}(X\beta, \sigma^2 S)$ où

$$X = \begin{bmatrix} x_{1,1} & \ldots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \ldots & x_{n,p} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{et} \quad S = \begin{bmatrix} f(x_{1,1}) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & f(x_{n,1}) \end{bmatrix},$$

avec $f(x_{i,1}) > 0$ pour tout $i = 1, \ldots, n$. On suppose que l'on observe la matrice X, que X est de rang plein et que la fonction f est connue. On souhaite estimer les paramètres inconnus $\beta \in \mathbb{R}^p$ et $\sigma^2 > 0$. Il s'agit d'un modèle linéaire gaussien *hétéroscédastique*, par opposition au modèle linéaire *homoscédastique* étudié dans le cours où $S = I_n$.

1. (1pt) Construisez une matrice M telle que $Z = MY$ suive un modèle linéaire homoscédastique. En déduire un estimateur $\widehat{\beta}$ de β en utilisant la procédure des moindres carrés sur Z.

2. (2pts) On peut aussi construire l'estimateur des moindres carrés directement sur Y en calculant $\widetilde{\beta} = \arg\min_{b\in\mathbb{R}^p} \|Y - Xb\|^2$. Comparez $\mathbb{E}\big[\|\widehat{\beta} - \beta\|^2\big]$ et $\mathbb{E}\big[\|\widetilde{\beta} - \beta\|^2\big]$ et conclure.