Université Paris I, Panthéon - Sorbonne

Première Année Master M.A.E.F. 2017 – 2018

Statistiques I

Examen final, janvier 2018

Examen de 3h. Tout document ou calculatrice est interdit.

1. **(Sur 12 points)** Soit la variable X qui suit une loi normale $\mathcal{N}(m\,,\,\sigma^2)$, où $m \in \mathbf{R}$ et $\sigma^2 \in]0,\infty[$ sont deux paramètres inconnus. Soit également une suite $(\varepsilon_k)_k$ de variables i.i.d. de loi $\mathcal{N}(0\,,\,\sigma^2)$, indépendantes de X. Au final, on note $\theta = {}^t(m,\sigma^2)$ et on observe $(X_1,\cdots,X_n)$ défini par
$$X_k = X + \varepsilon_k \quad \text{pour tout } k \in \mathbf{N}.$$
   (a) Déterminer le modèle statistique paramétrique dominé induit par $(X_1,\cdots,X_n)$ après avoir montré que la loi de probabilité de $(X_1,\cdots,X_n)$ est gaussienne **(1.5pts)**.
   (b) Déterminer la loi de X_i pour tout i **(0.5pts)**.
   (c) Démontrer que Σ_n matrice de variance-covariance de $(X_1,\cdots,X_n)$ est une matrice avec $2\sigma^2$ sur la diagonale et σ^2 partout ailleurs **(0.5pts)**. Montrer que $\det(\Sigma_n) = (n+1)\sigma^{2n}$ **(1.5pts)**.
   (d) Démontrer que Σ_n^{-1} est la matrice avec $\frac{1}{\sigma^2}\,\frac{n}{n+1}$ sur la diagonale et $-\frac{1}{\sigma^2}\,\frac{1}{n+1}$ partout ailleurs **(1pt)**.
   (e) En déduire la vraisemblance de $(X_1,\cdots,X_n)$ **(1pt)**. Montrer que le modèle appartient à la famille exponentielle **(2.5pts)** et démontrer que $\big(\sum_{i=1}^n X_i\,,\, n\sum_{i=1}^n X_i^2 - 2\sum_{1\leq i<j\leq n} X_iX_j\big)$ est une statistique exhaustive complète pour ce modèle **(0.5pts)**.
   (f) Démontrer que $\widehat{m}_n = \frac{1}{n}\sum_{i=1}^n X_i$ est l'estimateur de m sans biais uniformément de variance minimale parmi les estimateurs sans biais de m **(1pt)**. Déterminer la loi de $\widehat{m}_n$ **(1pt)**. Est-ce un estimateur convergent **(1pt)**?

2. **(Sur 18 points)** Soit une suite $(\varepsilon_{ij})_{i,j\in\mathbf{N}}$ de v.a.i.i.d. de loi Bernoulli de paramètre $p \in]0,1[$. Pour $m \in \mathbf{N}$, on définit
$$X_j = \sum_{i=1}^m \varepsilon_{ij} \quad \text{pour } j \in \mathbf{N}^*$$
(par convention $\sum_{i=1}^0 \varepsilon_{ij} = 0$). On observe $(X_1,\ldots,X_n)$ où $n \in \mathbf{N}^*$.
   (a) Pour $m \in \mathbf{N}$ fixé, démontrer que $(X_j)_{j\in\mathbf{N}}$ forme une suite de v.a.i.i.d. de loi à préciser **(1pt)**.
   (b) Tout d'abord, on suppose que m est connu et p inconnu. Déterminer le modèle statistique induit par $(X_1,\ldots,X_n)$ **(0.5pts)**, montrer qu'il est dominé **(0.5pts)** et appartient à la famille exponentielle **(0.5pts)**. Déterminer un estimateur de p sans biais et efficace **(1pt)** en précisant la borne de Cramèr-Rao **(0.5pts)**.
   (c) On suppose maintenant que m est également inconnu et on note $\theta = {}^t(p,m)$. Déterminer le modèle statistique induit par $(X_1,\ldots,X_n)$ **(0.5pts)**, montrer qu'il est dominé **(0.5pts)** et qu'il n'appartient pas à la famille exponentielle **(0.5pts)**. Soit l'estimateur $\widehat{\theta}_n = {}^t(\widehat{p}_n,\widehat{m}_n)$ avec $\widehat{m}_n = \max_{1\leq i\leq n} X_i$ et $\widehat{p}_n = \frac{1}{\widehat{m}_n n}\sum_{i=1}^n X_i$. Montrer que $\mathbb{P}(\max_{1\leq i\leq n} X_i = m) = 1-(1-p^m)^n$ **(1.5pts)**. En déduire que $\widehat{m}_n$ est un estimateur convergent de m **(0.5pts)**. En déduire que $\widehat{p}_n$ est également un estimateur convergent de p (utiliser par exemple la fonction de répartition de $\widehat{p}_n$...) **(2.5pts)**.

(d) On suppose enfin que

$$X_j = \sum_{i=1}^{T_j} \varepsilon_{ij} \quad \text{pour } j \in \mathbf{N}^*$$

où les (T_j) sont des v.a.i.i.d. de loi de Poisson de paramètre $\lambda > 0$ inconnu, indépendantes des (ε_{ij}). Le vecteur de paramètre inconnu est maintenant $\theta = {}^t(p, \lambda)$. Déterminer le modèle statistique induit par $(X_1, \ldots, X_n)$ **(0.5pts)**, montrer qu'il est dominé **(0.5pts)**. Montrer que pour $m \geq k$ alors $\mathbb{P}(X_j = k \cap T_j = m) = e^{-\lambda} \frac{\lambda^m}{k!(m-k)!} p^k (1-p)^{m-k}$ **(1.5pts)**. En déduire que $(X_j)_j$ est une suite de v.a.i.i.d. de loi de Poisson de paramètre $\lambda\, p$ (on pourra faire le changement d'indice $m' = m-k$...) **(1.5pts)**. En déduire que le modèle est exponentiel **(1pt)**. Montrer que $\sum_{i=1}^n X_i$ est une statistique exhaustive et complète pour le modèle **(1pt)**. Démontrer cependant qu'il n'y a pas unicité de l'estimateur du maximum de vraisemblance de $\theta = {}^t(p, \lambda)$ et non convergence de n'importe lequel d'entre eux **(2pts)**.