Université Paris I, Panthéon - Sorbonne

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

Statistiques I

Contrôle continu n°1, novembre 2017

Examen de 1h30. Tout document ou calculatrice est interdit.

1. **(Sur 12 points)** Soit $(X_1, \ldots, X_n)$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant une loi de Weibull, c'est-à-dire une loi dont la fonction de répartition est:
$$F(x) = 1 - \exp\Big(- \Big(\frac{x}{\lambda}\Big)^{\beta}\Big) \quad \text{pour } x \geq 0,$$
avec $\theta = (\lambda, \beta) \in]0, \infty[^2$ inconnu.
   (a) Déterminer $F(x)$ pour $x < 0$ **(0.5pts)**.
   (b) Déterminer le modèle statistique paramétrique induit par $(X_1, \ldots, X_n)$ **(0.5pts)** et montrer qu'il est dominé **(1pt)**.
   (c) Préciser la vraisemblance de ce modèle **(1pt)**.
   (d) Le modèle fait-il partie de la famille exponentielle? (justifier...) **(3pts)**
   (e) Montrer que la statistique $\widehat{T} = (X_1, X_2, \ldots, X_n)$ est exhaustive minimale pour ce modèle **(4pts)**.
   (f) On suppose que β est connu. Préciser alors le modèle statistique et déterminer une statistique exhaustive complète pour ce modèle **(2pts)**.

2. **(Sur 16 points)** Soit une suite $(\varepsilon_i)_{n\geq 1}$ de variables aléatoires indépendantes et identiquement distribuées suivant une loi de Bernoulli de paramètre $p \in]0, 1[$ inconnu. On définit également la suite $(X_n)_{n\geq 1}$ par
$$X_{n+1} = \varepsilon_{n+1}\, X_n \quad \text{pour } n \in \mathbf{N}^*, \text{ et} \quad X_1 = \varepsilon_1.$$
On observe $(X_1, \ldots, X_n)$ où $n \in \mathbf{N}^*$.
   (a) Pour $k \in \mathbf{N}^*$ fixé, écrire X_k en fonction des ε_i **(0.5pts)**. Quelles sont les valeurs possibles pour X_k **(0.5pts)**? Quelle est sa loi **(0.5pts)**? Les variables (X_k) sont-elles indépendantes **(1.5pts)**? Identiquement distribuées **(0.5pts)**?
   (b) Montrer que le vecteur $(X_1, \ldots, X_n)$ ne peut prendre que $(n + 1)$ valeurs que l'on précisera **(1pt)**. En déduire le modèle statistique induit par ce vecteur (on n'explicitera pas sa mesure de probabilité) **(0.5pts)**.
   (c) Démontrer que pour $(x_1, \ldots, x_n)$ appartenant au modèle statistique **(1pt)**:
$$\mathbb{P}\Big((X_1, \ldots, X_n) = (x_1, \ldots, x_n)\Big) = \mathbb{P}(X_1 = x_1) \prod_{k=2}^{n} \mathbb{P}\Big(X_k = x_k \;|(X_1, \ldots, X_{k-1}) = (x_1, \ldots, x_{k-1})\Big).$$
   (d) Montrer que $\mathbb{P}\Big(X_k = x_k \;|(X_1, \ldots, X_{k-1}) = (x_1, \ldots, x_{k-1})\Big) = \mathbb{P}\Big(X_k = x_k \;|X_{k-1} = x_{k-1}\Big)$ **(1pt)**, puis que $\mathbb{P}\Big(X_k = x_k \;|X_{k-1} = x_{k-1}\Big) = p^{x_k}(1 - p)^{x_{k-1} - x_k}$ pour $k \geq 2$ **(2pts)**.
   (e) En déduire l'expression de la vraisemblance du modèle **(1pt)**.
   (f) Le modèle appartient-il à la famille exponentielle **(1pt)**?
   (g) Peut-on en déduire que $\widehat{T} = (\sum_{i=1}^{n} X_i \,,\, X_n)$ est une statistique exhaustive complète pour le modèle **(3pts)**?
   (h) Déterminer l'estimateur par maximum de vraisemblance de p (c'est-à-dire la valeur de $p \in]0, 1[$ qui maximise la vraisemblance du modèle prise en $(X_1, \ldots, X_n)$) et vérifier que c'est une fonction de $\widehat{T}$ **(2pts)**.