Université de Strasbourg
L3 - Mathématique 4 janvier 2011

Examen de Statistique Mathématique (3h)

Les téléphones portables doivent être éteints et rangés.
L'usage des calculatrices n'est pas autorisé.
Les notes de cours et autres documents sont interdits.
La qualité de la rédaction sera très largement prise en compte.

Exercice 1.
L'entreprise Granulex distribue un aliment pour chat dans un contenant métallique dont le poids après remplissage, représenté par une variable aléatoire X, est calibré à une valeur nominale de 350g. On prélève un échantillon de 100 boites afin de vérifier la calibration de la chaine de production. On obtient $\sum_{i=1}^{100} X_i = 36000g$ et $\sum_{i=1}^{100} X_i^2 = 12\,970\,000g^2$. Des études antérieures de la chaine de production avaient déjà montré que le poids est distribué selon une loi normale.

1. Proposer un estimateur des paramètres de la loi normale en justifiant votre choix. Déterminer les estimations associées.
2. Estimer la probabilité qu'un contenant choisi au hasard de la production ait un poids entre 340g et 360g ?
3. L'entreprise met en production 1000 contenants. Estimer le nombre moyen de boites de poids inférieur à 330g.
4. Estimer la probabilité que le contenant le plus lourd parmi 100 contenants prélevés soit inférieur à 360g.
5. Estimer la probabilité que le contenant le plus léger parmi 100 contenants prélevés soit supérieur à 330g.

NB : les résultats sont à exprimer en fonction de Φ la fonction de répartition de la loi normale centrée réduite.

Exercice 2.
Soit $(X_1, ..., X_n)$ un échantillon de variables aléatoires i.i.d. de variable parente X dont la distribution admet la densité suivante par rapport à la mesure de Lebesgue sur $\mathbb{R}$:

$$f_\mu(x) = \frac{1}{2} \exp\left(-|x-\mu|\right) \quad \text{avec } \mu \in \mathbb{R}.$$

1. Vérifier qu'il s'agit bien d'une densité de probabilité.
2. Déterminer l'espérance, la variance et la médiane de X.
3. Déterminer l'information de Fisher du modèle probabiliste (à une observation).
4. Le modèle appartient-il à la famille exponentielle ?
5. Déterminer $\widetilde{\mu}_n$ l'estimateur de μ obtenu par la méthode des moments. Etablir son comportement asymptotique.
6. Déterminer $\widehat{\mu}_n$ l'estimateur de μ obtenu par la méthode du maximum de vraisemblance.

Exercice 3.
On souhaite estimer le ratio de l'espérance de vie moyenne des non-fumeurs par rapport à l'espérance de vie moyenne des fumeurs :

$$\theta := \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}$$

en notant X la variable aléatoire représentant la durée de vie des non-fumeurs et Y la variable aléatoire représentant la durée de vie des fumeurs.
Soit $(X_1, ..., X_{n_1})$ un échantillon i.i.d. distribué comme X et soit $(Y_1, ..., Y_{n_2})$ un échantillon i.i.d. distribué comme Y, indépendant de $(X_1, ..., X_{n_1})$.

1. Proposer un estimateur $\widehat{\theta} = T(X_1, ..., X_{n_1}, Y_1, ..., Y_{n_2})$ de θ et justifier votre démarche.
2. On rappelle que la delta-méthode affirme que si
$$\sqrt{n}\left((U_n, V_n) - (u, v)\right) \xrightarrow{\mathcal{D}} (U, V)$$
alors, pour toute fonction Ψ continûment différentiable sur le domaine de (U, V),
$$\sqrt{n}\left(\Psi(U_n, V_n) - \Psi(u, v)\right) \xrightarrow{\mathcal{D}} D\Psi(u, v).\begin{pmatrix} U \\ V \end{pmatrix}$$
Déterminer la loi asymptotique de $\widehat{\theta}$ lorsque $n_1 = n_2 = n$.
3. En déduire la construction d'un intervalle de confiance bilatéral asymptotique pour θ de niveau de confiance $(1 - \alpha)$.

Exercice 4.
Soient X et Y des variables aléatoires représentant des quantités que l'on suspecte être liées. Plus particulièrement, on se demande s'il existe une influence de $\log(X)$ sur Y. Soient (X_i, Y_i) pour $i = 1, ..., n$ des triplets i.i.d. distribués comme (X, Y).

1. On souhaite ajuster un modèle de régression linéaire de la forme $Y = a\log(X) + b + \varepsilon$ où ε est une variable aléatoire centrée indépendante de X et de variance σ^2. Comment interpréter ce modèle ? Que valent $\mathbb{E}[Y|X]$ et $\mathrm{Var}(Y|X)$?
2. Déterminer $\widehat{a}$ et $\widehat{b}$ les estimateurs de a et b respectivement obtenus par la méthode des moindres carrés i.e. satisfaisant le problème d'optimisation suivant :
$$(\widehat{a}, \widehat{b}) = \mathrm{argmin}_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a\log(X_i) - b)^2.$$

Exercice 5.

1. Les vecteurs suivants sont-ils bien des vecteurs gaussiens de $\mathbb{R}^3$?
$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \simeq \mathcal{N}\left(\begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 6 & -3 & -2 \\ -3 & 3 & 2 \\ -2 & -2 & 3 \end{pmatrix}\right); \quad \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \simeq \mathcal{N}\left(\begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 1 & 2 \\ 1 & 9 & -1 \\ 2 & -1 & -3 \end{pmatrix}\right);$$
$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \simeq \mathcal{N}\left(\begin{pmatrix} 2 \\ 1 \\ 5 \end{pmatrix}, \begin{pmatrix} 6 & -3 & -2 \\ -3 & 3 & 2 \\ -2 & 2 & 3 \end{pmatrix}\right).$$
2. Soit $\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$ un vecteur aléatoire de loi $\mathcal{N}\left(\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 6 & -3 & 0 \\ -3 & 3 & 2 \\ 0 & 2 & 3 \end{pmatrix}\right)$.
Calculer $\mathbb{E}[X - Z|Y - X]$.En déduire la loi de $\mathbb{E}[X - Z|Y - X]$.