Université Paris-Sud **Préparation à l'agrégation**
Département de mathématiques **Option Probabilités-Statistiques**

Devoir de Statistiques 1

Exercice 1 [Loi de Poisson]

Soit $X_1, \ldots, X_n$ un échantillon de loi de Poisson de paramètre θ, où θ est un réel strictement positif. La densité correspondante est notée $p_\theta(x)$.

(1) A quels estimateurs pouvez-vous penser pour estimer le paramètre θ ?

(2) Calculez la log-vraisemblance associée aux observations pour le paramètre θ et en déduire que le maximum de vraisemblance $\hat{\theta}_n$ est égal à la moyenne empirique.

(3) Ecrivez un programme simulant N réalisations de l'estimateur $\hat{\theta}_n$ et illustrez sa consistance, que vous vérifierez également par le calcul.

(4) Notons $\dot{l}_\theta(x)$ la dérivée partielle en θ de $(\theta, x) \to \log p_\theta(x)$ et $I_\theta = \mathbb{E}(\dot{l}_\theta(X)^2)$ l'information de Fisher (où X suit la loi $\mathcal{P}(\theta)$). Montrez que I_θ vaut $1/\theta$.

(5) Le N-échantillon construit en (3) a-t-il une variance proche de $(nI_\theta)^{-1}$? Pourquoi ?

(6) Prouvez et illustrez la normalité asympototique de l'estimateur. On pourra tracer sur un même graphique la densité de la loi normale centrée réduite et un histogramme de $\sqrt{1000/\theta}(\hat{\theta}_{1000} - \theta)$.

Partie facultative (questions 7 à 12) :

Il est connu qu'une suite de variables aléatoires de lois binomiales $\mathcal{B}(n, p_n)$ converge en loi vers une Poisson de paramètre θ dès que $np_n \to \theta$. C'est ce que l'on appelle l'approximation binomiale par Poisson. On a en plus la vitesse de convergence suivante. Soit B_n une suite de variables aléatoires de lois $B(n, \theta/n)$, alors :

$$D(n) = \sup_{k \geq 0} \left| \mathbb{P}(B_n = k) - \frac{e^{-\theta}\theta^k}{k!} \right| \leq \frac{2\theta^2}{n}.$$

(7) Montrer (théoriquement) que lorsque n tend vers l'infini, $\mathbb{P}(B_n = k)$ tend vers $e^{-\theta}\theta^k/k!$

(8) Illustrer graphiquement la proximité de la loi de B_n avec la loi $\mathcal{P}(\theta)$ pour $\theta = 5$ et les valeurs de n suivantes : 10, 50, 100, 150.

(9) A partir d'un échantillon $B_n^{(1)}, \ldots, B_n^{(N)}$, estimer $\mathbb{P}(B_n = k)$ pour $0 \leq k \leq n$. Puis, à partir du même échantillon, donner un estimateur de $D(n)$ (en supposant ici θ connu).

(10) Donner un intervalle de confiance pour l'estimation de $D(n)$ (n fixé) à l'aide du N-échantillon précédent. On pourra utiliser l'inégalité de Hoeffding.

(11) Verifier numériquement ou graphiquement si la distance entre les deux lois est bien inférieure à la borne $2\theta^2/n$.

(12) Pour $n = 100$ et $n = 300$, à partir d'un échantillon $B_n^{(1)}, \ldots, B_n^{(N)}$, effectuez un test du χ^2 de niveau 1% pour déterminer si cet échantillon suit la loi de Poisson $\mathcal{P}(\theta)$.

Exercice 2 [**Mélange de lois**] (voir exercice 5 du TP2)

On se donne deux fonctions de répartition F_1, F_2 de deux variables aléatoires Z_1, Z_2 et un nombre positif $p \in]0,1[$. Alors $F(t) = pF_1(t) + (1-p)F_2(t)$ est encore une fonction de répartition. On dit que la nouvelle loi obtenue est le *mélange* de la loi de Z_1 et de la loi de Z_2. Attention, cela n'a rien à voir avec la loi de $pZ_1 + (1-p)Z_2$.

(1) Prouver que F est la fonction de répartition de la v.a. X suivante :

$$X = \begin{cases} Z_1 & \text{si } Y = 1 \\ Z_2 & \text{si } Y = 0 \end{cases}$$

où Y est une variable de Bernoulli de paramètre p, indépendante du couple (Z_1, Z_2).

On se place maintenant dans le cas où $Z_1 \sim \mathcal{U}([0,1])$ et $Z_2 \sim \mathcal{U}([0,\theta])$ avec $\theta \geq 1$. On note $F_{\theta,p}$ la fonction de répartition de X dans ce cas.

(2) Ecrire une fonction Scilab qui permette de simuler un n-échantillon de fonction de répartition $F_{\theta,p}$. Contrôlez graphiquement que votre fonction simule bien la loi $F_{\theta,p}$.

(3) On suppose qu'on observe un n-échantillon $X_1, \ldots, X_n$ de même loi que X, et qu'on cherche à estimer à la fois p et θ. Montrer que si l'on exclut une valeur de θ à préciser, ce modèle statistique est identifiable.

(4) Montrer que $\hat{\theta} = \max_{1 \leq i \leq n} X_i$ est un estimateur faiblement consistant.

(5) Quelle est la densité de X ? Calculer $\mathbb{E}(X)$, en déduire un estimateur $\hat{p}$ de p. Montrer que si $\theta > 1$, $\hat{p}$ converge en probabilité vers p. Que se passe-t-il dans le cas $\theta = 1$?

(6) Ecrire une fonction Scilab qui prend en argument un vecteur (supposé être un échantillon de loi $F_{\theta,p}$) et fournit un estimateur de θ et un estimateur de p. Vérifier en prenant comme entrée une sortie de votre fonction implémentée à la question (2).

(7) Pour $\theta > 1$, montrer que $n(\theta - \hat{\theta})$ converge en loi vers une loi exponentielle $\mathcal{E}((1-p)/\theta)$. En déduire un intervalle de confiance asymptotique pour θ de niveau $1 - \alpha$.

(8) Pour $n = 30$, déterminer numériquement le niveau réel de cet intervalle de confiance pour $\alpha = 0.05$, dans le cas $p = 0.7, \theta = 2$.

(9) **(Facultatif)** Donner une procédure de test pour tester $\theta = 1$ contre $\theta > 1$.