Université Bordeaux I – 2010/11 **MASTER 1**
Probabilités-Statistiques **M.-L. Chabanol**

Devoir surveillé
Mars 2011 – 3 heures. Aucun document autorisé

Dans tout le sujet, on pourra faire intervenir la fonction quantile q de la loi normale centrée réduite, ainsi que la fonction quantile $q^{\chi^2_n}$ d'une loi du χ_2 à n degrés de liberté. On pourra utiliser le cas échéant $q_{0.975} \simeq 1.96$ et $q_{0.95} \simeq 1.65$.
D'autre part on rappelle le développement asymptotique $\sum_{k=1}^n \frac{1}{k} = \ln(n) + \gamma + o(1)$ où γ est la constante d'Euler, ainsi que l'égalité $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$

<u>Problème 1</u>. Soit X une variable aléatoire de loi exponentielle de paramètre λ (donc de densité $\lambda \exp(-\lambda x) 1_{\mathbf{R}^+}(x)$)

1. Déterminer la loi de la partie entière de X.
2. Compléter la fonction ci-dessous pour qu'elle simule une variable aléatoire qui suit une loi géométrique de paramètre p
   fonction X = rgeom(p)
   U=rand(1)
   ...

   end

<u>Problème 2</u>. On dispose d'une balance dont les résultats ont une erreur normale centrée de variance σ^2. Pour peser deux objets de poids a et b on effectue trois pesées : d'abord chaque objet séparément, puis les deux objets simultanément. On modélise ceci en disant que le résultat des pesées est un vecteur aléatoire $Y = m + \sigma\epsilon$, où $m = \begin{pmatrix} a \\ b \\ a+b \end{pmatrix}$ et où ϵ suit une loi $\mathcal{N}(0, I_3)$ (on a noté I_3 la matrice identité 3×3). Y est donc un vecteur gaussien de loi $\mathcal{N}(m, \sigma^2 I_3)$.
On note $\langle ., . \rangle$ le produit scalaire, $u = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $v = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, H le plan engendré par u et v,
et $w = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ un vecteur unitaire normal à H.
On note P_H la projection orthogonale sur H.

1. Quelle est la loi de $||P_H(\epsilon)||^2$?
2. Quelle est la loi du produit scalaire $\langle w, Y \rangle$? Quelle est la loi de $\frac{\langle w, Y \rangle^2}{\sigma^2}$?
3. Donner la matrice dans la base canonique de la projection orthogonale P_H sur H (on pourra remarquer que si $X \in \mathbf{R}^3, P_H(X) = X - \langle w, X \rangle w$).
4. Quelle est la loi de $P_H(Y)$?
5. On suppose qu'on ne connaît pas a et b, et on prend comme estimateur de a, $\hat{a} = \dfrac{2Y_1 - Y_2 + Y_3}{3}$ et comme estimateur de b, $\hat{b} = \dfrac{-Y_1 + 2Y_2 + Y_3}{3}$. Ces estimateurs sont-ils sans biais ? Donner la loi de $\hat{a}$. Quel est le risque quadratique de $\hat{a}$?
6. On suppose qu'on connaît σ^2. Proposer un intervalle de confiance pour a de risque 0.05.
7. $\hat{a}$ et $\hat{b}$ sont-ils indépendants ? $\hat{a}$ et $\langle w, Y \rangle$ sont-ils indépendants ?

8. On suppose qu'on a perdu le manuel de la balance et qu'on ne connaît pas σ^2. On prend comme estimateur de σ^2 $\hat{S} = \langle w, Y\rangle^2$. Est-il sans biais ? Si $\alpha \in]0,1[$, proposer un intervalle de confiance pour σ de risque α.
9. Quelle est la loi de $\frac{\hat{a}}{\sqrt{\hat{S}}}$?

<u>PROBLÈME 3</u>. Soit $(X_k)_{k\in\mathbf{N}^*}$ une suite de variables aléatoires indépendantes de même loi de Pareto de densité $f_\alpha(x) = \frac{\alpha}{x^{\alpha+1}} 1_{x>1}$, où $\alpha > 0$ est un paramètre.

1. Déterminer $\{q \in \mathbf{R}^{+*}, X_1^q \text{ est intégrable}\}$.
2. Déterminer la fonction de répartition de X_1. Compléter la fonction matlab ci-dessous pour qu'elle fournisse un n-échantillon de loi de Pareto de paramètre α.
   fonction $X = pareto(n, \alpha)$
   U=rand(...)
   ...

   end

   On suppose désormais qu'on ne connaît pas α.
3. Calculer si elles existent $E_\alpha(X_1)$ et $Var_\alpha(X_1)$.
4. On suppose dans cette question seulement $\alpha > 1$. Montrer que $\tilde{\alpha}_n = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n X_i - n}$ est un estimateur convergent de α.
5. Donner la vraisemblance du modèle, et montrer que l'estimateur de maximum de vraisemblance de α est $\hat{\alpha}_n = \frac{n}{\sum_{i=1}^n \ln(X_i)}$.
6. Montrer que $\ln(X_1)$ suit une loi exponentielle dont on déterminera le paramètre.
7. L'estimateur $\hat{\alpha}_n$ est-il convergent ?
8. Montrer que $\sqrt{n}(\hat{\alpha}_n - \alpha)$ converge en loi vers une loi normale dont on donnera les paramètres.
9. Examiner la convergence en loi de $\sqrt{n}\dfrac{(\hat{\alpha}_n - \alpha)}{\hat{\alpha}_n}$.
10. Soit $\eta \in]0,1[$. Déduire de la question précédente un intervalle de confiance asymptotique pour α au risque η.

<u>PROBLÈME 4</u>. Votre petit frère collectionne les images d'animaux qu'on trouve dans des tablettes de chocolat. On suppose qu'il existe n images différentes (numérotées de 1 à n), que chaque tablette de chocolat contient une image, et qu'elles sont distribuées de manière équiprobable (il n'y a pas d'animal plus rare qu'un autre), et de manière indépendante sur chaque tablette.
On note X_i le numéro de l'image contenue dans la tablette i (les X_i sont donc supposées i.i.d. de loi uniforme sur $\{1, \ldots, n\}$), et N_k le nombre de tablettes achetées avant d'avoir k images différentes : $N_k = \inf\{j \geq 1, \mathrm{Card}\{X_i, i \leq j\} = k\}$. On note $T_k = N_k - N_{k-1}$, avec par convention $T_1 = N_1 = 1$.

1. (a) Soit T une variable géométrique de paramètre p. Donner la fonction caractéristique $E[e^{itT}]$. Que vaut $E[T]$? On pourra ultérieurement utiliser sans calcul $Var(T) = \frac{1-p}{p^2}$.
   (b) Soit $l_2 \in \mathbf{N}^*$. Exprimer l'événement $(T_2 = l_2)$ en fonction des $(X_i)_{i\in\mathbf{N}^*}$ et en déduire que T_2 suit une loi géométrique de paramètre $1 - \frac{1}{n}$.
   (c) Soit $(l_2, l_3) \in (\mathbf{N}^*)^2$. Exprimer l'événement $(T_2 = l_2 \text{ et } T_3 = l_3)$ en fonction des $(X_i)_{i\in\mathbf{N}^*}$ puis calculer $P(T_2 = l_2 \text{ et } T_3 = l_3)$.
   (d) En déduire que T_3 suit une loi géométrique de paramètre $\frac{n-2}{n}$, et qu'elle est indépendante de T_2.
   (e) Retrouver la loi de T_3 en exprimant T_3 comme instant de premier succès, puis montrer de même que pour tout $2 \leq k \leq n$, T_k suit une loi géométrique dont on donnera le paramètre. On admettra que les variables $T_1, \ldots, T_n$ sont indépendantes.
2. (a) Montrer $E[N_n] = n \sum_{j=1}^n \frac{1}{j}$ et en déduire que lorsque n tend vers $+\infty$, $E[N_n] = n\ln(n) + O(n)$.
   (b) Montrer $Var(N_n) = \frac{\pi^2}{6} n^2 + o(n^2)$.

(c) Montrer que $\frac{N_n}{n\ln(n)}$ converge en probabilité vers 1.

3. Soit $k \in \mathbf{N}$, fixé. Montrer que la suite $(\frac{T_{n-k+1}}{n}, n \geq k)$ converge en loi vers une loi exponentielle de paramètre k.
4. (Question bonus) Montrer plus précisément que si on note $\psi_{n,k}^{(T)}$ la fonction caractéristique de $\frac{T_{n-k+1}}{n}$ et $\psi_k^{(exp)}$ la fonction caractéristique d'une loi exponentielle de paramètre k on a

$$\forall u \in \mathbf{R}, \exists (n_0, C) \in \mathbf{N} \times \mathbf{R}^+, (n \geq \max(n_0, k) \Rightarrow |\psi_{n,k}^{(T)}(u) - \psi_k^{(exp)}(u)| \leq \frac{C}{nk})$$

5. Soit $(X_1, \ldots, X_n)$ un n-échantillon de loi exponentielle de paramètre 1. On note $M_n = \max(X_i)_{1 \leq i \leq n}$.
   (a) Calculer la fonction de répartition de M_n.
   (b) Montrer que la suite de terme général $M_n - \ln(n)$ converge en loi vers une variable aléatoire Z dont on donnera la fonction de répartition F_Z. Quelle est la densité de Z ?
   (c) Soit $0 < \alpha < 1$. Déterminer deux réels r_1 et r_2 tels que $P(Z \leq r_1) = P(Z > r_2) = \alpha$.

   On admettra pour la suite que M_n a même loi que $\sum_{j=1}^n E_j$ où les E_j sont des variables aléatoires indépendantes de loi exponentielle de paramètre j, et on notera $\psi_n^{(M)}$ sa fonction de répartition.
6. On note $\psi_n^{(N/n)}$ la fonction caractéristique de $\frac{N_n}{n}$.En utilisant les questions 4 et 5 ainsi que l'inégalité $|\prod_{j=1}^n a_j - \prod_{j=1}^n b_j| \leq \sum_{j=1}^n |a_j - b_j|$ valable pour tous complexes $(a_j)_{1\leq j\leq n}, (b_j)_{1\leq j\leq n}$ de modules inférieur ou égal à 1, montrer que pour tout réel u, il existe un entier n_0 et un réel C tels que pour $n \geq n_0$,

$$|\psi_n^{(N/n)}(u) - \psi_n^{(M)}(u)| \leq C\frac{\ln(n)}{n}$$

7. Montrer que la suite de terme général $\frac{N_n - n\ln(n)}{n}$ converge en loi vers la variable aléatoire Z.
8. Donner un intervalle de confiance I_n de risque asymptotique 0.05 pour N_n.
9. On suppose $n = 250$. Evaluer le nombre de plaquettes de chocolat que votre petit frère devra acheter (et potentiellement manger...) pour compléter sa collection (on suppose qu'il ne fait pas d'échange avec ses copains).