CENTRE DE TECHNIQUES DE PLANIFICATION ET D'ECONOMIE APPLIQUEE (CTPEA)

Cours : Probabilités I / Prof. : J-B ANTENORD / Examen Intra / Date: Mardi 4 octobre 2011

==

Documentation autorisée : Tables **Deuxième Année** **Durée : 2 heures 30**

==

<u>Exercices théoriques (60 points)</u>

1. Soit X une variable aléatoire continue de fonction de densité $f_x(x) = \begin{cases} \sqrt{\dfrac{\lambda}{2\pi x^3}} e^{-\lambda\frac{(x-\mu)^2}{2\mu x^2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$ et $\lambda > 0$.

Démontrer que $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ et $T = \dfrac{n}{\sum_{i=1}^{n}(1/X_i) - (1/\overline{X})}$ sont deux statistique exhaustives et complètes (15 pts).

2. Soit X une variable aléatoire continue de fonction de masse $f_x(x) = \begin{cases} e^{-\left(\frac{x-\theta}{\theta}\right)} \dfrac{1}{\theta}, & x \ge \theta \\ 0, & x < \theta \end{cases}$ et *θ>0*. On considère un n-échantillon $X_1, X_2, \ldots\ldots, X_n$ de variables aléatoires *iid* de même loi que X. (Lecoutre et Legait-Maille, pp.139, 163-164)

2.1 Déterminer l'estimateur des moments $\hat{\theta}_n^M$ de *θ*. Est-il sans biais ? Convergent ? Efficace ? Justifier rigoureusement. (15 points)

2.2 Déterminer la loi limite de $\hat{\theta}_n^M$ pour n très grand et en déduire un intervalle bilatéral symétrique pour θ de niveau voisin de 0.95. On donne : $\sum_{i=1}^{100} x_i = 660$. (5 points)

3. Soit X une variable aléatoire discrète de fonction de masse $p_x(x) = \begin{cases} \left(1 - \dfrac{1}{\theta}\right)^{x-1} \dfrac{1}{\theta}, & x \in IN^* \\ 0, & x \notin IN^* \end{cases}$ et *θ>1*. On admet que $E(X) = \theta$ et $V(X) = \theta(\theta - 1)$. On considère un n-échantillon $X_1, X_2, \ldots\ldots, X_n$ de variables aléatoires *iid* de même loi que X.

a) Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n^{MV}$ de *θ*. Vérifier qu'il s'agit effectivement d'un estimateur du maximum de vraisemblance. (5 points)
b) Montrer que $\hat{\theta}_n^{MV}$ obtenu en 1.1 est sans biais et convergent. Pour n grand, construire un intervalle symétrique à 95% de confiance pour θ . (10 points)
c) Montrer que $\hat{\theta}_n^{MV}$ est efficace au sens de la borne FDCR. (5 points)
d) On pose $p = \dfrac{1}{\theta}$. Déterminer un estimateur $\hat{p}_n$ de *p*. De quelle loi de probabilité connue s'agit il ? (5 points)

Problème (40 points)

Un coureur automobile doit choisir entre deux types de pneus A et B dont on a étudié l'usure. Le nombre de kilomètres que peut parcourir un pneu A (resp. B) avant usure est une variable aléatoire X (resp. Y) qui suit une loi normale de paramètres (μ_A, σ_A^2) (resp. (μ_B, σ_B^2)). Soit $(X_1, X_2, ..., X_n)$ et $(Y_1, Y_2, ..., Y_m)$ des échantillons de variables aléatoires *iid* de même loi respective que X et Y. On dispose des moyennes échantillonnales suivantes : $\bar{x}_9 = 20000$et $\bar{y}_{21} = 25000$.

1. Sachant que $\sigma_A = 2000$, $\sigma_B = 5000$, construire pour $\mu_A - \mu_B$, à un niveau de 0.95, un intervalle de confiance

1.1 Bilatéral symétrique (5 points)
1.2 Bilatéral dissymétrique avec $\alpha_1 = 0.03$ et $\alpha_2 = 0.02$ (8 points)
1.3 Unilatéral à droite de la forme $[c\ ;\ +\infty[$. (5 points)

2. Sachant que $s_A^2 = 2560000$et $s_B^2 = 25000000$, σ_A et σ_B étant inconnus, construire à un niveau de 0.95, un intervalle de confiance symétrique pour
2.1 $\mu_A - \mu_B$, (6 points)
2.2 σ_B / σ_A. (6 points)

3. Une enquête réalisée sur 500 pneus de type A et 450 de type B a permis de vérifier que 100 pneus de type A et 80 de type B ont présenté un certain défaut avant la fin d'une période de garantie donnée. Construire, à un niveau de 0.90 , un intervalle de confiance symétrique pour $p_A - p_B$ où p_A et p_B représentent respectivement la proportion de pneus de type A et de type B présentant un certain défaut avant la fin d'une période de garantie donnée. (10 points)

CENTRE DE TECHNIQUES DE PLANIFICATION ET D'ÉCONOMIE APPLIQUÉE (CTPEA)

Cours : Stat. mathématique/Professeurs : J-B ANTÉNORD et S. ST FLEUR/Examen final/2019-2020

Documentation autorisée : Tables **2e Année** **Durée : 2 heures 45**

Exercice 1 (30 points @ 5 points / question : vrai ouf faux (1 point) ; justification (4 points))

1. On considère une variable aléatoire $X \sim B(n, p)$. L'estimateur du maximum de vraisemblance de p noté $\hat{p}$ est donné par $\hat{p} = \frac{X}{n}$. Vrai ou faux ? Justifiez.
2. L'estimateur du maximum de vraisemblance de p noté $\hat{p}$ donné par $\hat{p} = \frac{X}{n}$ $\left(X \sim B(n, p)\right)$ est efficace au sens de la borne FDCR. Vrai ou faux ? Justifiez.
3. Un estimateur du maximum de vraisemblance est toujours efficace au sens de la borne FDCR.
4. Soit le *r-uplet* de variables aléatoires discrètes $(X_1, X_2, ..., X_r) \sim \mathcal{M}\ (n\ ;\ p_1, p_2, ..., p_r)$ tel que $X_j \sim B(n, p_j)$ où $E(X_j) = np_j$ et $V(X_j) = np_j(1 - p_j)$. Vrai ou faux et justifiez : la variable aléatoire

$$K = \sum_{j=1}^{r} \left\{ \frac{\left(X_j - np_j\right)^2}{np_j} \right\} \xrightarrow{n \longrightarrow \infty} \chi_r^2 \ .$$

5. Soit la fonction de répartition empirique $F_n : \mathbb{R}^n \longrightarrow [0,1]$ tel que $\forall x \in \mathbb{R}^n, F_n(x) = \hat{P}(X \leq x) \equiv \frac{\#(X_j \leq x)}{n}$. La variable aléatoire $F_n(x) \sim \mathcal{B}\big(n, F(x)\big)$, $F(x)$ étant la fonction de répartition théorique de la variable aléatoire X. Vrai ou faux ? Justifiez.
6. On définit la variance empirique corrigée par $s_n^2 = \frac{1}{n-1}\sum_{1=1}^{n}(X_i - \overline{X})^2$. Pour tout n-uplet de variables aléatoires telles que $X_i \sim iid(\mu;\sigma^2)$, on a : $Var(s_n^2) = \frac{2\sigma^4}{n-1}$, $\sigma^4 = (\sigma^2)^2 \equiv \mu_2^2 = [E(X_i - \mu))^2]^2$. Vrai ou faux ? Justifiez.

<u>Exercice 2</u> (35 points)

Soit une loi de probabilité de densité (par rapport à la mesure de Lebesgue sur R).

$$f(\theta, x) = \frac{2}{\sqrt{\pi\theta}} exp^{\left(\frac{-x^2}{\theta}\right)}\; I(x > 0), \forall\, x \in \text{R}$$

où $\theta > 0$ est un paramètre inconnu. On dispose d'un n-échantillon $X_1, X_2, \ldots X_n$ de cette loi. On admet que pour tout $\theta > 0$, $f()$ est bien une densité sur $\mathbb{R}$ par rapport à la mesure de Lebesgue.

On vous demande de répondre aux questions suivantes :

1. Expliciter l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ. Montrer qu'il est sans biais.
2. Calculer la variance de l'estimateur $\hat{\theta}$ (on rappelle que $E(Z^4) = 3$ pour $Z \sim N(0,1)$).
3. Calculer l'information de Fisher associée à ce modèle. Comparer la avec la variance de $\hat{\theta}$. Conclusion. Mettre ce résultat en perspective avec la Borne de Cramer-Rao.
4. Déterminer la loi limite quand $n \to \infty$ de $\sqrt{n}\big(\hat{\theta} - \theta\big)$.
5. Soit $\alpha \in]0,1[$. Proposer un test de niveau asymptotique α de l'hypothèse $H_0 : \theta < 3$ contre l'alternative $H_1 : \theta > 3$.

<u>Exercice 3 (20 points)</u>

Une entreprise d'assurance s'interroge sur la liaison entre l'intérêt pour un produit d'assurance vie et l'âge des individus. Elle dispose d'un fichier contenant un grand nombre de clients potentiels classés suivant leur âge et suivant leur intérêt pour l'assurance proposée. Elle note que parmi les 100 personnes de moins de 35 ans, 48 sont intéressées, parmi les 206 personnes d'au moins 35 ans et moins de 70 ans, 128 déclarent être intéressées. Enfin, parmi les 80 personnes de 70 ans et plus, 20 ne souhaitent pas acheter le produit.

1. Construire un tableau de contingence associé aux deux variables «âge» et «intérêt pour l'assurance proposée».
2. Construire le tableau des effectifs théoriques sous l'hypothèse d'indépendance des deux variables. En déduire que l'on peut utiliser une loi du chi-deux pour effectuer un test d'indépendance des deux variables.
3. Au seuil de 10%, l'intérêt des personnes pour l'assurance dépend-il alors de l'âge des personnes interrogées ?

Exercice 4 (15 points)

On compte les arrivées de voitures à un poste de péage sur l'autoroute, pendant la durée de temps unitaire d'une minute. Le tableau suivant résume les 200 arrivées observées.

Nombre de voitures	0	1	2	3	4	5	6	7	8	9	10	11
Fréquence	1	15	30	46	38	30	16	13	5	3	2	1

On vous demande de tester l'adéquation de la loi empirique observée à une loi de Poisson de paramètre λ, avec un risque de première espèce $\alpha = 0.05$.

CENTRE DE TECHNIQUES DE PLANIFICATION ET D'ECONOMIE APPLIQUEE (CTPEA)

Cours: Stat Math, Semestre 2 / Prof. : J-B ANTENORD / Examen Semestriel / 28 septembre 2016

Deuxième Année

===

Document autorisé : tables ***Définissez clairement vos événements et variables aléatoires*** ***Durée : 2h55***

Exercices théoriques (40 points)

1. Démontrer le théorème suivant:
Pour tout $\theta_0 \in \Theta$, on définit $R(\theta_0)$ la région d'acceptation d'un test de niveau α sous $H_0 : \theta = \theta_0$. Pour chaque $x \in D_X$, on définit dans l'espace des paramètres, un ensemble $C(x) = \{\theta_0 : x \in R(\theta_0)\}$. Alors l'ensemble aléatoire *C(X)* est un intervalle de confiance. Inversement, soit *C(X)* un intervalle de confiance de niveau $1-\alpha$. Si pour tout $\theta_0 \in \Theta$, on définit $R(\theta_0) = \{x : \theta_0 \in C(x)\}$, alors $R(\theta_0)$ est la région d'acceptation d'un test de niveau α sous $H_0 : \theta = \theta_0$. (15 points)

2. On considère un échantillon aléatoire indépendant $X_1, X_2, \ldots\ldots, X_n$ tel que $E(X_i) = \mu$ et $Var(X_i) = \sigma^2, \forall i$. Construire un estimateur BLUE pour μ. (10 points)

3. L'équation d'ANOVA est donnée par:

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{\overline{X}})^2 = n\sum_{i=1}^{k}(\overline{X}_i - \overline{\overline{X}})^2 + \sum_{j=1}^{n}\sum_{i=1}^{k}(X_{ij} - \overline{X}_i)^2$$

$$SCT \quad = \quad SCE \quad + \quad SCR$$

Soit $T = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}$ et $T_i = \sum_{j=1}^{n} X_{ij}$. Démontrer que : a) $SCT = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \frac{T^2}{kn}$;

b) $SCE = \frac{1}{n}\sum_{i=1}^{k} T_i^2 - \frac{T^2}{kn}$

Problèmes (60 points)

Problème 1 . Une banque souhaite mettre en place sur Internet une procédure systématique d'attribution de crédit à la consommation. A partir d'un .chier de clients existants, ayant déjà souscrit un prêt, elle dé.nit plusieurs classes d'individus correspondants de profils de risque différents. Parmi celles-ci, elle distingue la catégorie des artisans, âgés de plus de 45 ans, célibataires, et travaillant dans le secteur du bâtiment. La question posée par la direction des risques de la banque est de savoir s.il convient ou non d'accorder systématiquement un prêt à la consommation à tous les individus de cette catégorie.

Pour répondre à cette question, vous disposez d'un n-échantillon aléatoire d'anciens clients appartenant à cette catégorie. On note Y_i pour i = 1,2,...,n la variable dichotomique telle que:

$Y_i = 1$ si le client a connu un incident de paiement et $Y_i = 0$ sinon.

On suppose que les Y_i sont *i.i.d.* tels que $P(Y_i = 1) = p \in]0;1[$. La banque souhaite mettre en place une règle de décision quant à l'attribution ou non du crédit à la consommation fondée sur cette probabilité de défaut.

2.1 Proposez à la banque (i) un test d'hypothèse simple contre hypothèse simple sur p, puis (ii) un test d'hypothèse simple contre hypothèse multiple unilatérale répondant à son attente. Vous expliquerez la signification économique de l'hypothèse nulle et justifierez votre choix. (5 points)

2.2 On considère le test:

$$\begin{cases} \mathrm{H}_0 : p = p_0 \\ \mathrm{H}_1 : p = p_1 \end{cases} \text{ avec } p_1 > p_0$$

Expliquez la signification économique du risque de première espèce et de la puissance de ce test au responsable du département des risques. (5 points)

2.3 Soit $M_n = (1/n)\sum_{i=1}^{n} Y_i$ la fréquence empirique des défauts associés au n-échantillon $\{Y_1, Y_2, ..., Y_n\}$. On admet que la région critique du test UPP de niveau α de l'hypothèse $\begin{cases} \mathrm{H}_0 : p = p_0 \\ \mathrm{H}_1 : p = p_1 > p_0 \end{cases}$ est de la forme $W = \{Y_1, ..., Y_n / M_n > C\}$ où C désigne une constante déterminée par le risque de première espèce. En utilisant le théorème central limite, démontrez que dans le cas d'un échantillon de taille asymptotique le seuil critique C est défini par : $C = p_0 + F_Z^{-1}(1-\alpha)\sqrt{\frac{p_0(1-p_0)}{n}}$ où F_Z désigne la fonction de répartition de la loi normale centrée réduite. (5 points)

2.4 La banque vous demande un premier diagnostic quant à la décision qu'elle doit prendre d'accorder ou non de façon systématique un prêt à la consommation à tous les individus de la catégorie étudiée. La banque souhaite limiter son risque de première espèce à 10%. Pour cela, elle vous fournit les éléments suivants :

$$p_0 = 0.2,\ p_1 = 0.3,\ \text{n=100},\ \sum_{i=1}^{n} Y_i = 23.$$

Quelle décision conseillez vous à la banque concernant l'octroi de crédits à la consommation aux clients appartenant de cette catégorie ? (5 points)

2.5 partir des éléments de la question 4, déterminez la puissance du test proposé. Expliquez la signification économique de ce résultat. Pensez vous que ce résultat satisfera les responsables de la banque ? (5 points)

2.6 Le responsable du département des risques souhaite que la règle de décision permette d'atteindre un niveau donné de puissance, noté γ pour un niveau donné de risque de première espèce α. Montrez lui que pour cela la nouvelle règle de décision doit être fondée sur un échantillon de taille n vérifiant :

$$n = \left[\frac{F_Z^{-1}(1-\gamma)\sqrt{p_1(1-p_1)} - F_Z^{-1}(1-\alpha)\sqrt{p_0(1-p_0)}}{p_0 - p_1} \right]^2$$

Quelle taille d'échantillon convient-il d'adopter pour garantir 90% de puissance pour un niveau de risque de première espèce de 10% ? (5 points)

2.7 On considère à présent un test d'hypothèse simple contre une hypothèse multiple unilatérale :

$$\begin{cases} \mathrm{H}_0 : p = p_0 \\ \mathrm{H}_1 : p > p_0 \end{cases}$$

Pour un échantillon de taille n = 100, quelle est la région critique du test UPP de niveau α = 10% ? (5 points)

2.8 Déterminez la fonction puissance de ce test en fonction des valeurs de p et des paramètres α, p0 et n. Pour p0 = 0.2, n = 100 et α = 10%; si la vraie valeur de la probabilité de défaut de paiement est égale à 28%, quelle est la puissance de votre test ? (5 points)

2.9 On vous demande enfin de tester la valeur de la probabilité de défaut comme suit :

$$\begin{cases} \mathrm{H}_0 : p = 0.2 \\ \mathrm{H}_1 : p \neq 0.2 \end{cases}$$

Pour α = 5%; n=100 et $\sum_{i=1}^{n} Y_i = 23$, que concluez vous? (5 points)

Problème 2 (15 points)

On souhaite tester si une variable X suit une loi binomiale négative de paramètre n et p, notée Binomiale Négative (*n, p*). On rappelle que la loi binomiale négative est une distribution de probabilité discrète. Elle décrit la situation suivante: une expérience consiste en une série de tirages indépendants, donnant un succès avec probabilité p et un échec avec une probabilité complémentaire. Cette expérience se poursuit jusqu'à l'obtention d'un nombre donné r de succès. La variable aléatoire représentant le nombre d'échecs (avant l'obtention du nombre donné r de succès) suit alors une loi binomiale négative. Ainsi, si X suit une Binomiale Négative(*n, p*), n > 0 et 0 < p < 1, alors la probabilité que X soit égale à *x* est égale à : $P(X = x) \equiv p_X(x) = C_{x+n-1}^{x} p^n (1-p)^x$, $\forall x \in \{0,1,2,.....\}$.

On considère un échantillon de 100 variables aléatoires indépendantes $\{X_1, X_2, ..., X_n\}$de même loi que X. Pour ces 100 variables, on observe les effectifs empiriques suivants :

X_i	0	1	2	3	4	5	6	7
Nombre d'observations	16	18	19	16	11	9	6	5

Testez au seuil de 5%, l'hypothèse selon laquelle la variable X suit une loi binomiale négative de paramètres p = 0.5 et n = 3.