Université de Cergy-Pontoise 2011– 2012
L3 Mathématiques, Statistiques
E. Löcherbach

Examen. Avril 2012

Les documents et calculatrices ne sont pas autorisés. Durée de l'épreuve : 3 heures.

Questions de cours

1. Vrai ou faux ? Le risque quadratique d'un estimateur est la variance.
   Donnez une justification pour votre réponse.

2. Pour estimer la variance inconnue d'un $n-$échantillon $X_1, \ldots X_n$ d'une loi inconnue, les deux estimateurs suivants sont utilisés :
$$\hat{s}_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i,$$
   et
$$\Sigma_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - m)^2, \quad m = E(X_1).$$
   Quand est-ce qu'on utilise le premier estimateur, quand le deuxième?

3. Donner la définition d'un modèle statistique dominé et de la vraisemblance d'un modèle. Comment est défini l'estimateur du maximum de vraisemblance? Pouvez-vous donner une motivation pour la définition de cet estimateur?

4. Donner la définition d'un modèle exponentiel et de la statistique exhaustive associée. Donner un exemple d'un tel modèle. Que peut-on dire sur l'estimation d'un paramètre inconnu dans un modèle exponentiel? (Donner l'énoncé du théorème de Lehmann-Scheffé.) Quelle est l'hypothèse essentielle sur la fonction $\alpha(\vartheta)$?

Exercice 1
On considère une variable aléatoire réelle X dont la densité par rapport à la mesure de Lebesgue est donnée par
$$f(x) = ce^{-\vartheta|x|} \text{ avec } \vartheta > 0 \text{ et } c > 0.$$

1. Calculer c pour que f soit bien une densité de probabilité.

2. Expliquer - sans faire des calculs - pourquoi $E(X) = 0$.

3. Calculer $E(|X|)$.

4. Calculer la variance de X.

5. On observe un $n-$échantillon $X_1, \ldots, X_n$ de la loi de X. Ecrire la vraisemblance de $X_1, \ldots, X_n$.

6. Donner la statistique exhaustive du modèle.

7. Est-ce que le modèle est régulier? Calculer l'information de Fisher.

8. Calculer l'estimateur $\hat{\vartheta}_n$ du maximum de vraisemblance de ϑ.

9. Justifier pourquoi $\hat{\vartheta}_n$ converge vers $1/E(|X_1|)$ presque sûrement lorsque n tend vers l'infini.

10. Commentaire?

11. Donner un intervalle de confiance asymptotique pour $1/\vartheta$. Rappel : Si U suit une loi $\mathcal{N}(0,1)$, alors
$$P(-1,96 \leq U \leq 1,96) = 0,95.$$

Exercice 2
Sur 9 individus issus d'une population donnée, le dosage du calcium sanguin a donné les résultats $x_1, \dots, x_9$ avec $x_1 + \dots + x_9 = 21,6$ mmol/L et $x_1^2 + \dots + x_9^2 = 51,9048$ mmol^2/L^2. On suppose que la loi de la probabilité de la calcémie dans cette population est une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec les paramètres μ et σ^2 inconnus.

1) Calculer - à la main - la moyenne et la variance empirique de l'échantillon $x_1, \dots, x_9$.

2) Donner les estimations sans biais pour les paramètres μ et σ^2.

3) Trouver un intervalle de confiance pour μ au risque de 5%.
Rappel : si U_8 suit une loi de Student à 8 ddl, alors
$$P(-2,306 \leq U_8 \leq 2,306) = 0,95,$$
et si U_9 suit une loi de Student à 9 ddl, alors
$$P(-2,2622 \leq U_9 \leq 2,2622) = 0,95.$$
Laquelle des deux valeurs faut-il prendre? Donner l'intervalle de confiance sans faire les calculs - les formules sont suffisantes.

Exercice 3
Soient $X_1, \dots, X_n$ i.i.d. avec $X_i \sim U(0,\theta)$ où $U(0,\theta)$ est la loi uniforme sur l'intervalle $(0,\theta)$. On souhaite utiliser le théorème de Neyman-Pearson pour construire un test UPP pour $H_0 := \theta \leq 1$ contre $H_1 := \theta > 1$, au niveau $\alpha \in (0,1)$.

1. Ecrire la vraisemblance du modèle.

2. Fixer $\theta_0 = 1 < \theta_1$ et construire le test de Neyman-Pearson pour $\theta_0 = 1$ contre θ_1.

3. Reécrire ce test en termes de $T := \max(X_1, \dots, X_n)$.

4. Determiner la constante qui intervient dans la définition du test pour obtenir un niveau α.

5. Montrer que le test ainsi construit ne dépend pas de θ_1.

6. Montrer que pour $\theta < 1$, $E_\theta(\phi) \leq E_1(\phi)$. Conclure.

L'exercice suivant est supplémentaire et peut être ignoré si vous n'avez pas le temps de le traiter.

Exercice 4
(Exercice théorique).
Soient $X_1, \dots, X_n$ des variables aléatoires i.i.d., de moyenne m et de variance finie σ^2. On suppose que $\bar{X}_n$ et la variance empirique $\hat{s}_n = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ sont indépendants. Le but de cet exercice est de montrer que la loi de X_i est alors nécessairement la loi gaussienne $\mathcal{N}(m, \sigma^2)$.

On note ψ la fonction caractéristique de X_1 et on suppose que $m = 0$.

1. Calculer $E(\hat{s}_n)$ en fonction de σ^2. Montrer que pour tout réel t,
$$E[(n-1)\hat{s}_n e^{it\bar{X}_n}] = (n-1)\psi(t)^n \sigma^2.$$
2. Développer le terme $E[(n-1)\hat{s}_n e^{it\bar{X}_n}]$ et en déduire que ψ est solution de l'équation différentielle
$$\frac{\psi''}{\psi} - \left(\frac{\psi'}{\psi}\right)^2 = -\sigma^2, \psi(0) = 1, \psi'(0) = 0.$$
3. Soit $\varphi(t) = \psi'(t)/\psi(t)$. Montrer que $\varphi(t) = -\sigma^2 t$ (il convient de dériver φ). En déduire une forme explicite pour $\psi(t)$.
4. Rappelons que $U \sim \mathcal{N}(0, \sigma^2)$ si et seulement si $E(e^{itU}) = e^{-\sigma^2 t^2/2}$. (Ce résultat du cours de probabilités est admis.) Conclure.