HEC Paris – ESCP Europe

Mathématiques II

Mathématiques approfondies

Sujet adapté – 2008

Table des matières

1 Préliminaire **2**

Thèmes abordés : intégrales généralisées ; variables à densité. 2

2 Loi du χ^2 centré **3**

Thèmes abordés : variables à densité ; estimations. 3

3 Loi du χ^2 décentré **4**

Thèmes abordés : variables à densité. 4

4 Nombre aléatoire de degrés de liberté **4**

Thèmes abordés : probabilités discrètes ; variables à densité. 4

5 Estimateur de James–Stein **5**

Thèmes abordés : estimations ; variables à densité. 5

Les parties du sujet contenant des éléments n'étant plus au programme depuis 2022 ont été adaptées pour convenir aux nouveaux programmes de la filière Mathématiques approfondies en vigueur depuis 2023.

Toutes les variables aléatoires qui interviennent dans ce problème sont supposées définies sur un même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. Sous réserve d'existence, on note $\mathbb{E}(X)$ et $\mathbb{V}(X)$ respectivement l'espérance et la variance d'une variable aléatoire réelle X. Pour toute variable aléatoire réelle X admettant une densité f_X sur $\mathbb{R}$, on note D_X l'ensemble des réels s pour lesquels la variable aléatoire e^{sX} admet une espérance, et Φ_X la fonction définie sur D_X par $\Phi_X(s) = \mathbb{E}(e^{sX})$.

On admet les résultats suivants :

— si deux variables aléatoires X et Y sont telles que Φ_X et Φ_Y coïncident sur un intervalle ouvert non vide, alors X et Y ont la même loi ;

— si n est un entier naturel non nul et $X_1, \dots, X_n$ des variables aléatoires réelles mutuellement indépendantes, alors, pour tout entier $p \in [1, n-1]$ et pour toutes fonctions réelles continues g_1 et g_2, les variables aléatoires $g_1(X_1, \dots, X_p)$ et $g_2(X_{p+1}, \dots, X_n)$ sont indépendantes ;

— si X et Y sont des variables aléatoires indépendantes admettant une espérance, alors XY admet une espérance et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

La fonction exponentielle est également notée exp. On rappelle que

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x = \sqrt{2\pi}.$$

Dans tout le problème, U désigne une variable aléatoire qui suit la loi normale centrée réduite.

1 Préliminaire

Thèmes abordés : intégrales généralisées ; variables à densité.

On rappelle que, pour tout $s \in D_X$, on a

$$\Phi_X(s) = \int_{-\infty}^{+\infty} \exp(sx) f_X(x) \, \mathrm{d}x.$$

1. Soit a un réel non nul et b un réel quelconque.

   a) Montrer que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} \exp(-ax^2) \, \mathrm{d}x$ est convergente si et seulement si $a > 0$, et vaut alors $\sqrt{\pi/a}$.

   b) En déduire que l'intégrale $\displaystyle\int_{-\infty}^{+\infty} \exp(-ax^2 + bx) \, \mathrm{d}x$ est convergente si et seulement si $a > 0$, puis calculer sa valeur dans ce cas.

2. a) Déterminer D_U et, pour tout $s \in D_U$, calculer $\Phi_U(s)$.

   b) On pose $Z = U^2$. Établir que $D_Z =]-\infty, \frac{1}{2}[$, puis montrer, à l'aide du théorème de transfert, que, pour tout réel $s \in D_Z$, on a $\Phi_Z(s) = (1-2s)^{-1/2}$.

3. Soit X une variable aléatoire réelle à densité, et soient μ et β deux réels quelconques.

   a) Montrer qu'un réel s appartient à $D_{\mu X+\beta}$ si et seulement si μs appartient à D_X, et que, dans ce cas,

   $$\Phi_{\mu X+\beta}(s) = \exp(\beta s)\Phi_X(\mu s).$$

b) On suppose que X suit une loi gamma de paramètre ν, où ν est un réel strictement positif. Montrer que $D_X =]-\infty, 1[$ et que, pour tout $s \in D_X$, $\Phi_X(s) = (1-s)^{-\nu}$. De même, déterminer D_{2X} et, pour tout $s \in D_{2X}$, calculer $\Phi_{2X}(s)$.

2 Loi du χ^2 centré

Thèmes abordés : variables à densité ; estimations.

Pour $a > 0$ et $\theta > 0$, on rappelle que la loi gamma à deux paramètres $\Gamma(a, \theta)$ possède la densité

$$x \longmapsto \frac{x^{a-1}}{\theta^a \Gamma(a)} \exp\left(-\frac{x}{\theta}\right) \mathbf{1}_{]0,+\infty[}(x).$$

Si $G \sim \gamma(a)$ désigne une loi gamma à un paramètre de forme a, alors $\theta G \sim \Gamma(a, \theta)$.

Soit r un entier supérieur ou égal à 1. On considère une variable aléatoire X_r suivant la loi $\Gamma\left(\frac{r}{2}, 2\right)$, c'est-à-dire que X_r possède une densité f_{X_r} donnée par

$$f_{X_r}(x) = \begin{cases} \dfrac{x^{\frac{r}{2}-1}}{2^{r/2}\Gamma\left(\frac{r}{2}\right)} \exp\left(-\dfrac{x}{2}\right) & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

On dit que X_r suit une loi du χ^2 centré à r degrés de liberté, et on note $X_r \sim \chi^2(r)$.

1. Étudier les variations de f_{X_r} et tracer sa courbe représentative dans un repère orthogonal du plan.
2. a) Montrer que la variable aléatoire $\frac{X_r}{2}$ suit une loi gamma de paramètre $\frac{r}{2}$. En déduire $\mathbb{E}(X_r)$ et $\mathbb{V}(X_r)$.
   b) Déterminer D_{X_r} et, pour tout $s \in D_{X_r}$, calculer $\Phi_{X_r}(s)$.
3. Soit n un entier de $\mathbb{N}^*$. On considère n variables aléatoires indépendantes $U_1, U_2, \ldots, U_n$ de même loi que U. Pour tout $i \in [1, n]$, on pose $Z_i = U_i^2$.
   a) Vérifier que Z_1 et U^2 sont de même loi.
   b) On pose $W_n = \sum_{i=1}^{n} Z_i$. Quelle est la loi de probabilité de W_n ?
   c) Déterminer D_{W_n} et, pour tout $s \in D_{W_n}$, exprimer $\Phi_{W_n}(s)$ en fonction de s et de n. Établir une relation entre $\Phi_{\sum_{i=1}^n Z_i}(s)$ et $\Phi_{Z_1}(s), \ldots, \Phi_{Z_n}(s)$.
4. Soit T une variable aléatoire qui suit la loi normale centrée de variance σ^2 inconnue, σ étant un réel strictement positif. Pour n entier supérieur ou égal à 2, on dispose d'un n-échantillon indépendant et identiquement distribué $T_1, T_2, \ldots, T_n$ de la loi de T. On considère la variable aléatoire
$$S_n = \frac{1}{n} \sum_{i=1}^{n} T_i^2.$$
   a) Montrer que S_n est un estimateur sans biais et convergent du paramètre σ^2.
   b) Soit α un réel vérifiant $0 < \alpha < 1$, et soit k_α le réel strictement positif tel que $\mathbb{P}([W_n \geq k_\alpha]) = 1 - \alpha$. Montrer que l'intervalle $\left]0, \dfrac{nS_n}{k_\alpha}\right]$ est un intervalle de confiance de σ^2 au risque α.

3 Loi du χ^2 décentré

Thèmes abordés : variables à densité.

On considère une suite $(M_j)_{j\geq 1}$ de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, mutuellement indépendantes, telles que, pour tout $j \in \mathbb{N}^*$, M_j suive la loi normale d'espérance m_j ($m_j \in \mathbb{R}$) et de variance égale à 1. Pour $n \in \mathbb{N}^*$, on pose

$$Y_n = \sum_{j=1}^{n} M_j^2 \qquad \text{et} \qquad \lambda_n = \sum_{j=1}^{n} m_j^2.$$

On dit que Y_n suit une loi du χ^2 décentré à n degrés de liberté, de paramètre de décentrage λ_n, et on note $Y_n \sim \chi^2(n, \lambda_n)$.

1. Dans cette question uniquement, on suppose que l'entier n est égal à 1.
   a) Montrer les deux égalités suivantes : $\mathbb{E}(U^3) = 0$ et $\mathbb{E}(U^4) = 3$.
   b) En déduire, en fonction de λ_1, les valeurs respectives de $\mathbb{E}(Y_1)$ et de $\mathbb{V}(Y_1)$.
   c) Montrer que $D_{Y_1} =]-\infty, \frac{1}{2}[$ et établir, pour tout réel $s \in D_{Y_1}$, la formule
   $$\Phi_{Y_1}(s) = (1-2s)^{-1/2} \exp\left(\frac{\lambda_1 s}{1-2s}\right).$$
2. Soit n un entier de $\mathbb{N}^*$.
   a) Calculer $\mathbb{E}(Y_n)$ et $\mathbb{V}(Y_n)$ en fonction de n et de λ_n.
   b) On admet que $D_{Y_n} =]-\infty, \frac{1}{2}[$. Pour tout $s \in D_{Y_n}$, exprimer $\Phi_{Y_n}(s)$ en fonction de s, n et λ_n.

4 Nombre aléatoire de degrés de liberté

Thèmes abordés : probabilités discrètes ; variables à densité.

Sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$, on considère une variable aléatoire X à valeurs dans $\mathbb{R}$ admettant une espérance $\mathbb{E}(X)$, et une variable aléatoire K à valeurs dans $\mathbb{N}$. On note N_K l'ensemble des entiers naturels k vérifiant $\mathbb{P}([K=k]) > 0$, et on suppose que, pour tout entier $k \in N_K$, la variable aléatoire X admet une espérance pour la probabilité conditionnelle $\mathbb{P}_{[K=k]}$, notée $\mathbb{E}(X \mid K=k)$. On admet alors l'égalité suivante :

$$(*) \qquad \mathbb{E}(X) = \sum_{k \in N_K} \mathbb{E}(X \mid K=k)\mathbb{P}([K=k]).$$

Soit g l'application définie sur $\mathbb{N}$ par

$$g(k) = \begin{cases} \mathbb{E}(X \mid K=k) & \text{si } k \in N_K, \\ 0 & \text{sinon.} \end{cases}$$

1. **Vérification de la formule $(*)$ sur un exemple.**
   Soit $(J_i)_{i\geq 1}$ une suite de variables aléatoires définies sur $(\Omega, \mathcal{A}, \mathbb{P})$, indépendantes et de même loi uniforme sur l'intervalle $[0,1]$. Pour tout $k \in \mathbb{N}^*$, on pose $X_k = \sup_{1\leq i\leq k} J_i$; autrement dit, pour tout $\omega \in \Omega$, $X_k(\omega) = \max_{1\leq i\leq k} J_i(\omega)$. Soit K une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$, qui suit la loi uniforme discrète sur $[1,n]$ ($n \in \mathbb{N}^*$). On suppose que K est indépendante des variables aléatoires de la suite $(J_i)_{i\geq 1}$. Pour tout $\omega \in \Omega$, on pose $X(\omega) = \max_{1\leq i\leq K(\omega)} J_i(\omega)$, et on admet que X est une variable aléatoire définie sur $(\Omega, \mathcal{A}, \mathbb{P})$.

a) Établir, pour tout entier $k \in [1, n]$ et pour tout réel x, la relation

$$\mathbb{P}_{[K=k]}([X \leq x]) = \mathbb{P}([X_k \leq x]).$$

b) Déterminer la fonction de répartition F_X de la variable aléatoire X.

c) En déduire que X est une variable aléatoire à densité qui admet une espérance $\mathbb{E}(X)$ que l'on exprimera en fonction de $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n} \frac{k}{k+1}$.

d) Vérifier l'égalité $(*)$: $\mathbb{E}(X) = \mathbb{E}(g(K))$.

2. Soit $(U_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes qui suivent la loi normale centrée réduite. Soit K une variable aléatoire indépendante des variables aléatoires de la suite $(U_i)_{i\geq 1}$, qui suit la loi de Poisson de paramètre $\frac{\lambda}{2}$, où λ est strictement positif. Pour $n \in \mathbb{N}^*$, on pose

$$H_n = U_1^2 + U_2^2 + \cdots + U_{n+2K}^2.$$

On admet que H_n est une variable aléatoire à densité à valeurs positives, et que $D_{H_n} =]-\infty, \frac{1}{2}[$. Soit $s \in]-\infty, \frac{1}{2}[$.

a) Montrer que, pour tout $k \in \mathbb{N}$, la loi conditionnelle de H_n, sachant $[K = k]$, est la loi de la variable aléatoire W_{n+2k} définie dans la question I.3.b.

b) En posant $X = e^{sH_n}$, déterminer, pour tout entier $k \in \mathbb{N}$, l'expression de $g(k)$ en fonction de k.

c) Établir la formule suivante :

$$\mathbb{E}(g(K)) = (1-2s)^{-n/2} \exp\left(\frac{\lambda s}{1-2s}\right).$$

d) En utilisant l'égalité $(*)$, admise au début de cette partie, avec $X = e^{sH_n}$, déterminer la loi de H_n.

e) À l'aide de la question III.2.a, montrer que, pour tout entier n supérieur ou égal à 3,

$$\mathbb{E}\left(\frac{1}{H_n}\right) = \mathbb{E}\left(\frac{1}{n-2+2K}\right).$$

5 Estimateur de James–Stein

Thèmes abordés : estimations ; variables à densité.

Dans cette partie, le risque quadratique scalaire d'un estimateur θ^* du paramètre vectoriel $\theta = (\theta_1, \ldots, \theta_p)$ est défini par

$$R(\theta^*, \theta) = \mathbb{E}\left(\sum_{j=1}^{p} (\theta_j^* - \theta_j)^2\right).$$

Soit p un entier supérieur ou égal à 3. On suppose qu'un modèle aléatoire défini sur $(\Omega, \mathcal{A}, \mathbb{P})$ comporte p paramètres réels inconnus $\theta_1, \ldots, \theta_p$, non tous nuls. Un échantillon d'observations statistiques permet d'exhiber des estimateurs $\widehat{\theta}_1, \ldots, \widehat{\theta}_p$, sans biais des paramètres $\theta_1, \ldots, \theta_p$ respectivement. On suppose que les variables aléatoires $\widehat{\theta}_1, \ldots, \widehat{\theta}_p$ sont indépendantes et suivent une loi normale de variance égale à 1.

On pose $\theta = (\theta_1, \ldots, \theta_p)$, $\widehat{\theta} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_p)$,

$$B_p = \sum_{j=1}^{p} \widehat{\theta}_j^2 \qquad \text{et} \qquad b_p = \sum_{j=1}^{p} \theta_j^2.$$

Le vecteur aléatoire $\widehat{\theta}$ est un estimateur sans biais du paramètre vectoriel θ.

Dans cette partie, on cherche un estimateur θ^* de θ, représenté par un vecteur aléatoire $(\theta_1^*, \ldots, \theta_p^*)$, dont le risque $R(\theta^*, \theta)$ est strictement inférieur à $R(\widehat{\theta}, \theta)$.

1. Justifier que la variable aléatoire B_p suit la loi $\chi^2(p, b_p)$, et qu'elle constitue un estimateur biaisé de b_p.

2. On pose $\theta^* = \left(1 - \dfrac{c}{B_p}\right)\widehat{\theta}$, où c est un paramètre réel strictement positif. Soit K une variable aléatoire qui suit la loi de Poisson de paramètre $\frac{b_p}{2}$.

   a) En admettant que l'on a

   $$\mathbb{E}\left(\frac{1}{B_p}\sum_{j=1}^{p}\theta_j\widehat{\theta}_j\right) = \mathbb{E}\left(\frac{2K}{p-2+2K}\right),$$

   établir l'égalité suivante :

   $$R(\theta^*, \theta) - R(\widehat{\theta}, \theta) = (c^2 - 2c(p-2))\mathbb{E}\left(\frac{1}{p-2+2K}\right).$$

   b) Montrer que l'inégalité $R(\theta^*, \theta) < R(\widehat{\theta}, \theta)$ est vérifiée si et seulement si $0 < c < 2(p-2)$. Déterminer en fonction de p la valeur de c pour laquelle $R(\theta^*, \theta) - R(\widehat{\theta}, \theta)$ est minimale. Comment s'écrit alors l'estimateur θ^* ?