



PRÉPASIMPLE

prepasimple.fr

HEC — 2008

Sujet de mathématiques approfondies

Mathématiques approfondies — Prépa ECG

Annales de concours · Sujet

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

On admet les résultats suivants :

- si n est un entier naturel non nul, et $X_1, X_2, \dots, X_n$ des variables aléatoires réelles quelconques, mutuellement indépendantes, alors, pour tout entier p de $\llbracket 1, n-1 \rrbracket$ et pour toutes fonctions réelles continues φ_1 et φ_2 , les variables aléatoires $\varphi_1(X_1, \dots, X_p)$ et $\varphi_2(X_{p+1}, \dots, X_n)$ sont indépendantes ;
- si X et Y sont des variables aléatoires indépendantes admettant une espérance, alors XY admet une espérance, et $E(XY) = E(X)E(Y)$.

La fonction exponentielle est également notée $\exp$. On rappelle que :

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

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

Preliminaire

On rappelle que, pour tout s de $\mathcal{D}_X$, on a : $\Phi_X(s) = \int_{-\infty}^{+\infty} \exp(sx) f_X(x) dx$.

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

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

b) En déduire que l'intégrale $\int_{-\infty}^{+\infty} \exp(-ax^2 + bx) dx$ est convergente si et seulement

si $a > 0$, puis montrer que dans ces conditions, on a :

$$\int_{-\infty}^{+\infty} \exp(-ax^2 + bx) dx = \sqrt{\frac{\pi}{a}} \exp\left(\frac{b^2}{4a}\right).$$

2. a) Déterminer $\mathcal{D}_U$; pour tout s de $\mathcal{D}_U$, calculer $\Phi_U(s)$.

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

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

a) Montrer qu'un réel s appartient à $\mathcal{D}_{\mu X + \beta}$ si et seulement si μs appartient à $\mathcal{D}_X$, et que dans ce cas, on a : $\Phi_{\mu X + \beta}(s) = \exp(\beta s) \Phi_X(\mu s)$.

b) On suppose que X suit une loi γ de paramètre ν , où ν est un réel strictement positif.

Montrer que : $\mathcal{D}_X =]-\infty, 1[$; pour tout s de $\mathcal{D}_X$, établir la formule : $\Phi_X(s) = (1 - s)^{-\nu}$.

De même, déterminer $\mathcal{D}_{2X}$; pour tout s de $\mathcal{D}_{2X}$, calculer $\Phi_{2X}(s)$.

Partie I. Loi du χ^2 centré

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

$$f_{X_r}(x) = \frac{1}{2^2 \Gamma(\frac{r}{2})} x^{\frac{r}{2}-1} \exp\left(-\frac{x}{2}\right) \text{ si } x > 0, \text{ et } f_{X_r}(x) = 0 \text{ si } x \leq 0.$$

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

1. Étudier les variations de f_{X_4} 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 γ de paramètre $\frac{r}{2}$. En déduire $E(X_r)$ et $V(X_r)$.

b) Déterminer $\mathcal{D}_{X_r}$; pour tout s de $\mathcal{D}_{X_r}$, calculer $\Phi_{X_r}(s)$.

3. Soit n un entier de $\mathbb{N}^*$. On considère n variables aléatoires indépendantes $U_1, U_2, \dots, U_n$ de même loi que U . Pour tout i de $\llbracket 1, n \rrbracket$, on pose : $Z_i = U_i^2$.

a) Vérifier que X_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 $\mathcal{D}_{W_n}$, et pour tout s de $\mathcal{D}_{W_n}$, exprimer $\Phi_{W_n}(s)$ en fonction de s et de n .
Établir une relation entre $\Phi_{W_n}(s)$ et $\Phi_{Z_1}(s), \Phi_{Z_2}(s), \dots, \Phi_{Z_n}(s)$.

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, identiquement distribué (i.i.d.), $T_1, T_2, \dots, T_n$ de la loi de T . On considère la variable aléatoire S_n définie par :

$$S_n = \frac{1}{n} \sum_{i=1}^n T_i^2.$$

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 :

$P([W_n \geq k_\alpha]) = 1 - \alpha$. Montrer que l'intervalle $\left] 0, \frac{nS_n}{k_\alpha} \right]$ est un intervalle de

confiance de σ^2 au risque α .

Partie II. Loi du χ^2 décentré

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

Pour n entier de $\mathbb{N}^*$, on pose :

$$Y_n = \sum_{j=1}^n M_j^2$$

et

$$\lambda_n = \sum_{j=1}^n m_j^2.$$

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 : $Y_n \hookrightarrow \chi^2(n, \lambda_n)$.

1. Dans cette question uniquement, on suppose que l'entier n est égal à 1.

a) Montrer les deux égalités suivantes : $E(U^3) = 0$ et $E(U^4) = 3$.

b) En déduire, en fonction de λ_1 , les valeurs respectives de $E(Y_1)$ et de $V(Y_1)$.

c) Montrer que : $\mathcal{D}_{Y_1} = \left] -\infty, \frac{1}{2} \right[$ et établir, pour tout réel s de $\mathcal{D}_{Y_1}$, la formule suivante :

$$\Phi_{Y_1}(s) = (1 - 2s)^{-1/2} \times \exp\left(\frac{s\lambda_1}{1 - 2s}\right)$$

2. Soit n un entier de $\mathbb{N}^*$.

a) Calculer $E(Y_n)$ et $V(Y_n)$ en fonction de n et λ_n .

b) On admet que l'on a : $\mathcal{D}_{Y_n} = \left] -\infty, \frac{1}{2} \right]$. Pour tout s de $\mathcal{D}_{Y_n}$, exprimer $\Phi_{Y_n}(s)$ en fonction de s , n et λ_n .

Partie III. Nombre aléatoire de degrés de liberté

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

On admet alors l'égalité suivante :

$$(\star) \quad E(X) = \sum_{k \in N_K} E(X/K = k) P([K = k])$$

Soit g l'application définie sur $\mathbb{N}$ par : $g(k) = E(X/K = k)$ si $k \in N_K$, et $g(k) = 0$ sinon.

1. Vérification de la formule $(\star)$ sur un exemple.

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

Pour tout ω de Ω , 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 $(\Omega, \mathcal{A}, P)$.

a) Établir, pour tout entier k de $\llbracket 1, n \rrbracket$ et pour tout réel x , la relation :

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

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 $E(X)$ que l'on exprimera en fonction de

$$\sum_{k=1}^n \frac{k}{k+1}.$$

d) Vérifier l'égalité $(\star) : E(X) = E(g(K))$.

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 $\frac{\lambda}{2}$ strictement positif.

Pour n entier de $\mathbb{N}^*$, on pose : $H_n = U_1^2 + U_2^2 + \dots + U_{n+2K}^2$. On admet que H_n est une variable aléatoire à densité à valeurs positives, et que $\mathcal{D}_{H_n} = \left] -\infty, \frac{1}{2} \right[$.

Soit s un réel de $\left] -\infty, \frac{1}{2} \right[$.

a) Montrer que pour tout k de $\mathbb{N}$, 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 de $\mathbb{N}$, l'expression de $g(k)$ en fonction de k .

c) Établir la formule suivante :

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

- d) On admet le résultat suivant : pour tout entier $n \geq 1$ et tout réel $\lambda > 0$, si une variable aléatoire H à densité vérifie $\left] -\infty, \frac{1}{2} \right[\subset \mathcal{D}_H$ et, pour tout réel $s < \frac{1}{2}$, $\Phi_H(s) = (1 - 2s)^{-n/2} \exp\left(\frac{\lambda s}{1 - 2s}\right)$, alors H suit la loi $\chi^2(n, \lambda)$, c'est-à-dire que H a même loi que Y_n pour tout choix des réels $m_1, \dots, m_n$ tels que $\lambda_n = \lambda$.

En utilisant l'égalité $(\star)$, 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, on a :

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

Partie IV. Estimateur de James-Stein

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

On pose : $\theta = (\theta_1, \dots, \theta_p)$, $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$,

$$B_p = \sum_{j=1}^p \hat{\theta}_j^2$$

et

$$b_p = \sum_{j=1}^p \theta_j^2.$$

On dit que le vecteur aléatoire $\hat{\theta}$ est un estimateur sans biais du paramètre vectoriel θ , et

$E(\hat{\theta})$ est alors le vecteur θ .

On définit le risque quadratique scalaire d'un estimateur θ^* de θ , noté $R(\theta^*, \theta)$, par :

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

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

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 : $\theta^* = \left(1 - \frac{c}{B_p}\right) \times \hat{\theta}$, 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 :

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

établir l'égalité suivante :

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

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