



# BANQUE COMMUNE D'ÉPREUVES

CONCOURS D'ADMISSION DE 2011

Conceptions : H.E.C. – E.S.C.P. / EUROPE

Code épreuve :

283

OPTION SCIENTIFIQUE

## MATHEMATIQUES II

Lundi 9 mai 2011, de 8 h. à 12 h.

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction, la clarté et la précision des raisonnements entreront pour une part importante dans l'appréciation des copies.

Les candidats sont invités à encadrer dans la mesure du possible les résultats de leurs calculs.

Ils ne doivent faire usage d'aucun document : l'utilisation de toute calculatrice et de tout matériel électronique est interdite. Seule l'utilisation d'une règle graduée est autorisée.

Si au cours de l'épreuve, un candidat repère ce qui lui semble être une erreur d'énoncé, il la signalera sur sa copie et poursuivra sa composition en expliquant les raisons des initiatives qu'il sera amené à prendre

Toutes les variables aléatoires qui apparaissent dans ce problème sont supposées définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

On note  $F_Z$  la fonction de répartition d'une variable aléatoire  $Z$  et, si cette variable aléatoire admet une densité, on note  $f_Z$  une densité de  $Z$ .

Sous réserve d'existence, on note  $E(Z)$  et  $V(Z)$  respectivement, l'espérance et la variance d'une variable aléatoire réelle  $Z$ , et  $\text{Cov}(Z_1, Z_2)$  la covariance de deux variables aléatoires  $Z_1$  et  $Z_2$ .

La fonction exponentielle est notée  $\exp$  et la partie entière d'un réel  $x$  est notée  $[x]$ .

**On admet** les résultats suivants :

- la définition et les propriétés de la covariance et du coefficient de corrélation linéaire de deux variables aléatoires discrètes, s'appliquent au cas de variables aléatoires à densité ;
- si  $Z_1$  et  $Z_2$  sont deux variables aléatoires indépendantes admettant une espérance, alors la variable aléatoire  $Z_1 Z_2$  admet une espérance et  $E(Z_1 Z_2) = E(Z_1)E(Z_2)$ .

**Dans tout le problème**, on considère une variable aléatoire  $X$  de fonction de répartition  $F_X$  et admettant une densité  $f_X$ .

Les solutions éventuelles de l'équation  $F_X(x) = \frac{1}{2}$  s'appellent *les médianes théoriques de  $X$* .

Pour  $n$  entier de  $\mathbb{N}^*$ , on considère un  $n$ -échantillon  $(X_1, X_2, \dots, X_n)$  i.i.d. (indépendant, identiquement distribué)

de la loi de  $X$  et on définit la variable aléatoire :  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ , qui est *la moyenne empirique de l'échantillon*  $(X_1, X_2, \dots, X_n)$ .

On admet l'existence de variables aléatoires à densité  $Y_1, Y_2, \dots, Y_n$  telles que, pour tout  $\omega$  de  $\Omega$ , les réels  $Y_1(\omega), Y_2(\omega), \dots, Y_n(\omega)$  constituent un réarrangement par ordre croissant des réels  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ , de telle sorte que, pour tout  $\omega$  de  $\Omega$  :  $Y_1(\omega) \leq Y_2(\omega) \leq \dots \leq Y_n(\omega)$ .

En particulier,  $Y_1 = \inf(X_1, X_2, \dots, X_n)$  et  $Y_n = \sup(X_1, X_2, \dots, X_n)$ . Plus généralement, pour tout  $k$  de  $[1, n]$ , il existe une fonction  $\psi_k$  définie et continue sur  $\mathbb{R}^n$  à valeurs réelles, telle que  $Y_k = \psi_k(X_1, X_2, \dots, X_n)$ .

Si  $n$  est un entier impair ( $n = 2\ell + 1$ , avec  $\ell \in \mathbb{N}$ ), alors la variable aléatoire  $Y_{\ell+1}$  est appelée la médiane empirique de l'échantillon  $(X_1, X_2, \dots, X_n)$ .

La partie II du problème est indépendante de la partie I.

### Partie I. Quelques propriétés des statistiques d'ordre

Pour tout réel  $x$  et tout entier  $k$  de  $[1, n]$ , on note  $J_k(x)$  la variable aléatoire de Bernoulli définie par :

$$J_k(x) = \begin{cases} 1 & \text{si } [X_k \leq x] \text{ est réalisé} \\ 0 & \text{si } [X_k > x] \text{ est réalisé} \end{cases} ; \quad \text{on pose : } S_n(x) = \sum_{k=1}^n J_k(x)$$

1. a) Montrer que les fonctions  $f_{Y_1}$  et  $f_{Y_n}$  définies pour tout  $x$  réel par :  $f_{Y_1}(x) = n(1 - F_X(x))^{n-1}f_X(x)$  et  $f_{Y_n}(x) = n(F_X(x))^{n-1}f_X(x)$ , sont des densités de  $Y_1$  et  $Y_n$  respectivement.

b) Quelle est la loi de probabilité de la variable aléatoire  $S_n(x)$  ?

c) Justifier l'égalité entre événements suivante :  $[Y_k \leq x] = [S_n(x) \geq k]$ .

d) Établir la relation : pour tout  $x$  réel,  $F_{Y_k}(x) = \sum_{j=k}^n \binom{n}{j} (F_X(x))^j (1 - F_X(x))^{n-j}$ .

e) En déduire que pour tout  $k$  de  $[1, n]$ , la fonction  $f_{Y_k}$  définie pour tout  $x$  réel par :

$$f_{Y_k}(x) = k \binom{n}{k} (F_X(x))^{k-1} (1 - F_X(x))^{n-k} f_X(x)$$

est une densité de  $Y_k$ .

f) Montrer que si  $X$  admet un moment d'ordre  $r$  ( $r \in \mathbb{N}^*$ ), alors pour tout  $k$  de  $[1, n]$ ,  $Y_k$  admet un moment d'ordre  $r$ .

**Exemple.** Dans les questions 2 à 4, on suppose que la fonction de répartition  $F_X$  est donnée par :

$$F_X(x) = \begin{cases} 1 - \frac{1}{\sqrt{x}} & \text{si } x \geq 1 \\ 0 & \text{si } x < 1 \end{cases}$$

2. a) Tracer la courbe représentative de  $F_X$  dans le plan rapporté à un repère orthonormé.

Préciser la demi-tangente à droite au point d'abscisse  $x = 1$ .

Justifier que  $X$  est une variable aléatoire à densité et préciser une densité  $f_X$  de  $X$ .

b) Montrer que  $X$  n'admet aucun moment.

c) Établir l'unicité de la médiane théorique  $M$  de  $X$ . Calculer  $M$ .

d) Expliciter, pour tout  $k$  de  $[1, n]$  et pour tout  $x$  réel, l'expression  $f_{Y_k}(x)$  d'une densité de  $Y_k$ .

En déduire un équivalent de  $f_{Y_k}(x)$  lorsque  $x$  tend vers  $+\infty$ .

3. On suppose dans cette question que  $n \geq 3$ .

a) Montrer que pour tout  $k$  de  $[1, n-2]$ ,  $Y_k$  admet une espérance.

b) En justifiant l'emploi du changement de variable  $t = \frac{1}{\sqrt{x}}$ , établir pour tout  $k$  de  $[1, n-2]$ ,

$$\text{la formule : } E(Y_k) = k \binom{n}{k} \int_0^1 t^{n-k-2} (1-t)^{k-1} dt.$$

c) Pour tout couple  $(r, s)$  de  $(\mathbb{N}^*)^2$ , on pose :  $I_{r,s} = \int_0^1 t^{r-1} (1-t)^{s-1} dt$ .

$$\text{Montrer que pour tout couple } (r, s) \text{ de } (\mathbb{N}^*)^2, \text{ on a : } I_{r,s} = \frac{(r-1)!(s-1)!}{(r+s-1)!}.$$

d) En déduire l'expression de  $E(Y_k)$  pour tout  $k$  de  $[1, n-2]$ .

e) On suppose que  $n$  est impair et supérieur ou égal à 5, et on pose  $n = 2\ell + 1$ . Justifier la définition de

la médiane empirique  $Y_{\ell+1}$  d'un échantillon, et établir l'égalité :  $E(Y_{\ell+1}) = 4 + \frac{6}{\ell-1}$ . Commenter.

4. On pose pour tout  $n$  de  $\mathbb{N}^*$  :  $Z_n = \frac{1}{n^2} \sup(X_1, X_2, \dots, X_n) = \frac{Y_n}{n^2}$ .

a) Calculer pour tout  $x$  réel,  $F_{Z_n}(x)$ .

b) On définit la fonction  $\varphi_Z$  par :  $\varphi_Z(x) = \begin{cases} \exp\left(-\frac{1}{\sqrt{x}}\right) & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$ .

Montrer que  $\varphi_Z$  est la fonction de répartition d'une variable aléatoire  $Z$  à densité.

c) Montrer que la suite de variables aléatoires  $(Z_n)_{n \in \mathbb{N}^*}$  converge en loi vers  $Z$ .

## Partie II. Existence et unicité d'un estimateur optimal

Dans cette partie,  $X$  suit la loi normale d'espérance  $\theta$  et de variance égale à 1. On suppose que le paramètre réel  $\theta$  est inconnu.

On note  $\Phi$  la fonction de répartition de la loi normale centrée réduite.

On rappelle que pour  $n$  entier de  $\mathbb{N}^*$ ,  $(X_1, X_2, \dots, X_n)$  est un  $n$ -échantillon i.i.d. de la loi de  $X$ .

5. Quelle est la loi de  $\bar{X}_n$ ? Montrer que  $\bar{X}_n$  est un estimateur sans biais et convergent du paramètre  $\theta$ .

6. Soit  $\alpha$  un réel tel que  $0 < \alpha < 1$ . On appelle *marge d'erreur* associée à un intervalle de confiance de  $\theta$  au risque  $\alpha$ , le réel positif noté  $\mu(\alpha)$ , égal à la demi-longueur de cet intervalle.

a) Justifier l'existence de la fonction réciproque  $\Phi^{-1}$  de la fonction  $\Phi$ .

b) Déterminer un intervalle de confiance du paramètre  $\theta$  au risque  $\alpha$  dont le milieu est  $\bar{X}_n$ .

Vérifier que  $\mu(\alpha) = -\frac{\Phi^{-1}(\alpha/2)}{\sqrt{n}}$ .

c) On considère un risque  $\beta$  ( $\beta \neq \alpha$ ) tel que  $\mu(\beta) = b\mu(\alpha)$ , avec  $0 < b < 1$ . Exprimer  $\beta$  en fonction de  $\alpha$ .

Comparer  $\alpha$  et  $\beta$ . Commenter.

7. On note  $\mathcal{E}_\theta$  l'ensemble des statistiques  $U_n = g_n(X_1, X_2, \dots, X_n)$ , où  $g_n$  est une fonction de  $\mathbb{R}^n$  dans  $\mathbb{R}$ , qui sont des estimateurs sans biais de  $\theta$  et qui admettent une variance.

Sous réserve d'existence, on dit qu'un élément  $Z_n$  de  $\mathcal{E}_\theta$  est un estimateur optimal dans  $\mathcal{E}_\theta$ , si pour tout  $U_n$  de  $\mathcal{E}_\theta$ , on a :  $V(Z_n) \leq V(U_n)$ .

On admet que pour tout  $U_n$  de  $\mathcal{E}_\theta$ , on a :  $\text{Cov}(\bar{X}_n, U_n - \bar{X}_n) = 0$ .

a) Montrer que  $\mathcal{E}_\theta$  n'est pas vide.

b) Montrer que  $\bar{X}_n$  est optimal dans  $\mathcal{E}_\theta$ .

c) Soit  $Z_n$  un estimateur optimal dans  $\mathcal{E}_\theta$ . On pose pour tout  $U_n$  de  $\mathcal{E}_\theta$  et pour tout  $\lambda$  réel :

$$A_n(\lambda) = Z_n + \lambda(U_n - Z_n).$$

Montrer que  $A_n(\lambda)$  est un élément de  $\mathcal{E}_\theta$ . Calculer  $V(A_n(\lambda))$ . En déduire que  $\text{Cov}(Z_n, U_n - Z_n) = 0$ .

d) On suppose l'existence de deux estimateurs optimaux  $\bar{X}_n$  et  $Z_n$  dans  $\mathcal{E}_\theta$ .

Montrer que  $Z_n = \bar{X}_n$  presque sûrement. Conclure.

8. a) Justifier l'existence et l'unicité de la médiane théorique  $M$  de  $X$ , et exprimer  $M$  en fonction de  $\theta$ .

b) Calculer  $f_X(M)$ . Montrer que pour tout réel  $x$ , on a :  $F_X(2M - x) = 1 - F_X(x)$ .

En déduire une relation entre  $f_X(2M - x)$  et  $f_X(x)$ .

c) Établir pour tout  $k$  de  $[1, n]$ , la relation :  $E(Y_k - M) = E(M - Y_{n-k+1})$ .

d) En supposant que  $n = 2\ell + 1$  ( $\ell \in \mathbb{N}$ ), calculer  $E(Y_{\ell+1})$ , puis justifier que  $V(Y_{\ell+1}) \geq \frac{1}{n}$ . Commenter.

## Partie III. Résultats asymptotiques

Le contexte de cette partie est identique à celui de la partie II.

Dans cette partie, on note  $T$  une variable aléatoire qui suit la loi normale centrée réduite.

Si  $U$  est une variable aléatoire et  $s$  un réel tels que la variable aléatoire  $\exp(sU)$  admette une espérance, on pose :  $L_U(s) = \exp(sU)$ .

9. Soit  $J$  une variable aléatoire qui suit la loi de Bernoulli de paramètre  $p$  ( $0 < p < 1$ ).

- Calculer pour tout  $s$  réel,  $L_J(s)$ .
- Établir pour tout  $s$  réel, l'existence de  $L_T(s)$ .
- Calculer pour tout  $s$  réel,  $L_T(s)$ . En déduire que pour tout couple  $(\theta, \sigma)$  de  $\mathbb{R} \times \mathbb{R}^{**}$  et pour tout  $s$  réel, on a :  $L_{\sigma T + \theta}(s) = \exp\left(\sigma^2 \frac{s^2}{2} + \theta s\right)$ .

Dans les questions 10 et 11,  $x$  est un réel fixé.

10. On pose pour tout  $n$  de  $\mathbb{N}^*$  :  $y_n = M + \frac{x}{\sqrt{n}}$ ,  $q_n = F_X(y_n)$  et  $k(n) = \lfloor n/2 \rfloor + 1$ .

- Montrer que l'on a :  $k(n) = \frac{n}{2} + o(\sqrt{n})$  lorsque  $n$  tend vers  $+\infty$ .
- En appliquant la formule de Taylor-Young à la fonction  $F_X$  au voisinage de  $M$ , justifier la relation :

$$q_n = \frac{1}{2} + \frac{x}{\sqrt{2\pi n}} + o\left(\frac{1}{\sqrt{n}}\right)$$

- Soit  $(u_n)_{n \in \mathbb{N}^*}$  la suite définie par : pour tout  $n$  de  $\mathbb{N}^*$ ,  $u_n = \frac{1}{\sqrt{n}}(k(n) - nq_n)$ .

Montrer que la suite  $(u_n)_{n \in \mathbb{N}^*}$  est convergente et déterminer sa limite  $u$ .

11. On pose pour tout  $n$  de  $\mathbb{N}^*$  :  $W_n = \frac{1}{\sqrt{n}}(S_n(y_n) - nq_n)$ , où  $S_n(y_n)$  a été définie dans le préambule de la partie I.

- Établir pour tout réel  $s$ , la relation :

$$L_{W_n}(s) = \exp(-s\sqrt{n}q_n) \left(1 + q_n \exp\left(\frac{s}{\sqrt{n}}\right) - q_n\right)^n$$

- En utilisant un développement limité à l'ordre 2, montrer que l'on a :  $\ln(L_{W_n}(s)) = \frac{s^2}{8} + o(1)$  lorsque  $n$  tend vers  $+\infty$ . En déduire l'égalité :  $\lim_{n \rightarrow +\infty} L_{W_n}(s) = L_{\frac{T}{2}}(s)$ .

On admet alors que la suite de variables aléatoires  $(W_n)_{n \in \mathbb{N}^*}$  converge en loi vers la variable aléatoire  $\frac{T}{2}$ .

12. On suppose que  $x = 0$ . Quels sont les arguments qui permettent d'obtenir directement le résultat final de la question 11.b) ?

13. a) Établir pour tout  $n$  de  $\mathbb{N}^*$  et pour tout  $x$  réel, les égalités d'événements suivantes :

$$[\sqrt{n}(Y_{k(n)} - M) \leq x] = [S_n(y_n) \geq k(n)] = [W_n \geq u_n]$$

- Montrer l'égalité :  $\lim_{n \rightarrow +\infty} P([W_n \geq u_n]) = P\left(\left[\sqrt{\frac{\pi}{2}}T \leq x\right]\right)$ .

- En déduire que la suite de variables aléatoires  $\left(\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M)\right)_{n \in \mathbb{N}^*}$  converge en loi et préciser sa limite.

14. On suppose dans cette question que  $n$  est impair et on pose  $n = 2\ell + 1$  ( $\ell \in \mathbb{N}$ ).

On note  $\rho_n$  le coefficient de corrélation linéaire de  $Y_{k(n)}$  et  $\bar{X}_n$ .

- Que vaut  $k(n)$  ?
- Préciser la valeur de  $\lim_{n \rightarrow +\infty} E(Y_{k(n)})$ .

- On admet sans démonstration que la suite réelle de terme général  $E\left(\left(\sqrt{\frac{2n}{\pi}}(Y_{k(n)} - M)\right)^2\right)$

converge vers  $E(T^2)$ . En déduire un équivalent de  $V(Y_{k(n)})$  lorsque  $n$  tend vers  $+\infty$ .

- À l'aide des questions 5,7,8 et 14.c), déterminer la limite de  $\rho_n$  lorsque  $n$  tend vers  $+\infty$ .