



BANQUE COMMUNE D'ÉPREUVES

CONCOURS D'ADMISSION DE 2009

Conceptions : H.E.C. – E.S.C.P. – E.A.P

283

OPTION SCIENTIFIQUE

CCIP_M2_S

MATHEMATIQUES II

Mardi 5 mai 2009, 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

Dans tout le problème, N désigne un entier supérieur ou égal à 1.

On note $E(X)$ et $V(X)$ respectivement, l'espérance et la variance lorsqu'elles existent, de toute variable aléatoire réelle X définie sur un espace probabilisé.

Soit $(U_n)_{n \geq 1}$ une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, mutuellement indépendantes et de même loi uniforme discrète sur $[1, N]$.

On pose, pour tout n de $\mathbb{N}^*$: $T_n = \sup(U_1, U_2, \dots, U_n)$ et $Z_n = \inf(U_1, U_2, \dots, U_n)$. On admet que T_n et Z_n sont des variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$. Ainsi, pour tout ω de Ω , on a :

$$T_n(\omega) = \max(U_1(\omega), U_2(\omega), \dots, U_n(\omega)) \text{ et } Z_n(\omega) = \min(U_1(\omega), U_2(\omega), \dots, U_n(\omega))$$

On rappelle que si C désigne un élément de $\mathcal{A}$, on note 1_C la variable aléatoire indicatrice de l'événement C , définie sur $(\Omega, \mathcal{A}, P)$ par :

$$1_C(\omega) = \begin{cases} 1 & \text{si } \omega \in C \\ 0 & \text{si } \omega \notin C \end{cases}$$

$$\text{On pose, pour tout } n \text{ de } \mathbb{N}^* : d_n(N) = \begin{cases} \sum_{k=1}^{N-1} \left(\frac{k}{N}\right)^n & \text{si } N \geq 2 \\ 0 & \text{si } N = 1 \end{cases}$$

Préliminaire

1. Soit Y une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$, à valeurs dans $[1, N]$. Établir les deux relations suivantes :

$$E(Y) = \sum_{k=0}^{N-1} P([Y > k]) \quad \text{et} \quad E(Y^2) = \sum_{k=0}^{N-1} (2k+1)P([Y > k])$$

Partie I. Inf et Sup

2. Rappeler, sans démonstration, les valeurs respectives de $E(U_1)$ et de $V(U_1)$.
3. a) Calculer, pour tout k de $[1, N]$, $P([T_n \leq k])$.
b) En déduire la loi de probabilité de T_n .
4. a) Montrer que la suite $(d_n(N))_{n \geq 1}$ est convergente et calculer sa limite.
b) Exprimer $E(T_n)$ en fonction de N et $d_n(N)$. En déduire la valeur de $\lim_{n \rightarrow +\infty} E(T_n)$.
c) Établir la formule suivante : $V(T_n) = (2N - 1)d_n(N) - 2Nd_{n+1}(N) - d_n^2(N)$.
En déduire la valeur de $\lim_{n \rightarrow +\infty} V(T_n)$.
d) Montrer que si $N \geq 2$, on a : $\lim_{n \rightarrow +\infty} \frac{d_{n+1}(N)}{d_n(N)} = 1 - \frac{1}{N}$; en déduire que, lorsque n tend vers $+\infty$, on a : $V(T_n) \sim d_n(N)$.
5. Déterminer la loi de Z_n . Calculer $E(Z_n)$ et $V(Z_n)$.
6. On rappelle que la fonction Pascal `random(N)` permet de simuler une variable aléatoire suivant la loi uniforme sur $[0, N - 1]$. Écrire une fonction Pascal d'en-tête `simulmax(n : integer) : integer` qui simule la variable aléatoire T_n .

Partie II. Couple (Inf, Sup)

7. On pose, pour tout n de $\mathbb{N}^*$ et pour tout couple (k, ℓ) de $\mathbb{N}^2$: $\phi_n(k, \ell) = P([T_n \leq k] \cap [Z_n \leq \ell])$.
a) Montrer, pour tout (k, ℓ) de $[1, N]^2$, la relation suivante :
$$\phi_n(k, \ell) = \begin{cases} \left(\frac{k}{N}\right)^n & \text{si } k \leq \ell \\ \left(\frac{k}{N}\right)^n - \left(\frac{k-\ell}{N}\right)^n & \text{si } k > \ell \end{cases}$$
- b) Établir, pour tout (k, ℓ) de $[1, N]^2$, la formule suivante :
$$P([T_n = k] \cap [Z_n = \ell]) = \phi_n(k, \ell) + \phi_n(k-1, \ell-1) - \phi_n(k-1, \ell) - \phi_n(k, \ell-1)$$
- c) En déduire, en distinguant les trois cas $k < \ell$, $k = \ell$ et $k > \ell$, l'expression de $P([T_n = k] \cap [Z_n = \ell])$ en fonction de k et ℓ .
8. On donne, pour tout couple (m, n) de $(\mathbb{N}^*)^2$, les deux relations suivantes :
i) $\sum_{j=1}^m [(j+1)^n - 2j^n + (j-1)^n] = (m+1)^n - m^n - 1$;
ii) $\sum_{j=1}^m j[(j+1)^n - 2j^n + (j-1)^n] = m(m+1)^n - (m+1)m^n$.
a) En déduire, pour tout n de $\mathbb{N}^*$, la formule suivante : $E(T_n Z_n) = N(1 + d_{n+1}(N))$.
b) On note, pour tout n de $\mathbb{N}^*$, ρ_n le coefficient de corrélation linéaire entre T_n et Z_n .
Calculer $\lim_{n \rightarrow +\infty} \rho_n$ lorsque $N \geq 2$.
9. a) Pour tout n de $\mathbb{N}^*$ et pour tout couple (k, ℓ) de $[1, N]^2$, calculer la probabilité conditionnelle $P_{[T_n=k]}([Z_n = \ell])$.
b) En déduire, pour tout n de $\mathbb{N}^*$ et pour tout k de $[1, N]$, l'expression de l'espérance conditionnelle $E(Z_n/[T_n = k])$ de Z_n sachant $[T_n = k]$.

Partie III. Pr vision

Pour n entier de $\mathbb{N}^*$, on dispose d'un $(n+1)$ - chantillon ind pendant identiquement distribu  (i.i.d.) $(U_1, U_2, \dots, U_{n+1})$ de la loi uniforme sur $[1, N]$.

On pose : $T_n = \sup(U_1, U_2, \dots, U_n)$ et $T_{n+1} = \sup(U_1, U_2, \dots, U_{n+1}) = \sup(T_n, U_{n+1})$.

Pour tout $t = (t_1, t_2, \dots, t_N)$ de $\mathbb{R}^N$, on pose : $W_t(T_n) = \sum_{k=1}^N t_k \times \mathbf{1}_{[T_n=k]}$.

Dans cette partie, on se propose de d terminer la valeur de t pour laquelle les deux conditions suivantes sont v rifi es :

- i) $E(W_t(T_n)) = E(T_{n+1})$;
- ii) $E[(T_{n+1} - W_t(T_n))^2]$ est minimale.

10. Montrer, pour tout k de $[1, N]$, la relation : $P([W_t(T_n) = t_k]) = P([T_n = k])$.

11.  tablir, pour tout k de $[1, N]$, la formule suivante :

$$E(T_{n+1} \times \mathbf{1}_{[T_n=k]}) = E(T_{n+1}/[T_n = k]) \times P([T_n = k])$$

12. a) Calculer, pour tout couple (k, j) de $[1, N]^2$, $P([T_n = k] \cap [T_{n+1} = j])$.

b) En d duire, pour tout couple (k, j) de $[1, N]^2$, la probabilit  conditionnelle $P_{[T_n=k]}([T_{n+1} = j])$.

c) D terminer, pour tout k de $[1, N]$, l'expression de l'esp rance conditionnelle $E(T_{n+1}/[T_n = k])$ de T_{n+1} sachant $[T_n = k]$.

d) En appliquant la formule de l'esp rance totale, d duire de la question pr c dente la relation suivante :

$$E(T_{n+1}) = \frac{N+1}{2} + \frac{1}{2N} (E(T_n^2) - E(T_n))$$

13.  tablir l' galit  suivante : $(W_t(T_n))^2 = \sum_{k=1}^N t_k^2 \times \mathbf{1}_{[T_n=k]}$.

14. Soit g la fonction d finie sur $\mathbb{R}^N$   valeurs r elles par :

$$g(t_1, t_2, \dots, t_N) = E[(T_{n+1} - W_t(T_n))^2]$$

a)   l'aide des r sultats des questions 11, 12 et 13, expliciter g en fonction des variables $t_1, t_2, \dots, t_N$.

b) Montrer que g admet un minimum global sur $\mathbb{R}^N$ atteint en un point $\theta = (\theta_1, \theta_2, \dots, \theta_N)$ que l'on d terminera en fonction de $E(T_{n+1}/[T_n = 1]), E(T_{n+1}/[T_n = 2]), \dots, E(T_{n+1}/[T_n = N])$.

15.  tablir les deux relations suivantes :

$$E(W_\theta(T_n)) = E(T_{n+1}) \quad \text{et} \quad V(W_\theta(T_n)) \leq V(T_{n+1})$$

16. a)  tablir, pour tout i de $\mathbb{N}^*$, l' galit  suivante : $\sum_{k=1}^N k^i \times \mathbf{1}_{[T_n=k]} = (T_n)^i$.

b) En d duire la relation suivante : $W_\theta(T_n) = \frac{N+1}{2} + \frac{1}{2N} (T_n^2 - T_n)$.

Partie IV. Estimation

Soit U une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, de loi uniforme discrète sur $[1, N]$. On suppose que le paramètre N est inconnu.

Cette partie a pour objet la détermination d'un estimateur ponctuel de N , sans biais et de variance minimale.

Pour n entier supérieur ou égal à 1, soit $(U_1, U_2, \dots, U_n)$ un n -échantillon i.i.d. de la loi de U .

17. Soit ε un réel strictement positif. On pose :

$$A_n(\varepsilon) = [|T_n - N| \geq \varepsilon] \text{ et } B_n(\varepsilon) = [|T_n - E(T_n)| + |d_n(N)| \geq \varepsilon]$$

- Peut-on dire que $T_n + d_n(N)$ est un estimateur sans biais de N ?
- Montrer que la suite $(T_n)_{n \geq 1}$ est une suite d'estimateurs asymptotiquement sans biais du paramètre N .
- Montrer que $A_n(\varepsilon) \subset B_n(\varepsilon)$ et qu'il existe un entier naturel n_0 tel que, pour tout $n > n_0$, on a : $B_n(\varepsilon) \subset [|T_n - E(T_n)| \geq \varepsilon/2]$.
- En déduire que la suite d'estimateurs $(T_n)_{n \geq 1}$ est convergente.

18. a) Calculer, pour tout n -uplet $(u_1, u_2, \dots, u_n)$ de $[1, N]^n$, $P\left(\bigcap_{i=1}^n [U_i = u_i]\right)$.

b) En déduire que, pour tout k de $[1, N]$, la loi conditionnelle du vecteur aléatoire $(U_1, U_2, \dots, U_n)$ sachant $[T_n = k]$ est donnée par :

$$P_{[T_n=k]}\left(\bigcap_{i=1}^n [U_i = u_i]\right) = \begin{cases} \frac{1}{k^n - (k-1)^n} & \text{si pour tout } i \text{ de } [1, n], 1 \leq u_i \leq N \text{ et } \max_{1 \leq i \leq n} (u_i) = k \\ 0 & \text{sinon} \end{cases}$$

On remarquera que cette loi conditionnelle ne dépend pas du paramètre N .

19. On pose, pour n entier de $\mathbb{N}^*$: $S_n = T_n + Z_n - 1$ et, pour tout k de $[1, N]$: $\psi_n(k) = \frac{k^{n+1} - (k-1)^{n+1}}{k^n - (k-1)^n}$.

- Montrer que S_n est un estimateur sans biais de N .
- Établir, pour tout k de $[1, N]$, l'égalité : $\psi_n(k) = E(S_n/[T_n = k])$.
- En déduire que $\psi_n(T_n)$ est un estimateur sans biais de N .
- On pose, pour tout k de $[1, N]$: $\varphi_n(k) = E(S_n^2/[T_n = k])$.
Établir, pour tout k de $[1, N]$, l'inégalité : $\psi_n^2(k) \leq \varphi_n(k)$ (on pourra utiliser la fonction définie sur $\mathbb{R}$ par : $\lambda \mapsto E((S_n - \lambda)^2/[T_n = k])$). En déduire que $V(\psi_n(T_n)) \leq V(S_n)$.
- Calculer $V(S_n)$. En déduire que $\psi_n(T_n)$ est un estimateur convergent de N .

20. Soit, pour n entier de $\mathbb{N}^*$, un estimateur sans biais R_n du paramètre N .

On pose, pour tout k de $[1, N]$: $f_n(k) = E(R_n/[T_n = k])$.

- En utilisant une méthode analogue à celle de la question 19.d, montrer que : $V(f_n(T_n)) \leq V(R_n)$.
- Soit F une fonction réelle. Montrer que, pour n fixé dans $\mathbb{N}^*$, la condition « pour tout N de $\mathbb{N}^*$, $E(F(T_n)) = N$ » est vérifiée, si et seulement si, pour tout k de $[1, N]$, on a : $F(k) = \psi_n(k)$.
- En déduire que dans l'ensemble des estimateurs sans biais de N , l'estimateur $\psi_n(T_n)$ est optimal, dans le sens où $V(\psi_n(T_n))$ est minimale.

La partie IV constitue une démonstration du théorème de Lehmann-Scheffé dans le cas particulier d'une loi uniforme sur $[1, N]$, avec N inconnu.

MATHEMATIQUES 2 S
(épreuve n°283)

Epreuve conçue par CCIP

Voie scientifique

1. Le sujet

Le problème de mathématiques 2 de cette année avait pour objet principal l'étude du couple $(\inf, \sup)$ de variables aléatoires suivant la loi uniforme sur $\{1, 2, \dots, N\}$. Il se prolongeait ensuite par la recherche du meilleur prédicteur du $\sup$ d'un $(n+1)$ -échantillon et s'achevait par la mise en évidence d'un estimateur optimal du paramètre N (amélioré de Rao-Blackwell et démonstration du théorème de Lehmann-Scheffé dans le cas particulier d'une loi uniforme).

Le contenu du sujet permettait de couvrir une large partie du programme de probabilités discrètes et d'algorithmique : loi d'un couple, probabilité et espérance conditionnelles, formule de l'espérance totale, estimation, étude de suites, minimisation, fonctions de plusieurs variables, programmation Pascal....

2. Les résultats statistiques

Sur les 3256 candidats ayant composé dans cette épreuve, la note moyenne s'établit à 9,94 (comparable à celle du concours 2008) et l'écart-type est de 4,55. Il y a 14 candidats qui ont obtenu la note maximale de 20 et 9% de l'ensemble des candidats se voient attribuer une note supérieure à 16.

Le poids de chacune des parties du problème était de 5% pour le préliminaire, 22% pour la partie I, 22% pour la partie II, 24% pour la partie III et 27% pour la partie IV.

La partie III et la seconde moitié de la partie IV ont été peu abordées, l'essentiel du travail des candidats se limitant aux deux premières parties.

3. Commentaires détaillés et erreurs les plus fréquentes

Les remarques générales qui ressortent de l'examen des copies sont les suivantes :

- très souvent, les candidats négligent de justifier l'application d'un théorème du cours en ne vérifiant pas l'adéquation de ses hypothèses au cas de la question traitée ;
 - la justification précise de l'utilisation de tel ou tel concept probabiliste est fréquemment absente ;
 - la présentation des copies reste globalement satisfaisante, seules quelques copies étant encore illisibles (écriture, ratures, etc.).
1. Cette première question, très proche du cours, a été assez discriminante. Finalement, elle n'aura été correctement traitée que dans de la moitié des copies, ce qui a fortement surpris les correcteurs.
 2. Ces valeurs sont connues même si on relève quelques erreurs sur la variance.
 3. Cette question est en général bien résolue.
 4. La limite de la suite $(d_n(N))_n$ n'a pas été trouvée dans toutes les copies. Il y a eu confusion entre n et N . La détermination de la variance $V(T_n)$ et de sa limite révèle la capacité de certains candidats à conduire correctement un calcul à son terme (questions c et d) ; d'autres, sans reproduire leur calcul sur la copie, affirment le résultat !

5. Très peu de candidats connaissent le lien entre sup et inf. Son utilisation aurait permis un gain de temps précieux.
6. Question d'informatique abordée dans la moitié des copies. Confusion entre $\text{Random}(N) + 1$ et $\text{random}(N+1)$. Rappelons que l'on attribue un poids non négligeable dans le barème de notation à l'écriture correcte d'un programme, fût-il très court.
7. La question a) a été « devinée », puisque la réponse était donnée dans l'énoncé, mais rarement justifiée correctement. Une probabilité égale à N ne surprend pas ! La question b) a été littéralement « massacrée ».
8. Là encore, on mesure les faibles capacités de calcul de certains candidats. Trouver la limite d'un coefficient de corrélation linéaire égale à l'infini ne semble pas gêner...
Un peu de recul est parfois nécessaire...
9. Pour calculer une probabilité conditionnelle, il suffit souvent (et c'était le cas ici) de s'en tenir à la définition. De même pour l'espérance conditionnelle. Trop souvent, on trouve des pseudo- raisonnements « alambiqués » et faux.
10. Un oubli dans cette question (il faut supposer les t_i distincts) sans conséquence pour la suite du problème, a été repéré par une petite minorité de candidats qui s'est vue récompensée par des points de bonification.
11. Le maniement des fonctions indicatrices (présentes dans plusieurs problèmes de concours ces dernières années) est toujours aussi troublante pour les candidats.
12. Question qui a été mal traitée car l'énoncé ne fournissait pas le résultat de la question a).
13. à 16. Seule, une infime minorité de candidats a abordé ces questions.
17. Question correctement traitée, excepté la a). Une dizaine de candidats connaissent la définition d'un estimateur.
18. La question a) a été convenablement résolue, mais la b) rarement justifiée.
- 19 et 20. Questions non abordées.

4. Recommandations aux futurs candidats

Les développements mathématiques doivent faire l'objet de justifications rigoureuses s'appuyant sur des raisonnements et des théorèmes qui s'expriment dans un langage clair et soigné. Une réponse qui n'est pas correctement argumentée ne rapporte aucun point.

Les correcteurs sont évidemment à même de juger si une réponse est démontrée ou seulement *devinée* ; dans cette dernière éventualité, l'impression générale qui se dégage est négative et ils seront moins cléments pour des questions plus difficiles. De même, ils ne sont pas censés corriger des copies dans lesquelles il est nécessaire de déchiffrer des notations de cours écrites dans un style sténographique.