MATHEMATIQUES II (S) 2007 (épreuve n° 283)

Epreuve conçue par CCIP

Voie Scientifique

	NBRE CANDIDATS	MOYENNES	ECARTS-TYPE
RESULTATS GLOBAUX	2 914	9,93	5,16

VOIES PREPARATOIRES			
Scientifique	2 914	9,93	5,16

ECOLES UTILISATRICES			
HEC	2 006	11,63	4,77
ESSEC	2 085	11,54	4,77
ESCP-EAP	2 376	10,96	4,92
EM Lyon	2 821	10,03	5,15

MATHÉMATIQUES II – Option Scientifique

Le sujet

Le problème proposé cette année avait pour objet l'étude de l'approximation d'une somme finie de variables aléatoires $X_i, i \in [\![1, n]\!]$, lorsque X_i suit une loi de Bernoulli de paramètre p_i, par une variable aléatoire suivant la loi de Poisson de paramètre $\sum_{i=1}^{n} p_i$ (inégalité de Le Cam).

La méthode utilisée repose essentiellement sur des outils probabilistes (équation de Chen–Stein) (voir annexe).

On remarquera que le sujet de l'épreuve de mathématiques I concernait la même problématique, mais en utilisant des techniques liées aux semi-groupes d'opérateurs.

Résultats statistiques

La note moyenne obtenue par les 2914 candidats à cette épreuve est de 9,93, avec un écart-type élevé de 5,16. L'histogramme des résultats ainsi que les caractéristiques statistiques précédentes, révèlent une épreuve discriminante qui a très certainement permis de sélectionner les meilleurs candidats.

Notons que des candidats, certes en minorité, ont produit d'excellentes copies, démontrant une bonne compréhension de la thématique du problème, une grande habileté dans les calculs et faisant preuve de précision et de rigueur dans les raisonnements.

Les résultats par école sont :

- HEC (2006 candidats) – moyenne : 11,63, écart-type : 4,77
- ESCP-EAP (2376 candidats) – moyenne : 10,96, écart-type : 4,92

Erreurs les plus fréquentes

Partie I (29% de la note totale)

La partie I, d'orientation essentiellement calculatoire, consistait à appliquer des résultats démontrés dans la partie III.

1. Question bien traitée dans la grande majorité des copies. La justification de la non indépendance des variables X_i et X_j est trop souvent omise.

2. Certains pensent que W_n suit une loi binomiale... Le calcul de $V(W_n)$ n'est pas mené à son terme. Bien entendu, la vérification demandée en c) en pâtit.

3. Rarissimes sont les candidats qui traitent cette question, et aucun ne gère correctement la partie entière, alors qu'il s'agissait de réaliser un développement limité à l'ordre 2 d'une fonction du type $e^{u \ln v}$.
Dans la question c), on lit trop souvent des raisonnements du type : « W_n tend en loi vers T_n », oubliant ainsi le sens du paramètre n.

4. La présence du paramètre $e^{-\theta}$ donnait la possibilité aux candidats de corriger les erreurs de la question précédente. Très peu l'ont fait.
La notion d'intervalle de confiance est mal comprise puisque la plupart des candidats ont fait apparaître le paramètre inconnu dans les bornes de l'intervalle demandé.

Partie II (34% de la note totale)

La partie II mettait en place les outils nécessaires aux démonstrations des deux parties suivantes (équation de Chen-Stein caractérisant la loi de Poisson). C'est cette partie qui a été la mieux traitée par les candidats. La grande majorité des candidats abordent avec succès les questions 1, 2 et 3. La question 4, plus délicate, est moins bien traitée, avec des raisonnements faux. La question 5 est correctement résolue lorsqu'elle est abordée. La question 6.a) n'a été comprise que par une minorité de candidats.

Partie III (23% de la note totale)

Dans la partie III, on démontre l'inégalité de Le Cam dans le cas où les variables X_i sont indépendantes. Le principal outil utilisé est la formule de l'espérance totale.

La question 1.a) a souvent été discriminante, de même que la justification de l'indépendance dans la question b).
Les questions 2 et 3 n'ont été traitées correctement que par les candidats ayant réellement intégré l'utilisation et la manipulation correcte de la formule de l'espérance totale.
Malgré son caractère classique et sa relative facilité, la question 6 a donné lieu à des résultats très surprenants.

Partie IV (14% de la note totale)

Enfin, dans la partie IV, on démontre l'inégalité de Le Cam, dans le cas où les variables X_i ne sont pas nécessairement indépendantes, en utilisant l'espérance conditionnelle et la formule de l'espérance totale. Cette partie a été abordée par un nombre infime de candidats.

Recommandations aux futurs candidats

Pour ce qui concerne la forme, le jury demande aux candidats de lire attentivement le texte préliminaire qui précède toute épreuve de mathématiques, dans lequel il est notamment précisé que *la lisibilité et la qualité de la rédaction entrent pour une part non négligeable dans l'appréciation des copies.*
Il est également conseillé de bien numéroter ses questions et d'encadrer ses résultats. De plus, les raisonnements doivent être clairs et précis, les affirmations étant étayées par une argumentation solide. Par exemple, le recours trop fréquent à des phrases du type *il est clair que..., par une récurrence immédiate...* doit être évité au profit d'une justification correcte, fondée sur un apprentissage très sérieux et une connaissance approfondie du cours.

Enfin le jury recommande aux futurs candidats de prendre le temps de lire l'ensemble du sujet, non seulement pour s'en imprégner, mais aussi pour pointer les questions classiques ou faciles à résoudre, lesquelles ne se situent pas obligatoirement dans la première partie du problème.

Annexe.

En 1972, C. Stein a proposé une nouvelle méthode permettant de majorer la distance entre une variable aléatoire X et une variable aléatoire Z suivant la loi normale centrée réduite, à l'aide d'une *fonction de test* h. Il a observé que sous certaines conditions, il existe une fonction f telle que

$$(\star) \qquad E(h(X)-h(Z)) = E(f'(X) - Xf(X))$$

Par exemple, lorsque $h = 1_A$, alors

$$E(h(X)-h(Z)) = P(X\in A) - \frac{1}{\sqrt{2\pi}}\int_A e^{-x^2/2}dx = E(f'(X)-Xf(X))$$

Ainsi, lorsque la fonction f est *simple*, la partie droite de l'équation $(\star)$ est plus facilement manipulable que la partie gauche.

En 1975, L.H.Y Chen a démontré que la méthode précédente s'appliquait également lorsque Z suivait une loi de Poisson de paramètre λ, l'équation $(\star)$ devenant

$$E(h(X) - h(Z)) = E(\lambda f(X+1) - Xf(X))$$

avec f et h définies sur $\mathbb{N}$.

En fait, on peut démontrer que cette équation *caractérise* la loi de Poisson dans le sens suivant : une variable aléatoire Z suit la loi de Poisson de paramètre λ si et seulement si pour toute fonction $g : \mathbb{N} \to \mathbb{R}$ bornée, on a : $E(\lambda g(Z+1) - Zg(Z)) = 0$.

Cette proposition induit donc l'équation de Stein-Chen

$$h(x) - \sum_{k=0}^{\infty} h(k)\frac{\lambda^k}{k!} = \lambda f(x+1) - xf(x)$$

Lorsque $h(x) = 1_{\{x \in A\}}$, pour $A \subset \mathbb{N}$, on cherche une fonction f_A satisfaisant l'équation

$$1_{\{x \in A\}} - \sum_{k \in A} \frac{\lambda^k}{k!} = \lambda f_A(x+1) - xf_A(x)$$

C'est la fonction déterminée dans la partie II du problème.

Cette méthode permet ainsi de généraliser l'approximation d'une loi binomiale $\mathcal{B}(n, \lambda/n)$ par la loi de Poisson $\mathcal{P}(\lambda)$, au cas des variables de Bernoulli de paramètres différents, indépendantes, ou non.