1.2B – MATHEMATIQUES I – filière PC

Présentation du sujet

Soient $X_1, \ldots, X_n$ des variables de Bernoulli mutuellement indépendantes de paramètres respectifs $p_1, \ldots, p_n$, $\lambda = p_1 + \ldots + p_n$. Un problème classique est de comparer la loi de la somme $S = X_1 + \ldots + X_n$ à une loi de Poisson de paramètre λ. Un des premiers résultats généraux dans cette direction est l' *inégalité de Le Cam* (1960), qui assure que la distance en variation totale entre la loi de S et $P(\lambda)$ est majorée par

$$2\sum_{i=1}^{n} p_i^{\,2}.$$

Cette inégalité a reçu beaucoup de démonstrations. Celle qui se généralise le mieux à des situations de dépendance faible est fondée sur la méthode dite de Stein-Chen et fait l'objet du sujet. Le coeur de la partie analytique de la preuve est l'estimation de la constante de Lipschitz des solutions de l'équation fonctionnelle

$$\forall n \in \mathbb{N}, \qquad \lambda f(n+1) - nf(n) = h(n) - e^{-\lambda} \sum_{k=0}^{+\infty} h(k) \frac{\lambda^k}{k!},$$

où l'inconnue est f et h une fonction bornée de $\mathbb{N}$ dans $\mathbb{R}$.

Pour des renseignements complémentaires, on renvoie à l'article de de J.M. Steele Le Cam's Inequality and Poisson Approximations (American Mathematical Monthly, 101(1), pages 48-54, 1994).

Commentaires généraux

Le sujet était long, mais progressif. Après des questions préliminaires simples, on passait à une caractérisation de la distribution de Poisson de paramètre λ, qui donne l'idée de base de la méthode de Stein-Chen. Venaient ensuite la résolution de l'équation (1) et l'estimation de la constante de Lipschitz des solutions. C'est seulement dans la dernière partie qu'apparaissait la motivation probabiliste du problème.

Les connaissances utilisées étaient assez limitées (séries numériques, variables aléatoires discrètes). Le sujet demandait en revanche une certaine habileté en analyse, notamment une bonne maîtrise des inégalités.

Le sujet a atteint son but en permettant un bon étalonnement des notes. Les correcteurs ont cependant été surpris par le nombre relativement élevé de copies très faibles. Les candidats en question manifestent, soit par l'absence de réponse, soit par des réponses aberrantes dont nous citerons plus loin quelques exemples, une ignorance des éléments de l'analyse et un manque de rigueur intellectuelle très surprenants après deux années de classes préparatoires.

Conseils aux futurs candidats

Ce sujet valorise le travail en profondeur du cours et de la technique de calcul. C'est seulement par une pratique importante de gammes très simples que l'on s'approprie les gestes de base de l'analyse (majoration, estimations, preuves de convergence). Beaucoup de candidats ont des réflexes très faibles en ces matières.

Les premières questions du sujet sont faciles. Elles doivent être rédigées de façon très soignée afin de convaincre les correcteurs de l'honnêteté et de la solidité mathématique du candidat.

D'autre part, l'attitude consistant à dissimuler l'incompréhension d'une question par un discours imprécis ou un passage en force influence de manière très négative la perception qu'a le correcteur de l'ensemble de la copie et est lourdement pénalisée. On relève cette année une tendance accrue à faire des calculs dépourvus de signification, tant le rôle des objets est mal compris. Conseillons donc aux candidats de prendre le temps de comprendre le sens des questions !

La rédaction est systématiquement évaluée par les correcteurs. Beaucoup trop de candidats rédigent en empilant des formules, sans même déclarer les objets. Cette pratique est également sanctionnée. Il est indispensable de rédiger de manière claire, précise et si possible non diluée. Enfin, la présentation se doit d'être claire et agréable et de mettre en évidence des résultats.

Analyse détaillée des questions

Nous ne parlerons que des questions 1 à 20, qui sont celles ayant été traitées par un nombre significatif de candidats.

Q1. Traitée dans l'immense majorité des copies. Quelques réponses fantaisistes cependant, pouvant même dépendre de n.

Q2. Il suffit d'appliquer la définition de d et d'évaluer l'expression dépendant de A. Cette définition n'est pas souvent comprise et la question s'est révélée sélective. À noter la réponse fréqente 0, qui atteste un certain manque de bon sens vu le sens intuitif du mot distance.

Q3.Q4. Beaucoup d'erreurs sur ces questions simples. Il suffisait d'établir la convergence absolue en utilisant le caractère borné de f. De nombreux candidats se contentent d'arguments qui ne prouvent rien : majoration (sans valeur absolue) du terme général, ou majoration de la valeur absolue de la somme partielle (qui n'est pas la somme partielle des valeurs absolues). D'autres erreurs ont été commises (une série dont le terme général tend vers 0 converge, application de la règle de d'Alembert en prétendant que $(f(n+1)/f(n))$ est bornée ...).

Q5. Question en général réussie.

Q6. Question immédiate en testant l'égalité sur les indicatrices de singletons. Cette démarche demande cependant une vraie compréhension des objets. Les correcteurs ont lu beaucoup d'argument faux (identification de coefficients de séries entières mal définies, choix $f(n)=1/n$ suivie de l' argument selon lequel une série convergente de somme nulle a tous ses termes nuls ...).

Q7. Le caractère infini de S_h, qui résulte simplement du fait que $f(0)$ peut être choisi arbitrairement, est très mal compris. Beaucoup de candidats ne voient pas que h est fixée, d'autres invoquent des arguments vagues de dimension, voire un théorème de type Cauchy linéaire pour les équations différentielles.

La seconde partie de la question est mieux traitée. Cependant, nombre de candidats se contentent de vérifier que l'expression proposée de $f(n)$ donne un élément de S_h, ce qui ne prouve rien. Parmi ceux qui raisonnent par récurrence, beaucoup oublient l'intialisation.

Q8. Cette question peut se déduire de la formule de Q7, ou être démontrée directement. L'erreur mentionnée plus haut (vérification du fait que l'expression proposée vérifie S_h) se retrouve souvent ici.

Q9. Cette question délicate reçoit très rarement une solution correcte. Le fond de l'affaire est que

$$\sum_{k=n}^{+\infty} \frac{\lambda^k}{k!} : \frac{\lambda^n}{n!}.$$

Une estimation plus simple, où l'équivalent est remplacé par une inégalité, suffit en fait. Il est judicieux, pour y voir clair, d'expliciter les premiers termes de la série de l'énoncé.

Q10.Q11. Ces questions se déduisent de Q6 et Q7. Encore faut-il expliciter l'argument, sans se contenter d'une référence incantatoire. Curieusement, beaucoup de candidats oublient au moins un des deux signes demandés.

Q12. Cette question sélective nécessite un certain soin dans le traitement des inégalités. Elle est souvent abordée, rarement vraiment résolue.

Q13. Comme on ne suppose pas $f(0)=0$, la première formule demandée n'est pas nécessairement vraie. Les correcteurs ont tenu compte de cette petite erreur de texte et noté avec générosité. La seconde formule demande un peu de persévérance dans les caluls. Elle est correctement établie dans un nombre conséquent de copies ; pas mal de tentatives de passage en force toutefois.

Q14. La majoration de $\Delta f_m(n)$ par $\Delta f_m(m)$ est souvent vue, mais rares sont les candidats qui tirent parti du second item de Q13 pour conclure.

Q15. Un calcul simple, souvent mené de façon aveugle, alors que le fait que $h-h_+$ soit constante permet de conclure rapidement.

Q16. De nombreuses confusions dans cette question, notamment quant aux rôles de m et n.

Q17. L'indicatrice $1_{\{m\}}$ est rarement utilisée.

Q18. Cette question, qui demande de la lucidité, est peu abordée.

Q19. Cette question est assez souvent abordée ; malheureusement, beaucoup de candidats se contentent peu ou prou de recopier l'énoncé. La première égalité de Q19 est immédiate par disjonction des cas $X_k=0, X_k=1$. Pour la seconde, il faut justifier l'indépendance de X_k et W_k, qui résulte de l'indépendance (mutuelle) de $X_1,\ldots,X_k$ et utiliser le théorème du produit des espérances.

Q20. Question peu réussie. La complexité apparente de la réponse cache cependant un argument très simple.