1.2.3. Mathématiques I — PC

Présentation du problème

Ce problème a pour but d'établir le résultat suivant dû à Komlòs (1967).

Théorème. Soit, pour $n \in \mathbb{N}^*$, $(X_{i,j}^n)_{1\leq i,j\leq n}$ une famille de variables aléatoires de Rademacher mutuellement indépendantes. Soit M_n la matrice aléatoire $M_n = \left(X_{i,j}^n\right)_{1\leq i,j\leq n}$. Alors

$$P(M_n \in \mathrm{GL}_n(\mathbb{R})) \underset{n\to+\infty}{\longrightarrow} 1.$$

Le sujet commence par quelques questions simples, dont certaines sont utiles dans la suite, qui recouvrent l'ensemble des thèmes abordés dans le problème (inégalités, asymptotique, algèbre linéaire, probabilités et combinatoire). Les questions 5, 6, 7 permettent de se familiariser avec les matrices M_n dans le cas simple n = 2. À partir de la question 8, on rentre dans la démonstration du théorème de Komlòs.

Nous allons décrire dans ce préambule les grandes lignes de cette démonstration.

Nous noterons L^n_1 ,... , L^n_n les lignes de M_n. Le point de départ est l'inégalité suivante, qui résulte immédiatement de la sous-additivité de P :

$$P((M_n \notin \mathrm{GL}_n(\mathbb{R})) \leq \sum_{j=1}^{n} P(L_j^n \in \mathrm{Vect}(L_1^n, \ldots, L_{j-1}^n)).$$

On combine alors les arguments suivants :

- Un énoncé déterministe d'algèbre linéaire assure qu'un sous-espace de dimension d de R^n contient au plus 2^d vecteurs à coordonnées ±1. L'indépendance des L^n_j permet d'en déduire, pour j dans {1,... , n}, la majoration

$$P(L_j^n \in \mathrm{Vect}(L_1^n, \ldots, L_{j-1}^n) \leq \frac{1}{2^{n-j+1}}.$$

Ce fait est établi à la question 12 du problème.

- La majoration précédente n'est pas bonne lorsque j est proche de n. Il faut utiliser dans ce cas un argument plus subtil, qui tienne compte du caractère aléatoire des espaces

$\mathrm{Vect}(L^n_1,\ldots,L^n_j)$ pour $1 \leq j \leq n$.

On exploite l'aléa en utilisant un résultat combinatoire classique. Appelons antichaîne sur {1,...,n} toute sous-partie de *P*({1,...,n}) dont les éléments sont deux à deux incomparables pour l'inclusion.

Théorème (lemme de Sperner). Le cardinal maximal d'une antichaîne sur $\{1, \ldots, n\}$ est $\displaystyle\binom{n}{\lfloor n/2\rfloor}$.

Littlewwod et Offord ont déduit du lemme de Sperner le résultat suivant.

Théorème. Soient $x_1, \ldots, x_n$ des nombres réels de valeurs absolues supérieures ou égales à 1, I un intervalle semi-ouvert de $\mathbb{R}$ de longueur 2. Le nombre de n-uplets $(\varepsilon_1, \ldots, \varepsilon_n)$ de $\{\pm 1\}^n$ tels que $\sum_{i=1}^{n} \varepsilon_i x_i \in I$ est majoré par $\binom{n}{\lfloor n/2 \rfloor}$.

Ce résultat déterministe se reformule en une inégalité probabiliste exprimant un phénomène d'« anti-concentration ».

Théorème. Soient $(X_i)_{1 \leq i \leq n}$ une famille de variables de Rademacher indépendantes, $\lambda_1, \ldots, \lambda_n$ des nombres réels non nuls, I un intervalle semi-ouvert de $\mathbb{R}$ de longueur 2. Alors

$$P\left(\sum_{i=1}^{n} \lambda_i X_i \in I\right) \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n}.$$

En particulier

$$P\left(\sum_{i=1}^{n} \lambda_i X_i = 0\right) \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n}.$$

L'ensemble des faits ci-dessus est l'objet des questions 14 à 20.
L'inégalité d'anti-concentration précédente est la clé de la suite. Elle suggère de distinguer, parmi les sous-espaces de R^n, ceux dont l'orthogonal contient au moins un vecteur ayant p coordonnées non nulles[1]. On met alors en forme des idées suivantes. Si Vect(L^n_1 , . . . , L^n_j) est un sous-espace vérifiant la condition précédente pour une grande valeur de *p*, la probabilité

$$P\,(L^n_{j+1} \in \mathrm{Vect}(L^n_1, \ldots, L^n_j))$$

est petite grâce à l'inégalité d'anti-concentration ; il est par ailleurs peu probable que Vect(L^n_1 , . . . , L^n_j) ne vérifie pas la condition précédente. Les détails de cet argument occupent les dernières questions du problème.

Le théorème de Komlòs suggère l'étude asymptotique de la suite

$$\left(P(M_n \notin \mathrm{GL}_n(\mathbb{R}))\right)_{n \geq 1}. \quad (1)$$

La démonstration proposée dans le sujet conduit à une majoration très grossière. En fait, on conjecture que

$$\forall \varepsilon > 0, \qquad P\left(M_n \notin \mathrm{GL}_n(\mathbb{R})\right) \underset{n \to +\infty}{=} O\left(\left(\frac{1}{2} + \varepsilon\right)^n\right).$$

On ne peut pas améliorer le nombre 1/2 : en effet,

$$P\left(M_n \notin \mathrm{GL}_n(\mathbb{R})\right) \geq P(L_1^n = L_2^n) \quad \text{et} \quad P(L_1^n = L_2^n) = \frac{1}{2^{n-1}}.$$

On sait depuis 1998 que la convergence de la suite (1) est géométrique. Le meilleur résultat actuel semble être une estimation de Bourgain, Vu et Wood (2010) :

[1] À vrai dire, il s'agit surtout d'une question de dualité ; la structure euclidienne ne joue aucun rôle ici.

$$P\left(M_n \notin \mathrm{GL}_n(\mathbb{R})\right) \underset{n \to +\infty}{=} O\left(\frac{1}{\sqrt{2}^n}\right).$$

Commentaires généraux

Le sujet abordait un grand nombre de notions du programme : probabilités, combinatoire, analyse asymptotique, algèbre linéaire et espaces euclidiens. Cette diversité thématique semble avoir déconcerté une bonne partie des candidats. À regarder de près les questions, peu d'entre elles étaient réellement délicates et peu vraiment simples ; beaucoup, en particulier, demandaient un certain soin dans la rédaction. L'ensemble était par ailleurs assez long.

Malgré son intérêt mathématique incontestable, le sujet s'est donc révélé un peu difficile ; il a en outre le défaut de s'appuyer exclusivement sur le programme de première année.

L'épreuve a cependant permis de mettre en évidence un nombre significatif de très bonnes copies et un lot important de copies satisfaisantes, ayant traité correctement une petite moitié des questions. En revanche, les correcteurs déplorent un contingent assez fort de copies presque vides et une quantité surprenante de copies superficielles, qui donnent à beaucoup de questions des réponses sans aucun contenu.

Conseils aux futurs candidats

Comme d'habitude, ce sujet récompensait le travail en profondeur du cours. Le caractère atypique de l'épreuve a valorisé les candidats capables de prendre de la hauteur et de rédiger avec soin et précision. Nous incitons les candidats à apprendre leur cours de manière réfléchie et à ne pas manipuler aveuglément les objets mathématiques. Nous leur recommandons également, face un sujet difficile, de bien traiter une partie des questions plutôt que de produire un discours inconsistant pour chacune d'entre elles : les tentatives de bluff n'apportent aucun point et préviennent très défavorablement le correcteur quant à l'ensemble de la copie.

D'un point de vue plus technique, on souligne les points suivants.

- Les probabilités appellent une rédaction aussi précise que les autres parties des mathématiques ; par ailleurs, même dans le cadre �ni qui est celui du sujet, elles ne se réduisent pas à la combinatoire.
- Les bases de l'analyse (majorations, estimations asymptotiques) sont au coeur d'une grande part des mathématiques ; on ne les acquiert que par une pratique assidue du calcul.
- Les notions fondamentales de l'algèbre linéaire de première année sont également indispensables dans une grande variété de domaines et doivent être maîtrisées.

Rappelons pour conclure l'importance de la présentation. Les copies peu lisibles sont pénalisées ; on recommande aux candidats d'employer une encre foncée, qui reste bien visible sur les copies scannées. Une présentation soignée (écriture nette, absence de ratures, résultats encadrés) dispose très favorablement le correcteur.

Analyse détaillée des questions

Q1. Des calculs qui n'aboutissent pas toujours, avec pas mal d'erreurs. Un grand flou sur l'intervalle où est choisi k : certains candidats établissent la croissance hors de l'ensemble d'étude! Par ailleurs, oubli fréquent de la symétrie dans la fin de la question.

Q2. Il était indispensable, pour répondre correctement, de dissocier les cas n pair et n impair, ce qui a très rarement été fait. Beaucoup de candidats maîtrisent mal la notion d'équivalent, d'où des calculs abusifs dans la première partie de la question (on remplace sans vergogne ⌊n/2⌋ par n/2) et des affirmations trop rapides dans la seconde partie.

Q3. Cette question reposait sur des inégalités très élémentaires. Beaucoup de candidats se sont perdus dans tentatives incomplètes de raisonnement par récurrence. Une fraction non négligeable a tenté

d'utiliser la question précédente, ce qui est voué à l'échec dans la mesure où Q2 ne donne qu'un renseignement asymptotique.
Q4. La première partie de la question a été en général résolue. La seconde partie n'a pas été traitée dans la grande majorité des copies ; il suffisait de voir que v et les $v - 2e_i$ sont dans $\Omega_{1,n}$, mais la notion de sous-espace engendré semble souvent mal comprise.
Q5.6.7. Ces questions ont souvent été traitées de manière combinatoire, de manière plus ou moins convaincante selon les copies. Peu de candidats ont noté que Q5 se faisait immédiatement avec les propriétés de l'espérance. La variance était donnée dans Q6, ce qui a conduit à un certain nombre de tentatives d'escroquerie. Dans Q7, certaines copies trouvent des probabilités strictement su- périeures à 1 : on conseille aux candidats de prendre un peu de recul !
Q8. La rédaction de cette question est souvent approximative et ne fait pas nettement apparaître les arguments probabilistes sous-jacents : indépendance, incompatibilité, monotonie et sous-additivité de P. Par ailleurs, beaucoup de candidats calculent $P(L_1 = L_2)$ et $P(L_1 = -L_2)$ sans répondre vraiment à la question.
Q9. Dans la première partie de la question, les deux sens sont rarement traités. On note un certain flou quant à la notion de famille liée, ainsi que des erreurs de logique. Pour la deuxième partie de la question, on relève à nouveau, dans nombre de copies, une maîtrise insuffisante du formalisme des probabilités ; dans un certain nombre de copies, le lien entre la nullité du déterminant et le caractère lié de la famille des lignes n'apparaît pas.
Q10. Cette question a donné lieu à beaucoup de réponses dénuées de sens ; ce sont ici des difficultés de logique qui sont en cause (le quantificateur existentiel précède l'équivalence)
Q11. Le lien entre cette question et la question 10 a rarement été perçu ; la question a reçu très peu de bonnes réponses.
Q12. Question rarement traitée.
Q13. Cette question n'a pratiquement jamais été bien traitée ; elle est à vrai dire assez difficile à résoudre lors d'une épreuve en temps limité avec le programme de la filière PC.
Q14. Cette question a souvent été traitée. Cependant, la rédaction de la première partie est souvent filandreuse. Il suffit de dire nettement qu'une partie E d'un ensemble fini F ayant même cardinal que F ne peut être que F.
Q15. Cette petite question de dénombrement, ouverte, s'est révélée sélective. À noter qu'un certain nombre de candidats donnent la réponse correcte sans aucune explication, ce qui ne peut valoir tous les points.
Q16. Question souvent abordée, traitée dans un certain nombre de copies ; la rédaction n'est pas toujours claire et beaucoup de candidats semble confondre « incomparables » et « disjoints ».
Q17. Cette question combinatoire difficile n'a été bien traitée que dans les meilleures copies.
Q18. Q19 Ces questions simples ont permis à la plupart des candidats qui les ont abordées de récupérer quelques points.
Q21. Cette question facile nécessitait une maîtrise correcte de la logique et de la syntaxe ensembliste. Assez souvent abordée, elle a connu des fortunes très diverses.
Les questions suivantes ne concernent qu'une faible fraction des candidats.

1.2.4. Mathématiques II — PC

Le sujet de la deuxième épreuve de mathématiques PC était consacré aux propriétés classiques des fonctions harmoniques définies sur un ouvert de R^2 et à valeurs réelles ou complexes. De facture certes classique, ce problème était tout à fait dans l'esprit du programme PC, et conçu pour aborder un très grand nombre de chapitres du programme d'analyse.