Mathématiques 2

Présentation du sujet

Le problème porte sur les conditions sous lesquelles une variable aléatoire à valeurs dans $\mathbb{N}$ peut être écrite comme somme de deux variables entières non constantes et indépendantes (variable décomposable), ou bien comme somme de variables discrètes indépendantes et de même loi (variable divisible).

La première partie s'intéresse à quelques exemples de variables décomposables ; on y établit notamment les conditions sous lesquelles une variable binomiale, ou une variable uniforme, sont décomposables.

Dans la deuxième partie, on introduit la notion de variables infiniment divisibles et on étudie des exemples : variables constantes, variables bornées, variables de Poisson et somme de multiples entiers de telles variables. On termine cette partie en établissant qu'une telle somme pondérée de variables de Poisson converge, sous certaines hypothèses, vers une variable infiniment divisible.

La troisième partie propose une caractérisation des variables à valeurs dans $\mathbb{N}$ infiniment divisibles ; une étude très guidée, faisant intervenir le logarithme de la fonction génératrice, permettait de voir que ces variables sont exactement celles qui ont été construites en fin de deuxième partie.

Analyse globale des résultats

Une très grande partie des candidats confond les notions les plus importantes en probabilités, notamment les variables aléatoires et leurs lois. Dans ce contexte, la propriété d'indépendance est très mal comprise ; et son importance capitale dans les définitions proposées par le sujet n'a pas toujours été perçue.

Peu de candidats font l'effort nécessaire pour s'approprier les notions du sujet, en vérifiant scrupuleusement les définitions à l'aide des premiers exemples traités. Ceci va de pair avec une maitrise insuffisante de la langue : il est souvent bien difficile, par exemple, de distinguer une hypothèse supplémentaire d'une étape (censée avoir été démontrée) dans le raisonnement.

Les questions d'analyse (séries entières) et d'algèbre (polynômes) ont été décevantes ; le jury note un manque de rigueur général sur ces sujets, sans parler de la tendance à inventer de nouveaux « théorèmes » en cas de difficulté dans une question.

Les meilleurs candidats sont ceux qui ont abordé le problème en citant précisément les hypothèses du sujet et les théorèmes du cours qui permettaient de résoudre les questions posées. Les questions plus délicates leur ont permis d'exprimer des idées pertinentes, et même si leur mise en œuvre n'était pas parfaite, elles ont été valorisées.

Commentaires sur les réponses apportées et conseils aux futurs candidats

Afin d'aider la préparation des futurs candidats et de préciser les attentes du jury, un corrigé commenté est disponible en annexe de ce rapport.

Partie I

Pour la première question, les candidats oublient presque toujours de préciser que la rayon de convergence de G_X est strictement positif ; il n'est pas possible d'utiliser l'unicité des coefficients sans cette hypothèse.

Trop souvent, on lit que les fonctions génératrices intervenant dans cette partie sont des polynômes, sans qu'une explication en soit donnée.

Dans **I.A.3** et suivantes, et également dans **II.B.3**, il faut rappeler au candidats que l'indépendance est une notion qui s'applique aux variables aléatoires, pas aux lois ; on ne peut pas déduire du choix des lois de deux variables qu'elles sont indépendantes. Et l'indépendance est essentielle dans la définition de décomposition, il faut absolument en faire état, pour montrer qu'on a compris la notion.

Trop souvent, dans cette question et dans **I.B.2a**, le caractère indécomposable d'une variable est confondu avec la simple irréductibilité de sa fonction génératrice, lorsque c'est un polynôme.

Dans **I.A.4a**, l'énoncé invite explicitement à considérer tous les cas possibles pour les degrés de U et V ; traiter le cas $(1,3)$ et expliquer que l'autre cas $(2,2)$ est similaire est alors un gain de temps, mais une perte de points. Deux polynômes qui factorisent un polynôme unitaires ne sont pas forcément unitaires aussi. Dans le cas où l'on voit que $A(-1) = 0$, ce n'est pas parce qu'un polynôme unitaire de degré 1 divise A qu'il s'agit forcément de $T + 1$; il y a une autre racine réelle... Pour résoudre la question en raisonnant avec les deux racines réelles, donc ici avec les trois facteurs irréductibles dans $\mathbb{R}[X]$, il faut être très clair, rigoureux, et surtout exhaustif.

Dans **I.A.4b**, indiquer qu'on choisit X suivant la loi $\mathcal{B}(2, 1/2)$ permet d'éviter un verbiage inutile.

La question **I.B.1a** n'a pas été comprise ; le théorème de la division euclidienne permettait de définir les applications Q et R (ce point précis a rarement été fait explicitement) et il fallait vérifier que ces applications étaient les seules qui convenaient. Moins d'un candidat sur 100 s'est intéressé ensuite à la démonstration du fait que Q et R sont des variables aléatoires ; c'est pourtant explicitement demandé.

La question **I.B.2a** n'a pas toujours été comprise ; un candidat a éclairé les correcteurs, en affirmant que la question n'avait aucun sens, puisqu'on ne demandait pas de prouver un résultat, mais de prouver qu'il suffisait de prouver un autre résultat ! Le lien entre U, V, G_Y et G_Z n'a pas toujours été bien explicité et très peu de candidats se sont intéressés aux hypothèses imposées à U et V (positivité des coefficients, polynômes unitaires).

Plus de la moitié des candidats traitent la question (difficile) **I.B.2b**, en inventant un théorème d'identification : si on a deux factorisations $A = UV = U_1V_1$ avec égalité des degrés (même pas d'égalité des coefficients dominants), alors $U = U_1$ et $V = V_1$. Il est clair qu'une telle affirmation est lourdement pénalisante pour la suite de la copie.

Partie II

Beaucoup de candidats ne traitent pas correctement la première question, en omettant d'évoquer l'indépendance, ou bien en oubliant de choisir la même loi pour les variables qui interviennent.

La question **II.A.2a** a montré une grande confusion dans l'esprit des candidats : plus de 80% de ceux qui traitent la question pensent que $X_1 + \cdots + X_n$ a même loi que nX_1, puisque les X_i sont indépendantes et de même loi. Aucun d'entre eux ne semble réagir en se disant que dans ces conditions, la notion de divisibilité ne présenterait aucun intérêt.

L'existence de la variance dans **II.A.2b** n'a que rarement été mentionnée ; on y a souvent vu par contre l'enchainement absurde $\mathbb{V}(X_1) \leqslant \mathbb{V}(M/n) = M^2/n^2$.

En **II.A.2c**, il fallait démontrer explicitement que la nullité de la variance entrainait que la variable est presque sûrement constante. L'inégalité de Bienaymé-Tchebitchev, même avec $\mathbb{V}(X) = 0$, ne suffisait pas, seule, pour conclure.

La question **II.B.1** est souvent résolue en faisant référence aux résultats de la partie I, sans s'apercevoir que le contexte n'est plus le même (les variables ne sont plus à valeurs entières). Les cas particuliers $p = 0$ et $p = 1$ sont oubliés.

La question **II.B.2** est une réussite, mais les candidats doivent démontrer la formule donnant la fonction caractéristique d'une variable de Poisson. En procédant par récurrence, les étudiants oublient souvent d'établir l'indépendance des variables à l'aide du lemme des coalitions.

La question difficile **II.B.4** a permis aux candidats ayant le plus de recul de proposer une solution, fortement valorisée, bien que la démonstration de l'indépendance des variables soit souvent incomplète.

La formule des probabilités totales n'est qu'exceptionnellement citée pour le calcul de la probabilité $\mathbb{P}(X \neq Y)$ dans la question **II.C.1b**.

Dans la question **II.C.2a**, il est souvent fait mention du théorème de continuité décroissante, alors qu'il ne permet pas de conclure ; c'est l'inégalité de Boole qu'il fallait utiliser ici et le théorème de continuité à la question suivante.

Dans la question **II.C.2c**, les candidats croient le plus souvent que $\lim S_n = S$ entraine automatiquement que $\lim \mathbb{P}(S_n \neq S) = 0$.

Partie III

Cette partie a été peu abordée. La première question a souvent donné lieu a des rédactions très longues, alors qu'il suffisait d'expliciter (λ_k) comme suite récurrente pour conclure en quelques lignes. Dans la question **III.A.6**, les candidats prennent le logarithme des fonctions génératrices à l'intérieur du disque de convergence sans s'inquiéter de la positivité des fonctions ; d'autres croient pouvoir conclure sans faire référence à l'indépendance des variables, ni en fait à un lien quelconque entre G_{X+Y}, G_X et G_Y.

Conclusion

Faire des probabilités, c'est avant tout faire des mathématiques ; l'intuition y joue certes un rôle accru, mais elle ne remplace pas une vraie démonstration explicitant les hypothèses et les théorèmes utilisés.

Il convient de s'approprier rapidement et surtout précisément les notions du sujet, afin de bien mettre en valeur les points les plus importants des démonstrations. Ici par exemple, l'indépendance des variables était centrale, et il convenait de la mentionner explicitement à chaque fois que c'était nécessaire.

Rappelons pour finir quelques évidences. Une copie rédigée correctement permet de suivre sans effort la démarche proposée par le candidat, et ceci est toujours apprécié et valorisé. À l'inverse, une copie difficilement lisible, écrite dans un français approximatif, présentant de nombreuses ratures ou fautes d'orthographe, ne mettant pas en valeur les résultats démontrés, est forcément sanctionnée, plus ou moins consciemment.