Centrale Maths 2 MP 2017 — Corrigé

Ce corrigé est proposé par Alban Levy (docteur en mathématiques) ; il a été relu par Yvon Vignaud (professeur en CPGE) et Florian Metzger (docteur en mathématiques).

Ce problème d'analyse propose l'étude de la représentation de la loi d'une variable aléatoire comme loi d'une somme de variables aléatoires discrètes indépendantes. En particulier, l'égalité des fonctions génératrices de deux variables aléatoires discrètes étant équivalente à l'égalité de leurs lois, l'étude des distributions de probabilité est souvent ramenée à une étude des fonctions génératrices. La première partie est largement indépendante des deux autres.

- Dans la première partie, la décomposition étudiée est une somme de deux variables à valeurs entières. Quelques lois classiques y sont vues, comme les lois binomiale et uniforme. Les polynômes et fonctions génératrices y jouent un rôle important.
- Dans la deuxième partie, on étudie quelques exemples de variables aléatoires à valeurs dans $\mathbb{N}$ ayant la propriété d'être infiniment divisibles, c'est-à-dire s'écrivant comme une somme de m variables aléatoires mutuellement indépendantes et de même loi, pour $m \in \mathbb{N}^*$ quelconque. Les variables aléatoires proposées sont bornées ou de Poisson.
- La troisième partie étudie la cas général des variables infiniment divisibles. Le résultat principal est la caractérisation de ces variables de deux façons.

Il s'agit d'un sujet ambitieux qui navigue astucieusement entre algèbre et analyse, utilisant comme ligne directrice les lois de probabilités classiques mais incluant de nombreux raisonnements autour des polynômes, séries numériques et séries entières. Outre une bonne maîtrise des probabilités au programme de MP, il faut donc être à l'aise avec une importante partie du programme d'analyse. On trouvera par exemple une équation différentielle, des produits infinis ou un polynôme cyclotomique. On démontre aussi le lemme de Borel-Cantelli. Bien que les trois derniers ne soient pas au programme, il est utile de les avoir déjà étudiés.

La résolution de ce sujet est gratifiante car elle tisse des liens entre des concepts mathématiques habituellement éloignés, requérant parfois des raisonnements purement algébriques.

Indications

Partie I

I.A.1 Calculer les dérivées n-ièmes de G_X en 0 pour exprimer la loi de X.

I.A.2 Exprimer G_X comme une espérance pour utiliser l'indépendance de Y et Z.

I.A.3 Séparer les cas $n = 1$ (par l'absurde) et $n \geqslant 2$ (somme de binomiales).

I.A.4.a Supposer que le degré de U vaut 1 ou 2 et déterminer les coefficients de U.

I.A.4.b Voir que $A(T)/4$ est la fonction génératrice du carré d'une variable aléatoire de loi bien connue.

I.B.1.a Utiliser l'existence et l'unicité de la division euclidienne.

I.B.1.b Obtenir la loi du couple (Q, R) par unicité d'une division euclidienne puis sommer pour obtenir les lois marginales.

I.B.1.c Montrer l'indépendance de Q et R en tant que variables aléatoires.

I.B.2.a Raisonner par l'absurde.

I.B.2.b Décomposer U en produit de polynômes de degré 1.

I.B.2.c Écrire la dérivée r-ième de UV en 0 de deux façons.

I.B.2.d Procéder par récurrence sur k allant de 1 à r, en utilisant la dérivée $(k+1)$-ième de UV en 0 pour montrer que les propriétés $u_i \in \{0, 1\}$ et $v_i \in \{0, 1\}$ sont vraies pour $i \in [\![1\,; k+1]\!]$.

I.B.2.e Regarder $(UV)(1)$ droit dans les yeux : un produit d'entiers se cache dans ce terme.

Partie II

II.A.1 Utiliser le résultat admis dans le préambule pour avoir m copies de X/m.

II.A.2.a Établir une première inégalité entre $\mathbb{P}(X_1 > M/n)^n$ et $\mathbb{P}(X > M)$ et une autre entre $\mathbb{P}(|X_1| > M/n)$ et $\mathbb{P}(X_1 > M/n) + \mathbb{P}(X_1 < -M/n)$.

II.A.2.b Majorer $\mathbb{V}(X_1)$ par $\mathbb{E}[(X_1)^2]$.

II.A.3 Montrer que $\mathbb{V}(X) = 0$.

II.B.1 Se ramener à une variable aléatoire bornée.

II.B.2 Se référer à la question où a été établi le résultat sur la fonction génératrice.

II.B.3 Sommer m variables aléatoires de lois judicieusement choisies.

II.B.4 Décomposer chacun des r termes X_i en m variables aléatoire indépendantes, et s'assurer que ces rm variables aléatoires peuvent être choisies mutuellement indépendantes.

II.C.1.a Partir de $A = (A \cap B) \cup (A \cap \overline{B})$ et aller vers une majoration de $\mathbb{P}(A \cap B)$ par $\mathbb{P}(A)$ et $\mathbb{P}(B)$.

II.C.1.b Utiliser les définitions des fonctions génératrices puis majorer en utilisant la question II.C.1.a.

II.C.2.a Majorer $\mathbb{P}(Z_n)$ par le reste d'une série convergente faisant intervenir les U_i.

II.C.2.b Relier l'ensemble $\{i \in \mathbb{N}^* | U_i \neq 0\}$ aux variables Z_m pour $m \in \mathbb{N}$.

II.C.2.c Utiliser les questions II.C.2.b et II.C.1.b et relier l'événement $(S \neq S_n)$ à Z_{n+1}.

II.C.3.a Noter que $\mathbb{P}(X_i \neq 0) = 1 - e^{-\lambda_i}$ puis utiliser une comparaison de séries.

II.C.3.b Étudier la limite de la suite de fonctions génératrices $G_{X_1+\cdots+X_n}$ puis conclure par unicité de la limite.

II.C.3.c Vérifier que les hypothèses de la question II.C.2 s'appliquent puis utiliser le résultat de la question II.B.4.

Partie III

III.A.1 Procéder par récurrence sur k pour montrer l'existence et l'unicité de la famille $(\lambda_i)_{i=1,\ldots,k}$.

III.A.2 Majorer certaines probabilités en jeu par $\mathbb{P}(X \geqslant 1)$.

III.A.3 Procéder par récurrence sur $k \in \mathbb{N}^*$.

III.A.4.a Invoquer la règle de d'Alembert.

III.A.5 Utiliser un produit de Cauchy puis le théorème de Cauchy-Lipschitz.

III.B.1 Se servir de la formule construisant $\lambda_1, \ldots, \lambda_k$.

III.B.2 Prouver la convergence normale de la série définissant H_X sur $[-1\,;1]$.

III.B.3 Montrer l'égalité des fonctions génératrices de X et de la somme de $\sum iX_i$.

III.C.1.a Il suffit que tous les $X_{n,i}$ pour $n \in \mathbb{N}^*$ et $i \in [\![1\,;n]\!]$ soient négatifs pour que X le soit.

III.C.1.b Calculer la valeur de $\mathbb{P}(X_{n,1} = 0)$.

III.C.1.c Si la variable X_1 possède une valeur non entière et que les autres variables X_j pour $j \geqslant 2$ sont nulles, alors la variable X n'est pas entière.

III.C.2.a On connaît $\mathbb{P}(X_{n,1} = 0)$.

III.C.3.b Identifier les coefficients des deux séries entières définissant nH_n et H_X.

III.C.4 Utiliser les questions III.C.2.a et b.

III.C.5.a Rechercher trois implications dans les questions précédentes.

III.C.5.b Se servir de $X - m_X$ où m_X est le minimum presque sûr de X.

III.C.5.c Calculer explicitement la famille $(\lambda_k)_{k\in\mathbb{N}^*}$.

I. Variables aléatoires entières décomposables

I.A.1 Si $\mathrm{X} \sim \mathrm{X}'$, alors pour tout $n \in \mathbb{N}$, $\mathbb{P}(\mathrm{X} = n) = \mathbb{P}(\mathrm{X}' = n)$ d'où l'égalité de chaque terme général des séries entières définissant, pour tout $t \in \mathbb{R}$, $\mathrm{G_X}(t)$ et $\mathrm{G_{X'}}(t)$. Ainsi, leurs sommes sont égales sur leur domaine de définition.

Réciproquement, supposons que $\mathrm{G_X} = \mathrm{G_{X'}}$. Les coefficients de la série entière définissant $\mathrm{G_X}$ étant des probabilités, donc compris entre 0 et 1, par comparaison de séries à termes positifs, cette série a un rayon de convergence supérieur ou égal à celui de la série $\sum t^n$, c'est-à-dire 1. La fonction $\mathrm{G_X}$ est de plus de classe $\mathscr{C}^\infty$ au voisinage de 0 et

$$\forall n \in \mathbb{N} \qquad {\mathrm{G_X}}^{(n)}(0) = n!\, \mathbb{P}(\mathrm{X} = n)$$

et de même pour $\mathrm{G_{X'}}$. L'égalité de $\mathrm{G_X}$ et de $\mathrm{G_{X'}}$ entraînant l'égalité de leurs dérivées en zéro, on en déduit que

$$\boxed{\mathrm{X} \sim \mathrm{X}' \iff \mathrm{G_X} = \mathrm{G_{X'}}}$$

I.A.2 Fixons $t \in \mathbb{R}$ tel que la somme définissant $\mathrm{G_X}(t)$ soit convergente, ce qui inclut l'intervalle $]\,-1\,;1\,[$ d'après la question I.A.1. Réécrivons alors $\mathrm{G_X}(t)$ comme une espérance grâce au théorème de transfert :

$$\mathrm{G_X}(t) = \sum_{n=0}^{+\infty} \mathbb{P}(\mathrm{X} = n) t^n = \mathbb{E}[t^{\mathrm{X}}]$$

La décomposition $\mathrm{X} \sim \mathrm{Y} + \mathrm{Z}$ donne de plus l'égalité $\mathbb{E}[t^{\mathrm{X}}] = \mathbb{E}[t^{\mathrm{Y}+\mathrm{Z}}] = \mathbb{E}[t^{\mathrm{Y}}\, t^{\mathrm{Z}}]$. Enfin, l'indépendance de Y et Z entraînant celle de t^{Y} et t^{Z}, on peut écrire

$$\mathbb{E}[t^{\mathrm{Y}}\, t^{\mathrm{Z}}] = \mathbb{E}[t^{\mathrm{Y}}]\, \mathbb{E}[t^{\mathrm{Z}}] = \mathrm{G_Y}(t)\, \mathrm{G_Z}(t)$$

en reconnaissant l'expression de $\mathrm{G_Y}$ et $\mathrm{G_Z}$. Ainsi,

$$\boxed{\text{Pour Y et Z indépendants, l'égalité } \mathrm{G_{Y+Z}} = \mathrm{G_Y}\,\mathrm{G_Z} \text{ est vraie sur le domaine de définition commun à } \mathrm{G_Y} \text{ et } \mathrm{G_Z}. \text{ En particulier, } \mathrm{G_X} = \mathrm{G_Y}\,\mathrm{G_Z} \text{ sur }]\,-1\,;1\,[.}$$

I.A.3 • Si $n = 1$, $\mathrm{X} \sim \mathcal{B}(n,p)$ suit une loi de Bernoulli de paramètre p. Montrons par l'absurde que X n'est pas décomposable. Supposons donc que $\mathrm{X} \sim \mathrm{Y} + \mathrm{Z}$ avec Y et Z indépendants, à valeurs dans $\mathbb{N}$ et presque sûrement non constantes. Comme

$$(\mathrm{Y} \geqslant 1) \cap (\mathrm{Z} \geqslant 1) \subset (\mathrm{Y} + \mathrm{Z} \geqslant 2)$$

et

$$\mathbb{P}(\mathrm{Y} \geqslant 1, \mathrm{Z} \geqslant 1) = \mathbb{P}(\mathrm{Y} \geqslant 1)\, \mathbb{P}(\mathrm{Z} \geqslant 1)$$

par indépendance de Y et Z, on a

$$\mathbb{P}(\mathrm{X} \geqslant 2) = \mathbb{P}(\mathrm{Y} + \mathrm{Z} \geqslant 2) \geqslant \mathbb{P}(\mathrm{Y} \geqslant 1, \mathrm{Z} \geqslant 1) = \mathbb{P}(\mathrm{Y} \geqslant 1)\, \mathbb{P}(\mathrm{Z} \geqslant 1)$$

Comme Y est à valeurs dans $\mathbb{N}$ et non constante presque sûrement, $\mathbb{P}(\mathrm{Y} = 0) < 1$. Il s'ensuit par additivité de $\mathbb{P}$ que $\mathbb{P}(\mathrm{Y} \geqslant 1) = \mathbb{P}(\mathrm{Y} \in \mathbb{N}) - \mathbb{P}(\mathrm{Y} = 0) > 0$, et de même $\mathbb{P}(\mathrm{Z} \geqslant 1) > 0$. Or, $\mathbb{P}(\mathrm{X} \geqslant 2) = 0$ car X suit une loi de Bernoulli. On obtient alors la contradiction suivante montrant finalement que X n'est pas décomposable :

$$0 = \mathbb{P}(\mathrm{X} \geqslant 2) \geqslant \mathbb{P}(\mathrm{Y} \geqslant 1)\, \mathbb{P}(\mathrm{Z} \geqslant 1) > 0$$

• Si $n \geqslant 2$, X suit une loi binomiale $\mathcal{B}(n,p)$. Le résultat donné par l'énoncé garantit l'existence de n variables aléatoires indépendantes $\mathrm{X}_1, \ldots, \mathrm{X}_n$ suivant une loi de Bernoulli de paramètre $p \in \,]\,0\,;1\,[$. Vérifions que $\mathrm{Y} = \mathrm{X}_1$ et $\mathrm{Z} = \mathrm{X}_2 + \cdots + \mathrm{X}_n$ forment une décomposition de X. Leur indépendance est garantie par construction car les $\mathrm{X}_1, \ldots, \mathrm{X}_n$ sont mutuellement indépendants. De plus, la somme de n variables aléatoires de Bernoulli indépendantes et de même paramètre p suit une loi binomiale