Corrigé et barème – Lois de probabilité et statistique à deux variables — niveau ★★★

Durée : 120 min – Calculatrice autorisée – Sur 20 points · Maths complémentaires Terminale

Exercice 1 – Démontrer *(5 pts)*

a) Relation de Pascal avec $k=1$: pour tout entier $m \geqslant 2$, $\binom{m}{1}+\binom{m}{2}=\binom{m+1}{2}$, soit $\binom{m+1}{2}=\binom{m}{2}+m$. Ainsi $\binom{2}{2}=1$, $\binom{3}{2}=1+2$, $\binom{4}{2}=1+2+3$, et, de proche en proche jusqu'à $m=n-1$: $\binom{n}{2}=1+2+\cdots+(n-1)$. C'est la somme des $n-1$ premiers termes d'une suite arithmétique : $\binom{n}{2}=\dfrac{(n-1)\big(1+(n-1)\big)}{2}=\dfrac{n(n-1)}{2}$. (0,5 pt pour l'utilisation de la relation de Pascal, 0,5 pt pour la somme) (1)

b) $\dfrac14-p(1-p)=p^2-p+\dfrac14=\left(p-\dfrac12\right)^2\geqslant 0$, avec égalité pour $p=\dfrac12$. Donc $V(X)=np(1-p)\leqslant \dfrac{n}{4}$ et, la fonction racine carrée étant croissante, $\sigma(X)=\sqrt{V(X)}\leqslant\sqrt{\dfrac{n}{4}}=\dfrac{\sqrt{n}}{2}$. Par exemple, pour $n=100$, $\sigma(X)\leqslant 5$ quel que soit p. (0,5 pt pour l'inégalité, 0,5 pt pour la conséquence) (1)

c) $P\big(X\in\{2\,;\dots\,;2n\}\big)=\displaystyle\sum_{k=1}^{n}(1-p)^{2k-1}p=p(1-p)\Big[1+(1-p)^2+\cdots+\big((1-p)^2\big)^{n-1}\Big]$. C'est la somme de n termes d'une suite géométrique de raison $(1-p)^2\neq 1$: $p(1-p)\times\dfrac{1-(1-p)^{2n}}{1-(1-p)^2}$, et $1-(1-p)^2=p(2-p)$, d'où la formule. Comme $0<(1-p)^2<1$, $(1-p)^{2n}=\big((1-p)^2\big)^n$ tend vers 0 : la limite vaut $\dfrac{1-p}{2-p}$. Or $\dfrac{1-p}{2-p}<\dfrac12 \iff 2-2p<2-p \iff p>0$: un rang impair est plus probable (cohérent avec $P(X=1)=p$, la plus grande des probabilités). (0,5 pt pour la somme géométrique, 0,5 pt pour la limite et la conclusion) (1)

d) Pour $x\geqslant 2$: $\displaystyle\int_2^x \frac{k}{t^2}\,dt=k\Big[-\frac1t\Big]_2^x=k\Big(\frac12-\frac1x\Big)$, qui tend vers $\dfrac{k}{2}$ quand x tend vers $+\infty$. f est continue sur $[2\,;+\infty[$, positive si $k>0$, et l'aire totale vaut 1 si et seulement si $k=2$. Alors $P(X\leqslant 10)=2\Big(\dfrac12-\dfrac{1}{10}\Big)=0{,}8$ et $P(X>10)=0{,}2$. (0,5 pt pour k, 0,5 pt pour la probabilité) (1)

e) $\displaystyle\int_2^x t\times\frac{2}{t^2}\,dt=\int_2^x\frac{2}{t}\,dt=2(\ln x-\ln 2)$, et $\ln x$ tend vers $+\infty$: l'intégrale tend vers $+\infty$. La limite qui définit l'espérance n'est pas un nombre réel : X n'a pas d'espérance. Une variable à densité n'en a donc pas toujours : ici, les grandes valeurs sont rares ($P(X>100)=\dfrac{2}{100}=0{,}02$) mais pèsent assez pour que l'aire sous $t\mapsto t\,f(t)$ soit infinie. (0,5 pt pour le calcul de l'intégrale et sa limite, 0,5 pt pour la conclusion) (1)

Méthode — Pour démontrer une formule portant sur une somme de probabilités, je fais apparaître une suite géométrique (ou arithmétique) en factorisant, je compte ses termes, puis j'applique la formule de la somme.

Exercice 2 – Problème — La première panne d'une éolienne *(5 pts)*

a) 18 % des éoliennes ont leur première panne la première année : $P(D \leqslant 1) = 1 - \mathrm{e}^{-\lambda} = 0{,}18$, donc $\mathrm{e}^{-\lambda} = 0{,}82$ et $\lambda = -\ln 0{,}82 \approx 0{,}198\,5$. $E(D) = \dfrac{1}{\lambda} \approx 5{,}04$ ans. (0,5 pt pour $\mathrm{e}^{-\lambda}$ et λ, 0,5 pt pour la durée moyenne) (1)

b) $\{D > 5\}$ est inclus dans $\{D > 3\}$, donc $P_{D>3}(D > 5) = \dfrac{P(D > 5)}{P(D > 3)} = \dfrac{\mathrm{e}^{-5\lambda}}{\mathrm{e}^{-3\lambda}} = \mathrm{e}^{-2\lambda} = 0{,}82^2 = 0{,}672\,4$. C'est la probabilité qu'une éolienne neuve n'ait pas de panne en 2 ans : la loi exponentielle est sans mémoire. (0,5 pt pour le quotient justifié, 0,5 pt pour la valeur) (1)

c) Pour chaque éolienne, le succès « pas de panne pendant 2 ans » a pour probabilité $P(D > 2) = \mathrm{e}^{-2\lambda} = 0{,}672\,4$; les 20 épreuves sont identiques et indépendantes, et K compte les succès : K suit $\mathcal{B}(20\,;0{,}672\,4)$. À la calculatrice, $P(K \geqslant 15) = 1 - P(K \leqslant 14) \approx 1 - 0{,}684 = 0{,}316$. (0,5 pt pour la loi, 0,5 pt pour la probabilité) (1)

d) La condition est `d > 2`. À chaque tour de boucle, `d` simule la durée sans panne d'une éolienne, et l'on ajoute 1 à `K` si elle dépasse 2 ans. Les 20 tirages sont indépendants et chacun réussit avec la probabilité $P(D > 2) = \mathrm{e}^{-2\lambda}$: la valeur renvoyée suit la loi binomiale $\mathcal{B}\left(20\,;\mathrm{e}^{-2\lambda}\right)$. (0,5 pt pour la condition, 0,5 pt pour la loi) (1)

e) K suit $\mathcal{B}(20\,;p)$. La calculatrice donne $P(K \geqslant 15) \approx 0{,}890 < 0{,}9$ pour $p = 0{,}83$ et $P(K \geqslant 15) \approx 0{,}913 \geqslant 0{,}9$ pour $p = 0{,}84$ (et davantage au-delà) : il faut $p = 0{,}84$. Alors $\mathrm{e}^{-2\lambda} = 0{,}84 \iff \lambda = -\dfrac{\ln 0{,}84}{2} \approx 0{,}087\,2$, soit une durée moyenne sans panne de $\dfrac{1}{\lambda} \approx 11{,}5$ ans. Il faut donc plus que doubler la durée moyenne sans panne du modèle actuel (5,04 ans) : l'exigence est forte. (0,5 pt pour p justifié par les deux valeurs, 0,5 pt pour λ et la durée moyenne commentée) (1)

Astuce — Quand l'énoncé donne une proportion de pannes sur une durée, je la traduis par $P(D \leqslant t) = 1 - \mathrm{e}^{-\lambda t}$; souvent, la valeur de $\mathrm{e}^{-\lambda}$ suffit pour la suite, sans arrondir λ.

Exercice 3 – Vrai ou faux ? *(5 pts)*

a) Vrai. Y compte les échecs du même schéma de Bernoulli : en appelant succès l'ancien échec, de probabilité $1-p$, on a encore n épreuves identiques et indépendantes. Par le calcul : $P(Y=k) = P(X=n-k) = \binom{n}{n-k}p^{n-k}(1-p)^k = \binom{n}{k}(1-p)^k p^{n-k}$, grâce à la symétrie des coefficients binomiaux. (0,5 pt pour la réponse, 0,5 pt pour la justification) (1)

b) Faux. Comme $U \geqslant 0$, $P(U^2 \leqslant 0{,}25) = P(U \leqslant 0{,}5) = 0{,}5$, alors qu'une variable de loi uniforme sur $[0\,;1]$ est inférieure ou égale à 0,25 avec la probabilité 0,25. Plus généralement, pour $0 < x < 1$, $P(U^2 \leqslant x) = P(U \leqslant \sqrt{x}) = \sqrt{x} \neq x$. (0,5 pt pour la réponse, 0,5 pt pour le contre-exemple) (1)

c) Vrai. Pour une droite $\mathcal{D}$, notons $S(\mathcal{D})$ la somme des carrés des résidus de l'ancienne série et e le résidu du point ajouté : pour la nouvelle série, la somme vaut $S(\mathcal{D}) + e^2$. Pour l'ancienne droite des moindres carrés $\mathcal{D}_0$, $e = 0$: la nouvelle somme vaut $S(\mathcal{D}_0)$. Pour toute droite $\mathcal{D}$, $S(\mathcal{D}) + e^2 \geqslant S(\mathcal{D}) \geqslant S(\mathcal{D}_0)$. $\mathcal{D}_0$ rend donc minimale la nouvelle somme et, la droite des moindres carrés étant unique, c'est celle de la nouvelle série. (0,5 pt pour la réponse, 0,5 pt pour l'argument de minimalité) (1)

d) Faux. Avec $E(T) = \dfrac{1}{\lambda}$: $P\big(T > 2E(T)\big) = \mathrm{e}^{-2} \approx 0{,}135$, alors que $\dfrac{1}{2}P\big(T > E(T)\big) = \dfrac{\mathrm{e}^{-1}}{2} \approx 0{,}184$. En fait, $P\big(T > 2E(T)\big) = \mathrm{e}^{-2} = \Big(P\big(T > E(T)\big)\Big)^2$: la probabilité est élevée au carré, pas divisée par 2. (0,5 pt pour la réponse, 0,5 pt pour le calcul) (1)

e) Faux. À la calculatrice, $P(45 \leqslant X \leqslant 55) = P(X \leqslant 55) - P(X \leqslant 44) \approx 0{,}864 - 0{,}136 = 0{,}729 < 0{,}95$. Il faut un intervalle plus large : $P(40 \leqslant X \leqslant 60) \approx 0{,}965$. (0,5 pt pour la réponse, 0,5 pt pour le calcul) (1)

Erreur fréquente — Une affirmation se réfute par un seul contre-exemple chiffré, comme U^2 avec $x = 0{,}25$; elle se démontre en toute généralité : la vérifier sur un cas ne prouve rien.

Exercice 4 – Problème — Combien de temps reste un lecteur ? *(5 pts)*

a) Le pourcentage de lecteurs encore présents au bout de t secondes est $y = 100\,P(T > t) = 100\,\mathrm{e}^{-\lambda t}$, donc $\ln y = \ln 100 - \lambda t$: c'est une fonction affine de t, de coefficient directeur $-\lambda$ et d'ordonnée à l'origine $\ln 100 \approx 4{,}605$. (0,5 pt pour $y = 100\,\mathrm{e}^{-\lambda t}$, 0,5 pt pour la forme affine) (1)

b) $z_i \approx 4{,}431$; $4{,}248$; $4{,}078$; $3{,}892$; $3{,}555$; $2{,}996$. La calculatrice donne $z = -0{,}017\,8\,t + 4{,}610$ et $r \approx -0{,}999\,9$. Les points $(t_i\,;z_i)$ sont presque alignés et l'ordonnée à l'origine est très proche de $\ln 100 \approx 4{,}605$: le modèle exponentiel est adapté, avec $\lambda \approx 0{,}017\,8$. (0,5 pt pour les z_i, 0,5 pt pour la droite, r et la conclusion) (1)

c) $\lambda = 0{,}017\,8$ et $E(T) = \dfrac{1}{0{,}017\,8} \approx 56$ s. La médiane m vérifie $\mathrm{e}^{-\lambda m} = 0{,}5$, soit $m = \dfrac{\ln 2}{\lambda} \approx 39$ s. C'est cohérent avec le tableau : la moitié des lecteurs part entre 30 s (59 % encore présents) et 40 s (49 %). (0,5 pt pour la moyenne, 0,5 pt pour la médiane et sa vérification) (1)

d) Tableau : $\dfrac{20}{59} \approx 0{,}339$. Modèle : $\{T > 90\}$ est inclus dans $\{T > 30\}$, donc $P_{T>30}(T > 90) = \dfrac{\mathrm{e}^{-90\lambda}}{\mathrm{e}^{-30\lambda}} = \mathrm{e}^{-60\lambda} = \mathrm{e}^{-1{,}068} \approx 0{,}344$, c'est-à-dire $P(T > 60)$. Les deux valeurs sont proches : les données sont compatibles avec l'absence de mémoire (un lecteur resté 30 s n'est ni plus ni moins « accroché » qu'un lecteur qui arrive). (0,5 pt pour les deux valeurs, 0,5 pt pour la propriété) (1)

e) Le nombre de lecteurs encore présents à l'instant t suit $\mathcal{B}\big(40\,;\mathrm{e}^{-\lambda t}\big)$ (lecteurs indépendants), donc $P(\text{au moins un}) = 1 - \big(1-\mathrm{e}^{-\lambda t}\big)^{40}$. Or $1 - \big(1-\mathrm{e}^{-\lambda t}\big)^{40} \geqslant 0{,}99 \iff 40\ln\big(1-\mathrm{e}^{-\lambda t}\big) \leqslant \ln 0{,}01$ $\iff 1-\mathrm{e}^{-\lambda t} \leqslant \mathrm{e}^{\frac{\ln 0{,}01}{40}} \approx 0{,}891\,3 \iff \mathrm{e}^{-\lambda t} \geqslant 0{,}108\,7 \iff t \leqslant \dfrac{-\ln 0{,}108\,7}{\lambda} \approx 124{,}6$. Contrôle : pour $t = 124$, la probabilité vaut environ $0{,}990\,5$; pour $t = 125$, environ $0{,}989\,7 < 0{,}99$. La plus grande durée est 124 s. (0,5 pt pour l'inéquation, 0,5 pt pour la valeur contrôlée) (1)

Piège — Le changement de variable $z = \ln y$ porte sur les pourcentages mesurés, pas sur t ; le paramètre λ est l'opposé du coefficient directeur, et la réponse finale revient toujours à la variable d'origine.

Ce document est la propriété intellectuelle de Ben, professeur particulier (benprofparticulier.fr), protégée par le Code de la propriété intellectuelle (art. L. 111-1 et L. 122-4). Il est mis à disposition pour un usage personnel de révision. Toute reproduction, diffusion, mise en ligne ou revente, totale ou partielle, sans autorisation écrite est interdite et constitue une contrefaçon (art. L. 335-2).