

Variables aléatoires.

1. On lance un dé tétraédral dont les faces sont numérotées de 1 à 4 et un dé octaédral dont les faces sont numérotées de 1 à 8. Calculer la loi de la somme S , du produit P et du plus grand M des deux nombres obtenus.

Solution de l'exercice 1. Il y a 32 lancers équiprobables. Il faut compter le nombre de manière d'obtenir chacun des résultats. On obtient :

$$\mathbb{P}(S = 2) = \mathbb{P}(S = 12) = 1/32, \quad \mathbb{P}(S = 3) = \mathbb{P}(S = 11) = 1/16,$$

$$\mathbb{P}(S = 4) = \mathbb{P}(S = 10) = 3/32,$$

$$\mathbb{P}(S = 5) = \mathbb{P}(S = 6) = \mathbb{P}(S = 7) = \mathbb{P}(S = 8) = \mathbb{P}(S = 9) = 1/8.$$

$$\mathbb{P}(P = 1) = \mathbb{P}(P = 5) = \mathbb{P}(P = 7) = \mathbb{P}(P = 9) = \mathbb{P}(P = 10) = \mathbb{P}(P = 14) = \mathbb{P}(P = 15) = \mathbb{P}(P = 18) = \mathbb{P}(P = 21) = \mathbb{P}(P = 28) = \mathbb{P}(P = 32) = 1/32,$$

$$\mathbb{P}(P = 2) = \mathbb{P}(P = 3) = \mathbb{P}(P = 16) = \mathbb{P}(P = 24) = 1/16,$$

$$\mathbb{P}(P = 4) = \mathbb{P}(P = 6) = \mathbb{P}(P = 8) = 3/32, \quad \mathbb{P}(P = 12) = 1/8.$$

$$\mathbb{P}(M = 1) = 1/32, \quad \mathbb{P}(M = 2) = 3/32, \quad \mathbb{P}(M = 3) = 5/32,$$

$$\mathbb{P}(M = 4) = 7/32, \quad \mathbb{P}(M = 5) = \mathbb{P}(M = 6) = \mathbb{P}(M = 7) = \mathbb{P}(M = 8) = 1/8.$$

2. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Déterminer la loi de $-\log U$.

Solution de l'exercice 2. On remarque que $\mathbb{P}(-\log U \leq 0) = 0$. De plus, si $x \geq 0$,

$$\mathbb{P}(-\log U \leq x) = \mathbb{P}(U \geq \exp(-x)) = 1 - \exp(-x).$$

On reconnaît la fonction de répartition d'une variable aléatoire exponentielle de paramètre 1. Comme la fonction de répartition d'une variable aléatoire réelle caractérise la loi, on conclut que $-\log U$ suit une loi exponentielle de paramètre 1.

3. Soit X une variable aléatoire de loi exponentielle de paramètre θ . Soit $c > 0$ un réel. Déterminer la loi de cX .

Solution de l'exercice 3. Pour tout réel x , on a

$$\mathbb{P}(cX \leq x) = \mathbb{P}(X \leq \frac{x}{c}) = \left(1 - \exp\left(-\frac{\theta x}{c}\right)\right) \mathbb{1}_{[0, +\infty[}(x).$$

On reconnaît la fonction de répartition d'une variable aléatoire exponentielle de paramètre $\frac{\theta}{c}$. La fonction de répartition caractérisant la loi, on en déduit que cX suit une loi exponentielle de paramètre $\frac{\theta}{c}$.

4. Soient X, Y, Z trois variables aléatoires à valeurs dans \mathbb{N} . On suppose que X et Y ont même loi. Soit $f : \mathbb{N} \rightarrow \mathbb{N}$ une fonction.

Est-il vrai que $f(X)$ et $f(Y)$ ont même loi ? Est-il vrai que $X + Z$ et $Y + Z$ ont même loi ?

Solution de l'exercice 4. $f(X)$ et $f(Y)$ ont bien la même loi, car

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(f(X) = k) = \mathbb{P}(X \in f^{-1}(\{k\})) = \mathbb{P}(Y \in f^{-1}(\{k\})) = \mathbb{P}(f(Y) = k).$$

En revanche, $X + Z$ et $Y + Z$ peuvent avoir des lois différentes. Par exemple, on peut déterminer les valeurs des 3 variables aléatoires en fonction du résultat d'un même lancer d'une pièce de monnaie équilibrée :

$$\mathbb{P}(X = 0, Y = 1, Z = 2) = 1/2, \quad \mathbb{P}(X = 1, Y = 0, Z = 5) = 1/2.$$

X et Y suivent alors la même loi, mais $\mathbb{P}(X + Z = 2) = 1/2$ tandis que $\mathbb{P}(Y + Z = 2) = 0$.

5. On étudie des variables aléatoires qui ont une propriété d'absence de mémoire.

a) Soit T une variable aléatoire à valeurs dans \mathbb{N} . On suppose que pour tous $n, m \geq 0$ entiers, on a

$$\mathbb{P}(T \geq n + m) = \mathbb{P}(T \geq n)\mathbb{P}(T \geq m).$$

Que peut-on dire de la loi de T ?

b) Soit S une variable aléatoire à valeurs dans \mathbb{R} . On suppose que pour tous $a, b \geq 0$ réels, on a

$$\mathbb{P}(S > a + b) = \mathbb{P}(S > a)\mathbb{P}(S > b).$$

Que peut-on dire de la loi de S ?

Solution de l'exercice 5.

a) Par récurrence sur n , on voit immédiatement que $\mathbb{P}(T \geq n) = \mathbb{P}(T \geq 1)^n$ pour tout entier $n \geq 0$. On en déduit que pour tout $n \geq 0$, $\mathbb{P}(T = n) = \mathbb{P}(T \geq n) - \mathbb{P}(T \geq n + 1) = \mathbb{P}(T \geq 1)^n(1 - \mathbb{P}(T \geq 1))$. Ainsi T suit une loi géométrique de paramètre $\mathbb{P}(T \geq 1)$.

b) On remarque que $t := \mathbb{P}(S > 0)$ vérifie $t = t^2$. La première possibilité est que $t = 0$, auquel cas S est une variable aléatoire négative, et on ne peut rien dire de plus car toutes les variables aléatoires telles que $\mathbb{P}(S > 0) = 0$ conviennent.

La seconde possibilité est que $t = 1$. Dans ce cas S est une variable aléatoire strictement positive avec probabilité 1. On introduit $L(x) := \log \mathbb{P}(S > x)$. Par continuité à droite de la fonction de répartition, on sait que pour $x > 0$ assez petit, $\mathbb{P}(S > x) > 0$. On montre alors par récurrence que pour tout $n \in \mathbb{N}$, $L(nx) = nL(x)$. Si $m \geq 1$ est aussi un entier, on a par la même égalité (pour $x' = x/m$, en remarquant que $\mathbb{P}(S > x') > 0$ équivaut à $\mathbb{P}(S > x) > 0$), $L(x/m) = L(x)m$. En combinant les deux, il vient $L(nx/m) = L(x)n/m$.

Autrement dit, pour tout rationnel r positif, $L(rx) = rL(x)$. Ce qui par continuité à droite de L entraîne que la même égalité est vraie pour tout r réel strictement positif. On en déduit que $\mathbb{P}(S > rx) = \mathbb{P}(S > x)^r$. En prenant $x=1$, on constate que S est une variable aléatoire exponentielle de paramètre $\theta = -\log \mathbb{P}(S > 1)$.

6. Un chimpanzé tape à la machine à écrire en appuyant chaque seconde sur une touche choisie au hasard. Quelle est la probabilité qu'il parvienne à écrire *Hamlet*, c'est-à-dire qu'à un certain moment il écrive d'une traite le texte de cette pièce ?

Solution de l'exercice 6. Soit n la longueur, en caractères, de la pièce *Hamlet*. La probabilité p qu'il tape la pièce du premier coup est faible, mais strictement positive. Pour tout entier naturel k , on définit une variable aléatoire X_k qui vaut 1 si les caractères $nk + 1$ à $n(k + 1)$ correspondent au texte de la pièce. Les X_k sont les mêmes que dans un jeu de pile ou face biaisé. Or on sait dans ce cas qu'il finira par sortir un pile, ce qui correspond ici à écrire le texte de la pièce d'une traite.

La probabilité que le chimpanzé écrive *Hamlet* au bout d'un certain nombre (aléatoire, mais fini) de tentatives vaut donc 1. Néanmoins, en pratique, il y a fort à parier que le chimpanzé (ou la machine à écrire) arrivera à épuisement bien avant.

*On rappelle que la mesure de Lebesgue, notée **Leb**, est l'unique mesure borélienne sur \mathbb{R} qui, pour tous $a, b \in \mathbb{R}$ tels que $a < b$, associe à l'intervalle ouvert $]a, b[$ la mesure $b - a$.*

Pour tout $x \in \mathbb{R}$, on appelle partie entière de x et on note $\lfloor x \rfloor$ l'unique entier relatif tel que $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$. On note $\{x\} = x - \lfloor x \rfloor$ la partie fractionnaire de x .

7. On considère l'espace de probabilités $([0, 1[, \mathcal{B}_{[0,1]}, \mathbf{Leb})$. Pour tout $n \geq 1$, on définit une variable X_n à valeurs réelles en posant

$$\forall x \in [0, 1[, X_n(x) = \lfloor 2\{2^{n-1}x\} \rfloor.$$

- a) Représenter le graphe de X_1, X_2, X_3 vues comme fonctions de $[0, 1[$ dans \mathbb{R} .
- b) Déterminer la loi de X_n pour tout $n \geq 1$.
- c) Soit $n \geq 1$ un entier. Soient $\varepsilon_1, \dots, \varepsilon_n \in \{0, 1\}$. On pose $a = \sum_{k=1}^n 2^{-k} \varepsilon_k$ et $b = a + 2^{-n}$. Montrer que

$$\{x \in [0, 1[: X_1(x) = \varepsilon_1, \dots, X_n(x) = \varepsilon_n\} = [a, b[.$$

En déduire la loi de la variable aléatoire (X_1, \dots, X_n) à valeurs dans $\{0, 1\}^n$.

La suite des variables aléatoires $(X_n)_{n \geq 1}$, définie sur l'espace de probabilités $([0, 1[, \mathcal{B}_{[0,1]}, \mathbf{Leb})$, constitue donc un jeu de pile ou face infini.

- d) Montrer que la plus petite tribu sur $[0, 1[$ qui contienne tous les ensembles $\{x \in [0, 1[: X_n(x) = \varepsilon\}$ pour $n \geq 1$ et $\varepsilon \in \{0, 1\}$ est la tribu borélienne.

Solution de l'exercice 7.

a) On remarque que pour tout $x \in [0, 1[$, $\lfloor 2x \rfloor = \mathbb{1}_{[1/2, 1[}(x)$, ce qui permet de mieux comprendre la définition des X_n .

$$X_1 = \mathbb{1}_{[1/2, 1[}, \quad X_2 = \mathbb{1}_{[1/4, 1/2[\cup [3/4, 1[}, \quad X_3 = \mathbb{1}_{[1/8, 1/4[\cup [3/8, 1/2[\cup [5/8, 3/4[\cup [7/8, 1[}.$$

b) On va montrer que pour tout $n \geq 1$, $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = 0) = 1/2$.

On fixe $n \geq 1$. On partage $[0, 1[$ en 2^n intervalles $I_k = [k2^{-n}, (k+1)2^{-n}[$, $k = 0, \dots, 2^n - 1$. Soit $x \in [0, 1[$. Il existe un unique k tel que $x \in I_k$. On distingue deux cas :

Si $k = 2i$ est pair, alors $2^{n-1}x \in [i, i + \frac{1}{2}[$, $\{2^{n-1}x\} \in [0, \frac{1}{2}[$ and $X_n(x) = 0$.

Si $k = 2i + 1$ est impair, alors $2^{n-1}x \in [i + \frac{1}{2}, i + 1[$, $\{2^{n-1}x\} \in [\frac{1}{2}, 1[$ and $X_n(x) = 1$.

Comme il y a exactement 2^{n-1} intervalles I_k avec k pair et autant avec k impair, et que chaque I_k a pour mesure de Lebesgue 2^{-n} , la disjonction ci-dessus donne le résultat annoncé.

c) Soit $x \in [a, b[$ et k un entier tel que $1 \leq k \leq n$.

Alors $2^{k-1}x \in [2^{k-1}a, 2^{k-1}b[$. Or

$$(2^{k-1}a = 2^{k-2}\varepsilon_1 + \dots + 2^1\varepsilon_{k-2} + \varepsilon_{k-1}) + 2^{-1}\varepsilon_k + (\dots + 2^{k-1-n}\varepsilon_n) =: E_k^n(x) + 2^{-1}\varepsilon_k + R_k^n(x)$$

où, avec des notations évidentes, $E_k^n(a)$ est entier et $0 \leq R_k^n(a) \leq \frac{1}{2} - 2^{k-1-n}$. De plus, on a la même écriture pour b avec $E_k^n(b) = E_k^n(a)$ mais $2^{k-1-n} \leq R_k^n(b) \leq \frac{1}{2}$.

Comme $a \leq x < b$, on en déduit que $\{2^{k-1}x\} \in [\varepsilon_k/2, (1 + \varepsilon_k)/2[$, et donc $X_k(x) = \varepsilon_k$.

On a montré que pour chaque choix de $\varepsilon_1, \dots, \varepsilon_n$, et pour a et b qui en résultent comme dans l'énoncé, l'intervalle $[a, b[$ est inclus dans $\{x \in [0, 1[: X_1(x) = \varepsilon_1, \dots, X_n(x) = \varepsilon_n\}$. Montrons maintenant par l'absurde que ces inclusions sont en fait des égalités. Supposons que ce soit faux, on dispose alors de $\varepsilon_1, \dots, \varepsilon_n$ et du $[a, b[$ associé par l'énoncé, et d'un $x \in [0, 1[$ tel que $X_1(x) = \varepsilon_1, \dots, X_n(x) = \varepsilon_n\}$ serait inclus dans un autre intervalle dyadique $[a', b'[$, avec $a' = \sum_{k=1}^n 2^{-k}\varepsilon'_k$. Avec $(\varepsilon_1, \dots, \varepsilon_n) \neq (\varepsilon'_1, \dots, \varepsilon'_n)$. Mais d'après ce qu'on vient de voir, on devrait avoir $[a', b'[\subset \{x \in [0, 1[: X_1(x) = \varepsilon_1, \dots, X_n(x) = \varepsilon_n\}$, et x fournit une contradiction. Ce qui prouve l'égalité demandée : pour tout choix de $\varepsilon_1, \dots, \varepsilon_n$, en notant $a = \sum_{k=1}^n 2^{-k}\varepsilon_k$ et $b = a + 2^{-n}$, on a

$$\{x \in [0, 1[: X_1(x) = \varepsilon_1, \dots, X_n(x) = \varepsilon_n\} = [a, b[.$$

En prenant la mesure de Lebesgue de l'ensemble ci-dessus, on obtient la loi jointe des X_k :

$$\mathbb{P}(X_1 = \varepsilon_1, \dots, X_n = \varepsilon_n) = 2^{-n}.$$

Il s'agit bien de la loi d'un jeu de pile ou face, après n lancers. Comme la suite de variables aléatoires $(X_k)_{k \geq 1}$ est bien définie sur l'espace de probabilité $([0, 1[, \mathcal{B}_{[0, 1[}, \text{Leb})$, on a construit un jeu de pile ou face infini, et ainsi démontré le théorème 2.1.1 du polycopié de cours.

d) On sait déjà que la tribu engendrée par les $\{x \in [0, 1[: X_n(x) = \varepsilon\}$ est incluse dans $\mathcal{B}_{[0, 1[}$ puisque ces ensembles sont boréliens (on a vu au b) qu'il s'agissait de réunions

finies d'intervalles). Il reste à montrer l'autre inclusion. Par le c), il suffit de vérifier que les intervalles dyadiques engendrent la tribu borélienne. Il suffit même de montrer que la tribu qu'ils engendrent contiennent tous les intervalles $[0, c[$, $c \in]0, 1[$ (puisque ceux-ci engendrent la tribu borélienne). C'est bien vrai, puisqu'un intervalle $[0, c[$, $c \in]0, 1[$, est la réunion des intervalles dyadiques élémentaires (autrement dit, ceux considérés au c)) qu'il contient. Il s'agit essentiellement de démontrer que si $c \in]0, 1[$, et si $x \in [0, c[$, alors on peut trouver n , et $\varepsilon_1, \dots, \varepsilon_n$, tels que, en notant $a = \sum_{k=1}^n 2^{-k} \varepsilon_k$ et $b = a + 2^{-n}$, on ait

$$a \leq x < b \leq c.$$

Pour cela, on choisit n tel que $2^{-n} = b - a < c - x$. Ceci entraîne que dès qu'on aura $a \leq x$, on aura aussi, automatiquement, $b \leq c$. Il ne reste alors plus qu'à prendre $\varepsilon_1 = X_1(x), \dots, \varepsilon_n = X_n(x)$, et $x \in [a, b[$ a été démontré au c).