Chapitre 7

Espérance

7.1 Introduction

L'ESPÉRANCE d'une variable aléatoire est, lorsqu'elle existe, la *moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation.* On voit bien comment traduire cette définition informelle dans le cas d'une variable aléatoire discrète X en posant :

$$\mathbf{E}X := \sum_{x \in X(\Omega)} xP(X = x). \tag{7.1}$$

Cette formule n'a de sens que si la famille de réels $\{xP(X = x)\ ;\ x \in X(\Omega)\}$ est sommable, ce qui se traduit par la condition suivante pour l'existence de l'espérance de la v.a. discrète X :

$$\sum_{x \in X(\Omega)} |x|P(X = x) < +\infty. \tag{7.2}$$

Tant que l'on reste dans le cadre des variables aléatoires discrètes, cette définition est satisfaisante et permet d'établir toutes les propriétés de l'espérance [14, Chap. 5]. En bonne place parmi ces propriétés, figure *l'additivité* de l'espérance : si X et Y définies sur le même $(\Omega, \mathcal{F}, P)$ ont une espérance, il en va de même pour $X + Y$ et

$$\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y. \tag{7.3}$$

Essayons de traduire la définition informelle ci-dessus dans le cas d'une variable aléatoire à densité f. Partant de (7.1), on remplace $P(X = x)$ par $P(X \in [x, x + \mathrm{d}x])$, probabilité « valant [1] $f(x)\,\mathrm{d}x$ » et on remplace la somme (ou série) par une intégrale, ce qui conduit à :

$$\mathbf{E}X := \int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x, \tag{7.4}$$

1. Nous ne prétendons pas donner un sens rigoureux à cette probabilité d'appartenance à un « intervalle infinitésimal », il s'agit juste d'une approche intuitive.

la condition d'existence de l'espérance étant tout simplement la convergence absolue de cette intégrale généralisée, ce qui vu la positivité de f, se traduit par

$$\int_{-\infty}^{+\infty} |x| f(x) \,\mathrm{d}x < +\infty. \tag{7.5}$$

Cette définition malgré son analogie formelle avec (7.1) est loin d'offrir la même souplesse pour établir les propriétés de l'espérance. Par exemple la preuve de l'additivité est *complètement hors de portée.* En effet, si X et Y sont à densité, $X + Y$ peut n'être ni discrète ni à densité[2], cf. l'exercice 6.13 pour un exemple, et alors le premier membre de (7.3) n'est même pas défini pour la v.a. $Z = X + Y$.

La solution donnée à ce problème par la théorie moderne des probabilités est la définition dans le cas général, de l'espérance de X comme une intégrale abstraite sur Ω, relativement à la mesure P :

$$\mathbf{E}X := \int_\Omega X(\omega) \,\mathrm{d}P(\omega), \quad \text{si} \int_\Omega |X(\omega)| \,\mathrm{d}P(\omega) < +\infty. \tag{7.6}$$

On peut donner une première idée de ce qu'est cette intégrale abstraite en considérant le cas d'une variable aléatoire X telle que $X(\Omega) = \{x_1, \dots, x_n\}$. Alors en notant $A_k := \{X = x_k\} = \{\omega \in \Omega \,;\, X(\omega) = x_k\}$,

$$\int_\Omega X(\omega) \,\mathrm{d}P(\omega) = \sum_{k=1}^{n} x_k P(A_k), \tag{7.7}$$

ce qui traduit bien la définition informelle de $\mathbf{E}X$ comme la moyenne des valeurs de X pondérées par leurs probabilités de réalisation. Le passage au cas d'une variable aléatoire X quelconque revient précisément à construire une intégrale au sens de Lebesgue sur $(\Omega, \mathcal{F}, P)$ et cette théorie sort du cadre de ce livre.

Il nous faut donc trouver une autre définition de $\mathbf{E}X$. Cette définition doit permettre un traitement unifié de toutes les lois[3]. Rappelons qu'il existe des lois qui ne sont ni discrètes ni à densité et que la description la plus générale des lois de variables aléatoires réelles est donnée par leur fonction de répartition, cf. le théorème 5.30 et la remarque 6.17. Il est donc naturel de chercher à définir $\mathbf{E}X$ *à partir de la fonction de répartition* $F : t \mapsto P(X \leqslant t)$. Nous allons motiver cette définition en nous restreignant au cas des variables aléatoires positives et en partant du cas simple où X est discrète avec $X(\Omega) = \{x_1, \dots, x_n\}$ partie *finie* de $\mathbb{R}_+$. Dans ce cas, la définition informelle de $\mathbf{E}X$ se traduit par la formule $\mathbf{E}X = \sum_{k=1}^{n} x_k P(X = x_k)$. Les figures 7.1 et 7.2 nous montrent comment exprimer cette moyenne pondérée à l'aide de F. Rappelons que dans ce cas, F présente en chaque x_k un saut d'amplitude $P(X = x_k)$. L'interprétation graphique en terme d'aires donnée par la figure 7.2 nous permet d'écrire $\mathbf{E}X$ comme l'intégrale de Riemann ordinaire : $\mathbf{E}X = \int_0^{x_n} (1 - F(t)) \,\mathrm{d}t$ et aussi comme la fausse intégrale généralisée $\int_0^{+\infty} (1 - F(t)) \,\mathrm{d}t$.

2. Alors que la somme de deux variables aléatoires discrètes est *toujours* une variable aléatoire discrète.

3. La définition informelle de $\mathbf{E}X$ nous fait pressentir que $\mathbf{E}X$ ne doit dépendre que de la loi de X, ce qui est bien le cas dans les formules (7.1) et (7.4).

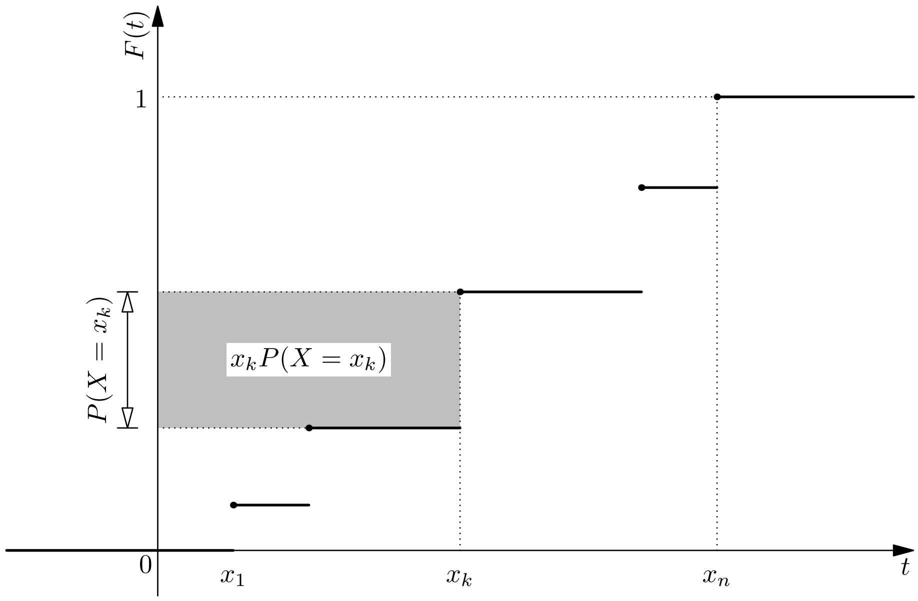

FIGURE 7.1 – Interprétation graphique des $x_k P(X = x_k)$, pour $x_k \geqslant 0$

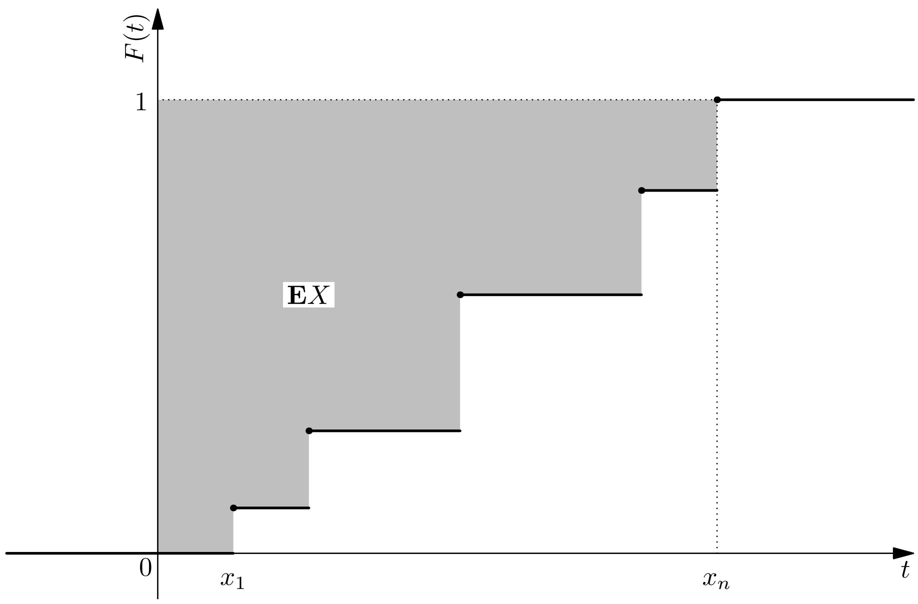

FIGURE 7.2 – Interprétation graphique de $\mathbf{E}X = \sum_{k=1}^{n} x_k P(X = x_k)$, les $x_k \geqslant 0$.

Si on passe maintenant au cas d'une variable aléatoire positive quelconque, il paraît alors naturel de considérer que $\mathbf{E}X$ est l'aire (éventuellement infinie) délimitée par le segment vertical $t = 0$, $y \in [0, 1]$, la demi droite « asymptote » $y = 1$, $t \geqslant 0$ et le graphe de F, ce qui nous conduit à la formule

$$\mathbf{E}X := \int_0^{+\infty} (1 - F(t))\,\mathrm{d}t = \int_0^{+\infty} P(X > t)\,\mathrm{d}t, \quad \text{pour toute v.a. positive } X.$$

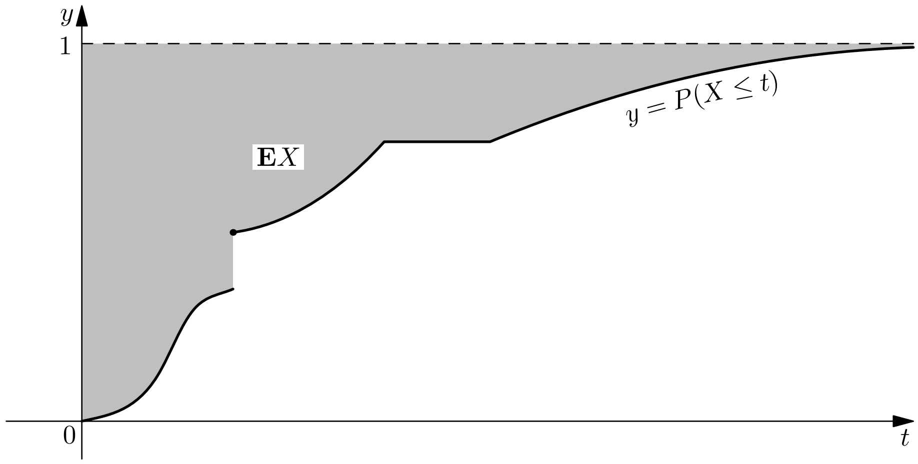

FIGURE 7.3 – Interprétation graphique de $\mathbf{E}X$ *via* la f.d.r. de X v.a. positive.

Nous verrons que cette définition permet d'établir en toute généralité les propriétés de l'espérance. Bien sûr, nous devrons retrouver à partir de cette définition, les formules (7.1) et (7.4) pour X discrète ou à densité.

7.2 Espérance d'une variable aléatoire positive

Dans toute la suite de ce chapitre, on fixe un espace probabilisé $(\Omega, \mathcal{F}, P)$. Toutes les variables aléatoires considérées seront, sauf mention explicite du contraire, définies sur cet espace et leur loi sera la loi sous P.

Définition 7.1 (espérance d'une v.a. positive)**.** *Soit X une variable aléatoire positive*[4] *sur $(\Omega, \mathcal{F})$. On appelle* espérance *de X (ou espérance de X sous P) la quantité*

$$\mathbf{E}X := \int_0^{+\infty} P(X > t)\,\mathrm{d}t, \tag{7.8}$$

qui est un élément de $\overline{\mathbb{R}}_+$.

4. C'est-à-dire une application $\Omega \to \mathbb{R}_+$, mesurable $\mathcal{F}$ - $\mathrm{Bor}(\mathbb{R}_+)$.

Pour justifier l'existence de $\mathbf{E}X$, on commence par noter que l'application $G : \mathbb{R}_+ \to [0,1]$, $t \mapsto G(t) := P(X > t)$ est *décroissante* sur $\mathbb{R}_+$, donc *Riemann intégrable* sur $[0,b]$ pour tout $b \in \mathbb{R}_+^*$, cf. proposition 3.10. L'intégrale $\int_0^b G(t)\,\mathrm{d}t = \int_0^b P(X > t)\,\mathrm{d}t$ existe donc bien et est un réel positif pour tout b. Comme c'est une fonction croissante de sa borne supérieure b, elle converge dans $\overline{\mathbb{R}}_+$ quand b tend vers $+\infty$.

Dans cette section, nous utiliserons l'interprétation graphique de $\mathbf{E}X$ *via* la fonction de survie $t \mapsto P(X > t)$, cf. figure 7.4, plutôt que *via* la f.d.r. $F : t \mapsto P(X \leqslant t)$. On passe évidemment d'une représentation à l'autre en effectuant une symétrie orthogonale par rapport à la droite $y = 1/2$, puisque $G = 1 - F$. Cette symétrie conserve les aires, cf. prop. 5.14.

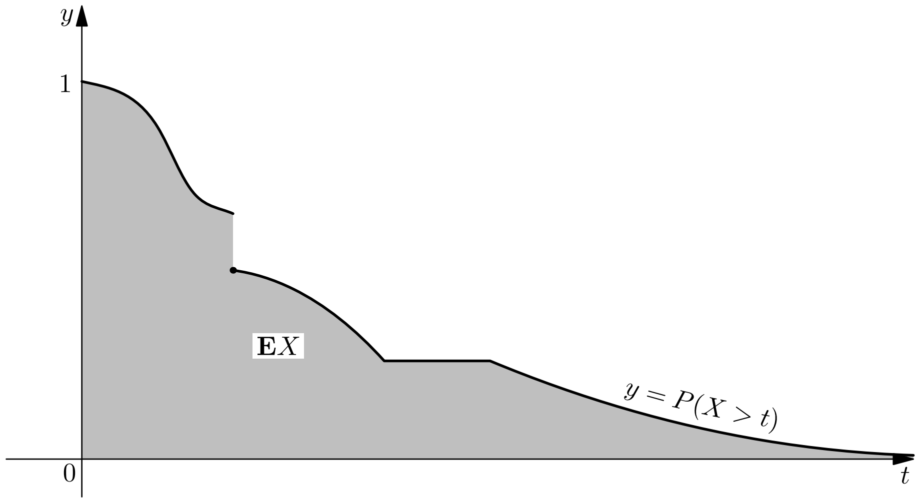

FIGURE 7.4 – Interprétation graphique de $\mathbf{E}X$ *via* la fonction de survie de X v.a. positive.

Remarque 7.2. $\mathbf{E}X$ ne dépend que de la loi de X, il serait donc plus correct de parler de l'espérance de la loi de X sous P au lieu de l'espérance de X. L'usage donne néanmoins la préférence à cette dernière appellation *quand il n'y a pas d'ambiguïté sur P.*

Remarque 7.3 (espérance d'une v.a. presque sûrement positive)**.** Dans les exercices, la variable aléatoire X n'est pas toujours donnée explicitement, il arrive assez souvent que l'on ne connaisse que sa loi P_X. Si $P_X(\mathbb{R}_+) = 1$, on s'autorisera une généralisation de la définition 7.1 en considérant que la formule (7.8) reste valable. Il s'agit bien d'une généralisation car on peut avoir $P(X \geqslant 0) = 1$ sans que $X(\omega)$ soit positif ou nul pour tout ω, par exemple si $\Omega = \mathbb{R}$, $\mathcal{F} = \mathrm{Bor}(\mathbb{R})$, P est la loi uniforme sur $[0,1]$ et $X : \omega \mapsto \omega$ est l'identité sur $\mathbb{R}$. On a alors $X(\omega) < 0$ pour une infinité non dénombrable de ω et $P(X \geqslant 0) = 1$. Cette généralisation est cohérente avec la remarque 7.2 car si X est une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ et telle que $P(X \geqslant 0) = 1$, on peut toujours trouver un espace probabilisé $(\Omega', \mathcal{F}', P')$ et une variable aléatoire positive X' définie sur cet espace tels que X et X' aient même loi. Il suffit de prendre $\Omega' = \mathbb{R}_+$, $\mathcal{F}' = \mathrm{Bor}(\mathbb{R}_+)$, $P' = P_X$ et $X' : \mathbb{R}_+ \to \mathbb{R}$, $\omega \mapsto \omega$ égale à l'identité sur $\mathbb{R}_+$. On a alors pour tout borélien B de $\mathbb{R}$, $P'_{X'}(B) = P'(X' \in B) = P'(\{\omega \in \Omega' ; X'(\omega) \in B\} = P'(B \cap \mathbb{R}_+) = P_X(B \cap \mathbb{R}_+) = P_X(B)$ car $P_X(B \cap \mathbb{R}_-^*) \leqslant P_X(\mathbb{R}_-^*) = 0$. Ceci montre que X et X' ont même loi [5].

Définition 7.4 (intégrabilité d'une v.a. positive)**.** *On dit que la variable aléatoire positive X est* intégrable *si*

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t < +\infty. \tag{7.9}$$

Exemple 7.5. Si la variable aléatoire positive X est bornée, c.-à-d. s'il existe une constante c telle que pour tout $\omega \in \Omega$, $0 \leqslant X(\omega) \leqslant c$, alors elle est intégrable. En effet pour $t \geqslant c$, $P(X > t) = 0$, ce qui réduit l'intégrale généralisée définissant $\mathbf{E}X$ à une intégrale de Riemann ordinaire $\int_0^c P(X > t)\,\mathrm{d}t$ donc finie (et majorée par c).

Plus généralement, si la loi de X, v.a. positive, vérifie $P(X > t) \leqslant Ct^{-\alpha}$ pour un certain $\alpha > 1$ et tout $t \geqslant t_0 > 0$, ou si $P(X > t) \leqslant t^{-1}(\ln t)^{-\beta}$ pour un $\beta > 1$ et tout $t \geqslant t_0 > 0$, alors X est intégrable. Réciproquement, l'intégrabilité de X nous donne un renseignement sur la vitesse de convergence [6] vers 0 de $P(X > t)$ quand t tend vers $+\infty$. C'est l'inégalité de Markov que nous verrons ci-dessous (proposition 7.16).

Voyons maintenant quelques exemples simples de calcul d'espérance de variables aléatoires positives.

Exemple 7.6 (espérance d'une constante positive)**.** Si la variable aléatoire X est une constante positive c, c.-à-d. $X(\omega) = c$ pour tout $\omega \in \Omega$, alors $\mathbf{E}X = c$. En effet :

$$P(X > t) = \begin{cases} 1 & \text{si } t < c \\ 0 & \text{si } t \geqslant c \end{cases} = \mathbf{1}_{]-\infty, c[}(t),$$

d'où

$$\mathbf{E}X = \int_0^{+\infty} \mathbf{1}_{]-\infty, c[}(t)\,\mathrm{d}t = \int_0^c \mathbf{1}_{[0, c[}(t)\,\mathrm{d}t = \int_0^c 1\,\mathrm{d}t = c.$$

L'exemple suivant est d'une grande importance car il permet d'écrire toute probabilité d'évènement comme une espérance. Nous le formulons sous forme de proposition.

Proposition 7.7 (espérance d'une indicatrice d'évènement)**.** *Pour tout évènement $A \in \mathcal{F}$,*

$$\mathbf{E}(\mathbf{1}_A) = P(A). \tag{7.10}$$

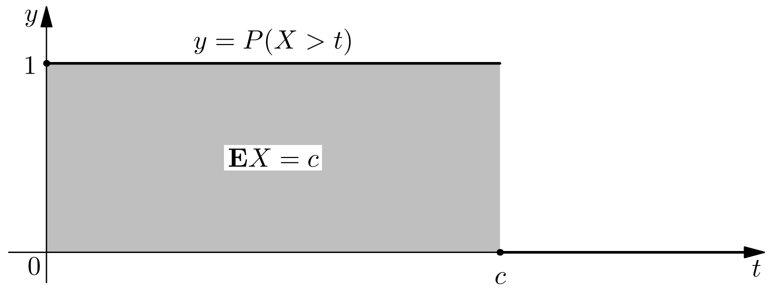

FIGURE 7.5 – Espérance de la v.a. constante $X = c$

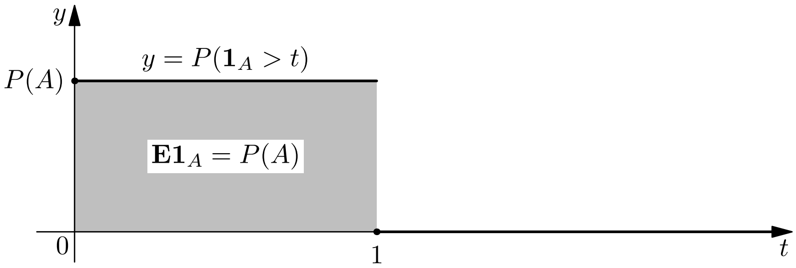

FIGURE 7.6 – Espérance de la v.a. indicatrice $X = \mathbf{1}_A$

Preuve. La variable aléatoire positive $\mathbf{1}_A$ prend la valeur 1 sur l'évènement A et 0 sur A^c, elle suit la loi de Bernoulli de paramètre $p = P(A)$. L'évènement $\{\mathbf{1}_A > t\}$ est donc égal à A si $0 \leqslant t < 1$ et à l'ensemble vide si $t \geqslant 1$. On en déduit que

$$P(\mathbf{1}_A > t) = \begin{cases} P(A) & \text{si } 0 \leqslant t < 1, \\ 0 & \text{si } t \geqslant 1. \end{cases}$$

Par conséquent,

$$\mathbf{E}\big(\mathbf{1}_A\big) = \int_0^1 P(A)\,\mathrm{d}t = P(A).$$

□

Dans ce chapitre, les variables aléatoires discrètes X ne prenant qu'un nombre fini de valeurs jouent un rôle important car elles vont nous permettre d'établir par passage à la limite les principales propriétés de l'espérance. Il est commode de les dénommer comme suit.

5. La tribu borélienne de $\mathbb{R}_+$ est la plus petite tribu contenant tous les ouverts de $\mathbb{R}_+$. On peut vérifier qu'un sous-ensemble B' de $\mathbb{R}_+$ est un borélien de $\mathbb{R}_+$ si et seulement s'il s'écrit $B \cap \mathbb{R}_+$ où B est un borélien de $\mathbb{R}$.

6. Pour n'importe quelle variable aléatoire X, $P(X > t)$ tend vers 0 quand t tend vers $+\infty$, car $P(X > t) = 1 - F(t)$, où F est la f.d.r. de X qui tend toujours vers 1 en $+\infty$.

Définition 7.8 (variable aléatoire simple)**.** *On dit que la variable aléatoire réelle X définie sur $(\Omega, \mathcal{F})$ est* simple *ou* étagée *si $X(\Omega)$ est fini. En notant $X(\Omega) = \{x_1, \dots, x_n\}$, les x_i étant tous distincts, X admet la décomposition*

$$X = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}, \quad \text{où } A_k := \{X = x_k\},\ k \in [\![1, n]\!], \tag{7.11}$$

les évènements A_k formant une partition de Ω.

Proposition 7.9 (espérance d'une v.a. positive simple)**.** *Si X est une variable aléatoire positive simple avec $X(\Omega) = \{x_1, \dots, x_n\}$,*

$$\mathbf{E}X = \sum_{k=1}^{n} x_k P(X = x_k). \tag{7.12}$$

On retrouve ainsi la formule (7.1) de l'introduction dans le cas particulier où $X(\Omega)$ est fini ; voir aussi (7.7).

Preuve. Notons en préliminaire qu'il nous faut résister ici à la tentation de dire « c'est immédiat en utilisant la décomposition (7.11), la proposition 7.7 et la linéarité de l'espérance », car *nous n'avons pas encore prouvé* que l'espérance est linéaire. En fait la proposition 7.9 est l'un des ingrédients de la preuve de la linéarité de l'espérance. Il nous faut donc vérifier (7.12) par un *calcul direct* basé sur la définition 7.1.

Quitte à réindexer, on peut toujours supposer que les x_k sont rangés par ordre croissant. Notons $p_i := P(X = x_i)$ et $s_k := \sum_{1 \leqslant i \leqslant k} p_i$. La fonction de répartition F peut alors s'écrire (cf. la proposition 6.20) :

$$F(t) = \sum_{k=1}^{n-1} s_k \mathbf{1}_{[x_k, x_{k+1}[}(t) + s_n \mathbf{1}_{[x_n, +\infty[}(t). \tag{7.13}$$

Notons que pour $t \geqslant x_n$, $F(t) = s_n = 1$, donc $P(X > t) = 1 - F(t) = 0$. Ainsi $\mathbf{E}X = \int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{x_n} P(X > t)\,\mathrm{d}t$. On peut alors calculer $\mathbf{E}X$ en utilisant la décomposition (7.13) et les propriétés de l'intégrale de Riemann sur l'intervalle fermé borné $[0, x_n]$:

$$\mathbf{E}X = \int_0^{x_n} (1 - F(t))\,\mathrm{d}t = x_n - \int_0^{x_n} F(t)\,\mathrm{d}t = x_n - \sum_{k=1}^{n-1} \int_{x_k}^{x_{k+1}} s_k\,\mathrm{d}t,$$

d'où

$$\begin{aligned}
\mathbf{E}X = x_n - \sum_{k=1}^{n-1} (x_{k+1} - x_k) s_k &= x_n - \sum_{j=2}^{n} x_j s_{j-1} + \sum_{j=1}^{n-1} x_j s_j \\
&= x_n - x_n s_{n-1} + \sum_{j=2}^{n-1} x_j (s_j - s_{j-1}) + x_1 s_1 \\
&= x_n p_n + \sum_{j=2}^{n-1} x_j p_j + x_1 p_1 = \sum_{j=1}^{n} p_k x_k.
\end{aligned}$$

Ceci établit (7.12). □

Proposition 7.10 (espérance d'une v.a. positive à densité)**.** *Si la variable aléatoire positive X a pour densité f,*

$$\mathbf{E}X = \int_0^{+\infty} xf(x)\,\mathrm{d}x. \tag{7.14}$$

Dans cette formule, $\mathbf{E}X$ *peut prendre la valeur* $+\infty$ *si l'intégrale généralisée diverge.*

Preuve. Si X admet pour densité f, $P(X > t) = \int_t^{+\infty} f(x)\,\mathrm{d}x$ pour tout t. En reportant cette égalité dans la définition de $\mathbf{E}X$, on obtient :

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{+\infty} \left\{ \int_t^{+\infty} f(x)\,\mathrm{d}x \right\} \mathrm{d}t = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x)\mathbf{1}_{[t,+\infty[}(x)\,\mathrm{d}x \right\} \mathrm{d}t.$$

Notons que pour $t \geqslant 0$, $\mathbf{1}_{[t,+\infty[}(x) = \mathbf{1}_{[0,x]}(t)$. L'intégrande $(x,t) \mapsto \mathbf{1}_{[0,x]}(t)f(x)$ étant *positive*, le théorème de Fubini-Tonelli légitime l'interversion des intégrations [7], ce qui donne :

$$\mathbf{E}X = \int_0^{+\infty} \left\{ \int_0^{+\infty} f(x)\mathbf{1}_{[0,x]}(t)\,\mathrm{d}t \right\} \mathrm{d}x = \int_0^{+\infty} f(x) \left\{ \int_0^{+\infty} \mathbf{1}_{[0,x]}(t)\,\mathrm{d}t \right\} \mathrm{d}x.$$

Comme pour $x \geqslant 0$, $\int_0^{+\infty} \mathbf{1}_{[0,x]}(t)\,\mathrm{d}t = \int_0^x \mathrm{d}t = x$, on en déduit (7.14). □

Remarque 7.11. Notons que dans la démonstration ci-dessus, nous n'avons utilisé à aucun moment la positivité de la variable aléatoire X. On peut donc appliquer ce calcul à toute variable aléatoire réelle X ayant une densité f pour obtenir :

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{+\infty} xf(x)\,\mathrm{d}x \quad (\text{égalité dans } \overline{\mathbb{R}}_+). \tag{7.15}$$

Attention à ne pas écrire $\mathbf{E}X$ au premier membre de (7.15), cette quantité n'étant pour l'instant définie que pour X *positive.* La vraie formule pour $\mathbf{E}X$ lorsque la v.a. réelle X est à densité est donnée à la proposition 7.32.

Proposition 7.12. *Si X est une variable aléatoire positive et c une constante réelle strictement positive, on a*

$$\mathbf{E}(cX) = c\mathbf{E}X.$$

Cette égalité reste vraie pour $c = 0$ si X est de plus intégrable.

Preuve. Puisque X est une variable aléatoire positive et c une constante positive, $cX : \omega \mapsto (cX)(\omega) := cX(\omega)$ est une variable aléatoire positive. En lui appliquant la définition 7.1, on obtient :

$$\mathbf{E}(cX) = \int_0^{+\infty} P(cX > t)\,\mathrm{d}t = \int_0^{+\infty} P\Big(X > \frac{t}{c}\Big)\,\mathrm{d}t.$$

7. Même si les intégrales valent $+\infty$. Pour une preuve ne reposant pas sur un théorème admis, voir l'exercice 4.11.

Dans cette intégrale généralisée d'une fonction *positive* localement intégrable sur $[0,+\infty[$, on peut effectuer le changement de variable $s = t/c$, cf. proposition 4.41-ii), qui nous donne :

$$\mathbf{E}(cX) = \int_0^{+\infty} P\Big(X > \frac{t}{c}\Big)\,\mathrm{d}t = c\int_0^{+\infty} P(X > s)\,\mathrm{d}s = c\mathbf{E}X.$$

Dans le cas particulier $c = 0$, cette méthode n'est plus valable (on ne peut déjà plus écrire « $P(cX > t) = P(X > t/c)$ ») mais la formule est vraie trivialement, à condition que $\mathbf{E}X$ soit *finie*, puisqu'alors $\mathbf{E}(0 \times X) = \mathbf{E}(0) = 0$ et $0 \times \mathbf{E}X = 0$. □

Proposition 7.13 (croissance de l'espérance)**.** *Si X et Y sont deux variables aléatoires positives définies sur le même $(\Omega, \mathcal{F}, P)$ et si $X \leqslant Y$ c.-à-d. $X(\omega) \leqslant Y(\omega)$ pour tout $\omega \in \Omega$, alors $\mathbf{E}X \leqslant \mathbf{E}Y$.*

Preuve. Si $X(\omega) > t$, alors comme $Y(\omega) \geqslant X(\omega)$, on a aussi $Y(\omega) > t$. Ceci justifie l'inclusion d'évènements $\{X > t\} \subset \{Y > t\}$, puis l'inégalité $P(X > t) \leqslant P(Y > t)$. Cette dernière inégalité étant vérifiée pour tout t, on peut l'intégrer entre 0 et $+\infty$, pour obtenir [8] :

$$\mathbf{E}X = \int_0^{+\infty} P(X > t)\,\mathrm{d}t \leqslant \int_0^{+\infty} P(Y > t)\,\mathrm{d}t = \mathbf{E}Y.$$

□

La proposition suivante nous donne un peu de confort pour l'expression des espérances de variables aléatoires positives et nous sera utile pour l'inégalité de Markov.

Proposition 7.14. *Pour toute variable aléatoire positive X,*

$$\int_0^{+\infty} P(X > t)\,\mathrm{d}t = \int_0^{+\infty} P(X \geqslant t)\,\mathrm{d}t \quad \textit{(égalité dans } \overline{\mathbb{R}}_+\textit{)}. \tag{7.16}$$

Avant d'en donner la preuve, notons que (7.16) n'a rien d'évident car $P(X > t)$ et $P(X \geqslant t)$ peuvent différer pour certaines valeurs de t (au plus pour une infinité dénombrable de valeurs de t).

Preuve. Notons respectivement I et J le premier et le deuxième membre de (7.16). On prouve leur égalité en montrant l'inégalité dans les deux sens. L'inégalité $I \leqslant J$ s'obtient par intégration de l'inégalité $P(X > t) \leqslant P(X \geqslant t)$ vraie pour tout t.

Pour montrer que $J \leqslant I$, fixons $\varepsilon > 0$ quelconque. L'intégrande dans J est une fonction *positive* localement Riemann intégrable sur $[0,+\infty[$. On peut donc effectuer

8. La croissance de l'intégrale de Riemann (cf. prop. 3.22 ii)) passe aux intégrales généralisées de fonctions *positives*. En effet, si f et g sont positives et localement Riemann intégrables sur $[0,+\infty[$ et telles que $f \leqslant g$ sur $[0,+\infty[$, alors $\int_0^x f(t)\,\mathrm{d}t \leqslant \int_0^x g(t)\,\mathrm{d}t$ pour tout $x \geqslant 0$ et cette inégalité entre deux fonctions *croissantes* de x se conserve dans $\overline{\mathbb{R}}_+$ par passage à la limite quand x tend vers $+\infty$.

dans J le changement de variable « translation » $t = s + \varepsilon$, cf. proposition 4.41-i), qui nous donne :

$$J = \int_{-\varepsilon}^{+\infty} P(X \geqslant s + \varepsilon)\,\mathrm{d}s = \int_{-\varepsilon}^{0} P(X \geqslant s + \varepsilon)\,\mathrm{d}s + \int_{0}^{+\infty} P(X \geqslant s + \varepsilon)\,\mathrm{d}s.$$

En majorant $P(X \geqslant s + \varepsilon)$ par 1 sur $[-\varepsilon, 0]$ et par $P(X > s)$ sur $[0, +\infty[$, on en déduit que

$$J \leqslant \varepsilon + \int_{0}^{+\infty} P(X > s)\,\mathrm{d}s = \varepsilon + I.$$

L'inégalité $J \leqslant I + \varepsilon$ étant ainsi vérifiée pour tout $\varepsilon > 0$, on en déduit en faisant tendre ε vers 0 que $J \leqslant I$. □

Remarque 7.15. Dans la démonstration ci-dessus, la positivité de X ne joue aucun rôle. Donc (7.16) reste valable pour n'importe quelle variable aléatoire réelle X. Une adaptation facile de la preuve ci-dessus montre que l'on a aussi

$$\int_{-\infty}^{0} P(X < t)\,\mathrm{d}t = \int_{-\infty}^{0} P(X \leqslant t)\,\mathrm{d}t. \tag{7.17}$$

Proposition 7.16 (inégalité de Markov)**.** *Si X est une variable aléatoire positive,*

$$\forall x > 0, \quad P(X \geqslant x) \leqslant \frac{\mathbf{E}X}{x}. \tag{7.18}$$

Remarques 7.17. Cette inégalité n'a d'intérêt que lorsque le second membre est inférieur à 1, c'est-à-dire lorsque $\mathbf{E}X < +\infty$ *et* $x > \mathbf{E}X$.

D'autre part il peut sembler un peu incongru de vouloir contrôler $P(X \geqslant x)$ à l'aide de $\mathbf{E}X$, puisque le calcul de cette espérance par la définition 7.1 présuppose la connaissance des $P(X > t)$ pour $t \geqslant 0$ (dont on déduit facilement les $P(X \geqslant t)$). Il se trouve qu'il arrive souvent en pratique que l'on sache calculer $\mathbf{E}X$ sans connaître, ou sans avoir besoin de calculer, la loi de X. C'est le cas par exemple quand X est une somme finie de variables aléatoires d'espérances connues. On peut aussi savoir majorer $\mathbf{E}X$ sans connaître la loi de X. Dans ces situations, l'inégalité de Markov est très utile. Pour ne citer qu'un exemple, l'inégalité de Markov est l'un des outils pour établir des « lois des grands nombres ».

Voici maintenant 3 preuves de l'inégalité de Markov, libre au lecteur de choisir celle qu'il préfère.

Preuve n° 1. C'est la preuve « muette » donnée par la figure 7.7.

□

Preuve n° 2. Cette preuve ne fait que traduire explicitement la preuve graphique n° 1. Fixons $x > 0$, la quantité $P(X \geqslant x)$ devenant ainsi une constante. À partir de cette constante, définissons la fonction $h : [0, +\infty[\to \mathbb{R}_+,\ t \mapsto P(X \geqslant x)\mathbf{1}_{[0,x]}(t)$. Par

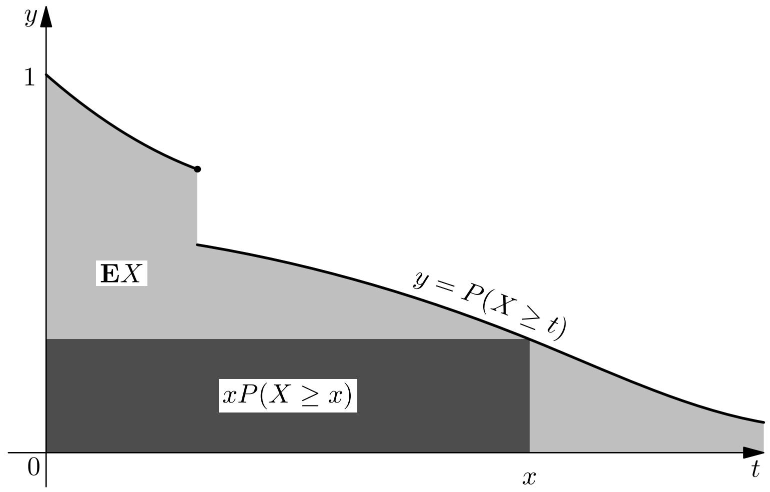

FIGURE 7.7 – Inégalité de Markov : $xP(X \geqslant x) \leqslant \int_0^{+\infty} P(X \geqslant t)\,\mathrm{d}t = \mathbf{E}X$.

décroissance de la fonction $t \mapsto P(X \geqslant t)$, on a $h(t) \leqslant P(X \geqslant t)$ pour tout $t \in [0, x]$. D'autre part cette inégalité est aussi vérifiée pour tout $t > x$ car alors $h(t) = 0$. En intégrant sur $[0, +\infty[$ l'inégalité $h(t) \leqslant P(X \geqslant t)$, on obtient compte-tenu de (7.16) :

$$\int_0^{+\infty} h(t)\,\mathrm{d}t \leqslant \int_0^{+\infty} P(X \geqslant t)\,\mathrm{d}t = \mathbf{E}X.$$

D'autre part, puisque h est nulle sur $]x, +\infty[$,

$$\int_0^{+\infty} h(t)\,\mathrm{d}t = \int_0^{x} P(X \geqslant x)\,\mathrm{d}t = xP(X \geqslant x).$$

Par conséquent

$$xP(X \geqslant x) \leqslant \mathbf{E}X,$$

ce qui nous donne (7.18) puisque $x > 0$. □

Preuve no 3. Cette preuve plus abstraite exploite les propriétés déjà connues de l'espérance des v.a. positives. On fixe x qui joue donc le rôle d'une constante dans toute la preuve. On part de l'inégalité entre v.a. positives : $x\mathbf{1}_{\{X \geqslant x\}} \leqslant X$ (vérifiez) dont on déduit par croissance de $\mathbf{E}$ (proposition 7.13) :

$$\mathbf{E}\big(x\mathbf{1}_{\{X \geqslant x\}}\big) \leqslant \mathbf{E}X,$$

puis grâce aux propositions 7.12 et 7.7, $xP(X \geqslant x) \leqslant \mathbf{E}X$. On conclut en divisant par $x > 0$. □