

L'espérance mathématique d'une variable aléatoire

Gérard Letac¹

¹Professeur émérite, Institut de Mathématiques de Toulouse., Toulouse

25 mai 2022



Espérance d'une variable et théorème de transport.

1 L'espérance mathématique d'une variable aléatoire

1.1 Les variables aléatoires étagées.

DÉFINITION 0.1 espérance mathématique de X

Soit (Ω, \mathcal{A}, P) un espace de probabilité. Désignons par \mathcal{E} l'ensemble de toutes les variables aléatoires réelles étagées définies sur Ω . A tout élément X de \mathcal{E} nous associons un nombre appelé *espérance mathématique de X* , noté $\mathbb{E}(X)$, et défini ainsi : si la loi de X est

$$P_X = p_1\delta_{a_1} + \cdots + p_N\delta_{a_N},$$

alors

$$\mathbb{E}(X) = p_1a_1 + \cdots + p_Na_N.$$

En fait, \mathcal{E} est un espace vectoriel et $X \mapsto \mathbb{E}(X)$ est une forme linéaire positive dessus, comme le montre le théorème suivant :

THÉORÈME 0.1

(Linéarité et positivité de l'espérance)

Si X et Y sont des v.a. étagées sur Ω alors $\lambda X + \mu Y$, pour des réels λ et μ , est encore une v.a. étagée. De plus $\mathbb{E}(\lambda X + \mu Y) = \lambda\mathbb{E}(X) + \mu\mathbb{E}(Y)$. Enfin $\mathbb{E}(X) \geq \mathbb{E}(Y)$ si $X \geq Y$.

Démonstration Introduisons les lois de X et Y :

$$P_X = p_1\delta_{a_1} + \dots + p_N\delta_{a_N}, \quad P_Y = q_1\delta_{b_1} + \dots + q_M\delta_{b_M},$$

notons $X^{-1}(\{a_i\}) = A_i, Y^{-1}(\{b_j\}) = B_j$ et $C_{ij} = A_i \cap B_j$ et $r_{ij} = P(C_{ij})$. La matrice (r_{ij}) a pour somme des lignes le vecteur ligne (q_1, \dots, q_M) et pour somme des colonnes le vecteur colonne ${}^t(p_1, \dots, p_N)$. Les valeurs prises par $Z = \lambda X + \mu Y$ sont les $c_{ij} = \lambda a_i + \mu b_j$ et comme $Z^{-1}(\{c_{ij}\}) = C_{ij} \in \mathcal{A}$, on en déduit que Z est aussi une v.a. Sa loi est

$$P_Z = \sum_{ij} r_{ij} \delta_{c_{ij}},$$

et est donc d'espérance

$$\begin{aligned} \mathbb{E}(Z) &= \sum_{ij} r_{ij} c_{ij} = \sum_{ij} r_{ij} (\lambda a_i + \mu b_j) = \\ &= \lambda \sum_i a_i \sum_j r_{ij} + \mu \sum_j b_j \sum_i r_{ij} = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y). \end{aligned}$$

Quant à l'inégalité, il suffit d'observer que $\mathbb{E}(X - Y) \geq 0$ par définition de l'espérance et d'appliquer ensuite la linéarité qu'on vient de démontrer.

DÉFINITION 0.2 Variable aléatoire de Bernoulli

Variable aléatoire de Bernoulli : Un exemple particulièrement simple et important de v.a étagée est celui où X ne prend que les valeurs 0 et 1, c'est à dire où la loi de X est

$$P_X = (1 - p)\delta_0 + p\delta_1,$$

où $p \in [0, 1]$. Sa loi est appelée une **loi de Bernoulli**. p est appelé le **paramètre de la loi de Bernoulli**.

PROPOSITION 0.2

L'espérance d'une loi de Bernoulli X de paramètre p est p . Si X est définie sur l'espace de probabilité (Ω, \mathcal{A}, P) , soit $A = \{\omega ; X(\omega) = 1\}$ alors $X = \mathbf{1}_A$ est l'indicateur de A , et on a donc

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

Inversement, un indicateur a toujours une loi de Bernoulli.

Nous allons utiliser le théorème précédent et les indicateurs pour terminer la démonstration du théorème . On veut donc montrer que si $B_j \in \mathcal{A}_j = \{\emptyset, A_j, A_j^c, \Omega\}$ et si les A_j sont indépendants, alors

$$P(\cap_{j=1}^N B_j) = \prod_{j=1}^N P(B_j).$$

On le montre en remarquant d'abord que dans les 4 cas possibles pour B_j , il existe deux nombres a_j et b_j tels que

$$\mathbf{1}_{B_j} = a_j + b_j \mathbf{1}_{A_j};$$

on prend en effet $a_j = b_j = 0$ si B_j est vide, $a_j = 1, b_j = 0$ si B_j est plein, $a_j = 0, b_j = 1$ si $B_j = A_j$, $a_j = 1, b_j = -1$ si $B_j = A_j^c$. D'où le calcul :

$$\begin{aligned} P(\cap_{j=1}^N B_j) &= \mathbb{E}(\prod_{j=1}^N \mathbf{1}_{B_j}) = \mathbb{E}(\prod_{j=1}^N (a_j + b_j \mathbf{1}_{A_j})) = \mathbb{E}[\sum_I (\prod_{j \in I^c} a_j) (\prod_{j \in I} b_j \mathbf{1}_{A_j})] = \\ &= \sum_I (\prod_{j \in I^c} a_j) (\prod_{j \in I} b_j) \mathbb{E}(\prod_{j \in I} \mathbf{1}_{A_j}) = \sum_I (\prod_{j \in I^c} a_j) (\prod_{j \in I} b_j) P(\cap_{j \in I} A_j) = \\ &= \sum_I (\prod_{j \in I^c} a_j) (\prod_{j \in I} b_j) (\prod_{j \in I} P(A_j)) = \prod_{j=1}^N (a_j + b_j P(A_j)) = \prod_{j=1}^N \mathbb{E}(\mathbf{1}_{B_j}) = \prod_{j=1}^N P(B_j). \end{aligned}$$

Dans cette chaîne de 9 égalités, la première, la cinquième et les 2 dernières s'appuient sur le fait que l'espérance de l'indicateur est la probabilité, la deuxième sur la définition des a_j et b_j , la troisième et la septième sur un développement algébrique ; enfin, surtout, la quatrième s'appuie sur le théorème précédent et la sixième sur l'indépendance des A_j .

1.2 Espérance d'une variable aléatoire quelconque.

Toutes les variables aléatoires ne sont pas étagées, mais toutes sont approchables par des v.a. étagées, et cela va permettre de définir l'espérance d'une v.a. quelconque. Plus précisément, on a le théorème suivant :

THÉORÈME 0.3

Soit (Ω, \mathcal{A}, P) un espace de probabilité, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire *positive*. Alors

1. Il existe une suite croissante de v.a. étagées (X_n) telle $X = \lim_{n \rightarrow +\infty} X_n$.
2. Si la suite (X_n) ci dessus est telle que $\mathbb{E}(X_n)$ soit bornée, alors le nombre

$$\lim_{n \rightarrow +\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$$

ne dépend que de X et non de la suite particulière (X_n) , dans le sens que si (X'_n) a les propriétés demandées à (X_n) au 1), alors la suite $\mathbb{E}(X'_n)$ a la même limite. $\mathbb{E}(X)$ est l'**espérance** de la variable aléatoire positive X .

3. Si Y est une autre v.a positive sur (Ω, \mathcal{A}, P) telle que $E(Y)$ existe, et si λ et μ sont des nombres ≥ 0 , alors $\mathbb{E}(\lambda X + \mu Y)$ existe et est égale à $\lambda \mathbb{E}(X) + \mu \mathbb{E}(Y)$.
4. Si $0 \leq X \leq Y$ et si $\mathbb{E}(Y)$ existe, alors $\mathbb{E}(X)$ existe et $\mathbb{E}(X) \leq \mathbb{E}(Y)$.
5. Si $X \geq 0$, alors $\mathbb{E}(X) = 0$ si et seulement si la loi de X est la probabilité de Dirac en 0.

Nous omettons la démonstration, bien que celle ci ne soit pas difficile. Il faut insister sur le fait que l'espérance de cette v.a. positive n'existe pas toujours.

Ce théorème définit donc $\mathbb{E}(X)$ pour des v.a positives. Pour passer au cas d'une v.a de signe quelconque, voici la démarche à suivre :

DÉFINITION 0.3 espérance

On considère une v.a. X définie sur (Ω, \mathcal{A}, P) et on écrit cette fonction de $\omega \in \Omega$ comme différence de deux fonctions positives $X = X_+ - X_-$, où a_+ signifie $\max(a, 0)$ et $a_- = (-a)_+$ (rappelons que cela implique $a = a_+ - a_-$ et $|a| = a_+ + a_-$). Donc $|X| = X_+ + X_-$. On dira que $\mathbb{E}(X)$ existe si, au sens du théorème, l'espérance de $|X|$ existe. Dans ces conditions, d'après le 2) du théorème, $\mathbb{E}(X_+)$ et $\mathbb{E}(X_-)$ existent, et on définit l'espérance de X par $\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-)$.

On a alors l'importante extension du théorème de linéarité et de positivité :

COROLLAIRE 0.4

Soit (Ω, \mathcal{A}, P) un espace de probabilité, soit \mathcal{L}_1 l'ensemble des variables aléatoires X sur cet espace telles que $\mathbb{E}(X)$ existe (ou, de façon équivalente, telles que $\mathbb{E}(|X|)$ soit finie). Alors \mathcal{L}_1 est un espace vectoriel et $X \mapsto \mathbb{E}(X)$ est une forme linéaire sur \mathcal{L}_1 , telle que de plus $\mathbb{E}(X) \geq \mathbb{E}(Y)$ si $X \geq Y$.

Appliquons cela à deux cas particuliers importants, celui où X est discrète et positive et celui où la loi de X a une densité.

PROPOSITION 0.5

Soit X une v.a. discrète avec

$$P_X = \sum_{j=1}^{\infty} p_j \delta_{a_j}$$

où $\sum_{j=1}^{\infty} p_j = 1$. Alors l'espérance de X , $E(X)$ existe si et seulement si la série $\sum_{j=1}^{\infty} p_j a_j$ est absolument convergente. S'il en est ainsi, alors

$$\mathbb{E}(X) = \sum_{j=1}^{\infty} p_j a_j.$$

Démonstration Montrons le d'abord si les a_n sont positifs ou nuls. Alors puisque $X = \sum_{j=1}^{\infty} a_j \mathbf{1}_{A_j}$, où les événements $A_j = \{X = j\}$ sont deux à deux disjoints dans Ω , il suffit de considérer la v.a. étagée $X_n = \sum_{j=1}^n a_j \mathbf{1}_{A_j}$, qui est nulle sur $\cup_{j=n+1}^{\infty} A_j$, et qui définit une suite ayant les propriétés requises au théorème. Le résultat est alors clair.

Si les a_n ne sont pas positifs on écrit $a_n = (a_n)_+ - (a_n)_-$ et les deux séries $\sum_{j=1}^{\infty} p_j (a_j)_+$ et $\sum_{j=1}^{\infty} p_j (a_j)_-$ convergent si et seulement si $\sum_{j=1}^{\infty} p_j a_j$ est absolument convergente. Cela permet de conclure facilement.

PROPOSITION 0.6

Supposons que la loi de la v.a. X ait une densité f avec un nombre fini de points de discontinuités $a_1 < \dots < a_N$. Alors l'espérance de X , $E(X)$ existe si et seulement si $\int_{-\infty}^{\infty} x f(x) dx$ est absolument convergente. S'il en est ainsi, alors

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

Démonstration Contentons nous de donner les idées de la démonstration quand X est positive et quand sa densité f est continue. L'extension aux hypothèses du théorème sera alors standard. On découpe $[0, n]$ en $n2^n$ intervalles égaux par les points $x_k = \frac{k}{2^n}$, avec $k = 0, 1, \dots, n2^n$, on convient $x_{n2^n+1} = +\infty$ et on définit la variable aléatoire étagée $X_n = x_k$ quand $x_k \leq X < x_{k+1}$. Ceci est bien une suite croissante et on a bien $\lim_{n \rightarrow +\infty} X_n = X$.

Si $\int_0^\infty xf(x)dx$ converge, notons

$$D_n = \int_0^\infty xf(x)dx - \mathbb{E}(X_n) = \sum_{k=0}^{n2^n} \int_{x_k}^{x_{k+1}} (x - x_k)f(x)dx.$$

Soit $\epsilon > 0$. Il existe un entier A tel que $\int_A^\infty xf(x)dx \leq \epsilon$. Soit alors K tel que $x_K = A$ et soit F la fonction de répartition de X . On partage alors D_n en deux sommes A_n et B_n , avec

$$A_n = \sum_{k=K}^{n2^n} \int_{x_k}^{x_{k+1}} (x - x_k)f(x)dx \leq 2 \int_A^{+\infty} xf(x)dx \leq 2\epsilon,$$

$$B_n = \sum_{k=0}^{K-1} \int_{x_k}^{x_{k+1}} (x - x_k)f(x)dx = - \int_0^A F(x)dx + \sum_{k=0}^{K-1} (x_{k+1} - x_k)F(x_{k+1}),$$

la dernière égalité étant obtenue par intégration par parties en posant $u = (x - x_k)$ et $v' = f$. Notons que les symboles x_K et K sont des fonctions de n . Si n tend vers l'infini, (B_n) tend vers zéro, comme suite des différences entre une intégrale et les sommes de Riemann de cette intégrale. On voit donc que (D_n) tend vers 0. Le cas où $\int_0^\infty xf(x)dx$ diverge est similaire.

Exercices sur

1. Calculer l'espérance d'une variable aléatoire de loi

$$\sum_{n=1}^{\infty} \frac{4}{n(n+1)(n+2)} \delta_n.$$

2. Pour quelles valeurs de $a > 0$ la variable aléatoire X ayant pour fonction de répartition $F_X(x) = 1 - \frac{1}{(1+x)^a}$ si $x > 0$, et $F_X(x) = 0$ si $x \leq 0$, possède-t-elle une espérance ?

1.3 Théorème du transport.

Il arrive souvent qu'on ait besoin de calculer, non l'espérance de la variable aléatoire X , mais l'espérance d'une fonction $Y = g(X)$ de celle-ci. Si on applique la définition de l'espérance, cela suppose qu'on calcule la loi de Y , ce qui peut être très incommode. Le résultat suivant simplifie ce problème.

THÉORÈME 0.7 (du transport)

Soit X une v.a. sur l'espace de probabilité (Ω, \mathcal{A}, P) . Soit $x \mapsto y = g(x)$ une fonction mesurable de \mathbb{R} dans \mathbb{R} . Si X est étagée ou discrète et de loi

$$P_X = \sum_{j \geq 1} p_j \delta_{a_j},$$

alors l'espérance de X , $\mathbb{E}(g(X))$ existe si et seulement si

$$\sum_{j \geq 1} p_j g(a_j)$$

converge absolument et dans ce cas $\mathbb{E}(g(X))$ est égale à cette somme.

Si X a une densité f , alors de même $\mathbb{E}(g(X))$ existe si et seulement si

$$\int_{-\infty}^{\infty} g(x)f(x)dx$$

est absolument convergente, et dans ce cas $\mathbb{E}(g(X))$ est égale à la somme de l'intégrale.

Démonstration On montre d'abord le résultat quand X est étagée, puis quand X est positive en appliquant la définition de l'espérance d'une variable aléatoire positive, et on passe facilement au cas où X est de signe quelconque.

Exercices sur

1. Soit une variable aléatoire X de densité $\frac{1}{2}\exp(-|x|)$. Soit z un nombre réel et soit $g(x) = \exp(zx)$. Pour quelles valeurs de z $Y = g(X)$ a-t-elle une espérance? La calculer quand elle existe.
2. X une variable aléatoire de densité $\frac{1}{2}\mathbf{1}_{[-1,1]}(x)$ et soit $Y = \tan(\frac{\pi}{2}X)$. Etudier de deux manières l'existence éventuelle de $\mathbb{E}(Y)$: soit à l'aide du théorème du transport, soit en calculant la densité de Y : pour cela, écrire d'abord la fonction de répartition de Y puis dériver.

1.4 Variables aléatoires indépendantes et espérance du produit.

DÉFINITION 0.4 famille indépendante

Soit (X_1, \dots, X_N) une suite de v.a. sur (Ω, \mathcal{A}, P) . On se rappelle que si \mathcal{B} est la tribu de Borel, alors par définition des variables aléatoires $X_j^{-1}(\mathcal{B}) = \mathcal{A}_j$ est une sous tribu de \mathcal{A} .

Nous dirons que c'est une suite de variables aléatoires indépendantes si la famille de sous tribus $\{\mathcal{A}_1, \dots, \mathcal{A}_N\}$ est une *famille indépendante*.

Ceci entraîne un fait simple et utile : si les X_j sont des v.a. indépendantes, et si f_j est une fonction réelle quelconque, alors les $Y_j = f_j(X_j)$ sont des v.a. indépendantes aussi.

Dans le théorème suivant, qui sert à caractériser l'indépendance pratiquement, contentons nous de $N = 2$: la généralisation $N > 2$ est évidente.

THÉORÈME 0.8

Soit X et Y deux variables aléatoires sur (Ω, \mathcal{A}, P) . Alors elles sont indépendantes si et seulement si pour tous x et y réels on a

$$P(X \leq x; Y \leq y) = F_X(x)F_Y(y) = P(X \leq x)P(Y \leq y).$$

En particulier, si elles sont discrètes de lois respectives

$$P_X = \sum_{i \geq 1} p_i \delta_{a_i}, \quad P_Y = \sum_{j \geq 1} q_j \delta_{b_j},$$

alors elles sont indépendantes si et seulement si pour tout couple (i, j) on a

$$P(X = a_i; Y = b_j) = p_i q_j = P(X = a_i)P(Y = b_j).$$

Démonstration *Partie \Rightarrow . Introduisons les évènements $A = \{X \leq x\} \in X^{-1}(\mathcal{B})$ et $B = \{Y \leq y\} \in X^{-1}(\mathcal{B})$. Par hypothèse ils sont indépendants.*

Partie \Leftarrow . Elle n'est pas élémentaire et sera montrée en 3^{ème} année.

Toutefois, dans le cas discret de la seconde partie la démonstration directe est facile.

Voici enfin un théorème d'une importance considérable.

THÉORÈME 0.9

Soit (X_1, \dots, X_N) une suite de v.a. indépendantes sur (Ω, \mathcal{A}, P) . Alors le produit $X_1 \cdots X_N$ a une espérance si et seulement si chaque X_j a une espérance. Dans ces conditions l'espérance du produit est le produit des espérances :

$$\mathbb{E}(X_1 \cdots X_N) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_N).$$

Démonstration *On le démontre d'abord pour $N = 2$, et une récurrence permet de passer au cas de N quelconque. Pour $N = 2$, notons $X = X_1$ et $Y = X_2$ pour simplifier. On le démontre d'abord dans le cas où X et Y sont étagées. Ceci fait, on suppose ensuite que X et Y sont positives. Il est facile de construire deux suites croissantes (X_n) et (Y_n) de v.a. étagées qui sont de plus indépendantes. Comme $(X_n Y_n)$ est à son tour une suite de v.a. qui croit vers XY , on arrive au résultat. Quant au passage au cas où les X et Y ne sont plus positives, il est standard.*

Exercices sur

1. Soit X et Y deux variables aléatoires indépendantes à valeurs dans les entiers ≥ 0 de lois respectives données par $P(X = n) = (1 - p)^n p$ et $P(Y = n) = (1 - q)^n q$, où p et q sont dans $]0, 1[$. Montrer à l'aide de la deuxième partie du Th. que $U = X - Y$ et $V = \min(X, Y)$ sont indépendantes.
2. Soit une matrice carrée d'ordre 2 dont les coefficients sont des variables aléatoires indépendantes et de même loi $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$. Calculer l'espérance du carré du déterminant de cette matrice.

Références