

## 5 variables aléatoires

### 5.1 définition

Une **variable aléatoire** sur un espace probabilisé  $(\Omega, p)$  est une application  $X : \Omega \rightarrow \mathbb{R}$ .

**exemples :** le résultat obtenu en jetant un dé, la mesure d'une tige métallique, le nombre de pannes quotidiennes d'une machine, ... sont des variables aléatoires.

Chaque variable aléatoire définit une probabilité  $p_X$  sur  $\mathbb{R}$  de la manière suivante :

pour tout  $A \subset \mathbb{R}$ ,  $p_X(A) = p(X^{-1}(A)) = p(X \in A) = p(\{x \in \Omega \mid X(x) \in A\})$ .

$p_X(A)$  est donc la probabilité pour que la variable aléatoire  $X$  prenne ses valeurs dans un ensemble  $A$  fixé. Et la **loi** d'une variable aléatoire  $X$  est la donnée des valeurs  $p_X(A)$  pour tous les  $A \subset \mathbb{R}$ .

La **fonction de répartition**  $F_X$  d'une variable aléatoire  $X$  est la fonction définie par  $F_X(x) = p(X^{-1}(]-\infty; x[)) = p(X < x)$ .

$F_X$  est une fonction croissante, à valeurs dans  $[0; 1]$ . De plus, comme  $p(\emptyset) = 0$  et  $p(\Omega) = 1$ , la limite de  $F_X$  en  $-\infty$  est 0, sa limite en  $+\infty$  est 1, et donc  $p(X \geq x) = 1 - F_X(x)$ .

Bien sûr, donner la loi d'une variable aléatoire en donnant chaque  $p_X(A)$  est fastidieux, et souvent impossible si  $X(\Omega)$  est infini. Mais nous allons voir que l'on peut être plus efficace en décrivant la loi seulement pour certains événements « simples » : les événements élémentaires (pour une variable discrète) ou les événements du type  $] -\infty; x[$  (pour une variable continue), ce qui revient à donner seulement la fonction de répartition.

### 5.2 cas des univers finis ou infinis discrets

#### 5.2.1 variables aléatoires

Si les valeurs prises par  $X$  forment un ensemble discret  $X(\Omega) = \{x_1, x_2, \dots, x_n, \dots\}$  on peut décrire la loi de la variable  $X$  en donnant seulement les  $p_X(\{x_i\})$ , probabilité que les éléments de  $\Omega$  « prennent » la valeur  $x_i$ . En effet pour tout événement  $A$ ,  $p_X(A)$  sera la somme des  $p_X(\{a\})$  pour  $a \in A$ .

**remarque :** la notation  $p_X(\{x_i\})$  est lourde et abrège en  $p_X(x_i)$  ou  $p(X = x_i)$ .

Ainsi la probabilité de chaque événement élémentaire  $x_i$  est un nombre  $p(X = x_i) = p_i$  tel que  $\sum_{i=1}^{\text{card } \Omega} p_i = 1$ , et décrire la loi de  $X$  revient à donner l'ensemble des valeurs  $p_i$ .

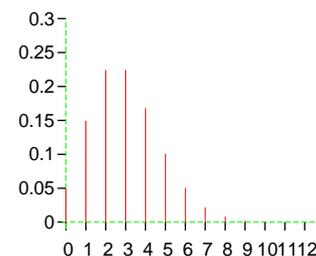
**exemple 1 :** on joue à pile ou face, et on appelle  $X$  la variable aléatoire qui vaut 0 en cas de résultat pile et 1 en cas de résultat face. Alors  $\Omega = \{\text{pile}, \text{face}\}$ , et  $X(\Omega) = \{0; 1\}$ .

Si la pièce est bien équilibrée on a donc  $p_X(0) = p(\text{pile}) = 1/2$ ,  $p_X(1) = p(\text{face}) = 1/2$ .

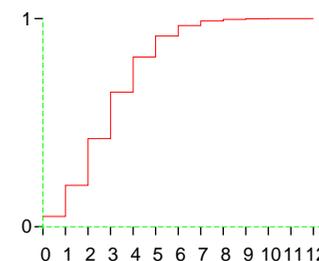
**exemple 2 :** on étudie une expérience de désintégration nucléaire. Soit  $X$  la variable aléatoire représentant le nombre d'atomes désintégrés en 1 seconde. Alors on a déjà vu que  $p(X = k)$  vaut  $\frac{\lambda^k}{k!} e^{-\lambda}$  où  $\lambda$  est une constante qui dépend du composé radioactif.

On peut représenter graphiquement les  $p_k$  et la fonction de répartition (avec ici  $\lambda = 3$ ) :

valeurs des  $p_k = p(X = k)$



fonction de répartition



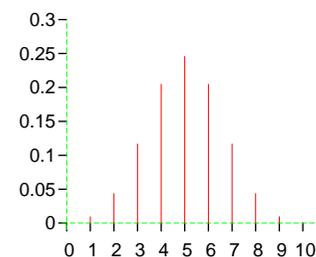
On peut déduire de ces probabilités élémentaires les probabilités des autres événements. Ainsi,  $p(\text{le nombre d'atomes désintégrés est compris entre 2 et 4}) = p(X \in [2; 4]) = p(X = 2) + p(X = 3) + p(X = 4) = F_X(5) - F_X(2)$  (car  $F_X(5) = p(X < 5) = p(X \leq 4)$ ,  $F_X(2) = p(X < 2)$ ). De même,  $p(\text{deux atomes au moins se sont désintégrés}) = p(X \in [2; +\infty[) = 1 - p(X < 2) = 1 - F_X(2)$ .

**exemple 3 :** on joue dix fois de suite à pile ou face, et on compte le nombre de « face ».  $\Omega$  est ici constitué des listes de 10 résultats successifs « pile » ou « face ». Que vaut  $X(\Omega)$  ? C'est  $\{0; 1; 2; \dots; 10\}$  : on peut obtenir face 0, 1, 2, ..., 9 ou 10 fois.

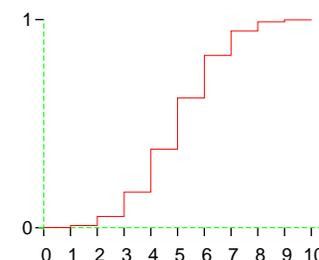
Et si  $k$  désigne un nombre entre 0 et 10,  $p_X(k) = \frac{\text{nombre de listes avec } k \text{ fois « face »}}{\text{nombre total de listes}}$  ; mais le nombre total de listes de 10 résultats pile ou face est  $2^{10}$ , alors que le nombre de listes avec exactement  $k$  résultats face est  $\binom{10}{k}$ . Ainsi,  $p(X = k) = \frac{\binom{10}{k}}{2^{10}}$ .

On obtient les représentations graphiques :

valeurs des  $p_k = p(X = k)$



fonction de répartition



On en déduit de même ici la probabilité d'autres événements, par exemple  $p(\text{le nombre de résultats pile est impair}) = p(X \in \{1, 3, 5, 7, 9\}) = p(X = 1) + p(X = 3) + p(X = 5) + p(X = 7) + p(X = 9)$ , ou bien  $p(X > 3) = 1 - p(X \leq 3) = 1 - F_X(4)$ .

### 5.2.2 espérance, variance, écart-type

On définit pour une variable aléatoire discrète  $X$  qui prend les valeurs  $\{x_1, x_2, \dots, x_n, \dots\}$  avec les probabilités  $\{p_1, p_2, \dots, p_n, \dots\}$  (où  $p_i = p(X = x_i)$ ) les trois quantités suivantes :

$$\text{l'espérance de } X \text{ est } E(X) = \sum_{i=1}^n p_i x_i = p_1 x_1 + p_2 x_2 + \dots,$$

la **variance** de  $X$  est

$$\text{Var}(X) = \sum_{i=1}^n p_i (x_i - E(X))^2 = p_1 (x_1 - E(X))^2 + p_2 (x_2 - E(X))^2 + \dots,$$

(à chaque fois les sommes ci-dessus peuvent être finies ou infinies)

$$\text{l'écart-type de } X \text{ est } \sigma(X) = \sqrt{\text{Var}(X)}.$$

Le terme espérance vient de la théorie des jeux, pour laquelle ont été développées les probabilités. Il s'agit en quelque sorte du « gain moyen » au cours du jeu : dans une loterie, avec une chance sur 10 de gagner 20€ et neuf chances sur dix de perdre la mise de 3€, l'espérance de gain, ou gain moyen sur un grand nombre de parties, sera bien de  $\frac{1}{10}20 + \frac{9}{10}(-3) = -\frac{7}{10}$  : l'organisateur du jeu est favorisé !

La variance (et l'écart-type) mesurent, eux, la dispersion des valeurs autour de cette espérance : ils sont d'autant plus élevés que les probabilités des valeurs éloignées de l'espérance (valeur moyenne) sont importantes.

Citons quelques-unes des propriétés remarquables de ces quantités :

L'espérance est linéaire :  $E(X + Y) = E(X) + E(Y)$  et  $E(aX) = aE(X)$

$$E(X + b) = E(X) + b$$

$$\text{Var}(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$$

$$\text{Var}(aX) = a^2 \text{Var}(X) \text{ et } \sigma(aX) = |a| \sigma(X)$$

$$\text{Var}(X + b) = \text{Var}(X) \text{ et } \sigma(X + b) = \sigma(X)$$

Une variable aléatoire  $X$  est dite **centrée** si  $E(X) = 0$ , et **réduite** si  $\text{Var}(X) = 1$ . Des propriétés précédentes on déduit que

si  $X$  est une variable aléatoire quelconque,

$$Z = \frac{X - E(X)}{\sigma(X)} \text{ est une variable aléatoire centrée réduite.}$$

### 5.3 variables aléatoires continues

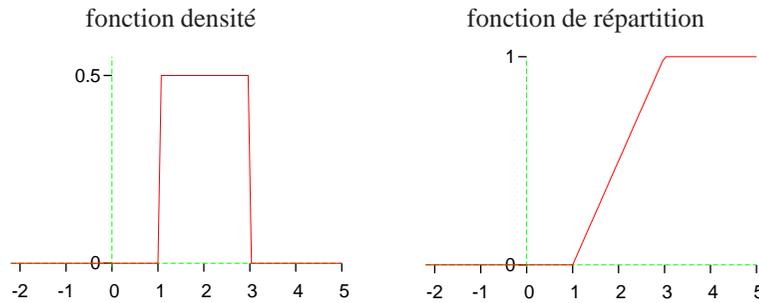
On dit qu'une variable aléatoire  $X$  admet une densité  $f_X$  si pour tout  $x \in \mathbb{R}$   $F_X(x) = p(X < x) = \int_{-\infty}^x f_X(t) dt$  : la densité est la dérivée de la fonction de répartition.

On a en particulier  $p(a \leq X < b) = F_X(b) - F_X(a)$ , et aussi  $p(X = a) = \int_a^a f_X(t) dt = 0$  : la probabilité d'obtenir une valeur donnée doit être considérée comme nulle, et ce sont les probabilités d'obtenir un résultat dans un segment  $[a, b]$  qui nous intéresseront, et permettront de calculer les probabilités de tous les événements.

**exemple** : si  $X$  est une variable aléatoire uniformément répartie sur  $[a; b]$ , on détermine

$$f_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{si } x > b \end{cases} \text{ et } F_X(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}.$$

que l'on peut représenter graphiquement, pour  $a = 1, b = 3$  :



**remarque :** nous ne présentons ici qu'un cas particulier, suffisant pour une première approche : en fait une variable aléatoire continue n'admet pas forcément de densité et on parle, si cette densité existe, de variables aléatoires absolument continues. Toutes les variables continues que nous rencontrerons seront en fait absolument continues.

On définit alors pour une variable aléatoire continue  $X$ , par analogie avec le cas discret :

l'**espérance** de  $X$  :  $E(X) = \int_I x f_X(x) dx$

la **variance** de  $X$  :  $\text{Var}(X) = \int_I f_X(x) (x - E(X))^2 dx$

l'**écart-type** de  $X$  :  $\sigma(X) = \sqrt{\text{Var}(X)}$ .

Ces formules sont bien entendu analogues aux formules vues dans le cas discret.

Plus précisément, les probabilités élémentaires  $p_i = p(X = x_i)$  sont remplacées par la fonction densité  $f_X(x)$ , que l'on peut interpréter en disant que  $f_X(x)dx$  est la probabilité de l'événement infinitésimal  $p(x \leq X \leq x + dx)$ .

Dans le cas discret, on obtient la probabilité de l'événement  $A$  par la formule  $p(A) = \sum_{a \in A} p(X = a)$  (la probabilité d'un événement est la somme des probabilités des événements élémentaires qui le composent). Dans le cas continu, la somme est remplacée par une intégrale, et on a la formule  $p(A) = \int_A f_X(x)dx$ .

De même, l'espérance  $E(X) = \sum_i x_i p_i$ , qui est la somme des produits de la forme (valeur de la variable aléatoire  $X$ )  $\times$  (probabilité d'obtenir cette valeur), devient  $\int_{x \in \mathbb{R}} x f(x)dx$  : on somme ici encore le produit de la valeur  $x$  par la probabilité  $f(x)dx$  que  $X$  prenne une valeur infiniment proche de  $x$ .

## 5.4 l'inégalité de Tchébycheff

Citons ce résultat qui justifiera, plus tard, la loi des grands nombres faisant le lien entre probabilités et statistiques :

Si  $X$  est une variable aléatoire quelconque, on a pour tout  $\lambda > 0$  :

$$p( E(X) - \lambda\sigma(X) < X < E(X) + \lambda\sigma(X) ) \geq 1 - 1/\lambda^2,$$

soit encore :

$$p( \frac{|X - E(X)|}{\sigma(X)} \geq \lambda ) \leq 1/\lambda^2.$$

Autrement dit, si  $\lambda$  tend vers l'infini, la probabilité que  $X$  prenne une valeur dans  $[E(X) - \lambda\sigma(X); E(X) + \lambda\sigma(X)]$  devient proche de 1.

Pour  $\lambda = 2$  par exemple, la probabilité que  $X$  prenne une valeur dans  $[E(X) - 2\sigma(X); E(X) + 2\sigma(X)]$  est au moins de  $3/4$ . Pour  $\lambda = 3$ , la probabilité que  $X$  prenne une valeur dans  $[E(X) - 3\sigma(X); E(X) + 3\sigma(X)]$  est au moins de  $8/9$ .

Ces valeurs sont, la plupart du temps, loin d'être optimales. Mais elles ont l'avantage d'être vérifiées sans aucune hypothèse sur la loi de  $X$ . Quand cette loi est connue, nous verrons plus loin comment améliorer ces résultats.

## 6 Couples de variables aléatoires

On peut vouloir étudier plusieurs variables définies sur une même population, et les liens entre ces variables. Par exemple : l'âge  $X$ , la taille  $Y$ , le revenu mensuel  $Z$ , ... Dans ce but nous allons introduire la notion de couple de variables aléatoires.

On considère donc deux variables  $X$  et  $Y$  définies sur un même univers  $\Omega$ .  $(X, Y)$  est une nouvelle variable aléatoire, à deux dimensions : c'est une application de  $\mathcal{P}(\Omega)$  dans  $\mathbb{R}^2$ .

### 6.1 loi d'un couple de variable aléatoires discrètes

Si  $\Omega$  est discret, pour connaître la **loi conjointe**, la loi de  $(X, Y)$ , il suffit de la connaître sur les événements élémentaires  $(x, y)$  : on doit donc connaître les  $p_{xy} = p(X = x \text{ et } Y = y)$ . Alors si  $A$  est une partie quelconque de  $\mathbb{R}^2$ , on a  $p((X, Y) \in A) = \sum_{(x,y) \in A} p((X, Y) = (x, y)) = \sum_{(x,y) \in A} p_{xy}$ . En particulier, les  $p_{xy}$  sont tous positifs, et leur somme vaut 1.

La loi d'un couple de variables aléatoires  $(X, Y)$  étant donnée, on définit la **loi marginale** de  $X$  par  $p(X = x) = \sum_y p(X = x, Y = y) = \sum_y p_{xy}$  ; et de même la **loi marginale**

de  $Y$  est donnée par  $p(Y = y) = \sum_x p(X = x, Y = y) = \sum_x p_{xy}$ .

On représente en général ces informations sous forme d'un tableau : les cases centrales donnent les probabilités des événements élémentaires, la somme de chaque ligne ou chaque colonne fournissant les probabilités marginales.

**exemple** : dans une urne contenant 4 boules indiscernables au toucher marquées 1 à 4, on en tire simultanément deux.

On appelle  $X$  le plus petit des deux numéros sortis, et  $Y$  le plus grand. Donner la loi conjointe et les lois marginales de  $(X, Y)$ .

$X Y$	2	3	4	loi marginale en $X$
1	1/6	1/6	1/6	1/2
2	0	1/6	1/6	1/3
3	0	0	1/6	1/6
loi marginale en $Y$	1/6	1/3	1/2	

**Indépendance** : on dit que deux variables discrètes  $X$  et  $Y$  sont indépendantes si tous les événements  $X = x$  et  $Y = y$  sont indépendants, autrement dit si l'on a  $p(X = x, Y = y) = p(X = x) \times p(Y = y)$  pour tous  $x, y$ .

L'intérêt d'avoir des variables indépendantes est là : la connaissance des lois marginales suffit pour connaître la loi conjointe.

**exemple** : étudier l'indépendance des variables  $X$  et  $Y$  de l'exercice précédent.

### 6.2 loi d'un couple de variables aléatoires continues : densité

Si  $X$  et  $Y$  sont des variables continues, on appelle **densité conjointe** de  $X$  et  $Y$  une fonction  $f(x, y)$  positive telle que  $\iint f(x, y) \, dx dy = 1$  et  $p((a < X < b) \cap (c < Y < d)) = \int_a^b \int_c^d f(x, y) \, dx dy$ .

Les **densités marginales** de  $X$  et  $Y$  sont respectivement  $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) \, dy$  et  $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) \, dx$ .

**Indépendance** : on dit que deux variables continues  $X$  et  $Y$  sont indépendantes si tous les événements  $a < X < b$  et  $c < Y < d$  sont indépendants, autrement dit si l'on a  $p(a < X < b, c < Y < d) = p(a < X < b) \times p(c < Y < d)$  pour tous les intervalles  $[a, b]$  et  $[c, d]$ . Cela équivaut à demander que la densité conjointe soit égale au produit des densités marginales :  $f(x, y) = f_X(x)f_Y(y)$ .

**exemple** : on peut voir un couple de variables aléatoires  $(X, Y)$  comme la détermination d'une répartition de masse totale de 1 sur  $\mathbb{R}^2$ . Dans le cas discret, il s'agit de répartir des masses ponctuelles. Dans le cas continu, des masses continues, la densité de probabilité correspond à une densité « physique ».

Par exemple quelle est la densité associée à la répartition de la masse uniformément sur un triangle de sommets  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 1)$  ?

Le triangle  $T$  est déterminé par les équations  $y \leq 1$ ,  $x \geq 0$  et  $y \geq x$ . On cherche une fonction nulle en dehors de  $T$  et constante sur  $T$ .

Sa valeur  $k$  doit donc vérifier  $\int_D k = 1$ , soit  $\int_0^1 \int_x^1 k \, dx dy = 1$ , soit encore  $\int_0^1 (1-x)k = k/2 = 1$ , donc  $k = 2$ .

La densité marginale en  $x$  vaut 0 si  $x < 0$  ou si  $x > 1$ , et sinon  $f_X(x) = \int_{y=x}^1 2 \, dy = 2(1-x)$  ; elle est, logiquement, plus importante quand  $x$  est proche de 0 que de 1.

De même la densité marginale en  $y$  vaut 0 en dehors de  $[0; 1]$  et sur  $[0; 1]$  :  $f_Y(y) = \int_{x=0}^y 2 \, dx = 2y$ .

### 6.3 propriétés

On peut citer les propriétés suivantes des couples de variables aléatoires, qu'elles soient discrètes ou continues :

pour deux variables aléatoires  $X$  et  $Y$  quelconques :  
$$E(X + Y) = E(X) + E(Y)$$

et si les variables  $X$  et  $Y$  sont indépendantes, alors :  
$$E(XY) = E(X)E(Y)$$

et  
$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

(attention, il ne suffit pas que ces relations soient vérifiées pour que les variables aléatoires soient indépendantes !!)

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y)$$

Si  $X$  et  $Y$  sont indépendantes,  $\text{cov}(X, Y) = \rho(X, Y) = 0$

$$|\text{cov}(X, Y)| \leq \sigma(X) \sigma(Y)$$

$$-1 \leq \rho(X, Y) \leq 1$$

Pour tous  $a, b$  réels,  $\text{Var}(aX + bY) = a^2\text{Var}(X) + 2abcov(X, Y) + b^2\text{Var}(Y)$

**remarque :** en toute généralité, la formule donnant  $E(XY)$  est

dans le cas discret,  
$$E(XY) = \sum_{x,y} p(X = x, Y = y)xy$$

dans le cas continu,  
$$E(XY) = \int \int f(x, y)xy \, dx dy,$$

dont le calcul sera souvent laborieux et parfois délicat : si les variables sont indépendantes, la formule  $E(XY) = E(X)E(Y)$  est bien plus efficace !

## 6.4 covariances

On appelle **covariance** de  $X$  et  $Y$  le nombre  $\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$   
et **coefficient de corrélation** de  $X$  et  $Y$  le nombre  $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma(X)\sigma(Y)}$ .

On a alors les propriétés suivantes :

Ces deux nombres mesurent l'importance de la dépendance entre les variables  $X$  et  $Y$  ;  $\rho(X, Y)$  a l'avantage d'être un nombre sans dimension, toujours compris entre 0 et 1.

**remarque :**  $\rho(X, Y) = 0$  n'implique pas que  $X$  et  $Y$  sont indépendantes !

Si  $\rho(X, Y) = \pm 1$ ,  $X$  et  $Y$  sont liées par une relation de dépendance linéaire :  $Y$  est de la forme  $Y = aX + b$ .

**exemple :** en reprenant l'exemple précédent du tirage de deux boules,  $X$  étant la plus petite des deux valeurs et  $Y$  la plus grande, on trouve  $E(X) = 1/2 \times 1 + 1/3 \times 2 + 1/6 \times 3 = 5/3$  et  $E(Y) = 1/6 \times 2 + 1/3 \times 3 + 1/2 \times 4 = 10/3$ .

Comme de plus  $E(XY) = 1/6 \times 1 \times 2 + 1/6 \times 1 \times 3 + 1/6 \times 1 \times 4 + 1/6 \times 2 \times 3 + 1/6 \times 2 \times 4 + 1/6 \times 3 \times 4 = 35/6$ , la covariance  $\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 5/18$  est bien non nulle : on retrouve le fait que les variables  $X$  et  $Y$  ne sont pas indépendantes.