

# Probabilités

## Variables Aléatoires Discrètes

---

**Mohamad GHASSANY**

EFREI PARIS

## Mohamad GHASSANY

- ▶ Associate Professor at EFREI Paris, head of Data & Artificial Intelligence Master program.
- ▶ Phd in Computer Science Université Paris 13.
- ▶ Master 2 in Applied Mathematics & Statistics from Université Grenoble Alpes.
- ▶ Personal Website: [mghassany.com](http://mghassany.com)



Introduction aux probabilités

Notion de variable aléatoire réelle

Variables aléatoires discrètes

Moments d'une variable aléatoire discrète

Couple de variables aléatoires discrètes

## Introduction aux probabilités

---

## Hasard

**Exemple fondamental:** Considérons le jeu du lancé d'un dé.

- ▶ **Expérience aléatoire**  $\varepsilon$  : "lancer un dé équilibré"  $\longleftarrow$  **Action**.
- ▶ **Univers**: l'ensemble de tous les résultats possibles de cette expérience aléatoire

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

- ▶ **Événements**: Dans cette expérience aléatoire, on peut s'intéresser à des événements plus complexes qu'un simple résultat élémentaire.
- ▶ L'**ensemble de parties de  $\Omega$** , appelé  $\mathcal{P}(\Omega)$ , est l'ensemble des sous-ensembles de  $\Omega$ .
- ▶ Une **famille  $\mathcal{A}$  de parties** (i.e. de sous ensembles) de  $\Omega$ . Ces parties sont appelées des événements. On dit que l'événement  $A$  s'est réalisé si et seulement si le résultat  $\omega$  de  $\Omega$  qui s'est produit appartient à  $A$ .
- ▶ **Tribu**: On appelle **tribu** sur  $\Omega$ , toute famille  $\mathcal{A}$  de parties de  $\Omega$  vérifiant:
  1.  $\Omega \in \mathcal{A}$ .
  2. si  $A \in \mathcal{A}$ , alors  $\bar{A} \in \mathcal{A}$ .
  3. si  $(A_n)_{n \in \mathbb{N}}$  est une suite d'éléments de  $\mathcal{A}$ , alors  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$ .
- ▶  $(\Omega, \mathcal{A})$  est un **espace probibilisable**.

- ▶ Soit  $(\Omega, \mathcal{A})$  un espace probablisable:
  - L'ensemble  $\mathcal{A}$  est appelé **tribu des événements**. Les éléments de  $\mathcal{A}$  s'appellent les **événements**.
  - L'événement  $\Omega$  est appelé **événement certain**. L'événement  $\emptyset$  est appelé **événement impossible**.
- ▶ **Opérations sur les événements**. Soient  $A$  et  $B$  deux événements:
  - $\bar{A}$  est l'événement **contraire** de  $A$  (on note aussi  $A^c$ ).  $\bar{A} = \Omega \setminus A$ .  
 $\bar{A}$  se réalise si et seulement si  $A$  ne se réalise pas.
  - $A \cap B$  est l'événement « $A$  **et**  $B$ ».  
 $A \cap B$  se réalise lorsque les deux événements se réalisent.
  - $A \cup B$  est l'événement « $A$  **ou**  $B$ ».  
 $A \cup B$  se réalise lorsque au moins un des deux événements se réalise.
- ▶ **Incompatibilité**:  $A$  et  $B$  sont incompatibles si leur réalisation simultanée est impossible:  $A \cap B = \emptyset$ .
- ▶ **Implication**:  $A$  implique  $B$  signifie que si  $A$  se réalise, alors  $B$  se réalise aussi:  $A \subset B$ .

- Soit  $(\Omega, \mathcal{A})$  un espace probabilisable. On appelle **probabilité** sur  $(\Omega, \mathcal{A})$ , toute application

$$P : \mathcal{A} \rightarrow \mathbb{R}$$

vérifiant:

1.  $\forall A \in \mathcal{A}, P(A) \geq 0$ .
2.  $P(\Omega) = 1$ .
3.  $\forall (A_n)_{n \in \mathbb{N}^*} \in \mathcal{A}^{\mathbb{N}^*}$ , une suite d'éléments de  $\mathcal{A}$  deux à deux incompatibles, on a:

$$P\left(\bigcup_{n \in \mathbb{N}^*} A_n\right) = \sum_{n=1}^{+\infty} P(A_n)$$

- Le triplet  $(\Omega, \mathcal{A}, P)$  est appelé **espace probabilisé**.

1.  $P(\emptyset) = 0$ .
2.  $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$ .
3. Si  $A_1$  et  $A_2$  sont incompatibles,  $A_1 \cap A_2 = \emptyset$ ,  $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ .
4.  $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$ .
5.  $P(\bar{A}) = 1 - P(A)$ .
6.  $P(B \setminus A) = P(B) - P(B \cap A)$ .
7.  $A \subset B \Rightarrow P(A) \leq P(B)$ .

## Probabilité uniforme sur $\Omega$ fini

- Soit  $\Omega$  un univers fini. On dit que  $P$  est la **probabilité uniforme** sur l'espace probabilisable  $(\Omega, P(\Omega))$  si:

$$\forall \omega, \omega' \in \Omega, \quad P(\{\omega\}) = P(\{\omega'\})$$

On dit aussi qu'il y a **équiprobabilité** des événements élémentaires.

- Soit  $(\Omega, \mathcal{P}(\Omega), P)$  un espace probabilisé fini. Si  $P$  est la probabilité uniforme, alors

$$\forall A \in \mathcal{A}, \quad P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}$$



- Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $B \in \mathcal{A}$  tel que  $P(B) > 0$ . L'application  $P_B$  définie sur  $\mathcal{A}$  par:

$$P_B(A) = P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \forall A \in \mathcal{A}$$

est une probabilité sur  $(\Omega, \mathcal{A})$ ; elle est appelée la **probabilité conditionnelle** sachant  $B$ . C'est la probabilité pour que l'événement  $A$  se produise sachant que l'événement  $B$  s'est produit.

- **Remarque:**  $(A|B)$  n'est pas un événement! On utilise la notation  $P(A|B)$  par simplicité, mais c'est  $P_B(A)$  qui est correcte.
- **Formule des probabilités composées:**

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

- **Formule des probabilités totales:**
- $\forall A \in \mathcal{A}, \quad P(A) = P(A \cap B) + P(A \cap \bar{B})$
  - On appelle **système complet d'événements (SCE)**, toute partition dénombrable de  $\Omega$  formée d'éléments de  $\mathcal{A}$ ; c-à-d tout ensemble dénombrable d'événements, deux à deux incompatibles et dont l'union dénombrable est l'événement certain.
  - Soit  $(B_n)_{n \geq 0}$  un SCE de  $\Omega$ . On a:

$$\forall A \in \mathcal{A}, \quad P(A) = \sum_{n \geq 0} P(A \cap B_n)$$

- **Indépendance:** Les événements  $A$  et  $B$  sont indépendants ssi  $P(A \cap B) = P(A)P(B)$ .

### Première formule de Bayes

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé. Pour tous événements  $A$  et  $B$  tels que  $P(A) \neq 0$  et  $P(B) \neq 0$ , on a :

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

### Deuxième formule de Bayes

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $(B_n)_{n \geq 0}$  une SCE de  $\Omega$  t.q. pour tout  $n \geq 0$   $P(B_n) \neq 0$ . On a pour tout  $A \in \mathcal{A}$  t.q.  $P(A) \neq 0$

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{n \geq 0} P(A|B_n)P(B_n)} \quad \forall i \geq 0$$

## Notion de variable aléatoire réelle

---

### Définition

*Soient  $\varepsilon$  une expérience aléatoire et  $(\Omega, \mathcal{A}, P)$  un espace probabilisé lié à cette expérience. Dans de nombreuses situations, on associe à chaque résultat  $\omega \in \Omega$  un nombre réel noté  $X(\omega)$ ; on construit ainsi une application  $X : \Omega \rightarrow \mathbb{R}$ . Historiquement,  $\varepsilon$  était un jeu et  $X$  représentait le gain du joueur.*

### Exemple: Jeu de dé

Un joueur lance un dé équilibré à 6 faces numérotées de 1 à 6, et on observe le numéro obtenu.

- ▶ Si le joueur obtient 1, 3 ou 5, il gagne 1 euro.
- ▶ S'il obtient 2 ou 4, il gagne 5 euros.
- ▶ S'il obtient 6, il perd 10 euros.

## Analyse

- ▶  $\varepsilon$ : “lancer d'un dé équilibré”.
- ▶  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .
- ▶  $\mathcal{A} = \mathcal{P}(\Omega)$ .
- ▶  $P$  l'équiprobabilité sur  $(\Omega, \mathcal{A})$ .

Soit  $X$  l'application de  $\Omega$  dans  $\mathbb{R}$  qui à tout  $\omega \in \Omega$  associe le gain correspondant.

On a donc

- ▶  $X(1) = X(3) = X(5) = 1$
- ▶  $X(2) = X(4) = 5$
- ▶  $X(6) = -10$

On dit que  $X$  est une **variable aléatoire** sur  $\Omega$ .

On peut s'intéresser à la probabilité de gagner 1 euro:

$$\Rightarrow X(\omega) = 1.$$

- ▶ ce qui se réalise si et seulement si  $\omega \in \{1, 3, 5\}$ .
- ▶ La probabilité cherchée est donc  $P(\{1, 3, 5\}) = 1/2$ .
- ▶ On écrira aussi  $P(X = 1) = 1/2$ .

On pourra donc considérer l'événement:

$$\{X = 1\} = \{\omega \in \Omega / X(\omega) = 1\} = \{\omega \in \Omega / X(\omega) \in \{1\}\} = X^{-1}(\{1\}) = \{1, 3, 5\}.$$

On aura du même:

- ▶  $P(X = 5) = 1/3$ .
- ▶  $P(X = -10) = 1/6$ .

On peut présenter les probabilités précédentes dans un tableau:

$x_i$	-10	1	5
$p_i = P(X = x_i)$	1/6	1/2	1/3

Cela revient à considérer un nouvel ensemble d'événements élémentaires:

$$\Omega_X = X(\Omega) = \{-10, 1, 5\}$$

et à munir cet ensemble de la probabilité  $P_X$  définie par le tableau des  $P(X = x_i)$  ci dessus. Cette nouvelle probabilité s'appelle **loi de la variable aléatoire**  $X$ .

Remarquer que

$$P\left(\bigcup_{x_i \in \Omega_X} \{X = x_i\}\right) = \sum_{x_i \in \Omega_X} P(X = x_i) = 1$$

Dans ce chapitre:

- ▶ Nous traitons le cas où  $X(\Omega)$  est dénombrable.
- ▶ La variable aléatoire est alors dite *discrète*.
- ▶ Nous définirons sa loi de probabilité par les probabilités individuelles.
- ▶ **NB:** Sa loi de probabilité peut être toujours définie par sa fonction de répartition.
- ▶ Nous définirons les deux caractéristiques numériques principales d'une variable aléatoire discrète:
  - L'**Espérance**: caractéristique de valeur centrale.
  - La **Variance**: caractéristique de dispersion.
- ▶ Nous définirons aussi les *couples* de variables aléatoires.



## **Variables aléatoires discrètes**

---

**Définition**

On dit qu'une variable aléatoire réelle (v.a.r.)  $X$  est **discrète** (v.a.r.d.) si l'ensemble des valeurs que prend  $X$  est fini ou infini dénombrable.

Si on suppose  $X(\Omega)$  l'ensemble des valeurs de  $X$  qui admet un plus petit élément  $x_1$ . Alors la v.a.r.d.  $X$  est entièrement **définie** par:

- ▶ L'ensemble  $X(\Omega)$  des valeurs prises par  $X$ , rangées par ordre croissant:  $X(\Omega) = \{x_1, x_2, \dots, x_i, \dots\}$  avec  $x_1 \leq x_2 \leq \dots \leq x_i \leq \dots$
- ▶ La **loi de probabilité** définie sur  $X(\Omega)$  par

$$p_i = p(x_i) = P(X = x_i) \quad \forall i = 1, 2, \dots$$

**Remarques:**

- ▶  $B \subset \mathbb{R}$ ,  $P(X \in B) = \sum_{i/x_i \in B} p(x_i)$ .
- ▶  $P(a < X \leq b) = \sum_{i/a < x_i \leq b} p(x_i)$ .
- ▶  $p(x_i) \geq 0$  et  $\sum_{i=1}^{\infty} p(x_i) = 1$ .
- ▶ Si  $X$  ne prend qu'un petit nombre de valeurs, cette loi est généralement présentée dans un tableau.

**Définition**

On appelle fonction de répartition de la v.a.  $X$ , qu'on note  $F(\alpha)$  de la v.a.r.d.  $X$ , ou  $F_X(\alpha)$ , la fonction définie pour tout réel  $\alpha$ ,  $-\infty < \alpha < \infty$ , par

$$F(\alpha) = P(X \leq \alpha) = \sum_{i/x_i \leq \alpha} P(X = x_i)$$

Cette valeur représente la probabilité de toutes les réalisations inférieures ou égales au réel  $\alpha$ .

**Propriétés**

1. C'est une fonction en escalier (constante par morceaux).
2.  $F(\alpha) \leq 1$  car c'est une probabilité.
3.  $F(\alpha)$  est continue à droite.
4.  $\lim_{\alpha \rightarrow -\infty} F(\alpha) = 0$  et  $\lim_{\alpha \rightarrow \infty} F(\alpha) = 1$

La fonction de répartition caractérise la loi de  $X$ , autrement dit:  $F_X = F_Y$  si et seulement si les variables aléatoires  $X$  et  $Y$  ont la même loi de probabilité.

Tous les calculs de probabilité concernant  $X$  peuvent être traités en termes de fonction de répartition. Par exemple,

$$P(a < X \leq b) = F(b) - F(a) \quad \text{pour tout } a < b$$

On peut mieux s'en rendre compte en écrivant  $\{X \leq b\}$  comme union des deux événements incompatibles  $\{X \leq a\}$  et  $\{a < X \leq b\}$ , soit

$$\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$$

et ainsi

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

ce qui établit l'égalité ci dessus.

### Remarque

On peut déduire de  $F$  les probabilités individuelles par:

$$p_i = P(X = x_i) = F(x_i) - F(x_{i-1}) \quad \text{pour } 1 \leq i \leq n$$

## Exemple

On joue trois fois à pile ou face  $\Rightarrow$

- ▶  $\Omega = \{P, F\}^3$ .
- ▶  $\text{card}(\Omega) = |\Omega| = 2^3 = 8$ .

Soit  $X$  la variable aléatoire "nombre de pile obtenus"  $\Rightarrow X(\Omega) = \{0, 1, 2, 3\}$ .

- ▶ Calculons par exemple  $P(X = 1)$ .
  - ▶  $X^{-1}(1) = \{(P, F, F), (F, P, F), (F, F, P)\}$ .
- $\Rightarrow P(X = 1) = \frac{3}{8}$

En procédant de la même façon, on obtient la loi de probabilité de  $X$ :

$k$	0	1	2	3
$P(X = k)$	1/8	3/8	3/8	1/8

La fonction de répartition de  $X$  est donc donnée par:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1/8 & \text{si } 0 \leq x < 1 \\ 1/2 & \text{si } 1 \leq x < 2 \\ 7/8 & \text{si } 2 \leq x < 3 \\ 1 & \text{si } x \geq 3 \end{cases}$$

On peut présenter la fonction de répartition dans le tableau de la loi de probabilité de  $X$ :

$k$	0	1	2	3
$P(X = k)$	1/8	3/8	3/8	1/8
$F_X(x)$	1/8	1/2	7/8	1

Le graphe de cette dernière est représentée dans la figure suivante:

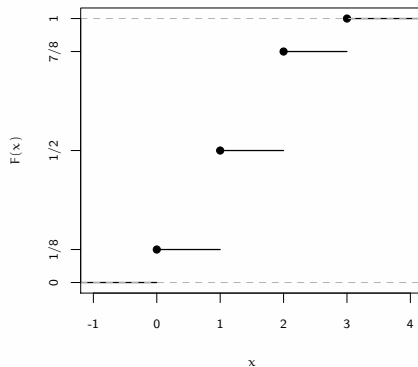


Figure 1: Fonction de répartition

Une autre représentation de la fonction de répartition:

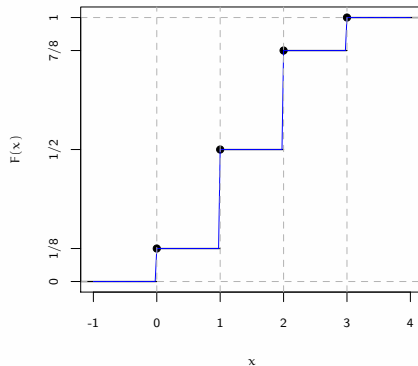


Figure 2: Fonction de répartition



**Définition**

Soit  $A$  un événement quelconque. On appelle *variable aléatoire indicatrice* de cet événement  $A$ , la variable aléatoire notée  $X = \mathbb{1}_A$  et définie par:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in \bar{A} \end{cases}$$

Ainsi:

- ▶  $P(X = 1) = P(A) = p$
- ▶  $P(X = 0) = P(\bar{A}) = 1 - p$

La fonction de répartition de  $X$  est donc donnée par:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

## Exemple

- ▶ Soit  $\mathcal{U}$  une urne contenant 2 boules blanches et 3 boules noires.
- ▶ On tire une boule au hasard.
- ▶ Soit  $A$  : “obtenir une boule blanche”.
- ▶ Soit  $X$  la variable indicatrice de  $A$ .

Déterminez la loi de probabilité de  $X$  ainsi que sa Fonction de répartition.

La loi de probabilité de  $X$  est

$k$	0	1
$P(X = k)$	$\frac{3}{5}$	$\frac{2}{5}$

et sa fonction de répartition est:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 3/5 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

## Moments d'une variable aléatoire discrète

---

**Définition**

Pour une variable aléatoire discrète  $X$  de loi de probabilité  $p(\cdot)$ , on définit l'espérance de  $X$ , notée  $E(X)$ , par l'expression

$$E(X) = \sum_{i \in \mathbb{N}} x_i p(x_i)$$

En termes concrets, l'espérance de  $X$  est la moyenne pondérée des valeurs que  $X$  peut prendre, les poids étant les probabilités que ces valeurs soient prises.

**Exemples**

1. Dans l'exemple où on joue 3 fois à pile ou face. L'espérance de  $X$  = "nombre de pile obtenus" est égal à :

$$E(X) = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1.5$$

2. Pour la variable aléatoire indicatrice de  $A$  :

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = P(A) = p$$

Ceci signifie que l'espérance de la variable indicatrice pour l'événement  $A$  est égale à la probabilité que  $A$  se produise.

**Théorème**

Si  $X$  est une variable aléatoire discrète pouvant prendre ses valeurs parmi les valeurs  $x_i$ ,  $i \geq 1$ , avec des probabilités respectives  $p(x_i)$ , alors pour toute fonction réelle  $g$  on a

$$E(g(X)) = \sum_i g(x_i)p(x_i)$$

**Exemple**

Soit  $X$  une variable aléatoire qui prend une des trois valeurs  $\{-1, 0, 1\}$  avec les probabilités respectives

$$P(X = -1) = 0.2 \quad P(X = 0) = 0.5 \quad P(X = 1) = 0.3$$

Calculer  $E(X^2)$ .

**Solution**

*Première approche:* Soit  $Y = X^2$ . La distribution de  $Y$  est donnée par

$$P(Y = 1) = P(X = -1) + P(X = 1) = 0.5$$

$$P(Y = 0) = P(X = 0) = 0.5$$

Donc

$$E(X^2) = E(Y) = 1(0.5) + 0(0.5) = 0.5$$

*Deuxième approche:* En utilisant le théorème

$$E(X^2) = (-1)^2(0.2) + 0^2(0.5) + 1^2(0.3)$$

$$= 1(0.2 + 0.3) + 0(0.5) = 0.5$$

**Remarque**

$$0.5 = E(X^2) \neq (E(X))^2 = 0.01$$

## Propriétés

1.  $E(X + a) = E(X) + a$ ,  $a \in \mathbb{R}$

résultat qui se déduit de:

$$\sum_i p_i (x_i + a) = \sum_i p_i x_i + \sum_i a p_i = \sum_i p_i x_i + a \sum_i p_i = \sum_i p_i x_i + a$$

2.  $E(aX) = aE(X)$ ,  $a \in \mathbb{R}$

il suffit d'écrire:

$$\sum_i p_i a x_i = a \sum_i p_i x_i$$

3.  $E(X + Y) = E(X) + E(Y)$ ,  $X$  et  $Y$  étant deux variables aléatoires.

On peut résumer ces trois propriétés en disant que l'espérance mathématique est linéaire:

$$E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y), \quad \forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R}.$$

**Définition**

Il s'agit d'un indicateur mesurant la dispersion des valeurs  $x_i$  que peut prendre la v.a.  $X$  et son espérance  $E(X)$ . On appelle **variance** de  $X$ , que l'on note  $V(X)$ , la quantité, lorsqu'elle existe,

$$V(X) = E[(X - E(X))^2]$$

C'est l'espérance mathématique du carré de la v.a. centrée  $X - E(X)$ .

**Remarque**

On peut établir une autre formule pour le calcul de  $V(X)$ :

$$V(X) = E(X^2) - E^2(X)$$

Or:

$$\begin{aligned} V(X) &= E[X^2 - 2XE(X) + E^2(X)] \\ &= E(X^2) - E[2XE(X)] + E[E^2(X)] \\ &= E(X^2) - 2E^2(X) + E^2(X) \\ &= E(X^2) - E^2(X) \end{aligned}$$



## Exemple

On cherche  $V(X)$  où  $X$  est le nombre obtenu lors du jet d'un dé équilibré.

On a vu précédemment que  $E(X) = \frac{7}{2}$ . De plus,

$$\begin{aligned} E(X^2) &= \sum_i x_i^2 p(x_i) \\ &= 1^2 \left(\frac{1}{6}\right) + 2^2 \left(\frac{1}{6}\right) + 3^2 \left(\frac{1}{6}\right) + 4^2 \left(\frac{1}{6}\right) + 5^2 \left(\frac{1}{6}\right) + 6^2 \left(\frac{1}{6}\right) \\ &= \left(\frac{1}{6}\right)(91) = \frac{91}{6} \end{aligned}$$

Et donc

$$\begin{aligned} V(X) &= E(X^2) - E^2(X) \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \end{aligned}$$

## Propriétés

1.  $V(X) \geq 0$

2.  $\forall a \in \mathbb{R}, \quad V(X + a) = V(X)$

en effet:

$$\begin{aligned} V(X + a) &= E[ [X + a - E(X + a)]^2 ] \\ &= E[ [X + a - E(X) - a]^2 ] \\ &= E[ [X - E(X)]^2 ] = V(X) \end{aligned}$$

3.  $\forall a \in \mathbb{R}, \quad V(aX) = a^2 V(X)$

en effet:

$$\begin{aligned} V(aX) &= E[ [aX - E(aX)]^2 ] \\ &= E[ [aX - aE(X)]^2 ] \\ &= E[ a^2 [X - E(X)]^2 ] \\ &= a^2 E[ [X - E(X)]^2 ] = a^2 V(X) \end{aligned}$$

**Définition**

La racine carrée de  $V(X)$  est appelée l'**écart-type** de  $X$ , qui se note  $\sigma_X$ <sup>1</sup>. On a

$$\sigma_X = \sqrt{V(X)}$$

$\sigma_X$  s'exprime dans les mêmes unités de mesure que la variable aléatoire  $X$ .

- ▶ L'écart type sert à mesurer la dispersion d'un ensemble de données.
- ▶ Plus il est faible, plus les valeurs sont regroupées autour de la moyenne.
- ▶ **Exemple**: La répartition des notes d'une classe. Plus l'écart type est faible, plus la classe est homogène.
- ▶ L'espérance et l'écart-type sont reliés par l'*inégalité de Bienaymé-Tchebychev*.

---

<sup>1</sup>ou  $\sigma(X)$ , ou tout simplement  $\sigma$ .

**Théorème**

Soit  $X$  une variable aléatoire d'espérance  $\mu$  et de variance  $\sigma^2$ . Pour tout  $\varepsilon > 0$ , on a l'inégalité suivante:

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

**Remarque**

On peut l'écrire autrement. Soit  $k = \varepsilon/\sigma$ .

$$P(|X - E(X)| \geq k\sigma) \leq \frac{1}{k^2}$$

**Importance**

Cette inégalité relie la probabilité pour  $X$  de s'écarter de sa moyenne  $E(X)$ , à sa variance qui est justement un indicateur de dispersion autour de la moyenne de la loi. Elle montre quantitativement que "plus l'écart type est faible, plus la probabilité de s'écarter de la moyenne est faible".

**Définition**

On appelle *moment non centré* d'ordre  $r \in \mathbb{N}^*$  de  $X$  la quantité, lorsqu'elle existe:

$$m_r(X) = \sum_{i \in \mathbb{N}} x_i^r p(x_i) = E(X^r).$$

**Définition**

Le *moment centré* d'ordre  $r \in \mathbb{N}^*$  est la quantité, lorsqu'elle existe:

$$\mu_r(X) = \sum_{i \in \mathbb{N}} p_i [x_i - E(X)]^r = E[X - E(X)]^r.$$

**Remarque**

Les premiers moments sont:

- ▶  $m_1(X) = E(X), \quad \mu_1(X) = 0.$
- ▶  $\mu_2(X) = V(X) = m_2(X) - m_1^2(X).$

## Couple de variables aléatoires discrètes

---

Nous avons traité jusqu'ici des variables isolées. Or, il est souvent nécessaire de considérer des événements relatifs à deux variables simultanément, ou même à plus de deux variables.

### Définition

Soit  $X$  et  $Y$  deux variables aléatoires réelles discrètes, définies sur un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et que  $X(\Omega) = \{x_1, x_2, \dots, x_l\}$  et  $Y(\Omega) = \{y_1, y_2, \dots, y_k\}$ ,  $l$  et  $k \in \mathbb{N}$ .

La **loi du couple**  $(X, Y)$  est entièrement définie par les probabilités:

$$p_{ij} = P(X = x_i; Y = y_j) = P(\{X = x_i\} \cap \{Y = y_j\})$$

On a

$$p_{ij} \geq 0 \quad \text{et} \quad \sum_{i=1}^l \sum_{j=1}^k p_{ij} = 1$$

Le couple  $(X, Y)$  s'appelle variable aléatoire à deux dimensions et peut prendre  $l \times k$  valeurs.

Les probabilités  $p_{ij}$  peuvent être présentées dans un tableau à deux dimensions qu'on appelle table de probabilité conjointe:

*Table 1: Table de probabilité conjointe*

$X \backslash Y$	$y_1$	$y_2$	$\dots$	$y_j$	$\dots$	$y_k$
$x_1$	$p_{11}$	$p_{12}$		$p_{1j}$		$p_{1k}$
$x_2$	$p_{21}$	$p_{22}$		$p_{2j}$		$p_{2k}$
$\vdots$						
$x_i$	$p_{i1}$	$p_{i2}$		$p_{ij}$		$p_{ik}$
$\vdots$						
$x_l$	$p_{l1}$	$p_{l2}$		$p_{lj}$		$p_{lk}$

A la première ligne figure l'ensemble des valeurs de  $Y$  et à la première colonne figure l'ensemble des valeurs de  $X$ . La probabilité  $p_{ij} = P(X = x_i; Y = y_j)$  est à l'intersection de la  $i^e$  et de la  $j^e$  colonne.



### Exemple

On tire au hasard 3 boules d'une urne contenant 3 boules rouges, 4 blanches et 5 noires.  $X$  et  $Y$  désignent respectivement le nombre de boules rouges et celui de boules blanches tirées. Déterminer la loi de probabilité conjointe du couple  $(X, Y)$ .

### Solution

- ▶  $\varepsilon$ : "tirer 3 boules d'une urne contenant 12 boules".
- ▶  $|\Omega| = C_{12}^3 = 220$ .
- ▶  $X(\Omega) = \{0, 1, 2, 3\}$  et  $Y(\Omega) = \{0, 1, 2, 3\}$ .
- ▶  $p(X = 0, Y = 0) = p(0, 0) = C_5^3 / C_{12}^3 = \frac{10}{220}$ .
- ▶  $p(0, 1) = C_4^1 C_5^2 / C_{12}^3 = \frac{40}{220}$ .
- ▶  $p(1, 0) = C_3^1 C_5^2 / C_{12}^3 = \frac{30}{220}$ .

## Exemple

On tire au hasard 3 boules d'une urne contenant 3 boules rouges, 4 blanches et 5 noires.  $X$  et  $Y$  désignent respectivement le nombre de boules rouges et celui de boules blanches tirées. Déterminer la loi de probabilité conjointe du couple  $(X, Y)$ .

## Solution

*Table 2: Table de probabilité conjointe*

$X \backslash Y$	0	1	2	3
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0
3	$\frac{1}{220}$	0	0	0

Lorsqu'on connaît la loi conjointe des variables aléatoires  $X$  et  $Y$ , on peut aussi s'intéresser à la loi de probabilité de  $X$  seule et de  $Y$  seule. Ce sont les lois de probabilité marginales.

► Loi marginale de  $X$ :

$$p_{i.} = P(X = x_i) = P[\{X = x_i\} \cap \Omega] = \sum_{j=1}^k p_{ij} \quad \forall i = 1, 2, \dots, l$$

► Loi marginale de  $Y$ :

$$p_{.j} = P(Y = y_j) = P[\Omega \cap \{Y = y_j\}] = \sum_{i=1}^l p_{ij} \quad \forall j = 1, 2, \dots, k$$

On peut calculer les lois marginales directement depuis la table de la loi conjointe.

*Table 3: Table de probabilité conjointe avec les lois marginales*

$X \backslash Y$	$y_1$	$y_2$	$\dots$	$y_j$	$\dots$	$y_k$	Marginale de $X$
$x_1$	$p_{11}$	$p_{12}$		$p_{1j}$		$p_{1k}$	$p_{1.}$
$x_2$	$p_{21}$	$p_{22}$		$p_{2j}$		$p_{2k}$	$p_{2.}$
$\vdots$							
$x_i$	$p_{i1}$	$p_{i2}$		$p_{ij}$		$p_{ik}$	$p_{i.}$
$\vdots$							
$x_l$	$p_{l1}$	$p_{l2}$		$p_{lj}$		$p_{lk}$	$p_{l.}$
Marginale de $Y$	$p_{.1}$	$p_{.2}$		$p_{.j}$		$p_{.k}$	1

## Exemple

On tire au hasard 3 boules d'une urne contenant 3 boules rouges, 4 blanches et 5 noires.  $X$  et  $Y$  désignent respectivement le nombre de boules rouges et celui de boules blanches tirées. Déterminer les lois marginales de  $X$  et  $Y$ .

## Solution

*Table 4: Table de probabilité conjointe*

$X \backslash Y$	0	1	2	3	$p_{i.} = P(X = x_i)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
$p_{.j} = P(Y = y_j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

**Définition**

Pour chaque valeur  $y_j$  de  $Y$  telle que  $p_{\cdot j} = P(Y = y_j) \neq 0$  on peut définir la loi conditionnelle de  $X$  sachant  $Y = y_j$  par

$$p_{i/j} = P(X = x_i / Y = y_j) = \frac{P(X = x_i; Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}} \quad \forall i = 1, 2, \dots, l$$

De même on définit la loi de  $Y$  sachant  $X = x_i$  par

$$p_{j/i} = P(Y = y_j / X = x_i) = \frac{P(X = x_i; Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i\cdot}} \quad \forall j = 1, 2, \dots, k$$

**Définition**

*On dit que deux v.a.r.d sont indépendantes si et seulement si*

$$P(X = x_i; Y = y_j) = P(X = x_i)P(Y = y_j) \quad \forall i = 1, 2, \dots, l \text{ et } j = 1, 2, \dots, k$$

*On montre que*

$$P(\{X \in A\} \cap \{Y \in B\}) = P(\{X \in A\})P(\{Y \in B\}) \quad \forall A \text{ et } B \in \mathcal{A}$$

**Propriétés**

Soit deux v.a.r.d.  $X$  et  $Y$ ,

1.  $E(X + Y) = E(X) + E(Y)$
2. Si  $X$  et  $Y$  sont indépendantes alors  $E(XY) = E(X)E(Y)$ . Mais la réciproque n'est pas toujours vraie.

**Définition**

Soit  $X$  et  $Y$  deux v.a.r.d. On appelle **covariance** de  $X$  et de  $Y$  la valeur si elle existe de

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = \sum_i \sum_j (x_i - E(X))(y_j - E(Y))p_{ij}$$

qu'on peut calculer en utilisant la formule suivante

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

**Propriétés**

- ▶  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- ▶  $\text{Cov}(aX_1 + bX_2, Y) = a\text{Cov}(X_1, Y) + b\text{Cov}(X_2, Y)$
- ▶  $V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X, Y)$
- ▶ Si  $X$  et  $Y$  sont indépendantes alors
  - $\text{Cov}(X, Y) = 0$  (la réciproque n'est pas vraie)
  - $V(X + Y) = V(X) + V(Y)$  (la réciproque n'est pas vraie)



**Définition**

On appelle coefficient de corrélation linéaire de  $X$  et de  $Y$  la valeur définie par

$$\rho = \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

On peut montrer que

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

**Interprétation de  $\rho$** 

- ▶ Le coefficient de corrélation est une mesure du degré de linéarité entre  $X$  et  $Y$ .
- ▶ Les valeurs de  $\rho$  proches de 1 ou  $-1$  indiquent une linéarité quasiment rigoureuse entre  $X$  et  $Y$ .
- ▶ Les valeurs de  $\rho$  proche de 0 indiquent une absence de toute relation linéaire.
- ▶ Lorsque  $\rho(X, Y)$  est positif,  $Y$  a tendance à augmenter si  $X$  en fait autant.
- ▶ Lorsque  $\rho(X, Y) < 0$ ,  $Y$  a tendance à diminuer si  $X$  augmente.
- ▶ Si  $\rho(X, Y) = 0$ , on dit que ces deux statistiques sont non corrélées.

