

# Cours de probabilités pour l'informatique

5 décembre 2013

# Chapitre 1

## Ensembles, Cardinaux, Dénombrément

### 1 Ensembles

Un ensemble  $\Omega$  est une collection d'objets appelés éléments de cet ensemble. On écrit  $\omega \in \Omega$  pour exprimer que  $\omega$  est un élément de  $\Omega$  : C'est la relation d'appartenance. Deux éléments  $\omega$  et  $\omega'$  de  $\Omega$ , sont égaux et on écrit  $\omega = \omega'$ , ou différents et on écrit  $\omega \neq \omega'$ . L'ensemble vide, noté  $\emptyset$  ne contient aucun élément.

**Définition 1.** Un ensemble  $A$  est appelé sous-ensemble ou partie de  $\Omega$  si tout élément de  $A$  est élément de  $\Omega$ . On écrit  $A \subset \Omega$  pour exprimer que  $A$  est une partie de  $\Omega$  : C'est la relation d'inclusion. Un élément  $\omega \in \Omega$  peut appartenir ou ne pas appartenir à la partie  $A$ . On écrit alors respectivement  $\omega \in A$  ou  $\omega \notin A$ . Le complémentaire de  $A$ , noté  $A^c$ , est l'ensemble des éléments  $x \in \Omega$  tels que  $x \notin A$ .

**Définition 2.** La fonction indicatrice d'une partie  $A$  est l'application  $1_A : \Omega \rightarrow \{0; 1\}$  définie par  $1_A(\omega) = 1$  si  $\omega \in A$  et  $1_A(\omega) = 0$  si  $\omega \notin A$ .

**Définition 3.** L'ensemble des parties de  $\Omega$  est noté  $P(\Omega)$ . Une famille  $F$  de parties de  $\Omega$  est une partie de  $P(\Omega)$ . Les éléments de  $F$  sont des parties de  $\Omega$ .

*Remarque 1.* Si  $A$  et  $B$  sont deux parties de  $\Omega$  (i.e. deux éléments de  $P(\Omega)$ ).

1. On a les relations  $A \subset B$  ou  $B \subset A$  ou ( $A \not\subset B$  et  $B \not\subset A$ ).  $A \subset B$  s'écrit aussi  $B \supset A$ .
2. On peut définir  $A \times B$  l'ensemble des couples  $(a, b)$   $a \in A, b \in B$ , l'intersection  $A \cap B$ , l'union  $A \cup B$ , le complémentaire de  $B$  dans  $A - B = A \cap B^c$  et la différence symétrique  $A \Delta B = (A - B) \cup (B - A)$ .
3.  $A$  et  $B$  sont disjoints si  $A \cap B = \emptyset$ .

#### 1.1 Cardinaux

Soient  $A$  et  $B$  deux ensembles.

**Proposition 1.** *Il existe une application injective  $f : A \rightarrow B$  si et seulement s'il existe une application surjective  $g : B \rightarrow A$ . On écrit alors  $\text{Card}(A) \leq \text{Card}(B)$ . On a à la fois  $\text{Card}(A) \leq \text{Card}(B)$  et  $\text{Card}(A) \geq \text{Card}(B)$  si et seulement s'il existe une application bijective  $f : A \rightarrow B$ . On dit alors que  $A$  et  $B$  sont équipotents ou ont même cardinal et on écrit  $\text{Card}(A) = \text{Card}(B)$ .*

## 2 Ensembles finis

On dit qu'un ensemble  $A$  est fini s'il est vide et on écrit  $\text{Card}(A) = 0$  ou s'il existe un entier  $n \in \mathbb{N}^*$ , qui est alors unique, et une bijection  $f : \llbracket 1, n \rrbracket \rightarrow A$ . On écrit alors  $\text{Card}(A) = n$  ou  $|A| = n$ . Ainsi, pour un ensemble fini  $A$ , l'entier  $n = \text{Card}(A)$  représente le nombre d'éléments de  $A$  et la bijection  $f : \llbracket 1, n \rrbracket \rightarrow A$  permet de représenter les éléments de  $A$  à l'aide d'une suite finie et d'écrire  $A = \{x_k; 1 \leq k \leq n\}$ .

## 3 Opérations sur les ensembles finis et leurs cardinaux

**Proposition 2.** *Toute partie d'un ensemble fini est finie et toute intersection d'ensembles finis est finie.*

**Proposition 3.** *Un produit fini d'ensembles finis est fini. Si  $A$  et  $B$  sont finis,  $\text{Card}(A \times B) = \text{Card}(A)\text{Card}(B)$ . Plus généralement, si  $(A_k)_{1 \leq k \leq n}$  est une suite finie d'ensembles finis, on a*

$$\text{Card}\left(\prod_{k=1}^n A_k\right) = \prod_{k=1}^n \text{Card}(A_k).$$

**Proposition 4.** *L'ensemble des applications d'un ensemble  $A$  vers un ensemble  $B$  est souvent noté  $B^A$ . Si  $A$  et  $B$  sont finis, il en est de même de  $B^A$  et l'on a*

$$\text{Card}(B^A) = \text{Card}(B)^{\text{Card}(A)}.$$

**Proposition 5.** *(Principe du berger). Soit  $f : A \rightarrow B$  une application d'un ensemble fini  $A$  vers un ensemble fini  $B$ . On suppose qu'il existe  $m \in \mathbb{N}^*$  tel que, pour tout  $b \in B$ ;  $\text{Card}(f^{-1}b) = m$ . Alors,  $\text{Card}(A) = m\text{Card}(B)$ .*

**Proposition 6.** *Une réunion finie d'ensembles finis est finie et l'on a  $\text{Card}(\cup_{k=1}^n A_k) \leq \sum_{k=1}^n \text{Card}(A_k)$ , l'égalité ayant lieu si et seulement si les ensembles  $A_k$  sont deux à deux disjoints.*

Le résultat suivant traite le cas d'ensembles non disjoints.

**Proposition 7.** *Si  $A$  et  $B$  sont finis,*

$$\text{Card}(A \cup B) = \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B).$$

Plus généralement, si  $(A_k)_{1 \leq k \leq n}$  est une suite finie d'ensembles finis, on a la relation suivante

$$\text{Card}(\cup_{k=1}^n A_k) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \text{Card}(A_{i_1} \cap \dots \cap A_{i_k}).$$

Cette relation est connue sous le nom de Formule du crible ou de Poincaré ou encore Principe d'inclusion-exclusion.

*Exemple 1.* Sur 30 clients ayant commandé un article  $A \cup B$ , 20 l'ont fait par écrit ( $A$ ) et 17 l'ont fait par téléphone ( $B$ ). Le nombre de ceux qui ont utilisé les deux moyens est  $|A \cap B| = |A| + |B| - |A \cup B| = 20 + 17 - 30 = 7$ .

*Exemple 2.* Le nombre de délégations de 3 membres comprenant un représentant de chacun des trois groupes de 10; 15 et 13 étudiants est  $10 \cdot 15 \cdot 13 = 1950$ .

*Exemple 3.* On lance un dé à 6 faces puis une pièce de monnaie. Le nombre de résultats possibles est  $6 \cdot 2 = 12$ .

*Exemple 4.* En considérant la bijection entre les parties d'un ensemble et leurs fonctions indicatrices, on voit que le nombre de parties d'un ensemble à  $n$  éléments est  $2^n$ .

## 4 Ensembles infinis

Un ensemble  $A$  qui n'est pas fini est dit infini.

### 4.1 Ensembles infinis dénombrables

**Définition 4.** Un ensemble infini  $A$  est dénombrable s'il existe une bijection  $f : \mathbb{N} \rightarrow A$ .

On peut représenter les éléments d'un ensemble dénombrable  $A$  à l'aide d'une suite infinie en écrivant  $A = \{x_n; n \geq 1\}$ . Les ensembles  $\mathbb{N}$ ;  $\mathbb{Z}$  et  $\mathbb{Q}$  sont infinis dénombrables. Les ensembles finis et dénombrables serviront de base aux probabilités discrètes.

### 4.2 Ensembles infinis non dénombrables

Les ensembles qui n'appartiennent pas aux catégories précédentes (finis ou infinis dénombrables) sont dits infinis non dénombrables. Par exemple,  $\mathbb{R}$  et  $\mathbb{C}$ ,  $[a, b]$ ,  $a < b$  sont infinis non dénombrables.

## 5 Dénombrement

On appelle dénombrement l'ensemble des techniques qui permettent de déterminer le cardinal d'une partie finie ou d'une famille finie de parties ou d'applications ayant une propriété donnée.

### 5.1 Arrangement

**Définition 5.** On appelle *arrangement* de  $p$  éléments parmi  $n$  (ou de  $n$  éléments de pris  $p$  à  $p$ ), toute suite ordonnée de longueur  $p$  (ou  $p$ -liste) dont les termes sont des éléments *distincts* d'un ensemble  $\Omega$  de cardinal  $n$ . En d'autres termes, un tel arrangement est une injection  $f : \llbracket 1, p \rrbracket \rightarrow \Omega$ . Le nombre de ces arrangements, noté  $A_n^p$ , est

$$A_n^p = \frac{n!}{(n-p)!} = n(n-1) \cdots (n-p+1).$$

On pose  $A_n^p = 0$  si  $p > n$  ou si  $p < 0$ .

On considère des arrangements quand on parle de "tirage ordonné sans remise" ou "tirage ordonné d'éléments distincts".

*Exemple 5.* Le nombre de codes à 4 chiffres distincts est  $A_{10}^4 = 10 \times 9 \times 8 \times 7 = 5040$ .

*Exemple 6.* Le nombre de bureaux composés d'un président, d'un secrétaire et d'un trésorier qu'une assemblée de 17 personnes peut élire est  $A_{17}^3 = 17!/14! = 17 \times 16 \times 15 = 4080$ .

### 5.2 Permutation

**Définition 6.** On appelle permutation des éléments d'un ensemble  $\Omega$  de cardinal  $n$  toute suite ordonnée de longueur  $n$  (ou  $n$ -liste) dont les termes sont les éléments distincts de  $\Omega$ . En d'autres termes, une permutation est une bijection  $f : \llbracket 1, n \rrbracket \rightarrow \Omega$ . Le nombre de ces permutations est  $A_n^n = n!$ .

*Exemple 7.* Sur une carte, le nombre de façons de colorier 6 régions avec 6 couleurs est  $6! = 720$ .

*Exemple 8.* Une urne contient 7 jetons numérotés de 1 à 7. Ces jetons sont tirés successivement sans remise. Le nombre de tirages possibles est  $7! = 5040$ .

### 5.3 Combinaison

**Définition 7.** On appelle *combinaison de  $p$  éléments parmi  $n$*  (ou de  $n$  éléments de pris  $p$  à  $p$ ), tout ensemble de cardinal  $p$  dont les éléments appartiennent à un ensemble  $\Omega$  de cardinal  $n$ . En d'autres termes, une telle combinaison est une partie à  $p$  éléments de  $\Omega$ . Le nombre de ces combinaisons, noté  $C_n^p$  ou  $\binom{n}{p}$ , est

$$C_n^p = \frac{n!}{p!(n-p)!} = \frac{n(n-1)\cdots(n-p+1)}{p!}.$$

On pose  $C_n^p = 0$  si  $p > n$  ou si  $p < 0$ .

On considère des combinaisons quand on parle de "tirage non-ordonné sans remise" ou "tirage non-ordonné d'éléments distincts".

*Exemple 9.* Le nombre de listes classées par ordre alphabétique de 4 noms parmi 11 est  $C_{11}^4 = \frac{11 \times 10 \times 9 \times 8}{4!} = 330$ .

*Exemple 10.* Le nombre de choix possibles pour désigner 5 places visiteurs dans un parking de 31 places est  $C_{31}^5 = \frac{31!}{5!(26)!} = 169911$ .

### 5.4 Propriétés des coefficients $C_n^p$

**Proposition 8.** *Formule du binôme de Newton.*  $(x + y)^n = \sum_{k=0}^n C_n^k x^k y^{n-k}$ ,  $x$  et  $y$  des nombres complexes (ou deux éléments d'un anneau unitaire tels que  $xy = yx$ ).

En raison de cette formule, les nombres  $C_n^k$  sont appelés *coefficients du binôme*.

**Proposition 9.** *On a les relations :*  $C_n^0 = C_n^n = 1$ ,  $C_n^p = C_n^{n-p}$ ,  $pC_n^p = nC_{n-1}^{p-1}$ , et le triangle de Pascal.

$$C_n^p = C_{n-1}^p + C_{n-1}^{p-1}.$$

### 5.5 Arrangement avec répétition

**Définition 8.** Soit  $\Omega = \{a_1, \dots, a_n\}$  un ensemble à  $n$  éléments,  $p$  un entier et  $(r_1; \dots; r_n)$  une suite d'entiers telle que  $\sum_{k=1}^n r_k = p$ . On appelle *arrangement avec répétition*, de longueur  $p$  et d'ordre  $(r_1; \dots; r_n)$  toute suite ordonnée de longueur  $p$  (ou  $p$ -liste) dont les termes sont des éléments de  $\Omega$  où, pour tout  $k \in \llbracket 1, n \rrbracket$ , chaque  $a_k$  figure  $r_k$  fois. En d'autres termes, un tel arrangement est une application  $f : \llbracket 1, p \rrbracket \rightarrow \Omega$  telle que pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $Card(f^{-1}(a_k)) = r_k$ . Le nombre des ces arrangements est  $\frac{p!}{r_1! \dots r_n!}$ . C'est aussi le nombre de  $p$ -tirages avec remise dans une population de taille  $n$ , pour lesquels le même individu de la population apparaît  $r_k$  fois en tenant compte de l'ordre d'apparition.

Le nombre total de tous les arrangements avec répétition de  $p$  objets parmi  $n$  est  $n^p$ . Plus généralement, on a la formule du multinôme :

$$\sum_{r_1 + \dots + r_n = p, r_i \geq 0} \frac{p!}{r_1! \dots r_n!} x_1^{r_1} \dots x_n^{r_n} = (x_1 + \dots + x_n)^p.$$

*Exemple 11.* On veut ranger sur un rayon dix ouvrages comprenant 5 livres de chimie, 3 livres de mathématiques et 2 livres de physique. Le nombre de rangements possibles est celui des arrangements avec répétition d'ordre  $(5; 3; 2)$  de 10 objets parmi 3. Ce nombre est  $\frac{10!}{5!3!2!} = 2620$ .

*Exemple 12.* Dans une compétition, le nombre de classements où figurent 3 français, 4 anglais et 2 allemands est celui des arrangements avec répétition d'ordre  $(3; 4; 2)$  de 9 objets parmi 3. Ce nombre est  $\frac{9!}{3!4!2!} = 1260$ .

## 5.6 Combinaison avec répétition

**Définition 9.** Soit  $\Omega = \{a_1, \dots, a_n\}$  un ensemble à  $n$  éléments,  $p$  un entier et  $(r_1; \dots; r_n)$  une suite d'entiers telle que  $\sum_{k=1}^n r_k = p$ . On appelle *combinaison avec répétition*, de longueur  $p$  et d'ordre  $(r_1; \dots; r_n)$  tout ensemble de  $p$  objets où, pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $r_k$  objets sont égaux à  $a_k$ . Ce qui compte ici, c'est uniquement le nombre de fois où chaque objet apparaît, abstraction faite de l'ordre d'apparition. Il y a évidemment une seule combinaison avec répétition de ce type. Le nombre total de toutes les combinaisons avec répétition de  $p$  objets parmi  $n$ , noté  $K_n^p$ , est

$$K_n^p = C_{n+p-1}^p = C_{p+n-1}^{n-1}.$$

On peut noter qu'une combinaison avec répétition de  $p$  objets parmi  $n$  est l'image d'une application  $f : \llbracket 1, p \rrbracket \rightarrow \Omega$  telle que  $\sum_{k=1}^n \text{Card}(f^{-1}(a_k)) = p$ . C'est aussi le nombre de  $p$ -tirages avec remise dans une population de taille  $n$  où seul compte le nombre d'apparitions de chaque individu.

(Pour s'en souvenir, on comprend qu'une telle combinaison revient à séparer  $n$  classes (avec  $n-1$  délimiteurs) dans l'ensemble à  $p$  éléments, ce qui revient à choisir les  $n-1$  délimiteurs dans un ensemble à  $n-1+p$  éléments formés des délimiteurs et des éléments de base).

*Exemple 13.* On peut répartir 10 voix entre 3 candidats de  $K_{10}^3 = 66$  façons.

*Exemple 14.* On peut répartir 8 personnes selon leur groupe sanguin de  $K_8^4 = 55$  façons.

*Exemple 15.* Le développement (multinomial) de  $(a+b+c+d)^7$  comprend  $K_7^4 = 120$  termes.

## 5.7 Nombre de solutions de l'équation $\sum_{k=1}^n x_k = p$

**Proposition 10.** Le nombre de solutions de l'équation  $\sum_{k=1}^n x_k = p$  où, pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $x_k \in \mathbb{N}^*$ , est  $C_{p-1}^{n-1}$ .

*Démonstration.* Forcément  $p \geq n$  pour avoir une solution non nulle. Soit  $b_{n,p}$  le nombre de solutions. En faisant le produit de Cauchy de séries absolument convergentes (par récurrence) pour  $x < 1$  (série entière à l'intérieur du rayon de convergence) :

$$\frac{x^n}{(1-x)^n} = \sum_{x_1=1}^{\infty} \dots \sum_{x_n=1}^{\infty} x^{\sum_{k=1}^n x_k} = \sum_{p=n}^{\infty} \sum_{\sum_{k=1}^n x_k=p, x_k \geq 1} x^{\sum_{k=1}^n x_k} = \sum_{p=n}^{\infty} b_{n,p} x^p.$$

D'où par la formule de Taylor pour  $q \geq 0$

$$b_{n,q+n} = \left( \frac{1}{q!} \left( \frac{d}{dx} \right)^q \frac{1}{(1-x)^n} \right)_{x=0} = \left( \frac{1}{q!} \left( \frac{d}{dx} \right)^{p-1} \frac{n}{(1-x)^{n+1}} \right)_{x=0} = \frac{n(n+1)\dots(n+q-1)}{q!} = C_{n+q-1}^q = C_{n+q-1}^{n-1}.$$

□

**Proposition 11.** Le nombre de solutions de l'équation  $\sum_{k=1}^n x_k = p$  où, pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $x_k \in \mathbb{N}$ , est  $K_n^p = C_{n+p-1}^{n-1}$ .

*Démonstration.* Forcément  $p \geq 0$  pour avoir une solution non nulle. Soit  $b'_{n,p}$  le nombre de solutions. En faisant le produit de Cauchy de séries absolument convergentes (par récurrence) pour  $x < 1$  (série entière à l'intérieur du rayon de convergence) :

$$\frac{1}{(1-x)^n} = \sum_{x_1=0}^{\infty} \dots \sum_{x_n=0}^{\infty} x^{\sum_{k=1}^n x_k} = \sum_{p=0}^{\infty} \sum_{\sum_{k=1}^n x_k=p, x_i \geq 0} x^{\sum_{k=1}^n x_k} = \sum_{p=0}^{\infty} b'_{n,p} x^p.$$

D'où par la formule de Taylor pour  $q \geq 0$

$$b'_{n,q} = \left(\frac{1}{q!} \left(\frac{d}{dx}\right)^q \frac{1}{(1-x)^n}\right)_{x=0} = \left(\frac{1}{q!} \left(\frac{d}{dx}\right)^{p-1} \frac{n}{(1-x)^{n+1}}\right)_{x=0} = \left(\frac{n(n+1)\dots(n+q-1)}{q!}\right) = C_{n+q-1}^q.$$

□

**Proposition 12.** *Le nombre de solutions de l'équation  $\sum_{k=1}^n x_k = p$  où, pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $x_k \geq s_k$ , est  $C_{p-s+n-1}^{n-1}$  avec  $s = \sum_{k=1}^n s_k$ .*

*Démonstration.* Forcément  $p \geq s$  pour avoir une solution non nulle. Soit  $b''_{n,p}$  le nombre de solutions. En faisant le produit de Cauchy de séries absolument convergentes (par récurrence) pour  $x < 1$  (série entière à l'intérieur du rayon de convergence) :

$$\frac{x^s}{(1-x)^n} = \sum_{x_1=s_1}^{\infty} \dots \sum_{x_n=s_n}^{\infty} x^{\sum_{k=1}^n x_k} = \sum_{p=s}^{\infty} \sum_{\sum_{k=1}^n x_k=p, x_i \geq 0} x^{\sum_{k=1}^n x_k} = \sum_{p=s}^{\infty} b''_{n,p} x^p.$$

On déduit le résultat des questions précédentes  $b''_{n,p} = b'_{n,p-s}$ .

□

## 5.8 Application de la méthode des séries génératrices.

Il s'agit de considérer la série entière  $\sum_{n=1}^{\infty} b_n x^n$  pour calculer  $b_n$ .

*Exemple 16.* (Cas polynomiale) Calculons  $\sum_{k_1+k_2+k_3=n, k_i \geq 0} k_1 \frac{n!}{k_1! \dots k_n!}$ . Comme par la formule du multinôme

$$\sum_{r_1+\dots+r_n=p, r_i \geq 0} \frac{p!}{r_1! \dots r_n!} x^{r_1} = (x+n-1)^p.$$

En dérivant par rapport à  $x$  on obtient :

$$\sum_{r_1+\dots+r_n=p, r_i \geq 0} \frac{p!}{r_1! \dots r_n!} r_1 x^{r_1-1} = p(x+n-1)^{p-1}.$$

Donc en évaluant en 1 on obtient :

$$\sum_{k_1+k_2+k_3=n, k_i \geq 0}^n k_1 \frac{n!}{k_1! k_2! k_3!} = n3^{n-1}.$$

*Exemple 17.* Soit  $C_n$  le nombre de parenthésages à  $n$  parenthèses ouvrantes. La première parenthèse doit se refermer après avoir ouvert entre 0 (si elle se referme tout de suite) et  $n-1$  parenthèses, d'où la relation de récurrence :

$$C_n = \sum_{k=0}^{n-1} C_k C_{n-k}, C_0 = 1.$$

Si on pose  $T(x) = \sum_{k=0}^{\infty} C_k x^k$  cette série a rayon de convergence au moins  $1/4$  car  $C_n \leq 2^{2n}$  (le nombre de parenthésage bien ou mal formé de  $2n$  est  $C_{2n} \leq 2^{2n}$ .)

De plus la relation devient

$$T(x) = \sum_{n=0}^{\infty} \sum_{k=0}^{n-1} C_k C_{n-k} x^k x^{n-k} = 1 + x \sum_{m=0}^{\infty} \sum_{k=0}^m C_k C_{n-k} x^k x^{m-k} = 1 + xT(x)^2$$

(en prenant  $m = n - 1$  et en regardant le produit de Cauchy de  $T$  et lui même).  
 En résolvant l'équation du second degré, on obtient :

$$\begin{aligned}
 T(x) &= \frac{1 - \sqrt{1 - 4x}}{2x} \\
 &= -\frac{1}{2x} \sum_{n=1}^{\infty} \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\dots(-\frac{2n+3}{2})}{n!} (-4x)^n \\
 &= \frac{1}{2x} \sum_{n=0}^{\infty} \frac{(1.3\dots(2n-3))}{n!} (2x)^n \\
 &= \sum_{n=1}^{\infty} \frac{(2n-2)!}{n!(n-1)!} x^{n-1}.
 \end{aligned}$$

Donc  $C_n = \frac{1}{n+1} C_{2n}^n$ .



# Chapitre 2

## Espaces probabilisés

### 1 Expérience aléatoire, Introduction

On appelle expérience aléatoire toute expérience, provoquée ou non, reproductible, mais dont la répétition, dans des conditions supposées être les mêmes pour un observateur ou un expérimentateur, conduit à un ensemble de résultats dont aucun ne peut être prévu à l'avance. Cette définition se rattache à la notion de phénomènes aléatoires qui se manifestent toujours avec un certain degré d'indétermination, ce qui empêche de prévoir exactement le résultat de leurs observations, par opposition aux phénomènes déterministes régis par des lois qui en déterminent le déroulement et le résultat. On peut cependant remarquer que tout phénomène comporte une part d'aléatoire qui traduit l'impossibilité de connaître ou de maîtriser tous les paramètres qui le définissent. En fait, notre environnement comporte de nombreux phénomènes aléatoires qui interviennent dans toutes les branches de l'activité humaine. D'où la nécessité et l'importance de disposer et de développer des moyens et des techniques qui aident à comprendre ces phénomènes, à déceler et dégager des propriétés de régularité et des tendances quantifiées afin de maîtriser, prévoir et agir. A travers cette quantification, on obtient un calcul déterministe décrivant les phénomènes aléatoires.

Deux approches s'imposent et se complètent : L'observation et la modélisation. Mais un traitement qui ne s'appuie pas sur un modèle solide atteint vite ses limites et peut conduire à des résultats incohérents appelés paradoxes. La modélisation fournit un cadre d'étude fondé sur une théorie cohérente qui possède et permet de développer des outils efficaces et rationnels. L'objet fondamental de cette modélisation est l'espace probabilisé qui est un triplet  $(\Omega, \mathcal{T}, P)$  dont les termes sont présentés dans ce chapitre.

Par ailleurs, l'utilisation d'aléa permet la réalisation d'algorithmes parfois plus efficaces ou l'obtention plus aisée d'algorithmes efficaces. On proposera donc la simulation de la plupart des objets introduits en utilisant un générateur de nombres aléatoires comme point de départ.

### 2 L'espace des réalisations

L'espace des réalisations  $\Omega$  d'une expérience aléatoire est l'ensemble des résultats possibles de cette expérience. Cet espace est aussi appelé univers, référentiel, ensemble fondamental ou espace des épreuves.

Les éléments  $\omega \in \Omega$  sont appelés réalisations de l'expérience en question.

L'ensemble  $\Omega$  peut être fini, infini dénombrable ou infini non dénombrable.

Selon le point de vue et l'objectif de l'étude, on peut considérer des ensembles différents pour une

même expérience. Par exemple, on peut s'intéresser aux six nombres amenés par le lancer d'un dé ou uniquement à leur parité. Dans le premier cas, on prendra  $\Omega = \{1, 2, 3, 4, 5, 6\}$  et dans le second  $\Omega = \{0, 1\}$ .

Selon les cas, l'ensemble  $\Omega$  peut être plus ou moins simple ou plus ou moins compliqué. Parfois, il n'est pas aisé de définir explicitement  $\Omega$  mais il est fondamental de pouvoir le faire, au moins dans les cas simples. Lorsqu'une idée correcte de la nature et de l'importance de  $\Omega$  est acquise, on peut utiliser efficacement le modèle avec une définition implicite de cet espace.

### 3 Événements et tribu d'événements

Dans l'étude d'une expérience aléatoire on s'intéresse à des ensembles de réalisations appelés événements. Un événement lié à une expérience aléatoire et à son espace  $\Omega$  s'identifie donc à une partie de  $\Omega$ . Les événements élémentaires s'identifient aux singletons  $\{\omega\}$ .

**Définition 10.** Soit  $A \subset \Omega$  un événement et  $\omega \in \Omega$  une réalisation de l'expérience. On dit que l'événement  $A$  s'est réalisé lorsque  $\omega \in A$  et ne s'est pas réalisé lorsque  $\omega \notin A$ .

*Remarque 2.* Tout événement étant une partie de  $\Omega$ , une correspondance s'établit alors entre le langage naturel des événements et celui des sous-ensembles. Nous en donnons ici quelques exemples.

- L'espace  $\Omega$  représente l'événement certain et l'événement impossible correspond à l'ensemble vide  $\emptyset$ .
- L'événement contraire à l'événement  $A$  est représenté par son complémentaire  $A^c$ .
- Deux événements  $A$  et  $B$  sont dits *incompatibles* lorsqu'ils sont disjoints, i.e. lorsque  $A \cap B = \emptyset$ .

**Définition 11.** On appelle *système complet d'événements* (SCE) toute suite finie ou infinie  $(A_n)_{n \geq 1}$  d'événements deux à deux incompatibles telle que  $\cup_{n \geq 1} A_n = \Omega$ . (On parle aussi parfois de partition).

Nous verrons plus loin que cette notion est très utile car elle ramène à des situations où on dispose de plus d'informations pour traiter les événements étudiés.

**Définition 12.** La famille  $\mathcal{T}$  des événements constitue une partie de  $P(\Omega)$  ayant la structure d'une tribu qui, par définition, possède les propriétés suivantes :

1.  $\emptyset \in \mathcal{T}, \Omega \in \mathcal{T}$ .
2. Si  $A \in \mathcal{T}$  alors  $A^c \in \mathcal{T}$ .
3. Pour toute suite finie ou infinie (dénombrable)  $(A_n)_{n \geq 1}$  de parties de  $\mathcal{T}$ , alors  $\cup_{n \geq 1} A_n \in \mathcal{T}$ .

Par construction (ou hypothèse), la tribu  $\mathcal{T}$  contient tous les événements auxquels on s'intéresse lors de l'étude d'une expérience aléatoire. La structure de tribu est importante pour l'efficacité du modèle et pour donner un sens aux différentes opérations que l'on est amené à faire sur les événements.

En pratique, on peut supposer que toutes ces propriétés sont satisfaites sans référence explicite à la tribu, d'autant plus que, *lorsque  $\Omega$  est fini ou dénombrable, on peut prendre pour tribu  $\mathcal{T}$  la famille  $P(\Omega)$  de toutes les parties de  $\Omega$ .*

Cependant, il faut toujours avoir conscience de la tribu retenue pour l'étude. En effet, lorsque  $\Omega$  est infini non dénombrable (par exemple  $\Omega = \mathbb{R}$ ) la tribu  $P(\Omega)$  est trop grande et soulève des difficultés d'ordre technique. Il y a alors un choix judicieux à faire et la tribu considérée sera assez riche pour contenir tous les événements qui interviennent lors de l'étude entreprise.

## 4 Probabilité sur une tribu d'événements

Ayant défini les événements et leur famille, on est amené à quantifier la vraisemblance de chacun d'eux et à évaluer ses chances de réalisation au cours d'une expérience aléatoire. Ceci se fait en associant à chaque événement un nombre compris entre 0 et 1 appelé sa probabilité, en respectant quelques règles inspirées par l'intuition. Ceci nous conduit à la définition suivante

**Définition 13** (Définition d'une probabilité). On appelle *probabilité* sur une tribu  $\mathcal{T}$  toute application  $P : \mathcal{T} \rightarrow [0, 1]$  ayant les propriétés suivantes :

1.  $P(\emptyset) = 0, P(\Omega) = 1$ .
2. Pour toute suite finie ou infinie (dénombrable)  $(A_n)_{n \geq 1}$  d'événements deux à deux incompatibles,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

La propriété 2., appliquée à une suite finie, est l'additivité d'une probabilité et, appliquée à une suite infinie, est la  $\sigma$ -additivité d'une probabilité. Une probabilité est aussi appelée *loi* ou *distribution*.

## 5 Premières propriétés d'une probabilité

Malgré la simplicité apparente de sa définition, une probabilité possède des propriétés très intéressantes qu'il est essentiel de connaître car elles nous permettent d'accéder à la probabilité d'événements variés et très complexes.

Notons que tous les événements considérés dans les relations suivantes appartiennent à la tribu  $\mathcal{T}$  sur laquelle la probabilité  $P$  est définie.

**Proposition 13.** 1.  $P(A^c) = 1 - P(A)$ .

2. Si  $A \subset B$  alors  $P(B) = P(A) + P(B - A)$ . En particulier, si  $A \subset B$  alors  $P(A) \leq P(B)$ .

3. Si  $A \cap B = \emptyset$ , alors  $P(A \cup B) = P(A) + P(B)$ . Dans le cas général,  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Plus généralement encore,

4.

$$P\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} P(A_{i_1} \cap \dots \cap A_{i_k}).$$

( Formule du crible ou de Poincaré ).

5. Si  $(A_n)_{n \geq 1}$  est une suite croissante,

$$P\left(\bigcup_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) = \sup_{n \geq 1} P(A_n).$$

6. Si  $(A_n)_{n \geq 1}$  est une suite décroissante,

$$P\left(\bigcap_{n \geq 1} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) = \inf_{n \geq 1} P(A_n).$$

**Définition 14.** Un événement  $A$  est dit *presque certain* (ou *presque sûr*) si  $P(A) = 1$  et *presque impossible* si  $P(A) = 0$ .

**Définition 15.** On appelle système presque complet d'événements (SPCE) toute suite finie ou infinie (dénombrable)  $(A_n)_{n \geq 1}$  d'événements deux à deux incompatibles telle que

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) = 1.$$

**Proposition 14.** Soit  $(A_n)_{n \geq 1}$  un SPCE. Alors, pour tout événement  $B$ , on a

$$P(B) = \sum_{n \geq 1} P(A_n \cap B).$$

Il est important de noter qu'il n'existe pas de probabilité unique et prédéfinie. Le choix et la construction d'une probabilité pour modéliser une expérience aléatoire doit tenir compte de ses propriétés spécifiques et refléter autant que possible la réalité du phénomène. Un mauvais choix conduit à un modèle qui a sa cohérence interne mais complètement inadapté au phénomène qu'il est censé décrire. Il fournit des résultats en contradiction avec les observations que l'on peut faire. Il est donc inutile.

Le paragraphe suivant donne la méthode générale de construction d'une probabilité sur un ensemble fini ou dénombrable.

## 6 Probabilité sur un ensemble fini ou dénombrable

Soit  $\Omega$  un ensemble fini ou dénombrable. La donnée d'une probabilité  $P$  sur  $\mathcal{P}(\Omega)$  équivaut à la donnée d'une fonction  $p : \Omega \rightarrow [0, 1]$  telle que  $\sum_{\omega \in \Omega} p(\omega) = 1$ . La correspondance se fait de la manière suivante : Lorsque  $P$  est donnée, on définit  $p$  par les relations  $p(\omega) = P(\{\omega\})$  et lorsque  $p$  est donnée, on définit  $P$  par la relation

$$P(A) = \sum_{\omega \in A} p(\omega).$$

Cette construction est explicitée davantage dans ce qui suit.

### 6.1 Cas des ensembles finis

Une probabilité sur (la famille de toutes les parties d') un ensemble fini  $\Omega = \{\omega_1, \dots, \omega_n\}$  est complètement déterminée par la donnée de  $n$  nombres positifs  $(p_k)_{1 \leq k \leq n}$  tels que  $\sum_{k=1}^n p_k = 1$ . Le nombre  $p_k$  représente alors la probabilité de la réalisation  $\omega_k$  et, pour tout  $A \subset \Omega$ , on a  $P(A) = \sum_{k \in A} p_k$ .

### 6.2 La probabilité uniforme $\mathcal{U}(n)$

Lorsque toutes les réalisations sont équiprobables, tous les nombres  $p_k$  sont égaux et de somme 1. Il en résulte que, pour tout  $k \in \llbracket 1, n \rrbracket$ ,  $p_k = \frac{1}{n}$ . La probabilité correspondante  $P$  est la probabilité uniforme sur un ensemble de cardinal  $n$ , notée  $\mathcal{U}(n)$ . Cette probabilité modélise toute expérience aléatoire où les réalisations sont tirées ou choisies ou apparaissent "au hasard" parmi  $n$  sans qu'aucune ne soit privilégiée par rapport aux autres. Dans ce cas, et uniquement dans ce cas, la probabilité d'un événement est donnée par la formule

$$P(A) = \frac{\text{Card}(A)}{\text{Card}(\Omega)}.$$

qui se déduit de  $P(A) = \sum_{k \in A} \frac{1}{n}$ . Cette relation s'exprime en énonçant que la probabilité d'un événement est le rapport du nombre de cas favorables au nombre de cas possibles.

### 6.3 Cas d'un ensemble dénombrable

Une probabilité sur (la famille de toutes les parties d') un ensemble infini dénombrable  $\Omega = \{\omega_n\}_{n \geq 1}$  est complètement déterminée par la donnée d'une suite  $(p_n)_{n \geq 1}$  de nombres positifs telle que  $\sum_{n=1}^{\infty} p_n = 1$ . Le nombre  $p_n$  représente alors la probabilité de la réalisation  $\omega_n$  et, pour tout  $A \in \Omega$ , on a

$$P(A) = \sum_{n \in A} p_n.$$

Il n'existe pas de probabilité uniforme sur un ensemble infini dénombrable.

### 6.4 Quelques lois usuelles sur un ensemble fini ou dénombrable

La loi uniforme  $\mathcal{U}(n)$ , de paramètre  $n \in \mathbb{N}^*$  a été présentée ci-dessus.

**Définition 16.** La loi binomiale  $B(n, p)$  de paramètres  $(n, p) \in \mathbb{N}^* \times [0, 1]$ . C'est la probabilité sur  $\llbracket 0, n \rrbracket$  définie par les nombres  $p_k = C_n^k p^k (1-p)^{n-k}$ . C'est la loi du nombre de succès en  $n$  essais identiques et indépendants, lorsque la probabilité du succès à chaque essai est  $p$ . La loi  $B(1, p)$  est aussi notée  $B(p)$  et appelée loi de Bernoulli de paramètre  $p$ .

**Définition 17.** La loi hypergéométrique  $H(N, n, p)$ ,  $n \in \llbracket 1, N \rrbracket$ ,  $p = m/N$ . C'est la probabilité sur  $\llbracket 0, n \rrbracket$  définie par les nombres  $p_k = \frac{C_m^k C_{N-m}^{n-k}}{C_N^n}$ . Notons que certains  $p_k$  peuvent être nuls. C'est la loi du nombre de succès (c'est-à-dire, le nombre de boules d'une catégorie) obtenus en  $n$  tirages sans remise, lorsque le nombre initial des boules de cette catégorie est  $m$  et le nombre total de boules dans l'urne est initialement  $N$ .

**Définition 18.** La loi de Poisson  $P(\lambda)$ , de paramètre  $\lambda > 0$ . C'est la loi sur  $\mathbb{N}$  définie par les nombres  $p_n = e^{-\lambda} \frac{\lambda^n}{n!}$ . C'est la loi du nombre de succès associé à un phénomène discret qui se déroule dans le temps ou dans l'espace. C'est la limite de  $B(n, \frac{\lambda}{n})$  quand  $n \rightarrow \infty$

**Définition 19.** La loi géométrique  $G(p)$ , de paramètre  $p \in [0, 1]$ . C'est la loi sur  $\mathbb{N}^*$  définie par les nombres  $p_n = pq^{n-1}$  où  $q = 1-p$ . C'est la loi d'attente du premier succès dans une succession infinie dénombrable d'essais identiques et indépendants, lorsque la probabilité du succès à chaque essai est  $p$ .

**Définition 20.** La loi géométrique  $G_0(p)$ , de paramètre  $p \in [0, 1]$ . C'est la loi sur  $\mathbb{N}$  définie par les nombres  $p_n = pq^n$  où  $q = 1-p$ . C'est la loi du nombre d'échecs précédant le premier succès dans une succession infinie dénombrable d'essais identiques et indépendants, lorsque la probabilité du succès à chaque essai est  $p$ .

## 7 Probabilité uniforme sur la droite et dans le plan

**Définition 21.** Soit  $\Omega$  un segment  $[a, b]$  de  $\mathbb{R}$  de longueur non nulle. La probabilité uniforme sur  $\Omega$ , notée  $\mathcal{U}(\Omega)$ , modélise le choix "au hasard" d'un point dans  $\Omega$ . Cette probabilité  $P$  associe à tout événement  $A \subset \Omega$  le nombre

$$P(A) = \frac{\text{longueur}(A)}{\text{longueur}(\Omega)}.$$

**Définition 22.** Soit  $\Omega$  un compact du plan d'aire non nulle. La probabilité uniforme sur  $\Omega$ , notée  $\mathcal{U}(\Omega)$ , modélise le choix "au hasard" d'un point dans  $\Omega$ . Cette probabilité  $P$  associe à tout événement  $A \subset \Omega$  le nombre

$$P(A) = \frac{\text{aire}(A)}{\text{aire}(\Omega)}.$$

## 8 Simulation

On appellera plus loin variable aléatoire une fonction  $X : \Omega \rightarrow \mathbb{R}$ . Par exemple, il sera pratique de représenter un événement  $A$  par sa fonction indicatrice  $1_A$  vu comme variable aléatoire.

Pour décrire la simulation d'évènements ou de variables aléatoires, nous utiliserons un générateur de nombre aléatoire, disons noté *ALEA*, qui donnera un nombre de  $[0, 1]$  selon la loi uniforme sur le segment  $[0, 1]$ . Autrement  $P(\textit{ALEA} \in [a, b]) = b - a$  pour  $0 \leq a \leq b \leq 1$ . Nous dirons plus loin que *ALEA* représente une variable aléatoire de loi uniforme sur  $[0, 1]$ .

On peut construire tous les objets de ce cours à partir de *ALEA*. Par exemple, si on veut un événement de probabilité  $P(A)$ .

On peut utiliser l'algorithme générique :

Soit  $U := \textit{ALEA}$ , Si  $(U < P(A))$ ,  $X := 1$  Sinon  $X := 0$  FinSi.

$X$  est alors une variable aléatoire représentant la fonction indicatrice  $1_A$  d'un événement de probabilité  $P(A)$ .

*Exemple 18.* 'Simulation d'un dé à  $n$  faces]. La procédure *dé(n)*, suivante simule un dé à  $n$  face non-biaisé. On suppose que *Ent* retourne la partie entière :

Soit  $U = \textit{ALEA}$ , ( $X := \textit{Ent}(nU) + 1$ ).

En effet, on a :

$$P(X = k) = P(\textit{Ent}(nU) = k - 1) = P((k - 1)/n \leq U \leq k/n) = \frac{\text{longeur}([\frac{k-1}{n}, \frac{k}{n}])}{\text{longeur}([0, 1])} = \frac{1}{n}.$$

# Chapitre 3

## Probabilité conditionnelle. Indépendance

### 1 Probabilité conditionnelle

Dans la modélisation d'une expérience aléatoire, l'espace probabilisé tient compte des informations qui apparaissent ou qui modifient le déroulement de cette expérience. Ainsi, l'information concernant la réalisation d'un événement  $A$  sera présente dans la définition d'une nouvelle probabilité appropriée. Ceci conduit à la notion de probabilité conditionnelle qui permet par ailleurs de développer des outils très utiles pour le traitement des expériences aléatoires.

**Définition 23.** Soit  $(\Omega, \mathcal{T}, P)$  un espace probabilisé et  $A$  un événement tel que  $P(A) > 0$ . La probabilité conditionnelle sachant  $A$ , notée  $P(\cdot|A)$ , est la probabilité sur  $\mathcal{T}$  définie par la relation

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

On parle de "probabilité de  $B$  sachant  $A$ ".

On vérifie qu'il s'agit bien d'une probabilité qui restreint l'univers à  $A$  pour tenir compte de sa réalisation. En effet  $P(\emptyset|A) = 0/P(A) = 0$ ,  $P(\Omega|A) = P(A)/P(A) = 1$  et si  $B_n$  sont disjoints, de même  $B_n \cap A$  sont disjoints donc

$$P\left(\bigcup_{n=1}^{\infty} B_n|A\right) = \frac{1}{P(A)} P\left(\bigcup_{n=1}^{\infty} (B_n \cap A)\right) = \frac{1}{P(A)} \sum_{n=1}^{\infty} P(B_n \cap A) = \sum_{n=1}^{\infty} P(B_n|A).$$

Les propriétés qui suivent sont en fait des présentations différentes de la définition, mais elles sont très utiles.

**Proposition 15** (Formule des probabilités composées). *Soit  $A$  et  $B$  deux événements de probabilité non nulle. Alors, on a*

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

*Plus généralement, si  $(A_k)_{1 \leq k \leq n}$  est une suite finie d'événements telle que  $P(\bigcap_{1 \leq k \leq n-1} A_k) > 0$ , alors*

$$P\left(\bigcap_{k=1}^n A_k\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n | \bigcap_{1 \leq k \leq n-1} A_k).$$

**Proposition 16** (Théorème des probabilités totales). Soit  $(A_n)_{n \geq 1}$  un système presque complet d'événements, tous de probabilité non nulle. Alors, pour tout événement  $B$ , on a

$$P(B) = \sum_{n=1}^{\infty} P(B|A_n)P(A_n).$$

*Exemple 19.* On note  $(N_1, N_2)$ , le résultat du lancer d'une paire de dés équilibrés à  $n$  faces.  $\Omega = \{1, \dots, n\}^2$  avec la probabilité uniforme. On cherche la probabilité de l'évènement  $N_1 > N_2$ . Par la formule :

$$P(N_1 > N_2) = \sum_{k=1}^n P(N_1 > N_2 | N_1 = k)P(N_1 = k).$$

Par l'uniformité  $P(N_2 = k) = \frac{1}{n}$  et

$$P(N_1 > N_2 | N_1 = k) = P(k > N_2 \cap N_1 = k) / P(N_1 = k) = P(k > N_2) = \frac{k-1}{n}$$

(ce qui sera justifié en vérifiant que les lancers de  $N_1$  et  $N_2$  sont indépendants.) Donc :

$$P(N_1 > N_2) = \sum_{k=1}^n \frac{k-1}{n^2} = \frac{n(n-1)}{2n^2} = \frac{n-1}{2n}.$$

Pour  $n = 6$ , on obtient  $5/12$ .

**Proposition 17** (Formule de Bayes). Soit  $(A_n)_{n \geq 1}$  un système presque complet d'événements, tous de probabilité non nulle et  $B$  un événement de probabilité non nulle également. Alors, pour tout  $k \geq 1$ ,

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{n \geq 1} P(B|A_n)P(A_n)}.$$

D'après l'interprétation qu'on peut lui donner, cette formule est aussi connue sous le nom de formule des probabilités des causes.

*Exemple 20.* Un laboratoire a mis au point un test pour déceler des souris malades. Des essais prouvent que :

- 96 fois sur 100, le test donne un résultat positif quand la souris est effectivement malade.
- 94 fois sur 100, le test donne un résultat négatif quand la souris n'est pas malade.

Dans une population de souris comprenant 3 % de malades, on pratique le test sur une souris choisie au hasard et on constate que le test donne un résultat positif. Quelle est la probabilité que la souris soit malade ?

On pose les évènements  $T$  : le test est positif et  $M$  : la souris est malade. L'énoncé nous apprend que  $P(T|M) = 0,96$  et  $P(T^c|M^c) = 0,94$ , donc  $P(T|M^c) = 0,06$ . Par ailleurs on a  $P(M) = 3/100$ , la formule de Bayes nous permet de calculer la probabilité qu'une souris soit malade sachant que le test a été positif :

$$P(M|T) = \frac{P(T|M)P(M)}{P(T|M)P(M) + P(T|M^c)P(M^c)} = \frac{0,96 \times 0,03}{0,96 \times 0,03 + 0,06 \times 0,97} = \frac{288}{870} \simeq 0,331.$$

Bien que le test semblait efficace, il garantit très peu le caractère malade si il est positif. Cela vient de ce que  $P(M) \ll P(M^c) \sim 1$ . Si on n'a pas pour compenser  $P(T|M^c) \ll P(M)$  (même avec  $P(T|M) \sim 1$ ) mais comme dans l'énoncé  $P(T|M^c) \simeq 2P(M)$  on obtient donc  $P(M|T) \simeq \frac{P(M)}{P(M)+P(T|M^c)} \simeq \frac{1}{3}$ .



## 2 Indépendance

Intuitivement, on peut dire que deux événements  $A$  et  $B$  sont indépendants si la probabilité de l'un d'eux ne change pas quand on sait que l'autre s'est réalisé. En traduisant cette idée avec la définition de la probabilité conditionnelle, on obtient la définition suivante :

**Définition 24.** Soit  $(\Omega, \mathcal{T}, P)$  un espace probabilisé. Deux événements  $A$  et  $B$  sont indépendants (pour la probabilité  $P$ ) lorsque

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

Autrement dit si  $P(B) > 0$  :  $P(A|B) = P(A)$ .

*Exemple 21.* On note  $(N_1, N_2)$ , le résultat du lancer d'une paire de dés équilibrés à  $n$  faces.  $\Omega = \{1, \dots, n\}^2$  avec la probabilité uniforme. Montrons que les événements  $A = \{N_1 = i\}$  et  $B = \{N_2 = j\}$ ,  $i, j \in \{1, \dots, n\}$  sont indépendants. En effet on obtient

$$P(A \cap B) = P((N_1, N_2) = (i, j)) = \frac{1}{n^2}.$$

$$P(A) = \sum_{j=1}^n P((N_1, N_2) = (i, j)) = \frac{n}{n^2} = \frac{1}{n}$$

et de même  $P(B) = \frac{1}{n}$  d'où :

$$P(A \cap B) = \frac{1}{n^2} = P(A)P(B).$$

Cette définition se généralise de la manière suivante :

**Définition 25.** Soit  $n \in \mathbb{N}^*$ . Les événements de la suite  $(A_k)_{1 \leq k \leq n}$  sont *deux à deux indépendants* si pour tout  $(i, j) \in \llbracket 1, n \rrbracket^2$  tel que  $i \neq j$ ,

$$P(A_i \cap A_j) = P(A_i)P(A_j).$$

La notion la plus utile est la notion suivante :

**Définition 26.** Soit  $n \in \mathbb{N}^*$ . Les événements de la suite  $(A_k)_{1 \leq k \leq n}$  sont *indépendants* (ou mutuellement indépendants) si pour toute partie  $J \subset \llbracket 1, n \rrbracket$ ,

$$P(\cap_{k \in J} A_k) = \prod_{k \in J} P(A_k).$$

**Définition 27.** Les événements de la suite infinie  $(A_k)_{k \geq 1}$  sont *indépendants* si, pour tout  $n \in \mathbb{N}^*$ , les événements des suites  $(A_k)_{1 \leq k \leq n}$  le sont.

*Remarque 3.* 1. La notion d'indépendance n'est pas absolue et dépend essentiellement de la probabilité définie sur la tribu des événements. Deux événements définis par les mêmes termes dans deux modèles différents peuvent être indépendants dans l'un des modèles sans l'être dans l'autre.

2. Un événement presque impossible ou presque certain est indépendant de tout événement et un événement est indépendant de lui-même si et seulement s'il est presque certain ou presque impossible. Cette propriété montre à nouveau que la notion d'indépendance dépend de la probabilité considérée.

3. Il ne faut pas confondre les notions d'indépendance et d'incompatibilité, la dernière étant déjà une notion spécifique des événements et elle contredit en quelque sorte la première. En fait, deux événements incompatibles sont indépendants si et seulement si l'un d'eux est presque impossible.
4. L'indépendance des événements de l'un des quatre couples  $(A; B); (A^c; B^c); (A; B^c)$  ou  $(A^c; B)$  est équivalente à l'indépendance des événements de chacun de ces couples. Ceci se généralise au cas de  $n$  événements.
5. L'indépendance deux à deux de  $n$  événements se traduit par  $C_n^2 = \frac{n(n-1)}{2}$  égalités qui sont déjà présentes dans la définition de l'indépendance de  $n$  événements. Mais celle-ci requiert  $2^n - n - 1$  égalités. Ainsi, l'indépendance entraîne l'indépendance deux à deux mais la réciproque est fausse.

*Exemple 22.* On lance 25 dés équilibrés indépendamment. Calculons la probabilité de l'évènement  $A$  "avoir au moins un 1". L'évènement contraire est  $A^c$  n'avoir aucun 1 qui est l'intersection des événements indépendants, ne pas avoir un 1 au  $i$ -ème lancer  $N_i \neq 1$ , donc :

$$P(A) = 1 - \prod_{i=1}^{25} P(N_i \neq 1) = 1 - \left(\frac{5}{6}\right)^{25}.$$

## 2.1 Retour sur l'interprétation des lois géométriques et binomiales

*Exemple 23* (Loi Géométrique). On considère une suite  $(X_i)$  de lancer de pile ou face biaisé indépendant avec  $P(X_i = P) = p \in [0, 1]$ . On se demande qu'elle est la distribution de probabilité de l'indice  $N$  du premier pile obtenu. (On dira très vite que  $N$  est une variable aléatoire)

L'évènement  $\{N = n\} = \{X_1 = F\} \cap \dots \cap \{X_{n-1} = F\} \cap \{X_n = P\}$ , et comme ces événements sont indépendants par hypothèse, on obtient :

$$P(N = n) = P(X_1 = F) \dots P(X_{n-1} = F) P(X_n = P) = (1 - p)^{n-1} p.$$

On retrouve les probabilités de la loi géométrique  $G(p)$ .

*Exemple 24* (Application à la simulation des probabilités conditionnelles). Dans la suite, on supposera souvent que

*Si l'on a une procédure donnant un résultat aléatoire, la répétition de la procédure donne des tirages indépendants.*

Montrons le résultat suivant.

*Lemme 18.* Si on suit l'algorithme

*Répéter Epreuve*

*$X :=$  Résultat de l'épreuve*

*Jusqu'à (Condition réalisée) On obtient  $P(X \in B) = P(\text{"Résultat de l'épreuve dans B"} | \text{Condition})$ .*

*Démonstration.* Soit  $B_n$  l'évènement "le résultat est dans  $B$  à l'épreuve  $n$ ". Soit  $C_n$  l'évènement "la condition  $C$  se réalise à l'épreuve  $n$ ".

Alors

$$\{X \in B\} = \bigcup_{n=1}^{\infty} C_1^c \cap \dots \cap C_{n-1}^c \cap C_n \cap B_n,$$

donc comme les événements sont disjoints et par indépendances des lancers :

$$P(X \in B) = \sum_{n=1}^{\infty} P(C_1^c) \dots P(C_{n-1}^c) P(C_n \cap B_n) = \frac{P(C \cap B)}{1 - P(C^c)} = P(B|C).$$

□

Par exemple,

Si on effectue :

Répéter

$U := \text{dé}(8)$

Jusqu'à ( $U \leq 6$ )

$X := U$

On obtient la simulation d'un dé à six faces à partir d'un dé à 8 faces.

En effet pour  $k = 1, \dots, 6$ ,

$$P(X = k) = P(U = k | U \leq 6) = \frac{1/8}{6/8} = \frac{1}{6}.$$

*Exemple 25.* (Loi binomiale) On considère comme à l'exemple 23 une suite  $(X_i)$  de lancer de pile ou face biaisé indépendant avec  $P(X_i = P) = p \in [0, 1]$ .

On note  $N_n$  le nombre de pile obtenu parmi les  $n$  premiers lancers.

$$\{N_n = k\} = \cup_{P \in P_k(\llbracket 1, n \rrbracket)} \cap_{i \in P} \{X_i = P\} \cap \cap_{i \notin P} \{X_i = F\}.$$

Par union disjointe et indépendance, on déduit

$$P(N_n = k) = \sum_{P \in P_k(\llbracket 1, n \rrbracket)} P(X_i = P)^k P(X_i = F)^{n-k} = \text{Card}(P_k(\llbracket 1, n \rrbracket)) p^k (1-p)^{n-k} = C_n^k p^k (1-p)^{n-k}.$$

On retrouve la loi binomiale.

# Chapitre 4

## Variables aléatoires

### 1 Introduction

Le plus souvent, les questions liées à l'étude d'une expérience aléatoire s'expriment en termes d'une fonction numérique ou vectorielle des réalisations. Il est donc naturel de s'intéresser à ce type de fonctions et c'est ce qui nous conduit à la notion de variable et vecteur aléatoire. Nous considérons ici uniquement les variables aléatoires. A l'aide de ces variables, on dispose à la fois d'une interprétation concrète de l'objet d'étude et d'un moyen de mettre à profit des outils analytiques puissants. Leur utilité est telle qu'il est toujours recommandé, au moyen d'un codage approprié s'il en est besoin, d'associer une variable ou un vecteur aléatoire aux questions considérées. L'étude d'une variable ou vecteur aléatoire porte principalement sur sa loi (sa répartition) et ses paramètres caractéristiques.

### 2 Exemples

*Exemple 26.* Considérons le lancer de deux dés dont l'ensemble des réalisations est  $\Omega = \llbracket 1, 6 \rrbracket^2$ . A chaque réalisation  $\omega = (i, j)$ , on peut associer la somme  $S(\omega) = i + j$  des points obtenus, le maximum  $M(\omega) = \max(i, j)$  des points obtenus, etc. On définit ainsi des variables aléatoires.

*Exemple 27.* Considérons une succession de tirages au hasard d'une boule dans une urne contenant 4 boules noires et 6 boules rouges. Le rang d'apparition de la première boule rouge est une variable aléatoire.

*Exemple 28.* Le nombre de portables vendus dans une journée dans un magasin est une variable aléatoire.

*Exemple 29.* Un schéma général largement utilisé est la répétition d'une expérience aléatoire élémentaire à deux issues appelées succès et échec. Au  $k$ -ème essai, on associe la variable aléatoire  $X_k$  qui vaut 1 lorsque l'issue correspondante est succès et qui vaut 0 dans le cas contraire. Alors, pour tout entier  $n \in \mathbb{N}^*$ ,  $S_n = \sum_{k=1}^n X_k$  est la variable aléatoire égale au nombre de succès en  $n$  essais.

*Exemple 30.* En choisissant au hasard un point dans le disque de centre  $O$  et de rayon  $R$ , on peut s'intéresser à la variable aléatoire  $D$  égale à la distance du point choisi au centre  $O$  du disque.

*Exemple 31.* La hauteur des précipitations en un mois dans une région donnée est une variable aléatoire.

*Exemple 32.* Le temps de service d'un client à un guichet est une variable aléatoire.

La notion de variable aléatoire et l'étude de sa répartition se font à partir de la construction d'un espace probabilisé dont le premier terme est  $\mathbb{R}$ . Ceci est précisé dans les définitions suivantes :

### 3 Définitions. Loi d'une variable aléatoire

Soit  $(\Omega, \mathcal{T}, P)$  un espace probabilisé. 1. Pour définir la tribu des événements de  $\mathbb{R}$ , on peut penser à la famille  $I$  des intervalles de  $\mathbb{R}$  mais on s'aperçoit vite que celle-ci n'est pas une tribu. Le bon choix se fait en considérant la plus petite tribu contenant les intervalles.

**Définition 28.** Par définition, la plus petite tribu contenant les intervalles, notée  $\mathcal{B}(\mathbb{R})$  ou  $\mathcal{B}$ , est la tribu borélienne de  $\mathbb{R}$ . Un ensemble  $B \subset \mathbb{R}$  tel que  $B \in \mathcal{B}$  est appelé borélien de  $\mathbb{R}$ .

Dans la pratique, toute partie  $B$  de  $\mathbb{R}$  qui apparaît "de manière naturelle" dans l'étude d'une expérience aléatoire sera une partie borélienne et nous la considérons toujours comme telle sans se poser des questions sur sa nature. Le plus souvent,  $B$  sera un intervalle ou une réunion d'intervalles. Par défaut,  $\mathbb{R}$  sera toujours muni de sa tribu borélienne  $\mathcal{B}$ .

**Définition 29.** Soit  $S \subset \mathbb{R}$  un borélien. La tribu borélienne de  $S$ , notée  $\mathcal{B}(S)$  est constituée de tous les boréliens  $B$  de  $\mathbb{R}$  tels que  $B \subset S$ .

Là aussi, toute partie de  $S$  qui apparaît dans l'étude sera borélienne.

**Définition 30.** Une variable aléatoire (v.a.) définie sur  $\Omega$  est une application  $X : \Omega \rightarrow \mathbb{R}$  (ou  $X : \Omega \rightarrow \overline{\mathbb{R}}$ ) telle que, pour tout borélien  $B$  de  $\mathbb{R}$ , l'ensemble

$$\{X \in B\} \equiv X^{-1}(B) \equiv \{\omega \in \Omega : X(\omega) \in B\}$$

est un événement, i.e.  $\{X \in B\} \in \mathcal{T}$ . Pour cela, on démontre qu'il faut et il suffit que cette propriété soit satisfaite lorsque  $B$  est un intervalle de  $\mathbb{R}$ .

Cette définition s'explique naturellement de la manière suivante. Du point de vue probabiliste, la caractéristique essentielle d'une v.a. est sa répartition qui permet de déterminer la probabilité avec laquelle cette variable appartient à telle ou telle région et ceci suppose que l'on soit en mesure de calculer la probabilité des ensembles signalés ci-dessus tels que  $\{X \in B\}$  et pour cela il faut que ces ensembles appartiennent à la tribu des événements sur laquelle la probabilité est définie.

**Définition 31.** Une application  $X : \Omega \rightarrow \mathbb{C}$  est une v.a. si sa partie réelle et sa partie imaginaire le sont.

Là encore, toute application  $X : \Omega \rightarrow \mathbb{R}$  ou  $X : \Omega \rightarrow \mathbb{C}$  qui apparaît dans l'étude d'une expérience aléatoire sera considérée d'office comme une variable aléatoire.

**Définition 32.** Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a. On appelle loi (distribution, ou répartition) de  $X$  la probabilité sur  $\mathcal{B}$ , notée  $P_X$  ou  $\mu_X$ , définie par

$$P_X(B) = P(\{X \in B\}), B \in \mathcal{B}$$

On vérifie que  $P_X$  ainsi définie est effectivement une probabilité.

Soit  $S = X(\Omega)$  l'ensemble des valeurs de  $X$ . Lorsque  $S$  est distinct de  $\mathbb{R}$ , on a, par additivité, pour tout  $B \in \mathcal{B}$ ,  $P_X(B) = P_X(B \cap S) + P_X(B \cap S^c) = P_X(B \cap S)$  car  $P_X(B \cap S^c) = P(X \in B \cap S^c) = P(\emptyset) = 0$ . Cette constatation montre que la loi de  $X$  est complètement définie par ses valeurs sur les boréliens de  $S$ .

Ainsi, on peut considérer la loi de  $X$  comme une probabilité sur  $S$ . Mais lorsque l'on fait l'étude de plusieurs v.a. à la fois, il est plus commode de considérer que toutes les lois sont définies sur  $\mathbb{R}$ .

Avec ce qui précède, on voit qu'une variable aléatoire  $X$  permet de construire un nouvel espace probabilisé  $(\mathbb{R}, \mathcal{B}, P_X)$  qui va permettre d'établir toutes les caractéristiques de  $X$ . Les probabilités sont surtout l'étude des questions sur les variables aléatoires qui se ramènent à l'étude de la loi, indépendamment de la construction concrète de  $\Omega$ .

Voici deux notions qui peuvent intervenir quand on est en présence de deux variables aléatoires.

**Définition 33** (Égalité en loi). . Deux v.a.  $X$  et  $Y$  ayant même loi  $P_X = P_Y$  sont dites équidistribuées ou égales en loi ou *identiquement distribuées*, ce que l'on peut noter  $X =_{\mathcal{L}} Y$ . Insistons sur le fait que l'égalité en loi est distincte de l'égalité ponctuelle.

**Définition 34** (Indépendance). . Deux v.a.  $X$  et  $Y$  sont dites indépendantes si l'on a, pour tous boréliens  $A$  et  $B$  de  $\mathbb{R}$ ,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

Pour cela, il faut et il suffit que cette condition soit satisfaite lorsque  $A$  et  $B$  sont des intervalles quelconques de  $\mathbb{R}$ .

*Remarque 4.* 1.  $X$  et  $Y$  sont donc indépendants si  $\{X \in A\}$  et  $\{Y \in B\}$  sont des évènements indépendants pour TOUT borélien (ou intervalle)  $A$  et  $B$ .

2. Lorsque  $X$  et  $Y$  sont deux v.a. indépendantes,  $f$  et  $g$  deux fonctions boréliennes (i.e. des v.a. sur  $(\mathbb{R}, \mathcal{B})$ ), alors les v.a.  $f(X)$  et  $g(Y)$  sont encore indépendantes.

3. Les définitions précédentes se généralisent au cas de plusieurs v.a. Une suite  $(X_n)_{n \geq 1}$  de v.a. sont dites indépendantes si pour toute suite de boréliens  $(A_n)_{n \geq 1}$ , la suite d'évènements  $(\{X_n \in A_n\})_{n \geq 1}$  est indépendante.

Les v.a. usuelles sont de deux types et pour chacun d'eux, on dispose d'outils adaptés pour faire le calcul des lois et de leurs caractéristiques. On présente d'abord le premier type, les variables discrètes, puis le second, les variables continues.

## 4 Variables aléatoires discrètes

**Définition 35.** Une variable aléatoire  $X : \Omega \rightarrow \mathbb{R}$  (ou  $X : \Omega \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{\infty, -\infty\}$  ou  $X : \Omega \rightarrow \mathbb{C}$ ) est dite *discrète (v.a.d.)* si l'ensemble  $S = X(\Omega)$  de ses valeurs est une partie finie ou dénombrable de  $\mathbb{R}$  ( de  $\overline{\mathbb{R}}$  ou de  $\mathbb{C}$  ).

Dans cette partie, on se limitera aux v.a.d. et, sauf mention expresse du contraire, elles seront à valeurs dans  $\mathbb{R}$ .

### 4.1 Loi d'une v.a.d.

La loi d'une v.a.d.  $X$  est donc la probabilité  $P_X$ , définie sur l'ensemble fini ou dénombrable  $S = X(\Omega)$  de ses valeurs. D'après le résultat du chapitre 2, on voit que  $P_X$  est complètement définie par la suite des nombres positifs  $p_s = P(X = s)$ , lorsque  $s$  parcourt  $S$ . Bien entendu,  $\sum_{s \in S} p_s = 1$ . Alors, par additivité ou  $\sigma$ -additivité, on a pour toute partie (borélienne)  $A \subset \mathbb{R}$ ,

$$P(X \in A) = \sum_{s \in S \cap A} p_s.$$

et ceci définit entièrement la loi de  $X$ .

*Ainsi, lorsque la v.a. est discrète, on considère que sa loi est correctement définie quand on précise l'ensemble de ses valeurs et la probabilité avec laquelle chacune de ses valeurs est prise .*

*Exemple 33.* Dans l'exemple 27, on considère des tirages au hasard et sans remise et note  $X$  le rang d'apparition de la première boule rouge. On a  $S = X(\Omega) = \llbracket 1, 5 \rrbracket$ .  $\text{Card}(\Omega) = C_{10}^6$   $\Omega = P_4(\llbracket 1, 10 \rrbracket)$  (les positions des boules rouges dans un tirage complet. La loi de  $X$  est définie par  $p_1 = \frac{C_9^5}{C_{10}^6} = \frac{3}{5}$ ,  $p_2 = \frac{C_8^5}{C_{10}^6} = \frac{4}{15}$ ,  $p_3 = \frac{C_7^5}{C_{10}^6} = \frac{1}{10}$ ,  $p_4 = \frac{C_6^5}{C_{10}^6} = \frac{1}{35}$  et  $p_5 = \frac{C_5^5}{C_{10}^6} = \frac{1}{210}$ .

*Exemple 34.* Dans l'exemple 29, on note  $p$  la probabilité du succès à chaque essai. Alors,  $S_n \sim B(n, p)$ .

Dans les énoncés qui suivent,  $X$  désigne une v.a.d. dont la loi est définie par la suite des nombres  $(p_s)_{s \in S}$ .

## 4.2 Fonction de répartition (f.r.) d'une v.a.d.

**Définition 36.** La fonction de répartition de  $X : \Omega \rightarrow \mathbb{R}$  est la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par

$$F_X(x) = P(X \leq x) = \sum_{s \leq x} p_s.$$

Quand il n'y a pas de confusion à craindre, on note plus simplement  $F$  la f.r. de  $X$ .

**Proposition 19.** Toute fonction de répartition  $F_X$  est croissante, continue à droite et vérifie  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ . De plus, pour tout  $x \in \mathbb{R}$ ,  $P(X = x) = F_X(x) - F_X(x-)$ .

La dernière propriété montre que la f.r. d'une v.a.d. détermine complètement la loi de cette variable.

Divers paramètres et fonctions permettent de caractériser la loi d'une v.a.d. On commence par la moyenne qui donne une idée de l'ordre de grandeur de la variable en question.

## 4.3 Moyenne ou Espérance ou Moment d'ordre 1

**Définition 37.** La v.a.d.  $X$  est dite d'ordre 1 lorsque la série  $\sum_{s \in S} |s|p_s$  est convergente. Dans ce cas, on appelle Moyenne ou Espérance ou Moment d'ordre 1 de  $X$  la quantité

$$m = E[X] = \sum_{s \in S} sp_s.$$

Remarquons que si  $X$  est bornée (et, en particulier, si  $S = X(\Omega)$  est fini), la v.a.  $X$  est d'ordre 1. On note  $L_d^1(\Omega)$  l'ensemble des v.a.d. d'ordre 1. C'est un espace vectoriel.

*Exemple 35.* Si  $X$  est la v.a. de l'exemple 33, on a  $E[X] = 1 \frac{3}{5} + 2 \frac{4}{15} + 3 \frac{1}{10} + 4 \frac{1}{35} + 5 \frac{1}{210} = \frac{11}{7}$ .

*Exemple 36.* Si  $S_n$  est la v.a. de l'exemple 34, on a  $E[S_n] = np$ .

La moyenne est une forme linéaire croissante sur l'espace vectoriel  $L_d^1(\Omega)$  des v.a.d. d'ordre 1. En d'autres termes, on a

**Proposition 20.** Si  $X$  et  $Y$  sont deux v.a.d. d'ordre 1, alors

1. Pour tout couple  $(a, b)$  de scalaires,  $aX + bY$  est d'ordre 1 et  $E(aX + bY) = aE(X) + bE(Y)$ .
2.  $X \leq Y \Rightarrow E[X] \leq E[Y]$ . En particulier,  $X \geq 0 \Rightarrow E[X] \geq 0$ .
3.  $Z$  v.a.d. et  $|Z| \leq |X|$  implique  $Z$  est d'ordre 1.
4.  $|E(X)| \leq E(|X|)$ .
5.  $E[c] = c$ ,  $a \leq X \leq b \Rightarrow a \leq E[X] \leq b$ .

6. Si  $X$  et  $Y$  sont indépendantes,  $XY$  est d'ordre 1 et  $E(XY) = E(X)E(Y)$ .

On peut aussi s'intéresser au calcul de la moyenne d'une fonction quelconque d'une v.a.d. en généralisant la relation précédente. Ceci permettra en même temps d'obtenir différentes caractérisations de la loi d'une v.a.d. à travers le calcul de la moyenne d'une fonction bien choisie. Ceci nous conduit au théorème de transfert.

#### 4.4 Théorème de transfert

**Théorème 21.** Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a.d.,  $S = X(\Omega)$  et  $h : S \rightarrow \mathbb{C}$  une fonction telle que  $\sum_{s \in S} |h(s)|p_s$  soit convergente. Alors, la v.a.d.  $h(X)$  admet une moyenne définie par

$$E[h(X)] = \sum_{s \in S} h(s)p_s.$$

Remarquons que lorsque  $X$  et  $Y$  sont égales en loi, on voit que, pour toute fonction  $h$ ,  $E[h(X)] = E[h(Y)]$ , pourvu que l'un des deux membres existe. En particulier,  $X$  et  $Y$  ont mêmes paramètres (moyenne, variance, f.r., f.c. etc).

**Proposition 22.**  $X$  et  $Y$  v.a.d. sont indépendantes si et seulement si

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)],$$

pour toute fonction  $h, g$  telle que  $E[|h(X)|] < \infty, E[|g(Y)|] < \infty$ . De plus, pour caractériser l'indépendance il suffit de prendre  $h, g$  bornées.

En plus de la f.r., les deux fonctions suivantes constituent des outils intéressants pour caractériser la loi de  $X$ .

#### 4.5 Fonctions caractéristique et génératrice d'une v.a.d.

Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a.d. et  $S = X(\Omega)$ .

**Définition 38.** La fonction caractéristique (f.c.) de  $X$  est définie par

$$\varphi_X(t) = E[e^{itX}] = \sum_{s \in S} e^{its} p_s,$$

pour tout  $t \in \mathbb{R}$ .

**Définition 39.** La fonction génératrice (f.g.) de  $X$  est définie par

$$g_X(t) = E(t^X) = \sum_{s \in S} t^s p_s,$$

pour tout  $t \in \mathbb{R}$  tel que la somme est finie. Le plus souvent  $g_X$  est définie lorsque  $S \subset \mathbb{N}$ , c'est alors une série entière.

En considérant les fonctions puissances, le théorème de transfert nous conduit à considérer les différentes notions suivantes.



## 4.6 Moments d'une v.a.d.

**Définition 40.** Soit  $k$  un entier positif. La v.a.d.  $X$  est dite d'ordre  $k$  si  $\sum_{s \in S} |s|^k p_s$  est convergente.

On voit que si  $X$  est d'ordre  $k$ , alors elle est d'ordre  $j$  pour tout  $j \leq k$ . Remarquons que si  $X$  est bornée (et, en particulier, si  $S = X(\Omega)$  est fini), la v.a.  $X$  est d'ordre  $k$  pour tout  $k$ . On note  $L_d^k(\Omega)$  l'ensemble des v.a.d. d'ordre  $k$ . C'est un espace vectoriel.

**Définition 41.** Soit  $X$  une v.a. d'ordre  $k$ . Alors Le *moment d'ordre  $k$*  de  $X$  est défini par la quantité

$$m_k = E[X^k] = \sum_{s \in S} s^k p_s.$$

## 4.7 Variance et Écart-type d'une v.a.d. Variable centrée réduite

On suppose que  $X$  est une v.a.d. d'ordre 2 de moyenne  $m$ .

**Définition 42.** La *variance* de  $X$  est définie par la quantité  $v = \sigma^2 = V[X] = E[(X - m)^2] = E[X^2] - m^2$ . L'*écart-type* de  $X$  est la quantité  $\sigma = (V[X])^{1/2}$ .

La variable centrée réduite associée à  $X$  est  $U = \frac{X-m}{\sigma}$ . On a  $E[U] = 0$  et  $V[U] = 1$ .

*Remarque 5.* 1. La moyenne et l'écart-type d'une v.a.  $X$  constituent un résumé intéressant de la répartition de  $X$ . La moyenne donne une idée de l'ordre de grandeur de la variable et l'écart-type une idée de la dispersion des valeurs de  $X$  autour de leur moyenne. Ces deux paramètres interviennent dans plusieurs inégalités.

2. Le recours à la variable centrée réduite permet et facilite certains calculs. Il permet aussi de comparer entre elles de façon cohérente plusieurs variables.

La variance est définie sur l'espace vectoriel  $L_d^2(\Omega)$  des v.a.d. d'ordre 2.

**Proposition 23.** si  $X$  et  $Y$  sont deux v.a.d. d'ordre 2, on a les propriétés suivantes :

1.  $V[c] = 0$ ,  $V[X] \geq 0$  et  $V[X] = 0 \Leftrightarrow X = m$ .
2.  $V[X] = E[X^2] - (E[X])^2$ .
3.  $V(aX + b) = a^2 V(X)$ .
4. Si  $X$  et  $Y$  sont indépendantes,  $V(X + Y) = V(X) + V(Y)$ .

## 4.8 Inégalités de Markov et de Bienaymé-Tchébychev

**Proposition 24** (Inégalité de Markov). . Soit  $Z$  une v.a. positive et  $a > 0$ . Alors,

$$P(Z \geq a) \leq \frac{E[Z]}{a}.$$

**Proposition 25** ( Inégalité de Bienaymé-Tchébychev). Soit  $X$  une v.a. de moyenne  $m$  et d'écart-type  $\sigma$ . Alors, pour tout  $\epsilon > 0$ ,

$$P(|X - m| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

## 5 Variables aléatoires continues (ou à densité)

Les notions relatives aux variables aléatoires continues (v.a.c.) sont essentiellement les mêmes que celles des v.a.d., seule l'apparence du mode de calcul change, le signe  $\sum$  étant remplacé par le signe  $\int$ .

Nous suivrons le plan d'étude d'une v.a.d. en apportant les modifications nécessaires et en signalant les différences éventuelles. Lorsque  $X$  est une v.a.d.,  $S = X(\Omega)$  est fini ou dénombrable et la loi de  $X$  est définie à partir d'une suite à termes positifs c'est-à-dire une fonction sur  $\mathbb{N}$  (ou  $S$ ) positive et de somme 1. Dans le cas d'une v.a.c.  $X$ ,  $S = X(\Omega)$  n'est pas dénombrable et la loi de  $X$  sera définie à partir d'une fonction sur  $\mathbb{R}$  positive et d'intégrale 1 et cela nous amène à parler de densité.

### 5.1 Densité d'une probabilité

Une densité de probabilité  $f$  (sur  $\mathbb{R}$ ) est une fonction positive  $f : \mathbb{R} \rightarrow \mathbb{R}_+$ , intégrable telle que  $\int_{-\infty}^{\infty} f(x)dx = 1$ . (Par exemple  $f$  continue par morceaux et tout segment et  $\int_{-\infty}^{\infty} f(x)dx := \lim_{K \rightarrow \infty} \int_{-K}^K f(x)dx = 1$ )

**Définition 43.** Une variable aléatoire  $X : \Omega \rightarrow R$  est dite continue (v.a.c.) si sa loi  $P_X$  admet une densité  $f$ , c'est-à-dire si l'on a

$$\text{pour toute partie borélienne } B \subset \mathbb{R}, P(X \in B) = \int_B f(x)dx.$$

Cette densité est alors unique et on dit que  $f$  est la densité de  $X$  ou que  $X$  admet la densité  $f$ .

De façon équivalente il suffit de vérifier :

$$\text{pour tout intervalle } I = [a, b] \subset \mathbb{R}, P(X \in [a, b]) = \int_a^b f(x)dx.$$

*Remarque 6.* Soit  $X$  une v.a.c. de densité  $f$

1. Pour tout  $a \in \mathbb{R}$ ,  $P(X = a) = \int_{\{a\}} f(x)dx = 0$ . Ainsi, contrairement à une v.a.d., une v.a.c. prend avec la probabilité 0 toute valeur unique (singleton).
2. Soit  $S = X(\Omega)$ . Dans les cas usuels,  $S$  est un intervalle et, quand il est distinct de  $\mathbb{R}$ , on a  $\int_{S^c} f(x)dx = P(X \in S^c) = P(\emptyset) = 0$ . Ceci impose à  $f$  d'être nulle en dehors de  $S$ .

Ainsi, lorsque la v.a. est continue, on considère que sa loi est correctement définie quand on précise sa densité qui doit être nulle en dehors de l'ensemble de ses valeurs.

### 5.2 Fonction de répartition (f.r.) d'une v.a.c.

**Définition 44.** La fonction de répartition de  $X : \Omega \rightarrow \mathbb{R}$  de densité  $f$  est la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Quand il n'y a pas de confusion à craindre, on note plus simplement  $F$  la f.r. de  $X$ .

**Proposition 26.** Toute fonction de répartition  $F_X$  d'une v.a.c. est croissante, **continue** et vérifie  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

*Remarque 7.* Soit  $X$  une v.a. dont la densité est inconnue mais dont les propriétés disponibles permettent de calculer sa f.r.  $F$ . Alors, lorsque  $F$  est continue et dérivable par morceaux, on peut affirmer que  $X$  est une v.a.c. qui admet la densité  $f$  obtenue en dérivant  $F$  aux points où elle est dérivable

$$f(x) = F'(x),$$

et en donnant à  $f$  des valeurs arbitraires aux points où  $F$  n'est pas dérivable.

En fait on peut montrer que la f.r. d'une v.a.c. détermine complètement la loi de cette variable.

### 5.3 Lois à densité classiques

*Exemple 37.* La **loi uniforme sur le segment**  $[a, b]$ , notée  $\mathcal{U}([a, b])$  est la loi de densité  $f(x) = \frac{1}{b-a}1_{[a,b]}(x)$ .

*Exemple 38.* La **loi exponentielle de paramètre**  $\lambda$ , notée  $\mathcal{E}(\lambda)$ . a pour densité  $f(x) = \lambda e^{-\lambda x}1_{[0,\infty[}(x)$ . Elle modélise la durée de vie d'un matériel ou un processus sans mémoire car  $P(X - t > s | X > t) = P(X > s)$ . En effet, sa fonction de répartition est  $F_X(x) = 1_{[0,\infty[}(x)(1 - e^{-\lambda x})$ .

*Exemple 39.* La **loi normale**  $\mathcal{N}(m, \sigma)$  (ou loi gaussienne) a pour densité :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Elle intervient dans le théorème central limite.

*Exemple 40.* La **loi semicirculaire** (ou de Wigner) a pour densité :

$$f(x) = 1_{[-R,R]}(x) \frac{2}{\pi R^2} \sqrt{R^2 - x^2}.$$

Elle intervient dans la loi des valeurs propres des grandes matrices aléatoires.

Une v.a.c. a les mêmes paramètres qu'une v.a.d. Ces paramètres s'obtiennent par une méthode de calcul appropriée comme il sera précisé dans la suite, où  $X$  désigne v.a.c. de densité  $f$  et  $S = X(\Omega)$ .

### 5.4 Moyenne ou Espérance ou Moment d'ordre 1

**Définition 45.** La v.a.d.  $X$  est dite d'ordre 1 lorsque la série  $\int_{-\infty}^{\infty} |x|f(x)dx = \lim_{K \rightarrow \infty} \int_{-K}^K |x|f(x)dx < \infty$ . Dans ce cas, on appelle Moyenne ou Espérance ou Moment d'ordre 1 de  $X$  la quantité

$$m = E[X] = \int_{-\infty}^{\infty} x f(x) dx = \lim_{K \rightarrow \infty} \int_{-K}^K x f(x) dx.$$

Remarquons que si  $X$  est bornée, la v.a.  $X$  est d'ordre 1. On note  $L^1(\Omega)$  l'ensemble des v.a. d'ordre 1 (tels que  $E(|X|) < \infty$ ). C'est un espace vectoriel.

*Exemple 41.* Si  $X \sim \mathcal{U}([a, b])$

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}.$$

*Exemple 42.* Si  $X \sim \mathcal{E}(\lambda)$

$$E(X) = \int_0^{\infty} \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

*Exemple 43.* Si  $X \sim \mathcal{N}(m, \sigma^2)$   $E(X) = m$ .

La moyenne est une forme linéaire croissante sur l'espace vectoriel  $L^1(\Omega)$  des v.a. d'ordre 1. En d'autres termes, on a

**Proposition 27.** Si  $X$  et  $Y$  sont deux v.a. d'ordre 1, alors

1. Pour tout couple  $(a, b)$  de scalaires,  $aX + bY$  est d'ordre 1 et  $E(aX + bY) = aE(X) + bE(Y)$ .

2.  $X \leq Y \Rightarrow E[X] \leq E[Y]$  . En particulier,  $X \geq 0 \Rightarrow E[X] \geq 0$ .
3.  $Z$  v.a.d. et  $|Z| \leq |X|$  implique  $Z$  est d'ordre 1.
4.  $|E(X)| \leq E(|X|)$ .
5.  $E[c] = c$ ,  $a \leq X \leq b \Rightarrow a \leq E[X] \leq b$ .
6. Si  $X$  et  $Y$  sont indépendantes,  $XY$  est d'ordre 1 et  $E(XY) = E(X)E(Y)$ .

On peut aussi s'intéresser au calcul de la moyenne d'une fonction quelconque d'une v.a.d. en généralisant la relation précédente. Ceci permettra en même temps d'obtenir différentes caractérisations de la loi d'une v.a.d. à travers le calcul de la moyenne d'une fonction bien choisie. Ceci nous conduit au théorème de transfert.

## 5.5 Théorème de transfert

**Théorème 28.** Soit  $X : \Omega \rightarrow \mathbb{R}$  une v.a.c.,  $S = X(\Omega)$  et  $h : S \rightarrow \mathbb{C}$  une fonction telle que  $\int_{-\infty}^{\infty} |h(s)|f(s)ds$  soit convergente. Alors, la v.a.  $h(X)$  admet une moyenne et

$$E[h(X)] = \int_{-\infty}^{\infty} h(s)f(s)ds.$$

Remarquons que lorsque  $X$  et  $Y$  sont égales en loi, on voit que, pour toute fonction  $h$ ,  $E[h(X)] = E[h(Y)]$ , pourvu que l'un des deux membres existe. En particulier,  $X$  et  $Y$  ont mêmes paramètres (moyenne, variance, f.r., f.c. etc).

**Proposition 29.**  $X$  et  $Y$  v.a.d. sont indépendantes si et seulement si

$$E[h(X)g(Y)] = E[h(X)]E[g(Y)],$$

pour toute fonction  $h, g$  telle que  $E[|h(X)|] < \infty$ ,  $E[|g(Y)|] < \infty$ . De plus, pour caractériser l'indépendance il suffit de prendre  $h, g$  bornées.

En plus de la f.r., la fonction suivante constitue un outil intéressant pour caractériser la loi de  $X$ .

**Définition 46.** La fonction caractéristique (f.c.) de  $X$  est définie par

$$\varphi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{its} f(s)ds,$$

pour tout  $t \in \mathbb{R}$ .

la fonction  $\varphi_X$  caractérise la loi de  $X$ .

## 5.6 Moments d'une v.a.c.

**Définition 47.** Soit  $k$  un entier positif. La v.a.c.  $X$  est dite d'ordre  $k$  si  $\int_{-\infty}^{\infty} |s|^k f(s)ds$  est convergente.

On voit que si  $X$  est d'ordre  $k$ , alors elle est d'ordre  $j$  pour tout  $j \leq k$ . Remarquons que si  $X$  est bornée, la v.a.  $X$  est d'ordre  $k$  pour tout  $k$  et la loi est caractérisée par la donnée de tous les moments  $E[X^k]$ ,  $k \in \mathbb{N}$ . On note  $L_d^k(\Omega)$  l'ensemble des v.a.d. d'ordre  $k$ . C'est un espace vectoriel.

**Définition 48.** Soit  $X$  une v.a. d'ordre  $k$ . Alors Le moment d'ordre  $k$  de  $X$  est défini par la quantité

$$m_k = E[X^k] = \int_{-\infty}^{\infty} s^k f(s)ds.$$

## 5.7 Variance et Écart-type d'une v.a.c. Variable centrée réduite

On suppose que  $X$  est une v.a.d. d'ordre 2 de moyenne  $m$ .

**Définition 49.** La *variance* de  $X$  est définie par la quantité  $v = \sigma^2 = V[X] = E[(X - m)^2] = E[X^2] - m^2$ . L'*écart-type* de  $X$  est la quantité  $\sigma = (V[X])^{1/2}$ .

La variable centrée réduite associée à  $X$  est  $U = \frac{X-m}{\sigma}$ . On a  $E[U] = 0$  et  $V[U] = 1$ .

*Exemple 44.* La loi  $\mathcal{N}(m, \sigma^2)$  a pour moyenne  $m$  et variance  $\sigma^2$ , sa variable centrée réduite a pour loi  $\mathcal{N}(0, 1)$  (loi normale centrée réduite).

*Remarque 8.* 1. La moyenne et l'écart-type d'une v.a.  $X$  constituent un résumé intéressant de la répartition de  $X$ . La moyenne donne une idée de l'ordre de grandeur de la variable et l'écart-type une idée de la dispersion des valeurs de  $X$  autour de leur moyenne. Ces deux paramètres interviennent dans plusieurs inégalités.

2. Le recours à la variable centrée réduite permet et facilite certains calculs. Il permet aussi de comparer entre elles de façon cohérente plusieurs variables.

La variance est définie sur l'espace vectoriel  $L^2(\Omega)$  des v.a. d'ordre 2.

**Proposition 30.** *si  $X$  et  $Y$  sont deux v.a.c. d'ordre 2, on a les propriétés suivantes :*

1.  $V[c] = 0$ ,  $V[X] \geq 0$  et  $V[X] = 0 \Leftrightarrow X = m$ .
2.  $V[X] = E[X^2] - (E[X])^2$ .
3.  $V(aX + b) = a^2V(X)$ .
4. Si  $X$  et  $Y$  sont indépendantes,  $V(X + Y) = V(X) + V(Y)$ .

## 5.8 Inégalités de Markov et de Bienaymé-Tchébychev

**Proposition 31** (Inégalité de Markov). *. Soit  $Z$  une v.a. positive et  $a > 0$ . Alors,*

$$P(Z \geq a) \leq \frac{E[Z]}{a}.$$

**Proposition 32** ( Inégalité de Bienaymé-Tchébychev). *Soit  $X$  une v.a. de moyenne  $m$  et d'écart-type  $\sigma$ . Alors, pour tout  $\epsilon > 0$ ,*

$$P(|X - m| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

*Exemple 45.* Soit  $X$  le nombre de rames de métro qui passent entre 16H et 17H en une station donnée. On a observé que le nombre moyen de rames est  $E(X) = 55$  et que la variance est  $V(X) = 20$ . La probabilité que le nombre de rame soit compris entre 50 et 60 est au moins  $P(X \in [50, 60]) = P(|X - E(X)| \leq 5) \geq 1 - P(|X - E(X)| \geq 5) \geq 1 - \frac{V(X)}{5^2} = 1 - 20/25 = 1/5$ .

Voici une conséquence importante, la loi faible des grands nombres, que l'on renforcera au prochain chapitre.

**Proposition 33.** *Soit  $(X_n)_{n \geq 1}$  une suite de variables indépendantes de même loi, telle que  $X_n \in L^2(\Omega)$  est d'ordre 2 et  $m = E(X_n)$ . Alors pour tout  $\epsilon > 0$*

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - m\right| > \epsilon\right) \rightarrow_{n \rightarrow \infty} 0.$$

*Démonstration.* En effet en appliquant l'inégalité de Tchébychev à  $Y_n = \frac{X_1 + \dots + X_n}{n}$  qui vérifie  $E(Y_n) = m$ ,  $V(Y_n) = \frac{1}{n^2}(\sum_{i=1}^n)V(X_i) = \frac{V(X_1)}{n}$  par indépendance et équidistribution, on obtient :

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - m\right| > \epsilon\right) \leq \frac{V(Y_n)}{\epsilon^2} = \frac{V(X_1)}{n\epsilon^2} \rightarrow 0.$$

□

## 6 Simulation

On résume quelques méthodes de simulation

*Exemple 46.* A partir de la fonction de répartition  $F_X$ , on pose

$$F_X^{-1}(t) = \inf\{x \in \mathbb{R}, F_X(x) \geq t\}.$$

Alors si  $U := ALEA$  de loi uniforme sur  $[0, 1]$ ,  $X := F_X^{-1}(U)$  est une v.a. de loi de fonction de répartition  $F_X$ .

En effet, l'idée de la preuve que l'on ne détaillera pas est la suivante :  $P(F_X^{-1}(U) \leq t) = P(U \leq F_X(t)) = F_X(t)$ .

Cette méthode peut être assez coûteuse en pratique, mais c'est la seule méthode générale.

*Exemple 47.* (Changement de variables) Si  $X$  est v.a. réelle admettant une densité  $f_X$  et  $\varphi$  dérivable strictement croissante de  $\mathbb{R} \rightarrow \mathbb{R}$ , alors  $\varphi(X)$  admet pour densité :

$$f_Y(y) = \frac{1}{\varphi'(\varphi^{-1}(y))} f_X(\varphi^{-1}(y)).$$

*Exemple 48.* Obtention de variables indépendantes. Si  $U := ALEA$ , la suite  $X_n := Ent(2^n U) - 2 \times Ent(2^{n-1} U)$  ( $n \in \mathbb{N}^*$ ,  $X_n$  est le  $n$ -ième terme du développement décimal) forme une suite de v.a. de Bernoulli  $B(\frac{1}{2})$  indépendantes. Réciproquement, étant donné une suite de v.a. de Bernoulli indépendante :

$$X = \sum_{n=1}^{\infty} \frac{X_n}{2^n}.$$

est une loi uniforme sur  $[0, 1]$ .

Par exemple

$$X_1 = \sum_{n=1}^{\infty} \frac{X_{2n}}{2^n}, X_2 = \sum_{n=1}^{\infty} \frac{X_{2n+1}}{2^n}$$

sont des lois  $\mathcal{U}([0, 1])$  indépendantes. On produit ainsi autant de v.a. indépendantes que l'on souhaite.

*Exemple 49.* (Sur les générateurs de nombres *pseudo*-aléatoires) Ce sont des suites déterministes produisant des suites longues de nombres presque indépendantes et uniformes.

La méthode la plus commune est de prendre une suite

$$x_{n+1} = ax_n + b \text{ modulom}$$

en partant d'une donnée initiale  $x_0$  et on considère  $x_{n+1}/m$  comme la valeur aléatoire.

Autrement dit, on calcule le reste de la division euclidienne de  $ax_n + b = mq + r$  et on retourne  $x_{n+1}/m = r/m$ .

Cette suite est toujours périodique de période au plus  $m$ , pour  $a$  et  $b$  bien choisis, la période est en effet proche de  $m$  et les sous-suites de longueurs autour de  $m/10$  simulent bien des variables aléatoires et indépendantes. Choisir  $a$  et  $b$  est un problème difficile. Par exemple, la fonction *rand* de scilab utilise  $a = 84331861, b = 453816693, m = 2^{31}$ . (la fonction *grand*, meilleure, a une période  $2^{19937} - 1$ .)

# Chapitre 5

## Couples, Vecteurs aléatoires, Théorèmes limites

Le but de ce chapitre est surtout de décrire la loi d'une paire de variables aléatoires  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  simultanément dans le cas où elles ne sont pas indépendantes.

Plus généralement, on peut considérer le cas des vecteurs aléatoires  $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ .

Par ailleurs, à titre préliminaire de résultats sur les limites de variables aléatoires indépendantes très utilisées en statistique, nous énoncerons les principaux théorèmes limites pour les suites de v.a. indépendantes.

### 1 Couples aléatoires

On considère la tribu borélienne  $\mathcal{B}(\mathbb{R}^2)$  (engendrée par les produits d'intervalles  $[a, b] \times [c, d]$  qui contient donc les unions dénombrables de tels produits).

Comme pour les variables aléatoires, on va définir la loi du couple comme probabilité, ayant une fonction de répartition, une fonction caractéristique etc. permettant de la décrire. On verra ensuite le cas particulier des v.a. discrètes et continues.

**Définition 50.** La loi du couple de v.a.  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  est la probabilité  $P_{(X,Y)}$  sur  $\mathcal{B}(\mathbb{R}^2)$  définie par :

$$P_{(X,Y)}(B) = P((X, Y) \in B).$$

Elle est caractérisée par le cas  $B = [a, b] \times [c, d]$

$$P_{(X,Y)}([a, b] \times [c, d]) = P(X \in [a, b], Y \in [c, d]).$$

Les lois  $P_X$  et  $P_Y$  sont dites lois marginales et sont déterminées par :

$$P_X(A) = P_{(X,Y)}(A \times \mathbb{R}), P_Y(B) = P_{(X,Y)}(\mathbb{R} \times B).$$

**Définition 51.** La fonction de répartition de  $(X, Y) : \Omega \rightarrow \mathbb{R}^2$  est la fonction  $F_{(X,Y)} : \mathbb{R}^2 \rightarrow [0, 1]$  définie par

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y).$$



**Proposition 34.**  $x \mapsto F_{(X,Y)}(x, y)$  et  $y \mapsto F_{(X,Y)}(x, y)$  sont croissantes et

$$\lim_{x \rightarrow \infty} F_{(X,Y)}(x, y) = F_Y(y), \lim_{y \rightarrow \infty} F_{(X,Y)}(x, y) = F_X(x).$$

$F_{(X,Y)}$  caractérise  $P_{(X,Y)}$ .

**Définition 52.** La fonction caractéristique (f.c.) de  $(X, Y)$  est définie par

$$\varphi_{(X,Y)}(t, s) = E[e^{itX} e^{isY}],$$

pour tout  $t, s \in \mathbb{R}^2$ .  $\varphi_{(X,Y)}$  caractérise  $P_{(X,Y)}$ .

**Proposition 35.** Les assertions suivantes sont équivalentes :

1.  $X$  et  $Y$  sont indépendantes

2.

$$\forall (t, s) \in \mathbb{R}^2, \varphi_{(X,Y)}(t, s) = \varphi_X(t)\varphi_Y(s),$$

3.

$$\forall (t, s) \in \mathbb{R}^2, F_{(X,Y)}(t, s) = F_X(t)F_Y(s).$$

*Exemple 50.* Soit  $(X, Y)$  prenant les valeurs  $(1, 1), (1, 2), (2, 1)$  avec les probabilités  $\frac{1}{2}, \frac{1}{4}, \frac{1}{4}$ . On a donc

$$P(X = 1) = P(X = 1, Y = 1) + P(X = 1, Y = 2) = \frac{3}{4},$$

et

$$P(Y = 1) = P(X = 1, Y = 1) + P(X = 2, Y = 1) = \frac{3}{4},$$

Or

$$P(X = 1, Y = 1) = \frac{1}{2} \neq \left(\frac{3}{4}\right)^2 = P(X = 1)P(Y = 1)$$

donc  $X$  et  $Y$  ne sont pas indépendantes. On a surtout introduit les couples pour étudier le cas des v.a. non indépendantes. Plus généralement la loi marginale est donnée par  $P(X = 1) = \frac{3}{4}$

On résume souvent cela dans le tableau :

	1	2	Y
1	0.5	0.25	0.75
2	0.25	0	0.25
X	0.75	0.25	1

Le tableau suivant avec les mêmes marginales correspond à des variables indépendantes. On voit au passage que les marginales ne suffisent pas à caractériser la loi du couple (le tableau contient plus d'informations que les dernières lignes et colonnes).

	1	2	Y'
1	0.5625	0.1875	0.75
2	0.1875	0.0625	0.25
X'	0.75	0.25	1

**Définition 53.** Soient  $X, Y \in L^1(\Omega)$  des variables d'ordre 1, l'espérance de  $(X, Y)$  est le vecteur  $(E(X), E(Y))$ .

Soient  $X, Y \in L^2(\Omega)$  des variables d'ordre 2, la matrice de covariance de  $(X, Y)$  est la matrice  $2 \times 2$  (symétrique) :

$$\begin{pmatrix} E((X - E(X))^2) & E((X - E(X))(Y - E(Y))) \\ E((X - E(X))(Y - E(Y))) & E((Y - E(Y))^2) \end{pmatrix}.$$

On note  $Cov(X, Y) = E((X - E(X))(Y - E(Y)))$ .

## 1.1 Couples de v.a. discrètes

Si  $X, Y$  sont des v.a. discrètes, le couple  $(X, Y)$  est à valeur  $(X, Y)(\Omega) \subset X(\Omega) \times Y(\Omega)$  qui est encore fini ou dénombrable.

La loi est donc caractérisée par les valeurs :

$$P(X = i, Y = j), (i, j) \in (X, Y)(\Omega).$$

Elles vérifient bien sûr

$$\sum_{(i,j) \in (X,Y)(\Omega)} P(X = i, Y = j) = 1.$$

Un Théorème de transfert énonce alors que pour toute fonction  $h : (X, Y)(\Omega) \rightarrow \mathbb{R}$ , on a :

$$E(h(X, Y)) = \sum_{(i,j) \in (X,Y)(\Omega)} h(i, j)P(X = i, Y = j).$$

Cela permet de trouver tous les paramètres du couple.

*Exemple 51.* Une urne contient  $r$  boules rouges et  $v$  boules vertes, soit  $b = r + v$  boules au total. On tire de façon successive avec remise  $n$  boules de l'urne. On note  $X$  le nombre de boules rouges tirées,  $Y$  le nombre de boules vertes tirées. Calculons la loi du couple  $(X, Y)$ . On a  $(X, Y)(\Omega) = \{(k, l) \in \mathbb{N}^{*2}, k + l = n\}$ . On calcule  $\Omega = \llbracket 1, b \rrbracket^2$ ,  $Card(\Omega) = b^n$ ,  $Card(X = k, Y = l) = r^k v^l C_k^l$  (choix des boules rouges tirées  $r^k$ , des vertes  $v^l$  et des positions des rouges et des vertes  $C_k^l$ ) donc la loi de  $(X, Y)$  est déterminée par :

$$P(X = k, Y = l) = \frac{k!}{l!(k-l)!} \left(\frac{r}{b}\right)^k \left(\frac{v}{b}\right)^l.$$

La loi marginale de  $X$  est donnée par  $P(X = k) = P(X = k, Y = n - k)$ , c'est donc la loi binomiale  $\mathcal{B}(n, \frac{r}{b})$ , on a  $Y = n - X$ .

## 1.2 Couples de v.a. continues

**Définition 54.** Un couple de v.a.  $(X, Y)$  est dit à *densité* s'il existe une fonction (densité)  $f_{(X,Y)} : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  intégrable telle que  $\int \int_{\mathbb{R}^2} f_{(X,Y)}(s, t) dt ds = 1$  et avec :

$$F_{(X,Y)}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{(X,Y)}(t, s) dt ds.$$

**Proposition 36** (Transfert). *On a alors pour tout  $C \in \mathcal{B}(\mathbb{R}^2)$*

$$P((X, Y) \in C) = \int \int_C f_{(X,Y)}(t, s) dt ds,$$

*et pour tout  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  telle que  $\int \int_{\mathbb{R}^2} |h(x, y)| f_{(X,Y)}(x, y) < \infty$ , alors  $h(X, Y) \in L^1(\Omega)$  et*

$$E(h(X, Y)) = \int \int_{\mathbb{R}^2} h(x, y) f_{(X,Y)}(x, y).$$

*En particulier, les lois marginales  $P_X$  de  $X$  et  $P_Y$  de  $Y$  ont des densités :*

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy, \quad f_Y(y) = \int_{\mathbb{R}} f(x, y) dx.$$

*De plus  $X$  et  $Y$  sont **indépendantes** si et seulement si :*

$$f_{(X,Y)}(x, y) = f_X(x) f_Y(y)$$

*Exemple 52* (Loi uniforme sur le disque unité). On considère la loi du couple  $(X, Y)$  de densité  $f(x, y) = \frac{1}{\pi} 1_{[0,1]}(x^2 + y^2)$ . On vérifie que  $\int \int_{\mathbb{R}^2} f(x, y) dx dy$  vaut l'aire du disque unité divisée par  $\pi$ , soit 1. La loi marginale  $P_X$  est la loi semicirculaire avec  $R = 1$ , en effet

$$f_X(x) = 1_{[-1,1]}(x) \int_{|y| \leq \sqrt{1-x^2}} \frac{1}{\pi} = 1_{[-1,1]}(x) \frac{2}{\pi} \sqrt{1-x^2}.$$

## 2 Vecteurs aléatoires

On considère la tribu borélienne  $\mathcal{B}(\mathbb{R}^n)$  (engendrée par les produits d'intervalles).

**Définition 55.** La loi du vecteur aléatoire  $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$  est la probabilité  $P_{(X_1, \dots, X_n)}$  sur  $\mathcal{B}(\mathbb{R}^n)$  définie par :

$$P_{(X_1, \dots, X_n)}(B) = P((X_1, \dots, X_n) \in B).$$

Elle est caractérisée par la fonction de répartition  $F_{(X_1, \dots, X_n)} : \mathbb{R}^n \rightarrow [0, 1]$  définie par

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n),$$

ou la fonction caractéristique définie par

$$\varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = E[e^{i\langle t, X \rangle}],$$

pour tout  $t = (t_1, \dots, t_n) \in \mathbb{R}^n$  et  $\langle t, X \rangle := \sum_{i=1}^n t_i X_i$ .

**Définition 56.** Soient  $X_1, \dots, X_n \in L^1(\Omega)$  des variables d'ordre 1, l'espérance de  $(X_1, \dots, X_n)$  est le vecteur  $(E(X_1), \dots, E(X_n))$ .

Soient  $X_1, \dots, X_n \in L^2(\Omega)$  des variables d'ordre 2, la matrice de covariance de  $(X_1, \dots, X_n)$  est la matrice  $n \times n$  (symétrique)  $(Cov(X_i, X_j))_{ij}$ .

Si  $X_i$  sont des v.a. discrètes, la loi est donc caractérisée par les valeurs :

$$P(X_1 = x_1, \dots, X_n = x_n), (x_1, \dots, x_n) \in (X_1, \dots, X_n)(\Omega).$$

Un Théorème de transfert énonce alors que pour toute fonction  $h : (X_1, \dots, X_n)(\Omega) \rightarrow \mathbb{R}$ , on a :

$$E(h(X_1, \dots, X_n)) = \sum_{(x_1, \dots, x_n) \in (X_1, \dots, X_n)(\Omega)} h(x_1, \dots, x_n) P(X = x_1, \dots, X_n = x_n).$$

**Définition 57.** Un v.a.  $(X_1, \dots, X_n)$  est dit à loi continue s'il existe une fonction (densité)  $f : \mathbb{R}^n \rightarrow \mathbb{R}_+$  intégrable telle que  $\int \dots \int_{\mathbb{R}^n} f(x) dx = 1$  et avec :

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \dots dt_n.$$

Alors on a la formule de transfert, pour tout  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  telle que  $\int \dots \int_{\mathbb{R}^n} |h(x_1, \dots, x_n)| f(x_1, \dots, x_n) < \infty$ , alors  $h(X_1, \dots, X_n) \in L^1(\Omega)$

$$E(h(X_1, \dots, X_n)) = \int \dots \int_{\mathbb{R}^n} h(x_1, \dots, x_n) f(x_1, \dots, x_n).$$

### 3 Théorèmes limites (facultatifs)

Les deux résultats suivants, la loi forte des grands nombres et le théorème central limite (TCL) sont très importants en statistique et en probabilité. Ils disent respectivement qu'une suite de variables aléatoires indépendantes et de même loi (i.i.d) ont une moyenne empirique qui tend presque sûrement (avec probabilité 1) vers leur moyenne commune  $m$  et que si on retranche la moyenne, l'écart à la moyenne correctement normalisé se comporte comme une loi gaussienne.

**Théorème 37** (Loi forte des grands nombres). *Soit  $(X_i)_{i \geq 1}$  une suite de v.a. d'ordre 1 indépendantes et identiquement distribuées ( $P_{X_i} = P_{X_1}$ ), alors*

$$P(\{\omega : \frac{X_1(\omega) + \dots + X_n(\omega)}{n} \rightarrow E(X_1)\}) = 1.$$

**Théorème 38** (Théorème central limite TCL). *Soit  $(X_i)_{i \geq 1}$  une suite de v.a. d'ordre 2 indépendantes et identiquement distribuées ( $P_{X_i} = P_{X_1}$ ) de moyenne  $m = E(X_1)$  et de variance  $\sigma^2 = V(X_1)$  et soit  $Y$  une v.a. de loi normale  $N(0, \sigma^2)$ , alors*

$$P(\{\omega : \frac{X_1(\omega) + \dots + X_n(\omega) - nm}{\sqrt{n}} \in [a, b]\}) \xrightarrow{n \rightarrow \infty} P(Y \in [a, b]) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2\sigma^2}} dt.$$