

# Chapitre 1

## Éléments d'analyse combinatoire

### 1.1 Quelques définitions

Disposition sans répétition : c'est une disposition où un élément peut apparaître 0 ou 1 fois.

Disposition avec répétition : un élément peut figurer plus d'une fois.

Disposition ordonnée : l'ordre d'obtention d'un élément est important.

Ex. les éléments constituant la plaque minéralogique d'un véhicule.

Disposition non-ordonnée : l'ordre d'obtention d'un élément n'est pas important, on n'en tient pas compte dans la caractérisation de la disposition.

Ex. Les numéros issus d'un tirage du loto.

Exemple 1 : On considère un ensemble à deux éléments  $\{a, b\}$ . Avec deux tirages sans répétition, on peut obtenir  $\{a, b\}$  ou  $\{b, a\}$ ; Avec deux tirages avec répétition, on peut obtenir  $\{a, a\}$ ,  $\{a, b\}$ ,  $\{b, a\}$  ou  $\{b, b\}$ . Cela correspond à un tirage avec remise.

Exemple 2 : Prenons un jeu de dé à 6 faces (éléments discernables) numérotées par  $\Omega = \{1; 2; 3; 4; 5; 6\}$ . Après 3 jets, nous obtenons la réalisation  $A = (2; 5; 1)$ ; nous réitérons les jets et nous obtenons  $B = (5; 1; 2)$ .  $A$  et  $B$  sont équivalents si nous considérons que les dispositions sont non-ordonnées. En revanche, ils ne sont pas équivalents si nous sommes dans le cadre d'une disposition ordonnée.

La valeur *Factorielle*( $n$ ), notée  $n!$  est définie par  $n! = 1 \cdot 2 \cdots n = \prod_{i=1}^n i$ . Par convention  $0! = 1$ . Nous pouvons également utiliser une définition récursive

$$n! = n \cdot (n - 1)!$$

### 1.2 Arrangement avec répétition

Soit  $\Omega$  un ensemble composé de  $n$  éléments :  $\text{card}(\Omega) = n$ . Nous constituons un échantillon  $E$  de taille  $p$  ( $\text{card}(E) = p$ ) à partir des éléments de  $\Omega$ . Si nous avons à choisir  $p$  éléments parmi  $n$  dans une disposition ordonnée (les places sont distinctes) et avec répétition (on peut choisir le même élément plusieurs fois), on dit qu'on a un arrangement de  $p$  éléments parmi  $n$ . Le nombre d'arrangement avec répétition est  $n^p$ .

N.B. Dans ce cas, il est possible que  $p > n$ .

Réaliser un arrangement avec répétition des éléments de  $\Omega$ , c'est aussi définir une application d'un ensemble  $E$  à  $p$  éléments dans  $\Omega$ . L'ensemble des applications de  $E$  dans  $\Omega$  sera noté  $\Omega^E$  et on a  $\#(\Omega^E) = (\#\Omega)^{\#E}$ .

### 1.3 Arrangement sans répétition

Soit  $\Omega$  un ensemble avec  $\text{card}(\Omega) = n$ . On constitue un échantillon de taille  $p$  ( $p \leq n$ ), la disposition est ordonnée et sans répétition. On dit qu'on a un arrangement sans répétition de  $p$  éléments parmi  $n$ . Le nombre de  $p$ -arrangements d'un ensemble à  $n$  éléments est :

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

Réaliser un arrangement sans répétition des éléments de  $\Omega$ , c'est déterminer un  $p$ -uplet  $(x_1, \dots, x_p)$  d'éléments de  $\Omega$  deux à deux distincts. C'est aussi définir une application injective d'un ensemble  $E$  à  $p$  éléments dans  $\Omega$  à  $n$  éléments.

### 1.4 Permutation sans répétition

C'est un arrangement sans répétition de  $n$  éléments parmi  $n$ .

$$P_n = A_n^n = \frac{n!}{(n-n)!} = n!$$

Réaliser une permutation des éléments de  $\Omega$ , c'est réaliser un tirage exhaustif sans remise des éléments de  $\Omega$  en tenant compte de l'ordre du tirage. C'est aussi définir une bijection de ensemble  $\Omega$  sur lui-même. L'ensemble des permutations d'un ensemble à  $n$  éléments s'appelle le groupe symétrique d'ordre  $n$  et se note  $S_n$ . On a  $\#S_n = n!$ .

### 1.5 Permutation avec répétition

On appelle permutation avec répétition de  $p$  éléments où  $n$  sont distincts ( $n \leq p$ ), une disposition ordonnée de l'ensemble de ces  $p$  éléments où le premier figure  $p_1$  fois, le second  $p_2$  fois, etc., tel que  $p_1 + p_2 + \dots + p_n = p$ . Le nombre de permutation avec répétitions est  $\frac{p!}{p_1!p_2!\dots p_n!}$

Démonstration : (Voir préalablement la définition d'une Combinaison sans répétition)

Pour construire un  $p$ -uplet correspondant à une combinaison contenant  $p_1$  fois  $x_1$ ,  $p_2$  fois  $x_2$ , ...,  $p_n$  fois  $x_n$ , il suffit :

- de choisir les  $p_1$  emplacements des  $x_1$ , parmi  $p_1 + p_2 + \dots + p_n$  places disponibles,
- de choisir les  $p_2$  emplacements des  $x_2$ , parmi les  $p_2 + \dots + p_n$  places restantes,
- etc.
- de choisir les  $p_n$  emplacements des  $x_n$ , parmi les  $p_n$  places restantes.

Au total, il y a

$$C_{p_1+p_2+\dots+p_n}^{p_1} \cdot C_{p_2+\dots+p_n}^{p_2} \cdot \dots \cdot C_{p_n}^{p_n} = \frac{p!}{p_1!p_2!\dots p_n!}$$

Exemple [Nombre d'anagrammes du mot MATHÉMATIQUE] : nous voyons qu'en échangeant les deux lettres A, le mot reste identique, et par contre en transposant les lettres Ê et E nous obtenons un mot différent. (M :2;A :2;T :2;H :1;Ê :1;I :1;Q :1;U :1;E :1) :  $\# \text{Anagrammes} = 12!/(2!2!2!)$

Exemple 2 : Nombre de quartets binaires de poids de Hamming égal à 2;

Il y en a  $6 = 4!/(2!2!) : (0011), (0101), (0110), (1001), (1010), (1100)$ .

## 1.6 Combinaison sans répétition

On considère un ensemble  $\Omega$  constitué de  $n$  éléments tous discernables. On forme un échantillon de taille  $p$ . Si la disposition est non-ordonnée et sans répétition, on dit que l'on a une combinaison sans répétition de  $p$  éléments parmi  $n$ . Le nombre de ces combinaisons se note  $C_n^p$  ou  $\binom{n}{p}$ .

$$C_n^p = \frac{n!}{p!(n-p)!}$$

Propriétés :

1.  $C_n^0 = C_n^n = 1$
2.  $C_n^p = C_n^{n-p}$  (complémentaire)
3.  $C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$  (triangle de Pascal)
4.  $C_n^p = \frac{A_n^p}{p!}$

Preuve que  $C_n^p = C_{n-1}^p + C_{n-1}^{p-1}$  :

$$\begin{aligned} C_{n-1}^p + C_{n-1}^{p-1} &= \frac{(n-1)!}{p!(n-p-1)!} + \frac{(n-1)!}{(p-1)!(n-p)!} \\ &= \frac{(n-1)! \cdot (n-p)}{p!(n-p)!} + \frac{p \cdot (n-1)!}{p!(n-p)!} \\ &= \frac{n \cdot (n-1)!}{p!(n-p)!} = C_n^p \end{aligned}$$

**Proposition 1.6.1 (Formule du binôme)**

$$(a+b)^n = \sum_{p=0}^n C_n^p \cdot a^p \cdot b^{n-p}.$$

Exercice : preuve de la formule du binôme par récurrence sur  $n$

Preuve :

$$\begin{aligned} (a+b)^{n+1} &= (a+b)(a+b)^n \\ &= (a+b) \sum_{p=0}^n C_n^p \cdot a^p \cdot b^{n-p} \\ &= \sum_{p=0}^n C_n^p \cdot a^{p+1} \cdot b^{n-p} + \sum_{p=0}^n C_n^p \cdot a^p \cdot b^{n+1-p} \\ &= \sum_{p'=1}^{n+1} C_n^{p'-1} \cdot a^{p'} \cdot b^{n+1-p'} + \sum_{p=0}^n C_n^p \cdot a^p \cdot b^{n+1-p} \\ &= \left( \sum_{p=1}^n C_n^{p-1} \cdot a^p \cdot b^{n+1-p} + C_n^n a^{n+1} b^0 \right) + \left( C_n^0 a^0 b^{n+1} + \sum_{p=1}^n C_n^p \cdot a^p \cdot b^{n+1-p} \right) \\ &= a^{n+1} + \sum_{p=1}^n \underbrace{(C_n^{p-1} + C_n^p)}_{C_{n+1}^p} \cdot a^p \cdot b^{n+1-p} + b^{n+1} \\ (a+b)^{n+1} &= \sum_{p=0}^{n+1} C_{n+1}^p \cdot a^p \cdot b^{n+1-p}. \end{aligned}$$

## 1.7 Combinaison avec répétition

C'est une disposition non-ordonnée de  $p$  éléments, à choisir parmi  $n$  éléments discernables, avec répétition. Le nombre de combinaisons avec répétitions de  $n$  objets pris  $p$  à  $p$  est :

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

Exemple : [jeu de domino] Les pièces sont constituées en disposant côte à côte deux éléments de l'ensemble  $\{\text{blanc}, 1, 2, 3, 4, 5, 6\}$ . Si nous retournons un domino, nous changeons l'ordre des deux éléments, mais le domino reste identique (C'est donc une disposition non-ordonnée). Nous avons une combinaison avec répétition de 2 éléments pris parmi les 7, et au total il y a  $K_7^2 = 28$  dominos dans un jeu.

Toute  $p$ -combinaison avec répétition peut s'écrire :

$$x_1 : k_1 \text{ fois}, \dots, x_n : k_n \text{ fois}$$

avec  $0 \leq k_i \leq p$  et  $\sum_{i=1}^n k_i = p$ .

On peut ainsi mettre en bijection l'ensemble des  $p$ -combinaisons avec répétition des  $n$  éléments de  $E$  avec les applications  $f : E \rightarrow \mathbb{N}$  telles que

$$\begin{array}{l} x_1 \mapsto f(x_1) = k_1 \\ \dots \\ x_n \mapsto f(x_n) = k_n \end{array} \quad \text{vérifiant} \quad \sum_{i=1}^n f(x_i) = p$$

Exemple : Dans un jeu de dominos, un domino est une 2-combinaison avec répétition de l'ensemble  $E = \{\text{blanc}, 1, 2, 3, 4, 5, 6\}$ . Chaque domino peut être représenté par une application de  $E$  dans  $\{0, 1, 2\}$  qui associe à chaque élément de  $E$  le nombre de fois où l'élément apparaît sur le domino. Ainsi le domino [blanc, blanc], est représenté par l'application  $f$  définie par

$$f(\text{blanc}) = 2, f(1) = 0, f(2) = 0, f(3) = 0, f(4) = 0, f(5) = 0, f(6) = 0$$

et le domino [blanc, 1] par l'application  $f$  définie par

$$f(\text{blanc}) = 1, f(1) = 1, f(2) = 0, f(3) = 0, f(4) = 0, f(5) = 0, f(6) = 0.$$

On peut aussi mettre cet ensemble en bijection avec l'ensemble des manières de placer  $p$  objets dans  $n$  boîtes :

$$\begin{array}{cccccc} \text{boîte} & 1 & \dots & i & \dots & n \\ x_1 & \dots & x_i & \dots & x_n \\ k_1 & \dots & k_i & \dots & k_n \end{array}$$

Mais placer  $p$  objets dans  $n$  boîtes c'est aussi se donner  $n + p - 1$  objets et décider que  $n - 1$  d'entre eux seront des cloisons :

$$\underbrace{0 \dots 0}_{k_1} | \underbrace{0 \dots 0}_{k_2} | \dots | \underbrace{0 \dots 0}_{k_n}.$$

Inversement, à toute façon de choisir  $n - 1$  objets qui seront des cloisons, on peut associer une et une seule façon de placer  $p$  objets dans  $n$  boîtes.

Il y a une bijection entre l'ensemble des  $p$ -combinaisons avec répétition de  $E$  et l'ensemble des  $p$ -uplets croissants d'éléments de  $E$ , ou encore des applications croissantes (au sens large) de  $\{1, 2, \dots, p\}$  dans  $E$ .

Propriété :  $K_n^p = K_n^{p-1} + K_{n-1}^p$ .

Preuve :  $C_{n+p-1}^p = C_{n+p-2}^{p-1} + C_{n+p-2}^p$

# Chapitre 2

## Probabilités

### 2.1 Espace probabilisé

#### 2.1.1 Événement et ensemble fondamental

Une épreuve est une expérience dont l'issue n'est pas prévisible car répétée dans des conditions identiques, elle peut donner lieu à des résultats différents ou aléatoires (expérience aléatoire). L'ensemble des résultats possibles s'appelle l'**ensemble fondamental** (ou référentiel, univers des possibles) et sera noté  $\Omega$ .

Un **événement** est un ensemble de résultats (un sous-ensemble de l'univers) d'une expérience aléatoire. Comme l'événement est une affirmation concernant le résultat d'une expérience, nous devons pouvoir dire, pour tout résultat de l'univers, si l'événement se réalise ou non. Un événement donné, souvent défini par une proposition, est identifié à la partie de l'univers pour laquelle il est réalisé.

On exige que la collection  $\mathcal{C}$  des événements dispose de la structure d'une algèbre de Boole :

1.  $\Omega \in \mathcal{C}; \emptyset \in \mathcal{C}$ .
2. si  $A \in \mathcal{C}; \Rightarrow \bar{A} \in \mathcal{C}$ ;
3. si  $A, B \in \mathcal{C} \Rightarrow A \cup B \in \mathcal{C}$  et  $A \cap B \in \mathcal{C}$ .

On peut préciser le calcul de probabilités d'un événement  $E$ . De manière simplifiée, la probabilité théorique vaut

$$P(E) = \frac{\text{nombre de cas favorables}}{\text{nombre total de cas}}.$$

Exemple 1 : Si on lance un dé à 6 faces, le référentiel est composé des six faces  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

Exemple 2 : Si on lance trois fois une pièce, le référentiel est composé des  $2^3$  arrangements avec répétition des 2 faces distinctes notées  $P$  et  $F$  :  $\Omega = \{PPP, PPF, PFP, PFF, FPP, FPF, FFP, FFF\}$ .

Exemple 3 : Si on lance trois pièces identiques simultanément, le référentiel est composé des 3-combinaisons avec répétition des 2 faces distinctes notées  $P$  et  $F$  :  $\Omega = \{PPP, PPF, FFP, FFF\}$ . de cardinal  $K_2^3$ .

Question : "On lance trois pièces de monnaie. Quelle est la probabilité que toutes trois retombent du même côté, que ce soit pile ou face?"

**Définition 1** Deux événements  $A$  et  $B$  sont dits incompatibles s'ils ne peuvent se réaliser simultanément c'est-à-dire lorsque l'intersection des sous-ensembles  $A$  et  $B$  est vide :  $A \cap B = \emptyset$ .

### 2.1.2 Axiomatique de Kolmogorov

A chaque événement, on associe un nombre positif compris entre 0 et 1, sa probabilité.

La théorie moderne des probabilités repose sur l'axiomatique suivante :

**Définition 2** On appelle probabilité sur  $(\Omega, \mathcal{C})$  (où  $\Omega$  est l'ensemble des événements et  $\mathcal{C}$  une classe de parties de  $\Omega$ ), ou loi de probabilité, une application  $P$  de  $\mathcal{C}$  dans  $[0, 1]$  telle que :

1. Pour tout événement  $E$ ,  $0 \leq P(E) \leq 1$ .
2.  $P(\Omega) = 1$
3. pour tout ensemble dénombrable d'événements incompatibles  $A_1, A_2, \dots, A_n$ , on a

$$P(\cup A_i) = \sum P(A_i). \quad (\sigma\text{-additivité de } P)$$

**Définition 3** On appelle espace probabilisé le triplé  $(\Omega, \mathcal{C}, P)$  où  $\Omega$  est l'ensemble fondamental,  $\mathcal{C}$  est une collection de sous-ensembles de  $\Omega$  (la collection des événements), qui possède la structure précédente de  $\sigma$ -algèbre de Boole et  $P : \mathcal{C} \rightarrow [0, 1]$  est une mesure de probabilité sur  $\mathcal{C}$ .

Propriétés élémentaires : de l'axiomatique de Kolmogorov, on peut déduire les propriétés suivantes :

1.  $P(\emptyset) = 0$
2.  $P(\bar{A}) = 1 - P(A)$
3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4.  $P(A) \leq P(B)$  si  $A \subset B$  (inégalité de Boole)
5.  $P(\cup_i A_i) \leq \sum_i P(A_i)$  (Il n'y a stricte égalité que si les événements  $A_i$  sont deux à deux incompatibles.)
6. Si la suite  $(A_n)$  croît vers  $A$  (c'est-à-dire  $\forall n, A_n \subset A_{n+1}$  et  $\cup A_n = A$ ) alors  $\lim P(A_n) = P(A)$ .
7. Continuité monotone séquentielle. Soient  $A_1 \supset A_2 \supset \dots \supset A_n \supset \emptyset$ . Si  $\lim_{n \rightarrow \infty} A_n = \emptyset$  alors  $\lim_{n \rightarrow \infty} P(A_n) = 0$

Démonstration :

1. Soit  $E$  un événement quelconque. Comme  $E \cup \emptyset = E$ ,  $P(E \cup \emptyset) = P(E)$ . D'autre part, on sait que  $E \cap \emptyset = \emptyset$  (tout événement est incompatible avec l'événement impossible) et d'après le 3ème axiome,  $P(E \cup \emptyset) = P(E) + P(\emptyset)$ . Des deux égalités, on obtient  $P(\emptyset) = 0$ .
2.  $A \cup \bar{A} = \Omega$  et  $A \cap \bar{A} = \emptyset$   $P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$  d'où  $P(\bar{A}) = 1 - P(A)$
3. On découpe selon une partition de  $A \cup B$  : on a  $P(A \cup B) = P((A \cap \bar{B}) \cup (B \cap \bar{A}) \cup (A \cap B))$ . Ces ensembles sont deux à deux incompatibles d'où  $P(A \cup B) = P(A \cap \bar{B}) + P(B \cap \bar{A}) + P(A \cap B)$ . De plus,  $P(A) = P(A \cap \bar{B}) + P(A \cap B)$  et  $P(B) = P(B \cap \bar{A}) + P(A \cap B)$ , d'où  $P(A \cap \bar{B}) = P(A) - P(A \cap B)$  et  $P(B \cap \bar{A}) = P(B) - P(A \cap B)$ , valeurs que l'on remplace dans la première égalité obtenue.
4. D'après la propriété précédente et la positivité de la probabilité, on a  $P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq P(A) + P(B)$ ,
5. Formule précédente que l'on peut généraliser à un nombre quelconque d'événements :  $P(\cup_i A_i) \leq \sum_i P(A_i)$  avec égalité si des événements sont deux à deux incompatibles.
6. On pose  $B_1 = A_1$  et  $\forall n \geq 2, B_n = A_n \setminus A_{n-1}$ . Les  $B_i$  sont disjoints et vérifient  $\cup_{n \geq 1} B_n = \cup_{n \geq 1} A_n$  et  $\forall n \geq 1, \cup_{k=1}^n B_k = A_n$ . Par la propriété de  $\sigma$ -additivité,  $\sum_{n \geq 1} P(B_n) = P(\cup_{n \geq 1} A_n)$  et  $\forall n \geq 1, \sum_{k=1}^n P(B_k) = P(A_n)$ . Ainsi  $\lim_{n \rightarrow \infty} P(A_n) = P(\cup_n A_n) = P(A)$ .
7. On note  $A = \cap A_n$ . Comme  $A_n \supset A_{n+1}$ ,  $\bar{A}_n \subset \bar{A}_{n+1}$ . On pose  $B_1 = \bar{A}_1$  et  $B_{n+1} = \bar{A}_{n+1} \setminus \bar{A}_n$ . Ainsi  $\cup_n B_n = \cup_n \bar{A}_n = \bar{A}$  et  $\cup_{k=1}^n B_k = \bar{A}_n$ . Ainsi  $\lim_{n \rightarrow \infty} P(\bar{A}_n) = \lim_n P(\cup B_k) = P(\bar{A})$ . En passage au complémentaire,  $\lim_{n \rightarrow \infty} P(A_n) = P(A)$ . On peut prendre  $A = \emptyset$ .

**Théorème 2.1.1 (Théorème des probabilités totales)** Soit  $\Omega = \cup B_i$  un système complet d'événements (i.e. tel que les  $\{B_i\}$  constituent une partition de  $\Omega$ ). Alors

$$\forall A : P(A) = \sum_i P(A \cap B_i).$$

Exemples de construction :

1. Si  $\Omega = \{x_1, \dots, x_n\}$  est fini, on définit une probabilité sur  $\mathcal{P}(\Omega)$  en se donnant  $n$  nombres  $p_i$  tels que  $\sum_i p_i = 1$  en posant  $P(\{x_i\}) = p_i$ . On parle d'équiprobabilité si pour tout  $i$ ,  $P(\{x_i\}) = \frac{1}{n}$ . Dans ce cas,  $P(A) = \frac{\text{card}(A)}{n}$ .
2. Si  $\Omega = \{x_n, n \in \mathbb{N}\}$  est dénombrable, on définit une probabilité sur  $\mathcal{P}(\Omega)$  en se donnant une série de terme général  $p_n$  convergente à termes positifs et de somme égale à 1 en posant  $P(\{x_i\}) = p_i$ . (par exemple :  $p_n = \frac{1}{2^{n+1}}$ )
3. Si  $\Omega$  est un intervalle  $]a, c[$ , où  $a, c$  appartiennent à  $\mathbb{R} \cup \{-\infty, \infty\}$ , on peut définir une probabilité sur tous les intervalles inclus dans  $\Omega$  en définissant une probabilité sur des intervalles de la forme  $]a, b[$ . Par exemple,  $P(]a, b[) = \int_a^b \varphi(t) dt$  où  $\varphi$  est une fonction positive qui vérifie  $\int_a^c \varphi(t) dt = 1$ .
4. Si  $\Omega = ]0, 1]^2$ , on obtient une probabilité en posant  $P(A) = \text{surface de } A$ .

## 2.2 Probabilité conditionnelle

On considère la réalisation de deux événements  $A$  et  $B$ .

Que peut-on déduire sur la probabilité de l'événement  $B$  sachant que l'événement  $A$  est réalisé? Cette probabilité est appelée probabilité conditionnelle de  $A$  sachant  $B$  et se note  $P(A/B)$  ou  $P_B(A)$ .

Par définition, on a

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

Cette probabilité a posteriori est dite Probabilité de Bayes.

Exercice : montrer que  $P_B$  est une probabilité sur  $\Omega$ .

### 2.2.1 Formule de Bayes

Comme  $P(B/A) = \frac{P(A \cap B)}{P(A)}$ , on a  $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B/A)P(A)}{P(B)}$ . La formule de Bayes est

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

On remarque que  $P(B) = P(A \cap B) + P(\bar{A} \cap B) = P(B/A)P(A) + P(B/\bar{A})P(\bar{A})$ , ainsi

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P(B/\bar{A})P(\bar{A})}.$$

Plus généralement si  $\{A_j\}$  est une partition de l'ensemble des possibles, pour tout  $i$ ,

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_j P(B/A_j)P(A_j)}.$$

Exemple : Soit un événement  $A$  qui peut dépendre de  $N$  causes  $C_i$  différentes et incompatibles deux à deux (on ne peut avoir deux causes réalisées simultanément). Etant donnée la réalisation de l'événement  $A$ , quelle est la probabilité que ce soit  $C_i$  qui en soit la cause?

### 2.2.2 Formule des probabilités composées

**Proposition 2.2.1 (Formule des probabilités composées)** Soient  $n$  événements  $A_1, \dots, A_n$  tels que  $P(A_1 \cap \dots \cap A_n) \neq 0$ . Alors :

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

Exemple : Une urne contient initialement 7 boules noires et 3 boules blanches. On tire successivement 3 boules : si on tire une noire, on l'enlève, si on tire une blanche, on la retire, et on ajoute une noire à la place. Quelle est la probabilité de tirer 3 blanches à la suite ?

On note  $B_i$  l'événement "La  $i$ -ème boule tirée est blanche". La probabilité recherchée est :

$$P(B_1 \cap B_2 \cap B_3) = P(B_3|B_1 \cap B_2) \cdot P(B_2|B_1) \cdot P(B_1).$$

Clairement,  $P(B_1) = 3/10$ . Maintenant, si  $B_1$  est réalisé, avant le 2ème tirage, l'urne est constituée de 8 boules noires et 2 blanches. On a donc :  $P(B_2|B_1) = 2/10$ . Si  $B_1$  et  $B_2$  sont réalisés, avant le 3ème tirage, l'urne est constituée de 9 boules noires et 1 blanche. On en déduit  $P(B_3|B_1 \cap B_2) = 1/10$ . Finalement  $P(B_1 \cap B_2 \cap B_3) = 6/1000 = 3/500$ .

**Proposition 2.2.2 (Formule des probabilités totales)** Soit  $\{A_n; n \in \mathbb{N}\}$  un système complet d'événements, tous de probabilité non nulle. Soit  $B$  un événement. Alors :

$$P(B) = \sum_{n \in \mathbb{N}} P(A_n)P(B|A_n).$$

Cette formule permet de calculer la probabilité d'un événement  $B$  en le décomposant suivant un système complet d'événements (En effet,  $B$  est égal à la réunion disjointes des  $B \cap A_n$ ).

### 2.2.3 Événements indépendants

Soient 2 événements  $A$  et  $B$ . Ils sont indépendants si la réalisation de  $A$  n'affecte pas la réalisation de  $B$ , et inversement. On peut alors écrire

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B) \end{aligned}$$

On dit encore que  $A$  et  $B$  sont indépendants si et seulement la probabilité de réalisation simultanée de ces événements est égal au produit de leurs probabilités individuelles :

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

Si deux événements  $A$  et  $B$  sont indépendants, alors il en est de même pour  $A$  et  $B^c$ ,  $A^c$  et  $B$ ,  $A^c$  et  $B^c$ .

**Définition 4** Un ensemble d'événements  $A_1, A_2, \dots, A_n$  est dit totalement indépendant si pour tout sous-ensemble  $I \subset \{1, 2, \dots, n\}$

$$P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i).$$

Les événements sont deux à deux indépendants si pour tous indices  $i, j$  ( $i \neq j$ ),

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

Que pensez-vous des deux définitions ? Sont-elles équivalentes ?

Exemple : On jette deux dés équilibrés. Soient les événements

$$A = \{\text{la somme des dés vaut } 7\},$$

$$B = \{\text{le premier dé affiche } 4\}$$

$$C = \{\text{le second dé affiche } 3\}.$$

Calculer  $P(A \cap B)$ ,  $P(A \cap C)$ ,  $P(B \cap C)$  et  $P(A|B \cap C)$ .



# Chapitre 3

## Variables aléatoires

### 3.1 Définition d'une variable aléatoire

Une variable aléatoire est une fonction définie sur l'ensemble des éventualités, c'est-à-dire l'ensemble des résultats possibles d'une expérience aléatoire.

On s'intéressera aux valeurs prises  $x_i$  par une variable aléatoire  $X$ , événement noté  $(X = x_i)$ , et plus particulièrement à la probabilité d'obtenir ces valeurs  $P(X = x_i)$ .

C'est à peu près la même chose qu'une variable statistique sauf que dans le cas d'une variable statistique on évalue un comportement réalisé (moyenne, etc) alors que dans le cas de variables aléatoires on suppose un comportement futur (Dans ce cas, on parle d'espérance plutôt que de moyenne par exemple) ou théorique. Les variables aléatoires sont utilisées pour modéliser le résultat d'un mécanisme non-déterministe.

#### 3.1.1 Différents types de variables aléatoires

**Définition 5** Une variable aléatoire (ou v.a.) est une application  $X : \Omega \rightarrow \mathbb{R}$ . Si  $X(\Omega)$  est au plus dénombrable, on dit que  $X$  est un v.a. discrète sinon on dit qu'elle est continue.

#### Variable aléatoire discrète

Si une variable aléatoire  $X$  prend un nombre de valeurs fini ou dénombrable (son ensemble de définition est inclus dans  $\mathbb{N}$ ), on parle de variable discrète.

On s'intéresse à définir l'ensemble des valeurs possibles et leurs probabilités associées.

Quelques exemples :

- nombre de "face" dans un lancer de 3 pièces :  $X(\omega)$  de 0 à 3.
- nombre de lancers avant d'obtenir "6" avec un dé :  $X(\omega)$  de 0 à l'infini ;
- nombre de clients attendant au service après-vente :  $X(\omega)$  de 0 à 10.
- nombre de cycles lecture/écriture sur une clé USB :  $X(\omega)$  de 10000 à 100000.
- nombre d'appels arrivant à un standard téléphonique en une minute de 0 à 10.

### Variable aléatoire continue

Une variable aléatoire est dite continue si elle peut prendre toutes les valeurs d'un intervalle. En particulier, dans le cas où la variable aléatoire peut prendre toute valeur réelle (son ensemble de définition contient un intervalle de  $\mathbb{R}$ ), on parle de variable aléatoire réelle.

Dans ce cas, il ne s'agira plus de calculer une probabilité d'apparition d'une valeur donnée mais d'un intervalle.

Quelques exemples :

- temps d'attente pour avoir le bus :  $X(\omega) \in [0, 10']$
- longueur de cheveux :  $X(\omega) \in [0, 4m]$
- intervalle entre deux averses :  $X(\omega) \in [1', 20 \text{ ans}]$
- moyenne des tailles de 20 étudiants pris au hasard :  $X(\omega) \in [\alpha, \beta]$

### 3.1.2 Loi de probabilité

Une variable aléatoire est totalement définie par sa loi de probabilité. Cette dernière est caractérisée par :

- l'ensemble des valeurs qu'elle peut prendre (son domaine de définition) ;
- les probabilités attribuées à chacune des valeurs potentiellement prises  $P(X = x)$ .

Dans ce cas, la loi de la variable aléatoire est la loi de probabilité sur l'ensemble des valeurs possibles de  $X$  qui affecte la probabilité  $P(X = x_k)$  au singleton  $\{x_k\}$ .

Soit  $X : \Omega \rightarrow \mathbb{R}$ . Dans le cas où  $X$  prend ses valeurs dans un intervalle réel, on cherche à exprimer par exemple la probabilité que  $X$  prenne ses valeurs dans  $] \alpha, \beta [$ . On remarque que  $X(\omega) \in ] \alpha, \beta [ \iff \omega \in X^{-1}(] \alpha, \beta [)$  et donc on pose  $P(X \in ] \alpha, \beta [) = P(X^{-1}(] \alpha, \beta [))$ .

### 3.1.3 Fonction de répartition

**Définition 6** La fonction de répartition d'une v.a.  $X$  est l'application  $F$  de  $\mathbb{R}$  dans  $[0, 1]$  définie par

$$F(x) = P(X < x) = P(X^{-1}(] - \infty, x [)) .$$

On s'intéresse souvent à la probabilité cumulée. Par exemple dans le cas de probabilités sur  $\mathbb{N}$  :

$$P(X < n) = P(X = 0 \text{ ou } X = 1 \text{ ou } \dots \text{ ou } X = n - 1).$$

Les événements étant incompatibles entre eux, on obtient

$$P(X < n) = \sum_{j=0}^{n-1} P(X = j),$$

et de façon plus générale, avec  $x \in \mathbb{R}$ ,

$$P(X < x) = \sum_{j=0}^{\lceil x-1 \rceil} P(X = j).$$

Ainsi pour  $X$  une v.a. discrète prenant les valeurs classées  $x_i$  avec les probabilités  $p_i$ , on a

$$F(x) = \sum_{x_i < x} p_i.$$

(Dans ce cas,  $F$  est une fonction en escaliers, continue à gauche, ayant pour limite 0 en  $-\infty$  et 1 en  $+\infty$ ).

## Propriétés

1.  $F$  est non décroissante.
2.  $F$  est continue à gauche.
3.  $\forall x_0 \in \mathbb{R}, P(X = x_0) = \lim_{x \rightarrow x_0^+} F(x) - F(x_0)$ .
4.  $F(-\infty) = 0$  et  $F(+\infty) = 1$
5.  $P(a \leq X < b) = F(b) - F(a)$

## Preuves :

1. Montrons que  $F$  est croissante ( $x < y \implies F(x) \leq F(y)$ ). On a la réunion disjointe  $] - \infty, y[ = ] - \infty, x[ \cup [x, y[$  et  $F(y) = P(X^{-1}(] - \infty, y]) = P(X^{-1}(] - \infty, x]) \cup X^{-1}([x, y])$  et en utilisant les propriétés de  $P$ ,  $F(y) = F(x) + P(X \in [x, y]) \geq F(x)$ .
2. Soit  $x_0 \in \mathbb{R}$ .  $X^{-1}([x_0 - 1/n, x_0])$  décroît vers  $\emptyset$  quand  $n$  tend vers l'infini, donc  $F(x_0) - F(x_0 - 1/n)$  tend vers 0. Comme  $F$  est croissante, cela implique que  $F$  est continue à gauche.
3. De même,  $X^{-1}([x_0, x_0 + 1/n])$  décroît vers  $X^{-1}(\{x_0\})$  donc la différence  $F(x_0 + 1/n) - F(x_0)$  tend vers  $P(X^{-1}\{x_0\})$  quand  $n$  tend vers l'infini.
4.  $F$  étant croissante,  $F(-\infty) = \lim_{n \rightarrow \infty} F(-n)$ . Or  $] - \infty, -n[$  décroît vers  $\emptyset$  quand  $n$  tend vers l'infini ; ainsi  $F(-n) = P(X^{-1}(] - \infty, -n])$  décroît vers 0. De même,  $] - \infty, n[$  croît vers  $\mathbb{R}$  quand  $n$  tend vers l'infini et  $F(n) = P(X^{-1}(] - \infty, n])$  croît vers  $P(X \in \mathbb{R}) = 1$ .
5.  $X^{-1}(] - \infty, b])$  est la réunion disjointe de  $X^{-1}(] - \infty, a])$  et de  $X^{-1}([a, b])$  donc  $F(b) = P(X \in [a, b]) + F(a)$ .

Remarque :  $F$  est continue à droite dans le cas des v.a. continues.

Preuve : D'après la propriété 4 des fonction de répartition,  $F$  est continue si et seulement si :  $\forall x \in \mathbb{R}, P(X = x) = 0$ .

Remarque : Probabilité ponctuelle pour une variable continue.

La vraie distinction entre variables continues et discrètes tient dans le calcul de la probabilité ponctuelle. La probabilité d'un point  $c$  situé entre  $a$  et  $b$  serait  $\lim_{b \rightarrow a} P(a < X < b) = 0$ . Ainsi, la probabilité d'un point est par définition nulle pour les variables continues. Ainsi  $P(a < X < b) = P(a \leq X \leq b)$ .

En réalité, il s'agit bien souvent d'un problème d'échelle ou de précision de mesure. On travaille toujours sur des intervalles (par exemple la précision de la mesure si on a une mesure au centimètre près la valeur  $x = 160$  correspond à  $x = 160 \pm 0.5$  soit un intervalle de travail  $160 - 0.5 < x < 160 + 0.5$ ).

### 3.1.4 Densité de probabilité

Pour une variable continue, on travaille la plupart du temps avec un ensemble de définition sur les réels. La probabilité ponctuelle  $P(X = x) = f(x)$  est la fonction de densité. La fonction de répartition  $F(x) = P(X < x)$  est définie par :

$$F(x) = \int_{-\infty}^x f(t) dt$$

La densité de probabilité d'une variable aléatoire continue est la dérivée première par rapport à  $x$  de la fonction de répartition. Cette dérivée prend le nom de fonction de densité, notée  $f(x) = \frac{dF(x)}{dx}$ . Elle est équivalente à  $P(X = x)$  dans le cas des variables discrètes.

Pour calculer la probabilité  $P(a \leq X < b)$  dans le cas d'une variable continue, le calcul est le suivant  $\int_a^b f(x) dx$ . Développons cette expression :

$$\begin{aligned} P(a \leq X < b) &= P(X < b) - P(X < a) \\ &= \int_{-\infty}^b f(t) dt - \int_{-\infty}^a f(t) dt \\ &= F(b) - F(a) \end{aligned}$$

Graphiquement, cela se traduit par la surface comprise entre  $a$  et  $b$  en dessous de la courbe de densité.

Propriété : De la même façon que  $p_i \geq 0$  et  $\sum_i p_i = 1$ , on a  $f(x) \geq 0$  et  $\int_{-\infty}^{+\infty} f(x)dx = 1$ .

## 3.2 Caractéristiques d'une variable aléatoire

### 3.2.1 Tendances centrale

#### Les fractiles

On appelle quantile ou fractile d'ordre  $\alpha$  ( $0 < \alpha < 1$ ) d'une variable aléatoire  $X$  dont la fonction de répartition est  $F(x)$ , la valeur  $x_\alpha$  telle que  $F(x_\alpha) = \alpha$ . La valeur  $x_\alpha$  s'appelle quantile d'ordre  $\alpha$ .

Remarque Dans le cas où  $X$  est une variable discrète,  $F(x_\alpha) = \alpha$  s'entend  $P(X < x_\alpha) = \alpha$ .

Nous énumérons ici quelques quantiles particuliers.

**La médiane :** La médiane est le quantile d'ordre  $\alpha = 1/2$ , en d'autres termes la médiane  $Med$  est définie par  $F(Med) = 0,5$ . La médiane partage la population en deux parties égales, c'est une caractéristique de tendance centrale.

**Les quartiles :** les quartiles, notés  $Q_i$  (respectivement  $i = 1; 2; 3$ ) correspondent aux quantiles d'ordre ( $\alpha = 0,25; 0,5; 0,75$ ). Notons que  $Q_2 = Med$ .

**Les déciles** Le  $k$ -ième décile ( $k = 1$  à  $9$ ) est le quantile d'ordre  $k/10$ . En particulier, le 5-ième décile correspond à la médiane.

#### Le mode

On appelle mode (valeur dominante, valeur la plus probable) d'une variable aléatoire, la valeur  $Mode$  pour laquelle l'histogramme de fréquence présente son maximum.

Dans le cas des variables discrètes, le Mode est la valeur de  $X$  associée à la plus grande probabilité, d'où l'appellation valeur la plus probable.

Lorsque la variable aléatoire  $X$  est continue, avec une fonction de densité pourvue d'une dérivée première et d'une dérivée seconde, le mode  $M$  est un maximum de la fonction densité et satisfait à ce titre à  $f'(M) = 0$  et  $f''(M) < 0$  (concavité vers le bas).

#### Espérance mathématique

Soit  $X$  une v.a. discrète prenant ses valeurs dans  $\{x_1, \dots, x_n\}$  et dont les probabilités associées sont  $P(X = x_i) = p_i$ . On définit l'espérance mathématique de  $X$ , notée  $E(X)$  par

$$E(X) = \sum_{i=1}^n p_i x_i.$$

Cette quantité n'est définie que si la série de terme général  $[p_i x_i]$  converge.

C'est la moyenne théorique de  $X$ . Cette moyenne est à rapprocher de la moyenne expérimentale où chaque événement  $X = x_i$  se réalise  $n_i$  fois dans un échantillon de taille  $N = \sum_i n_i$ . La moyenne expérimentale vaut  $\bar{X} = \frac{1}{N} \sum_i n_i x_i = \sum_i f_i x_i$ , où  $f_i = \frac{n_i}{N}$  est la fréquence observée dans chaque classe d'événement  $X = x_i$ .

Dans le cas continu,  $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$  où  $f(x)$  est la densité de probabilité de  $X$ . Cette quantité n'existe que si  $\int_{-\infty}^{+\infty} x f(x) dx$  est absolument convergente.

De façon générale, pour  $Y$  une fonction de  $X$ , on a

$$E(Y) = \sum_{i=1}^n p_i y_i \quad \text{ou} \quad E(Y) = \int_{-\infty}^{+\infty} y(x) f(x) dx$$

Par exemple, pour  $Y = X^2$ , on a  $E(X^2) = \sum_{i=1}^n p_i x_i^2$  ou  $E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx$ .

Propriétés : (Espérance d'une constante).  $E(a) = a$ . Preuve.

$$E(a) = \int_D a f(x) dx = a \int_D f(x) dx = a \times 1 = a.$$

On peut en déduire que  $E[E(X)] = E(X)$ , puisque  $E(X)$  n'est pas une variable aléatoire.

Caractéristique (opérateur linéaire).  $E(aX + bY) = aE(X) + bE(Y)$ . De manière générale,

$$E\left(\sum_i a_i X_i + b\right) = \left(\sum_i a_i E(X_i)\right) + b$$

### 3.2.2 Paramètres de dispersion

#### Moments

- Un moment non-centré d'ordre  $r$  est défini de la manière suivante :

$$m_r(X) = E(X^r)$$

Application : pour une v.a. (variable aléatoire) discrète,  $m_r(X) = \sum_i p_i x_i^r$ , où  $p_i = P(X = x_i)$  et pour une v.a. continue,  $m_r(X) = \int_D x^r f(x) dx$ .

Remarque (Moments non-centrés empiriques, statistique descriptive). Rappelons qu'en statistique descriptive, ce moment non-centré, pour une v.a. discrète par exemple, est obtenue avec la formule  $m_r = \sum_i f_i x_i^r$ , où  $f_i$  est la fréquence observée de la valeur  $x_i$ .

Remarque (Cas particuliers).

( $r = 0$ )  $m_0(X) = 1$  ;

( $r = 1$ )  $m_1(X) = E(X)$ , espérance mathématique ;

- Un moment centré d'ordre  $r$  est défini de la manière suivante :

$$\mu_r(X) = E[(X - E(X))^r]$$

Soit pour une v.a. discrète :  $\mu_r(X) = \sum_i (x_i - E(X))^r p_i$

et pour une v.a. continue :  $\mu_r(X) = \int_D (x - E(X))^r f(x) dx$

Remarque (Statistique descriptive). En statistique descriptive, pour une variable discrète, le moment centré d'ordre  $r$  est obtenu avec  $\mu_r = \sum_{i=0}^n f_i (x_i - \bar{x})^r$

Remarque (Cas particuliers).

( $r = 0$ )  $\mu_0(X) = E[(X - E(X))^0] = E(1) = 1$

( $r = 1$ )  $\mu_1(X) = E[(X - E(X))] = E(X) - E[E(X)] = E(X) - E(X) = 0$

( $r = 2$ )  $\mu_2(X) = E[(X - E(X))^2] = V(X)$  (c'est la variance de  $X$ ).

#### Variance

**Définition 7** On appelle variance de  $X$ , noté  $V(X)$ , le moment centré d'ordre 2 de  $X$  (s'il existe).

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

La variance mesure la dispersion autour de la moyenne. L'écart-type est la racine carrée de la variance :

$$\sigma(X) = \sqrt{V(X)}.$$

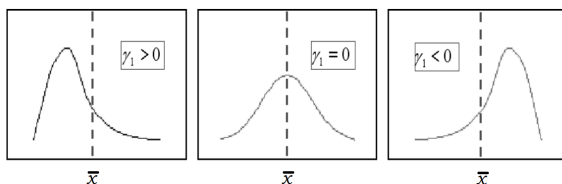
Remarque (Formule de Koenig). Il est possible d'exprimer la variance à partir des moments non centrés.

$$\begin{aligned} V(X) &= E[(X - E(X))^2] = E[X^2 - 2XE(X) + E(X)^2] \\ &= E(X^2) + E(X)^2 - 2E[XE(X)] \\ &= E(X^2) + E(X)^2 - 2E(X)E(X) \\ &= E(X^2) - E(X)^2 \\ V(X) &= m_2(X) - m_1(X)^2 \end{aligned}$$

### 3.2.3 Caractéristiques de forme

#### Le coefficient d'asymétrie (skewness)

Graphiquement, il s'agit de l'étalement à gauche ou à droite de l'histogramme des fréquences de la variable statistique.



**Le coefficient d'asymétrie de Fisher :** Outil banal de la statistique descriptive, il s'agit du moment centré d'ordre 3 normalisé par le cube de l'écart-type, c'est-à-dire (« gamma un ») :

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{1}{\sigma^3} \sum_i f_i (x_i - \bar{x})^3.$$

Remarque :  $\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}}$ .

Comme c'est un nombre sans dimension, il permet de comparer des distributions même si leurs échelles diffèrent. Lorsque l'étalement est à gauche (moyenne inférieure à la médiane), le coefficient d'asymétrie est négatif et vice versa.

**Le coefficient d'asymétrie de Pearson  $\beta_1$  (« beta un »)** est le carré du coefficient de Fisher, soit :

$$\beta_1 = \left( \frac{\mu_3}{\sigma^3} \right)^2.$$

**Le coefficient d'asymétrie de Yule et Kendall** On a juste besoin des quartiles pour le calculer.

$$u = \frac{(Q_3 - Q_2) - (Q_3 - Q_1)}{(Q_3 - Q_2) + (Q_3 - Q_1)}.$$

Comme il n'existe pas de table, donc pas de critère précis de séparation entre symétrie et asymétrie, on utilisera plutôt ce coefficient comme élément de comparaison entre deux distributions.

### Le coefficient d'aplatissement (kurtosis)

Généralement, on observe le coefficient d'aplatissement (kurtosis) en même temps que celui d'asymétrie.

Le coefficient d'aplatissement de Pearson  $\beta_2$  est la valeur obtenue par le calcul suivant :

$$\beta_2 = \frac{\mu_4}{\sigma^4}.$$

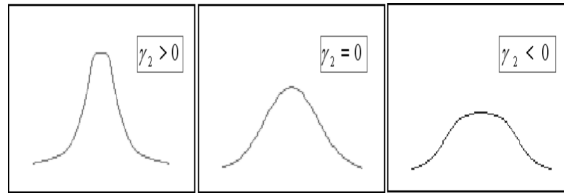
Un coefficient d'aplatissement élevé indique que la distribution est plutôt pointue en sa moyenne, et des queues de distribution épaisses.

Pour une variable aléatoire suivant une loi normale centrée réduite (loi vue par la suite), ce coefficient d'aplatissement vaut 3. C'est pour cela que l'on normalise la valeur pour mesurer l'excès d'aplatissement pour obtenir le coefficient d'aplatissement de Fisher.

Le coefficient d'aplatissement de Fisher  $\gamma_2$  est la valeur obtenue par le calcul suivant :

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3.$$

Remarque :  $\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$ .



### 3.2.4 Inégalité de Bienaymé-Chebyshev

**Proposition 3.2.1 (Inégalité de Bienaymé-Chebyshev)** Soit  $X$  une v.a. d'espérance mathématique  $\mu$  et de variance  $\sigma^2$ . L'inégalité de Bienaymé-Chebyshev indique que pour tout nombre réel positif  $t$ , la probabilité que  $X$  s'écarte de son espérance mathématique d'une grandeur supérieure à  $t$ , a pour limite supérieure  $\sigma^2/t^2$  :

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

Preuve :

1. Dans le cas où  $X$  est discrète : On a  $\sigma^2 = \sum_i p_i (x_i - \mu)^2$ . En enlevant les termes, positifs, correspondant aux  $x_i$  pour lesquels  $|x_i - \mu| < t$ , on obtient  $\sigma^2 \geq \sum_{|x_i - \mu| \geq t} p_i (x_i - \mu)^2$ . Dans cette somme, les  $x_i$  vérifient tous  $|x_i - \mu|^2 \geq t^2$  donc en remplaçant  $\sigma^2 \geq t^2 \sum_{|x_i - \mu| \geq t} p_i = t^2 P(|X - \mu| \geq t)$ .
2. Dans le cas où  $X$  est continue :

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \geq \int_{-\infty}^{\mu-t} (x - \mu)^2 f(x) dx + \int_{\mu+t}^{+\infty} (x - \mu)^2 f(x) dx$$

car pour tout  $x$ ,  $(x - \mu)^2 f(x)$  est positive ou nulle. Dans les intervalles considérés  $x \notin [\mu - t, \mu + t]$ , on a  $(x - \mu)^2 \geq t^2$  donc

$$\sigma^2 \geq t^2 \left( \int_{-\infty}^{\mu-t} f(x) dx + \int_{\mu+t}^{+\infty} f(x) dx \right) = t^2 (1 - F(\mu+t) + F(\mu-t)) = t^2 P(|X - \mu| \geq t)$$

Remarque (Loi des grands nombres). Cette inégalité est importante car elle permet de démontrer la loi dite des "grands nombres". Elle stipule qu'il suffit d'extraire un échantillon d'un effectif suffisant dans une population pour estimer l'espérance mathématique à l'aide de la moyenne arithmétique, ou encore pour estimer une probabilité à l'aide d'une fréquence. On parle alors de convergence en probabilité.

Remarque 2 (Une expression plus générale de l'inégalité de Bienaymé-Chebychev). La portée de l'inégalité de Bienaymé-Chebychev est bien plus large que celle couramment présentée dans les ouvrages. Elle peut être définie pour une fonction  $g(X)$  telle que

$$P(g(X) \geq t) \leq \frac{E(g(X)^k)}{t^k}$$

La présentation usuelle correspond à  $k = 2$  et  $g(X) = |X - \mu|$ .

### 3.2.5 Fonctions génératrices

#### Fonction caractéristique

La fonction caractéristique d'une variable aléatoire  $X$  est définie sur  $\mathbb{R}$  par

$$\varphi_X(t) = E(e^{itX}),$$

où  $i$  est l'unité imaginaire ( $i^2 = -1$ ). Ainsi, pour une variable discrète,  $\varphi_X(t) = \sum_k e^{itx_k} p_k$  et pour une variable continue,  $\varphi_X(t) = \int e^{itx} f(x) dx$ .

Propriétés de la fonction caractéristique :

1.  $\varphi_X(t)$  est bien définie pour tout  $t$  réel.
2. La relation suivante sert, par exemple, à calculer la fonction caractéristique d'une variable centrée réduite, à partir de la fonction caractéristique de la variable de départ : pour tous  $a, b$  réels,

$$\varphi_{aX+b}(t) = \varphi_X(at)e^{itb}.$$

3. Il y a aussi une relation entre les moments et la fonction caractéristique d'une variable aléatoire. Lorsque les moments existent et que la série converge :  $\varphi_X(t) = \sum_{k=0}^{\infty} \frac{i^k \mu_k}{k!} t^k$  où  $\mu_k$  est le moment d'ordre  $k$ . Cette relation sert parfois pour calculer la moyenne (premier moment) et la variance d'une variable aléatoire. Plus explicitement,  $1 = \varphi_X(0)$ ,  $E(X) = -i\varphi'_X(0)$ ,  $E(X^2) = -\varphi''_X(0)$  et  $\text{Var}(X) = -\varphi''_X(0) + \varphi_X''(0)$  ou encore  $\varphi_X^{(k)}(0) = i^k E(X^k)$ .
4. Elle détermine de façon unique la loi d'une variable aléatoire au sens où  $\varphi_X = \varphi_Y$  (égalité de fonctions) équivaut à " $X$  et  $Y$  ont la même loi."

**Preuve** 1. Dans le cas discret,  $\varphi_X(t) = \sum_k e^{itx_k} p_k$ . C'est la somme d'une série absolument convergente car  $|e^{itx_k} p_k| = p_k$ . Dans le cas continu,  $\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx$ . C'est une intégrale définie absolument convergente car  $|e^{itx} f(x)| = f(x)$  dont l'intégrale sur  $\mathbb{R}$  est égale à 1.

2. Utiliser les propriétés de la fonction exponentielle et de l'espérance.

3. On a  $\varphi_X(0) = E(1) = 1$ . En supposant les bonnes conditions de convergence, on a

$$\varphi_X(t) = \sum_n e^{itx_n} p_n = \sum_n \sum_{k=0}^{\infty} \frac{(itx_n)^k}{k!} p_n = \sum_{k=0}^{\infty} \frac{i^k}{k!} \left( \sum_n p_n x_n^k \right) t^k = \sum_{k=0}^{\infty} \frac{i^k}{k!} (m_k) t^k$$

De même dans le cas continu :

$$E(e^{itX}) = \int_{-\infty}^{+\infty} \left[ \sum_{k=0}^{\infty} \frac{(itx)^k}{k!} \right] f(x) dx = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \int_{-\infty}^{+\infty} x^k f(x) dx = \sum_{k=0}^{\infty} m_k \frac{(it)^k}{k!}.$$

Par identification avec le développement de  $\varphi_X(t)$  en série de Taylor-Mac-Laurin, on a bien la propriété proposée.

4. (admis)

**Définition 8** *On appelle fonction génératrice des moment de la v.a.  $X$ , si elle existe, la fonction :*

$$M_X(t) = E(e^{tX}).$$



# Chapitre 4

## Lois discrètes usuelles

### 4.1 Loi uniforme discrète

L'ensemble des valeurs possibles est  $\{1, 2, 3, \dots, n\}$ ,  $n$  étant un paramètre de la loi. Chaque valeur reçoit la même probabilité  $1/n$  (Uniformité). On obtient la loi de probabilité suivante :

|       |       |       |       |         |       |
|-------|-------|-------|-------|---------|-------|
| $x_i$ | 1     | 2     | 3     | $\dots$ | $n$   |
| $p_i$ | $1/n$ | $1/n$ | $1/n$ | $\dots$ | $1/n$ |

Pour une v.a.  $X$  qui suit cette loi, on a :

$$\begin{aligned} E(X) &= \frac{n+1}{2} \\ V(X) &= (n^2 - 1)/12 \end{aligned}$$

En effet,

$$E(X) = \sum_{i=1}^n p_i x_i = \frac{1}{n} \sum_{k=1}^n k = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

et

$$\begin{aligned} V(X) &= \sum_i p_i x_i^2 - (E(X))^2 = \frac{1}{n} \sum_{k=1}^n k^2 - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{1}{n} \cdot \frac{n}{6} (n+1)(2n+1) - \frac{(n+1)^2}{4} \\ &= \frac{4(n+1)(2n+1) - 6(n+1)(n+1)}{24} \\ &= \frac{(n+1)(8n+4-6n-6)}{24} = \frac{(n+1)(n-1)}{12} \\ V(X) &= \frac{n^2 - 1}{12} \end{aligned}$$

On peut également déterminer sa fonction caractéristique :

$$\begin{aligned}
 \varphi_X(u) &= \sum_k e^{iuk} \cdot \frac{1}{n} \\
 &= \frac{1}{n} \sum_k (e^{iu})^k \\
 &= \frac{1}{n} \cdot e^{iu} \frac{1 - e^{iun}}{1 - e^{iu}} \quad (\text{somme des premiers termes d'une suite géométrique})
 \end{aligned}$$

Or

$$\begin{aligned}
 e^{iu} - 1 &= e^{iu/2+iu/2} - e^{iu/2-iu/2} = e^{iu/2} \left( \underbrace{e^{iu/2} - e^{-iu/2}}_{2i \sin(u/2)} \right) \\
 e^{iun} - 1 &= e^{inu/2+inu/2} - e^{inu/2-inu/2} = e^{inu/2} 2i \sin(nu/2)
 \end{aligned}$$

D'où

$$\varphi_X(u) = \frac{1}{n} \frac{e^{iun/2} \sin(nu/2)}{e^{iu/2} \sin(u/2)} e^{iu} = \frac{e^{iu \frac{n+1}{2}}}{n} \cdot \frac{\sin(nu/2)}{\sin(u/2)}$$

## 4.2 Loi de Bernoulli

C'est une des lois les plus simples. Elle prend que deux valeurs possibles Vrai/Faux, codées 1 et 0. On note  $p$  la probabilité associée à la valeur 1 (ce sera le paramètre de la loi). Evidemment la probabilité associée à la valeur 0 sera  $1 - p$  (parfois notée  $q$  pour plus de lisibilité dans les formules). On notera cette loi  $\mathcal{B}(p)$ .

Caractéristiques :  $E(X) = p$ ,  $V(X) = pq$ ,  $\varphi_X(u) = \sum_{k=0}^1 P(X = k) e^{iuk} = P(X = 0) + P(X = 1) e^{iu} = q + p e^{iu}$ .

## 4.3 Loi binomiale

Considérons une épreuve aléatoire qui ne conduit qu'à deux éventualités exclusives : l'une succès ( $V$ ) et l'autre échec  $F$ . L'univers associé à cette épreuve est donc  $\Omega = \{V; F\}$ .

Soient  $p$  la probabilité de l'événement  $\{V\}$  et  $q$  la probabilité de l'événement  $\{F\}$  (on a  $q = 1 - p$ ).

L'expérience consistant à répéter  $n$  fois cette épreuve **de façon indépendante**, est appelée suite d'épreuves de Bernoulli, ou schéma de Bernoulli.

On s'intéresse au nombre  $X$  de succès obtenus au cours de cette suite d'épreuves, la probabilité de l'événement : " on obtient dans un ordre quelconque  $k$  succès et  $n - k$  échecs " est égal à

$$P(X = k) = C_n^k p^k q^{n-k} \quad \text{avec } k \in [0, \dots, n].$$

En effet, notons  $A_k$  l'événement "A se réalise exactement  $k$  fois durant les  $n$  expériences".  $A_k$  peut se réaliser de plusieurs manières chacune deux à deux incompatibles, par exemple  $A$  peut se réaliser durant les  $k$  premières expériences aléatoires et ne pas se réaliser durant les  $n - k$  dernières expériences aléatoires. Il y a  $C_n^k$  façons de "placer" les  $k$  réalisations de  $A$  parmi les  $n$  expériences aléatoires. La probabilité d'une de ces façons est égale à  $p^k (1 - p)^{n-k}$ . Ce qui donne :  $P(A_k) = C_n^k p^k q^{n-k}$ .

**Définition 9 (Loi binomiale)** On dit qu'une variable aléatoire  $X$ , à valeurs dans  $\{0; 1; \dots; n\}$ , suit une loi binomiale si sa loi de probabilité est  $P(X = k) = C_n^k p^k q^{n-k}$  avec  $k \in \{0, 1, 2, \dots, n\}$ , où  $n$  est un entier naturel fixé et où  $p$  est un réel de  $]0; 1[$ . Les valeurs  $n$  et  $p$  sont les deux paramètres de cette loi, que l'on notera  $\mathcal{B}(n, p)$

Caractéristiques :

- L'espérance mathématique d'une variable aléatoire suivant une loi  $\mathcal{B}(n, p)$  est :  $E(X) = np$
- La variance mathématique d'une variable aléatoire suivant une loi  $\mathcal{B}(n, p)$  est :  $V(X) = npq = np(1-p)$
- Sa fonction caractéristique est  $\varphi_X(u) = (q + pe^{iu})^n$ .

Exercices : Retrouver que  $\sum_k P(X = k) = 1$ ,  $E(X)$  et  $V(X)$  par calcul direct et en utilisant la fonction caractéristique.

**Théorème 4.3.1 (Stabilité de la loi binomiale)** Si  $X_n$  et  $X_m$  sont deux variables indépendantes suivant des lois binomiales respectivement  $X_n \hookrightarrow \mathcal{B}(n, p)$  et  $X_m \hookrightarrow \mathcal{B}(m, p)$  alors  $X_n + X_m \hookrightarrow \mathcal{B}(n + m, p)$ .

Exemple : On dispose d'une urne avec 1 boule blanche et 9 noires. On effectue 20 tirages avec remise. Soit  $X$  le nombre de sortie de la boule blanche à l'issue des 20 tirages. La variable aléatoire  $X$  suit  $\mathcal{B}(n, p)$ .

## 4.4 Loi hypergéométrique

Considérons une population d'effectif  $N$  dont on sait qu'un pourcentage  $p$  d'éléments possèdent un caractère étudié  $C$ . On extrait au hasard un échantillon de  $n$  éléments, tirage exhaustif de  $n$  éléments (c'est-à-dire  $n$  tirages sans remise). Quelle est alors la probabilité qu'exactly  $k$  d'entre eux possèdent le caractère  $C$  ?

Si  $m$  désigne le nombre d'éléments possédant le caractère  $C$ , alors  $p = m/N$  et on peut reformuler le problème en remplaçant la connaissance de  $p$  par celle de  $m$  et considérer le problème en termes de tirages aléatoires de boules dans une urne : il s'agit d'un tirage simultané de  $n$  objets parmi  $N$  (équivalent à  $n$  tirages sans remise) et on s'intéresse à la variable aléatoire  $X$  égale au nombre  $k$  ( $k \leq m$ ) d'apparitions d'éléments ayant le caractère étudié sachant que leur effectif dans la population est  $m$ .

Loi de probabilité : Parmi les  $n$  objet tirés,  $k$  sont souhaités et  $n - k$  ne le sont pas. Il y a  $C_m^k$  façons de constituer des lots de  $k$  objets parmi les  $m$  présentant le caractère étudié et  $C_{N-m}^{n-k}$  façons de choisir les autres. Le nombre de cas possibles est  $C_N^n$ . Finalement, la loi de probabilité est fournie par la formule :

$$P(X = k) = \frac{C_m^k \cdot C_{N-m}^{n-k}}{C_N^n} = \frac{C_{Np}^k \cdot C_{N(1-p)}^{n-k}}{C_N^n},$$

avec  $0 \leq k \leq [Np]$ .

On note  $\mathcal{H}(N, n, p)$  la loi hypergéométrique de paramètre  $N$ ,  $n$  et  $p$ .

Caractéristiques : On peut montrer que l'espérance mathématique de  $X \hookrightarrow \mathcal{H}(N, n, p)$  est  $E(X) = np$  (comme dans le cas de la loi binomiale). Sa variance est  $V(X) = \frac{N-n}{N-1} npq$ . Sa fonction caractéristique est compliquée.

Convergence : On remarque que si  $n$  (la taille de l'échantillon) est petit devant  $N$ , alors la variance est sensiblement  $npq$ , c'est-à-dire celle de la loi binomiale. Ce résultat n'est pas un hasard... :

La limite pour  $N$  infini de sorte que  $m/N$  tende vers une limite finie  $p$  de la loi  $\mathcal{H}(N, n, p)$  est la loi binomiale  $\mathcal{B}(n, p)$ .

## 4.5 Loi géométrique

La loi géométrique est la loi du nombre d'essais nécessaires pour faire apparaître un événement de probabilité  $p$ .

On dit qu'une variable aléatoire  $X$  suit la loi géométrique de paramètre  $p$ , ce que l'on note  $X \hookrightarrow \mathcal{G}(p)$  si :

1.  $X(\Omega) = \mathbb{N}^*$
2.  $P(X = k) = q^{k-1}p$  où  $q = 1 - p$ .

Caractéristiques :

$$E(X) = \frac{1}{p} \quad V(X) = \frac{q}{p^2} \quad \varphi_X(u) = \frac{pe^{iu}}{1 - qe^{iu}}$$

Loi de Pascal d'ordre  $r$  : C'est la loi du nombre d'essais nécessaires pour observer exactement  $r$  fois un événement de probabilité  $p$ . Cette loi est la somme de  $r$  lois géométriques indépendantes.

On dit qu'une variable aléatoire  $X$  suit la loi de Pascal de paramètres  $r$  et  $p$ , ce que l'on note  $X \hookrightarrow \mathcal{P}(r, p)$  si :

1.  $X(\Omega) = \{r, r+1, \dots\}$
2.  $P(X = k) = C_{k-1}^{r-1} p^r q^{k-r}$  où  $q = 1 - p$ .

Caractéristiques :  $X$  admet alors une espérance et une variance :

$$E(X) = \frac{r}{p} \quad V(X) = \frac{r(1-p)}{p^2}.$$

## 4.6 Loi de Poisson

Soit  $\lambda > 0$ . On dit qu'une variable aléatoire  $X$  suit la loi de Poisson de paramètre  $\lambda$ , ce que l'on note  $X \hookrightarrow \mathcal{P}(\lambda)$  si :

1.  $X(\Omega) = \mathbb{N}$
2.  $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ .

La loi de Poisson est la loi des événements rares (de petite probabilité).

Caractéristiques :

$$E(X) = \lambda \quad V(X) = \lambda.$$

$$\varphi_X(u) = e^{\lambda(\cos u + i \sin u - 1)}.$$

Remarque : la loi de Poisson a été introduite en 1838 par Siméon-Denis Poisson (1781-1840), qui lui a donné son nom. Aucun rapport avec la loi de Fisher.

Exemple : [Le célèbre exemple de Von Bortkiewicz]

Von Bortkiewicz a étudié le nombre de morts par ruade de cheval dans l'armée prussienne de 1875 à 1894 dans 200 corps de cavalerie : pendant 20 ans, il a étudié 10 corps de cavalerie par an

|                              |     |    |    |   |   |
|------------------------------|-----|----|----|---|---|
| Nombre de morts par an       | 0   | 1  | 2  | 3 | 4 |
| Nombre de corps de cavalerie | 109 | 65 | 22 | 3 | 1 |

Calculer la moyenne  $\lambda$  du nombre de morts par an. Comparer la distribution réelle à la distribution résultant de l'application de la loi de Poisson de paramètre  $\lambda$ .

Exercices : retrouver  $E(X)$ ,  $V(X)$ ,  $\varphi_X(u)$ .

(Solutions sur [http://fr.wikipedia.org/wiki/Loi\\_de\\_Poisson](http://fr.wikipedia.org/wiki/Loi_de_Poisson))

## 4.7 Approximation de $\mathcal{B}$ par $\mathcal{P}$

Lorsque  $n$  devient grand, le calcul des probabilités d'une loi binomiale devient très fastidieux. On va donc, sous certaines conditions, trouver une approximation de  $P(X = k)$  plus manipulable. On constate le comportement asymptotique : si  $n \rightarrow \infty$  et  $p \rightarrow 0$ , alors  $X : \mathcal{B}(n, p) \rightarrow \mathcal{P}(\lambda)$  avec  $np = \lambda$ .

Remarque : cette approximation est correcte dès que  $n > 30$  et  $np < 5$  ou dès que  $n > 50$  et  $p < 0,1$ .

Preuve : Montrons que  $P(X = k) = C_n^k p^k q^{n-k}$  tend vers  $e^{-\lambda} \frac{\lambda^k}{k!}$  avec  $\lambda = np$ . Comme  $\lambda = np$ , on a  $p = \frac{\lambda}{n}$  et  $q = 1 - \frac{\lambda}{n}$  que l'on remplace dans la définition de la probabilité binomiale :

$$P(X = k) = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Or  $\left(1 - \frac{\lambda}{n}\right)^{n-k} = \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} = \left(1 - \frac{\lambda}{n}\right)^n \frac{1}{q^k}$ , d'où

$$P(X = k) = \frac{n(n-1)\cdots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \frac{1}{q^k} \left(1 - \frac{\lambda}{n}\right)^n = \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{q^k} \frac{1}{k!} \left(1 - \frac{\lambda}{n}\right)^n$$

Si  $n$  est assez grand ( $n \geq 50$  et  $p$  proche de 0 donc  $q$  proche de 1, on peut faire les approximations suivantes :

1.  $\frac{n(n-1)\cdots(n-k+1)}{n^k} = 1 \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \approx 1$
2.  $\frac{\lambda^k}{q^k} \approx \lambda^k$
3.  $\left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}$

Ainsi  $P(X = k) \approx e^{-\lambda} \frac{\lambda^k}{k!}$ .



# Chapitre 5

## Couple de variables aléatoires

### 5.1 Couple de v.a. discrètes

#### 5.1.1 Loi d'un couple de variables aléatoires discrètes

Soient  $X$  et  $Y$  deux variables aléatoires discrètes définies sur le même espace probabilisé  $(\Omega, A, P)$ . On notera  $X(\Omega) = \{x_i, i \in I\}$  et  $Y(\Omega) = \{y_j, j \in J\}$ , l'ensemble des valeurs, ordonnées, prises respectivement par  $X$  et  $Y$  (où  $I$  et  $J$  sont des ensembles d'entiers).

On appelle couple  $(X, Y)$  l'application

$$(X, Y) : \Omega \longrightarrow \mathbb{R}^2 \\ \omega \longmapsto (X(\omega), Y(\omega)).$$

Alors, l'ensemble  $(X, Y)(\Omega)$  des valeurs prises par le couple  $(X, Y)$  est inclus dans l'ensemble des couples de réels suivants  $\{(x_i, y_j), (i, j) \in I \times J\}$ .

**Définition 10** On appelle loi conjointe ou loi du couple  $(X, Y)$ , l'ensemble des couples

$$\{((x_i, y_j), p_{ij}), (i, j) \in I \times J\}$$

où  $p_{ij} = P((X = x_i) \cap (Y = y_j)) = P((X, Y)^{-1}(\{(x_i, y_j)\}))$ . On pourra représenter cette loi par un tableau à double entrée.

**Proposition 5.1.1**  $\{(x_i, y_j), p_{ij}), (i, j) \in I \times J\}$  est la loi d'un couple de variables discrètes si et seulement si  $p_{ij} \geq 0$  pour tout  $(i, j) \in I \times J$  et  $\sum_{(i,j) \in I \times J} p_{ij} = 1$ .

Remarque : Dans ce contexte,  $\sum_{(i,j) \in I \times J} p_{ij} = \sum_{\substack{i \in I \\ j \in J}} p_{ij} = \sum_{i \in I} \left( \sum_{j \in J} p_{ij} \right) = \sum_{j \in J} \left( \sum_{i \in I} p_{ij} \right)$ . On peut commencer par sommer sur les indices  $i$  puis les indices  $j$  ou inversement.

#### 5.1.2 Lois marginales

Les variables  $X$  et  $Y$  sont appelées variables marginales du couple  $(X, Y)$  et leur loi, appelée loi marginale de  $X$  (resp. de  $Y$ ) peut être obtenue de la façon suivante :

Supposons connue la loi du couple  $(X, Y) : \{(x_i, y_j), p_{ij}, (i, j) \in I \times J\}$ . On cherche maintenant à connaître la loi de  $X$  i.e. l'ensemble des couples  $\{(x_i, P(X = x_i)), i \in I\}$ . Or la famille des événements  $\{(Y = y_j), j \in J\}$  forme un système complet d'événements, donc d'après la formule des probabilités totales appliquée à ce système complet d'événements, on obtient :

$$p_{i\bullet} := P(X = x_i) = \sum_{j \in J} P((X = x_i) \cap (Y = y_j)) = \sum_{j \in J} p_{ij}.$$

De même, la loi de  $Y$  s'obtient à l'aide de la formule des probabilités totales appliquée au système complet d'événements  $\{(X = x_i), i \in I\}$  :

$$p_{\bullet j} := P(Y = y_j) = \sum_{i \in I} P((X = x_i) \cap (Y = y_j)) = \sum_{i \in I} p_{ij}.$$

### 5.1.3 Lois conditionnelles

On définit les lois conditionnelles par

$$P(X = x_i / Y = y_j) = P_{X|Y=y_j}(x_i) = \frac{p_{ij}}{p_{\bullet j}}$$

et

$$P(Y = y_j / X = x_i) = P_{Y|X=x_i}(y_j) = \frac{p_{ij}}{p_{i\bullet}}$$

### 5.1.4 Espérance, Variance

Sous réserve d'existence (dans le cas infini dénombrable), on a alors dans le cas discret

$$\begin{aligned} E(X) &= \sum_i P(X = x_i) x_i = \sum_i p_{i\bullet} x_i \\ V(X) &= \sigma^2(X) = \sum_i P(X = x_i) (x_i - E(X))^2 = \sum_i p_{i\bullet} x_i^2 - (E(X))^2, \dots \end{aligned}$$

et de façon plus générale, nous pouvons définir la notion d'espérance mathématique d'une fonction  $g(X, Y)$  du couple :

$$E(g(X, Y)) = \sum_i \sum_j p_{ij} g(x_i, y_j) \text{ dans le cas discret,}$$

Ainsi par exemple

$$\begin{aligned} E(XY) &= \\ E(X^2) &= \end{aligned}$$

### 5.1.5 Fonction de répartition

**Définition 11** Soit  $(X, Y)$  un couple de variables aléatoires. On appelle **fonction de répartition conjointe** de  $(X, Y)$  la fonction  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  définie par

$$F(x, y) = P(X < x \text{ et } Y < y).$$

Dans le cas de deux variables discrètes,

$$F(x, y) = \sum_{\substack{\{i / x_i < x\} \\ \{j / y_j < y\}}} p_{ij}.$$

Propriétés :

1.  $P(X = x_i \text{ et } Y < y_j) = F(x_{i+1}, y_j) - F(x_i, y_j)$ .
2. Les fonctions de répartition  $F_X$  et  $F_Y$  des lois marginales vérifient  $F_X(x) = \lim_{y \rightarrow \infty} F(x, y)$  et  $F_Y(y) = \lim_{x \rightarrow \infty} F(x, y)$ .

Preuve :

1. On découpe en une réunion disjointe  $(X, Y)^{-1}([-\infty, x_{i+1}[ \times ]-\infty, y_j]) = (X, Y)^{-1}([-\infty, x_i[ \times ]-\infty, y_j]) \cup (X, Y)^{-1}(\{x_i\} \times ]-\infty, y_j])$ .
2.  $X^{-1}([-\infty, x]) = \cup_{n \in \mathbb{N}} (X, Y)^{-1}([-\infty, x] \times ]-\infty, n])$ .

### 5.1.6 Indépendance

**Définition 12** Les v.a.  $X$  et  $Y$  sont indépendantes si pour tous  $i, j$ , les événements  $\{X = x_i\}$  et  $\{Y = y_j\}$  sont indépendants, c'est-à-dire  $p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$  ou encore  $P((X, Y)^{-1}(\{x_i\} \times \{y_j\})) = P(X^{-1}(\{x_i\})) \times P(Y^{-1}(\{y_j\}))$ .

**Théorème 5.1.2**  $X$  et  $Y$  sont indépendantes si et seulement si pour tous  $x$  et  $y$  réels,

$$F(x, y) = F_X(x) F_Y(y),$$

c'est-à-dire si les événements  $\{X < x\}$  et  $\{Y < y\}$  sont indépendants.

Preuve : Supposons  $X$  et  $Y$  indépendantes. Alors

$$\begin{aligned} F(x, y) &= \sum_{\substack{\{i / x_i < x\} \\ \{j / y_j < y\}}} p_{i\bullet} \cdot p_{\bullet j} = \sum_{\{i / x_i < x\}} \sum_{\{j / y_j < y\}} p_{i\bullet} \cdot p_{\bullet j} \\ &= \sum_{\{i / x_i < x\}} p_{i\bullet} \sum_{\{j / y_j < y\}} p_{\bullet j} = \sum_{\{i / x_i < x\}} p_{i\bullet} F_Y(y) = F_Y(y) \sum_{\{i / x_i < x\}} p_{i\bullet} \\ F(x, y) &= F_X(x) F_Y(y). \end{aligned}$$

Réciproquement, si pour tous  $x, y$ ,  $F(x, y) = F_X(x) F_Y(y)$  alors, en remarquant que

$$X = x_i \text{ et } Y < y_{j+1} \iff (X, Y) = (x_i, y_j) \text{ ou } X = x_i \text{ et } Y < y_j,$$

$$\begin{aligned} F(x_{i+1}, y_{j+1}) - F(x_i, y_{j+1}) &= P(X = x_i \text{ et } Y < y_{j+1}) = p_{ij} + P(X = x_i \text{ et } Y < y_j) \\ &= p_{ij} + F(x_{i+1}, y_j) - F(x_i, y_j). \end{aligned}$$

En appliquant l'hypothèse de départ, on a  $F(x_{i+1}, y_{j+1}) = F_X(x_{i+1}) F_Y(y_{j+1})$ ,  $F(x_i, y_{j+1}) = F_X(x_i) F_Y(y_{j+1})$ ,  $F(x_{i+1}, y_j) = F_X(x_{i+1}) F_Y(y_j)$  et  $F(x_i, y_j) = F_X(x_i) F_Y(y_j)$  d'où

$$p_{ij} = (F_X(x_{i+1}) - F_X(x_i)) (F_Y(y_{j+1}) - F_Y(y_j)) = p_{i\bullet} \cdot p_{\bullet j}.$$

**Théorème 5.1.3**  $X$  et  $Y$  sont indépendantes si et seulement si pour tout couple d'intervalles réels  $[a, a']$  et  $[b, b']$ , on a

$$P(X \in [a, a'] \text{ et } Y \in [b, b']) = P(X \in [a, a']) \cdot P(Y \in [b, b']).$$

(indépendance des événements  $X \in [a, a']$  et  $Y \in [b, b']$ )

Preuve : Si  $X$  et  $Y$  sont indépendantes,

$$\sum_{\substack{\{i / a \leq x_i < a'\} \\ \{j / b \leq y_j < b'\}}} p_{ij} = \sum_{\{i / a \leq x_i < a'\}} p_{i\bullet} \sum_{\{j / b \leq y_j < b'\}} p_{\bullet j}.$$

Réciproquement, on particularise le couple  $(a, a')$  et  $(b, b')$  en  $(x_i, x_{i+1})$  et  $(y_j, y_{j+1})$  respectivement.

$$\begin{aligned} p_{ij} &= P(x_i \leq X < x_{i+1} \text{ et } y_j \leq Y < y_{j+1}) = P(x_i \leq X < x_{i+1})P(y_j \leq Y < y_{j+1}) \\ &= P(X = x_i)P(Y = y_j) = p_{i\bullet} p_{\bullet j}. \end{aligned}$$

**Proposition 5.1.4** *Les v.a.  $X$  et  $Y$  sont indépendantes si et seulement si toutes les lois conditionnelles sont identiques aux lois marginales.*

Preuve :  $X, Y$  indépendantes alors  $p_{ij} = p_{i\bullet} p_{\bullet j}$  d'où

$$P(Y = y_j / X = x_i) = \frac{p_{ij}}{p_{i\bullet}} = \frac{p_{i\bullet} p_{\bullet j}}{p_{i\bullet}} = p_{\bullet j} = P(Y = y_j).$$

(même chose pour  $X$ )

Réciproquement, on suppose que pour tous  $i, j$ ,  $P(Y = y_j / X = x_i) = P(Y = y_j)$ .

$$P(X = x_i \text{ et } Y = y_j) = P(X = x_i / Y = y_j)P(X = x_i) = P(Y = y_j)P(X = x_i).$$

### 5.1.7 Covariance et Corrélation

**Définition 13** *Soient  $X$  et  $Y$  deux variables aléatoires. On appelle **covariance** de  $X$  et de  $Y$ , l'expression*

$$\text{Cov}(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))].$$

*En particulier, pour une variable discrète,*

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

Propriétés :

1.  $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \sum_i \sum_j p_{ij} x_i y_j - E(X) \cdot E(Y)$ .
2. Pour tout  $\lambda$  réel,  $V(X + \lambda Y) = V(X) + 2\lambda \text{Cov}(X, Y) + \lambda^2 V(Y)$ .
3. Si  $X$  et  $Y$  sont indépendantes alors  $E(XY) = E(X)E(Y)$ ,  $\text{Cov}(X, Y) = 0$  et  $V(X + Y) = V(X) + V(Y)$ .

Preuve :

1. Voir théorème de König
2. Par définition de la variance

$$\begin{aligned} V(X + \lambda Y) &= E([X + \lambda Y - E(X + \lambda Y)]^2) = E([X - E(X) + \lambda(Y - E(Y))]^2) \\ &= E([X - E(X)]^2 + 2\lambda[X - E(X)][Y - E(Y)] + \lambda^2[Y - E(Y)]^2) \\ &= E([X - E(X)]^2) + 2\lambda E([X - E(X)][Y - E(Y)]) + \lambda^2 E([Y - E(Y)]^2) \\ &= V(X) + 2\lambda \text{Cov}(X, Y) + \lambda^2 V(Y) \end{aligned}$$

$$3. E(XY) = \sum_i \sum_j p_{ij} x_i y_j \stackrel{(ind)}{=} \sum_i \sum_j p_{i\bullet} p_{\bullet j} x_i y_j = \sum_i p_{i\bullet} x_i \sum_j p_{\bullet j} y_j = E(X)E(Y).$$

La réciproque est ....

**Définition 14** *Soient  $X$  et  $Y$  deux variables aléatoires. On appelle **coefficient de corrélation linéaire** entre  $X$  et de  $Y$ , la quantité*

$$r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Propriétés :

1.  $r(X, Y) \in [-1, 1]$ .
2. Pour tous réels  $a, b, c, d$  ( $a, c \neq 0$ ),  $r(aX + b, cY + d) = r(X, Y)$ .
3. Si  $X$  et  $Y$  sont indépendantes alors  $r(X, Y) = 0$ .

Preuve de (1) :  $V(X + \lambda Y) \geq 0$  quel que soit  $\lambda$ . Cela implique que le polynôme en  $\lambda$ ,  $\lambda^2 V(Y) + 2\lambda \text{Cov}(X, Y) + V(X)$  a un déterminant  $\Delta' = \text{Cov}(X, Y)^2 - V(X)V(Y)$  négatif ou nul. Ainsi  $\text{Cov}(X, Y)^2 \leq V(X)V(Y)$  et  $r(X, Y) = \frac{\text{Cov}(X, Y)^2}{V(X)V(Y)} \leq 1$ .

## 5.2 Couple de v.a. continues

Soient  $X$  et  $Y$  deux variables aléatoires définies sur le même espace probabilisé  $(\Omega, P)$ . La loi de couple sera définie à partir de sa fonction de répartition :

$$F(x, y) = P((X, Y)^{-1}([-\infty, x[ \times ]-\infty, y]) = P(X < x \text{ et } Y < y).$$

Propriétés :

1.  $F$  est totalement croissante au sens large, c'est-à-dire que les fonctions partielles  $F(x, \cdot)$  et  $F(\cdot, y)$  sont croissantes pour tous  $x$  et  $y$  réels.
2.  $F(x, \cdot)$  et  $F(\cdot, y)$  sont continues à gauche
3.  $\lim_{x \rightarrow -\infty} F(x, y) = \lim_{y \rightarrow -\infty} F(x, y) = 0$ ,  $\lim_{n \rightarrow +\infty} F(n, n) = 1$

Preuve :

1. Soit  $y_1 > y_0$ . Alors l'ensemble  $] - \infty, x[ \times ] - \infty, y_1[$  se décompose en  $(] - \infty, x[ \times ] - \infty, y_0[) \cup (] - \infty, x[ \times [y_0, y_1[)$ . Ainsi  $F(x, y_1) - F(x, y_0) = P((X, Y)^{-1}([-\infty, x[ \times [y_0, y_1])) \geq 0$ .
2. On considère l'ensemble  $] - \infty, x[ \times ] - \infty, y_0[$  qui correspond à la réunion croissante des ensembles  $] - \infty, x[ \times ] - \infty, y_0 - 1/n[$ . On en déduit que  $F(x, y_0) = \lim_{n \rightarrow \infty} F(x, y_0 - 1/n) = \lim_{y \rightarrow y_0^-} F(x, y)$  puisque  $F(x, \cdot)$  est croissante.
3. L'intersection décroissante des ensembles  $] - \infty, -n[ \times ] - \infty, y[$  est vide et la réunion croissante des  $] - \infty, n[ \times ] - \infty, n[$  est  $\mathbb{R}^2$  tout entier.

La probabilité pourra être calculée sur tout type d'intervalle à partir de la fonction de répartition :

$$\begin{aligned} P(X \in [a, a'[ \text{ et } Y < y) &= F(a', y) - F(a, y) \\ P(X \in [a, a'[ \text{ et } Y \in [b, b'[) &= F(a', b') - F(a, b') + F(a, b) - F(a', b) \end{aligned}$$

Preuve : Il suffit de décomposer en ensembles disjoints :

$$\begin{aligned} ] - \infty, a'[ \times ] - \infty, y[ &= (] - \infty, a[ \times ] - \infty, y[) \cup ([a, a'[ \times ] - \infty, y[) \\ [a, a'[ \times ] - \infty, b'[ &= [a, a'[ \times ] - \infty, b[ \cup [a, a'[ \times [b, b'[ \end{aligned}$$

### 5.2.1 Fonction densité de probabilité

**Définition 15** Soient  $X$  et  $Y$  deux variables aléatoires définies sur le même espace probabilisé  $(\Omega, P)$ . La loi du couple  $(X, Y)$  est dite absolument continue s'il existe une fonction positive  $f$  de  $\mathbb{R}^2$  dans  $\mathbb{R}$ , telle que pour tous  $x$  et  $y$  réels,  $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, du \, dv$ . La fonction  $f$  est dite densité de probabilité du couple  $(X, Y)$ .

Propriétés :

1. Pour tout  $A$  élément de  $\mathcal{P}(\Omega^2)$ , alors  $P((X, Y) \in A) = \iint_A f(u, v) du dv$ .  
En particulier,  $\iint_{\mathbb{R}^2} f(u, v) du dv = 1$ .
2. En tout point où  $f$  est continue,  $f(x_0, y_0) = \frac{\partial^2 F}{\partial x \partial y}(x_0, y_0)$ .
3.  $X$  et  $Y$  sont aussi absolument continues, et leurs densités de probabilités respectives sont

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

Calcul de l'espérance et de la variance :

|                                                           |                                                          |
|-----------------------------------------------------------|----------------------------------------------------------|
| $E(X) = \iint_{\mathbb{R}^2} x f(x, y) dx dy$             | $E(Y) = \iint_{\mathbb{R}^2} y f(x, y) dx dy$            |
| $E(g(X, Y)) = \iint_{\mathbb{R}^2} g(x, y) f(x, y) dx dy$ | $E(XY) =$                                                |
| $V(X) = \iint_{\mathbb{R}^2} (x - E(X))^2 f(x, y) dx dy$  | $V(Y) = \iint_{\mathbb{R}^2} (y - E(Y))^2 f(x, y) dx dy$ |

Lorsque les quantités sont définies, on pose, comme dans le cas discret,

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) \quad \text{et} \quad r(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

### 5.2.2 Lois marginales et lois conditionnelles

◇ Si l'on s'intéresse à un événement sur  $X$  quelle que soit la valeur prise par  $Y$ , on obtient la loi de la v.a.  $X$  qui, dans le contexte d'un couple de v.a., est appelée (comme auparavant) loi marginale. Comme pour le couple  $(X, Y)$ , les lois marginales de  $X$  et  $Y$  sont connues par leur fonction de répartition :

$$F_X(x) = P(X < x) = P(X < x, Y \in \mathbb{R}) = \lim_{y \rightarrow +\infty} F(x, y).$$

De même,

$$F_Y(y) = P(Y < y) = P(X \in \mathbb{R}, Y < y) = \lim_{x \rightarrow +\infty} F(x, y).$$

◇ Soit  $(X, Y)$  un couple aléatoire absolument continu, de densité de probabilité  $f$ . Soit  $f_X$  la densité de probabilité de  $X$  et un réel  $x$  tel que  $f_X(x) \neq 0$ . La loi conditionnelle de  $Y$  liée par la condition  $X = x$  est définie par sa densité de probabilité  $f_x(y) = \frac{f(x, y)}{f_X(x)}$ .

### 5.2.3 Indépendance

**Définition 16** Les variables  $X$  et  $Y$  sont indépendantes si et seulement si la fonction de répartition du couple est égale au produit des fonctions de répartitions des lois marginales :  $F(x, y) = F_X(x)F_Y(y)$  pour tous  $x$  et  $y$  réels.

**Proposition 5.2.1** Les variables  $X$  et  $Y$  sont indépendantes si et seulement si pour tous  $x$  et  $y$  réels,  $f(x, y) = f_X(x)f_Y(y)$ . Dans ce cas,  $\text{Cov}(X, Y) = 0$ .

**Preuve** Si  $X$  et  $Y$  sont indépendantes, alors en dérivant  $F(x, y) = F_X(x)F_Y(y)$  par rapport à  $x$  puis par rapport à  $y$ , on obtient  $\frac{\partial F}{\partial x}(x, y) = f_X(x)F_Y(y)$  puis  $\frac{\partial^2 F}{\partial x \partial y}(x, y) = f_X(x)f_Y(y)$  et on utilise la propriété 2 :  $\frac{\partial^2 F}{\partial x \partial y}(x, y) = f(x, y)$ .

Réciproquement, si  $f(u, v) = f_X(u) f_Y(v)$ , alors

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = \int_{-\infty}^x f_X(u) \left( \int_{-\infty}^y f_Y(v) dv \right) du = F_X(x) F_Y(y).$$

De plus, si  $f(x, y) = f_X(x) f_Y(y)$ ,  $E(XY) = \iint_{\mathbb{R}^2} xy f_X(x) f_Y(y) dx dy = E(X)E(Y)$  et ainsi

$$\text{Cov}(X, Y) = 0.$$

**Proposition 5.2.2** *Les variables  $X$  et  $Y$  sont indépendantes si et seulement si pour tous couples d'intervalles réels  $[a, a'[, [b, b'[,$  on a*

$$P((X, Y)^{-1}([a, a'[\times[b, b'[)) = P(X^{-1}([a, a'[)) \cdot P(Y^{-1}([b, b'[)).$$

**Preuve** Si  $X$  et  $Y$  sont indépendantes, alors

$$\begin{aligned} P((X, Y)^{-1}([a, a'[\times[b, b'[)) &= F(a, b) + F(a', b') - F(a, b') - F(a', b) \\ &= F_X(a)F_Y(b) + F_X(a')F_Y(b') - F_X(a)F_Y(b') - F_X(a')F_Y(b) \\ &= (F_X(a') - F_X(a))(F_Y(b') - F_Y(b)) \\ &= P(X^{-1}([a, a'[)) \cdot P(Y^{-1}([b, b'[)) \end{aligned}$$

Réciproquement, comme  $] - \infty, a'[\times] - \infty, b'[,$  est la réunion croissante des  $[-n, a'[\times[-n, b'[,$  en posant  $a = b = -n$ ,  $a' = x$  et  $b' = y$  dans

$$P((X, Y)^{-1}([a, a'[\times[b, b'[)) = P(X^{-1}([a, a'[)) \cdot P(Y^{-1}([b, b'[)),$$

On a  $F(x, y) = \lim_{n \rightarrow \infty} P((X, Y)^{-1}([-n, x[\times[-n, y[)) = \lim_{n \rightarrow \infty} P(X^{-1}([-n, x[)) \cdot P(Y^{-1}([-n, y[)) = \lim_{n \rightarrow \infty} (F_X(x) - F_X(-n))(F_Y(y) - F_Y(-n)) = F_X(x) F_Y(y).$

**Proposition 5.2.3** *Les variables  $X$  et  $Y$  sont indépendantes si et seulement si les lois conditionnelles sont identiques aux lois marginales (En particulier,  $F_x(y)$  ne dépendra pas de  $x$ ).*

## 5.3 Somme et Produit

### 5.3.1 Somme

On considère la variable aléatoire  $Z = X + Y$ . On peut calculer  $E(Z)$  et  $V(Z)$  :

$$\begin{array}{ll} E(X + Y) &= E(X) + E(Y) && \text{(vraie si X et Y sont indépendantes ou non)} \\ V(X + Y) &= V(X) + V(Y) && \text{si X et Y indépendantes} \\ \text{et } V(X + Y) &= V(X) + V(Y) + 2 \text{Cov}(X, Y) && \text{sinon} \end{array}$$

Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes,  $\varphi_{X+Y} = \varphi_X \varphi_Y$ . Plus généralement, si  $X_1, \dots, X_n$  sont des variables aléatoires indépendantes dans leur ensemble, alors .

$$\varphi_{X_1 + \dots + X_n} = \varphi_{X_1} \cdots \varphi_{X_n}.$$

Remarque (admis) : En appliquant alors la transformée de Fourier à  $\varphi_{X+Y}$  cela permet de retrouver la loi de  $X + Y$ .

Calcul de la fonction densité de  $Z = X + Y$ .

On suppose que  $X$  et  $Y$  sont indépendantes et absolument continues, de densités respectives  $f_X$  et  $f_Y$ . Alors la fonction de répartition  $F$  de  $Z = X + Y$  est définie par :

$$\begin{aligned} F(z) &= P(X + Y < z) = \int_{-\infty}^{+\infty} f_X(x) \left( \int_{-\infty}^{z-x} f_Y(y) dy \right) dx \\ &= \int_{-\infty}^{+\infty} f_X(x) F_Y(z-x) dx = \int_{-\infty}^{+\infty} f_Y(y) F_X(z-y) dy \quad (\text{par symétrie}) \end{aligned}$$

Par dérivation (théorème admis) par rapport à  $z$ , la densité de probabilité de  $Z$  est définie par

$$f(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} f_Y(y) f_X(z-y) dy.$$

$f$  est le produit de convolution de  $f_X$  et  $f_Y$ , noté  $f_X * f_Y$ .

### 5.3.2 Produit

L'espérance du produit de deux variables aléatoires est donné par la formule  $E(XY) = E(X) E(Y) + Cov(X, Y)$ ,  $Cov()$  est la covariance entre les variables. En particulier, lorsque  $X$  et  $Y$  sont indépendantes,  $E(XY) = E(X) E(Y)$ .

### 5.3.3 Fonction de variable aléatoire

Soit  $\varphi$  une fonction dérivable de  $\mathbb{R}$  dans  $\mathbb{R}$ . En posant  $Y = \varphi \circ X$ , on obtient une nouvelle variable aléatoire, notée  $\varphi(X)$ , que l'on étudiera à l'aide de sa fonction de répartition.

Exemple : Soit  $X$  exprimant la consommation en litres aux 100 kilomètres d'une voiture. Aux Etats-Unis, on s'intéresse plus à la notion de distance parcourue avec un plein, que l'on retranscrit sous la forme  $Z$  est le nombre de miles parcourus avec un gallon d'essence (Plus précisément  $Z = 235/X$ ).

1. Cas où  $\varphi$  est monotone croissante

Soit  $F$  la fonction de répartition de  $X$ . La fonction de répartition  $G$  de  $Y$  est définie, pour  $y$  réel, par :  $G(y) = P(Y < y) = P(X < \varphi^{-1}(y)) = F(\varphi^{-1}(y))$ , soit encore  $F(x) = G(\varphi(x))$ . Si  $X$  est absolument continue,  $Y$  aussi et leurs densités de probabilité respectives,  $f$  et  $g$  sont liées par :  $f(x) = g(\varphi(x))\varphi'(x)$  ou  $g(y) = f(x)/\varphi'(x) = f(\varphi^{-1}(y))/\varphi'(\varphi^{-1}(y))$ .

Exemple :  $Y = e^X$  a pour densité de probabilité  $g(y) = f(x)e^{-x} = f(\ln y)/y$ .

2. Cas où  $\varphi$  est décroissante

Alors  $X > x$  équivaut à  $Y < \varphi(x)$ , donc  $G(y) = 1 - F(\varphi^{-1}(y))$ , que l'on peut écrire  $F(x) = 1 - G(\varphi(x))$ . Dans le cas absolument continu,  $g(y) = -f(x)/\varphi'(x) = -f(\varphi^{-1}(y))/\varphi'(\varphi^{-1}(y))$ .

Exemple :  $Y = c/X$  où  $c > 0$  et  $X$  est à valeurs dans  $]0, +\infty[$ .  $G(y) = P(Y < y) = P(\frac{c}{X} < y) = P(X > \frac{c}{y}) = 1 - F(\frac{c}{y})$  pour  $y > 0$  et  $G(y) = 0$  pour  $y \leq 0$ .

3. Cas  $\varphi$  quelconque

On résout alors au cas par cas l'inéquation  $Y < y$  afin de trouver la fonction de répartition de  $Y$ .

Exemple pour  $Y = X^2$  :

- pour  $y < 0$ ,  $Y < y$  est impossible ainsi  $G(y) = 0$ .
- pour  $y \geq 0$ ,  $Y < y$  correspond à  $-\sqrt{y} < X < \sqrt{y}$  alors  $G(Y) = F(\sqrt{y}) - F(-\sqrt{y})$ . Dans le cas où  $X$  est absolument continu,  $g(y) = \frac{1}{2\sqrt{y}}(f(\sqrt{y}) + f(-\sqrt{y}))$  sur  $\mathbb{R}^+$ .

Calcul de l'espérance :  $E(Y) = \sum_i p_i \varphi(x_i)$  dans le cas discret et  $E(Y) = \int_{\mathbb{R}} \varphi(x) f(x) dx$  dans le cas continu.

Calcul de la variance :  $V(Y) =$  dans le cas discret et  $V(Y) =$  dans le cas continu.

# Chapitre 6

## Lois continues usuelles

### 6.1 Loi continue uniforme

La variable aléatoire  $U$  est distribuée uniformément sur l'intervalle  $[a, b]$  si sa densité de probabilité est constante sur cet intervalle :

$$f(x) = \begin{cases} 1/(b-a) & \text{si } x \in [a, b] \\ 0 & \text{sinon} \end{cases}$$

On dit que  $U$  suit la loi uniforme et on note  $U \hookrightarrow \mathcal{U}(a, b)$ . Par conséquent, sa fonction de répartition est donnée par :

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ (x-a)/(b-a) & \text{si } x \in [a, b] \\ 1 & \text{si } x > b \end{cases}$$

$$E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}, \quad \varphi_X(t) = \frac{e^{ibt} - e^{iat}}{it(b-a)} = e^{it(a+b)/2} \frac{\sin(\frac{b-a}{2}t)}{\frac{b-a}{2}t}$$

Exercice : calcul de  $\gamma_1$  et  $\gamma_2$ .

### 6.2 Loi exponentielle

On souhaite modéliser l'intervalle de temps séparant deux occurrences successives d'un processus de Poisson. Ainsi la probabilité qu'il n'y ait aucune occurrence dans un intervalle de temps de longueur  $t$  est égale à  $p_0(t) = e^{-\lambda t}$  (absence de mémoire de la loi exponentielle) où  $\lambda > 0$  constituera le paramètre de la loi. Cette loi permet entre autres de modéliser la durée de vie de la radioactivité ou d'un composant électronique.

La fonction de répartition de la loi exponentielle  $\mathcal{E}(\lambda)$  est

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

et sa fonction de densité

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Caractéristiques :  $E(X) = 1/\lambda$ ,  $V(X) = 1/\lambda^2$ ,  $\varphi_X(t) = 1/(1 - it/\lambda)$ .

## 6.3 Loi normale

### 6.3.1 Rappel : calcul de l'intégrale de Gauss

Soient

$$G = \int_0^{+\infty} e^{-x^2} dx \quad \text{et} \quad H = \iint_{\mathbb{R}_+ \times \mathbb{R}_+} e^{-(x^2+y^2)} dx dy.$$

Compte tenu de ce que les variables  $x$  et  $y$  se séparent, le théorème de Fubini donne :

$$H = \iint_{\mathbb{R}_+ \times \mathbb{R}_+} e^{-x^2} e^{-y^2} dx dy = \left( \int_0^{+\infty} e^{-x^2} dx \right) \left( \int_0^{+\infty} e^{-y^2} dy \right) = G^2.$$

On passe en coordonnées polaires en posant  $x = r \cos \theta$  et  $y = r \sin \theta$  ; les variables  $r$  et  $\theta$  se séparent elles aussi :

$$H = \iint_{\mathbb{R}_+ \times [0, \frac{\pi}{2}[} e^{-r^2} r dr d\theta = \left( \int_0^{+\infty} e^{-r^2} r dr \right) \left( \int_0^{\frac{\pi}{2}} d\theta \right) = \frac{1}{2} \cdot \frac{\pi}{2}$$

car  $\int_0^{+\infty} e^{-r^2} r dr = \frac{1}{2} \int_0^{+\infty} e^{-u} du = \frac{1}{2}$  (par le changement de variable  $r = \sqrt{u}$ ). On en déduit :  $G^2 = \frac{\pi}{4}$ , d'où  $G = \frac{1}{2}\sqrt{\pi}$  puisque  $G \geq 0$ , et enfin :  $\int_{-\infty}^{+\infty} e^{-x^2} dx = 2G = \sqrt{\pi}$  par parité.

### 6.3.2 Gaussienne

On appelle loi normale (ou gaussienne) centrée réduite la loi définie par la densité de probabilité  $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$  définie par :

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}.$$

On peut vérifier qu'elle est continue et que son intégrale sur  $\mathbb{R}$  est égale à 1 : On sait que  $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$  (intégrale de Gauss) et en posant  $x = t/\sqrt{2}$ , on trouve que  $\int_{\mathbb{R}} \varphi(t) dt = 1$ .

Remarques :

1. la densité  $\varphi$  est une fonction paire ;
2. elle est indéfiniment dérivable et vérifie, pour tout  $t \in \mathbb{R}$ , l'identité  $\varphi'(t) = -t\varphi(t)$ .

La représentation graphique de cette densité est une courbe en cloche (ou courbe de Gauss).

On démontre par la suite que la loi définie par cette densité de probabilité admet une espérance nulle et une variance égale à 1. Sa fonction caractéristique vaut  $\varphi_X(u) = e^{-u^2/2}$  (à ne pas confondre avec la fonction densité). Le calcul se fait de la façon suivante :

$$\varphi_X(u) = \int_{-\infty}^{+\infty} e^{itu} \frac{e^{-t^2}}{\sqrt{2\pi}} dt.$$

Or  $-\frac{t^2}{2} + iut = -\frac{1}{2}(t^2 - 2iut) = -\frac{1}{2}(t^2 - 2iut + (-iut)^2 - (-iut)^2) = -\frac{1}{2}((t - iu)^2 + u^2)$ . On pose  $x = t - iu$ , ainsi

$$\varphi_X(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2 - u^2/2} dx = \frac{e^{-u^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = e^{-u^2/2}.$$

### 6.3.3 Moments

Les moments de cette loi existent tous. Pour tout  $r \in \mathbb{N}$ , le moment d'ordre  $r$  par rapport à l'origine est :

$$m_r = \int_{-\infty}^{+\infty} t^r \varphi(t) dt.$$

En raison de la parité de l'intégrande, tous les moments d'ordre impair sont nuls :

$$m_{2k+1} = 0.$$

Supposons à présent  $r$  pair :  $r = 2k$ , où  $k \in \mathbb{N}$ . Si  $k \geq 1$ , une intégration par parties donne :

$$m_{2k} = \int_{-\infty}^{+\infty} t^{2k-1} t \varphi(t) dt = - \int_{-\infty}^{+\infty} t^{2k-1} \varphi'(t) dt = (2k-1) \int_{-\infty}^{+\infty} t^{2k-2} \varphi(t) dt$$

ce qui fournit la relation de récurrence :

$$m_{2k} = (2k-1)m_{2k-2}.$$

De cette relation, on déduit, comme  $m_0 = 1$ , que :

$$m_{2k} = 1 \cdot 3 \cdots (2k-1) = \frac{(2k)!}{2^k k!}.$$

En particulier,  $m_1 = 0$  et  $m_2 = 1$ , ainsi l'espérance est nulle (la loi est donc dite centrée) et la variance vaut  $m_2 - m_1^2 = 1$  (la loi est donc dite réduite). Ceci justifie l'appellation de loi normale centrée réduite. Pour la suite on supposera  $\mu = 0$  et  $\sigma^2 = 1$ .

Des formules précédentes, on déduit encore :  $m_3 = 0$  et  $m_4 = 3$ .

La loi étant réduite, les moments centrés sont tous égaux aux moments par rapport à l'origine de même rang ; en particulier :  $\mu_2 = \sigma^2 = 1$ ,  $\mu_3 = 0$  et  $\mu_4 = 3\sigma^4$ . On en déduit l'asymétrie (skewness) :  $\gamma_1 = \frac{\mu_3}{\sigma^3} = 0$  et l'aplatissement (kurtosis) :  $\beta_2 = \frac{\mu_4}{\sigma^4} = 3$ .

### 6.3.4 Fonction de répartition

On note  $\Phi$  la fonction de répartition de la loi normale centrée réduite. Elle est définie, pour tout réel  $x$ , par :

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

$\Phi$  est la primitive de  $\varphi$  qui tend vers 0 en  $-\infty$  ; **cette primitive ne s'exprime pas à l'aide des fonctions usuelles** (exponentielle, etc.) mais devient elle-même une fonction usuelle, importante, pour quiconque pratique le calcul des probabilités ou les statistiques. Les valeurs de cette fonction peuvent donc se trouver sous la forme d'une table ou directement dans des logiciels de calcul statistique.

Propriétés de  $\Phi$  :

1. Elle est indéfiniment dérivable, et  $\Phi' = \varphi$
2. Elle est strictement croissante, tend vers 0 en  $-\infty$  et vers 1 en  $+\infty$ . (c'est donc une bijection  $\mathbb{R} \rightarrow ]0, 1[$  : pour tout  $p \in ]0, 1[$ , il existe un unique  $x \in \mathbb{R}$ , noté  $\Phi^{-1}(p)$ , tel que  $\Phi(x) = p$ )
3. Pour tout  $x \in \mathbb{R}$ ,  $\Phi(-x) = 1 - \Phi(x)$  (ceci résulte de la parité de la fonction densité) ; en particulier,  $\Phi(0) = 0,5$ . La table de valeurs sera donc établie pour les valeurs  $x$  positives.

### 6.3.5 Loi Normale ou de Laplace-Gauss

Plus généralement, on dit qu'une variable aléatoire réelle  $X$  suit une loi normale (ou loi normale gaussienne, loi de Laplace-Gauss) d'espérance  $\mu$  et d'écart type  $\sigma$  strictement positif si cette variable aléatoire réelle  $X$  admet pour densité de probabilité la fonction suivante définie, pour tout nombre réel  $x$ , par :

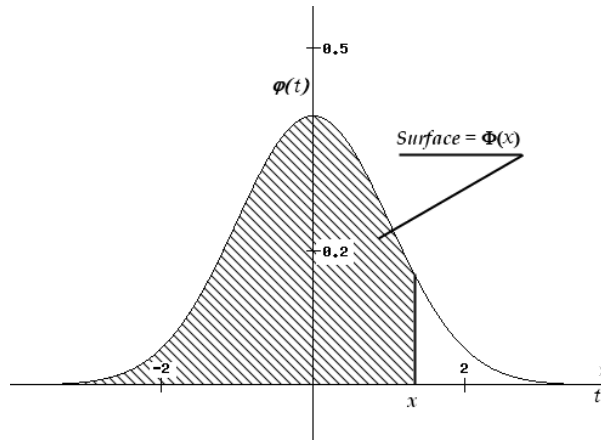
$$d(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Une telle variable aléatoire est appelée variable gaussienne. On notera  $X \hookrightarrow \mathcal{N}(\mu, \sigma)$ . On aura alors  $E(X) = \mu$ ,  $V(X) = \sigma^2$  et  $\varphi_X(u) = e^{i\mu u - \frac{\sigma^2 u^2}{2}}$ .

On pourrait étudier cette loi précisément mais on se ramène au cas précédent (loi normale centrée réduite) en considérant le théorème suivant :

**Théorème 6.3.1** *Si  $X$  suit  $\mathcal{N}(\mu, \sigma)$ , alors  $Z = \frac{X-\mu}{\sigma}$  suit  $\mathcal{N}(0, 1)$ .*

Ainsi une seule loi sera tabulée (celle de la loi normale centrée réduite), les autres pourront être déduites. Cette loi est une des plus importantes dans la théorie des probabilités.



### 6.3.6 Somme de deux variables gaussiennes

**Théorème 6.3.2** *Si  $X$  et  $Y$  sont deux variables indépendantes suivant des lois normales de moyennes respectives  $\mu_1$  et  $\mu_2$  et de variances  $\sigma_1^2$  et  $\sigma_2^2$  alors  $X + Y$  suit une loi normale de moyenne  $\mu = \mu_1 + \mu_2$  et de variance  $\sigma^2 = \sigma_1^2 + \sigma_2^2$ .*

Ce théorème se généralise bien sûr à toute somme finie de variables aléatoires normales indépendantes.

Preuve : Il s'agit de déterminer la loi de  $Z = X + Y$ . Calculons sa fonction caractéristique :

$$\begin{aligned} \varphi_Z(t) &= \varphi_X(t) \varphi_Y(t) \\ &= e^{i\mu_1 t - \frac{(\sigma_1 t)^2}{2}} e^{i\mu_2 t - \frac{(\sigma_2 t)^2}{2}} \\ &= e^{i(\mu_1 + \mu_2)t - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}} \end{aligned}$$

On en déduit que  $Z$  suit  $\mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ .

### 6.3.7 Somme de carrés de variables gaussiennes

Soient  $k$  variables aléatoires indépendantes  $X_1, \dots, X_k$  de même loi normale centrée et réduite, alors par définition la variable  $X$ , telle que

$$X = \sum_{i=1}^k X_i^2$$

suit une loi du  $\chi^2$  à  $k$  degrés de liberté. La loi du  $\chi^2$  (prononcer « khi-deux ») est une loi caractérisée par un paramètre dit degrés de liberté à valeur dans l'ensemble des entiers naturels (non nuls). Pour une variable aléatoire suivant une loi du  $\chi^2$  à  $k$  degrés de liberté, on notera la loi de  $\chi^2(k)$  la loi de  $X$ . Alors la densité de notée sera :

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

pour tout  $x$  positif où  $\Gamma$  (gamma) est la fonction

$$\Gamma : z \mapsto \int_0^{+\infty} t^{z-1} e^{-t} dt.$$

L'espérance mathématique de  $X$  vaut  $k$  et sa variance vaut  $2k$ .

### 6.3.8 Approximation de $\mathcal{B}$ par $\mathcal{N}$

(Voir Chap convergences pour justification)

Pour  $n$  assez grand, la loi binomiale se comporte comme une loi normale gaussienne d'espérance  $np$  et de variance  $npq$ . Plus précisément, le théorème de Moivre-Laplace précise que si  $\Phi$  est la fonction de répartition de la loi normale centrée réduite et si  $X_n$  suit une loi binomiale de paramètres  $n$  et  $p$ , on a alors, pour tout réel  $t$  :

$$\lim_{n \rightarrow \infty} P\left(\frac{X_n - np}{\sqrt{npq}} \leq t\right) = \Phi(t).$$

Le théorème de Berry-Esseen fournit une majoration de l'erreur commise quand on remplace  $P(X_n \leq x)$  par  $P(Y_n \leq x)$  où  $Y_n$  suit une loi normale d'espérance  $np$  et de variance  $npq$  : l'erreur commise est inférieure à  $\frac{C}{\sqrt{npq}}$  où  $C < 0,4784$  (Korolev & Shevtsova (2010)).

En pratique, on remplace une loi binomiale par une loi normale pour  $n$  grand et  $p$  pas trop proche de 0 ni de 1 (par exemple pour  $n > 30$ ,  $np > 5$  et  $nq > 5$  ou pour  $npq > 9$ )

### 6.3.9 Simulation

Pour simuler la loi normale, on peut utiliser la méthode de Box-Muller : Si  $U_1$  et  $U_2$  sont des variables aléatoires indépendantes qui suivent la loi uniforme sur  $]0,1[$ , alors on démontre assez aisément que les variables aléatoires :

$$\begin{aligned} T_1 &= \sqrt{-2 \ln U_1} \cos(2\pi U_2) \\ T_2 &= \sqrt{-2 \ln U_1} \sin(2\pi U_2) \end{aligned}$$

suivent toutes deux la loi normale centrée réduite (et sont indépendantes). On peut simuler toute loi normale  $\mathcal{N}(\mu, \sigma)$ , en construisant la variable  $Y = \mu + \sigma T_1$ .

## 6.4 Loi de Weibull

C'est une loi de probabilité continue appliquée aux durées de vie. C'est donc dans le contrôle de fiabilité que les entreprises ont tendance à l'utiliser, et plus précisément lorsque le taux de défaillance évolue comme une puissance du temps (ce qui est le cas le plus courant). Rappelons que lorsque ce taux est constant, on utilise la loi exponentielle, forme particulière de celle de Weibull, et lorsque le taux augmente proportionnellement au temps, c'est la distribution de Rayleigh qui est employée.

La loi de Weibull repose sur deux paramètres positifs, l'un de forme et l'autre d'échelle de temps. La loi de Weibull à trois paramètres prend en compte la "localisation", c'est-à-dire un éventuel décalage du départ de la courbe par rapport à l'origine (soit à gauche soit à droite).

On prendra  $\alpha$  comme paramètre de forme, et  $\lambda$  étant celui de temps.

Le paramètre  $\alpha$  est habituellement supérieur à 1 : le taux de défaillance croît avec le temps. S'il est inférieur, c'est pendant le rodage que les risques de défaillance sont élevés et s'il est égal à 1, on retombe sur la loi exponentielle.

Sa fonction de répartition

$$F(x) = \begin{cases} 1 - e^{-\lambda x^\alpha} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

et sa fonction de densité

$$f(x) = \begin{cases} \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

Caractéristiques :  $E(X) = \Gamma(1 + 1/\alpha)/\lambda^{1/\alpha}$ ,  $V(X) = \frac{\Gamma(1+2/\alpha) - \Gamma(1+1/\alpha)^2}{\lambda^{2/\alpha}}$  où  $\Gamma$  est la fonction Gamma d'Euler.

**Proposition 6.4.1** *Si  $X$  suit une loi exponentielle de paramètre  $\lambda$  alors  $X^{1/\alpha}$  suit une loi de Weibull  $W(\lambda, \alpha)$ .*

## 6.5 Loi de Pareto

L'économiste italien Vilfredo Pareto (1848-1923) observa au début du XX<sup>e</sup> siècle que 20% de la population italienne possédait 80% de la richesse nationale d'où le nom de la loi 80-20 ou 20-80.

La loi de Pareto admet deux paramètres  $(c, \alpha)$ . Le premier paramètre ( $c > 0$ ) tronque la distribution : le domaine de définition de  $X$  suivant cette loi est alors restreint à  $]c, +\infty[$  (introduction de la contrainte  $x > c$ ). Le deuxième paramètre est le paramètre de forme  $\alpha > 0$ . Alors la distribution est caractérisée par :

$$P(X > x) = \left(\frac{x}{c}\right)^{-\alpha}$$

avec  $x > c$ . La fonction de densité est alors

$$f(x) = \frac{\alpha}{c} \left(\frac{c}{x}\right)^{\alpha+1}$$

et la fonction de répartition est

$$F(x) = 1 - \left(\frac{c}{x}\right)^\alpha$$

Caractéristiques :  $E(X) = \frac{\alpha c}{\alpha-1}$  pour  $\alpha > 1$ ,  $V(X) = \frac{\alpha}{(\alpha-1)^2(\alpha-2)} c^2$  pour  $\alpha > 2$ ,  $E(X^k) = \frac{\alpha}{\alpha-k} c^k$  pour  $\alpha > k$ . Fonction caractéristique :  $\alpha(-ict)^\alpha \Gamma(-\alpha, -ict)$

## 6.6 Loi de Gumbel

C'est une loi de modélisation de valeurs extrêmes dont la fonction de répartition est la suivante :

$$F(x) = e^{-e^{-(x-\alpha)/\beta}}.$$

Sa fonction de densité est

$$f(x) = e^{-x-e^{-x}}$$

Caractéristiques de la loi :  $E(X) = \alpha + \gamma\beta$  (où  $\gamma = 0.5772156\dots$  est la constante d'Euler ),  $V(X) = \beta^2\pi^2/6$ , mode =  $\alpha$ . Fonction caractéristique :  $\varphi_X(t) = \Gamma(1 - i\beta t)e^{i\alpha t}$



# Chapitre 7

## Convergences

### 7.1 Convergence en probabilité

#### Rappel : Inégalité de Bienaymé-Chebyshev

Soit  $X$  une variable aléatoire admettant une espérance  $E(X)$  et de variance finie  $\sigma^2$  (l'hypothèse de variance finie garantit l'existence de l'espérance).

L'inégalité de Bienaymé-Tchebychev s'énonce de la façon suivante : pour tout réel  $\varepsilon$  strictement positif,

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

**Définition 17 (Convergence en probabilité)** On considère une suite  $(X_n)$  d'une v.a. définie sur  $\Omega$ ,  $X$  une autre v.a. définie sur  $\Omega$ .

On dit que la suite  $(X_n)$  converge en probabilité vers une constante réelle  $\ell$  si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - \ell| > \varepsilon) = 0.$$

On dit que la suite  $(X_n)$  converge en probabilité vers  $X$  si

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

Conséquence : Pour que  $(X_n)$  converge en probabilité vers  $X$ , il faut et il suffit que  $E(X_n - X) \rightarrow 0$  et  $V(X_n - X) \rightarrow 0$  lorsque  $n \rightarrow \infty$  (la démonstration passe par l'inégalité de Bienaymé-Chebyshev).

#### 7.1.1 Exemple de la loi binomiale

On réalise  $n$  expériences indépendantes et on suppose que lors de chacune de ces expériences, la probabilité d'un événement appelé "succès" est  $p$ . Soit  $S_n$  le nombre de succès obtenus lors de ces  $n$  expériences. La variable aléatoire  $S_n$ , somme de  $n$  variables de Bernoulli indépendantes, de même paramètre  $p$ , suit une loi binomiale :  $S_n \hookrightarrow \mathcal{B}(n, p)$ .

On s'intéresse alors à la variable aléatoire  $\frac{S_n}{n}$ , proportion de succès sur  $n$  expériences, a donc pour espérance  $E(\frac{S_n}{n}) = p$  et pour variance  $V(\frac{S_n}{n}) = \frac{1}{n^2} V(S_n) = \frac{p(1-p)}{n}$ . Comme  $p(1-p)$  atteint son maximum lorsque  $p = 1/2$ , on a ainsi  $p(1-p) \leq 1/4$ . En appliquant l'inégalité de Bienaymé-Chebyshev, il vient

$$P(|S_n/n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Ainsi pour tout  $\varepsilon > 0$ , il existe  $\eta > 0$  (plus précisément  $\eta > \frac{1}{4n\varepsilon^2}$ ) tel que  $P(|S_n/n - p| \geq \varepsilon) < \eta$  ou encore  $\lim_{n \rightarrow \infty} P(|S_n/n - p| \geq \varepsilon) = 0$ . La variable aléatoire  $\frac{S_n}{n}$  converge en probabilité vers  $p$ .

### 7.1.2 Convergence en probabilité

**Théorème 7.1.1** Soit  $(X_n)$  une suite de variables aléatoires sur le même espace probabilisé  $(\Omega, P)$  admettant des espérances et des variances vérifiant

$$\lim_{n \rightarrow \infty} E(X_n) = \ell \quad \text{et} \quad \lim_{n \rightarrow \infty} V(X_n) = 0,$$

alors les  $(X_n)$  convergent en probabilité vers  $\ell$ .

**Preuve** Soit  $\varepsilon > 0$ . Posons  $E(X_n) = \ell + u_n$  avec  $\lim u_n = 0$ . Alors il existe  $N \in \mathbb{N}$  tel que :

$$n \geq N \Rightarrow |u_n| < \varepsilon/2$$

et donc à partir du rang  $N$ ,

$$|X_n - E(X_n)| < \varepsilon/2 \Rightarrow |X_n - \ell| < \varepsilon, \quad (7.1)$$

car  $|X_n - \ell| = |X_n - E(X_n) + E(X_n) - \ell| \leq |X_n - E(X_n)| + |E(X_n) - \ell|$ .

L'implication (7.1) peut être encore écrite sous la forme

$$|X_n - \ell| \geq \varepsilon \Rightarrow |X_n - E(X_n)| \geq \varepsilon/2.$$

Par conséquent, en utilisant l'inégalité de Bienaymé-Chebyshev,

$$P(|X_n - \ell| \geq \varepsilon) \leq P(|X_n - E(X_n)| \geq \varepsilon/2) \leq \frac{V(X_n)}{(\varepsilon/2)^2},$$

qui tend vers 0 quand  $n$  tend vers l'infini.

### 7.1.3 Loi faible des grands nombres

**Théorème 7.1.2** Soit  $(X_n)$  une suite de variables aléatoires indépendantes sur le même espace probabilisé  $(\Omega, P)$  ayant une même espérance mathématique  $\ell$  et des variances vérifiant  $\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 = 0$ . On pose  $S_n = X_1 + \dots + X_n$  alors  $\frac{S_n}{n}$  converge en probabilité vers  $\ell$ .

Si on considère une suite de variables aléatoires  $(X_n)$  indépendantes définies sur un même espace probabilisé, ayant même espérance et même variance finie notées respectivement  $E(X)$  et  $V(X)$ . La loi faible des grands nombres stipule que, pour tout réel  $\varepsilon$  strictement positif, la probabilité que la moyenne empirique  $\frac{S_n}{n}$  s'éloigne de l'espérance d'au moins  $\varepsilon$ , tend vers 0 quand  $n$  tend vers l'infini. La moyenne  $\frac{S_n}{n}$  converge en probabilité vers l'espérance commune  $E(X)$ .

## 7.2 Convergence en loi

**Définition 18** Soient  $(X_n)$  et  $X$  des variables aléatoires sur un même espace probabilité  $(\Omega, P)$ , de fonctions de répartition respectives  $F_n$  et  $F$  ; on dit que les  $(X_n)$  convergent vers  $X$  en loi (et on note  $X_n \xrightarrow{L} X$ ) si en tout point  $x$  où  $F$  est continue, les  $F_n(x)$  convergent vers  $F(x)$ .

Propriétés : (admisses)

1. La convergence en probabilité entraîne la convergence en loi.
2. Si les  $(X_n)$  et  $X$  sont des variables aléatoires discrètes, alors  $X_n$  converge en loi vers  $X$  si et seulement si

$$\forall x \in \mathbb{R}, \lim_{n \rightarrow \infty} P(X_n = x) = P(X = x).$$

**Proposition 7.2.1 (Convergence de la loi hypergéométrique vers la loi binomiale)** Soit  $(X_N)$  une suite de variables aléatoires sur un même espace probabilisé, de loi hypergéométrique :  $X_N \hookrightarrow \mathcal{H}(N, n, p)$  où  $n$  et  $p$  sont supposés constants. Alors  $(X_N)$  convergent en loi, quand  $N$  tend vers l'infini, vers  $X$  de loi binomiale  $\mathcal{B}(n, p)$  (mêmes valeurs de paramètres).

**Preuve** La probabilité ponctuelle de  $X_N$  est

$$P(X_N = k) = \frac{C_{Np}^k C_{Nq}^{n-k}}{C_N^n}.$$

Lorsque  $N$  tend vers l'infini avec  $n$  constant,

$$C_N^n = \frac{N(N-1) \cdots (N-n+1)}{n!} = N^n \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{n-1}{N}\right) \frac{1}{n!} \equiv \frac{N^n}{n!}$$

car  $(1 - \frac{m}{N}) \equiv 1$  lorsque  $N$  tend vers l'infini. De même, lorsque  $N$  tend vers l'infini avec  $p$  et  $k$  fixes, alors

$$C_{Np}^k \equiv \frac{(Np)^k}{k!} \quad \text{et} \quad C_{N(1-p)}^{n-k} \equiv \frac{(N(1-p))^{n-k}}{(n-k)!}.$$

Finalement,

$$P(X_N = k) \equiv \frac{p^k (1-p)^{n-k} n!}{k!(n-k)!} = C_n^k p^k (1-p)^{n-k},$$

ce qui correspond à la probabilité ponctuelle d'une variable aléatoire qui suit la loi binomiale  $\mathcal{B}(n, p)$ .

C'est pour cela que lorsque la population (de taille  $N$ ) est très grande, on peut assimiler la loi d'une variable aléatoire comptant le nombre de réussite sur un tirage sans remise (loi hypergéométrique) à une loi binomiale (tirage avec remise).

**Proposition 7.2.2 (Convergence de la loi binomiale vers une loi de Poisson)** Soit  $(X_n)$  une suite de variables aléatoires binomiales sur un même espace probabilisé : pour tout  $n$ ,  $X_n$  suit  $\mathcal{B}(n, p_n)$ . On suppose que  $\lim_{n \rightarrow +\infty} p_n = 0$  et  $\lim_{n \rightarrow +\infty} np_n = \lambda$ . Alors  $(X_n)$  convergent en loi, quand  $n$  tend vers l'infini, vers une loi de Poisson de paramètre  $\lambda$ .

**Preuve** Pour  $k$  fixé,

$$\begin{aligned} P(X_n = k) &= \frac{n(n-1) \cdots (n-k+1)}{k!} p_n^k (1-p_n)^{n-k} \\ &= \frac{(np_n)^k}{k!} (1-p_n)^n \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) (1-p_n)^{-k} \end{aligned}$$

On cherche la limite de  $(1-p_n)^n = \exp(n \ln(1-p_n)) = \exp(n \ln(1-np_n/n))$ . Comme  $\lim_{n \rightarrow +\infty} np_n = \lambda$ , on pose  $np_n = \lambda + \varepsilon_n$  avec  $\lim_{n \rightarrow +\infty} \varepsilon_n = 0$  et ainsi  $\ln(1-np_n/n) \sim_{\infty} -\lambda/n$  donc  $\lim_{n \rightarrow +\infty} (1-p_n)^n = e^{-\lambda}$ . Comme  $k$  est fixé,  $\lim_{n \rightarrow +\infty} (1 - \frac{1}{n}) \cdots (1 - \frac{k-1}{n}) (1-p_n)^{-k} = 1$

Ainsi

$$\lim_{n \rightarrow +\infty} P(X_n = k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

ce qui correspond à la probabilité ponctuelle d'une variable aléatoire qui suit une loi de Poisson  $\mathcal{P}(\lambda)$ . Il s'agit donc d'une convergence en loi en appliquant le point 2 des propriétés.

**Corollaire 7.2.3 (Application pratique)** On peut remplacer  $\mathcal{B}(n, p)$  par  $\mathcal{P}(\lambda)$  avec  $\lambda = np$  pour  $n$  très grand ( $n > 50$ ) et  $p$  très petit ( $p < 0,1$ ).

## 7.3 Convergence des fonctions caractéristiques

### 7.3.1 Continuité

**Théorème 7.3.1 (théorème de continuité de Levy)** Soit  $(X_n)$  une suite de variables aléatoires de fonctions caractéristiques  $\varphi_{X_n}$  et  $X$  une variable aléatoire de fonction caractéristique  $\varphi_X$ , toutes sur un même espace probabilisé. Si les  $(X_n)$  convergent en loi vers  $X$  alors la suite de fonctions  $(\varphi_{X_n})$  converge uniformément vers  $\varphi_X$  sur tout intervalle  $[-a, a]$ .

Inversement si les  $(\varphi_{X_n})$  convergent vers une fonction  $\varphi$  dont la partie réelle est continue en 0, alors  $\varphi$  est la fonction caractéristique d'une variable aléatoire  $X$  vers laquelle les  $X_n$  convergent en loi.

On peut le résumer ainsi :

$$\{\forall t \in \mathbb{R}; \varphi_{X_n}(t) \rightarrow \varphi_X(t)\} \Leftrightarrow \{X_n \xrightarrow{L} X\}$$

### 7.3.2 Théorème central limite

**Corollaire 7.3.2 (Théorème central limite)** Soit une suite  $(X_n)$  de variables aléatoires définies sur le même espace de probabilité, suivant la même loi  $D$  et dont l'espérance  $\mu$  et l'écart-type  $\sigma$  communes existent et soient finis ( $\sigma \neq 0$ ). On suppose que les  $(X_n)$  sont indépendantes. Considérons la somme  $S_n = X_1 + \dots + X_n$ . Alors l'espérance de  $S_n$  est  $n\mu$  et son écart-type vaut  $\sigma\sqrt{n}$  et  $\frac{S_n - n\mu}{\sigma\sqrt{n}}$  converge en loi vers une variable aléatoire normale centrée réduite.

**Preuve** Posons  $Y_i = \frac{X_i - \mu}{\sigma\sqrt{n}}$ . Alors

$$\varphi_{Y_i}(t) = \varphi_{\frac{X_i - \mu}{\sigma\sqrt{n}}}(t) = \varphi_{X_i - \mu}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Pour  $t$  fixé, lorsque  $n$  tend vers l'infini,  $\frac{t}{\sigma\sqrt{n}}$  est infiniment petit. Ecrivons le développement limité, au voisinage de 0, de la fonction caractéristique d'une variable aléatoire  $W$  :

$$\begin{aligned} \varphi_W(u) &= \varphi_W(0) + u \varphi'_W(0) + \frac{t^2}{2} \varphi''_W(0) + u^2 \varepsilon(u) \\ &= 1 + i u E(W) - \frac{u^2}{2} E(W^2) + u^2 \varepsilon(u) \end{aligned}$$

En posant  $W = X_i - \mu$ ,  $u = t/(\sigma\sqrt{n})$ , on a  $E(W) = E(X_i - \mu) = 0$  et  $E(W^2) = E((X_i - \mu)^2) = V(X_i) = \sigma^2$  d'où

$$\varphi_{X_i - \mu}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2}{2\sigma^2 n} \sigma^2 + \frac{1}{n} \varepsilon(t^3/\sigma^3\sqrt{n}) = 1 - \frac{t^2}{2n} + \frac{1}{n} \varepsilon_i(n)$$

avec  $\lim_{n \rightarrow +\infty} \varepsilon_i(n) = 0$ .

Maintenant, posons  $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \sum_{i=1}^n Y_i$ .

L'indépendance des  $X_n$  entraîne celle des  $Y_i$  et ainsi

$$\begin{aligned} \varphi_{Z_n}(t) &= \prod_{i=1}^n \varphi_{Y_i}(t) \\ &= \exp\left(\sum_{i=1}^n \ln\left(1 - \frac{t^2}{2n} + \frac{1}{n} \varepsilon_i(n)\right)\right) \end{aligned}$$

et  $\lim_{n \rightarrow +\infty} \varphi_{Z_n}(t) = e^{-t^2/2}$  qui est la fonction caractéristique de  $\mathcal{N}(0, 1)$ .

Ce théorème établit une propriété générale, qui va justifier l'importance considérable de la loi normale, à la fois comme modèle pour décrire des situations pratiques, mais aussi comme outil théorique. Il s'énonce ainsi :

« Soit  $X_1, \dots, X_i, \dots, X_n$ , une suite de  $n$  variables aléatoires indépendantes, de moyennes  $\mu_1, \dots, \mu_i, \dots, \mu_n$ , et de variances  $s_1^2, \dots, s_i^2, \dots, s_n^2$ , et de lois de probabilité quelconques, leur somme suit une loi qui, lorsque  $n$  augmente, tend vers une loi normale de moyenne  $\mu = \sum_{i=1}^n \mu_i$  et de variance  $s^2 = \sum_{i=1}^n s_i^2$ . Il y a une seule condition restrictive, c'est que les variances soient finies et qu'aucune ne soit prépondérante devant les autres. »

La loi normale comme modèle : prenons l'exemple du fonctionnement d'un tour d'usinage du bois. Le réglage du tour a pour but d'obtenir des pièces présentant une cote bien définie ; mais on sait que de multiples causes perturbatrices agissent au cours de l'usinage d'une pièce : vibrations, usures, variations de courant ... Or si les causes perturbatrices sont nombreuses, si leurs effets interviennent de façon additive, enfin si la dispersion provoquée par chacune d'elles reste faible par rapport à la dispersion totale, alors le théorème central limite signifie qu'on doit observer une fluctuation globale très voisine de la loi normale. Et, comme ce mécanisme d'intervention de causes perturbatrices est très répandu dans la nature, il en résulte que la loi normale occupe en statistique une place privilégiée.

### 7.3.3 convergence de $\mathcal{P}$ vers $\mathcal{N}$

**Corollaire 7.3.3** Soit  $(X_n)$  une suite de variables aléatoires suivants des lois de Poisson de paramètres  $\lambda_n$ . Si  $\lim_{n \rightarrow +\infty} \lambda_n = \infty$ , alors  $\frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$  converge en loi vers  $\mathcal{N}(0, 1)$ .

**Preuve** On utilise la fonction caractéristique de la loi de Poisson de paramètre  $\lambda$  :

$$\varphi_X(t) = e^{\lambda(\cos t + i \sin t - 1)}.$$

En utilisant les propriétés de la fonction caractéristique ( $\varphi_{aX}(t) = \varphi(at)$  et  $\varphi_{X+b}(t) = e^{itb} \varphi_X(t)$ ), il vient  $\varphi_{X-\lambda}(t) = e^{-it\lambda} e^{\lambda(\cos t + i \sin t - 1)}$  puis  $\varphi_{\frac{X-\lambda}{\sqrt{\lambda}}}(t) = e^{\lambda(\cos \frac{t}{\sqrt{\lambda}} + i \sin \frac{t}{\sqrt{\lambda}} - 1)} e^{i \frac{t}{\sqrt{\lambda}}(-\lambda)}$ . Or, lorsque  $\lambda$  tend vers l'infini,  $1/\lambda$  est au voisinage de 0 et

$$\begin{aligned} \cos(t/\sqrt{\lambda}) &\sim 1 - \frac{(t/\sqrt{\lambda})^2}{2} + \frac{1}{\lambda} \varepsilon(\lambda) \\ \sin(t/\sqrt{\lambda}) &\sim (t/\sqrt{\lambda}) + \frac{1}{\lambda} \varepsilon(\lambda) \end{aligned}$$

avec  $\lim_{\lambda \rightarrow \infty} \varepsilon(\lambda) = 0$ . Ou encore le développement de l'exposant avec  $1/\lambda$  au voisinage de 0 est

$$e^{it/\sqrt{\lambda}} - 1 = \frac{it}{\sqrt{\lambda}} + \frac{(it)^2}{2\lambda} + \frac{1}{\lambda} \varepsilon(\lambda).$$

Ainsi

$$\lambda(\cos(t/\sqrt{\lambda}) + i \sin(t/\sqrt{\lambda}) - 1) - i\sqrt{\lambda}t \sim -t^2/2$$

et  $\varphi_{\frac{X-\lambda}{\sqrt{\lambda}}}(t) \sim e^{-t^2/2}$ , fonction caractéristique de  $\mathcal{N}(0, 1)$ .

Application pratique : Pour  $\lambda$  suffisamment grand (disons  $\lambda > 1000$ ), la distribution normale de moyenne  $\lambda$  et de variance  $\lambda$  est une excellente approximation de la distribution de Poisson de paramètre  $\lambda$ . Si  $\lambda$  est plus grand que 10, alors la distribution normale est une bonne approximation si une correction de continuité est appliquée, c'est-à-dire  $P(X \leq x)$  lorsque  $x$  est un entier positif ou nul est remplacé par  $P(X \leq x + 0, 5)$ .

### 7.3.4 convergence de $\mathcal{B}$ vers $\mathcal{N}$

**Corollaire 7.3.4 (Théorème de Moivre-Laplace)** Soit  $(X_n)$  une suite de variables aléatoires telles que  $(X_n) \in \mathcal{B}(n, p)$ . Alors  $\frac{X_n - np}{\sqrt{npq}}$  converge en loi vers la variable centrée réduite  $Z \in \mathcal{N}(0, 1)$  ou encore  $X_n$  converge en loi vers  $\mathcal{N}(np, \sqrt{npq})$ .

**Preuve** On rappelle que l'on a défini une variable de Bernoulli comme une variable qui prend la valeur 1 avec la probabilité  $p$ , et la valeur 0 avec la probabilité  $(1 - p)$ , et montré que sa moyenne est égale à  $p$  et sa variance à  $p(1 - p)$ . Or on peut considérer une variable binomiale comme la somme de  $n$  variables de Bernoulli. Il résulte du théorème central limite que, si  $n$  est suffisamment grand (en pratique à partir de  $n = 50$ ), la loi binomiale peut être approximée par une loi normale de moyenne  $np$  et de variance  $np(1 - p)$ . C'est pourquoi les tables de la loi binomiale s'arrêtent généralement à  $n = 50$ .

Application pratique : on peut assimiler une loi binomiale à une loi normale dès que  $np > 15$  et  $nq > 15$  ou  $n > 30, np > 5, nq > 5$ .