

# Probabilités

Jean-Romain Heu

2014



# Chapitre 1

## Mesures de probabilité

### 1.1 Définitions et exemples

**Définition 1.1.1.** Soit  $\Omega$  un ensemble.

Une **mesure de probabilité** (ou **distribution de probabilité**) est une application définie sur une partie  $\mathcal{F}$  de  $\mathcal{P}(\Omega)$  de la forme  $\mathbf{P} : \mathcal{F} \rightarrow [0, 1]$  et telle que

- $\mathbf{P}(\Omega) = 1$ .
- Si  $A$  et  $B$  sont deux parties disjointes de  $\Omega$ , alors  $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$ .
- Et plus généralement, si  $(A_n)$  est une suite de parties deux à deux disjointes de  $\Omega$ , alors

$$\mathbf{P}\left(\bigcup_{n \in \mathbf{N}} A_n\right) = \sum_{n \in \mathbf{N}} \mathbf{P}(A_n).$$

On dit que  $(\Omega, \mathcal{F}, \mathbf{P})$  est un **espace probabilisé**. L'ensemble  $\Omega$  est appelé **univers** et les éléments de  $\mathcal{F}$  sont appelés **événements**.

**Définition 1.1.2.** On dit qu'une mesure de probabilité sur  $\Omega$  est **discrète** si l'ensemble  $\Omega$  est un ensemble fini ou dénombrable.

On parle de mesure de probabilité **continue** pour les mesures définies sur  $\mathbf{R}$  à partir d'une fonction de densité. Ce sont les mesures de probabilités  $\mathbf{P}$  de la forme suivante : pour tout intervalle  $[a, b]$  de  $\mathbf{R}$ ,

$$\mathbf{P}([a, b]) = \int_a^b f(x) dx,$$

où  $f$  est une fonction positive telle que  $\int_{-\infty}^{+\infty} f(x) dx = 1$ . La fonction  $f$  est appelée **densité** de probabilité de  $\mathbf{P}$ .

**Proposition 1.1.1.** Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace probabilisé et soient  $A$  et  $B$  des événements.

- $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A)$
- $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$

**Exemple 1.** Soient  $0 \leq k \leq n$  deux entiers. On lance une pièce biaisée  $n$  fois de suite. Quelle est la probabilité d'obtenir exactement  $k$  fois « pile » ?

Il faut commencer par se donner un espace probabilisé. Un événement est ici le résultat de  $n$  lancers. L'univers  $\Omega$  est donc l'ensemble des  $n$ -uplets à valeurs dans  $\{P, F\}$ . Il contient  $2^n$  éléments.

Il faut ensuite définir une mesure de probabilité sur  $\mathcal{P}(\Omega)$ . Comme  $\Omega$  est un ensemble discret (et même fini), il suffit de définir la probabilité de chaque élément de  $\Omega$ . Notons  $p$  la probabilité d'obtenir pile pour un lancer de la pièce. Alors  $1-p$  est la probabilité d'obtenir face. On peut considérer que les  $n$  lancers sont indépendants. Alors la probabilité d'obtenir la séquence PFPF est  $p^2(1-p)p$ . On définit alors la probabilité d'un  $n$ -uplet  $X = (x_1, x_2, \dots, x_n) \in \Omega$  par  $\mathbf{P}(X) = p^j(1-p)^{n-j}$  où  $j$  est le nombre de  $P$  dans le  $n$ -uplet  $X$  et  $n-j$  le nombre de  $F$ . On a ainsi défini un espace probabilisé : connaissant la probabilité de chaque élément on connaît la probabilité de chaque partie de  $\Omega$ .

Répondons maintenant à la question. Notons  $A_k$  l'évènement « il y a exactement  $k$  piles ». Il y a  $\binom{n}{k}$  façons de choisir les  $k$  lancers faisant  $P$  parmi les  $n$  lancers, les autres lancers faisant  $F$ . Il y a donc  $\binom{n}{k}$   $n$ -uplets contenant exactement  $k$  fois  $P$ . Chacun de ces  $n$ -uplets a une probabilité  $p^k(1-p)^{n-k}$ . Ainsi  $A_k$  est une partie de  $\Omega$  ayant  $\binom{n}{k}$  éléments ayant tous la même probabilité. Donc la probabilité d'obtenir exactement  $k$  fois pile en  $n$  lancers est  $\mathbf{P}(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$ .

Remarque : chaque  $n$ -uplet de  $\Omega$  contient exactement  $k$  piles où  $k$  est un entier compris entre 0 et  $n$ . Ainsi  $\Omega = \bigsqcup_{k=0}^n A_k$  et  $\mathbf{P}(\Omega) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (1 + (1-p))^n = 1$  en utilisant la formule du binôme de Newton. Le résultat obtenu est cohérent !

## 1.2 Probabilités conditionnelles et événements indépendants

**Définition 1.2.1.** Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace probabilisé. Soient  $A$  et  $B$  des événements avec  $\mathbf{P}(B) \neq 0$ .

On appelle **probabilité conditionnelle** de  $A$  sachant  $B$  le nombre

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

**Remarque 1.2.1.** L'application  $A \mapsto \mathbf{P}_B(A)$  définit une nouvelle mesure de probabilité sur  $\Omega$  qui est portée par  $B$ . Et elle permet de définir une probabilité sur  $\mathcal{P}(B)$ . Cette mesure de probabilité est la restriction de la probabilité  $\mathbf{P}$  à  $B$  renormalisée par la mesure de  $B$ .

**Proposition 1.2.1.** Formule de Bayes :

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

$$\mathbf{P}(A) = \mathbf{P}(B)\mathbf{P}(A|B) + \mathbf{P}(\bar{B})\mathbf{P}(A|\bar{B}).$$

**Définition 1.2.2.** Deux événements  $A$  et  $B$  sont **indépendants** si

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B), \quad \text{ou encore si } \mathbf{P}_B(A) = \mathbf{P}(A).$$

On peut le reformuler ainsi : si  $A$  et  $B$  sont indépendants, le fait d'être dans  $B$  ou pas n'a pas d'influence sur la probabilité d'être dans  $A$  (et de même en inversant  $A$  et  $B$ ).

# Chapitre 2

## Variables aléatoires

### 2.1 Définitions

**Définition 2.1.1.** Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace probabilisé. On appelle **variable aléatoire** toute application  $X : \Omega \rightarrow \mathbf{R}$ .

Une telle variable permet de définir une mesure de probabilité  $\mathbf{P}_X$  sur  $\mathbf{R}$  par

$$\forall A \subset \mathbf{R}, \mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(X \in A).$$

**Remarque 2.1.1.** La plupart des distributions de probabilités sont définies sur des parties  $\Omega$  de  $\mathbf{R}$  et définissent naturellement une variable aléatoire par  $X(\omega) = \omega$ .

*Exemple : le résultat d'un lancer de dé.*

Une variable aléatoire est parfaitement définie par l'ensemble des valeurs qu'elle prend et leurs probabilités associées. On dispose de deux façons de représenter une variable aléatoire :

**Définition 2.1.2.** Soit  $X$  une variable aléatoire. On appelle **fonction de répartition** de  $X$  la fonction

$$\begin{aligned} F_X : \mathbf{R} &\rightarrow [0, 1] \\ x &\mapsto \mathbf{P}_X(-\infty, x] = \mathbf{P}(X \leq x). \end{aligned}$$

**Remarque 2.1.2.**  $F_X$  est une fonction croissante avec  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  et  $\lim_{x \rightarrow +\infty} F_X(x) = 1$ .

Si la variable  $X$  est discrète,  $F_X$  est constante par morceaux ; si  $X$  est continue,  $F_X$  est continue.

**Définition 2.1.3.** Soit  $X$  une variable discrète.

On appelle **fonction de densité** de  $X$  la fonction définie par

$$\begin{aligned} f_X : \mathbf{R} &\rightarrow [0, 1] \\ x &\mapsto \mathbf{P}_X(x). \end{aligned}$$

Soit  $X$  une variable continue.

La densité de  $X$  est la fonction positive  $f_X$  telle que  $\forall a, b \in \mathbf{R}$

$$\mathbf{P}_X([a, b]) = \int_a^b f_X(x) dx.$$

**Remarque 2.1.3.** Lorsque  $X$  est une variable continue, la probabilité de prendre une valeur précise est nulle :  $\mathbf{P}(X = x) = 0$ . Mais en terme de calcul infinitésimal, on peut interpréter la densité de  $X$  ainsi :  $\mathbf{P}(X = x) = f_X(x)dx$ .

**Proposition 2.1.1.** Si  $X$  est une variable continue, alors

$$F'_X = f_X.$$

## 2.2 Propriétés

### 2.2.1 Espérance

**Définition 2.2.1.** Soit  $X$  une variable aléatoire discrète à valeurs dans  $\{x_i; i \in I\}$ . On appelle **espérance** de  $X$  le nombre réel défini par

$$\mathbf{E}(X) = \sum_{i \in I} x_i \mathbf{P}(x_i)$$

Si  $X$  est continue, son espérance est définie par

$$\mathbf{E}(X) = \int_{\mathbf{R}} x f(x) dx,$$

où  $f$  est la densité de  $X$ .

Ce nombre n'est défini que si l'intégrale ou la somme sont définies et convergentes.

L'espérance représente la valeur moyenne de la variable  $X$ . C'est la moyenne des valeurs prises par  $X$  pondérées par leurs probabilités.

**Proposition 2.2.1.** Soient  $X$  et  $Y$  des variables aléatoires d'espérances finies et  $\lambda$  un réel. Alors

$$\mathbf{E}(X + Y) = \mathbf{E}(X) + \mathbf{E}(Y), \quad \mathbf{E}(\lambda X) = \lambda \mathbf{E}(X).$$

### Théorème 2.2.2. de transfert

Soit  $X$  une variable aléatoire de densité  $f$  et soit  $\varphi : \mathbf{R} \rightarrow \mathbf{R}$ . Alors la composée  $\varphi(X)$  est une variable aléatoire dont l'espérance est donnée par

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

### 2.2.2 Variance

**Définition 2.2.2.** Soit  $X$  une variable discrète d'espérance finie.

On appelle **variance** de  $X$  le nombre positif défini par

$$\mathbf{V}(X) = \sigma^2(X) = \sum_{i \in I} (x_i - \mathbf{E}(X))^2 \mathbf{P}(x_i).$$

Si  $X$  est continue de densité  $f$ , sa variance est donnée par

$$\mathbf{V}(X) = \sigma^2(X) = \int_{-\infty}^{+\infty} (x - \mathbf{E}(X))^2 f(x) dx.$$

Autrement dit,  $\mathbf{V}(X) = \mathbf{E}((X - \mathbf{E}(X))^2)$ .

On appelle **écart-type** de  $X$  la racine carrée de sa variance :  $\sigma(X) = \sqrt{\mathbf{V}(X)}$ .

L'écart-type mesure l'écart moyen quadratique entre les valeurs prises par  $X$  et sa moyenne.

**Proposition 2.2.3.** Soit  $X$  une variable admettant une variance et  $a$  et  $b$  des réels. Alors

$$\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2, \quad \mathbf{V}(aX + b) = a^2\mathbf{V}(X).$$

Si  $X$  et  $Y$  sont indépendantes (voir plus bas),

$$\mathbf{V}(X + Y) = \mathbf{V}(X) + \mathbf{V}(Y)$$

(et aussi  $\mathbf{V}(X - Y) = \mathbf{V}(X) + \mathbf{V}(Y)$ ).

### 2.2.3 Couples de variables

Soit  $(\Omega, \mathcal{F}, \mathbf{P})$  un espace probabilisé. Un **couple** de variables aléatoires est la donnée de deux variables définies sur  $\Omega$ .

**Définition 2.2.3.** Soient  $X$  et  $Y$  deux variables aléatoires. Elles sont **indépendantes** si,

$$\forall A, B \subset \mathbf{R}, \quad \mathbf{P}(X \in A \text{ et } Y \in B) = \mathbf{P}(X \in A) \times \mathbf{P}(Y \in B).$$

**Proposition 2.2.4.** Si deux variables continues  $X$  et  $Y$  de densités  $f_X$  et  $f_Y$  sont indépendantes, alors la loi du couple  $(X, Y)$  est donnée par la fonction  $(x, y) \mapsto f_X(x)f_Y(y)$ . Autrement dit, pour tout ensemble  $A$  de  $\mathbf{R}^2$ ,

$$\mathbf{P}(A) = \iint_A f_X(x)f_Y(y)dx dy.$$





## Chapitre 3

# Mesures de probabilités classiques

On dit que deux variables aléatoires  $X$  et  $Y$  suivent la même loi si elles ont la même fonction de densité. Cela ne signifie pas qu'elles prennent les mêmes valeurs simultanément mais que chaque valeur a la même probabilité d'apparaître pour  $X$  et pour  $Y$ .

Nous présentons dans ce chapitre les fonctions de densité (autrement dit les lois de probabilités) les plus classiques à partir desquelles on décrit la majorité des phénomènes aléatoires.

### 3.1 Lois discrètes

#### 3.1.1 Loi de Bernouilli

Soit  $p \in [0, 1]$ .

On dit qu'une variable aléatoire  $X$  suit la loi de Bernouilli de paramètre  $p$  si sa densité est donnée par

$$f(1) = p, \quad f(0) = 1 - p.$$

On note  $X \sim \mathcal{B}(p)$ .

On a

$$\mathbf{E}(X) = p, \quad \mathbf{V}(X) = p(1 - p).$$

La loi de Bernouilli est la loi permettant de décrire le jeu de pile ou face avec une pièce biaisée.

#### 3.1.2 Loi binomiale

Soit  $p \in [0, 1]$  et  $n \in \mathbf{N}^*$ .

On dit qu'une variable aléatoire  $X$  suit la loi binomiale de paramètres  $n$  et  $p$  si sa densité est donnée par

$$\forall k \in \{0, 1, \dots, n\}, \quad f(k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

On note  $X \sim \mathcal{B}(n, p)$ .

On a

$$\mathbf{E}(X) = np, \quad \mathbf{V}(X) = np(1 - p).$$

Si on lance une pièce biaisée  $n$  fois (les lancers sont indépendants), et si on note  $X$  le nombre de « pile » obtenus, alors  $X$  suit une loi binomiale.

De manière générale, on dit que la loi binomiale est la loi permettant de décrire le nombre de « succès » obtenus lors d'une expérience répétée.

Cette loi présente cependant un défaut : lorsque  $n$  est très grand, les coefficients binomiaux sont difficiles à calculer et la loi binomiale devient numériquement difficile à utiliser. On l'approche alors par d'autres lois de probabilité.

### 3.1.3 Loi géométrique

Soit  $p \in [0, 1]$ .

On dit qu'une variable aléatoire  $X$  suit la loi géométrique de paramètre  $p$  si sa densité est donnée par

$$\forall n \in \mathbf{N}^*, \quad f(n) = p(1 - p)^{n-1}.$$

On note  $X \sim \mathcal{G}(p)$ .

On a

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

Si on lance une pièce biaisée jusqu'à ce qu'on obtienne pile et si on note  $X$  le nombre de lancers effectués, alors  $X$  suit une loi géométrique.

De manière générale, on dit que la loi géométrique est la « loi du premier succès ».

### 3.1.4 Loi de Poisson

Soit  $\lambda \in \mathbf{R}_+^*$ .

On dit qu'une variable aléatoire  $X$  suit la loi de Poisson de paramètre  $\lambda$  si sa densité est donnée par

$$\forall n \in \mathbf{N}, \quad f(n) = e^{-\lambda} \frac{\lambda^n}{n!}.$$

On note  $X \sim \mathcal{P}(\lambda)$ .

On a

$$\mathbf{E}(X) = \lambda, \quad \mathbf{V}(X) = \lambda.$$

La loi de Poisson permet de décrire dans plusieurs contextes le nombre d'événements qui peuvent se produire durant un intervalle de temps donné. Exemples : le nombre d'utilisateurs arrivant à un guichet de la Poste entre 9h et 10h ; les nombres d'atomes radioactifs se désintégrant durant une minute.

La loi de Poisson est également appelée « loi des événements rares ». La raison est la suivante : si  $p$  est petit, la loi binomiale  $\mathcal{B}(n, p)$  est proche de la loi de Poisson  $\mathcal{P}(np)$ . En pratique, pour  $n$  grand et  $p$  petit, les termes de la densité de la loi binomiale sont difficiles à calculer et on peut les approcher par les termes de la densité de la loi de Poisson, plus faciles à utiliser.

## 3.2 Lois continues

### 3.2.1 Loi uniforme

Soient  $a$  et  $b$  dans  $\mathbf{R}$  avec  $a < b$ .

On dit qu'une variable aléatoire  $X$  suit la loi uniforme sur l'intervalle  $[a, b]$  si sa densité est donnée par

$$\forall x \in \mathbf{R}, \quad f(x) = \frac{1}{b-a} \mathbf{1}_{[a,b]}(x).$$

On note  $X \sim \mathcal{U}([a, b])$ .

On a

$$\mathbf{E}(X) = \frac{a+b}{2}, \quad \mathbf{V}(X) = \frac{(b-a)^2}{12}.$$

La fonction *rand* des logiciels de calcul permet de simuler des variables suivant des lois uniformes. Mais elle permet également de simuler n'importe quelle loi continue pour peu que l'on connaisse sa fonction de répartition :

**Proposition 3.2.1.** *Soit  $F$  une fonction de répartition continue. Soit  $X$  une variable aléatoire suivant la loi uniforme  $\mathcal{U}([0, 1])$ .*

*Soit  $Y = F^{-1}(X)$ . Alors  $Y$  suit la loi de probabilité de fonction de répartition  $F$ .*

Ainsi, si on veut simuler une variable aléatoire de fonction de répartition  $F$ , on tire des nombres  $x_1, \dots, x_n$  avec la fonction *rand*. Alors les nombres  $F^{-1}(x_1), \dots, F^{-1}(x_n)$  donnent un échantillon simulé avec la loi voulue.

### 3.2.2 Loi exponentielle

Soit  $\lambda \in \mathbf{R}_+^*$ .

On dit qu'une variable aléatoire  $X$  suit la loi exponentielle de paramètre  $\lambda$  si sa densité est donnée par

$$\forall x \in \mathbf{R}, \quad f(x) = e^{-\lambda x} \mathbf{1}_{\mathbf{R}_+}(x).$$

On note  $X \sim \mathcal{E}(\lambda)$ .

On a

$$\mathbf{E}(X) = \frac{1}{\lambda}, \quad \mathbf{V}(X) = \frac{1}{\lambda^2}.$$

La loi exponentielle est une loi sans mémoire :

$$\forall t' > t > 0, \quad \mathbf{P}(X > t' | X > t) = \mathbf{P}(X > t' - t).$$

Elle permet de modéliser la durée de vie de composants électroniques. Elle permet également de modéliser des temps d'attente. On peut modéliser par une loi exponentielle la variable  $T$  donnant le temps passé à un guichet de la Poste. On peut également modéliser par une loi exponentielle le temps séparant les arrivées à la Poste de deux usagers successifs. La loi exponentielle est naturellement liée à la loi de Poisson.

### 3.2.3 Loi gaussienne

Soient  $m \in \mathbf{R}$  et  $\sigma \in \mathbf{R}_+^*$ .

On dit qu'une variable aléatoire  $X$  suit la loi gaussienne de paramètres  $m$  et  $\sigma^2$  si sa densité est donnée par

$$\forall x \in \mathbf{R}, \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Cette loi est également appelée loi normale. On note  $X \sim \mathcal{N}(m, \sigma^2)$ .

On a

$$\mathbf{E}(X) = m, \quad \mathbf{V}(X) = \sigma^2.$$

Si  $m = 0$ , on parle de loi normale centrée et la loi  $\mathcal{N}(0, 1)$  est appelée loi normale centrée réduite.

On dispose de tables numériques pour la loi normale centrée réduite. On peut s'y ramener facilement :

$$\text{Si } X \sim \mathcal{N}(m, \sigma^2), \text{ alors } \frac{X - m}{\sigma} \sim \mathcal{N}(0, 1).$$

La loi normale intervient dans le théorème central limite. Pour résumer, ce théorème affirme que la loi d'une somme de variables indépendantes de même loi est proche (quand le nombre de variables aléatoires est grand) d'une loi normale. Par exemple, la loi binomiale  $\mathcal{B}(n, p)$  est très proche, pour  $n$  grand de la loi normale  $\mathcal{N}(np, np(1-p))$ .

### 3.2.4 Loi du khi 2

Soit  $n \in \mathbf{N}^*$ . Soient  $X_1, X_2, \dots, X_n$   $n$  variables indépendantes de même loi  $\mathcal{N}(0, 1)$ . Soit  $X = \sum_{i=1}^n X_i^2$ .  
On appelle loi du  $\chi^2$  à  $n$  degrés de liberté la loi de la variable  $X$ .

La densité de cette loi est définie par

$$f_X(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} t^{\frac{k}{2}-1} e^{-\frac{t}{2}}.$$

On dispose de tables numériques pour ces lois. On a

$$\mathbf{E}(X) = n, \quad \mathbf{V}(X) = 2n$$

La loi du  $\chi^2$  est très utilisée en statistiques. Elle permet d'estimer des variances, d'effectuer des tests d'adéquation à une loi ou encore des tests d'indépendance entre deux variables aléatoires.

# Chapitre 4

## Théorèmes limites

On effectue un grand nombre de tirages indépendants d'une même variable aléatoire. On souhaite décrire les comportements limites des tirages ainsi obtenus. Par exemple, lorsqu'on lance un grand nombre de fois une pièce, on s'attend à obtenir à peu près autant de « pile » que de « face ». Nous allons justifier cela et mesurer de manière probabiliste ce « à peu près autant ».

### 4.1 Théorèmes

#### **Théorème 4.1.1. Inégalités de Markov et de Bienaymé-Tchebychev**

Soit  $X$  une variable aléatoire à valeurs positives d'espérance finie. Soit  $a > 0$ . Alors

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}(X)}{a}.$$

Soit  $X$  une variable aléatoire ayant une espérance et une variance finies. Soit  $a > 0$ . Alors

$$\mathbf{P}(|X - \mathbf{E}(X)| > a) \leq \frac{\mathbf{V}(X)}{a^2}.$$

Autrement dit, la probabilité que  $X$  soit très éloignée de sa moyenne ne peut pas être très grande.

#### **Théorème 4.1.2. Loi faible des grands nombres**

Soit  $(X_k)_{k \in \mathbf{N}^*}$  une suite de variables aléatoires indépendantes et de même loi (i.i.d.). On les suppose d'espérance  $m$  et de variance  $\sigma^2$  finies. Alors, pour tout  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow +\infty} \mathbf{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - m\right| > \varepsilon\right) = 0.$$

#### **Théorème 4.1.3. Loi forte des grands nombres**

Soit  $(X_k)_{k \in \mathbf{N}^*}$  une suite de variables aléatoires indépendantes et de même loi. Soit  $m$  leur espérance. Alors

$$\lim_{n \rightarrow +\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = m \quad \text{presque sûrement.}$$

Précisément, cela signifie

$$\mathbf{P}\left(\lim_{n \rightarrow +\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = m\right) = 1.$$

**Théorème 4.1.4. Théorème central limite**

Soit  $(X_k)_{k \in \mathbf{N}^*}$  une suite de variables aléatoires i.i.d. d'espérance  $m$  et de variance  $\sigma^2$  finies. Notons  $S_n = X_1 + X_2 + \dots + X_n$ .

Quand  $n$  tend vers  $+\infty$ , la suite  $\frac{S_n - nm}{\sigma\sqrt{n}}$  converge en loi vers la loi normale  $\mathcal{N}(0, 1)$ .

Reformulations : le théorème signifie qu'à la limite, les probabilités sont données par la densité  $f_{\mathcal{N}}$  de la loi normale :

$$\lim_{n \rightarrow +\infty} \mathbf{P} \left( a \leq \frac{S_n - nm}{\sigma\sqrt{n}} \leq b \right) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Ou encore :

$$\lim_{n \rightarrow +\infty} \mathbf{P} \left( \frac{S_n - nm}{\sigma\sqrt{n}} \leq \alpha \right) = F_{\mathcal{N}}(\alpha),$$

où  $F_{\mathcal{N}}$  désigne la fonction de répartition de la loi normale  $\mathcal{N}(0, 1)$ . La table de la loi normale donne les valeurs de  $F_{\mathcal{N}}$ .

La loi forte des grands nombres signifie que la suite des moyennes de tirages aléatoires indépendants d'une même loi converge vers la moyenne théorique de cette loi. Le théorème central limite permet de mesurer (avec des probabilités) l'écart entre les termes de la suite et cette limite.

Utilisation du théorème : en pratique, quand  $n$  est assez grand ( $n \geq 30$ ), le théorème permet d'approcher la loi de  $S_n$  par une loi normale :

$$S_n \sim \mathcal{N}(nm, n\sigma^2) \quad \text{et} \quad \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\frac{S_n}{n} - m}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0, 1).$$

Par exemple, la loi binomiale étant une somme de Bernoulli indépendantes, on peut approcher pour  $n$  grand, cette loi par une loi normale :

$$\mathcal{B}(n, p) \sim \mathcal{N}(np, np(1-p)).$$

Ainsi, si  $S \sim \mathcal{B}(100, \frac{1}{2})$ , on peut approcher la loi de  $S$  par  $S \sim \mathcal{N}(50, 25)$  et on obtient alors  $\frac{S-50}{5} \sim \mathcal{N}(0, 1)$ . Alors

$$\begin{aligned} \mathbf{P}(40 \leq S \leq 60) &= \mathbf{P} \left( \frac{40-50}{5} \leq \frac{S-50}{5} \leq \frac{60-50}{5} \right) \\ &\approx F_{\mathcal{N}}(2) - F_{\mathcal{N}}(-2) = 2F_{\mathcal{N}}(2) - 1 = 0,9545. \end{aligned}$$

La valeur exacte est  $\sum_{k=40}^{60} \frac{1}{2^{100}} \binom{100}{k}$ ; elle est plus difficile à calculer.

## 4.2 Applications

Parmi les applications classiques de la loi forte des grands nombres, citons la méthode de Monte Carlo pour le calcul numérique d'intégrale.

**Théorème 4.2.1.** Soit  $f$  une fonction intégrable définie sur un intervalle  $[a, b]$ . Soient  $x_1, x_2, \dots, x_n, \dots$  des tirages indépendants de la loi uniforme sur  $[a, b]$ .

Alors la suite  $\frac{1}{n} \sum_{i=1}^n f(x_i)$  converge presque sûrement vers  $\frac{1}{b-a} \int_a^b f(x) dx$ .

Le théorème central limite est le résultat théorique qui permet en statistiques de définir la notion d'intervalle de confiance. L'application la plus usuelle est sans doute le sondage.