

# MATHF-105 : Probabilités

## Résumé

R. Petit

Année académique 2015 - 2016

### Contents

<b>1</b>	<b>Rappels</b>	<b>1</b>
1.1	Rappel sur les séries . . . . .	1
1.1.1	Exemple sur les séries . . . . .	1
1.1.2	Conclusion de la suite géométrique . . . . .	1
1.2	Rappels d'analyse . . . . .	1
<b>2</b>	<b>Espaces de probabilités</b>	<b>3</b>
2.1	Définition . . . . .	3
2.1.1	Loi uniforme sur un ensemble fini (ou dénombrable) . . . . .	3
2.1.2	Loi uniforme sur un ensemble infini (intervalle) . . . . .	4
2.2	Modèles . . . . .	5
2.2.1	Modèles discrets . . . . .	5
2.2.2	Modèles continus (à densité) . . . . .	6
2.2.3	Divergence sur la fonction Gamma d'Euler . . . . .	7
2.2.4	Retour aux modèles stochastiques . . . . .	8
2.3	Notion de variables aléatoires . . . . .	8
2.3.1	Cas discret . . . . .	8
2.3.2	Cas absolument continu . . . . .	9
2.4	Théorème de de Moivre-Laplace . . . . .	10
2.5	Convergence en loi . . . . .	11
<b>3</b>	<b>Espérance</b>	<b>12</b>
3.1	Pari de pascal . . . . .	12
3.2	Espérance et variables aléatoires . . . . .	12
3.3	Définition de l'espérance . . . . .	12
3.3.1	Cas positif . . . . .	12
3.3.2	Cas général . . . . .	13
3.4	Exemples d'espérance . . . . .	14
3.5	Espérance de fonctions de variables aléatoires . . . . .	15
3.6	Variance . . . . .	16
3.6.1	Définitions . . . . .	16
3.7	Moments de variables aléatoires . . . . .	18
3.7.1	Cas discret . . . . .	18
3.7.2	Cas absolument continu . . . . .	19
3.8	Fonctions génératrices . . . . .	19
3.8.1	Cas discret . . . . .	19

3.8.2	Lien entre moments entiers et fonctions génératrices . . . . .	20
3.8.3	Cas absolument continu . . . . .	21
3.8.4	Aperçu de la convergence en loi . . . . .	22
<b>4</b>	<b>Conditionnement et indépendance</b>	<b>26</b>
4.1	Événements indépendants . . . . .	26
4.2	Variables aléatoires indépendantes . . . . .	27
4.3	Formules de convolution . . . . .	27
4.3.1	Cas discret . . . . .	27
4.3.2	Cas absolument continu . . . . .	29

# 1 Rappels

## 1.1 Rappel sur les séries

Les fonctions logarithmique et exponentielle ont un développement de Taylor exact. Pour la fonction logarithmique, on a, pour  $x \in (-1, 1)$  :

$$\log(1-x) = - \sum_{k \geq 1} \frac{x^k}{k}.$$

Si on pose  $S_n := \sum_{k=1}^n u_k$ , on a  $(S_n)_{n \in \mathbb{N}}$ , la suite des sommes partielles, et  $n \mapsto S_n$ , une application croissante si  $(u_n)$  est une suite positive. Il y a donc deux situations distinctes possibles :

- $(S_n)$  est une suite bornée ( $\exists M \in \mathbb{R} \text{ t. q. } \forall n \in \mathbb{N} : S_n \leq M$ ) et donc converge vers  $S \in \mathbb{R}$  ;
- $(S_n)$  n'est pas bornée ( $\forall M \in \mathbb{R} : \exists n \in \mathbb{N} \text{ t. q. } S_n > M$ ) et donc diverge vers  $+\infty$ .

### 1.1.1 Exemple sur les séries

Prenons  $u_n := x^n$ , avec  $x > 0$ .

- Si  $x = 1$ , on a  $n \rightarrow +\infty \Rightarrow S_n \rightarrow +\infty$  ;
- si  $x \neq 1$ , on a  $(1-x)S_n = x - x^{n+1}$ , et donc :

$$S_n := x \frac{1-x^{n+1}}{1-x}.$$

- Si  $x < 1$ , alors  $x^n \rightarrow 0$  pour  $n \rightarrow +\infty$ , et donc  $S_n \rightarrow \frac{x}{1-x}$  ;
- si  $x > 1$ , alors  $x^n \rightarrow +\infty$  pour  $n \rightarrow +\infty$ , et donc  $S_n \rightarrow +\infty$ .

### 1.1.2 Conclusion de la suite géométrique

On voit alors :

$$\sum_{n \geq 1} x^n = \begin{cases} \frac{x}{1-x} & \text{si } x \in [0, 1) \\ +\infty & \text{sinon} \end{cases}.$$

Si la suite commence à l'indice 0, on a :

$$\sum_{n \geq 0} x^n = 1 + \sum_{n \geq 1} x^n = \begin{cases} 1 + \frac{x}{1-x} = \frac{1}{1-x} & \text{si } x \in [0, 1) \\ +\infty & \text{sinon} \end{cases}.$$

## 1.2 Rappels d'analyse

**Définition 1.1.** Une fonction  $f : X \rightarrow Y$  est dite mesurable si :

$$\forall A \subset \mathcal{B}(Y) : \{\omega \in \Omega \text{ t. q. } X(\omega) \in A\} \in \mathcal{F},$$

où  $\mathcal{B}(Y)$  représente la tribu des boréliens (voir définition 2.9).

**Théorème 1.2.** Dans  $\mathbb{R}$ , toute série absolument convergente est convergente.

**Théorème 1.3.** *Dans  $\mathbb{R}$ , toute intégrale impropre absolument convergente est convergente.*

## 2 Espaces de probabilités

### 2.1 Définition

**Définition 2.1.** L'ensemble  $\Omega$  est l'**espace des chances**, l'ensemble des résultats possibles d'un phénomène aléatoire.

*Remarque.*

- $\Omega$  peut être fini (dénombrable) ou infini ;
- $\Omega = \{0, 1\}^{\mathbb{N}}$  est l'ensemble des suites à valeur dans  $\{0, 1\}$  ;
- $\Omega$  peut être un espace dit *fonctionnel* quand le résultat d'une expérience est une fonction.

**Définition 2.2.** Un événement  $E$  est un ensemble de réalisations possibles à une expérience tel que  $E \subseteq \Omega$ .

*Remarque.* L'ensemble  $\mathcal{P}(\Omega)$  n'est pas toujours dénombrable. Et donc l'ensemble  $\mathcal{P}(\Omega)$  est-il le bon ensemble pour décrire les événements ?

- Si  $|\Omega| \in \mathbb{N}$  : oui ;
- si  $|\Omega| \notin \mathbb{N}$  : non.

**Définition 2.3.**  $\mathcal{F}$  est la **classe des événements**. On mesure la *probabilité d'occurrence* d'un événement  $A \in \mathcal{F}$ . On introduit une fonction d'ensemble  $\mathbb{P}$  où :

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1] : A \mapsto \mathbb{P}(A).$$

On impose :

- (i)  $\mathbb{P}(\emptyset) = 0$  ;
- (ii)  $\mathbb{P}(\Omega) = 1$  ;
- (iii)  $\forall A, B \in \mathcal{F} : A \cap B = \emptyset \Rightarrow \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .

**Proposition 2.4.** Soient  $A_1, \dots, A_n \in \mathcal{F}$ . On a :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} \mathbb{P}\left(\bigcap_{\gamma=1}^i A_{k_\gamma}\right).$$

#### 2.1.1 Loi uniforme sur un ensemble fini (ou dénombrable)

**Définition 2.5.** Soient  $m < n \in \mathbb{N}$ . On définit l'**intervalle entier**  $\llbracket m, n \rrbracket$  par :

$$\llbracket m, n \rrbracket : \{x \in \mathbb{N} \text{ t. q. } m \leq x \leq n\}.$$

**Définition 2.6.** Soit  $\Omega = \llbracket 1, n \rrbracket$ . Soit  $A \subseteq \Omega$ . La loi uniforme est donnée par :

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|A|}{n}.$$

*Remarque.* Il arrive que  $|A|$  soit difficile à déterminer et qu'il faille aller chercher du côté de l'analyse combinatoire.

### 2.1.2 Loi uniforme sur un ensemble infini (intervalle)

**Définition 2.7.** Soit  $\Omega = [0, 1]$  et soit  $A = [a, b] \subseteq \Omega$ . La loi uniforme est donnée par :

$$\mathbb{P}(A) = (b - a).$$

*Remarque.* La définition de loi uniforme sur un intervalle fait intervenir la notion de mesure et donc de mesurabilité. Or il existe des parties de  $\Omega$  sur lesquelles la mesure n'a pas de sens. En général,  $\mathcal{P}(\Omega)$  est *trop grand*, et il faut donc remplacer l'utilisation de l'ensemble des parties par la notion de tribu.

**Définition 2.8.** Soit  $\Omega$  un ensemble de chances et  $\mathcal{F} \subseteq \mathcal{P}(\Omega)$  une famille de parties de  $\Omega$ . On dit que  $\mathcal{F}$  est une tribu s'il respecte les trois propriétés suivantes :

- $\emptyset \in \mathcal{F}$  ;
- $\forall A : A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$  ;
- $\forall A_1, \dots, A_n, \dots : A_1, \dots, A_n, \dots \in \mathcal{F} \Rightarrow \bigcup_{k \geq 1} A_k \in \mathcal{F}$ .

Une autre appellation pour une tribu est une  $\sigma$ -algèbre.

*Remarque.*

- On remarque que  $\mathcal{P}(\Omega)$  est une tribu, mais une tribu trop grande pour être intéressante ;
- Soit  $A \in \mathcal{P}(\Omega)$ . Alors  $\mathcal{T} := \{\emptyset, A, A^c, \mathcal{P}(\Omega)\}$  est une tribu.  $\mathcal{T}$  est la plus petite tribu contenant  $A$ , et on l'appelle la **tribu engendrée par  $A$** , que l'on note  $\sigma(A)$ .

**Définition 2.9.** Soit  $\mathcal{I}$  une partie de  $\mathcal{P}(\Omega)$ . On appelle la *tribu engendrée par  $\mathcal{I}$*  la plus petite tribu contenant  $\mathcal{I}$  et on la note  $\sigma(\mathcal{I})$ .

En prenant  $\mathcal{I} := \{\text{intervalles ouverts de } [0, 1]\}$ , on obtient  $\sigma(\mathcal{I})$  que l'on appelle **tribu des boréliens**.<sup>1</sup>

**Définition 2.10.** Soit  $\Omega$  un ensemble de chances et  $\mathcal{F} \subset \mathcal{P}(\Omega)$  une tribu sur  $\Omega$ . Une probabilité sur  $(\Omega, \mathcal{F}, \mathbb{P})$  est une fonction  $\mathbb{P}$  définie par :

$$\mathbb{P} : \mathcal{F} \rightarrow [0, 1] : A \mapsto \mathbb{P}(A),$$

où  $\mathbb{P}$  satisfait :

- (i)  $\mathbb{P}(\emptyset) = 0$  ;
- (ii)  $\forall A \in \mathcal{F} : \mathbb{P}(A) + \mathbb{P}(A^c) = 1$  ;
- (iii)  $\forall A_1, \dots, A_n, \dots$  disjoints deux à deux, on a :

$$\mathbb{P}\left(\bigcup_{k \geq 1} A_k\right) = \sum_{k \geq 1} \mathbb{P}(A_k).$$

**Définition 2.11.** On appelle  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilités.

*Remarque.* Probabiliser une expérience revient à déterminer :

- $\Omega$ , l'espace des chances ;
- $\mathcal{F}$ , la classe des événements ;
- $\mathbb{P}$ , la fonction d'ensembles sur  $\mathcal{F}$ .

---

<sup>1</sup>Le nom de *borélien* vient du mathématicien français Émile Borel suite à ses travaux sur la théorie de la mesure.

## 2.2 Modèles

### 2.2.1 Modèles discrets

*Remarque.* On prend  $\Omega$  un ensemble fini ou dénombrable. On prend également  $\mathcal{F} = \mathcal{P}(\Omega)$ .

Si  $\Omega$  est fini, on parle de tirages, et si  $\Omega$  est infini dénombrable, on parle de populations.

On pose :

$$\mathbb{P} : \{k\} \mapsto p_k \in [0, 1],$$

où :

$$\sum_{k \in \Omega} p_k = 1$$

et pour  $A = \{k_1, \dots, k_n\} \in \mathcal{F}$  :

$$\mathbb{P}(A) = \sum_{\gamma=1}^n p_{k_\gamma}.$$

**Définition 2.12** (Modèle de Bernoulli). On prend  $\Omega = \{0, 1\}$  où :

$$\begin{cases} p_0 &= 1 - p \\ p_1 &= p \end{cases}.$$

*Remarque.* Il est évident que  $p + (1 - p) = 1 = \mathbb{P}(\Omega)$ .

**Définition 2.13** (Modèle binomial). On prend  $\Omega = \llbracket 0, N \rrbracket$  (et donc  $\mathcal{F} = \mathcal{P}(\Omega)$ ) et  $p \in [0, 1]$ . Le modèle binomial est défini par  $p_k = \binom{N}{k} p^k (1 - p)^{N-k}$  pour tout  $k \in \llbracket 0, N \rrbracket$ .

*Remarque.* On remarque que  $\sum_{k \geq 1} p_k = 1$  car les  $p_k$  représentent les termes du binôme de Newton  $(p + (1 - p))^N = 1^N = 1$ .

**Définition 2.14** (Modèle géométrique). On prend  $\Omega = \mathbb{N}$ ,  $\mathcal{F} = \mathcal{F}(\Omega) \simeq \mathbb{R}$ , et  $p \in (0, 1)$ . Le modèle géométrique est défini par  $p_k = (1 - p)^{k-1} p$  pour tout  $k \in \mathbb{N}$ .

*Remarque.* On remarque que :

$$\sum_{k \geq 1} p_k = \sum_{k \geq 1} p(1 - p)^{k-1} = p \sum_{k \geq 0} (1 - p)^k = p \frac{1}{1 - (1 - p)} = \frac{p}{p} = 1,$$

où on utilise la formule de la somme des termes d'une suite géométrique  $u$  définie par  $u_n = u_{n-1}q$  pour  $n \geq 1$  (avec  $0 < q < 1$ ) qui donne :

$$\sum_{k=0}^N u_k = u_0 \frac{1 - q^{N+1}}{1 - q},$$

et pour la série, il suffit de passer à la limite :

$$\lim_{N \rightarrow +\infty} \sum_{k=0}^N u_k = \lim_{N \rightarrow +\infty} u_0 \frac{1 - q^{N+1}}{1 - q} = u_0 \frac{1}{1 - q}.$$

**Définition 2.15** (Modèle de Poisson). On prend  $\Omega = \mathbb{N}$ ,  $\mathcal{F} = \mathcal{P}(\Omega)$ , et un paramètre  $\lambda \in \mathbb{R}_0^+$ . Le modèle poissonien est défini par  $p_k = \exp(-\lambda) \frac{\lambda^k}{k!}$  pour tout  $k \in \mathbb{N}$ .

*Remarque.* On remarque que  $\mathbb{P}(\Omega) = 1$  en utilisant la formule de Taylor de l'exponentielle :

$$\exp(x) = \sum_{k \geq 0} \frac{x^k}{k!}.$$

On a effectivement :

$$\mathbb{P}(\Omega) = \sum_{k \geq 0} \mathbb{P}(\{k\}) = \sum_{k \geq 0} p_k = \sum_{k \geq 0} \exp(-\lambda) \frac{\lambda^k}{k!} = \exp(-\lambda) \exp(\lambda) = 1.$$

## 2.2.2 Modèles continus (à densité)

*Remarque.* On prend  $\Omega$  un intervalle (fini ou infini<sup>2</sup>) sur  $\mathbb{R}$ , et  $\mathcal{F} = \mathcal{B}(I)$ , la tribu des boréliens sur  $I^3$ .

**Définition 2.16.** Soit  $f : I \rightarrow \mathbb{R}^+$  une fonction intégrable telle que  $\int_{\mathbb{R}} f(x) dx = 1$ . Soit  $A \in \mathcal{F}$ , on pose  $\mathbb{P}(A) = \int_A f(x) dx$ .  $f$  est appelée fonction de densité de modèle stochastique.

**Définition 2.17** (Loi uniforme continue). On prend  $I = [a, b]$  avec  $a < b \in \mathbb{R}$ . Le modèle uniforme est défini par  $f$  constante :

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

*Remarque.* On remarque effectivement  $\int_{\mathbb{R}} f(x) dx = 1$  :

$$\int_{\mathbb{R}} f(x) dx = \int_{-\infty}^a f(x) dx + \int_a^b f(x) dx + \int_b^{+\infty} f(x) dx = 0 + \frac{1}{b-a} \int_a^b dx + 0 = 1.$$

**Définition 2.18** (Modèle exponentiel). <sup>4</sup> On prend  $I = \mathbb{R}^+$  et  $\lambda > 0$ . Le modèle exponentiel est défini par :

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}.$$

*Remarque.* On peut calculer l'intégrale impropre comme suit :

$$\begin{aligned} \int_{\mathbb{R}} f(x) dx &= \int_{-\infty}^0 f(x) dx + \int_0^{+\infty} f(x) dx = 0 + \lim_{M \rightarrow +\infty} \int_0^M f(x) dx \\ &= \lim_{M \rightarrow +\infty} [-\exp(-\lambda x)]_0^M = \lim_{M \rightarrow +\infty} (1 - \exp(-\lambda M)) = 1. \end{aligned}$$

**Définition 2.19** (Modèle gaussien). <sup>5</sup> On prend  $I = \mathbb{R}$ , et  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_0^+$ . Le modèle gaussien est défini par :

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

<sup>2</sup>On parle d'intervalle fini pour  $[a, b]$ , avec  $a < b \in \mathbb{R}$  et d'intervalle semi-infini pour  $(-\infty, b]$  ou  $[a, +\infty)$  et d'intervalle infini pour  $(-\infty, +\infty) = \mathbb{R}$ .

<sup>3</sup>Ou encore la tribu engendrée par les intervalles de  $I$ .

<sup>4</sup>Également appelé *modèle des files d'attente*.

<sup>5</sup>Également appelé *modèle des erreurs* ou encore *modèle normal*.



*Remarque.* Pour que  $\mathbb{P}$  soit une probabilité, il faut que  $f$  soit définie positive. Or  $f$  est une exponentielle multipliée par un coefficient positif. Il faut également  $\int_{\mathbb{R}} f(x) dx = 1$ , ce qui peut se vérifier par :

$$\int_{\mathbb{R}} f(x) dx,$$

en posant  $y := x - \mu$ , et donc  $dy = dx$  :

$$\int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2}\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{y^2}{2\sigma^2}\right).$$

En posant  $z := \frac{y}{\sigma}$  (et donc  $dz = \frac{dx}{\sigma}$ ), on obtient :

$$\int_{\mathbb{R}} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{z^2}{2}\right) dz.$$

Une primitive de  $\exp\left(-\frac{z^2}{2}\right)$  est :

$$\int_{-\infty}^z \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}} = \text{Erf}(z).$$

On écrit alors :

$$\begin{aligned} \mathbb{P}(\Omega)^2 &= \left( \int_{\mathbb{R}} \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}} \right) \left( \int_{\mathbb{R}} \exp\left(-\frac{y^2}{2}\right) \frac{dy}{\sqrt{2\pi}} \right) \\ &= \iint_{\mathbb{R}^2} \exp\left(-\frac{x^2 + y^2}{2}\right) \frac{dx dy}{2\pi}. \end{aligned}$$

En passant en coordonnées polaires, on obtient :

$$\mathbb{P}(\Omega)^2 = \int_{-\pi}^{+\pi} \int_{\mathbb{R}} \exp\left(-\frac{r^2}{2}\right) \frac{r dr d\theta}{2\pi} = \int_{-\pi}^{+\pi} \frac{d\theta}{2\pi} \int_{\mathbb{R}} r \exp\left(-\frac{r^2}{2}\right) dr = \left[ -\exp\left(-\frac{r^2}{2}\right) \right]_0^{+\infty} = 1.$$

On en déduit alors  $\mathbb{P}(\Omega) = 1$  également.  $\mathbb{P}$  est donc bien une probabilité.

**Définition 2.20.** On a défini une probabilité sur  $(\mathbb{R}^+, (\mathbb{R}^+))$  via la fonction  $f(r) = r \exp\left(-\frac{r^2}{2}\right)$ . On l'appelle la *probabilité de Rayleigh*.

### 2.2.3 Divergence sur la fonction Gamma d'Euler

**Définition 2.21** (Fonction Gamma d'Euler). La fonction Gamma d'Euler est définie comme suit :

$$\Gamma : \mathbb{R}_0^+ \rightarrow \mathbb{R} : x \mapsto \int_0^{+\infty} \exp(-x) x^{t-1} dx.$$

*Remarque.* On note  $\gamma := -\Gamma'(1) > 0$  la constante d'Euler-Mascheroni. La question  $\gamma \stackrel{?}{\in} \mathbb{Q}$  est toujours ouverte.

**Proposition 2.22.**  $\forall t > 0 : \Gamma(t + 1) = t\Gamma(t)$ .

*Démonstration.* Soit  $t > 0$ . Par l'intégration par parties, on a :

$$\Gamma(t + 1) = \int_0^{+\infty} \exp(-x)x^t dx = [-x^t \exp(-x)]_0^{+\infty} + t \int_0^{+\infty} \exp(-x)x^{t-1} dx = t\Gamma(t).$$

□

*Remarque.* Par la proposition 2.22, on peut définir la factorielle de tout nombre naturel par :

$$\forall n \in \mathbb{N}^* : n! = \Gamma(n + 1)$$

**Proposition 2.23** (Formule des compléments). Soit  $t \in (0, 1)$ . Alors :

$$\Gamma(t)\Gamma(1 - t) = \frac{\pi}{\sin(\pi t)}.$$

## 2.2.4 Retour aux modèles stochastiques

**Définition 2.24** (Modèle Gamma). <sup>6</sup> On prend  $\Omega = \mathbb{R}^+$ . Le modèle Gamma est défini par :

$$f_t(x) = \frac{x^t - \exp(-x)}{\Gamma(t)}.$$

## 2.3 Notion de variables aléatoires

### 2.3.1 Cas discret

**Définition 2.25.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité. Une variable aléatoire discrète<sup>7</sup> est une application  $X : \Omega \rightarrow E$  où  $E$  est un ensemble fini ou infini dénombrable. On demande à cette application d'être mesurable.

*Remarque.*

- Bien souvent, on a  $E = \Omega$ , et  $X(\omega) = \omega$ . Dans ce cas, on *identifie* l'espace des chances avec l'espace d'arrivée. La probabilité  $\mathbb{P}$  s'appelle alors la **loi** de la variable aléatoire  $X$ .
- Il arrive parfois que l'espace de probabilités soit plus gros que l'espace d'état.

**Définition 2.26.** Plus formellement, la **loi** d'une v.a.d.  $X$  est l'ensemble :

$$\{\mathbb{P}(X = x) \text{ t. q. } x \in E\}.$$

**Définition 2.27.** Pour toute valeur  $k \in E$  que peut prendre la variable aléatoire  $X$ , on note  $\mathbb{P}(X = k)$  la probabilité que la variable  $X$  prenne la valeur  $k$ . C'est équivalent à  $\mathbb{P}(X(\omega) = k)$  pour  $\omega \in \Omega$ .

**Définition 2.28.** Lorsqu'une v.a.d.  $X$  suit une certaine loi  $\mathcal{L}$ , on note  $X \sim \mathcal{L}$ .

Par exemple, une variable  $Y$  suivant une poisson de paramètre  $\lambda$  se note  $Y \sim \mathcal{P}(\lambda)$ .

<sup>6</sup>Le modèle  $\Gamma$  est une généralisation du modèle exponentiel (définition 2.18).

<sup>7</sup>Souvent écrite v.a.d. ou V.A.-D.

### 2.3.2 Cas absolument continu

**Définition 2.29.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité. Une variable aléatoire absolument continue<sup>8</sup> est une application  $X : \Omega \rightarrow \mathbb{R}$  mesurable au sens où :

$$\forall A \in \mathcal{B}(\mathbb{R}) : \{\omega \in \Omega \text{ t.q. } X(\omega) \in A\} \in \mathcal{F},$$

et *absolument continue* au sens où :

$$\exists f_X : \mathbb{R} \rightarrow \mathbb{R}^+$$

mesurable et telle que :

$$\int_{\mathbb{R}} f_X(x) dx = 1,$$

avec :

$$\mathbb{P}(X \in A) = \int_A f_X(x) dx. \quad (1)$$

**Définition 2.30.** On appelle  $f_X$  la **densité** de  $X$ .

*Remarque.* La loi de  $X$  est donnée par (1).

**Définition 2.31.** On note  $F_X(t) = \mathbb{P}(X \leq t)$ , ou encore  $F_X(t) = \int_{-\infty}^t f(x) dx$  (en prenant  $A = (-\infty, t]$ ).

*Remarque.* La fonction  $t \mapsto F_X(t)$  est continue et est (presque) partout dérivable avec :

$$\frac{\partial F_X}{\partial t}(t) = f_X(t) \geq 0.$$

Donc  $F_X$  est croissante avec :

$$\lim_{t \rightarrow -\infty} F_X(t) = 0,$$

et :

$$\lim_{t \rightarrow +\infty} F_X(t) = 1.$$

*Remarque.* On peut associer une fonction de répartition  $F_X$  à toute variable aléatoire  $X$ , même si  $X$  est une v.a.d. Dans ce cas, on construit  $F_X$  constante par morceaux (et présente donc des points de discontinuité).

**Définition 2.32.** Si  $F_X$  est continue, on dit que  $X$  est continue.

*Remarque.* Donc si  $X$  est continue, alors  $\mathbb{P}(X = x) = F_X(x) - \lim_{y \rightarrow x} F_X(y) = 0$ . Ce résultat peut également être observé en utilisant le fait que  $\mathbb{P}(X = x) = \int_x^x f(x) dx$ , et une intégration sur un point est nulle.

*Remarque.* Il existe des fonction continues nulle part dérivables. On peut donc avoir  $F_X(t)$  continue mais pas sous la forme suivante :

$$F_X(t) = \int_{-\infty}^t f(x) dx, \quad (2)$$

pour une fonction  $f_X$  donnée.

**Définition 2.33.** On dit qu'une variable fonction  $f : \mathbb{R} \rightarrow \mathbb{R}$  est **absolument continue** si elle admet une représentation intégrale de type (2).

**Définition 2.34.** Soit  $E$  un ensemble. La fonction  $1_E$  est appelée **fonction indicatrice** est est définie telle que :

$$\forall x : 1_E(x) = \begin{cases} 1 & \text{si } x \in E \\ 0 & \text{sinon} \end{cases}.$$

---

<sup>8</sup>Souvent écrite v.a.c. ou V.A.-C.

## Exemples

1. Si  $X_1 \sim \mathcal{U}_{[a,b]}$  est une v.a.c. uniforme sur  $[a, b]$ , alors :

$$F_{X_1}(t) = \begin{cases} 0 & \text{si } t \leq a \\ t - a & \text{si } a < t < b \\ 1 & \text{si } t > b \end{cases}.$$

2. Si  $X_2 \sim \text{Exp}(\lambda)$  est une v.a.c. exponentielle de paramètre  $\lambda$ , alors :

$$F_{X_2}(t) = \int_{-\infty}^t \lambda \exp(-\lambda x) 1_{(0,+\infty)}(x) dx = 1 - \exp(-\lambda t) 1_{(0,+\infty)}(t).$$

3. Si  $X_3 \sim \mathcal{N}(\mu, \sigma^2)$  est une v.a.c. normale de moyenne  $\mu$  et de variance  $\sigma^2$ , alors :

$$F_{X_3}(t) = \int_{-\infty}^t f(x) dx = \text{Erf} \left( \frac{t - \mu}{\sigma} \right).$$

4. Si  $X_4 \sim \mathcal{C}$  est une v.a.c. de Cauchy de densité donnée par :

$$f_{X_4}(x) = \frac{1}{\pi(1+x^2)},$$

alors :

$$F_{X_4}(t) = \frac{1}{2} + \frac{1}{\pi} \arctan(t).$$

## 2.4 Théorème de de Moivre-Laplace

Soient  $p \in (0, 1)$  et  $n \geq 1$ . On pose  $X_{n,p} \sim \mathcal{B}(n, p)$ .

Soit  $Y_{n,p}$  défini par :

$$Y_{n,p} := \frac{X_{n,p} - np}{\sqrt{np(1-p)}}.$$

On remarque que  $Y_{n,p}$  est une binomiale renormalisée.

**Théorème 2.35** (Théorème de de Moivre-Laplace). *Si  $t \in \mathbb{R}$ , alors :*

$$\mathbb{P}(Y_{n,p} \leq t) \xrightarrow{n \rightarrow +\infty} F_{\mathcal{N}(0,1)}(t).$$

*Remarque.* La signification de ce théorème est qu'une binomiale renormalisée se comporte comme une gaussienne  $\mathcal{N}(0, 1)$  lorsque  $n \rightarrow +\infty$ .

**Proposition 2.36** (Formule de Stirling).

$$n! \sim \sqrt{2\pi n} \left( \frac{n}{e} \right)^n.$$

## 2.5 Convergence en loi

**Définition 2.37.**

- Soit  $Z$  une v.a.c. Soit  $\{Z_n, n \geq 1\}$  une suite de v.a. quelconques. On dit que  $Z_n$  *converge en loi vers*  $Z$  si :

$$\forall x : F_{Z_n}(x) = \mathbb{P}(Z_n \leq x) \xrightarrow{n \rightarrow +\infty} F_Z(x).$$

On note cela :

$$Z_n \xrightarrow{\mathcal{D}} Z.$$

- Soient  $Z$  une v.a.d. et une  $\{Z_n, n \geq 1\}$  une suite de variables aléatoires discrètes. On dit que  $Z_n$  *converge en loi vers*  $Z$  si :

$$\forall x \in E : \mathbb{P}(Z_n = x) \xrightarrow{n \rightarrow +\infty} \mathbb{P}(Z = x).$$

On note cela :

$$Z_n \xrightarrow{\mathcal{D}} Z.$$

*Remarque.* Un exemple typique de convergence en loi de variables discrètes est :

$$\mathcal{B}\left(n, \frac{\lambda}{n}\right) \xrightarrow{\mathcal{D}} \mathcal{P}(\lambda).$$

## 3 Espérance

### 3.1 Pari de pascal

Le terme *espérance* vient de Blaise Pascal et de son traitement de la question « Faut-il croire en Dieu ? ». On pose la variable  $X$  qui décrit le résultat de l'existence de Dieu définie comme suit :

$$X = \begin{cases} 0 & \text{si Dieu n'existe pas} \\ +\infty & \text{sinon} \end{cases}.$$

On a alors  $\mathbb{P}(X = 0) = p$  et  $\mathbb{P}(X = +\infty) = 1 - p$ . Prenons  $p < 1$  (car si  $p = 1$ , on suppose que Dieu n'existe pas). Alors  $\bar{X}$ , la valeur moyenne de  $X$  est donnée par :

$$\bar{X} = p \cdot 0 + (1 - p) \cdot +\infty.$$

Blaise Pascal a appelé cette valeur **espérance** et l'a noté  $\mathbb{E}(X)$ .

### 3.2 Espérance et variables aléatoires

*Remarque.* Il existe plusieurs méthodes pour décrire le comportement d'une variable aléatoire. On s'intéresse ici aux **indicateurs de position**. Il existe d'autres types d'indicateurs dont les **indicateurs de répartition** qui seront vus plus loin.

**Définition 3.1.** On considère  $X$  une variable aléatoire sur un espace de probabilité  $(\Omega, \mathcal{F}, \mathbb{P})$ .

1. Méthode de la médiane :

On évalue le nombre  $\tilde{x}$  tel que  $\mathbb{P}(X \leq \tilde{x}) = \mathbb{P}(X \geq \tilde{x}) = \frac{1}{2}$ .

Lorsque  $X$  est continue, la médiane existe toujours. Si  $X$  est discrète, la médiane n'existe pas obligatoirement et n'est pas forcément unique.

2. Méthode de l'espérance :

On évalue une *moyenne pondérée* des valeurs que peut prendre  $X$  par leur probabilité.

### 3.3 Définition de l'espérance

#### 3.3.1 Cas positif

**Définition 3.2** (Cas discret). Soit  $X$  une v.a.d. à valeurs positives. On note  $p_k := \mathbb{P}(X = x_k)$ . On pose :

$$\mathbb{E}(X) = \sum_{k \geq 0} p_k x_k.$$

*Remarque.* Dans ce cas, l'espérance fait toujours sens et existe toujours mais peut valoir  $+\infty$ .

**Définition 3.3** (Cas absolument continu). Soit  $X$  une v.a.c. définie positive de densité  $f_X$  et de répartition  $F_X$ . On note :

$$\begin{cases} F_X(t) = \mathbb{P}(X \leq t) \\ f_X(t) = \frac{\partial}{\partial t} F_X(t) \end{cases}.$$

On définit alors :

$$\mathbb{E}(X) = \int_0^{+\infty} xf(x) dx.$$

*Remarque.*

- L'intégrale démarre en 0 car la variable aléatoire  $X$  est définie positive ;
- à nouveau, l'espérance existe toujours mais peut valoir  $+\infty$ .

### 3.3.2 Cas général

**Définition 3.4** (Cas discret). Soit  $X$  une v.a.d. à valeurs dans  $E = \{x_0, \dots, x_n\} \subset \mathbb{R}$  fini ou infini dénombrable (typiquement  $E = \mathbb{Z}$ ). On pose  $p_n := \mathbb{P}(X = x_n)$ . On considère la série à termes positifs :

$$\sum_{x_n \in E} |x_n| p_n.$$

Si la série vaut  $+\infty$ , on dit que  $X$  n'est pas intégrable et on ne peut pas définir son espérance.

Si la série est finie, alors le théorème 1.2 entraîne que la série :

$$\sum_{x_n \in E} x_n p_n$$

converge également.

On définit alors :

$$\mathbb{E}(X) = \sum_{x_n \in E} x_n p_n. \quad (3)$$

**Définition 3.5** (Cas absolument continu (à densité)). Soit  $X$  une v.a.c. de densité  $f_X$  sur  $\mathbb{R}$  telle que  $\forall A \in \mathcal{B}(\mathbb{R})$  :

$$\mathbb{P}(X \in A) = \int_A f(x) dx.$$

On considère :

$$I = \int_{-\infty}^{+\infty} |x| f(x) dx. \quad (4)$$

Si  $I = +\infty$ , on dit que  $X$  n'est pas intégrable et on ne peut pas définir son espérance.

Si  $I < +\infty$ , le théorème 1.3 entraîne que l'intégrale :

$$\int_{-\infty}^{+\infty} xf_X(x) dx$$

converge également.

On définit alors :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x) dx. \quad (5)$$

### 3.4 Exemples d'espérance

*Exemple 1.* (exemple de 3.2.) Soit  $X \sim \mathcal{B}(n, p)$  une binomiale. Par définition, on évalue :

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} k = \sum_{k=1}^n \binom{n}{k} p^k (1-p)^{n-k} k = np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{\gamma=0}^{n-1} \binom{n-1}{\gamma} p^\gamma (1-p)^{n-1-\gamma} = np(p + (1-p))^{n-1} = np.\end{aligned}$$

*Exemple 2.* (exemple de 3.2.) Soit  $X \sim \mathcal{P}(\lambda)$  une poisson de paramètre  $\lambda$ . On évalue :

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k \geq 0} \mathbb{P}(X = k) k = \sum_{k \geq 0} \exp(-\lambda) \frac{\lambda^k}{k!} k = \exp(-\lambda) \sum_{k \geq 1} \frac{\lambda^k}{k!} k \\ &= \exp(-\lambda) \lambda \sum_{k \geq 1} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \exp(-\lambda) \sum_{\gamma \geq 0} \frac{\lambda^\gamma}{\gamma!} = \lambda \exp(-\lambda) \exp(\lambda) = \lambda.\end{aligned}$$

*Exemple 3.* (exemple de 3.2.) Soit  $X \sim \text{B\AA}$ . La loi de  $X$  est donnée par  $\mathbb{P}(X = k) = \frac{6}{(\pi k)^2}$  pour tout  $k \geq 1$ . On évalue :

$$\mathbb{E}(X) = \sum_{k \geq 1} \mathbb{P}(X = k) k = \sum_{k \geq 1} \frac{6}{\pi^2} \sum_{k \geq 1} \frac{1}{k^2} k = \frac{6}{\pi^2} \sum_{k \geq 1} \frac{1}{k} = +\infty.^9$$

*Exemple 4.* (exemple de 3.3.) Soit  $X \sim \text{Exp}(\lambda)$  une exponentielle négative<sup>10</sup>. On sait :

$$f_X(t) = \lambda \exp(-\lambda t),$$

et donc on calcule :

$$\mathbb{E}(X) = \int_0^{+\infty} x f_X(x) dx = \int_0^{+\infty} x \lambda \exp(-\lambda x) dx.$$

On pose  $y := \lambda x$  (et donc  $dy = \lambda dx$ ), et on obtient :

$$\mathbb{E}(X) = \frac{1}{\lambda} \int_0^{+\infty} y \exp(-y) dy = \frac{1}{\lambda}.$$

*Exemple 5.* (exemple de 3.3.) Soit  $X \sim \mathcal{U}_{(a,b)}$  une uniforme sur  $(a, b)$  où  $0 \leq a < b \in \mathbb{R}$ . On sait :

$$f_X(t) = \frac{1}{b-a} 1_{(a,b)}(t),$$

et donc, on calcule :

$$\mathbb{E}(X) = \int_0^{+\infty} x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \left[ \frac{x^2}{2(b-a)} \right]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}.$$

*Exemple 6.* (exemple de 3.3.) Soit  $X \sim \frac{1}{2}\mathcal{C}$  une demi-Cauchy. On sait (pour  $t \in [0, +\infty)$ ) :

$$f_X(t) = \frac{2}{\pi(1+t^2)},$$

<sup>9</sup>La série  $\sum_{k \geq 1} \frac{1}{k} = +\infty$  se démontre en utilisant le fait que  $\sum_{i=2^\alpha}^{2^{\alpha+1}} \frac{1}{i} \geq \frac{1}{2} \forall \alpha \in \mathbb{N}$  et donc en faisant tendre  $\alpha \rightarrow +\infty$ , on obtient  $+\infty$ .

<sup>10</sup>la notion d'exponentielle *négative* vient du fait que le paramètre de la fonction exponentielle est négatif, mais la fonction exponentielle est définie positive.



et donc, on calcule :

$$\mathbb{E}(X) = \int_0^{+\infty} x f_X(x) dx = \int_0^{+\infty} \frac{2x}{(1+x^2)} \frac{dx}{\pi} = \frac{1}{\pi} \left[ \log(1+x^2) \right]_0^{+\infty} = +\infty$$

*Exemple 7.* (exemple de 3.5.) Soit  $X \sim \mathcal{C}$  une Cauchy. On sait :

$$f_X(t) = \frac{1}{\pi(1+t^2)},$$

et donc, on calcule :

$$\int_{-\infty}^{+\infty} x f_X(x) dx = 2 \int_0^{+\infty} |x| f_X(x) dx = 2\infty = +\infty.$$

**Proposition 3.6** (Critère de d'Alembert). Soit  $f$  une fonction définie positive sur  $[1, +\infty)$ . On suppose  $f(x) \sim \frac{c}{x^a}$  quand  $x \rightarrow +\infty$  avec  $a, c \in \mathbb{R}$ .

Si  $a < 1$ , alors :

$$\int_1^{+\infty} f(x) dx < +\infty,$$

et si  $a \geq 1$ , alors :

$$\int_1^{+\infty} f(x) dx = +\infty.$$

*Remarque.* Dans les cas des Cauchy, on a  $a = 1$  et  $c = \frac{1}{\pi}$ . Donc, par d'Alembert, on sait que l'intégrale est infinie. On n'a pas besoin de primitive explicite.

*Remarque.* Pour les variables aléatoires continues n'ayant pas d'espérance, on peut s'intéresser à la médiane  $m \in \mathbb{R}$  telle que :

$$\int_{-\infty}^m f(x) dx = \int_m^{+\infty} f(x) dx = \frac{1}{2}.$$

Si  $f$  s'annule en certaines  $x$ , il se peut que  $m$  ne soit pas unique.

Une variable  $X \sim \mathcal{C}$  Cauchy est paire, et donc  $m = 0$ .

### 3.5 Espérance de fonctions de variables aléatoires

**Définition 3.7.** Soit  $X$  une v.a.c. réelle et  $g : \mathbb{R} \rightarrow \mathbb{R}$  une fonction mesurable. La quantité  $Y = g(X)$  est une variable aléatoire car c'est une application :  $Y : \Omega \rightarrow \mathbb{R} : \omega \mapsto (g \circ X)(\omega)$ .

**Théorème 3.8** (Principe de transfert). Soient  $X$  une v.a. et  $Y := g(X)$ . On suppose  $\mathbb{E}(|Y|) < +\infty^{11}$ , alors :

$$\mathbb{E}(Y) = \begin{cases} \sum_{x_n \in E} g(x_n) \mathbb{P}(X = x_n) & \text{si } X \text{ est discrète} \\ \int_{\mathbb{R}} g(x) f_X(x) dx & \text{si } X \text{ est absolument continue} \end{cases}. \quad (6)$$

*Remarque.* Ce théorème signifie que pour déterminer l'espérance de  $g(X)$ , on intègre  $g(x)$  le long de la loi de  $X$ .

---

<sup>11</sup> Ainsi,  $\mathbb{E}(Y)$  a un sens.

*Exemple 8* (Calcul du moment d'ordre 2). On prend  $g : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x^2$ .

Cas discret : soit  $X$  à valeurs dans  $E := \{x_0, x_1, \dots\}$ . On prend  $Y := g(X)$ . Alors, l'espérance est donnée par :

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_{x_n \in E} (x_n)^2 \mathbb{P}(X = x_n).$$

Si  $X \sim \mathcal{P}(\lambda)$  est une poisson de paramètre  $\lambda$ , on a  $E = \mathbb{N}$ . Dès lors, on considère  $x_n = n$ . L'espérance est alors :

$$\begin{aligned} \mathbb{E}(Y) &= \mathbb{E}(X^2) = \sum_{n \geq 0} n^2 \mathbb{P}(X = n) = \sum_{n \geq 1} n^2 \exp(-\lambda) \frac{\lambda^n}{n!} = \exp(-\lambda) \sum_{n \geq 1} n \frac{\lambda^n}{(n-1)!} \\ &= \exp(-\lambda) \sum_{n \geq 1} ((n-1) + 1) \frac{\lambda^n}{(n-1)!} = \exp(-\lambda) \left[ \lambda \sum_{n \geq 1} \frac{\lambda^{n-1}}{(n-1)!} + \sum_{n \geq 1} (n-1) \frac{\lambda^n}{(n-1)!} \right] \\ &= \lambda + \exp(-\lambda) \sum_{n \geq 2} \frac{\lambda^n}{(n-2)!} = \lambda + \lambda^2 \exp(-\lambda) \sum_{n \geq 2} \frac{\lambda^{n-2}}{(n-2)!} \\ &= \lambda + \lambda^2. \end{aligned}$$

Cas absolument continu : soit  $X$  à valeurs dans  $\mathbb{R}$  de densité  $f_X$ . On prend  $Y := g(X)$ . Alors, l'espérance est donnée par :

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{\mathbb{R}} x^2 f_X(x) dx.$$

Si  $X \sim \text{Exp}(\lambda)$  est une exponentielle négative de paramètre  $\lambda$ , on a (en posant  $y := \lambda x$  et donc  $dy = \lambda dx$ ) :

$$\mathbb{E}(Y) = \mathbb{E}(X^2) = \int_{\mathbb{R}} x^2 \lambda \exp(-\lambda x) dx = \frac{1}{\lambda^2} \int_{\mathbb{R}} y^2 \exp(-y) dy = \frac{\Gamma(3)}{\lambda^2} = \frac{2!}{\lambda^2} = \frac{2}{\lambda^2}.$$

**Définition 3.9.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilités. On note :

$$\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}) := \{X \text{ v.a. t. q. } \mathbb{E}(|X|) < +\infty\}. \quad (7)$$

**Théorème 3.10** (Propriété fondamentale de l'espérance). *L'espace  $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$  est un espace vectoriel réel de dimension infinie. Donc :*

$$\forall X, Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}), \lambda, \mu \in \mathbb{R} : \lambda X + \mu Y \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P}).$$

De plus,  $\mathbb{E}(\lambda X + \mu Y) = \lambda \mathbb{E}(X) + \mu \mathbb{E}(Y)$ .

*Remarque.* On dit alors que l'espérance est un opérateur linéaire.

**Théorème 3.11.** *Si  $X \geq 0$  est une v.a. définie positive, alors son espérance est positive.*

## 3.6 Variance

### 3.6.1 Définitions

**Définition 3.12.** Soit  $X \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ . On pose  $Y := (X - \mathbb{E}(X))^2$ . Par définition,  $Y$  est positive, et donc on peut définir son espérance. On pose :

$$\text{Var}(X) := \mathbb{E}(Y). \quad (8)$$

*Remarque.* Il est possible que  $\text{Var}(X) = \mathbb{E}(Y) = +\infty$ . Dans ce cas, on dit que  $X$  est de variance infinie.

*Remarque.* La variance est un indicateur de répartition par rapport à la moyenne. Dans le cas où la variance est infinie, on regarde les quantités  $\mathbb{P}(|X - \mathbb{E}(X)| \geq x)$  pour  $x > 0$  par analogie à la médiane.

**Proposition 3.13.** Soit  $X$  une v.a. de variance  $\text{Var}(X) < +\infty$ . Alors :

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (9)$$

*Démonstration.* On observe que :

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

□

*Remarque.* La variance est une valeur positive. Donc on a  $\mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0$ , et donc  $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$ .

**Proposition 3.14.** La proposition 3.13 peut se retrouver à l'aide de l'inégalité de Cauchy-Schwartz.

*Démonstration.* Dans le cas discret, on pose  $y_n := x_n \sqrt{\mathbb{P}(X = x_n)}$ , et  $z_n := \sqrt{\mathbb{P}(X = x_n)}$ . L'inégalité de Cauchy-Schwartz implique :

$$\left| \sum_{n=1}^N y_n z_n \right| \leq \sqrt{\left( \sum_{n=1}^N y_n^2 \right) \left( \sum_{n=1}^N z_n^2 \right)}.$$

Les valeurs étant positives, on peut passer au carré. En faisant tendre  $N \rightarrow +\infty$ , on obtient :

$$\left( \sum_{n \geq 1} y_n z_n \right)^2 \leq \left( \sum_{n \geq 1} y_n^2 \right) \left( \sum_{n \geq 1} z_n^2 \right).$$

On sait que :

$$\begin{cases} \sum_{n \geq 1} z_n^2 = \sum_{n \geq 1} \mathbb{P}(X = x_n) = 1, \\ \sum_{n \geq 1} y_n^2 = \sum_{n \geq 1} (x_n)^2 \mathbb{P}(X = x_n) = \mathbb{E}(X^2), \\ \sum_{n \geq 1} y_n z_n = \sum_{n \geq 1} y_n z_n = \sum_{n \geq 1} x_n \mathbb{P}(X = x_n) = \mathbb{E}(X). \end{cases}$$

On a bien :

$$\mathbb{E}(X)^2 \leq \mathbb{E}(X^2),$$

qui est l'inégalité (9).

Dans le cas absolument continu, on a  $X$  une v.a. de densité  $f_X$  sur  $\mathbb{R}$ . On pose  $g(x) = \sqrt{f_X(x)}$  et  $h(x) = x\sqrt{f_X(x)}$ . L'inégalité de Cauchy-Schwartz implique :

$$\left| \int_{\mathbb{R}} h(x)g(x) dx \right| \leq \sqrt{\left( \int_{\mathbb{R}} h(x)^2 dx \right) \left( \int_{\mathbb{R}} g(x)^2 dx \right)}.$$

À nouveau, en mettant au carré, on obtient :

$$\begin{aligned} \left( \int_{\mathbb{R}} x f_X(x) dx \right)^2 &\leq \left( \int_{\mathbb{R}} \left( x \sqrt{f_X(x)} \right)^2 dx \right) \left( \int_{\mathbb{R}} \left( \sqrt{f_X(x)} \right) dx \right) \\ &= \left( \int_{\mathbb{R}} x^2 f_X(x) dx \right) \left( \int_{\mathbb{R}} f_X(x) dx \right). \end{aligned}$$

On sait que :

$$\begin{cases} \left( \int_{\mathbb{R}} x f_X(x) dx \right)^2 = \mathbb{E}(X)^2 \\ \int_{\mathbb{R}} x^2 f_X(x) dx = \mathbb{E}(X^2) \\ \int_{\mathbb{R}} f_X(x) dx = 1. \end{cases}$$

On a bien :

$$\mathbb{E}(X)^2 \leq \mathbb{E}(X^2),$$

qui est l'inégalité (9). □

**Définition 3.15.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilités. On pose :

$$\mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P}) := \{V \text{ v.a. t. q. } \mathbb{E}(X^2) < +\infty\}.$$

**Théorème 3.16.** L'espace  $\mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$  est un espace vectoriel réel de dimension infinie. Donc :

$$\forall X, Y \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P}), \lambda, \mu \in \mathbb{R} : \lambda X + \mu Y \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P}).$$

*Remarque.* L'application  $X \mapsto \mathbb{E}(X^2)$  est une forme quadratique et donc n'est pas linéaire.

**Théorème 3.17.** L'espace  $\mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$  est inclus dans l'espace  $\mathcal{L}_1(\Omega, \mathcal{F}, \mathbb{P})$ . Donc si  $\mathbb{E}(X^2) < +\infty$ , alors  $\mathbb{E}(|X|) < +\infty$ .

**Théorème 3.18** (Inégalité de Cauchy-Schwartz sur les espaces  $\mathcal{L}_i$ ). Soient  $X, Y \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ . Alors :

$$|\mathbb{E}(XY)| \leq \mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \quad (10)$$

## 3.7 Moments de variables aléatoires

### 3.7.1 Cas discret

**Définition 3.19.** Soit  $E = x_{i \in \mathbb{N}}$  un espace d'état fini ou infini dénombrable. Si  $X$  est une variable aléatoire à valeurs dans  $E$ , alors sa loi est donnée par  $\{p_i = \mathbb{P}(X = x_i)\}_{i \in \mathbb{N}}$ . Les moments de  $X$  sont les valeurs moyennes de  $F(X)$  où  $F : E \rightarrow \mathbb{R}$  est une fonction donnée.

**Définition 3.20.** On considère la suite de terme général  $|F(x_i)| p_i$ .

(i) Si  $\sum_{i \in \mathbb{N}} |F(x_i)| p_i$  converge, alors le F-moment de  $X$  existe et vaut :

$$\mathbb{E}(F(X)) = \sum_{i \in \mathbb{N}} F(x_i) \mathbb{P}(X = x_i);$$

(ii) si  $\sum_{i \in \mathbb{N}} |F(x_i)| p_i$  diverge, alors le F-moment de  $X$  n'a pas de sens et donc n'est pas défini.

### 3.7.2 Cas absolument continu

**Définition 3.21.** Soient  $E = \mathbb{R}$  et  $f_X : \mathbb{R} \rightarrow \mathbb{R}^+$  mesurable telle que  $\int_{\mathbb{R}} f_X(x) dx$ .  $f_X$  est appelée *densité de probabilité* sur  $\mathbb{R}$ . La variable aléatoire  $X$  associée est telle que :

$$\forall A \in \mathcal{B}(\mathbb{R}) : \mathbb{P}(X \in A) = \int_A f_X(x) dx = \int_{\mathbb{R}} f_X(x) 1_A(x) dx.$$

*Remarque.* Soit  $A \in \mathcal{B}(\mathbb{R})$ . Si  $\int_A dx =: \text{Leb}(A) = 0$ , alors  $\mathbb{P}(X \in A) = 0$ .

**Définition 3.22.** Soit  $F : \mathbb{R} \rightarrow \mathbb{R}$  mesurable.

(i) Si  $\int_{\mathbb{R}} |F(x)| f_X(x) dx$  converge, alors le F-moment de  $X$  existe et vaut :

$$\mathbb{E}(F(X)) = \int_{\mathbb{R}} F(x) f_X(x) dx ;$$

(ii) si  $\int_{\mathbb{R}} |F(x)| f_X(x) dx$  diverge, alors le F-moment de  $X$  n'a pas de sens et donc n'est pas défini.

*Remarque.* Le calcul *effectif* de  $\mathbb{E}(F(X))$  est une intégrale impropre qui peut s'avérer compliquée.

## 3.8 Fonctions génératrices

L'objectif est de trouver un certain type de fonctions qui permet de retrouver la loi d'une variable aléatoire  $X$ .

### 3.8.1 Cas discret

**Définition 3.23.** Soit  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$ . On pose  $p_i := \mathbb{P}(X = i)$  pour tout  $i \geq 1$ . On regarde :

$$F : \mathbb{R} \rightarrow \mathbb{R} : x \mapsto t^x \quad \text{avec } t \in [0, 1].$$

Pour tout  $x$ , on sait  $F(x) \in [0, 1]$ , et donc la fonction  $F$  est définie positive. Dès lors, la série :

$$\sum_{i \geq 1} t^i p_i$$

est majorée par :

$$\sum_{i \geq 1} 1^i p_i = \sum_{i \geq 1} p_i = 1.$$

La série étant convergente, on peut définir pour tout  $t \in [0, 1]$  l'espérance suivante :

$$\mathbb{E}(t^X) := \sum_{k \geq 1} t^k p_k.$$

On pose  $G_X$  telle que :

$$G_X : [0, 1] \rightarrow [0, 1] : t \mapsto \mathbb{E}(t^X).$$

Donc  $G_X$  envoie  $t$  sur le F-moment de  $X$ . On appelle  $G_X$  la *fonction génératrice* de  $X$ .

**Proposition 3.24.** Soient  $X$  une variable aléatoire à valeurs dans  $\mathbb{N}$  et  $G_X$  sa fonction génératrice. Alors pour tout  $i \in \mathbb{N}^*$  :

$$\left. \frac{\partial^i}{\partial t^i} \right|_{t=0} G_X(t) = i! \cdot \mathbb{P}(X = i). \quad (11)$$

*Démonstration.* On sait que  $G_X(t)$  s'écrit sous la forme suivante :

$$G_X(t) = \sum_{k \geq 1} \mathbb{P}(X = k) t^k.$$

Soit  $i \in \mathbb{N}^*$ . On peut « découper » la somme en :

$$G_X(t) = \sum_{k=1}^{i-1} \mathbb{P}(X = k) t^k + \mathbb{P}(X = i) t^i + \sum_{k \geq i} \mathbb{P}(X = k) t^k.$$

En prenant la dérivée  $i$ ème selon  $t$ , on obtient :

$$\frac{\partial^i}{\partial t^i} G_X(t) = \sum_{k=1}^{i-1} \frac{\partial^i}{\partial t^i} t^k \mathbb{P}(X = k) + \frac{\partial^i}{\partial t^i} t^i \mathbb{P}(X = i) + \sum_{k \geq i} \frac{\partial^i}{\partial t^i} t^k \mathbb{P}(X = k) = 0 + i! \mathbb{P}(X = i) + \sum_{k \geq i} \frac{k!}{(k-i)!} t^{k-i} \mathbb{P}(X = k).$$

En évaluant tout cela en  $t = 0$ , on obtient bien :

$$\left. \frac{\partial^i}{\partial t^i} \right|_{t=0} G_X(t) = 0 + i! \cdot \mathbb{P}(X = i) + 0,$$

qui est bien l'égalité (11) □

*Exemple 9.* Soit  $X \sim \mathcal{G}(p)$ , une géométrique de paramètre  $p \in (0, 1)$ . On pose  $q := 1 - p$  et  $p_i := q^{i-1}p$  pour tout  $i \geq 1$ . On calcule sa fonction génératrice  $G_X$  :

$$G_X(t) = \mathbb{E}(t^X) = \sum_{k \geq 1} t^k \mathbb{P}(X = k) = \sum_{k \geq 1} t^k q^{k-1} p = p \sum_{k \geq 1} t^k q^{k-1} = p t \sum_{k \geq 1} (t q)^{k-1} = p t \sum_{k \geq 0} (q t)^k = \frac{p t}{1 - q t},$$

par la formule de somme d'une suite géométrique. On peut réorganiser la formule afin d'obtenir :

$$G_X(t) = \frac{q}{q} \frac{p t}{1 - q t} = \frac{p}{q} \frac{q t}{1 - q t} = \frac{p}{q} \left( -1 + \frac{1}{1 - q t} \right).$$

En dérivant et en instanciant en  $t = 0$ , on a bien :

$$\left. \frac{\partial^i}{\partial t^i} \right|_{t=0} G_X(t) = \frac{p}{q} \left. \frac{\partial^i}{\partial t^i} \right|_{t=0} \frac{1}{1 - q t} = \frac{p}{q} i! \left( \frac{q^i}{(1 - q t)^{i+1}} \right) = \frac{i! p q^{i-1}}{(1 - q t)^{i+1}} = i! p q^{i-1} = i! \cdot \mathbb{P}(X = i).$$

### 3.8.2 Lien entre moments entiers et fonctions génératrices

En prenant une dérivée formelle (terme à terme) de  $G_X(t) = \mathbb{E}(t^X)$ , on obtient :

$$\frac{\partial}{\partial t} G_X(t) = \sum_{k \geq 1} k t^{k-1} \mathbb{P}(X = k) = \mathbb{E}(X t^{X-1}).$$

Le théorème de convergence monotone affirme que si  $t$  croît vers 1, alors  $Xt^{X-1}$  croît vers  $X$  et  $\mathbb{E}(Xt^{X-1})$  croît vers  $\mathbb{E}(X)$ . On a alors :

$$\frac{\partial^n}{\partial t^n} G_X(t) = \mathbb{E} \left[ t^{X-n} \prod_{i=0}^{n-1} (X-i) \right],$$

qui croît vers :

$$\mathbb{E} \left[ \prod_{i=0}^{n-1} (X-i) \right],$$

quand  $t$  croît vers 1.

*Remarque.* On remarque que  $\{\mathbb{E}(X^n)\}_{n \in \mathbb{N}^*}$  se retrouve par linéarité à partir de  $\{\mathbb{E} \left( \prod_{i=0}^{n-1} (X-i) \right)\}$ .

Les moments entiers peuvent donc s'écrire comme une combinaison linéaire d'instanciations de dérivées de fonction génératrices en  $t = 1$ .

*Exemple 10.* Prenons par exemple  $\mathbb{E}(X^2)$ . Par linéarité de l'opérateur  $\mathbb{E}$ , on sait que :

$$\mathbb{E}(X^2) = \mathbb{E}(X^2 - X + X) = \mathbb{E}(X^2 - X) + \mathbb{E}(X) = \mathbb{E}((X-0)(X-1)) + \mathbb{E}(X) = \frac{\partial^2}{\partial t^2} G_X(t) \Big|_{t=1} + \frac{\partial}{\partial t} G_X(t) \Big|_{t=1}.$$

### 3.8.3 Cas absolument continu

**Définition 3.25.** On considère la fonction  $F : x \mapsto \exp(itx)$ . Par définition de l'exponentielle complexe, on sait que  $|F(x)| = 1$ . Donc l'intégrale :

$$\int_{\mathbb{R}} |F(x)| f_X(x) dx = \int_{\mathbb{R}} f_X(x) dx = 1.$$

On peut donc définir :

$$\mathbb{E}(F(X)) = \mathbb{E}(\exp(itX)).$$

On définit alors la fonction génératrice de  $X$  par :

$$G_X(t) : \mathbb{R} \rightarrow \mathbb{C} : t \mapsto \mathbb{E}(\exp(itX)).$$

*Remarque.* On appelle également  $G_X$  la fonction caractéristique de  $X$ .

**Théorème 3.26** (Théorème de Lévy). *La fonction caractéristique  $G_X$  de la variable aléatoire  $X$  caractérise la loi de  $X$ .*

**Théorème 3.27** (Formule d'inversion de Fourier). *Soit  $X$  une variable aléatoire de densité  $f_X$ . On suppose  $\int_{\mathbb{R}} |G_X(t)| dt < +\infty$ . Alors :*

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx) G_X(t) dt.$$

*Remarque.* Il existe également une formule générale d'inversion due à Lévy qui permet de retrouver  $x \mapsto \mathbb{P}(X \leq x)$  en connaissant  $G_X(t)$ . Dès lors, on déduit que deux variables aléatoires  $X$  et  $Y$  sont égales en loi si et seulement si leur fonction caractéristique respective est identique.

$$X \stackrel{\mathcal{D}}{=} Y \iff G_X \equiv G_Y.$$

### 3.8.4 Aperçu de la convergence en loi

**Définition 3.28.** Soit  $\{X_k\}_{k \geq 1}$  une famille de variables aléatoires à valeurs dans  $\mathbb{N}$ . On dit que  $\{X_k\}$  converge en loi si :

$$\forall n \geq 0 : (k \rightarrow +\infty) \Rightarrow (\mathbb{P}(X_k = n) \rightarrow \mathbb{P}(X = n)).$$

La convergence en loi se note :

$$X_k \xrightarrow{\mathcal{D}} X.^{12}$$

**Théorème 3.29.** Soit  $\{X_k\}_{k \geq 1}$  une suite de variables aléatoires et  $X$  une v.a. On pose  $G_{X_k} := \mathbb{E}(t^{X_k})$ . Alors les assertions suivantes sont équivalentes :

1.  $X_k \xrightarrow{\mathcal{D}} X$  ;
2.  $\forall n \geq 0 : p_n^{(k)} := \mathbb{P}(X_k = n) \rightarrow \mathbb{P}(X = n) =: p_n$  quand  $k \rightarrow +\infty$  ;
3.  $\forall n \geq 0, t \in [0, 1] : p_n^{(k)} t^n \rightarrow p_n t^n$  quand  $k \rightarrow +\infty$ .

*Remarque.* Par convergence normale, on déduit  $\forall t \in [0, 1]$  :

$$G_{X_k}(t) \xrightarrow{k \rightarrow +\infty} G_X(t). \quad (12)$$

**Définition 3.30** (Convergence simple). Soit  $(F_n)_{n \in \mathbb{N}}$  une suite de fonctions telles que  $F_n : X \rightarrow Y$ . On dit que la suite  $(F_n)$  converge simplement en l'application  $F$  si :

$$\forall x \in X : (F_n(x))_n \rightarrow F(x) \in Y.$$

*Remarque.* On remarque que  $F_{X_k} \rightarrow F_X$  simplement quand  $k \rightarrow +\infty$ . Dans ce cas, les dérivées convergent également simplement :

$$\forall n \in \mathbb{N}^* : \frac{\partial^n}{\partial t^n} G_{X_k} \rightarrow \frac{\partial^n}{\partial t^n} G_X \text{ simplement.}$$

Et par  $\mathbb{P}(X = i) = \frac{1}{i!} \frac{\partial^i}{\partial t^i} \Big|_{t=0} F_X(t)$ , on déduit :

$$X_k \xrightarrow{\mathcal{D}} X \iff G_{X_k} \rightarrow G_X \text{ simplement.}$$

*Exemple 11* (Loi des événements rares<sup>13</sup>). Soit  $X_k \sim \mathcal{B}(k, p(k))$  où  $p(k) := \lambda k$  pour  $\lambda > 0$  fixé. Par définition, on sait pour tout  $n \leq k$  :

$$\mathbb{P}(X_k = n) = \binom{k}{n} p(k)^n (1 - p(k))^{k-n},$$

et donc :

$$\begin{aligned} \mathbb{E}(t^{X_k}) &= \sum_{n=0}^k \binom{k}{n} p(k)^n (1 - p(k))^{k-n} t^n = \sum_{n=0}^k \binom{k}{n} (p(k)t)^n (1 - p(k))^{k-n} \\ &= (1 - p(k) + p(k)t)^k = \left(1 + \frac{\lambda(t-1)}{k}\right)^k = \exp \left[ k \log \left(1 + \frac{\lambda(t-1)}{k}\right) \right]. \end{aligned}$$

On sait cependant que le développement de Taylor d'ordre 1 de  $x \mapsto \log(1+x)$  autour de  $x=0$  est :

$$T_1(x \mapsto \log(1+x), 0) = x + o(x).$$

<sup>12</sup>Le  $\mathcal{D}$  vient du terme *distribution*.

<sup>13</sup>Distribution de Poisson.



Et donc :

$$\mathbb{E} \left( t^{X_k} \right) = \exp \left[ k \left( \frac{\lambda(t-1)}{k} + o \left( \frac{1}{k} \right) \right) \right].$$

En faisant tendre  $k \rightarrow +\infty$ , on sait que  $k o \left( \frac{1}{k} \right) \rightarrow 0$ , et donc on trouve :

$$\mathbb{E} \left( t^{X_k} \right) \xrightarrow{k \rightarrow +\infty} \exp(\lambda(t-1)).$$

Soit  $X \sim \mathcal{P}(\lambda)$  une poisson de ce même paramètre  $\lambda$ . On calcule également sa fonction génératrice :

$$\mathbb{E} \left( t^X \right) = \sum_{k \geq 1} \mathbb{P}(X = k) t^k = \exp(-\lambda) \sum_{k \geq 1} \frac{\lambda^k}{k!} t^k = \exp(-\lambda) \exp(\lambda t) = \exp(\lambda(t-1)).$$

On en déduit alors que  $\forall t \in [0, 1] : G_{X_k}(t) \xrightarrow{k \rightarrow +\infty} G_X(t)$ . Et par (??), on a :

$$X_k \xrightarrow{k \rightarrow +\infty} X,$$

ce qui est équivalent à :

$$\mathcal{B} \left( k, \frac{\lambda}{k} \right) \xrightarrow{k \rightarrow +\infty} \mathcal{P}(\lambda).$$

**Théorème 3.31** (Théorème de Lévy). Soient  $X$  une variable aléatoire réelle de fonction caractéristique  $G_X$  et une suite de v.a. réelles  $\{X_k\}_{k \in \mathbb{N}}$  de fonction caractéristique respective  $G_{X_k}$ . Alors :

$$X_k \xrightarrow{\mathcal{D}} X \iff G_{X_k} \rightarrow G_X \text{ simplement.}$$

*Exemple 12* (Démontrer De Moivre-Laplace grâce à Lévy). Soient  $p \in (0, 1)$  et  $X_n \sim \mathcal{B}(n, p)$ . On pose :

$$Z_n := \frac{X_n - np}{\sqrt{np(1-p)}}.$$

La fonction caractéristique de  $Z_n$  est donnée par :

$$\begin{aligned} G_{Z_n}(t) &= \mathbb{E} [\exp(itZ_n)] = \mathbb{E} \left[ \exp \left( it \frac{X_n}{\sqrt{np(1-p)}} \right) \right] \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\ &= \left[ (1-p) + p \exp \left( \frac{it}{\sqrt{np(1-p)}} \right) \right]^n \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\ &= \left[ 1 + p \left( \exp \left( \frac{it}{\sqrt{np(1-p)}} \right) - 1 \right) \right]^n \exp \left( -it \sqrt{\frac{np}{1-p}} \right). \end{aligned}$$

On regarde maintenant l'approximation de Taylor de  $x \mapsto \exp(ix)$  autour de  $x = 0$  qui vaut :

$$T_1(x \mapsto \exp(ix), 0) = 1 - \frac{x^2}{2} + o(x^2) + i \left( x + o(x^2) \right) = 1 + ix - \frac{x^2}{2} + o(x^2).$$

On peut donc ré-écrire la fonction caractéristique comme :

$$\begin{aligned} G_{Z_n} &= \left[ 1 + p \left( \left( 1 + i \frac{t}{\sqrt{np(1-p)}} - \frac{t^2}{2np(1-p)} + o(n^{-1}) \right) - 1 \right) \right]^n \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\ &= \left[ 1 + \left( i \frac{t\sqrt{p}}{\sqrt{n(1-p)}} - \frac{t^2}{2n(1-p)} + o(n^{-1}) \right) \right]^n \exp \left( -it \sqrt{\frac{np}{1-p}} \right). \end{aligned}$$

En posant :

$$\alpha := i \frac{t\sqrt{p}}{\sqrt{n(1-p)}} - \frac{t^2}{2n(1-p)} + o(n^{-1}),$$

on peut approximer  $G_{Z_n}(t)$  par :

$$\begin{aligned} G_{Z_n}(t) &= [1 + \alpha]^n \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\ &= \exp \left( \log([1 + \alpha]^n) \right) \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\ &= \exp(n \log(1 + \alpha)) \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\ &= \exp \left( \alpha - \frac{\alpha^2}{2} + o(\alpha^3) \right) \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \end{aligned}$$

On calcule donc  $\alpha - \frac{\alpha^2}{2}$  :

$$\alpha - \frac{\alpha^2}{2} = i \frac{t\sqrt{p}}{\sqrt{n(1-p)}} - \frac{t^2}{2n(1-p)} + o(n^{-1}) - \frac{1}{2} \left( \frac{i^2 t^2 p}{n(1-p)} + \frac{t^4}{4n^2(1-p)^2} + o(n^{-2}) - 2 \frac{it^3 \sqrt{p}}{2n(1-p)\sqrt{n(1-p)}} \right).$$

Or,

$$\frac{t^4}{4n^2(1-p)^2} + o(n^{-2}) - 2 \frac{it^3 \sqrt{p}}{2n(1-p)\sqrt{n(1-p)}} = o(n^{-1}).$$

Donc, on peut simplifier en :

$$\begin{aligned} \alpha - \frac{\alpha^2}{2} &= i \frac{t\sqrt{p}}{\sqrt{n(1-p)}} - \frac{t^2}{2n(1-p)} + o(n^{-1}) - \frac{i^2 t^2 p}{2n(1-p)} + o(n^{-1}) \\ &= it \frac{\sqrt{p}}{\sqrt{n(1-p)}} - \frac{t^2}{2n} \left( \frac{1}{1-p} - \frac{p}{1-p} \right) + o(n^{-1}) \\ &= it \frac{\sqrt{p}}{\sqrt{n(1-p)}} - \frac{t^2}{2n} + o(n^{-1}). \end{aligned}$$

En remettant cela dans la formule précédente, on obtient :

$$\begin{aligned}
G_{Z_n}(t) &= \exp \left[ n \left( it \frac{\sqrt{p}}{\sqrt{n(1-p)}} - \frac{t^2}{2n} + o(n^{-1}) \right) \right] \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\
&= \exp \left[ it \sqrt{\frac{np}{1-p}} - \frac{t^2}{2} + no(n^{-1}) \right] \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\
&= \exp \left( it \sqrt{\frac{np}{1-p}} \right) \exp \left( -\frac{t^2}{2} \right) \exp(no(n^{-1})) \exp \left( -it \sqrt{\frac{np}{1-p}} \right) \\
&= \exp \left( -\frac{t^2}{2} + no(n^{-1}) \right).
\end{aligned}$$

Or, on a :

$$\exp \left( -\frac{t^2}{2} + no(n^{-1}) \right) \xrightarrow{n \rightarrow +\infty} \exp \left( -\frac{t^2}{2} \right) = G_{\mathcal{N}(0,1)}.$$

De là, par Lévy, on sait que  $\mathbb{E}(\exp(itZ_n)) \xrightarrow{n \rightarrow +\infty} \mathbb{E}(\exp(it\mathcal{N}(0,1)))$ , ce qui implique  $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ .

□

## 4 Conditionnement et indépendance

### 4.1 Événements indépendants

**Définition 4.1.** Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilités. Soit  $A \in \mathcal{F}$  un événement de probabilité non-nulle. On note :

$$\mathbb{P}_A : \mathcal{F} \rightarrow [0, 1] : B \mapsto \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

la probabilité conditionnelle selon  $A$ .

*Remarque.* Les deux notations suivantes :  $\mathbb{P}(B|A)$  et  $\mathbb{P}_A(B)$  sont strictement équivalentes.

**Proposition 4.2.** La fonction  $\mathbb{P}_A$  définit une probabilité sur  $(\Omega, \mathcal{F})$ .

**Théorème 4.3** (Formule de Bayes). Soit  $(\Omega, \mathcal{F}, \mathbb{P})$  un espace de probabilité, et soient  $A, B \in \mathcal{F}$  deux événements de probabilité non nulle. Alors :

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

**Théorème 4.4** (Formule des probabilités totales). Soit  $A \in \mathcal{F}$  un événement tel que  $0 < \mathbb{P}(A) < 1$ . Alors, pour tout  $B \in \mathcal{F}$ , on a :

$$\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B|A) + \mathbb{P}(A^c)\mathbb{P}(B|A^c).$$

**Définition 4.5.** Soient  $A, B \in \mathcal{F}$  deux événements tels que  $\mathbb{P}(A), \mathbb{P}(B) \in (0, 1)$ . On dit que  $A$  et  $B$  sont indépendants si :

$$\mathbb{P}(B|A) = \mathbb{P}(B) \quad \text{et} \quad \mathbb{P}(A|B) = \mathbb{P}(A).$$

La notation usuelle est  $A \sqcup B$ .

*Remarque.* Deux événements  $A$  et  $B$  sont indépendants si et seulement si :

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

Cela permet d'étendre la notion d'indépendance à tout  $\mathcal{F}$ .

**Définition 4.6.** Soit  $\{A_i\}_{1 \leq i \leq n} \subset \mathcal{F}$  une famille d'événements. On dit que les  $A_i$  sont *mutuellement indépendants* si :

$$\forall k \in \mathbb{N}, 1 \leq i_1 < \dots < i_k \leq n : \mathbb{P}\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k \mathbb{P}(A_{i_j}).$$

*Exemple 13.* Pour  $n = 3$ , on a  $A, B, C \in \mathcal{F}$ . Si :

- (i)  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$  ;
- (ii)  $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$  ;
- (iii)  $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$  ;
- (iv)  $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ ,

alors  $A, B$ , et  $C$  sont mutuellement indépendantes.

**Définition 4.7.** Soit  $\{A_i\} \subset \mathcal{F}$  une famille d'événements. On dit que les  $A_i$  sont indépendants deux à deux si :

$$\forall i \neq j : \mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

*Remarque.* On remarque que l'indépendance mutuelle implique l'indépendance deux à deux, mais que la réciproque est fausse.

## 4.2 Variables aléatoires indépendantes

**Définition 4.8.** Soient  $X, Y$  deux variables aléatoires. On dit que  $X$  et  $Y$  sont indépendantes si :

$$\forall A, B \in \mathcal{B}(\mathbb{R}) : \mathbb{P}[(X \in A) \cap (Y \in B)] = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

**Définition 4.9.** Soient  $\{X_i\}_{1 \leq i \leq n}$  des variables aléatoires réelles. On dit que les  $X_i$  sont mutuellement indépendants si les événements  $\{X_i \in A_i\}$  sont mutuellement indépendants  $\forall \{A_i\}_{i \in I} \subset \mathcal{B}(\mathbb{R})$ .

## 4.3 Formules de convolution

### 4.3.1 Cas discret

Soient  $X, Y$  deux variables aléatoires à valeurs dans  $\mathbb{N}$ . On suppose  $X \sqcup Y$ . On a donc :

$$\forall \omega \in \Omega : X(\omega) \in \mathbb{N} \ni Y(\omega),$$

ce qui implique :

$$(X + Y)(\omega) \in \mathbb{N}.$$

On doit calculer  $\mathbb{P}(X + Y = n)$  avec  $n \in \mathbb{N}$ . On remarque que :

$$\{X + Y = n\} = \bigcup_{k=0}^n (\{X = k\} \cap \{Y = n - k\}),$$

qui est une union disjointe. On peut donc y appliquer l'axiomatique des probabilités :

$$\mathbb{P}(X + Y = n) = \mathbb{P}\left(\bigcup_{k=0}^n (\{X = k\} \cap \{Y = n - k\})\right) = \sum_{k=0}^n \mathbb{P}(\{X = k\} \cap \{Y = n - k\}).$$

Or, par indépendance (supposée par hypothèse) de  $X$  et  $Y$ , on peut encore séparer la probabilité :

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k).$$

**Proposition 4.10** (Formule de convolution discrète). Soient  $X, Y$  deux v.a.d. indépendantes, alors :

$$\forall n \in \mathbb{N} : \mathbb{P}(X + Y = n) = \sum_{k=0}^n \mathbb{P}(X = k)\mathbb{P}(Y = n - k). \quad (13)$$

*Exemple 14* (Somme indépendante de poissos). Soient  $X_1, X_2$  deux variables aléatoires de loi respective  $\mathcal{P}(\lambda_1)$  et  $\mathcal{P}(\lambda_2)$ . On calcule, par la formule (13) :

$$\begin{aligned}\mathbb{P}(X_1 + X_2 = n) &= \sum_{k=0}^n \mathbb{P}(X_1 = k) \mathbb{P}(X_2 = n - k) = \sum_{k=0}^n \exp(-\lambda_1) \frac{\lambda_1^k}{k!} \exp(-\lambda_2) \frac{\lambda_2^{n-k}}{(n-k)!} \\ &= \exp(-\lambda_1 - \lambda_2) \sum_{k=0}^n \frac{n!}{n!} \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!} \\ &= \exp(-(\lambda_1 + \lambda_2)) \frac{1}{n!} \sum_{k=0}^n \binom{n}{k} \lambda_1^k \lambda_2^{n-k} \\ &= \exp(-(\lambda_1 + \lambda_2)) \frac{1}{n!} (\lambda_1 + \lambda_2)^n.\end{aligned}$$

On a donc  $X + y \sim \mathcal{P}(\lambda_1 + \lambda_2)$ . Cela se note :

$$\mathcal{P}(\lambda_1) + \mathcal{P}(\lambda_2) = \mathcal{P}(\lambda_1 + \lambda_2).$$

**Définition 4.11.** Soient  $X_1, X_2$  deux variables aléatoires de même loi, paramétrisés respectivement par le vecteur  $\theta_1$  et  $\theta_2$ . Donc  $X_1 \sim \mathcal{L}(\theta_1)$ , et  $X_2 \sim \mathcal{L}(\theta_2)$ . Si  $X_1 + X_2$  est de même loi, paramétrisé par une transformation de  $\theta_1$  et  $\theta_2$ , on dit que la somme est *interne* :

$$(X_1 \sim \mathcal{L}(\theta_1)) \wedge (X_2 \sim \mathcal{L}(\theta_2)) \Rightarrow (X_1 + X_2) \sim \mathcal{L}(F(\theta_1, \theta_2)) \iff + \text{ interne},$$

où  $F$  est une fonction quelconque.

*Remarque.* Une réciproque a été montrée dans les années 1930 par Raïkov : Soit  $X \sim \mathcal{P}(\lambda)$ . Supposons qu'il existe  $Y, Z$  indépendantes telles que  $X = Y + Z$ . Alors il existe  $\mu \in (0, \lambda)$  tel que  $Y = \mathcal{P}(\mu)$  et  $Z = \mathcal{P}(\lambda - \mu)$ .

**Théorème 4.12** (Formule de Chu-Vanermonde). Soient  $\alpha, \beta, n \in \mathbb{N}$  tels que  $\alpha + \beta \geq n$ . Alors :

$$\sum_{k=0}^n \binom{\alpha}{k} \binom{\beta}{n-k} = \binom{\alpha + \beta}{n}. \quad (14)$$

*Exemple 15* (Somme indépendante de binomiales). Soient  $X \sim \mathcal{B}(\alpha, p)$ , et  $Y \sim \mathcal{B}(\beta, p)$  deux binomiales de paramètre  $p$ . On calcule  $\mathbb{P}(X + Y = n)$  :

$$\mathbb{P}(X + Y = n) = \sum_{k=0}^n \binom{\alpha}{k} p^k (1-p)^{\alpha-k} \binom{\beta}{n-k} p^{n-k} (1-p)^{\beta-n+k} = p^n (1-p)^{\alpha+\beta-n} \sum_{k=0}^n \binom{\alpha}{k} \binom{\beta}{n-k},$$

où par la formule de Vandermonde ((14)), la somme de coefficients binomiaux est le coefficient binomial suivante :  $\binom{\alpha+\beta}{n}$ . On a donc :

$$\mathbb{P}(X + Y = n) = \binom{\alpha + \beta}{n} p^n (1-p)^{\alpha+\beta-n},$$

ou encore :

$$\mathcal{B}(\alpha, p) + \mathcal{B}(\beta, p) = \mathcal{B}(\alpha + \beta, p).$$

*Remarque.* L'exemple 15 peut également se montrer par une somme de  $N$  variables indépendantes de Bernoulli  $\mathcal{B}(p)$ .

### 4.3.2 Cas absolument continu

Soient  $X, Y$  deux v.a. réelle de densité respective  $f_X$  et  $f_Y$ . On a donc :

$$\mathbb{P}(X + Y \leq x) = \mathbb{P}(X \leq x - Y).$$

En intégrant cette valeur selon les valeurs prises par  $y$ , on obtient :

$$\mathbb{P}(X + Y \leq x) = \int_{\mathbb{R}} \mathbb{P}(X \leq x - Y) f_Y(y) dy.$$

On dérive ensuite (formellement, sous l'intégrale) afin d'obtenir :

$$\frac{\partial}{\partial x} \mathbb{P}(X + Y \leq x) = \frac{\partial}{\partial x} \int_{\mathbb{R}} \mathbb{P}(X \leq x - y) f_Y(y) dy = \int_{\mathbb{R}} f_X(x - y) f_Y(y) dy.$$

**Proposition 4.13** (Formule de convolution absolument continue). Soient  $X, Y$  deux variables aléatoires réelles à densité respective  $f_X$  et  $f_Y$ . Alors :

$$f_{X+Y}(x) = \int_{\mathbb{R}} f_X(x - y) f_Y(y) dy. \quad (15)$$

*Exemple 16* (Somme indépendante de normales). Soient  $X_1 \sim \mathcal{N}(0, \sigma_1^2)$  et  $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ . Par la formule (15), on obtient :

$$\begin{aligned} f_{X+Y}(x) &= \int_{\mathbb{R}} f_X(x - y) f_Y(y) dy = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(x - y)^2}{2\sigma_1^2}\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{y^2}{2\sigma_2^2}\right) dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{x^2 + y^2 - 2xy}{2\sigma_1^2}\right) \exp\left(-\frac{y^2}{2\sigma_2^2}\right) dy \\ &= \frac{\exp\left(-\frac{x^2}{2\sigma_1^2}\right)}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{y^2}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) + \frac{xy}{\sigma_1^2}\right) dy \\ &= \frac{\exp\left(-\frac{x^2}{2\sigma_1^2}\right)}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma_1^2\sigma_2^2}\left(y^2(\sigma_1^2 + \sigma_2^2) - 2xy\sigma_2^2\right)\right) dy. \end{aligned}$$

On cherche ensuite à faire apparaître un carré parfait dans l'exponentielle dans l'intégrale. Pour cela, on cherche  $\alpha, \beta \in \mathbb{R}$  tels que

$$\alpha^2 - 2\alpha\beta = y^2(\sigma_1^2 + \sigma_2^2) - 2xy\sigma_2^2.$$

Pour cela, on trouve :

$$\alpha = y\sqrt{\sigma_1^2 + \sigma_2^2}.$$

Dès lors, on sait :

$$-2xy\sigma_2^2 = -2\alpha\beta = -2y\sqrt{\sigma_1^2 + \sigma_2^2}\beta,$$

ou encore :

$$\beta = \frac{x\sigma_2^2}{\sqrt{\sigma_1^2 + \sigma_2^2}}.$$

Dès lors, on peut réécrire la densité de  $X + Y$  comme :

$$\begin{aligned}
f_{X+Y}(x) &= \frac{\exp\left(-\frac{x^2}{2\sigma_1^2}\right)}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma_1^2\sigma_2^2}(\alpha - \beta)^2 + \frac{\beta^2}{2\sigma_1^2\sigma_2^2}\right) \\
&= \frac{\exp\left(-\frac{x^2}{2\sigma_1^2}\right)}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{1}{2\sigma_1^2\sigma_2^2}(\alpha - \beta)^2 + \frac{x^2\sigma_2^4}{2\sigma_1^2\sigma_2^2(\sigma_1^2 + \sigma_2^2)}\right) \\
&= \frac{\exp\left(-\frac{x^2}{2\sigma_1^2}\right)}{2\pi\sigma_1\sigma_2} \exp\left(\frac{x^2\sigma_2^2}{2\sigma_1^2(\sigma_1^2 + \sigma_2^2)}\right) \int_{\mathbb{R}} \exp\left(-\frac{(\alpha - \beta)^2}{2\sigma_1^2\sigma_2^2}\right) dy.
\end{aligned}$$

On remarque :

$$\begin{aligned}
\Delta &:= \exp\left(-\frac{x^2}{2\sigma_1^2}\right) \exp\left(\frac{x^2\sigma_2^2}{2\sigma_1^2(\sigma_1^2 + \sigma_2^2)}\right) \\
&= \exp\left(-\frac{x^2}{2\sigma_1^2} \left(1 - \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\right) \\
&= \exp\left(-\frac{x^2}{2\sigma_1^2} \left(\frac{\sigma_1^2 + \sigma_2^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\right) \\
&= \exp\left(-\frac{x^2}{2\sigma_1^2} \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right) \\
&= \exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right).
\end{aligned}$$

Dès lors, on peut réécrire :

$$f_{X+Y}(x) = \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{(\alpha - \beta)^2}{2\sigma_1^2\sigma_2^2}\right) dy.$$

En posant :

$$\xi := \frac{\alpha - \beta}{\sigma_1\sigma_2},$$

on a également :

$$d\xi := \frac{\sqrt{\sigma_1^2 + \sigma_2^2}}{\sigma_1\sigma_2} dy,$$

et on peut intégrer :

$$\begin{aligned}
f_{X+Y}(x) &= \frac{\exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right)}{2\pi\sigma_1\sigma_2} \int_{\mathbb{R}} \exp\left(-\frac{\xi^2}{2}\right) d\xi \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{x^2}{2(\sigma_1^2 + \sigma_2^2)}\right).
\end{aligned}$$



On retrouve finalement  $f_{X+Y}(x) = \mathcal{N}(0, \sqrt{\sigma_1^2 + \sigma_2^2})$ . On peut en conclure que la gaussienne est stable par l'addition :

$$\mathcal{N}(0, \sigma_1) + \mathcal{N}(0, \sigma_2) = \mathcal{N}(0, \sqrt{\sigma_1^2 + \sigma_2^2}).$$