MATHF-105 : Probabilités
Résumé

R. Petit

Année académique 2015 - 2016

Contents

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 $x \in (-1, 1)$:

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

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 ($\exists M \in \mathbb{R}$ t. q. $\forall n \in \mathbb{N} : S_n \leqslant M$) et donc converge vers $S \in \mathbb{R}$;
- (S_n) n'est pas bornée ($\forall M \in \mathbb{R} : \exists n \in \mathbb{N}$ 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 $n \to +\infty \Rightarrow S_n \to +\infty$;
- si $x \neq 1$, on a $(1-x)S_n = x - x^{n+1}$, et donc :

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

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

1.1.2 Conclusion de la suite géométrique

On voit alors :

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

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

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

1.2 Rappels d'analyse

Définition 1.1. Une fonction $f : X \to Y$ 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 Ω est l'**espace des chances**, l'ensemble des résultats possibles d'un phénomène aléatoire.

Remarque.

- Ω peut être fini (dénombrable) ou infini ;
- $\Omega = \{0,1\}^{\mathbb{N}}$ est l'ensemble des suites à valeur dans $\{0,1\}$;
- Ω 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 :

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

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 $A_1, \ldots, A_n \in \mathcal{F}$. On a :*

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

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** $[\![m,n]\!]$ par :

$$[\![m,n]\!] : \{x \in \mathbb{N} \text{ t. q. } m \leqslant x \leqslant n\}.$$

Définition 2.6. Soit $\Omega = [\![1,n]\!]$. 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 Ω 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 Ω un ensemble de chances et $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ une famille de parties de Ω. On dit que $\mathcal{F}$ est une tribu s'il respecte les trois propriétés suivantes :

- $\emptyset \in \mathcal{F}$;
- $\forall A : A \in \mathcal{F} \Rightarrow A^{\complement} \in \mathcal{F}$;
- $\forall A_1, \ldots, A_n, \ldots : A_1, \ldots, A_n, \ldots \in \mathcal{F} \Rightarrow \bigcup_{k \geqslant 1} A_k \in \mathcal{F}$.

Une autre appellation pour une tribu est une σ-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 $T \coloneqq \{\emptyset, A, A^{\complement}, \mathcal{P}(\Omega)\}$ est une tribu. 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 I une partie de $\mathcal{P}(\Omega)$. On appelle la *tribu engendrée par* I la plus petite tribu contenant I et on la note $\sigma(I)$.

En prenant $I \coloneqq \{$ intervalles ouverts de $[0, 1]\}$, on obtient $\sigma(I)$ que l'on appelle **tribu des boréliens**. [1]

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

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

où $\mathbb{P}$ satisfait :

(i) $\mathbb{P}(\emptyset) = 0$;

(ii) $\forall A \in \mathcal{F} : \mathbb{P}(A) + \mathbb{P}(A^{\complement}) = 1$;

(iii) $\forall A_1, \ldots, A_n, \ldots$ disjoints deux à deux, on a :

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

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

Remarque. Probabiliser un expérience revient à déterminer :

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

[1] 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 Ω un ensemble fini ou dénombrable. On prend également $\mathcal{F} = \mathcal{P}(\Omega)$.

Si Ω est fini, on parle de tirages, et si Ω 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 $\Omega = [\![0, N]\!]$ (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 $k \in [\![0, N]\!]$.

Remarque. On remarque que $\sum_{k \geqslant 1} p_k = 1$ 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 $\Omega = \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\Omega) \simeq \mathbb{R}$, 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 :

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

où on utilise la formule de la somme des termes d'une suite géométrique u définie par $u_n = u_{n-1} q$ pour $n \geqslant 1$ (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 :

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

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 :

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

On a effectivement :

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

2.2.2 Modèles continus (à densité)

Remarque. On prend Ω un intervalle (fini ou infini[2]) sur $\mathbb{R}$, et $\mathcal{F} = \mathcal{B}(I)$, la tribu des boréliens sur I[3].

Définition 2.16. Soit $f : I \to \mathbb{R}^+$ 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)**.** [4] On prend $I = \mathbb{R}^+$ et $\lambda > 0$. Le modèle exponentiel est défini par :

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

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

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

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

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

[2] 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 $(-\infty, +\infty) = \mathbb{R}$.

[3] Ou encore la tribu engendrée par les intervalles de I.

[4] Également appelé *modèle des files d'attente*.

[5] 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 $y \coloneqq x - \mu$, et donc $dy = dx$:

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

En posant $z \coloneqq \frac{y}{\sigma}$ (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 :

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

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 :

$$\mathbb{P}(\Omega)^2 = \int_{-\pi}^{+\pi} \int_{0}^{+\infty} \exp\left(-\frac{r^2}{2}\right) \frac{r\,dr\,d\theta}{2\pi} = \int_{-\pi}^{+\pi} \frac{d\theta}{2\pi} \int_{0}^{+\infty} r \exp\left(-\frac{r^2}{2}\right) dr = \left[-\exp\left(-\frac{r^2}{2}\right)\right]_0^{+\infty} = 1.$$

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 $\left(\mathsf{R}^+, \mathcal{B}(\mathsf{R}^+)\right)$ 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 :

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

Remarque. On note $\gamma \coloneqq -\Gamma'(1) > 0$ la constante d'Euler-Mascheroni. La question $\gamma \overset{?}{\in} \mathbb{Q}$ 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 :

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

□

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)**.** [6] 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[7] est une application $X : \Omega \to E$ 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 $\mathbb{P}\left(X(\omega) = k\right)$ 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 λ se note $Y \sim \mathcal{P}(\lambda)$.

[6] Le modèle Γ est une généralisation du modèle exponentiel (définition 2.18).
[7] 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[8] est une application $X : \Omega \to \mathbb{R}$ 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 :

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

mesurable et telle que :

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

avec :

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

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 $F_X(t) = \mathbb{P}(X \leqslant t)$, 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 :

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

Donc F_X est croissante avec :

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

et :

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

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 $\mathbb{P}(X = x) = F_X(x) - \lim_{y \to x} F_X(y) = 0$. 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 :

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

pour une fonction f_X donnée.

Définition 2.33. On dit qu'une fonction $f : \mathbb{R} \to \mathbb{R}$ 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 :

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

[8] Souvent écrite v.a.c. ou V.A.-C.

Exemples

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

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

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

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

3. Si $x_3 \sim \mathcal{N}(\mu, \sigma^2)$ est une v.a.c. normale de moyenne μ est de variance σ^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 $n \geqslant 1$. 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 :*

$$\mathbb{P}(Y_{n,p} \leqslant t) \overset{n\to+\infty}{\to} F_{\mathcal{N}(0,1)}(t).$$

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

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

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

2.5 Convergence en loi

Définition 2.37.

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

$$\forall x : F_{Z_n}(x) = \mathbb{P}(Z_n \leqslant x) \overset{n \to +\infty}{\to} F_Z(x).$$

  On note cela :

$$Z_n \overset{\mathcal{D}}{\to} Z.$$

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

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

  On note cela :

$$Z_n \overset{\mathcal{D}}{\to} Z.$$

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

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

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 $\widetilde{x}$ tel que $\mathbb{P}(X \leqslant \widetilde{x}) = \mathbb{P}(X \geqslant \widetilde{x}) = \frac{1}{2}$.

   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 :

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

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 :

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

On définit alors :

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

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 $p_n \coloneqq \mathbb{P}(X = x_n)$. 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 entraine que la série :

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

converge également.

On définit alors :

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

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 :

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

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

Si $E < +\infty$, le théorème 1.3 entraine que l'intégrale :

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

converge également.

On définit alors :

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\, dx. \tag{5}$$

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 λ. On évalue :

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

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

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

Exemple 4. (exemple de 3.3.) Soit $X \sim \text{Exp}(\lambda)$ une exponentielle négative[10]. 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 :

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

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

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

et donc, on calcule :

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

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

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

[9] La série $\sum_{k \geqslant 1} \frac{1}{k} = +\infty$ se démontre en utilisant le fait que $\sum_{i=2^{\alpha}}^{2^{\alpha+1}} \frac{1}{i} \gneqq \frac{1}{2} \forall \alpha \in \mathbb{N}$ et donc en faisant tendre $\alpha \to +\infty$, on obtient $+\infty$.

[10] 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 :

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

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* $x \to +\infty$ *avec* $a, c \in \mathbb{R}$*.*

Si $a < 1$*, alors :*

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

et si $a \geqslant 1$*, 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 $g : \mathbb{R} \to \mathbb{R}$ une fonction mesurable. La quantité $Y = g(X)$ est une variable aléatoire car c'est une application : $Y : \Omega \to \mathbb{R} : \omega \mapsto (g \circ X)(\omega)$.

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

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

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

[11] Ainsi, $\mathbb{E}(Y)$ a un sens.

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

<u>Cas discret :</u> soit X à valeurs dans $E \coloneqq \{x_0, x_1, \ldots\}$. On prend $Y \coloneqq g(X)$. Alors, l'espérance est donnée par :

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

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

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

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

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

Si $X \sim \operatorname{Exp}(\lambda)$ est une exponentielle négative de paramètre λ, on a (en posant $y \coloneqq \lambda x$ 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 :

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

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 $X \geqslant 0$ 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 $Y \coloneqq (X - \mathbb{E}(X))^2$. Par définition, Y est positive, et donc on peut définir son espérance. On pose :

$$\operatorname{Var}(X) \coloneqq \mathbb{E}(Y). \tag{8}$$

Remarque. Il est possible que $\mathrm{Var}(X) = \mathbb{E}(Y) = +\infty$. 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 $\mathbb{P}(|X - \mathbb{E}(X)| \geqslant x)$ pour $x > 0$ par analogie à la médiane.

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

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

Démonstration. On observe que :

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

□

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

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 :

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

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

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

On sait que :

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

On a bien :

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

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-Scwhartz implique :

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

À nouveau, en mettant au carré, on obtient :

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

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 :

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

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 :*

$$\left|\mathbb{E}(XY)\right| \leqslant \mathbb{E}\left(|XY|\right) \leqslant \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}. \tag{10}$$

3.7 Moments de variables aléatoires

3.7.1 Cas discret

Définition 3.19. Soit $E = \{x_i\}_{i\in\mathbb{N}}$ 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 $F : E \to \mathbb{R}$ est une fonction donnée.

Définition 3.20. On considère la série de terme général $\left|F(x_i)\right| p_i$.

(i) Si $\sum_{i\in\mathbb{N}} \left|F(x_i)\right| p_i$ 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 $\sum_{i\in\mathbb{N}} \left|F(x_i)\right| p_i$ 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 $f_X : \mathbb{R} \to \mathbb{R}^+$ mesurable telle que $\int_{\mathbb{R}} f_X(x)\,dx = 1$. 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 $\int_A dx =: \mathrm{Leb}(A) = 0$, alors $\mathbb{P}(X \in A) = 0$.

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

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

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

(ii) si $\int_{\mathbb{R}} \left|F(x)\right| f_X(x)\,dx$ 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 $p_i \coloneqq \mathbb{P}(X = i)$ pour tout $i \geqslant 1$. On regarde :

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

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

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

est majorée par :

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

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

$$\mathbb{E}(t^X) \coloneqq \sum_{k \geqslant 1} t^k p_k.$$

On pose G_X telle que :

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

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* $i \in \mathbb{N}^*$ *:*

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

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

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

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

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

En prenant la dérivée ième selon t, on obtient :

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

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

$$\frac{\partial^i}{\partial t^i}\Big|_{t=0} G_x(t) = 0 + i! \cdot \mathbb{P}(X = i) + 0,$$

qui est bien l'égalité (11) $\square$

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 $i \geqslant 1$. On calcule sa fonction génératrice G_X :

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

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

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

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

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

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 :

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

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 :

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

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 :

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

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 $x \mapsto \mathbb{P}(X \leqslant x)$ 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 $\{X_k\}_{k\geqslant 1}$ une famille de variables aléatoires à valeurs dans $\mathbb{N}$. On dit que $\{X_k\}$ converge en loi vers X si :

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

La convergence en loi se note :

$$X_k \overset{\mathcal{D}}{\to} X.$$ [12]

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

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

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

$$G_{X_k}(t) \overset{k\to+\infty}{\to} G_X(t). \tag{12}$$

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

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

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

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

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

$$X_k \overset{\mathcal{D}}{\to} X \iff G_{X_k} \to G_X \text{ simplement.}$$

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

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

et donc :

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

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

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

[12] Le $\mathcal{D}$ vient du terme *distribution*.
[13] 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 $k \to +\infty$, on sait que $ko\left(\frac{1}{k}\right) \to 0$, et donc on trouve :

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

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

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

On en déduit alors que $\forall t \in [0,1] : G_{X_k}(t) \overset{k\to+\infty}{\to} G_X(t)$. Et par (12), on a :

$$X_k \overset{k\to+\infty}{\to} X,$$

ce qui est équivalent à :

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

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 :*

$$X_k \overset{\mathcal{D}}{\to} X \iff G_{X_k} \to G_X \quad \textit{simplement}.$$

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

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

La fonction caractéristique de Z_n est donnée par :

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

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-écire 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 :

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

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

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

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

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

Or,

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

Donc, on peut simplifier en :

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

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 :

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

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

$\square$

4 Conditionnement et indépendance

4.1 Événements indépendants

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

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

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 :*

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

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 :

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

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

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

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 indépendants.

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 :

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

Définition 4.9. Soient $\{X_i\}_{1 \leqslant i \leqslant n}$ 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 :

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

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

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

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 :*

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

Exemple 14 (Somme indépendante de poissons). 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) :

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

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 θ_1 et θ_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 θ_1 et θ_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* $\alpha+\beta \geqslant n$*. Alors :*

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

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 :

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

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

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

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

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

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

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

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 $\alpha, \beta \in \mathbb{N}$ 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 :

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

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 :

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

on a également :

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

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 :

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

4.3.3 Application de la formule de convolution

On suppose $X \sim \Gamma_t$ et $Y \sim \Gamma_s$ deux v.a. réelles indépendantes. On sait :

$$\begin{cases} f_X(x) & = \frac{1}{\Gamma(t)} x^{t-1} \exp(-x) 1_{\{x>0\}}, \\ f_Y(x) & = \frac{1}{\Gamma(s)} x^{s-1} \exp(-x) 1_{\{x>0\}}. \end{cases}$$

Montrons que la somme de Gamma est interne.

Définition 4.14. On définit l'intégrale abélienne de seconde espèce $B(t, s)$ par :

$$B(t, s) := \int_0^1 (1-z)^{t-1} z^{s-1} \, dz.$$

Remarque. L'intégrale abélienne $B(t, s)$ est une généralisation de :

$$\int_0^1 (1-z)^{t-1} \, dz = \frac{1}{t} \qquad \text{et} \qquad \int_0^1 z^{s-1} \, dz = \frac{1}{s}.$$

Proposition 4.15. *La somme indépendante de $X \sim \Gamma_t$ et $Y \sim \Gamma_s$ est une Γ_{t+s}.*

Démonstration. Par la formule de convolution continue (15), on a :

$$f_{X+Y}(x) = \int_0^x \left(\frac{(x-y)^{t-1}}{\Gamma(t)} \exp(y-x) \right) \left(\frac{x^{s-1}}{\Gamma(s)} \exp(-y) \right) dy = \frac{\exp(-x)}{\Gamma(s)\Gamma(t)} \int_0^x (x-y)^{t-1} y^{s-1} \, dy.$$

On pose ensuite $y := xz$, et donc $dy = x \, dz$. En substituant, on obtient :

$$f_{X+Y}(x) = \frac{\exp(-x)}{\Gamma(t)\Gamma(s)} \int_0^1 (x - xz)^{t-1} (xz)^{s-1} x \, dz = \frac{\exp(-x)}{\Gamma(t)\Gamma(s)} x^{(t-1)+(s-1)+1} \int_0^1 (1-z)^{t-1} z^{s-1} \, dz.$$

On pose :

$$\Psi := \frac{1}{\Gamma(t)\Gamma(s)} B(t, s).$$

On a donc :

$$f_{X+Y}(x) = \Psi x^{t+s-1} \exp(-x).$$

Étant donné que f_{X+Y} est une fonction de densité, on peut imposer

$$1 = \int_{\mathbb{R}} f_{X+Y}(x) \, dx = \Psi \int_0^{+\infty} x^{t+s-1} \exp(-x) = \Psi \Gamma(t+s).$$

On en déduit alors :

$$\Psi = \frac{1}{\Gamma(t+s)}.$$

En replaçant Ψ dans l'équation de f_{X+Y}, on obtient :

$$f_{X+Y}(x) = \frac{1}{\Gamma(t+s)} \exp(-x) x^{t+s-1}.$$

Ce qui prouve que $f_{X+Y} = f_{\Gamma(t+s)}$. $\square$

Remarque. On peut trouver une valeur pour $B(t,s)$ étant donné que :

$$\frac{B(t,s)}{\Gamma(t)\Gamma(s)} = \Psi = \frac{1}{\Gamma(t+s)},$$

on détermine :

$$B(t,s) = \frac{\Gamma(t)\Gamma(s)}{\Gamma(t+s)}.$$

On a donné une *preuve probabiliste* du calcul exact de $B(t,s)$. La preuve *classique* repose sur de l'intégration curviligne complexe (c.f. cours d'analyse de Dieudonné) et est beaucoup plus longue.

Remarque. Il a été mentionné plus haut que $B(t,s)$ est une généralisation de deux intégrales abéliennes séparées. On observe en réalité :

- $B(1,s) = \frac{\Gamma(1)\Gamma(s)}{\Gamma(s+1)} = \frac{\Gamma(s)}{s\Gamma(s)} = \frac{1}{s}$;
- $B(t,1) = \frac{\Gamma(t)\Gamma(1)}{\Gamma(t+1)} = \frac{\Gamma(t)}{t\Gamma(t)} = \frac{1}{t}$;
- $B(0.5, 0.5) = \frac{\Gamma(0.5)^2}{\Gamma(1)} = \frac{\sqrt{\pi}^2}{1} = \pi$ (intégrale de Gauss pour $\Gamma(0.5) = \sqrt{\pi}$).

On sait cependant que :

$$\int_{\mathbb{R}} f_{\Gamma_t + \Gamma_s}(x) = \frac{1}{\Gamma(t+s)} \int_0^1 (1-z)^{t-1} z^{s-1} \, dz.$$

En prenant $t = s = \frac{1}{2}$, on peut déterminer :

$$1 = \frac{1}{\Gamma\left(\frac{1}{2} + \frac{1}{2}\right)} \int_0^1 (1-z)^{\frac{1}{2}} z^{\frac{1}{2}} \, dz = \frac{1}{\pi} \int_0^1 \frac{dz}{\sqrt{z(1-z)}}.$$

On peut donc définir une nouvelle densité de probabilité :

$$z \mapsto \frac{1}{\pi\sqrt{z(1-z)}} 1_{(0,1)}(z).$$

Effectivement, on peut poser $u^2 = z$ (et donc $dz = 2u \, du$), ce qui permet d'écrire :

$$\frac{1}{\pi} \int_0^1 \frac{2u \, du}{\sqrt{u^2(1-u^2)}} = \frac{2}{\pi} \int_0^1 \frac{du}{\sqrt{1-u^2}} = \frac{2}{\pi} \left[\arcsin(u)\right]_0^1 = \frac{2}{\pi}\frac{\pi}{2} = 1,$$

qui donne donc bien une probabilité.

4.4 Indépendance et fonctions caractéristiques

Proposition 4.16. *Soit* X *une variable aléatoire à valeurs dans un ensemble* E. *Soit* $A \subseteq E$. *Alors* $\mathbb{E}(1_A(X)) = \mathbb{P}(X \in A)$.

Démonstration. Commençons par le cas discret. Soient $E = \{x_n\}_{n \in I}$ et X une v.a.d. à valeurs dans E. On sait alors :

$$\mathbb{P}(X \in A) = \sum_{x_n \in A} \mathbb{P}(X = x_n) = \sum_{x_n \in A} 1 \cdot \mathbb{P}(X = x_n) + \sum_{x_n \notin A} 0 \cdot \mathbb{P}(X = x_n) = \sum_{x \in E} 1_A(x)\mathbb{P}(X = x_n) = \mathbb{E}(1_A(X)).$$

Intéressons-nous maintenant au cas absolument continu. Soient $E \subset \mathbb{R}$ et X une v.a.c. de densité $f_X(x)$. On sait alors :

$$\mathbb{P}(X \in A) = \int_A f_X(x)\,dx = \int_{\mathbb{R}} 1_A(x) f_X(x)\,dx = \mathbb{E}(1_A(X)).$$

□

Théorème 4.17. *Soient* X, Y *deux variables aléatoires à valeurs respectivement dans* A *et* B*. Alors :*

$$X \sqcup Y \iff \mathbb{E}\left(1_A(X)1_B(Y)\right) = \mathbb{E}(1_A(X))\mathbb{E}(1_B(Y)).$$

Remarque. Prenons $A_1, \ldots, A_n, B_1, \cdots, B_m \in \mathcal{B}(\mathbb{R})$ et $\lambda_1, \cdots, \lambda_n, \mu_1, \cdots, \mu_m \in \mathbb{R}$. On pose ensuite :

$$f_n(x) = \sum_{i=1}^{n} \lambda_i 1_{A_i}(x),$$
$$g_m(y) = \sum_{j=1}^{m} \mu_j 1_{B_j}(y).$$

Par linéarité, on trouve alors :

$$\mathbb{E}\left(f_n(X)g_m(Y)\right) = \mathbb{E}(f_n(X))\mathbb{E}(g_m(Y)). \tag{16}$$

Théorème 4.18. *Toute fonction* f *mesurable (au sens où* $\forall A \in \mathcal{B}(\mathbb{R}) : f^{-1}(A) \in \mathcal{B}(\mathbb{R})$*) et bornée est limite monotone de fonctions étagées, c'est-à-dire des fonctions du type :*

$$\sum_{i=1}^{n} \lambda_i 1_{A_i}(x).$$

Théorème 4.19 (Principe de convergence monotone)**.** *En passant à la limite pour* $n, m \to +\infty$ *dans* (16)*, on a :*

$$\forall f, g \text{ mesurables bornées }, X, Y \text{ v.a. } : X \sqcup Y \iff \mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)). \tag{17}$$

Remarque. On dit que (17) est la *caractérisation fonctionnelle de l'indépendance*.

Théorème 4.20 (Théorème de Lévy)**.** *Soient* X, Y *deux v.a. réelles. Alors :*

$$\forall t \in \mathbb{R} : X \sqcup Y \iff \mathbb{E}(\exp(itX + Y)) = \mathbb{E}(\exp(itX))\mathbb{E}(\exp(itY)).$$

Remarque. Le théorème 4.20 est un cas particulier du principe de convergence monotone où la classe des fonctions mesurables bornées a été réduite à $f(x) = g(x) = \exp(itx)$.

Exemple 17 (Application de Lévy à deux Gamma). Soient $X \sqcup Y$ telles que $X \sim \Gamma_t$ et $Y \sim \Gamma_s$. Soit $z \in \mathbb{R}$, on calcule :

$$\mathbb{E}(\exp(iz\Gamma_t)) = \int_0^{+\infty} \exp(izu)\exp(-u)u^{t-1}\frac{du}{\Gamma(t)} = \int_0^{+\infty} \exp(u(1 - iz))u^{t-1}\frac{du}{\Gamma(t)}.$$

On pose ensuite $\nu \coloneqq u(1 - iz)$, et donc $d\nu = (1 - it)\,du$. Ce qui amène :

$$\mathbb{E}(\exp(it\Gamma_t)) = \int_0^{+\infty} \exp(-\nu)\frac{\nu^{t-1}}{(1 - iz)^{t-1}}\frac{d\nu}{\Gamma(t)} = \frac{1}{(1 - iz)^t}\int_0^{+\infty} \exp(-\nu)\frac{\nu^{t-1}\,d\nu}{\Gamma(t)} = \frac{1}{(1 - iz)^t}.$$

De manière similaire, on obtient :

$$\mathbb{E}(\exp(iz\Gamma_s)) = \frac{1}{(1-iz)^s}.$$

Et également :

$$\mathbb{E}(\exp(iz\Gamma_{t+s})) = \frac{1}{(1-iz)^{t+s}} = \frac{1}{(1-iz)^t}\frac{1}{(1-iz)^s} = \mathbb{E}(\exp(iz\Gamma_t))\mathbb{E}(\exp(iz\Gamma_s)).$$

Les fonctions caractéristiques caractérisent les lois, on a donc :

$$\Gamma_t + \Gamma_s \overset{\mathcal{D}}{=} \Gamma_{t+s}.$$ [14]

4.5 Lien entre indépendance et De Moivre-Laplace

Rappel et topo Si $X_{n,p} \sim \mathcal{B}(n,p)$, on a vu que De Moivre-Laplace implique pour $n \to +\infty$:

$$\frac{X_{n,p} - np}{\sqrt{np(1-p)}} \overset{\mathcal{D}}{\to} \mathcal{N}(0,1).$$

On sait également que :

$$X_{n,p} = \sum_{i=1}^{n} X_p^{(i)},$$

où les $X_p^{(i)} \sim \mathcal{B}(p)$ sont des Bernoulli de paramètre p. L'espérance et la variance peuvent être aisément déterminées :

$$\begin{cases} \mathbb{E}(X_p^{(i)}) & = p, \\ \mathrm{Var}(X_p^{(i)}) & = p(1-p). \end{cases}$$

On a donc :

$$\frac{\sum_{i=1}^{n} X_p^{(i)} - n\mathbb{E}(X_p^{(1)})}{\sqrt{n\,\mathrm{Var}(X_p^{(1)})}} \overset{\mathcal{D}}{\to} \mathcal{N}(0,1).$$

Le but est donc de généraliser ce résultat.

Théorème 4.21 (Théorème central limite (TCL))**.** *Soit* $X \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ *une v.a. réelle. On pose* $\mu \coloneqq \mathbb{E}(X)$ *et* $\sigma^2 \coloneqq \mathrm{Var}(X)$*. Soient* $\{X_i\}_{1 \leqslant i \leqslant n}$ *des variables aléatoires mutuellement indépendantes de même loi. Alors :*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \overset{n\to+\infty}{\to} \mathcal{N}(0,1),$$

où $S_n = X_1 + X_2 + \cdots X_n$.

Démonstration. On pose :

$$\varphi_n(t) \coloneqq \mathbb{E}\left(\exp\left(it\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)\right).$$

[14] Ce qui est un résultat déjà obtenu plus haut.

On sait que $\mathbb{E}(it\mathcal{N}(0,1)) = \exp\left(-\frac{t^2}{2}\right)$. Par linéarité de l'espérance, on a :

$$\varphi_n(t) = \exp\left(\frac{-it\mu\sqrt{n}}{\sigma}\right)\mathbb{E}\left(\frac{itS_n}{\sigma\sqrt{n}}\right) = \exp\left(\frac{-it\mu\sqrt{n}}{\sigma}\right)\prod_{j=1}^{n}\mathbb{E}\left(\exp\left(\frac{it}{\sigma\sqrt{n}}X_j\right)\right).$$

Or, tous les X_j ont la même loi. Dès lors :

$$\varphi_n(t) = \exp\left(\frac{-it\mu\sqrt{n}}{\sigma}\right)\left(\exp\left(\frac{it}{\sigma\sqrt{n}}X_1\right)\right)^n.$$

On observe que :

$$\forall u \in \mathbb{R} : \left|\exp(iu) - 1 + iu - \frac{u^2}{2}\right| \leqslant u^2\min\{u, 1\}.$$

Donc en posant $\psi(u) \leqslant u^2\min\{u, 1\}$, on peut trouver ψ tel que :

$$\exp(iu) = 1 + iu - \frac{u^2}{2} + \psi(u).$$

Et donc :

$$\begin{aligned}
\mathbb{E}\left(\exp\left(it\frac{X_1}{\sigma\sqrt{n}}\right)\right) &= \left[1 + \frac{it\mu}{\sigma\sqrt{n}} - \frac{t^2\mathbb{E}(X_1^2)}{2\sigma^2 n} + \psi\left(\frac{tX_1}{\sigma\sqrt{n}}\right)\right]^n \\
&= \left[1 + \frac{it\mu}{\sigma\sqrt{n}} - \frac{t^2(\mathbb{E}(X_1)^2 + \mathrm{Var}(X_1))}{2\sigma^2 n} + \psi\left(\frac{tX_1}{\sigma\sqrt{n}}\right)\right]^n \\
&= \left[1 + \frac{it\mu}{\sigma\sqrt{n}} - \frac{t^2\mu^2}{2\sigma^2 n} - \frac{t^2}{2n} + \psi\left(\frac{tX_1}{\sigma\sqrt{n}}\right)\right]^n
\end{aligned}$$

En posant :

$$\alpha \coloneqq \frac{it\mu}{\sigma\sqrt{n}} - \frac{t^2\mu^2}{2\sigma^2 n} - \frac{t^2}{2n} + \psi\left(\frac{tX_1}{\sigma\sqrt{n}}\right),$$

on peut déterminer :

$$\mathbb{E}\left(\exp\left(it\frac{X_1}{\sigma\sqrt{n}}\right)\right) = \exp\left(n\log(1+\alpha)\right).$$

On détermine :

$$n\log(1+\alpha) = n\alpha - \frac{n\alpha^2}{2} + no(\alpha^3).$$

En regroupant tous les termes en $o(n^{-\frac{1}{2}})$, on trouve :

$$\begin{aligned}
n\log(1+\alpha) &= n\left[\frac{it\mu}{\sigma\sqrt{n}} - \frac{t^2\mu^2}{2\sigma^2 n} - \frac{t^2}{2n} - \frac{(it\mu)^2}{2\sigma^2 n} + \psi\left(\frac{tX_1}{\sigma\sqrt{n}}\right) + o(n^{-\frac{1}{2}})\right] \\
&= \left[\frac{it\mu\sqrt{n}}{\sigma} - \frac{t^2}{2} + n\psi\left(\frac{tX_1}{\sigma\sqrt{n}} + no(n^{-\frac{1}{2}})\right)\right].
\end{aligned}$$

On retrouve alors :

$$\varphi_n(t) = \exp(n\log(1+\alpha)) = \exp\left(-\frac{t^2}{2} + o(n^{-\frac{1}{2}})\right)\mathbb{E}\left(\exp\left(in\psi\left(\frac{tX_1}{\sigma\sqrt{n}}\right)\right)\right).$$

Et on sait que :

$$\beta \coloneqq \left|n\psi\left(\frac{tX_1}{\sigma\sqrt{n}}\right)\right| \leqslant \frac{t^2X_1^2}{\sigma^2}\min\left\{1, \frac{tX_1}{\sigma\sqrt{n}}\right\} \overset{n\to+\infty}{\to} 0,$$

ce qui fait tendre l'espérance de $\exp(\beta)$ vers 1, il reste donc :

$$\varphi_n(t) = \exp\left(-\frac{t^2}{2} + o(n^{-\frac{1}{2}})\right) \overset{n\to+\infty}{\to} \exp\left(-\frac{t^2}{2}\right).$$

On voit bien que $\varphi_n(t)$ tend vers la fonction caractéristique d'une $\mathcal{N}(0,1)$ et donc, par Lévy (théorème de convergence), on a bien :

$$\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n\,\mathrm{Var}(X_1)}} \overset{\mathcal{D}}{\to} \mathcal{N}(0,1).$$

$\square$

5 Loi des grands nombres

5.1 Énoncé du résultat faible

Définition 5.1. Soit $(X_1, \dots, X_n)$ une suite de v.a. réelles. On dit que les X_i sont identiquement distribuées si elles suivent toutes une même loi $\mathfrak{L}$.

Si les X_i sont également indépendantes, on dit qu'elles sont iid (indépendantes et identiquement distribuées), et si elles sont *mutuellement* indépendantes, on dit qu'elles sont miid (**mutuellement** iid).

Définition 5.2. Soient $X_1, \dots, X_n$ des variables aléatoires. On appelle la *somme de Césaro* la quantité :

$$\overline{X_n} \coloneqq \frac{1}{n}\sum_{k=1}^{n} X_k,$$

qui représente une moyenne *empirique*.

Définition 5.3 (Convergence en probabilités). Soit $\{Z_n\}_n$ une famille de variables aléatoires à valeur dans E. Soit $Z \in E$. On dit que Z_n *converge en probabilités vers* Z si :

$$\forall \epsilon > 0 : \mathbb{P}\left(|Z_n - E| > \epsilon\right) \overset{n\to+\infty}{\to} 0.$$

On note cela :

$$Z_n \overset{(p)}{\to} Z.$$

Théorème 5.4 (Théorème de Bieaymé-Tchebychev[15]). *Soit* $Z \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$. *Alors :*

$$\forall x : \mathbb{P}\left(\left|Z - \mathbb{E}(Z)\right| > x\right) \leqslant \frac{\mathrm{Var}(Z)}{x^2}.$$

Démonstration. On pose :

$$\begin{cases} A_x & \coloneqq \{\omega \in \Omega \text{ t. q.} \left|Z(\omega) - \mathbb{E}(Z)\right| > x\} \\ A_x^{\complement} & \coloneqq \{\omega \in \Omega \text{ t. q.} \left|Z(\omega) - \mathbb{E}(Z)\right| \leqslant x\}. \end{cases}$$

On peut donc exprimer la variance de Z comme :

$$\mathrm{Var}(Z) = \mathbb{E}\left((Z - \mathbb{E}(Z))^2\right) = \mathbb{E}\left(\left|Z - \mathbb{E}(X)\right|^2 1_{A_x}(Z)\right) + \mathbb{E}\left(\left|Z - \mathbb{E}(Z)\right|^2 1_{A_x^{\complement}}(Z)\right).$$

Par positivité, on peut déduire :

$$\mathrm{Var}(Z) \geqslant \mathbb{E}\left(\left|Z - \mathbb{E}(Z)\right|^2 1_{A_x}(x)\right).$$

Soit $\omega \in \Omega$. Si $\omega \in A_x$, alors :

$$\left|Z(\omega) - \mathbb{E}(Z)\right| > x,$$

par définition. On peut donc exprimer :

$$\mathrm{Var}(Z) \geqslant \mathbb{E}\left(x^2 1_{A_x}(Z)\right) = x^2\mathbb{E}(1_{A_x}(Z)) = x^2\mathbb{P}(X \in A_x) = x^2\mathbb{P}\left(\left|Z - \mathbb{E}(Z)\right| > x\right).$$

En réarrangeant, on obtient :

$$\mathbb{P}\left(\left|Z - \mathbb{E}(Z)\right| > x\right) \leqslant \frac{\mathrm{Var}(Z)}{x^2}.$$

□

[15] L'adaptation orthographique en alphabet latin est assez hasardeuse et variable…

Proposition 5.5. *Soit* $\{X_i\}_{1\leqslant i\leqslant n} \subset \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ *une famille de variables aléatoires mutuellement indépendantes. Alors la variance de la somme des* X_i *est égale à la somme des variances :*

$$\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \operatorname{Var}(X_k).$$

Démonstration. On pose :

$$\Delta \coloneqq \sum_{k=1}^{n} X_k.$$

Par définition, on calcule :

$$\operatorname{Var}(\Delta) = \mathbb{E}\left[(\Delta - \mathbb{E}(\Delta))^2\right] = \mathbb{E}\left[\left(\left(\sum_{k=1}^{n} X_k\right) - \mathbb{E}\left(\sum_{k=1}^{n} X_k\right)\right)^2\right] = \mathbb{E}\left[\left(\sum_{k=1}^{n}\left(X_k - \mathbb{E}(X_k)\right)\right)^2\right].$$

On pose $Y_k \coloneqq X_k - \mathbb{E}(X_k)$. Ce qui nous donne :

$$\operatorname{Var}(\Delta) = \mathbb{E}\left[\left(\sum_{k=1}^{n} Y_k\right)^2\right].$$

Le carré d'une somme donne :

$$\left(\sum_{k=1}^{n} Y_k\right)^2 = \sum_{k=1}^{n} (Y_k)^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Y_i Y_j.$$

On peut donc utiliser la linéarité de l'espérance pour réécrire :

$$\operatorname{Var}(\Delta) = \mathbb{E}\left[\sum_{k=1}^{n} (Y_k)^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Y_i Y_j\right] = \sum_{k=1}^{n} \mathbb{E}\left((Y_k)^2\right) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \mathbb{E}(Y_i Y_j).$$

Par l'indépendance des Y_i, on sait que pour tout $i \neq j$, on a :

$$\mathbb{E}(Y_i Y_j) = \mathbb{E}(Y_i)\mathbb{E}(Y_j) = \mathbb{E}(X_i - \mathbb{E}(X_i))\mathbb{E}(X_j - \mathbb{E}(X_j)) = \left(\mathbb{E}(X_i) - \mathbb{E}(X_i)\right)\left(\mathbb{E}(X_j) - \mathbb{E}(X_j)\right) = 0.$$

On trouve donc :

$$\operatorname{Var}(\Delta) = \sum_{k=1}^{n} \mathbb{E}\left((Y_k)^2\right) = \sum_{k=1}^{n} \mathbb{E}\left((X_k - \mathbb{E}(k))^2\right) = \sum_{k=1}^{n} \operatorname{Var}(X_k).$$

□

Remarque. La proposition 5.5 n'est pas vraie sans l'hypothèse d'indépendance (indépendance deux à deux suffit). Par exemple, en prenant $X_1 = X$ et $X_2 = -X$, on a $X_1 + X_2 = 0$, et donc $\operatorname{Var}(X_1 + X_2) = 0$, alors que $\operatorname{Var}(X_1) + \operatorname{Var}(X_2) = 2\operatorname{Var}(X) \gneqq 0$.

Théorème 5.6 (Loi faible des grands nombres)**.** *Soient* $X_1, \cdots, X_n \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$ *des variables miid. Alors :*

$$\overline{X_n} \overset{(p)}{\to} \mathbb{E}(X).$$

Démonstration. On regarde $\mathbb{P}\left(\left|\overline{X_n} - \mathbb{E}(X)\right| > \epsilon\right)$. On remarque que :

$$\overline{X_n} - \mathbb{E}(X) = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbb{E}(X_k)) = \frac{1}{n}\sum_{k=1}^{n} Y_k.$$

Dès lors, par Bienaymé-Tchebychev, on trouve :

$$\mathbb{P}\left(\left|\overline{X_n} - \mathbb{E}(X)\right| > \epsilon\right) = \mathbb{P}\left(\left|\sum_{k=1}^{n} Y_k\right| > n\epsilon\right) \leqslant \frac{\operatorname{Var}\left(\sum_{k=1}^{n} Y_k\right)}{(n\epsilon)^2}.$$

Par la proposition 5.5, on trouve :

$$\mathbb{P}\left(\left|\overline{X_n} - \mathbb{E}(X)\right| > \epsilon\right) \leqslant \frac{\sum_{k=1}^{n} \operatorname{Var}(Y_k)}{(n\epsilon)^2}.$$

Or, les Y_k suivent la même loi, ils ont donc la même variance (qui d'ailleurs est $\operatorname{Var}(X)$). On trouve alors :

$$\mathbb{P}\left(\left|\overline{X_n} - \mathbb{E}(X)\right| > \epsilon\right) \leqslant \frac{n \operatorname{Var}(X)}{(n\epsilon)^2} = \frac{\operatorname{Var}(X)}{n\epsilon^2} \overset{n\to+\infty}{\to} 0.$$

Ce qui signifie $\overline{X_n} - \mathbb{E}(X) \overset{(p)}{\to} 0$, ou encore $\overline{X_n} \overset{(p)}{\to} \mathbb{E}(X)$. □

Si on peut majorer la variance (ce qui n'est pas toujours applicable), on peut établir un *intervalle de confiance* sur $\mathbb{E}(X)$.

Définition 5.7. Soit X une v.a. Un *intervalle de confiance* sur $\mathbb{E}(X)$ est un intervalle sous la forme $I = [\mathbb{E}(X) \pm \beta(\sigma, n)]$, où $\beta : \mathbb{R} \to \mathbb{R}$ est une fonction, tel que $\mathbb{P}(\overline{X_n} \in I)$ soit suffisamment élevé.

Exemple 18. Soit $X \sim \mathcal{B}(p)$ une Bernoulli de paramètre p et $X_n \sim \mathcal{B}(n, p)$ une binomiale de même paramètre p et de n tirages. On peut borner la variance d'une Bernoulli par :

$$\operatorname{Var}(X) = p(1-p) \leqslant \frac{1}{4}.$$

Dès lors, on sait que $\mathbb{P}\left(\left|\overline{X_n} - p\right| > \epsilon\right) \leqslant \frac{1}{4n\epsilon^2}$. On peut alors fixer un petit ϵ pour borner notre approximation (par exemple $\epsilon = 10^{-2}$), ce qui nous permet de déterminer un nombre de tirages n tel que la probabilité que l'écart entre l'estimation $\overline{X_n}$ et le paramètre effectif p soit supérieure à ϵ soit $\frac{1}{4n\epsilon^2}$. Par exemple, bornons la probabilité par 5%. On a alors :

$$\frac{1}{4n\epsilon^2} = 5\% \iff n = 50\,000.$$

L'intervalle de confiance nous dit donc qu'en réalisant 50 mille tirages, la probabilité que l'estimation $\overline{X_n}$ s'écarte de p de plus de 0.01 est inférieure à 5%.

5.2 Lien entre le TCL et la loi faible des grands nombres

Soit $(X_i)_{1\leqslant i\leqslant n}$ un échantillon de variables aléatoires réelles d'une même loi dans $\mathcal{L}_2(\Omega, \mathcal{F}, \mathbb{P})$. Le TCL dit :

$$\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n \operatorname{Var}(X_1)}} \overset{\mathcal{D}}{\to} \mathcal{N}(0,1),$$

que l'on peut réécrire en :

$$\sqrt{n}\left(\overline{X_n} - \mathbb{E}(X_1)\right) \overset{\mathcal{D}}{\to} \sqrt{\mathrm{Var}(X_1)}\mathcal{N}(0,1) \overset{\mathcal{D}}{=} \mathcal{N}(0, \mathrm{Var}(X_1)).$$

Donc :

$$\overline{X_n} - \mathbb{E}(X) = \frac{1}{\sqrt{n}}\left[\sqrt{n}\left(\overline{X_n} - \mathbb{E}(X_1)\right)\right] \overset{\mathcal{D}}{\to} \frac{1}{\sqrt{n}}\mathcal{N}(0, \mathrm{Var}(X_1)) \overset{\mathcal{D}}{\to} 0.$$

On a donc un résultat proche qui dit :

$$\overline{X_n} - \mathbb{E}(X_1) \overset{\mathcal{D}}{\to} 0. \tag{18}$$

Lemme 5.8. *Soit* $\{X_n\}_{n\geqslant 1}$ *une famille de v.a. réelles, et soit* $c \in \mathbb{R}$ *une constante déterministe. On a :*

$$X_n \overset{\mathcal{D}}{\to} c \iff X_n \overset{(p)}{\to} c.$$

Remarque. Une conséquence de ce lemme que la formule (18) devient $X_n \overset{(p)}{\to} 0$, qui est la loi faible. On a donc retrouvé la loi faible à partir du TCL.