Université de Bordeaux | **Mathématiques**
Préparation à l'agrégation | **Année 2021/2022**

Rappels de Probabilités

---

1 Espaces de Probabilités

Définition 1. Une mesure de probabilités μ est une mesure définie sur un espace mesurable $(E, \mathcal{A})$ telle que $\mu(E) = 1$. On dira alors que $(E, \mathcal{A}, \mu)$ est un espace de probabilités.

Exemple 1.

- Si $(E, \mathcal{A})$ est un espace mesurable, la mesure de Dirac en $x \in E$ est définie par $\delta_x(A) = \mathbf{1}_{x \in A}$.
- Si E est un ensemble fini muni de la tribu discrète, alors $\mu = \frac{1}{|E|} \sum_{x \in E} \delta_x$ est une mesure de probabilité qu'on appelle la mesure uniforme sur E notée $\mathcal{U}(E)$. Par définition, $\mu(A) = \frac{|A|}{|E|}$ ce qui correspond à une règle d'équiprobabilité (nombre de cas favorables / nombre de cas possibles).
- Si $I = [\alpha, \beta]$ est un intervalle borné de $\mathbb{R}$, il existe une unique mesure μ sur I telle que
$$\forall a, b \in I, \quad \mu([a, b]) = b - a \ .$$
Ainsi, $\frac{1}{\beta - \alpha}\mu$ est appelée *mesure uniforme* sur $[\alpha, \beta]$ et notée $\mathcal{U}([\alpha, \beta])$.

Dans la suite, on fixera un espace de probabilités $(\Omega, \mathcal{A}, \mathbb{P})$. Dans ce contexte,

- Ω est appelé *univers*,
- un élément $\omega \in \Omega$ est appelé *éventualité* (La variable courante ω est parfois appelé l'*aléa*),
- un élément $A \in \mathcal{A}$ est appelé *évènement*.

Exemple 2.

- Pour modéliser le lancer de deux dés équilibrés, on utilisera l'espace de probabilité $\Omega = \{1, \ldots, 6\}^2$ muni de sa tribu discrète et de la loi uniforme.
- Pour modéliser le lancer d'une infinité de dés équilibrés, on utilisera l'espace de probabilité $\Omega = \{1, \ldots, 6\}^{\mathbb{N}^*}$ composé des suites $\omega = (\omega_n)_{n \geqslant 1}$ de nombres entiers entre 1 et 6. Il est muni de la tribu engendrée par les évènements de la forme
$$A_{i_1, \ldots, i_n} = \{ \ \omega \in \Omega \mid \omega_1 = i_1, \omega_2 = i_2, \ldots, \omega_n = i_n \ \} \ , \quad (i_1, \ldots, i_n) \in \{1, \ldots, 6\}^n \ , \ n \in \mathbb{N} \ .$$
On peut montrer qu'il existe une unique mesure de probabilité sur Ω telle que
$$\forall n \in \mathbb{N}, \ \forall (i_1, \ldots, i_n) \in \{1, \ldots, 6\}^n, \quad \mathbb{P}(A_{i_1, \ldots, i_n}) = \frac{1}{6^n} \ .$$
- Si l'on veut observer le niveau de pluie, la température, la pression à Bordeaux à un instant précis, on pourra utiliser l'univers $\Omega = \mathbb{R}^3$ muni de la tribu borélienne. Par contre, on ne connaît pas *a priori* la loi de probabilité sous-jacente.

Remarque 1. De l'art de bien choisir sa tribu... Sur $\Omega = \{1, \ldots, 6\}^2$, on peut aussi poser la tribu $\mathcal{A}$ formée des évènements de la forme $A \times \emptyset$ ou $A \times \{1, \ldots, 6\}$ avec $A \subset \{1, \ldots, 6\}$. Cette tribu permet d'observer précisément le premier dé, mais pas le deuxième dé.

Proposition 1. *Soient $A, B \in \mathcal{A}$ et $A_n \in \mathcal{A}$ pour tout $n \in \mathbb{N}$. Alors*

- $A \subset B \implies \mathbb{P}(A) \leqslant \mathbb{P}(B)$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- *Si (A_n) est croissante, alors* $\mathbb{P}(\cup A_n) = \lim \mathbb{P}(A_n)$.
- *Si (A_n) est décroissante, alors* $\mathbb{P}(\cap A_n) = \lim \mathbb{P}(A_n)$.

Lemme 1. *Si deux mesures de probabilités μ, ν sur $(E, \mathcal{A})$ coïncident sur une collection C d'événements stable par intersection finie telle que $E \in C$ et $\mathcal{A} = \sigma(C)$, alors $\mu = \nu$.*

En particulier, si deux mesures de probabilité sur $\mathbb{R}$ coïncident sur les intervalles $]-\infty, x]$ avec $x \in \mathbb{R}$, alors elles sont égales. De même, si deux mesures de probabilité sur le produit $E \times F$ coïncident sur les événements de la forme $A \times B$ avec $A \in \mathcal{A}$ et $B \in \mathcal{B}$, alors elles sont égales.

2 Variables Aléatoires

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilités.

Définition 2. Une variable aléatoire à valeurs dans un espace mesurable $(E, \mathcal{E})$ est une fonction mesurable $X : \Omega \to E$, c'est-à-dire que $X^{-1}(B) \in \mathcal{A}$ dès que $B \in \mathcal{E}$. La loi de X est la mesure image P_X de $\mathbb{P}$ par X, c'est-à-dire

$$\forall B \in \mathcal{E}, \quad P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B) \ .$$

Dans l'écriture précédente, on utilise la notation usuelle $\{X \in B\}$ pour référer à l'image réciproque

$$X^{-1}(B) = \{ \ \omega \in \Omega \mid X(\omega) \in B \ \} \ .$$

On notera que P_X est une mesure de probabilité sur E.

Définition 3. Soit $X_i : (\Omega, \mathcal{A}) \to (E_i, \mathcal{E}_i), i \in I$ une famille de variables aléatoires. La tribu engendrée par $(X_i)_{i \in I}$ est la plus petite tribu rendant mesurables toutes les X_i, c'est-à-dire la tribu engendrée par les évènements de la forme $X_i^{-1}(A_i)$ où $A_i \in \mathcal{E}_i$.

Définition 4. On dira que X est une **variable aléatoire discrète** lorsqu'elle est à valeurs dans un ensemble E fini ou dénombrable (naturellement muni de sa tribu discrète). La loi de X s'écrit alors

$$\forall B \in \mathcal{E}, \quad P_X(B) = \sum_{x \in B} \mathbb{P}(X = x) \ .$$

Définition 5. On dira que X est une **variable aléatoire à densité** losqu'elle est à valeurs dans $\mathbb{R}^d$ muni de sa tribu borélienne et qu'il existe une fonction mesurable $f : \mathbb{R}^d \to \mathbb{R}_+$ telle que

$$\forall B \in \mathcal{E}, \quad P_X(B) = \int_B f(x) dx \ .$$

Cette fonction, unique à un ensemble de mesure nulle près, est appelée la densité de X et notée f_X.

Définition 6. Si X est une v.a.r., la fonction de répartition de X est la fonction $F_X : \mathbb{R} \to [0, 1]$ définie par

$$\forall t \in \mathbb{R}, \quad F_X(t) = \mathbb{P}(X \leqslant t) \ .$$

Elle est croissante, continue à droite, admet des limites à gauche en tout point. De plus, elle admet 0 pour limite en $-\infty$ et 1 pour limite en $+\infty$. Lorsque X admet une densité continue f_X, alors $F_X' = f_X$; plus précisément, F_X est l'unique primitive de f_X qui admet 0 pour limite en $-\infty$.

Théorème 2. *La fonction de répartition d'une v.a.r. caractérise sa loi.*

Autrement dit, si deux v.a.r. X et Y ont même fonction de répartition, alors $P_X = P_Y$.

Définition 7. Soit X, Y deux v.a. définies sur Ω. La loi jointe de X, Y est la loi $P_{(X,Y)}$ du couple (X, Y). On dit que P_X, P_Y sont les lois marginales de $P_{(X,Y)}$. Cette définition s'étend naturellement à un nombre quelconque de variables aléatoires $X_1, X_2, \ldots, X_d$.

3 Lois Usuelles

Soit X une variable aléatoire réelle. On présente ci-dessous des lois usuelles, en donnant à chaque fois les valeurs des espérances et variances (même si ces notions sont définies dans la section suivante).

3.1 Lois discrètes

Définition 8. On dit que X suit la loi uniforme sur un ensemble **fini** E lorsque

$$\forall a \in E, \quad \mathbb{P}(X = a) = \frac{1}{|E|} \, .$$

Cette loi modélise une situation d'équiprobabilité où tous les éléments de E ont même chance de se réaliser.

Définition 9. On dit que X suit la loi de Bernoulli $\mathcal{B}(p)$ de paramètre $p \in [0, 1]$ si

$$\mathbb{P}(X = 0) = 1 - p \quad \text{et} \quad \mathbb{P}(X = 1) = p \, .$$

Ainsi, X modélise le résultat d'une expérience aléatoire à deux issues (expérience élémentaire) avec probabilité p de réussite. On a

$$\mathbb{E}[X] = p \quad \text{et} \quad \mathrm{Var}(X) = p(1 - p) \, .$$

Définition 10. On dit que X suit la loi binomiale $\mathcal{B}(n, p)$ si

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

Ainsi, X modélise le nombre de succès dans n répétitions indépendantes d'une même expérience élémentaire avec probabilité p de réussite. On a

$$\mathbb{E}[X] = np \quad \text{et} \quad \mathrm{Var}(X) = np(1 - p) \, .$$

Définition 11. On dit que X suit la loi géométrique $\mathcal{G}(p)$ de paramètre $p \in]0, 1[$ si

$$\forall k \in \mathbb{N}^*, \quad \mathbb{P}(X = k) = (1 - p)^{k-1} p \, .$$

Ainsi, X modélise l'instant du premier succès lors d'une infinité de répétitions indépendantes d'une même expérience élémentaire de probabilité p de réussite. On a

$$\mathbb{E}[X] = \frac{1}{p} \quad \text{et} \quad \mathrm{Var}(X) = \frac{1 - p}{p^2} \, .$$

Définition 12. On dit que X suit la loi de Poisson $\mathcal{P}(\lambda)$ de paramètre $\lambda > 0$ si

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \, .$$

Dans un cours ultérieur, on s'intéressera au processus de Poisson qui modélise les instants d'apparition d'événements aléatoires après $t = 0$ (par exemple les instants de passages d'un bus à arrêt donné) ; le paramètre λ est lié à la fréquence d'apparition des évènements ; on l'appellera paramètre d'intensité. Dans ce cadre, X modélise le nombre d'évènements se produisant dans l'intervalle de temps $[0, 1]$. On a alors

$$\mathbb{E}[X] = \lambda \quad \text{et} \quad \mathrm{Var}(X) = \lambda \, .$$

Définition 13. On dit que X suit la loi hypergéométrique de paramètres (n, r, s) (entiers $\geqslant 1$) si

$$\forall k \in \{0, \dots, \min(n, s)\}, \quad \mathbb{P}(X = k) = \frac{\binom{s}{k}\binom{r-s}{n-k}}{\binom{r}{n}} \, .$$

Ainsi, X modélise le nombre de boules rouges obtenues lorsqu'on tire sans remise n boules dans une urne contenant r boules dont s rouges. On a

$$\mathbb{E}[X] = \frac{ns}{r} \, .$$

Définition 14. La loi multinomiale intervient lorsqu'on effectue n répétitions indépendantes d'une expérience ayant k issues $i_1, \dots, i_k$ de probabilités respectives $p_1, \dots, p_k$. Elle modélise alors la probabilité d'avoir obtenu exactement m_1 fois i_1, ..., m_k fois i_k. Par exemple, dans une succession de n tirages avec remise de boules ayant k couleurs possibles présentes respectivement en proportions $p_1, \dots, p_k$, la loi multinomiale est la loi du k-uplet $(X_1, \dots, X_k)$ donnant le nombre de boules tirées de chaque couleur. Ainsi, si X suit la loi multinomiale de paramètres $(n, p_1, \dots, p_k)$, alors pour tout $m \in \mathbb{N}^k$ tel que $m_1 + \dots + m_k = n$, on a

$$\mathbb{P}\big(X = (m_1, \dots, m_k)\big) = \frac{n!}{m_1! \dots m_k!} p_1^{m_1} \dots p_k^{m_k} .$$

Définition 15. La loi hypergéométrique généralisée intervient dans un contexte de tirage *sans remise* de n objets dans une urne contenant N objets, ces N objets ayant un caractère pouvant prendre k valeurs possibles $i_1, \dots, i_k$. La loi hypergéométrique généralisée, de paramètre $(n, N, (N_1, \dots, N_k))$ (où l'on a nécessairement $N_1 + \dots + N_k = N$) est la loi du k-uplet $(X_1, \dots, X_k)$ comptant le nombre d'objets de chaque caractère possible : pour tout $m \in \mathbb{N}^k$ tel que $m_1 + \dots + m_k = n$, on a

$$\mathbb{P}\big(X = (m_1, \dots, m_k)\big) = \frac{\binom{N_1}{m_1} \dots \binom{N_k}{m_k}}{\binom{N}{n}} .$$

3.2 Lois continues

Définition 16. La loi uniforme $\mathcal{U}(A)$ sur un borélien A de $\mathbb{R}^d$ dont la mesure de Lebesgue $\lambda(A) \in]0, \infty[$ est la loi dont la densité est $\frac{1}{\lambda(A)} \mathbf{1}_A$. Ceci modèle un cas continu d'équiprobabilité où aucune partie de A n'est privilégiée. Lorsque A est un intervalle $[a, b]$ de $\mathbb{R}$ avec $a < b$, alors la densité s'écrit $\frac{1}{b-a} \mathbf{1}_{[a,b]}$ et on a alors

$$\mathbb{E}[X] = \frac{a+b}{2} \quad \text{et} \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12} .$$

Définition 17. La loi gaussienne $\mathcal{N}(m, \sigma^2)$ de paramètre $m \in \mathbb{R}$ et $\sigma^2 > 0$ est la loi de densité

$$x \longmapsto \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) .$$

Cette loi apparaît dans de nombreux contextes et notamment dans le théorème central limite. On a

$$\mathbb{E}[X] = m \quad \text{et} \quad \mathrm{Var}(X) = \sigma^2 .$$

Définition 18. La loi exponentielle $\mathcal{E}(\lambda)$ de paramètre $\lambda > 0$ est la loi de densité $x \mapsto \lambda e^{-\lambda x} \mathbf{1}_{x>0}$. La loi exponentielle a une propriété d'absence de mémoire qui la rend apte à modéliser la durée de vie d'un appareil dans un contexte idéal "sans vieillissement".

$$\mathbb{E}[X] = \frac{1}{\lambda} \quad \text{et} \quad \mathrm{Var}(X) = \frac{1}{\lambda^2} .$$

Définition 19. La loi Gamma $\Gamma(a, \lambda)$ de paramètres $a > 0, \lambda > 0$ est la loi à densité

$$x \longmapsto \frac{\lambda}{\Gamma(a)} e^{-\lambda x} (\lambda x)^{a-1} \mathbf{1}_{\mathbb{R}_+} \quad \text{où} \quad \Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx .$$

Elle généralise la loi exponentielle et le paramètre $a > 0$ permet de tenir compte du vieillissement de l'appareil. On a

$$\mathbb{E}[X] = \frac{a}{\lambda} \quad \text{et} \quad \mathrm{Var}(X) = \frac{a}{\lambda^2} .$$

4 Espérance

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité.

Définition 20. Soit X une v.a.r.

- Si X est positive, on définit

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) \in [0, +\infty] .$$

- En général, on dira que X est intégrable si $\int_\Omega |X(\omega)| d\mathbb{P}(\omega) < \infty$ et dans ce cas on définit

$$\mathbb{E}[X] = \int_\Omega X(\omega) d\mathbb{P}(\omega) \in \mathbb{R} .$$

On notera que si $a, b \in \mathbb{R}$, $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.
Rappelons immédiatement trois théorèmes fondamentaux provenant de la théorie de la mesure.

- **Convergence monotone :** Si (X_n) est une suite croissante de v.a. positives de limite X, alors

$$\mathbb{E}[X_n] \to \mathbb{E}[X] .$$

- **Lemme de Fatou :** Si (X_n) est une suite de v.a. positives, alors

$$\mathbb{E}[\liminf X_n] \leqslant \liminf \mathbb{E}[X_n] .$$

- **Convergence dominée :** Si (X_n) est une suite de v.a.r. convergeant presque sûrement vers X et s'il existe une v.a. intégrable Z vérifiant pour tout n, $|X_n| \leqslant Z$ presque sûrement, alors

$$\mathbb{E}[X_n] \to \mathbb{E}[X] .$$

Le cours d'intégration fournit aussi la définition et propriétés des espaces $L^p(\Omega)$, $p \in [1, \infty]$.

On notera que

$$\forall A \in \mathcal{A}, \quad \mathbb{E}[\mathbf{1}_A] = \int_\Omega \mathbf{1}_A(\omega) d\mathbb{P}(\omega) = \mathbb{P}(A) .$$

De plus, par définition de la loi de X, dès que X est intégrable, on a

$$\mathbb{E}[X] = \int_\mathbb{R} x dP_X(x) .$$

Plus généralement, on a le théorème suivant (théorème de transfert).

Théorème 3. *Si X est une v.a. à valeurs dans $(E, \mathcal{E})$ et si $\psi : E \to \mathbb{R}$ est mesurable positive, alors*

$$\mathbb{E}[\psi(X)] = \int_E \psi(x) dP_X(x) .$$

Cette définition s'étend au cas où ψ est mesurable de signe quelconque, à condition que $\psi(X)$ soit intégrable.

Exemple 3. Lorsque X est une variable aléatoire discrète à valeurs dans E, alors

$$\mathbb{E}[\psi(X)] = \sum_{x \in E} \psi(x) \mathbb{P}(X = x) .$$

Exemple 4. Lorsque X est une variable aléatoire à densité f sur $\mathbb{R}^d$, alors

$$\mathbb{E}[\psi(X)] = \int_{\mathbb{R}^d} \psi(x) f(x) dx .$$

Remarque 2. (très importante) En fait, les quantités $\mathbb{E}[\psi(X)]$ pour ψ mesurable positive caractérisent la loi de X. Ceci se voit quand on écrit $P_X(B) = \mathbb{E}[\mathbf{1}_B(X)]$ comme dans la preuve précédente. Cette remarque est utilisée régulièrement en pratique pour calculer la loi d'une variable aléatoire.

5 Moments, variance, covariance

Définition 21. On dit qu'une v.a.r. X admet un moment d'ordre $k \in \mathbb{N}^*$ si $\mathbb{E}[|X|^k] < \infty$, et dans ce cas, le moment d'odre k est $\mathbb{E}[X^k]$.

Si X est intégrable, on définit la variance de X et son écart-type par

$$\mathrm{Var}(X) = \mathbb{E}\big[(X - \mathbb{E}[X])^2\big] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \quad , \quad \sigma_X = \sqrt{\mathrm{Var}(X)} \ .$$

Lorsque X admet un moment d'ordre 2, et que $\sigma_X > 0$, la variable centrée réduite associée est $X^* = \frac{X-\mathbb{E}[X]}{\sigma_X}$.

Proposition 2 (Inégalité de Markov)**.** *Si X est une v.a. positive et p>0,*

$$\forall a > 0, \quad \mathbb{P}(X \geqslant a) \leqslant \frac{1}{a^p}\mathbb{E}[X^p] \ .$$

Proposition 3 (Inégalité de Bienaymé-Tchebytchev)**.** *Si $X \in L^2(\Omega)$, alors*

$$\forall a > 0, \quad \mathbb{P}\big(|X - \mathbb{E}[X]| \geqslant a\big) \leqslant \frac{1}{a^2}\mathrm{Var}(X) \ .$$

Définition 22. La covariance d'un couple (X, Y) de v.a.r. de carrés intégrables est définie par

$$\mathrm{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big] \ .$$

Si de plus $\sigma_X\sigma_Y > 0$, on définit aussi le coefficient de corrélation de X, Y par $\rho(X, Y) = \dfrac{\mathrm{Cov}(X, Y)}{\sigma_X\sigma_Y}$. Plus généralement, si $X = (X_1, \dots, X_d)^T$ est un vecteur aléatoire de $\mathbb{R}^d$ avec $X_i \in L^2(\mathbb{R})$ pour tout i, la matrice de covariance de X notée Γ_X est la matrice formée par les $\mathrm{Cov}(X_i, X_j)$ pour $i, j = 1, \dots, d$. Pour ce vecteur aléatoire, on peut en fait écrire directement sous forme matricielle

$$\mathbb{E}[X] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_d] \end{pmatrix} \quad \text{et} \quad \Gamma_X = \mathbb{E}\big[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T\big] = \big(\ \mathrm{Cov}(X_i, X_j)\ \big)_{1\leqslant i,j\leqslant d} \ .$$

6 Fonction Génératrice

Définition 23. Soit X une v.a. à valeurs dans $\mathbb{N}$. La fonction génératrice de X est définie par

$$G_X(s) = \mathbb{E}[s^X] = \sum_{k=0}^{\infty} \mathbb{P}(X = k)s^k \quad (s \in [0, 1]) \ .$$

Remarquons que la série entière définissant $G_X(s)$ a un rayon de convergence $\geqslant 1$ puisqu'elle converge en $s = 1$. Par conséquent, G_X est au moins définie sur le disque unité ouvert du plan complexe, et y est holomorphe. L'expression de ses dérivées sur le disque ouvert $D(0, 1)$ est

$$G_X^{(m)}(s) = \sum_{k=m}^{\infty} \mathbb{P}(X = k)k(k-1)\dots(k-m+1)s^{k-m} = \mathbb{E}\big[X(X-1)\dots(X-m+1)s^{X-m}\big] \ ,$$

Théorème 4. *La fonction génératrice d'une v.a. à valeurs dans $\mathbb{N}$ caractérise sa loi.*

Proposition 4. *Soit X une v.a. à valeurs dans $\mathbb{N}$. La fonction G_X est m fois dérivable à gauche en 1 si et seulement si X a un moment d'ordre m fini. Dans ce cas, on a*

$$G_X^{(m)}(1) = \mathbb{E}\big[X(X-1)\dots(X-m+1)\big] \ .$$

Exemple 5. La fonction génératrice de la loi binomiale de paramètre (n, p) est

$$G(s) = (1 - p + ps)^n \ .$$

Exemple 6. La fonction génératrice de la loi de Poisson de paramètre λ est

$$G(s) = e^{\lambda(s-1)} \ .$$

7 Indépendance

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

7.1 Définitions et premières propriétés

Définition 24. Soient deux évènements $A, B \in \mathcal{A}$.

- On dit que A et B sont indépendants si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
- Lorsque $\mathbb{P}(B) > 0$, on définit la probabilité conditionnelle

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

On note que si $\mathbb{P}(B) > 0$, alors l'indépendance de A et B se traduit par $\mathbb{P}(A|B) = \mathbb{P}(A)$, autrement dit, le fait de savoir que B se réalise n'apporte aucune information supplémentaire sur A.

Profitons-en pour rappeler deux formules fondamentales. Si $(A_1, \dots, A_n)$ est un système complet d'évènements, c'est-à-dire une partition $\Omega = \sqcup_{i=1}^n A_i$ de l'univers en n évènement disjoints tels que $\mathbb{P}(A_i) > 0$, alors pour tout évènement B, la formule des probabilités totales s'écrit

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(A_i \cap B) = \sum_{i=1}^n \mathbb{P}(B|A_i)\mathbb{P}(A_i) \ .$$

Ceci s'étend immédiatement à un système complet d'évènements dénombrable $(A_n)_{n\in\mathbb{N}}$. Maintenant, si A, B sont des évènements tels que $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, la formule de Bayes permet d'inverser les conditionnements

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

Définition 25. On dit que les évènements $A_1, \dots, A_n$ sont indépendants si pour tout sous-ensemble non-vide $\{j_1, \dots, j_p\}$ de $\{1, \dots, n\}$, on a

$$\mathbb{P}(A_{j_1} \cap \dots \cap A_{j_p}) = \mathbb{P}(A_{j_1}) \dots \mathbb{P}(A_{j_p}) \ .$$

Ceci revient à dire que pour tous $B_i \in \sigma(A_i) = \{\varnothing, A_i, A_i^c, \Omega\}$,

$$\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1) \dots \mathbb{P}(B_n) \ .$$

Remarque 3. Attention, l'indépendance de la famille $(A_1, \dots, A_n)$ dans son ensemble est plus forte que l'indépendance deux à deux des évènements $A_1, \dots, A_n$.

Définition 26. On dit que les tribus $\mathcal{B}_1, \dots, \mathcal{B}_n \subset \mathcal{A}$ sont indépendantes si

$$\forall A_1 \in \mathcal{B}_1, \dots, A_n \in \mathcal{B}_n, \quad \mathbb{P}(A_1 \cap \dots \cap A_n) = \mathbb{P}(A_1) \dots \mathbb{P}(A_n) \ .$$

Définition 27. On dit que les v.a. $X_1, \dots, X_n$ à valeurs dans $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$ sont indépendantes si les tribus $\sigma(X_1), \dots, \sigma(X_n)$ le sont, ce qui revient à dire que

$$\forall F_1 \in \mathcal{E}_1, \dots, \forall F_n \in \mathcal{E}_n, \quad \mathbb{P}\big(\{X_1 \in F_1\} \cap \dots \cap \{X_n \in F_n\}\big) = \mathbb{P}(X_1 \in F_1) \dots \mathbb{P}(X_n \in F_n) \ .$$

Théorème 5. *Soient $X_1, \dots, X_n$ des v.a. à valeurs dans $(E_1, \mathcal{E}_1), \dots, (E_n, \mathcal{E}_n)$.*

- *Les v.a. $X_1, \dots, X_n$ sont indépendantes si et seulement si la loi $P_{(X_1,\dots,X_n)}$ de $(X_1, \dots, X_n)$ est la loi produit $P_{X_1} \otimes \dots \otimes P_{X_n}$. Ceci équivaut à dire que pour toutes fonctions mesurables positives $\psi_i : E_i \to \mathbb{R}, 1 \leqslant i \leqslant n$, on a*

$$\mathbb{E}\big[\psi_1(X_1) \dots \psi_n(X_n)\big] = \mathbb{E}[\psi_1(X_1)] \dots \mathbb{E}[\psi_n(X_n)] \ .$$

- *Si les X_i sont des v.a. discrètes, elles sont indépendantes si et seulement si*

$$\forall x_1 \in E_1, \dots, \forall x_n \in E_n, \quad \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \ .$$

- *Si les X_i sont des v.a.r. à densité f_i, elles sont indépendantes si et seulement si la loi de $(X_1, \dots, X_n)$ admet pour densité le produit des densité*

$$f(x_1, \dots, x_n) = f_1(x_1) \dots f_n(x_n) \quad p.p. \quad (x_1, \dots, x_n) \in \mathbb{R}^n \ .$$

Définition 28. On dit qu'une famille infinie $(X_i)_{i \in I}$ de v.a. est indépendante si et seulement si l'on a indépendance de toute sous-famille finie $(X_{i_1}, \dots, X_{i_n})$ avec $\{i_1, \dots, i_n\} \subset I$.

7.2 Somme de v.a. indépendantes

Définition 29. Si μ, ν sont deux mesures positives sur $\mathbb{R}^d$, la convolution $\mu * \nu$ est la mesure image du produit $\mu \otimes \nu$ par l'application $(x, y) \mapsto x + y$. Elle vérifie donc que pour toute fonction $\psi : \mathbb{R}^d \to \mathbb{R}$ mesurable positive,

$$\int_{\mathbb{R}^d} \psi(z)(\mu * \nu)(dz) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \psi(x + y)\mu(dx)\nu(dy) \ .$$

Proposition 5. *Si X, Y sont deux v.a. indépendantes dans $\mathbb{R}^d$, la loi de $X + Y$ est la convolution $P_X * P_Y$.*

Par exemple, lorsque X et Y sont indépendantes et à densité sur $\mathbb{R}^d$, on retrouve que $X + Y$ admet pour densité

$$f_X * f_Y(z) = \int_{\mathbb{R}^d} f_X(x) f_Y(z - x) dx \ .$$

Proposition 6. *Si X, Y sont deux v.a. indépendantes à valeurs dans $\mathbb{N}$, alors $X + Y$ admet pour fonction génératrice*

$$G_{X+Y}(s) = G_X(s) G_Y(s) \ .$$

Exemple 7. La loi de Bernoulli $\mathcal{B}(p)$ admet pour fonction génératrice $1 - p + ps$. La somme de n v.a. de Bernoulli indépendantes de paramètre p admet donc pour fonction génératrice $(1 - p + ps)^n$, qui est la fonction génératrice de la loi binomiale. On retrouve que $\mathcal{B}(n, p)$ est la loi de la somme de n v.a. $\mathcal{B}(p)$ indépendantes.

Théorème 6. *Soit (X_n) une suite de v.a. i.i.d. à valeurs dans $\mathbb{N}$. Soit N une v.a. à valeurs dans $\mathbb{N}$.*
On suppose que N est indépendante des (X_n). Alors la fonction génératrice de $S = \sum_{i=1}^{N} X_i$ s'écrit

$$G_S(s) = G_N(G_{X_1}(s)) \ .$$

7.3 Lemme de Borel-Cantelli

Définition 30. Soit (A_n) une suite d'évènements.

- On définit

$$\liminf_{n \to \infty} A_n = \bigcup_{N \in N} \bigcap_{n \geqslant N} A_n \ .$$

  Autrement dit, l'évènement $\liminf A_n$ se réalise si et seulement si les évènements A_N se réalisent tous à partir d'un certain rang N.

- On définit

$$\limsup_{n \to \infty} A_n = \bigcap_{N \in N} \bigcup_{n \geqslant N} A_n \ .$$

  Autrement dit, l'évènement $\limsup A_n$ se réalise si et seulement une infinité d'évènements A_n se réalise, ou plus précisément, s'il existe une infinité d'indices n pour lesquels A_n se réalise.

Remarquons que

$$\mathbf{1}_{\liminf A_n} = \liminf \mathbf{1}_{A_n} \quad \text{et} \quad \mathbf{1}_{\limsup A_n} = \limsup \mathbf{1}_{A_n} \ .$$

Théorème 7 (Lemme de Borel-Cantelli)**.** *Soit (A_n) une suite d'évènements.*

1. *Si $\sum \mathbb{P}(A_n) < \infty$, alors*
$$\mathbb{P}(\limsup A_n) = 0 \ .$$
2. *Si $\sum \mathbb{P}(A_n) = \infty$ et si les (A_n) sont indépendants, alors*
$$\mathbb{P}(\limsup A_n) = 1 \ .$$

Démonstration. 1. Remarquons que la suite $\left(\bigcup_{n=N}^{\infty} A_n\right)$ est décroissante. Par conséquent,
$$\mathbb{P}(\limsup A_n) = \lim_{N\to\infty} \mathbb{P}\left(\bigcup_{n=N}^{\infty} A_n\right)$$
Or pour tout N,
$$\mathbb{P}\left(\bigcup_{n=N}^{\infty} A_n\right) \leqslant \sum_{n=N}^{\infty} \mathbb{P}(A_n)$$
et le membre de droite tend vers zéro si $\sum \mathbb{P}(A_n)$ converge. D'où le premier point.

2. On a pour tout N,
$$\mathbb{P}\left(\bigcup_{n=N}^{\infty} A_n\right) = \lim_{M\to\infty} \mathbb{P}\left(\bigcup_{n=N}^{M} A_n\right) \ .$$
Or pour tous M, N,
$$\mathbb{P}\left(\bigcup_{n=N}^{M} A_n\right) = 1 - \mathbb{P}\left(\bigcap_{N\leqslant n\leqslant M} A_n^c\right) \ .$$
Comme les A_n sont supposés indépendants,
$$\mathbb{P}\left(\bigcup_{n=N}^{M} A_n\right) = 1 - \prod_{n=N}^{M}(1-\mathbb{P}(A_n)) \geqslant 1 - \exp\left(-\sum_{n=N}^{M} \mathbb{P}(A_n)\right)$$
où l'inégalité provient de $1 - x \leqslant e^{-x}, \forall x \in \mathbb{R}$. Si $\sum \mathbb{P}(A_n) = \infty$, alors à N fixé, le membre de droite tend vers 1 quand $M \longrightarrow \infty$. Ainsi,
$$\mathbb{P}\left(\bigcup_{n=N}^{\infty} A_n\right) = 1 \quad \forall N \in \mathbb{N}$$
et donc en faisant $N \longrightarrow \infty$, il vient $\mathbb{P}(\limsup A_n) = 1$.

□

Exemple 8. Soit (X_n) une suite de v.a. i.i.d. de loi $\mathcal{N}(0,1)$ et $S_n = X_1 + \ldots + X_n$. On a
$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log n}} \leqslant 1 \quad \text{p.s.}$$

Terminons par un résultat qui vient en quelque sorte généraliser le lemme de Borel-Cantelli.

Définition 31. Soit $(\mathcal{T}_n)$ une suite de tribus sur $(\Omega, \mathcal{A}, \mathbb{P})$. On appelle tribu terminale la tribu
$$\bigcap_{n\in\mathbb{N}} \sigma(\mathcal{T}_n, \mathcal{T}_{n+1}, \ldots) \ .$$

Théorème 8 (Loi du $0-1$ de Kolmogorov)**.** *Si $(\mathcal{T}_n)$ est une suite de tribus indépendantes sur $(\Omega, \mathcal{A}, \mathbb{P})$, alors tout évènement A de la tribu terminale vérifie $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.*

8 Fonction Caractéristique

Définition 32. Si X est un vecteur aléatoire dans $\mathbb{R}^d$, on définit sa fonction caractéristique par

$$\forall \xi \in \mathbb{R}^d, \quad \Phi_X(\xi) = \mathbb{E}[e^{i\xi X}] = \int_{\mathbb{R}^d} e^{i\xi x} P_X(dx) \; .$$

Proposition 7. *La fonction caractéristique Φ_X est une fonction continue bornée sur $\mathbb{R}^d$. De plus, si X est à densité, elle tend vers zéro à l'infini.*

Remarque 4. Au signe près sur ξ, Φ_X est la transformée de Fourier de P_X. Par exemple, si P_X est à densité f,

$$\Phi_X(-\xi) = \int_{\mathbb{R}^d} e^{-i\xi x} f(x) dx = \widehat{f}(\xi) \; ,$$

et dans ce cas, Φ_X tend vers zéro à l'infini.

Exemple 9.

- La fonction caractéristique de la loi $\mathcal{E}(\lambda)$ est $\xi \mapsto \frac{\lambda}{\lambda - i\xi}$.
- La fonction caractéristique de la loi $\mathcal{N}(m, \sigma^2)$ est $\xi \mapsto e^{im\xi} e^{-\frac{\sigma^2 x^2}{2}}$.

Proposition 8. *Si X, Y sont deux vecteurs aléatoires de $\mathbb{R}^d$ indépendants, alors*

$$\forall \xi \in \mathbb{R}^d, \quad \Phi_{X+Y}(\xi) = \Phi_X(\xi)\Phi_Y(\xi) \; .$$

Proposition 9. *Soit X une v.a.r. admettant un moment d'ordre $k \in \mathbb{N}$. Alors Φ_X est de classe $\mathscr{C}^k$ et*

$$\Phi_X^{(k)}(0) = i^k \mathbb{E}[X^k] \; .$$

Remarque 5. Les moments de X sont donc à $\frac{i^k}{k!}$ près les coefficients du d.l. en zéro de Φ_X.

Théorème 9. *La fonction caractéristique d'un vecteur aléatoire de $\mathbb{R}^d$ caractérise sa loi.*

Conséquence : Si (X, Y) est un vecteur aléatoire de $\mathbb{R}^n \times \mathbb{R}^p$, alors X et Y sont indépendantes si et seulement si

$$\forall \xi \in \mathbb{R}^n, \; \forall \zeta \in \mathbb{R}^p, \quad \Phi_{(X,Y)}(\xi, \zeta) = \Phi_X(\xi)\Phi_Y(\zeta) \; .$$

Définition 33. Soient X un vecteur aléatoire de $\mathbb{R}^d$, $m \in \mathbb{R}^d$ et $\Gamma \in \mathbb{R}^{d\times d}$ une matrice symétrique positive.
On dit que X est un vecteur gaussien de moyenne m et de matrice de covariance Γ si

$$\forall \xi \in \mathbb{R}^d, \quad \Phi_X(\xi) = \exp\left(i\xi^T m - i\frac{\xi^T \Gamma \xi}{2} \right) \; .$$

Dans le cas où Γ est inversible (i.e. lorsqu'elle est symétrique définie positive), le vecteur gaussien X admet une densité qui s'écrit

$$\frac{1}{\sqrt{\det(2\pi\Gamma)}} \exp\left(-\frac{1}{2}(x-m)^T \Gamma^{-1}(x-m) \right) \; .$$

Dans le cas contraire il est dit dégénéré.

Proposition 10.

- *Un vecteur aléatoire X de $\mathbb{R}^d$ est gaussien si et seulement si toute combinaison linéaire $\xi \cdot X$ avec $\xi \in \mathbb{R}^d$ suit une loi gaussienne (unidimensionnelle), éventuellement dégénérée (i.e. constante p.s.).*
- *Soit un vecteur aléatoire gaussien (X, Y) de $\mathbb{R}^n \times \mathbb{R}^p$. Alors X et Y sont indépendantes si et seulement si la matrice de covariance du couple s'écrit*

$$\Gamma_{(X,Y)} = \left(\begin{array}{c|c} \Gamma_X & 0 \\ \hline 0 & \Gamma_Y \end{array} \right) \; .$$

9 Convergence de Suites de Variables Aléatoires

9.1 Convergence p.s., en probabilité, et dans L^p

Soient $(X_n), X$ des v.a. définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $\mathbb{R}$ (ou $\mathbb{R}^d$).

Définition 34. On dit que

- $X_n \to X$ p.s. si $\mathbb{P}(X_n \to X) = 1$.
- $X_n \to X$ en probabilité si pour tout $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow[n\to\infty]{} 0$.
- $X_n \to X$ dans L^p si $\|X_n - X\|_p \xrightarrow[n\to\infty]{} 0$.

Remarque 6.

- $\{X_n \to X\}$ est bien un évènement car il s'écrit $\displaystyle\bigcap_{p\in\mathbb{N}^*} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geqslant N} \left\{ |X_n - X| < \frac{1}{p} \right\}$.
- On rappelle que si $p < \infty$, $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$.
  En revanche, si $p = \infty$, $\|X\|_\infty$ est la plus petite constante qui majore $|X|$ presque sûrement.

Proposition 11. *Notons $L^0(\Omega, \mathcal{A}, \mathbb{P})$ l'espace des v.a.r. quotienté par la relation d'égalité presque sûre. Alors $L^0(\Omega, \mathcal{A}, \mathbb{P})$ est un espace complet pour la distance*

$$d(X, Y) = \mathbb{E}\Big[\min(1, |X - Y|)\Big] .$$

De plus, la notion de convergence associée est la convergence en probabilité.

Démonstration. D'abord d est bien une distance. On vérifie aisément la symétrie et l'inégalité triangulaire. Ensuite, si $d(X, Y) = 0$, alors $\min(1, |X - Y|) = 0$ p.s c'est-à-dire $X = Y$ p.s. et donc $X = Y$ dans L^0.

Ensuite L^0 est complet pour cette distance. En effet, si (X_n) est une suite de Cauchy pour d, on peut extraire une sous-suite $(X_{\varphi(n)})$ de telle sorte que pour tout n

$$d(X_{\varphi(n+1)}, X_{\varphi(n)}) \leqslant \frac{1}{2^n} .$$

donc $\sum \min(1, |X_{\varphi(n+1)} - X_{\varphi(n)}|) < \infty$ p.s. d'où $\sum |X_{\varphi(n+1)} - X_{\varphi(n)}| < \infty$ p.s. Autrement dit, presque sûrement la série $\sum X_{\varphi(n+1)} - X_{\varphi(n)}$ converge absolument, donc converge, ce qui donne que $(X_{\varphi(n)})$ converge p.s. vers une v.a. X. Par convergence dominée on a

$$d(X_{\varphi(n)}, X) = \mathbb{E}\Big[\min(1, |X_{\varphi(n)} - X|)\Big] \longrightarrow 0 .$$

Comme (X_n) est de Cauchy, ceci suffit à montrer que (X_n) tend vers X dans L^0.

Il reste à montrer que $d(X_n, X) \to 0$ équivaut à dire que $X_n \to X$ en probabilité. Supposons que $d(X_n, X) \to 0$. Alors, pour tout $\varepsilon > 0$,

$$d(X_n, X) \geqslant \mathbb{E}[\min(1, |X_n - X|)\mathbf{1}_{|X_n - X| > \varepsilon}] \geqslant \varepsilon \mathbb{P}(|X_n - X| > \varepsilon)$$

donc le membre de droite tend vers zéro quand $n \to \infty$. Réciproquement, si $X_n \to X$ en probabilité, alors en prenant $\varepsilon > 0$,

$$d(X_n, X) = \mathbb{E}[\min(1, |X_n - X|)\mathbf{1}_{|X_n - X| > \varepsilon}] + \mathbb{E}[\min(1, |X_n - X|)\mathbf{1}_{|X_n - X| \leqslant \varepsilon}] \leqslant \mathbb{P}(|X_n - X| > \varepsilon) + \varepsilon$$

et donc il existe n_0 tel que pour $n > n_0$, $d(X_n, X) < 2\varepsilon$, ce qui prouve que $d(X_n, X) \to 0$. $\square$

Corollaire 1. *Si $X_n \to X$ en probabilité, alors il existe une sous-suite $(X_{\varphi(n)})$ qui converge vers X p.s.*

Proposition 12.

1. *Si $X_n \to X$ p.s., alors $X_n \to X$ en probabilité.*
2. *Si $X_n \to X$ dans L^p, alors $X_n \to X$ en probabilité.*

3. *Si $X_n \to X$ en probabilité et si (X_n) est bornée dans L^r avec $r \in]1, \infty[$, alors $X_n \to X$ dans L^p, $\forall p \in [1, r[$.*

Démonstration.

1. Si $X_n \to X$ p.s. alors $d(X_n, X) = \mathbb{E}[\min(1, |X_n - X|)] \to 0$ par convergence dominée.
2. Il suffit de voir que $d(X, Y) \leqslant \mathbb{E}[|X - Y|] \leqslant \|X - Y\|_p$.
3. Pour $\varepsilon > 0$ fixé, en décomposant sur $\{|X_n - X| > \varepsilon\}$ et son complémentaire, on peut majorer

$$\|X_n - X\|_p^p \leqslant \mathbb{E}[|X_n - X|^p \mathbf{1}_{|X_n - X| > \varepsilon}] + \varepsilon^p \ .$$

Mais alors, l'inégalité de Hölder nous donne

$$\begin{aligned}\mathbb{E}[|X_n - X|^p \mathbf{1}_{|X_n - X| > \varepsilon}] &\leqslant \left(\mathbb{E}\left[|X_n - X|^{p\frac{r}{p}}\right]\right)^{\frac{p}{r}} \mathbb{P}(|X_n - X| > \varepsilon)^{1-\frac{p}{r}} \\ &= \|X_n - X\|_r^p \mathbb{P}(|X_n - X| > \varepsilon)^{1-\frac{p}{r}}\end{aligned}$$

Le premier terme est en fait borné car (X_n) est bornée dans L^r et car X est aussi dans L^r. En effet, on peut extraire une sous-suite $(X_{\varphi(n)})$ qui converge p.s. vers X, et alors le lemme de Fatou donne

$$\mathbb{E}[|X|^r] \leqslant \liminf \mathbb{E}[|X_{\varphi(n)}|^r] \ .$$

□

Proposition 13. *Supposons que $(X_n - X)$ soient des v.a. indépendantes.*
Alors $X_n \to X$ p.s. si et seulement si pour tout $\varepsilon > 0$, $\sum \mathbb{P}\big(|X_n - X| > \varepsilon\big) < \infty$.

Exemple 10. On considère des v.a. (X_n) indépendantes, $X_n \sim \mathcal{B}(\varepsilon_n)$, avec $0 < \varepsilon_n < 1$.

- $X_n \to 0$ en probabilité et dans L^p ($p < \infty$) si et seulement si $\varepsilon_n \to 0$.
- $X_n \to 0$ p.s. si et seulement si $\sum \varepsilon_n < \infty$.

Proposition 14. *Soient (X_n), (Y_n) deux suites de v.a. qui convergent respectivement vers X, Y en probabilité. Alors $X_n + Y_n \to X + Y$ en probabilité et $X_n Y_n \to XY$ en probabilité.*

Proposition 15. *Supposons que les X_n sont des v.a.r. L^2 et qu'il existe une constante a telle que*

$$\mathbb{E}[X_n] \longrightarrow a \quad et \quad \mathrm{Var}(X_n) \longrightarrow 0 \ .$$

Alors $X_n \to a$ dans L^2.

Démonstration. Il suffit d'écrire

$$\mathbb{E}[(X_n - a)^2] = \mathbb{E}[(X_n - \mathbb{E}[X_n] + \mathbb{E}[X_n] - a)^2] = \mathrm{Var}(X_n) + (\mathbb{E}[X_n] - a)^2$$

□

9.2 Lois des Grands Nombres

Théorème 10 (Loi forte des grands nombres)**.** *Soit (X_n) une suite de v.a. i.i.d. intégrables d'espérance m. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{} m \quad p.s.$$

Remarque 7. En statistique, on reformule ce résultat en disant que $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ est un estimateur fortement consistant de m.

Théorème 11 (Loi faible des grands nombres)**.** *Soit (X_n) une suite de v.a.r. i.i.d. dans L^2. Alors*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{} m \quad dans\ L^2 \ .$$

Démonstration. On a
$$\mathbb{E}\left[(\overline{X}_n - m)^2\right] = \mathrm{Var}(\overline{X}_n) = \frac{1}{n}\mathrm{Var}(X_1) \to 0 \ .$$
□

Remarque 8. Ce résultat se généralise facilement au cas où les X_n sont des v.a.r. de même espérance avec (X_n) bornée dans L^2 et telles que $\mathrm{Cov}(X_i, X_j) = 0$ si $i \neq j$.

Théorème 12. *Soit (X_n) une suite de v.a.r. i.i.d.*
On suppose que les X_n sont dans L^4 et on note $m = \mathbb{E}[X_i]$, alors
$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{} m \quad p.s.$$

Démonstration. On commence par se restreindre au cas où les X_i sont centrées. Il s'agit ensuite de majorer
$$\mathbb{E}\left[\left(\frac{X_1 + \cdots + X_n}{n}\right)^4\right] = \frac{1}{n^4}\sum_{1\leqslant i,j,k,\ell\leqslant n} \mathbb{E}[X_i X_j X_k X_\ell] \leqslant \frac{C}{n^2} \ .$$
ce qui entraîne que $\frac{X_1+\cdots+X_n}{n} \to 0$ p.s. □

Définition 35. Soit (X_n) une suite de v.a.r. i.i.d. de même loi de fonction de répartition F. On appelle fonction de répartition empirique associée au n-échantillon $(X_1, \ldots, X_n)$ la fonction F_n définie par
$$\forall (x,\omega) \in \mathbb{R}\times\Omega, \quad F_n(x,\omega) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i(\omega)\leqslant x} \ .$$

Théorème 13 (Glivenko-Cantelli)**.** *Soit (X_n) une suite de v.a.r. i.i.d. de même loi de fonction de répartition F. Alors presque sûrement, $F_n(\cdot,\omega)$ converge uniformément vers F, c-est-à-dire que pour presque tout ω*
$$\sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| \xrightarrow[n\to\infty]{} 0 \ .$$

9.3 Convergence en loi

Soient $(X_n), X$ des vecteurs aléatoires de $\mathbb{R}^d$ (non nécessairement définis sur le même Ω).

Définition 36. On dit qu'une suite (μ_n) de mesures de probabilité sur $\mathbb{R}^d$ converge étroitement vers une mesure de probabilité μ sur $\mathbb{R}^d$ si
$$\forall \varphi \in \mathscr{C}_b(\mathbb{R}^d), \quad \int \varphi d\mu_n \xrightarrow[n\to\infty]{} \int \varphi d\mu \ .$$

Définition 37. On dit que $X_n \to X$ en loi si $P_{X_n} \to P_X$ étroitement, ce qui revient à dire que pour toute fonction $\varphi : \mathbb{R}^d \to \mathbb{R}$ continue bornée,
$$\mathbb{E}[\varphi(X_n)] \longrightarrow \mathbb{E}[\varphi(X)] \ .$$

Théorème 14. *Soit H un sous-espace de $\mathscr{C}_b(\mathbb{R}^d)$ dont l'adhérence pour la norme uniforme contient $\mathscr{C}_c(\mathbb{R}^d)$. Alors $X_n \to X$ en loi si et seulement si*
$$\forall \varphi \in H, \quad \mathbb{E}[\varphi(X_n)] \longrightarrow \mathbb{E}[\varphi(X)] \ .$$

Théorème 15. *Supposons que les X_n et X sont à valeurs dans $\mathbb{R}$. Alors $X_n \to X$ en loi si et seulement si*
$$F_{X_n}(t) \xrightarrow[n\to\infty]{} F_X(t)$$
en tout point t où F_X est continue.

Démonstration. Supposons que $X_n \to X$ en loi fixons $t \in \mathbb{R}$ où F_X est continue. Donnons-nous $\varepsilon > 0$ et considérons une fonction φ_ε qui vaut 1 sur $]-\infty, t]$, qui vaut 0 sur $[t+\varepsilon, +\infty[$ et qui est affine sur $[t, t+\varepsilon]$. On a alors

$$\mathbb{P}(X_n \leqslant t) \leqslant \mathbb{E}[\varphi_\varepsilon(X_n)]$$

donc en passant à la limite

$$\limsup \mathbb{P}(X_n \leqslant t)\mathbb{E}[\varphi_\varepsilon(X)] \leqslant \mathbb{P}(X \leqslant t+\varepsilon) = F_X(t+\varepsilon) \ .$$

De même on traite la liminf, et on en déduit

$$F_X(t-\varepsilon) \leqslant \liminf \mathbb{P}(X_n \leqslant t) \leqslant \limsup \mathbb{P}(X_n \leqslant t) \leqslant F_X(t+\varepsilon) \ .$$

Comme F_X est continue en t, en faisant $\varepsilon \to 0$, on en déduit que $\lim \mathbb{P}(X_n \leqslant t) = F_X(t)$.

Réciproquement, considérons $\varphi \in \mathscr{C}_c(\mathbb{R})$ à support inclus dans $[a,b]$. Sur $[a,b]$, φ est limite uniforme de fonctions en escalier. De plus, comme F est croissante, elle admet au plus une infinité dénombrable de points de discontinuité. Par suite, on peut construire une fonction en escalier $g = \sum_{j=1}^k a_j \mathbf{1}_{]x_j, x_{j+1}]}$ telle que $\|\varphi - g\|_\infty < \varepsilon$ et telle que les x_j sont des points de continuité de F_X. Alors

$$|\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| \leqslant 2\varepsilon + |\mathbb{E}[g(X_n) - g(X)]| \ .$$

Mais alors

$$\mathbb{E}[g(X_n)] = \sum_{j=1}^k a_j F_{X_n}(x_{j+1}) - F_{X_n}(x_j) \longrightarrow \sum_{j=1}^k F_X(x_{j+1}) - F_X(x_j) = \mathbb{E}[g(X)] \ .$$

Par suite,

$$\limsup |\mathbb{E}[\varphi(X_n)] - \mathbb{E}[\varphi(X)]| \leqslant 2\varepsilon \ ,$$

et comme ε est arbitraire, ceci donne que $\mathbb{E}[\varphi(X_n)] \to \mathbb{E}[\varphi(X)]$. □

Exemple 11. Soit (U_n) une suite de v.a. i.i.d. de loi $\mathcal{U}([0,1])$ et $M_n = \max_{1 \leqslant i \leqslant n} U_i$.

Alors $n(1-M_n) \longrightarrow \mathcal{E}(1)$ en loi.

Remarque 9. Si $X_n = \frac{1}{n}$ p.s., alors F_{X_n} converge partout sauf en zéro vers la fonction $\mathbf{1}_{[0,+\infty[}$ qui est bien la fonction de répartition de la v.a. égale à 0 p.s.

Proposition 16. *Supposons que les X_n et X sont à valeurs dans $\mathbb{N}$. Alors $X_n \to X$ en loi si et seulement si*

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X_n = k) \xrightarrow[n\to\infty]{} \mathbb{P}(X = k) \ .$$

Exemple 12. Soit X_n de loi binomiale $\mathcal{B}(n, p_n)$ avec $np_n \to \lambda > 0$.
Alors X_n converge en loi vers une v.a. de loi $\mathcal{P}(\lambda)$. Pour cette raison, on considère que $\mathcal{P}(np)$ est une bonne approximation de $\mathcal{B}(n,p)$ lorsque n est grand, et p assez petit.

Proposition 17.

1. *Si $X_n \to X$ en probabilité, alors $X_n \to X$ en loi.*
2. *La réciproque est vraie si X est une constante.*

Théorème 16 (Théorème de Lévy)**.** *On a $X_n \to X$ en loi si et seulement si*

$$\forall \xi \in \mathbb{R}^d, \quad \Phi_{X_n}(\xi) \xrightarrow[n\to\infty]{} \Phi_X(\xi) \ .$$

On prendra garde au fait que la convergence en loi n'est en général pas compatible avec l'addition. En effet, $X_n \to X$ en loi et $Y_n \to Y$ en loi n'implique pas que $X_n + Y_n \to X+Y$ en loi, ni que $(X_n, Y_n) \to (X,Y)$ en loi. Néanmoins...

Proposition 18. *Si $X_n \to X$ en loi, alors $f(X_n) \to f(X)$ en loi dès que f est continue sur $\mathbb{R}^d$.*

Théorème 17 (Slutsky)**.** *Soient (X_n), (Y_n) deux suites de v.a.r. définies sur $(\Omega, \mathcal{A}, \mathbb{P})$ telles que $X_n \longrightarrow X$ en loi et $Y_n \longrightarrow a$ en loi (avec a une constante). Alors*

- $(X_n, Y_n) \longrightarrow (X, a)$ *en loi,*
- $X_n + Y_n \longrightarrow X + a \quad$ *et* $\quad X_n Y_n \longrightarrow aX$ *en loi.*

Démonstration. Le deuxième point s'obtient à partir du premier par composition avec les fonctions continues "somme" et "produit". Il suffit donc de montrer le premier point, ce qui se fait avec la fonction caractéristique du couple

$$\left|\mathbb{E}[e^{isX_n+itY_n}] - \mathbb{E}[e^{isX+ita}]\right| \leqslant \mathbb{E}[|e^{itY_n} - e^{ita}|] + |\mathbb{E}[e^{itX_n} - e^{itX}]|$$

□

9.4 Théorème Central Limite

Théorème 18. *Soient (X_n) une suite de v.a.r. i.i.d. dans L^2 d'espérance m et de variance σ^2. Alors*

$$\frac{\sum_{k=1}^n X_k - nm}{\sqrt{n}} \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma^2) \quad \textit{en loi.}$$

Démonstration. D'abord on peut supposer X_n centrée en ôtant son espérance. Ensuite on applique le théorème de Lévy en calculant la fonction caractéristique de

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \ .$$

Avec l'indépendance des X_k, le calcul donne que pour tout $\xi \in \mathbb{R}$,

$$\Phi_{Z_n}(\xi) = \prod_{k=1}^n \mathbb{E}\left[e^{i\xi\frac{X_k}{\sqrt{n}}}\right] = \left(\Phi_{X_1}\Big(\frac{\xi}{\sqrt{n}}\Big)\right)^n \ .$$

De plus, X_1 étant L^2, Φ_{X_1} est deux fois dérivable et son développement limité en zéro donne s'écrit

$$\Phi_{X_1}\left(\frac{\xi}{\sqrt{n}}\right) = 1 - \frac{\sigma^2\xi^2}{2n} + o\left(\frac{1}{n}\right) \ .$$

En particulier, à ξ fixé, ces termes sont dans $D(1,1)$ à partir d'un certain rang. En utilisant la déterminant principale du logarithme complexe, on a donc

$$\log \Phi_{Z_n}(\xi) = n \log\left(1 - \frac{\sigma^2\xi^2}{2n} + o\Big(\frac{1}{n}\Big)\right) = -\frac{\sigma^2\xi^2}{2} + o(1)$$

ce qui donne que

$$\Phi_{Z_n}(\xi) \longrightarrow e^{-\frac{\sigma^2\xi^2}{2}} \ .$$

On reconnait la fonction caractéristique de la loi $\mathcal{N}(0,1)$. □

Remarque 10. Autrement dit, $\sqrt{n}(\overline{X}_n - m) \longrightarrow \mathcal{N}(0, \sigma^2)$ en loi ce qui se reformule par

$$\forall a > 0, \quad \mathbb{P}\left(m \in \left[\overline{X}_n - \frac{a\sigma}{\sqrt{n}}, \overline{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-a}^{a} e^{\frac{-x^2}{2}} dx \ .$$

En statistique, si l'on connaît la valeur de σ, en prenant a le quantile d'ordre 0.975 de la loi normale ($a \approx 1.96$), on en déduit que $\left[\overline{X}_n - \frac{a\sigma}{\sqrt{n}}, \overline{X}_n + \frac{a\sigma}{\sqrt{n}}\right]$ un intervalle de confiance asymptotique sur m de niveau 0.95.

Exemple 13. Considérons $Y_n \sim \mathcal{B}(n, p)$. Alors Y_n a même loi qu'une somme $X_1 + \ldots + X_n$ de v.a. i.i.d. de loi $\mathcal{B}(p)$. De la conclusion du théorème central limite appliqué aux X_i, on déduit que $\frac{Y_n - np}{\sqrt{n}}$ converge vers la loi $\mathcal{N}(0, p(1-p))$. Pour cette raison, on considère que $\mathcal{N}(np, np(1-p))$ est une bonne approximation de $\mathcal{B}(p)$ lorsque n est grand, et p ni trop proche de 0 ou de 1.

Remarque 11. Le théorème central limite se reformule en disant que la fonction de répartition F_n de

$$Z_n = \frac{\sum_{k=1}^{n} X_k - nm}{\sqrt{n}}$$

converge simplement vers la fonction de répartition de la loi $\mathcal{N}(0, 1)$. Comme toutes ces fonctions sont croissantes et admettent respectivement 0 et 1 pour limites en $-\infty$ et $+\infty$, en utilisant le théorème de Dini on peut montrer qu'il y a toujours convergence *uniforme* de F_n vers F.