A

Rappels de probabilités

Cet appendice rappelle les définitions et énonce sans démonstration les résultats enseignés dans un cours d'initiation aux probabilités en première année d'école d'ingénieur ou en troisième année du cycle Licence. On peut trouver les démonstrations de ces résultats dans les ouvrages généraux [2, 6 et 7] ou les ouvrages plus spécialisés [1, 3 et 4].

A.1 Variables aléatoires

A.1.1 Espace de probabilité

Définition A.1.1. *Soit Ω un ensemble. Un sous-ensemble $\mathcal{A}$ de l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω est une tribu si*

- $\emptyset, \Omega \in \mathcal{A}$.
- *$\mathcal{A}$ est stable par passage au complémentaire : si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$.*
- *$\mathcal{A}$ est stable par réunion et intersection dénombrables : si pour tout $i \in \mathbb{N}$, $A_i \in \mathcal{A}$ alors $\bigcup_{i\in\mathbb{N}} A_i$ et $\bigcap_{i\in\mathbb{N}} A_i$ sont dans $\mathcal{A}$.*

Exemple A.1.2.

- $\mathcal{P}(\Omega)$ est une tribu appelée tribu discrète.
- On appelle tribu borélienne de $\mathbb{R}^d$, et on note $\mathcal{B}(\mathbb{R}^d)$, la plus petite tribu contenant tous les ouverts de $\mathbb{R}^d$. Comme l'intersection d'une famille quelconque de tribus est une tribu, $\mathcal{B}(\mathbb{R}^d)$ s'obtient comme l'intersection de toutes les tribus contenant les ouverts de $\mathbb{R}^d$.

$\diamondsuit$

Définition A.1.3. *Soit Ω un ensemble muni d'une tribu $\mathcal{A}$. On appelle probabilité sur $(\Omega, \mathcal{A})$ une application $\mathbb{P} : \mathcal{A} \to [0, 1]$ qui vérifie les deux propriétés suivantes :*

- $\mathbb{P}(\Omega) = 1$,
- *Pour toute famille $(A_i)_{i\in I}$ dénombrable d'éléments disjoints de $\mathcal{A}$,*

$$\mathbb{P}\Big(\bigcup_{i\in I} A_i\Big) = \sum_{i\in I} \mathbb{P}(A_i) \text{ \textit{(propriété appelée σ-additivité).}}$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ s'appelle un espace de probabilité.

Exemple A.1.4. Si E est un ensemble fini, on appelle probabilité uniforme sur $(E, \mathcal{P}(E))$ la probabilité $\mathbb{P}$ qui à $A \in \mathcal{P}(E)$ associe $\mathbb{P}(A) = \text{Card}\,(A)/\text{Card}\,(E)$. ♢

Remarque A.1.5. La propriété de σ-additivité implique la propriété de monotonie suivante : si $(A_n, n \in \mathbb{N})$ est une suite d'éléments de $\mathcal{A}$ croissante au sens où pour tout $n \in \mathbb{N}$, $A_n \subset A_{n+1}$, alors

$$\mathbb{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \lim_{n\to+\infty} \nearrow \mathbb{P}(A_n).$$

♢

A.1.2 Variables aléatoires

Définition A.1.6. *Soit F et G deux ensembles munis respectivement d'une tribu $\mathcal{F}$ et d'une tribu $\mathcal{G}$. Une application $f : F \to G$ est dite mesurable de $(F, \mathcal{F})$ dans $(G, \mathcal{G})$ si pour tout $B \in \mathcal{G}$, l'image réciproque $f^{-1}(B)$ de B par f appartient à $\mathcal{F}$.*

Exemple A.1.7. Toute application continue de $\mathbb{R}^d$ dans $\mathbb{R}^k$ est mesurable de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ dans $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. ♢

Définition A.1.8. *Soit Ω muni d'une tribu $\mathcal{A}$. On appelle variable aléatoire à valeurs dans $\mathbb{R}^d$ toute application X mesurable de $(\Omega, \mathcal{A})$ dans $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Lorsque $d = 1$, on parle de variable aléatoire réelle.*

Exemple A.1.9. Si $A \in \mathcal{A}$, la fonction $\mathbf{1}_A$ définie par

$$\forall \omega \in \Omega,\ \mathbf{1}_A(\omega) = \begin{cases} 1 \text{ si } \omega \in A, \\ 0 \text{ sinon,} \end{cases}$$

est une variable aléatoire réelle. ♢

Remarque A.1.10. Comme la composée de deux applications mesurables est mesurable, si X est une variable aléatoire à valeurs dans $\mathbb{R}^d$, et f une application mesurable de $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ dans $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, alors $Y = f(X)$ est une variable aléatoire à valeurs dans $\mathbb{R}^k$. ♢

A.1.3 Espérance

Désormais on suppose que l'on s'est donné un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. Les variables aléatoires que l'on considère sont définies sur $(\Omega, \mathcal{A})$.

Cas des variables aléatoires positives

Définition A.1.11.

- *On appelle variable aléatoire positive une application X de Ω dans $\mathbb{R} \cup \{+\infty\}$ telle que pour tout B dans la plus petite tribu contenant $\mathcal{B}(\mathbb{R})$ et $\{+\infty\}$, $X^{-1}(B) \in \mathcal{A}$ et telle que $\mathbb{P}(X \in \mathbb{R}_+ \cup \{+\infty\}) = 1$.*
- *Si une variable aléatoire positive X prend un nombre fini de valeurs $\{x_1, \ldots, x_k\} \subset \mathbb{R}_+ \cup \{+\infty\}$ au sens où $\mathbb{P}(X \in \{x_1, \ldots, x_k\}) = 1$, on définit son espérance par*

$$\mathbb{E}[X] = \sum_{i=1}^{k} x_i \mathbb{P}(X = x_i).$$

Remarque A.1.12. La notion d'espérance formalise l'idée de valeur moyenne. Elle prolonge également la notion de probabilité car pour tout $A \in \mathcal{A}$, $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$. $\Diamond$

On étend l'espérance à toute variable aléatoire X positive en «approchant» X par la suite croissante $(X_n, n \in \mathbb{N}^*)$ où, pour $n \in \mathbb{N}^*$,

$$X_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbf{1}_{\{\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\}} + n\mathbf{1}_{\{X \geq n\}}$$

prend un nombre fini de valeurs. La variable aléatoire X_n a pour espérance

$$\mathbb{E}[X_n] = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{P}\left(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right) + n\mathbb{P}(X \geq n).$$

La suite $(\mathbb{E}[X_n], n \in \mathbb{N}^*)$ est croissante et on définit l'espérance de X par

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n].$$

L'espérance de X peut être égale à $+\infty$. C'est le cas par exemple si $\mathbb{P}(X = +\infty) > 0$ puisqu'alors pour tout $n \in \mathbb{N}^*$, $\mathbb{E}[X_n] \geq n\mathbb{P}(X = +\infty)$. Par contraposée, une variable aléatoire positive d'espérance finie est finie avec probabilité 1.
Il est facile de vérifier que si deux variables aléatoires positives X et Y sont telles que $\mathbb{P}(X \leq Y) = 1$ alors pour tout $n \in \mathbb{N}^*$, $\mathbb{E}[X_n] \leq \mathbb{E}[Y_n]$ ce qui implique que $\mathbb{E}[X] \leq \mathbb{E}[Y]$ (propriété de croissance).

Cas des variables aléatoires de signe quelconque

Pour $x \in \mathbb{R}$, on pose $x^+ = \max(x, 0)$ et $x^- = \max(-x, 0)$. Notons que $x = x^+ - x^-$ et que $\max(x^+, x^-) = |x|$.

Définition A.1.13. *Une variable aléatoire réelle X est dite intégrable si $\mathbb{E}[|X|] < +\infty$. Dans ce cas, $\mathbb{E}[X^+] < +\infty$ et $\mathbb{E}[X^-] < +\infty$ et on définit l'espérance de X par*

$$\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-].$$

Le théorème suivant regroupe des propriétés importantes de l'espérance.

Théorème A.1.14.

1. *L'espérance est linéaire : si X et Y sont deux variables aléatoires réelles intégrables alors pour tout $\lambda \in \mathbb{R}$, $X + \lambda Y$ est intégrable et*

$$\mathbb{E}[X + \lambda Y] = \mathbb{E}[X] + \lambda \mathbb{E}[Y].$$

2. *Si Y est une variable aléatoire réelle intégrable, alors toute variable aléatoire réelle X telle que $\mathbb{P}(|X| \leq Y) = 1$ est intégrable.*
3. *L'espérance est croissante : si X et Y sont deux variables aléatoires réelles intégrables telles que $\mathbb{P}(X \leq Y) = 1$, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.*

A.1.4 Convergence des espérances

Définition A.1.15. *Une suite $(X_n, n \in \mathbb{N}^*)$ de variables aléatoires réelles* ***converge presque sûrement*** *(p.s.) vers une variable aléatoire réelle X si et seulement si $\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$.*

Lorsque $(X_n, n \in \mathbb{N}^*)$ converge presque sûrement vers X, on est souvent confronté à la question naturelle de savoir si $(\mathbb{E}[X_n], n \in \mathbb{N}^*)$ converge vers $\mathbb{E}[X]$. Les théorèmes suivants énoncent des conditions sous lesquelles la réponse est affirmative.
Comme l'espérance d'une variable aléatoire positive X a été introduite comme la limite des espérances de variables aléatoires X_n qui croissent vers X lorsque n tend vers l'infini, le résultat suivant est naturel.

Théorème A.1.16 (Convergence monotone). *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires* ***positives*** *croissante au sens où pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(X_n \leq X_{n+1}) = 1$. Alors on a*

$$\mathbb{E}\left[\lim_{n\to\infty} \nearrow X_n\right] = \lim_{n\to\infty} \nearrow \mathbb{E}[X_n],$$

où les deux membres peuvent prendre simultanément la valeur $+\infty$.

On en déduit le résultat suivant.

Corollaire A.1.17 (Lemme de Fatou). *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires* ***positives****. On a :*

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Théorème A.1.18 (Convergence dominée). *Soit Y une variable aléatoire positive telle que $\mathbb{E}[Y] < \infty$ et $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires réelles dominées par Y au sens où pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(|X_n| \leq Y) = 1$. Si la suite $(X_n, n \in \mathbb{N}^*)$ converge presque sûrement vers X, alors X est intégrable et*

$$\lim_{n\to+\infty} \mathbb{E}[|X - X_n|] = 0,$$

ce qui implique en particulier que $\boxed{\mathbb{E}[X] = \lim_{n\to+\infty} \mathbb{E}[X_n].}$

A.1.5 Indépendance

Définition A.1.19.

1. *Des variables aléatoires $X_1, \ldots, X_n$ à valeurs respectives dans $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_n}$ sont dites indépendantes si pour toutes fonctions $f_i : \mathbb{R}^{d_i} \to \mathbb{R}$ mesurables et bornées*
$$\mathbb{E}\left[\prod_{i=1}^{n} f_i(X_i)\right] = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)]. \tag{A.1}$$
2. *Une famille quelconque de variables aléatoires est dite indépendante si toute sous-famille finie est indépendante.*

Remarque A.1.20. Si les variables aléatoires $X_1, \ldots, X_n$ sont indépendantes et si $f_i(X_i)$ est intégrable pour $1 \leq i \leq n$, alors $\prod_{i=1}^{n} f_i(X_i)$ est intégrable et (A.1) reste vraie. ♢

Proposition A.1.21. *Soit X et Y deux variables aléatoires indépendantes à valeurs respectives dans $\mathbb{R}^d$ et $\mathbb{R}^k$ et $f : \mathbb{R}^{d+k} \to \mathbb{R}$ une fonction mesurable telle que pour tout x dans $\mathbb{R}^d$, $f(x, Y)$ est intégrable d'espérance $\psi(x)$.*
Alors la fonction ψ est mesurable. En outre, si $f(X, Y)$ est intégrable, $\psi(X)$ est également intégrable et

$$\mathbb{E}[f(X, Y)] = \mathbb{E}[\psi(X)].$$

A.1.6 Variance

Définition A.1.22.

- *Une variable aléatoire réelle X est dite de carré intégrable si X^2 est intégrable (ce qui implique que X est intégrable). On définit alors sa variance notée* $\mathrm{Var}(X)$ *par*

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

- *Soit X, Y deux variables aléatoires réelles de carré intégrable. On définit leur covariance notée* $\mathrm{Cov}(X, Y)$ *par*

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

– *Pour $X = (X_1, \ldots, X_d)$ une variable aléatoire à valeurs dans $\mathbb{R}^d$ dont les coordonnées sont intégrables, on appelle espérance de X et on note $\mathbb{E}[X]$ le vecteur dont les coordonnées sont les espérances des coordonnées de X. Lorsque les coordonnées de X sont de carré intégrable, on appelle matrice de covariance de X et on note $\mathrm{Cov}(X, X)$ la matrice symétrique positive $\Sigma = (\Sigma_{ij}, 1 \le i, j \le d)$ définie par*

$$\Sigma_{ij} = \mathrm{Cov}(X_i, X_j), \ 1 \le i, j \le d.$$

Remarque A.1.23.

– On appelle écart-type d'une variable aléatoire réelle la racine carrée de sa variance.
– La variance et l'écart-type d'une variable aléatoire réelle X mesurent l'étalement de cette variable aléatoire autour de son espérance. En particulier, l'inégalité de Bienaymé-Tchebychev permet de majorer la probabilité pour que X diffère de $\mathbb{E}[X]$ de plus de $a \in \mathbb{R}_+^*$ en fonction de $\mathrm{Var}(X)$:

$$\mathbb{P}(|X - \mathbb{E}[X]| \ge a) \le \frac{\mathrm{Var}(X)}{a^2}. \tag{A.2}$$

– Si X et Y sont des variables aléatoires réelles intégrables indépendantes, alors on a $\mathrm{Cov}(X, Y) = 0$.
– Si X et Y sont de carré intégrable alors XY est intégrable et on a l'inégalité de Cauchy-Schwarz

$$|\mathbb{E}[XY]| \le \sqrt{\mathbb{E}[X^2]} \times \sqrt{\mathbb{E}[Y^2]}.$$

$\lozenge$

Proposition A.1.24. *Soit $X_1, \ldots, X_n$ des variables aléatoires réelles de carré intégrable et $S = X_1 + \cdots + X_n$ leur somme. Alors S est de carré intégrable et*

$$\mathrm{Var}(S) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \le j < i \le n} \mathrm{Cov}(X_i, X_j).$$

En particulier si les variables X_i sont **indépendantes***, on a*

$$\boxed{\mathrm{Var}(S) = \sum_{i=1}^{n} \mathrm{Var}(X_i).}$$

A.1.7 Fonction caractéristique

Définition A.1.25. *Soit $X = (X_1, \ldots, X_d)$ une variable aléatoire à valeurs dans $\mathbb{R}^d$. On appelle fonction caractéristique de X la fonction $\psi_X : \mathbb{R}^d \to \mathbb{C}$ définie par*

$$\boxed{\forall u = (u_1, \ldots, u_d) \in \mathbb{R}^d, \ \psi_X(u) = \mathbb{E}\left[\mathrm{e}^{i(u_1 X_1 + \ldots + u_d X_d)}\right].}$$

Notons que $\forall u \in \mathbb{R}^d$, $|\psi_X(u)| \leq 1$. Comme son nom l'indique, la fonction caractéristique caractérise la loi.

Définition A.1.26. *Soit X et Y deux variables aléatoires à valeurs $\mathbb{R}^d$. On dit que X et Y ont même loi ou sont identiquement distribuées si*

$$\forall f : \mathbb{R}^d \to \mathbb{R} \text{ mesurable bornée}, \ \mathbb{E}[f(X)] = \mathbb{E}[f(Y)].$$

Théorème A.1.27. *Deux variables aléatoires X et Y à valeurs dans $\mathbb{R}^d$ ont même loi si et seulement si elles ont même fonction caractéristique.*

A.1.8 Transformée de Laplace

Définition A.1.28. *Soit X une variable aléatoire positive. On appelle transformée de Laplace de X la fonction*

$$\alpha \in \mathbb{R}_+ \to \mathbb{E}\left[\mathrm{e}^{-\alpha X}\right] \in \mathbb{R}_+,$$

où on utilise la convention $\mathrm{e}^{-\alpha x} = 1$ pour $\alpha = 0$ et $x = +\infty$.

La transformée de Laplace $\alpha \to \mathbb{E}\left[\mathrm{e}^{-\alpha X}\right]$ d'une variable aléatoire X positive est une fonction C^∞ sur $]0, +\infty[$. Enfin si X est finie p.s., la dérivée de la transformée de Laplace en $\alpha \in]0, +\infty[$ est donnée par $-\mathbb{E}\left[X\,\mathrm{e}^{-\alpha X}\right]$.

Une autre propriété très utile de la transformée de Laplace est qu'elle caractérise la loi des variables aléatoires positives.

Théorème A.1.29. *Si X et Y sont deux variables aléatoires positives qui ont même transformée de Laplace, alors elles ont même loi.*

A.1.9 Probabilités conditionnelles

La notion de probabilité conditionnelle d'un événement A sachant un événement B permet de prendre en compte l'information que l'événement B est réalisé dans le poids que l'on donne à l'événement A :

Définition A.1.30. *Soit $A, B \in \mathcal{A}$ et X une variable aléatoire intégrable.*

- *La probabilité conditionnelle de l'événement A sachant l'événement B est notée $\mathbb{P}(A|B)$ et définie par*

$$\boxed{\mathbb{P}(A|B) = \begin{cases} \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} & \text{si } \mathbb{P}(B) > 0, \\ \mathbb{P}(A) & \text{sinon.} \end{cases}} \tag{A.3}$$

- *L'espérance conditionnelle de la variable aléatoire X sachant l'événement B est notée $\mathbb{E}[X|B]$ et définie par*

$$\boxed{\mathbb{E}[X|B] = \begin{cases} \dfrac{\mathbb{E}[X \mathbf{1}_B]}{\mathbb{P}(B)} & \text{si } \mathbb{P}(B) > 0, \\ \mathbb{E}[X] & \text{sinon.} \end{cases}}$$

Remarque A.1.31.

- $\mathbb{P}(\cdot|B)$ est une probabilité sur $(\Omega, \mathcal{A})$ et $\mathbb{E}[\cdot|B]$ l'espérance correspondant à cette probabilité.
- Il est facile de vérifier que si $A, B, C \in \mathcal{A}$ et $\mathbb{P}(C) > 0$, alors

$$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|B \cap C)\mathbb{P}(B|C). \tag{A.4}$$

$\Diamond$

A.2 Lois usuelles

A.2.1 Lois discrètes usuelles

Définition A.2.1. *Soit E un ensemble dénombrable. On appelle variable aléatoire discrète à valeurs dans E toute application X mesurable de $(\Omega, \mathcal{A})$ dans $(E, \mathcal{P}(E))$. On appelle loi de X la famille des nombres $(\mathbb{P}(X = x), x \in E)$. Dans le cas où $E \subset \mathbb{N}$, la variable aléatoire X est dite entière.*

La proposition suivante est très utile pour calculer l'espérance d'une fonction réelle d'une variable aléatoire discrète.

Proposition A.2.2. *Soit X une variable aléatoire discrète à valeurs dans E et f une application mesurable de $(E, \mathcal{P}(E))$ dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Alors la variable aléatoire réelle $f(X)$ est intégrable si et seulement si $\sum_{x \in E} |f(x)|\mathbb{P}(X = x) < +\infty$ et dans ce cas,*

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

Définition A.2.3. *Soit X une variable aléatoire entière. On appelle fonction génératrice de X la fonction $G_X : [0, 1] \to [0, 1]$ définie par*

$$\forall s \in [0, 1],\ G_X(s) = \mathbb{E}\left[s^X\right] = \sum_{k \in \mathbb{N}} s^k \mathbb{P}(X = k).$$

Le théorème suivant regroupe plusieurs propriétés utiles de la fonction génératrice :

Théorème A.2.4.

- *La fonction génératrice G_X est continue, croissante et convexe sur $[0, 1]$.*
- *Elle est C^∞ sur $[0, 1[$.*
- *Elle caractérise la loi de X car pour tout $k \in \mathbb{N}$, $\mathbb{P}(X = k) = G_X^{(k)}(0)/k!$ où $G_X^{(k)}$ désigne la dérivée d'ordre k de G_X.*
- *Pour $n \in \mathbb{N}^*$, l'espérance $\mathbb{E}[X^n] = \sum_{k \in \mathbb{N}} k^n \mathbb{P}(X = k)$ est finie si et seulement si G_X est n fois continûment différentiable sur $[0, 1]$. Dans ce cas, sa dérivée n-ième en 1 est*

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

Définition A.2.5. *On dit que la variable aléatoire entière X suit la loi*
- **de Bernoulli de paramètre** p*, où* $p \in [0,1]$*, si*

$$\mathbb{P}(X=1) = 1 - \mathbb{P}(X=0) = p,$$

- **binomiale de paramètre** (n,p)*, où* $n \in \mathbb{N}^*$ *et* $p \in [0,1]$*, si*

$$\forall k \in \{0,1,\ldots,n\},\ \mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ où } \binom{n}{k} = \frac{n!}{k!(n-k)!},$$

- **de Poisson de paramètre** θ*, où* $\theta > 0$*, si*

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

- **géométrique de paramètre** p*, où* $p \in]0,1]$*, si*

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

Remarque A.2.6. On rappelle que la loi binomiale de paramètre (n,p) est la loi de la somme de n variables de Bernoulli de paramètre p indépendantes et que la loi géométrique de paramètre p est la loi du temps de premier succès dans une suite d'expériences aléatoires indépendantes avec même probabilité de succès p. $\diamondsuit$

Le tableau A.1 rappelle la fonction génératrice, l'espérance, la variance et la fonction caractéristique de chacune de ces lois (pour retrouver l'espérance et la variance des lois géométrique et de Poisson, on peut commencer par calculer leur fonction génératrice et utiliser le théorème A.2.4).

Remarque A.2.7. La transformée de Laplace d'une variable aléatoire entière X est reliée à sa fonction génératrice de moments $G_X(s)$ par l'égalité

$$\forall \alpha \geq 0,\ \mathbb{E}\left[\mathrm{e}^{-\alpha X}\right] = G_X(\mathrm{e}^{-\alpha}).$$

Si par exemple, X suit la loi géométrique de paramètre $p \in]0,1]$, on a pour $\alpha \geq 0$, $\mathbb{E}\left[\mathrm{e}^{-\alpha X}\right] = \dfrac{p\,\mathrm{e}^{-\alpha}}{1-(1-p)\,\mathrm{e}^{-\alpha}}$. $\diamondsuit$

Définition A.2.8. *Soit* $n,k \in \mathbb{N}^*$ *et* $(p_1,\ldots,p_k) \in [0,1]^k$ *tels que* $p_1 + \cdots + p_k = 1$*. On dit que la variable aléatoire* $X = (X_1,\ldots,X_k)$ *à valeurs dans* $\mathbb{N}^k$ *suit la loi multinomiale de paramètre* $(n,p_1,\ldots,p_k)$ *si pour tout* $x = (x_1,\ldots,x_k)$ *dans* $\{0,\ldots,n\}^k$*,*

$$\mathbb{P}(X=x) = \mathbf{1}_{\{x_1+\ldots+x_k=n\}} \frac{n!}{x_1!\cdots x_k!} p_1^{x_1}\cdots p_k^{x_k}.$$

Tableau A.1. Fonction génératrice, espérance, variance et fonction caractéristique des lois discrètes usuelles

Loi	fonction génératrice $G_X(s)$	espérance $\mathbb{E}[X]$	variance $\mathrm{Var}(X)$	fonction caractéristique $\psi_X(u)$
Bernoulli p	$1-p+ps$	p	$p(1-p)$	$1-p+p\,\mathrm{e}^{iu}$
binomiale (n,p)	$(1-p+ps)^n$	np	$np(1-p)$	$(1-p+p\,\mathrm{e}^{iu})^n$
Poisson θ	$\mathrm{e}^{\theta(s-1)}$	θ	θ	$\mathrm{e}^{\theta(\mathrm{e}^{iu}-1)}$
géométrique p	$\dfrac{ps}{1-(1-p)s}$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{p\,\mathrm{e}^{iu}}{1-(1-p)\,\mathrm{e}^{iu}}$

Remarque A.2.9.

– La loi multinomiale de paramètre $(n, p_1, \ldots, p_k)$ est la loi du vecteur constitué des nombres d'objets dans les boîtes d'indices $1, \ldots, k$ lorsque n objets sont rangés indépendamment dans ces boites suivant la probabilité $(p_1, \ldots, p_k)$.
– Si $X = (X_1, \ldots, X_k)$ suit la loi multinomiale de paramètre $(n, p_1, \ldots, p_k)$ alors pour $i \in \{1, \ldots, k\}$, X_i suit la loi binomiale de paramètre (n, p_i). Comme $X_1 + \cdots + X_k = n$, les variables X_i ne sont pas indépendantes.

$\diamondsuit$

A.2.2 Lois à densité usuelles

Définition A.2.10. *On dit que la variable aléatoire X à valeurs dans $\mathbb{R}^d$ possède la densité p où $p : \mathbb{R}^d \to \mathbb{R}$ est une fonction mesurable positive d'intégrale 1 si*

$$\forall f : \mathbb{R}^d \to \mathbb{R} \textit{ mesurable bornée, } \mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x)p(x)dx.$$

Définition A.2.11. *On dit que la variable aléatoire réelle X suit la loi*

– **uniforme sur** $[a,b]$*, où $a < b \in \mathbb{R}$, si X possède la densité*

$$p(x) = \frac{1}{b-a}\mathbf{1}_{\{a \le x \le b\}},$$

– **exponentielle de paramètre** λ, *où* $\lambda > 0$, *si* X *possède la densité*

$$p(x) = \lambda\, \mathrm{e}^{-\lambda x}\, \mathbf{1}_{\{x>0\}},$$

– **de Cauchy de paramètre** a, *où* $a > 0$, *si* X *possède la densité*

$$p(x) = \frac{a}{\pi(x^2+a^2)},$$

– **gaussienne ou normale de paramètre** (m, σ^2) **notée** $\mathcal{N}(m, \sigma^2)$, *où* $m \in \mathbb{R}$ *et* $\sigma > 0$, *si* X *possède la densité*

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-\frac{(x-m)^2}{2\sigma^2}},$$

– **gamma de paramètre** (λ, α) *où* $\lambda, \alpha > 0$ *si* X *possède la densité*

$$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, \mathrm{e}^{-\lambda x}\, \mathbf{1}_{\{x>0\}}$$

où

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}\, \mathrm{e}^{-x}\, dx, \tag{A.5}$$

– **béta de paramètre** (a, b) *où* $a, b > 0$ *si* X *possède la densité*

$$p(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbf{1}_{\{0<x<1\}}.$$

Le tableau A.2 rappelle l'espérance, la variance et la fonction caractéristique de chacune de ces lois.

Remarque A.2.12. On vérifie facilement (par exemple en utilisant la fonction caractéristique) que la somme de n variables exponentielles de paramètre λ indépendantes suit la loi gamma de paramètre (λ, n). ♢

Définition A.2.13. *Soit* $m \in \mathbb{R}^d$ *et* $\Sigma \in \mathbb{R}^{d\times d}$ *une matrice symétrique positive. On dit que la variable aléatoire* $X = (X_1, \ldots, X_d)$ *à valeurs dans* $\mathbb{R}^d$ *suit la loi gaussienne ou normale en dimension* d *de paramètre* (m, Σ), *notée* $\mathcal{N}(m, \Sigma)$, *si*

$$\forall u \in \mathbb{R}^d,\ \psi_X(u) = \mathrm{e}^{i(u,m)-\frac{1}{2}(u,\Sigma u)},$$

où $(.,.)$ *désigne le produit scalaire dans* $\mathbb{R}^d$.

Remarque A.2.14. Si $X = (X_1, \ldots, X_d)$ suit la loi gaussienne de paramètre (m, Σ), alors pour tout $\lambda \in \mathbb{R}^d$, la variable aléatoire (λ, X) suit la loi gaussienne en dimension 1 de paramètre $((\lambda, m), (\lambda, \Sigma\lambda))$. En outre le paramètre s'interprète de la façon suivante : $m = \mathbb{E}[X]$ où $\mathbb{E}[X]$ désigne le vecteur constitué des espérances de chacune des coordonnées de X et $\Sigma = \mathrm{Cov}(X, X)$. ♢

Tableau A.2. Espérance, variance et fonction caractéristique des lois à densité usuelles

Loi	espérance $\mathbb{E}[X]$	variance $\mathrm{Var}(X)$	fonction caractéristique $\psi_X(u)$
uniforme $[a,b]$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{\sin((b-a)u/2)}{(b-a)u/2}\ \mathrm{e}^{iu\frac{a+b}{2}}$
exponentielle λ	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\dfrac{\lambda}{\lambda - iu}$
gaussienne $\mathcal{N}(m,\sigma^2)$	m	σ^2	$\mathrm{e}^{ium-\frac{\sigma^2u^2}{2}}$
Cauchy a	non définie	non définie	$\mathrm{e}^{-a\lvert u\rvert}$
gamma (λ,α)	$\dfrac{\alpha}{\lambda}$	$\dfrac{\alpha}{\lambda^2}$	$\left(\dfrac{\lambda}{\lambda-iu}\right)^{\alpha}$
béta (a,b)	$\dfrac{a}{a+b}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$	Pas de formule explicite

A.2.3 Simulation

La plupart des langages de programmation comportent un générateur de nombres pseudo-aléatoires qui génère une suite de nombres à valeurs dans $[0,1]$ imitant une réalisation d'une suite de variables aléatoires uniformes sur $[0,1]$ indépendantes. Afin de pouvoir simuler des variables aléatoires distribuées suivant les lois usuelles à partir d'un tel générateur, nous rappelons dans le Tableau A.3 comment générer une variable aléatoire de loi donnée à partir d'une variable aléatoire U uniforme sur $[0,1]$ ou bien d'une suite $(U_i, i \geq 1)$ de variables indépendantes et identiquement distribuées suivant la loi uniforme sur $[0,1]$.

En dehors des techniques spécifiques présentées dans ce tableau, on peut citer la méthode d'inversion de la fonction de répartition en dimension 1 qui repose sur la proposition C.5 et la méthode du rejet. Cette dernière permet de simuler un vecteur aléatoire de densité p sur $\mathbb{R}^d$ majorée à une constante multiplicative près par une densité q sur $\mathbb{R}^d$ suivant laquelle on sait simuler. Elle repose sur le résultat suivant :

Tableau A.3. Simulation suivant les lois usuelles

Loi	Méthode de simulation
Bernoulli p	$\mathbf{1}_{\{U \le p\}}$
binomiale (n,p)	$\sum_{i=1}^{n} \mathbf{1}_{\{U_i \le p\}}$
géométrique p	$1 + \left[\dfrac{\log(U)}{\log(1-p)}\right]$ où $[x]$ désigne la partie entière de x
Poisson θ	$\inf\{n \in \mathbb{N} :\ \prod_{k=1}^{n+1} U_k \le \mathrm{e}^{-\theta}\}$
uniforme sur $[a,b]$	$a + (b-a)U$
exponentielle λ	$-\dfrac{1}{\lambda}\log(U)$
Cauchy a	$a\tan(\pi U)$
gaussienne $\mathcal{N}(m,\sigma^2)$	$m + \sigma\sqrt{-2\log(U_1)}\ \cos(2\pi U_2)$

Proposition A.2.15. *Soit p, q deux densités sur $\mathbb{R}^d$ telles que*

$$\exists k > 0,\ \forall x \in \mathbb{R}^d,\ p(x) \le kq(x)$$

et $((Y_i, U_i), i \ge 1)$ une suite de vecteurs aléatoires indépendants et identiquement distribués avec Y_1 de densité q et U_1 uniforme sur $[0,1]$ indépendantes. Alors la variable aléatoire $N = \inf\{i \ge 1 : kq(Y_i)U_i \le p(Y_i)\}$ suit la loi géométrique de paramètre $1/k$ (elle est donc finie avec probabilité 1). Elle est indépendante du couple $(Y_N, kq(Y_N)U_N)$ qui est uniformément réparti sur $\{(x,z) \in \mathbb{R}^d \times \mathbb{R} :\ 0 \le z \le p(x)\}$.
En particulier $X = Y_N$, possède la densité p.

Pour plus de détails sur les techniques de simulation, nous renvoyons aux livres de Bouleau [2], de Fishman [5] et d'Ycart [8].

A.3 Convergence et théorèmes limites

A.3.1 Convergence de variables aléatoires

Nous commençons par définir la convergence en probabilité et étendre la définition A.1.15 de la convergence presque sûre au cas vectoriel.

Définition A.3.1. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$.*

- *On dit que la suite $(X_n, n \in \mathbb{N}^*)$ converge presque sûrement vers X à valeurs dans $\mathbb{R}^d$ et on note $X_n \xrightarrow[n\to\infty]{\textit{p.s.}} X$ si*

$$\mathbb{P}\left(\lim_{n\to+\infty} X_n = X\right) = 1.$$

- *On dit que la suite $(X_n, n \in \mathbb{N}^*)$ converge en probabilité vers une variable aléatoire X à valeurs dans $\mathbb{R}^d$ si et seulement si pour tout $\varepsilon > 0$, on a*

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

où $|X_n - X|$ désigne la norme euclidienne de $X_n - X$.

Exemple A.3.2. Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires de Bernoulli avec X_n de paramètre p_n. Si $\lim_{n\to\infty} p_n = 0$, alors la suite converge en probabilité vers 0. En effet, pour tout $\varepsilon \in]0,1[$, on a $\mathbb{P}(|X_n| > \varepsilon) = p_n$. ♢

Proposition A.3.3. *La convergence presque sûre entraîne la convergence en probabilité.*

Définition A.3.4. *On dit qu'une suite de variables aléatoires $(X_n, n \in \mathbb{N}^*)$ à valeurs dans $\mathbb{R}^d$ converge en loi vers X à valeurs dans $\mathbb{R}^d$ si pour toute fonction $g : \mathbb{R}^d \to \mathbb{R}$ continue et bornée, on a*

$$\lim_{n\to\infty} \mathbb{E}\left[g(X_n)\right] = \mathbb{E}\left[g(X)\right].$$

On dit alors également que la suite des lois des variables X_n converge étroitement vers la loi de X.
Dans le cas où X suit la loi gaussienne en dimension d de paramètre (m, Σ), on note

$$X_n \xrightarrow[n\to\infty]{\textit{Loi}} \mathcal{N}(0, \Sigma).$$

Exemple A.3.5. Soit X_n de loi uniforme sur $\left\{0, \frac{1}{n}, \cdots, \frac{n-1}{n}\right\}$. Alors la suite $(X_n, n \in \mathbb{N}^*)$ converge en loi vers U de loi uniforme sur $[0,1]$. En effet, si g est continue bornée, on déduit de la convergence des sommes de Riemann que

$$\mathbb{E}[g(X_n)] = \frac{1}{n}\sum_{k=0}^{n-1} g\left(\frac{k}{n}\right) \underset{n\to\infty}{\longrightarrow} \int_{[0,1]} g(x)\, dx = \mathbb{E}[g(U)].$$

$\diamondsuit$

On peut étendre la convergence des espérances à des fonctions discontinues.

Proposition A.3.6. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ qui converge en loi vers X. Soit $h : \mathbb{R}^d \to \mathbb{R}$ une fonction mesurable bornée et C l'ensemble de ses points de continuité.*

$$\textit{Si } \mathbb{P}(X \in C) = 1, \textit{ alors on a } \quad \lim_{n\to\infty} \mathbb{E}\left[h(X_n)\right] = \mathbb{E}\left[h(X)\right].$$

On en déduit le corollaire suivant.

Corollaire A.3.7. *Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}^d$ qui converge en loi vers X. Soit $h : \mathbb{R}^d \to \mathbb{R}$ une fonction mesurable. On note C l'ensemble des points où elle est continue. Si $\mathbb{P}(X \in C) = 1$, alors la suite de variables aléatoires $(h(X_n), n \in \mathbb{N}^*)$ converge en loi vers $h(X)$.*

La convergence en loi est équivalente à la convergence ponctuelle des fonctions caractéristiques, résultat très utile en pratique.

Théorème A.3.8. *Une suite $(X_n, n \in \mathbb{N}^*)$ de variables aléatoires à valeurs dans $\mathbb{R}^d$ converge en loi vers X si et seulement si*

$$\forall u \in \mathbb{R}^d,\ \psi_{X_n}(u) \underset{n\to\infty}{\longrightarrow} \psi_X(u).$$

Pour des variables aléatoires réelles positives, on peut également caractériser la convergence en loi à l'aide de la transformée de Laplace.

Théorème A.3.9. *Une suite $(X_n, n \in \mathbb{N}^*)$ de variables aléatoires réelles positives converge en loi vers X si et seulement si*

$$\forall \alpha \geq 0, \quad \mathbb{E}\left[\mathrm{e}^{-\alpha X_n}\right] \underset{n\to\infty}{\longrightarrow} \mathbb{E}\left[\mathrm{e}^{-\alpha X}\right].$$

Proposition A.3.10. *La convergence en probabilité implique la convergence en loi.*

Ainsi la convergence presque sûre entraîne la convergence en probabilité qui elle-même implique la convergence en loi. Les réciproques sont fausses en général. Signalons la réciproque partielle suivante.

Proposition A.3.11. *Si la suite de variables aléatoires* $(X_n, n \in \mathbb{N}^*)$ *converge en loi vers une constante* a*, alors elle converge également en probabilité vers* a*.*

Le résultat suivant est très utile pour construire des intervalles de confiance.

Théorème A.3.12 (Théorème de Slutsky). *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires à valeurs dans* $\mathbb{R}^{d_1}$ *qui converge en loi vers* X *et* $(Y_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires à valeurs dans* $\mathbb{R}^{d_2}$ *qui converge en loi vers une* ***constante*** a*. Alors la suite* $((X_n, Y_n), n \in \mathbb{N}^*)$ *converge en loi vers* (X, a)*.*

A.3.2 Loi forte des grands nombres

Théorème A.3.13 (Loi forte des grands nombres). *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires réelles ou vectorielles indépendantes, de même loi et intégrables (*$\mathbb{E}[|X_1|] < \infty$*). Alors on a*

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}[X_1].$$

En outre, $\lim_{n\to\infty} \mathbb{E}\left[\left|\bar{X}_n - \mathbb{E}[X_1]\right|\right] = 0.$

On en déduit le corollaire suivant.

Corollaire A.3.14. *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires positives, indépendantes et de même loi. Alors la moyenne empirique* $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ *converge presque sûrement vers* $\mathbb{E}[X_1] \in [0, \infty]$ *quand* n *tend vers l'infini.*

A.3.3 Théorème central limite

Le théorème central limite (TCL) précise la vitesse de convergence de la loi forte des grands nombres.

Théorème A.3.15 (Théorème central limite). *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires réelles indépendantes de même loi et de carré intégrable (*$\mathbb{E}[X_n^2] < \infty$*). On pose* $\mu = \mathbb{E}[X_n]$*,* $\sigma^2 = \mathrm{Var}(X_n)$ *et* $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$*. La suite de variables aléatoires* $(\sqrt{n}[\bar{X}_n - \mu], n \in \mathbb{N}^*)$ *converge en loi vers une variable aléatoire de loi gaussienne* $\mathcal{N}(0, \sigma^2)$ *:*

$$\sqrt{n}[\bar{X}_n - \mu] = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{Loi}} \mathcal{N}(0, \sigma^2).$$

Ce résultat admet la version vectorielle suivante.

Proposition A.3.16. *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires à valeurs dans* $\mathbb{R}^d$ *indépendantes, de même loi et de carré intégrable (*$\mathbb{E}[|X_n|^2] < \infty$*). Soit* $\mu = \mathbb{E}[X_n]$ *et* $\Sigma = \mathrm{Cov}(X_n, X_n)$ *la matrice de covariance de* X_n*. La suite de variables aléatoires* $(\sqrt{n}(\bar{X}_n - \mu), n \in \mathbb{N}^*)$*, où* $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$*, converge en loi vers un vecteur gaussien de loi* $\mathcal{N}(0, \Sigma)$ *:*

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{\textit{Loi}} \mathcal{N}(0, \Sigma).$$

On peut également préciser le comportement asymptotique de $g(\bar{X}_n)$ lorsque la fonction g a de bonnes propriétés.

Proposition A.3.17. *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires indépendantes et de même loi, à valeurs dans* $\mathbb{R}^d$*. On suppose que* X_n *est de carré intégrable. On note* $\mu = \mathbb{E}[X_n]$ *sa moyenne et* $\Sigma = \mathrm{Cov}(X_n, X_n)$ *sa matrice de covariance. Soit* g *une fonction de* $\mathbb{R}^d$ *dans* $\mathbb{R}^p$ *mesurable. On suppose de plus que* g *est continue et différentiable en* μ*. Sa différentielle au point* μ *est la matrice* $\dfrac{\partial g}{\partial x'}(\mu)$ *de taille* $p \times d$ *définie par :*

$$\frac{\partial g}{\partial x'}(\mu)_{i,j} = \frac{\partial g_i}{\partial x_j}(\mu)\,; \quad 1 \le i \le p, \quad 1 \le j \le d.$$

On notera $\dfrac{\partial g'}{\partial x}(\mu)$ *sa transposée. On a alors*

$$\boxed{g(\bar{X}_n) \xrightarrow[n\to\infty]{\textit{p.s.}} g(\mu) \quad \text{et} \quad \sqrt{n}\left[g(\bar{X}_n) - g(\mu)\right] \xrightarrow[n\to\infty]{\textit{Loi}} \mathcal{N}(0, V),}$$

où $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^n X_k$ *désigne la moyenne empirique et* $V = \dfrac{\partial g}{\partial x'}(\mu)\ \Sigma\ \dfrac{\partial g'}{\partial x}(\mu)$*.*

A.3.4 Intervalles de confiance

Soit $(X_n, n \in \mathbb{N}^*)$ une suite de variables aléatoires réelles de carré intégrable indépendantes et identiquement distribuées. On note $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ la moyenne empirique et on pose $\mu = \mathbb{E}[X_n]$, $\sigma^2 = \mathrm{Var}(X_n)$.

Comme la loi gaussienne est une loi à densité, elle ne charge pas les points de discontinuité de la fonction indicatrice $\mathbf{1}_{[-a\sigma, a\sigma]}$. Après avoir remarqué que

$$\mathbb{E}\left[\mathbf{1}_{[-a\sigma,a\sigma]}(\sqrt{n}[\bar{X}_n - \mu])\right] = \mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right),$$

on déduit du théorème central limite A.3.15 et de la proposition A.3.6 le corollaire suivant.

Corollaire A.3.18. *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires réelles indépendantes, identiquement distribuées et de carré intégrable. On pose* $\mu = \mathbb{E}[X_n], \sigma^2 = \mathrm{Var}(X_n)$ *et* $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. *Alors, pour* $a > 0$, *on a*

$$\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} \int_{-a}^{a} \mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}}.$$

Si l'on désire donner une approximation de μ, à l'aide de la moyenne empirique $\bar{X}_n$, on peut fournir un intervalle aléatoire $I_n = \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]$ qui contient la valeur de la moyenne, μ, avec une probabilité asymptotique $\alpha = \mathbb{P}(|Z| \leq a)$, où Z suit la loi $\mathcal{N}(0, 1)$. L'intervalle I_n est appelé **intervalle de confiance** pour μ de **niveau asymptotique** α. Les valeurs les plus couramment utilisées sont $\alpha = 95\,\%$ avec $a \simeq 1.96$ et $\alpha = 99\,\%$ avec $a \simeq 2.58$.

En général, quand on désire estimer la moyenne μ, on ne connaît pas la variance σ^2. Il faut donc remplacer σ dans l'intervalle de confiance par une estimation. Comme $\sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, on déduit de la loi forte des grands nombres que la suite $(\sigma_n^2, n \in \mathbb{N}^*)$, définie par

$$\sigma_n^2 = \frac{1}{n}\sum_{k=1}^n X_k^2 - \bar{X}_n^2,$$

converge p.s. vers σ^2.

Le théorème de Slutsky A.3.12 assure que $((\sqrt{n}(\bar{X}_n - \mu), \sigma_n^2), n \in \mathbb{N}^*)$ converge en loi vers $(\sigma Z, \sigma^2)$, où Z est de loi gaussienne centrée réduite $\mathcal{N}(0, 1)$. Enfin, la fonction $f(x, y) = x/\sqrt{|y|}$ si $y \neq 0$ et $f(x, y) = 0$ sinon, admet $\mathbb{R} \times \mathbb{R}_+^*$ comme ensemble de points de continuité. Si $\sigma^2 > 0$, on a $\mathbb{P}((\sigma Z, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*) = 1$. On déduit donc du corollaire A.3.7, que la suite $(f(X_n, Y_n), n \in \mathbb{N}^*)$ converge en loi vers Z qui a pour loi $\mathcal{N}(0, 1)$. En particulier, une nouvelle utilisation de la proposition A.3.6, avec la fonction $h(r) = \mathbf{1}_{[-a,a]}(r)$, assure que si $\sigma \neq 0$,

$$\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right]\right) \xrightarrow[n\to\infty]{} \int_{-a}^{a} \mathrm{e}^{-x^2/2}\ \frac{dx}{\sqrt{2\pi}}.$$

Ainsi pour n grand, la moyenne μ appartient à l'intervalle de confiance aléatoire $\left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right]$ avec une probabilité proche du niveau asymptotique $\alpha = \mathbb{P}(|Z| \leq a)$.

On peut enfin se poser la question de la validité de l'intervalle de confiance : on a $\mathbb{P}\left(\mu \in \left[\bar{X}_n - \frac{a\sigma_n}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma_n}{\sqrt{n}}\right]\right) \simeq \mathbb{P}(|Z| \leq a)$, mais quelle est la précision de cette approximation ? Le résultat suivant apporte une réponse à cette question.

Théorème A.3.19 (Théorème de Berry-Esséen). *Soit* $(X_n, n \in \mathbb{N}^*)$ *une suite de variables aléatoires réelles indépendantes et identiquement* *distribuées. On suppose* $\mathbb{E}[|X_n|^3] < \infty$*. On note* $\mu = \mathbb{E}[X_n]$, $\sigma^2 = \mathrm{Var}(X_n)$ *et* $\mu_3 = \mathbb{E}[|X_n - \mu|^3]$*. Alors pour tout* $a \in \mathbb{R}$, $n \geq 1$, *on a*

$$\left| \mathbb{P}\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \leq a\right) - \int_{-\infty}^{a} \mathrm{e}^{-x^2/2} \, \frac{dx}{\sqrt{2\pi}} \right| \leq \frac{C\mu_3}{\sigma^3\sqrt{n}}, \tag{A.6}$$

où la constante C *est universelle (i.e. indépendante de* a *et de la loi de* X_1*) avec* $(2\pi)^{-1/2} \leq C < 0.8$.

On connaît de meilleures majorations si a est grand. En effet, on peut alors remplacer C par $C(a)$, où $\lim_{|a|\to\infty} C(a) = 0$. Enfin, la majoration (A.6) reste vraie si l'on remplace σ dans le membre de gauche par son estimation σ_n, avec une autre constante universelle C' remplaçant C.

Si le rapport μ_3/σ^3 (ou une approximation de μ_3/σ^3) est élevé, cela suggère que la convergence du théorème central limite peut être lente. Par exemple, pour des variables de Bernoulli de paramètre p, le rapport est équivalent à $1/\sqrt{p}$ lorsque p tend vers 0. Dans ce cas, il est plus judicieux d'utiliser l'approximation donnée par la loi des petits nombres (voir le paragraphe 6.3.1) et non pas la loi forte des grands nombres et le TCL.

Enfin remarquons que la majoration du théorème A.3.19 est indépendante de la loi des variables (la constante C est universelle). Elle est dans bien des cas très grossière. Elle donne cependant le bon ordre de grandeur pour des variables de Bernoulli.

Références

1. A.A. Borovkov. *Mathematical statistics.* Gordon and Breach Science Publishers, Amsterdam, 1998.
2. N. Bouleau. *Probabilités de l'ingénieur. Variables aléatoires et simulation*, volume 1418 d'*Actualités Scientifiques et Industrielles.* Hermann, Paris, 1986.
3. L. Breiman. *Probability*, volume 7 de *Classics in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
4. P. Brémaud. *Point processes and queues. Martingale dynamics.* Series in Statistics. Springer-Verlag, New York - Heidelberg - Berlin, 1981.
5. G.S. Fishman. *Monte Carlo. Concepts, algorithms, and applications.* Springer Series in Operations Research. Springer-Verlag, New York, 1996.
6. G.R. Grimmett et D.R. Stirzaker. *Probability and random processes.* Oxford University Press, New York, troisième édition, 2001.
7. G. Saporta. *Probabilités, Statistique et Analyse des Données.* Technip, 1990.
8. B. Ycart. *Modèles et algorithmes markoviens*, volume 39 de *Mathématiques & Applications.* Springer, Berlin, 2002.