

Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées
Bât. M2, F-59655 Villeneuve d'Ascq Cedex

Initiation à la Statistique IS-Math314

Chapitres 1–4

Charles SUQUET

Table des matières

1	Théorème limite central	7
1.1	Convergence en loi	7
1.2	Normalité asymptotique	12
1.2.1	Sommes de variables aléatoires i.i.d.	12
1.2.2	Vitesse de convergence dans le TLC	14
1.2.3	Intervalle de confiance pour une probabilité inconnue	16
1.2.4	Généralisation du TLC	18
1.3	Théorème limite central vectoriel	20
1.3.1	Espérance et covariance d'un vecteur aléatoire	20
1.3.2	Vecteurs aléatoires gaussiens	23
1.3.3	TLC vectoriel	27
1.4	Compléments sur la convergence en loi et le TLC	29
1.4.1	Outillage pour la convergence en loi	29
1.4.2	Démonstration du TLC	36
2	Simulation de variables et vecteurs aléatoires	43
2.1	Introduction	43
2.2	Méthode théorique pour simuler une v.a.r.	44
2.3	Méthodes particulières pour lois usuelles	48
2.3.1	Lois discrètes à support fini	48
2.3.2	Lois binomiales et multinomiales	50
2.3.3	Lois de Poisson	52
2.3.4	Lois géométriques	54
2.3.5	Lois gaussiennes	55
2.4	Algorithmes de rejet	57
2.4.1	Simulation de lois uniformes par rejet	57
2.4.2	Simulation de lois à densité par rejet	61
2.4.3	Simulation d'une loi discrète par rejet	64
2.5	Simulation de vecteurs aléatoires par transformation	67
2.5.1	Loi uniforme par transformation affine	67
2.5.2	Vecteur gaussien de covariance donnée	73
	Bibliographie	75

3	Échantillons et statistiques	77
3.1	Modélisation statistique	77
3.2	Mesure empirique	82
3.2.1	Une loi construite à partir des observations	82
3.2.2	Convergence de la f.d.r. empirique vers la f.d.r. théorique	85
3.2.3	Application au test de Kolmogorov-Smirnov	92
3.3	Moments empiriques	94
3.3.1	Moments observés et moments empiriques	94
3.3.2	Espérance et variance des moments empiriques	95
3.4	Lois des moments empiriques	97
3.4.1	Échantillon de grande taille	97
3.4.2	Échantillon gaussien	101
4	Estimation	107
4.1	Estimateurs	107
4.1.1	Exemples	107
4.1.2	Généralités	109
4.1.3	Erreur quadratique moyenne	110
4.2	Maximum de vraisemblance	118
4.2.1	Exercice introductif	118
4.2.2	Cas discret	119
4.2.3	Cas à densité	120
A	Tables statistiques	123
A.1	Loi normale standard	123
A.2	Lois du khi2	125
A.3	Lois de Student	127
A.4	Test de Kolmogorov Smirnov	129

Index

- acceptation-rejet, 57
- algorithme du rejet, 61
- biais, 110
- Box Muller, 56
- condition
 - de Liapounov, 18, 38
 - de Lindeberg, 19
- consistance
 - faible, 109
 - forte, 109
- convergence
 - étroite, 8
 - en loi, 7
 - par image continue, 10
 - en loi (pièges), 8
 - en loi dans \mathbb{R}^d , 8
 - formes linéaires, 27
- covariance, 22
- Cramér-Rao, 114
 - borne de, 115
- Cramér-Wold, 27
- dérivation sous le signe somme, 115
- échantillon, 81
- efficace (estimateur), 115
- EQM, 110
 - décomposition, 111
- équidistribution, 18
- équitension, 31
- erreur quadratique moyenne, 110
- espérance d'un vecteur aléatoire, 21
 - et applications linéaires, 21
 - linéarité, 21
- espacements, 71
- estimateur, 109
 - asymptotiquement sans biais, 110
 - biaisé, 110
 - efficace, 115
 - faiblement consistant, 109
 - fortement consistant, 109
 - sans biais, 110
- estimation
 - fonctionnelle, 107
 - ponctuelle, 107
- factorielle, 15
- fonction de répartition empirique, 84
- fonction quantile, 45
- formule
 - de Stirling, 15
- fourchette, 16
- générateur congruentiel, 43
- inégalité
 - de Cramér-Rao, 114
 - de Jensen, 39
- intervalle de confiance, 16
- inverse généralisée d'une f.d.r., 45
- lemme
 - de Slutsky, 98
- loi
 - de Gumbel, 10
 - de Poisson, 52
 - de Student, 104
 - de Weibull, 48
 - de Zipf, 66
 - du khi-deux (χ^2), 104
 - formes linéaires, 23
 - géométrique, 54
 - Gamma, 63
 - multinomiale, 28, 51

- normale
 - tables, 124
- matrice de covariance, 22
- mesure empirique, 82
 - espérance, 83
 - variance, 83
- modèle statistique, 77
- moment empirique, 95
- moyenne empirique, 83, 95
 - théorème limite central, 97
- quantile, 45
- record, 9
- simplexe, 70
- simulation
 - loi binomiale, 50
 - loi de Cauchy, 47
 - loi de Poisson, 52
 - loi de Zipf, 66
 - loi discrète, 48
 - loi géométrique, 55
 - loi Gamma, 63
 - loi multinomiale, 51
 - loi uniforme sur ellipsoïde, 73
 - loi uniforme sur un simplexe, 71
 - lois de Weibull, 48
 - lois exponentielles, 48
 - rejet
 - densité, 61
 - loi discrète, 64
 - v.a. gaussiennes, 56
 - vecteur gaussien, 73
- sondage, 16
- statistique, 81
- statistiques d'ordre, 71
- tableau triangulaire, 19
- tables
 - loi de Student, 128
 - loi du khi-deux, 126
 - loi normale, 124
 - test de Kolmogorov Smirnov, 129
- test
 - Kolmogorov Smirnov, 92
- théorème
 - dérivation sous le signe somme, 115
 - de Berry-Esséen, 14
 - de de Moivre-Laplace, 12
 - de Glivenko-Cantelli, 86
 - de Katz Petrov, 15
 - de Kolmogorov Smirnov, 93
 - de Liapounov, 18, 38
 - de Lindeberg, 19
 - de Student, 101
 - limite central, 12
- théorème limite central
 - cas i.i.d., 12
 - dans \mathbb{R}^d , 27
 - de de Moivre-Laplace, 12
 - de Liapounov, 18, 38
 - de Lindeberg, 19
 - et lois multinomiales, 28
 - pour la moyenne empirique, 97
 - vectorel, 27
 - vitesse de convergence, 14, 15
- variance empirique, 83, 95
 - formule de Koenig, 95
- vecteur aléatoire
 - espérance, 21
 - gaussien, 23
 - paramètres, 24
 - intégrable, 21
- vitesse de convergence
 - loi forte des grands nombres, 14
 - théorème limite central, 14, 15, 38
- vraisemblance, 119

Chapitre 1

Théorème limite central

Le théorème limite central nous dit qu'une somme d'un grand nombre de variables aléatoires indépendantes, de carré intégrable, convenablement normalisée, se comporte asymptotiquement « en loi » comme une v.a. gaussienne. Il explique l'importance centrale des lois gaussiennes dans la théorie des probabilités et la statistique. Il complète la loi des grands nombres en donnant une sorte de vitesse de convergence, permettant notamment de construire des « intervalles de confiance » pour l'estimation d'un paramètre.

Pour donner un sens mathématique précis à cette notion de « comportement asymptotique en loi », il nous faut d'abord introduire la convergence en loi.

1.1 Convergence en loi

Nous admettrons l'équivalence des deux définitions suivantes de la convergence en loi.

Définition 1.1 (convergence en loi). *Notons F_n et F les fonctions de répartition respectives des variables aléatoires réelles Y_n ($n \geq 1$) et Y . On dit que la suite $(Y_n)_{n \geq 1}$ converge en loi vers Y si*

$$\forall x \text{ point de continuité de } F, \quad F_n(x) \xrightarrow[n \rightarrow +\infty]{} F(x). \quad (1.1)$$

Rappelons que x est point de continuité de la f.d.r. F si et seulement si $F(x-) = F(x)$ ou encore $P(Y = x) = 0$.

Définition 1.2 (convergence en loi). *On dit que la suite $(Y_n)_{n \geq 1}$ de variables aléatoires réelles converge en loi vers la variable aléatoire réelle Y si*

$$\forall h \text{ continue bornée } \mathbb{R} \rightarrow \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}h(Y). \quad (1.2)$$

Remarquons que si h est continue bornée, les $h(Y_n)$ et $h(Y)$ sont des v.a. bornées, donc intégrables. Nous noterons la convergence en loi de Y_n vers Y par

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y.$$

La définition 1.1 est la plus concrète, surtout lorsque F est continue sur tout \mathbb{R} , cas souvent rencontré en pratique. En effet dans ce cas, la convergence en loi équivaut à la convergence simple sur \mathbb{R} des fonctions de répartition et nous donne, pour tous réels $a < b$, la convergence des $P(Y_n \in I(a, b))$ vers les $P(Y \in I(a, b))$, où $I(a, b)$ désigne n'importe lequel des 4 intervalles d'extrémités a et b .

La définition 1.2 est souvent plus commode pour établir les propriétés de la convergence en loi et a l'intérêt d'une généralisation immédiate aux vecteurs aléatoires de \mathbb{R}^d .

Définition 1.3 (convergence en loi de vecteurs aléatoires). *On dit que la suite $(Y_n)_{n \geq 1}$ de vecteurs aléatoires de \mathbb{R}^d converge en loi vers le vecteur aléatoire Y de \mathbb{R}^d si*

$$\forall h \text{ continue bornée } \mathbb{R}^d \rightarrow \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow{n \rightarrow +\infty} \mathbf{E}h(Y). \quad (1.3)$$

Remarques 1.4 (les pièges de la convergence en loi). Pointons d'emblée des différences importantes entre la convergence en loi et les autres modes de convergence vus jusqu'ici.

1. Il n'est pas nécessaire, pour la convergence en loi de Y_n vers Y , que ces variables aléatoires soient définies *sur le même* (Ω, \mathcal{F}, P) .
2. Il n'y a pas unicité de la v.a. limite en loi. Si $(Y_n)_{n \geq 1}$ converge en loi vers Y , elle converge aussi en loi vers *n'importe quelle variable aléatoire Z ayant même loi* que Y (éventuellement définie sur un autre espace probabilisé). Ceci se voit facilement sur chacune des deux définitions de la convergence en loi¹. Réciproquement si Y_n converge en loi vers Y et aussi vers Z , alors Y et Z ont même loi. En effet en utilisant la définition 1.2 et l'unicité de la limite d'une suite convergente de réels, on voit que $\mathbf{E}h(Y) = \mathbf{E}h(Z)$ pour toute $h : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée. Par la caractérisation des lois par leurs h -moments, cf. cours d'I.P.É., on en déduit que Y et Z ont même loi. En résumé, s'il n'y a pas unicité de la v.a. limite en loi, il y a unicité de sa loi, que l'on appellera *loi limite*².
3. La convergence en loi n'est pas compatible avec l'addition. Si X_n converge en loi vers X et si Y_n converge en loi vers Y , il est faux en général que $X_n + Y_n$ converge en loi vers $X + Y$. En effet si c'était le cas, comme X_n converge en loi vers *n'importe quel* X' ayant même loi que X , $X_n + Y_n$ devrait converger aussi en loi vers $X' + Y$. Le *hic* c'est que $X + Y$ n'a pas forcément même loi que $X' + Y$.

Après ces mises en garde, voyons un exemple assez typique où la convergence en loi est le concept pertinent pour décrire le comportement asymptotique d'une suite de variables aléatoires.

1. Cette non-unicité de la limite est bien plus générale que pour les autres modes de convergence vus jusqu'ici où l'on avait convergence vers *n'importe quelle* Z égale p.s. à Y . Bien sûr, si Y et Z sont définies sur le même espace et sont égales p.s., elles ont même loi, mais la réciproque est grossièrement fausse. Quand on lance deux dés, on n'est pas sûr d'obtenir un double !

2. Ceci incite à voir la convergence en loi de Y_n vers Y comme la *convergence de la loi* P_{Y_n} vers la loi P_Y . On pourrait d'ailleurs, en sortant nettement du programme de ce cours, donner un sens mathématique précis à cette convergence, appelée *convergence étroite des mesures de probabilité* en notant que $\mathbf{E}h(Y_n)$ ne dépend que de h et de P_{Y_n} .

Exemple 1.5 (une loi limite de records). Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires indépendantes et de même loi avec fonction de répartition commune F . Définissons la suite de variables aléatoires « records » $(M_n)_{n \geq 1}$ par :

$$M_n := \max_{1 \leq k \leq n} X_k, \quad n \in \mathbb{N}^*. \quad (1.4)$$

Connaissant F , il est facile d'obtenir la fonction de répartition G_n de M_n :

$$G_n(x) = P(M_n \leq x) = P(\forall k \in \{1, \dots, n\}, X_k \leq x) = P\left(\bigcap_{k=1}^n \{X_k \leq x\}\right).$$

En utilisant l'indépendance des X_k , puis le fait qu'elles ont même loi, on en déduit :

$$G_n(x) = \prod_{k=1}^n P(X_k \leq x) = (F(x))^n. \quad (1.5)$$

Supposons désormais que les X_k ont pour loi commune la loi exponentielle de paramètre a , alors

$$F(x) = 1 - e^{-ax} \text{ si } x \geq 0, \quad F(x) = 0 \text{ si } x < 0;$$

$$G_n(x) = (1 - e^{-ax})^n \text{ si } x \geq 0, \quad G_n(x) = 0 \text{ si } x < 0.$$

Donc pour x réel fixé, on a $\lim_{n \rightarrow +\infty} G_n(x) = 0$. La signification intuitive de ce résultat est que le record M_n finira par dépasser n'importe quel niveau x fixé pour n assez grand³. Afin de préciser cette idée, on cherche une suite non aléatoire tendant vers $+\infty$ à la même vitesse que M_n . On peut vérifier que $\mathbf{E}M_n = \frac{1}{a}(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n})$, donc $\mathbf{E}M_n \sim a^{-1} \ln n$, cf. par exemple le corrigé de l'examen d'I.P.É. de janvier 2006. Ceci nous amène à étudier le comportement asymptotique de $P(M_n - a^{-1} \ln n \leq x)$:

$$P\left(M_n - \frac{\ln n}{a} \leq x\right) = G_n\left(x + \frac{\ln n}{a}\right) = (1 - e^{-ax - \ln n})^n = \left(1 - \frac{e^{-ax}}{n}\right)^n. \quad (1.6)$$

On en déduit que :

$$\lim_{n \rightarrow +\infty} P\left(M_n - \frac{\ln n}{a} \leq x\right) = \exp(-e^{-ax}). \quad (1.7)$$

Le calcul (1.6) est valable pour $\ln n \geq -ax$, donc pour tout $n \in \mathbb{N}^*$ et tout $x \geq 0$. Pour $x < 0$ fixé, on aura $\ln n \geq -ax$ pour $n \geq n_0(x)$ donc (1.7) est valable pour tout x réel. On peut donc dire qu'*asymptotiquement*, M_n est de l'ordre de grandeur de $a^{-1} \ln n$ et que la dispersion aléatoire de M_n autour de cette valeur est donnée par la loi de fonction de répartition :

$$H(x) = \exp(-e^{-ax}), \quad x \in \mathbb{R}. \quad (1.8)$$

3. N'appliquez pas cette remarque au sport, même avec dopage. Cette « convergence en probabilité vers l'infini » de M_n n'est possible que parce que chaque X_k peut elle même prendre une valeur supérieure à x avec une probabilité non nulle. Si on prend pour X_k des variables de loi uniforme sur $[0, 1]$, la suite des records restera bornée par 1.

On vérifie immédiatement que H est continue sur \mathbb{R} , croissante (comme composée de deux fonctions décroissantes) avec pour limites 0 en $-\infty$ et 1 en $+\infty$. C'est donc bien une fonction de répartition. La loi de f.d.r. H est une loi de Gumbel.

D'après la définition 1.1, on peut reformuler la conclusion en disant que la suite de variables aléatoires $M_n - a^{-1} \ln n$ converge en loi vers une v.a. suivant la loi de Gumbel de f.d.r. H donnée par (1.8).

Une propriété bien commode de la convergence en loi est sa conservation par image continue.

Proposition 1.6 (convergence en loi par image continue). *Si Y_n converge en loi vers Y , alors pour toute f continue $\mathbb{R} \rightarrow \mathbb{R}$, $f(Y_n)$ converge en loi vers $f(Y)$.*

Noter que l'on ne suppose pas f bornée sur \mathbb{R} .

Preuve. D'après la définition 1.2, il nous faut vérifier que pour toute fonction continue bornée $g : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbf{E}g(f(Y_n))$ tend vers $\mathbf{E}g(f(Y))$ quand n tend vers $+\infty$. Or la fonction $g \circ f$ est continue sur \mathbb{R} par composition et bornée sur \mathbb{R} par $\sup_{t \in \mathbb{R}} |g(t)|$. On sait par hypothèse que $\mathbf{E}h(Y_n)$ converge vers $\mathbf{E}h(Y)$ pour toute h continue bornée sur \mathbb{R} . En appliquant ceci avec $h = g \circ f$, on obtient la conclusion souhaitée. \square

La preuve ci-dessus se généralise immédiatement aux vecteurs aléatoires.

Proposition 1.7 (convergence en loi de vecteurs par image continue). *Si les Y_n et Y sont des vecteurs aléatoires de \mathbb{R}^d tels que Y_n converge en loi vers Y , alors pour toute f continue $\mathbb{R}^d \rightarrow \mathbb{R}^j$, $f(Y_n)$ converge en loi vers $f(Y)$ dans \mathbb{R}^j .*

Le diagramme des convergences de la figure 1.1 indique que la convergence en loi est la plus faible des convergences de suites de variables aléatoires. Cette affirmation se justifie par le résultat suivant.

Proposition 1.8. *La convergence en probabilité implique la convergence en loi : si les Y_n ($n \geq 1$) et Y sont des variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{F}, P) telles que Y_n converge en probabilité vers Y , alors Y_n converge aussi en loi vers Y .*

Nous allons prouver la proposition en utilisant la définition 1.2, cette méthode ayant l'avantage de se généraliser immédiatement au cas des vecteurs aléatoires⁴ de \mathbb{R}^d . Nous aurons besoin du lemme élémentaire d'analyse suivant.

Lemme 1.9 (convergence par sous-suites). *La suite de réels $(u_n)_{n \geq 1}$ converge vers le réel ℓ si de toute sous-suite de $(u_n)_{n \geq 1}$ on peut extraire une nouvelle sous-suite convergeant vers ℓ .*

4. La convergence en probabilité de Y_n vers Y dans \mathbb{R}^d se définit comme en dimension 1, mais en remplaçant $|Y_n - Y|$ par $\|Y_n - Y\|$ après le choix d'une norme dans \mathbb{R}^d . Peu importe laquelle, puisqu'en dimension finie elles sont toutes équivalentes.

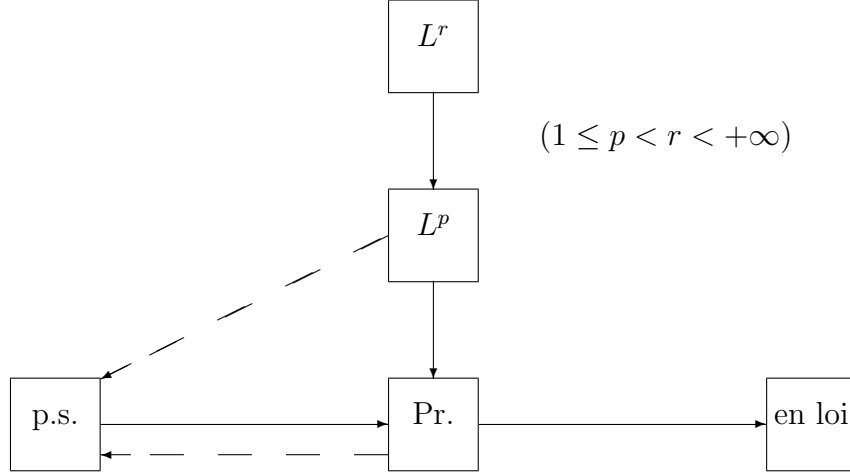


FIG. 1.1 – Diagramme des convergences des suites de v.a.

Preuve. Par confort typographique, nous noterons une sous-suite de $(u_n)_{n \in \mathbb{N}^*}$ comme une suite $(u_n)_{n \in A}$, où A est une partie infinie de \mathbb{N}^* . Une sous-suite de $(u_n)_{n \in A}$ s'écrira alors $(u_n)_{n \in B}$, pour une partie infinie B de A . La convergence de cette sous-suite sera notée :

$$u_n \xrightarrow[n \in B, n \rightarrow +\infty]{} \ell.$$

L'hypothèse du lemme s'écrit donc

$$\forall A \text{ infini } \subset \mathbb{N}^*, \quad \exists B \text{ infini } \subset A, \quad u_n \xrightarrow[n \in B, n \rightarrow +\infty]{} \ell. \quad (1.9)$$

Supposons que $(u_n)_{n \in \mathbb{N}^*}$ ne converge pas vers ℓ . Il existe alors un $\varepsilon > 0$ tel que

$$\forall j \in \mathbb{N}^*, \quad \exists n \geq j, \quad |u_n - \ell| \geq \varepsilon. \quad (1.10)$$

Autrement dit, il existe une infinité d'entiers n tels que $|u_n - \ell| \geq \varepsilon$. Notons A l'ensemble de ces entiers. Par l'hypothèse (1.9), il existe une partie infinie B de cet ensemble A telle que la sous-suite $(u_n)_{n \in B}$ converge vers ℓ . On peut alors trouver $n \in B$ assez grand pour que $|u_n - \ell| < \varepsilon$. Mais comme cet n est aussi dans A , on aboutit à une contradiction. On peut donc conclure à la convergence de $(u_n)_{n \in \mathbb{N}^*}$ vers ℓ . \square

Preuve de la proposition 1.8. Par hypothèse, Y_n converge en probabilité vers Y . Soit $h : \mathbb{R} \rightarrow \mathbb{R}$ continue bornée quelconque, il s'agit de prouver que $\mathbf{E}h(Y_n)$ converge vers $\mathbf{E}h(Y)$. Nous allons utiliser pour cela le lemme 1.9 avec $u_n = \mathbf{E}h(Y_n)$ et $\ell = \mathbf{E}h(Y)$. Comme h est bornée, il existe un réel $b > 0$ tel que $h(x) \in [-b, b]$ pour tout $x \in \mathbb{R}$. On en déduit l'inégalité entre variables aléatoires $-b \leq h(Y_n) \leq b$. Soit A une partie infinie quelconque de \mathbb{N}^* . Puisque $(Y_n)_{n \in A}$ converge en probabilité vers Y , on peut en extraire une sous-suite $(Y_n)_{n \in B}$ qui converge p.s. vers Y . Par continuité de h , on en déduit que

$$h(Y_n) \xrightarrow[n \in B, n \rightarrow +\infty]{\text{p.s.}} h(Y).$$

Comme $|h(Y_n)| \leq b$, on en déduit par le théorème de convergence dominée, la v.a. constante b étant évidemment intégrable, que

$$\mathbf{E}h(Y_n) \xrightarrow[n \in B, n \rightarrow +\infty]{} \mathbf{E}h(Y).$$

Comme A était quelconque, on conclut par le lemme 1.9 que c'est toute la suite $(\mathbf{E}h(Y_n))_{n \in \mathbb{N}^*}$ qui converge vers $\mathbf{E}h(Y)$. Ceci étant vrai pour toute fonction continue bornée h , la proposition 1.8 est démontrée. \square

1.2 Normalité asymptotique

1.2.1 Sommes de variables aléatoires i.i.d.

Théorème 1.10 (théorème limite central, cas i.i.d.). *Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{F}, P) , indépendantes, de même loi et de carré intégrable (et non p.s. constantes). Notons $\mu := \mathbf{E}X_1$, $\sigma^2 := \text{Var } X_1$ avec $\sigma > 0$ et $S_n = \sum_{k=1}^n X_k$. Alors*

$$S_n^* := \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var } S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z, \quad (1.11)$$

où Z est une variable de loi gaussienne $\mathfrak{N}(0, 1)$.

Il est possible d'énoncer ce théorème de manière plus élémentaire, sans parler de convergence en loi, ni même de loi gaussienne. En exploitant la *continuité* sur \mathbb{R} de la f.d.r. Φ de la loi $\mathfrak{N}(0, 1)$ et la définition 1.1 de la convergence en loi, on voit en effet qu'une formulation équivalente de la conclusion (1.11) du théorème est :

$$\forall x \in \mathbb{R}, \quad P(S_n^* \leq x) \xrightarrow[n \rightarrow +\infty]{} \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt. \quad (1.12)$$

Une conséquence pratique de (1.12) est

$$\forall a < b \in \mathbb{R}, \quad P(S_n^* \in I(a, b)) \xrightarrow[n \rightarrow +\infty]{} \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt, \quad (1.13)$$

où $I(a, b)$ est n'importe lequel des quatre intervalles d'extrémités a et b . Noter que sous cette forme on pourrait énoncer une version du théorème limite central compréhensible par un public ne connaissant que la notion d'intégrale de Riemann ordinaire⁵.

Corollaire 1.11 (théorème de de Moivre-Laplace). *Si S_n est une variable aléatoire de loi binomiale de paramètres n et $p \in]0, 1[$, on a avec $q := 1 - p$,*

$$S_n^* := \frac{S_n - np}{\sqrt{npq}} = \sqrt{\frac{n}{pq}} \left(\frac{S_n}{n} - p \right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z,$$

où Z est une variable de loi gaussienne $\mathfrak{N}(0, 1)$.

5. Typiquement un élève de terminale. On peut même laisser tomber la forme intégrale de la limite dans (1.13) en se contentant de dire que Φ est une fonction croissante continue que l'on a tabulée.

Preuve. C'est une conséquence immédiate du théorème 1.10 en remarquant que S_n a même loi⁶ que $X_1 + \dots + X_n$, où les X_k sont des variables aléatoires de Bernoulli indépendantes et de même paramètre p et en rappelant que l'espérance et la variance de la loi $\text{Bin}(n, p)$ sont respectivement np et npq . \square

La démonstration historique du théorème de de Moivre-Laplace repose sur un bon contrôle des coefficients binomiaux *via* la formule de Stirling. L'intérêt de cette approche « élémentaire » est de donner une idée de la vitesse de convergence qui est en $O(n^{-1/2})$, voir ICP⁷ chapitre 7.

Nous admettrons le théorème 1.10 dont la démonstration n'est pas au programme (les lecteurs curieux ou « avancés » pourront consulter la section 1.4). Ce théorème a de multiples applications, notamment en statistique. À ce stade, on peut souligner deux idées.

D'abord, on peut noter que le comportement asymptotique en loi de S_n^* ne dépend pas de la loi de X_1 . La seule condition pour que la loi de S_n^* soit approximativement gaussienne pour les grandes valeurs de n est que X_1 soit de carré intégrable. Ceci donne un caractère universel aux lois gaussiennes et explique la fréquence de l'utilisation de ces lois en modélisation⁸. On peut dire que le comportement asymptotique en loi de sommes S_n^* et donc aussi de S_n « oublie » tout de la loi des X_i , sauf le paramètre de localisation $\mu = \mathbf{E}X_1$ et le paramètre de dispersion $\sigma^2 = \text{Var } X_1$. C'est l'une des raisons de l'importance donnée à ces deux paramètres en théorie des probabilités.

La deuxième idée importante est que le théorème limite central donne une idée de la vitesse de convergence dans la loi des grands nombres. *Grosso modo*, on peut dire que dans le bon cas où $\mathbf{E}X_1^2 < +\infty$, cette vitesse est en $O(n^{-1/2})$. Précisons le sens de cette affirmation. Par (1.13) appliqué avec $a = -t$, $b = t$, $t > 0$, on obtient :

$$\lim_{n \rightarrow +\infty} P(S_n^* \in [-t, t]) = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1, \quad (1.14)$$

en utilisant la relation $\Phi(-t) = 1 - \Phi(t)$ due à la parité de la densité de $\mathfrak{N}(0, 1)$. En remarquant maintenant que

$$S_n^* = \frac{S_n - n\mathbf{E}X_1}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma} \left(\frac{S_n}{n} - \mathbf{E}X_1 \right), \quad (1.15)$$

on peut réécrire (1.14) sous la forme

$$P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \leq \frac{\sigma t}{\sqrt{n}}\right) = 2\Phi(t) - 1 + \varepsilon_n, \quad (1.16)$$

6. Il est clair d'après la définition de la convergence en loi que si Y_n converge en loi vers Y et si pour chaque n , Y'_n a même loi que Y_n , Y'_n converge aussi en loi vers Y .

7. Nous désignons par « ICP » le document *Introduction au Calcul des Probabilités*, Ch. Suquet, photocopié de deuxième année, Université des Sciences et Technologies de Lille.

http://math.univ-lille1.fr/~ipeis/cours_ipeis.html

8. D'autant plus qu'il existe de nombreuses généralisations du théorème limite central, avec des v.a. indépendantes mais de lois différentes, avec des vecteurs aléatoires, avec des v.a. « faiblement dépendantes » ...

où ε_n est une suite de réels (pas forcément positifs), convergente vers 0. Pour tout $\delta > 0$, on peut choisir un $t = t(\delta)$ assez grand pour que $2\Phi(t) - 1 > 1 - \delta/2$ car $2\Phi(t) - 1$ tend vers 1 quand t tend vers $+\infty$. Ensuite pour $n \geq n_0(\delta)$, on aura $|\varepsilon_n| < \delta/2$ et finalement

$$\forall \delta > 0, \exists t(\delta), n(\delta), \forall n \geq n(\delta), \quad P\left(\left|\frac{S_n}{n} - \mathbf{E}X_1\right| \leq \frac{\sigma t(\delta)}{\sqrt{n}}\right) > 1 - \delta. \quad (1.17)$$

C'est au sens de (1.17) que l'on peut dire que S_n/n converge vers $\mathbf{E}X_1$ avec une vitesse en $O(n^{-1/2})$. On peut résumer (1.17) par l'écriture $|n^{-1}S_n - \mathbf{E}X_1| = O_P(n^{-1/2})$, dont le deuxième membre se lit « grand O en probabilité de $n^{-1/2}$ ».

Dans l'utilisation pratique du théorème 1.10, on travaille souvent avec n « grand » fixé et on approxime la loi de S_n^* par la loi $\mathfrak{N}(0, 1)$, ou ce qui revient au même, *on approxime la loi de S_n par la loi gaussienne $\mathfrak{N}(n\mathbf{E}X_1, \sigma\sqrt{n})$ de même espérance et même variance que S_n* . Plus précisément, en notant que $g_n : x \mapsto \frac{x - n\mathbf{E}X_1}{\sigma\sqrt{n}}$ est une bijection croissante de \mathbb{R} sur \mathbb{R} et en posant pour $a < b$ réels,

$$a_n = g_n(a) = \frac{a - n\mathbf{E}X_1}{\sigma\sqrt{n}}, \quad b_n = g_n(b) = \frac{b - n\mathbf{E}X_1}{\sigma\sqrt{n}},$$

on a

$$P(a \leq S_n \leq b) = P(a_n \leq S_n^* \leq b_n) = \Phi(b_n) - \Phi(a_n) + \varepsilon_n. \quad (1.18)$$

On néglige alors le terme d'erreur ε_n et on termine le calcul en utilisant la table des valeurs de Φ .

1.2.2 Vitesse de convergence dans le TLC

La question qui se pose dans le calcul précédent est « que signifie n grand ? », ou encore « comment peut-on contrôler l'erreur ε_n ? », autrement dit, quelle est la vitesse de convergence vers 0 de ε_n ? La réponse est que dans le « bon cas » où X_1 a un moment d'ordre 3, la vitesse de convergence dans le théorème limite central est en $O(n^{-1/2})$.

Théorème 1.12 (Berry-Esséen, 1941–42). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbf{E}|X_i|^3 < +\infty$. On note $\sigma^2 := \text{Var } X_1$, $\rho^3 := \mathbf{E}|X_1 - \mathbf{E}X_1|^3$, avec $\sigma > 0$ et $\rho > 0$. Il existe alors une constante universelle $C > 0$ telle que pour tout $n \geq 1$,*

$$\Delta_n := \sup_{x \in \mathbb{R}} |P(S_n^* \leq x) - \Phi(x)| \leq C \frac{\rho^3}{\sigma^3} \frac{1}{\sqrt{n}}.$$

L'obtention de la meilleure constante C a été l'objet d'une longue quête. La valeur initiale de Esséen était $C = 7,59$. Une valeur plus moderne et proche de l'optimale est $C = 0,7975$ (Van Beek (1972)).

Il est intéressant de regarder ce que donne le théorème de Berry-Esséen pour le cas de de Moivre-Laplace, donc avec des X_i suivant la loi de Bernoulli de paramètre p . On trouve alors (vérifiez !)

$$\Delta_n \leq C \frac{p^2 + q^2}{\sqrt{pq}} \frac{1}{\sqrt{n}}, \quad q := 1 - p.$$

On a en fait des résultats plus précis de Uspensky concernant ce cas particulier, cf. [ICP, Chap. 7].

Voici un exemple tout à fait élémentaire permettant de comprendre qu'il n'y a pas lieu d'espérer une vitesse de convergence meilleure que $O(n^{-1/2})$ pour Δ_n . Prenons X_1 de loi de Bernoulli de paramètre $1/2$. On a alors

$$S_{2n} \sim \text{Bin}(2n, \frac{1}{2}), \quad \mathbf{E}S_{2n} = 2n\frac{1}{2} = n.$$

On cherche un équivalent de $P(S_{2n}^* \leq 0) - \Phi(0)$. Remarquons d'abord que

$$\{S_{2n}^* < 0\} = \{0 \leq S_{2n} < n\} \quad \text{et} \quad \{S_{2n}^* > 0\} = \{n < S_{2n} \leq 2n\}.$$

En raison de la symétrie des coefficients binomiaux ($C_{2n}^k = C_{2n}^{2n-k}$),

$$P(S_{2n}^* < 0) = \sum_{k=0}^{n-1} C_{2n}^k 2^{-2n} = \sum_{j=n+1}^{2n} C_{2n}^j 2^{-2n} = P(S_{2n}^* > 0).$$

On a ainsi $2P(S_{2n}^* < 0) + P(S_{2n}^* = 0) = 1$ d'où l'on tire $P(S_{2n}^* < 0) = \frac{1}{2} - \frac{1}{2}P(S_{2n}^* = 0)$ et $P(S_{2n}^* \leq 0) = \frac{1}{2} + \frac{1}{2}P(S_{2n}^* = 0)$. En rappelant que $\Phi(0) = \frac{1}{2}$, on aboutit à

$$P(S_{2n}^* \leq 0) - \Phi(0) = \frac{1}{2}P(S_{2n}^* = 0) = \frac{1}{2}P(S_{2n} = n) = C_{2n}^n 2^{-2n-1}.$$

Par la formule de Stirling ($n! \sim \sqrt{2\pi n} n^{n+1/2} e^{-n}$), on obtient l'équivalent

$$P(S_{2n}^* \leq 0) - \Phi(0) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2n}}.$$

Comme $(2\pi)^{-1/2} > 0,398\,9$, on a pour $n \geq n_0$, $|P(S_{2n}^* \leq 0) - \Phi(0)| \geq 0,398(2n)^{-1/2}$, minorant à comparer avec le majorant uniforme $\Delta_{2n} \leq 0,798(2n)^{-1/2}$ fourni dans ce cas par le théorème de Berry-Esséen⁹.

Revenons à la situation générale du théorème 1.12. Que se passe-t-il dans la zone intermédiaire où le théorème limite central est vérifié parce que X_1 a un moment d'ordre 2, mais X_1 n'a pas de moment d'ordre 3? On a toujours une vitesse de convergence, pourvu que l'intégrabilité de X_1 soit un peu plus forte que la seule existence d'un moment d'ordre 2. Le résultat précis est le suivant.

Théorème 1.13 (Katz 1963, Petrov 1965). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. telle que $\mathbf{E}X_1 = 0$, $\mathbf{E}(X_1^2 g(X_1)) < +\infty$, où g est une fonction positive, paire, croissante sur $]0, +\infty[$ et telle que $x/g(x)$ soit croissante sur $]0, +\infty[$. Il existe alors une constante universelle $A > 0$ telle que pour tout $n \geq 1$,*

$$\Delta_n := \sup_{x \in \mathbb{R}} |P(S_n^* \leq x) - \Phi(x)| \leq A \frac{\mathbf{E}(X_1^2 g(X_1))}{\sigma^2 g(\sigma\sqrt{n})}.$$

9. Signalons cependant que dans le cas de la loi de Bernoulli de paramètre $p = 1/2$, il est possible d'obtenir une vitesse de convergence en $O(n^{-1})$ pour une variante de la distance Δ_n obtenue en appliquant la *correction de continuité*, voir [ICP, chap. 7].

En particulier si $\mathbf{E}|X_1|^{2+\delta} < +\infty$ pour un $\delta \in]0, 1]$,

$$\Delta_n \leq A \frac{\mathbf{E}|X_1|^{2+\delta}}{\sigma^{2+\delta}} \frac{1}{n^{\delta/2}}.$$

1.2.3 Intervalle de confiance pour une probabilité inconnue

Voyons maintenant une première application statistique du théorème limite central à l'estimation d'une probabilité inconnue p à partir de l'observation d'un *échantillon* de n variables de Bernoulli indépendantes de paramètre p . Par exemple on se trouve en présence d'une urne dont on sait seulement qu'elle contient des boules rouges et des vertes, la proportion de chaque couleur étant inconnue. On cherche à estimer la proportion p de boules rouges en effectuant n tirages avec remise et en notant à chaque tirage si la couleur de la boule est rouge ou non. On pose alors $X_i = 1$ si le i^e tirage donne une boule rouge et $X_i = 0$ sinon. $S_n = X_1 + \dots + X_n$ est le nombre aléatoire de boules rouges sorties en n tirages et S_n/n la fréquence observée de sortie d'une boule rouge. Cette situation est comparable à un sondage avant le deuxième tour d'une élection opposant deux candidats A et B . On interroge n personnes et on pose $X_i = 1$ si la i^e personne interrogée annonce son intention de voter pour A . Ici S_n/n est la fréquence des intentions de votes pour A dans l'échantillon des sondés. La proportion p d'électeurs ayant l'intention de voter A dans la population totale est inconnue avant l'élection et c'est elle que l'on cherche à estimer¹⁰.

Nous allons voir comment le théorème de de Moivre-Laplace permet de construire des *intervalles de confiance* pour p (ce que les medias appellent la « fourchette » des instituts de sondage). Considérons pour $t > 0$ l'événement

$$A_{n,t} := \left\{ \omega \in \Omega; -t \leq \sqrt{\frac{n}{pq}} \left(\frac{S_n(\omega)}{n} - p \right) \leq t \right\}.$$

Le théorème de de Moivre-Laplace nous dit que pour n assez grand, on peut utiliser l'approximation :

$$P(A_{n,t}) \simeq \Phi(t) - \Phi(-t) = 2\Phi(t) - 1.$$

Ceci peut se réécrire

$$P\left(\frac{S_n}{n} - t\sqrt{\frac{pq}{n}} \leq p \leq \frac{S_n}{n} + t\sqrt{\frac{pq}{n}} \right) = 2\Phi(t) - 1 + \varepsilon_n.$$

On ignore la valeur de p , donc a fortiori celle de \sqrt{pq} . Heureusement, il est possible de la majorer car $p(1-p)$ est maximal pour $p = 1/2$. D'où

$$\sqrt{pq} \leq \sqrt{\frac{1}{4}} = \frac{1}{2}, \quad (1.19)$$

10. En toute rigueur, les 2 modèles ne sont pas équivalents car dans l'échantillon des sondés, on n'interroge jamais deux fois la même personne tandis que lors des tirages avec remise, la même boule peut sortir plus d'une fois. Néanmoins en raison du théorème de convergence de la loi hypergéométrique vers la loi binomiale, on considère généralement que le modèle d'urne ci-dessus est une bonne représentation du sondage.

de sorte qu'en notant

$$B_{n,t} := \left\{ \omega \in \Omega; \frac{S_n(\omega)}{n} - \frac{t}{2\sqrt{n}} \leq p \leq \frac{S_n(\omega)}{n} + \frac{t}{2\sqrt{n}} \right\},$$

l'inclusion $A_{n,t} \subset B_{n,t}$ nous donne :

$$P(B_{n,t}) \geq 2\Phi(t) - 1 + \varepsilon_n. \quad (1.20)$$

En pratique, n est fixé et on a observé des valeurs numériques explicites x_1, \dots, x_n que l'on interprète comme les valeurs de $X_1(\omega), \dots, X_n(\omega)$ pour *un même* ω tiré au sort (suivant P). On est donc en présence d'une valeur numérique explicite, $S_n(\omega)/n = (x_1 + \dots + x_n)/n$, disons pour fixer les idées $S_n(\omega)/n = 0,53$. Proposer pour le paramètre inconnu p l'intervalle de confiance

$$I_{n,t} = \left[0,53 - \frac{t}{2\sqrt{n}}; 0,53 + \frac{t}{2\sqrt{n}} \right],$$

c'est faire le *pari* que le ω observé est bien dans $B_{n,t}$. La probabilité de gagner ce pari est minorée par $2\Phi(t) - 1 + \varepsilon_n$. On dit que $I_{n,t}$ est un intervalle de confiance pour p avec un *niveau*¹¹ d'au moins $2\Phi(t) - 1 + \varepsilon_n$. En pratique, on laisse tomber le ε_n et on détermine t de façon approchée grâce à la tabulation de Φ . Par exemple pour un niveau de confiance de 95%, on est ramené à la résolution de l'équation $\Phi(t) = 1,95/2 = 0,975$ d'où $t \simeq 1,96$, ce qui nous donne l'intervalle

$$I_n = \left[\frac{S_n(\omega)}{n} - \frac{1,96}{2\sqrt{n}}; \frac{S_n(\omega)}{n} + \frac{1,96}{2\sqrt{n}} \right], \quad \text{au niveau de confiance 95\%}.$$

En fait les statisticiens préfèrent une variante de cette méthode pour obtenir des intervalles de confiance plus étroits, notamment quand p n'est pas trop proche de $1/2$. L'idée est de remplacer la variance inconnue pq de X_1 par un *estimateur* au lieu de la majorer de façon certaine par (1.19). Ainsi en estimant pq par $M_n(1 - M_n)$ où $M_n := S_n/n$, on obtient au niveau de confiance 95% l'intervalle

$$J_n = \left[M_n(\omega) - 1,96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}}; M_n(\omega) + 1,96\sqrt{\frac{M_n(\omega)(1 - M_n(\omega))}{n}} \right].$$

Cette deuxième méthode sera justifiée ultérieurement (voir l'exemple 3.20, p. 101).

11. Il y a ici un piège sémantique : supposons qu'on ait trouvé $I_{n,t} = [0,51; 0,55]$ avec un niveau de confiance de 95%. Il est tout à fait incorrect d'écrire « $P(p \in [0,51; 0,55]) \geq 0,95$ ». En effet, p n'a rien d'aléatoire, c'est une constante. L'aléatoire concerne notre ignorance sur sa valeur. Même si on considère p comme une variable aléatoire constante, la probabilité de son appartenance à $[0,51; 0,55]$ vaut 0 ou 1, et comme on ne peut pas exclure le premier cas, on ne peut pas minorer cette probabilité par 0,95.

1.2.4 Généralisation du TLC

Le théorème limite central a un domaine de validité bien plus vaste que celui des suites de variables i.i.d. de carré intégrable. Nous allons en voir deux généralisations. Mais auparavant signalons que dans le cas des v.a. i.i.d., le théorème 1.10 est optimal : la convergence en loi de S_n^* équivaut à la condition $\mathbf{E}X_1^2 < +\infty$ (noter l'analogie avec la LFGN dans le cas i.i.d. où la convergence presque-sûre de S_n/n équivaut à la condition $\mathbf{E}|X_1| < +\infty$). On peut en effet démontrer le résultat suivant.

Théorème 1.14. *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires i.i.d. définies sur le même espace probabilisé et telle que S_n/\sqrt{n} converge en loi vers $\mathfrak{N}(0, 1)$. Alors X_1 est de carré intégrable, $\mathbf{E}X_1^2 = 1$ et $\mathbf{E}X_1 = 0$.*

Ce théorème contient essentiellement la réciproque du théorème 1.10. Plus précisément on en déduit que s'il existe $\mu \in \mathbb{R}$ et $\sigma > 0$ constantes¹² telles que $\sigma n^{-1/2}(S_n - n\mu)$ converge en loi vers¹³ une v.a. Z de loi $\mathfrak{N}(0, 1)$, alors $\mathbf{E}X_1^2 < +\infty$, $\mathbf{E}X_1 = \mu$ et $\text{Var } X_1 = \sigma^2$. En effet, il suffit d'appliquer le théorème 1.14 aux variables aléatoires $X'_i = (X_i - \mu)/\sigma$.

Passons aux généralisations du TLC. Au vu du théorème 1.14, on pressent¹⁴ que si on laisse tomber l'hypothèse d'équidistribution des X_i (i.e. l'hypothèse qu'elles ont même loi), on risque fort de le payer plus cher en termes d'intégrabilité des X_i , donc de devoir supposer que les X_i ont des moments d'ordre supérieur à 2. C'est exactement ce qui se passe avec la première généralisation du cas i.i.d., le théorème de Liapounov.

Théorème 1.15 (Liapounov). *Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé, indépendantes (mais pas forcément de même loi), centrées, ayant toutes un moment d'ordre $2+\delta$ (pour une certaine constante $\delta > 0$). On note $s_n^2 := \text{Var}(S_n)$ et on suppose $s_n > 0$. On suppose de plus vérifiée la condition de Liapounov :*

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^n \mathbf{E}|X_k|^{2+\delta} \xrightarrow{n \rightarrow +\infty} 0. \quad (1.21)$$

Alors S_n/s_n converge en loi vers Z de loi $\mathfrak{N}(0, 1)$.

Bien entendu, ce théorème n'est énoncé avec des variables centrées que pour alléger les notations. Si les X_k ne sont pas centrées, on se ramène à des variables centrées en appliquant le théorème aux $X'_k := X_k - \mathbf{E}X_k$, ce qui donne la convergence en loi vers Z de $(S_n - \mathbf{E}S_n)/s_n$.

En fait, on peut se passer de l'hypothèse d'existence de moments d'ordre strictement supérieur à 2, mais au prix d'une hypothèse qui peut sembler de prime abord un peu artificielle, la *condition de Lindeberg*. Les courageux et les curieux désireux de

12. Bien sûr sans supposer *a priori* que ces constantes ont un lien avec d'éventuels moments de X_1 !

13. En fait, on peut même se passer de l'hypothèse que la loi limite est gaussienne.

14. O.K., là j'exagère un peu et cette remarque a l'air un peu trop subjective. Mais si vous avez la curiosité d'aller regarder la preuve du théorème 1.14, vous verrez que l'hypothèse d'équidistribution des X_i y joue un rôle important.

comprendre sa provenance sont invités à lire la section 1.4. Le résultat suivant peut être considéré comme la généralisation optimale du TLC dans le cas de variables aléatoires indépendantes non équidistribuées.

Théorème 1.16 (Lindeberg). *Considérons le « tableau triangulaire » de variables aléatoires de ligne numéro n ($n \in \mathbb{N}^*$) :*

$$X_{n,1}, \dots, X_{n,i}, \dots, X_{n,k_n},$$

où ces k_n variables sont définies sur le même $(\Omega_n, \mathcal{F}_n, P_n)$, indépendantes, centrées, de carré intégrable. On note

$$S_n := \sum_{i=1}^{k_n} X_{n,i}, \quad \sigma_{n,i}^2 := \text{Var } X_{n,i}, \quad s_n^2 := \text{Var } S_n = \sum_{i=1}^{k_n} \sigma_{n,i}^2.$$

On suppose $s_n > 0$ pour tout n . Si de plus le tableau vérifie la condition de Lindeberg :

$$\forall \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}) \xrightarrow{n \rightarrow +\infty} 0, \quad (1.22)$$

alors S_n/s_n converge en loi vers une v.a. Z de loi $\mathfrak{N}(0, 1)$.

On peut montrer que la condition de Lindeberg implique que $\lim_{n \rightarrow +\infty} k_n = +\infty$, ce qui explique que l'on s'abstienne de mentionner que k_n tend vers l'infini dans les hypothèses. L'appellation « tableau triangulaire » est un abus de langage commode, la forme triangulaire n'apparaissant réellement que dans le cas particulier où $k_n = cn$. Remarquons qu'avec un tableau triangulaire, toutes les v.a. termes de la somme S_n peuvent changer quand on passe de S_n à S_{n+1} , c'est pour cela qu'il n'est pas nécessaire de supposer que des variables situées sur des lignes différentes du tableau sont définies sur le même espace probabilisé. L'espérance utilisée dans la formule (1.22) est définie relativement à la mesure de probabilité P_n , il serait plus correct de la noter \mathbf{E}_n au lieu de \mathbf{E} . Pour toute v.a. Y positive définie sur $(\Omega_n, \mathcal{F}_n, P_n)$, $\mathbf{E}_n Y := \int_0^{+\infty} P_n(Y > t) dt$, etc.

Nous allons vérifier à titre d'exercice que le théorème de Lindeberg contient à la fois le TLC cas i.i.d. (th. 1.10) et le théorème de Liapounov.

Commençons par le cas d'une suite $(X_i)_{i \geq 1}$, de variables aléatoires i.i.d. de carré intégrable. Considérons le tableau triangulaire dont la ligne n est formée des $k_n = n$ variables aléatoires X_1, X_2, \dots, X_n (i.e. $X_{n,i} = X_i$) définies sur $(\Omega_n, \mathcal{F}_n, P_n) = (\Omega, \mathcal{F}, P)$. Ces variables étant indépendantes et de même loi, $s_n^2 = n\sigma^2$ et les $X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}$ ont même loi. Par conséquent, la quantité censée tendre vers 0 dans la condition de Lindeberg s'écrit pour tout $\varepsilon > 0$,

$$\frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}) = \frac{1}{n\sigma^2} n \mathbf{E}(X_1^2 \mathbf{1}_{\{|X_1| > \varepsilon \sigma n^{1/2}\}}) = \frac{1}{\sigma^2} \mathbf{E}(X_1^2 \mathbf{1}_{\{|X_1| > \varepsilon \sigma n^{1/2}\}}).$$

La convergence vers 0 de cette dernière expression quand n tend vers $+\infty$ découle de l'hypothèse $\mathbf{E}(X_1^2) < +\infty$ via le théorème de convergence dominée (exercice). De façon

plus élémentaire, on peut aussi remarquer que si u_n est une suite de réels positifs tendant vers $+\infty$ et Y une v.a. positive

$$\mathbf{E}(Y \mathbf{1}_{\{Y > u_n\}}) = \int_0^{+\infty} P(Y \mathbf{1}_{\{Y > u_n\}} > t) dt \stackrel{*}{=} u_n P(Y > u_n) + \int_{u_n}^{+\infty} P(Y > t) dt.$$

Si Y est intégrable, cette dernière expression ¹⁵ tend vers 0 quand n tend vers $+\infty$ en raison de la convergence de l'intégrale généralisée $\int_0^{+\infty} P(Y > t) dt$ et par application de l'inégalité de Markov raffinée à Y , cf. cours d'I.P.É. Ensuite on prend $Y = X_1^2$ et $u_n = n\varepsilon^2\sigma^2$ en notant que $\{X_1^2 > u_n\} = \{|X_1| > u_n^{1/2}\}$. Nous avons donc bien vérifié que le théorème de Lindeberg implique le théorème 1.10.

Considérons maintenant le cas d'une suite $(X_i)_{i \geq 1}$, de variables aléatoires indépendantes, d'espérance nulle et vérifiant pour une certaine constante $\delta > 0$, $\mathbf{E}|X_i|^{2+\delta} < +\infty$ pour tout $i \geq 1$. Supposons de plus vérifiée la condition de Liapounov (1.21). Nous allons voir que l'on peut alors appliquer le théorème de Lindeberg au tableau triangulaire où $k_n = n$, la ligne n étant formée des variables aléatoires X_1, X_2, \dots, X_n (i.e. $X_{n,i} = X_i$) définies sur $(\Omega_n, \mathcal{F}_n, P_n) = (\Omega, \mathcal{F}, P)$. En notant que sur l'évènement $\{|X_i| \geq \varepsilon s_n\}$ on a $|X_i|^\delta / (\varepsilon s_n)^\delta \geq 1$, la majoration

$$\mathbf{E}(X_i^2 \mathbf{1}_{\{|X_i| \geq \varepsilon s_n\}}) \leq \mathbf{E}\left(X_i^2 \left(\frac{|X_i|}{\varepsilon s_n}\right)^\delta \mathbf{1}_{\{|X_i| \geq \varepsilon s_n\}}\right) \leq \frac{1}{\varepsilon^\delta s_n^\delta} \mathbf{E}|X_k|^{2+\delta},$$

nous montre que la condition de Lindeberg est vérifiée *via* l'inégalité

$$\frac{1}{s_n^2} \sum_{i=1}^n \mathbf{E}(X_i^2 \mathbf{1}_{\{|X_i| > \varepsilon s_n\}}) \leq \frac{1}{\varepsilon^\delta s_n^{2+\delta}} \sum_{k=1}^n \mathbf{E}|X_k|^{2+\delta}.$$

En conclusion, S_n/s_n converge en loi vers $\mathfrak{N}(0, 1)$ et nous avons ainsi vérifié que le théorème de Lindeberg implique celui de Liapounov.

1.3 Théorème limite central vectoriel

Il est courant en statistique d'observer des données de nature vectorielle. On a souvent recours dans ce contexte à des approximations gaussiennes dont la justification mathématique repose sur un théorème limite central vectoriel. Avant de présenter la version du TLC pour des suites i.i.d. de vecteurs aléatoires, il convient d'apporter quelques précisions sur les vecteurs aléatoires de carré intégrable et les vecteurs aléatoires gaussiens.

1.3.1 Espérance et covariance d'un vecteur aléatoire

En dimension 1, l'espérance et la variance d'une variable aléatoire, lorsqu'elles existent, permettent de se faire une idée de la localisation de la loi et de sa dispersion. Elles jouent un rôle important notamment dans le théorème limite central, cf. page 13. Nous allons étendre ces notions au cas des vecteurs aléatoires.

15. À vous de justifier soigneusement l'égalité étoilée.

Définition 1.17 (espérance d'un vecteur aléatoire). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de \mathbb{R}^d . On dit qu'il est intégrable si la variable aléatoire positive $\|X\|$ est intégrable ($\mathbf{E}\|X\| < +\infty$), ce qui équivaut à l'intégrabilité de chacune des composantes ($\forall i = 1, \dots, d, \mathbf{E}|X_i| < +\infty$). Dans ce cas on appelle espérance de X ou vecteur d'espérances de X le vecteur

$$\mathbf{E}X := (\mathbf{E}X_1, \dots, \mathbf{E}X_d). \quad (1.23)$$

C'est délibérément que nous n'avons pas précisé le choix de la norme dans cette définition. En effet toutes les normes sur \mathbb{R}^d sont équivalentes. Si donc $\|\cdot\|$ est une norme sur \mathbb{R}^d elle est équivalente en particulier à la norme $\|\cdot\|_1$ définie par $\|x\|_1 := |x_1| + \dots + |x_d|$. On en déduit l'existence de deux constantes a et b strictement positives telles que

$$a(|X_1| + \dots + |X_d|) \leq \|X\| \leq b(|X_1| + \dots + |X_d|). \quad (1.24)$$

De la première inégalité on tire $|X_i| \leq a^{-1}\|X\|$, ce qui montre que l'intégrabilité de $\|X\|$ implique celle de chaque X_i . De la seconde inégalité on déduit que si les v.a. X_i sont toutes intégrables, les $|X_i|$ le sont aussi, ainsi que leur somme finie indexée par i , d'où l'intégrabilité de $\|X\|$. Nous venons de vérifier que l'intégrabilité de $\|X\|$ équivaut à celle de toutes les X_i , ce qui montre aussi que l'intégrabilité de X ne dépend pas du choix de la norme.

Une propriété importante de l'espérance des v.a. est la *linéarité*. Sa généralisation aux vecteurs aléatoires est immédiate : si X et Y sont des vecteurs aléatoires intégrables de \mathbb{R}^d , a et b des scalaires quelconques, le vecteur aléatoire $aX + bY$ est intégrable et $\mathbf{E}(aX + bY) = a\mathbf{E}X + b\mathbf{E}Y$. Pour le vérifier, il suffit d'appliquer composante par composante, la linéarité de l'espérance des variables aléatoires réelles.

Le résultat suivant nous dit *grosso modo* que l'espérance commute avec les applications linéaires.

Proposition 1.18 (espérance et applications linéaires). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire intégrable de \mathbb{R}^d .

a) Pour toute forme linéaire $u : \mathbb{R}^d \rightarrow \mathbb{R}$, la v.a. réelle $u(X)$ est intégrable et

$$\mathbf{E}u(X) = u(\mathbf{E}X). \quad (1.25)$$

b) $\mathbf{E}X$ est le seul vecteur z de \mathbb{R}^d vérifiant $\mathbf{E}u(X) = u(z)$ pour toute forme linéaire u sur \mathbb{R}^d .

c) Si $A : \mathbb{R}^d \rightarrow \mathbb{R}^j$, $x \mapsto A(x)$ est une application linéaire, le vecteur aléatoire AX de \mathbb{R}^j est intégrable et

$$\mathbf{E}A(X) = A(\mathbf{E}X). \quad (1.26)$$

Preuve. Soit $u : \mathbb{R}^d \rightarrow \mathbb{R}$, $x \mapsto a_1x_1 + \dots + a_dx_d$ une forme linéaire quelconque. L'intégrabilité de la v.a. $u(X)$ résulte immédiatement¹⁶ de l'intégrabilité des X_i puisque $|u(X)| \leq |a_1||X_1| + \dots + |a_d||X_d|$. Par linéarité de l'espérance des v.a. réelles, on a

$$\mathbf{E}u(X) = \mathbf{E}(a_1X_1 + \dots + a_dX_d) = a_1\mathbf{E}X_1 + \dots + a_d\mathbf{E}X_d = u(\mathbf{E}X),$$

16. On pourrait aussi remarquer que $|u(X)| \leq \|u\| \|X\|$, d'où $\mathbf{E}|u(X)| \leq \|u\| \mathbf{E}\|X\| < +\infty$.

ce qui justifie (1.25).

D'après le a), il existe au moins un vecteur (non aléatoire) z de \mathbb{R}^d tel que pour toute u , $\mathbf{E}u(X) = u(z)$, c'est $z = \mathbf{E}X$. Si $z' \in \mathbb{R}^d$ a la même propriété, nécessairement $u(z) = u(z')$ pour toute forme linéaire u sur \mathbb{R}^d . Ceci s'écrit encore $u(z - z') = 0$ pour toute u , donc en particulier pour les d formes coordonnées $(x_1, \dots, x_d) \mapsto x_i$. Le vecteur $z - z'$ a ainsi toutes ses coordonnées nulles, donc $z - z' = 0$ et $z = z'$.

Soit $A : \mathbb{R}^d \rightarrow \mathbb{R}^j$, $x \mapsto A(x)$ une application linéaire. On sait que

$$\forall x \in \mathbb{R}^d, \quad \|A(x)\| \leq \|A\| \|x\|,$$

où $\|A\|$ désigne la « norme opérateur » de l'application linéaire A , définie par

$$\|A\| := \sup_{\|x\|=1} \frac{\|A(x)\|}{\|x\|}.$$

On en déduit immédiatement que si X est un vecteur aléatoire intégrable de \mathbb{R}^d , le vecteur aléatoire $Y = A(X)$ de \mathbb{R}^j est intégrable puisque

$$\mathbf{E}\|Y\| = \mathbf{E}\|A(X)\| \leq \mathbf{E}(\|A\| \|X\|) = \|A\| \mathbf{E}\|X\| < +\infty.$$

Ceci justifie l'existence de $\mathbf{E}Y$. Soit maintenant v une forme linéaire quelconque sur \mathbb{R}^j . D'après le a) appliqué avec \mathbb{R}^j et son dual, on a

$$v(\mathbf{E}Y) = \mathbf{E}v(Y) = \mathbf{E}v(A(X)) = \mathbf{E}u(X) \quad \text{avec } u := v \circ A.$$

Il est clair que u est une forme linéaire sur \mathbb{R}^d , donc d'après le a) appliqué cette fois avec \mathbb{R}^d

$$\mathbf{E}u(X) = u(\mathbf{E}X) = (v \circ A)(\mathbf{E}X) = v(A(\mathbf{E}X)).$$

Nous venons ainsi de vérifier que pour toute forme linéaire v sur \mathbb{R}^j , $v(\mathbf{E}Y) = v(A(\mathbf{E}X))$. On en déduit que $\mathbf{E}Y = A(\mathbf{E}X)$, ce qui établit (1.26). \square

Définition 1.19 (matrice de covariance). Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de carré intégrable, c.à.d. $\mathbf{E}\|X\|^2 < +\infty$. On appelle matrice de covariance de X la matrice carrée K de terme général

$$K_{i,j} := \text{Cov}(X_i, X_j), \quad i, j = 1, \dots, d.$$

En utilisant (1.24) comme ci-dessus on a $|X_i| \leq a^{-1}\|X\|$ ce qui montre que les composantes X_i d'un vecteur de carré intégrable sont des v.a. de carré intégrable. Ceci justifie l'existence des $\text{Cov}(X_i, X_j)$. La connaissance de la matrice de covariance K de X permet le calcul de $\text{Var } u(X)$ pour toute forme linéaire u sur \mathbb{R}^d .

Proposition 1.20. Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire de carré intégrable et $u : x = (x_1, \dots, x_d) \mapsto a_1x_1 + \dots + a_dx_d$ une forme linéaire sur \mathbb{R}^d . Alors la v.a. réelle $u(X)$ est de carré intégrable et

$$\text{Var } u(X) = \sum_{i,j=1}^d a_i a_j \text{Cov}(X_i, X_j). \quad (1.27)$$

Preuve. L'inégalité $|u(X)| \leq \|u\| \|X\|$ nous donne $\mathbf{E}|u(X)|^2 \leq \|u\|^2 \mathbf{E}\|X\|^2 < +\infty$, donc $u(X)$ est de carré intégrable et sa variance est définie. En appliquant la formule générale pour la variance d'une somme et la bilinéarité de la covariance, cf. cours d'I.P.É., on obtient :

$$\text{Var } u(X) = \text{Var} \left(\sum_{i=1}^d a_i X_i \right) = \sum_{i,j=1}^d \text{Cov}(a_i X_i, a_j X_j) = \sum_{i,j=1}^d a_i a_j \text{Cov}(X_i, X_j).$$

□

Pour l'étude des vecteurs aléatoires et de leur convergence en loi, les formes linéaires sur \mathbb{R}^d sont un outil très pratique car elles permettent de se ramener à la dimension 1. Nous admettrons le résultat suivant.

Proposition 1.21. *Soient X et Y deux vecteurs aléatoires de \mathbb{R}^d tels que pour toute forme linéaire $u : \mathbb{R}^d \rightarrow \mathbb{R}$, les variables aléatoires réelles $u(X)$ et $u(Y)$ aient même loi. Alors les vecteurs aléatoires X et Y ont même loi.*

Remarque 1.22. Attention à l'hypothèse « pour toute u », on ne peut pas l'affaiblir en la remplaçant par « pour toute u d'une base du dual de \mathbb{R}^d ». Sinon en prenant les formes coordonnées $u : x = (x_1, \dots, x_d) \mapsto x_i$, on en déduirait que si pour tout $i = 1, \dots, d$, X_i a même loi que Y_i , alors les vecteurs $X = (X_1, \dots, X_d)$ et $Y = (Y_1, \dots, Y_d)$ ont même loi. Or on sait, cf. cours d'I.P.É., que ceci est faux.

1.3.2 Vecteurs aléatoires gaussiens

Il est commode d'élargir la famille des lois gaussiennes sur \mathbb{R} en considérant qu'une variable aléatoire constante $X = c$, ou $X = c$ p.s., suit la loi gaussienne de paramètres $m = c$ et $\sigma = 0$. Bien sûr, cette loi est la mesure de Dirac au point c et n'a donc pas de densité par rapport à λ_1 . Avec cette convention, nous pouvons donner la définition suivante.

Définition 1.23 (vecteur aléatoire gaussien). *On dit que le vecteur aléatoire $X = (X_1, \dots, X_d)$ de \mathbb{R}^d est gaussien si pour toute forme linéaire u sur \mathbb{R}^d , la variable aléatoire réelle $u(X)$ suit une loi gaussienne sur \mathbb{R} , autrement dit si*

$$\forall (a_1, \dots, a_d) \in \mathbb{R}^d, \quad a_1 X_1 + \dots + a_d X_d \text{ est une v.a. gaussienne.} \quad (1.28)$$

Remarque 1.24. En prenant pour formes linéaires u les formes coordonnées $x \mapsto x_i$, $i = 1, \dots, d$, on voit immédiatement que si X est un vecteur gaussien, ses composantes X_1, \dots, X_d sont des variables aléatoires gaussiennes¹⁷. *La réciproque est fausse.* Pour s'en convaincre, on pourra étudier en exercice l'exemple suivant. On définit $X = (X_1, X_2)$ en prenant X_1 gaussienne $\mathfrak{N}(0, 1)$ et en posant $X_2 := X_1 \mathbf{1}_{\{|X_1| \leq 1\}} - X_1 \mathbf{1}_{\{|X_1| > 1\}}$. On peut alors vérifier que

17. Au sens élargi donné ci-dessus, certaines pouvant être constantes p.s.

- a) X_2 suit la loi $\mathfrak{N}(0, 1)$ (en distinguant les trois cas $x < -1$, $-1 \leq x \leq 1$ et $x > 1$, vérifiez que $P(X_2 \leq x) = \Phi(x) = P(X_1 \leq x)$) ;
- b) $X_1 + X_2$ ne suit pas une loi gaussienne, donc le vecteur (X_1, X_2) ne peut pas être gaussien.

Proposition 1.25 (paramètres d'un vecteur gaussien). *La loi d'un vecteur aléatoire gaussien $X = (X_1, \dots, X_d)$ est caractérisée par son vecteur d'espérances*

$$m = \mathbf{E}X := (\mathbf{E}X_1, \dots, \mathbf{E}X_d) \quad (1.29)$$

et sa matrice de covariance

$$K = [\text{Cov}(X_i, X_j)]_{1 \leq i, j \leq d}. \quad (1.30)$$

La loi de X sera notée $\mathfrak{N}(m, K)$.

Preuve. Soient deux vecteurs aléatoires gaussiens $X = (X_1, \dots, X_d)$ et $Y = (Y_1, \dots, Y_d)$ ayant même vecteur d'espérances et même matrice de covariance. D'après (1.25) et (1.27), pour toute forme linéaire u sur \mathbb{R}^d , $\mathbf{E}u(X) = u(\mathbf{E}X) = u(\mathbf{E}Y) = \mathbf{E}u(Y)$ et $\text{Var } u(X) = \text{Var } u(Y)$. Les deux variables aléatoires réelles gaussiennes $u(X)$ et $u(Y)$ ont donc mêmes paramètres, donc même loi $\mathfrak{N}(\mathbf{E}u(X), \text{Var}^{1/2} u(X))$. Ceci étant vrai pour toute forme linéaire u , on en déduit grâce à la proposition 1.21, que les vecteurs aléatoires X et Y ont même loi. La loi d'un vecteur gaussien X est donc bien caractérisée par les seuls paramètres $\mathbf{E}X$ et K . \square

Proposition 1.26 (famille des vecteurs gaussiens). *La famille des vecteurs aléatoires gaussiens de \mathbb{R}^d est stable par addition de vecteur constant, multiplication par un scalaire constant et addition de vecteurs aléatoires gaussiens indépendants. L'image d'un vecteur aléatoire gaussien de \mathbb{R}^d par une application linéaire $\mathbb{R}^d \rightarrow \mathbb{R}^j$ est un vecteur gaussien de \mathbb{R}^j .*

Preuve. On vérifie facilement que la famille des variables aléatoires gaussiennes est stable par transformations affines : si Z est gaussienne de loi $\mathfrak{N}(m, \sigma)$, $aZ + b$ est gaussienne¹⁸ de loi $\mathfrak{N}(am + b, |a|\sigma)$.

Soit $z = (z_1, \dots, z_d)$ un vecteur (non aléatoire) de \mathbb{R}^d et X un vecteur aléatoire gaussien. Pour vérifier que le vecteur aléatoire $Y := X + z$ est gaussien, on prend une forme linéaire quelconque u et on note que $u(Y) = u(X) + u(z)$ est la somme de la variable aléatoire gaussienne $u(X)$ et de la constante réelle $u(z)$, c'est donc une v.a. gaussienne ayant même variance que $u(X)$ et d'espérance $\mathbf{E}u(X) + u(z)$. Ainsi $u(Y)$ est une v.a. gaussienne pour toute forme linéaire u , donc Y est un vecteur gaussien. De même si c est une constante réelle, cX est gaussien car $u(cX) = cu(X)$ est une v.a. gaussienne de loi $\mathfrak{N}(c\mathbf{E}u(X), c\text{Var}^{1/2} u(X))$.

^{18.} Démontrez le en exercice de révision du cours d'I.P.É., en cherchant une relation entre la f.d.r. de Z et celle de $aZ + b$ à l'aide d'un changement de variable dans l'intégrale de la densité (cas $a \neq 0$ et Z non constante)...

Soient X et Y deux vecteurs aléatoires gaussiens de \mathbb{R}^d indépendants. Pour toute forme linéaire u sur \mathbb{R}^d , $u(X + Y) = u(X) + u(Y)$ est la somme de deux variables aléatoires gaussiennes $u(X)$ et $u(Y)$ qui ont hérité de l'indépendance de X et Y . Or on peut vérifier, cf. lemme 1.28 ci-dessous, que la somme de deux variables aléatoires gaussiennes *indépendantes* est encore une v.a. gaussienne. Ainsi $u(X + Y)$ est une v.a. gaussienne pour toute u , donc le vecteur $X + Y$ est gaussien.

Enfin si X est un vecteur gaussien de \mathbb{R}^d et $F : \mathbb{R}^d \rightarrow \mathbb{R}^j$ est une application linéaire, $Y = F(X)$ est un vecteur gaussien de \mathbb{R}^j car pour toute forme linéaire $v : \mathbb{R}^j \rightarrow \mathbb{R}$, $v(Y) = (v \circ F)(X) = u(X)$ avec $u := v \circ F$ forme linéaire sur \mathbb{R}^d . \square

Corollaire 1.27. *Si les variables aléatoires réelles X_1, \dots, X_d sont indépendantes et gaussiennes, alors le vecteur aléatoire $X = (X_1, \dots, X_d)$ est gaussien dans \mathbb{R}^d .*

Preuve. On raisonne par récurrence sur la dimension d de l'espace. D'abord pour $d = 2$, si X_1 et X_2 sont deux variables aléatoires gaussiennes¹⁹ indépendantes, elles engendrent par transformation linéaire les vecteurs gaussiens $(X_1, 0)$ et $(0, X_2)$ de \mathbb{R}^2 qui héritent de leur indépendance. Par conséquent leur somme $((X_1, 0) + (0, X_2) = (X_1, X_2))$ est encore un vecteur gaussien de \mathbb{R}^2 . Supposons maintenant le corollaire vrai pour \mathbb{R}^j pour tout $j = 2, \dots, d$ et montrons qu'il est encore vrai pour $j = d + 1$. En effet par transformations linéaires, $(X_1, \dots, X_d, 0)$ et $(0, \dots, 0, X_{d+1})$ sont des vecteurs gaussiens de \mathbb{R}^{d+1} qui héritent de l'indépendance de (X_1, \dots, X_d) et X_{d+1} , par hérédité de l'indépendance pour des fonctions mesurables de blocs disjoints, cf. cours d'I.P.É. Donc leur somme $(X_1, \dots, X_d, X_{d+1})$ est un vecteur gaussien de \mathbb{R}^{d+1} . \square

Lemme 1.28. *La somme de deux variables aléatoires gaussiennes indépendantes est encore une variable aléatoire gaussienne.*

Preuve. Soient Y_1 et Y_2 deux v.a. gaussiennes indépendantes de paramètres respectifs (m_1, σ_1) et (m_2, σ_2) . Si ces deux v.a. sont constantes, $\sigma_1 = \sigma_2 = 0$, leur somme est la constante $m_1 + m_2$ donc est encore gaussienne. Si une seule des deux est constante, disons Y_2 , on sait que $Y_1 + Y_2 = Y_1 + m_2$ est gaussienne de paramètres $m_1 + m_2$ et σ_1 . Il nous reste à traiter le cas où $\sigma_1 > 0$ et $\sigma_2 > 0$. En posant $X_1 := Y_1 - m_1$ et $X_2 := Y_2 - m_2$, on obtient deux nouvelles v.a. gaussiennes indépendantes de loi respective $\mathfrak{N}(0, \sigma_1)$ et $\mathfrak{N}(0, \sigma_2)$. Il suffit de vérifier que $X_1 + X_2$ est une v.a. gaussienne $\mathfrak{N}(0, \sigma)$ et on en déduira que $Y_1 + Y_2$ a pour loi $\mathfrak{N}(m_1 + m_2, \sigma)$.

Puisque σ_1 et σ_2 sont non nuls, X_1 et X_2 ont chacune une densité donnée par :

$$f_i(t) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma_i^2}\right), \quad i = 1, 2.$$

On sait alors qu'en raison de l'indépendance de X_1 et X_2 , la v.a. $X_1 + X_2$ admet elle aussi une densité g qui est la convoluée $f_1 * f_2$ définie par $g(t) = \int_{-\infty}^{+\infty} f_1(s)f_2(t-s) ds$, soit ici :

$$\forall t \in \mathbb{R}, \quad g(t) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} \exp\left(-\frac{s^2}{2\sigma_1^2} - \frac{(t-s)^2}{2\sigma_2^2}\right) ds. \quad (1.31)$$

^{19.} Une variable aléatoire gaussienne peut être vue comme un vecteur aléatoire gaussien de dimension 1. Justifiez cette affirmation.

Pour vérifier que g est bien une densité gaussienne, nous allons transformer la forme quadratique en s, t à l'intérieur de l'exponentielle pour la mettre sous la forme

$$\frac{s^2}{2\sigma_1^2} + \frac{(t-s)^2}{2\sigma_2^2} = c((s-at)^2 + bt^2).$$

Ensuite le changement de variable $x = s - at$ dans l'intégrale (on intègre relativement à s , donc at joue le rôle d'une constante) nous permettra de sortir tout ce qui dépend de t de l'intégrale pour aboutir à $g(t) = c' \exp(-b't^2)$, l'intégrale devenant alors une constante absorbée dans le c' .

Passons aux détails du calcul. Il est commode de poser $\sigma^2 := \sigma_1^2 + \sigma_2^2$ et $\tau_i := \sigma_i^2/\sigma^2$.

$$\begin{aligned} \frac{s^2}{2\sigma_1^2} + \frac{(t-s)^2}{2\sigma_2^2} &= \frac{\sigma_2^2 s^2 + \sigma_1^2 (t-s)^2}{2\sigma_1^2 \sigma_2^2} = \frac{\sigma^2 s^2 - 2\sigma_1^2 s t + \sigma_1^2 t^2}{2\sigma_1^2 \sigma_2^2} \\ &= \frac{\sigma^2}{2\sigma_1^2 \sigma_2^2} (s^2 - 2\tau_1 s t + \tau_1 t^2) \\ &= \frac{\sigma^2}{2\sigma_1^2 \sigma_2^2} ((s - \tau_1 t)^2 + (\tau_1 - \tau_1^2) t^2) \\ &= \frac{\sigma^2}{2\sigma_1^2 \sigma_2^2} (s - \tau_1 t)^2 + \frac{t^2}{2\sigma^2}, \end{aligned}$$

en notant que $\tau_1 + \tau_2 = 1$, d'où $\tau_1 - \tau_1^2 = \tau_1 \tau_2 = \sigma_1^2 \sigma_2^2 \sigma^{-4}$. En reportant ce calcul dans (1.31), on a donc

$$g(t) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{t^2}{2\sigma^2}\right) \int_{-\infty}^{+\infty} \exp\left(-\frac{\sigma^2}{2\sigma_1^2\sigma_2^2}(s - \tau_1 t)^2\right) ds.$$

Le changement de variable $x = \sigma(\sigma_1\sigma_2)^{-1}(s - \tau_1 t)$ dans cette intégrale généralisée nous permet de nous ramener après simplifications à l'intégrale de la densité gaussienne standard :

$$g(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

On en déduit que $X_1 + X_2$ suit la loi gaussienne $\mathfrak{N}(0, \sigma)$ avec $\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2}$. \square

Remarque 1.29. Jusqu'à présent nous n'avons pas donné d'exemple de vecteur gaussien en dimension $d > 1$. Le corollaire 1.27 nous donne une première famille d'exemples, les vecteurs gaussiens construits en prenant leurs composantes gaussiennes indépendantes. En combinant ceci avec la proposition 1.26, on obtient une classe d'exemples bien plus large, celle des vecteurs aléatoires Y qui peuvent s'écrire

$$Y = A(X) + b \tag{1.32}$$

où $X = (X_1, \dots, X_d)$ est gaussien à composantes indépendantes, A est une application linéaire quelconque de \mathbb{R}^d dans \mathbb{R}^d et $b = (b_1, \dots, b_d)$ un vecteur constant.

En fait *tout* vecteur gaussien de \mathbb{R}^d admet une représentation de la forme (1.32). Pour une preuve, allez voir la méthode de simulation d'un vecteur gaussien dans le chapitre 2 et complétez les détails.

1.3.3 TLC vectoriel

Nous admettons le résultat suivant qui permet de ramener la convergence en loi en dimension d à de la convergence en loi en dimension 1. Dans la littérature anglo-américaine, ce résultat est connu sous le sobriquet de *Cramér-Wold device*, ce que l'on pourrait traduire par « truc de Cramér-Wold » ou de façon moins irrévérencieuse, par « astuce de Cramér-Wold ».

Lemme 1.30 (Cramér-Wold). *La suite $(Y_n)_{n \geq 1}$ de vecteurs aléatoires dans \mathbb{R}^d converge en loi dans \mathbb{R}^d vers le vecteur aléatoire Y si et seulement si pour toute forme linéaire $u : \mathbb{R}^d \rightarrow \mathbb{R}$, la suite de variables aléatoires réelles $(u(Y_n))_{n \geq 1}$ converge en loi dans \mathbb{R} vers $u(Y)$.*

Théorème 1.31. *Soit (X_k) une suite de vecteurs aléatoires de \mathbb{R}^d , indépendants, de même loi et de carré intégrable (i.e. $\mathbf{E}\|X_1\|^2 < +\infty$) et $S_n := \sum_{k=1}^n X_k$. Alors*

$$\frac{S_n - \mathbf{E}S_n}{\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z, \quad (1.33)$$

où Z est gaussien de loi $\mathfrak{N}(0, K)$, K étant la matrice de covariance de X_1 .

Malgré la ressemblance formelle de cet énoncé avec celui du théorème 1.10, il n'est pas possible ici de modifier la normalisation \sqrt{n} par $\sigma\sqrt{n}$ pour avoir toujours la même loi limite $\mathfrak{N}(0, I)$, où I est la matrice identité par rapport à la base canonique de \mathbb{R}^d .

Preuve. Par le lemme de Cramér-Wold, il suffit de vérifier que pour toute forme linéaire u sur \mathbb{R}^d , $u(n^{-1/2}(S_n - \mathbf{E}S_n))$ converge en loi vers $u(Z)$ où Z désigne un vecteur aléatoire gaussien de loi $\mathfrak{N}(0, K)$. Notons d'emblée que $\mathbf{E}u(Z) = u(\mathbf{E}Z) = u(0) = 0$ et $\text{Var } u(Z) = \text{Var } u(X_1)$ puisque Z et X_1 ont même matrice de covariance, cf. proposition 1.20.

Par linéarité de u ,

$$u\left(\frac{S_n - \mathbf{E}S_n}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}} \sum_{k=1}^n (u(X_k) - \mathbf{E}u(X_k)).$$

Les $u(X_k)$ sont des variables aléatoires réelles indépendantes, de même loi, de carré intégrable ($\mathbf{E}|u(X_1)|^2 \leq \|u\|^2 \mathbf{E}\|X_1\|^2$), on peut donc appliquer le théorème 1.10 qui donne ici

$$u\left(\frac{S_n - \mathbf{E}S_n}{\sqrt{n}}\right) \xrightarrow[n \rightarrow +\infty]{\text{loi}} W,$$

où W désigne n'importe quelle v.a. gaussienne de loi $\mathfrak{N}(0, \text{Var } u(X_1))$. Cette loi est celle de $u(Z)$, donc $u(n^{-1/2}(S_n - \mathbf{E}S_n))$ converge en loi vers $u(Z)$. \square

Une application importante du théorème 1.31 est la convergence d'une loi multinomiale vers une loi gaussienne, ce qui en un certain sens, généralise le théorème de de Moivre Laplace.

Rappelons que la loi multinomiale sert à modéliser le total des résultats observés pour chaque type dans une suite d'épreuves répétées indépendantes ayant chacune d types

de résultats possibles. Par exemple si on lance 200 fois un dé, on obtient un vecteur de dimension 6 dont la i -ème composante est le nombre total d'apparitions de la face numéro i au cours des 200 lancers. Ce vecteur suit la loi multinomiale de paramètres 200 et $(p_1, p_2, p_3, p_4, p_5, p_6)$, où les p_i valent tous $1/6$ si le dé est équilibré. Plus formellement, le vecteur aléatoire N suit la loi multinomiale de paramètres n et $p = (p_1, \dots, p_d)$ où $n \in \mathbb{N}^*$ et les p_i sont strictement positifs et de somme 1, si pour tout d -uplet (j_1, j_2, \dots, j_d) d'entiers tels que $j_1 + j_2 + \dots + j_d = n$,

$$P(N = (j_1, j_2, \dots, j_d)) = \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}$$

et $P(N = (j_1, j_2, \dots, j_d)) = 0$ si $j_1 + j_2 + \dots + j_d \neq n$.

Théorème 1.32. Si $(N_n)_{n \geq 1}$ est une suite de vecteurs aléatoires de \mathbb{R}^d de loi multinomiale de paramètres n et $p = (p_1, \dots, p_d)$,

$$\frac{1}{\sqrt{n}}(N_n - np) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z \quad \text{de loi } \mathfrak{N}(0, K), \quad (1.34)$$

où K est la matrice de covariance de terme général

$$K_{i,j} = p_i \delta_{i,j} - p_i p_j, \quad 1 \leq i, j \leq d.$$

Preuve. Le vecteur aléatoire N_n a même loi que $S_n := \sum_{k=1}^n X_k$, où les X_k sont des vecteurs aléatoires de \mathbb{R}^d , indépendants et de même loi donnée par

$$P\{X_k = (0, \dots, 0, \underset{\substack{\uparrow \\ i}}{1}, 0, \dots, 0)\} = p_i, \quad 1 \leq i \leq d.$$

Autrement dit, on se ramène à un modèle d'épreuves répétées indépendantes et le vecteur aléatoire X_k est un codage binaire du résultat de la k^e épreuve. En notant $X_{k,i}$ la i^e composante de X_k , on voit que cette variable aléatoire suit la loi de Bernoulli de paramètre p_i et que $\sum_{i=1}^d X_{k,i} = 1$. L'espérance de X_k est clairement le vecteur p . Le vecteur X_k étant borné est clairement de carré intégrable. On est ainsi dans les conditions de validité du théorème 1.31 qui nous donne

$$T_n := \frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - p) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z \quad \text{de loi } \mathfrak{N}(0, K), \quad (1.35)$$

où le terme général de la matrice de K par rapport à la base canonique de \mathbb{R}^d est

$$K_{i,j} = \text{Cov}(X_{1,i}, X_{1,j}) = \mathbf{E}(X_{1,i} X_{1,j}) - p_i p_j.$$

Si $i \neq j$, $X_{1,i} X_{1,j} = 0$ comme produit d'indicatrices d'événements disjoints. Pour $i = j$, $X_{1,i}^2 = X_{1,i}$ car $X_{1,i}$ ne peut prendre que la valeur 0 ou 1. On a donc bien dans les deux cas $K_{i,j} = p_i \delta_{i,j} - p_i p_j$ et on conclut en remarquant que T_n et $n^{-1/2}(N_n - np)$ ont même loi. \square

Remarque 1.33. La loi limite dans ce théorème est un exemple naturel de loi gaussienne sur \mathbb{R}^d sans densité par rapport à λ_d , la mesure de Lebesgue sur \mathbb{R}^d . On montre en effet que

$$P(Z_1 + Z_2 + \cdots + Z_d = 0) = 1 \quad \text{où} \quad Z = (Z_1, \dots, Z_d) \sim \mathfrak{N}(0, K). \quad (1.36)$$

Pour cela, considérons la forme linéaire

$$u(x) := x_1 + x_2 + \cdots + x_d, \quad x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d.$$

La continuité de u et (1.35) entraînent la convergence en loi de $u(T_n)$ vers $u(Z)$. D'autre part, $u(T_n) = n^{-1/2} \sum_{k=1}^n (u(X_k) - u(p))$ par linéarité de u . Comme $u(X_k) = 1 = u(p)$, $u(T_n)$ est la variable aléatoire nulle, sa loi est la masse de Dirac en 0 et il en est de même pour la loi limite quand n tend vers $+\infty$. Mais dire que $u(Z)$ a pour loi δ_0 revient exactement à (1.36). Ainsi toute la masse de la loi de Z est portée par l'hyperplan $x_1 + x_2 + \cdots + x_d = 0$ qui est de λ_d -mesure nulle et la loi de Z ne peut avoir de densité par rapport à λ_d .

1.4 Compléments sur la convergence en loi et le TLC

Cette section développe les aspects techniques de la convergence en loi dans le but de fournir autant que possible des preuves de résultats admis au cours de ce chapitre. En première lecture on pourra se contenter de jeter un coup d'oeil aux énoncés, quitte à y revenir plus tard en cas de besoin.

1.4.1 Outillage pour la convergence en loi

Avant de nous focaliser sur l'étude de la convergence en loi par les moments fonctionnels donc au sens de la définition 1.2, nous montrons que celle-ci implique la convergence en loi au sens de la définition 1.1. La réciproque est vraie mais plus difficile à démontrer et n'est de toutes façons quasiment pas utilisée dans ce cours. On s'en dispensera donc²⁰.

Proposition 1.34. Notons F_n et F les fonctions de répartition respectives des variables aléatoires réelles Y_n ($n \geq 1$) et Y . On suppose que

$$\forall h \text{ continue bornée } \mathbb{R} \rightarrow \mathbb{R}, \quad \mathbf{E}h(Y_n) \xrightarrow{n \rightarrow +\infty} \mathbf{E}h(Y). \quad (1.37)$$

Alors

$$\forall x \text{ point de continuité de } F, \quad F_n(x) \xrightarrow{n \rightarrow +\infty} F(x). \quad (1.38)$$

Preuve. Fixons $x \in \mathbb{R}$ et remarquons que $F_n(x) = P(Y_n \leq x) = \mathbf{E}\mathbf{1}_{]-\infty, x]}(Y_n)$ et $F(x) = \mathbf{E}\mathbf{1}_{]-\infty, x]}(Y)$. La fonction $f : t \mapsto \mathbf{1}_{]-\infty, x]}(t)$ est bornée sur \mathbb{R} mais malheureusement, à cause de son unique discontinuité au point x , on ne peut pas appliquer directement (1.37) pour conclure.

Nous allons approximer f en l'encadrant par deux fonctions continues bornées, affines par morceaux, cf. figure 1.2. De façon plus formelle, si avec un paramètre $\delta > 0$, on définit

²⁰. Les curieux peuvent aller voir la preuve du théorème 9.32 dans le cours d'IFP 2004 à <http://math.univ-lille1.fr/~suquet/ens/IFP/indexIFP.html>

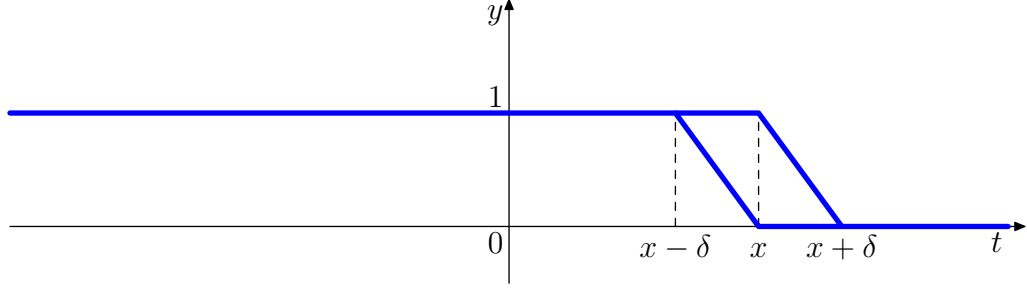


FIG. 1.2 – Encadrement de $f = \mathbf{1}_{]-\infty, x]}$ par 2 fonctions continues

la fonction g par

$$g(t) = \begin{cases} 1 & \text{si } t \in]-\infty, x] \\ 1 - (t - x)/\delta & \text{si } t \in]x, x + \delta] \\ 0 & \text{si } t \in]x + \delta, +\infty[\end{cases}$$

on a pour tout $t \in \mathbb{R}$, $g(t + \delta) \leq f(t) \leq g(t)$, d'où

$$\forall t \in \mathbb{R}, \quad f(t + \delta) \leq g(t + \delta) \leq f(t) \leq g(t) \leq f(t - \delta). \quad (1.39)$$

Grâce à ces encadrements nous pouvons contrôler pour Z v.a. quelconque, la différence $|P(Z \leq x) - P(Y \leq x)|$ à l'aide des moments fonctionnels des v.a. Y et Z . En effet

$$\begin{aligned} P(Z \leq x) - P(Y \leq x) &= \mathbf{E}f(Z) - \mathbf{E}f(Y) \\ &\leq \mathbf{E}g(Z) - \mathbf{E}f(Y) \\ &= \mathbf{E}g(Z) - \mathbf{E}g(Y) + \mathbf{E}g(Y) - \mathbf{E}f(Y) \\ &\leq \mathbf{E}g(Z) - \mathbf{E}g(Y) + \mathbf{E}f(Y - \delta) - \mathbf{E}f(Y) \\ &= \mathbf{E}g(Z) - \mathbf{E}g(Y) + F(x + \delta) - F(x), \end{aligned}$$

en rappelant que F est la f.d.r. de Y . De même

$$\begin{aligned} P(Y \leq x) - P(Z \leq x) &= \mathbf{E}f(Y) - \mathbf{E}f(Z) \\ &\leq \mathbf{E}f(Y) - \mathbf{E}g(Z + \delta) \\ &= \mathbf{E}f(Y) - \mathbf{E}g(Y + \delta) + \mathbf{E}g(Y + \delta) - \mathbf{E}g(Z + \delta) \\ &\leq \mathbf{E}f(Y) - \mathbf{E}f(Y + \delta) + \mathbf{E}g(Y + \delta) - \mathbf{E}g(Z + \delta) \\ &= F(x) - F(x - \delta) + \mathbf{E}g(Y + \delta) - \mathbf{E}g(Z + \delta). \end{aligned}$$

On en déduit la majoration

$$|P(Z \leq x) - P(Y \leq x)| \leq \max(|\mathbf{E}g(Z) - \mathbf{E}g(Y)|; |\mathbf{E}g(Y + \delta) - \mathbf{E}g(Z + \delta)|) + \Delta(F, x, \delta), \quad (1.40)$$

avec

$$\Delta(F, x, \delta) := \max(F(x) - F(x - \delta); F(x + \delta) - F(x)). \quad (1.41)$$

Cette inégalité est valable pour tout x réel, tout $\delta > 0$, toute v.a. Z , avec la fonction g dépendant de x et δ . Si x est un point de continuité de F , $\Delta(F, x, \delta)$ tend vers 0 quand δ tend vers 0. Fixons $\varepsilon > 0$ quelconque. On peut alors trouver un δ assez petit pour que $\Delta(F, x, \delta) < \varepsilon$. Une fois choisi ce δ , les fonctions continues bornées g et $g(\cdot + \delta)$ sont fixées. En prenant alors $Z = Y_n$ dans (1.40), l'hypothèse (1.37) appliquée avec $h = g$ et $h = g(\cdot + \delta)$ nous donne l'existence d'un entier n_0 tel que

$$\forall n \geq n_0, \quad \max(|\mathbf{E}g(Y_n) - \mathbf{E}g(Y)|; |\mathbf{E}g(Y + \delta) - \mathbf{E}g(Y_n + \delta)|) < \varepsilon.$$

Ainsi pour tout $n \geq n_0$, $|P(Y_n \leq x) - P(Y \leq x)| < 2\varepsilon$. Comme $\varepsilon > 0$ était arbitraire, la convergence de $F_n(x)$ vers $F(x)$ est établie. \square

La définition 1.2 de la convergence en loi par les moments fonctionnels $\mathbf{E}h(Y_n)$ pour toute fonction h continue bornée est bien commode pour démontrer des propriétés de cette convergence, cf. la preuve de la proposition 1.6. Par contre, quand on veut vérifier pratiquement qu'une suite donnée Y_n de v.a. ou de vecteurs aléatoires converge en loi vers Y , on a tout intérêt à pouvoir réduire la classe des fonctions h considérée. Nous débutons l'étude de ce type de réduction par l'importante notion d'équitension.

Définition 1.35 (équitension). Une famille $\{P_i, i \in I\}$ de mesures de probabilité sur \mathbb{R}^d est dite équitendue si

$$\forall \varepsilon > 0, \exists K \text{ compact de } \mathbb{R}^d \text{ tel que } \forall i \in I, P_i(\mathbb{R}^d \setminus K) < \varepsilon. \quad (1.42)$$

Une suite $(X_n)_{n \geq 1}$ de vecteurs aléatoires de \mathbb{R}^d est dite équitendue si la suite des lois $(P_{X_n})_{n \geq 1}$ est équitendue au sens précédent. Ceci se traduit par

$$\forall \varepsilon > 0, \exists K \text{ compact de } \mathbb{R}^d \text{ tel que } \forall n \geq 1, P(X_n \notin K) < \varepsilon, \quad (1.43)$$

ou de manière équivalente par

$$\forall \varepsilon > 0, \exists K \text{ compact de } \mathbb{R}^d \text{ tel que } \forall n \geq 1, P(X_n \in K) > 1 - \varepsilon. \quad (1.44)$$

Remarque 1.36. On sait que tout compact est borné et qu'en dimension finie, la fermeture de tout ensemble borné est compacte. On pourrait donc sans dommage remplacer dans la définition ci-dessus K compact par K borné et même plus simplement remplacer K par $[-a, a]^d$ pour $a > 0$. L'importance de prendre vraiment K compact au lieu de borné n'apparaît que lorsqu'on étudie les vecteurs aléatoires en dimension infinie. En pratique, une condition nécessaire et suffisante d'équitension de $(X_n)_{n \geq 1}$ dans \mathbb{R}^d est donc

$$\forall \varepsilon > 0, \exists a > 0, \quad \forall n \geq 1, P(\|X_n\| > a) < \varepsilon. \quad (1.45)$$

Bien entendu, le choix de la norme n'a ici aucune importance, autre que de confort, puisque toutes les normes sur \mathbb{R}^d sont équivalentes.

Remarque 1.37. Une famille finie $\{P_i, i \in I\}$ de mesures de probabilité sur \mathbb{R}^d est toujours équitendue. Vérifions le d'abord dans le cas où la famille est réduite à une seule mesure²¹ P_0 . Posons $C_k := [-k, k]^d$. Les C_k sont des compacts de \mathbb{R}^d et la suite $(C_k)_{k \geq 1}$ est croissante de réunion \mathbb{R}^d . Donc $P_0(C_k) \uparrow 1$ quand k tend vers l'infini, par continuité séquentielle croissante de P_0 . Pour tout $\varepsilon > 0$, on a alors un $k_0 = k_0(\varepsilon)$ tel que $P_0(C_{k_0}) > 1 - \varepsilon$, ce qui équivaut à $P_0(\mathbb{R}^d \setminus C_{k_0}) < \varepsilon$. Maintenant si I est un ensemble fini quelconque d'indices, on peut construire comme ci-dessus pour chaque $i \in I$ un $k_i = k_i(\varepsilon)$ tel que $P_i(C_{k_i}) > 1 - \varepsilon$. En prenant le plus grand élément de cet ensemble fini d'entiers k_i , appelons le m , tous les C_{k_i} sont inclus dans le compact C_m et on a $P_i(C_m) > 1 - \varepsilon$ pour tout $i \in I$, la famille est donc équitendue. Évidemment ce raisonnement s'écroule si I est infini, car alors rien ne garantit que m , défini maintenant comme le sup des $k_i(\varepsilon)$, soit fini. Or si m vaut $+\infty$, $C_m = \mathbb{R}^d$ n'est plus compact.

Remarque 1.38. La réunion de deux familles équitendues $\{P_i, i \in I\}$ et $\{Q_j, j \in J\}$ de mesures de probabilité sur \mathbb{R}^d est encore équitendue. En effet si K et K' sont des compacts tels que pour tout i , $P_i(K) > 1 - \varepsilon$ et pour tout j , $Q_j(K') > 1 - \varepsilon$, alors $K \cup K'$ est un compact vérifiant $P_i(K \cup K') > 1 - \varepsilon$ et $Q_j(K \cup K') > 1 - \varepsilon$ pour tout $i \in I$ et tout $j \in J$. En particulier, compte-tenu de la remarque 1.37, si $(X_n)_{n \geq 1}$ est une suite équitendue de vecteurs aléatoires de \mathbb{R}^d et X un vecteur aléatoire *quelconque* de \mathbb{R}^d , la famille $\{X_n, n \geq 1\} \cup \{X\}$ est équitendue.

Voici maintenant une première utilisation de l'équitension pour établir un résultat intermédiaire.

Lemme 1.39. Soient Z_n ($n \geq 1$) et Z des variables aléatoires réelles vérifiant

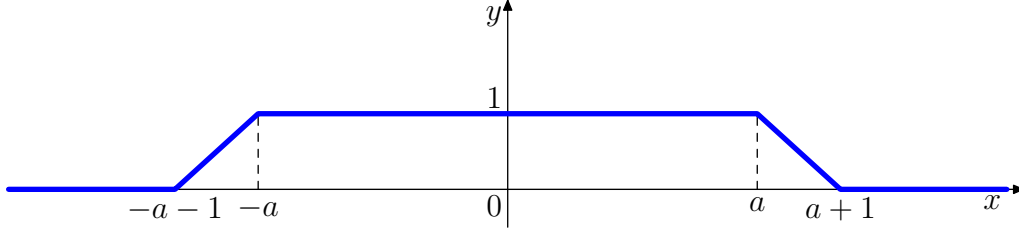
- i) $(Z_n)_{n \geq 1}$ est équitendue,
- ii) pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact²², $\mathbf{E}g(Z_n)$ converge vers $\mathbf{E}g(Z)$ quand n tend vers l'infini.

Alors Z_n converge en loi vers Z .

Preuve. L'hypothèse ii) est *a priori* plus faible que la convergence en loi. En effet toute fonction continue à support compact est évidemment continue bornée mais la réciproque est fautive. Il nous faut montrer que pour toute fonction *continue bornée* $h : \mathbb{R} \rightarrow \mathbb{R}$, $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$ quand n tend vers l'infini. Soit donc h une telle fonction et notons $\|h\|_\infty := \sup_{x \in \mathbb{R}} |h(x)|$. Fixons un $\varepsilon > 0$ arbitraire. Par l'hypothèse i) et la remarque 1.38, la famille $\{P_{Z_n}, n \geq 1\} \cup \{P_Z\}$ est équitendue. On dispose donc d'un réel $a > 0$ tel que

$$\forall n \geq 1, \quad P(|Z_n| > a) < \varepsilon \quad \text{et} \quad P(|Z| > a) < \varepsilon. \quad (1.46)$$

Notons f_a la fonction « trapèze » définie²³ par la figure 1.3 et posons $g := hf_a$. Alors g est continue comme produit de deux fonctions continues et nulle en dehors du compact


 FIG. 1.3 – Fonction « trapèze » f_a

$[-a-1, a+1]$. L'erreur commise en approximant $\mathbf{E}h(Z_n)$ par $\mathbf{E}g(Z_n)$ est égale par linéarité de l'espérance à $\mathbf{E}(h-g)(Z_n)$. Pour majorer cette quantité, on remarque que $|h-g|(Z_n) \leq \|h\|_\infty \mathbf{1}_{\{|Z_n|>a\}}$, car $0 \leq |h-g| = |h|(1-f_a) \leq |h|$ et $h-g$ est nulle sur $[-a, a]$. Par conséquent, compte-tenu de (1.46),

$$|\mathbf{E}(h-g)(Z_n)| \leq \|h\|_\infty \mathbf{E}\mathbf{1}_{\{|Z_n|>a\}} = \|h\|_\infty P(|Z_n| > a) < \|h\|_\infty \varepsilon. \quad (1.47)$$

Il est clair que la même inégalité est vérifiée avec Z à la place de Z_n .

D'autre part grâce à l'hypothèse ii), il existe un n_0 dépendant seulement de g et de ε tel que

$$\forall n \geq n_0, \quad |\mathbf{E}g(Z_n) - \mathbf{E}g(Z)| < \varepsilon. \quad (1.48)$$

Finalement en utilisant (1.47) appliquée à Z_n et à Z et (1.48), on voit que pour tout $n \geq n_0$,

$$\begin{aligned} |\mathbf{E}h(Z_n) - \mathbf{E}h(Z)| &\leq |\mathbf{E}(h-g)(Z_n)| + |\mathbf{E}(h-g)(Z)| + |\mathbf{E}g(Z_n) - \mathbf{E}g(Z)| \\ &\leq (2\|h\|_\infty + 1)\varepsilon. \end{aligned}$$

Comme ε était arbitraire, la convergence de $\mathbf{E}h(Z_n)$ vers $\mathbf{E}h(Z)$ est établie. \square

Proposition 1.40. Soient Z_n ($n \geq 1$) et Z des variables aléatoires réelles telles que pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ continue à support compact, $\mathbf{E}g(Z_n)$ converge vers $\mathbf{E}g(Z)$ quand n tend vers l'infini. Alors Z_n converge en loi vers Z .

Preuve. D'après le lemme 1.39, il suffit de prouver l'équitension de $(Z_n)_{n \geq 1}$. Fixons $\varepsilon > 0$ arbitraire. La variable aléatoire Z étant tendue (remarque 1.37), il existe un réel $a > 0$ tel que $P(|Z| \leq a) > 1 - \varepsilon$. La f.d.r. de $|Z|$ n'ayant qu'un ensemble au plus dénombrable de points de discontinuité, a au moins un point de continuité c dans l'ensemble non dénombrable $[a, +\infty[$. On a alors

$$P(|Z| \leq c) > 1 - \varepsilon \quad \text{et} \quad \lim_{x \rightarrow c} P(|Z| \leq x) = P(|Z| \leq c). \quad (1.49)$$

Pour $0 < \delta < c$, soit g_δ la fonction trapèze nulle en dehors de $[-c, c]$, valant 1 sur $[-c + \delta, c - \delta]$, cf. figure 1.4, donc vérifiant l'encadrement $\mathbf{1}_{[-c+\delta, c-\delta]} \leq g_\delta \leq \mathbf{1}_{[-c, c]}$. De

21. On dit alors que cette mesure est « tendue », le préfixe « équi » sous-entendant qu'il y a plusieurs mesures dans la famille. De même on parle de fonction continue et de famille équicontinue de fonctions.

22. Donc continue sur \mathbb{R} et nulle en dehors d'un certain compact de \mathbb{R} .

23. Pour ceux qui pensent qu'une figure ne définit pas une fonction et qui préfèrent les formules obscures, disons $f_a(x) := \mathbf{1}_{[0, a]}(|x|) + (a+1-|x|)\mathbf{1}_{[a, a+1]}(|x|)$.

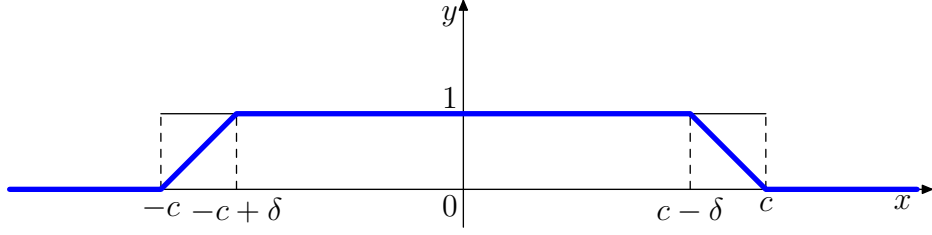


FIG. 1.4 – Fonction « trapèze » g_δ vérifiant $\mathbf{1}_{[-c+\delta, c-\delta]} \leq g_\delta \leq \mathbf{1}_{[-c, c]}$

cet encadrement on déduit que

$$P(|Z| \leq c - \delta) = \mathbf{E}\mathbf{1}_{[-c+\delta, c-\delta]}(Z) \leq \mathbf{E}g_\delta(Z) \leq \mathbf{E}\mathbf{1}_{[-c, c]}(Z) = P(|Z| \leq c).$$

D'après (1.49) on en déduit qu'en prenant δ suffisamment petit,

$$\mathbf{E}g_\delta(Z) \geq P(|Z| \leq c - \delta) > P(|Z| \leq c) - \varepsilon > 1 - 2\varepsilon. \quad (1.50)$$

Fixons désormais δ vérifiant (1.50). Comme g_δ est continue à support compact, l'hypothèse nous donne un n_0 tel que

$$\forall n \geq n_0, \quad \mathbf{E}g_\delta(Z_n) > \mathbf{E}g_\delta(Z) - \varepsilon. \quad (1.51)$$

Comme $g_\delta \leq \mathbf{1}_{[-c, c]}$, $P(|Z_n| \leq c) = \mathbf{E}\mathbf{1}_{[-c, c]}(Z_n) \geq \mathbf{E}g_\delta(Z_n)$, d'où en combinant cette minoration avec (1.50) et (1.51),

$$\forall n \geq n_0, \quad P(|Z_n| \leq c) > 1 - 3\varepsilon. \quad (1.52)$$

D'autre part la famille finie $\{P_{X_n}, n < n_0\}$ est équitendue, cf. remarque 1.37, on peut donc trouver un réel $c' > 0$ tel que

$$\forall n < n_0, \quad P(|Z_n| \leq c') > 1 - 3\varepsilon. \quad (1.53)$$

Finalement en posant $c'' := \max(c, c')$, nous déduisons de (1.52) et (1.53) que

$$\forall n \in \mathbb{N}^*, \quad P(|Z_n| \leq c'') > 1 - 3\varepsilon.$$

Comme ε était arbitraire, ceci établit l'équitension de $(X_n)_{n \geq 1}$ et termine la preuve. \square

Corollaire 1.41. Soient Z_n ($n \geq 1$) et Z des variables aléatoires réelles telles que pour toute fonction $g : \mathbb{R} \rightarrow \mathbb{R}$ continue tendant vers 0 à l'infini, $\mathbf{E}g(Z_n)$ converge vers $\mathbf{E}g(Z)$ quand n tend vers l'infini. Alors Z_n converge en loi vers Z .

Preuve. L'espace $\mathcal{C}_c(\mathbb{R})$ des fonctions continues à support compact est inclus dans l'espace $\mathcal{C}_0(\mathbb{R})$ des fonctions continues tendant vers 0 à l'infini. \square

Corollaire 1.42. Si $\mathbf{E}h(Z_n)$ converge quand n tend vers $+\infty$ vers $\mathbf{E}h(Z)$ pour toute $h \in \mathcal{H}$, où \mathcal{H} est une famille dense dans $\mathcal{C}_0(\mathbb{R})$ pour la norme $\|\cdot\|_\infty$, alors Z_n converge en loi vers Z .

Preuve. On utilise le corollaire 1.41 en montrant que pour toute $g \in \mathcal{C}_0(\mathbb{R})$, $\mathbf{E}g(Z_n)$ converge vers $\mathbf{E}g(Z)$. Soit donc g quelconque dans $\mathcal{C}_0(\mathbb{R})$. Par densité de \mathcal{H} dans $\mathcal{C}_0(\mathbb{R})$, il existe pour tout $\varepsilon > 0$, une fonction $h \in \mathcal{H}$ telle que $\|h - g\|_\infty < \varepsilon$. Cette inégalité s'écrit encore $\sup_{x \in \mathbb{R}} |(g - h)(x)| < \varepsilon$, ce qui nous permet de voir que pour toute v.a. réelle Y ,

$$|\mathbf{E}(g - h)(Y)| \leq \mathbf{E}|(g - h)(Y)| \leq \varepsilon. \quad (1.54)$$

En écrivant maintenant $\mathbf{E}g(Z_n) = \mathbf{E}h(Z_n) + \mathbf{E}(g - h)(Z_n)$ et de même avec Z au lieu de Z_n , en appliquant (1.54) avec $Y = Z_n$, puis avec $Y = Z$, on obtient

$$\forall n \in \mathbb{N}^*, \quad |\mathbf{E}g(Z_n) - \mathbf{E}g(Z)| \leq |\mathbf{E}h(Z_n) - \mathbf{E}h(Z)| + 2\varepsilon.$$

Par hypothèse $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$, donc pour $n \geq n_0$, $|\mathbf{E}h(Z_n) - \mathbf{E}h(Z)| < \varepsilon$. Finalement

$$\forall n \geq n_0, \quad |\mathbf{E}g(Z_n) - \mathbf{E}g(Z)| < 3\varepsilon.$$

Comme ε était arbitraire, on a bien vérifié la convergence de $\mathbf{E}g(Z_n)$ vers $\mathbf{E}g(Z)$. La fonction $g \in \mathcal{C}_0(\mathbb{R})$ était elle aussi arbitraire, on conclut donc par le corollaire 1.41 que Z_n converge en loi vers Z . \square

Corollaire 1.43. *Si $\mathbf{E}f(Z_n)$ converge quand n tend vers $+\infty$ vers $\mathbf{E}f(Z)$ pour toute $f \in \mathcal{H}'$, où \mathcal{H}' est une famille totale dans $\mathcal{C}_0(\mathbb{R})$ pour la norme $\|\cdot\|_\infty$, alors Z_n converge en loi vers Z .*

Preuve. Dire que \mathcal{H}' est totale dans $\mathcal{C}_0(\mathbb{R})$ signifie que la famille \mathcal{H} des combinaisons linéaires de fonctions de \mathcal{H}' est dense dans $\mathcal{C}_0(\mathbb{R})$. Il suffit donc d'après le corollaire 1.42 de vérifier que $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$ pour toute h de la forme

$$h = \sum_{i \in I} a_i f_i, \quad I \text{ fini, les } a_i \in \mathbb{R}, \text{ les } f_i \in \mathcal{H}'.$$

Comme \mathcal{H}' est incluse dans $\mathcal{C}_0(\mathbb{R})$, les f_i sont bornées donc les v.a. $f_i(Z_n)$ et $f_i(Z)$ sont intégrables. Par linéarité de l'espérance on a donc

$$\mathbf{E}h(Z_n) = \sum_{i \in I} a_i \mathbf{E}f_i(Z_n), \quad \mathbf{E}h(Z) = \sum_{i \in I} a_i \mathbf{E}f_i(Z).$$

Par hypothèse $\mathbf{E}f_i(Z_n)$ converge vers $\mathbf{E}f_i(Z)$ pour chaque $i \in I$. Comme I est fini, on en déduit que $\mathbf{E}h(Z_n)$ converge vers $\mathbf{E}h(Z)$. \square

Corollaire 1.44. *Si $\mathbf{E}f(Z_n)$ converge quand n tend vers $+\infty$ vers $\mathbf{E}f(Z)$ pour toute $f \in C^\infty$ tendant vers 0 à l'infini, Z_n converge en loi vers Z .*

Preuve. L'espace $\mathcal{C}_0^\infty(\mathbb{R})$ des fonctions indéfiniment dérivables sur \mathbb{R} et tendant vers 0 à l'infini est dense dans $\mathcal{C}_0(\mathbb{R})$, cf. cours d'analyse²⁴. On peut donc appliquer le corollaire 1.42. \square

24. On peut le démontrer soit en faisant de la régularisation par convolution, soit en utilisant le théorème de Stone-Weierstrass après compactification de \mathbb{R} par l'ajout d'un seul point à l'infini.

1.4.2 Démonstration du TLC

Dans toutes les versions du TLC dans \mathbb{R} exposées ci-dessus, on a une somme de variables aléatoires indépendantes que l'on centre et divise par l'écart type de la somme et c'est cette somme normalisée qui converge en loi vers une gaussienne $\mathfrak{N}(0, 1)$. D'après le corollaire 1.44, il suffit de montrer la convergence des f -moments pour toute $f \in \mathcal{C}_0^\infty(\mathbb{R})$. Nous allons étudier l'erreur commise en remplaçant ces f -moments d'une somme par les f -moments de gaussiennes ayant même espérance et même variance que la somme²⁵. Dans cette approche le lemme suivant nous sera très utile.

Lemme 1.45. *Soit $f \in \mathcal{C}_0^3(\mathbb{R})$ et Y une v.a. réelle. Soient U, V des variables aléatoires réelles indépendantes de Y telles que $\mathbf{E}|U|^3 < +\infty$, $\mathbf{E}|V|^3 < +\infty$, $\mathbf{E}U = \mathbf{E}V$ et $\mathbf{E}U^2 = \mathbf{E}V^2$. Alors*

$$|\mathbf{E}f(Y + U) - \mathbf{E}f(Y + V)| \leq c(\mathbf{E}|U|^3 + \mathbf{E}|V|^3), \quad (1.55)$$

où c est une constante ne dépendant que de f .

Preuve. La formule de Taylor au point y pour un accroissement t s'écrit

$$f(y + t) = f(y) + tf'(y) + \frac{t^2}{2}f''(y) + r(y, t), \quad (1.56)$$

où le reste $r(y, t)$ est de la forme

$$r(y, t) = \frac{t^3}{6}f^{(3)}(y + \theta t), \quad \theta = \theta(f, y, t) \in [0, 1].$$

On a donc la majoration

$$|r(y, t)| \leq \frac{1}{6}\|f^{(3)}\|_\infty|t|^3. \quad (1.57)$$

En prenant $y = Y(\omega)$, $t = U(\omega)$ pour ω quelconque dans Ω , ces égalités et cette majoration passent aux variables aléatoires :

$$f(Y + U) = f(Y) + Uf'(Y) + \frac{U^2}{2}f''(Y) + r(Y, U), \quad (1.58)$$

$$|r(Y, U)| \leq \frac{1}{6}\|f^{(3)}\|_\infty|U|^3. \quad (1.59)$$

Les variables aléatoires $f(Y + U)$, $f(Y)$, $f'(Y)$, $f''(Y)$ sont bornées et U est de cube intégrable, de plus U et Y sont indépendantes. On a donc $\mathbf{E}|Uf'(Y)| = \mathbf{E}|U|\mathbf{E}|f'(Y)| < +\infty$, $\mathbf{E}|U^2f''(Y)| = \mathbf{E}U^2\mathbf{E}|f''(Y)| < +\infty$. On peut donc prendre l'espérance des deux membres dans (1.58) et dans (1.59) pour obtenir

$$\mathbf{E}f(Y + U) = \mathbf{E}f(Y) + \mathbf{E}U\mathbf{E}f'(Y) + \frac{1}{2}\mathbf{E}U^2\mathbf{E}f''(Y) + \mathbf{E}r(Y, U), \quad (1.60)$$

25. La technique de preuve du TLC exposée dans cette section n'est pas la plus répandue dans la littérature où on utilise souvent les fonctions caractéristiques. Nous suivons d'assez près l'exposé donné dans le livre de D. Pollard, *Convergence of stochastic processes*, Springer 1984.

avec

$$\mathbf{E}|r(Y, U)| \leq \frac{1}{6} \|f^{(3)}\|_{\infty} \mathbf{E}|U|^3. \quad (1.61)$$

Évidemment (1.60) et (1.61) restent vérifiées avec V à la place de U . On en déduit que

$$\mathbf{E}f(Y+U) - \mathbf{E}f(Y+V) = (\mathbf{E}U - \mathbf{E}V)\mathbf{E}f'(Y) + \frac{1}{2}(\mathbf{E}U^2 - \mathbf{E}V^2)\mathbf{E}f''(Y) + \mathbf{E}r(Y, U) - \mathbf{E}r(Y, V),$$

d'où compte tenu des égalités de moments $\mathbf{E}U = \mathbf{E}V$ et $\mathbf{E}U^2 = \mathbf{E}V^2$,

$$|\mathbf{E}f(Y+U) - \mathbf{E}f(Y+V)| = |\mathbf{E}r(Y, U) - \mathbf{E}r(Y, V)| \leq \frac{1}{6} \|f^{(3)}\|_{\infty} (\mathbf{E}|U|^3 + \mathbf{E}|V|^3),$$

ce qui nous donne (1.55) avec $c = \frac{1}{6} \|f^{(3)}\|_{\infty}$. \square

Le lemme 1.45 nous permet de majorer l'erreur commise en approximant le f -moment d'une somme centrée normalisée de v.a. indépendantes par le f -moment de la loi gaussienne $\mathfrak{N}(0, 1)$.

Lemme 1.46. Soient X_1, \dots, X_n des v.a. indépendantes vérifiant $\mathbf{E}|X_i|^3 < +\infty$, $\mathbf{E}X_i = 0$, $\sum_{i=1}^n \sigma_i^2 = 1$ avec $\sigma_i^2 := \mathbf{E}X_i^2$. On note $S_n^* := \sum_{i=1}^n X_i$. Soient Z_1, \dots, Z_n des v.a. telles que Z_i ait pour loi $\mathfrak{N}(0, \sigma_i)$ et que la suite $X_1, \dots, X_n, Z_1, \dots, Z_n$ soit indépendante. Soit Z de loi $\mathfrak{N}(0, 1)$. Alors pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$,

$$|\mathbf{E}f(S_n^*) - \mathbf{E}f(Z)| \leq c \sum_{i=1}^n (\mathbf{E}|X_i|^3 + \mathbf{E}|Z_i|^3), \quad (1.62)$$

avec $c = \frac{1}{6} \|f^{(3)}\|_{\infty}$.

Preuve. Notons d'abord que Z a même loi que $W_n := Z_1 + \dots + Z_n$. En effet par indépendance des Z_i , le vecteur (Z_1, \dots, Z_n) est gaussien, cf. corollaire 1.27. Donc la combinaison linéaire $W_n = Z_1 + \dots + Z_n$ est une v.a. gaussienne. Sa loi est déterminée par son espérance qui vaut 0 puisque les Z_i sont centrées et sa variance $\sigma^2 = \text{Var } W_n = \sum_{i=1}^n \sigma_i^2 = 1$ à cause de l'indépendance des Z_i . Donc W_n suit comme Z , la loi $\mathfrak{N}(0, 1)$. Par conséquent, pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$, $\mathbf{E}f(Z) = \mathbf{E}f(W_n)$.

L'idée de la preuve est de faire du chaînage pour passer progressivement de $\mathbf{E}f(S_n^*)$ à $\mathbf{E}f(W_n)$ en remplaçant à chaque pas un X_i par un Z_i . Le lemme 1.46 permet de majorer l'erreur d'approximation commise à chaque pas. Définissons les sommes « trouées » T_j (T_j trouée au rang j) en posant :

$$T_j := X_1 + \dots + X_{j-1} + Z_{j+1} + \dots + Z_n, \quad 1 \leq j \leq n.$$

Pour bien comprendre le mécanisme de chaînage que nous allons utiliser, voyons le à l'oeuvre dans le cas $n = 3$. On a alors $T_1 = Z_2 + Z_3$, $T_2 = X_1 + Z_3$ et $T_3 = X_1 + X_2$. On part de la somme télescopique

$$\begin{aligned} f(X_1 + X_2 + X_3) - f(Z_1 + Z_2 + Z_3) &= f(X_1 + X_2 + X_3) - f(X_1 + X_2 + Z_3) \\ &\quad + f(X_1 + X_2 + Z_3) - f(X_1 + Z_2 + Z_3) \\ &\quad + f(X_1 + Z_2 + Z_3) - f(Z_1 + Z_2 + Z_3), \end{aligned}$$

que l'on réécrit sous la forme

$$\begin{aligned} f(X_1 + X_2 + X_3) - f(Z_1 + Z_2 + Z_3) &= f(T_3 + X_3) - f(T_3 + Z_3) \\ &\quad + f(T_2 + X_2) - f(T_2 + Z_2) \\ &\quad + f(T_1 + X_1) - f(T_1 + Z_1), \end{aligned}$$

en utilisant les égalités $T_2 + X_2 = T_3 + Z_3$ et $T_1 + X_1 = T_2 + Z_2$.

Pour le cas général, il est commode de poser $T_0 := W_n$ et $X_0 := 0$. On note alors que

$$\forall j = 1, \dots, n, \quad T_j + Z_j = T_{j-1} + X_{j-1}. \quad (1.63)$$

On peut alors écrire la somme télescopique

$$f(S_n^*) - f(W_n) = f(T_n + X_n) - f(T_0 + X_0) = \sum_{j=1}^n (f(T_j + X_j) - f(T_{j-1} + X_{j-1}))$$

En prenant l'espérance et en utilisant (1.63), il vient

$$\mathbf{E}f(S_n^*) - \mathbf{E}f(W_n) = \sum_{j=1}^n (\mathbf{E}f(T_j + X_j) - \mathbf{E}f(T_j + Z_j)).$$

Pour tout j , X_j et Z_j sont indépendantes de T_j , sont de cube intégrable et ont mêmes moments d'ordre 1 et 2. On peut donc appliquer le lemme 1.45 avec $Y = T_j$, $U = X_j$ et $V = Z_j$ à chacun des termes de la somme ci-dessus, ce qui donne

$$|\mathbf{E}f(S_n^*) - \mathbf{E}f(W_n)| \leq \sum_{j=1}^n |\mathbf{E}f(T_j + X_j) - \mathbf{E}f(T_j + Z_j)| \leq c \sum_{j=1}^n (\mathbf{E}|X_j|^3 + \mathbf{E}|Z_j|^3),$$

avec $c = \frac{1}{6} \|f^{(3)}\|_\infty$. □

Le lemme 1.46 permet de démontrer une version du théorème de Liapounov pour des v.a. de cube intégrable et d'explicitier dans le cas i.i.d. une vitesse de convergence en $O(n^{-1/2})$ pour les f -moments.

Théorème 1.47 (Liapounov). *Soit $(X_k)_{k \geq 1}$ une suite de variables aléatoires définies sur le même espace probabilisé, indépendantes (mais pas forcément de même loi), centrées, ayant toutes un moment d'ordre 3 ($\mathbf{E}|X_k|^3 < +\infty$). On note $S_n := X_1 + \dots + X_n$, $\sigma_k^2 := \text{Var } X_k$, $s_n^2 := \text{Var}(S_n) = \sigma_1^2 + \dots + \sigma_n^2$ et on suppose $s_n > 0$. Si de plus la condition de Liapounov :*

$$\frac{1}{s_n^3} \sum_{k=1}^n \mathbf{E}|X_k|^3 \xrightarrow{n \rightarrow +\infty} 0 \quad (1.64)$$

est vérifiée, alors S_n/s_n converge en loi vers Z de loi $\mathfrak{N}(0, 1)$.

Dans le cas où les X_k sont de plus i.i.d., on a pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$,

$$|\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leq Cn^{-1/2}, \quad (1.65)$$

où la constante C vaut

$$C = \frac{1}{6} \left(\frac{\mathbf{E}|X_1|^3}{\sigma_1^3} + \frac{4}{\sqrt{2\pi}} \right) \|f^{(3)}\|_\infty. \quad (1.66)$$

Preuve. D'après le corollaire 1.44 et l'inclusion $\mathcal{C}_0^\infty(\mathbb{R}) \subset \mathcal{C}_0^3(\mathbb{R})$, il suffit de montrer la convergence des f -moments de S_n/s_n vers ceux de Z pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$ pour obtenir la convergence en loi de S_n/s_n vers Z . Posons $X_{n,k} := s_n^{-1}X_k$, $\sigma_{n,k}^2 := \text{Var } X_{n,k} = s_n^{-2}\mathbf{E}X_k^2$. En appliquant le lemme 1.46 aux $X_{n,k}$, notons qu'alors $S_n^* = S_n/s_n$, on obtient

$$\forall f \in \mathcal{C}_0^3(\mathbb{R}), \quad |\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leq c \sum_{k=1}^n \left(\frac{\mathbf{E}|X_k|^3}{s_n^3} + \left(\frac{\sigma_k}{s_n} \right)^3 \mathbf{E}|Z|^3 \right). \quad (1.67)$$

En effet la gaussienne centrée $Z_{n,k}$ de variance $\sigma_{n,k}^2$ a même loi que $\sigma_{n,k}Z$, donc même moment absolu d'ordre 3, d'où $\mathbf{E}|Z_{n,k}|^3 = \sigma_{n,k}^3 \mathbf{E}|Z|^3 = \sigma_k^3 s_n^{-3} \mathbf{E}|Z|^3$. Compte-tenu de la condition de Liapounov (1.64) et de (1.67), il ne nous reste plus pour obtenir la convergence en loi qu'à vérifier la convergence :

$$\frac{1}{s_n^3} \sum_{k=1}^n \sigma_k^3 \xrightarrow{n \rightarrow +\infty} 0. \quad (1.68)$$

Pour pouvoir exploiter la condition de Liapounov, on souhaite disposer d'une majoration du type $\sigma_k^3 \leq a \mathbf{E}|X_k|^3$, avec a constante, qui ferait immédiatement découler (1.68) de (1.64). L'inégalité de Jensen, non étudiée en cours d'I.P.É., nous dit que si φ est une fonction convexe sur un intervalle I de \mathbb{R} et Y une v.a. intégrable telle que $P(Y \in I) = 1$ et $\mathbf{E}|\varphi(Y)| < +\infty$, $\varphi(\mathbf{E}Y) \leq \mathbf{E}\varphi(Y)$. En appliquant ceci à la v.a. $Y = X^2$ et avec la fonction convexe $\varphi : [0, +\infty[\rightarrow \mathbb{R}_+$, $y \mapsto y^{3/2}$, on obtient pour toute v.a. de cube intégrable :

$$(\mathbf{E}(X^2))^{3/2} \leq \mathbf{E}((X^2)^{3/2}) = \mathbf{E}(|X|^3).$$

On en déduit pour chaque X_k la majoration annoncée avec $a = 1$ puisque $\sigma_k = (\mathbf{E}(X_k^2))^{1/2}$.

Voici une variante « manuelle » évitant le recours à l'inégalité de Jensen au prix d'une moins bonne constante a . On commence par vérifier que pour tout $t > 0$, on a

$$\mathbf{E}X^2 = \mathbf{E}(X^2 \mathbf{1}_{\{|X| \leq t\}}) + \mathbf{E}(X^2 \mathbf{1}_{\{|X| > t\}}) \leq t^2 + \frac{\mathbf{E}|X|^3}{t}. \quad (1.69)$$

En effet $\mathbf{E}(X^2 \mathbf{1}_{\{|X| \leq t\}})$ se majore naturellement par $t^2 P(|X| \leq t) \leq t^2$. Pour le deuxième terme, on remarque que sur l'évènement $\{|X| > t\}$, on a $|X|^3 > t|X|^2$, d'où $|X|^2 < t^{-1}|X|^3$. Cette inégalité vraie seulement sur $\{|X| > t\}$ implique l'inégalité entre variables aléatoires, vraie sur tout Ω :

$$X^2 \mathbf{1}_{\{|X| > t\}} \leq \frac{1}{t} |X|^3 \mathbf{1}_{\{|X| > t\}} \leq \frac{1}{t} |X|^3,$$

puis en prenant les espérances, $\mathbf{E}(X^2 \mathbf{1}_{\{|X| > t\}}) \leq t^{-1} \mathbf{E}|X|^3$. Puisque (1.69) est ainsi vérifiée pour tout $t > 0$, on a intérêt à choisir la valeur de t qui minimise le majorant

obtenu. Un simple calcul de dérivée nous montre que la fonction $g : t \mapsto t^2 + \frac{\mathbf{E}|X|^3}{t}$ atteint son minimum sur $]0, +\infty[$ au point $t = 2^{-1/3}(\mathbf{E}|X|^3)^{1/3}$ et donc que la valeur minimale de g est $3 \times 2^{-2/3}(\mathbf{E}|X|^3)^{2/3}$. On en déduit l'inégalité $\sigma_k^3 \leq 3^{3/2} \times 2^{-1} \mathbf{E}|X|^3$, ce qui est moins bon que l'inégalité de Jensen puisque ici $a = 3^{3/2} \times 2^{-1} \simeq 2,599$.

Enfin dans le cas i.i.d., on a $s_n = n^{1/2}\sigma_1$. En revenant à (1.67) et en notant que $\mathbf{E}|Z|^3 = 4(2\pi)^{-1/2}$ (vérifiez!), on obtient :

$$|\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leq \frac{c}{\sigma_1^3 n^{3/2}} (n\mathbf{E}|X_1|^3 + n\sigma_1^3 \mathbf{E}|Z|^2) = \frac{c}{n^{1/2}} \left(\frac{\mathbf{E}|X_1|^3}{\sigma_1^3} + \frac{4}{\sqrt{2\pi}} \right),$$

d'où (1.65) et (1.66). \square

Jusqu'ici nous n'avons prouvé le TLC qu'en supposant l'existence de moments d'ordre 3. Nous allons maintenant nous affranchir de cette hypothèse en revenant à la formule de Taylor (1.56) et en notant que le reste $r(y, t)$ s'écrit aussi :

$$r(y, t) = (f(y+t) - f(y) - tf'(y)) - \frac{t^2}{2}f''(y) = \frac{t^2}{2}(f''(y+\tau t) - f''(y)),$$

pour un certain $\tau = \tau(f, y, t) \in [0, 1]$. Par conséquent,

$$\forall y, t \in \mathbb{R}, \quad |r(y, t)| \leq t^2 \|f''\|_\infty. \quad (1.70)$$

Cette nouvelle majoration de $|r(y, t)|$ est meilleure que (1.57) lorsque $|t|$ tend vers l'infini, par contre elle est moins bonne lorsque t tend vers 0. Pour garder le meilleur des deux majorations, on fait un compromis en utilisant (1.57) pour $|t| < \varepsilon$ et (1.70) pour $|t| \geq \varepsilon$, ce qui donne pour tout $\varepsilon > 0$,

$$\forall y, t \in \mathbb{R}, \quad |r(y, t)| \leq C|t|^3 \mathbf{1}_{\{|t| < \varepsilon\}} + Ct^2 \mathbf{1}_{\{|t| \geq \varepsilon\}} \leq C\varepsilon t^2 \mathbf{1}_{\{|t| < \varepsilon\}} + Ct^2 \mathbf{1}_{\{|t| \geq \varepsilon\}}, \quad (1.71)$$

avec

$$C = \max \left(\|f''\|_\infty, \frac{1}{6} \|f^{(3)}\|_\infty \right). \quad (1.72)$$

Par conséquent, pour toute v.a. Y et toute v.a. U de carré intégrable,

$$\mathbf{E}|r(Y, U)| \leq C\varepsilon \mathbf{E}U^2 \mathbf{1}_{\{|U| < \varepsilon\}} + C\mathbf{E}(U^2 \mathbf{1}_{\{|U| \geq \varepsilon\}}) \leq C\varepsilon \mathbf{E}U^2 + C\mathbf{E}(U^2 \mathbf{1}_{\{|U| \geq \varepsilon\}}).$$

Reprenant la preuve des lemmes (1.45) et (1.46) avec cette nouvelle majoration de $\mathbf{E}|r(Y, U)|$, on obtient la clé de la preuve du théorème de Lindeberg.

Lemme 1.48. Soient X_1, \dots, X_n des v.a. indépendantes vérifiant $\mathbf{E}X_i^2 < +\infty$, $\mathbf{E}X_i = 0$, $\sum_{i=1}^n \sigma_i^2 = 1$ avec $\sigma_i^2 := \mathbf{E}X_i^2$. On note $S_n^* := \sum_{i=1}^n X_i$. Soient Z_1, \dots, Z_n des v.a. telles que Z_i ait pour loi $\mathfrak{N}(0, \sigma_i)$ et que la suite $X_1, \dots, X_n, Z_1, \dots, Z_n$ soit indépendante. Soit Z de loi $\mathfrak{N}(0, 1)$. Alors pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$ et tout $\varepsilon > 0$,

$$|\mathbf{E}f(S_n^*) - \mathbf{E}f(Z)| \leq C \sum_{i=1}^n \left(\varepsilon \sigma_i^2 + \mathbf{E}(X_i^2 \mathbf{1}_{\{|X_i| \geq \varepsilon\}}) + \mathbf{E}|Z_i|^3 \right), \quad (1.73)$$

avec C donnée par (1.72).

Notez le traitement dissymétrique des X_i et des Z_i , dû au fait que les Z_i étant gaussiennes ont des moments d'ordre 3. Il est alors plus commode de leur appliquer (1.57) pour contrôler $|r(Y, Z_i)|$, tandis que pour les X_i , on tient à n'utiliser que des majorations valables même si ces variables sont seulement de carré intégrable.

Preuve du théorème de Lindeberg. Avec les notations et les hypothèses du théorème 1.16, il suffit de montrer la convergence de $\mathbf{E}f(S_n/s_n)$ vers $\mathbf{E}f(Z)$ pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$. En appliquant le lemme (1.48) aux k_n variables $X_i = X_{n,i}/s_n$, on dispose de la majoration suivante valable pour tout $\varepsilon > 0$:

$$\begin{aligned} |\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| &\leq C \sum_{i=1}^{k_n} \left(\varepsilon \frac{\sigma_{n,i}^2}{s_n^2} + \frac{1}{s_n^2} \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}) + \frac{1}{s_n^3} \mathbf{E}|Z_{n,i}|^3 \right) \\ &= C\varepsilon + \frac{C}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| > \varepsilon s_n\}}) + \frac{1}{s_n^3} \sum_{i=1}^{k_n} \mathbf{E}|Z_{n,i}|^3. \end{aligned} \quad (1.74)$$

Dans ce majorant, le deuxième terme tend vers 0 quand n tend vers l'infini par la condition de Lindeberg, cf. hypothèse (1.22). Il peut donc être rendu inférieur à ε pour tout $n \geq n_1$. Pour le troisième terme, on rappelle que la v.a. gaussienne $Z_{n,i}$ a même loi que $\sigma_{n,i}Z$, d'où

$$\frac{1}{s_n^3} \sum_{i=1}^{k_n} \mathbf{E}|Z_{n,i}|^3 = \frac{\mathbf{E}|Z|^3}{s_n^3} \sum_{i=1}^{k_n} \sigma_{n,i}^3 \leq \frac{\mathbf{E}|Z|^3}{s_n^3} \left(\max_{1 \leq i \leq k_n} \sigma_{n,i} \right) \sum_{i=1}^{k_n} \sigma_{n,i}^2 = \frac{4}{\sqrt{2\pi}} \frac{1}{s_n} \max_{1 \leq i \leq k_n} \sigma_{n,i}.$$

Il nous reste à montrer la convergence vers 0 de $s_n^{-1} \max_{1 \leq i \leq k_n} \sigma_{n,i}$, ce qui équivaut à celle de $s_n^{-2} \max_{1 \leq i \leq k_n} \sigma_{n,i}^2$. On se rappelle opportunément que $\sigma_{n,i}^2 = \mathbf{E}X_{n,i}^2$, ce qui va nous permettre de réexploiter la condition de Lindeberg. En effet

$$\mathbf{E}X_{n,i}^2 = \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| < s_n \varepsilon\}}) + \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| \geq s_n \varepsilon\}}) \leq \varepsilon^2 s_n^2 + \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| \geq s_n \varepsilon\}}).$$

D'où

$$\begin{aligned} \frac{1}{s_n^2} \max_{1 \leq i \leq k_n} \sigma_{n,i}^2 &\leq \varepsilon^2 + \frac{1}{s_n^2} \max_{1 \leq i \leq k_n} \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| \geq s_n \varepsilon\}}) \\ &\leq \varepsilon^2 + \frac{1}{s_n^2} \sum_{i=1}^{k_n} \mathbf{E}(X_{n,i}^2 \mathbf{1}_{\{|X_{n,i}| \geq s_n \varepsilon\}}). \end{aligned}$$

Le deuxième terme de ce majorant tend vers 0 quand n tend vers $+\infty$ par la condition de Lindeberg. Il est donc inférieur à ε^2 pour tout $n \geq n_2$. Par conséquent,

$$\forall n \geq n_2, \quad \frac{1}{s_n} \max_{1 \leq i \leq k_n} \sigma_{n,i} \leq \varepsilon \sqrt{2}.$$

En prenant $n_0 := \max(n_1, n_2)$ et en revenant à (1.74), on obtient finalement

$$\forall n \geq n_0, \quad |\mathbf{E}f(S_n/s_n) - \mathbf{E}f(Z)| \leq \left(C + 1 + \frac{4}{\sqrt{\pi}} \right) \varepsilon.$$

Comme $\varepsilon > 0$ était arbitraire, on a bien établi la convergence de $\mathbf{E}f(S_n/s_n)$ vers $\mathbf{E}f(Z)$. Ceci étant vérifié pour toute $f \in \mathcal{C}_0^3(\mathbb{R})$, S_n/s_n converge en loi vers Z . \square

Chapitre 2

Simulation de variables et vecteurs aléatoires

2.1 Introduction

La simulation informatique du hasard a de multiples applications : simulation de phénomènes physiques, méthodes de Monte-Carlo pour le calcul d'intégrales, étude de tests statistiques ou d'estimateurs, simulation de fonctionnements de réseaux ou de systèmes complexes, cryptographie, imagerie, algorithmes probabilistes,...

Théoriquement, la génération de nombres aléatoires suivant une loi donnée se ramène à la génération de suites de variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. On peut montrer en effet que si les X_i sont des variables de Bernoulli indépendantes de même paramètre $p = 1/2$, la v.a. $U := \sum_{k=1}^{+\infty} X_k 2^{-k}$ suit la loi uniforme sur $[0, 1]$. Le problème se ramène donc à la génération d'une suite de « bits » aléatoires indépendants pouvant prendre chacun la valeur 0 ou la valeur 1 avec même probabilité $1/2$. En d'autres termes, il suffirait de réaliser un jeu de pile ou face infini avec une pièce parfaitement équilibrée¹. Cette méthode n'est évidemment pas réaliste et en pratique on a recours à l'informatique pour « simuler » une telle suite. Pourquoi employer ici le mot « simuler » ? Parce qu'une suite de nombres générée par un algorithme n'est pas vraiment aléatoire. Si on connaît les valeurs d'initialisation et l'algorithme, on peut calculer (et donc prévoir) les termes de la suite. Néanmoins on considèrera que l'on a un bon générateur de nombres aléatoires si on ne parvient pas à distinguer la suite de nombres pseudo aléatoires produite d'une suite véritablement aléatoire. La signification précise de cette phrase demanderait tout un développement amenant à s'interroger sur la notion même de hasard. On pourra utilement consulter à ce sujet [2]. Pour l'utilisation en statistique, nous nous contenterons de dire qu'un générateur est acceptable s'il passe avec succès une batterie de tests statistiques courants.

Les fonctions `random` des principaux langages de programmation ou logiciels sont bâties sur des algorithmes arithmétiques dont le plus simple correspond au générateur

1. Si $p \neq 1/2$, la loi de U est une loi singulière à fonction de répartition continue mais n'ayant pas de densité par rapport à la mesure de Lebesgue.

congruentiel linéaire. Il s'agit de générer une suite de nombres $(X_n)_{n \geq 1}$ vérifiant une relation de récurrence

$$X_{n+1} = aX_n + c \pmod{M} \quad (2.1)$$

et d'en déduire une suite $(U_n)_{n \geq 1}$ à valeurs dans $[0, 1[$ en prenant $U_n = X_n/M$. Par exemple la fonction `rand` de Scilab utilise (2.1) avec $M = 2^{31}$, $a = 843\,314\,861$ et $c = 453\,816\,693$. La suite (U_n) ainsi construite est complètement déterministe, car périodique. Cependant sa période est tellement grande qu'on peut en pratique la considérer comme une suite aléatoire, du moins pour des besoins statistiques courants (son usage est déconseillé en cryptographie). Remarquons d'ailleurs que même si U_n était ici aléatoire, ses valeurs seraient de la forme $k2^{-31}$ et on obtiendrait la loi uniforme discrète sur $D_{31} = \{k2^{-31}; 0 \leq k < 2^{31}\}$ au lieu de la loi uniforme sur $[0, 1[$. Ceci n'est pas trop gênant pour les deux raisons suivantes. D'une part la loi uniforme μ_n sur $D_n = \{k2^{-n}; 0 \leq k < 2^n\}$ converge *étroitement* vers la loi uniforme sur $[0, 1[$ quand n tend vers l'infini² et d'autre part les nombres réels sont représentés en machine par des rationnels dyadiques de la forme $k2^{-j}$, de sorte que tous les réels de $[k2^{-j}, (k+1)2^{-j}[$ sont confondus.

Dans ce document, nous nous situons en aval du problème de la construction d'un générateur de nombres aléatoires. On suppose que l'on sait générer une suite i.i.d. (U_n) de variables aléatoires de loi uniforme sur $[0, 1]$. On se propose de construire et de justifier mathématiquement des algorithmes permettant à partir de là de simuler une variable aléatoire ou un vecteur aléatoire de loi donnée. On donnera une traduction de certains de ces algorithmes en Scilab à titre d'illustration. Scilab est un logiciel libre développé par l'INRIA. Il existe en version Linux et Windows. On peut le télécharger depuis le site de l'INRIA à l'URL :

<http://scilabsoft.inria.fr/>

2.2 Méthode théorique pour simuler une v.a.r.

Soit X une variable aléatoire réelle de fonction de répartition F , définie par

$$F(x) := P(X \leq x). \quad (2.2)$$

Rappelons que F est croissante, continue à droite et limitée à gauche en tout point de \mathbb{R} , que l'ensemble de ses discontinuités est au plus dénombrable et que F a pour limite 0 en $-\infty$ et 1 en $+\infty$. Dans le cas particulier où F est continue et strictement croissante sur tout \mathbb{R} , elle réalise une bijection de \mathbb{R} sur $]0, 1[$ et admet donc un inverse $F^{-1} :]0, 1[\rightarrow \mathbb{R}$ au sens classique. Si U est une variable aléatoire de loi uniforme sur $]0, 1[$, alors $Y := F^{-1}(U)$ a même loi que X . On le vérifie facilement en calculant la fonction de répartition de Y :

$$\forall x \in \mathbb{R}, \quad P(Y \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x). \quad (2.3)$$

2. Ce qui signifie que U_n converge en loi vers U de loi uniforme sur $[0, 1]$, cf. la note 2 page 8. L'erreur commise sur la f.d.r. de μ_n en la remplaçant par la f.d.r. de la loi uniforme sur $[0, 1]$ est majorée par 2^{-n} et $2^{-31} \simeq 4,7 \times 10^{-10}$ est suffisamment petit pour un usage courant de cette approximation.

Dans cette suite d'égalités, la deuxième repose sur la croissance de F et de F^{-1} qui légitiment l'équivalence $F^{-1}(U) \leq x \Leftrightarrow U \leq F(x)$. La dernière égalité est due au calcul de la fonction de répartition de la loi uniforme sur $]0, 1[$ (qui coïncide sur $[0, 1]$ avec l'identité) et au fait que $F(x) \in [0, 1]$.

Cette propriété permet donc, à partir d'un générateur aléatoire fournissant des réalisations d'une v.a. uniforme, de simuler une v.a. de même loi que X , pourvu que F^{-1} existe et soit calculable simplement. C'est le cas par exemple pour la loi de Cauchy, cf. exemple 2.4. Mais il arrive souvent même pour des lois usuelles que la f.d.r. ne soit pas strictement croissante sur tout \mathbb{R} , ou ne soit pas continue en tout point de \mathbb{R} . Dans un but de généralisation, ceci nous conduit à poser la définition suivante.

Définition 2.1 (fonction quantile). *Si F est la fonction de répartition d'une variable aléatoire, son inverse généralisée, appelée aussi fonction quantile, est définie par*

$$F^{-1} :]0, 1[\rightarrow \mathbb{R}, \quad \forall u \in]0, 1[, \quad F^{-1}(u) := \inf\{x \in \mathbb{R}; F(x) \geq u\}. \quad (2.4)$$

L'équation $F(x) = u$ peut avoir soit une solution unique, soit aucune solution, soit une infinité de solutions. La détermination graphique de $F^{-1}(u)$ dans chacune de ces configurations est illustrée figure 2.1. On vous laisse en exercice le soin de vérifier que pour F continue strictement croissante son inverse au sens classique (bijection inverse de $]0, 1[\rightarrow \mathbb{R}$) et son inverse généralisée coïncident.

Théorème 2.2. *Soient X une variable aléatoire réelle de fonction de répartition F et U une variable aléatoire de loi uniforme sur $]0, 1[$. Alors X et $F^{-1}(U)$ ont même loi.*

Preuve. Pour éviter d'avoir à éliminer un ensemble de probabilité nulle sur lequel $F^{-1}(U)$ ne serait pas défini, nous considérerons que U est non seulement de loi uniforme sur $]0, 1[$, mais que pour tout $\omega \in \Omega$, $U(\omega) \in]0, 1[$. La fonction de répartition de U est donnée par

$$P(U \leq t) = \begin{cases} 0 & \text{si } t < 0, \\ t & \text{si } 0 \leq t \leq 1, \\ 1 & \text{si } t > 1. \end{cases} \quad (2.5)$$

Or pour tout $x \in \mathbb{R}$, $F(x) \in [0, 1]$, d'où en prenant $t = F(x)$ dans (2.5),

$$\forall x \in \mathbb{R}, \quad P(U \leq F(x)) = F(x). \quad (2.6)$$

Admettons pour un instant que l'on sache trouver une fonction $g :]0, 1[\rightarrow \mathbb{R}$ vérifiant :

$$\forall u \in]0, 1[, \forall x \in \mathbb{R}, \quad (u \leq F(x)) \Leftrightarrow (g(u) \leq x). \quad (2.7)$$

Alors le problème de la simulation d'une variable aléatoire Y ayant même loi que X serait résolu. En effet en posant $Y := g(U)$, on aurait :

$$\begin{aligned} \forall x \in \mathbb{R}, \quad P(Y \leq x) &= P(g(U) \leq x) \\ &= P(U \leq F(x)) \quad \text{grâce à (2.7),} \\ &= F(x) \quad \text{par (2.6).} \end{aligned}$$

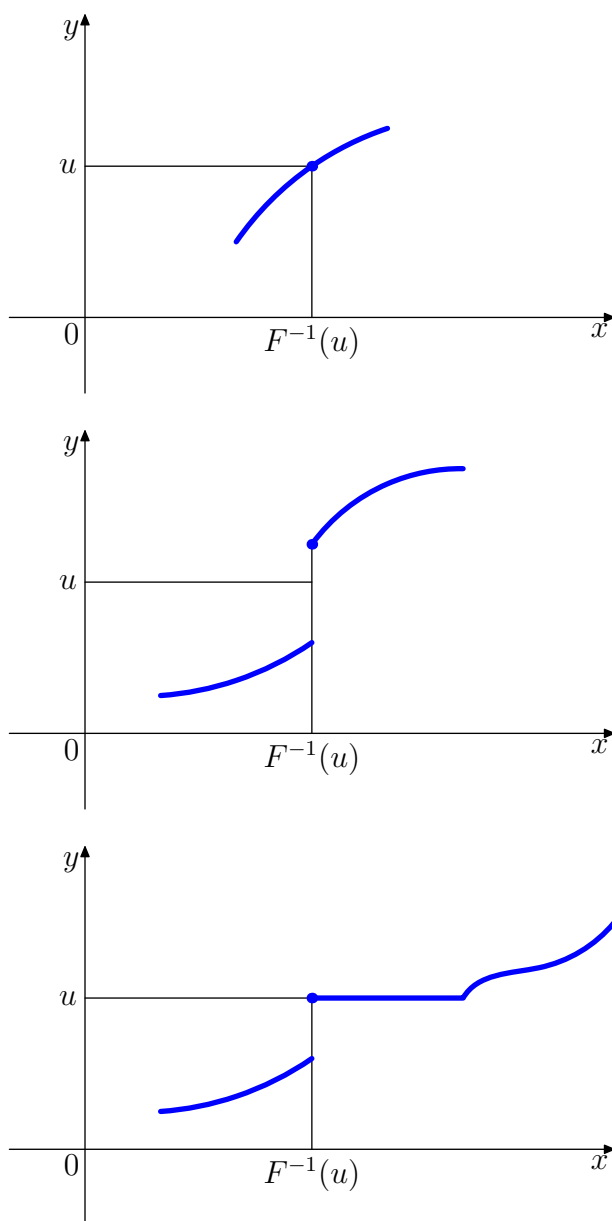


FIG. 2.1 – Détermination graphique de l'inverse généralisé

Ainsi Y aurait même f.d.r. F que X , donc même loi.

Recherche d'une fonction g vérifiant (2.7). Définissons les ensembles de réels I_u par :

$$\forall u \in]0, 1[, \quad I_u := \{x \in \mathbb{R}; u \leq F(x)\}.$$

Comme u est strictement inférieur à 1 et $F(t)$ tend vers 1 quand t tend vers $+\infty$, on voit que $F(t) > u$ pour t assez grand, donc I_u n'est pas vide. Si x est un élément quelconque de I_u , on a

$$\forall x' > x, \quad F(x') \geq F(x) \geq u, \quad \text{donc } x' \in I_u,$$

la première inégalité venant de la croissance de F et la deuxième de l'appartenance de x à I_u . Nous venons ainsi de vérifier que

$$\forall x \in I_u, \quad [x, +\infty[\subset I_u.$$

Ceci nous montre que I_u est un *intervalle* de la forme $[b, +\infty[$ ou $]b, +\infty[$. Notons que la borne inférieure b de cet intervalle I_u ne peut pas être $-\infty$. En effet si $b = -\infty$, alors $I_u = \mathbb{R}$. Or u est strictement positif et $F(t)$ tend vers 0 quand t tend vers $-\infty$, donc pour $t \leq t_0$, on a $F(t) < u$ d'où $t \notin I_u$, d'où $I_u \neq \mathbb{R}$.

Vérifions maintenant que I_u est toujours de la forme $[b, +\infty[$ avec $b \in \mathbb{R}$. On vient de voir que b est forcément un réel. Tout x strictement supérieur à la borne inférieure b de l'intervalle I_u est dans cet intervalle, donc $x \in I_u$ et $F(x) \geq u$. L'inégalité large

$$\forall x > b, \quad F(x) \geq u,$$

se conserve par passage à la limite à droite de F au point b . Or la f.d.r. F est continue à droite en b donc

$$F(b) = \lim_{x \rightarrow b+} F(x) \geq u,$$

ce qui établit l'appartenance de b à I_u . Finalement, $I_u = [b, +\infty[$ et on a montré l'existence d'un réel b tel que

$$I_u := \{x \in \mathbb{R}; u \leq F(x)\} = [b, +\infty[.$$

Ceci légitime les équivalences

$$(F(x) \geq u) \Leftrightarrow (x \in I_u) \Leftrightarrow (x \geq b)$$

En posant $g(u) := b = \inf I_u = F^{-1}(u)$, au sens de (2.4), la fonction $g :]0, 1[\rightarrow \mathbb{R}$ ainsi définie vérifie (2.7), ce qui achève la preuve³. \square

Remarque 2.3. Il était commode dans la démonstration de considérer la loi uniforme sur l'intervalle *ouvert* $]0, 1[$. Comme $\{0\}$ et $\{1\}$ sont de mesure de Lebesgue nulle, cette loi est la même que la loi uniforme sur $[0, 1[$ ou $]0, 1]$ ou $[0, 1]$. Du point de vue informatique, il y a cependant une nuance car la loi uniforme simulée par la machine est la loi uniforme discrète sur $\{k2^{-d}; 0 \leq k < 2^d\}$ ou sur $\{k2^{-d}; 0 < k \leq 2^d\}$.

Voici des exemples simples d'utilisation de F^{-1} pour la simulation.

Exemple 2.4 (loi de Cauchy). La loi de Cauchy $\text{Cau}(0, 1)$ de densité $t \mapsto \frac{1}{\pi(1+t^2)}$ a pour f.d.r.

$$F : x \mapsto \frac{1}{\pi} \left(\frac{\pi}{2} + \arctan x \right).$$

Cette fonction de répartition est continue strictement croissante sur \mathbb{R} , donc réalise une bijection de \mathbb{R} sur $]0, 1[$. Il est facile de l'inverser. On en déduit que si U suit la loi uniforme sur $]0, 1[$, $Y := \tan(\pi(U - 1/2))$ suit la loi de Cauchy $\text{Cau}(0, 1)$.

3. En toute rigueur, il nous reste à vérifier la mesurabilité de l'application $g \circ U : \Omega \rightarrow \mathbb{R}$ pour pouvoir parler de la « variable aléatoire $Y := g(U)$ ». Cette mesurabilité s'obtient par composition en notant que g est croissante, donc borélienne.

Exemple 2.5 (lois exponentielles). Si U suit la loi uniforme sur $]0, 1[$, $Y := -\frac{\ln U}{a}$ suit la loi exponentielle de paramètre a . En fait ici $F(x) = 1 - \exp(-ax)$ s'inverse en $F^{-1}(u) = -\frac{\ln(1-u)}{a}$, mais on exploite le fait que $1 - U$ a même loi que U .

Exemple 2.6 (lois de Weibull). Les lois de Weibull sont très utilisées en fiabilité. La loi $\text{Weib}(a, b, c)$ de paramètres $a > 0$, $b \geq 0$ et $c > 0$ est caractérisée par sa fonction de survie $G(x) = 1 - F(x)$ donnée par

$$G_{a,b,c}(x) = \exp\left(-\left(\frac{x-b}{c}\right)^a\right), \text{ pour } x \geq b.$$

Clairement b est un paramètre de localisation et c un paramètre d'échelle de sorte que $\text{Weib}(a, b, c)$ se déduit par translation et changement d'échelle de la loi $\text{Weib}(a) = \text{Weib}(a, 0, 1)$ de fonction de survie

$$G_a(x) = \exp(-x^a), \text{ pour } x \geq 0.$$

La simulation de la loi $\text{Weib}(a, b, c)$ se ramène ainsi à celle de $\text{Weib}(a)$. En exploitant à nouveau le fait que U et $1-U$ ont même loi, on voit immédiatement que $Y := (-\ln U)^{1/a}$ suit la loi $\text{Weib}(a)$.

2.3 Méthodes particulières pour lois usuelles

2.3.1 Lois discrètes à support fini

Soit X une variable aléatoire discrète dont l'ensemble des valeurs possibles est fini :

$$X(\Omega) = \{x_1, \dots, x_d\}.$$

Notons

$$p_k := P(X = x_k), \quad s_0 := 0, \quad s_k := \sum_{i \leq k} p_i, \quad 1 \leq k \leq d.$$

Les points s_0, s_1, \dots, s_d induisent une partition en intervalles consécutifs de $[0, 1]$ et si U est une variable de loi uniforme sur $[0, 1]$, $P(U \in [s_{k-1}, s_k]) = s_k - s_{k-1} = p_k$. On en déduit que

$$Y := \sum_{k=1}^d x_k \mathbf{1}_{[s_{k-1}, s_k]}(U) \tag{2.8}$$

a même loi que X . L'écriture (2.8) est commode pour un mathématicien, mais il serait maladroit de la programmer telle quelle, car les d multiplications $x_k \mathbf{1}_{[s_{k-1}, s_k]}(U)$ et la somme de leurs d résultats sont inutiles. En pratique, il suffit de trouver pour $U(\omega)$ donné, l'unique indice $k = k(\omega)$ tel que $s_{k-1} \leq U(\omega) < s_k$ et de décider alors que $Y(\omega) = x_k$. C'est exactement ce que fait la fonction Scilab `discr1.sci` dont voici le code.


```

function [y]=discr1(x,p)
//
// simule une variable aléatoire discrète
// d'ensemble de valeurs possibles x_1,...,x_d
// avec probabilités respectives p_1,...,p_d
// x=( x_1,...,x_d), p=(p_1,...,p_d)
//
if sum(p) ~= 1 then
    error('La somme des probabilités doit valoir 1');
end
rand('uniform');
d=length(p);
pp=[0 p(1:(d-1))];
cpp=cumsum(pp); cp=cumsum(p);
U=rand(1,1);
k=find((cpp<= U)&(U<cp));
y=x(k)
endfunction

```

Si $X(\Omega)$ est infini dénombrable, la formule (2.8) reste valable avec une série au lieu d'une somme finie. On ne peut évidemment pas la programmer sous cette forme ! Une possibilité serait de réindexer les p_k (et donc aussi les x_k) de façon à obtenir une suite décroissante. Ensuite on testerait l'appartenance de $U(\omega)$ à $[s_{k-1}, s_k[$ en *s'arrêtant* dès que l'on obtient une réponse positive. Ainsi l'algorithme fournira une valeur $x_{k(\omega)}$ pour $Y(\omega)$ en un *temps fini*. Cette idée de commencer par les plus grosses valeurs des p_k peut d'ailleurs être utilisée pour optimiser la simulation de X dans le cas fini, particulièrement lorsque d est grand. Par exemple si $d = 1001$ et $p_{1001} = 1/2$, les autres p_k valant $1/2000$, il serait maladroit d'utiliser la fonction `discr1` telle qu'elle est programmée. Voici un deuxième code où l'on a essayé de minimiser le nombre de tests utilisés pour produire $Y(\omega)$.

```

function [y]=discr2(x,p)
//
// simule une variable aléatoire discrète
// d'ensemble de valeurs possibles x_1,...,x_d
// avec probabilités respectives p_1,...,p_d
// x=( x_1,...,x_d), p=(p_1,...,p_d)
// on optimise le nombre de tests en réarrangeant p
// par ordre décroissant et en quittant dès que la bonne valeur de
// k est trouvée
//
if sum(p) ~= 1 then
    error('La somme des probabilités doit valoir 1');
end

```

```

rand('uniform');
d=length(p);
[pr,i]=sort(p); //réarrangement de p
xr=x(i(:)); // réindexation correspondante pour x
cpr=cumsum(pr);
U=rand(1,1);
k=1;
while U>=cpr(k), k=k+1; end
y=xr(k)
endfunction

```

2.3.2 Lois binomiales et multinomiales

Pour simuler une variable aléatoire X de loi binomiale $\text{Bin}(n, p)$, plutôt que d'utiliser l'algorithme précédent, il est préférable de remarquer que la somme S_n de n variables de Bernoulli indépendantes et de même paramètre p suit la loi $\text{Bin}(n, p)$. Pour générer ces variables de Bernoulli, il suffit de les prendre égales à $\mathbf{1}_{\{U_i \leq p\}}$ où les U_i sont i.i.d. de loi uniforme sur $[0, 1]$. Ainsi X a même loi que

$$S_n := \sum_{k=1}^n \mathbf{1}_{\{U_k \leq p\}}.$$

On évite ainsi l'inconvénient du calcul des valeurs de la f.d.r. de X qui font intervenir des coefficients binomiaux et des puissances de p et $1 - p$.

En Scilab, un moyen commode de programmer une indicatrice est d'utiliser la fonction `bool2s` (dont le nom signifie *Boolean to string*). Elle prend en argument un booléen (vrai %T ou faux %F) et retourne la valeur 1 pour vrai et 0 pour faux. D'où le code très simple suivant :

```

function [Y] = simbin1(n,p)
//
// simule une v.a. de loi Bin(n,p)
//
rand('uniform');U=rand(1,n);
Y=sum(bool2s(U<=p))
endfunction

```

Pour les débutants en Scilab, noter que dans ce code, U est un vecteur de longueur n , il représente une réalisation $(U_1(\omega), \dots, U_n(\omega))$ du vecteur aléatoire (U_1, \dots, U_n) . De même $U \leq p$ est un vecteur booléen contenant les résultats des tests $U_1 \leq p, \dots, U_n \leq p$.

Cette méthode pour simuler une variable aléatoire de loi binomiale se généralise à la simulation d'un *vecteur aléatoire* de loi multinomiale. Rappelons que le vecteur aléatoire N suit la loi multinomiale de paramètres n et (p_1, \dots, p_d) où $n \in \mathbb{N}^*$ et les p_i sont

strictement positifs et de somme 1, si pour tout d -uplet (j_1, j_2, \dots, j_d) d'entiers tels que $j_1 + j_2 + \dots + j_d = n$,

$$P\{N = (j_1, j_2, \dots, j_d)\} = \frac{n!}{j_1! j_2! \dots j_d!} p_1^{j_1} p_2^{j_2} \dots p_d^{j_d}.$$

Un coup d'oeil sur cette formule devrait vous convaincre de l'intérêt d'éviter le calcul de ces probabilités. On remarque alors opportunément que le vecteur aléatoire N a *même loi* que $\sum_{k=1}^n X_k$, où les X_k sont des vecteurs aléatoires discrets de \mathbb{R}^d , indépendants et de même loi donnée par

$$P\{X_k = v_i\} = p_i, \quad 1 \leq i \leq d,$$

où l'on a posé

$$v_i := (0, \dots, 0, \underset{\substack{\uparrow \\ i}}{1}, 0, \dots, 0).$$

Autrement dit, on se ramène à un modèle d'épreuves répétées indépendantes et le vecteur aléatoire X_k est un codage binaire du résultat de la k^{e} épreuve. Pour simuler les X_k , on adapte de manière évidente (2.8) au cas d'un vecteur aléatoire discret. Notons donc $s_0 := 0$ et $s_i = p_1 + \dots + p_i$ pour $i = 1, \dots, d$. Les U_k désignant toujours des variables i.i.d. de loi uniforme sur $[0, 1]$, on voit finalement que N a même loi que

$$S_n := \sum_{k=1}^n \sum_{i=1}^d v_i \mathbf{1}_{\{s_{i-1} \leq U_k < s_i\}}.$$

Voici un code Scilab inspiré de cette formule (notez la permutation des sommations qui permet de construire le vecteur aléatoire $Y = S_n$ composante par composante, évitant l'utilisation de v_i).

```
function [Y]=simultin1(n,p)
//
// simulation d'un vecteur aléatoire de loi multinomiale de paramètres
// n (nombre d'épreuves) et p vecteur des probabilités de résultats
// élémentaires pour une épreuve.
// Retourne un vecteur colonne.
//
d=length(p);
s=[0 cumsum(p)]; // graduation de [0,1] en d intervalles de longueur p(i)
rand("uniform"); U=rand(1:n);
Y=zeros(d,1); // initialisation
for i=1:d,
Y(i,1)=sum(bool2s((s(i)<=U)&(U<s(i+1))))); // attention s(1)=0, s(d+1)=1
end
endfunction
```

2.3.3 Lois de Poisson

La variable aléatoire discrète X suit la loi de Poisson de paramètre α ($\alpha \in \mathbb{R}_+^*$) si

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

L'algorithme de simulation que nous allons proposer pour cette loi repose sur le lemme suivant.

Lemme 2.7. *Soit $(E_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes et de même loi exponentielle de paramètre α . Notons $S_1 := E_1$ et pour $n \geq 2$, $S_n := E_1 + \dots + E_n$. On a alors*

$$\forall n \geq 1, \quad P(S_n \leq 1 < S_{n+1}) = \frac{e^{-\alpha} \alpha^n}{n!}. \quad (2.9)$$

Preuve. Le vecteur aléatoire $V_{n+1} := (E_1, \dots, E_n, E_{n+1})$ a pour densité :

$$f_{n+1} : (x_1, \dots, x_{n+1}) \longmapsto \alpha^{n+1} \exp(-\alpha(x_1 + \dots + x_{n+1})) \mathbf{1}_{\mathbb{R}_+^{n+1}}(x_1, \dots, x_{n+1}),$$

parce que ses composantes sont indépendantes et de même densité f_1 . Pour exprimer $P(S_n \leq 1 < S_{n+1})$ à l'aide de la loi de V_{n+1} , on introduit le borélien

$$A_{n+1} := \{x \in \mathbb{R}^{n+1}; x_1 + \dots + x_n \leq 1 < x_1 + \dots + x_n + x_{n+1}\}$$

et on remarque que

$$P(S_n \leq 1 < S_{n+1}) = P(V_{n+1} \in A_{n+1}) = \int_{A_{n+1} \cap \mathbb{R}_+^{n+1}} f_{n+1}(x_1, \dots, x_{n+1}) dx_1 \dots dx_{n+1}.$$

Pour calculer cette intégrale, il est commode d'utiliser le changement de variable linéaire bijectif

$$\varphi : (x_1, \dots, x_{n+1}) \longmapsto (s_1, \dots, s_{n+1}), \quad \text{où} \quad s_k := \sum_{i \leq k} x_i, \quad k = 1, \dots, n+1.$$

On voit immédiatement que l'application φ^{-1} est donnée par $x_1 = s_1$ et $x_k = s_k - s_{k-1}$ pour $k \geq 2$ et que son déterminant vaut 1. En notant pour alléger $A_{n+1}^+ := A_{n+1} \cap \mathbb{R}_+^{n+1}$, la formule de changement de variable pour les bijections linéaires s'écrit donc

$$P(S_n \leq 1 < S_{n+1}) = \int_{\varphi(A_{n+1}^+)} \alpha^{n+1} \exp(-\alpha s_{n+1}) ds_1 \dots ds_{n+1}.$$

Pour déterminer $\varphi(A_{n+1}^+)$, on note que A_{n+1}^+ est caractérisé par les inéquations :

$$\forall i = 1, \dots, n+1, \quad x_i \geq 0 \quad \text{et} \quad x_1 + \dots + x_n \leq 1 < x_1 + \dots + x_n + x_{n+1}.$$

En remplaçant x par $\varphi^{-1}(s)$ dans ces inéquations, on obtient la caractérisation de $\varphi(A_{n+1}^+)$, à savoir : $s_1 \geq 0$, $\forall i = 2, \dots, n+1$, $s_i - s_{i-1} \geq 0$ et $s_n \leq 1 < s_{n+1}$. On voit ainsi que

$$\varphi(A_{n+1}^+) = \{s \in \mathbb{R}^{n+1}; 0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq 1 < s_{n+1}\} = B_n \times]1, +\infty[.$$

où l'on a noté B_n le simplexe

$$B_n := \{s \in \mathbb{R}^n; 0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq 1\}.$$

Cette écriture de $\varphi(A_{n+1}^+)$ en produit cartésien nous permet d'appliquer le théorème de Fubini Tonelli pour obtenir

$$\begin{aligned} P(S_n \leq 1 < S_{n+1}) &= \alpha^n \left\{ \int_{B_n} ds_1 \dots ds_n \right\} \left\{ \int_{]1, +\infty[} \alpha \exp(-\alpha s_{n+1}) ds_{n+1} \right\} \\ &= \alpha^n \lambda_n(B_n) \int_1^{+\infty} \alpha \exp(-\alpha t) dt \\ &= \alpha^n \lambda_n(B_n) e^{-\alpha}. \end{aligned}$$

Il ne reste plus qu'à vérifier que $\lambda_n(B_n) = 1/n!$ pour achever la preuve de (2.9). On peut le voir de manière géométrique en notant $C_n := [0, 1]^n$ et C'_n le sous-ensemble de C obtenu en supprimant dans C tous les points ayant au moins deux coordonnées égales. Comme $C_n \setminus C'_n$ est inclus dans une réunion finie d'hyperplans (d'équation $s_i = s_j$) tous de λ_n -mesure nulle, $\lambda_n(C_n) = \lambda_n(C'_n)$. On partitionne alors C'_n en $n!$ simplexes se déduisant de $B'_n := \{s \in \mathbb{R}^n; 0 \leq s_1 < s_2 < \dots < s_n \leq 1\}$ par une permutation de coordonnées. La mesure de Lebesgue λ_n étant invariante par permutation de coordonnées, on en déduit $\lambda_n(C'_n) = n! \lambda_n(B'_n)$, puis $1 = \lambda_n(C_n) = n! \lambda_n(B_n)$. \square

Pour compléter le lemme 2.7, remarquons que l'on peut étendre (2.9) au cas particulier $n = 0$. En posant $S_0 := 0$, (2.9) résulte dans ce cas de l'égalité

$$P(1 < S_1) = P(E_1 \in]1, +\infty[) = \int_1^{+\infty} \alpha \exp(-\alpha t) dt = e^{-\alpha}.$$

Il résulte immédiatement du lemme 2.7 que la variable aléatoire

$$Y := \sum_{k=1}^{+\infty} k \mathbf{1}_{\{S_k \leq 1 < S_{k+1}\}}$$

suit la loi de Poisson de paramètre α . Pour déduire de cette formule un algorithme effectif, on simule les E_i par la méthode de l'exemple 2.5, ce qui revient à poser $E_i = -\alpha^{-1} \ln U_i$, les U_i étant i.i.d. de loi uniforme sur $[0, 1]$. On dispose alors des équivalences

$$\begin{aligned} S_k \leq 1 < S_{k+1} &\Leftrightarrow \frac{-1}{\alpha} \sum_{i=1}^k \ln U_i \leq 1 < \frac{-1}{\alpha} \sum_{i=1}^{k+1} \ln U_i \\ &\Leftrightarrow \prod_{i=1}^{k+1} U_i < e^{-\alpha} \leq \prod_{i=1}^k U_i. \end{aligned} \tag{2.10}$$

L'algorithme est donc le suivant. On génère l'une après l'autre les variables U_i et on compare leur produit à $e^{-\alpha}$, en s'arrêtant lorsque (2.10) est vérifiée. On attribue alors à Y la valeur correspondante de k . Voici une traduction en Scilab de cet algorithme.

```
function [Y]=simpois1(alpha)
//
// simule une variable aléatoire suivant
// la loi de Poisson de paramètre alpha
//
rand('uniform');
a=exp(-alpha);
k=0; // initialisation
M=rand(1,1);
while (M>=a) do
    M=M*rand(1,1); k=k+1;
end
Y=k;
endfunction
```

Comme la suite de terme général $M_n := U_1 \dots U_n$ converge presque sûrement vers 0 (exercice!), on est sûr d'aboutir en un nombre fini de pas. De façon plus quantitative, on peut dire aussi que le nombre « moyen » de variables U_i à générer pour que l'algorithme aboutisse est égal à $1 + \alpha$ puisque $\mathbf{E}Y = \alpha$ (si l'algorithme répond $Y(\omega) = k$, c'est qu'il a calculé $k + 1$ variables U_i pour arriver à (2.10)). Notons N_n le nombre total de variables aléatoires U_i à générer pour simuler un échantillon de taille n de variables aléatoires indépendantes et de même loi que Y . En remarquant que $\text{Var}(1 + Y) = \text{Var} Y = \alpha$, on a par l'inégalité de Tchebycheff

$$P(|N_n - n(1 + \alpha)| \geq t\sqrt{n\alpha}) \leq \frac{1}{t^2}.$$

Par exemple, la simulation de 400 v.a. de Poisson de paramètre $\alpha = 4$ « consomme » un nombre de variables uniformes compris avec une probabilité d'au moins 99% entre 1 600 et 2 400 .

2.3.4 Loïs géométriques

Rappelons que X suit la loi géométrique de paramètre $p \in]0, 1[$ si

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

C'est la loi du temps d'attente d'un premier succès dans une suite d'épreuve répétées indépendantes ayant chacune même probabilité de succès p . Par conséquent si $(U_i)_{i \geq 1}$ est une suite i.i.d. de variables suivant la loi uniforme sur $[0, 1]$, la variable aléatoire

$$Y := \min\{k \in \mathbb{N}^*; U_k < p\} \tag{2.11}$$

suit la loi géométrique⁴ de paramètre p . L'algorithme correspondant à (2.11) consiste donc à générer une v.a. U_k de loi uniforme tant que la valeur obtenue est supérieure ou

4. En toute rigueur, il faudrait adopter la convention $\min \emptyset = +\infty$ et considérer Y comme variable aléatoire discrète $(\Omega, \mathcal{F}) \rightarrow \mathbb{N}^* \cup \{+\infty\}$. Pour en faire une variable discrète classique, il faut la définir sur $\Omega' := \Omega \setminus \{Y = +\infty\}$ muni de la tribu trace \mathcal{F}' de \mathcal{F} . Comme $P(Y = +\infty) = P(\forall k \in \mathbb{N}^*, U_k \geq p) = 0$ (exercice!), cela ne change pas la loi de $Y \dots$

égale à p et à s'arrêter à la première valeur de U_k strictement inférieure à p en retournant son indice.

```
function [Y]=simgeom1(p)
//
// simule une v.a. suivant la loi géométrique de paramètre p
// comme temps d'attente du premier succès dans un schéma de Bernoulli
//
rand('uniform');
k=1; // initialisation
U=rand(1,1);
while (U>=p) do
    U=rand(1,1); k=k+1;
end
Y=k;
endfunction
```

Cet algorithme s'arrête en un temps fini car $p > 0$ et $\min_{i \leq n} U_i$ converge p.s. vers 0 (exercice). Le nombre moyen de variables aléatoires uniformes utilisées est $\mathbf{E}Y = 1/p$, ce qui peut être coûteux pour les petites valeurs de p . De ce point de vue, le deuxième algorithme proposé ci-dessous, quoique moins naturel, est plus économique puisqu'il n'utilise qu'une seule variable aléatoire uniforme. Il repose sur le lemme suivant dont la vérification est immédiate.

Lemme 2.8. *Si E est une v.a. de loi exponentielle de paramètre a , alors*

$$\forall n \geq 1, \quad P(n-1 < E \leq n) = (e^{-a})^{n-1} (1 - e^{-a}). \quad (2.12)$$

Comme $0 < a < +\infty$, e^{-a} et $1 - e^{-a}$ sont dans $]0, 1[$. Choissant alors a dans (2.12) tel que $1 - e^{-a} = p$, c.-à-d. $a := -\ln(1 - p)$, on voit que la partie entière supérieure de E suit la loi géométrique de paramètre p .

En simulant E par la méthode de l'exemple 2.5, on aboutit au code Scilab suivant (ceil est la fonction partie entière supérieure) :

```
function [Y]=simgeom2(p)
//
// simule une v. a. suivant la loi géométrique de paramètre p
// comme partie entière supérieure d'une v.a. de loi
// exponentielle de paramètre a=-ln(1-p)
//
Y=ceil(log(rand(1,1,'uniform'))./log(1-p));
endfunction
```

2.3.5 Lois gaussiennes

La plupart des logiciels utilisant un générateur de nombres aléatoires ont une option « gaussiennes » permettant de simuler des variables gaussiennes. Par exemple en Scilab

l'instruction `rand('normal')` est un commutateur qui fait passer le générateur en mode gaussien. Après cette instruction, `rand(m,n)` génère une matrice $m \times n$ dont les termes sont considérés comme des $X_{i,j}(\omega)$, les $X_{i,j}$ étant des variables aléatoires de même loi $\mathfrak{N}(0,1)$ et indépendantes⁵. Cette fonctionnalité du générateur ne devrait pas nous dispenser de réfléchir un instant à la façon dont on peut programmer la simulation d'une variable aléatoire gaussienne à partir de la génération de variables uniformes.

En notant Φ la fonction de répartition de $\mathfrak{N}(0,1)$, qui est continue et strictement croissante sur \mathbb{R} , la méthode générale exposée à la section 2.2 nous propose de prendre $\Phi^{-1}(U)$. L'ennui c'est qu'on ne connaît pas d'expression analytique pour Φ ni pour Φ^{-1} et que l'inversion numérique de Φ serait assez coûteuse. Le lemme suivant fournit immédiatement un algorithme bien plus simple, connu sous le nom de méthode de Box Muller.

Lemme 2.9. *Si U_1 et U_2 sont indépendantes et de loi uniforme sur $]0,1[$, les variables aléatoires*

$$X := (-2 \ln U_1)^{1/2} \cos(2\pi U_2) \quad \text{et} \quad Y := (-2 \ln U_1)^{1/2} \sin(2\pi U_2)$$

sont indépendantes et de même loi $\mathfrak{N}(0,1)$.

Preuve. Le vecteur aléatoire (U_1, U_2) suit la loi uniforme sur $]0,1[^2$, de densité $f_{U_1, U_2} = \mathbf{1}_{]0,1[^2}$. Le changement de variable

$$g : (u_1, u_2) \mapsto (x, y) \quad \text{avec} \quad \begin{cases} x = (-2 \ln u_1)^{1/2} \cos(2\pi u_2) \\ y = (-2 \ln u_1)^{1/2} \sin(2\pi u_2) \end{cases}$$

réalise un C^1 -difféomorphisme de l'ouvert $D =]0,1[^2$ sur l'ouvert $D' = \mathbb{R}^2 \setminus \{(0,0)\}$. On sait alors, cf. cours d'I.P.É. chapitre 5, que le vecteur aléatoire $(X, Y) = g(U_1, U_2)$ a une densité $f_{X,Y}$ donnée par

$$f_{X,Y}(x, y) = f_{U_1, U_2}(g^{-1}(x, y)) |\text{Jac}(g^{-1})(x, y)| \mathbf{1}_{D'}(x, y).$$

Pour tout $(x, y) \in \mathbb{R}^2 \setminus \{(0,0)\}$, $g^{-1}(x, y)$ appartient à $]0,1[^2$ et comme f_{U_1, U_2} vaut 1 sur $]0,1[^2$, le premier facteur dans la formule ci-dessus s'écrit $f_{U_1, U_2}(g^{-1}(x, y)) = 1$. Pour calculer $\text{Jac}(g^{-1})(x, y)$, on calcule $\text{Jac}(g)(u_1, u_2)$ et on utilise la formule

$$\text{Jac}(g^{-1})(x, y) = \frac{1}{\text{Jac}(g)(g^{-1}(x, y))}.$$

En notant $g = (g_1, g_2)$, les dérivées partielles de g sont

$$\begin{aligned} \frac{\partial g_1}{\partial u_1} &= -u_1^{-1}(-2 \ln u_1)^{-1/2} \cos(2\pi u_2), & \frac{\partial g_1}{\partial u_2} &= -2\pi(-2 \ln u_1)^{1/2} \sin(2\pi u_2), \\ \frac{\partial g_2}{\partial u_1} &= -u_1^{-1}(-2 \ln u_1)^{-1/2} \sin(2\pi u_2), & \frac{\partial g_2}{\partial u_2} &= 2\pi(-2 \ln u_1)^{1/2} \cos(2\pi u_2). \end{aligned}$$

5. Au risque d'enfoncer une porte ouverte, rappelons à ce propos que si X a pour loi $\mathfrak{N}(0,1)$, $\sigma X + m$ a pour loi $\mathfrak{N}(m, \sigma)$ (pour tous $\sigma \in \mathbb{R}_+$ et $m \in \mathbb{R}$). Ainsi pour simuler n'importe quelle loi gaussienne (en dimension 1), il suffit de savoir simuler $\mathfrak{N}(0,1)$.

On en déduit que

$$\text{Jac}(g)(u_1, u_2) = \left| \begin{array}{cc} \frac{\partial g_1}{\partial u_1} & \frac{\partial g_1}{\partial u_2} \\ \frac{\partial g_2}{\partial u_1} & \frac{\partial g_2}{\partial u_2} \end{array} \right| = \frac{2\pi}{u_1} \cos^2(2\pi u_2) + \frac{2\pi}{u_1} \sin^2(2\pi u_2) = \frac{2\pi}{u_1}.$$

Pour exprimer u_1 en fonction de (x, y) , on remarque que $x^2 + y^2 = -2 \ln u_1$, d'où $u_1 = \exp(-(x^2 + y^2)/2)$ et

$$|\text{Jac}(g^{-1})(x, y)| = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right).$$

Finalement la densité du couple (X, Y) est donnée par

$$f_{X,Y}(x, y) = \frac{1}{2\pi} \exp\left(-\frac{x^2 + y^2}{2}\right) \mathbf{1}_{D'}(x, y).$$

Comme $\mathbf{1}_{D'}$ et $\mathbf{1}_{\mathbb{R}^2}$ ne diffèrent que sur le singleton $\{(0, 0)\}$ qui est un ensemble de mesure nulle, (X, Y) admet aussi pour densité⁶ $\tilde{f}_{X,Y}(x, y) = (2\pi)^{-1} \exp(-(x^2 + y^2)/2)$ qui est la forme classique de la densité d'un couple de v.a. i.i.d. $\mathfrak{N}(0, 1)$. Donc X et Y sont gaussiennes $\mathfrak{N}(0, 1)$ indépendantes. \square

2.4 Algorithmes de rejet

La méthode du rejet (appelée aussi d'acceptation-rejet) peut être décrite abstraitement comme suit. On suppose que l'on sait générer un vecteur aléatoire M_1 de \mathbb{R}^d suivant une certaine loi μ . On génère alors l'un après l'autre, les vecteurs de la suite i.i.d. M_1, \dots, M_n, \dots en s'arrêtant au premier d'entre eux qui vérifie une certaine condition (\mathcal{H}_0) . Soit T l'indice (aléatoire) correspondant. On a ainsi fabriqué un vecteur (doublement) aléatoire M_T . Comme T est aléatoire, la loi de ce vecteur n'est pas celle de M_1 , c'est une *nouvelle* loi ν . Si la simulation de M_1 et le test de (\mathcal{H}_0) sont facilement programmables, on dispose ainsi d'une méthode pour générer un vecteur aléatoire de loi ν . Nous allons voir comment fonctionne ce principe général d'abord pour la simulation d'un vecteur aléatoire de loi uniforme sur un borélien de \mathbb{R}^d , puis pour celle d'un vecteur aléatoire de densité connue et enfin pour simuler certaines lois discrètes.

2.4.1 Simulation de lois uniformes par rejet

Commençons par rappeler la définition de la loi uniforme sur un borélien.

Définition 2.10. Soient (Ω, \mathcal{F}, P) un espace probabilisé et B un borélien de \mathbb{R}^d tel que $0 < \lambda_d(B) < +\infty$, λ_d désignant la mesure de Lebesgue sur \mathbb{R}^d . Le vecteur aléatoire $M : \Omega \rightarrow \mathbb{R}^d$ suit la loi uniforme sur B si sa loi P_M est la probabilité uniforme sur B , c.-à-d.

$$\forall A \in \text{Bor}(\mathbb{R}^d), \quad P_M(A) = P(M \in A) = \frac{\lambda_d(A \cap B)}{\lambda_d(B)}. \quad (2.13)$$

6. On peut remplacer $f_{X,Y}$ par $\tilde{f}_{X,Y}$ dans les intégrales sans changer la valeur de ces intégrales.

Rappelons aussi que si le vecteur aléatoire M de \mathbb{R}^d suit la loi uniforme sur B , il admet pour densité

$$f_M = \frac{1}{\lambda_d(B)} \mathbf{1}_B.$$

À partir d'une variable U de loi uniforme sur $[0, 1]$, on fabrique facilement V de loi uniforme sur $[a, b]$ en prenant $V := a + (b - a)U$. On en déduit la construction d'un vecteur aléatoire $M = (V_1, \dots, V_d)$ de loi uniforme sur le pavé $B = [a_1, b_1] \times \dots \times [a_d, b_d]$, en prenant $V_i := a_i + (b_i - a_i)U_i$, les U_i étant i.i.d. de loi uniforme sur $[0, 1]$. En effet la loi P_{V_i} de V_i a pour densité $c_i \mathbf{1}_{[a_i, b_i]}$ avec $c_i = (b_i - a_i)^{-1}$. Les V_i héritent de l'indépendance des U_i puisque $V_i = h_i(U_i)$ avec h_i mesurable. On en déduit que P_M a pour densité la fonction f donnée par

$$\begin{aligned} f(x_1, \dots, x_d) &= c_1 \mathbf{1}_{[a_1, b_1]}(x_1) \dots c_d \mathbf{1}_{[a_d, b_d]}(x_d) = (c_1 \dots c_d) \mathbf{1}_{[a_1, b_1] \times \dots \times [a_d, b_d]}(x_1, \dots, x_d) \\ &= \lambda_d(B)^{-1} \mathbf{1}_B(x_1, \dots, x_d), \end{aligned}$$

autrement dit que P_M est la loi uniforme sur B .

En dehors de ce cas particulier où B est un pavé et de ceux qui s'y ramènent par transformation affine, par exemple loi uniforme sur un parallélogramme, la simulation d'un vecteur aléatoire de loi uniforme sur un borélien demande un peu plus de travail. L'algorithme du rejet est souvent une bonne solution à ce problème. On suppose que

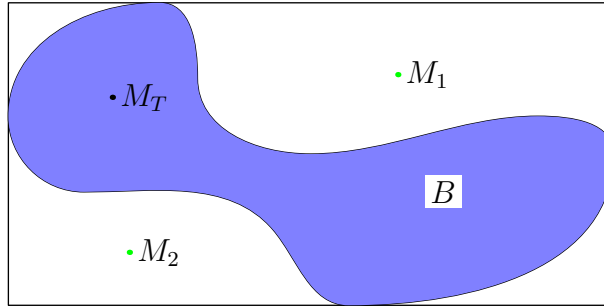


FIG. 2.2 – Simulation par rejet de la loi uniforme sur B , ici $T(\omega) = 3$

l'on sait générer un vecteur aléatoire M_1 de loi uniforme sur un borélien C contenant B (c'est le cas notamment lorsque B est borné en prenant pour C un pavé assez grand). On génère alors séquentiellement les vecteurs i.i.d. M_1, \dots, M_n, \dots en s'arrêtant au premier d'entre eux qui vérifie la condition $M_i \in B$. Soit T l'indice (aléatoire) correspondant. T est donc le numéro du premier point « tombé » dans B . Alors le vecteur aléatoire M_T suit la loi uniforme sur B .

La justification de cet algorithme repose sur le résultat suivant, qu'il est commode d'énoncer dans un cadre un peu plus général.

Proposition 2.11. *Soit $(M_n)_{n \in \mathbb{N}^*}$ une suite de vecteurs aléatoires $\Omega \rightarrow \mathbb{R}^d$, indépendants et de même loi μ . Soit B un borélien de \mathbb{R}^d tel que $\mu(B) > 0$. Pour tout $\omega \in \Omega$, on pose*

$$T(\omega) := \inf\{i \in \mathbb{N}^*; M_i(\omega) \in B\},$$

avec la convention $\inf \emptyset = +\infty$. On définit $M_T : \Omega \rightarrow \mathbb{R}^d$ par

$$M_T(\omega) := \begin{cases} M_{T(\omega)}(\omega) & \text{si } T(\omega) < +\infty, \\ 0 & \text{si } T(\omega) = +\infty. \end{cases}$$

Dans ces conditions,

- a) $P(T < +\infty) = 1$ et $T' := T\mathbf{1}_{\{T < +\infty\}}$ est une variable aléatoire suivant la loi géométrique de paramètre $p = \mu(B)$.
- b) M_T est un vecteur aléatoire de loi ν donnée par

$$\forall A \in \text{Bor}(\mathbb{R}^d), \quad \nu(A) = P(M_T \in A) = \frac{\mu(A \cap B)}{\mu(B)}.$$

Autrement dit, ν est la probabilité conditionnelle $\mu(\cdot \mid B)$.

En particulier lorsque μ est la loi uniforme sur un borélien C contenant B et tel que $0 < \lambda_d(B) < \lambda_d(C)$, ν est la loi uniforme sur B .

Preuve du a). L'introduction de T' a pour seul but d'éviter de travailler avec une v.a. à valeurs dans $\overline{\mathbb{N}}$. Nous verrons dans un instant que de toutes façons T et T' sont « p.s. égales ». Commençons par justifier la mesurabilité de l'application $T' : \Omega \rightarrow \mathbb{N}$. En raison de la dénombrabilité de \mathbb{N} , il suffit de vérifier⁷ que

$$\forall k \in \mathbb{N}, \quad T'^{-1}(\{k\}) \in \mathcal{F}. \quad (2.14)$$

Si $k \in \mathbb{N}^*$, on a $\{T' = k\} = \{T = k\}$ d'où

$$\{T' = k\} = \{\forall i < k, M_i \notin B \text{ et } M_k \in B\} = \left(\bigcap_{1 \leq i < k} M_i^{-1}(B^c) \right) \cap M_k^{-1}(B), \quad (2.15)$$

tandis que dans le cas particulier $k = 0$,

$$\{T' = 0\} = \{T = +\infty\} = \{\forall i \in \mathbb{N}^*, M_i \notin B\} = \bigcap_{i \in \mathbb{N}^*} M_i^{-1}(B^c). \quad (2.16)$$

Les M_i étant des vecteurs aléatoires dans \mathbb{R}^d , donc mesurables \mathcal{F} - $\text{Bor}(\mathbb{R}^d)$, l'image réciproque par M_i d'un borélien quelconque de \mathbb{R}^d est un élément de \mathcal{F} . Il résulte alors de (2.15) et de (2.16) que $T'^{-1}(\{k\})$ est intersection finie ou dénombrable d'éléments de \mathcal{F} , ce qui établit (2.14). Ainsi T' est bien une variable aléatoire discrète à valeurs dans \mathbb{N} .

Sa loi est caractérisée par les $P(T' = k)$. Posons $p = \mu(B) = P(M_i \in B)$. Si $k \in \mathbb{N}^*$, la décomposition (2.15) et l'indépendance des M_i nous donnent

$$\forall k \in \mathbb{N}^*, \quad P(T = k) = (1 - p)^{k-1}p. \quad (2.17)$$

Calculons $P(T' = 0) = P(T = +\infty)$. Par hypothèse $\mu(B) > 0$, donc $1 - p = \mu(B^c) = P(M_i^{-1}(B^c))$ est strictement inférieur à 1. En remarquant que pour tout $n \in \mathbb{N}^*$ on

7. Voir la définition d'une v.a. discrète dans le cours d'I.P.É.

a l'inclusion $\cap_{i \in \mathbb{N}^*} M_i^{-1}(B^c) \subset \cap_{i \leq n} M_i^{-1}(B^c)$, on a par indépendance $P(T = +\infty) \leq (1 - p)^n$, d'où en faisant tendre n vers l'infini,

$$P(T' = 0) = P(T = +\infty) = 0. \quad (2.18)$$

D'après (2.17), T' suit la loi géométrique de paramètre $p = \mu(B)$. \square

Preuve du b). Commençons par vérifier que M_T est un vecteur aléatoire, c'est-à-dire une application \mathcal{F} -Bor(\mathbb{R}^d) mesurable, en montrant que $M_T^{-1}(A) \in \mathcal{F}$ pour tout $A \in \text{Bor}(\mathbb{R}^d)$. En partitionnant Ω suivant les évènements $\{T = k\}$, on a la décomposition :

$$M_T^{-1}(A) = \{\omega \in \Omega; M_{T(\omega)}(\omega) \in A\} = \bigcup_{k \in \mathbb{N}^*} (\{M_T \in A\} \cap \{T = k\}).$$

Cette union étant dénombrable, il suffit de vérifier que chacun de ses termes est un élément de \mathcal{F} . Si $k \in \mathbb{N}^*$, ceci résulte de la mesurabilité des M_i via la décomposition

$$\begin{aligned} \{M_T \in A\} \cap \{T = k\} &= \{\omega \in \Omega; T(\omega) = k \text{ et } M_k(\omega) \in A\} \\ &= \bigcap_{i < k} \{M_i \notin B\} \cap \{M_k \in B\} \cap \{M_k \in A\} \\ &= \bigcap_{i < k} M_i^{-1}(B^c) \cap M_k^{-1}(A \cap B). \end{aligned} \quad (2.19)$$

Dans le cas particulier $k = +\infty$,

$$\{M_T \in A\} \cap \{T = +\infty\} = \bigcap_{i \in \mathbb{N}^*} M_i^{-1}(B^c) \cap \{M_\infty \in A\}.$$

M_∞ étant le vecteur aléatoire constant 0, $\{M_\infty \in A\}$ vaut Ω ou \emptyset selon que 0 appartient ou non à A . On a donc une intersection dénombrable d'éléments de \mathcal{F} .

Maintenant que nous voilà rassurés sur la mesurabilité de M_T , on peut s'intéresser à sa loi ν que l'on détermine en calculant $\nu(A) = P(M_T \in A)$ pour $A \in \text{Bor}(\mathbb{R}^d)$. En partitionnant par les $\{T = k\}$ et compte-tenu de (2.18), on obtient :

$$\nu(A) = \sum_{k \in \mathbb{N}^*} P(M_T \in A \text{ et } T = k). \quad (2.20)$$

La décomposition (2.19) et l'indépendance des M_i nous donnent

$$P(M_T \in A \text{ et } T = k) = (1 - p)^{k-1} \mu(A \cap B),$$

d'où en reportant dans (2.20) et en notant que $0 < 1 - p < 1$,

$$\nu(A) = \mu(A \cap B) \sum_{k \in \mathbb{N}^*} (1 - p)^{k-1} = \mu(A \cap B) \frac{1}{1 - (1 - p)} = \frac{\mu(A \cap B)}{\mu(B)}.$$

On a donc bien $\nu = \mu(\cdot \mid B)$. Dans le cas particulier où μ est la loi uniforme sur un borélien C contenant B et tel que $0 < \lambda_d(B) < \lambda_d(C)$, on obtient grâce à (2.13),

$$\forall A \in \text{Bor}(\mathbb{R}^d), \quad \nu(A) = \frac{\frac{\lambda_d(A \cap B \cap C)}{\lambda_d(C)}}{\frac{\lambda_d(B \cap C)}{\lambda_d(C)}} = \frac{\lambda_d(A \cap B)}{\lambda_d(B)},$$

ce qui montre que ν est la loi uniforme sur B . \square

2.4.2 Simulation de lois à densité par rejet

La méthode du rejet permet aussi de simuler des variables ou des vecteurs aléatoires dont la loi est à densité par rapport à la mesure de Lebesgue, en s'aidant d'une autre densité dont on sait simuler la loi. Commençons par la description de l'algorithme.

On voudrait simuler une variable aléatoire Z , de densité f . On suppose que l'on sait simuler une v.a. X de densité g et trouver une constante c telle que $f \leq cg$ (nécessairement $c \geq 1$, pourquoi?). Les X_i sont des v.a. indépendantes de même loi ayant pour densité g , les U_i des variables aléatoires indépendantes de même loi uniforme sur $[0, 1]$ et les suites $(X_i)_{i \geq 1}$ et $(U_i)_{i \geq 1}$ sont indépendantes. On génère la suite des $M_i := (X_i, cg(X_i)U_i)$ en s'arrêtant au premier indice i_0 (aléatoire) tel que $cg(X_{i_0})U_{i_0} \leq f(X_{i_0})$. On pose alors $Z =: X_{i_0}$ et Z a pour densité f .

La justification de cet algorithme repose sur la proposition 2.11 combinée avec les deux propositions suivantes dont nous différons légèrement la preuve.

Proposition 2.12. *Soit f une densité de probabilité sur \mathbb{R} et G son hypographe :*

$$G := \{(x, y) \in \mathbb{R} \times \mathbb{R}; 0 \leq y \leq f(x)\}.$$

Soit $M = (Z, Y)$ un vecteur aléatoire de $\mathbb{R} \times \mathbb{R}$ de loi uniforme sur G . Alors la loi de la variable aléatoire Z a pour densité f .

Proposition 2.13. *Soit X une variable aléatoire réelle de densité g . Posons*

$$M := (X, cg(X)U),$$

où U est une variable aléatoire de loi uniforme sur $[0, 1]$, indépendante de X et $c > 0$ une constante. On note H l'hypographe de cg :

$$H := \{(x, y) \in \mathbb{R} \times \mathbb{R}; 0 \leq y \leq cg(x)\}.$$

Alors M suit la loi uniforme sur H .

Avant de les démontrer, voyons d'abord comment on utilise les propositions 2.12 et 2.13.

Justification de l'algorithme. On conserve les notations des propositions 2.12 et 2.13. Les fonctions f et g sont des densités donc $\int_{-\infty}^{+\infty} f(t) dt = \int_{-\infty}^{+\infty} g(t) dt = 1$. D'autre part $f \leq cg$, donc G est inclus dans H .

En utilisant l'interprétation de l'intégrale généralisée d'une fonction positive comme aire de son hypographe, cf. cours I.P.É. Annexe B, on a

$$1 = \int_{-\infty}^{+\infty} f(t) dt = \lambda_2(G) \leq \lambda_2(H) = \int_{-\infty}^{+\infty} cg(t) dt = c,$$

ce qui montre que $c \geq 1$ et que $\lambda_2(H \setminus G) = c - 1$. Donc en imposant à la constante c de vérifier à la fois $f \leq cg$ et $c > 1$, on est sûr que $0 < \lambda_2(G) < \lambda_2(H) < +\infty$. Par hypothèse, on sait simuler les suites indépendantes $(X_i)_{i \geq 1}$ et $(U_i)_{i \geq 1}$, la première étant

i.i.d. avec les X_i de densité g et la seconde i.i.d. avec les U_i de loi uniforme sur $[0, 1]$. En raison de toutes ces indépendances, la suite des vecteurs $M_i := (X_i, cg(X_i)U_i)$ est i.i.d. et par la proposition 2.13, les M_i suivent la loi uniforme sur H .

Posons maintenant

$$T(\omega) := \inf\{i \in \mathbb{N}^*; M_i(\omega) \in G\}$$

et

$$M_T(\omega) := \begin{cases} M_{T(\omega)}(\omega) & \text{si } T(\omega) < +\infty, \\ 0 & \text{si } T(\omega) = +\infty. \end{cases}$$

Par la proposition 2.11, M_T est un vecteur aléatoire de \mathbb{R}^2 , de loi uniforme sur G . La proposition 2.12 nous permet alors de conclure que sa première composante $Z = X_T$ a pour densité f . \square

Signalons pour les curieux que les propositions 2.12 et 2.13 et donc l'algorithme de simulation par rejet d'une loi à densité donnée se généralisent⁸ à la dimension d en remplaçant partout $\mathbb{R} \times \mathbb{R}$ par $\mathbb{R}^d \times \mathbb{R}$, Z et X devenant alors des *vecteurs aléatoires* de \mathbb{R}^d .

Preuve de la proposition 2.12. Pour identifier la loi de Z lorsque $M = (Z, Y)$ suit la loi uniforme sur G , nous allons calculer $P(Z \in [a, b])$ pour tout intervalle $[a, b] \subset \mathbb{R}$. Notons $G_{a,b}$ l'intersection de l'hypographe G avec la « bande verticale » $[a, b] \times \mathbb{R}$, cf. figure 2.3. On peut aussi voir $G_{a,b}$ comme l'hypographe de la restriction de f à $[a, b]$. On s'appuie

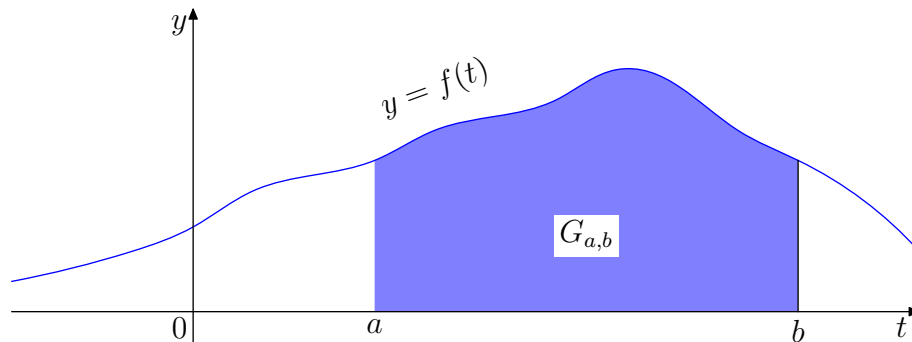


FIG. 2.3 – Ensemble $G_{a,b}$

sur notre connaissance de la loi de M en notant que

$$P(Z \in [a, b]) = P((Z, Y) \in [a, b] \times \mathbb{R}) = \frac{\lambda_2(G \cap ([a, b] \times \mathbb{R}))}{\lambda_2(G)} = \lambda_2(G_{a,b}), \quad (2.21)$$

car $\lambda_2(G) = \int_{-\infty}^{+\infty} f(t) dt = 1$. L'aire de l'hypographe d'une fonction positive sur $[a, b]$ étant égale à l'intégrale de Riemann (éventuellement généralisée) de cette fonction sur $[a, b]$, on a

$$P(Z \in [a, b]) = \lambda_2(G_{a,b}) = \int_a^b f(t) dt.$$

8. Voir <http://math.univ-lille1.fr/~suquet/ens/Agr/simul06.pdf>

Cette formule étant valable pour tous réels a et b tels que $a < b$, on en déduit que Z a pour densité f . \square

Preuve de la proposition 2.13. Pour identifier la loi de $M := (X, cg(X)U)$, nous allons calculer $\mathbf{E}h(M)$ pour $h : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ continue à support compact⁹ quelconque. M étant une fonction du couple (X, U) , il est commode d'exprimer $\mathbf{E}h(M)$ à l'aide de la densité de ce couple. Par indépendance de X et U , cette densité s'écrit $(x, u) \mapsto g(x)\mathbf{1}_{[0,1]}(u)$. Ainsi

$$\begin{aligned} \mathbf{E}h(M) &= \int_{\mathbb{R}^2} h(x, cg(x)u) g(x) \mathbf{1}_{[0,1]}(u) \, dx \, du \\ &= \int_{-\infty}^{+\infty} \frac{1}{c} \left\{ \int_0^1 h(x, cg(x)u) cg(x) \, du \right\} \, dx. \end{aligned} \quad (2.22)$$

Notons $I(x)$ l'intégrale entre accolades. Si $g(x) > 0$, le changement de variable $y = cg(x)u$ à x fixé nous donne

$$I(x) = \int_0^{cg(x)} h(x, y) \, dy = \int_{-\infty}^{+\infty} h(x, y) \mathbf{1}_{[0, cg(x)]}(y) \, dy = \int_{-\infty}^{+\infty} h(x, y) \mathbf{1}_H(x, y) \, dy. \quad (2.23)$$

Remarquons que si $g(x) = 0$, $[0, cg(x)] = \{0\}$ et $\int_{-\infty}^{+\infty} h(x, y) \mathbf{1}_{[0, cg(x)]}(y) \, dy = 0$. On peut donc retenir la formule (2.23) comme l'expression générale de $I(x)$, y compris lorsque $g(x) = 0$. Par report de cette expression dans (2.22), on obtient

$$\mathbf{E}h(M) = \int_{\mathbb{R}^2} h(x, y) \frac{1}{c} \mathbf{1}_H(x, y) \, dx \, dy = \int_{\mathbb{R}^2} h(x, y) \frac{1}{\lambda_2(H)} \mathbf{1}_H(x, y) \, dx \, dy.$$

Cette formule étant vraie pour toute $h : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ continue bornée, on en déduit que le vecteur aléatoire M suit la loi de densité $\lambda_2(H)^{-1} \mathbf{1}_H$, autrement dit que M suit la loi uniforme sur H . \square

Exemple 2.14 (lois Gamma). La loi Gamma de paramètre $a > 0$ est définie par sa densité

$$f(t) = \frac{1}{\Gamma(a)} t^{a-1} e^{-t} \mathbf{1}_{]0, +\infty[}(t).$$

Nous examinons le cas $0 < a < 1$. Le cas $a = 1$ est celui de la loi exponentielle et pour $a \geq 1$, nous renvoyons à [3] où plusieurs algorithmes sont proposés. L'intérêt de ce cas est de fournir un exemple où les densités f et g ne sont pas bornées. Pour contrôler l'explosion de f en zéro, il nous faut utiliser une densité g tendant vers $+\infty$ au moins aussi vite. Une bonne candidate est la densité g de la loi de Weibull de paramètre a , qu'il est facile de simuler, cf. exemple 2.6. En effet en dérivant la f.d.r. $F = 1 - G_a$, on trouve

$$g(t) = at^{a-1} \exp(-t^a) \mathbf{1}_{]0, +\infty[}(t).$$

9. Il est facile de voir que la loi d'un vecteur aléatoire de \mathbb{R}^2 est caractérisée par la famille de ces moments fonctionnels, par exemple en montrant que l'indicatrice d'un pavé ouvert $]a, b[\times]c, d[$ est limite d'une suite croissante de telles fonctions (exercice).

On vérifie par un simple calcul de maximum que pour tout $t > 0$,

$$\frac{f(t)}{g(t)} = \frac{\exp(t^a - t)}{a\Gamma(a)} \leq c := \frac{\exp(b(1-a))}{\Gamma(a+1)}, \quad \text{où } b = a^{a/(1-a)}.$$

Voici une implémentation en Scilab de l'algorithme du rejet pour simuler une loi Gamma de paramètre $a \in]0, 1[$. Pour des raisons de lisibilité, le code a été scindé en trois fonctions regroupées dans le même fichier.

```
function [y]=gamdns(a,t)
//
// densité de la loi Gamma(a)
//
y=exp(-t).*t.^(a-1)./gamma(a);
endfunction

function [y]=weibdns(a,t)
//
// densité de la loi de Weibull de paramètre a
//
y=a.*t.^(a-1).*exp(-t.^a);
endfunction

function [Z]=simgamm(a)
//
// simule une variable de loi Gamma(a) pour 0<a<1
// par la méthode du rejet en utilisant la densité de Weib(a)
//
if a>1 then error("erreur, le paramètre doit être <1"); end
// Calcul de la constante de rejet
b=a.^(a./(1-a)); c=exp(b.*(1-a))./gamma(1+a);
test = %T;
rand('uniform');
while test do
    U=rand();V= rand();
    X = (-log(V)).^(1/a) ; // simulation d'une v.a. de loi Weib(a)
    test =( c.*U.*weibdns(a,X) > gamdns(a,X) );
end
Z = X;
endfunction
```

2.4.3 Simulation d'une loi discrète par rejet

La portée de la méthode du rejet ne se limite pas aux lois à densité. On peut aussi appliquer les idées exposées ci-dessus à la simulation de lois discrètes. Pour ne pas

alourdir les notations, nous nous limiterons au cas de lois discrètes dont l'ensemble des valeurs possibles est inclus dans \mathbb{N} .

On se propose de simuler une variable aléatoire discrète Z dont la loi est donnée par l'ensemble $Z(\Omega) \subset \mathbb{N}$ et la fonction $f : \mathbb{N} \rightarrow [0, 1]$, $k \mapsto f(k) = p_k = P(Z = k)$. On suppose pour cela que l'on sait simuler une variable aléatoire discrète X de loi donnée par $X(\Omega) \subset \mathbb{N}$ et $g : \mathbb{N} \rightarrow [0, 1]$, $k \mapsto g(k) = q_k = P(X = k)$ et que l'on connaît une constante $c > 1$ telle que $f \leq cg$ (ceci suppose que $Z(\Omega) \subset X(\Omega)$).

Proposition 2.15. *Avec les notations ci-dessus, soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires discrètes indépendantes de même loi que X et $(U_i)_{i \geq 1}$ une suite i.i.d. de variables aléatoires uniformes sur $[0, 1]$, les deux suites étant indépendantes. On note $(M_i)_{i \geq 1}$ la suite de vecteurs aléatoires*

$$M_i := (X_i, cg(X_i)U_i), \quad i \in \mathbb{N}^*$$

et on définit T par

$$T(\omega) := \inf\{i \in \mathbb{N}^*; cg(X_i)U_i \leq f(X_i)\},$$

avec la convention $\inf \emptyset = +\infty$. On définit $X_T : \Omega \rightarrow \mathbb{R}^d$ par

$$X_T(\omega) := \begin{cases} X_{T(\omega)}(\omega) & \text{si } T(\omega) < +\infty, \\ 0 & \text{si } T(\omega) = +\infty. \end{cases}$$

Alors X_T est une v.a. discrète de même loi que Z : $P(X_T = k) = f(k)$ pour tout $k \in \mathbb{N}$.

Preuve. Notons $M := (X, cg(X)U)$ avec U uniforme sur $[0, 1]$ et indépendante de X et définissons M_T sur le même modèle que X_T ci-dessus. Avant de chercher la loi de M_T , intéressons nous à la loi μ de M . Pour la caractériser, il suffit de connaître¹⁰ $\mu(\{k\} \times [0, y])$ pour tout $k \in \mathbb{N}$ et tout $y \in \mathbb{R}_+$. Cette connaissance va découler du calcul suivant où l'on suppose $g(k) > 0$, le cas $g(k) = 0$ étant trivial puisque $\mu(\{k\} \times [0, y]) \leq \mu(\{k\} \times \mathbb{R}_+) = P(X = k) = g(k)$.

$$\begin{aligned} P(M \in \{k\} \times [0, y]) &= P(X = k \text{ et } cg(X)U \leq y) \\ &= P\left(X = k \text{ et } U \leq \frac{y}{cg(k)}\right) \\ &= P(X = k)P\left(U \leq \frac{y}{cg(k)}\right) \quad (X \text{ et } U \text{ indépendantes}) \\ &= g(k) \min\left(\frac{y}{cg(k)}, 1\right). \end{aligned} \tag{2.24}$$

Considérons le borélien $B := \{(k, y) \in \mathbb{N} \times \mathbb{R}_+; 0 \leq y \leq f(k)\}$. Il est facile de voir que $\mu(B)$ n'est pas nul. En effet grâce à (2.24) et en rappelant que $f \leq cg$ et $c > 1$, on obtient

$$\mu(B) = P(M \in B) = \sum_{k \in \mathbb{N}} g(k) \min\left(\frac{f(k)}{cg(k)}, 1\right) = \sum_{k \in \mathbb{N}} \frac{f(k)}{c} = \frac{1}{c}. \tag{2.25}$$

10. Vérifiez le en montrant par exemple, que pour tout pavé $C =]a, b] \times]a', b']$, la connaissance des $\mu(\{k\} \times [0, y])$ détermine la valeur de $\mu(C)$.

Ainsi les conditions d'application de la proposition 2.11 sont satisfaites¹¹. On sait alors que M_T a pour loi $\mu(\cdot \mid B)$. Ceci nous permet de calculer la loi de X_T .

$$\begin{aligned} P(X_T = k) &= P(M_T \in \{k\} \times \mathbb{R}_+) = P(M \in \{k\} \times \mathbb{R}_+ \mid M \in B) \\ &= \frac{P(M \in (\{k\} \times \mathbb{R}_+) \cap B)}{\mu(B)} \\ &= cP(M \in \{k\} \times [0, f(k)]) \\ &= cg(k) \min\left(\frac{f(k)}{cg(k)}, 1\right) \\ &= f(k). \end{aligned}$$

La variable aléatoire discrète X_T a donc même loi que Z . □

Exemple 2.16 (lois de Zipf). Ces lois utilisées en linguistique et en sciences sociales sont construites à partir de la fonction ζ de Riemann $\zeta(a) = \sum_{k \geq 1} k^{-a}$ en posant

$$p_k = \frac{1}{\zeta(a)k^a}, \quad k \in \mathbb{N}^*.$$

Examinons d'abord le cas où $a = 2$. La valeur de $\zeta(2)$ étant bien connue, nous avons donc ici

$$p_k = f(k) = \frac{6}{\pi^2 k^2}, \quad k \in \mathbb{N}^*$$

et nous proposons de prendre

$$q_k = g(k) = \frac{1}{k(k+1)}, \quad k \in \mathbb{N}^*.$$

La raison de ce choix est que la loi correspondante est facile à simuler puisque c'est celle de $X := [U^{-1}]$ où U est uniforme sur $[0, 1]$: en effet

$$P(X = k) = P(k \leq U^{-1} < k+1) = P\left(\frac{1}{k+1} < U \leq \frac{1}{k}\right) = \frac{1}{k} - \frac{1}{k+1} = q_k.$$

La meilleure constante c possible est

$$c = \sup_{k \geq 1} \frac{p_k}{q_k} = \frac{6}{\pi^2} \sup_{k \geq 1} \frac{k+1}{k} = \frac{12}{\pi^2}.$$

Dans le cas général où le paramètre a est un réel quelconque de $]1, +\infty[$, on s'inspire de la même idée en prenant $X = [U^{-1/(a-1)}]$, voir [3, pp. 550–552] pour les détails.

Exemple 2.17 (loi du nombre de coïncidences ou *matching distribution*). Soit Z le nombre de points fixes d'une permutation aléatoire sur $\{1, \dots, n\}$, choisie selon la loi uniforme sur l'ensemble de toutes les permutations de $\{1, \dots, n\}$ (voir par exemple

11. Même si cette proposition n'a été appliquée jusqu'ici qu'avec μ loi uniforme sur un borélien, elle est valide pour n'importe quelle loi de probabilité μ , revoyez la preuve si vous en doutez.

l'exercice 1.12 *Le problème des appariements* dans [6]). On peut vérifier que la loi μ_n de Z est donnée par

$$P(Z = k) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}, \quad 0 \leq k \leq n.$$

On voit immédiatement que $P(Z = k) \leq 1/k!$, on peut alors prendre pour loi de X la loi de Poisson de paramètre 1 et pour constante $c = e$. Le nombre d'itérations T suit la loi géométrique de paramètre $\mu_n(B_n)$ donc d'espérance $1/\mu_n(B_n)$ et d'après (2.25), $\mu_n(B_n) = 1/c = e^{-1}$. En moyenne on génère e variables de loi $\text{Pois}(1)$ pour simuler Z et ce quelle que soit la valeur de n . Le coût global de l'algorithme du rejet est donc ici en $O(n)$ puisque pour chacune des variables de Poisson générée, on doit calculer $f(k)$ (addition de $n - k + 1$ termes¹²).

Ceci explique que l'algorithme du rejet soit concurrentiel avec l'algorithme plus simple qui consiste à générer un n -échantillon de la loi uniforme sur $[0, 1]$, à le trier par ordre croissant et à compter le nombre de valeurs dont l'indice après le tri est resté le même. Comme le coût des tris les plus performants est en $O(n \ln n)$, on voit que la méthode du rejet est plus économique, au moins pour les grandes valeurs de n .

2.5 Simulation de vecteurs aléatoires par transformation

Supposons que l'on sache simuler un vecteur aléatoire M de \mathbb{R}^d ayant une certaine loi $\mu = P_M$. Si h est une transformation mesurable, on a *ispso facto* un moyen de simuler un vecteur de loi $\nu = \mu \circ h^{-1}$, il suffit de prendre le vecteur $h(M)$. Bien sûr, en pratique, c'est le problème inverse qui se pose : étant donnée ν , trouver μ simulable et h telles que $\nu = \mu \circ h^{-1}$. Nous allons examiner plusieurs situations où cette méthode est pertinente, en commençant par le cas important des lois uniformes.

2.5.1 Loi uniforme par transformation affine

Principe

La simulation de lois uniformes par transformation affine repose sur le résultat suivant.

Lemme 2.18. *Soient h une application affine bijective $\mathbb{R}^d \rightarrow \mathbb{R}^d$ et M un vecteur aléatoire de loi uniforme sur un borélien B de \mathbb{R}^d . Alors $h(M)$ suit la loi uniforme sur $h(B)$.*

12. En programmant de manière économique le calcul des factorielles, *i.e.* en utilisant une seule multiplication pour passer de $j!$ à $(j+1)!$

Preuve. Commençons par chercher la mesure image de λ_d par h . On peut écrire $h = \varphi \circ \tau$, où τ est une translation et φ une bijection linéaire de $\mathbb{R}^d \rightarrow \mathbb{R}^d$. On peut montrer¹³ que

$$\lambda_d \circ \varphi^{-1} = |\det(\varphi^{-1})| \lambda_d.$$

Ceci est lié à la formule de changement de variable linéaire bijectif que nous avons admise en cours d'I.P.É. La mesure de Lebesgue étant invariante par translations, on en déduit que

$$\lambda_d \circ h^{-1} = (\lambda_d \circ \tau^{-1}) \circ \varphi^{-1} = \lambda_d \circ \varphi^{-1} = |\det(\varphi^{-1})| \lambda_d = c \lambda_d,$$

la constante c ne dépendant que de h .

Soit A' un borélien quelconque de \mathbb{R}^d . Puisque M suit la loi uniforme sur B , on a :

$$P(h(M) \in A') = P(M \in h^{-1}(A')) = \frac{\lambda_d(h^{-1}(A') \cap B)}{\lambda_d(B)} \quad (2.26)$$

Posons $B' := h(B)$. Comme h est bijective, on a $B = h^{-1}(B')$ (vérifiez !). On peut alors écrire $h^{-1}(A') \cap B = h^{-1}(A') \cap h^{-1}(B') = h^{-1}(A' \cap B')$. En reportant dans (2.26), on obtient

$$P(h(M) \in A') = \frac{(\lambda_d \circ h^{-1})(A' \cap B')}{(\lambda_d \circ h^{-1})(B')} = \frac{c \lambda_d(A' \cap B')}{c \lambda_d(B')} = \frac{\lambda_d(A' \cap B')}{\lambda_d(B')},$$

ce qui montre que $h(M)$ suit la loi uniforme sur B' , puisque A' était quelconque. \square

Loi uniforme sur un parallélogramme

Pour simuler une loi uniforme sur un parallélogramme $ABCD$ ($\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{AD}$), on part d'un vecteur $M = (U_1, U_2)$ de loi uniforme sur $[0, 1]^2$. On détermine l'application linéaire φ par $\varphi(1, 0) = \overrightarrow{AB}$ et $\varphi(0, 1) = \overrightarrow{AD}$. On prend pour τ la translation de vecteur \overrightarrow{OA} . L'image du carré $[0, 1]^2$ par la transformation affine $h = \tau \circ \varphi$ est le parallélogramme $ABCD$ (figure 2.4) et le vecteur aléatoire $M' = h(M)$ suit la loi uniforme sur $ABCD$.

On détermine les coordonnées V_1, V_2 de M' matriciellement par :

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} x_B - x_A & x_D - x_A \\ y_B - y_A & y_D - y_A \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \end{bmatrix} + \begin{bmatrix} x_A \\ y_A \end{bmatrix}.$$

Loi uniforme sur un triangle ou un polygone

Pour simuler un vecteur aléatoire de loi uniforme sur un triangle donné $T = ABC$, on part d'un vecteur de loi uniforme sur un triangle T_0 construit sur les vecteurs unitaires de la base canonique de \mathbb{R}^2 et on détermine une transformation affine h qui transforme T_0 en T . Notons $O = (0, 0)$, $I = (1, 0)$, $J = (0, 1)$ et $K = (1, 1)$. Soit U_1 et U_2 deux variables

13. Voir par exemple le cours d'I.F.P. 2004, chapitre 5
<http://math.univ-lille1.fr/~suquet/ens/IFP/Cours/cours04/Chap5ifp04.pdf>

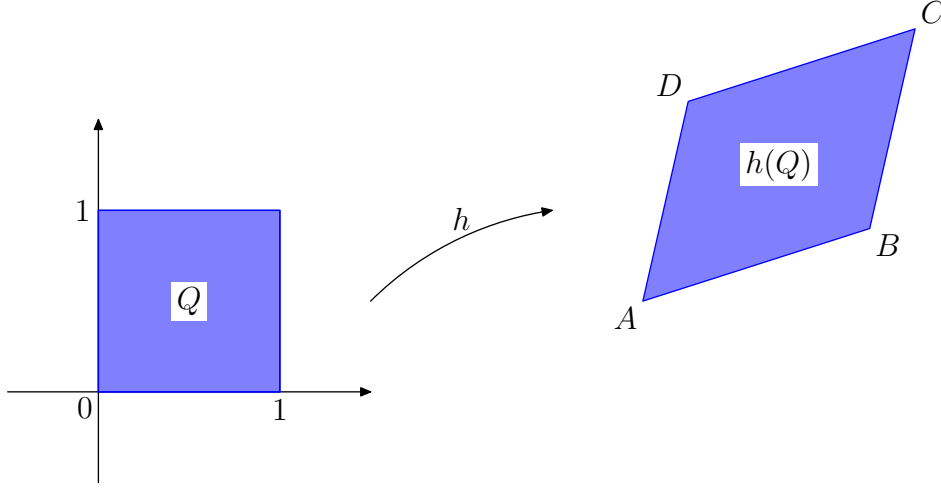


FIG. 2.4 – Parallélogramme image du carré unité par transformation affine

aléatoires indépendantes et de même loi uniforme sur $[0, 1]$. Définissons les statistiques d'ordre du vecteur (U_1, U_2) par

$$U_{2:1} = \min(U_1, U_2), \quad U_{2:2} = \max(U_1, U_2).$$

Selon un exercice classique auquel vous n'avez probablement pas échappé, $(U_{2:1}, U_{2:2})$ suit la loi uniforme sur le triangle OKJ et $(U_{2:2}, U_{2:1})$ la loi uniforme sur OIK . On peut donc choisir l'un de ces deux triangles pour T_0 . Si on prend OKJ , on détermine alors h en écrivant que $h(O) = A$, $h(J) = B$ et $h(K) = C$ (ceci détermine l'une des 6 applications affines transformant OKJ en ABC).

En fait pour simplifier la détermination de h (et même pour l'éviter), il est préférable de choisir pour T_0 le triangle OIJ . L'astuce est d'écrire T_0 comme l'enveloppe convexe de ses sommets :

$$T_0 = OIJ = \{M = p_1O + p_2I + p_3J; p_1, p_2, p_3 \geq 0, p_1 + p_2 + p_3 = 1\}.$$

En raison de la conservation du barycentre par transformation affine, si $M \in T_0$ a pour coordonnées barycentriques (p_1, p_2, p_3) relativement à OIJ et si h est l'unique transformation affine telle que $h(O) = A$, $h(I) = B$ et $h(J) = C$, le point $h(M)$ a mêmes coordonnées barycentriques (p_1, p_2, p_3) , mais relativement à ABC . Il suffit donc de savoir choisir (p_1, p_2, p_3) aléatoires de telle sorte que M suive la loi uniforme sur T_0 . Pour cela on prend les *espacements* associés au vecteur (U_1, U_2) , c'est à dire les longueurs des trois segments de la subdivision de $[0, 1]$ générée par U_1 et U_2 :

$$p_1(\omega) = U_{2:1}(\omega), \quad p_2(\omega) = U_{2:2}(\omega) - U_{2:1}(\omega), \quad p_3(\omega) = 1 - U_{2:2}(\omega).$$

On laisse en exercice¹⁴ la vérification du fait que la loi de

$$M = p_1O + p_2I + p_3J = (U_{2:2} - U_{2:1}, 1 - U_{2:2})$$

14. La proposition 2.19 ci-après donne une solution dans un cadre plus général.

est bien la loi uniforme sur OIJ . Nous pouvons maintenant proposer l'algorithme suivant pour la simulation d'un vecteur aléatoire $h(M)$ de loi uniforme sur ABC .

1. Génération de U_1 et U_2 indépendants de loi uniforme sur $[0, 1]$.
2. Tri de (U_1, U_2) et calcul des espacements $U_{2:1}$, $U_{2:2} - U_{2:1}$ et $1 - U_{2:2}$.
3. Calcul de $h(M) = U_{2:1}A + (U_{2:2} - U_{2:1})B + (1 - U_{2:2})C$.

On peut immédiatement réinvestir cette technique pour la simulation d'un vecteur aléatoire M de loi uniforme sur un polygone Q du plan (pas forcément convexe). On commence par *triangler* Q , autrement dit à le découper en un nombre fini de triangles T_1, \dots, T_n (on a donc $Q = \cup_{1 \leq i \leq n} T_i$ et si $i \neq j$, $\lambda_2(T_i \cap T_j) = 0$). Une fois cette triangulation faite¹⁵, on utilise l'algorithme suivant.

1. Calculer pour $i = 1, \dots, n$ les $a_i := \lambda_2(T_i)$ et $a := \lambda_2(Q) = a_1 + \dots + a_n$.
2. Générer une variable discrète X de loi sur $\{1, \dots, n\}$ donnée par $P(X = k) = a_k/a$.
3. Pour $k = X(\omega)$ ainsi obtenu, générer indépendamment de X un vecteur aléatoire M de loi uniforme sur le triangle T_k .

La justification de cet algorithme est laissée en exercice et sa traduction en Scilab est un bon sujet de T.P. Un rappel peut-être pas superflu : l'aire d'un triangle ABC peut se calculer par la formule

$$\lambda_2(ABC) = \frac{1}{2} |\det(\overrightarrow{AB}, \overrightarrow{AC})|.$$

Loi uniforme sur un simplexe de \mathbb{R}^d

La simulation par transformation affine de la loi uniforme sur un triangle se généralise à la dimension d au cas du *simplexe*. Un simplexe de \mathbb{R}^d est l'enveloppe convexe de $d + 1$ points A_0, A_1, \dots, A_d « en position générale », *i.e.* si $d = 2$ les trois points ne sont pas alignés, si $d = 3$ les quatre points ne sont pas coplanaires et pour d quelconque A_1, \dots, A_{d+1} n'appartiennent pas à un même hyperplan de \mathbb{R}^d . Un simplexe de \mathbb{R}^3 est un *tétraèdre*. Soit donc T un simplexe de sommets A_0, \dots, A_d et T_0 le simplexe de sommets I_0, I_1, \dots, I_d , où $I_0 = O$ et pour $j = 1, \dots, d$, I_j est le point défini par $\overrightarrow{OI_j} = e_j$, j^e vecteur de la base canonique de \mathbb{R}^d . On voit immédiatement que si (p_0, p_1, \dots, p_d) sont les coordonnées barycentriques de M dans le simplexe T_0 , ses coordonnées cartésiennes usuelles sont (p_1, \dots, p_d) .

$$T_0 = \left\{ M = \sum_{j=0}^d p_j I_j; \forall j, p_j \geq 0, \sum_{j=0}^d p_j = 1 \right\} = \left\{ (p_1, \dots, p_d) \in \mathbb{R}_+^d; \sum_{j=1}^d p_j \leq 1 \right\}.$$

D'autre part, si h est l'unique application affine telle que $H(I_j) = A_j$ pour $j = 0, 1, \dots, d$, alors $h(T_0) = T$ et les coordonnées barycentriques de $h(M)$ dans T sont les

15. Il semble plus facile de la faire « à la main » que de la programmer dans le cas le plus général.

mêmes que celles de M dans T_0 . Il suffit donc de savoir choisir p_1, \dots, p_d aléatoires de telle sorte que $M = (p_1, \dots, p_d)$ suive la loi uniforme sur T_0 pour que

$$h(M) = p_0 A_0 + p_1 A_1 + \dots + p_d A_d$$

suive la loi uniforme sur $T = h(T_0)$. Bien sûr dans cette formule, p_0 est donné par $p_0 + (p_1 + \dots + p_d) = 1$.

Comme en dimension 2, nous aurons recours aux *espacements* d'un d -échantillon uniforme sur $[0, 1]$ pour choisir les p_j . Soit (U_1, \dots, U_d) un échantillon de variables aléatoires indépendantes de même loi uniforme sur $[0, 1]$. Notons $(U_{d:1}, \dots, U_{d:d})$ le vecteur des statistiques d'ordre obtenu par réarrangement croissant de l'échantillon :

$$\{U_1, \dots, U_d\} = \{U_{d:1}, \dots, U_{d:d}\} \quad \text{et} \quad U_{d:1} \leq U_{d:2} \leq \dots \leq U_{d:d}.$$

Posons $U_{d:0} = 0$ et $U_{d:d+1} = 1$. Définissons les espacements S_j associés par

$$S_j = U_{d:j+1} - U_{d:j}, \quad j = 0, \dots, d.$$

Dit autrement et sans formules, les espacements S_0, S_1, \dots, S_d sont les longueurs des $d+1$ intervalles découpés dans $[0, 1]$ par les nombres U_1, \dots, U_d , ces segments étant numérotés de gauche à droite.

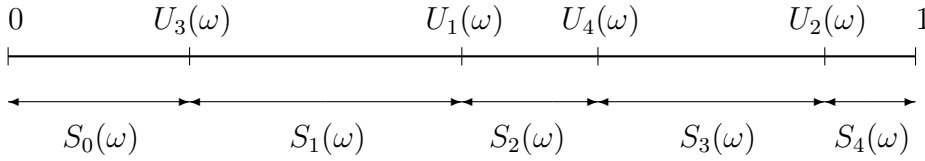


FIG. 2.5 – Espacements d'un 4-échantillon

Proposition 2.19. *Soit T un simplexe de \mathbb{R}^d de sommets A_0, \dots, A_d . Soit (U_1, \dots, U_d) un échantillon de variables aléatoires indépendantes de même loi uniforme sur $[0, 1]$ et S_j , $j = 0, 1, \dots, d$ les espacements associés. Alors le point aléatoire*

$$\sum_{j=0}^d S_j A_j$$

suit la loi uniforme sur T .

Preuve. Compte tenu des explications données ci-dessus, la seule chose à vérifier est que le vecteur aléatoire $S = (S_1, \dots, S_d)$ suit la loi uniforme sur T_0 . Pour cela notons g une application continue à support compact $\mathbb{R}^d \rightarrow \mathbb{R}_+$ arbitraire et calculons $\mathbf{E}g(S)$. Avec des notations similaires à celles des statistiques d'ordre,

$$\mathbf{E}g(S) = \int_{[0,1]^d} g(u_{d:2} - u_{d:1}, \dots, u_{d:d+1} - u_{d:d}) du_1 \dots du_d.$$

Découpons $[0, 1]^d$ en les $d!$ simplexes qui se déduisent par permutation de coordonnées de

$$\Delta := \{(u_1, \dots, u_d) \in [0, 1]^d; 0 \leq u_1 \leq u_2 \leq \dots \leq u_d \leq 1\},$$

Les frontières de ces simplexes sont incluses dans une réunion finie d'hyperplans d'équation $x_i = x_j$, donc de mesure nulle. Des changements de variables par permutation sur les coordonnées envoient tous ces simplexes sur Δ et sont linéaires bijectifs de déterminant 1 en valeur absolue. On en déduit que

$$\mathbf{E}g(S) = d! \int_{\Delta} g(u_2 - u_1, u_3 - u_2, \dots, u_d - u_{d-1}, 1 - u_d) du_1 \dots du_d.$$

Effectuons maintenant le changement de variable $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}^d, (u_1, \dots, u_d) \mapsto (t_1, \dots, t_d)$ donné par

$$t_1 = u_2 - u_1, t_2 = u_3 - u_2, \dots, t_{d-1} = u_d - u_{d-1}, t_d = 1 - u_d.$$

On vérifie facilement que φ est une bijection affine d'inverse donné par les égalités

$$u_j = 1 - \sum_{i=j}^d t_i, \quad j = 1, \dots, d.$$

Pour déterminer $\varphi(\Delta)$, on remarque en remplaçant les u_j par leur expression en fonction des t_i dans les inégalités définissant Δ , que $(t_1, \dots, t_d) \in \varphi(\Delta)$ si et seulement si

$$0 \leq 1 - (t_1 + \dots + t_d) \leq 1 - (t_2 + \dots + t_d) \leq \dots \leq 1 - t_d \leq 1.$$

Ceci équivaut à

$$t_1 + \dots + t_d \leq 1 \quad \text{et} \quad \forall j = 1, \dots, d, \quad t_j \geq 0.$$

Ainsi $\varphi(\Delta) = T_0$. D'autre part il est clair que le déterminant de l'application linéaire associée à la transformation affine φ vaut 1 en valeur absolue. Le changement de variable dans l'intégrale \int_{Δ} donne finalement

$$\begin{aligned} \mathbf{E}g(S) &= d! \int_{T_0} g(t_1, \dots, t_d) dt_1, \dots, dt_d \\ &= \int_{\mathbb{R}^d} g(t_1, \dots, t_d) d! \mathbf{1}_{T_0}(t_1, \dots, t_d) dt_1, \dots, dt_d. \end{aligned} \quad (2.27)$$

Soit maintenant S' un vecteur aléatoire de loi uniforme sur T_0 . Sa densité est $d! \mathbf{1}_{T_0}$. En effet on peut vérifier que

$$\lambda_d(T_0) = \int_{T_0} dt_1, \dots, dt_d = \int_{\Delta} |\det \varphi^{-1}| dt_1, \dots, dt_d = \int_{\Delta} dt_1, \dots, dt_d = \lambda_d(\Delta)$$

et $\lambda_d(\Delta) = 1/d!$, voir la justification de l'égalité $\lambda_n(B_n) = 1/n!$ dans la preuve du lemme 2.7. On a alors

$$\mathbf{E}g(S') = \int_{\mathbb{R}^d} g(t_1, \dots, t_d) d! \mathbf{1}_{T_0}(t_1, \dots, t_d) dt_1, \dots, dt_d = \mathbf{E}g(S)$$

et ceci étant vrai pour toute g positive continue à support compact, on en déduit que S et S' ont même loi, donc que S suit la loi uniforme sur T_0 . \square

Loi uniforme sur un ellipsoïde

Un ellipsoïde E de \mathbb{R}^d peut toujours être vu comme image par une transformation affine bijective h de la boule unité euclidienne

$$B_d := \{(x_1, \dots, x_d) \in \mathbb{R}^d; x_1^2 + \dots + x_d^2 \leq 1\}.$$

Supposons résolu le problème de la simulation de la loi uniforme sur B_d , on sait déjà le faire au moins par rejet à partir de la loi uniforme sur l'hypercube $[-1, 1]^d$, même si cette méthode n'est pas très performante en dimension élevée¹⁶. La simulation de la loi uniforme sur E se réduit alors à la détermination de h affine telle que $h(B_d) = E$.

Supposons d'abord que l'on connaisse le centre C de E et ses demi-axes $\overrightarrow{CA_1}, \dots, \overrightarrow{CA_d}$ (vecteurs orthogonaux). Il suffit de générer un vecteur colonne X de loi uniforme sur B_d et de calculer matriciellement $Y = C + HX$, où H est la matrice ayant pour colonnes les vecteurs $\overrightarrow{CA_i}$ écrits dans la base canonique de \mathbb{R}^d . Le vecteur colonne Y ainsi obtenu suit la loi uniforme sur E .

L'ellipsoïde E peut aussi être donné par son inéquation cartésienne qui après réduction peut s'écrire sous la forme matricielle, en notant la transposition par un prime :

$$E = \{Y \in \mathbb{R}^d; (Y - C)'Q(Y - C) \leq 1\},$$

où Y et C sont des vecteurs colonnes (non aléatoires!), C désignant les coordonnées du centre de E et Q est une matrice symétrique définie positive. On cherche alors une matrice régulière H ($\det H \neq 0$) telle que l'application $X \mapsto C + HX$ transforme B_d en E . Ceci se produit si l'inéquation $(Y - C)'Q(Y - C) \leq 1$ est équivalente à $X'X \leq 1$ (inéquation de B_d). Or

$$(Y - C)'Q(Y - C) = X'H'QH X,$$

il suffit donc de trouver H telle que $H'QH = I$, matrice identité. Ceci revient à trouver H telle que $Q = H'^{-1}H^{-1}$ ou encore $Q = R'R$, en posant $R = H^{-1}$. L'algorithme de Cholesky [5, p. 95] permet de trouver une matrice triangulaire supérieure R ayant cette propriété. Il est implémenté en Scilab (fonction `chol`).

2.5.2 Vecteur gaussien de covariance donnée

Nous savons déjà simuler une suite de variables aléatoires gaussiennes indépendantes (cf. lemme 2.9). Ceci résout le problème de la simulation d'un vecteur aléatoire gaussien à composantes indépendantes (donc à matrice de covariance diagonale). Il nous reste à étudier le cas général du vecteur gaussien ayant une matrice de covariance Q quelconque, c'est-à-dire dans ce contexte, symétrique semi-définie positive. La méthode utilisée repose sur le résultat suivant (voir par exemple [4, Th. 16.2, p. 135]).

16. Voir <http://math.univ-lille1.fr/~suquet/ens/Agr/simul06.pdf>

Proposition 2.20. Soit X un vecteur (colonne) gaussien de \mathbb{R}^d , de vecteur moyenne μ . Il existe des variables aléatoires réelles indépendantes Y_1, \dots, Y_d , chaque Y_j ayant une loi $\mathfrak{N}(0, \sigma_j)$ avec $\sigma_j \geq 0$ et une matrice $d \times d$ orthogonale A telles que $X = \mu + AY$, en notant $Y = (Y_1, \dots, Y_d)'$.

Notons que certains des σ_j peuvent être nuls et rappelons que si $\sigma = 0$, la loi gaussienne $\mathfrak{N}(m, \sigma)$ est la masse de Dirac au point m , autrement dit, la loi de la v.a. constante m . La démonstration de la proposition 2.20 va nous indiquer l'algorithme pour simuler X .

Preuve. La matrice de covariance Q étant symétrique semi-définie positive, on sait qu'on peut l'écrire

$$Q = ASA', \quad (2.28)$$

où S est une matrice diagonale à termes diagonaux σ_j positifs et A une matrice orthogonale (donc vérifiant $A' = A^{-1}$). On pose alors $Y = A'(X - \mu)$. Toute combinaison linéaire des composantes de Y peut s'écrire comme une combinaison linéaire des composantes de X plus une constante, donc est une v.a. gaussienne, puisque X est un vecteur gaussien par hypothèse. Ainsi Y est gaussien. De plus comme $A' = A^{-1}$, $X = \mu + AY$. Il nous reste à vérifier que la matrice de covariance de Y est S (ce qui entraîne l'indépendance des Y_j). Ceci résulte du calcul suivant :

$$\mathbf{E}(YY') = \mathbf{E}(A'(X - \mu)(X - \mu)'A) = A'\mathbf{E}((X - \mu)(X - \mu)')A \quad (2.29)$$

$$\begin{aligned} &= A'QA \\ &= A^{-1}(ASA^{-1})A \\ &= S. \end{aligned} \quad (2.30)$$

L'égalité (2.29) repose sur la propriété suivante. Soit Z une matrice aléatoire $d \times d$ dont toutes les composantes ont une espérance et B une matrice $d \times d$ non aléatoire. Alors $\mathbf{E}(BZ) = B\mathbf{E}Z$ et $\mathbf{E}(ZB) = (\mathbf{E}Z)B$. La vérification est immédiate en écrivant le produit des matrices et en utilisant la linéarité de l'espérance. L'égalité (2.30) exploite (2.28) et l'égalité entre la transposée de A et son inverse, due à l'orthogonalité de A . \square

Ce qui précède suggère l'algorithme suivant pour simuler le vecteur gaussien X de vecteur moyenne μ et de matrice de covariance Q .

1. Calculer la décomposition (2.28) pour Q .
2. Générer le vecteur colonne Y en simulant des gaussiennes $\mathfrak{N}(0, \sigma_j)$ pour les indices j tels que $\sigma_j > 0$ et compléter les autres composantes de Y par des zéros.
3. Calculer $X = \mu + AY$.

Pour réaliser pratiquement la première étape, Scilab nous fournit la fonction `svd` (*singular value decomposition*). Voici un extrait de la documentation en ligne (les noms des paramètres ont été modifiés pour éviter la confusion avec les notations ci-dessus) :

`[A,S,B]=svd(Q)`

DESCRIPTION

produces a diagonal matrix S , of the same dimension as Q and with nonnegative diagonal elements in decreasing order, and unitary matrices A and B so that $X = A*S*B'$.

En appliquant `svd` à une matrice Q symétrique semi-définie positive, on obtient $A = B$ matrice orthogonale et la diagonale de S est constituée de réels positifs rangés par ordre décroissant. Si d_0 est l'indice du dernier d'entre eux non nul, on génère les Y_1, \dots, Y_{d_0} indépendantes de lois $\mathcal{N}(0, \sigma_j)$ et on prend $Y = (Y_1, \dots, Y_{d_0}, 0, \dots, 0)'$ (si $d_0 < d$).

Bibliographie

- [1] N. BOULEAU. *Probabilités de l'Ingénieur, variables aléatoires et simulation*, Hermann, 2002.
- [2] J.-P. DELAHAYE. Aléas du hasard informatique. *Revue Pour la Science*, n° 245, mars 1998, pp. 92–97.
- [3] L. DEVROYE. *Non-Uniform Random Variate Generation*. Springer, 1986.
- [4] J. JACOD et P. PROTTER. *L'essentiel en théorie des probabilités*. Cassini, 2003.
- [5] D. SERRE. *Les matrices, théorie et pratique*. Dunod, 2001.
- [6] Ch. SUQUET. *Introduction au Calcul des Probabilités*, polycopié de Deug MIAS et MASS, Université des Sciences et Technologies de Lille.
<http://math.univ-lille1.fr/~ipeis/Cours/ICP0106.pdf>

Chapitre 3

Échantillons et statistiques

Nous abordons maintenant la problématique statistique. En deux mots, il s'agit à partir de l'observation d'un *échantillon* provenant d'une loi inconnue, de deviner quelle est cette loi (problème d'*estimation*) ou plus généralement de prendre une décision en contrôlant au mieux le risque d'erreur (problème de *test*). Le but de ce chapitre est d'introduire un modèle permettant de « mathématiser » ces questions et de mettre en place un outillage basique autour de la notion d'échantillon. Le résultat important est le théorème de Glivenko-Cantelli, appelé quelquefois *théorème fondamental de la statistique*. Ce théorème permet de justifier l'idée intuitive que l'on peut reconstruire une loi inconnue à partir d'observations, avec une approximation d'autant meilleure que le nombre d'observations est plus grand. L'estimation et les tests font l'objet des deux chapitres suivants.

3.1 Modélisation statistique

En guise d'exemple introductif, considérons le problème d'urne suivant.

Exemple 3.1. On se trouve en présence d'une urne dont on sait seulement qu'elle contient des boules vertes en proportion inconnue $p \in [0, 1]$ et des boules rouges en proportion $1 - p$. On effectue n tirages d'une boule avec remise, en notant sa couleur. Que peut-on dire de p au vu de ces observations¹ ?

Définition 3.2 (modèle statistique). *On appelle modèle statistique la donnée d'un espace mesurable (Ω, \mathcal{F}) et d'une famille $(P_\theta, \theta \in \Theta)$ de mesures de probabilité sur (Ω, \mathcal{F}) . Θ est appelé ensemble des paramètres.*

1. Bien sûr cet exemple a surtout un intérêt métaphorique. Si on veut vraiment connaître avec exactitude la proportion de boules vertes, le plus simple est sans doute de renverser l'urne sur la table et de les compter ! Dans l'enseignement des probabilités les problèmes d'urnes sont en fait déjà une première forme de modélisation. Imaginez qu'au lieu d'une urne on ait ici un étang peuplé par deux espèces de poissons. Un tirage sans remise correspond alors à la pêche non destructive d'un poisson remis à l'eau après observation. La solution qui consisterait à assécher l'étang pour compter les poissons n'est en général pas acceptable...

Cette définition situe d'emblée la statistique mathématique dans le prolongement de la théorie des probabilités, mais il convient de noter une différence fondamentale. En théorie des probabilités, on travaille généralement avec *un seul* espace probabilisé (Ω, \mathcal{F}, P) de référence, réputé modéliser correctement une expérience aléatoire. Ω est l'ensemble des « événements élémentaires », \mathcal{F} est une tribu de parties de Ω appelées par commodité « événements » et P est une mesure de probabilité sur (Ω, \mathcal{F}) . Les notions importantes de loi d'une v.a., d'espérance, d'indépendance sont toujours relatives à cette mesure P , même si on omet généralement de le préciser. Dans le modèle statistique, on dispose de *plusieurs espaces probabilisés* $(\Omega, \mathcal{F}, P_\theta)$, éventuellement une infinité, on ignore lequel est « le bon », et on les met en concurrence au vu des observations. Voici une question de probabilités avec une urne du type de l'exemple 3.1 : je *sais* qu'il y a 17% de boules vertes dans l'urne, j'en tire 20 avec remise, quelle probabilité ai-je d'obtenir 5 fois la couleur verte ? Voici maintenant une question de statistique : ayant effectué 20 tirages avec remise, j'ai *observé* la sortie de 5 boules vertes, en quoi cette information modifie-t-elle mon degré d'ignorance sur la proportion inconnue de boules vertes dans l'urne ?

Regardons comment associer un modèle statistique à l'exemple 3.1. Ici le paramètre θ est égal à p , proportion de boules vertes dans l'urne. Si l'on connaît le nombre total N de boules dans l'urne, l'ensemble des paramètres est évidemment l'ensemble fini $\Theta = \{k/N; 0 \leq k \leq N\}$. Si l'on ne connaît pas N , on sait seulement que $p = k/N$, avec $N \in \mathbb{N}^*$ et $0 \leq k \leq N$, est un nombre rationnel compris entre 0 et 1 et on a un ensemble de paramètres infini $\Theta = [0, 1] \cap \mathbb{Q}$. Quel Ω peut-on prendre ? Supposons d'abord que l'on soit décidé à ne faire que n tirages. Comme on ne s'intéresse qu'à la couleur de la boule obtenue à chaque tirage, on peut prendre pour événements élémentaires les n -uplets $\omega = (u_1, \dots, u_n)$ avec u_i valant r ou v, pour $1 \leq i \leq n$. On choisit alors $\Omega = \Omega_n$ et $\mathcal{F} = \mathcal{F}_n$ donné par

$$\Omega_n := \{r, v\}^n, \quad \mathcal{F}_n := \mathcal{P}(\Omega_n). \quad (3.1)$$

Pour toute valeur θ de la proportion de boules vertes, on choisit la probabilité $P_\theta = P_{\theta,n}$ sur $(\Omega_n, \mathcal{F}_n)$ telle que les tirages soient P_θ mutuellement indépendants et que pour tout i la probabilité d'obtenir une boule verte au i^{e} tirage soit θ . Comme Ω_n est fini, il suffit de donner la valeur $P_\theta(\{\omega\})$ pour chaque événement élémentaire $\omega = (u_1, \dots, u_n) \in \Omega_n$. Au vu des conditions énoncées ci-dessus, il est clair que la seule solution est de prendre

$$P_{\theta,n}(\{\omega\}) = P_{\theta,n}(\{(u_1, \dots, u_n)\}) = \theta^j (1 - \theta)^{n-j}, \quad j = \text{card} \{i \leq n; u_i = v\}. \quad (3.2)$$

On peut alors définir sur $(\Omega_n, \mathcal{F}_n)$ la suite finie de variables aléatoires $(X_i)_{1 \leq i \leq n}$ par

$$\forall i = 1, \dots, n, \quad X_i(\omega) = X_i(u_1, \dots, u_n) = \begin{cases} 1 & \text{si } u_i = v, \\ 0 & \text{si } u_i = r. \end{cases} \quad (3.3)$$

Remarquons que cette définition des X_i ne fait intervenir aucune des mesures de probabilité² P_θ . Maintenant, il est facile de voir que pour *chaque valeur de la proportion* θ de

2. Ceci correspond au fait que l'on n'a pas besoin de connaître la proportion de boules vertes pour effectuer les n tirages sans remise et noter les observations.

boules vertes, les v.a. X_i sont P_θ -indépendantes et de même loi sous P_θ , à savoir la loi de Bernoulli de paramètre θ . La variable aléatoire $S_n := X_1 + \dots + X_n$ compte le nombre de boules vertes observées en n tirages. Sa loi sous P_θ est binomiale de paramètres (n, θ) . La fréquence observée de sortie des boules vertes est $M_n := n^{-1}S_n$. Elle suit sous P_θ une loi qui a les mêmes masses que $\text{Bin}(n, \theta)$, mais localisées sur les rationnels $0, 1/n, 2/n, \dots, 1$, au lieu des entiers $0, 1, 2, \dots, n$. En notant cette loi P_{θ, M_n} , on a

$$P_{\theta, M_n} = \sum_{i=0}^n C_n^i \theta^i (1 - \theta)^{n-i} \delta_{i/n}.$$

L'espérance sous P_θ de cette loi est $n^{-1}\mathbf{E}_\theta(S_n) = n^{-1}n\theta = \theta$. Quand n est grand, cette loi est bien concentrée dans un petit voisinage de θ , voir les diagrammes en bâtons³ de la figure 3.1 pour une illustration. Supposons que l'on ait réalisé 100 tirages et obtenu 13

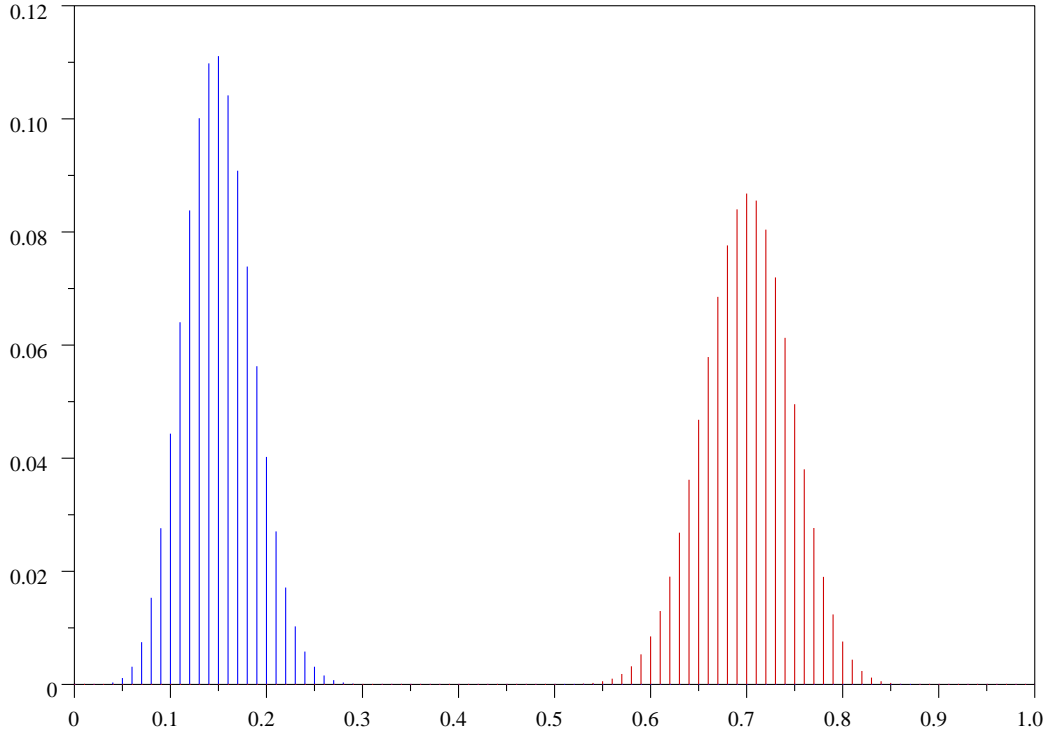


FIG. 3.1 – Loi de M_{100} sous $P_{0,15}$ et sous $P_{0,7}$

3. Pour représenter graphiquement la loi discrète $\mu = \sum p_i \delta_{x_i}$, on trace à partir de chaque point $(x_i, 0)$ un segment vertical de hauteur proportionnelle à p_i . La figure 3.1 affiche théoriquement 101 bâtons pour chacune des deux lois de M_{100} , sous $P_{0,15}$ et sous $P_{0,7}$. En pratique seuls sont visibles ceux dont la hauteur est supérieure à l'épaisseur d'un trait d'imprimante.

boules vertes. La *valeur observée* $M_{100}(\omega)$ est donc ici 0,13. Si on met en concurrence les lois $P_{0,15}$ et $P_{0,7}$, la figure 3.1 nous incite clairement à donner la préférence à $P_{0,15}$. Par contre, si on met en concurrence $P_{0,12}$ et $P_{0,14}$, le choix est beaucoup moins clair et on se dit qu'il faudrait peut-être davantage de tirages pour les départager. Intuitivement, plus θ_0 et θ_1 sont proches, plus il faudra d'observations pour les séparer. On peut étayer cette affirmation en rappelant que les intervalles de confiance obtenus par le TLC ou par l'inégalité de Tchebycheff ont une longueur en $O(n^{-1/2})$.

Ceci nous amène à proposer un autre modèle, ne dépendant plus du nombre de tirages n et permettant d'envisager une suite infinie de tirages avec remise⁴. On pose alors

$$\Omega := \{r, v\}^{\mathbb{N}^*}. \quad (3.4)$$

Les événements élémentaires $\omega \in \Omega$ sont donc les suites infinies $\omega = (u_i)_{i \geq 1}$ de terme général u_i valant r ou v . On munit Ω de la tribu \mathcal{F} engendrée par la famille \mathcal{C} des événements dont la réalisation ne dépend que d'un *nombre fini d'épreuves*⁵. Cette tribu est plus petite que $\mathcal{P}(\Omega)$. Notons pour tout $n \in \mathbb{N}$, $\Omega_n := \{j \in \mathbb{N}; j > n\}$. Avec Ω_n défini par (3.1), les éléments de \mathcal{C} admettent la représentation suivante :

$$C \in \mathcal{C} \Leftrightarrow \exists n \in \mathbb{N}^*, \exists A \subset \Omega_n, \quad C = A \times \{r, v\}^{\Omega_n}. \quad (3.5)$$

On définit ensuite pour chaque $\theta \in \Theta$ la probabilité P_θ sur \mathcal{C} en posant :

$$P_\theta(A \times \{r, v\}^{\Omega_n}) := P_{\theta,n}(A), \quad (3.6)$$

où $P_{\theta,n}$ est la probabilité sur Ω_n donnée par (3.2). Comme la représentation (3.5) n'est pas unique, cela pose un problème de compatibilité qu'il n'est pas difficile de résoudre, voir le chapitre 2 du cours d'I.P.É. On admet ensuite que P_θ ainsi définie sur \mathcal{C} se prolonge de manière unique en une probabilité sur la tribu $\mathcal{F} = \sigma(\mathcal{C})$.

On définit sur (Ω, \mathcal{F}) la suite infinie de variables aléatoires $(X_i)_{i \geq 1}$ par

$$\forall i \in \mathbb{N}^*, \quad X_i(\omega) = X_i((u_j)_{j \geq 1}) = \begin{cases} 1 & \text{si } u_i = v, \\ 0 & \text{si } u_i = r. \end{cases} \quad (3.7)$$

Pour toute valeur de θ , cette suite est i.i.d. sous P_θ , chaque X_i suivant la loi $\text{Bern}(\theta)$. En posant à nouveau $S_n := X_1 + \dots + X_n$, on peut appliquer la loi forte des grands nombres *sur chacun des espaces probabilisés* $(\Omega, \mathcal{F}, P_\theta)$, ce qui s'écrit ici :

$$\forall \theta \in \Theta, \quad \frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{P_\theta\text{-p.s.}} \mathbf{E}_\theta X_1 = \theta. \quad (3.8)$$

On pourrait de même appliquer pour chaque θ le TLC, c'est d'ailleurs ce que l'on a fait implicitement dans la construction des intervalles de confiance par le TLC vue au chapitre 1.

Pour compléter notre présentation du modèle statistique, il nous reste à définir les notions d'*échantillon* et de *statistique*. Nous avons déjà utilisé le mot « échantillon » au chapitre 2 au sens de suite finie de variables aléatoires indépendantes et de même loi. Il convient maintenant d'être plus précis.

4. Seulement par la pensée bien sûr !

5. Voir la discussion sur « une question de dés », cours d'I.P.É. chap. 2.

Définition 3.3 (échantillon d'une loi). Soit (Ω, \mathcal{F}, P) un espace probabilisé et Q une mesure de probabilité sur \mathbb{R} muni de sa tribu borélienne. Pour $n \geq 2$, on appelle n -échantillon de la loi Q , associé à (Ω, \mathcal{F}, P) , toute suite finie X_1, \dots, X_n de variables aléatoires définies sur (Ω, \mathcal{F}, P) , P -indépendantes et de même loi Q , c.-à-d. $\forall i = 1, \dots, n$, $P_{X_i} = P \circ X_i^{-1} = Q$.

Quand il n'apparaît pas utile de préciser l'espace (Ω, \mathcal{F}, P) concerné, on parle plus simplement d'échantillon de la loi Q . C'est ce que nous avons fait au chapitre 2. Notons d'ailleurs qu'étant donné une loi quelconque Q , autrement dit une mesure de probabilité sur \mathbb{R} muni de sa tribu borélienne, il est toujours possible de construire un espace probabilisé (Ω, \mathcal{F}, P) et une suite X_1, \dots, X_n de v.a. définies sur cet espace qui constituent un n -échantillon de la loi Q . Il suffit de prendre $\Omega = \mathbb{R}^n$, muni de sa tribu borélienne et de la probabilité produit $Q^{\otimes n}$ définie à partir des probabilités des pavés $C =]a_1, b_1] \times \dots \times]a_n, b_n]$ en posant $Q^{\otimes n}(C) := Q([a_1, b_1]) \dots Q([a_n, b_n])$. En définissant pour $i = 1, \dots, n$, X_i comme la projection $\mathbb{R}^n \rightarrow \mathbb{R}$ sur la i^{e} coordonnée, on vérifie facilement que X_1, \dots, X_n est un n -échantillon de la loi Q (pourquoi?).

Définition 3.4 (échantillon associé à un modèle statistique). Soit $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique. On appelle n -échantillon associé à ce modèle toute suite X_1, \dots, X_n de variables aléatoires définies sur (Ω, \mathcal{F}) qui sont pour tout $\theta \in \Theta$, P_θ -indépendantes et de même loi Q_θ sous P_θ :

$$\forall \theta \in \Theta, \forall i = 1, \dots, n, \quad P_{\theta, X_i} := P_\theta \circ X_i^{-1} = Q_\theta.$$

Avec les deux modèles statistiques proposés ci-dessus pour les tirages avec remise de l'exemple 3.1, les variables aléatoires X_1, \dots, X_n définies par (3.3) ou (3.7) constituent un n -échantillon associé au modèle. Pour chaque valeur de θ , elle sont P_θ -indépendantes et de même loi $\text{Bern}(\theta)$.

Définition 3.5 (statistique). Soit X_1, \dots, X_n un n -échantillon associé à un modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$. On appelle statistique associée à cette échantillon, toute v.a. Y_n de la forme

$$Y_n = f(X_1, \dots, X_n)$$

où $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $(t_1, \dots, t_n) \mapsto f(t_1, \dots, t_n)$ est une application borélienne ne dépendant pas de θ .

Au risque d'insister lourdement, notons que le point important dans cette définition est la possibilité de calculer Y_n à partir des X_i sans avoir besoin de connaître la valeur de θ . Bien sûr la loi de Y_n (sous P_θ) dépend en général de θ , mais la fonction f qui elle, n'a rien à voir avec le modèle ni avec les X_i , ne doit pas dépendre de θ . Voici quatre exemples simples de statistiques.

1. La moyenne empirique $\bar{X} := \frac{X_1 + \dots + X_n}{n} = \frac{S_n}{n}$.
2. La variance empirique $\frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{S_n}{n} \right)^2$.

$$3. \min_{1 \leq i \leq n} X_i.$$

$$4. \max_{1 \leq i \leq n} X_i.$$

Par contre $Y_n := n^{-1} \sum_{i=1}^n (X_i - \mathbf{E}X_i)^2$ n'est pas une statistique en général car $\mathbf{E}X_i$ doit en fait s'écrire $\mathbf{E}_\theta X_i$ et la fonction f correspondante donnée par $f(t_1, \dots, t_n) = n^{-1} \sum_{i=1}^n (t_i - \mathbf{E}_\theta X_i)^2$ n'est pas calculable sans la connaissance de $\mathbf{E}_\theta X_1$, donc du paramètre⁶ θ .

3.2 Mesure empirique

3.2.1 Une loi construite à partir des observations

Un outil fondamental en statistique est la *mesure empirique* que l'on peut décrire de manière informelle comme suit. On a des observations x_1, \dots, x_n , que l'on interprète comme les $X_1(\omega), \dots, X_n(\omega)$, où les v.a. X_i suivent la loi inconnue μ . Faute d'information supplémentaire, on décide de « ne pas faire de jalouses » parmi les observations en attribuant à chacune la probabilité $1/n$. On construit ainsi une nouvelle mesure de probabilité μ_n , dépendant des observations et on se sert de μ_n pour estimer la mesure de probabilité inconnue μ .

Définition 3.6 (mesure empirique). *Soit X_1, \dots, X_n un échantillon⁷. On appelle mesure empirique associée à cet échantillon, la « mesure aléatoire »*

$$\mu_n : \omega \mapsto \mu_n(\omega) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}.$$

Rappelons que la mesure de Dirac δ_x au point x est définie par

$$\forall B \in \text{Bor}(\mathbb{R}), \quad \delta_x(B) := \begin{cases} 0 & \text{si } x \notin B \\ 1 & \text{si } x \in B \end{cases} = \mathbf{1}_B(x).$$

On en déduit que

$$\begin{aligned} \forall \omega \in \Omega, \forall B \in \text{Bor}(\mathbb{R}), \quad \mu_n(\omega)(B) &= \mu_n(\omega, B) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_B(X_i(\omega)) \\ &= \frac{1}{n} \text{card} \{X_i(\omega); X_i(\omega) \in B\}. \end{aligned} \quad (3.9)$$

Autrement dit, $\mu_n(\omega, B)$ est la proportion de valeurs de l'échantillon appartenant à B ou *fréquence des observations* dans B .

6. Sauf dans le cas particulier où $\mathbf{E}_\theta X_i$ est constante par rapport à θ , par exemple si $\Theta =]0, +\infty[$ et si pour tout $\theta \in \Theta$, les X_i suivent sous P_θ la loi gaussienne $\mathfrak{N}(3, \theta)$.

7. que ce soit au sens de la définition 3.3 ou de la définition 3.4.

- Pour B borélien fixé, l'application $\mu_n(B) := \mu_n(\cdot, B) : \omega \mapsto \mu_n(\omega, B)$ est une variable aléatoire discrète. La loi de la v.a. $n\mu_n(B)$ est la binomiale de paramètres n et $p = P_{X_1}(B) = \mu(B)$. Elle dépend donc de la loi μ des X_i .
- Pour $\omega \in \Omega$ fixé, l'application $\mu_n(\omega) = \mu_n(\omega, \cdot) : B \mapsto \mu_n(\omega, B)$ est une mesure de probabilité sur \mathbb{R} muni de sa tribu borélienne. Si $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$ sont les observations, $\mu_n(\omega)$ est la loi discrète qui attribue une masse $1/n$ à chacune de ces observations. S'il n'y a pas d'*ex-æquo* parmi les observations, $\mu_n(\omega)$ est donc la loi uniforme discrète sur l'ensemble fini $\{x_1, \dots, x_n\}$.

Remarque 3.7 (espérance et variance de la mesure empirique). Rappelons que si I est un ensemble fini, l'espérance de la loi discrète $\sum_{i \in I} p_i \delta_{x_i}$ est $m = \sum_{i \in I} p_i x_i$. En appliquant ceci à la loi discrète $\mu_n(\omega)$, on voit que celle-ci a pour espérance $n^{-1} \sum_{i=1}^n X_i(\omega) = S_n(\omega)/n$. Autrement dit : *l'espérance de la mesure empirique est la moyenne arithmétique des valeurs de l'échantillon ou moyenne empirique*. De même la variance de la loi discrète $\sum_{i \in I} p_i \delta_{x_i}$ est $s^2 = \sum_{i \in I} p_i (x_i - m)^2$. En appliquant ceci à $\mu_n(\omega)$, on voit qu'elle a pour variance $n^{-1} \sum_{i=1}^n (X_i(\omega) - S_n(\omega)/n)^2$. Ainsi *la variance de la mesure empirique est la variance empirique de l'échantillon*.

Puisque $\mu_n(\omega)$ est une mesure de probabilité sur \mathbb{R} , elle est caractérisée par sa fonction de répartition.

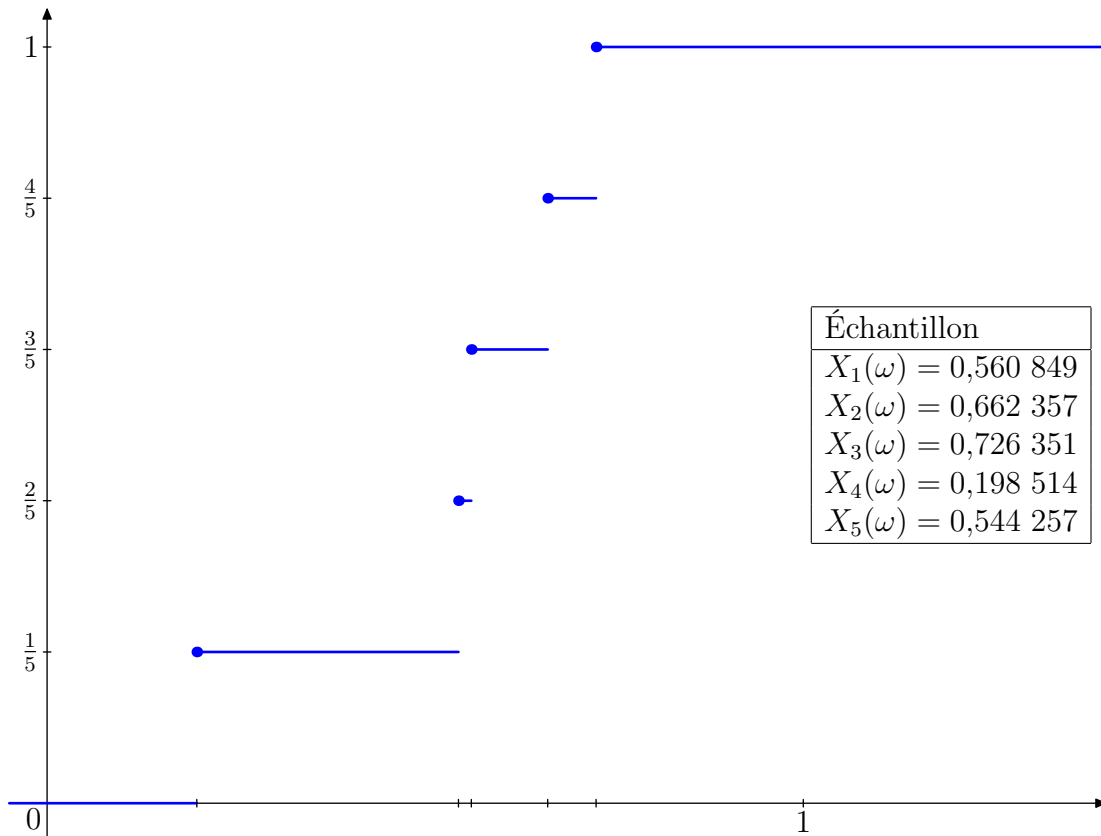
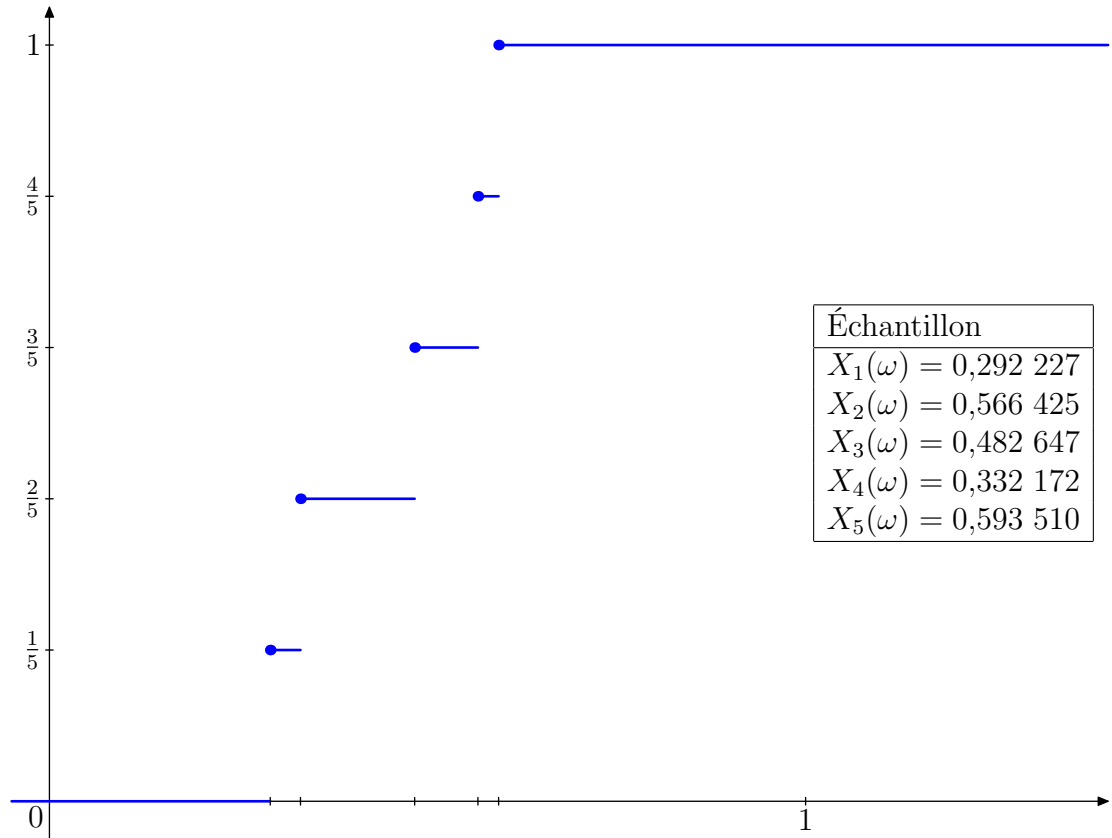


FIG. 3.2 – F.d.r. empirique d'un 5-échantillon de la loi uniforme sur $[0, 1]$


 FIG. 3.3 – F.d.r. empirique d'un autre 5-échantillon de la loi uniforme sur $[0, 1]$

Définition 3.8 (fonction de répartition empirique). Soit X_1, \dots, X_n un échantillon et μ_n la mesure empirique associée. On appelle fonction de répartition empirique de l'échantillon, la fonction de répartition F_n de μ_n . Plus formellement on pose :

$$\forall \omega \in \Omega, \forall x \in \mathbb{R}, \quad F_n(\omega, x) := \mu_n(\omega,] - \infty, x]). \quad (3.10)$$

En reprenant (3.9) avec un borélien B de la forme $B =] - \infty, x]$, on voit que :

$$\begin{aligned} \forall \omega \in \Omega, \forall x \in \mathbb{R}, \quad F_n(\omega, x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(X_i(\omega)) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i(\omega) \leq x\}} \\ &= \frac{1}{n} \text{card} \{X_i(\omega); X_i(\omega) \leq x\}. \end{aligned} \quad (3.11)$$

- Pour $\omega \in \Omega$ fixé, $F_n(\omega) := F_n(\omega, \cdot) : x \mapsto F_n(\omega, x)$ est la fonction de répartition de la loi de probabilité discrète $\mu_n(\omega) = n^{-1} \sum_{i=1}^n \delta_{X_i(\omega)}$.
- Pour x fixé, $F_n(x) := F_n(\cdot, x) : \omega \mapsto F_n(\omega, x)$ est la variable aléatoire discrète

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} \quad (3.12)$$

autrement dit la v.a. « proportion de valeurs inférieures ou égales à x dans l'échantillon ». Si F désigne la f.d.r. commune aux X_i , la loi de cette v.a. $F_n(x)$ est donnée par

$$\forall k = 0, 1, \dots, n, \quad P\left(F_n(x) = \frac{k}{n}\right) = C_n^k F(x)^k (1 - F(x))^{n-k}. \quad (3.13)$$

3.2.2 Convergence de la f.d.r. empirique vers la f.d.r. théorique

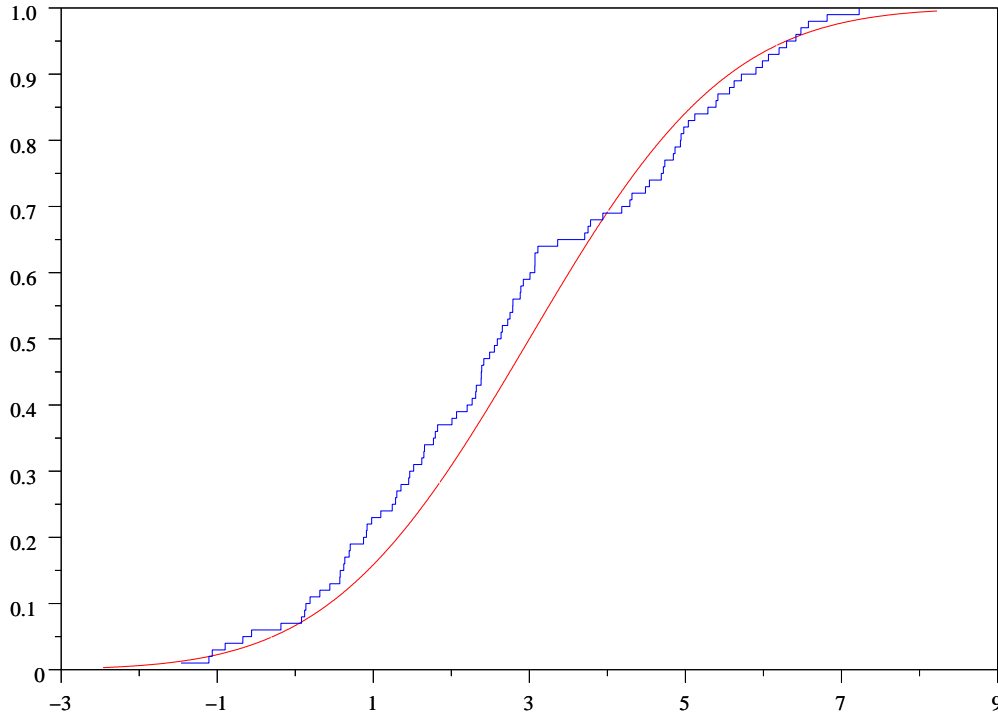


FIG. 3.4 – Approximation de la f.d.r. de $\mathfrak{N}(3, 2)$ par une f.d.r. empirique, $n = 100$

Le théorème suivant nous dit que presque-sûrement, F_n converge uniformément vers F sur \mathbb{R} . C'est en quelque sorte une loi forte des grands nombres fonctionnelle pour la suite des f.d.r. empiriques vues comme des fonctions aléatoires. La signification pratique est que si l'on a observé un échantillon de grande taille d'une loi inconnue de f.d.r. F , la fonction de répartition empirique peut être prise comme approximation de F . À titre d'illustration, les figures 3.4 et 3.5 montrent la f.d.r. de la loi $\mathfrak{N}(3, 2)$ et des f.d.r. empiriques construites sur un n -échantillon de la loi $\mathfrak{N}(3, 2)$ pour $n = 100$ et $n = 400$,

simulation réalisée avec Scilab. Les sauts des f.d.r. empiriques sont représentés ici par des segments verticaux en trait plein.

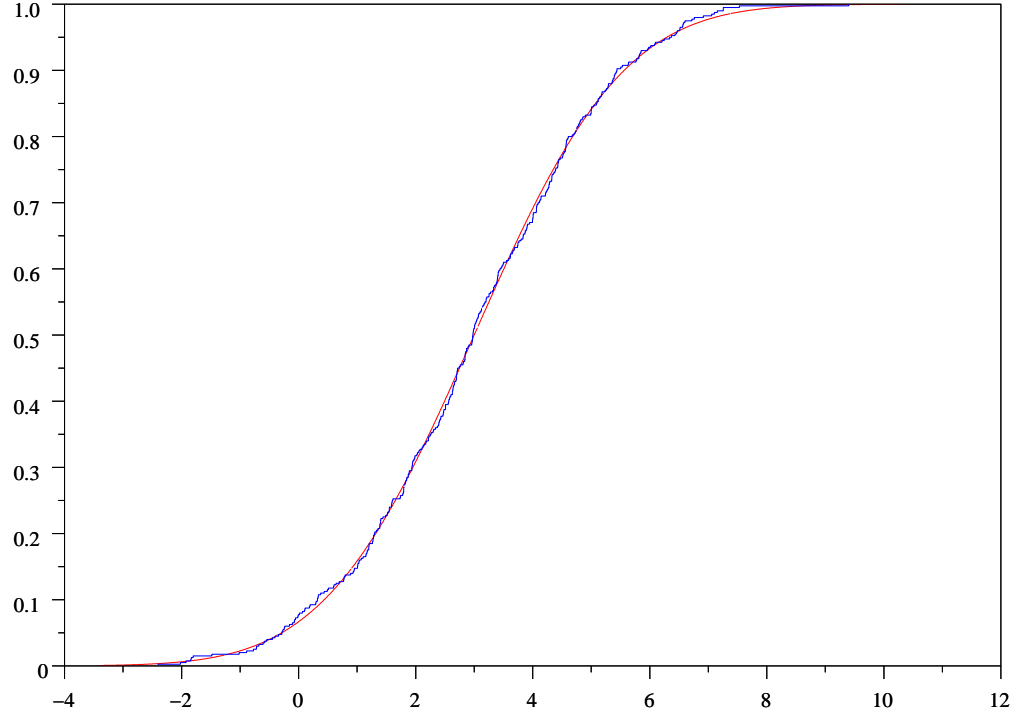


FIG. 3.5 – Approximation de la f.d.r. de $\mathfrak{N}(3, 2)$ par une f.d.r. empirique, $n = 400$

Théorème 3.9 (Glivenko-Cantelli). *Soit F_n la f.d.r. empirique d'un échantillon X_1, \dots, X_n , où les X_i ont pour f.d.r. F . Alors*

- a) $\forall x \in \mathbb{R}, \quad F_n(x) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} F(x);$
- b) $\|F_n - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0.$

Preuve du a). Fixons $x \in \mathbb{R}$ et posons $Y_i := \mathbf{1}_{\{X_i \leq x\}}$. En rappelant que si A est un évènement, $\mathbf{1}_A$ est une v.a. de Bernoulli de paramètre $p = P(A)$ et d'espérance $p = P(A)$, on voit que les Y_i sont des v.a. i.i.d. de Bernoulli de paramètre

$$p = P(X_1 \leq x) = F(x).$$

En appliquant la loi forte des grands nombres à la suite i.i.d. de v.a. bornées (donc intégrables) Y_i , on obtient

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{E}Y_1 = p = F(x),$$

ce qui établit le a). □

Avant d'attaquer la preuve du b), il convient de se demander si $\|F_n - F\|_\infty$ est bien une variable aléatoire, c'est à dire une application mesurable $(\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \text{Bor}(\mathbb{R}))$. En effet pour x fixé, $|F_n(x) - F(x)|$ est une variable aléatoire, mais $\|F_n - F\|_\infty$ est un supremum indexé par l'ensemble *non dénombrable* \mathbb{R} de v.a. $|F_n(x) - F(x)|$. Il n'est donc pas évident que ce supremum hérite de la mesurabilité des v.a. $|F_n(x) - F(x)|$. Le lemme suivant apporte une réponse positive.

Lemme 3.10. *La f.d.r. empirique F_n d'un échantillon X_1, \dots, X_n , où les X_i ont pour f.d.r. F vérifie*

$$\forall \omega \in \Omega, \quad \sup_{x \in \mathbb{R}} |F_n(\omega, x) - F(x)| = \sup_{x \in \mathbb{Q}} |F_n(\omega, x) - F(x)|. \quad (3.14)$$

En conséquence, $\|F_n - F\|_\infty$ est une variable aléatoire réelle.

Preuve du lemme 3.10. Fixons ω quelconque dans Ω et notons pour alléger

$$\tau := \|F_n(\omega, \cdot) - F\|_\infty = \sup_{x \in \mathbb{R}} |F_n(\omega, x) - F(x)|.$$

Comme $F_n(\omega, x)$ et $F(x)$ sont toujours deux réels de $[0, 1]$, ce supremum τ est fini. Alors pour tout $\varepsilon > 0$, il existe un x_ε tel que

$$\tau - \varepsilon < |F_n(\omega, x_\varepsilon) - F(x_\varepsilon)| \leq \tau.$$

Les f.d.r. $F_n(\omega, \cdot)$ et F étant continues à droite au point x_ε , la valeur absolue de leur différence l'est aussi. Il existe donc un $\delta > 0$ tel que

$$\forall t \in]x_\varepsilon, x_\varepsilon + \delta[, \quad |F_n(\omega, t) - F(t)| > |F_n(\omega, x_\varepsilon) - F(x_\varepsilon)| - \varepsilon$$

Dans l'intervalle $]x_\varepsilon, x_\varepsilon + \delta[$, il y a au moins un nombre *rationnel* t . Ce rationnel vérifiant l'inégalité ci-dessus, on en déduit

$$\sup_{r \in \mathbb{Q}} |F_n(\omega, r) - F(r)| > \tau - 2\varepsilon.$$

Le premier membre ne dépendant pas de ε et $\varepsilon > 0$ étant arbitraire, on en déduit :

$$\sup_{r \in \mathbb{Q}} |F_n(\omega, r) - F(r)| \geq \tau = \sup_{r \in \mathbb{R}} |F_n(\omega, r) - F(r)|.$$

Puisque ω était quelconque, ceci vaut pour tout $\omega \in \Omega$. L'inégalité dans l'autre sens est évidente, donc l'égalité (3.14) est démontrée pour tout $\omega \in \Omega$. Elle permet de voir l'application $\|F_n - F\|_\infty : \Omega \rightarrow \mathbb{R}, \omega \mapsto \|F_n(\omega, \cdot) - F\|_\infty$ comme le sup d'une famille *dénombrable* de variables aléatoires réelles. Cette application $\|F_n - F\|_\infty$ hérite ainsi de la mesurabilité de ces v.a., c'est donc elle-même une variable aléatoire. □

Preuve du b) dans le cas de v.a. uniformes sur $]0, 1[$. Nous verrons ci-dessous comment on peut réduire la preuve du b) au cas particulier où les X_i sont des v.a. U_i de loi uniforme sur $]0, 1[$, cette réduction ayant son intérêt propre. Notons G_n la f.d.r. empirique bâtie sur l'échantillon U_1, \dots, U_n et G la f.d.r. de la loi uniforme sur $]0, 1[$. Ces deux f.d.r. valent 0 sur $] - \infty, 0[$ et 1 sur $]1, +\infty[$, donc leur différence est nulle sur ces intervalles. De plus sur $]0, 1[$, $G(t) = t$. On a donc

$$\sup_{t \in \mathbb{R}} |G_n(t) - G(t)| = \sup_{t \in [0,1]} |G_n(t) - t|. \quad (3.15)$$

Commençons par fixer un entier M arbitraire et par vérifier que

$$\max_{0 \leq k \leq M} \left| G_n\left(\frac{k}{M}\right) - \frac{k}{M} \right| \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0. \quad (3.16)$$

En effet en appliquant le a) avec $t = k/M$, on obtient :

$$\forall k = 0, 1, \dots, M, \exists \Omega_{M,k} \in \mathcal{F}, P(\Omega_{M,k}) = 1, \forall \omega \in \Omega_{M,k}, \left| G_n\left(\omega, \frac{k}{M}\right) - \frac{k}{M} \right| \xrightarrow[n \rightarrow +\infty]{} 0.$$

En posant $\Omega_M := \cap_{0 \leq k \leq M} \Omega_{M,k}$, on a encore $P(\Omega_M) = 1$. Comme le max de $M+1$ suites convergentes vers 0 est encore une suite convergente vers 0, on a

$$\forall \omega \in \Omega_M, \max_{0 \leq k \leq M} \left| G_n\left(\omega, \frac{k}{M}\right) - \frac{k}{M} \right| \xrightarrow[n \rightarrow +\infty]{} 0, \quad (3.17)$$

ce qui établit (3.16) puisque $P(\Omega_M) = 1$.

Ensuite par *croissance* de la f.d.r. G_n on a

$$\forall t \in \left] \frac{k-1}{M}, \frac{k}{M} \right], \quad G_n\left(\frac{k-1}{M}\right) - \frac{k}{M} \leq G_n(t) - t \leq G_n\left(\frac{k}{M}\right) - \frac{k-1}{M},$$

ce qui s'écrit encore

$$\forall t \in \left] \frac{k-1}{M}, \frac{k}{M} \right], \quad G_n\left(\frac{k-1}{M}\right) - \frac{k-1}{M} - \frac{1}{M} \leq G_n(t) - t \leq G_n\left(\frac{k}{M}\right) - \frac{k}{M} + \frac{1}{M}.$$

On en déduit que

$$\forall t \in \left] \frac{k-1}{M}, \frac{k}{M} \right], \quad |G_n(t) - t| \leq \max_{0 \leq j \leq M} \left| G_n\left(\frac{j}{M}\right) - \frac{j}{M} \right| + \frac{1}{M}.$$

Comme ce majorant ne dépend pas de t ni de k , on en déduit

$$\sup_{t \in [0,1]} |G_n(t) - t| \leq \max_{0 \leq j \leq M} \left| G_n\left(\frac{j}{M}\right) - \frac{j}{M} \right| + \frac{1}{M}. \quad (3.18)$$

Notons que (3.18) est vraie sur tout Ω . Elle est donc vraie en particulier sur l'évènement Ω_M de probabilité 1. Compte-tenu de (3.17), on en déduit :

$$\forall \omega \in \Omega_M, \quad \limsup_{n \rightarrow +\infty} \sup_{t \in [0,1]} |G_n(\omega, t) - t| \leq 0 + \frac{1}{M}. \quad (3.19)$$

Nous utilisons ici la limite sup et pas la limite car s'il est clair par (3.17) que le second membre de (3.18) a une limite égale à $1/M$ quand n tend vers $+\infty$, nous ne savons aucunement à ce stade si son premier membre a une limite. On contourne cette difficulté en utilisant la limite sup qui a l'avantage de toujours exister⁸.

Pour finir, il est temps de se souvenir que l'entier M était arbitraire, donc (3.19) est vraie pour tout $M \in \mathbb{N}^*$. L'évènement $\Omega' := \cap_{M \geq 1} \Omega_M$ est de probabilité 1 comme intersection *dénombrable* d'évènements de probabilité 1. On a donc

$$\forall \omega \in \Omega', \forall M \geq 1, \quad 0 \leq \limsup_{n \rightarrow +\infty} \sup_{t \in [0,1]} |G_n(\omega, t) - t| \leq \frac{1}{M}.$$

En faisant maintenant tendre M vers $+\infty$, on en déduit que

$$\forall \omega \in \Omega', \quad 0 \leq \limsup_{n \rightarrow +\infty} \sup_{t \in [0,1]} |G_n(\omega, t) - t| \leq 0.$$

On en déduit que pour tout $\omega \in \Omega'$, donc presque-sûrement, la limite quand n tend vers $+\infty$ de $\sup_{t \in [0,1]} |G_n(\omega, t) - t|$ existe et vaut 0. En rappelant (3.15), on conclut que le théorème de Glivenko-Cantelli est démontré pour la f.d.r. empirique G_n . \square

Pour compléter la preuve du théorème de Glivenko-Cantelli, il nous reste à justifier la réduction au cas d'un échantillon de la loi Unif] $0, 1$ [. C'est l'objet du lemme suivant.

Lemme 3.11. *Soit $(X_i)_{i \geq 1}$ une suite i.i.d. de f.d.r. F et $(U_i)_{i \geq 1}$ une suite i.i.d. de variables aléatoires à valeurs dans $]0, 1[$ et de loi uniforme sur $]0, 1[$. On note F_n la f.d.r. empirique construite sur X_1, \dots, X_n et F^{-1} l'inverse généralisé de F défini sur $]0, 1[$ par $F^{-1}(u) := \inf\{x \in \mathbb{R}; F(x) \geq u\}$. On pose pour tout $i \geq 1$, $Y_i := F^{-1}(U_i)$ et on note H_n la f.d.r. empirique construite sur Y_1, \dots, Y_n . Alors*

1. *pour tout $n \geq 1$ et tout $\varepsilon > 0$, $P(\|F_n - F\|_\infty > \varepsilon) = P(\|H_n - F\|_\infty > \varepsilon)$;*
2. *$P\left(\|F_n - F\|_\infty \xrightarrow{n \rightarrow +\infty} 0\right) = P\left(\|H_n - F\|_\infty \xrightarrow{n \rightarrow +\infty} 0\right)$;*
3. *pour tout $\omega \in \Omega$,*

$$\|H_n(\omega, \cdot) - F\|_\infty \leq \sup_{t \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i(\omega) \leq t\}} - t \right|, \quad (3.20)$$

avec égalité si $F(\mathbb{R}) \supset]0, 1[$, autrement dit si F est continue sur \mathbb{R} .

En première lecture, on pourra admettre les points 1 et 2 et se contenter de regarder la preuve du point 3.

Preuve du point 1. Par le théorème 2.2, pour tout $i \geq 1$, X_i et $Y_i = F^{-1}(U_i)$ ont même loi. Par indépendance de leurs composantes, on en déduit que pour tout $n \geq 1$, les

8. Et qui, comme la limite, conserve les inégalités larges.

vecteurs aléatoires (X_1, \dots, X_n) et (Y_1, \dots, Y_n) ont même loi. D'après (3.14), on a pour tout $\omega \in \Omega$,

$$\|F_n(\omega, \cdot) - F\|_\infty = \sup_{x \in \mathbb{Q}} |F_n(\omega, x) - F(x)|, \quad (3.21)$$

$$\|H_n(\omega, \cdot) - F\|_\infty = \sup_{x \in \mathbb{Q}} |H_n(\omega, x) - F(x)|, \quad (3.22)$$

relisez la preuve de (3.14) si vous doutez de la deuxième égalité. Fixons une numérotation $j \mapsto r_j$ de \mathbb{Q} par les entiers de sorte que $\mathbb{Q} = \{r_j, j \in \mathbb{N}\}$. Introduisons pour tout $k \geq 1$ les variables aléatoires

$$T_k := \max_{0 \leq j \leq k} |F_n(r_j) - F(r_j)|, \quad T'_k := \max_{0 \leq j \leq k} |H_n(r_j) - F(r_j)|.$$

T_k est une fonction mesurable du vecteur (X_1, \dots, X_n) et T'_k est la même fonction mesurable du vecteur (Y_1, \dots, Y_n) . Par conséquent les v.a. T_k et T'_k ont même loi. En particulier,

$$\forall k \geq 1, \forall \varepsilon > 0, \quad P(T_k > \varepsilon) = P(T'_k > \varepsilon). \quad (3.23)$$

La suite de v.a. positives $(T_k)_{k \geq 1}$ est croissante de limite $\sup_{x \in \mathbb{Q}} \|F_n(x) - F(x)\|$. Il en résulte que la suite d'évènements $(\{T_k > \varepsilon\})_{k \geq 1}$ est croissante de réunion l'évènement $\{\sup_{x \in \mathbb{Q}} \|F_n(x) - F(x)\| > \varepsilon\}$. De même $\{T_k > \varepsilon\} \uparrow \{\sup_{x \in \mathbb{Q}} \|H_n(x) - F(x)\| > \varepsilon\}$. Par continuité séquentielle croissante de P , en passant à la limite dans (3.23) et en tenant compte de (3.21) et (3.22), on obtient $P(\|F_n - F\|_\infty > \varepsilon) = P(\|H_n - F\|_\infty > \varepsilon)$, ce qui établit le point 1 du lemme. \square

Preuve du point 2. Posons

$$A := \left\{ \|F_n - F\|_\infty \xrightarrow{n \rightarrow +\infty} 0 \right\}, \quad A' := \left\{ \|H_n - F\|_\infty \xrightarrow{n \rightarrow +\infty} 0 \right\}.$$

Choisissons une suite $(\varepsilon_i)_{i \geq 0}$ décroissante vers 0, par exemple $\varepsilon_i := 2^{-i}$. En discrétisant la définition de la limite à l'aide de cette suite $(\varepsilon_i)_{i \geq 0}$, revoyez si nécessaire le chapitre 6 du cours d'I.P.É., on obtient :

$$A = \bigcap_{i \geq 0} \bigcup_{j \geq 1} \bigcap_{k \geq j} \{ \|F_k - F\|_\infty \leq \varepsilon_i \}, \quad A' = \bigcap_{i \geq 0} \bigcup_{j \geq 1} \bigcap_{k \geq j} \{ \|H_k - F\|_\infty \leq \varepsilon_i \}.$$

Posons pour $i \in \mathbb{N}$ et $l \geq j$,

$$A_{i,j,l} := \bigcap_{k=j}^{k=l} \{ \|F_k - F\|_\infty \leq \varepsilon_i \}, \quad A'_{i,j,l} := \bigcap_{k=j}^{k=l} \{ \|H_k - F\|_\infty \leq \varepsilon_i \}.$$

Par le même raisonnement qu'au point 1, on voit que

$$\forall i \geq 0, \forall j \geq 1, \forall l \geq j, \quad P(A_{i,j,l}) = P(A'_{i,j,l}), \quad (3.24)$$

puisque la réalisation de $A_{i,j,l}$, resp. $A'_{i,j,l}$, ne dépend que des valeurs prises par l'échantillon (X_1, \dots, X_l) , resp. (Y_1, \dots, Y_l) et ces deux vecteurs aléatoires ont même loi. Ensuite

en utilisant 3 fois la continuité séquentielle de P , on en déduit que $P(A) = P(A')$ ce qui établit le point 2. Essayez d'abord d'écrire la justification par vous même avant de lire éventuellement les détails que voici.

Dans ce qui suit nous définirons les notations $A_{i,j}$ puis A_i . Pour alléger les écritures, convenons que $A'_{i,j}$ et A'_i seront définies de la même façon en remplaçant F_k par H_k . Pour tout couple (i, j) fixé, la suite $(A_{i,j,l})_{l \geq j}$ est décroissante pour l'inclusion, d'où par continuité séquentielle décroissante de P :

$$A_{i,j,l} \downarrow A_{i,j} := \bigcap_{k \geq j} \{ \|F_k - F\|_\infty \leq \varepsilon_i \} \Rightarrow P(A_{i,j,l}) \downarrow P(A_{i,j}), \quad (l \uparrow +\infty).$$

Il en va de même avec $A'_{i,j,l}$ et $A'_{i,j}$. En passant à la limite dans (3.24) quand l tend vers l'infini, i et j restant fixés, on obtient :

$$\forall i \geq 0, \forall j \geq 1, \quad P(A_{i,j}) = P(A'_{i,j}). \quad (3.25)$$

Pour i fixé, la suite $(A_{i,m})_{m \geq 1}$ est croissante pour l'inclusion, donc par continuité séquentielle croissante de P :

$$A_{i,m} \uparrow A_i := \bigcup_{j \geq 1} A_{i,j} = \bigcup_{j \geq 1} \bigcap_{k \geq j} \{ \|F_k - F\|_\infty \leq \varepsilon_i \} \Rightarrow P(A_{i,m}) \uparrow P(A_i), \quad (m \uparrow +\infty).$$

De même $P(A'_{i,m}) \uparrow P(A'_i)$. Réécrivons (3.25) avec m à la place de j et passons à la limite quand m tend vers l'infini, i restant fixé, pour obtenir :

$$\forall i \geq 0, \quad P(A_i) = P(A'_i). \quad (3.26)$$

Pour tout $n \in \mathbb{N}$, $\varepsilon_{n+1} \leq \varepsilon_n$ donc $A_{n+1} \subset A_n$. La suite $(A_n)_{n \geq 0}$ est ainsi décroissante pour l'inclusion et par continuité séquentielle décroissante de P :

$$A_n \downarrow \bigcap_{i \geq 0} A_i = A \Rightarrow P(A_n) \downarrow P(A), \quad (n \uparrow +\infty).$$

De même $P(A'_n) \downarrow P(A')$ et on déduit de (3.26) que $P(A) = P(A')$. □

Preuve du point 3. On rappelle, voir (2.7) et la fin de la preuve du théorème 2.2, que

$$\forall u \in]0, 1[, \forall x \in \mathbb{R}, \quad (u \leq F(x)) \Leftrightarrow (F^{-1}(u) \leq x).$$

De cette équivalence, on tire immédiatement l'égalité suivante vraie sur tout Ω :

$$\|H_n(\omega, \cdot) - F\|_\infty = \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{F^{-1}(U_i) \leq x\}} - F(x) \right| = \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i \leq F(x)\}} - F(x) \right|.$$

Comme F est une f.d.r., on a l'inclusion $F(\mathbb{R}) \subset [0, 1]$, de sorte qu'en posant $t = F(x)$ dans le dernier supremum ci-dessus on obtient :

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i \leq F(x)\}} - F(x) \right| \leq \sup_{t \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i \leq t\}} - t \right|. \quad (3.27)$$

Ceci établit l'inégalité (3.20).

Pour trouver une condition suffisante d'égalité dans (3.20), on remarque qu'en notant G_n la f.d.r. empirique bâtie sur U_1, \dots, U_n , $G_n(t) - t$ s'annule en $t = 0$ et en $t = 1$. En effet on a pour tout $\omega \in \Omega$, $U_i(\omega) \in]0, 1[$, d'où $G_n(0) = 0$ et $G_n(1) = 1$. Par conséquent $\sup_{t \in [0,1]} |G_n(t) - t| = \sup_{t \in]0,1[} |G_n(t) - t|$. Pour avoir l'égalité dans (3.27), il suffit donc que $F(\mathbb{R}) \supset]0, 1[$, ce qui équivaut à la continuité de F sur \mathbb{R} car F est une f.d.r. Plus précisément, si F a une discontinuité en un point a , alors $0 \leq F(a-) < F(a) \leq 1$ et par croissance de F , aucun $y \in]F(a-), F(a)[$ ne peut être l'image d'un réel x par F . Ceci empêche clairement que $F(\mathbb{R})$ contienne tout l'intervalle $]0, 1[$. Si F est continue, comme elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$, le théorème des valeurs intermédiaires nous donne l'inclusion $]0, 1[\subset F(\mathbb{R})$. \square

Fin de la preuve du théorème de Glivenko-Cantelli. En combinant les points 2 et 3 du lemme 3.11, on obtient :

$$1 \geq P\left(\|F_n - F\|_\infty \xrightarrow{n \rightarrow +\infty} 0\right) \geq P\left(\sup_{t \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{U_i \leq t\}} - t \right| \xrightarrow{n \rightarrow +\infty} 0\right).$$

La preuve du b) dans le cas de v.a. uniformes sur $]0, 1[$ nous a montré que cette dernière probabilité vaut 1, ce qui nous permet de conclure définitivement à la convergence p.s. de $\|F_n - F\|_\infty$ vers zéro. \square

3.2.3 Application au test de Kolmogorov-Smirnov

On peut utiliser la distance uniforme entre la f.d.r. empirique F_n et F , pour tester l'hypothèse que les observations $X_1(\omega), \dots, X_n(\omega)$ proviennent ou non d'un échantillon de la loi de f.d.r. F continue et connue. L'idée du test est la suivante. Posons

$$D_n := \|F_n - F\|_\infty.$$

Nous avons vu ci-dessus que D_n est bien une variable aléatoire positive. On se fixe un seuil ε , disons $\varepsilon = 0,05$, et on détermine un réel t_ε tel que si les X_i ont vraiment pour f.d.r. F , $P(D_n > t_\varepsilon) = \varepsilon$. On calcule ensuite la valeur $D_n(\omega)$ à partir des observations $X_1(\omega), \dots, X_n(\omega)$. Si on trouve une valeur supérieure à t_ε , on rejette l'hypothèse que la loi commune des X_i a pour f.d.r. F . Si $D_n(\omega) \leq t_\varepsilon$, on l'accepte. Les trois points clé sont ici les suivants.

1. Si les X_i ont vraiment pour f.d.r. F et si F est continue, la loi de D_n ne dépend pas de F .
2. On sait calculer la f.d.r. de D_n , de manière exacte pour les petites valeurs de n (cf. tables p. 129) ou *via* un théorème de convergence en loi quand n tend vers l'infini pour les grandes valeurs de n .
3. Il est possible de calculer exactement la valeur de $D_n(\omega)$, de manière simple et économique à partir des observations $X_1(\omega), \dots, X_n(\omega)$.

Le point 1 découle clairement du lemme 3.11, en notant que si F est continue, on a égalité dans (3.20).

Pour le point 2, voici ce que l'on peut dire et que nous admettrons. On peut montrer que la fonction de répartition de D_n se calcule par la formule suivante :

$$\forall t > 0, \quad P(D_n \leq t) = n! \det [c_{i,j}],$$

où $[c_{i,j}]$ est la matrice $n \times n$ dont les termes sont définis par :

$$c_{i,j} = \begin{cases} 0 & \text{si } j+1 < i \leq n, \\ \frac{(b_i - a_j)_+^{j-i+1}}{(j-i+1)!} & \text{si } 1 \leq i \leq j+1, \end{cases}$$

où $(b_i - a_j)_+$ désigne la partie positive de $b_i - a_j$ et

$$a_j = \max \left(-t + \frac{j}{n}; 0 \right), \quad b_i = \min \left(t + \frac{i-1}{n}; 1 \right).$$

En pratique, on utilise une table donnant pour les valeurs de ε les plus couramment utilisées, les valeurs de t_ε telles que $P(D_n > t_\varepsilon) = \varepsilon$. Pour les grandes valeurs de n , on a plutôt recours au résultat suivant qui exprime la convergence en loi de $\sqrt{n}D_n$.

Théorème 3.12 (Kolmogorov Smirnov). *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes et de même loi de fonction de répartition continue F . Soit F_n la fonction de répartition empirique construite sur X_1, \dots, X_n et $D_n = \|F_n - F\|_\infty$. On a alors*

$$\forall x > 0, \quad \lim_{n \rightarrow +\infty} P(n^{1/2} D_n \leq x) = K(x) := 1 + 2 \sum_{j=1}^{+\infty} (-1)^j \exp(-2j^2 x^2). \quad (3.28)$$

Lorsque x n'est pas trop proche de 0, la série définissant K converge très rapidement et il suffit de calculer un petit nombre de termes pour avoir une excellente précision⁹.

Voyons maintenant le calcul pratique de $D_n(\omega)$. Notons $(X_{n:1}, \dots, X_{n:n})$ le vecteur aléatoire des statistiques d'ordre, obtenu par réarrangement croissant¹⁰ de l'échantillon :

$$\{X_1, \dots, X_n\} = \{X_{n:1}, \dots, X_{n:n}\} \quad \text{et} \quad X_{n:1} \leq X_{n:2} \leq \dots \leq X_{n:n}.$$

En particulier, $X_{n:1} = \min_{1 \leq i \leq n} X_i$ et $X_{n:n} = \max_{1 \leq i \leq n} X_i$. Il est commode de poser aussi, compte-tenu des limites des f.d.r. en $\pm\infty$:

$$X_{n:0} := -\infty, F(X_{n:0}) = F_n(X_{n:0}) = 0, \quad X_{n:n+1} := +\infty, F(X_{n:n+1}) = F_n(X_{n:n+1}) = 1.$$

La fonction de répartition empirique F_n est une fonction en escaliers, constante sur chaque intervalle $[X_{n:i}, X_{n:i+1}[$, $0 \leq i \leq n$ et valant i/n sur cet intervalle.

9. Il ne saute pas aux yeux que K est bien une fonction de répartition. La vérification de cette propriété de K est un bon exercice d'analyse.

10. L'ordre de réarrangement dépend de ω .

Proposition 3.13. Avec les notations ci-dessus, lorsque la f.d.r. F est continue, la statistique $D_n = \|F_n - F\|_\infty$ se calcule par la formule :

$$D_n = \max_{0 \leq i \leq n} \max \left(\frac{i}{n} - F(X_{n:i}); F(X_{n:i+1}) - \frac{i}{n} \right). \quad (3.29)$$

Preuve. Comme $F(-\infty) = F_n(-\infty) = 0$, le supremum de $|F_n - F|$ sur \mathbb{R} est le même que sur $[-\infty, +\infty[$. Cet intervalle étant la réunion disjointe des $[X_{n:i}, X_{n:i+1}[$ pour $0 \leq i \leq n$, on en déduit que

$$D_n = \sup_{\mathbb{R}} |F_n - F| = \max_{0 \leq i \leq n} Z_i, \quad \text{où} \quad Z_i := \sup_{[X_{n:i}, X_{n:i+1}[} |F_n - F|.$$

Le problème se réduit donc à vérifier que

$$\forall i \in \{0, 1, \dots, n\}, \quad Z_i = \max \left(\frac{i}{n} - F(X_{n:i}); F(X_{n:i+1}) - \frac{i}{n} \right). \quad (3.30)$$

Traisons d'abord le cas « général » $1 \leq i < n$, où l'intervalle $I_i := [X_{n:i}, X_{n:i+1}[$ est borné. Il y a trois configurations possibles selon la position de la restriction à I_i du graphe de F relativement au segment horizontal d'équation $y = i/n$ (graphe de la restriction de F_n à I_i). La vérification ci-dessous utilise la croissance de F et dans les cas 2 et 3, sa continuité à gauche au point $X_{n:i+1}$.

1. Si $F(X_{n:i+1}) \leq i/n$, le supremum de $|F_n - F|$ sur I_i est atteint en $X_{n:i}$, borne gauche de I_i et vaut $F_n(X_{n:i}) - F(X_{n:i}) = i/n - F(X_{n:i})$. D'autre part, $F(X_{n:i+1}) - i/n$ est négatif. Ainsi l'égalité dans (3.30) est vérifiée dans ce cas.
2. Si $F(X_{n:i}) \leq i/n \leq F(X_{n:i+1})$, le supremum de $|F_n - F|$ sur I_i est atteint en l'une des bornes de cet intervalle, soit $X_{n:i}$ soit $X_{n:i+1}$ et (3.30) est encore vérifiée.
3. Si $i/n < F(X_{n:i})$, le supremum de $|F_n - F|$ sur I_i est atteint en $X_{n:i+1}$, borne droite de I_i et vaut $F(X_{n:i+1}) - i/n$. D'autre part $i/n - F(X_{n:i})$ est négatif. L'égalité (3.30) est encore vérifiée dans ce cas.

Dans le cas particulier $i = 0$, $I_0 = [-\infty, X_{n:1}[$ et $Z_0 = F(X_{n:1})$ par croissance de F et continuité à gauche en $X_{n:1}$. Là encore (3.30) est vérifiée. Pour $i = n$, $I_n = [X_{n:n}, +\infty[$ et $Z_n = 1 - F(X_{n:n})$, ce qui vérifie aussi (3.30) puisque $F(X_{n:n+1}) - n/n = 1 - 1 = 0$. \square

3.3 Moments empiriques

3.3.1 Moments observés et moments empiriques

Nous avons déjà vu, cf. remarque 3.7, que si l'on a des observations x_1, \dots, x_n , interprétées comme des réalisations $X_1(\omega), \dots, X_n(\omega)$, les statistiques $\bar{x} := n^{-1}(x_1 + \dots + x_n)$ et $s^2 := n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ appelées moyenne observée et variance observée peuvent être interprétées comme l'espérance et la variance de la mesure de probabilité $\mu_n(\omega) = n^{-1} \sum_{i=1}^n \delta_{x_i}$, c.-à-d. de la mesure empirique. Quand on laisse varier ω , on obtient ainsi des variables aléatoires $\bar{X} = n^{-1}(X_1 + \dots + X_n)$ et $S^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ appelées respectivement moyenne empirique et variance empirique. Ces notions se généralisent aux moments d'ordre k .

Définition 3.14 (moments empiriques). Soit $k \in \mathbb{N}^*$. On appelle moment empirique d'ordre k associé à l'échantillon X_1, \dots, X_n le moment d'ordre k de la mesure empirique μ_n , c.-à-d. la variable aléatoire

$$M_k := \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Pour ω fixé, on a des observations numériques $x_1 = X_1(\omega), \dots, x_n = X_n(\omega)$, le moment observé d'ordre k est alors $M_k(\omega) = n^{-1} \sum_{i=1}^n x_i^k$.

On vous propose de montrer en exercice les convergences suivantes.

$$\text{Si } \mathbf{E}|X_1| < +\infty, \quad \overline{X} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbf{E}X_1. \quad (3.31)$$

$$\text{Si } \mathbf{E}X_1^2 < +\infty, \quad S^2 \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \text{Var } X_1. \quad (3.32)$$

Ceci montre que pour des échantillons de grande taille, la moyenne empirique et la variance empirique peuvent être prises comme approximations respectives de l'espérance théorique $\mathbf{E}X_1$ et de la variance théorique $\text{Var } X_1$. Attention aux notations trompeuses, \overline{X} et S^2 dépendent de n . Pour prouver (3.32), la formule de Koenig pour la variance empirique, qui a aussi son intérêt propre, peut être utile :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}^2 = M_2 - M_1^2. \quad (3.33)$$

Pour la vérifier, il suffit de noter que pour tout $\omega \in \Omega$ fixé, $S^2(\omega)$ est la variance de la loi de probabilité $\mu_n(\omega)$, $M_1(\omega) = \overline{X}(\omega)$ est son espérance et $M_2(\omega)$ est l'espérance du carré d'une v.a. de loi $\mu_n(\omega)$. Donc en appliquant la formule de Koenig pour la variance de la loi $\mu_n(\omega)$, il vient $S^2(\omega) = M_2(\omega) - M_1(\omega)^2$. Ceci étant vrai pour tout $\omega \in \Omega$, on en déduit (3.33). Si cette méthode ne vous a pas plu, libre à vous de retrouver (3.33) par un calcul algébrique.

3.3.2 Espérance et variance des moments empiriques

Les moments empiriques étant des variables aléatoires peuvent avoir des...moments ! Nous allons examiner les deux exemples les plus importants, la moyenne empirique et la variance empirique. Insistons encore une fois : moyenne empirique et variance empirique sont des moments relatifs à la mesure de probabilité *aléatoire* μ_n , donc sont des variables aléatoires. Ce qui nous nous intéresse maintenant, ce sont les moments de ces variables aléatoires, en particulier espérance et variance, relativement à la mesure de probabilité P .

Proposition 3.15. Soient \overline{X} et S^2 la moyenne et la variance empiriques de l'échantillon X_1, \dots, X_n .

a) Si $\mathbf{E}|X_1| < +\infty$,

$$\mathbf{E}\overline{X} = \mathbf{E}X_1. \quad (3.34)$$

b) Si $\mathbf{E}X_1^2 < +\infty$,

$$\mathrm{Var}(\overline{X}) = \frac{1}{n} \mathrm{Var} X_1. \quad (3.35)$$

$$\mathbf{E}(S^2) = \left(1 - \frac{1}{n}\right) \mathrm{Var} X_1. \quad (3.36)$$

c) Si $\mathbf{E}(X_1^4) < +\infty$, en notant $\sigma^2 := \mathrm{Var} X_1$, $\tau := \mathbf{E}(X_1 - \mathbf{E}X_1)^4$ et si $\tau \neq \sigma^4$,

$$\mathrm{Var}(S^2) \sim \frac{1}{n}(\tau - \sigma^4), \quad \text{quand } n \rightarrow +\infty. \quad (3.37)$$

Preuve. Si $\mathbf{E}|X_1| < +\infty$, $\mathbf{E}X_1$ existe et comme les X_i ont même loi, les $\mathbf{E}X_i$ existent et sont toutes égales à $\mathbf{E}X_1$. Par linéarité de l'espérance, on en déduit que $\mathbf{E}(\overline{X})$ existe et se calcule comme suit :

$$\mathbf{E}\overline{X} = \mathbf{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}X_i = \frac{1}{n} n \mathbf{E}X_1 = \mathbf{E}X_1.$$

Supposons maintenant que $\mathbf{E}X_1^2 < +\infty$, alors tous les X_i ont des moments d'ordre 2, ce qui légitime l'existence de $\mathbf{E}S^2$ et de $\mathrm{Var} \overline{X}$. Par équidistribution¹¹ des X_i , on a pour tout i , $\mathbf{E}X_i^2 = \mathbf{E}X_1^2$ et $\mathrm{Var} X_i = \mathrm{Var} X_1$. En utilisant la formule de calcul de la variance d'une somme de v.a. indépendantes, on obtient :

$$\mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathrm{Var} X_i = \frac{1}{n^2} n \mathrm{Var} X_1 = \frac{1}{n} \mathrm{Var} X_1,$$

ce qui établit (3.35). Pour calculer $\mathbf{E}(S^2)$, on prend l'espérance des deux membres dans (3.33), ce qui donne

$$\mathbf{E}(S^2) = \mathbf{E}M_2 - \mathbf{E}(M_1^2) = \mathbf{E}\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) - \mathbf{E}(\overline{X}^2) = \mathbf{E}(X_1^2) - \mathbf{E}(\overline{X}^2),$$

par équidistribution des X_i . On applique ensuite la formule de Koenig « à l'envers », c.-à-d. $\mathbf{E}(Y^2) = \mathrm{Var} Y + (\mathbf{E}Y)^2$ avec la v.a. $Y = \overline{X}$ pour obtenir :

$$\begin{aligned} \mathbf{E}(S^2) &= \mathbf{E}(X_1^2) - \mathrm{Var} \overline{X} - (\mathbf{E}\overline{X})^2 \\ &= \mathbf{E}(X_1^2) - \frac{1}{n} \mathrm{Var} X_1 - (\mathbf{E}X_1)^2 \quad (\text{voir (3.35) et (3.34)}) \\ &= \left(1 - \frac{1}{n}\right) \mathrm{Var} X_1, \end{aligned}$$

ce qui établit (3.36).

Pour vérifier (3.37), on peut montrer après un calcul assez pénible que

$$\mathrm{Var}(S^2) = \frac{1}{n}(\tau - \sigma^4) + \frac{2}{n^2}(2\sigma^4 - \tau) + \frac{1}{n^3}(\tau - 3\sigma^4).$$

□

11. Rappelons qu'une suite de v.a. est dite équidistribuée lorsque ses termes ont même loi.

3.4 Lois des moments empiriques

Quand on veut approximer les moments d'une loi par ses moments empiriques, il est utile d'avoir des informations sur l'erreur commise. La proposition 3.15 nous donne une première information. Nous allons maintenant essayer d'en savoir plus dans deux situations typiques. Si on a un grand échantillon, on utilise la convergence en loi de l'erreur commise (après normalisation). Pour un échantillon de petite taille, le calcul exact de la loi de l'erreur est le plus souvent impossible, sauf dans le cas important d'un échantillon gaussien.

3.4.1 Échantillon de grande taille

Par (3.31), on sait déjà que \bar{X} converge p.s. vers $\mathbf{E}X_1$ lorsque n tend vers l'infini. Pour le comportement en loi de l'erreur d'approximation $\bar{X} - \mathbf{E}X_1$, on s'attend à pouvoir utiliser un théorème limite central. Supposons donc que $\mathbf{E}X_1^2 < +\infty$. Par le TLC version i.i.d., en posant comme d'habitude $S_n := X_1 + \dots + X_n$ et en notant que $S_n = n\bar{X}$, la somme centrée réduite S_n^* s'écrit

$$S_n^* = \frac{S_n - \mathbf{E}S_n}{\sqrt{\text{Var } S_n}} = \frac{n\bar{X} - n\mathbf{E}X_1}{\sqrt{n\sigma^2}} = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mathbf{E}X_1).$$

Le TLC appliqué à \bar{X} s'écrit donc comme suit.

Théorème 3.16 (TLC pour la moyenne empirique). *Si \bar{X} est la moyenne empirique d'un échantillon X_1, \dots, X_n tel que $\mathbf{E}X_1^2 < +\infty$ et $\sigma^2 := \text{Var } X_1 > 0$,*

$$\frac{\sqrt{n}}{\sigma}(\bar{X} - \mathbf{E}X_1) \xrightarrow[n \rightarrow +\infty]{\text{loi}} Z, \quad (3.38)$$

où Z suit la loi gaussienne standard $\mathfrak{N}(0, 1)$.

L'interprétation pratique de cet énoncé est que pour les grandes valeurs de n , la loi de \bar{X} est approximativement gaussienne $\mathfrak{N}(\mathbf{E}X_1, \sigma n^{-1/2})$. Si on connaît σ , on peut utiliser ceci pour proposer des intervalles de confiance pour le paramètre inconnu $\theta = \mathbf{E}X_1$. Par exemple le théorème 3.16 légitime pour n grand l'approximation :

$$P\left(\frac{\sqrt{n}}{\sigma}|\bar{X} - \mathbf{E}X_1| \leq 2\right) \simeq 2\Phi(2) - 1 \simeq 0,9544,$$

ce qui peut aussi s'écrire

$$P(\bar{X} - 2\sigma n^{-1/2} \leq \theta \leq \bar{X} + 2\sigma n^{-1/2}) \simeq 0,9544.$$

On en déduit l'intervalle de confiance suivant pour θ au niveau 95,4% :

$$[\bar{X} - 2\sigma n^{-1/2}, \bar{X} + 2\sigma n^{-1/2}].$$

En général, on ne connaît pas σ et on le remplace par $S := \sqrt{S^2}$. La légitimation théorique de cette recette est fournie par le théorème suivant.

Théorème 3.17 (TLC avec autonormalisation). Si \bar{X} et S^2 sont respectivement la moyenne et la variance empiriques d'un échantillon X_1, \dots, X_n tel que $\mathbf{E}X_1^2 < +\infty$ et $\sigma^2 := \text{Var } X_1 > 0$,

$$T_n := \sqrt{n} \frac{\bar{X} - \mathbf{E}X_1}{\sqrt{S^2}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y, \quad (3.39)$$

où Y suit la loi gaussienne standard $\mathfrak{N}(0, 1)$.

Preuve. En toute rigueur, cet énoncé est incorrect car il est possible que S^2 prenne la valeur 0 avec probabilité non-nulle et alors T_n n'est pas définie. Pour simplifier, nous supposons dans un premier temps qu'il existe un rang $n_0 \geq 2$ (pourquoi 2?) à partir duquel $\{S^2 = 0\} = \emptyset$, en sorte que pour tout $n \geq n_0$, T_n est bien définie sur tout Ω et est une variable aléatoire¹².

L'idée est de traiter séparément \bar{X} et S^2 , le premier grâce au théorème 3.16, le deuxième grâce à la LFGN. Pour cela on commence par écrire :

$$\sqrt{n} \frac{\bar{X} - \mathbf{E}X_1}{\sqrt{S^2}} = \frac{\sqrt{n}}{\sigma} (\bar{X} - \mathbf{E}X_1) \sqrt{\frac{\sigma^2}{S^2}} = S_n^* \sqrt{\frac{\sigma^2}{S^2}}.$$

Par le théorème 3.16, S_n^* converge en loi vers Y gaussienne $\mathfrak{N}(0, 1)$.

Par (3.32) qui n'est qu'une simple application de la LFGN à la suite i.i.d. et intégrable $(X_i^2)_{i \geq 1}$, S^2 converge presque-sûrement vers σ^2 . On utilise alors la remarque suivante. Si $(V_n)_{n \geq 1}$ est une suite de variables aléatoires qui converge p.s. vers une constante c et f une fonction définie sur un voisinage de c et continue au point c , alors $f(V_n)$ converge presque sûrement vers $f(c)$. La justification est laissée en exercice. En appliquant ceci à la fonction $f : x \mapsto \sigma x^{-1/2}$ qui est continue sur \mathbb{R}^* , donc en particulier continue au point $c = \sigma^2 > 0$, on obtient :

$$\sqrt{\frac{\sigma^2}{S^2}} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 1. \quad (3.40)$$

Arrivés là, nous avons besoin d'un lemme qui est un outil important dans les problèmes de convergence en loi. Nous différons légèrement sa preuve.

Lemme 3.18 (Slutsky). Soient $(Y_n)_{n \geq 1}$ et $(Z_n)_{n \geq 1}$ deux suites de v.a. définies sur le même espace probabilisé et telles que

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y \quad \text{et} \quad Z_n \xrightarrow[n \rightarrow +\infty]{\text{Pr}} c,$$

où c est une constante. Alors

$$(Y_n, Z_n) \xrightarrow[n \rightarrow +\infty]{\text{loi}} (Y, c). \quad (3.41)$$

12. Par la loi forte des grands nombres, S^2 converge p.s. vers σ^2 , donc presque sûrement à partir d'un certain rang *aléatoire* $n_1 = n_1(\omega)$, on a $S^2 > \sigma^2/2 > 0$. Malheureusement, comme ce n_1 est aléatoire, on ne peut pas utiliser cet argument pour dire que T_n est bien définie comme variable aléatoire (sauf peut-être sur un événement de probabilité nulle) à partir d'un certain rang.

Nous appliquons le lemme de Slutsky à la convergence en loi d'un produit. En effet la convergence en loi des vecteurs aléatoires étant conservée par image continue, cf. proposition 1.7, $g(Y_n, Z_n)$ converge en loi vers $g(Y, c)$ pour toute fonction continue $g : \mathbb{R}^2 \rightarrow \mathbb{R}$. En appliquant ceci avec $Y_n = S_n^*$, $Z_n = \sigma S^{-1}$, $c = 1$, $g : (y, z) \mapsto yz$ et en rappelant que la convergence p.s. implique la convergence en probabilité, on obtient (3.39).

Pour compléter la preuve, il nous reste à définir proprement T_n sans utiliser l'hypothèse $\{S^2 = 0\} = \emptyset$ pour $n \geq n_0$ et à vérifier la convergence en loi de ce « nouveau » T_n vers Y . Posons donc pour tout $\omega \in \Omega$,

$$T_n(\omega) := \begin{cases} \sqrt{n} \frac{\bar{X}(\omega) - \mathbf{E}X_1}{\sqrt{S^2(\omega)}} & \text{si } S^2(\omega) \neq 0, \\ 0 & \text{si } S^2(\omega) = 0. \end{cases}$$

Avec la convention habituelle sur les indicatrices dans les variables aléatoires la valeur nulle d'une indicatrice s'impose toujours lorsqu'elle est facteur d'une quantité infinie ou non définie, en clair :

$$\sqrt{\frac{\sigma^2}{S^2}} \mathbf{1}_{\{S^2 \neq 0\}} = \begin{cases} \sqrt{\frac{\sigma^2}{S^2}} & \text{sur } \{S^2 \neq 0\}, \\ 0 & \text{sur } \{S^2 = 0\}. \end{cases}$$

avec cette convention, le T_n redéfini proprement ci-dessus peut s'écrire :

$$T_n = \frac{\sqrt{n}}{\sigma} (\bar{X} - \mathbf{E}X_1) \left(\sqrt{\frac{\sigma^2}{S^2}} \mathbf{1}_{\{S^2 \neq 0\}} \right).$$

Il est alors immédiat d'adapter la démonstration donnée dans le cas où $\{S^2 = 0\}$ est vide en notant que $\mathbf{1}_{\{S^2 \neq 0\}}$ converge p.s. vers 1 en raison de la loi forte des grands nombres. \square

Dans la preuve du lemme de Slutsky, nous aurons besoin du résultat suivant qui peut être vu comme un cas particulier de « convergence dominée en probabilité ».

Lemme 3.19. *Soit $(W_n)_{n \geq 1}$ une suite de variables aléatoires réelles vérifiant*

a) il existe une constante $M \in \mathbb{R}_+$ telle que

$$\forall \omega \in \Omega, \forall n \in \mathbb{N}^*, \quad |W_n(\omega)| \leq M;$$

b) W_n converge en probabilité vers 0 quand n tend vers $+\infty$.

Alors $\mathbf{E}|W_n|$ tend vers 0 quand n tend vers $+\infty$.

Preuve. Le a) nous assure que $P(|W_n| > t) = 0$ pour tout $t > M$. En écrivant la définition de l'espérance de la v.a. positive $|W_n|$, on en déduit que

$$\mathbf{E}|W_n| = \int_0^{+\infty} P(|W_n| > t) dt = \int_0^M P(|W_n| > t) dt.$$

Fixant ε arbitraire dans $]0, M[$, nous avons donc

$$\mathbf{E}|W_n| = \int_0^\varepsilon P(|W_n| > t) dt + \int_\varepsilon^M P(|W_n| > t) dt.$$

Majorons la probabilité dans \int_0^ε par 1 et majorons la dans \int_ε^M par $P(|W_n| > \varepsilon)$ puisqu'il s'agit d'une fonction décroissante de t . Il vient ainsi

$$\mathbf{E}|W_n| \leq \int_0^\varepsilon dt + \int_\varepsilon^M P(|W_n| > \varepsilon) dt = \varepsilon + (M - \varepsilon)P(|W_n| > \varepsilon).$$

L'hypothèse b) de convergence en probabilité nous donne pour tout $\delta > 0$ un entier $N = N(\delta)$ tel que pour tout $n \geq N$, $P(|W_n| > \varepsilon) < \delta$. Ceci est vrai en particulier avec $\delta = \varepsilon$, d'où

$$\forall n \geq N(\varepsilon), \quad \mathbf{E}|W_n| < \varepsilon + M\varepsilon = (M + 1)\varepsilon.$$

Comme ceci est vrai pour tout $\varepsilon \in]0, M[$, on conclut à la convergence vers 0 de $\mathbf{E}|W_n|$. \square

À titre d'exercice, vous pouvez démontrer la version plus générale du lemme 3.19 obtenue en remplaçant l'hypothèse a) par « pour tout $n \geq 1$, $|W_n| \leq M$ presque-sûrement ».

Preuve du lemme de Slutsky. En revenant à la définition de la convergence en loi des vecteurs aléatoires, il s'agit de montrer que si

$$Y_n \xrightarrow[n \rightarrow +\infty]{\text{loi}} Y \quad \text{et} \quad Z_n \xrightarrow[n \rightarrow +\infty]{\text{Pr}} c, \quad (c \text{ constante})$$

alors

$$\forall h \in \mathcal{C}_b(\mathbb{R}^2), \quad \mathbf{E}h(Y_n, Z_n) \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}h(Y, c), \quad (3.42)$$

où $\mathcal{C}_b(\mathbb{R}^2)$ désigne l'espace des fonctions continues bornées sur \mathbb{R}^2 . En fait il suffit de vérifier (3.42) pour des fonctions h de la forme particulière $h(y, z) = f(y)g(z)$ avec $f, g \in \mathcal{C}_b(\mathbb{R})$. Nous admettrons ce point qui pourrait se démontrer en utilisant l'idée de la preuve du corollaire 1.43 combinée avec un corollaire du théorème de Stone Weierstrass, après compactification de \mathbb{R}^2 par l'ajout d'un seul point à l'infini. On réduit ainsi la preuve du lemme de Slutsky à celle de la convergence suivante :

$$\forall f \in \mathcal{C}_b(\mathbb{R}), \forall g \in \mathcal{C}_b(\mathbb{R}) \quad \mathbf{E}(f(Y_n)g(Z_n)) \xrightarrow[n \rightarrow +\infty]{} \mathbf{E}(f(Y)g(c)). \quad (3.43)$$

Comme f et g sont bornées, les v.a. $f(Y_n)$, $f(Y)$, $g(Z_n)$ ainsi que leurs produits et leurs différences le sont aussi. Ceci légitime l'existence de toutes les espérances figurant dans les calculs ci-dessous. En utilisant la linéarité de l'espérance et la constance de $g(c)$, on peut écrire :

$$\begin{aligned} \mathbf{E}(f(Y_n)g(Z_n)) - \mathbf{E}(f(Y)g(c)) &= \mathbf{E}\left(f(Y_n)(g(Z_n) - g(c)) + g(c)(f(Y_n) - f(Y))\right) \\ &= \mathbf{E}\left(f(Y_n)(g(Z_n) - g(c))\right) + g(c)(\mathbf{E}f(Y_n) - \mathbf{E}f(Y)). \end{aligned}$$

On en déduit la majoration

$$|\mathbf{E}(f(Y_n)g(Z_n)) - \mathbf{E}(f(Y)g(c))| \leq \|f\|_\infty \mathbf{E}|g(Z_n) - g(c)| + |g(c)| |\mathbf{E}f(Y_n) - \mathbf{E}f(Y)|.$$

Comme g est continue, la convergence en probabilité de Z_n vers c implique celle de $g(Z_n)$ vers $g(c)$, d'après un exercice vu en travaux dirigés¹³. Cette convergence en probabilité équivaut à celle de $W_n := g(Z_n) - g(c)$ vers 0 et comme g est bornée, on peut appliquer le lemme 3.19 avec $M = 2\|g\|_\infty$ pour obtenir :

$$\|f\|_\infty \mathbf{E}|g(Z_n) - g(c)| \xrightarrow{n \rightarrow +\infty} 0. \quad (3.44)$$

D'autre part il résulte immédiatement de la définition de la convergence en loi de Y_n vers Y que

$$|g(c)| |\mathbf{E}f(Y_n) - \mathbf{E}f(Y)| \xrightarrow{n \rightarrow +\infty} 0. \quad (3.45)$$

De (3.44) et (3.45) on déduit la convergence (3.43), ce qui achève la preuve. \square

Exemple 3.20. Revenons sur la question de l'intervalle de confiance pour une probabilité inconnue p . Dans ce cas, l'intervalle de confiance au niveau 95% pour p résultant de l'application du TLC avec autonormalisation est

$$J_n = [\bar{X} - 1,96Sn^{-1/2}, \bar{X} + 1,96Sn^{-1/2}],$$

où S est la racine carrée positive de la variance empirique de l'échantillon de taille n de v.a. de Bernoulli X_i (de paramètre inconnu p). Comme chaque X_i ne peut prendre que les valeurs 0 ou 1, $X_i = X_i^2$, d'où

$$S^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i - (\bar{X})^2 = \bar{X} - (\bar{X})^2 = \bar{X}(1 - \bar{X}).$$

On retrouve bien l'intervalle de confiance J_n proposé à la page 17, puisque $M_n = \bar{X}$.

3.4.2 Échantillon gaussien

Dans le cas d'un échantillon gaussien, les lois de \bar{X} et de S^2 sont calculables, de même que celle du vecteur (\bar{X}, S^2) .

Théorème 3.21 (Student, 1908). *Soit X_1, \dots, X_n un n -échantillon de la loi $\mathfrak{N}(m, \sigma)$. Alors*

- a) *La moyenne empirique \bar{X} et la variance empirique S^2 sont indépendantes.*
- b) *\bar{X} suit la loi gaussienne $\mathfrak{N}(m, \sigma n^{-1/2})$.*
- c) *$\frac{nS^2}{\sigma^2}$ suit la loi du khi-deux (χ^2) de Pearson à $n-1$ degrés de liberté, notée $\chi^2(n-1)$. Cette loi est tabulée.*

13. Sinon essayez de le démontrer tout seuls avant de demander à vos enseignant(e)s.

Nous admettrons le a) et le c).

Preuve du b). Puisque X_1, \dots, X_n sont indépendantes et gaussiennes, le vecteur aléatoire (X_1, \dots, X_n) est gaussien, cf. corollaire 1.27. Il en résulte d'après la définition des vecteurs gaussiens que la combinaison linéaire

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

de ses composantes est une v.a. gaussienne, cf. définition 1.23. Les paramètres de sa loi sont alors l'espérance et l'écart type de \bar{X} , données par (3.34) et (3.35) :

$$\mathbf{E}\bar{X} = \mathbf{E}X_1 = m, \quad (\text{Var } \bar{X})^{1/2} = \left(\frac{\text{Var } X_1}{n} \right)^{1/2} = \sigma n^{-1/2}.$$

La loi de \bar{X} est donc bien $\mathfrak{N}(m, \sigma n^{-1/2})$. □

Corollaire 3.22. Soit X_1, \dots, X_n un échantillon de la loi $\mathfrak{N}(m, \sigma)$. Posons

$$U := \frac{\sqrt{n-1}}{S} (\bar{X} - m).$$

Alors U suit la loi de Student à $n-1$ degrés de liberté. Cette loi est tabulée.

D'où sort cette v.a. U ? L'idée de départ est de trouver une variable aléatoire de la forme $V = f_n(X_1, \dots, X_n)$, permettant de construire des intervalles de confiance pour le paramètre inconnu m . Pour cela il est nécessaire que la loi de cette variable ne dépende pas du ou des paramètres inconnus. Lorsque l'on connaît la valeur de σ , la candidature de $V := \sqrt{n/\sigma^2}(\bar{X} - m)$ est assez naturelle. En effet V est gaussienne comme combinaison linéaire des composantes X_i du vecteur gaussien (X_1, \dots, X_n) . Son espérance est $\sqrt{n/\sigma^2} \mathbf{E}(\bar{X} - m) = 0$. Sa variance est $n\sigma^{-2} \text{Var}(\bar{X} - m) = n\sigma^{-2} \text{Var } \bar{X} = n\sigma^{-2} n^{-2} n\sigma^2 = 1$. Ainsi V suit la loi $\mathfrak{N}(0, 1)$. Notons d'ailleurs que V est exactement le S_n^* du théorème 3.16. La différence est que dans ce théorème, S_n^* n'est généralement pas gaussienne, c'est seulement sa limite en loi qui l'est. Ici V est gaussienne pour tout $n \geq 1$. L'intérêt de V est le même que celui de S_n^* : permettre la construction d'intervalles de confiance pour m inconnue lorsque σ est connu. L'avantage du cas gaussien est que cette construction est pertinente même avec de petits échantillons. Le cas où σ^2 est inconnu nous amène à nouveau à le remplacer par un *estimateur* de la variance. Dans le théorème limite central avec autonormalisation, nous avons remplacé σ^2 par la variance empirique S^2 . Nous avons vu que S^2 converge p.s. vers σ^2 et que $\mathbf{E}S^2 = (1 - 1/n)\sigma^2$. Ici on travaille avec n fixé, éventuellement petit et on préfère remplacer S^2 par $(1 - 1/n)^{-1}S^2$. L'estimateur de σ^2 ainsi obtenu est considéré comme meilleur que le précédent parce que son espérance est exactement σ^2 . On parle d'estimateur *sans biais*, cf. le chapitre sur l'estimation. Ceci nous amène à choisir

$$f_n(X_1, \dots, X_n) = \sqrt{\frac{n}{\frac{n}{n-1}S^2}} (\bar{X} - m) = \frac{\sqrt{n-1}}{S} (\bar{X} - m) = U.$$

Voici une application du théorème de Student et de son corollaire à la construction d'intervalles de confiance pour les paramètres inconnus d'une loi gaussienne.

Exemple 3.23 (intervalle de confiance pour m avec σ inconnu, cas gaussien). Pour construire un tel intervalle de confiance au niveau $1 - \varepsilon$, par exemple au niveau 95%, on commence par chercher en utilisant la table de la loi de Student, un intervalle $[a, b]$ le plus court possible tel que

$$P(U \in [a, b]) \geq 1 - \varepsilon. \quad (3.46)$$

En revenant à la définition de U dans le corollaire 3.22, on voit que l'on a

$$a \leq U \leq b \Leftrightarrow \frac{aS}{\sqrt{n-1}} \leq \bar{X} - m \leq \frac{bS}{\sqrt{n-1}},$$

d'où

$$P\left(\frac{aS}{\sqrt{n-1}} \leq \bar{X} - m \leq \frac{bS}{\sqrt{n-1}}\right) \geq 1 - \varepsilon.$$

On en déduit qu'un intervalle de confiance au niveau $1 - \varepsilon$ pour m est

$$\left[\bar{X} - \frac{bS}{\sqrt{n-1}}, \bar{X} - \frac{aS}{\sqrt{n-1}}\right].$$

Cet intervalle peut se calculer à partir des seules observations $X_1(\omega), \dots, X_n(\omega)$, sans avoir besoin de connaître l'autre paramètre inconnu σ . On voit bien sur cette formule l'intérêt de choisir l'intervalle $[a, b]$ de longueur minimale parmi tous les intervalles vérifiant (3.46). En effet la longueur de l'intervalle de confiance est proportionnelle à $(b - a)$. Nous verrons ci-dessous que toute loi de Student a une densité paire, maximale en zéro. On en déduit facilement que l'intervalle $[a, b]$ de longueur minimale vérifiant (3.46) est symétrique autour de 0, donc de la forme $[-c, c]$. Le meilleur intervalle de confiance au niveau $1 - \varepsilon$ pour m obtenu par cette méthode est donc

$$\left[\bar{X} - \frac{cS}{\sqrt{n-1}}, \bar{X} + \frac{cS}{\sqrt{n-1}}\right], \text{ avec } P(|U| \leq c) = 2P(U \leq c) - 1 = 1 - \varepsilon.$$

Exemple 3.24 (intervalle de confiance pour σ avec m inconnu, cas gaussien). Cette fois nous utilisons le c) du théorème de Student. Notons Z une v.a. de loi $\chi^2(n-1)$. Avec la table de la loi $\chi^2(n-1)$, on cherche un intervalle $[\alpha, \beta] \subset \mathbb{R}_+^*$ tel que $P(Z \in [\alpha, \beta]) \geq 1 - \varepsilon$. En prenant grâce au c), $Z = nS^2\sigma^{-2}$, on en déduit que

$$P\left(\frac{nS^2}{\sigma^2} \in [\alpha, \beta]\right) \geq 1 - \varepsilon. \quad (3.47)$$

Il suffit ensuite de « résoudre » l'encadrement $\alpha \leq nS^2\sigma^{-2} \leq \beta$ en un encadrement de σ^2 pour réécrire (3.47) sous la forme

$$P\left(\frac{nS^2}{\beta} \leq \sigma^2 \leq \frac{nS^2}{\alpha}\right) \geq 1 - \varepsilon.$$

Ceci nous permet de proposer pour σ l'intervalle de confiance

$$[Sn^{1/2}\beta^{-1/2}, Sn^{1/2}\alpha^{-1/2}]$$

au niveau $1 - \varepsilon$. Pour obtenir le meilleur intervalle de confiance au niveau $1 - \varepsilon$, il faudrait donc trouver le couple (α, β) de réels positifs qui minimise $\alpha^{-1/2} - \beta^{-1/2}$ parmi ceux vérifiant $P(Z \in [\alpha, \beta]) \geq 1 - \varepsilon$. Ce problème est beaucoup moins simple que son analogue de l'exemple 3.23. On peut le résoudre numériquement. En pratique on se contente du choix sous-optimal de α, β tels que

$$P(Z \leq \alpha) = \frac{\varepsilon}{2}, \quad P(Z > \beta) = \frac{\varepsilon}{2}.$$

Il est temps de faire plus ample connaissance avec les lois du χ^2 et de Student.

Définition 3.25 (lois de χ^2). *On appelle loi du khi-deux à d degrés de liberté et on note $\chi^2(d)$ la loi de la v.a.*

$$Z = Y_1^2 + \cdots + Y_d^2,$$

où les d variables aléatoires Y_i sont i.i.d. $\mathfrak{N}(0, 1)$.

Sans en savoir plus, on peut déjà voir que

$$\text{si } Z \text{ a pour loi } \chi^2(d), \quad \mathbf{E}Z = d. \quad (3.48)$$

En effet, $\mathbf{E}Y_1^2 = \text{Var } Y_1 = 1$ car Y_1 est centrée et de variance 1, d'où par linéarité de l'espérance et équidistribution des Y_i^2 , $\mathbf{E}Z = d\mathbf{E}Y_1^2 = d$.

On peut montrer que la loi $\chi^2(d)$ admet pour densité

$$f : t \mapsto \frac{1}{2^{d/2}\Gamma(d/2)} t^{d/2-1} e^{-t/2} \mathbf{1}_{\mathbb{R}_+}(t), \quad (3.49)$$

avec

$$\Gamma(\alpha) := \int_0^{+\infty} t^{\alpha-1} e^{-t} dt, \quad \alpha \in]0, +\infty[. \quad (3.50)$$

Définition 3.26 (lois de Student). *La loi de Student à d degrés de libertés est la loi de la v.a.*

$$T := \frac{Y_0}{\sqrt{\frac{1}{d}(Y_1^2 + \cdots + Y_d^2)}},$$

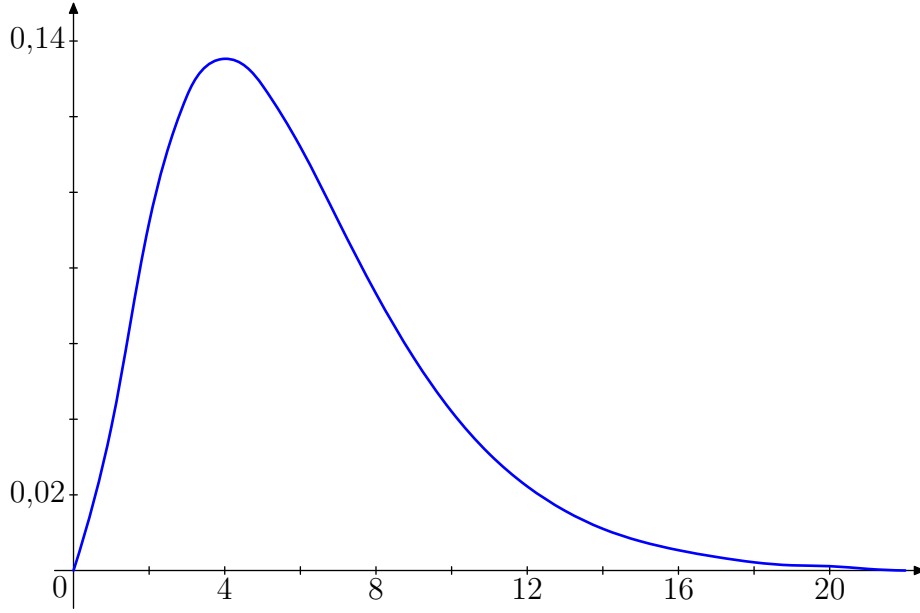
où les Y_0, Y_1, \dots, Y_d sont i.i.d. gaussiennes $\mathfrak{N}(0, 1)$.

Là encore, il est possible de calculer simplement l'espérance de T à partir de cette définition. Pour cela nous admettrons que

$$\forall d \geq 2, \quad \mathbf{E}((Y_1^2 + \cdots + Y_d^2)^{-1/2}) < +\infty.$$

Par indépendance de Y_0 et (Y_1, \dots, Y_d) et nullité de l'espérance de la gaussienne standard Y_0 , on en déduit :

$$\forall d \geq 2, \quad \mathbf{E}T = \mathbf{E}Y_0 \mathbf{E}(d^{1/2}(Y_1^2 + \cdots + Y_d^2)^{-1/2}) = 0. \quad (3.51)$$

FIG. 3.6 – Densité de la loi $\chi^2(6)$

On peut montrer que pour tout $d \in \mathbb{N}^*$, la loi de Student à d degrés de liberté admet pour densité

$$g : t \mapsto \frac{1}{\sqrt{d\pi}} \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} \left(1 + \frac{t^2}{d}\right)^{-\frac{d+1}{2}}, \quad (3.52)$$

où Γ est définie en (3.50). Rappelons que $\Gamma(1/2) = \sqrt{\pi}$, d'où

$$\forall k \in \mathbb{N}^*, \quad \Gamma\left(k + \frac{1}{2}\right) = \left(k - \frac{1}{2}\right)\left(k - \frac{3}{2}\right) \dots \frac{1}{2}\sqrt{\pi}, \quad (3.53)$$

en utilisant la relation bien connue $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ vérifiée pour tout réel $\alpha > 1$. Ceci permet d'expliciter le calcul de la constante $\Gamma((d+1)/2)/\Gamma(d/2)$ dans (3.52). On distingue pour cela les cas d pair et d impair.

$$\text{Si } d = 2k, k \in \mathbb{N}^*, \quad \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} = \frac{\Gamma(k + \frac{1}{2})}{\Gamma(k)} = \frac{(k - \frac{1}{2})(k - \frac{3}{2}) \dots \frac{1}{2}\sqrt{\pi}}{(k-1)!}. \quad (3.54)$$

$$\text{Si } d = 2k + 1, k \in \mathbb{N}^*, \quad \frac{\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2})} = \frac{\Gamma(k + 1)}{\Gamma(k + \frac{1}{2})} = \frac{k!}{(k - \frac{1}{2})(k - \frac{3}{2}) \dots \frac{1}{2}\sqrt{\pi}}. \quad (3.55)$$

Pour $d = 1$, on a directement $\Gamma((d+1)/2)/\Gamma(d/2) = \Gamma(1)/\Gamma(1/2) = 0!/\sqrt{\pi} = \pi^{-1/2}$. Ainsi la loi de Student à 1 degré de liberté n'est autre que la loi de Cauchy $\text{Cau}(0, 1)$. Cette loi n'a pas d'espérance. Dans tous les cas la loi de T est *symétrique*, ce qui signifie que T et $-T$ ont même loi. C'est clair sur la définition 3.26 ou par parité de g .

Chapitre 4

Estimation

Nous revenons maintenant à la question posée en introduction au chapitre 3 : à partir de l'observation d'un échantillon X_1, \dots, X_n associé à un modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$, comment « deviner quel est le bon θ » ? Plus précisément on s'intéresse généralement à un objet $\gamma = g(\theta)$ qui peut être de nature fonctionnelle (densité de la loi des X_i , f.d.r. des X_i , fonction de survie, ...) ou numérique (espérance, variance, médiane, ...) ou vectorielle fini-dimensionnelle (couple (m, σ) dans un modèle gaussien, bornes d'un intervalle de support d'une loi, triplet des paramètres d'une loi de Weibull, ...). Dans le premier cas on parle *d'estimation fonctionnelle*. Nous en avons vu un exemple important avec la f.d.r. empirique qui est un estimateur fonctionnel de la f.d.r. F des X_i . Par le théorème de Glivenko Cantelli, cet estimateur fonctionnel converge p.s. uniformément vers F . Nous nous occupons dans ce chapitre du cas où γ est fini-dimensionnel et pour simplifier, nous nous limitons dans les énoncés au cas où g est l'identité et $\gamma = \theta$ est un réel. En cas de besoin, par exemple si dans un modèle gaussien paramétré par $\theta = (m, \sigma)$, on s'intéresse à la variance $g(\theta) = \sigma^2$, il sera facile d'adapter les énoncés.

Nous avons déjà vu sur des exemples une technique d'estimation, celle des intervalles de confiance où, plutôt que de tenter de deviner la « vraie valeur » du paramètre inconnu θ , on essaie de l'encadrer avec une grande probabilité de succès entre des bornes calculées à partir de l'échantillon. On parle alors *d'estimation par intervalle de confiance*. Dans ce chapitre, nous étudions *l'estimation ponctuelle*. Il s'agit cette fois non plus de « parier » sur un intervalle censé contenir la bonne valeur θ , mais de proposer directement une valeur calculée à partir de l'échantillon pour représenter (plus ou moins bien) la valeur inconnue de θ .

4.1 Estimateurs

4.1.1 Exemples

Avant de donner les définitions relatives à l'estimation, il est utile de passer en revue quelques exemples introductifs. Pour chacun de ces exemples, on suppose que l'on a

un modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$ et on note X_1, \dots, X_n un échantillon associé à ce modèle.

Exemple 4.1 (estimation de l'espérance). Dans le cas où le paramètre θ inconnu est l'espérance de l'échantillon ($\mathbf{E}_\theta X_i = \theta$), un estimateur usuel de θ est la moyenne empirique

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Une des raisons de ce choix est que \bar{X} converge presque sûrement quand n tend vers l'infini vers $\mathbf{E}_\theta X_1 = \theta$ par la loi forte des grands nombres. On dit que \bar{X} est un estimateur *fortement consistant* de θ . Notons aussi que pour tout n , $\mathbf{E}_\theta \bar{X} = \theta$. On dit que \bar{X} est un estimateur *sans biais* de θ .

Exemple 4.2 (estimation de la variance). Lorsque le paramètre inconnu est la variance σ^2 , on peut l'estimer par la variance empirique :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2.$$

Nous avons vu que S^2 converge p.s. vers σ^2 , donc là encore cet estimateur est fortement consistant. Par contre, cf. proposition 3.15,

$$\forall \theta \in \Theta, \quad \mathbf{E}_\theta(S^2) = \left(1 - \frac{1}{n}\right) \text{Var}_\theta X_1 \neq \text{Var}_\theta X_1.$$

Ici l'espérance de l'estimateur n'est pas égale au paramètre à estimer, on dit que l'estimateur est *biaisé*.

Exemple 4.3 (estimation du support d'une loi uniforme). Soit X_1, \dots, X_n un échantillon de la loi uniforme sur $[0, \theta]$, où $\theta \in]0, +\infty[$ est inconnu. Voici un premier estimateur T_n de θ défini par

$$T_n := \max_{1 \leq i \leq n} X_i.$$

Il est facile de vérifier que T_n est un estimateur fortement consistant et biaisé¹. Voici un deuxième estimateur T'_n fortement consistant et sans biais :

$$T'_n = 2\bar{X} = \frac{2}{n} \sum_{i=1}^n X_i.$$

En effet si X_i suit la loi uniforme sur $[0, \theta]$ son espérance vaut $\theta/2$.

1. Il est assez probable que vous ayez déjà vu en exercice que T_n converge p.s. vers θ et que $\mathbf{E}_\theta T_n < \theta$. Sinon dépêchez vous de combler cette lacune.

4.1.2 Généralités

Dans les exemples vus ci-dessus, l'estimateur proposé est à chaque fois, une fonction des observations « proche », au moins pour les grandes valeurs de n , du paramètre qu'il est censé estimer. Cette idée de proximité est néanmoins trop imprécise pour être incorporée à la définition mathématique d'un estimateur, laquelle doit être valable pour toute valeur de n . On y renonce donc et il ne reste plus que la notion de fonction des observations, autrement dit de *statistique* au sens de la définition 3.5.

Définition 4.4 (estimateur). *Soit X_1, \dots, X_n un n -échantillon associé à un modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$, où Θ est une partie de \mathbb{R} . On appelle estimateur de θ associé à cet échantillon, toute v.a. T_n de la forme*

$$T_n = f_n(X_1, \dots, X_n)$$

où $f_n : \mathbb{R}^n \rightarrow \mathbb{R}$, $(t_1, \dots, t_n) \mapsto f_n(t_1, \dots, t_n)$ est une application borélienne ne dépendant pas de θ .

D'un point de vue formel, les définitions 3.5 et 4.4 sont équivalentes. La seule différence, non mathématique, est le *contexte* d'utilisation. L'expression *statistique* est plus générale car elle recouvre aussi bien les estimateurs que les statistiques de test.

Il faut bien avouer qu'à son niveau de généralité, la définition 4.4 a quelque chose de choquant car il semble finalement que T_n puisse n'avoir aucun rapport avec θ . À y regarder de plus près, on voit que le seul rapport de T_n avec θ , c'est que sa loi dépend de θ via la loi du vecteur aléatoire (X_1, \dots, X_n) . C'est néanmoins bien peu et on se dépêche de compléter la définition 4.4 en définissant des propriétés qui permettent de dire que certains estimateurs sont moins mauvais que d'autres.

Définition 4.5 (estimateur faiblement consistant). *Soit T_n un estimateur de θ . On dit qu'il est faiblement consistant s'il converge en probabilité vers θ quand n tend vers l'infini.*

Bien sûr, cette définition contient un grossier abus de langage. Il faudrait dire « la suite d'estimateurs $(T_n)_{n \geq 1}$ est faiblement consistante si... ». D'autre part, rappelons que nous travaillons avec un modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$. Dans ce cadre, la convergence en probabilité de T_n vers θ signifie très précisément :

$$\forall \theta \in \Theta, \quad \forall \varepsilon > 0, \quad P_\theta(|T_n - \theta| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Avec le même abus de langage que ci-dessus, on définit la consistance forte.

Définition 4.6 (estimateur fortement consistant). *Soit T_n un estimateur de θ . On dit qu'il est fortement consistant s'il converge presque-sûrement vers θ quand n tend vers l'infini.*

Ici la convergence presque-sûre de T_n vers θ signifie :

$$\forall \theta \in \Theta, \quad P_\theta\left(\lim_{n \rightarrow +\infty} T_n = \theta\right) = 1.$$

Fixons pour un moment n et notons $T = T_n$. L'erreur d'estimation est la v.a. $T - \theta$. On suppose ici que $\mathbf{E}_\theta|T| < +\infty$ pour tout $\theta \in \Theta$, ce qui entraîne l'existence de $\mathbf{E}_\theta T$. On peut alors décomposer l'erreur d'estimation comme suit :

$$T - \theta = (T - \mathbf{E}_\theta T) + (\mathbf{E}_\theta T - \theta). \quad (4.1)$$

Le premier terme d'erreur $T - \mathbf{E}_\theta T$ est aléatoire et provient inévitablement des fluctuations de la v.a. T « autour » de son espérance. Le deuxième terme $(\mathbf{E}_\theta T - \theta)$ est déterministe et représente une *erreur systématique*, dont on pourrait se débarrasser en ajoutant une constante convenable² à T .

Définition 4.7 (biais). *Soit T un estimateur de θ . Si $\mathbf{E}_\theta T$ existe pour tout $\theta \in \Theta$, on appelle biais de l'estimateur T la quantité $(\mathbf{E}_\theta T - \theta)$. De plus*

- si pour tout $\theta \in \Theta$, $\mathbf{E}_\theta T = \theta$, on dit que T est un estimateur sans biais de θ ;
- si $\mathbf{E}_\theta T \neq \theta$ pour au moins un $\theta \in \Theta$, on dit que T est un estimateur biaisé de θ ;
- si $(T_n)_{n \geq 1}$ est une suite d'estimateurs telle que pour tout $\theta \in \Theta$, $\mathbf{E}_\theta T_n$ converge vers θ quand n tend vers l'infini, on dit que T_n est asymptotiquement sans biais.

Cette définition du biais contient encore un abus de langage, puisqu'on considère le biais comme un nombre réel, alors qu'il s'agit en fait de la fonction :

$$b : \Theta \rightarrow \mathbb{R}, \quad \theta \mapsto b(\theta) = \mathbf{E}_\theta T - \theta.$$

Dans l'exemple 4.2, la variance empirique S^2 est un estimateur biaisé mais asymptotiquement sans biais de σ^2 . Il en va de même pour l'estimateur T_n de la borne θ dans l'exemple 4.3.

4.1.3 Erreur quadratique moyenne

Pour mesurer la proximité de l'estimateur T au paramètre θ , on utilise une *fonction de perte*, le plus souvent de la forme $g(\|T - \theta\|)$, où g est une fonction croissante et $\|\cdot\|$ une norme sur l'espace des variables aléatoires. L'une des fonctions de perte les plus courantes est l'*erreur quadratique moyenne*.

Définition 4.8 (erreur quadratique moyenne EQM). *On suppose que l'estimateur T du paramètre θ vérifie $\mathbf{E}_\theta T^2 < +\infty$ pour tout $\theta \in \Theta$. On appelle erreur quadratique moyenne de T , en abrégé EQM, la quantité*

$$\text{EQM}(T) := \mathbf{E}_\theta (T - \theta)^2.$$

Là encore, il serait plus correct de parler de la *fonction* :

$$\text{EQM} : \Theta \rightarrow \mathbb{R}_+, \quad \theta \mapsto \mathbf{E}_\theta (T - \theta)^2.$$

Dans la littérature anglo-américaine, l'EQM est notée MSE (*mean square error*).

2. À condition d'être capable de calculer $\mathbf{E}_\theta T - \theta$, ce qui n'est pas toujours le cas en pratique !

Proposition 4.9 (décomposition de l'EQM). *L'EQM d'un estimateur est égale à la somme de sa variance et du carré du biais :*

$$\forall \theta \in \Theta, \quad \text{EQM}(T) = \mathbf{E}_\theta(T - \theta)^2 = \text{Var}_\theta T + (\mathbf{E}_\theta T - \theta)^2. \quad (4.2)$$

En conséquence, parmi tous les estimateurs sans biais de θ , les plus précis au sens de l'EQM sont ceux de variance minimale.

Preuve. En utilisant la décomposition 4.1 de l'erreur d'estimation et en notant le biais $b(\theta)$, on obtient pour tout $\theta \in \Theta$:

$$\begin{aligned} \mathbf{E}_\theta(T - \theta)^2 &= \mathbf{E}_\theta((T - \mathbf{E}_\theta T) + b(\theta))^2 \\ &= \mathbf{E}_\theta((T - \mathbf{E}_\theta T)^2 + b(\theta)^2 + 2b(\theta)(T - \mathbf{E}_\theta T)) \\ &= \mathbf{E}_\theta((T - \mathbf{E}_\theta T)^2) + b(\theta)^2 + 2b(\theta)\mathbf{E}_\theta(T - \mathbf{E}_\theta T) \\ &= \text{Var}_\theta(T) + b(\theta)^2, \end{aligned}$$

en rappelant que $b(\theta)$ n'est pas aléatoire et en notant que $T - \mathbf{E}_\theta T$ a une \mathbf{E}_θ -espérance nulle par linéarité de \mathbf{E}_θ . \square

Puisque l'EQM est une mesure de la précision d'un estimateur, il est naturel de chercher à la minimiser. Nous allons donner sous certaines hypothèses une borne inférieure pour l'EQM, qui sera en quelque sorte le prix plancher à payer (en monnaie EQM) pour notre ignorance de la vraie valeur de θ . Nous allons voir que dans les bons cas cette borne inférieure est de la forme $c(\theta)n^{-1}$ pour des estimateurs sans biais. C'est l'inégalité de Cramér-Rao. Pour établir ce résultat, il nous faut introduire quelques notations et faire des hypothèses techniques. Nous allons d'abord travailler avec un vecteur aléatoire (X_1, \dots, X_n) défini sur un modèle statistique pour minorer la variance d'une v.a. $T = h(X_1, \dots, X_n)$, cf. lemme 4.11. Ensuite nous appliquerons ce résultat au cas où (X_1, \dots, X_n) est un échantillon, autrement dit les X_i sont P_θ -i.i.d. pour tout θ , afin d'en déduire l'inégalité de Cramér-Rao proprement dite. Les lecteurs pressés peuvent sauter directement à l'énoncé du théorème 4.12.

Nous supposons ici que $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$ est un modèle statistique ayant pour espace Θ de paramètres un ouvert de \mathbb{R} . Nous limiterons notre étude aux deux cas suivants.

Cas 1. Pour tout $\theta \in \Theta$, (X_1, \dots, X_n) a une loi discrète sous P_θ .

Cas 2. Pour tout $\theta \in \Theta$, (X_1, \dots, X_n) a une loi à densité³ p_θ sous P_θ .

Pour unifier les notations relatives à ces deux cas, on définit $f_n : \mathbb{R}^n \times \Theta \rightarrow \mathbb{R}_+$ par

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n, \forall \theta \in \Theta, \quad f_n(x, \theta) := \begin{cases} P_\theta(X_1 = x_1, \dots, X_n = x_n) & (\text{cas 1}) \\ p_\theta(x_1, \dots, x_n) & (\text{cas 2}). \end{cases}$$

Voici les 4 hypothèses dont nous aurons besoin.

3. Parmi toutes les densités possibles, qui diffèrent entre elles d'un ensemble de mesure nulle, on en choisit définitivement une, de préférence la plus lisse possible.

- (H1) L'ensemble $A_n := \{x \in \mathbb{R}^n; f_n(x, \theta) > 0\}$ ne dépend pas de θ .
 (H2) Pour tout $x \in A_n$, $f_n(x, \cdot)$ est dérivable par rapport à θ sur l'ouvert Θ .
 (H3) Définissons pour tout $\theta \in \Theta$, la fonction $u_\theta : \mathbb{R}^n \rightarrow \mathbb{R}$ et la v.a. U_θ par

$$u_\theta(x) := \begin{cases} \frac{\partial}{\partial \theta} \ln f_n(x, \theta) & \text{si } x \in A_n \\ 0 & \text{si } x \notin A_n \end{cases} \quad U_\theta := u_\theta(X_1, \dots, X_n).$$

On suppose alors que $0 < \mathbf{E}_\theta(U_\theta^2) < +\infty$ et que

$$\forall \theta \in \Theta, \quad \mathbf{E}_\theta(U_\theta) = 0. \quad (4.3)$$

- (H4) On suppose que $T = h(X_1, \dots, X_n)$ est une v.a. telle que $\mathbf{E}_\theta T^2 < +\infty$ pour tout $\theta \in \Theta$, que $g : \theta \mapsto g(\theta) := \mathbf{E}_\theta T$ est dérivable sur Θ et que

$$\forall \theta \in \Theta, \quad g'(\theta) = \mathbf{E}_\theta(TU_\theta). \quad (4.4)$$

Remarque 4.10. Les conditions (4.3) et (4.4) peuvent sembler assez exigeantes à première vue. En fait elles sont automatiquement vérifiées chaque fois que l'on a le droit d'intervertir dérivée $\partial/\partial\theta$ et série multiple (cas 1) ou intégrale sur \mathbb{R}^n (cas 2). Pour s'en convaincre, examinons le cas 2, laissant le 1 en exercice. Dans ce cas, (X_1, \dots, X_n) a pour densité p_θ nulle en dehors de A_n et l'espérance de $T = h(X_1, \dots, X_n)$ s'exprime par une intégrale multiple :

$$g(\theta) = \mathbf{E}_\theta T = \int_{\mathbb{R}^n} h(x) p_\theta(x) dx = \int_{\mathbb{R}^n} h(x) f_n(x, \theta) dx = \int_{A_n} h(x) f_n(x, \theta) dx.$$

En admettant que l'interversion entre l'intégrale et la dérivation par rapport à θ est légitime, on a donc :

$$\begin{aligned} g'(\theta) &= \frac{\partial}{\partial \theta} \int_{A_n} h(x) f_n(x, \theta) dx = \int_{A_n} h(x) \frac{\partial}{\partial \theta} f_n(x, \theta) dx \\ &= \int_{A_n} h(x) \frac{\frac{\partial}{\partial \theta} f_n(x, \theta)}{f_n(x, \theta)} f_n(x, \theta) dx \\ &= \int_{A_n} h(x) \frac{\partial}{\partial \theta} (\ln f_n(x, \theta)) f_n(x, \theta) dx \\ &= \int_{A_n} h(x) u_\theta(x) f_n(x, \theta) dx = \mathbf{E}_\theta(TU_\theta), \end{aligned}$$

ce qui nous donne (4.4). Dans le cas particulier où h est la fonction constante 1 et donc T est la v.a. constante 1, on a $g(\theta) = \mathbf{E}_\theta(1) = 1$, d'où $g'(\theta) = 0$ et on obtient (4.3).

Lemme 4.11. *Sous les hypothèses (H1), (H2), (H3) et (H4), on a l'inégalité*

$$\text{Var}_\theta(T) \geq \frac{g'(\theta)^2}{I(\theta)}, \quad \text{où } I(\theta) := \mathbf{E}_\theta(U_\theta^2). \quad (4.5)$$

Preuve. Puisque T et U_θ sont de carré intégrable, leur covariance existe et peut s'écrire par la formule de Koenig

$$\text{Cov}_\theta(T, U_\theta) = \mathbf{E}_\theta(TU_\theta) - \mathbf{E}_\theta(T)\mathbf{E}_\theta(U_\theta) = \mathbf{E}_\theta(TU_\theta), \quad (4.6)$$

à cause de (4.3). En utilisant successivement (4.4), (4.6), l'inégalité de Cauchy-Schwarz pour les covariances et (4.3), on obtient

$$g'(\theta)^2 = (\text{Cov}_\theta(T, U_\theta))^2 \leq \text{Var}_\theta(T) \text{Var}_\theta(U_\theta) = \text{Var}_\theta(T) \mathbf{E}_\theta(U_\theta^2),$$

ce qui nous donne la minoration (4.5) après division par $\mathbf{E}_\theta(U_\theta^2) > 0$. \square

Voyons maintenant ce que donne le lemme 4.11 dans le cas où le vecteur aléatoire (X_1, \dots, X_n) est un échantillon associé au modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$.

Dans le cas 2, notons provisoirement $p_{\theta,1}$ la densité de X_1 sous P_θ . Pour donner une expression de $f_n(x, \theta)$ unifiant les cas 1 et 2, notons f la fonction

$$f : \mathbb{R} \times \Theta, \quad (t, \theta) \longmapsto f(t, \theta) := \begin{cases} P_\theta(X_1 = t) & (\text{cas 1}) \\ p_{\theta,1}(t) & (\text{cas 2}). \end{cases} \quad (4.7)$$

Alors par indépendance et équidistribution (sous chaque P_θ) des X_i , on a

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad \forall \theta \in \Theta, \quad f_n(x, \theta) = f(x_1, \theta) \cdots f(x_n, \theta). \quad (4.8)$$

Cette factorisation de f_n va nous permettre de simplifier les hypothèses (H1)–(H4). Posons

$$A := \{t \in \mathbb{R}; f(t, \theta) > 0\}.$$

Il résulte alors immédiatement de la factorisation de $f_n(x, \theta)$ que

$$A_n = A \times \cdots \times A = A^n.$$

Ainsi l'hypothèse (H1) se réduit à « A ne dépend pas de θ ».

En raison de la règle de dérivation d'un produit de fonctions, l'hypothèse (H2) se réduit à la dérivabilité par rapport à θ de $f(t, \theta)$.

Pour simplifier (H3), commençons par noter que par (4.8),

$$\forall x \in A^n, \quad u_\theta(x) = \sum_{i=1}^n \frac{\partial}{\partial \theta} (\ln f(x_i, \theta)),$$

d'où $U_\theta = \sum_{i=1}^n Y_{\theta,i}$ avec $Y_{\theta,i} := \frac{\partial}{\partial \theta} (\ln f(X_i, \theta)) \mathbf{1}_A(X_i)$. Par équidistribution des $Y_{\theta,i}$, pour que (H3) soit vérifiée, il suffit que $0 < \mathbf{E}_\theta(Y_{\theta,1}^2) < +\infty$ et que $\mathbf{E}_\theta Y_{\theta,1} = 0$. Pour cette dernière condition, il suffit de pouvoir intervertir dérivation par rapport à θ et

sommation sur A (cas 1) ou intégration sur A (cas 2). Vérifions le dans le cas 2 :

$$\begin{aligned}
 \mathbf{E}_\theta(Y_{\theta,1}) &= \int_A \frac{\partial}{\partial \theta} (\ln f(t, \theta)) f(t, \theta) dt = \int_A \left(\frac{\partial f(t, \theta)}{\partial \theta} \right) \frac{1}{f(t, \theta)} f(t, \theta) dt \\
 &= \int_A \left(\frac{\partial f(t, \theta)}{\partial \theta} \right) dt \\
 &= \frac{\partial}{\partial \theta} \int_A f(t, \theta) dt \\
 &= \frac{\partial}{\partial \theta} \int_{\mathbb{R}} f(t, \theta) dt \\
 &= \frac{\partial}{\partial \theta} (1) = 0.
 \end{aligned}$$

Notons au passage que si $\mathbf{E}_\theta(Y_{\theta,1}^2) < +\infty$ et si l'interversion ci-dessus est légitime, alors $\mathbf{E}_\theta(U_\theta^2) = \text{Var}_\theta U_\theta$, d'où par indépendance et équidistribution des $Y_{\theta,i}$,

$$\mathbf{E}_\theta(U_\theta^2) = n \text{Var}_\theta(Y_{\theta,1}) = n \mathbf{E}_\theta \left| \frac{\partial}{\partial \theta} (\ln f(X_1, \theta)) \right|^2. \quad (4.9)$$

Dans cette formule, nous avons délibérément omis de recopier $\mathbf{1}_A(X_1)$ en explicitant $Y_{\theta,1}$. Cet abus n'est pas bien grave car $P_\theta(X_1 \in A) = 1$ pour tout θ . Nous ferons systématiquement cet abus d'écriture dans la suite pour toutes les v.a. construites à l'aide d'une dérivée logarithmique. Cela revient à considérer que ces v.a. sont définies sur Ω privé d'un événement de probabilité nulle pour P_θ , cet événement ne dépendant pas de θ . On peut d'ailleurs convenir que l'on complète leur définition en les prenant égales à zéro sur cet événement de probabilité nulle.

Enfin, concernant (H4), la factorisation (4.8) n'apporte pas de simplification spectaculaire et nous nous contentons de reprendre la remarque 4.10.

Tout ceci aboutit à l'inégalité de Cramér-Rao suivante. Nous reprenons dans l'énoncé toutes les notations et hypothèses développées ci-dessus à l'intention des lecteurs ayant sauté la marche d'approche.

Théorème 4.12 (inégalité de Cramér-Rao). *Soit $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique ayant pour espace Θ de paramètres un ouvert de \mathbb{R} . Soit (X_1, \dots, X_n) un échantillon associé à ce modèle et relevant de l'un des deux cas suivants.*

Cas 1. *Pour tout $\theta \in \Theta$, X_1 a une loi discrète sous P_θ .*

Cas 2. *Pour tout $\theta \in \Theta$, X_1 a une loi à densité $p_{\theta,1}$ sous P_θ .*

On note f la fonction

$$f : \mathbb{R} \times \Theta, \quad (t, \theta) \mapsto f(t, \theta) := \begin{cases} P_\theta(X_1 = t) & (\text{cas 1}) \\ p_{\theta,1}(t) & (\text{cas 2}). \end{cases}$$

On suppose vérifiées les hypothèses suivantes.

(H1) *L'ensemble $A := \{t \in \mathbb{R}; f(t, \theta) > 0\}$ ne dépend pas de θ .*

(H2) Pour tout $t \in A$, la fonction $f(t, \cdot)$ est dérivable sur Θ .

(H3) Pour tout $\theta \in \Theta$, $0 < \mathbf{E}_\theta \left| \frac{\partial}{\partial \theta} \ln f(X_1, \theta) \right|^2 < +\infty$ et

$$\mathbf{E}_\theta \left(\frac{\partial}{\partial \theta} \ln f(X_1, \theta) \right) = 0. \quad (4.10)$$

Pour que (4.10) soit vérifiée, il suffit que l'on puisse intervertir

- dans le cas 1 : $\sum_{t \in A}$ et $\frac{\partial}{\partial \theta}$ dans $\sum_{t \in A} \frac{\partial f(t, \theta)}{\partial \theta}$;
- dans le cas 2 : \int_A et $\frac{\partial}{\partial \theta}$ dans $\int_A \frac{\partial f(t, \theta)}{\partial \theta} dt$.

(H4) $T = h(X_1, \dots, X_n)$ est un estimateur sans biais de $g(\theta)$ tel que $\mathbf{E}_\theta T^2 < +\infty$ pour tout $\theta \in \Theta$, avec g dérivable sur Θ . De plus,

$$g'(\theta) = \sum_{i=1}^n \mathbf{E}_\theta \left(T \frac{\partial}{\partial \theta} \ln f(X_i, \theta) \right). \quad (4.11)$$

Pour que (4.11) soit vérifiée, il suffit que l'on puisse intervertir :

- dans le cas 1 : $\sum_{x \in A^n}$ et $\frac{\partial}{\partial \theta}$ dans $\sum_{x \in A^n} h(x_1, \dots, x_n) \frac{\partial}{\partial \theta} (f(x_1, \theta) \dots f(x_n, \theta))$;
- dans le cas 2 : \int_A et $\frac{\partial}{\partial \theta}$ dans $\int_{A^n} h(x_1, \dots, x_n) \frac{\partial}{\partial \theta} (f(x_1, \theta) \dots f(x_n, \theta)) dx_1 \dots dx_n$.

Sous ces conditions, on a

$$\text{Var}_\theta(T) \geq \frac{g'(\theta)^2}{n \mathbf{E}_\theta \left| \frac{\partial}{\partial \theta} (\ln f(X_1, \theta)) \right|^2}.$$

En particulier, si T est un estimateur sans biais de θ , on a

$$\text{EQM}(T) \geq \frac{c(\theta)}{n}, \quad \text{avec} \quad \frac{1}{c(\theta)} := \mathbf{E}_\theta \left| \frac{\partial}{\partial \theta} (\ln f(X_1, \theta)) \right|^2.$$

La quantité $c(\theta)/n$ est appelée borne de Cramér-Rao associée au modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$ et à l'échantillon (X_1, \dots, X_n) .

Définition 4.13 (estimateur efficace). Un estimateur sans biais est dit efficace si son EQM atteint la borne de Cramér-Rao pour tout $\theta \in \Theta$.

Les hypothèses (H1)–(H4) sont appelées *conditions de régularité* pour l'inégalité de Cramér-Rao. En pratique pour vérifier (H3) et (H4), on a parfois besoin d'un théorème d'interversion entre dérivation et sommation⁴. Nous en proposons un ci-dessous, issu du théorème de convergence dominée de Lebesgue. Sa démonstration sortirait du cadre de ce cours⁵.

Théorème 4.14 (dérivation sous le signe somme). Soient $D \subset \mathbb{R}^n$, $n \geq 1$, Θ un ouvert de \mathbb{R} , et $F : D \times \Theta \rightarrow \mathbb{R}$, $(t, \theta) \mapsto F(t, \theta)$. On suppose que pour tout $t \in D$, $F(t, \cdot)$ est dérivable sur Θ .

4. Essayez d'abord une vérification directe de (4.10) et (4.11), comme dans l'exemple 4.15.

5. Voir par exemple le cours d'I.F.P. 2004, chapitre 4

<http://math.univ-lille1.fr/~suquet/ens/IFP/Cours/cours04/Chap4ifp04.pdf>

Cas 1. D est dénombrable et pour tout $\theta \in \Theta$, la famille $\{F(t, \theta); t \in D\}$ est sommable. On suppose de plus que pour tout $\theta_0 \in \Theta$, il existe un voisinage V_0 de θ_0 dans Θ et une fonction $M_0 : D \rightarrow \mathbb{R}_+$ telle que

$$\forall \theta \in V_0, \forall t \in D, \quad \left| \frac{\partial F(t, \theta)}{\partial \theta} \right| \leq M_0(t) \quad (4.12)$$

et

$$\sum_{t \in D} M_0(t) < +\infty. \quad (4.13)$$

Alors $\left\{ \frac{\partial F(t, \theta)}{\partial \theta}; t \in D \right\}$ est sommable pour tout $\theta \in \Theta$, sa somme est dérivable comme fonction de θ et on a l'interversion

$$\forall \theta \in \Theta, \quad \sum_{t \in D} \frac{\partial F(t, \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \sum_{t \in D} F(t, \theta). \quad (4.14)$$

Cas 2. D est un borélien⁶ de \mathbb{R}^n et pour tout $\theta \in \Theta$, l'intégrale $\int_D F(t, \theta) dt$ est absolument convergente. On suppose de plus que pour tout $\theta_0 \in \Theta$, il existe un voisinage V_0 de θ_0 dans Θ , un borélien $D_0 \subset D$ tel que $\lambda_n(D \setminus D_0) = 0$ et une fonction $M_0 : D_0 \rightarrow \mathbb{R}_+$ telle que

$$\forall \theta \in V_0, \forall t \in D_0, \quad \left| \frac{\partial F(t, \theta)}{\partial \theta} \right| \leq M_0(t) \quad (4.15)$$

et

$$\int_{D_0} M_0(t) dt < +\infty. \quad (4.16)$$

Alors l'intégrale $\int_D \frac{\partial F(t, \theta)}{\partial \theta} dt$ est absolument convergente, est dérivable comme fonction de θ et on a l'interversion

$$\forall \theta \in \Theta, \quad \int_D \frac{\partial F(t, \theta)}{\partial \theta} dt = \frac{\partial}{\partial \theta} \int_D F(t, \theta) dt. \quad (4.17)$$

Exemple 4.15 (estimation du paramètre d'une loi de Poisson). Dans cet exemple, $\Theta =]0, +\infty[$ et chaque X_i suit, sous P_θ , la loi de Poisson de paramètre θ . On se propose d'estimer θ . Comme $\mathbf{E}_\theta X_i = \theta$, la moyenne empirique \bar{X} est un estimateur sans biais de θ . Nous allons vérifier les conditions de régularité pour l'inégalité de Cramér-Rao avec ce modèle et $T = \bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. D'abord il est clair que l'on est dans le cas 1 et que

$$\forall \theta \in]0, +\infty[, \quad f(t, \theta) = \begin{cases} \frac{\theta^t e^{-\theta}}{t!} & \text{si } t \in \mathbb{N}, \\ 0 & \text{si } t \in (\mathbb{R} \setminus \mathbb{N}). \end{cases}$$

6. En pratique, on utilisera ceci avec D produit cartésien d'intervalles de \mathbb{R} ou union finie d'ensembles de ce type.

On voit immédiatement que $A = \{t \in \mathbb{R}; f(t, \theta) > 0\} = \mathbb{N}$ ne dépend pas de θ , ce qui satisfait (H1). Clairement $f(t, \cdot)$ est dérivable sur $]0, +\infty[$ pour tout $t \in \mathbb{N}$, ce qui satisfait (H2). La dérivée par rapport à θ s'écrit :

$$\forall t \in \mathbb{N}, \forall \theta \in]0, +\infty[, \quad \frac{\partial f(t, \theta)}{\partial \theta} = \frac{e^{-\theta}}{t!} (t\theta^{t-1} - \theta^t) = \frac{\theta^{t-1}e^{-\theta}}{t!} (t - \theta).$$

La dérivée logarithmique a une expression particulièrement simple :

$$\forall t \in \mathbb{N}, \forall \theta \in]0, +\infty[, \quad \frac{\partial \ln f(t, \theta)}{\partial \theta} = \frac{t - \theta}{\theta}. \quad (4.18)$$

On en déduit que pour tout $\theta \in]0, +\infty[$,

$$\mathbf{E}_\theta \left| \frac{\partial \ln f(X_1, \theta)}{\partial \theta} \right|^2 = \mathbf{E}_\theta \frac{(X_1 - \theta)^2}{\theta^2} = \theta^{-2} \text{Var}_\theta(X_1) = \frac{1}{\theta} < +\infty,$$

en utilisant le fait que l'espérance et la variance d'une loi $\text{Pois}(\theta)$ valent θ . Ainsi la condition d'intégrabilité dans (H3) est vérifiée. La vérification de (4.10) est immédiate grâce à (4.18) :

$$\mathbf{E}_\theta \left(\frac{\partial}{\partial \theta} \ln f(X_1, \theta) \right) = \mathbf{E}_\theta \left(\frac{X_1 - \theta}{\theta} \right) = \frac{1}{\theta} (\mathbf{E}_\theta X_1 - \theta) = 0.$$

Pour (H4), on commence par remarquer que $\mathbf{E}_\theta(T^2) = \mathbf{E}_\theta(\bar{X}^2) < +\infty$, parce que la v.a. X_1 de loi $\text{Pois}(\theta)$ est de carré intégrable quel que soit $\theta > 0$. Vérifions (4.11). Ici $g(\theta) = \mathbf{E}_\theta \bar{X} = \theta$, donc $g'(\theta) = 1$. En utilisant à nouveau (4.18), on obtient

$$\begin{aligned} \sum_{i=1}^n \mathbf{E}_\theta \left(T \frac{\partial}{\partial \theta} \ln f(X_i, \theta) \right) &= \sum_{i=1}^n \mathbf{E}_\theta (\theta^{-1} (X_i - \theta) \bar{X}) \\ &= \frac{1}{n\theta} \sum_{i=1}^n \mathbf{E}_\theta \left((X_i - \theta) \sum_{j=1}^n X_j \right) \\ &= \frac{1}{n\theta} \sum_{i,j=1}^n \mathbf{E}_\theta ((X_i - \theta) X_j). \end{aligned} \quad (4.19)$$

Si $i \neq j$, par P_θ -indépendance de $X_i - \theta$ et X_j , on a $\mathbf{E}_\theta((X_i - \theta) X_j) = \mathbf{E}_\theta(X_i - \theta) \mathbf{E}_\theta X_j = 0$, car $\mathbf{E}_\theta X_i = \theta$. Si $j = i$, $\mathbf{E}_\theta((X_i - \theta) X_i) = \mathbf{E}_\theta(X_i^2) - \theta \mathbf{E}_\theta X_i = \mathbf{E}_\theta(X_i^2) - (\mathbf{E}_\theta X_i)^2 = \text{Var}_\theta X_i = \theta$. En reportant ces résultats dans (4.19), il vient

$$\sum_{i=1}^n \mathbf{E}_\theta \left(T \frac{\partial}{\partial \theta} \ln f(X_i, \theta) \right) = \frac{1}{n\theta} \sum_{i=1}^n \theta = \frac{1}{n\theta} n\theta = 1,$$

ce qui établit (4.11). Ainsi toutes les conditions de régularité pour l'inégalité de Cramér-Rao sont vérifiées.

Puisque $\mathbf{E}_\theta \left| \frac{\partial \ln f(X_1, \theta)}{\partial \theta} \right|^2 = \theta^{-1}$, la constante $c(\theta)$ est simplement θ et la borne de Cramér-Rao vaut ici θn^{-1} . D'autre part, puisque \bar{X} est un estimateur sans biais de θ , son EQM est égale à sa variance. Or nous savons que $\text{Var}_\theta(\bar{X}) = n^{-1} \text{Var}_\theta(X_1) = \theta n^{-1}$. Ainsi l'EQM de \bar{X} est égale à la borne de Cramér-Rao, pour tout $\theta \in]0, +\infty[$. L'estimateur $T = \bar{X}$ est donc un estimateur *efficace* du paramètre θ .

4.2 Maximum de vraisemblance

4.2.1 Exercice introductif

La méthode d'estimation par maximum de vraisemblance repose sur une idée assez naturelle. L'exercice suivant devrait vous en convaincre. Il est librement adapté d'un sujet de baccalauréat, donné dans les années 70 (académie de Rennes). Essayez d'en rédiger la solution de manière détaillée, avant de regarder les réponses succinctes ci-dessous.

Un étang contient un nombre N , inconnu mais défini de poissons. L'objet du problème est de proposer une évaluation de N basée sur des hypothèses bien définies.

1) On pêche dans différents endroits de l'étang; on en sort 20 poissons que l'on marque et que l'on remet vivants dans l'étang après avoir été marqués.

Quelques jours plus tard, on effectue une nouvelle pêche dans des endroits variés du même étang, en pêchant les poissons un par un, en notant pour chaque poisson pris s'il est marqué ou non et en le rejetant à l'eau avant de pêcher le suivant. On prend ainsi 50 poissons dont 4 sont marqués.

On suppose qu'entre les deux pêches, la population de l'étang n'a pas varié et que lors de la seconde pêche à chaque coup, il y a équiprobabilité de sortie pour chacun des N poissons de l'étang.

Avant la seconde pêche, on pouvait se poser le problème : quelle est la probabilité de sortir k poissons marqués sur les 50 pêchés ? Cette probabilité dépendant de N sera notée $P_N(S_{50} = k)$ où S_{50} désigne la variable aléatoire « nombre de poissons marqués parmi les 50 poissons pêchés ». Donner l'expression générale de $P_N(S_{50} = k)$.

2) Soit f , la fonction qui à x réel supérieur à 20 associe le nombre réel

$$\left(\frac{20}{x}\right)^4 \left(1 - \frac{20}{x}\right)^{46}.$$

Montrer que f a un maximum. Pour quelle valeur a de x ce maximum est-il atteint ?

3) Quelle valeur doit on attribuer à N pour maximiser la probabilité $P_N(S_{50} = 4)$? Cette façon d'estimer le nombre inconnu N conduit à la même valeur qu'une autre méthode très simple que l'on aurait pu utiliser pour évaluer N . Laquelle ?

Voici les réponses. La proportion inconnue de poissons marqués dans le lac est $p_N = 20/N$ et S_{50} suit, sous P_N , la loi binomiale de paramètres 50 et p_N , d'où

$$P_N(S_{50} = k) = C_{50}^k \left(\frac{20}{N}\right)^k \left(1 - \frac{20}{N}\right)^{50-k}, \quad k = 0, 1, \dots, 50.$$

Pour voir que f a un maximum, il y a au moins 3 méthodes : dériver brutalement f et étudier le signe de f' , poser $t = 20/x$ et étudier les variations de la fonction $g(t)$ ainsi obtenue, calculer la dérivée logarithmique de f qui a même signe que f' . On vérifie ainsi que f a un unique maximum, atteint en $a = 250$. Comme $P_N(S_{50} = 4) = C_{50}^4 f(N)$, la valeur de N qui maximise $P_N(S_{50} = 4)$ est la même que celle qui maximise $f(N)$, soit

$N = 250$. On aurait pu arriver à la même valeur estimée pour N en considérant que la proportion de poissons marqués parmi les 50 pêchés est la même que dans tout l'étang, $\frac{4}{50} = \frac{20}{N}$, soit $N = \frac{20 \times 50}{4} = 250$. Cela revient d'ailleurs à estimer $p_N = 20/N = \mathbf{E}_N X_i$ par la moyenne empirique $\bar{X} = S_{50}/50$, où l'on a noté X_i la v.a. de Bernoulli valant 1 si le i^e poisson pêché est marqué et 0 sinon.

4.2.2 Cas discret

Soit $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique avec $\Theta \subset \mathbb{R}^d$ et (X_1, \dots, X_n) un échantillon associé à ce modèle. On suppose que la loi de X_i sous chaque P_θ est discrète et on pose

$$\forall t \in \mathbb{R}, \quad f(t, \theta) = P_\theta(X_1 = t).$$

Par indépendance et équidistribution des X_i , on a

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n, \quad L(x, \theta) := P_\theta(X_1 = x_1, \dots, X_n = x_n) = f(x_1, \theta) \dots f(x_n, \theta).$$

La fonction L est appelée *fonction de vraisemblance* (L à cause de « likelihood », vraisemblance en anglais). Lorsque la fonction $\theta \mapsto L(x, \theta)$ a un unique maximum atteint en $\hat{\theta} = g_n(x_1, \dots, x_n)$, on peut utiliser cette valeur $\hat{\theta}$ pour estimer θ . On dit alors que

$$T := g_n(X_1, \dots, X_n)$$

est l'estimateur par maximum de vraisemblance de θ . En pratique, la recherche de ce maximum se fait par dérivation de L relativement à θ et en raison de la factorisation ci-dessus, il est souvent plus commode d'utiliser au lieu de L la fonction $\ln L$, appelée log-vraisemblance, qui est maximale pour le même $\hat{\theta}$ que L .

Exemple 4.16 (estimation du paramètre d'une loi de Bernoulli). Ici $\Theta =]0, 1[$ et les X_i suivant sous P_θ la loi $\text{Bern}(\theta)$. On a

$$f(t, \theta) = P_\theta(X_1 = t) = \begin{cases} \theta & \text{si } t = 1, \\ 1 - \theta & \text{si } t = 0, \\ 0 & \text{si } t \neq 0 \text{ et } t \neq 1. \end{cases}$$

Posons pour $x = (x_1, \dots, x_n) \in \{0, 1\}^n$, $s_n := x_1 + \dots + x_n$. Ainsi s_n est le nombre de 1 dans l'échantillon observé (x_1, \dots, x_n) et le nombre de zéros dans cet échantillon est $n - s_n$. La vraisemblance et la log-vraisemblance s'écrivent alors pour tout $x \in \{0, 1\}^n$ sous la forme :

$$L(x, \theta) = \theta^{s_n} (1 - \theta)^{n-s_n}, \quad \ln L(x, \theta) = s_n \ln \theta + (n - s_n) \ln(1 - \theta).$$

Pour tout $\theta \in]0, 1[$,

$$\frac{\partial \ln L(x, \theta)}{\partial \theta} = \frac{s_n}{\theta} - \frac{n - s_n}{1 - \theta} = \frac{s_n(1 - \theta) - \theta(n - s_n)}{\theta(1 - \theta)} = \frac{s_n - n\theta}{\theta(1 - \theta)},$$

donc $\frac{\partial \ln L(x, \theta)}{\partial \theta}$ a même signe que $s_n - n\theta$. La log vraisemblance et donc aussi la vraisemblance ont ainsi un unique maximum atteint en $\theta = \hat{\theta} = n^{-1}s_n = n^{-1}(x_1 + \dots + x_n)$. En revenant aux v.a. X_i , on en déduit que l'estimateur de θ par maximum de vraisemblance est

$$T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

On retrouve là un estimateur bien connu du paramètre de la loi Bern(θ).

4.2.3 Cas à densité

Soit $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in \Theta})$ un modèle statistique avec $\Theta \subset \mathbb{R}^d$ et (X_1, \dots, X_n) un échantillon associé à ce modèle. On suppose que la loi de X_i sous chaque P_θ admet une densité continue $f(\cdot, \theta)$. Si on veut définir comme précédemment la vraisemblance comme la quantité $P_\theta(X_1 = x_1, \dots, X_n = x_n)$, cela n'a plus grand sens car cette probabilité est toujours nulle. Voici une heuristique pour légitimer la définition alternative ci-dessous. Au lieu de $P_\theta(X_1 = x_1, \dots, X_n = x_n)$, on cherche à maximiser

$$P_\theta(D_\varepsilon) := P_\theta(X_1 \in [x_1 - \varepsilon, x_1 + \varepsilon], \dots, X_n \in [x_n - \varepsilon, x_n + \varepsilon])$$

pour ε « petit ». Cette probabilité s'écrit

$$P_\theta(D_\varepsilon) = \int_{x_1 - \varepsilon}^{x_1 + \varepsilon} \dots \int_{x_n - \varepsilon}^{x_n + \varepsilon} f(t_1, \theta) \dots f(t_n, \theta) dt_1 \dots dt_n.$$

En raison de la continuité de $f(\cdot, \theta)$, on en déduit :

$$P_\theta(D_\varepsilon) \sim (2\varepsilon)^n f(x_1, \theta) \dots f(x_n, \theta) \quad (\varepsilon \rightarrow 0).$$

Donc pour ε « petit », maximiser $P_\theta(D_\varepsilon)$ revient⁷ à maximiser $f(x_1, \theta) \dots f(x_n, \theta)$. Ceci nous conduit à prendre pour fonction vraisemblance

$$L(x, \theta) := f(x_1, \theta) \dots f(x_n, \theta).$$

Nous utiliserons cette définition de la vraisemblance même quand la densité $f(\cdot, \theta)$ présente un nombre fini de points de discontinuité (ce qui est le cas par exemple de la densité d'une loi uniforme sur un segment ou d'une loi exponentielle). Une fois définie la vraisemblance, la méthode d'estimation par maximum de vraisemblance fonctionne exactement comme dans le cas discret. Nous l'illustrons ci-dessous par des exemples d'estimation des paramètres d'une gaussienne.

Exemple 4.17 (loi gaussienne d'espérance inconnue et d'écart-type connu). Ici la loi de X_i sous P_θ est $\mathfrak{N}(\theta, \sigma)$, avec σ connu et θ inconnu, $\theta \in \Theta = \mathbb{R}$. La densité de X_i sous P_θ s'écrit

$$f(t, \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t - \theta)^2}{2\sigma^2}\right),$$

7. Ceci ne prétend pas être une justification rigoureuse, c'est juste une heuristique.

d'où

$$L(x_1, \dots, x_n, \theta) = (2\pi)^{-n/2} \sigma^{-n} \exp \left(\frac{-1}{2\sigma^2} \sum_{i=1}^n (x_i - \theta)^2 \right).$$

Il est clair que maximiser cette quantité équivaut à minimiser $\sum_{i=1}^n (x_i - \theta)^2$. En considérant les x_i comme des points de la droite réelle, il est bien connu que le point réalisant le minimum de la somme des carrés des distances aux x_i est l'isobarycentre des x_i , soit $n^{-1}(x_1 + \dots + x_n)$. Pour ceux qui ne sont pas convaincus par cet argument géométrique, il suffit de calculer la dérivée :

$$\frac{\partial}{\partial \theta} \sum_{i=1}^n (x_i - \theta)^2 = -2 \sum_{i=1}^n (x_i - \theta) = 2n\theta - 2 \sum_{i=1}^n x_i,$$

pour arriver à la même conclusion. L'estimateur de θ par maximum de vraisemblance est donc ici :

$$T = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}.$$

Exemple 4.18 (loi gaussienne d'espérance connue et d'écart-type inconnu). Ici la loi de X_i sous P_θ est $\mathfrak{N}(m, \theta)$, avec m connu et θ inconnu, $\theta \in \Theta =]0, +\infty[$. La densité de X_i sous P_θ s'écrit

$$f(t, \theta) = \frac{1}{\theta \sqrt{2\pi}} \exp \left(\frac{-(t - m)^2}{2\theta^2} \right),$$

d'où

$$L(x_1, \dots, x_n, \theta) = (2\pi)^{-n/2} \theta^{-n} \exp \left(\frac{-1}{2\theta^2} \sum_{i=1}^n (x_i - m)^2 \right).$$

Ici il est plus commode d'utiliser la log-vraisemblance et de poser pour alléger les écritures $v_n^2 := n^{-1} \sum_{i=1}^n (x_i - m)^2$, $v_n \geq 0$, en notant que v_n ne dépend pas de θ :

$$\ln L(x_1, \dots, x_n, \theta) = \frac{-n}{2} \ln(2\pi) - n \ln \theta - \frac{nv_n^2}{2\theta^2}.$$

Dérivons par rapport à θ :

$$\frac{\partial}{\partial \theta} \ln L(x_1, \dots, x_n, \theta) = -\frac{n}{\theta} - \frac{nv_n^2}{2} \frac{(-2)}{\theta^3} = \frac{n}{\theta} \left(\frac{v_n^2}{\theta^2} - 1 \right).$$

Ainsi la dérivée logarithmique est positive pour $\theta < v_n$ et négative pour $\theta > v_n$. La fonction $\theta \mapsto L(x, \theta)$ a donc un unique maximum atteint en $\hat{\theta} = v_n$. L'estimateur par maximum de vraisemblance de l'écart-type inconnu est donc ici

$$T = \left(\frac{1}{n} \sum_{i=1}^n (X_i - m)^2 \right)^{1/2}.$$

Annexe A

Tables statistiques

A.1 Loi normale standard

La table de la page 124 donne les valeurs de la fonction de répartition Φ de la loi normale $\mathfrak{N}(0, 1)$. Rappelons que

$$\forall x \in \mathbb{R}, \quad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

Il n'y a pas d'expression analytique pour les primitives de la densité gaussienne et on ne sait calculer cette intégrale que numériquement. Pour en déduire les valeurs de la f.d.r d'une v.a. Y de loi gaussienne quelconque $\mathfrak{N}(m, \sigma)$, on utilise le fait que $(Y - m)/\sigma$ suit la loi $\mathfrak{N}(0, 1)$, d'où

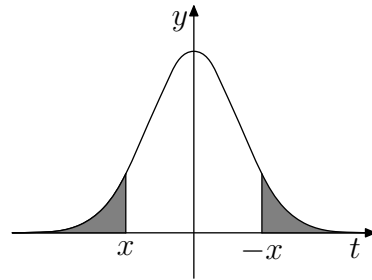
$$P(Y \leq y) = P(\sigma^{-1}(Y - m) \leq \sigma^{-1}(y - m)) = \Phi(\sigma^{-1}(y - m)).$$

La table donne les valeurs de $\Phi(x)$ pour x positif. Lorsque x est négatif, on utilise la relation

$$\Phi(x) = 1 - \Phi(-x)$$

qui résulte de la parité de la densité gaussienne $\mathfrak{N}(0, 1)$.

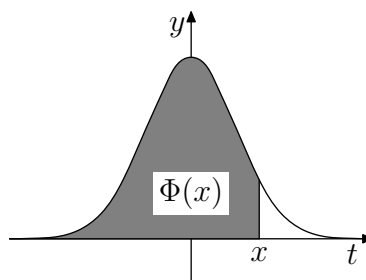
Exemple : pour $x = -1,8$, on trouve :
 $\Phi(x) = 1 - 0,9641 = 0,0359$.



Pour les « très grandes valeurs de x », (i.e. $|x| \geq 4$), on dispose du résultat suivant qui donne une évaluation de la « queue » de la loi normale.

Pour tout $x > 0$, on a l'encadrement :

$$\left(\frac{1}{x} - \frac{1}{x^3}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) < 1 - \Phi(x) < \frac{1}{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

Table des valeurs de Φ , f.d.r. de la loi normale standard $\mathfrak{N}(0, 1)$ 

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6627	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7122	0.7156	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7356	0.7389	0.7421	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8414	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8622
1.1	0.8643	0.8665	0.8687	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9083	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9193	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9895	0.9898	0.9901	0.9903	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9926	0.9928	0.9930	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9944	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9958	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Table pour les grandes valeurs de x

x	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.8	4.0	4.5
$\Phi(x)$	0.99865	0.99904	0.99931	0.99952	0.99966	0.99976	0.999841	0.999928	0.999968	0.999997

A.2 Lois du khi2

La table de la page 126 donne pour X de loi $\chi^2(d)$ et certaines valeurs courantes de la probabilité q , la valeur du réel x_q tel que

$$P(X > x_q) = q.$$

C'est donc une table de *quantiles* de la loi du χ^2 . Pour $d = 1$, les trois premières valeurs données par la table sont inférieures à 10^{-3} . Des valeurs plus précises sont $x_{0,999} = 16.10^{-7}$, $x_{0,995} = 393.10^{-7}$, $x_{0,99} = 1571.10^{-7}$.

À titre d'exemple, voici comment utiliser cette table pour déterminer deux réels α et β tels que $P(\alpha < X \leq \beta) = 0,95$, pour X suivant la loi $\chi^2(7)$. On cherche α tel que $P(X \leq \alpha) = (1 - 0,95)/2 = 0,025$ et β tel que $P(X > \beta) = (1 - 0,95)/2 = 0,025$. On trouve donc que $\beta = x_{0,025}$ et il suffit de lire sa valeur à l'intersection de la ligne 7 et de la colonne 0,025, d'où $\beta = 16,013$. Pour α , on remarque que $P(X \leq \alpha) = 1 - P(X > \alpha)$, d'où $P(X > \alpha) = 1 - 0,025 = 0,975$. Donc $\alpha = x_{0,975}$ que l'on trouve dans la table à l'intersection de la ligne 7 et de la colonne 0,975, soit $\alpha = 1,690$.

Table des valeurs du réel x_q tel que $P(X > x_q) = q$, pour X de loi $\chi^2(d)$

$d \backslash q$	0,999	0,995	0,990	0,975	0,95	0,90	0,50	0,10	0,05	0,025	0,01	0,005	0,001
1	0,000	0,000	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879	10,828
2	0,002	0,010	0,020	0,051	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597	13,816
3	0,024	0,072	0,115	0,216	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838	16,266
4	0,091	0,207	0,297	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860	18,467
5	0,210	0,412	0,554	0,831	1,145	1,610	4,351	9,236	11,070	12,833	15,086	16,750	20,515
6	0,381	0,676	0,872	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548	22,458
7	0,598	0,989	1,239	1,690	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278	24,322
8	0,857	1,344	1,646	2,180	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955	26,124
9	1,152	1,735	2,088	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589	27,877
10	1,479	2,156	2,558	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188	29,588
11	1,834	2,603	3,053	3,816	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757	31,264
12	2,214	3,074	3,571	4,404	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300	32,909
13	2,617	3,565	4,107	5,009	5,892	7,042	12,340	19,812	22,362	24,736	27,688	29,819	34,528
14	3,041	4,075	4,660	5,629	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319	36,123
15	3,483	4,601	5,229	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801	37,697
16	3,942	5,142	5,812	6,908	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267	39,252
17	4,416	5,697	6,408	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718	40,790
18	4,905	6,265	7,015	8,231	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156	42,312
19	5,407	6,844	7,633	8,907	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582	43,820
20	5,921	7,434	8,260	9,591	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997	45,315
21	6,447	8,034	8,897	10,283	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401	46,797
22	6,983	8,643	9,542	10,982	12,338	14,041	21,337	30,813	33,924	36,781	40,289	42,796	48,268
23	7,529	9,260	10,196	11,689	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181	49,728
24	8,085	9,886	10,856	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,559	51,179
25	8,649	10,520	11,524	13,120	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928	52,620
26	9,222	11,160	12,198	13,844	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290	54,052
27	9,803	11,808	12,879	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645	55,476
28	10,391	12,461	13,565	15,308	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,993	56,892
29	10,986	13,121	14,256	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,336	58,301
30	11,588	13,787	14,953	16,791	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672	59,703

A.3 Lois de Student

La table de quantiles de la loi de Student page [128](#) donne pour un choix de probabilités q usuelles les valeurs des réels x_q tels que

$$P(U > x_q) = q,$$

où U suit la loi de Student à d degrés de liberté.

Table des valeurs x_q telles que $P(U > x_q) = q$, pour U de loi de Student à d degrés de liberté

$d \backslash q$	0,45	0,4	0,35	0,3	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704
50	0,126	0,255	0,388	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,403	2,678
100	0,126	0,254	0,386	0,526	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626

A.4 Test de Kolmogorov Smirnov

La table ci-dessous donne la valeur critique t_ε telle que $P(D_n > t_\varepsilon) = \varepsilon$ pour le test de Kolmogorov Smirnov, avec $D_n := \|F_n - F\|_\infty$, F_n étant la fonction de répartition empirique.

$n \backslash \varepsilon$	0,2	0,1	0,05	0,02	0,01
4	0,492 6	0,565 2	0,623 9	0,688 9	0,734 2
5	0,447 0	0,509 4	0,563 3	0,627 2	0,668 5
6	0,410 4	0,468 0	0,519 3	0,577 4	0,616 6
7	0,381 5	0,436 1	0,483 4	0,538 4	0,575 8
8	0,358 3	0,409 6	0,454 3	0,506 5	0,541 8
9	0,339 1	0,387 5	0,430 0	0,479 6	0,513 3
10	0,322 6	0,368 7	0,409 3	0,456 6	0,488 9
11	0,308 3	0,352 4	0,391 2	0,436 7	0,467 7
12	0,295 8	0,338 2	0,375 4	0,419 2	0,449 1
13	0,284 7	0,325 5	0,361 4	0,403 6	0,432 5
14	0,274 8	0,314 2	0,348 9	0,389 7	0,417 6
15	0,265 9	0,304 0	0,337 6	0,377 1	0,404 2
16	0,257 8	0,294 7	0,327 3	0,365 7	0,392 0
17	0,250 4	0,286 3	0,318 0	0,355 3	0,380 9
18	0,243 6	0,278 5	0,309 4	0,345 7	0,370 6
19	0,237 4	0,271 4	0,301 4	0,336 8	0,361 2
20	0,231 6	0,264 7	0,294 1	0,328 7	0,352 4
21	0,226 2	0,258 6	0,287 2	0,321 0	0,344 3
22	0,221 2	0,252 8	0,280 9	0,313 9	0,336 7
23	0,216 5	0,247 5	0,274 9	0,307 3	0,329 5
24	0,212 0	0,242 4	0,269 3	0,301 0	0,322 9
25	0,207 9	0,237 7	0,264 0	0,295 2	0,316 6
26	0,204 0	0,233 2	0,259 1	0,289 6	0,310 6
27	0,200 3	0,229 0	0,254 4	0,284 4	0,305 0
28	0,196 8	0,225 0	0,249 9	0,279 4	0,299 7
29	0,193 5	0,221 2	0,245 7	0,274 7	0,294 7
30	0,190 3	0,217 6	0,241 7	0,270 2	0,289 9
31	0,187 3	0,214 1	0,237 9	0,266 0	0,285 3
32	0,184 5	0,210 9	0,234 2	0,261 9	0,280 9
33	0,181 7	0,207 7	0,230 8	0,258 0	0,276 8
34	0,179 1	0,204 7	0,227 4	0,254 3	0,272 8
35	0,176 6	0,201 9	0,224 3	0,250 7	0,269 0
36	0,174 2	0,199 1	0,221 2	0,247 3	0,265 3
37	0,171 9	0,196 5	0,218 3	0,244 0	0,261 8
38	0,169 7	0,193 9	0,215 4	0,240 9	0,258 4
39	0,167 5	0,191 5	0,212 7	0,237 9	0,255 2

$\varepsilon \backslash n$	0,2	0,1	0,05	0,02	0,01
40	0,165 5	0,189 1	0,210 1	0,234 9	0,252 0
41	0,163 5	0,186 9	0,207 6	0,232 1	0,249 0
42	0,161 6	0,184 7	0,205 2	0,229 4	0,246 1
43	0,159 7	0,182 6	0,202 8	0,226 8	0,243 3
44	0,158 0	0,180 5	0,200 6	0,224 3	0,240 6
45	0,156 2	0,178 6	0,198 4	0,221 8	0,238 0
46	0,154 6	0,176 7	0,196 3	0,219 4	0,235 4
47	0,153 0	0,174 8	0,194 2	0,217 2	0,233 0
48	0,151 4	0,173 0	0,192 2	0,214 9	0,230 6
49	0,149 9	0,171 3	0,190 3	0,212 8	0,228 3
50	0,148 4	0,169 6	0,188 4	0,210 7	0,226 0
51	0,147 0	0,168 0	0,186 6	0,208 6	0,223 9
52	0,145 6	0,166 4	0,184 8	0,206 7	0,221 7
53	0,144 2	0,164 8	0,183 1	0,204 7	0,219 7
54	0,142 9	0,163 3	0,181 4	0,202 9	0,217 7
55	0,141 6	0,161 9	0,179 8	0,201 1	0,215 7
56	0,140 4	0,160 4	0,178 2	0,199 3	0,213 8
57	0,139 2	0,159 1	0,176 7	0,197 6	0,212 0
58	0,138 0	0,157 7	0,175 2	0,195 9	0,210 2
59	0,136 9	0,156 4	0,173 7	0,194 3	0,208 4
60	0,135 7	0,155 1	0,172 3	0,192 7	0,206 7
65	0,130 5	0,149 1	0,165 7	0,185 2	0,198 8
70	0,125 9	0,143 8	0,159 8	0,178 6	0,191 7
75	0,121 7	0,139 0	0,154 4	0,172 7	0,185 3
80	0,117 9	0,134 7	0,149 6	0,167 3	0,179 5
85	0,114 4	0,130 7	0,145 2	0,162 4	0,174 2
90	0,111 3	0,127 1	0,141 2	0,157 9	0,169 4
95	0,108 3	0,123 7	0,137 5	0,153 7	0,164 9
100	0,105 6	0,120 7	0,134 0	0,149 9	0,160 8