



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

# Lois des grands nombres

Charles SUQUET

2004–2005

# Lois des grands nombres

Notations usuelles : les  $X_k$  sont des variables aléatoires réelles indépendantes et

$$S_n := \sum_{k=1}^n X_k.$$

On s'intéresse à la convergence des moyennes  $n^{-1}S_n$ . En préambule, il convient de mentionner la loi du zéro-un de Kolmogorov.

**Théorème 1 (Loi 0-1)** Soit  $(X_k)$  une suite de variables aléatoires indépendantes. On définit sa tribu d'événements asymptotiques

$$\mathcal{F}_\infty := \bigcap_{n \in \mathbb{N}} \sigma(X_k; k \geq n).$$

Si  $A \in \mathcal{F}_\infty$ ,  $P(A) = 0$  ou  $P(A) = 1$ .

**Preuve :** Voir Billingsley [2], Barbe Ledoux [1], Revuz [9]. □

L'évènement  $\{S_n/n \text{ converge}\}$  est dans  $\mathcal{F}_\infty$ , on sait donc dès la départ que sa probabilité vaut zéro ou 1. Dans le cas où elle vaut 1, la variable aléatoire limite  $S$  est  $\mathcal{F}_\infty$ -mesurable. En particulier pour tout  $x \in \mathbb{R}$ ,  $\{S \leq x\} \in \mathcal{F}_\infty$ . Donc  $P(S \leq x) = 0$  ou 1. Ceci implique que la fonction de répartition de  $S$  est de la forme  $F = \mathbf{1}_{[c, +\infty[}$  pour une certaine constante  $c$ . Autrement dit,  $S = c$  p.s., la limite lorsqu'elle existe ne peut être qu'une v.a. constante.

## 1 Cas des variables aléatoires bornées

### 1.1 Une inégalité exponentielle

**Théorème 2** On suppose les variables aléatoires réelles  $X_k$  indépendantes et identiquement distribuées, centrées ( $\mathbb{E} X_1 = 0$ ) et bornées ( $\exists c > 0; |X_1| \leq c$  p.s.). Alors

$$\forall \varepsilon > 0, \quad P\left(\left|\frac{S_n}{n}\right| \geq \varepsilon\right) \leq 2 \exp\left(-n \frac{\varepsilon^2}{2c^2}\right). \quad (1)$$

**Commentaires :** Pour comprendre la signification de ce théorème, comparons avec ce que l'on obtient lorsque les  $X_k$  sont gaussiennes  $N(0, 1)$ . Dans ce cas, la loi de  $S_n^* := S_n/\sqrt{n}$  est aussi  $N(0, 1)$ , d'où  $P(|S_n|/n \geq \varepsilon) = P(|S_n^*| \geq \varepsilon\sqrt{n}) \leq \exp(-n\varepsilon^2/2)$  en utilisant l'inégalité élémentaire<sup>1</sup>  $P(|X| \geq t) \leq \exp(-t^2/2)$  pour tout  $t > 0$  lorsque

---

1. Par changement de variable  $x = t + u$  dans  $P(X \geq t) = \int_t^{+\infty} \exp(-x^2/2) \frac{dx}{\sqrt{2\pi}}$  et  $\exp(-ut) \leq 1 \dots$

$X \sim N(0, 1)$ . Ainsi lorsque les  $X_k$  sont bornées, le comportement asymptotique de  $S_n$  est analogue à celui du cas gaussien. [ Par ailleurs, le théorème central limite nous fait pressentir qu'on ne peut espérer mieux. Remarquer aussi que chaque  $S_n$  est une v.a. bornée, mais que la suite  $(S_n)$  n'est pas bornée p.s.]. On pourra trouver le théorème 2 dans Ouvrard [8, Ex. 10.11, p. 132] ou Toulouse [11, Th. 1.4, p. 14].

**Preuve :** L'idée est d'exploiter l'existence de moments exponentiels  $\mathbb{E} \exp(tS_n)$  en faisant de l'optimisation par rapport au paramètre  $t$ . On remarque d'abord que pour tout  $t > 0$ ,

$$P\left(\frac{S_n}{n} \geq \varepsilon\right) = P(tS_n \geq nt\varepsilon) = P(\exp(tS_n) \geq e^{nt\varepsilon}).$$

L'inégalité de Markov puis l'indépendance et l'équidistribution des  $X_i$  nous donnent alors :

$$P\left(\frac{S_n}{n} \geq \varepsilon\right) \leq \frac{\mathbb{E} \exp(tS_n)}{\exp(nt\varepsilon)} = e^{-nt\varepsilon} (\mathbb{E} \exp(tX_1))^n. \quad (2)$$

Ceci nous amène à chercher une bonne majoration de  $\mathbb{E} \exp(tX_1)$ . En représentant tout  $x \in [-c, c]$  sous la forme  $x = -cu + c(1 - u)$  avec  $u \in [0, 1]$ , la convexité de  $\exp(t \cdot) : x \mapsto \exp(tx)$  nous donne

$$\exp(tx) \leq ue^{-ct} + (1 - u)e^{ct}. \quad (3)$$

En appliquant le paramétrage de  $[-c, c]$  à  $x = X_1(\omega)$  avec le  $u = U(\omega)$  correspondant, on voit que la variable aléatoire  $U$  vérifie  $2U = 1 - X_1/c$ , d'où  $\mathbb{E} U = 1/2$  puisque  $\mathbb{E} X_1 = 0$ . Compte tenu de (3), il vient

$$\mathbb{E} \exp(tX_1) \leq \mathbb{E} U e^{-ct} + (1 - \mathbb{E} U) e^{ct} = \text{ch}(ct). \quad (4)$$

En raison de l'exposant  $n$  dans le deuxième membre de (2), il est commode de majorer  $\text{ch}(ct)$  par une exponentielle bien choisie. Le développement en série entière

$$\text{ch}(ct) = 1 + \frac{c^2 t^2}{2} + \sum_{k=2}^{+\infty} \frac{(ct)^{2k}}{(2k)!}$$

nous suggère de choisir  $\exp(c^2 t^2 / 2)$ . L'inégalité

$$\text{ch}(ct) \leq \exp(c^2 t^2 / 2), \quad \forall t \in \mathbb{R}, \quad (5)$$

peut se vérifier en comparant terme à terme les développements en série entière. En effet

$$\frac{(ct)^{2k}}{(2k)!} \leq \frac{1}{k!} \left(\frac{c^2 t^2}{2}\right)^k \Leftrightarrow \frac{1}{(2k)!} \leq \frac{1}{2^k k!} \Leftrightarrow 2^k \leq (k+1)(k+2) \cdots (k+k)$$

et cette dernière inégalité est clairement vérifiée dès que  $k \geq 1$ . En revenant à (2), on a donc montré que pour tout  $t > 0$ ,  $P(n^{-1}S_n \geq \varepsilon) \leq \exp(-nt\varepsilon + nc^2 t^2 / 2)$ . Comme le premier membre de cette inégalité ne dépend pas de  $t$ , on optimise en écrivant

$$P(n^{-1}S_n \geq \varepsilon) \leq \inf_{t>0} \exp(-nt\varepsilon + nc^2 t^2 / 2) = \exp\left(n \inf_{t>0} (-t\varepsilon + c^2 t^2 / 2)\right).$$

Le minimum étant atteint en  $t = \varepsilon/c^2$ , on obtient

$$\forall \varepsilon > 0, \quad P\left(\frac{S_n}{n} \geq \varepsilon\right) \leq \exp\left(-n \frac{\varepsilon^2}{2c^2}\right). \quad (6)$$

En remplaçant  $X_k$  par  $-X_k$  dans la démonstration précédente on a immédiatement

$$\forall \varepsilon > 0, \quad P\left(\frac{S_n}{n} \leq -\varepsilon\right) \leq \exp\left(-n \frac{\varepsilon^2}{2c^2}\right), \quad (7)$$

ce qui joint à (6), donne (1).  $\square$

Comme sous-produit de la démonstration précédente, on a établi au passage le résultat suivant (noter que la convexité de  $x \mapsto \exp(tx)$  ne dépend pas du signe de  $t$ ).

**Lemme 3** *Si  $\mathbb{E} X = 0$  et s'il existe  $c$  constante telle que  $P(|X| \leq c) = 1$ , alors*

$$\forall t \in \mathbb{R}, \quad \mathbb{E} \exp(tX) \leq \exp\left(\frac{c^2 t^2}{2}\right). \quad (8)$$

## 1.2 LFGN pour des variables aléatoires i.i.d. bornées

Le théorème 2 donne facilement<sup>2</sup> la loi forte des grands nombres suivante par une simple utilisation du premier lemme de Borel-Cantelli et la discrétisation du  $\varepsilon$ .

**Théorème 4** *Soit  $(X_k)_{k \geq 1}$  une suite de variables aléatoires indépendantes, de même loi telle que pour une constante  $c$ ,  $|X_1| \leq c$  presque sûrement. Alors*

$$\frac{S_n}{n} \rightarrow \mathbb{E} X_1 \quad \text{p.s.} \quad (9)$$

Une application importante de ce théorème est la convergence des fréquences de succès dans une suite d'épreuves répétées de Bernoulli indépendantes. Ce résultat explique *a posteriori* l'approche fréquentiste dans la définition d'une probabilité. À titre d'exemple historique, on peut mentionner le problème de l'aiguille de Buffon. Le théorème 4 a une traduction statistique fondamentale : il permet de justifier la convergence de la fonction de répartition empirique. Considérons une suite  $(Y_k)$  de variables aléatoires indépendantes et de même loi de fonction de répartition  $F$ . On définit la *fonction de répartition empirique*  $F_n$  construite sur l'échantillon  $Y_1, \dots, Y_n$  par

$$F_n(x) := \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{Y_k \leq x\}}, \quad x \in \mathbb{R}. \quad (10)$$

Le théorème 4 appliqué aux variables aléatoires bornées  $X_k = \mathbf{1}_{\{Y_k \leq x\}}$  nous donne immédiatement pour tout  $x \in \mathbb{R}$  la convergence presque sûre de  $F_n(x)$  vers  $F(x)$  en remarquant que  $\mathbb{E} X_1 = P(Y_1 \leq x) = F(x)$ . Ainsi une loi inconnue peut être reconstituée approximativement à partir de l'observation d'un échantillon de grande taille. En fait, on peut obtenir mieux que la convergence simple presque sûre de  $F_n$  vers  $F$ .

---

2. Pour une preuve détaillée, voir Th. 23 dans l'annexe A.

**Théorème 5 (Glivenko-Cantelli)** Soit  $(Y_k)$  une suite de variables aléatoires indépendantes, de même loi et  $(F_n)$  la suite de fonctions de répartition empiriques associées. Alors

$$\|F_n - F\|_\infty := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0, \quad \text{p.s.} \quad (11)$$

**Preuve :** Voir Billingsley [2], Th. 20.6 p. 269. Voir aussi Ouvrard [8] pp. 115–121 incluant une digression sur le test de Kolmogorov-Smirnov dans le cas non asymptotique ( $n$  petit).  $\square$

La LFGN pour des variables aléatoires bornées donne aussi immédiatement la convergence presque sûre des fonctions caractéristiques empiriques.

**Proposition 6** Soit  $(Y_k)$  une suite de vecteurs aléatoires dans  $\mathbb{R}^d$ , indépendants et de même loi de fonction caractéristique  $\varphi$  définie par  $\varphi(u) := \mathbb{E} \exp(i\langle u, Y_1 \rangle)$ ,  $u \in \mathbb{R}^d$ . Alors la fonction caractéristique empirique

$$\varphi_n(u) := \frac{1}{n} \sum_{k=1}^n \exp(i\langle u, Y_k \rangle)$$

converge ponctuellement presque sûrement sur  $\mathbb{R}^d$  vers  $\varphi$ .

**Preuve :** Il suffit d'appliquer le théorème 4 aux variables aléatoires  $X'_k = \cos(\langle u, Y_k \rangle)$  et  $X''_k = \sin(\langle u, Y_k \rangle)$ .  $\square$

## 1.3 Autres applications et illustrations du cas borné

### 1.3.1 Entonneurs déterministes pour les fréquences

L'inégalité exponentielle (1) permet une approche « quantitative » de la convergence p.s. des moyennes  $S_n/n$ . En effet, (1) nous donne un contrôle explicite du reste de série  $\sum_{k>n} P(|S_k/k| > \varepsilon)$ . En prenant  $\varepsilon = \varepsilon_n$  tendant vers 0 à une vitesse adéquate, on peut avec une probabilité  $1 - \delta$ , encadrer  $S_n/n$  à partir d'un rang *déterministe*  $n_0 = n_0(\delta)$  et ce *jusqu'à l'infini* entre les deux suites déterministes  $\mathbb{E} X_1 - \varepsilon_n$  et  $\mathbb{E} X_1 + \varepsilon_n$ . Pour préciser cette idée, nous allons considérer le cas où  $S_n$  suit la loi binomiale  $\text{Bin}(n, p)$ . En vue des simulations, on va d'abord donner une version affinée du théorème 2, en effet la constante  $c$  dans (1) intervenant à l'intérieur de l'exponentielle, il n'est pas du tout indifférent en pratique de pouvoir la minimiser.

**Théorème 7** Si les  $X_k$  sont indépendantes, identiquement distribuées et s'il existe des constantes  $a$  et  $b$  telles que  $P(a \leq X_1 \leq b) = 1$ , alors

$$\forall \varepsilon > 0, \quad P\left(\left|\frac{S_n - \mathbb{E} S_n}{n}\right| \geq \varepsilon\right) \leq 2 \exp\left(-n \frac{2\varepsilon^2}{(b-a)^2}\right). \quad (12)$$

Remarquons qu'avec des bornes symétriques  $a = -c$  et  $b = c$ , on retrouve exactement (1). Le gain est dans le cas non symétrique (pour le voir, comparer les majorants fournis par (1) et (12) lorsque  $X_1$  suit une loi de Bernoulli de paramètre  $p \neq 1/2$ ). La preuve est tout à fait analogue à celle du théorème 2, le seul point méritant d'être explicité est la version améliorée du lemme 3 que l'on pourra trouver dans [4, Th. I.2 p. 41].

**Lemme 8** *Si la variable aléatoire réelle  $X$  est telle que  $P(a \leq X \leq b) = 1$ ,*

$$\forall t \in \mathbb{R}, \quad \mathbb{E} \exp(t(X - \mathbb{E} X)) \leq \exp\left(\frac{(b-a)^2 t^2}{8}\right). \quad (13)$$

**Preuve :** Comme le majorant cherché ne dépend des bornes  $a$  et  $b$  que par  $b - a$ , on ne perd pas de généralité en se ramenant au cas où  $\mathbb{E} X = 0$  (cela revient à remplacer  $a$  par  $a' = a - \mathbb{E} X$ ,  $b$  par  $b' = b - \mathbb{E} X$  et ne change pas le résultat final). De plus, quitte à remplacer  $t$  par  $(b - a)t$  et  $X$  par  $X/(b - a)$ , on voit qu'il suffit de prouver le résultat pour  $b - a = 1$ . L'argument de convexité utilisé pour obtenir (4) donne ici :

$$\mathbb{E} \exp(tX) \leq b \exp(at) - a \exp(bt) =: f(t)$$

(noter que comme  $\mathbb{E} X = 0$ ,  $a \leq 0$  et  $b \geq 0$ ). Posons  $g(t) := \ln(f(t))$ . Pour montrer que  $f(t) \leq \exp(t^2/8)$ , on va vérifier que  $g''(t) \leq 1/4$ . On obtient successivement

$$\begin{aligned} g'(t) &= ab \frac{e^{at} - e^{bt}}{f(t)}; \\ g''(t) &= ab \frac{(-a^2 - b^2 + 2ab)e^{(a+b)t}}{(be^{at} - ae^{bt})^2} \\ &= \frac{-abe^{(a+b)t}}{(be^{at} - ae^{bt})^2}. \end{aligned}$$

L'inégalité  $(x + y)^2 \geq 4xy$  avec  $x = -a \exp(bt)$  et  $y = b \exp(at)$  donne  $g''(t) \leq 1/4$ . Comme  $g'(0) = 0$  et  $g(0) = 0$ , on en déduit par intégration que  $g(t) \leq t^2/8$ .  $\square$

On peut utiliser l'inégalité (12) pour étudier quantitativement les fluctuations asymptotiques de  $S_n/n$  autour de  $\mathbb{E} X_1$ . Pour simplifier, on suppose désormais  $a = 0$  et  $b = 1$ . On vérifie alors facilement (faites le!) que pour tout entier  $N \geq 2$  et tout  $\alpha > 1/2$ ,

$$\mathbf{P} \left\{ \forall k > N, \left| \frac{S_k}{k} - \mathbb{E} X_1 \right| \leq \sqrt{\frac{\alpha \ln k}{k}} \right\} \geq 1 - \frac{2}{2\alpha - 1} N^{1-2\alpha}. \quad (14)$$

Par exemple avec  $\alpha = 1$ ,

$$\mathbf{P} \left\{ \forall k > 200, \mathbb{E} X_1 - \sqrt{\frac{\ln k}{k}} \leq \frac{S_k}{k} \leq \mathbb{E} X_1 + \sqrt{\frac{\ln k}{k}} \right\} \geq 0,99.$$

La représentation graphique nous donne un entonnoir déterministe qui avec une probabilité d'au moins 0,99 encadre *jusqu'à l'infini* la ligne polygonale de sommets  $(k, S_k/k)$ . Les figures 1.1 et 1.2 ont été réalisées à partir d'un échantillon simulé de taille 32 000 ( voir TP en Scilab).

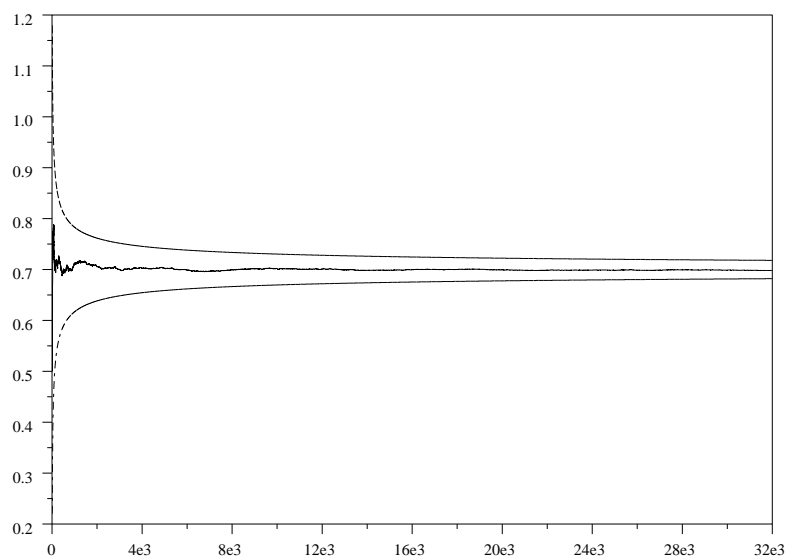


FIGURE 1.1 – Entonnoir pour  $X_1 \sim \text{Ber}(0.7)$  et  $\alpha = 1$

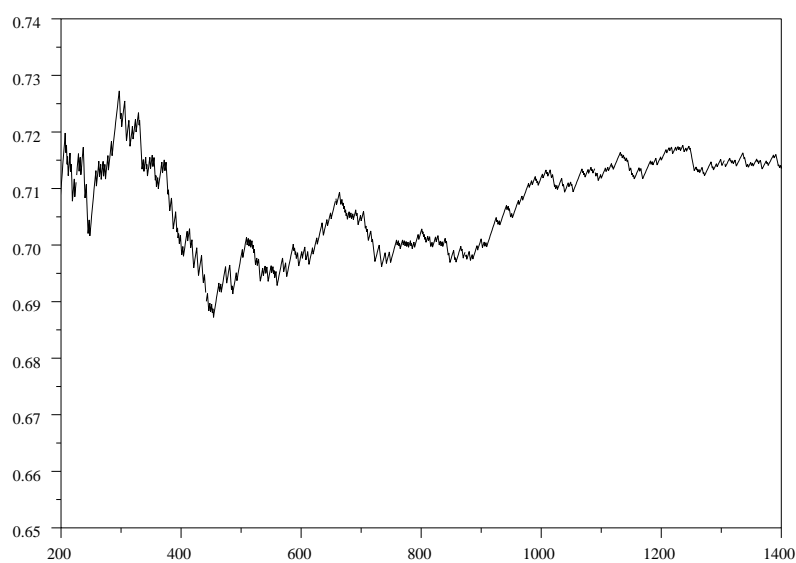


FIGURE 1.2 – Zoom sur l'entonnoir de la figure 1.1

### 1.3.2 Nombres normaux de Borel et mesures singulières

Thème assez classique, que l'on retrouve dans de nombreux ouvrages, par exemple [1, ex. 6.14 p. 154].

**Ex 1.3.1** Les  $X_k$  sont des v.a. de Bernoulli indépendantes de même paramètre  $p$ . La série  $\sum_{k \geq 1} 2^{-k} X_k$  convergeant p.s. (pourquoi?), on note  $U$  sa somme. La loi de cette variable aléatoire  $U$  qui est donc une mesure de probabilité sur  $[0, 1]$  sera notée  $\mu_p$ . Pour  $x$  réel de  $[0, 1]$ , on note  $\sum_{k \geq 1} x_k 2^{-k}$  son développement propre en base 2.

1) En utilisant la loi forte des grands nombres pour les fréquences, montrer que pour  $\mu_p$ -presque tout  $x$  de  $[0, 1]$ , la proportion de 1 dans le développement propre de  $x$  en base 2 tend vers  $p$ . En déduire que les lois  $\mu_p$  sont étrangères les unes aux autres.

2) Montrer que  $\mu_{1/2}$  est la mesure de Lebesgue (ou loi uniforme) sur  $[0, 1]$ . *Indication* : calculer la fonction caractéristique de  $U_n := \sum_{1 \leq k \leq n} 2^{-k} X_k$ , en utilisant l'identité :

$$\sin t = 2 \sin(t/2) \cos(t/2) = 2^2 \sin(t/4) \cos(t/2) \cos(t/4) = \dots = 2^n \sin(2^{-n}t) \prod_{k=1}^n \cos(2^{-k}t).$$

3) On suppose  $0 < p < 1$ . Montrer que la mesure  $\mu_p$  n'a pas de masse ponctuelle ( $\forall x \in [0, 1], \mu_p(\{x\}) = 0$ ). On a ainsi construit une infinité de mesures singulières à fonctions de répartition continues.

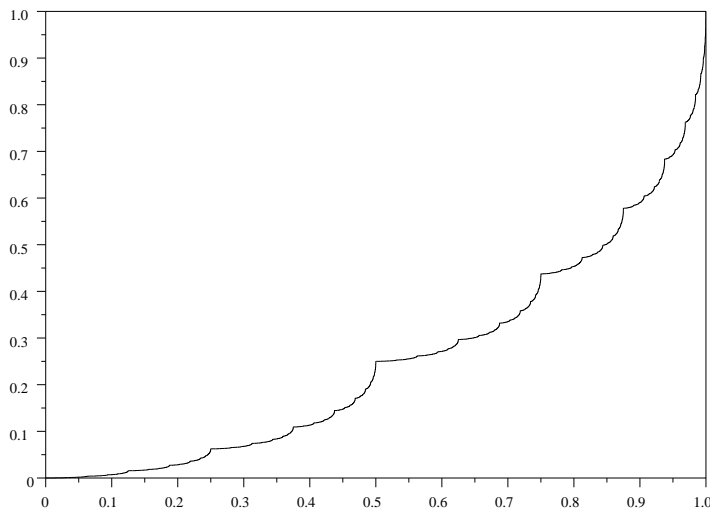


FIGURE 1.3 – Fonction de répartition de  $U$  pour  $p = 3/4$



### 1.3.3 Vitesse de convergence des polynômes de Bernstein

Thème un peu à la limite de cette leçon, donné sous forme d'exercices.

**Ex 1.3.2** Une preuve probabiliste d'un théorème d'analyse

Le but de cet exercice est de présenter une démonstration probabiliste d'un célèbre théorème d'analyse (Bernstein-Weierstrass-Stone) : toute fonction continue sur  $[0, 1]$  est limite *uniforme* sur cet intervalle d'une suite de polynômes. La méthode utilisée ici est due à Bernstein et donne une construction explicite de la suite de polynômes. Les trois dernières questions sont consacrées à la vitesse de convergence. On note  $\mathcal{C}[0, 1]$  l'espace des fonctions continues sur  $[0, 1]$  muni de la norme :

$$\|f\|_\infty = \sup_{x \in [0, 1]} |f(x)|.$$

La convergence suivant cette norme n'est autre que la convergence uniforme. Si  $f \in \mathcal{C}[0, 1]$ , on définit son polynôme de Bernstein d'ordre  $n$  par :

$$B_n f(x) = \sum_{k=0}^n C_n^k f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}, \quad n \geq 1.$$

1) Justifier la relation :

$$f(x) = \sum_{k=0}^n C_n^k f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

2) Pour  $x \in [0, 1]$  fixé, considérons la variable aléatoire  $S_n$  de loi binomiale  $\mathcal{B}(n, x)$ . Vérifier que :

$$\mathbb{E} f\left(\frac{S_n}{n}\right) = B_n f(x).$$

3) Justifier les inégalités :

$$\begin{aligned} |f(x) - B_n f(x)| &\leq \sum_{k: |f(x) - f(k/n)| < \varepsilon} \varepsilon C_n^k x^k (1-x)^{n-k} \\ &\quad + \sum_{k: |f(x) - f(k/n)| \geq \varepsilon} 2\|f\|_\infty C_n^k x^k (1-x)^{n-k} \\ &\leq \varepsilon + 2\|f\|_\infty P\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right). \end{aligned} \quad (15)$$

4) La fonction  $f$  est uniformément continue sur  $[0, 1]$  (pourquoi?). On a donc :

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \text{tel que } |x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon,$$

$\delta$  ne dépendant que de  $f$  et  $\varepsilon$ , mais pas de  $x$ . En déduire que

$$P\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \leq P(|S_n - nx| \geq n\delta),$$

puis en appliquant l'inégalité de Tchebycheff :

$$P\left(\left|f(x) - f\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \leq \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2}.$$

5) En reportant cette majoration dans (15), on obtient finalement :

$$\forall n \geq 1, \forall x \in [0, 1], \quad |f(x) - B_n f(x)| \leq \varepsilon + \frac{\|f\|_\infty}{2\delta^2 n} \quad (16)$$

Conclure.

6) On s'intéresse maintenant à la vitesse de convergence. Supposons d'abord que  $f$  est lipschitzienne : il existe une constante  $a$  telle que

$$\forall x, y \in [0, 1], \quad |f(x) - f(y)| \leq a|x - y|.$$

On peut alors prendre  $\delta = \varepsilon/a$  dans l'écriture de la continuité uniforme de  $f$ . En choisissant convenablement  $\varepsilon$  en fonction de  $n$  dans (16), en déduire que  $\|f - B_n f\|_\infty = O(n^{-1/3})$ .

7) Plus généralement, on suppose  $f$  hölderienne d'exposant  $\alpha$  : il existe des constantes  $0 < \alpha \leq 1$  et  $a > 0$  telles que :

$$\forall x, y \in [0, 1], \quad |f(x) - f(y)| \leq a|x - y|^\alpha.$$

Montrer qu'alors  $\|f - B_n f\|_\infty = O\left(n^{-\alpha/(2+\alpha)}\right)$ .

**Ex 1.3.3** *Vitesse de convergence des polynômes de Bernstein*

L'inégalité (12) permet d'améliorer les résultats de l'exercice précédent sur la vitesse de convergence uniforme des polynômes de Bernstein d'une fonction continue. L'utilisation de (12) à la place de l'inégalité de Tchebycheff nous donne en effet la majoration :

$$\|f - B_n f\|_\infty \leq \varepsilon + 4\|f\|_\infty \exp(-2n\delta^2). \quad (17)$$

1) On suppose  $f$  lipschitzienne. Vérifier que le choix  $\varepsilon = cn^{-\beta}$  dans (17) donne une vitesse de convergence en  $O(n^{-\beta})$  pour tout  $\beta < 1/2$ , mais que la même méthode ne permet pas d'obtenir la vitesse  $O(n^{-1/2})$ .

2) Toujours avec  $f$  lipschitzienne, comment choisir  $c$  minimal pour obtenir avec  $\varepsilon = c(\ln n/n)^{1/2}$  la vitesse  $O((\ln n/n)^{1/2})$  ?

3) On suppose maintenant  $f$  hölderienne d'exposant  $\alpha$ . Montrer qu'avec un choix judicieux de  $\varepsilon$ , on obtient la vitesse  $O((\ln n/n)^{\alpha/2})$ .

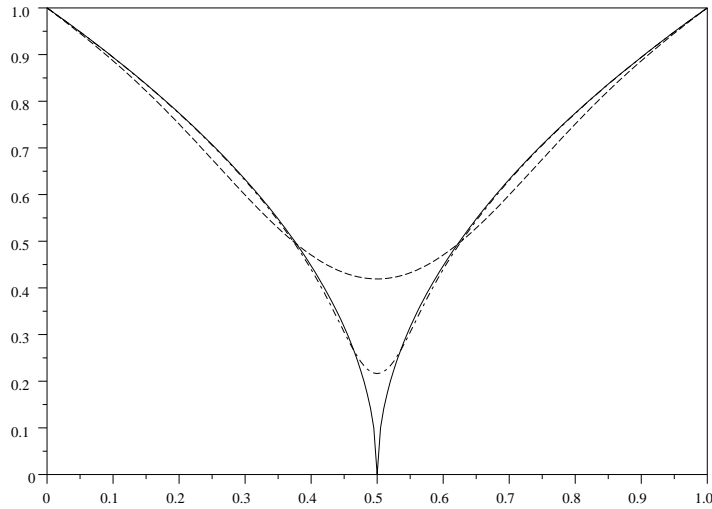


FIGURE 1.4 – Fonctions  $f$ ,  $B_{10}f$  et  $B_{200}f$ , pour  $f(t) = \sqrt{|2t - 1|}$

## 2 Une loi faible des grands nombres

La version naïve de la loi faible des grands nombres présente la convergence en probabilité de  $S_n/n$  vers  $\mathbb{E} X_1$  comme un corollaire immédiat de l'inégalité de Tchebycheff. Ceci a l'inconvénient de supposer l'existence d'un moment d'ordre 2 pour  $X_1$  (d'ailleurs c'est en fait la convergence  $L^2$  que l'on obtient)<sup>3</sup>. La version de la loi faible étudiée ci-dessous suppose seulement l'existence de l'espérance de  $X$ . On suit l'approche de Toulouse [11]. On pourra consulter également Foata Fuchs [7, p. 227–228] pour une version plus générale supposant seulement l'indépendance 2 à 2 et l'existence de  $\mathbb{E} X_1$ .

**Lemme 9** *Soit  $X$  une variable aléatoire d'espérance nulle. Pour tout  $\delta > 0$ , il existe une variable aléatoire bornée  $Y$  telle que  $\mathbb{E} Y = 0$  et  $\mathbb{E} |X - Y| \leq \delta$ .*

**Preuve :** Comme l'espérance de  $X$  existe,  $\mathbb{E} |X|$  est fini et par convergence dominée,  $\lim_{t \rightarrow +\infty} \mathbb{E} |X| \mathbf{1}_{\{|X| > t\}} = 0$ . Ceci nous fournit un  $t$  tel que

$$\mathbb{E} |X| \mathbf{1}_{\{|X| > t\}} < \frac{\delta}{2}.$$

On pose alors  $Z := X \mathbf{1}_{\{|X| \leq t\}}$  et  $Y := Z - \mathbb{E} Z$ . Clairement  $|Y| \leq 2t$  et  $\mathbb{E} Y = 0$ . D'autre part comme  $X = Z + X \mathbf{1}_{\{|X| > t\}}$ ,

$$\mathbb{E} |X - Y| = \mathbb{E} |X \mathbf{1}_{\{|X| > t\}} + \mathbb{E} Z| \leq \frac{\delta}{2} + |\mathbb{E} Z|.$$

3. L'avantage de cette méthode est que l'indépendance *deux à deux* ou plus généralement l'orthogonalité des  $X_k$  suffit.

La construction de  $Z$  nous assure que  $\mathbb{E} Z$  diffère de  $\mathbb{E} X$  d'au plus  $\delta/2$  :

$$|\mathbb{E} X - \mathbb{E} Z| = |\mathbb{E} X \mathbf{1}_{\{|X|>t\}}| \leq \mathbb{E} |X| \mathbf{1}_{\{|X|>t\}} < \frac{\delta}{2}.$$

Comme  $\mathbb{E} X = 0$ , on en déduit  $|\mathbb{E} Z| < \delta/2$  et  $\mathbb{E} |X - Y| < \delta$ .  $\square$

Notons que la variable bornée  $Y$  peut s'écrire  $Y = f_\delta(X)$ , où

$$f_\delta(x) = x \mathbf{1}_{[-t,t]}(x), \quad x \in \mathbb{R}.$$

La fonction *mesurable*  $f_\delta$  ne dépend que de  $\delta$  et de la loi de  $X$  (via le choix de  $t$ ).

**Théorème 10** *Si  $(X_k)$  est une suite de v. a. réelles i.i.d. et si  $\mathbb{E} |X_1| < +\infty$ ,  $n^{-1}S_n$  converge dans  $L^1(\Omega)$  vers  $\mathbb{E} X_1$  (et donc aussi en probabilité).*

**Commentaires :** Ce théorème n'est pas un simple corollaire de la LFGN de Kolmogorov-Khintchine (th. 11 ci-dessous) puisque la convergence p.s. implique la convergence en probabilité, mais n'implique pas la convergence dans  $L^1(\Omega)$ . La preuve de cette LFGN par les martingales permet d'obtenir les deux convergences p.s. et  $L^1(\Omega)$ , cf. Williams [12, chap. 14].

**Preuve :** Par centrage, on se ramène au cas où  $\mathbb{E} X_1 = 0$ . Fixons  $\delta > 0$ . Par le lemme 9, les variables aléatoires  $Y_k := f_\delta(X_k)$ , forment une suite i.i.d. de variables bornées ( $|Y_k| \leq 2t$ ), d'espérances nulles et telles que  $\mathbb{E} |X_k - Y_k| < \delta$  pour tout  $k \geq 1$ . Posons

$$T_n := \sum_{k=1}^n Y_k, \quad n \geq 1.$$

Par le théorème 4,  $n^{-1}T_n$  converge p.s. vers 0. Ceci joint à l'inégalité  $|n^{-1}T_n| \leq 2t$  nous donne par convergence dominée :

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left| \frac{T_n}{n} \right| = 0.$$

Par ailleurs,

$$\mathbb{E} \left| \frac{S_n}{n} - \frac{T_n}{n} \right| \leq \frac{1}{n} \sum_{k=1}^n |X_k - Y_k| \leq \delta.$$

On en déduit

$$\mathbb{E} \left| \frac{S_n}{n} \right| \leq \mathbb{E} \left| \frac{T_n}{n} \right| + \delta,$$

puis

$$\limsup_{n \rightarrow +\infty} \mathbb{E} \left| \frac{S_n}{n} \right| \leq 0 + \delta.$$

Cette inégalité étant valable pour tout  $\delta > 0$  et le premier membre ne dépendant pas de  $\delta$ , il en résulte que

$$\limsup_{n \rightarrow +\infty} \mathbb{E} \left| \frac{S_n}{n} \right| = 0.$$

Ceci peut se réécrire  $\lim_{n \rightarrow +\infty} \mathbb{E} |n^{-1}S_n| = 0$ , ce qui est exactement la convergence dans  $L^1$  de  $n^{-1}S_n$  vers 0.  $\square$

### 3 Lois fortes des grands nombres

On discute dans cette section les lois fortes des grands nombres dans le cas général où les  $X_k$  ne sont pas supposées bornées.

#### 3.1 Cas i.i.d.

**Théorème 11 (Kolmogorov-Khintchine)** *Si  $(X_k)$  est une suite de variables aléatoires réelles i.i.d.,*

$$\frac{S_n}{n} \text{ converge p.s.} \Leftrightarrow \mathbb{E}|X_1| < +\infty. \quad (18)$$

*Lorsqu'il y a convergence, la limite est  $\mathbb{E}X_1$ .*

Pour la partie  $\mathbb{E}|X_1|$  finie implique  $n^{-1}S_n$  converge p.s. vers  $\mathbb{E}X_1$ , une bonne référence est Billingsley [2, Th. 22.1]. On peut aussi voir Revuz [9] pour l'équivalence (méthode inspirée de techniques de martingales. On peut proposer à la démonstration la nécessité de  $\mathbb{E}|X_1|$  fini que nous détaillons ci-dessous (cf. Barbe Ledoux [1, Th. 5.2, p. 140]).

**Preuve de la nécessité de l'intégrabilité de  $X_1$  :** Par hypothèse, il existe une constante  $c$  telle que  $n^{-1}S_n$  converge p.s. vers  $c$ . Alors

$$\frac{X_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n} \times \frac{S_{n-1}}{n-1} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} c - c = 0.$$

En fixant  $\varepsilon > 0$ , on en déduit

$$P\left\{\omega \in \Omega; \exists n_0 = n_0(\omega), \forall n \geq n_0, \frac{|X_n(\omega)|}{n} < \varepsilon\right\} = 1,$$

soit en passant au complémentaire

$$P\left\{\frac{|X_n|}{n} \geq \varepsilon \text{ une infinité de fois}\right\} = 0.$$

Par le second lemme de Borel Cantelli<sup>4</sup> on a alors

$$\sum_{n \geq 1} P\left\{\frac{|X_n|}{n} \geq \varepsilon\right\} < +\infty.$$

Les  $X_i$  ayant même loi, ceci s'écrit

$$\sum_{n=1}^{+\infty} P(|X_1| \geq n\varepsilon) < +\infty. \quad (19)$$

---

4. Si  $(A_n)_{n \geq 1}$  est une suite d'événements *indépendants* telle que  $\sum_{n \geq 1} P(A_n) = +\infty$ , alors  $P(\limsup_{n \rightarrow +\infty} A_n) = 1$ .

Pour finir la preuve, on observe que

$$\begin{aligned} \mathbb{E} |X_1| &\leq \mathbb{E} \left( \sum_{n=0}^{+\infty} (n+1)\varepsilon \mathbf{1}_{\{n\varepsilon \leq |X_1| < (n+1)\varepsilon\}} \right) \\ &= \sum_{n=0}^{+\infty} (n+1)\varepsilon P(n\varepsilon \leq |X_1| < (n+1)\varepsilon) \\ &= \varepsilon \sum_{n=0}^{+\infty} P(n\varepsilon \leq |X_1|) < +\infty, \end{aligned}$$

la dernière ligne s'obtenant par sommation triangulaire à partir des décompositions en unions disjointes

$$\{|X_1| \geq n\varepsilon\} = \bigcup_{k \geq n} \{n\varepsilon \leq |X_1| < (n+1)\varepsilon\}.$$

□

### 3.2 LFGN sans équidistribution

**Théorème 12 (Kolmogorov)** *Soit  $(X_k)$  une suite de variables aléatoires indépendantes vérifiant :*

- a) pour tout  $k \geq 1$ ,  $\mathbb{E} X_k^2 < +\infty$  ;
- b) il existe une suite  $(a_k)$  de réels strictement positifs qui tend en croissant vers  $+\infty$  telle que

$$\sum_{k=1}^{+\infty} \frac{\text{Var } X_k}{a_k^2} < +\infty.$$

Alors  $(S_n - \mathbb{E} S_n)/a_n$  converge presque sûrement vers 0. Si de plus  $a_n^{-1} \mathbb{E} S_n \rightarrow m$ ,  $S_n/a_n$  converge p.s. vers  $m$ .

Les conditions de moments d'ordre 2 sont plus sévères qu'au théorème 11, mais il faut noter qu'on ne suppose plus les  $X_k$  de même loi. Une bonne référence pour la preuve de ce théorème est Feller [6], VII.8, Th. 2 et 3. On peut esquisser le schéma de la preuve qui repose sur les trois résultats suivants dont chacun a son intérêt propre.

**Théorème 13 (Inégalité maximale de Kolmogorov)** *Si les  $X_k$  sont indépendantes, centrées et de carrés intégrables*

$$\forall t > 0, \quad P\left(\max_{1 \leq k \leq n} |S_k| \geq t\right) \leq \frac{1}{t^2} \text{Var } S_n.$$

**Preuve :** Voir Billingsley [2], Th. 22.4 p. 287 ou Ouvrard [8], Th. 10.13 p. 101. □

Cette inégalité pour les maxima des sommes partielles permet d'établir une condition suffisante de convergence p.s. d'une série de v.a. indépendantes.

**Théorème 14** *Si les  $Y_k$  sont indépendantes, centrées et si  $\sum_{k \geq 1} \mathbb{E} Y_k^2 < +\infty$ , alors  $\sum_{k=1}^{+\infty} Y_k$  converge p.s.*

**Preuve :** Voir Billingsley [2], Th. 22.6 p. 289 ou Feller [6], VII.8, Th. 2. □

À ce stade, l'hypothèse b) du théorème 12 nous donne la convergence presque sûre de la série de terme général  $(X_k - \mathbb{E} X_k)/a_k$ . On complète la preuve du théorème 12 grâce au lemme d'analyse suivant.

**Lemme 15 (Kronecker)** *Soient  $(x_k)$  une suite de réels et  $(a_k)$  une suite de réels strictement positifs qui tend en croissant vers  $+\infty$ , telles que  $\sum_{k=1}^{+\infty} x_k/a_k$  converge. Alors*

$$\frac{1}{a_n} \sum_{k=1}^n x_k \rightarrow 0.$$

**Preuve :** Voir Feller [6], VII.8, Lemme 1 ou Ouvrard [8] Lemme 10.16 p. 105 qui note  $x_k$  ce que nous avons noté  $x_k/a_k$ . □

Il est intéressant de regarder ce que donne le théorème 12 dans le cas i.i.d.

**Corollaire 16** *Si les v.a.  $X_k$  sont i.i.d., centrées et de carrés intégrables,*

$$\frac{S_n}{\sqrt{n}(\ln n)^\beta} \rightarrow 0, \quad \text{p.s.}, \quad (20)$$

pour tout réel  $\beta > 1/2$ .

**Preuve :** Choisir  $a_k = k^{1/2}(\ln k)^\beta$  dans le théorème 12. □

La signification de ce corollaire est que si  $\mathbb{E} X_1^2 < +\infty$ , on a une vitesse de convergence dans la LFGN de Kolmogorov-Khintchine : en effet on peut alors écrire  $n^{-1}S_n(\omega) = \varepsilon_n(\omega)n^{-1/2}(\ln n)^\beta$ , avec  $\varepsilon_n \rightarrow 0$  presque sûrement. Remarquons qu'on ne peut espérer supprimer le facteur logarithmique dans (20) puisque qu'en raison du théorème central limite  $S_n/\sqrt{n}$  ne tend pas vers 0 en loi.

## 4 Fluctuations des sommes partielles

La loi forte des grands nombres nous donne dans le cas i.i.d. l'estimation  $S_n - \mathbb{E} S_n = o(n)$  avec probabilité 1. On a vu avec le corollaire 16 que s'il y a un moment d'ordre 2 cette estimation peut être améliorée. On examine plus précisément dans cette section la relation entre l'intégrabilité de  $X_1$  et les fluctuations asymptotiques de  $S_n$ . Une bonne référence pour toute cette section est Stout [10], p. 126–137.

**Théorème 17 (LFGN de Marcinkiewicz)** *Soit  $(X_k)_{k \geq 1}$  une suite de variables aléatoires i.i.d.*

- a) Si  $\mathbb{E}|X_1|^p < +\infty$  pour un  $p \in ]0, 1[$ ,  $\frac{S_n}{n^{1/p}} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ .
- b) Si  $\mathbb{E}|X_1|^p < +\infty$  pour un  $p \in [1, 2[$ ,  $\frac{S_n - n \mathbb{E} X_1}{n^{1/p}} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ .
- c) S'il existe un  $p \in ]0, 2[$  et une suite de constantes  $(b_n)$  tels que  $\frac{S_n - b_n}{n^{1/p}} \xrightarrow[n \rightarrow +\infty]{p.s.} 0$ , alors  $\mathbb{E}|X_1|^p < +\infty$ .

**Commentaires :** On notera l'exclusion du cas  $p = 2$  (et *a fortiori*  $p > 2$ ) en liaison avec le théorème de limite centrale. La LFGN de Marcinkiewicz dans le cas  $1 < p < 2$  donne une vitesse de convergence dans la LFGN de Kolmogorov-Khintchine (pourquoi?).

Lorsque  $X_1$  a un moment d'ordre 2, on a un résultat très précis sur les fluctuations de  $S_n - \mathbb{E} S_n$ , c'est le théorème suivant connu sous le nom *loi du log itéré*.

**Théorème 18 (Hartman Wintner, 1941)** *On suppose les  $X_k$  i.i.d. et de carré intégrable. On note  $\sigma^2 = \text{Var } X_1$  ( $\sigma > 0$ ). Alors presque sûrement,*

$$\liminf_{n \rightarrow +\infty} \frac{S_n - \mathbb{E} S_n}{\sigma \sqrt{2n \log \log n}} = -1 \quad \text{et} \quad \limsup_{n \rightarrow +\infty} \frac{S_n - \mathbb{E} S_n}{\sigma \sqrt{2n \log \log n}} = +1.$$

**Commentaires :** Ce théorème signifie que presque sûrement pour tout  $c \in ]0, 1[$ , la suite des  $S_n - \mathbb{E} S_n$  sortira une infinité de fois par le bas et une infinité de fois par le haut du segment  $[-c\sigma\sqrt{2n \log \log n}, +c\sigma\sqrt{2n \log \log n}]$  et qu'elle restera définitivement à partir d'un certain rang (aléatoire) dans  $[-c'\sigma\sqrt{2n \log \log n}, +c'\sigma\sqrt{2n \log \log n}]$  pour tout  $c' > 1$ . La loi du log itéré nous donne *avec probabilité 1* des entonnoirs déterministes de la forme

$$\mathbb{E} X_1 - c'\sigma\sqrt{\frac{2 \log \log k}{k}} \leq \frac{S_k}{k} \leq \mathbb{E} X_1 + c'\sigma\sqrt{\frac{2 \log \log k}{k}},$$

pour tout  $k \geq N$  avec  $N$  aléatoire et nous dit que ces entonnoirs sont les meilleurs possibles. Les entonnoirs déterministes du type (14) obtenus par des techniques élémentaires dans le cas des variables bornées sont asymptotiquement moins précis. Ils ont néanmoins l'avantage de donner un résultat quantitatif avec un  $N$  déterministe. Une très bonne lecture pour la loi du log itéré dans le cas du jeu de pile ou face est le chapitre 19 de Foata Fuchs [7]. Enfin on notera que la loi du log itéré fournit un exemple « naturel » de suite qui converge en probabilité mais pas presque sûrement. En effet, en raison du théorème de limite centrale, on vérifie (exercice!) que

$$\frac{S_n - \mathbb{E} S_n}{\sigma \sqrt{2n \log \log n}} \xrightarrow[n \rightarrow +\infty]{\text{Pr}} 0.$$

## 5 Applications

### 5.1 La méthode de Monte Carlo

La loi des grands nombres fournit une méthode de calcul approché d'intégrales, intéressante lorsque la fonction à intégrer est très irrégulière ou lorsque la dimension de



l'espace est élevée. Supposons que l'on veuille effectuer un calcul approché de

$$I := \int_{[0,1]^d} f(x) dx,$$

où  $f$  est Lebesgue intégrable sur  $[0, 1]^d$ . Soit  $(U_i)_{i \geq 1}$ , une suite de variables aléatoires indépendantes de même loi uniforme sur  $[0, 1]$ . On déduit facilement de la LFGN de Kolmogorov-Khintchine que :

$$\frac{1}{n} \sum_{k=1}^n f(U_{(k-1)d+1}, U_{(k-1)d+2}, \dots, U_{kd}) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E} f(U_1, \dots, U_d) = \int_{[0,1]^d} f(x) dx.$$

Le théorème de limite centrale permet ensuite d'obtenir un *intervalle de confiance* pour  $I$  si l'on a des hypothèses supplémentaires permettant de contrôler la variance de  $f(U_1, \dots, U_d)$ , par exemple  $f$  bornée...

## 5.2 Estimation de paramètres

La LFGN permet de définir des estimateurs convergents de paramètres d'une loi inconnue  $\mu$  (ou partiellement inconnue, par exemple on sait qu'il s'agit d'une loi de Poisson paramètre  $\alpha$  dont on ignore la valeur). Pour cela on utilise une suite d'observations indépendantes  $X_1(\omega), \dots, X_n(\omega)$  où les  $X_i$  sont i.i.d. de même loi  $\mu$ . On souhaite estimer un paramètre  $\theta$  de la forme  $\theta = \int_{\mathbb{R}} H d\mu$ . L'idée est de remplacer la mesure déterministe mais inconnue  $\mu$  par la mesure *aléatoire*  $\mu_n$  calculable à partir des observations :

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

Cette mesure est appelée *mesure empirique*. La fonction de répartition empirique déjà vue en (10) est simplement sa fonction de répartition :  $F_n(x) = \mu_n(]-\infty, x])$ . On propose d'estimer  $\theta$  par

$$\hat{\theta}_n := \int_{\mathbb{R}} H d\mu_n = \frac{1}{n} \sum_{i=1}^n H(X_i).$$

La définition de  $\theta$  suppose implicitement que  $H$  est  $\mu$  intégrable. Cette intégrabilité s'écrit encore  $\mathbb{E} |H(X_1)| < +\infty$ . Ainsi par la loi forte des grands nombres,

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E} H(X_1) = \int_{\mathbb{R}} H d\mu = \theta.$$

On dit que  $\hat{\theta}_n$  est un estimateur *fortement consistant* de  $\theta$ . Il est aussi *sans biais* puisque  $\mathbb{E} \hat{\theta}_n = \theta$ .

Cette méthode permet notamment d'estimer les moments de  $\mu$  : en prenant  $H(x) = x^r$ ,  $\theta = \mathbb{E} X_1^r = \int_{\mathbb{R}} x^r \mu(dx)$ . Le cas  $r = 1$  revêt une importance particulière. L'estimateur  $\hat{\theta}_n$  est alors simplement la *moyenne empirique*

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i.$$

On peut ainsi estimer notamment

- le paramètre  $p$  d'une loi de Bernoulli car  $\mathbb{E} X_1 = p$ ;
- le paramètre  $\alpha$  d'une loi de Poisson car  $\mathbb{E} X_1 = \alpha$ ;
- le paramètre  $m$  d'une loi  $N(m, \sigma^2)$  car  $\mathbb{E} X_1 = m$ ;
- le paramètre  $\theta$  d'une loi uniforme<sup>5</sup> sur  $[0, \theta]$  car  $\theta = 2 \mathbb{E} X_1$ .

Dans le même ordre d'idées, on peut estimer le paramètre  $a$  d'une loi exponentielle de densité  $f(t) = a \exp(-at) \mathbf{1}_{\mathbb{R}^+}(t)$  par  $\tilde{\theta}_n = 1/\bar{X}_n$ . En effet,  $\mathbb{E} X_1 = 1/a$ . On garde un estimateur fortement consistant, mais il n'est plus sans biais car  $\mathbb{E}(1/\bar{X}_n) \neq 1/\mathbb{E} X_1$ .

On peut de même estimer la variance  $\sigma^2$  d'une loi  $\mu$  d'espérance connue  $m$ . Il suffit de prendre  $H(x) = (x - m)^2$  et on obtient l'estimateur fortement consistant et sans biais

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2.$$

Quand  $m$  est inconnu, on l'estime par  $\bar{X}_n$  et la variance  $\sigma^2$  est estimée par la *variance empirique*

$$V_n = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2,$$

la dernière égalité résultant simplement de la formule de Koenig pour la variance de la loi de probabilité  $\mu_n(\omega)$  qui est exactement  $V_n(\omega)$ . On a toujours un estimateur fortement consistant par la LFGN, par contre il n'est plus qu'*asymptotiquement* sans biais puisque :

$$\mathbb{E} V_n = \mathbb{E} X_1^2 - \mathbb{E} \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2 = \dots = \frac{n-1}{n} \sigma^2.$$

Ceci explique pourquoi pour les petites valeurs de  $n$  on préfère l'estimateur sans biais  $\frac{n}{n-1} V_n$  noté souvent  $\sigma_{n-1}^2$  par un de ces abus d'écriture qui font le charme si particulier de la littérature statistique...

### 5.3 Estimation de la densité par fonctions orthogonales

On va estimer cette fois un paramètre de nature fonctionnelle (donc vivant a priori dans un espace de dimension infinie) de la loi inconnue  $\mu$  : sa densité  $f$ , en supposant qu'elle existe et qu'elle est dans  $L^2(\mathbb{R})$ . Le schéma général de la méthode est le suivant. On choisit une base hilbertienne  $(e_k)_{k \in \mathbb{N}}$  de  $L^2(\mathbb{R})$  et on fixe une version de chaque  $e_k$ . En pratique, celà ne pose pas de problème puisque les  $e_k$  sont des fonctions de Haar ou des fonctions bien régulières (base trigonométrique si  $f$  est à support compact, polynômes orthogonaux, ...). On a ainsi

$$f = \sum_{k=0}^{+\infty} a_k e_k, \quad \text{avec} \quad a_k = \int_{\mathbb{R}} e_k(t) f(t) dt,$$

---

5. En fait dans ce cas, un meilleur estimateur est  $\tilde{\theta}_n = \max_{1 \leq i \leq n} X_i$ , affaire à suivre...

la convergence de la série ayant lieu au sens  $L^2(\mathbb{R})$ .

On commence par projeter  $f$  sur s.e.v.  $\{e_0, \dots, e_N\}$ . On obtient ainsi une approximation déterministe

$$f_N = \sum_{k=0}^N a_k e_k, \quad \text{avec} \quad f_N \xrightarrow[N \rightarrow +\infty]{L^2} f.$$

Des convergences plus fortes sont possibles moyennant des hypothèses de régularité de  $f$  (penser au cas des séries de Fourier).

Ensuite on estime  $f_N$  en estimant ses  $N + 1$  coefficients  $a_k$  :

$$\widehat{f}_N := \sum_{k=0}^N \widehat{a}_{k,n} e_k \quad \text{où} \quad \widehat{a}_{k,n} := \frac{1}{n} \sum_{i=1}^n e_k(X_i).$$

Il serait plus correct de noter  $\widehat{f}_{N,n}$ , mais dans la suite on fera dépendre  $N$  de  $n$ , d'où l'abus de notation. Pour l'instant remarquons que par la LFGN,

$$\widehat{a}_{k,n} \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E} e_k(X_1) = \int_{\mathbb{R}} e_k(t) f(t) dt = a_k.$$

Ainsi pour  $N$  fixé,  $\widehat{f}_N$  converge p.s. vers  $f_N$  dans  $L^2(\mathbb{R})$  (en dimension finie, il suffit pour cela d'avoir la convergence de chaque composante sur la base).

La suite du jeu consiste à prendre  $N = N(n)$  tendant vers  $+\infty$  avec  $n$  (intuitivement beaucoup plus lentement) et à discuter le choix de  $N(n)$  en fonction d'hypothèses supplémentaires sur  $f$  (régularité, intégrabilité, ...) pour obtenir diverses convergences de  $\widehat{f}_N$  vers  $f$ . On pourra consulter à ce sujet Bosq Lecoutre [4, Chap. 9]. On peut envisager une illustration expérimentale de cette méthode avec Scilab (cf. TP).

# Annexe A

## Loi des grands nombres\*

Les inégalités de moment (Markov, Tchebycheff) ont d'importantes applications à la convergence de la moyenne arithmétique :

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

des  $n$  premiers termes d'une suite de v.a. indépendantes et de même loi. Ce type de résultat est connu sous le nom de loi des grands nombres. Nous en donnons un premier aperçu<sup>1</sup>.

### A.1 Deux modes de convergence

Pour commencer, il convient de préciser ce que l'on entend par convergence d'une suite de variables aléatoires  $(X_n)$  vers une v.a.  $X$ . Comme les  $X_n$  sont des applications de  $\Omega$  dans  $\mathbb{R}$ , le premier mode de convergence auquel on pense est la convergence *pour tout*  $\omega \in \Omega$  de la suite de réels  $X_n(\omega)$  vers le réel  $X(\omega)$ . Ceci correspond à la convergence simple d'une suite d'applications en analyse. Malheureusement pour le type de résultat que nous avons en vue, ce mode de convergence est trop restrictif. Pour la loi des grands nombres, même dans le cas le plus favorable<sup>2</sup>, on ne peut empêcher que la suite étudiée diverge pour une infinité de  $\omega$ . Ce qui sauve la situation est que l'ensemble de ces  $\omega$  a une probabilité nulle. Ceci nous amène à définir la convergence *presque sûre* :

#### Définition 19 (Convergence presque sûre)

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires et  $X$  une v.a. définies sur le même espace probabilisé  $(\Omega, \mathcal{F}, P)$ . On dit que  $X_n$  converge presque sûrement vers  $X$  si l'ensemble des

---

\*. Reproduction du chapitre 6 de *Introduction au Calcul des Probabilités*, cours de Deug, Ch. Suquet.

1. Seuls sont au programme du DEUG dans ce chapitre, la convergence en probabilité et la loi faible des grands nombres avec ses applications. La convergence presque sûre et la loi forte des grands nombres sont destinés aux lecteurs plus curieux ou plus avancés. Ils pourront être considérés comme une introduction au cours de Licence. Néanmoins ils ont été rédigés en n'utilisant que des outils mathématiques du DEUG.

2. Voir la discussion à propos de la loi forte des grands nombres pour les fréquences section A.5.

$\omega$  tels que  $X_n(\omega)$  converge vers  $X(\omega)$  a pour probabilité 1.

Notation :  $X_n \xrightarrow[n \rightarrow +\infty]{p.s.} X$ .

Rappelons qu'un événement de probabilité 1 n'est pas forcément égal à  $\Omega$ , il peut même y avoir une infinité d'éléments dans son complémentaire (par exemple si A et B lancent un dé à tour de rôle, le gagnant étant le premier à obtenir « six », l'évènement « il n'y a pas de gagnant » a une probabilité nulle mais est constitué d'une infinité non dénombrable d'évènements élémentaires, c'est  $\{1, \dots, 5\}^{\mathbb{N}^*}$ ). Remarquons aussi que l'ensemble  $\Omega'$  des  $\omega$  tels que  $X_n(\omega)$  converge vers  $X(\omega)$  est bien un événement observable (vu en exercice), c'est-à-dire un événement de la famille  $\mathcal{F}$ . Il est donc légitime de parler de sa probabilité.

Dans la convergence presque sûre, le rang  $n_0$  à partir duquel on peut approximer  $X_n(\omega)$  par  $X(\omega)$  avec une erreur inférieure à  $\varepsilon$  dépend à la fois de  $\varepsilon$  et de  $\omega \in \Omega'$  :  $n_0 = n_0(\varepsilon, \omega)$ . On ne sait pas toujours expliciter la façon dont  $n_0(\varepsilon, \omega)$  dépend de  $\omega$ . D'autre part on peut très bien avoir  $\sup\{n_0(\varepsilon, \omega), \omega \in \Omega'\} = +\infty$ . Ceci fait de la convergence presque sûre en général un résultat essentiellement théorique<sup>3</sup>. Supposons que la valeur de  $X_n$  dépende du résultat de  $n$  épreuves répétées (ou de  $n$  observations). Savoir que  $X_n$  converge presque sûrement vers  $X$  ne permet pas de *prédire* le nombre non aléatoire  $n$  d'épreuves (ou d'observations) à partir duquel on aura  $|X_n(\omega) - X(\omega)| < \varepsilon$  (sinon pour tous les  $\omega \in \Omega'$ , du moins avec une probabilité supérieure à un seuil fixé à l'avance par exemple 95%, 99%,...). Or cette question a une grande importance pratique pour le statisticien. C'est l'une des raisons de l'introduction de la *convergence en probabilité* qui permet de répondre à cette question lorsque l'on connaît la vitesse de convergence selon ce mode.

### Définition 20 (Convergence en probabilité)

Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires et  $X$  une v.a. définies sur le même espace probabilisé  $(\Omega, \mathcal{F}, P)$ . On dit que  $X_n$  converge en probabilité vers  $X$  si :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

Notation :  $X_n \xrightarrow[n \rightarrow +\infty]{Pr} X$ .

La convergence presque sûre implique la convergence en probabilité, la réciproque est fautive (exercices). Pour cette raison, la convergence en probabilité de la suite  $M_n$  définie en introduction s'appelle une loi *faible* des grands nombres, sa convergence presque sûre une loi *forte* des grands nombres.

## A.2 Loi faible des grands nombres

**Théorème 21** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires deux à deux indépendantes, de même loi ayant un moment d'ordre 2. Alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{Pr} \mathbb{E} X_1.$$

3. Sauf si l'on connaît la loi de la v.a.  $\omega \mapsto n_0(\varepsilon, \omega)$ , ou au moins si l'on sait majorer  $P(n_0 > t)$ ...

**Preuve :** Ici, la v.a. limite est la constante  $\mathbb{E} X_1$  (ou n'importe quel  $\mathbb{E} X_i$ , puisque les  $X_i$  ayant même loi ont même espérance). Il s'agit donc de vérifier que :

$$\forall \varepsilon > 0, \quad \lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mathbb{E} X_1\right| \geq \varepsilon\right) = 0.$$

Posons  $M_n = \frac{1}{n} \sum_{i=1}^n X_i$ . On a :

$$\mathbb{E} M_n = \frac{1}{n} \sum_{i=1}^n \mathbb{E} X_i = \mathbb{E} X_1. \quad (\text{A.1})$$

D'autre part, les  $X_i$  étant deux à deux indépendantes et de même loi on a :

$$\text{Var} M_n = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} (n \text{Var} X_1) = \frac{1}{n} \text{Var} X_1. \quad (\text{A.2})$$

L'inégalité de Tchebycheff appliquée à chaque  $M_n$  nous dit que pour  $\varepsilon > 0$  fixé :

$$\forall n \in \mathbb{N}^*, \quad P(|M_n - \mathbb{E} M_n| \geq \varepsilon) \leq \frac{\text{Var} M_n}{\varepsilon^2}.$$

D'où compte tenu du calcul de  $\mathbb{E} M_n$  et  $\text{Var} M_n$  :

$$\forall n \in \mathbb{N}^*, \quad P(|M_n - \mathbb{E} X_1| \geq \varepsilon) \leq \frac{\text{Var} X_1}{n\varepsilon^2}. \quad (\text{A.3})$$

En faisant tendre  $n$  vers  $+\infty$  ( $\varepsilon$  restant fixé) on en déduit :

$$\lim_{n \rightarrow +\infty} P(|M_n - \mathbb{E} X_1| \geq \varepsilon) = 0.$$

Ce raisonnement est valable pour tout  $\varepsilon > 0$ . □

Remarque : Nous avons en fait démontré un peu plus que la seule convergence en probabilité. Nous avons d'après (A.3) une *vitesse de convergence* en  $O(1/n)$ . Si l'on connaît  $\text{Var} X_1$  ou si on sait le majorer, on peut donc répondre à la question posée page 20 lors de l'introduction de la convergence en probabilité.

### Corollaire 22 (Loi faible des g. n. pour les fréquences)

Si  $(X_n)_{n \geq 1}$  est une suite de v.a. de Bernoulli indépendantes de même paramètre  $p$ , alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{Pr} p.$$

**Preuve :** Il suffit d'appliquer la loi faible des grands nombres en notant qu'ici  $\mathbb{E} X_1 = p$ . □

Interprétation : Considérons une suite d'épreuves répétées indépendantes. Pour chaque épreuve la probabilité d'un « succès » est  $p$ . Notons  $X_i$  l'indicatrice de l'événement *succès* à la  $i$ -ème épreuve. Alors :

$S_n = \sum_{i=1}^n X_i$  est le nombre de succès en  $n$  épreuves et  $M_n = n^{-1} S_n$  est la *fréquence* des succès au cours des  $n$  premières épreuves. Remarquons que pour tout  $\omega$ ,  $0 \leq M_n(\omega) \leq 1$ .

### A.3 Estimation d'une proportion inconnue

On se propose d'estimer le paramètre  $p$  inconnu d'une loi de Bernoulli à partir des observations  $X_i(\omega)$ ,  $1 \leq i \leq n$ , les  $X_i$  étant des v.a. de Bernoulli indépendantes de même paramètre  $p$ .

**Exemple 1** On a une urne comportant des boules rouges en proportion inconnue  $p$  et des boules vertes (en proportion  $q = 1 - p$ ). On effectue  $n$  tirages d'une boule avec remise. Notons :

$$X_i = \mathbf{1}_{\{\text{rouge au } i\text{-ème tirage}\}}$$

et comme ci-dessus désignons par  $M_n$  la moyenne arithmétique des  $X_i$  ou fréquence d'apparition du rouge au cours des  $n$  premiers tirages. D'après la loi faible des grands nombres pour les fréquences,  $M_n$  converge en probabilité vers  $p$ . Comme on s'attend à ce que  $M_n$  soit proche de  $p$  pour les grandes valeurs de  $n$ , il est naturel d'estimer  $p$  par  $M_n$ . En fait on observe *une* valeur particulière  $M_n(\omega)$  calculée à partir des résultats des  $n$  tirages *réellement* effectués. La question pratique qui se pose est de donner une « fourchette » pour l'approximation de  $p$  par la valeur observée  $M_n(\omega)$ . L'inégalité de Tchebycheff (A.3) pour  $M_n$  s'écrit ici :

$$P(|M_n - p| \geq t) \leq \frac{\text{Var } X_1}{nt^2} = \frac{p(1-p)}{nt^2}. \quad (\text{A.4})$$

Comme  $p$  est inconnu, on ne peut pas utiliser directement ce majorant. On remplace alors  $p(1-p)$  par :

$$\sup_{x \in [0,1]} x(1-x) = \frac{1}{4}$$

(la parabole d'équation  $y = x(1-x)$  a sa concavité tournée vers les  $y$  négatifs, les deux zéros du trinôme sont  $x_1 = 0$  et  $x_2 = 1$ ; par symétrie, le sommet a pour abscisse  $(x_1 + x_2)/2 = 1/2$  et pour ordonnée  $1/2(1 - 1/2) = 1/4$ ). En reportant dans (A.4), on obtient *quelle que soit la valeur inconnue*  $p$  :

$$P(|M_n - p| \geq t) \leq \frac{\text{Var } X_1}{nt^2} = \frac{1}{4nt^2} \quad (\text{A.5})$$

d'où en passant à l'événement complémentaire :

$$P(M_n - t < p < M_n + t) \geq 1 - \frac{1}{4nt^2}. \quad (\text{A.6})$$

En pratique on remplace  $M_n$  par la valeur réellement observée  $M_n(\omega)$  et on dit que  $I = ]M_n(\omega) - t, M_n(\omega) + t[$  est un intervalle de confiance (ou fourchette) pour  $p$ . Le deuxième membre de (A.5) peut s'interpréter comme un majorant de la probabilité de se tromper lorsque l'on déclare que  $p$  est dans  $I$ . On dit aussi que  $I$  est un intervalle de confiance au niveau  $\alpha \geq 1 - 1/(4nt^2)$ .  $\square$

**Exemple 2 (Sondage)** Avant le second tour d'une élection présidentielle opposant les candidats  $A$  et  $B$ , un institut de sondage interroge au hasard 1 000 personnes dans la rue<sup>4</sup>. On note  $p$  la proportion d'électeurs décidés à voter pour  $A$  dans la population totale. Dans l'échantillon sondé, cette proportion est égale à 0.54. Proposer un intervalle de confiance pour  $p$  au niveau 0.95.

Le sondage peut être assimilé à un tirage avec remise (en admettant qu'une personne interrogée plusieurs fois accepte de répondre à chaque fois) et on est ramené à la situation de l'exemple précédent. Ici la fréquence *observée* réellement est  $M_n(\omega) = 0.54$  et l'inégalité (A.6) nous dit que l'on peut prendre comme intervalle de confiance :

$$I = ]0.54 - t, 0.54 + t[ \quad \text{avec un niveau} \quad \alpha \geq 1 - \frac{1}{4nt^2}.$$

Comme on souhaite que  $\alpha$  soit au moins égal à 0.95, il suffit de choisir la plus petite valeur de  $t$  telle que :

$$1 - \frac{1}{4000t^2} \geq 0.95 \Leftrightarrow t \geq \frac{1}{10\sqrt{2}} \simeq 0.0707.$$

En prenant  $t = 0.071$ , on obtient :  $I = ]0.469, 0.611[$ . On remarque qu'une partie de cet intervalle correspond à  $p < 1/2$ . Ainsi, bien que le sondage donne 54% d'intentions de vote en faveur de  $A$ , l'inégalité (A.6) ne nous permet pas de pronostiquer sa victoire avec une probabilité d'erreur inférieure à 5%.  $\square$

**Exemple 3 (Sondage, suite)** L'institut de sondage désire présenter à ses clients une fourchette à  $\pm 1\%$  avec un niveau de confiance égal au moins à 0.95%. Combien de personnes doit-il interroger ?

On repart de (A.6). Cette fois on impose  $t = 0.01$  et on cherche  $n$  *minimal* tel que :

$$\frac{1}{4n \times 0.01^2} \leq 0.05$$

On trouve  $n = 50\,000$ , ce qui donne au sondage un coût prohibitif<sup>5</sup>. Nous reviendrons sur ce problème au chapitre suivant.  $\square$

## A.4 Convergence presque sûre des fréquences

On peut représenter graphiquement la suite  $M_n(\omega)$  des fréquences de succès dans une suite d'épreuves de Bernoulli par la ligne brisée dont les sommets ont pour coordonnées

---

4. Ceci est une simplification volontaire permettant d'assimiler la situation à un tirage avec remise : une même personne peut ainsi être interrogée plusieurs fois au cours du sondage. En pratique les méthodes utilisées par les instituts de sondage pour sélectionner un échantillon sont un peu plus compliquées. . .

5. Les sondages ordinaires sont faits sur des échantillons de 500 ou 1 000 personnes. Pour les élections présidentielles, les instituts interrogent des échantillons de 5 000 personnes. La petite étude ci-dessus montre que pour gagner une décimale sur la précision du sondage (i.e. diviser par 10 la longueur de l'intervalle de confiance), il faut multiplier la taille de l'échantillon et donc le coût du sondage par 100. . .



$(n, M_n(\omega))$ . A chaque  $\omega$  correspond ainsi une ligne brisée infinie que nous appellerons *trajectoire*. La loi faible des grands nombres nous donne le comportement asymptotique de ces trajectoires *dans leur ensemble*. Elle signifie grosso modo que pour  $n$  grand fixé ( $n \geq n_0(\varepsilon)$ ) la plupart des trajectoires vont traverser le segment vertical d'extrémités  $(n, p - \varepsilon)$  et  $(n, p + \varepsilon)$ . Elle ne nous dit rien sur le comportement *individuel* de chaque trajectoire. Une trajectoire qui traverse  $]p - \varepsilon, p + \varepsilon[$  à la verticale de  $n$  peut très bien sortir de la bande horizontale engendrée par ce segment au delà de  $n$ . Une question naturelle est alors : existe-t-il des trajectoires qui à partir d'un certain rang  $n_0 = n_0(\omega, \varepsilon)$  restent dans la bande  $\{(x, y) \in \mathbb{R}^2, x \geq n_0 \text{ et } p - \varepsilon < y < p + \varepsilon\}$ ? Nous allons montrer que l'ensemble des trajectoires qui vérifient cette propriété pour tout  $\varepsilon > 0$  a pour probabilité 1, autrement dit que  $M_n$  converge presque sûrement vers  $p$ .

**Théorème 23 (Loi forte des g. n. pour les fréquences)**

Si  $(X_n)_{n \geq 1}$  est une suite de v.a. de Bernoulli indépendantes de même paramètre  $p$ , alors :

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{p.s.} p.$$

**Preuve :** Comme précédemment, nous notons :

$$S_n = \sum_{i=1}^n X_i \quad \text{et} \quad M_n = \frac{S_n}{n}.$$

Les deux ingrédients principaux de la démonstration sont :

- L'écriture de l'événement  $\{M_n \text{ converge vers } p\}$  à l'aide d'opérations ensemblistes dénombrables sur les événements  $\{|M_n - p| \geq \varepsilon\}$  dont on sait majorer les probabilités.
- L'obtention d'une vitesse de convergence vers 0 de ces mêmes probabilités suffisante pour que :

$$\sum_{n=1}^{+\infty} P(|M_n - p| \geq \varepsilon) < +\infty. \tag{A.7}$$

Remarquons que l'inégalité de Tchebycheff est ici trop faible puisqu'elle nous donne seulement une vitesse en  $O(n^{-1})$ . En fait, on peut obtenir une vitesse de convergence exponentielle grâce à l'inégalité suivante :

$$P(|M_n - p| \geq \varepsilon) \leq 2 \exp(-2n\varepsilon^2). \tag{A.8}$$

Nous admettons provisoirement cette inégalité dont une preuve est proposée en exercice <sup>6</sup>. A partir de maintenant, la démonstration se développe en 7 « pas » élémentaires.

1<sup>er</sup> pas : On rappelle la traduction automatique des quantificateurs. Si  $I$  est un ensemble quelconque d'indices,  $(\mathcal{P}_i)$  une propriété dépendant de l'indice  $i$  et  $A_i$  l'ensemble des

---

6. Dans le polycopié de Deug. Pour le présent document, voir théorème 7.

$\omega \in \Omega$  vérifiant  $(\mathcal{P}_i)$ , on a :

$$\begin{aligned} \{\omega \in \Omega, \forall i \in I, \omega \text{ vérifie } (\mathcal{P}_i)\} &= \bigcap_{i \in I} A_i \\ \{\omega \in \Omega, \exists i = i(\omega) \in I, \omega \text{ vérifie } (\mathcal{P}_i)\} &= \bigcup_{i \in I} A_i \end{aligned}$$

Ainsi le quantificateur  $\forall$  peut toujours se traduire par une intersection et le quantificateur  $\exists$  par une réunion.

2<sup>e</sup> pas : Considérons l'ensemble :

$$C = \{\omega \in \Omega, \lim_{n \rightarrow +\infty} M_n(\omega) = p\}.$$

On peut exprimer  $C$  à l'aide des événements  $\{|M_n - p| < \varepsilon\}$  en écrivant la définition de la limite :

$$\omega \in C \Leftrightarrow \forall \varepsilon > 0, \exists k = k(\omega, \varepsilon), \forall n \geq k, |M_n(\omega) - p| < \varepsilon, \quad (\text{A.9})$$

et en appliquant la règle de traduction automatique des quantificateurs :

$$C = \bigcap_{\varepsilon > 0} \bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} \{|M_n - p| < \varepsilon\}.$$

L'inconvénient de cette décomposition est que le «  $\varepsilon > 0$  » dans la première intersection est une indexation par l'ensemble  $I = ]0, +\infty[$  qui n'est pas dénombrable. On ne peut donc pas appliquer les propriétés de  $\sigma$ -additivité ou de continuité monotone séquentielle à ce stade.

3<sup>e</sup> pas : Il est facile de remédier à cet inconvénient : il suffit de *discrétiser* le  $\varepsilon$  dans la définition de la limite. On sait qu'on obtient une définition équivalente remplaçant dans (A.9) le «  $\forall \varepsilon > 0$  » par «  $\forall \varepsilon_j$  » où  $(\varepsilon_j)_{j \in \mathbb{N}}$  est une suite strictement décroissante de réels tendant vers 0. On peut choisir par exemple  $\varepsilon_j = 10^{-j}$ . En appliquant à nouveau la traduction des quantificateurs, nous obtenons :

$$C = \bigcap_{j \in \mathbb{N}} \bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} \{|M_n - p| < \varepsilon_j\}.$$

Remarquons au passage que, sous cette forme, il est clair que l'ensemble  $C$  est en fait un événement, c'est-à-dire un membre de la famille  $\mathcal{F}$  de parties de  $\Omega$  sur laquelle est définie la fonction d'ensemble  $P$ . En effet,  $M_n$  étant une variable aléatoire, les  $\{|M_n - p| < \varepsilon_j\}$  sont des événements et  $C$  s'obtient par des opérations ensemblistes dénombrables sur ces événements. Il est donc légitime de parler de la probabilité de  $C$ . Nous allons montrer que  $P(C) = 1$ .

4<sup>e</sup> pas : Nous venons de passer d'une infinité non dénombrable de  $\varepsilon$  à une suite  $(\varepsilon_j)$ . Le lemme suivant va nous permettre de travailler avec une seule valeur de  $\varepsilon$ .

**Lemme 24** *Si  $(A_j)_{j \in \mathbb{N}}$  est une suite d'événements ayant chacun une probabilité 1, alors leur intersection a aussi une probabilité 1.*

**Preuve :** Par passage au complémentaire, il suffit de prouver que la réunion des  $A_j^c$  a une probabilité nulle. Or :

$$0 \leq P\left(\bigcup_{j \in \mathbb{N}} A_j^c\right) \leq \sum_{j \in \mathbb{N}} P(A_j^c) = 0,$$

puisque chaque  $P(A_j^c)$  est nul par hypothèse.  $\square$

Si l'on prouve que pour chaque  $\varepsilon > 0$  fixé,  $P(C_\varepsilon) = 1$  où

$$C_\varepsilon = \bigcup_{k \in \mathbb{N}} \bigcap_{n \geq k} \{|M_n - p| < \varepsilon\},$$

il suffira d'appliquer le lemme avec  $A_j = C_{\varepsilon_j}$  pour obtenir  $P(C) = 1$ .

5<sup>e</sup> pas : Soit donc  $\varepsilon > 0$  fixé. Pour montrer que  $C_\varepsilon$  a une probabilité 1, on travaille sur son complémentaire que nous noterons  $B$ .

$$B = \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} \{|M_n - p| \geq \varepsilon\}.$$

On a :

$$B = \bigcap_{k \in \mathbb{N}} B_k \quad \text{avec} \quad B_k = \bigcup_{n \geq k} \{|M_n - p| \geq \varepsilon\}.$$

Donc  $B$  est inclus dans chaque  $B_k$ , d'où :

$$\forall k \in \mathbb{N}, \quad 0 \leq P(B) \leq P(B_k). \quad (\text{A.10})$$

6<sup>e</sup> pas : On majore  $P(B_k)$  en utilisant la sous-additivité de  $P$  pour les unions dénombrables :

$$0 \leq P(B_k) = P\left(\bigcup_{n \geq k} \{|M_n - p| \geq \varepsilon\}\right) \leq \sum_{n \geq k} P(|M_n - p| \geq \varepsilon).$$

D'après (A.8), ce majorant est le reste de rang  $k$  d'une série convergente. Il tend donc vers 0 quand  $k$  tend vers  $+\infty$ . Il en est donc de même pour  $P(B_k)$ .

7<sup>e</sup> pas, conclusion : En passant à la limite quand  $k$  tend vers  $+\infty$  dans (A.10), on en déduit  $P(B) = 0$ . En passant à l'événement complémentaire on a donc montré que  $P(C_\varepsilon) = 1$ . Comme la seule hypothèse faite sur  $\varepsilon$  pour obtenir ce résultat était  $\varepsilon > 0$ , on a donc  $P(C_\varepsilon) = 1$  pour tout  $\varepsilon > 0$ . D'après le 4<sup>e</sup> pas ceci entraîne  $P(C) = 1$ , autrement dit :  $M_n$  converge presque sûrement vers  $p$ .  $\square$

Comme sous-produit de la démonstration que nous venons d'achever, nous avons montré au passage que la convergence en probabilité avec une vitesse suffisante implique la convergence presque sûre, plus précisément :

**Théorème 25 (Condition suffisante de convergence p.s.)**

Si  $(Y_n)_{n \geq 1}$  et  $Y$  sont des variables aléatoires vérifiant :

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} P(|Y_n - Y| > \varepsilon) < +\infty, \quad (\text{A.11})$$

alors  $Y_n$  converge presque sûrement vers  $Y$ .

**Preuve :** Il suffit de remplacer  $|M_n - p|$  par  $|Y_n - Y|$  dans la démonstration ci-dessus.  $\square$

## A.5 Discussion

Considérons une urne contenant 10 boules numérotées de 0 à 9. La loi forte des grands nombres pour les fréquences nous dit que si l'on effectue une suite illimitée de tirages avec remise d'une boule, la fréquence d'apparition du chiffre 7 va converger vers  $1/10$  avec probabilité 1. Pour démontrer ce théorème, nous avons admis implicitement l'existence d'un espace probabilisé  $(\Omega, \mathcal{F}, P)$  modélisant cette expérience (suite infinie de tirages avec remise). La construction mathématique rigoureuse d'un tel modèle présente une réelle difficulté qui est au coeur de la *théorie de la mesure* et relève du programme de la licence de mathématiques. Nous nous contenterons de quelques considérations élémentaires<sup>7</sup> sur cet espace probabilisé, utiles pour notre exploration de la loi forte des grands nombres.

L'espace  $\Omega$  doit être assez « riche » pour « supporter » une suite infinie  $(Y_i)_{i \geq 1}$  de v. a. indépendantes et de même loi uniforme sur  $\{0, 1, 2, \dots, 8, 9\}$ . La variable aléatoire  $Y_i$  s'interprète comme le numéro obtenu lors du  $i$ -ième tirage. On pose alors  $X_i = \mathbf{1}_{\{Y_i=7\}}$  et  $M_n = n^{-1} \sum_{i=1}^n X_i$  est la fréquence d'apparition du 7 en  $n$  tirages.

Nous allons examiner deux choix possibles pour  $\Omega$ . Le premier et le plus naturel est de prendre :

$$\Omega = \{0, 1, 2, \dots, 8, 9\}^{\mathbb{N}^*}.$$

Autrement dit un élément quelconque  $\omega$  de  $\Omega$  est une suite  $(c_i)_{i \geq 1}$  de chiffres décimaux. Le choix de la famille  $\mathcal{F}$  d'événements observables est plus délicat. On ne peut pas prendre l'ensemble de toutes les parties de  $\Omega$  car on ne pourrait pas attribuer une probabilité à chacune de ces parties de façon compatible avec ce que l'on sait déjà sur les tirages finis. Il est clair que  $\mathcal{F}$  doit contenir les événements dont la réalisation ne dépend que d'un nombre fini de tirages (c'est bien le cas des événements du type  $\{|M_n - p| > \varepsilon\}$  auxquels on sait attribuer une probabilité (au moins théoriquement puisque l'on sait écrire une formule donnant  $P(n(p - \varepsilon) < S_n < n(p + \varepsilon))$  à l'aide de la loi binomiale). On prend pour  $\mathcal{F}$  la plus petite famille d'événements observables<sup>8</sup> parmi celles qui contiennent les événements dont la réalisation ne dépend que d'un nombre fini d'épreuves. Pour définir la fonction d'ensemble  $P$  sur  $\mathcal{F}$ , on utilise un théorème de prolongement de la théorie de la mesure. On peut alors voir qu'avec ce modèle, *chaque événement élémentaire  $\omega$  doit avoir une probabilité nulle*. En effet, fixons  $\omega_0 = (u_1, u_2, \dots, u_n, \dots) \in \Omega$ . On a :

$$\forall n \geq 1, \quad \{\omega_0\} \subset \{Y_1 = u_1\} \cap \{Y_2 = u_2\} \cap \dots \cap \{Y_n = u_n\},$$

d'où

$$P(\{\omega_0\}) \leq P\left(\bigcap_{i=1}^n \{Y_i = u_i\}\right) = \prod_{i=1}^n P(Y_i = u_i) = \left(\frac{1}{10}\right)^n,$$

---

7. Tout est relatif. . .

8. i.e. la plus petite tribu.

en utilisant la nécessaire indépendance des  $Y_i$ . Ainsi :

$$\forall n \geq 1, \quad 0 \leq P(\{\omega_0\}) \leq 10^{-n}.$$

En faisant tendre  $n$  vers l'infini on en déduit  $P(\{\omega_0\}) = 0$ . Ce raisonnement est valable pour tout  $\omega_0$  de  $\Omega$ .

Notons que la nullité de  $P(\{\omega\})$  pour tout  $\omega \in \Omega$  ne contredit pas l'égalité  $P(\Omega) = 1$ . En effet on n'a pas le droit d'écrire ici «  $P(\Omega) = \sum_{\omega \in \Omega} P(\{\omega\})$  » car l'ensemble d'indexation  $\Omega$  n'est pas dénombrable (il est en bijection avec l'intervalle  $[0, 1]$  de  $\mathbb{R}$ ).

Si  $E$  est un événement dénombrable, les événements élémentaires qui le composent peuvent être indexés par  $\mathbb{N}$  :  $E = \{\omega_0, \omega_1, \dots, \omega_n, \dots\}$  et  $P(E) = \sum_{n \in \mathbb{N}} P(\{\omega_n\}) = 0$ . Ceci est valable a fortiori pour les événements finis.

Donc si un événement a une probabilité non nulle dans ce modèle, il est nécessairement composé d'une infinité non dénombrable d'événements élémentaires. La réciproque est fautive. Considérons en effet l'événement  $B$  défini comme l'obtention à chacun des tirages des seuls chiffres 0 ou 1. Dans notre modèle  $B$  est l'ensemble des suites de 0 et de 1, il n'est pas dénombrable (puisqu'en bijection avec  $[0, 1]$ ). Par ailleurs :

$$B = \bigcap_{i \in \mathbb{N}^*} \{Y_i = 0 \text{ ou } 1\}.$$

On a donc pour tout  $n \geq 1$ ,  $B \subset B_n = \bigcap_{i=1}^n \{Y_i = 0 \text{ ou } 1\}$ , d'où

$$\forall n \geq 1, \quad 0 \leq P(B) \leq P(B_n) = \left(\frac{2}{10}\right)^n.$$

En faisant tendre  $n$  vers l'infini, on en déduit  $P(B) = 0$ . Notons d'autre part que si  $\omega \in B$ ,  $\omega$  ne contient aucun « 7 » parmi ses termes donc  $M_n(\omega) = 0$  et  $B$  est inclus dans l'événement  $\{M_n \rightarrow 0\}$  (ce qui prouve d'une autre façon que  $P(B) = 0$  grâce à la loi forte des grands nombres). Ainsi le complémentaire de l'événement de probabilité 1  $\{M_n \rightarrow 1/10\}$  contient l'événement  $B$  et est donc lui même infini non dénombrable.

La situation est même encore plus surprenante : on peut faire converger  $M_n(\omega)$  vers n'importe quel rationnel  $r$  fixé de  $[0, 1]$  et ce, pour tous les  $\omega$  d'un événement  $C_r$  non dénombrable et de probabilité nulle (si  $r \neq 1/10$ ). Voici comment faire. On pose  $r = k/l$ ,  $k \in \mathbb{N}$ ,  $l \in \mathbb{N}^*$  et on définit  $C_r$  comme l'ensemble des suites de la forme :

$$\omega = (\underbrace{7, \dots, 7}_k, \underbrace{u_{k+1}, \dots, u_l}_{l-k}, \underbrace{7, \dots, 7}_k, \underbrace{u_{l+k+1}, \dots, u_{2l}}_{l-k}, \underbrace{7, \dots, 7}_k, \dots)$$

en répétant indéfiniment l'alternance de blocs de  $k$  chiffres 7 consécutifs et des blocs de  $l - k$  chiffres  $u_i$  pouvant prendre seulement les valeurs 0 ou 1. Il est immédiat de vérifier que la fréquence des 7 dans une telle suite converge vers  $k/l$ , donc  $C_r \subset \{M_n \rightarrow r\}$ . Il est aussi clair que  $C_r$  est en bijection avec  $\{0, 1\}^{\mathbb{N}^*}$  (la bijection s'obtient en effaçant les 7 et sa réciproque en intercalant des blocs de  $k$  chiffres 7 consécutifs tous les  $l - k$  chiffres binaires).

En adaptant ce procédé, on peut faire converger  $M_n(\omega)$  vers *n'importe quel réel*  $x$  de  $[0, 1]$  sur un événement  $C_x$  non dénombrable et de probabilité nulle si  $x \neq 1/10$  (exercice).

On peut aussi construire des événements non dénombrables et de probabilité nulle sur lesquels  $M_n$  ne converge vers aucune limite. A titre d'exemple voici comment construire un événement  $E$  tel que  $\forall \omega \in E$  :

$$\liminf_{n \rightarrow +\infty} M_n(\omega) = 0, \quad \text{et} \quad \limsup_{n \rightarrow +\infty} M_n(\omega) = 1. \quad (\text{A.12})$$

Commençons par construire une suite particulière  $\omega_0 = (c_i)_{i \geq 1}$  vérifiant (A.12) :

$$\omega_0 = ( \underbrace{7, 7}_2, \underbrace{8, 8, 8, 8}_{2^2}, \underbrace{7, \dots, 7}_{6^2}, \underbrace{8, \dots, 8}_{42^2}, \underbrace{7, \dots, 7}_{(42+42^2)^2}, \dots ).$$

et ainsi de suite en alternant indéfiniment des bloc de 7 consécutifs et de 8 consécutifs. La longueur de chaque bloc est le carré de la somme des longueurs de tous les blocs précédents. Avec cette construction, l'indice du dernier chiffre de chaque bloc est un entier de la forme  $m + m^2$ . A chaque étape, le dernier bloc placé écrase quasiment tout le passé et ainsi  $M_n(\omega_0)$  va osciller indéfiniment entre 0 et 1. Plus précisément, si le bloc considéré se termine par un 8, il contient au moins  $m^2$  chiffres 8 et donc au plus  $m$  chiffres 7 donc  $M_{m^2+m}(\omega_0) \leq m/(m+m^2)$  et ce majorant tend vers 0 quand  $m$  tend vers  $+\infty$ . Si le bloc finit sur un 7, il contient au moins  $m^2$  chiffres 7, donc  $M_{m^2+m}(\omega_0) \geq (m^2)/(m+m^2)$  et ce minorant tend vers 1 quand  $m$  tend vers  $+\infty$ . On a ainsi pu extraire de la suite  $(M_n(\omega_0))_{n \in \mathbb{N}^*}$  une sous suite convergeant vers 0 et une autre convergeant vers 1. Comme  $0 \leq M_n(\omega_0) \leq 1$  pour tout  $n$ , on en déduit que  $\omega_0$  vérifie (A.12).

Pour obtenir une infinité non dénombrable de suites  $\omega$  ayant la même propriété, il suffit de modifier légèrement la construction de  $\omega_0$  en :

$$\omega = ( \underbrace{7, *}_2, \underbrace{8, 8, 8, *}_{2^2}, \underbrace{7, \dots, 7, *}_{6^2}, \underbrace{8, \dots, 8, *}_{42^2}, \underbrace{7, \dots, 7, *}_{(42+42^2)^2}, \dots ).$$

où le dernier chiffre de chaque bloc de  $\omega_0$  est remplacé au choix par un 0 ou un 1 (représenté par l'astérisque ci-dessus).

En travaillant encore un peu, on pourrait de même montrer pour tout couple de réels  $(a, b)$  de  $[0, 1]$  tels que  $a < b$ , l'existence d'événements  $E_{a,b}$  non dénombrables et de probabilité nulle sur lesquels  $M_n$  a pour limite inférieure  $a$  et pour limite supérieure  $b$ . . .

Tous ces exemples montrent que l'événement  $\{M_n \text{ ne converge pas vers } 1/10\}$  a une structure très complexe. Ainsi l'aspect naturel et intuitif de la loi forte des grands nombres pour les fréquences masque un résultat plus profond qu'il n'y paraît. Le *presque sûrement* qui figure dans l'énoncé de cette loi n'est pas une finasserie artificielle de puriste mais est bien inhérent au problème étudié.

On est naturellement tenté d'interpréter les résultats précédents du point de vue de la théorie des nombres en considérant les suites de chiffres décimaux sur lesquelles nous venons de travailler comme des développements décimaux illimités de nombres réels de  $[0, 1]$ . Notre second modèle sera donc  $(\Omega', \mathcal{F}', P')$  où :

$$\Omega' = [0, 1]$$

et  $\mathcal{F}'$  et  $P'$  restent à définir.

Cependant il se présente ici une difficulté qui fait que ce nouveau modèle ne se réduit pas à une traduction automatique du précédent. Si  $(c_i)_{i \geq 1}$  est une suite de chiffres décimaux, la série :

$$\sum_{i=1}^{+\infty} \frac{c_i}{10^i} \quad (\text{A.13})$$

converge et sa somme est un réel  $x$  de  $[0, 1]$  que l'on peut noter

$$x = 0.c_1c_2 \dots c_i \dots = \sum_{i=1}^{+\infty} \frac{c_i}{10^i}.$$

Réciproquement, tout réel de  $[0, 1]$  admet un développement décimal du type (A.13). Ce développement est unique lorsque  $x$  n'est pas un *nombre décimal* (i.e.  $x$  n'est pas de la forme  $k/10^n$ ,  $k \in \mathbb{N}$ ,  $n \in \mathbb{N}^*$ ). Par contre si  $x$  est décimal, il possède deux développements décimaux distincts. Ceci provient de la sommation de série géométrique suivante :

$$\forall n \geq 1, \quad \sum_{i=n}^{+\infty} \frac{9}{10^i} = \frac{9}{10^n} \sum_{j=0}^{+\infty} \frac{1}{10^j} = \frac{9}{10^n} \frac{1}{(1 - \frac{1}{10})} = \frac{1}{10^{n-1}}. \quad (\text{A.14})$$

Cette relation permet de voir que si un développement décimal illimité ne comporte plus que des 9 à partir d'un certain rang  $n$  (le  $(n - 1)$ -ème chiffre n'étant pas un 9), on ne change pas la somme de la série en remplaçant tous ces 9 par des 0 et en augmentant d'une unité le  $(n - 1)$ -ème chiffre. On a ainsi la propagation d'une *retenue* depuis l'infini. Par exemple :

$$0.5972999999 \dots = 0.5973000000 \dots = \frac{5973}{10^4}$$

(il ne s'agit pas d'une égalité approchée, mais d'une égalité rigoureuse, les points de suspension représentant la répétition indéfinie du chiffre 9 ou 0 respectivement). Le développement ne comportant que des 9 à partir d'un certain rang est appelé *développement décimal impropre*, celui ne comportant que des 0 est appelé *développement décimal propre*.

En revenant aux tirages illimités dans notre urne à dix boules, on voit que si l'on choisit  $\Omega' = [0, 1]$ , les deux suites de résultats qui correspondent à un même réel *décimal* seront représentées par le même réel  $\omega$ . Par exemple  $(5, 9, 7, 2, 9, 9, 9, \dots)$  et  $(5, 9, 7, 3, 0, 0, 0, \dots)$  seront représentées par l'événement élémentaire  $\omega = 5973/10000$ .

Pour surmonter cette difficulté, nous « dédoublons » la suite  $(Y_i)_{i \geq 1}$ . Pour tout  $i \geq 1$ , on définit les deux variables aléatoires  $Y_i$  et  $Y'_i$  comme suit. Si  $\omega \in [0, 1]$  n'est pas décimal,  $Y_i(\omega) = Y'_i(\omega)$  est le  $i$ -ème chiffre décimal de l'unique développement décimal de  $\omega$ . Si  $\omega$  est un décimal de  $[0, 1]$ ,  $Y_i(\omega)$  est le  $i$ -ème chiffre de son développement propre,  $Y'_i(\omega)$  le  $i$ -ème chiffre décimal de son développement impropre. On requiert, comme dans le premier modèle que chacune de ces deux suites soit indépendante et que chacune des variables  $Y_i$  et  $Y'_i$  suive la loi uniforme sur  $\{0, 1, \dots, 8, 9\}$ . Ceci permet de montrer que chaque événement élémentaire  $\omega$  doit avoir une probabilité  $P'$  nulle. D'autre part,  $Y_i$

et  $Y'_i$  diffèrent seulement sur l'ensemble  $D$  des décimaux de  $[0, 1]$  qui est dénombrable (vu en exercice), donc de probabilité  $P'$  nulle. Ainsi les deux suites  $(Y_i)_{i \geq 1}$  et  $(Y'_i)_{i \geq 1}$  sont égales  $P'$ -presque sûrement. Il est donc quand même possible d'interpréter la suite illimitée de tirages dans l'urne comme le choix aléatoire d'un réel  $\omega$  de  $[0, 1]$  suivant la loi de probabilité  $P'$ .

On peut maintenant examiner les conséquences de notre cahier des charges (les conditions sur les suites de v.a.  $(Y_i)_{i \geq 1}$  et  $(Y'_i)_{i \geq 1}$ ) sur la construction de  $(\mathcal{F}', P')$ . La condition d'indépendance de la suite  $(Y_i)_{i \geq 1}$  avec même loi uniforme sur  $\{0, 1, \dots, 8, 9\}$  pour tout  $Y_i$  peut s'écrire comme suit. Pour tout  $n \geq 1$ , et tout  $n$ -uplet  $(c_1, \dots, c_n)$  de chiffres décimaux,

$$P'(Y_1 = c_1, Y_2 = c_2, \dots, Y_n = c_n) = \prod_{i=1}^n P'(Y_i = c_i) = \frac{1}{10^n}.$$

En notant que l'on a exclu les développements impropres dans la définition des  $Y_i$ , on a l'équivalence :

$$Y_1(\omega) = c_1, Y_2(\omega) = c_2, \dots, Y_n(\omega) = c_n \Leftrightarrow \omega \in [\alpha_n, \alpha_n + 10^{-n}[,$$

où l'on a posé :  $\alpha_n = c_1 10^{-1} + \dots + c_n 10^{-n}$ . Lorsque le  $n$ -uplet  $(c_1, \dots, c_n)$  prend toutes les valeurs possibles (à  $n$  fixé),  $\alpha_n$  décrit exactement l'ensemble des décimaux pouvant s'écrire sous la forme  $k 10^{-n}$ . La condition sur la suite  $(Y_i)_{i \geq 1}$  peut donc se traduire par :

$$\forall n \geq 1, \forall k = 0, 1, \dots, 10^n - 1, \quad P' \left( \left[ \frac{k}{10^n}, \frac{k+1}{10^n} \right] \right) = \frac{1}{10^n}.$$

L'utilisation de la suite  $(Y'_i)$  à la place de  $(Y_i)$  dans le raisonnement ci-dessus nous aurait donné la même conclusion mais avec des intervalles ouverts à gauche et fermés à droite. Notons que dans les deux cas la probabilité  $P'$  de l'intervalle concerné est égale à sa *longueur*. On peut aussi utiliser chacun de ces deux résultats pour redémontrer que la probabilité d'un événement élémentaire  $\omega$  est forcément nulle. Finalement, grâce à l'additivité de  $P'$  on en déduit facilement que la condition sur la suite  $(Y_i)$  équivaut à :

$$\forall a, b \in [0, 1] \cap D \quad (a < b), \quad P'([a, b]) = b - a \quad (\text{A.15})$$

(ou à chacune des conditions obtenues avec  $[a, b[$ ,  $]a, b]$  ou  $]a, b[$ ). Par continuité monotone de  $P'$ , on en déduit que (A.15) s'étend au cas de réels  $a, b > a$  quelconques de  $[0, 1]$  : il suffit de considérer deux suites de décimaux  $a_n \uparrow a$  et  $b_n \downarrow b$  et de noter que  $[a, b] = \bigcap_{n \geq 1} [a_n, b_n]$  (détails laissés en exercice).

Nous voyons maintenant que le problème de la construction de  $(\mathcal{F}', P')$  est exactement celui de la construction d'une fonction d'ensemble  $\sigma$  - *additive* prolongeant la fonction *longueur d'un intervalle*. Ce problème est celui de la construction de la *mesure* de Lebesgue. On peut le résoudre en prenant pour  $\mathcal{F}'$  la plus petite famille d'événements observables contenant les intervalles. On arrive ainsi à définir la longueur ou mesure de Lebesgue des sous ensembles de  $[0, 1]$  qui sont dans  $\mathcal{F}'$ . Si un tel sous ensemble est de la forme  $B = \cup_{i \geq 1} ]a_i, b_i[$  où les suites  $(a_i)$  et  $(b_i)$  vérifient pour tout  $n$  :



$0 \leq a_n < b_n \leq a_{n+1} < b_{n+1} \leq 1$ , alors  $B$  est une réunion disjointe d'intervalles et sa probabilité  $P'$  ou longueur est évidemment la série de terme général la longueur de  $]a_i, b_i[$ . Malheureusement, tous les éléments de la famille  $\mathcal{F}'$  sont loin d'avoir une structure aussi simple et le calcul explicite de leur longueur n'est pas toujours possible (on sait qu'elle existe et on connaît ses propriétés). Nous connaissons déjà un exemple d'élément de  $\mathcal{F}'$  qui ne peut pas s'écrire comme réunion dénombrable d'intervalles disjoints, c'est l'événement  $C_7 = \{\text{convergence de la fréquence du chiffre 7 vers } 1/10\}$ . En effet par densité des décimaux, tout intervalle contient au moins un décimal (en fait une infinité) et si  $\omega$  est décimal,  $Y_i(\omega) = 0$  à partir d'un certain rang (de même  $Y_i'(\omega) = 9$ ) par conséquent  $M_n(\omega)$  converge vers 0 donc  $\omega \notin C_7$ . Ainsi  $C_7$  ne peut s'écrire comme réunion dénombrable d'intervalles disjoints. Nous savons pourtant calculer sa *longueur* par la loi forte des grands nombres : elle vaut 1.

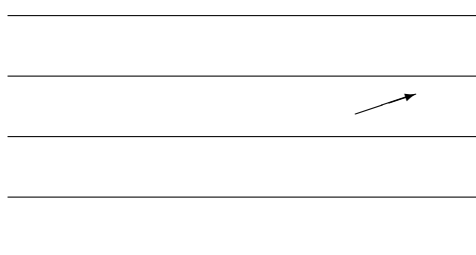
Dans toute cette section nous nous sommes intéressés à la fréquence d'apparition du 7. Bien sûr ce chiffre n'a été choisi que pour fixer les idées et n'importe quel autre chiffre décimal aurait tout aussi bien fait l'affaire. Pour généraliser un peu définissons  $M_{n,j}$  comme la fréquence d'apparition du chiffre  $j$  ( $j \in \{0, 1, \dots, 8, 9\}$ ) au cours des  $n$  premiers tirages. Notons de même  $C_j$  l'événement  $\{M_{n,j} \text{ converge vers } 1/10\}$ . Par la loi forte des grands nombres, chaque  $C_j$  a une *longueur* (i.e. une probabilité  $P'$ ) égale à 1. Par le lemme 24, l'intersection de ces dix ensembles a aussi une longueur 1.

Convenons d'appeler *nombre normal* tout réel de  $[0, 1]$  tel que la fréquence de chacun des 10 chiffres décimaux  $0, 1, \dots, 9$  dans le développement décimal illimité de ce nombre converge vers  $1/10$ . Nous avons ainsi obtenu un résultat de théorie des nombres qui s'énonce ainsi : *l'ensemble de tous les nombres normaux de  $[0, 1]$  a pour longueur 1* (on dit aussi *presque tout* nombre de  $[0, 1]$  est normal). Ce résultat est dû à Borel. On pourrait maintenant traduire tous les exemples étudiés dans le cadre du premier modèle et voir ainsi que l'ensemble de longueur nulle des nombres non normaux a une structure très complexe. Là encore, le théorème de Borel est plus profond qu'il n'y paraît à première vue...

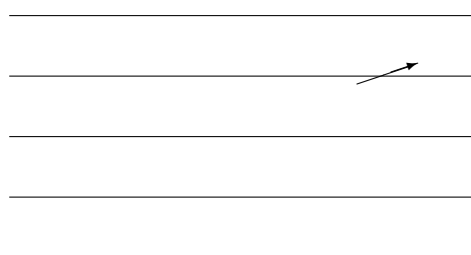
# Annexe B

## L'aiguille de Buffon\*

Dans cette expérience inventée par Buffon (1777) on trace sur une surface plane horizontale des droites parallèles équidistantes, séparées par une distance  $a$  (on peut par exemple utiliser les rainures d'un parquet). On laisse tomber sur cette surface une aiguille de longueur  $\ell \leq a$  et une fois l'aiguille immobilisée, on observe si elle coupe l'une des droites du réseau. On répète l'expérience en notant la fréquence des intersections. Lorsque le nombre d'expériences augmente indéfiniment, cette fréquence converge selon Buffon vers  $p = \frac{2\ell}{\pi a}$  permettant ainsi d'obtenir une estimation expérimentale du nombre  $\pi$ .



« Echec »



« Succès »

Le document de la page 36 représente les résultats de 1200 lancers réalisés avec une allumette et un réseau tracé sur une feuille de format A4. On a ici  $\ell = a = 4,5$  cm et  $p = \frac{2}{\pi} \approx 0,637$ .

Cherchons une modélisation de cette expérience. On note  $Y$  la distance du milieu de l'aiguille à la droite la plus proche.  $Y$  prend ses valeurs dans  $[0, \frac{a}{2}]$ . On note  $\Phi$  une mesure de l'angle entre les droites (toutes orientées dans le même sens) et l'aiguille orientée du chas vers la pointe.  $\Phi$  prend ses valeurs dans  $[0, 2\pi]$  (par exemple)<sup>1</sup>.

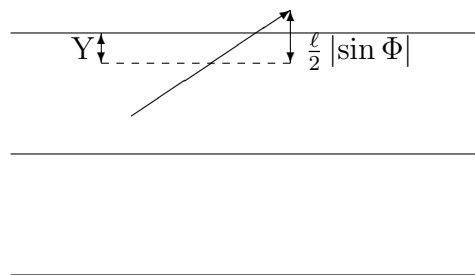
---

\*. Extrait de Mathématiques pour l'Enseignement Secondaire (M.E.S. 1), Probabilités, option de Maîtrise, Ch. Suquet, Lille 1992.

1. On pourrait aussi utiliser les angles de droites,  $\Phi$  serait alors à valeurs dans un intervalle de longueur  $\pi$ .

$Y$  et  $\Phi$  sont des variables aléatoires. La connaissance du couple  $(Y(\omega), \Phi(\omega))$  suffit pour savoir s'il y a ou non intersection. En effet en notant  $E$  l'événement ■ l'aiguille coupe l'une des droites du réseau■, on a :

$$E = \left\{ Y \leq \frac{\ell}{2} |\sin \Phi| \right\}$$



Nous ferons les hypothèses suivantes sur les variables aléatoires  $Y$  et  $\Phi$  :

( $H_1$ )  $Y$  suit la loi uniforme sur  $[0, \frac{a}{2}]$ .

( $H_2$ )  $\Phi$  suit la loi uniforme sur  $[0, 2\pi]$ .

( $H_3$ )  $Y$  et  $\Phi$  sont indépendantes.

Compte tenu de ces trois hypothèses, la loi du couple  $(\Phi, Y)$  est la loi uniforme sur le rectangle  $[0, 2\pi] \times [0, \frac{a}{2}]$ .

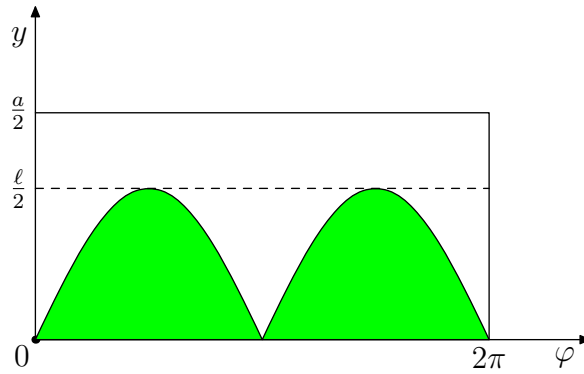
Remarquons<sup>2</sup> que nous n'avons pas précisé  $(\Omega, \mathcal{A}, \mathbf{P})$ . Si on souhaite considérer chaque position précise de l'aiguille comme un événement élémentaire, on peut prendre  $\Omega = \mathbf{R}^2 \times [0, 2\pi]$  où  $\omega = ((u, v), \varphi)$  représente la position de l'aiguille lorsque son centre est le point de coordonnées  $(u, v)$  et qu'elle forme un angle  $\varphi$  avec les droites orientées du réseau. La tribu  $\mathcal{A}$  associée peut être choisie de la manière suivante. On note  $g$  l'application de  $\Omega$  dans  $[0, 2\pi] \times [0, \frac{a}{2}]$  définie par  $g(\omega) = (\varphi, y)$  où  $y$  est la distance du point  $(u, v)$  à la droite la plus proche du réseau. On note  $\mathcal{B}$  la tribu borélienne de  $[0, 2\pi] \times [0, \frac{a}{2}]$ . Pour que  $E$  soit bien un événement dans ce modèle, il suffit qu'il soit un élément de  $\mathcal{A}$ . Il suffit pour cela de prendre  $\mathcal{A} = g^{-1}(\mathcal{B})$ . Autrement dit  $\mathcal{A}$  est la tribu des événements qui ne dépendent que de  $Y$  et  $\Phi$ . Si  $\mathbf{P}$  est une probabilité sur cette tribu, on a par le théorème de transfert en notant  $A$  un élément quelconque de  $\mathcal{A}$  et  $A' = g(A)$  :

$$\mathbf{P}(A) = \int_{\Omega} \mathbf{1}_A d\mathbf{P} = \int_{\mathbf{R}^2} \mathbf{1}_{A'} dP_{(\Phi, Y)} = \int_{[0, 2\pi] \times [0, \frac{a}{2}]} \mathbf{1}_{A'}(\varphi, y) dP_{(\Phi, Y)}(\varphi, y).$$

On voit ainsi que si  $P_{(\Phi, Y)}$  est la loi uniforme sur le rectangle  $[0, 2\pi] \times [0, \frac{a}{2}]$ , alors en définissant  $\mathbf{P}$  par la formule de transfert ci-dessus,  $(\Omega, \mathcal{A}, \mathbf{P})$  vérifie bien les hypothèses ( $H_1$ ), ( $H_2$ ), ( $H_3$ ). On pourrait faire la même construction avec tout  $\Omega$  suffisamment riche pour décrire tous les résultats possibles de l'expérience.

Finalement, tout revient pour le calcul de  $\mathbf{P}(E)$  à remplacer l'espace  $\Omega$  par  $\Omega' = [0, 2\pi] \times [0, a/2]$  et  $\mathbf{P}$  par la loi uniforme sur  $\Omega'$  et  $E$  par l'ensemble  $E' = \{(\varphi, y) \in \Omega', y \leq \frac{\ell}{2} |\sin \varphi|\}$ .

2. Ce paragraphe peut être sauté en première lecture.



On obtient ainsi :

$$\mathbf{P}(E) = \frac{\lambda_2(E')}{\lambda_2(\Omega')} = \frac{1}{2\pi \frac{a}{2}} \int_0^{2\pi} \frac{\ell}{2} |\sin \varphi| \, d\varphi = \frac{\ell}{\pi a} \int_0^\pi \sin \varphi \, d\varphi = \frac{2}{\pi} \left( \frac{\ell}{a} \right).$$

Remarquons que le choix *a priori* des hypothèses  $(H_1)$ ,  $(H_2)$ ,  $(H_3)$  ne peut guère être guidé que par des considérations du genre : « on ne voit pas pourquoi certaines valeurs de  $\Phi$  ou  $Y$  devraient être avantagées, on ne voit pas pourquoi il devrait y avoir un lien entre  $Y$  et  $\Phi \dots$  ». Ou plus cyniquement : « Ces hypothèses conduisent à des calculs simples que l'on sait faire » !

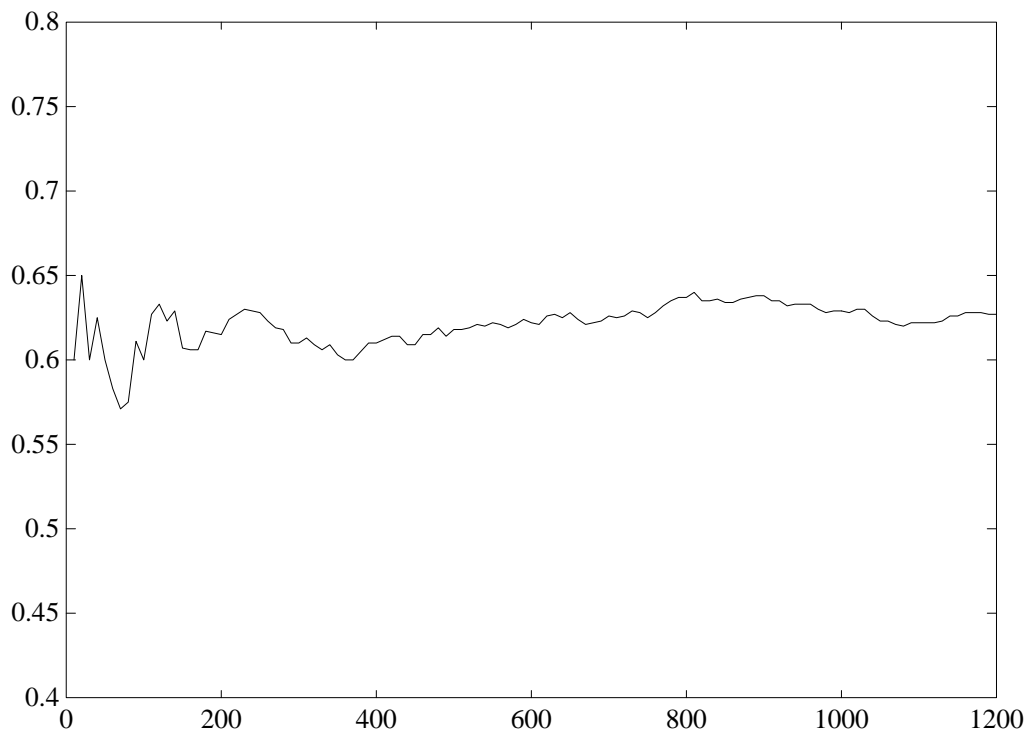
## Résultats de 1200 lancers

0111110001	6	6	1111111011	9	69	1011111110	8	131
0101111011	7	13	1110100111	7	76	1001110111	7	138
0011100110	5	18	0111100010	5	81	1101101110	7	145
1110011101	7	25	1110111100	7	88	1001011110	6	151
1110000011	5	30	0000100101	3	91	0111110001	6	157
0001111001	5	35	0001111011	6	97	1101001001	5	162
1101000101	5	40	1110011001	6	103	1101001100	5	167
1101111000	6	46	1011111110	8	111	1110100101	6	173
1111011111	9	55	1101010101	6	117	0110001100	4	177
1000100111	5	60	1100100111	6	123	1011010110	6	183
1011101011	7	190	0111110011	7	251	1101100110	6	315
1101100010	5	195	0110110111	7	258	1011011101	7	322
1100000111	5	200	1011110010	6	264	1100101111	7	329
0001111111	7	207	0000111001	4	268	1111010010	6	335
0010001101	4	211	0111100101	6	274	1110001111	7	342
0101001011	5	216	1111111101	9	283	0101001111	6	348
0100111011	6	222	1101010101	6	289	1001100101	5	353
0111111101	8	230	0111111110	8	297	1101111001	7	360
1111111010	8	238	0011101000	4	301	1010111111	8	368
1110001011	6	244	0111111101	8	309	0111101000	5	373
1001110011	6	379	1011011100	6	444	0111111110	8	518
1111011111	9	388	1011111010	7	451	0100010010	3	521
0110011111	7	395	1011110111	8	459	0101101011	6	527
1000011101	5	400	0011001111	6	465	0011111011	7	534
1111100111	8	408	1000001101	4	469	1100100101	5	539
1010010001	4	412	1011011111	8	477	1010110110	6	545
0100001110	4	416	1111111111	10	487	1111101101	8	553
1111011010	7	423	1011111101	8	495	1110111101	8	561
1010111110	7	430	1011111101	8	503	1110110110	7	568
1101011111	8	438	1110011101	7	510	0111001101	6	574
1000010110	4	578	1101001001	5	634	1100110011	6	690
0110110110	6	584	1110111111	9	643	1010111101	7	697
1000101100	4	588	0110110110	6	649	1010111110	7	704
1011011101	7	595	1000001000	2	651	1111111111	10	714
1100110110	6	601	1000011000	3	654	0011011110	6	720
1111011010	7	608	0111001011	6	660	0111111110	8	728
0001000110	3	611	1100010100	4	664	1001111101	7	735
0001110001	4	615	0101110110	6	670	0100111101	6	741
1111100111	8	623	1111111001	8	678	0100110101	5	746
0010111110	6	629	0000111111	6	684	1110101010	6	752

Tableau des fréquences observées

10k	0	10	20	30	40	50	60	70	80	90
		0,600	0,650	0,600	0,625	0,600	0,583	0,571	0,575	0,611
100	0,600	0,627	0,633	0,623	0,629	0,607	0,606	0,606	0,617	0,616
200	0,615	0,624	0,627	0,630	0,629	0,628	0,623	0,619	0,618	0,610
300	0,610	0,613	0,609	0,606	0,609	0,603	0,600	0,600	0,605	0,610
400	0,610	0,612	0,614	0,614	0,609	0,609	0,615	0,615	0,619	0,614
500	0,618	0,618	0,619	0,621	0,620	0,622	0,621	0,619	0,621	0,624
600	0,622	0,621	0,626	0,627	0,625	0,628	0,624	0,621	0,622	0,623
700	0,626	0,625	0,626	0,629	0,628	0,625	0,628	0,632	0,635	0,637
800	0,637	0,640	0,635	0,635	0,636	0,634	0,634	0,636	0,637	0,638
900	0,638	0,635	0,635	0,632	0,633	0,633	0,633	0,630	0,628	0,629
1000	0,629	0,628	0,630	0,630	0,626	0,623	0,623	0,621	0,620	0,622
1100	0,622	0,622	0,622	0,623	0,626	0,626	0,628	0,628	0,628	0,627
1200	0,627									

Courbe des fréquences observées





# Bibliographie

- [1] P. BARBE et M. LEDOUX. *Probabilité*. Espaces 34, Belin, 1998.
- [2] P. BILLINGSLEY. *Probability and measure*. Wiley, third edition 1995.
- [3] E. BOREL. *Probabilité et certitude*. Que sais-je? N° 445 P.U.F.
- [4] D. BOSQ et J.-P. LECOUTRE. *Théorie de l'Estimation Fonctionnelle*. Collection « Économie et Statistiques Avancées », Economica, 1987.
- [5] W. FELLER. *An Introduction to Probability Theory and its Applications*, Vol. I. Wiley.
- [6] W. FELLER. *An Introduction to Probability Theory and its Applications*, Vol. II. Wiley.
- [7] D. FOATA et A. FUCHS. *Calcul des Probabilités*. Dunod, 1998.
- [8] J.-Y. OUVRARD. *Probabilités tome 2, Maîtrise–Agrégation*. Cassini, 2000.
- [9] D. REVUZ. *Probabilités*. Hermann, 1997.
- [10] W. F. STOUT. *Almost sure convergence*. Academic Press, 1974.
- [11] P. S. TOULOUSE. *Thèmes de Probabilités et Statistique*. Dunod, 1999.
- [12] WILLIAMS *Probability with martingales*. Cambridge University Press, Cambridge (1991).