Le problème des moments et le théorème de Berry–Esseen

Mémoire de stage de recherche

Jérémy BETTINGER

École Normale Supérieure de Rennes

Sous la direction de Nicolas Juillet

IRIMAS, Université de Haute-Alsace,
Laboratoire de recherche en mathématiques,
Analyse et Probabilités

Introduction

Le problème des moments

Soient I un intervalle fermé de $\mathbb{R}$ et $(m_n) \in \mathbb{R}^{\mathbb{N}}$ telle que $m_0 = 1$. Résoudre le problème des moments sur I consiste à trouver des conditions d'existence et d'unicité d'une mesure borélienne positive μ portée par I telle que :

$$\forall n \in \mathbb{N}, \qquad m_n = \int x^n \, \mathrm{d}\mu(x).$$

La normalisation $m_0 = 1$ impose à μ d'être une mesure de probabilité sur I.

Ainsi, cela revient à chercher des conditions d'existence, et d'unicité en loi, d'une variable aléatoire X presque sûrement à valeurs dans I, telle que pour tout n dans $\mathbb{N}$, $m_n = \mathbb{E}(X^n)$.

L'objectif de cette partie du mémoire est de donner des conditions sous lesquelles la réponse est de savoir sous quelles conditions il y a existence ou unicité d'une telle mesure.

Nous établirons quelques généralités et un critère particulièrement simple pour résoudre le problème de l'unicité, quelque soit l'intervalle I, dans les deux premières sections. Cette section sera notamment riche en exemples de lois connues.

On rencontre principalement trois cas dans lesquels ce problème est résolu. Le cas où $I = [0, 1]$ est le problème des moments de Hausdorff, il sera traité en troisième section. Il suivra, le cas où $I = \mathbb{R}_+$, c'est le problème des moments de Stieltjes, et viendra le problème des moments de Hamburger, celui où $I = \mathbb{R}$.

Nous étudierons enfin dans les deux dernières sections de cette partie la condition intégrale de Krein et celle de Lin. Ces résultats entraînent par exemple que sur $\mathbb{R}$ la gaussienne est caractérisée par ses moments, tandis que sur $\mathbb{R}_+$ l'exponentielle de la gaussienne (loi log-normale) ne l'est pas.

Le théorème de Berry–Esseen

Le théorème central limite énonce que si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne $\mathbb{E}(X)$, de variance $Var(X) = \sigma^2$, alors en notant $S_n := \sum_{k=1}^{n} X_k$, la variable aléatoire $(S_n - n\mathbb{E}(X))/(\sigma\sqrt{n})$ converge en loi vers une loi Normale centrée réduite. En d'autres termes, pour $t \in \mathbb{R}$,

$$\mathbb{P}\left(\frac{S_n - n\mathbb{E}(X)}{\sigma\sqrt{n}} \leqslant t\right) \xrightarrow[n\to+\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2} \, \mathrm{d}x.$$

Le théorème de Berry–Esseen fournit un contrôle sur la vitesse de convergence uniforme des fonctions de répartition vers la fonction de répartition normale dans le théorème central limite.

Le théorème de Berry-Esseen (1942) s'énonce comme tel : Soit $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes identiquement distribuée de moyenne $\mathbb{E}(X)$, de variance $Var(X) = \sigma^2$, et admettant un moment absolu d'ordre 3 $\mathbb{E}(|X|^3) < \infty$. On pose $S_n = X_1 + \cdots + X_n$ pour $n \geqslant 1$. Alors, pour tout $n \geqslant 1$,

$$\boxed{\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - n\mathbb{E}(X)}{\sigma\sqrt{n}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2} \, \mathrm{d}x \right| \leqslant \frac{C \ \mathbb{E}(|X - \mathbb{E}(X)|^3)}{\sigma^3\sqrt{n}}}$$

où C désigne une constante universelle strictement positive.

Le théorème de Berry-Esseen a été découvert indépendamment par les mathématiciens Andrew Berry en 1941 et Carl-Gustav Esséen en 1942, chacun fournissant une constante différente. La recherche de la constante optimale C est encore un sujet de recherche actuel. La constante la plus optimale est, à ce jour, celle découverte par Asof en 2011 et vaut $C = 0,4784$ alors que la valeur initiale du théorème de 1942 était proche de 8.

Nous effectuerons deux démonstrations de ce théorème, une par couplage basée sur une idée de Lindeberg mais qui aboutie à un théorème plus faible avec une majoration en $n^{-1/8}$. Ensuite, nous démontrons le théorème de Berry–Esseen à l'aide d'outils de transformée de Fourier.

Nous terminerons avec un exemple de l'inégalité de Berry–Esseen avec la loi de Bernoulli et nous quantifierons, pour certaines probabilités, la vitesse de convergence à l'aide de l'inégalité de Le Cam.

Remerciements

Je tiens à remercier l'ensemble des membres du laboratoire de mathématiques de Mulhouse pour son accueil chaleureux et sa sympathie. J'y ai passé de très bons moments mathématiques et humains.

Bien sûr, un grand merci à Nicolas Juillet pour le temps consacré à encadrer mon stage, pour les réponses à mes questions, pour les nombreux conseils et pour tous ces bons moments !

Table des matières

Première partie
Le problème des moments

1 Introduction

1.1 Conventions

Soient μ une mesure de probabilité sur $\mathbb{R}$ et n un entier naturel. On définit les moments et les moments absolus de μ respectivement par :

$$m_n = \int_{\mathbb{R}} x^n \, \mathrm{d}\mu(x), \qquad \mu_n = \int_{\mathbb{R}} |x|^n \, \mathrm{d}\mu(x).$$

Nous allons nous poser la question de l'existence et de l'unicité des lois ayant pour moments des réels donnés, c'est le « problème des moments ».

1.2 Exemple fondamental

Introduisons tout d'abord la loi Log-Normale et établissons quelques propriétés sur cette dernière.

Définition-Proposition 1.1 (Loi Log-Normale)**.** *Soit N une variable aléatoire de loi $\mathcal{N}(0,1)$. Alors la variable aléatoire e^N suit la loi Log-Normale et est à densité $x \mapsto f(x) := \mathrm{e}^{-(\ln x)^2/2}/(\sqrt{2\pi}x)\, \mathbb{1}_{x>0}$ par rapport à la mesure de Lebesgue.*

Démonstration. Soit h une fonction de classe $\mathcal{C}^\infty$ et à support compact. Calculons $\mathbb{E}(h(\mathrm{e}^N))$ à l'aide de la formule de transfert et du changement de variable $y = \mathrm{e}^x$ qui est un $\mathcal{C}^1$ difféomorphisme de $\mathbb{R}$ à $\mathbb{R}_+^*$.

$$\begin{aligned}
\mathbb{E}(h(\mathrm{e}^N)) &= \int_{\mathbb{R}} h(\mathrm{e}^x)\, \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} \, \mathrm{d}x \\
&= \int_{\mathbb{R}_+^*} h(y)\, \frac{\mathrm{e}^{-(\ln(y))^2/2}}{\sqrt{2\pi}} \frac{\mathrm{d}y}{y} \\
&= \int_{\mathbb{R}} h(y) \frac{\mathrm{e}^{-(\ln y)^2/2}}{\sqrt{2\pi}y} \, \mathbb{1}_{y>0} \, \mathrm{d}y \\
&= \int_{\mathbb{R}} h(y)\, f(y) \, \mathrm{d}y.
\end{aligned}$$

ce qui prouve que e^N a pour densité f. □

Proposition 1.2. *Les moments d'une variable aléatoire X suivant la Loi-Normale sont donnés par la formule :*

$$\forall n \in \mathbb{N}, \qquad \mathbb{E}(X^n) = \mathrm{e}^{n^2/2}.$$

Démonstration. En exploitant ce qui précède, puis en utilisant une identité remarquable et en effectuant le changement de variable $u = x - n$ on obtient les égalités :

$$\mathbb{E}(X^n) = \int_{\mathbb{R}} y^n\, f(y) \, \mathrm{d}y = \int_{\mathbb{R}} \mathrm{e}^{xn}\, \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}} \, \mathrm{d}x = \mathrm{e}^{n^2/2} \int_{\mathbb{R}} \frac{\mathrm{e}^{-(x-n)^2/2}}{\sqrt{2\pi}} \, \mathrm{d}x = \mathrm{e}^{n^2/2} \int_{\mathbb{R}} \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}} \, \mathrm{d}u = \mathrm{e}^{n^2/2}.$$

□

Exemple 1.3 (Exemple fondamental). *En toute généralité, les moments ne caractérisent pas la loi. La loi Log-Normale fournit un tel exemple.*

Démonstration. Notons f la fonction de densité de la loi Log-Normale. Nous allons montrer que la loi à densité $x \mapsto g(x) := f(x)(1 + \sin(2\pi \ln(x))$ a les mêmes moments que f mais n'est pas égale à f. Montrons alors qu'elles ont les mêmes moments et que g est bien une densité.
Soit $k \in \mathbb{N}$. Effectuons le changement de variable $y = \ln(x)$ pour calculer l'intégrale suivante :

$$\begin{aligned}\int_0^{+\infty} x^k f(x) \sin(2\pi \ln(x))\ \mathrm{d}x &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{ky}\ \mathrm{e}^{-y^2/2} \sin(2\pi y)\ \mathrm{d}y = \frac{\mathrm{e}^{k^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-(y-k)^2/2} \sin(2\pi y)\ \mathrm{d}y \\ &= \frac{\mathrm{e}^{k^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-u^2/2} \sin(2\pi u)\ \mathrm{d}u \quad \text{(par } 2\pi \text{ périodicité de sinus)} \\ &= 0 \quad \text{(par imparité de sinus).}\end{aligned}$$

Ainsi f et g ont les mêmes moments.
De plus, g est bien une densité car elle est positive et d'intégrale égale à celle de f (calcul précédent avec $k = 0$) et est différente de f. Ainsi, les moments ne caractérisent pas la loi. □

2 Condition suffisante d'unicité de Fourier

2.1 Régularité de la fonction caractéristique

Il s'agit de donner un critère simple pour savoir si une loi est caractérisée par ses moments.

Proposition 2.1. *Si* $\mathbb{E}(|X|^n)$ *est fini pour tout* $n \in \mathbb{N}$, *alors la fonction caractéristique de* X *notée* φ *est de classe* $\mathcal{C}^n$, *de dérivée* k*-ième* $\varphi^{(k)}(t) = i^k \mathbb{E}(X^k \mathrm{e}^{itX})$ *pour tout* $t \in \mathbb{R}$. *En particulier,* $\varphi^{(k)}(0) = i^k \mathbb{E}(X^k)$.

Démonstration. On peut obtenir ce résultat en appliquant le théorème de régularité sous le signe intégral. Notons pour $x, t \in \mathbb{R}$: $f(x,t) = \mathrm{e}^{itx}$. On a pour $n \in \mathbb{N}$:

- $\forall x \in \mathbb{R},\ t \mapsto f(x,t) \in \mathcal{C}^n(\mathbb{R})$.
- $\forall t \in \mathbb{R},\ \forall k \leqslant n-1,\ x \mapsto \partial_t^k f(x,t) = (ix)^k \mathrm{e}^{itx} \in L^1(\mathrm{d}\mu)$.
- $\forall x, t \in \mathbb{R},\ |\partial_t^n f(x,t)| = |(ix)^n \mathrm{e}^{itx}| = |x|^n \in L^1(\mathrm{d}\mu)$ par hypothèse (majorant indépendant en t).

Alors $\varphi : t \mapsto \displaystyle\int_{\mathbb{R}} \mathrm{e}^{itx}\ \mathrm{d}\mu(x)$ est définie et de classe $\mathcal{C}^n(\mathbb{R})$. De plus on a l'égalité :

$$\forall k \leqslant n,\ \forall t \in \mathbb{R}, \quad \varphi^{(k)}(t) = \int_{\mathbb{R}} (ix)^k \mathrm{e}^{itx}\ \mathrm{d}\mu(x) = i^k \mathbb{E}(X^k \mathrm{e}^{itX}).$$

□

2.2 Critère des moments de Fourier et conditions suffisantes

Théorème 2.2 (Critère des moments de Fourier). *Si* μ *est une mesure de probabilité sur* $\mathbb{R}$ *ayant tous ses moments absolus* μ_n *finis alors si la quantité* $\overline{\lim_n} (\mu_n)^{1/n}/n$ *est finie,* μ *est déterminée par ses moments.*

Démonstration. Soit X une variable aléatoire admettant pour moments absolus les μ_n.
On va montrer que sa fonction caractéristique φ est analytique.
La proposition 2.1 assure que φ est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$ et qu'on a de plus que $\varphi^{(n)}(0) = i^n m_n$.

Soient $x, t, h \in \mathbb{R}$. Appliquons la formule de Taylor avec reste intégral à la fonction $y \mapsto \mathrm{e}^{ixy}$ qui est de classe $\mathcal{C}^\infty$ entre 0 et h.
Alors :

$$\mathrm{e}^{ixh} = \sum_{p=0}^{n} \frac{(ix)^p}{p!} h^p + \int_0^h \frac{(h-s)^n}{n!} (ix)^{n+1} \mathrm{e}^{isx} \, \mathrm{d}s.$$

Puis en multipliant par e^{itx} on obtient

$$\mathrm{e}^{ix(t+h)} = \sum_{p=0}^{n} \frac{(ix)^p}{p!} \mathrm{e}^{itx} h^p + \int_0^h \frac{(h-s)^n}{n!} (ix)^{n+1} \mathrm{e}^{ix(t+s)} \, \mathrm{d}s$$

et en intégrant sur $\mathbb{R}$ pour la mesure μ :

$$\varphi(t+h) - \sum_{p=0}^{n} \frac{h^p}{p!} \int_{\mathbb{R}} (ix)^p \mathrm{e}^{itx} \, \mathrm{d}\mu(x) = \int_{\mathbb{R}} \int_0^h \frac{(h-s)^n}{n!} (ix)^{n+1} \mathrm{e}^{ix(s+t)} \, \mathrm{d}s \, \mathrm{d}\mu(x).$$

Par inégalité triangulaire on obtient :

$$\left| \varphi(t+h) - \sum_{p=0}^{n} \frac{h^p}{p!} \varphi^{(p)}(t) \right| \leqslant \int_{\mathbb{R}} \int_0^{|h|} \frac{(h-s)^n}{n!} |x|^{n+1} \, \mathrm{d}s \, \mathrm{d}\mu(x) = \frac{|h|^{n+1}}{(n+1)!} \mu_{n+1}. \tag{1}$$

Notons $r := \overline{\lim_n} \, (\mu_n)^{1/n}/n$ qui est fini par hypothèse.
Alors il existe $N \in \mathbb{N}$ tel que $\forall n \geqslant N, \ \frac{(\mu_n)^{1/n}}{n} \leqslant r+1$. Ainsi : $\forall n \geqslant N, \ \mu_n \leqslant n^n (r+1)^n$.
Le développement en série de e^n donne $\mathrm{e}^n = \displaystyle\sum_{k=0}^{\infty} \frac{n^k}{k!} \geqslant \frac{n^n}{n!}$.
Prenons h vérifiant $|h| \leqslant 1/(2\mathrm{e}(r+1))$.
On obtient alors :

$$\forall n \geqslant N, \qquad \frac{|h|^n}{n!} \mu_n \leqslant \frac{1}{(2\mathrm{e}(r+1))^n} \frac{n^n}{n!} (r+1)^n \leqslant \frac{1}{2^n}.$$

Cela fournit ainsi :

$$\forall n \geqslant N, \qquad \left| \varphi(t+h) - \sum_{p=0}^{n} \frac{h^p}{p!} \varphi^{(p)}(t) \right| \leqslant \frac{1}{2^{n+1}} \xrightarrow[n \to +\infty]{} 0$$

c'est-à-dire :

$$\forall t \in \mathbb{R}, \ \forall |h| \leqslant \frac{1}{2e(r+1)}, \quad \varphi(t+h) = \sum_{p=0}^{\infty} \frac{\varphi^{(p)}(t)}{p!} h^p.$$

Cela montre que φ est analytique sur $\mathbb{R}$.
Puisque $\varphi^{(n)}(0) = i^n m_n$, on obtient en particulier pour $t = 0$:

$$\forall |h| \leqslant \frac{1}{2e(r+1)}, \quad \varphi(h) = \sum_{p=0}^{\infty} \frac{i^p m_p}{p!} h^p.$$

Ainsi on en déduit que μ est caractérisée par ses moments.
En effet si ν est une autre probabilité vérifiant les mêmes hypothèses, alors les deux fonctions caractéristiques coïncident au voisinage de 0 et sont analytiques donc égales. Comme la fonction caractéristique caractérise la loi, on peut conclure que $\mu = \nu$. $\square$

Remarque 2.3 (Analytique)**.** *On a montré lors de la démonstration que si la fonction caractéristique d'une variable aléatoire est analytique alors cette variable aléatoire est caractérisée par ses moments.*

Remarque 2.4. *On peut encore affaiblir les hypothèses en ne supposant que $\overline{\lim\limits_{n}}\,(\mu_{2n})^{1/2n}/(2n)$ fini.*

Théorème 2.5 (Critère des moments de Fourier amélioré)**.** *Si μ est une mesure de probabilité sur $\mathbb{R}$ ayant tous ses moments absolus μ_n finis alors si la quantité $\overline{\lim\limits_{n}}\,(\mu_{2n})^{1/2n}/(2n)$ est finie, μ est déterminée par ses moments.*

Démonstration. Par l'inégalité de Cauchy–Schwarz,

$$\mu_{2n+1} = \int |x|^n |x|^{n+1}\ \mathrm{d}\mu(x) \leqslant \sqrt{\int |x|^{2n}\ \mathrm{d}\mu(x)}\sqrt{\int |x|^{2n+2}\ \mathrm{d}\mu(x)} = \sqrt{\mu_{2n}}\sqrt{\mu_{2n+2}}.$$

Ainsi,

$$(\mu_{2n+1})^{1/2n+1} \leqslant (\sqrt{\mu_{2n}\mu_{2n+2}})^{1/2n+1} \leqslant (\mu_{2n})^{1/2n+1} + (\mu_{2n+2})^{1/2n+1}.$$

Puis,

$$\frac{(\mu_{2n+1})^{1/2n+1}}{2n+1} \leqslant \frac{(\mu_{2n})^{1/2n+1}}{2n+1} + \frac{(\mu_{2n+2})^{1/2n+1}}{2n+1}. \tag{2}$$

On a

$$(\mu_{2n})^{1/2n+1} = \left((\mu_{2n})^{1/2n}\right)^{2n/(2n+1)} \leqslant \max\left\{(\mu_{2n})^{1/2n}, \sqrt{(\mu_{2n})^{1/2n}}\right\} \text{ car } 1/2 \leqslant 2n/(2n+1) \leqslant 1.$$

Ce qui fournit

$$\frac{(\mu_{2n})^{1/2n+1}}{2n+1} \leqslant \frac{2n}{2n+1}\frac{(\mu_{2n})^{1/2n}}{2n} + \frac{\sqrt{2n}}{2n+1}\sqrt{\frac{(\mu_{2n})^{1/2n}}{2n}} \leqslant \frac{(\mu_{2n})^{1/2n}}{2n} + \sqrt{\frac{(\mu_{2n})^{1/2n}}{2n}} \tag{3}$$

puisque pour $a, b \geqslant 0, \max\{a, b\} \leqslant a + b$ et $\sqrt{2n} \leqslant 2n \leqslant 2n + 1$ pour tout $n \geqslant 1$.
De même, écrivons

$$\frac{(\mu_{2n+2})^{1/2n+1}}{2n+1} = \left(\frac{(\mu_{2n+2})^{1/2n+2}}{2n+2}\right)^{(2n+2)/(2n+1)} \frac{(2n+2)^{(2n+2)/(2n+1)}}{2n+1}$$

et

$$\frac{(2n+2)^{(2n+2)/(2n+1)}}{2n+1} = \frac{2n+2}{2n+1}(2n+2)^{1/(2n+1)} \leqslant 2 \times 2 = 4.$$

En effet, $(2n+2)^{1/(2n+1)} \leqslant 2 \iff 2n+2 \leqslant 2^{2n+1} \iff n+1 \leqslant 4^n$ ce qui est toujours vrai.

Ainsi

$$\frac{(\mu_{2n+2})^{1/2n+1}}{2n+1} \leqslant 4\max\left\{\frac{(\mu_{2n+2})^{1/(2n+2)}}{2n+2}, \left(\frac{(\mu_{2n+2})^{1/(2n+2)}}{2n+2}\right)^2\right\} \text{ car } 1 \leqslant (2n+2)/(2n+1) \leqslant 2. \quad (4)$$

Notons $u_n = (\mu_{2n})^{1/2n}/(2n)$ alors en utilisant que pour $a, b \geqslant 0, \max\{a,b\} \leqslant a+b$ dans (4), on obtient avec (3) et (4) l'inégalité suivante issue de (2) :

$$\frac{(\mu_{2n+1})^{1/2n+1}}{2n+1} \leqslant u_n + \sqrt{u_n} + 4u_{n+2} + 4u_{n+2}^2.$$

On passe à la limite supérieure pour obtenir, par continuité de $x \mapsto \sqrt{x}$ et $x \mapsto x^2$:

$$\overline{\lim_n} \frac{(\mu_{2n+1})^{1/2n+1}}{2n+1} \leqslant \overline{\lim_n} u_n + \sqrt{\overline{\lim_n} u_n} + 4\overline{\lim_n} u_{n+2} + 4\left(\overline{\lim_n} u_{n+2}\right)^2 < +\infty \text{ car } \overline{\lim_n} u_n < +\infty.$$

Ainsi comme $\overline{\lim_n} \frac{(\mu_{2n})^{1/2n}}{2n}$ et $\overline{\lim_n} \frac{(\mu_{2n+1})^{1/2n+1}}{2n+1}$ sont finis, $\overline{\lim_n} \frac{(\mu_n)^{1/n}}{n}$ est fini. □

Proposition 2.6 (Conditions suffisantes de Fourier équivalentes). *Soit μ est une mesure de probabilité sur $\mathbb{R}$ ayant tous ses moments absolus μ_n finis alors si une des conditions équivalentes*

(i) La quantité $R^{-1} := \overline{\lim_n} \left(\frac{\mu_n}{n!}\right)^{1/n}$ est finie.

(ii) La quantité $r := \overline{\lim_n} \frac{(\mu_n)^{1/n}}{n}$ est finie.

est vraie alors μ est déterminée par ses moments.

Dans ce cas, la fonction caractéristique φ est analytique sur $\mathbb{R}$ et particulier au voisinage de 0 :

$$\forall t \in]-R, R[, \quad \varphi(t) = \sum_{n \geqslant 0} \frac{(it)^n}{n!} \mathbb{E}(X^n).$$

Démonstration. Il suffit de remarquer que

$$\frac{n!}{n^n} = \frac{n \times (n-1) \times \cdots \times 1}{n \times n \times \cdots \times n} = \prod_{k=0}^{n-1}\left(1 - \frac{k}{n}\right) \leqslant 1 \quad \text{car chaque terme est plus petit que 1}$$

et d'utiliser l'inégalité $\left(\frac{n!}{n^n}\right)^{1/n} \geqslant e^{-1}$ issue du développement en série entière de e pour obtenir ainsi :

$$\mathrm{e}^{-1} \overline{\lim_n} \left(\frac{\mu_n}{n!}\right)^{1/n} \leqslant \overline{\lim_n} \frac{(\mu_n)^{1/n}}{n} = \overline{\lim_n} \frac{(\mu_n)^{1/n}}{(n!)^{1/n}} \left(\frac{n!}{n^n}\right)^{1/n} \leqslant \overline{\lim_n} \left(\frac{\mu_n}{n!}\right)^{1/n}$$

c'est à dire $\mathrm{e}^{-1}R^{-1} \leqslant r \leqslant R^{-1}$ ce qui prouve que R^{-1} est fini si et seulement si r est fini.
Le théorème 2.2 conclut. □

Remarque 2.7. *Si on connaît la fonction caractéristique φ de X alors, si elle est développable en série entière, il existe au plus une loi μ.*

Corollaire 2.8 (Condition suffisante)**.** *S'il existe $\alpha > 0$ tel que $\mathbb{E}(\mathrm{e}^{\alpha|X|})$ est fini alors la fonction caractéristique de X, notée φ, est analytique. Ainsi μ est déterminée par ses moments.*

Démonstration. Soit $|h| < \alpha$. On a l'expression

$$\mathbb{E}(\mathrm{e}^{|hX|}) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{(|hX|)^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{\mathbb{E}(|X|^k)}{k!}|h|^k$$

par le théorème de convergence monotone puisque tous les termes sont positifs.
Comme le membre de gauche est fini pour $|h| < \alpha$ alors celui de droite l'est aussi.
Ainsi le rayon de convergence de la série $\sum \frac{\mu_k}{k!}x^k$ est d'au moins α.
Or ce rayon de convergence R vérifie aussi $R^{-1} = \varlimsup_n \left(\frac{\mu_n}{n!}\right)^{1/n}$ par la formule de Cauchy-Hadamard.
Puisque $R \geqslant \alpha$ on a alors $\varlimsup_n \left(\frac{\mu_n}{n!}\right)^{1/n} = \frac{1}{R} \leqslant \frac{1}{\alpha} < +\infty$ car $\alpha \neq 0$.
D'après la proposition 2.6, φ est analytique. Ainsi μ est déterminée par ses moments. □

2.3 Exemples avec des lois usuelles

2.3.1 Le cas de la loi Log-Normale

Exemple 2.9 (Fonction caractéristique non analytique)**.** *La fonction caractérisque d'une variable aléatoire suivant la loi Log-Normale n'est pas analytique.*

Démonstration. Soit X une variable aléatoire suivant la loi Log-Normale. Supposons par l'absurde que φ soit analytique. Alors d'après la proposition 2.1, φ s'écrirait au voisinage de 0 :

$$\exists\varepsilon > 0,\ \forall t \in]-\varepsilon, \varepsilon[, \quad \varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!}\mathbb{E}(X^n).$$

Le rayon de convergence R de la série vérifie $R^{-1} := \varlimsup_n (|i^n m_n|/(n!))^{1/n}$ d'après la formule de Cauchy-Hadamard. En procédant comme dans la démonstration de la proposition 2.6, on a que R^{-1} est fini si et seulement si $r := \varlimsup_n (|m_n|)^{1/n}/n$ est fini. De plus, comme X suit une loi Log-Normale, $\mathbb{E}(X^n) = \mathrm{e}^{n^2/2}$ d'après la proposition 1.2. Ainsi on obtient l'expression explicite de r : $r = \varlimsup_n \mathrm{e}^{n/2}/n = +\infty$.
Cela donne alors $R^{-1} = +\infty$, et fournit $R = 0$.

Donc la série est de rayon de convergence nul, ce qui contredit l'existence de ε et conclut que φ n'est pas analytique. □

Deuxième démonstration. Supposons par l'absurde que φ est analytique. Or d'après la remarque 2.7, la loi de X serait caractérisée par ses moments, ce qui n'est pas le cas d'après l'exemple fondamental 1.3. Par conséquent φ n'est pas analytique. □

Remarque 2.10. *Cet exemple donne une illustration d'une fonction de classe $\mathcal{C}^\infty$ qui n'est pas analytique.*

En effet, par ce qui précède, on peut donc dire que

$$\varphi : t \mapsto \int_0^{+\infty} \mathrm{e}^{itx} \frac{\mathrm{e}^{-\ln(x)^2}}{\sqrt{2\pi}x}\,\mathrm{d}x$$

n'est pas analytique, bien que de classe $\mathcal{C}^\infty$ d'après la proposition 2.1 étant donnée qu'elle admet des moments absolus à tout ordre.

2.3.2 La loi Normale

Proposition 2.11. *La loi Normale est caractérisée par ses moments.*

Démonstration. En centrant et en réduisant une variable aléatoire suivant une loi normale on se ramène à une loi normale de paramètres 0 et 1.
Calculons l'intégrale suivante par une intégration par parties pour $n \geqslant 2$:

$$\begin{aligned}
\mu_n &= \int_{\mathbb{R}} |x|^n \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x = 2\int_0^{+\infty} x^n \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x \\
&= \left[-2x^{n-1}\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\right]_0^{+\infty} - \int_0^{+\infty} -2(n-1)x^{n-2}\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x \\
&= 2(n-1)\int_0^{+\infty} x^{n-2}\frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x \\
&= (n-1)\mu_{n-2}.
\end{aligned}$$

Ainsi, si n est pair, on a : $\mu_n = (n-1)\times(n-3)\times\ldots\times 1 \times \mu_0 \leqslant C_1\,(n!)$ où $C_1 > 0$ et si n est impair on a : $\mu_n = (n-1)\times(n-3)\times\ldots\times 2\times\mu_1 \leqslant C_2\,(n!)$ où $C_2 > 0$.
Par conséquent, il existe $C > 0$ tel que $\mu_n \leqslant C\,(n!)$. Donc $\overline{\lim_n}\,(\mu_n/n!)^{1/n} \leqslant \overline{\lim_n}\,C^{1/n} < +\infty$.
D'après la proposition 2.6, il y a unicité de la loi ayant ces moments : c'est la loi normale. □

Remarque 2.12. *On a montré pour $p \in \mathbb{N}$: $\mu_{2p} = m_{2p} = (2p)!/(2^p p!)$, $\mu_{2p+1} = 2^{p+1}p!\,\mu_1$ et $m_{2p+1} = 0$.*

2.3.3 La loi Gamma

Définition 2.13. *On définit la densité d'une loi Gamma de paramètres $\lambda > 0, a > 0$ par*

$$f : x \mapsto \frac{\lambda^a}{\Gamma(a)}\mathrm{e}^{-\lambda x}x^{a-1}\mathbb{1}_{x\geqslant 0}.$$

Lemme 2.14. *Les moments d'une loi Gamma sont donnés par $m_n = a\times(a+1)\times\cdots\times(a+n-1)/\lambda^n$.*

Démonstration. Pour calculer m_n, effectuons le changement de variable $y = \lambda x$:

$$\begin{aligned}
\mathbb{E}(X^n) &= \frac{\lambda^a}{\Gamma(a)}\int_0^{+\infty} x^{a-1+n}\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{\lambda^a}{\Gamma(a)}\int_0^{+\infty}\left(\frac{y}{\lambda}\right)^{a-1+n}\mathrm{e}^{-y}\,\frac{\mathrm{d}y}{\lambda} = \frac{1}{\Gamma(a)\lambda^n}\int_0^{+\infty} y^{a-1+n}\mathrm{e}^{-y}\,\mathrm{d}y \\
&= \frac{\Gamma(a+n)}{\Gamma(a)\,\lambda^n} = \frac{a\times(a+1)\times\cdots\times(a+n-1)}{\lambda^n}.
\end{aligned}$$

□

Proposition 2.15. *La loi Gamma est caractérisée par ses moments.*

Démonstration (en calculant les moments). Utilisons le lemme 2.14.
Tout d'abord les quantités m_n et μ_n coïncident puisque X est à support dans $\mathbb{R}_+$.
Remarquons que pour n grand on a : $a \leqslant n$, on a ainsi pour $0 \leqslant k \leqslant n-1$ et n grand : $a+k \leqslant 2n$.
Cela fournit alors pour n grand :

$$m_n \leqslant \frac{(2n)^n}{\lambda^n} = \left(\frac{2n}{\lambda}\right)^n \quad \text{et} \quad \overline{\lim_n} \frac{(\mu_n)^{1/n}}{n} \leqslant \overline{\lim_n} \frac{2n}{\lambda n} = \frac{2}{\lambda} < +\infty.$$

D'après le théorème 2.2, la loi Gamma est caractérisée par ses moments. $\square$

Intéressons nous également à la fonction caractéristique de la loi Gamma, cela donnera une seconde preuve de la proposition 2.15.

Lemme 2.16. *La fonction caractéristique φ d'une variable aléatoire suivant une loi Gamma de paramètres λ et a est $\varphi : t \mapsto (\lambda/(\lambda - it))^a$ pour tout $t \in \mathbb{R}$ et est analytique sur $\mathbb{R}$.*

Démonstration. Tout d'abord, calculons la transformée de Laplace de X pour $t < \lambda$ en effectuant le changement de variable $y = (\lambda - t)x$:

$$\mathbb{E}(\mathrm{e}^{tX}) = \frac{\lambda^a}{\Gamma(a)} \int_0^{+\infty} x^{a-1}\, \mathrm{e}^{-(\lambda-t)x}\, \mathrm{d}x = \frac{\lambda^a}{\Gamma(a)} \int_0^{+\infty} \frac{y^{a-1}}{(\lambda-t)^a}\, \mathrm{e}^{y}\, \mathrm{d}y = \frac{\lambda^a}{\Gamma(a)} \frac{\Gamma(a)}{(\lambda-t)^a} = \left(\frac{\lambda}{\lambda-t}\right)^a. \quad (5)$$

Désormais notons $D = \{z \in \mathbb{C}, \Re(z) < \lambda\}$ et posons :

$$\forall z \in D, \quad F(z) := \frac{\lambda^a}{\Gamma(a)} \int_0^{+\infty} x^{a-1}\, \mathrm{e}^{(z-\lambda)x}\, \mathrm{d}x \qquad \text{et} \qquad \forall z \in D, x > 0, \quad f(x,z) = x^{a-1}\, \mathrm{e}^{(z-\lambda)x}.$$

Montrons que F holomorphe sur D. Appliquons le théorème d'holomorphie sous l'intégrale.

- $\forall z \in D,\ x \mapsto f(x,z)$ est mesurable.
- $\forall x > 0,\ z \mapsto f(x,z)$ est holomorphe.
- Soient $D_\varepsilon = \{z \in \mathbb{C},\ \Re(z) - \lambda \leqslant \varepsilon\}$ et K un compact de D. Alors il existe $\varepsilon > 0$ tel que $K \subset D_\varepsilon$. Pour $z \in K$, on a $|f(x,z)| = x^{a-1}\, \mathrm{e}^{(\Re(z)-\lambda)x} \leqslant x^{a-1}\, \mathrm{e}^{-\varepsilon x} \in L^1(\mathbb{R}_+)$.

Ainsi F est holomorphe sur K pour tout compact $K \subset D$, donc F est holomorphe sur D.
Or d'après (5) :

$$\forall z \in D \cap \mathbb{R},\ F(z) = \left(\frac{\lambda}{\lambda - z}\right)^a.$$

De plus, la fonction $z \mapsto (\lambda/\lambda - z)^a$ est holomorphe sur D car elle s'écrit $z \mapsto \lambda^a \mathrm{e}^{-a\log(\lambda - z)}$ qui est holomorphe, la fonction log étant prise en détermination principale étant donné que $\Re(\lambda - z) > 0$.

Ainsi les fonctions F et $z \mapsto (\lambda/\lambda - z)^a$, holomorphes sur D, coïncident sur l'ouvert connexe $D \cap \mathbb{R}$ possédant un point d'accumulation (1 par exemple).
Par le théorème de prolongement analytique, elles sont donc égales sur D.
On obtient ainsi l'expression de φ

$$\forall t \in \mathbb{R},\ \varphi(t) = F(it) = \left(\frac{\lambda}{\lambda - it}\right)^a$$

et que φ est analytique sur $\mathbb{R}$. $\square$

Démonstration de la proposition 2.15 (avec la fonction caractéristique). D'après le lemme 2.16, φ est analytique sur $\mathbb{R}$, ce qui est une condition suffisante d'après la remarque 2.7 pour affirmer que la loi Gamma est caractérisée par ses moments. □

Remarque 2.17. *Comme la loi Gamma est caractérisée par ses moments, c'est en particulier le cas pour les lois exponentielles et du chi-deux.*

2.3.4 La loi Géométrique

Proposition 2.18. *La loi Géométrique est caractérisée par ses moments.*

Démonstration. Considérons une loi géométrique de paramètre $0 < p < 1$.
Utilisons le corollaire 2.8. Pour cela cherchons s'il existe $\alpha > 0$ tel que $\mathbb{E}(\mathrm{e}^{\alpha|X|})$ soit fini.

$$\mathbb{E}(\mathrm{e}^{\alpha|X|}) = \sum_{k=1}^{+\infty} \mathrm{e}^{\alpha k}(1-p)^{k-1}p = \frac{p}{1-p}\sum_{k=1}^{+\infty} \left(\mathrm{e}^{\alpha}(1-p)\right)^k.$$

Le membre de droite est fini si et seulement si $\mathrm{e}^{\alpha}(1-p) < 1$, ce qui est le cas pour $\alpha \in]0, -\ln(1-p)[$.
Prenons par exemple $\alpha = -\ln(1-p)/2 > 0$. Le corollaire 2.8 permet de conclure. □

2.3.5 La loi de Poisson

Proposition 2.19. *La loi de Poisson est caractérisée par ses moments.*

Démonstration. Considérons une loi de Poisson de paramètre $\lambda > 0$.
Utilisons le corollaire 2.8. Pour cela cherchons s'il existe $\alpha > 0$ tel que $\mathbb{E}(\mathrm{e}^{\alpha|X|})$ soit fini.

$$\mathbb{E}(\mathrm{e}^{\alpha|X|}) = \sum_{k=0}^{+\infty} \mathrm{e}^{\alpha k}\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} = \mathrm{e}^{-\lambda}\sum_{k=0}^{+\infty}\frac{(\mathrm{e}^{\alpha}\lambda)^k}{k!} = \mathrm{e}^{-\lambda}\mathrm{e}^{\mathrm{e}^{\alpha}\lambda} = \mathrm{e}^{\lambda(\mathrm{e}^{\alpha}-1)} < +\infty.$$

Prenons par exemple $\alpha = 1$. Le corollaire 2.8 permet de conclure. □

3 Problème de Hausdorff

Le problème de Hausdorff est le problème des moments lorsque l'intervalle est un segment $[a, b]$.

3.1 Théorème de convergence de Weierstrass

Introduisons les polynômes de Bernstein et le théorème d'approximation de Weierstrass.

Définition 3.1 (Polynômes de Bernstein)**.** *Pour tout entier naturel non nul n on définit le polynôme de Bernstein de f de degré n par :*

$$\forall x \in [0,1] \qquad B_n f(x) := \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}.$$

Théorème 3.2 (Approximation de Weierstrass). *Toute fonction f réelle continue sur un segment est limite uniforme de fonctions polynomiales. De plus, les polynômes de Bernstein de f fournissent une telle suite approchant f. On a l'inégalité suivante qui fournit la vitesse de convergence vers f :*

$$\|f - B_n f\|_\infty \leqslant \frac{3}{2}\,\omega\left(\frac{1}{\sqrt{n}}\right)$$

où ω est le module de continuité de f :

$$\begin{array}{rccc} \omega & : & \mathbb{R}_+ & \to & \mathbb{R}_+ \\ & & \delta & \mapsto & \displaystyle\sup_{|u-v|\leqslant\delta} |f(u)-f(v)| \end{array}.$$

Démonstration. Quitte à dilater par $t \mapsto a+(b-a)t$, on peut supposer $[a,b]=[0,1]$. Soit f une fonction continue sur le segment $[0,1]$. On définit ω comme ci-dessus, il est bien défini par uniforme continuité. En effet, d'après le théorème de Heine une fonction continue sur un segment est uniformément continue. De plus, ω vérifie les propriétés suivantes :

(i) ω est croissante.

(ii) $\omega(0)=0$ et ω est continue en 0.

(iii) Pour $\lambda \geqslant 0$ et $\delta \geqslant 0$, on a $\omega(\lambda\delta) \leqslant (\lambda+1)\,\omega(\delta)$.

Le point (i) est évident.

Pour (ii), le fait que $\omega(0)=0$ est évident. Fixons $\varepsilon>0$, par uniforme continuité de f il existe $\eta>0$ tel que $\omega(\eta)<\varepsilon$. Ainsi par croissance, $\omega \leqslant \varepsilon$ sur $]0,\eta]$, ce qui prouve que ω est continue en 0 avec $\omega(0)=0$.

Enfin pour (iii), prenons $\lambda, \delta > 0$ puis $u<v$ tels que $|v-u| \leqslant \lambda\delta$. Considérons la subdivision $x_i = u+i\delta$ si $u+i\delta<v$ puis $x_m = v$ dès que $u+m\delta \geqslant v$. Nécessairement $m \leqslant \lambda+1$. En effet, dans le pire des cas, $v-u=\lambda\delta$ et alors $x_{\lfloor\lambda\rfloor+1} = u+\lfloor\lambda\rfloor+1\delta \geqslant u+\lambda\delta = v$ ce qui prouve que $m \leqslant \lfloor\lambda\rfloor+1 \leqslant \lambda+1$.
Ainsi :

$$|f(v)-f(u)| \leqslant \sum_{i=0}^{m-1} |f(x_{i+1})-f(x_i)| \leqslant m\,\omega(\delta) \leqslant (\lambda+1)\,\omega(\delta).$$

Utilisons ces propriétés pour montrer le résultat. Soit $x \in [0,1]$. On considère une suite de variables aléatoires $(X_i)_{i\in\mathbb{N}}$ indépendantes suivant une loi de Bernoulli de paramètre x. On pose $S_n = \sum_{i=1}^n X_i$. Ainsi les polynômes de Bernstein s'expriment :

$$B_n f(x) := \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]$$

d'après le théorème de transfert appliqué à S_n.
On a par croissance de l'espérance :

$$\begin{array}{rcll} |f(x)-B_nf(x)| & = & \left|\mathbb{E}(f(x)) - \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]\right| = \left|\mathbb{E}\left[f(x)-f\left(\frac{S_n}{n}\right)\right]\right| & \text{(linéarité)} \\ & \leqslant & \mathbb{E}\left[\left|f(x)-f\left(\frac{S_n}{n}\right)\right|\right] & \text{(inégalité triangulaire)} \\ & \leqslant & \mathbb{E}\left[\omega\left(\left|x-\frac{S_n}{n}\right|\right)\right] & \text{(définition de } \omega) \\ & \leqslant & \mathbb{E}\left[\left(1+\sqrt{n}\left|x-\frac{S_n}{n}\right|\right)\omega\left(\frac{1}{\sqrt{n}}\right)\right] & \text{(d'après } (iii)) \\ & \leqslant & \left(1+\sqrt{n}\,\mathbb{E}\left[\left|x-\frac{S_n}{n}\right|\right]\right)\omega\left(\frac{1}{\sqrt{n}}\right). & (\star) \end{array}$$

Or d'après l'inégalité de Cauchy–Schwarz sur $[0,1]$:

$$\mathbb{E}\left[\left|x-\frac{S_n}{n}\right|\right] \leqslant \sqrt{\mathbb{E}\left[\left(x-\frac{S_n}{n}\right)^2\right]}. \tag{6}$$

Or

$$\mathbb{E}\left[x-\frac{S_n}{n}\right] = \mathbb{E}\left[x\right] - \mathbb{E}\left[\frac{S_n}{n}\right] = \mathbb{E}\left[x\right] - \mathbb{E}\left[X_1\right] = 0.$$

Donc

$$\mathbb{E}\left[\left(x-\frac{S_n}{n}\right)^2\right] = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{Var(X_1)}{n} = \frac{x(1-x)}{n} \leqslant \frac{1}{4n} \tag{7}$$

par indépendance des X_i et en utilisant l'inégalité classique $x(1-x) \leqslant 1/4$ pour $x \in [0,1]$.

Ainsi, la relation $(\star)$ de la série d'inégalités précédentes devient à l'aide des inégalités (6) et (7)

$$|f(x) - B_n f(x)| \leqslant \left(1+\sqrt{n}\,\mathbb{E}\left[\left|x-\frac{S_n}{n}\right|\right]\right)\omega\left(\frac{1}{\sqrt{n}}\right) \leqslant \left(1+\sqrt{n}\sqrt{\frac{1}{4n}}\right)\omega\left(\frac{1}{\sqrt{n}}\right) = \frac{3}{2}\,\omega\left(\frac{1}{\sqrt{n}}\right)$$

ce qui est valable pour tout $x \in [0,1]$ et fournit :

$$\|f - B_n f\|_\infty \leqslant \frac{3}{2}\,\omega\left(\frac{1}{\sqrt{n}}\right).$$

Puisque $1/\sqrt{n}$ tend vers 0 lorsque n tend vers l'infini, le point (ii) montre que le membre de droite tend vers 0 lorsque n tend vers l'infini. Cela prouve la convergence uniforme des polynômes de Bernstein $B_n f$ vers la fonction continue f. □

Remarque 3.3. *On ne peut pas étendre ce résultat à un intervalle non borné.*
En effet, si $(P_n)_n \in \mathbb{R}[X]$ converge uniformément sur $\mathbb{R}$ vers f continue, alors $(P_n)_n$ est de Cauchy pour la norme infinie sur $\mathbb{R}$. Ainsi, pour $\varepsilon > 0$ fixé, on a à partir d'un certain rang N_0 : $\|P_{n+1} - P_n\|_\infty \leqslant \varepsilon$. Or les seuls polynômes bornés sur $\mathbb{R}$ sont les constantes d'où pour $n \geqslant N_0$, $P_{n+1} - P_n$ est constant égal à la constante c_n. Alors $f = P_{N_0} + \sum_{n \geqslant N_0} c_n$ est aussi un polynôme.

Proposition 3.4. *L'inégalité du théorème 3.2 est optimale dans le sens où il existe une fonction f continue sur $[0,1]$ et $\delta > 0$ tels que :*

$$\forall n \in \mathbb{N}, \qquad \|f - B_n f\|_\infty \geqslant \delta\ \omega\left(\frac{1}{\sqrt{n}}\right).$$

Démonstration. Considérons $f : x \mapsto \left|x - \frac{1}{2}\right|$ sur $[0,1]$. Cette fonction est continue, 1-lipschitzienne et vérifie $\omega(\delta) \leqslant \delta$ car $|f(x) - f(y)| = \left|\,\left|x-\frac{1}{2}\right| - \left|\frac{1}{2} - y\right|\,\right| \leqslant |x-y| \leqslant \delta$ par inégalité triangulaire.

On a avec les notations précédentes :

$$\|f - B_n f\|_\infty \geqslant \left|f\left(\frac{1}{2}\right) - B_n f\left(\frac{1}{2}\right)\right| = \left|B_n f\left(\frac{1}{2}\right)\right| = \mathbb{E}\left[\left|\frac{S_n}{n} - \frac{1}{2}\right|\right] = \frac{1}{2n}\mathbb{E}\left[\,|2S_n - n|\,\right]. \tag{8}$$

Écrivons $2S_n - n = \sum_{i=1}^{n}(2X_i - 1) = \sum_{i=1}^{n} \varepsilon_i$ où les ε_i suivent une loi de Rademacher de paramètre $1/2$.
Posons ensuite $Y = \displaystyle\prod_{j=1}^{n}\left(1 + i\frac{\varepsilon_j}{\sqrt{n}}\right)$.
On a d'une part :

$$|Y|^2 = \prod_{j=1}^{n}\left(1 + \frac{\varepsilon_j^2}{n}\right) = \prod_{j=1}^{n}\left(1 + \frac{1}{n}\right) \leqslant \prod_{j=1}^{n} \mathrm{e}^{1/n} = \mathrm{e} \quad \text{car } 1 + x \leqslant \mathrm{e}^x.$$

Ce qui prouve

$$\mathbb{E}\left[\,|2S_n - n|\,|Y|\,\right] \leqslant \sqrt{e}\,\mathbb{E}[\,|2S_n - n|\,]. \tag{9}$$

Et d'autre part puisque les ε_j sont indépendantes et centrées :

$$\begin{aligned}
\mathbb{E}\left[\,(2S_n - n)\,Y\,\right] &= \mathbb{E}\left[\sum_{j=1}^{n} \varepsilon_j\,Y\right] = \sum_{j=1}^{n} \mathbb{E}\left[\,\varepsilon_j\,Y\,\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[\varepsilon_j\,\left(1 + i\frac{\varepsilon_j}{\sqrt{n}}\right)\prod_{\substack{k=1\\k\neq j}}^{n}\left(1 + i\frac{\varepsilon_k}{\sqrt{n}}\right)\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[\varepsilon_j\,\left(1 + i\frac{\varepsilon_j}{\sqrt{n}}\right)\right]\prod_{\substack{k=1\\k\neq j}}^{n} \mathbb{E}\left[\left(1 + i\frac{\varepsilon_k}{\sqrt{n}}\right)\right] \\
&= \sum_{j=1}^{n} \mathbb{E}\left[\left(\varepsilon_j + i\frac{\varepsilon_j^2}{\sqrt{n}}\right)\right]\prod_{\substack{k=1\\k\neq j}}^{n}\left(1 + i\frac{\mathbb{E}\left[\varepsilon_k\right]}{\sqrt{n}}\right) \\
&= \sum_{j=1}^{n} \mathbb{E}\left[\left(\varepsilon_j + i\frac{1}{\sqrt{n}}\right)\right]\prod_{\substack{k=1\\k\neq j}}^{n} 1 \\
&= \sum_{j=1}^{n}\left(\mathbb{E}\left[\varepsilon_j\right] + \frac{i}{\sqrt{n}}\right) = \sum_{j=1}^{n}\frac{i}{\sqrt{n}} = i\sqrt{n}.
\end{aligned}$$

Cela prouve l'inégalité suivante par inégalité triangulaire :

$$\mathbb{E}\left[\,|2S_n - n|\,|Y|\,\right] \geqslant |\mathbb{E}\left[\,(2S_n - n)\,Y\,\right]| = \left|i\sqrt{n}\right| = \sqrt{n}. \tag{10}$$

Les inégalités (8), (9) et (10) fournissent ainsi :

$$\|f - B_n f\|_\infty \geqslant \frac{1}{2n}\,\mathbb{E}[\,|2S_n - n|\,] \geqslant \frac{1}{2n}\sqrt{\frac{n}{e}} = \frac{1}{2\sqrt{e}}\frac{1}{\sqrt{n}} \geqslant \frac{1}{2\sqrt{e}}\,\omega\left(\frac{1}{\sqrt{n}}\right)$$

puisque $\delta \geqslant \omega(\delta)$ par construction, ce qui prouve l'inégalité recherchée. □

Remarque 3.5. *Si f est k-lipschitzienne, on a $\omega(\delta) \leqslant k\delta$ donc la vitesse de convergence est en $1/\sqrt{n}$.*

Remarque 3.6 (Cas particulier des fonctions puissances)**.** *Soient $k \geqslant 1$ et $f : x \mapsto x^k$ définie sur $[0,1]$. Puisque f est de classe $\mathcal{C}^1$ alors on a par l'inégalité des accroissements finis pour $x, y \in [0,1], x < y$:*

$$|f(x) - f(y)| \leqslant \sup_{x<z<y} |f'(z)||x-y| \leqslant \sup_{0<z<1} k\,|z^{k-1}|\,|x-y| \leqslant k\,|x-y|$$

ce qui reste valable pour $k = 0$. Ainsi la vitesse de convergence des polynômes de Bernstein de f vers f est d'après le théorème 3.2 de Weierstrass d'au moins $k/\sqrt{n}$.

3.2 Unicité de la mesure

Dans le cas du problème de Hausdorff, il y a unicité d'une loi μ à support dans $[a,b]$ ayant pour moments les réels m_n. On peut montrer cela de deux manières différentes, c'est ce qui fait l'objet des deux démonstrations suivantes.

Théorème 3.7 (Unicité de la mesure)**.** *Il existe au plus une mesure solution du problème de Hausdorff.*

Démonstration (avec la condition suffisante de Fourier). Soit X une variable aléatoire de loi μ. Comme μ est à support dans un intervalle $[a,b]$, on a $|X| \leqslant \max\{|a|,|b|\} := m$ et donc $\mathbb{E}(\mathrm{e}^{|X|}) \leqslant \mathbb{E}(\mathrm{e}^m) < +\infty$. D'après la condition suffisante énoncée au corollaire 2.8 (pour $\alpha = 1$), μ est caractérisée par ses moments. □

Démonstration (par approximation). Soit μ et ν deux mesures de probabilités concentrées sur un intervalle $[a,b]$ ayant les mêmes moments :

$$\forall p \geqslant 0, \quad \int x^p \,\mathrm{d}\mu(x) = \int x^p \,\mathrm{d}\nu(x).$$

Cela fournit par linéarité que pour tout polynôme $P \in \mathbb{R}[X]$:

$$\int P(x)\,\mathrm{d}\mu(x) = \int P(x)\,\mathrm{d}\nu(x).$$

Alors, d'après le théorème d'approximation uniforme de Weierstrass 3.2, pour toute fonction f réelle continue, il existe une suite de polynômes $(P_n)_n$ de $\mathbb{R}[X]$ convergeant uniformément vers f sur le segment $[a,b]$. Ainsi on obtient par convergence uniforme l'égalité pour toute fonction continue f :

$$\begin{aligned}\int f(x)\,\mathrm{d}\mu(x) &= \int \lim_n P_n(x)\,\mathrm{d}\mu(x) = \lim_n \int P_n(x)\,\mathrm{d}\mu(x) = \lim_n \int P_n(x)\,\mathrm{d}\nu(x) \\ &= \int \lim P_n(x)\,\mathrm{d}\nu(x) = \int f(x)\,\mathrm{d}\nu(x).\end{aligned}$$

Cela est valable en particulier pour la fonction continue $x \in [a,b] \mapsto f_k(x) := \min\{1, k\,\mathrm{dist}\,(x, I^c)\}$ où I est un intervalle de $\mathbb{R}$ ouvert et $k > 0$.
Remarquons que cette fonction est croissante, positive et que $\lim_{k\to+\infty} f_k(x) = 1$ si $x \in I$, 0 sinon.
D'après le théorème de convergence monotone :

$$\begin{aligned}\mu(I) &= \int \mathbb{1}_I(x)\,\mathrm{d}\mu(x) = \int \lim_{k\to+\infty} f_k(x)\,\mathrm{d}\mu(x) = \lim_{k\to+\infty} \int f_k(x)\,\mathrm{d}\mu(x) \\ &= \lim_{k\to+\infty} \int f_k(x)\,\mathrm{d}\nu(x) = \lim_{k\to+\infty} \int f_k(x)\,\mathrm{d}\nu(x) = \int \mathbb{1}_I(x)\,\mathrm{d}\nu(x) = \nu(I).\end{aligned}$$

Comme l'égalité $\mu(I) = \nu(I)$ est valable pour tout intervalle ouvert I, et que la tribu engendrée par les intervalles ouverts est la tribu borélienne, on obtient que les mesures sont égales, ce qui conclut. □

Exemple 3.8. *Les lois bêta, uniforme, binomiale, hypergéométrique et du demi-cercle sont des lois à support borné. Ainsi, elles sont caractérisées par leurs moments.*

Il reste ainsi à se poser la question de l'existence d'une telle mesure de probabilité.

3.3 Existence de la mesure

Le but des prochaines sous-parties sera de démontrer le théorème suivant, donnant l'existence d'une mesure satisfaisant au problème de Hausdorff. Énonçons-le et donnons ensuite quelques définitions.

Théorème 3.9 (Existence de la mesure)**.** *Il existe une mesure solution du problème de Hausdorff ayant les moments m_n si et seulement si la suite $(m_n)_n$ est complètement monotone.*

3.3.1 Définitions et conventions

Dans toute la suite, on supposera pour plus de commodité l'intervalle $[a, b]$ comme étant $[0, 1]$.
On impose alors $m_0 = 1$ puisque

$$m_0 = \int_0^1 x^0 \ \mathrm{d}\mu(x) = \mu([0,1]) = 1. \tag{11}$$

Définition 3.10 (Opérateur de différence)**.** *On définit l'opérateur de différence Δ comme :*

$$\begin{array}{cccc} \Delta & : & \mathbb{R}^{\mathbb{N}} & \to & \mathbb{R}^{\mathbb{N}} \\ & & (a_n)_n & \mapsto & (a_{n+1} - a_n)_n \end{array}$$

et l'opérateur de différence itéré Δ^{k+1} défini pour tout $k \in \mathbb{N}$ comme $\Delta^{k+1} = \Delta^k \circ \Delta$ et $\Delta^0 = id$.

Définition 3.11 (Suite complètement monotone)**.** *Une suite $a = (a_n)_n$ est dite complètement monotone si : $\forall k \in \mathbb{N},\ n \in \mathbb{N},\ \ \Big\{(-1)^k \Delta^k(a)\Big\}_n \geqslant 0$.*

L'opérateur Δ est régulièrement utilisé dans la littérature mais pour des raisons pratiques, nous introduisons l'opérateur opposé de différence D ci-après.

Définition 3.12 (Opérateur opposé de différence)**.** *On notera aussi dans la suite D l'opérateur $-\Delta$. Par conséquent, une suite $a = (a_n)_n$ est dite complètement monotone si : $\forall k \in \mathbb{N},\ n \in \mathbb{N},\ (D^k(a))_n \geqslant 0$.*

Établissons une expression explicite de $(\mathrm{D}^n m)_i$ qui servira de nombreuses fois.

Proposition 3.13. *Pour une suite $m = (m_n)_n$ on a l'égalité :*

$$\forall n \in \mathbb{N},\ \forall i \in \mathbb{N}, \qquad (D^n m)_i = \sum_{j=0}^{n} \binom{n}{j} (-1)^j \ m_{i+j}. \tag{12}$$

Démonstration. Effectuons une récurrence sur $n \in \mathbb{N}$.
Pour $n = 0$, c'est vrai car le membre de droite de (12) vaut m_i et $(\mathrm{D}^0 m)_i = m_i$.
Supposons désormais l'égalité (12) vraie pour un certain $n \in \mathbb{N}$ et montrons qu'elle est vraie pour $n+1$.
Soit $i \in \mathbb{N}$. On a par définition de D et par hypothèse de récurrence :

$$
\begin{aligned}
(\mathrm{D}^{n+1}m)_i &= (\mathrm{D}^n m)_i - (\mathrm{D}^n m)_{i+1} \\
&= \sum_{j=0}^{n} \binom{n}{j} (-1)^j \; m_{i+j} - \sum_{j=0}^{n} \binom{n}{j} (-1)^j \; m_{i+1+j} \\
&= \sum_{j=0}^{n} \binom{n}{j} (-1)^j \; m_{i+j} - \sum_{j=1}^{n+1} \binom{n}{j-1} (-1)^{j-1} \; m_{i+j} \quad \text{(changement de variable } j \leftarrow j-1) \\
&= m_i - (-1)^n m_{i+n+1} + \sum_{j=1}^{n} \left(\binom{n}{j} + \binom{n}{j-1} \right) (-1)^j \; m_{i+j} \\
&= m_i + (-1)^{n+1} m_{i+n+1} + \sum_{j=1}^{n} \binom{n+1}{j} (-1)^j \; m_{i+j} \qquad \text{(formule de Pascal)} \\
&= \sum_{j=0}^{n+1} \binom{n+1}{j} (-1)^j \; m_{i+j}.
\end{aligned}
$$

Ce qui achève la preuve par récurrence. □

3.3.2 Démonstration de la condition nécessaire

L'objectif de cette partie est de montrer la condition nécessaire du théorème d'existence 3.9, explicitée dans le théorème ci-dessous.

Théorème 3.14. *Si il existe une mesure solution du problème de Hausdorff ayant les moments m_n alors la suite des moments $(m_n)_n$ est complètement monotone.*

Démonstration. Supposons qu'il existe une mesure solution du problème de Hausdorff. Cela signifie que la suite $m := (m_n)_n$ est une suite de moments d'une variable aléatoire X à support dans $[0,1]$.
Montrons par récurrence sur $k \in \mathbb{N}$ la proposition suivante :

$$\mathcal{P}(k) : \quad \forall n \in \mathbb{N}, \; \left\{ (-1)^k \Delta^k(m) \right\}_n = \mathbb{E}((1-X)^k X^n).$$

La propriété $\mathcal{P}(0)$ est vraie car $\left\{(-1)^k \Delta^k(m)\right\}_n = \{m\}_n = m_n = \mathbb{E}((1-X)^k X^n)$.
Supposons $\mathcal{P}(k)$ vraie et montrons $\mathcal{P}(k+1)$. Soit $n \in \mathbb{N}$. On a l'égalité :

$$
\begin{aligned}
\left\{ (-1)^{k+1} \Delta^{k+1}(m) \right\}_n &= (-1)^{k+1} \left\{ \Delta \left(\Delta^k(m) \right) \right\}_n \\
&= (-1)^{k+1} \left(\{\Delta^k(m)\}_{n+1} - \{\Delta^k(m)\}_n \right) \\
&= (-1)^k \{\Delta^k(m)\}_n - (-1)^k \{\Delta^k(m)\}_{n+1} \\
&= \mathbb{E}((1-X)^k X^n) - \mathbb{E}((1-X)^k X^{n+1}) \quad \text{(par hypothèse de récurrence)} \\
&= \mathbb{E}((1-X)^k X^n (1-X)) \quad \text{(par linéarité)} \\
&= \mathbb{E}((1-X)^{k+1} X^n).
\end{aligned}
$$

Ce qui prouve $\mathcal{P}(k+1)$ et achève la récurrence.
Puisque $X \in [0,1]$, $(1-X)^k X^n \geqslant 0$ quelque soit les entiers k et n.
Cela fournit d'après la propriété $\mathcal{P}(k)$: $\forall k \in \mathbb{N},\ \forall n \in \mathbb{N},\ \mathbb{E}((1-X)^k X^n) = \left\{(-1)^k \Delta^k(m)\right\}_n \geqslant 0$.
C'est à dire que la suite des moments $(m_n)_n$ est complètement monotone. □

3.3.3 Démonstration de la condition suffisante

L'objectif de cette partie est de montrer la condition suffisante du théorème d'existence 3.9, explicitée dans le théorème suivant.

Théorème 3.15. *Si la suite $(m_n)_n$ est complètement monotone alors il existe une mesure solution du problème de Hausdorff ayant les moments m_n.*

Proposition 3.16. *Supposons que la suite $(m_n)_n$ est complètement monotone avec $m_0 = 1$. Alors la fonction F_n définie par $F_n(x) := \sum_{i \leqslant nx} \binom{n}{i} (D^{n-i}m)_i$ pour $x \in [0,1]$ définit une fonction de répartition.*

Démonstration. Supposons que la suite $m := (m_n)_n$ est complètement monotone, avec $m_0 = 1$ (Cf. 11). Montrons que F_n est croissante, continue à droite et vérifie $F_n(0) = 0$ et $F_n(1) = 1$.
Puisque la suite m est complètement monotone, chaque terme de la somme est positif d'après 3.12. Par conséquent la fonction F_n est croissante. Elle est de plus continue à droite et vérifie $F_n(0) = 0$. Montrons que $F_n(1) = 1$, c'est-à-dire

$$\sum_{i \leqslant n} \binom{n}{i} (\mathrm{D}^{n-i}m)_i = m_0 = 1. \tag{13}$$

Remarquons tout d'abord que tirer i boules dans un sac de n boules et les colorier en bleu puis tirer encore j boules et les colorier en rouge est similaire à tirer $i+j$ boules dans le sac de n boules et d'en choisir i à colorier en bleu et le reste en rouge. Cela se traduit mathématiquement par l'identité :

$$\binom{n}{i}\binom{n-i}{j} = \binom{n}{i+j}\binom{i+j}{i}. \tag{14}$$

Montrons (13) à l'aide de la remarque ci-dessus et de la proposition 3.13. En effet on a :

$$\begin{aligned}
\sum_{i \leqslant n} \binom{n}{i} (\mathrm{D}^{n-i}m)_i &= \sum_{i=0}^{n} \binom{n}{i} \sum_{j=0}^{n-i} \binom{n-i}{j} (-1)^j\ m_{i+j} \\
&= \sum_{i=0}^{n} \sum_{j=0}^{n-i} \binom{n}{i+j}\binom{i+j}{i} (-1)^j\ m_{i+j} \\
&= \sum_{\ell=0}^{n} \sum_{i=0}^{\ell} \binom{n}{\ell}\binom{\ell}{i} (-1)^{\ell-i}\ m_\ell \quad \text{(changement de variable } \ell = i+j\text{)} \\
&= \sum_{\ell=0}^{n} \binom{n}{\ell}\ m_\ell\ \sum_{i=0}^{\ell} \binom{\ell}{i} (-1)^{\ell-i} \\
&= \sum_{\ell=0}^{n} \binom{n}{\ell}\ m_\ell\ (1-1)^\ell = \binom{n}{0} m_0 = 1.
\end{aligned}$$

Ainsi F_n est une fonction de répartition. □

Il s'agit à présent de calculer les moments de la variable aléatoire associée à F_n.

Définition 3.17 (Moments de F_n)**.** *On définit le k-ième moment de F_n noté $m_{n,k}$:*

$$\forall n \in \mathbb{N},\ \forall k \in \mathbb{N}, \quad m_{n,k} = \int_0^1 x^k dF_n(x).$$

Proposition 3.18. *On a les expressions :*

$$m_{n,k} = \sum_{i=0}^{n} \left(\frac{i}{n}\right)^k \binom{n}{i} (D^{n-i}m)_i = \sum_{\ell=0}^{n} \frac{1}{n^k} \binom{n}{\ell} m_\ell \left(\sum_{i=0}^{\ell} \binom{\ell}{i} i^k (-1)^{\ell-i} \right).$$

Démonstration. Remarquons que si une variable aléatoire X_n est à valeurs dans $\{0, 1/n, 2/n, \dots, 1\}$ et a pour probabilité $P(X_n = i/n) = \binom{n}{i}(\mathrm{D}^{n-i}m)_i$ alors sa fonction de répartition est F_n. De plus, comme F_n est une fonction de répartition, alors il existe une telle variable aléatoire et puisque la fonction de répartition caractérise la loi, on peut associer à la fonction de répartition F_n la variable aléatoire X_n. Ainsi,

$$m_{n,k} = \sum_{i=0}^{n} \left(\frac{i}{n}\right)^k \binom{n}{i} (\mathrm{D}^{n-i}m)_i.$$

En remplaçant $(\mathrm{D}^{n-i}m)_i$ par sa formule explicite (3.13), on obtient à l'aide de l'identité (14) :

$$m_{n,k} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \left(\frac{i}{n}\right)^k \binom{n}{i} \binom{n-i}{j} (-1)^j\ m_{i+j} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} \left(\frac{i}{n}\right)^k \binom{n}{i+j} \binom{i+j}{i} (-1)^j\ m_{i+j}$$

et en effectuant le changement de variable $\ell = i + j$:

$$m_{n,k} = \sum_{\ell=0}^{n} \frac{1}{n^k} \binom{n}{\ell} m_\ell \left(\sum_{i=0}^{\ell} \binom{\ell}{i} i^k (-1)^{\ell-i} \right).$$

□

Proposition 3.19. *Les moments de F_n notés $m_{n,k}$ convergent quand n tend vers l'infini vers m_k.*

Démonstration. Notons $b_k^\ell := \sum_{i=0}^{\ell} \binom{\ell}{i} i^k (-1)^{\ell-i}$ de telle façon à avoir, d'après la proposition 3.18, l'égalité :

$$m_{n,k} = \sum_{\ell=0}^{n} \frac{1}{n^k} \binom{n}{\ell} m_\ell\, b_k^\ell\,.$$

Posons

$$f(x_1, \dots, x_k) = \sum_{i=0}^{\ell} \binom{\ell}{i} (-1)^{\ell-i}\, x_1^i \dots x_k^i = (x_1 \dots x_k - 1)^\ell$$

et remarquons que $b_k^\ell = \partial^k_{x_1 \dots x_k} f(1, \dots, 1)$.
De plus, en dérivant le membre de droite de l'égalité définissant f, on obtient que si $\ell > k$ alors $b_k^\ell = 0$. Ainsi,

$$m_{n,k} = \sum_{\ell=0}^{k} \frac{1}{n^k} \binom{n}{\ell} m_\ell\, b_k^\ell\,.$$

Or

$$\binom{n}{\ell} = \frac{n!}{\ell!\,(n-\ell)!} = \frac{n(n-1)\dots(n-\ell+1)}{\ell!} \underset{+\infty}{\sim} \frac{n^\ell}{\ell!}.$$

Donc si $\ell < k$ alors $\dfrac{1}{n^k}\dbinom{n}{\ell} \underset{+\infty}{\sim} \dfrac{n^{\ell-k}}{\ell!} \xrightarrow[n\to+\infty]{} 0.$

Ainsi tous les termes de la somme définissant $m_{n,k}$ tendent vers 0, sauf le k-ième :

$$\lim_{n\to+\infty} m_{n,k} = \lim_{n\to+\infty} \frac{1}{n^k}\binom{n}{k} m_k\, b_k^k = \frac{b_k^k}{k!}\, m_k.$$

Calculons b_k^k. On a :

$$b_k^k = \partial^k_{x_1\dots x_k}(x_1\dots x_k - 1)^k\big|_{x_1=\dots=x_k=1}.$$

Cette dérivée se met sous la forme

$$\partial^k_{x_1\dots x_k}(x_1\dots x_k-1)^k = g(x_1,\dots,x_k) + k!\,x_1\dots x_k$$

où $g(1,\dots,1) = 0$.
D'où $b_k^k = k!$ et on a alors montré : $\lim\limits_{n\to+\infty} m_{n,k} = m_k$. □

Remarque 3.20 (Heuristique d'une autre méthode de démonstration)**.**
On aurait pu démontrer ce résultat à l'aide du théorème d'extension de Marcel Riesz.
Ce dernier énonce que dans un espace vectoriel réel E, F un sous-espace vectoriel et K un cône convexe, si on a $E = F+K$ et une forme linéaire $\varphi : F \to \mathbb{R}$, K-positive : $\forall x \in K\cap F,\ \varphi(x) \geqslant 0$ alors cette dernière s'étend en un prolongement K-positif, noté $\psi : E \to \mathbb{R}$ vérifiant : $\psi|_F = \varphi$ et $\forall x\in K,\ \psi(x)\geqslant 0$.
Ainsi en prenant $\varphi : X^i \mapsto m_i$, on peut obtenir que φ s'étend en une forme linéaire positive sur l'espace de fonctions continues $\mathcal{C}([a,b])$. On peut ensuite utiliser les polynômes de Bernstein B_nf associés à la fonction $f : x \mapsto x^k$ et remarquer que d'après l'expression explicite 3.18 de $m_{n,k}$ et par continuité et définition de ψ on a $m_{n,k} = \varphi(B_nf(X)) = \psi((B_nf(X)) \to_n \psi(f(X)) = \psi(X^k) = m_k$.

Lemme 3.21 (Darboux–Froda)**.** *Les discontinuités d'une fonction monotone à valeurs réelles forment un ensemble au plus dénombrable.*

Démonstration. Quitte à considérer $-f$ on peut supposer f croissante. Étudions d'abord le cas où est f est définie sur un intervalle $[a,b]$. Notons pour $x \in]a,b[$, $\delta(x) = \lim\limits_{y\to x^+} f(y) - \lim\limits_{y\to x^-} f(y)$ le saut de f en x.

Montrons que pour $n\in\mathbb{N}^*$, $E_n = \{x \in]a,b[;\ \delta(x) > 1/n\}$ est fini.
Soit $x_1 < \dots < x_p$ des éléments distincts de E_n et prenons $y_0,\dots,y_p$ des réels vérifiant

$$a \leqslant y_0 < x_1 < y_1 < x_2 < \dots < x_p < y_p \leqslant b.$$

Alors, par définition de E_n, et puisque la fonction f est croissante, on sait que $f(y_i) - f(y_{i-1}) \geqslant \frac{1}{n}$ pour tout $i = 1,\dots,p$. On en déduit que

$$f(b) - f(a) \geqslant f(y_p) - f(y_0) = \sum_{i=1}^p f(y_i) - f(y_{i-1}) \geqslant \frac{p}{n}.$$

On a donc $p \leqslant n\big(f(b) - f(a)\big)$ ce qui prouve bien que le cardinal de E_n est fini. L'ensemble des points de discontinuité de f est contenu dans la réunion des E_n pour n parcourant $n \in \mathbb{N}^*$, plus éventuellement a et/ou b. S'écrivant comme réunion dénombrable d'ensembles finis, cet ensemble est au plus dénombrable. Notons Ω l'ensemble des points de discontinuité de f dans $\mathbb{R}$ et Ω_n l'ensemble des points de discontinuité de f situés dans le segment $[-n, n]$. Alors puisque $\Omega = \bigcup_{n\in\mathbb{N}^*} \Omega_n$ et puisque chaque Ω_n est au plus dénombrable, il en est de même de Ω. $\square$

Théorème 3.22 (Théorème de Helly–Prokhorov)**.** *Soit $(\mu_n)_n$ une suite de probabilité de $[0,1]$ et $(F_n)_n$ la suite de fonctions de répartition associée. Alors il existe une extraction $(\alpha_n)_n$ et une fonction de répartition $F : [0,1] \to [0,1]$ telle qu'en tout point de continuité x de F, on a :*

$$F_{\alpha_n}(x) \xrightarrow[n\to+\infty]{} F(x).$$

Démonstration. Soit $q \in \mathbb{Q}$, la suite $(F_n(q))_n$ est une suite de $[0,1]$, on peut donc extraire une sous-suite qui converge vers un réel qu'on note $\tilde{F}(q) \in [0,1]$. Par extraction diagonale, il existe une extraction $(\alpha_n)_n$ telle que :

$$\forall q \in \mathbb{Q}, \quad F_{\alpha_n}(q) \xrightarrow[n\to+\infty]{} \tilde{F}(q).$$

Alors on peut définir, si $x \in \mathbb{R}$:

$$F(x) := \inf_{q\in\mathbb{Q}, x<q} \tilde{F}(q).$$

F est croissante par construction. Montrons qu'elle est continue à droite. Soient $x \in \mathbb{R}$ et $\varepsilon > 0$. Il existe $q > x$ tel que $\tilde{F}(q) \leqslant F(x) + \varepsilon$. Ainsi, si $y \in [x, q]$, on a :

$$F(x) \leqslant F(y) \leqslant \tilde{F}(q) \leqslant F(x) + \varepsilon$$

ce qui prouve que F est continue à droite.

Soit maintenant x un point de continuité de F et $\varepsilon > 0$. Par continuité, il existe $y < x$ tel que :

$$F(x) - \varepsilon \leqslant F(y).$$

Soient $r, s \in \mathbb{Q}$ tels que : $y < r < x < s$ et $\tilde{F}(s) \leqslant F(x) + \varepsilon$. On a alors :

$$F(x) - \varepsilon \leqslant F(y) \leqslant \tilde{F}(r) \leqslant \tilde{F}(s) \leqslant F(x) + \varepsilon.$$

Or par définition on a :

$$F(x) - \varepsilon \leqslant \tilde{F}(r) = \lim_n F_{\alpha_n}(r) \leqslant \underline{\lim}_n F_{\alpha_n}(x) \leqslant \overline{\lim}_n F_{\alpha_n}(x) \leqslant \overline{\lim}_n F_{\alpha_n}(s) = \tilde{F}(s) \leqslant F(x) + \varepsilon.$$

En faisant tendre ε vers 0, on obtient $F_{\alpha_n}(x) \xrightarrow[n\to+\infty]{} F(x)$ en tout point de continuité de F.

Montrons que $F(0) = 0$ et $F(1) = 1$. Fixons $\varepsilon > 0$, il existe α, β tels que $[\alpha, \beta] \subset [0,1]$, on peut les supposer comme points de continuité de F puisque les discontinuités d'une fonction monotone forment un ensemble au plus dénombrable, tels que :

$$\forall n \in \mathbb{N}, \quad \mu_n([\alpha, \beta]) \geqslant 1 - \varepsilon.$$

En passant à la limite, on obtient :

$$F(\beta) - F(\alpha) \geqslant 1 - \varepsilon$$

puis par croissance de F on a : $1 \geqslant F(1) \geqslant F(\beta) \geqslant 1 - \varepsilon$ et $\varepsilon \geqslant F(\alpha) \geqslant F(0) \geqslant 0$. En faisant tendre ε vers 0 on obtient alors que $F(1) = 1$ et $F(0) = 0$. Cela conclut que F est une fonction de répartition. $\square$

Démontrons maintenant l'objet principal de cette partie, à savoir la condition suffisante du théorème d'existence (3.15).

Démonstration du théorème 3.15 d'existence.
Soit F_n la fonction de répartition précédemment construite (3.16). Le théorème 3.22 de Helly-Prokhorov fournit une sous suite (n_i) telle que F_{n_i} converge faiblement vers une fonction de répartition F.
Soit F est l'unique valeur d'adhérence de (F_n), soit il en existe une autre notée G.
S'il existe une autre valeur d'adhérence, on a par définition de la convergence faible ($x \mapsto x^k$ est continue) :

$$\int_0^1 x^k \, \mathrm{d}F(x) = \lim_{n_i \to +\infty} \int_0^1 x^k \, \mathrm{d}F_{n_i}(x) = \lim_{n_i \to +\infty} m_{n_i,k} = m_k \quad \text{(d'après la proposition 3.19)} \tag{15}$$

et

$$\int_0^1 x^k \, \mathrm{d}G(x) = \lim_{n_j \to +\infty} \int_0^1 x^k \, \mathrm{d}G_{n_j}(x) = \lim_{n_j \to +\infty} m_{n_j,k} = m_k$$

soit

$$\int_0^1 x^k \, \mathrm{d}F(x) = \int_0^1 x^k \, \mathrm{d}G(x).$$

D'après le théorème 3.7 d'unicité, puisque F et G ont mêmes moments, $F = G$. Donc F_n converge faiblement vers une fonction de répartition F ayant pour moments m_k d'après l'équation (15) c'est à dire que la variable aléatoire associée à F a pour moments m_k, ce qui répond au problème de Hausdorff. □

On a au passage démontré le théorème suivant :

Théorème 3.23. *Soit $(X_n)_n$ une suite de variables aléatoires ayant pour support un segment $[a, b]$ et admettant des moments à tout ordre. Soit X une variable aléatoire sur $[a, b]$ admettant également des moments à tout ordre et telle que la loi de X est caractérisée par ses moments. On suppose que :*

$$\forall k \in \mathbb{N}, \quad m_{n,k} := \mathbb{E}(X_n^k) \xrightarrow[n \to +\infty]{} \mathbb{E}(X^k) =: m_k.$$

Alors $(X_n)_n$ converge en loi vers X.

Démonstration. Soit une suite de variables aléatoires $(X_n)_n$ de moments $m_{n,k}$ et de fonctions de répartition F_n telle que $\lim_{n \to +\infty} m_{n,k} = m_k$. D'après la démonstration précédente, on a prouvé qu'il existe une unique fonction de répartition F de moments m_k étant l'unique valeur d'adhérence de F_n lorsque F est caractérisée par ses moments, ce qui est le cas par hypothèse. Ainsi, F est par unicité la fonction de répartition de X et de plus F_n converge vers F ce qui prouve que $(X_n)_n$ converge en loi vers X. □

Remarque 3.24. *Le théorème 3.23 reste vrai même si les variables aléatoires n'ont pas pour support un segment.*

Remarque 3.25. *Dans la continuité de la remarque 3.20, il est également possible d'utiliser le théorème de représentation de Riesz-Markov à la place des polynômes de Bernstein, en effet, il énonce que l'existence d'un tel prolongement positif, évoqué dans la remarque 3.20, pour φ équivaut à l'existence d'une mesure μ telle que pour tout polynôme P, $\varphi(P) = \int P \, \mathrm{d}\mu$, en particulier pour $P = X^i$ ce qui fournit directement l'existence de μ telle que $m_i = \varphi(X^i) = \int x^i \, \mathrm{d}\mu(x)$.*

4 Problème de Stieltjes

Le problème de Stieltjes est le problème des moments lorsque l'intervalle est $[0, +\infty[$.
On donne dans cette sous-partie des résultats sans démonstrations.

Théorème 4.1 (Existence)**.** *Soit $(m_n)_n$ une suite réelle. $(m_n)_n$ est la suite des moments d'une mesure positive sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ portée par $\mathbb{R}_+$ si et seulement si :*

$$\forall N \in \mathbb{N},\ \forall \beta_0, \ldots, \beta_N \in \mathbb{C}, \qquad \sum_{k=0}^{N}\sum_{l=0}^{N} \beta_k \bar{\beta}_l m_{k+l} \geqslant 0 \quad \textit{et} \quad \sum_{k=0}^{N}\sum_{l=0}^{N} \beta_k \bar{\beta}_l m_{k+l+1} \geqslant 0.$$

Théorème 4.2 (Unicité)**.** *Si $m_0 = 1$, $m_n \geqslant 0$ pour tout $n \in \mathbb{N}$ et si $\sum_{n=1}^{\infty} m_n^{-1/n} = \infty$ alors le problème de Stieltjes associé à $(m_n)_n$ a au plus une solution.*

Remarque 4.3. *On avait fournit des conditions suffisantes d'unicité dans la section 2 Condition suffisante d'unicité de Fourier, elles sont parfois plus pratiques.*

On va encore énoncer un théorème très utile puisqu'il donne des conditions suffisantes au fait que la loi soit ou non déterminée par ses moments.

Théorème 4.4 (Conditions de Krein et de Lin sur $\mathbb{R}_+$)**.** *Soit μ une mesure de probabilité définie par une densité strictement positive f sur $\mathbb{R}_+$.*

1. *Si la condition suivante, dite de Krein sur $\mathbb{R}_+$,*

$$\int_0^{+\infty} \frac{-\ln f(t^2)}{1+t^2}\,\mathrm{d}t \ < \infty$$

est vérifiée alors μ n'est pas déterminée par ses moments.

2. *Si au contraire*

$$\int_0^{+\infty} \frac{-\ln f(t^2)}{1+t^2}\,\mathrm{d}t \ = \infty$$

et si de plus f vérifie la condition de Lin sur $\mathbb{R}_+$, à suivre, alors μ est déterminée par ses moments. La condition de Lin est : f est dérivable et, asymptotiquement, $-xf'(x)/f(x)$ tend en croissant vers $+\infty$ lorsque x tend vers $+\infty$.

Exemple 4.5 (Application à la loi Log-Normale)**.** *La loi Log-Normale n'est pas caractérisée par ses moments.*

Démonstration. Rappelons que la loi log normale a pour densité $f : x \mapsto \frac{1}{\sqrt{2\pi}\,x} \mathrm{e}^{-(\ln(x))^2/2}\ \mathbb{1}_{x>0}$ et est à support dans $\mathbb{R}_+$.
On a

$$\ln(f(x^2)) = \ln\left(\frac{1}{\sqrt{2\pi}\,x^2}\mathrm{e}^{-(\ln(x^2))^2/2}\right) = \ln\left(\frac{1}{\sqrt{2\pi}\,x^2}\mathrm{e}^{-2(\ln(x))^2}\right) = -\ln(\sqrt{2\pi}\,x^2) - 2(\ln(x))^2$$

soit

$$\forall t > 0, \qquad g(t) := \frac{-\ln(f(t^2))}{1+t^2} = \frac{\ln(\sqrt{2\pi}\,t^2)}{1+t^2} + \frac{2(\ln(t))^2}{1+t^2}.$$

Or $\ln(\sqrt{2\pi}\,t^2) + 2(\ln(t))^2 = o(\sqrt{t})$ en $+\infty$, donc $-\ln(f(t^2))/(1+t^2) = o\left(\sqrt{t}/(1+t^2)\right) = o(t^{-3/2})$ et $t \mapsto t^{-3/2}$ est intégrable au voisinage de $+\infty$, ce qui prouve que g est intégrable au voisinage de l'infini.

De plus, au voisinage de 0, $\ln(\sqrt{2\pi}\,t^2) + 2(\ln(t))^2 = o(1/\sqrt{t})$ car $\sqrt{t}\,(\ln(\sqrt{2\pi}\,t^2) + 2(\ln(t))^2)$ tend vers 0 en 0. Or $t \mapsto 1/\sqrt{t}$ est intégrable au voisinage de 0, ce qui prouve que g est intégrable au voisinage de 0.

Ainsi, puisque g est continue sur $]0,+\infty[$ et intégrable en 0 et en $+\infty$, g est intégrable.

D'après le premier point du théorème 4.4 de Krein, la loi Log-Normale n'est pas caractérisée par ses moments. □

5 Problème de Hamburger

Le problème de Hamburger est le problème des moments lorsque l'intervalle est $\mathbb{R}$ tout entier.
On donne dans cette sous-partie des résultats sans démonstrations.

Théorème 5.1 (Existence). *Soit $(m_n)_n$ une suite réelle. $(m_n)_n$ est la suite des moments d'une mesure positive sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ si et seulement si :*

$$\forall N \in \mathbb{N},\ \forall \beta_0, \ldots, \beta_N \in \mathbb{C}, \qquad \sum_{k=0}^{N} \sum_{l=0}^{N} \beta_k \bar{\beta}_l m_{k+l} \geqslant 0.$$

Théorème 5.2 (Unicité). *Si $m_0 = 1$, $m_{2n} \geqslant 0$ pour tout $n \in \mathbb{N}$ et si $\sum_{n=1}^{\infty} m_{2n}^{-1/2n} = \infty$ alors le problème de Hamburger associé à $(m_n)_n$ a au plus une solution.*

Théorème 5.3 (Conditions de Krein et de Lin sur $\mathbb{R}$). *Soit μ une mesure de probabilité définie par une densité strictement positive f sur $\mathbb{R}$.*

1. *Si la condition suivante, dite de Krein sur $\mathbb{R}_+$,*
$$\int_{-\infty}^{+\infty} \frac{-\ln f(t)}{1+t^2}\,\mathrm{d}t \ < \infty$$
est vérifiée alors μ n'est pas déterminée par ses moments.
2. *Si au contraire*
$$\int_{-\infty}^{+\infty} \frac{-\ln f(t)}{1+t^2}\,\mathrm{d}t \ = \infty$$
et si de plus f vérifie la condition de Lin sur $\mathbb{R}$ alors μ est déterminée par ses moments.
La condition de Lin est : f est dérivable et, asymptotiquement, $-xf'(x)/f(x)$ tend en croissant vers $+\infty$ lorsque x tend vers $+\infty$.

Exemple 5.4 (Loi Normale). *La loi Normale est caractérisée par ses moments.*

Démonstration. Soit $f : x \mapsto \mathrm{e}^{-x^2/2}/\sqrt{2\pi}$ la densité de la loi Normale à support $\mathbb{R}$.
On a $-\ln(f(t)) = t^2/2 + \ln(\sqrt{2\pi})$ ainsi $-\ln(f(t))/(1+t^2)$ n'est pas intégrable au voisinage de l'infini car équivalent à $t^2/(t^2+1)$, équivalent à 1.

De plus, f est dérivable, de dérivée $f'(t) = -t\mathrm{e}^{-t^2/2}/\sqrt{2\pi} = -tf(t)$. Ainsi, $-tf'(t)/f(t) = t^2$ et tend en croissant vers $+\infty$ lorsque x tend vers $+\infty$.

Par conséquent, les conditions de Krein et de Lin sont vérifiées, on en conclut que la loi Normale est caractérisée par ses moments. □

Deuxième partie
Le théorème de Berry–Esseen

Le théorème de Berry–Esseen fournit un contrôle sur la vitesse de convergence uniforme des fonctions de répartition vers la fonction de répartition normale dans le théorème central limite. Pour plus de commodité, il est énoncé pour des variables aléatoires centrées.

Théorème 5.5 (Berry–Esseen)**.** *Soit $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes identiquement distribuée de moyenne nulle $\mathbb{E}(X) = 0$, de variance $Var(X) = \mathbb{E}(X^2) = \sigma^2 > 0$, et admettant un moment absolu d'ordre 3, $\mathbb{E}(|X|^3) < \infty$. On pose $S_n = X_1 + \cdots + X_n$ pour $n \geqslant 1$.*
Alors, pour tout $n \geqslant 1$,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2}\ \mathrm{d}x \right| \leqslant \frac{C\ \mathbb{E}(|X|^3)}{\sigma^3\sqrt{n}}$$

où C désigne une constante universelle strictement positive.

6 Une démonstration par couplage

Lemme 6.1 (Inégalité de couplage de Lindeberg)**.** *Soient $X_1, Y_1, \dots, X_n, Y_n$ des variables aléatoires indépendantes telles que $\mathbb{E}(|X_k|^3) < +\infty$ et $Y_k \sim \mathcal{N}(\mathbb{E}(X_k), Var(X_k))$ pour tout $1 \leqslant k \leqslant n$. Alors pour toute fonction $f \in \mathcal{C}^3(\mathbb{R})$, avec f, f', f'' et $f^{(3)}$ bornées, on obtient en posant $\tau_k^3 = \mathbb{E}(|X_k - \mathbb{E}(X_k)|^3)$:*

$$|\mathbb{E}\left(f\left(X_1 + \cdots + X_n\right)\right) - \mathbb{E}\left(f\left(Y_1 + \cdots + Y_n\right)\right)| \leqslant \frac{\tau_1^3 + \cdots + \tau_n^3}{2} \|f^{(3)}\|_\infty.$$

Démonstration. Quitte à translater f on peut se placer dans le cas où $\mathbb{E}[X_k] = 0$ pour tout $1 \leqslant k \leqslant n$.
Fixons $n \geqslant 1$ et posons $Z_k := X_1 + \cdots + X_{k-1} + Y_{k+1} + \cdots + Y_n$ pour tout $1 \leqslant k \leqslant n$.
Notons :

$$S_k := \sum_{i=1}^{k} X_k \quad \text{et} \quad P_k := \sum_{i=k}^{n} Y_i$$

avec pour convention que $S_0 = 0$ et $P_{n+1} = 0$. Ainsi $Z_k + X_k = S_k + P_{k+1}$ et $Z_k + Y_k = S_{k-1} + P_k$.

Effectuons une transformation d'Abel :

$$\begin{aligned}
\sum_{k=1}^{n} \mathbb{E}\left(f(Z_k + X_k) - f(Z_k + Y_k)\right) &= \sum_{k=1}^{n} \mathbb{E}\left(f(S_k + P_{k+1}) - f(S_{k-1} + P_k)\right) \\
&= \sum_{k=1}^{n} \mathbb{E}\left(f(S_k + P_{k+1})\right) - \sum_{k=0}^{n-1} \mathbb{E}\left(f(S_k + P_{k+1})\right) \\
&= \mathbb{E}\left(f(S_n + P_{n+1}))\right) - \mathbb{E}\left(f(S_0 + P_1))\right) \\
&= \mathbb{E}\left(f\left(X_1 + \cdots + X_n\right)\right) - \mathbb{E}\left(f\left(Y_1 + \cdots + Y_n\right)\right)
\end{aligned}$$

La formule de Taylor-Lagrange appliquée à f à l'ordre 2 en Z_k donne :

$$\left| f(Z_k + X_k) - f(Z_k) - f'(Z_k)X_k - f''(Z_k)\frac{X_k^2}{2!} \right| \leqslant \frac{|X_k|^3}{3!} \|f^{(3)}\|_\infty \tag{16}$$

et

$$\left| f(Z_k + Y_k) - f(Z_k) - f'(Z_k)Y_k - f''(Z_k)\frac{Y_k^2}{2!} \right| \leqslant \frac{|Y_k|^3}{3!} \, \|f^{(3)}\|_\infty \tag{17}$$

$$\begin{aligned}\mathbb{E}\left(f(Z_k + X_k) - f(Z_k + Y_k)\right) = \mathbb{E}&\left(f(Z_k + X_k) - f(Z_k) - f'(Z_k)X_k - f''(Z_k)\frac{X_k^2}{2!} \right) \\ &+ \mathbb{E}\left(f(Z_k) + f'(Z_k)X_k + f''(Z_k)\frac{X_k^2}{2!} - f(Z_k + Y_k) \right).\end{aligned} \tag{18}$$

Or puisque Z_k, X_k et Y_k sont indépendants et que $\mathbb{E}(Y_k) = \mathbb{E}(X_k)$ et $Var(Y_k) = Var(X_k)$:

$$\begin{aligned}\mathbb{E}\left(f'(Z_k)X_k + f''(Z_k)\frac{X_k^2}{2!} \right) &= \mathbb{E}\left(f'(Z_k)\right)\mathbb{E}(X_k) + \mathbb{E}\left(f''(Z_k)\right)\frac{\mathbb{E}\left(X_k^2\right)}{2!} \\ &= \mathbb{E}\left(f'(Z_k)\right)\mathbb{E}(Y_k) + \mathbb{E}\left(f''(Z_k)\right)\frac{\mathbb{E}\left(Y_k^2\right)}{2!} \\ &= \mathbb{E}\left(f'(Z_k)Y_k + f''(Z_k)\frac{Y_k^2}{2!} \right).\end{aligned}$$

Ainsi on obtient à l'aide des inégalités (16), (17), l'égalité (18) et ce qui précède, les inégalités :

$$\begin{aligned}|\mathbb{E}\left(f(Z_k + X_k) - f(Z_k + Y_k)\right)| &\leqslant \left| \mathbb{E}\left(f(Z_k + X_k) - f(Z_k) - f'(Z_k)X_k - f''(Z_k)\frac{X_k^2}{2!} \right) \right| \\ &\quad + \left| \mathbb{E}\left(f(Z_k) + f'(Z_k)X_k + f''(Z_k)\frac{X_k^2}{2!} - f(Z_k + Y_k) \right) \right| \\ &= \left| \mathbb{E}\left(f(Z_k + X_k) - f(Z_k) - f'(Z_k)X_k - f''(Z_k)\frac{X_k^2}{2!} \right) \right| \\ &\quad + \left| \mathbb{E}\left(f(Z_k) + f'(Z_k)Y_k + f''(Z_k)\frac{Y_k^2}{2!} - f(Z_k + Y_k) \right) \right| \\ &\leqslant \mathbb{E}\left(\left| f(Z_k + X_k) - f(Z_k) - f'(Z_k)X_k - f''(Z_k)\frac{X_k^2}{2!} \right| \right) \\ &\quad + \mathbb{E}\left(\left| f(Z_k + Y_k) - f(Z_k) - f'(Z_k)Y_k - f''(Z_k)\frac{Y_k^2}{2!} \right| \right) \\ &\leqslant \frac{\mathbb{E}(|X_k|^3 + |Y_k|^3)}{3!} \, \|f^{(3)}\|_\infty.\end{aligned}$$

Dans la suite, nous utiliserons la notation $\mathbb{E}(|\mathcal{N}(0,1)|^3)$ pour désigner l'espérance de la valeur absolue du cube d'une variable aléatoire suivant une loi normale centrée réduite.
Notons de plus $\sigma(X_k) = \sqrt{Var(X_k)} = \sqrt{\mathbb{E}(X_k^2)}$ car X_k est centré. Ainsi on a $Y_k = \sigma(X_k)\,\mathcal{N}(0,1)$ et alors $\mathbb{E}(|Y_k|^3) = \sigma(X_k)^3\,\mathbb{E}(|\mathcal{N}(0,1)|^3)$. Or d'après la remarque 2.12, on a l'égalité $\mathbb{E}(|\mathcal{N}(0,1)|^3) = 2\,\mathbb{E}(|\mathcal{N}(0,1)|)$. De plus,

$$\begin{aligned}\sqrt{2\pi}\,\mathbb{E}(|\mathcal{N}(0,1)|) &= \int_{\mathbb{R}} |x| \; \mathrm{e}^{-x^2/2} \; \mathrm{d}x = \int_0^{+\infty} x \; \mathrm{e}^{-x^2/2} \; \mathrm{d}x - \int_{-\infty}^0 x \; \mathrm{e}^{-x^2/2} \; \mathrm{d}x \\ &= \left[-\mathrm{e}^{-x^2/2}\right]_0^{+\infty} - \left[-\mathrm{e}^{-x^2/2}\right]_{-\infty}^0 = 2.\end{aligned}$$

Ce qui fournit :

$$\mathbb{E}(|Y_k|^3) = \sigma(X_k)^3\ \mathbb{E}(|\mathcal{N}(0,1)|^3) = \sigma(X_k)^3\ \mathbb{E}(|\mathcal{N}(0,1)|^3) = \frac{4}{\sqrt{2\pi}}\,\sigma(X_k)^3 \leqslant 2\,\sigma(X_k)^3 = 2\,\mathbb{E}(|X_k|^2)^{3/2}.$$

Or d'après l'inégalité de Jensen, pour ϕ une fonction convexe on a : $\phi(\mathbb{E}(X)) \leqslant \mathbb{E}(\phi(X))$.
Alors pour $\phi(x) = x^{3/2}$ on obtient :

$$\sigma(X_k)^3 = \mathbb{E}(|X_k^2|)^{3/2} = \phi(\mathbb{E}(|X_k^2|)) \leqslant \mathbb{E}(\phi(|X_k|^2)) = \mathbb{E}(|X_k|^3).$$

D'où

$$\mathbb{E}(|X_k|^3 + |Y_k|^3) \leqslant \mathbb{E}(|X_k|^3) + 2\,\mathbb{E}(|X_k|^2)^{3/2} \leqslant \mathbb{E}(|X_k|^3) + 2\,\mathbb{E}(|X_k|^3) = 3\,\mathbb{E}(|X_k|^3).$$

Ainsi

$$|\mathbb{E}\left(f(Z_k + X_k) - f(Z_k + Y_k)\right)| \leqslant \frac{\mathbb{E}(|X_k|^3 + |Y_k|^3)}{3!}\,\|f^{(3)}\|_\infty \leqslant \frac{\mathbb{E}(|X_k|^3)}{2}\,\|f^{(3)}\|_\infty = \frac{\tau_k}{2}\,\|f^{(3)}\|_\infty.$$

En sommant de 1 à n on obtient avec ce qui précède :

$$\begin{aligned}
|\mathbb{E}\left(f\left(X_1 + \cdots + X_n\right)\right) - \mathbb{E}\left(f\left(Y_1 + \cdots + Y_n\right)\right)| &= \left|\sum_{k=1}^n \mathbb{E}\left(f(Z_k + X_k) - f(Z_k + Y_k)\right)\right| \\
&\leqslant \sum_{k=1}^n |\mathbb{E}\left(f(Z_k + X_k) - f(Z_k + Y_k)\right)| \\
&\leqslant \sum_{k=1}^n \frac{\tau_k}{2}\,\|f^{(3)}\|_\infty
\end{aligned}$$

ce qui achève la démonstration. □

Corollaire 6.2. *Soient $X_1, \dots, X_n$ des variables aléatoires indépendantes et identiquement distribuées telles que $\mathbb{E}(|X_k|^3) < +\infty$. Alors pour toute fonction $f \in \mathcal{C}^3(\mathbb{R})$, avec f, f', f'' et $f^{(3)}$ bornées, on a :*

$$\left|\mathbb{E}\left(f\left(\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}\,\sigma(X_1)}\right)\right) - \mathbb{E}\left(f\left(Y\right)\right)\right| \leqslant \frac{\mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)}{2\sqrt{n}\,\sigma(X_1)^3}\,\|f^{(3)}\|_\infty.$$

Démonstration. Considérons des variables aléatoires X_k, $(1 \leqslant k \leqslant n)$, indépendantes et identiquement distribuées ayant un moment absolu d'ordre 3 fini. Définissons ensuite les variables aléatoires centrées $Z_k := (X_k - \mathbb{E}(X_1))/(\sqrt{n}\,\sigma(X_1))$. Alors les Z_k ont les mêmes propriétés que les X_k énoncées ci-dessus. Par conséquent, en appliquant l'inégalité de couplage de Lindeberg (6.1) aux variables Z_k, et aux variables aléatoires indépendantes $Y_k \sim \mathcal{N}(\mathbb{E}(Z_k), Var(Z_k)) = \mathcal{N}(0, 1/n)$ on obtient pour $f \in \mathcal{C}^3 \cap L^\infty$ ayant ses trois premières dérivées bornées :

$$|\mathbb{E}\left(f\left(Z_1 + \cdots + Z_n\right)\right) - \mathbb{E}\left(f\left(Y_1 + \cdots + Y_n\right)\right)| \leqslant \frac{\tau_1^3 + \cdots + \tau_n^3}{2}\,\|f^{(3)}\|_\infty$$

et en notant $S_n = \sum_{k=1}^n X_k$ et $Y = \sum_{k=1}^n Y_k \sim \mathcal{N}(0,1)$ puisque les Y_k sont indépendantes :

$$\left|\mathbb{E}\left(f\left(\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}\,\sigma(X_1)}\right)\right) - \mathbb{E}\left(f\left(Y\right)\right)\right| \leqslant \frac{\tau_1^3 + \cdots + \tau_n^3}{2}\,\|f^{(3)}\|_\infty. \tag{19}$$

Or

$$\tau_k = \mathbb{E}(|Z_k - \mathbb{E}(Z_k)|^3) = \mathbb{E}(|Z_k|^3) = \frac{\mathbb{E}(|X_k - \mathbb{E}(X_k)|^3)}{n^{3/2}\,\sigma(X_k)^3} = \frac{\mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)}{n^{3/2}\,\sigma(X_1)^3}.$$

Ce qui donne dans (19) :

$$\left|\mathbb{E}\left(f\left(\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}\,\sigma(X_1)}\right)\right) - \mathbb{E}\left(f\left(Y\right)\right)\right| \leqslant \frac{\mathbb{E}(|X_1 - \mathbb{E}(X_1)|^3)}{2\sqrt{n}\,\sigma(X_1)^3}\,\|f^{(3)}\|_\infty.$$

□

Démontrons une version faible du théorème de Berry-Esseen à l'aide du couplage de Lindeberg :

Théorème 6.3 (Berry–Esseen, version faible)**.** *Soit $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes identiquement distribuée de moyenne nulle $\mathbb{E}(X) = 0$, de variance $Var(X) = \mathbb{E}(X^2) = \sigma^2 > 0$, et admettant un moment absolu d'ordre 3, $\mathbb{E}(|X|^3) < \infty$. On pose $S_n = X_1 + \cdots + X_n$ pour $n \geqslant 1$. Alors, pour tout $n \geqslant 1$,*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \mathrm{e}^{-x^2/2}\,\mathrm{d}x\right| \leqslant \left(\frac{8\,\mathbb{E}(|X|^3)}{2\sqrt{n}\,\sigma^3}\right)^{1/4}.$$

Démonstration. Plaçons nous dans le cadre des hypothèses énoncées ci-dessus. Pour tout $a \in \mathbb{R}$, on pose $I_a =]-\infty, a[$. Pour tout $\varepsilon > 0$, on peut construire $f_{a,\varepsilon} \in \mathcal{C}^3(\mathbb{R},\mathbb{R})$ telle que $f_{a,\varepsilon}, f'_{a,\varepsilon}, f''_{a,\varepsilon}$ et $f^{(3)}_{a,\varepsilon}$ soient bornées et que $\mathbb{1}_{a-\varepsilon} \leqslant f_{a,\varepsilon} \leqslant \mathbb{1}_{a+\varepsilon}$ et $\|f^{(3)}_{a,\varepsilon}\|_\infty \leqslant \varepsilon^{-3}$. Cela fournit d'après (19) :

$$\left|\mathbb{P}\left(\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}\,\sigma(X_1)} \leqslant a\right) - \mathbb{E}\left(f_{a,\varepsilon}\left(Y\right)\right)\right| \leqslant \frac{\mathbb{E}(|X_1|^3)}{2\sqrt{n}\,\sigma(X_1)^3\,\varepsilon^3}. \tag{20}$$

De plus par croissance de l'espérance on a par construction de $f_{a,\varepsilon}$:

$$\mathbb{E}(\mathbb{1}_{a-\varepsilon}(Y) - \mathbb{1}_a(Y)) \leqslant \mathbb{E}\left(f_{a,\varepsilon}\left(Y\right) - \mathbb{1}_a(Y)\right) \leqslant \mathbb{E}(\mathbb{1}_{a+\varepsilon}(Y) - \mathbb{1}_a(Y)).$$

Or

$$|\mathbb{E}(\mathbb{1}_\alpha(Y) - \mathbb{1}_\beta(Y))| = \left|\int_{-\infty}^{\alpha} \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x - \int_{-\infty}^{\beta} \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x\right| = \left|\int_{\beta}^{\alpha} \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x\right| \leqslant \frac{1}{\sqrt{2\pi}}|\alpha - \beta| \leqslant |\alpha - \beta|$$

ce qui donne :

$$|\mathbb{E}\left(f_{a,\varepsilon}\left(Y\right) - \mathbb{1}_a(Y)\right)| \leqslant \varepsilon. \tag{21}$$

L'inégalité triangulaire permet d'obtenir à l'aide de (20) et (21) :

$$\left|\mathbb{P}\left(\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}\,\sigma(X_1)} \leqslant a\right) - \mathbb{P}\left(Y \leqslant a\right)\right| \leqslant \frac{\mathbb{E}(|X_1|^3)}{2\sqrt{n}\,\sigma(X_1)^3\,\varepsilon^3} + \varepsilon. \tag{22}$$

Notons $b := \mathbb{E}(|X_1|^3)/(2\sqrt{n}\sigma(X_1)^3)$. Optimisons le membre de droite de (22) en posant $\varepsilon = b^{1/4}$, on obtient alors : $\varepsilon + b\varepsilon^{-3} = b\,b^{-3/4} + b^{1/4} = 2b^{1/4}$.
Ainsi, on obtient l'inégalité :

$$\left|\mathbb{P}\left(\frac{S_n - n\mathbb{E}(X_1)}{\sqrt{n}\,\sigma(X_1)} \leqslant a\right) - \mathbb{P}\left(Y \leqslant a\right)\right| \leqslant 2\left(\frac{\mathbb{E}(|X_1|^3)}{2\sqrt{n}\,\sigma(X_1)^3}\right)^{1/4} = \left(\frac{8\,\mathbb{E}(|X_1|^3)}{2\sqrt{n}\,\sigma(X_1)^3}\right)^{1/4}$$

ce qui prouve le théorème de Berry-Esseen faible puisque l'inégalité est valable pour tout $a \in \mathbb{R}$. □

Remarque 6.4. *On a montré que la vitesse de convergence est d'au moins $n^{1/8}$, en réalité c'est $\sqrt{n}$.*

7 Une démonstration par transformation de Fourier

Théorème 7.1 (Berry–Esseen)**.** *Soit $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes identiquement distribuées de moyenne nulle $\mathbb{E}(X) = 0$, de variance $Var(X) = \mathbb{E}(X^2) = \sigma^2 > 0$, et admettant un moment absolu d'ordre 3, $\mathbb{E}(|X|^3) < \infty$. On pose $S_n = X_1 + \cdots + X_n$ pour $n \geqslant 1$.*
Alors, pour tout $n \geqslant 1$,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2} \, \mathrm{d}x \right| \leqslant \frac{C\ \mathbb{E}(|X|^3)}{\sigma^3\sqrt{n}}$$

où C désigne une constante universelle strictement positive.

7.1 Formule d'inversion

Théorème 7.2 (Formule d'inversion)**.** *Soit $\varphi : t \mapsto \int_{\mathbb{R}} \mathrm{e}^{itx}\ \mu(dx)$ où μ est une mesure de probabilité. Alors si $a < b$,*

$$\frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it} \varphi(t) \ \mathrm{d}t = \mu([a,b]) + \frac{1}{2}\mu(\{a,b\}).$$

Démonstration. Soit

$$I_T := \int_{-T}^{T} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it} \varphi(t) \ \mathrm{d}t = \int_{-T}^{T} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it} \int_{\mathbb{R}} \mathrm{e}^{itx}\ \mu(\mathrm{d}x) \ \mathrm{d}t.$$

Or

$$\left| \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it} \right| = \left| \int_a^b \mathrm{e}^{-ity} \ \mathrm{d}y \right| \leqslant |b-a|$$

et comme μ est une mesure de probabilité et que $[-T,T]$ est un intervalle, on peut utiliser le théorème de Fubini :

$$I_T = \int_{\mathbb{R}} \int_{-T}^{T} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it} \mathrm{e}^{itx} \ \mathrm{d}t \, \mu(\mathrm{d}x).$$

Remarquons que :

$$\begin{aligned} \frac{\mathrm{e}^{it(x-a)} - \mathrm{e}^{it(x-b)}}{it} &= \frac{\mathrm{e}^{it(x-a)} - \mathrm{e}^{-it(x-a)}}{2it} + \frac{\mathrm{e}^{it(x-a)} + \mathrm{e}^{-it(x-a)}}{2it} - \frac{\mathrm{e}^{it(x-b)} - \mathrm{e}^{-it(x-b)}}{2it} - \frac{\mathrm{e}^{it(x-b)} + \mathrm{e}^{-it(x-b)}}{2it} \\ &= \frac{\sin(t(x-a))}{t} + \frac{\cos(t(x-a))}{it} - \frac{\sin(t(x-b))}{t} - \frac{\cos(t(x-b))}{it} \end{aligned}$$

et

$$\int_{-T}^{T} \frac{\cos(t(x-a)) - \cos(t(x-b))}{it} \ \mathrm{d}t = 0$$

par imparité.
Ainsi

$$\int_{-T}^{T} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it} \mathrm{e}^{itx} \ \mathrm{d}t = \int_{-T}^{T} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} \ \mathrm{d}t.$$

Notons $R(\theta, T) := \int_{-T}^{T} \sin(\theta t)/t \ \mathrm{d}t$ et alors

$$I_T = \int_{\mathbb{R}} \left(R(x-a,T) - R(x-b,T)\right) \mu(\mathrm{d}x).$$

Notons également
$$S(T) := \int_0^T \frac{\sin(x)}{x}\ \mathrm{d}x.$$
On a pour $\theta > 0$,
$$R(\theta, T) = 2\int_0^T \frac{\sin(\theta t)}{t}\ \mathrm{d}t = 2\int_0^{\theta T} \frac{\sin(u)}{u}\ \mathrm{d}u = 2S(\theta T)$$
et si $\theta < 0$, $R(\theta, T) = -2S(-\theta T)$. Soit pour tout θ, $R(\theta, T) = 2\, sgn(\theta)\, S(T|\theta|)$.
Et quand $[T \to +\infty]$, $S(T) \to \pi/2$ (intégrale de Dirichlet) donc $R(\theta, T) \to \pi\, sgn(\theta)$.
Ainsi,
$$R(x-a, T) - R(x-b, T) \xrightarrow[T\to+\infty]{} \pi\,(sgn(x-a) - sgn(x-b)) = \begin{cases} 2\pi & \text{si } a < x < b \\ \pi & \text{si } x = a \text{ ou } x = b \\ 0 & \text{sinon} \end{cases}.$$
On a la majoration $|R(\theta, T)| = 2|S(T|\theta|)| \leqslant \sup_{y\in\mathbb{R}} |S(y)|$ qui est fini et intégrable car μ est une probabilité.
Cela nous permet d'appliquer le théorème de convergence dominée, et d'obtenir :
$$\begin{aligned} \int_{-\infty}^{+\infty} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it}\,\varphi(t)\ \mathrm{d}t = \lim_{T\to+\infty} I_T &= \int_{\mathbb{R}} \lim_{T\to+\infty} \left(R(x-a, T) - R(x-b, T)\right)\mu(\mathrm{d}x) \\ &= \int_{\mathbb{R}} 2\pi\, \mathbb{1}_{a<x<b} + \pi\, \mathbb{1}_{x=a\cup x=b}\, \mu(\mathrm{d}x) \\ &= 2\pi\, \mu\{[a,b]\} + \pi\, \mu\{a,b\}. \end{aligned}$$
soit le résultat voulu en divisant par 2π. $\square$

Corollaire 7.3. *Si $\int |\varphi| < +\infty$ alors μ admet une densité continue et bornée f telle que :*
$$\forall y \in \mathbb{R}, \qquad f(y) := \frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-ity}\,\varphi(t)\ \mathrm{d}t.$$

Démonstration. D'après le théorème d'inversion 7.2, on a pour $b > a$:
$$\mu([a,b]) + \frac{1}{2}\mu(\{a,b\}) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it}\,\varphi(t)\ \mathrm{d}t. \tag{23}$$
Or puisque
$$\left|\frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it}\right| = \left|\int_a^b \mathrm{e}^{-ity}\ \mathrm{d}y\right| \leqslant b - a$$
on a
$$\frac{1}{2\pi}\int_{\mathbb{R}} \left|\frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itb}}{it}\,\varphi(t)\right|\ \mathrm{d}t \ \leqslant\ \frac{(b-a)}{2\pi}\int_{\mathbb{R}} |\varphi(t)|\ \mathrm{d}t < +\infty$$
ce qui montre que l'intégrale de (23) converge absolument.
En faisant $[b \to a]$ on obtient que $\mu(\{a\}) = 0$ pour tout a. Ainsi μ ne charge pas les singletons et alors :
$$\begin{aligned} \mu([x, x+h]) &= \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\mathrm{e}^{-itx} - \mathrm{e}^{-it(x+h)}}{it}\,\varphi(t)\ \mathrm{d}t \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \left(\int_x^{x+h} \mathrm{e}^{-ity}\ \mathrm{d}y\right)\varphi(t)\ \mathrm{d}t \\ &= \int_x^{x+h} \left(\frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{-ity}\,\varphi(t)\ \mathrm{d}t\right)\ \mathrm{d}y \end{aligned}$$

d'après le théorème de Fubini. Cela montre que la mesure μ a une densité f telle que :

$$f(y) := \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-ity}\, \varphi(t)\ \mathrm{d}t.$$

De plus, f est uniformément bornée :

$$|f(y)| \leqslant \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi(t)|\ \mathrm{d}t$$

et est continue par application du théorème de convergence dominée. □

Lemme 7.4 (Riemann–Lebesgue)**.** *Soit* $f \in L^1(\mathbb{R})$*. On a :*

$$\lim_{x\to\infty} \int_{\mathbb{R}} f(t)\, \mathrm{e}^{ixt}\ \mathrm{d}t = 0.$$

Démonstration. On sait que les fonctions de classe $C^1(\mathbb{R})$ à support compact, noté $\mathcal{C}_c^1(\mathbb{R})$, sont denses dans $L^1(\mathbb{R})$. Soit alors f_n une fonction de $\mathcal{C}_c^1(\mathbb{R})$ convergeant dans $L^1(\mathbb{R})$ vers f.
Soit $g_n \in \mathcal{C}_c^1(\mathbb{R})$. Disons que g_n est à support dans $[a, b]$. Alors :

$$\int_{\mathbb{R}} g_n'(t)\ \mathrm{d}t = \int_a^b g_n'(t)\ \mathrm{d}t = g_n(b) - g_n(a) = 0.$$

Si $f_n \in \mathcal{C}_c^1(\mathbb{R})$ alors $g_n : t \mapsto f_n(t)\, \mathrm{e}^{-ixt} \in \mathcal{C}_c^1(\mathbb{R})$ pour tout x fixé, donc

$$0 = \int_{\mathbb{R}} g_n'(t)\ \mathrm{d}t = \int_{\mathbb{R}} f_n'(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t - ix \int_{\mathbb{R}} f_n(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t.$$

Soit :

$$\int_{\mathbb{R}} f_n'(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t = ix \int_{\mathbb{R}} f_n(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t$$

pour tout $x \in \mathbb{R}$.
De plus, le membre de gauche est borné, en effet :

$$\left| \int_{\mathbb{R}} f_n'(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t \right| \leqslant \int_{\mathbb{R}} |f_n'(t)|\ \mathrm{d}t = \|f_n'\|_1.$$

Ainsi, pour x non nul on a la majoration

$$\left| \int_{\mathbb{R}} f_n(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t \right| \leqslant \frac{\|f_n'\|_1}{|x|}.$$

Par convergence de $f_n \in \mathcal{C}_c^1(\mathbb{R})$ vers $f \in L^1(\mathbb{R})$ on obtient que pour tout $\varepsilon > 0$ fixé et n suffisamment grand l'inégalité $\|f - f_n\|_1 \leqslant \varepsilon$ et alors les inégalités

$$\begin{aligned} \left| \int_{\mathbb{R}} f(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t \right| &\leqslant \left| \int_{\mathbb{R}} (f(t) - f_n(t))\ \mathrm{e}^{-ixt}\ \mathrm{d}t \right| + \left| \int_{\mathbb{R}} f_n(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t \right| \\ &\leqslant \|f - f_n\|_1 + \frac{\|f_n'\|_1}{|x|} \end{aligned}$$

soit l'inégalité

$$\lim_{x\to\infty} \left| \int_{\mathbb{R}} f(t)\, \mathrm{e}^{-ixt}\ \mathrm{d}t \right| \leqslant \|f - f_n\|_1 \leqslant \varepsilon$$

valable pour tout $\varepsilon > 0$, ce qui prouve le lemme de Riemann–Lebesgue. □

Lemme 7.5. *Soient K_1 et K_2 deux fonctions de répartitions dont leurs variables aléatoires associées sont de moyennes nulles, et de fonctions caractéristiques intégrables respectives κ_1 et κ_2.*
On a l'égalité pour tout $x \in \mathbb{R}$:

$$K_1(x) - K_2(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-itx} \, \frac{\kappa_2(t) - \kappa_1(t)}{it} \, \mathrm{d}t.$$

Démonstration. Puisque les κ_j sont intégrables, la formule d'inversion 7.3 montre que les densités k_j s'écrivent :

$$k_j(y) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-ity} \, \kappa_j(t) \, \mathrm{d}t.$$

Effectuons $k_1 - k_2$, intégrons entre a et x, et posons $\Delta K := K_1 - K_2$:

$$\begin{aligned}
\Delta K(x) - \Delta K(a) &= \int_a^x k_1(y) - k_2(y) \, \mathrm{d}y \\
&= \frac{1}{2\pi} \int_a^x \int_{\mathbb{R}} \mathrm{e}^{-ity} \left(\kappa_1(t) - \kappa_2(t)\right) \, \mathrm{d}t \, \mathrm{d}y \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_a^x \mathrm{e}^{-ity} \, \mathrm{d}y \right) \left(\kappa_1(t) - \kappa_2(t)\right) \, \mathrm{d}t \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathrm{e}^{-ita} - \mathrm{e}^{-itx}}{it} \left(\kappa_1(t) - \kappa_2(t)\right) \, \mathrm{d}t
\end{aligned}$$

par application du théorème de Fubini. Ce dernier est licite car les κ_i sont intégrables en t et on a considéré un intervalle borné en y. De plus, comme κ_1 et κ_2 sont des fonctions caractéristiques on a d'après la formule de Taylor-Young pour t proche de 0 :

$$\kappa_j(t) = \kappa_j(0) + \kappa_j'(0) \, t + O\left(t^2\right)$$

et d'après la proposition 2.1, la moyenne de κ_j est égale à $\kappa_j'(0)/i$, c'est à dire que $\kappa_j'(0)$ est nulle par hypothèse. Notons que $\kappa_j(0) = 1$ car κ_j est une fonction caractéristique. Ainsi :

$$\kappa_j(t) = 1 + O\left(t^2\right)$$

et alors

$$\frac{\kappa_1(t) - \kappa_2(t)}{it} = O(t)$$

ce qui prouve que $t \mapsto (\kappa_1(t) - \kappa_2(t))/(it)$ est intégrable au voisinage de 0.
Ajoutons que $t \mapsto (\kappa_1(t) - \kappa_2(t))/(it)$ est intégrable partout ailleurs puisque κ_1 et κ_2 sont intégrables par hypothèse.
Nous pouvons alors appliquer le lemme de Riemann–Lebesgue 7.4 et faire $[a \to -\infty]$ en remarquant que $\lim\limits_{a \to -\infty} \Delta K(a) = 0$ car les K_j sont des fonctions de répartitions pour obtenir :

$$\Delta K(x) = \frac{1}{2\pi} \int_{\mathbb{R}} -\mathrm{e}^{-itx} \, \frac{\kappa_1(t) - \kappa_2(t)}{it} \, \mathrm{d}t = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-itx} \, \frac{\kappa_2(t) - \kappa_2(t)}{it} \, \mathrm{d}t.$$

ce qui conclut. $\square$

7.2 Fonction de Polya

Définition-Proposition 7.6 (Fonction caractéristique et densité de Polya)**.** *La fonction caractéristique de Polya de paramètre $L > 0$ est définie par $\omega_L : t \mapsto (1 - |t/L|)^+$. La mesure sous-jacente admet une densité :*

$$h_L : x \mapsto \frac{1 - \cos(Lx)}{\pi L x^2}.$$

Démonstration. La fonction ω_L est intégrable puisqu'elle est continue et à support fini. D'après le corollaire 7.3, μ admet une densité continue bornée f telle que :

$$\forall y \in \mathbb{R}, \qquad f(y) := \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-ity}\, \omega_L(t)\ \mathrm{d}t = \frac{1}{2\pi} \int_{-L}^{L} \mathrm{e}^{-ity}\, (1 - |t/L|)\ \mathrm{d}t.$$

Effectuons le changement de variable $u = t/L$:

$$\int_{-L}^{L} \mathrm{e}^{-ity}\, (1 - |t/L|)\ \mathrm{d}t = \int_{-1}^{1} L\, \mathrm{e}^{-iLuy}\, (1 - |u|)\ \mathrm{d}u$$

soit par intégrations par parties :

$$\begin{aligned}
\frac{2\pi}{L} f(y/L) &= \int_{-1}^{1} \mathrm{e}^{-iuy}\, (1 - |u|)\ \mathrm{d}u = \int_{-1}^{0} \mathrm{e}^{-iuy}\, (1 + u)\ \mathrm{d}u + \int_{0}^{1} \mathrm{e}^{-iuy}\, (1 - u)\ \mathrm{d}u \\
&= \left[\frac{\mathrm{e}^{-iuy}}{-iy}(1+u)\right]_{-1}^{0} - \int_{-1}^{0} \frac{\mathrm{e}^{-iuy}}{-iy}\ \mathrm{d}u + \left[\frac{\mathrm{e}^{-iuy}}{-iy}(1-u)\right]_{0}^{1} + \int_{0}^{1} \frac{\mathrm{e}^{-iuy}}{-iy}\ \mathrm{d}u \\
&= \left[\frac{\mathrm{e}^{-iuy}}{(-iy)^2}\right]_{0}^{1} - \left[\frac{\mathrm{e}^{-iuy}}{(-iy)^2}\right]_{-1}^{0} = \frac{-1}{y^2}\left(\mathrm{e}^{-iy} - 1 - 1 + \mathrm{e}^{iy}\right) = \frac{2 - 2\cos(y)}{y^2}
\end{aligned}$$

ce qui fournit :

$$f(y) = \frac{L}{2\pi} \frac{2 - 2\cos(Ly)}{L^2 y^2} = \frac{1 - \cos(Ly)}{\pi L y^2}$$

et achève la démonstration. $\square$

Remarque 7.7. *On a que pour tout $t \in \mathbb{R},\ 0 \leqslant \omega_L(t) \leqslant 1$. De plus, son support est inclus dans $[-L, L]$.*

Lemme 7.8. *Soient F et G des fonctions de répartition avec G dérivable vérifiant : $G'(x) \leqslant \lambda < +\infty$. On pose :*

$$\Delta(x) = F(x) - G(x), \quad \eta = \sup_{x \in \mathbb{R}} |\Delta(x)|, \quad \Delta_L = \Delta \star h_L \quad \textit{et} \quad \eta_L = \sup_{x \in \mathbb{R}} |\Delta_L(x)|.$$

Alors on a l'inégalité

$$\eta_L \geqslant \frac{\eta}{2} - \frac{12\lambda}{\pi L}.$$

Démonstration. Comme F et G sont des fonctions de répartitions, Δ tend vers 0 en l'infini. De plus, G est dérivable donc continue et F est continue à droite. Ainsi il existe un x_0 tel que $|\Delta|$ atteint son supremum tel que $\Delta(x_0) = \eta$ ou $\Delta(x_0^-) = -\eta$. En considérant $-\Delta$, on peut supposer sans perdre de généralité que $\Delta(x_0) = \eta$.
Comme $G'(x) \leqslant \lambda$ et F est croissante, on a $\Delta(x_0 + s) \geqslant \eta - \lambda s$ pour tout $s \geqslant 0$.
En effet, on a :

$$G(x_0 + s) - G(x_0) = \int_{x_0}^{x_0+s} G'(t)\ \mathrm{d}t \leqslant \lambda s$$

soit

$$\Delta(x_0 + s) = F(x_0 + s) - G(x_0 + s) \geqslant F(x_0 + s) - G(x_0) - \lambda s \geqslant F(x_0) - G(x_0) - \lambda s = \Delta(x_0) - \lambda s.$$

Posons $\delta = \eta/(2\lambda)$ et $t = x_0 + \delta$. On a ainsi :

$$\Delta(t - x) \geqslant \begin{cases} \eta/2 + \lambda x & \text{si } |x| \leqslant \delta \\ -\eta & \text{sinon} \end{cases}.$$

Remarquons ensuite qu'en notant $I = \mathbb{R} \setminus [-\delta, \delta]$ et par parité :

$$\int_I h_L(x)\ \mathrm{d}x = 2\int_\delta^\infty h_L(x)\ \mathrm{d}x \leqslant 2\int_\delta^\infty \frac{2}{\pi L x^2}\ \mathrm{d}x = \frac{4}{\pi L}\left[\frac{-1}{x}\right]_\delta^\infty = \frac{4}{\pi L \delta}.$$

Par parité de h_L, $x \mapsto x h_L(x)$ est impaire donc son intégrable sur $]-\delta, \delta[$ est nulle.
On obtient alors pour tout $t \in \mathbb{R}$:

$$\begin{aligned}
\Delta_L(t) &= \int_{\mathbb{R}} \Delta(t - x) h_L(x)\ \mathrm{d}x \\
&= \int_{-\delta}^{\delta} \Delta(t - x) h_L(x)\ \mathrm{d}x + \int_I \Delta(t - x) h_L(x)\ \mathrm{d}x \\
&\geqslant \int_{-\delta}^{\delta} \left(\frac{\eta}{2} + \lambda x\right) h_L(x)\ \mathrm{d}x + \int_I -\eta\, h_L(x)\ \mathrm{d}x \\
&= \frac{\eta}{2}\int_{-\delta}^{\delta} h_L(x)\ \mathrm{d}x - \eta\int_I h_L(x)\ \mathrm{d}x \\
&= \frac{\eta}{2}\left(1 - \int_I h_L(x)\ \mathrm{d}x\right) - \eta\int_I h_L(x)\ \mathrm{d}x \\
&\geqslant \frac{\eta}{2}\left(1 - \frac{4}{\pi L\delta}\right) - \eta\frac{4}{\pi L \delta} \\
&= \frac{\eta}{2} - \frac{6\eta}{\pi L \delta} = \frac{\eta}{2} - \frac{12\lambda}{\pi L}.
\end{aligned}$$

ce qui prouve le lemme puisque $\eta_L \geqslant \Delta_L(t)$. $\square$

7.3 Démonstration du théorème de Berry–Esseen

Lemme 7.9. *Soient F et G deux fonctions de répartitions dont leurs variables aléatoires associées sont de moyennes nulles. Notons φ_F et φ_G leurs fonctions caractéristiques qu'on suppose intégrables. Alors on a l'inégalité pour tout $L > 0$:*

$$\sup_{x\in\mathbb{R}} |F(x) - G(x)| \leqslant \frac{1}{\pi} \int_{-L}^{L} \frac{|\varphi_F(t) - \varphi_G(t)|}{|t|}\, \mathrm{d}t + \frac{24\lambda}{\pi L}$$

où $\lambda = \sup_{x\in\mathbb{R}} G'(x)$.

Démonstration. Soient F et G deux fonctions de répartitions dont leurs variables aléatoires associées sont de moyennes nulles. Notons φ_F et φ_G leurs fonctions caractéristiques qu'on suppose intégrables. Appliquons le lemme 7.5 à $F_L = F \star h_L$ et $G_L = G \star h_L$. On obtient :

$$F_L(x) - G_L(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{-itx}\, \frac{\varphi_{G_L}(t) - \varphi_{F_L}(t)}{it}\, \mathrm{d}t. \tag{24}$$

Or la fonction caractéristique d'une convolée est le produit des fonctions caractéristiques, de sorte que : $\varphi_{F_L} = \varphi_F\, \omega_L$ et $\varphi_{G_L} = \varphi_G\, \omega_L$ où ω_L désigne la fonction caractéristique de Polya (7.6).
On obtient ainsi, en utilisant la remarque 7.7, les inégalités :

$$\begin{aligned} |F_L(x) - G_L(x)| &\leqslant \frac{1}{2\pi} \int_{\mathbb{R}} \frac{|\varphi_F(t)\, \omega_L(t) - \varphi_G(t)\, \omega_L(t)|}{|t|}\, \mathrm{d}t \\ &\leqslant \frac{1}{2\pi} \int_{-L}^{L} \frac{|\varphi_F(t) - \varphi_G(t)|}{|t|}\, \mathrm{d}t. \end{aligned}$$

Ainsi par définition de η_L :

$$\eta_L \leqslant \frac{1}{2\pi} \int_{-L}^{L} \frac{|\varphi_F(t) - \varphi_G(t)|}{|t|}\, \mathrm{d}t$$

soit d'après le lemme 7.8

$$\frac{1}{2\pi} \int_{-L}^{L} \frac{|\varphi_F(t) - \varphi_G(t)|}{|t|}\, \mathrm{d}t \geqslant \eta_L \geqslant \frac{\eta}{2} - \frac{12\lambda}{\pi L}$$

ce qui fournit

$$\eta := \sup_{x\in\mathbb{R}} |F(x) - G(x)| \leqslant \frac{1}{\pi} \int_{-L}^{L} \frac{|\varphi_F(t) - \varphi_G(t)|}{|t|}\, \mathrm{d}t + \frac{24\lambda}{\pi L}$$

où $\lambda = \sup_{x\in\mathbb{R}} G'(x)$. □

Considérons désormais $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes identiquement distribuée de moyenne nulle $\mathbb{E}(X) = 0$, de variance $Var(X) = \mathbb{E}(X^2) = \sigma^2 = 1$ et admettant un moment absolu d'ordre 3, $\rho := \mathbb{E}(|X|^3) < \infty$. On pose $S_n = X_1 + \cdots + X_n$ pour $n \geqslant 1$.

Dans toute la suite nous noterons F_n la fonction de répartition de $S_n/\sqrt{n}$ et $\mathcal{N}$ celle de la loi Normale centrée réduite. On notera également φ la fonction caractéristique de X_1 et ψ celle de la loi Normale centrée réduite.

Corollaire 7.10. *Avec les notations précédentes, on a l'inégalité suivante :*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \mathcal{N}(x)| \leqslant \frac{1}{\pi} \int_{-L}^{L} \frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|} \,\mathrm{d}\theta + \frac{24}{\pi L\sqrt{2\pi}}.$$

Démonstration. Puisque les X_i sont centrées, la variable aléatoire $S_n/\sqrt{n}$ est de moyenne nulle.
Calculons la fonction caractéristique de $S_n/\sqrt{n}$.
Soit $t \in \mathbb{R}$.

$$\begin{aligned}
\varphi_{S_n/\sqrt{n}}(t) &= \mathbb{E}\left(\mathrm{e}^{itS_n/\sqrt{n}}\right) = \mathbb{E}\left(\prod_{k=1}^{n} \mathrm{e}^{iX_k t/\sqrt{n}}\right) \\
&= \prod_{k=1}^{n} \mathbb{E}\left(\mathrm{e}^{iX_k t/\sqrt{n}}\right) \quad \text{car les } X_k \text{ sont indépendants} \\
&= \prod_{k=1}^{n} \mathbb{E}\left(\mathrm{e}^{iX_1 t/\sqrt{n}}\right) \quad \text{car les } X_k \text{ sont identiquement distribués} \\
&= \mathbb{E}\left(\mathrm{e}^{iX_1 t/\sqrt{n}}\right)^n = \varphi(t/\sqrt{n})^n.
\end{aligned}$$

On obtient alors l'inégalité suivante avec le lemme 7.9 :

$$\sup_{x\in\mathbb{R}} |F_n(x) - \mathcal{N}(x)| \leqslant \frac{1}{\pi} \int_{-L}^{L} \frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|} \,\mathrm{d}\theta + \frac{24\lambda}{\pi L}.$$

où $\lambda = \sup_{x\in\mathbb{R}} \mathcal{N}'(x)$.
La dérivée de la fonction de répartition d'une loi gaussienne est la densité gaussienne.
Son maximum étant en 0, $\lambda = (2\pi)^{-1/2}$, ce qui achève la démonstration. $\square$

Proposition 7.11. *La fonction caractéristique de la loi normale centrée réduite est* $\psi : x \mapsto \mathrm{e}^{-x^2/2}$.

Démonstration. On a pour $x \in \mathbb{R}$,

$$\psi(x) = \int_{-\infty}^{+\infty} \mathrm{e}^{ixu} \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}} \,\mathrm{d}u.$$

Dérivons ψ pour obtenir une équation différentielle sur ψ.
On a

$$\psi'(x) = \int_{-\infty}^{+\infty} iu\mathrm{e}^{ixu} \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}} \,\mathrm{d}u$$

car $\left| iu\mathrm{e}^{ixu}\, \mathrm{e}^{-u^2/2}/\sqrt{2\pi} \right| \leqslant |u|\, \mathrm{e}^{-u^2/2}/\sqrt{2\pi} \in L^1(\mathbb{R})$.

Effectuons une intégration par parties.

$$\begin{aligned}
\psi'(x) &= \frac{-i}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} -u\mathrm{e}^{-u^2/2}\, \mathrm{e}^{ixu} \,\mathrm{d}u \\
&= \frac{-i}{\sqrt{2\pi}} \left[\mathrm{e}^{-u^2/2}\, \mathrm{e}^{ixu}\right]_{-\infty}^{+\infty} + \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{-u^2/2}\, ix\, \mathrm{e}^{ixu} \,\mathrm{d}u \\
&= i^2\, x \int_{-\infty}^{+\infty} \mathrm{e}^{ixu} \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}} \,\mathrm{d}u \\
&= -x\, \psi(x).
\end{aligned}$$

Cela fournit $\psi'(x) + x\psi(x) = 0$. En multipliant par $\mathrm{e}^{x^2/2}$ on obtient $\partial_x(\psi(x)\,\mathrm{e}^{x^2/2}) = 0$.
Donc $x \mapsto \psi(x)\,\mathrm{e}^{x^2/2}$ est constante soit $\psi(x)\,\mathrm{e}^{x^2/2} = \psi(x)\,\mathrm{e}^{x^2/2}\big|_{x=0} = \varphi(0) = 1$ (densité).
Ainsi, $\psi(x) = \mathrm{e}^{-x^2/2}$ pour tout $x \in \mathbb{R}$. □

Proposition 7.12. *On a pour* $|t| < \sigma^2/\rho = 1/\rho$ *l'inégalité :*

$$\left|\log(\varphi(t)) + \frac{t^2}{2}\right| \leqslant \frac{5}{12}\,\rho|t|^3.$$

Démonstration. Soit $T > 0$ un paramètre à déterminer tel que sur l'intervalle $[-T, T]$ $\log(\varphi(t))$ soit bien défini. D'après (1) on a l'inégalité :

$$|\varphi(t) - \varphi(0) - \varphi'(0)t| \leqslant \sup_{0<h<t} |\varphi''(h)|\,\frac{t^2}{2}$$

La proposition 2.1 fournit l'égalité $\varphi^{(k)}(0) = i^k\mathbb{E}(X^k)$ pour tout $k \in \mathbb{N}$ et on obtient l'inégalité suivante puisque la variable aléatoire sous jacente est par hypothèse centrée réduite :

$$|\varphi(t) - 1| \leqslant \sigma^2\,\frac{t^2}{2} = \frac{t^2}{2} \tag{25}$$

De plus, d'après l'inégalité de Jensen pour la fonction convexe $f : x \mapsto |x|^{3/2}$ appliqué à la variable aléatoire X^2 on obtient :

$$\sigma^3 = f(\mathbb{E}(X^2)) \leqslant \mathbb{E}(f(X^2)) = \mathbb{E}(|X|^3) = \rho.$$

Ainsi l'inégalité (25) devient pour $t \leqslant \sigma^2/\rho$:

$$|\varphi(t) - 1| \leqslant \sigma^2\,\frac{t^2}{2} \leqslant \frac{1}{2}\left(\frac{\sigma^3}{\rho}\right)^2 \leqslant \frac{1}{2}. \tag{26}$$

Posons alors $T = \sigma^2/\rho$ et sur $[-T, T]$ on a bien convergence du développement en série entière de $\log\varphi$ puisque :

$$\left|\sum_{k\geqslant 1} \frac{1}{k}\,(1 - \varphi(t))^k\right| \leqslant \sum_{k\geqslant 1} |1 - \varphi(t)|^k \leqslant \sum_{k\geqslant 1} \frac{1}{2^k} < \infty.$$

On obtient alors du développement en série entière de $\log\varphi$ l'égalité :

$$-\log(\varphi(t)) - \frac{t^2}{2} = \left(1 - \varphi(t) - \frac{t^2}{2}\right) + \sum_{k\geqslant 2} \frac{1}{k}(1 - \varphi(t))^k. \tag{27}$$

En procédant de la même façon qu'auparavant on a l'inégalité :

$$\left|\varphi(t) - 1 + \frac{t^2}{2}\right| = \left|\varphi(t) - \varphi(0) - \varphi'(0)t - \varphi''(0)\,\frac{t^2}{2}|\right| \leqslant \sup_{0<h<t} |\varphi^{(3)}(h)|\,\frac{t^3}{6} \leqslant \rho\,\frac{|t|^3}{6}. \tag{28}$$

On obtient alors avec l'égalité (27) et l'inégalité (28) :

$$\begin{aligned}
\left|\log(\varphi(t)) + \frac{t^2}{2}\right| &= \left|\varphi(t) - 1 + \frac{t^2}{2}\right| + \left|\sum_{k\geqslant 2} \frac{1}{k}(1-\varphi(t))^k\right| \\
&\leqslant \rho\,\frac{|t|^3}{6} + \sum_{k\geqslant 2} \frac{1}{k}|1-\varphi(t)|^k \\
&\leqslant \rho\,\frac{|t|^3}{6} + \sum_{k\geqslant 2} \frac{1}{2}|1-\varphi(t)|^k \\
&\leqslant \rho\,\frac{|t|^3}{6} + \frac{|1-\varphi(t)|^2}{2(1-|1-\varphi(t)|)}
\end{aligned}$$

Or puisque $|1-\varphi(t)| \leqslant 1/2$, on a $1/2 \leqslant 1 - |1-\varphi(t)|$ soit $1/(1-|1-\varphi(t)|) \leqslant 2$, ce qui fournit

$$\left|\log(\varphi(t)) + \frac{t^2}{2}\right| \leqslant \rho\,\frac{|t|^3}{6} + |1-\varphi(t)|^2 \leqslant \rho\,\frac{|t|^3}{6} + \sigma^4\,\frac{t^4}{4}$$

d'après l'inégalité (26).

En utilisant l'inégalité $\sigma^3 \leqslant \rho$ on a ainsi pour $t < T = \sigma^2/\rho$:

$$\begin{aligned}
\left|\log(\varphi(t)) + \frac{t^2}{2}\right| &\leqslant \rho\,\frac{|t|^3}{6} + \sigma^4\,\frac{t^4}{4} \\
&\leqslant \rho|t|^3\left(\frac{1}{6} + |t|\,\frac{\sigma^4}{4\rho}\right) \\
&\leqslant \rho|t|^3\left(\frac{1}{6} + \frac{\sigma^6}{4\rho^2}\right) \\
&\leqslant \rho|t|^3\left(\frac{1}{6} + \frac{1}{4}\right) = \frac{5}{12}\,\rho|t|^3.
\end{aligned}$$

□

Corollaire 7.13. *On a la majoration*

$$\int_{-L}^{L} \frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|}\,\mathrm{d}\theta \leqslant \frac{5}{12\,L}\int_{-\infty}^{\infty} \theta^2\,\mathrm{e}^{-\theta^2/12}\,\mathrm{d}\theta$$

en prenant $L = (\sigma^3/\rho)\sqrt{n} = \sqrt{n}/\rho$.

Démonstration. Soit $L = (\sigma^3/\rho)\sqrt{n} = \sqrt{n}/\rho$. Pour $|\theta| < L$ on a $|\theta/\sqrt{n}| < 1/\rho = T$.
D'après la propositon 7.12, on a pour $|\theta| < L$ l'inégalité :

$$\left|\log(\varphi(\theta/\sqrt{n})) + \frac{\theta^2}{2n}\right| \leqslant \frac{5}{12}\,\rho\,\frac{|\theta|^3}{n^{3/2}} = \frac{5}{12L}\,\frac{|\theta|^3}{n}.$$

Exploitons cette inégalité et l'expression explicite de ψ donnée par la proposition 7.11.

On a alors pour $|\theta| < L$:

$$\begin{aligned}
\frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|} &= \frac{|\varphi^n(\theta/\sqrt{n}) - \mathrm{e}^{-\theta^2/2}}{|\theta|} \\
&= \frac{\mathrm{e}^{-\theta^2/2}}{|\theta|} \left| \exp\left\{ n \log\left(\varphi(\theta/\sqrt{n})\right) + \theta^2/2 \right\} - 1 \right| \\
&= \frac{\mathrm{e}^{-\theta^2/2}}{|\theta|} \left| \exp\left\{ n \left(\log\left(\varphi(\theta/\sqrt{n})\right) + \theta^2/(2n) \right) \right\} - 1 \right| .
\end{aligned}$$

Or on a l'inégalité pour tout $a \in \mathbb{R}$: $|\mathrm{e}^a - 1| = \left| \int_0^a \mathrm{e}^t \, \mathrm{d}t \right| \leqslant |a| \, \mathrm{e}^{|a|}$.

Et en particulier pour $a = n\left(\log\left(\varphi(\theta/\sqrt{n})\right) + \theta^2/(2n)\right) \leqslant 5|\theta|^3/(12L)$:

$$\begin{aligned}
\frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|} &\leqslant \frac{\mathrm{e}^{-\theta^2/2}}{|\theta|} \, \frac{5}{12L} \, |\theta|^3 \, \exp\left\{ \frac{5}{12L} \, |\theta|^3 \right\} \\
&= \frac{5}{12L} \, \theta^2 \, \exp\left\{ \frac{5}{12L} \, |\theta|^3 - \frac{\theta^2}{2} \right\}
\end{aligned}$$

et comme $|\theta| < L$ on a :

$$\frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|} \leqslant \frac{5}{12L} \, \theta^2 \, \exp\left\{ \frac{5}{12} \, |\theta|^2 - \frac{\theta^2}{2} \right\} = \frac{5}{12L} \, \theta^2 \, \exp\left\{ -\frac{|\theta|^2}{12} \right\}.$$

Ainsi on obtient en intégrant sur $[-L, L]$:

$$\int_{-L}^{L} \frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|} \, \mathrm{d}\theta \leqslant \frac{5}{12\,L} \int_{-L}^{L} \theta^2 \, \mathrm{e}^{-\theta^2/12} \, \mathrm{d}\theta \leqslant \frac{5}{12\,L} \int_{-\infty}^{\infty} \theta^2 \, \mathrm{e}^{-\theta^2/12} \, \mathrm{d}\theta$$

puisque l'intégrande est positive, ce qui conclut. □

Lemme 7.14. *On a l'égalité*

$$\int_{-\infty}^{+\infty} \theta^2 \, \mathrm{e}^{-\theta^2/12} \, \mathrm{d}\theta = 12\sqrt{3\pi}.$$

Démonstration. Effectuons le changement de variable $u = \theta/\sqrt{12}$ et faisons une intégration par parties.

$$\begin{aligned}
\int_{-\infty}^{+\infty} \theta^2 \, \mathrm{e}^{-\theta^2/12} \, \mathrm{d}\theta &= \int_{-\infty}^{+\infty} 12u^2 \, \mathrm{e}^{-u^2} \sqrt{12} \, \mathrm{d}u \\
&= \frac{12\sqrt{12}}{-2} \int_{-\infty}^{+\infty} -2u^2 \, \mathrm{e}^{-u^2} \, \mathrm{d}u \\
&= -6\sqrt{12} \int_{-\infty}^{+\infty} u \, \frac{\partial}{\partial u}(\mathrm{e}^{-u^2}) \, \mathrm{d}u \\
&= -6\sqrt{12} \left[u \, \mathrm{e}^{-u^2} \right]_{-\infty}^{+\infty} + 6\sqrt{12} \int_{-\infty}^{+\infty} \mathrm{e}^{-u^2} \, \mathrm{d}u \\
&= 6\sqrt{12}\sqrt{\pi} \\
&= 12\sqrt{3\pi}.
\end{aligned}$$

□

Théorème 7.15 (Berry–Esseen). *Soit $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes identiquement distribuée de moyenne nulle $\mathbb{E}(X) = 0$, de variance $Var(X) = \mathbb{E}(X^2) = \sigma^2 = 1$, et admettant un moment absolu d'ordre 3, $\mathbb{E}(|X|^3) < \infty$. On pose $S_n = X_1 + \cdots + X_n$ pour $n \geqslant 1$.*
Alors, pour tout $n \geqslant 1$,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n}{\sqrt{n}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2}\ \mathrm{d}x \right| \leqslant \frac{C\ \mathbb{E}(|X|^3)}{\sqrt{n}}$$

où C désigne une constante universelle strictement positive.

Démonstration. D'après le corollaire 7.13 et le lemme 7.14 on a l'inégalité :

$$\int_{-L}^{L} \frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|}\ \mathrm{d}\theta \ \leqslant\ \frac{5}{12\,L}\int_{-\infty}^{\infty} \theta^2\,\mathrm{e}^{-\theta^2/12}\ \mathrm{d}\theta \ \leqslant\ \frac{5}{12\,L} 12\sqrt{3\pi} \ =\ \frac{5\sqrt{3\pi}}{L}$$

en prenant $L = (\sigma^3/\rho)\sqrt{n} = \sqrt{n}/\rho$.

Le corollaire 7.10 affirme

$$\sup_{x\in\mathbb{R}} |F_n(x) - \mathcal{N}(x)| \leqslant \frac{1}{\pi}\int_{-L}^{L} \frac{|\varphi^n(\theta/\sqrt{n}) - \psi(\theta)|}{|\theta|}\ \mathrm{d}\theta + \frac{24}{\pi L\sqrt{2\pi}}$$

ce qui fournit avec la majoration précédente :

$$\sup_{x\in\mathbb{R}} |F_n(x) - \mathcal{N}(x)| \leqslant \frac{5\sqrt{3\pi}}{\pi L} + \frac{24}{\pi L\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}L}\left(5\sqrt{6} + \frac{24}{\pi}\right) \approx \frac{7,934}{L} < \frac{8}{L}.$$

soit

$$\sup_{x\in\mathbb{R}} |F_n(x) - \mathcal{N}(x)| \leqslant \frac{8\rho}{\sqrt{n}}.$$

□

Remarque 7.16. *Ici on a trouvé la constante $C = 8$, des recherches plus récentes ont montré qu'on pouvait diminuer cette constante jusqu'à $0,4784$ (Asof, 2011).*

Remarque 7.17. *Si X n'est pas réduite, on se ramène au cas réduit en appliquant l'inégalité de Berry–Esseen établie ci-dessus à X/σ. De même si X n'est pas centrée réduite, par $(X - \mathbb{E}(X))/\sigma$.*
On peut alors énoncer le théorème de Berry–Esseen dans le cas le plus général :

Soit $(X_n)_{n\in\mathbb{N}}$, une suite de variables aléatoires indépendantes identiquement distribuées de moyenne $\mathbb{E}(X)$, de variance $Var(X) = \sigma^2$, et admettant un moment absolu d'ordre 3 $\mathbb{E}(|X|^3) < \infty$.
On pose $S_n = X_1 + \cdots + X_n$ pour $n \geqslant 1$. Alors, pour tout $n \geqslant 1$,

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - n\mathbb{E}(X)}{\sigma\sqrt{n}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2}\ \mathrm{d}x \right| \leqslant \frac{C\ \mathbb{E}(|X - \mathbb{E}(X)|^3)}{\sigma^3\sqrt{n}}$$

où C désigne une constante universelle strictement positive.

8 Conséquences et inégalité de Le Cam

8.1 Le théorème de Berry–Esseen pour la loi de Bernoulli

Soit X une variable aléatoire suivant une loi de Bernoulli de paramètre p. Alors l'inégalité de Berry–Esseen fournit en posant $q = 1 - p$:

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2}\ \mathrm{d}x \right| \leqslant \frac{C\ \mathbb{E}(|X-p|^3)}{(pq)^{3/2}\sqrt{n}}$$

Or $\mathbb{E}(|X-p|^3) = q|-p|^3 + p|1-p|^3 = qp^3 + pq^3 = pq(p^2+q^2) = pq(p^2+1-2p+p^2) = pq(1-2pq)$ soit :

$$\sup_{t\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} \leqslant t\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2}\ \mathrm{d}x \right| \leqslant \frac{C\,(1-2pq)}{\sqrt{npq}}$$

Notons que S_n suit une loi binomiale de paramètres n, p. Cette approximation de la loi binomiale par la loi normale est d'autant plus bonne que $C\,(1-2p(1-p))/\sqrt{np(1-p)}$ est petit. À n fixé, cette borne est minimale pour $p = 1/2$ mais explose quand p se rapproche de 0 ou de 1. Ainsi, lorsque p est trop proche de 0 ou de 1, l'approximation gaussienne du théorème central limite n'est pas très bonne.
Il est alors naturel de chercher à quantifier cette convergence dans l'esprit de l'inégalité de Berry–Esseen. En effet, de manière plus précise, pour certaines petites probabilités, la convergence a lieu vers une loi de Poisson et l'inégalité de le Cam quantifie cette vitesse de convergence.

8.2 Inégalité de Le Cam

Établissons tout d'abord un lemme pour démontrer l'inégalité de Le Cam.

Lemme 8.1. *Si le support de deux variables aléatoires X et Y est inclus dans $\mathbb{N}$ alors :*

$$\forall A \subset \mathbb{N}, \quad |\mathbb{P}(X\in A) - \mathbb{P}(Y \in A)| \leqslant \mathbb{P}(X \neq Y).$$

Démonstration. On a d'une part :

$$\mathbb{P}(X\in A) = \mathbb{P}(X \in A, X \neq Y) + \mathbb{P}(X\in A, X = Y) \leqslant \mathbb{P}(X\neq Y) + \mathbb{P}(Y \in A). \tag{29}$$

Et de même,

$$\mathbb{P}(Y \in A) \leqslant \mathbb{P}(X \neq Y) + \mathbb{P}(X\in A). \tag{30}$$

Ainsi (29) et (30) fournissent :

$$-\mathbb{P}(X\neq Y) \leqslant \mathbb{P}(X\in A) - \mathbb{P}(Y\in A) \leqslant \mathbb{P}(X \neq Y)$$

ce qui montre l'inégalité recherchée. $\square$

Théorème 8.2 (Inégalité de Le Cam)**.** *Soient $n \in \mathbb{N}^*$ et $\lambda > 0$ tels que $2\lambda < n$.*
Alors si $X_n \sim \mathcal{B}\left(n, \frac{\lambda}{n}\right)$ et $Y \sim \mathcal{P}(\lambda)$ on a :

$$\sum_{k=0}^{+\infty} |\mathbb{P}(X_n = k) - P(Y=k)| \leqslant \frac{4\lambda^2}{n}.$$

Démonstration. Notons tout d'abord $p = \lambda/n$. On définit la loi $\mathcal{M}_p$ du couple de variables aléatoires (X, Y) sur $\mathbb{N} \times \{0, 1\}$ par :

$P(X = k, Y = i)$	$k = 0$	$k = 1$	$k \in \mathbb{N},\ k \geqslant 2$	$k \in \mathbb{N}$
$i = 0$	$\mathrm{e}^{-p} - p\,(1 - \mathrm{e}^{-p})$	0	$\mathrm{e}^{-p}\,p^k/(k!)$	$1 - p$
$i = 1$	$p\,(1 - \mathrm{e}^{-p})$	$p\,\mathrm{e}^{-p}$	0	p
$i \in \{0, 1\}$	e^{-p}	$p\,\mathrm{e}^{-p}$	$\mathrm{e}^{-p}\,p^k/(k!)$	1

On vérifie que cela définit bien une loi de probabilité. Puisque $0 < p \leqslant 1/2$, toutes les probabilités énoncées ci-dessus sont bien définies. En effet, $\mathrm{e}^{-p} - p(1 - \mathrm{e}^{-p}) = \mathrm{e}^{-p}(1 + p) - p \geqslant \mathrm{e}^{-1/2} - \frac{1}{2} \approx 0,1 > 0$. Ainsi la loi de (X, Y) est bien définie, notée $\mathcal{M}_p$.

De plus, si $(X, Y) \sim \mathcal{M}_p$ alors $X \sim \mathcal{P}(p)$ et $Y \sim \mathcal{B}(p)$.
En effet, cela se vérifie facilement à l'aide de la formule des probabilités totales et du tableau ci-dessus.

Soient $(X_1, Y_1), \ldots, (X_n, Y_n)$ des couples de variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{M}_p$. Alors $X := X_1 + \cdots + X_n \sim \mathcal{P}(np) = \mathcal{P}(\lambda)$ et $Y := Y_1 + \ldots Y_n \sim \mathcal{B}(n, p) = \mathcal{B}\left(n, \frac{\lambda}{n}\right)$.
Notons $A := \{k \in \mathbb{N},\ \mathbb{P}(X = k) \geqslant \mathbb{P}(Y = k)\}$. On obtient ainsi :

$$\begin{aligned}
\sum_{k \geqslant 0} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| &= \sum_{k \in A} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| + \sum_{k \notin A} |\mathbb{P}(X = k) - \mathbb{P}(Y = k)| \\
&= \sum_{k \in A} \mathbb{P}(X = k) - \mathbb{P}(Y = k) + \sum_{k \notin A} - (\mathbb{P}(X = k) - \mathbb{P}(Y = k)) \\
&= \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) - (\mathbb{P}(X \notin A) - \mathbb{P}(Y \notin A)) \\
&\leqslant \mathbb{P}(X \neq Y) + \mathbb{P}(X \neq Y) = 2\mathbb{P}(X \neq Y) \ \text{ d'après le lemme 8.1} \\
&= 2\mathbb{P}(\exists i, X_i \neq Y_i) \quad \text{car sinon } X = Y \\
&= 2 \sum_{i=1}^{n} \mathbb{P}(X_i \neq Y_i) = 2n - 2 \sum_{i=1}^{n} \mathbb{P}(X_i = Y_i) \\
&= 2n - 2 \sum_{i=1}^{n} (\mathbb{P}(X_i = Y_i = 0) + \mathbb{P}(X_i = Y_i = 1)) \\
&= 2n - 2 \sum_{i=1}^{n} (\mathbb{P}(X_i = Y_i = 0) + \mathbb{P}(X_i = Y_i = 1)) \\
&= 2n - 2 \sum_{i=1}^{n} \left(\mathrm{e}^{-p} - p(1 - \mathrm{e}^{-p}) + p\mathrm{e}^{-p}\right) \\
&= 2n - 2n\left((2p + 1)\mathrm{e}^{-p} - p\right) \\
&\leqslant 2n - 2n\left((2p + 1)(1 - p) - p\right) \ \text{ car } 1 + x \leqslant \mathrm{e}^x \\
&= 2n - 2n(1 - 2p^2) = 4np^2 = 4n\frac{\lambda^2}{n^2} \\
&= 4\frac{\lambda^2}{n}.
\end{aligned}$$

Comme l'inégalité obtenue ne dépend que des lois de X et de Y et pas de leur couplage, cela prouve l'inégalité de Le Cam. $\square$

Corollaire 8.3 (Approximation poissonnienne)**.** *On a la convergence en loi :*

$$\mathcal{B}\left(n, \frac{\lambda}{n}\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{P}(\lambda).$$

Démonstration. Soient X_n et Y deux variables aléatoires suivant les lois : $X_n \sim \mathcal{B}\left(n, \frac{\lambda}{n}\right)$ et $Y \sim \mathcal{P}(\lambda)$ et soit f une fonction continue bornée. Alors on a :

$$\begin{aligned} |\mathbb{E}(f(X_n)) - \mathbb{E}(f(Y))| &= \left| \sum_{k=0}^{+\infty} f(k)\mathbb{P}(X_n = k) - \sum_{k=0}^{+\infty} f(k)\mathbb{P}(Y = k) \right| \\ &\leqslant \|f\|_\infty \sum_{k=0}^{+\infty} |\mathbb{P}(X_n = k) - \mathbb{P}(Y = k)| \\ &\leqslant \|f\|_\infty \frac{4\lambda^2}{n} \end{aligned}$$

d'après l'inégalité de Le Cam.

En faisant $[n \to +\infty]$, on obtient que pour toute fonction continue bornée f : $\mathbb{E}(f(X_n)) \xrightarrow[n \to +\infty]{} \mathbb{E}(f(Y))$.

Cela prouve, d'après une caractérisation usuelle de la convergence en loi, que X_n converge en loi vers Y, c'est-à-dire

$$\mathcal{B}\left(n, \frac{\lambda}{n}\right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{P}(\lambda).$$

□

Références

[BL07] Philippe Barbe et Michel Ledoux. *Probabilité*. EDP SCIENCES, 2007.

[Wil91] David Williams. *Probability With Martingales*. Cambridge University Press, 1991.

[ST43] James Alexander Shohat et Jacob David Tamarkin. *The Problem of Moments*. American Mathematical Society, 1943.

[QZ13] Hervé Queffélec et Claude Zuily. *Analyse pour l'agrégation*. 4e édition. Dunod, 2013.

[Fel66] William Feller. *An Introduction to Probability Theory and its Applications, Volume 1 and 2*. John Wiley & Sons, 1964 & 1966.

[GK19] Olivier Garet et Aline Kurtzmann. *De l'intégration aux probabilités*. Ellipses, 2019.

[CZ13] Djalil Chafaï et Pierre-André Zitt. *Probabilités*. 2013.