Sur une forme quantitative du théorème central limite

Erik Thomas*

Résumé

Dans cette note, nous donnons une forme quantitative du théorème central limite (i.e. à l'aide d'un terme de reste). Le résultat établi ressemble à une inégalité de Berry-Esseen.

1 Introduction

Dans toute cette note, les variables aléatoires sont définies sur un univers probabilisé $(\Omega, \mathscr{A}, \mathbf{P})$.

1.1 Notations et rappels

- $\mathscr{D}(\mathbf{R})$ est l'ensemble des fonctions de classe C^∞ à support compact ;
- On définit l'espace de Sobolev $W^{4,1}(\mathbf{R})$ par :

$$W^{4,1}(\mathbf{R}) := \left\{ u \in L^1(\mathbf{R}),\ u', u'', u^{(3)} \text{ et } u^{(4)} \in L^1(\mathbf{R}) \right\},$$

  où les dérivées sont au sens des distributions. Cet espace est muni de la norme $\|\cdot\|_{W^{4,1}(\mathbf{R})}$ par :

$$\forall u \in W^{4,1}(\mathbf{R}), \quad \|u\|_{W^{4,1}(\mathbf{R})} = \|u\|_1 + \|u'\|_1 + \|u''\|_1 + \|u^{(3)}\|_1 + \|u^{(4)}\|_1.$$

  On rappelle que $\mathscr{D}(\mathbf{R})$ est dense dans $W^{4,1}(\mathbf{R})$.
- On note γ_1 la loi normale centrée réduite dont la densité est donnée par $\varphi(x) := \dfrac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi}}$.

1.2 But de cette note

Le but de cette note est de donner une forme quantitative élémentaire du théorème central limite que l'on rappelle ci-dessous.

Théorème 1.1. *Théorème central limite.*
Si $(X_n)_{n\in\mathbf{N}^\star}$ est une suite de variables aléatoires i.i.d. centrées (espérance nulle) et réduites (variance égale à 1), alors

$$\lim_{n\to+\infty} \frac{X_1 + \cdots + X_n}{\sqrt{n}} \underset{\text{loi}}{=} \gamma_1.$$

De telle forme quantitative existe déjà, par exemple citons l'inégalité de Berry-Esseen :

Théorème 1.2. *Inégalité de Berry-Esseen.*
Soit $n \in \mathbf{N}^\star$ et soit $X_1, \ldots, X_n$ une suite de variables aléatoires i.i.d., centrées et réduites. On suppose qu'il existe $\tau \geq 0$ tel que pour tout $i \in [\![1, n]\!]$, $\mathbf{E}\,|X_i|^3 \leq \tau$. Alors, si l'on pose $S_n := \dfrac{X_1 + \cdots + X_n}{\sqrt{n}}$, on a :

$$\sup_{x\in\mathbf{R}} |\mathbf{P}(S_n \leq x) - \Phi(x)| \leq \frac{7\tau}{\sqrt{n}}.$$

*erik.thomas@ens-rennes.fr

Nous renvoyons à [2] pour une preuve de ce résultat. Le but de cette note est de prouver la proposition suivante.

Proposition 1.1. *Soit* $(X_n)_{n\in\mathbf{N}^\star}$ *est une suite de variables aléatoires i.i.d. centrées (espérance nulle) et réduites (variance égale à* 1*) ayant un moment d'ordre* 3 *avec* $\mathbf{E}(X_1^3)=0$*.*

Soit Y *une variable aléatoire suivant une loi normale centrée, réduite. Alors, pour toute fonction* $f\in W^{4,1}(\mathbf{R})$*, on a :*

$$\left|f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)-f\left(Y\right)\right|\leq\frac{1}{6n^{1/2}}\|f\|_{W^{1,4}(\mathbf{R})}\left(\mathbf{E}\left|X_1\right|^3+\mathbf{E}\left|Y_1\right|^3\right).$$

Remarque 1. On remarquera que la décroissance du reste en $\frac{1}{\sqrt{n}}$ est la même que celle de l'inégalité de Berry-Esseen.

2 Preuve du résultat principal

On prouve la proposition 1.1. La preuve proposée ici est une preuve « à la Lindeberg ». Il est le premier à utiliser les formules de Taylor pour une preuve d'un nouveau théorème central limite (voir [3]) pour une suite de variables aléatoires indépendantes satisfaisant une condition appelée condition de Lindeberg.

Ici, nous suivons l'approche de [1] qui donne une preuve du théorème central limite à la Lindeberg.

Démonstration. Soit $(Y_n)_{n\in\mathbf{N}^\star}$ une suite i.i.d. de variables aléatoires qui suivent une loi normale centrée réduite.

Soit $f\in\mathscr{D}(\mathbf{R})$. On pose

$$I_n(f):=\mathbf{E}\left(f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right)-\mathbf{E}\left(f\left(Y\right)\right).$$

Par stabilité de la loi normale par somme $\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\underset{\text{loi}}{=}Y$, on a donc :

$$I_n(f)=\mathbf{E}\left(f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right)-\mathbf{E}\left(f\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\right)\right).$$

On pose $U_k:=X_1+\cdots+X_k+Y_{k+1}+\cdots+Y_n$. On a :

$$\begin{aligned}f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)-f\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\right)&=f\left(\frac{U_n}{\sqrt{n}}\right)-f\left(\frac{U_0}{\sqrt{n}}\right)\\&=\sum_{k=1}^{n}\left(f\left(\frac{U_k}{\sqrt{n}}\right)-f\left(\frac{U_{k-1}}{\sqrt{n}}\right)\right).\end{aligned}$$

Pour tout $k\in[\![1,n]\!]$, on écrit :

$$f\left(\frac{U_k}{\sqrt{n}}\right)-f\left(\frac{U_{k-1}}{\sqrt{n}}\right)=f\left(Z_k+\frac{X_k}{\sqrt{n}}\right)-f\left(Z_k+\frac{Y_k}{\sqrt{n}}\right)$$

où l'on a posé $Z_k:=\frac{1}{\sqrt{n}}\left(X_1+\cdots+X_{k-1}+Y_{k+1}+\cdots+Y_n\right)$. D'après la formule de Taylor-Lagrange, il existe des quantités S_k et T_k telles que :

$$\begin{aligned}&f\left(Z_k+\frac{X_k}{\sqrt{n}}\right)-f\left(Z_k+\frac{Y_k}{\sqrt{n}}\right)\\=\;&f'\left(Z_k\right)\frac{X_k}{\sqrt{n}}+\frac{f''\left(Z_k\right)}{2}\frac{X_k^2}{n}+\frac{f^{(3)}(S_k)}{6}\frac{X_k^3}{n^{3/2}}-f'\left(Z_k\right)\frac{Y_k}{\sqrt{n}}-\frac{f''\left(Z_k\right)}{2}\frac{Y_k^2}{n}-\frac{f^{(3)}(T_k)}{6}\frac{Y_k^3}{n^{3/2}}\\=\;&f'\left(Z_k\right)\frac{X_k-Y_k}{\sqrt{n}}+\frac{f''\left(Z_k\right)}{2}\frac{X_k^2-Y_k^2}{n}+\frac{1}{6n^{3/2}}\left(f^{(3)}(S_k)X_k^3-f^{(3)}(T_k)Y_k^3\right).\end{aligned}$$

En faisant apparaître $f^{(3)}(Z_k)$, la ligne précédente peut se réécrire en

$$\begin{aligned} & f\left(Z_k + \frac{X_k}{\sqrt{n}}\right) - f\left(Z_k + \frac{Y_k}{\sqrt{n}}\right) \\ = \quad & f'\left(Z_k\right)\frac{X_k - Y_k}{\sqrt{n}} + \frac{f''\left(Z_k\right)}{2}\frac{X_k^2 - Y_k^2}{n} + \frac{f^{(3)}(Z_k)}{6}\frac{X_k^3 - Y_k^3}{n^{3/2}} + R_k, \end{aligned}$$

on l'on a posé

$$R_k := \frac{1}{6n^{3/2}}\left(\left(f^{(3)}(S_k) - f^{(3)}\left(Z_k\right)\right)X_k^3 - \left(f^{(3)}\left(T_k\right) - f^{(3)}\left(Z_k\right)\right)Y_k^3\right). \tag{1}$$

Comme $f^{(4)} \in L^1(\mathbf{R})$, on peut écrire

$$|R_k| \leq \frac{1}{6n^{3/2}}\|f^{(4)}\|_1\left(|X_k|^3 + |Y_k|^3\right).$$

En utilisant l'indépendance des variables X_k, Y_k et Z_k et en utilisant le fait que $\mathbf{E}(X_k) = \mathbf{E}(Y_k)$, $\mathbf{E}(X_k^2) = \mathbf{E}(Y_k^2)$ et $\mathbf{E}(X_k^3) = \mathbf{E}(X_k^3) = 0$, on a :

$$\left|\mathbf{E}\left(f\left(\frac{U_k}{\sqrt{n}}\right)\right) - \mathbf{E}\left(f\left(\frac{U_{k-1}}{\sqrt{n}}\right)\right)\right| \leq \frac{1}{6n^{3/2}}\|f^{(4)}\|_1\left(\mathbf{E}\left|X_k\right|^3 + \mathbf{E}\left|Y_k\right|^3\right) = \frac{1}{6n^{3/2}}\|f^{(4)}\|_1\left(\mathbf{E}\left|X_1\right|^3 + \mathbf{E}\left|Y_1\right|^3\right). \tag{2}$$

On en déduit que

$$|I_n(f)| \leq \frac{1}{6n^{1/2}}\|f^{(4)}\|_1\left(\mathbf{E}\left|X_1\right|^3 + \mathbf{E}\left|Y_1\right|^3\right) \leq \frac{1}{6n^{1/2}}\|f\|_{W^{1,4}(\mathbf{R})}\left(\mathbf{E}\left|X_1\right|^3 + \mathbf{E}\left|Y_1\right|^3\right).$$

Il s'ensuit que I_n est une forme linéaire continue sur $\mathscr{D}\left(\mathbf{R}\right)$ muni de la norme $\|\cdot\|_{W^{4,1}(\mathbf{R})}$. Par densité de $\mathscr{D}\left(\mathbf{R}\right)$ dans $W^{4,1}(\mathbf{R})$ pour la norme $\|\cdot\|_{W^{4,1}(\mathbf{R})}$ (théorème de Meyers-Serrin), on peut prolonger I_n en une application continue sur $W^{4,1}(\mathbf{R})$ de sorte que

$$\forall f \in W^{4,1}\left(\mathbf{R}\right), \quad |I_n(f)| \leq \frac{1}{6n^{1/2}}\|f\|_{W^{1,4}(\mathbf{R})}\left(\mathbf{E}\left|X_1\right|^3 + \mathbf{E}\left|Y_1\right|^3\right).$$

□

Références

[1] J. Depauw, *Une preuve élémentaire du théorème central limite.* RMS 120-1, 2009.

[2] C.-G. Esseen, *On the Liapunoff limit of error in the theory of probability.* Arkiv for Matematik Astronomi och Fysik, vol. A28, no. 9, p. 1-19, 1942.

[3] J. W. Lindeberg. *Eine neue Herleitung des Exponentialgezetzes in der Wahrscheinlichkeitsrechnung.* Mathematische Zeitschrift, 15, p. 211-225, 1922.