CHAPITRE 8

LA CONVERGENCE en LOI et LE THÉORÈME LIMITE CENTRAL

I) INTRODUCTION

Etant donné deux v.a. X et Y de lois respectives μ_1 et μ_2, il est intéressant de disposer d'une notion mathématique pour pouvoir exprimer que les lois μ_1 et μ_2 sont "proches". On a déjà ressenti ce besoin lorsqu'on a établi l'approximation de la loi binomiale $\mathcal{B}(n,p)$ par la loi de Poisson $\mathcal{P}(\lambda)$ dans le cas où "n est grand", "p petit" et $np = \lambda$ "ni trop grand ni trop petit". On a exprimé rigoureusement cette idée par un théorème de convergence (voir la proposition 3.7. du chapitre 3) des lois binomiales μ_n de paramètres (n, p_n) telles que $np_n \to \lambda$, vers la loi de Poisson μ de paramètre λ, en montrant que pour tout $k \in \mathbb{N}$,

$$\mu_n(k) \longrightarrow \mu(k) \quad n \to +\infty \quad (*)$$

où $\mu_n(k) = \mathbb{P}(X_n = k)$ est la "masse" de μ_n en k et $\mu(k) = \mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ la "masse" de μ en k

Ce cas est un peu particulier car les v.a. binomiales X_n et la v.a. de Poisson X prennent leurs valeurs dans $\mathbb{N}$.

Considérons maintenant une suite (X_n) une suite de v.a. de lois respectives (μ_n) et X une v.a. de loi μ. Pour exprimer que (μ_n) converge vers μ une première idée pour généraliser la condition $(*)$ précédente serait de dire que

$$\forall B \in \mathcal{B}_1 \ , \ \mu_n(B) \longrightarrow \mu(B) \text{ si } n \to +\infty . (**)$$

$(\mathcal{B}_1 =$ la tribu des boréliens de $\mathbb{R})$.

Mais <u>cette condition n'est pas bonne</u>; elle est trop exigeante. Elle ne permettrait pas par exemple de dire que la suite des mesures de Dirac δ_{x_n} converge vers la mesure de Dirac δ_x si $x_n \to x$ (En effet considérons par exemple les mesures de probabilités $\delta_{-1/n}$. Le bon sens exige que cette suite converge vers δ_0; or si $B =]-\infty, 0[$, on a $\delta_{-1/n}(B) = 1$ pour tout entier $n \in \mathbb{N}^*$ et $\delta_0(B) = 0$).

Plutôt que de considérer l'action des mesures sur des ensembles, il est plus fructueux de considérer l'action d'une mesure sur des fonctions boréliennes convenables via l'intégration. Ainsi pour $f : \mathbb{R} \to \mathbb{R}$ borélienne intégrable relativement à μ, on notera

$$\langle \mu, f \rangle = \int_{\mathbb{R}} f \, d\mu = \int_{\mathbb{R}} f(x) \, d\mu(x)$$

son intégrale. Notons qu'elle existe toujours si f est borélienne <u>bornée</u>.

Une condition moins exigeante que (**) serait donc de supposer que

$$\langle \mu_n, f \rangle \longrightarrow \langle \mu, f \rangle \quad (n \to +\infty)$$

pour toute f dans un sous ensemble convenable de fonctions boréliennes bornées.

II) <u>CONVERGENCE EN LOI</u>

Soit $C_b(\mathbb{R})$ l'espace des fonctions $f : \mathbb{R} \to \mathbb{R}$ <u>continues</u> et <u>bornées</u> sur $\mathbb{R}$ et $C_0(\mathbb{R})$ le sous-espace de $C_b(\mathbb{R})$ des fonctions qui tendent vers zéro à l'infini (i.e. $\lim_{|x| \to +\infty} f(x) = 0$).

Soit (X_n) une suite de v.a. et X une v.a. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$. On note μ_n (resp. μ) la loi de probabilité de X_n (resp. de X).

<u>Définition</u> : on dit que la suite (X_n) <u>converge en loi</u> vers la v.a. X si

$$\forall f \in \mathcal{C}_b(\mathbb{R}), \quad \langle \mu_n, f \rangle \longrightarrow \langle \mu, f \rangle \ (n \to +\infty) \quad (*)$$

On écrit alors $\lim\limits_{n\to\infty} X_n \stackrel{\mathcal{L}}{=} X$ ou $X_n \xrightarrow{\mathcal{L}} X \ (n \to +\infty)$.

<u>Remarque</u> : la condition $(*)$ porte sur les mesures μ_n et μ. On dit que la suite (μ_n) <u>converge étroitement vers μ</u>. Ce vocabulaire est utilisé dans la plupart des livres qui traitent de la convergence des mesures. Pour nous, convergence étroite et convergence en loi sont synonymes.

Il y a une autre notion de convergence des mesures d'une grande utilité en analyse fonctionnelle connue sous le nom de <u>convergence faible</u> :

Considérons (μ_n) et μ des mesures bornées sur $(\mathbb{R}, \mathcal{B}_1)$ (i.e. de masse finie, pas forcément des probabilités). On dit que (μ_n) converge faiblement vers μ si

$$\forall f \in \mathcal{C}_0(\mathbb{R}), \quad \lim_{n \to +\infty} \langle \mu_n, f \rangle = \langle \mu, f \rangle \quad (**)$$

Le lien entre convergence étroite et convergence faible est donné par le résultat suivant :

Soient (μ_n) une suite de <u>mesures bornées</u> et μ une mesure bornée sur $\mathbb{R}$ (<u>pas forcément des probabilités</u>) on dira que (μ_n) converge étroitement (resp. faiblement vers μ) si elle satisfait $(*)$ (resp. $(**)$).

<u>Théorème 1</u> : les conditions suivantes sont équivalentes :

1) (μ_n) converge étroitement vers μ

2) (μ_n) converge faiblement vers μ <u>et</u>
$\lim_{n\to\infty} \langle \mu_n, 1\rangle = \langle \mu, 1\rangle$ (conservation de la masse)

<u>dém^{on}</u> : 1) $\Rightarrow$ 2) évident

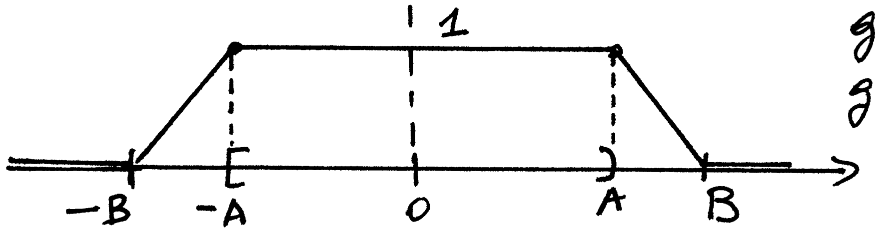
2) $\Rightarrow$ 1) : Supposons 2) vérifié. Pour toute $f \in C_b(\mathbb{R})$ et toute $g \in C_b(\mathbb{R})$ telle que $0 \leq g \leq 1$, remarquons qu'en écrivant $f = fg + f(1-g)$, on a

$$|\langle \mu_n, f\rangle - \langle \mu, f\rangle| \leq |\langle \mu_n, fg\rangle - \langle \mu, fg\rangle| + \|f\|_\infty \big(\langle \mu_n, 1-g\rangle + \langle \mu, 1-g\rangle\big) \quad (1)$$

Soit $\varepsilon > 0$. Il existe un compact $K = [-A, A]$ tel que $\mu(K^c) \leq \varepsilon$ et une fonction $g \in C_K(\mathbb{R})$, $0 \leq g \leq 1$ telle que

$$\mathbb{1}_K \leq g \quad \text{et} \quad \langle \mu, 1-g\rangle \leq \varepsilon$$

On peut prendre g trapézoïdale du type :



$g = 1$ sur $[-A, A]$
$g = 0$ sur $[-B, B]^C$

En appliquant l'inégalité (1) avec cette fonction g, et comme $\lim_{n\to\infty} \langle \mu_n, 1-g \rangle = \langle \mu, 1-g \rangle$ (par hypothèse) et que

$$\lim_{n\to\infty} \langle \mu_n, fg \rangle = \lim_{n\to\infty} \langle \mu, fg \rangle$$

(puisque $fg \in C_0(\mathbb{R})$ puisque nulle en dehors de $[-B,B]$)), si on passe à la limsup quand $n \to +\infty$ dans (1), on obtient :

$$\limsup_{n\to+\infty} \left| \langle \mu_n, f \rangle - \langle \mu, f \rangle \right| \leq 2\varepsilon \|f\|_\infty .$$

Mais $\varepsilon > 0$ est arbitraire donc la limsup précédente est nulle et 1) est démontré <u>CQFD</u>.

<u>Remarque</u> : Si on sait <u>a priori</u> que (μ_n) et μ sont des mesures de probabilité, la condition de conservation de la masse est vérifiée puisque pour tout n, $\langle \mu_n, 1 \rangle = \mu_n(\mathbb{R}) = 1 = \mu(\mathbb{R})$.

Cette condition de conservation de la masse est <u>fondamentale</u> car une suite de probabilités (μ_n) peut converger faiblement vers une mesure μ sans que μ soit nécessairement une probabilité (par exemple $\mu_n = \delta_n$).

Voyons maintenant le lien entre la convergence en loi et la convergence des fonctions de répartition

Soit (X_n) une suite de v.a. de lois respectives μ_n et de fonction de répartition $F_n(t) = \mu_n(]-\infty, t])$ et soit X une v.a. de loi μ et de fonction de répartition F.

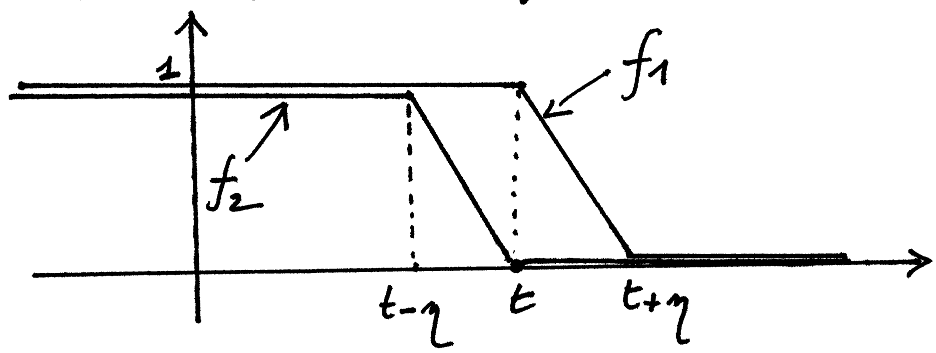
Théorème 2 $X_n \xrightarrow{\mathcal{L}} X$ $(n \to +\infty)$ si et seulement si on a

$$\lim_{n\to\infty} F_n(t) = F(t)$$

pour tout point $t \in \mathbb{R}$ où F est continue.

dém^n:

$(\Longrightarrow)$ supposons que $\mu_n \to \mu$ étroitement et soit t un point de continuité de F. Considérons $\eta > 0$ et $f_1, f_2 \in C_b(\mathbb{R})$ des fonctions de la forme suivante :



$(f_1 = f_2 = 1$ sur $]-\infty, t-\eta]$ et $f_1 = f_2 = 0$ sur $[t+\eta, +\infty[)$

On a $\qquad f_2 \leq \mathbb{1}_{]-\infty, t]} \leq f_1 \quad$ et

$$|\langle \mu, f_1\rangle - \langle \mu, f_2\rangle| = \left|\int_{\mathbb{R}} (f_1 - f_2)\, d\mu\right| \leq \int_{\mathbb{R}} \mathbb{1}_{]t-\eta, t+\eta]}\, d\mu$$
$$= F(t+\eta) - F(t-\eta) \quad (*)$$

Pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que la quantité (*) soit $\leq \varepsilon$. Mais, on a les inégalités suivantes :

$$\langle \mu_n, f_2 \rangle \leq F_n(t) \leq \langle \mu_n, f_1 \rangle \quad (1)$$

$$F(t) - \varepsilon \leq \langle \mu, f_2 \rangle \leq F(t) \leq \langle \mu, f_1 \rangle \leq F(t) + \varepsilon \quad (2)$$

Comme $\langle \mu_n, f_2 \rangle \to \langle \mu, f_2 \rangle$ et $\langle \mu_n, f_1 \rangle \to \langle \mu, f_1 \rangle$ en passant à la limite dans (1) quand $n \to +\infty$, on obtient :

$$\langle \mu, f_2 \rangle \leq \liminf F_n(t) \leq \limsup F_n(t) \leq \langle \mu, f_1 \rangle$$

Puis en utilisant (2) on a :

$$F(t) - \varepsilon \leq \liminf F_n(t) \leq \limsup F_n(t) \leq F(t) + \varepsilon$$

Mais $\varepsilon > 0$ est arbitraire ce qui implique

$$\liminf F_n(t) = \limsup_n F_n(t) = F(t)$$

autrement dit $F_n(t) \to F(t)$ d'où 2).

2) $\Rightarrow$ 1) : L'ensemble D des points discontinuité de F est fini ou dénombrable. Il en résulte que D^c est dense dans $\mathbb{R}$. Pour $I =]a, b]$ avec $a, b \in D^c$, on a

$$\lim_{n\to\infty} \langle \mu_n, \mathbb{1}_I \rangle = \lim_{n\to\infty} F_n(b) - F_n(a) = F(b) - F(a)$$
$$= \langle \mu, \mathbb{1}_I \rangle$$

Toute $f \in C_0(\mathbb{R})$ est limite uniforme d'une suite f_k de fonctions en escalier $f_k = \sum_{\text{finie}} \alpha_i \mathbb{1}_{I_i}$ où les intervalles I_i sont de la forme $]a_i, b_i]$ avec $a_i, b_i \in D^c$ (exercice facile en utilisant l'uniforme continuité de f). Par l'inégalité triangulaire on peut écrire

$$|\langle \mu_n, f\rangle - \langle \mu, f\rangle| \leq |\langle \mu_n, f\rangle - \langle \mu_n, f_k\rangle| \quad (1)$$
$$+ |\langle \mu_n, f_k\rangle - \langle \mu_n, f_k\rangle| \quad (2)$$
$$+ |\langle \mu_n, f_k\rangle - \langle \mu, f\rangle| \quad (3)$$

Soit $\varepsilon > 0$. Il existe $N (= N(\varepsilon)) \in \mathbb{N}^*$ tel que $k \geq N \Rightarrow \|f - f_k\|_\infty \leq \varepsilon/3$. Les quantités (1) et (3) sont donc inférieures à $\varepsilon/3$. Dans l'inégalité précédente en fixant $k = N$, on peut rendre $(2) \leq \varepsilon/3$ pour n assez grand ($n \geq N_1$ par exemple). Alors $n \geq N_1 \Rightarrow |\langle \mu_n, f\rangle - \langle \mu, f\rangle| \leq 3 \cdot \frac{\varepsilon}{3} = \varepsilon$. Donc μ_n converge faiblement vers μ. Comme μ_n et μ sont des probas $\langle \mu_n, 1\rangle \to \langle \mu, 1\rangle$ est évidente donc μ_n converge étroitement vers μ par le théorème 1 <u>CQFD</u>.

III LE THÉORÈME LIMITE CENTRAL

Soit (X_n) une suite de v.a. indépendantes et de même loi ayant un moment d'ordre 2. On note $m = \mathbb{E}(X_1)$ et $\sigma^2 = \operatorname{Var} X_1$.

Soit $S_n = \sum_{i=1}^{n} X_i$ et $\tilde{S}_n = \frac{S_n - nm}{\sigma\sqrt{n}}$ la v.a. centrée réduite associée. Le résultat suivant est central dans la théorie des probabilités.

Théorème : $\tilde{S}_n \xrightarrow{\mathcal{L}} N(0,1)$

où $N(0,1)$ est une v.a. normale centrée réduite

C'est ce résultat connu sous le nom de Théorème limite central qui explique l'importance de la loi normale dans le calcul des probabilités.

Ce résultat a d'abord été démontré par De Moivre et Laplace dans le cas particulier où les v.a. X_n sont de loi de Bernoulli $\mathcal{B}(1,p)$ (De Moivre pour $p=\frac{1}{2}$ en 1733 puis Laplace pour les autres valeurs de $p \in]0,1[$ vers 1800) Naturellement leur énoncé est différent et ne parle pas de convergence en loi (notion moderne introduite dans les années 1930 pour simplifier les formulations anciennes difficilement compréhensibles). Gauss a également obtenu des résultats voisins dans ses études de la distribution des erreurs de mesures mais en utilisant un point de vue différent. La première démonstration satisfaisante du théorème date de 1901, elle est de Lyapounov. Selon nos critères actuels la première démonstration rigoureuse est de Paul Lévy (vers 1930) et utilise la transformation de Fourier des mesures. Nous étudierons ce point de vue très puissant mathématiquement dans le cours de probabilités 2.

La démonstration que nous donnons du TLC est due à Trotter (1959). Elle utilise la notion d'opérateur dans un espace de Banach

1) Opérateur associé à une v.a. : Soit X une v.a. et μ_X sa loi de probabilité. On sait (cours d'Analyse) que $\mathcal{C}_0(\mathbb{R})$ muni de la norme $\|\cdot\|_\infty$ ($\|f\|_\infty = \sup_{t\in\mathbb{R}} |f(t)|$) est un espace de Banach. Pour toute $f \in \mathcal{C}_0(\mathbb{R})$ considérons la fonction $T_X f$ définie par

$$(T_X f)(t) = \int_{\mathbb{R}} f(t+x)\, d\mu_X(x) = \mathbb{E}(f(t+X))$$

En utilisant le théorème de convergence dominée, il est facile de voir que $T_X f \in \mathcal{C}_0(\mathbb{R})$. De plus par la linéarité de l'intégrale, on voit que l'application $f \mapsto T_X f$ est linéaire de $C_0(\mathbb{R})$ dans $C_0(\mathbb{R})$ et que

$$\forall f \in C_0(\mathbb{R}), \quad \|T_X f\|_\infty \leq \|f\|_\infty$$

Donc T_X est un opérateur continu ($|||T_X||| \leq 1$) et plus précisément une contraction de $C_0(\mathbb{R})$.

<u>Lemme 1</u> : si X et Y sont des v.a. indépendantes,

$$\boxed{T_X \circ T_Y = T_Y \circ T_X = T_{X+Y}}$$

<u>démⁿ</u>: Soit $f \in C_0(\mathbb{R})$, calculons $T_X \circ T_Y$:

$$(T_X \circ T_Y f)(t) = T_X(T_Y f)(t) = \int_{\mathbb{R}} T_Y f(t+x)\, d\mu_X(x)$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(t+x+y)\, d\mu_Y \right) d\mu_X(x) =$$

$$\iint_{\mathbb{R}^2} f(t+x+y) \underbrace{d\mu_X(x)\, d_Y(y)}_{d\mu_X \otimes \mu_Y (x,y) = d\mu_{(X,Y)}(x,y)} \quad (\text{Fubini})$$

(indépendance de X et Y)

$$= \mathbb{E}(f(t+X+Y)) \quad (\text{Formule du transfert})$$

$$= T_{X+Y} f(t)$$

De même $T_Y \circ T_X f(t) = T_{X+Y} f(t)$ <u>CQFD</u>.

<u>Lemme 2</u> Si X et Y sont des v.a. indépendantes, pour toute $f \in C_0(\mathbb{R})$, on a

$$\| T_X^n f - T_Y^n f \|_\infty \leq n \| T_X f - T_Y f \|_\infty .$$

(où $T_X^n = T_X \circ T_X \circ \ldots \circ T_X$ (n fois) est la puissance n de composition de l'opérateur T_X)

<u>dém^n</u> : on utilise la relation purement algébrique

$$T_X^n f - T_Y^n f = \sum_{i=1}^{n-1} T_X^{n-i+1} \circ (T_X - T_Y) \circ T_Y^i f$$

$$= \sum_{i=1}^{n-1} T_X^{n-i+1} \circ T_Y^i \circ (T_X - T_Y) f$$

(car T_X et T_Y commutent). Il suffit de prendre les normes et d'utiliser l'inégalité triangulaire pour obtenir le résultat.

<u>Lemme 3</u> : Si $N_1, N_2, \ldots, N_k$ sont des v.a. indépendantes de lois normales $\mathcal{N}(m_1, \sigma_1^2), \ldots, \mathcal{N}(m_k, \sigma_k^2)$, alors la somme $N_1 + \cdots + N_k$ est de loi normale $\mathcal{N}(m, \sigma^2)$ où $m = \sum_{i=1}^{k} m_i$ et $\sigma^2 = \sum_{i=1}^{k} \sigma_i^2$.

<u>dém^n</u> : exercice (il suffit de faire le calcul pour $k=2$).

<u>Remarque</u> : si N est une v.a. normale $\mathcal{N}(0,1)$,

pour tout entier $n \geq 1$, on a

$$T_N = T_{\frac{N}{\sqrt{n}}}^{\,n}$$

en effet soient $N_1, \ldots, N_n$ des v.a. indépendantes de même loi $\mathcal{N}(0,1)$. La v.a. $\frac{N_i}{\sqrt{n}}$ est de loi $\mathcal{N}(0, \frac{1}{n})$ et $\frac{1}{\sqrt{n}}(N_1 + \cdots + N_n)$ de loi $\mathcal{N}(0,1)$ d'après le lemme 3.

D'où $\quad T_{\frac{1}{\sqrt{n}}(N_1+\cdots+N_n)} = T_{\frac{1}{\sqrt{n}}N_1} \circ T_{\frac{1}{\sqrt{n}}N_2} \circ \cdots \circ T_{\frac{1}{\sqrt{n}}N_n} = T_{\frac{N_1}{\sqrt{n}}}^{\,n} = T_{\frac{N}{\sqrt{n}}}^{\,n}$

puisque l'opérateur T_X ne dépend que de la loi de X.

<u>Lemme 4</u> Soient X_n, $n \geq 1$ et X des v.a. Une condition suffisante pour que $X_n \xrightarrow{\mathcal{L}} X$ est que l'on ait :

$$\lim_{n\to\infty} \| T_{X_n} f - T_X f \|_\infty = 0$$

pour toute $f \in C_0(\mathbb{R})$ telle que f' et f'' existent et appartiennent aussi à $C_0(\mathbb{R})$.

<u>dém^n</u> : La condition implique pour toute $f \in C_0(\mathbb{R})$ avec $f', f'' \in C_0(\mathbb{R})$ on a $(T_{X_n} f)(0) - (T_X f)(0) \underset{(n\to\infty)}{\longrightarrow} 0$

i.e. $\lim\limits_{n\to\infty} \langle \mu_{X_n}, f \rangle = \langle \mu_X, f \rangle$ pour toute $f \in C_0(\mathbb{R})$ avec f' et $f'' \in C_0(\mathbb{R})$. Mais ces fonctions sont denses dans $C_0(\mathbb{R})$ donc la limite précédente est vraie pour toute

$f \in C_0(\mathbb{R})$ d'où la convergence étroite de μ_{X_n} vers μ_X (la conservation de la masse est assurée puisque μ_{X_n} et μ_X sont des probabilités).

<u>Démonstration du théorème limite central</u> : Il suffit de supposer $m=0$ et $\sigma^2=1$. Alors

$\tilde{S}_n = \frac{1}{\sqrt{n}}(X_1+\cdots+X_n)$ et soit N indépendante des X_i et de loi $N(0,1)$.

$$T_{\tilde{S}_n} = T_{\frac{X_1}{\sqrt{n}}}^{\,n} \quad , \quad T_N = T_{\frac{N}{\sqrt{n}}}^{\,n}$$

(voir la remarque suivant le lemme 3)

Soit $f \in C_0(\mathbb{R})$ avec f' et $f'' \in C_0(\mathbb{R})$. La formule de Taylor donne :

$$(*) \quad f(t+u) = f(t) + u f'(t) + \frac{u^2}{2} f''(t) + \frac{u^2}{2}\big(f''(\eta) - f''(t)\big)$$

où η est un nombre entre t et $t+u$. Alors en intégrant en u la relation $(*)$ par rapport à la loi de $\frac{X_1}{\sqrt{n}}$ (qui est centrée), on a :

$$T_{\frac{X_1}{\sqrt{n}}} f(t) = f(t) + \frac{1}{2n} f''(t) + \frac{1}{2n}\int_{\mathbb{R}} \big(f''(\tilde{\eta}) - f''(t)\big) u^2 \mu_{X_1}(u)\,du$$

où $\tilde{\eta} \in [t, t+\frac{u}{\sqrt{n}}]$. On a aussi la même expression pour $T_{\frac{N}{\sqrt{n}}} f(t)$ en remplaçant dans le membre de droite la mesure μ_{X_1} par la mesure μ_N.

Soit maintenant $\varepsilon > 0$ et $\delta > 0$ telles que

$$|f''(\tilde{\eta}) - f''(t)| \leq \varepsilon \text{ dès que } |\tilde{\eta} - t| \leq \delta.$$

En décomposant l'intégrale $\frac{1}{2n}\int_{\mathbb{R}} (f''(\tilde{\eta}) - f''(t)) u^2 \mu_{X_1}(u) du$

en deux intégrales, l'une sur l'ensemble $\{|u| < \delta\sqrt{n}\}$ et l'autre sur $\{|u| \geq \delta\sqrt{n}\}$, la première est inférieure à $\frac{\varepsilon}{2n}$ et la seconde est inférieure à

$$C\int_{\{|u| \geq \delta\sqrt{n}\}} u^2 \mu_{X_1}(du)$$

donc tend vers zéro (si $n \to +\infty$) car μ_{X_1} a un moment d'ordre 2. On obtient donc uniformément pour $t \in \mathbb{R}$:

$$\left| T_{\frac{X_1}{\sqrt{n}}} f(t) - f(t) - \frac{1}{2n} f''(t) \right| \leq \frac{\varepsilon}{n}$$

pour tout entier n assez grand. On obtient aussi la même inégalité avec N au lieu de X_1. En additionnant les deux inégalités obtenues, on voit que

$$\left| T_{\frac{X_1}{\sqrt{n}}} f(t) - T_{\frac{N}{\sqrt{n}}} f(t) \right| \leq \frac{2\varepsilon}{n}$$

uniformément en $t \in \mathbb{R}$ pour tout n assez grand. Compte tenu du lemme 2, on a montré que

$$\| T_{\tilde{S}_n} f - T_N f \|_\infty \leq 2\varepsilon$$

pour tout n assez grand. Ce qui prouve le théorème compte tenu du lemme 4. CQFD

Remarque : Le fait (utilisé dans le lemme 4) que les fonctions $f \in C_0(\mathbb{R})$ deux fois dérivables et telles que f' et $f'' \in C_0(\mathbb{R})$ sont denses dans $C_0(\mathbb{R})$ peut se démontrer en utilisant une approximation de l'identité de classe $\mathcal{C}^\infty$ à support compact.

Remarque : La démonstration précédente est extraite de l'article original de H.F. Trotter : « An elementary proof of the central limit theorem » paru dans Archiv der Math. N°10 (1959) p. 226-234.

Généralisations du TLC : Il y a un grand nombre de généralisations du TLC avec diverses modifications des hypothèses principalement à des cas où les v.a. X_n sont de lois différentes. Voici par exemple un tel résultat (admis) :

Théorème (de Lindeberg-Feller) : Soit (X_n) une suite de v.a. indépendantes telles qu'il existe $a > 0$ tel que :

$$\mathbb{P}(X_n \in [-a, a]) = 1 \text{ pour tout } n$$

(on dit que les X_n sont uniformément bornées).

On suppose que X_i a une variance σ_i^2 ($i \in \mathbb{N}$) et que

$$\sum_{i=1}^{n} \sigma_i^2 \to +\infty \quad \text{si } n \to +\infty$$

Alors la v.a. centrée réduite associée à $S_n = X_1 + \cdots + X_n$, converge en loi vers la loi normale $\mathcal{N}(0,1)$

Noter que ce résultat s'applique lorsque les v.a. indépendantes X_n sont uniformément bornées et ont des variances égales ou voisines.

Exemple: la veille d'un week-end 200 personnes font la queue à un distributeur de billets pour s'approvisionner en argent liquide. On sait par expérience que les sommes prélevées sont des v.a. indépendantes de moyenne 200 et d'écart type 50. Le distributeur contient 42000 unités. Quelle est la probabilité que cette somme suffise à payer tous les clients ?

Solution: Soit X la somme demandée. On a $E(X) = 40000$ et $\sigma_X^2 = 200 \cdot 25 \cdot 10^2$ ie $\sigma_X = 707$. Si les conditions du TLC sont satisfaits: $\mathbb{P}(X \leq 42000) = \mathbb{P}\left(\frac{X - 40000}{707} \leq 2{,}82\right) = 0{,}9976$.