

## Fragment 2

# Théorèmes limites classiques

Dans ce chapitre, on s'intéresse aux « théorèmes limites » classiques comme la loi des grands nombres (LGN) et le théorème central limite (TCL), ainsi qu'à certains de leurs raffinements et extensions.

### 2.1 Lois des grands nombres & théorème de Glivenko-Cantelli

Pour des variables aléatoires centrées et de carré intégrable, l'indépendance deux à deux entraîne l'orthogonalité. La somme de  $n$  vecteurs orthogonaux de même norme étant de l'ordre de  $\sqrt{n}$ , on s'attend donc à ce que la moyenne arithmétique des v.a. soit de l'ordre de  $1/\sqrt{n}$ , et donc à ce qu'elle converge vers 0. C'est ce que confirme la loi des grands nombres (LGN) :

**Théorème 2.1.1 (Loi des Grands Nombres).** *Soit  $(X_n, n \in \mathbb{N}^*)$  une suite de variables aléatoires i.i.d. de loi  $\mu$  sur  $\mathbb{R}$ , alors leur moyenne arithmétique*

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

*converge  $\mu$ -presque-surement si et seulement si les  $X_n$  sont intégrables. De plus, dans ce cas, la limite vaut  $m_\mu := \mathbf{E}(X_1)$  et la convergence a également lieu dans  $\mathbf{L}^1$ .*

Ce théorème est très facile à démontrer si l'on renforce les hypothèses. Supposons par exemple que les v.a.  $(X_n, n \in \mathbb{N}^*)$  sont i.i.d., dans  $\mathbf{L}^4$  et indépendantes 4 à 4. On peut supposer sans perdre de généralité qu'elles sont centrées. On a alors :

$$\mathbf{E}\left((S_n)^4\right) = \frac{1}{n^3} \mathbf{E}(X_1^4) + \frac{4n(n-1)}{n^4} \mathbf{E}(X_1^2),$$

d'où  $\sum_{n=1}^{+\infty} \mathbf{E}\left((S_n)^4\right) < +\infty$ , or d'après le théorème de convergence monotone :

$$\sum_{n=1}^{+\infty} \mathbf{E}\left((S_n)^4\right) = \mathbf{E}\left(\sum_{n=1}^{+\infty} (S_n)^4\right).$$

Cette espérance étant finie, la v.a.  $\sum_{n=1}^{+\infty} (S_n)^4$  est finie p.s. et donc  $(S_n)^4$  converge p.s. vers 0 en tant que terme général d'une série convergente, d'où la convergence p.s. vers 0 de  $S_n$ , CQFD.

*Remarque 2.1.2.* On parle souvent de « loi faible des grands nombres » lorsque seule la convergence en probabilité est énoncée. Pour les lois discrètes, cette version faible est ancienne et remonte sans doute aux travaux de Markov. D'autre part, le fait qu'il y ait convergence  $\mathbf{L}^1$  dans la LGN est souvent désigné sous le nom de « loi des grands nombres de Kolmogorov ».

*Remarque 2.1.3.* Les hypothèses faites sur la suite des v.a. dans la LGN peuvent être considérablement affaiblies, au détriment de la preuve qui est alors plus difficile (par exemple le théorème d'Etemadi qui ne requière que l'indépendance deux à deux [Bor95]). De façon générale, la LGN reste encore vraie lorsque les v.a. n'ont pas la même loi – il faut toute fois contrôler certains moments. Nous avons évoqué une sorte de LGN pour le processus de Poisson (équation (6.2) page 101) et pour le mouvement brownien (équation (5.1) page 97). Nous verrons plus généralement dans le chapitre 5 page 81 que des théorèmes du type LGN et TCL existent pour des suites de variables aléatoires « dépendantes ».

### 2.1.1 Mesure empirique

Soit  $(X_i, n \in \{1, \dots, n\})$  une suite finie de variables aléatoires i.i.d. de loi  $\mu$  sur un espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ . En statistique, on parle de *n-échantillon* de loi  $\mu$ . On note  $F : \mathbb{R} \rightarrow [0, 1]$  la fonction de répartition de la loi  $\mu$ , donnée par  $F(t) := \mu(-\infty, t]$ . À l'échantillon  $(X_i, n \in \{1, \dots, n\})$ , on associe une *mesure empirique*  $L_n(\omega)$  définie par

$$L_n(\omega) := \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)},$$

et une *fonction de répartition empirique*  $F_n(\omega)(\cdot) : \mathbb{R} \rightarrow [0, 1]$  par

$$F_n(\omega)(t) := \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{\{X_i \leq t\}}(\omega).$$

En regroupant suivant les valeurs des  $X_i(\omega)$ , on obtient les écritures suivantes faisant apparaître les *effectifs cumulés* et les *fréquences empiriques* :

$$L_n(\omega) = \sum_{x \in \{X_1(\omega), \dots, X_n(\omega)\}} \frac{\#\{1 \leq i \leq n, X_i(\omega) = x\}}{n} \delta_x,$$

et :

$$F_n(\omega)(t) = \frac{1}{n} \#\{1 \leq i \leq n, X_i(\omega) \leq t\}.$$

Notons que  $L_n$  est une variable aléatoire à valeurs dans l'ensemble des lois de probabilité sur  $\mathbb{R}$ , et que  $F_n$  est une variable aléatoire à valeurs dans l'ensemble des fonctions de répartition sur  $\mathbb{R}$ . Ainsi, pour tout  $\omega \in \Omega$ ,  $L_n(\omega)$  est une loi sur  $\mathbb{R}$ , de fonction de répartition  $F_n(\omega)(\cdot)$ .

Une conséquence de la LGN est que la mesure empirique  $L_n(\omega)$  converge étroitement vers la loi  $\mu$ . On rappelle qu'une suite de mesures de probabilités  $(\mu_n)_n$  converge étroitement vers une mesure de probabilité  $\mu$  lorsque pour toute fonction continue bornée  $f$ ,  $\mathbf{E}_{\mu_n}(f)$  converge vers  $\mathbf{E}_{\mu}(f)$  quand  $n \rightarrow \infty$ . Une suite de variables aléatoires converge en loi si et seulement si la suite de leurs lois converge étroitement. Dans le cas des mesures empiriques, on a pour  $\mu$ -presque tout  $\omega$  :

$$L_n(\omega) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mu.$$

Pour le voir, il suffit d'écrire, pour une fonction  $f \in \mathcal{C}_b(\mathbb{R}, \mathbb{R})$ ,

$$\mathbf{E}_{L_n(\omega)}(f) = \frac{1}{n} \sum_{i=1}^n f(X_i(\omega)).$$

Les variables aléatoires  $Y_i := f(X_i)$  sont i.i.d. et la loi des grands nombres nous assure la convergence presque sûre de  $\mathbf{E}_{L_n(\omega)}(f)$  vers  $\mathbf{E}(Y_1) = \mathbf{E}_{\mu}(f)$ . La permutation des quantificateurs universels sur  $f$  et  $\omega$

se fait en utilisant la séparabilité de  $\mathbb{R}$ . Ce résultat de convergence étroite des mesures empiriques peut être affiné en une convergence uniforme des fonctions de répartition, comme le précise le théorème de Glivenko-Cantelli :

**Théorème 2.1.4 (Glivenko-Cantelli).** Soit  $(X_n, n \in \mathbb{N}^*)$  une suite de variables aléatoires i.i.d. de loi  $\mu$  sur  $\mathbb{R}$ , de fonction de répartition  $F$ , alors on a pour  $\mu$ -presque-tout  $\omega$

$$\lim_{n \rightarrow \infty} \left( \sup_{t \in \mathbb{R}} |F_n(\omega)(t) - F(t)| \right) = 0, \quad (2.1)$$

où  $F_n(\omega)(\cdot)$  désigne la fonction de répartition empirique des  $X_1(\omega), \dots, X_n(\omega)$ .

*Exercice 2.1.5 (Polynômes de Bernstein & théorème de Weierstrass).* Le théorème de Weierstrass sur la densité de  $\mathbb{R}[X]$  dans  $\mathcal{C}([0, 1], \|\cdot\|_\infty)$  peut-être établi en utilisant une forme faible de la loi des grands nombres (convergence en probabilité de  $S_n$  vers la moyenne de la loi  $\mu$ ).

En effet, si  $f : [0, 1] \rightarrow \mathbb{R}$  est continue, le polynôme de Bernstein

$$P_n(f)(x) := \sum_{k=0}^n x^k (1-x)^{n-k} C_n^k f(k/n)$$

converge uniformément vers  $f$  sur  $[0, 1]$ . Pour le voir, il suffit d'exprimer  $P_n(f)(x)$  comme la moyenne de  $f$  sous la loi binomiale de paramètre  $x$  sur  $\{0, 1/n, \dots, 1\}$ , puis de découper autour de la moyenne  $p$  et d'utiliser l'inégalité de Chebychev et l'absolue continuité de  $f$  (théorème de Heine). Pour illustrer ce qui précède, créer un fichier `ber.m` contenant le code suivant :

```
function Y=ber(X,n,f)
%BER Approximation de Bernstein sur [0,1]
% Y=ber(X,n,f)
% Renvoie l'approximation de Bernstein y=Pn(f)(X) de f(X).
% X est un vecteur quelconque a valeurs dans [0,1]
% Fait appel a StixBox pour les coef. binomiaux
%
% Exemple : X=[.1:.1:.9]; f=inline('x.^4.*sin(x)'); ber(X,100,f)
%
m=length(X);
D= repmat([0:n],m,1); % Réplication de D
B=bincoef(D,n); % Calcul des coefficients binomiaux
F=repmat(feval(f,[0,[1:n]/n]),m,1); % Calcul et réplication de f
X=reshape(X,1,m); % Remise en forme de X
XX=repmat(X',1,n+1); % Réplication de X
P=(XX.^D) .* ((1-XX).^(n-D)); % Calcul des coefficients en X
Y=sum((B.*F.*P)'); % Calcul de la valeur du polynôme
```

Voici maintenant un petit programme faisant appel à cette fonction, dont la sortie graphique est donnée par la figure 2.1 :

```
% Trace la fonction log(1+x)+x^4 sin(10*x) sur [0,1] ainsi que
% les polynômes de Berntein associés de degre 5, 10, et 20.
% Utilise la fonction ber pour le calcul des polynômes de Berstein
clf;
clear;
f=inline('log(1+x)+x.^4.*sin(10*x)');
S=[.01:.01:.99]; % subdivision de [0,1]
```

```

F=feval(f,S); % image de la subdivision par notre fonction
B1=ber(S,5,f);
B2=ber(S,10,f);
B3=ber(S,20,f);
plot(S,F,'k-',S,B1,'r.',S,B2,'b+',S,B3,'g*');
legend('f','B(5,f)','B(10,f)','B(20,f)');

```

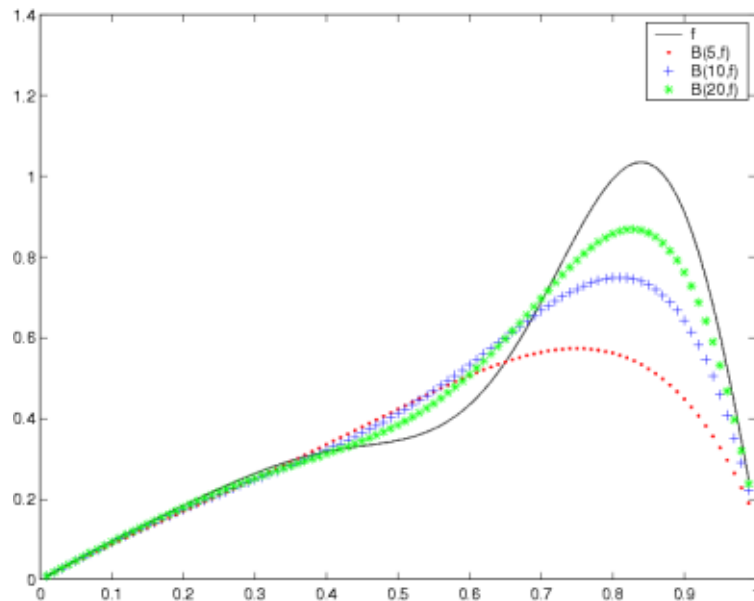


FIG. 2.1 – Illustration du théorème de Weierstrass avec les polynômes de Bernstein.

La méthode de Monte-Carlo pour le calcul approché d'intégrales est également basée sur la loi des grands nombres, cf. fragment 3.2 page 64.

## 2.2 Théorème Central Limite & principe d'invariance

Poursuivant l'analogie avec les vecteurs orthogonaux d'un espace euclidien (cf. section 2.1 page 41), on s'attend à ce que la somme des  $n$  premiers vecteurs divisée par  $\sqrt{n}$  « n'explose pas » lorsque  $n$  tend vers l'infini. Cette intuition est confirmée par le Théorème Central Limite (TCL<sup>1</sup>) :

**Théorème 2.2.1 (Théorème Central Limite).** *Si  $(X_n, n \in \mathbb{N}^*)$  est une suite de v.a.r. i.i.d. de carré intégrable de moyenne  $m$  et de variance  $\sigma^2$ , alors*

$$Y_n := \frac{X_1 + \cdots + X_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

<sup>1</sup>On parle également parfois de « Théorème de la Limite Centrale ». En anglais, on dit tout simplement « Central Limit Theorem », que l'on abrège en CLT.

De manière équivalente, si  $Q_n$  désigne la fonction de répartition de la v.a.r.  $Y_n$  et  $\Phi$  la fonction de répartition de la loi normale  $\mathcal{N}(0, 1)$  :

$$\Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left(-\frac{x^2}{2}\right) dx,$$

alors on a pour tout  $t$  dans  $\mathbb{R}$  :

$$Q_n(t) \xrightarrow[n \rightarrow +\infty]{} \Phi(t).$$

On peut également utiliser le théorème de Lévy de caractérisation de la convergence en loi par la transformée de Fourier :

$$(\forall s \in \mathbb{R}) \quad \mathbf{E}(e^{isY_n}) \xrightarrow[n \rightarrow +\infty]{} e^{-s^2/2}.$$

Pour établir le TCL, on peut supposer sans perdre de généralité que les v.a. sont centrées réduites, on a alors :

$$\mathbf{E}(e^{isY_n}) = \left(\mathbf{E}(e^{isn^{-1/2}X_1})\right)^n,$$

qui donne le résultat par un développement en série entière, mais il faut prendre garde au fait que l'on a affaire à une fonction complexe dont il faut choisir la détermination...

De même que pour la LGN, de nombreuses versions du TCL existent, notamment avec des hypothèses plus faibles sur les lois des v.a. On peut par exemple affaiblir l'indépendance requise ou encore se passer du fait que les v.a. aient la même loi, moyennant une hypothèse sur leurs moments. Historiquement, le premier TCL concernait les lois de Bernoulli :

**Théorème 2.2.2 (Théorème de Moivre-Laplace).** *Soit  $(X_n, n \in \mathbb{N}^*)$  des v.a. i.i.d. de loi de Bernoulli sur  $\{0, 1\}$  et de paramètre  $p \in [0, 1]$ , alors, pour tout  $a \in \mathbb{R}$  et  $b \in \mathbb{R}$  avec  $a < b$ , on a :*

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{npq}} \leq b\right) = (2\pi)^{-1/2} \int_a^b e^{-x^2/2} dx,$$

où  $q := 1 - p$ .

On en trouvera une preuve dans [App96, cor. 8.9] et dans [BL98, expl. V.5.5(i) pages 145-146]. Voici un exemple de code Matlab illustrant le TCL, dont la sortie graphique est donnée par la figure 2.2 :

```
% Illustration du TLC avec un échantillon de lois uniformes
% Fait appel à StixBox pour pnorm
clear;
i=0;
p=input('Nombre de v. a. uniformes (par exemple 10) = ');
n=input('Taille de l' échantillon (par exemple 50) = ');
d=input('Discrétisation (par exemple 100) = ');
% On simule n fois la variable Z:=(U_1+...+U_p-p/2)/(p/12)
X=rand(p,n);
Z=sum(X)-p/2;      % Centrage
Z=Z/sqrt(p/12);    % Réduction
dz=range(Z)/d;
% Fonction de Répartition Simulée
[Effectifs, Classes]=hist(Z,d);
FRS=cumsum(Effectifs)/n;
FRS(d+1)=1;
% Tracé de la fonction de répartition simulée
% et de celle de la loi normale
```

```

Abscisses=[min(Z)+dz/2:dz:max(Z)+dz/2];
figure; clf; hold on;
plot(Abscisses,FRS,'b+-');
FRN=pnorm(Abscisses);
plot(Abscisses,FRN,'r-');
legend('Fonc. de rep. simulee','Fonc. de rep. Normale',4);

```

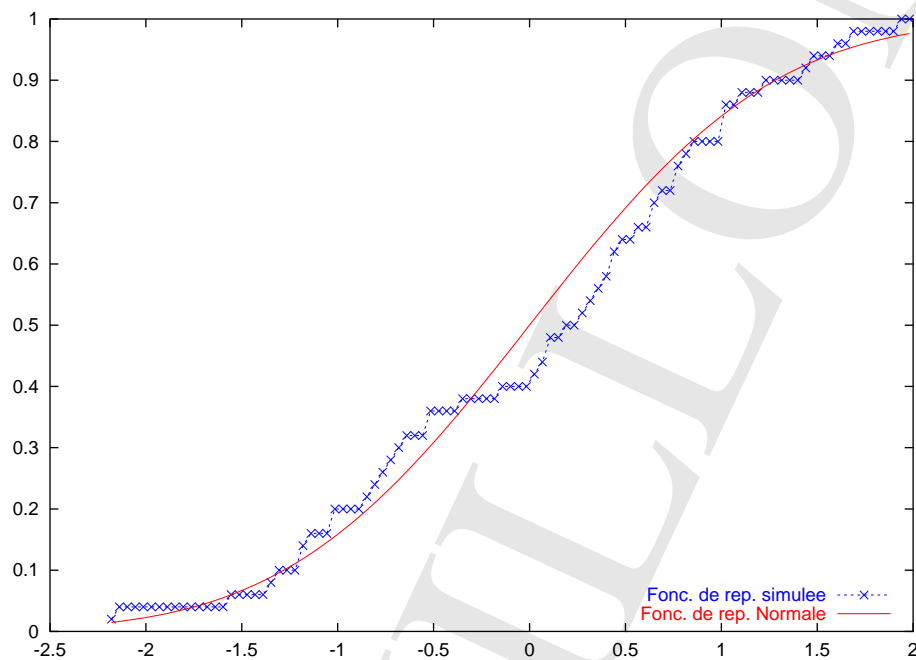


FIG. 2.2 – Illustration du TCL (lois uniformes).

Le théorème suivant est une conséquence du TCL. Il affirme que les courbes affines par morceaux construites à partir des sommes partielles renormalisées d'une suite de v.a. i.i.d., vues comme des courbes aléatoires, « convergent en loi vers la loi d'un mouvement brownien ». Il constitue donc une justification de l'intuition selon laquelle une marche aléatoire est un analogue discret du mouvement brownien :

**Théorème 2.2.3 (Principe d'invariance de Donsker-Prohorov).** *Soit  $(X_n, n \in \mathbb{N}^*)$  une suite de v.a.r. i.i.d. centrées de variance  $\sigma^2$ . On définit la suite de v.a.  $(Y_n, n \in \mathbb{N}^*)$  à valeurs dans  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$  muni de sa tribu borélienne associée à la topologie cylindrique par :*

$$Y_n(t) := \frac{X_1 + \cdots + X_{[nt]} + (nt - [nt])X_{[nt]+1}}{\sigma\sqrt{n}}.$$

où  $[r]$  désigne la partie entière de  $r$ . Alors, la suite de v.a.  $(Y_n, n \in \mathbb{N}^*)$  converge faiblement vers une loi de probabilité  $\mu^*$  sur  $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$  telle que la v.a.  $(B_t, t \geq 0)$  définie par

$$\begin{aligned} B_t : (\mathcal{C}(\mathbb{R}_+, \mathbb{R}), \mu^*) &\rightarrow \mathbb{R} \\ \omega &\mapsto B_t(\omega) := \omega(t) \end{aligned}$$

soit un mouvement brownien standard sous  $\mu^*$ , c'est-à-dire un processus gaussien à trajectoires presque-sûrement continues, centré, à accroissements indépendants, tel que  $B_0 = 0$  et  $\mathbb{E}(B_t - B_s)^2 = |t - s|$ . En particulier, pour tout  $0 \leq t_1 < \dots < t_n$ , on a :

$$(Y_n(t_1), \dots, Y_n(t_n)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} (B(t_1), \dots, B(t_n)).$$

Voici une illustration de ce théorème en Matlab, dont la sortie graphique est donnée par la figure 2.3 :

```

%% Illustration du théorème de Donsker.
clear;
% On veut simuler Y_n
n=1000;
mint=0; % Valeur maximale de t
maxt=1; % Valeur minimale de t
Dt=80; % Précision du tracé en t
% Approximation de Y_n
dt=(maxt-mint)/Dt; % Pas de la subdivision en t
T=[mint:dt:maxt]; % Subdivision en t de [mint,maxt]
X=rand(1,1+floor(n*maxt)); % Génération des v.a. X_1,...,X_n
X=X-mean(X); % Centrage empirique (théorique=1/2)
sigma=std(X); % Sigma empirique (théorique=1/12)
for i=1:Dt,
    Y(i)=sum(X(1:floor(i*T(i))))+(i*T(i)-floor(i*T(i)))*X(1+floor(i*T(i)));
end
Y(Dt+1)=0;
Y=Y/(sigma*sqrt(n));
figure
clf;
hold on
plot(T,Y,'r-')
% Approximation au dixième Y_(n/10)
Dt=floor(Dt/10); % Précision du tracé en t
dt=(maxt-mint)/Dt; % Pas de la subdivision en t
T=[mint:dt:maxt]; % Subdivision en t
for i=1:Dt,
    Z(i)=sum(X(1:floor(i*T(i))))+(i*T(i)-floor(i*T(i)))*X(1+floor(i*T(i)));
end
Z(Dt+1)=0;
Z=Z/(sigma*sqrt(n));
plot(T,Z,'b-o')
legend('Y(n)', 'Y(n/10)')
xlabel('t')
ylabel('Yn')

```

## 2.3 Vitesse de convergence dans le TCL

Considérons une suite  $(X_n, n \in \mathbb{N}^*)$  de v.a. i.i.d. de carré intégrable de moyenne  $m$  et de variance  $\sigma^2$ . Soit  $G_n$  la fonction de répartition de la v.a.r.

$$Y_n := \frac{X_1 + \dots + X_n - nm}{\sigma\sqrt{n}}$$

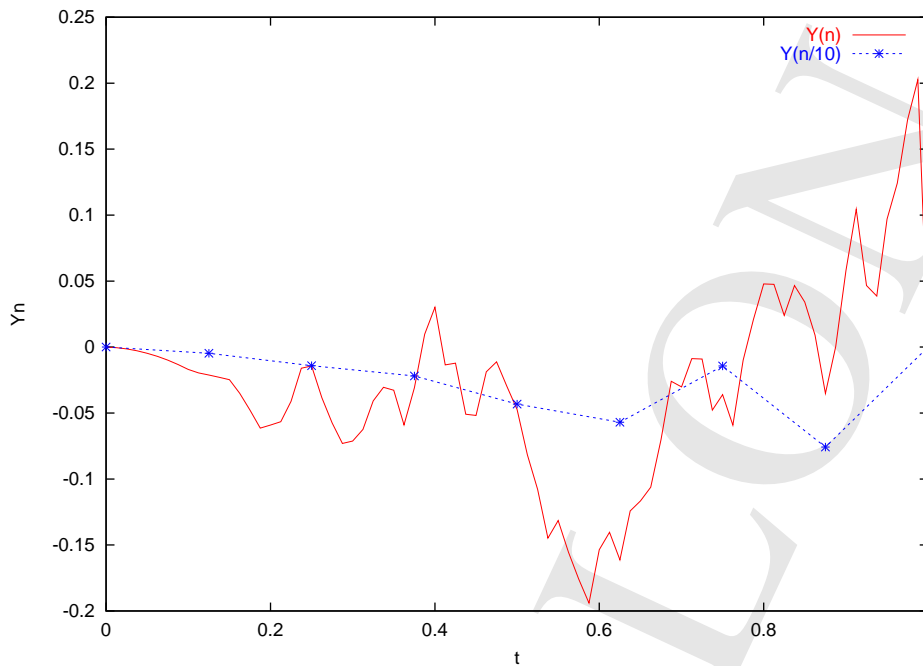


FIG. 2.3 – Illustration du théorème de Donsker.

et  $\Phi$  la fonction de répartition de la gaussienne standard  $\mathcal{N}(0,1)$ . Le théorème central limite (TCL) indique que pour tout  $t \in \mathbb{R}$  :

$$\lim_{n \rightarrow +\infty} |G_n(t) - \Phi(t)| = 0$$

Le théorème suivant donne une estimée uniforme sur la vitesse de cette convergence, cf. [Pet95] pour une preuve :

**Théorème 2.3.1 (Berry-Essen).** *Soit  $(X_n, n \in \mathbb{N}^*)$  une suite de v.a.r. i.i.d. de variance  $\sigma^2 > 0$  et possédant un moment d'ordre 3. Alors on a :*

$$\sup_{t \in \mathbb{R}} |G_n(t) - \Phi(t)| \leq \frac{C\rho}{\sigma^3 \sqrt{n}},$$

où  $\rho := \mathbf{E}(|X_1 - \mathbf{E}(X_1)|^3)$  et  $C$  est une constante positive qui ne dépend pas de la loi des  $X_n$ .

Il existe des versions dites « locales », du TCL, qui donnent des estimées sur les densités plutôt que sur les fonctions de répartition. Elles sont plus fortes que celles du type « Berry-Essen ». La plus élémentaire est sans doute la version forte du théorème de Moivre-Laplace 2.2.2, dont on trouvera une preuve faisant appel à la formule de Stirling dans [DCD82b, thm 2.2.4 page 35] :

**Théorème 2.3.2 (Théorème de Moivre-Laplace « fort »).** *Soit  $(S_n, n \in \mathbb{N}^*)$  des v.a. i.i.d. de loi binomiale  $\mathcal{B}(n, p)$  de taille  $n$  et de paramètre  $p \in ]0, 1[$ . On pose  $q := 1 - p$ . Alors, pour tout couple  $(a, b) \in \mathbb{R}^2$  avec  $a < b$ , on a :*

$$\sup_{x \in I_n \cap [a, b]} \left| \sqrt{npq} \mathbb{P}\left(\frac{S_n - np}{\sqrt{npq}} = x\right) - (2\pi)^{-1/2} e^{-x^2/2} \right| \xrightarrow[n \rightarrow +\infty]{} 0,$$



où  $I_n$  désigne le support de la v.a. discrète  $(S_n - np)/\sqrt{npq}$ .

## 2.4 Vitesse de convergence dans le théorème de Glivenko-Cantelli

Reprenons à présent le cadre du Théorème de Glivenko-Cantelli 2.1.4 page 43 et supposons que  $F$  est continue, et donc que  $\mu$  n'a pas de masses ponctuelles. Alors, on montre que :

$$\mathcal{L}\left(\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|\right) \xrightarrow[n \rightarrow +\infty]{\text{étr.}} \mathcal{L}_{\mathbf{F}}, \quad (2.2)$$

où  $\mathcal{L}_{\mathbf{F}}$  est la loi de probabilité sur  $\mathbb{R}_+$ , appelée « distribution de Kolmogorov-Smirnov », dont la fonction de répartition  $\mathbf{F} : \mathbb{R} \rightarrow [0, 1]$  est donnée par :

$$\mathbf{F}(u) := \begin{cases} 0 & \text{si } u \leq 0, \\ 1 + 2 \sum_{k=1}^{+\infty} (-1)^k e^{-2k^2 u^2} & \text{si } u > 0. \end{cases} \quad (2.3)$$

La bibliothèque Stixbox fournit la fonction Matlab `pks` qui permet de la calculer. Comme nous allons le voir ultérieurement dans la section 3.4.3 page 73, cette convergence en loi permet de construire un test d'adéquation appelé *test de Kolmogorov-Smirnov*.

Si  $(B_t, t \in [0, 1])$  est un mouvement brownien standard, on appelle *pont brownien sur  $[0, 1]$*  le processus  $(X_t, 0 \leq t \leq 1) := (B_t - tB_1, 0 \leq t \leq 1)$ . Il doit son nom au fait qu'il s'interprète comme un « passage brownien » entre les deux extrémités de l'intervalle  $[0, 1]$  en lesquelles il est nul. On montre que la v.a.r.  $\sup_{0 \leq t \leq 1} |X_t|$  a pour loi  $\mathcal{L}_{\mathbf{F}}$ . De plus, lorsque la fonction de répartition  $F$  de  $\mu$  est continue, on a :

$$\sqrt{n} \sup_{x \in \mathbb{R}} (F_n(x) - F(x)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \sup_{0 \leq t \leq 1} X_t, \quad (2.4)$$

et que :

$$\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \sup_{0 \leq t \leq 1} |X_t|. \quad (2.5)$$

La convergence en loi (2.5) n'est qu'une réécriture de (2.2). En voici une illustration en Matlab, dont la sortie graphique est donnée par la figure 2.4 :

```
% Illustration de la vitesse de convergence dans
% le théorème de Glivenko-Cantelli.
% Tracé des fonctions de répartitions à n fixé.
% Fait appel à Stixbox pour rnorm, pnorm et pks
clf; clear;
n=input('Taille des échantillons (par exemple 30)=? ');
p=input('Nombre d' échantillons (par exemple 30)=? ');
for i=1:p, % Parcours des échantillons
    % On simule un échantillon de taille n de loi normale
    X=rnorm(n);
    % Fonction de Répartition Empirique de l'échantillon
    [Effectifs, Classes]=hist(X,n);
    FRE=cumsum(Effectifs)/n;
    % Calcul de la norme uniforme de l'écart
    % à la fonction de répartition théorique.
    du=range(X)/n; % i.e. (max(X)-min(X))/n
    Z(i)=sqrt(n)*max(abs(FRE-pnorm(Classes+du/2)))
```

```

end
% Tracé
% Fonction de Répartition Simulée
[Effectifs , Classes]=hist(Z,p);
FRS=cumsum(Effectifs)/p;
% Tracé de la fonction de répartition empirique
% et de celle du sup de la valeur absolue du pont brownien
% appelée aussi distribution de Kolmogorov-Smirnov.
figure
du=range(Z)/p;
plot(Classes+du/2,FRS,'b+-');
hold on;
Abscisses=[min(Z)+du/2:du:max(Z)+du/2];
plot(Abscisses , pks(Abscisses) , 'r-');
legend('Fonction_de_repartition_simulee',\
'Fonction_de_repartition_de_Kolmogorov-Smirnov',4);

```

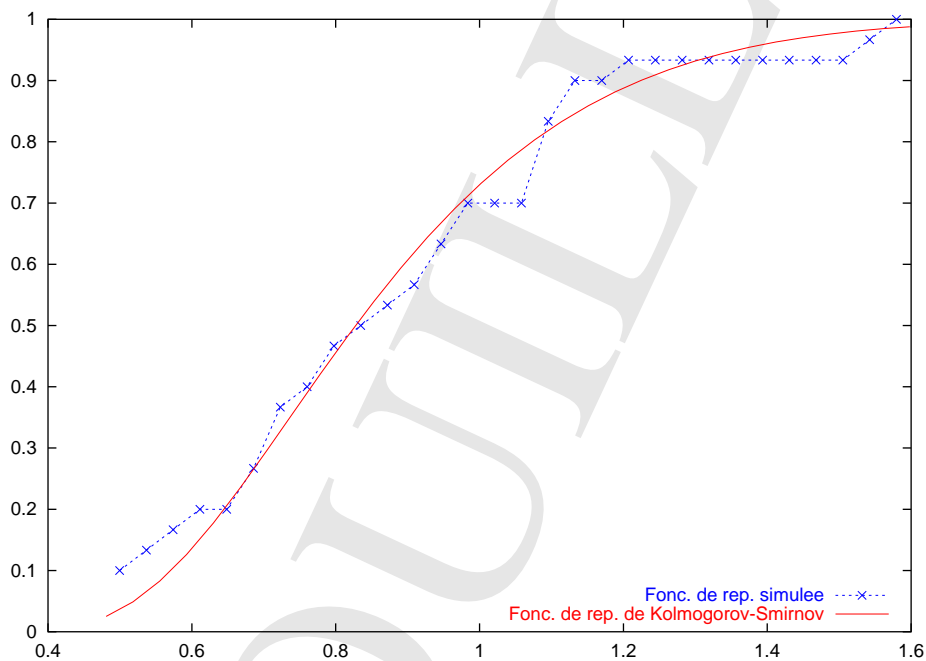


FIG. 2.4 – Convergence vers la distribution de Komogorov-Smirnov pour le théorème de Glivenko-Cantelli (loi normale).

## 2.5 Principes de Grandes Déviations

La loi des grands nombres indique que la moyenne arithmétique  $S_n$  de v.a. i.i.d. de loi  $\mu$ , intégrables, à valeurs dans  $\mathbb{R}^n$ , se « concentre » autour de leur moyenne  $m_\mu$ . Ainsi, on s'attend à ce que la probabilité de l'évènement  $\{S_n \in A\}$  tende vers 0 lorsque  $A$  est « loin » de la moyenne  $m_\mu$ . Prenons pour simplifier

$m_\mu = 0$ ,  $n = 1$  et  $A = [r, +\infty[$ . On a pour tout  $\theta \in \mathbb{R}^+$ , d'après l'inégalité de Markov :

$$\begin{aligned} \mathbb{P}(S_n \in [r, +\infty[) &= \mathbb{P}(\exp(\theta(X_1 + \dots + X_n)) \geq nr\theta) \\ &\leq \exp(-nr\theta - \log \mathbf{E}(\exp(\theta(X_1 + \dots + X_n)))) \\ &= \exp(-n(r\theta - \log \mathbf{E}(\exp(\theta X_1)))). \end{aligned}$$

Cette inégalité étant valable pour tout  $\theta > 0$ , on s'attend donc à un comportement du type :

$$\mathbb{P}(S_n \in [r, +\infty[) \simeq \exp\left(-n \sup_{\theta} (r\theta - \log \mathbf{E}(\exp(\theta X_1)))\right).$$

Le théorème de Cramér confirme cette intuition. Il affirme que si  $\mu$  possède des moments exponentiels de tous ordres, c'est-à-dire que  $\mathbf{E}_\mu(\exp(\theta \cdot X_1)) < +\infty$  pour tout  $\theta \in \mathbb{R}$ , alors on a pour tout borélien  $A$  :

$$-\inf_{int(A)} \Lambda_\mu^* \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(S_n \in A) \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(S_n \in A) \leq -\inf_{adh(A)} \Lambda_\mu^*,$$

où  $\Lambda_\mu^*$  désigne la transformée de Cramér<sup>2</sup> de  $\mu$ , définie par :

$$\Lambda_\mu^*(x) := \sup_{\theta \in \mathbb{R}^n} \left[ x \cdot \theta - \log \int e^{z \cdot \theta} d\mu(z) \right].$$

On dit alors que la suite de v.a. satisfait à un *principe de grandes déviations* par rapport à leur moyenne (PGD), de *vitesse*  $1/n$  et de *fonction de taux*  $\Lambda_\mu^*$ . On montre que la fonction de taux est positive, semi-continue inférieurement, convexe, qu'elle décroît sur  $] -\infty, m_\mu]$ , croît sur  $[m_\mu, +\infty[$ , qu'elle tend vers  $+\infty$  en  $\pm\infty$  et qu'elle admet un unique minimum en  $m_\mu$  pour lequel elle est nulle. Quelques exemples de transformées de Cramér sont donnés dans la table 2.1.

Bernoulli $p\delta_1 + (1-p)\delta_0$ avec $0 < p < 1$	$x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$ si $0 \leq x \leq 1$ et $+\infty$ sinon
Poisson de paramètre $\lambda > 0$	$\lambda - x + x \log \frac{x}{\lambda}$ si $x > 0$ et $+\infty$ sinon
Exponentielle de paramètre $\lambda > 0$	$\lambda x - 1 - \log(\lambda x)$ si $x > 0$ et $+\infty$ sinon
Gaussienne de moyenne $m$ et de variance $\sigma^2 > 0$	$(x - m)^2 / 2\sigma^2$

TAB. 2.1 – Quelques transformées de Cramér

De nombreux autres théorèmes autour des PGD existent, notamment celui de Sanov qui concerne la convergence de la mesure empirique vers la loi  $\mu$  :

**Théorème 2.5.1 (Théorème de Sanov).** *Soit  $(X_n, n \in \mathbb{N}^*)$  une suite de v.a. i.i.d. intégrables, de loi commune  $\mu$  sur  $\mathbb{R}^n$ . On note  $(L_n, n \in \mathbb{N}^*)$  les mesures empiriques associées et on munit l'ensemble  $\mathcal{P}(\mathbb{R}^n)$  des lois de probabilité sur  $\mathbb{R}^n$  de la topologie de la convergence étroite (i.e. en loi). Alors, pour tout  $A \subset \mathcal{P}(\mathbb{R}^n)$ , on a :*

$$-\inf_{\nu \in int(A)} \mathbf{Ent}(\nu | \mu) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(L_n \in A) \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} \log \mathbb{P}(L_n \in A) \leq -\inf_{\nu \in adh(A)} \mathbf{Ent}(\nu | \mu),$$

où  $\mathbf{Ent}(\nu | \mu)$  désigne l'entropie relative<sup>3</sup> de  $\nu$  par rapport à  $\mu$ , définie par :

$$\mathbf{Ent}(\nu | \mu) := \begin{cases} \mathbf{E}_\mu \left( \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} \right) & \text{si } \nu \ll \mu, \\ +\infty & \text{sinon.} \end{cases} \quad (2.6)$$

<sup>2</sup>Ça n'est rien d'autre que la transformée de Legendre du logarithme de la transformée de Laplace de  $\mu$ .

<sup>3</sup>On parle également d'entropie de Kullback-Leibler.

Le symbole  $\ll$  désigne l'absolue continuité. On peut retrouver le théorème de Cramér à partir d'un théorème du « type Sanov » pour les lois de probabilité sur  $\{1, \dots, n\}$ . L'entropie relative se muant en transformée de Cramér.

*Exemple 2.5.2.* Considérons une suite  $(X_n, n \in \mathbb{N}^*)$  de v.a. i.i.d. de loi de Bernoulli sur  $\{0, 1\}$  de paramètre  $p \in ]0, 1[$ . Soit  $r$  un réel positif. On note  $S_n := (X_1 + \dots + X_n)/n$  leur moyenne arithmétique. La v.a.  $S_n$  est de moyenne  $p$ . Comme  $nS_n$  suit une loi binomiale  $\mathcal{B}(n, p)$  de taille  $n$  et de paramètre  $p$ , on a, en notant  $F$  la fonction de répartition de  $\mathcal{B}(n, p)$  :

$$\log \mathbb{P}(S_n - p \geq r) = \log(1 - F(n(p+r))).$$

D'autre part, le TCL indique que la loi de  $\sqrt{n}(S_n - p)/\sqrt{pq}$  est approximativement  $\mathcal{N}(0, 1)$ . On a donc, en notant  $\Phi$  la fonction de répartition de la loi  $\mathcal{N}(0, 1)$  :

$$\log \mathbb{P}(S_n - p \geq r) \simeq \log \left( 1 - \Phi \left( r \sqrt{\frac{n}{pq}} \right) \right) = \log \Phi \left( -r \sqrt{\frac{n}{pq}} \right).$$

Enfin, d'après le théorème de Cramér, on a par le PGD et par la formule explicite pour la transformée de Cramér pour la loi de Bernoulli donnée par la table 2.1 :

$$\log \mathbb{P}(S_n - p \geq r) \simeq -n \left( (p+r) \log \frac{p+r}{p} + (1-p-r) \log \frac{1-p-r}{1-p} \right).$$

Voici un code Matlab illustrant ces trois estimées, dont la sortie graphique est donnée par la figure 2.5 :

```

%% Comparaison entre l'estimée de P(Sn - p >= r) par TLC et par
%% Principe de Grandes Déviations de Cramér lorsque Sn est la
%% moyenne arithmétique de v.a. de Bernoulli i.i.d.
%% de paramètre p. Les paramètres du programme sont p et r.
%% Il faut que 0 <= p+r <= 1
%% Exemple : pour (p,r)=(.975, .0125),
%% le PGD est meilleur que la LGN pour n >~ 3500
%% Fait appel à StixBox pour pnorm & pbinom
clf; clear
p=input('Paramètre p de la loi de Bernoulli = ');
r=input('Paramètre r de l''événement = ');
q=1-p;
x=p+r;
N=[200:20:5000];
for n=N,
    logpb(n)=log(1-pbinom(n*x, n, p))/n;
    tcl(n)=log(pnorm(-r*sqrt(n/(p*q)), 0, 1))/n;
    pgd(n)=-((x*log(x/p)+(1-x)*log((1-x)/q)));
end
plot(N, logpb(N), 'k', N, tcl(N), 'r-', N, pgd(N), 'b-');
legend('Log_proba', 'TCL', 'PGD', 4);
xlabel('Taille de l''échantillon');
ylabel('Approximations de log(P(Sn > p+r))');

```

## 2.6 Loi du logarithme itéré

À mi-chemin entre la LGN et le TLC, la loi du logarithme itéré précise ce qui se passe lorsque l'on divise la somme partielle d'une suite de v.a.r. i.i.d. par une suite réelle de vitesse intermédiaire entre  $\sqrt{n}$

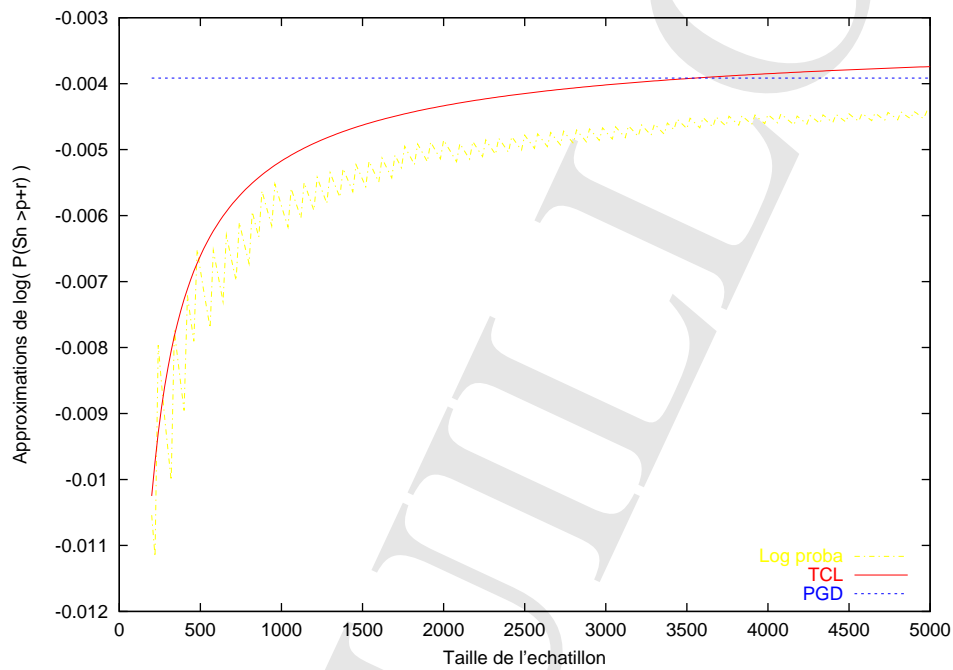


FIG. 2.5 – Comparaison des approximations par TCL et PGD pour un écart à la moyenne de la loi binomiale.

et  $n$ . En effet, si  $(X_n, n \in \mathbb{N}^*)$  est une suite de v.a.r. i.i.d. de loi  $\mu$ , intégrables, de moyenne  $m$  et de variance  $\sigma^2$ , alors on a  $\mu$ -presque sûrement :

$$\overline{\lim}_{n \rightarrow +\infty} Y_n = +\sigma \quad \text{et} \quad \underline{\lim}_{n \rightarrow +\infty} Y_n = -\sigma,$$

où

$$Y_n := \frac{X_1 + \cdots + X_n - nm}{\sqrt{2n \log \log n}}.$$

D'après Strassen, on a plus précisément :

$$\text{adh}((Y_n)_{n \in \mathbb{N}^*}) = [-\sigma, +\sigma] \quad \text{et} \quad \text{dist}(Y_n, [-\sigma, +\sigma]) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} 0.$$

De nombreuses sophistications existent, avec des suites  $(a_n)_n$  quelconques. Il est sans doute vain de vouloir illustrer par une simulation la convergence des valeurs d'adhérence de  $(Y_n, n \in \mathbb{N}^*)$  car la propriété est « trop asymptotique » pour être vue sur un échantillon de taille finie. En effet, ici, la suite  $(Z_n, n \in \mathbb{N}^*)$  qui converge p.s. par exemple vers  $+\sigma$  est donnée par la définition de  $\lim$  :

$$Z_n := \inf_{n \geq 1} \sup_{k \geq n} \frac{X_1 + \cdots + X_k - km}{\sqrt{2k \log \log k}},$$

qui fait déjà intervenir un nombre infini de  $X_i$  et n'est donc pas accessible numériquement puisque les échantillons dont on dispose sont finis. Voici un petit bout de code pour s'en convaincre (à exécuter plusieurs fois...)

```
%%% Tentative vaine d'illustration de la loi du logarithme itéré
clear
n=100000;
moy=1/2; sig=sqrt(1/12);
N=[1:n];
X=rand(1,n);
S=cumsum(X);
Y=(S-N*moy)/sqrt(2*N*log(log(N)));
clf; plot(N,Y,'b.',N,sig*ones(1,n),'r.-',-sig*ones(1,n),'r.-');
xlabel('Taille_de_l_echantillon');
```

En revanche, on peut visualiser la convergence vers 0 de la distance de la suite  $(Y_n, n \in \mathbb{N}^*)$  à l'intervalle  $[-\sigma, +\sigma]$ . Voici un code Matlab dont la sortie graphique est donnée par la figure 2.6 :

```
%%% Illustration de la décroissance vers zéro de la distance à
%%% l'intervalle [-sig,+sig] dans le théorème de Strassen
%%% (loi du logarithme itéré)
clear
n=1000;
n0=3;
moy=1/2; sig=sqrt(1/12);
N=[1:n];
S=cumsum(rand(1,n));
Y=(S(n0:n)-N(n0:n)*moy)/sqrt(2*N(n0:n)*log(log(N(n0:n))));
Z=((Y>sig).*(Y-sig)+(Y<-sig).*(-Y-sig))/sig;
clf;
plot(N(n0:n),Z,'b-.');
xlabel('Taille_n_de_l_echantillon');
ylabel('Distance(Y_n,[-sig,sig])/sig');
legend('Distance_dans_LLI_de_Strassen');
```

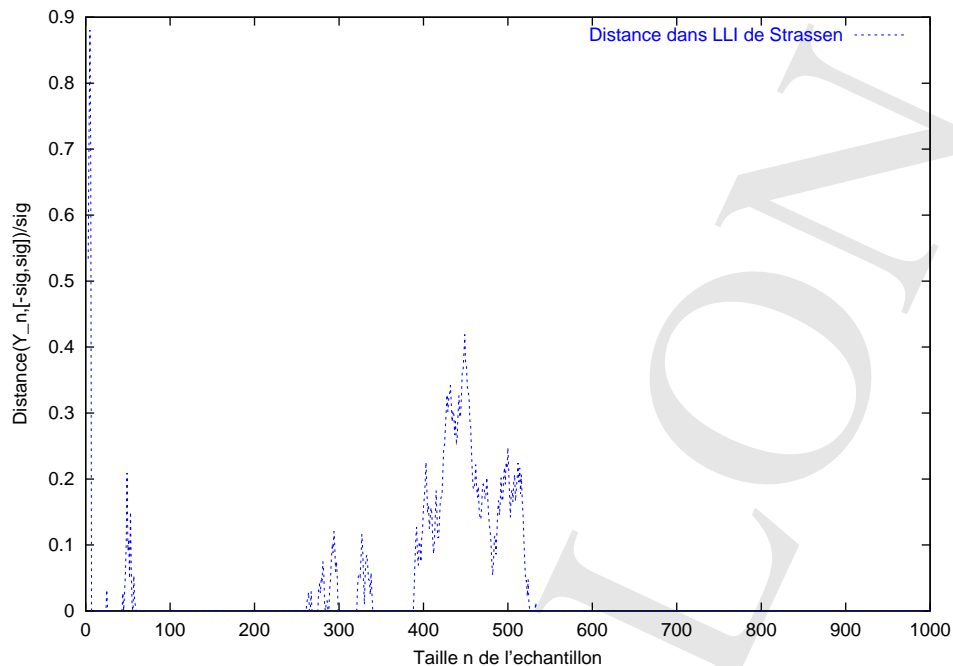


FIG. 2.6 – Illustration de la loi du log. itéré de Strassen (distance à  $[-\sigma, +\sigma]$ ).

*Remarque 2.6.1 (Sens des simulations).* Le code précédent pose le problème de ce qu'apporte une simulation numérique d'un point de vue rationnel. Contrairement au calcul formel, une simulation numérique ne prouve rien. C'est plutôt la théorie qui permet de justifier son utilisation dans la pratique. En revanche, la simulation peut s'avérer très utile pour expérimenter et guider l'intuition. Enfin, elle permet parfois de mettre jour des comportements étranges ou inattendus qui font progresser la théorie en posant des questions nouvelles.

En fait, les « preuves » de théorèmes obtenues par calcul formel sur ordinateurs ne font pas l'unanimité dans la communauté des mathématiciens. Malgré les problèmes philosophiques posés, cela n'est peut-être après tout qu'une manifestation d'un certain conservatisme. Les progrès futurs de l'informatique nous promettent sans doute de nombreux débats sur le sujet... À la fin du dix-neuvième siècle, les preuves non constructives, fussent-elles dues à Hilbert, étaient jugées irrecevables par de nombreux mathématiciens, comme Kronecker par exemple. Les choses ont un peu changé depuis, comme vous l'avez constaté au cours de votre cursus mathématique!

## 2.7 Convergence des histogrammes

Simuler  $n$  réalisations indépendantes d'une loi  $\mu$  sur  $\mathbb{R}$  correspond à la donnée, pour un  $\omega \in \Omega$ , de la suite  $X_1(\omega), \dots, X_n(\omega)$  où les  $(X_n, n \in \mathbb{N}^*)$  sont des variables aléatoires indépendantes de même loi  $\mu$ . On note ici  $F$  la fonction de répartition de  $\mu$ ,  $L_n$  la mesure empirique et  $F_n$  la fonction de répartition empirique.

Si l'on considère un intervalle  $I_n = [a_n, b_n]$  de  $\mathbb{R}$  dont les bornes peuvent dépendre a priori de  $n$ , le

théorème de Glivenko-Cantelli (2.1) page 43 nous permet d'écrire, pour  $\mu$ -presque tout  $\omega$  :

$$\overline{\lim}_{n \rightarrow \infty} |L_n(I_n)(\omega) - \mu(I_n)| \leq \overline{\lim}_{n \rightarrow \infty} (|F_n(b_n) - F(b_n)| + |F_n(a_n) - F(a_n)|) = 0.$$

On rappelle que la quantité

$$L_n(I_n)(\omega) = \frac{\#\{i \leq n, X_i(\omega) \in I_n\}}{n}$$

représente la fréquence empirique de l'intervalle  $I_n$ . Partant des réalisations  $X_1(\omega), \dots, X_n(\omega)$  obtenues par simulation, on subdivise l'intervalle  $[\min_i X_i(\omega), \max_i X_i(\omega)]$  en  $m$  sous-intervalles disjoints  $A_{i,n,m}(\omega)$ ,  $i = 1, \dots, m$  de longueurs égales. L'histogramme en  $m$  classes associé aux valeurs simulées correspond au tracé de la fonction constante par morceaux  $H_{n,m}(\omega)$  définie par

$$H_{n,m}(\omega) := \sum_{i=1}^m \#\{j, X_j(\omega) \in A_{i,n,m}(\omega)\} I_{A_{i,n,m}(\omega)}.$$

Les  $A_{i,n,m}(\omega)$  étant de même longueur  $l_{n,m}(\omega)$ , la surface de l'histogramme est donnée par la formule  $n(\max_i X_i(\omega) - \min_i X_i(\omega))/m$ . On définit à présent la fonction  $D_{n,m}(\omega)$  comme la densité de probabilité sur  $\mathbb{R}$  obtenue en renormalisant l'histogramme

$$D_{n,m}(\omega) = \frac{m}{n(\max_i X_i(\omega) - \min_i X_i(\omega))} H_{n,m}(\omega).$$

Les  $A_{i,n,m}(\omega)$  étant de même longueur, on a aussi

$$D_{n,m}(\omega) := \frac{1}{n} \sum_{i=1}^m \frac{\#\{j, X_j(\omega) \in A_{i,n,m}(\omega)\}}{l_{n,m}(\omega)} I_{A_{i,n,m}(\omega)},$$

et  $D_{n,m}(\omega)$  apparaît alors comme une combinaison linéaire convexe de densités uniformes sur les  $A_{i,n,m}(\omega)$ . On voit bien que la mesure de probabilité de densité  $D_{n,m}(\omega)$  et la mesure empirique  $L_n(\omega)$  affectent la même probabilité aux  $A_{i,n,m}(\omega)$ , et l'on peut montrer que pour tout  $n$ , la mesure de probabilité de densité  $D_{n,m}(\omega)$  converge étroitement vers la mesure empirique  $L_n(\omega)$ .

Il existe donc une suite  $m_n(\omega)$  telle que la mesure de probabilité de densité  $D_{n,m_n(\omega)}(\omega)$  converge étroitement vers  $\mu$  quand  $n \rightarrow \infty$ .

### 2.7.1 Estimées uniformes sur les densités

Nous avons vu que l'on avait convergence faible. Il est plus délicat d'obtenir des estimées uniformes sur la proximité des densités de  $\mu$  et de  $D_{n,m}$ . Les  $A_{i,n,m}(\omega)$  sont tous de même longueur  $l_{n,m}(\omega)$ , et l'on peut donc écrire, en omettant la dépendance à  $\omega$  dans les notations

$$D_{n,m} = \frac{1}{l_{n,m}} \sum_{i=1}^m L_n(A_{i,n,m}) 1_{A_{i,n,m}}.$$

Si l'on définit à présent la densité approchée en  $m$  escaliers de  $\mu$  par

$$f_m := \frac{1}{l_{n,m}} \sum_{i=1}^m \mu(A_{i,n,m}) 1_{A_{i,n,m}},$$

on peut écrire

$$|D_{n,m} - f_m| \leq \frac{2 \sup_{t \in I} |F_n(t) - F(t)|}{l_{n,m}} \sum_{i=1}^m 1_{A_{i,n,m}} \leq \frac{2 \sup_{t \in I} |F_n(t) - F(t)|}{l_{n,m}}.$$



Lorsque  $\mu$  a pour support un intervalle compact  $I$ ,  $m l_{n,m} = \max_{1 \leq i \leq n} X_i - \min_{1 \leq i \leq n} X_i$  tend vers  $|I|$  presque sûrement quand  $n$  tend vers l'infini (pourquoi?). Malheureusement, le théorème de Glivenko-Cantelli ne donne pas la vitesse de convergence de  $\sup_{t \in I} |F_n(t) - F(t)|$  vers 0. Pour aller plus loin, il faudrait faire appel à un théorème plus fort du type Berry-Essen par exemple (hors programme), qui donne une vitesse de l'ordre de  $\sqrt{n}$ , ce qui conduit à prendre  $m = m_n = \sqrt{n}$  pour obtenir la convergence uniforme de  $|D_{n,m} - f_m|$  vers 0.

Notons que si  $\mu$  admet une densité  $f$  sur  $I$  de classe  $\mathcal{C}^2$ , alors, le théorème de Heine entraîne que  $f_m$  converge uniformément vers  $f$  lorsque  $m$  tendent vers l'infini (méthodes des rectangles!).

### 2.7.2 Un exemple avec une illustration du théorème central limite

Soient  $(U_n, n \in \mathbb{N}^*)$  une suite de v.a. i.i.d. suivant une loi uniforme sur  $[0, 1]$ . On rappelle que la loi uniforme sur  $[0, 1]$  a pour moyenne  $1/2$  et variance  $1/12$ . En vertu du théorème central limite, la variable aléatoire

$$\frac{U_1 + \dots + U_p - p/2}{\sqrt{p/12}} \quad (2.7)$$

converge en loi vers la loi normale centrée réduite lorsque  $p \rightarrow \infty$ . Pour  $p = 12$ , la formule précédente a une expression très simple ( $U_1 + \dots + U_{12} - 6$ ) et fournit déjà une « bonne » approximation de la loi normale.

On effectue alors  $n$  simulations indépendantes  $X_1(\omega), \dots, X_n(\omega)$  de la v.a. (2.7) et l'on trace sur un même graphique l'histogramme  $H_{n,m_n}$  et la densité de la loi normale centrée réduite correctement dilatée pour être comparée à l'histogramme. Pour obtenir un joli graphique, on choisit  $m_n$  plus petit que  $n$ , par exemple  $m_n = \sqrt{n}$ . Ce facteur  $\sqrt{n}$  ne représente pas la normalisation du Théorème Central Limite mais plutôt la vitesse de convergence dans le théorème de Glivenko-Cantelli.

```
% Illustration du TLC avec un échantillon de lois uniformes
% Fait appel à StixBox pour dnorm & histo
clf      % Nettoyage des graphiques
clear    % Nettoyage des variables
i=0;
p=input('Nombre de v. a. uniformes (par exemple 12) = ');
n=input('Taille de l' échantillon (par exemple 1000) = ');
% On simule n fois la variable Z:=U_1+...+U_p,
X=rand(p,n);
Z=sum(X)-p/2; % Centrage
Z=Z/sqrt(p/12); % Réduction
% Z est proche de la N(0,1) d'après le TCL
% Nombre de classes pour histogramme
nc=sqrt(n);
% On trace l'histogramme normalisé en nc classes
% E contient les effectifs pour nos classes
% C contient les extrémités des classes
[E,C]=histo(Z,nc,0,1);
hold on;
% On trace la densité normale
plot(C,dnorm(C),'r-');
legend('Histogramme','Loi normale');
```

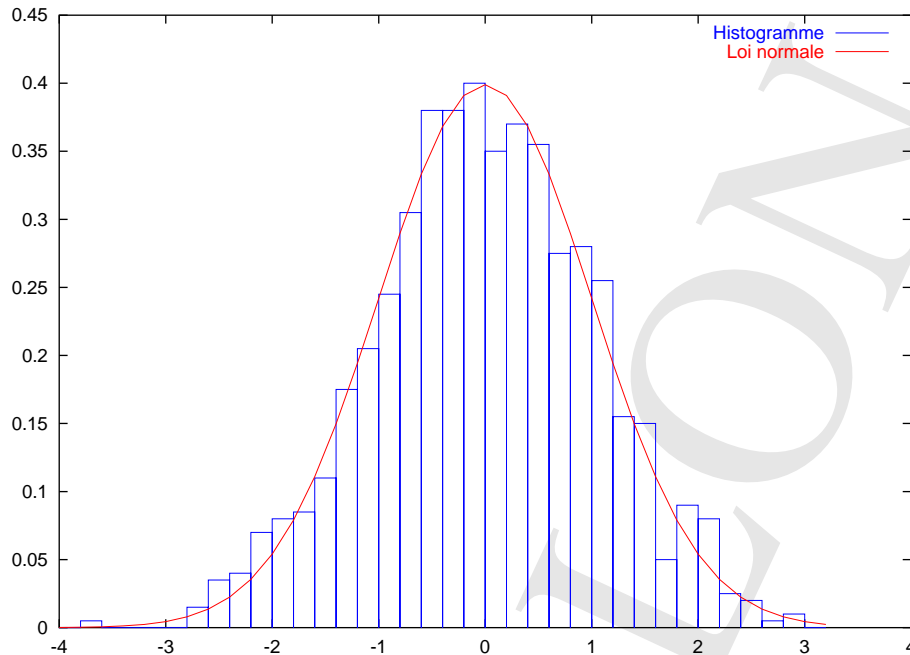


FIG. 2.7 – Approximation de la loi normale par TCL (lois uniformes).

## 2.8 Le théorème de Wigner

Le théorème de Wigner, qui ressemble à la loi des grands nombres, est ignoré par la plupart des cours de probabilités. Il est pourtant très simple à énoncer et sa démonstration n'est pas si difficile et fait appel à un peu de combinatoire. Il en existe de nombreuses versions avec des hypothèses plus faibles que celles requises ici. Considérons un tableau symétrique de taille infinie dont les entrées sont des v.a.r. :

$$\begin{pmatrix} w_{1,1} & w_{1,2} & \cdots \\ w_{2,1} & w_{2,2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

On note  $W_n$  la matrice aléatoire carrée symétrique réelle de taille  $n$  obtenue en tronquant le tableau à partir de son coin supérieur gauche. Les valeurs propres de  $W_n$  sont réelles d'après le théorème spectral.

**Théorème 2.8.1 (Wigner).** *Supposons que les v.a.  $(w_{i,j}, 1 \leq i < j)$  sont i.i.d. de variance finie  $\sigma^2$  et que les v.a.  $(w_{i,i}, 1 \leq i)$  sont i.i.d. Si  $(\lambda_{n,1}(\omega), \dots, \lambda_{n,n}(\omega))$  désigne le spectre de  $n^{-1/2} W_n(\omega)$ , alors, pour presque tout  $\omega$ , la mesure empirique*

$$\frac{1}{n} \sum_{i=1}^n \delta_{\lambda_{i,n}(\omega)}$$

*converge étroitement (i.e. en loi) vers la loi du demi-cercle de paramètre  $\sigma$ , de densité par rapport à la mesure de Lebesgue  $dx$  sur  $\mathbb{R}$  :*

$$\frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2} \mathbb{I}_{\{-2\sigma \leq x \leq 2\sigma\}}(x).$$

Remarquons que les v.a.  $\lambda_{n,i}$  ne sont pas i.i.d. et on ne peut donc pas appliquer la loi des grands nombres. Voici un code Matlab pour illustrer ce théorème, dont la sortie graphique est donnée par la figure 2.8 :

```

%%% Illustration du théorème de Wigner
clf;
clear;
dim=100;      % Dimension des matrices
nreal=100;    % Nombre de réalisations i.i.d. de ces matrices
spectres=zeros(dim*nreal,1);
for i=[0:nreal-1],
    M=randn(dim,dim);
    spectres(1+i*dim:i*dim+dim)=eig(M-triu(M,1)+tril(M,-1)')/sqrt(dim);
end
nc=sqrt(nreal*dim);
% On utilise l'histogramme normalisé en nc classes
% E contient les effectifs pour nos classes
% C contient les extrémités des classes
[E,C]=histo(spectres,nc,0,1);
hold on;
% On trace la densité de la loi semi-circulaire
plot(C,sqrt(4-C.^2)/(2*pi),'r-');
legend('Histogramme','Loi semi-circulaire');

```

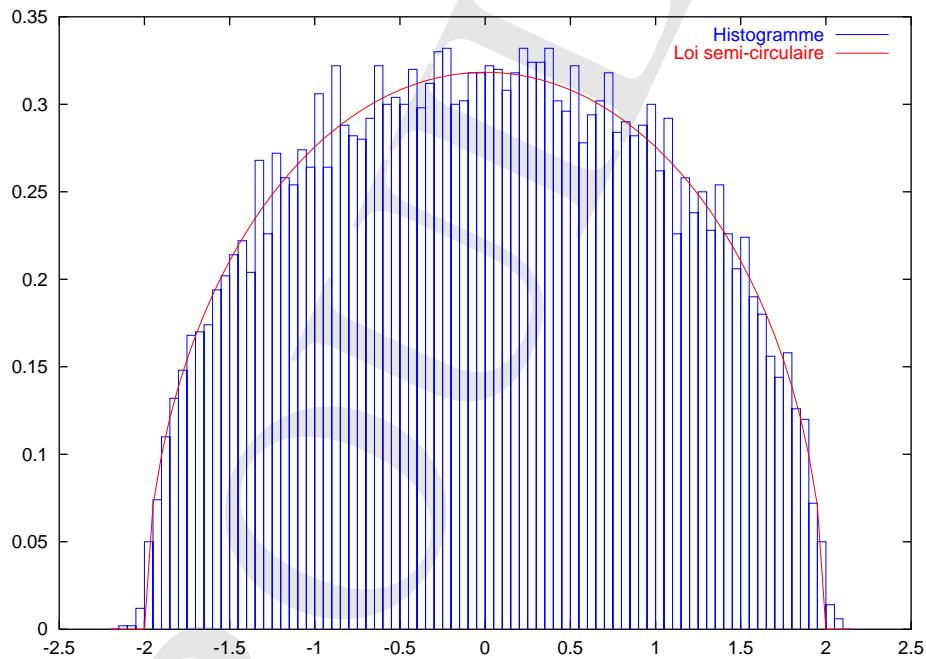


FIG. 2.8 – Illustration du théorème de Wigner.