Les parties marquées d'une astérisque (*) sont facultatives.

Chapitre 5. Théorèmes limites

1. Motivation
2. Convergence (divers types)
3. Loi faible des grands nombres
4. Loi forte des grands nombres*
5. Théorème Central limite

1 Motivation

Une part essentielle de l'inférence statistique est le comportement des suites de variables aléatoires : c'est la théorie asymptotique (en probabilité).

Idée de base : étant donné une suite de variables aléatoires $X_1, \cdots, X_n$, que peut-on dire de son comportement quand n devient suffisamment grand (tend vers l'infini) ? En analyse mathématique, les choses sont simples : on dit que la suite numérique $\{x_n\}_{n\geq 1}$ converge vers un nombre x si pour tout $\epsilon > 0$, on a $|x_n - x| < \epsilon$ dès que n est suffisamment grand. Par exemple, si $x_1 = x_2 = \ldots = x_n = \ldots = x$, alors on a trivialement

$$\lim_{n\to\infty} x_n = x$$

Mais en probabilité, les choses sont un peu plus compliquées. Soit par exemple $X_1, \cdots, X_n$ iid $\sim N(0,1)$. Peut-on dire que

$$\lim_{n\to\infty} X_n = X \sim N(0,1) \ ?$$

Mais nous avons

$$P(X_n = X) \overset{PT}{=} \int_{\mathbb{R}} P(X_n = x) f_{X_n}(x)\, dx = 0 \qquad \forall n$$

Il faut donc introduire des définitions spécifiques (rappelons qu'une variable aléatoire est une fonction $\Omega \to \mathbb{R}$).

Deux cas importants seront à noter : **la Loi Faible des Grands Nombres** et **le Théorème Central Limite**.

Au-delà de l'aspect purement mathématique, pourquoi, d'un point de vue pratique, est-il important d'étudier le comportement de suites de variables aléatoires ? Souvent, on dispose d'un échantillon iid

$$X_1, \cdots, X_n$$

et on s'intéresse à une fonction

$$Y_n = \varphi(X_1, \cdots, X_n)$$

de cet échantillon. Des exemples courants sont la moyenne empirique

$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

ou la variance empirique

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} \left(X_k - \overline{X}_n \right)^2$$

On voudrait connaître, idéalement, la loi de Y_n (appelée distribution d'échantillonnage, cf. chapitre ultérieur) et, parfois on peut la trouver de manière exacte. Par exemple,

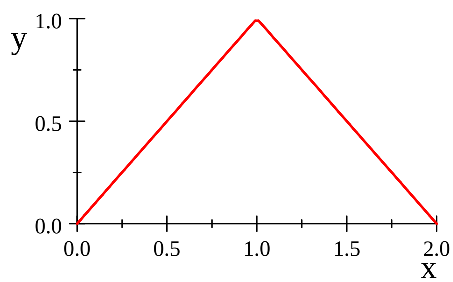
$$\text{si les } X_i \sim N\left(\mu, \sigma^2\right), \qquad \text{alors} \qquad \overline{X}_n \sim N\left(\mu, \sigma^2/n\right).$$

Malheureusement, la plupart du temps, il est impossible d'obtenir la loi de Y_n au vu de sa complexité. Ceci nous mène à voir son comportement asymptotique (quand la taille de l'échantillon augmente indéfiniment). On utilisera alors cette loi limite comme approximation de la loi de Y_n (par exemple, le théorème Central Limit nous dit que pour n assez grand, la loi binomiale peut être approximée par la loi normale). Mais même cette approche n'est parfois pas possible pour donner une approximation utile (taille de l'échantillon trop petite, problème trop compliqué, etc.). Dans ce cas, on utilise une simulation (méthode de Monte-Carlo) car la plupart des ordinateurs d'aujourd'hui sont assez puissants pour une telle approche. Enfin, signalons que dans les applications concrètes, on ne sait pas grand chose sur les variables X_k. Nous reviendrons sur ces sujets ultérieurement (cf. chapitre estimation). Ce chapitre-ci est plus concerné par l'aspect mathématique (outils théoriques) du problème. D'autres définitions (et résultats) seront données plus tard.

Exemple 1 *Soit $X_1, \cdots, X_n$ iid $\sim U[0;1]$ et soit $Y_n = X_1 + \cdots + X_n$. Trouver les lois de Y_2, de Y_3 et de Y_4.*

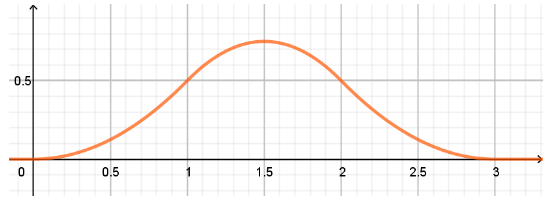
Réponses *Voici les densités respectives*

$$f_{Y_2}(y) = \begin{cases} y & si \quad y \in [0;1] \\ 2-y & si \quad y \in [1;2] \\ 0 & ailleurs \end{cases} \qquad (loi\ triangulaire)$$



$f_{Y_2}(y)$

$$f_{Y_3}(y) = \begin{cases} y^2/2 & si \quad y \in [0;1] \\ -y^2+3y-3/2 & si \quad y \in [1;2] \\ y^2/2-3y+9/2 & si \quad y \in [2;3] \\ 0 & ailleurs \end{cases}$$



$f_{Y_3}(y)$

$$f_{Y_4}(y) = \begin{cases} y^3/6 & si & y \in [0;1] \\ -y^3/2 + 2y^2 - 2y + 2/3 & si & y \in [1;2] \\ y^3/2 - 4y^2 + 10y - 22/3 & si & y \in [2;3] \\ -y^3/6 + 2y^2 - 8y + 32/3 & si & y \in [3;4] \\ 0 & ailleurs & \end{cases}$$

$f_{Y_4}(y)$

On se rend vite compte que plus n augmente, plus la loi devient compliquée. Il n'y a pas si longtemps, on a utilisé $Z = Y_{12} - 6$ pour simuler une variable $N(0.1)$: on génère 12 nombres aléatoires selon une loi uniforme sur $[0;1]$, on les additionne puis on soustrait 6 au total. Pouvez-vous donner une justification ?

Exemple 2 *Soit $X_1, \cdots, X_n$ iid (discrètes) distribuées selon la loi*

X	1	2	3	*Total*
p_X	$1/2$	$1/4$	$1/4$	1
	$E(X) = 7/4$	$Var(X) = 11/16$		

et prenons $Y_n = \sqrt[n]{X_1 \cdots X_n}$ (moyenne géométrique). Trouver la loi de $Y_2 = \sqrt{X_1 X_2}$, de $Y_3 = \sqrt[3]{X_1 X_2 X_3}$:$\frac{(\ln 2)^2 + (\ln 3)^2}{4} - \left(\frac{\ln 2 + \ln 3}{4}\right)^2 = \frac{3}{16}\ln^2 2 - \frac{1}{8}\ln 2 \ln 3 + \frac{3}{16}\ln^2 3 = 0.2212$: 0.44794

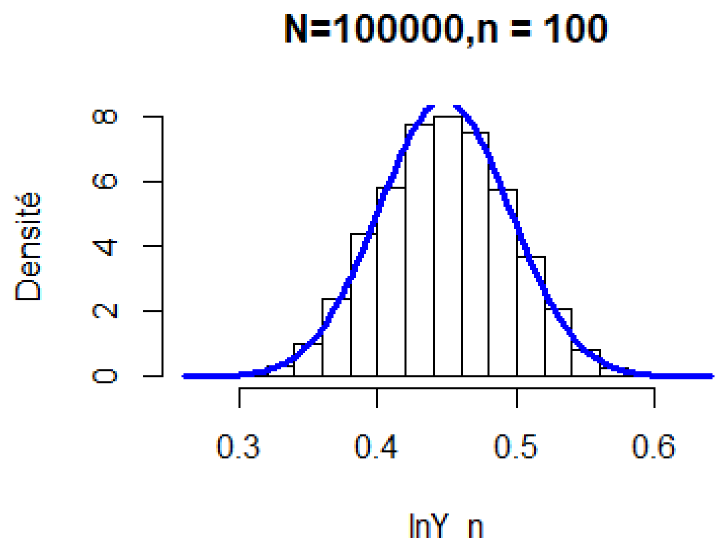
Y_2	1	$\sqrt{2}$	$\sqrt{3}$	2	$\sqrt{6}$	3	*Total*
p_{Y_2}	$4/16$	$4/16$	$4/16$	$1/16$	$2/16$	$1/16$	1

Y_3	1	$\sqrt[3]{2}$	$\sqrt[3]{3}$	$\sqrt[3]{4}$	$\sqrt[3]{6}$	2	$\sqrt[3]{9}$	$\sqrt[3]{12}$	$\sqrt[3]{18}$	3	*Total*
p_{Y_3}	$8/64$	$12/64$	$12/64$	$6/64$	$12/64$	$1/64$	$6/64$	$3/64$	$3/64$	$1/64$	1

Pour les mêmes raisons que dans l'exemple précédent, la loi de Y_n devient trop compliquée à mesure que la taille de l'échantillon augmente. Cependant, si on prend le log :

$$\ln Y_n = \frac{1}{n}\sum_{k=1}^{n} \ln X_k = \overline{\ln X}$$

on pense au Théorème Central Limite et donc, pour n assez grand, $\ln Y_n$ suit approximativement une loi normale $N\left(\mu = 0.447\,94, \sigma^2 = 0.221\,2/n\right)$. On peut également faire une simulation comme dans l'exemple précédent : on génère n (disons $n = 100$) nombres aléatoires $x_1, \cdots, x_n$ selon la loi des X_i, puis on prend leur moyenne géométrique $y_n = \sqrt[n]{x_1 \cdots x_n}$ et on répète le processus un grand nombre N de fois (disons $N = 100000$). L'histogramme des valeurs obtenues sera une approximation de la loi de Y_n.



1.1 Exercice

Exercice 1 *Soit $X_1, \cdots, X_n$ iid $\sim U[0;1]$ et soit $Y_n = \sqrt[n]{X_1 \cdots X_n}$. Trouver la loi de Y_2.*

Pour conclure cette introduction, si on s'intéresse à la probabilité d'un événement particulier $P(Y_n \in E)$, il faut trouver $\varphi^{-1}(E) = \{(x_1, \cdots, x_n) \in \mathbb{R}^n : \varphi(x_1, \cdots, x_n) \in E\}$, puis calculer la probabilité de cet événement. Ceci est généralement trop compliqué et il faut passer par une approximation ou une simulation.

2 Convergence (divers types)

2.1 Convergence en probabilité et loi faible des grands nombres

Définition 1 *Convergence en probabilité (ou stochastique). La suite de variables aléatoires $(X_n)_{n\geq 1}$ (toutes définies sur le même espace Ω) converge en probabilité vers une constante c si*

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} P(|X_n - c| \geq \epsilon) = 0$$

Plus généralement, la suite $(X_n)_{n\geq 1}$ converge en probabilité vers la variable aléatoire X si

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

Notation :

$$X_n \xrightarrow{P} c$$
$$X_n \xrightarrow{P} X$$

Remarque 1 *En d'autres termes,* $X_n \xrightarrow{P} X$ *si*

$$\forall \epsilon, \delta > 0, \quad \exists N(\epsilon, \delta) : n > N \Rightarrow P(|X_n - X| \geq \epsilon) < \delta$$

Ceci nous dit que l'événement $E_n = \{|X_n - X| < \epsilon\}$ *est réalisé avec une probabilité supérieure à* $1 - \delta$ *à partir d'un certain rang* N*. Ceci <u>ne nous dit pas</u> que les événements* E_n *sont réalisés <u>simultanément</u> avec une probabilité supérieure à* $1 - \delta$*. Cette exigence est satisfaite par la convergence presque sûre (plus loin).*

Remarque 2 *La convergence en probabilité est à la base de la justification de l'utilisation de la méthode des sondages pour estimer un pourcentage.*

Exemple 3 *Soit* $X_1, \ldots, X_n$ *iid* $\sim U[0;1]$ *et soit* $Y_n = X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}$*. Alors* $Y_n \xrightarrow{P} 0$*. En effet, soit* $\epsilon > 0$*. Alors, de toute évidence,*

$$P(|Y_n| \geq \epsilon) = P(Y_n \geq \epsilon) = 0 \qquad si\ \epsilon \geq 1$$

Prenons donc $0 < \epsilon < 1$*. Alors*

$$P(Y_n \geq \epsilon) = P(tous\ les\ X_i \geq \epsilon) \overset{ind.}{=} (1-\epsilon)^n \underset{n\to\infty}{\to} 0$$

Exemple 4 *Soit* $X_1, \ldots, X_n$ *iid* $\sim U[0;\theta]$ *et soit* $Y_n = X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}$*. Nous allons montrer que* $Y_n \xrightarrow{P} \theta$*. Soit* $\epsilon > 0$*. Alors,*

$$P(|Y_n - \theta| \geq \epsilon) = P(\theta - Y_n \geq \epsilon) = P(Y_n \leq \theta - \epsilon) = 0 \ si\ \epsilon > \theta$$

Prenons $0 < \epsilon \leq \theta$*. Alors*

$$P(Y_n \leq \theta - \epsilon) = P(tous\ les\ X_i \leq \theta - \epsilon) \overset{ind.}{=} \left(1 - \frac{\epsilon}{\theta}\right)^n \underset{n\to\infty}{\to} 0$$

Théorème 1 *<u>Loi faible des grands nombres</u> (LFGN). Soit* $X_1, \ldots, X_n$ *iid avec* $E(X_i) = \mu$ *et* $VAR(X_i) = \sigma^2$*. Posons* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$*. Alors* $\overline{X}_n \xrightarrow{P} \mu$*. En d'autres termes,*

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| \geq \epsilon\right) = 0$$

Preuve. On a (inégalité de Chebyshev) :

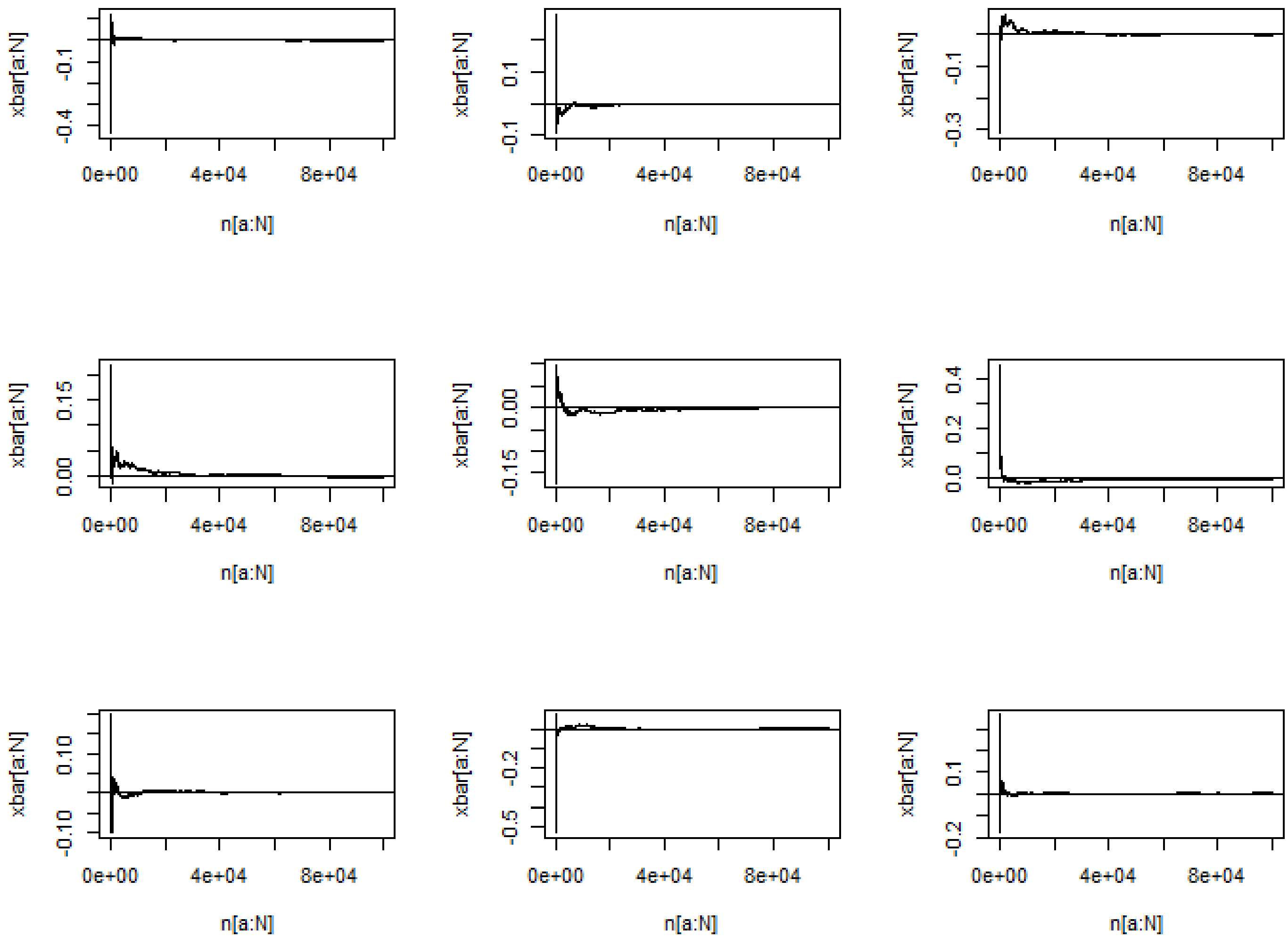
$$P\left(\left|\overline{X}_n - \mu\right| > \epsilon\right) \leq \frac{VAR\left(\overline{X}_n\right)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \underset{n\to\infty}{\to} 0$$

■

Remarque 3 *La condition d'existence de la variance n'est pas nécessaire. C'est la loi faible des grands nombres de Khinchine.*

Exemple 5 *Simulons la loi faible pour une variable normale centrée réduite. Voici un code R :*

```
## Générons N = 100000 tirages d'une loi normale centrée réduite et examinons
## les moyennes empiriques pour n = 1 :N. Faisons 9 expériences indépendentes
par(mfrow=c(3,3))
N <- 100000 ; a <- 10
for (i in 1 :9) {
  z <- rnorm(N,0,1)
  n <- c(1 :N)
  ## Calculons la somme cumulée et divisons par la taille de l'échantillon
 xbar<-cumsum(z)/n
 # Traçons le vecteur des N moyennes empiriques contre la taille de l'échantillon
 plot(n[a :N],xbar[a :N],type="l")
 # Traçons une droite horizontale en y = 0
 abline(h=0)
}
```



Exemple 6 *Méthode de Monte-Carlo. Le but de la méthode est de trouver une approximation de l'intégrale*

$$I(g) = \int_0^1 g(x)\,dx = E(g(X))$$

où $X \sim U[0;1]$. *On génère une suite iid* $X_i \sim U[0;1]$ $(i = 1, \ldots n)$ *et on calcule l'approximation*

$$I(g) \approxeq \widehat{I}(g) = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

car $\widehat{I}(g) \xrightarrow{P} I(g)$ *(en supposant* g *continue).*

Noter (pour ceux qui ont suivi un cours de méthodes numériques) qu'il y a une branche des mathématiques dédié à des calculs de ce genre (calcul scientifique, analyse numérique). Des logiciels (comme MATLAB) ou des languages informatiques (comme R ou Python) intègrent ces techniques de calcul.

Remarque 4 *Le fait que les X_i aient la même loi n'est pas une condition nécessaire. On a la version plus générale suivante (preuve similaire).*

Théorème 2 *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes (toutes définies sur le même espace Ω) telles que $E(X_i) = \mu_i$ et $VAR(X_i) = \sigma_i^2$. Posons*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

et

$$\overline{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$$

Si

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} \sigma_i^2 = 0$$

alors

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} P\left(\left|\overline{X}_n - \overline{\mu}_n\right| \geq \epsilon\right) = 0$$

Théorème 3 *Conditions suffisantes de convergence en probabilité. Si la suite de variables aléatoires $(X_n)_{n\geq 1}$ vérifie*

$$\begin{aligned} &1. \quad \lim_{n\to\infty} E(X_n) = c \\ &2. \quad \lim_{n\to\infty} VAR(X_n) = 0 \end{aligned}$$

Alors

$$X_n \xrightarrow{P} c$$

Preuve. *Conséquence immédiate de l'inégalité de Chebyshev.* ■

Exemple 7 *Supposons que $X_n \sim Bin(n,p)$. Alors la fréquence relative du nombre de "succès" converge en probabilité vers p :*

$$\frac{X_n}{n} \xrightarrow{P} p$$

En effet,

$$\begin{aligned} &1. \quad E(X_n/n) = p \underset{n\to\infty}{\to} p \\ &2. \quad VAR(X_n/n) = \frac{p(1-p)}{n} \underset{n\to\infty}{\to} 0 \end{aligned}$$

Ce résultat justifie l'approche fréquentielle des probabilités (à opposer à l'approche bayésienne, par exemple).

Remarque 5 *Attention, si $X_n \xrightarrow{P} c$, ceci n'entraîne pas que $E(X_n) \underset{n\to\infty}{\to} c$ comme le montre l'exemple suivant. Prenons*

X_n	0	n^2
P	$1-\frac{1}{n}$	$\frac{1}{n}$

On voit bien que $X_n \xrightarrow{P} 0$, mais que $E(X_n) = n \underset{n\to\infty}{\to} \infty$.

Exemple 8 *Autre exemple.*

X_n	1	n
P	$1-\frac{1}{n}$	$\frac{1}{n}$

Ici, $X_n \xrightarrow{P} 1$, *mais* $E(X_n) = 2 - 1/n \underset{n\to\infty}{\to} 2$.

Remarque 6 *Une formulation plus générale de la convergence en probabilité est la suivante. On dit que la suite de variables aléatoires* X_n *converge en probabilité vers la variable aléatoire* X *si la suite* $X_n - X$ *converge en probabilité vers* 0 :

$$X_n \xrightarrow{P} X \Leftrightarrow X_n - X \xrightarrow{P} 0$$

Dans ce cas, les conditions suffisantes du théorème ci-dessus deviennent :

$$\begin{array}{ll} 1. & E(X_n - X) \underset{n\to\infty}{\to} 0 \\ 2. & VAR(X_n - X) \underset{n\to\infty}{\to} 0 \end{array} \quad \Rightarrow X_n \xrightarrow{P} X$$

Théorème 4 *de continuité (Slutsky).*

$$\begin{array}{ll} 1. & X_n \xrightarrow{P} c \\ 2. & \varphi \text{ est continue en } c \end{array} \quad \Rightarrow \varphi(X_n) \xrightarrow{P} \varphi(c)$$

Preuve. *Rappelons que* φ *est continue en* c *signifie que* $\forall \epsilon > 0, \exists \delta > 0$ *tel que*

$$|x - c| < \delta \Rightarrow |\varphi(x) - \varphi(c)| < \epsilon$$

Ceci entraîne

$$P(|\varphi(X_n) - \varphi(c)| < \epsilon) \geq P(|X_n - c| < \delta)$$

pour n *assez grand (rappelons que* $A \subset B \Rightarrow P(A) \leq P(B)$*). Maintenant,*

$$X_n \xrightarrow{P} c \Rightarrow \forall \delta > 0 : 1 \geq \lim_{n\to\infty} P(|\varphi(X_n) - \varphi(c)| < \epsilon) \geq \lim_{n\to\infty} P(|X_n - c| < \delta) = 1$$

■

Remarque 7 *Ce théorème est valide dans le cas vectoriel (deux variables ou plus). On a alors :*

$$\begin{array}{ll} 1. & X_n \xrightarrow{P} X \\ 2. & Y_n \xrightarrow{P} Y \\ 3. & \varphi \text{ est continue en } (X,Y) \end{array} \quad \Rightarrow \varphi(X_n, Y_n) \xrightarrow{P} \varphi(X,Y)$$

Ainsi, si on a par exemple $X_n \xrightarrow{P} c$ *et* $Y_n \xrightarrow{P} d$, *alors*

1. $aX_n + bY_n \xrightarrow{P} ac + bd$
2. $X_n Y_n \xrightarrow{P} cd$
3. $\dfrac{X_n}{c} \xrightarrow{P} 1$ $(c \neq 0)$
4. $\dfrac{1}{X_n} \xrightarrow{P} \dfrac{1}{c}$ $(P(X_n = 0) = 0$ *et* $c \neq 0)$
5. $\sqrt{X_n} \xrightarrow{P} \sqrt{c}$ $(X_n \geq 0$ *pp*$)$

Exemple 9 *Supposons que $X_n \sim Bin(n,p)$. On a vu plus haut que*

$$\frac{X_n}{n} \xrightarrow{P} p$$

On a également

$$\frac{X_n}{n}\left(1-\frac{X_n}{n}\right) \xrightarrow{P} p(1-p)$$

Exemple 10 *Supposons que $X_1,\ldots,X_n iid \sim U[0;\theta]$. On a vu plus haut que $X_{(n)} = \max(X_1,\ldots,X_n) \xrightarrow{P} \theta$. Soit $\varphi(x) = \sqrt{x}$. Puisque c'est une fonction continue, on en déduit que*

$$\varphi\left(X_{(n)}\right) = \sqrt{X_{(n)}} \xrightarrow{P} \sqrt{\theta}$$

Voyons comment on peut démontrer ceci directement (sans utiliser Slutsky) à partir de la définition. Soit donc $\epsilon > 0$. On a

$$\begin{aligned}
P\left(\left|\sqrt{X_{(n)}}-\sqrt{\theta}\right| > \epsilon\right) &= P\left(\sqrt{\theta}-\sqrt{X_{(n)}} > \epsilon\right) \\
&= P\left(\left(\sqrt{\theta}-\sqrt{X_{(n)}}\right)\left(\sqrt{\theta}+\sqrt{X_{(n)}}\right) > \epsilon\left(\sqrt{\theta}+\sqrt{X_{(n)}}\right)\right) \\
&= P\left(\theta - X_{(n)} > \epsilon\left(\sqrt{\theta}+\sqrt{X_{(n)}}\right)\right) \\
&\leq P\left(\theta - X_{(n)} > \epsilon\sqrt{\theta}\right) \underset{n\longrightarrow\infty}{\longrightarrow} 0
\end{aligned}$$

puisque nous savons que $X_{(n)} \xrightarrow{P} \theta$.

Exemple 11 *Fonction de répartition empirique.*

$$F_n(x) \xrightarrow{P} F(x)$$

Ce résultat est à la base du test de Kolmogorov-Smirnov.

Remarque 8 *Rappelons que si $x_1,\ldots,x_n$ sont des données obtenues à partir d'un échantillon aléatoire de taille n d'une loi $F(x)$, et $x_{(1)} < \ldots < x_{(n)}$ sont les données ordonnées, alors*

$$p_n(x) = \begin{cases} 0 & si & x < x_{(1)} \\ i/n & si & x_{(i)} \leq x < x_{(i+1)} \\ 1 & si & x_{(n)} \leq x \end{cases}$$

2.1.1 Exercices

Exercice 2 *Soit $X_n \sim U\left[0;1+1/n^2\right]$. Montrer que $X_n \xrightarrow{P} X \sim U[0;1]$.*

Exercice 3 *Évaluer l'intégrale suivante par la méthode de Monte-Carlo :*

$$I = \int_0^1 \cos\left(x^2\right)\sin\left(x^4\right)dx$$

Exercice 4 *Soit $X_1,\cdots,X_n$ iid $\sim Exp(\lambda=3)$. Montrer que pour n suffisamment grand, on a $P(X_1+\cdots+X_n < n/2) > 0.999$. Vérifier avec une simulation.*

2.2 Convergence en moyenne d'ordre k

Soit $k \in]0; \infty[$. La suite de variables aléatoires $\{X_n\}_{n\geq 1}$ converge en moyenne d'ordre k vers la variable aléatoire X si

$$\lim_{n\to\infty} E\left(|X_n - X|^k\right) = 0$$

On écrit alors

$$X_n \xrightarrow{k} X \qquad \text{ou} \qquad X_n \xrightarrow{L^k} X$$

Un cas particulier important est $k = 2$. On dit alors qu'on a une convergence en *moyenne quadratique* (d'ordre 2). Noter que dans ce cas ($k = 2$) :

$$E\left[(X_n - X)^2\right] = VAR\,(X_n - X) + [E\,(X_n - X)]^2$$

On retrouve ainsi les deux conditions suffisantes de convergence en probabilité vues plus haut. On a ainsi le

Théorème 5

$$X_n \xrightarrow{L^2} X \Leftrightarrow \begin{array}{ll} 1. & E\,(X_n - X) \underset{n\to\infty}{\rightarrow} 0 \\ 2. & VAR\,(X_n - X) \underset{n\to\infty}{\rightarrow} 0 \end{array}$$

Exemple 12 *Soit $X_n iid \sim Bernoulli\,(p)$ et $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ leur moyenne empirique. Alors $\overline{X}_n \xrightarrow{L^2} p$.*

Remarque 9 *En fait, il peut être facilement montré que s'il existe $k > 0$ tel que $X_n \xrightarrow{L^k} X$, alors $X_n \xrightarrow{P} X$. En effet, il suffit d'appliquer l'inégalité de Markov :*

$$P\,(|X_n - X| \geq \epsilon) = P\left(|X_n - X|^k \geq \epsilon^k\right) \leq \frac{E\left(|X_n - X|^k\right)}{\epsilon^k} \underset{n\to\infty}{\rightarrow} 0$$

2.3 Convergence presque sûre*

Définition 2 *Soit donné une suite de variables aléatoires $\{X_n\}_{n\geq 1}$ et une variable aléatoire X toutes définies sur le même espace probabilisé $(\Omega, \mathcal{E}, P)$. Soit l'événement*

$$A = \left\{\omega \in \Omega : X_n\,(\omega) \underset{n\to\infty}{\rightarrow} X\,(\omega)\right\}$$

(convergence ponctuelle). Si $P\,(A) = 1$, on dit que la suite $\{X_n\}$ converge presque sûrement (ou avec probabilité égale à 1) vers la variable X. Notation :

$$X_n \xrightarrow{PS} X$$

Remarque 10 *Le théorème de contituité (Slustky) est valide pour la convergence PS : si $X_n \xrightarrow{PS} X$ et $\varphi\,(x)$ est continue, alors $\varphi\,(X_n) \xrightarrow{PS} \varphi\,(X)$.*

On admettra le résultat suivant.

Théorème 6 *La convergence presque sûre entraîne la convergence en probabilité.*

$$X_n \overset{PS}{\to} X \Rightarrow X_n \overset{P}{\to} X$$

Remarque 11 *Attention, la réciproque de ce résultat est fausse comme le montre l'exemple suivant.*

Exemple 13 *Soit $X_1, X_2, \ldots, X_n$ iid de même loi*

X_n	0	1
P	$1 - 1/n$	$1/n$

Soit $0 < \epsilon < 1$. On a

$$P(|X_n| \geq \epsilon) = P(X_n = 1) = \frac{1}{n} \underset{n\to\infty}{\to} 0$$

Donc

$$X_n \overset{P}{\to} 0$$

Noter qu'on peut aussi utiliser

$$\begin{array}{ll} 1. & E(X_n) = \dfrac{1}{n} \underset{n\to\infty}{\to} 0 \\ 2. & VAR(X_n) = \dfrac{1}{n} - \dfrac{1}{n^2} \underset{n\to\infty}{\to} 0 \end{array}$$

La suite $\{X_n(\omega)\}$ est donc juste une suite de 0 et de 1. Pour qu'elle converge vers 0, il faut qu'il n'y ait qu'un nombre fini de 1 (en d'autres termes, à partir d'un certain rang, il n'y a que des 0). Posons

$$E_n = \{\omega : X_n(\omega) = 1\}$$

On a

$$\sum_{n\geq 1} P(E_n) = \sum_{n\geq 1} \frac{1}{n} = \infty$$

(série harmonique). Un théorème (Borel-Cantelli) nous dit qu'alors une infinité de E_n sont réalisés. Ceci entraîne que

$$P(X_n \text{ converge}) = 0 \text{ et } X_n \overset{PS}{\nrightarrow} X$$

(ne converge pas PS).

On a les résultats suivants.

Théorème 7 *La suite $X_n \overset{PS}{\to} X \Leftrightarrow \forall k = 1, 2, 3, \ldots$*

$$\lim_{N\to\infty} P\left(|X_n - X| \geq \frac{1}{k} \text{ pour certains } n \geq N\right) = 0$$

ou, de manière équivalente,

$$\lim_{N\to\infty} P\left(\sup_{n\geq N} |X_n - X| \geq \frac{1}{k}\right) = 0$$

Théorème 8 **des trois séries de Kolmogorov. Soit $\{X_n\}_{n\geq 1}$ une suite de variables aléatoires indépendantes et soit Y_n la troncation de X_n de niveau $c > 0$:*

$$Y_n = \begin{cases} X_n & si \quad |X_n| \leq c \\ \\ 0 & si \quad |X_n| > c \end{cases}$$

Alors $\sum_{n=1}^{\infty} X_n$ converge PS $\Leftrightarrow \exists c > 0$:

$$\begin{array}{ll} a. & \displaystyle\sum_{n=1}^{\infty} P(|X_n| > c) < \infty \\ b. & \displaystyle\sum_{n=1}^{\infty} E(Y_n) < \infty \\ c. & \displaystyle\sum_{n=1}^{\infty} VAR(Y_n) < \infty \end{array}$$

Exemple 14 *Nous savons que la série harmonique diverge :*

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots + \frac{1}{n} + \ldots = +\infty$$

et que la série harmonique alternée converge :

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots + \frac{(-1)^{n-1}}{n} + \ldots = \ln 2$$

Que se passe-t-il si on attribue les signes $+$ et $-$ de façon aléatoire en jettant une pièce de monnaie équilibrée ? En d'autres termes, on demande la convergence (ou divergence) de

$$X_1 + X_2 + X_3 + X_4 + \ldots + X_n + \ldots$$

où $X_n = \pm 1/n$ avec probabilité $\frac{1}{2}$ et sont indépendantes. Prenons $c = 2$, par exemple (dans le théorème). On a alors $Y_n = X_n$ pour tout n. Ainsi, tous les termes de la série a. sont nuls. De plus $E(Y_n) = 0$ (condition b.). Enfin, (condition c.)

$$VAR(Y_n) = E\left(Y_n^2\right) = \frac{1}{n^2} \Rightarrow \sum_{n=1}^{\infty} VAR(Y_n) < \infty$$

Donc la série $\sum_{n=1}^{\infty} X_n$ converge PS : c'est un événement certain.

2.4 Convergence en loi

Définition 3 *Soit c une constante réelle et soit la fonction*

$$F(x) = \begin{cases} 0 & si \quad x < c \\ \\ 1 & si \quad x \geq c \end{cases}$$

Alors on dit que $F(x)$ est la fonction de répartition d'une variable aléatoire dégénérée (ou loi de Dirac) X

$$P(X = c) = 1$$

(toute la masse de probabilité est concentrée en c).

Définition 4 *Convergence en loi. Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires avec fonction de répartition* $F_n(x)$*. Soit* X *une variable aléatoire avec fonction de répartition* $F(x)$*. Alors*

$$X_n \overset{Loi}{\to} X \Leftrightarrow \lim_{n\to\infty} F_n(x) = F(x)$$

aux points de continuité de $F(x)$*.*

Remarque 12 *Supposons que* $X_n \overset{Loi}{\to} X$*.Si* $F(x)$ *est continue sur tout* $\mathbb{R}$*, il peut être montré que la convergence de la suite* $F_n(x)$ *est uniforme. En d'autres termes :*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0$$

Remarque 13 *Dire que la suite* X_n *converge en loi vers la variable dégénérée* $X = c$ *équivaut à dire que pour toute fonction continue et bornée* g*,*

$$\lim_{n\to\infty} E(g(X_n)) = E(g(c)) = g(c)$$

Voir également le théorème de Slutsky.

Exemple 15 *Soit* $X_n \sim U[0;1]$ *iid et* $Y_n = X_{(n)} = \max\{X_1, \ldots, X_n\}$*. On a*

$$F_{Y_n}(x) = F_n(x) = \begin{cases} x^n & si \quad x \in [0;1] \\ 0 & si \quad x \leq 0 \\ 1 & si \quad x \geq 1 \end{cases}$$

On a

$$Y_n \overset{Loi}{\to} Y$$

$$\lim_{n\to\infty} F_n(x) = F_Y(x) = \begin{cases} 1 & si \quad x \geq 1 \\ & \\ 0 & si \quad x < 1 \end{cases}$$

On voit que Y *est dégénérée* ($P(Y=1)=1$)*. Noter qu'il n'y a pas de convergence en* $x = 1$ *(ce n'est pas un point de continuité de* $F_Y(x)$*).*

Exemple 16 *Reprenons l'exemple précédent* ($X_n \sim U[0;1]$ *iid*) *et* $Y_n = n\left(1 - X_{(n)}\right)$*. On a*

$$\begin{aligned} F_{Y_n}(x) &= P\left(n\left(1 - X_{(n)}\right) \leq x\right) = P\left(X_{(n)} \geq 1 - \frac{x}{n}\right) \\ &= 1 - P\left(X_{(n)} < 1 - \frac{x}{n}\right) = \begin{cases} 0 & si \quad x < 0 \\ & \\ 1 - \left(1 - \frac{x}{n}\right)^n & si \quad x \geq 0 \end{cases} \end{aligned}$$

On voit immédiatement (pourquoi ?) que $Y_n \overset{Loi}{\to} Y \backsim Exp(\lambda = 1)$*. Le graphe de* $F_{Y_n}(x)$ *montre que la convergence est très rapide.*

Exemple 17 *Soit* $X_n \sim N\left(0, \sigma^2 = 1/n\right)$. *Alors* $X_n \overset{Loi}{\to} 0$. *En effet,*

$$F_n(x) = P(X_n \leq x) = P\left(\frac{X_n}{1/\sqrt{n}} \leq \frac{x}{1/\sqrt{n}}\right) = \Phi\left(\sqrt{n}x\right) \underset{n\to\infty}{\to} F(x) = \begin{cases} 1 & si \quad x > 0 \\ \\ 0 & si \quad x \leq 0 \end{cases}$$

Noter que $x = 0$ *est un point de discontinuité de* $F(x)$ *et que* $F_n(0) = \frac{1}{2} \neq F(0) = 0 \ \forall n$.

Exemple 18 $X_n \sim N\left(\mu, \sigma^2\right)$. *Alors* $Y_n = \sum_{k=1}^{n} X_k$ *ne converge pas en loi. Notons d'abord que* $Y_n \sim N\left(n\mu, n\sigma^2\right)$. *On a :*

$$F_{Y_n}(y) = P(Y_n \leq y) = \Phi\left(\frac{y - n\mu}{\sqrt{n}\sigma}\right)$$

On voit que la limite, quand n *tend vers l'infini, dépend du signe de* μ. *Considérons les trois cas possibles.*

<u>*Cas 1 :* $\mu > 0$</u>. *Alors*

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} \Phi\left(\frac{y - n\mu}{\sqrt{n}\sigma}\right) = \Phi(-\infty) = 0$$

<u>*Cas 2 :* $\mu = 0$</u>. *Alors*

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} \Phi\left(\frac{y}{\sqrt{n}\sigma}\right) = \Phi(0) = \frac{1}{2}$$

<u>*Cas 3 :* $\mu < 0$</u>. *Alors*

$$\lim_{n\to\infty} F_{Y_n}(y) = \lim_{n\to\infty} \Phi\left(\frac{y - n\mu}{\sqrt{n}\sigma}\right) = \Phi(+\infty) = 1$$

On voit que dans les trois cas, il n'y a pas de convergence en loi (voir la définition d'une variable dégénérée plus haut).

Remarque 14 *Ne pas confondre la convergence en loi avec la convergence en probabilité. Voir l'exemple suivant.*

Exemple 19 *Soit* $X_n = X \sim N(0,1)$ *et* $Y = -X$. *Il est évident que (pourquoi ?)* $Y \sim N(0,1)$. *On a, pour tout* $x \in \mathbb{R}$,

$$F_{X_n}(x) = \Phi(x) = \Phi(-x) = F_Y(x)$$

et alors

$$X_n = X \overset{Loi}{\to} Y = -X$$

Mais $X_n = X$ *ne converge pas en probabilité vers* $Y = -X$. *En effet, soit* $\epsilon > 0$ *donné,*

$$\begin{aligned} \lim_{n\to\infty} P(|X_n - Y| > \epsilon) &= \lim_{n\to\infty} P(|2X| > \epsilon) = \lim_{n\to\infty} P(|X| > \epsilon/2) \\ &= 2 \lim_{n\to\infty} (1 - \Phi(\epsilon/2)) = 2(1 - \Phi(\epsilon/2)) \neq 0 \end{aligned}$$

Le lien entre la convergence en loi et la convergence en probabilité est fourni par le résultat suivant.

Théorème 9 *Convergence en* ***probabilité*** $\Rightarrow$ ***convergence en loi*** *(dans le cas dégénéré :* $X_n \stackrel{Loi}{\to} c \Leftrightarrow X_n \stackrel{P}{\to} c$*).*

Avant de démontrer ce théorème, rappelons que pour tous évènements A et B, nous avons

$$\begin{aligned} A &= (A\cap B) \cup \left(A \cap \overline{B}\right) \Rightarrow P(A) = P(A\cap B) + P\left(A\cap \overline{B}\right) \\ &\leq P(B) + P\left(A\cap\overline{B}\right) \quad (\text{puisque } A\cap B \subset B) \end{aligned}$$

Maintenant, soit X et Y deux variables aléatoires et c et ϵ deux nombres, avec $\epsilon > 0$. Posons

$$A = \text{“}X \leq c\text{”},\ B = \text{“}Y \leq c+\epsilon\text{”}$$

Alors, en tenant compte de l'inégalité ci-dessus,

$$\begin{aligned} P(X\leq c) &\leq P(Y\leq c+\epsilon) + P(X \leq c \cap Y > c+\epsilon) \\ &= P(Y\leq c+\epsilon) + P(X\leq c\cap X - Y < -\epsilon) \\ &\leq P(Y\leq c+\epsilon) + P(|X-Y| > \epsilon) \end{aligned}$$

Passons à la démonstration du théorème.

Preuve. Soit $(X_n)_{n\geq 1}$ telle que $X_n \stackrel{P}{\to} X$. On veut montrer que $X_n \stackrel{Loi}{\to} X$. On a (en utilisant l'inégalité et en adaptant les notations)

$$\begin{aligned} F_n(x) = P(X_n \leq x) \leq P(X\leq x+\epsilon) + P(|X_n - X| > \epsilon) = F(x+\epsilon) + P(|X_n - X| > \epsilon) \\ F_n(x) = P(X_n \leq x) \geq P(X\leq x-\epsilon) + P(|X_n - X| > \epsilon) = F(x-\epsilon) + P(|X_n - X| > \epsilon) \end{aligned}$$

où x est un point de continuité de $F(x)$. On a alors

$$\begin{aligned} &\lim_{n\to\infty} F_n(x) \leq F(x+\epsilon) + \lim_{n\to\infty} P(|X_n - X| > \epsilon) = F(x+\epsilon) \\ &\qquad\qquad et \\ &\lim_{n\to\infty} F_n(x) \geq F(x-\epsilon) + \lim_{n\to\infty} P(|X_n - X| > \epsilon) = F(x-\epsilon) \end{aligned}$$

En d'autres termes,

$$F(x-\epsilon) \leq \lim_{n\to\infty} F_n(x) \leq F(x+\epsilon)$$

Puisque x est un point de continuité de F, on déduit que

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

■

On a les résultats utiles suivants.

Théorème 10
1. *(Slutsky)* $X_n \stackrel{Loi}{\to} X \Rightarrow g(X_n) \stackrel{Loi}{\to} g(X)$ *(où g est continue). Ce résultat est valide dans le cas de plusieurs variables.*
2. $X_n - Y_n \stackrel{P}{\to} 0$ *et* $Y_n \stackrel{Loi}{\to} Y \Rightarrow X_n \stackrel{Loi}{\to} Y$.

3. $X_n \overset{Loi}{\to} X$ et $Y_n \overset{P}{\to} c \Rightarrow X_n + Y_n \overset{Loi}{\to} X + c$ et $X_n Y_n \overset{Loi}{\to} cX$.

4. $X_n \overset{PS}{\to} X \Rightarrow X_n \overset{Loi}{\to} X$.

Preuve. *Admis.* ■

Exemple 20 $X_k \sim Pareto\,(1,1)$ *iid* :

$$F_k(x) = \begin{cases} 1 - \frac{1}{1+x} & si \quad x > 0 \\ \\ 0 & sinon \end{cases}$$

Soit $Y_n = nX_{(1)} = n \min\{X_1, \ldots, X_n\}$. *On a*

$$P\left(Y_n > \frac{x}{n}\right) = P\left(nX_{(1)} > x\right) = P\left(X_{(1)} > \frac{x}{n}\right) = 1 - F_k(x)^n = \frac{1}{\left(1 + \frac{x}{n}\right)^n} \qquad pour\ x > 0$$

Ceci donne

$$F_{Y_n}(x) = \begin{cases} 1 - \dfrac{1}{\left(1 + \frac{x}{n}\right)^n} & si \quad x > 0 \\ \\ 0 & sinon \end{cases}$$

On a, pour tout $x > 0$,

$$\lim_{n\to\infty} F_{Y_n}(x) = \lim_{n\to\infty} \left[1 - \frac{1}{\left(1 + \frac{x}{n}\right)^n}\right] = 1 - e^{-x} = F_Y(x)$$

où $Y \sim Exp\,(\lambda = 1)$ *et on a*

$$Y_n \overset{Loi}{\to} Y$$

On a la conséquence utile suivante.

Corollaire 1 *Si* $a_n \uparrow +\infty$ *(suite de réels) et* $a_n(X_n - c) \overset{Loi}{\to} X$, *alors* $X_n \overset{P}{\to} c$ *(en pratique, souvent* $a_n = \sqrt{n}$ *: TCL).*

Preuve. *On a*

$$X_n - c = \frac{1}{a_n} a_n (X_n - c)$$

Appliquons le théorème précédent (point 3) :

$$\begin{matrix} a_n(X_n - c) \overset{Loi}{\to} X \\ \frac{1}{a_n} \to 0 \end{matrix} \quad \Rightarrow \quad (X_n - c) \overset{Loi}{\to} 0 \quad \Rightarrow \quad X_n \overset{P}{\to} c$$

■

Remarque 15 *Contrairement à la convergence presque sûre et à la convergence en probabilité, la convergence en loi n'est pas stable par addition, c-à-d si* $X_n \overset{Loi}{\to} X$ *et* $Y_n \overset{Loi}{\to} Y$, *cela n'entraîne pas que* $X_n + Y_n \overset{Loi}{\to} X + Y$. *Pour le voir, prenons* $X_n = Y_n = X$ *pour tout* n *et* $Y = -X$. *Il est évident que* $X_n \overset{Loi}{\to} X$ *et* $Y_n \overset{Loi}{\to} Y$, *mais que* $X_n + Y_n$ *ne converge pas en loi vers* $X + Y = 0$.

Théorème 11 *(de continuité). Soit $F_n(x)$ une suite de fonctions de répartition avec FGM $M_n(t)$. Soit $F(x)$ une fonction de répartition avec FGM $M(t)$. Alors, si*

$$\lim_{n\to\infty} M_n(t) = M(t)$$

dans un voisinage de $t = 0$, on a

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

aux points de continuité de $F(x)$ (i.e. $X_n \overset{Loi}{\to} X$).

Exemple 21 *Soit $M_n(t) = \left(q + pe^t\right)^n$ où $q = 1 - p$ (donc $X_n \sim Bin(n,p)$). On a*

$$M_n(t) = \left(q + pe^t\right)^n = \left(1 - p + pe^t\right)^n = \left(1 + p\left(e^t - 1\right)\right)^n$$

Supposons que $\lim_{\substack{n\to\infty \\ p\to 0}} (np) = \lambda > 0$. Alors,

$$M_n(t) \approxeq \left(1 + \frac{\lambda}{n}\left(e^t - 1\right)\right)^n \underset{n\to\infty}{\to} e^{\lambda\left(e^t-1\right)} = M(t)$$

Donc

$$X_n \overset{Loi}{\to} X \sim Poi(\lambda)$$

Exemple 22 *Soit $X_n \sim Poi(\lambda_n)$ où λ_n est une suite croissante ($\lambda_n \to \infty$). On a $E(X_n) = VAR(X_n) = \lambda_n$. Posons*

$$Z_n = \frac{X_n - E(X_n)}{\sqrt{VAR(X_n)}} = \frac{X_n - \lambda_n}{\sqrt{\lambda_n}}$$

On a

$$M_{X_n}(t) = e^{\lambda_n\left(e^t-1\right)} \Rightarrow M_{Z_n}(t) = e^{-t\sqrt{\lambda_n}} M_{X_n}\left(\frac{t}{\sqrt{\lambda_n}}\right) = e^{-t\sqrt{\lambda_n}} \exp\left(\lambda_n\left(e^{t/\sqrt{\lambda_n}} - 1\right)\right)$$

Prenons le log :

$$\begin{aligned}
\ln M_{Z_n}(t) &= -t\sqrt{\lambda_n} + \lambda_n\left(e^{\left(\frac{t}{\sqrt{\lambda_n}}\right)} - 1\right) \\
&= -\lambda_n - t\sqrt{\lambda_n} + \lambda_n e^{\frac{t}{\sqrt{\lambda_n}}}
\end{aligned}$$

Utilisant

$$e^x = \sum_{k\geq 0} \frac{x^k}{k!}$$

on a

$$\begin{aligned}
\ln M_{Z_n}(t) &= -\lambda_n - t\sqrt{\lambda_n} + \lambda_n\left(1 + \frac{t}{\sqrt{\lambda_n}} + \frac{1}{2!}\frac{t^2}{\left(\sqrt{\lambda_n}\right)^2} + \frac{1}{3!}\frac{t^3}{\left(\sqrt{\lambda_n}\right)^3} + \ldots\right) \\
&= -\lambda_n - t\sqrt{\lambda_n} + \lambda_n + t\sqrt{\lambda_n} + \frac{t^2}{2!} + \frac{1}{3!}\frac{t^3}{\sqrt{\lambda_n}} + \frac{1}{4!}\frac{t^4}{\lambda_n} + \ldots \\
&= \frac{t^2}{2!} + \frac{1}{3!}\frac{t^3}{\sqrt{\lambda_n}} + \frac{1}{4!}\frac{t^4}{\lambda_n} + \ldots
\end{aligned}$$

Puisque $\lambda_n \underset{n\to\infty}{\to} \infty$, *on trouve*

$$\lim_{n\to\infty} \ln M_{Z_n}(t) = \frac{t^2}{2} \Rightarrow \lim_{n\to\infty} M_{Z_n}(t) = e^{t^2/2}$$

Donc $Z_n \overset{Loi}{\to} Z \sim N(0,1)$.

2.4.1 Exercices

Exercice 5 *Étudier la convergence en loi de la suite* $X_n \sim U[-1/n; 1/n]$.

Exercice 6 *La définition de la convergence en loi est basée sur les fonctions de répartition. Qu'en est-il des densités ou des fonctions de masse ? Considérons la situation suivante : soit la suite de fonctions de masse (de variables dégénérées)*

$$p_n(x) = \begin{cases} 1 & si & x = 2 + 1/n \\ 0 & ailleurs & \end{cases}$$

On a bien

$$F_n(x) = \begin{cases} 0 & si & x < 2 + 1/n \\ 1 & si & x \geq 2 + 1/n \end{cases}$$

Il est clair que

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0 & si & x < 2 \\ 1 & si & x \geq 2 \end{cases}$$

Nous avons donc une convergence en loi vers une variable dégénérée concentrée en $x = 2$. *Pourtant, pour tout* $x \in \mathbb{R}$,

$$\lim_{n\to\infty} p_n(x) = 0$$

qui n'est pas une fonction de masse.

Exercice 7 *Soit* X_n *de fonction de densité*

$$f_n(x) = \begin{cases} 1 - \cos(2\pi n x) & si & x \in [0;1] \\ 0 & ailleurs & \end{cases}$$

1. *Vérifier que pour tout* $0 < x < 1$, *la suite* $f_n(x)$ *est divergente.*
2. *Montrer que*

$$X_n \overset{Loi}{\to} X \sim U[0;1]$$

3 Loi forte des grands nombres*

On a les deux résultats suivants (dus à Kolmogorov).

Théorème 12 *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires indépendantes telles que* $E(X_i) = \mu_i$ *et* $VAR(X_i) = \sigma_i^2$. *Si*

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$$

(converge), alors

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_i)\overset{PS}{\to}0$$

Théorème 13 *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires iid avec $E(X_i)=\mu$. Posons $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$. Alors*

$$\overline{X}_n\overset{PS}{\to}\mu$$

Remarque 16 *Si $E(X_n)$ ($X_n$iid) diverge, la suite $\overline{X}_n$ présente des variations brusques et ne converge pas en général. Cependant, si $X_n\geq 0$ et $E(X_n)=+\infty$, on peut montrer que $\overline{X}_n\overset{PS}{\to}+\infty$. Faire une simulation (avec Cauchy par exemple).*

4 Théorème Central Limite

Nous donnons ici la version la plus simple (il y en a plusieurs).

Théorème 14 *Central limite (de Lindeberg et Levy). Soit $X_1,\ldots,X_n$ iid avec $E(X_i)=\mu$ et $VAR(X_i)=\sigma^2$. Posons $\overline{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$. Alors, $\forall x\in\mathbb{R}$,*

$$\lim_{n\to\infty}P\left(\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\leq x\right)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du$$

En d'autres termes,

$$\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\overset{Loi}{\to}Z\sim N(0;1)$$

Preuve. *On peut utiliser la FGM des X_i si elle existe. Voir Rice pour les détails. Noter que l'existence de la FGM n'est pas une condition requise dans le théorème.* ■

Remarque 17 *Remarquer qu'avec les hypothèses de théorème on a $\overline{X}_n\overset{P}{\to}\mu$. De plus, puisque $\Phi(x)$ est partout continue, la convergence en loi est uniforme sur $\mathbb{R}$.*

Remarque 18 *Le TCL nous dit que $Z_n=\dfrac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}$ est approximativement normalement distribuée $N(0;1)$ ou, ce qui revient au même, que $\overline{X}_n$ est approximativement normalement distribuée $N\left(\mu;\sigma^2/n\right)$ pour n suffisamment grand. De plus, ce théorème nous dit que la vitesse de convergence est approximativement égale à $\sqrt{n}$ car*

$$\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}=\sqrt{n}\frac{\overline{X}_n-\mu}{\sigma}$$

Comme on connaît rarement σ (et μ), nous verrons plus tard qu'on peut remplacer (sous certaines conditions) σ^2 par la variance empirique (échantillonnale)

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

On peut aussi (ce n'est pas évident) remplacer σ par $S_n=\sqrt{S_n^2}$. On a alors la version suivante.

Théorème 15 *Avec les hypothèses du TCL,*

$$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \xrightarrow{Loi} Z \sim N(0;1)$$

Exemple 23 *Soit $X_i \sim U[0;1]$. On a donc $E(X_i) = \frac{1}{2}$ et $VAR(X_i) = \frac{1}{12}$. Posons $S_n = \sum_{i=1}^{n} X_i$. Pour n suffisamment grand, S_n suit approximativement la loi $N\left(\frac{n}{2}; \frac{n}{12}\right)$. En particulier, $S_{12} - 6 \approxeq\sim N(0;1)$. En pratique, on génère 12 nombres aléatoires iid $U[0;1]$. Ils nous donnent 1 nombre aléatoire $\approxeq\sim N(0;1)$. Cette technique a été utilisée jusqu'à une période récente.*

Exemple 24 *Reprenons l'exemple précédent ($X_i \sim U[0;1]$) et soit $Y_i = g(X_i)$ où $g(x)$ est une fonction donnée telle que $\int_0^1 g(x)^2 dx < \infty$ (par exemple $g(x) = x^3$, $g(x) = \frac{1}{1+x}$, $g(x) = \ln(2+x)$, etc.). Alors le TCL nous dit que si n est assez grand, la variable $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ suit approximativement une loi $N(\mu, \sigma^2)$ avec $\mu = \int_0^1 g(x)\,dx$ et $\sigma^2 = \int_0^1 (g(x) - \mu)^2 dx$. En pratique, si la fonction $g(x)$ est trop compliquée, on procède par une simulation de Monte-Carlo.*

Exemple 25 *Approximation de la loi binomiale. La première version historique du TCL est spécifique à cette situation (théorème de de Moivre-Laplace). Soit $X_i$iid $\sim$ Bernoulli (p) et $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ la moyenne empirique du nombre de succès. Pour n assez grand et $c > 0$, on a*

$$P\left(\sqrt{n}\left|\frac{\overline{X}_n - p}{\sqrt{p(1-p)}}\right| \geq c\right) \approxeq P(Z_n \geq c) = 2(1 - \Phi(c))$$

Puisque $p(1-p) \leq 1/4$, alors

$$\begin{aligned} P\left(\left|\overline{X}_n - p\right| \geq \frac{c}{2\sqrt{n}}\right) &= P\left(\sqrt{n}\left|\frac{\overline{X}_n - p}{1/2}\right| \geq c\right) \leq P\left(\sqrt{n}\left|\frac{\overline{X}_n - p}{\sqrt{p(1-p)}}\right| \geq c\right) \\ &= P\left(\left|\overline{X}_n - p\right| \geq \frac{c\sqrt{p(1-p)}}{\sqrt{n}}\right) \approxeq 2(1 - \Phi(c)) \end{aligned}$$

c-à-d

$$\boxed{P\left(\left|\overline{X}_n - p\right| \geq \frac{c}{2\sqrt{n}}\right) \leq 2(1 - \Phi(c))}$$

Cette borne est toujours meilleure que celle fournie par l'inégalité de Hoeffding (cf chapitre4) qui dit

$$P\left(\left|\overline{X}_n - p\right| \geq \frac{c}{2\sqrt{n}}\right) \leq 2\exp\left(-c^2/2\right)$$

Faire un graphe superposant $2(1 - \Phi(c))$ et $2\exp\left(-c^2/2\right)$.

Exemple 26 *Un événement E a une probabilité égale à $p = \frac{1}{2}$ de se réaliser. On veut répéter une expérience aléatoire (de Bernoulli) $n = 20$ fois. Trouver la probabilité que l'événement se réalise au moins $k = 9$ fois :*

$$P(S_{20} \geq 9) = 1 - P(S_{20} \leq 8) = 1 - \sum_{k=0}^{8} \binom{20}{k} \frac{1}{2^{20}} = \frac{392\,313}{524\,288} \approxeq 0.748\,28$$

$$\stackrel{R}{=} pbinom(8, 20, 0.5, FALSE) = 0.7482777$$

Avec la loi normale (en utilisant la correction de continuité) :

$$P(S_{20} \geq 9) = P(S_{20} \geq 8.5) = 1 - P(S_{20} \leq 8.5) = 1 - P\left(\frac{S_{20} - 10}{\sqrt{5}} \leq \frac{8.5 - 10}{\sqrt{5}}\right)$$
$$\approxeq 1 - \Phi\left(\frac{8.5 - 10}{\sqrt{5}}\right) = 1 - 0.251\,17 = \underline{0.748}\,83 \quad \textit{(3 décimales)}$$
$$\stackrel{R}{=} pnorm(8.5, mu, s, FALSE) = 0.7488325$$

Exemple 27 *Approximation de la loi de Poisson. Soit $X \sim Poi\,(\lambda = 20)$. On a*

$$P(20 \leq X \leq 30) = e^{-20} \sum_{k=20}^{30} \frac{20^k}{k!} \approxeq 0.516\,27$$
$$\stackrel{R}{=} ppois(30, 20) - ppois(19, 20) = 0.5162681$$

Maintenant,

$$P(20 \leq X \leq 30) = P(19.5 \leq X \leq 30.5) = P\left(\frac{19.5 - 20}{\sqrt{20}} \leq \frac{X - 20}{\sqrt{20}} \leq \frac{30.5 - 20}{\sqrt{20}}\right)$$
$$\approxeq \Phi\left(\frac{30.5 - 20}{\sqrt{20}}\right) - \Phi\left(\frac{19.5 - 20}{\sqrt{20}}\right) = 0.990\,56 - 0.455\,49 = \underline{0.5}35\,07$$
$$\stackrel{R}{=} pnorm(30.5, 20, sqrt\,(20)) - pnorm(19.5, 20, sqrt\,(20)) = 0.5350698$$

Exemple 28 *Mesures répétées. On veut mesurer une certaine constante inconnue μ. On la mesure n fois : $X_1, \cdots, X_n$. En supposant l'absence de biais dans l'instrument de mesure, on a $E(X_i) = \mu$. On sait (LFGN) que $\overline{X}_n \xrightarrow{P} \mu$. L'inégalité de Chebyshev donne une borne de l'erreur commise, mais le TCL fait beaucoup mieux :*

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) = P\left(\frac{\left|\overline{X}_n - \mu\right|}{\sigma/\sqrt{n}} < \frac{\epsilon}{\sigma/\sqrt{n}}\right) \approxeq \Phi\left(\frac{\epsilon}{\sigma/\sqrt{n}}\right) - \Phi\left(-\frac{\epsilon}{\sigma/\sqrt{n}}\right)$$
$$= 2\Phi\left(\frac{\epsilon}{\sigma/\sqrt{n}}\right) - 1$$

Prenons $n = 30, \sigma = 1, \epsilon = 0.5$. Alors

$$P\left(\left|\overline{X}_n - \mu\right| < \epsilon\right) \approxeq 2\Phi\left(\frac{0.5}{1/\sqrt{30}}\right) - 1 = 2\,(0.996\,92) - 1 = 0.993\,84$$

Si on veut fixer une valeur maximale tolérée (marge d'erreur) pour ϵ, on peut alors calculer la valeur minimale de n pour réaliser l'objectif.

Exemple 29 *Soit $X_1, \ldots, X_{15}$ un échantillon de taille $n = 15$ issu d'une loi de densité*

$$f(x) = 3x^2 \text{ pour } x \in [0; 1] \text{ (0 ailleurs).}$$

Trouver une approximation de $P\left(3/5 < \overline{X} < 4/5\right)$ où $\overline{X}$ est la moyenne empirique.

Que dit TCL pour un échantillon de taille aussi petite ($n = 15$) ? On a (petit calcul)

$$\begin{aligned} E(X_1) &= 3/4 \\ Var(X_1) &= 3/80 \\ &\Rightarrow E(\overline{X}) = 3/4 \text{ et } Var(\overline{X}) = 1/20^2 \end{aligned}$$

On a alors

$$\begin{aligned} P\left(3/5 < \overline{X} < 4/5\right) &= P\left(\frac{3/5 - 3/4}{1/20} < \frac{\overline{X} - 3/4}{1/20} < \frac{4/5 - 3/4}{1/20}\right) \\ &= P\left(-3 < \frac{\overline{X} - 3/4}{1/20} < 1\right) \\ &\overset{TCL}{\approxeq} \Phi(1) - \Phi(-3) \\ &= 0.840 \end{aligned}$$

Puisque l'échantillon est plutôt petit, faisons une simulation R :

```
## Faisons une simulation de x-bar N = 100000 fois
## et sauvegardons les résultats
N <- 100000 ; n <- 15 ; xbar<-c(1 :N)
for( i in 1 :N) {
  sample <- runif(n)
  sample <- sample^(1/3)
  sampmean <- mean(sample)
  xbar[i] <- sampmean
}
## Cherchons la proportion des moyennes empiriques
## qui se trouve dans l'intervalle (3/5,4/5)
upper <- xbar < (4/5)
lower <- xbar > (3/5)
mean(upper*lower)
> 0.83517
```

Remarque 19 *Il existe une version multivariée du TCL.*

4.1 Méthode delta

Théorème 16 *(Mann-Wald). Soit (a_n) une suite de réels tels que $a_n \uparrow +\infty$ et (X_n) une suite de variables aléatoires telles que*

$$a_n (X_n - c) \overset{Loi}{\to} X$$

où c est une constante et X une variable aléatoire. Soit g une fonction de classe C^1 (continûment différentiable) dans un voisinage de c. Alors

$$a_n (g(X_n) - g(c)) \overset{Loi}{\to} g'(c) X$$

En particulier, si $a_n = \sqrt{n}$ *et* $X \sim N\left(0, \sigma^2\right)$*, alors*

$$\boxed{\sqrt{n}\left(g\left(X_n\right) - g\left(c\right)\right) \overset{Loi}{\to} N\left(0, \sigma^2 g'\left(c\right)^2\right)}$$

ou, de manière équivalente,

$$\boxed{\sqrt{n}\left(\frac{g\left(X_n\right) - g\left(c\right)}{\sigma g'\left(c\right)}\right) \overset{Loi}{\to} N\left(0, 1\right)}$$

Preuve. *Faisons un développement limité (Taylor) d'ordre 1 dans un voisinage de c :*

$$g\left(x\right) = g\left(c\right) + \left(x - c\right) g'\left(c\right) + R\left(x\right) \qquad (R = \text{ reste})$$

avec $\lim_{x \to c} R\left(x\right) = 0$*. Puisque* $X_n \overset{P}{\to} c$*, alors (théorème de continuité)*

$$g\left(X_n\right) \overset{P}{\to} R\left(c\right) = 0$$

On a ainsi le développement

$$g\left(X_n\right) = g\left(c\right) + g'\left(c\right)\left(X_n - c\right) + R\left(X_n\right)$$

avec

$$g'\left(c\right) + R\left(X_n\right) \overset{P}{\to} g'\left(c\right)$$

et (Slutsky)

$$a_n\left(g\left(X_n\right) - g\left(c\right)\right) = a_n\left(X_n - c\right)\left(g'\left(c\right) + R\left(X_n\right)\right) \overset{Loi}{\to} g'\left(c\right) X$$

■

Exemple 30 *Soit* X_n *la fréquence relative du nombre de succès dans* n *essais indépendants de Bernoulli. Trouvons la distribution asymptotique de* $g\left(X_n\right) = 1/X_n$*.*

Posons $p = P\left(\text{succès}\right)$*. On a* $g'\left(p\right) = -1/p^2$ *et (TCL)*

$$\sqrt{n}\left(X_n - p\right) \overset{Loi}{\to} N\left(0, p\left(1 - p\right)\right)$$

Donc

$$\sqrt{n}\left(\frac{1}{X_n} - \frac{1}{p}\right) \overset{Loi}{\to} N\left(0, \frac{1 - p}{p^3}\right)$$

Exemple 31 *Soit* $X_1, \cdots, X_n$ *iid* $\sim Exp\left(\lambda = 1/\theta\right)$*, ie* $f\left(x\right) = \frac{1}{\theta} e^{-x/\theta}$*,* $x \geq 0$*,* $\theta > 0$*. On sait que* $E\left(X_i\right) = \theta$ *et* $Var\left(X_i\right) = \theta^2$*. Donc (TCL)*

$$\sqrt{n}\left(\overline{X} - \theta\right) \overset{Loi}{\to} N\left(0, \theta^2\right)$$

Considérons maintenant la transformation $g\left(x\right) = \sqrt{x}$ *et* $Y_n = g\left(\overline{X}\right) = \sqrt{\overline{X}}$*. Puisque* $g'\left(x\right) = 1/\left(2\sqrt{x}\right)$*, alors*

$$\sqrt{n}\left(\sqrt{\overline{X}} - \sqrt{\theta}\right) \overset{Loi}{\to} N\left(0, \theta/4\right)$$

Remarque 20 *Si $g'(c) = 0$ dans le théorème, alors $g'(c) X = 0$ et la loi limite est dégénérée (Dirac), ce qui nous dit que $g(X_n)$ converge plus vite vers $g(c)$. Il suffit alors de faire le développement limité à d'ordre supérieur à 1, jusqu'au premier terme non nul. Par exemple, supposons que*

$$\sqrt{n}\,(X_n - 1) \overset{Loi}{\rightarrow} N(0,1)$$

avec $g'(1) = 0$ et $g''(1) \neq 0$. Alors (preuve heuristique),

$$X_n \approxeq 1 + \frac{1}{\sqrt{n}} N(0,1)$$

et

$$g(1+h) \approxeq g(1) + \frac{1}{2} g''(1)\, h^2$$

D'où

$$\begin{aligned} g(X_n) &\approxeq g\left(1 + \frac{1}{\sqrt{n}} N(0,1)\right) \approxeq g(1) + \frac{1}{2} g''(1) \left(\frac{1}{\sqrt{n}} N(0,1)\right)^2 \\ &= g(1) + \frac{1}{2n} g''(1)\, \chi_1^2 \end{aligned}$$

Ou encore,

$$n\,(g(X_n) - g(1)) \approxeq \frac{1}{2} g''(1)\, \chi_1^2$$

Finalement,

$$\frac{2n}{g''(1)}\,(g(X_n) - g(1)) \overset{Loi}{\rightarrow} \chi_1^2$$

La loi χ_1^2 (Khi-carré à un degré de liberté) sera vue dans un chapitrer ultérieur.

Remarque 21 *Il existe une version multivariée de la méthode delta. Nous l'omettrons ici.*