Convergence stochastique : rappels et compléments

Statistique mathématique
M2 santé publique, université Paris-Sud

17 octobre 2017

1. Motivation : fonctions de variables aléatoires.
2. Convergence stochastique
   - Comment fonctionne la convergence d'une variable aléatoire ?
   - Convergence en probabilité et convergence en loi
3. Théorèmes utiles
   - Convergence faible de variables aléatoires
4. Notions de convergence forte
5. Les deux grands théorèmes

Fonctions de variables aléatoires

Soit $X_1, \ldots, X_n$ des variables aléatoires iid où $\mathbb{E}\big(X_i\big) = \mu$ et $\mathrm{Var}\big(X_i\big) = \sigma^2$. On considère :

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

- Si $X_i \sim \mathcal{N}(\mu, \sigma^2)$ ou $X_i \sim \exp\big(1/\mu\big)$ alors on connait la loi de $\bar{X}_n$.
- X_i peut avoir une loi moins classique que les précédentes.
- Souvent, il est difficile de calculer la loi jointe des X_i

On veut dire quelque chose à propos de $\bar{X}_n$ même dans ces cas

Difficile à n fixé, mais si on faisait tendre n vers l'infini ? (faire de l'asymptotique)

Fonctions de variables aléatoires

Quand $n \to \infty$ on comprend mieux la loi de $\bar{X}_n$

- Grossièrement, $\bar{X}_n$ se concentre autour de μ

$$\forall \varepsilon > 0, \mathbb{P}\left[|\bar{X}_n - \mu| > \varepsilon\right] \approx 0, \text{ quand } n \to \infty$$

- Peut-être qu'il est plus intéressant de regarder

$$\mathbb{P}\left[\sqrt{n}\left(\bar{X}_n - \mu\right) \leq x\right] \overset{n\to\infty}{\approx} ? \text{ pour obtenir } \mathbb{P}\left[\bar{X}_n \leq x\right]$$

Plus généralement $\to$ on veut étudier la loi de $Y = g\left(X_1, \ldots, X_n\right)$ pour une une certaine fonction générique g

- Souvent difficile à n fixé
- Recours aux approximations asymptotiques pour comprendre la loi de Y.

Convergence de variables aléatoires

On a besoin de préciser ce que veut dire :

- Y_n se *concentre* autour de μ quand $n \to \infty$
- Plus généralement, ce qu'on entend par Y_n *se comporte comme* Y quand n est grand
- Loi de $g\big(X_1, \dots, X_n\big)$ quand $n \to \infty$.

$\hookrightarrow$ notions appropriées de convergence de variables aléatoires

Petit rappel : Les variables aléatoires sont des fonctions entre espaces mesurables.

$\Rightarrow$ Il y a différents modes de convergence de variables aléatoires

- Convergence en probabilité (convergence en mesure)
- Convergence en loi (convergence faible)
- Convergence avec probabilité 1 (convergence presque sûre)
- Convergence du $p^{\text{ième}}$ moment (convegence $\mathbb{L}^p$)

Convergence en probabilité

Définition (Convergence en probabilité)

Soit $\{X_n\}_{n\geq 1}$ et X des variables aléatoires définies sur le même espace de probabilité. On dit que X_n converge en probabilité vers X quand $n \to \infty$ et on écrit ($X_n \xrightarrow{p} X$) si pour tout $\varepsilon > 0$, nous avons

$$\mathbb{P}\big[|X_n - X| > \varepsilon\big] \to 0, \text{ quand } n \to \infty$$

Intuitivement, si $X_n \xrightarrow{p} X$, alors $X_n \approx X$ avec une forte probabilité quand n tend vers l'infini

Exemple

Soit $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{U}[0,1]$, on définit $M_n = \max\{X_1, \ldots, X_n\}$. Alors,

$$\begin{aligned} F_{M_n}(x) = x^n \Rightarrow \mathbb{P}\big[|M_n - 1| > \varepsilon\big] &= \mathbb{P}\big[M_n < 1 - \varepsilon\big] \\ &= (1-\varepsilon)^n \overset{n\to\infty}{\longrightarrow} 0 \end{aligned}$$

pour tout $0 < \varepsilon < 1$. On déduit $M_n \xrightarrow{p} 1$.

Convergence en loi

Définition (convergence en loi)

Soit $\{X_n\}_{n\geq 1}$ et X des variables aléatoires. On dit que X_n converge en loi vers X quand $n \to \infty$ (on écrit $X_n \overset{d}{\to} X$) si

$$\mathbb{P}\big[X_n \leq x\big] \overset{n\to\infty}{\longrightarrow} \mathbb{P}\big[X \leq x\big],$$

en tout point de continuité de $F_X(x) = \mathbb{P}\big[X \leq x\big]$.

Exemple 1

Soit $X_1, \dots, X_n$ des v.a iid telles que

$$\mathbb{P}(X_i = j) = \frac{1}{10} \text{ pour } j = 0, 1, 2, \dots, 9$$

et on définit

$$U_n = \sum_{i=1}^{n} \frac{X_i}{10^i}.$$

- On s'intéresse à la loi limite de U_n, dont les valeurs possibles sont $\frac{j}{10^n}$ où $j = 0, 1, 2, \dots, 10^n - 1$ et

$$\mathbb{P}\left(U_n = \frac{j}{10^n}\right) = \frac{1}{10^n} \text{ pour } j = 0, 1, 2, \dots, 10^n - 1.$$

  - Calculer la fonction de répartition de $\mathbb{P}(U_n \leq x)$ pour $\frac{j}{10^n} \leq x < \frac{j+1}{10^n}$.
  - Majorer $\left|\mathbb{P}(U_n \leq x) - x\right|$ et conclure.

Exemple 1 : suite

- $$\mathbb{P}\big(U_n \leq x\big) = \frac{j+1}{10^n} \text{ pour } \frac{j}{10^n} \leq x < \frac{j+1}{10^n}.$$

- $$\Big|\mathbb{P}\big(U_n \leq x\big) - x\Big| \leq 10^{-n} \to 0 \text{ quand } n \to \infty.$$

Exemple 2

Soit $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{U}[0,1]$, on définit $M_n = \max\{X_1, \dots, X_n\}$, et $Q_n = n(1 - M_n)$.

- Calculer la loi limite de Q_n.
  - Calculer la fonction de répartition de M_n
  - Calculer la fonction de répartition de Q_n
  - Faire tendre n vers l'infini

Exemple 2

Soit $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{U}[0,1]$, on définit $M_n = \max\{X_1, \ldots, X_n\}$, et $Q_n = n(1 - M_n)$.

- Calculer la loi limite de Q_n.
  - Calculer la fonction de répartition de M_n
  - Calculer la fonction de répartition de Q_n
  - Faire tendre n vers l'infini

Exemple 2 : suite

$$\mathbb{P}[Q_n \leq x] = \mathbb{P}\Big[M_n \geq 1 - \frac{x}{n}\Big] = 1 - \Big(1 - \frac{x}{n}\Big)^n \overset{n\to\infty}{\longrightarrow} 1 - e^{-x}$$

pour tout $x \geq 0$. On a donc, $Q_n \xrightarrow{d} Q$, où $Q \sim \mathcal{E}(1)$.

Quelques commentaires sur $\xrightarrow{p}$ et $\xrightarrow{d}$

- La convergence en probabilité implique la convergence en loi
- La convergence en loi n'implique pas la convergence en probabilité

$$\text{Considérons } X \sim \mathcal{N}(0,1), -X + \frac{1}{n} \xrightarrow{d} X \text{ mais } -X + \frac{1}{n} \xrightarrow{p} -X$$

- $\xrightarrow{d}$ fait le lien entre des fonctions de répartition
  - On peut l'utiliser pour approcher des fonctions de répartitions (erreur d'approximation ?)
- Les deux modes de convergence sont ***métrisables***
  - *i.e.* il existe des métriques sur les espaces de variables aléatoires ...
  - On peut utiliser l'inégalité triangulaire etc ...
- $\xrightarrow{d}$ est appelée aussi ***convergence faible***

Définition équivalente

$X_n \xrightarrow{d} X \iff \mathbb{E}f(X_n) \to \mathbb{E}f(X)$ pour tout fonction f continue bornée

Quelques propriétés

Théorème

(a) $X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$

(b) $X_n \xrightarrow{d} c \Longrightarrow X_n \xrightarrow{p} c, \quad c \in \mathbb{R}$

Quelques propriétés

Théorème

(a) $X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$

(b) $X_n \xrightarrow{d} c \Longrightarrow X_n \xrightarrow{p} c, \quad c \in \mathbb{R}$

(a) Soit x un point de continuité de F_X, montrer les deux inégalités suivantes

(i) $\mathbb{P}\big[X_n \leq x\big] \leq \mathbb{P}\big[X \leq x + \varepsilon\big] + \mathbb{P}\big[|X_n - X| > \varepsilon\big]$

(ii) $\mathbb{P}\big[X \leq x - \varepsilon\big] - \mathbb{P}\big[|X_n - X| > \varepsilon\big] \leq \mathbb{P}\big[X_n \leq x\big]$

(b) Écrire la fonction de répartition d'une variable aléatoire constante et majorer $\mathbb{P}\big[|X_n - c| > \varepsilon\big]$.

Preuve

(a) Soit x un point de continuité de F_X et $\varepsilon > 0$. On a

(i)

$$\begin{aligned}\mathbb{P}\big[X_n \leq x\big] &= \mathbb{P}\big[X_n \leq x, |X_n - X| \leq \varepsilon\big] + \mathbb{P}\big[X_n \leq x, |X_n - X| > \varepsilon\big] \\ &\leq \mathbb{P}\big[X \leq x + \varepsilon\big] + \mathbb{P}\big[|X_n - X| > \varepsilon\big],\end{aligned}$$

car $\big\{X_n \leq x, |X_n - X| \leq \varepsilon\big\} \subseteq \big\{X \leq x + \varepsilon\big\}$.

(ii) On a aussi,

$$\begin{aligned}\mathbb{P}\big[X \leq x - \varepsilon\big] &= \mathbb{P}\big[X \leq x - \varepsilon, |X_n - X| \leq \varepsilon\big] \\ &\quad + \mathbb{P}\big[X \leq x - \varepsilon, |X_n - X| > \varepsilon\big] \\ &\leq \mathbb{P}\big[X_n \leq x\big] + \mathbb{P}\big[|X_n - X| > \varepsilon\big].\end{aligned}$$

On obtient $\mathbb{P}\big[X \leq x - \varepsilon\big] - \mathbb{P}\big[|X_n - X| > \varepsilon\big] \leq \mathbb{P}\big[X_n \leq x\big]$.

On déduit (a) des deux inégalités précédentes (i) et (ii).

preuve : suite

(b) Soit F la fonction de répartition d'une variable aléatoire constante c,

$$F(x) = \mathbb{P}[c \leq x] = \begin{cases} 1 & \text{si } x \geq c, \\ 0 & \text{si } x < c. \end{cases}$$

$$\begin{aligned} \mathbb{P}\big[|X_n - c| > \varepsilon\big] &= \mathbb{P}\Big[\big\{X_n - c > \varepsilon\big\} \cup \big\{c - X_n > \varepsilon\big\}\Big] \\ &= \mathbb{P}\big[X_n > c + \varepsilon\big] + \mathbb{P}\big[X_n < c - \varepsilon\big] \\ &\leq 1 - \mathbb{P}\big[X_n \leq c + \varepsilon\big] + \mathbb{P}\big[X_n \leq c - \varepsilon\big] \\ &\xrightarrow{n \to \infty} 1 - F(\underbrace{c + \varepsilon}_{\geq c}) + F(\underbrace{c - \varepsilon}_{< c}) = 0 \end{aligned}$$

car $X_n \xrightarrow{d} c$.

Transformation continue

Théorème

Soit $g : \mathbb{R} \to \mathbb{R}$ une application continue.

(a) $X_n \xrightarrow{p} X \Longrightarrow g(X_n) \xrightarrow{p} g(X)$

(b) $Y_n \xrightarrow{d} Y \Longrightarrow g(Y_n) \xrightarrow{d} g(Y)$

Théorème de Slutsky

Slutsky

Soit $X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{d} c \in \mathbb{R}$. Alors,

(a) $X_n + Y_n \xrightarrow{d} X + c$

(b) $X_n Y_n \xrightarrow{d} cX$

Théorème de Slutsky

Slutsky

Soit $X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{d} c \in \mathbb{R}$. Alors,

(a) $X_n + Y_n \xrightarrow{d} X + c$

(b) $X_n Y_n \xrightarrow{d} cX$

(a) Prendre $c = 0$ et x un point de continuité de F_X montrer que
- $\mathbb{P}\big(X_n + Y_n \leq x\big) \leq \mathbb{P}\big(X_n \leq x + \varepsilon\big) + \mathbb{P}\big(|Y_n| > \varepsilon\big)$
- $\mathbb{P}\big(X_n \leq x - \varepsilon\big) - \mathbb{P}\big(|Y_n| > \varepsilon\big) \leq \mathbb{P}\big(X_n + Y_n \leq x\big)$.

(b) Soit $\varepsilon > 0$ et $M > 0$, il suffit de montrer que

$$\mathbb{P}\big(|X_n Y_n| > \varepsilon\big) \leq \mathbb{P}\big(|X_n| > \varepsilon M\big) + \mathbb{P}\big(|Y_n| > 1/M\big).$$

Slutsky : preuve

Nous avons

$$\begin{aligned}\mathbb{P}\big[X_n + Y_n \le x\big] &= \mathbb{P}\big[X_n + Y_n \le x, |Y_n| \le \varepsilon\big] + \mathbb{P}\big[X_n + Y_n \le x, |Y_n| > \varepsilon\big] \\ &\le \mathbb{P}\big(X_n \le x + \varepsilon\big) + \mathbb{P}\big(|Y_n| > \varepsilon\big)\end{aligned}$$

De manière similaire,
$\mathbb{P}\big[X_n \le x - \varepsilon\big] - \mathbb{P}\big[|Y_n| > \varepsilon\big] \le \mathbb{P}\big[X_n + Y_n \le x\big]$ On l'encadrement suivant

$$\begin{aligned}\mathbb{P}\big[X_n \le x - \varepsilon\big] - \mathbb{P}\big[|Y_n| > \varepsilon\big] &\le \mathbb{P}\big[X_n + Y_n \le x\big] \\ &\le \mathbb{P}\big[X_n \le x + \varepsilon\big] + \mathbb{P}\big[|Y_n| > \varepsilon\big]\end{aligned}$$

Il suffit de faire tendre $n \to \infty$ et $\varepsilon \to 0$

Slutsky : preuve (suite)

Nous avons

$$\begin{aligned}\mathbb{P}\big[|X_nY_n| > \varepsilon\big] &\leq \mathbb{P}\big[|X_nY_n| > \varepsilon, |Y_n| \leq 1/M\big] + \mathbb{P}\big[|Y_n| > 1/M\big] \\ &\leq \mathbb{P}\big[|X_n| > \varepsilon M\big] + \mathbb{P}\big[|Y_n| > 1/M\big] \\ &\to \mathbb{P}\big[|X_n| > \varepsilon M\big] + 0 \text{ quad } n \text{ tend vers l'infini}\end{aligned}$$

Il suffit de tendre M suffisamment grand pour que ce terme tend vers 0.

Théorème de Slutsky

(version générale)

Soit $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ une fonction continue et supposons que $X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{d} c \in \mathbb{R}$. Alors

$$g\big(X_n, Y_n\big) \xrightarrow{d} g(X, c).$$

quand $n \to \infty$.

Théorème de Slutsky

(version générale)

Soit $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ une fonction continue et supposons que $X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{d} c \in \mathbb{R}$. Alors

$$g\big(X_n, Y_n\big) \xrightarrow{d} g(X, c).$$

quand $n \to \infty$.

- Cette version n'est pas une conséquence du théorème de l'application continue. Il faut que le couple $\big(X_n, Y_n\big)$ (la loi jointe) converge en loi vers couple (X, c).
- Ici, on a supposé la convergence des lois marginales seulement ($X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{d} c$ séparément).
- Le point clé de la preuve : lorsque $Y_n \xrightarrow{d} c$ où c une constante, convergence des marges $\Longleftrightarrow$ convergence de la loi jointe.
- Ce théorème n'est pas valable lorsque $X_n \xrightarrow{d} X$ et $Y_n \xrightarrow{d} Y$.

la delta méthode

δ-méthode

Soit $Z_n = a_n\big(X_n - \theta\big) \xrightarrow{d} Z$ où $\theta \in R$ et $a_n \uparrow \infty$. Soit $g(\cdot)$ une fonction continue dérivable en θ. Alors

$$a_n\Big(g\big(X_n\big) - g(\theta)\Big) \xrightarrow{d} g'(\theta)Z.$$

la delta méthode : preuve

Preuve

Développement de Taylor autour de θ, nous avons

$$g\big(X_n\big) = g(\theta) + g'\big(\theta_n^*\big)\big(X_n - \theta\big) \text{ où } \theta_n^* \text{ est entre } \theta \text{ et } X_n.$$

On a $|\theta_n^* - \theta| < |X_n - \theta| = a_n^{-1}|a_n(X_n - \theta)| = a_n^{-1}|Z_n| \xrightarrow{p} 0$ (par Slutsky). Ainsi, $\theta_n^* \xrightarrow{p} \theta$.
Par le théorème de l'application continue, nous avons $g'\big(\theta_n^*\big) \xrightarrow{p} g'(\theta)$.
On déduit

$$\begin{aligned} a_n\Big(g\big(X_n\big) - g(\theta)\Big) &= a_n\Big(g(\theta) + g'\big(\theta_n^*\big)\big(X_n - \theta\big) - g(\theta)\Big) \\ &= g'\big(\theta_n^*\big)a_n\big(X_n - \theta\big) \xrightarrow{d} g'(\theta)Z. \end{aligned}$$

On peut utiliser la delta méthode lorsque $g'(\theta)$ n'est pas dérivable (la preuve fait appel au théorème de représentation de Skorokhod).

Convergence de l'espérance

Convergence de l'espérance

Si $|X_n| < M < \infty$ et $X_n \xrightarrow{d} X$, alors

$$\mathbb{E}\big(X_n\big) \to \mathbb{E}\big(X\big) \text{ quand } n \to \infty.$$

Convergence de l'espérance : Preuve

Convergence de l'espérance : preuve

Supposons un premier temps que les variables X_n sont positives ou nulles $\forall n$. Utiliser le fait que pour toute v.a Y positive ou nulle, nous avons

$$\mathbb{E}(Y) = \int_0^\infty \mathbb{P}\left[Y > y\right] \mathrm{d}y,$$

pour montrer le théorème.

Remarques sur la convergence faible

- Il est souvent difficile de montrer la convergence faible en utilisant sa définition.
- Quand F_n est connue, on y arrive.
- Nous avons besoin d'autres conditions suffisantes mais commodes.

Remarques sur la convergence faible

- Il est souvent difficile de montrer la convergence faible en utilisant sa définition.
- Quand F_n est connue, on y arrive.
- Nous avons besoin d'autres conditions suffisantes mais commodes.

Théorème de Scheffé

Supposons que X_n a pour densité f_n (ou une fonction de masse dans le cas discret) et f la densité de X. Alors

$$f_n(x) \xrightarrow{n\to\infty} f(x) \text{ pour tout } x, \Rightarrow X_n \xrightarrow{d} X.$$

Remarques sur la convergence faible

- Il est souvent difficile de montrer la convergence faible en utilisant sa définition.
- Quand F_n est connue, on y arrive.
- Nous avons besoin d'autres conditions suffisantes mais commodes.

Théorème de Scheffé

Supposons que X_n a pour densité f_n (ou une fonction de masse dans le cas discret) et f la densité de X. Alors

$$f_n(x) \overset{n\to\infty}{\longrightarrow} f(x) \text{ pour tout } x, \Rightarrow X_n \overset{d}{\to} X.$$

Attention : l'inverse du théorème de Scheffé n'est pas vrai.

Théorème de continuité ou théorème de Lévy

Attention : à ne pas confondre avec le théorème de l'application continue ! ! !

Théorème de continuité ou théorème de Lévy

Soit $\varphi_n(t) = \mathbb{E}\left[e^{itX_n}\right]$ et $\varphi(t) = \mathbb{E}\left[e^{itX}\right]$ les fonctions caractéristiques de X_n et X respectivement.

(a) $X_n \overset{d}{\to} X \iff \varphi_n \overset{n\to\infty}{\longrightarrow} \varphi$ ponctuellement.

(b) Si $\varphi_n(t)$ converge ponctuellement vers une certaine limite $\psi(t)$ continue en 0. Alors

   (i) Il existe une mesure de probabilité ν ayant $\psi(t)$ comme fonction caractéristique.

   (ii) $F_{X_n} \overset{w}{\to} \nu$ (w pour *weak*).

Convergence faible pour les vecteurs aléatoires

Définition

Soit $\{\mathbf{X}_n\}$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$ et $\mathbf{X}$ un vecteur aléatoire de $\mathbb{R}^d$ où $\mathbf{X}_n = \left(X_n^1, X_n^2, \ldots, X_n^d\right)^\top$ et $\mathbf{X} = \left(X^1, X^2, \ldots, X^d\right)^\top$.
On définit les fonctions de répartitions

$$F_{X_n}(\mathbf{x}) = \mathbb{P}\left[X_n^1 \leq x^1, X_n^2 \leq x^2, \ldots, X_n^d \leq x^d\right]$$

et

$$F(\mathbf{x}) = \mathbb{P}\left[X^1 \leq x^1, X^2 \leq x^2, \ldots, X^d \leq x^d\right]$$

pour tout $\mathbf{x} = \left(x^1, x^2, \ldots, x^d\right)^\top \in \mathbb{R}^d$.
On dit que $\mathbf{X}_n$ converge en loi vers $\mathbf{X}$ quand $n \to \infty$ (et on écrit $\mathbf{X}_n \xrightarrow{d} \mathbf{X}$) si pour tout point de continuité de F_X, nous avons

$$F_{\mathbf{X}_n}(\mathbf{x}) \overset{n\to\infty}{\longrightarrow} F_{\mathbf{X}}(\mathbf{x}).$$

Lien entre les deux versions de la convergence faible (scalaire et vectorielle)

Théorème de Cramér-Wold

Soit $\{\mathbf{X}_n\}$ une suite de vecteurs aléatoires dans $\mathbb{R}^d$. Alors

$$\mathbf{X}_n \xrightarrow{d} \mathbf{X} \Longleftrightarrow \boldsymbol{\theta}^\top \mathbf{X}_n \xrightarrow{d} \boldsymbol{\theta}^\top \mathbf{X}, \forall \boldsymbol{\theta} \in \mathbb{R}^d.$$

Lien entre les deux versions de la convergence faible (scalaire et vectorielle)

Théorème de Cramér-Wold

Soit $\{\mathbf{X}_n\}$ une suite de vecteurs aléatoires dans $\mathbb{R}^d$. Alors

$$\mathbf{X}_n \xrightarrow{d} \mathbf{X} \Longleftrightarrow \boldsymbol{\theta}^\top \mathbf{X}_n \xrightarrow{d} \boldsymbol{\theta}^\top \mathbf{X}, \forall \boldsymbol{\theta} \in \mathbb{R}^d.$$

Nous allons maintenant définir deux autres modes de convergence plus forts que les précédents ! ! !

Convergence presque sûre

Définition (convergence presque sûre)

Soit $\{X_n\}_{n\geq 1}$ et X des variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Soit

$$A := \{\omega \in \Omega : X_n(\omega) \overset{n\to\infty}{\longrightarrow} X(\omega)\}.$$

On dit que X_n converge presque sûrement vers X quand $n \to \infty$ (on écrit $X_n \xrightarrow{p.s} X$) si

$$\mathbb{P}[A] = 1.$$

De manière abusive, on dit que $X_n \xrightarrow{p.s} X$ si $\mathbb{P}[X_n \to X] = 1$.

Convergence dans $\mathbb{L}^p$

Définition (Convergence dans $\mathbb{L}^p$)

Soit $\{X_n\}_{n\geq 1}$ et X des variables aléatoires définies sur le même espace de probabilité. On dit que X_n vers X dans $\mathbb{L}^p$ quand $n \to \infty$ (on écrit $X_n \xrightarrow{\mathbb{L}^p} X$) si

$$\mathbb{E}|X_n - X|^p \overset{n\to\infty}{\longrightarrow} 0.$$

Notons que $\|X\|_{\mathbb{L}^p} = \left(\mathbb{E}\left|X\right|^p\right)^{1/p}$ définit une norme ..... (lorsque cette espérance existe!!!)

Relations entre les différents modes de convergence

- $X_n \xrightarrow{p.s} X \Longrightarrow X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$
- $X_n \xrightarrow{\mathbb{L}^p} X$, pour $p > 0 \Longrightarrow X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$
- Pour $p \geq q, \quad X_n \xrightarrow{\mathbb{L}^p} X \Longrightarrow \quad X_n \xrightarrow{\mathbb{L}^q} X$
- Il n'existe pas de relation *implicative* entre $\xrightarrow{p.s}$ et $\xrightarrow{\mathbb{L}^p}$

Théorème de représentation de Skorokhod

Théorème de représentation de Skorokhod

Soit $\{X_n\}_{n\geq 1}$ et X des variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ où $X_n \xrightarrow{d} X$. Alors, il existe des variables aléatoires $\{Y_n\}_{n\geq 1}$ et Y des variables aléatoires définies sur le même espace de probabilité $(\Omega', \mathcal{G}, \mathbb{Q})$ telles que

(i) $Y \stackrel{d}{=} X \quad \& \quad Y_n \stackrel{d}{=} X_n$, pour tout $n \geq 1$,

(ii) $Y_n \xrightarrow{p.s} Y$

Théorème de représentation de Skorokhod

Théorème de représentation de Skorokhod

Soit $\{X_n\}_{n\geq 1}$ et X des variables aléatoires définies sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ où $X_n \stackrel{d}{\to} X$. Alors, il existe des variables aléatoires $\{Y_n\}_{n\geq 1}$ et Y des variables aléatoires définies sur le même espace de probabilité $(\Omega', \mathcal{G}, \mathbb{Q})$ telles que

(i) $Y \stackrel{d}{=} X \quad \& \quad Y_n \stackrel{d}{=} X_n$, pour tout $n \geq 1$,

(ii) $Y_n \xrightarrow{p.s} Y$

Ce théorème est utilisé pour monter que $X_n \stackrel{d}{\to} X \Longrightarrow g(X_n) \stackrel{d}{\to} g(X)$ pour g continue.

Loi faible des grands nombres

- Rappeler l'inégalité de Tchebychev.
- Montrer que si $\{X_n\}_{n\geq 1}$ une suite de v.a iid où $\mathbb{E}X_k = \mu$ et $\mathbb{E}X_k^2 < \infty$ pour tout k, alors

$$X_n \xrightarrow{p} \mu.$$

Loi forte des grands nombres

Loi forte des grands nombres

Soit $\{X_n\}_{n\geq 1}$ une suite de variables aléatoires iid où $\mathbb{E}(X_k) = \mu$ et $\mathbb{E}|X_k| < \infty$ pour tout $k \geq 1$, alors

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p.s} \mu$$

- **forte** est en opposition à **faible** qui nécessite $\mathbb{E}X_k^2 < \infty$ au lieu de $\mathbb{E}|X_k| < \infty$ et donne une convergence $\xrightarrow{p}$ au lieu de $\xrightarrow{p.s}$.

Théorème de la limite centrale

Théorème centrale limite

Soit $\{\mathbf{X}_n\}_{n\geq 1}$ une suite de vecteurs aléatoires iid à valeurs dans $\mathbb{R}^d$ de moyenne $\boldsymbol{\mu}$ et de matrice de covariance Σ, on définit $\bar{\mathbf{X}}_n := \dfrac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$. Alors

$$\sqrt{n}(\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \sim \mathcal{N}_d\Big(0_d, \Sigma\Big).$$

Vitesse de convergence

- Souvent, montrer la convergence ne suffit pas
- À quelle vitesse ?
- Ça revient à mesurer la qualité d'approximation

Vitesse de convergence

- Souvent, montrer la convergence ne suffit pas
- À quelle vitesse ?
- Ça revient à mesurer la qualité d'approximation

- Loi des grands nombres : variance finie, vitesse $n^{-1/2}$ dans $\mathbb{L}^2$

Vitesse de convergence

- Souvent, montrer la convergence ne suffit pas
- À quelle vitesse ?
- Ça revient à mesurer la qualité d'approximation

- Loi des grands nombres : variance finie, vitesse $n^{-1/2}$ dans $\mathbb{L}^2$

Théorème de Berry-Esseen

Soit $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ des vecteurs aléatoires iid à valeurs dans $\mathbb{R}^d$ tels que $\mathbb{E}\left[\mathbf{X}_i\right] = 0_d$ et $\mathrm{Cov}\left[\mathbf{X}_i\right] = \mathrm{I}_d$. On définit

$$\mathbf{S} = \frac{1}{\sqrt{n}}\left(\mathbf{X}_1 + \cdots + \mathbf{X}_n\right).$$

Si $\mathcal{A}$ est la classe des sous-ensembles convexes de $\mathbb{R}^d$, alors pour $\mathbf{Z} \sim \mathcal{N}_d\left(\mathbf{0}, \mathrm{I}_d\right)$,

$$\sup_{A \in \mathcal{A}} \left| \mathbb{P}\left[\mathbf{S}_n \in A\right] - \mathbb{P}\left[\mathbf{Z} \in A\right] \right| \leq C \frac{d^{1/4} \mathbb{E}\|\mathbf{X}_i\|^3}{\sqrt{n}}.$$