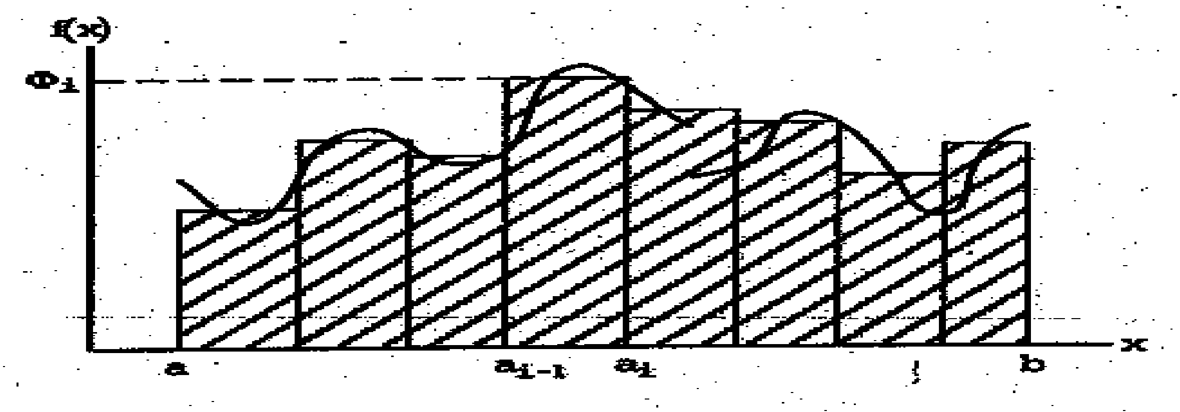
Rappels de probabilités

Romuald Elie

Septembre 2004

1 Rappels de Mesure

1.1 De l'intégrale de Riemann à l'intégrale de Lebesgue

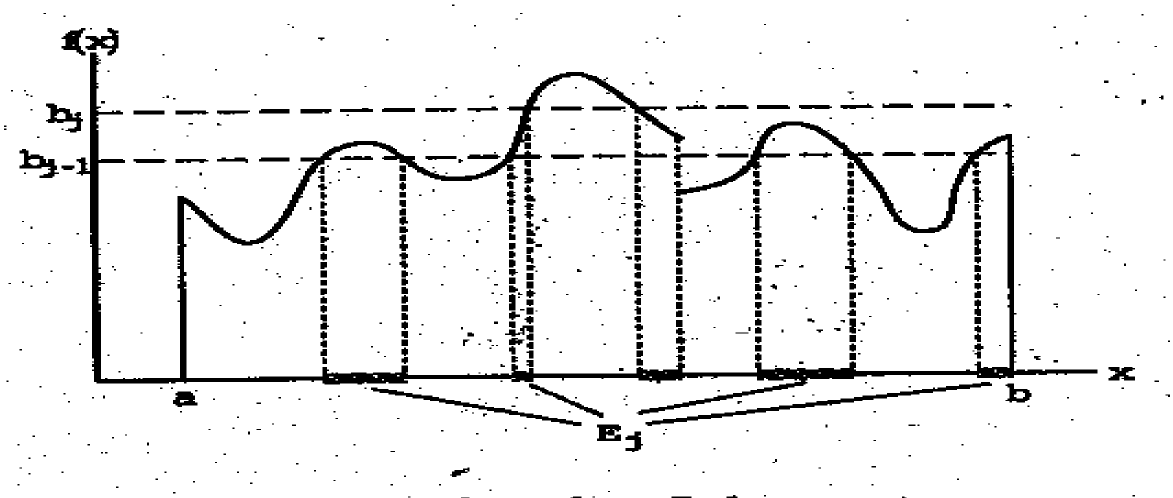


Intégrale selon Riemann

$$\int_a^b f \approx \sum_i \phi_i(a_i - a_{i-1})$$

- La classe des fonctions Riemann-intégrables sur $[a, b]$ est stable par limite uniforme mais pas par limite simple.

- La classe des fonctions Riemann-intégrables (continues) sur $[a, b]$ munie de la semi-norme (norme) $N_1(f) = \int_a^b |f(x)|dx$ n'est pas complet.



Intégrale selon Lebesgue

$$\int_a^b f \approx \sum_j \frac{b_{j-1} + b_j}{2} \times longueur(\{x/b_{j-1} \leq f(x) \leq b_j\})$$

$\Rightarrow$ **Il faut donner un sens à longueur**$(\{\mathbf{x/b_{j-1} \leq f(x) \leq b_j}\})$

1.2 Tribu

Définition 1.1 *On appelle **tribu** $\mathcal{A}$ sur Ω toute famille $\mathcal{A}$ de parties de Ω vérifiant :*

- $\emptyset \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$
- $(\forall n \in \mathbb{N} \quad A_n \in \mathcal{A}) \Rightarrow \bigcup_n A_n \in \mathcal{A}$

Définition 1.2 *La **tribu** $\sigma(\mathcal{C})$ **engendrée** par une famille $\mathcal{C}$ de parties de Ω est la plus petite tribu contenant $\mathcal{C}$. C'est également l'intersection de toutes les tribus contenant $\mathcal{C}$.*

Exemple 1.3 *La tribu des boréliens $\mathcal{B}(\mathbb{R}^n)$ sur $\mathbb{R}^n$ est la tribu engendrée par les ouverts de $\mathbb{R}^n$.*

1.3 Mesure

Définition 1.4 *Une **mesure** μ est une application $\mu : \mathcal{A} \to \overline{\mathbb{R}}^+$ vérifiant*

- $\mu(\emptyset) = 0$
- *Pour $(A_n)_{n\geq 1}$ suite d'éléments de $\mathcal{A}$, deux à deux disjoints,* $\mu(\bigcup_{n\geq 1} A_n) = \sum_{n\geq 1} \mu(A_n)$

1.4 Application mesurable

Définition 1.5 *Une application $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est dite **mesurable (ici borélienne)** si $f^{-1}(\mathcal{B}(\mathbb{R})) \subset \mathcal{A}$*

Remarque 1.6 *Pour montrer qu'une application est borélienne, il suffit de montrer que pour tout $b \in \mathbb{R}$, $\{x, f(x) < b\} \subset \mathcal{A}$.*

Exemple 1.7 *Les fonctions indicatrices et les fonctions continues sont boréliennes.*

Proposition 1.8 *L'ensemble des fonctions mesurables est stable par addition, multiplication, multiplication par un scalaire, passage au sup dénombrable, passage à l'inf dénombrable, passage à la limite sup, à la limite inf.*
*Donc **une limite simple de fonctions mesurables est mesurable**.*

1.5 Construction de l'intégrale de Lebesgue

$$f = \sum_i a_i \mathbf{1}_{A_i} \geq 0 \Rightarrow \int f d\mu = \sum_i a_i \mu(A_i)$$

La construction de l'intégrale de Lebesgue se fait tout d'abord sur les fonctions positives, en commençant par les indicatrices, puis en passant aux fonctions étagées pour arriver enfin par limite croissante aux fonctions mesurables positives. Elle se généralise aux fonctions pouvant être négatives par la décomposition en $f = f^+ - f^-$ avec $f^+ = \max(f, 0)$ et $f^- = \max(-f, 0)$.

L'intégrale obtenue est alors linéaire et monotone.

Toute fonction mesurable positive est dite intégrable, quitte à ce que son intégrale soit infinie. Par, contre pour f uniquement mesurable, on impose $\int |f| < \infty$ pour assurer $\int f^- < \infty$ et $\int f^+ < \infty$ et que leur différence ait un sens.

Dans le cas de variables discrètes, l'intégrale de Lebesgue est simplement une somme.

La notion d'intégrale de Lebesgue prolonge la notion d'intégrale de Riemann.

1.6 Les théorèmes de convergence

Théorème 1.9 ***Théorème de convergence monotone***
Si une suite ***croissante*** *de fonctions mesurables* ***positives*** $(f_n)_{n\in\mathbb{N}}$ *converge vers* f*, alors* $\int f_n$ *converge vers* $\int f$ *(qui est à valeur dans* $\overline{\mathbb{R}}$*).*

Théorème 1.10 ***Lemme de Fatou***
Pour toute suite $(f_n)_{n\in\mathbb{N}}$ *de fonctions mesurables* ***positives****, on a :*

$$\int \underline{lim} f_n \leq \underline{lim} \int f_n$$

.

Démonstration 1.11 .
$\underline{lim} f_n = sup_k h_k$ *où* $h_k := inf_{n\geq k} f_n$.
On a $h_k \leq f_k$ *donc* $\int h_k \leq \int f_k$.
La suite $(h_k)_{k\in\mathbb{N}}$ *est croissante, donc le théorème de convergence monotone donne :*

$$\int \underline{lim} f_n = \int sup_k h_k = \int lim_k h_k = lim_k \int h_k \leq \underline{lim} \int f_n$$

Théorème 1.12 ***Théorème de convergence dominée (ou théoreme de Lebesgue)***
Soit une suite de fonctions mesurables positives $(f_n)_{n\in\mathbb{N}}$ *et* g *une fonction intégrable telles que* $|f_n| < g$ *p.s.*
Alors, si $f_n \to f$ *p.s.,* f *est intégrable et* $\int f_n \to \int f$.

Démonstration 1.13 .
Appliquons Fatou à $\varepsilon_n := 2g - |f_n - f| \geq 0$ *:*

$$\int 2g = \int \underline{lim} \varepsilon_n \leq \underline{lim} \int \varepsilon_n = \int 2g - \int |f_n - f| \Rightarrow \overline{lim} \int |f_n - f| \leq 0$$

Donc $lim \int |f_n - f|$ *existe et est nulle.*
L'inégalité $|\int f_n - \int f| \leq \int |f_n - f|$ *permet alors de conclure.*

2 Espace de probabilité

2.1 Probabilité, variable aléatoire...

Un espace de Probabilité est un triplet $(\Omega, \mathcal{A}, \mathbb{P})$ où Ω est un ensemble, $\mathcal{A}$ une tribu sur Ω et $\mathbb{P}$ une probabilité sur $\mathcal{A}$.

Une **probabilité** $\mathbb{P}$ est une mesure telle que $\mathbb{P}(\Omega) = 1$.
Alors, une fonction mesurable $X : \mathcal{A} \to \mathbb{R}^p$ est appelée **variable aléatoire** (v.a.).
$\mu_X : B \mapsto \mathbb{P}[\omega \in \Omega / X(\omega) \in B]$ est la **loi** de X.

Dans la cas où la v.a. est absolument continue par rapport à la mesure de Lebesgue λ, on appelle **densité** de X, l'application $f_X : \mathbb{R}^p \to \mathbb{R}$ telle que $d\mu_X(x) = f_X(x)d\lambda(x)$. (i.e. $\mu_X(A) = \int_A f d\lambda \quad \forall A \in \mathcal{B}(\mathbb{R}^d)$).

$F_X : x \mapsto \mathbb{P}[X \leq x]$ est la **fonction de répartition** de X.
$\phi_X : \lambda \mapsto \mathbb{E}[e^{i\lambda X}]$ est la **fonction caractéristique** de X.

La **tribu engendrée** par la v.a. X est $X^{-1}(\mathcal{B}(\mathbb{R}^d)) = \{X^{-1}(A) : A \in \mathcal{B}(\mathbb{R}^d)\}$.
On la note $\sigma(X)$ et c'est la plus petite tribu qui rende X mesurable.
En effet une v.a. X est $\mathcal{B}$-mesurable ssi $X^{-1}(\mathcal{B}(\mathbb{R}^d)) C \mathcal{B}.//$

Une v.a. Y est $\sigma(X)$-mesurable ssi elle est de la forme $f(X)$ avec f borelienne.

2.2 Espérance, Variance...

Soient X et Y deux variables aléatoires.

L'**espérance** de X est donnée par $\mathbb{E}(X) := \int_\Omega X(\omega) d\mathbb{P}(\omega) = \int_\mathbb{R} x d\mu_X(x) (= \int x f_X(x) dx$ lorsque f_X existe).

Son **moment d'ordre** p est $\mathbb{E}(X^p)$.

Sa **variance** est $Var(X) := \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

Son **écart type** est donné par $\sigma(X) = \sqrt{Var(X)}$.

La **covariance** de X et Y : $cov(X, Y) := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$.
X et Y sont dites non corrélées ssi $cov(X, Y) = 0$.

2.3 Espace $\mathbb{L}^p$

Pour $p > 0$, on note $\mathbb{L}^p$ l'ensemble des variables aléatoires telles que

$$||X||_p \;:=\; \mathbb{E}[|X^p|]^{\frac{1}{p}} < \infty. \tag{1}$$

Alors $(\mathbb{L}^\mathbf{p}, ||.||_\mathbf{p})$ **est un espace vectoriel complet** .

2.4 Indépendance

Deux événements A et B sont **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A).\mathbb{P}(B)$.

Les n **événements** $A_1, \ldots, A_n$ sont **indépendants** si pour tout $J \subset \{1, \ldots, n\}$ $\mathbb{P}(\cap_{i \in J} A_i) = \prod\limits_{i \in J} \mathbb{P}(A_i)$.

Les **tribus** $\mathcal{A}_1, \ldots, \mathcal{A}_n$ sont **indépendantes** si pour tout $(A_1, \ldots, A_n) \in \mathcal{A}_1 \times \ldots \times \mathcal{A}_n \quad \mathbb{P}(A_1 \cap \ldots \cap A_n) = \prod_{i=1}^n \mathbb{P}(A_i)$.

On dit que les **n variables aléatoires** $X_1, \ldots, X_n$ sont **indépendantes** si les tribus $\sigma(X_1), \ldots, \sigma(X_n)$ sont indépendantes.

Une **suite quelconque de variables aléatoires** $(X_i)_{i \in I}$ est **indépendante** si toute suite finie extraite de cette suite est composée de v.a. indépendantes.

2.5 Vecteur gaussien

Une v.a. X suit une loi $\mathcal{N}(\mu, \sigma^2)$ si elle a une densité qui vaut

$$f_X \;:=\; \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{2}$$

On dit alors que X est une **variable gaussienne**.

Un vecteur de v.a. $(X_1, \ldots, X_n)$ est appelé **vecteur gaussien** si toute combinaison linéaire de ses coordonnées est gaussienne.

Proposition 2.1 *Si $X_1, \ldots, X_n$ sont des gaussiennes indépendantes, alors le vecteur $(X_1, \ldots, X_n)$ est gaussien.*

Proposition 2.2 *Si (X_1, X_2) est un vecteur gaussien avec $cov(X_1, X_2) = 0$, alors X_1 et X_2 sont indépendants.*

3 Inégalités utiles

Inégalité de Jensen

Soient X une v.a. et $\phi : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Alors, on a :

$$\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(X)]$$

Inégalité de Hölder

Soient $p > 1$ et $q > 1$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Soient $f \in \mathbb{L}^p$ et $g \in \mathbb{L}^q$. Alors $fg \in \mathbb{L}^1$ et

$$||fg||_1 \leq ||f||_p.||g||_q$$

Inégalité de Minkowski

Soient $p \geq 1$, $f \in \mathbb{L}^p$ et $g \in \mathbb{L}^p$. Alors $f + g \in \mathbb{L}^p$ et on a :

$$||f + g||_p \leq ||f||_p + ||g||_p$$

Inégalité de Chebyshev

Soient X une v.a., $p > 0$ et $\lambda > 0$. Alors on a :

$$\mathbb{P}(|X| > \lambda) \leq \frac{\mathbb{E}(|X|^p)}{\lambda^p}$$

Croissance en p de la norme $\mathbb{L}^p$

Soit X une v.a., p et q tels que $0 < p < q \leq \infty$. Alors on a :

$$||X||_p \leq ||X||_q$$

4 Espérance conditionnelle

4.1 Probabilités et espérances conditionnelles

Définition 4.1 ***Probabilité conditionnelle***
Soient A et B deux événements. La probabilité conditionnelle de A par rapport à B est définie par $\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Théorème 4.2 ***Espérance conditionnelle par rapport à une tribu***
Soit une variable aléatoire $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et $\mathcal{B} \subset \mathcal{A}$ alors il existe une unique variable aléatoire $\mathcal{B}$-mesurable $\mathbb{E}(X|\mathcal{B})$ appelée ***espérance conditionnelle de X sachant $\mathcal{B}$*** *telle que*

$$\forall B \in \mathcal{B} \qquad \mathbb{E}(X\mathbf{1}_B) = \mathbb{E}(\mathbb{E}[X|\mathcal{B}]\mathbf{1}_B)$$

Remarque 4.3 *Cette caractérisation est équivalente*

$$\forall Y \ \textit{v.a. } \mathcal{B}\textit{-mesurable} \qquad \mathbb{E}(X\mathbf{1}_B) = \mathbb{E}(\mathbb{E}[X|\mathcal{B}]\mathbf{1}_B)$$

Remarque 4.4 *L'espérance conditionnelle est linéaire et monotone.*

4.2 L'espérance conditionnelle comme une projection

Quand $X \in \mathbb{L}^2$, $\mathbb{E}(X|\mathcal{B})$ est la **projection de X sur $\mathbb{L}^2(\Omega, \mathcal{B})$ au sens de la norme** $||.||_2$.
En effet, on a :

$$\begin{aligned} X - \mathbb{E}[X|\mathcal{B}] \perp \mathbb{L}^2(\Omega, \mathcal{B}) &\iff < X - \mathbb{E}[X|\mathcal{B}], \mathbf{1}_B >= 0 \qquad \forall B \in \mathcal{B} \\ &\iff \mathbb{E}[(X - \mathbb{E}[X|\mathcal{B}]) \times \mathbf{1}_B] = 0 \qquad \forall B \in \mathcal{B} \end{aligned}$$

Remarque 4.5 *Par abus de notation, pour X et Y v.a., $\mathbb{E}[X|\sigma(Y)]$ est noté $\mathbb{E}[X|Y]$.*

4.3 Propriétés de l'espérance conditionnelle

Théorème des 3 perpendiculaires (ou Tower Property)

$$\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \mathcal{A} \;\Rightarrow\; \mathbb{E}[\mathbb{E}(X|\mathcal{B}_2)|\mathcal{B}_1] = \mathbb{E}[X|\mathcal{B}_1]$$

Conservation de l'espérance : Pour X v.a. et $\mathcal{B} \subseteq \mathcal{A}$,on a :

$$\mathbb{E}[\mathbb{E}[X|\mathcal{B}]] = \mathbb{E}[X].$$

Sort ce que tu connais : Soit $\mathcal{B} \subseteq \mathcal{A}$, X $\mathcal{A}$-mesurable et Y $\mathcal{B}$-mesurable tels que $\mathbb{E}[|X|.(1+|Y|)] < \infty$. Alors, on a :

$$\mathbb{E}[XY|\mathcal{B}] = Y.\mathbb{E}[X|\mathbf{B}].$$

L'indépendance rend le conditionnement inutile : Soient $\mathbf{B} \subseteq \mathbf{A}$ et X v.a. tels que $\sigma(X)$ et $\mathbf{B}$ sont indépendants. Alors, on a :

$$\mathbb{E}[X|\mathbf{B}] = \mathbb{E}[X].$$

Décroissance en $\mathbb{L}^\mathbf{p}$: Soient $\mathbf{B} \subseteq \mathbf{A}$, X une v.a. et $p \geq 1$. Alors :

$$||\mathbb{E}[X|\mathbf{B}]||_p \leq ||X||_p.$$

Jensen conditionnel : Soient $\mathbf{B} \subseteq \mathbf{A}$, X une v.a. et $\phi : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Alors :

$$\phi(\mathbb{E}[X|\mathbf{B}]) \leq \mathbb{E}[\phi(X)|\mathbf{B}].$$

Théorème de convergence monotone conditionnel : Soit $\mathbf{B} \subseteq \mathbf{A}$. Si une suite **croissante** de v.a. positives $(X_n)_{n\in\mathbb{N}}$ dans $\mathbb{L}^1$ converge vers X, alors $\mathbb{E}[X_n|\mathbf{B}]$ converge vers $\mathbb{E}[X|\mathbf{B}]$.

Lemme de Fatou conditionnel : Pour $\mathbf{B} \subseteq \mathbf{A}$ toute suite $(X_n)_{n\in\mathbb{N}}$ de v.a. positives, on a :

$$\mathbb{E}[\underline{lim} X_n|\mathbf{B}] \leq \underline{lim}\mathbb{E}[X_n|\mathbf{B}].$$

Théorème de convergence dominée conditionnel : Soient $\mathbf{B} \subseteq \mathbf{A}$, $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. positives et Y une v.a. dans $\mathbb{L}^1$ telles que $|X_n| < Y$ p.s.
Alors, si $X_n \to X$ p.s., X est dans $\mathbb{L}^1$ et $\mathbb{E}[X_n|\mathbf{B}] \to \mathbb{E}[X|\mathbf{B}]$.

5 Les différents types de convergence

Soient $(X_n)_{n\in\mathbb{N}}$ une suite de v.a. et X une autre v.a. Alors $(X_n)_{n\in\mathbb{N}}$ peut converger vers X de plusieurs façons différentes.

Convergence presque sûre

$$\begin{aligned} & \mathbb{P}(X_n(\omega) \to X(\omega)) = 1 \\ \iff \quad & \forall \epsilon > 0, \quad \mathbb{P}\left[\, \overline{lim}\ \{|X_n - X| \ > \ \epsilon\} \,\right] \ = \ 0 \end{aligned}$$

Convergence en probabilité

$$\mathbb{P}[|X_n - X| > \epsilon] \to 0 \qquad \forall \epsilon > 0$$

Convergence dans $\mathbb{L}^p$

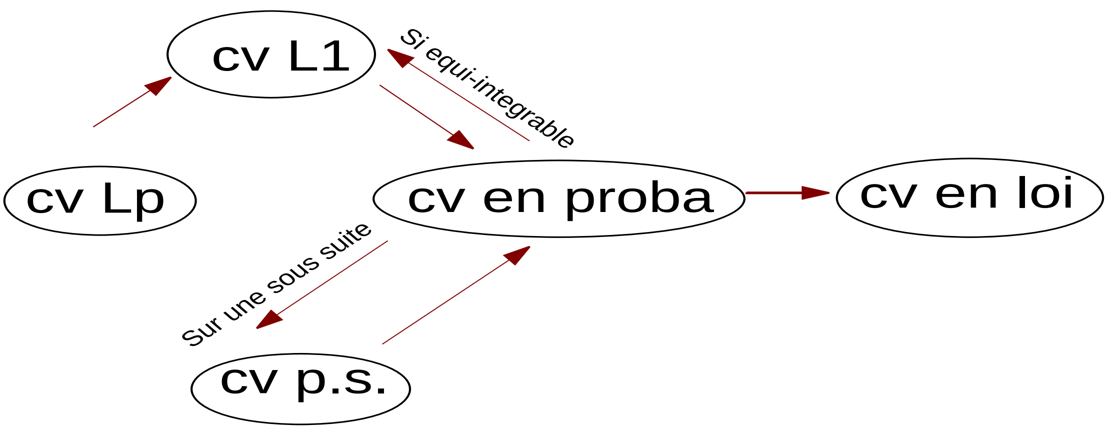
$$\mathbb{E}[|X_n - X|^p] \to 0$$

Convergence en loi

$$\begin{aligned} & \mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \qquad \textit{pour toute } f : \mathbb{R} \to \mathbb{R} \textit{ continue bornee} \\ \iff \quad & F_{X_n}(x) \to F_X(x) \quad \forall x \in \mathbb{C}_F := \{\textit{ensemble des points de continuite de } F\} \\ \iff \quad & \Phi_{X_n}(\lambda) \to \Phi_X(\lambda) \quad \forall \lambda \in \mathbb{R}. \end{aligned}$$

Stabilité de convergence

$$g \text{ continue} \quad \Rightarrow \quad \left\{ \begin{array}{l} X_n \xrightarrow{proba} X \quad \Rightarrow \quad g(X_n) \xrightarrow{proba} g(X) \\ X_n \xrightarrow{loi} X \quad \Rightarrow \quad g(X_n) \xrightarrow{loi} g(X) \end{array} \right.$$



6 Théorèmes limites

La loi des grands nombres (Kolmogorov)

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires dans $\mathbb{L}^1$ indépendantes et de même loi (i.i.d.). Alors, on a :

$$\frac{X_1 + \ldots + X_n}{n} \to \mathbb{E}[X_1] \qquad p.s.$$

Le théorème central limite

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires dans $\mathbb{L}^2$ (i.e. de variance finie) indépendantes et de même loi. Alors, on a :

$$\sqrt{n}\Big(\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}[X_1]\Big) \to \mathcal{N}(0, Var(X_1)) \qquad en\ loi$$

Bibliographie

Mesure et Intégration

Théorie de l'intégration, Marc Briane & Gilles Pages, Vuibert supérieur 1998

Probabilités

Probabilités, D. Revuz Herrmann 1997

En passant par hasard : les probabilités de tous les jours, G. Pages, Vuibert (pour le fun)

Calcul stochastique

Introduction au calcul stochastique appliqué à la finance, D. Lamberton & B. Lapeyre, Ellipses 1999 (vivement conseillé)

Stochastic differential equations, An introduction with applications, Bernt Oksendal, Springer 1995 (difficile mais abordable)

Brownian Motion and Stochastic Calculus, I. Karatzas & S. Shreeve Springer (Bon courage...)