Calcul Stochastique - MMA

Partie 1 : Rappels en Probabilités

Pr. Hamza El Mahjour

Faculté
Polydisciplinaire
Larache
Université Abdelmalek Essaâdi

Points Principaux

Éléments de la théorie de la mesure

Tribus ($\sigma-$algèbre)

Soit E un ensemble quelconque. Soit $\mathcal{F}$ un famille de parties de E. Une tribu est une famille de parties de E vérifiant

- $E \in \mathcal{F}$,
- $A \in \mathcal{F} \implies A^C \in \mathcal{F}$,
- $(A_n)_{n\in\mathbb{N}} \in \mathcal{F} \implies \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

Exemple

- $\mathcal{F}_0 = \{\emptyset, E\}$ est une tribu (dite triviale)
- $\mathcal{P}(E)$ est aussi un tribu (mais pas utilisée en pratique) sauf pour des ensembles dénombrables

Le couple $(E, \mathcal{F})$ est appelée "espace mesurable".

On peut en déduire les choses suivantes à partir des trois hypothèses

Propriétés

(i) $\emptyset \in \mathcal{F}$,

(ii) $\bigcap_{n\in\mathbb{N}} \in \mathcal{F}$,

(iii) Stable par intersections/unions finies.

On peut en déduire les choses suivantes à partir des trois hypothèses

Propriétés

(i) $\emptyset \in \mathcal{F}$,

(ii) $\bigcap_{n\in\mathbb{N}} \in \mathcal{F}$,

(iii) Stable par intersections/unions finies.

Voici une notion importante :
Soit $A \in \mathcal{P}(E)$, on a

$$\sigma(A) = \bigcap \{\text{Toutes les} \quad \text{tribus} \quad \text{contenant } A\}$$

On peut en déduire les choses suivantes à partir des trois hypothèses

Propriétés

(i) $\emptyset \in \mathcal{F}$,

(ii) $\bigcap_{n\in\mathbb{N}} \in \mathcal{F}$,

(iii) Stable par intersections/unions finies.

Voici une notion importante :
Soit $A \in \mathcal{P}(E)$, on a

$$\sigma(A) = \bigcap\{\text{Toutes les} \;\; \text{tribus} \;\; \text{contenant } A\}$$

Souvent, on parle de tribus engendrées par un ensemble

Tribu borélienne

En topologie, on connaît la famille des ouverts d'un espace topologique.

La plus petite tribu contenant tous les **ouverts** d'un espace topologique s'appelle une **tribu borélienne**.

La topologie de $\mathbb{R}$ est engendrée par les intervalles $[a, b]$, $]a, b[$, $[a, +\infty[$ ...

Définition

Supposons que E est un espace topologique et $\mathcal{O}$ la classe des ouverts de E. La tribu $\sigma(\mathcal{O})$ est appelée tribu borélienne et notée $\mathcal{B}(E)$. On appelle les éléments de $\mathcal{B}(E)$ les **boréliens**.

Tribu produit

$(E_1, {}_1)$, $(E_2, {}_2)$ espaces mesurables.
La tribu produit sur $E_1 \times E_2$ définie par

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(A_1 \times A_2; A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2).$$

On a

$$\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}).$$

(exercice)

Mesures positives

On définit sur l'ensemble mesurable $(E, \mathcal{A})$ l'application $\mu : \mathcal{A} \longrightarrow [0, +\infty]$ vérifiant

(i) $\mu(\emptyset) = 0$

(ii) Pour toute famille (A_n) dans $\mathcal{A}$ avec $A_i \bigcap A_j = \emptyset$ on a

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

La propriété (ii) est appelée $\sigma-$additivité.

Mesures positives

On définit sur l'ensemble mesurable $(E, \mathcal{A})$ l'application $\mu : \mathcal{A} \longrightarrow [0, +\infty]$ vérifiant

(i) $\mu(\emptyset) = 0$

(ii) Pour toute famille (A_n) dans $\mathcal{A}$ avec $A_i \bigcap A_j = \emptyset$ on a

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

La propriété (ii) est appelée $\sigma-$additivité.

Est-ce qu'on peut déduire que la $\sigma-$additivité est valable pour des familles finies ?

On énonce quelques conventions de calculs dans $[0, \infty]$ avec $a \in \mathbb{R}^+$

- $a + \infty = \infty, \qquad \infty + a = \infty.$
- $a \times \infty = \infty (a > 0), \quad 0 \times \infty = 0.$

On énonce quelques conventions de calculs dans $[0, \infty]$ avec $a \in \mathbb{R}^+$

- $a + \infty = \infty, \qquad \infty + a = \infty.$
- $a \times \infty = \infty (a > 0), \quad 0 \times \infty = 0.$

Pour quelles raisons insérer l'∞ ?

On énonce quelques conventions de calculs dans $[0, \infty]$ avec $a \in \mathbb{R}^+$

- $a + \infty = \infty, \qquad \infty + a = \infty.$
- $a \times \infty = \infty (a > 0), \quad 0 \times \infty = 0.$

Pour quelles raisons insérer l'∞ ?

Parmi les choses qu'on gagne le fait que toute suite croissante de $[0, \infty]$ va être toujours convergente. On peut même prolonger les propriétés habituelles d'addition et de multiplication des limites de suites ...

Démontrons les propriétés suivantes :

Propriétés

1. Pour A, B quelconques dans $\mathcal{A}$,

$$\mu(A) + \mu(B) = \mu(A \cup B) + \mu(A \cap B)$$

2. Pour $A_n \in \mathcal{A}$ famille croissante

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mu(A_n)$$

3. Pour $A_n \in \mathcal{A}$ famille décroissante et $\mu(A_0) < \infty$.

$$\mu(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mu(A_n)$$

4. Pour $A_n \in \mathcal{A}$ famille quelconque

$$\mu(\bigcup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} A_n$$

Exemple

- **Mesure de comptage** : $E = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ alors $\mu(A) = \mathrm{card}(A)$. (prop 4 cdt)
- **Dirac** : $(E, \mathcal{A})$ quelconque avec $x \in E$,

$$\delta_x(A) = \begin{cases} 1 & \text{si} \quad x \in A \\ 0 & \text{si} \quad x \notin A \end{cases}$$

- **L'exemple des exemples : Lebesgue !** $E = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, notée généralement λ, la mesure de Lebesgue est l'unique mesure telle que $\lambda(]a, b[) = b - a$.

Noms ...

- $\mu(E) \longrightarrow$ **la masse totale.**
- $x \in E, \mu(\{x\}) > 0 \longrightarrow$ **atome** de E.
- Une mesure est diffuse $\longrightarrow$ pas d'atomes.

- μ est finie si $\mu(E) < \infty$.
- μ est de probabilité si $\mu(E) = 1$.
- μ est dite $\sigma-$finie s'il existe une suite croissante de parties mesurables E_n telles que $E = \bigcup_{n\in\mathbb{N}} E_n$ et $\mu(E_n) < \infty$ pour tout n.

Fonctions mesurables

Soit f une application de $(E, \mathcal{A})$ vers $(F, \mathcal{B})$.
f est **mesurable** si

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) \in \mathcal{A}.$$

Exemple

Toute application constante est mesurable !

Propriétés

- f, g mesurables $\longrightarrow \quad g \circ f$ mesurable.
- Si $\mathcal{C} \subset \mathcal{B} = \sigma(\mathcal{C})$. Il suffit que $f^{-1}(B) \in \mathcal{A}$ pour $B \in \mathcal{C}$.

- Sur $(F, \mathcal{B}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ il suffit de montrer que les $f^{-1}(]a, b[)$ sont mesurables.

Entre deux espaces topologiques avec des tribus boréliennes, une application continue est aussi une application mesurable.

Corollaire

Soient $f, g : (E, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *mesurables. Alors les fonctions* $f+g$, fg, $\inf(f, g)$, $f^+ \sup(f, 0)$, $f^- = \sup(-f, 0)$ *sont mesurables.*

On rappelle

$$\limsup a_n = \lim_{n\to\infty} \downarrow \left(\sup_{k\geq n} a_k \right)$$

$$\liminf a_n = \lim_{n\to\infty} \uparrow \left(\inf_{k\geq n} a_k \right)$$

Alors, "lim sup" et "lim inf" sont respectivement la plus grande et la plus petite valeur d'adhérence d'une suite.

Mesure Image

Soit $f : (E, \mathcal{A}) \longrightarrow (F, \mathcal{B})$ une application mesurable et soit μ une mesure positive sur $(E, \mathcal{A})$. La mesure image de μ par f, notée $f(\mu)$ est la mesure positive $(F, \mathcal{B})$ définie par

$$f(\mu)(B) = \mu\big(f^{-1}(B)\big)$$

Montrer que f_μ est bien une mesure.

Remarque

μ est $\sigma-$finie $\quad \not\Rightarrow \quad$ μ_f est $\sigma-$finie

Soit $(E, \mathcal{A})$ un espace mesurable muni d'une mesure positive μ. Une fonction mesurable $f : E \longrightarrow \mathbb{R}$ est **étagée** si elle s'écrit

$$f(x) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}(x), \quad x \in E,$$

où $n \in \mathbb{N}^*$ et $\alpha_1 < \alpha_2 \ldots < a_n$, et pour tout $i = 1 \ldots n$, $A_i = f^{-1}(\{\alpha_i\}) \in \mathcal{A}$.

Définition

Soit f étagée à valeurs positives. L'intégrale de f par rapport à μ est

$$\int f d\mu = \int_E f d\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i) \in [0, \infty],$$

Cas particulier : Si $f = \mathbb{1}_A$ avec $A \in \mathcal{A}$, alors $\int \mathbb{1}_A d\mu = \mu(A)$.

L'intégrale ainsi définie bien sûr respecte les propriétés de linéarité et d'inégalité.

Éventuellement, elle peut valoir l'∞.

Définition

Soit $f : E \longrightarrow [0, \infty]$ une fonction mesurable. On pose

$$\int f d\mu := \sup \left\{ \int g d\mu : g \leq f, \quad g \text{ étagée positive} \right\}$$

Dans ce qui suit on discutera un peu les théorèmes de convergence des intégrales !

Théorème (Beppo-Lévi)

Soit $f_n : (E, \mathcal{F}, \mu) \to [0, +\infty]$ *une suite croissante de fonctions mesurables positives minorant* f*, et soit* $f = \lim f_n$ *la limite ponctuelle des* f_n*. Alors*

$$\int f d\mu = \lim_{n \to +\infty} \int f_n d\mu.$$

limite ponctuelle $\longrightarrow x$ fixé, $\lim_{n\to\infty} f_n(x) = f(x)$.

Démonstration.

- $\int f d\mu \geq \lim \int f_n$ (évident)
- $B_n = \{x \in E, ag(x) \leq f_n(x)\}$, $(0 \leq a < 1)$ ... pour obtenir le deuxième sens d'inégalité. ■

Théorème

Soit $(f_n)_{n\geq 1}$ une suite quelconque de fonctions mesurables positives. Alors $\int \sum_{n\geq 1} f_n d\mu = \sum \int f_n d\mu$.

Preuve : il suffit d'appliquer le théorème de Beppo-Lévi pour la suite des somme partielles $(\sum_{i=1}^{n} f_n)_{n\geq 1}$ qui est une suite croissante.

Théorème (Approximation d'une fonction mesurable)

Soit $f : E \longrightarrow [0, +\infty[$ une fonction mesurable. Il existe une suite croissante de fonctions étagées positives qui converge vers f.

Terminologie : On dit que qu'une propriété est vraie μ presque partout (μ.p.p) si sa négation est fausse sur un ensemble de mesure nulle. Par exemple $f = g$ p.p si

$$\mu(\{x \in E, \quad f(x) \neq g(x)\}) = 0.$$

Propriétés

Soit f une fonction mesurable positive

(i) $\int f d\mu < \infty \implies f < \infty$ p.p.

(ii) $\int f d\mu = 0 \iff f = 0$ p.p.

Démonstration.

(i) Prouver l'inégalité de Markov : $\mu(\{f(x) \geq a\}) \leq \frac{1}{a} \int f d\mu$
$A_n = \{x \in E, f(x) \geq n\}$. $(A_n) \downarrow$, et $\bigcap_{n \geq 1} A_n = \{x \in E : f(x) = \infty\}$.
Utiliser l'inégalité de Markov sur les A_n pour conclure
(ii) Poser $B_n = \{f(x) \geq \frac{1}{n}\}$. Utiliser (i).
$(B_n) \uparrow$ donc $\mu(\bigcup B_n) = \lim \mu(B_n) = 0$. ■

Théorème (Lemme de Fatou)

Soit (f_n) *une suite de fonctions mesurables positives. Alors*

$$\int \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int f_n d\mu.$$

Théorème (Convergence dominée)

Soit (f_n) *une suite de fonctions mesurables vérifiant*

(i) *Convergence p.p :* $\mu\big(\lim_{n\to\infty} f_n \textit{ existe }\big) = 1.$

(ii) *Domination :* $|f_n| \leq g$ *avec* $\int g d\mu < +\infty$

Alors, $\lim_{n\to\infty} f_n d\mu = \int \lim_{n\to\infty} f_n d\mu.$

Intégrale à une mesure image/ à densité

On connaît déjà ce qu'on appelle une mesure image de μ par φ ($\mu_\varphi = \mu \circ \varphi^{-1}$) On dit que f est μ_φ-intégrable ssi f est $\mu \circ \varphi$ intégrable

$$\int f d\mu_\varphi = \int f \circ \varphi d\mu$$

Théorème (Radon-Nikodym)

Soit $(E, \mathcal{F}, \mu)$ un espace mesuré et $f : E \longrightarrow [0, +\infty]$ mesurable. L'application $\nu : \mathcal{F} \to [0, +\infty]$ définie par

$$\nu(A) = \int_A f d\mu = \int f \delta_A d\mu.$$

est aussi une mesure sur $(E, \mathcal{F})$, appelé la mesure de densité f par rapport à μ. De plus, $\varphi : E \to \mathbb{R}$ est ν-intégrable ssi φf est μ-intégrable.

$$\int \varphi d\nu = \int \varphi f d\mu$$

On peut écrire "symboliquement" $d\nu = f d\mu$ ou $d\nu / d\mu = f$.

Probabilités

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé
$\Omega \longrightarrow$ Éventualités possibles,
$\mathcal{F} \longrightarrow$ Événements (parties de Ω),
$\mathbb{P} \longrightarrow$ Mesure de probabilité.

Exemple

- Jet d'une pièce de monnaie non truquée

$$\mathbb{P}(\mathbf{Face}) = \mathbb{P}(\mathbf{Pile}) = 50\% = \frac{1}{2}.$$

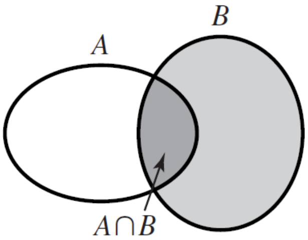
- Lotterie tirage de six nombres (sans ordre). Chaque nombre varie entre 1 et 30. Nombre de tirages possibles : $C_{30}^{6} = \frac{30!}{24!6!} = 593775$

$$\mathbb{P}(\text{Ticket Gagnant}) = \frac{1}{593775} = 1.6841 \times 10^{-6}.$$

Probabilité conditionnelle et indépendance

Une utilisation majeure de la probabilité dans l'inférence statistique est la mise à jour des probabilités lorsque certains événements sont observés. La probabilité mise à jour de l'événement A après avoir appris que l'événement B s'est produit est la probabilité conditionnelle de A sachant B.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{si } \mathbb{P}(B) > 0.$$



Une personne choisit un ticket de loterie contenant les nombres suivants

$$12 - 14 - 27 - 30 - 11 - 07$$

Il allume sa télévision pour suivre les résultats et voit que le nombre 11 a bien été choisi ! Soudain, une panne électrique survient ...

Une personne choisit un ticket de loterie contenant les nombres suivants

$$12 - 14 - 27 - 30 - 11 - 07$$

Il allume sa télévision pour suivre les résultats et voit que le nombre 11 a bien été choisi ! Soudain, une panne électrique survient ...

Quelle est maintenant la probabilité qu'il soit gagnant ?

Une personne choisit un ticket de loterie contenant les nombres suivants

$$12 - 14 - 27 - 30 - 11 - 07$$

Il allume sa télévision pour suivre les résultats et voit que le nombre 11 a bien été choisi ! Soudain, une panne électrique survient ...

Quelle est maintenant la probabilité qu'il soit gagnant ? Appelons
$B = \{\text{un des nombres tirés est } 15\}$ et
$A = \{\text{les nombres tirés sont } 12 - 14 - 27 - 30 - 11 - 07\}$. On remarque que $A \subset B$ donc $\mathbb{P}(A \cap B) = \mathbb{P}(A)$. En on a
$P(B) = \frac{C_{29}^5}{C_{30}^6} = 0,2.$

Une personne choisit un ticket de loterie contenant les nombres suivants

$$12 - 14 - 27 - 30 - 11 - 07$$

Il allume sa télévision pour suivre les résultats et voit que le nombre 11 a bien été choisi ! Soudain, une panne électrique survient ...

Quelle est maintenant la probabilité qu'il soit gagnant ? Appelons $B = \{\text{un des nombres tirés est } 15\}$ et $A = \{\text{les nombres tirés sont } 12 - 14 - 27 - 30 - 11 - 07\}$. On remarque que $A \subset B$ donc $\mathbb{P}(A \cap B) = \mathbb{P}(A)$. En on a $P(B) = \frac{C_{29}^{5}}{C_{30}^{6}} = 0,2$.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Définition (Indépendance de deux événements)

On dit que deux événement sont indépendants si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Autrement dit si $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Exemple

- Si on jette un dé, la probabilité d'obtenir un 5 sachant que dans le jet précédent on a eu 3 est tout simplement $1/6$.
- Si on tire avec remise des boules numérotées de 1 à n d'une urne. La porbabilité d'obtenir k sachant que j'ai tirée l est $\frac{1}{n}$.

Définition (Variable Aléatoire {v.a})

Une **variable aléatoire** est une application $X : \Omega \longrightarrow E$ mesurable où Ω est issu d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ et $(E, \mathcal{A})$ est un espace mesurable.

Définition (Loi d'une v.a)

La loi de la variable aléatoire X est la mesure-image de $\mathbb{P}$ par X. C'est donc la mesure de probabilité sur $(E, \mathcal{A})$ notée $\mathbb{P}_X$, définie par

$$\mathbb{P}_X(B) = \mathbb{P}\left(X^{-1}(B)\right), \quad \forall B \in \mathcal{A}.$$

En pratique,

$$\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \{\omega \in \Omega, \quad X(\omega) \in B\}.$$

Exemple

Lancement de deux dés équilibrés et calcul de la somme des dés.

Types de v.a

- **Discrète** : $X : \Omega \longrightarrow E$ et E dénombrable avec $\mathcal{A} = \mathcal{P}(E)$. Loi de X

$$P_X = \sum_{x \in E} \mathbb{P}(X = x)\delta_x.$$

- **À densité** : $X : \Omega \longrightarrow \mathbb{R}^d$ et $\mathcal{A} = \mathcal{B}(\mathbb{R}^d)$. Si $\mathbb{P}_X$ est absolument continue par rapport à la mesure de Lebesgue λ. Dans ce cas, grâce à Radon-Nikodym, il existe $p : \mathbb{R}^d \longrightarrow \mathbb{R}^+$ borélienne telle que

$$P_X(B) = \int_B p(x)dx.$$

la fonction p s'appelle la densité de (la loi de) X. En dimension 1

$$\mathbb{P}(a \leq X \leq b) = \int_a^b p(x)dx$$

- **mixte** :

Lois de probabilité discrètes

- *Loi uniforme :* Si E est un ensemble fini, $\mathrm{card}(E) = n$,

$$\mathbb{P}(\{X = x\}) = \frac{1}{n}, \quad \forall x \in E.$$

- *Loi de Bernoulli :* de paramètre $p \in [0, 1]$, $X \in \{0, 1\}$.

$$\mathbb{P}(\{X = 1\}) = p, \quad \mathbb{P}(\{X = 0\}) = 1 - p$$

(pièce de monnaie truquée qui tombe sur pile)

- *Loi binômiale* $\mathcal{B}(n, p)$. $X \in \{1, \dots, n\}$ telle que

$$\mathbb{P}(\{X = k\}) = C_n^k \, p^k (1 - p)^{n-k}.$$

(nombre de piles obtenus en n lancers avec une pièce truquée)

- *Loi géométrique* de paramètre $p \in]0,1[$. $X \in \{0,1,2,\ldots\}$.

$$\mathbb{P}(\{X = k\}) = (1-p)p^k$$

(nombre de piles obtenus avant le premier face)

- *Loi de Poisson :* de paramètre $\lambda > 0$. $X \in \{0,1,2,\ldots\}$.

$$\mathbb{P}(\{X = k\}) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad \forall k \in \mathbb{N}.$$

(Intuition : nombre d'événements rares produits durant une longue période)

Lois continues

Soit une v.a $X : \Omega \longrightarrow \mathbb{R}$ à densité $p(x)$

- *Loi uniforme :* sur $[a, b]$

$$p(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

- *Loi exponentielle :* de paramètre $\lambda > 0$

$$p(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}^+}(x).$$

- *Loi gaussienne (normale) :* $\mathcal{N}(m, \sigma^2)$ avec $m \in \mathbb{R}$ et $\sigma > 0$.

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Espérance mathématique

Définition

Soit X une v.a à valeurs dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On note

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega).$$

qui est bien définie

- si $X \geq 0$ ($\mathbb{E}[X] \in [0, \infty]$).
- si X est signée et $\mathbb{E}[|X|] = \int |X| d\mathbb{P} < \infty$

Exemple

On lance deux dés X_1, X_2 et on calcule leur somme $Y = X_1 + X_2$, $Y \in \{2, 3, \dots, 12\}$.

$$\begin{aligned} \mathbb{E}[Y] &= \int_\Omega X d\mathbb{P} \\ &= \sum_{i=1}^{11} y_i p_i = \frac{2\cdot1+3\cdot2+4\cdot3+5\cdot4+6\cdot5+7\cdot6+8\cdot5+9\cdot4+10\cdot3+11\cdot2+12\cdot1}{36} = 7. \end{aligned}$$

$\mathbb{E}[X]$ s'interprète souvent comme la moyenne de la v.a X.

Définition

- Si $k \in \mathbb{N}^*$ alors $\mathbb{E}[X^k]$ s'appelle le **moment d'ordre** k de X.
- Le **premier moment** $\mathbb{E}[X]$ s'appelle généralement une **moyenne** et se note μ ou m.
- Si $\mathbb{E}[X^2] < \infty$ alors la **variance** de X est

$$\mathbf{var}[X] = \mathbb{E}[(X-\mu)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2, = \mathbb{E}[X^2] - \mu^2.$$

À partir de ce qui précède on a : $\mathbb{E}[X^2] \leq \mathbb{E}[X]^2$.
En utilisant les propriétés de l'espérance on peut écrire

$$\mathbf{var}[aX+b] = a^2\mathbf{var}[X]$$

Fonctions de répartition

Définition

Soit X une v.a réelle. La **fonction de répartition** de X est la fonction $F_X : \mathbb{R} \longrightarrow [0, 1]$ définie par

$$F_X(t) = \mathbb{P}(X \leq t) = P_X(]-\infty, t]), \quad \forall t \in \mathbb{R}.$$

Propriétés

La fonction F_X vérifie les propriétés suivante :

- Elle est croissante.
- Elle est continue à droite.
- $\lim\limits_{X \to -\infty} F_X(t) = 0, \quad \lim\limits_{X \to \infty} F_X(t) = 1.$

Tribu engendrée par X

Définition

Soit X une v.a. à valeurs dans $(E, \mathcal{A})$. La tribu engendrée par X, notée $\sigma(X)$ est la plus petite tribu qui rend X mesurable.

$$\sigma(X) = \{A = X^{-1}(B) : B \in \mathcal{A}\}.$$

Proposition

Soit X une v.a à valeurs dans $(E, \mathcal{A})$ et soit Y une v.a réelle.

$$Y \text{ est } \sigma(X)-\text{ mesurable} \iff \exists f : (E, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \text{ mesurable, } Y = f(X)$$

Démonstration.

$\Leftarrow$ évident (composée de fonctions mesurables).
$\Rightarrow$ 1) Y est $\sigma(X)-$mesurable. Supposons $Y = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ (étagée) [$A_i \in \sigma(X)$ et $\alpha_i \in \mathbb{R}$]. Pour chaque i on peut trouver $B_i \in \mathcal{A}$ tel que $A_i = X^{-1}(B_i)$, et on a

$$Y = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i} = \sum_{\lambda_i} \mathbb{1}_{B_i} \circ X = f \circ X,$$

où $f := \sum_{i=1}^{n} \alpha_i \mathbb{1}_{B_i}$ est $\mathcal{A}-$mesurable.
2) Y mesurable donc limite d'une suite de fonctions étagées ! d'après 1) , pour tout n $Y_n = f(X_n)$ où $f_n : E \to \mathbb{R}$ est mesurable. Posons pour tout $x \in E$

$$f(x) = \begin{cases} \lim\limits_{n\to\infty} f_n(x) & \text{si la limite existe.} \\ 0 & \text{sinon.} \end{cases}$$

■

Indépendance des tribus

Définition

- Deux tribus $\mathcal{F}$ et $\mathcal{G}$ sur un même espace Ω sont indépendantes si pour tout $A \in \mathcal{F}$ $B \in \mathcal{G}$ on a

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

- Une famille de tribus $(\mathcal{F}_i)_{i \in I}$ est dite indépendante si pour tous les $A_i \in \mathcal{F}_i$, la famille d'évènements $(A_i)_{i \in I}$ est indépendante.

Exemple

Lancé d'un dé. $\mathcal{F}_1 = \{\emptyset, \Omega, \{1,2\}, \{3,4,5,6\}\}$ et $\mathcal{F}_2 = \{\emptyset, \Omega, \{1,3\}, \{2,4,5,6\}\}$. Soit $A = \{1,2\} \in \mathcal{F}_1$ et $B = \{1,3\} \in \mathcal{F}_2$. On a $\mathbb{P}(A \cap B) = \mathbb{P}(\{1\}) = \frac{1}{6} \neq \mathbb{P}(A) \cdot \mathbb{P}(B) = \frac{2}{6} \cdot \frac{2}{6} = \frac{4}{36}$. Donc $\mathcal{F}_1$ et $\mathcal{F}_2$ ne sont pas indépendants mais dépendantes.

Définition (Indépendance de deux v.a)

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit $(E, \mathcal{A})$ un espace mesurable. Soit $X, Y : \Omega \longrightarrow E$ deux v.a. On dit que X et Y sont indépendants ssi

$$\forall B_1, B_2 \in \mathcal{F}, \quad \mathbb{P}[\{X \in A_1\} \cap \{Y \in A_2\}] = \mathbb{P}[\{X \in A_1\}]\mathbb{P}[\{Y \in A_2\}]$$

Proposition

Toute suite (X_n) de v.a. est indépendante si et seulement si leur tribus associées sont indépendantes.

Convergences ...

Soit $(X_n)_{n\geq}$ une suite de variables aléatoires sur un même espace $(\Omega, \mathcal{F}, \mathbb{P})$.

Définition (Convergence presque sûre)

On dit que X_n converge vers X **presque sûrement** (p.s) ssi

$$\exists A \in \mathcal{F}, \mathbb{P}(A) = 1, \forall \omega \in A, \quad \lim_{n\to\infty} X_n(\omega) = X(\omega).$$

et on note $X_n \xrightarrow{p.s} X$.

Proposition

$$\mathbb{P}\Big(\bigcap_{k\geq 1} \bigcup_{m\geq 1} \bigcap_{m\geq n} \{|X_n - X| < \epsilon\}\Big) = 1.$$

Démonstration.

Réécrire la définition comme suit

$$\mathbb{P}\Big(\bigcap_{k\geq 1}\bigcup_{m\geq 1}\bigcap_{m\geq n}\{|X_n - X| < \epsilon\}\Big) = 1.$$

Définition (Convergence en probabilité)

$$X_n \xrightarrow{\mathbb{P}} X \iff \lim_{n\to\infty} \mathbb{P}(\{|X_n - X| > \epsilon\}) = 0.$$

Lemma

$$X_n \xrightarrow{p.s} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$$

Démonstration.

On a ... ∎

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

On dit que X_n converge vers X en $\mathbb{L}^p$ pour $p > 0$ si

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|^p\right] = 0.$$

Lemma

$$X_n \xrightarrow{\mathbb{L}^p} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X$$

Proposition

Si la suite X_n converge en probabilité vers X alors il existe une sous-suite X_{n_k} qui converge vers X p.s.

Résumé - convergences

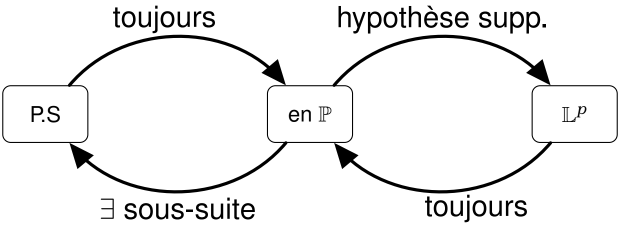


FIGURE – Implications entre les différents types de convergence des suites de variables aléatoires.

Lois des grands nombres

L'histoire d'une convergence dangereuse !

On veut établir un résultat concernant la convergence d'une somme de v.a.r du style

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \longrightarrow \quad ?$$

L'histoire d'une convergence dangereuse !

On veut établir un résultat concernant la convergence d'une somme de v.a.r du style

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \longrightarrow \quad ?$$

C'est pourquoi nous allons petit à petit explorer cette idée !

Somme de v.a indépendantes

Nous allons ici donner une autre caractérisation de l'indépendance entre des variables aléatoires.

Théorème

On dit que n variables aléatoires $X_1, X_2, \ldots X_n$ sont indépendantes si la loi du n-uplet $(X_1, X_2, \ldots, X_n)$ est produits des lois de X_i, c-à-d

$$\mathbb{P}_{(X_1,\ldots,X_n)} = \mathbb{P}_{X_1} \otimes \ldots \mathbb{P}_{X_n}$$

de plus, si f_i sont des fonctions mesurables positives sur chaque $(E_i, \mathcal{F}_i)$ alors

$$\mathbb{E}[\prod_{i=1}^{n} f_i(X_i)] = \prod_{i=1}^{n} \mathbb{E}[f_i(X_i)].$$

Définition

Si μ et ν sont deux mesures de probabilité sur $\mathbb{R}^d$ sur $\mathbb{R}^d$, on note $\mu * \nu$ la mesure-image de $\mu \otimes \nu$ par l'application $(x, y) \mapsto x + y$: pour toute fonction mesurable positive sur $\mathbb{R}^d$:

$$\int_{\mathbb{R}^d} \varphi(z) \mu * \nu(dz) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \varphi(x + y) \mu(dx) \nu(dy)$$

Proposition

Soient X et Y deux v.a. indépendantes dans $\mathbb{R}^d$.

(i) *La loi de $X + Y$ est $\mathbb{P}_X * \mathbb{P}_Y$. En particulier si X et et Y sont à densités p_x et p_y alors la loi de $X + Y$ a une densité $p_x * p_y$.*

(ii) *Si $\mathbb{E}[|X|^2], \mathbb{E}[|Y|^2] < \infty$ et si on note K_X et K_Y les matrices de covariance de X et de Y alors $K_{X+Y} = K_X + K_Y$. Quand $d = 1$ on a $\mathbf{var}(X + Y) = \mathbf{var}(X) + \mathbf{var}(Y)$.*

Démonstration : (esquisse).

(i) On applique la définition de $\mathbb{P}_X * \mathbb{P}_Y$.

(ii) Il suffit de remarquer que $X = (X_1, \dots, X_d)$ et $Y = (Y_1, \dots, Y_d)$ sont indépendants implique que $cov(X_i, Y_j) = 0$ pour tout $i,j = 1...d$. Et par bilinéarité $cov(X_i + Y_i, X_j + Y_j) = cov(X_i, X_j) + cov(Y_i, Y_j)$.

$\blacksquare$

Lois faibles des grands nombres

Le théorème suivant est un résultat préliminaire sur la loi faible des grands nombres. Il existe d'autres versions de lois faibles, mais c'est une passerelle pour arriver à voir la loi forte des grands nombres.

Théorème

*Soit (X_n) une suite de variables aléatoires **non corrélées** avec $\mathbb{E}[X_i] = \mu$ et $\mathbf{var}[X_i] < \infty$. Si $S_n = \sum_{i=1}^{n} X_i$ alors $\frac{S_n}{n}$ converge vers μ en $\mathbb{L}^2$ et en probabilité.*

Lois faibles des grands nombres

Le théorème suivant est un résultat préliminaire sur la loi faible des grands nombres. Il existe d'autres versions de lois faibles, mais c'est une passerelle pour arriver à voir la loi forte des grands nombres.

Théorème

Soit (X_n) une suite de variables aléatoires ***non corrélées*** *avec $\mathbb{E}[X_i] = \mu$ et $\mathbf{var}[X_i] < \infty$. Si $S_n = \sum_{i=1}^{n} X_i$ alors $\frac{S_n}{n}$ converge vers μ en $\mathbb{L}^2$ et en probabilité.*

Avant de démontrer, nous avons besoin de :

Lois faibles des grands nombres

Le théorème suivant est un résultat préliminaire sur la loi faible des grands nombres. Il existe d'autres versions de lois faibles, mais c'est une passerelle pour arriver à voir la loi forte des grands nombres.

Théorème

Soit (X_n) une suite de variables aléatoires ***non corrélées*** *avec $\mathbb{E}[X_i] = \mu$ et $\mathbf{var}[X_i] < \infty$. Si $S_n = \sum_{i=1}^{n} X_i$ alors $\frac{S_n}{n}$ converge vers μ en $\mathbb{L}^2$ et en probabilité.*

Avant de démontrer, nous avons besoin de :

Définition (Corrélation)

Soit X_i une famille de v.a avec $\mathbb{E}[X_i^2] < \infty$. On dit que la famille X_i n'est pas corrélée si $\mathbb{E}[X_iX_j] = \mathbb{E}[X_i]\mathbb{E}[X_j]$ pour chaque $i \neq j$

Théorème (Résultat 1 : Variance des sommes)

Soit $X_i, i = 1 \ldots n \quad (\mathbb{E}[X_i^2] < \infty)$ *une famille de v.a non corrélées* $\mathbf{var}[X_1 + \ldots + X_n] = \mathbf{var}[X_1] + \ldots + \mathbf{var}[X_n]$*.*

Démonstration.

- Utiliser la définition de la variance
- Écrire le carré de la somme comme un produit et développer
- Écrire

$$\mathbf{var}[S_n] = \sum_{i=1}^{n} \mathbb{E}[(X_i - \mu_i)^2] + 2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)]$$

Observez que $\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = 0$ pour $i \neq j$. ■

Proposition (Résultat 2)

$$\mathbf{var}[cY] = c^2\mathbf{var}[Y]$$

Proposition (Résultat 3)

Si $p > 0$ *et* $\mathbb{E}[|X_n|^p]$ *converge vers* 0 *alors* X_n *converge vers 0 en proba.*

Démonstration.

Remarquez que $\mathbb{E}\left[\frac{S_n}{n}\right] = \mu$, donc $\mathbb{E}\left[\frac{S_n}{n} - \mu\right] = \mathbf{var}(\frac{S_n}{n}) = \frac{1}{n^2}\left(\mathbf{var}(X_1) + \ldots + \mathbf{var}(X_n)\right) \leq \frac{Cn}{n^2}$ qui tend vers 0 si n tend vers l'∞. ∎

Lois fortes de grands nombres

Remarque

Convergence en $\mathbb{P}$ ==> lois faibles ; Convergence p.s ==> lois fortes.

On introduit le théorème de Borel-Cantelli pour obtenir une première version de la loi faible des grands nombres.

Théorème (lemme de Borel-Cantelli)

Si $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$ *alors*

$$\mathbb{P}\left(\bigcap_i A_i, \quad \textit{infiniment souvent}\right) = 0.$$

En d'autres termes, le théorème de Borel-Cantelli que si la somme des probabilités d'une suite $(A_i$ d'événement alors la probabilité qu'une infinité d'entre eux se réalisent simultanément est nulle.

Une version de la loi forte des grands nombres (conditions non-optimales)

Théorème (version 1)

Soit $(X_n)_{n\geq 1}$ *une de v.a.r qui sont i.i.d. Si* $\mathbb{E}[X_1^4] < \infty$*, alors*

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

Preuve(esquisse).

- On pose $X_i' = X_i - \mu$ et $S_n = X_1 + \ldots + X_n$.
- Supposer (sans perte de généralité) que $\mu = 0$.
- Remarquer que $\left(\mathbb{E}\left[\sum_{i=1}^n X_i\right]\right)^4 = \mathbb{E}[\sum_{1\leq i,j,k,l\leq n} X_iX_jX_kX_l]$
- En fait, $\mathbb{E}\left[X_i^3X_j\right]$, $\mathbb{E}\left[X_i^2X_jX_k\right]$ et $\mathbb{E}\left[X_iX_jX_kX_l\right]$ sont nuls si i, j, k, l sont différents deux à deux.
- Chebyshev nous donne $\mathbb{P}(|S_n| > n\epsilon) \leq \mathbb{E}[S_n^4]/(n\epsilon)^4 \leq C/(n^2\epsilon^4)$
- Conclure par Borel-Cantelli. $\blacksquare$

Ce théorème aide à démontrer la loi forte des grands nombres (version optimale)

Théorème (Loi du tout ou rien)

Soit $(X_n)_{n\geq 1}$ *une suite de v.a à valeurs dans des espaces mesurables quelconques. Pour tout* $n \geq 1$ *on définit la tribu* $\mathcal{B}_n = \sigma(X_k, k \geq n)$*. Alors la tribu asymptotique* $\mathcal{B}_\infty = \bigcap_{n=1}^{\infty} \mathcal{B}_n$ *est "grossière" dans le sens où* $\forall B \in \mathcal{B}_i nfty, \quad \mathbb{P}(B) = 1$ *ou* $\mathbb{P}(B) = 0$ *!*

Démonstration.

Posons $\mathcal{D}_n = \sigma(X_k, k \leq n)$. On va admettre que $\mathcal{B}_{n+1}$ est indépendante de $\mathcal{D}_n$ (c'est pas difficile à démontrer) Donc $\mathcal{D}_n$ est indépendante de $\mathcal{B}_\infty$. Ainsi

$$\forall A \in \bigcup_{n=1}^{\infty} \mathcal{D}_n, \forall B \in \mathcal{B}_\infty, \quad \mathbb{P}(A \cap B) = \mathbb{P}(A) \cap \mathbb{P}(B).$$

■

Démonstration.

Puisque $\bigcup_{n=1}^{\infty} \mathcal{D}_n$ est stable par intersections finies. On conclut que $\mathcal{B}_\infty$ est indépendante de $\bigcup_{n=1}^{\infty} \mathcal{D}_n = \sigma(X_n, n \geq 1)$. En particulier $\mathcal{B}_\infty$ est indépendante d'elle même ! ! Donc

$$\forall B \in \mathcal{B}_\infty, \quad \mathbb{P}(B) = \mathbb{P}(B \cap B) = \mathbb{P}(B) \cdot \mathbb{P}(B) = \mathbb{P}(B)^2.$$

Donc $\mathbb{P}(B)$ = 1 ou 0 pour tout $B \in \mathcal{B}_\infty$. ■

Théorème (Loi forte des grands nombres)

Soit (X_n) une suite de variables aléatoires i.i.d intégrables (de $\mathbb{L}^1$) alors

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

Je ne vais pas démontrer le résultat précédent.
Voici quelques remarques ...

Supposons que nous permettons que certains $\mathbb{E}[X_i] = \infty$.
Comment procéder ? On décompose les X_i de sorte que $\mathbb{E}[X_i^+] = \infty$ et $\mathbb{E}[X_i^+] < \infty$. Soit $M > 0$ et $X_i^M = \inf(X_i, M)$. Les v.a X_i^M ainsi formées sont $i.i.d$ aussi avec $\mathbb{E}[X_i^M] < \infty$ pour tout i. Donc on peut appliquer la loi forte des grands nombres ! Posons $S_n^M = \sum_{k=0}^n X_k^M$ alors

$S_n^M \xrightarrow[n\to\infty]{p.s} \mathbb{E}[X_1^M]$. Et par construction $X_i^M \leq X_i$ c'est à dire $S_n \leq S_n^M$,

donc $\mathbb{E}\left[X_i^M\right] = \liminf_{n\to\infty} \frac{S_n^M}{n} \leq \liminf_{n\to\infty} \frac{S_n}{n}$. En employant le théorème de convergence monotone quand $M \to \infty$. Donc
$\mathbb{E}[X_i^M]\mathbb{E}[X_i^M]^+ \to \mathbb{E}[X_i^+] = \infty$