memo de probabilités et de statistiques

Simon Queric

2022-2023

# 1 Tribu, $\pi$-système, $\lambda$-système

Une **tribu** $\mathcal{F}$ sur un ensemble $\Omega$ est une famille de parties de $\Omega$ qui contient $\Omega$, stable par passage au complémentaire et stable par union dénombrable.

Un **$\pi$-système** $\mathcal{C}$ sur un ensemble $\Omega$ est un ensemble de parties de $\Omega$ non vide et stable par intersection finie.

Un **$\lambda$-système** $\Lambda$ sur un ensemble $\Omega$ est un ensemble de parties de $\Omega$ stables par passage au complémentaire et par union dénombrable de suites d'éléments deux à deux disjoints de $\Lambda$.

**Théorème de classe monotone** Si $\mathcal{C} \subseteq \Lambda$ alors $\sigma(\mathcal{C}) \subseteq \Lambda$

# 2 Espérance, Variance, Covariance

Soit $X$ une v.a à valeurs complexes. Quand elle existe on définie son espérance par $\mathbb{E}(X) = \int_\Omega X \mathrm{d}\mathbb{P}$

**Théorème de transfert** $\displaystyle\int_\Omega f \circ X \mathrm{d}\mathbb{P} = \int_{\mathrm{X}} f \mathrm{d}P^X$

$\mathrm{Cov}(X) = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))^H)$ (pour le cas où $X$ est complexe/réelle il s'agit de la variance de $X$)

$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^H)$

$\mathbb{E}(AX + b) = A\,\mathbb{E}(X) + b,\ \mathrm{Cov}(AX + b) = A\mathrm{Cov}(X)A^H$

**Inégalité de Markov** $\mathbb{P}(|X| \geqslant a) \leqslant \dfrac{\mathbb{E}(|X|)}{a}$

Soit $X, Y$ deux variables aléatoires réelles.

**Inégalité de Cauchy-Schwartz** $\mathbb{E}(|XY|) \leqslant \mathbb{E}(X^2)^{1/2}\,\mathbb{E}(Y^2)^{1/2}$

**Inégalité de Hölder** $\mathbb{E}(|XY|) \leqslant \mathbb{E}(|X|^p)^{1/p}\,\mathbb{E}(|Y|^q)^{1/q}$ pour $\dfrac{1}{p} + \dfrac{1}{q} = 1$

**Inégalité de Jensen** Si $\phi$ est convexe et que $\phi(X)$ est $L^1$ alors $\phi(\mathbb{E}(X)) \leqslant \mathbb{E}(\phi(X))$

# 3 Espérance conditionnelle

Soit $\mathcal{G} \subseteq \mathcal{F}$ une sous tribu. Si $X$ est $L^1$, il existe une unique variable aléatoire $Y$ dans $L^1(\Omega, \mathcal{G}, \mathbb{P})$ tq $\forall A \in \mathcal{G}\ \mathbb{E}(\mathbb{1}_A Y) = \mathbb{E}(\mathbb{1}_A X)$ On note cette v.a $\mathbb{E}(X|\mathcal{G})$

Dans le cas où $X$ est $L^2$ on a $\mathbb{E}(X|\mathcal{G}) = \mathrm{proj}(X|L^2(\Omega, \mathcal{G}, \mathbb{P}))$

### Propriétés

L'espérance conditionnelle est linéaire et positive. Si $X$ est $\mathcal{G}$ mesurable $\mathbb{E}(X|\mathcal{G}) = X$. Si $X \perp\!\!\!\perp \mathcal{G}$ ou si $\mathcal{G} = \{\Omega, \emptyset\}$ $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.

(*i*) $\mathbb{E}(|\,\mathbb{E}(X|\mathcal{G})|) \leqslant \mathbb{E}(|X|)$

(*ii*) $\mathbb{E}(X|\mathcal{G}) \vee \mathbb{E}(Y|\mathcal{G}) \leqslant \mathbb{E}(X \vee Y|\mathcal{G})$

(*iii*) $\mathbb{E}(X|\mathcal{G})_+ \leqslant \mathbb{E}(X_+|\mathcal{G})$

(*iv*) $|\,\mathbb{E}(X|\mathcal{G})| \leqslant \mathbb{E}(|X||\mathcal{G})$

(*v*) (Tower property ) Si $\mathcal{G} \subseteq \mathcal{H}$ alors $\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})$

(*vi*) $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$

(*vii*) Si $X$ est $\mathcal{G}$ mesurable et $XY$ est $L^1$ alors : $\mathbb{E}(XY|\mathcal{G}) = X\,\mathbb{E}(Y|\mathcal{G})$

(*viii*) Si $\phi$ est une fonction convexe et si $\phi(X)$ est $L^1$ alors $\phi(\mathbb{E}(X|\mathcal{G})) \leqslant \mathbb{E}(\phi(X)|\mathcal{G})$

# 4 Noyau et loi conditionnelle

Soit $(E_1, \mathcal{E}_1), (E_2, \mathcal{E}_2)$ deux espaces mesurables. Un noyau est une application $N : E_2 \times \mathcal{E}_1 \to [0, +\infty]$ tel que : (i) $\forall x \in E_2$ $N(x, .)$ est une mesure et (ii) $\forall A \in \mathcal{E}_1$ $N(., A)$ est mesurable.

De plus, $N$ est un noyau de probabilité si $\forall x \in E_2$ $N(x, .)$ est une mesure de probabilité.

Loi sachant $\mathcal{G}$ $\mathbb{P}^{\mathcal{G}}$ : noyau de probabilité $\Omega \times \mathcal{F}$ tel que $\mathbb{P}^{\mathcal{G}}(., A) = \mathbb{E}(\mathbb{1}_A | \mathcal{G})$

Loi de $X$ sachant $\mathcal{G}$ $\mathbb{P}^{X|\mathcal{G}}$ : noyau de probabilité $\Omega \times \mathcal{X}$ tel que $\mathbb{P}^{X|\mathcal{G}}(., A) = \mathbb{E}(\mathbb{1}_A(X) | \mathcal{G})$

Loi de $X$ sachant $Y$ $\mathbb{P}^{X|Y}$ : noyau de probabilité $\mathrm{Y} \times \mathcal{X}$ tel que $\mathbb{P}^{X|Y}(Y, A) = \mathbb{E}(\mathbb{1}_A(X) | Y)$

On définit la densité conditionnelle de $X$ sachant $Y$ par $f^{X|Y}(x|y) = \dfrac{f^{(X,Y)}(x, y)}{f^Y(y)}$

Soit $N$ un noyau sur $\mathbf{X} \times \mathcal{Y}$ et $\mu \in \mathbb{M}_+(\mathbf{X}, \mathcal{X})$ on peut définir la mesure $\mu \otimes N$ sur $\mathcal{X} \otimes \mathcal{Y}$ par $\mu \otimes N(C) = \displaystyle\int \left( \int \mathbb{1}_C(x, y) N(x, \mathrm{d}y) \right) \mu(\mathrm{d}x)$

**Théorème de désintégration** Soient $X$ et $Y$ deux variables aléatoires de $(\Omega, \mathcal{F}, \mathbb{P})$ dans $\mathbf{X}$ et $\mathbf{Y}$ respectivement. Si $\mathbb{P}^{(X,Y)} = \mu \otimes N$ alors $\mu = \mathbb{P}^X$ et $\mathbb{P}^{Y|X} = N$

# 5 Difféomorphisme et changement de variable

Soit $\phi : U \to V$ un difféomorphisme. On note $J_\phi = \left( \dfrac{\partial \phi_i}{\partial x_j} \right)_{1 \leqslant i,j \leqslant n}$ le jacobien de $\phi$. On a $\displaystyle\int_U f = \int_V \frac{f \circ \phi^{-1}}{|(\det J_\phi) \circ \phi^{-1}|}$

Si $Y = \phi(X)$ alors $f_Y = \dfrac{f_X \circ \phi^{-1}}{|(\det J_\phi) \circ \phi^{-1}|} \mathbb{1}_V$

# 6 Lois usuelles

## 6.1 Lois discrètes

| Nom | Mesure de probabilité $\mu$ | Fonction caractéristique | $m$ | $\sigma^2$ |
|---|---|---|---|---|
| Loi de Bernoulli $\mathcal{B}(p)$ | $p\delta_0 + q\delta_1$ | $pe^{it} + q$ | $p$ | $pq$ |
| Loi Binomiale $\mathcal{B}(n, p)$ | $\displaystyle\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \delta_k$ | $(pe^{it} + q)^n$ | $np$ | $npq$ |
| Loi de Poisson $\mathcal{P}(\lambda)$ | $\displaystyle\sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} \delta_k$ | $\exp(\lambda(e^{it} - 1))$ | $\lambda$ | $\lambda$ |
| Loi géométrique $\mathcal{G}(p)$ sur $\mathbb{N}$ | $\displaystyle\sum_{n=0}^{+\infty} pq^n \delta_n$ | $\dfrac{p}{1 - qe^{it}}$ | $\dfrac{q}{p}$ | $\dfrac{q}{p^2}$ |
| Loi géométrique $\mathcal{G}(p)$ sur $\mathbb{N}^*$ | $\displaystyle\sum_{n=1}^{+\infty} pq^{n-1} \delta_n$ | $\dfrac{pe^{it}}{1 - qe^{it}}$ | $\dfrac{1}{p}$ | $\dfrac{q}{p^2}$ |

### 6.2 Lois à densité par rapport à la mesure de Lebesgue

| Nom | Densité | Fonction caractéristique | $m$ | $\sigma^2$ |
|---|---|---|---|---|
| Loi uniforme $\mathcal{U}([a,b])$ | $\frac{1}{b-a}\mathbb{1}_{[a,b]}$ | $\frac{e^{itb}-e^{ita}}{it(b-a)}$ | $\frac{b-a}{2}$ | $\frac{(b-a)^2}{12}$ |
| Loi de Cauchy | $\frac{1}{\pi}\frac{1}{1+x^2}$ | $e^{-\lvert t\rvert}$ | n'existe pas | n'existe pas |
| Loi normale $\mathcal{N}(m,\sigma^2)$ | $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-m)^2}{2\sigma^2}}$ | $e^{itm-\frac{t^2\sigma^2}{2}}$ | $m$ | $\sigma^2$ |
| Loi exponentielle $\mathcal{E}(\lambda)$ | $\mathbb{1}_{\mathbb{R}_+}(x)\lambda e^{-\lambda x}$ | $\left(1-\frac{it}{\lambda}\right)^{-1}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Loi Gamma $\Gamma(a,b)$ | $\mathbb{1}_{\mathbb{R}_+}(x)\frac{b^a}{\Gamma(a)}e^{-bx}x^{a-1}$ | $\left(1-\frac{it}{b}\right)^{-a}$ | $\frac{a}{b}$ | $\frac{a}{b^2}$ |
| Loi du chi-deux à $n$ degrés de liberté $\chi^2_n$ | $\mathbb{1}_{\mathbb{R}_+}(x)\frac{e^{-\frac{x}{2}}x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}$ | $(1-2it)^{-\frac{n}{2}}$ | $n$ | $2n$ |

# 7 Vecteurs gaussiens

**Définition :** $X=(X_1,\cdots,X_d)$ est un vecteur gaussien si pour tout $t\in\mathbb{R}^d$ $\langle t,X\rangle$ suit une loi normale.

Si $X\sim\mathcal{N}(m,\Gamma)$ alors $\phi_X(t)=\mathbb{E}(e^{i\langle t,X\rangle})=e^{i\langle t,m\rangle-\frac{1}{2}t^T\Gamma t}$ et si $\Gamma\in GL_n(\mathbb{R})$ alors :

$$f_X(x_1,\cdots,x_d)=\frac{1}{\sqrt{(2\pi)^d\lvert\det\Gamma\rvert}}\exp\Big(-\frac{1}{2}\langle\Gamma^{-1}(x-m),x-m\rangle\Big)$$

Soient $(X_1,\cdots,X_n)$ $n$ variables aléatoires. Les deux assertions sont équivalentes :

*(i)* $X_1,\cdots,X_n$ *sont* $n$ *variables gaussiennes indépendantes*

*(ii)* $(X_1,\cdots,X_n)$ *est un vecteur gaussien dont la matrice de covariance est diagonale*

Autrement dit, pour un vecteur gaussien, la décorrelation implique l'indépendance.

# 8 Statistiques

Un **modèle statistique** est une famille de mesures de probabilité. Une **statistique** $S$ est une fonction mesurable des observations. Un **estimateur** est une statistique qui cherche à estimer un paramètre.

On considère des modèles **dominés** i.e tels qu'il existe une mesure $\mu$ telle que pour toute probabilité $P$ du modèle $P$ admette une densité par rapport à $\mu$. Les modèles considérés sont également **paramétrisés** i.e qu'il existe une bijection $\theta\in\Theta\mapsto P_\theta\in\mathcal{P}$.

L'**erreur quadratique moyenne** est :

$$EQM(\hat{g},\theta):=\mathbb{E}_\theta((\hat{g}(X)-g(\theta))^2)=\mathbb{E}_\theta(\hat{g}(X)-g(\theta))^2+V_\theta(\hat{g}(X))=b(\hat{g},\theta)^2+V_\theta(\hat{g}(X))=\text{biais}^2+\text{variance}$$

Une **statistique suffisante** est une statistique $S$ telle que pour toute probabilité du modèle $P\in\mathcal{P}$ la loi de $X|S$ ne dépend pas de $P$.

**Théorème de factorisation de Fisher :** Si $S = g(X)$ et que $\forall \theta \in \Theta \; p_\theta = h \times f_\theta \circ g$ alors $S$ est une statistique suffisante.

# 9 Intuitions

Une statistique suffisante est une fonction des observations qui conserve l'information nécessaire à la determination du paramètre $\theta$. En théorie de l'information : $I(X, \theta) = I(g(X), \theta)$.

Une loi conditionnelle est une probabilité paramétrée. C'est une loi de probabilité qui est elle même "aléatoire". La notation $K(x, A)$ correspond a l'intuition $Y \in A$ sachant $X = x$.

# 10 Chaînes de Markov

**Définition** Une chaîne de Markov est un processus filtré $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ tel que :

$$\forall A \in \mathcal{F} \;\; \mathbb{P}(X_{n+1} \in A \mid \mathcal{F}_n) = \mathbb{P}(X_{n+1} \in A \mid X_n)$$

Une chaîne de Markov est caractérisée par sa loi initiale $\nu$ et ses noyaux de transition $(P_k)_{k\in\mathbb{N}}$. Une chaîne de Markov est **homogène** s'il existe un noyau $P$ tel que pour tout $k \in \mathbb{N}$ et tout $A \in \mathcal{X}$ on ait $\mathbb{P}(X_{k+1} \in A \mid X_k) = P(X_k, A)$. La loi de $X_k$ est alors $\nu P^k$

# 11 Martingales à temps discret

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_n)_n, \mathbb{P})$ un espace de probabilité filtré.

**Définition** Un processus $L^1$ adapté $(X_n)_{n\in\mathbb{N}}$ est une martingale si $\forall n \in \mathbb{N}$ $\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$.

**Exemple 1 (Marche aléatoire)** Soit $(X_k)_k$ une suite de variables intégrables, indépendantes de moyenne nulle. Alors le processus $(S_n)_n$ définit par $S_0 = 0$ et $S_n = \sum_{k=1}^n X_k$ est une martingale (pour la filtration $\mathcal{F}_0 = \{\emptyset, \Omega\}$ et $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$)

**Exemple 2 (Martingale produit)** Soit $(X_k)_k$ une suite de variables intégrables, positive de moyenne égale à 1. Alors le processus $(M_n)_n$ définit par $M_n = \prod_{k=0}^n X_k$ est une martingale (pour la filtration $\mathcal{F}_n = \sigma(X_0, \cdots, X_n)$)

**Exemple 3 (Accumulation d'informations sur une variable aléatoire au cours du temps)** Soit $Y$ une variable aléatoire $L^1$. Alors $(\mathbb{E}(Y| \mathcal{F}_n), \mathcal{F}_n)_{n\in\mathbb{N}}$ est une martingale.

**Exemple 4 (Processus de branchement)**

Soit $(\xi_{i,j})_{(i,j)\in\mathbb{N}^2}$ des variables i.i.d à valeurs dans $\mathbb{N}$ de moyenne $m$. On pose $\mathcal{F}_0 = \{\emptyset, \Omega\}$ et $\mathcal{F}_n = \sigma(\xi_{i,j}\ \forall i \leqslant n-1)$.

On définit par récurrence le processus $(X_n)_n$ par $X_{n+1} = \displaystyle\sum_{j=1}^{X_n} \xi_{n,j}$. Le processus $Z_n = \dfrac{X_n}{m^n}$ est une martingale.

**Exemple 5 (Statistique de test pour le Sequential Probability Ratio Test)**

**Proposition** Si $(X_n)_n$ est une martingale alors $\mathbb{E}(X_n \mid \mathcal{F}_m) = X_m\ \forall m \leqslant n$

**Définition** Un processus $(A_n)_n$ est $(\mathcal{F}_n)_n$ prévisible si $\forall n \in \mathbb{N}^*$ $A_n$ est $\mathcal{F}_{n-1}$ mesurable.

**Proposition** Si $X$ est une martingale et $A$ un processus prévisible bornée presque sûrement (i.e pour tout $n$, $A_n$ est borné presque sûrement) alors $(A \cdot X)$ est une martingale. Si $A \geqslant 0$ presque sûrement et que $X$ est une sur/sous martingale alors $(A \cdot X)$ est une sur/sous-martingale.

**Corollaire**

Si $(X_n)_n$ est une martingale alors pour tout temps d'arrêt $\tau$, $(X_{n\wedge\tau})_n$ est une martingale.

**Interprétations des martingales en terme de jeu de hasard**

$(i)$ $X_n - X_{n-1}$

$(ii)$ $C_n$ (mise au temps $n$)

$(iii)$ $Y_n = (C \cdot X)_n = \sum_{k=1}^n C_k(X_k - X_{k-1})$ (gain total à l'instant $n$)

**Stratégie**

$C_1 = \mathbb{1}_{\{X_1<a\}}$ et $C_n = \mathbb{1}_{\{C_{n-1}=1\}}\mathbb{1}_{\{X_1\leqslant b\}} + \mathbb{1}_{\{C_{n-1}=0\}}\mathbb{1}_{\{X_1<a\}}$

**Doob's Upcrossing Lemma**

Soit $a < b$.

Si $(X_n)$ est une sous-martingale alors $(b-a)\,\mathbb{E}(U_n^X(a,b)) \leqslant \mathbb{E}((X_n-a)_+) - \mathbb{E}((X_0-a)_+)$

Si $(X_n)$ est une sur-martingale alors $(b-a)\,\mathbb{E}(U_n^X(a,b)) \leqslant \mathbb{E}((X_n-a)_-)$

**Théorème d'arrêt**

Soit $\underline{\nu} \leqslant \overline{\nu}$ deux temps d'arrêt bornés presque sûrement. Alors $\mathbb{E}(X_{\overline{\nu}} \mid \mathcal{F}_{\underline{\nu}}) = X_{\underline{\nu}}$

**Corollaire** $(X_n)_n$ est une martingale si et seulement si pour tout temps d'arrêt $L^\infty$ $\tau$ on a $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$

**Identité de Wald** Soit $\sigma$ un temps d'arrêt $L^1$ et $S_n = \sum_{k=1}^n X_k$ une marche aléatoire. Alors $\mathbb{E}(S_\sigma) = \mathbb{E}(\sigma)\mathbb{E}(X_1)$

**Inégalités maximales**

On pose $X_n^* = \max(X_0, X_1, \ldots, X_n)$.

$(i)$ Si $X$ est une sous-martingale, alors pour tout $a \geqslant 0$ et pour tout $n \in \mathbb{N}$ on a $a\mathbb{P}(X_n^* \geqslant a) \leqslant \mathbb{E}(X_n \mathbb{1}_{\{X_n^* \geqslant a\}})$

$(ii)$ Si $X$ est une sous-martingale positive, alors pour tout $a \geqslant 0$ et $n \in \mathbb{N}$ : $a\mathbb{P}(X_n^* \geqslant a) \leqslant \mathbb{E}(X_n)$.

$(iii)$ Si $X$ est une sur-martingale positive alors pour tout $a \geqslant 0$ on a : $a\mathbb{P}(\sup X_n^* \geqslant a) \leqslant \mathbb{E}(X_0)$

$(iv)$ (Maximal Doob inequality) Si $|X|$ est une sous-martingale alors pour tout $n \in \mathbb{N}$ et $p > 1$ : $\| |X|_n^* \|_p \leqslant \dfrac{p}{p-1} \|X_n\|_p$

**Théorème de décomposition de Doob**

Tout processus $L^1$ $(X_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ se décompose comme la somme d'une martingale et d'un processus prévisible : $X_n = M_n + A_n$ et $A_n = A_{n-1} + \mathbb{E}(\Delta X_n |\, \mathcal{F}_{n-1})$ i.e $\Delta A_n = \mathbb{E}(\Delta X_n | \mathcal{F}_{n-1})$

## Théorèmes de convergence

## Martingale bornée dans $L^1$

Soit $(X_n)_n$ une suite de v.a $L^1$. On dit que $(X_n)_n$ est bornée dans $L^1$ si $\sup_{n\in\mathbb{N}} \mathbb{E}(\|X_n\|) < +\infty$.

Une martingale/sur-martingale ou sous-martingale bornée dans $L^1$ converge dans $L^1$ et presque sûrement et la limite est la même.

## Martingale bornée dans $L^2$

Soit $(X_n)_n$ une suite de v.a $L^2$. On dit que $(X_n)_n$ est bornée dans $L^2$ si $\sup_{n\in\mathbb{N}} \mathbb{E}(\|X_n\|^2) < +\infty$.

Une martingale/sur-martingale ou sous-martingale bornée dans $L^2$ converge dans $L^2, L^1$, presque sûrement et la limite est la même.

### Uniforme Intégrabilité (U.I)

**Définition** Une famille $(X_t)_{t\in T}$ de variables aléatoires est dite uniformément intégrable si

$$\lim_{c\to+\infty}\sup_{t\in T}\ \mathbb{E}(|X_t|\mathbb{1}_{\{|X_t|\geq c\}})=0$$

Une famille de v.a uniformément intégrable est bornée dans $L^1$.

#### Théorème

#### Loi 0-1 de Kolmogorov

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes. On pose $\mathcal{T}_\infty := \bigcap\limits_{n\in\mathbb{N}} \sigma(X_{n+k}, k\geq 0)$ la tribu asymptotique. Pour tout $A\in\mathcal{T}_\infty$ on a $\mathbb{P}(A)=0$ ou $\mathbb{P}(A)=1$.

### Lois des grands nombres

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires $L^1$ i.i.d. Alors $\displaystyle\frac{1}{n}\sum_{k=1}^{n}X_k \longrightarrow \mathbb{E}(X_1)$ $\mathbb{P}$ presque partout.

## 12 Convergence des variables aléatoires

### Lemme de Borel-Cantelli

Soit $(A_n)_n\in\mathcal{F}^{\mathbb{N}}$. Si $\sum\mathbb{P}(A_n)<+\infty$ alors $\mathbb{P}(\limsup A_n)=0$

Soit $(A_n)_n\in\mathcal{F}^{\mathbb{N}}$ des évènements **INDÉPENDANTS**. Si $\sum\mathbb{P}(A_n)=+\infty$ alors $\mathbb{P}(\limsup A_n)=1$ .

### Convergence en probabilité

#### Définition

Une suite $(X_n)_n$ de variables aléatoires converge en probabilité vers $X$ si $\forall\varepsilon>0$ la suite $\mathbb{P}(\|X_n-X\|>\varepsilon)$ converge vers 0.

#### Lemme

Si $\forall\varepsilon>0,\ \sum\mathbb{P}(\|X_n-X\|>\varepsilon)<+\infty$ alors $X_n$ converge presque sûrement vers $X$.

**Proposition**

Une suite $(X_n)$ converge en probabilité vers $X$ si et seulement si il existe une extractrice $(n_k)_k$ telle que $X_{n_k}$ converge vers $X$ presque sûrement.

**Proposition**

Soit $(X_n)_n, X : \Omega \to \mathcal{X}$ des variables aléatoires et $f : \mathcal{X} \to \mathcal{Y}$ une fonction continue.

$(i)$ Si $X_n \longrightarrow X$ p.s alors $f(X_n) \longrightarrow f(X)$ presque sûrement.

$(ii)$ Si $X_n \longrightarrow X$ en probabilité alors $f(X_n) \longrightarrow f(X)$ en probabilité.

## Convergence en loi

**Définition**

Une suite de mesure $(\mu_n)_n$ sur $\mathbb{R}^d$ converge étroitement vers une mesure $\mu$ si pour toute fonction $f : \mathbb{R}^d \to \mathbb{R}$ continue et bornée on a $\mu_n(f) \longrightarrow \mu(f)$.

Une suite de variables aléatoires $(X_n)_n$ converge en loi vers une variable $X$ si $P_{X_n}$ converge étroitement vers $P_X$.

**Théorème**

La convergence en probabilité implique la convergence en loi.

**Tension**

**Définition** Soit $(\mu_n)_n$ une famille de mesures de probabilité. On dit que cette suite est tendue si pour tout $\varepsilon > 0$ il existe un compact $K$ tel que pour tout entier $n$ $\mu_n(K) \geq 1 - \varepsilon$. Cela signifie que toute la masse des mesures de probabilité est localisée dans ce compact à $\varepsilon$ près.

**Théorème de Portmanteau**

Les assertions suivantes sont équivalentes.

$(i)$ $\mu_n \longrightarrow \mu$. $(ii)$ Pour toute fonction $f$ continue bornée, $\mu_n(f) \longrightarrow \mu(f)$.
$(iii)$ Pour toute fonction $f$ lipschitzienne, continue, bornée $\mu_n(f) \longrightarrow \mu(f)$.
$(iv)$ Pour tout $A \in \mathcal{B}(\mathcal{X})$ tel que $\mu(\partial A) = 0, \mu_n(A) \longrightarrow \mu(A)$.

**Théorème de Portmanteau pour les variables aléatoires**

$(i)$ $X_n \xrightarrow{\mathcal{L}} X$.

$(ii)$ Pour toute fonction $f$ continue bornée, $\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X))$.

$(iii)$ Pour toute fonction $f$ lipschitzienne, continue, bornée $\mathbb{E}(f(X_n)) \longrightarrow \mathbb{E}(f(X))$.

$(iv)$ Pour tout $A \in \mathcal{B}(\mathcal{X})$ tel que $\mathbb{P}(X \in \partial A) = 0, \mathbb{P}(X_n \in A) \longrightarrow \mathbb{P}(X \in A)$.

$(v)$ Pour tout $x$, point de continuité de $F_X$ on a $F_{X_n}(x) \longrightarrow F_X(x)$.

**Théorème de Prokhorov**

Soit $(\mu_n)_n$ une suite tendue de mesures de probabilités. Il existe une extraction de cette suite qui converge étroitement vers une mesure $\mu$.

**Théorème de représentation de Skorokhod**

Soient $(X_n)_n$ des variables aléatoires convergeant en loi vers $X$. Il existe un espace $(E, \mathcal{E}, \mathbb{P})$ et des variables $(Y_n)_n, Y : E \to \mathcal{X}$ telles que $Y_n$ et $X_n$ ont même loi, $Y$ et $X$ ont même loi et $Y_n$ converge presque sûrement vers $Y$.

**Théorème de Lévy**

Soit $(\mu_n)_n$ une suite de mesures sur $\mathbb{R}^d$. Alors $(\mu_n)_n$ converge étroitement vers $\mu$ si et seulement si pour tout $t \in \mathbb{R}^d$, $\phi_{\mu_n}(t) \longrightarrow \phi_\mu(t)$.

# Références


[1] Probabilités avancées Cours de Master Avancé 1, ENS Lyon de Christophe Garban

[2] L'essentiel en théorie des probabilités de Jean Jacod et Philip Protter

[3] Probability with martingales by David Williams

[4] Polycopié du cours de Probabilités de M1 de François Roueff, Télécom Paris