# MACS203 : Martingales

## 1 Théorie de la mesure

les théorèmes de convergence monotone, dominée, et le lemme de Fatou, - les inégalités de Markov, Chebysev, Cauchy-Schwarz, Hölder, Minkowsky, et de Jensen,

### Espaces mesurables et mesures

**Def.** Soit $\mathcal{A} \subset \mathcal{P}(\Omega)$. On dit que

(i) $\mathcal{A}$ est une **algèbre** sur $\Omega$ si $\Omega \in \mathcal{A}$ et est stable par passage au complémentaire et par réunion,

(ii) $\mathcal{A}$ est une $\sigma$-algèbre si c'est une algèbre stable par union dénombrable. On dit alors que $(\Omega, \mathcal{A})$ est un **espace mesurable**.

**Def.** Soit $\mathcal{I} \subset \mathcal{P}(\Omega)$. On dit que $\mathcal{I}$ est un $\pi$**-système** s'il est stable par intersection finie.

**Def.** Soit $\mathcal{A}$ une algèbre sur $\Omega$ et $\mu\colon \mathcal{A} \to \mathbf{R}_+$.

(i) $\mu$ est dite **additive** si $\mu(\varnothing) = 0$ et $\forall A, B \in \mathcal{A}, A \cap B = \varnothing \implies \mu(A \cup B) = \mu(A) + \mu(B)$.

(ii) $\mu$ est dite $\sigma$**-additive** si $\mu(\varnothing) = 0$ et $\forall (A_n)_{n\in\mathbf{N}} \subset \mathcal{A}$, si les $A_n$ sont disjoints, $\mu(\cup_n A_n) = \sum_n \mu(A_n)$.

(iii) Une fonction $\sigma$-additive $\mu$ sur un espace mesurable $(\Omega, \mathcal{A})$ est appelée **mesure** et on dit que $(\Omega, \mathcal{A}, \mu)$ est un **espace mesuré**.

(iv) Un espace mesuré $(\Omega, \mathcal{A}, \mu)$ est dit **fini** si $\mu(\Omega) < \infty$, et $\sigma$**-fini** s'il existe $(\Omega_n)_{n\in\mathbf{N}} \subset \mathcal{A}$ telle que $\mu(\Omega_n) < \infty$ et $\bigcup_{n\in\mathbf{N}} \Omega_n = \Omega$.

**Prop.** *Soit $\mathcal{I}$ un $\pi$-système, et $\mu, \nu$ deux mesures finies sur $(\Omega, \sigma(\mathcal{I}))$. Si $\mu = \nu$ sur $\mathcal{I}$ alors $\mu = \nu$ sur $\sigma(\mathcal{I})$.*

**Th** (Extension de Carathéodory). *Soit $\mathcal{A}_0$ une algèbre sur $\Omega$ et $\mu\colon \mathcal{A}_0 \to \mathbf{R}_+$ $\sigma$-additive. Alors il existe une mesure $\mu$ sur $\mathcal{A} := \sigma(\mathcal{A}_,)$ telle que $\mu = \mu_0$ sur $\mathcal{A}_0$. Si de plus $\mu_0(\Omega) < \infty$ alors une telle extension est unique.*

**Def.** (i) Sur un espace mesuré $(\Omega, \mathcal{A}, \mu)$, $N \in \mathcal{A}$ est dit **négligeable** si $\mu(N) = 0$.

(ii) Soit $P(\omega)$ une propriété qui ne dépend que de $\omega \in \Omega$. On dit que $P$ est vraie $\mu$-presque partout si $\{w \in \Omega \mid P(\omega)$ n'est pas vraie$\}$ est inclus dans un ensemble négligeable.

**Prop.** *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(A_i)_{i\leqslant n} \subset \mathcal{A}$. Alors,*

*(i) $\mu(\cup_{i\leqslant n} A_i) \leqslant \sum_{i=1}^n \mu(A_i)$,*

*(ii) Si de plus $\mu(\Omega) < \infty$, on a $\mu(\cup_{i\leqslant n} A_i) = \sum_{k\leqslant n} (-1)^{k-1} \sum_{i_1 < \ldots < i_k \leqslant n} \mu(A_{i_1} \cap \cdots \cap A_{i_k})$.*

**Prop.** *Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $(A_i)_n \subset \mathcal{A}$. Alors,*

*(i) $A_n \uparrow A \implies \mu(A_n) \uparrow \mu(A)$,*

*(ii) $(A_n \downarrow A \wedge (\exists k, \mu(A_k) < \infty)) \implies \mu(A_n) \downarrow \mu(A)$.*

**Lem** (de Fatou pour les ensembles). *Soit $(A_n)_n$ une suite dans $\mathcal{A}$. Alors, $\mu(\liminf A_n) \leqslant \liminf \mu(A_n)$.*

**Lem** (inverse Fatou pour les ensembles). *Supposons $(\Omega, \mathcal{A}, \mu)$ fini. Soit $(A_n)_n$ une suite dans $\mathcal{A}$. Alors, $\mu(\limsup A_n) \geqslant \limsup \mu(A_n)$.*

**Lem** (de Borel-Cantelli). $\sum_n \mu(A_n) < \infty \implies \mu(\limsup A_n) = 0$.

### L'intégrale de Lebesgue

**Def.** On dit qu'une fonction $f\colon (\Omega, \mathcal{A}) \to (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ est **mesurable** si l'image réciproque de tout ensemble borélien est dans $\mathcal{A}$. On note $\mathcal{L}^0(\mathcal{A})$ l'ensemble des fonctions mesurables, $\mathcal{L}^0_+(\mathcal{A})$ si elles sont positives et $\mathcal{L}^\infty(\mathcal{A})$ si elles sont bornées.

*Rem.* Si $f\colon \Omega \to \mathbf{R}$ est continue avec $\Omega$ un espace topologique, alors $f$ est $\mathcal{B}(\Omega)$-mesurable et on dit qu'elle est **borélienne**.

**Prop.** *(i) Pour $f, g \in \mathcal{L}^0(\mathcal{A})$, $h \in \mathcal{L}^0(\mathcal{B}(\mathbf{R}))$, $\lambda \in \mathbf{R}$, on a $f + g, \lambda f, fg, f \circ g \in \mathcal{L}_0(\mathcal{A})$.*

*(ii) Pour une suite $(f_n)_n \subset \mathcal{L}_0(\mathcal{A})$, on a $\inf f_n, \liminf f_n, \sup f_n, \limsup f_n \in \mathcal{L}^0(\mathcal{A})$.*

**Th (des classes monotones).** *Soit $\mathcal{H}$ une classe de fonctions réelles bornées sur $\Omega$ vérifiant*

**H1** *$\mathcal{H}$ est un espace vectoriel contenant la fonction constante $\mathbf{1}$,*

**H2** *pour toute suite croissante $(f_n)_n \subset \mathcal{H}$ de fonctions positives dont la limite $f := \lim \uparrow f_n$ est bornée, on a $f \in \mathcal{H}$.*

*Soit $\mathcal{I}$ un $\pi$-système tel que $\{\mathbf{1}_A, A \in \mathcal{I}\} \subset \mathcal{H}$. Alors $\mathcal{L}^\infty(\sigma(\mathcal{I})) \subset \mathcal{H}$.*

*Not.* L'intégrale $\int f \, \mathrm{d}\mu$ sera aussi notée $\mu(f)$ par abus de notation.

**Def.** Pour $f \in \mathcal{L}^0_+(\mathcal{A})$, l'intégrale de $f$ par rapport à $\mu$ est définie par $\mu(f) := \sup\{\mu(g) \mid g \in \mathcal{S}^+, g \leqslant f\}$ où $\mathcal{S}^+$ contient les fonctions de la forme $g = \sum_i a_i \mathbf{1}_{A_i}, a_i \in \bar{\mathbf{R}}_+$ et $\mu(g) = \sum_i a_i \mu(A_i)$.

**Lem.** $\forall f_1, f_2 \in \mathcal{L}_0^+(\mathcal{A}, f_1 \leqslant f_2 \implies 0 \leqslant \mu(f_1) \leqslant \mu(f_2)$ *et* $\mu(f_1) = 0 \iff f_1 \overset{\mu\text{-p.p.}}{=} 0$.

**Th (convergence monotone).** *Soit* $(f_n)_n \subset \mathcal{L}_+^0(\mathcal{A})$ *une suite croissante* $\mu$*-p.p., i.e.* $\forall n, f_n \overset{\mu\text{-p.p.}}{\leqslant} f_{n+1}$. *Alors* $\mu(\lim \uparrow f_n) = \lim \uparrow \mu(f_n)$.

**Lem (Fatou).** *Soit* $(f_n)_n \subset \mathcal{L}_+^0(\mathcal{A})$. *On a* $\mu(\liminf f_n) \leqslant \liminf \mu(f_n)$.

**Def.** $f \in \mathcal{L}_0(\mathcal{A})$ est dite $\mu$**-intégrable** si $\mu(|f|) < \infty$ et son intégrale est définie par $\mu(f) := \mu(f^+) - \mu(f^-)$. On note $\mathcal{L}^1(\mathcal{A}, \mu)$ leur ensemble.

**Th (convergence dominée).** *Soit* $(f_n)_n \subset \mathcal{L}_0(\mathcal{A})$ *une suite telle que* $f_n \overset{\mu\text{-p.p.}}{\longrightarrow} f \in \mathcal{L}_0(\mathcal{A})$. *Si* $\sup_n |f_n| \in \mathcal{L}^1(\mathcal{A}, \mu)$, *alors* $f_n \to f$ *dans* $\mathcal{L}^1(\mathcal{A}, \mu$, *i.e.* $\mu(|f_n - f|) \to 0$. *En particulier,* $\mu(f_n) \to \mu(f)$.

**Lem (Scheffé).** *Soit* $(f_n)_n \subset \mathcal{L}^1(\mathcal{A}, \mu)$ *telle que* $f_n \overset{\mu\text{-p.p.}}{\longrightarrow} f \in \mathcal{L}_1(\mathcal{A}, \mu)$. *Alors* $f_n \to f$ *dans* $\mathcal{L}_1(\mathcal{A}, \mu)$ *ssi* $\mu(|f_n|) \to \mu(|f|)$.

### Transformées de mesures

**Def.** Soit $(\Omega_1, \mathcal{A}_1, \mu_1)$ un espace mesuré, $(\Omega_2, \mathcal{A}_2)$ un espace mesurable et $f \colon \Omega_1 \to \Omega_2$ une fonction mesurable. Alors $\mu_2 = \mu_1 \circ f^{-1}$, notée $\mu_1 f^{-1}$, définit une mesure appelée **mesure image**.

**Th** (transfert). *Soit* $\mu_2 = \mu_1 f^{-1}$ *et* $h \in \mathcal{L}^0(\mathcal{A}_2)$. *Alors* $h \in \mathcal{L}^1(\mathcal{A}_2, \mu_2) \iff h \circ f \in \mathcal{L}^1(\mathcal{A}_2, \mu_2)$. *Dans ces conditions on a* $\int_{\Omega_2} h \, \mathrm{d}(\mu_1 f^{-1}) = \int_{\Omega_1} h \circ f \, \mathrm{d}\mu_1$.

**Def.** Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $f \in \mathcal{L}_+^0(\mathcal{A})$. On définit $\forall A \in \mathcal{A}, \nu(A) := \mu(f \mathbf{1}_A) = \int_A f \, \mathrm{d}\mu$.

(i) $\nu = f \cdot \nu$ est une mesure appelée mesure de **densité** $f$ par rapport à $\mu$.

(ii) Soit $\mu_1, \mu_2$ deux mesures sur un espace mesurable $(\Omega, \mathcal{A})$. On dit que $\mu_2$ est **absolument continue** par rapport à $\mu_1$, $\mu_2 \prec \mu_1$, si $\forall A \in \mathcal{A}, \mu_2(A) = 0 \implies \mu_1(A) = 0$. Sinon on dit que $\mu_2$ est étrangère à $\mu_1$.

(iii) Si $\mu_2 \prec \mu_1$ et $\mu_1 \prec \mu_2$, on dit que $\mu_1$ et $\mu_2$ sont **équivalentes**, $\mu_1 \sim \mu_2$. Si $\mu_2 \not\prec \mu_1$ et $\mu_1 \not\prec \mu_2$, on dit que $\mu_1$ et $\mu_2$ sont **singulières**.

**Th.** *(i) Pour* $g \colon \Omega \to \bar{\mathbf{R}}_+$ $\mathcal{A}$ *mesurable, on a* $(f \cdot \mu)(g) = \mu(fg)$.

*(ii) Pour* $g \in \mathcal{L}_+^0(\mathcal{A})$, *on a* $g \in \mathcal{L}^1(\mathcal{A}, f \cdot \mu)$ *ssi* $fg \in \mathcal{L}^1(\mathcal{A}, \mu)$ *et alors* $(f \cdot \mu)(g) = \mu(fg)$.

### Inégalités remarquables

**Th.** *Soit* $f$ *une fonction* $\mathcal{A}$*-mesurable, et* $g \colon \mathbf{R} \to \mathbf{R}_+$ *une fonction borélienne croissante positive.*

*(i)* $g \circ f$ *est mesurable et* $\forall c \in \mathbf{R}, \mu(g \circ f) \geqslant g(c)\mu(\{f \geqslant c\})$ ***(Inégalité de Markov)****,*

*(ii) Si* $f^2 \in \mathcal{L}_1(\mathcal{A}, \mu)$, $\forall c > 0, c^2\mu(\{|f| \geqslant c\}) \leqslant \mu(f^2)$ ***(inégalité de Tchebyshev)****.*

### Espaces produits

**Th (Fubini).** *L'application* $\mu \colon A \mapsto \int(\int \mathbf{1}_A \, \mathrm{d}\mu_1) \, \mathrm{d}\mu_2 = \int(\int \mathbf{1}_1 \, \mathrm{d}\mu_2) \, \mathrm{d}\mu_1$ *sur* $\mathcal{A}_1 \otimes \mathcal{A}_2$ *est une mesure sur* $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, *appelée* ***mesure produit*** *de* $\mu_1$ *et* $\mu_2$, *et notée* $\mu_1 \otimes \mu_2$. *C'est l'unique mesure sur* $\Omega_1 \times \Omega_2$ *vérifiant* $\forall (A_1, A_2) \in \mathcal{A}_1 \times \mathcal{A}_2, \mu(A_1 \times A_2) = \mu_1(A_1)\mu(A_2)$.

*De plus, pour tout* $f$ *dans* $\mathcal{L}_+^0(\mathcal{A}_1 \times \mathcal{A}_2)$ *ou* $\mathcal{L}^1(\mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$, $\int f \, \mathrm{d}\mu = \int(\int f \, \mathrm{d}\mu_1) \, \mathrm{d}\mu_2 = \int(\int f \, \mathrm{d}\mu_2) \, \mathrm{d}\mu_1 \in \bar{\mathbf{R}}_+$.

Soit $g \colon \Omega_1 \to \Omega_2$ avec $\Omega_1$ et $\Omega_2$ des ouverts de $\mathbf{R}^n$. Si $g$ est différentiable en $x$, on note $Dg(x) := \left(\frac{\partial g_i}{\partial x_j}\right)_{1 \leqslant i,j \leqslant n}$ sa matrice jacobienne en $x$.

$g$ est un $\mathcal{C}^1$-difféomorphisme si $g$ est une bijection telle que $g$ et $g^{-1}$ sont de classe $\mathcal{C}^1$. Dans ce cas $\det[Dg^{-1}(y)] = \frac{1}{\det[Dg \circ g^{-1}(y)]}$.

**Th.** *Soit* $\mu_1$ *une mesure sur* $(\Omega_1, \mathcal{B}(\Omega_1))$ *de densité par rapport à la mesure de Lebesgue* $f_1 \in \mathcal{L}_0^+(\mathcal{B}(\Omega_1))$, *i.e.* $\mu_1(\mathrm{d}x) = \mathbf{1}_{\Omega_1} f_1(x) \cdot \mathrm{d}x$. *Soit* $g$ *un* $\mathcal{C}^1$*-difféomorphisme. La mesure image* $\mu_2 = \mu_1 g^{-1}$ *est absolument continue par rapport à la mesure de Lebesgue, de densité* $f_2 \colon y \mapsto \mathbf{1}_{\Omega_2}(y) f_1(g^{-1}) \left|\det[Dg^{-1}(y)]\right|$ *et pour toute fonction* $h \colon \Omega_2 \to \mathbf{R}$ *positive ou* $\mu_2$*-intégrable,* $\int_{\Omega_1} h \circ g(x) f_1(x) \, \mathrm{d}x = \int_{\Omega_2} h(y) f_2(y) \, \mathrm{d}y$.

## 2 Préliminaires de la théorie des probabilités

### Variables aléatoires

**Def.** Soit $\mathbf{T}$ un ensemble et $\{X_\tau, \tau \in \mathbf{T}\}$ une famille quelconque de v.a. La $\sigma$-algèbre $\mathcal{X}$ engendrée par cettefamille est la plus petite $\sigma$-algèbre sur $\Omega$ telle que $X_\tau$ est $\mathcal{X}$-mesurable pour tout $\tau \in \mathbf{T}$, i.e.

$$\mathcal{X} = \sigma(X_\tau, \tau \in \mathbf{T}) = \sigma(\{X_\tau^{-1}(A) \mid \tau \in \mathbf{T}, A \in \mathcal{B}(\mathbf{R})\}) \,.$$

**Lem.** *Soit* $X$ *et* $Y$ *deux v.a. sur* $(\Omega, \mathcal{A}, \mathbf{P})$ *à valeurs respectivement dans* $\mathbf{R}$ *et* $\mathbf{R}^n$. *Alors* $X$ *est* $\sigma(Y)$*-mesurable ssi* $\exists f \colon \mathbf{R}^n \to \mathbf{R}, X = f(Y)$.

**Espérance de variables aléatoires**

**Th.** *Soit $X \in \mathcal{L}^1(\mathcal{A}, \mathbf{P})$ et $g\colon \mathbf{R}^d \to \bar{\mathbf{R}}$ une fonction convexe telle que $\mathbf{E}(|g(X)|) < \infty$. Alors $\mathbf{E}(g(X)) \geqslant g(\mathbf{E}(X))$.*

**Def.** Soit $X$ une v.a. à valeurs dans $\mathbf{R}^d$. Sa fonction caractéristique est $\Phi_X\colon \begin{array}{ccc} \mathbf{R}^d & \to & \mathbf{C} \\ u & \mapsto & \mathbf{E}\left[e^{i\langle u|X\rangle}\right] \end{array}$.

**Lem.** *$\Phi_X(0) = 1$ et $\Phi_X$ est continue bornée (par 1) sur $\mathbf{R}^d$.*

**Prop.** *Soit $X \sim \mathcal{N}(b, V)$. On a $\Phi_X(u) = e^{\langle u|b\rangle - \frac{1}{2}\langle u|Vu\rangle}$.*

**Prop.** *Soit $X$ réelle avec $\mathbf{E}(|X|^p) < \infty$ pour un certain $p \in \mathbf{N}^*$. Alors $\Phi_X$ est $p$ fois dérivable et $\forall k \in [\![1\,;p]\!], \Phi_X^{(k)}(0) = i^k \mathbf{E}(X^k)$.*

**Espaces $\mathcal{L}^p$ et convergences fonctionnelles des v.a.**

La corrélation entre deux v.a. $X$ et $Y$ est $\mathrm{Cor}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\|X\|_2\|Y\|_2}$.
Le théorème de Pythagore s'écrit

$$\mathbf{E}(XY) = 0 \implies \mathbf{E}[(X+Y)^2] = \mathbf{E}\left[X^2\right] + \mathbf{E}\left[Y^2\right] \quad \text{ou} \quad \mathrm{Cov}(X,Y) = 0 \implies \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)\,.$$

et la loi du parallélogramme s'écrit $\|X+Y\|_2^2 + \|X-Y\|_2^2 = 2\,\|X\|_2^2 + 2\,\|Y\|_2^2$.

**Def.** Soit $(X_n)_n$ et $X$ dans $\mathcal{L}^0$. On dit que $(X_n)_n$ **converge en probabilité** vers $X$ si $\forall \varepsilon > 0, \lim_{n\to\infty} \mathbf{P}[|X_n - X| \geqslant 0] = 0$.

**Lem.** *La convergence p.s. ou la convergence en norme dans $L^p$ impliquent la convergence en probabilité.*

**Lem.** *La convergence en probabilité est équivalente à la convergence au sens de la distance $D\colon (X,Y) \mapsto \mathbf{E}(|X-Y| \wedge 1)$*

**Th.** *$(L^0, D)$ est un espace métrique complet.*

**Th.** *Soit $(X_n)_n$ et $X$ des v.a. dans $\mathcal{L}^0$.*

*(i) $X_n \longrightarrow X$ p.s. ssi $\sup_{m \geqslant n} |X_m - X| \longrightarrow 0$ en probabilité.*

*(ii) $X_n \longrightarrow X$ en probabilité ssi de toute suite croissantes $(n_k)_k \subset \mathbf{N}$, on peut extraire une sous-suite $(n_{k_j})_j$ telle que $X_{n_{k_j}} \longrightarrow X$ p.s.*

**Cor** (Slutsky). *Soit $\phi$ continue. Si $X_n \longrightarrow X$ en probabilités, alors $\phi(X_n) \longrightarrow \phi(X)$ en probabilité.*

**Def.** Une famille $C$ de v.a. est dite **uniformément intégrable** (U.I.) si $\lim_{c\to\infty} \sup_{X\in C} \mathbf{E}\left[|X|\, \mathbf{1}_{|X|\geqslant c}\right] = 0$.

**Th.** *Soit $(X_n)_n$ et $X$ des v.a. dans $\mathcal{L}^1$. Alors $X_n \longrightarrow X$ dans $L^1$ si et seulement si $X_n \longrightarrow X$ en probabilité et $(X_n)_n$ est U.I.*

**Convergence en loi**

## 3 Espérance conditionnelle

Soit un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et $\mathcal{F}, \mathcal{G}$ des sous-$\sigma$-algèbres de $\mathcal{A}$.

**Th.** *Pour toute v.a. $X \in Ł^1(\mathcal{A}, \mathbf{P})$, il existe une v.a. $Z$ telle que*

*(i) $Z$ est $\mathcal{F}$-mesurable,*

*(ii) $\mathbf{E}|Z| < \infty$,*

*(iii) pour tout événement $F \in \mathcal{F}$, on a $\mathbf{E}(X\,\mathbf{1}_F) = \mathbf{E}(Z\,\mathbf{1}_F)$.*

*De plus $Z$ est unique p.s.*

**Def.** La v.a. vérifiant les propriétés ce-dessus est appeleé **version de l'espérance conditionnelle** de $X$ sachant $\mathcal{F}$, notée $\mathbf{E}(X \mid \mathcal{F})$. Si $\mathcal{F} = \sigma(Y_1, \ldots, Y_n)$, on écrit simplement $\mathbf{E}(X \mid Y_1, \ldots, Y_n)$.

*Ex.* On a $\mathbf{E}(X \mid \{\varnothing, \Omega\}) = \mathbf{E}(X)$ et $\mathbf{E}(X \mid \sigma(X)) = X$.

**Prop.** *L'opérateur $\mathbf{E}(\cdot \mid \mathcal{F})$ est linéaire et $\forall X \in Ł^1(\mathcal{A}, \mathbf{P})$ on a :*

*(i) $\mathbf{E}(\mathbf{E}(X \mid \mathcal{F})) = \mathbf{E}(X)$*

*(ii) si $X$ est $\mathcal{F}$-mesurable, $\mathbf{E}(X \mid \mathcal{F}) \overset{p.s.}{=} X$,*

*(iii) si $X \geqslant 0$, $\mathbf{E}(X \mid \mathcal{F}) \overset{p.s.}{\geqslant} 0$,*

*(iv) si $g\colon \mathbf{R}^n \to \mathbf{R}$ est convexe et $\mathbf{E}(|g(X)|) < \infty$ alors $\mathbf{E}(g(X) \mid \mathcal{F}) \geqslant g(\mathbf{E}(X \mid \mathcal{F}))$,*

*(v) si $\mathcal{F} \subset \mathcal{G}$, $\mathbf{E}(\mathbf{E}(X \mid \mathcal{G}) \mid \mathcal{F}) = \mathbf{E}(X \mid \mathcal{F})$,*

*(vi) si $\mathcal{G}$ est indépendante de $\sigma(\sigma(X), \mathcal{F})$, $\mathbf{E}(X \mid \sigma(\mathcal{F}, \mathcal{G})) = \mathbf{E}(X \mid \mathcal{F})$.*

**Prop.** *La convergence monotone, le lemme de Fatou et la convergence dominée restent vraies pour l'espérance conditionnelle.*

**Prop.** *Soit $X \in Ł^0(\mathcal{A})$ et $Y \in Ł^0(\mathcal{F})$. On suppose $\mathbf{E}(|X|) < \infty$ et $\mathbf{E}(|XY|) < \infty$. Alors $\mathbf{E}(XY \mid \mathcal{F}) = Y \cdot \mathbf{E}(X \mid \mathcal{F})$.*

**Prop.** *Soit $X, Y$ deux v.a. à valeurs dans $\mathbf{R}^n$ et $\mathbf{R}^m$ respectivement, et $g\colon \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ une fonction telle que $\mathbf{E}(|g(X,Y)|) < \infty$. Si $X$ et $Y$ sont indépendantes alors $\mathbf{E}(g(X,Y) \mid X) = G(X)$ où $\forall x \in \mathbf{R}^n, G(x) := \mathbf{E}(g(x,Y))$.*

# 4 Vecteurs gaussiens

**Def.** $X$ est un **vecteur gaussien** (ou variable gaussienne multivariée ou variable normale multivariée) si et seulement si $\forall a \in \mathbf{R}^d$, la loi de $\langle a \mid X\rangle$ est une loi gaussienne (éventuellement de variance nulle).

**Th.** *$X$ est un vecteur gaussien d'espérance $m$ et de matrice de covariance $\Gamma$ si et seulement si sa fonction caractéristique est $t \mapsto \exp\left(i\langle t \mid m\rangle - \frac{1}{2}t^\intercal \Gamma t\right)$. On écrit $X \sim \mathcal{N}_d(m, \Gamma)$.*

**Prop.** *Soit $(X, Y)$ un vecteur gaussien à valeurs dans $\mathbf{R}^n \times \mathbf{R}^m$, de moyenne $\mu = \begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}$ et de matrice de variances-covariances $V = \begin{pmatrix} V_X & V_{XY}^\intercal \\ V_{XY} & V_Y\end{pmatrix}$. Supposons que $\mathrm{Var}(Y) = V_Y$ est inversible. Alors la loi conditionnelle de $X$ sachant $Y = y$ est gaussienne de moyenne $\mathbf{E}(X \mid Y = y) = \mu_X + V_{XY}V_Y^{-1}(y - \mu_Y)$ et variance $\mathrm{Var}(X \mid Y = y) = V_X - V_{XY}V_Y^{-1}V_{XY}^\intercal$.*

# 5 Processus aléatoires et structure d'information

**Def.** Un **processus** est une suite $(X_n)_n$ de v.a. sur $(\Omega, \mathcal{A})$ à valeurs dans un ensemble mesuré $(E, \mathcal{E})$.

**Def.** Une **filtration** de $\mathcal{A}$ est une suite croissante $\mathbf{F} = (\mathcal{F}_n)_{n\geqslant 0}$ de sous-$\sigma$-algèbres de $\mathcal{A}$. On dit que $(\Omega, \mathcal{A}, \mathbf{F})$ est un espace probabilisable filtré et $(\Omega, \mathcal{A}, \mathbf{F}, \mathbf{P})$ un espace probabilisé filtré.

*Ex.* La suite $(\mathcal{F}_n^X)_{n\in\mathbf{N}} = (\sigma(X_i, i \leqslant n))_{n\in\mathbf{N}}$ est une filtration de $\mathcal{A}$ appelée **filtration naturelle** de $X$.

**Def.** Soit $X = (X_n)_n$ un processus aléatoire et $(\mathcal{F}_n)_n$ une filtration de $\mathcal{A}$. On dit que $X$ est :
- **F-adapté** si $\forall n \in \mathbf{N}$, $X_n$ est $\mathcal{F}_n$-mesurable,
- **F-prévisible** si $\forall n \in \mathbf{N}$, $X_n$ est $\mathcal{F}_{n-1}$-mesurable, où $\mathcal{F}_{-1} := \{\varnothing, \Omega\}$.

**Def.** Un **temps d'arrêt** $\nu$ est une variable aléatoires à valeurs dans $\mathbf{N} \cup \{\infty\}$ telle que $\forall n \in \mathbf{N}, \{\nu = n\} \in \mathcal{F}_n$. On note $\mathcal{T}$ l'ensemble des temps d'arrêt.

**Prop.** *Soit $(X_n)_{n\in\mathbf{N}}$ un processus $\mathbf{F}$-adapté à valeurs dans $(E, \mathcal{E})$. Pour tout $A \in \mathcal{E}$, on définit le **premier temps d'atteinte** $T_A := \inf\{n \in \mathbf{N} \mid X_n \in A\}$, avec la convention $\inf \varnothing = \infty$. Alors $T_A$ est un temps d'arrêt.*

**Prop.** *Soit $\tau, \theta, (\tau_n)_{n\in\mathbf{N}}$ des temps d'arrêt.*

*(i)* *$\tau \wedge \theta$, $\tau \vee \theta$ et $\tau + \theta$ sont des temps d'arrêt,*

*(ii)* *soit $c \geqslant 0$ une constante, alors $\tau + c$ et $(1 + c)\tau$ sont des temps d'arrêt,*

*(iii)* *$\liminf_n \tau_n$ et $\limsup_n \tau_n$ sont des temps d'arrêt.*

**Prop.** *Soit $(X_n)_n$ un processus aléatoire à valeurs dans un espace mesuré $(E, \mathcal{E})$ et $\tau$ un temps d'arrêt. Alors $X_\tau : \omega \in \Omega \mapsto X_{\tau(\omega)}(\omega)$ est une v.a.*

**Prop.** *Pour tout temps d'arrêt $\tau \in \mathcal{T}$, $\mathcal{F}_\tau \subset \mathcal{A}$ est une sous-$\sigma$-algèbre de $\mathcal{A}$. Si $X$ est un processus aléatoire $\mathbf{F}$-adapté, $X_\tau$ est $\mathcal{F}_\tau$-mesurable.*

**Def.** L'**information disponible à un temps d'arrêt** est $\mathcal{F}_\tau := \{A \in \mathcal{A} \mid \forall n \in \mathbf{N}, A \cap \{\tau = n\} \in \mathcal{F}_n\}$.

**Prop.** *Pour tout temps d'arrêt $\tau \in \mathcal{T}$, $\mathcal{F}_\tau$ est une sous-$\sigma$-algèbre de $\mathcal{A}$. Si $X$ est un processus aléatoire $\mathbf{F}$-adapté, $X_\tau$ est $\mathcal{F}_\tau$-mesurable.*

**Prop.** *Soit $\tau$ et $\theta$ deux temps d'arrêt. Alors $\{\tau \leqslant \theta\}$, $\{\tau \geqslant \theta\}$ et $\{\tau = \theta\}$ appartiennent à $\mathcal{F}_\tau \cap \mathcal{F}_\theta$, et pour toute v.a. $X$ intégrable, on a $\mathbf{E}(\mathbf{E}(X \mid \mathcal{F}_\tau) \mid \mathcal{F}_\theta) = \mathbf{E}(\mathbf{E}(X \mid \mathcal{F}_\theta) \mid \mathcal{F}_\tau) = \mathbf{E}(X \mid \mathcal{F}_{\tau\wedge\theta})$.*

# 6 Chaînes de Markov

Soit $X = (X_n)_{n\in\mathbf{N}}$ un processus stochastique défini sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ et à valeurs dans un espace d'états discret $E$, fini ou dénombrable.

*Not.* $\pi_n$ est la distribution marginale de $X_n$ : $\forall x \in E, \pi_n(x) := \mathbf{P}(X_n = x)$.

**Def.** On dit que $X$ est un **chaîne de Markov** si $\forall A \subset E, \forall n \in \mathbf{N}^*, \mathbf{P}[X_n \in A \mid \mathcal{F}_{n-1}^X] = \mathbf{P}[X_n \in A \mid X_{n-1}]$.

Les **probabilités de transition** sont représentées par les **matrices de transition** $P_n$ définies par $\forall n \in \mathbf{N}, \forall x, y \in E, P_n(x, y) := \mathbf{P}[X_n = y \mid X_{n-1} = x]$. Ce sont des matrices stochastiques : leurs composantes sont positives et leurs lignes somment à l'unité.

**Prop.** *Soit $(P_n)_{n\in\mathbf{N}^*}$ une suite de matrices stochastiques sur $E$. Pour toute distribution initiale $\pi_0$ il existe une chaîne de Markov de loi initiale $\pi_0$ et de matrices de transition $(P_n)_{n\in\mathbf{N}^*}$.*

Les probabilités marginales $\pi_n$ se déduisent par $\forall n \in \mathbf{N}^*, \pi_n = \pi_0 P_1 \dots P_n$, où $\pi_0$ est un vecteur ligne de taille $\mathrm{Card}(E)$.

**Prop** (Formule de Chapman-Kolmogorov). $\forall x, y \in E, \forall k \in [0\,;n], \mathbf{P}(X_n = y \mid X_0 = x) = \sum_{z\in E} \mathbf{P}(X_n = y \mid X_k = z)\mathbf{P}(X_k = z \mid X_0 = x)$.

*Not.* $\mathbf{P}_x$ est la probabilité conditionelle sachant $X_0 = x$, et $\mathbf{E}_x$ est l'espérance associée.

**Th** (Propriété de Markov forte). *Soit* $(X_n)_{n\in\mathbf{N}}$ *une chaîne de Markov et* $\tau$ *un temps d'arrêt à valeurs dans* $\mathbf{N}$*. Alors* $\forall A\subset E, \forall n\in\mathbf{N}^*, \mathbf{P}(X_{\tau+n}\in A \mid \mathcal{F}^X_{\tau+n-1}) = \mathbf{P}(X_{\tau+n}\in A\mid X_{\tau+n-1})$.

**Def.** Une chaîne de Markov est dite **homogène** si sa matrice de transition $P_n$ est indépendante de $n$.

# 7 Lois invariantes et classification des états

Soit $X$ une chaîne de Markov homogène de matrice de transition $P$.

**Def.** Une probabilité $\nu$ sur $E$ est représentée par un vecteur ligne $(\nu(x))_{x\in E}$. On dit que $\nu$ est une probabilité invariante pour $X$ si $\nu P=\nu$.

**Th.** *Soit* $E$ *un espace d'état fini. Alors il existe au moins une probabilité invariante.*

Si $\forall x\in E, \forall n\in\mathbf{N}\pi_n(x)>0$ on définit $Q_n(x,y) := \mathbf{P}(X_n=y\mid X_{n+1}=x) = \frac{P(y,x)\pi_n(y)}{\pi_{n+1}(x)}$.

**Def.** On dit que $X$ (homogène) est **réversible** par rapport à une mesure de probabilité $\nu$ si $\forall x,y\in E, \nu(x)P(x,y) = \nu(y)P(y,x)$, i.e. si les lois marginales $\pi_n$ sont données par $\nu$, $Q_n=P$ pour tout $n$.

**Prop.** *Soit* $\nu$ *une mesure de probabilité par rapport à laquelle* $X$ *est invariant. Alors* $\nu$ *est une probabilité invariante.*

**Def.** Soit $x\in E$. On définit le temps d'arrêt de premier retour à $x$ : $R_x := R_1^x = \inf\{n\in\mathbf{N}^*\mid X_n=x\}$. $x$ est dit **récurrent** si $\mathbf{P}(R_x<\infty)=1$, dont **récurrent positif** si $\mathbf{E}_x(R_x)<\infty$ et **récurrent nul** si $\mathbf{E}_x(R_x)=\infty$. Sinon on dit que $x$ est **transitoire** ou **transient**.

On introduit les mesures à valeurs dans $[0\,;\infty]$ définies par $\forall x,y\in E, \mu_x(y) = \mathbf{E}_x\left[\sum_{n=0}^{R_x-1}\mathbf{1}_{\{X_n=y\}}\right] = \sum_{n\in\mathbf{N}}\mathbf{P}_x(R_x>n, X_n=y)$.

**Prop.** *Soit* $x\in E$*. Alors,*

*(i)* $\mu_xP=\mu_x$ *ssi* $x$ *est un état récurrent,*

*(ii)* $\mu_x$ *est une mesure finie ssi* $x$ *est récurrent positif, dans ce cas* $\nu_x = \frac{\mu_x}{\mathbf{E}_x(R_x)}$ *est une probabilité invariante.*

**Def.** Soit $x,y\in E$. On dit que :

- $x$ **communique** avec $y$, noté $x\leftarrow y$ si $\exists n\in\mathbf{N}, x_1,\dots,x_n, P(x,x_1)\cdots P(x_n,y)>0$,
- $x$ et $y$ communiquent, noté $x\leftrightarrow y$, si $x\leftarrow y$ et $y\leftarrow x$.

**Def.** Une classe $E_0\subset E$ est dite **irréductible** si $\forall x,y\in E_0, x\leftarrow y$. $X$ est dite irréductible si $E$ est irréductible. Une classe $E_0\subset E$ est dite **fermée** si $\forall x,y\in E, (x\in E_0\wedge x\leftarrow y) \implies y\in E_0$. Si $\{x_0\}$ est fermée, on dit que $x_0$ est absorbant.

On introduit le **nombre de visite d'un état** $x$ : $N^x := \sum_{n\in\mathbf{N}}\mathbf{1}_{\{X_n=x\}}$.

**Prop.** *Soit* $x,y\in E$.

*(i) Si* $x\leftarrow y$ *et* $x$ *est récurrent, alors* $y$ *est récurrent et* $N^y=\infty$, $\mathbf{P}_x$*-p.s.*

*(ii) Si* $x\leftrightarrow y$*, alors* $x$ *et* $y$ *sont simultanément soit transitoires soit récurrents.*

**Th.** *Supposons* $X$ *irréductible. Alors* $X$ *est récurrente positive si et seulement si* $X$ *admet une loi invariante* $\nu$*. De plus,* $\nu$ *est unique, strictement positive, donnée par* $\forall x\in E, \nu(x) = \frac{1}{\mathbf{E}_x(R_x)}$.

**Prop.** *Soit* $X$ *une chaîne de Markov sur un espace d'état dénombrable* $E$*, et* $x\in E$ *récurrent. Alors, pour toute mesure* $\nu$ *sur* $E$, $\nu\geqslant\nu P \implies \nu=\nu(x)\mu_x$.

# 8 Théorèmes ergodiques

### Théorèmes ergodiques

**Th.** *Soit* $X$ *une chaîne de Markov irréductible,* $\forall x,y\in E, \frac{1}{n}N_n^y := \frac{1}{n}\sum_{i=0}^n\mathbf{1}_{\{X_i=y\}} \longrightarrow \frac{1}{\mathbf{E}_y(R^y)}$, $\mathbf{P}_x$*-p.s.*

En particulier il vient $\forall x,y\in E, \frac{1}{n}\sum_{i=0}^n\mathbf{P}(X_i=x) = \frac{1}{n}\sum_{i=0}^n\pi_i(x) \longrightarrow \nu(x)$, $\mathbf{P}_y$-p.s. avec $\nu$ une loi invariante.

**Th.** *Soit* $X$ *une chaîne de Markov irréductible et récurrente positive sur* $E$ *dénombrable, de matrice de transition* $P$ *et d'unique loi invariante* $\nu$*. Alors, pour toute fonction* $g\colon E\times E\to\mathbf{R}$ *positive ou telle que* $\mathbf{E}_\nu\left[|g(X_0,X_1)|\right]<\infty$*, on a* $\forall\pi_0, \frac{1}{n}\sum_{i=1}^n g(X_{i-1},X_i) \xrightarrow{p.s.} \mathbf{E}_\nu\left[g(X_0,X_1)\right] = \sum_{x\in E}\nu(x)\sum_{y\in E}P(x,y)g(x,y)$.

**Th.** *Soit* $X$ *et* $g$ *comme précedemment. Supposons qu'il existe* $x\in E$ *tel que*

$$s(x)^2 := \mathbf{E}_x\left[\sum_{i=1}^{R_x}\left(g(X_{i-1},X_i)-\mathbf{E}_\nu(g(X_0,X_1))\right)^2\right]<\infty\,.$$

*Alors* $\sigma^2 := \nu(x)s(x)^2$ *est une constante (indépendante de* $x$*) et*

$$\sqrt{x}\left(\frac{1}{n}\sum_{i=1}^n g(X_{i-1},X_i)-\mathbf{E}_\nu(g(X_0,X_1))\right)\longrightarrow\mathcal{N}(0,\sigma^2) \text{ en loi.}$$

### Convergence des lois marginales et apériodicité

*Not.* Pour tout état $x \in E$ on définit $I(x) := \{n \in \mathbf{N}^* \mid P^n(x,x) > 0\}$ et $\mathbf{p}(x) := \operatorname{pgcd}(I(x))$.

**Prop.** *Soit $X$ une chaîne de Markov irréductible. Alors la fonction $\mathbf{p}(x) = \mathbf{p}_X$ est constante.*

**Def.** Soit $X$ une chaîne de Markov irréductible. On dit que $X$ est **apériodique** si $\mathbf{p}_X = 1$.

**Lem.** *Pour $x \in E$, $\mathbf{p}(x) = 1 \iff \exists \mathbf{n}(x) \in \mathbf{N}, \forall n \geqslant \mathbf{n}(x), P^n(x,x) > 0$.*

**Th.** *Soit $X$ une chaîne de Markov irréductible, apériodique et récurrente positive d'unique loi invariante $\nu$. Alors $\forall x \in E, \pi_n(x) \longrightarrow \nu(x)$.*

**Prop.** *Soit $X^1$ et $X^2$ deux chaînes de Markov indépendantes de même matrice de transition $P$ irréductible apériodique. Alors la chaîne produit $Y := (X^1, X^2)$ est irréductible apériodique. Si de plus $P$ est récurrente positive, il en est de même pour $Y$.*

**Prop.** *Soit $X$ une chaîne de Markov irréductible apériodique sur $E$ fini. Alors sa matrice de transition $P$ vérifie la **condition de Dobelin** : il existe $k \in \mathbf{N}, \epsilon > 0$ et une loi $\delta$ sur $E$ tels que $\forall x, y \in E, P^k(x,y) \geqslant \epsilon \cdot \delta(y)$.*

**Th.** *Soit $P$ une matrice de transition vérifiant la condition de Dobelin. Alors il existe une unique loi invariante $\nu \geqslant \epsilon \cdot \delta$ vérifiant*

$$\sup_{x \in E} \sum_{y \in E} |P^n(x,y) - \nu(y)| \leqslant 2(1-\epsilon)^{\lfloor n/k \rfloor}.$$

## 9 Martingales en temps discret

### Martingales et temps d'arrêt

**Def.** Soit $(X_n)_{n \geqslant 0}$ un processus aléatoire adapté sur l'espace probabilisé filtré $(\Omega, \mathcal{A}, \mathbf{F}, \mathbf{P})$. On dit que $X$ est une **surmartingale** (resp. **sous-martingale**) si $X_n$ est $\mathbf{P}$-intégrable pour tout $n$ et $\forall n \in \mathbf{N}^*, \mathbf{E}[X_n \mid \mathcal{F}_{n-1}] \leqslant$ (resp. $\geqslant$)$X_{n-1}$. $X$ est une **martingale** s'il est à la fois surmartingale et sous-martingale.

**Def.** Pour un processus aléatoire $X = (X_n)_{n \geqslant 0}$, on définit le **processus arrêté** au temps d'arrrêt $\nu$ par $\forall n \in \mathbf{N}, X_n^\nu := X_{n \wedge \nu}$.

**Lem.** *Soit $X$ une surmartingale (resp. sous-martingale, martingale) et $\nu$ un temps d'arrêt. Alors le processus arrêté $X^\nu$ est une surmartingale (resp. sous-martingale, martingale).*

**Th.** *Soit $X$ une martingale (resp. surmartingale) et $\underline{\nu}$, $\overline{\nu}$ deux temps d'arrêt bornés dans $\mathcal{T}$ vérifiant $\underline{\nu} \leqslant \overline{\nu}$ p.s. Alors $\mathbf{E}[X_{\overline{\nu}} \mid \mathcal{F}_{\underline{\nu}}] =$ (resp. $\leqslant$)$X_{\underline{\nu}}$.*

**Prop.** *Soit $X = (X_n)_n$ un processus aléatoire $\mathbf{F}$-adapté, $\forall n \in \mathbf{N}, \mathbf{E}(|X_n|) < \infty$. Alors $X$ est une martingale ssi $\mathbf{E}[X_\nu] = \mathbf{E}[X_0]$ pour tout temps d'arrêt $n$ borné.*

**Def.** Une martingale $(X_n)_n$ est **fermée** s'il existe une v.a. réelle intégrable $Y$ telle que $\forall n \in \mathbf{N}, X_n = \mathbf{E}(Y \mid X_n)$.

**Th.** *Toute martingale fermée est uniformément intégrable.*

**Th** (Inégalité maximale de Doob)**.** *Soit $M = (M_n)_{n \geqslant 0}$ une sous-martingale, et $M_n^* := \sup_{k \leqslant n} M_k$ son processus de maximum courant.*

*(i)* $\forall c > 0, \forall n \in \mathbf{N}, c\mathbf{P}(M_n^* \geqslant c) \leqslant \mathbf{E}(M_n \mathbf{1}_{M_n^* \geqslant c})$

*(ii)* *Soit $p > 1$ et supposons que la sous-martingale $M$ est positive et $\forall n \in \mathbf{N}, M_n \in \mathcal{L}^p$. Alors $M_n^* \in \mathcal{L}^p$ et $\|M_n^*\|_p \leqslant \frac{p}{p-1} \|M_n\|_p$.*

**Prop** (Décomposition de Doob)**.** *Soit $(X_n)_n$ un processus aléatoire intégrable. Il existe une martingale $(M_n)_n$ et un processus $\mathbf{F}$-prévisible $(V_n)_n$ tels que $M_0 = V_0 = 0$ et $\forall n \geqslant 0, X_n = X_0 + M_n + V_n$. Cette décomposition est unique.*

*Rem.* On voit que : $X$ est une surmartingale ssi $V$ est décroissant, $X$ est une sous-martingale ssi $V$ est croissant et $X$ est une martingale ssi $V = 0$.

**Prop.** *Soit $X = (X_n)_n$ une martingale de carré intégrable, et $\Delta X_n := X_n - X_{n-1}$. Alors $X_n^2 = X_0^2 + N_n + [X]_n$ où $N_n := 2\sum_{i=1}^n X_{i-1}\Delta X_i$, $[X]_n := \sum_{i=1}^n (\Delta X_i)^2$ et $N_0 = [X]_0 = 0$. Dans cette décomposition, $(N_n)_n$ est une martingale nulle en zéro, et $([X]_n)_n$ est un processus $\mathbf{F}$-adapté croissant intégrable appelé **variation quadratique** de la martingale $X$.*

**Def.** Un processus $X = (X_n)_{n \in \mathbf{N}}$ est une **martingale locale** s'il existe une suite de temps d'arrêt $(\tau_n)_n$ telle que $\tau_n \longrightarrow \infty$ $\mathbf{P}$-p.s. et le processus arrêté $X^{\tau_n}$ est une martingale pour tout $n$.

**Lem.** *Soit $X = \{X_n, n \in [\![0\,;N]\!]\}$ une martingale locale telle que $\mathbf{E}[X_N^-] < \infty$. Alors $X$ est une martingale.*