

MDI 104 : Probabilités

Chaînes de Markov

Def. Chaîne de Markov : suite $(X_n)_{n \in \mathbb{N}}$ de v.a. à valeurs dans $\llbracket 1, N \rrbracket$ telle que $\forall n \in \mathbb{N}, \forall x_0, \dots, x_{n+1} \in \llbracket 1, N \rrbracket, p(x_{n+1} \mid x_0, \dots, x_n) = p(x_{n+1} \mid x_n)$.

Prop. $\forall n \in \mathbb{N}, \forall k \geq 1, \forall x_0, \dots, x_{n+1} \in \llbracket 1, N \rrbracket, p(x_{n+1}, \dots, x_{n+k} \mid x_0, \dots, x_n) = p(x_{n+1}, \dots, x_{n+k} \mid x_n)$.

Prop. $\forall n \in \mathbb{N}, \forall k \geq 1, \forall x_n, \dots, x_{n+k} \in \llbracket 1, N \rrbracket, p(x_{n+1}, \dots, x_{n+k} \mid x_n) = p(x_{n+1} \mid x_n) \cdots p(x_{n+k} \mid x_{n+k-1})$.

Def. Une chaîne de Markov est dite **homogène** si $\forall n \in \mathbb{N}, \mathbf{P}(X_{n+1} = j \mid X_n = i) = \mathbf{P}(X_1 = j \mid X_0 = i)$. L'évolution de la chaîne de Markov ne dépend alors que de sa matrice de transition P , définie par $\forall i, j \in \llbracket 1, N \rrbracket^2, P_{ij} = p(j \mid i)$.

C'est une matrice stochastique, i.e. $\forall i \in \llbracket 1, N \rrbracket, \sum_j P_{ij} = 1$. En notant la loi de X_n comme un vecteur ligne $\pi(n)$ de dimension N , avec $\pi(n) \mathbf{1} = 1$ il vient $\forall n \in \mathbb{N}, \pi(n) = \pi(0)P^n$. On parle de **loi stationnaire** si $\pi = \pi P$.

Def. Une chaîne de Markov est dite **irréductible** si son graphe de transition est fortement connexe, i.e. $\forall i, j, i \neq j$, il existe un chemin de i vers j dans le graphe de transition. Cette propriété ne dépend que de la structure du graphe de transition et non des poids.

Prop. Une chaîne de Markov est irréductible si et seulement si $\exists n \in \mathbb{N}^*, (I + P)^n > 0$ (positivité sur les coefficients).

Th (Perron-Frobenius partiel). Une chaîne de Markov irréductible admet une unique loi stationnaire.

Cor. La loi stationnaire π d'une chaîne de Markov irréductible vérifie $\forall i \in \llbracket 1, N \rrbracket, \pi_i > 0$.

Def. Période d'un état : pgcd des longueurs des cycles du graphe de transition passant par cet état. Un état est apériodique lorsqu'il est de période 1.

Mesure et intégration

Def. Tribu sur Ω : partie \mathcal{A} de $\mathcal{P}(\Omega)$ qui possède les propriétés suivantes :

- Élément neutre : $\Omega \in \mathcal{A}$;
- Stabilité par passage au complémentaire : $\forall A \in \mathcal{A}, \bar{A} \in \mathcal{A}$;
- σ -additivité : si $(A_i)_{i \in I}$ est une famille d'éléments de \mathcal{A} , alors $\bigcup_{i \in I} A_i \in \mathcal{A}$.

Prop. On a $\emptyset \in \mathcal{A}$ et $\forall A_1, \dots \in \mathcal{A}, \bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

Lem. L'intersection de 2 tribus est une tribu.

Th. Soit $\mathcal{C} \subset \mathcal{P}(\Omega)$. L'intersection $\sigma(\mathcal{C})$ de toutes les tribus sur Ω contenant \mathcal{C} est une tribu sur Ω appelée tribu engendrée par \mathcal{C} sur Ω .

Def (Tribu borélienne). La tribu de Borel sur \mathbf{R}^d , notée $\mathcal{B}(\mathbf{R}^d)$, est la tribu engendrée par les pavés (ouverts). Ses éléments sont appelés boréliens.

Th (Éléments caractéristiques de $\mathcal{B}(\mathbf{R})$). $\mathcal{B}(\mathbf{R})$ contient : $\{a\}, [a; b],]-\infty; b], [a; +\infty[, \mathbf{Q}$ et $\mathbf{R} \setminus \mathbf{Q}$.

Def. Une application $X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ est dite \mathcal{F}/\mathcal{E} -mesurable si $\forall H \in \mathcal{E}, X^{-1}(H) \in \mathcal{F}$.

Th. Soit (E', \mathcal{E}') un espace mesurable, $X : \Omega \rightarrow E$ une application \mathcal{F}/\mathcal{E} -mesurable et $f : E \rightarrow E'$ une application \mathcal{E}/\mathcal{E}' -mesurable. Alors $f \circ X$ est \mathcal{F}/\mathcal{E}' -mesurable.

Th. $(\forall i, X_i \text{ est } \mathcal{F}/\mathcal{B}(\mathbf{R})\text{-mesurable}) \iff ((X_1, \dots, X_d) \text{ est } \mathcal{F}/\mathcal{B}(\mathbf{R}^d)\text{-mesurable})$.

Th. Une application $f : \mathbf{R}^d \rightarrow \mathbf{R}^n$ continue est $\mathcal{B}(\mathbf{R}^d)/\mathcal{B}(\mathbf{R}^n)$ -mesurable. On dit alors qu'elle est borélienne.

Cor. Sont mesurables : $f \circ X, X + Y, XY, X \vee Y, X \wedge Y, \sup_n X_n, \inf_n X_n, \liminf_n X_n, \limsup_n X_n$ et $\lim_n X_n$ si les fonctions en jeu sont mesurables.

Def. Mesure sur (E, \mathcal{E}) : fonction $\mu : \mathcal{E} \rightarrow [0; +\infty]$ telle que $\mu(\emptyset) = 0$ et, pour des événements $(A_n)_{n \in \mathbb{N}}$ deux à deux disjoints, $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$.

Lem (Borel-Cantelli). $\sum_{n \geq 1} \mathbf{P}(A_n) < +\infty \implies \mathbf{P}(\limsup_n A_n) = 0$

Not. \mathcal{I}_d est l'ensemble des pavés (fermés) de \mathbf{R}^d .

Def. Une mesure μ sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ est localement finie si $\forall x \in \mathbf{R}^d, \exists I \subset \mathcal{I}_d, x \in I, \mu(I) < \infty$.

Th. Deux mesures localement finies sur \mathbf{R}^d qui coïncident sur \mathcal{I}_d sont égales.

Th (Mesure de Lebesgue, Kolmogorov). Il existe une unique mesure λ sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ qui coïncide avec la mesure de volume sur les pavés. Elle est invariante par translation.

Th. Il existe une unique mesure $\mu \otimes \nu$, dite mesure produit de μ et ν sur $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ telle que $\forall A \in \mathcal{E}_1, \forall B \in \mathcal{E}_2, \mu \otimes \nu(A \times B) = \mu(A)\nu(B)$.

Not. Une propriété est vraie μ -presque-partout (μ -p.p.) ou μ -presque-sûrement (μ -p.s.) lorsque son complémentaire est de mesure nulle.

Th. Soit f v.a. positive, de μ -intégrale nulle. Alors f est nulle μ -p.p.

Th (Convergence monotone). Soit $(f_n)_{n \geq 1}$ mesurables positives qui convergent en croissant μ -p.p. vers f . Alors $\lim_n \int f_n d\mu = \int f d\mu$.

Lem (Lemme de Fatou). Soit f_n des fonctions mesurables positives. Alors $\int (\liminf f_n) \leq \liminf (\int f_n)$.

Th (Convergence dominée). Soit des fonctions mesurables (f_n) , $f_n \xrightarrow{p.p.} f$. S'il existe g positive intégrable telle que $\forall n, |f_n| \leq g$ alors f est intégrable et $\int f_n \rightarrow \int f$.

Th (Fubini-Tonelli). Soit $f: E \times F \rightarrow \mathbf{R}_+$ sur $(E \times F, \mathcal{E} \otimes \mathcal{F}, \mu \otimes \nu)$. Alors

$$\iint_{E \times F} f(x, y) d\mu \otimes \nu(x, y) = \int_F \left[\int_E f(x, y) d\mu(x) \right] d\nu(y) = \int_E \left[\int_F f(x, y) d\nu(y) \right] d\mu(x).$$

Th (Fubini). Le résultat précédent est encore vrai en supposant f intégrable et non plus positive.

Th (Continuité sous le signe somme). La fonction $f: t \mapsto \int f(x, t) d\mu(x)$ est continue sous les hypothèses suivantes : $f(\cdot, t)$ est mesurable, $f(x, \cdot)$ est continue et $|f(x, \cdot)| \leq g(x)$ avec g intégrable.

Th. On a $\frac{d}{dt} (\int f(x, t) d\mu(x)) = \int \frac{\partial f}{\partial t}(x, t) d\mu(x)$ et la dérivabilité sous les hypothèses suivantes : $f(\cdot, t)$ est mesurable, $f(x, \cdot)$ est dérivable p.p. et $\left| \frac{\partial f}{\partial t}(x, \cdot) \right| \leq g(x)$ avec g intégrable.

Th. On a $\mathbf{P}_1 = \mathbf{P}_2$ sur \mathbf{R}^d si et seulement si, $\forall f$ mesurable positive (ou mesurable bornée), $\int f d\mathbf{P}_1 = \int f d\mathbf{P}_2$.

Th. On dit que ν est absolument continue par rapport à μ si $\forall A, \mu(A) = 0 \implies \nu(A) = 0$. On écrit $\nu \ll \mu$.

Th (Radon-Nykodym). $(\nu \ll \mu) \iff (\exists f \geq 0, \forall g \text{ mesurable bornée}, \int g d\nu = \int g \cdot f d\mu)$. f s'appelle densité de ν par rapport à μ .

Domaine	Densité	Expression de $f(x)$	Notation
\mathbf{R}^d	Densité uniforme sur $A \subset \mathbf{R}^d$	$\frac{\mathbf{1}_A(x)}{\int \mathbf{1}_A}$	$\mathcal{U}(A)$
\mathbf{R}	Densité exponentielle, $\alpha > 0$	$\alpha e^{-\alpha x} \mathbf{1}_{\mathbf{R}_+}(x)$	$\mathcal{E}(\alpha)$
\mathbf{R}	Densité gaussienne, $m \in \mathbf{R}, \sigma^2 > 0$	$\frac{e^{-(x-m)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}$	$\mathcal{N}(m, \sigma^2)$
\mathbf{R}^d	Gaussienne multivariée, $m \in \mathbf{R}^d, \Sigma \in \mathfrak{M}_{r,r}(\mathbf{R})$ définie positive	$\frac{e^{-\frac{1}{2}(x-m)\Sigma^{-1}(x-m)}}{\sqrt{(2\pi)^d \det(\Sigma)}}$	$\mathcal{N}_d(m, \Sigma)$
\mathbf{R}	Densité de Cauchy, $m \in \mathbf{R}, \alpha > 0$	$\frac{1}{\pi} \frac{\alpha}{(x-m)^2 + \alpha^2}$	
\mathbf{R}	Densité Gamma, $a > 0, b > 0$	$x^{a-1} \frac{b^a e^{-bx}}{\Gamma(a)}$	$\Gamma(a, b)$

Fonction gamma d'Euler : $\Gamma: a \mapsto \int_0^{+\infty} x^{a-1} e^{-x} dx$.

Variables et vecteurs aléatoires réels

Def. Variable aléatoire X sur E : application mesurable de Ω dans E .

Def. Loi de X : fonction $\mathbf{P} \circ X^{-1} = \mathbf{P}_X: \mathcal{E} \rightarrow [0; 1]$ $H \mapsto \mathbf{P}(X^{-1}(H))$.

Def. Fonction de répartition de X v.a.r. : $F_X: \mathbf{R} \rightarrow [0; 1]$ $x \mapsto \mathbf{P}(X \leq x)$.

Def. Une fonction F est dite absolument continue si $\forall \epsilon > 0, \exists \delta > 0, \forall [a_1; b_1], \dots, [a_p; b_p]$ disjoints, $(\sum_{k=1}^p (b_k - a_k) < \delta) \implies (\sum_{k=1}^p |F(b_k) - F(a_k)| < \epsilon)$.

Th (v.a à densité). Soit X tel que $\mathbf{P}_X \ll \lambda, \exists f \geq 0$ intégrable tel que $\forall g$ mesurable bornée, $d\mathbf{P}_X(x) = f(x) dx$. C'est équivalent au fait que F_X soit absolument continue.

Def. Espérance de X : $\mathbf{E}(X) := \int X d\mathbf{P}$. En particulier $\mathbf{P}(X \in A) = \mathbf{E}(\mathbf{1}_A(X))$.

Th (Théorème de transfert). Soit $X: \Omega \rightarrow E$ et $g: E \rightarrow \mathbf{R}$ mesurables. Alors $\mathbf{E}(g(X)) = \int_{\Omega} g(X(\omega)) d\mathbf{P}(\omega) = \int_{\mathbf{R}} g(x) d\mathbf{P}_X(x) = (\int g(x) f_X(x) dx)$.

Th (Inégalité de Markov). $\mathbf{P}(|X| > \epsilon) \leq \frac{\mathbf{E}(|X|^p)}{\epsilon^p}$.

Th (Inégalité de Hölder). Soit $p, q \geq 0, \frac{1}{p} + \frac{1}{q} = 1$, alors $\mathbf{E}(|XY|) \leq \mathbf{E}(|X|^p)^{\frac{1}{p}} \mathbf{E}(|Y|^q)^{\frac{1}{q}}$.

Th (Inégalité de Jensen). Soit $\varphi: \mathbf{R} \rightarrow \mathbf{R}$ convexe et X une v.a.r. telle que $\mathbf{E}(|X|) < \infty$ et $\mathbf{E}(|\varphi(X)|) < \infty$. Alors $\varphi(\mathbf{E}(X)) \leq \mathbf{E}(\varphi(X))$.

Def. Moment d'ordre $p \geq 0$: $\mathbf{E}(X^p)$ lorsque $\mathbf{E}(|X|^p) < \infty$.

Prop. Une variable d'ordre p possède tous ses moments d'ordre inférieur.

Def. Variance de X d'ordre 2 : $\text{Var}(X) := \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$. Écart-type : $\sigma_X := \sqrt{\text{Var}(X)}$.

Def. Covariance de X et Y d'ordre 2 : $\text{Cov}(X, Y) := \mathbf{E}((X - \mathbf{E}(X))(Y - \mathbf{E}(Y))) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$.

Coefficient de corrélation : $\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$. X et Y sont décorrélées si $\text{Cov}(X, Y) = 0$. C'est le produit scalaire associé à Var . En particulier $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$.

Def. Sur \mathbf{R}^d on a $F_X(x_1, \dots, x_d) = \mathbf{P}(X_1 \leq x_1, \dots, X_d \leq x_d) = \mathbf{P}_X \left(\prod_{k=1}^d]-\infty; x_k] \right)$.

Th. Les propositions suivantes sont équivalentes :

- X_1, \dots, X_d sont indépendantes,
- $\forall x_1, \dots, x_d \in \mathbf{R}, F_X(x_1, \dots, x_d) = \prod_{k=1}^d F_{X_k}(x_k)$,
- pour toutes fonctions h_1, \dots, h_d mesurables $\mathbf{R} \rightarrow \mathbf{R}$ telles que les v.a. $h(X_i)$ sont toutes positives ou toutes intégrables, $\mathbf{E} \left(\prod_{k=1}^d h_k(X_k) \right) = \prod_{k=1}^d \mathbf{E}(h_k(X_k))$.
- même égalité avec des h_i positives continues et à support compact,
- (si les X_i admettent des densités) $\exists f_X, \forall x_1, \dots, x_d \in \mathbf{R}, f_X(x_1, \dots, x_d) = \prod_{k=1}^d f_{X_k}(x_k)$.

Def. Une famille de v.a. est dite indépendante si toute sous-famille finie l'est.

Lem. Soit $(X_i)_{i \in I}$ des v.a. indépendantes et J, K disjoints inclus dans I . Alors $(X_i)_{i \in J}$ et $(X_i)_{i \in K}$ sont indépendantes.

Def. Soit $\phi: \begin{matrix} O \subset \mathbf{R}^n & \rightarrow & \mathbf{R}^n \\ x & \mapsto & (\phi_1(x), \dots, \phi_d(x)) \end{matrix}$ avec O ouvert. Si toutes les dérivées partielles existent sur O ,

la jacobienne de ϕ en x est la matrice $D_x(\phi) = \left(\frac{\partial \phi_i}{\partial x_j}(x) \right)_{1 \leq i, j \leq n}$. Le jacobien de ϕ est donné par $\forall x, J_\phi(x) := \det(D_x(\phi))$.

Def. Soit $O \subset \mathbf{R}^n$ ouvert et $\phi: O \rightarrow \mathbf{R}^n$. C'est un \mathcal{C}^1 -difféomorphisme de O sur $\Delta \subset \mathbf{R}^n$ lorsque : les dérivées partielles de ϕ existent et sont continues sur O , ϕ est bijective de O sur Δ et J_ϕ ne s'annule pas sur O . Alors ϕ^{-1} est également continuellement dérivable et on a $J_{\phi^{-1}} = \frac{1}{J_\phi \circ \phi^{-1}}$.

Th (Formule du changement de variables). Soit U et V deux ouverts de \mathbf{R}^d , $\phi: U \rightarrow V$ un difféomorphisme et $f: V \rightarrow \mathbf{R}_+$. Alors $\int_U f \circ \phi = \int_V f |J_{\phi^{-1}}|$ ou encore $\int_U f \circ \phi \cdot |J_\phi| = \int_V f$.

Loi conditionnelle

Def. Noyau de probabilité : application $\mathcal{Q}: \begin{matrix} \mathbf{R}^n \times \mathcal{B}(\mathbf{R}^p) & \rightarrow & [0; 1] \\ (x, A) & \mapsto & \mathcal{Q}_x(A) \end{matrix}$ telle que $\forall x \in \mathbf{R}^n$, \mathcal{Q}_x est une probabilité sur $\mathcal{B}(\mathbf{R}^p)$ et $\forall A \in \mathcal{B}(\mathbf{R}^p)$, $\mathcal{Q}(\cdot, A)$ est mesurable.

Def. Support d'une probabilité : complémentaire du plus grand ouvert de probabilité nulle.

Def. Soit X et Y des v.a. dans \mathbf{R}^n et \mathbf{R}^p et E le support de X . La loi conditionnelle de Y sachant X est l'unique famille de noyaux $(\mathcal{Q}_x)_{x \in E}$ définie pour toutes fonctions ψ et θ mesurables bornées par

$$\mathbf{E}(\psi(X)\theta(Y)) = \mathbf{E} \left(\psi(X) \int_{\mathbf{R}^p} \theta(y) d\mathcal{Q}_X(y) \right) = \int_{\mathbf{R}^n} \psi(x) \int_{\mathbf{R}^p} \theta(y) d\mathcal{Q}_x(y) d\mathbf{P}_X(x)$$

Th. Pour toute fonction $\psi: \mathbf{R}^n \times \mathbf{R}^p \rightarrow \mathbf{R}$ mesurable bornée, $\mathbf{E}(\psi(X, Y)) = \int_{\mathbf{R}^n \times \mathbf{R}^p} \psi(x, y) d\mathcal{Q}_X(y) d\mathbf{P}_X(x)$.

Th. Si X et Y sont indépendantes, $\forall x \in E$, $\mathcal{Q}_x = \mathbf{P}_Y$. Si $Y = T(X)$, $\forall x \in E$, $\mathcal{Q}_x = \delta_{T(x)}$.

Th. Soit X et Y deux v.a. discrètes de supports respectifs E_X et E_Y . Alors $\forall x \in E_X$, $\mathcal{Q}_x = \sum_{y \in E_Y} \mathbf{P}(Y = y | X = x) \delta_y$.

Th. Soit $X: \Omega \rightarrow \mathbf{R}^n$ et $Y: \Omega \rightarrow \mathbf{R}^p$, de densité jointe $f_{(X,Y)}$. La loi conditionnelle de Y sachant X est donnée par $d\mathcal{Q}_x(y) = \frac{f_{(X,Y)}(x,y)}{\int_{\mathbf{R}^p} f_{(X,Y)}(x,y) dy} dy$.

Def. On appelle **espérance conditionnelle** de $\theta(Y)$ sachant $X = x$ la quantité $\mathbf{E}(\theta(Y) | X = x) = \int \theta(y) d\mathcal{Q}_x(y)$.

Prop. On a $\mathbf{E}(\mathbf{E}(\theta(Y) | X)) = \mathbf{E}(\theta(Y))$.

Th. Soit X et Y deux v.a. discrètes et E_Y le support de Y . Alors $\mathbf{E}(\theta(Y) | X = x) = \sum_{y \in E_Y} \theta(y) \mathbf{P}(Y = y | X = x)$.

Vecteurs Gaussiens

Def. $\mathbf{E}(X) = {}^t(\mathbf{E}(X_1), \dots, \mathbf{E}(X_d))$ est bien définie si et seulement si toutes ses composantes le sont.

Def. Matrice de covariance d'un vecteur X d'ordre 2 : $\text{Cov}(X) := (\text{Cov}(X_i, X_j))_{1 \leq i, j \leq d} \in \mathcal{S}_d^+(\mathbf{R})$. La diagonale de $\text{Cov}(X)$ est égale au vecteur des variances et, dans le cas où les X_i sont décorréliées, $\text{Cov}(X)$ est diagonale.

Prop. Soit $A \in \mathcal{M}_{n,d}(\mathbf{R})$ et $b \in \mathcal{M}_{d,1}(\mathbf{R})$. On a $\mathbf{E}(AX + b) = A\mathbf{E}(X) + b$, $\text{Cov}(AX + b) = A\text{Cov}(X)^t A$, et $\text{Cov}(X) = \mathbf{E}(X_c {}^t X_c)$ où $X_c := X - \mathbf{E}(X)$ est le vecteur recentré.

On utilisera $m \in \mathbf{R}^d$ et $\Gamma \in \mathcal{S}_d^+(\mathbf{R})$.

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 $\langle a | X \rangle$ est une loi gaussienne (éventuellement de variance nulle).

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