A

Rappels de probabilités

Ce document est un complément au livre [CM15]. Il regroupe des éléments sur l'intégration, les martingales, les chaînes de Markov, et le calcul stochastique. Le contenu est inspiré en partie du livre [BC07]. Les démonstrations sont omises.

A.1 Propriétés élémentaires

Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements alors

$$\varlimsup_n A_n := \cap_n \cup_{m\geqslant n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ pour une infinité de valeurs de } n\},$$

et

$$\varliminf_n A_n := \cup_n \cap_{m\geqslant n} A_m = \{\omega \in \Omega : \omega \in A_n \text{ à partir d'un certain rang sur } n\}.$$

On a

$$\varlimsup_n A_n^c = (\varliminf_n A_n)^c \quad \text{et} \quad \varliminf_n A_n^c = (\varlimsup_n A_n)^c,$$

et

$$\mathbb{1}_{\varlimsup_n A_n} = \varlimsup_n \mathbb{1}_{A_n} \quad \text{et} \quad \mathbb{1}_{\varliminf_n A_n} = \varliminf_n \mathbb{1}_{A_n}.$$

Lemme A.1 (Borel-Cantelli)**.** *Soit $(A_n)_{n\in\mathbb{N}}$ est une famille d'événements.*

- *(Cantelli) Si $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$ alors $\mathbb{P}(\varlimsup_n A_n) = 0$;*
- *(Loi du zéro-un de Borel) Si les $(A_n)_{n\in\mathbb{N}}$ sont indépendants alors $\mathbb{P}(\varlimsup_n A_n) \in \{0,1\}$ et $\mathbb{P}(\varlimsup_n A_n) = 0$ ssi $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$.*

Lemme A.2 (Loi du zéro-un de Kolmogorov)**.** *Si $(\mathcal{F}_n)_{n\geqslant 1}$ est une suite de tribus indépendantes alors*

$$\mathbb{P}(A) \in \{0,1\} \textit{ pour tout } A \in \varlimsup_n \mathcal{F}_n := \cap_{n\geqslant 1} \sigma(\cup_{m\geqslant n} \mathcal{F}_m).$$

On dit que $\overline{\lim}_n \mathcal{F}_n$ est la tribu *aymptotique* ou tribu *terminale*[1].

Théorème A.3 (Inégalités de Markov et de Tchebychev)**.** *Si X est une variable aléatoire positive et si Y est une v.a. dans L^2, alors pour tout réel $r > 0$, l'inégalité de Markov et l'inégalité de Tchebychev s'écrivent*

$$\mathbb{P}(X \geqslant r) \leqslant \frac{\mathbb{E}(X)}{r} \quad \textit{et} \quad \mathbb{P}(|Y - \mathbb{E}(Y)| \geqslant r) \leqslant \frac{\mathrm{Var}(Y)}{r^2}.$$

Il est temps d'introduire les deux notions de convergence suivantes :

— On dit qu'une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires converge *presque sûrement* («p.s.») vers une variable aléatoire X_∞ lorsque

$$\mathbb{P}(\lim_{n\to\infty} X_n = X_\infty) = 1,$$

et on note

$$X_n \xrightarrow[n\to\infty]{\text{p.s.}} X_\infty.$$

— On dit qu'une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires converge *en probabilité* vers une variable aléatoire X_∞ lorsque

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X_\infty| \geqslant \varepsilon) = 0,$$

et on note

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}} X_\infty.$$

Lemme A.4 (Lemme de Fatou)**.** *Si $(X_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires positives alors $\mathbb{E}(\underline{\lim}_{n\to\infty} X_n) \leqslant \underline{\lim}_{n\to\infty} \mathbb{E}(X_n)$.*

Théorème A.5 (Convergence monotone)**.** *Si $(X_n)_{n\in\mathbb{N}}$ est une suite croissante de v.a. positives alors $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(\lim_{n\to\infty} X_n)$ dans $\overline{\mathbb{R}}_+$. En particulier $\lim_{n\to\infty} X_n \in L^1$ ssi $\lim_{n\to\infty} \mathbb{E}(X_n) = \sup_{n\in\mathbb{N}} \mathbb{E}(X_n) < \infty$.*

Théorème A.6 (Convergence dominée)**.** *Si $(X_n)_{n\in\mathbb{N}}$ et X_∞ sont des v.a. avec $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X_\infty$ et $\sup_{n\in\mathbb{N}} |X_n| \in L^1$ alors $X_\infty \in L^1$ et $X_n \xrightarrow[n\to\infty]{L^1} X_\infty$.*

En pratique, le théorème de convergence dominée est utilisé quand on a $X_n \to X_\infty$ presque sûrement, et on recherche une variable aléatoire $Y \in L^1$ telle que $|X_n| \leqslant Y$ pour tout $n \in \mathbb{N}$ pour assurer que $\sup_{n\in\mathbb{N}} |X_n| \in L^1$.

Théorème A.7 (Intégrabilité uniforme)**.** *Pour toute famille de variables aléatoires intégrables $(X_i)_{i\in I} \subset L^1$, les propriétés suivantes sont équivalentes :*

1. *Intégrabilité uniforme : $\lim_{x\to\infty} \sup_{i\in I} \mathbb{E}(|X_i| \mathbb{1}_{\{|X_i| \geqslant x\}}) = 0$;*
2. *Critère epsilon-delta : la famille est bornée dans L^1, c'est-à-dire que $\sup_{i\in I} \mathbb{E}(|X_i|) < \infty$, et de plus, pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que pour tout événement A, si $\mathbb{P}(A) \leqslant \delta$ alors $\sup_{i\in I} \mathbb{E}(|X_i| \mathbb{1}_A) \leqslant \varepsilon$;*

1. « *Tail sigma field* » en anglais.

3. *Critère de de la Vallée Poussin : il existe $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ convexe avec*

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = \infty \quad \textit{et} \quad \sup_{i\in I} \mathbb{E}(\varphi(|X_i|)) < \infty;$$

4. *Critère de Dunford-Pettis : l'ensemble $(X_i)_{i\in I}$ est (aussi séquentiellement) relativement compact dans L^1 pour la topologie faible $\sigma(L^1, L^\infty)$.*

Les parties finies de L^1 sont toutes uniformément intégrables. Si $X \in L^1$ alors le critère de de la Vallée Poussin pour le singleton $\{X\} \subset L^1$ donne $\varphi(X) \in L^1$ pour φ convexe et sur-linéaire qui dépend de X. Pour comprendre ce phénomène, au cœur de la notion d'intégrabilité uniforme, on peut penser à la condition de sommabilité des séries de Riemman, qui est ouverte.

Le critère de de la Vallée Poussin entraîne que si $(X_i)_{i\in I}$ est bornée dans L^p avec $p > 1$, c'est-à-dire que $\sup_{i\in I} \mathbb{E}(|X_i|^p) < \infty$, alors $(X_i)_{i\in I}$ est uniformément intégrable. On prendra garde à ne pas confondre la bornitude dans L^p avec la condition de domination $\sup_{i\in I} |X_i| \in L^p$, qui est plus forte.

L'intégrabilité uniforme remplace avantageusement la condition de domination du théorème de convergence dominée. Si $(X_n)_{n\in\mathbb{N}}$ et X_∞ sont des variables aléatoires dans L^1 alors $\lim_{n\to\infty} X_n = X_\infty$ dans L^1 si et seulement si $\lim_{n\to\infty} X_n = X$ en probabilité et $(X_n)_{n\in\mathbb{N}}$ est uniformément intégrable.

Convergence en loi

Soit L une loi de probabilité de fonction de répartition F_∞, et soit X_∞ une variable aléatoire de loi L. On dit qu'une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires converge *en loi* vers L, et on note

$$X_n \xrightarrow[n\to\infty]{\text{loi}} L$$

lorsque $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X_\infty))$ pour tout $f \in \mathcal{F}$ pour une classe de fonctions $\mathcal{F}$ parmi les classes suivantes :
- (fonctions continues et bornées) $\mathcal{F} = \mathcal{C}_b$;
- (fonction caractéristique) $\mathcal{F} = \{x \mapsto e^{itx} : t \in \mathbb{R}\}$;
- (fonction de répartition) $\mathcal{F} = \{\mathbb{1}_{]-\infty,x]} : x$ point de continuité de $F_\infty\}$.

Théorème A.8 (Lien entre les notions de convergence)**.**

$$\begin{array}{c} CV\ L^{p>1} \\ \Downarrow \\ CV\ L^1 \\ \Downarrow^3 \\ CV\ p.s. \Rightarrow^2 CV\ en\ \mathbb{P} \Rightarrow^1 CV\ en\ loi \end{array}$$

La réciproque dans 1 *a lieu lorsque la limite est constante, dans* 2 *le long de sous-suites, et dans* 3 *si la suite est uniformément intégrable.*

La *fonction caractéristique* d'une variable aléatoire X est définie par $\varphi_X : t \in \mathbb{R} \mapsto \varphi_X(t) := \mathbb{E}(e^{itx}) \in \mathbb{C}$. Elle ne dépend que de la loi de X.

Théorème A.9 (Fonctions caractéristique)**.**

— *(Caractérisation) Pour toutes v.a. X et Y on a $\varphi_X = \varphi_Y$ ssi $X \overset{\text{loi}}{=} Y$;*
— *(Paul Lévy) Si $(X_n)_{n\in\mathbb{N}}$ sont des v.a. telles que $\lim_{n\to\infty} \varphi_{X_n}(t) = \varphi(t)$ pour tout $t \in \mathbb{R}$, avec $t \in \mathbb{R} \mapsto \varphi(t) \in \mathbb{C}$ continue en 0, alors φ est continue sur $\mathbb{R}$, il existe une variable aléatoire X_∞ telle que $\varphi_X = \varphi$, et $(X_n)_{n\in\mathbb{N}}$ converge en loi vers la loi de X_∞ ;*
— *(Bochner) Une fonction continue $\varphi : \mathbb{R} \to \mathbb{C}$ est la fonction caractéristique d'une v.a. ssi $\varphi(0) = 1$ et φ est définie positive c'est-à-dire que pour tout $n \geqslant 1$, tous $t_1, \ldots, t_n \in \mathbb{R}$, et tout $c \in \mathbb{C}^n$,*

$$\sum_{j=1}^{n} \sum_{k=1}^{n} c_j \varphi(t_j - t_k) \overline{c_k} \geqslant 0.$$

Lemme A.10 (Slutsky)**.** *Si $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ sont des suites de v.a. avec*

$$X_n \xrightarrow[n\to\infty]{\text{loi}} X_\infty \quad \textit{et} \quad Y_n \xrightarrow[n\to\infty]{\text{loi}} c$$

où c est une constante, alors

$$(X_n, Y_n) \xrightarrow[n\to\infty]{\text{loi}} \mathrm{Loi}(X_\infty, c).$$

et en particulier, pour toute fonction continue $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, on a

$$f(X_n, Y_n) \xrightarrow[n\to\infty]{\text{loi}} \mathrm{Loi}(f(X_\infty, c)).$$

A.2 Théorèmes limites classiques

Théorème A.11 (Loi des grands nombres (LGN))**.** *Si $(X_n)_{n\in\mathbb{N}}$ sont des v.a. i.i.d. intégrables de moyenne m, alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{p.s.}} m.$$

Théorème A.12 (Loi du logarithme itéré de Strassen (LLI))**.** *Si $(X_n)_{n\in\mathbb{N}}$ sont i.i.d. de carré intégrable, de moyenne m et de variance σ^2, alors p.s.*

$$\varliminf_{n\to\infty} \frac{X_1 + \cdots + X_n - nm}{\sqrt{2n \log(\log(n))}} = -\sigma \quad \textit{et} \quad \varlimsup_{n\to\infty} \frac{X_1 + \cdots + X_n - nm}{\sqrt{2n \log(\log(n))}} = +\sigma.$$

La loi du logarithme itéré se situe entre LGN et TLC.

Théorème A.13 (Théorème limite central (TLC))**.** *Si* $(X_n)_{n\in\mathbb{N}}$ *sont des v.a. i.i.d. de carré intégrable, de moyenne* m *et de variance* σ^2 *alors*

$$\sqrt{n}\left(\frac{X_1+\cdots+X_n}{n}-m\right)=\frac{X_1-m+\cdots+X_n-m}{\sqrt{n}}\xrightarrow[n\to\infty]{\text{loi}}\mathcal{N}(0,\sigma^2).$$

Le TLC entraîne la LGN faible (convergence en $\mathbb{P}$) par le lemme de Slutsky.

Théorème A.14 (Berry-Esseen (vitesse dans TLC))**.** *Si* $(X_n)_{n\in\mathbb{N}^*}$ *est une suite de v.a. i.i.d. dans* L^3 *de moyenne* m *et de variance* σ^2*, alors, en notant* $\tau^3 := \mathbb{E}(|X-m|^3)$*, on a, pour tout* $n \geqslant 1$*,*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(\frac{X_1-m+\cdots+X_n-m}{\sqrt{n}\sigma}\leqslant t\right)-\int_{-\infty}^{t}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx\right|\leqslant\frac{\tau^3}{\sqrt{n}\sigma^3}.$$

Le théorème A.14 de Berry-Esseen donne une majoration de la vitesse de convergence en loi du TLC, au niveau des fonctions de répartition.

LGN et TLC uniformes

Théorème A.15 (Glivenko-Cantelli (LGN uniforme))**.** *Si* $(X_n)_{n\in\mathbb{N}}$ *sont des v.a. i.i.d. de fonction de répartition commune* F*, et si, pour tout* $n \geqslant 1$*,* F_n *est la fonction de répartition de la mesure empirique* $\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$*, alors*

$$\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{p.s.}} 0.$$

Théorème A.16 (Kolmogorov-Smirnov (TLC uniforme))**.** *Si* $(X_n)_{n\in\mathbb{N}}$ *sont des v.a. i.i.d. de fonction de répartition* F *continue, et si pour tout* $n \geqslant 1$*,* F_n *désigne la fonction de répartition de la mesure empirique* $\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$*, alors pour tout* $n \geqslant 1$*, la loi de* $\|F_n - F\|_\infty$ *est libre (elle ne dépend pas de* F*), et*

$$\sqrt{n}\|F_n - F\|_\infty \xrightarrow[n\to\infty]{\text{loi}} \mathrm{KS}$$

où KS *est la loi de Kolmogorov-Smirnov, qui est la loi de* $\sup_{t\in[0,1]}|P_t|$ *où* $(P_t)_{t\in[0,1]}$ *est le pont brownien* $(P_t)_{t\in[0,1]} := (B_t - tB_1)_{t\in[0,1]}$ *où* $(B_t)_{t\in\mathbb{R}_+}$ *est un mouvement brownien standard. Enfin* KS *a pour fonction de répartition*

$$\mathrm{KS}(]-\infty,t]) = 1 + 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2t^2}, \quad t > 0.$$

Le théorème A.16 de Kolmogorov-Smirnov est au théorème A.15 de Glivenko-Cantelli ce que le TLC est à la LGN.

Principes de grandes déviations

Les principes de grandes déviations (PGD) [2] sont des raffinements asymptotiques de la LGN, qui ne sont pas de même nature que le TLC.

Théorème A.17 (Principe de grandes déviations de Cramér)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une suite de v.a. i.i.d. de moyenne* m*, telles que la transformée de Laplace de leur loi commune* $t \in \mathbb{R} \mapsto L(t) := \mathbb{E}(e^{tX_1}) \in [0,\infty]$ *est finie sur un voisinage de l'origine. Soit* Ψ *la transformée de Legendre de* $\log(L)$*, donnée par*

$$\Psi(\theta) := \sup_{t\in\mathbb{R}}(t\theta - \log(L(t))), \quad \theta \in \mathbb{R}.$$

Alors Ψ *prend ses valeurs dans* $\overline{\mathbb{R}}_+$*, est semi-continue inférieurement, convexe, décroissante sur* $]-\infty, m]$*, nulle en* m*, croissante sur* $[m,\infty[$*. De plus, si*

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

alors pour tout borélien $B \subset \mathbb{R}$*, quand* $n \gg 1$*,*

$$\mathbb{P}(S_n \in B) \approx \exp(-n \inf_B \Psi),$$

au sens où en notant $\operatorname{int}(B)$ *et* $\operatorname{adh}(B)$ *l'intérieur et l'adhérence de* B*,*

$$-\inf_{\operatorname{int}(B)} \Psi \leqslant \varliminf_{n\to\infty} \frac{\log \mathbb{P}(S_n \in B)}{n} \leqslant \varlimsup_{n\to\infty} \frac{\log \mathbb{P}(S_n \in B)}{n} \leqslant -\inf_{\operatorname{adh}(B)} \Psi.$$

Il est possible de retrouver la LGN à partir du théorème A.17 de Cramér en utilisant le lemme de Borel-Cantelli. La fonction Ψ qui intervient dans le théorème A.17 est appelée *transformées de Cramér*. En voici quelques exemples :

$\mathrm{Loi}(X_n)$	Ψ	$\{\theta\in\mathbb{R} : \Psi(\theta) < \infty\}$
$\mathrm{Ber}(p)$	$\theta\log\big(\frac{\theta}{p}\big) + (1-\theta)\log\big(\frac{1-\theta}{1-p}\big)$	$[0,p]$
$\mathrm{Poi}(\lambda)$	$\lambda - \theta + \theta\log\big(\frac{\theta}{\lambda}\big)$	$[0,\infty[$
$\mathrm{Exp}(\lambda)$	$\lambda\theta - 1 - \log(\lambda\theta)$	$]0,\infty[$
$\mathcal{N}(m,\sigma^2)$	$\frac{(\theta-m)^2}{2\sigma^2}$	$\mathbb{R}$

Soit $\mathcal{M}_1$ l'ensemble des mesures de probabilités sur $\mathbb{R}$ équipé de la topologie de la convergence étroite, qui est la convergence faible par rapport aux fonctions test continues et bornées. Cette topologie est métrisable par la distance de Fortet-Mourier [3] définie pour tous $\mu,\nu \in \mathcal{M}_1$ par

$$d_{\mathrm{FM}}(\mu,\nu) := \sup_f \left| \int f\,d\mu - \int f\,d\nu \right|$$

2. «*Large Deviations Principles (LDP)*» en anglais.
3. «*Bounded-Lipschitz distance*» en anglais.

où le supremum porte sur l'ensemble des fonctions test f continues telles que $\|f\|_\infty := \sup_x |f(x)| \leqslant 1$ et $\|f\|_{\mathrm{Lip}} := \sup_{x\neq y} |f(x)-f(y)|/|x-y| \leqslant 1$.

Pour tous $\mu,\nu \in \mathcal{F}$ on définit l'*entropie relative* de ν par rapport à ν, ou divergence de Kullback-Leibler, par

$$\mathrm{Ent}(\nu\,|\,\mu) := \begin{cases} \displaystyle\int \frac{d\nu}{d\mu}\log\frac{d\nu}{d\mu}\,d\mu & \text{si } \nu \ll \mu, \\ +\infty & \text{si } \nu \not\ll \mu. \end{cases}$$

L'inégalité de Jensen pour la fonction strictement convexe $x \in \mathbb{R}_+ \mapsto x\log(x)$ indique que $\mathrm{Ent}(\nu\,|\mu) \geqslant 0$ avec égalité si et seulement si $\mu = \nu$.

Théorème A.18 (Principe de grandes déviations de Sanov)**.** *Soit* $(X_n)_{n\in\mathbb{N}^*}$ *une suite de v.a. i.i.d. de loi* μ*, et* $\Psi := \mathrm{Ent}(\cdot\,|\,\mu)$*. Alors la mesure empirique*

$$\mu_n := \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$$

vérifie, pour tout borélien B *de* $\mathcal{M}_1$*, quand* $n \gg 1$*,*

$$\mathbb{P}(\mu_n \in B) \approx \exp(-n\inf_B \Psi)$$

au sens où

$$-\inf_{\mathrm{int}(B)}\Psi \leqslant \varliminf_{n\to\infty}\frac{\log\mathbb{P}(\mu_n\in B)}{n} \leqslant \varlimsup_{n\to\infty}\frac{\log\mathbb{P}(\mu_n\in B)}{n} \leqslant -\inf_{\mathrm{adh}(B)}\Psi$$

où $\mathrm{int}(B)$ *et* $\mathrm{adh}(B)$ *sont l'intérieur et l'adhérence de* B*.*

On peut retrouver la LGN exprimée sur les mesures empiriques à partir du théorème A.18 de Sanov en utilisant le lemme de Borel-Cantelli en prenant par exemple $A_n = \{\nu \in \mathcal{M}_1 : d_{\mathrm{FM}}(\mu,\nu) > \varepsilon\}$ pour $\varepsilon > 0$ arbitraire.

Extrêmes

Que devient le TLC si $\frac{1}{n}(x_1+\cdots+x_n)$ est remplacée par $\max(x_1,\ldots,x_n)$?

Théorème A.19 (des extrêmes de Gnedenko-Fréchet-Fisher-Tippet)**.** *Si* $(X_n)_{n\in\mathbb{N}^*}$ *sont des v.a. i.i.d. et s'il existe une suite* $(a_n)_{n\in\mathbb{N}^*}$ *dans* $]0,\infty[$ *et* $(b_n)_{n\in\mathbb{N}^*}$ *dans* $\mathbb{R}$ *et une loi* L *telles que*

$$\frac{\max(X_1,\ldots,X_n)-b_n}{a_n} \xrightarrow[n\to\infty]{\text{loi}} L,$$

alors soit L *est une masse de Dirac, soit* L *s'obtient par translation et dilation à partir de l'un des trois types de lois dont les fonctions de répartition sont de la forme*

1. *(Fréchet)* $F(t) = e^{-t^{-a}}\mathbb{1}_{t\in\mathbb{R}_+}, \quad c > 0$ *;*

2. *(Gumbel)* $F(t) = e^{-e^{-t}}$;
3. *(Weibul)* $F(t) = \mathbb{1}_{t\in\mathbb{R}_+} + e^{-(-t)^a}\mathbb{1}_{t\notin\mathbb{R}_+}, \quad c > 0.$

Si F est la fonction de répartition de X_1 alors pour tout $x \in \mathbb{R}$,

$$F^n(a_n x + b_n) \underset{n\to\infty}{\longrightarrow} L(]-\infty, x]).$$

Le bassin d'attraction de chacune des trois lois des extrêmes dépend du comportement au bord droit du support de la loi des $(X_n)_{n\in\mathbb{N}^*}$. Exemples :

Loi(X_n)	Queue	L	a_n	b_n
Cauchy	lourde	Fréchet $c = 1$	n/π	0
Exp(1)	légère	Gumbel	1	$\log(n)$
Unif$([0,1])$	nulle	Weibull $c = 1$	$1/n$	1

Les suites $(a_n)_{n\in\mathbb{N}^*}$ et $(b_n)_{n\in\mathbb{N}^*}$ ne sont pas toujours simples, comme le montre l'exemple de la loi $\mathcal{N}(0,1)$ pour lequel L est la loi de Gumbel et

$$a_n = \frac{1}{\sqrt{2\log(n)}} \quad \text{et} \quad b_n = \sqrt{2\log(n)} - \frac{\log(\log(n)) + \log(4\pi)}{2\sqrt{2\log(n)}}.$$

Espérance conditionnelle

Théorème A.20 (Espérance conditionnelle)**.** *Si $X \in L^1$ et $\mathcal{G}$ est une sous-tribu de $\mathcal{F}$ alors il existe une unique v.a. Y, notée $\mathbb{E}(X \,|\, \mathcal{G})$, telle que*
- *$Y \in L^1(\mathcal{G})$;*
- *$\mathbb{E}(YZ) = \mathbb{E}(XZ)$ pour toute v.a. Z mesurable pour $\mathcal{G}$.*

Si de plus $X \in L^2$ alors Y est la projection orthogonale de X sur $L^2(\mathcal{G})$.

On note $\mathbb{E}(X \,|\, Y) := \mathbb{E}(X \,|\, \mathcal{G})$ où $\mathcal{G}$ est la tribu engendrée par Y.

Théorème A.21 (Propriétés de l'espérance conditionnelle)**.**
- *$X \mapsto \mathbb{E}(X \,|\, \mathcal{G})$ a les propriétés d'une espérance : linéarité, positivité, normalisation, convergence monotone, inégalité de Jensen ;*
- *$\mathbb{E}(X \,|\, \mathcal{G}) = \mathbb{E}(X)$ si X est indépendante de $\mathcal{G}$;*
- *$\mathbb{E}(XZ \,|\, \mathcal{G}) = Z\mathbb{E}(X \,|\, \mathcal{G})$ si Z est mesurable pour $\mathcal{G}$;*
- *$\mathbb{E}(f(X,Z) \,|\, \mathcal{G}) = g(Z)$ où $g(z) := \mathbb{E}(f(X,z) \,|\, \mathcal{G})$ si X est indépendante de $\mathcal{G}$ et Z est mesurable pour $\mathcal{G}$;*
- *$\mathbb{E}(\mathbb{E}(X \,|\, \mathcal{G})) = \mathbb{E}(X)$, et plus généralement, si $\mathcal{G} \subset \mathcal{H}$ alors*[4]

$$\mathbb{E}(\mathbb{E}(X \,|\, \mathcal{G}) \,|\, \mathcal{H}) = \mathbb{E}(\mathbb{E}(X \,|\, \mathcal{H}) \,|\, \mathcal{G}) = \mathbb{E}(X \,|\, \mathcal{G}).$$

Théorème A.22 (Loi conditionnelle)**.** *Si $X, Y \in L^1$ alors pour tout $y \in \mathbb{R}$ il existe une mesure de probabilité μ_y notée également* Loi$(X \,|\, Y = y)$*, mesurable en y, telle que pour toute fonction $f : \mathbb{R} \to \mathbb{R}$ vérifiant $f(X) \in L^1$,*

4. La «plus petite gagne», comme pour les projections orthogonales imbriquées.

$$\mathbb{E}(f(X)\,|\,Y) = \int f(x)\mu_Y(dx).$$

Ainsi $\mathbb{E}(X\,|\,Y) = m(Y)$ *où* $m(y)$ *est la moyenne de* μ_y*, notée* $\mathbb{E}(X\,|\,Y = y)$*.*

Soit $(X_i)_{i\in I}$ une famille de variables aléatoires et $I_1, I_2 \subset I$. On dit que les familles de variables aléatoires $X_{I_1} := (X_i)_{i\in I_1}$ et $X_{I_2} := (X_i)_{i\in I_2}$ sont *conditionnellement indépendantes* par rapport à une tribu $\mathcal{G}$ lorsque

$$\mathbb{E}(f_1(X_{I_1})f_2(X_{I_2})\,|\,\mathcal{G}) = \mathbb{E}(f(X_{I_1})\,|\,\mathcal{G})\mathbb{E}(f_2(X_{I_2})\,|\,\mathcal{G})$$

pour toutes fonctions $f_1 : \mathbb{R}^{I_1} \to \mathbb{R}$ et $f_2 : \mathbb{R}^{I_2} \to \mathbb{R}$ mesurables et bornées.

Échangeabilité et théorèmes à la de Finetti

Une mesure de probabilité sur un espace produit est *échangeable* lorsqu'elle est invariante par toute permutation d'un nombre fini de coordonnées. On dit qu'une suite de variables aléatoires est échangeable lorsque sa loi est échangeable. L'échangeabilité est un affaiblissement de la notion d'indépendance.

Toute mesure de probabilité produit est échangeable. Tout mélange de mesures de probabilités produit est échangeable. Le théorème de Diaconis-Freedman (respectivement de Hewitt-Savage) fournit une réciproque : si une suite finie (respectivement infinie) est échangeable alors sa loi est proche (respectivement égale) à un mélange de mesures de probabilités produit.

Théorème A.23 (Diaconis-Freedman)**.** *Si* $(X_1, \ldots, X_n)$ *est un vecteur aléatoire de loi* L_n *échangeable, alors il existe une mesure de probabilité* P_n*, mélange de mesures de probabilité produit, telle que pour tout* $k \in \{1, \ldots, n\}$*, les lois marginales* $L_{n,k}$ *et* $P_{n,k}$ *de dimension* k *de* L_n *et* P_n *vérifient*

$$d_{\mathrm{VT}}\left(L_{n,k}, P_{n,k}\right) := \sup_{A\in\mathcal{F}} |L_{n,k}(A) - P_{n,k}(A)| \leqslant \frac{k(k-1)}{n}.$$

Comme $d_{\mathrm{VT}}(\cdot,\cdot) \in [0,1]$, le théorème A.23 de Diaconis-Freedman n'est utile que lorsque $k \leqslant \sqrt{n}$. D'autre part, pour $k = 1$, il donne $L_{n,1} = P_{n,1}$.

Théorème A.24 (Hewitt-Savage)**.** *Si* $(X_n)_{n\in\mathbb{N}}$ *est une suite de v.a. telle que la loi du vecteur* $(X_0, \ldots, X_n)$ *est échangeable pour tout* $n \in \mathbb{N}$*, alors il existe une tribu* $\mathcal{G}$ *telle que les v.a.* $(X_n)_{n\in\mathbb{N}}$ *sont indépendantes conditionnellement à* $\mathcal{G}$ *: la loi de* $(X_n)_{n\in\mathbb{N}}$ *est donc un mélange de mesures de probabilité produit.*

Si $\mathcal{F}_n$ est la tribu engendrée par les fonctions mesurables symétriques des variables aléatoires $X_0, \ldots, X_n$, alors $\mathcal{G} = \overline{\lim}_n \mathcal{F}_n := \cap_{n\geqslant 1}\sigma(\cup_{m\geqslant n}\mathcal{F}_m)$.

Le théorème A.24 est attribué à de Finetti lorsque les v.a. sont de Bernoulli.

Théorème A.25 (Loi du zéro-un de Hewitt-Savage)**.** *Si* $(X_n)_{n\in\mathbb{N}}$ *est une suite de v.a. i.i.d. et si* $A \in \sigma(X_0, X_1, \ldots)$ *est un événement invariant par toute permutation des indices à support fini, alors* $\mathbb{P}(A) \in \{0,1\}$*.*

A.3 Martingales

Cette section est consacrée aux martingales à temps discret.

Sous-martingales, martingales, et sur-martingales

Une *filtration* $(\mathcal{F}_n)_{n\in\mathbb{N}}$ est une famille croissante de sous-tribus de $\mathcal{F}$, c'est-à-dire que $\mathcal{F}_n$ est une tribu sur Ω et $\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \mathcal{F}$ pour tout $n \in \mathbb{N}$. Un processus à temps discret (suite de variables aléatoires) $(X_n)_{n\in\mathbb{N}}$ est *adapté* à une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ lorsque X_n est mesurable pour $\mathcal{F}_n$ pour tout $n \in \mathbb{N}$. La *filtration naturelle* d'un processus $(X_n)_{n\in\mathbb{N}}$ est la plus petite filtration adaptée, donnée par $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ pour tout $n \in \mathbb{N}$.

Si $(M_n)_{n\in\mathbb{N}}$ est un processus adapté à une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ et vérifie $M_n \in L^1$ pour tout $n \in \mathbb{N}$, alors on dit que $(M_n)_{n\in\mathbb{N}}$ est une...

— *martingale* lorsque tout $n \in \mathbb{N}$

$$\mathbb{E}(M_{n+1} - M_n \,|\, \mathcal{F}_n) = 0,$$

et ceci implique une constance en moyenne : $\mathbb{E}(M_n) = \mathbb{E}(M_0)$ pour tout $n \in \mathbb{N}$ (à comprendre comme une loi de conservation) ;

— *sous-martingale* si pour tout $n \in \mathbb{N}$

$$\mathbb{E}(M_{n+1} - M_n \,|\, \mathcal{F}_n) \geqslant 0,$$

et ceci implique une croissance en moyenne : $\mathbb{E}(M_{n+1}) \geqslant \mathbb{E}(M_n)$ pour tout $n \in \mathbb{N}$ (reste *en-dessous* de la limite possible) ;

— *sur-martingale* si pour tout $n \in \mathbb{N}$

$$\mathbb{E}(M_{n+1} - M_n \,|\, \mathcal{F}_n) \leqslant 0,$$

et ceci implique une décroissance en moyenne : $\mathbb{E}(M_{n+1}) \leqslant \mathbb{E}(M_n)$ pour tout $n \in \mathbb{N}$ (reste *au-dessus* de la limite possible).

Notons que M est une sous-martingale ssi $-M$ est une sur-martingale, et que M est une martingale ssi M est une sous-martingale et une sur-martingale.

Si $\Phi : \mathbb{R} \to \mathbb{R}$ est convexe et si $M = (M_n)_{n\in\mathbb{N}}$ est une martingale pour la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, telle que $\Phi(M_n) \in L^1$ pour tout $n \in \mathbb{N}$, alors $(\Phi(M_n))_{n\in\mathbb{N}}$ est une sous-martingale pour $(\mathcal{F}_n)_{n\in\mathbb{N}}$ (inégalité de Jensen !). En particulier $(|M_n|)_{n\in\mathbb{N}}$ est une sous-martingale, ainsi que $(M_n^2)_{n\in\mathbb{N}}$ lorsque M est *de carré intégrable* c'est-à-dire lorsque $M_n \in L^2$ pour tout $n \in \mathbb{N}$.

Par définition, une martingale de carré intégrable est une somme de v.a.r. conditionnellement orthogonales. Cette structure, inspirée par les sommes de v.a.r. indépendantes, est généralisée par les martingales au cadre L^1.

Temps d'arrêt et théorème d'arrêt

Une variable aléatoire T à valeurs dans $\{0, 1, 2, \ldots\} \cup \{\infty\}$ est un *temps d'arrêt* pour une filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ lorsque pour tout $n \in \mathbb{N}$, l'événement $\{T \leqslant n\} \in \mathcal{F}_n$, ou de manière équivalente, pour tout $n \in \mathbb{N}$, $\{T = n\} \in \mathcal{F}_n$. Un exemple typique de temps d'arrêt est le temps d'atteinte

$$\tau_A := \inf\{n \in \mathbb{N} : X_n \in A\}$$

d'un borélien A de $\mathbb{R}$ par un processus $(X_n)_{n\in\mathbb{N}}$ adapté pour $(\mathcal{F}_n)_{n\in\mathbb{N}}$.

Un événement $A \in \mathcal{F}$ est *antérieur* à T lorsque $A \cap \{T = n\} \in \mathcal{F}_n$ pour tout $n \in \mathbb{N}$. La tribu $\mathcal{F}_T$ engendrée par les événements antérieurs à T est appelée *tribu d'arrêt* associée à T. Le temps d'arrêt T est mesurable pour sa tribu d'arrêt $\mathcal{F}_T$. Si S et T sont des temps d'arrêt avec $S \leqslant T$ alors $\mathcal{F}_S \subset \mathcal{F}_T$.

Théorème A.26 (d'arrêt de Doob)**.** *Si $(M_n)_{n\in\mathbb{N}}$ et T sont une martingale et un temps d'arrêt pour $(\mathcal{F}_n)_{n\in\mathbb{N}}$, alors le* processus arrêté $(M_{T\wedge n})_{n\in\mathbb{N}}$ *est une martingale pour $(\mathcal{F}_n)_{n\in\mathbb{N}}$. En particulier, pour tout $n \in \mathbb{N}$,*

$$\mathbb{E}(M_{T\wedge n}) = \mathbb{E}(M_0).$$

En outre $\mathbb{E}(M_T) = \mathbb{E}(M_0)$ dès que l'une des propriétés suivantes est vérifiée :
- *T est borné c'est-à-dire que $T \leqslant c$ pour une constante $c \geqslant 0$;*
- *$\mathbb{P}(T < \infty) = 1$ et M est uniformément intégrable ;*
- *$\mathbb{E}(T) < \infty$ et M est à accroissements bornés :*

$$\sup_{n\in\mathbb{N}} \|M_{n+1} - M_n\|_\infty < \infty.$$

Dans la pratique, on invoque souvent directement le théorème de convergence monotone ou le théorème de convergence dominée pour obtenir que $\mathbb{E}(M_T) = \mathbb{E}(M_0)$ à partir de $\mathbb{E}(M_{n\wedge T}) = \mathbb{E}(M_0)$. Par ailleurs, des propriétés analogues ont lieu pour les sous-martingales (appliquer le lemme de Fatou !).

Inégalité maximale pour les sous-martingales

Un processus $X = (X_n)_{n\in\mathbb{N}}$ est *positif* lorsque $X_n \geqslant 0$ pour tout $n \in \mathbb{N}$.

Théorème A.27 (Inégalité maximale de Doob)**.** *Si $M = (M_n)_{n\in\mathbb{N}}$ est une sous-martingale positive bornée dans L^p avec $p > 1$, alors pour tout $n \geqslant 0$, la norme L^p du maximum de M sur $[0, n] \cap \mathbb{N}$ est majorée, à une constante universelle près qui ne dépend que de p, par la norme L^p au temps terminal :*

$$\mathbb{E}\Big(\Big|\max_{0\leqslant k\leqslant n} M_k\Big|^p\Big) \leqslant \Big(\frac{p}{p-1}\Big)^p \mathbb{E}(|M_n|^p).$$

En particulier, pour tout réel $r > 0$,

$$\mathbb{P}\Big(\Big|\max_{0\leqslant k\leqslant n} M_k\Big| \geqslant r\Big) \leqslant \Big(\frac{p}{p-1}\Big)^p \frac{\mathbb{E}(|M_n|^p)}{r^p}.$$

Dans le très courant cas hilbertien $p = 2$, on trouve $(p/(p-1))^p = 4$.

Théorèmes de convergence des martingales

En analyse, une suite croissante et majorée, ou décroissante et minorée, est convergente. Un analogue de ce phénomène a lieu pour les martingales.

Théorème A.28 (de Doob sur les sous-martingales bornées dans L^1)**.** *Si une sous-martingale $M = (M_n)_{n\in\mathbb{N}}$ a sa partie positive bornée dans L^1 :*

$$\sup_{n\in\mathbb{N}} \mathbb{E}(\max(0, M_n)) < \infty,$$

alors il existe une variable aléatoire notée $M_\infty \in L^1$ telle que presque sûrement

$$\lim_{n\to\infty} M_n = M_\infty.$$

En particulier, toute martingale positive, toute sous-martingale majorée par une constante, ainsi que toute sur-martingale minorée par une constante, converge presque sûrement vers une variable aléatoire intégrable.

Notons que $\mathbb{P}(M_\infty < \infty) = 1$ car $M_\infty \in L^1$. Attention, la convergence de M vers M_∞ n'a pas forcément lieu dans L^1. Le lemme de Scheffé affirme que la convergence a lieu dans L^1 si et seulement si $\lim_{n\to\infty} \mathbb{E}(|M_0|) = \mathbb{E}(|M_\infty|)$, condition qui se réduit à $\mathbb{E}(M_0) = \mathbb{E}(M_\infty)$ lorsque M est une martingale positive ! Plus généralement, une martingale bornée dans L^1 converge dans L^1 si et seulement si elle est uniformément intégrable.

Théorème A.29 (Convergence des martingales bornées dans L^p, $p > 1$)**.** *Si $M = (M_n)_{n\in\mathbb{N}}$ est une martingale bornée dans L^p avec $p > 1$, alors il existe une variable aléatoire $M_\infty \in L^p$ telle que presque sûrement et dans L^p*

$$\lim_{n\to\infty} M_n = M_\infty.$$

De plus, si $(\mathcal{F}_n)_{n\in\mathbb{N}}$ est la filtration de M alors, pour tout $n \in \mathbb{N}$,

$$M_n = \mathbb{E}(M_\infty \mid \mathcal{F}_n).$$

Le théorème A.29 est souvent utilisé dans le cas hilbertien $p = 2$. D'autre part, la convergence $M_n \to M_\infty$ dans L^p entraîne que $\mathbb{E}(M_n) = \mathbb{E}(M_\infty)$ pour tout $n \in \mathbb{N}$ et que $\lim_{n\to\infty} \mathbb{E}(|M_n|^r) = \mathbb{E}(|M_\infty|^r)$ pour tout $q \in [1, p]$.

Corollaire A.30. *Soit $(X_n)_{n\in\mathbb{N}^*}$ des v.a.r. indépendantes et centrées.*

1. *Si $\sum_{n\in\mathbb{N}^*} \mathbb{E}(X_n^2) < \infty$ alors $\sum_{n\in\mathbb{N}^*} X_n$ converge presque sûrement ;*
2. *Si $\sum_{n\in\mathbb{N}^*} \frac{1}{n^2}\mathbb{E}(X_n^2) < \infty$ alors $\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow[n\to\infty]{\text{p.s.}} 0$.*

Le second résultat du corollaire peut se déduire du premier grâce au lemme de Kronecker : si $(b_n)_{n\in\mathbb{N}^*}$ est une suite de $]0, \infty[$ croissante qui tend vers $+\infty$ et si $\sum_{n\in\mathbb{N}^*} x_n$ est une série convergente, alors

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = 0.$$

Par une transformation d'Abel, le lemme de Kronecker se déduit du lemme de Toeplitz : si $(a_n)_{n\in\mathbb{N}^*}$ est une suite de $]0,\infty[$ avec $\sum_{n\in\mathbb{N}^*} a_n = \infty$ et si $(x_n)_{n\in\mathbb{N}^*}$ est une suite de réels telle que $\lim_{n\to\infty} x_n = \ell \in \mathbb{R}$ alors

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n} a_k x_k}{\sum_{k=1}^{n} a_k} = \ell.$$

On retrouve le lemme de Cesàro lorsque $(a_n)_{n\in\mathbb{N}^*}$ est constante.

Décomposition de Doob, LGN, TLC

Cette section est consacrée aux martingales de carré intégrable, ainsi qu'à des théorèmes limite (loi des grands nombres et théorème limite central).

On dit qu'une martingale $(M_n)_{n\in\mathbb{N}}$ est une *martingale de carré intégrable* lorsque $M_n \in L^2$ pour tout $n \in \mathbb{N}$. Dans ce cas, l'inégalité de Jensen indique que $(M_n^2)_{n\in\mathbb{N}}$ est une sous-martingale.

Théorème A.31 (Décomposition de Doob)**.** *Si $M = (M_n)_{n\in\mathbb{N}}$ est une martingale pour la filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$, de carré intégrable, alors*

— *le processus $(\langle M\rangle_n)_{n\in\mathbb{N}}$ défini par $\langle M\rangle_0 := 0$ puis pour tout $n \in \mathbb{N}$ par*

$$\langle M\rangle_{n+1} := \langle M\rangle_n + \mathbb{E}((M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n),$$

est positif, croissant, et prévisible[5] *;*

— *le processus $(M_n^2 - \langle M\rangle_n)_{n\in\mathbb{N}}$ est une martingale pour $(\mathcal{F}_n)_{n\in\mathbb{N}}$;*

— *la martingale $(M_n)_{n\in\mathbb{N}}$ est bornée dans L^2 ssi $\langle M\rangle_\infty \in L^1$ où*

$$\langle M\rangle_\infty := \lim_{n\to\infty} \langle M\rangle_n = \sum_{n\in\mathbb{N}} \mathbb{E}((M_{n+1} - M_n)^2 \,|\, \mathcal{F}_n).$$

Le processus $(\langle M\rangle_n)_{n\in\mathbb{N}}$ est appelé *compensateur* de la sous-martingale $(M_n^2)_{n\in\mathbb{N}}$ ou *processuss croissant* de la martingale de carré intégrable $(M_n)_{n\in\mathbb{N}}$.

La formule $M_n^2 = (M_n^2 - \langle M\rangle_n) + \langle M\rangle_n$ pour tout $n \in \mathbb{N}$ est la *décomposition de Doob* de la sous-martingale $(M_n^2)_{n\in\mathbb{N}}$ en somme d'une martingale et d'un processus positif, croissant, et prévisible. Ce type de décomposition reste valable en général pour les sous-martingales. Cette décomposition est remarquablement explicite dans le cas des carrés de martingales de carré intégrable.

Théorème A.32 (Loi des grands nombres pour les martingales L^2)**.** *Si $(M_n)_{n\in\mathbb{N}}$ est une martingale de carré intégrable alors*

1. *sur $\{\langle M\rangle_\infty < \infty\}$, $\lim_{n\to\infty} M_n = M_\infty$ presque sûrement, où $M_\infty \in L^2$;*

---

5. C'est-à-dire que $\langle M\rangle_{n+1}$ est mesurable pour $\mathcal{F}_n$ pour tout $n \in \mathbb{N}$.

2. sur $\{\langle M\rangle_\infty = \infty\}$, $\lim_{n\to\infty} \frac{M_n}{\langle M\rangle_n} = 0$ *presque sûrement.*

Théorème A.33 (Théorème limite central pour les martingales L^2)**.** *Si* $(M_n)_{n\in\mathbb{N}}$ *est une martingale de carré intégrable, et si* $(a_n)_{n\in\mathbb{N}}$ *est une suite de de réels strictement positifs, croissante et tendant vers* ∞*, vérifiant :*

1. *(convergence du crochet) il existe un réel* $\ell \geqslant 0$ *tel que* $\frac{\langle M\rangle_n}{a_n} \xrightarrow[n\to\infty]{\mathbb{P}} \ell$ *;*
2. *(condition de Lindeberg) pour tout* $\varepsilon > 0$

$$\frac{1}{a_n}\sum_{k=1}^{n} \mathbb{E}((M_k - M_{k-1})^2 \mathbb{1}_{\{|M_k - M_{k-1}|\geqslant \varepsilon\sqrt{a_n}\}} \mid \mathcal{F}_{k-1}) \xrightarrow[n\to\infty]{\mathbb{P}} 0\ ;$$

alors[6]

$$\frac{M_n}{\sqrt{a_n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,\ell), \quad \textit{et} \quad \sqrt{a_n}\frac{M_n}{\langle M\rangle_n} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0,\ell^{-1}) \quad (\textit{si } \ell > 0).$$

A.4 Chaînes de Markov

Cette section est consacrée aux chaînes de Markov à temps discret et à espace d'états fini ou infini dénombrable.

Dans toute cette section E désigne un ensemble fini ou infini dénombrable.

Noyau de transition et récurrence aléatoire

Un *noyau de transition*[7] sur E est une application $\mathbf{P} : E \times E \to [0,1]$ telle que $\sum_{y\in E}\mathbf{P}(x,y) = 1$ pour tout $x \in E$. Une *chaîne de Markov* de *noyaux de transition* $(P_n)_{n\in\mathbb{N}}$ et de *loi initiale* μ_0 est un processus $X = (X_n)_{n\in\mathbb{N}}$ à valeurs dans E tel que $X_0 \sim \mu_0$ et, pour tout $n \in \mathbb{N}$ et tous $x_0,\ldots,x_n \in E$,

$$\mathbb{P}(X_0 = x_0,\ldots,X_n = x_n) = \mu_0(x_0)\prod_{k=0}^{n-1}\mathbf{P}_k(x_k,x_{k+1}).$$

La formule précédente montre que la loi de la chaîne (sur les trajectoires) est entièrement caractérisée par la loi initiale et les noyaux de transition. En identifiant les mesures à des vecteurs lignes et les noyaux à des matrices, la formule s'écrit de manière condensée $X_n \sim \mu_0\mathbf{P}_1\cdots\mathbf{P}_n$ pour tout $n \in \mathbb{N}$. On dit que E est l'*espace d'états* de la chaîne. La chaîne est *homogène* si $\mathbf{P}_n$ ne dépend pas de n, et *inhomogène* dans le cas contraire. Sauf mention explicite du contraire, les chaînes considérées par la suite sont toutes homogènes.

À un noyau de transition $\mathbf{P}$ sur E on peut associer un graphe complet orienté dont l'ensemble des arêtes est E et dans lequel l'arête (x,y) porte le poids $\mathbf{P}(x,y)$ pour tous $x,y \in E$. On parle de *graphe des transitions*.

6. La seconde convergence se déduit de la première et du lemme de Slutsky.
7. Ou encore *matrice de transition*, *matrice markovienne*, ou *matrice stochastique*.

Théorème A.34 (Chaîne de Markov $\Leftrightarrow$ Suite récurrente aléatoire)**.** *Soit μ_0 une loi sur E et $(U_n)_{n\in\mathbb{N}^*}$ une suite de v.a.r. i.i.d. de loi uniforme sur $[0,1]$.*

- *Soit $X_0 \sim \mu_0$ et soit $f : E \times [0,1] \to E$ une fonction quelconque. Alors la suite récurrente aléatoire*[8] [9] *$(X_n)_{n\in\mathbb{N}}$ définie par*

$$X_{n+1} = f(X_n, U_{n+1}), \quad n \in \mathbb{N},$$

est une chaîne de Markov d'espace d'états E, de loi initiale μ_0, et de noyau de transition $\mathbf{P}(x,y) := \mathbb{P}(f(x,U_1) = y)$;

- *Réciproquement, soit $\mathbf{P}$ un noyau de transition sur E. Pour tout $x \in E$, considérons une partition $[0,1] = \cup_{y\in E} I_{x,y}$ de l'intervalle $[0,1]$ telle que*

$$\mathbb{P}(U_1 \in I_{x,y}) = |I_{x,y}| = \mathbf{P}(x,y), \quad x, y \in E.$$

Soit $f : E \times [0,1] \mapsto E$ définie pour tous $x, y \in E$ et $u \in I_{x,y}$ par

$$f(x,u) := y.$$

Alors la suite récurrente aléatoire $(X_n)_{n\in\mathbb{N}}$ définie par $X_0 \sim \mu_0$ et

$$X_{n+1} = f(X_n, U_{n+1}), \quad n \in \mathbb{N},$$

est une chaîne de Markov de loi initiale μ_0 et de noyau $\mathbf{P}$.

La structure de récurrence aléatoire est facile à repérer en général dans les modèles stochastiques, et indique donc la présence d'une chaîne de Markov.

Formulation matricielle et équation d'évolution

Un noyau de transition est une matrice qui opère à gauche sur les mesures vues comme des vecteurs ligne, et à droite sur les fonctions vues comme des vecteurs colonne. Si $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov sur E de noyau $\mathbf{P}$ et de loi initiale μ_0, alors, pour toute fonction $f : E \to \mathbb{R}$, on a

$$\begin{aligned}
\mathbb{E}(f(X_1)) &= \mathbb{E}(\mathbb{E}(f(X_1) \mid X_0)) \\
&= \sum_{x\in E} \mu_0(x) \sum_{y\in E} \mathbf{P}(x,y) f(y) \\
&= \sum_{x,y\in E} \mu_0(x)\mathbf{P}(x,y) f(y) \\
&= \mu_0 \mathbf{P} f.
\end{aligned}$$

Posons $\mathbf{P}^0 := \mathbf{I}$ où $\mathbf{I}$ est la matrice identité définie par $\mathbf{I}(x,y) = \mathbb{1}_{x=y}$ pour tous $x, y \in E$. Pour tout $n \in \mathbb{N}$, on note $\mathbf{P}^n$ le noyau de transition obtenu par produit matriciel, défini par $\mathbf{P}^0 = \mathbf{I}$ et pour tout $n \in \mathbb{N}^*$ et tous $x, y \in E$ par

8. C'est aussi un processus autorégressif non-linéaire d'ordre 1.
9. C'est aussi un système dynamique avec bruit.

$$\mathbf{P}^n(x,y) := \sum_x \mathbf{P}(x_1,x_2)\cdots\mathbf{P}(x_n,x_{n+1}),$$

où la somme porte sur les chemins $(x_1,\ldots,x_{n+1})$ dans E d'extrémités $x_1 := x$ et $x_{n+1} := y$. La suite de noyaux $(\mathbf{P}^n)_{n\in\mathbb{N}}$ est un semi-groupe pour le produit matriciel. On dit qu'il s'agit d'un semi-groupe de Markov à temps discret.

De plus, pour tout $n \in \mathbb{N}$, si μ_n désigne la loi de X_n, alors :

— la suite $(\mu_n)_{n\in\mathbb{N}}$ vérifie la relation de récurrence linéaire, appelée équation d'évolution de Chapman-Kolmogorov, suivante :

$$\mu_{n+1} = \mu_n\mathbf{P} = \mu_0\mathbf{P}^{n+1}, \quad n \in \mathbb{N}\,;$$

— pour tous $x,y \in E$ et tout $n \in \mathbb{N}$,

$$\mathbb{P}(X_n = y \,|\, X_0 = x) = \mathbf{P}^n(x,y)\,;$$

— pour toute fonction $f : E \to \mathbb{R}$ et tout $n \in \mathbb{N}$,

$$\mathbb{E}(f(X_n) \,|\, X_0 = x) = (\mathbf{P}^n f)(x)\,;$$

— pour toute fonction $f : E \to \mathbb{R}$ et tout $n \in \mathbb{N}$,

$$\mathbb{E}(f(X_n) = \mu_n f = \mu_0\mathbf{P}^n f = \mu_0\mathbf{P}^n f.$$

Propriété de Markov

Conditionnellement au présent, passé et futur sont indépendants :

Théorème A.35 (Propriété de Markov)**.** *Soit $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov de noyau $\mathbf{P}$, et soit $(\mathcal{F}_n)_{n\in\mathbb{N}}$ la filtration naturelle de X. Soit $x \in E$.*

— Propriété de Markov faible. *Pour tout $m \in \mathbb{N}$, et conditionnellement à $\{X_m = x\}$, la suite $(X_{m+n})_{n\in\mathbb{N}}$ est une chaîne de Markov de loi initiale δ_x et de noyau $\mathbf{P}$, indépendante de $\mathcal{F}_m$;*

— Propriété de Markov forte. *Pour tout temps d'arrêt τ pour $\mathcal{F}$, et conditionnellement à $\{\tau < \infty, X_\tau = x\}$, la suite $(X_{\tau+n})_{n\in\mathbb{N}}$ est une chaîne de Markov de loi initiale δ_x et de noyau $\mathbf{P}$, indépendante de $\mathcal{F}_\tau$.*

La propriété de Markov faible se déduit de la forte en prenant $\tau = m$.

Récurrence, transience, et irréductibilité

Soit $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov sur E de noyau $\mathbf{P}$, et $F \subset E$. Le *temps de premier passage* en F et le *temps d'atteinte* de F sont définis par

$$T_F = \inf\{n \in \mathbb{N}^* : X_n \in F\} \quad \text{et} \quad \tau_F = \inf\{n \in \mathbb{N} : X_n \in F\}.$$

Ces variables prennent leurs valeurs dans $\{1,2,\ldots\}\cup\{\infty\}$ et $\{0,1,2,\ldots\}\cup\{\infty\}$ respectivement. Ce sont des temps d'arrêt pour la filtration naturelle de X.

Elles sont égales sur $\{X_0 \notin F\}$. Sur $\{X_0 \in F\}$, la variable T_F est également le *temps de retour* en F. Le *nombre de passages* en F est défini par

$$N_F := \sum_{n\in\mathbb{N}} \mathbb{1}_{\{X_n=x\}}.$$

Pour tout $x \in E$, on note $\mathbb{P}_x := \mathbb{P}(\cdot \mid X_0 = x)$ et $\mathbb{E}_x := \mathbb{E}(\cdot \mid X_0 = x)$, et on abrège $T_{\{x\}}, \tau_{\{x\}}, N_{\{x\}}$ en T_x, τ_x, N_x.

Théorème A.36 (Dichotomie)**.** *Pour tout $x \in E$, deux cas sont possibles :*

x *récurrent*	x *transitoire*
$\mathbb{P}_x(T_x < \infty) = 1$	$\mathbb{P}_x(T_x < \infty) < 1$
$\mathbb{P}_x(N_x = \infty) = 1$	$\mathbb{P}_x(N_x = \infty) = 0$
$\mathbb{E}_x(N_x) = \sum_{n=0}^{\infty} \mathbf{P}^n(x,x) = \infty$	$\mathbb{E}_x(N_x) = \sum_{n=0}^{\infty} \mathbf{P}^n(x,x) < \infty$

Si $\mathbf{P}$ est un noyau de transition sur E alors :

— *Communication.* Pour tous $x, y \in E$, on dit que x *conduit* à y, et on note $x \to y$ s'il existe $n \in \mathbb{N}$ tel que $\mathbf{P}^n(x,y) > 0$; on dit que x et y *communiquent*, et on note $x \leftrightarrow y$ si $x = y$ ou si à la fois $x \to y$ et $y \to x$. La relation binaire $\leftrightarrow$ est une relation d'équivalence qui partitionne E en classes d'équivalence appelées *classes de communication*.

— *Irréductibilité.* On dit que $\mathbf{P}$ ou par extension X est *irréductible* lorsqu'il existe une unique classe de communication, qui couvre donc E tout entier (tous les états communiquent entre eux).

— *Ensemble clos.* Un ensemble $F \subset E$ est *clos* lorsque la chaîne ne s'en échappe pas : $\mathbf{P}(x,F) = 1$ pour tout $x \in F$. Un état $x \in E$ est dit *absorbant* lorsque $\{x\}$ est clos, c'est-à-dire que $\mathbf{P}(x,x) = 1$. Tout ensemble clos est réunion de classes de communication closes.

Théorème A.37 (Communication)**.** *Le caractère transitoire ou récurent est constant sur les classes de communication : on parle de classes de récurrence (ou classes récurrentes) et de classes transitoires. Toute classe de récurrence est close, et toute classe close finie est une classe de récurrence.*

Les ensembles clos comme par exemple les classes de récurrence sont absorbants. Presque sûrement, une chaîne partant d'un état récurrent repasse une infinité de fois par son état initial, ne s'échappe jamais de sa classe de récurrence, et visite une infinité de fois tous les états qui constituent cette classe de récurrence. Le cas des états absorbants est singulier puisque leur classe de récurrence est réduite à eux-mêmes. Presque sûrement, une chaîne partant d'un état transitoire ne peut repasser qu'un nombre fini de fois par son état initial, et peut être capturée par une classe de récurrence ou visiter d'autres états transitoires (un nombre fini de fois pour chacun). L'ensemble des états transitoires peut être infini, et la chaîne peut ne jamais être capturée par une classe de récurrence (c'est toujours le cas lorsqu'il n'y a pas d'état récurrent). Contrairement aux classes de récurrence, les classes transitoires peuvent ne pas être closes. Des passages (à sens unique) peuvent exister entre elles, ainsi que vers les classes de récurrence. Sur le graphe des transitions, les flèches qui pénètrent dans une classe close proviennent toujours d'états transitoires.

$$\text{irréductible}\begin{cases}\text{(1) une seule classe de récurrence (noyau récurrent)}\\ \text{(2) une seule classe transitoire (noyau transitoire)}\end{cases}$$

$$\text{réductible}\begin{cases}\text{(3) plusieurs classes de récurrence et aucun état transitoire}\\ \text{(4) au moins une classe de récurrence et des états transitoires}\\ \text{(5) plusieurs classes transitoires}\end{cases}$$

L'étude des classes de récurrence des cas (3-4) se ramène au cas (1). L'étude des classes transitoires closes des cas (4-5) se ramène au cas (2).

Atteinte, absorption, harmonicité, martingales

Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov sur E, de noyau $\mathbf{P}$. Si $F \subset E$ est clos, alors le *temps d'atteinte* τ_F devient un *temps d'absorption*.

Théorème A.38 (Probabilités et temps moyens d'atteinte)**.** *Si* $(X_n)_{n\in\mathbb{N}}$ *est une chaîne de Markov sur* E *de noyau* $\mathbf{P}$*, et si* $F \subset E$*, alors :*

1. *le vecteur* $(a_F(x))_{x\in E}$ $(a_F(x))_{x\in E}$ *des* probabilités d'atteinte *défini par*

$$a_F(x) := \mathbb{P}_x(\tau_F < \infty)$$

est solution positive ou nulle minimale[10] *du système linéaire*

$$a_F(x) = \begin{cases} 1 & \textit{si } x \in F,\\ \sum_{y\in E} \mathbf{P}(x,y)a_F(y) & \textit{si } x \notin F,\end{cases}$$

(s'écrit $\mathbf{P}a_F = a_F$ *où* a_F *est un vecteur colonne lorsque* F *est clos) ;*

2. *le vecteur* $(m_F(x))_{x\in E}$ *des temps moyens d'atteinte défini par*

$$m_F(x) := \mathbb{E}_x(\tau_F)$$

est solution positive ou nulle minimale du système linéaire

$$m_F(x) = \begin{cases} 0 & \textit{si } x \in F,\\ 1 + \sum_{y\in E} \mathbf{P}(x,y)m_F(y) & \textit{si } x \notin F.\end{cases}$$

On a $a_F = 0$ et $m_F = \infty$ sur tout ensemble clos G disjoint de F. Seule la somme sur F^c intervient dans l'équation de m_F car $m_F(x) = 0$ si $x \in F$.

10. Si a est solution positive ou nulle, alors $a_F(x) \leqslant a(x)$ pour tout $x \in E$.

Théorème A.39 (Martingale et formule d'Itô discrète)**.** *Si $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov sur E de noyau $\mathbf{P}$ et de loi initiale ν, si $f : E \to \mathbb{R}$ est telle que $f(X_n) \in L^1$ pour tout $n \in \mathbb{N}$, alors la suite $(M_n)_{n\in\mathbb{N}}$ définie par*

$$M_0 := 0 \quad et \quad M_n := f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (\mathbf{L}f)(X_k), \quad n \in \mathbb{N}^*,$$

où $\mathbf{L} := \mathbf{P} - \mathbf{I}$, est une martingale adaptée à la filtration naturelle de X.

On dit que $\mathbf{L}$ est le *générateur* de X. Une fonction $f : E \to \mathbb{R}$ est *harmonique* pour $\mathbf{P}$ lorsque $\mathbf{L}f = 0$, ou de manière équivalente lorsque $\mathbf{P}f = f$. Dans ce cas, la martingale se réduit à $(f(X_n))_{n\in\mathbb{N}}$. Le vecteur a_F des probabilités d'atteinte d'un ensemble clos F est harmonique et borné.

Mesures invariantes

Une mesure μ sur E est *invariante* pour un noyau de transition $\mathbf{P}$ lorsque

$$\mu\mathbf{P} = \mu,$$

c'est-à-dire lorsque $\mu^\top$ est un vecteur propre (à coordonnées $\geqslant 0$) associé à la valeur propre 1 pour la matrice $\mathbf{P}^\top$. Une mesure de probabilité μ invariante pour $\mathbf{P}$ est toujours *stationnaire* : si $X_0 \sim \mu$ et si $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov de noyau $\mathbf{P}$ alors $X_n \sim \mu$ pour tout $n \in \mathbb{N}$, et la loi de la suite $(X_n)_{n\in\mathbb{N}}$ est stationnaire (invariante par translation). On dit aussi à ce propos que la chaîne de Markov est *à l'équilibre*. Pour une mesure de probabilité, l'invariance est une propriété de *point fixe en loi* dans la formulation récursive :

$$X \overset{\text{loi}}{=} f(X, U)$$

où $U \sim \mathrm{Unif}([0,1])$ et $X \sim \mu$ sont des v.a. indépendantes.

Mesures réversibles

Une mesure μ sur E est *réversible* pour un noyau de transition $\mathbf{P}$ lorsque les propriétés équivalentes suivantes sont réalisées :

— condition de balance détaillée [11] : pour tous $x, y \in E$,

$$\mu(x)\mathbf{P}(x,y) = \mu(y)\mathbf{P}(y,x);$$

— les cycles ont un poids qui ne dépend pas du sens de parcours : pour tout $n \in \mathbb{N}$ et tous $x_0, \ldots, x_n \in E$, avec la convention $x_{n+1} = x_0$,

$$\mu(x_0) \prod_{k=0}^{n} \mathbf{P}(x_k, x_{k+1}) = \mu(x_0) \prod_{k=0}^{n} \mathbf{P}(x_{k+1}, x_k);$$

(cette condition [12] a l'avantage de ne pas vraiment faire intervenir μ)

11. «*Detailed balance condition*» en anglais.
12. On parle parfois de critère de Kolmogorov.

— identité en loi par renversement du temps : si $X = (X_n)_{n\in\mathbb{N}}$ est une chaîne de noyau $\mathbf{P}$ et de loi initiale μ alors pour tout $n \in \mathbb{N}$,

$$(X_0, \ldots, X_n) \overset{\text{loi}}{=} (X_n, \ldots, X_0).$$

(condition valable uniquement lorsque μ est une mesure de probabilité).

Une mesure réversible est toujours invariante, mais la réciproque est fausse. Les mesures réversibles sont plus faciles à rechercher que les mesures invariantes, mais sont plus rares. La mesure uniforme sur E est…

— invariante pour $\mathbf{P}$ ssi $\mathbf{P}^\top$ est un noyau de transition [13] ;

— réversible pour $\mathbf{P}$ ssi $\mathbf{P}$ est une matrice symétrique.

Récurrence positive, récurrence nulle, et invariance

Un état x est absorbant si et seulement si la masse de Dirac δ_x est invariante. Plus généralement, pour tout état récurrent x, on introduit le nombre moyen de passages en y avant le retour en x, noté $\mu_x(y)$ et donné par

$$\mu_x(y) := \mathbb{E}_x\left(\sum_{n=0}^{T_x-1} \mathbb{1}_{\{X_n=y\}}\right).$$

On a $\mu_x(y) = \delta_x(y)$ si x est absorbant, et $\mu_x(y) > 0$ si et seulement si $x \to y$.

Théorème A.40 (Invariance et noyaux irréductibles récurrents)**.** *Si* $\mathbf{P}$ *est un noyau irréductible et récurrent sur* E*, alors pour tout* $x \in E$*,* μ_x *est une mesure invariante pour* $\mathbf{P}$*, qui charge tous les états et qui vérifie* $\mu_x(x) = 1$*.*

La formule $T_x = \sum_{y\in E}\sum_{n=0}^{T_x-1} \mathbb{1}_{\{X_n=y\}}$ donne $\mu_x(E) = \mathbb{E}_x(T_x)$. On dit qu'un état récurrent x est *récurrent positif* lorsque

$$\mathbb{E}_x(T_x) = \mu_x(E) < \infty,$$

et *récurrent nul* lorsque $\mathbb{E}_x(T_x) = \mu_x(E) = \infty$. La propriété de Markov forte entraîne que la nature de la récurrence ainsi définie est constante sur les classes de récurrence. Un état absorbant est toujours récurrent positif.

$$\text{classe de récurrence}\begin{cases}(1) \text{ la classe est récurrente positive}\\(2) \text{ la classe est récurrente nulle}\end{cases}$$

Théorème A.41 (Invariance et noyaux irréductibles)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une chaîne de Markov sur* E *de noyau* $\mathbf{P}$ *irréductible. Alors toute mesure invariante* μ *pour* $\mathbf{P}$ *charge tous les états, et vérifie* $\mu(y) \geqslant \mu(x)\mu_x(y)$ *pour tous* x *et* y *dans* E*. De plus, il y a trois cas distincts.*

1. $\mathbf{P}$ *est transitoire. Alors* $\mu(E) = \infty$ *pour toute mesure invariante* μ*,* E *est infini, et il n'y a pas de mesure de probabilité invariante ;*

13. On dit alors que $\mathbf{P}$ est doublement stochastique ou bistochastique.

2. $\mathbf{P}$ *est récurrent. Alors* μ_x *est invariante pour tout* $x \in E$ *; les mesures invariantes sont toutes proportionnelles ; et si* μ *est invariante alors* $\mu(y) = \mu(x)\mu_x(y)$ *pour tous* $x, y \in E$*. En particulier,* $\mu_x(y)\mu_y(x) = 1$.
   a) $\mathbf{P}$ *est récurrent positif. Alors il existe une unique mesure de probabilité invariante* μ *donnée par* $\mu(x) = 1/\mathbb{E}_x(T_x)$ *pour tout* x *;*
   b) $\mathbf{P}$ *est récurrent nul. Alors* $\mu(E) = \infty$ *pour toute mesure invariante* μ*,* E *est infini, et il n'y a pas de mesure de probabilité invariante.*

Le cas 2)a) du théorème A.41 est le seul cas où il existe une mesure de probabilité invariante (seul cas possible si E est fini et $\mathbf{P}$ est irréductible).

Théorème A.42 (de Perron-Frobenius dans le cas fini)**.** *Soit* $\mathbf{P}$ *un noyau de transition sur un espace d'états* E *fini. Alors il existe une classe récurrente, et toutes les classes de récurrence sont positives. De plus, si* $(\mathcal{C}_i)_{i \in I}$ *désigne la famille finie des classes de récurrence, alors chaque* $\mathcal{C}_i$ *porte une unique mesure de probabilité invariante notée* μ_i*, et l'ensemble des mesures de probabilité invariantes est l'enveloppe convexe de l'ensemble fini* $\{\mu_i : i \in I\}$*. En particulier les mesures invariantes ne chargent pas les états transitoires.*

Convergence des fréquences de passage

Si μ est une mesure sur E et si $h : E \to \mathbb{R}$ est une fonction, alors $h \in L^1(\mu)$ signifie que $\sum_{x \in E} |h(x)|\mu(x) < \infty$, et dans ce cas on a

$$\int h\, d\mu := \sum_{x \in E} h(x)\mu(x).$$

Le théorème ergodique exprime le fait que la moyenne temporelle le long des trajectoires converge vers la moyenne en espace pour la mesure invariante, quelque soit la loi initiale. Ainsi la chaîne oublie son point de départ et converge vers une quantité qui ne dépend que de son noyau de transition.

Théorème A.43 (Théorème ergodique ou loi des grands nombres)**.** *Si* $(X_n)_{n \in \mathbb{N}}$ *est une chaîne de Markov sur* E*, de noyau* $\mathbf{P}$ *irréductible récurrent positif et d'unique mesure de probabilité invariante* μ*, alors, quelque soit la loi initiale, pour toute fonction* $h : E \to \mathbb{R}$ *dans* $L^1(\mu)$*, on a*

$$\frac{h(X_0) + \cdots + h(X_{n-1})}{n} \xrightarrow[n \to \infty]{\text{p.s.}} \int_E h d\mu.$$

En particulier, pour tout $x \in E$*, la fréquence de passage en* x *avant le temps* n *converge quand* $n \to \infty$ *vers l'inverse du temps moyen de retour en* x *:*

$$\frac{\mathrm{card}\{0 \leqslant k \leqslant n-1 : X_k = x\}}{n} \xrightarrow[n \to \infty]{\text{p.s.}} \mu(x) = \frac{1}{\mathbb{E}_x(T_x)}.$$

Lorsque les lignes de $\mathbf{P}$ sont toutes égales à μ et où $X_0 \sim \mu$, alors $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires i.i.d. de loi μ et le théorème A.43 se réduit à une loi des grands nombres pour variables aléatoires i.i.d. !

Théorème A.44 (de Chacon-Ornstein)**.** *Si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov sur E, de noyau $\mathbf{P}$ irréductible récurrent, alors, pour toute mesure invariante μ, et quelque soit la loi initiale, pour tous $x, y \in E$,*

$$\frac{\mathrm{card}\{0 \leqslant k \leqslant n-1 : X_k = x\}}{\mathrm{card}\{0 \leqslant k \leqslant n-1 : X_k = y\}} \xrightarrow[n\to\infty]{\text{p.s.}} \frac{\mu(x)}{\mu(y)},$$

et, avec la convention $1/\infty = 0$ lorsque x est récurrent nul,

$$\frac{\mathrm{card}\{0 \leqslant k \leqslant n-1 : X_k = x\}}{n} \xrightarrow[n\to\infty]{\text{p.s.}} \frac{1}{\mathbb{E}_x(T_x)}.$$

Le théorème A.44 est surtout utile lorsque la chaîne est récurrente nulle.

Théorème A.45 (Théorème limite central dans le cas fini)**.** *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov d'espace d'états E fini, de noyau de transition $\mathbf{P}$ irréductible (donc récurrent positif car E est fini), et d'unique mesure invariante μ. Alors, quelque soit la loi initiale, pour toute fonction $h : E \to \mathbb{R}$,*

$$\sqrt{n}\left(\frac{h(X_0) + \cdots + h(X_{n-1})}{n} - \int h\, d\mu\right) \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, \sigma_h^2)$$

où

$$\sigma_h^2 := \int (\mathbf{P}(g^2) - (\mathbf{P}g)^2)\, d\mu,$$

et où g est la fonction de μ-moyenne nulle solution de l'équation de Poisson

$$\mathbf{L}g = h - \int h\, d\mu \quad \textit{où} \quad \mathbf{L} := \mathbf{P} - \mathbf{I}.$$

Convergence en loi des chaînes apériodiques

La *période* d'un état $x \in E$ pour un noyau $\mathbf{P}$ sur E est l'entier

$$d(x) := \mathrm{PGCD}\{n \in \mathbb{N}^* : \mathbf{P}^n(x,x) > 0\} \in \mathbb{N}^*.$$

On dit que x est *apériodique* lorsque $d(x) = 1$, ce qui est toujours le cas lorsque $\mathbf{P}^n(x,x) > 0$ pour un entier $n \in \mathbb{N}^*$. La période est une propriété de classe : elle est constante sur les classes de communication.

Théorème A.46 (Périodicité des trajectoires)**.** *Si $\mathbf{P}$ est un noyau de transition irréductible sur un espace d'états E, de période d, alors le noyau $\mathbf{P}^d$ possède d classes de communication $E_0, \ldots, E_{d-1}$, qui fournissent la partition $E = E_0 \cup \cdots \cup E_{d-1}$, et de plus, pour tous $x, y \in E$, on a $\mathbf{P}(x,y) > 0$ ssi $(x,y) \in E_i \times E_{i+1}$ pour un $i \in \{0, \ldots, d-1\}$, avec la convention $E_{d+1} := E_0$.*

Théorème A.47 (Espace d'états fini)**.** *Si $\mathbf{P}$ est un noyau irréductible sur E fini alors les propriétés suivantes sont équivalentes*

— *le noyau $\mathbf{P}$ est apériodique ;*

— *il existe* $x \in E$ *et* $n \in \mathbb{N}^*$ *tels que* $\mathbf{P}^n(x,x) > 0$ *;*
— *il existe* $n \in \mathbb{N}^*$ *tel que* $\mathbf{P}^n(x,y) > 0$ *pour tous* $x, y \in E$ *;*
— *la matrice* $\mathbf{P}$ *possède une unique valeur propre de module* 1*, qui est* 1*.*

Théorème A.48 (Convergence en loi)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une chaîne de Markov d'espace d'états* E *et de noyau* $\mathbf{P}$ *irréductible, récurrent, et apériodique.*

— *Si* $\mathbf{P}$ *est récurrent positif (toujours le cas si* E *est fini), d'unique mesure de probabilité invariante* μ*, alors quelque soit la loi initiale,*

$$X_n \xrightarrow[n\to\infty]{\text{loi}} \mu.$$

En d'autres termes, pour tous $x, y \in E$*,*

$$\mathbf{P}^n(x,y) \underset{n\to\infty}{\longrightarrow} \mu(y).$$

— *Si* $\mathbf{P}$ *est récurrent nul (*E *est donc infini), alors pour tous* $x, y \in E$*,*

$$\lim_{n\to\infty} \mathbf{P}^n(x,y) = 0.$$

Théorème A.49 (Critère de Doeblin pour convergence géométrique)**.** *Soit* $(X_n)_{n\in\mathbb{N}}$ *une chaîne de Markov sur* E *de noyau* $\mathbf{P}$*. S'il existe une mesure de probabilité* ν *sur* E*, un réel* $\alpha > 0$*, et* $m \in \mathbb{N}^*$ *tels que pour tous* $x, y \in E$*,*

$$\mathbf{P}^m(x,y) \geqslant \alpha\nu(y),$$

alors le noyau $\mathbf{P}$ *est irréductible, récurrent positif, et il existe des constantes* $c > 0$ *et* $\rho \in]0,1[$ *telles que pour tout* $n \in \mathbb{N}$ *et tout* $x \in E$*, en notant* μ *l'unique mesure de probabilité invariante de* $\mathbf{P}$*,*

$$d_{\mathrm{VT}}\left(\mathrm{Loi}(X_n \mid X_0 = x), \mu\right) := \frac{1}{2}\sum_{y\in E} |\mathbf{P}^n(x,y) - \mu(y)| \leqslant c\rho^n.$$

La condition du théorème A.49 exprime le fait que les lignes de la matrice $\mathbf{P}$ sont toutes minorables par une mesure de référence $\alpha\nu$. Cela est toujours vrai par exemple lorsque E est fini et $\min_{x,y\in E} \mathbf{P}(x,y) > 0$.

Théorème A.50 (Critère de Foster-Lyapunov)**.** *Si* $\mathbf{P}$ *est un noyau de transition irréductible et apériodique d'espace d'états au plus dénombrable* E *et s'il existe une fonction (dite de Lyapunov)* $V : E \to \mathbb{R} \cup \{-\infty\}$*, un ensemble fini* $F \subset E$*, et une constante* $c > 0$ *tels que* $V \in L^1(\mathbf{P}(x,\cdot))$ *pour tout* $x \in F$ *et*

$$(\mathbf{P} - \mathbf{I})V(x) := \sum_{y\in E} \mathbf{P}(x,y)V(y) - V(x) \leqslant -c$$

pour tout $x \notin F$*, alors* $\mathbf{P}$ *est récurrent positif.*

Le critère de Foster-Lyapunov exploite la présence éventuelle d'une force de rappel vers un ensemble fini, en d'autres termes une sorte de tension.

A.5 Processus de Poisson

Soit $(E_n)_{n\in\mathbb{N}^*}$ des v.a. i.i.d. de loi $\mathrm{Exp}(\lambda)$, où $\lambda > 0$. Soit

$$T_0 := 0 \quad \text{et} \quad T_n := T_{n-1} + E_n = E_1 + \cdots + E_n,\ n \in \mathbb{N}^*.$$

Presque sûrement la suite $(T_n)_{n\in\mathbb{N}}$ est strictement croissante et

$$\sup_{n\in\mathbb{N}} T_n = \lim_{n\to\infty} T_n = \sum_{n=1}^{\infty} E_n = \infty.$$

Le processus de Poisson $N = (N_t)_{t\in\mathbb{R}_+}$ est défini pour tout $t \in \mathbb{R}_+$ par

$$N_t := \sum_{n=1}^{\infty} \mathbb{1}_{\{T_n\in[0,t]\}} = \sup\{n \in \mathbb{N} : T_n \leqslant t\} \in \mathbb{N}.$$

C'est le processus de comptage de tops survenant aux temps $(T_n)_{n\in\mathbb{N}^*}$, espacés par des durées $(E_n)_{n\in\mathbb{N}^*} := (T_n - T_{n-1})_{n\in\mathbb{N}^*}$. Il est issu de 0 car $N_0 = 0$, et d'intensité λ car $\mathbb{E}(T_n) = n/\lambda$ (les sauts de N sont d'autant plus rapides que λ est grand). Presque sûrement la trajectoire $t \mapsto N_t$ est issue de 0, constante par morceaux, croissante et continue à droite, avec des sauts de $+1$.

Théorème A.51 (Processus de Poisson)**.** *Si $N := (N_t)_{t\in\mathbb{R}_+}$ est un processus à valeurs dans $\mathbb{N}$ avec $N_0 = 0$ et avec des trajectoires $t \mapsto N_t$ presque sûrement croissantes et continues à droite*[14]*, alors pour tout réel $\lambda > 0$, les propriétés suivantes sont équivalentes (avec la convention $T_0 := 0$) :*

1. Comptage. *N est le processus de comptage de tops $(T_n)_{n\in\mathbb{N}^*}$ espacés par des durées $(E_n)_{n\in\mathbb{N}^*} := (T_n - T_{n-1})_{n\in\mathbb{N}^*}$ i.i.d. de loi* $\mathrm{Exp}(\lambda)$ *;*
2. Sauts. *Les temps inter-sauts $(E_n)_{n\in\mathbb{N}^*} := (T_n - T_{n-1})_{n\in\mathbb{N}^*}$ de N sont i.i.d. de loi* $\mathrm{Exp}(\lambda)$ *et les sauts sont d'amplitude* $+1$ *;*
3. Accroissements. *Pour tous $0 \leqslant t_0 \leqslant t_1 \leqslant \cdots \leqslant t_n$, les accroissements $N_{t_n} - N_{t_{n-1}}, \ldots, N_{t_1} - N_{t_0}$ de N sont indépendants et de lois de Poisson* $\mathrm{Poi}(\lambda(t_n - t_{n-1})), \ldots, \mathrm{Poi}(\lambda(t_1 - t_0))$ *respectivement ;*
4. Infinitésimal. *Les accroissements de N sont indépendants et*

$$\sup_{t\in\mathbb{R}_+} \mathbb{P}(N_{t+\varepsilon} - N_t = 0) = 1 - \lambda\varepsilon + o(\varepsilon)$$

et

$$\sup_{t\in\mathbb{R}_+} \mathbb{P}(N_{t+\varepsilon} - N_t = 1) = \lambda\varepsilon + o(\varepsilon).$$

Notons que pour tout $n \in \mathbb{N}$ et tout $t \in \mathbb{R}_+$,

$$N_{T_n} = n \quad \text{et} \quad T_{N_t} \leqslant t < T_{N_t+1}.$$

14. donc avec limites à gauche, on dit «càdlàg», même en anglais.

Pour tout $n \in \mathbb{N}^*$ et tout $t \in \mathbb{R}_+$,

$$T_n = E_1 + \cdots + E_n \sim \mathrm{Gamma}(n, \lambda) \quad \text{et} \quad N_t \sim \mathrm{Poi}(\lambda t).$$

On dit que N a des *accroissements stationnaires* car leurs loi ne dépend que des écarts de temps. Conditionnellement à l'événement $\{T_{n+1} = t\}$, les coordonnées du vecteur aléatoire $(T_1, \ldots, T_n)$ sont distribuées comme des v.a. i.i.d. de loi uniforme sur $[0, 1]$. Idem pour l'événement plus gros $\{N_t = n\}$.

Théorème A.52 (Propriété de Markov)**.** *Si $N = (N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson issu de 0 d'intensité λ, alors :*

— P. de Markov faible. *pour tout $s \in \mathbb{R}_+$, $(N_{s+t} - N_s)_{t\in\mathbb{R}_+}$ est un processus de Poisson issu de 0 et d'intensité λ, indépendant de $(N_u)_{u\in[0,s]}$;*

— P. de Markov forte. *pour tout temps d'arrêt S pour la filtration naturelle de N, conditionnellement à $\{S < \infty\}$, le processus $(N_{S+t} - N_S)_{t\in\mathbb{R}_+}$ est un processus de Poisson issu de 0 et d'intensité λ, indépendant de la tribu des événements antérieurs $\mathcal{F}_S$.*

Théorème A.53 (Stabilité par superposition et par amincissement[15])**.**

— Superposition. *Si $(N_t)_{t\in\mathbb{R}_+}$ et $(N'_t)_{t\in\mathbb{R}_+}$ sont des processus de Poisson indépendants issus de 0 et d'intensités λ et λ', alors $(N_t + N'_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson simple issu de 0 d'intensité $\lambda + \lambda'$;*

— Amincissement. *Si $(N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson issu de 0 et d'intensité λ, et de temps de sauts $(T_n)_{n\in\mathbb{N}^*}$, et si $(Y_n)_{n\in\mathbb{N}^*}$ est une suite de variables aléatoires i.i.d. de loi de Bernoulli $\mathrm{Ber}(p)$, indépendante de $(T_n)_{n\in\mathbb{N}^*}$, alors les processus de comptage des suites amincies*

$$(T_n)_{n\in\mathbb{N}^*:Y_n=1} \quad \textit{et} \quad (T_n)_{n\in\mathbb{N}^*:Y_n=0}$$

sont des processus de Poisson indépendants issus de 0 et d'intensités

$$p\lambda \quad \textit{et} \quad (1-p)\lambda.$$

Le n^{e} temps de saut du premier processus aminci s'écrit $T'_n := E_1 + \cdots + E_G$ où $G \sim \mathrm{Geo}(p)$ est indépendante de $(T_n)_{n\in\mathbb{N}^*}$, d'où[16] $T'_n \sim \mathrm{Exp}(\lambda p)$.

Pour tous $s < t$ la variable aléatoire $N_t - N_s$ donne le nombre de tops (sauts du processus) dans l'intervalle $]s, t] = [0, t] \setminus [0, s]$. Pour tout borélien $A \subset \mathbb{R}_+$, on note $N(A)$ le nombre de tops dans A. C'est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{\infty\}$. C'est la mesure de A pour la mesure ponctuelle aléatoire $\sum_{n\in\mathbb{N}^*} \delta_{T_n}$. On note $|A|$ la mesure de Lebesgue de A.

Théorème A.54 (Structure par localisation)**.** *Soit $(N_t)_{t\geqslant 0}$ un processus de Poisson issu de 0 et d'intensité λ. Les deux propriétés suivantes ont lieu :*

— *si $I \subset \mathbb{R}_+$ est un intervalle borné non-vide alors $N(I) \sim \mathrm{Poi}(\lambda|I|)$, et pour tout $n \in \mathbb{N}^*$, conditionnellement à $\{N(I) = n\}$, les n tops dans I suivent la même loi que n v.a.r. i.i.d. de loi uniforme sur I ;*

15. « *Thinning* » en anglais.
16. Mélange géométrique de lois Gamma !

— *si* $I_1, \dots, I_d$ *sont des intervalles bornés et disjoints de* $\mathbb{R}_+$*, alors les composantes du vecteur aléatoire* $C = (N(I_1), \dots, N(I_d))$ *sont indépendantes de lois* $\mathrm{Poi}(\lambda|I_1|), \dots, \mathrm{Poi}(\lambda|I_d|)$*, et de plus, pour tout* $n \in \mathbb{N}^*$*, en notant* $I := I_1 \cup \dots \cup I_d$ *et* $(p_1, \dots, p_d) := (|I_1|/|I|, \dots, |I_d|/|I|)$*,*

$$\mathrm{Loi}(C \mid N(I) = n) = \mathrm{Mul}(n, (p_1, \dots, p_d)).$$

Cette loi conditionnelle ne dépend pas de l'intensité λ*.*

Théorème A.55 (Loi des grands nombres et théorème limite central)**.** *Si* $(N_t)_{t \in \mathbb{R}_+}$ *est un processus de Poisson issu de* 0 *et d'intensité* λ*, alors*

$$\frac{N_t}{t} \xrightarrow[t \to \infty]{\text{p.s.}} \lambda \quad \text{et} \quad \sqrt{t}\left(\frac{N_t}{t} - \lambda\right) \xrightarrow[t \to \infty]{\text{loi}} \mathcal{N}(0, \lambda).$$

Certaines propriétés asymptotiques subsistent pour les processus de comptage de tops espacés par des durées avec mémoire (non-exponentielles) :

Théorème A.56 (Renouvellement[17])**.** *Soit* $N = (N_t)_{t \in \mathbb{R}_+}$ *le processus de comptage issu de* 1 *d'une suite de tops espacés par des durées aléatoires* $(E_n)_{n \in \mathbb{N}^*}$ *i.i.d. de loi* ν *sur* $\mathbb{R}_+$ *possédant une moyenne* $m > 0$ *:*

$$N_t := \sum_{n=0}^{\infty} \mathbb{1}_{\{T_n \leqslant t\}}, \quad T_n := T_{n-1} + E_n, \quad T_0 := 0.$$

Alors $m\mathbb{E}(N_t) \geqslant t$ *pour tout* $t \in \mathbb{R}_+$*, et*

$$\lim_{t \to \infty} \frac{\mathbb{E}(N_t)}{t} = \frac{1}{m}.$$

De plus, si ν *possède une variance* $\sigma^2 < \infty$ *alors*

$$\lim_{t \to \infty} \frac{\mathrm{Var}(N_t)}{t} = \frac{\sigma^2}{m^3}, \quad \text{et} \quad \sqrt{t}\left(\frac{N_t}{t} - \frac{1}{m}\right) \xrightarrow[t \to \infty]{\text{loi}} \mathcal{N}\left(0, \frac{\sigma^2}{m^3}\right).$$

D'autre part, la mesure $R := \sum_{n=0}^{\infty} \nu^{*n} = \delta_0 + \nu + \nu * \nu + \cdots$ *est solution de l'équation (dite de renouvellement)* $\delta_0 + \nu * R = R$*. Enfin, si le support de* ν *n'est pas arithmétique*[18]*, c'est-à-dire que* $\nu(r\mathbb{Z}) < 1$ *pour tout* $r \in \mathbb{Z}$*, alors*

$$\lim_{t \to \infty} \mathbb{E}(N_t) - \frac{t}{m} = \frac{\sigma^2 + m^2}{2m^2},$$

et les formules (dites de Blackwell) suivantes ont lieu : pour tous $s, t \in \mathbb{R}_+$*,*

$$\mathbb{E}(N_t) = R([0,t]) \quad \text{et} \quad \lim_{t \to \infty} R(]t, t+s]) = \lim_{t \to \infty} \mathbb{E}(N_{t+s} - N_t) = \frac{s}{m}.$$

17. «*Renewal*» en anglais.
18. En anglais on dit en général «*non-lattice*».

Processus de Bernoulli	**Processus de Poisson**
Espace d'état $\mathbb{N}$ et loi initiale δ_0	
Processus $(B_n)_{n\in\mathbb{N}}$ avec $B_0 = 0$	Processus $(N_t)_{t\in\mathbb{R}_+}$ avec $N_0 = 0$
Temps discret $n \in \mathbb{N}$	Temps continu $t \in \mathbb{R}_+$
Paramètre $p \in [0,1]$	Paramètre $\lambda \in \mathbb{R}_+$
$T_0 = 0$ et $T_m = E_1 + \cdots + E_m$	
$(E_m)_{m\in\mathbb{N}^*}$ i.i.d. Geo(p)	$(E_m)_{m\in\mathbb{N}^*}$ i.i.d. Exp(λ)
$B_n = \sup\{m \in \mathbb{N}; T_m \leqslant n\}$	$N_t = \sup\{m \in \mathbb{N}; T_m \leqslant t\}$
$T_m \sim \mathrm{Geo}(p)^{*m} = \mathrm{NegBin}(m,p)$	$T_m \sim \mathrm{Exp}(\lambda)^{*m} = \mathrm{Gamma}(n,\lambda)$
$B_n \sim \mathrm{Bin}(n,p)$	$N_t \sim \mathrm{Poi}(\lambda t)$
$\mathbb{E}(B_n) = np$	$\mathbb{E}(N_t) = \lambda t$
$B_n = B_{n_k} - B_{n_{k-1}} + \cdots + B_{n_1} - B_{n_0}$	$N_t = N_{t_k} - N_{t_{k-1}} + \cdots + N_{t_1} - N_{t_0}$
$0 \leqslant n_0 \leqslant n_1 \leqslant \cdots \leqslant n_k = n$	$0 \leqslant t_0 \leqslant t_1 \leqslant \cdots \leqslant t_k = t$
$(B_{n_k} - B_{n_{k-1}})_{k\in\mathbb{N}}$ indépendantes	$(N_{t_k} - N_{t_{k-1}})_{k\in\mathbb{N}}$ indépendantes
$B_{n_k} - B_{n_{k-1}} \sim \mathrm{Bin}(n_k - n_{k-1}, p)$	$N_{t_k} - N_{t_{k-1}} \sim \mathrm{Poi}(\lambda(t_k - t_{k-1}))$
$\lim_{n\to\infty} \mathrm{Bin}(\lfloor tk \rfloor, p_k) = \mathrm{Poi}(\lambda t)$ quand $\lim_{k\to\infty} kp_k = \lambda$	

Tableau A.1. Tableau analogique entre processus de Bernoulli et de Poisson.

Théorème A.57 (Processus de Poisson et de Bernoulli)**.** *Soit $\lambda > 0$ un réel et $(p_k)_{k>0}$ une suite de réels dans $[0,1]$ telle que $\lim_{k\to\infty} kp_k = \lambda$. Pour tout entier $k > 0$, soit $(B_n^k)_{n\in\mathbb{N}}$ un processus de Bernoulli issu de 0 et de paramètre p_k. Alors pour tous $0 \leqslant t_1 \leqslant \cdots \leqslant t_n$, la suite de vecteurs aléatoires $((B^k_{\lfloor t_1 k\rfloor}, \ldots, B^k_{\lfloor t_n k\rfloor}))_{k>0}$ converge en loi vers la loi de $(N_{t_1}, \ldots, N_{t_n})$, où $(N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson simple issu de 0 et d'intensité λ.*

Théorème A.58 (Paradoxe de l'autobus ou de l'inspection)**.** *Si $(N_t)_{t\in\mathbb{R}_+}$ est un processus de Poisson issu de 0 et d'intensité λ, de temps de saut $(T_n)_{n\in\mathbb{N}^*}$, et si $T_0 := 0$, alors pour tout $t \in \mathbb{R}_+$, on a*

$$T_{N_t} \leqslant t < T_{N_t+1},$$

et si $U_t := t - T_{N_t}$ et $V_t := T_{N_t+1} - t$ sont les longueurs dans la dichotomie

$$[T_{N_t}, T_{N_t+1}] = [T_{N_t}, t] \cup [t, T_{N_t+1}],$$

alors les variables aléatoires U_t et V_t sont indépendantes, avec $V_t \sim \mathrm{Exp}(\lambda)$ et U_t de même loi que $\min(E,t)$ où $E \sim \mathrm{Exp}(\lambda)$. En particulier

$$\mathbb{E}(T_{N_t+1} - T_{N_t}) = \mathbb{E}(U_t) + \mathbb{E}(V_t) = \frac{1-e^{-\lambda t}}{\lambda} + \frac{1}{\lambda} > \frac{1}{\lambda}.$$

En moyenne, l'intervalle de temps $[T_{N_t}, T_{N_t+1}]$ délimité par les deux tops encadrant un temps déterministe fixé t est plus grand que la durée moyenne $1/\lambda$ séparant deux tops consécutifs. Le rapport croît avec t et tend 2 lorsque $t \to \infty$. Cependant, $\lim_{t\to 0} \mathbb{E}(T_{N_t+1} - T_{N_t}) = 1/\lambda$. Si les tops représentent l'arrivée d'un autobus à la station, l'écart moyen entre le dernier passage avant le temps t et le prochain passage après le temps t est plus grand que la durée moyenne des passages des autobus à la station. Idem si les tops représentent les durées de vie de composants et t un temps d'inspection du composant.

Théorème A.59 (Chaîne de Markov poissonisée)**.** *Soit $Z := (Z_n)_{n\in\mathbb{N}}$ une chaîne de Markov d'espace d'état au plus dénombrable E et de noyau de transition $\mathbf{P}$, et soit $N := (N_t)_{t\in\mathbb{R}_+}$ un processus de Poisson issu de 0 et d'intensité λ, de temps de saut $(T_n)_{n\in\mathbb{N}^*}$, indépendant de Z. Alors le processus à temps continu $X := (Z_{N_t})_{t\in\mathbb{R}_+}$ à valeurs dans E vérifie, pour tous $n \in \mathbb{N}$ et $t \in \mathbb{R}_+$,*

$$Z_{N_t} = Z_n \quad \textit{si } T_n \leqslant t < T_{n+1}$$

avec $T_0 := 0$, et pour tous $x, y \in E$,

$$\mathbb{P}(Z_{N_t} = y \,|\, Z_0 = x) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \mathbf{P}^n(x, y).$$

De plus, les temps de saut $(S_n)_{n\in\mathbb{N}^*}$ de X définis pour tout $n \in \mathbb{N}^*$ par

$$S_n := \inf\{t > S_{n-1} : X_t \neq X_{S_{n-1}}\},$$

avec la convention $S_0 := 0$, sont séparés par des durées $(S_n - S_{n-1})_{n\in\mathbb{N}^*}$ aléatoires indépendantes de loi exponentielles[19]*.*

Si Z est une marche aléatoire, alors $Z_n = Z_0 + \xi_1 + \cdots + \xi_n$, et donc $X_t = Z_0 + \sum_{n=1}^{N_t} \xi_n$, et on dit que X est un *processus de Poisson composé.*

A.6 Chaînes de Markov à temps continu

Dans cette section on s'intéresse à des processus $(X_t)_{t\in\mathbb{R}_+}$ indexés par un temps continu $t \in \mathbb{R}_+$ et à valeurs dans un espace discret au plus dénombrable. Ces processus sont obtenus en insérant des durées aléatoires exponentielles entre les actions successives d'un noyau de transition. Ces temps exponentiels peuvent dépendre de la position spatiale, ce qui va donc plus loin que la simple poissonisation d'une chaîne à temps discret abordée dans le théorème A.59.

On considère dans cette section les ingrédients suivants :

— une chaîne de Markov à temps discret $Z := (Z_n)_{n\in\mathbb{N}}$ de noyau de transition $\mathbf{P}$ sur un espace d'état au plus dénombrable E ;

— une fonction $\boldsymbol{\lambda} : E \to]0, \infty[$ appelée *fonction de temporisation* ;

— des v.a. $(E_n^{(z)})_{z\in E, n\in\mathbb{N}^*}$ indépendantes, indépendantes de Z, telles que $((E^{(z)})_n)_{n\in\mathbb{N}}$ sont i.i.d. de loi $\mathrm{Exp}(\boldsymbol{\lambda}(z))$ pour tout $z \in E$.

On construit la suite $(T_n)_{n\in\mathbb{N}}$ en posant

$$T_0 := 0 \quad \text{et} \quad T_n := E_1^{(Z_0)} + \cdots + E_n^{(Z_{n-1})}, \; n \in \mathbb{N}^*.$$

19. Mélange géométrique de lois Gamma !

Presque sûrement, la suite $(T_n)_{n\in\mathbb{N}}$ est strictement croissante, mais rien ne garantit qu'elle tend vers l'infini. On définit le *temps d'explosion* T_∞ par

$$T_\infty := \sup_{n\in\mathbb{N}} T_n = \lim_{n\to\infty} T_n = \sum_{n=1}^{\infty} E_n^{(Z_{n-1})} \in \overline{\mathbb{R}}_+.$$

Il a *explosion* lorsque $\mathbb{P}(T_\infty = \infty) < 1$. Pour tout $t \in [0, T_\infty[$, soit

$$N_t := \sup\{n \in \mathbb{N} : T_n \leqslant t\} \quad \text{et} \quad X_t := Z_{N_t}.$$

On a $N_t = n$ et $X_t = Z_n$ sur $\{T_n \leqslant t < T_{n+1}\}$. La trajectoire $t \mapsto N_t$ et donc la trajectoire $t \mapsto X_t$ est définie presque sûrement pour tout temps $t \in \mathbb{R}_+$ si et seulement si il n'y pas explosion :

$$\mathbb{P}(T_\infty = \infty) = 1.$$

On dit que $X := (X_t)_{t\in\mathbb{R}_+}$ est une *chaîne de Markov à temps continu* de noyau $\mathbf{P}$ et de temporisation $\boldsymbol{\lambda}$. Lorsque la temporisation $\boldsymbol{\lambda}$ est constante, alors N est un processus de Poisson indépendant de Z, il n'y a pas explosion, et X est une chaîne poissonisée comme dans le théorème A.59.

Théorème A.60 (Critère de non explosion)**.** *Le temps d'explosion T_∞ vérifie*

$$\mathbb{P}(T_\infty = \infty) = \mathbb{P}\left(\sum_{n=0}^{\infty} \frac{1}{\boldsymbol{\lambda}(Z_n)} = \infty\right).$$

En particulier $\mathbb{P}(T_\infty = \infty) = 1$ lorsque l'une des propriétés suivantes a lieu :
— *la fonction de temporisation $\boldsymbol{\lambda}$ est bornée (toujours le cas si E fini !) ;*
— *la loi initiale de Z ne charge pas d'état transitoire de $\mathbf{P}$;*
— *le noyau $\mathbf{P}$ de Z est récurrent.*

Dans toute la suite on suppose qu'il n'y a pas explosion :

$$\mathbb{P}(T_\infty = \infty) = 1.$$

Le *générateur infinitésimal*[20] $\mathbf{L} : E \times E \to \mathbb{R}$ est défini pour $x, y \in E$ par

$$\mathbf{L}(x,y) = \begin{cases} \boldsymbol{\lambda}(x)\mathbf{P}(x,y) & \text{si } x \neq y, \\ \boldsymbol{\lambda}(x)(\mathbf{P}(x,x) - 1) & \text{si } x = y, \end{cases}$$

c'est-à-dire qu'en terme de matrices,

$$\mathbf{L} = \mathbf{D}_{\boldsymbol{\lambda}}(\mathbf{P} - \mathbf{I}) \quad \text{où} \quad \mathbf{D}_{\boldsymbol{\lambda}} := \mathrm{Diag}((\boldsymbol{\lambda}(x))_{x\in E}).$$

Le générateur $\mathbf{L}$ vérifie les propriétés suivantes :
— $\mathbf{L}(x,x) \leqslant 0$ pour tout x dans E ;

20. On parle aussi de *Q-matrice*.

— $\mathbf{L}(x,y) \geqslant 0$ pour tous x et y dans E avec $x \neq y$;
— $\sum_{y\in E} \mathbf{L}(x,y) = 0$ pour tout x dans E.

Théorème A.61 (Le générateur fait la loi)**.** *Soit* $(S_n)_{n\in\mathbb{N}^*}$ *la suite des temps de saut de* X *à valeurs dans* $\overline{\mathbb{R}}_+$ *définie par*

$$S_0 := 0 \quad \textit{et} \quad S_n = \inf\{t > S_{n-1}; X_t \neq X_{S_{n-1}}\},\ n \in \mathbb{N}^*.$$

Alors $(S_n)_{n\in\mathbb{N}}$ *est strictement croissante, puis éventuellement constante et égale à* ∞*, et converge presque sûrement vers* ∞*. De plus, pour tous* $x \in E$ *et* $n \in \mathbb{N}$*, conditionnellement à* $\{X_{S_n} = x\}$*,*

1. *si* $\mathbf{L}(x,x) = 0$*, alors* $S_{n+1} = \infty$*, et on dit que* x *est* absorbant *;*
2. *si* $\mathbf{L}(x,x) \neq 0$*, alors* S_{n+1} *est presque sûrement finie, les variables aléatoires* $S_{n+1} - S_n$ *et* $X_{S_{n+1}}$ *sont indépendantes, et pour tout* y *dans* E*,*

$$\mathrm{Loi}(S_{n+1} - S_n \,|\, X_{S_n} = x) = \mathrm{Exp}(-\mathbf{L}(x,x))$$

et

$$\mathrm{Loi}(X_{S_{n+1}} = y \,|\, X_{S_n} = x) = -\frac{\mathbf{L}(x,y)}{\mathbf{L}(x,x)}.$$

L'écriture de $\mathbf{L}$ sous la forme $\mathbf{L} = \mathbf{D}_{\boldsymbol{\lambda}}(\mathbf{P} - \mathbf{I})$ n'est pas unique. Plusieurs choix sont possibles. Il est commode de prendre pour $\boldsymbol{\lambda}$ l'opposé en signe de la diagonale de $\mathbf{L}$, et pour $\mathbf{P}$ le noyau donné pour tous x et y dans E par

$$\mathbf{P}(x,y) = \begin{cases} 0 & \text{si } x = y \text{ et } \mathbf{L}(x,x) < 0, \\ 1 & \text{si } x = y \text{ et } \mathbf{L}(x,x) = 0, \\ -\frac{\mathbf{L}(x,y)}{\mathbf{L}(x,x)} & \text{si } x \neq y \text{ et } \mathbf{L}(x,x) < 0, \\ 0 & \text{si } x \neq y \text{ et } \mathbf{L}(x,x) = 0. \end{cases}$$

Ce noyau est appelé *noyau inclus*. La *chaîne incluse* saute à chaque étape vers un état différent, sauf bien sûr si l'état courant est absorbant. La diagonale de $\mathbf{P}$ est constituée de 0 et de 1 (ces derniers correspondent aux états absorbants). Avec cette décomposition les temps $(T_n)_{n\in\mathbb{N}^*}$ liés à $\boldsymbol{\lambda}$ dans la construction de $(X_t)_{t\in\mathbb{R}_+}$ coïncident avec les temps de saut $(S_n)_{n\in\mathbb{N}^*}$ de $(X_t)_{t\in\mathbb{R}_+}$.

Une autre paramétrisation de $\mathbf{L}$ sous la forme $\mathbf{L} = \mathbf{D}_{\boldsymbol{\lambda}}(\mathbf{P} - \mathbf{I})$ s'obtient en faisant en sorte que $\boldsymbol{\lambda}$ soit constant et égal à $\sup_x -\mathbf{L}(x,x)$. Cela n'est possible que si ce supremum est fini. Le noyau $\mathbf{P}$ associé est appelé *noyau harmonisé*, et la chaîne associée est appelée *chaîne harmonisée*.

Théorème A.62 (Semi-groupe de Markov)**.** *Pour tout* $t \in \mathbb{R}_+$ *on définit le noyau de transition* $\mathbf{P}_t$ *en posant, pour tous* $x, y \in E$*,*

$$\mathbf{P}_t(x,y) := \mathbb{P}(X_t = y \,|\, X_0 = x).$$

Alors $(\mathbf{P}_t)_{t\in\mathbb{R}_+}$ *est un semi-groupe :*

$$\mathbf{P}_0 = \mathbf{I} \quad et \quad \mathbf{P}_{s+t} = \mathbf{P}_s\mathbf{P}_t, \ s,t \in \mathbb{R}_+.$$

De plus la fonction $t \in \mathbb{R}_+ \mapsto \mathbf{P}_t(x,y)$ *est différentiable et constitue la solution positive minimale des deux équations de Chapman-Kolmogorov suivantes :*

$$\frac{d}{dt}\mathbf{P}_t(x,y) = (\mathbf{L}\mathbf{P}_t)(x,y) \quad et \quad \frac{d}{dt}\mathbf{P}_t(x,y) = (\mathbf{P}_t\mathbf{L})(x,y)$$

On a également

$$\lim_{t\to 0^+} \frac{\mathbf{P}_t(x,y) - \mathbf{P}_0(x,y)}{t} = \mathbf{L}(x,y).$$

Enfin, si E *est fini alors* $\mathbf{P}_t = \exp(t\mathbf{L})$ *où* exp *est l'exponentielle de matrice.*

Nous avons ainsi une *description triple* de la loi de X :
— par la donnée de $(\mathbf{P}_t)_{t\in\mathbb{R}_+}$;
— par la donnée de $\mathbf{L}$;
— par la donnée de $(\boldsymbol{\lambda}, \mathbf{P})$.
Ces trois objets sont liés par $\partial_{t=0}\mathbf{P}_t = \mathbf{L} = \mathbf{D}_{\boldsymbol{\lambda}}(\mathbf{P} - \mathbf{I})$. Les deux dernières descriptions permettent de simuler directement les trajectoires du processus, qui sont constantes par morceaux, continues à droite avec limites à gauche.

Théorème A.63 (Propriété de Markov)**.** *Pour tout* $s \in \mathbb{R}_+$ *et tout* $x \in E$, *conditionnellement à l'événement* $\{X_s = x\}$, *le processus* $(X_{s+t})_{t\in\mathbb{R}_+}$ *est une chaîne de Markov à temps continu sur* E *de même générateur infinitésimal que* X *et de loi initiale* δ_x, *indépendante de* $(X_t)_{t\in[0,s]}$.

Pour une chaîne de Markov à temps continu, les notions de communication, ainsi que celles d'états récurrents et transitoires sont héritées de $\mathbf{P}$.

Théorème A.64 (Communication)**.** *Pour tous* $x, y \in E$, *les propriétés suivantes sont équivalentes :*
— $x \to y$ *pour* $\mathbf{P}$ *;*
— $\mathbb{P}(\exists t \in \mathbb{R}_+ : X_t = y \,|\, X_0 = x) > 0$ *;*
— *pour tout* $t \in \mathbb{R}_+^*$ *on a* $\mathbf{P}_t(x,y) > 0$ *;*
— *il existe* $t \in \mathbb{R}_+^*$ *tel que* $\mathbf{P}_t(x,y) > 0$.

Théorème A.65 (Classification des états)**.** *Les propriétés suivantes ont lieu :*
— *si* $x \in E$ *est récurrent pour* $\mathbf{P}$ *alors*

$$\mathbb{P}(\{t \in \mathbb{R}_+ : X_t = x\} \text{ est non borné} \,|\, X_0 = x) = 1;$$

— *si* $x \in E$ *est transitoire pour* $\mathbf{P}$ *alors*

$$\mathbb{P}(\{t \in \mathbb{R}_+ : X_t = x\} \text{ est non borné} \,|\, X_0 = x) = 0.$$

Théorème A.66 (Temps de premier retour)**.** *Pour tout* $x \in E$, *soit*

$$T_x := \inf\{t \geqslant S_1 : X_t = x\}$$

le temps de premier passage en x. *Alors on a la dichotomie suivante :*

1. *si* $\mathbf{L}(x,x) = 0$ *ou* $\mathbb{P}(T_x < \infty \mid X_0 = x) = 1$*, alors* x *est récurrent et*
$$\int_0^\infty \mathbf{P}_t(x,x)\,dt = \infty;$$
2. *si* $\mathbf{L}(x,x) < 0$ *et* $\mathbb{P}(T_x < \infty \mid X_0 = x) < 1$*, alors* x *est transitoire et*
$$\int_0^\infty \mathbf{P}_t(x,x)\,dt < \infty.$$

Pour une chaîne de Markov à temps continu, les mesures invariantes sont héritées de celles de $\mathbf{P}$ à une déformation près par la temporisation.

Théorème A.67 (Mesures invariantes)**.** *Pour toute mesure* μ *sur* E*, les propriétés suivantes sont équivalentes :*
— $\mu\mathbf{L} = 0$ *;*
— $\mu\mathbf{P}_t = \mu$ *pour tout* $t \in \mathbb{R}_+$ *;*
— $\nu\mathbf{P} = \nu$ *où* $\nu(x) := \mu(x)\boldsymbol{\lambda}(x)$ *pour tout* $x \in E$*.*

Notons que comme en temps discret, si μ est une mesure sur E réversible pour $\mathbf{L}$ c'est-à-dire que $\mu(x)\mathbf{L}(x,y) = \mu(y)\mathbf{L}(y,x)$ pour tous $x, y \in E$, alors μ est invariante pour $\mathbf{L}$ c'est-à-dire que $\mu\mathbf{L} = 0$.

Si $x \in E$ est récurrent, on dit que...
— x est récurrent positif si $\mathbf{L}(x,x) = 0$ (absorbant) ou si $\mathbb{E}_x(T_x) < \infty$;
— x est récurrent nul sinon.

Théorème A.68 (Récurrence positive et mesure de probabilité invariante)**.** *Si* $\mathbf{P}$ *est irréductible, alors les propriétés suivantes sont équivalentes :*

1. *tous les états sont récurrents positifs ;*
2. *il existe un état récurrent positif ;*
3. *il existe une mesure de probabilité invariante* μ *qui charge tous les états.*

Dans ce cas $\mu(x) = -\dfrac{1}{\mathbf{L}(x,x)\mathbb{E}_x(T_x)}$ *pour tout* $x \in E$*.*

Théorème A.69 (Convergence en loi)**.** *Si* $\mathbf{P}$ *est irréductible et récurrent positif et si* μ *est la mesure de probabilité invariante de* $\mathbf{L}$*, alors*
$$X_t \xrightarrow[t\to\infty]{\text{loi}} \mu,$$
quelque soit la loi initiale. En d'autres termes, pour tous $x, y \in E$*,*
$$\mathbf{P}_t(x,y) = \mathbb{P}(X_t = y \mid X_0 = x) \underset{t\to\infty}{\longrightarrow} \mu(y) = -\frac{1}{\mathbf{L}(y,y)\mathbb{E}_y(T_y)}.$$

Théorème A.70 (Théorème ergodique ou loi des grands nombres)**.** *Si* $\mathbf{P}$ *est irréductible récurrent alors quelque soit la loi initiale, pour tout* $x \in E$*,*

$$\frac{1}{t}\int_0^t \mathbb{1}_{\{X_s=x\}}\,ds \xrightarrow[t\to\infty]{\text{p.s.}} -\frac{1}{\mathbf{L}(x,x)\mathbb{E}_x(T_x)},$$

avec la convention $1/\infty = 0$ *dans le cas récurrent nul. De plus, dans le cas récurrent positif, on a, pour toute fonction bornée* $f : E \to \mathbb{R}$,

$$\frac{1}{t}\int_0^t f(X_s)\,ds \xrightarrow[t\to\infty]{\text{p.s.}} \int f\,d\mu = -\sum_{i\in E}\frac{f(x)}{\mathbf{L}(x,x)\mathbb{E}_x(T_x)}.$$

Le tableau A.2 compare temps discret et temps continu, en particulier :
— la notion de période est spécifique au temps discret ;
— la notion de temps d'explosion est spécifique au temps continu ;
— la notion de générateur est commune au cas discret et continu.

Remarque A.71 (Processus de naissance et mort [21]). *Ces processus importants sont les chaînes de Markov à temps continu sur* $E = \mathbb{N}$ *dont le générateur infinitésimal vérifie* $\mathbf{L}(x,y) = 0$ *si* $|y-x| > 1$ *pour tous* $x,y \in E$. *Le mouvement ne se fait qu'aux plus proches voisins. On dit que* $p(x) := \mathbf{L}(x,x+1)$ *avec* $x \in E$ *sont les taux de naissance tandis que* $q(x) := \mathbf{L}(x,x-1)$ *avec* $x \in E$ *sont les taux de mort (avec* $q(0) := 0$*). Le processus reste en* x *pendant un temps exponentiel de paramètre* $-\mathbf{L}(x,x) = p(x)+q(x)$*, puis il saute vers* $x+1$ *ou* $x-1$ *avec probabilités respectives* $p(x)/(p(x)+q(x))$ *et* $q(x)/(p(x)+q(x))$.
— *Le processus est irréductible ssi* $p > 0$ *sur* $\mathbb{N}$ *et* $q > 0$ *sur* $\mathbb{N}^*$ *;*
— *Le processus est non explosif ssi*

$$\sum_{x=1}^{\infty}\left(\frac{1}{p(x)} + \frac{q(x)}{p(x)p(x-1)} + \cdots + \frac{q(x)\cdots q(1)}{p(x)p(x-1)\cdots p(0)}\right) = \infty;$$

— *Le processus est récurrent positif ssi*

$$s = \sum_{x=1}^{\infty}\prod_{y=1}^{x}\frac{p(y-1)}{q(y)} < \infty,$$

et la mesure de probabilité invariante μ *est alors réversible, donnée par*

$$\mu(0) = \frac{1}{1+s} \quad \textit{et} \quad \mu(x) = \frac{1}{1+s}\prod_{y=1}^{x}\frac{p(y-1)}{q(y)}.$$

Voici enfin deux cas spéciaux importants :
— *Si* $p(x) = p$ *et* $q(x) = q < p$ *alors* $\mu = \mathrm{Geo}_{\mathbb{N}}(p/q)$ *;*
— *Si* $p(x) = p$ *et* $q(x) = qx$ *alors* $\mu = \mathrm{Poi}(p/q)$.

21. «*Birth and death processes*» en anglais.

Chaîne à temps discret	Chaîne à temps continu
Espace d'états au plus dénombrable E et loi initiale ν	
Processus $(X_n)_{n\in\mathbb{N}}$ avec $X_0 \sim \nu$	Processus $(X_t)_{t\in\mathbb{R}_+}$ avec $X_0 \sim \nu$
$\mathbf{P}^n(x,y) = \mathbb{P}(X_n = y \mid X_0 = x)$	$\mathbf{P}_t(x,y) = \mathbb{P}(X_t = y \mid X_0 = x)$
Rien d'équivalent	Temps d'explosion $T_\infty \in \overline{\mathbb{R}}_+$
Suite $(\mathbf{P}^n)_{n\in\mathbb{N}}$	Famille $(\mathbf{P}_t)_{t\in\mathbb{R}_+}$
Générateur $\mathbf{L} = \mathbf{P} - \mathbf{I}$	Générateur $\mathbf{L} = \lim_{t\to 0^+} \frac{1}{t}(\mathbf{P}_t - \mathbf{I})$
$\mathbb{E}(f(X_n)) = \nu\mathbf{P}^n f$	$\mathbb{E}(f(X_t)) = \nu\mathbf{P}_t f$
Notion de période	Rien d'équivalent
$S_{n+1} = \inf\{n > S_n; X_n > X_{S_n}\}$	$S_{n+1} = \inf\{t > S_n; X_t > X_{S_n}\}$
$\mathrm{Loi}(S_{n+1} - S_n \mid X_{S_n} = x)$	$\mathrm{Loi}(S_{n+1} - S_n \mid X_{S_n} = x)$
$= \mathrm{Geo}(-\mathbf{L}(x,x))$	$= \mathrm{Exp}(-\mathbf{L}(x,x))$
Chaîne incluse $\mathbb{P}(X_{S_{n+1}} = y \mid X_{S_n} = x) = -\frac{\mathbf{L}(x,y)}{\mathbf{L}(x,x)}$	
x absorbant ssi $\mathbf{L}(x,x) = 0$	
μ invariante ssi $\mu\mathbf{L} = 0$	
μ réversible ssi $\mu(x)\mathbf{L}(x,y) = \mu(y)\mathbf{L}(y,x)$	

Tableau A.2. Tableau analogique entre chaînes à temps discret et à temps continu.

A.7 Vecteurs aléatoires gaussiens

Si $X = (X_1,\dots,X_d)^\top$ est un vecteur colonne aléatoire de $\mathbb{R}^d$ alors on dit que $X \in L^1$ lorsque ses composantes sont dans L^1. Sa moyenne est le vecteur colonne $m = \mathbb{E}(X) = (\mathbb{E}(X_1),\dots,\mathbb{E}(X_d))^\top$. De même on dit que $X \in L^2$ lorsque ses composantes sont dans L^2. L'inégalité de Schwarz donne alors $X_jX_k \in L^1$ pour tous j,k ce qui permet de définir la matrice de covariance

$$C := \mathbb{E}((X-m)(X-m)^\top) = \mathbb{E}(XX^\top) - mm^\top = (\mathrm{Cov}(X_j,X_k))_{1\leqslant j,k\leqslant d}.$$

Théorème A.72 (Matrices de covariance des vecteurs aléatoires)**.** *Pour tout $d \geqslant 1$, l'ensemble des matrices de covariance des vecteurs aléatoires de $\mathbb{R}^d$ est égal à l'ensemble $\mathcal{S}_d^+$ des matrices $d\times d$ symétriques à spectre dans $\mathbb{R}_+$.*

Théorème A.73 (Déformations affines des vecteurs aléatoires)**.** *Si X est un vecteur colonne aléatoire de $\mathbb{R}^d$ de moyenne m et de matrice de covariance C alors pour tout $a \in \mathbb{R}^p$ et toute $A \in \mathcal{M}_{p,d}(\mathbb{R})$, le vecteur colonne aléatoire $a + AX$ de $\mathbb{R}^p$ a pour moyenne $a + Am$ et matrice de covariance $ACA^\top$.*

Si X est un vecteur aléatoire de $\mathbb{R}^d$ à coordonnées i.i.d. de loi $\mathcal{N}(0,1)$ alors X a pour moyenne 0 et matrice de covariance I_d, et on note $\mathcal{N}(0,I_d)$ sa loi. Pour tous $m \in \mathbb{R}^d$ et $A \in \mathcal{M}_d(\mathbb{R})$, le vecteur aléatoire $m+AX$ a pour moyenne m et matrice de covariance $C := AA^\top$ et on note $\mathcal{N}(m,C)$ sa loi.

Les notions de convergence presque sûre, en probabilité, et en loi, s'étendent directement aux vecteurs aléatoires. La fonction caractéristique $\varphi_X : \mathbb{R}^d \to \mathbb{C}$

d'un vecteur aléatoire X de $\mathbb{R}^d$ est définie par $\varphi_X(t) := \mathbb{E}(\exp(i\langle t, X\rangle))$ pour tout $t \in \mathbb{R}^d$, et jouit des mêmes propriétés (caractérisation et convergence).

Si $X \sim \mathcal{N}(m, C)$ avec $m \in \mathbb{R}^d$ et $C \in \mathcal{S}_d^+$ alors

$$\varphi_X(t) := \mathbb{E}\Big(e^{i\langle t,X\rangle}\Big) = e^{i\langle t,m\rangle - \frac{1}{2}\langle t,Ct\rangle}, \quad t \in \mathbb{R}^d.$$

Il en découle que la loi $\mathcal{N}(m, C)$ ne dépend que de m et C (d'où la notation !).

Théorème A.74 (LGN et TLC vectoriels)**.** *Si $(X_n)_{n\in\mathbb{N}^*}$ est une suite de vecteurs aléatoires i.i.d. de $\mathbb{R}^d$ dans L^1, de moyenne $m \in \mathbb{R}^d$, alors*

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{\text{p.s.}} m$$

si de plus les vecteurs sont dans L^2, de matrice de covariance $C \in \mathcal{S}_d^+$, alors

$$\sqrt{n}\left(\frac{X_1 + \cdots + X_n}{n} - m\right) = \frac{X_1 - m + \cdots + X_n - m}{\sqrt{n}} \xrightarrow[n\to\infty]{\text{loi}} \mathcal{N}(0, C).$$

Théorème A.75 (Vecteurs gaussiens)**.** *Si X est un vecteur aléatoire de $\mathbb{R}^d$ de moyenne m et de matrice de covariance C, alors les propriétés suivantes sont équivalentes et on dit dans ce cas que X est un vecteur gaussien :*

- *$X \sim \mathcal{N}(m, C)$;*
- *Pour tout $t \in \mathbb{R}^d$ la v.a. $\langle t, X\rangle$ suit une loi gaussienne sur $\mathbb{R}$;*
- *La fonction caractéristique de X est donnée pour tout $t \in \mathbb{R}^d$ par*

$$\varphi_X(t) := \mathbb{E}\Big(e^{i\langle t,X\rangle}\Big) = e^{i\langle t,m\rangle - \frac{1}{2}\langle t,Ct\rangle}.$$

Si $X \sim \mathcal{N}(m, C)$ alors pour toute $A \in \mathcal{M}_n(\mathbb{R})$, $AX \sim \mathcal{N}(Am, ACA^\top)$.

Théorème A.76 (Existence de densité)**.** *La loi gaussienne $\mathcal{N}(m, C)$ sur $\mathbb{R}^d$ admet une densité de probabilité par rapport à la mesure de Lebesgue de $\mathbb{R}^d$ si et seulement si C est inversible, et dans ce cas la densité est donnée par*

$$x \in \mathbb{R}^d \mapsto \frac{\exp\left(-\frac{1}{2}\big\langle (x-m), C^{-1}(x-m)\big\rangle\right)}{\sqrt{(2\pi)^d \det(C)}}.$$

La loi $\mathcal{N}(m, C)$ est portée par le sous-espace affine $m + \mathrm{Im}(C)$ de $\mathbb{R}^d$.

Lorsque $d = 1$ et $C = \sigma^2$ on retrouve la formule $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$.

Théorème A.77 (Caractérisation de l'indépendance)**.** *Si X et Y sont deux vecteurs aléatoires de $\mathbb{R}^p$ et $\mathbb{R}^q$ respectivement et si Z est le vecteur aléatoire de $\mathbb{R}^{p+q}$ obtenu par concaténation, alors X et Y sont indépendants si et seulement si $\varphi_Z(s, t) = \varphi_X(s)\varphi_Y(t)$ pour tout $(s, t) \in \mathbb{R}^p \times \mathbb{R}^q$.*

Théorème A.78 (Caractérisation de l'indépendance)**.** *Pour tout vecteur gaussien X de $\mathbb{R}^d$, les trois propriétés suivantes sont équivalentes :*

— *Les composantes $X_1, \ldots, X_d$ sont mutuellement indépendantes ;*
— *Les composantes $X_1, \ldots, X_d$ sont deux à deux indépendantes ;*
— *La matrice de covariance C de X est diagonale.*

Si $Z \sim \mathcal{N}(0,1)$ alors pour tout $n \in \mathbb{N}$ on a $\mathbb{E}(Z^{2n+1}) = 0$ par symétrie tandis que par une récurrence sur n et des intégrations par parties on obtient [22]

$$\mathbb{E}(Z^{2n}) = (2n-1)(2n-3)\cdots 3 = \frac{(2n)!}{2n(2n-2)\cdots 2} = \frac{(2n)!}{2^n n!}.$$

Le théorème suivant généralise cela aux vecteurs gaussiens.

Théorème A.79 (Formule de Wick)**.** *Si X est un vecteur gaussien centré de $\mathbb{R}^d$ et si $f_1, \ldots, f_{2n}$ est un nombre pair de formes linéaires sur $\mathbb{R}^d$, alors*

$$\mathbb{E}\left(\prod_{k=1}^{2n} f_k(X)\right) = \sum_{\mathcal{A}_n} \prod_{r=1}^{n} \mathbb{E}(f_{i_r}(X) f_{j_r}(X)),$$

où la somme porte sur l'ensemble $\mathcal{A}_n$ des appariements de $\{1, \ldots, 2n\}$, formé des suites non ordonnées de paires non ordonnées $\{\{i_1, j_1\}, \ldots, \{i_n, j_n\}\}$ telles que $\{i_1, \ldots, i_n, j_1, \ldots, j_n\} = \{1, \ldots, 2n\}$.

Il y a $\mathrm{card}(\mathcal{A}_n) = (2n)!/(2^n n!)$ appariements (formule d'Isserlis). Si $d = 1$ et $X = Z \sim \mathcal{N}(0,1)$ et $f_k(x) = x$ pour tout $k \in \{1, \ldots, 2n\}$ alors la formule de Wick se réduit au résultat déjà connu $\mathbb{E}(Z^{2n}) = \mathrm{card}(\mathcal{A}_n) = (2n)!/(2^n n!)$.

La formule de Wick ramène le calcul des moments d'un vecteur gaussien à l'utilisation de la matrice de covariance (membre de droite de la formule).

Les notions d'espérance conditionnelle et de loi conditionnelle se généralisent sans difficulté aux vecteurs aléatoires. Il se trouve que les lois conditionnelles des vecteurs gaussiens ont une structure remarquable.

Théorème A.80 (Lois conditionnelles des vecteurs gaussiens)**.** *Considérons un vecteur gaussien $Z = (X, Y)$ de $\mathbb{R}^d = \mathbb{R}^p \times \mathbb{R}^q$, de matrice de covariance*

$$C = \begin{pmatrix} \mathrm{Cov}(X,X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Cov}(Y,Y) \end{pmatrix} = \begin{pmatrix} K & J \\ J^\top & L \end{pmatrix}.$$

Si C est inversible, alors K et L le sont. De plus :

— *X et Y sont des vecteurs gaussiens de $\mathbb{R}^p$ et $\mathbb{R}^q$ de matrices de covariance respectives K et L, indépendants ssi $J = 0$;*
— *Pour tout $1 \leqslant j \leqslant q$, l'espérance conditionnelle $\mathbb{E}(Y_j \mid X)$ est la projection orthogonale dans L^2 de Y_j sur $\{a + b^\top X : (a,b) \in \mathbb{R} \times \mathbb{R}^p\}$. En particulier $\mathbb{E}(Y \mid X)$ est une fonction affine de X ;*

22. On note au passage que $\|Z\|_{2n} = \mathbb{E}(Z^{2n})^{1/(2n)} \sim_{n\to\infty} c\sqrt{n}$ avec $c = \sqrt{2/e}$.

— *Si K est inversible, alors* $\mathrm{Loi}(Y \mid X)$ *est gaussienne sur* $\mathbb{R}^q$ *de moyenne*

$$\mathbb{E}(Y \mid X) = \mathbb{E}(Y) + J^\top K^{-1}(X - \mathbb{E}(X))$$

et de matrice de covariance $L - J^\top K^{-1} J$. *En particulier le vecteur* $Y - \mathbb{E}(Y \mid X)$ *est indépendant de* X *et suit la loi* $\mathcal{N}(0, L - J^\top K^{-1} J)$.

La matrice carrée $S := L - J^\top K^{-1} J$ est le *complément de Schur* de K dans la matrice par blocs de C. La matrice C est symétrique à spectre strictement positif ssi K et son complément de Schur S dans C le sont. On a $\det(C) = \det(K)\det(S)$, et la formule d'inversion par blocs

$$\begin{aligned} C^{-1} &= \begin{pmatrix} I_p & -K^{-1}J \\ 0 & I_q \end{pmatrix} \begin{pmatrix} K^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ -J^\top K^{-1} & I_q \end{pmatrix} \\ &= \begin{pmatrix} K^{-1} + K^{-1} J S^{-1} J^\top K^{-1} & -K^{-1} J S^{-1} \\ -S^{-1} J^\top K^{-1} & S^{-1} \end{pmatrix}. \end{aligned}$$

Les zéros dans la matrice de covariance d'un vecteur gaussien indiquent une indépendance de coordonnées. Il se trouve que les zéros dans l'inverse de la matrice de covariance indiquent une indépendance conditionnelle !

Théorème A.81 (Indépendance conditionnelle)**.** *Si Z est un vecteur gaussien de $\mathbb{R}^d$ de matrice de covariance inversible C, alors deux composantes Z_i et Z_j avec $1 \leqslant i \neq j \leqslant d$ sont conditionnellement indépendantes sachant les autres composantes de Z ssi $(C^{-1})_{i,j} = 0$. Plus généralement, si $I \subset \{1, \ldots, d\}$, alors les composantes de $(Z_k)_{k \in I}$ sont conditionnellement indépendantes sachant les composantes de $(Z_k)_{k \notin I}$ ssi la sous-matrice $(C^{-1})_{I,I}$ de C^{-1} est diagonale.*

Voici une sorte de théorème de Pythagore pour les vecteurs gaussiens.

Théorème A.82 (de Cochran)**.** *Si X est un vecteur colonne aléatoire de $\mathbb{R}^n$ de loi $\mathcal{N}(m, \sigma^2 \mathrm{I}_n)$ avec $\sigma^2 > 0$, et si $\mathbb{R}^n = E_1 \oplus \cdots \oplus E_p$ est une décomposition de $\mathbb{R}^n$ en somme directe de p sous-espaces vectoriels orthogonaux de dimensions $d_1, \ldots, d_p$ avec $d_1 + \cdots + d_p = n$, alors, en notant $\mathbf{P}_k$ la matrice du projecteur orthogonal sur E_k et $Y_k = \mathbf{P}_k X$ la projection orthogonale :*

— *les v.a. $Y_1, \ldots, Y_p$ sont des vecteurs gaussiens indépendants et*

$$Y_k \sim \mathcal{N}(\mathbf{P}_k m, \sigma^2 \mathbf{P}_k), \quad 1 \leqslant k \leqslant p;$$

— *les v.a. $\|Y_1 - \mathbf{P}_1 m\|^2, \ldots, \|Y_p - \mathbf{P}_p m\|^2$ sont indépendantes et*

$$\frac{\|Y_k - \mathbf{P}_k m\|^2}{\sigma^2} \sim \chi^2(d_k), \quad 1 \leqslant k \leqslant p.$$

Corollaire A.83 (Échantillons gaussiens)**.** *Soit $X_1, \ldots, X_n$ sont des v.a. i.i.d. de loi $\mathcal{N}(m, \sigma^2)$ avec $\sigma^2 > 0$, et alors la* moyenne empirique $\widehat{m}_n := \frac{1}{n}\sum_{k=1}^n X_k$ *et la* variance empirique $\widehat{\sigma}_n^2 := \frac{1}{n-1}\sum_{k=1}^n (X_k - \widehat{m}_n)^2$ *sont indépendantes et*

$$\widehat{m}_n \sim \mathcal{N}\left(m, \frac{\sigma^2}{n}\right), \quad \frac{(n-1)}{\sigma^2}\widehat{\sigma}_n^2 \sim \chi^2(n-1), \quad \sqrt{n}\left(\frac{\widehat{m}_n - m}{\widehat{\sigma}_n}\right) \sim \mathrm{t}(n-1)$$

où $t(n-1)$ désigne la loi t de Student à $n-1$ degrés de liberté.

A.8 Calcul stochastique et formule d'Itô

Cette section est un survol très succinct et partiel du calcul stochastique.

Mouvement brownien

On dit qu'un processus $B := (B_t)_{t\in\mathbb{R}_+}$ est un mouvement brownien standard sur $\mathbb{R}$ lorsque l'une des propriétés équivalentes suivantes est réalisée :

— B est gaussien, centré et de covariance $\mathbb{E}(B_tB_s) = \min(s,t)$, $s,t \in \mathbb{R}_+$;

— $B_0 = 0$ et pour tout $n \in \mathbb{N}$ et tous $0 \leqslant t_1 \leqslant \cdots \leqslant t_n$, les accroissements

$$B_{t_1} - B_{t_0}, \ldots, B_{t_n} - B_{t_{n-1}}$$

sont indépendants, gaussiens et stationnaires, de lois respectives

$$\mathcal{N}(0, t_1 - t_0), \ldots, \mathcal{N}(0, t_n - t_{n-1}).$$

Le mouvement brownien peut être construit de plusieurs manières. Dans toute la suite, on se donne un mouvement brownien $B := (B_t)_{t\in\mathbb{R}_+}$. L'adaptation des processus est relative à la tribu naturelle de ce processus B.

Théorème A.84 (Martingale)**.** *Le processus $B = (B_t)_{t\in\mathbb{R}_+}$ est une martingale de carré intégrable et $(M_t)_{t\in\mathbb{R}_+} := (B_t^2 - t)_{t\in\mathbb{R}_+}$ est une martingale.*

Théorème A.85 (Régularité des trajectoires)**.** *Presque sûrement, une trajectoire du mouvement brownien est höldérienne d'indice α pour tout $\alpha \in]0, 1/2[$ uniformément sur les compacts, mais ne l'est pas pour l'indice $1/2$, et en particulier, elle n'est dérivable nulle part.*

Théorème A.86 (p-variation)**.** *Soient $T > 0$ et $p > 0$ des réels, et Δ une subdivision de l'intervalle $[0, T]$ donnée par $0 = t_0 < t_1 < \cdots < t_n = T$. Soit $|\Delta| := \max_{0\leqslant i\leqslant n-1}(t_{i+1} - t_i)$ le pas de la subdivision Δ, et*

$$V_p(\Delta) := \sum_{i=0}^{n-1} \left|B_{t_{i+1}} - B_{t_i}\right|^p.$$

Alors

$$\sup\left\{V_p(\Delta) \;:\; |\Delta| \leqslant \frac{1}{n}\right\} \xrightarrow[n\to\infty]{\text{p.s.}} \begin{cases} 0 & \text{si } p > 2, \\ T & \text{si } p = 2, \\ +\infty & \text{si } p < 2. \end{cases}$$

Intégrale stochastique

On peut définir l'intégrale d'une fonction déterministe $f \in L^2_{\mathrm{loc}}(\mathbb{R}_+)$ contre le mouvement brownien B sur l'intervalle $[0, t]$, que l'on note

$$M_t = \int_0^t f(s)\, dB_s.$$

Théorème A.87 (Intégrande déterministe)**.** *Le processus* $(M_t)_{t\in\mathbb{R}_+}$ *est une martingale continue, et un processus à accroissements indépendants, gaussien centré de covariance donnée pour tous* $s,t\in\mathbb{R}_+$ *par*

$$\mathbb{E}(M_sM_t)=\int_0^{\min(s,t)} f^2(u)\,du.$$

Si $f(s)=1$ pour tout $s\in\mathbb{R}_+$ alors $M=B$. Le cas $f(s)=e^s$ permet de construire le processus d'Ornstein-Uhlenbeck $(X_t)_{t\in\mathbb{R}_+}$ en posant

$$X_t=xe^{-t}+e^{-t}\int_0^t e^s\,dB_s,\quad t\in\mathbb{R}_+.$$

C'est un processus gaussien, et $\mathbb{E}(X_t)=xe^{-t}$ et $\mathrm{Cov}(X_s,X_t)=e^{-2|t-s|}$.

On peut étendre la définition de l'intégrale stochastique $(M_t)_{t\in\mathbb{R}_+}$ à des intégrandes dans l'ensemble Λ des processus $(\theta_t)_{t\in\mathbb{R}_+}$ adaptés continus à droite avec limite à gauche (càdlàg) tels que pour tout $t\in\mathbb{R}_+$

$$\mathbb{E}\int_0^t \theta_s^2\,ds<\infty.$$

Le processus M n'est plus gaussien en général. Cependant, le processus M est toujours une martingale à trajectoires continues de carré intégrable et le processus $(N_t)_{t\in\mathbb{R}_+}$ suivant est une martingale :

$$N_t=\left(\int_0^t \theta_s\,dB_s\right)^2-\int_0^t\theta_s^2\,ds,\quad t\in\mathbb{R}_+.$$

Semi-martingale

Si $(\beta_t)_{t\in\mathbb{R}_+}$ est un processus adapté à variation finie et si $(\gamma_t)_{t\geqslant 0}$ est un processus appartenant à l'ensemble des intégrandes admissibles Λ, alors pour tout $x\in\mathbb{R}$, le processus $(X_t)_{t\in\mathbb{R}_+}$ défini pour tout $t\in\mathbb{R}_+$ par

$$X_t:=x+\int_0^t\beta_s ds+\int_0^t\gamma_s dB_s$$

est appelé une *semi-martingale* associée à (β,γ). Les semi-martingales possèdent une propriété de stabilité remarquable par composition avec une fonction de classe $\mathcal{C}^2$, qui constitue une sorte de théorème fondamental du calcul intégral pour des processus à variation quadratique finie :

Théorème A.88 (Formule d'Itô)**.** *Si* X *est une semi-martingale associée à* (β,γ) *et si* f *est une fonction de classe* $\mathcal{C}^2$ *alors le processus* $(f(X_t))_{t\in\mathbb{R}_+}$ *est encore une semi-martingale donnée par*

$$\begin{aligned}f(X_t)&=f(x)+\int_0^t f'(X_s)\,dX_s+\int_0^t\frac{1}{2}\gamma_s^2 f''(X_s)\,ds\\&=f(x)+\int_0^t\left(\beta_s f'(X_s)+\frac{1}{2}\gamma_s^2f''(X_s)\right)ds+\int_0^t f'(X_s)\gamma_s\,dB_s.\end{aligned}$$

Équation différentielle stochastique

Si $b : \mathbb{R} \mapsto \mathbb{R}$ et $\sigma : \mathbb{R} \mapsto \mathbb{R}_+$ sont des fonctions lipschitziennes, alors pour tout $x \in \mathbb{R}$, on peut définir le processus $(X_t)_{t\in\mathbb{R}_+}$ solution de l'équation différentielle stochastique (ÉDS)

$$X_t = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dB_s, \quad t \in \mathbb{R}_+,$$

que l'on note également

$$X_0 = x, \quad dX_t = b(X_t)dt + \sigma(X_t)dB_t.$$

Les fonctions b et σ sont respectivement appelés coefficients de dérive et de diffusion. Le processus $(X_t)_{t\in\mathbb{R}_+}$ est un processus de Markov. Son semi-groupe $(P_t)_{t\geqslant 0}$ est donné pour toute fonction $f : \mathbb{R} \to \mathbb{R}$ mesurable et et bornée, pour tout $x \in \mathbb{R}$, et pour tout $t \in \mathbb{R}_+$, par

$$P_t f(x) := \mathbb{E}(f(X_t) \,|\, X_0 = x)$$

Son générateur infinitésimal L est l'opérateur différentiel linéaire du second ordre sans terme constant donné pour toute fonction f de classe $\mathcal{C}^2$ par

$$Lf(x) := b(x)f'(x) + \frac{\sigma^2(x)}{2}f''(x).$$

On lit sur le générateur infinitésimal L la dynamique du processus : le terme du premier ordre correspond à une dérive déterministe tandis que celui du second ordre traduit une excitation brownienne locale déformée par σ.

La formule d'Itô assure que, pour $0 \leqslant r < t$,

$$f(X_t) = f(X_r) + \int_r^t Lf(X_s)ds + \int_r^t f'(X_s)\sigma(X_s)dB_s.$$

Si $\mu_t := \mathrm{Loi}(X_t)$ et si $\mu_0 := \mathrm{Loi}(X_0)$ est donnée, alors, en prenant l'espérance dans les deux membres de la formule précédente, et en utilisant la définition de μ_t et la nature martingale de l'intégrale stochastique, il vient

$$\int_{\mathbb{R}} f d\mu_t = \int_{\mathbb{R}} f d\mu_r + \int_{\mathbb{R}} \int_r^t Lf d\mu_s.$$

Si les coefficients b et σ sont assez réguliers, la loi μ_t de X_t admet une densité u_t de classe $\mathcal{C}^\infty$. La relation ci-dessus peut s'écrire, au moins formellement,

$$\int_{\mathbb{R}} \left(u_t - u_r - \int_r^t L^* u_r dr \right)(x) f(x) dx = 0.$$

La fonction $(t,x) \mapsto u_t(x)$ est solution de l'équation aux dérivées partielles (ÉDP) dite de Fokker-Planck

$$\partial_t u = L^* u$$

où L^* est l'opérateur adjoint de l'opérateur L pour la mesure de Lebesgue.

Une mesure μ est invariante pour X si pour toute fonction f régulière

$$\int_{\mathbb{R}} \mu(dx) Lf(x) = 0.$$

Dans le cas où $\sigma(x) = \sqrt{2}$ et $b(x) = -U'(x)$ pour tout $x \in \mathbb{R}$, les mesures invariantes ont une densité proportionnelle à e^{-U}. L'une d'elles est une mesure de probabilité si e^{-U} est d'intégrale finie. Le processus d'Ornstein-Uhlenbeck correspond au cas $b(x) = -x$ pour tout $x \in \mathbb{R}$, et son unique mesure de probabilité invariante est la loi gaussienne standard $\mathcal{N}(0, 1)$.

Une mesure μ est réversible pour le processus X si l'opérateur L est auto-adjoint dans $L^2(\mu)$: pour toutes fonctions f et g régulières,

$$\int_{\mathbb{R}} \mu(dx) g(x) Lf(x) = \int_{\mathbb{R}} \mu(dx) f(x) Lg(x).$$

Toute mesure réversible est invariante. La réciproque est vraie pour les diffusions sur $\mathbb{R}$ mais fausse pour la généralisation des diffusion sur $\mathbb{R}^d$, $d \geqslant 2$.

Littérature

BC07. B. Bercu et D. Chafaï – *Modélisation stochastique et simulation, cours et applications*, Collection Sciences Sup, Dunod, 2007, Mathématiques appliquées/SMAI. 1

CM15. D. Chafaï et F. Malrieu – « Recueil de modèles aléatoires », à parraître dans la collection Mathématiques & Applications, Springer, 2015. 1