Le problème des moments

Références : P. Billingsley, *Probability and Measure* – R. Durrett, Probability : *Theory and Examples* – `http://perso.eleves.ens-rennes.fr/people/rudy.morel/agregation.html`

Leçons : 243, 260, 261.

Définition 1

Soit μ une probabilité sur $\mathbb{R}$. Soit $n \in \mathbb{N}$, on définit les moments absolus et les moments (lorsque cela est possible) de μ respectivement par :

$$\mu_n = \int_{\mathbb{R}} |x|^n d\mu(x).$$

et

$$m_n = \int_{\mathbb{R}} x^n d\mu(x),$$

L'objet du développement est le théorème suivant.

Théorème 2

Soit μ une probabilité sur $\mathbb{R}$ admettant des moments (finis) à tout ordre. On suppose que :

$$\limsup_n \frac{(\mu_n)^{1/n}}{n} < +\infty.$$

Alors μ est caractérisée par ses moments.

Démonstration : Le but est de montrer l'analyticité de la fonction caractéristique ce qui suffira à conclure. On note ϕ la fonction caractéristique de la loi μ. Le théorème de régularité sous l'intégrale assure que ϕ est de classe C^∞ sur $\mathbb{R}$ (ce qui n'assure pas l'analyticité ...), de plus on a :

$$\forall n \in \mathbb{N},\ \frac{\phi^{(n)}(0)}{n!} = i^n m_n.$$

- **Étape 1 :** On se donne $x, t, h \in \mathbb{R}$ et on applique la formule de Taylor à la fonction $s \mapsto e^{isx}$, qui est de classe C^∞, entre 0 et h :

$$e^{ixh} = \sum_{m=0}^{n} \frac{(ix)^m}{m!} h^m + \int_0^h \frac{(h-s)^n}{n!} (ix)^{n+1} e^{isx}\, ds.$$

  On déduit donc que :

$$e^{ix(t+h)} = \sum_{m=0}^{n} \frac{(ix)^m}{m!} e^{itx} h^m + \int_0^h \frac{(h-s)^n}{n!} (ix)^{n+1} e^{i(s+t)x}\, ds.$$

On intègre ensuite sur $\mathbb{R}$ pour la mesure μ :

$$\phi(t+h) - \sum_{m=0}^{n} \frac{h^m}{m!} \int_{\mathbb{R}} (ix)^m e^{itx}\, d\mu(x) = \int_{\mathbb{R}} \int_0^h \frac{(h-s)^n}{n!} (ix)^{n+1} e^{i(s+t)x}\, ds\, d\mu(x).$$

On suppose $h \geq 0$, les calculs étant analogues dans l'autre cas. On a par inégalité triangulaire et d'après le théorème de Fubini :

$$\begin{aligned} \left| \phi(t+h) - \sum_{m=0}^{n} \frac{\phi^{(m)}(t)}{m!} h^m \right| &\leq \int_{\mathbb{R}} \int_0^h \frac{|h-s|^n}{n!} |x|^{n+1}\, ds\, d\mu(x) \\ &= \frac{|h|^{n+1}}{(n+1)!} \mu_{n+1}. \end{aligned}$$

- **Etape 2 :** On note $R = \limsup_n \frac{(\mu_n)^{1/n}}{n} < +\infty$. On dispose donc de $N \in \mathbb{N}$ tel que :

$$\forall n \geq N,\ \frac{(\mu_n)^{1/n}}{n} \leq R+1.$$

On a donc :

$$\forall n \geq N,\ \mu_n \leq n^n (R+1)^n.$$

Ainsi si on suppose que $|h| \leq \frac{1}{2e(R+1)}$, on a :

$$\forall n \geq N,\ \frac{|h|^n}{n!} \mu_n \leq \frac{n^n}{n!} |h|^n (R+1)^n.$$

Le développement en série de e^n montre que : $\frac{n^n}{n!} \leq e^n$. Cela assure que :

$$\forall n \geq N,\ \frac{|h|^n}{n!} \mu_n \leq \left(\frac{1}{2}\right)^n.$$

Ainsi :

$$\forall |h| \leq \frac{1}{2e(R+1)}, \quad \phi(t+h) = \sum_{m=0}^{\infty} \frac{\phi^{(m)}(t)}{m!} h^m,$$

ce qui est exactement le caractère analytique de ϕ.

- **Étape 3** On conclut : la fonction ϕ est analytique sur $\mathbb{R}$ sous les hypothèses précédentes. Puisque les dérivées de ϕ en 0 dépendent uniquement des moments de μ on déduit μ est caractérisée par ses moments. En effet si ν est une autre probabilité vérifiant les mêmes hypothèses, alors les deux fonctions caractéristiques coïncident autour de 0 et sont analytiques donc égales. Comme la fonction caractéristique caractérise la loi, on peut conclure que $\mu = \nu$.

$\square$

Corollaire 3

Une probabilité à support compact est caractérisée par ses moments, c'est en particulier le cas des lois à support fini.

Démonstration : Soit $M > 0$ tel que $\mu([-M, M]) = 1$. Alors $\mu_n \leq M^n$ dont on déduit :

$$\limsup_n \frac{(\mu_n)^{1/n}}{n} = 0.$$

$\square$

Donnons maintenant un exemple et un contre-exemple.

Proposition 4

La loi $\mathcal{N}(0,1)$ est caractérisée par ses moments.

Démonstration : On a pour tout n par intégration par partie :

$$\begin{aligned}
\mu_n &= 2\int_0^\infty x^n e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \\
&= 2\int_0^\infty (n-1)x^{n-2} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} \\
&= (n-1)\mu_{n-2}.
\end{aligned}$$

On en déduit qu'il existe une constante C (en réalité on a une expression exacte) telle que : $\forall n \in \mathbb{N}$, $\mu_n \leq C(n!) \leq Cn^n$. Cela assure que : $\limsup_n \frac{(\mu_n)^{1/n}}{n} \leq 1$.

$\square$

Proposition 5

Soit N une variable aléatoire de loi $\mathcal{N}(0,1)$. Alors e^N est à densité $x \mapsto f(x) = \frac{1}{\sqrt{2\pi}x} e^{-\frac{ln(x)^2}{2}} \mathbf{1}_{x>0}$ par rapport à la mesure de Lebesgue. La loi à densité $x \mapsto g(x) = f(x)(1 + \sin(2\pi ln(x)))$ a les mêmes moments que f, mais n'est pas égale à f. Cette loi n'est donc pas caractérisée par ses moments.

Démonstration : Soit $k \in \mathbb{N}$. On a, en posant $y = ln(x)$:

$$\begin{aligned}
\int_0^\infty x^k f(x) \sin(2\pi ln(x))\, dx &= \int_{\mathbb{R}} e^{ky} e^{-\frac{y^2}{2}} \sin(2\pi y) \frac{dy}{\sqrt{2\pi}} \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{y^2-2yk}{2}} \sin(2\pi y)\, dy \\
&= \frac{e^{\frac{k^2}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{(y-k)^2}{2}} \sin(2\pi y)\, dy \\
&= \frac{e^{\frac{k^2}{2}}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{u^2}{2}} \sin(2\pi u)\, du \\
&= 0
\end{aligned}$$

f et g ont les mêmes moments mais ne sont pas égales. $\square$

Remarque 6. - *Pour que le développement tienne en 15 minutes, on peut se contenter de démontrer le théorème en énonçant le corollaire et selon le temps qu'il reste traiter l'exemple de la gaussienne voire également le contre exemple si on est rapide !*

- *Le fait que la fonction caractéristique caractérise la loi peut se justifier rapidement par inversion de Fourier sur les distributions tempérées.*
- *En utilisant l'inégalité de Cauchy-Schwarz, on obtient :*

$$\mu_{2n+1} = \int_{\mathbb{R}} |x|^{n+1}|x|^n \, d\mu(x) \leq \sqrt{\mu_{2n}\mu_{2n+2}}.$$

Cela permet de montrer que la condition suivante suffit à assurer la caractérisation de la loi par les moments :

$$\limsup_n \frac{(\mu_{2n})^{1/2n}}{2n} < +\infty.$$

Annexe 1

La suite est plutôt " culturelle " bien que très intéressante notamment pour la leçon 262 ! Le théorème nécessite le critère de convergence étroite d'une suite tendue qui a été démontré dans le développement sur les théorèmes de Prokhorov et de Lévy mais que je remets en Annexe 2. On commence d'abord par des résultats généraux sur l'uniforme intégrabilité qu'on pourra trouver dans le livre de O. Garet et A. Kurtzmann : *De l'intégration aux probabilités*. Ces résultats trouvent tout à fait leur place dans la leçon 234 !

Définition 7

Une famille de variables aléatoires réelles $(X_i)_i$ est dite uniformément intégrable si :

$$\lim_{M \to +\infty} \quad \sup_i \mathbb{E}(|X_i| \mathbf{1}_{|X_i| \geq M}) = 0.$$

Lemme 8

Soit $(X_n)_n$ une suite de variables aléatoires qui converge en loi vers X. Alors :

$$\mathbb{E}(|X|) \leq \liminf_n \mathbb{E}(|X_n|).$$

Démonstration : D'après le théorème de Fubini-Tonelli, on a :

$$\mathbb{E}(|X_n|) = \int_{\mathbb{R}^+} \mathbb{P}(|X_n| > t)\, dt.$$

Or en tout point t de continuité de la fonction de répartition de X, on a [1] :

$$\mathbb{P}(|X_n| > t) \to \mathbb{P}(|X| > t).$$

Comme l'ensemble des discontinuités d'une fonction monotone est dénombrable, on peut appliquer le lemme de Fatou :

$$\mathbb{E}(|X|) = \int_{\mathbb{R}^+} \liminf_n \mathbb{P}(|X_n| > t)\, dt \leq \liminf_n \mathbb{E}(|X_n|).$$

□

Théorème 9

Soit $(X_n)_n$ une suite de variables aléatoires uniformément intégrable qui converge en loi vers X. Alors $X \in L^1$ et on a :

$$\mathbb{E}(X_n) \to \mathbb{E}(X).$$

1. Pour la preuve on peut consulter le développement sur les théorèmes de Prokhorov et de Lévy

Démonstration : Puisqu'une famille uniformément intégrable est bornée dans L^1, on sait déjà que X est intégrable par le lemme précédent. Soit $\epsilon > 0$. On dispose de $M > 0$ tel que :

$$\sup_n \mathbb{E}(|X_n|\mathbf{1}_{|X_n| \geq M}) \leq \epsilon \quad \text{et} \quad \mathbb{E}(|X|\mathbf{1}_{|X| \geq M}) \leq \epsilon.$$

On a alors successivement, en notant $\wedge$ le minimum :

$$\begin{aligned} |\mathbb{E}(X_n) - \mathbb{E}(X)| &= |\mathbb{E}(X_n \wedge M) - \mathbb{E}(X \wedge M) + \mathbb{E}((X_n - M)\mathbf{1}_{X_n \geq M}) - \mathbb{E}((X - M)\mathbf{1}_{X \geq M}) \\ &\leq |\mathbb{E}(X_n \wedge M) - \mathbb{E}(X \wedge M)| + \mathbb{E}(|X_n - M|\mathbf{1}_{X_n \geq M}) + \mathbb{E}(|X - M|\mathbf{1}_{X \geq M}) \\ &\leq |\mathbb{E}(X_n \wedge M) - \mathbb{E}(X \wedge M)| + \mathbb{E}(|X_n|\mathbf{1}_{X_n \geq M}) + \mathbb{E}(|X|\mathbf{1}_{X \geq M}) \\ &\leq |\mathbb{E}(X_n \wedge M) - \mathbb{E}(X \wedge M)| + 2\epsilon. \end{aligned}$$

La fonction $x \mapsto x \wedge M$ étant continue bornée, on a : $\mathbb{E}(X_n \wedge M) \to \mathbb{E}(X \wedge M)$. Ainsi :

$$\limsup_n |\mathbb{E}(X_n) - \mathbb{E}(X)| \leq 2\epsilon,$$

D'où le résultat. □

J'énonce ici un corollaire que je n'utilise pas dans la suite mais qui est un résultat " classique ".

Théorème 10

Soit $(X_n)_n$ une suite de variables aléatoires uniformément intégrable qui converge en probabilité vers X. Alors $X \in L^1$ et on a :

$$X_n \underset{L^1}{\to} X.$$

Démonstration : Par hypothèse la suite $(|X_n - X|)_n$ converge en probabilité vers 0 donc également en loi. Il suffit d'appliquer le théorème précédent pour conclure. □

Passons maintenant au résultat de convergence grâce aux moments.

Théorème 11 (*Méthode des moments*)

Soit $(X_n)_n$ une suite de variables aléatoires admettant des moments à tout ordre. Soit X une variable aléatoire admettant également des moments à tout ordre et telle que la loi μ de X est caractérisée par ses moments. On suppose que :

$$\forall r \in \mathbb{N},\ \mathbb{E}(X_n^r) \to \mathbb{E}(X^r) := m_r.$$

Alors $(X_n)_n$ converge en loi vers X.

Démonstration : On note μ_n la loi de X_n et μ celle de X. D'après le corollaire de l'annexe, pour prouver le théorème, il suffit de montrer que la suite $(\mu_n)_n$ est tendue et admet μ comme unique valeur d'adhérence (pour la convergence en loi).

- **Tension :** Il suffit de voir que la suite $(X_n)_n$ est bornée dans L^1 par hypothèse. On note $M = \sup_n \mathbb{E}(|X_n|)$. L'inégalité de Markov assure que si $x > 0$:

$$\forall n \in \mathbb{N},\, \mathbb{P}(|X_n| \geq x) \leq \frac{M}{x},$$

ce qui implique directement la tension de la suite $(\mu_n)_n$.

- μ **est la seule valeur d'adhérence :** Soit ν une autre valeur d'adhérence de la suite $(\mu_n)_n$ et Y une variable aléatoire de loi ν. On dispose d'une extraction $(n_k)_k$ telle que $(X_{n_k})_k$ converge en loi vers Y. Si on montre l'uniforme intégrabilité de la suite $(|X_n|^r)_n$ pour tout $r \in \mathbb{N}$, on aura, d'après un résultat précédent :

$$\mathbb{E}(X_{n_k}^r) \underset{k \to +\infty}{\to} \mathbb{E}(Y^r)$$

puisque $(X_n^r)_n$ converge en loi vers X^r. Par unicité de la limite, on a alors :

$$\forall r \in \mathbb{N},\, \mathbb{E}(Y^r) = m_r.$$

Ainsi puisque X est caractérisée par ses moments, on obtient que $\nu = \mu$, ce qui achèvera la preuve. On se donne donc $r \in \mathbb{N}$. On note $M_{2r} = \sup_n \mathbb{E}(|X_n|^{2r})$. On a alors pour tout $C > 0$:

$$\mathbb{E}(|X_n|^r \mathbf{1}_{|X_n|^r \geq C}) \leq \frac{1}{C^r}\mathbb{E}(|X_n|^{2r}) \leq \frac{M_{2r}}{C^r}.$$

Cette majoration assure l'uniforme intégrabilité de $(X_n^r)_n$ pour tout $r \in \mathbb{N}$.

$\square$

Annexe 2

Je remets ici la preuve du critère de convergence étroite d'une suite tendue utilisé, on pourra trouver l'intégralité dans le développement sur les théorèmes de Prokhorov et de Lévy ou dans *De l'intégration aux probabilités.*

Définition 12

Soient $(\mu_n)_n$ une suite de probabilités sur $\mathbb{R}^n$ et μ une probabilité sur $\mathbb{R}^n$. On dit que cette suite converge étroitement vers μ si :

$$\forall \phi \in \mathcal{C}_b^0(\mathbb{R}^n, \mathbb{R}) \int_{\mathbb{R}^n} \phi(x)\, d\mu_n(x) \to \int_{\mathbb{R}^n} \phi(x)\, d\mu(x).$$

Définition 13 (*Tension*)

Soit $(\mu_n)_n$ une suite de probabilités sur $\mathbb{R}^n$. On dit que cette suite est tendue si pour tout $\epsilon > 0$, il existe un compact K tel que :

$$\forall n \in \mathbb{N},\ \mu_n(K) \geq 1 - \epsilon.$$

Le théorème suivant est un critère de compacité pour la convergence étroite de probabilités. Nous en aurons besoin pour démontrer le théorème de Lévy.

Théorème 14 (*Théorème de Prokhorov*)

Soit $(\mu_n)_n$ une suite tendue de probabilités sur $\mathbb{R}$. Alors il existe une extraction de cette suite qui converge étroitement.

Avant de démontrer ce théorème on aura besoin du lemme suviant.

Théorème 15 (*Théorème de Helly*)

Soit $(F_n)_n$ une suite de fonctions de répartition. Alors il existe une extraction $(\alpha_n)_n$ et une fonction $F : \mathbb{R} \to [0,1]$ croissante et continue à droite telle qu'en tout point de continuité x de F, on a :

$$F_{\alpha_n}(x) \underset{n\to\infty}{\to} F(x).$$

Démonstration : Soit $q \in \mathbb{Q}$, la suite $(F_n(q))_n$ est une suite de $[0,1]$, on peut donc extraire une sous-suite qui converge vers un réel qu'on note $\tilde{F}(q) \in [0,1]$. Par extraction diagonale, il existe une extraction $(\alpha_n)_n$ telle que :

$$\forall q \in \mathbb{Q},\ F_{\alpha_n}(q) \underset{n\to\infty}{\to} \tilde{F}(q).$$

On peut alors définir, si $x \in \mathbb{R}$:

$$F(x) = \inf_{q\in\mathbb{Q},\, x<q} \{\tilde{F}(q)\}.$$

F est croissante par construction. Montrons qu'elle est continue à droite. Soit $x \in \mathbb{R}$ et $\epsilon > 0$. Il existe $q > x$ tel que $\tilde{F}(q) \leq F(x) + \epsilon$. Ainsi, si $y \in [x, q]$, on a :

$$F(x) \leq F(y) \leq \tilde{F}(q) \leq F(x) + \epsilon.$$

Soit maintenant x un point de continuité de F et $\epsilon > 0$. Par continuité, il existe $y < x$ tel que :

$$F(x) - \epsilon \leq F(y).$$

Soient $r, s \in \mathbb{Q}$ tels que : $y < r < x < s$ et $\tilde{F}(s) \leq F(x) + \epsilon$. On a alors :

$$F(x) - \epsilon \leq F(y) \leq \tilde{F}(r) \leq \tilde{F}(s) \leq F(x) + \epsilon.$$

Or par définition, on a :

$$F(x) - \epsilon \leq \tilde{F}(r) = \lim_{n} F_{\alpha_n}(r) \leq \underline{\lim}\, F_{\alpha_n}(x) \leq \overline{\lim}\, F_{\alpha_n}(x) \leq \overline{\lim}\, F_{\alpha_n}(s) = \tilde{F}(s) \leq F(x) + \epsilon.$$

En laissant tendre ϵ vers 0, on obtient $F_{\alpha_n}(x) \underset{n \to \infty}{\to} F(x)$. □

Démonstration : (Théorème de Prokhorov).

On note F_n la fonction de répartition de μ_n. Il suffit de montrer que la fonction F obtenue est une fonction de répartition. Par caractérisation de la convergence étroite avec les fonctions de répartition, on aura le résultat. Si on montre que F tend vers 0 en $-\infty$ et vers 1 en $+\infty$, F sera une fonction de répartition grâce à l'inverse généralisé. Fixons $\epsilon > 0$, Par hypothèse il existe $M > 0$, qu'on peut supposé point de continuité de F puisque les discontinuités d'une fonction monotone forment un ensemble au plus dénombrable, tel que :

$$\forall n,\ \mu_n([-M, M]) \geq 1 - \epsilon.$$

En passant à la limite suivant l'extraction du lemme de Helly, on obtient :

$$F(M) - F(-M) \geq 1 - \epsilon.$$

Compte tenu de la monotonie de F, cela assure que F est une fonction de répartition.

□

Le corollaire suivant est un critère de convergence étroite (analogue au critère pour les suites réelles bornés, d'ailleurs la preuve en découle).

Corollaire 16

Si $(\mu_n)_n$ est tendue et admet une unique valeur d'adhérence pour la convergence étroite, alors $(\mu_n)_n$ convergence étroitement vers μ.

Démonstration : Soit f continue bornée. La suite $\left(\int f\, d\mu_n\right)_n$ converge vers $\int f\, d\mu$ puisque de toute extraction de cette suite, on peut extraire une sous-suite qui converge vers $\int f\, d\mu$ grâce au théorème de Prokhorov et car il y a une unique valeur d'adhérence. □