Théorème de Lévy et théorème central limite

Mon développement

Dans tout le développement, on notera $\langle ., . \rangle$ le produit scalaire usuel sur $\mathbb{R}^d$ et $\mathcal{M}_b(\mathbb{R}^d)$ l'ensemble des mesures de masse *inférieure ou* égale à b sur $\mathbb{R}^d$. De plus, on notera $\widehat{f}$, resp. $\widehat{\mu}$, la transformée de Fourier d'une fonction f, resp. d'une mesure μ.
On commence par rappeler une propriété importante, qu'on utilisera deux fois par la suite.

Lemme. Soient $(\mu_n)_{n\in\mathbb{N}}$, μ des mesures de $\mathcal{M}_b(\mathbb{R}^d)$ telles que $(\mu_n)_{n\in\mathbb{N}}$ converge faiblement vers μ. Alors, $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ si et seulement si

$$\lim_{n\to+\infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d) \ .$$

Le sens direct est évident en considérant la fonction constante égale à 1, qui est continue et bornée sur $\mathbb{R}^d$.
Réciproquement, on suppose que $(\mu_n)_{n\in\mathbb{N}}$ converge faiblement vers μ et que

$$\lim_{n\to+\infty} \mu_n(\mathbb{R}^d) = \mu(\mathbb{R}^d) \ .$$

Soit f une fonction continue et bornée sur $\mathbb{R}^d$. Soit $(g_p)_{p\in\mathbb{N}}$ une suite de fonctions continues sur $\mathbb{R}^d$, positives, de limite nulle à l'infini, qui converge en croissant vers la fonction constante égale à 1. Pour tous $n, p \in \mathbb{N}$, on a :

$$\begin{aligned}
\left|\int f d\mu_n - \int f d\mu\right| &\leq \left|\int f d\mu_n - \int f g_p d\mu_n\right| + \left|\int f g_p d\mu_n - \int f g_p d\mu\right| + \left|\int f g_p d\mu - \int f d\mu\right| \\
&\leq \|f\|_\infty \int (1-g_p) d\mu_n + \left|\int f g_p d\mu_n - \int f g_p d\mu\right| + \|f\|_\infty \int (1-g_p) d\mu \\
&= \|f\|_\infty \left(\mu_n(\mathbb{R}^d) - \int g_p d\mu_n\right) + \left|\int f g_p d\mu_n - \int f g_p d\mu\right| + \|f\|_\infty \int (1-g_p) d\mu
\end{aligned}$$

Comme les g_p et les fg_p sont continues et de limite nulle à l'infini et $\mu_n(\mathbb{R}^d)$ tend vers $\int 1 d\mu$, en faisant tendre $n \to +\infty$, on obtient

$$\limsup_{n\to+\infty} \left|\int f d\mu_n - \int f d\mu\right| \leq 2\|f\|_\infty \int (1-g_p) d\mu$$

pour tout $p \in \mathbb{N}$. D'après le théorème de convergence dominée (domination par la fonction constante égale à 1), en faisant tendre $p \to +\infty$, on obtient

$$\limsup_{n\to+\infty} \left|\int f d\mu_n - \int f d\mu\right| \leq 0$$

donc $\int f d\mu_n$ converge de limite $\int f d\mu$ quand $n \to +\infty$.
$(\mu_n)_{n\in\mathbb{N}}$ converge donc étroitement vers μ. $\square$

Remarque : En particulier, si les μ_n et μ sont des probabilités, il y a équivalence entre convergence faible et convergence étroite.

Théorème (Lévy). Soient $(\mu_n)_{n\in\mathbb{N}}$ une suite de mesures de $\mathcal{M}_b(\mathbb{R}^d)$.
- Si $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers une mesure $\mu \in \mathcal{M}_b(\mathbb{R}^d)$, alors $(\widehat{\mu_n})_{n\in\mathbb{N}}$ converge simplement de limite $\widehat{\mu}$.
- Si $(\widehat{\mu_n})_{n\in\mathbb{N}}$ converge simplement vers une fonction φ continue en 0, alors il existe une unique mesure $\mu \in \mathcal{M}_b(\mathbb{R}^d)$ telle que $\widehat{\mu} = \varphi$; de plus, $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ.

Le premier point provient de la définition de la convergence étroite, en constatant que pour tout $t \in \mathbb{R}^d$, la fonction $x \mapsto e^{i\langle t,x\rangle}$ est continue et bornée sur $\mathbb{R}^d$.

On suppose maintenant que $(\widehat{\mu_n})_{n\in\mathbb{N}}$ converge simplement vers une fonction φ continue en 0.
Comme l'ensemble $\mathcal{M}_b(\mathbb{R}^d)$ est métrisable et compact pour la topologie faible, $(\mu_n)_{n\in\mathbb{N}}$ admet une valeur d'adhérence faible, notée μ, et il existe une suite extraite $(\mu_{\psi(n)})_{n\in\mathbb{N}}$ convergeant faiblement vers μ.

Par hypothèse, quand $n \to +\infty$, $\mu_{\psi(n)}(\mathbb{R}^d) = \widehat{\mu_{\psi(n)}}(0)$ tend vers $\varphi(0)$. Montrons que $\varphi(0) = \widehat{\mu}(0)\ \big(= \mu(\mathbb{R}^d)\big)$.
Soit $\varepsilon > 0$. Par la formule de dualité de la transformée de Fourier, pour toute mesure bornée ν, on a

$$\int_{\mathbb{R}^d} \widehat{\nu}(t)\, \mathbf{1}_{[0,\varepsilon]^d}(t)dt = \int_{\mathbb{R}^d} \widehat{\mathbf{1}_{[0,\varepsilon]^d}}(t)d\nu(t)\ . \tag{1}$$

Comme la fonction $\widehat{\mathbf{1}_{[0,\varepsilon]^d}}$ est continue de limite nulle à l'infini, l'utilisation de la convergence faible et de (1) donne que, quand $n \to +\infty$:

$$\int_{\mathbb{R}^d} \widehat{\mu_{\psi(n)}}(t)\, \mathbf{1}_{[0,\varepsilon]^d}(t)dt = \int_{\mathbb{R}^d} \widehat{\mathbf{1}_{[0,\varepsilon]^d}}(t)d\mu_{\psi(n)}(t) \to \int_{\mathbb{R}^d} \widehat{\mathbf{1}_{[0,\varepsilon]^d}}(t)d\mu(t) = \int_{\mathbb{R}^d} \widehat{\mu}(t)\, \mathbf{1}_{[0,\varepsilon]^d}(t)dt$$

Par ailleurs, d'après le théorème de convergence dominée (domination par $b\,\mathbf{1}_{[0,\varepsilon]^d}$), on a

$$\lim_{n\to+\infty} \int_{\mathbb{R}^d} \widehat{\mu_{\psi(n)}}(t)\, \mathbf{1}_{[0,\varepsilon]^d}(t)dt = \int_{\mathbb{R}^d} \varphi(t)\, \mathbf{1}_{[0,\varepsilon]^d}(t)dt$$

Donc :

$$\int_{\mathbb{R}^d} \widehat{\mu}(t)\, \mathbf{1}_{[0,\varepsilon]^d}(t)dt = \int_{\mathbb{R}^d} \varphi(t)\, \mathbf{1}_{[0,\varepsilon]^d}(t)dt\ .$$

En divisant par ε^d et en effectuant le changement de variables linéaire « $u = \frac{t}{\varepsilon}$ », on obtient

$$\int_{\mathbb{R}^d} \widehat{\mu}(u\varepsilon)\, \mathbf{1}_{[0,1]^d}(u)du = \int_{\mathbb{R}^d} \varphi(u\varepsilon)\, \mathbf{1}_{[0,1]^d}(u)du\ .$$

En faisant tendre $\varepsilon \to 0$, le théorème de convergence dominée donne alors $\widehat{\mu}(0) = \varphi(0)$.

On a finalement montré que $\lim\limits_{n\to+\infty} \mu_{\psi(n)}(\mathbb{R}^d) = \mu(\mathbb{R}^d)$. Par le lemme ci-dessus, $(\mu_{\psi(n)})_{n\in\mathbb{N}}$ converge étroitement vers μ. D'après le premier point du théorème de Lévy, $(\widehat{\mu_{\psi(n)}})_{n\in\mathbb{N}}$ converge donc simplement de limite $\widehat{\mu}$. Par unicité de la limite simple, $\widehat{\mu} = \varphi$.

Par ailleurs, par injectivité de la transformation de Fourier, il existe au plus une mesure bornée dont la transformée de Fourier est égale à φ. Cela prouve donc que $(\mu_n)_{n\in\mathbb{N}}$ admet une unique valeur d'adhérence faible, notée μ. Donc $(\mu_n)_{n\in\mathbb{N}}$ converge faiblement vers μ.

De plus, par convergence simple, quand $n \to +\infty$, $\mu_n(\mathbb{R}^d) = \widehat{\mu_n}(0)$ converge de limite $\widehat{\mu}(0) = \mu(\mathbb{R}^d)$.
Finalement, par le lemme ci-dessus, $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement de limite μ. Comme on a montré précédemment que $\widehat{\mu} = \varphi$, le théorème est donc démontré. □

Remarque : si les μ_n sont des probabilités, alors μ aussi car $\widehat{\mu}(0) = \varphi(0) = \lim \widehat{\mu_n}(0)$.

Remarque : on peut traduire cet énoncé en terme de variables aléatoires. Soit $(Y_n)_{n\in\mathbb{N}}$ une suite de v.a. dans $\mathbb{R}^d$. Si la suite des fonctions caractéristiques $(\varphi_{Y_n})_{n\in\mathbb{N}}$ converge simplement vers une fonction φ continue en 0, alors il existe une unique probabilité μ telle que $\widehat{\mu} = \varphi$ et $Y_n \overset{\text{loi}}{\to} \mu$.

Lemme. Soit X une variable aléatoire réelle admettant un moment d'ordre 2. Quand $t \to 0$, on a :

$$\varphi_X(t) = 1 + it\,\mathbb{E}(X) - \frac{t^2}{2}\,\mathbb{E}(X^2) + o(t^2)\ .$$

- La fonction $t \mapsto e^{itX}$ est de classe $\mathcal{C}^2$ presque sûrement, de dérivées première et seconde $t \mapsto iXe^{itX}$ et $t \mapsto -X^2e^{itX}$ respectivement.
- Pour tout $t \in \mathbb{R}$, on a $|e^{itX}| \leq 1$, $|iXe^{itX}| \leq |X|$ et $|-X^2e^{itX}| \leq X^2$ p.s., où 1, $|X|$ et X^2 sont dans L^1, puisque X admet un moment d'ordre 2.

D'après le théorème de dérivation sous le signe intégral, la fonction caractéristique φ_X est de classe $\mathcal{C}^2$ sur $\mathbb{R}$, de dérivées première et seconde $t \mapsto i\,\mathbb{E}(Xe^{itX})$ et $t \mapsto -\mathbb{E}(X^2e^{itX})$ respectivement.
Or, d'après la formule de Taylor-Young, quand $t \to 0$, on a

$$\varphi_X(t) = \varphi_X(0) + t\varphi_X'(0) + \frac{t^2}{2}\varphi_X''(0) + o(t^2)\ ,$$

ce qui donne la formule annoncée. □

Théorème (central limite). Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de v.a. i.i.d. à valeurs dans $\mathbb{R}^d$, admettant un moment d'ordre 2. On note m la moyenne de X_1 et Σ sa matrice de covariance. Alors :

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - m) \xrightarrow{\text{loi}} \mathcal{N}_d(0, \Sigma)\ .$$

Pour tout $n \in \mathbb{N}^*$, on note

$$Y_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - m)\ .$$

Comme les X_j sont i.i.d., pour tous $n \in \mathbb{N}^*$ et $t \in \mathbb{R}^d$, on a :

$$\begin{aligned}
\varphi_{Y_n}(t) &= \mathbb{E}\left(\exp\left(i\langle t, \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(X_j - m)\rangle\right)\right) \\
&= \mathbb{E}\left(\prod_{j=1}^{n}\exp\left(i\langle t, X_j - m\rangle\frac{1}{\sqrt{n}}\right)\right) \\
&= \left(\mathbb{E}\left(\exp\left(i\langle t, X_1 - m\rangle\frac{1}{\sqrt{n}}\right)\right)\right)^n \\
&= \left(\varphi_{\langle t, X_1 - m\rangle}\left(\frac{1}{\sqrt{n}}\right)\right)^n
\end{aligned}$$

Or, pour tout $t \in \mathbb{R}^d$, la variable aléatoire réelle $\langle X_1 - m, t\rangle$ est centrée et admet un moment d'ordre 2 d'après l'inégalité de Cauchy-Schwarz. Donc, d'après le second lemme, quand $n \to +\infty$,

$$\begin{aligned}
\varphi_{Y_n}(t) &= \left(1 - \frac{1}{2n}\mathbb{E}(\langle X_1 - m, t\rangle^2) + o\left(\frac{1}{n}\right)\right)^n \\
&= \exp\left(n\ln\left(1 - \frac{1}{2n}\mathbb{E}(\langle X_1 - m, t\rangle^2) + o\left(\frac{1}{n}\right)\right)\right) \\
&= \exp\left(-\frac{1}{2}\mathbb{E}(\langle X_1 - m, t\rangle^2) + o(1)\right) \\
&= \exp\left(-\frac{1}{2}\langle \Sigma t, t\rangle + o(1)\right)
\end{aligned}$$

où on a utilisé que $\mathbb{E}(\langle X_1 - m, t\rangle^2) = \langle \Sigma t, t\rangle$ car $\Sigma = \mathbb{E}((X_1 - m).{}^t(X_1 - m))$.

Ainsi, $(\varphi_{Y_n})_{n\in\mathbb{N}}$ converge simplement vers la fonction caractéristique de la loi $\mathcal{N}_d(0, \Sigma)$. Par le théorème de Lévy, $Y_n \overset{\text{loi}}{\to} \mathcal{N}_d(0, \Sigma)$. □

Références

J'ai utilisé [Ouv09, pp. 204, 294, 299, 313]. On peut aussi consulter [CDGCM11, p. 117] et les livres classiques de probabilités.

Leçons correspondantes

J'utilise ce développement pour les leçons 235, 240, 242, 250, 251. On peut également l'utiliser pour les leçons 234, 241.

Remarques

- Dans un espace topologique, il n'y a pas équivalence entre valeur d'adhérence et sous-suite convergente. C'est par contre le cas dans un espace métrique.
- On peut en fait traiter seulement le cas de $\mathcal{M}_1(\mathbb{R}^d)$ au lieu de $\mathcal{M}_b(\mathbb{R}^d)$ pour $b > 0$. Avec les notations du développement, il faut d'ailleurs veiller à ne pas confondre l'ensemble des mesures de masse inférieure ou égale à 1 sur $\mathbb{R}^d$ et l'ensemble des probabilités sur $\mathbb{R}^d$, qui lui n'est pas compact !

– Pour démontrer le théorème central limite, on a en fait seulement besoin de la version faible du théorème de Lévy, c'est-à-dire que le second point est remplacé par : « s'il existe $\mu \in \mathcal{M}_b(\mathbb{R}^d)$ telle que $(\widehat{\mu_n})_{n\in\mathbb{N}}$ converge simplement vers $\widehat{\mu}$, alors $(\mu_n)_{n\in\mathbb{N}}$ converge étroitement vers μ ».
– Pour ne pas dépasser le temps imparti, il vaut mieux occulter les démonstrations des lemmes, qui ne sont pas le cœur du développement.
– À la rigueur, on peut utiliser ce développement dans les leçons 203, 218, 239, mais c'est un peu juste.
– Le théorème central limite permet de retrouver la formule de Stirling, en considérant des v.a. exponentielles (voir [FF12, p. 243], [GK11, p. 283]) ou de Poisson (voir [GL08, p. 194]). Pour d'autres applications, voir par exemple [CDGCM11].

Questions possibles

1. Définir les différents modes de convergence des mesures et leurs liens.
2. Expliquer comment on peut obtenir une suite $(g_p)_{p\in\mathbb{N}}$ comme dans le premier lemme.
3. Justifier l'injectivité de la transformation de Fourier.
4. Justifier que $\mathcal{M}_b(\mathbb{R}^d)$ est métrisable et compact.
5. Définir la fonction caractéristique d'une loi et calculer celle d'une gaussienne.
6. Montrer que quand $n \to +\infty$,
$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$$
tend vers $\frac{1}{2}$.

Références

[CDGCM11] Marie Cottrell, Christian Duhamel, Valentine Genon-Catalot et Thierry Meyre : *Exercices de probabilités.* Cassini, 2011.

[FF12] Dominique Foata et Aimé Fuchs : *Calcul des probabilités.* Dunod, 2012.

[GK11] Olivier Garet et Aline Kurtzmann : *De l'intégration aux probabilités.* Ellipses, 2011.

[GL08] Valérie Girardin et Nikolaos Limnios : *Probabilités en vue des applications : variables, vecteurs et suites aléatoires.* Vuibert, 2008.

[Ouv09] Jean-Yves Ouvrard : *Probabilités 2.* Cassini, 2009.