1.15 Théorème limite central et intervalles de confiance asymptotiques (218, 261, 262, 266) [4], [26]

Le théorème limite central est, comme son nom l'indique, le théorème limite central en probabilités, qui unifie le comportement de toutes les variables aléatoires indépendantes et de même loi en une unique loi de probabilité : la loi normale. Ce résultat est notamment utilisé en statistiques pour déterminer des intervalles de confiance asymptotiques pour des paramètres d'intérêt de certaines lois de probabilité.

Théorème 1.27 (Théorème limite central)**.** Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles de variance finie, indépendantes et identiquement distribuées. Alors, en notant m l'espérance commune de ces variables et σ^2 leur variance, on a :

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - m) \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}(0,\sigma^2).$$

Démonstration. La preuve de ce théorème se base sur une version faible du théorème de Lévy qui est la suivante :

Proposition 1.28 (Théorème de Lévy, version faible)**.** Soient $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires réelles et X une autre variable aléatoire, toutes définies sur le même espace de probabilité. Alors il y a équivalence entre :

1. $X_n \xrightarrow[n\to+\infty]{\mathscr{L}} X$,
2. $\varphi_{X_n} \xrightarrow[n\to+\infty]{\text{CVS}} \varphi_X$.

Démonstration de la version faible du théorème de Lévy. L'implication 1. ⇒ 2. est claire car les fonctions $f_t : x \mapsto e^{itx}$ sont continues et bornées. Montrons l'implication réciproque.

Étape 1 : 2. ⇒ $\mathbb{E}[f(X_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}[f(X)]$ pour toute fonction f continue tendant vers 0

Soient $\varepsilon > 0$ et soit $f \in \mathscr{C}_0^0(\mathbb{R})$. On a alors que f est uniformément continue sur $\mathbb{R}$ et donc elle est approchable par des fonctions régulières en norme infini, disons dans la classe de Schwartz : il existe $f_\varepsilon \in \mathscr{S}(\mathbb{R})$ telle que $\|f - f_\varepsilon\|_\infty < \varepsilon$. Ainsi, on a :

$$\begin{aligned}
|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leqslant |\mathbb{E}[f_\varepsilon(X_n)] - \mathbb{E}[f_\varepsilon(X)]| + |\mathbb{E}[(f-f_\varepsilon)(X_n)]| + |\mathbb{E}[(f-f_\varepsilon)(X)]| \\
&\leqslant |\mathbb{E}[f_\varepsilon(X_n)] - \mathbb{E}[f_\varepsilon(X)]| + 2\|f - f_\varepsilon\|_\infty \\
&< |\mathbb{E}[f_\varepsilon(X_n)] - \mathbb{E}[f_\varepsilon(X)]| + 2\varepsilon
\end{aligned}$$

Il suffira alors de montrer que $\mathbb{E}[f_\varepsilon(X_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}[f_\varepsilon(X)]$ pour conclure cette étape ! Pour cela, on utilise le théorème d'inversion de Fourier dans Schwartz :

$$\forall x \in \mathbb{R}, \quad f_\varepsilon(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}_\varepsilon(\xi) e^{ix\xi}\mathrm{d}\xi.$$

En notant $g := \frac{1}{2\pi}\hat{f}_\varepsilon$, on a que $g \in \mathscr{S}(\mathbb{R})$ et :

$$\mathbb{E}[f_\varepsilon(X_n)] = \mathbb{E}\left[\int_{\mathbb{R}} g(\xi)e^{i\xi X_n}\mathrm{d}\xi\right].$$

Le théorème de Fubini peut alors s'appliquer, car on peut borner l'intégrande par $|g|$ qui est intégrable sur $\mathbb{R}$ car dans $\mathscr{S}(\mathbb{R})$, et les constantes sont également intégrables pour les mesures de probabilité. On a donc :

$$\mathbb{E}[f_\varepsilon(X_n)] = \int_{\mathbb{R}} g(\xi)\varphi_{X_n}(\xi)\mathrm{d}\xi.$$

Là encore, puisque g est dans la classe de Schwartz, le théorème de convergence dominée peut s'appliquer et donc, par convergence simple des φ_{X_n} vers φ_X :

$$\mathbb{E}[f_\varepsilon(X_n)] \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}} g(\xi)\varphi_X(\xi)\mathrm{d}\xi = \int_{\mathbb{R}} g(\xi)\mathbb{E}\left[e^{i\xi X}\right]\mathrm{d}\xi = \mathbb{E}\left[\int_{\mathbb{R}} g(\xi)e^{iX\xi}\mathrm{d}\xi\right] = \mathbb{E}[f_\varepsilon(X)],$$

ce qui permet de conclure cette étape ! En effet, on a alors :

$$\limsup_{n\to+\infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leqslant \limsup_{n\to+\infty} |\mathbb{E}[f_\varepsilon(X_n)] - \mathbb{E}[f_\varepsilon(X)]| + 2\varepsilon = 2\varepsilon,$$

et ce, pour tout $\varepsilon > 0$.

Étape 2 : Le fait établi à l'étape 1 implique la convergence en loi

Maintenant, il ne reste plus qu'à montrer que si pour tout $f \in \mathscr{C}_0^0(\mathbb{R})$, $\mathbb{E}[f(X_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}[f(X)]$, alors pour tout $f \in \mathscr{C}_b^0(\mathbb{R})$, $\mathbb{E}[f(X_n)] \xrightarrow[n\to+\infty]{} \mathbb{E}[f(X)]$. Prenons alors $f \in \mathscr{C}_b^0(\mathbb{R})$. On va se servir du fait qu'on travaille avec une mesure de probabilité pour tronquer notre fonction f grâce à une fonction $\varphi \in \mathscr{C}_0^0(\mathbb{R})$ bien choisie en perdant le moins de masse possible. Soit $\varepsilon > 0$. Puisque $\mathbb{P}_X$ (loi de X) est une mesure de probabilité sur $\mathbb{R}$, il existe $A > 0$ tel que $\mathbb{P}_X(\mathbb{R} \setminus [-A, A]) < \varepsilon$. Si alors $\varphi \in \mathscr{C}_c^0(\mathbb{R})$ est une fonction telle que :

— $\varphi \equiv 1$ sur $[-A, A]$,
— $0 \leqslant \varphi \leqslant 1$ sur $\mathbb{R}$,
— $\varphi \equiv 0$ sur $\mathbb{R} \setminus [-2A, 2A]$,

alors on a :

$$\mathbb{E}[1 - \varphi(X)] \leqslant \mathbb{P}_X(\mathbb{R} \setminus [-A, A]) < \varepsilon.$$

Ainsi, on a :

$$|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leqslant |\mathbb{E}[f(X_n)(1 - \varphi(X_n))]| + \underbrace{|\mathbb{E}[f\varphi(X_n)] - \mathbb{E}[f\varphi(X)]|}_{\xrightarrow[n\to+\infty]{} 0 \text{ car } f\varphi \in \mathscr{C}_0^0(\mathbb{R})} + \underbrace{|\mathbb{E}[f(X)(\varphi(X) - 1)]|}_{\leqslant \|f\|_\infty \mathbb{E}[1-\varphi(X)] < \varepsilon\|f\|_\infty}$$

Les deux derniers termes ont été faciles à analyser comme vous avez pu le constater. Le premier n'est pas plus dur mais demande plus de rédaction :

$$|\mathbb{E}[f(X_n)(1 - \varphi(X_n))]| \leqslant \|f\|_\infty \mathbb{E}[1 - \varphi(X_n)] = \|f\|_\infty (1 - \mathbb{E}[\varphi(X_n)]) \xrightarrow[n\to+\infty]{} \|f\|_\infty (1 - \mathbb{E}[\varphi(X)]) < \varepsilon\|f\|_\infty.$$

En définitive, on obtient :

$$\limsup_{n\to+\infty} |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| \leqslant 2\varepsilon\|f\|_\infty,$$

et ce pour tout $\varepsilon > 0$! On a donc montré le théorème de Lévy ! □

On va pouvoir s'attaquer à la preuve du théorème limite central. Quitte à considérer les variables aléatoires $X_n - m$ qui sont i.i.d. et de variance σ^2, on peut supposer $m = 0$. Posons :

$$\forall n \in \mathbb{N}^*, \quad S_n := \sum_{k=1}^{n} X_k.$$

Le but est alors de montrer que $\varphi_{\frac{S_n}{\sqrt{n}}} \xrightarrow[n\to+\infty]{\text{CVS}} \varphi_{\mathcal{N}(0,\sigma^2)} = t \mapsto e^{-\frac{\sigma^2 t^2}{2}}$. Par inépendance des variables X_n, on a :

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t) = \mathbb{E}\left[\prod_{k=1}^{n} e^{it\frac{X_k}{\sqrt{n}}}\right] = \prod_{k=1}^{n} \mathbb{E}\left[e^{it\frac{X_k}{\sqrt{n}}}\right] = \prod_{k=1}^{n} \varphi_{X_k}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

étant donné que les X_k suivent la même loi. Or, $X_1 \in \mathrm{L}^2(\Omega, \mathbb{P})$. Ainsi, sa fonction caractéristique est de classe $\mathscr{C}^2$ en 0 et on peut donc utiliser la formule de Taylor-Young :

$$\forall t \in \mathbb{R}, \quad \varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \underbrace{\varphi_{X_1}(0)}_{=1} + \frac{t}{\sqrt{n}} \underbrace{\varphi_{X_1}'(0)}_{=i\mathbb{E}[X_1]=0} + \frac{t^2}{2n} \underbrace{\varphi_{X_1}''(0)}_{=-\mathbb{E}[X_1^2]=-\sigma^2} + \underset{n\to+\infty}{o}\left(\frac{1}{n}\right) = 1 - \frac{\sigma^2 t^2}{2n} + \underset{n\to+\infty}{o}\left(\frac{1}{n}\right)$$

On observe alors que pour $t \in \mathbb{R}$ fixé, $\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)$ prendra ses valeurs dans un disque ouvert centré en 1 et de rayon, disons, 1 et donc en particulier dans $\mathbb{C} \setminus \mathbb{R}^-$, où on a la branche principale du logarithme complexe. On peut donc définir, à partir d'un certain rang :

$$\log\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n\right) = n\log\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right) = n\log\left(1 - \frac{\sigma^2 t^2}{2n} + \underset{n\to+\infty}{o}\left(\frac{1}{n}\right)\right) = -\frac{\sigma^2 t^2}{2} + \underset{n\to+\infty}{o}(1).$$

Ainsi, on obtient par continuité de l'exponentielle :

$$\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \exp\left(n\log\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)\right) \xrightarrow[n\to+\infty]{} e^{-\frac{\sigma^2 t^2}{2}}$$

ce qui montre le théorème central limite grâce au théorème de Lévy ! □

On peut, si le temps le permet, déterminer des intervalles de confiance asymptotiques pour le paramètre d'une loi de Bernoulli ou de Poisson. Ici, je détaille pour une loi de Bernoulli (Sacha Quayle fait pour une loi de Poisson, allez voir !)

Corollaire 1.29. On considère le modèle statistique suivant, associé à un n-échantillon $(X_1, \ldots, X_n)$ de loi de Bernoulli de paramètre inconnu p :

$$\left(\{0,1\}^n, \mathscr{P}\left(\{0,1\}^n\right), (\mathscr{B}(p))_{p\in(0,1)}\right).$$

Alors, si $\alpha \in (0,1)$, en notant q_α le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale $\mathscr{N}(0,1)$, l'intervalle :

$$\left[\overline{X}_n - q_\alpha\sqrt{\frac{\overline{X}_n\left(1-\overline{X}_n\right)}{n}}, \overline{X}_n + q_\alpha\sqrt{\frac{\overline{X}_n\left(1-\overline{X}_n\right)}{n}}\right]$$

constitue un intervalle de confiance de niveau asymptotique $1-\alpha$ du paramètre d'intérêt p. Ici, on a noté :

$$\overline{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Démonstration. Par le théorème limite central, on sait que, si notre n-échantillon suit la loi $\mathscr{B}(p)$:

$$\sqrt{n}\left(\overline{X}_n - p\right) \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}\left(0, p(1-p)\right).$$

Ainsi, en renormalisant, puisque $p \notin \{0,1\}$, on a :

$$\sqrt{\frac{n}{p(1-p)}}\left(\overline{X}_n - p\right) \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}\left(0,1\right).$$

Si on s'arrête là, on fera face à un problème : l'intervalle de confiance qu'on trouverait dépendra de p à cause du terme $\frac{n}{p(1-p)}$! Il faut donc pouvoir "éliminer" ce terme dans la limite en loi, pour trouver ce qu'on appelle une statistique "asymptotiquement pivotale", c'est-à-dire dont la loi limite ne dépend pas du paramètre d'intérêt p. On va alors utiliser l'artillerie classique des statistiques : la loi des grands nombres et le lemme de Slutsky. On sait par la loi forte des grands nombres que :

$$\overline{X}_n \xrightarrow[n\to+\infty]{\text{p.s.}} p.$$

Ainsi, étant donné que la convergence presque sûre passe aux fonctions continues, on a :

$$\sqrt{\overline{X}_n\left(1-\overline{X}_n\right)} \xrightarrow[n\to+\infty]{\text{p.s.}} \sqrt{p(1-p)},$$

qui est une constante ! Ainsi, par le lemme de Slutsky, on a la convergence en loi suivante :

$$\left(\sqrt{\frac{n}{p(1-p)}}\left(\overline{X}_n - p\right), \sqrt{\overline{X}_n\left(1-\overline{X}_n\right)}\right) \xrightarrow[n\to+\infty]{\mathscr{L}} \left(Z, \sqrt{p(1-p)}\right)$$

où Z est une variable aléatoire suivant la loi $\mathscr{N}(0,1)$. De plus, $p(1-p)$ est non-nul. Ainsi, par convergence presque sûre, il existe un événement presque sûr Ω' tel que :

$$\forall \omega \in \Omega',\ \exists n_0(\omega) \in \mathbb{N}^*,\ \forall n \geqslant n_0(\omega), \quad \overline{X}_n(\omega)\left(1-\overline{X}_n(\omega)\right) \neq 0.$$

Ainsi, on va pouvoir diviser et invoquer le *continuous mapping theorem* via l'application continue :

$$\begin{array}{rccc} g & : & \mathbb{R}\times\mathbb{R}^* & \longmapsto & \mathbb{R} \\ & & (x,y) & \longmapsto & \dfrac{x\sqrt{p(1-p)}}{y} \end{array}$$

pour pouvoir affirmer que :

$$\sqrt{\frac{n}{\overline{X}_n\left(1-\overline{X}_n\right)}}\left(\overline{X}_n - p\right) \xrightarrow[n\to+\infty]{\mathscr{L}} \mathscr{N}(0,1).$$

Cette étape m'a paru délicate à justifier alors je détaille : prenons $f \in \mathscr{C}_b^0(\mathbb{R})$ une fonction-test. On a alors :

$$\begin{aligned} \left|\mathbb{E}\left[f\left(\sqrt{\frac{n}{\overline{X}_n\left(1-\overline{X}_n\right)}}\left(\overline{X}_n - p\right)\right)\right] - \mathbb{E}[Z]\right| &\leqslant \left|\mathbb{E}\left[f\left(\sqrt{\frac{n}{\overline{X}_n\left(1-\overline{X}_n\right)}}\left(\overline{X}_n - p\right)\right)\mathbb{1}_{\overline{X}_n\notin\{0,1\}}\right] - \mathbb{E}[Z]\right| \\ &\quad +\|f\|_\infty \mathbb{E}\left[\mathbb{1}_{\overline{X}_n\notin\{0,1\}}\right] \\ &= \left|\mathbb{E}\left[f\left(\sqrt{\frac{n}{\overline{X}_n\left(1-\overline{X}_n\right)}}\left(\overline{X}_n - p\right)\right)\mathbb{1}_{\overline{X}_n\notin\{0,1\}}\right] - \mathbb{E}[Z]\right| \\ &\quad +\|f\|_\infty \mathbb{P}\left(\overline{X}_n \in \{0,1\}\right). \end{aligned}$$

Or, puisque $\overline{X}_n$ devient presque sûrement, à partir d'un certain rang aléatoire, différent de 0 ou 1, on a bien :

$$\mathbb{1}_{\overline{X}_n\notin\{0,1\}} \xrightarrow[n\to+\infty]{\text{p.s.}} 1.$$

Ainsi, par convergence dominée par exemple, on a bien que le premier terme tend vers 0, tout comme le deuxième terme. Bref. Une fois qu'on a cette convergence en loi, on a donc, par symétrie de la loi normale centrée réduite :

$$\mathbb{P}\left(\sqrt{\frac{n}{\overline{X}_n\left(1-\overline{X}_n\right)}}\left(\overline{X}_n-p\right)\in[-p_\alpha,p_\alpha]\right)\xrightarrow[n\to+\infty]{}1-\frac{\alpha}{2}-\frac{\alpha}{2}=1-\alpha.$$

En réarrangeant les termes dans la probabilité, on obtient donc :

$$\mathbb{P}\left(\overline{X}_n-p\in\left[-p_\alpha\sqrt{\frac{\overline{X}_n\left(1-\overline{X}_n\right)}{n}},p_\alpha\sqrt{\frac{\overline{X}_n\left(1-\overline{X}_n\right)}{n}}\right]\right)\xrightarrow[n\to+\infty]{}1-\alpha$$

ce qui donne bien :

$$\mathbb{P}\left(p\in\left[\overline{X}_n-p_\alpha\sqrt{\frac{\overline{X}_n\left(1-\overline{X}_n\right)}{n}},\overline{X}_n+p_\alpha\sqrt{\frac{\overline{X}_n\left(1-\overline{X}_n\right)}{n}}\right]\right)\xrightarrow[n\to+\infty]{}1-\alpha,$$

ce qu'on voulait démontrer ! $\square$