Théorèmes asymptotiques

Loi des grands nombres

Le premier résultat fondamental en probabilités concerne le comportement asymptotique de la moyenne empirique:

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \quad .$$

quand on observe n variables aléatoires i.i.d $X_1, \ldots, X_n$, ayant une espérance finie.

Théorème 0.1 (Loi forte des grands nombres)**.** *Soit $(X_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) dans $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Notons $\mu = \mathbb{E}[X_1]$. Alors $\bar{X}_n$ converge vers μ presque sûrement :*

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n}{n} \underset{n\to\infty}{\longrightarrow} \mu\right) = 1\,.$$

Interprétation: Intuitivement, la probabilité d'un événement A correspond à la fréquence d'apparition de A quand on répète une expérience qui fait intervenir cet événement. Par exemple, si on dispose une pièce truquée, on estimera la probabilité d'apparition du côté pile en lançant la pièce un grand nombre de fois et en comptant le nombre de pile obtenu. La loi des grands nombres justifie a posteriori cette intuition : si $X_1, \ldots, X_n$ sont i.i.d. de loi de Bernoulli de paramètre p, alors

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

Le membre de gauche correspond au nombre empirique de pile obtenu, celui de droite à la valeur théorique.

Remarque: Bien qu'assez intuitif, ce théorème est difficile à démontrer, cf.(Ouvrard 2008; Barbe et Ledoux 2006) ou encore (Williams 1991) pour une version de preuve avec des martingales.

Pour aller plus loin:

Quant p varie, à n fixé…les signaux générés sont très très proches, ce qui ne devrait pas être le cas sans structuration particulière de la génération. L'aléa est imparfait!

Théorème central limite (TCL)

Une fois la loi des grands nombres établie, on peut se demander quel est l'ordre suivant dans le développement asymptotique de $\bar{X}_n - \mu$, ou de manière équivalente de $S_n - n\mu$, où $S_n = X_1 + \cdots + X_n$. Le théorème suivant répond à cette question, en donnant une convergence en loi d'une transformation affine de la moyenne empirique:

Théorème 0.2 (Théorème central limite)**.** *Soit $X_1, \ldots, X_n$ une suite de variables aléatoires i.i.d de variance $\sigma^2 = \mathrm{var}(X_1) \in]0, \infty[$. On note $\mu = \mathbb{E}[X_1]$ leur espérance. Alors*

$$\sqrt{n}\left(\tfrac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow[n\to+\infty]{\mathcal{L}} N \ ,$$

où N suit une loi normale centrée réduite : $N \sim \mathcal{N}(0,1)$.

Preuve: cf.(Ouvrard 2008; Barbe et Ledoux 2006).

On peut interpréter ce théorème grossièrement de la façon suivante: la moyenne empirique de variables aléatoires i.i.d de variance σ^2 se comporte asymptotiquement comme une loi normale $\mathcal{N}(\mu, \frac{\sigma^2}{n})$, ce que l'on écrit avec un abus de notation:

$$\bar{X}_n \approx \mathcal{N}(\mu, \frac{\sigma^2}{n}) \ .$$

En termes de somme cumulée empirique, la convergence se réécrit

$$\tfrac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n\to+\infty]{\mathcal{L}} N \ .$$

Les hypothèses de ce théorème sont plutôt faibles (il suffit de supposer une variance finie). Pourtant, le résultat est universel : la loi de départ peut être aussi farfelue que l'on veut, elle se rapprochera toujours asymptotiquement d'une loi normale.

On rappelle que la convergence en loi est équivalente à la convergence des fonctions de répartition en tout point de continuité de la limite. Ainsi, le théorème central limite se réécrit de la manière suivante : pour tout $a < b$, notons $\alpha_n = \mathbb{P}\left(\bar{X}_n \notin [\mu + \frac{a\sigma}{\sqrt{n}}, \mu + \frac{b\sigma}{\sqrt{n}}]\right)$. Ainsi,

$$\begin{aligned}
1 - \alpha_n &= \mathbb{P}\left(\bar{X}_n \in [\mu + \tfrac{a\sigma}{\sqrt{n}}, \mu + \tfrac{b\sigma}{\sqrt{n}}]\right) \\
&= \mathbb{P}\left(\tfrac{\bar{X}_n - \mu}{\sigma} \in [\tfrac{a}{\sqrt{n}}, \tfrac{b}{\sqrt{n}}]\right) \\
&= \mathbb{P}\left(a \leq \sqrt{n}\left(\tfrac{\bar{X}_n - \mu}{\sigma}\right) \leq b\right) \\
&\underset{n\to\infty}{\longrightarrow} \int_a^b \varphi(x)\, dx \,.
\end{aligned} \tag{1}$$

où l'on note φ (resp. Φ) la densité (resp. la fonction de répartition) d'une loi normale centrée réduite, définie pour tout $x \in \mathbb{R}$ par $\varphi(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ (resp. $\Phi(x) = \int_{-\infty}^{x} \varphi(u)du$).

Dans le cas classique d'un intervalle de confiance à 95%, c'est-à-dire quand $\alpha_n = 0.05$, et en prenant un intervalle de confiance symétrique (alors $a = -t$ et $b = q$) on obtient $1 - \alpha_n = \int_{-q}^{q} \varphi(x)\, dx = \Phi(q) - \Phi(-q) = 2\Phi(q) - 1 \implies \boxed{q = \Phi^{-1}(1 - \frac{\alpha_n}{2})}$ et q est donc le quantile de niveau $1 - \frac{\alpha_n}{2}$ de la loi normale centrée réduite. Numériquement on peut facilement évaluer q et vérifier que $q \approx 1.96$ avec `scipy`:

```
from scipy.stats import norm
q = norm.ppf(1-0.05/2)
print(f"Gaussienne centrée réduite,\nQuantile de niveau (1- /2):\nq = {q:.2f}")
```

```
Gaussienne centrée réduite,
Quantile de niveau (1- /2):
q = 1.96
```

Exemple 0.1 (Loi de Bernoulli)**.** On considère des variables aléatoires $X_1, \dots, X_n$ i.i.d. suivant une loi de Bernoulli de paramètre $p \in]0, 1[$, dont l'espérance et la variance sont respectivemenbt p et $p(1-p)$. Le théorème central limite donne alors

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

avec $N \sim \mathcal{N}(0, 1)$. Cette convergence est illustrée dans le widget ci-dessous. Le contexte est le suivant. On répète t fois le processus, qui consiste à afficher $(\bar{X}_k)_{k\in[n]}$, où les n variables aléatoires sont i.i.d. et suivent une loi de Bernoulli de paramètre p.

Une autre illustration possible de la convergence donnée par le TCL est celle qui correspond au point de vue donnée par l'analyse. Pour cela supposons que l'on ait une suite de variables aléatoires réelles $X_1, \dots, X_n$, i.i.d. dont la fonction de densité commune est notée par f.

On rappelle quelques éléments de probabilités concernant les densités. Pour cela on rappelle la définition de la convolution deux fonctions. Pour cela prenons deux fonctions f et g définies sur $\mathbb{R}$ et qui sont intégrables au sens de Lebesgue. La **convolution** de f par g est alors la fonction $f * g$ suivante:

$$\begin{aligned} &\mathbb{R} \mapsto \mathbb{R} \\ &x \to \int_{-\infty}^{+\infty} f(x-y)g(y)dy\ . \end{aligned}$$

Note

On peut aussi obtenir $f * g(x)$ en calculant $\int_{\mathbb{R}^2} f(u)g(v)\mathbb{1}_{u+v=x} dudv$.

Théorème 0.3 (Loi de la somme et convolutions)**.** *Soient X et Y des v.a. indépendantes de densités f et g respectivement, la loi de $X+Y$ est donnée par la convolution $f * g$.*

Rappel: pour un scalaire $\alpha \neq 0$, la densité de αX est donnée par la fonction $x \mapsto \frac{1}{|\alpha|} \cdot f(\frac{x}{\alpha})$.

Corollaire 0.1 (Loi de la moyenne)**.** *Soient $X_1, \dots, X_n$ des v.a. i.i.d. de densité f, la densité de $\bar{X}_n$ est donnée par la fonction $x \mapsto n \cdot [f * \cdots * f](n \cdot x)$.*

Dessous, pour $X_1, \dots, X_n$, i.i.d., de densité f, on affiche la densité de la loi de $\bar{X}_n$.

Pour aller plus loin sur les convolutions, voir la vidéo de 3Blue1Brown à ce sujet: Convolutions | Why X+Y in probability is a beautiful mess

Barbe, Philippe, et Michel Ledoux. 2006. *Probabilités*.

Ouvrard, J.-Y. 2008. *Probabilités : Tome 1, Licence - CAPES*. 2 éd. Enseignement des mathématiques. Cassini.

Williams, D. 1991. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge: Cambridge University Press.