Théorème central limite et intervalle de confiance

Dorian Cacitti-Holland

2020-2021

Références.

1. Probabilités de Barbe et Ledoux (pour le lemme et le théorème)
2. Probabilités et statistiques (pour le corollaire)

Leçons.

1. 261 : Loi d'une variable aléatoire, caractérisations, exemples, applications
2. 262 : Convergences d'une suite de variables aléatoires, théorèmes limite, exemples et applications
3. 266 : Illustration de la notion d'indépendance en probabilités

Théorème. Soit $(X_n)_{n\in\mathbb{N}}$ variables aléatoires indépendantes identiquement distribuées tels que $\mathbb{E}(X_1^2) < +\infty$ et $Var(X_1) \neq 0$, alors

$$\frac{1}{\sqrt{nVar(X_1)}}\left(\sum_{k=1}^{n} X_k - n\mathbb{E}(X_1)\right) \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$$

Démonstration.
Etape 1 : Si $\mathbb{E}(X_1) = 0, Var(X_1) = 1$
Alors, par indépendance des X_n et identique distribution,

$$\forall t \in \mathbb{R}, \varphi_{\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n} X_k}(t) = \varphi_{\sum\limits_{k=1}^{n} X_k}\left(\frac{t}{\sqrt{n}}\right) = \left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Or $\mathbb{E}(X_1^2) < +\infty$, donc φ_X est deux fois dérivable avec

$$\varphi_X'(0) = \mathbb{E}(X) = 0, \varphi_X''(0) = -\mathbb{E}(X^2) = -1$$

Donc, par formule de Taylor-Young,

$$\varphi_X(u) \underset{u\to 0}{=} 1 - \frac{u^2}{2} + o(u^2)$$

Ainsi

$$\varphi_{\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n} X_k}(t) \underset{n\to+\infty}{=} \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n$$

Pour continuer et manipuler correctement ce o complexe, nous allons avoir besoin du lemme suivant :

Lemme. Soit $(z_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$ tel que $z_n \underset{n\to+\infty}{\longrightarrow} 0$, alors

$$\left(1+\frac{z_n}{n}\right)^n \underset{n\to+\infty}{\longrightarrow} 1$$

Démonstration. On a, pour $n \in \mathbb{N}^*$,

$$\left(1+\frac{z_n}{n}\right)^n - 1 = \sum_{k=1}^{n} \binom{n}{k} \left(\frac{z_n}{n}\right)^k$$

D'où

$$\left|\left(1+\frac{z_n}{n}\right)^n - 1\right| \leq \sum_{k=0}^{n-1} \binom{n}{k+1} \left(\frac{|z_n|}{n}\right)^{k+1} = \sum_{k=0}^{n-1} \binom{n-1}{k} \left(\frac{|z_n|}{n}\right)^k \frac{|z_n|}{k+1} \leq |z_n| \left(1+\frac{|z_n|}{n}\right)^{n-1}$$

Avec, comme $z_n \underset{n\to+\infty}{\longrightarrow} 0$,

$$\left(1+\frac{|z_n|}{n}\right)^{n-1} = e^{(n-1)ln\left(1+\frac{|z_n|}{n}\right)} \underset{n\to+\infty}{\sim} e^{(n-1)\frac{|z_n|}{n}} \underset{n\to+\infty}{\sim} e^{|z_n|}$$

D'où

$$\left(1+\frac{|z_n|}{n}\right)^{n-1} \underset{n\to+\infty}{\longrightarrow} 1$$

Puis, par l'inégalité précédente,

$$\left(1+\frac{z_n}{n}\right)^n \underset{n\to+\infty}{\longrightarrow} 1$$

□

On écrit alors

$$\left(1-\frac{t^2}{2n}+o\left(\frac{1}{n}\right)\right)^n = \left(1-\frac{t^2}{2n}+\frac{z_n}{n}\right)^n = \frac{1}{\left(1-\frac{t^2}{2n}\right)^n}\left(1+\frac{\frac{z_n}{1-\frac{t^2}{2n}}}{n}\right)^n = \frac{1}{\left(1-\frac{t^2}{2n}\right)^n}\left(1+\frac{z'_n}{n}\right)^n$$

Avec $z'_n = \frac{z_n}{1-\frac{t^2}{2n}} \underset{n\to+\infty}{\longrightarrow} 0$ (car $z_n \underset{n\to+\infty}{=} o(1)$) et $\frac{1}{\left(1-\frac{t^2}{2n}\right)^n} = e^{-nln\left(1-\frac{t^2}{2n}\right)} \underset{n\to+\infty}{\longrightarrow} e^{\frac{t^2}{2}}$, donc, d'après le lemme précédent,

$$\varphi_{\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n} X_k}(t) \underset{n\to+\infty}{\longrightarrow} e^{-\frac{t^2}{2}}$$

Or, pour $X \sim \mathcal{N}(0,1)$, $\varphi_X(t) = e^{-\frac{t^2}{2}}$, donc, par théorème de Lévy :

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$$

<u>Etape 2 : Cas général</u>

On considère les variables aléatoires $Y_n = \frac{X_n-\mathbb{E}(X_n)}{\sqrt{Var(X_n)}}$, on a les Y_n vérifient les hypothèses de l'étape 1, d'où

$$\frac{1}{\sqrt{nVar(X_1)}}\left(\sum_{k=1}^{n} X_k - n\mathbb{E}(X_1)\right) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$$

□

Corollaire. Si $\sigma^2 := Var(X_1)$ est connue, alors la moyenne empirique $\hat{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$ est un estimateur sans biais, fortement consistent et asymptotiquement normal de $m = \mathbb{E}(X_1)$ et

$$ICA_{1-\alpha}(m) = \left[\hat{X}_n - \sqrt{\frac{n}{\sigma^2}}q_{1-\frac{\alpha}{2}}, \hat{X}_n + \sqrt{\frac{n}{\sigma^2}}q_{1-\frac{\alpha}{2}}\right]$$

avec $q_{1-\frac{\alpha}{2}}$ quantile d'ordre $1-\frac{\alpha}{2}$ de $\mathcal{N}(0,1)$ et $Var(X_1)$ connue.
On obtient ainsi un test pour $\mathbb{E}(X_1)$: On ne rejette pas $H_0 : m = m_0$ contre $H_1 : m \neq m_0$ au niveau de confiance α si

$$-q_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}}{\sqrt{\sigma^2}}\left(\hat{X}_n - m_0\right) \leq q_{1-\frac{\alpha}{2}}$$

Démonstration.
Le caractère sans biais vient de l'identique distribution et de la linéarité de l'espérance.
La forte consistence vient de la loi forte des grands nombres.
L'asymptotique normalité vient du théorème précédent :

$$\frac{1}{\sqrt{n\sigma^2}}\left(\sum_{k=1}^{n} X_k - n\mathbb{E}(X_1)\right) \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$$

ie

$$\frac{\sqrt{n}}{\sqrt{\sigma^2}}\left(\hat{X}_n - \mathbb{E}(X_1)\right) \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,1)$$

Puis si on considère $Z \sim \mathcal{N}(0,1)$, alors on a

$$\lim_{n\to+\infty} \mathbb{P}\left(-q_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}}{\sqrt{\sigma^2}}\left(\hat{X}_n - \mathbb{E}(X_1)\right) \leq q_{1-\frac{\alpha}{2}}\right) = \mathbb{P}\left(-q_{1-\frac{\alpha}{2}} \leq Z \leq q_{1-\frac{\alpha}{2}}\right)$$

Or $Z \overset{loi}{=} -Z$, donc

$$\mathbb{P}\left(-q_{1-\frac{\alpha}{2}} \leq Z \leq q_{1-\frac{\alpha}{2}}\right) = 1 - \mathbb{P}(Z < -q_{1-\frac{\alpha}{2}}) - \mathbb{P}(Z > q_{1-\frac{\alpha}{2}}) = 1 - 2\left(1 - \frac{\alpha}{2}\right) = \alpha$$

D'où

$$\lim_{n\to+\infty} \mathbb{P}\left(\hat{X}_n - \sqrt{\frac{n}{\sigma^2}}q_{1-\frac{\alpha}{2}} \leq \mathbb{E}(X_1) \leq \hat{X}_n + \sqrt{\frac{n}{\sigma^2}}q_{1-\frac{\alpha}{2}}\right)$$
$$= \lim_{n\to+\infty} \mathbb{P}\left(-q_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}}{\sqrt{\sigma^2}}\left(\hat{X}_n - \mathbb{E}(X_1)\right) \leq q_{1-\frac{\alpha}{2}}\right) = \alpha$$

Ainsi

$$ICA_{1-\alpha}(\mathbb{E}(X_1)) = \left[\hat{X}_n - \sqrt{\frac{n}{\sigma^2}}q_{1-\frac{\alpha}{2}}, \hat{X}_n + \sqrt{\frac{n}{\sigma^2}}q_{1-\frac{\alpha}{2}}\right]$$

$\square$