# Théorème central limite et applications

Soit $X$ v.a. associée à une expérience aléatoire.

Soient $\mathbb{E}(X)=\mu$, $\mathrm{Var}(X)=\sigma^2$.

On répète l'exp. de façon indép. On obtient $X_1, \dots, X_n$, $n$ v.a. indép de même loi (i.i.d.)

Soient $Y = X_1 + \dots + X_n$, $\overline{X} = \dfrac{X_1 + \dots + X_n}{n}$

Propriétés $\mathbb{E}(Y) = n\mu$ $\quad \mathrm{Var}(Y) = n\sigma^2$ (par l'indép.)

$$\mathbb{E}(\overline{X}) = \mu \qquad \mathrm{Var}(\overline{X}) = \frac{1}{n^2}\left(\mathrm{Var}(Y)\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Rem. l'indép. de $X_1, \dots, X_n$ est importante.

Si $X_1 = X_2 = \dots = X_n = X$ (non indép.) alors $\mathrm{Var}(X_1 + \dots + X_n) = \mathrm{Var}(nX_1) = n^2\,\mathrm{Var}(X)$.

Maintenant on a $\overline{X} = \dfrac{X_1 + \dots + X_n}{n}$, avec $\mathbb{E}(\overline{X}) = \mu$, $\mathrm{Var}(\overline{X}) = \dfrac{\sigma^2}{n}$.

je la transforme en $Z = \dfrac{\overline{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ qui est centrée réduite, c-à-d $\mathbb{E}(Z) = 0$, $\mathrm{Var}(Z) = 1$.

Th. central limite $X_1, \dots, X_n$ i.i.d., t.q. $\mathbb{E}(X_i) = \mu$ et $\mathrm{Var}(X_i) = \sigma^2$.

Alors $$\mathbb{P}\left(a < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq b\right) \xrightarrow[n \to \infty]{} \int_a^b \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}t^2\right) dt$$

Cela signifie que pour $n$ grand, $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ suit à peu près $\mathcal{N}(0,1)$.

Équivalent : $Y = X_1 + \dots + X_n$ suit à peu près $\mathcal{N}(n\mu, n\sigma^2)$ ;

$\overline{X} = \dfrac{X_1 + \dots + X_n}{n}$ suit à peu près $\mathcal{N}\left(\mu, \dfrac{\sigma^2}{n}\right)$.

Application : approximation d'une loi binomiale

$X_1, \dots, X_n$ i.i.d. de loi $\mathcal{B}(p)$, donc $\mathbb{E}(X_i) = p$, $\mathrm{Var}(X_i) = p(1-p)$.

On fait la somme $Y = X_1 + \dots + X_n$ de loi $\mathcal{B}(n, p)$ $\quad \mathbb{E}(Y) = np$ $\quad \mathrm{Var}(\cancel{X_i}\, Y) = np(1-p)$

et la moyenne $\dfrac{Y}{n} = \dfrac{X_1 + \dots + X_n}{n}$ qui a $\mathbb{E}\left(\dfrac{Y}{n}\right) = p$, $\mathrm{Var}\left(\dfrac{Y}{n}\right) = \dfrac{p(1-p)}{n}$

TCL : $\dfrac{Y/n - p}{\sqrt{\frac{p(1-p)}{n}}}$ suit à peu près $\mathcal{N}(0,1)$ pour $n$ grand.

Équivalent : $Y = X_1 + \dots + X_n$ qui est $\mathcal{B}(n,p)$ suit à peu près $\mathcal{N}(np, np(1-p))$

Usage : aproximer $\mathcal{B}(n,p)$ qui est difficile à calculer, par $\mathcal{N}(np, np(1-p))$, qui se calcule avec les tables.

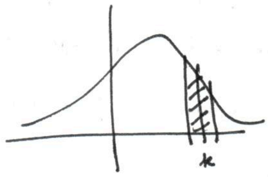
Rem. En pratique, $n$ grand signifie $n \geq 30$ et $p$ ne doit être ni trop proche de $0$ ni trop proche de $1$, en particulier si $np$ et $n(1-p) \geq 15$ ou $20$ c'est bon.

Rem. Correction de continuité

$X \sim \mathcal{B}(n,p)$ et on l'approche par $Z \sim \mathcal{N}(np, np(1-p))$

1) On veut calculer $\mathbb{P}(X=k)$ : ~~comme~~ puisque $Z$ continue, on a $\mathbb{P}(Z=k)=0$ $\forall k$,

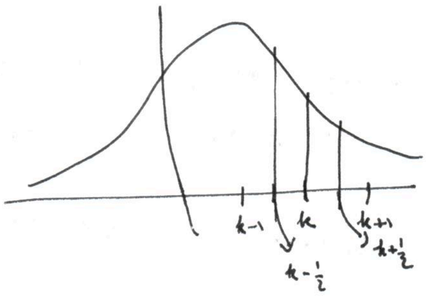
correction de continuité : $\mathbb{P}(X=k) \simeq \mathbb{P}\left(k-\frac{1}{2} < Z \leq k+\frac{1}{2}\right)$




2) On veut calculer $\mathbb{P}(X<k)$ : puisque $X$ discrète,

on a $\mathbb{P}(X<k) = \mathbb{P}(X \leq k-1)$. Mais si on approche $X$ par $Z$ qui est continue,

$\mathbb{P}(Z<k) \neq \mathbb{P}(Z \leq k-1)$.




Correction de continuité : $\mathbb{P}(X<k) \simeq \mathbb{P}\left(Z < k-\frac{1}{2}\right)$

$\mathbb{P}(X \leq k) \simeq \mathbb{P}\left(Z < k+\frac{1}{2}\right)$

$\mathbb{P}(X>k) \simeq \mathbb{P}\left(Z > k+\frac{1}{2}\right)$ , $\mathbb{P}(X \geq k) \simeq \mathbb{P}\left(Z > k-\frac{1}{2}\right)$.

et donc $\mathbb{P}(a < X \leq b) \simeq \mathbb{P}\left(a+\frac{1}{2} < Z < b+\frac{1}{2}\right) = F_Z\left(b+\frac{1}{2}\right) - F_Z\left(a+\frac{1}{2}\right)$.

Exemple $X$ binomiale $X \sim \mathcal{B}(50, 0{,}5)$

on approche $X$ par $Z \sim \mathcal{N}\left(25, \frac{25}{2}\right)$

On veut calculer $\mathbb{P}(24 \leq X \leq 26)$. On a :

calcul exact : $\mathbb{P}(X=24) + \mathbb{P}(X=25) + \mathbb{P}(X=26) \simeq 0.3282$

avec $Z \sim \mathcal{N}(50, 0.5)$ : $\mathbb{P}(24 \leq Z \leq 26) \simeq 0.2222$

avec $Z \sim \mathcal{N}(50, 0.5)$ corrigée par continuité : $\mathbb{P}(24-0.5 < Z < 26+0.5) \simeq 0.3286$

donc le résultat est bien meilleur avec la correction de continuité.