Probabilités – 3 IF

TD 7 : Théorèmes limites

Rappel de cours

On s'intéresse à une suite de variables aléatoires, entières ou réelles, $(X_i)_{i\in\mathbb{N}}$, représentant typiquement les résultats d'une expérience aléatoire répétée (ex. X_i représente la taille d'un individu pris au hasard dans une population). On peut alors s'intéresser à la moyenne des résultats :

$$\bar{X_n} = \frac{1}{n}\sum_{i\leq n} X_i$$

Chaque résultat étant aléatoire, la moyenne elle-même est une variable aléatoire. Néanmoins, lorsque le nombre d'expérience n devient grand, on peut sous certaines condition prévoir sa convergence.

<u>Loi faible des grands nombres</u>

Soient $(X_i)_{i\in\mathbb{N}}$ une suite de v.a. de même loi sur Ω, non corrélées.

non-corrélation : $\mathrm{Cov}(X,Y) = 0$

$\Uparrow$

indépendance : $\mathbb{P}[(X_i)_i = (x_i)_i] = \prod_i \mathbb{P}[X_i = x_i]$

$$\bar{X_n} \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}[X_i] \text{ i.e. } \forall\epsilon>0, \mathbb{P}[|\bar{X_n} - \mathbb{E}[X_i]| \geq \epsilon] \underset{n\to\infty}{\longrightarrow} 0$$

<u>Loi forte des grands nombres</u>

Soient $(X_i)_{i\in\mathbb{N}}$ une suite de v.a. de même loi sur Ω, indépendantes.

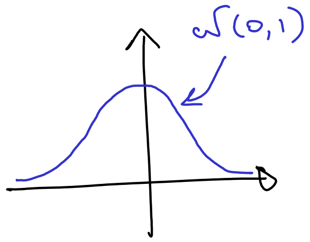
$$\bar{X_n} \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[X_i] \text{ i.e. } \mathbb{P}\left[\left\{(w_i)_i \in \Omega^{\mathbb{N}} \mid \bar{X_n}((w_i)_i) \underset{n\to\infty}{\longrightarrow} \mathbb{E}[X_i]\right\}\right] = 1$$

<u>Théorème central limite</u>

Soient $(X_i)_{i\in\mathbb{N}}$ une suite de v.a. de même loi sur Ω, indépendantes, d'espérance m et de variance σ^2.

$X = \sum_i X_i$

$$\frac{\bar{X_n} - m}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1) \text{ i.e. } \forall\alpha, \mathbb{P}\left[\frac{\bar{X_n} - m}{\frac{\sigma}{\sqrt{n}}} \leq \alpha\right] \underset{n\to\infty}{\longrightarrow} \mathrm{F}_{\mathcal{N}(0,1)}(\alpha)$$



Rmq : En pratique, pour des X_i suivant une loi de Bernoulli de paramètre p (succès avec probabilité p, échec avec probabilité $1-p$), on considère qu'on peut approximer $\bar{X_n}$ par une loi normale si $n > 30$, $np > 5$ et $n(1-p) > 5$.

Exercice 1 :

Lors du deuxième tour de l'élection présidentielle, un sondage "sortie des urnes" est effectué sur un échantillon de n personnes. On fait l'hypothèse que les réponses des sondés sont indépendantes, et que les sondés ne mentent pas. De plus on ne tiendra compte que des suffrages exprimés.

1. L'institut de sondage annonce la victoire de celui des deux candidats A et B qui a pour lui le plus grand nombre de voix des personnes sondées. Quelles sont les probabilités de se tromper dans le cas où A a 51% des suffrages et où n prend les valeurs n = 100, n = 500, n = 1000 ?

On note X le nombre de sondés ayant voté pour A. X suit une loi binomiale $\mathcal{B}(n, 0.51)$

Le sondage se trompe s'il annonce la victoire de B, c'est-à-dire si $X < \frac{n}{2}$.

$$\mathbb{P}\left[X < \frac{n}{2}\right] = \mathbb{P}\left[\frac{X - 0.51\,n}{\sqrt{np(1-p)}} < \frac{\frac{n}{2} - 0.51\,n}{\sqrt{np(1-p)}}\right] \xrightarrow[n\to\infty]{} F_{N(0,1)}\left(\frac{\frac{n}{2} - 0.51\,n}{\sqrt{np(1-p)}}\right)$$

$\simeq$

Or $n > 30$, $np = 100 \times 0.51 > 5$
$n(1-p) = 100 \times 0.49 > 5$

Donc on peut approximer $\mathbb{P}\left[X < \frac{n}{2}\right]$ par $F_{N(0,1)}\left(\frac{\frac{n}{2} - np}{\sqrt{np(1-p)}}\right)$

n	$\frac{n\left(\frac{1}{2} - p\right)}{\sqrt{np(1-p)}} = -0.02\sqrt{n}$	$\mathbb{P}\left[X < \frac{n}{2}\right]$
100	-0.2	0.42
500	-0.45	0.33
1000	-0.63	0.26

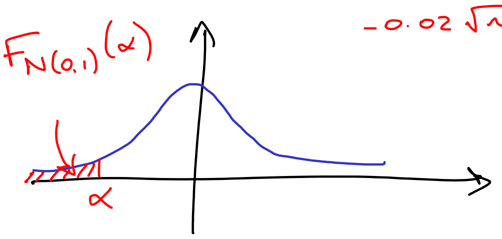
Quelle est la valeur minimale de n pour que la probabilité de se tromper soit inférieure 5% ?

On cherche n tel que :

$$F_{N(0,1)}\Bigg(\underbrace{\frac{n\left(\frac{1}{2} - p\right)}{\sqrt{np(1-p)}}}_{-0.02\sqrt{n}}\Bigg) < 0.05 \iff -0.02\sqrt{n} < -1.65$$

$$\iff n > \left(\frac{1.65}{0.02}\right)^2$$

$$\iff \boxed{n > 6803}.$$



2. Sur les 40 premiers électeurs interrogés, 26 ont voté pour A. Est-ce la peine de poursuivre le sondage ?

On fait un test statistique : On suppose que B a gagné l'élection, et on calcule sous cette hypothèse la probabilité de notre observation (26 sondés sur 40 ont voté pour A). Si cette probabilité est suffisamment faible, on peut en conclure que l'hypothèse est hautement improbable.

On fixe un seuil d'erreur à 5%.

On suppose que B a gagné l'élection, c'est à dire que $p \leq 0.5$.

Prenons le cas le plus défavorable, c'est-à-dire $p = 0.5$.

Alors la probabilité que 26 sondés sur 40 aient voté pour A vaut :

$$\mathbb{P}[X \geq 26] = 1 - \mathbb{P}[X < 26]$$

$$= 1 - \mathbb{P}\Bigg[\underbrace{\frac{X - np}{\sqrt{np(1-p)}}}_{\sim \mathcal{N}(0,1)} < \frac{26 - np}{\sqrt{np(1-p)}}\Bigg] = 1 - F_{N(0,1)}\Bigg(\underbrace{\frac{26 - 40 \times 0.5}{\sqrt{40 \times 0.5^2}}}_{1.9}\Bigg) = 0.03$$

La probabilité que 26 sondés sur 40 aient voté pour A est inférieure au seuil de 5%. Donc on peut rejeter l'hypothèse que B ait gagné l'élection, et arrêter le sondage.

Exercice 2 : (Méthode de Monte Carlo de calcul d'une intégrale)

Soit h une fonction continue sur [0, 1] et soit (X n , n ∈ N) une suite de variables aléatoires indépendantes de même loi uniforme sur [0, 1].

~~Montrez que~~
~~1~~
~~[h(X 1) + · · · + h(X n)]~~
~~n~~
~~P~~
~~Z~~
~~−→~~
~~n→+∞~~
~~1~~
~~h(x) dx,~~
~~0~~

$$\frac{1}{n}\left[h(X_1)+\dots+h(X_n)\right] \xrightarrow[n\to\infty]{P} \int_0^1 h(x).dx.$$

On voudrait appliquer la loi forte des grands nombre. les $(X_i)_i$ sont indépendantes ; donc les $(\underbrace{h(X_i)}_{Y_i})_i$ aussi :

$$\frac{1}{n}\left(h(X_1)+\dots+h(X_n)\right) \xrightarrow[n\to\infty]{P} E\left(h(X_i)\right)$$

$$\boxed{E\left[h(X_i)\right] = \int_{-\infty}^{+\infty} h(x) f_X(x).dx = \int_{-\infty}^{+\infty} h(x)\, 1_{0\le x\le 1}.dx = \int_0^1 h(x).dx}$$

A quoi peut servir cette propriété ?

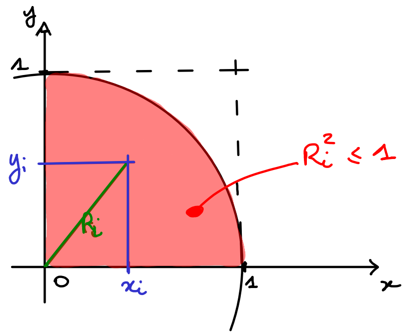
à calculer l'intégrale numériquement en calculant $h(x_i)$ pour des valeurs de x_i tirées aléatoirement uniformément et indépendamment puis en calculant $\frac{1}{n}\sum_{i=1}^{n} h(x_i)$.

Exercice 3 (DS 2013)

On s'intéresse au calcul du nombre π = 3,14 . . . par une méthode dite de Monte-Carlo, c'est-à-dire à l'aide de simulations de lois de probabilité.

Soient X 1 , X 2 , . . . , X n et Y 1 , Y 2 , . . . , Y n des variables indépendantes de loi U[0; 1]. Notons $R_i^2 = X_i^2 + Y_i^2$. Alors on peut montrer que la variable aléatoire $Z_i = 1_{R^2\le 1}$ suit une i loi de Bernoulli B(π/4).

1. Justifiez graphiquement que $P(R_i^2 \le 1) = \pi/4$.



$$P\left[R_i^2 \le 1\right] = \frac{1}{4}\,\mathcal{A}_{C(1)} = \frac{\pi}{4}$$

$P_n = \frac{4}{n}\sum_{i=1}^{n} Z_i$

2. Soit . Montrer que P_n converge presque-sûrement vers π/3. Soit α > 0. A l'aide de l'inégalité de Tchebychev, déterminer n_α tel que pour tout n supérieur à n_α , P(|P n − π| > α) ≤ 0,05.

2) $Z_i \sim \mathcal{B}(\pi/4)$

Donc, comme les Z_i sont indépendants (car les X_i et Y_i le sont) on peut appliquer la loi forte des grands nombres :

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow[n\to\infty]{p.s.} \mathbb{E}[Z_i] = \frac{\pi}{4} \quad \text{d'où} \quad P_n = \frac{4}{n}\sum_{i=1}^{n} Z_i \xrightarrow[n\to\infty]{p.s.} \pi$$

3) D'après l'inégalité de Tchebychev :

$$\mathbb{P}\left[|P_n - \mathbb{E}[P_n]| \geqslant \alpha\right] \leqslant \frac{\sigma^2(P_n)}{\alpha^2}$$

or $\mathbb{E}[P_n] = \frac{4}{n}\sum_{i=1}^{n}\mathbb{E}(Z_i) = 4\,\mathbb{E}[Z_i] = \pi$

$$\sigma^2(P_n) = \frac{16}{n^2}\sum_{i=1}^{n} Var(Z_i) \quad (\text{par indépendance})$$
$$= \frac{16}{n}\,\frac{\pi}{4}\times\left(1-\frac{\pi}{4}\right)$$
$$= \frac{\pi(4-\pi)}{n}$$

et $\mathbb{P}\left[|P_n - \pi| \geqslant \alpha\right] \leqslant \frac{\pi(4-\pi)}{n\,\alpha^2}$

on veut n_α tel que $\frac{\pi(4-\pi)}{n_\alpha \alpha^2} = 0.05$ donc $n_\alpha = \frac{0.05\,\alpha^2}{\pi(4-\pi)}$

3. Reprendre la question précédente avec le théorème de la limite centrale. Commentez.

D'après le théorème central limite :

$$\frac{P_n - \pi}{\sqrt{\frac{\pi(4-\pi)}{n}}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$$

Donc si on approxime la loi de $\frac{P_n - \pi}{\sqrt{\frac{\pi(4-\pi)}{n}}}$ par la loi normale centrée réduite :

$$\mathbb{P}\left[|P_{n'_\alpha} - \pi| \geqslant \alpha\right] = 2\times\mathbb{P}\left[\frac{P_{n'_\alpha} - \pi}{\sqrt{\frac{\pi(4-\pi)}{n'_\alpha}}} > \frac{\alpha}{\sqrt{\frac{\pi(4-\pi)}{n'_\alpha}}}\right] = 2\left(1 - F_{\mathcal{N}(0,1)}\left(\alpha\sqrt{\frac{n'_\alpha}{\pi(4-\pi)}}\right)\right)$$

et $\mathbb{P}\left[|P_{n'_\alpha} - \pi| \geqslant \alpha\right] \leqslant 0.05 \iff F_{\mathcal{N}(0,1)}\left(\alpha\sqrt{\frac{n'_\alpha}{\pi(4-\pi)}}\right) \geqslant 0.975$

$$\iff \alpha\sqrt{\frac{n'_\alpha}{\pi(4-\pi)}} \geqslant 1.96$$

$$\iff \boxed{n'_\alpha \geqslant 1.96^2\,\frac{\pi(4-\pi)}{\alpha^2}}$$

← n'_α est 5 fois plus faible que n_α → inégalité plus puissante

4. Ecrire un algorithme qui retourne une valeur comprise dans l'intervalle $[\pi - 10^{-10}; \pi + 10^{-10}]$ avec une probabilité de 95%. On supposera que l'on dispose d'une fonction Alea() qui retourne une réalisation d'une loi uniforme sur [0; 1].

```
α ← 10^-10
n_α ← 1.96^2 π(4-π)/α^2
P = 0
For i in range(n_α) :
    x ← alea()
    y ← alea()
    If x^2 + y^2 ≤ 1 :
        P ← P+1
return 4P/n
```