Chapitre 13

Loi des Grands Nombres et Théorème Central Limite

La Loi des Grands Nombres

Abordons maintenant le premier théorème fondamental des probabilités.

On a vu précédemment que la proba d'un certain événement pouvait être intuitivement interprété comme la **fréquence** de cet événement sur le long terme, si l'**expérience est répétée un grand nombre de fois**

On a aussi défini mathématiquement une proba. comme la valeur de la fonction de distribution de la variable aléatoire représentant l'expérience

La Loi des Grands Nombres montre que ce modèle mathématique est cohérent avec l'interprétation en termes de fréquence des probabilités.

L'inégalité de Tchebychev

Afin d'étudier la Loi des Grands Nombres, on a tout d'abord besoin d'une inégalité importante appelée "Inégalité de Tchebychev":

Théorème : Inégalité de Tchebychev

Soit X une variable aléatoire d'espérance $\mu = \mathbb{E}(X)$ et de variance $\sigma^2 = V(X)$ finie. Alors pour tout $\epsilon > 0$:

$$\mathbb{P}(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

Intuition : La probabilité de s'éloigner de la moyenne est d'autant plus petite que la variance est petite

Démonstration

Par définition $\mathbb{P}(|X-\mu| \geq \epsilon) = \int_{|x-\mu|\geq\epsilon} f(x)dx$

$$\text{Or } \sigma^2 = \int_x (x-\mu)^2 f(x)dx$$
$$\geq \int_{|x-\mu|\geq\epsilon} (x-\mu)^2 f(x)dx$$

puisque tous les termes sont positifs et qu'on restreint l'espace d'intégration.

Par ailleurs comme on considère uniquement les x tels que $|x-\mu| > \epsilon$, on a dans l'intégrale $(x-\mu)^2 > \epsilon^2$ et donc

$$\sigma^2 \geq \int_{|x-\mu|\geq\epsilon} \epsilon^2 f(x)dx = \epsilon^2 \int_{|x-\mu|\geq\epsilon} f(x)dx = \epsilon^2 \mathbb{P}(|X-\mu| \geq \epsilon)$$

Ainsi, $\mathbb{P}(|X-\mu| \geq \epsilon) \leq \dfrac{\sigma^2}{\epsilon^2}$

Exemple

Soit X une variable aléatoire avec $\mathbb{E}(X) = \mu$ et $V(X) = \sigma^2$.

Avec $\epsilon = k\sigma$, l'inégalité de Tchebychev indique que

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

Ainsi, pour tout variable aléatoire, la probabilité d'une déviation de la moyenne de plus de k écart-type est inférieurs à $1/k^2$.

Par exemple si $k = 5$, $1/k^2 = 0.04$

La Loi des Grands Nombres

Théorème

Soit $X_1, X_2, ..., X_n$ une suite de v.a. indépendantes, de même loi d'espérance $\mu = \mathbb{E}(X_i)$ et de variance $\sigma^2 = V(X_i)$. Notons $S_n = X_1 + ... + X_n$. Alors pour tout $\epsilon > 0$:

$$\lim_{n \to +\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Ou de manière équivalente

$$\lim_{n \to +\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) = 1$$

Intuition : La moyenne arithmétique $\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$ converge (en probabilité) vers l'espérance statistique $\mu = \mathbb{E}(X_i)$

Démonstration

$X_1, X_2, ..., X_n$ étant indépendantes et ayant la même distrib,
$\mathbb{E}(S_n) = \mathbb{E}(X_1) + ... + \mathbb{E}(X_n) = n\mu$ et
$V(S_n) = V(X_1) + ... + V(X_n) = n\sigma^2$.
Ainsi $\mathbb{E}\left(\frac{S_n}{n}\right) = \mu$ et $V\left(\frac{S_n}{n}\right) = \frac{\sigma^2}{n}$.
D'après l'inégalité de Tchebychev, on a donc $\forall \epsilon$:

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon}$$

Et donc

$$\lim_{n\to+\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Remarques

Notons que $\frac{S_n}{n}$ est une moyenne d'"issues individuelles". Ainsi on appelle souvent les Loi des Grands Nombres, "Loi de la moyenne".

Il est frappant de voir qu'on peut commencer avec une expérience dont on ne connait pas grand chose et, en prenant la moyenne, obtenir une expérience dont on peut prédire le résultat avec un fort niveau de certitude.

La Loi des Grands Nombres que nous venons d'étudier est souvent appelé "Loi Faible des Grands Nombres" pour la différencier de la "Loi Forte des Grands Nombres"

Loi Forte des Grands Nombres

La loi que nous venons de démontrer ne correspond pas exactement au fait que la moyenne arithmétique converge vers l'espérance, i.e. à l'intuition qu'en lançant une pièce un grand nombre de fois la proportion de "face" atteigne véritablement $1/2$.

Pour cela, il faut utiliser la "Loi Forte des Grands Nombres" qui indique que

$$\mathbb{P}\left(\frac{S_n}{n} \to \mu\right) = 1$$

on parlera alors de convergence presque sure

Exemple 1 : Expériences de Bernouilli

Considérons le cas important de n expériences de Bernoulli avec une probabilité p de succès. $X_i = 1$ si la i$^{\text{ème}}$ issue est un succès, 0 sinon. Alors $S_n = X_1 + ... + X_n$ est le nombre de succès en n épreuves et $\mu = \mathbb{E}(X_1) = p$. La Loi (Faible) des Grands Nombres indique que

$$\lim_{n \to +\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) = 1$$

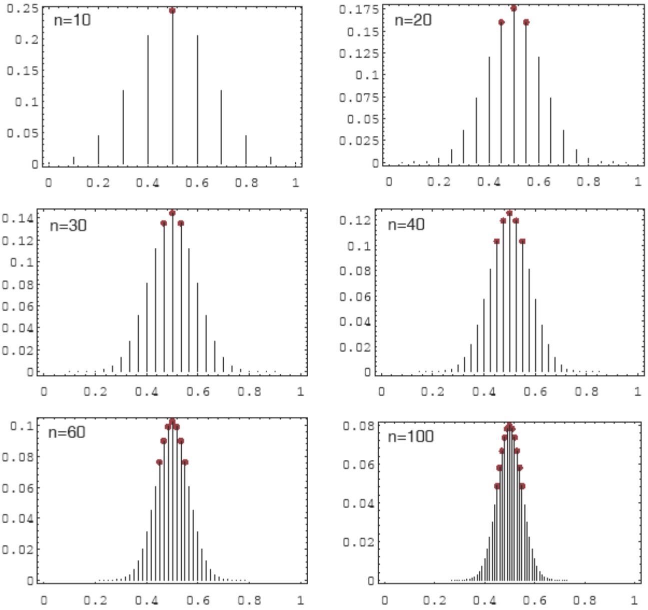
Ainsi, si l'expérience est répétée un grand nombre de fois, on peut s'attendre à ce que la proportion de succès soit proche de p.

Cela montre que le modèle mathématique de probabilité est cohérant avec l'interprétation en termes de fréquence des probabilités

Considérons le cas particulier de n "Pile ou Face" avec S_n le nombre de "Face". La variable aléatoire $\frac{S_n}{n}$ représente alors la proportion de "Face" et est comprise entre 0 et 1. La Loi des Grands Nombres prédit que le résultat de cette variable aléatoire devrait, pour n grand, être proche de $1/2$

Sur le graphique suivant, on a représenté la distribution de cette variable en augmentant la valeur de n et on a marqué d'un point rouge les valeurs entre $0,45$ et $0,55$.

On voit alors qu'en augmentant n la distribution devient plus concentrée autour de $0,5$ et qu'un pourcentage de plus en plus élevé de l'aire est contenu dans l'intervalle $[0,45; 0,55]$, comme prédit par la Loi des Grands Nombres


n=10
n=20
n=30
n=40
n=60
n=100
0
0.2
0.4
0.6
0.8
1

Exemple 2 : La distribution Uniforme

Imaginons qu'on choisisse n nombres dans l'intervalle $[0,1]$ avec une distribution uniforme. Si X_i représente le i$^{\text{ème}}$ choix :

$$\begin{aligned}
\mu &= \mathbb{E}(X_i) = \int_0^1 x dx = \frac{1}{2} \\
\sigma^2 &= V(X) = \int_0^1 x^2 dx - \mu = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}
\end{aligned}$$

Ainsi, $\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{2}$, $V\left(\frac{S_n}{n}\right) = \frac{1}{12n}$, et pour tout $\epsilon > 0$:

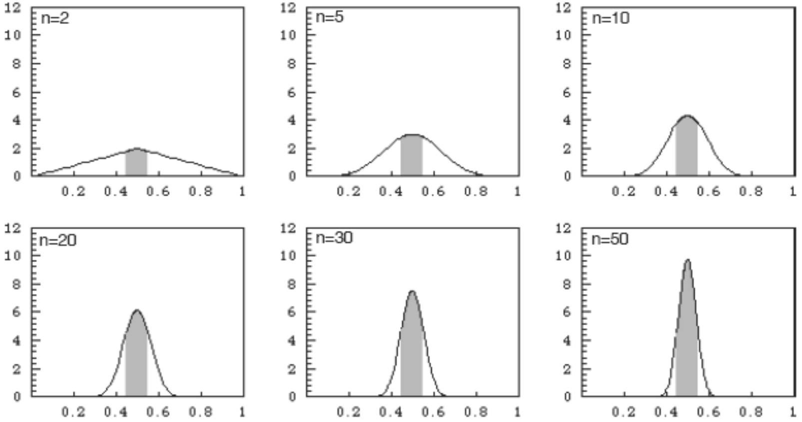
$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{12n\epsilon^2}$$

Ainsi, si on choisit n nombres au hasard entre 0 et 1, il y a plus de $1 - \frac{1}{12n\epsilon^2}$ chance que la différence $\left|\frac{S_n}{n} - \frac{1}{2}\right|$ soit inférieur à ϵ.

Remarquez que ϵ joue le rôle de l'ampleur des erreurs que nous sommes prêts à tolérer.

Si on choisit $\epsilon = 0,1$, les chances que $\left|\frac{S_n}{n} - \frac{1}{2}\right|$ soit inférieur à $0,1$ sont supérieur à $1 - 100/(12n)$. Ainsi pour $n = 100$, elles sont supérieures à $0,92$; supérieures à $0,99$ si $n = 1000$ et à $0,999$ avec $n = 10.000$

On peut représenter graphiquement ce que la Loi des Grands Nombres implique dans cet exemple. Sur la figure suivante, on a représenté S_n/n pour plusieurs valeurs de n et assombrit l'aire pour laquelle S_n/n est compris entre $0,45$ et $0,55$



Ainsi, quand on augmente n, une partie de plus importante de l'aire totale est sombre. La LGN indique qu'on peut obtenir une proportion assombie aussi importante que l'on souhaite en choisissant un n assez grand.

Théorème Central Limite

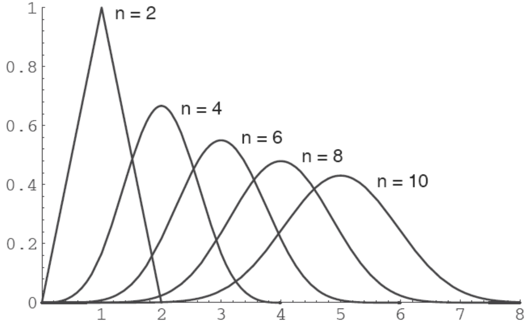
Le deuxième théorème fondamental des probabilités est le Théorème Central Limite.

Ce théorème indique que si S_n est la somme de n v.a. mutuellement indépendantes et identiquement distribuées, alors la distribution de S_n peut être correctement approximé par la densité normale (et ce quelque soit la distribution initiale!).

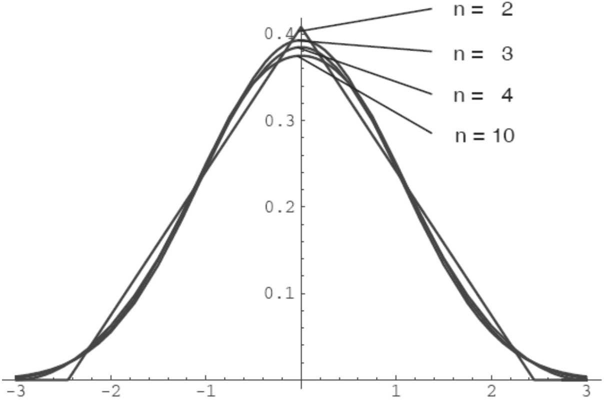
Avant de définir formellement le Théorème Central Limite (TCL) étudions quelques exemples.

Exemple 1 : La distribution uniforme

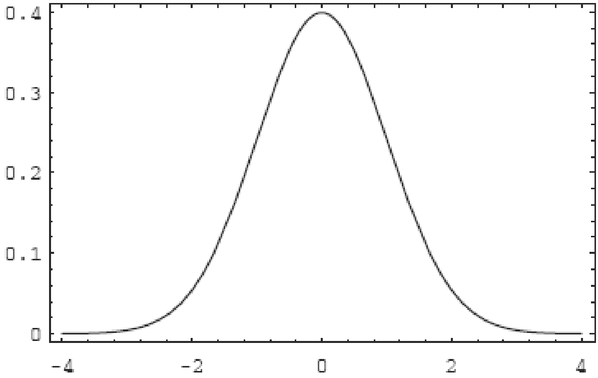
Soient n nombres choisit aléatoirement dans l'intervalle $[0, 1]$ avec une densité uniforme. Notons ces choix $X_1, X_2, ..., X_n$ et leur somme $S_n = X_1 + X_2 + ... + X_n$. En graphant la densité de S_n en fonction de n, on remarque que cette somme à une forme de loi normale mais est centrée en $n/2$ et s'applatit quand n augmente.



Afin de comparer la forme de ces densités pour différentes valeurs de n, il est utile de centrer et réduire S_n en définissant $S_n^* = \dfrac{S_n - n\mu}{\sqrt{n}\sigma}$. Ainsi pour tout n, on a $\mathbb{E}(S_n) = 0$ et $V(S_n) = 1$. La fonction de densité de S_n^* est alors une version centrée-réduite de la densité de S_n



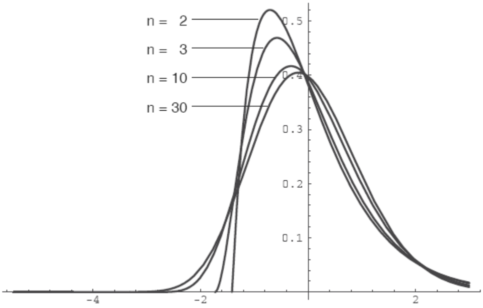
On remarque que, quand n augmente, S_n^* tend vers une densité normale centrée-réduite qui pour rappel se représente comme suit



Exemple 2 : La distribution exponentielle

Reproduisons la même "expérience" en choisissant des nombres dans l'intervalle $[0, +\infty[$ selon une loi exponentielle de paramètre λ. On a vu en TD qu'alors $\mu = \mathbb{E}(X_i) = \frac{1}{\lambda}$ et $\sigma^2 = V(X_i) = \frac{1}{\lambda^2}$.

La(les) densité(s) de S_n^* se représente(nt) alors comme suit



Ces exemples illustrent le fait que la densité de la variable centrée réduite S_n^* resemble beaucoup à celle d'une loi normale $\mathcal{N}(0,1)$ pour de grandes valeurs de n et ce quelque soit la distribution initiale (elle peut même être discrête).

Le Théorème Central Limite précise cette observation.

Théorème Central Limite

Soit $S_n = X_1 + ... + X_n$ la somme de n variables aléatoires indépendantes ayant la même densité d'espérance μ et de variance σ^2. Soit $S_n^* = \dfrac{S_n - \mu n}{\sqrt{n}\sigma}$. Alors, pour tout $a < b$:

$$\lim_{n\to+\infty} \mathbb{P}\left(a < S_n^* < b\right) = \frac{1}{\sqrt{2\Pi}} \int_a^b e^{\frac{-x^2}{2}} dx$$

En d'autre termes, on a $\boxed{\lim_{n \to +\infty} F_{S_n^*}(t) = \Phi_{0,1}(t)}$

(on parle alors de "convergence en loi")

Ainsi, le TCL indique que si S_n est la somme de n variables aléatoires étant

- mutuellement independentes
- identiquement distribuées

et si on "centre-réduit" S_n, la v.a. résultante : S_n^* a des caractéristiques très spéciales. En particulier, la fonction de distrib de S_n^* approxime la distribution normale centrée-réduite **quelque soit la distribution des v.a. originales**. De plus l'approximation est d'autant plus précise que n est grand.

Ainsi le processus de sommation et de standardisation (centrer-réduire) "lave" les caractéristiques originelles des v.a. individuelles pour les remplacer par celles d'une loi normale centrée réduite.