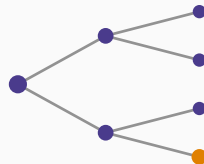


Probabilités

Chapitre 21 : Le théorème central limite



Autour de la limite

Le théorème

Pourquoi la cloche est partout

Les usages

Ce qu'il faut retenir

Autour de la limite

Converger ne suffit pas

La loi des grands nombres dit que $\overline{X}_n$ s'approche de m . Elle ne dit pas **comment** les valeurs se répartissent autour de m .

Or c'est précisément ce qu'il faut pour construire un intervalle, tester une hypothèse, calculer un risque. Sans la forme de la distribution, la convergence est inexploitable.

La réponse, et sa surprise

La distribution de $\overline{X}_n$ tend vers une **loi normale**, quelle que soit la loi de départ.

Cette dernière précision est le cœur du théorème : la loi des X_i peut être n'importe laquelle — discrète, dissymétrique, bornée, sans aucune ressemblance avec une cloche.

Le théorème

Théorème central limite

Soit $X_1, \dots, X_n$ indépendantes, de même loi, d'espérance m et d'écart-type σ finis. Alors, quand n grandit :

$$\overline{X}_n \approx \mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right)$$

soit, sous forme standardisée :

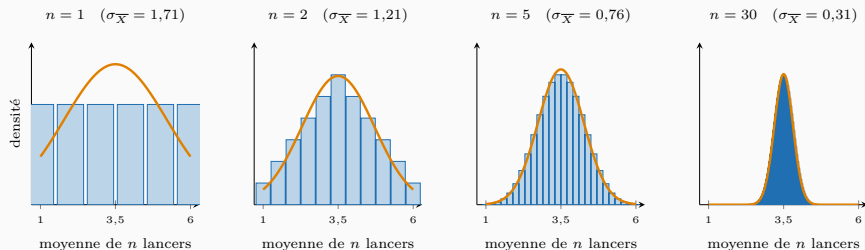
$$Z_n = \frac{\overline{X}_n - m}{\sigma/\sqrt{n}} \longrightarrow \mathcal{N}(0; 1)$$

Aucune hypothèse sur la loi de départ

Le théorème ne demande que trois choses : indépendance, loi commune, variance finie.

Il ne demande rien sur la **forme** de cette loi. C'est ce qui en fait le théorème le plus utilisé de toute la statistique.

Un dé n'a rien d'une cloche



Un dé n'a rien d'une cloche — ses six faces sont équiprobables. Pourtant, dès qu'on en moyenne quelques-uns, la cloche apparaît, et à trente lancers elle est parfaite.

$$\bar{X}_n \approx \mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right) \quad \text{quelle que soit la loi de départ.}$$

Deux choses se produisent en même temps : la forme devient normale, et l'étalement se resserre en $1/\sqrt{n}$. Quadrupler l'échantillon ne divise l'erreur que par deux.

Deux phénomènes simultanés

1. La **forme** devient normale : c'est le théorème central limite.
2. L'**étalement** se resserre en $\sigma/\sqrt{n}$: c'est la loi des grands nombres.

Les deux théorèmes se lisent sur la même série de vignettes.

La vitesse de convergence

Une loi de départ symétrique — comme le dé — converge très vite : dès $n = 5$ l'ajustement est bon.

Une loi très dissymétrique demande davantage. La règle usuelle est $n \geq 30$, et elle est prudente pour la plupart des cas rencontrés en gestion.

Pourquoi la cloche est partout

Le mécanisme

Une grandeur qui résulte de **l'addition de nombreuses causes indépendantes**, aucune n'étant prépondérante, suit approximativement une loi normale.

Ce n'est pas une observation empirique : c'est une conséquence du théorème.

Quatre exemples

- Une **taille** additionne des dizaines de facteurs génétiques et nutritionnels.
- Une **erreur de mesure** additionne de multiples imprécisions indépendantes.
- Un **rendement mensuel** additionne des variations quotidiennes.
- Le **nombre de défauts** d'un portefeuille additionne des indicatrices : c'est une binomiale, donc une somme.

Et pourquoi elle n'est pas toujours là

Trois conditions, trois façons d'échouer

- Les causes se **multiplient** au lieu de s'additionner : on obtient une loi log-normale, pas une normale.
- Une cause **domine** toutes les autres : la somme lui ressemble, non à une cloche.
- Les causes sont **dépendantes** : en cas de krach, tout baisse ensemble, et la somme a des queues bien plus épaisses.

Le point aveugle de la finance

Les rendements boursiers ne sont pas normaux : les krachs sont bien trop fréquents pour une loi de Gauss.

Un mouvement à -5 écarts-types devrait survenir une fois tous les sept mille ans. Il en est survenu plusieurs en un siècle. **La cause : les dépendances, précisément ce que le théorème exclut.**

Les usages

La construction

Puisque $\bar{X}_n \approx \mathcal{N}(m; \sigma/\sqrt{n})$ et que $P(-1,96 \leq Z \leq 1,96) = 0,95$:

$$P\left(\bar{X}_n - 1,96 \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X}_n + 1,96 \frac{\sigma}{\sqrt{n}}\right) = 0,95$$

Un sondage

Sur 1 000 personnes, 52 % déclarent une intention favorable. Pour une proportion, $\sigma = \sqrt{p(1-p)} \approx 0,5$.

$$0,52 \pm 1,96 \times \frac{0,5}{\sqrt{1\,000}} = 0,52 \pm 0,031$$

L'intervalle $[0,489; 0,551]$ contient 0,50 : le résultat ne permet pas de conclure à une majorité.

Le fameux « plus ou moins trois points »

$$1,96 \times \frac{0,5}{\sqrt{1\,000}} = 0,031$$

Il n'y a aucun mystère : trois points, c'est **1,96** fois un demi divisé par la racine de mille. Le chiffre annoncé à la télévision se recalcule en une ligne.

Ce que la marge ne couvre pas

Elle ne mesure **que** l'aléa d'échantillonnage. Elle ignore les refus de répondre, les questions mal posées, les déclarations inexactes.

L'erreur statistique est souvent la plus petite des erreurs d'une enquête — et c'est pourtant la seule que l'on publie.

Le retour du portefeuille

Les 200 crédits, une dernière fois

$X \sim \mathcal{B}(200; 0,04)$ est une somme de 200 indicatrices indépendantes : le théorème s'applique.

$$X \approx \mathcal{N}(8; 2,77) \quad \text{et} \quad q_{0,99} = 8 + 2,326 \times 2,77 = 14,4$$

Valeur exacte au chapitre 15 : 15 défauts. L'approximation est un peu optimiste, mais du bon ordre de grandeur.

Pourquoi les modèles de risque l'utilisent

Un portefeuille réel compte des dizaines de milliers de lignes. Calculer la loi exacte serait long; l'approximation normale donne le quantile en une formule.

$$\text{VaR}_{99\%} = m + 2,326 \sigma$$

Toute la Value at Risk paramétrique tient dans cette ligne — et dans l'hypothèse d'indépendance qui la soutient.

Ce qu'il faut retenir

Sept points

1. **Théorème central limite** : $\bar{X}_n \approx \mathcal{N}(m; \sigma/\sqrt{n})$, quelle que soit la loi de départ.
2. Trois hypothèses seulement : indépendance, loi commune, variance finie.
3. Deux effets simultanés : la forme devient normale, l'étalement se resserre en $1/\sqrt{n}$.
4. En pratique, $n \geq 30$ suffit; moins si la loi de départ est symétrique.
5. Il **explique** l'omniprésence de la loi normale : une somme de causes indépendantes.
6. Il **échoue** si les causes se multiplient, si l'une domine, ou si elles sont dépendantes.
7. Intervalles de confiance et Value at Risk paramétrique en découlent directement.

La suite

Le cours théorique est achevé : de la fréquence observée à la fréquence expliquée, la boucle est complète.

Le chapitre 22 met tout cela en œuvre sur machine — les lois, les quantiles, les simulations et les deux théorèmes, en quelques lignes de R et de Python.