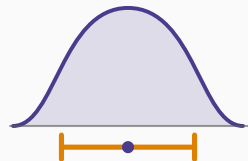


Statistique inférentielle

Chapitre 8 : Construire un estimateur



Jaouad Madkour

Faculté d'Économie et de Gestion de Tanger · 27 août 2026

Deux méthodes

La méthode des moments

Le maximum de vraisemblance

Ce qu'il faut retenir

Deux méthodes

L'évidence s'arrête vite

Pour estimer une moyenne, on prend la moyenne : personne n'hésite.

Mais pour estimer le paramètre λ d'une loi exponentielle, la borne supérieure d'une loi uniforme, ou la probabilité de défaut d'un portefeuille observé partiellement ? L'intuition ne suffit plus.

Les deux méthodes du cours

La **méthode des moments** : on égale les moments théoriques aux moments observés.

Le **maximum de vraisemblance** : on retient la valeur du paramètre qui rendait les données les plus plausibles.

La méthode des moments

Une équation par paramètre

On écrit l'espérance théorique en fonction du paramètre, on l'égalé à la moyenne observée, et on résout.

$$E(X) = g(\theta) \implies \bar{x} = g(\hat{\theta}) \implies \hat{\theta} = g^{-1}(\bar{x})$$

Trois applications immédiates

Loi	Relation	Estimateur
Bernoulli $\mathcal{B}(p)$	$E(X) = p$	$\hat{p} = \bar{x} = f$
Poisson $\mathcal{P}(\lambda)$	$E(X) = \lambda$	$\hat{\lambda} = \bar{x}$
Exponentielle	$E(X) = 1/\lambda$	$\hat{\lambda} = 1/\bar{x}$

Simple, mais rustique

Elle ne demande aucun calcul difficile, et donne souvent le bon résultat du premier coup.

Mais elle n'utilise que les premiers moments — donc pas toute l'information de l'échantillon. Elle peut produire des estimateurs manifestement inefficaces, voire absurdes.

Un cas où elle échoue

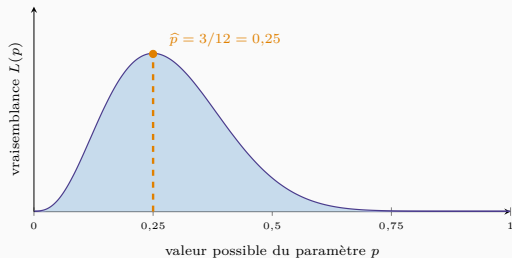
Sur une loi uniforme sur $[0 ; b]$, elle donne $\hat{b} = 2\bar{x}$.

Si l'échantillon contient la valeur 9 mais que $2\bar{x} = 8$, l'estimateur affirme que le maximum possible est 8 — **alors qu'on a observé 9**. Une estimation contredite par les données mêmes.

Le maximum de vraisemblance

Retourner la question des probabilités

sur 12 crédits, 3 ont fait défaut : quelle valeur de p rend cela le plus plausible ?



Le maximum de vraisemblance

On retourne la question des probabilités.

Le cours de probabilités demandait : p étant donné, quelle chance d'observer 3 défauts ?

L'inférence demande : 3 défauts étant observés, quelle valeur de p rendait cela le plus vraisemblable ?

La réponse est la fréquence observée.
Ce n'est pas une coïncidence : pour Bernoulli, le maximum de vraisemblance est la proportion empirique.

En pratique on maximise $\ln L$, qui a le même maximum et transforme les produits en sommes.

Fonction de vraisemblance

Pour un échantillon $x_1, \dots, x_n$ indépendant, la **vraisemblance** est la probabilité d'observer exactement ces données, vue comme une fonction du paramètre :

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$$

L'estimateur du maximum de vraisemblance est la valeur $\hat{\theta}$ qui maximise L .

Le renversement, en une phrase

Le cours de probabilités demandait : *le paramètre étant donné, quelle chance d'observer ces données ?*

L'inférence demande : *ces données étant observées, quelle valeur du paramètre les rendait les plus vraisemblables ?* Même fonction, lue dans l'autre sens.

On maximise le logarithme

$$\ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta)$$

Le logarithme est croissant : il a le même maximum. Et il transforme le produit en **somme**, ce qui rend la dérivation praticable.

Le cas de Bernoulli, entièrement traité

Sur $n = 12$ crédits, $k = 3$ défauts :

$$L(p) = p^3(1-p)^9 \quad \ln L(p) = 3 \ln p + 9 \ln(1-p)$$

$$\frac{d \ln L}{dp} = \frac{3}{p} - \frac{9}{1-p} = 0 \quad \Rightarrow \quad \hat{p} = \frac{3}{12} = 0,25$$

La fréquence observée n'était pas un choix arbitraire

Le maximum de vraisemblance **redonne** la proportion empirique. Ce que le chapitre 1 des probabilités posait comme une définition intuitive se révèle être la solution d'un problème d'optimisation.

De même, pour une population normale, la méthode redonne $\bar{x}$ pour m . Les estimateurs « évidents » sont ceux que la théorie construit.

Ses propriétés, et pourquoi il domine

1. **Convergent** — il s'approche du vrai paramètre.
2. **Asymptotiquement sans biais** — le biais s'efface quand n grandit.
3. **Asymptotiquement efficace** — aucun autre ne fait mieux à grand échantillon.
4. **Invariant** — l'estimateur de $g(\theta)$ est $g(\hat{\theta})$, sans recalcul.

Trois réserves à connaître

- Il peut être **biaisé à distance finie** : sur une population normale, il estime σ^2 en divisant par n , pas par $n - 1$.
- Il exige de **postuler une loi**. Si la loi supposée est fausse, l'estimateur est faux — et rien ne le signale.
- Le maximum n'a pas toujours de forme explicite : il faut alors le chercher numériquement.

Sa portée réelle

C'est pourtant la méthode d'estimation la plus employée au monde : régression logistique, modèles GARCH, économétrie des données de panel, apprentissage statistique.

Tout ce qui n'est pas une régression linéaire ordinaire est, presque toujours, du maximum de vraisemblance.

Ce qu'il faut retenir

Six points

1. Un estimateur se **construit**; il ne se devine pas.
2. **Méthode des moments** : égaliser moments théoriques et moments observés. Simple, parfois grossière.
3. **Maximum de vraisemblance** : retenir la valeur qui rendait les données les plus plausibles.
4. On maximise $\ln L$: même maximum, produits transformés en sommes.
5. Pour Bernoulli il redonne f , pour la normale il redonne $\bar{x}$: les estimateurs usuels sont les siens.
6. Convergent, asymptotiquement sans biais et efficace, invariant — mais il exige de **postuler une loi**.

La suite

Une estimation ponctuelle est un nombre unique, et elle est presque sûrement fausse : la probabilité que $\bar{x}$ vaille exactement m est nulle.

Annoncer 7 005,80 sans rien d'autre est donc une information incomplète. La quatrième partie ajoute ce qui manque : **la précision**.