Notes de Inference Statistique

Yehor Korotenko

January 27, 2026

Contents

CHAPTER 1

Lecture 1

Evaluation

- 0.4 CC +0.6 Examen.
- Répartition : 80% partiel, 20% Interro (prévue le 26/01).

1.1 Modèle Statistique

Definition 1.1.1 *(Modèle Statistique)* — Un modèle statistique est un espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$ où $\mathcal{P}$ est une famille de lois de probabilité $\{P_\theta; \theta \in \Theta\}$.

- Si $\exists p \in \mathbb{N}^*, \Theta \subset \mathbb{R}^p$: modèle paramétrique.
- Sinon : modèle non paramétrique.

Example 1.1.2 *(Familles de lois)* —
- Lois de Poisson : $\mathcal{P} = \{P(\lambda); \lambda > 0\}$.
- Densité régulière : $\mathcal{P} = \{\mathbb{P}; \mathbb{P}$ dont la densité admet une dérivée seconde bornée$\}$.

Definition 1.1.3 *(Observation)* — Une observation est une variable aléatoire (v.a.) dont la loi appartient à $\{P_\theta, \theta \in \Theta\}$. Notre observation aura une structure de n-échantillons $X_1, \ldots, X_n$ i.i.d. (indépendants et identiquement distribués) de loi commune $\in \{P_\theta, \theta \in \Theta\}$.

Remark 1.1.4 — $(X_1, \ldots, X_n)$ est de loi $P_\theta^{\otimes n}$. L'échantillon contient toute l'information sur P_θ, donc sur θ.

Definition 1.1.5 *(Identifiabilité)* — Un modèle est identifiable si et seulement si (ssi) l'application $\theta \mapsto P_\theta$ est injective.

1.2 Estimateurs

Hypothèse : On observe $X_1, \ldots, X_n$ i.i.d. de loi commune $\in \{P_\theta, \theta \in \Theta \subset \mathbb{R}^p\}$ (modèle paramétrique identifiable). Soit θ^* la vraie valeur inconnue telle que $P_{X_i} = P_{\theta^*}$.

Definition 1.2.1 *(Estimateur)* — Un estimateur de θ est une fonction de l'échantillon $(X_1, \ldots, X_n)$ mesurable et indépendante de θ (calculable à partir des données).

Notation : $\hat{\theta} = \hat{\theta}_n = h(X_1, \dots, X_n)$. C'est une variable aléatoire.
Exemples : $\hat{\theta} = \bar{X}$, $\hat{\theta} = X_1 - X_3$, etc.

Questions fondamentales :

1. Comment définir un bon estimateur ?
2. Comment construire un bon estimateur ?

1.3 Risque quadratique

Idée : En moyenne, $\hat{\theta}$ doit être proche de θ. On regarde $\mathbb{E}[\hat{\theta} - \theta]$.

Definition 1.3.1 *(Biais)* — Le biais de $\hat{\theta}$ est défini par :

$$B(\hat{\theta}, \theta) = \mathbb{E}[\hat{\theta}] - \theta$$

On dit que $\hat{\theta}$ est **sans biais** si $B(\hat{\theta}, \theta) = 0$.

Definition 1.3.2 *(Risque quadratique / MSE)* —

$$R(\hat{\theta}, \theta) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

C'est la **Mean Squared Error (MSE)** en anglais.

On dit que $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ ssi $R(\hat{\theta}_1, \theta) \leq R(\hat{\theta}_2, \theta)$.

Exemple : Modèle de Poisson

Soit $X_1, \dots, X_n$ de loi P_θ de Poisson, $\theta > 0$. On cherche un estimateur de $\theta = \mathbb{E}[X_i]$.

Proposons : $\hat{\theta} = \overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$.

Calcul du Biais :

$$\begin{aligned} B(\hat{\theta}, \theta) &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] - \theta \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] - \theta \quad \text{(par linéarité)} \\ &= \frac{1}{n} \cdot n \cdot \mathbb{E}[X_1] - \theta \\ &= \theta - \theta = 0 \end{aligned}$$

Donc $\mathbb{E}[\overline{X}] = \theta$, est l'estimateur sans biais.

Calcul du Risque :

$$\begin{aligned} R(\hat{\theta}, \theta) &= \mathbb{E}[(\overline{X} - \theta)^2] = \mathbb{E}[(\overline{X} - \mathbb{E}[\overline{X}])^2] \\ &= Var(\overline{X}) = Var\left(\frac{1}{n} \sum X_i\right) \\ &= \frac{1}{n^2} \sum Var(X_i) \quad \text{(car i.i.d)} \\ &= \frac{1}{n^2} \cdot n \cdot Var(X_1) = \frac{Var(X_1)}{n} = \frac{\theta}{n} \end{aligned}$$

Proposition 1.3.3 *(Décomposition Biais-Variance du risque)* —

$$R(\hat{\theta}, \theta) = (B(\hat{\theta}, \theta))^2 + Var(\hat{\theta})$$

Proof —

$$\begin{aligned}
R(\hat{\theta}, \theta) &= \mathbb{E}[(\hat{\theta} - \theta)^2] \\
&= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta)^2] \\
&= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \theta)^2] + 2\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])(\mathbb{E}[\hat{\theta}] - \theta)] \\
&= Var(\hat{\theta}) + (B(\hat{\theta}, \theta))^2 + 2(\mathbb{E}[\hat{\theta}] - \theta)\underbrace{\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]]}_{0} \\
&= Var(\hat{\theta}) + B(\hat{\theta}, \theta)^2
\end{aligned}$$

□

1.4 Consistance

Propriété asymptotique. On ne considère que des estimateurs consistants.

Definition 1.4.1 *(Consistance)* — Soit $(X_1, \ldots, X_n)$ i.i.d. de loi P_θ. Soit $\hat{\theta}_n = h(X_1, \ldots, X_n)$. $\hat{\theta}_n$ est un estimateur consistant (ou convergent) de θ ssi :

$$\hat{\theta}_n \xrightarrow[n \to +\infty]{\mathbb{P}} \theta$$

Remark 1.4.2 — $\hat{\theta}_n$ est fortement consistant ssi $\hat{\theta}_n \xrightarrow[n \to +\infty]{p.s.} \theta$.

Exemple : Retour au modèle de Poisson

$\Theta = \mathbb{R}_+^*$, $\hat{\theta}_n = \overline{X}$.

- On peut invoquer la Loi des Grands Nombres (LGN) : $\overline{X} \xrightarrow{\mathbb{P}} \mathbb{E}[X_i] = \theta$.
- Via le risque quadratique :

  $$R(\hat{\theta}_n, \theta) = Var(\overline{X}) = \frac{\theta}{n} \xrightarrow{n \to +\infty} 0$$

  D'après l'inégalité de Bienaymé-Tchebychev :

  $$P(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{\mathbb{E}[(\hat{\theta}_n - \theta)^2]}{\varepsilon^2} = \frac{R(\hat{\theta}_n, \theta)}{\varepsilon^2} \to 0$$

Méthode "Plug-in"

Soit $(X_1, \ldots, X_n)$ i.i.d. Poisson(θ). On veut estimer $\beta = P(X_i = 0) = e^{-\theta}$.

$$\hat{\beta} = e^{-\hat{\theta}} = e^{-\overline{X}}$$

$\hat{\beta}$ est consistant pour estimer β.

Lemma 1.4.3 *(Lemme de l'application continue)* — Si $Z_n \xrightarrow{\mathbb{P}} Z$, alors $h(Z_n) \xrightarrow{\mathbb{P}} h(Z)$ pour toute fonction continue h.

CHAPTER 2

Lecture 2

2.1 Cadre paramétrique

2.1.1 Modèle statistique paramétrique

On dispose d'une observation $(X_1, \ldots, X_n)$, un échantillon de variable aléatoire i.i.d (indépendantes, identiquement distribuées) de loi commune P appartenant à une famille de lois de probabilités paramétrée $\{P_{\theta, \theta \in \Theta \subset \mathbb{R}^p}\}$.

Remark 2.1.1 — Si $\Theta \subset$ espace de dimension infinie $\rightarrow$ modèle non-paramétré.

Estimer P c'est estimer $\theta \in \mathbb{R}^p$

Example 2.1.2 — Bernolli(θ), Exponentielle(θ), $\mathcal{N}(\mu, \sigma^2)$, loi de densité $f_\theta(x) = \theta x^{\theta-1} 1_{x \in [0,1]}$

Notation. $E_{\theta_n}[h(X_1, \ldots, X_n)]$, $Var_\theta[h(X_1, \ldots, X_n)]$

Loi de $(X_1, \ldots, X_n) \rightarrow P_\theta^{\otimes n}$

Definition 2.1.3 *(Estimateur)* —

$$\hat{\theta} = \hat{\theta_n} = h(X_1, \ldots, X_n)$$

Definition 2.1.4 *(Qualité)* —

- Risque

$$R(\hat{\theta}, \theta) = E_\theta[(\hat{\theta} - \theta)^2]$$

- Consistance

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{P} \theta$$

Definition 2.1.5 *(Modèle identifiable)* —

$$\theta \rightarrow P_\theta \quad \text{injective}$$

2.2 Méthode des moments

Definition 2.2.1 — On appelle **moment théorique** de la loi de X_i d'ordre k:

$$\mu_k = E[X_i^k], \quad k \geq 1$$

Definition 2.2.2 — On appelle **moment empirique** de la loi des X_i d'ordre k:

$$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

Par la loi des grands nombres $\hat{\mu}_k \xrightarrow[n\to+\infty]{P} \mu_k$

La méthode des moments: si on peut écrire θ ou $g(\theta)$ paramètre d'intérêt comme une fonction des k premiers moments théoriques.

$$\theta = \mathcal{L}(\mu_1, \dots, \mu_k)$$

alors l'estimateur

$$\hat{\theta} = \mathcal{L}(\hat{\mu_1}, \dots, \mu_k)$$

est obtenu par la méthode.

Example 2.2.3 — Des calculs des estimateurs en utilisant la méthode des moments.

- $X_i \sim \text{Bernoulli}(\theta)$ à valeurs 0-1, $\theta = P(X_i = 1) = E[X_i] \to \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$
- $X_i \sim \text{Exp}(\theta)$, $f_\theta(x) = \theta e^{-\theta x}1_{x\geq 0}$, $E[X] = \frac{1}{\theta} \Leftrightarrow \theta = \frac{1}{\mu_1}$, par la méthode des moments,

$$\hat{\theta} = \frac{1}{\hat{\mu}_1} = \frac{1}{\overline{X}}$$

$$\begin{aligned} Var_\theta(X_i) = \frac{1}{\theta^2} \Leftrightarrow \theta^2 &= \frac{1}{E[X_i^2] - E[X_i]^2} \\ \Leftrightarrow \theta &= \frac{1}{\sqrt{\mu_2 - \mu_1^2}} \Rightarrow \hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2}} \end{aligned}$$

- $X_1, \dots, X_n$ i.i.d. de la loi P_θ de densité

$$f_\theta(x) = \theta x^{\theta-1}1_{x\in[0,1]}$$

$$E_\theta[X_i] = \theta\int_0^1 x^\theta\, dx = \frac{\theta}{\theta+1}$$

Méthode des moments:

$$\begin{aligned} (\theta+1)\mu_1 = \theta \Leftrightarrow \theta(1-\mu_1) = \mu_1 \Leftrightarrow \theta &= \frac{E[X_i]}{1-E[X_i]} \\ \Rightarrow \hat{\theta}_{MM} = \frac{\overline{X}}{1-\overline{X}}&, P_\theta(\overline{X} = 1) = P_\theta(X_1 = X_2 = \ldots = X_n = 1) = 0 \end{aligned}$$

2.3 Rendu sur le L.A.C.

(L.A.C = lemme des applications continues) $(X_n)_{n\geq 1}$ suite de variables aléatoires. Si X_n converge vers X, que peut-on dire de $g(X_n)_{n\geq 1}$? Si g continue, LAC.

- si $X_n \xrightarrow{P} X$ alors $g(X_n) \xrightarrow{P} g(X)$
- si $X_n \xrightarrow{\mathcal{L}} X$ alors $g(X_n) \xrightarrow{\mathcal{L}} g(X)$

Remark 2.3.1 *(Condition suffisante)* —

$$D_g = \{\text{points de discontinuité de } g\}$$

si $P(X \in D_g) = 0$, le LAC est vrai.

EXAMPLE 2.3.2 —

$$g(x) = \frac{x}{1-x}$$

LGN: $\overline{X} \xrightarrow{P} E[X]$

- LAC: $g(\overline{X}) = \hat{\theta}_n \xrightarrow[n\to+\infty]{P} g(E[X]) = \theta$

LAC pour des couples de suites de variables aléatoires:

- si $(X_n, Y_n) \xrightarrow{P} (X, Y)$, alors $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$, si $g : \mathbb{R}^2 \to \mathbb{R}$ ou $\mathbb{R}^2$ continue
- si $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$, alors $g(X_n, Y_n) \xrightarrow{\mathcal{L}} g(X, Y)$

EXAMPLE 2.3.3 —

$$\hat{\theta}_2 = \frac{1}{\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2 - (\overline{X})^2}} \quad \text{consistant?}$$

LGN:

- $\overline{X} \xrightarrow{P} \mu_1$
- $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \mu_2$

donc

$$\begin{bmatrix} \overline{X} \\ \frac{1}{n}\sum_{i=1}^n X_i^2 \end{bmatrix} \xrightarrow{P} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$g(x, y) = \frac{1}{\sqrt{y-x^2}} \Rightarrow \hat{\theta}^{MM}$ constant de θ, g continue sauf en $\{(x, y) \in \mathbb{R}^2, y = x^2\}$ de mesure nulle.

Mais c'est faux pour une converge en loi.

PROPOSITION 2.3.4 (*CONVERGENCE DE COUPLES*) —

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} X \\ Y \end{pmatrix} \text{ ssi } \begin{cases} X_n \xrightarrow{P} X \\ Y_n \xrightarrow{P} Y \end{cases}$$

Proof —

- $\Rightarrow$ alors LAC $g(x, y) = x$ continue donc $X_n \to X$ et $Y_n \to Y$
- $\Leftarrow$ convergence du couple?

$$\forall \varepsilon > 0, P(|X_n - X| + |Y_n - Y| > \varepsilon) \leq \underbrace{P(|X_n - X| > \frac{\varepsilon}{2})}_{\to 0} + \underbrace{P(|Y_n - Y| > \frac{\varepsilon}{2})}_{\to 0}$$

<u>Cette réciproque est fausse pour la converge en loi!</u>

□

2.3.1 Variance empirique

Si la X_i admettent une esperance μ et une variance σ^2, on appelle variance empirique

$$\begin{aligned} \hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 + \frac{1}{n}\sum_{i=1}^n \overline{X}^2 - \frac{2}{n}\sum_{i=1}^n X_i\overline{X} \\ &= \frac{1}{n}\sum_{i=1}^n X_i^2 + \overline{X}^2 - 2\overline{X}\overline{X} = \tilde{\sigma}^2 \end{aligned}$$

estimateur des moments:

$$\sigma^2 = E[X_i^2] - E[X_i]^2$$

On remplace les moments théoriques par les moments impiriques

$$\rightarrow \tilde{\sigma^2}^{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2$$

Consistance: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\overline{X})^2$,

$$\begin{cases} \overline{X} \xrightarrow{P} E[X] \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} E[X^2] \end{cases} \overset{\text{cv en proba}}{\Rightarrow} \begin{pmatrix} \overline{X} \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 \end{pmatrix} \overset{LAC}{\rightarrow} \hat{\sigma}^2 \text{ consistant de } Var(X) = E[X^2] - E[X]^2$$

Exercise 2.3.1.

- calculer le biais de $\hat{\sigma_n}^2$
- calculer le risque de $\hat{\sigma_n}^2$

2.4 Méthode de maximum de vraisemblance

2.4.1 Modèle donné

$(P_\theta)_{\theta\in\Theta}$ est donné s'il existe une mesure μ(positive σ définie $\rightarrow X_i$ à valeurs dans E, $E = \cup E_n$ avec $\mu(E_n)$ finie) telle que $\forall\theta, P_\theta$ admet une densité par rapport à μ.

2.4.2 En pratique

- soit E au plus dénombrable: μ = mesure de comptage. Si $\exists\{a_1, a_2, \dots\}$ tq $\sum_{k\geq 1} P_\theta(X_i = a_k) = 1$, alors $\mu = \sum_{k\geq 1}\delta_{a_k}$ avec $\delta_a(\{a\}) = 1$ mesure de dirac.

> EXAMPLE 2.4.1 — Bernoulli(θ), $X_i = 1$, probas $\theta \rightarrow \mu = \delta_0 + \delta_1$ On écrira
>
> $$f_\theta(x) = \overset{=1-\theta}{P_\theta(\{x\})} - P_\theta(X_i = x) \text{ avec } x \in \{a_1, a_2, \dots\}$$

- soit $E = \mathbb{R}^p$, alors f_θ est la densité usuelle

f_θ densité de P_θ

> DEFINITION 2.4.2 — On appelle vraisemblance de l'échantillon $(X_1, \dots, X_n)$ la fonction
>
> $$\theta \rightarrow L_n(\theta) = \prod_{i=1}^{n} f_\theta(X_i) \text{ (variable aléatoire)}$$

> DEFINITION 2.4.3 — Un estimateur du max de vraisemblance $\hat{\theta}_{MV}$ est définie par:
>
> $$\forall\theta\in\Theta, L_n(\theta) \leq L_n(\hat{\theta})$$

On travaille souvent avec la **log-vraisemblance**

$$\log L_n(\theta) = \sum_{i=1}^{n} \ln f_\theta(X_i) \text{ somme de variables aléatoires}$$

$$\log L_n(\hat{\theta}) = \sup_{\theta\in\Theta} \log L_n(\theta)$$

REMARK 2.4.4 — $\hat{\theta}$ est une variable aléatoire

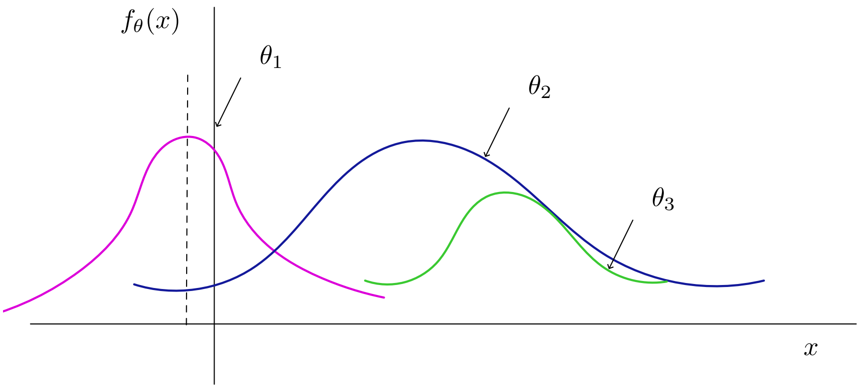

Figure 2.1: remarque-max-vraisemblance-1

EXAMPLE 2.4.5 —

- Bernoulli(θ), $f_\theta(x) = \theta^x(1-\theta)^{1-x}$, X_i à valeurs 0-1

$$L_n(\theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum_{i=1}^n X_i}(1-\theta)^{n-\sum_{i=1}^n X_i}$$

$$\log L_n(\theta) = (\sum_{i=1}^{n} X_i)\ln\theta + (n - \sum_{i=1}^{n} X_i)\ln(1-\theta)$$

$$(\log L_n)'(\theta) = \frac{\sum_{i=1}^n X_i}{\theta} - \frac{n - \sum_{i=1}^n X_i}{1-\theta} = \frac{\sum_{i=1}^n X_i - n\theta}{\theta(1-\theta)}(\overline{X} - \theta)$$

Equation de vraisemblance:

$$(\log L_n)'(\theta) = 0 \Leftrightarrow (1-\theta)\sum_{i=1}^{n} X_i = (n - \sum_{i=1}^{n} X_i)\theta) \Leftrightarrow \sum_{i=1}^{n} X_i = n\theta \Rightarrow \theta = \frac{\sum_{i=1}^n X_i}{n}$$

le point critique, est-il un maximum?

- La dérivée change de signe en $\overline{X} \to$ on a bien un max $\to \hat{\theta}^{MV} = \overline{X}$
- Condition du 2nd ordre, si $(\log L_n)''(\theta) < 0$ pour tout $\theta \Rightarrow \log L_n$ est concave $\Rightarrow$ max global

$$(\log L_n)''(\theta) = -\frac{\sum_{i=1}^n X_i}{\theta^2} - \frac{n - \sum_{i=1}^n X_i}{(1-\theta)^2} < 0, \forall\theta$$