17.11.2025
lundi, 17 novembre 2025 08:08

Rappel: Loi des grds nombres : Si $X_1, X_2, \ldots$ i.i.d.,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s.} E[X_1] \quad \text{dès que } E[|X_1|] < \infty$$

Thm central limit: Même setup,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - E[X_i]) \xrightarrow{\text{loi}} \mathcal{N}(0, \mathrm{Var}(X_1))$$

$$\text{si } E[X_1^2] < \infty.$$

Chapitre 4: Intro à la statistique

Motivation: On se donne une suite de résultats d'une même experience $\rightsquigarrow$ résultat d'une suite de nombres aléatoires, et on veut comprendre la loi de ces nombres.

Déf: Soit $\mathbb{P}$ une mesure de proba sur $\mathbb{R}^d$. Soit $n \geq 1$.
Un n-échantillon de loi $\mathbb{P}$ est une famille indep. de vect. al. $X_1, X_2, \ldots, X_n \sim \mathbb{P}$.
Une réalisation d'un n-échantillon $X_1, X_2, \ldots, X_n$, est une suite de vecteurs $(x_1, x_2, \ldots, x_n) \in (\mathbb{R}^d)^n$

But: trouver $\mathbb{P}$ à partir de $(x_1, \ldots, x_n)$.

— [illegible] à partir de $(x_1, \dots, x_n)$.

Déf: Une <u>statistique</u> sur un n-échantillon $(X_1, \dots, X_n)$ est une fonction $Y : (\mathbb{R}^d)^n \to \mathbb{R}$.

↳ • $Y(X_1, X_2, \dots, X_n)$ sera une v.c.

• $Y(x_1, x_2, \dots, x_n)$ est un nombre

Expl: On lance 10 pièces de monnaie et on note 0 pour pile, 1 pour face.

X_i = résultat du $i^{\text{ème}}$ lancé.

On fait les lancées et on trouve

0, 0, 1, 0 0, 1, 1, 1, 0, 1

$(X_1, \dots, X_{10})$ sera un 10-échantillon de loi $\mathbb{P}$ (qu'on ne connait pas)

$(0, 0, 1, 0 0, 1, 1, 1, 0, 1)$ est une réalisation du 10-éch.

<u>Expl de statistiques:</u> · $\frac{1}{10} \sum_{i=1}^{10} X_i$ est une stat.

· $X_1 + X_2 + X_3$ est une stat

· X_5 est une stat.

<u>Un premier exemple d'estimation paramétrique</u>:

<u>Expérience</u>: lancé de pièce. 0 pour pile

<u>Expérience</u> : lancé de pièce : 0 pour pile
1 pour face.

$(X_1, X_2, \ldots, X_n)$ un n-échantillon de loi $\mathbb{P}$ à déterminer.

⚠ $\mathbb{P}$ est une mesure de proba sur $\mathbb{R}$

a priori $\mathbb{P} \in \{ \mathbb{P} : \mathcal{P}(\mathbb{R}) \to [0,1] : \text{mesure de proba} \}$.

<u>Idée 1</u> : réduire l'espace : ici on sait que

$\mathbb{P}(\{0,1\}) = 1$ et donc $\mathbb{P}(\{0\}) + \mathbb{P}(\{1\}) = 1$

↳ pour comprendre $\mathbb{P}$, il suffit de calculer $p = \mathbb{P}(\{1\}) = \mathbb{P}(X_i = 1)$.

⇝ on fait l'hypothèse qu'il existe $p \in [0,1]$ t.q.

$\mathbb{P} = \text{Bern}(p)$. ⊛

<u>Idée 2</u> : On cherche une statistique qui approxime p.

$$\hat{p}_n = \hat{p}_n(X_1, \ldots X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

<u>Questions</u> : ① est-ce que $\hat{p}_n$ donne la bonne valeur en moyenne ?

② est-ce que augmenter n améliore

② est-ce que augmenter n améliore mon estimation ?

③ à n fixé, quelle est la proba que mon estimation soit mauvaise ?

① $$\mathbb{E}\left[\hat{p}_n(X_1,\dots,X_n)\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] \overset{\text{hyp}}{=} \frac{1}{n}\sum_{i=1}^{n} p = p$$

<u>Rem:</u> vrai pour tout n...

② la loi forte des grands nombres dit

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s.} \mathbb{E}[X_1] = p.$$

↳ augmenter n nous rapproche bien de la valeur voulue.

③ Proba de ~~se~~ planter : $\varepsilon > 0$ erreur d'approximation

On cherche à calculer

$$\mathbb{P}\left(|\hat{p}_n - p| \geq \varepsilon\right) \underset{\mathbb{E}[X_i]=p}{=} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i]\right| \geq \varepsilon\right)$$

$$= \mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])\right| \geq \sqrt{n}\,\varepsilon\right)$$

$$= \mathbb{P}\left(\left|\underbrace{\frac{1}{\sqrt{np(1-p)}}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])}_{n \text{ grand},\ \simeq \mathcal{N}(0,1)\ \text{par CLT}}\right| \geq \frac{\sqrt{n}\,\varepsilon}{\sqrt{p(1-p)}}\right)$$

n grand, $\approx \mathcal{N}(0,1)$ par CLT

$$\approx \mathbb{P}\left(|\mathcal{N}(0,1)| \geq \frac{\sqrt{n}\,\varepsilon}{\sqrt{p(1-p)}} \right)$$

$$= 2\mathbb{P}\left(\mathcal{N}(0,1) \geq \frac{\sqrt{n}\,\varepsilon}{\sqrt{p(1-p)}} \right) \quad (**)$$

On ne connait pas p, on prend la valeur qui maximise cette proba d'erreur :

$$(**) \leq 2\mathbb{P}\left(\mathcal{N}(0,1) \geq 2\sqrt{n}\,\varepsilon \right) \qquad (p = 1/2)$$

$$= \frac{2}{\sqrt{2\pi}} \int_{2\sqrt{n}\varepsilon}^{\infty} e^{-x^2/2}\, dx \leq \frac{1000}{\sqrt{n}\,\varepsilon} e^{-\varepsilon^2 n/2}$$

Récap :

- Experience + mesure $\rightsquigarrow$ vect. al. X de loi $\mathbb{P}$: $\mathbb{P}(A) = \mathbb{P}(X \in A)$.
- On répète n fois l'exp. $\rightsquigarrow$ n-échantillon $(X_1, \dots, X_n)$ avec $X_i \sim \mathbb{P}$
- On trouve une réalisation $(x_1, \dots, x_n)$ de $(X_1, \dots, X_n)$ et on cherche à déterminer $\mathbb{P}$ à partir de $(x_1, \dots, x_n)$.

Estimateurs :

Déf. (estimation paramétrique) On suppose que $\mathbb{P}$ appartient à une famille $(\mathbb{P}_\theta)_{\theta \in \Theta}$ paramétrisée par à une famille $(\mathbb{P}_\theta)_{\theta \in \Theta}$ paramétrisée par $\theta \in \Theta \subset \mathbb{R}^m$ pour un $m \geq 1$ fixé.

θ est le <u>paramètre</u> et Θ est l'<u>espace des paramètres</u>.

<u>Expl</u>: 1) On tire au hasard un humain dans la population suisse et on mesure sa masse.

↳ On pourrait se dire que la masse suit une loi $\mathcal{N}(\mu, \sigma^2)$.

↳ $\theta = (\mu, \sigma^2)$, $\Theta = \mathbb{R} \times [0, +\infty)$.

2) Durée de vie d'un ordinateur en salle d'info à l'EPFL.

↳ On pourrait prendre une loi $\mathrm{Exp}(\lambda)$:

$\theta = \lambda$, $\Theta = (0, +\infty)$.

3) Lancé d'un dé à 20 faces (non-équilibré) :

$$\Theta = \left\{ p \in [0,1]^{\{1, \dots, 20\}} : \sum_{i=1}^{20} p_i = 1 \right\}$$

$\theta \in \Theta \subset \mathbb{R}^{20}$

<u>Déf</u>: Soit $(X_1, \dots, X_n)$ un n-échantillon de loi $\mathbb{P}_\theta$, $\theta \in \Theta$. Soit $f : \Theta \to \mathbb{R}$. Un <u>estimateur</u> de $f(\theta)$ est une <u>statistique</u> $\hat{f} : (\mathbb{R}^d)^n \to \mathbb{R}$ qui ne dépend pas de θ.

<u>Idée</u>: on approxime $f(\theta)$ par $\hat{f}(X_1, \dots, X_n)$.

Idée : on approxime $f(\theta)$ par $\hat{f}(X_1, \dots, X_n)$.

Pourquoi f plutôt que θ ?

expl : $E[X]$ n'est pas forcément $=$ à θ.

mais $E[X] = f(\theta)$ pour f bien choisit.

Expl d'estimateur : $\frac{1}{n} \sum_{i=1}^{n} X_i$ est un estimateur de $E[X_1]$.

Déf. Soit $f : \Theta \to \mathbb{R}$ une quantité à estimer. Pour $n \geq 1$, on se donne un n-échantillon de loi $\mathbb{P}_\theta$, $\theta \in \Theta$.
Une suite $\hat{f}_n(X_1, \dots, X_n)$ d'estimateurs de $f(\theta)$ est dite convergente si $\forall \varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}\left(|\hat{f}_n(X_1, \dots, X_n) - f(\theta)| \geq \varepsilon \right) = 0.$$

(i.e. $\hat{f}_n \xrightarrow{\text{proba}} f(\theta)$).

Déf : Soit $(X_1, \dots, X_n)$ un n-échantillon de loi $\mathbb{P}_\theta$.
Soit $f(\theta)$ une quantité à estimer et soit $\hat{f}_n(X_1, \dots, X_n)$ un estimateur de $f(\theta)$. On définit le biais de $\hat{f}_n$ par :

$$E\left[\hat{f}_n(X_1, \dots, X_n) \right] - f(\theta) =: \text{Biais}_\theta(\hat{f}_n)$$

$E[\hat{g}_n(X_1,\dots,X_n)] - g(\theta) =: \text{Biais}_\theta(\hat{g}_n)$

On dit qu'une suite d'estimateurs $(\hat{g}_n)_{n\geq 1}$ de $g(\theta)$ est <u>asymptotiquement sans biais</u> si

$$\lim_{n\to\infty} \text{Biais}_\theta(\hat{g}_n) = 0.$$

La <u>variance</u> d'un estimateur $\hat{g}_n$ de $g(\theta)$ est $\text{Var}(\hat{g}_n(X_1,\dots,X_n)) =: \text{Var}_\theta(\hat{g}_n)$.

⚠ La dépendance des quantités en θ est souvent implicite mais elle est bien présente : $X_1,\dots,X_n \sim \mathbb{P}_\theta$.

<u>Lemme</u> : Si $(\hat{g}_n)_{n\geq 1}$ est une suite d'estimateur de $g(\theta)$ telle que :

i) $(\hat{g}_n)_{n\geq 1}$ est asymptotiquement ss. biais,

ii) $\text{Var}_\theta(\hat{g}_n) \xrightarrow{n\to\infty} 0$,

alors $(\hat{g}_n)_{n\geq 1}$ est une suite convergente d'estimateurs.

*<u>Pv</u> : Par Chebychev, $\forall \varepsilon > 0$

$$\mathbb{P}(|\hat{g}_n - g(\theta)| \geq \varepsilon) \leq \frac{E[(\hat{g}_n - g(\theta))^2]}{\varepsilon^2}$$

$$\frac{\qquad\qquad}{\varepsilon^2}$$

$$= \frac{E\left[\left(\hat{f}_n - E[\hat{f}_n] + E[\hat{f}_n] - f(\theta)\right)^2\right]}{\varepsilon^2}$$

$$= \frac{1}{\varepsilon^2}\left(E\left[\left(\hat{f}_n - E[\hat{f}_n]\right)^2\right] + \left(E[\hat{f}_n] - f(\theta)\right)^2 + \underbrace{2\,\overbrace{E\left[\hat{f}_n - E[\hat{f}_n]\right]}^{=0}\left(E[\hat{f}_n] - f(\theta)\right)}_{0} \right)$$

$$= \frac{1}{\varepsilon^2}\left(\underbrace{\mathrm{Var}_\theta(\hat{f}_n)}_{\substack{\to 0 \\ ii)}} + \underbrace{\mathrm{Biais}_\theta(\hat{f}_n)^2}_{\substack{\to 0 \\ i)}} \right)$$

$\boxtimes$