Chapitre I : Introduction

MATH-F-207

Davy Paindaveine

Bloc 2, Bachelier en sciences mathématiques
Université libre de Bruxelles

2023–2024

Contenu du chapitre

Le lien entre probabilités et statistique

La notion de modèle statistique

Trois exemples récurrents

Contenu du chapitre

Une expérience aléatoire

Considérons l'expérience consistant à servir un verre de bière.
La quantité X (en cl) contenue dans le verre est une variable aléatoire.

Rappel : Une v.a. X sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ est une fonction mesurable de Ω vers $\mathbb{R}$.

Rappel : La distribution P^X de X est définie par

$$\mathrm{P}^X[B] := \mathrm{P}[X^{-1}(B)] = \mathrm{P}[\{\omega \in \Omega : X(\omega) \in B\}]$$

pour tout B dans la σ-algèbre de Borel $\mathcal{B}$ sur $\mathbb{R}$.

Différentes quantités d'intérêt

Ici, X est continue, donc admet une fonction de densité f^X.

Ceci permet de calculer

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f^X(x)\, dx,$$

ou

$$\mathrm{P}[X > 20] = \int_{20}^{\infty} f^X(x)\, dx.$$

Probabilités vs Statistique

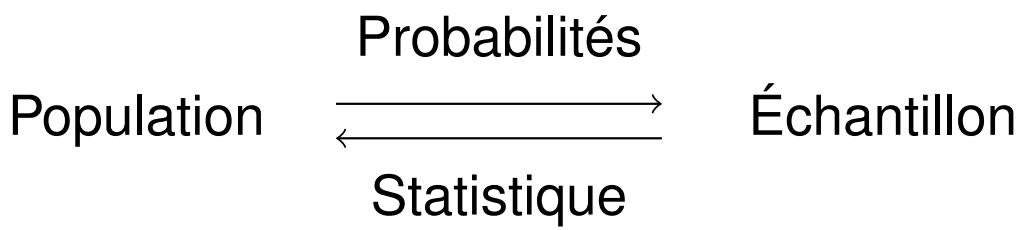
Les probabilités permettent d'évaluer ces quantités sur base de P^X.
Mais, en pratique, P^X est inconnue...

Sur base d'un échantillon $x_1, \ldots, x_n$ (des réalisations de v.a. $X_1, \ldots, X_n$ i.i.d. avec la même distribution que X), la statistique vise à obtenir des informations sur $\mathrm{E}[X]$ et $\mathrm{P}[X > 20]$, ou plus généralement sur P^X.

On parlera de processus inférentiel ou d'inférence statistique.

Probabilités vs Statistique (2)

Les deux processus sont donc, par nature, antagonistes:



Contenu du chapitre

Modèle statistique

Soit $X^{(n)} = (X_1, \ldots, X_n)$ un vecteur de v.a. X_i à valeurs dans $\mathbb{R}$.
On appellera souvent $X^{(n)}$ l'observation.

La distribution jointe de $X^{(n)}$, notée $\mathrm{P}^{(n)}$, est inconnue.

$\rightsquigarrow$ On considérera un modèle statistique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)}\right),$$

où $\mathcal{P}^{(n)}$ est une collection de mesures de probabilité sur $(\mathbb{R}^n, \mathcal{B}^n)$,
dont une seule constitue la "vraie" distribution de $X^{(n)}$.

Modèle d'échantillonnage paramétrique

Un modèle *d'échantillonnage paramétrique* suppose que $X_1, \ldots, X_n$ sont *i.i.d.*, de distribution commune appartenant à

$$\mathcal{P} = \Big\{ \mathrm{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k \Big\}.$$

Modèle d'échantillonnage paramétrique

Un modèle *d'échantillonnage paramétrique* suppose que $X_1, \ldots, X_n$ sont *i.i.d.*, de distribution commune appartenant à

$$\mathcal{P} = \left\{ \mathrm{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k \right\}.$$

Le terme "échantillonnage" précise que les observations X_i sont i.i.d.

Modèle d'échantillonnage paramétrique

Un modèle *d'échantillonnage paramétrique* suppose que $X_1, \ldots, X_n$ sont *i.i.d.*, de distribution commune appartenant à

$$\mathcal{P} = \left\{ \mathrm{P}_\theta : \theta \in \Theta \subset \mathbb{R}^k \right\}.$$

Le terme "échantillonnage" précise que les observations X_i sont i.i.d.

Le qualificatif "paramétrique" impose que la famille de distributions est indicée par un paramètre fini-dimensionnel.

Modèle d'échantillonnage paramétrique: exemple

Le modèle d'échantillonnage paramétrique gaussien

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{ \mathrm{P}_{\mu,\sigma^2}^{\mathcal{N}(n)} : \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_0^+ \right\}\right)$$

suppose que $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$.

Pour ce modèle paramétrique,

$$\theta = \begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \quad \Theta = \mathbb{R} \times \mathbb{R}_0^+ \quad \text{et} \quad k = 2.$$

Modèle d'échantillonnage non paramétrique

Un modèle d'échantillonnage non paramétrique est

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_f^{(n)} : f \in \mathcal{F}\right\}\right),$$

où $\mathrm{P}_f^{(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. avec une loi commune admettant la fonction de densité f.

L'ensemble $\mathcal{F}$ des densités sur $\mathbb{R}$ est un espace de fonctions qui, si on ne fait pas d'hypothèses sur la forme de f, est de dimension infinie.

Modèles sans échantillonnage

Si les observations ne sont pas i.i.d., il faudra plutôt adopter un modèle paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{ \mathrm{P}_\theta^{(n)} : \theta \in \Theta \subset \mathbb{R}^k \right\}\right)$$

ou un modèle non paramétrique

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^{(n)} = \left\{ \mathrm{P}_f^{(n)} : f \in \mathcal{F}^{(n)} \right\}\right),$$

qui prévoient un ensemble de lois possibles pour la distribution jointe de $X^{(n)}$.

(Ici, $\mathcal{F}^{(n)}$ désigne l'ensemble de toutes les densités sur $\mathbb{R}^n$).

Contenu du chapitre

L'exemple du verre de bière

Soit X la quantité de bière (en cl) servie dans un verre.

Nous adopterons le modèle d'échantillonnage paramétrique gaussien

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{ \mathrm{P}_{\mu,\sigma^2}^{\mathcal{N}(n)} : \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_0^+ \right\}\right),$$

où $\mathrm{P}_{\mu,\sigma^2}^{\mathcal{N}(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathcal{N}(\mu, \sigma^2)$.

Parfois, nous considérerons le sous-modèle pour lequel la variance prend une valeur σ_0^2 connue. Le paramètre sera alors $\theta = \mu \in \Theta = \mathbb{R}$.

L'exemple du bus

Supposons qu'un bus se présente toutes les $\theta(>0)$ minutes à l'arrêt.

Le temps d'attente (en min) d'une personne arrivant à un moment arbitraire à cet arrêt est alors une v.a. X de loi uniforme sur $[0, \theta]$ (on écrira $X \sim \mathrm{Unif}([0, \theta])$).

Si cette personne monte systématiquement dans le bus dès qu'un bus se présente, la valeur de θ reste inconnue.

Rappel : $X \sim \mathrm{Unif}([a, b]) \Leftrightarrow X$ admet la densité

$$x \mapsto f^X(x) = \frac{1}{b-a} \mathbb{I}[a \leq x \leq b].$$

L'exemple du bus (2)

Des répétitions indépendantes de cette expérience mènent au modèle d'échantillonnage uniforme

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_\theta^{(n)} : \theta \in \mathbb{R}_0^+ \subset \mathbb{R}\right\}\right),$$

où $\mathrm{P}_\theta^{(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathrm{Unif}([0, \theta])$. Contrairement à l'exemple précédent, le paramètre est scalaire plutôt que vectoriel.

L'exemple électoral

Considérons, dans un scrutin électoral entre deux candidats,
la proportion p (inconnue) de la population en faveur du candidat A.

Si on sonde un individu au hasard, sa préférence est représentée
par une v.a. X de loi de Bernoulli de paramètre p ($X \sim \mathrm{Bern}(p)$):

- $X = 1$ (être en faveur du candidat A) avec probabilité p, et
- $X = 0$ (être en faveur du candidat B) avec probabilité $1 - p$.

Rappel : $X \sim \mathrm{Bern}(p) \Leftrightarrow P[X = x] = p^x (1-p)^{1-x} \mathbb{I}[x \in \{0, 1\}]$.

L'exemple électoral (2)

Si on sonde n personnes de façon indépendante dans la rue sur leurs intentions de vote, on obtient le modèle d'échantillonnage de Bernoulli

$$\left(\mathbb{R}^n, \mathcal{B}^n, \mathcal{P}^n = \left\{\mathrm{P}_p^{(n)} : p \in [0,1] \subset \mathbb{R}\right\}\right),$$

où $\mathrm{P}_p^{(n)}$ désigne la distribution de $X^{(n)} = (X_1, \ldots, X_n)$ lorsque $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathrm{Bern}(p)$.

Contrairement aux deux exemples précédents,
(i) la distribution commune des observations X_i est discrète
(ii) l'espace paramétrique ($\Theta = [0,1]$) est compact.

Résumé du chapitre

- Introduction des différents types de modèles statistiques
- Description des trois exemples principaux du cours