

---

# Table des matières

---

<b>1</b>	<b>Introduction à l'estimation paramétrique</b>	<b>2</b>
1.1	Prérequis . . . . .	2
1.1.1	Définition de base . . . . .	2
1.1.2	Convergence . . . . .	4
1.2	Modélisation statistique . . . . .	4
1.2.1	Modèle paramétrique . . . . .	4
1.3	Estimateurs . . . . .	5

# INTRODUCTION À L'ESTIMATION PARAMÉTRIQUE

## Sommaire

1.1	Prérequis . . . . .	2
1.2	Modélisation statistique . . . . .	4
1.3	Estimateurs . . . . .	5

## 1.1 Prérequis

### 1.1.1 Définition de base

**Définition :** *espace probabilisé*, espace probabilisable.

Un espace probabilisé est un triplet  $(\Omega, \mathcal{F}, \mathbb{P})$ .  $\Omega$  est l'univers, ou l'ensemble des événements possibles,  $\mathcal{F}$  est une  $\sigma$ -algèbre et  $\mathbb{P}$  est une mesure de probabilité. De même on appelle espace probabilisable un couple  $(\Omega, \mathcal{F})$

**Définition :**  $\sigma$ -Algèbre.

Soit  $\Omega$  un ensemble, une  $\sigma$ -algèbre  $\mathcal{F}$  est un ensemble de partie de  $\Omega$  vérifiant trois règles :

- $\Omega \in \mathcal{F}$
- $\forall A \in \mathcal{F}, A^c \in \mathcal{F}$
- $\forall A_1, \dots, A_n, \dots \in \Omega, \bigcup_{i>1} A_i \in \mathcal{F}$

#### Exemples

- La tribu pleine :  $\mathcal{P}(\Omega)$
- la tribu triviale  $\{\emptyset, \Omega\}$
- la tribu borélienne  $\mathcal{B}(\mathbb{R}) = \{]a, b[, ]a, b[ \subset \mathbb{R}\}$  formée des ouverts de  $\mathbb{R}$

#### Exercice :

- Soit un jeu de pile ou face effectué deux fois, quelle est l'univers et la tribu résultante
- Soit un lancé de dé effectué deux fois, quelle est l'univers et la tribu résultante

**Définition :** *probabilité*.

On appelle loi de probabilité, ou tout simplement probabilité une mesure  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  vérifiant :

- $\mathbb{P}(\Omega) = 1$  la probabilité de l'événement certain est 1

—  $\mathbb{P}(\bigcup_{i>1} A_i) = \sum_i \mathbb{P}(A_i)$  pour des événements deux à deux disjoints.

**Remarque** : Si  $A$  est un événement de  $\Omega$  alors  $\mathbb{P}(A) = \frac{\text{Cardinal}(A)}{\text{Cardinal}(\Omega)}$ .

**Définition** : *indépendance de deux événements.*

Deux événements  $A, B$  sont indépendants si  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

**Définition** : *probabilité conditionnelle.*

Soit  $A, B \in \Omega$  deux événements. On appelle probabilité conditionnelle de  $A$  sachant  $B$  et on note  $\mathbb{P}(A|B)$  la quantité suivante :  $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

**Définition** : *variable aléatoire.*

Soit un espace probabilisable  $(\Omega, \mathcal{F})$  et un espace mesurable  $(E, \mathcal{E})$ . Une variable aléatoire, abrégée v.a., est une fonction mesurable  $X$  de  $(\Omega, \mathcal{F})$  à valeur dans  $(E, \mathcal{E})$ . Autrement dit c'est une fonction qui vérifie la propriété suivante :

$$\forall A \in \mathcal{E}, X^{-1}(A) \in \mathcal{F}$$

Lorsque  $E = \mathbb{N}$  on parle de variable aléatoire discrète, si  $E = \mathbb{R}$ , on parle de variable aléatoire continue. Lorsque  $E$  est un espace de plusieurs dimension, on parle de vecteur aléatoire, ou de variable aléatoire vectorielle. Une variable aléatoire  $X$  définit une loi de probabilité sur  $(\Omega, \mathcal{F})$  souvent noté  $P_X$ . Ainsi  $\forall B \in \mathcal{E} P_X(B) = \mathbb{P}(X^{-1}(B))$ . En fait variable aléatoire est simplement une fonction pour représenter des événements aléatoires sous forme numérique.

**Exemple(s)** : Dans le cadre d'un jeu de pile ou face l'univers  $\Omega = \{P, F\}$ , typiquement, une v.a. va associer chaque événement à un nombre par exemple  $\{0, 1\}$ . On peut alors attribuer une valeur à chaque événement à l'aide d'une v.a. Ainsi  $\mathbb{P}(X = 0) = \mathbb{P}(P)$  et  $\mathbb{P}(X = 1) = \mathbb{P}(F)$ . Une loi permet de modéliser cet expérience : il s'agit de la loi de Bernoulli.

**Exercice** : Donnez la loi associée au lancé d'un dé équilibré.

**Définition** : *fonction de répartition.*

Soit  $X$  une v.a. définie sur un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ . On appelle fonction caractéristique associée à  $X$  la fonction  $F_X(x) := \mathbb{P}(X < x)$ .

**Définition** : *espérance d'une v.a.*

Soit  $X$  une v.a. définie sur  $(\Omega, \mathcal{F}, \mathbb{P})$ , on appelle espérance de  $X$  la quantité suivante lorsqu'elle est définie :

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) \quad \text{si } X \text{ une v.a. discrète} \quad (1.1)$$

$$\mathbb{E}[X] = \int_{\omega} X(\omega) d\mathbb{P}(\omega) \quad \text{si } X \text{ une v.a. continue} \quad (1.2)$$

**Définition** : *variance d'une v.a.*

On appelle variance d'une v.a  $X$  la quantité suivante :

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

### 1.1.2 Convergence

Dans la suite on se place dans un espace probabilisé  $(\Omega, \mathcal{F}, \mathbb{P})$ . Soit  $(X_n)$  une suite de v.a et  $X$  une v.a à définir dans l'espace probabilisé. Nous allons énoncer plusieurs types de convergence utiles par la suite.

**Définition :** *convergence presque sûre.*

$(X_n)$  converge presque sûrement vers  $X$  et on note  $X_n \xrightarrow{p.s} X$ , lorsque la convergence est vraie avec une probabilité égale à 1.

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1$$

Autrement dit, la convergence est presque sûre si les événements  $\omega$  pour lesquels la suite  $(X_n)$  ne converge pas ont une probabilité nulle.

$$\mathbb{P}(\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)) = 1$$

**Définition :** *convergence en moyenne quadratique.*

$(X_n)$  converge en moyenne quadratique vers  $X$  et on note  $X_n \xrightarrow{L^2} X$  lorsque :

$$\lim_{n \rightarrow \infty} \mathbb{E}[(X_n - X)^2] = 0$$

**Définition :** *convergence en probabilité.*

$(X_n)$  converge en probabilité vers  $X$  et on note  $X_n \xrightarrow{p} X$  lorsque :

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

**Définition :** *convergence en loi.*

Soit  $(F_n)$  la suite de fonction de répartition associé à la suite de v.a  $(X_n)$  et  $F$  la fonction de répartition associé à  $X$ . On dit que  $X_n$  converge en loi vers  $X$  et on note  $X_n \xrightarrow{\mathcal{L}} X$  lorsque pour tout point  $x$  où  $F$  est continue,  $F_n(x)$  converge vers  $F(x)$

## 1.2 Modélisation statistique

### 1.2.1 Modèle paramétrique

Un modèle statistique est un objet mathématique associé à l'observation des données provenant d'un phénomène aléatoire.

**Définition :** *modèle statistique.*

Un modèle statistique est la donnée d'un espace probabilisable  $(\Omega, \mathcal{F})$  et d'une famille de loi de probabilité  $\mathcal{L}$ . Lorsque cette famille de loi peut s'écrire sous la forme  $\mathbb{P}_\Theta := \{\mathbb{P}_\theta, \theta \in \Theta\}$  ( $\Theta$  est l'espace des paramètres) on parle de modèle statistique paramétrique ou plus simplement de modèles paramétriques. Sinon on parle de modèles non-paramétriques.

Une expérience statistique consiste à observer des données souvent représentées par une suite  $x_1, \dots, x_n$  et à les interpréter avec un modèle statistique. Cela revient à voir les données  $x_i$  comme des réalisations de v.a  $X_i$ . Dans ce cours nous nous concentrerons exclusivement sur les modèles paramétriques.

**Définition :** *expérience statistique.*

Une expérience statistique est la donnée d'un modèle paramétrique  $(\Omega, \mathcal{F}, \mathbb{P}_\Theta)$  et d'une variable aléatoire  $X$  à définir dans  $(\Omega, \mathcal{F})$ .

**Définition :** *identifiabilité d'un modèle paramétrique*

Un modèle paramétrique est dit identifiable si  $\forall \theta, \theta' \in \Theta$

$$\mathbb{P}_\theta = \mathbb{P}_{\theta'} \implies \theta = \theta'$$

**Exercice :** Montrer que le modèle  $\mathcal{P} = \{\mathcal{N}(\theta, 1), \theta \in \mathbb{R}\}$  est identifiable.

Deux types d'estimation existent : l'estimation ponctuelle et l'estimation par intervalle de confiance. Nous allons nous concentrer dans les premiers chapitres sur l'estimation ponctuelle. Vers la fin du cours nous ferons une ouverture sur l'estimation par intervalle de confiance.

## 1.3 Estimateurs

Dans l'interprétation fréquentiste des phénomènes aléatoires on considère que la loi est connue, mais que le paramètre est inconnu mais fixe. Un estimateur est donc une fonction permettant d'estimer le paramètre inconnu à partir des données.

**Définition :** *statistique.*

Soit un espace mesurable  $(E, \mathcal{E})$ , une statistique est une fonction mesurable  $T$  quelconque défini sur  $(E, \mathcal{E})$ .

**Exemple(s) :** Soit  $x_1, \dots, x_n \in E$

- $T(x_1, \dots, x_n) = \max(x_1, \dots, x_n)$
- $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$

**Remarque :** Les définitions de variable aléatoire et de statistique sont pratiquement les mêmes. En pratique cependant, une statistique est définie sur l'espace de sortie des variables aléatoires.

**Définition :** *estimateur.*

Soit une expérience statistique associée à un modèle paramétrique  $(\Omega, \mathcal{F}, \mathbb{P}_\Theta)$  et  $X_1, \dots, X_n$

des v.a identiquement et indépendamment distribuées. Un estimateur  $T_n$  du paramètre  $\theta$  est une statistique définie sur  $\Omega, \mathcal{F}$  à valeur dans l'espace des paramètres possibles  $\Theta$ .

$$T_n : \Omega^n \rightarrow \Theta \quad (1.3)$$

$$(x_1, \dots, x_n) \mapsto t_n \quad (1.4)$$

Une estimation de  $\theta$  est une réalisation  $t_n$  de l'estimateur  $T_n$ . Il faut bien distinguer un estimateur qui est une v.a et une estimation qui est une valeur déterministique. En effet un estimateur est une v.a car c'est une fonction de v.a, donc ses réalisations varient en fonction des valeurs d'entrée. Les méthodes classiques pour construire les estimateurs sont la méthode des moments et la méthode du maximum de vraisemblance. Nous nous intéresseront à cette dernière méthode dans ce cours. Dans la suite on considère que les données sont identiquement et indépendamment distribuées i.i.d. C'est à dire que les réalisations sont de même loi et indépendantes.

**Exercice :** Exprimez la différence entre estimateur et estimation en utilisant l'exemple discuté dans le TP : 'Notions de base de modélisation statistiques'.

### Exemple(s)

Soit  $X_1, \dots, X_n$  une suite de v.a

- L'estimateur de la moyenne souvent noté  $\bar{X}_n$  est donné par la formule suivante :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

- L'estimateur de la variance empirique est la moyenne des carrés des écarts à la moyenne empirique :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- L'estimateur de la variance empirique corrigé :

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

**Définition :** *fonction de vraisemblance.*

Soit  $(x_1, \dots, x_n)$  un échantillon issu d'une expérience statistique. On appelle fonction de vraisemblance ou tout simplement vraisemblance la fonction suivante

$$\mathcal{L}(x_1, \dots, x_n; \theta) = \mathbb{P}_\theta(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i)$$

Ainsi la vraisemblance est la probabilité d'observer les données en fixant le paramètre  $\theta$ .

**Définition :** *estimateur du maximum de vraisemblance.*

L'estimateur du maximum de vraisemblance  $\hat{\theta}_n$  est un estimateur permettant de trouver la valeur de  $\theta$  rendant la vraisemblance maximale.

$$\hat{\theta}_n = \arg \max_{\theta} \mathcal{L}(x_1, \dots, x_n; \theta) = \arg \max_{\theta} \log \mathcal{L}(x_1, \dots, x_n; \theta)$$

En pratique on utilise maximizer log-vraisemblance, en effet le logarithme est une fonction croissante.

**Définition :** *score*.

On appelle score le gradient de la log-vraisemblance.

$$s(\theta) = \frac{\partial \log \mathcal{L}(x_1, \dots, x_n; \theta)}{\partial \theta}$$

En pratique on utilise le score pour trouver le maximum de vraisemblance en cherchant la valeur de  $\theta$  annulant le score.

**Exercice :** Soit  $x \sim \mathcal{N}(\mu, \sigma^2)$ .

1. Donnez l'expression du score .
2. Montrer que l'espérance du score est nul de manière générale.
3. On considère un jeu de données  $x_1, \dots, x_n$  i.i.d suivant  $\mathcal{N}(\mu, \sigma^2)$ . Exprimez la vraisemblance de cette loi et donnez une estimation du maximum de vraisemblance pour  $\mu$  et  $\sigma^2$ .
4. Refaire la question 3 pour une loi de Bernoulli.