Statistique

Notes de Cours

Licence L3 — 2025–2026

Yaë Ulrich Gaba

« Il est facile de mentir avec des statistiques, mais il est plus facile encore de mentir sans elles. »
— Frederick Mosteller

25 mars 2026

Table des matières

Préface **vii**

1 Statistique Descriptive — Résumé et Visualisation **1**
1.1 Exemple motivant . . . 1
1.2 Population, échantillon, variables . . . 1
1.3 Mesures de tendance centrale . . . 2
1.4 Mesures de dispersion . . . 2
1.5 Mesures de forme . . . 2
1.6 Représentations graphiques . . . 3
1.6.1 Histogramme . . . 3
1.6.2 Boîte à moustaches (boxplot) . . . 3
1.6.3 Diagramme en bâtons et diagramme circulaire . . . 3
1.6.4 Fonction de répartition empirique . . . 4
1.7 Données bivariées . . . 4
1.8 Implémentation Python . . . 4
1.9 Exercices . . . 6

2 Modèle Statistique et Échantillonnage **7**
2.1 Exemple motivant . . . 7
2.2 Modèle statistique . . . 7
2.3 Échantillon aléatoire . . . 8
2.4 Statistiques et distributions d'échantillonnage . . . 8
2.5 Loi des grandes statistiques d'échantillonnage . . . 8
2.5.1 Moyenne empirique . . . 8
2.5.2 Loi des grands nombres . . . 9
2.5.3 Théorème central limite . . . 9
2.6 Distribution de la variance empirique . . . 9
2.7 Théorème de Glivenko–Cantelli . . . 10
2.8 Implémentation Python . . . 10
2.9 Exercices . . . 11

3 Estimation Ponctuelle — EMV et Méthode des Moments **13**
3.1 Exemple motivant . . . 13
3.2 Cadre général . . . 13
3.3 Méthode des moments . . . 13
3.4 Maximum de vraisemblance . . . 14
3.4.1 Exemples fondamentaux . . . 14
3.5 Propriétés de l'EMV . . . 15
3.6 Information de Fisher . . . 15

3.7 Implémentation Python . . . 15
3.8 Exercices . . . 17

4 Propriétés des Estimateurs — Biais, Convergence, Efficacité 19
4.1 Exemple motivant . . . 19
4.2 Biais d'un estimateur . . . 19
4.3 Erreur quadratique moyenne . . . 20
4.4 Convergence (consistance) . . . 20
4.5 Efficacité et borne de Cramér–Rao . . . 21
4.6 Estimateur UMVUE et théorème de Rao–Blackwell . . . 21
4.7 Implémentation Python . . . 22
4.8 Exercices . . . 23

5 Intervalles de Confiance 25
5.1 Exemple motivant . . . 25
5.2 Définition et interprétation . . . 25
5.3 IC pour la moyenne . . . 25
5.3.1 Variance connue . . . 25
5.3.2 Variance inconnue . . . 26
5.4 IC pour la variance . . . 26
5.5 IC pour une proportion . . . 26
5.6 IC pour la différence de deux moyennes . . . 27
5.7 Taille d'échantillon . . . 27
5.8 Implémentation Python . . . 27
5.9 Exercices . . . 29

6 Tests d'Hypothèses — Cadre Général 31
6.1 Exemple motivant . . . 31
6.2 Formulation des hypothèses . . . 31
6.3 Erreurs de première et deuxième espèce . . . 31
6.4 Structure d'un test . . . 32
6.5 Puissance d'un test . . . 32
6.6 Lemme de Neyman–Pearson . . . 33
6.7 Test du rapport de vraisemblance généralisé . . . 33
6.8 Implémentation Python . . . 33
6.9 Exercices . . . 35

7 Tests Classiques : z, t, χ^2, F 37
7.1 Exemple motivant . . . 37
7.2 Test z (variance connue) . . . 37
7.2.1 Test z pour deux moyennes . . . 37
7.2.2 Test z pour une proportion . . . 37
7.2.3 Test z pour deux proportions . . . 38
7.3 Test t de Student . . . 38
7.3.1 Test t pour deux échantillons indépendants . . . 38
7.3.2 Test t apparié . . . 38
7.4 Test du χ^2 . . . 38
7.4.1 Test sur la variance . . . 38
7.4.2 Test d'adéquation du χ^2 . . . 39

7.4.3 Test d'indépendance du χ^2 . . . 39
7.5 Test F de Fisher . . . 39
7.6 Résumé : choix du test . . . 39
7.7 Implémentation Python . . . 40
7.8 Exercices . . . 41

8 Régression Linéaire Simple **43**
8.1 Exemple motivant . . . 43
8.2 Le modèle . . . 43
8.3 Estimation par moindres carrés ordinaires (MCO) . . . 43
8.4 Décomposition de la variance . . . 44
8.5 Inférence sous normalité . . . 44
8.5.1 Prédiction . . . 45
8.6 Diagnostic des résidus . . . 45
8.7 Implémentation Python . . . 45
8.8 Exercices . . . 47

9 Régression Linéaire Multiple et ANOVA **49**
9.1 Exemple motivant . . . 49
9.2 Le modèle matriciel . . . 49
9.3 Estimation MCO . . . 49
9.4 Inférence . . . 50
9.4.1 Test individuel . . . 50
9.4.2 Test global de Fisher . . . 50
9.4.3 Test de sous-modèle . . . 50
9.5 ANOVA à un facteur . . . 50
9.5.1 Comparaisons multiples . . . 51
9.6 Implémentation Python . . . 51
9.7 Exercices . . . 52

10 Introduction à la Statistique Bayésienne **55**
10.1 Exemple motivant . . . 55
10.2 Le paradigme bayésien . . . 55
10.3 Choix de la loi a priori . . . 56
10.4 Exemples fondamentaux . . . 56
10.5 Estimateurs bayésiens . . . 56
10.6 Implémentation Python . . . 57
10.7 Exercices . . . 58

11 Méthodes Non-Paramétriques **59**
11.1 Exemple motivant . . . 59
11.2 Pourquoi le non paramétrique ? . . . 59
11.3 Test des signes . . . 59
11.4 Test de Wilcoxon pour un échantillon (rangs signés) . . . 59
11.5 Test de Mann–Whitney / Wilcoxon pour deux échantillons . . . 60
11.6 Test de Kruskal–Wallis . . . 60
11.7 Test de Kolmogorov–Smirnov . . . 60
11.8 Test du χ^2 d'adéquation . . . 61
11.9 Estimation non paramétrique de la densité . . . 61

11.10 Corrélation de Spearman 61
11.11 Implémentation Python 61
11.12 Exercices 63

Formulaire récapitulatif **65**
.1 Lois de probabilité usuelles 65
.2 Formules d'estimation 65
.3 Intervalles de confiance 65
.4 Statistiques de test 66

Tables statistiques **67**

Préface

La statistique mathématique est la science qui permet de tirer des conclusions rigoureuses à partir de données incertaines. Partout — en médecine, en physique, en économie, en intelligence artificielle — les décisions reposent sur l'analyse de données. Ce cours fournit les fondements théoriques et pratiques de cette discipline.

Prérequis. Théorie des probabilités (variables aléatoires, lois classiques, convergences), analyse réelle I (séries, intégrales), algèbre linéaire (espaces vectoriels, matrices).

Organisation. Chaque chapitre s'ouvre sur un exemple motivant, développe la théorie avec rigueur, propose des implémentations Python et se conclut par des exercices gradués :

⋆ Exercices de calcul et programmation

⋆⋆ Exercices théoriques

⋆⋆⋆ Projets

Références principales.

- Saporta, G. — *Probabilités, Analyse des Données et Statistique*, Technip.
- Casella, G. & Berger, R.L. — *Statistical Inference*, 2e éd., Cengage.
- Wasserman, L. — *All of Statistics*, Springer.
- Wackerly, Mendenhall & Scheaffer — *Mathematical Statistics with Applications*.

Chapitre 1

Statistique Descriptive — Résumé et Visualisation

« On ne peut améliorer que ce que l'on mesure. » — Lord Kelvin

1.1 Exemple motivant

Une entreprise possède les salaires mensuels (en euros) de ses 20 employés :

1800, 1950, 2100, 2100, 2200, 2300, 2350, 2400, 2500, 2600,

2650, 2700, 2800, 2900, 3100, 3200, 3500, 3800, 4500, 8000.

Questions naturelles : quel est le salaire « typique » ? Les salaires sont-ils dispersés ? Y a-t-il des valeurs aberrantes ? Pour y répondre, il faut *résumer* et *visualiser* les données.

1.2 Population, échantillon, variables

Définition 1.1 (Population et échantillon)**.** La **population** est l'ensemble complet des individus étudiés. Un **échantillon** de taille n est un sous-ensemble $(x_1, \ldots, x_n)$ de la population.

Définition 1.2 (Types de variables)**.**

- **Variable qualitative** (catégorielle) : valeurs dans un ensemble fini de catégories.
  - *Nominale* : pas d'ordre naturel (couleur, sexe).
  - *Ordinale* : ordre naturel (mention au bac : passable $<$ AB $<$ B $<$ TB).
- **Variable quantitative** : valeurs numériques.
  - *Discrète* : valeurs dans un ensemble dénombrable ($\mathbb{N}$, nombre d'enfants).
  - *Continue* : valeurs dans un intervalle de $\mathbb{R}$ (taille, poids).

1.3 Mesures de tendance centrale

Définition 1.3 (Moyenne arithmétique)**.** Pour un échantillon $(x_1, \ldots, x_n)$, la **moyenne** est :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Définition 1.4 (Médiane)**.** La **médiane** Me est la valeur qui sépare l'échantillon ordonné en deux moitiés égales. Si $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ désigne l'échantillon ordonné :

$$\text{Me} = \begin{cases} x_{((n+1)/2)} & \text{si } n \text{ est impair,} \\ \frac{1}{2}\left(x_{(n/2)} + x_{(n/2+1)}\right) & \text{si } n \text{ est pair.} \end{cases}$$

Définition 1.5 (Mode)**.** Le **mode** est la valeur la plus fréquente de l'échantillon.

Intuition statistique

La moyenne est sensible aux valeurs extrêmes ; la médiane est **robuste**. Dans l'exemple des salaires, $\bar{x} = 3007{,}50$ € est tirée vers le haut par le salaire de 8000 €, tandis que Me = 2625 € reflète mieux le salaire « typique ».

1.4 Mesures de dispersion

Définition 1.6 (Variance et écart-type)**.** La **variance empirique** (corrigée) et l'**écart-type** sont :

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad s = \sqrt{s^2}.$$

La version non corrigée utilise n au dénominateur. Le facteur $n-1$ (correction de Bessel) garantit que s^2 est un estimateur sans biais de la variance de la population (cf. chapitre 4).

Définition 1.7 (Étendue et écart interquartile)**.**

$$\text{Étendue} = x_{(n)} - x_{(1)}, \tag{1.1}$$

$$\text{IQR} = Q_3 - Q_1, \tag{1.2}$$

où Q_1 et Q_3 sont les premier et troisième quartiles (médianes des moitiés inférieure et supérieure).

Définition 1.8 (Coefficient de variation)**.** Le **coefficient de variation** est le rapport $\text{CV} = s/\bar{x}$ (pour $\bar{x} \neq 0$). Il permet de comparer la dispersion de variables d'unités différentes.

1.5 Mesures de forme

Définition 1.9 (Coefficient d'asymétrie (skewness))**.**

$$\gamma_1 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3.$$

$\gamma_1 > 0$: queue à droite (asymétrie positive). $\gamma_1 < 0$: queue à gauche. $\gamma_1 = 0$: distribution symétrique.

Définition 1.10 (Coefficient d'aplatissement (kurtosis))**.**

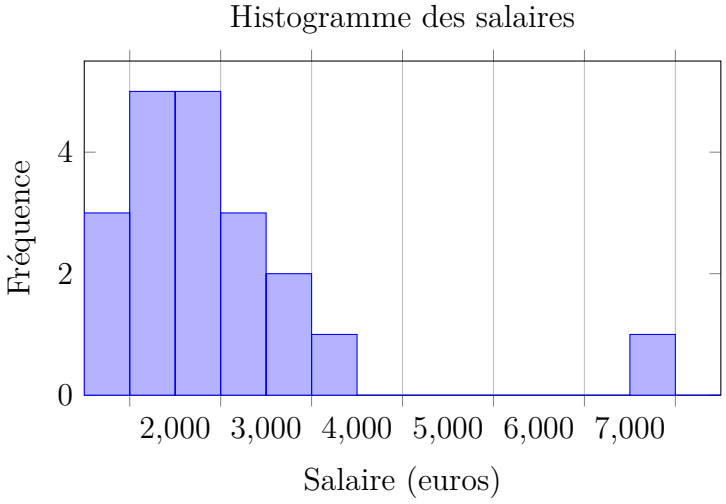
$$\gamma_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3.$$

$\gamma_2 = 0$ pour une distribution normale (mesosurtique). $\gamma_2 > 0$: queues lourdes (leptokurtique). $\gamma_2 < 0$: queues légères (platykurtique).

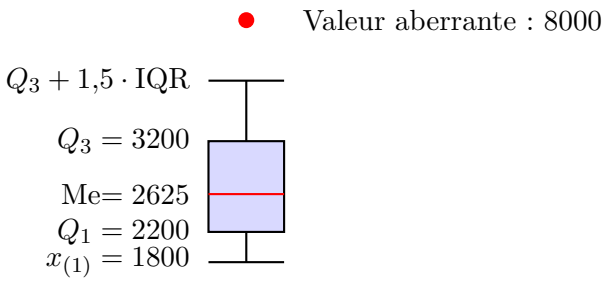
1.6 Représentations graphiques

1.6.1 Histogramme

L'histogramme représente la distribution d'une variable quantitative continue en regroupant les valeurs en classes.



1.6.2 Boîte à moustaches (boxplot)



1.6.3 Diagramme en bâtons et diagramme circulaire

Le diagramme en bâtons convient aux variables discrètes ; le diagramme circulaire (camembert) aux variables qualitatives nominales avec peu de modalités.

1.6.4 Fonction de répartition empirique

Définition 1.11. La **fonction de répartition empirique** est :

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{x_i \le x}.$$

C'est une fonction en escalier qui approche la vraie fonction de répartition F (théorème de Glivenko–Cantelli, cf. chapitre 2).

1.7 Données bivariées

Définition 1.12 (Covariance et corrélation)**.** Pour deux variables $(x_1, y_1), \ldots, (x_n, y_n)$:

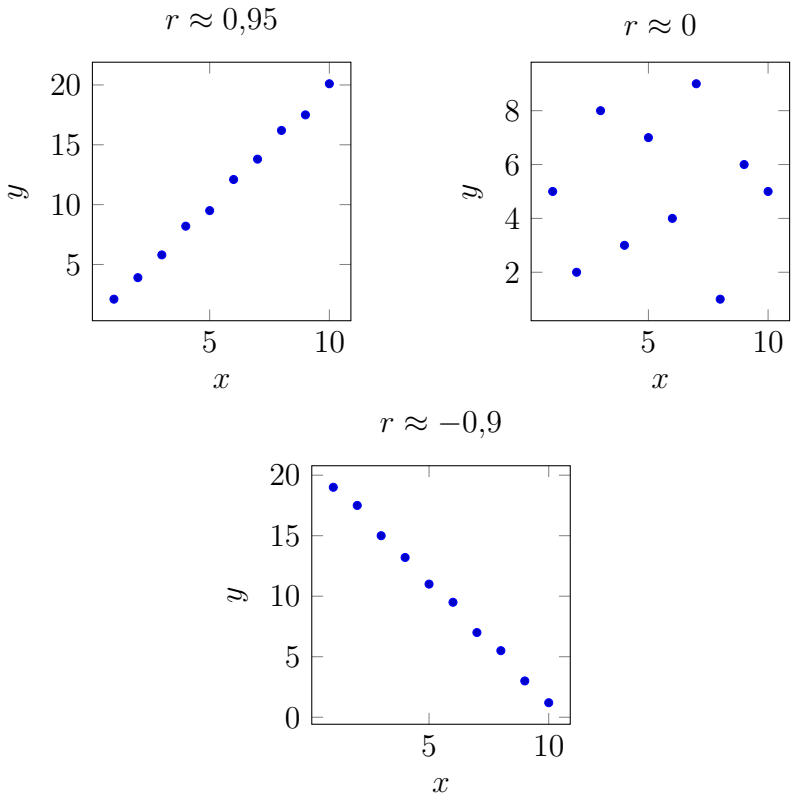
$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \tag{1.3}$$

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} \in [-1, 1]. \tag{1.4}$$

r_{xy} mesure la *liaison linéaire* entre x et y.

Attention — Piège classique

Corrélation n'implique pas causalité. Deux variables peuvent être fortement corrélées sans qu'il y ait de lien causal (variable confondante, coïncidence).



1.8 Implémentation Python

Implémentation Python

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats

# Donnees
salaires = np.array([1800, 1950, 2100, 2100, 2200, 2300, 2350, 2400,
                     2500, 2600, 2650, 2700, 2800, 2900, 3100, 3200,
                     3500, 3800, 4500, 8000])

# Mesures de tendance centrale
print(f"Moyenne    : {np.mean(salaires):.2f}")
print(f"Mediane    : {np.median(salaires):.2f}")
print(f"Mode       : {stats.mode(salaires, keepdims=True).mode[0]}")

# Mesures de dispersion
print(f"Ecart-type : {np.std(salaires, ddof=1):.2f}")
print(f"Variance   : {np.var(salaires, ddof=1):.2f}")
print(f"IQR        : {stats.iqr(salaires):.2f}")
print(f"CV         : {np.std(salaires, ddof=1)/np.mean(salaires):.4f}")

# Mesures de forme
print(f"Skewness   : {stats.skew(salaires):.4f}")
print(f"Kurtosis   : {stats.kurtosis(salaires):.4f}")  # excess kurtosis

# Visualisations
fig, axes = plt.subplots(1, 3, figsize=(14, 4))
axes[0].hist(salaires, bins=8, edgecolor='black', alpha=0.7)
axes[0].set_title("Histogramme")
axes[0].set_xlabel("Salaire (EUR)")

axes[1].boxplot(salaires, vert=True)
axes[1].set_title("Boite a moustaches")

# Fonction de repartition empirique
x_sorted = np.sort(salaires)
y_ecdf = np.arange(1, len(x_sorted)+1) / len(x_sorted)
axes[2].step(x_sorted, y_ecdf, where='post')
axes[2].set_title("ECDF")
axes[2].set_xlabel("Salaire (EUR)")
axes[2].set_ylabel("F_n(x)")

plt.tight_layout()
plt.savefig("ch01_descriptive.pdf")
plt.show()
```

1.9 Exercices

Exercice 1.1 ($\star$ – Calculs de base)**.** On mesure les températures (en °C) sur 10 jours : $12, 14, 15, 13, 16, 18, 14, 15, 17, 16$.

1. Calculer la moyenne, la médiane et le mode.
2. Calculer la variance (corrigée), l'écart-type et l'IQR.
3. Tracer l'histogramme et la boîte à moustaches avec Python.

Exercice 1.2 ($\star$ – Données bivariées)**.** Un cours de statistique a les notes suivantes pour le partiel (P) et l'examen final (E) de 8 étudiants :

Étudiant	1	2	3	4	5	6	7	8
P	8	12	14	10	16	6	18	11
E	9	13	15	12	17	8	19	13

Calculer r_{PE} et interpréter.

Exercice 1.3 ($\star\star$ – Propriétés de la moyenne)**.**

1. Montrer que $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.
2. Montrer que $\bar{x}$ minimise $\sum_{i=1}^{n}(x_i - c)^2$ sur $c \in \mathbb{R}$.
3. En déduire que $s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$.

Exercice 1.4 ($\star\star$ – Inégalité de Tchebychev empirique)**.** Montrer que pour tout $k > 0$, la proportion d'observations x_i vérifiant $|x_i - \bar{x}| > ks$ est au plus $1/k^2$.

Exercice 1.5 ($\star\star\star$ – Projet : analyse exploratoire)**.** Télécharger un jeu de données réel (par ex. `tips` de `seaborn` ou un fichier CSV de l'INSEE). Produire un rapport complet : statistiques descriptives, visualisations (histogrammes, boxplots, scatter plots), identifier les valeurs aberrantes, commenter la forme de la distribution.

Formules clés

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\mathrm{IQR} = Q_3 - Q_1 \qquad r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$\gamma_1 = \frac{1}{n}\sum\left(\frac{x_i - \bar{x}}{s}\right)^3 \qquad \gamma_2 = \frac{1}{n}\sum\left(\frac{x_i - \bar{x}}{s}\right)^4 - 3$$

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{x_i \le x} \qquad \mathrm{CV} = s/\bar{x}$$

Chapitre 2

Modèle Statistique et Échantillonnage

« Tous les modèles sont faux, certains sont utiles. » — George Box

2.1 Exemple motivant

Un laboratoire pharmaceutique teste un nouveau médicament. Sur $n = 100$ patients, 68 guérissent. Le médicament est-il efficace ? Pour répondre, il faut un *modèle* : on suppose que chaque patient guérit indépendamment avec une probabilité p inconnue. Les 100 résultats forment un **échantillon aléatoire** de loi $\mathcal{B}(1, p)$. Toute la statistique inférentielle repose sur cette formalisation.

2.2 Modèle statistique

Définition 2.1 (Modèle statistique)**.** Un **modèle statistique** est un triplet $(\mathcal{X}, \mathcal{A}, \mathcal{P})$ où :

- $\mathcal{X}$ est l'**espace des observations** (espace d'échantillonnage),
- $\mathcal{A}$ est une tribu sur $\mathcal{X}$,
- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ est une famille de lois de probabilité indexée par le **paramètre** $\theta \in \Theta$.

Définition 2.2 (Modèle paramétrique vs. non paramétrique)**.**

- **Paramétrique** : $\Theta \subseteq \mathbb{R}^d$ pour un d fini. Exemple : $\mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$.
- **Non paramétrique** : Θ est de dimension infinie. Exemple : $\mathcal{P}$ = ensemble de toutes les lois continues sur $\mathbb{R}$.
- **Semi-paramétrique** : composante paramétrique + composante non paramétrique.

Définition 2.3 (Identifiabilité)**.** Le modèle est **identifiable** si l'application $\theta \mapsto P_\theta$ est injective : $P_{\theta_1} = P_{\theta_2} \implies \theta_1 = \theta_2$.

2.3 Échantillon aléatoire

Définition 2.4 (Échantillon aléatoire)**.** Un **échantillon aléatoire** de taille n issu de la loi P_θ est un n-uplet $(X_1, \dots, X_n)$ de variables aléatoires **indépendantes et identiquement distribuées** (i.i.d.) de loi P_θ.

On note $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} P_\theta$.

La **réalisation** $(x_1, \dots, x_n)$ est l'échantillon observé.

Remarque 2.5. La distinction entre X_i (variable aléatoire) et x_i (valeur observée) est fondamentale. Une **statistique** est une fonction des X_i ; son évaluation en $(x_1, \dots, x_n)$ donne une valeur numérique.

2.4 Statistiques et distributions d'échantillonnage

Définition 2.6 (Statistique)**.** Une **statistique** est une fonction mesurable $T = T(X_1, \dots, X_n)$ de l'échantillon qui ne dépend *pas* du paramètre θ.

Définition 2.7 (Statistique exhaustive (suffisante))**.** T est **exhaustive** (ou suffisante) pour θ si la loi conditionnelle de $(X_1, \dots, X_n)$ sachant T ne dépend pas de θ.

Théorème 2.8 (Critère de factorisation de Fisher–Neyman)**.** *$T(X_1, \dots, X_n)$ est exhaustive pour θ si et seulement s'il existe des fonctions g et h telles que :*

$$f(x_1, \dots, x_n; \theta) = g\big(T(x_1, \dots, x_n), \theta\big) \cdot h(x_1, \dots, x_n),$$

où f désigne la densité (ou la fonction de masse) jointe.

Démonstration. (Cas discret.) Si T est exhaustive, $f(\mathbf{x}; \theta) = P_\theta(T = T(\mathbf{x})) \cdot P(X = \mathbf{x} \mid T = T(\mathbf{x}))$. On pose $g(t, \theta) = P_\theta(T = t)$ et $h(\mathbf{x}) = P(X = \mathbf{x} \mid T = T(\mathbf{x}))$.

Réciproquement, si la factorisation tient, alors :

$$P_\theta(\mathbf{X} = \mathbf{x} \mid T = t) = \frac{g(t, \theta) \cdot h(\mathbf{x})}{\sum_{\mathbf{y}: T(\mathbf{y}) = t} g(t, \theta) \cdot h(\mathbf{y})} = \frac{h(\mathbf{x})}{\sum_{\mathbf{y}: T(\mathbf{y}) = t} h(\mathbf{y})},$$

qui ne dépend pas de θ. □

Exemple 2.9. Soit $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}(1, p)$. La vraisemblance est :

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n - \sum x_i}.$$

On identifie $g(t, p) = p^t(1-p)^{n-t}$ avec $t = \sum x_i$ et $h \equiv 1$. Donc $T = \sum X_i$ est exhaustive pour p.

2.5 Loi des grandes statistiques d'échantillonnage

2.5.1 Moyenne empirique

Théorème 2.10 (Espérance et variance de $\bar{X}$)**.** *Si $X_1, \dots, X_n \overset{i.i.d.}{\sim}$ avec $\mathbb{E}[X_i] = \mu$ et $\mathrm{Var}(X_i) = \sigma^2$, alors :*

$$\mathbb{E}[\bar{X}] = \mu, \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

Démonstration. $\mathbb{E}[\bar{X}] = \frac{1}{n}\sum \mathbb{E}[X_i] = \mu$. Par indépendance, $\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\sum \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$. □

2.5.2 Loi des grands nombres

Théorème 2.11 (Loi faible des grands nombres). *Si $X_1, \dots, X_n$ sont i.i.d. avec $\mathbb{E}[X_i] = \mu$ et $\mathrm{Var}(X_i) < \infty$, alors :*

$$\bar{X} \xrightarrow{\mathbb{P}} \mu \quad (n \to \infty).$$

Théorème 2.12 (Loi forte des grands nombres). *Sous les mêmes hypothèses (ou seulement $\mathbb{E}[|X_1|] < \infty$) :*

$$\bar{X} \xrightarrow{p.s.} \mu \quad (n \to \infty).$$

2.5.3 Théorème central limite

Théorème 2.13 (Théorème central limite (TCL)). *Si $X_1, \dots, X_n$ sont i.i.d. avec $\mathbb{E}[X_i] = \mu$ et $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$, alors :*

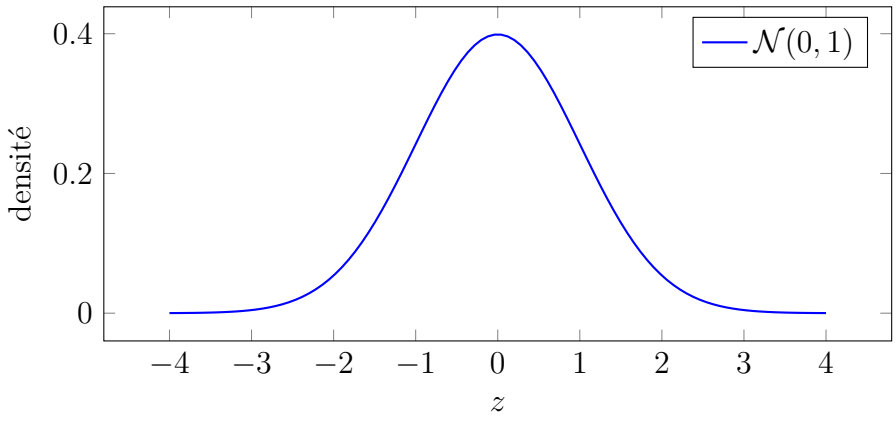
$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Autrement dit, $\bar{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ pour n grand.

Intuition statistique

Le TCL explique pourquoi la loi normale est omniprésente : dès qu'une quantité résulte de la **somme de nombreux petits effets indépendants**, elle est approximativement gaussienne.

Illustration du TCL : convergence vers $\mathcal{N}(0, 1)$



2.6 Distribution de la variance empirique

Théorème 2.14 (Loi de S^2 sous normalité). *Si $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, alors :*

1. *$\bar{X}$ et S^2 sont **indépendants** (théorème de Cochran),*
2. $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

Corollaire 2.15. *Sous les hypothèses du théorème 2.14 :*

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Démonstration. On écrit $\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}} = \frac{Z}{\sqrt{V/(n-1)}}$ où $Z \sim \mathcal{N}(0,1)$ et $V = (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ sont indépendants, donc le rapport suit $t(n-1)$ par définition de la loi de Student. □

2.7 Théorème de Glivenko–Cantelli

Théorème 2.16 (Glivenko–Cantelli)**.** *Si $X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$, alors :*

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{p.s.} 0.$$

La fonction de répartition empirique converge uniformément vers la vraie fonction de répartition.

2.8 Implémentation Python

Implémentation Python

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

# --- Illustration du TCL ---
np.random.seed(42)
n_values = [1, 2, 5, 30]
fig, axes = plt.subplots(1, 4, figsize=(16, 3.5))

for ax, n in zip(axes, n_values):
    means = [np.mean(np.random.exponential(1, n)) for _ in range(10000)]
    ax.hist(means, bins=40, density=True, alpha=0.7, edgecolor='black')
    ax.set_title(f"n = {n}")
    if n >= 5:
        x = np.linspace(min(means), max(means), 100)
        ax.plot(x, stats.norm.pdf(x, 1, 1/np.sqrt(n)), 'r-', lw=2)
plt.suptitle("TCL : moyennes d'exponentielles", fontsize=13)
plt.tight_layout()
plt.savefig("ch02_tcl.pdf")
plt.show()

# --- Glivenko-Cantelli ---
np.random.seed(0)
X = np.random.normal(0, 1, 50)
x_sorted = np.sort(X)
ecdf = np.arange(1, len(X)+1) / len(X)
```

```
x_grid = np.linspace(-3, 3, 200)

plt.figure(figsize=(8, 5))
plt.step(x_sorted, ecdf, where='post', label=r'$\hat{F}_{50}(x)$', lw=2)
plt.plot(x_grid, stats.norm.cdf(x_grid), 'r-', label=r'$\Phi(x)$', lw=2)
plt.legend()
plt.title("Glivenko-Cantelli : ECDF vs CDF normale")
plt.xlabel("x")
plt.ylabel("F(x)")
plt.savefig("ch02_glivenko.pdf")
plt.show()

# --- Statistique exhaustive : Bernoulli ---
n, p_true = 100, 0.68
sample = np.random.binomial(1, p_true, n)
T = sample.sum()
print(f"Statistique exhaustive T = sum(Xi) = {T}")
print(f"Estimation naturelle p_hat = T/n = {T/n:.2f}")
```

2.9 Exercices

Exercice 2.1 ($\star$ – Échantillonnage)**.** On prélève un échantillon de taille $n = 25$ d'une loi $\mathcal{N}(10, 4)$. Calculer $\mathbb{E}[\bar{X}]$, $\mathrm{Var}(\bar{X})$ et $\mathbb{P}(|\bar{X} - 10| > 0.5)$.

Exercice 2.2 ($\star$ – TCL en pratique)**.** Générer 10 000 moyennes d'échantillons de taille $n = 30$ d'une loi $\mathcal{U}([0, 1])$. Vérifier graphiquement que la distribution est approximativement gaussienne. Calculer la moyenne et la variance empiriques et comparer aux valeurs théoriques.

Exercice 2.3 ($\star\star$ – Exhaustivité)**.**

1. Soit $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{P}(\lambda)$. Montrer que $T = \sum X_i$ est exhaustive pour λ.

2. Soit $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnus. Montrer que $T = (\sum X_i, \sum X_i^2)$ est exhaustive pour (μ, σ^2).

Exercice 2.4 ($\star\star$ – Preuve du TCL par fonctions caractéristiques)**.** Soit $X_1, \ldots, X_n$ i.i.d. avec $\mathbb{E}[X_i] = 0$, $\mathrm{Var}(X_i) = 1$. On note φ la fonction caractéristique de X_1.

1. Montrer que $\varphi(t) = 1 - \frac{t^2}{2} + o(t^2)$ au voisinage de 0.

2. Calculer la fonction caractéristique de $\bar{X}_n\sqrt{n}$.

3. En déduire la convergence vers $\mathcal{N}(0, 1)$.

Exercice 2.5 ($\star\star\star$ – Projet : simulations Monte Carlo)**.** Étudier la vitesse de convergence du TCL pour différentes lois mères (exponentielle, Bernoulli, uniforme, Cauchy). Pour chaque loi, estimer la taille n minimale pour laquelle l'approximation normale est acceptable (test de Kolmogorov–Smirnov). Que se passe-t-il pour la loi de Cauchy ?

Formules clés

$$\mathbb{E}[\bar{X}] = \mu \qquad \mathrm{Var}(\bar{X}) = \sigma^2/n$$

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1) \qquad \sup_x |\hat{F}_n(x) - F(x)| \xrightarrow{\text{p.s.}} 0$$

Fisher–Neyman : $f(\mathbf{x};\theta) = g(T(\mathbf{x}),\theta)\cdot h(\mathbf{x}) \iff T$ exhaustive.

Chapitre 3

Estimation Ponctuelle — EMV et Méthode des Moments

« L'art de la statistique est de choisir le bon résumé numérique. »

3.1 Exemple motivant

On mesure la durée de vie (en heures) de 10 ampoules LED :

1200, 1350, 980, 1500, 1100, 1450, 1300, 1050, 1400, 1250.

On modélise par une loi exponentielle $\mathcal{E}(\lambda)$. Comment estimer λ ? Deux méthodes classiques s'offrent à nous : la **méthode des moments** et le **maximum de vraisemblance**.

3.2 Cadre général

Définition 3.1 (Estimateur)**.** Soit $(X_1, \dots, X_n) \overset{\text{i.i.d.}}{\sim} P_\theta$. Un **estimateur** de θ (ou d'une fonction $g(\theta)$) est une statistique $\hat{\theta} = T(X_1, \dots, X_n)$ à valeurs dans Θ.

Définition 3.2 (Estimation)**.** L'**estimation** $\hat{\theta} = T(x_1, \dots, x_n)$ est la valeur numérique obtenue en remplaçant les v.a. par les observations.

3.3 Méthode des moments

Définition 3.3 (Méthode des moments (MoM))**.** On égale les k premiers moments théoriques aux moments empiriques :

$$\mathbb{E}_\theta[X^j] = \frac{1}{n}\sum_{i=1}^{n} X_i^j, \quad j = 1, \dots, k,$$

où $k = \dim(\theta)$, puis on résout le système en θ.

Exemple 3.4 (Loi normale)**.** Soit $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$. On a $\mathbb{E}[X] = \mu$ et $\mathbb{E}[X^2] = \mu^2 + \sigma^2$. La MoM donne :

$$\hat{\mu}_{\text{MoM}} = \bar{X}, \qquad \hat{\sigma}^2_{\text{MoM}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Exemple 3.5 (Loi exponentielle). $X \sim \mathcal{E}(\lambda)$: $\mathbb{E}[X] = 1/\lambda$. Donc $\hat{\lambda}_{\text{MoM}} = 1/\bar{X}$.

Exemple 3.6 (Loi Gamma). $X \sim \Gamma(\alpha, \beta)$ avec $\mathbb{E}[X] = \alpha/\beta$ et $\text{Var}(X) = \alpha/\beta^2$. On obtient :

$$\hat{\beta}_{\text{MoM}} = \frac{\bar{X}}{S^2}, \qquad \hat{\alpha}_{\text{MoM}} = \frac{\bar{X}^2}{S^2}.$$

Intuition statistique

La MoM est simple et donne souvent de bons estimateurs, mais elle n'utilise pas toute l'information contenue dans l'échantillon. Le maximum de vraisemblance est généralement préférable.

3.4 Maximum de vraisemblance

Définition 3.7 (Fonction de vraisemblance). La **vraisemblance** de l'échantillon $(x_1, \ldots, x_n)$ est :

$$L(\theta) = L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta),$$

où $f(\cdot; \theta)$ est la densité (cas continu) ou la fonction de masse (cas discret).

La **log-vraisemblance** est $\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta)$.

Définition 3.8 (Estimateur du maximum de vraisemblance (EMV)). L'**EMV** est :

$$\hat{\theta}_{\text{EMV}} = \arg\max_{\theta \in \Theta} L(\theta) = \arg\max_{\theta \in \Theta} \ell(\theta).$$

Remarque 3.9. En pratique, on résout souvent l'**équation de vraisemblance** :

$$\frac{\partial \ell}{\partial \theta}(\theta) = 0,$$

en vérifiant que la solution est bien un maximum (condition d'ordre 2 ou argument de convexité).

3.4.1 Exemples fondamentaux

Exemple 3.10 (Loi de Bernoulli). $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}(1, p)$. On a $\ell(p) = \sum x_i \ln p + (n - \sum x_i) \ln(1 - p)$. L'équation $\ell'(p) = 0$ donne :

$$\frac{\sum x_i}{p} - \frac{n - \sum x_i}{1 - p} = 0 \implies \hat{p}_{\text{EMV}} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}.$$

Exemple 3.11 (Loi normale). $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Les équations $\partial \ell / \partial \mu = 0$ et $\partial \ell / \partial \sigma^2 = 0$ donnent :

$$\hat{\mu}_{\text{EMV}} = \bar{X}, \qquad \hat{\sigma}^2_{\text{EMV}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Attention — Piège classique

L'EMV de σ^2 sous le modèle normal est $\frac{1}{n}\sum(X_i-\bar{X})^2$ (diviseur n), qui est **biaisé**. La version corrigée S^2 (diviseur $n-1$) est sans biais (cf. chapitre 4).

Exemple 3.12 (Loi exponentielle)**.** $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathcal{E}(\lambda)$. On a $\ell(\lambda) = n\ln\lambda - \lambda\sum x_i$. L'équation $\ell'(\lambda)=0$ donne $\hat{\lambda}_{\text{EMV}} = n/\sum X_i = 1/\bar{X}$.

Exemple 3.13 (Loi de Poisson)**.** $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathcal{P}(\lambda)$. On a $\ell(\lambda) = \sum x_i \ln\lambda - n\lambda - \sum\ln(x_i!)$. L'équation $\ell'(\lambda)=0$ donne $\hat{\lambda}_{\text{EMV}} = \bar{X}$.

Exemple 3.14 (Loi uniforme)**.** $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0,\theta])$. La vraisemblance est $L(\theta) = \theta^{-n}\mathbf{1}_{\theta\geq x_{(n)}}$, décroissante en θ pour $\theta \geq x_{(n)}$. Donc $\hat{\theta}_{\text{EMV}} = X_{(n)} = \max_i X_i$.

Ici, l'équation de vraisemblance ne s'applique pas (le maximum est atteint au bord).

3.5 Propriétés de l'EMV

Théorème 3.15 (Propriété d'invariance)**.** *Si $\hat{\theta}$ est l'EMV de θ, alors pour toute fonction g, l'EMV de $g(\theta)$ est $g(\hat{\theta})$.*

Théorème 3.16 (Consistance de l'EMV)**.** *Sous des conditions de régularité (famille exponentielle, etc.), l'EMV est* ***convergent*** *:*

$$\hat{\theta}_{\text{EMV}} \xrightarrow{\mathbb{P}} \theta_0 \quad (n\to\infty).$$

Théorème 3.17 (Normalité asymptotique de l'EMV)**.** *Sous des conditions de régularité :*

$$\sqrt{n}(\hat{\theta}_{\text{EMV}} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{1}{I(\theta_0)}\right),$$

*où $I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right]$ est l'****information de Fisher*** *(cf. chapitre 4).*

3.6 Information de Fisher

Définition 3.18 (Information de Fisher)**.** Pour un modèle régulier, l'**information de Fisher** d'une observation est :

$$I(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right] = -\mathbb{E}_\theta\left[\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right].$$

Pour un échantillon de taille n : $I_n(\theta) = nI(\theta)$.

Exemple 3.19. $X\sim\mathcal{N}(\mu,\sigma^2)$ avec σ^2 connu. $\ln f = -\frac{(x-\mu)^2}{2\sigma^2} + \text{cst}$, donc $\frac{\partial^2 \ln f}{\partial\mu^2} = -\frac{1}{\sigma^2}$ et $I(\mu) = \frac{1}{\sigma^2}$.

Exemple 3.20. $X\sim\mathcal{E}(\lambda)$. $\ln f = \ln\lambda - \lambda x$, $\frac{\partial^2 \ln f}{\partial\lambda^2} = -\frac{1}{\lambda^2}$, donc $I(\lambda) = \frac{1}{\lambda^2}$.

3.7 Implémentation Python

Implémentation Python

```python
import numpy as np
from scipy import stats
from scipy.optimize import minimize_scalar
import matplotlib.pyplot as plt

# --- Donnees : duree de vie d'ampoules ---
data = np.array([1200, 1350, 980, 1500, 1100, 1450, 1300, 1050, 1400,
↪  1250])

# --- Methode des moments : Exp(lambda) ---
lambda_mom = 1 / np.mean(data)
print(f"MoM: lambda_hat = {lambda_mom:.6f}")

# --- EMV : Exp(lambda) ---
lambda_mle = 1 / np.mean(data)  # coincide avec MoM pour l'exponentielle
print(f"EMV: lambda_hat = {lambda_mle:.6f}")

# --- EMV pour la loi normale ---
mu_mle = np.mean(data)
sigma2_mle = np.var(data, ddof=0)  # diviseur n (EMV)
sigma2_unbiased = np.var(data, ddof=1)  # diviseur n-1 (sans biais)
print(f"EMV Normale: mu = {mu_mle:.2f}, sigma^2 = {sigma2_mle:.2f}")
print(f"Sans biais:  sigma^2 = {sigma2_unbiased:.2f}")

# --- EMV avec scipy (methode generale) ---
# Ajustement d'une loi exponentielle
loc, scale = stats.expon.fit(data, floc=0)  # floc=0 fixe le decalage
lambda_scipy = 1 / scale
print(f"scipy EMV: lambda = {lambda_scipy:.6f}")

# --- Visualisation de la log-vraisemblance ---
lambdas = np.linspace(0.0005, 0.002, 200)
log_lik = np.array([np.sum(stats.expon.logpdf(data, scale=1/lam)) for lam
↪  in lambdas])

plt.figure(figsize=(8, 5))
plt.plot(lambdas, log_lik, 'b-', lw=2)
plt.axvline(lambda_mle, color='r', linestyle='--', label=f'EMV =
↪  {lambda_mle:.5f}')
plt.xlabel(r'$\lambda$')
plt.ylabel(r'$\ell(\lambda)$')
plt.title('Log-vraisemblance exponentielle')
plt.legend()
plt.savefig("ch03_loglik.pdf")
plt.show()

# --- EMV loi uniforme U([0, theta]) ---
data_unif = np.array([0.3, 0.7, 0.5, 0.9, 0.2, 0.8, 0.6, 0.4, 0.95, 0.1])
theta_mle = np.max(data_unif)
```

```
print(f"EMV U([0,theta]): theta_hat = {theta_mle}")

# --- Information de Fisher ---
# Pour Exp(lambda) : I(lambda) = 1/lambda^2
n = len(data)
I_fisher = 1 / lambda_mle**2
var_asymp = 1 / (n * I_fisher)
print(f"Variance asymptotique de l'EMV: {var_asymp:.8f}")
print(f"IC asymptotique 95%: [{lambda_mle - 1.96*np.sqrt(var_asymp):.5f},
↪  "
      f"{lambda_mle + 1.96*np.sqrt(var_asymp):.5f}]")
```

3.8 Exercices

Exercice 3.1 ($\star$ – Calcul d'EMV)**.** Calculer l'EMV pour chacun des modèles suivants :

1. $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{G}omtrique(p)$ (loi géométrique).
2. $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([a, b])$ avec a et b inconnus.

Exercice 3.2 ($\star$ – MoM)**.** Soit $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, \beta)$. On rappelle que $\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta}$ et $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$. Trouver les estimateurs des moments pour α et β.

Exercice 3.3 ($\star\star$ – Invariance)**.** $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$. En utilisant l'invariance de l'EMV, déterminer l'EMV de :

1. σ (l'écart-type),
2. $\mathbb{P}(X > c)$ pour c fixé,
3. le coefficient de variation σ/μ.

Exercice 3.4 ($\star\star$ – Information de Fisher)**.**

1. Calculer $I(\theta)$ pour $X \sim \mathcal{B}(1, p)$.
2. Calculer $I(\theta)$ pour $X \sim \mathcal{P}(\lambda)$.
3. Calculer la matrice d'information de Fisher $I(\mu, \sigma^2)$ pour $X \sim \mathcal{N}(\mu, \sigma^2)$.

Exercice 3.5 ($\star\star$ – EMV non régulier)**.** Soit $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{U}([0, \theta])$.

1. Montrer que l'EMV est $\hat{\theta} = X_{(n)}$.
2. Trouver la loi de $X_{(n)}$ et calculer $\mathbb{E}[X_{(n)}]$. L'EMV est-il sans biais ?
3. Proposer un estimateur sans biais basé sur $X_{(n)}$.

Exercice 3.6 ($\star\star\star$ – Projet : comparaison MoM vs EMV)**.** Pour la loi Gamma(α, β) :

1. Calculer les estimateurs MoM et EMV (numériquement pour l'EMV).
2. Par simulation (1000 répétitions, $n = 20, 50, 200$), comparer le biais et l'EQM des deux méthodes.
3. Tracer les distributions d'échantillonnage des estimateurs et conclure.

Formules clés

$$\text{MoM} : \mathbb{E}_\theta[X^j] = \frac{1}{n}\sum X_i^j,\ j = 1, \dots, \dim(\theta)$$

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

$$\hat{\theta}_{\text{EMV}} = \arg\max_\theta \ell(\theta)$$

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ln f}{\partial \theta^2}\right] = \mathbb{E}\left[\left(\frac{\partial \ln f}{\partial \theta}\right)^2\right]$$

$$\sqrt{n}(\hat{\theta}_{\text{EMV}} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/I(\theta))$$

Chapitre 4

Propriétés des Estimateurs — Biais, Convergence, Efficacité

« Un bon estimateur doit viser juste et ne pas trop se disperser. »

4.1 Exemple motivant

Deux tireurs visent une cible. Le premier touche en moyenne le centre mais ses tirs sont dispersés. Le second groupe ses tirs mais décalés systématiquement à droite. Lequel est « meilleur » ? Cette question se pose exactement pour les estimateurs : il faut comprendre le **biais**, la **variance** et leur compromis.

4.2 Biais d'un estimateur

Définition 4.1 (Biais)**.** Le **biais** de l'estimateur $\hat{\theta}$ de θ est :

$$\text{Biais}(\hat{\theta}) = \mathbb{E}_\theta[\hat{\theta}] - \theta.$$

$\hat{\theta}$ est dit **sans biais** si $\text{Biais}(\hat{\theta}) = 0$ pour tout $\theta \in \Theta$, i.e. $\mathbb{E}_\theta[\hat{\theta}] = \theta$.

Exemple 4.2. $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim}$ avec $\mathbb{E}[X_i] = \mu$. Alors $\bar{X}$ est sans biais pour μ : $\mathbb{E}[\bar{X}] = \mu$.

Proposition 4.3. *$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ est un estimateur sans biais de σ^2.*

Démonstration.

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

Or $\mathbb{E}[X_i^2] = \text{Var}(X_i) + (\mathbb{E}[X_i])^2 = \sigma^2 + \mu^2$ et $\mathbb{E}[\bar{X}^2] = \text{Var}(\bar{X}) + \mu^2 = \frac{\sigma^2}{n} + \mu^2$. Donc :

$$\mathbb{E}\Big[\sum (X_i - \bar{X})^2\Big] = n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) = (n-1)\sigma^2.$$

D'où $\mathbb{E}[S^2] = \sigma^2$. □

Définition 4.4 (Biais asymptotique)**.** $\hat{\theta}_n$ est **asymptotiquement sans biais** si $\text{Biais}(\hat{\theta}_n) \to 0$ quand $n \to \infty$.

4.3 Erreur quadratique moyenne

Définition 4.5 (EQM / MSE)**.** L'**erreur quadratique moyenne** est :

$$\mathrm{EQM}(\hat{\theta}) = \mathbb{E}_\theta[(\hat{\theta} - \theta)^2].$$

Théorème 4.6 (Décomposition biais-variance)**.**

$$\mathrm{EQM}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \big[\mathrm{Biais}(\hat{\theta})\big]^2.$$

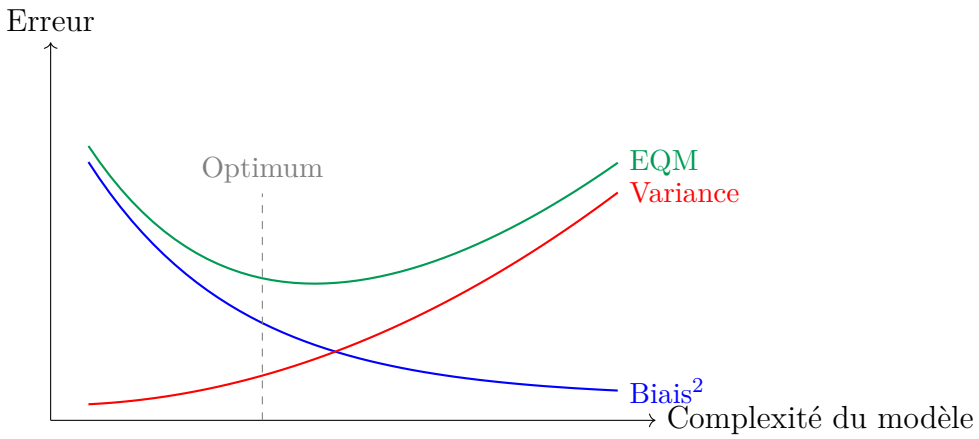
Démonstration. Posons $b = \mathbb{E}[\hat{\theta}] - \theta$. Alors :

$$\begin{aligned}\mathrm{EQM}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}] + b)^2] \\ &= \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] + 2b\,\mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]] + b^2 = \mathrm{Var}(\hat{\theta}) + b^2.\end{aligned}$$

□

Intuition statistique

L'EQM combine deux sources d'erreur. Un estimateur biaisé peut avoir un EQM plus faible qu'un estimateur sans biais si sa variance est suffisamment réduite. C'est le **compromis biais-variance**, fondamental en apprentissage statistique.



4.4 Convergence (consistance)

Définition 4.7 (Estimateur convergent (consistant))**.** $\hat{\theta}_n$ est **convergent** (ou consistant) pour θ si :

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta \quad (n \to \infty),$$

c'est-à-dire $\forall \varepsilon > 0$, $\mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) \to 0$.

Proposition 4.8 (Condition suffisante)**.** *Si* $\mathrm{Biais}(\hat{\theta}_n) \to 0$ *et* $\mathrm{Var}(\hat{\theta}_n) \to 0$ *quand* $n \to \infty$*, alors* $\hat{\theta}_n$ *est convergent.*

Démonstration. Par l'inégalité de Tchebychev :

$$\mathbb{P}(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{\mathrm{EQM}(\hat{\theta}_n)}{\varepsilon^2} = \frac{\mathrm{Var}(\hat{\theta}_n) + \mathrm{Biais}^2(\hat{\theta}_n)}{\varepsilon^2} \to 0.$$

□

Exemple 4.9. $\bar{X}$ est convergent pour μ : $\mathrm{Biais}(\bar{X}) = 0$ et $\mathrm{Var}(\bar{X}) = \sigma^2/n \to 0$.

4.5 Efficacité et borne de Cramér–Rao

Définition 4.10 (Score et conditions de régularité)**.** Le **score** est $S(\theta) = \frac{\partial}{\partial\theta} \ln f(X;\theta)$. Les conditions de régularité exigent que :

1. Le support de $f(\cdot;\theta)$ ne dépend pas de θ.
2. On peut dériver sous le signe intégrale.
3. $0 < I(\theta) < \infty$.

Théorème 4.11 (Inégalité de Cramér–Rao)**.** *Sous les conditions de régularité, pour tout estimateur sans biais $\hat{\theta}$ de $g(\theta)$:*

$$\mathrm{Var}_\theta(\hat{\theta}) \geq \frac{[g'(\theta)]^2}{nI(\theta)}.$$

En particulier, si $\hat{\theta}$ est sans biais pour θ ($g = \mathrm{id}$) :

$$\mathrm{Var}_\theta(\hat{\theta}) \geq \frac{1}{nI(\theta)}.$$

Démonstration. Par l'inégalité de Cauchy–Schwarz. Posons $U = \hat{\theta}$ et $V = \sum_{i=1}^n \frac{\partial}{\partial\theta} \ln f(X_i;\theta)$ (score total). On sait que $\mathbb{E}[V] = 0$ (dérivation sous l'intégrale) et $\mathrm{Var}(V) = nI(\theta)$.

Comme $\hat{\theta}$ est sans biais pour $g(\theta)$: $\mathbb{E}_\theta[\hat{\theta}] = g(\theta)$. En dérivant :

$$g'(\theta) = \frac{\partial}{\partial\theta} \int \hat{\theta} \prod f(x_i;\theta)\, d\mathbf{x} = \mathbb{E}_\theta[\hat{\theta} \cdot V] = \mathrm{Cov}(\hat{\theta}, V).$$

Par Cauchy–Schwarz : $[g'(\theta)]^2 = [\mathrm{Cov}(\hat{\theta}, V)]^2 \leq \mathrm{Var}(\hat{\theta}) \cdot \mathrm{Var}(V) = \mathrm{Var}(\hat{\theta}) \cdot nI(\theta)$. □

Définition 4.12 (Estimateur efficace)**.** $\hat{\theta}$ est **efficace** s'il atteint la borne de Cramér–Rao : $\mathrm{Var}(\hat{\theta}) = \frac{1}{nI(\theta)}$.

Exemple 4.13. Pour $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ (σ^2 connu) : $I(\mu) = 1/\sigma^2$ et $\mathrm{Var}(\bar{X}) = \sigma^2/n = 1/(nI(\mu))$. Donc $\bar{X}$ est **efficace**.

Définition 4.14 (Efficacité relative)**.** L'**efficacité relative** de $\hat{\theta}_1$ par rapport à $\hat{\theta}_2$ est :

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{\mathrm{Var}(\hat{\theta}_2)}{\mathrm{Var}(\hat{\theta}_1)}.$$

Si $e > 1$, $\hat{\theta}_1$ est plus efficace.

4.6 Estimateur UMVUE et théorème de Rao–Blackwell

Définition 4.15 (UMVUE)**.** Un estimateur sans biais $\hat{\theta}^*$ de $g(\theta)$ est **UMVUE** (Uniformly Minimum Variance Unbiased Estimator) si pour tout autre estimateur sans biais $\hat{\theta}$ de $g(\theta)$:

$$\mathrm{Var}_\theta(\hat{\theta}^*) \leq \mathrm{Var}_\theta(\hat{\theta}) \quad \forall\theta \in \Theta.$$

Théorème 4.16 (Rao–Blackwell). *Soit $\hat{\theta}$ un estimateur sans biais de $g(\theta)$ et T une statistique exhaustive. Alors $\hat{\theta}^* = \mathbb{E}[\hat{\theta} \mid T]$ est :*

1. *Sans biais pour $g(\theta)$,*
2. $\mathrm{Var}(\hat{\theta}^*) \leq \mathrm{Var}(\hat{\theta})$, *avec égalité ssi $\hat{\theta}$ est déjà fonction de T.*

Démonstration. (1) $\mathbb{E}[\hat{\theta}^*] = \mathbb{E}[\mathbb{E}[\hat{\theta} \mid T]] = \mathbb{E}[\hat{\theta}] = g(\theta)$ (tour de l'espérance).
(2) Par la formule de décomposition de la variance :

$$\mathrm{Var}(\hat{\theta}) = \mathbb{E}[\mathrm{Var}(\hat{\theta} \mid T)] + \mathrm{Var}(\mathbb{E}[\hat{\theta} \mid T]) = \mathbb{E}[\mathrm{Var}(\hat{\theta} \mid T)] + \mathrm{Var}(\hat{\theta}^*).$$

Comme $\mathbb{E}[\mathrm{Var}(\hat{\theta} \mid T)] \geq 0$, on a $\mathrm{Var}(\hat{\theta}) \geq \mathrm{Var}(\hat{\theta}^*)$. □

Théorème 4.17 (Lehmann–Scheffé). *Si T est exhaustive et **complète**, alors tout estimateur sans biais fonction de T est UMVUE.*

4.7 Implémentation Python

Implémentation Python

```python
import numpy as np
import matplotlib.pyplot as plt

np.random.seed(42)

# --- Comparaison biais et EQM de deux estimateurs de sigma^2 ---
mu_true, sigma2_true = 0, 4
n = 10
n_sim = 50000

estimates_biased = []   # diviseur n (EMV)
estimates_unbiased = []  # diviseur n-1

for _ in range(n_sim):
    sample = np.random.normal(mu_true, np.sqrt(sigma2_true), n)
    estimates_biased.append(np.var(sample, ddof=0))
    estimates_unbiased.append(np.var(sample, ddof=1))

estimates_biased = np.array(estimates_biased)
estimates_unbiased = np.array(estimates_unbiased)

print("=== Estimateur EMV (diviseur n) ===")
print(f"  E[hat_sigma^2] = {np.mean(estimates_biased):.4f} (biais = "
      f"{np.mean(estimates_biased) - sigma2_true:.4f})")
print(f"  Var = {np.var(estimates_biased):.4f}")
print(f"  EQM = {np.mean((estimates_biased - sigma2_true)**2):.4f}")

print("\n=== Estimateur S^2 (diviseur n-1) ===")
print(f"  E[S^2] = {np.mean(estimates_unbiased):.4f} (biais = "
      f"{np.mean(estimates_unbiased) - sigma2_true:.4f})")
```

```
print(f"  Var = {np.var(estimates_unbiased):.4f}")
print(f"  EQM = {np.mean((estimates_unbiased - sigma2_true)**2):.4f}")

# --- Borne de Cramer-Rao ---
# Pour N(mu, sigma^2=4 connu) : I(mu) = 1/sigma^2 = 0.25
# Borne CR pour n obs : 1/(n*I) = sigma^2/n
print(f"\nBorne de Cramer-Rao pour mu: {sigma2_true/n:.4f}")
print(f"Variance empirique de Xbar:
↪  {np.var([np.mean(np.random.normal(0,2,n)) for _ in
↪  range(n_sim)]):.4f}")

# --- Visualisation ---
fig, ax = plt.subplots(figsize=(8, 5))
ax.hist(estimates_biased, bins=50, alpha=0.5, density=True, label='EMV
↪  (n)')
ax.hist(estimates_unbiased, bins=50, alpha=0.5, density=True,
↪  label='$S^2$ (n-1)')
ax.axvline(sigma2_true, color='red', lw=2,
↪  label=f'$\\sigma^2={sigma2_true}$')
ax.set_xlabel(r'$\hat{\sigma}^2$')
ax.set_ylabel('Densite')
ax.set_title(f'Distribution des estimateurs de $\\sigma^2$ (n={n})')
ax.legend()
plt.savefig("ch04_biais_variance.pdf")
plt.show()
```

4.8 Exercices

Exercice 4.1 ($\star$ – Biais)**.** Soit $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{E}(\lambda)$. L'estimateur $\hat{\lambda} = 1/\bar{X}$ est-il sans biais pour λ ? (Indication : étudier la loi de $n\bar{X}$.)

Exercice 4.2 ($\star\star$ – Cramér–Rao)**.**

1. Calculer la borne de CR pour l'estimation sans biais de p dans le modèle $\mathcal{B}(1, p)$.
2. Montrer que $\bar{X}$ atteint cette borne.
3. En déduire que $\bar{X}$ est UMVUE pour p.

Exercice 4.3 ($\star\star$ – Rao–Blackwell)**.** Soit $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{P}(\lambda)$. On considère l'estimateur $\hat{\theta} = \mathbf{1}_{X_1=0}$ de $g(\lambda) = e^{-\lambda} = \mathbb{P}(X = 0)$.

1. Vérifier que $\hat{\theta}$ est sans biais.
2. $T = \sum X_i$ est exhaustive pour λ. Calculer $\hat{\theta}^* = \mathbb{E}[\hat{\theta} \mid T]$.
3. Comparer $\mathrm{Var}(\hat{\theta})$ et $\mathrm{Var}(\hat{\theta}^*)$.

Exercice 4.4 ($\star\star$ – Compromis biais-variance)**.** Soit $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$. On considère l'estimateur $\hat{\theta}_c = c \cdot S^2$ de σ^2 pour $c > 0$.

1. Calculer le biais et l'EQM de $\hat{\theta}_c$ en fonction de c.

2. Trouver la valeur c^* qui minimise l'EQM. Est-ce $c = 1/(n-1)$?

Exercice 4.5 ($\star\star\star$ – Projet : efficacité par simulation)**.** Pour la loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = 3$:

1. Vérifier numériquement que $\bar{X}$ est efficace (atteint la borne CR).
2. Comparer par simulation l'EQM de $\bar{X}$, de la médiane et de la variance empirique comme estimateurs de λ.
3. Tracer l'efficacité relative en fonction de n.

Formules clés

$$\text{Biais}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$
$$\text{EQM}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [\text{Biais}(\hat{\theta})]^2$$
$$\text{Cramér–Rao} : \text{Var}(\hat{\theta}) \geq \frac{[g'(\theta)]^2}{nI(\theta)}$$
$$\text{Rao–Blackwell} : \hat{\theta}^* = \mathbb{E}[\hat{\theta} \mid T],\ \text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta})$$

Consistance : Biais $\to 0$ et Var $\to 0 \implies \hat{\theta} \xrightarrow{\mathbb{P}} \theta$.

Chapitre 5

Intervalles de Confiance

« Une estimation sans mesure d'incertitude est une estimation incomplète. »

5.1 Exemple motivant

Un sondage interroge $n = 1000$ personnes : 520 sont favorables à une réforme. On estime la proportion p par $\hat{p} = 0{,}52$. Mais quelle est la *précision* de cette estimation ? Un intervalle de confiance répond : « avec 95% de confiance, $p \in [0{,}489;\ 0{,}551]$ ».

5.2 Définition et interprétation

Définition 5.1 (Intervalle de confiance)**.** Un **intervalle de confiance** (IC) de niveau $1-\alpha$ pour θ est un intervalle aléatoire $[L(X_1,\dots,X_n),\ U(X_1,\dots,X_n)]$ tel que :

$$\mathbb{P}_\theta(L \le \theta \le U) \ge 1-\alpha \quad \forall \theta \in \Theta.$$

Attention — Piège classique

L'interprétation correcte est **fréquentiste** : si on répète l'expérience un grand nombre de fois, $(1-\alpha)\times 100\%$ des IC construits contiendront la vraie valeur θ.

Erreur courante : « Il y a 95% de chances que θ soit dans l'IC ». Non ! θ est fixé ; c'est l'intervalle qui est aléatoire. Après observation, l'IC contient θ ou non.

Définition 5.2 (Quantité pivotale)**.** Une **quantité pivotale** est une fonction $Q(X_1,\dots,X_n,\theta)$ dont la loi ne dépend pas de θ. On inverse la relation $\mathbb{P}(a \le Q \le b) = 1-\alpha$ pour obtenir l'IC.

5.3 IC pour la moyenne

5.3.1 Variance connue

Théorème 5.3. *Soit $X_1,\dots,X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu,\sigma^2)$ avec σ^2 connu. Un IC de niveau $1-\alpha$ pour μ est :*

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right],$$

où $z_{\alpha/2}$ est le quantile d'ordre $1-\alpha/2$ de $\mathcal{N}(0,1)$.

Démonstration. La quantité pivotale est $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$. On a $\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$, et on isole μ. □

5.3.2 Variance inconnue

Théorème 5.4. *Soit $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec σ^2 inconnu. Un IC de niveau $1-\alpha$ pour μ est :*

$$\left[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}},\ \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right],$$

où $t_{\alpha/2,n-1}$ est le quantile d'ordre $1-\alpha/2$ de la loi $t(n-1)$.

Démonstration. La quantité pivotale est $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ (cf. théorème 2.14). □

Remarque 5.5. Pour n grand ($n \geq 30$ en pratique), $t_{\alpha/2,n-1} \approx z_{\alpha/2}$ et les deux IC coïncident. De plus, par le TCL, l'IC de Student est approximativement valide même sans normalité pour n grand.

5.4 IC pour la variance

Théorème 5.6. *Sous le modèle normal, un IC de niveau $1-\alpha$ pour σ^2 est :*

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}},\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right],$$

où $\chi^2_{\alpha/2,n-1}$ et $\chi^2_{1-\alpha/2,n-1}$ sont les quantiles de la loi $\chi^2(n-1)$.

Démonstration. La quantité pivotale est $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. □

5.5 IC pour une proportion

Théorème 5.7. *Soit $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{B}(1,p)$. Pour n grand, un IC approximatif de niveau $1-\alpha$ pour p est :*

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right],$$

où $\hat{p} = \bar{X}$.

Démonstration. Par le TCL, $\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$. On remplace p par $\hat{p}$ dans l'écart-type (méthode de Wald). □

Attention — Piège classique

L'IC de Wald peut être mauvais pour p proche de 0 ou 1, ou pour n petit. L'IC de Wilson (ou Agresti–Coull) est préférable :

$$\tilde{p} = \frac{\hat{p}n + z^2/2}{n+z^2}, \quad \tilde{p} \pm z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+z^2}}.$$

5.6 IC pour la différence de deux moyennes

Théorème 5.8. *Soient* $X_1, \ldots, X_{n_1} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_1, \sigma^2)$ *et* $Y_1, \ldots, Y_{n_2} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_2, \sigma^2)$ *indépendants, avec variance commune* σ^2 *inconnue. Un IC pour* $\mu_1 - \mu_2$ *est :*

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2,\nu} \cdot S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

où $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ *est la variance poolée et* $\nu = n_1 + n_2 - 2$.

5.7 Taille d'échantillon

Proposition 5.9. *Pour estimer* μ *avec une marge d'erreur* E *au niveau* $1 - \alpha$ *:*

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{E} \right)^2.$$

Pour une proportion : $n \geq \frac{z_{\alpha/2}^2 \cdot p(1-p)}{E^2} \leq \frac{z_{\alpha/2}^2}{4E^2}$.

5.8 Implémentation Python

Implémentation Python

```python
import numpy as np
from scipy import stats

# --- IC pour la moyenne (variance inconnue) ---
data = np.array([12.1, 11.8, 12.5, 12.3, 11.9, 12.7, 12.0, 12.4, 11.7,
↪  12.2])
n = len(data)
xbar = np.mean(data)
s = np.std(data, ddof=1)
alpha = 0.05
t_crit = stats.t.ppf(1 - alpha/2, df=n-1)

ic_lower = xbar - t_crit * s / np.sqrt(n)
ic_upper = xbar + t_crit * s / np.sqrt(n)
print(f"IC 95% pour mu: [{ic_lower:.4f}, {ic_upper:.4f}]")

# Avec scipy directement
ci = stats.t.interval(1-alpha, df=n-1, loc=xbar, scale=s/np.sqrt(n))
print(f"scipy:          [{ci[0]:.4f}, {ci[1]:.4f}]")

# --- IC pour la variance ---
chi2_lower = stats.chi2.ppf(alpha/2, df=n-1)
chi2_upper = stats.chi2.ppf(1-alpha/2, df=n-1)
ic_var = ((n-1)*s**2/chi2_upper, (n-1)*s**2/chi2_lower)
print(f"IC 95% pour sigma^2: [{ic_var[0]:.4f}, {ic_var[1]:.4f}]")
```

```python
# --- IC pour une proportion ---
n_sondage, x_fav = 1000, 520
p_hat = x_fav / n_sondage
z_crit = stats.norm.ppf(1 - alpha/2)
margin = z_crit * np.sqrt(p_hat * (1-p_hat) / n_sondage)
print(f"IC 95% pour p (Wald):   [{p_hat-margin:.4f},
↪  {p_hat+margin:.4f}]")

# IC de Wilson
p_tilde = (x_fav + z_crit**2/2) / (n_sondage + z_crit**2)
margin_w = z_crit * np.sqrt(p_tilde*(1-p_tilde)/(n_sondage+z_crit**2))
print(f"IC 95% pour p (Wilson): [{p_tilde-margin_w:.4f},
↪  {p_tilde+margin_w:.4f}]")

# --- Taille d'echantillon ---
E = 0.03  # marge d'erreur souhaitee
n_needed = (z_crit**2 * 0.25) / E**2  # p(1-p) <= 0.25
print(f"Taille minimale pour marge {E} sur proportion: n >=
↪  {int(np.ceil(n_needed))}")

# --- Visualisation : couverture de l'IC ---
import matplotlib.pyplot as plt
mu_true = 12.0
sigma_true = 0.3
n_ic = 10
n_rep = 50
np.random.seed(1)
fig, ax = plt.subplots(figsize=(10, 6))
for i in range(n_rep):
    sample = np.random.normal(mu_true, sigma_true, n_ic)
    m, se = np.mean(sample), np.std(sample, ddof=1)/np.sqrt(n_ic)
    t_c = stats.t.ppf(0.975, df=n_ic-1)
    lo, hi = m - t_c*se, m + t_c*se
    color = 'blue' if lo <= mu_true <= hi else 'red'
    ax.plot([lo, hi], [i, i], color=color, lw=1.5)
    ax.plot(m, i, 'o', color=color, markersize=3)
ax.axvline(mu_true, color='green', lw=2, linestyle='--',
↪  label=f'$\\mu={mu_true}$')
ax.set_xlabel('Valeur')
ax.set_ylabel('Repetition')
ax.set_title('50 IC a 95% : les rouges ne contiennent pas $\\mu$')
ax.legend()
plt.tight_layout()
plt.savefig("ch05_couverture_ic.pdf")
plt.show()
```

5.9 Exercices

Exercice 5.1 ($\star$ – IC pratique)**.** On mesure le poids (en g) de 15 sachets de thé : $\bar{x} = 2{,}03$, $s = 0{,}12$. Construire un IC à 99% pour le poids moyen, en supposant la normalité.

Exercice 5.2 ($\star$ – Proportion)**.** Sur 500 pièces produites, 23 sont défectueuses. Construire un IC à 95% pour le taux de défauts (Wald et Wilson).

Exercice 5.3 ($\star\star$ – Taille d'échantillon)**.** On veut estimer la taille moyenne des étudiants avec une précision de 1 cm au niveau 95%. On suppose $\sigma \approx 8$ cm. Quelle taille d'échantillon faut-il ?

Exercice 5.4 ($\star\star$ – IC pour le rapport de variances)**.** Soient deux échantillons indépendants de lois normales. Construire un IC pour σ_1^2/σ_2^2 en utilisant la loi de Fisher.

Exercice 5.5 ($\star\star\star$ – Projet : couverture réelle)**.** Par simulation Monte Carlo, comparer le taux de couverture réel des IC de Wald et de Wilson pour $p = 0{,}05$ et $n = 20, 50, 100, 500$. Lequel est le plus fiable pour petits n ?

Formules clés

$$\text{Moyenne } (\sigma \text{ connu}) : \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\text{Moyenne } (\sigma \text{ inconnu}) : \bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$$

$$\text{Proportion (Wald)} : \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\text{Variance} : \left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right]$$

$$\text{Taille} : n \geq \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

Chapitre 6

Tests d'Hypothèses — Cadre Général

« L'absence de preuve n'est pas la preuve de l'absence. » — Carl Sagan

6.1 Exemple motivant

Un fabricant affirme que ses vis ont une résistance moyenne de 50 kN. Un client prélève $n = 25$ vis et mesure $\bar{x} = 48{,}2$ kN avec $s = 4{,}5$ kN. Cette différence est-elle significative ou simplement due au hasard de l'échantillonnage ?

6.2 Formulation des hypothèses

Définition 6.1 (Hypothèses)**.**

- H_0 : **hypothèse nulle** — l'affirmation que l'on cherche à remettre en question.
- H_1 : **hypothèse alternative** — ce que l'on soupçonne si H_0 est fausse.

Formes classiques pour un paramètre θ :

Type	H_0	H_1
Bilatéral	$\theta = \theta_0$	$\theta \neq \theta_0$
Unilatéral gauche	$\theta \geq \theta_0$	$\theta < \theta_0$
Unilatéral droit	$\theta \leq \theta_0$	$\theta > \theta_0$

6.3 Erreurs de première et deuxième espèce

Définition 6.2 (Types d'erreurs)**.**

	H_0 vraie	H_0 fausse
Rejeter H_0	**Erreur de type I** (α)	Décision correcte
Ne pas rejeter H_0	Décision correcte	**Erreur de type II** (β)

- $\alpha = \mathbb{P}(\text{rejeter } H_0 \mid H_0 \text{ vraie})$: **niveau** (ou seuil) du test.
- $\beta = \mathbb{P}(\text{ne pas rejeter } H_0 \mid H_1 \text{ vraie})$.
- $1 - \beta$: **puissance** du test.

Intuition statistique

α est le risque de « fausse alarme ». β est le risque de « manquer » un vrai effet. En général, on fixe α (souvent 5%) et on cherche à maximiser la puissance.

6.4 Structure d'un test

Définition 6.3 (Test statistique)**.** Un **test** de H_0 contre H_1 au niveau α est une règle de décision :

1. Calculer une **statistique de test** $T = T(X_1, \ldots, X_n)$.
2. Déterminer la **région critique** W telle que $\mathbb{P}_{\theta_0}(T \in W) \leq \alpha$.
3. **Décision** : rejeter H_0 si $T \in W$.

Définition 6.4 (p-valeur)**.** La **p-valeur** est la probabilité, sous H_0, d'observer une valeur de la statistique de test au moins aussi extrême que celle observée :

$$p = \mathbb{P}_{H_0}(T \geq t_{\text{obs}}) \quad \text{(test unilatéral droit)}.$$

On rejette H_0 si et seulement si $p \leq \alpha$.

Attention — Piège classique

La p-valeur n'est PAS la probabilité que H_0 soit vraie. C'est la probabilité d'observer des données aussi extrêmes **si** H_0 était vraie.
Autres pièges :

- « $p > 0{,}05$ » ne signifie pas que H_0 est vraie, seulement que les données ne fournissent pas de preuve suffisante contre H_0.
- La significativité statistique ($p < 0{,}05$) n'implique pas la significativité **pratique** (l'effet peut être négligeable).

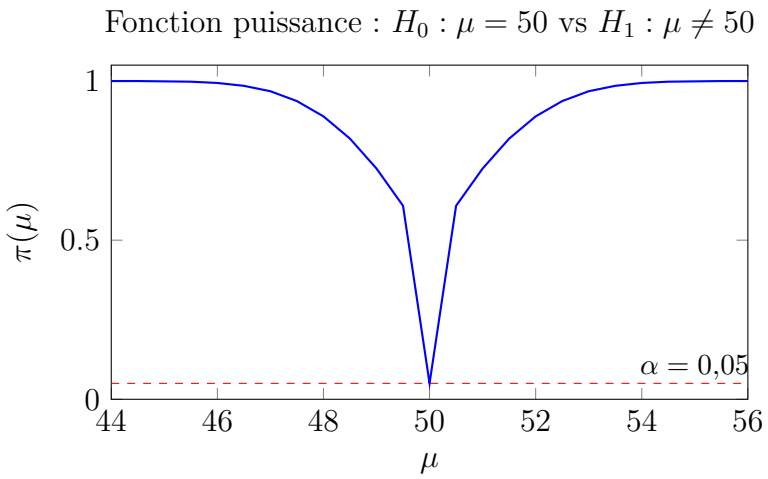
6.5 Puissance d'un test

Définition 6.5 (Fonction puissance)**.** La **fonction puissance** du test est :

$$\pi(\theta) = \mathbb{P}_\theta(\text{rejeter } H_0).$$

- Pour $\theta \in \Theta_0$ (sous H_0) : $\pi(\theta) \leq \alpha$ (contrôle du niveau).
- Pour $\theta \in \Theta_1$ (sous H_1) : $\pi(\theta) = 1 - \beta(\theta)$ (puissance).

Un bon test a une puissance élevée pour les valeurs de θ sous H_1.



6.6 Lemme de Neyman–Pearson

Théorème 6.6 (Lemme de Neyman–Pearson)**.** *Pour tester $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ (hypothèses simples), le test le plus puissant de niveau α rejette H_0 lorsque le* ***rapport de vraisemblance*** *est grand :*

$$\Lambda(\mathbf{x}) = \frac{L(\theta_1; \mathbf{x})}{L(\theta_0; \mathbf{x})} > k_\alpha,$$

où k_α est choisi tel que $\mathbb{P}_{\theta_0}(\Lambda > k_\alpha) = \alpha$.

Esquisse de preuve. Soit φ la règle du test NP et φ' tout autre test de niveau $\leq \alpha$. On veut montrer $\mathbb{E}_{\theta_1}[\varphi] \geq \mathbb{E}_{\theta_1}[\varphi']$.

Sur la région de rejet NP : $\varphi = 1$, $L(\theta_1) \geq k \cdot L(\theta_0)$. Donc $(\varphi - \varphi')(L(\theta_1) - kL(\theta_0)) \geq 0$ partout (les deux facteurs ont le même signe). En intégrant :

$$\mathbb{E}_{\theta_1}[\varphi - \varphi'] \geq k\, \mathbb{E}_{\theta_0}[\varphi - \varphi'] = k(\alpha - \alpha') \geq 0. \qquad \square$$

6.7 Test du rapport de vraisemblance généralisé

Définition 6.7 (TRV généralisé)**.** Pour $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1 = \Theta \setminus \Theta_0$, la statistique du **test du rapport de vraisemblance** (TRV) est :

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta; \mathbf{x})}{\sup_{\theta \in \Theta} L(\theta; \mathbf{x})}, \qquad 0 \leq \lambda \leq 1.$$

On rejette H_0 pour $\lambda \leq c_\alpha$ (petites valeurs de λ).

Théorème 6.8 (Théorème de Wilks)**.** *Sous H_0 et des conditions de régularité :*

$$-2 \ln \lambda(\mathbf{X}) \xrightarrow{\mathcal{L}} \chi^2(r),$$

où $r = \dim(\Theta) - \dim(\Theta_0)$ est le nombre de contraintes imposées par H_0.

6.8 Implémentation Python

Implémentation Python

```python
import numpy as np
from scipy import stats

# --- Test sur la moyenne (sigma inconnu) ---
# H0: mu = 50, H1: mu != 50
data = np.array([48.2, 47.5, 49.1, 46.8, 50.3, 48.7, 47.9, 49.5,
                 46.2, 48.8, 49.0, 47.3, 48.1, 50.1, 47.0,
                 49.4, 48.6, 46.5, 49.8, 47.2, 48.3, 49.7,
                 47.8, 48.5, 46.9])
n = len(data)
mu0 = 50
xbar = np.mean(data)
s = np.std(data, ddof=1)
t_obs = (xbar - mu0) / (s / np.sqrt(n))
p_value = 2 * stats.t.cdf(t_obs, df=n-1)  # bilateral, t_obs < 0
print(f"t_obs = {t_obs:.4f}, p-valeur = {p_value:.6f}")

# Avec scipy
t_stat, p_val = stats.ttest_1samp(data, mu0)
print(f"scipy: t = {t_stat:.4f}, p = {p_val:.6f}")

if p_val < 0.05:
    print("Conclusion: on rejette H0 au niveau 5%.")
else:
    print("Conclusion: on ne rejette pas H0 au niveau 5%.")

# --- Calcul de puissance ---
from scipy.stats import norm

def power_z_test(mu_true, mu0, sigma, n, alpha=0.05):
    """Puissance d'un test z bilateral."""
    z_alpha = norm.ppf(1 - alpha/2)
    delta = (mu_true - mu0) / (sigma / np.sqrt(n))
    power = 1 - norm.cdf(z_alpha - delta) + norm.cdf(-z_alpha - delta)
    return power

mu_values = np.linspace(44, 56, 200)
powers = [power_z_test(mu, 50, 4.5, 25) for mu in mu_values]

import matplotlib.pyplot as plt
plt.figure(figsize=(8, 5))
plt.plot(mu_values, powers, 'b-', lw=2)
plt.axhline(0.05, color='red', linestyle='--', label=r'$\alpha=0.05$')
plt.axvline(50, color='gray', linestyle=':')
plt.xlabel(r'$\mu$ (vraie valeur)')
plt.ylabel(r'$\pi(\mu)$ (puissance)')
plt.title('Fonction puissance du test z bilateral')
plt.legend()
plt.savefig("ch06_puissance.pdf")
```

```
plt.show()

# --- Taille d'echantillon pour puissance cible ---
def sample_size_z(mu_alt, mu0, sigma, alpha=0.05, power=0.8):
    z_alpha = norm.ppf(1 - alpha/2)
    z_beta = norm.ppf(power)
    n = ((z_alpha + z_beta) * sigma / (mu_alt - mu0))**2
    return int(np.ceil(n))

n_needed = sample_size_z(48, 50, 4.5)
print(f"Taille pour detecter mu=48 avec puissance 80%: n = {n_needed}")
```

6.9 Exercices

Exercice 6.1 ($\star$ – Test de base)**.** Un enseignant affirme que la note moyenne est 12. Sur 20 copies, on observe $\bar{x} = 11{,}2$ et $s = 2{,}5$. Tester $H_0 : \mu = 12$ vs $H_1 : \mu < 12$ au niveau 5%.

Exercice 6.2 ($\star\star$ – Neyman–Pearson)**.** Soit $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$. Appliquer le lemme de Neyman–Pearson pour $H_0 : \mu = 0$ vs $H_1 : \mu = 1$. En déduire la région critique optimale.

Exercice 6.3 ($\star\star$ – Puissance)**.** Pour le test $H_0 : \mu = 100$ vs $H_1 : \mu \neq 100$ avec $\sigma = 15$ et $\alpha = 0{,}05$:

1. Calculer la puissance pour $n = 25$ et $\mu_1 = 105$.
2. Quelle taille d'échantillon faut-il pour une puissance de 90% à $\mu_1 = 105$?

Exercice 6.4 ($\star\star$ – TRV)**.** $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ avec σ^2 inconnu. Calculer λ pour $H_0 : \mu = \mu_0$ et montrer que $-2 \ln \lambda$ est une fonction croissante de $|T|$ où T est la statistique de Student.

Exercice 6.5 ($\star\star\star$ – Projet : simulations de puissance)**.** Pour le test t unilatéral, étudier par simulation : (a) le taux d'erreur de type I pour différentes distributions (normale, exponentielle, uniforme), (b) la puissance en fonction de la taille d'effet $\delta = (\mu_1 - \mu_0)/\sigma$, (c) comparer la puissance du test t et du test de Wilcoxon.

Formules clés

$$\alpha = \mathbb{P}(\text{rejeter } H_0 \mid H_0) \qquad \beta = \mathbb{P}(\text{ne pas rejeter } H_0 \mid H_1)$$

$$p\text{-valeur} = \mathbb{P}_{H_0}(T \geq t_{\text{obs}}) \qquad \pi(\theta) = \mathbb{P}_\theta(\text{rejeter } H_0)$$

$$\text{NP : rejeter si } \frac{L(\theta_1)}{L(\theta_0)} > k_\alpha \qquad \lambda = \frac{\sup_{\Theta_0} L}{\sup_{\Theta} L}$$

Wilks : $-2 \ln \lambda \xrightarrow{\mathcal{L}} \chi^2(r)$.

Chapitre 7

Tests Classiques : z, t, χ^2, F

« La théorie sans la pratique est stérile ; la pratique sans la théorie est aveugle. »

7.1 Exemple motivant

Deux machines A et B produisent des pièces. On veut comparer leur précision (variance) et savoir si elles produisent en moyenne la même longueur. Pour la machine A : $n_1 = 20$, $\bar{x}_1 = 50{,}2$, $s_1 = 1{,}5$. Pour B : $n_2 = 25$, $\bar{x}_2 = 49{,}5$, $s_2 = 2{,}1$. Quels tests utiliser ?

7.2 Test z (variance connue)

Théorème 7.1 (Test z pour la moyenne)**.** $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, σ^2 *connu. Pour* $H_0 : \mu = \mu_0$ *:*

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \text{ sous } H_0.$$

H_1	*Région critique*	*p-valeur*
$\mu \neq \mu_0$	$\lvert Z\rvert > z_{\alpha/2}$	$2\Phi(-\lvert z_{\text{obs}}\rvert)$
$\mu > \mu_0$	$Z > z_\alpha$	$1 - \Phi(z_{\text{obs}})$
$\mu < \mu_0$	$Z < -z_\alpha$	$\Phi(z_{\text{obs}})$

7.2.1 Test z pour deux moyennes

$X_1, \ldots, X_{n_1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$ et $Y_1, \ldots, Y_{n_2} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$ indépendants, variances connues :

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}, \quad H_0 : \mu_1 - \mu_2 = \delta_0.$$

7.2.2 Test z pour une proportion

$\hat{p} = \bar{X}$ avec $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{B}(1, p)$, n grand :

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}, \quad H_0 : p = p_0.$$

7.2.3 Test z pour deux proportions

$H_0 : p_1 = p_2$:

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_{\text{pool}}(1-\hat{p}_{\text{pool}})(1/n_1 + 1/n_2)}}, \quad \hat{p}_{\text{pool}} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}.$$

7.3 Test t de Student

Théorème 7.2 (Test t pour un échantillon)**.** *$X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$, σ^2 inconnu. Pour $H_0 : \mu = \mu_0$:*

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \text{ sous } H_0.$$

Régions critiques analogues au test z avec les quantiles de $t(n-1)$.

7.3.1 Test t pour deux échantillons indépendants

Variances égales :

$$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t(n_1+n_2-2), \quad S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}.$$

Variances inégales (Welch) :

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t(\nu),$$

avec ν calculé par la formule de Welch–Satterthwaite :

$$\nu = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}.$$

7.3.2 Test t apparié

Pour des données appariées (X_i, Y_i), on travaille sur les différences $D_i = X_i - Y_i$:

$$T = \frac{\bar{D}}{S_D/\sqrt{n}} \sim t(n-1), \quad H_0 : \mu_D = 0.$$

7.4 Test du χ^2

7.4.1 Test sur la variance

Théorème 7.3 (Test χ^2 pour la variance)**.** *$X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$. Pour $H_0 : \sigma^2 = \sigma_0^2$:*

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \text{ sous } H_0.$$

7.4.2 Test d'adéquation du χ^2

Théorème 7.4 (Test d'adéquation de Pearson)**.** *Soit k catégories avec effectifs observés O_i et effectifs théoriques $E_i = np_i$:*

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \xrightarrow{\mathcal{L}} \chi^2(k-1-r),$$

où r est le nombre de paramètres estimés. On rejette H_0 pour $\chi^2 > \chi^2_{\alpha,k-1-r}$.

7.4.3 Test d'indépendance du χ^2

Théorème 7.5 (Test d'indépendance)**.** *Pour un tableau de contingence $r \times c$ avec effectifs O_{ij} :*

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}, \quad E_{ij} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{n},$$

suit approximativement $\chi^2((r-1)(c-1))$ sous H_0 : indépendance.

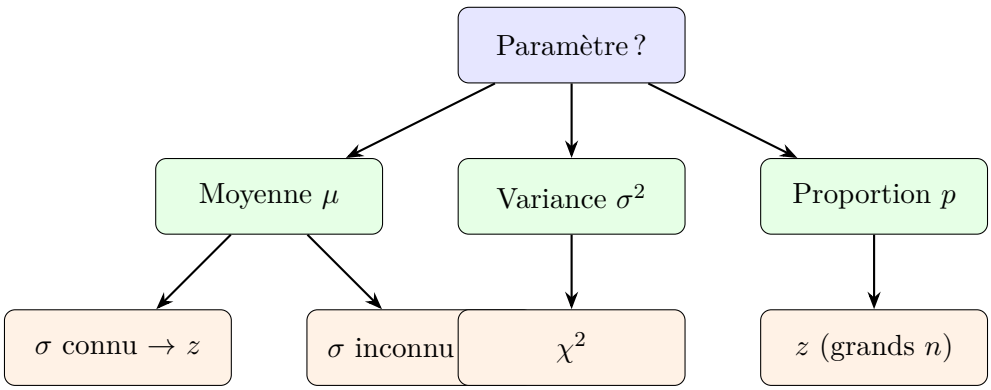
7.5 Test F de Fisher

Théorème 7.6 (Test F pour le rapport de variances)**.** $X_1, \ldots, X_{n_1} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$ *et* $Y_1, \ldots, Y_{n_2} \overset{i.i.d.}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$ *indépendants. Pour $H_0 : \sigma_1^2 = \sigma_2^2$:*

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1) \text{ sous } H_0.$$

On rejette H_0 si $F > F_{\alpha/2,n_1-1,n_2-1}$ ou $F < F_{1-\alpha/2,n_1-1,n_2-1}$.

7.6 Résumé : choix du test



7.7 Implémentation Python

Implémentation Python

```python
import numpy as np
from scipy import stats

# --- Test t pour un echantillon ---
data = np.array([50.2, 49.8, 51.1, 48.9, 50.5, 49.3, 50.8, 49.1,
                 50.0, 49.7, 50.3, 49.5, 50.6, 49.2, 50.4])
t_stat, p_val = stats.ttest_1samp(data, 50)
print(f"Test t (H0: mu=50): t={t_stat:.4f}, p={p_val:.4f}")

# --- Test t pour deux echantillons ---
A = np.array([50.2, 49.8, 51.1, 48.9, 50.5, 49.3, 50.8, 49.1,
              50.0, 49.7, 50.3, 49.5, 50.6, 49.2, 50.4,
              50.1, 49.6, 50.7, 49.4, 50.2])
B = np.array([49.5, 48.2, 50.1, 47.8, 49.0, 48.5, 50.3, 47.5,
              49.2, 48.8, 49.7, 48.0, 49.4, 48.3, 49.1,
              48.7, 49.6, 47.9, 49.3, 48.1, 49.8, 48.4,
              49.0, 48.6, 49.5])

# Variances egales
t_eq, p_eq = stats.ttest_ind(A, B, equal_var=True)
print(f"Test t (egal var): t={t_eq:.4f}, p={p_eq:.4f}")

# Welch (variances inegales)
t_w, p_w = stats.ttest_ind(A, B, equal_var=False)
print(f"Test Welch:        t={t_w:.4f}, p={p_w:.4f}")

# --- Test F pour egalite des variances ---
f_stat = np.var(A, ddof=1) / np.var(B, ddof=1)
df1, df2 = len(A)-1, len(B)-1
p_f = 2 * min(stats.f.cdf(f_stat, df1, df2), 1-stats.f.cdf(f_stat, df1,
↪  df2))
print(f"Test F: F={f_stat:.4f}, p={p_f:.4f}")

# --- Test chi2 d'adequation ---
# De de 60 lancers
obs = np.array([8, 12, 10, 11, 9, 10])  # observes
exp = np.array([10, 10, 10, 10, 10, 10])  # attendus (de equilibre)
chi2, p_chi = stats.chisquare(obs, exp)
print(f"chi2 adequation: chi2={chi2:.4f}, p={p_chi:.4f}")

# --- Test chi2 d'independance ---
# Tableau de contingence
tableau = np.array([[30, 20, 10],
                    [25, 30, 15],
                    [15, 20, 35]])
chi2_ind, p_ind, dof, expected = stats.chi2_contingency(tableau)
```

```
print(f"chi2 independance: chi2={chi2_ind:.4f}, p={p_ind:.4f},
↪  ddl={dof}")

# --- Test de proportion ---
from statsmodels.stats.proportion import proportions_ztest
count = np.array([520])
nobs = np.array([1000])
z_prop, p_prop = proportions_ztest(count, nobs, value=0.5)
print(f"Test proportion (H0: p=0.5): z={z_prop:.4f}, p={p_prop:.4f}")
```

7.8 Exercices

Exercice 7.1 ($\star$ – Tests sur la moyenne)**.** Les temps de réaction (en ms) de 12 sujets après un café sont : 210, 195, 220, 205, 198, 215, 208, 200, 225, 192, 218, 202. Le temps moyen normal est 220 ms. Le café réduit-il significativement le temps de réaction ? (Test t unilatéral, $\alpha = 0{,}05$.)

Exercice 7.2 ($\star$ – Test χ^2)**.** On lance un dé 120 fois et on observe : face 1 : 25, face 2 : 17, face 3 : 22, face 4 : 18, face 5 : 19, face 6 : 19. Le dé est-il équilibré ? ($\alpha = 0{,}05$.)

Exercice 7.3 ($\star\star$ – Deux échantillons)**.** Comparer les moyennes de deux traitements : Traitement A ($n_1 = 15$, $\bar{x}_1 = 72{,}3$, $s_1 = 8{,}5$) et Traitement B ($n_2 = 18$, $\bar{x}_2 = 68{,}1$, $s_2 = 9{,}2$). D'abord tester l'égalité des variances (test F), puis choisir le test t adapté.

Exercice 7.4 ($\star\star$ – Test apparié)**.** Dix patients ont leur pression artérielle mesurée avant et après un traitement. Avant : 140, 135, 150, 145, 138, 142, 155, 148, 137, 143. Après : 132, 130, 142, 138, 135, 136, 148, 140, 133, 138. Le traitement est-il efficace ?

Exercice 7.5 ($\star\star\star$ – Projet : robustesse des tests)**.** Par simulation, étudier la robustesse du test t à la non-normalité. Générer des échantillons de lois exponentielle, log-normale et uniforme. Pour $n = 5, 10, 30, 100$, comparer le taux d'erreur de type I réel au niveau nominal 5%.

Formules clés

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \qquad t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \qquad F = \frac{S_1^2}{S_2^2}$$

$$\chi^2_{\text{adéq}} = \sum \frac{(O_i - E_i)^2}{E_i} \qquad \chi^2_{\text{ind}} = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2} \qquad z_{\text{prop}} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Chapitre 8

Régression Linéaire Simple

« La régression est l'outil le plus utile et le plus mal utilisé de la statistique. »

8.1 Exemple motivant

On dispose de données sur la surface habitable (en m^2) et le prix de vente (en k€) de 15 appartements. On veut modéliser la relation Prix $= f($Surface$)$ et prédire le prix pour une surface donnée.

8.2 Le modèle

Définition 8.1 (Modèle de régression linéaire simple)**.**

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

avec les hypothèses :

1. $\mathbb{E}[\varepsilon_i] = 0$ (erreurs centrées),
2. $\mathrm{Var}(\varepsilon_i) = \sigma^2$ (homoscédasticité),
3. $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ pour $i \neq j$ (erreurs non corrélées),
4. (optionnel) $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ (normalité).

β_0 : ordonnée à l'origine. β_1 : pente (effet marginal de x sur Y).

8.3 Estimation par moindres carrés ordinaires (MCO)

Théorème 8.2 (Estimateurs MCO)**.** *Les estimateurs qui minimisent* $\sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$ *sont :*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}, \tag{8.1}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}. \tag{8.2}$$

Démonstration. Posons $Q(\beta_0, \beta_1) = \sum (Y_i - \beta_0 - \beta_1 x_i)^2$. Les conditions du premier ordre sont :

$$\frac{\partial Q}{\partial \beta_0} = -2\sum (Y_i - \beta_0 - \beta_1 x_i) = 0, \quad \frac{\partial Q}{\partial \beta_1} = -2\sum x_i (Y_i - \beta_0 - \beta_1 x_i) = 0.$$

La première donne $\beta_0 = \bar{Y} - \beta_1 \bar{x}$. Substitution dans la seconde :

$$\sum x_i (Y_i - \bar{Y} + \beta_1 \bar{x} - \beta_1 x_i) = 0 \implies \beta_1 \sum (x_i - \bar{x})^2 = \sum (x_i - \bar{x})(Y_i - \bar{Y}). \qquad \square$$

Théorème 8.3 (Propriétés des MCO). *1.* $\mathbb{E}[\hat{\beta}_1] = \beta_1$ *et* $\mathbb{E}[\hat{\beta}_0] = \beta_0$ *(sans biais).*

2. $\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$ *et* $\mathrm{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$.

3. $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{\mathrm{SCR}}{n-2}$ *est sans biais pour* σ^2.

Théorème 8.4 (Gauss–Markov). *Sous les hypothèses (1)–(3) du modèle, les MCO sont les* ***meilleurs estimateurs linéaires sans biais*** *(BLUE) : parmi tous les estimateurs linéaires et sans biais, ils ont la plus petite variance.*

8.4 Décomposition de la variance

Théorème 8.5 (Décomposition SCT = SCE + SCR).

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\mathrm{SCT}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\mathrm{SCE}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\mathrm{SCR}},$$

où SCT = somme des carrés totale, SCE = somme des carrés expliquée, SCR = somme des carrés résiduelle.

Définition 8.6 (Coefficient de détermination).

$$R^2 = \frac{\mathrm{SCE}}{\mathrm{SCT}} = 1 - \frac{\mathrm{SCR}}{\mathrm{SCT}} \in [0, 1].$$

R^2 mesure la proportion de la variance de Y expliquée par le modèle. En régression simple, $R^2 = r_{xy}^2$.

8.5 Inférence sous normalité

Sous $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$:

Théorème 8.7 (Tests et IC pour les coefficients).

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{S_{xx}}} \sim t(n-2), \qquad \frac{\hat{\beta}_0 - \beta_0}{\hat{\sigma}\sqrt{1/n + \bar{x}^2/S_{xx}}} \sim t(n-2).$$

- ***Test*** $H_0 : \beta_1 = 0$ *(pas de relation linéaire) :* $T = \hat{\beta}_1 / \mathrm{SE}(\hat{\beta}_1) \sim t(n-2)$.
- ***IC 95%*** *pour* β_1 *:* $\hat{\beta}_1 \pm t_{0{,}025, n-2} \cdot \mathrm{SE}(\hat{\beta}_1)$.

Théorème 8.8 (Test global de Fisher).

$$F = \frac{\mathrm{SCE}/1}{\mathrm{SCR}/(n-2)} = \frac{R^2/(1)}{(1-R^2)/(n-2)} \sim F(1, n-2) \text{ sous } H_0 : \beta_1 = 0.$$

En régression simple, $F = T^2$ *où* T *est la statistique du test de* β_1.

8.5.1 Prédiction

Définition 8.9 (Intervalle de confiance vs. de prédiction)**.** Pour $x = x_0$:

$$\text{IC pour } \mathbb{E}[Y|x_0] : \hat{Y}_0 \pm t_{\alpha/2,n-2} \cdot \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \tag{8.3}$$

$$\text{IP pour } Y_{\text{new}} : \hat{Y}_0 \pm t_{\alpha/2,n-2} \cdot \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}. \tag{8.4}$$

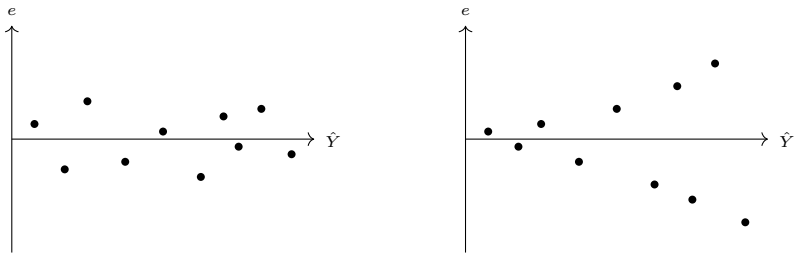
L'intervalle de prédiction est toujours plus large (incertitude additionnelle de ε).

8.6 Diagnostic des résidus

Définition 8.10 (Résidus)**.** Les résidus sont $e_i = Y_i - \hat{Y}_i$. On examine :

1. **Résidus vs. valeurs ajustées** : détecte non-linéarité et hétéroscédasticité.
2. **QQ-plot des résidus** : vérifie la normalité.
3. **Résidus vs. ordre** : détecte l'autocorrélation.

Bon (homoscédastique) Mauvais (hétéroscédastique)



8.7 Implémentation Python

Implémentation Python

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import statsmodels.api as sm

# --- Donnees : surface et prix ---
surface = np.array([25, 30, 35, 40, 45, 50, 55, 60, 65, 70,
                    75, 80, 90, 100, 120])
prix = np.array([95, 110, 125, 140, 148, 160, 175, 185, 195, 210,
                 220, 240, 260, 290, 350])

# --- Regression avec scipy ---
```

```
slope, intercept, r_value, p_value, std_err = stats.linregress(surface,
↪  prix)
print(f"beta_1 = {slope:.4f} (SE={std_err:.4f}, p={p_value:.6f})")
print(f"beta_0 = {intercept:.4f}")
print(f"R^2 = {r_value**2:.4f}")

# --- Regression avec statsmodels (plus complet) ---
X = sm.add_constant(surface)
model = sm.OLS(prix, X).fit()
print(model.summary())

# --- Graphique de regression ---
fig, axes = plt.subplots(2, 2, figsize=(12, 10))

# 1. Droite de regression
x_pred = np.linspace(20, 130, 100)
y_pred = intercept + slope * x_pred
axes[0, 0].scatter(surface, prix, color='blue', label='Donnees')
axes[0, 0].plot(x_pred, y_pred, 'r-', lw=2, label='Regression')
axes[0, 0].set_xlabel('Surface (m$^2$)')
axes[0, 0].set_ylabel('Prix (kEUR)')
axes[0, 0].set_title('Regression lineaire simple')
axes[0, 0].legend()

# 2. Residus vs valeurs ajustees
y_hat = intercept + slope * surface
residuals = prix - y_hat
axes[0, 1].scatter(y_hat, residuals)
axes[0, 1].axhline(0, color='red', linestyle='--')
axes[0, 1].set_xlabel('Valeurs ajustees')
axes[0, 1].set_ylabel('Residus')
axes[0, 1].set_title('Residus vs Ajustees')

# 3. QQ-plot
stats.probplot(residuals, plot=axes[1, 0])
axes[1, 0].set_title('QQ-plot des residus')

# 4. Histogramme des residus
axes[1, 1].hist(residuals, bins=6, edgecolor='black', alpha=0.7)
axes[1, 1].set_title('Distribution des residus')

plt.tight_layout()
plt.savefig("ch08_regression.pdf")
plt.show()

# --- Prediction et intervalle ---
x_new = 85
predictions = model.get_prediction(sm.add_constant([x_new]))
pred_summary = predictions.summary_frame(alpha=0.05)
print(f"\nPrediction pour {x_new} m^2:")
```

```r
print(pred_summary)
```

8.8 Exercices

Exercice 8.1 ($\star$ – Calcul MCO)**.** Calculer à la main $\hat{\beta}_0$, $\hat{\beta}_1$ et R^2 pour les données :

x	1	2	3	4	5
y	2.1	3.8	6.2	7.9	10.3

Exercice 8.2 ($\star\star$ – Gauss–Markov)**.** Montrer que parmi les estimateurs de la forme $\hat{\beta}_1 = \sum c_i Y_i$ qui sont linéaires et sans biais pour β_1, le MCO a la plus petite variance.

Exercice 8.3 ($\star\star$ – Diagnostics)**.** Réaliser une régression de la consommation de carburant sur le poids du véhicule (données `mtcars`). Analyser les résidus. La relation est-elle vraiment linéaire ?

Exercice 8.4 ($\star\star\star$ – Projet)**.** Collecter des données (immobilier, météo, etc.), ajuster un modèle de régression simple, tester $\beta_1 = 0$, analyser les résidus, construire des intervalles de prédiction. Rapport complet avec code Python.

Formules clés

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}$$

$$R^2 = 1 - \frac{\text{SCR}}{\text{SCT}} = r_{xy}^2$$

$$T = \frac{\hat{\beta}_1}{\hat{\sigma}/\sqrt{S_{xx}}} \sim t(n-2)$$

$$F = \frac{\text{SCE}/1}{\text{SCR}/(n-2)} \sim F(1, n-2)$$

Chapitre 9

Régression Linéaire Multiple et ANOVA

« En science, la réalité est rarement unidimensionnelle. »

9.1 Exemple motivant

On veut prédire le prix d'un appartement en fonction de la surface, du nombre de pièces, de l'étage et de la distance au centre-ville. Un seul prédicteur ne suffit plus : il faut la régression **multiple**.

9.2 Le modèle matriciel

Définition 9.1 (Modèle linéaire général)**.**

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

où $\mathbf{Y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times p}$ (matrice de design, rang p), $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

Explicitement : $Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i$.

9.3 Estimation MCO

Théorème 9.2 (Solution MCO)**.** *L'estimateur MCO est :*

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

Il minimise $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}\|^2$.

Démonstration. $Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$. La condition $\nabla_{\boldsymbol{\beta}} Q = \mathbf{0}$ donne $-2\mathbf{X}^\top(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$, soit $\mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} = \mathbf{X}^\top \mathbf{Y}$ (**équations normales**). □

Théorème 9.3 (Propriétés)**.**

1. $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ *(sans biais).*
2. $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$.
3. $\hat{\sigma}^2 = \frac{\|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2}{n-p}$ *est sans biais.*

4. $\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1})$.

Définition 9.4 (Matrice de projection)**.** La **matrice chapeau** est $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. On a $\hat{\mathbf{Y}} = \mathbf{HY}$ et $\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$.

9.4 Inférence

9.4.1 Test individuel

Pour $H_0 : \beta_j = 0$:

$$T_j = \frac{\hat{\beta}_j}{\mathrm{SE}(\hat{\beta}_j)} \sim t(n-p), \quad \mathrm{SE}(\hat{\beta}_j) = \hat{\sigma}\sqrt{[(\mathbf{X}^\top\mathbf{X})^{-1}]_{jj}}.$$

9.4.2 Test global de Fisher

$H_0 : \beta_1 = \cdots = \beta_{p-1} = 0$:

$$F = \frac{(\mathrm{SCT} - \mathrm{SCR})/(p-1)}{\mathrm{SCR}/(n-p)} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)} \sim F(p-1, n-p).$$

Définition 9.5 (R^2 ajusté)**.**

$$R^2_{\mathrm{adj}} = 1 - \frac{n-1}{n-p}(1-R^2).$$

Pénalise l'ajout de variables inutiles.

9.4.3 Test de sous-modèle

Pour comparer un modèle complet (p paramètres) et un modèle réduit ($q < p$ paramètres) :

$$F = \frac{(\mathrm{SCR}_{\text{réd}} - \mathrm{SCR}_{\text{comp}})/(p-q)}{\mathrm{SCR}_{\text{comp}}/(n-p)} \sim F(p-q, n-p).$$

9.5 ANOVA à un facteur

Définition 9.6 (Modèle ANOVA)**.** Soit k groupes avec n_j observations dans le groupe j :

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \quad i = 1, \ldots, n_j,\ j = 1, \ldots, k,$$

avec $\sum n_j\alpha_j = 0$ et $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.

$H_0 : \alpha_1 = \cdots = \alpha_k = 0$ (toutes les moyennes sont égales).

Théorème 9.7 (Tableau ANOVA)**.**

Source	***SC***	***ddl***	***CM***	F
Facteur	$\mathrm{SCE} = \sum n_j(\bar{Y}_j - \bar{Y})^2$	$k-1$	CME	$F = \frac{\mathrm{CME}}{\mathrm{CMR}}$
Résidu	$\mathrm{SCR} = \sum\sum(Y_{ij} - \bar{Y}_j)^2$	$n-k$	CMR	
Total	$\mathrm{SCT} = \sum\sum(Y_{ij} - \bar{Y})^2$	$n-1$		

Sous H_0 *:* $F \sim F(k-1, n-k)$.

Intuition statistique

L'ANOVA est un cas particulier de régression multiple avec des variables explicatives indicatrices (dummy variables). Le test F de l'ANOVA est exactement le test global de Fisher du modèle correspondant.

9.5.1 Comparaisons multiples

Attention — Piège classique

Si H_0 est rejetée, on veut savoir *quels* groupes diffèrent. Les tests de comparaison deux à deux sans correction augmentent le risque d'erreur de type I global. On utilise :

- **Bonferroni** : $\alpha_{\text{corrigé}} = \alpha/m$ où $m = \binom{k}{2}$.
- **Tukey HSD** : basé sur la loi de l'étendue studentisée.

9.6 Implémentation Python

Implémentation Python

```python
import numpy as np
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols
from scipy import stats
import matplotlib.pyplot as plt

# --- Regression multiple ---
np.random.seed(42)
n = 50
surface = np.random.uniform(25, 120, n)
pieces = np.random.randint(1, 6, n)
etage = np.random.randint(0, 10, n)
prix = 30 + 2.5*surface + 15*pieces + 3*etage + np.random.normal(0, 15,
↪  n)

df = pd.DataFrame({'prix': prix, 'surface': surface,
                   'pieces': pieces, 'etage': etage})

# Ajustement
X = sm.add_constant(df[['surface', 'pieces', 'etage']])
model = sm.OLS(df['prix'], X).fit()
print(model.summary())

# R^2 ajuste
print(f"\nR^2 = {model.rsquared:.4f}")
print(f"R^2 ajuste = {model.rsquared_adj:.4f}")
```

```python
# --- ANOVA a un facteur ---
# Rendement de 3 types d'engrais
engrais_A = [20.1, 21.5, 19.8, 22.0, 20.5, 21.2]
engrais_B = [23.4, 24.1, 22.8, 23.9, 24.5, 23.2]
engrais_C = [19.5, 20.2, 18.9, 20.8, 19.1, 20.5]

f_stat, p_val = stats.f_oneway(engrais_A, engrais_B, engrais_C)
print(f"\nANOVA: F={f_stat:.4f}, p={p_val:.6f}")

# Avec statsmodels (plus de details)
df_anova = pd.DataFrame({
    'rendement': engrais_A + engrais_B + engrais_C,
    'engrais': ['A']*6 + ['B']*6 + ['C']*6
})
model_anova = ols('rendement ~ C(engrais)', data=df_anova).fit()
anova_table = sm.stats.anova_lm(model_anova, typ=2)
print("\nTableau ANOVA:")
print(anova_table)

# Comparaisons multiples (Tukey)
from statsmodels.stats.multicomp import pairwise_tukeyhsd
tukey = pairwise_tukeyhsd(df_anova['rendement'], df_anova['engrais'],
↪  alpha=0.05)
print("\nComparaisons de Tukey:")
print(tukey)

# --- Diagnostic regression multiple ---
fig, axes = plt.subplots(1, 2, figsize=(12, 5))
residuals = model.resid
fitted = model.fittedvalues

axes[0].scatter(fitted, residuals, alpha=0.6)
axes[0].axhline(0, color='red', linestyle='--')
axes[0].set_xlabel('Valeurs ajustees')
axes[0].set_ylabel('Residus')
axes[0].set_title('Residus vs Ajustees')

stats.probplot(residuals, plot=axes[1])
axes[1].set_title('QQ-plot des residus')

plt.tight_layout()
plt.savefig("ch09_diagnostic.pdf")
plt.show()
```

9.7 Exercices

Exercice 9.1 ($\star$ – Régression multiple)**.** Ajuster un modèle de régression du prix d'une voiture en fonction de l'année, du kilométrage et de la puissance (données simulées ou

réelles). Interpréter les coefficients.

Exercice 9.2 ($\star\star$ – Démonstration)**.** Montrer que $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$ et que $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ est idempotente et symétrique.

Exercice 9.3 ($\star\star$ – ANOVA)**.** Quatre méthodes pédagogiques sont testées sur des groupes de 10 étudiants. Réaliser l'ANOVA, vérifier les hypothèses (normalité, homogénéité) et effectuer les comparaisons de Tukey.

Exercice 9.4 ($\star\star\star$ – Projet : sélection de modèle)**.** Sur un jeu de données réel, comparer les modèles par AIC, BIC et R^2_{adj}. Utiliser la validation croisée pour évaluer le pouvoir prédictif.

Formules clés

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y} \\
\mathrm{Cov}(\hat{\boldsymbol{\beta}}) &= \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} \\
F_{\text{global}} &= \frac{R^2/(p-1)}{(1-R^2)/(n-p)} \\
R^2_{\text{adj}} &= 1 - \frac{n-1}{n-p}(1-R^2) \\
F_{\text{ANOVA}} &= \frac{\text{CME}}{\text{CMR}} \sim F(k-1, n-k)
\end{aligned}$$

Chapitre 10

Introduction à la Statistique Bayésienne

« La probabilité est le langage de l'incertitude. » — Pierre-Simon de Laplace

10.1 Exemple motivant

Un médecin dispose d'un test de dépistage dont la sensibilité est 95% et la spécificité 90%. Le patient est positif. Quelle est la probabilité qu'il soit vraiment malade, sachant que la prévalence est de 1% ? Le théorème de Bayes donne une réponse surprenante : environ 8,7%, bien loin des 95% naïvement attendus.

10.2 Le paradigme bayésien

Définition 10.1 (Approche bayésienne)**.** En statistique bayésienne, le paramètre θ est traité comme une **variable aléatoire**. On lui associe :

- Une **loi a priori** $\pi(\theta)$: croyances avant observation.
- Les **données** : $\mathbf{x} = (x_1, \ldots, x_n)$, de vraisemblance $L(\theta; \mathbf{x})$.
- La **loi a posteriori** : mise à jour des croyances.

Théorème 10.2 (Théorème de Bayes)**.**

$$\pi(\theta \mid \mathbf{x}) = \frac{L(\theta; \mathbf{x}) \cdot \pi(\theta)}{\int L(\theta; \mathbf{x}) \pi(\theta)\, d\theta} \propto L(\theta; \mathbf{x}) \cdot \pi(\theta).$$

$$\boxed{\textit{Postérieur} \propto \textit{Vraisemblance} \times \textit{Prieur}}$$

Intuition statistique

L'approche bayésienne permet d'incorporer de l'information préalable et produit une distribution complète d'incertitude sur θ, pas juste une estimation ponctuelle.

10.3 Choix de la loi a priori

Définition 10.3 (Prieur conjugué)**.** La famille $\pi(\theta)$ est **conjuguée** à la vraisemblance si la loi a posteriori appartient à la même famille que le prieur.

Vraisemblance	**Prieur conjugué**	**Postérieur**
$\mathcal{B}(1,p)$	$\text{Beta}(\alpha,\beta)$	$\text{Beta}(\alpha+\sum x_i,\ \beta+n-\sum x_i)$
$\mathcal{P}(\lambda)$	$\Gamma(\alpha,\beta)$	$\Gamma(\alpha+\sum x_i,\ \beta+n)$
$\mathcal{N}(\mu,\sigma^2)$ (σ^2 connu)	$\mathcal{N}(\mu_0,\tau^2)$	$\mathcal{N}(\mu_n,\tau_n^2)$
$\mathcal{E}(\lambda)$	$\Gamma(\alpha,\beta)$	$\Gamma(\alpha+n,\ \beta+\sum x_i)$

Définition 10.4 (Prieur non informatif)**.** Un prieur « vague » qui laisse les données dominer. Exemples :

- **Prieur uniforme** : $\pi(\theta) \propto 1$ (Laplace).
- **Prieur de Jeffreys** : $\pi(\theta) \propto \sqrt{I(\theta)}$ (invariant par reparamétrisation).

10.4 Exemples fondamentaux

Exemple 10.5 (Modèle Beta-Binomial)**.** $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}(1,p)$, prieur $p \sim \text{Beta}(\alpha,\beta)$.
Postérieur : $p \mid \mathbf{x} \sim \text{Beta}(\alpha+s,\ \beta+n-s)$ où $s=\sum x_i$.
Moyenne a posteriori :

$$\mathbb{E}[p \mid \mathbf{x}] = \frac{\alpha+s}{\alpha+\beta+n} = \underbrace{\frac{n}{\alpha+\beta+n}}_{\text{poids données}} \cdot \frac{s}{n} + \underbrace{\frac{\alpha+\beta}{\alpha+\beta+n}}_{\text{poids prieur}} \cdot \frac{\alpha}{\alpha+\beta}.$$

C'est une **moyenne pondérée** entre l'EMV s/n et la moyenne a priori $\alpha/(\alpha+\beta)$.

Exemple 10.6 (Modèle Normal-Normal)**.** $X_1,\dots,X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu,\sigma^2)$ (σ^2 connu), prieur $\mu \sim \mathcal{N}(\mu_0,\tau^2)$.
Postérieur : $\mu \mid \mathbf{x} \sim \mathcal{N}(\mu_n,\tau_n^2)$ avec :

$$\mu_n = \frac{\frac{\mu_0}{\tau^2}+\frac{n\bar{x}}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}, \qquad \frac{1}{\tau_n^2} = \frac{1}{\tau^2}+\frac{n}{\sigma^2}.$$

La précision a posteriori = précision a priori + précision des données.

10.5 Estimateurs bayésiens

Définition 10.7 (Estimateurs ponctuels bayésiens)**.**

- **Moyenne a posteriori** : $\hat{\theta}_{\text{Bayes}} = \mathbb{E}[\theta \mid \mathbf{x}]$ (minimise le risque quadratique).
- **Médiane a posteriori** : minimise le risque en valeur absolue.
- **Mode a posteriori** (MAP) : $\hat{\theta}_{\text{MAP}} = \arg\max_\theta \pi(\theta \mid \mathbf{x})$.

Définition 10.8 (Intervalle de crédibilité)**.** Un **intervalle de crédibilité** à $100(1-\alpha)\%$ est $[a,b]$ tel que $\mathbb{P}(\theta \in [a,b] \mid \mathbf{x}) = 1-\alpha$. L'intervalle HPD (Highest Posterior Density) est le plus court.

Attention — Piège classique

L'intervalle de crédibilité a une interprétation directe : « la probabilité que $\theta \in [a, b]$ est $1 - \alpha$ », contrairement à l'IC fréquentiste. Mais cette interprétation repose sur le choix du prieur.

10.6 Implémentation Python

Implémentation Python

```python
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt

# --- Modele Beta-Binomial ---
# Prieur : Beta(2, 2) (leger a priori pour p ~ 0.5)
alpha_prior, beta_prior = 2, 2
n, s = 100, 68  # 68 succes sur 100

# Posterieur
alpha_post = alpha_prior + s
beta_post = beta_prior + n - s

p_grid = np.linspace(0, 1, 500)
prior = stats.beta.pdf(p_grid, alpha_prior, beta_prior)
likelihood = stats.binom.pmf(s, n, p_grid)
posterior = stats.beta.pdf(p_grid, alpha_post, beta_post)

fig, ax = plt.subplots(figsize=(10, 6))
ax.plot(p_grid, prior/prior.max(), 'b--', lw=2, label='Prieur Beta(2,2)')
ax.plot(p_grid, likelihood/likelihood.max(), 'g:', lw=2,
↪  label='Vraisemblance')
ax.plot(p_grid, posterior/posterior.max(), 'r-', lw=2, label=f'Post.
↪  Beta({alpha_post},{beta_post})')
ax.axvline(s/n, color='gray', linestyle=':', label=f'EMV = {s/n:.2f}')
ax.set_xlabel('p')
ax.set_ylabel('Densite (normalisee)')
ax.set_title('Inference bayesienne : modele Beta-Binomial')
ax.legend()
plt.savefig("ch10_beta_binomial.pdf")
plt.show()

# Estimateurs bayesiens
mean_post = alpha_post / (alpha_post + beta_post)
median_post = stats.beta.median(alpha_post, beta_post)
mode_post = (alpha_post - 1) / (alpha_post + beta_post - 2)
print(f"Moyenne posterieure : {mean_post:.4f}")
print(f"Mediane posterieure : {median_post:.4f}")
print(f"Mode (MAP)          : {mode_post:.4f}")
```

```python
print(f"EMV                    : {s/n:.4f}")

# Intervalle de credibilite 95%
ci_bayes = stats.beta.interval(0.95, alpha_post, beta_post)
print(f"IC credibilite 95%  : [{ci_bayes[0]:.4f}, {ci_bayes[1]:.4f}]")

# --- Modele Normal-Normal ---
sigma2 = 4  # variance connue
mu0, tau2 = 0, 10  # prieur N(0, 10)
n_obs = 20
x_data = np.random.normal(3, np.sqrt(sigma2), n_obs)
xbar = np.mean(x_data)

prec_prior = 1/tau2
prec_data = n_obs/sigma2
prec_post = prec_prior + prec_data
mu_post = (mu0 * prec_prior + xbar * prec_data) / prec_post
tau2_post = 1 / prec_post

print(f"\nNormal-Normal:")
print(f"Moyenne posterieure : {mu_post:.4f}")
print(f"Variance posterieure: {tau2_post:.4f}")
print(f"Xbar (EMV)          : {xbar:.4f}")
```

10.7 Exercices

Exercice 10.1 ($\star$ – Beta-Binomial)**.** Un joueur de basket marque 7 paniers sur 10 lancers. Avec un prieur Beta$(1, 1)$ (uniforme), calculer la loi a posteriori, la moyenne et un intervalle de crédibilité à 90%.

Exercice 10.2 ($\star\star$ – Normal-Normal)**.** Dériver la loi a posteriori dans le modèle Normal-Normal en complétant le carré dans l'exposant.

Exercice 10.3 ($\star\star$ – Prieur de Jeffreys)**.** Calculer le prieur de Jeffreys pour : (a) $\mathcal{B}(1, p)$, (b) $\mathcal{P}(\lambda)$, (c) $\mathcal{N}(\mu, \sigma^2)$ (σ^2 connu).

Exercice 10.4 ($\star\star\star$ – Projet)**.** Comparer l'approche fréquentiste (EMV + IC) et l'approche bayésienne (moyenne postérieure + IC de crédibilité) pour l'estimation de p dans un modèle binomial, en variant le prieur et la taille d'échantillon. Visualiser la convergence vers le même résultat quand $n \to \infty$.

Formules clés

$$\pi(\theta \mid \mathbf{x}) \propto L(\theta; \mathbf{x}) \cdot \pi(\theta)$$

$$\text{Beta-Binomial} : \text{Beta}(\alpha, \beta) \to \text{Beta}(\alpha + s, \beta + n - s)$$

$$\text{Normal-Normal} : \mu_n = \frac{\mu_0/\tau^2 + n\bar{x}/\sigma^2}{1/\tau^2 + n/\sigma^2}$$

$$\text{IC crédibilité} : \mathbb{P}(\theta \in [a, b] \mid \mathbf{x}) = 1 - \alpha$$

Chapitre 11

Méthodes Non-Paramétriques

« Ne faites pas d'hypothèses inutiles. » — Guillaume d'Ockham

11.1 Exemple motivant

Un psychologue compare les scores de bien-être de deux groupes (méditation vs. contrôle) mais les données sont ordinales (échelle de 1 à 10) et clairement non normales. Les tests t et F ne sont pas adaptés. Les méthodes non paramétriques offrent des alternatives valides sans hypothèse de normalité.

11.2 Pourquoi le non paramétrique ?

- **Données ordinales** ou de rang.
- **Petits échantillons** où la normalité ne peut être vérifiée.
- **Distributions fortement asymétriques** ou avec valeurs aberrantes.
- **Distribution inconnue** : on ne veut pas imposer de modèle paramétrique.

Attention — Piège classique

Les tests non paramétriques sont en général **moins puissants** que leurs homologues paramétriques quand les hypothèses paramétriques sont satisfaites. Mais ils sont plus **robustes**.

11.3 Test des signes

Définition 11.1 (Test des signes)**.** Pour tester H_0 : médiane $= m_0$, on compte $S^+ = \#\{i : X_i > m_0\}$. Sous H_0 : $S^+ \sim \mathcal{B}(n, 1/2)$.

11.4 Test de Wilcoxon pour un échantillon (rangs signés)

Définition 11.2 (Test de Wilcoxon signé)**.** 1. Calculer $D_i = X_i - m_0$ et éliminer les $D_i = 0$.

2. Ranger les $|D_i|$ par ordre croissant et affecter les rangs R_i.

3. $W^+ = \sum_{\{i:D_i>0\}} R_i$.

Sous H_0 : médiane $= m_0$, W^+ a une distribution connue (tables ou approximation normale pour n grand).

11.5 Test de Mann–Whitney / Wilcoxon pour deux échantillons

Définition 11.3 (Test de Mann–Whitney)**.** Pour comparer deux échantillons indépendants $X_1, \ldots, X_{n_1}$ et $Y_1, \ldots, Y_{n_2}$:

1. Combiner les deux échantillons et ranger toutes les observations.

2. $U_1 = \sum_{i=1}^{n_1} R_i - \frac{n_1(n_1+1)}{2}$, où R_i est le rang de X_i.

3. Sous $H_0 : F_X = F_Y$, U a une distribution connue.

Théorème 11.4 (Approximation normale)**.** *Pour n_1, n_2 grands :*

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

11.6 Test de Kruskal–Wallis

Définition 11.5 (Test de Kruskal–Wallis)**.** Généralisation du Mann–Whitney à k groupes (H_0 : toutes les distributions sont identiques) :

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1),$$

où R_j est la somme des rangs du groupe j. Sous H_0 : $H \sim \chi^2(k-1)$ asymptotiquement.

11.7 Test de Kolmogorov–Smirnov

Définition 11.6 (Test KS pour un échantillon)**.** Pour $H_0 : F = F_0$ (loi spécifiée) :

$$D_n = \sup_x |\hat{F}_n(x) - F_0(x)|.$$

La loi de D_n sous H_0 est tabulée (distribution de Kolmogorov).

Définition 11.7 (Test KS pour deux échantillons)**.**

$$D_{n_1,n_2} = \sup_x |\hat{F}_{n_1}(x) - \hat{G}_{n_2}(x)|.$$

Teste si les deux échantillons proviennent de la même distribution.

11.8 Test du χ^2 d'adéquation

Rappel du théorème 7.4 : ce test est déjà non paramétrique au sens où il ne suppose pas de forme paramétrique. Il compare les effectifs observés aux effectifs attendus.

11.9 Estimation non paramétrique de la densité

Définition 11.8 (Estimateur à noyau (KDE))**.**

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right),$$

où K est un **noyau** (fonction intégrant à 1, typiquement gaussien) et $h > 0$ est la **fenêtre** (bandwidth).

Intuition statistique

h petit $\rightarrow$ estimateur bruité (faible biais, forte variance). h grand $\rightarrow$ estimateur lissé (fort biais, faible variance). C'est le compromis biais-variance appliqué à l'estimation de densité.

11.10 Corrélation de Spearman

Définition 11.9 (Coefficient de Spearman)**.** Le coefficient de corrélation de Spearman r_s est le coefficient de corrélation de Pearson appliqué aux **rangs** :

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

où $d_i = \text{rang}(x_i) - \text{rang}(y_i)$. Il mesure la monotonie de la relation (pas nécessairement linéaire).

11.11 Implémentation Python

Implémentation Python

```python
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt

np.random.seed(42)

# --- Test de Wilcoxon signe ---
avant = np.array([5.2, 4.8, 6.1, 5.5, 4.9, 5.8, 6.3, 5.1, 4.7, 5.4])
apres = np.array([4.8, 4.2, 5.5, 5.0, 4.3, 5.1, 5.7, 4.5, 4.1, 4.9])
stat_w, p_w = stats.wilcoxon(avant, apres, alternative='greater')
print(f"Wilcoxon signe: W={stat_w:.1f}, p={p_w:.4f}")
```

```python
# --- Test de Mann-Whitney ---
groupe_A = np.array([23, 25, 28, 30, 27, 24, 26, 29, 31, 22])
groupe_B = np.array([18, 20, 22, 19, 21, 17, 23, 20, 19, 18])
u_stat, p_mw = stats.mannwhitneyu(groupe_A, groupe_B,
↪  alternative='two-sided')
print(f"Mann-Whitney: U={u_stat:.1f}, p={p_mw:.4f}")

# --- Test de Kruskal-Wallis ---
g1 = [6.4, 7.1, 6.8, 7.3, 6.9]
g2 = [8.2, 7.9, 8.5, 8.1, 8.4]
g3 = [5.5, 6.2, 5.8, 6.0, 5.7]
h_stat, p_kw = stats.kruskal(g1, g2, g3)
print(f"Kruskal-Wallis: H={h_stat:.4f}, p={p_kw:.4f}")

# --- Test de Kolmogorov-Smirnov ---
data = np.random.exponential(2, 50)
ks_stat, p_ks = stats.kstest(data, 'norm', args=(np.mean(data),
↪  np.std(data)))
print(f"KS (normalite): D={ks_stat:.4f}, p={p_ks:.4f}")

ks_stat2, p_ks2 = stats.kstest(data, 'expon', args=(0, np.mean(data)))
print(f"KS (exponentielle): D={ks_stat2:.4f}, p={p_ks2:.4f}")

# --- KDE ---
data_kde = np.concatenate([np.random.normal(-2, 0.8, 100),
                           np.random.normal(2, 1.2, 150)])
x_grid = np.linspace(-6, 7, 300)

fig, ax = plt.subplots(figsize=(10, 5))
ax.hist(data_kde, bins=30, density=True, alpha=0.3, label='Histogramme')
for bw in [0.2, 0.5, 1.0]:
    kde = stats.gaussian_kde(data_kde, bw_method=bw)
    ax.plot(x_grid, kde(x_grid), lw=2, label=f'KDE h={bw}')
ax.set_title("Estimation a noyau (KDE)")
ax.legend()
plt.savefig("ch11_kde.pdf")
plt.show()

# --- Correlation de Spearman ---
x = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
y = np.array([1, 4, 9, 16, 25, 36, 49, 64, 81, 100])  # y = x^2
r_pearson, _ = stats.pearsonr(x, y)
r_spearman, p_sp = stats.spearmanr(x, y)
print(f"Pearson: r={r_pearson:.4f}")
print(f"Spearman: r_s={r_spearman:.4f}, p={p_sp:.4f}")
```

11.12 Exercices

Exercice 11.1 ($\star$ – Mann–Whitney)**.** Deux algorithmes sont testés sur 8 jeux de données. Temps (en s) : A = [2.3, 3.1, 2.8, 3.5, 2.9, 3.2, 2.7, 3.0], B = [3.5, 4.2, 3.8, 4.0, 3.7, 4.1, 3.9, 3.6]. L'algorithme A est-il significativement plus rapide ?

Exercice 11.2 ($\star\star$ – Comparaison paramétrique / non paramétrique)**.** Générer 1000 échantillons de taille $n = 15$ d'une loi log-normale. Pour chaque échantillon, tester H_0 : médiane $= 1$ avec le test t et le test de Wilcoxon. Comparer les taux d'erreur de type I.

Exercice 11.3 ($\star\star$ – KDE)**.** Implémenter un estimateur à noyau gaussien. Tester sur un mélange de deux gaussiennes et étudier l'effet de h sur la qualité de l'estimation (MISE).

Exercice 11.4 ($\star\star\star$ – Projet)**.** Collecter des données réelles non normales (par ex. revenus, durées d'attente). Appliquer les méthodes non paramétriques vues dans ce chapitre. Comparer avec les résultats des tests paramétriques et discuter.

Formules clés

$$U = \sum R_i - \frac{n_1(n_1+1)}{2} \qquad H = \frac{12}{n(n+1)} \sum \frac{R_j^2}{n_j} - 3(n+1)$$

$$D_n = \sup_x |\hat{F}_n(x) - F_0(x)| \qquad r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$\hat{f}_h(x) = \frac{1}{nh} \sum K\left(\frac{x - X_i}{h}\right)$$

Formulaire récapitulatif

.1 Lois de probabilité usuelles

Loi	Paramètres	Espérance	Variance
$\mathcal{B}(n,p)$	$n \in \mathbb{N}^*,\ p \in [0,1]$	np	$np(1-p)$
$\mathcal{P}(\lambda)$	$\lambda > 0$	λ	λ
$\mathcal{U}([a,b])$	$a < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathcal{E}(\lambda)$	$\lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathcal{N}(\mu,\sigma^2)$	$\mu \in \mathbb{R},\ \sigma > 0$	μ	σ^2
$\chi^2(k)$	$k \in \mathbb{N}^*$	k	$2k$
$t(k)$ (Student)	$k \in \mathbb{N}^*$	$0\ (k > 1)$	$\frac{k}{k-2}\ (k > 2)$
$F(d_1,d_2)$ (Fisher)	$d_1, d_2 \in \mathbb{N}^*$	$\frac{d_2}{d_2-2}$	—
$\Gamma(\alpha,\beta)$	$\alpha, \beta > 0$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$

.2 Formules d'estimation

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1}$$

$$\hat{\theta}_{\text{EMV}} = \arg\max_{\theta} \sum_{i=1}^{n} \ln f(x_i;\theta) \qquad \hat{\theta}_{\text{moments}} : \mathbb{E}_\theta[X^k] = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{2}$$

.3 Intervalles de confiance

$$\text{Moyenne } (\sigma \text{ connu}) : \bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{3}$$

$$\text{Moyenne } (\sigma \text{ inconnu}) : \bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \tag{4}$$

$$\text{Proportion} : \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{5}$$

$$\text{Variance} : \left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}},\ \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right] \tag{6}$$

.4 Statistiques de test

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1) \tag{7}$$

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \tag{8}$$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \tag{9}$$

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1) \tag{10}$$

Tables statistiques

Les tables de la loi normale, de Student, du χ^2 et de Fisher sont disponibles en annexe numérique. En pratique, on utilise Python :

```
from scipy import stats
# Quantiles
stats.norm.ppf(0.975)        # z_{0.025} = 1.96
stats.t.ppf(0.975, df=10)    # t_{0.025,10}
stats.chi2.ppf(0.95, df=5)   # chi2_{0.05,5}
stats.f.ppf(0.95, 3, 20)     # F_{0.05,3,20}
```

Bibliographie

[1] Saporta, G. (2011). *Probabilités, Analyse des Données et Statistique*. Technip, 3e édition.

[2] Casella, G. & Berger, R.L. (2002). *Statistical Inference*. 2e édition, Cengage Learning.

[3] Wasserman, L. (2004). *All of Statistics : A Concise Course in Statistical Inference*. Springer.

[4] Wackerly, D., Mendenhall, W. & Scheaffer, R. (2008). *Mathematical Statistics with Applications*. 7e édition, Cengage.