Statistiques : compléments théoriques

Juste Raimbault [1]

2025-2026

[1]LaSTIG, IGN-ENSG-UGE

ENSG Géomatique
ÉCOLE NATIONALE DES SCIENCES GÉOGRAPHIQUES

Contenu du cours

Estimateur statistique

Propriétés asymptotiques des estimateurs

Décomposition biais-variance

Intervalles de confiance et bootstrap

Estimateur statistique

Définition d'un estimateur

- **Paramètre ou propriété de la population (θ):** caractéristique fixe et inconnue de la population (représentée par une variable aléatoire X avec loi sous-jacente): paramètre de la loi μ, σ^2, β, moments, etc.
- **Echantillon:** observations $(X_1, \ldots, X_n)$ indépendantes et identiquement distribuées.

Definition

Un **estimateur** ($\hat{\Theta}$) est une fonction mesurable des observations X_i, utilisée pour calculer une approximation de θ.

$$\hat{\Theta} = g(X_1, X_2, \ldots, X_n)$$

$\rightarrow$ l'échantillonnage X_i est aléatoire, donc $\hat{\Theta}$ est une **Variable Aléatoire**.

Estimateur vs valeur estimée

Estimateur ($\hat{\Theta}$)

- La **formule** de calcul
- Une **Variable aléatoire** (avec une distribution d'échantillonnage)
- **Exemple:** $\bar{X} = \frac{1}{n} \sum X_i$

Estimate ($\hat{\theta}$)

- La **valeur** spécifique calculée sur un échantillon observé
- **Nombre fixe**
- Tests et intervalles de confiance associés

Propriétés d'un estimateur

Un estimateur $\hat{\Theta}$ est évalué sur les propriétés de sa distribution d'échantillonnage par rapport au vrai paramètre θ.

- **Biais :** estimation correcte en moyenne ?
  - $\text{Bias}(\hat{\Theta}) = \mathbb{E}[\hat{\Theta}] - \theta$
  - Non-biaisé si $\mathbb{E}[\hat{\Theta}] = \theta$.
- **Efficacité (Variance) :** quelles fluctuations entre différents échantillonnages ?
  - $\mathrm{Var}(\hat{\Theta}) = \mathbb{E}\left[(\hat{\Theta} - \mathbb{E}[\hat{\Theta}])^2\right]$
  - Une variance faible est préférée.
- **Convergence :** comportement en fonction de la taille de l'échantillon n (cf. loi des grands nombres et Théorème Central Limite)
  - $\hat{\Theta} \xrightarrow{p} \theta \quad$ avec $n \to \infty$

Biais de l'Estimateur de la Variance

- **Hypothèses:** Échantillon i.i.d. $X_1, \ldots, X_n$ avec $\mathbb{E}[X] = \mu$ et $\mathrm{Var}(X) = \sigma^2$.

Estimateur Biaisé

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

$$\mathbb{E}[S_b^2] = \frac{n-1}{n} \sigma^2$$

$\rightarrow$ sous-estimation de σ^2

Estimateur Non Biaisé (correction de Bessel)

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

Propriétés asymptotiques des estimateurs

Loi forte des Grands Nombres

Soit $(X_i)_{i \geq 1}$ des Variables Aléatoires indépendantes et identiquement distribuées (i.i.d.) telles que $\mathbb{E}[X_i] = \mu < \infty$.

La moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ converge **presque sûrement** vers μ :

$$\mathbb{P}\left(\lim_{n \to \infty} \bar{X}_n = \mu\right) = 1$$

Loi Faible des Grands Nombres (LFGN)

- **Hypothèses:** $X_1, \ldots, X_n$ i.i.d. avec $\mathbb{E}[X_i] = \mu$ et **variance finie** $\mathrm{Var}(X_i) = \sigma^2 < \infty$.
- Soit $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ la moyenne d'échantillon.

Loi Faible des Grands Nombres

La moyenne d'échantillon $\bar{X}_n$ converge en probabilité vers l'espérance μ.

$$\forall \epsilon > 0, \quad \lim_{n\to\infty} P\left(|\bar{X}_n - \mu| > \epsilon\right) = 0$$

Remarque : convergence p.s. $\implies$ convergence en probabilité

Démonstration de la LFGN

Inégalité de Tchebychev

Pour toute V.A. Z avec $\mathbb{E}[Z] = \mu_Z$ et $\mathrm{Var}(Z) = \sigma_Z^2$ finie, $\forall \epsilon > 0$:

$$P(|Z - \mu_Z| \geq \epsilon) \leq \frac{\mathrm{Var}(Z)}{\epsilon^2}$$

$\rightarrow$ application à $Z = \bar{X}_n$: comme $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2/n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \xrightarrow[n\to\infty]{} 0$$

d'où la LFGN par le théorème des gendarmes.

- **Principe:** La distribution de la moyenne de V.A. i.i.d. de variance finie σ^2 tend vers une **Loi Normale**, quelles que soient leurs distributions initiales.

Théorème de la Limite Centrale

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1) \quad \text{quand } n \to \infty$$

$\rightarrow$ permet l'inférence statistique (tests d'hypothèses, intervalles de confiance) quelle que soit la distribution des données d'origine.

Moment d'ordre k : $m_k = \mathbb{E}[X^k]$ estimé par $\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

Convergence des Moments (conséquence de la LFGN)

Si m_k existe et est fini, alors son estimateur empirique $\hat{m}_k$ converge en probabilité vers lui :

$$\hat{m}_k \xrightarrow[n\to\infty]{P} m_k$$

- TCL approximation d'ordre 1
- termes correctifs basés sur les moments d'ordre supérieur

Formule de l'Expansion (Ordre 1)

La densité de Z_n est mieux approximée par :

$$f_{Z_n}(z) \approx \phi(z) \left[1 + \frac{\gamma_3}{6\sqrt{n}} H_3(z) + \frac{\gamma_4}{24n} H_4(z) + o\left(\frac{1}{n}\right) \right]$$

- $\phi(z)$: densité de la $\mathcal{N}(0,1)$
- γ_3 : **Skewness** (asymétrie) de la population
- γ_4 : **Kurtosis** (excès d'aplatissement) de la population
- $H_k(z)$: Polynômes d'Hermite

Théorème Central Limite Généralisé (TCLG)

$\rightarrow$ cas avec variance infinie ? (ex: Loi de Pareto avec $\alpha \leq 2$, Loi de Cauchy)

Théorème (TCLG)

La limite d'une somme de variables i.i.d. (correctement normalisée) est une **Distribution Stable** (ou “Loi Stable de Lévy”) avec un indice de stabilité $\alpha \in (0, 2]$

- **Cas 1:** $\alpha = 2 \implies$ **Loi Normale**. On retrouve le TCL standard (le seul cas où la variance est finie).
- **Cas 2:** $\alpha < 2 \implies$ Variance infinie, queues lourdes.
- **Cas 3:** $\alpha = 1 \implies$ **Loi de Cauchy** (avec $\beta = 0$).

Décomposition biais-variance

Erreur de prédiction d'un modèle

Contexte : minimisation de la MSE du modèle $\hat{f}(x)$ cherchant à approximer Y.

Décomposition de la MSE

$$\mathbb{E}\left[\left(Y - \hat{f}(x)\right)^2\right] = \textbf{Biais}^2 + \textbf{Variance} + \textbf{Bruit intrinsèque}$$

$\rightarrow$ propriété sous-jacente lors de la selection de modèle et le contrôle du sur-apprentissage

Décomposition biais-variance

En supposant $Y = f(x) + \epsilon$ avec $\mathbb{E}[\epsilon] = 0$ et $\mathrm{Var}(\epsilon) = \sigma^2$, on obtient la décomposition :

$$\mathbb{E}\left[\left(Y - \hat{f}(x)\right)^2\right] = \underbrace{\left(\mathbb{E}[\hat{f}(x)] - f(x)\right)^2}_{\textbf{Biais}^2} + \underbrace{\mathbb{E}\left[\left(\hat{f}(x) - \mathbb{E}[\hat{f}(x)]\right)^2\right]}_{\textbf{Variance}} + \underbrace{\mathbb{E}\left[\epsilon^2\right]}_{\textbf{Bruit intrinsèque}}$$

En pratique, $\mathbb{E}[\hat{f}(x)]$ correspond à l'erreur de prédiction moyenne sur l'ensemble des jeux de données possibles.

Compromis entre biais et variance

1. **Bias (Bias^2):** erreur systématique
2. **Variance ($\mathrm{Var}(\hat{f}(x))$):** variabilité au regard des données d'entrainement
3. **Erreur intrinsèque (σ^2):** bruit non capturé par le modèle

Compromis :

- **Haute complexité** $\implies$ faible biais, forte variance (Overfitting)
- **Basse complexité** $\implies$ fort biais, faible variance (Underfitting)

Intervalles de confiance et bootstrap

Approche par bootstrap

$\rightarrow$ approches des intervalles de confiance font des hypothèses sur la distribution

Definition

La méthode **Bootstrap** est une méthode computationnelle non-paramétrique qui utilise B nouveaux échantillons $(\mathcal{D}^*)$ par **rééchantillonnage avec remise** des données initiales $\mathcal{D}$.

$\rightarrow$ Pour chaque nouvel échantillon $\mathcal{D}_b^*$, calcul de la statistique $\hat{\theta}_b^*$ et utilisation de sa distribution empirique

Justification théorique

Théorème de Glivenko-Cantelli : le ré-échantillonnage préserve les propriétés asymptotiques de l'estimateur

1. **Statistique:** calcul de $\hat{\theta}$ avec $\mathcal{D}$
2. **Echantillon de bootstrap :** tirage de B échantillons $\mathcal{D}_b^*$
3. **Statistiques de bootstrap empiriques:** calcul de chaque $\hat{\theta}_b^*$
4. **IC :** distribution empirique de $\{\hat{\theta}_b^*\}$ donne l'IC, avec par exemple les quantiles pour un IC à $100(1-\alpha)\%$:
   - tri des réplications $\hat{\theta}_{(1)}^* \leq \cdots \leq \hat{\theta}_{(B)}^*$.
   - IC donné par :

$$\mathrm{IC}_{1-\alpha} = \left[\mathrm{Percentile}_{\alpha/2}, \mathrm{Percentile}_{1-\alpha/2}\right]$$