Statistique Inférentielle - Résumé Complet

I. ÉCHANTILLONNAGE ET MODÉLISATION

Échantillon aléatoire : $(X_1, \ldots, X_n)$ v.a. i.i.d. de même loi que X

Modèle statistique : famille de lois $(P_\theta)_{\theta\in\Theta}$

Modèle paramétrique: $\Theta \subset \mathbb{R}^d$, identifiable si $\theta \mapsto P_\theta$ injective.

n-échantillon: $X = (X_1, \ldots, X_n)$ i.i.d. de loi p_θ. Modèle: $(\Omega = \Psi^n, \mathcal{A} = \mathcal{B}^{\otimes n}, P_\theta = (p_\theta)^{\otimes n})$.

Démarche statistique :

1. **Modélisation** : $X \sim P_\theta$ avec θ inconnu
2. **Inférence** : estimer θ à partir des observations

II. MODES DE CONVERGENCE

Convergence en probabilité : $X_n \xrightarrow{P} X$

$$\forall\varepsilon > 0, \quad P(|X_n - X| \geq \varepsilon) \xrightarrow{n\to\infty} 0$$

Convergence presque sûre : $X_n \xrightarrow{p.s.} X$

$$P\left(\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = 1$$

Convergence en loi : $X_n \xrightarrow{\mathcal{L}} X$

$$\forall\varphi \text{ continue bornée}, \quad E[\varphi(X_n)] \to E[\varphi(X)]$$

Convergence L^2 : $X_n \xrightarrow{\|\cdot\|_2} X$

$$E[|X_n - X|^2] \to 0$$

0.1 Hiérarchie des convergences

$$\text{p.s.} \Rightarrow \text{proba} \Rightarrow \text{loi}$$

$$L^2 \Rightarrow \text{proba}$$

Réciproque fausse sauf si limite = constante

0.2 Critères pour convergence en loi

- **Paul Lévy** : $\Phi_{X_n}(t) \to \Phi_X(t)\ \forall t$
- **Fonctions répartition** : $F_{X_n}(x) \to F_X(x)$ aux points de continuité

III. INÉGALITÉS ET THÉORÈMES ASYMPTOTIQUES

0.3 Inégalité de Markov

$$P(|X| \geq c) \leq \frac{E[|X|^p]}{c^p}$$

0.4 Inégalité de Bienaymé-Tchebychev

$$P(|X - E[X]| \geq c) \leq \frac{\text{Var}(X)}{c^2}$$

0.5 Inégalité de Hoeffding

$X_1, \ldots, X_n$ indép., $a_i \leq X_i \leq b_i$, $S_n = \sum X_i$

$$P(|S_n - E[S_n]| \geq c) \leq 2\exp\left(-\frac{2c^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Hoeffding : décroissance exponentielle vs polynomiale (Markov)

Application Borel-Cantelli : Si $\sum P(|X_n - X| \geq \varepsilon) < \infty$ alors $X_n \xrightarrow{p.s.} X$

0.6 Loi des Grands Nombres

$X_1, \ldots, X_n$ i.i.d., $E[X_1] = m < \infty$

$$\frac{S_n}{n} \xrightarrow{p.s.} m$$

0.7 Théorème Central Limite

$X_1, \ldots, X_n$ i.i.d., $\text{Var}(X_1) = \sigma^2 < \infty$

$$\sqrt{n}\left(\frac{S_n}{n} - m\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

TCL : base des intervalles de confiance asymptotiques

IV. OPÉRATIONS SUR LES LIMITES

0.8 Théorème de continuité

Si $X_n \to X$ et $P(X \in D_g) = 0$ (D_g = points de discontinuité)

$$g(X_n) \text{ hérite du mode de convergence}$$

0.9 Théorème de Slutsky

$X_n \xrightarrow{\mathcal{L}} X$, $Y_n \xrightarrow{P} a$ (constante)

$$X_n + Y_n \xrightarrow{\mathcal{L}} X + a$$

$$X_n Y_n \xrightarrow{\mathcal{L}} aX$$

0.10 Delta méthode

$v_n(X_n - a) \xrightarrow{\mathcal{L}} X$, g dérivable en a

$$v_n(g(X_n) - g(a)) \xrightarrow{\mathcal{L}} g'(a)X$$

Utilité : propager incertitude par transformations

V. VECTEURS GAUSSIENS

0.11 Définition

$X = (X_1, \dots, X_d)'$ gaussien si $\forall u \in \mathbb{R}^d$, $u'X \sim \mathcal{N}$

0.12 Propriétés fondamentales

- Caractérisé par $\mu = E[X]$ et $\Sigma = \mathrm{Var}(X)$
- Notation : $X \sim \mathcal{N}_d(\mu, \Sigma)$
- Fonction caractéristique :

$$\Phi(t) = \exp(it'\mu - \frac{1}{2}t'\Sigma t)$$

0.13 Propriétés importantes

- **Indépendance** : composantes indép. $\Leftrightarrow$ Σ diagonale
- **Transformation linéaire** : $A + BX \sim \mathcal{N}_q(A + B\mu, B\Sigma B')$
- **Densité** (si $\det \Sigma \neq 0$) :

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

0.14 Loi du χ^2

$X_1, \dots, X_k$ i.i.d. $\mathcal{N}(\mu_i, 1)$

$$\|X\|^2 \sim \chi^2(k, \|\mu\|^2)$$

Si $\mu = 0$: $\|X\|^2 \sim \chi^2(k)$

0.15 Théorème de Cochran

$X \sim \mathcal{N}_d(\mu, I_d)$, $E_1, \dots, E_p$ sous-espaces orthogonaux

- $\Pi_{E_i}(X)$ gaussiens indépendants
- $\|\Pi_{E_i}(X)\|^2 \sim \chi^2(r_i, \|\Pi_{E_i}\mu\|^2)$

Application : construction de statistiques en cadre gaussien

VI. ESTIMATION

Estimateur: Statistique $T = h(X)$ pour estimer $g(\theta)$.
Biais: $b_\theta(\hat{g}) = \|g(\theta) - \mathbb{E}_\theta \hat{g}\|$. Sans biais si $\mathbb{E}_\theta \hat{g} = g(\theta)$.
Risque quadratique: $R_\theta(\hat{g}) = \mathbb{E}_\theta[\|\hat{g} - g(\theta)\|^2]$.
Décomposition biais-variance:

$$R_\theta(\hat{g}) = \mathrm{Var}(\hat{g}) + b_\theta(\hat{g})^2$$

Convergence:

- Asympt. sans biais: $\mathbb{E}_\theta \hat{g}_n \to g(\theta)$
- Convergence en moyenne quad.: $R_\theta(\hat{g}_n) \to 0$
- Consistant: $\hat{g}_n \xrightarrow{P_\theta} g(\theta)$
- Fortement consistant: $\hat{g}_n \xrightarrow{p.s.} g(\theta)$

Vitesse: $v_n(\hat{g}_n - g(\theta)) \xrightarrow{\mathcal{L}} \ell(\theta)$ non dégénérée. Si $\ell = \mathcal{N}$ et $v_n = \sqrt{n}$: asympt. normal.

0.16 Méthode des Moments

Estimer $\phi(\theta) = \mathbb{E}_\theta[h(X_1)]$ par $\hat{\phi} = \frac{1}{n}\sum h(X_i)$. Inverser si possible.

0.17 Maximum de Vraisemblance

Fonction de vraisemblance: $L_n(\theta) = \prod_{i=1}^n s_\theta(X_i)$, $\ell_n(\theta) = \ln L_n(\theta)$.
EMV: $\hat{\theta} \in \arg\max_{\theta \in \Theta} L_n(\theta)$.
Score: $\ell_n'(X, \theta) = \frac{\partial}{\partial \theta} \ln L_n(X, \theta)$.
Information de Fisher:

$$I_n(\theta) = \mathbb{E}_\theta[(\ell_n'(X, \theta))^2] = -\mathbb{E}_\theta[\ell_n''(X, \theta)]$$

Efficacité asymptotique: Sous régularité, si $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(\theta))$, borne inf: $\sigma^2(\theta) \geq 1/I_1(\theta)$. EMV asympt. efficace si égalité.

VII. ESTIMATION GAUSSIENNE

Modèle: $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$.
Estimateurs:

$$\bar{X}_n = \frac{1}{n}\sum X_i, \quad S_n^2 = \frac{1}{n-1}\sum (X_i - \bar{X}_n)^2$$

Fisher-Cochran:

1. $\bar{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$
2. $S_n^2 \sim \frac{\sigma^2}{n-1}\chi^2(n-1)$
3. $\bar{X}_n$ et S_n^2 indépendantes
4. $\sqrt{n}\frac{\bar{X}_n - \mu}{S_n} \sim \tau(n-1)$

VIII. INTERVALLES DE CONFIANCE

IC de niveau $1-\alpha$: $\hat{C}$ tel que $P_\theta(\theta \in \hat{C}) \geq 1 - \alpha$ $\forall \theta$.

0.18 Cas Gaussien

IC pour μ (σ inconnu):

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

où $t_{n-1,1-\alpha/2}$: quantile Student.

IC pour σ^2:

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

IC asymptotique pour μ (var. finie):

$$\left[\bar{X}_n \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}_n^2}{n}}\right]$$

Réalisation: L'IC est aléatoire, sa réalisation est fixe.

IX. TESTS D'HYPOTHÈSES

Test: $\Phi : X \mapsto \{0, 1\}$, $\Phi = 1 =$ rejet H_0.
Hypothèses: $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$.
Erreurs:

- 1ère espèce: $e_{pr} = \sup_{\theta \in \Theta_0} P_\theta(\text{rejeter } H_0)$
- 2ème espèce: $e_{sc} = \sup_{\theta \in \Theta_1} P_\theta(\text{accepter } H_0)$

Niveau: Test de niveau α si $e_{pr} \leq \alpha$.
Puissance: $\pi(\theta) = P_\theta(\text{rejeter } H_0)$.
Uniformément plus puissant: $\pi(\theta) \geq \pi'(\theta)\ \forall \theta \in \Theta_1$.

0.19 Tests Gaussiens

Test de Student (μ):

$$T_n = \sqrt{n}\frac{\bar{X}_n - \mu_0}{S_n} \sim \tau(n-1) \text{ sous } H_0$$

Rejet si $|T_n| > t_{n-1,1-\alpha/2}$ (bilatère) ou $T_n > t_{n-1,1-\alpha}$ (unilatère).

Test variance:

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

Test 2 échantillons:

- Comparaison variances: $F = S_X^2/S_Y^2 \sim \mathcal{F}(n_x-1, n_y-1)$ sous H_0
- Comparaison moyennes ($\sigma_x = \sigma_y$):

$$T = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{1/n_x + 1/n_y}} \sim \tau(n_x + n_y - 2)$$

Test proportion:

$$U_n = \sqrt{n}\frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Test de Wald: Si $\sqrt{n}(\hat{\theta} - \theta)/S \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$, rejet si $|W| > z_{1-\alpha/2}$.

0.20 p-value

p-value: Plus petit α pour lequel on rejette H_0.
Interprétation:

- p-value $< \alpha$: rejet H_0
- p-value grande: pas de preuve contre H_0
- Résultat significatif si p-value très faible

X. RÉGRESSION LINÉAIRE

0.21 Modèle Linéaire

Modèle de régression simple:

$$Y_i = ax_i + b + \varepsilon_i, \quad i = 1, \ldots, n$$

avec ε_i i.i.d., $\mathbb{E}\varepsilon_i = 0$, $\text{Var}\varepsilon_i = \sigma^2$.
Modèle de régression multiple:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

où $\mathbf{X}_{n\times p}$ matrice du design, $\boldsymbol{\theta} \in \mathbb{R}^p$.
Hypothèses du Modèle:

- A1: $\mathbb{E}\boldsymbol{\varepsilon} = 0$ (erreurs centrées)
- A2: $\text{Var}\varepsilon_i = \sigma^2$ (homoscédasticité)
- A3: ε_i indépendantes
- A4: $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ (normalité)

Modèle régulier: $\mathbf{X}$ injective ($rg(\mathbf{X}) = p$).
Modèle singulier: $\mathbf{X}$ non injective.

0.22 Estimation par Moindres Carrés

Estimateur MC: Minimise $\|\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}\|^2$
Équations normales:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\theta} = \mathbf{X}'\mathbf{Y}$$

Cas régulier:

$$\hat{\boldsymbol{\theta}}_{MC} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

Propriétés:

- Sans biais: $\mathbb{E}\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$
- Variance: $\text{Var}\hat{\boldsymbol{\theta}} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
- Optimal (Gauss-Markov): Variance minimale parmi estimateurs linéaires sans biais

0.23 Valeurs Prédites et Résidus

Valeurs prédites: $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\theta}} = P_{\text{Im}(\mathbf{X})}\mathbf{Y}$
Résidus: $\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = P_{\text{Im}(\mathbf{X})^\perp}\mathbf{Y}$
Somme des carrés résiduels:

$$SC_{res} = \|\hat{\boldsymbol{\varepsilon}}\|^2 = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$$

Estimateur de σ^2:

$$S^2 = \frac{SC_{res}}{n-p} \quad \text{(sans biais)}$$

0.24 Estimateur du Maximum de Vraisemblance

Sous A1-A4 (normalité):

- $\hat{\boldsymbol{\theta}}_{MV} = \hat{\boldsymbol{\theta}}_{MC}$
- $\hat{\sigma}^2_{MV} = \frac{SC_{res}}{n}$ (biaisé)

0.25 Théorème de Fisher-Cochran (Régression)

Sous A1-A4:

1. $\hat{\mathbf{Y}} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 P_{\mathrm{Im}(\mathbf{X})})$
2. $\frac{SC_{res}}{\sigma^2} \sim \chi^2(n-p)$
3. $\hat{\mathbf{Y}}$ et S^2 indépendants
4. Si modèle régulier: $\hat{\boldsymbol{\theta}} \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$

0.26 Régression Simple

Estimateurs:

$$\hat{a} = \frac{\mathrm{Cov}(x,Y)}{\mathrm{Var}(x)}, \quad \hat{b} = \bar{Y} - \hat{a}\bar{x}$$

Matrice de covariance:

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n\mathrm{Var}(x)} \begin{bmatrix} \bar{x}^2 + \mathrm{Var}(x) & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}$$

0.27 Intervalles de Confiance

Pour θ_j:

$$\hat{\theta}_j \pm S\sqrt{w_{jj}}\, t_{n-p,1-\alpha/2}$$

où $w_{jj} = [(\mathbf{X}'\mathbf{X})^{-1}]_{jj}$

Pour $\mathbb{E}(Y_x)$ (nouvelle observation):

$$\hat{Y}_x \pm S\sqrt{\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}\, t_{n-p,1-\alpha/2}$$

Intervalle de prédiction:

$$\hat{Y}_x \pm S\sqrt{1 + \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}\, t_{n-p,1-\alpha/2}$$

0.28 Tests d'Hypothèses

Test de Student pour θ_j:

$$T_j = \frac{\hat{\theta}_j}{S\sqrt{w_{jj}}} \sim \tau(n-p) \text{ sous } H_0 : \theta_j = 0$$

Règle de décision: Rejet si $|T_j| > t_{n-p,1-\alpha/2}$

F-test pour modèles emboîtés: **Sous-modèle**: $\mathrm{Im}(\mathbf{X}_\omega) \subset \mathrm{Im}(\mathbf{X}_\Omega)$, $rg(\mathbf{X}_\omega) = q < p$

$$F_{\omega|\Omega} = \frac{(SC_\omega - SC_\Omega)/(p-q)}{SC_\Omega/(n-p)} \sim \mathcal{F}(p-q, n-p) \text{ sous } H_0$$

Règle de décision: Rejet si $F_{\omega|\Omega} > f_{p-q,n-p,1-\alpha}$

0.29 Régions de Confiance

Ellipsoïde de confiance:

$$\{\mathbf{u} \in \mathbb{R}^p : (\hat{\boldsymbol{\theta}} - \mathbf{u})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\theta}} - \mathbf{u}) \leq pS^2 f_{p,n-p,1-\alpha}\}$$

Méthode de Bonferroni:

$$\hat{I}_0 \times \cdots \times \hat{I}_{p-1} \text{ avec } 1 - \alpha/p \text{ par paramètre}$$

0.30 Robustesse

Sans hypothèse A4 (normalité): Propriétés conservées:

- $\hat{\boldsymbol{\theta}}_{MC}$ sans biais
- $\mathrm{Var}\hat{\boldsymbol{\theta}}_{MC} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$
- Optimalité Gauss-Markov

Propriétés perdues:

- Lois exactes (χ^2, Student, Fisher)
- Tests exacts de niveau α

Comportement asymptotique: Sous A1-A3, si $\frac{1}{n}\mathbf{X}'\mathbf{X} \to Q$ définie positive:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, \sigma^2 Q^{-1})$$

Tests asymptotiques: F-test valide pour grands échantillons

0.31 ANOVA à un facteur

Modèle: $Y_{ij} = \mu_i + \varepsilon_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n_i$

Écriture matricielle:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

avec $\mathbf{X}$ matrice indicatrices des groupes