MA101
Statistique

Maxime Ossonce

ENSTA – 1A

2018/2019

Outline

Bibliographie

- J. Pagès. Statistique générales pour utilisateurs. Presses Universitaires de Rennes, 2005.
- A. Monfort. Cours de statistique mathématique. Economica, 1997.
- S. Morgenthaler. Introduction à la statistique. PPUR, 2013.

Inférence statistique

Le terme **statistique** a plusieurs acceptions :

- observation et description d'un échantillon (statistique descriptive) ;
- recueil et interprétation des données ;
- discipline mathématique (**inférence statistique**) ;
- objet mathématique servant à l'estimation.

Inférence statistique

Les caractéristiques (**variables**) de la populations sont régies par un modèle (*cf.* cours de probabilités) partiellement connu.

- Parmi les lois de probabilités *possibles* appartenant à $\mathcal{P}$, laquelle régit la population ?
- Exemple : sondage pour prédire le résultat d'une élection.
- Estimation paramétrique, pour $\Theta \subset \mathbb{R}^p$: $\mathcal{P} = (\mathbb{P}_\theta)_{\theta \in \Theta}$.
  - Exemple : $\Theta = \mathbb{R} \times \mathbb{R}_+$ et $\mathcal{P}$ est l'ensemble de lois à densité gaussienne sur $\mathbb{R}$.

Population

Une **population** est un ensemble d'**individus** sur lequel on observe les **variables**.

- Si la population est finie de taille N, on effectue soit un recensement, soit un **sondage**, avec ou sans remise.
- L'**échantillonnage** est le procédé d'extraction d'un sous-ensemble de la population.
- Quand $N \gg 1$, l'échantillonnage avec ou sans remise sont équivalents.

Echantillon

Un ***n*-échantillon** est un vecteur $X = (X_1, \dots, X_n)$ de n variables aléatoires (v.a.) indépendantes et identiquement distribuées (i.i.d.)

- La loi $\mathbb{P}$ de X_1 est appelée loi **parente** de l'échantillon.
- $(X_1, \dots, X_n) \sim \mathbb{P}^{\otimes n}$.
- Une **série statistique** est une réalisation de X :

$$(x_1, \dots, x_n) = (X_1(\omega), \dots, X_n(\omega)).$$

- Le **modèle statistique paramétrique**, noté $\left(\mathbb{R}, \mathcal{R}, (\mathbb{P}_\theta)_{\theta\in\Theta}\right)^n$ est

$$\left(\mathbb{R}^n, \mathcal{R}^n, \left(\mathbb{P}_\theta^{\otimes n}\right)_{\theta\in\Theta}\right).$$

Statistique

Etant donné un n-échantillon, une **statistique** T_n est une fonction mesurable de $(\mathbb{R}^n, \mathcal{R}^n)$ dans $(\mathbb{R}^k, \mathcal{R}^k)$.

- T_n est une v.a. sur $(\mathbb{R}^n, \mathcal{R}^n)$.
- $T_n(X_1, \dots, X_n)$ est une v.a. sur $(\Omega, \mathcal{F})$.
- Exemple, la moyenne empirique de l'échantillon :

$$\begin{aligned} T_n \colon (\mathbb{R}^n, \mathcal{R}^n) &\to (\mathbb{R}, \mathcal{R}) \\ (x_1, \dots, x_n) &\to \overline{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k. \end{aligned}$$

- $\overline{X}_n$ est une v.a.

Estimation

Un estimateur sera une statistique dont l'objet sera de se faire une idée de la loi parente $\mathbb{P}$ élément de $\mathcal{P}$.

- Si le modèle est paramétrique, l'estimation sera dite paramétrique.
- Si une valeur estimée de θ est souhaitée, on parlera d'estimation ponctuelle (par exemple, par maximum de vraisemblance).
- Si un ensemble de valeurs est souhaitée, on aura affaire à une estimation par **intervalle de confiance**.
- Les tests d'hypothèse permettent de répondre à la question : la *vraie* valeur de θ est-elle dans l'ensemble $\Theta_0 \subset \Theta$ ou dans $\Theta \backslash \Theta_0$?

Convergenge presque sûre

On dit que la suite de variable aléatoire réelle (v.a.r.) (X_n) converge presque sûrement vers une v.a.r. X, et on note $X_n \xrightarrow{\text{p.s.}} X$, si il existe $N \in \mathcal{F}$ négligeable tel que

$$X_n(\omega) \to X(\omega) \ ; \quad \forall \omega \in N^{\mathsf{c}}.$$

Convergence en probabilité

On dit que la suite de v.a.r. (X_n) converge en probabilité vers une v.a.r. X, et on note $X_n \xrightarrow{\mathbb{P}} X$, si on a

$$\lim_n \mathbb{P}(|X_n - X| > \epsilon) = 0 \ ; \quad \forall \epsilon > 0.$$

- La convergence en probabilité est plus *faible* que la convergence presque sûrement (p.s.) :

$$X_n \xrightarrow{\text{p.s.}} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$$

- La réciproque est fausse.

Convergence dans $\mathcal{L}^1$

On dit que la suite de v.a.r. **intégrables** (X_n) converge dans $\mathcal{L}^1$ vers une v.a.r. X intégrable, et on note $X_n \xrightarrow{\mathcal{L}^1} X$, si

$$\mathbb{E}[|X_n - X|] \to 0.$$

- La convergence en probabilité est plus *faible* que la convergence dans $\mathcal{L}^1$:
$$X_n \xrightarrow{\mathcal{L}^1} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X.$$
- Si on a $X_n \xrightarrow{\text{p.s.}} X$ et $|X_n| \leq g$ pour tout n avec g intégrable alors X_n est intégrable pour tout n et $X_n \xrightarrow{\mathcal{L}^1} X$.

Convergence en loi

On dit que la suite de v.a.r. (X_n) converge en loi vers une v.a.r. X, et on note $X_n \xrightarrow{\mathscr{L}} X$, si pour toute fonction **continue bornée** φ à valeurs dans $\mathbb{R}$ on a :

$$\lim \mathbb{E}[\varphi(X_n)] = \mathbb{E}[\varphi(X)].$$

- Pour une suite de v.a.r., la convergence en loi est équivalente à la convergence ponctuelle des fonctions de répartition F_{X_n} vers F_X aux points de continuité de F_X.
- La v.a. X peut ne pas être définie sur le même espace probabilisé que (X_n).
- La convergence en loi est plus *faible* que la convergence en probabilité :

$$X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\mathscr{L}} X.$$

Loi faible des grands nombres

La loi des grands nombres (LGN) sera utilisée pour construire des estimateurs.

Théorème 1.1

Si (X_n) est une suite de v.a.r. i.i.d. d'espérance μ, alors la variable aléatoire $\overline{X}_n = \frac{X_1+\cdots+X_n}{n}$ converge en probabilité vers μ :

$$\mathbb{P}\left(\left|\overline{X}_n - \mu\right| > \epsilon\right) \xrightarrow[n\to\infty]{} 0 \ ; \quad \forall \epsilon > 0.$$

TCL

Le théorème de la limite centrale (TCL) permet contrôler la rapidité de convergence dans la loi faible des grands nombres.

Théorème 1.2 (TCL)

Si (X_n) est une suite de v.a.r. i.i.d. d'espérance μ et de variance σ^2 alors la variable aléatoire

$$Z_n = \frac{\sqrt{n}}{\sigma} \cdot (\overline{X}_n - \mu)$$

converge en ***loi*** *vers $Z \sim \mathcal{N}(0,1)$.*

- ▷ On a donc $\lim_{n\to\infty} \mathbb{P}(Z_n \leq z) = \Phi(z)$.

Lemme de Slutsky

X_n et Y_n sont des suites de v.a. à valeurs respectivement dans $\mathbb{R}^n$ et $\mathbb{R}^p$.

Théorème 1.3 (Lemme de Slutsky)

Si $X_n \xrightarrow{\mathscr{L}} X$ et $Y_n \xrightarrow{\mathscr{L}} c$ alors

$$(X_n, Y_n) \xrightarrow{\mathscr{L}} (X, c).$$

- ▷ Remarque : il y a équivalence entre convergence en loi et en probabilité quand la limite est une constante.

Outline

Estimateur

On se donne g fonction définie sur Θ à valeurs dans $\mathbb{R}^k$.

- Un **estimateur** T_n du paramètre $g(\theta)$ est une statistique à valeurs dans $\mathbb{R}^k$.
- Exemple, $\widehat{\theta}_n = \overline{X}_n$ est un estimateur **non-biaisé** de $\mu = \mathbb{E}[X_1]$.

Premier exemple

On cherche à connaître μ l'espérance de la loi parente $\mathbb{P}$. On a à disposition un n-échantillon $(X_1, \dots, X_n)$.

- ▷ On peut prendre $\widehat{\theta}_n = \overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$.
- ▷ On note que $\mathbb{E}[\widehat{\theta}_n] = \mu$.
- ▷ On dira que l'estimateur est sans **biais.**

Biais

Soit $\widehat{\nu}_n$ un estimateur de $\nu(\theta)$, on notera $b_\theta(\widehat{\nu}_n)$ le **biais** de l'estimateur :

$$b_\theta(\widehat{\nu}_n) := \mathbb{E}[\widehat{\nu}_n] - \nu(\theta).$$

- ▷ $b_\theta(\widehat{\nu}_n)$ est une fonction de θ.
- ▷ Si $b_\theta(\widehat{\nu}_n) = 0$ alors l'estimateur est dit non biaisé.
- ▷ Si $\lim_{n\to+\infty} b_\theta(\widehat{\nu}_n) = 0$ alors l'estimateur est asymptotiquement sans biais.

Risque quadratique

Le risque **quadratique** de l'estimateur $\widehat{\nu}_n$ de $\nu(\theta)$ est :

$$R_\theta(\widehat{\nu}_n) \coloneqq \mathbb{E}\left[(\widehat{\nu}_n - \nu(\theta))^2\right].$$

- ▷ Le risque quadratique est associé à la fonction de perte quadratique.
- ▷ Décomposition biais-variance (exercice 1 du sujet 1) :

$$R_\theta(\widehat{\nu}_n) = b_\theta^2(\widehat{\nu}_n) + \mathrm{Var}(\widehat{\nu}_n).$$

Relation d'ordre

L'estimateur $\widehat{\theta}_n$ sera dit meilleur que $\widetilde{\theta}_n$, et on notera $\widehat{\theta}_n \preceq \widetilde{\theta}$ ssi :

$$R_\theta(\widehat{\theta}_n) \leq R_\theta(\widetilde{\theta}_n) \ ; \quad \forall\theta \in \Theta$$

- ▷ $\preceq$ n'est pas une relation d'ordre total sur l'ensemble des estimateurs de $g(\theta)$ (voir exercice 8 du sujet 1).
- ▷ Un estimateur $\widehat{\theta}_n$ sera **admissible** si aucun autre estimateur n'est strictement meilleur que lui *i.e.* il n'existe pas d'estimateur $\widetilde{\theta}_n$ tel que $\widetilde{\theta}_n \preceq \widehat{\theta}_n$ et $\exists\theta\ R_\theta(\widetilde{\theta}_n) < R_\theta(\widehat{\theta}_n)$.

Convergence

On dit que T_n est **consistant** pour $g(\theta)$ si il converge en probabilité vers $g(\theta)$:

$$T_n \xrightarrow{\mathbb{P}} g(\theta).$$

- ▷ Si T_n est un estimateur (asymptotiquement) sans biais de $g(\theta)$ et si $\mathrm{Var}(T_n) \to 0$ quand $n \to \infty$ alors T_n est un estimateur consistant pour $g(\theta)$.
- ▷ Si la convergence de T_n vers $g(\theta)$ est p.s. alors l'estimateur sera fortement consistant.

Application du TCL

La **loi** de la moyenne empirique d'un n-échantillon de loi parente de variance finie peut être approximée par une loi **normale** :

$$\overline{X}_n \underset{appr}{\sim} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Par exemple si $X \sim \mathcal{B}(\theta)$, $\sigma^2 = \theta\,(1-\theta)$

$$\overline{X}_n \underset{appr}{\sim} \mathcal{N}\left(\theta, \frac{\theta\,(1-\theta)}{n}\right)$$

$$\sqrt{n}\frac{\overline{X}_n - \theta}{\sqrt{\theta\,(1-\theta)}} \underset{appr}{\sim} \mathcal{N}(0,1).$$

Vitesse de convergence

Un estimateur $\widehat{\nu}_n$ de $\nu(\theta)$ a un comportement **asymptotiquement normal** à la **vitesse** v_n si $v_n \uparrow +\infty$ et

$$v_n \cdot (\widehat{\nu}_n - \nu) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \Sigma).$$

- L'estimateur de la moyenne empirique converge à la vitesse $\sqrt{n}$ (TCL).
- On peut avoir des convergences vers d'autres lois à d'autres vitesses (exercice 5 du sujet 2).

Construire un estimateur

Trois méthodes :

- Méthode empirique.
- Méthode des moments

$$\widehat{\mu}_n(p) := \frac{1}{n} \sum_{k=1}^{n} X_k^p.$$

- Maximum de vraisemblance : le n-échantillon a une densité

$$f_\theta(x_1, \dots, x_n).$$

Moments

Le moment d'ordre p, $\mathbb{E}[X^p]$, est une fonction de θ.

- ▷ On suppose $\mathbb{E}[|X|^p] < \infty$
- ▷ On suppose $\mu_\theta(p) \coloneqq \mathbb{E}[X^p] = f(\theta)$ avec f une fonction inversible continue.

On pose $\widehat{\theta}_n \coloneqq f^{-1}(\widehat{\mu}_n(p))$.

- ▷ $\widehat{\theta}_n \xrightarrow{\mathbb{P}} \theta$.
- ▷ $\widehat{\theta}_n$ est un estimateur convergent de θ.

Delta méthode

La delta méthode permet d'établir la vitesse de convergence de l'estimateur $\widehat{\theta}_n = f^{-1}(\widehat{\mu}_n(p))$.

- ▷ On suppose que $\mathbb{E}[|X|^{2p}] < +\infty$ et $f \in \mathcal{C}^1(\Theta, \mathbb{R})$

$$\sqrt{n}\,(\widehat{\theta}_n - \theta) \xrightarrow{\mathscr{L}} \mathcal{N}\left(0, \frac{\mathrm{Var}(X^p)}{|f'(\theta)|^2}\right)$$

- ▷ On note $\widehat{\sigma}_n^2(p)$ l'estimateur empirique de $\mathrm{Var}(X^p)$.

$$\frac{\sqrt{n}\,(\widehat{\theta}_n - \theta)}{\widehat{\sigma}_n(p)} \xrightarrow{\mathscr{L}} \mathcal{N}\left(0, \frac{1}{|f'(\theta)|^2}\right).$$

En dimension supérieure

On suppose un estimateur $\widehat{\theta}_n$ de θ asymptotiquement normal à la vitesse v_n :

$$v_n \cdot (\widehat{\theta}_n - \theta) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \Sigma).$$

Alors, pour $\nu : \Theta \to \mathbb{R}^d$ une fonction **différentiable**, $\widehat{\nu}_n = \nu(\widehat{\theta}_n)$ est un estimateur asymptotiquement normal de $\nu(\theta)$;

$$v_n \cdot (\widehat{\nu}_n - \nu) \xrightarrow{\mathscr{L}} \mathcal{N}(0, D_\nu \Sigma D_\nu^\top).$$

- D_ν est le jacobien de ν au point θ : $D_{\nu ij} = \frac{\partial \nu_i}{\partial \theta_j}(\theta)$

Généralisation de la méthode des moments

On suppose :

$$\nu(\theta) = \phi(\mu_\theta(1), \dots, \mu_\theta(q)).$$

- L'estimateur des moments est

$$\widehat{\nu}_n = \phi(\widehat{\mu}_n(1), \dots, \widehat{\mu}_n(q))$$

- Si $\mathbb{E}[|X|^q] < \infty$, l'estimateur des moments est consistant.
- Extension : $\mu_\theta(p) = \mathbb{E}[g_p(X)]$ et $\widehat{\mu}_\theta(p) = \frac{1}{n} \sum_{k=1}^n g_p(X_k)$.

Modèle dominé

On suppose que les lois $(\mathbb{P}_\theta)_{\theta\in\Theta}$ admettent une **densité** f_θ par rapport à une mesure commune, *e.g.* :

- ▷ la mesure de comptage μ,
- ▷ la mesure de Lebesgue λ.

La densité f_θ est :

- ▷ la distribution de probabilité dans le cas d'une domination par μ, $f_\theta(x_1,\ldots,x_n) = \mathbb{P}_\theta(X = x)$.
- ▷ la fonction de densité dans le cas où $\mathbb{P}_\theta \ll \lambda$.

Définition 2.1 (Vraisemblance)

*Dans un modèle **paramétrique dominé**, la vraisemblance est une fonction de la variable θ définie pour toute réalisation du n-échantillon $x = (x_1,\ldots,x_n)$ et qui à θ associe la valeur $f_\theta(x_1,\ldots,x_n)$:*

$$\theta \mapsto L_n(\theta; x_1,\ldots,x_n) = f_\theta(x_1,\ldots,x_n)$$

- ▷ La vrasemblance est une fonction du paramètre θ.
- ▷ La densité est, elle, une fonction de la variable x.
- ▷ Log-vraisemblance : $\ell_n(\theta; x) = \sum_{k=1}^{n} \log f_\theta(x_i)$.

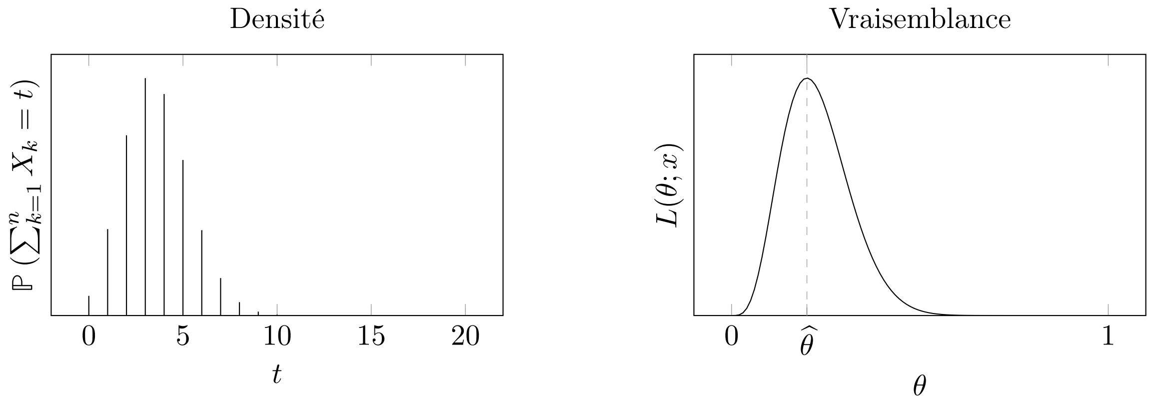

FIGURE – Densité de probabilité d'un n-échantillon de loi parente $\mathcal{B}(\theta^*)$ pour $\theta^* = 0{,}18$, $n = 20$ et vrasiemblance pour $\sum_{k=1}^{n} x_k = 4$.

Définition 2.2 (Estimateur du maximum de vraisemblance)

On considère un modèle paramétrique dominé. L'EMV est la ***v.a.*** $\widehat{\theta}_n$ *maximisant la vraisemblance :*

$$\widehat{\theta}_n \in \arg\max_{\theta \in \Theta} \{L_n(\theta; X)\}.$$

- Exemple, $\mathbb{P}_\theta = \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$:
  - $\ell_n(\theta; x1, \ldots x_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{k=1}^{n} \frac{(x_k-\mu)^2}{2\sigma^2}$.
  - $\widehat{\theta}_n = \left(\overline{X}_n, \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \overline{X}_n\right)^2\right)$.

Outline

Moyenne empirique

La méthode des moments fait intervenir, $\overline{X}_n$, $\widehat{\sigma}_n^2$.

- ▷ Loi de $\overline{X}_n$?
- ▷ Loi de $\widehat{\sigma}_n^2$?

TCL : $\sqrt{n}\frac{\overline{X}_n - \mathbb{E}[X]}{\widehat{\sigma}_n} \xrightarrow{\mathscr{L}} \mathcal{N}(0,1)$.

- ▷ Loi non asymptotique ?

Définition 3.1 (Vecteur gaussien)

Un ***vecteur aléatoire*** *X à valeurs dans $\mathbb{R}^d$ est dit gaussien si toute combinaison linéaire de ses coordonnées est une v.a.r. suivant une loi normale.*

- Les coordonnées d'un vecteur gaussien sont des v.a.r. gaussiennes.
- La réciproque est fausse.

Loi du vecteur gaussien

La loi d'un vecteur gaussien est déterminée par :

- son espérance $\mu = (\mu_1, \ldots, \mu_d)$;
- sa matrice de variance-covariance Σ.

On note $X \sim \mathcal{N}(\mu, \Sigma)$.

- Un n-échantillon gaussien **centré réduit** est un vecteur aléatoire $Z \sim \mathcal{N}(0, I_d)$.
- La matrice de variance-covariance d'un n-échantillon gaussien est de la forme $\sigma^2 I_d$.

Si Σ est inversible alors X est à densité et

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right).$$

Application linéaire

- Si $X \sim \mathcal{N}(\mu, \Sigma)$ alors $Y := AX \sim \mathcal{N}(A\mu, A\Sigma A^\top)$.
- Si $X \sim \mathcal{N}(0, \sigma^2 I_d,)$ et A est orthonormale alors

$$Y = AX \sim \mathcal{N}(0, \sigma^2 I_d).$$

- Si X est un n-échantillon gaussien ($i.e.$ $X_1 \sim \mathcal{N}(0, \sigma^2)$) alors

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Loi du khi-deux

Soit Z un n-échantillon gaussien **centré réduit**.

Définition 3.2

La loi du carré de la norme euclidienne du vecteur est appelée loi du khi-deux (centré) à n degrés de liberté. On note

$$\|Z\|^2 = \sum_{k=1}^{n} Z_k^2 \sim \chi^2(n).$$

Son espérance est n, sa variance $2n$.

- Si $Z \sim \mathcal{N}(\mu, I_n)$ alors $\|Z\|^2$ suit une loi du khi-deux décentré notée $\chi^2(n, \|\mu\|^2)$.
- Son espérance est $n + \|\mu\|^2$, sa variance $2(n + 2\|\mu\|^2)$.

n-échantillon

Soit X un n-échantillon gaussien. $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. On a

$$\sum_{k=1}^{n} \left(\frac{X_k - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

▷ Loi de l'estimateur empirique de la variance à espérance connue.

Variance empirique

Soit X un n-échantillon gaussien. $X_1 \sim \mathcal{N}(\mu, \sigma^2)$. On a

$$\sum_{k=1}^{n} \left(\frac{X_k - \overline{X}_n}{\sigma} \right)^2 \sim \chi^2(n-1).$$

▷ Loi de l'estimateur **sans biais** de la variance :

$$\widehat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} (X_k - \overline{X}_n)^2 \sim \frac{\sigma^2}{n-1} \chi^2(n-1).$$

▷ Les v.a.r. $\overline{X}_n$ et $\widehat{\sigma}_n^2$ sont **indépendantes**.

▷ $\overline{X}_n$ et $X - \overline{X}_n$ sont des projections orthogonales sur deux sous-espaces orthogonaux (Cochran).

Loi de Student

Z et K **indépendantes** avec $Z \sim \mathcal{N}(0,1)$ et $K \sim \chi^2(p)$ alors
$T := \frac{Z}{\sqrt{K/p}}$ suit une loi de **Student** à p degrés de liberté, notée $\mathcal{T}(p)$.

- Permet de calculer des intervalles de confiance sur $\overline{X}_n$:

$$T_n := \sqrt{n}\frac{\overline{X}_n - \mu}{\widehat{\sigma}_n}.$$

- $T_n \sim \mathcal{T}(n-1)$ est appelée **statistique de student** de l'échantillon.
- $T_n^2 \sim \mathcal{F}(1,n)$.
- La loi de Fisher $\mathcal{F}(m,n)$ est la loi de $\frac{K_1 n_2}{K_2 n_1}$ quand $K_i \sim \chi^2(n_i)$ et K_1 et K_2 indépendantes.

Outline

Comportements asymptotiques

Estimation **ponctuelle** de $\nu(\theta)$

- $\mathbb{E}[\widehat{\nu}_n] \xrightarrow[n\to+\infty]{} \nu$;
- $\widehat{\nu}_n \xrightarrow{\mathbb{P}} \nu$;
- $R_\theta(\widehat{\nu}_n) \xrightarrow[n\to+\infty]{} 0$;
- $v_n\left(\widehat{\nu}_n - \nu\right) \xrightarrow{\mathscr{L}} \mu_\nu$.
  - v_n est la vitesse de convergence ;
  - si μ est indépendante de ν : loi **pivotale**.

Définition 4.1 (Intervalle de confiance)

Un intervalle de confiance (IC) de niveau de confiance $1-\alpha$ pour θ est un intervalle dont les bornes sont des ***statistiques*** *telles que :*

$$\mathbb{P}\left(\left[\widehat{\theta}_n^-, \widehat{\theta}_n^+\right] \ni \theta\right) \geq 1-\alpha.$$

- $\mathbb{P}\left(\left[\widehat{\theta}_n^-, \widehat{\theta}_n^+\right] \ni \theta\right) = \mathbb{P}(\{\widehat{\theta}_n^- \leq \theta\} \cap \{\widehat{\theta}_n^+ \geq \theta\})$.
- α est un **risque** de première espèce.

Quantiles

Le **quantile** q_α d'ordre $\alpha \in [0, 1]$ d'une loi réelle μ est le réel tel que, si $X \sim \mu$,

$$\mathbb{P}(X \leq q_\alpha) = \alpha.$$

- ▷ q_α est la fonction réciproque de la fonction de répartition F,
- ▷ obtenue par interpolation linéaire dans le cas discret.
- ▷ Lorsque la loi est symétrique :

$$\mathbb{P}(|X| > q_{1-\alpha/2}) = \alpha.$$

- ▷ On note q^*_α le quantile de $\mathcal{N}(0, 1)$.

Exemple gaussien I

X un n-échantillon gaussien, $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ à espérance et variance inconnues.

$$T_n = \sqrt{n}\frac{\overline{X} - \mu}{\widehat{\sigma}_n} \sim \mathcal{T}(n-1).$$

Exemple gaussien II

On note t_n le quantile de $\mathcal{T}(n-1)$ d'ordre $1-\frac{\alpha}{2}$.

$$\begin{aligned}1-\alpha &= \mathbb{P}(-t_n \leq T_n \leq t_n) \\ &= \mathbb{P}\left(-\frac{\widehat{\sigma}_n t_n}{\sqrt{n}} \leq \mu - \overline{X}_n \leq \frac{\widehat{\sigma}_n t_n}{\sqrt{n}}\right) \\ &= \mathbb{P}\left(\overline{X}_n - \frac{\widehat{\sigma}_n t_n}{\sqrt{n}} \leq \mu \leq \overline{X}_n + \frac{\widehat{\sigma}_n t_n}{\sqrt{n}}\right).\end{aligned}$$

- ▷ Pour $\alpha = 0{,}01$, $t_{20} = 2{,}83$, $t_{40} = 2{,}70$.
- ▷ Il s'agit de l'IC de Student.

Définition 4.2 (Intervalle de confiance asymptotique)

*Un IC **asymptotique** de niveau de confiance $1-\alpha$ pour θ est une suite d'intervalles dont les bornes sont des **statistiques** telles que :*

$$\mathbb{P}\left(\left[\widehat{\theta}_n^-, \widehat{\theta}_n^+\right] \ni \theta\right) \xrightarrow[n\to+\infty]{} 1-\alpha.$$

- ▷ Dans le cas **non gaussien** on construira l'IC asymptotique pour l'espérance par le TCL.
- ▷ On considérera $\mathbb{P}(I_n \ni \theta) \approx 1-\alpha$ pour $n \geq 30$.

Exemple

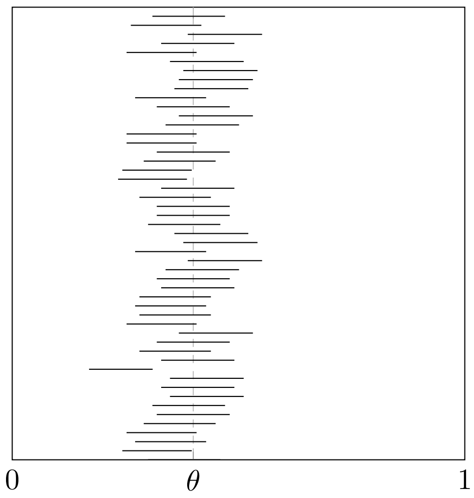

Figure – Exemples d'IC obtenus pour $\alpha = 0{,}1$ et un modèle statistique de Bernoulli ($\theta = 0{,}4$, $n = 100$).

Outline

Principe

Choix entre deux hypothèses sur le paramètre θ dans le cadre d'un modèle statistique paramétrique :

- hypothèse **nulle** (H_0) ;
- hypothèse **alternative** (H_1).
- Exemple :
  - $(H_0) : \mu = \mu_0$
  - $(H_1) : \mu \neq \mu_0$.

L'hypothèse (H_0) est supposée vraie : α, le **risque** de première espèce est la probabilité de **rejeter** à tort (H_0).

Construction

Lors de la construction d'un test, on définit :

- le modèle ;
- les hypothèses ;
- une statistique $T(X)$;
- α le risque de première espèce et la **région critique** :

$$\mathbb{P}_{(H_0)}(T_n \in \mathcal{R}_\alpha) = \alpha.$$

Puissance

La **puissance** du test est sa capacité à détecter l'hypothèse alternative.

- Le risque de **deuxième** espèce est noté β :

$$\mathbb{P}_{(H_1)}(T_n \in \mathcal{R}_\alpha^\mathsf{c}) = \beta$$

- La puissance du test est $\pi = 1 - \beta$.

Exemple

X n-échantillon gaussien. Test d'hypothèse sur l'espérance.

- $(H_0) : \mu = \mu_0$, $(H_1) : \mu \neq \mu_0$.
- $T_n \sim_{(H_0)} \mathcal{T}(n-1)$.
- $\mathcal{R}_\alpha^\mathsf{c} = [-t_n^*, t_n^*]$.
- $\mathbb{P}_{(H_0)}(T_n \in \mathcal{R}_\alpha) = \alpha$.

(H_0) est conservée si $t_n \in [-t_n^*, t_n^*]$.

- La puissance $\pi(\mu)$ du test est une fonction de μ.

$$\pi(\mu) = \mathbb{P}_{(H_1)}(T_n \notin [-t_n^*, t_n^*].)$$

Hypothèses composites

Un hypothèse composite porte sur un sous-ensemble de Θ :

- test **bilatère** pour hypothèse nulle simple, hypothèse alternative composite $(H_0) : \theta = \theta_0$ contre $(H_1) : \theta \neq \theta_0$
- test **unilatère** pour hypothèse nulle composite, hypothèse alternative composite $(H_0) : \theta < \theta_0$ contre $(H_1) : \theta \geq \theta_0$

Pour un test à hypothèse nulle composite, la **taille** du test est $\sup_{\theta \in \Theta_0} \alpha(\theta)$.

- Le test est de niveau α si $\sup_{\theta \in \Theta_0} \alpha(\theta) \leq \alpha$.

Biais et consistance

- Un test est dit sans biais si $\pi \geq \alpha$.
- Un test est dit consistant si $\pi \xrightarrow[n \to +\infty]{} 1$.

Test UPP

Un test est dit **uniformément plus puissant (UPP)** si il est plus puissant que tout autre test de même niveau pour toute valeur du paramètre θ.

Rapport de vraisemblance

Théorème 5.1 (Neyman-Pearson)

Dans le cas de deux hypothèses simples $\theta = \theta_0$ contre $\theta = \theta_1$ un test UPP est de la forme :

$$\mathcal{R}_\alpha = \left\{ x \in \mathbb{R}^n \,:\, \frac{L(\theta_1; x)}{L(\theta_0; x)} > k_\alpha \right\}.$$

- Pour des tests composites unilatères ($\theta = \theta_0$ contre $\theta > \theta_0$) le test UPP est test le du rapport de vraisemblance si $\mathcal{R}_\alpha$ ne dépend pas de θ_1 dans le test à hypothèses simples.

p-value

- La p-value est calculée à partir de l'observation t_n de la statistique du test.
- C'est le plus petit α qui rejette (H_0) étant donné t_n observé.
- Si $\theta_1 > \theta_0$, c'est $\mathbb{P}_{(H_0)}(T_n \geq t_n)$.

Tests d'adéquation

Les tests du khi-deux sont des tests asymptotiques non paramétriques. Ils sont construits avec une statistique $K_n \sim_{(H_0)} \chi^2(p)$:

- test d'ajustement ;
- test d'homogénéité ;
- test d'indépendance.

Loi multinomiale

On considère une loi catégorielle à K catégories caractérisée pas la distribution de probabilités (d.d.p.) $\pi = (\pi_1, \ldots, \pi_n)$.

- On note $O = (O_1, \ldots, O_K)$ le vecteur du décompte de chacune des catégories au sein d'un n-échantillon X de cette loi.

$$O_k = \sum_{i=1}^{n} \mathbb{1}_{X_i = k}.$$

- $O \sim \mathcal{M}(n, \pi)$.
- $\mathbb{P}(O = o) = n! \prod_{k=1}^{K} \frac{\pi_k}{n_k!}$.

Ajustement

$$k_n = \sum_{i=1}^{p} \frac{(o_i - e_i)^2}{e_i}.$$

- o_i effectif observé de la catégorie i
- e_i effectif théorique de la catégorie i
- Sous (H_0), $K_n \xrightarrow{\mathscr{L}} \chi^2(p-1)$.
- Loi asymptotique admise si $\min_i e_i > 5$.
- Si $k_n > q_{1-\alpha}$ alors l'hypthèse nulle (d'ajustement) est rejetée.

Homogénéité

$$k_n = \sum_{i=1}^{p}\sum_{j=1}^{m} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}.$$

- ▷ On dispose de m échantillons (de n_j individus).
- ▷ o_{ij} effectif observée de la catégorie i dans l'échantillon j.
- ▷ e_{ij} effectif théorique de la catégorie i dans l'échantillon j :

$$e_{ij} = \frac{\sum_{j=1}^{m} o_{ij} \sum_{i=1}^{p} o_{ij}}{N}$$

- ▷ $K_n \sim_{(H_0)} \chi^2((p-1)(m-1))$.
- ▷ Si $k_n > q_{1-\alpha}$ alors l'hypothèse nulle est rejetée.

Indépendance

Deux variables à p et m modalités.

$$k_n = \sum_{i=1}^{p}\sum_{j=1}^{m} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}.$$

- ▷ o_{ij} effectif observée de la catégorie i et j.
- ▷ e_{ij} effectif théorique de la catégorie i et j :

$$e_{ij} = \frac{\sum_{j=1}^{m} o_{ij} \sum_{i=1}^{p} o_{ij}}{N}$$

- ▷ $K_n \sim_{(H_0)} \chi^2((p-1)(m-1))$.

Si $k_n > q_{1-\alpha}$ alors l'hypthèse nulle est rejetée.

Outline

Motivation

- En statistique fréquentiste, θ est une constante inconnue.
- En statistique bayésienne l'inconnu sur θ est formalisé par une v.a.r.
- La densité de θ résume l'information sur θ.

Lois à densité

$$\pi(\theta|x) = \frac{\pi(\theta)f(x|\theta)}{f(x)}.$$

- ▷ $\pi(\theta|x)$ est la probabilité **a posteriori**.
- ▷ $\pi(\theta)$ est l'**a priori** sur θ.
- ▷ $f(x|\theta)$ est la **vraisemblance**.
- ▷ $f(x)$ n'intervient pas lors de l'estimation par maximum a posteriori (MAP).

Estimation ponctuelle

- ▷ $L(d, \theta)$ est la fonction de **coût** – conséquences de considérer d plutôt que θ.
- ▷ L'estimation ponctuelle minimise le coût moyen :

$$\delta(x) = \min_d \int_\Theta \pi(\theta|x) L(d, \theta)\, d\theta.$$

- ▷ Si la fonction de coût est quadratique :

$$\delta(x) = \int_\Theta \pi(\theta|x)\theta\, d\theta$$

Région de crédibilité

$$\mathcal{R}_k(x) = \{\theta \,:\, \pi(\theta|x) \geq k\}$$

- ▷ Permet de créer des régions de volume minimal.

Facteur de Bayes

Deux modèles concurrents, $\mathcal{M}_1$ et $\mathcal{M}_2$.

- ▷ Un a priori différent par modèle.
- ▷ Une vraisemblance par modèle.
- ▷ On construit le facteur de Bayes :

$$B_{12} = \frac{\mathbb{P}(\mathcal{M}_1|x)\mathbb{P}(\mathcal{M}_2)}{\mathbb{P}(\mathcal{M}_2|x)\mathbb{P}(\mathcal{M}_1)}$$

On dispose d'un n-échantillon X.

- On souhaite connaître la densité f de X.
- Estimation paramétrique : $f \in \left\{ f(x,\theta) \, : \, \theta \in \Theta \subset \mathbb{R}^k \right\}$.
- Estimation non paramétrique :$f \in \mathcal{F}$.

Fonction de répartition

$$\widehat{F}_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{x_i \leq y} \xrightarrow{\text{p.s.}} F(y) = \mathbb{P}(X_1 \leq y)$$

Densité

La densité est

$$f(y) = F'(y) \approx \frac{F(y+h) - F(y)}{h}.$$

- ▷ Estimateur de Rosenblatt :

$$\widehat{f}_n(y) = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{1}_{y \leq x_i < y+h}.$$

- ▷ Estimateur à noyau :

$$\widehat{f}_n(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - x_i}{h}\right).$$