Amphi de révision

Statistiques et Apprentissage

Basé sur les cours de Paul-Henri COURNÈDE et Laurent LE BRUSQUET

Par Mathïs Fédérico le 16/05/2019

Deux courants de pensée

Le Fréquentisme

En observant des ***fréquences***, on veut trouver la théorie qui ***maximise la vraisemblance*** des données :

$$T_{best} = \arg\max_{t} \mathbb{P}(D|t)$$

Le Bayésianisme

En observant des ***événements***, on ***compare la crédence*** des théories sachant des données :

$$\mathbb{P}(T|D) = \frac{\mathbb{P}(D|T)\mathbb{P}(T)}{\mathbb{P}(D|T)\mathbb{P}(T) + \sum_{A \neq T} \mathbb{P}(D|A)\mathbb{P}(A)}$$

Le Fréquentisme : Les statistiques classiques

Le Fréquentisme

Donnée
$X(\omega)$

Modèle fréquentiste
(VA iid)
Pp 9-18

Modèle paramétrique
$\mathcal{M}_\theta = \{p_\theta \,/\, \theta \in \Theta\}$
Pp 18-22

$$T_{best} = \arg\max_t \mathbb{P}(D|t) \qquad \theta^* = \arg\max_\theta \mathbb{P}(D|T(\theta))$$

Le Bayésianisme : L'ampleur de la formule de Bayes

Le Bayésianisme

Vraisemblance des données D
dans la théorie T

Préjugé
en la théorie T
A Priori

$$\mathbb{P}(T|D) = \frac{\mathbb{P}(D|T)\mathbb{P}(T)}{\mathbb{P}(D|T)\mathbb{P}(T) + \sum_{A \neq T} \mathbb{P}(D|A)\mathbb{P}(A)}$$

Crédence de la théorie T
sachant les données D
A Posteriori

Vraisemblance des données D
dans les théories A

Préjugé
en les théories A

La crédence d'une théorie est relative aux théories alternatives

Le Bayésianisme paramétrique

Le Bayésianisme paramétrique

$\theta \rightsquigarrow ?$

On considère simultanément toutes les valeurs possible de θ:
Densité de paramètre θ

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int_{\Theta} p(x|\theta)p(\theta)d\theta}$$

Vraisemblance des données x dans le modèle de paramètre θ

Densité de paramètre θ **A Priori**

Densité de paramètre θ sachant les données x **A Posteriori**

Normalisation relative aux autres paramètres (Non calculé -> Proportionnalité)

Construction des estimateurs

Estimateur fréquentiste

Méthode des moments
(substitution)

$$\widehat{\mathbb{E}}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i = \mathbb{E}_{\hat{\theta}}(X)$$

$$\begin{cases} \widehat{\mathbb{M}^p}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^p = \mathbb{M}_{\hat{\theta}}^p(X) \end{cases}$$

Ou

Maximum de vraisemblance

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} p_\theta(x_i)$$

Pp 28-30

Estimateur bayésien

Distribution à postériori
(Bayes)

$$p(\theta|x) \propto p(x|\theta)p(\theta) = p(x,\theta)$$

Suffisant si on tombe sur une loi connue sinon on doit calculer la normalisation

Espérance conditionnelle
(pour un risque quadratique)

$$\hat{\theta} = \mathbb{E}(\theta|X)$$

Pp 50

Qualité des estimateurs

Estimateur $\hat{\theta}$

Consistance
Faible : $\mathbb{P}$
Forte: $p.s$
Vitesse: $a_N(Y_N - c) \xrightarrow{\mathcal{L}} Z$
Pp 40-41

Biais
$b_{\theta^*}(\hat{\theta}) = \mathbb{E}_{\theta^*}(\hat{\theta}) - \theta^*$
Pp 32

Régularité C1-C4
Pp 23

Score de vraisemblance
$S_\theta(X) = \nabla_\theta \ln(p_\theta(X))$
Pp 24

Information de Fisher
$I(\theta) = \mathbb{V}_\theta(S_\theta(X))$
Si régularité C1 à C4:
$I_{i,j}(\theta) = -\mathbb{E}_\theta\left(\frac{\partial^2 \ln(p_\theta(X))}{\partial\theta_i \partial\theta_j}\right)$
Pp 25

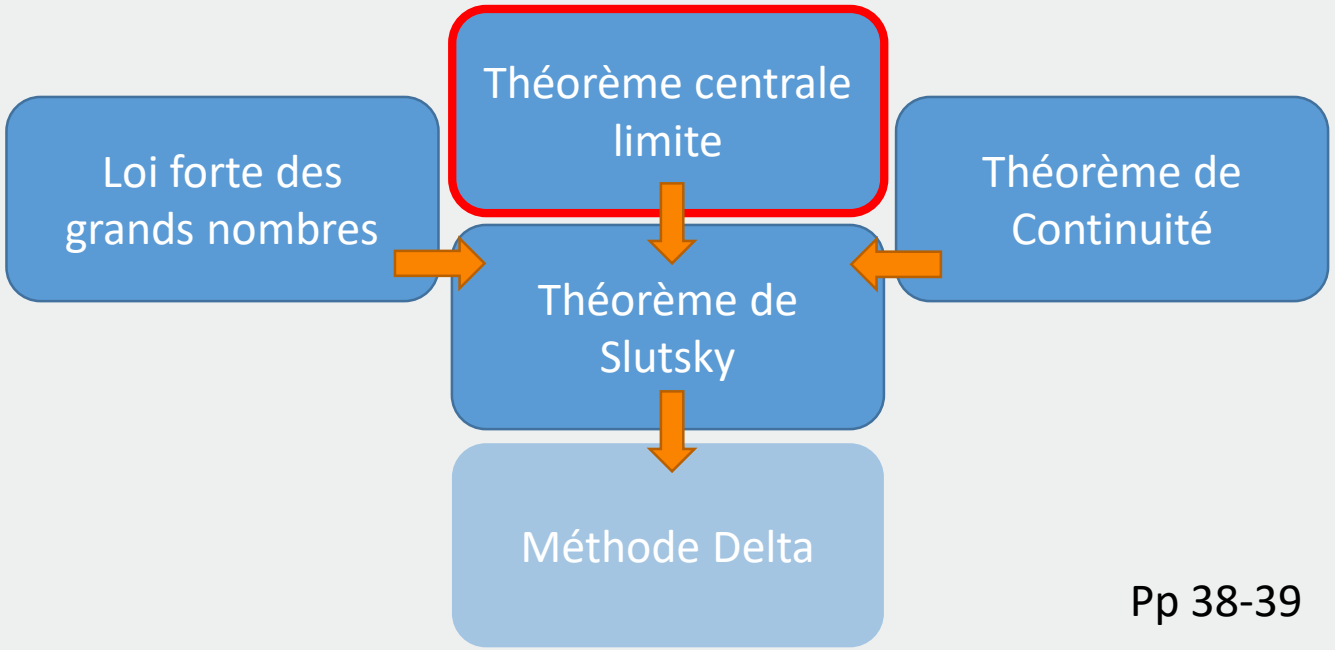
Risque
$R = f(X, \theta)$
Quadratique
$R = \| \hat{\theta} - \theta \|_2^2$
Pp 32

Efficacité ($\mathbb{V}$)
(Sans biais)
Efficace :
$\mathbb{V}_\theta = I^{-1}(\theta)$
Pp 34-36

Asymptotiquement normal
Pp 42

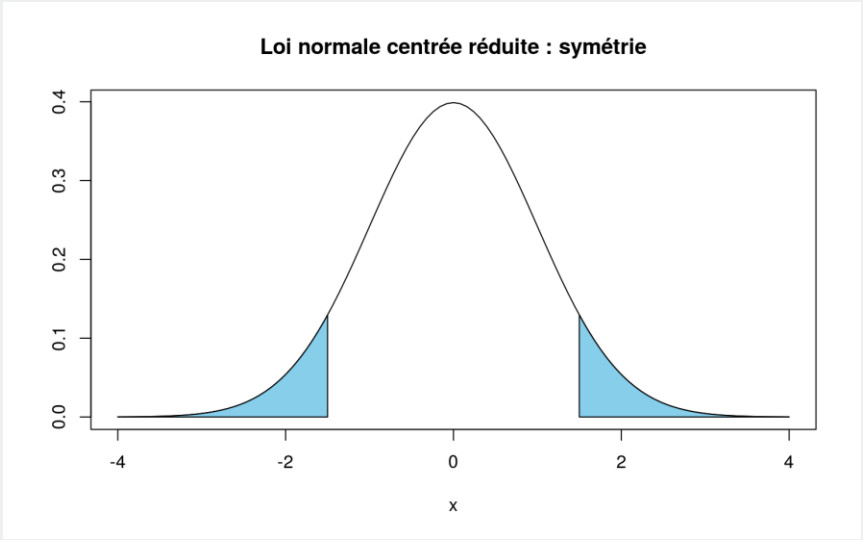
Calcul d'intervalle de confiance

Fonction pivotale



Pp 38-39

Intervalle de confiance



Pp 52-54 et 58-60

Test d'hypothèses

Vocabulaire des hypothèses

$H0$ et $H1$ — Pp 64

$Simple$ $\Theta_0 = \{\theta_0\}$ — Pp 67

$Complexe$ $\Theta_0 = \{\theta_0 \geq 0\}$ — Pp 67

Statistiques de test Et région de rejet — Pp 64

$\mathbb{P}(H0)$

$\mathbb{P}(H1)$

Θ_0

Θ_1

$\searrow \alpha$ 1ère espèce — Pp 64

β 2nd espèce — Pp 64

$\nearrow \pi$ Puissance — Pp 64

Fonction puissance — Pp 68

$Taille$ sup — Pp 65

UPP $\forall \theta \in \Theta_1$ — Pp 68

$p - value$ inf — Pp 66

Algorithme : Statistique de test

Construction

- **Simple**
  - Rapport de Vraisemblance (Neyman-Pearson) — Pp 71
    - Test plus simple — Pp 72
- **Composé**
  - Zone de rejet arbitraire
    - Kolmogorov-Smirnov — Pp 77
    - Chi-2 de Pearson — Pp 74

Evaluation

- **Seul**
  - $Taille$: $\sup \alpha$
  - π : Puissance
  - p-value : (1 – F(T(X)) si Simple)
- **Comparatif**
  - UPP : $\forall \theta \in \Theta_1$