STT 1000 - STATISTIQUES

ARTHUR CHARPENTIER

Puissance

Puissance

La fonction puissance d'un test statistique d'un paramètre θ est une fonction notée γ

$$\gamma(\theta) = \mathbb{P}\big[\text{rejeter } H_0 \mid \theta = \theta_1\big] = \mathbb{P}\big[\text{accepter } H_1 \mid \theta = \theta_1\big]$$

La fonction puissance permet d'évaluer la probabilité de *détecter* H_1 pour toute valeur de θ_1.
$\gamma(\theta_1) = 1 - \beta(\theta_1)$ pour tout θ_1,

Puissance

Si $H_1 : \theta > \theta_0$ alors

- $\gamma(\theta_0) = \alpha$ (par construction)
- $\gamma(\theta_0) < \alpha$ pour $\theta_1 < \theta_0$
- (on espére que) $\gamma(\theta_0) >> \alpha$ pour $\theta_1 > \theta_0$

α est le premier paramètre utilisé pour construire la règle de décision

la fonction puissance est là pour mesurer la qualité du test si nous étions sous H_1

Puissance

Pour un test de moyenne dans un échantillon Gaussien

$$\gamma(\mu_1) = 1 - \Phi\left(u_\alpha + \sqrt{n}\frac{\widehat{\mu} - \mu_1}{\sigma}\right)$$

Test du rapport de vraisemblance

On a vu que

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I(\theta)^{-1})$$

La Δ-méthode (formule de Taylor)

$$g(\widehat{\theta}) - g(\theta) \approx g'(\theta)[\widehat{\theta} - \theta]$$

avec $g = \log \mathcal{L}$ s'écrirait

$$\sqrt{n}\big(\log \mathcal{L}(\widehat{\theta}) - \log \mathcal{L}(\theta)\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \log \mathcal{L}'(\theta)^2 I(\theta)^{-1})$$

mais $\theta \mapsto \log \mathcal{L}'(\theta)^2 I(\theta)^{-1}$ s'annule en $\widehat{\theta}$... donc non !
Il faut une Δ-méthode au second ordre

$$g(\widehat{\theta}) - g(\theta) \approx \frac{g''(\theta)}{2}[\widehat{\theta} - \theta]$$

de telle sorte que

$$\frac{2\sqrt{n}}{I^{-1}(\theta)g''(\theta)}\big(g(\widehat{\theta}) - g(\theta)\big) \xrightarrow{\mathcal{L}} \chi^2(1)$$

Test du rapport de vraisemblance

Si $\widehat{\theta}$ est l'estimateur du maximum de vraisemblance

$$2\big(\log\mathcal{L}(\widehat{\theta})-\log\mathcal{L}(\theta)\big)\xrightarrow{\mathcal{L}}\chi^2(1)$$

- $H_0:\theta=\theta_0$ contre $H_1:\theta\neq\theta_0$

$$T=-2\big(\log\mathcal{L}(\theta_0)-\log\mathcal{L}(\widehat{\theta})\big)\xrightarrow{\mathcal{L}}\chi^2(1) \text{ si } H_0 \text{ est vraie}$$

- $H_0:\theta=\theta_0$ contre $H_1:\theta=\theta_1$

$$T=-2\big(\log\mathcal{L}(\theta_0)-\log\mathcal{L}(\theta_1)\big)$$

- $H_0:\theta\in\Theta_0$ contre $H_1:\theta\in\Theta_1$

$$T=-2\left(\sup_{\theta\in\Theta_0}\log\mathcal{L}(\theta)-\sup_{\theta\in\Theta_1}\log\mathcal{L}(\theta)\right)$$

Lemme de Neyman-Pearson

Si on cherche à tester

- $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$

alors le test le plus puissant de niveau α est celui dont la région critique est

$$= \mathcal{C} = \{\boldsymbol{x} : T(\boldsymbol{x}) > c_\alpha\} \text{ où } T(\boldsymbol{x}) = -2\big(\log\mathcal{L}(\theta_0;\boldsymbol{x}) - \log\mathcal{L}(\theta_1;\boldsymbol{x})\big)$$

où c_α vérifie $\mathbb{P}[\boldsymbol{X} \in \mathcal{C} | H_0] = \alpha$

Test du rapport de vraisemblance: cas Bernoulli

On considère $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ de loi $\mathcal{B}(p)$. On veut tester

$$H_0 : p = p_0 \text{ contre } H_1 : p = p_1, \text{ avec } p_0 < p_1$$

Ici $\log \mathcal{L}(p; \boldsymbol{x}) = s \log(p) + (n-s)\log(1-p)$, où $s = x_1 + \cdots + x_n$ donc

$$T = \big(s\log(p_1) + (n-s)\log(1-p_1) - s\log(p_0) + (n-s)\log(1-p_0)\big)$$

$$s \log\left[\frac{p_1(1-p_0)}{p_0(1-p_1)}\right] + n\log\left[\frac{(1-p_1)}{(1-p_0)}\right] > c_\alpha$$

donc

$$\boldsymbol{x} \in \mathcal{C} \text{ ssi } \sum_{i=1}^n x_i > \gamma_\alpha$$

où $\mathbb{P}[S > \gamma_\alpha | H_0] = \alpha$ avec $S = \sum_{i=1}^n X_i \sim \mathcal{B}(n, p_0)$

Test du rapport de vraisemblance: cas Beta

On considère $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ de loi $\mathcal{B}(p)$. On veut tester

$$H_0 : X_i \sim \mathcal{U}([0,1]) \text{ contre } H_1 : X_i \sim \mathcal{B}(2,1)$$

Ici $f_0(x) = \mathbf{1}_{[0,1]}(x)$ et $f_1(x) = 2x\mathbf{1}_{[0,1]}(x)$ donc

$$T = \big(\log \mathcal{L}_1\big) = \big(n \log 2 + \sum \log(x_i)\big)$$

donc

$$\boldsymbol{x} \in \mathcal{C} \text{ ssi } \sum_{i=1}^{n} \log(x_i) > \gamma_\alpha \text{ ssi } \sum_{i=1}^{n} -\log(x_i) < \delta_\alpha$$

où $\mathbb{P}[S < \delta_\alpha | H_0] = \alpha$ avec $S = -\sum_{i=1}^{n} \log[X_i] \sim \mathcal{G}(n,1)$, car $-\log[X_i] \sim \mathcal{E}(1)$ sous H_0.