

Sous certaines conditions de régularité, l'EMV $\hat{\theta}$ de $\theta \in \Theta \subset \mathbb{R}^p$ est

- ▶ consistant
- ▶ asymptotiquement normal

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}_p(0, I_1(\theta)^{-1})$$

où $I_1(\theta) = -\mathbb{E}_\theta[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X_1)]$

- ▶ si $\hat{V}_n I_n^{-1}(\theta) \xrightarrow{\mathcal{P}} Id_p$,

$$\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta^*) \overset{appr}{\sim} \mathcal{N}(0, Id_p)$$

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

- ▶ Calculer la loi asymptotique de l'estimateur d'une fonction $h(\theta)$: statistique de **Wald**, **delta-méthode**
- ▶ Méthodologie de construction d'un test (rappel)
- ▶ Propriétés des tests UPP dans les familles à **RV monotone**
- ▶ Construire le test du **rapport de vraisemblance**

Estimer $h(\theta)$

Tests

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Estimer une fonction $h(\theta)$ [poly 3.3.3]

Soit $\hat{\theta}_n$ un estimateur de $\theta \in \Theta \subset \mathbb{R}^p$ asymptotiquement normal

$$\hat{V}_n^{-1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id_p)$$

- pour $h(\theta) = A\theta \in \mathbb{R}^r$, avec A est une matrice de $\mathcal{M}_{r,p}(\mathbb{R})$ de rang r , on a :

$$T_n = (A\hat{V}_n A')^{-1/2}(A\hat{\theta}_n - A\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, Id_r)$$

$$W = T_n' T_n = [A(\hat{\theta}_n - \theta)]'(A\hat{V}_n A')^{-1} A(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \chi^2(r)$$

W est appelée **statistique de Wald**, dont la loi asymptotique est libre du paramètre inconnu

- et pour $h(\theta)$ non linéaire ?

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Proposition

Soit h une fonction de $\mathbb{R}^p$ dans $\mathbb{R}^q$, U_n une suite de variables aléatoires de $\mathbb{R}^p$ et a_n , $n \in \mathbb{N}$, une suite déterministe de réels. On suppose que

- ▶ $a_n \rightarrow +\infty$
- ▶ Il existe un vecteur déterministe $U \in \mathbb{R}^p$, et un vecteur aléatoire L tels que $a_n(U_n - U) \xrightarrow{\mathcal{L}} L$
- ▶ h est une fonction différentiable en U , de différentielle notée $Dh(U)$, matrice de taille $p \times q$.

On a alors la convergence en loi suivante :

$$a_n (h(U_n) - h(U)) \xrightarrow{\mathcal{L}} Dh(U) L$$

Estimer $h(\theta)$ h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Il suffit de montrer que

$$A_n = a_n (h(U_n) - h(U)) - Dh(U) [a_n(U_n - U)] \xrightarrow{\mathcal{P}} 0.$$

Pour tout $\varepsilon > 0$, montrons $\mathbb{P}(\|A_n\| \geq \varepsilon) \rightarrow 0$. Puisque h est différentiable, pour tout $k > 0$, il existe $\delta > 0$ tel que pour tout $V \in \mathbb{R}^p$

$$\|V - U\| \leq \delta \Rightarrow \|h(V) - h(U) - Dh(U) [V - U]\| \leq k\|V - U\|$$

Posons $k = \varepsilon/m$. On a

$$\begin{aligned} \mathbb{P}(\|A_n\| \geq \varepsilon) &\leq \mathbb{P}(\{\|A_n\| \geq \varepsilon\} \cap \{\|U_n - U\| \leq \delta\}) + \mathbb{P}(\|U_n - U\| > \delta) \\ &\leq \mathbb{P}\left(a_n \frac{\varepsilon}{m} \|U_n - U\| \geq \|A_n\| \geq \varepsilon\right) + \mathbb{P}(a_n \|U_n - U\| > a_n \delta) \\ &\leq 2 \mathbb{P}(a_n \|U_n - U\| \geq \min(m, a_n \delta)) \end{aligned}$$

d'où, pour tout m :

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\|A_n\| \geq \varepsilon) \leq \limsup_{n \rightarrow \infty} 2\mathbb{P}(a_n \|U_n - U\| \geq m) \leq 2\mathbb{P}(\|L\| \geq m)$$

Estimer $h(\theta)$ h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Dans le cas de l'EMV

- ▶ Invariance fonctionnelle : $h(\hat{\theta})$ est l'EMV de $h(\theta)$
- ▶ d'un modèle régulier : $U_n = \hat{\theta}_n$, $U = \theta$, $a_n = \sqrt{n}$ et $L \sim \mathcal{N}(0, V_1(\theta))$ avec $V_1(\theta) = I_1^{-1}$.

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_1(\theta)^{-1})$$

- ▶ d'où comportement asymptotiquement normal de l'EMV d'une fonction différentiable du paramètre :

$$\sqrt{n}(h(\hat{\theta}_n) - h(\theta)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta) = Dh(\theta) V_1(\theta) Dh(\theta)')$$

- ▶ on peut également définir une statistique de Wald

Exemple : Asymptotique de l'estimateur $\bar{X}(1 - \bar{X})$ de la variance d'une loi de Bernoulli

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Estimer $h(\theta)$

Tests

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Construire un test (rappel)

Choisir entre deux hypothèses (H_0) et (H_1), en calibrant le risque de choisir (H_1) à tort.

1. Définir le **modèle**
2. Définir les **hypothèses nulle** (H_0) et **alternative** (H_1)
3. Choisir une **statistique de test** $T(X)$, calculer sa **loi sous** (H_0)
4. Définir la **règle de décision** en calibrant la région de rejet $\mathcal{R}_\alpha$ de (H_0) suivant le risque de première espèce α

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

5. Calcul de la statistique observée et **décision**
6. Calcul de la puissance

$$\mathbb{P}_{(H_1)}(T(X) \in \mathcal{R}_\alpha) = \pi = 1 - \beta.$$

Les erreurs d'un test (rappel)

A l'issue d'un test, les quatre situations suivantes sont possibles

	Choix (H_0)	Choix (H_1)
(H_0) vraie	$1 - \alpha$ bonne décision	$\alpha = \mathbb{P}_{(H_0)}(T \in \mathcal{R})$ erreur première espèce mauvaise décision
(H_1) vraie	$\beta = \mathbb{P}_{(H_1)}(T \notin \mathcal{R})$ erreur seconde espèce mauvaise décision	$\pi = 1 - \beta$ puissance bonne décision

- Le rejet de (H_0) assure-t-il que (H_1) est vraie ?
- L'acceptation de (H_0) assure-t-elle que (H_0) est vraie ?
- Du point de vue du risque de la décision, est-il préférable de conserver (H_0) ?

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Risques d'un test (rappel)

Estimer $h(\theta)$

h linéaire

Delta-méthode

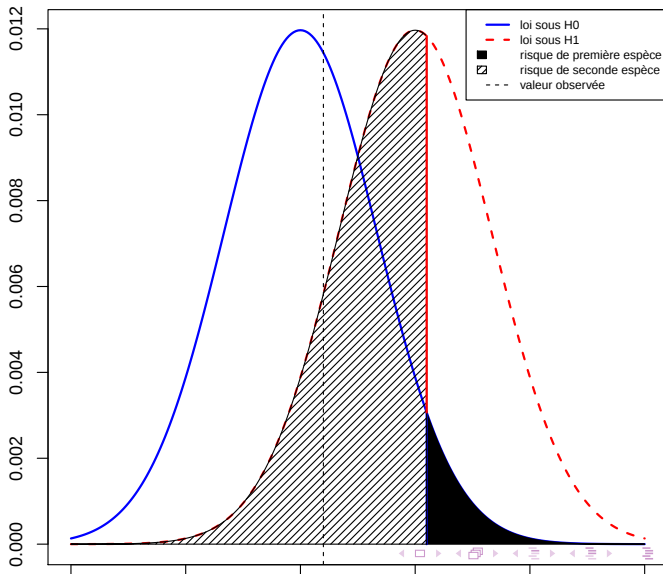
Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald



Définition

Un test est **uniformément plus puissant** (UPP) si, quelle que soit la valeur de $\theta \in \Theta_1$, sa puissance $\pi(\theta)$ est supérieure à la puissance de tout autre test de niveau α .

Théorème (Neyman-Pearson de deux hyp. simples)

La région critique optimale du test de $\theta = \theta_0$ vs $\theta = \theta_1$ de risque α est définie par

$$\mathcal{R}_{opt} = \left\{ x \in \mathbb{R}^n \mid \frac{L(\theta_1; x)}{L(\theta_0; x)} > k_\alpha \right\}; \quad \mathbb{P}(\{X \in \mathcal{R}_{opt}\}) = \alpha$$

Rem $RV(X) = L(\theta_1; x)/L(\theta_0; x)$ est bien une statistique parce que θ_1 et θ_0 sont donnés

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Test optimal de $\theta = \theta_0$ vs $\theta = \theta_1$ au niveau α

Propriétés : le test de Neyman-Pearson est UPP et

- ▶ sans biais
- ▶ constant

Si on dispose de plus d'une statistique **exhaustive** T , la région critique en dépend exclusivement et le test de NP se réduit à une région de rejet de la forme

$$\mathcal{R}_{opt} = \left\{ t \mid \frac{g(\theta_1; t)}{g(\theta_0; t)} > k_\alpha \right\}; \quad \mathbb{P}(\{T \in \mathcal{R}_{opt}\}) = \alpha$$

Exemple : le test de NP de l'espérance pour deux hypothèses simples dans un modèle iid $\mathcal{N}(\theta, 1)$ est le test de la moyenne usuel. Il est indépendant de θ_1 et est UPP.

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Quand l'hypothèse **alternative** est composite : la puissance est une fonction

- ▶ la **puissance** est une **fonction** de θ : pour $\theta_1 \in \Theta_1$,

$$\pi(\theta_1) = \mathbb{P}_{\theta_1}(T \in \mathcal{R}) = 1 - \beta(\theta_1).$$

Quand l'hypothèse **nulle** est composite :

- ▶ le risque de première espèce est une fonction de θ : pour $\theta \in \Theta_0$,

$$\alpha(\theta) = \mathbb{P}_{\theta}(T \in \mathcal{R})$$

- ▶ la **taille** du test est définie par : $\sup_{\theta \in \Theta_0} \alpha(\theta)$
- ▶ Un test est de **niveau** α si sa taille $\leq \alpha$

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Test entre deux hyp. composites au niveau α

Théorème (Lehman : rapport de vraisemblance monotone)

S'il existe une statistique T exhaustive minimale telle que pour tout couple (θ_1, θ_0) le rapport de vraisemblance $L(\theta_1; t)/L(\theta_0; t)$ soit une fonction monotone de t , alors il existe un test UPP pour les situations d'hypothèses unilatérales :

- ▶ $(H_0) : \theta \leq \theta_0$ et RV fct ↗ de $T : \mathcal{R} = \{T > k\}$
- ▶ $(H_0) : \theta \leq \theta_0$ et RV fct ↘ de $T : \mathcal{R} = \{T < k\}$
- ▶ $(H_0) : \theta \geq \theta_0$ et RV fct ↗ de $T : \mathcal{R} = \{T < k\}$
- ▶ $(H_0) : \theta \geq \theta_0$ et RV fct ↘ de $T : \mathcal{R} = \{T > k\}$

↔ cas des statistiques exhaustives des familles exponentielles de loi.

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

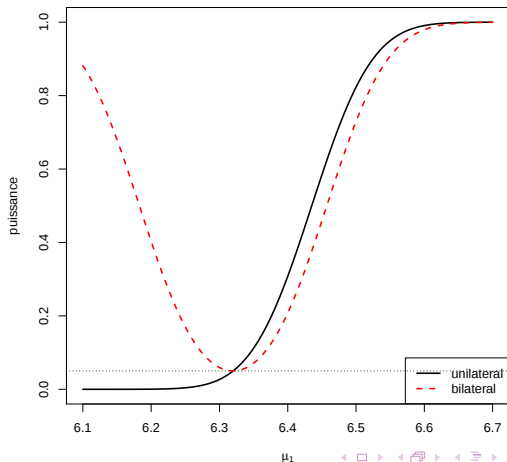
Neyman-Pearson

Test du RV

Test de Wald

un test pas UPP

Dans le modèle $\mathcal{N}(\theta, 1)$, $\theta = \theta_0$ contre $\theta \neq \theta_0$ de région de rejet $\mathcal{R} = \{|T| > k\}$ n'est pas UPP... mais le meilleur parmi les tests sans biais



Test du rapport des vraisemblances maximales

Cas d'hypothèses plus générales

Définition

Soit une famille paramétrique $\mathbb{P}_\theta, \theta \in \Theta$ et les hypothèses $(H_0) : \theta \in \Theta_0$ contre $(H_1) : \theta \in \Theta_1 = \Theta \setminus \Theta_0$. On appelle *rapport de vraisemblance généralisé*, la fonction $RV(X)$ telle que

$$RV(X) = \frac{\sup_{\theta \in \Theta_0} L(\theta; X)}{\sup_{\theta \in \Theta} L(\theta; X)}$$

Le *test du rapport de vraisemblance* (TRV) est le test défini par une région de rejet de la forme

$$\mathcal{R} = \{x; RV(x) < k_\alpha \leq 1\}.$$

Exemple : Test de $\mu = \mu_0$ contre $\mu \neq \mu_0$ dans un modèle gaussien. Lien avec le test classique

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald

Le test TRV n'a pas de propriétés d'optimalité notables, mais on constate dans des situations usuelles qu'il est UPP sans biais.

Théorème (asymptotique du RV)

Soit une famille paramétrique $\mathbb{P}_\theta, \theta \in \Theta$. Si Θ_0 définit une sous-hypothèse linéaire de Θ , $\dim(\Theta_0) = q$, $\dim(\Theta) = p$, et sous les conditions de régularité de l'EMV, alors, sous (H_0)

$$-2 \log(RV) \xrightarrow{\mathcal{L}} \chi^2(p - q)$$

La région de rejet $\{-2 \log(TRV) > q_{\chi^2(p-q)}(1 - \alpha)\}$ du test de rapport de vraisemblances maximales est asymptotiquement de niveau α .

Estimer $h(\theta)$

h linéaire

Delta-méthode

Tests

Intro

Neyman-Pearson

Test du RV

Test de Wald