

MAP 433 : Introduction aux méthodes statistiques. Cours 7

M. Hoffmann

28 mars 2014

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 7

M. Hoffmann

Estimation
optimale :
approche
asymptotique

Modèles
réguliers et
information de
Fisher

Tests
statistiques

1 Estimation optimale : approche asymptotique

2 Modèles réguliers et information de Fisher

- Construction de l'information de Fisher
- Modèle régulier
- Cadre général et interprétation géométrique
- Exemples, applications

3 Tests statistiques

- Notion de test et d'erreur de test
- Hypothèse simple contre alternative simple
- Lemme de Neyman-Pearson

Approche asymptotique

- Hypothèse simplificatrice : $\vartheta \in \Theta \subset \mathbb{R}$. On se restreint aux **estimateurs asymptotiquement normaux** c'est-à-dire vérifiant

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d} \mathcal{N}(0, v(\vartheta))$$

cf. théorèmes limites obtenus pour les Z -, M -estimateurs.

- Si $\hat{\vartheta}_{n,1}$ et $\hat{\vartheta}_{n,2}$ as. normaux de variance asymptotique $v_1(\vartheta) \leq v_2(\vartheta)$, alors la précision de $\hat{\vartheta}_{n,1}$ est **asymptotiquement meilleure** que celle de $\hat{\vartheta}_{n,2}$ au point ϑ :

$$\hat{\vartheta}_{n,1} = \vartheta + \sqrt{\frac{v_1(\vartheta)}{n}} \xi^{(n)}$$

$$\hat{\vartheta}_{n,2} = \vartheta + \sqrt{\frac{v_2(\vartheta)}{n}} \zeta^{(n)}$$

où $\xi^{(n)}$ et $\zeta^{(n)} \xrightarrow{d} \mathcal{N}(0, 1)$.

Comparaison d'estimateurs : cas asymptotique

- Si $v_1(\vartheta) < v_2(\vartheta)$, et si $\vartheta \rightsquigarrow v_i(\vartheta)$ est continue, on pose

$$\mathcal{C}_{n,\alpha}(\hat{\vartheta}_{n,i}) = \left[\hat{\vartheta}_{n,i} \pm \sqrt{\frac{v_i(\hat{\vartheta}_{n,i})}{n}} \Phi^{-1}(1 - \alpha/2) \right], \quad i = 1, 2$$

où $\alpha \in (0, 1)$ et $\Phi(\cdot)$ est la fonction de répartition de la loi normale standard.

- $\mathcal{C}_{n,\alpha}(\hat{\vartheta}_{n,i})$, $i = 1, 2$ sont deux **intervalles de confiance asymptotiquement de niveau $1 - \alpha$** et on a

$$\frac{|\mathcal{C}_{n,\alpha}(\hat{\vartheta}_{n,1})|}{|\mathcal{C}_{n,\alpha}(\hat{\vartheta}_{n,2})|} \xrightarrow{\mathbb{P}_{\vartheta}^n} \sqrt{\frac{v_1(\vartheta)}{v_2(\vartheta)}} < 1.$$

- La notion de **longueur minimale possible d'un intervalle de confiance** est en général difficile à manipuler.

Conclusion provisoire

- Il est **difficile en général** de comparer des estimateurs.
- Cadre asymptotique + normalité asymptotique \rightarrow comparaison de la **variance asymptotique** $\vartheta \rightsquigarrow v(\vartheta)$.
- Sous des hypothèses de régularité du modèle $\{\mathbb{P}_{\vartheta}^n, \vartheta \in \Theta\}$ alors
 - Il **existe** une variance asymptotique $v^*(\vartheta)$ **minimale** parmi les variances de la classe des M -estimateurs as. normaux.
 - Cette fonction est associée à une **quantité d'information intrinsèque** au modèle.
 - La variance asymptotique de l'**EMV** est $v^*(\vartheta)$.
- Ceci règle **partiellement** le problème de l'optimalité.

Régularité d'un modèle statistique et information

- Cadre simplificateur : modèle de densité

$$X_1, \dots, X_n \text{ i.i.d. de loi } \mathbb{P}_\vartheta$$

dans la famille $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$ avec $\Theta \subset \mathbb{R}$ pour simplifier.

- Notation :

$$f(\vartheta, x) = \frac{d\mathbb{P}_\vartheta}{d\mu}(x), \quad x \in \mathbb{R}, \vartheta \in \Theta.$$

- Hypothèse : la quantité

$$\mathbb{I}(\vartheta) = \mathbb{E}_\vartheta \left[(\partial_\vartheta \log f(\vartheta, X))^2 \right]$$

est bien définie.

Définition

- $\mathbb{I}(\vartheta) = \mathbb{E}_{\vartheta} \left[(\partial_{\vartheta} \log f(\vartheta, X))^2 \right]$ s'appelle *l'information de Fisher* de la famille $\{\mathbb{P}_{\vartheta}, \vartheta \in \Theta\}$ au point ϑ . Elle ne dépend pas de la mesure dominante μ .
- Le cadre d'intérêt est celui où

$$0 < \mathbb{I}(\vartheta) < +\infty.$$

- $\mathbb{I}(\vartheta)$ quantifie « l'information » qu'apporte chaque observation X_i sur le paramètre ϑ .

Remarque : on a $\mathbb{P}_{\vartheta} [f(\vartheta, X) > 0] = 1$, donc la quantité $\log f(\vartheta, X)$ est bien définie.

Information dans quel sens ? Origine de la notion

- Supposons l'EMV $\hat{\vartheta}_n^{\text{mv}}$ bien défini et **convergent**.
- Supposons l'application $(\vartheta, x) \rightsquigarrow f(\vartheta, x)$ possédant **toutes les propriétés de régularité et d'intégrabilité** voulues.
- Alors

$$\sqrt{n}(\hat{\vartheta}_n^{\text{mv}} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\vartheta)}\right)$$

en loi sous \mathbb{P}_{ϑ} , où encore

$$\hat{\vartheta}_n^{\text{mv}} \stackrel{d}{\approx} \vartheta + \frac{1}{\sqrt{n\mathbb{I}(\vartheta)}} \mathcal{N}(0, 1)$$

en loi sous \mathbb{P}_{ϑ} .

Construction de l'information + jeu d'hypothèses attendant

- Heuristique : on établira un jeu d'hypothèses justifiant **a posteriori** le raisonnement.
- Etape 1 : l'EMV $\hat{\vartheta}_n^{mv}$ **converge** :

$$\hat{\vartheta}_n^{mv} \xrightarrow{\mathbb{P}_{\vartheta}} \vartheta$$

via le théorème de convergence des M -estimateurs.

- Etape 2 : l'EMV $\hat{\vartheta}_n^{mv}$ est un **Z-estimateur** :

$$0 = \partial_{\vartheta} \left(\sum_{i=1}^n \log f(\vartheta, X_i) \right)_{\vartheta = \hat{\vartheta}_n^{mv}}.$$

Construction de $\mathbb{I}(\vartheta)$ cont.

- Etape 3 : développement asymptotique **autour de ϑ** :

$$0 \approx \sum_{i=1}^n \partial_{\vartheta} \log f(\vartheta, X_i) + (\hat{\vartheta}_n^{\text{mv}} - \vartheta) \sum_{i=1}^n \partial_{\vartheta}^2 \log f(\vartheta, X_i),$$

soit

$$\hat{\vartheta}_n^{\text{mv}} - \vartheta \approx - \frac{\sum_{i=1}^n \partial_{\vartheta} \log f(\vartheta, X_i)}{\sum_{i=1}^n \partial_{\vartheta}^2 \log f(\vartheta, X_i)}$$

- Etape 4 : le numérateur. Normalisation et convergence de $\frac{1}{n} \sum_{i=1}^n \partial_{\vartheta} \log f(\vartheta, X_i)$?

Numérateur

Lemme

On a

$$\mathbb{E}_{\vartheta} [\partial_{\vartheta} \log f(\vartheta, X)] = 0.$$

Preuve.

$$\begin{aligned}\mathbb{E}_{\vartheta} [\partial_{\vartheta} \log f(\vartheta, X)] &= \int_{\mathbb{R}} \partial_{\vartheta} \log f(\vartheta, x) f(\vartheta, x) \mu(dx) \\ &= \int_{\mathbb{R}} \frac{\partial_{\vartheta} f(\vartheta, x)}{f(\vartheta, x)} f(\vartheta, x) \mu(dx) \\ &= \int_{\mathbb{R}} \partial_{\vartheta} f(\vartheta, x) \mu(dx) \\ &= \partial_{\vartheta} \int_{\mathbb{R}} f(\vartheta, x) \mu(dx) = \partial_{\vartheta} 1 = 0.\end{aligned}$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 7

M. Hoffmann

Estimation
optimale :
approche
asymptotique

Modèles
réguliers et
information de
Fisher

Construction de
l'information de
Fisher

Modèle régulier
Cadre général et
interprétation
géométrique
Exemples,
applications

Tests
statistiques



Dénominateur

De même $\int_{\mathbb{R}} \partial_{\vartheta}^2 f(\vartheta, x) \mu(dx) = 0$. **Conséquence :**

$$\mathbb{I}(\vartheta) = \mathbb{E}_{\vartheta} [(\partial_{\vartheta} \log f(\vartheta, X))^2] = -\mathbb{E}_{\vartheta} [\partial_{\vartheta}^2 \log f(\vartheta, X)]$$

En effet

$$\begin{aligned} & \mathbb{E}_{\vartheta} [\partial_{\vartheta}^2 \log f(\vartheta, X)] \\ &= \int_{\mathbb{R}} \frac{\partial_{\vartheta}^2 f(\vartheta, x) f(\vartheta, x) - (\partial_{\vartheta} f(\vartheta, x))^2}{f(\vartheta, x)^2} f(\vartheta, x) \mu(dx) \\ &= \int_{\mathbb{R}} \partial_{\vartheta}^2 f(\vartheta, x) \mu(dx) - \int_{\mathbb{R}} \frac{(\partial_{\vartheta} f(\vartheta, x))^2}{f(\vartheta, x)} \mu(dx) \\ &= 0 - \int_{\mathbb{R}} \left(\frac{\partial_{\vartheta} f(\vartheta, x)}{f(\vartheta, x)} \right)^2 f(\vartheta, x) \mu(dx) = -\mathbb{E} [(\partial_{\vartheta} \log f(\vartheta, X))^2]. \end{aligned}$$

Conséquences

- Les $\partial_{\vartheta} \log f(\vartheta, X_i)$ sont i.i.d. et $\mathbb{E}_{\vartheta} [\partial_{\vartheta} \log f(\vartheta, X)] = 0$.
TCL :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \partial_{\vartheta} \log f(\vartheta, X_i) \xrightarrow{d} \mathcal{N}(0, \mathbb{E}_{\vartheta} [(\partial_{\vartheta} \log f(\vartheta, X))^2]) \\ = \mathcal{N}(0, \mathbb{I}(\vartheta)).$$

- Les $\partial_{\vartheta}^2 \log f(\vartheta, X_i)$ sont i.i.d. LGN :

$$\frac{1}{n} \sum_{i=1}^n \partial_{\vartheta}^2 \log f(\vartheta, X_i) \xrightarrow{\mathbb{P}_{\vartheta}} \mathbb{E}_{\vartheta} [\partial_{\vartheta}^2 \log f(\vartheta, X)] \\ \stackrel{\text{conséquence}}{=} -\mathbb{I}(\vartheta).$$

Conclusion

- En combinant les deux estimations + lemme de Slutsky :

$$\begin{aligned}\sqrt{n}(\hat{\vartheta}_n^{\text{mv}} - \vartheta) &\approx -\frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \partial_{\vartheta} \log f(\vartheta, X_i)}{\frac{1}{n} \sum_{i=1}^n \partial_{\vartheta}^2 \log f(\vartheta, X_i)} \\ &\xrightarrow{d} \frac{\mathcal{N}(0, \mathbb{I}(\vartheta))}{\mathbb{I}(\vartheta)} \\ &\stackrel{\text{loi}}{=} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\vartheta)}\right).\end{aligned}$$

- Le raisonnement est **rigoureux dès lors que** : i) on a la convergence de $\hat{\vartheta}_n^{\text{mv}}$, ii) on peut justifier le lemme et sa conséquence, iii) $\mathbb{I}(\vartheta)$ est bien définie et non dégénérée et iv) on sait contrôler le terme de reste dans le développement asymptotique, **partie la plus difficile**.

Modèle régulier

Définition

La famille de densités $\{f(\vartheta, \cdot), \vartheta \in \Theta\}$, par rapport à la mesure dominante μ , $\Theta \subset \mathbb{R}$, est **régulière** si

- Θ ouvert et $\{f(\vartheta, \cdot) > 0\} = \{f(\vartheta', \cdot) > 0\}$, $\forall \vartheta, \vartheta' \in \Theta$.
- μ -p.p. $\vartheta \rightsquigarrow f(\vartheta, \cdot)$, $\vartheta \rightsquigarrow \log f(\vartheta, \cdot)$ sont \mathcal{C}^2 .
- $\forall \vartheta \in \Theta, \exists \mathcal{V}_\vartheta \subset \Theta$ t.q. pour $a \in \mathcal{V}_\vartheta$

$$|\partial_a^2 \log f(a, x)| + |\partial_a \log f(a, x)| + (\partial_a \log f(a, x))^2 \leq g(x)$$

où

$$\int_{\mathbb{R}} g(x) \sup_{a \in \mathcal{V}(\vartheta)} f(a, x) \mu(dx) < +\infty.$$

- L'information de Fisher est non-dégénérée :

$$\forall \vartheta \in \Theta, \mathbb{I}(\vartheta) > 0.$$

Résultat principal

Proposition

- Si l'expérience engendrée par l'observation $X_1, \dots, X_n \sim_{i.i.d.} \mathbb{P}_\vartheta$ est associée à une famille de probabilités $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$ sur \mathbb{R} **régulière** au sens de la définition précédente, alors

$$\sqrt{n}(\hat{\vartheta}_n^{\text{mv}} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\vartheta)}\right).$$

- Si $\hat{\vartheta}_n$ est un Z -estimateur **régulier** asymptotiquement normal de variance $v(\vartheta)$, alors

$$\forall \vartheta \in \Theta, \quad v(\vartheta) \geq \frac{1}{\mathbb{I}(\vartheta)}.$$

Preuve de la proposition

- Le premier point consiste à **rendre rigoureux** le raisonnement précédent. **Point délicat** : le contrôle du terme de reste.
- **Optimalité de la variance de l'EMV parmi celle des Z-estimateurs** : on a vu que si $\hat{\vartheta}_n$ est un Z-estimateur régulier associé à la fonction ϕ , alors, sa variance asymptotique $v(\vartheta) = v_\phi(\vartheta)$ vaut

$$v_\phi(\vartheta) = \frac{\mathbb{E}_\vartheta [\phi(\vartheta, X)^2]}{(\mathbb{E}_\vartheta [\partial_\vartheta \phi(\vartheta, X)])^2}.$$

- **A montrer** : pour toute fonction ϕ :

$$\boxed{\frac{\mathbb{E}_\vartheta [\phi(\vartheta, X)^2]}{(\mathbb{E}_\vartheta [\partial_\vartheta \phi(\vartheta, X)])^2} \geq \frac{1}{\mathbb{I}(\vartheta)}}.$$

Preuve de l'inégalité

- Par construction

$$\partial_a \mathbb{E}_\vartheta [\phi(a, X)]|_{a=\vartheta} = 0.$$

- (avec $\dot{\phi}(\vartheta, x) = \partial_\vartheta \phi(\vartheta, x)$)

$$\begin{aligned} 0 &= \int_{\mathbb{R}} [\dot{\phi}(\vartheta, x) f(\vartheta, x) + \phi(\vartheta, x) \partial_\vartheta f(\vartheta, x)] \mu(dx) \\ &= \int_{\mathbb{R}} [\dot{\phi}(\vartheta, x) f(\vartheta, x) + \phi(\vartheta, x) \partial_\vartheta \log f(\vartheta, x) f(\vartheta, x)] \mu(dx). \end{aligned}$$

- Conclusion

$$\mathbb{E}_\vartheta [\dot{\phi}(\vartheta, X)] = - \mathbb{E}_\vartheta [\phi(\vartheta, X) \partial_\vartheta \log f(\vartheta, X)]$$

Preuve de l'inégalité (fin)

- On a

$$\mathbb{E}_{\vartheta} [\dot{\phi}(\vartheta, X)] = -\mathbb{E}_{\vartheta} [\phi(\vartheta, X) \partial_{\vartheta} \log f(\vartheta, X)]$$

- Cauchy-Schwarz :

$$(\mathbb{E}_{\vartheta} [\dot{\phi}(\vartheta, X)])^2 \leq \mathbb{E}_{\vartheta} [\phi(\vartheta, X)^2] \mathbb{E}_{\vartheta} [(\partial_{\vartheta} \log f(\vartheta, X))^2],$$

c'est-à-dire

$$v_{\phi}(\vartheta)^{-1} = \frac{(\mathbb{E}_{\vartheta} [\dot{\phi}(\vartheta, X)])^2}{\mathbb{E}_{\vartheta} [\phi(\vartheta, X)^2]} \leq \mathbb{I}(\vartheta).$$

Information de Fisher dans un modèle général

Définition

- **Situation** : suite d'expériences statistiques

$$\mathcal{E}^n = (\mathfrak{Z}^n, \mathcal{Z}^n, \{\mathbb{P}_\vartheta^n, \vartheta \in \Theta\})$$

dominées par μ_n , associées à l'observation $Z^{(n)}$,

$$f_n(\vartheta, z) = \frac{d\mathbb{P}_\vartheta^n}{d\mu^n}(z), \quad z \in \mathfrak{Z}^n, \vartheta \in \Theta \subset \mathbb{R}.$$

- **Information de Fisher** (si elle existe) de l'expérience au point ϑ :

$$\mathbb{I}(\vartheta | \mathcal{E}_n) = \mathbb{E}_\vartheta^n [(\partial_\vartheta \log f_n(\vartheta, Z^{(n)}))^2]$$

Le cas multidimensionnel

- **Même contexte** que précédemment, avec $\Theta \subset \mathbb{R}^d$, et $d \geq 1$.
- **Matrice d'information de Fisher**

$$\mathbb{I}(\vartheta) = \mathbb{E}_{\vartheta} \left[\nabla_{\vartheta} \log f(\vartheta, Z^n) \nabla_{\vartheta} \log f(\vartheta, Z^n)^T \right]$$

matrice symétrique positive.

- Si $\mathbb{I}(\vartheta)$ définie et si \mathcal{E}^n **modèle de densité**, en généralisant à la dimension d les conditions de régularité, on a

$$\sqrt{n}(\hat{\vartheta}_n^{\text{mv}} - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{I}(\vartheta)^{-1}\right).$$

Interprétation géométrique

- On pose $\mathbb{D}(a, \vartheta) = \mathbb{E}_{\vartheta} [\log f(a, X)]$. On a vu (inégalité d'entropie) que

$$\begin{aligned}\mathbb{D}(a, \vartheta) &= \int_{\mathbb{R}} \log f(a, x) f(\vartheta, x) \mu(dx) \\ &\leq \int_{\mathbb{R}} \log f(\vartheta, x) f(\vartheta, x) \mu(dx) = \mathbb{D}(\vartheta, \vartheta).\end{aligned}$$

- On a

$$\mathbb{I}(\vartheta) = \partial_a^2 \mathbb{D}(a, \vartheta) \big|_{a=\vartheta}.$$

- Si $\mathbb{I}(\vartheta)$ est « petite », le **rayon de courbure de $a \rightsquigarrow \mathbb{D}(a, \vartheta)$ est grand** dans un voisinage de ϑ : la stabilisation d'un maximum empirique (l'EMV) est plus difficile, rendant moins précis l'estimation.
- Si $\mathbb{I}(\vartheta)$ est « grande », le **rayon de courbure est petit** et le maximum de l'EMV est mieux localisé.

Information de Fisher et régression

- \mathcal{E}^n expérience engendrée par $(\mathbf{x}_1, Y_1), \dots, (\mathbf{x}_n, Y_n)$ avec

$$Y_i = r(\vartheta, \mathbf{x}_i) + \xi_i,$$

ξ_i : densité **g par rapport à la mesure de Lebesgue** +
« design » déterministe.

- Observation : $Z^n = (Y_1, \dots, Y_n)$, $\mu^n = dy_1 \dots dy_n$,
 $z = (y_1, \dots, y_n)$ et

$$f_n(\vartheta, Z^n) = \prod_{i=1}^n g(Y_i - r(\vartheta, \mathbf{x}_i))$$

- **Information de Fisher**

$$\mathbb{I}(\vartheta | \mathcal{E}^n) = \mathbb{E}_{\vartheta} [(\partial_{\vartheta} \log f_n(\vartheta, Z^n))^2]$$

Information de Fisher et régression

- **Formule explicite** pour la log-vraisemblance

$$\partial_{\vartheta} \log f_n(\vartheta, Z^n) = \sum_{i=1}^n \partial_{\vartheta} \log g(Y_i - r(\vartheta, \mathbf{x}_i))$$

- **Propriété analogue avec le modèle de densité :**
 $\mathbb{E}_{\vartheta} [\partial_{\vartheta} \log g(Y_i - r(\vartheta, \mathbf{x}_i))] = 0.$
- **Information de Fisher** par indépendance + centrage :

$$\begin{aligned} \mathbb{I}(\vartheta | \mathcal{E}^n) &= \sum_{i=1}^n \mathbb{E}_{\vartheta}^n [(\partial_{\vartheta} \log g(Y_i - r(\vartheta, \mathbf{x}_i)))^2] \\ &= \dots \end{aligned}$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 7

M. Hoffmann

Estimation
optimale :
approche
asymptotique

Modèles
réguliers et
information de
Fisher

Construction de
l'information de
Fisher

Modèle régulier
Cadre général et
interprétation
géométrique

Exemples,
applications

Tests
statistiques

Exemples et applications

A **titre d'exercice**, savoir calculer l'information de Fisher pour :

- L'estimation du paramètre d'une loi de Poisson dans le modèle de densité.
- L'estimation de la moyenne-variance pour un échantillon gaussien.
- **La régression logistique**
- L'estimation du paramètre d'une loi exponentielle **avec ou sans** censure.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 7

M. Hoffmann

Estimation
optimale :
approche
asymptotique

Modèles
réguliers et
information de
Fisher

Construction de
l'information de
Fisher

Modèle régulier
Cadre général et
interprétation
géométrique

Exemples,
applications

Tests
statistiques

Efficacité à un pas

- Dans un modèle régulier, le **calcul numérique** de l'EMV peut être difficile à réaliser.
- Si l'on dispose d'un estimateur $\hat{\vartheta}_n$ **asymptotiquement normal** et si les évaluations

$$\ell'_n(\vartheta) = \frac{1}{n} \sum_{i=1}^n \partial_{\vartheta} \log f(\vartheta, X_i), \quad \ell''_n(\vartheta) = \frac{1}{n} \sum_{i=1}^n \partial_{\vartheta}^2 \log f(\vartheta, X_i)$$

sont **faciles**, alors on peut **corriger** $\hat{\vartheta}_n$ de sorte d'avoir le même comportement asymptotique que l'EMV :

$$\tilde{\vartheta}_n = \hat{\vartheta}_n - \frac{\ell'_n(\hat{\vartheta}_n)}{\ell''_n(\hat{\vartheta}_n)} \quad (\text{algorithme de Newton})$$

satisfait

$$\sqrt{n}(\tilde{\vartheta}_n - \vartheta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\vartheta)}\right)$$

Exemple introductif

- On observe 10 lancers d'une pièce de monnaie et on obtient le résultat suivant :

$$(P, P, F, F, P, F, P, P, F, P).$$

La pièce est-elle équilibrée ?

- Répondre à cette question revient à **construire une procédure de décision** :

$$\varphi = \varphi(P, P, F, F, P, F, P, P, F, P)$$

$$= \begin{cases} 0 & \text{on accepte l'hypothèse « la pièce est équilibrée »} \\ 1 & \text{on rejette l'hypothèse « la pièce est équilibrée »} \end{cases}$$

- On associe l'expérience statistique (par exemple)

$$\mathcal{E}^{10} = (\{0, 1\}^{10}, \text{parties de}(\{0, 1\}^{10}), \{\mathbb{P}_\vartheta^{10}, \vartheta \in [0, 1]\}),$$

avec $(P = 0, F = 1)$

$$\mathbb{P}_\vartheta^{10} = (\vartheta \delta_0(dx) + (1 - \vartheta) \delta_1(dx))^{\otimes 10}.$$

- Hypothèse nulle : « la pièce est équilibrée »

$$H_0 : \vartheta = \frac{1}{2}$$

- Hypothèse alternative : « la pièce est truquée »

$$H_1 : \vartheta \neq \frac{1}{2}$$

Résolution (cont.)

- On note Z l'observation.
- On **construit** une **règle de décision simple** :

$$\varphi = 1_{\{Z \in \mathcal{R}\}} = \begin{cases} 0 & \text{on accepte l'hypothèse} \\ 1 & \text{on rejette l'hypothèse.} \end{cases}$$

- $\mathcal{R} \subset \mathfrak{Z}$ (espace des observables) : **zone de rejet** ou **région critique**.
- Exemple¹

$$\mathcal{R} = \{|\hat{\vartheta}(Z) - \tfrac{1}{2}| > t_0\}, \quad \hat{\vartheta}(Z) = \hat{\vartheta}_n^{\text{mv}} \left(\stackrel{\text{exemple}}{=} 0, 6 \right)$$

où t_0 est un seuil à choisir... **Comment ?**

1. léger abus de notation...

Erreur de décision

- Lorsque l'on prend la décision φ , on peut se tromper de deux manières :

Rejeter H_0 ($\varphi = 1$) alors que $\vartheta = \frac{1}{2}$

ou encore

Accepter H_0 ($\varphi = 0$) alors que $\vartheta \neq \frac{1}{2}$.

- Erreur de première espèce (=rejeter à tort)

$$\mathbb{P}_{\frac{1}{2}}^{10} [\varphi = 1]$$

- Erreur de seconde espèce (=accepter à tort)

$$(\mathbb{P}_{\vartheta}^{10} [\varphi = 0], \vartheta \neq \frac{1}{2}).$$

Conclusion provisoire

- Un « bon test » φ doit garantir **simultanément** des erreurs de première et seconde espèce **petites**.
- **Un test optimal existe-t-il ?**
- Si **non**, comment aborder la notion d'optimalité et comment construire un test optimal ?

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 7

M. Hoffmann

Estimation
optimale :
approche
asymptotique

Modèles
réguliers et
information de
Fisher

Tests
statistiques

Notion de test et
d'erreur de test

Hypothèse
simple contre
alternative
simple

Lemme de
Neyman-Pearson

Définition formelle

- Situation : $\mathcal{E} = (\mathcal{Z}, \mathfrak{Z}, \{\mathbb{P}_\vartheta, \vartheta \in \Theta\})$ engendrée par l'observation Z .
- **Hypothèse nulle et alternative** : $\Theta_0 \subset \Theta$ et $\Theta_1 \subset \Theta$ t.q.

$$\Theta_0 \cap \Theta_1 = \emptyset.$$

Définition (Test simple)

*Un test (simple) de l'hypothèse nulle $H_0 : \vartheta \in \Theta_0$ contre l'alternative $H_1 : \vartheta \in \Theta_1$ est une statistique $\varphi = \varphi(Z) \in \{0, 1\}$. (Fonction d') **erreur de première espèce** :*

$$\vartheta \in \Theta_0 \rightsquigarrow \mathbb{P}_\vartheta [\varphi = 1]$$

*(Fonction d') **erreur de seconde espèce***

$$\vartheta \in \Theta_1 \rightsquigarrow \mathbb{P}_\vartheta [\varphi = 0] = 1 - \text{puissance}_\varphi(\vartheta).$$

Hypothèse simple contre alternative simple

- Cas où $\Theta = \{\vartheta_0, \vartheta_1\}$ avec $\vartheta_0 \neq \vartheta_1$.
- Existe-t-il un test φ^* **optimal**, au sens où : $\forall \varphi$ test simple, on a **simultanément**

$$\mathbb{P}_{\vartheta_0} [\varphi^* = 1] \leq \mathbb{P}_{\vartheta_0} [\varphi = 1]$$

et

$$\mathbb{P}_{\vartheta_1} [\varphi^* = 0] \leq \mathbb{P}_{\vartheta_1} [\varphi = 0] \quad ?$$

- Si \mathbb{P}_{ϑ_0} et \mathbb{P}_{ϑ_1} ne sont pas **étrangères** (cf. Cours 6) un tel test φ^* **ne peut pas exister**.

Absence d'optimalité stricte

- **Equivalence** tests simples \iff estimateurs $\hat{\vartheta}$ de ϑ **via la représentation** :

$$\hat{\vartheta} = \vartheta_0 1_{\mathcal{R}^c} + \vartheta_1 1_{\mathcal{R}} \iff \varphi = 1_{\mathcal{R}}.$$

- **Fonction de risque**

$$\mathcal{P}(\varphi, \vartheta) = \mathbb{E}_{\vartheta} [1_{\hat{\vartheta} \neq \vartheta}], \quad \vartheta = \vartheta_0, \vartheta_1.$$

- La fonction de perte $\ell(\hat{\vartheta}, \vartheta) = 1_{\hat{\vartheta} \neq \vartheta}$ joue le même rôle que la perte quadratique $(\hat{\vartheta} - \vartheta)^2$ dans le Cours 6.
- Test optimal $\varphi^* \iff$ estimateur optimal ϑ^* pour \mathcal{P} .
- **Comme pour le cas du risque quadratique**, dès que \mathbb{P}_{ϑ_0} et \mathbb{P}_{ϑ_1} ne sont pas étrangères, un estimateur optimal **n'existe pas** (cf. Cours 6).

Riposte : principe de Neyman

- On « **disymétrise** » les hypothèses H_0 et H_1 : H_0 est « plus importante » que H_1 dans le sens suivant : on **impose** une **erreur de première espèce prescrite**.

Définition

Pour $\alpha \in [0, 1]$, un test $\varphi = \varphi_\alpha$ de l'hypothèse nulle $H_0 : \vartheta \in \Theta_0$ contre une alternative H_1 est de niveau α si

$$\sup_{\vartheta \in \Theta_0} \mathbb{P}_\vartheta [\varphi_\alpha = 1] \leq \alpha.$$

- Un test de niveau α ne dit **rien** sur l'erreur de seconde espèce (comportement sur l'alternative).

Principe de Neyman (cont.)

- Choix de la « disymétrisation » = choix de modélisation.
- Principe de Neyman : $\alpha \in (0, 1)$, parmi les test de niveau α , chercher celui (ou ceux) ayant **une erreur de seconde espèce minimale**.

Définition

*Un test de niveau α est dit **Uniformément Plus Puissant (UPP)** si son erreur de seconde espèce est minimale parmi celles des tests de niveau α .*

- Pour le cas d'une **hypothèse simple** contre une **alternative simple**, un test UPP existe.

Lemme de Neyman-Pearson

Proposition

Soit $\alpha \in [0, 1]$. S'il existe c_α solution de

$$\mathbb{P}_{\vartheta_0} [f(\vartheta_1, Z) > c_\alpha f(\vartheta_0, Z)] = \alpha$$

alors le test de région critique $\mathcal{R}_\alpha = \{f(\vartheta_1, Z) > c_\alpha f(\vartheta_0, Z)\}$ est de niveau α et UPP pour tester $H_0 : \vartheta = \vartheta_0$ contre $H_1 : \vartheta = \vartheta_1$.

- Si $U = f(\vartheta_1, Z)/f(\vartheta_0, Z)$ bien définie et $\mathcal{L}(U) \ll dx$ (sous \mathbb{P}_{ϑ_0}), alors $\mathbb{P}_{\vartheta_0} [U > c_\alpha] = \alpha$ admet une solution.

Exemple de mise en oeuvre

- On observe

$$Z = (X_1, \dots, X_n) \sim_{\text{i.i.d.}} \mathcal{N}(\vartheta, 1).$$

- Construction du test de N-P. de $H_0 : \vartheta = \vartheta_0$ contre $H_1 : \vartheta = \vartheta_1$, avec $\vartheta_0 < \vartheta_1$.

- **Mesure dominante** $\mu^n =$ mesure de Lebesgue sur \mathbb{R}^n et

$$f(\vartheta, Z) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n X_i^2 + n\vartheta \bar{X}_n - \frac{n\vartheta^2}{2}\right).$$

- Rapport de vraisemblance

$$\frac{f(\vartheta_1, Z)}{f(\vartheta_0, Z)} = \exp \left(n(\vartheta_1 - \vartheta_0) \overline{X}_n - \frac{n}{2} (\vartheta_1^2 - \vartheta_0^2) \right).$$

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 7

M. Hoffmann

Lemme de
Neyman-Pearson

Exemple (cont.)

- **Zone de rejet** du test de N-P. :

$$\begin{aligned} & \{f(\vartheta_1, Z) > cf(\vartheta_0, Z)\} \\ &= \left\{ n(\vartheta_1 - \vartheta_0) \bar{X}_n - \frac{n}{2}(\vartheta_1^2 - \vartheta_0^2) > \log c \right\} \\ &= \left\{ \bar{X}_n > \frac{\vartheta_0 + \vartheta_1}{2} + \frac{\log c}{n(\vartheta_0 - \vartheta_1)} \right\}. \end{aligned}$$

- **Choix de c** . On résout

$$\mathbb{P}_{\vartheta_0} \left[\bar{X}_n > \frac{1}{2}(\vartheta_0 + \vartheta_1) + \frac{\log c}{n(\vartheta_0 - \vartheta_1)} \right] = \alpha.$$

- **Approche standard** : on raisonne sous \mathbb{P}_{ϑ_0} . On a

$$\bar{X}_n = \vartheta_0 + \frac{1}{\sqrt{n}} \xi^{n, \vartheta_0},$$

où ξ^{n, ϑ_0} est une gaussienne standard $\mathcal{N}(0, 1)$ sous \mathbb{P}_{ϑ_0}
mais pas sous une autre probabilité \mathbb{P}_{ϑ} si $\vartheta \neq \vartheta_0$!

Exemple (fin)

■ Résolution de

$$\mathbb{P}_{\vartheta_0} \left[\vartheta_0 + \frac{1}{\sqrt{n}} \xi^{n, \vartheta_0} > \frac{1}{2}(\vartheta_0 + \vartheta_1) + \frac{\log c}{n(\vartheta_0 - \vartheta_1)} \right] = \alpha.$$

- Equivalent à $\mathbb{P}_{\vartheta_0} \left[\xi^{n, \vartheta_0} > \frac{\sqrt{n}}{2}(\vartheta_1 - \vartheta_0) + \frac{1}{\sqrt{n}} \frac{\log c}{\vartheta_0 - \vartheta_1} \right] = \alpha$,
soit

$$\frac{\sqrt{n}}{2}(\vartheta_1 - \vartheta_0) + \frac{1}{\sqrt{n}} \frac{\log c}{\vartheta_0 - \vartheta_1} = \Phi^{-1}(1 - \alpha),$$

où $\Phi(x) = \int_{-\infty}^x e^{-u^2/2} \frac{du}{\sqrt{2\pi}}.$

■ Conclusion

$$c_\alpha = \exp \left(n \frac{(\vartheta_1 - \vartheta_0)^2}{2} + \sqrt{n}(\vartheta_0 - \vartheta_1) \Phi^{-1}(1 - \alpha) \right)$$

Bilan provisoire

- Si l'on accepte **le principe de Neyman**, on sait résoudre le problème à deux points.
- Que faire si l'hypothèse nulle H_0 ou l'alternative H_1 sont **composites** ?
 - On peut proposer des extensions si l'on dispose de structures particulières sur la vraisemblance du modèle (Poly. Ch. 7.3, hors programme).
 - On sait dire **beaucoup de choses** dans le cas gaussien.
- **Critique méthodologique de l'approche de Neyman** \rightsquigarrow notion de p -valeur.
- On ne sait **toujours pas** répondre à la question de l'exemple introductif... **cadre asymptotique** Cours 8.

MAP 433 :
Introduction
aux méthodes
statistiques.
Cours 7

M. Hoffmann

Estimation
optimale :
approche
asymptotique

Modèles
réguliers et
information de
Fisher

Tests
statistiques

Notion de test et
d'erreur de test
Hypothèse
simple contre
alternative
simple

Lemme de
Neyman-Pearson