MAP 433 : Introduction aux méthodes statistiques. Cours 7

M. Hoffmann

28 mars 2014

Aujourd'hui

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
$$\sqrt{n}\big(\widehat{\vartheta}_n - \vartheta\big) \xrightarrow{d} \mathcal{N}\big(0, v(\vartheta)\big)$$
cf. théorèmes limites obtenus pour les Z-,M-estimateurs.
- Si $\widehat{\vartheta}_{n,1}$ et $\widehat{\vartheta}_{n,2}$ as. normaux de variance asymptotique $v_1(\vartheta) \leq v_2(\vartheta)$, alors la précision de $\widehat{\vartheta}_{n,1}$ est asymptotiquement meilleure que celle de $\widehat{\vartheta}_{n,2}$ au point ϑ :
$$\widehat{\vartheta}_{n,1} = \vartheta + \sqrt{\frac{v_1(\vartheta)}{n}}\xi^{(n)}$$
$$\widehat{\vartheta}_{n,2} = \vartheta + \sqrt{\frac{v_2(\vartheta)}{n}}\zeta^{(n)}$$
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

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

  où $\alpha \in (0,1)$ et $\Phi(\cdot)$ est la fonction de répartition de la loi normale standard.
- $\mathcal{C}_{n,\alpha}(\widehat{\vartheta}_{n,i})$, $i = 1, 2$ sont deux intervalles de confiance asymptotiquement de niveau $1 - \alpha$ et on a

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

- 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 $v^\star(\vartheta)$ 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 $v^\star(\vartheta)$.
- Ceci règle partiellement le problème de l'optimalité.

Régularité d'un modèle statistique et information

- Cadre simplificateur : modèle de densité

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

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

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

est bien définie.

Information de Fisher

Définition

- $\mathbb{I}(\vartheta) = \mathbb{E}_\vartheta\left[\big(\partial_\vartheta \log f(\vartheta, X)\big)^2\right]$ *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 $\mathbb{P}_\vartheta\left[f(\vartheta, X) > 0\right] = 1$, donc la quantité $\log f(\vartheta, X)$ est bien définie.

Information dans quel sens ? Origine de la notion

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

$$\boxed{\sqrt{n}\big(\widehat{\vartheta}_n^{\,\mathtt{mv}} - \vartheta\big) \xrightarrow{d} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\vartheta)}\Big)}$$

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

$$\widehat{\vartheta}_n^{\,\mathtt{mv}} \overset{d}{\approx} \vartheta + \frac{1}{\sqrt{n\mathbb{I}(\vartheta)}}\mathcal{N}(0,1)$$

en loi sous $\mathbb{P}_\vartheta$.

Construction de l'information + jeu d'hypothèses attenant

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

$$\widehat{\vartheta}_n^{\,\mathtt{mv}} \xrightarrow{\mathbb{P}_\vartheta} \vartheta$$

  via le théorème de convergence des M-estimateurs.
- Etape 2 : l'EMV $\widehat{\vartheta}_n^{\,\mathtt{mv}}$ est un Z-estimateur :

$$0 = \partial_\vartheta \Big(\sum_{i=1}^n \log f(\vartheta, X_i) \Big)_{\vartheta = \widehat{\vartheta}_n^{\,\mathtt{mv}}}.$$

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

- Etape 3 : développement asymptotique autour de ϑ :

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

soit

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

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

Numérateur

Lemme

On a

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

Preuve.

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

Dénominateur

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

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

En effet

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

$$= \int_{\mathbb{R}} \frac{\partial_\vartheta^2 f(\vartheta, x) f(\vartheta, x) - \big(\partial_\vartheta f(\vartheta, x)\big)^2}{f(\vartheta, x)^2} f(\vartheta, x)\mu(dx)$$

$$= \int_{\mathbb{R}} \partial_\vartheta^2 f(\vartheta, x)\mu(dx) - \int_{\mathbb{R}} \frac{\big(\partial_\vartheta f(\vartheta, x)\big)^2}{f(\vartheta, x)}\mu(dx)$$

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

Conséquences

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

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

$$= \mathcal{N}\big(0, \mathbb{I}(\vartheta)\big).$$

- Les $\partial^2_\vartheta \log f(\vartheta, X_i)$ sont i.i.d. LGN :

$$\frac{1}{n}\sum_{i=1}^{n} \partial^2_\vartheta \log f(\vartheta, X_i) \xrightarrow{\mathbb{P}_\vartheta} \mathbb{E}_\vartheta\left[\partial^2_\vartheta \log f(\vartheta, X)\right]$$

$$\overset{\text{conséquence}}{=} -\mathbb{I}(\vartheta).$$

Conclusion

- En combinant les deux estimations + lemme de Slutsky :

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

$$\xrightarrow{d} \frac{\mathcal{N}\big(0, \mathbb{I}(\vartheta)\big)}{\mathbb{I}(\vartheta)}$$

$$\stackrel{\text{loi}}{=} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\vartheta)}\Big).$$

- Le raisonnement est rigoureux dès lors que : i) on a la convergence de $\widehat{\vartheta}_n^{\,\mathtt{mv}}$, 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* $a\in\mathcal{V}_\vartheta$

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

  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* $X_1, \ldots, X_n \sim_{i.i.d.} \mathbb{P}_\vartheta$ *est associée à une famille de probabilités* $\{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$ *sur* $\mathbb{R}$ *régulière au sens de la définition précédente, alors*

$$\sqrt{n}\big(\widehat{\vartheta}_{\mathtt{n}}^{\,\mathtt{mv}} - \vartheta\big) \xrightarrow{d} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\vartheta)}\Big).$$

- *Si* $\widehat{\vartheta}_n$ *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 $\widehat{\vartheta}_n$ est un Z-estimateur régulier associé à la fonction ϕ, alors, sa variance asymptotique $v(\vartheta) = v_\phi(\vartheta)$ vaut

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

- A montrer : pour toute fonction ϕ :

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

Preuve de l'inégalité

- Par construction

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

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

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

- Conclusion

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

Preuve de l'inégalité (fin)

- On a

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

- Cauchy-Schwarz :

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

c'est-à-dire

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

Information de Fisher dans un modèle général

Définition

- *Situation : suite d'expériences statistiques*

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

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

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

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

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

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

$$\sqrt{n}\left(\widehat{\vartheta}_{\mathrm{n}}^{\,\mathtt{mv}} - \vartheta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{I}(\vartheta)^{-1}\right).$$

Interprétation géométrique

- On pose $\mathbb{D}(a,\vartheta) = \mathbb{E}_\vartheta\left[\log f(a,X)\right]$. 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

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

  - 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 $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_n, Y_n)$ avec

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

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

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

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

- Information de Fisher

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

Information de Fisher et régression

- Formule explicite pour la log-vraisemblance

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

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

$$\mathbb{I}(\vartheta|\mathcal{E}^n) = \sum_{i=1}^n \mathbb{E}_\vartheta^n\left[\Big(\partial_\vartheta \log g\big(Y_i - r(\vartheta, \mathbf{x}_i)\big)\Big)^2\right]$$

$$= \ldots$$

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.

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 $\widehat{\vartheta}_n$ 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 $\widehat{\vartheta}_n$ de sorte d'avoir le même comportement asymptotique que l'EMV :

$$\widetilde{\vartheta}_n = \widehat{\vartheta}_n - \frac{\ell_n'(\widehat{\vartheta}_n)}{\ell_n''(\widehat{\vartheta}_n)} \quad \text{(algorithme de Newton)}$$

satisfait

$$\boxed{\sqrt{n}\big(\widetilde{\vartheta}_n - \vartheta\big) \xrightarrow{d} \mathcal{N}\Big(0, \frac{1}{\mathbb{I}(\vartheta)}\Big)}$$

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 :

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

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

Résolution

- On associe l'expérience statistique (par exemple)

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

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

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

- Hypothèse nulle : « la pièce est équilibrée »

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

- Hypothèse alternative : « la pièce est truquée »

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

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.
- Exemple[1]

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

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 :

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

ou encore

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

- Erreur de première espèce (=rejeter à tort)

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

- Erreur de seconde espèce (=accepter à tort)

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

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 ?

Définition formelle

- Situation : $\mathcal{E} = \big(\mathcal{Z}, \mathfrak{Z}, \{\mathbb{P}_\vartheta, \vartheta \in \Theta\}\big)$ 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 :

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

(Fonction d') erreur de seconde espèce

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

Hypothèse simple contre alternative simple

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

$$\mathbb{P}_{\vartheta_0}\left[\varphi^\star = 1\right] \leq \mathbb{P}_{\vartheta_0}\left[\varphi = 1\right]$$

et

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

- Si $\mathbb{P}_{\vartheta_0}$ et $\mathbb{P}_{\vartheta_1}$ ne sont pas étrangères (cf. Cours 6) un tel test $\varphi^\star$ ne peut pas exister.

Absence d'optimalité stricte

- Equivalence tests simples $\rightsquigarrow$ estimateurs $\widehat{\vartheta}$ de ϑ via la représentation :

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

- Fonction de risque

$$\mathcal{P}(\varphi, \vartheta) = \mathbb{E}_\vartheta \left[1_{\widehat{\vartheta} \neq \vartheta} \right], \;\; \vartheta = \vartheta_0, \vartheta_1.$$

- La fonction de perte $\ell(\widehat{\vartheta}, \vartheta) = 1_{\widehat{\vartheta} \neq \vartheta}$ joue le même rôle que la perte quadratique $(\widehat{\vartheta} - \vartheta)^2$ dans le Cours 6.
- Test optimal $\varphi^\star \leftrightsquigarrow$ estimateur optimal $\vartheta^\star$ pour $\mathcal{P}$.
- Comme pour le cas du risque quadratique, dès que $\mathbb{P}_{\vartheta_0}$ et $\mathbb{P}_{\vartheta_1}$ 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

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

- 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.

Principe de construction

- $f(\vartheta, z) = \frac{d\,\mathbb{P}_\vartheta}{d\mu}(z)$, $z \in \mathfrak{Z}$, $\vartheta = \vartheta_0, \vartheta_1$, μ mesure dominante. L'EMV –si bien défini– s'écrit

$$\widehat{\vartheta}_n^{\,\mathtt{mv}} = \vartheta_0 1_{\{f(\vartheta_1, Z) < f(\vartheta_0, Z)\}} + \vartheta_1 1_{\{f(\vartheta_0, Z) < f(\vartheta_1, Z)\}}.$$

- On choisit une région critique de la forme

$$\boxed{\mathcal{R}(c) = \{f(\vartheta_1, Z) > cf(\vartheta_0, Z)\}, \quad c > 0}$$

et on calibre $c = c_\alpha$ de sorte que

$$\boxed{\mathbb{P}_{\vartheta_0}\big[Z \in \mathcal{R}(c_\alpha)\big] = \alpha.}$$

- Le test ainsi construit (si cette équation admet une solution) est de niveau α. On montre qu'il est UPP.

Lemme de Neyman-Pearson

Proposition

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

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

alors le test de région critique $\mathcal{R}_\alpha = \big\{f(\vartheta_1, Z) > c_\alpha f(\vartheta_0, Z)\big\}$ 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 $\mathbb{P}_{\vartheta_0}$), alors $\mathbb{P}_{\vartheta_0}\big[U > c_\alpha\big] = \alpha$ admet une solution.

Exemple de mise en oeuvre

- On observe

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

- 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 μ^n = mesure de Lebesgue sur $\mathbb{R}^n$ et

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

- Rapport de vraisemblance

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

Exemple (cont.)

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

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

- Choix de c. On résout

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

- Approche standard : on raisonne sous $\mathbb{P}_{\vartheta_0}$. On a

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

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

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

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.