

Mise à niveau en statistique

Correction de la feuille de TD 3 : IC et information de Fisher

M2 Mathématiques de l'Aléatoire – Université Paris-Saclay

Exercice 1 – Information de Fisher et inégalité de Cramér–Rao

Toutes les espérances de cet exercice sont prises sous la loi de paramètre θ . On suppose implicitement que les quantités considérées sont intégrables et que $\mathcal{I}(\theta) > 0$.

1. Le score est

$$V = \partial_{\theta} \log f(X, \theta) = \frac{\partial_{\theta} f(X, \theta)}{f(X, \theta)}.$$

Les hypothèses de dérivation sous le signe intégral donnent

$$\mathbb{E}_{\theta}[V] = \int \partial_{\theta} f(x, \theta) dx = \partial_{\theta} \int f(x, \theta) dx = 0.$$

Ainsi,

$$\begin{aligned} \text{Cov}_{\theta}(V, T) &= \mathbb{E}_{\theta}[VT] \\ &= \int T(x) \partial_{\theta} f(x, \theta) dx = \partial_{\theta} \int T(x) f(x, \theta) dx = \psi'(\theta). \end{aligned}$$

2. En dérivant une seconde fois la log-densité,

$$\partial_{\theta}^2 \log f = \frac{\partial_{\theta}^2 f}{f} - \left(\frac{\partial_{\theta} f}{f} \right)^2.$$

Après intégration,

$$\begin{aligned} \mathbb{E}_{\theta}[\partial_{\theta}^2 \log f(X, \theta)] &= \int \partial_{\theta}^2 f(x, \theta) dx - \mathbb{E}_{\theta}[V^2] \\ &= \partial_{\theta}^2 \int f(x, \theta) dx - \mathbb{E}_{\theta}[V^2] = -\mathbb{E}_{\theta}[V^2]. \end{aligned}$$

Comme $\mathbb{E}_{\theta}[V] = 0$,

$$\mathcal{I}(\theta) = \text{Var}_{\theta}(V) = \mathbb{E}_{\theta}[V^2] = -\mathbb{E}_{\theta}[\partial_{\theta}^2 \log f(X, \theta)].$$

3. L'inégalité de Cauchy–Schwarz appliquée à T et V donne

$$\text{Cov}_{\theta}(T, V)^2 \leq \text{Var}_{\theta}(T) \text{Var}_{\theta}(V).$$

En utilisant les deux questions précédentes,

$$\text{Var}_{\theta}(T) \geq \frac{\psi'(\theta)^2}{\mathcal{I}(\theta)}.$$

Lorsque T est sans biais pour θ , on a $\psi(\theta) = \theta$ et la borne devient $1/\mathcal{I}(\theta)$. L'égalité a lieu exactement lorsque $T - \psi(\theta)$ est proportionnel au score dans $L^2(P_{\theta})$.

4. Notons

$$\ell_n(u) = \sum_{i=1}^n \log f(X_i, u).$$

Puisque θ est intérieur à l'espace des paramètres et $\hat{\theta}_n$ est consistant, l'EMV est, avec une probabilité tendant vers un, un maximum intérieur; ainsi $\ell'_n(\hat{\theta}_n) = 0$. La formule de Taylor–Lagrange appliquée à ℓ'_n donne, pour un $\tilde{\theta}_n$ situé entre θ et $\hat{\theta}_n$,

$$0 = \ell'_n(\theta) + (\hat{\theta}_n - \theta)\ell''_n(\tilde{\theta}_n).$$

Par conséquent,

$$\sqrt{n}(\hat{\theta}_n - \theta) = -\frac{n^{-1/2}\ell'_n(\theta)}{n^{-1}\ell''_n(\tilde{\theta}_n)}.$$

Le numérateur est une somme de scores i.i.d., centrés et de variance $\mathcal{I}_1(\theta)$. Le théorème central limite donne donc

$$\frac{1}{\sqrt{n}}\ell'_n(\theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{I}_1(\theta)).$$

Une hypothèse suffisante pour traiter le dénominateur est l'existence d'un voisinage U de θ tel que $u \mapsto \partial_u^2 \log f(x, u)$ soit continue sur U pour presque tout x et

$$\mathbb{E}_\theta \left[\sup_{u \in U} |\partial_u^2 \log f(X_1, u)| \right] < \infty.$$

Une loi uniforme des grands nombres, la consistance de l'EMV et l'identité de Fisher entraînent alors

$$\frac{1}{n}\ell''_n(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} -\mathcal{I}_1(\theta).$$

Le lemme de Slutsky permet de conclure :

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathcal{I}_1(\theta)^{-1}).$$

Exercice 2 – EMV et modèle gaussien

On paramètre le modèle par (μ, v) , où $v = \sigma^2 > 0$.

1. Pour une observation,

$$\ell_{\mu,v}(x) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log v - \frac{(x - \mu)^2}{2v},$$

d'où

$$\partial_\mu \ell_{\mu,v}(x) = \frac{x - \mu}{v}, \quad \partial_v \ell_{\mu,v}(x) = -\frac{1}{2v} + \frac{(x - \mu)^2}{2v^2}.$$

Les deux composantes du score sont non corrélées, car $\mathbb{E}[(X - \mu)^3] = 0$. Comme $\mathbb{E}[(X - \mu)^2] = v$ et $\text{Var}((X - \mu)^2) = 2v^2$,

$$\mathcal{I}_1(\mu, v) = \begin{pmatrix} v^{-1} & 0 \\ 0 & (2v^2)^{-1} \end{pmatrix}.$$

L'information de l'échantillon vaut $\mathcal{I}_n = n\mathcal{I}_1$.

2. La log-vraisemblance de l'échantillon est

$$\ell_n(\mu, v) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log v - \frac{1}{2v} \sum_{i=1}^n (X_i - \mu)^2.$$

Les équations du score donnent

$$\hat{\mu}^{\text{MV}} = \bar{X}_n, \quad \hat{v}^{\text{MV}} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Par le théorème de Cochran,

$$\frac{n\hat{v}^{\text{MV}}}{v} \sim \chi_{n-1}^2.$$

Ainsi,

$$\mathbb{E}[\hat{v}^{\text{MV}}] = \frac{n-1}{n}v, \quad \text{Biais}(\hat{v}^{\text{MV}}) = -\frac{v}{n},$$

et

$$\text{Var}(\hat{v}^{\text{MV}}) = \frac{2(n-1)}{n^2}v^2.$$

La borne de Cramér–Rao $2v^2/n$ concerne les estimateurs sans biais de v (avec μ inconnu comme paramètre de nuisance). L’EMV étant biaisé, il n’est pas soumis à cette borne; sa variance peut donc lui être inférieure sans contradiction.

3. L’estimateur usuel sans biais est

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \hat{v}^{\text{MV}},$$

et

$$\frac{(n-1)S_n^2}{v} \sim \chi_{n-1}^2, \quad \text{Var}(S_n^2) = \frac{2v^2}{n-1}.$$

Son risque quadratique est donc

$$R(S_n^2, v) = \frac{2v^2}{n-1}.$$

Pour l’EMV,

$$R(\hat{v}^{\text{MV}}, v) = \text{Var}(\hat{v}^{\text{MV}}) + \text{Biais}(\hat{v}^{\text{MV}})^2 = \frac{2n-1}{n^2}v^2.$$

Pour tout $n \geq 2$,

$$\frac{2v^2}{n-1} > \frac{2n-1}{n^2}v^2.$$

L’estimateur sans biais a donc un risque quadratique plus grand que l’EMV. Enfin,

$$\text{Var}(S_n^2) = \frac{2v^2}{n-1} > \frac{2v^2}{n},$$

donc S_n^2 n’atteint pas la borne de Cramér–Rao à taille finie, même s’il est asymptotiquement efficace.

Exercice 3 – Loi de Poisson

1. Comme $\mathbb{E}[X_i] = \text{Var}(X_i) = \lambda$,

$$\mathbb{E}[\bar{X}_n] = \lambda, \quad \sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda).$$

2. L'identité

$$\sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n (X_i - \lambda)^2 - n(\bar{X}_n - \lambda)^2$$

donne

$$\mathbb{E}[S_n^2] = \frac{1}{n-1} \left(n\lambda - n\frac{\lambda}{n} \right) = \lambda.$$

Pour la loi limite, écrivons

$$\begin{aligned} \sqrt{n}(S_n^2 - \lambda) &= \frac{n}{n-1} \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \lambda)^2 - \lambda \right\} \\ &\quad - \frac{n}{n-1} \sqrt{n} (\bar{X}_n - \lambda)^2 + \frac{\sqrt{n}}{n-1} \lambda. \end{aligned}$$

Les deux derniers termes convergent en probabilité vers zéro. Or

$$\text{Var}((X_1 - \lambda)^2) = \mathbb{E}[(X_1 - \lambda)^4] - \lambda^2 = \lambda + 2\lambda^2.$$

Le théorème central limite et Slutsky donnent donc

$$\sqrt{n}(S_n^2 - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda + 2\lambda^2).$$

3. Les deux estimateurs sont sans biais et consistants, mais $\bar{X}_n$ a une variance asymptotique λ , strictement inférieure à $\lambda + 2\lambda^2$. On privilégie donc $\bar{X}_n$.

4. Comme $\bar{X}_n \xrightarrow{\mathbb{P}} \lambda$ et $S_n^2 \xrightarrow{\mathbb{P}} \lambda$, Slutsky donne

$$\sqrt{n} \frac{\bar{X}_n - \lambda}{\sqrt{\bar{X}_n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \quad \sqrt{n} \frac{\bar{X}_n - \lambda}{S_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Pour stabiliser la variance, la delta-méthode suggère $[g'(\lambda)]^2 \lambda = 1$. On peut prendre $g(x) = 2\sqrt{x}$, ce qui donne

$$\sqrt{n}(2\sqrt{\bar{X}_n} - 2\sqrt{\lambda}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

5. Notons $z = z_{1-\alpha/2}$. Les trois pivots précédents conduisent respectivement aux intervalles asymptotiques

$$\begin{aligned} I_{1,n} &= \left[\bar{X}_n - z\sqrt{\frac{\bar{X}_n}{n}}, \bar{X}_n + z\sqrt{\frac{\bar{X}_n}{n}} \right], \\ I_{2,n} &= \left[\bar{X}_n - z\frac{S_n}{\sqrt{n}}, \bar{X}_n + z\frac{S_n}{\sqrt{n}} \right], \end{aligned}$$

et

$$I_{3,n} = \left[\left(\max \left\{ 0, \sqrt{\bar{X}_n} - \frac{z}{2\sqrt{n}} \right\} \right)^2, \left(\sqrt{\bar{X}_n} + \frac{z}{2\sqrt{n}} \right)^2 \right].$$

Ils sont asymptotiquement équivalents. Lorsque la borne inférieure de $I_{3,n}$ n'est pas tronquée, $I_{1,n}$ et $I_{3,n}$ ont même longueur, et

$$\frac{|I_{1,n}|}{|I_{2,n}|} = \frac{\sqrt{\bar{X}_n}}{S_n} \xrightarrow{\mathbb{P}} 1.$$

Il n'y a donc pas de meilleur intervalle au premier ordre. $I_{1,n}$ ne dépend que de la statistique suffisante $\sum_i X_i$, tandis que $I_{3,n}$ respecte naturellement la contrainte $\lambda \geq 0$.

Un test asymptotique de niveau α de $H_0 : \lambda = \lambda_0$ contre $H_1 : \lambda \neq \lambda_0$ rejette H_0 lorsque λ_0 n'appartient pas à l'un de ces intervalles; par exemple, avec le premier pivot, lorsque

$$\left| \sqrt{n} \frac{\bar{X}_n - \lambda_0}{\sqrt{\bar{X}_n}} \right| > z.$$

Exercice 4 – Loi uniforme

Soit $X_1, \dots, X_n$ un échantillon i.i.d. uniforme sur $[0, \theta]$, avec $\theta > 0$, et $M_n = X_{(n)}$.

1. Le support $[0, \theta]$ dépend du paramètre. Pour tout $x > 0$, $\theta \mapsto \theta^{-1} \mathbf{1}_{[0, \theta]}(x)$ est discontinue en $\theta = x$. Le modèle n'est donc pas régulier et les résultats généraux sur l'efficacité asymptotique de l'EMV ne s'appliquent pas.

2. Puisque

$$\mathbb{E}[X_1] = \frac{\theta}{2}, \quad \text{Var}(X_1) = \frac{\theta^2}{12},$$

l'estimateur des moments est $\hat{\theta}_n^{\text{MM}} = 2\bar{X}_n$. Il est fortement consistant et

$$\sqrt{n}(\hat{\theta}_n^{\text{MM}} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\theta^2}{3}\right).$$

Par Slutsky, un intervalle asymptotique de niveau $1 - \alpha$ est

$$\left[\hat{\theta}_n^{\text{MM}} - z_{1-\alpha/2} \frac{\hat{\theta}_n^{\text{MM}}}{\sqrt{3n}}, \hat{\theta}_n^{\text{MM}} + z_{1-\alpha/2} \frac{\hat{\theta}_n^{\text{MM}}}{\sqrt{3n}} \right].$$

3. Sous $H_0 : \theta = \theta_0$,

$$Z_n = \sqrt{3n} \frac{\hat{\theta}_n^{\text{MM}} - \theta_0}{\theta_0} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

On rejette donc asymptotiquement H_0 au niveau α lorsque $|Z_n| > z_{1-\alpha/2}$.

4. La vraisemblance est

$$L_n(\theta) = \theta^{-n} \mathbf{1}_{[M_n, +\infty)}(\theta).$$

Elle est décroissante sur $[M_n, +\infty)$, donc

$$\hat{\theta}_n^{\text{MV}} = M_n.$$

Comme M_n/θ suit une loi bêta $\text{Beta}(n, 1)$,

$$\mathbb{E}[M_n] = \frac{n}{n+1}\theta, \quad \mathbb{E}[M_n^2] = \frac{n}{n+2}\theta^2.$$

Son risque quadratique vaut

$$R(M_n, \theta) = \mathbb{E}[(M_n - \theta)^2] = \frac{2\theta^2}{(n+1)(n+2)}.$$

5. Pour $x \geq 0$ et n assez grand pour que $x \leq n\theta$,

$$\mathbb{P}_\theta(n(\theta - M_n) \geq x) = \left(1 - \frac{x}{n\theta}\right)^n \longrightarrow e^{-x/\theta}.$$

Par conséquent,

$$n(\theta - \hat{\theta}_n^{\text{MV}}) \xrightarrow{\mathcal{L}} \text{Exp}(1/\theta),$$

où $\text{Exp}(1/\theta)$ désigne la loi exponentielle de taux $1/\theta$.

6. Puisque $M_n \leq \theta$ presque sûrement,

$$\begin{aligned}\mathbb{P}_\theta(\theta \leq c_{\alpha,n} M_n) &= 1 - \mathbb{P}_\theta\left(M_n < \frac{\theta}{c_{\alpha,n}}\right) \\ &= 1 - c_{\alpha,n}^{-n}.\end{aligned}$$

Le choix $c_{\alpha,n} = \alpha^{-1/n}$ donne donc l'intervalle exact

$$[M_n, \alpha^{-1/n} M_n]$$

de niveau $1 - \alpha$.

7. La médiane de X_1 vaut $\theta/2$. Si $\tilde{X}_n$ est une médiane empirique, on pose $\hat{\theta}_n^{\text{med}} = 2\tilde{X}_n$. Comme la densité vaut $1/\theta$ en la médiane, le théorème central limite pour les quantiles donne

$$\sqrt{n}(\tilde{X}_n - \theta/2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta^2/4),$$

donc

$$\sqrt{n}(\hat{\theta}_n^{\text{med}} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \theta^2).$$

Un intervalle asymptotique associé est

$$\left[\hat{\theta}_n^{\text{med}} - z_{1-\alpha/2} \frac{\hat{\theta}_n^{\text{med}}}{\sqrt{n}}, \hat{\theta}_n^{\text{med}} + z_{1-\alpha/2} \frac{\hat{\theta}_n^{\text{med}}}{\sqrt{n}} \right].$$

8. On choisira l'intervalle exact fondé sur M_n . Il est non asymptotique et sa longueur est d'ordre n^{-1} , car

$$\alpha^{-1/n} - 1 = e^{-\log(\alpha)/n} - 1 \sim -\frac{\log \alpha}{n},$$

alors que les intervalles fondés sur la moyenne ou la médiane ont une longueur d'ordre $n^{-1/2}$.

9. La notation λ dans l'énoncé est une coquille : le paramètre du modèle est θ . L'intervalle exact précédent fournit un test exact de niveau α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$. On rejette lorsque

$$\theta_0 \notin [M_n, \alpha^{-1/n} M_n],$$

c'est-à-dire lorsque

$$M_n > \theta_0 \quad \text{ou} \quad M_n < \alpha^{1/n} \theta_0.$$

Sous H_0 , le premier événement est impossible et le second a probabilité α ; le test est donc exactement de niveau α .

Exercice bonus – E-values et tests séquentiels

Posons

$$Z_i = \frac{(3/4)^{X_i} (1/4)^{1-X_i}}{(1/2)^{X_i} (1/2)^{1-X_i}}, \quad E_n = \prod_{i=1}^n Z_i, \quad E_0 = 1.$$

1. Chaque facteur est positif, donc $E_n \geq 0$. Sous H_0 ,

$$Z_i = \begin{cases} 3/2, & X_i = 1, \\ 1/2, & X_i = 0, \end{cases} \quad \mathbb{E}_0[Z_i] = \frac{1}{2} \frac{3}{2} + \frac{1}{2} \frac{1}{2} = 1.$$

Par indépendance, $\mathbb{E}_0[E_n] = \prod_i \mathbb{E}_0[Z_i] = 1$. Ainsi, E_n est une e-value pour tout horizon fixé n .

2. Si $S_n = \sum_{i=1}^n X_i$, il y a S_n succès et $n - S_n$ échecs, d'où

$$E_n = \left(\frac{3}{2}\right)^{S_n} \left(\frac{1}{2}\right)^{n-S_n}.$$

Il s'agit exactement du rapport entre la vraisemblance des observations sous $p = 3/4$ et leur vraisemblance sous $p = 1/2$.

3. Sous H_0 , l'inégalité de Markov donne, pour tout n fixé,

$$\mathbb{P}_0(E_n \geq 1/\alpha) \leq \alpha \mathbb{E}_0[E_n] = \alpha.$$

La règle proposée définit donc un test de niveau au plus α .

4. Si $n = 10$ et $S_{10} = 8$,

$$E_{10} = \left(\frac{3}{2}\right)^8 \left(\frac{1}{2}\right)^2 = \frac{6561}{1024} \simeq 6.41.$$

Au niveau 5%, le seuil est $1/\alpha = 20$. On ne rejette donc pas H_0 .

5. Comme $E_{n+1} = E_n Z_{n+1}$, E_n est $\mathcal{F}_n$ -mesurable et Z_{n+1} est indépendant de $\mathcal{F}_n$ sous H_0 ,

$$\mathbb{E}_0[E_{n+1} | \mathcal{F}_n] = E_n \mathbb{E}_0[Z_{n+1} | \mathcal{F}_n] = E_n.$$

Ainsi, $(E_n)_{n \geq 0}$ est une martingale positive sous H_0 .

6. L'inégalité de Ville donne directement

$$\mathbb{P}_0\left(\exists n \geq 1 : E_n \geq \frac{1}{\alpha}\right) = \mathbb{P}_0\left(\sup_{n \geq 0} E_n \geq \frac{1}{\alpha}\right) \leq \alpha.$$

On peut donc examiner les données après chaque observation et s'arrêter au temps

$$\tau = \inf\{n \geq 1 : E_n \geq 1/\alpha\}$$

sans inflation du risque de première espèce : $\mathbb{P}_0(\tau < \infty) \leq \alpha$. C'est la différence essentielle avec l'application répétée d'un test valide seulement à un horizon fixé.

7. Sous une vraie loi Bernoulli(p), la loi forte des grands nombres donne

$$\frac{1}{n} \log E_n \xrightarrow{\text{p.s.}} p \log(3/2) + (1-p) \log(1/2) = p \log 3 - \log 2.$$

La dérive est positive seulement si

$$p > \frac{\log 2}{\log 3} \simeq 0.631.$$

Ainsi, lorsque $p = 0.6$, $E_n \rightarrow 0$ presque sûrement : bien que la procédure reste parfaitement valide sous H_0 , elle est mal adaptée à cette alternative et n'a pas une puissance tendant vers un.

Pour ne pas privilégier une seule alternative, on peut choisir une loi de probabilité π sur $(1/2, 1]$ et mélanger les rapports de vraisemblance :

$$\tilde{E}_n = \int_{1/2}^1 \frac{\mathbb{P}_p(X_1, \dots, X_n)}{\mathbb{P}_{1/2}(X_1, \dots, X_n)} \pi(dp).$$

Par Tonelli,

$$\mathbb{E}_0[\tilde{E}_n] = \int_{1/2}^1 1 \pi(dp) = 1.$$

Cette statistique reste donc une e-value et, comme mélange de martingales positives, fournit également un test séquentiel valide via l'inégalité de Ville.