Stochastiek 2

Inleiding in de Mathematische Statistiek

Cramér-Rao

(Asymptotische) optimaliteit van schatters

We zullen zien dat de MLS onder voorwaarden voor $n \to \infty$ zuiver is (consistent) en dat we de asymptotische variantie kunnen bepalen.

Vraag: hoe klein kan de variantie van een zuiver schatter worden?

Fisher-informatie

Setting: waarneming X, positieve dichtheid p_θ, $\theta \in \Theta$, met $\Theta \subset \mathbb{R}$ een open interval. Aanname: $\theta \mapsto p_\theta$ is positief en “glad”.

Definieer:

$$\begin{aligned}
\ell_\theta &= \log p_\theta, \\
\dot{\ell}_\theta &= \frac{\partial}{\partial \theta} \log p_\theta = \frac{\dot{p}_\theta}{p_\theta}, \\
I_\theta &= Var_\theta \, \dot{\ell}_\theta(X).
\end{aligned}$$

Definitie.
I_θ heet de Fisher-informatie (voor θ in X).

Cramér-Rao

Stelling. [Cramér-Rao]

Zij T een zuivere schatter voor $g(\theta)$, voor $g : \Theta \to \mathbb{R}$. Onder "regulariteitsvoorwaarden" geldt dat voor alle $\theta \in \Theta$,

$$Var_\theta T \geq \frac{(g'(\theta))^2}{I_\theta}.$$

Definitie.

Het getal

$$\frac{(g'(\theta))^2}{I_\theta}$$

heet de Cramér-Rao ondergrens voor het schatten van $g(\theta)$.

Bewijs Cramér-Rao - 1

Aanname: X is continu verdeeld.

$T(X)$ is een zuivere schatter voor $g(\theta)$, i.e. voor alle θ

$$\int T(x)p_\theta(x)\,dx = g(\theta).$$

Differentiëren, verwisselen en $\dot{\ell}_\theta = \dot{p}_\theta / p_\theta$ geven dan

$$g'(\theta) = \int T(x)\dot{p}_\theta(x)\,dx = \int T(x)\dot{\ell}_\theta(x)p_\theta(x)\,dx = E_\theta T\dot{\ell}_\theta(X).$$

Omdat

$$E_\theta\dot{\ell}_\theta(X) = \int \frac{\dot{p}_\theta(x)}{p_\theta(x)}p_\theta(x)\,dx = \frac{\partial}{\partial\theta}\underbrace{\int p_\theta(x)\,dx}_{=1} = 0,$$

volgt dat

$$g'(\theta) = Cov_\theta(T, \dot{\ell}_\theta(X)).$$

Bewijs Cramér-Rao - 2

For s.v.'n U, V met eindige variantie geldt de Cauchy-Schwarz ongelijkheid:

$$Cov(U, V) \leq \sqrt{Var\,U}\sqrt{Var\,V}.$$

Toepassen met $U = T$ en $V = \dot{\ell}_\theta(X)$ geeft

$$(g'(\theta))^2 = |Cov_\theta(T, \dot{\ell}_\theta(X))|^2 \leq (Var_\theta T)(Var_\theta\, \dot{\ell}_\theta(X)) = (Var_\theta T) I_\theta.$$

$\square$

Fisher informatie en onafhankelijkheid

Vaak: waarneming is $X = (X_1, \ldots, X_n)$, met $X_1, \ldots, X_n$ o.o. en X_i heeft dichtheid p_θ.

Dan is de (simultane) dichtheid van de hele waarneming X:

$$p_\theta(x_1, \ldots, x_n) = p_\theta(x_1) \cdots p_\theta(x_n).$$

Volgt dat voor de Fisher-informatie in de hele waarneming geldt dat

$$I_\theta = Var_\theta \frac{\partial}{\partial\theta} \log \prod_{i=1}^{n} p_\theta(X_i) = \sum_{i=1}^{n} Var_\theta \frac{\partial}{\partial\theta} \log p_\theta(X_i) = n i_\theta,$$

met

$$i_\theta = Var_\theta \frac{\partial}{\partial\theta} \log p_\theta(X_1)$$

de Fisher-informatie in één waarneming X_1.

$\rightarrow$ vb. 6.28

Asymptotische optimaliteit van de MLS

Asymptotische normaliteit van de MLS

Setting: $X_1, \dots, X_n$ o.o. en X_i heeft positieve dichtheid p_θ die glad van $\theta \in \Theta \subset \mathbb{R}$ afhangt. Zij $\hat{\theta}_n$ de MLS en

$$i_\theta = Var_\theta \frac{\partial}{\partial\theta} \log p_\theta(X_1)$$

de Fisher-informatie in één waarneming X_1.

Stelling. (Zie Stelling 5.9 in het boek.)

Onder "regulariteitsvoorwaarden" geldt dat voor iedere $\theta_0 \in \Theta$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathrm{d}} N(0, i_{\theta_0}^{-1}),$$

onder P_{θ_0} als $n \to \infty$.

Schets van het bewijs - 1

Onder voorwaarden geldt voor de MLS dat $S_n(\hat{\theta}_n) = 0$, met

$$S_n(\theta) = \sum_{i=1}^{n} \dot{\ell}_\theta(X_i)$$

de score functie. Een Taylor expansie van S_n rond θ_0 geeft, met $\dot{S}_n(\theta) = \frac{\partial}{\partial\theta} S_n(\theta)$,

$$0 = S_n(\hat{\theta}_n) = S_n(\theta_0) + \dot{S}_n(\bar{\theta}_n)(\hat{\theta}_n - \theta_0),$$

met $\bar{\theta}_n$ een punt tussen θ_0 en $\hat{\theta}_n$. Er volgt dat

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \Big(-\frac{1}{n}\dot{S}_n(\bar{\theta}_n)\Big)^{-1} \frac{1}{\sqrt{n}} S_n(\theta_0).$$

Schets van het bewijs - 2

Er geldt (zie het bewijs van Cramér-Rao) $E_{\theta_0}\dot{\ell}_{\theta_0}(X_1) = 0$ en $Var_{\theta_0}\dot{\ell}_{\theta_0}(X_1) = i_{\theta_0}$. Dus vanwege de CLS

$$\frac{1}{\sqrt{n}} S_n(\theta_0) \xrightarrow{\mathrm{d}} N(0, i_{\theta_0})$$

onder P_{θ_0}.

Vanwege de WGA

$$-\frac{1}{n}\dot{S}_n(\bar{\theta}_n) \approx -\frac{1}{n}\dot{S}_n(\theta_0) \xrightarrow{P} -E_{\theta_0}\ddot{\ell}_{\theta_0}(X_1)$$

onder P_{θ_0}. Het differentiëren van de relatie $\int \dot{\ell}_\theta(x) p_\theta(x)\, dx = 0$ naar θ laat zien dat $-E_{\theta_0}\ddot{\ell}_{\theta_0}(X_1) = i_{\theta_0}$ (ga na!)

Combineer nu bovenstaande. $\square$

Asymptotische optimaliteit van de MLS

Als θ de ware parameter is, hebben we voor grote n dus de benadering

$$\hat{\theta}_n \overset{d}{\approx} N\Big(\theta, \frac{1}{n i_\theta}\Big).$$

In het bijzonder geldt dan voor iedere θ dat

$$E_\theta\, \hat{\theta}_n \approx \theta, \qquad Var_\theta\, \hat{\theta}_n \approx \frac{1}{n i_\theta} = \frac{1}{I_\theta},$$

met I_θ de Fisher informatie in de hele waarneming $X_1, \dots, X_n$.

Conclusie: Maximum likelihood-schatters zijn asymptotisch zuiver met minimale variantie!

N.B.

- Bayes-schatters hebben deze eigenschap ook! (Stelling 5.26)
- momentenschatters i.h.a. niet.