Stochastiek 2

Inleiding in de Mathematische Statistiek

UMVZ-schatters

UMVZ schatters

Definitie.
Een schatter T heet UMVZ voor θ als T zuiver is en voor alle zuivere schatters S en $\theta \in \Theta$: $Var_\theta T \leq Var_\theta S$.

Bij het zoeken naar UMVZ schatter kunnen we ons beperken tot schatters op basis van een voldoende statistiek:

Stelling. [Rao-Blackwell]
Zij T een schatter en V een voldoende statistiek. Dan bestaat er een schatter $T^* = T^*(V)$ zodanig dat voor alle $\theta \in \Theta$,

$$E_\theta T^* = E_\theta T, \qquad Var_\theta T^* \leq Var_\theta T$$

Volledigheid

Vraag: wanneer is een zuivere schatter op basis van een voldoende statistiek UMVZ?

Definitie.
Een statistiek V heet volledig als voor alle functies f:

$$\forall\theta : E_\theta f(V) = 0 \qquad \Longrightarrow \qquad \forall\theta : P_\theta(f(V) = 0) = 1.$$

Stelling.
Als V voldoende en volledig is en $T = T(V)$ is een zuiver schatter op basis van V, dan is T UMVZ.

Exponentiële families

Definitie. Een familie kansdichtheden p_θ, $\theta \in \Theta$, heet een k-dimensionale exponentiële familie als er functies c, h, Q_j en V_j bestaan z.d.d. voor alle x,

$$p_\theta(x) = c(\theta)h(x)e^{\sum_{j=1}^k Q_j(\theta)V_j(x)}.$$

N.B. Als $X \sim p_\theta$ met p_θ uit een k-dimensionale exponentiële familie, dan is $(V_1, \ldots, V_k)$ een voldoende statistiek.

Exponentiële families

Stelling.

$X \sim p_\theta$ met p_θ uit een k-dimensionale exponentiële familie.
Als

$$\Big\{(Q_1(\theta), \ldots, Q_k(\theta)) : \theta \in \Theta\Big\}$$

een inwendig punt bevat, dan is $(V_1, \ldots, V_k)$ voldoende en volledig.

→ vb. 6.22 t/m 6.25

Cramér-Rao

Fisher-informatie

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

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

Voor de MLS geldt onder "regulariteitsaannames" dat voor iedere $\theta \in \Theta$,

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

onder P_θ als $n \to \infty$, met i_θ de Fisher-informatie in één waarneming.

Voor grote n hebben we dus de benadering

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

Asymptotische optimaliteit

In het bijzonder geldt dan voor iedere θ dat

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

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

Conclusie: Maximum likelihood-schatters zijn asymptotisch UMVZ.

N.B.

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