Stochastiek 2

Inleiding in de Mathematische Statistiek

Betrouwbaarheidsgebieden - herhaling

Setting: waarneming $X \sim P_\theta$, $\theta \in \Theta$.

Definitie.
Een stochastische verzameling $G_X \subset \Theta$ die van X afhangt heet een betrouwbaarheidsgebied voor θ met onbetrouwbaarheid $\alpha \in (0, 1)$ als

$$\forall\ \theta \in \Theta : P_\theta(\theta \in G_X) \geq 1 - \alpha.$$

Als $\Theta \subset \mathbb{R}$ en G_X is een interval, dan heet G_X een betrouwbaarheidsinterval.

Constructies:

- Pivots (normale verwachting, binomiale succeskans, ...)
- MLS als bijna-pivot (Wald interval)

betrouwbaarheidsgebieden en toetsen

Stelling

Setting: waarneming $X \sim P_\theta$, $\theta \in \Theta$.

Willen: een betrouwbaarheidsgebied voor $g(\theta)$, voor een gegeven functie $g : \Theta \to g(\Theta)$ (meestal $g : \Theta \to \mathbb{R}$).

Toetsen: gegeven $\alpha \in (0,1)$ en voor iedere $\tau \in g(\Theta)$ een toets van niveau α voor $H_0 : g(\theta) = \tau$ tegen $H_1 : g(\theta) \neq \tau$, met kritiek gebied K_τ.

Stelling.

In deze setting is

$$\begin{aligned} G_X &= \{\tau \in g(\Theta) : X \notin K_\tau\} \\ &= \{\tau : \text{ ``}H_0 : g(\theta) = \tau\text{'' wordt niet verworpen bij waarneming } X\} \end{aligned}$$

een betrouwbaarheidsgebied voor $g(\theta)$ met onbetrouwbaarheid α.

$\to$ vb. 5.20, 5.22

likelihood-ratiogebieden

Likelihood-ratiostatistiek - asymptotische verdeling

Setting: o.o. waarnemingen $X_1, \ldots, X_n \sim p_\theta$, $\theta \in \Theta \subset \mathbb{R}^k$.
Hypothesen: $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$.

Aanname: stel dat voor $\theta_0 \in \Theta_0$:

$$\sqrt{n}(\Theta - \theta_0) \to \text{“}k\text{-dim. lineaire ruimte”},$$

$$\sqrt{n}(\Theta_0 - \theta_0) \to \text{“}k_0\text{-dim. lineaire ruimte”}$$

als $n \to \infty$.

“Stelling”

In deze setting geldt onder regulariteitsvoorwaarden dat onder P_{θ_0},

$$2 \log \frac{\prod_{i=1}^n p_{\hat{\theta}_n}(X_i)}{\sup_{\theta_0 \in \Theta_0} \prod_{i=1}^n p_{\theta_0}(X_i)} \xrightarrow{\mathrm{d}} \chi^2_{k-k_0}$$

als $n \to \infty$, met $\hat{\theta}_n$ de MLS.

Likelihood-ratiogebied - hele parameter

Stel nu, we willen een b.g. voor θ zelf ($g(\theta) = \theta$).

Onder regulariteitsvoorwaarden: LR-toets van benaderd niveau α voor $H_0 : \theta = \tau$ verwerpt H_0 als

$$2 \log \frac{\prod_{i=1}^n p_{\hat{\theta}_n}(X_i)}{\prod_{i=1}^n p_\tau(X_i)} \geq \chi^2_{k,1-\alpha}.$$

Benaderd b.g. voor θ met onbetrouwbaarheid α dus:

$$\Big\{\theta \in \Theta : \log \prod_{i=1}^n p_\theta(X_i) - \log \prod_{i=1}^n p_{\hat{\theta}_n}(X_i) \geq -\frac{1}{2}\chi^2_{k,1-\alpha}\Big\}.$$

Likelihood-ratiogebied - opmerkingen

- LR-gebied: gebied waar de log-likelihood "groot" is (relatief).
- De MLS zit altijd in het LR-gebied
- Als de log-likelihood meerdere lokale maxima heeft, kan het LR-gebied onsamenhangend zijn (uit meerdere gescheiden delen bestaan).

$\rightarrow$ vb. 5.23

Likelihood-ratiogebied - één component v.d. parameter

Stel nu, we willen een b.g. voor θ_1 ($g(\theta) = g(\theta_1, \ldots, \theta_k) = \theta_1$).

Onder regulariteitsvoorwaarden: LR-toets van benaderd niveau α voor $H_0 : \theta_1 = \tau$ verwerpt H_0 als

$$2 \log \frac{\prod_{i=1}^n p_{\hat{\theta}_n}(X_i)}{\sup_{\theta:\theta_1=\tau} \prod_{i=1}^n p_\theta(X_i)} \geq \chi^2_{1,1-\alpha}.$$

Benaderd b.g. voor θ_1 met onbetrouwbaarheid α dus:

$$\left\{ \tau \in \mathbb{R} : \log \sup_{\theta:\theta_1=\tau} \prod_{i=1}^n p_\theta(X_i) - \log \prod_{i=1}^n p_{\hat{\theta}_n}(X_i) \geq -\frac{1}{2}\chi^2_{1,1-\alpha} \right\}.$$

In rood: de profile likelihood.