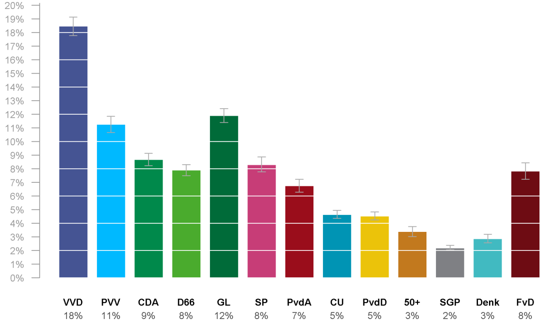
Stochastiek 2

Inleiding in de Mathematische Statistiek

Betrouwbaarheidsgebieden

Idee

- Een schatter T voor een parameter θ geeft één punt in de parameterruimte Θ.
- I.h.a. zal $T \neq \theta$ onder P_θ, we maken altijd een fout.
- Willen graag de afwijking/onzekerheid kwantificeren.



Peilingwijzer. Tom Louwerse, Universiteit Leiden. Bijgewerkt: 19-09-2018

Definitie

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.

Interpretatie: herhaling van experimenten! (p. 178 (172))

$\rightarrow$ vb. 5.2

Pivots

Definitie

Definitie.
Een pivot is een functie $T(X, \theta)$ van de waarneming X en de parameter θ, z.d.d. de verdeling van $T(X, \theta)$ onder P_θ niet van θ afhangt. M.a.w.

$$\forall B: \ P_\theta(T(X, \theta) \in B) \quad \text{is onafh. van } \theta.$$

Betrouwbaarheidsgebied op basis van pivot

- Fixeer een onbetrouwbaarheidsniveau $\alpha \in (0, 1)$.
- Bepaal B z.d.d.

$$P_\theta(T(X, \theta) \in B) \geq 1 - \alpha.$$

  N.B.: B hangt dan dus niet van θ af!

- Dan is

$$\{\theta \in \Theta : T(X, \theta) \in B\}$$

  een betrouwbaarheidsgebied met onbetrouwbaarheid α.

$\rightarrow$ vb. 5.4, 5.5, 5.6

Maximum likelihood-schatters als bijna-pivots

MLS en Fisher informatie

Setting: waarnemingen $X_1, \ldots, X_n$ o.o., X_i heeft dichtheid p_θ, $\theta \in \Theta \subset \mathbb{R}$.

Aanname: $\theta \mapsto p_\theta$ is positief en glad.

Definieer: $\ell_\theta(x) = \log p_\theta(x)$, $\dot{\ell}_\theta(x) = \frac{\partial}{\partial\theta}\ell_\theta(x)$.

Zij $\hat{\theta}_n$ de MLS: punt waar $\theta \mapsto \sum_{i=1}^n \ell_\theta(X_i)$ maximaal is.

Fisher-informatie in één waarneming X_1: $i_\theta = Var_\theta \dot{\ell}_\theta(X_1)$.

Asymptotiek van de MLS geeft benaderd b.i.

"Stelling".

Onder regulariteitsaannamen geldt dat onder P_θ:

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

In het bijzonder: $T(X_1, \ldots, X_n; \theta) = \sqrt{n i_\theta}(\hat{\theta}_n - \theta)$ is een bijna pivot voor grote n.

Volgt dat

$$\{\theta : -\xi_{1-\alpha/2} \leq \sqrt{n i_\theta}(\hat{\theta}_n - \theta) \leq \xi_{1-\alpha/2}\}$$

een benaderd betrouwbaarheidsgebied is voor θ met onbetrouwbaarheid α.

Wald-interval, schatters voor i_θ

Handig om i_θ te vervangen door een geschikte schatter $\widehat{i_\theta}$. We krijgen dan het Wald-interval

$$\theta = \hat{\theta}_n \pm \frac{1}{\sqrt{n\widehat{i_\theta}}}\xi_{1-\alpha/2}.$$

Geschikte/veelgebruikte schatters voor i_θ:

- plug-in schatter: $\widehat{i_\theta} = i_{\hat{\theta}_n}$, met $\hat{\theta}_n$ de MLS.
- waargenomen informatie:

$$\widehat{i_\theta} = -\frac{1}{n}\sum_{i=1}^{n} \ddot{\ell}_{\hat{\theta}_n}(X_i),$$

met $\ddot{\ell}_\theta = \frac{\partial}{\partial\theta}\dot{\ell}_\theta$. (Herinner: $i_\theta = -E_\theta\ddot{\ell}_\theta(X_1)$).

$\rightarrow$ vb. 5.11, 5.12, 5.13

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