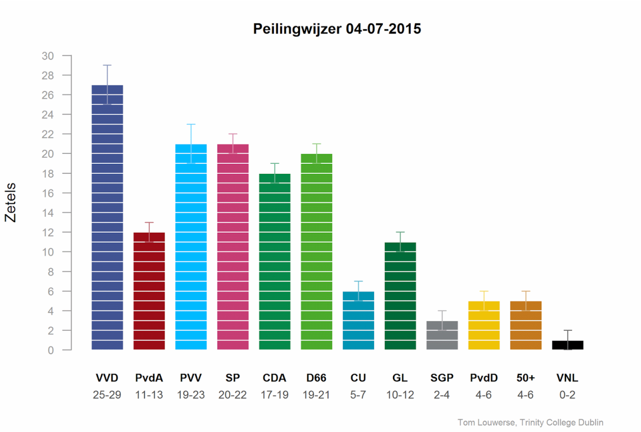
Stochastiek 2

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.



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. 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" (handout).

Onder regulariteitsaannamen geldt dat onder P_θ:

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

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

Volgt dat

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

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: $i_\theta = -E_\theta \ddot{\ell}_\theta(X_1)$ (handout)).

$\rightarrow$ vb. 5.11, 5.12, 5.13