Stochastiek 2

Inleiding in de Mathematische Statistiek

Statistische toetsen

Toetsen - idee

Veel voorkomende vraag: "is het geobserveerde effect significant?"

Voorbeeld:

We wassen 10 sokken in wasmiddel A en 10 in wasmiddel B. We meten hoeveel ze krimpen.

resultaten A: 0.56 0.03 1.30 1.00 0.70 0.32 0.57 0.50 1.90 0.71
resultaten B: 1.40 0.94 1.90 0.90 1.10 2.00 0.51 1.10 0.85 0.56

Is wasmiddel A beter dan wasmiddel B?...

Zien we slechts statistische variatie, of is er echt iets aan de hand?

Toetsen - formalisering - 1

Setting: observatie X in $\mathcal{X}$, model $\{P_\theta : \theta \in \Theta\}$.

Vertaal de twee hypothesen in termen van een disjuncte splitsing: $\Theta = \Theta_0 \cup \Theta_1$, met $\Theta_0 \cap \Theta_1 = \emptyset$. Hypothesen worden dan:

$$\begin{aligned} H_0 &: \theta \in \Theta_0 \quad &&\text{– nulhypothese,} \\ H_1 &: \theta \in \Theta_1 \quad &&\text{– alternatieve hypothese.} \end{aligned}$$

Doel: op basis van de data concluderen:

- H_0 verwerpen (en dus H_1 accepteren), of
- H_0 niet verwerpen (maar niet noodzakelijkerwijs H_0 accepteren, te weinig info).

Toetsen - formalisering - 2

Conventie: H_1 is wat we willen aantonen. Dit is de “sterke” uitspraak waarin we geïnteresseerd zijn. Als we dat niet kunnen, hoeven we niet per se te weten of H_0 waar is. Misschien is er dan voor geen van beide voldoende bewijs.

Mogelijke foute conclusies:

- Fout van de 1e soort: H_0 verwerpen terwijl H_0 wel waar is.
- Fout van de 2e soort: H_0 niet verwerpen terwijl H_0 wel incorrect is.

Fout van de 1e soort: ten onrechte de “sterke” conclusie.
Zeer ongewenst!

We verwerpen H_0 alleen als er sterke aanwijzingen zijn!

$\rightarrow$ vb. 4.1, 4.3

Toetsingsgrootheid, kritiek gebied

Formeel: een (niet-gerandomiseerde) toets is een afbeelding $\varphi : \mathcal{X} \to \{0, 1\}$.
Interpretatie: Als we $X = x$ waarnemen, dan verwerpen we H_0 als $\varphi(x) = 1$ en we verwerpen H_0 niet als $\varphi(x) = 0$.

Definitie.
Het kritieke gebied van de toets is de verzameling $K \subset \mathcal{X}$ zodanig dat we H_0 verwerpen als we een waarneming $x \in K$ hebben.
(M.a.w. $K = \{x \in \mathcal{X} : \varphi(x) = 1\}$).

Vaak baseren we een toets niet op de hele data X, maar op een toetsingsgrootheid $T(X)$.

$\rightarrow$ vb. 4.5, 4.6

Onbetrouwbaarheid, onderscheidend vermogen - 1

Setting: observatie X in $\mathcal{X}$, model $\{P_\theta : \theta \in \Theta\}$.
Hypothesen: $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$.

Beschouw een toets met kritiek gebied $K \subset \mathcal{X}$.

Definitie.
De functie $\pi(\cdot; K) : \Theta \to [0,1]$ gedefinieerd door

$$\pi(\theta; K) = P_\theta(X \in K), \qquad \theta \in \Theta,$$

heet het onderscheidend vermogen (power function) van de toets.

Interpretatie: $\pi(\theta; K)$ is de kans dat je H_0 verwerpt als θ de ware parameter is.

Construëren van een toets = kiezen van een kritiek gebied. Hoe?

Onbetrouwbaarheid, onderscheidend vermogen - 2

Een toets is "goed" als $\pi(\theta; K)$ "klein" is voor $\theta \in \Theta_0$ en "groot" voor $\theta \in \Theta_1$.

Definitie.
De onbetrouwbaarheid (size) van de toets is

$$\sup_{\theta\in\Theta_0} \pi(\theta; K) = \sup_{\theta\in\Theta_0} P_\theta(\text{“verwerp } H_0\text{”}).$$

De toets heeft niveau (level) α als de onbetrouwbaarheid $\leq \alpha$ is.

Conventie: Voor het construëren van een toets kiezen we een (klein) getal α, de onbetrouwbaarheidsdrempel. We beschouwen dan alleen toetsen van niveau α. Onder deze toetsen geven we de voorkeur aan de toetsen met een zo groot mogelijk onderscheidend vermogen op Θ_1.

$\rightarrow$ vb. 4.11, 4.12