10 Grundbegriffe der Testtheorie

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ Wahrscheinlichkeitsraum, $(\mathfrak{X}, \mathcal{B}, \{P_\vartheta : \vartheta \in \Theta\})$ statistischer Raum, $X: \ \Omega \to \mathfrak{X}$ Zufallsvariable, $\Theta = \Theta_0 + \Theta_1$ mit $\Theta_0, \Theta_1 \neq \emptyset$.
$(\Theta_0 \cap \Theta_1 = \emptyset)$

10.1 Definition

Die Aussage $H_0 : \vartheta \in \Theta_0$ heißt (Null-)Hypothese, $H_1 : \vartheta \in \Theta_1$ heißt Alternativhypothese oder Alternative.
$|\Theta_j| = 1 \Rightarrow \Theta_j$ heißt einfach, sonst zusammengesetzt

10.2 Definition

Ein **randomisierter Test** zur Prüfung von H_0 gegen H_1 ist eine messbare Abbildung $\varphi: \ \mathfrak{X} \to [0,1]$ mit der Interpretation

$$\varphi(x) = P(H_0 \text{ ablehnen}| \ X = x)$$

Gilt $\varphi(\mathfrak{X}) = \{0, 1\}$, so heißt φ **nicht randomisiert**. Mit $\mathcal{K} := \{x \in \mathfrak{X} : \varphi(x) = 1\}$ gilt dann $\varphi = \mathbf{1}_\mathcal{K}$ und die Testvorschrift lautet:

$$\begin{aligned} x \in \mathcal{K} \quad &\Rightarrow \quad H_0 \text{ ablehnen} \\ x \in \mathfrak{X} \backslash \mathcal{K} \quad &\Rightarrow \quad H_0 \text{ nicht ablehnen} \end{aligned}$$

$\mathcal{K}$ heißt kritischer Bereich (Ablehnbereich), $\mathfrak{X} \backslash \mathcal{K}$ heißt Annahmebereich.

10.3 Bemerkung

Falls $0 < \varphi(x) < 1$, so muss „externes“ Bernoulli-Experiment durchgeführt werden; man erhält also Realisierung y einer Zufallsvariablen Y mit
$Y \sim \mathrm{Bin}(1, \varphi(x))$.
In praktischen Anwendungen ist „Randomisierung“ unerwünscht.

10.4 Definition

Es sei $T:\ \mathfrak{X} \to \mathbb{R}$ eine messbare Abbildung. Häufig besitzt ein nicht randomisierter Test die Gestalt

$$(*) \quad \begin{array}{lcl} T(x) \geq c & \Rightarrow & H_0 \text{ ablehnen} \\ T(x) < c & \Rightarrow & \text{kein Widerspruch zu } H_0 \end{array}$$

(d.h. $\mathcal{K} = \{x \in \mathfrak{X}:\ T(x) \geq c\} = T^{-1}([c,\infty)))$

Dann heißt T Testgröße (Prüfgröße) und $c \in \mathbb{R}$ heißt kritischer Wert.
$(*)$ liefert Test mit **oberem Ablehnbereich**.

In $(*)$ $\geq$ durch $\leq$ und $<$ durch $>$ ersetzen $\hookrightarrow$ Test mit unterem Ablehnbereich

10.5 Beispiel

$(\mathfrak{X},\mathcal{B}) = (\mathbb{R}^{m+n},\mathcal{B}^{n+m})$, $X = (\underbrace{X_1,\ldots,X_m}_{\overset{uiv}{\sim} F},\underbrace{Y_1,\ldots,Y_n}_{\overset{uiv}{\sim} G})$, $X_1,\ldots,Y_n$ unabhängig, $\vartheta = (F,G)$, $\Theta = \{(F,G):\ F,G \text{ stetig}\}$, $\Theta_0 = \{(F,G) \in \Theta:\ F = G\}$

$$H_0:\ F = G$$

$$H_1:\ F \neq G$$

(nichtparametrisches 2-Stichproben-Problem mit allgemeiner Alternative)

Sei

$$\hat{F}_m(x) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\{X_i \leq x\},\ \hat{G}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{Y_j \leq x\}$$

Mögliche Prüfgröße (mit oberem Anlehnbereich):

$$T(X_1,\ldots,X_m,Y_1,\ldots,Y_n) = \sup_{x\in\mathbb{R}} |\hat{F}_m(x) - \hat{G}_n(x)|$$

(Kolmogorov-Smirnov-Testgröße)

10.6 Definition und Bemerkung

Ein Fehler 1. Art ist das Verwerfen von H_0, obwohl H_0 richtig ist.
Ein Fehler 2. Art ist das Nichtverwerfen von H_0, obwohl H_0 falsch ist.

Entscheidung	H_0 richtig	H_0 falsch
H_0 nicht verwerfen	richtige Entscheidung	Fehler 2. Art
H_0 verwerfen	Fehler 1. Art	richtige Entscheidung

Die Funktion

$$G_\varphi : \begin{array}{l} \Theta \to [0,1] \\ \vartheta \mapsto G_\varphi(\vartheta) := E_\vartheta[\varphi] = \int_{\mathfrak{X}} \varphi(x) P_\vartheta(dx) \end{array}$$

heißt **Gütefunktion** des Tests φ.
($\varphi = \mathbf{1}_{\mathcal{K}} \Rightarrow G_\varphi(\vartheta) = P_\vartheta(\mathcal{K})$, $\varphi = \mathbf{1}\{T(x) \geq c\} \Rightarrow G_\varphi(\vartheta) = P_\vartheta(T \geq c)$)

Ideale Gütefunktion wäre

$$G_\varphi(\vartheta) = \begin{cases} 1, \vartheta \in \Theta_1 \\ 0, \vartheta \in \Theta_0 \end{cases}$$

Sei $\alpha \in (0,1)$. φ heißt Test zum **Niveau** α $:\Leftrightarrow G_\varphi(\vartheta) \leq \alpha \ \forall \vartheta \in \Theta_0$[26]

In Praxis übliche Werte: $\alpha = 0,05;\ 0,01;\ 0,001$
Kleines α dient „Sicherung von H_1“.[27]

Die Zahl $\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta)$ heißt **Umfang** (size) von φ.

10.7 Definition

Sei

$$\Phi_\alpha = \{\varphi : \mathfrak{X} \to [0,1] | \sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) \leq \alpha\}$$

die Menge aller Niveau α-Tests.
$\Phi_\alpha \neq \emptyset$, da $\varphi \equiv \alpha \in \Phi_\alpha$.

Sei $\widetilde{\Phi}_\alpha \subset \Phi_\alpha$
$\varphi_1 \in \widetilde{\Phi}_\alpha$ heißt **gleichmäßig besser** als $\varphi_2 \in \widetilde{\Phi}_\alpha$ $:\Leftrightarrow$

$$G_{\varphi_1}(\vartheta) \geq G_{\varphi_2}(\vartheta) \ \forall \vartheta \in \Theta_1$$

$\varphi^* \in \widetilde{\Phi}_\alpha$ heißt (gleichmäßig) **bester Test** in $\widetilde{\Phi}_\alpha$ $:\Leftrightarrow$

$$G_{\varphi^*}(\vartheta) \geq G_\varphi(\vartheta) \ \forall \vartheta \in \Theta_1 \ \forall \varphi \in \widetilde{\Phi}_\alpha$$

Bezeichnung: UMP-Test („uniformly most powerfully“)

[26] Wahrscheinlichkeit für einen Fehler 1. Art ist $\leq \alpha$
[27] vgl. „Wahl der Nullhypothese“; das Verwerfen von H_0 ist „fast nie“ falsch, also in diesem Fall umgekehrt H_1 auch „fast immer“ richtig (...)