4.1 Grundbegriffe der Testtheorie

Hajo Holzmann

Philipps-Universität Marburg

Bsp.: Tea-tasting Lady

- Englische Lady: Trinkt Tee
- Behauptet zu erkennen: Erst Milch, dann Tee, oder erst Tee, dann Milch
- Versuch: n Mal zwei Tassen, eine erst Milch, dann Tee, die andere erst Tee, dann Milch. Soll richtig zuordnen.
- Modell: $X_1 = x_1, \ldots, X_n = x_n$, $X_i \sim Ber(p)$, $p \in [1/2, 1]$, unabhängig.

Bsp.: Tea-tasting Lady

- Englische Lady: Trinkt Tee
- Behauptet zu erkennen: Erst Milch, dann Tee, oder erst Tee, dann Milch
- Versuch: n Mal zwei Tassen, eine erst Milch, dann Tee, die andere erst Tee, dann Milch. Soll richtig zuordnen.
- Modell: $X_1 = x_1, \ldots, X_n = x_n$, $X_i \sim Ber(p)$, $p \in [1/2, 1]$, unabhängig.

Hypothese H: Die Lady rät nur.

Alternative K: Die Lady kann tendenziell richtig klassifizieren.

Formal:

$$H : p = 1/2 \qquad \text{gegen} \qquad K : p > 1/2.$$

Bsp.: Tea-tasting Lady

- Es ist $S_n = X_1 + \ldots + X_n \sim Bin(n, p)$.
- Angenommen, $S_n = s$ beobachtet. Etwa $n = 20$, $s = 13$.

Bsp.: Tea-tasting Lady

- Es ist $S_n = X_1 + \ldots + X_n \sim Bin(n, p)$.
- Angenommen, $S_n = s$ beobachtet. Etwa $n = 20$, $s = 13$.

Prinzip einer Testentscheidung

- Frage: Wie groß ist Wahrscheinlichkeit für die Beobachtung oder eine noch "extremere" Beobachtung unter Hypothese H.
- Hier: Wie groß ist W-keit von $\{S_n \geq s\}$ unter H, also

$$1 - F(s-1; 1/2) = \frac{1}{2^n} \sum_{k \geq s} \binom{n}{k}.$$

- Falls hinreichend klein, wird H zugunsten von K verworfen, sonst bleibe bei H.

Formal:

- Geben uns Fehlerniveau $\alpha > 0$ vor (meist: $\alpha = 0,05$ oder $0,01$).
- Verwerfen H zugunsten von K, falls $1 - F(s-1; 1/2) \leq \alpha$, sonst behalte H.
  $\rightarrow 1 - F(s-1; 1/2)$ P-Wert.
- Gilt H, so W-keit von max. α für Fehlentscheidung (Fehler 1. Art) zugunsten von K.

Äquivalentes Vorgehen:

- Geben uns Fehlerniveau $\alpha > 0$ vor (meist: $\alpha = 0,05$ oder $0,01$).
- Bestimmen kritischen Wert $k_\alpha \in \{0, 1, \dots, n\}$, also k_α minimal mit

$$P_{1/2}(S_n \geq k_\alpha) = 1 - F(k_\alpha - 1; 1/2) \leq \alpha.$$

- Verwerfen H zugunsten von K, falls $S_n = s \geq k_\alpha$ beobachtet, sonst behalte H.

- Bisher: Kontrollieren W-keit, H fälschlicher Weise zu verwerfen
  $\rightarrow \leq \alpha$ (Signifikanzniveau des Tests)
- Angenommen, es gilt K, Teste zum Signifikanzniveau $\alpha > 0$
  Wie groß ist W-keit, dann H zugunsten von K zu verwerfen.
  $\rightarrow$ Güte (Power, Macht) des Tests

Bsp.: Tea-tasting Lady

- Bisher: Kontrollieren W-keit, H fälschlicher Weise zu verwerfen
  $\rightarrow \leq \alpha$ (Signifikanzniveau des Tests)
- Angenommen, es gilt K, Teste zum Signifikanzniveau $\alpha > 0$
  Wie groß ist W-keit, dann H zugunsten von K zu verwerfen.
  $\rightarrow$ Güte (Power, Macht) des Tests

Für $p \in (1/2, 1]$, also unter K, ist

$$\beta(p) = P_p\big(S_n \geq k_\alpha\big) = 1 - F(k_\alpha - 1; p)$$

W-keit, H zugunsten von K zu verwerfen.

Kann zeigen

$$\begin{aligned}\beta(p) &= 1 - F(k_\alpha - 1; p)\\ &= \frac{n!}{(k_\alpha - 1)!(n - k_\alpha)!}\int_0^p t^{k_\alpha}(1-t)^{n-k_\alpha}\,dt\end{aligned}$$

also stetig, streng monoton wachsend in p.

$\rightarrow$ Power $\beta(p) > \alpha$ für alle $p \in (1/2, 1]$.

Theoretische Rahmen von Testproblem

$\mathcal{P} = (P_\theta)_{\theta \in \Theta}$, $\Theta \subset \mathbb{R}^k$: parametrisches Modell

Zerlege disjunkt

$$\Theta = \Theta_0 \cup \Theta_1 .$$

$H : \theta \in \Theta_0$ Hypothese (oder Nullhypothese).

$K : \theta \in \Theta_1$ Alternative (oder alternative Hypothese):

Gilt card H=1, so heißt H einfach, sonst zusammengesetzt.

Test

- Beobachte $X_1 = x_1, \dots, X_n = x_n$ u.i.v., $X_i \sim P_\theta$ für ein $\theta \in \Theta$.
- Aufgrund Beobachtungen entscheide, ob $\theta \in \Theta_0$ oder $\theta \in \Theta_1$.

Def.: Ein Test ist eine Abbildung $T : \mathbb{R}^n \to \{0, 1\}$.

Interpretation:

- Gilt $T(x_1, \dots x_n) = 0$, so behalte H,
- Gilt $T(x_1, \dots x_n) = 1$, so verwerfe H zugunsten von K.

Konstruktion eines Tests

Prüfgröße oder Teststatistik: $S : \mathbb{R}^n \to \mathbb{R}$,

Verwerfungsbereich: $I \subset \mathbb{R}$,

Dann Test T:

$$T(x_1, \ldots, x_n) = 1_{S(x_1,\ldots,x_n) \in I}.$$

Bsp.: Tea-Tasting Lady

$$S(x_1, \ldots, x_n) = x_1 + \ldots + x_n, \quad I = [k_\alpha, \infty).$$

Konstruktion eines Tests

Prüfgröße oder Teststatistik: $S : \mathbb{R}^n \to \mathbb{R}$,
Verwerfungsbereich: $I \subset \mathbb{R}$,
Dann Test T:

$$T(x_1, \dots, x_n) = 1_{S(x_1,\dots,x_n) \in I}.$$

Bsp.: Tea-Tasting Lady

$$S(x_1, \dots, x_n) = x_1 + \dots + x_n, \quad I = [k_\alpha, \infty).$$

Formen des Verwerfungsbereichs

einseitig: Für ein t: $I = [t, \infty)$ (oberer Verwerfungsbereich),
$I = (\infty, t]$ (unterer Verwerfungsbereich)

zweiseitig: Für $t_1 < t_2$: $I = (\infty, t_1] \cup [t_2, \infty)$.

Table Henze.

Gütefunktion

Sei $T(x_1, \ldots, x_n) = 1_{S(x_1,\ldots,x_n) \in I}$ ein Test.

Def.: Die Gütefunktion von T ist die Funktion

$$\begin{aligned} \beta : & \quad \Theta \to [0,1] \\ \beta(\theta) &= P_\theta\big(T(X_1, \ldots, X_n) = 1\big) \\ &= P_\theta\big(S(X_1, \ldots, X_n) \in I\big) \end{aligned}$$

Gütefunktion

Sei $T(x_1,\dots,x_n) = 1_{S(x_1,\dots,x_n)\in I}$ ein Test.

Def.: Die Gütefunktion von T ist die Funktion

$$\begin{aligned} \beta : & \quad \Theta \to [0,1] \\ \beta(\theta) &= P_\theta\big(T(X_1,\dots,X_n)=1\big) \\ &= P_\theta\big(S(X_1,\dots,X_n)\in I\big) \end{aligned}$$

Bem.:

- Für $\theta \in \Theta_0$ ist $\beta(\theta)$ die Wahrscheinlichkeit, H fälschlicher Weise zu verwerfen
  $\to$ W-keit für Fehler 1. Art.
- Für $\theta \in \Theta_1$ ist $\beta(\theta)$ die Wahrscheinlichkeit, die Alternative K zu erkennen
  $\to$ 1-(W-keit für Fehler 2. Art).

Signifikanzniveau und Güte

- Für $\theta \in \Theta_0$: $\quad \beta(\theta)$ möglichst klein.
- Für $\theta \in \Theta_1$: $\beta(\theta)$ möglichst groß.

  $\beta(\theta)$: Güte, Power, Macht des Tests bei $\theta \Theta_1$.

Vorgehen: Gib Schranke $\alpha \in (0,1)$ vor (typisch: $\alpha = 0,05$ oder $\alpha = 0,01$).

Signifikanzniveau und Güte

- Für $\theta \in \Theta_0$: $\beta(\theta)$ möglichst klein.
- Für $\theta \in \Theta_1$: $\beta(\theta)$ möglichst groß.

  $\beta(\theta)$: Güte, Power, Macht des Tests bei $\theta\Theta_1$.

Vorgehen: Gib Schranke $\alpha \in (0,1)$ vor (typisch: $\alpha = 0,05$ oder $\alpha = 0,01$).

Def.: Ein Test T hat Signifikanzniveau oder Niveau α, falls für seine Gütefunktion $\beta(\theta)$ gilt

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

Gilt $\sup_{\theta \in \Theta_0} \beta(\theta) < \alpha$, so heißt der Test T konservativ zum Niveau α.

Bem.: Wir beschränken also die maximale Wahrscheinlichkeit für einen Fehler 1. Art.

Modell: $X_1 = x_1, \ldots, X_n = x_n$, $X_i \sim Ber(p)$, $p \in (0,1)$, unabhängig. Sei $p_0 \in (0,1)$ fest (etwa $p_0 = 1/2$).

Testprobleme:

- Einseitig

$$H : p \leq p_0 \quad \text{gegen} \quad K : p > p_0,$$

oder

$$H : p \geq p_0 \quad \text{gegen} \quad K : p < p_0,$$

- Zweiseitig:

$$H : p = p_0 \quad \text{gegen} \quad K : p \neq p_0.$$

Binomialtest: Hypothese und Verwerfungsbereich

Zu $H : p \leq p_0$ gegen $K : p > p_0$

$\rightarrow$ einseitiger, oberer Verwerfungsbereich $I = [k_\alpha, \infty)$.

Dabei: k_α minimal mit

$$P_{p_0}(S_n \geq k_\alpha) = 1 - F(k_\alpha - 1; p_0) \leq \alpha.$$

Binomialtest: Hypothese und Verwerfungsbereich

Zu $H : p \leq p_0$ gegen $K : p > p_0$

$\rightarrow$ einseitiger, oberer Verwerfungsbereich $I = [k_\alpha, \infty)$.

Dabei: k_α minimal mit

$$P_{p_0}(S_n \geq k_\alpha) = 1 - F(k_\alpha - 1; p_0) \leq \alpha.$$

Dann für alle $p \in (0, 1/2]$

$$1 - F(k_\alpha - 1; p) = \beta(p) \leq \beta(p_0) \leq \alpha,$$

da $1 - F(k_\alpha - 1; p)$ streng monoton wachsend in p.

Binomialtest: Hypothese und Verwerfungsbereich

Zu $H : p \leq p_0$ gegen $K : p > p_0$

$\rightarrow$ einseitiger, oberer Verwerfungsbereich $I = [k_\alpha, \infty)$.

Dabei: k_α minimal mit

$$P_{p_0}(S_n \geq k_\alpha) = 1 - F(k_\alpha - 1; p_0) \leq \alpha.$$

Dann für alle $p \in (0, 1/2]$

$$1 - F(k_\alpha - 1; p) = \beta(p) \leq \beta(p_0) \leq \alpha,$$

da $1 - F(k_\alpha - 1; p)$ streng monoton wachsend in p.

Analog: Zu $H : p \geq p_0$ gegen $K : p < p_0$

$\rightarrow$ einseitiger, unteren Verwerfungsbereich $I = (-\infty, k_\alpha]$.

Binomialtest: Hypothese und Verwerfungsbereich

Zur zweiseitigen Hypothse

$$H : p = p_0 \quad \text{gegen} \quad K : p \neq p_0.$$

Wähle zweiseitigen Verwerfungsbereich

$$I = (-\infty, k_{\alpha/2,l}] \cup [k_{\alpha/2,u}, \infty),$$

Dabei: $k_{\alpha/2,u}$ minimal mit

$$P_{p_0}(S_n \geq k_{\alpha/2,u}) = 1 - F(k_{\alpha/2,u} - 1; p_0) \leq \alpha/2.$$

$k_{\alpha/2,l}$ maximal mit

$$P_{p_0}(S_n \leq k_{\alpha/2,l}) = 1 - F(k_{\alpha/2,l} - 1; p_0) \leq \alpha/2.$$

P-Wert

Inhaltlich: Bestes Niveau, zu dem ich bei gegebener Beobachtung $S = s$ der Prüfgröße verwerfen kann.

- würde oberen Verwerfungsbereich konstruieren, dann

$$P - Wert = \sup_{\theta \in \Theta_0} P_\theta(S \geq s).$$

- würde unteren Verwerfungsbereich konstruieren, dann

$$P - Wert = \sup_{\theta \in \Theta_0} P_\theta(S \leq s).$$

P-Wert

Inhaltlich: Bestes Niveau, zu dem ich bei gegebener Beobachtung $S = s$ der Prüfgröße verwerfen kann.

- würde oberen Verwerfungsbereich konstruieren, dann

$$P - Wert = \sup_{\theta \in \Theta_0} P_\theta(S \geq s).$$

- würde unteren Verwerfungsbereich konstruieren, dann

$$P - Wert = \sup_{\theta \in \Theta_0} P_\theta(S \leq s).$$

Verwerfe zum Niveau α, falls $p - Wert \leq \alpha$.

Bsp.: einseitiger Binomialtest

$$H : p \leq p_0 \quad \text{gegen} \quad K : p > p_0$$

$$P - Wert = P_{p_0}(S_n \geq s) = 1 - F(s-1; p_0).$$

P-Wert

Bsp.: einseitiger Binomialtest

$$H: p \leq p_0 \quad \text{gegen} \quad K: p > p_0$$

$$P - Wert = P_{p_0}(S_n \geq s) = 1 - F(s-1; p_0).$$

Bei zweiseitigem Test (Verwerfungsbereich)

$\rightarrow$ P-Wert schwerer anzugeben.

grob: (kleinere der einseitigen P-Werte) $\cdot 2$.

Benötigter Stichprobenumfang

Motivation:

- Lady 14 von 20 richtig klassifiziert
- können Hypothese $p \leq 1/2$ zum Niveau $\alpha = 0,05$ nicht verwerfen.
- Lady behauptet aber: Ich kanns mit einer W-keit von mind. $p = 0,7 = 14/20$.
- Power des Tests bei $p = 0,7$: $1 - F(15 - 1; 0,7) = 0,416$

Ziel: Wir wollen die Fähigkeit bei $p = 0,7$ mit W-keit $0,9$ erkennen.

$\rightarrow$: Power (grob) wachsend im Stichprobenumfang n.

Ist Roullette Rad richtig adjustiert?