Praktikum zur Statistik mit R

Till Breuer

Institut für Mathematische Statistik
Universität Münster

7. Oktober 2010

Gliederung

Gliederung

Gliederung

Geburten von Jungen und Mädchen

These:

> *Die Wahrscheinlichkeit für eine Mädchengeburt ist höher als die für eine Jungengeburt.*

Von der Stichprobe auf die Grundgesamtheit schließen

- Wir wollen von einer Stichprobe auf die Grundgesamtheit schließen.
- Dazu ziehen wir zunächst eine Stichprobe $(x_1, \ldots, x_n)$.
- Wir bezeichnen die Entscheidungsmöglichkeiten als *Hypothese* und *Alternative*
- In unserem Falle:
    - Hypothese $\cong$ Geburt eines Mädchens
    - Alternative $\cong$ Geburt eines Jungens

Mögliche Fehler, Optimalität

- Fehler 1. Art: Es liegt die Hypothese vor, wir entscheiden uns aber für die Alternative
- Fehler 2. Art: Es liegt die Alternative vor, wir entscheiden uns aber für die Hypothese

Es gilt, im Hinblick auf die beiden möglichen Fehlentscheidungen eine möglichst “gute” Entscheidung zu treffen.

Allgemeines Vorgehen in der Testtheorie

- Eine Stichprobe $x \in \mathfrak{X}$ ist gegeben
- Wir geben eine Schranke α für den Fehler 1. Art vor
- Wir wählen eine möglichst gute Testfunktion

  "Gut" bedeutet, dass die Testfunktion den "Verlust", bzw. die Wahrscheinlichkeit für die Fehler 1. und 2. Art in irgendeiner Form minimiert.

- Wir treffen mit Hilfe der gewählten Testfunktion abhängig von x eine Entscheidung für die Hypothese oder Alternative

  Können wir die Wahrscheinlichkeiten für beide Fehler zugleich minimieren?

Testfunktion

Wir wählen den Test nach folgender Optimalitätsregel:

Definition (Optimalität)

- Die Wahrscheinlichkeit für den Fehler 1. Art darf maximal $\alpha \in (0, 1)$ betragen
- Die Wahrscheinlichkeit für den Fehler 2. Art soll möglichst gering sein

Wie wählen wir Hypothese und Alternative?

Die Hypothese wird so gewählt, dass der Fehler 1. Art der "schlimmere" Fehler ist.

Beispiel (Diagnose)

Ein Test gibt eine Indikation über eine Erkrankung.

Hypothese $\cong$ der Patient ist krank

Die Alternative ist also die Aussage, deren fälschliche Annahme schlimmer ist.

Gliederung

Testtheorie vs. Schätztheorie

- Ein *Punktschätzer* ordnet einer Stichprobe x einen Wert aus dem Parameterraum zu.
  - Beispiel: Maximum Likelihood-Schätzer
- Ein *Bereichsschätzer* ordnet einer Stichprobe x eine Teilmenge des Parameterraums zu.
  - Beispiel: Konfidenzbereich zum Niveau $1 - \alpha$
- Eine *Testfunktion* wählt anhand einer Stichprobe x zwischen zwei Parameterbereichen, der Hypothese und der Alternative.

Bereichsschätzer / Konfidenzintervalle

Sei $\Theta \subset \mathbb{R}^d$ $(d \geq 1)$

- Ein Konfidenzintervall ist eine Abbildung $\mathfrak{X} \longrightarrow \mathfrak{P}(\mathbb{R}^d)$.
- Ein Konfidenzintervall zum Niveau α erfüllt die Bedingung

$$W_\theta(\{x \in \mathfrak{X} : \theta \in C(x)\}) \geq 1 - \alpha$$

Bemerkung

Hier besteht ein Zusammenhang zur Testtheorie!!

Beispiel: Normalverteilung

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$. Dann gilt

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Weiter gilt.

$$\begin{aligned} 1 - \alpha &= P\Big(- z_{1-\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\Big) \\ &= P\big(\overline{X} - z_{1-\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \overline{X} + z_{1-\alpha/2}\sigma/\sqrt{n}\big) \end{aligned}$$

Hierbei: $z_{1-\alpha/2} = q_{1-\alpha/2}(\mathcal{N}(0, 1))$.

In unserem Beispiel

- 100 Geburten werden untersucht, von denen 54 Mädchen und 46 Jungen sind
- Stimmt die Hypothese? Können wir dies “mit Sicherheit” sagen?

  Wieviele Jungengeburten erwarten wir, falls Mädchen- und Jungengeburten mit gleicher Wahrscheinlichkeit auftreten? Bei was für einer Stichprobe machen wir bei Wahl der Hypothese eher einen Fehler?

Gliederung

Statistisches Experiment

Definition

Ein *statistisches Experiment* ist ein Tripel $\mathcal{E} = (\mathfrak{X}, \mathfrak{A}, (P_\theta^X)_{\theta\in\Theta})$ mit

- einer nichtleeren Menge $\mathfrak{X}$, dem *Stichprobenraum*,
- einer σ-Algebra $\mathfrak{A}$ über $\mathfrak{X}$ und
- einer nichtleeren Familie $(P_\theta^X)_{\theta\in\Theta}$ von Wahrscheinlichkeitsverteilungen über $(\mathfrak{X}, \mathfrak{A})$ mit $P_\theta^X \neq P_{\theta'}^X$ für $\theta \neq \theta'$.

Im folgenden sei stets $W_\theta = P_\theta^X$.

Modellierung

In unserem Beispiel wählen wir:

- $\mathfrak{X} = \{0, 1\}^n$, wobei n die Anzahl der Versuchsbeobachtungen ist,
- $\mathfrak{A} = \mathfrak{P}(\mathfrak{X})$,
- $P_\theta^X = \bigotimes_{i=1}^n \mathfrak{B}(1, \theta)$, $\theta \in \Theta = (0, 1)$,

wobei

- $0 \mathrel{\widehat{=}}$ Mädchengeburt
- $1 \mathrel{\widehat{=}}$ Jungengeburt.

Während die Wahl von $\mathfrak{X}$ und $\mathfrak{A}$ kanonisch ist, liegen der Wahl von P_θ^X *Modellierungsannahmen* zugrunde.

Nullhypothese und Alternativthese

Beim Testen zerlegt man die Parametermenge Θ disjunkt in *Nullhypothese* H und *Alternative* K:

$$\Theta = H + K.$$

Beispiel

Im Beispiel der Mädchen- vs. Jungengeburten setzen wir $H = [0, 0.5]$ und $K = (0.5, 1]$.

In unserem Falle liegt ein ***einseitiges*** *Testproblem mit linksseitiger Hypothese vor.*

Tests

Definition

Jede messbare Funktion $\varphi : \mathfrak{X} \to [0, 1]$ heißt *Test* oder *Testfunktion*.
φ heißt genau dann randomisiert, wenn $\{\varphi \in (0, 1)\} \neq \emptyset$.

Bemerkung

Interpretation eines Testwerts $\varphi(x)$:

$$\varphi(x) = \begin{cases} 1, & \varphi \text{ rät, die Alternative zu wählen,} \\ 0, & \varphi \text{ rät, die Hypothese zu wählen,} \\ \gamma \in (0, 1) & \varphi \text{ rät, ein unabh. Zufallsexp. durchzu-} \\ & \text{führen, das mit W'keit } \gamma \text{ zur Wahl} \\ & \text{der Alternative führt.} \end{cases}$$

Ein optimaler Test

Eine optimale Test- bzw. Entscheidungsfunktion sähe so aus:

- Liegt die Hypothese vor, so liefert ϕ stets 0, ansonsten stets 1. Beide Fehler treten mit Wahrscheinlichkeit 0 auf.
- Formal:

$$\varphi(x) = \begin{cases} 1, & \text{falls } \theta \in K, \\ 0, & \text{falls } \theta \in H \end{cases}$$

Gibt es einen optimalen Test?

Fehlerwahrscheinlichkeiten

Bei einem gegebenen Test gelten:

$$\begin{aligned} E_\theta\varphi(X) &= \text{W'keit für den Fehler 1. Art, falls } \theta \in H, \\ 1 - E_\theta\varphi(X) &= \text{W'keit für den Fehler 2. Art, falls } \theta \in K. \end{aligned}$$

Die Gütefunktion

Definition

Für einen Test φ heißt die Funktion $\theta \mapsto E_\theta \varphi(X)$ die *Gütefunktion* oder *Operationscharakteristik (OC)* von φ.

Die Gütefunktion

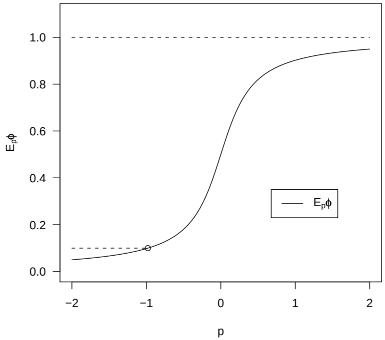

Abbildung: Gütefunktion eines einseitigen Tests z.N. α

Wahl des Tests

- Aus theoretischen Gründen gibt es i. A. keine gleichmäßig besten Tests.
- Daher wählt man für Testprobleme üblicherweise beste Tests innerhalb von Teilklassen der Menge der Tests aus.
- Eine der gängigsten Testklassen ist die der Tests zum Niveau α.

Tests zum Niveau α

Definition

Sei $\alpha \in [0, 1]$ ein vorgegebenes *Irrtums-* oder *Signifikanzniveau*. Dann heißt φ Test zum Niveau α, falls

$$E_\theta \varphi(X) \; \leq \; \alpha \quad \text{für alle } \theta \in H$$

gilt. Wir definieren Φ_α als die Menge der Tests zum Niveau α.

Gleichmäßig bester Test zum Niveau α

Definition

φ heißt gleichmäßig bester Test zum Niveau α, falls φ ein Test zum Niveau α ist, der die W'keit für einen Fehler 2. Art unter allen Test zum Niveau α gleichmäßig minimiert.

Wie finden wir einen gleichmäßig besten Test zum Niveau α?

Weitere Teilklassen der Menge der Tests

- Üblicherweise wählt man den gleichmäßig besten Test zum Niveau α für ein $\alpha \in \{0.01, 0.05, 0.1\}$.
- Manchmal (z. B. in zweiseitigen Testsituationen) gibt es keinen gleichmäßig besten Test z.N. α. Dann geht man z. B. zur Teilklasse der unverfälschten Tests zum Niveau α über.
- Manchmal (z. B. in der Situation des exakten Tests von Fisher oder des t-Tests) existieren keine gleichmäßig besten unverfälschten Tests zum Niveau α. Dann geht man zu noch kleineren Testklassen (z. B. J-ähnlichen Tests z. N. α) über.

Unverfälschte Tests zum Niveau α

Definition

Ein Test φ heißt *unverfälscht zum Niveau α (für H vs. K)*, falls $E_\theta\varphi(X) \leq \alpha$ für alle $\theta \in H$ und $E_\theta\varphi(X) \geq \alpha$ für alle $\theta \in K$ gilt.

Definition

Ein Test φ heißt *glm. bester unverfälschter Test zum Niveau α*, falls φ ein unverfälschter Test z. N. α ist und die Wahrscheinlichkeit für einen Fehler 2. Art unter allen unverfälschten Tests z. N. α glm. minimiert.

Warum macht man nicht Nägel mit Köpfen und gibt gleichzeitig "$1 - E_\theta\varphi(X) \geq \alpha$ für alle $\theta \in K$" als Schranke für den Fehler 2. Art vor?

Gliederung

Gliederung

Voraussetzungen

- Voraussetzung einer parametrischen Verteilungsfamilie mit Parametermenge $\Theta \subset \mathbb{R}$ oder $\Theta \subset \mathbb{R}^d$ $(d \geq 2)$
  - $B(1, \theta)$-Verteilung mit Parameterraum $(0, 1)$
  - $N(\mu, \sigma^2)$-Verteilung mit Parameterraum $\mathbb{R} \times (0, \infty)$
- Die Hypothese ist ein Intervall oder ein Punkt aus Θ

Wir suchen gleichmäßig beste Tests aus einer geeigneten Klasse von Testfunktionen.

Ein-Stichproben-Fall

Ein-Stichproben-Fall: Es wird ein einziges Merkmal X auf der Basis einer einfachen Zufallsstichprobe $(X_1, \ldots, X_n)$ bzgl. interessierender Fragestellungen getestet, z. B. die Lage von

- Mittelwert,
- Median,
- Varianz

betreffend etc.

Zwei-Stichproben-Fall

In diesem Fall wird ein Merkmal unter zwei Bedingungen untersucht oder man betrachtet zwei Merkmale, die am selben Merkmalsträger erhoben werden:

(a) Zwei unabhängige Zufallsstichproben $(X_{1,1}, \ldots, X_{1,n_1}), (X_{2,1}, \ldots, X_{2,n_2})$, $n_1, n_2 \in \mathbb{N}$, wobei sich die Randbedingungen bei der Entnahme der Stichproben in genau einer Randbedingung unterscheiden.

(b) Ein Merkmal unter zwei verschiedenen Bedingungen am selben Merkmalsträger: $(X_{1,1}, X_{1,2}), \ldots, (X_{n,1}, X_{n,2})$ (verbundene Stichproben, *matched pairs*).

(c) Zwei Merkmale X und Y am selben Merkmalsträger (unter jeweils gleichen Bedingungen): $(X_1, Y_1), \ldots, (X_n, Y_n)$ (verbundene Stichproben).

(Einseitiger) exakter Binomialtest

Definition

Der (einseitige) *exakte Binomialtest* φ^* für $H: \theta \leq \theta_0$ vs. $K: \theta > \theta_0$ ist durch

$$\varphi^*(x) = \begin{cases} 1, & \\ \gamma^* \in [0,1), & \text{falls } T(x) \gtreqqless c^*, \\ 0, & \end{cases}$$

definiert.

Bemerkung

Der (einseitige) *exakte Binomialtest* ist ein gleichmäßig bester Test zum Niveau α für ein einseitiges Testproblem, wie etwa unser Testproblem “Jungen- vs. Mädchengeburten”.

Prüfgröße

Definition

In obiger Situation ist $T(X)$ die Prüfgröße und c^* der kritische Wert.

Die Prüfgröße stellt eine Verdichtung der Daten dar, anhand derer über Verwerfen oder Beibehalten der Hypothese entschieden wird.

Welche Form haben die Prüfgröße und der kritische Wert?

Gliederung

Einfachster Fall: Neyman-Pearson-Lemma

Im einfachsten Fall gilt

- $H = \{\theta_0\}$ und $K = \{\theta_1\}$
- W_0 $(= W_{\theta_0})$ und W_1 $(= W_{\theta_1})$ haben bezüglich eines gemeinsamen dominierenden Maßes Dichten f_0 und f_1

Nach dem Neyman-Pearson-Lemma sind die folgenden Eigenschaften notwendig und hinreichend für einen gleichmäßig besten Test ψ zum Niveau α:

$$E_0\psi(X) = \int \psi dW_0 = \alpha \tag{1}$$

$$\psi(x) = \begin{cases} 1, \\ 0, \end{cases} \quad \text{falls } f_1(x) \gtrless kf_0(x) \ \mu\text{-f.ü.} \tag{2}$$

wobei $k \in [0, \infty)$ geeignet gewählt wird.

Der Neyman-Pearson-Test

Der gleichmäßig beste Test ψ im soeben beschriebenen Fall lässt sich schreiben als

$$\psi(x) \;=\; \begin{cases} 1, & \\ \gamma, & \text{falls } T(x) \gtreqless k \\ 0, & \end{cases}$$

mit

$$T(x) \;:=\; \begin{cases} \frac{f_1(x)}{f_0(x)}, & \text{falls } f_0(x) > 0 \\ \infty, & \text{falls } f_0(x) = 0. \end{cases}$$

In der Situation aus dem Neyman-Pearson-Lemma erhalten wir den gleichmäßig besten Test ψ zum Niveau α durch die Wahl

$$k := c(W_0^T, \alpha) = \inf\{y \in \mathbb{R} : W_0(T > y) \leq \alpha\}.$$

α-Fraktil

Definition

$c(W_0^T, \alpha)$ ist das α-Fraktil von T bezüglich W_0.

- Allgemein: $c(Q, \alpha) = \inf\{y \in \mathbb{R} : Q\big((y, \infty)\big) \leq \alpha\}$ ist das α-Fraktil eines W'Maßes Q.
- Ist F die Verteilungsfunktion von Q und F^{-1} die Quantilsfunktion, so ist $c(Q, \alpha) = q_{1-\alpha} := F^{-1}(1-\alpha)$.
- Mittels $c(W_0^T, \alpha)$ können wir einen Test mit den Eigenschaften (1) und (2) aus dem Neyman-Pearson-Lemma definieren.

Einseitige Testprobleme

Der nächstschwererere Fall.

Definition

Annahmen:	$(W_\theta)_{\theta\in\Theta}$ dominierte Familie Vtlg., wobei $\Theta \subset \mathbb{R}$
Hypothesen:	(a) $H_0\colon \theta \leq \theta_0$ vs. $K\colon \theta > \theta_0$, (b) $H_0\colon \theta \geq \theta_0$ vs. $K\colon \theta < \theta_0$.

Wie sieht ein gleichmäßig bester Test aus?

Familien mit monotonem Dichtequotienten

Das Neyman-Pearson-Lemma lässt sich verallgemeinern, wenn folgende Monotonieeigenschaft vorliegt:

Definition

Sei $\mathcal{E} = (\mathfrak{X}, \mathfrak{A}, (W_\theta)_{\theta\in\Theta})$ ein dominiertes statistisches Experiment mit dominierendem Maß μ. Dann heißt $(W_\theta)_{\theta\in\Theta}$ *Familie mit monotonem Dichtequotienten*, falls eine Statistik $T : \mathfrak{X} \to \mathbb{R}$ existiert, so dass für alle $\theta_0 < \theta_1$

$$\frac{f_{\theta_1}}{f_{\theta_0}} = g_{\theta_0,\theta_1} \circ T \qquad \mu\text{-f. ü.}$$

für eine monoton wachsende Funktion g_{θ_0,θ_1} gilt.

Bemerke: Beim N.P.-Lemma haben wir $T(x) = \frac{f_1(x)}{f_0(x)}$ gewählt.

Verbindung zum Neyman-Pearson-Lemma

Die Entscheidungsregel für den Neyman-Pearson-Test lässt sich für beliebige $\eta_0 < \eta_1$ verallgemeinern:

Für alle $x \in \mathfrak{X}$ mit $f_{\eta_0}(x) \neq 0$ oder $f_{\eta_1}(x) \neq 0$ gilt

$$\begin{aligned} & f_{\eta_1}(x) \gtreqless k(\eta_0, \eta_1) f_{\eta_0}(x), \\ \Rightarrow \quad & g_{\eta_0,\eta_1} \circ T(x) \gtreqless g_{\eta_0,\eta_1}(c^*) \qquad (*) \\ \Rightarrow \quad & T(x) \gtreqless c^* \end{aligned}$$

mit $k(\eta_0, \eta_1) := g_{\eta_0,\eta_1}(c^*)$.

Die letzte Ungleichung hängt nicht mehr von η_0 und η_1 ab. Wir benutzen sie zur Definition des Tests.

Gleichmäßig bester Test zum Niveau α

Mit

$$\varphi^*(x) = \begin{cases} 1 \\ \gamma^* \in [0,1) \\ 0 \end{cases}, \text{ falls } T(x) \substack{> \\ = \\ <} c^*,$$

wobei

$$c^* := c(W_{\theta_0}^T, \alpha) \text{ und } \gamma^* := \begin{cases} \frac{\alpha - W_{\theta_0}(T > c^*)}{W_{\theta_0}(T = c^*)}, & \text{falls } W_{\theta_0}(T = c^*) > 0, \\ 0, & \text{falls } W_{\theta_0}(T = c^*) = 0, \end{cases},$$

folgt

- $E_{\theta_0}\varphi^*(X) = \alpha$ (per Definition)
- $E_\theta\varphi^*(X) \leq \alpha$ für alle $\theta < \theta_0$
- $E_\theta\varphi^*(X) = \max_{\varphi:\varphi\in\Phi_\alpha} E_\theta\varphi(X)$ für alle $\theta > \theta_0$

Beispiel: Binomialtest

Für alle $x \in \mathfrak{X}$ sei $T(x) = s_n = \sum_{i=1}^{n} x_i$. Dann gilt

$$
\begin{aligned}
& f_{\eta_1}(x) \gtreqqless k(\eta_0, \eta_1) f_{\eta_0}(x) \\
\Leftrightarrow \; & \left(\frac{\eta_1}{\eta_0}\right)^{s_n} \left(\frac{1-\eta_1}{1-\eta_0}\right)^{n-s_n} \gtreqqless k(\eta_0, \eta_1) \\
\Leftrightarrow \; & g_{\eta_0,\eta_1}(s_n) \gtreqqless k(\eta_0, \eta_1) := g_{\eta_0,\eta_1}(c^*) \\
\Leftrightarrow \; & T(x) \gtreqqless c^*,
\end{aligned}
$$

wobei $g_{\eta_0,\eta_1}(t) = \left(\frac{\eta_1}{\eta_0}\right)^{t} \left(\frac{1-\eta_1}{1-\eta_0}\right)^{1-t}$.

Gleichmäßig bester Test zum Niveau α

Mit

$$\varphi^*(x) \ = \begin{cases} 1 \\ \gamma \in [0,1) \\ 0 \end{cases} \quad , \text{ falls } s_n \gtreqless c^*$$

erhalten wir im Falle $H = \{\theta < \theta_0\}$ einen glm. besten Test zum Niveau α, wobei hier

$$c^* = c(B(n, \theta_0), \alpha) \ \text{ und } \ \gamma := \frac{\alpha - W_{\theta_0}(S_n > c^*)}{W_{\theta_0}(S_n = c^*)}$$

gilt.

Beim Beweis müssen wir 2 mal das Neyman-Pearson-Lemma anwenden.

1. Anwendung: $\phi^* \in \Phi_\alpha$

Sei $\eta_0 := \theta \in H$ beliebig und $\eta_1 := \theta_0$. D.h. wir betrachten $\{\theta_0\}$ wie einen Alternativenparameter und η_0 als Hypothesenparameter.

Aus dem Neyman-Pearson-Lemma folgt mit $\beta = E_{\eta_0}\varphi(X)$

$$\alpha = E_{\theta_0}\varphi^*(X) = E_{\eta_1}\varphi^*(X) = \max_{\varphi, E_{\eta_0}\varphi(X) \leq \beta} E_{\eta_1}\varphi(X) \geq \beta,$$

also $\alpha \geq \beta$, wobei die Ungleichung durch Wahl von $\varphi \equiv \beta$ verifiziert werden kann.

2. Anwendung: $E_\theta \phi^*(X)$ maximal für alle $\theta \in K$

Sei $\theta \in K$ beliebig.

Aus dem Neyman-Pearson-Lemma folgt mit $W_0 = W_{\theta_0}$ und $W_1 = W_\theta$

$$E_\theta \varphi^*(X) = \max_{\varphi, E_{\theta_0}\varphi(X) \leq \alpha} E_\theta \varphi(X).$$

Rechtseitige Hypothesen

Beispiel: Produktionsprozess

Bei einer Gut-Schlecht-Prüfung soll anhand von 100 untersuchten Stücken eine Entscheidung über die Umstellung des Produktionsprozesses getroffen werden.

> *Wie sieht das Modell aus? Wählen Sie die Hypothese.*
>
> *Bei welcher Art von Stichprobe würde man eher zu einer Umstellung tendieren?*

Rechtseitige Hypothesen

Sei das Experiment $\mathfrak{E} = (\mathfrak{X}, \mathfrak{A}, (W_\theta)_{\theta\in\Theta})$ gegeben, wobei $(W_\theta)_{\theta\in\Theta})$ einen monotonen Dichtequotienten in T habe.

Analog zum linksseitigen Test hat ein gleichmäßig bester Test zum Niveau α die Gestalt:

$$\varphi^*(x) = \begin{cases} 1 \\ \gamma^* \in [0,1) \\ 0 \end{cases} , \text{ falls } T(x) \substack{< \\ = \\ >} c^*,$$

wobei

$$c^* := -c(W_{\theta_0}^{-T}, \alpha) \text{ und } \gamma^* := \begin{cases} \frac{\alpha - W_{\theta_0}(T<c^*)}{W_{\theta_0}(T=c^*)}, & \text{falls } W_{\theta_0}(T=c^*) > 0, \\ 0, & \text{falls } W_{\theta_0}(T=c^*) = 0, \end{cases},$$

Ablehnungsbereich bei rechtsseitiger Hypothese

- Bei rechtsseitigen Tests lehnen wir die Hypothese ab, falls $T < c^*$
- Im Falle $T = c^*$ wird die Entscheidung randomisiert
- $c^* = \sup\{t \in \mathbb{R} : W_{\theta_0}(T < t) \leq \alpha\} = -c(W_{\theta_0}^{-T}, \alpha)$
- Bei symmetrischen Verteilungen gilt $c^* = -c(W_{\theta_0}^{T}, \alpha)$
- Bei stetigen Verteilungen ist $c^* = q_\alpha$

Glm. bester Test z. N. α bei einseitigen Testproblemen mit monotonem DQ

Definition

Teststatistik:	$T(X)$
Verteilung unter θ_0:	$W_{\theta_0} = P_{\theta_0}^{T \circ X}$
Ablehnungsbereich:	(a) $T(X) > c(W_{\theta_0}^T, \alpha)$
	(b) $T(X) < -c(W_{\theta_0}^{-T}, \alpha)$

Hierbei liegt in Fall (a) eine linksseitige Hypothese und in Fall (b) eine rechtsseitige Hypothese vor.

Vorab: Einseitige Tests in R

binom.test(x = ,n = ,p = ,alternative="greater|less",conf.level=1-α)

Gliederung

Beispiel: Zweiseitiger Binomialtest

Ist die Münze, die für einen Losentscheid verwendet wird, gezinkt?

Bemerkung (Modell)

- $\mathcal{E} = (\{0,1\}, \mathfrak{P}(\{0,1\}), (B(1,\theta)_{\theta \in [0,1]}))$
- $\Theta = [0,1]$
- Hypothese: $H = \{\theta_0\}$ mit $\theta_0 = 0,5$

Zweiseitige Testprobleme

Sei $\Theta \subset \mathbb{R}$. Ein zweiseitiges Testproblem ist von der Form

- Zweiseitige Hypothesen

  $H = \{\theta \in \Theta : \theta \leq \theta_1 \text{ oder } \theta \geq \theta_2\}$ gegen $K = \{\theta \in \Theta : \theta_1 < \theta < \theta_2\}$

- Zweiseitige Alternativen

  $H = \{\theta \in \Theta : \theta_1 \leq \theta \leq \theta_2\}$ gegen $K = \{\theta \in \Theta : \theta < \theta_1 \text{ oder } \theta > \theta_2\}$

  $H = \{\theta_0\}$ gegen $K = \{\theta \in \Theta : \theta \neq \theta_0\}$.

Warum macht der Fall $K = \{\theta_0\}$ keinen Sinn?

In R unterstützte Hypothesen

In R lassen sich behandeln:

- Einseitige Testprobleme
- Testprobleme der Form $H = \{\theta_0\}$ gegen $K = \{\theta \in \Theta : \theta \neq \theta_0\}$

Exakter Binomialtest (zweiseitig)

Definition

Der (zweiseitige) *exakte Binomialtest* φ^* für zweiseitige Alternativen ist durch

$$\varphi^*(X) = \begin{cases} 1, & \text{falls } T(X) \notin [c_0^*, c_1^*], \\ \gamma_i^* \in [0,1), & \text{falls } T(X) = c_i^*, \\ 0, & \text{falls } T(X) \in (c_0^*, c_1^*) \end{cases} \tag{3}$$

für geeignete Konstanten γ_0^*, γ_1^*, c_0^* und c_1^* definiert. Dabei ist die Testgröße $T(X) := \sum_{i=1}^n X_i$.

Bei Wahl geeigneter Konstanten ist φ^* der gleichmäßig beste unverfälschte Test zum Niveau α.

Allgemeine Gestalt gleichmäßig bester unverfälschter Test

Die Testgestalt (3) ist die allgemeine Form eines gleichmäßig besten unverfälschten Tests zum Niveau α, wenn wir γ_i^* und c_i^* $(i = 1, 2)$ passend wählen (s. S. 106 Alsmeyer Skripten zur Statistik).

Beidseitige Tests in R

Hier exemplarisch die Eingabe des zweiseitigen Binomialtests

```
binom.test(... ,alternative=two.sided")
```

Gliederung

Gliederung

p-Wert und Prüfgröße

In Anwendungen vergleicht man häufig nicht die Prüfgröße $T(x)$ mit dem kritischen Wert, sondern den ***p*-Wert** mit dem Signifikanzniveau α.

Der *p*-Wert ist das kleinste Niveau α zu dem man bei Vorliegen der Stichprobe x und Prüfgrößenwert $T(x) = t$ die Nullhypothese noch ablehnen kann.

Definition des p-Wertes

Definition

Der *p-Wert* eines Tests ist die Wahrscheinlichkeit, unter der Nullhypothese W_{θ_0} den beobachteten Prüfwert $x \in \mathfrak{X}$ oder einen in Richtung der Alternative extremeren Wert zu erhalten.

Im Falle linksseitiger Hypothesen bedeutet dies formal $p = P_{\theta_0}(T(X) \geq T(x))$.

Bemerkung

Der Ablehnungsbereich der Hypothese lautet “$p \leq \alpha$”.

Eingabe von Tests

Worauf muss man bei der Eingabe eines Tests in R achten?

- Der Wert x repräsentiert die Realisierung der Prüfgröße
- R gibt neben dem p-Wert auch das Konfidenzintervall C zum Niveau α an.
- Ablehnung der Hypothese erfolgt anhand der Fragestellung "$p \leq \alpha$" oder $\theta_0 \notin C$.

Binomialtest in R

Beispiel

Der einseitige exakte Binomialtest eignet sich für das Ausgangsproblem „Jungen- vs. Mädchengeburten“.
Der Test wird in R mit dem Befehl `binom.test` aufgerufen.

```
> binom.test(x = 46, n = 100, p = 0.5, alternative
= "greater", conf.level = 0.95)

Exact binomial test

data: 46 and 100
number of successes = 46, number of trials = 100,
p-value = 0.8159
```

Der *p*-Wert ist größer als 0.05. Die Hypothese wird angenommen.

Gliederung

Beispiel: Sollwertmessung

Bei der Produktion von LKW-Achsen werden stichprobenartig die Achslängen x gemessen. Dazu untersucht man $n = 100$ produzierte Stücke.

Ein Test soll sicherstellen, dass die Achslängen innerhalb der Toleranz schwanken. Der Gaußtest gibt uns dafür ein Intervall, innerhalb dessen die mittlere Achslänge $\overline{x}$ schwanken darf. Die Wahrscheinlichkeit dafür, dass sich $\overline{x}$ innerhalb der Toleranz befindet, soll 99% betragen.

Damit $\overline{x}$ eine mögliche Prüfgröße ist, müssen wir die Varianz der Normalverteilung als bekannt voraussetzen! Die Varianz kennen wir dabei aus bisherigen Messungen.

Wie sieht das Modell aus? Wie sieht die Hypothese aus? Welches Signifikanzniveau liegt vor?

Modell für die Sollwertmessung

Statistisches Modell, Hypothesenwahl und Signifikanzniveau:

- $\mathfrak{E} = (\mathbb{R}^n, \mathfrak{B}^n, (N(\mu, \sigma^2)^n)_{\mu \in \mathbb{R}})$
- $H = \{\mu_0\}$ gegen $K = \{\mu \neq \mu_0\}$
- $\alpha = 0.01$

Der Gaußtest

Definition

Seien $X_1, \ldots, X_n$ u. i. v. Zufallsvariablen mit $X_i \sim N(\mu, \sigma^2)$, σ^2 bekannt, bzw. mit beliebiger stetiger Verteilung und $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, $n \geq 30$. Man betrachte folgende Testprobleme:

(a) $H_0\colon \mu = \mu_0$ vs. $K\colon \mu \neq \mu_0$,
(b) $H_0\colon \mu \leq \mu_0$ vs. $K\colon \mu > \mu_0$,
(c) $H_0\colon \mu \geq \mu_0$ vs. $K\colon \mu < \mu_0$.

Basierend auf der Prüfgröße $T(X) = \sqrt{n}\frac{\overline{X}-\mu_0}{\sigma} \sim_{P_{\mu_0}} \mathcal{N}(0,1)$ fällt die Entscheidung für K im Testproblem

(a) im Falle $|T(X)| > q_{1-\alpha/2}(\mathcal{N}(0,1))$,
(b) im Falle $T(X) > q_{1-\alpha}(\mathcal{N}(0,1))$,
(c) im Falle $T(X) < -q_{1-\alpha}(\mathcal{N}(0,1)) = q_{\alpha}(\mathcal{N}(0,1))$.

Sollwertmessung

Bei 100 Achsen ergibt sich für die Prüfgröße ein Wert von z (Generierung in R). Den Betrag dieses Wertes vergleichen wir mit dem 0.995-Quantil der Standarnormalverteilung, also mit $q_{0.995}(\mathcal{N}(0,1)) = 2.5758$

Gliederung

Approximativer Binomialtest

Definition

Gegeben seien folgende Testprobleme über den Parameter θ einer $\mathfrak{B}(n, \theta)$-Verteilung:

(a) $H_0\colon \theta = \theta_0$ vs. $K\colon \theta \neq \theta_0$,

(b) $H_0\colon \theta \leq \theta_0$ vs. $K\colon \theta > \theta_0$,

(c) $H_0\colon \theta \geq \theta_0$ vs. $K\colon \theta < \theta_0$.

Basierend auf der Prüfgröße $T(X) = \frac{\sum_{i=1}^n X_i - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \sim_{P_{\theta_0}} \mathcal{N}(0,1)$ und dem vorgegebenen Niveau α fällt die Entscheidung für K im Testproblem

(a) im Falle $|T(X)| > q_{1-\alpha/2}(\mathcal{N}(0,1))$,

(b) im Falle $|T(X)| > q_{1-\alpha}(\mathcal{N}(0,1))$,

(c) im Falle $|T(X)| < -q_{1-\alpha}(\mathcal{N}(0,1)) = q_\alpha(\mathcal{N}(0,1))$.

Danke

DANKE FÜR EURE AUFMERKSAMKEIT