

Skript zur Vorlesung

Mathematische Statistik

von Prof. Dr. Michael Kohler

Wintersemester 2014/15

Inhaltsverzeichnis

1	Einführung	4
1.1	W-Theorie und Statistik	4
1.2	Zwei (moderne) Anwendungsbeispiele	5
1.3	Drei (klassische) Problemstellungen	6
1.4	Klassische parametrische Statistik	9
1.5	Nichtparametrische Statistik	9
2	Schätzung von Verteilungen	11
2.1	Die empirische Verteilung	11
2.2	VC-Theorie	13
3	Dichteschätzung	21
3.1	Motivation	21
3.2	Der Kerndichteschätzer	28
3.3	Ein Konsistenzresultat	30
4	Punktschätzungen	39
4.1	Problemstellungen und Beispiele	39
4.2	Konstruktion von Punktschätzungen	41
4.3	Optimale Schätzverfahren	48
4.4	Der Begriff des optimalen erwartungstreuen Schätzers	49
4.5	Die Informationsungleichung von Cramér-Rao	51
4.6	Suffizienz	56
5	Statistische Testverfahren	63
5.1	Einführung	63
5.2	Das Fundamentallemma von Neyman und Pearson	71
5.3	Tests bei monotonen Dichtequotienten	78
5.4	Tests im Zusammenhang mit der Normalverteilung	85
5.5	Robustheit von Tests	96

5.6	Zwei nichtparametrische Tests	100
5.6.1	Der Zeichentest	100
5.6.2	Der Wilcoxon-Rangsummen-Test	101
5.7	Multiples Testen	103
6	Bereichsschätzungen	109
6.1	Einführung	109
6.2	Anwendungsbeispiele	110
6.3	Konstruktion von Bereichsschätzungen mit Hilfe von stochastischen Pivots	110
6.4	Konstruktion von Bereichsschätzungen mit Hilfe von statistischen Tests	114
7	Einige nichtparametrische Testverfahren	116
7.1	Der Test von Kolmogoroff-Smirnow	116
7.2	Der χ^2 -Anpassungstest	122

1 Einführung

1.1 W-Theorie und Statistik

W-Raum (Ω, \mathcal{A}, P)

(mit Grundmenge $\Omega \neq \emptyset$, σ -Algebra $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ und W-Maß $P : \mathcal{A} \rightarrow [0, 1]$)

Zufallsvariablen (kurz: ZVen) $X_n, X : \Omega \rightarrow \mathbb{R}$ ($n \in \mathbb{N}$) (d. h. X_n, X sind $\mathcal{A} - \mathcal{B}$ -messbare Abbildungen).

X (und analog X_n) wird das W-Maß

$$P_X : \mathcal{B} \rightarrow [0, 1]$$

$$P_X(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\})$$

zugeordnet.

Die ZVen X, X_1, X_2, \dots, X_n seien für alle $n \in \mathbb{N}$ unabhängig und identisch verteilt (kurz: u. i. v.), d. h.:

- identisch verteilt: $P_X = P_{X_1} = \dots = P_{X_n}$
- unabhängig: $P_{(X, X_1, \dots, X_n)} = P_X \otimes P_{X_1} \otimes \dots \otimes P_{X_n}$.

In diesem Fall heißt X_1, \dots, X_n **Stichprobe** von X bzw P_X (genauer: unabhängige Stichprobe).

Typische Fragestellung der W-Theorie:

Verteilung von X sei bekannt.

Wie verhält sich dann $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$ für $\omega \in \Omega$ (sog. **Realisierung** der X_1, \dots, X_n).

z. B.: X sei integrierbar mit Erwartungswert $EX \in \mathbb{R}$. Was kann man dann über

$$\frac{1}{n} \sum_{i=1}^n X_i(\omega)$$

aussagen?

Nach dem starken Gesetz der großen Zahlen (*SGdGZ*) gilt:

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow EX \quad f.s.,$$

d. h. es existiert $A \in \mathcal{A}$ mit $P(A) = 1$ und

$$\frac{1}{n} \sum_{i=1}^n X_i(\omega) \rightarrow EX \quad (n \rightarrow \infty) \text{ für alle } \omega \in A.$$

Typische Fragestellung der Statistik:

Verteilung von X sei unbekannt.

Realisierung x_1, \dots, x_n von X_1, \dots, X_n sei gegeben.

Was kann man daraus über P_X schließen?

z. B.: Wie groß ist der Erwartungswert von X ?

Naheliegend: "Schätze" EX durch

$$T(x_1, \dots, x_n) = \frac{x_1 + \dots + x_n}{n}.$$

Fragen:

- Welche Eigenschaften hat diese Schätzung?
- Gilt es "bessere" Schätzungen?
- Was sind "optimale" Schätzungen?

etc.

1.2 Zwei (moderne) Anwendungsbeispiele

Beispiel 1.1: Positionsbestimmung mittels GPS

Anwendungsgebiete:

- Navigationssysteme für Schiffe, Autos, etc.
- Erdbebenfrüherkennung (z. B. in Japan)
- Militärische Anwendungen

Idee: Bestimme (durch Schnitt von Kugeloberflächen) Standort ausgehend von Entfernungen zu drei bekannten Punkten im Raum.

Vorgehen: ca. 30 Satelliten umkreisen die Erde in 20.200 km Höhe und senden ihre Position und Signalausendezeit im Sekundentakt zur Erde.
Bestimme durch Vergleich der Signalausendezeit und der Empfangszeit (mittels Lichtgeschwindigkeit) die Entfernung zu den Satelliten.

Probleme: Messungenauigkeiten durch: • Uhrenfehler (beim Empfänger)
• Veränderungen in der Ionosphäre

Ausweg: Entfernung von 4 - 5 Satelliten bestimmen und statistische Verfahren verwenden ...

Beispiel 1.2: Analyse von DNA-Microarray-Daten

Stoffwechsel von Zellen wird gesteuert durch Proteine (Eiweiße). Bei DNA-Microarrays wird statt Aktivität der Proteine (schwierig zu messen!) die Aktivität von Genen (Abschnitten der DNA) simultan für ca. 3.000 - 20.000 verschiedener Gene gemessen.

Ausgehend von diesen Messungen (d. h. Vektor bestehend aus 3.000 - 20.000 reeller Zahlen) sollen dann z. B. bei Tumorzellen Vorhersagen gemacht werden bzgl.:

- Ansprechen auf verschiedene Therapiearten
- Überlebenszeit der Patienten

etc.

Als "Stichprobe" vorhanden:

beobachtete Daten zu in der Vergangenheit erkrankten Patienten (u. a. Überlebenszeit, gewählte Therapie) zusammen mit aus Zellproben der Tumore gewonnenen DNA-Microarray-Daten.

1.3 Drei (klassische) Problemstellungen

Beispiel 1.3

Zur Heilung einer bestimmten Krankheit wurde eine neue Behandlungsmethode I entwickelt. Bei Anwendung auf $n = 10$ Patienten ergab sich in 8 Fällen ein Heilerfolg, in 2 Fällen ein Misserfolg. Lässt sich aufgrund dieser 10 Überprüfungen sagen, dass die neue Methode I häufiger zum Erfolg führt als eine herkömmliche Methode II, deren Heilungschance erfahrungsgemäß 65 % beträgt?

Problem: Heilerfolg hängt nicht nur von Behandlungsmethode, sondern auch von vielen anderen (zufälligen) Faktoren ab. Also könnte Anwenden von Methode I auf 10 andere Patienten auch 9 oder 6 oder ... Heilerfolge geben.

Im Folgenden: Stochastische Modellierung

Beobachtungen werden als Realisierungen von ZVen aufgefasst. Aufgrund der beobachteten Werte machen wir Aussagen über die Verteilung dieser ZVen.

Dazu: Setze

$$x_i = \begin{cases} 1 & , \text{ falls Heilerfolg bei } i\text{-ten Patienten} \\ 0 & , \text{ sonst} \end{cases}$$

$(i = 1, \dots, 10)$.

Fasse x_1, \dots, x_{10} als Realisierung von u. i. v. ZVen X_1, \dots, X_{10} auf, die nur die Werte in $\{0, 1\}$ annehmen.

Dann sind die X_i $b(1, \vartheta)$ verteilt mit $\vartheta = P[X_i = 1]$ ($i = 1, \dots, 10$).

Problemstellung 1: Schätzproblem

Schätze den Zahlenwert der Erfolgs-Wahrscheinlichkeit ϑ z. B. durch

$$g_1(x) := g_1(x_1, \dots, x_{10}) = \frac{(x_1 + \dots + x_{10})}{10} =: \bar{x}$$

oder

$$g_2(x) = x_1$$

oder

$$g_3(x) = \frac{x_1 + x_3 + x_7}{3}$$

oder ...

Problemstellung 2: Bereichsschätzproblem

Bestimme eine möglichst kleine Menge $C(x) \subseteq [0, 1]$, die ϑ mit möglichst großer Wahrscheinlichkeit überdeckt.

Wegen

$$V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{\vartheta(1-\vartheta)}{n} \approx \frac{\bar{x}(1-\bar{x})}{n}$$

ist eine naheliegende Bereichsschätzung:

$$C(x) = \left[\bar{x} - k \cdot \sqrt{\bar{x}(1-\bar{x})/10}, \bar{x} + k \cdot \sqrt{\bar{x}(1-\bar{x})/10} \right]$$

$\stackrel{\text{hier}}{\approx} [0,8 - 0,13 \cdot k, 0,8 + 0,13 \cdot k]$

mit $k > 0$.

Hierbei:

k groß \Rightarrow Entscheidung $\vartheta \in C(x)$ mit großer Wahrscheinlichkeit richtig, Intervall $C(x)$ groß

k klein \Rightarrow Entscheidung $\vartheta \in C(x)$ nur mit kleiner Wahrscheinlichkeit richtig, Intervall $C(x)$ klein.

Anhaltspunkt für Wahl von k :

Fasse \bar{x} als Realisierung eines normalverteilten ZV N auf und beachte:

$$P[N \in [\mu - k \cdot \sigma, \mu + k \cdot \sigma]] \approx \begin{cases} 0,68 & , k = 1 \\ 0,95 & , k = 2 \\ 0,997 & , k = 3. \end{cases}$$

Problemstellung 3: Testproblem

Ist die Erfolgswahrscheinlichkeit ϑ der neu entwickelten Methode größer als 0,65 oder nicht?

Aufgrund der Beobachtungen x_1, \dots, x_{10} möchte man hier zwischen den beiden Hypothesen

$$H_0 : \vartheta \leq 0,65 \text{ und } H_1 : \vartheta > 0,65$$

entscheiden.

Mögliche Entscheidungsvorschriften sind z. B. Entscheidung für H_1 , falls

$$\sum_{i=1}^{10} x_i \geq c \text{ mit } c \geq 0 \text{ (z.B. } c = 8)$$

oder falls

$$x_1 = x_2 = x_3 = 1$$

oder falls ...

1.4 Klassische parametrische Statistik

Hier wird vorausgesetzt, dass die Verteilung der Daten bis auf einen endlichdimensionalen Parameter bekannt ist.

Dies lässt sich wie folgt formalisieren: Sei $\Theta \subseteq \mathbb{R}^l$ eine Parametermenge, und für jedes $\theta \in \Theta$ sei ein Wahrscheinlichkeitsmaß w_θ auf \mathcal{B} gegeben. Ausgehend von einer Stichprobe

$$X_1, \dots, X_n$$

von unabhängig identisch verteilten Zufallsvariablen, für deren Verteilung gilt

$$\mathbf{P}_{X_1} = w_\theta \quad \text{für ein } \theta \in \Theta,$$

sind “Aussagen” über θ gesucht.

Dabei auftretende Problemstellungen sind:

- (i) Konstruiere eine *Punktschätzung* $T_n(X_1, \dots, X_n) \in \Theta$ von θ .
- (ii) Konstruiere *Bereichsschätzungen* $I(X_1, \dots, X_n) \subseteq \Theta$ von θ .
- (iii) Entscheide zwischen Hypothesen wie

$$H_0 : \theta = \theta_0 \quad \text{und} \quad H_1 : \theta \neq \theta_0$$

mittels eines *statistischen Tests*.

Als Fragen dazu treten dann auf:

- Wie konstruiert man entsprechende Verfahren?
- Welche Eigenschaften haben diese Verfahren?
- Was sind optimale Verfahren?

Entsprechende Fragen wurden ansatzweise schon in der “Einführung in die Stochastik” behandelt.

1.5 Nichtparametrische Statistik

In der nichtparametrischen Statistik kann das zu schätzende Objekt nicht durch einen endlichdimensionalen Parameter beschrieben werden. Beispiele dafür sind:

a) Schätzung von Verteilungen

$X, X_1, X_2 \dots$ seien unabhängig identisch verteilte \mathbb{R}^d -wertige Zufallsvariablen. Ausgehend von

$$X_1, \dots, X_n$$

soll hier das Wahrscheinlichkeitsmaß

$$\mathbf{P}_X : \mathcal{B}_d \rightarrow [0, 1]$$

geschätzt werden.

Hier gilt

$$\mathbf{P}_X(B) = \mathbf{E}[1_B(X)],$$

und schätzt man den Erwartungswert wie oben durch ein Stichprobenmittel, so führt dies auf die Schätzung

$$\hat{\mathbf{P}}_X(B) = \frac{1}{n} \sum_{i=1}^n 1_B(X_i).$$

$\hat{\mathbf{P}}_X : \mathcal{B}_d \rightarrow [0, 1]$ heißt *empirische Verteilung* zu X_1, \dots, X_n .

b) Schätzung von Dichten

In a) sei nun $f : \mathbb{R}^d \rightarrow \mathbb{R}$ eine Dichte von X (bzgl. des LB-Maßes). Ausgehend von

$$X_1, \dots, X_n$$

soll dann f geschätzt werden, d.h. gesucht ist eine Schätzfunktion

$$f_n : \mathbb{R}^d \rightarrow \mathbb{R}, f_n(x) = f_n(x, X_1, \dots, X_n).$$

c) Schätzung von Regressionsfunktionen

Hier sind $(X, Y), (X_1, Y_1), (X_2, Y_2), \dots$ unabhängige identisch $\mathbb{R}^d \times \mathbb{R}$ -wertige Zufallsvariablen mit $\mathbf{E}(Y^2) < \infty$. Sei durch

$$m(x) = \mathbf{E}\{Y|X = x\} \quad (x \in \mathbb{R}^d)$$

die sog. *Regressionsfunktion* $m : \mathbb{R}^d \rightarrow \mathbb{R}$ definiert.

Ausgehend von

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

soll hier $m : \mathbb{R}^d \rightarrow \mathbb{R}$ geschätzt werden, d.h. gesucht ist eine Schätzung

$$m_n : \mathbb{R}^d \rightarrow \mathbb{R}, m_n(x) = m_n(x, (X_1, Y_1), \dots, (X_n, Y_n)).$$

2 Schätzung von Verteilungen

X_1, X_2, \dots, X_n seien u. i. v. \mathbb{R}^d -wertige ZVen.

μ sei die Verteilung von X_1 , d. h.

$$\mu : \mathcal{B}_d \rightarrow \mathbb{R}_+, \quad \mu(B) = P_{X_1}(B) = P(X_1^{-1}(B))$$

geg.: Realisierungen x_1, \dots, x_n von X_1, \dots, X_n

ges.: Schätzung

$$\hat{\mu}_n(\cdot) = \hat{\mu}_n(\cdot, x_1, \dots, x_n) : \mathcal{B}_d \rightarrow \mathbb{R}$$

von

$$\mu : \mathcal{B}_d \rightarrow \mathbb{R}_+.$$

Hierbei ist für jede $B \in \mathcal{B}_d$

$$\hat{\mu}_n(B) = \hat{\mu}_n(B, x_1, \dots, x_n) \in \mathbb{R}$$

eine Schätzung der Wahrscheinlichkeit $\mu(B) = P_{X_1}(B) = P[X_1 \in B]$

2.1 Die empirische Verteilung

Def. 2.1: Die Verteilung

$$\mu_n : \mathcal{B}_d \rightarrow \mathbb{R}_+$$

$$\mu_n(B) = \frac{1}{n} \sum_{i=1}^n I_B(x_i)$$

(mit

$$I_B(x_i) = \begin{cases} 1 & , \text{ falls } x_i \in B \\ 0 & , \text{ falls } x_i \notin B \end{cases} \quad)$$

heißt **empirische Verteilung** zu x_1, \dots, x_n .

Einfach zu sehen: μ_n ist W-Maß

(d. h. $\mu_n(\emptyset) = 0, \mu_n(\mathbb{R}^d) = 1, \mu_n\left(\bigcup_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} \mu_n(B_k)$ für paarweise disjunkte $B_1, B_2, \dots \in \mathcal{B}_d$).

Sind die Punkte x_1, \dots, x_n paarweise verschieden, so gilt

$$\mu_n(\{x_i\}) = \frac{1}{n} \quad (i = 1, \dots, n)$$

und

$$\mu_n(\mathbb{R}^d \setminus \{x_1, \dots, x_n\}) = 0,$$

d. h. jedem der x_1, \dots, x_n wird die Masse $\frac{1}{n}$ zugeteilt.

Allgemein gilt:

$$\mu_n(B) = \frac{\#\{1 \leq i \leq n : x_i \in B\}}{n}.$$

Ist μ_n die empirische Verteilung zu X_1, \dots, X_n , so gilt nach dem starken Gesetz der großen Zahlen:

$$(2.1) \quad \mu_n(B) = \frac{1}{n} \sum_{i=1}^n I_B(X_i) \xrightarrow{f.s.} E\{I_B(X_1)\} = P[X_1 \in B] \\ = \mu(B)$$

(da $I_B(X_1), I_B(X_2), \dots$ u. i. v. und integrierbar).

Im Folgenden: Verschärfung dieser Aussage.

Sei F die zu μ gehörende Verteilungsfunktion, d. h.

$$F : \mathbb{R}^d \rightarrow \mathbb{R}_+ \\ F(x) := \mu((-\infty, x]),$$

wobei für $x = (x^{(1)}, \dots, x^{(d)})$ gesetzt wird:

$$(-\infty, x] = (-\infty, x^{(1)}] \times (-\infty, x^{(2)}] \times \dots \times (-\infty, x^{(d)}]$$

Aus W-Theorie bekannt:

Das W-Maß μ ist durch seine Verteilungsfunktion F bereits eindeutig festgelegt, d. h.

$$\mu : \mathcal{B}_d \rightarrow \mathbb{R}_+, B \mapsto \mu(B)$$

ist eindeutig festgelegt durch

$$(-\infty, x] \mapsto \mu((-\infty, x]) \quad (x \in \mathbb{R}^d)$$

F kann geschätzt werden durch die zu μ_n gehörende Verteilungsfunktion.

Def. 2.2:

Die zur empirischen Verteilung μ_n gehörende Verteilungsfunktion

$$F_n : \mathbb{R}^d \rightarrow \mathbb{R}_+$$

$$F_n(x) = \mu_n((-\infty, x]) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(x_i)$$

heißt **empirische Verteilungsfunktion** zu x_1, \dots, x_n .

Ist F_n die empirische Verteilungsfunktion zu X_1, \dots, X_n , so gilt für alle $x \in \mathbb{R}^d$ analog zu (2.1):

$$(2.2) \quad F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i) \xrightarrow{f.s.} EI_{(-\infty, x]}(X_1) = P[X_1 \leq x] = F(x).$$

Diese Aussage lässt sich verschärfen:

Satz 2.1 (Satz von Glivenko-Cantelli bzw. Hauptsatz der Mathematischen Statistik)

Sind X_1, X_2, \dots u.i.v. \mathbb{R}^d -wertige ZVen mit Verteilungsfunktion F , und ist F_n die empirische Verteilungsfunktion zu X_1, \dots, X_n , so gilt:

$$\sup_{x \in \mathbb{R}^d} |F_n(x) - F(x)| \rightarrow 0 \quad \text{f.s.}$$

Der Beweis von Satz 2.1 erfolgt im allgemeineren Rahmen im nächsten Abschnitt.

2.2 VC-Theorie

Satz 2.1 lässt sich umformulieren zu

$$(2.3) \quad \sup_{A \in \mathcal{A}} |\mu(A) - \mu_n(A)| \rightarrow 0 \text{ f. s.}$$

für

$$\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}^d\}.$$

Im Folgenden leiten wir hinreichende Bedingungen für die Gültigkeit von (2.3) im Falle allgemeiner Mengensysteme $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^d)$ her. Dabei werden evtl. auftretende Messbarkeitsprobleme ignoriert.

Def. 2.3 Sei \mathcal{A} eine Klasse von Mengen $A \subseteq \mathbb{R}^d$, und sei $n \in \mathbb{N}$. Der **n-te Zerlegungskoeffizient** von \mathcal{A} ist

$$s(\mathcal{A}, n) = \max_{x_1, \dots, x_n \in \mathbb{R}^d} \#\{A \cap \{x_1, \dots, x_n\} : A \in \mathcal{A}\}.$$

Klar: $0 \leq s(\mathcal{A}, n) \leq 2^n =$ maximale Anzahl der Teilmengen einer n-elementigen Menge.

Beispiel 2.1

- a) Sei $d = 1$ und $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$. Sind $x_1, \dots, x_n \in \mathbb{R}$ mit $x_1 \leq x_2 \leq \dots \leq x_n$, so gilt

$$\{(-\infty, x] \cap \{x_1, \dots, x_n\} : x \in \mathbb{R}\} \subseteq \{\emptyset, \{x_1\}, \{x_1, x_2\}, \dots, \{x_1, \dots, x_n\}\}$$

(wobei Gleichheit für $x_1 < x_2 < \dots < x_n$ besteht).

Daraus folgt $s(\mathcal{A}, n) = n + 1$.

- b) Sei $d > 1$ und $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}^d\}$.

Dann gilt

$$s(\mathcal{A}, n) \leq (n + 1)^d.$$

Begründung: Seien $x_1, \dots, x_n \in \mathbb{R}^d$ fest.

Für $j \in \{1, \dots, d\}$ sei z_{1j}, \dots, z_{nj} Permutation von x_1, \dots, x_n mit

$$z_{1j}^{(j)} \leq z_{2j}^{(j)} \leq \dots \leq z_{nj}^{(j)}.$$

Wie oben gilt dann

$$\begin{aligned} & (\mathbb{R} \times \dots \times \mathbb{R} \times (-\infty, x^{(j)}] \times \mathbb{R} \times \dots \times \mathbb{R}) \cap \{x_1, \dots, x_n\} \\ & \in \{\emptyset, \{z_{1j}\}, \dots, \{z_{1j}, \dots, z_{nj}\}\} \end{aligned}$$

und mit

$$(-\infty, x] \cap \{x_1, \dots, x_n\} = \bigcap_{j=1}^d (\mathbb{R} \times \dots \times \mathbb{R} \times (-\infty, x^{(j)}] \times \mathbb{R} \times \dots \times \mathbb{R}) \cap \{x_1, \dots, x_n\}$$

folgt

$$(-\infty, x] \cap \{x_1, \dots, x_n\} \in \left\{ \left(\bigcap_{j_1=0}^d \{z_{1j_1}, z_{2j_1}, \dots, z_{nj_1}\} \right) \cap \dots \cap \left(\bigcap_{j_d=1}^d \{z_{1j_d}, z_{2j_d}, \dots, z_{nj_d}\} \right) \right\},$$

woraus folgt:

$$s(\mathcal{A}, n) \leq (n + 1)^d.$$

Das Hauptresultat dieses Abschnitts ist:

Satz 2.2

Seien X_1, X_2, \dots u.i.v. \mathbb{R}^d -wertige ZVen, $\mu = P_{X_1}$ und sei μ_n die empirische Verteilung zu X_1, \dots, X_n . Sei \mathcal{A} eine Klasse von Mengen $A \subseteq \mathbb{R}^d$.

Dann gilt für alle $n \in \mathbb{N}$ und alle $\varepsilon > 0$

$$P \left\{ \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| > \varepsilon \right\} \leq 8 \cdot s(\mathcal{A}, n) \cdot \exp \left(-\frac{n \cdot \varepsilon^2}{32} \right).$$

Korollar 2.3 (Vapnik und Chervonenkis (1971))

Unter den Voraussetzungen von Satz 2.2 gilt: Aus

$$(2.4) \quad \frac{\log s(\mathcal{A}, n)}{n} \rightarrow 0 \quad (n \rightarrow \infty)$$

folgt

$$\sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| \rightarrow 0 \quad \text{f.s.}$$

Satz 2.1 ergibt sich nun unmittelbar aus Korollar 2.3, da nach Beispiel 2.1 b) gilt:

$$\frac{\log s(\{(-\infty, x] : x \in \mathbb{R}^d\}, n)}{n} \leq \frac{\log((n+1)^d)}{n} \rightarrow 0 \quad (n \rightarrow \infty).$$

Beweis von Korollar 2.3

Setze

$$Z_n = \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)|.$$

Dann gilt für $\varepsilon > 0$:

$$\begin{aligned} \sum_{n=1}^{\infty} P\{|Z_n| > \varepsilon\} &\stackrel{\text{Satz 2.2}}{\leq} \sum_{n=1}^{\infty} 8 \cdot s(\mathcal{A}, n) \cdot \exp \left(-\frac{n \varepsilon^2}{32} \right) \\ &= \sum_{n=1}^{\infty} 8 \cdot \exp \left(\left(\frac{\log s(\mathcal{A}, n)}{n} - \frac{\varepsilon^2}{32} \right) \cdot n \right) \\ &< \infty, \end{aligned}$$

da nach (2.4) für n genügend groß gilt:

$$\frac{\log s(\mathcal{A}, n)}{n} - \frac{\varepsilon^2}{32} < -\frac{\varepsilon^2}{64}.$$

Daraus folgt

$$Z_n \rightarrow 0 \quad \text{f.s.}$$

Begründung:

Nach dem Lemma von Borel-Cantelli folgt aus $\sum_{n=1}^{\infty} P(|Z_n| > \varepsilon) < \infty$, dass gilt:

$$P[\overline{\lim}[|Z_n| > \varepsilon]] = 0.$$

Mit

$$\begin{aligned} \overline{\lim}[|Z_n| > \varepsilon] &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [|Z_k| > \varepsilon] \\ &= \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{\omega \in \Omega : |Z_k(\omega)| > \varepsilon\} \\ &\stackrel{(!)}{=} \{\omega \in \Omega : \overline{\lim}|Z_n(\omega)| > \varepsilon\} \end{aligned}$$

folgt daraus für beliebiges $k \in \mathbb{N}$ und mit $\varepsilon = \frac{1}{k}$:

Mit Wahrscheinlichkeit Eins gilt

$$\overline{\lim}_{n \rightarrow \infty} |Z_n| \leq \frac{1}{k}.$$

Also gilt mit Wahrscheinlichkeit Eins auch

$$\overline{\lim}_{n \rightarrow \infty} |Z_n| \leq 0 \quad \rightsquigarrow \text{Beh.} \quad \square$$

Beweis von Satz 2.2

OBdA $n \geq 8/\varepsilon^2$, da andernfalls linke Seite ≥ 1 .

Schritt 1: Symmetrisierung durch Einführung einer Geisterstichprobe.

Wir ersetzen

$$\mu(A) = \int_A 1 P_{X_1}(dx)$$

durch

$$\mu'_n(A) = \frac{1}{n} \sum_{i=1}^n I_A(X'_i),$$

wobei $X_1, \dots, X_1, X'_1, \dots, X'_n$ u. i. v.

Dazu setze

$$X_1^n = (X_1, \dots, X_n).$$

Wähle

$$A^* = A^*(X_1^n) \in \mathcal{A}$$

so, dass

$$|\mu_n(A^*) - \mu(A^*)| > \varepsilon,$$

falls eine solche Menge existiert; wähle $A^* \in \mathcal{A}$ beliebig, falls keine solche Menge existiert.

Gemäß der Ungleichung von Tschebyscheff gilt für jedes feste $A \in \mathcal{A}$

$$\begin{aligned} & P\{|\mu(A) - \mu'_n(A)| > \frac{\varepsilon}{2}\} \\ &= P\left\{\left|\frac{1}{n} \sum_{i=1}^n I_A(X'_i) - E\{I_A(X'_1)\}\right| > \frac{\varepsilon}{2}\right\} \\ &\leq \frac{V\left(\frac{1}{n} \sum_{i=1}^n I_A(X'_i)\right)}{\left(\frac{\varepsilon}{2}\right)^2} \\ &= \frac{\frac{1}{n^2} \sum_{i=1}^n V(I_A(X'_i))}{\frac{\varepsilon^2}{4}} \\ &= \frac{4}{n \varepsilon^2} \quad (\text{da } V(I_A(X_1)) \leq E(I_A(X_1)^2) \leq 1) \\ &\leq \frac{1}{2} \quad (\text{da nach Voraussetzung } n \geq \frac{8}{\varepsilon^2}), \end{aligned}$$

also gilt auch

$$P\left\{|\mu(A^*) - \mu'_n(A^*)| > \frac{\varepsilon}{2} \mid X_1^n\right\} \leq \frac{1}{2}.$$

Daraus folgt:

$$\begin{aligned} & P\left\{\sup_{A \in \mathcal{A}} |\mu_n(A) - \mu'_n(A)| > \frac{\varepsilon}{2}\right\} \\ &\geq P\{|\mu_n(A^*) - \mu'_n(A^*)| > \frac{\varepsilon}{2}\} \\ &\geq P\{|\mu_n(A^*) - \mu(A^*)| > \varepsilon, |\mu(A^*) - \mu'_n(A^*)| \leq \frac{\varepsilon}{2}\} \\ &= E\{P\{\dots \mid X_1^n\}\} \\ &\quad (\text{nach Definition der bedingten Wahrscheinlichkeit}) \\ &= E\left\{I_{\{|\mu_n(A^*) - \mu(A^*)| > \varepsilon\}} \cdot P\{|\mu(A^*) - \mu'_n(A^*)| \leq \frac{\varepsilon}{2} \mid X_1^n\}\right\} \\ &\quad (\text{da Indikatorfunktion (messbare) Funktion von } X_1^n \text{ ist}) \\ &\stackrel{s.o.}{\geq} E\left\{I_{\{|\mu_n(A^*) - \mu(A^*)| > \varepsilon\}} \cdot \frac{1}{2}\right\} \\ &= \frac{1}{2} \cdot P\{|\mu_n(A^*) - \mu(A^*)| > \varepsilon\} \\ &= \frac{1}{2} \cdot P\left\{\sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| > \varepsilon\right\} \\ &\quad (\text{nach Definition von } A^*). \end{aligned}$$

Also ist damit gezeigt:

$$\begin{aligned} & P \left\{ \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| > \varepsilon \right\} \\ & \leq 2 \cdot P \left\{ \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu'_n(A)| > \frac{\varepsilon}{2} \right\}. \end{aligned}$$

Schritt 2: Einführung zufälliger Vorzeichen.

Wähle Zufallsvariablen U_1, \dots, U_n mit

$$P\{U_i = 1\} = P\{U_i = -1\} = \frac{1}{2} \quad (i = 1, \dots, n)$$

und

$$X_1, \dots, X_n, X'_1, \dots, X'_n, U_1, \dots, U_n \quad \text{unabhängig.}$$

Die gemeinsame Verteilung von $(X_1, \dots, X_n, X'_1, \dots, X'_n)$ ändert sich nicht, wenn man Komponenten von (X_1, \dots, X_n) mit den entsprechenden Komponenten von (X'_1, \dots, X'_n) (zufällig (!)) vertauscht.

Daraus folgt:

$$\begin{aligned} & P \left\{ \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu'_n(A)| > \frac{\varepsilon}{2} \right\} \\ & = P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n (I_A(X_i) - I_A(X'_i)) \right| > \frac{\varepsilon}{2} \right\} \\ & \stackrel{(!)}{=} P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot (I_A(X_i) - I_A(X'_i)) \right| > \frac{\varepsilon}{2} \right\} \\ & \leq P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(X_i) \right| > \frac{\varepsilon}{4} \right\} \\ & \quad + P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(X'_i) \right| > \frac{\varepsilon}{4} \right\} \\ & = 2 \cdot P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(X_i) \right| > \frac{\varepsilon}{4} \right\}. \end{aligned}$$

Schritt 3: Festhalten der Werte der X_i 's.

Da (U_1, \dots, U_n) und (X_1, \dots, X_n) unabhängig sind, gilt nach dem Satz von Fubini:

$$\begin{aligned}
& P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(X_i) \right| > \frac{\varepsilon}{4} \right\} \\
&= \int P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(x_i) \right| > \frac{\varepsilon}{4} \right\} dP_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\
&\quad \text{(hier wird die Wk. als Integral bzgl. der gemeinsamen Verteilung von} \\
&\quad \text{(} U_1, \dots, U_n, X_1, \dots, X_n \text{) geschrieben, und dieses dann als iteriertes} \\
&\quad \text{Integral bzgl. } P_{(U_1, \dots, U_n)} \text{ und } P_{(X_1, \dots, X_n)} \text{ umgeschrieben).}
\end{aligned}$$

Für feste $x_1, \dots, x_n \in \mathbb{R}^d$ nimmt

$$(I_A(x_1), \dots, I_A(x_n)) \in \{0, 1\}^n \quad (*)$$

genau so viele verschiedene Werte an, wie es Mengen der Form

$$A \cap \{x_1, \dots, x_n\}$$

gibt. Daher nimmt $(*)$ höchstens $s(\mathcal{A}, n)$ verschiedene Werte an.

Also ist das obige Supremum in Wahrheit ein Maximum über $s(\mathcal{A}, n)$ verschiedene Zufallsvariablen, und mit

$$\begin{aligned}
& P \left\{ \max_{j=1, \dots, K} |Z_j| > \frac{\varepsilon}{4} \right\} \\
&\stackrel{\text{Def.}}{=} P \left(\{ \omega \in \Omega : \max_{j=1, \dots, K} |Z_j(\omega)| > \frac{\varepsilon}{4} \} \right) \\
&= P \left(\bigcup_{j=1}^K \{ \omega \in \Omega : |Z_j(\omega)| > \frac{\varepsilon}{4} \} \right) \\
&\leq \sum_{j=1}^K P \left(\{ \omega \in \Omega : |Z_j(\omega)| > \frac{\varepsilon}{4} \} \right) \\
&= \sum_{j=1}^K P \{ |Z_j| > \frac{\varepsilon}{4} \} \\
&\leq K \cdot \max_{j=1, \dots, K} P \{ |Z_j| > \frac{\varepsilon}{4} \}
\end{aligned}$$

folgt:

$$\begin{aligned}
& P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(x_i) \right| > \frac{\varepsilon}{4} \right\} \\
&\leq s(\mathcal{A}, n) \cdot \sup_{A \in \mathcal{A}} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(x_i) \right| > \frac{\varepsilon}{4} \right\}.
\end{aligned}$$

Schritt 4: Anwendung der Ungleichung von Hoeffding.

Die ZVen $U_1 \cdot I_A(x_1), \dots, U_n \cdot I_A(x_n)$ sind unabhängig und es gilt

$$E\{U_i \cdot I_A(x_i)\} = 0 \text{ und } -1 \leq U_i \cdot I_A(x_i) \leq 1.$$

Daher lässt sich das folgende Resultat anwenden:

Ungleichung von Hoeffding:

Sind Z_1, \dots, Z_n unabhängig mit $a_i \leq Z_i \leq b_i$ f.s. ($i = 1, \dots, n$), so gilt für jedes $\varepsilon > 0$:

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n (Z_i - EZ_i) \right| > \varepsilon \right\} \leq 2 \cdot \exp \left(- \frac{2n \varepsilon^2}{\frac{1}{n} \sum_{i=1}^n (b_i - a_i)^2} \right).$$

Damit folgt für beliebige $x_1, \dots, x_n \in \mathbb{R}^d$ und beliebiges $A \in \mathcal{A}$:

$$\begin{aligned} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(x_i) \right| > \frac{\varepsilon}{4} \right\} &\leq 2 \cdot \exp \left(- \frac{2n \left(\frac{\varepsilon}{4}\right)^2}{\frac{1}{n} \sum_{i=1}^n (1+1)^2} \right) \\ &= 2 \cdot \exp \left(- \frac{n\varepsilon^2}{32} \right). \end{aligned}$$

Die Behauptung folgt nun aus den Schritten 1 bis 4:

$$\begin{aligned} &P \left\{ \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu(A)| > \varepsilon \right\} \\ &\stackrel{\text{Schritt 1}}{\leq} 2 \cdot P \left\{ \sup_{A \in \mathcal{A}} |\mu_n(A) - \mu'_n(A)| > \frac{\varepsilon}{2} \right\} \\ &\stackrel{\text{Schritt 2}}{\leq} 4 \cdot P \left\{ \sup_{A \in \mathcal{A}} \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(X_i) \right| > \frac{\varepsilon}{4} \right\} \\ &\stackrel{\text{Schritt 3}}{\leq} 4 \cdot \int s(\mathcal{A}, n) \cdot \sup_{A \in \mathcal{A}} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n U_i \cdot I_A(x_i) \right| > \frac{\varepsilon}{4} \right\} dP_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &\stackrel{\text{Schritt 4}}{\leq} 4 \cdot s(\mathcal{A}, n) \cdot 2 \exp \left(- \frac{n\varepsilon^2}{32} \right). \quad \square \end{aligned}$$

3 Dichteschätzung

3.1 Motivation

X_1, \dots, X_n u. i. v. \mathbb{R}^d -wertige Zven, $\mu = P_{X_1}$. μ_n sei die empirische Verteilung zu X_1, \dots, X_n .

Nach Glivenko-Cantelli gilt:

$$\sup_{x \in \mathbb{R}^d} |\mu_n((-\infty, x]) - \mu((-\infty, x])| \rightarrow 0 \quad \text{f.s.}$$

für jede Verteilung μ auf $(\mathbb{R}^d, \mathcal{B}_d)$.

\rightsquigarrow Gute Vorhersage der Wahrscheinlichkeiten von Intervallen.

Frage: Auch gute Vorhersage von Wahrscheinlichkeiten beliebiger (messbarer) Mengen?

Antwort: Im allgemeinen leider nein, denn ist die Verteilungsfunktion F von μ stetig, so gilt $\mu(\{x\}) = 0$ für alle $x \in \mathbb{R}^d$, und daraus folgt:

$$\sup_{B \in \mathcal{B}_d} |\mu_n(B) - \mu(B)| \geq \underbrace{|\mu_n(\{X_1, \dots, X_n\})|}_{=1} - \underbrace{|\mu(\{X_1, \dots, X_n\})|}_{=0 \text{ (s.o.)}} \not\rightarrow 0 \quad \text{f.s.}$$

Man kann allgemeiner zeigen:

Satz 3.1. Es gibt keinen Schätzer

$$\hat{\mu}_n(\cdot) = \hat{\mu}_n(\cdot, X_1, \dots, X_n) : \mathcal{B}_d \rightarrow \mathbb{R}$$

mit

$$\sup_{B \in \mathcal{B}_d} |\hat{\mu}_n(B) - \mu(B)| \rightarrow 0 \quad \text{f.s.}$$

für alle Verteilungen μ auf $(\mathbb{R}^d, \mathcal{B}_d)$ und alle unabhängig identisch verteilten Zufallsvariablen X_1, X_2, \dots mit $\mathbf{P}_{X_1} = \mu$.

Beweis. oBdA $d=1$.

Wir zeigen:

Für jede Folge von Schätzfunktionen

$$\hat{\mu}_n(\cdot) = \hat{\mu}_n(\cdot, X_1, \dots, X_n) : \mathcal{B} \rightarrow \mathbb{R}$$

existiert eine Verteilung μ und unabhängig identisch verteilten Zufallsvariablen X_1, X_2, \dots mit $\mathbf{P}_{X_1} = \mu$ so, dass gilt:

$$\inf_{n \in \mathbb{N}} \sup_{A \in \mathcal{B}} |\hat{\mu}_n(A) - \mu(A)| \geq 0.45 \quad \text{f.s.}$$

Schritt 1: Wir definieren in Abhängigkeit eines Parameters

$$b = (b^{(j)})_{j \in \mathbb{N}} \in \{0, 1\}^{\mathbb{N}}$$

eine Verteilung μ_b und unabhängig identisch verteilten Zufallsvariablen X_1, X_2, \dots mit $\mathbf{P}_{X_1} = \mu_b$.

Dazu wählen wir unabhängige auf $\{0, 1, \dots, 9\}$ gleichverteilte Zufallsvariablen

$$Y^{(1)}, Y^{(2)}, \dots, Y_1^{(1)}, Y_1^{(2)}, \dots, Y_2^{(1)}, Y_2^{(2)}, \dots,$$

und setzen

$$Y = (Y^{(1)}, Y^{(2)}, \dots) \quad \text{und} \quad Y_j = (Y_j^{(1)}, Y_j^{(2)}, \dots) \quad (j \in \mathbb{N}),$$

und

$$X := X(Y, b) := \sum_{k=1}^{\infty} Y^{(k)} \cdot I_{\{b^{(k)}=1\}} \cdot \frac{1}{10^k}$$

sowie $X_1 = X(Y_1, b)$, $X_2 = X(Y_2, b)$, \dots

X ist also der zufällige Wert den man erhält, wenn man eine Zahl zwischen Null und Eins so erzeugt, dass man in ihrer Darstellung als Dezimalbruch alle Ziffern unabhängig voneinander zuerst rein zufällig wählt und dann alle die, an deren Position in b keine Eins steht, auf Null setzt. Enthält b genau L Nullen, so ist X gleichverteilt auf einer Menge vom LB-Maß $(1/2)^L$ (und damit stetig verteilt mit Dichte bzgl. des LB-Maßes). Enthält dagegen b genau L Einsen, so ist X gleichverteilt auf einer Menge der Kardinalität 2^L (und damit diskret verteilt). In allen anderen Fällen ist X weder stetig verteilt mit Dichte noch diskret verteilt (ohne Beweis).

Schritt 2: Wir verwenden die Schätzung

$$\hat{\mu}_n(\cdot) = \hat{\mu}_n(\cdot, X_1, \dots, X_n) : \mathcal{B} \rightarrow \mathbb{R}$$

von μ_b , um ausgehend von X_1, \dots, X_n die $b^{(1)}, b^{(2)}, \dots$ vorherzusagen.

Dazu setzen wir

$$A_k = \left\{ \sum_{j=1}^{\infty} \frac{x_j}{10^j} \in [0, 1] \quad : \quad x_i \in \{0, 1, \dots, 9\} \quad (i \in \mathbb{N}) \quad \text{und} \quad x_k = 0 \right\}$$

und beachten

$$\begin{aligned} \mu_b(A_k) &= \mathbf{P} \left[\sum_{j=1}^{\infty} Y^{(j)} \cdot I_{\{b^{(j)}=1\}} \cdot \frac{1}{10^j} \in A_k \right] \\ &= \mathbf{P} \left[Y^{(k)} \cdot I_{\{b^{(k)}=1\}} \cdot \frac{1}{10^k} = 0 \right] \\ &= \begin{cases} 1 & \text{falls } b^{(k)} = 0, \\ \frac{1}{10} & \text{falls } b^{(k)} = 1, \end{cases} \end{aligned}$$

wobei die zweite Gleichheit gilt da die dabei auftretenden Ereignisse mit Wahrscheinlichkeit Eins übereinstimmen.

Wir vergleichen nun die Vorhersage $\hat{\mu}_n(A_k)$ mit $\mu_b(A_k)$. Ist der vorhergesagte Wert näher an 1 als an $1/10$, so schätzen wir $b^{(k)}$ durch 0 und andernfalls durch 1. D.h., wir setzen

$$\hat{b}_{n,k} = \begin{cases} 0 & \text{falls } \hat{\mu}_n(A_k) > \frac{1+1/10}{2} = \frac{11}{20}, \\ \frac{1}{10} & \text{sonst.} \end{cases}$$

Dann gilt

$$|\hat{\mu}_n(A_k) - \mu_b(A_k)| \geq \frac{9}{20} \cdot I_{\{\hat{b}_{n,k} \neq b^{(k)}\}},$$

wie man sich wie folgt durch Fallunterscheidung klar macht:

Die obige Aussage ist trivial im Falle $\hat{b}_{n,k} = b^{(k)}$. Ist nun $\hat{b}_{n,k} = 1$ und $b^{(k)} = 0$, so führt zunächst $b^{(k)} = 0$ und dann $\hat{b}_{n,k} = 1$ auf

$$|\hat{\mu}_n(A_k) - \mu_b(A_k)| = |\hat{\mu}_n(A_k) - 1| \geq 1 - \frac{11}{20} = \frac{9}{20}.$$

Ist dagegen $\hat{b}_{n,k} = 0$ und $b^{(k)} = 1$, so führt zunächst $b^{(k)} = 1$ und dann $\hat{b}_{n,k} = 0$ auf

$$|\hat{\mu}_n(A_k) - \mu_b(A_k)| = \left| \hat{\mu}_n(A_k) - \frac{1}{10} \right| \geq \frac{11}{20} - \frac{1}{10} = \frac{9}{20}.$$

Damit erhalten wir insgesamt

$$\inf_{n \in \mathbb{N}} \sup_{A \in \mathcal{B}} |\hat{\mu}_n(A) - \mu_b(A)| \geq \inf_{n \in \mathbb{N}} \sup_{k \in \mathbb{N}} |\hat{\mu}_n(A_k) - \mu_b(A_k)| \geq \frac{9}{20} \cdot \inf_{n \in \mathbb{N}} \sup_{k \in \mathbb{N}} I_{\{\hat{b}_{n,k} \neq b^{(k)}\}}.$$

Schritt 3. Wir wählen den Wert von b als zufälligen Wert.

Dazu setzen wir

$$B = (B^{(1)}, B^{(2)}, \dots),$$

wobei $B^{(1)}, B^{(2)}, \dots$ unabhängige auf $\{0, 1\}$ gleichverteilte Zufallsvariablen sind, die auch unabhängig von allen $Y_i^{(j)}$ ($i, j \in \mathbb{N}$) sind. Diese zufällige Wahl von b führt auf

$$\inf_{n \in \mathbb{N}} \sup_{A \in \mathcal{B}} |\hat{\mu}_n(A) - \mu_B(A)| \geq \frac{9}{20} \cdot \inf_{n \in \mathbb{N}} \sup_{k \in \mathbb{N}} I_{\{\hat{b}_{n,k} \neq B^{(k)}\}} = \frac{9}{20} \cdot \inf_{n \in \mathbb{N}} Z_n$$

mit

$$Z_n = \sup_{k \in \mathbb{N}} I_{\{\hat{b}_{n,k} \neq B^{(k)}\}}.$$

Im Folgenden zeigen wir nun

$$Z_n = 1 \quad f.s. \quad \text{für alle } n \in \mathbb{N},$$

(was $Z_n = 1$ für alle $n \in \mathbb{N}$ f.s. impliziert), was den Beweis abschließt, denn dies zeigt, dass wir durch rein zufällige Wahl von b mit Wahrscheinlichkeit Eins einen Wert B erhalten mit

$$\inf_{n \in \mathbb{N}} \sup_{A \in \mathcal{B}} |\hat{\mu}_n(A) - \mu_B(A)| \geq \frac{9}{20} = 0.45,$$

was insbesondere die Existenz eines solchen Wertes nachweist.

Schritt 4: Abschluss des Beweises.

Wir zeigen für jedes $n \in \mathbb{N}$: $\mathbf{P}[Z_n = 1] = 1$.

Dazu beachten wir, dass aufgrund der Stetigkeit des W-Maßes von unten gilt:

$$\begin{aligned} \mathbf{P}[Z_n = 1] &= \mathbf{P} \left[\sup_{k \in \mathbb{N}} I_{\{\hat{b}_{n,k} \neq B^{(k)}\}} = 1 \right] \\ &= \mathbf{P} \left(\bigcup_{k \in \mathbb{N}} [\hat{b}_{n,k} \neq B^{(k)}] \right) \\ &= \lim_{N \rightarrow \infty} \mathbf{P} \left(\bigcup_{k=1}^N [\hat{b}_{n,k} \neq B^{(k)}] \right) \\ &= \lim_{N \rightarrow \infty} \mathbf{P} \left[(\hat{b}_{n,1}, \dots, \hat{b}_{n,N}) \neq (B^{(1)}, \dots, B^{(N)}) \right]. \end{aligned}$$

Im Folgenden leiten wir eine untere Schranke für die Wahrscheinlichkeit

$$\mathbf{P} \left[(\hat{b}_{n,1}, \dots, \hat{b}_{n,N}) \neq (B^{(1)}, \dots, B^{(N)}) \right]$$

her. Dazu beachten wir

$$\begin{aligned} &\mathbf{P} \left[(\hat{b}_{n,1}, \dots, \hat{b}_{n,N}) \neq (B^{(1)}, \dots, B^{(N)}) \right] \\ &= 1 - \mathbf{E} \left(\mathbf{P} \left[(\hat{b}_{n,1}, \dots, \hat{b}_{n,N}) = (B^{(1)}, \dots, B^{(N)}) \mid Y_i^{(k)} \cdot I_{\{B^{(k)}=1\}} \quad (1 \leq i \leq n, k \in \mathbb{N}) \right] \right) \end{aligned}$$

und die aus der Unabhängigkeit der $Y_i^{(j)}$, $B^{(j)}$ ($1 \leq i \leq n, j \in \mathbb{N}$) folgende Beziehung

$$\begin{aligned} &\mathbf{P} \left[(b_1, \dots, b_N) = (B^{(1)}, \dots, B^{(N)}) \mid Y_i^{(k)} \cdot I_{\{B^{(k)}=1\}} \quad (1 \leq i \leq n, k \in \mathbb{N}) \right] \\ &= \mathbf{P} \left[(b_1, \dots, b_N) = (B^{(1)}, \dots, B^{(N)}) \mid Y_i^{(k)} \cdot I_{\{B^{(k)}=1\}} \quad (1 \leq i \leq n, 1 \leq k \leq N) \right] \\ &= \prod_{k=1}^N \mathbf{P} \left[b_k = B^{(k)} \mid Y_i^{(k)} \cdot I_{\{B^{(k)}=1\}} \quad (1 \leq i \leq n) \right]. \end{aligned}$$

Da der letzte Ausdruck maximal wird, falls jeder einzelne Faktor maximal wird, können wir aus den obigen beiden Beziehungen folgern, dass gilt:

$$\mathbf{P} \left[(\hat{b}_{n,1}, \dots, \hat{b}_{n,N}) \neq (B^{(1)}, \dots, B^{(N)}) \right] \geq \mathbf{P} \left[(\bar{B}_{n,1}, \dots, \bar{B}_{n,N}) \neq (B^{(1)}, \dots, B^{(N)}) \right],$$

wobei

$$\bar{B}_{n,j} = \begin{cases} 1, & \text{falls } \mathbf{P} \left[B^{(j)} = 1 | Y_i^{(j)} \cdot I_{\{B^{(j)}=1\}} \quad (1 \leq i \leq n) \right] \geq \frac{1}{2}, \\ 0, & \text{sonst.} \end{cases}$$

Mit

$$\begin{aligned} & \mathbf{P} \left[B^{(j)} = 1 | Y_i^{(j)} \cdot I_{\{B^{(j)}=1\}} = y_i^{(j)} \cdot I_{\{b^{(j)}=1\}} \quad (1 \leq i \leq n) \right] \\ &= \frac{\mathbf{P} \left[B^{(j)} = 1, Y_i^{(j)} \cdot I_{\{B^{(j)}=1\}} = y_i^{(j)} \cdot I_{\{b^{(j)}=1\}} \quad (1 \leq i \leq n) \right]}{\mathbf{P} \left[Y_i^{(j)} \cdot I_{\{B^{(j)}=1\}} = y_i^{(j)} \cdot I_{\{b^{(j)}=1\}} \quad (1 \leq i \leq n) \right]} \\ &= \begin{cases} 1, & \text{falls } y_i^{(j)} \cdot I_{\{b^{(j)}=1\}} \neq 0 \quad \text{für ein } i \in \{1, \dots, n\}, \\ \frac{\frac{1}{2} \cdot \left(\frac{1}{10}\right)^n}{\frac{1}{2} + \frac{1}{2} \cdot \left(\frac{1}{10}\right)^n} = \frac{1}{10^n + 1}, & \text{falls } y_i^{(j)} \cdot I_{\{b^{(j)}=1\}} = 0 \quad \text{für alle } i \in \{1, \dots, n\} \end{cases} \end{aligned}$$

folgt

$$\bar{B}_{n,j} = \begin{cases} 1, & \text{falls } Y_i^{(j)} \cdot I_{\{B^{(j)}=1\}} \neq 0 \quad \text{für ein } i \in \{1, \dots, n\}, \\ 0, & \text{sonst.} \end{cases}$$

Erneute Anwendung der Unabhängigkeit der $Y_i^{(j)}, B^{(j)}$ ($1 \leq i \leq n, j \in \mathbb{N}$) liefert

$$\begin{aligned} \mathbf{P}[Z_n = 1] &\geq \lim_{N \rightarrow \infty} \mathbf{P} \left(\cup_{j=1}^N [\bar{B}_{n,j} \neq B^{(j)}] \right) \\ &= \lim_{N \rightarrow \infty} \left(1 - \mathbf{P} \left(\cap_{j=1}^N [\bar{B}_{n,j} = B^{(j)}] \right) \right) \\ &= \lim_{N \rightarrow \infty} \left(1 - \prod_{j=1}^N \mathbf{P} [\bar{B}_{n,j} = B^{(j)}] \right) \\ &= \lim_{N \rightarrow \infty} \left(1 - \prod_{j=1}^N \left(1 - \mathbf{P} [\bar{B}_{n,j} \neq B^{(j)}] \right) \right). \end{aligned}$$

Mit

$$\begin{aligned} & \mathbf{P} [\bar{B}_{n,j} \neq B^{(j)}] \\ &= \mathbf{P} [\bar{B}_{n,j} = 0, B^{(j)} = 1] + \mathbf{P} [\bar{B}_{n,j} = 1, B^{(j)} = 0] \\ &= \mathbf{P} [\bar{B}_{n,j} = 0, B^{(j)} = 1] + 0 \\ &= \mathbf{P} \left[Y_1^{(j)} = 0, \dots, Y_n^{(j)} = 0, B^{(j)} = 1 \right] \\ &= \left(\frac{1}{10} \right)^n \cdot \frac{1}{2} \end{aligned}$$

folgt

$$\mathbf{P}[Z_n = 1] \geq \lim_{N \rightarrow \infty} \left(1 - \left(1 - \frac{1}{2} \cdot \left(\frac{1}{10} \right)^n \right)^N \right) = 1,$$

w.z.z.w. □

Aber: Es gibt Schätzer $\hat{\mu}_n(\cdot) = \hat{\mu}(\cdot, X_1, \dots, X_n) : \mathcal{B}_d \rightarrow \mathbb{R}_+$ von μ mit

$$\sup_{B \in \mathcal{B}_d} |\hat{\mu}_n(B) - \mu(B)| \rightarrow 0 \quad \text{f.s.}$$

für alle Verteilungen μ auf $(\mathbb{R}^d, \mathcal{B}_d)$, die eine **Dichte** bzgl. des LB-Maßes besitzen, d. h. für die gilt:

$$\exists f : (\mathbb{R}^d, \mathcal{B}_d) \rightarrow (\mathbb{R}_+, \mathcal{B}_+) \text{ mit } \mu(B) = \int_B f(x) dx \quad (B \in \mathcal{B}_d).$$

Konstruktion solcher Schätzer mittelbar über Dichteschätzung möglich. Dies folgt aus:

Lemma 3.2 (Lemma von Scheffé)

Sind f, g Dichten auf $(\mathbb{R}^d, \mathcal{B}_d)$ (d. h. $f, g : (\mathbb{R}^d, \mathcal{B}_d) \rightarrow (\mathbb{R}_+, \mathcal{B})$ mit $\int_{\mathbb{R}^d} f(x) dx = 1 = \int_{\mathbb{R}^d} g(x) dx$), dann gilt:

$$\begin{aligned} \int |f(x) - g(x)| dx &= 2 \cdot \int (f(x) - g(x))_+ dx = 2 \cdot \int (g(x) - f(x))_+ dx \\ &= 2 \cdot \sup_{B \in \mathcal{B}_d} \left| \int_B f(x) dx - \int_B g(x) dx \right|, \end{aligned}$$

wobei

$$(y)_+ = \begin{cases} y & , \text{ falls } y \geq 0, \\ 0 & , \text{ sonst.} \end{cases}$$

Beweis:

Wegen

$$|f(x) - g(x)| = (f(x) - g(x))_+ + (g(x) - f(x))_+$$

gilt

$$\int |f(x) - g(x)| dx = \int (f(x) - g(x))_+ dx + \int (g(x) - f(x))_+ dx$$

Wegen

$$\begin{aligned} 0 &= \int f(x) dx - \int g(x) dx \\ &= \int (f(x) - g(x)) dx \\ &= \int (f(x) - g(x))_+ dx - \int (g(x) - f(x))_+ dx \\ &\quad (\text{da } f(x) - g(x) = (f(x) - g(x))_+ - (g(x) - f(x))_+) \end{aligned}$$

folgt

$$\int |f(x) - g(x)| dx = 2 \int (f(x) - g(x))_+ dx = 2 \int (f(x) - g(x))_- dx.$$

Darüber hinaus gilt

$$\begin{aligned} \int (f(x) - g(x))_+ dx &= \int_{\{t \in \mathbb{R}^d: f(t) \geq g(t)\}} (f(x) - g(x)) dx \\ &\leq \sup_{B \in \mathcal{B}_d} \left| \int_B f(x) dx - \int_B g(x) dx \right|, \end{aligned}$$

sowie für beliebiges $B \in \mathcal{B}_d$:

$$\begin{aligned} & \left| \int_B f(x) dx - \int_B g(x) dx \right| \\ &= \left| \underbrace{\int_{B \cap \{t: f(t) \geq g(t)\}} (f(x) - g(x)) dx}_{\geq 0} - \underbrace{\int_{B \cap \{t: f(t) < g(t)\}} (g(x) - f(x)) dx}_{\geq 0} \right| \\ &\leq \max \left\{ \int_{B \cap \{t: f(t) \geq g(t)\}} (f(x) - g(x)) dx, \int_{B \cap \{t: g(t) > f(t)\}} (g(x) - f(x)) dx \right\} \\ &\leq \max \left\{ \int_{\mathbb{R}^d} (f(x) - g(x))_+ dx, \int_{\mathbb{R}^d} (g(x) - f(x))_+ dx \right\} \\ &\stackrel{s.o.}{=} \int (f(x) - g(x))_+ dx \quad \rightsquigarrow \text{Beh.} \quad \square \end{aligned}$$

Folgerung:

Ist $f_n(\cdot) = f_n(\cdot, X_1, \dots, X_n)$ Folge von Schätzfunktionen mit

$$f_n(x) \geq 0 \quad (x \in \mathbb{R}^d) \quad \text{und} \quad \int_{\mathbb{R}^d} f_n(x) dx = 1$$

(d. h. f_n ist als Funktion von x eine Dichte) und

$$E \int |f_n(x) - f(x)| dx \rightarrow 0 \quad (n \rightarrow \infty),$$

so folgt für die Schätzung

$$\hat{\mu}_n(B) = \int_B \hat{f}_n(x) dx \quad (B \in \mathcal{B}_d)$$

von μ :

$$E \left(\sup_{B \in \mathcal{B}_d} |\hat{\mu}_n(B) - \mu(B)| \right) \rightarrow 0 \quad (n \rightarrow \infty),$$

wobei μ die zur Dichte f gehörende Verteilung ist.

Im Folgenden

Konstruktion von Dichteschätzern f_n mit

$$E \int |f_n(x) - f(x)| dx \rightarrow 0 \quad (n \rightarrow \infty)$$

für **jede** Dichte f .

3.2 Der Kerndichteschätzer

Zur Motivation des Dichteschätzers dient:

Lemma 3.3

Ist $f : (\mathbb{R}^d, \mathcal{B}_d) \rightarrow (\mathbb{R}, \mathcal{B})$ stetig in $x_0 \in \mathbb{R}$, so gilt

$$\frac{\int_{S_r(x_0)} f(x) dx}{\lambda(S_r(x_0))} \rightarrow f(x_0) \quad \text{für } r \rightarrow 0,$$

wobei

$$S_r(x_0) = \{x \in \mathbb{R}^d : \|x - x_0\| < r\}$$

die Kugel um x_0 mit Radius r ist, und λ das LB-Maß ist.

Beweis:

$$\begin{aligned} \left| \frac{\int_{S_r(x_0)} f(x) dx}{\lambda(S_r(x_0))} - f(x_0) \right| &= \left| \frac{\int_{S_r(x_0)} (f(x) - f(x_0)) dx}{\lambda(S_r(x_0))} \right| \\ &\leq \sup_{x: \|x - x_0\| < r} |f(x) - f(x_0)| \rightarrow 0 \quad \text{für } r \rightarrow 0, \text{ da } f \text{ stetig in } x_0. \quad \square \end{aligned}$$

Allgemeiner gilt:

Lemma 3.4 (Dichtetheorem von Lebesgue)

Ist $f : (\mathbb{R}^d, \mathcal{B}_d) \rightarrow (\mathbb{R}, \mathcal{B})$ eine Dichte (d. h. $f(x) \geq 0$ ($x \in \mathbb{R}^d$) und $\int_{\mathbb{R}^d} f(x) dx = 1$),

so gilt für λ -f.a. $x \in \mathbb{R}^d$:

$$\lim_{r \rightarrow 0} \frac{\int_{S_r(x)} f(u) du}{\lambda(S_r(x))} = f(x).$$

Beweis: Wird in der Maßtheorie behandelt. □

Für r klein ist also

$$\frac{\int_{S_r(x)} f(u) du}{\lambda(S_r(x))} = \frac{\mu(S_r(x))}{\lambda(S_r(x))}$$

(wobei f Dichte von μ ist) nahe bei $f(x)$.

Ausgehend von X_1, \dots, X_n (u. i. v. ZVen mit Verteilung μ und Dichte f) kann

$$\frac{\mu(S_r(x))}{\lambda(S_r(x))}$$

geschätzt werden durch

$$\begin{aligned} \frac{\mu_n(S_r(x))}{\lambda(S_r(x))} &= \frac{1}{\lambda(S_r(x))} \cdot \frac{1}{n} \sum_{i=1}^n I_{S_r(x)}(X_i) \\ &= \frac{1}{n \cdot r^d} \cdot \sum_{i=1}^n \frac{1}{\lambda(S_1(0))} \cdot I_{S_1(0)}\left(\frac{x-X_i}{r}\right). \end{aligned}$$

Hierbei gilt die letzte Gleichheit wegen

$$\lambda(S_r(x)) = r^d \cdot \lambda(S_1(0))$$

und

$$\begin{aligned} I_{S_r(x)}(X_i) = 1 &\Leftrightarrow X_i \in S_r(x) \Leftrightarrow \frac{x-X_i}{r} \in S_1(0) \\ &\Leftrightarrow I_{S_1(0)}\left(\frac{x-X_i}{r}\right) = 1. \end{aligned}$$

Dies führt auf den sogenannten **Kerndichteschätzer** (Rosenblatt (1956), Parzen (1962)):

$$f_n(x) = \frac{1}{n \cdot h_n^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)$$

mit

- **Kernfunktion** $K : \mathbb{R}^d \rightarrow \mathbb{R}$ integrierbar mit $\int K(x) dx = 1$ (oft wird K als Dichte bzgl des LB-Maßes gewählt)
- **Bandbreite** $h_n > 0$ (Parameter, der die "Glattheit" der Schätzung steuert).

Für $K = \frac{1}{\lambda(S_1(0))} \cdot I_{S_1(0)}$ (sog. **naiver Kern**) ergibt sich der obige Schätzer.

Für glatteres K , z. B.

- **Epanechnikov-Kern:** $K(u) = \text{const} \cdot (1 - \|u\|^2)_+$

- **Gauss-Kern:** $K(u) = \frac{1}{(2\pi)^{d/2}} \cdot e^{-\|u\|^2/2}$

ist die Schätzung glatter.

Der Kerndichteschätzer lässt sich deuten als Mittel von n um die Datenpunkte X_1, \dots, X_n zentrierte Dichten.

3.3 Ein Konsistenzresultat

Satz 3.5 (Schwache universelle Konsistenz des Kerndichteschätzers) X_1, \dots, X_n seien u. i. v. \mathbb{R}^d -wertige ZV mit Dichte f bzgl. des LB-Maßes. f_n sei der Kerndichteschätzer

$$f_n(x) = \frac{1}{n \cdot h_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$$

mit naivem Kern

$$K(u) = \frac{I_{S_1(0)}(u)}{\lambda(S_1(0))}$$

und Bandbreite $h_n > 0$.

Dann gilt: Aus

$$h_n \rightarrow 0 \ (n \rightarrow \infty) \text{ und } n \cdot h_n^d \rightarrow \infty \ (n \rightarrow \infty)$$

folgt

$$E\left\{\int |f_n(x) - f(x)| dx\right\} \rightarrow 0 \ (n \rightarrow \infty)$$

für **jede** Dichte f .

Beweis:

Wir zeigen zunächst, dass für λ -f.a. $x \in \mathbb{R}^d$ gilt:

$$E(|f_n(x) - f(x)|^2) \rightarrow 0 \ (n \rightarrow \infty). \quad (*)$$

Dazu beachten wir

$$\begin{aligned} E(|f_n(x) - f(x)|^2) &= E\{((f_n(x) - Ef_n(x)) + (Ef_n(x) - f(x)))^2\} \\ &= E\{(f_n(x) - Ef_n(x))^2\} + (Ef_n(x) - f(x))^2, \end{aligned}$$

da gilt

$$\begin{aligned} & E((f_n(x) - Ef_n(x)) \cdot (Ef_n(x) - f(x))) \\ &= (Ef_n(x) - f(x)) \cdot E\{f_n(x) - Ef_n(x)\} \\ &= (Ef_n(x) - f(x)) \cdot \underbrace{(Ef_n(x) - Ef_n(x))}_{=0} \\ &= 0. \end{aligned}$$

Hierbei ist

$$E\{(f_n(x) - Ef_n(x))^2\} = V(f_n(x))$$

der Varianzteil des Fehlers, während

$$Ef_n(x) - f(x)$$

als Bias (auf deutsch: Verzerrung) bezeichnet wird.

Unter Beachtung von

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n I_{S_{h_n}(x)}(X_i) / \lambda(S_{h_n}(x))$$

(vgl. Herleitung des Kerndichteschätzers für den naiven Kern) lassen sich Bias und Varianz einfach abschätzen.

Nach dem Lebesgueschen Dichtetheorem (Lemma 3.4) gilt wegen $h_n \rightarrow 0$ ($n \rightarrow \infty$) für λ -f.a. $x \in \mathbb{R}^d$

$$\begin{aligned} Ef_n(x) &= \frac{EI_{S_{h_n}(x)}(X_1)}{\lambda(S_{h_n}(x))} = \frac{\int_{S_{h_n}(x)} f(u) du}{\lambda(S_{h_n}(x))} \\ &\rightarrow f(x) \quad (n \rightarrow \infty), \end{aligned}$$

und wegen $n \cdot h_n^d \rightarrow \infty$ ($n \rightarrow \infty$) gilt darüber hinaus

$$\begin{aligned}
V(f_n(x)) &= V\left(\frac{1}{n} \sum_{i=1}^n I_{S_{h_n}(x)}(X_i) / \lambda(S_{h_n}(x))\right) \\
&= \frac{1}{n^2} \cdot \frac{1}{\lambda(S_{h_n}(x))^2} \cdot \sum_{i=1}^n V(I_{S_{h_n}(x)}(X_i)) \\
&\quad \text{(Rechenregeln für die Varianz und Unabhängigkeit} \\
&\quad \text{der } X_1, \dots, X_n) \\
&= \frac{1}{n \cdot \lambda(S_{h_n}(x))^2} V(I_{S_{h_n}(x)}(X_1)) \\
&\quad \text{(Identische Verteiltheit der } X_1, \dots, X_n) \\
&= \frac{1}{n \cdot \lambda(S_{h_n}(x))^2} E I_{S_{h_n}(x)}(X_1) \cdot (1 - E I_{S_{h_n}(x)}(X_1)) \\
&\quad \text{(da Varianz einer } b(1, p) \text{ - verteilten ZV gleich } p \cdot (1 - p) \text{ ist)} \\
&= \frac{1}{n \cdot h_n^d} \cdot \frac{1}{\lambda(S_1(0))} \cdot \frac{1}{\lambda(S_{h_n}(x))} \cdot E\{I_{S_{h_n}(x)}(X_1)\} \cdot (1 - E\{I_{S_{h_n}(x)}(X_1)\}) \\
&\rightarrow 0 \quad (n \rightarrow \infty) \text{ für } \lambda\text{-f.a. } x \in \mathbb{R}^d,
\end{aligned}$$

denn

- $\frac{1}{n \cdot h_n^d} \rightarrow 0 \quad (n \rightarrow \infty)$,
- $\frac{1}{\lambda(S_{h_n}(x))} \cdot E\{I_{S_{h_n}(x)}(X_1)\} \rightarrow f(x) \quad (n \rightarrow \infty)$,
für λ -f.a. $x \in \mathbb{R}^d$, und wegen $\int |f(x)| dx = 1$ ist $|f(x)| < \infty$ für L -f.a. $x \in \mathbb{R}^d$
- $1 - E\{I_{S_{h_n}(x)}(X_1)\}$ ist betragsmäßig durch 1 beschränkt.

Damit ist die Zwischenbehauptung (*) bewiesen.

Aus dieser folgt nun

$$f_n(x) \xrightarrow{P} f(x) \quad \text{für } \lambda\text{-f.a. } x,$$

was wiederum gemäß dem Satz von der majorisierten Konvergenz impliziert

$$E\{(f(x) - f_n(x))_+\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (**)$$

für λ -f.a. x .

Anwendung von Fubini und dem Lemma von Scheffe ergibt

$$\begin{aligned}
E \int |f_n(x) - f(x)| dx &= 2E \int (f(x) - f_n(x))_+ dx \\
&= 2 \cdot \int E\{(f(x) - f_n(x))_+\} dx \\
&\rightarrow 0 \quad (n \rightarrow \infty),
\end{aligned}$$

wobei die letzte Konvergenz (unter Beachtung von

$$E\{(f(x) - f_n(x))_+\} \leq f(x) \text{ und } \int_{\mathbb{R}^d} f(u) du = 1)$$

mit (**)) aus dem Satz von der majorisierten Konvergenz folgt. \square

Bemerkung: Satz 3.5 gilt wesentlich allgemeiner (siehe z. B. Devroye, Györfi (1985). Nonparametric density estimation: The L_1 view.)

Z. B. genügt es zu fordern, dass der Kern eine beschränkte Dichte mit kompaktem Support ist. In diesem Fall gilt sogar:

$$h_n \rightarrow 0 \text{ und } n \cdot h_n^d \rightarrow \infty (n \rightarrow \infty) \Leftrightarrow \int |f_n(x) - f(x)| dx \rightarrow 0 \text{ f.s. für jede Dichte } f.$$

Die Aussage von Satz 3.5 lässt sich verschärfen:

Satz 3.6. (*Starke universelle Konsistenz des Kerndichteschätzers*)

Unter den Voraussetzungen von Satz 3.5 gilt sogar:

$$\int_{\mathbb{R}^d} |f_n(x) - f(x)| dx \rightarrow 0 \quad \text{f.s.}$$

für **jede** Dichte $f : \mathbb{R}^d \rightarrow \mathbb{R}$.

Im Beweis verwenden wir:

Satz 3.7 (*Ungleichung von McDiarmid*).

Seien Z_1, \dots, Z_n unabhängige reelle Zufallsvariablen mit Werten in einer Menge $A \subseteq \mathbb{R}$ (bzw. genauer: $A \in \mathcal{B}$). Sei

$$f : A^n \rightarrow \mathbb{R}$$

eine (messbare) Funktion mit der Eigenschaft, dass für jedes $i \in \{1, \dots, n\}$ ein $c_i \in \mathbb{R}_+$ existiert mit

$$\sup_{z_1, \dots, z_n, z'_i} |f(z_1, \dots, z_n) - f(z_1, \dots, z_{i-1}, z'_i, z_{i+1}, \dots, z_n)| \leq c_i.$$

Dann gilt für jedes $\epsilon > 0$:

$$\mathbf{P} [f(Z_1, \dots, Z_n) - \mathbf{E} \{f(Z_1, \dots, Z_n)\} \geq \epsilon] \leq e^{-\frac{2 \cdot \epsilon^2}{\sum_{i=1}^n c_i^2}}$$

und

$$\mathbf{P} [\mathbf{E} \{f(Z_1, \dots, Z_n)\} - f(Z_1, \dots, Z_n) \geq \epsilon] \leq e^{-\frac{2 \cdot \epsilon^2}{\sum_{i=1}^n c_i^2}}.$$

Beweis von Satz 3.6:

Aufgrund von Satz 3.5 genügt es zu zeigen:

$$\int_{\mathbb{R}^d} |f_n(x) - f(x)| dx - \mathbf{E} \left\{ \int_{\mathbb{R}^d} |f_n(x) - f(x)| dx \right\} \rightarrow 0 \quad f.s.$$

(da wegen des Lemmas von Scheffe aus

$$\int_{\mathbb{R}^d} |f_n(x) - f(x)| dx \xrightarrow{\mathbf{P}} 0$$

mit dem Satz von der majorisierten Konvergenz auch

$$\mathbf{E} \left\{ \int_{\mathbb{R}^d} |f_n(x) - f(x)| dx \right\} \rightarrow 0 \quad (n \rightarrow \infty)$$

folgt).

Setze

$$f_n(x, x_1, \dots, x_n) = \frac{1}{n \cdot h_n^d} \cdot \sum_{i=1}^n K \left(\frac{x - x_i}{h_n} \right)$$

und

$$g(x_1, \dots, x_n) = \int_{\mathbb{R}^d} |f_n(x, x_1, \dots, x_n) - f(x)| dx.$$

Dann gilt

$$\int_{\mathbb{R}^d} |f_n(x) - f(x)| dx = g(X_1, \dots, X_n),$$

also ist zu zeigen:

$$g(X_1, \dots, X_n) - \mathbf{E} \{g(X_1, \dots, X_n)\} \rightarrow 0 \quad f.s.$$

Für $i \in \{1, \dots, n\}$, $x_1, \dots, x_n, x'_i \in \mathbb{R}$ gilt gemäß der zweiten Dreiecksungleichung

$$\begin{aligned}
& |g(x_1, \dots, x_n) - g(x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| \\
&= \left| \int_{\mathbb{R}^d} (|f_n(x, x_1, \dots, x_n) - f(x)| - |f_n(x, x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n) - f(x)|) dx \right| \\
&\leq \int_{\mathbb{R}^d} \left| |f_n(x, x_1, \dots, x_n) - f(x)| - |f_n(x, x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n) - f(x)| \right| dx \\
&\leq \int_{\mathbb{R}^d} |f_n(x, x_1, \dots, x_n) - f_n(x, x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n)| dx \\
&= \int_{\mathbb{R}^d} \left| \frac{1}{n \cdot h_n^d} \cdot K\left(\frac{x - x_i}{h_n}\right) - \frac{1}{n \cdot h_n^d} \cdot K\left(\frac{x - x'_i}{h_n}\right) \right| dx \\
&\leq \int_{\mathbb{R}^d} \frac{1}{n \cdot h_n^d} \cdot K\left(\frac{x - x_i}{h_n}\right) dx + \int_{\mathbb{R}^d} \frac{1}{n \cdot h_n^d} \cdot K\left(\frac{x - x'_i}{h_n}\right) dx \\
&= \frac{2}{n},
\end{aligned}$$

wobei wir bei der letzten Gleichheit benutzt haben, dass K eine Dichte ist.

Mit der Ungleichung von McDiarmid (Satz 3.7) folgt

$$\begin{aligned}
& \mathbf{P} \{|g(X_1, \dots, X_n) - \mathbf{E}\{g(X_1, \dots, X_n)\}| \geq \epsilon\} \\
&\leq \mathbf{P} \{g(X_1, \dots, X_n) - \mathbf{E}\{g(X_1, \dots, X_n)\} \geq \epsilon\} \\
&\quad + \mathbf{P} \{\mathbf{E}\{g(X_1, \dots, X_n)\} - g(X_1, \dots, X_n) \geq \epsilon\} \\
&\leq 2 \cdot \exp\left(-\frac{2 \cdot \epsilon^2}{\sum_{i=1}^n \frac{4}{n^2}}\right) = 2 \cdot \exp\left(-\frac{n \cdot \epsilon^2}{2}\right).
\end{aligned}$$

Daher gilt für jedes $\epsilon > 0$:

$$\sum_{n=1}^{\infty} \mathbf{P} \{|g(X_1, \dots, X_n) - \mathbf{E}\{g(X_1, \dots, X_n)\}| \geq \epsilon\} \leq \sum_{n=1}^{\infty} 2 \cdot \exp\left(-\frac{n \cdot \epsilon^2}{2}\right) < \infty,$$

was die Behauptung impliziert. \square

Zum Beweis von Satz 3.7 benötigen wir:

Lemma 3.8. Sei $(\Omega, \mathcal{A}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, sei Z eine \mathbb{R}^d -wertige und V eine (integrierbare) reelle Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbf{P})$, und sei $h : \mathbb{R}^d \rightarrow \mathbb{R}$ eine beschränkte (und messbare) Funktion. Es gelte

$$\mathbf{E}\{V|Z\} = 0$$

sowie für ein $c > 0$

$$h(Z) \leq V \leq h(Z) + c.$$

Dann gilt für alle $s > 0$:

$$\mathbf{E} \{e^{s \cdot V} | Z\} \leq e^{\frac{s^2 \cdot e^2}{8}} \quad f.s.$$

Beweis. Da $x \mapsto e^{s \cdot x}$ konvex ist, gilt für $a \leq x \leq b$:

$$e^{s \cdot x} = e^{\frac{x-a}{b-a} \cdot s \cdot b + \frac{b-x}{b-a} \cdot s \cdot a} \leq \frac{x-a}{b-a} \cdot e^{s \cdot b} + \frac{b-x}{b-a} \cdot e^{s \cdot a}.$$

Mit $x = V$, $a = h(Z)$ und $b = h(Z) + c$ folgt

$$e^{s \cdot V} \leq \frac{V - h(Z)}{c} \cdot e^{s \cdot c + s \cdot h(Z)} + \frac{h(Z) + c - V}{c} \cdot e^{s \cdot h(Z)}.$$

Unter Beachtung von $\mathbf{E}\{V|Z\} = 0$ erhalten wir daraus

$$\mathbf{E} \{e^{s \cdot V} | Z\} \leq -\frac{h(Z)}{c} \cdot e^{s \cdot c + s \cdot h(Z)} + \left(\frac{h(Z)}{c} + 1\right) \cdot e^{s \cdot h(Z)} \quad f.s.$$

bzw. mit $p = -h(Z)/c$

$$\begin{aligned} \mathbf{E} \{e^{s \cdot V} | Z\} &\leq p \cdot e^{s \cdot c - s \cdot p \cdot c} + (1 - p) \cdot e^{-s \cdot p \cdot c} \\ &= (1 - p + p \cdot e^{s \cdot c}) \cdot e^{-s \cdot p \cdot c} =: e^{\Phi(u)} \quad f.s., \end{aligned}$$

wobei

$$u = s \cdot c \quad \text{und} \quad \Phi(u) = -p \cdot u + \log(1 - p + p \cdot e^u).$$

Hierbei gilt $p \geq 0$ f.s., da nach Voraussetzung gilt

$$h(Z) = \mathbf{E}\{h(Z)|Z\} \leq \mathbf{E}\{V|Z\} = 0 \quad f.s.,$$

so dass der Logarithmus oben f.s. wohldefiniert ist.

Wegen

$$\Phi'(u) = -p + \frac{p \cdot e^u}{1 - p + p \cdot e^u}$$

und

$$\begin{aligned} \Phi''(u) &= p \cdot e^u \cdot (-1) \cdot (1 - p + p \cdot e^u)^{-2} \cdot p \cdot e^u + \frac{p \cdot e^u}{1 - p + p \cdot e^u} \\ &= \frac{-p^2 \cdot e^{2 \cdot u} + (1 - p) \cdot p \cdot e^u + p^2 \cdot e^{2 \cdot u}}{(1 - p + p \cdot e^u)^2} \\ &= \frac{(1 - p) \cdot p \cdot e^u}{(1 - p + p \cdot e^u)^2} \leq \frac{1}{4} \end{aligned}$$

(wobei die letzte Ungleichung aus $(a + b)^2 \geq 4 \cdot a \cdot b$ folgt) gilt

$$\Phi(0) = \Phi'(0) = 0 \quad \text{und} \quad \Phi''(v) \leq \frac{1}{4} \quad \text{für alle } v \in \mathbb{R}.$$

Mit Hilfe einer Taylorentwicklung erhalten wir daraus

$$\Phi(u) = \Phi(0) + \Phi'(0) \cdot u + \frac{1}{2} \cdot \Phi''(\xi) \cdot u^2 \leq \frac{u^2}{8} = \frac{s^2 c^2}{8},$$

was die Behauptung impliziert:

$$\mathbf{E} \{ e^{s \cdot V} | Z \} \leq e^{\Phi(u)} \leq e^{\frac{s^2 \cdot c^2}{8}} \quad f.s.$$

□

Beweis von Satz 3.7. Aus der Voraussetzung, dass sich der Wert von f nur um höchstens eine Konstante ändert, wenn man eines der Argumente abändert, folgt, dass f beschränkt ist.

Setze

$$V = f(Z_1, \dots, Z_n) - \mathbf{E} \{ f(Z_1, \dots, Z_n) \}$$

und

$$V_i = \mathbf{E} \{ f(Z_1, \dots, Z_n) | Z_1, \dots, Z_i \} - \mathbf{E} \{ f(Z_1, \dots, Z_n) | Z_1, \dots, Z_{i-1} \}$$

für $i \in \{1, \dots, n\}$. Dann gilt

$$V = \sum_{i=1}^n V_i,$$

was für beliebiges $s > 0$ aufgrund der Monotonie von $x \mapsto e^{s \cdot x}$ und der Ungleichung von Markov impliziert

$$\begin{aligned} & \mathbf{P} \{ \mathbf{E} \{ f(Z_1, \dots, Z_n) \} - f(Z_1, \dots, Z_n) \geq \epsilon \} \\ &= \mathbf{P} \left\{ \sum_{i=1}^n V_i \geq \epsilon \right\} \\ &\leq \mathbf{P} \left\{ \exp \left(s \cdot \sum_{i=1}^n V_i \right) \geq e^{s \cdot \epsilon} \right\} \\ &\leq e^{-s \cdot \epsilon} \cdot \mathbf{E} \left\{ \exp \left(s \cdot \sum_{i=1}^n V_i \right) \right\} \\ &= e^{-s \cdot \epsilon} \cdot \mathbf{E} \left\{ \exp \left(s \cdot \sum_{i=1}^{n-1} V_i \right) \cdot \mathbf{E} \{ \exp (s \cdot V_n) | Z_1, \dots, Z_{n-1} \} \right\}. \end{aligned}$$

Offensichtlich gilt

$$\begin{aligned} & \mathbf{E} \{V_n | Z_1, \dots, Z_{n-1}\} \\ &= \mathbf{E} \{ \mathbf{E} \{f(Z_1, \dots, Z_n) | Z_1, \dots, Z_n\} - \mathbf{E} \{f(Z_1, \dots, Z_n) | Z_1, \dots, Z_{n-1}\} | Z_1, \dots, Z_{n-1} \} \\ &= 0 \quad f.s. \end{aligned}$$

Weiter gilt für

$$\begin{aligned} h(z_1, \dots, z_{n-1}) &= \inf_{z \in \mathbb{R}} f(z_1, \dots, z_{n-1}, z) - \mathbf{E} \{f(Z_1, \dots, Z_n) | Z_1 = z_1, \dots, Z_{n-1} = z_{n-1}\} \\ &= \inf_{z \in \mathbb{R}} f(z_1, \dots, z_{n-1}, z) - \mathbf{E} \{f(z_1, \dots, z_{n-1}, Z_n)\} \end{aligned}$$

(hier ignorieren wir evt. Messbarkeitsprobleme (!)) aufgrund der Voraussetzung an f :

$$h(Z_1, \dots, Z_{n-1}) \leq V_n \leq h(Z_1, \dots, Z_{n-1}) + c_n.$$

Mit Lemma 3.8 folgt

$$\mathbf{E} \{ \exp(s \cdot V_n) | Z_1, \dots, Z_{n-1} \} \leq e^{\frac{s^2 \cdot c_n^2}{8}},$$

was impliziert:

$$\begin{aligned} & \mathbf{P} \{ \mathbf{E} \{f(Z_1, \dots, Z_n)\} - f(Z_1, \dots, Z_n) \geq \epsilon \} \\ & \leq e^{-s \cdot \epsilon} \cdot e^{\frac{s^2 \cdot c_n^2}{8}} \cdot \mathbf{E} \left\{ \exp \left(s \cdot \sum_{i=1}^{n-1} V_i \right) \right\} \\ & \leq e^{-s \cdot \epsilon} \cdot e^{\frac{s^2 \cdot c_n^2}{8}} \cdot \mathbf{E} \left\{ \exp \left(s \cdot \sum_{i=1}^{n-2} V_i \right) \cdot \mathbf{E} \{ \exp(s \cdot V_{n-1}) | Z_1, \dots, Z_{n-2} \} \right\}. \end{aligned}$$

Wiederholtes analoges Vorgehen wie oben liefert

$$\begin{aligned} \mathbf{P} \{ \mathbf{E} \{f(Z_1, \dots, Z_n)\} - f(Z_1, \dots, Z_n) \geq \epsilon \} & \leq e^{-s \cdot \epsilon} \cdot \prod_{i=1}^n e^{\frac{s^2 \cdot c_i^2}{8}} \\ & = \exp \left(-s \cdot \epsilon + s^2 \cdot \frac{\sum_{i=1}^n c_i^2}{8} \right). \end{aligned}$$

Mit

$$s = \frac{4 \cdot \epsilon}{\sum_{i=1}^n c_i^2}$$

folgt der erste Teil der Behauptung:

$$\mathbf{P} [f(Z_1, \dots, Z_n) - \mathbf{E} \{f(Z_1, \dots, Z_n)\} \geq \epsilon] \leq e^{-\frac{2 \cdot \epsilon^2}{\sum_{i=1}^n c_i^2}}.$$

Durch Vertauschung von f mit $(-1) \cdot f$ erhält man daraus auch den zweiten Teil der Behauptung. \square

4 Punktschätzungen

4.1 Problemstellungen und Beispiele

gegeben:

Realisierungen x_1, \dots, x_n von u. i. v. Zufallsvariablen X_1, \dots, X_n mit Werten in Menge \mathcal{X} .

P_{X_1} sei unbekannt.

Klasse $\{w_\vartheta : \vartheta \in \Theta\}$ von Verteilungen sei bekannt mit

$$P_{X_1} \in \{w_\vartheta : \vartheta \in \Theta\}.$$

Funktion $g : \Theta \rightarrow \mathbb{R}^k$ sei gegeben.

gesucht:

Schätzfunktion $T_n : \mathcal{X}^n \rightarrow \mathbb{R}^k$, mit der man, ausgehend von x_1, \dots, x_n , den unbekanntem Wert $g(\vartheta)$ durch

$$T_n(x_1, \dots, x_n)$$

schätzen kann.

Beispiel 4.1: $\Theta = \mathbb{R} \times \mathbb{R}_+$

$w_{(\mu, \sigma^2)}$ = Normalverteilung mit Erwartungswert μ und Varianz σ^2 .

$g : \Theta \rightarrow \mathbb{R}$, $g((\mu, \sigma^2)) = \sigma^2$.

Hier soll, ausgehend von einer Stichprobe einer Normalverteilung mit unbekanntem Erwartungswert und unbekannter Varianz, die Varianz geschätzt werden. Mögliche Schätzfunktion:

$$T_n(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ mit } \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j.$$

Beispiel 4.2 Die folgenden Daten beschreiben die Anzahl der Toten durch Hufschlag in 10 verschiedenen preußischen Kavallerieregimentern pro Regiment und Jahr, beobachtet über einen Zeitraum von 20 Jahren (insgesamt $n = 10 \cdot 20 = 200$ Datenpunkte)

Anzahl Toter im Jahr	0	1	2	3	4	≥ 5
Anzahl Regimenter	109	65	22	3	1	0

Die Zufallsvariable X_i ($i = 1, \dots, n$ mit $n = 200$) beschreibe die (zufällige) Anzahl der Toten durch Hufschlag in einem der 10 verschiedenen Regimentern und in einem der 20 verschiedenen Jahre.

Die Tabelle oben beschreibt dann

$$\#\{1 \leq i \leq n : x_i = j\}$$

für $j = 0, 1, 2, \dots$ für eine konkrete Realisierung der X_1, \dots, X_{200} .

Zur stochastischen Modellierung machen wir die folgenden Annahmen:

- X_1, \dots, X_n seien unabhängig und identisch verteilt.
- $P[X_1 = k] = \binom{m}{k} \cdot p^k \cdot (1 - p)^{m-k} \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda}$
mit $\lambda \approx m \cdot p$. Hier wird davon ausgegangen, dass das Regiment aus m Soldaten besteht, von denen jeder unbeeinflusst von den anderen mit Wahrscheinlichkeit p durch Hufschlag stirbt. Anschließend wird die Zähldichte der Binomialverteilung durch die einer Poisson-Verteilung approximiert.

Damit

$$\Theta = \mathbb{R}_+,$$

$$w_\vartheta = \text{Poisson-Verteilung mit Parameter } \vartheta,$$

$$g : \mathbb{R}_+ \rightarrow \mathbb{R} \text{ definiert durch } g(\vartheta) = \vartheta.$$

Die folgenden Daten beschreiben die Anzahl männlicher Krebstoter in verschiedenen Altersstufen in Hamburg im Jahr 1974:

Beispiel 4.3: Krebstote in Hamburg

Altersstufe	Anzahl männlicher Einwohner	Anzahl männlicher Krebstoter
[30,35)	76612	13
[35,40)	72615	21
[40,45)	50878	33
[45,50)	49059	65
[50,55)	45316	111
[55,60)	31274	118
[60,65)	47637	318
[65,70)	45792	514

Die (zufällige) Anzahl X_i der Krebstoten in Altersstufe $i \in \{1, 2, \dots, 8\}$ hängt von der (zufälligen) Anzahl N_i der männlichen Einwohner in Altersstufe i ab. Wir modellieren N_i durch eine Poisson-Verteilung (als Approximation einer Binomialverteilung) und bei gegebenem Wert von N_i die Zufallsvariable X_i durch eine Binomialverteilung. Dies führt auf die Modellannahme:

$$P[N_i = n, X_i = x] = \underbrace{\frac{\lambda_i^n}{n!} \cdot e^{-\lambda_i}}_{\text{Zähldichte von } \pi(\lambda_i)} \cdot \underbrace{\binom{n}{x} p_i^x \cdot (1 - p_i)^{n-x}}_{\text{Zähldichte von } b(n, p_i)}.$$

Damit

$$P[X_i = x] = \sum_{n=0}^{\infty} P[N_i = n, X_i = x] = \sum_{n=0}^{\infty} \frac{\lambda_i^n}{n!} \cdot e^{-\lambda_i} \cdot \binom{n}{x} \cdot p_i^x \cdot (1 - p_i)^{n-x}.$$

Gesucht sind hier Schätzungen von λ_i und p_i ausgehend von **einem** beobachteten Wert von (N_i, X_i) .

4.2 Konstruktion von Punktschätzungen

Sei μ_n die empirische Verteilung zu x_1, \dots, x_n .

Bei U - und V -Schätzern und vorausgesetzt:

$$g(\vartheta) = \int \dots \int h(z_1, \dots, z_l) w_{\vartheta}(dz_1) \dots w_{\vartheta}(dz_l).$$

Bei **V-Schätzern** wird w_ϑ durch μ_n geschätzt und das entsprechende Integral als Schätzer von $g(\vartheta)$ verwendet, d. h. $g(\vartheta)$ wird geschätzt durch:

$$\begin{aligned} T_n(x_1, \dots, x_n) &= \int \dots \int h(z_1, \dots, z_l) \mu_n(dz_1) \dots \mu(dz_l) \\ &= \frac{1}{n^l} \sum_{i_1=1}^n \dots \sum_{i_l=1}^n h(x_{i_1}, \dots, x_{i_l}), \end{aligned}$$

wobei die letzte Gleichheit aus

$$\int f(z) d\mu_n = \frac{1}{n} \sum_{i=1}^n \int f(z) d\delta_{x_i} = \frac{1}{n} \sum_{i=1}^n f(x_i)$$

(wobei δ_{x_i} das Dirac-Maß zum Punkt x_i ist) folgt (vgl. Übungen).

Beispiel 4.4: Der V-Schätzer der Varianz

$$\begin{aligned} g(\vartheta) &= V_\vartheta(X_1) = E_\vartheta(X_1 \cdot (X_1 - X_2)) \\ &= \int \int z_1 \cdot (z_1 - z_2) w_\vartheta(dz_1) w_\vartheta(dz_2) \end{aligned}$$

ist

$$\begin{aligned} T_n(x_1, \dots, x_n) &= \int \int z_1 \cdot (z_1 - z_2) \mu_n(dz_1) \mu_n(dz_2) \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ mit } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \end{aligned}$$

(vgl. Übungen).

Eine Variante des V-Schätzers ist der sogenannte **U-Schätzer** gegeben durch

$$\begin{aligned} \bar{T}_n(x_1, \dots, x_n) &= \frac{1}{n \cdot (n-1) \cdot \dots \cdot (n-l+1)} \sum_{i_1=1, \dots, n} \sum_{\substack{i_2=1, \dots, n \\ i_2 \neq i_1}} \dots \sum_{\substack{i_l=1, \dots, n \\ i_l \neq i_1, \dots, i_l \neq i_{l-1}}} h(x_{i_1}, \dots, x_{i_l}). \end{aligned}$$

U-Schätzer sind immer erwartungstreu (englisch: unbiased) im Sinne von:

Definition 4.1

Die Schätzung T_n heißt **erwartungstreu** für $g(\vartheta)$, falls für **alle** $\vartheta \in \Theta$ gilt:

$$E_\vartheta T_n(X_1, \dots, X_n) = g(\vartheta).$$

Dass U-Schätzer immer erwartungstreu für $g(\vartheta)$ sind, folgt aus

$$\begin{aligned}
 E_{\vartheta}[\bar{T}_n(X_1, \dots, X_n)] &= E_{\vartheta}h(X_1, \dots, X_l) \\
 &\quad (\text{da } P_{(X_1, \dots, X_l)} = P_{(X_{i_1}, \dots, X_{i_l})} \\
 &\quad \text{für alle Indizes } i_1, \dots, i_l \in \{1, \dots, n\} \\
 &\quad \text{mit } i_j \neq i_k \text{ für } j \neq k) \\
 &= \int \dots \int h(x_1, \dots, x_l) w_{\vartheta}(dx_1) \dots w_{\vartheta}(dx_l) \\
 &= g(\vartheta) \\
 &\quad (\text{nach Voraussetzung}).
 \end{aligned}$$

Im Falle der Schätzung der Varianz in Beispiel 4.4 ergibt sich als U-Schätzer:

$$\begin{aligned}
 \bar{T}_n(x_1, \dots, x_n) &= \frac{1}{n \cdot (n-1)} \sum_{i=1}^n \sum_{\substack{j=1, \dots, n \\ j \neq i}} x_i \cdot (x_i - x_j) \\
 &= \frac{n}{n-1} \cdot \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n x_i \cdot (x_i - x_j) \\
 &= \frac{n}{n-1} \cdot T_n(x_1, \dots, x_n).
 \end{aligned}$$

Da nach oben der U-Schätzer immer erwartungstreu ist, folgt aus

$$\bar{T}_n(x_1, \dots, x_n) = \frac{n}{n-1} \cdot T_n(x_1, \dots, x_n),$$

dass in Beispiel 4.4 der V-Schätzer **nicht** erwartungstreu ist. Also sind V-Schätzer im allgemeinen nicht erwartungstreu.

Ein zu V-Schätzern verwandtes Konstruktionsprinzip ist das sogenannte **Substitutionsprinzip für Erwartungswerte**.

Hierbei wird - sofern möglich - $g(\vartheta)$ als Funktion

$$h(E_{\vartheta}(X_1), E_{\vartheta}(X_1^2), \dots, E_{\vartheta}(X_1^l))$$

dargestellt, die Momente $E_{\vartheta}(X_1^p)$ werden durch ihre V-Schätzer

$$\frac{1}{n} \sum_{i=1}^n x_i^p$$

geschätzt, und dann wird

$$h\left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n x_i^2, \dots, \frac{1}{n} \sum_{i=1}^n x_i^l\right)$$

als Schätzung von $g(\vartheta)$ verwendet.

Abschließend werden noch **Maximum-Likelihood-Schätzungen** eingeführt. Dabei wird vorausgesetzt, dass w_ϑ eine Dichte f_ϑ bzgl. eines σ -endlichen Maßes ν hat (z. B. $\nu = \text{LB-Maß}$ und f_ϑ Dichte bzgl. des LB-Maßes, oder $\nu = \text{abzählendes Maß}$ und $f_\vartheta(x) = P_\vartheta[X = x]$).

Wegen der Unabhängigkeit und identischen Verteiltheit von X_1, \dots, X_n hat dann (X_1, \dots, X_n) die Dichte

$$g_\vartheta(x_1, \dots, x_n) = \prod_{i=1}^n f_\vartheta(x_i).$$

Die Idee beim **Maximum-Likelihood-Prinzip** ist nun, bei beobachteten x_1, \dots, x_n als Schätzung für ϑ (d.h. hier ist $g(\vartheta) = \vartheta$) dasjenige $\hat{\vartheta}$ zu verwenden, für das die sogenannte **Likelihood-Funktion**

$$L(\vartheta) = \prod_{i=1}^n f_\vartheta(x_i)$$

maximal wird.

(Die Heuristik dahinter ist, dass für eine kleine Umgebung $U(x_1, \dots, x_n)$ von (x_1, \dots, x_n)

$$(\otimes_{i=1}^n w_\vartheta)(U(x_1, \dots, x_n)) \approx \left(\prod_{i=1}^n f_\vartheta(x_i) \right) \cdot (\otimes_{i=1}^n \nu)(U(x_1, \dots, x_n))$$

genau dann groß ist, wenn $L(\vartheta)$ groß ist).

Statt $L(\vartheta)$ wird manchmal auch die sogenannte Log-Likelihood-Funktion

$$\ln(L(\vartheta)) = \sum_{i=1}^n \ln(f_\vartheta(x_i))$$

maximiert. Da $x \mapsto \ln(x)$ auf \mathbb{R}_+ streng monoton wachsend ist, ist dies im Falle $f_\vartheta(x) \neq 0$ für alle $\vartheta \in \Theta$ und alle x äquivalent zur Maximierung von $L(\vartheta)$. Die Definition eines Schätzers durch Maximierung der Log-Likelihood-Funktion lässt sich motivieren durch:

Lemma 4.1. Ist X \mathbb{R}^d -wertige Zufallsvariable mit Dichte f (bzgl. des LB-Maßes), so gilt für jede beliebige Dichte $g : \mathbb{R}^d \rightarrow \mathbb{R}$ (bzgl. des LB-Maßes):

$$\mathbf{E}[\log f(X)] \geq \mathbf{E}[\log g(X)].$$

Beweis: Sei g beliebige Dichte (bzgl. des LB-Maßes). Dann gilt oBdA $\mathbf{P}[g(X) = 0] = 0$, da andernfalls die rechte Seite gleich $-\infty$ ist. Weiter gilt (da f Dichte von X ist) auch:

$$\mathbf{P}[f(X) = 0] = \mathbf{P}_X(\{x \in \mathbb{R}^d : f(x) = 0\}) = \int_{\{x \in \mathbb{R}^d : f(x) = 0\}} f(z) dz = 0.$$

Durch Abändern der ZV X auf einer Nullmenge können wir darüberhinaus sogar voraussetzen, dass $g(X)$ und $f(X)$ nur positive Werte annehmen. Es genügt daher zu zeigen:

$$\mathbf{E} \left[\log \left(\frac{g(X)}{f(X)} \right) \right] \leq 0.$$

Wegen der Konkavität der Logarithmusfunktion und der Ungleichung von Jensen folgt dies aber aus:

$$\begin{aligned} \mathbf{E} \left[\log \left(\frac{g(X)}{f(X)} \right) \right] &\leq \log \left(\mathbf{E} \left[\frac{g(X)}{f(X)} \right] \right) \\ &\stackrel{f \text{ Dichte von } X}{=} \log \left(\int_{\mathbb{R}^d} \frac{g(x)}{f(x)} \cdot f(x) dx \right) \\ &\stackrel{\log \text{ monoton}}{\leq} \log \left(\int_{\mathbb{R}^d} g(x) dx \right) = \log(1) = 0. \end{aligned}$$

□

Hat nun die Verteilung w_θ von X eine Dichte f_θ (bzgl. des LB-Maßes) für $\theta \in \Theta$, so gilt

$$\theta = \arg \max_{\bar{\theta} \in \Theta} \mathbf{E}_\theta[\log f_{\bar{\theta}}(X)].$$

Beim Log-Likelihood-Schätzer wird nun $\mathbf{E}_\theta[\log f_{\bar{\theta}}(X)]$ durch das entsprechende Stichprobenmittel

$$\frac{1}{n} \sum_{i=1}^n \log f_{\bar{\theta}}(X_i)$$

geschätzt (wobei X, X_1, \dots, X_n unabhängig identisch verteilt sind mit Verteilung w_θ), und diese Schätzung wird maximiert, d.h.

$$\hat{\theta} = \arg \max_{\bar{\theta} \in \Theta} \frac{1}{n} \sum_{i=1}^n \log f_{\bar{\theta}}(X_i).$$

Bei Differenzierbarkeit hat man dabei die notwendige Bedingung

$$\frac{\partial}{\partial \vartheta} [l_n(L(\vartheta))] = 0.$$

Als Schätzung von $g(\vartheta)$ wird im Falle $g(\vartheta) \neq \vartheta$

$$g(\hat{\vartheta})$$

(mit $\hat{\vartheta}$ wie oben) verwendet.

Anwendung in Beispiel 4.1:

Hier ist

$$f_{\vartheta}(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \vartheta = (\mu, \sigma).$$

Maximierung von

$$L(\vartheta) = \prod_{i=1}^n f_{\vartheta}(x_i) = \frac{1}{(2\pi)^{n/2} \cdot \sigma^n} \cdot \exp\left(-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

führt auf

$$\begin{aligned} 0 &\stackrel{!}{=} \frac{\partial}{\partial \mu} \ln L(\vartheta) = 0 + \frac{\partial}{\partial \mu} \left[-\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu), \end{aligned}$$

also auf

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i,$$

sowie auf

$$0 \stackrel{!}{=} \frac{\partial}{\partial \sigma} \ln L(\vartheta) = -\frac{n}{\sigma} + \sum_{i=1}^n (x_i - \mu)^2 \cdot \sigma^{-3},$$

also

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \text{ mit } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Hier stimmt der Maximum-Likelihood-Schätzer mit dem V-Schätzer überein. Da der V-Schätzer in diesem Beispiel nicht erwartungstreu ist, sieht man: Maximum-Likelihood-Schätzer sind im allgemeinen nicht erwartungstreu.

Anwendung in Beispiel 4.2

Hier muss

$$L(\lambda) = \prod_{i=1}^n P_{\lambda}[X_i = x_i] = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} \cdot e^{-\lambda}$$

maximiert werden, was auf

$$\begin{aligned} 0 &\stackrel{!}{=} \frac{\partial}{\partial \lambda} \ln(L(\lambda)) = \frac{\partial}{\partial \lambda} \sum_{i=1}^n \left(\ln \frac{1}{x_i!} + x_i \cdot \ln(\lambda) - \lambda \right) \\ &= \sum_{i=1}^n x_i \cdot \frac{1}{\lambda} - n \end{aligned}$$

führt. Damit stimmt der Maximum-Likelihood-Schätzer

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i$$

hier wieder mit dem V-Schätzer überein.

Anwendung in Beispiel 4.3

Hier ist $\vartheta = (\lambda_i, p_i)$, Stichprobenumfang $n = 1$ und

$$L(\vartheta) = f_{\vartheta}(n_i, x_i) = \frac{\lambda_i^{n_i}}{n_i!} \cdot e^{-\lambda_i} \cdot \binom{n_i}{x_i} p_i^{x_i} (1 - p_i)^{n_i - x_i},$$

wobei n_i bzw. x_i die beobachteten Anzahlen von Männern bzw. männlichen Krebsstoten in Altersstufe i sind.

Maximieren von $L(\vartheta)$ führt auf

$$\begin{aligned} 0 &\stackrel{!}{=} \frac{\partial}{\partial \lambda_i} \ln(L(\vartheta)) = \frac{\partial}{\partial \lambda_i} \left[\ln \frac{1}{n_i!} + n_i \cdot \ln(\lambda_i) - \lambda_i \right] + 0 \\ &= \frac{n_i}{\lambda_i} - 1, \end{aligned}$$

also

$$\hat{\lambda}_i = n_i,$$

sowie

$$\begin{aligned} 0 &= \frac{\partial}{\partial p_i} \ln(L(\vartheta)) = \frac{\partial}{\partial p_i} [x_i \cdot \ln(p_i) + (n_i - x_i) \cdot \ln(1 - p_i)] + 0 \\ &= x_i \cdot \frac{1}{p_i} - (n_i - x_i) \cdot \frac{1}{1 - p_i}, \end{aligned}$$

was äquivalent ist zu

$$\begin{aligned} 0 &= x_i \cdot (1 - p_i) - (n_i - x_i) \cdot p_i \\ &= x_i - p_i \cdot x_i - (n_i - x_i) \cdot p_i \\ &= x_i - n_i \cdot p_i \end{aligned}$$

also

$$\hat{p}_i = \frac{x_i}{n_i}.$$

Problem: Schätzung von (λ_i, p_i) beruht nur auf einer Beobachtung der zugehörigen Zufallsvariablen.

Ausweg: Strukturelle Annahme (z. B. $\log \frac{p_i}{1-p_i} = \alpha + i \cdot \beta$ ($i = 1, \dots, 8$)) und alle Parameter simultan schätzen ...

4.3 Optimale Schätzverfahren

Frage: Was ist ein “optimales” Schätzverfahren?

Gegeben sei eine **Verlustfunktion** $l : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}_+$.

Bei Vorhersage von $g(\vartheta)$ durch $T(x_1, \dots, x_n)$ sei

$$l(T(x_1, \dots, x_n), g(\vartheta)) \geq 0$$

der auftretende Verlust.

Es gelte $l(v, v) = 0$ für alle $v \in \mathbb{R}^k$.

Beispiel für $k=1$: $l(u, v) = |u - v|^p$ für ein $p \geq 1$ (fest).

Bei Vorliegen des Parameters ϑ und wiederholter Vorhersage von $g(\vartheta)$ durch $T(x_1, \dots, x_n)$ tritt im Mittel der Verlust

$$r_T(\vartheta) = E_\vartheta[l(T(X_1, \dots, X_n), g(\vartheta))]$$

(sogenanntes **Risiko**) auf.

Wünschenswert: Schätzfunktion T_{opt} mit

$$r_{T_{opt}}(\vartheta) \leq r_T(\vartheta) \text{ für alle } \vartheta \in \Theta \text{ und alle } T.$$

Problem: Solch ein Verfahren existiert im allgemeinen nicht!

z. B. hat

$$T_{\vartheta_0}(x_1, \dots, x_n) = g(\vartheta_0) \text{ das Risiko } r_{T_{\vartheta_0}}(\vartheta_0) = 0,$$

also muss für T_{opt} von oben $r_{T_{opt}}(\vartheta) = 0$ für alle $\vartheta \in \Theta$ gelten, was im allgemeinen unmöglich ist.

Ausweg: Abschwächung des obigen Optimalitätskriteriums.

Möglichkeit 1: Minimax-Prinzip

Minimiere das maximale Risiko:

$$\sup_{\vartheta \in \Theta} r_{T_{opt}}(\vartheta) \stackrel{!}{=} \inf_T \sup_{\vartheta \in \Theta} r_T(\vartheta)$$

Möglichkeit 2: Bayes-Prinzip

Vorausgesetzt wird hier apriori Information über das Auftreten der einzelnen Parameterwerte. Dazu wird eine Verteilung ρ auf dem Parameterraum Θ vorgegeben.

Minimiert wird dann das mittlere Risiko bzgl. dieser Verteilung:

$$\int_{\Theta} r_{T_{opt}}(\vartheta) \rho(d\vartheta) \stackrel{!}{=} \inf_T \int_{\Theta} r_T(\vartheta) \rho(d\vartheta).$$

Möglichkeit 3: Einschränkung der Klasse der betrachteten Schätzfunktionen

Betrachte nur Schätzfunktionen T aus einer vorgegebenen Klasse Δ , d. h. Schätzfunktionen mit einer gewissen Eigenschaft, wie z. B. Erwartungstreue. Suche dann in dieser Klasse ein optimales Verfahren $T_{opt} \in \Delta$ mit:

$$r_{T_{opt}}(\vartheta) \leq r_T(\vartheta) \text{ für alle } \vartheta \in \Theta \text{ und alle } T \in \Delta.$$

Im Folgenden: Untersuchung von Möglichkeit 3.

4.4 Der Begriff des optimalen erwartungstreuen Schätzers

Ausgehend von u. i. v. Zufallsvariablen X_1, \dots, X_n mit

$$P_{X_1} \in \{w_{\vartheta} : \vartheta \in \Theta\}$$

soll

$$g(\vartheta)$$

für ein $g : \Theta \rightarrow \mathbb{R}$ geschätzt werden.

$T(X_1, \dots, X_n)$ sei eine Schätzfunktion.

Wünschenswert:

$$P_{T(X_1, \dots, X_n)} \text{ soll möglichst stark um } g(\vartheta) \text{ konzentriert sein.}$$

Mögliche Präzisierung:

(1) $T(X_1, \dots, X_n)$ erwartungstreu für $g(\vartheta)$, d. h.

$$E_{\vartheta} T(X_1, \dots, X_n) = g(\vartheta) \text{ für alle } \vartheta \in \Theta.$$

(2) $V_{\vartheta}(T(X_1, \dots, X_n))$ "möglichst klein" für alle $\vartheta \in \Theta$.

Definition 4.2: T heißt **gleichmäßig bester erwartungstreuer Schätzer** für $g(\vartheta)$, falls gilt:

(I) T ist erwartungstreu für $g(\vartheta)$.

(II) Für **alle** erwartungstreuen Schätzer \bar{T} für $g(\vartheta)$ und **alle** $\vartheta \in \Theta$ gilt:

$$V_{\vartheta}(\bar{T}(X_1, \dots, X_n)) \geq V_{\vartheta}(T(X_1, \dots, X_n)).$$

Bemerkung 1: Ist T gleichmäßig bester erwartungstreuer Schätzer für $g(\vartheta)$, dann gilt für alle $\vartheta \in \Theta$:

$$E_{\vartheta}[(T(X_1, \dots, X_n) - g(\vartheta))^2] = \min_{\bar{T} \text{ erwartungstreu für } g(\vartheta)} E_{\vartheta}[(\bar{T}(X_1, \dots, X_n) - g(\vartheta))^2]$$

Begründung:

Für einen beliebigen Schätzer \tilde{T} gilt

$$\begin{aligned} & E_{\vartheta}[(\tilde{T}(X_1, \dots, X_n) - g(\vartheta))^2] \\ &= E_{\vartheta}[(\tilde{T}(X_1, \dots, X_n) - E_{\vartheta}\tilde{T}(X_1, \dots, X_n)) + (E_{\vartheta}\tilde{T}(X_1, \dots, X_n) - g(\vartheta))]^2 \\ &= \underbrace{E_{\vartheta}[(\tilde{T}(X_1, \dots, X_n) - E_{\vartheta}\tilde{T}(X_1, \dots, X_n))^2]}_{=V_{\vartheta}(\tilde{T}(X_1, \dots, X_n))} \\ &\quad + (E_{\vartheta}\tilde{T}(X_1, \dots, X_n) - g(\vartheta))^2 + 0, \end{aligned}$$

da

$$\begin{aligned} & E_{\vartheta}[(\tilde{T}(X_1, \dots, X_n) - E_{\vartheta}\tilde{T}(X_1, \dots, X_n)) \cdot (E_{\vartheta}\tilde{T}(X_1, \dots, X_n) - g(\vartheta))] \\ &= (E_{\vartheta}\tilde{T}(X_1, \dots, X_n) - g(\vartheta)) \cdot \underbrace{E_{\vartheta}[\tilde{T}(X_1, \dots, X_n) - E_{\vartheta}\tilde{T}(X_1, \dots, X_n)]}_{=0} \\ &= 0 \end{aligned}$$

gilt.

Hierbei heißt $V_{\vartheta}(\tilde{T}(X_1, \dots, X_n))$ **Varianz** und

$$E_{\vartheta}\tilde{T}(X_1, \dots, X_n) - g(\vartheta) \quad \text{bias} \quad (\text{Verzerrung}) \text{ des Schätzers } \tilde{T}.$$

Ist nun \tilde{T} erwartungstreu, so ist der Bias gleich Null und man sieht: Für erwartungstreue Schätzer ist die Minimierung des mittleren quadratischen Vorhersagefehlers äquivalent zur Minimierung der Varianz. \square

Bemerkung 2: Erwartungstreue Schätzer existieren nicht immer (siehe nächstes Beispiel).

Beispiel 4.5:

Sei $\Theta = (0, 1)$, $w_{\vartheta} = b(1, \vartheta)$ -Verteilung und $n = 1$. Angenommen, $T(X_1)$ ist erwartungstreu für $\frac{1}{\vartheta}$, dann gilt:

$$\frac{1}{\vartheta} = E_{\vartheta}T(X_1) = T(0) \cdot P_{\vartheta}[X_1 = 0] + T(1) \cdot P_{\vartheta}[X_1 = 1] \quad (*)$$

also

$$\frac{1}{\vartheta} = T(0) \cdot (1 - \vartheta) + T(1) \cdot \vartheta.$$

Mit $\vartheta \rightarrow 0$ ergibt sich:

$$T(0) = \infty,$$

womit (*) nicht gelten kann.

4.5 Die Informationsungleichung von Cramér-Rao

Im Folgenden: Herleitung einer unteren Schranke für die Varianz von Schätzern.

Dazu beachten wir, dass gemäß der Ungleichung von Cauchy-Schwarz für beliebige Schätzer $T(X_1, \dots, X_n)$ und beliebige Zufallsvariablen V gilt:

$$\begin{aligned} & \text{Cov}_{\vartheta}(T(X_1, \dots, X_n), V) \\ & \stackrel{\text{Def.}}{=} E_{\vartheta}((T(X_1, \dots, X_n) - E_{\vartheta}T(X_1, \dots, X_n)) \cdot (V - E_{\vartheta}V)) \\ & \leq \sqrt{V_{\vartheta}(T(X_1, \dots, X_n))} \cdot \sqrt{V_{\vartheta}(V)}, \end{aligned}$$

woraus im Falle $V_{\vartheta}(V) > 0$ folgt:

$$(4.1) \quad V_{\vartheta}(T(X_1, \dots, X_n)) \geq \frac{(\text{Cov}(T(X_1, \dots, X_n), V))^2}{V_{\vartheta}(V)}.$$

Sei nun μ ein σ -endliches Maß und $f(\vartheta, \cdot)$ eine μ -Dichte von w_ϑ . Wir wählen dann speziell:

$$V = \frac{\partial}{\partial \vartheta} \log \left(\prod_{i=1}^n f(\vartheta, X_i) \right) = \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log(f(\vartheta, X_i)).$$

Sofern die mit (*) gekennzeichneten Umformungen dann zulässig sind, gilt:

(i)

$$\begin{aligned} E_\vartheta \left(\frac{\partial}{\partial \vartheta} \log(f(\vartheta, X_1)) \right) &= \int \frac{\partial}{\partial \vartheta} [\log f(\vartheta, x)] \cdot f(\vartheta, x) \cdot \mu(dx) \\ &= \int \frac{1}{f(\vartheta, x)} \cdot \frac{\partial}{\partial \vartheta} [f(\vartheta, x)] \cdot f(\vartheta, x) \cdot \mu(dx) \\ &= \int \frac{\partial}{\partial \vartheta} [f(\vartheta, x)] \cdot \mu(dx) \\ &\stackrel{(*)}{=} \frac{\partial}{\partial \vartheta} \int f(\vartheta, x) \cdot \mu(dx) \\ &= \frac{\partial}{\partial \vartheta} 1 = 0. \end{aligned}$$

(ii)

$$\begin{aligned} V_\vartheta(V) &= \sum_{i=1}^n V_\vartheta \left(\frac{\partial}{\partial \vartheta} [\log f(\vartheta, X_i)] \right) \\ &\quad \text{(wegen Unabhängigkeit)} \\ &= n \cdot V_\vartheta \left(\frac{\partial}{\partial \vartheta} [\log f(\vartheta, X_1)] \right) \\ &\quad \text{(wegen identischer Verteiltheit)} \\ &\stackrel{s.o.}{=} n \cdot E_\vartheta \left[\left[\frac{\partial}{\partial \vartheta} \log f(\vartheta, X_1) \right]^2 \right]. \quad (4.2) \end{aligned}$$

(iii)

$$\begin{aligned} &Cov_\vartheta(T(X_1, \dots, X_n), V) \\ &= E_\vartheta((T(X_1, \dots, X_n) - E_\vartheta(T(X_1, \dots, X_n))) \cdot (V - E_\vartheta V)) \\ &= E_\vartheta(T(X_1, \dots, X_n) \cdot V) \\ &\quad \text{(wegen } E_\vartheta V = 0, \text{ woraus auch } E_\vartheta(E_\vartheta(T(X_1, \dots, X_n)) \cdot V) = 0 \text{ folgt)} \\ &= E_\vartheta \left(T(X_1, \dots, X_n) \cdot \frac{\partial}{\partial \vartheta} [\log \prod_{i=1}^n f(\vartheta, X_i)] \right) \\ &= \int \dots \int T(x_1, \dots, x_n) \cdot \frac{\partial}{\partial \vartheta} [\log \prod_{i=1}^n f(\vartheta, x_i)] \cdot \prod_{i=1}^n f(\vartheta, x_i) \cdot \mu(dx_1) \dots \mu(dx_n) \\ &= \int \dots \int T(x_1, \dots, x_n) \cdot \frac{\partial}{\partial \vartheta} [\prod_{i=1}^n f(\vartheta, x_i)] \cdot \mu(dx_1) \dots \mu(dx_n) \\ &\stackrel{(*)}{=} \frac{\partial}{\partial \vartheta} [\int \dots \int T(x_1, \dots, x_n) \prod_{i=1}^n f(\vartheta, x_i) \cdot \mu(dx_1) \dots \mu(dx_n)] \\ &= \frac{\partial}{\partial \vartheta} [E_\vartheta T(X_1, \dots, X_n)]. \quad (4.3) \end{aligned}$$

Setzt man nun (4.2) und (4.3) in (4.1) ein, so hat man den folgenden Satz gezeigt:

Satz 4.2: Sei $\Theta \subseteq \mathbb{R}$ ein Intervall. Für $\vartheta \in \Theta$ besitze w_ϑ eine Dichte $f(\vartheta, \cdot)$ bzgl. eines σ -endlichen Maßes μ und es gelte:

$$(4.4) \quad \forall \vartheta \in \Theta : \mu\text{-f.ü. ist } \frac{\partial f(\vartheta, \cdot)}{\partial \vartheta} \text{ existent und endlich.}$$

$$(4.5) \quad \forall \vartheta \in \Theta : \int \frac{\partial}{\partial \vartheta} [f(\vartheta, x)] \mu(dx) = \frac{\partial}{\partial \vartheta} \int f(\vartheta, x) \mu(dx) = 0.$$

$$(4.6) \quad \forall \vartheta \in \Theta : 0 < I(\vartheta) := E_\vartheta \left\{ \frac{\partial}{\partial \vartheta} \log f(\vartheta, X_1) \right\}^2 < \infty.$$

Dann gilt für jede Schätzfunktion $T(X_1, \dots, X_n)$ mit

$$(4.7) \quad \forall \vartheta \in \Theta : \int \dots \int T(x_1, \dots, x_n) \frac{\partial}{\partial \vartheta} \left[\prod_{i=1}^n f(\vartheta, x_i) \right] \mu(dx_1) \dots \mu(dx_n) \\ = \frac{\partial}{\partial \vartheta} \int \dots \int T(x_1, \dots, x_n) \prod_{i=1}^n f(\vartheta, x_i) \mu(dx_1) \dots \mu(dx_n)$$

die folgende Abschätzung für die Varianz:

$$\forall \vartheta \in \Theta \quad V_\vartheta(T(X_1, \dots, X_n)) \geq \frac{\left(\frac{\partial}{\partial \vartheta} [E_\vartheta T(X_1, \dots, X_n)] \right)^2}{n \cdot I(\vartheta)}.$$

Im Spezialfall, dass T eine erwartungstreue Schätzung für $g(\vartheta)$ ist, gilt wegen

$$\frac{\partial}{\partial \vartheta} [E_\vartheta T(X_1, \dots, X_n)] = \frac{\partial}{\partial \vartheta} g(\vartheta) = g'(\vartheta)$$

insbesondere die sogenannte **Informationsungleichung von Cramér-Rao**:

$$\forall \vartheta \in \Theta : V_\vartheta(T(X_1, \dots, X_n)) \geq \frac{g'(\vartheta)^2}{n \cdot I(\vartheta)}.$$

Beweis: Folgt auf der obigen Herleitung, indem man (4.2) und (4.3) in (4.1) eingesetzt und beachtet, dass die mit (*) gekennzeichneten Umformungen aufgrund der Voraussetzungen des Satzes zulässig sind. \square

Definition 4.3:

a) $I(\vartheta) = E_\vartheta \left[\left(\frac{\partial}{\partial \vartheta} \log f(\vartheta, X_1) \right)^2 \right]$ heißt **Fisher-Information**.

- b) Ist T eine erwartungstreue Schätzfunktion für $g(\vartheta)$, so heißt die nach der Informationsungleichung im Intervall $[0, 1]$ liegende Zahl

$$\frac{(g'(\vartheta))^2 / (n \cdot I(\vartheta))}{V_{\vartheta}(T(X_1, \dots, X_n))}$$

Effizienz (oder Wirksamkeit) von T .

- c) Eine Schätzfunktion mit Effizienz = 1 für alle $\vartheta \in \Theta$ heißt **effizient**.

Klar: Erfüllen alle erwartungstreuen Schätzer die Voraussetzungen von Satz 4.1, so ist jeder erwartungstreue und effiziente Schätzer ein gleichmäßig bester erwartungstreuer Schätzer.

Aber: Effiziente Schätzer existieren nicht immer ...

Im Folgenden: Zwei Beispiele für effiziente Schätzverfahren, für die gezeigt wird, dass sie gleichmäßig beste erwartungstreue Schätzer sind.

Beispiel 4.6: $\Theta = (0, \infty)$, $w_{\vartheta} = \pi(\vartheta)$, d. h. $w_{\vartheta}(\{k\}) = \frac{\vartheta^k}{k!} \cdot e^{-\vartheta}$ ($k \in \mathbb{N}_0$)

Gesucht ist gleichmäßig bester erwartungstreuer Schätzer für $g(\vartheta) = \vartheta = \int x w_{\vartheta}(dx)$.

w_{ϑ} hat Dichte

$$f(\vartheta, x) = \frac{\vartheta^x}{x!} \cdot e^{-\vartheta} \cdot I_{\mathbb{N}_0}(x)$$

bzgl. des abzählenden Maßes μ .

Wir zeigen nun, dass Satz 4.1 anwendbar ist. Dazu beachten wir:

$\forall x \in \mathbb{N}_0$ existiert $\frac{\partial f(\vartheta, x)}{\partial \vartheta}$ und ist endlich, also ist (4.4) erfüllt.

Weiter gilt für $x \in \mathbb{N}_0$

$$\frac{\partial}{\partial \vartheta} [\log f(\vartheta, x)] = \frac{\partial}{\partial \vartheta} [x \cdot \log \vartheta - \log(x!) - \vartheta] = \frac{x}{\vartheta} - 1 = \frac{x - \vartheta}{\vartheta},$$

also ist

$$\begin{aligned} \int \frac{\partial}{\partial \vartheta} f(\vartheta, x) \mu(dx) &= \int \frac{\frac{\partial}{\partial \vartheta} f(\vartheta, x)}{f(\vartheta, x)} \cdot f(\vartheta, x) \mu(dx) \\ &= \int \frac{\partial}{\partial \vartheta} [\log f(\vartheta, x)] \cdot f(\vartheta, x) \mu(dx) \\ &= E_{\vartheta} \left(\frac{\partial}{\partial \vartheta} [\log f(\vartheta, X_1)] \right) \\ &\stackrel{s.o.}{=} E_{\vartheta} \left(\frac{X_1 - \vartheta}{\vartheta} \right) = \frac{E_{\vartheta}(X_1) - \vartheta}{\vartheta} = \frac{\vartheta - \vartheta}{\vartheta} = 0 \\ &= \frac{\partial}{\partial \vartheta} 1 = \frac{\partial}{\partial \vartheta} \int f(\vartheta, x) \mu(dx), \end{aligned}$$

womit auch (4.5) erfüllt ist.

Wegen

$$\begin{aligned} V_{\vartheta}(X_1) &= E_{\vartheta}(X_1 \cdot (X_1 - 1)) + E_{\vartheta}(X_1) - |(E_{\vartheta}(X_1))|^2 \\ &= \dots = \vartheta^2 + \vartheta - (\vartheta)^2 = \vartheta \end{aligned}$$

gilt

$$\begin{aligned} I(\vartheta) &= E_{\vartheta} \left(\left(\frac{\partial}{\partial \vartheta} [\log f(\vartheta, X_1)] \right)^2 \right) \\ &= E_{\vartheta} \left(\frac{|X_1 - \vartheta|^2}{\vartheta^2} \right) = \frac{1}{\vartheta^2} V_{\vartheta}(\vartheta) = \frac{1}{\vartheta}, \end{aligned}$$

woraus (4.6) folgt

Sei nun $T(X_1, \dots, X_n)$ ein beliebiger erwartungstreuer Schätzer. Zum Nachweis von (4.7) müssen wir

$$\begin{aligned} &\int \dots \int T(x_1, \dots, x_n) \prod_{i=1}^n f(\vartheta, x_i) \mu(dx_1) \dots \mu(dx_n) \\ &= \sum_{x_1=0}^{\infty} \dots \sum_{x_n=0}^{\infty} T(x_1, \dots, x_n) \cdot \prod_{i=1}^n f(\vartheta, x_i) \\ &= \sum_{x_1=0}^{\infty} \dots \sum_{x_n=0}^{\infty} T(x_1, \dots, x_n) \cdot \frac{\vartheta^{x_1 + \dots + x_n}}{x_1! \dots x_n!} \cdot e^{-\vartheta \cdot n} \end{aligned}$$

betrachten und zeigen, dass die Ableitung dieses Ausdrucks mit der gliedweisen Ableitung übereinstimmt.

Der obige Ausdruck lässt sich umschreiben zu

$$e^{-\vartheta \cdot n} \cdot \sum_{k=0}^{\infty} \left(\sum_{\substack{x_1, \dots, x_n \in \mathbb{N}_0, \\ x_1 + \dots + x_n = k}} \frac{T(x_1, \dots, x_n)}{x_1! \cdot \dots \cdot x_n!} \right) \cdot \vartheta^k.$$

Ist nun $V_{\vartheta}(T(X_1, \dots, X_n)) < \infty$, was $E_{\vartheta}|T(X_1, \dots, X_n)| < \infty$ impliziert, so ist das Produkt und damit auch der zweite Faktor absolut konvergent.

Damit kann der zweite Faktor (als absolut konvergente Potenzreihe) gliedweise abgeleitet werden, was impliziert, dass auch die ursprüngliche Reihe gliedweise abgeleitet werden kann. Also ist auch (4.7) (und damit die Informationsungleichung) für jedes $\vartheta \in \Theta$ mit $V_{\vartheta}(T(X_1, \dots, X_n)) < \infty$ nachgewiesen.

Also folgt aus der Informationsungleichung für **jeden** erwartungstreuen Schätzer

$$\forall \vartheta \in \Theta : V_{\vartheta}(T(X_1, \dots, X_n)) \geq \frac{1}{n \cdot I(\vartheta)} \stackrel{s.o.}{=} \frac{\vartheta}{n}.$$

Andererseits gilt für den erwartungstreuen Schätzer $T_0(X_1, \dots, X_n) = \frac{X_1 + \dots + X_n}{n}$:

$$V_{\vartheta}(T_0(X_1, \dots, X_n)) = \frac{1}{n} V_{\vartheta}(X_1) \stackrel{s.o.}{=} \frac{\vartheta}{n},$$

womit gezeigt ist:

$\frac{X_1 + \dots + X_n}{n}$ ist gleichmäßig bester erwartungstreuer Schätzer für ϑ .

Beispiel 4.7:

$\Theta = \mathbb{R}$, $w_\vartheta = \mathcal{N}(\vartheta, \sigma_0^2)$ -Verteilung.

Gesucht ist wieder ein gleichmäßig bester erwartungstreuer Schätzer für $g(\vartheta) = \vartheta = \int x w_\vartheta(dx)$.

w_ϑ hat die Dichte

$$f(\vartheta, x) = \frac{1}{\sqrt{2\pi\sigma_0}} \cdot e^{-\frac{(x-\vartheta)^2}{2\sigma_0^2}}.$$

Für die Fisher-Information gilt hier:

$$\begin{aligned} I(\vartheta) &= E_\vartheta \left(\left(\frac{\partial}{\partial \vartheta} [\log f(\vartheta, X_1)] \right)^2 \right) \\ &= E_\vartheta \left(\left(\frac{\partial}{\partial \vartheta} \left(\log \frac{1}{\sqrt{2\pi\sigma_0}} - \left(\frac{X_1 - \vartheta}{2\sigma_0^2} \right)^2 \right) \right)^2 \right) \\ &= E_\vartheta \left(\left(\frac{X_1 - \vartheta}{\sigma_0^2} \right)^2 \right) = \frac{V_\vartheta(X_1)}{\sigma_0^4} = \frac{1}{\sigma_0^2}. \end{aligned}$$

Man kann wieder zeigen, dass die Voraussetzungen von Satz 4.1 erfüllt sind, und erhält für jeden erwartungstreuen Schätzer für ϑ :

$$V_\vartheta(T(X_1, \dots, X_n)) \geq \frac{1}{n \cdot I(\vartheta)} = \frac{\sigma_0^2}{n} = V_\vartheta \left(\frac{X_1 + \dots + X_n}{n} \right).$$

Also ist wieder $\frac{X_1 + \dots + X_n}{n}$ ein gleichmäßig bester erwartungstreuer Schätzer für ϑ .

4.6 Suffizienz

Zwecks Vereinfachung der Notation betrachten wir nun folgendes Problem:

gegeben:

\mathbb{R}^n -wertige ZV X mit $P_X = w_\vartheta$ für ein $\vartheta \in \Theta \subseteq \mathbb{R}$.

gesucht:

Schätzung $T(X)$ von ϑ .

Wir versuchen nun, $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (möglichst mit $m < n$) so zu bestimmen, dass $S(X)$ alle zur Vorhersage von ϑ nötigen Informationen enthält.

Beispiel:

$X = (X_1, \dots, X_n)$ mit X_1, \dots, X_n unabhängig identisch $b(1, \vartheta)$ -verteilt mit $\vartheta \in [0, 1]$. Wegen der Unabhängigkeit der X_1, \dots, X_n spielt die Reihenfolge der auftretenden Nullen und Einsen keine Rolle, also ist es naheliegend, statt

$$(X_1, \dots, X_n)$$

nur

$$S(X_1, \dots, X_n) = \sum_{i=1}^n X_i$$

zu betrachten.

Wir fordern für S : Für beliebiges (messbares) $T : \mathbb{R}^n \rightarrow \mathbb{R}$ soll die Kenntnis von $T(X)$ bei gegebenem $S(X) = s$ keinen Rückschluss mehr auf ϑ erlauben, d.h. $P_\vartheta[T(X) \in A | S(X) = s]$ hängt nicht von ϑ ab für alle $A \in \mathcal{B}$. Da T beliebig ist, ist dies äquivalent zu $P_\vartheta[X \in B | S(X) = s]$ hängt nicht von ϑ ab für alle $B \in \mathcal{B}_n$.

Im Beispiel oben: Für $x = (x_1, \dots, x_n) \in \{0, 1\}^n$ und $s = S(x) = x_1 + \dots + x_n$ gilt

$$P_\vartheta[X = x | S(X) = s] = \frac{P_\vartheta[X = x, S(X) = s]}{P_\vartheta[S(X) = s]}$$

$$\stackrel{\text{da } S(x)=s}{=} \frac{P_\vartheta[(X_1, \dots, X_n) = (x_1, \dots, x_n)]}{P_\vartheta[(X_1 + \dots + X_n) = x_1 + \dots + x_n]}$$

$$\stackrel{X_1 + \dots + X_n \sim b(n, \vartheta)}{=} \frac{\prod_{i=1}^n \vartheta^{x_i} \cdot (1 - \vartheta)^{1-x_i}}{\binom{n}{x_1 + \dots + x_n} \cdot \vartheta^{x_1 + \dots + x_n} \cdot (1 - \vartheta)^{n - (x_1 + \dots + x_n)}} = \frac{1}{\binom{n}{x_1 + \dots + x_n}} = \frac{1}{\binom{n}{s}}$$

hängt nicht von ϑ ab (ebenso nicht im Fall $S(x) \neq s$, da dann obige Wk. Null ist).

Def. 4.4: Sei $\Theta \subseteq \mathbb{R}$ und $\{w_\vartheta : \vartheta \in \Theta\}$ Familie von W-Maßen auf $(\mathbb{R}^n, \mathcal{B}_n)$. Sei X \mathbb{R}^n -wertige ZV mit $P_X = w_\vartheta$ für ein $\vartheta \in \Theta$. Eine (messbare) Funktion $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$ heißt **suffizient** für ϑ , falls für alle $B \in \mathcal{B}_n$ eine von ϑ unabhängige Festlegung von

$$P_\vartheta[X \in B | S(X) = \cdot]$$

möglich ist.

Lemma 4.3

Sei S suffizient für ϑ , und sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ messbar mit $E_\vartheta[|f(X)|] < \infty$ für alle $\vartheta \in \Theta$. Dann existiert eine von ϑ unabhängige Festlegung von

$$E_\vartheta[f(X)|S(X) = \cdot].$$

Beweis: folgt für $f = 1_B$ unmittelbar aus Definition 4.4.

Dann Beweis schrittweise gemäß der schrittweisen Definition des Erwartungswertes, vgl. Übungen. \square

Der nächste Satz zeigt, dass für S suffizient für ϑ die Schätzung von ϑ durch Betrachtung von $S(X)$ anstelle von X erfolgen kann. Also kann in obigem Beispiel statt

$$T_n(X_1, \dots, X_n)$$

einfacher ein Schätzer

$$T_1 \left(\sum_{i=1}^n X_i \right)$$

von ϑ gesucht werden.

Satz 4.4 (Satz von Rao-Blackwell)

Sei S suffizient für ϑ , und $T(X)$ eine Schätzung von ϑ mit $E_\vartheta[|T(X)|] < \infty$ für alle $\vartheta \in \Theta$. Dann existiert Schätzung $\tilde{T}(S(X))$ mit

$$E_\vartheta[|\tilde{T}(S(X)) - \vartheta|^2] \leq E_\vartheta[|T(X) - \vartheta|^2]$$

für alle $\vartheta \in \Theta$.

Ist dabei $T(X)$ erwartungstreu für ϑ , so kann auch $\tilde{T}(S(X))$ erwartungstreu für ϑ gewählt werden.

Beweis: Setze

$$\tilde{T}(s) = E_\vartheta[T(X)|S(X) = s],$$

was gemäß Lemma 4.3 nicht von ϑ abhängt.

Dann gilt

$$\tilde{T}(S(X)) = E_{\vartheta}[T(X)|S(X)],$$

woraus folgt:

$$\begin{aligned} & E_{\vartheta}[|T(X) - \vartheta|^2] \\ &= E_{\vartheta}[|(T(X) - E_{\vartheta}[T(X)|S(X)] + (E_{\vartheta}[T(X)|S(X)] - \vartheta)|^2] \\ &= E_{\vartheta}[(T(X) - E_{\vartheta}[T(X)|S(X)])^2] + E_{\vartheta}[(E_{\vartheta}[T(X)|S(X)] - \vartheta)^2] \\ \text{da} \\ & E_{\vartheta}[(T(X) - E_{\vartheta}[T(X)|S(X)]) \cdot (E_{\vartheta}[T(X)|S(X)] - \vartheta)] \\ &= E_{\vartheta}[E_{\vartheta}[\dots |S(X)]] \\ &= E_{\vartheta}[(E_{\vartheta}[T(X)|S(X)] - \vartheta) \cdot \underbrace{(E_{\vartheta}[T(X)|S(X)] - E_{\vartheta}[T(X)S(X)])}_{=0}] \\ &= 0, \end{aligned}$$

und damit folgt

$$\begin{aligned} & E_{\vartheta}[|T(X) - \vartheta|^2] \\ &\geq E_{\vartheta}[(E_{\vartheta}[T(X)|S(X)] - \vartheta)^2] \\ &\stackrel{\text{s.ö.}}{=} E_{\vartheta}[(\tilde{T}(S(X)) - \vartheta)^2]. \end{aligned}$$

Ist nun $T(X)$ erwartungstreu für ϑ , so gilt auch

$$\begin{aligned} E_{\vartheta}[\tilde{T}(S(X))] &\stackrel{\text{s.ö.}}{=} E_{\vartheta}[E_{\vartheta}[T(X)|S(X)]] \\ &= E_{\vartheta}[T(X)] \\ &= \vartheta \quad \rightsquigarrow \text{Beh.} \quad \square \end{aligned}$$

Die Bestimmung suffizienter Funktionen $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ermöglicht:

Satz 4.5 (Neyman-Kriterium)

Sei $\{w_{\vartheta} : \vartheta \in \Theta\}$ Familie von W-Maßen auf $(\mathbb{R}^n, \mathcal{B}_n)$. Sei μ σ -endliches Maß auf $(\mathbb{R}^n, \mathcal{B}_n)$. Für $\vartheta \in \Theta$ existiere eine μ -Dichte $f_{\vartheta} : \mathbb{R}^n \rightarrow \mathbb{R}_+$ von w_{ϑ} . Sei $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$ messbar.

Existieren dann

$$g_{\vartheta} : \mathbb{R}^m \rightarrow \mathbb{R}_+ \quad (\vartheta \in \Theta)$$

und

$$r : \mathbb{R}^n \rightarrow \mathbb{R}_+$$

messbar mit

$$\forall \vartheta \in \Theta: \quad f_{\vartheta}(x) = g_{\vartheta}(S(x)) \cdot r(x) \text{ für } \mu - \text{f.a. } x \in \mathbb{R}^n,$$

so ist S suffizient für ϑ .

Bemerkung:

Man kann zeigen: Die Bedingung in Satz 4.5 ist nicht nur hinreichend, sondern auch notwendig für S suffizient für ϑ (vgl. Witting).

Beispiel:

$X = (X_1, \dots, X_n)$ mit X_1, \dots, X_n unabhängig $\mathcal{N}(\vartheta, \sigma_0^2)$ -verteilt für ein $\sigma_0^2 > 0$ fest. Dann hat X bzgl. des LB-Maßes die Dichte

$$\begin{aligned} f_{\vartheta}(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_0}} \cdot \exp\left(-\frac{(x_i - \vartheta)^2}{2\sigma_0^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma_0}}\right)^n \cdot \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma_0^2} + \frac{\sum_{i=1}^n x_i \cdot \vartheta}{\sigma_0^2} - \frac{n \cdot \vartheta^2}{2\sigma_0^2}\right) \\ &= g_{\vartheta}\left(\sum_{i=1}^n x_i\right) \cdot r(x) \end{aligned}$$

mit

$$g_{\vartheta}(u) = \exp\left(\frac{u \cdot \vartheta}{\sigma_0^2} - \frac{n}{2\sigma_0^2} \cdot \vartheta^2\right)$$

und

$$\begin{aligned} r(x) &= \left(\frac{1}{\sqrt{2\pi\sigma_0}}\right)^n \cdot \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma_0^2}\right). \\ &\rightsquigarrow \sum_{i=1}^n X_i \text{ suffizient für } (X_1, \dots, X_n). \end{aligned}$$

Beweis von Satz 4.5:

Definiere

$$\nu(A) = \int_A r(x) \mu(dx),$$

d.h. ν ist Maß auf $(\mathbb{R}^n, \mathcal{B}_n)$. Dann hat w_{ϑ} Dichte

$$h_{\vartheta}(x) = g_{\vartheta}(S(x)) \text{ bzgl. } \nu,$$

da für $A \in \mathcal{B}_n$ gilt:

$$\int_A h_\vartheta(x) \nu(dx) \stackrel{\text{Wahrscheinlichkeitstheorie, Integration bei Maß mit Dichte}}{=} \int_A h_\vartheta(x) \cdot r(x) \mu(dx) \\ \stackrel{\text{Vor.}}{=} \int_A f_\vartheta(x) \mu(dx) \stackrel{f_\vartheta \text{ Dichte von } w_\vartheta}{=} w_\vartheta(A).$$

Wir zeigen nun: Für alle $B \in \mathcal{B}_n$ können wir

$$P_\vartheta[X \in B | S(X)] = E_\vartheta[1_B(X) | S(X)]$$

unabhängig von $\vartheta \in \Theta$ festlegen (wobei $(P_\vartheta)_X = w_\vartheta$) (womit auch jeweils eine von ϑ unabhängige Faktorisierung dieser bed. Wahrscheinlichkeit existiert !)

Dazu setzen wir für $B \in \mathcal{B}_n$

$$k_B = E_\nu(1_B | S^{-1}(\mathcal{B}_m)) : \mathbb{R}^n \rightarrow \mathbb{R},$$

wobei

$$E_p(Z | \mathcal{F})$$

die bedingte Erwartung von $Z : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ bei gegebener σ -Algebra $\mathcal{F} \subseteq \mathcal{A}$ ist, wenn wir das Maß (nicht notwendigermaßen W-Maß) P auf (Ω, \mathcal{A}) zugrunde legen.

Wir zeigen dann, dass die (von ϑ unabhängige(!)) Funktion

$$k_B(X) : \Omega \rightarrow \mathbb{R}$$

eine Version von $E_\vartheta[1_B(X) | S(X)]$ ist für $\vartheta \in \Theta$ beliebig.

Da k_B nach Definition $S^{-1}(\mathcal{B}_m) - \mathcal{B}$ messbar ist, gilt für $A \in \mathcal{B}$:

$$(k_B \circ X)^{-1}(A) = X^{-1}(k_B^{-1}(A)) \\ = X^{-1}(S^{-1}(\bar{A})) = (S \circ X)^{-1}(\bar{A}) \in \mathcal{F}(S(X)),$$

(da $k_B^{-1}(A) \in S^{-1}(\mathcal{B}_m)$, also ex. $\bar{A} \in \mathcal{B}_m$ mit $k_B^{-1}(A) = S^{-1}(\bar{A})$)

d.h. $k_{B \circ X}$ ist $\mathcal{F}(S(X))$ -B messbar.

Weiter gilt für $C \in \mathcal{F}(S(X))$, also für $C = X^{-1}(S^{-1}(E)) = X^{-1}(D)$ für ein $E \in \mathcal{B}_m, D \in S^{-1}(\mathcal{B}_m)$

und beliebiges $\vartheta \in \Theta$:

$$\begin{aligned} \int_C k_B(X) dP_\vartheta &\stackrel{\text{Transformationssatz}}{=} \int_D k_B(x) (P_\vartheta)_X(dx) \\ &= \int_D k_B(x) w_\vartheta \stackrel{f_\vartheta \mu\text{-Dichte von } w_\vartheta}{=} \int_D k_B(x) \cdot f_\vartheta(x) \mu(dx) \\ &\stackrel{\text{Annahme}}{=} \int_D k_B(x) g_\vartheta(S(x)) \cdot r(x) \mu(dx) \\ &\stackrel{\text{Def. } \nu}{=} \int_D k_B(x) \cdot g_\vartheta(S(x)) \nu(dx) \end{aligned}$$

analog zu Integration von W-Maß mit Dichte

$$\begin{aligned} &\stackrel{\text{Def. } k_B}{=} \int_D E_\nu(1_B | S^{-1}(\mathcal{B}_m)) \cdot g_\vartheta \circ S d\nu \\ &\stackrel{g_\vartheta \circ S \text{ ist}}{=} \int_D E_\nu(1_B \cdot g_\vartheta \circ S | S^{-1}(\mathcal{B}_m)) d\nu \end{aligned}$$

da $g_\vartheta \circ S \quad S^{-1}(B_m) - \mathcal{B}$ -messbar.

Def. bedingte Erwartung, $D \in S^{-1}(B_m)$ (s.o.)

$$\begin{aligned} &\stackrel{\text{Def.}}{=} \int_D 1_B \cdot g_\vartheta \circ S d\nu \\ &\stackrel{\text{Def. } \nu, \text{ s.o.}}{=} \int_D 1_B(x) \cdot g_\vartheta(S(x)) \cdot r(x) \mu(dx) \\ &\stackrel{\text{Vor. an } f_\vartheta}{=} \int_D 1_B(x) f_\vartheta(x) \mu(dx) \\ &\stackrel{f_\vartheta \mu\text{-Dichte von } w_\vartheta}{=} \int_D 1_B(x) w_\vartheta(dx) \\ &= \int_D 1_B(x) \cdot (P_\vartheta)_X(dx) \\ &\stackrel{\text{Transformationssatz}}{=} \int_C 1_B(X) dP_\vartheta, \quad \text{w.z.z.w.} \quad \square \end{aligned}$$

5 Statistische Testverfahren

5.1 Einführung

Beispiel 5.1: "Qualitätsprüfung"

Betrachtet wird das Abfüllen von Mineralwasser in Flaschen. Dabei ergibt die zufällige Auswahl von n Flaschen Füllmengen x_1, \dots, x_n . Wie kann man daraus schließen, ob der Sollwert von 0,7 Liter eingehalten wird oder nicht?

Zahlenbeispiel: $n = 100$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0,71$, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0,003$.

Beispiel 5.2: "Marktforschung"

Im vergangenen Jahr betragen die Kosten für einen "Warenkorb" im Durchschnitt 312 Euro. Heutiger Einkauf des "Warenkorbs" in n zufällig ausgewählten Kaufhäusern ergibt x_1, \dots, x_n als zu zahlende Beträge. Wie kann man daraus schließen, ob der Preis gestiegen ist oder nicht?

Zahlenbeispiel: $n = 40$, $\bar{x} = 315$, $s^2 = 120$.

Beispiel 5.3: "Mietspiegel"

In der Stuttgarter Zeitung vom 31.05.2002 wurden $n = 10$ 4-Zimmerwohnungen zu Quadratmeterpreisen 7.52, 6.90, 9.05, 6.60, 7.97, 8.29, 7.48, 10.12, 7.47, 7.45, sowie $m = 5$ 5-6 Zimmerwohnungen zu Quadratmeterpreisen 6.92, 8.94, 9.31, 7.33, 8.13 (jeweils Kaltmiete pro Quadratmeter, in Euro) angeboten.

Kann man daraus schließen, dass sich die Quadratmeterpreise zwischen 4- und 5-6-Zimmerwohnungen unterscheiden?

Hier sind $\bar{x} = 7.89$ und $\bar{y} = 8.13$ die Durchschnittswerte der Preise für 4- bzw. 5-6-Zimmerwohnungen.

Stochastische Modellierung:

Wir fassen die beobachteten Daten als Realisierungen von unabhängigen Zufallsvariablen auf.

In den Beispielen 5.1 und 5.2:

x_1, \dots, x_n seien Realisierungen von u. i. v. ZVen X_1, \dots, X_n ("Einstichprobenproblem").

In Beispiel 5.3:

$x_1, \dots, x_n, y_1, \dots, y_m$ seien Realisierungen von unabhängigen ZVen $X_1, \dots, X_n, Y_1, \dots, Y_m$, wobei X_1, \dots, X_n u. i. v. und Y_1, \dots, Y_m u. i. v. (“**Zweistichprobenproblem**”).

Zwecks Vereinfachung der Problemstellung wird die Klasse der betrachteten Verteilungen eingeschränkt:

$$P_{X_1} \in \{w_\vartheta : \vartheta \in \Theta\} \text{ oder } P_{(X_1, Y_1)} \in \{w_\vartheta : \vartheta \in \Theta\}.$$

Mögliche Verteilungsklassen sind:

- In den Beispielen 5.1 und 5.2:
 - $w_\vartheta = \mathcal{N}(\mu, \sigma_0^2)$ mit $\vartheta = \mu \in \mathbb{R}$ unbekannt und $\sigma_0^2 = s_n^2$ bekannt.
 - $w_\vartheta = \mathcal{N}(\mu, \sigma^2)$ mit $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ unbekannt.
- In Beispiel 5.3:
 - $w_\vartheta = \mathcal{N}(\mu_X, \sigma^2) \otimes \mathcal{N}(\mu_Y, \sigma^2)$ mit $\vartheta = (\mu_X, \mu_Y, \sigma^2) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$ unbekannt. Hier wurde vereinfachend angenommen, dass die Varianz der X_i 's mit der der Y_j 's übereinstimmt.

Wir betrachten eine Aufteilung der Parametermenge in zwei Teile:

$$\Theta = \Theta_0 \cup \Theta_1 \text{ mit } \Theta_0 \cap \Theta_1 = \emptyset, \Theta_0 \neq \emptyset, \Theta_1 \neq \emptyset.$$

Aufgrund der beobachteten Stichprobe $x = (x_1, \dots, x_n)$ (bzw. $x = (x_1, \dots, x_n, y_1, \dots, y_m)$) in Beispiel 5.3) wollen wir uns dann zwischen den beiden Hypothesen

$$H_0 : \vartheta \in \Theta_0 \quad \dots \quad \text{sog. Nullhypothese}$$

bzw.

$$H_1 : \vartheta \in \Theta_1 \quad \dots \quad \text{sog. Alternativhypothese}$$

entscheiden.

In Beispiel 5.1:

Wir setzen $w_\vartheta = \mathcal{N}(\mu, s_n^2)$ mit $\vartheta = \mu \in \Theta = \mathbb{R}$ unbekannt voraus. Wir wollen uns dann entscheiden zwischen

$$H_0 : \vartheta < 0,7 \text{ und } H_1 : \vartheta \geq 0,7$$

d. h. hier ist $\Theta_0 = (-\infty, 0,7)$ und $\Theta_1 = [0,7, \infty)$.

Alternativ könnten wir uns auch zwischen

$$H_0 : \vartheta = 0,7 \text{ und } H_1 : \vartheta \neq 0,7$$

entscheiden, in diesem Falle wäre $\Theta_0 = \{0,7\}$, $\Theta_1 = \mathbb{R} \setminus \{0,7\}$.

Im ersten Fall interessieren Abweichungen von ϑ zu 0,7 nur in eine Richtung, daher spricht man von einem **einseitigen Testproblem**.

Im zweiten Fall interessieren Abweichungen von ϑ zu 0,7 sowohl nach oben als auch nach unten, was als **zweiseitiges Testproblem** bezeichnet wird.

Im Beispiel 5.3 möchte man sich zwischen

$$H_0 : \mu_X = \mu_Y \text{ und } H_1 : \mu_X \neq \mu_Y$$

entscheiden, also ist hier

$$\Theta_0 = \{(\mu_X, \mu_Y, \sigma^2) : \mu_X = \mu_Y\}$$

und

$$\Theta_1 = \{(\mu_X, \mu_Y, \sigma^2) : \mu_X \neq \mu_Y\}.$$

Ein **statistischer Test** ist eine Funktion

$$\varphi : \mathcal{X} \rightarrow [0, 1],$$

wobei \mathcal{X} der Wertebereich der Stichprobe ist ($\mathcal{X} = \mathbb{R}^n$ in den Beispielen 5.1 und 5.2, in Beispiel 5.3 gilt $\mathcal{X} = \mathbb{R}^{n+m}$).

Bei Vorliegen der Stichprobe x (also $x = (x_1, \dots, x_n)$ in den Beispielen 5.1 und 5.2) bedeutet

$\varphi(x) = 1 :$	Entscheidung für H_1
$\varphi(x) = 0 :$	Entscheidung für H_0
$\varphi(x) = p \in (0, 1) :$	Entscheidung für H_1 mit Wahrscheinlichkeit p und Entscheidung für H_0 mit Wahrscheinlichkeit $1 - p$.

Ein mögliches Vorgehen in Beispiel 5.1 zur Entscheidung zwischen

$$H_0 : \vartheta < 0,7 \text{ und } H_1 : \vartheta \geq 0,7$$

ist das folgende:

Schätze zunächst $\vartheta = \text{“Erwartungswert von } w_{\vartheta}\text{”}$ durch

$$T(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

und verwende dann als Test

$$\varphi(x) = \begin{cases} 1 & , \text{ falls } \frac{1}{n} \sum_{i=1}^n x_i > c \\ 0 & , \text{ falls } \frac{1}{n} \sum_{i=1}^n x_i \leq c \end{cases}$$

mit $c \in \mathbb{R}$ geeignet gewählt (und zwar gewählt in Abhängigkeit von $\vartheta_0 = 0.7, n$ und s_n^2).

Die genaue Wahl von c wird später erklärt. Beim obigen Test heißt T **Prüfgröße** oder **Teststatistik**, und c wird als **kritischer Wert** bezeichnet.

Beim Anwenden eines statistischen Test können die folgenden Fälle auftreten:

	Entscheidung für H_0	Entscheidung für H_1
H_0 richtig	richtig	falsch, sog. Fehler erster Art
H_1 richtig	falsch, sog. Fehler zweiter Art	richtig

Gesucht ist dann ein Test, für den die Wahrscheinlichkeiten des Auftretens eines Fehlers erster bzw. zweiter Art (sog. **Fehlerwahrscheinlichkeiten erster bzw. zweiter Art**) möglichst klein sind.

Problem: Im allgemeinen existiert kein Test, der in **allen** Situationen sowohl bzgl. der Fehlerwahrscheinlichen erster als auch zweiter Art optimal ist.

Beispiel: Setze $\varphi_1(x) = 1$ und $\varphi_2(x) = 0$ für alle $x \in \mathcal{X}$. Dann gilt:

Fehlerwahrscheinlichkeit erster Art von $\varphi_1 = 1$,
 Fehlerwahrscheinlichkeit zweiter Art von $\varphi_1 = 0$,

sowie

Fehlerwahrscheinlichkeit erster Art von $\varphi_2 = 0$,
Fehlerwahrscheinlichkeit zweiter Art von $\varphi_2 = 1$.

Ein im obigen Sinne optimaler Test müsste dann bei den Fehlerwahrscheinlichkeiten erster Art mindestens so gut wie φ_2 sein, und bei denen zweiter Art mindestens so gut wie φ_1 sein. Damit müssten bei diesem Test alle Fehlerwahrscheinlichkeiten Null sein, was im allgemeinen unmöglich ist.

Ausweg: Asymmetrische Betrachtungsweise der Entscheidung, z. B. in Beispiel 5.1: Ein Fehler erster Art (d. h. eine Entscheidung für $H_1 : \vartheta \geq 0,7$, obwohl in Wahrheit $H_0 : \vartheta < 0,7$ richtig ist) führt zur Herstellung unvollständig gefüllter Flaschen. Dies wird als schlimmer angesehen als der Fehler zweiter Art (Entscheidung für $H_0 : \vartheta < 0,7$, obwohl $H_1 : \vartheta \geq 0,7$ richtig ist), der zur unnötigen Überprüfung des Abfüllvorgangs führt.

Was man daher macht, ist, eine Schranke für die Fehlerwahrscheinlichkeiten 1. Art vorzugeben, und unter dieser Nebenbedingung die Fehlerwahrscheinlichkeiten 2. Art zu minimieren. (Dies entspricht der Einschränkung der Klasse der betrachteten Schätzverfahren in Abschnitt 4.3).

Genauer wird folgendermaßen vorgegangen:

Definiere die sogenannte **Gütefunktion** eines Test φ durch

$$\begin{aligned}\beta_\varphi &: \Theta \rightarrow [0, 1] \\ \beta_\varphi(\vartheta) &= E_\vartheta \varphi(X)\end{aligned}$$

wobei $X = (X_1, \dots, X_n)$ in den Beispielen 5.1 und 5.2, sowie

$$X = (X_1, \dots, X_n, Y_1, \dots, Y_m)$$

in Beispiel 5.3, und E_ϑ bedeutet, dass der Erwartungswert bei Vorliegen der Verteilung w_ϑ (für X_1 bzw. (X_1, Y_1)) berechnet wird.

$\beta_\varphi(\vartheta)$ gibt dann die Wahrscheinlichkeit für die Entscheidung für H_1 an, falls ϑ der wahre Parameter ist, d. h. falls w_ϑ die zugrundeliegende Verteilung ist. Dies folgt aus

$$P_\vartheta[\text{„Annahme von } H_1\text{“}] = E_\vartheta \left[\underbrace{P_\vartheta[\text{„Annahme von } H_1\text{“} | X]}_{\stackrel{\text{Def.}}{=} \varphi(X)} \right].$$

Damit ist

$$\beta_\varphi(\vartheta) = \text{Wahrscheinlichkeit für Fehler 1. Art, falls } \vartheta \in \Theta_0$$

und

$$1 - \beta_\varphi(\vartheta) = \text{Wahrscheinlichkeit für Fehler 2. Art, falls } \vartheta \in \Theta_1.$$

Definition 5.1: Sei $\alpha \in [0, 1]$.

a) φ heißt **Test zum Niveaus** α , falls gilt:

$$\beta_\varphi(\vartheta) \leq \alpha \quad \text{für alle } \vartheta \in \Theta_0$$

(d. h., falls die Fehlerwahrscheinlichkeiten erster Art immer kleiner oder gleich α sind).

b) φ heißt **gleichmäßig bester Test zum Niveau** α , falls φ Test zum Niveau α ist und falls für alle Test $\bar{\varphi}$ zum Niveau α gilt:

$$\underbrace{1 - \beta_\varphi(\vartheta)}_{\text{Fehlerwahrscheinlichkeit 2. Art von } \varphi \text{ für den Parameter } \vartheta} \leq \underbrace{1 - \beta_{\bar{\varphi}}(\vartheta)}_{\text{Fehlerwahrscheinlichkeit 2. Art von } \bar{\varphi} \text{ für den Parameter } \vartheta} \quad \text{für alle } \vartheta \in \Theta_1,$$

d. h. falls für alle $\vartheta \in \Theta_2$ gilt:

$$\beta_\varphi(\vartheta) \geq \beta_{\bar{\varphi}}(\vartheta).$$

In Beispiel 5.1:

X_1, \dots, X_n u. i. v., $P_{X_1} = \mathcal{N}(\mu, \sigma_0^2)$ mit $\sigma_0^2 = s_n^2$ gegeben.

$H_0 : \mu < 0.7, H_1 : \mu \geq 0.7$

Test ist

$$\varphi(X) = \begin{cases} 1 & , \quad \text{falls } \frac{1}{n} \sum_{i=1}^n X_i > c \\ 0 & , \quad \text{falls } \frac{1}{n} \sum_{i=1}^n X_i \leq c \end{cases}$$

mit $c \in \mathbb{R}$ fest. Wir berechnen nun die Gütefunktion dieses Tests:

$$\begin{aligned} \beta_\varphi(\mu) &= E_\mu[\varphi(x)] \\ &= P_\mu \left[\frac{1}{n} \sum_{i=1}^n X_i > c \right] \\ &= P_\mu \left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu) > (c - \mu) \right] \\ &= P_\mu \left[\frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (X_i - \mu) > \sqrt{n} \frac{c - \mu}{\sigma_0} \right]. \end{aligned}$$

Die Zufallsvariable auf der linken Seite im Innern der Wahrscheinlichkeit hat Erwartungswert Null und Varianz Eins. Da Linearkombinationen unabhängiger normalverteilter Zufallsvariablen wieder normalverteilt sind, ist sie folglich $\mathcal{N}(0, 1)$ -verteilt, also gilt:

$$\beta_\varphi(\mu) = 1 - \Phi\left(\sqrt{n} \frac{c - \mu}{\sigma_0}\right) = \Phi\left(\sqrt{n} \frac{\mu - c}{\sigma_0}\right),$$

wobei Φ die Verteilungsfunktion zu $\mathcal{N}(0, 1)$ ist, die die Symmetrieeigenschaft

$$1 - \Phi(x) = \Phi(-x) \quad (x \in \mathbb{R})$$

hat.

φ ist Test zum Niveau α , falls gilt

$$\beta_\varphi(\mu) \leq \alpha \text{ für alle } \mu \leq \mu_0 = 0,7,$$

was hier äquivalent ist zu

$$1 - \Phi\left(\sqrt{n} \frac{c - \mu}{\sigma_0}\right) \leq \alpha \text{ für alle } \mu \leq \mu_0,$$

bzw.

$$\Phi\left(\sqrt{n} \frac{c - \mu}{\sigma_0}\right) \geq 1 - \alpha \text{ für alle } \mu \leq \mu_0,$$

d. h. (wegen Monotonie von Φ):

$$\Phi\left(\sqrt{n} \frac{c - \mu_0}{\sigma_0}\right) \geq 1 - \alpha.$$

Die Fehlerwahrscheinlichkeiten 2. Art

$$1 - \beta_\varphi(\mu) = \Phi\left(\sqrt{n} \frac{c - \mu}{\sigma_0}\right) \quad (\text{für } \mu > \mu_0)$$

werden umso kleiner, je **kleiner** c ist.

Daher ist im Hinblick auf die Einhaltung des Niveaus und der Minimierung der Fehlerwahrscheinlichkeiten 2. Art die folgende Wahl von c naheliegend:

Wähle c so, dass gilt:

$$\Phi\left(\sqrt{n} \frac{c - \mu_0}{\sigma_0}\right) = 1 - \alpha,$$

d. h. das gilt

$$\sqrt{n} \frac{c - \mu_0}{\sigma_0} = u_\alpha,$$

wobei u_α das sogenannte α -Fraktile zu $\mathcal{N}(0, 1)$ ist, d. h. $\Phi(u_\alpha) = 1 - \alpha$.

Dies führt auf den sogenannten **einseitigen Gauß-Test**

$$\begin{aligned} \varphi(x) &= \begin{cases} 1 & , \text{ falls } \quad \frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (x_i - \mu_0) > u_\alpha \\ 0 & , \text{ sonst,} \end{cases} \\ &= \begin{cases} 1 & , \text{ falls } \quad \frac{1}{n} \sum_{i=1}^n x_i > \mu_0 + \frac{\sigma_0}{\sqrt{n}} \cdot u_\alpha \\ 0 & , \text{ sonst.} \end{cases} \end{aligned}$$

Nach Konstruktion ist φ Test zum Niveau α .

Wir werden im nächsten Abschnitt zeigen:

φ ist gleichmäßig bester Test zum Niveau α .

Bemerkung:

- a) Aus historischen Gründen wählt man für das Niveau meist $\alpha \in \{0.05, 0.01, 0.1\}$.
- b) Für die Verteilungsfunktion Φ von $\mathcal{N}(0, 1)$ gilt:

x	1.28	1.64	1.96	2.33
$\Phi(x)$	0.90	0.95	0.975	0.99.

Damit gilt für die α -Fraktile von $\mathcal{N}(0, 1)$:

α	0.1	0.05	0.01
u_α	1.28	1.64	2.33

Anwendung des einseitigen Gauß-Tests in Beispiel 5.1

Zu testen ist

$$H_0 : \mu \leq \mu_0 = 0.7 \quad \text{versus} \quad H_1 : \mu > 0.7.$$

Gegeben ist

$$n = 100, \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 0.71$$

sowie

$$\sigma_0^2 \stackrel{(!)}{=} s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.003.$$

Wir wählen $\alpha = 0.05$, also ist $u_\alpha = 1.64$.

Wegen

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{1}{\sigma_0} \sum_{i=1}^n (x_i - \mu_0) &= \frac{\sqrt{n}}{\sigma_0} \cdot (\bar{x} - \mu_0) \\ &= \frac{\sqrt{100}}{\sqrt{0.003}} (0.71 - 0.70) \\ &\approx 1.83 > u_\alpha \end{aligned}$$

kann hier H_0 zum Niveau $\alpha = 0,05$ abgelehnt werden.

Anwendung des einseitigen Gauß-Tests in Beispiel 5.2

Zu testen ist hier

$$H_0 : \mu \leq \mu_0 = 312 \text{ versus } H_1 : \mu > 312.$$

Gegeben ist $n = 40, \bar{x} = 315$ und $\sigma_0^2 \stackrel{(!)}{=} s_n^2 = 120$.

Wegen

$$\begin{aligned} \frac{1}{\sqrt{n}} \frac{1}{\sigma_0} \sum_{i=1}^n (x_i - \sigma_0) &= \frac{\sqrt{n}}{\sqrt{\sigma_0}} (\bar{x} - \mu_0) \\ &= \frac{\sqrt{40}}{\sqrt{120}} (315 - 312) \\ &= 1.73 > u_\alpha \end{aligned}$$

kann hier ebenfalls H_0 zum Niveau $\alpha = 0,05$ abgelehnt werden.

5.2 Das Fundamentallemma von Neyman und Pearson

Zur Konstruktion von Tests ist der folgende Begriff hilfreich.

Def. 5.2: Sei Q ein W-Maß auf \mathcal{B} mit Verteilungsfunktion $F : \mathbb{R} \rightarrow [0, 1]$ (d. h. $F(t) = Q((-\infty, t])$), und sei $\alpha \in (0, 1)$.

Dann heißt

$$Q_\alpha = \min\{t \in \mathbb{R} : F(t) \geq 1 - \alpha\}$$

α -Fraktile von Q .

Bemerkung: Wegen $F(t) \rightarrow 1$ ($t \rightarrow \infty$) und $F(t) \rightarrow 0$ ($t \rightarrow -\infty$) ist die Menge oben nicht leer und nach unten beschränkt. Das Minimum existiert wegen rechtsseitiger Stetigkeit von F .

Bemerkung: Es gilt $F(Q_\alpha) \geq 1 - \alpha$ und

$$F(Q_\alpha-) = \lim_{\substack{x \rightarrow Q_\alpha, \\ x < Q_\alpha}} F(x) \leq 1 - \alpha.$$

Ist daher X Zufallsvariable mit Verteilungsfunktion F , so gilt

$$\begin{aligned} P[X > Q_\alpha] &= 1 - F(Q_\alpha) \leq 1 - (1 - \alpha) = \alpha, \\ P[X \geq Q_\alpha] &= 1 - P[X < Q_\alpha] = 1 - F(Q_\alpha-) \geq 1 - (1 - \alpha) = \alpha, \\ \text{also: } P[X > Q_\alpha] &\leq \alpha \leq P[X \geq Q_\alpha] \end{aligned}$$

Im Folgenden betrachten wir das Testproblem:
Gegeben sei Realisierung x einer Zufallsvariablen

$$X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}_n)$$

mit

$$P_X = w_\vartheta \text{ für ein } \vartheta \in \Theta = \{\vartheta_0, \vartheta_1\}.$$

Die Verteilung w_ϑ besitze eine Dichte f_ϑ bzgl. eines Maßes μ auf $(\mathbb{R}^n, \mathcal{B}_n)$ (für $\vartheta \in \Theta$).

Zu testen sei

$$H_0 : \vartheta = \vartheta_0 \text{ versus } H_1 : \vartheta = \vartheta_1$$

(also $H_i : \vartheta \in \Theta_i$ mit $\Theta_0 = \{\vartheta_0\}$, $\Theta_1 = \{\vartheta_1\}$).

Satz 5.1: Fundamentallemma von Neyman und Pearson.

Betrachtet wird das obige Testproblem. Sei $\alpha \in (0, 1)$ beliebig.

a) Ein Test φ mit

$$E_{\vartheta_0}[\varphi(x)] = \alpha$$

ist gleichmäßig bester Test zum Niveau α genau dann, wenn für ein $k^* \in \mathbb{R}_+$ gilt:

$$(5.1) \quad \varphi(x) = \begin{cases} 1 & \text{für } x \in \mathbb{R}^n \text{ mit } f_{\vartheta_1}(x) > k^* \cdot f_{\vartheta_0}(x) \\ 0 & \text{für } x \in \mathbb{R}^n \text{ mit } f_{\vartheta_1}(x) < k^* \cdot f_{\vartheta_0}(x) \end{cases}$$

für μ -fast alle $x \in \mathbb{R}^n$.

- b) Es gibt einen gleichmäßig besten Test φ^* zum Niveau α . Dieser lässt sich folgendermaßen konstruieren:

Setze

$$T(x) = \frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} \quad (\text{mit } \frac{a}{0} = \infty \text{ für } a > 0, \frac{0}{0} = 0).$$

Wähle k^* als α -Fraktile der Verteilung von $T(X)$ bei wahren Parameter ϑ_0 und wähle $\gamma^* \in [0, 1]$ so, dass gilt:

$$(5.2) \quad P_{\vartheta_0}[T(X) > k^*] + \gamma^* \cdot P_{\vartheta_0}[T(X) = k^*] = \alpha.$$

Setze dann

$$\varphi^*(x) = \begin{cases} 1 & \text{falls } T(x) > k^*, \\ \gamma^* & \text{falls } T(x) = k^*, \\ 0 & \text{falls } T(x) < k^*. \end{cases}$$

Bemerkung:

- a) Wegen

$$\begin{aligned} P_{\vartheta_0}[T(x) > k^*] \leq \alpha &\leq P_{\vartheta_0}[T(x) \geq k^*] \\ &= P_{\vartheta_0}[T(X) > k^*] + 1 \cdot P_{\vartheta_0}[T(X) = k^*] \end{aligned}$$

existiert immer ein $\gamma^* \in [0, 1]$, für das (5.2) gilt.

- b) Wegen

$$\begin{aligned} E_{\vartheta_0}[\varphi^*(X)] &= 1 \cdot P_{\vartheta_0}[T(X) > k^*] + \gamma^* \cdot P_{\vartheta_0}[T(X) = k^*] \\ &\stackrel{(5.2)}{=} \alpha \end{aligned}$$

schöpft φ^* das Niveau an der Stelle $\vartheta = \vartheta_0$ voll aus.

Beweis von Satz 5.1

- a₁) **Wir zeigen:** Die Bedingung (5.1) ist hinreichend.

Sei dazu φ von der Form (5.1) mit $E_{\vartheta_0}[\varphi(X)] = \alpha$. Da die Werte von φ auf μ -Nullmengen sich nicht auf $E_{\vartheta_0}\varphi(X)$ auswirken, können wir dann oBdA sogar voraussetzen:

$$\varphi(x) = \begin{cases} 1 & \text{für } x \in \mathbb{R}^n \text{ mit } f_{\vartheta_1}(x) > k^* \cdot f_{\vartheta_0}(x), \\ \gamma(x) \in [0, 1] & \text{für } x \in \mathbb{R}^n \text{ mit } f_{\vartheta_1}(x) = k^* \cdot f_{\vartheta_0}(x), \\ 0 & \text{für } x \in \mathbb{R}^n \text{ mit } f_{\vartheta_1}(x) < k^* \cdot f_{\vartheta_0}(x) \end{cases}$$

für **alle** $x \in \mathbb{R}^n$.

Sei $\bar{\varphi}$ ein beliebiger Test zum Niveau α .

Dann gilt

$$E_{\vartheta_0}[\bar{\varphi}(X)] \leq \alpha = E_{\vartheta}[\varphi(X)].$$

Zu zeigen ist: Die Fehlerwahrscheinlichkeit 2. Art von $\bar{\varphi}$ an der Stelle ϑ_1 ist größer oder gleich als die entsprechende Fehlerwahrscheinlichkeit von φ , d. h. es gilt

$$1 - E_{\vartheta_1}[\bar{\varphi}(X)] \geq 1 - E_{\vartheta_1}[\varphi(X)],$$

was äquivalent ist zu

$$E_{\vartheta_1}[\bar{\varphi}(X)] \leq E_{\vartheta_1}[\varphi(X)].$$

Um dies zu zeigen, beachten wir, dass für alle $x \in \mathbb{R}^n$ gilt:

$$0 \leq (\varphi(x) - \bar{\varphi}(x)) \cdot (f_{\vartheta_1}(x) - k^* \cdot f_{\vartheta_0}(x))$$

(denn ist $f_{\vartheta_1}(x) - k^* \cdot f_{\vartheta_0}(x) > 0$, so gilt

$$\varphi(x) - \bar{\varphi}(x) = 1 - \bar{\varphi}(x) \geq 0,$$

ist dagegen $f_{\vartheta_1}(x) - k^* \cdot f_{\vartheta_0}(x) < 0$, so gilt

$$\varphi(x) - \bar{\varphi}(x) = 0 - \bar{\varphi}(x) \leq 0,$$

also ist das Produkt in beiden Fällen größer oder gleich Null).

Daraus folgt:

$$\begin{aligned} 0 &\leq \int (\varphi(x) - \bar{\varphi}(x)) \cdot (f_{\vartheta_1}(x) - k^* \cdot f_{\vartheta_0}(x)) \mu(dx) \\ &= \int \varphi(x) \cdot f_{\vartheta_1}(x) \mu(dx) - \int \bar{\varphi}(x) \cdot f_{\vartheta_1}(x) \mu(dx) \\ &\quad - k^* \cdot (\int \varphi(x) \cdot f_{\vartheta_0}(x) \mu(dx) - \int \bar{\varphi}(x) \cdot f_{\vartheta_0}(x) \mu(dx)) \\ &= E_{\vartheta_1}[\varphi(X)] - E_{\vartheta_1}[\bar{\varphi}(X)] - k^* \cdot (E_{\vartheta_0}[\varphi(X)] - E_{\vartheta_0}[\bar{\varphi}(X)]) \\ &= E_{\vartheta_1}[\varphi(X)] - E_{\vartheta_1}[\bar{\varphi}(X)] - k^* \cdot [\underbrace{\alpha - E_{\vartheta_0}[\bar{\varphi}(X)]}_{\leq \alpha}] \\ &\leq E_{\vartheta_1}[\varphi(X)] - E_{\vartheta_1}[\bar{\varphi}(X)], \end{aligned}$$

woraus die Behauptung von a_1) folgt.

b) Da – wie man sich leicht klar macht – gilt

$$\begin{aligned} - f_{\vartheta_1}(x) > k^* \cdot f_{\vartheta_0}(x) &\Rightarrow \frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} > k^* \\ - f_{\vartheta_1}(x) < k^* \cdot f_{\vartheta_0}(x) &\Rightarrow \frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} < k^* \end{aligned}$$

hat φ^* die Bauart (5.1). Weiter gilt nach Wahl von k^* und γ^* :

$$E_{\vartheta_0}[\varphi^*(X)] = P_{\vartheta_0}[T(X) > k^*] + \gamma^* \cdot P_{\vartheta_0}[T(X) = k^*] = \alpha.$$

Mit $a_1)$ folgt daher, dass der in b) konstruierte Test ein gleichmäßig bester Test zum Niveau α ist.

$a_2)$ **Wir zeigen:** Die Bedingung (5.1) ist notwendig.

Sei dazu φ ein gleichmäßig bester Test zum Niveau α mit

$$E_{\vartheta_0}[\varphi(X)] = \alpha$$

und sei φ^* der Test aus b).

Dann gilt:

$$E_{\vartheta_0}[\varphi(X)] = \alpha = E_{\vartheta_0}[\varphi^*(X)]$$

sowie – da beide Tests gleichmäßig beste Tests zum Niveau α sind –

$$E_{\vartheta_1}[\varphi(X)] = E_{\vartheta_1}[\varphi^*(X)].$$

Daraus folgt

$$\begin{aligned} & \int (\varphi^*(x) - \varphi(x)) \cdot (f_{\vartheta_1}(x) - k^* \cdot f_{\vartheta_0}(x)) \mu(dx) \\ & \stackrel{\text{vgl. } a_1)}{=} E_{\vartheta_1}[\varphi^*(X)] - E_{\vartheta_1}[\varphi(X)] - k^* \cdot (E_{\vartheta_0}[\varphi^*(X)] - E_{\vartheta_0}[\varphi(X)]) \\ & \stackrel{\text{s.o.}}{=} 0. \end{aligned}$$

Nach $a_1)$ ist der Integrand nicht negativ. Also gilt für μ -f.a. x

$$(\varphi^*(x) - \varphi(x)) \cdot (f_{\vartheta_1}(x) - k^* \cdot f_{\vartheta_0}(x)) = 0,$$

woraus für μ -f.a. x folgt:

$$\varphi(x) = \varphi^*(x) \text{ falls } f_{\vartheta_1}(x) \neq k^* \cdot f_{\vartheta_0}(x) \quad \square$$

Im Folgenden: Nachweis der Optimalität des einseitigen Gauß-Tests mit Hilfe von Satz 5.1

X_1, \dots, X_n seien u. i. v. ZVen mit $P_{X_1} = \mathcal{N}(\mu, \sigma_0^2)$.

Für $\sigma_0^2 > 0$ bekannt und $\mu_0 \in \mathbb{R}$ fest sei zu testen

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0.$$

Niveau sei $\alpha \in (0, 1)$.

Einseitiger Gauß-Test:

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & \text{falls } T_n(x_1, \dots, x_n) > u_\alpha, \\ 0 & \text{sonst,} \end{cases}$$

wobei

$$T_n(x_1, \dots, x_n) = \frac{1}{\sqrt{n} \cdot \sigma_0} \sum_{i=1}^n (X_i - \mu_0)$$

und

$$u_\alpha = \alpha - \text{Fraktil von } \mathcal{N}(0, 1),$$

d. h. $\Phi(u_\alpha) = 1 - \alpha$ mit $\Phi =$ Verteilungsfunktion zu $\mathcal{N}(0, 1)$.

Satz 5.2

In der obigen Situation ist der einseitige Gauß-Test ein gleichmäßig bester Test zum Niveau α .

Beweis:

1. Schritt: Wir zeigen φ ist Test zum Niveau α .

Dazu: Ist $\mu \leq \mu_0$, so gilt

$$\begin{aligned} \beta_\varphi(\mu) &= P_\mu \left[\frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (X_i - \mu_0) > u_\alpha \right] \\ &= P_\mu \left[\frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (X_i - \mu) > u_\alpha + \underbrace{\frac{\sqrt{n}}{\sigma_0}(\mu_0 - \mu)}_{>0} \right] \\ &\leq P_\mu \left[\frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (X_i - \mu) > u_\alpha \right] \\ &= \alpha, \end{aligned}$$

da $\frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (X_i - \mu)$ bei Vorliegen des Parameters μ standard-normalverteilt ist, und u_α das α -Fraktil zu $\mathcal{N}(0, 1)$ ist.

Insbesondere gilt

$$\beta_\varphi(\mu_0) \stackrel{s.o.}{=} P_{\mu_0} \left[\underbrace{\frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (X_i - \mu_0)}_{\mathcal{N}(0,1)\text{-verteilt}} > \mu_\alpha \right] = \alpha,$$

also schöpft φ das Niveau an der Stelle μ_0 voll aus.

2. Schritt: Sei $\mu_1 > \mu_0$ beliebig.

Wir setzen $X = (X_1, \dots, X_n)$ und betrachten das Problem, bei beobachtetem Wert von X

$$\overline{H}_0 : \mu = \mu_0 \text{ versus } \overline{H}_1 : \mu = \mu_1$$

zu testen.

Wir zeigen: φ ist gleichmäßig bester Test zum Niveau α für das obige Testproblem.

Dazu:

φ erfüllt

$$E_{\mu_0} \varphi(X) \stackrel{s.o.}{=} \alpha.$$

X hat bzgl. des LB-Maßes auf \mathbb{R}^n die Dichte

$$\begin{aligned} f_{\mu}(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_0} \cdot \exp\left(-\frac{(x_i - \mu)^2}{2\sigma_0^2}\right) \\ &= \frac{1}{(2\pi)^{n/2}\sigma_0^n} \cdot \exp\left(-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma_0^2\right). \end{aligned}$$

Für diese gilt:

$$f_{\mu_1}(x_1, \dots, x_n) > k^* \cdot f_{\mu_0}(x_1, \dots, x_n)$$

$$\begin{aligned} \Leftrightarrow \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu_1)^2}{2\sigma_0^2} + \frac{\sum_{i=1}^n (x_i - \mu_0)^2}{2\sigma_0^2}\right) &> k^* \\ \Leftrightarrow 2(\mu_1 - \mu_0) \sum_{i=1}^n x_i - n \cdot (\mu_1^2 - \mu_0^2) &> 2\sigma_0^2 \cdot \ln(k^*) \\ \Leftrightarrow \frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (x_i - \mu_0) &> \left(\frac{2\sigma_0^2 \cdot \ln(k^*) + n \cdot (\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)} - n \cdot \mu_0\right) \frac{1}{\sqrt{n}\sigma_0} \end{aligned}$$

und analog

$$\begin{aligned} f_{\mu_1}(x_1, \dots, x_n) &< k^* \cdot f_{\mu_0}(x_1, \dots, x_n) \\ \Leftrightarrow \frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (x_i - \mu_0) &< \left(\frac{2\sigma_0^2 \cdot \ln(k^*) + n \cdot (\mu_1^2 - \mu_0^2)}{2(\mu_1 - \mu_0)} - n \cdot \mu_0\right) \frac{1}{\sqrt{n}\sigma_0}. \end{aligned}$$

Daher hat φ die Bauart (5.1) aus Satz 5.1, und die Behauptung folgt aus Satz 5.1.

3. Schritt: Abschluss des Beweises.

Sei $\bar{\varphi}$ beliebiger Test zum Niveau α für

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0.$$

Sei $\mu_1 > \mu_0$ beliebig.

Wir zeigen: $\beta_{\varphi}(\mu_1) \geq \beta_{\bar{\varphi}}(\mu_1)$.

Dazu fassen wir φ und $\bar{\varphi}$ beide als Test zum Niveau α für

$$\bar{H}_0 : \mu = \mu_0 \text{ versus } \bar{H}_1 : \mu = \mu_1$$

auf. Dann ist φ nach Schritt 2 gleichmäßig bester Test für dieses Testproblem zum Niveau α , und da auch $\bar{\varphi}$ Test zum Niveau α für dieses Testproblem ist, folgt unmittelbar die Behauptung. \square

5.3 Tests bei monotonen Dichtequotienten

Entscheidend im Beweis von Satz 5.2 war, dass

$$\frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} > k$$

äquivalent war zu

$$T(x) > k^*$$

für ein nicht von $\vartheta_0 < \vartheta_1$ abhängendes T .

Existiert eine streng monoton wachsende Funktion $g_{\vartheta_0, \vartheta_1}$ mit

$$\frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} = g_{\vartheta_0, \vartheta_1}(T(x)),$$

so ist diese Äquivalenz immer gültig, falls k im Bild von $g_{\vartheta_0, \vartheta_1} \circ T$ liegt.

Def. 5.3. Eine Klasse $\{w_{\vartheta} : \vartheta \in \Theta\}$ von W-Maßen auf $(\mathbb{R}, \mathcal{B}_n)$ mit $\Theta \subseteq \mathbb{R}$ heißt **Klasse mit monotonen Dichtequotienten** $T : (\mathbb{R}^n, \mathcal{B}_n) \rightarrow (\mathbb{R}, \mathcal{B})$, falls gilt:

- (1) $w_{\vartheta_0} \neq w_{\vartheta_1}$ für $\vartheta_0 \neq \vartheta_1$ ("eindeutige Parametrisierung").
- (2) Es existiert ein σ -endliches Maß μ und μ -Dichten $f_{\vartheta} : \mathbb{R}^n \rightarrow \mathbb{R}$ von w_{ϑ} so, dass für alle $\vartheta_0, \vartheta_1 \in \Theta$ mit $\vartheta_0 < \vartheta_1$ eine streng monoton wachsende Funktion $g_{\vartheta_0, \vartheta_1} : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ existiert, für die gilt:

$$\frac{f_{\vartheta_1}(x)}{f_{\vartheta_0}(x)} = g_{\vartheta_0, \vartheta_1}(T(x)) \text{ für } (w_{\vartheta_0} + w_{\vartheta_1}) - \text{fast alle } x \in \mathbb{R}^n.$$

(Hierbei oBdA $\frac{0}{0} = 0$, da $\{x \in \mathbb{R}^n : f_{\vartheta_1}(x) = 0 = f_{\vartheta_0}(x)\}$ eine $(w_{\vartheta_0} + w_{\vartheta_1})$ -Nullmenge ist.)

Beispiel 5.4 Die Zufallsvariablen X_1, \dots, X_n seien u. i. v. mit $P_{X_1} = b(1, \vartheta)$, $\vartheta \in (0, 1) =: \Theta$.

Dann hat $X = (X_1, \dots, X_n)$ bzgl. dem abzählenden Maß μ auf $\{0, 1\}^n$ (eingebettet in \mathbb{R}^n , d. h.

$$\mu(A) = |A \cap \{0, 1\}^n|$$

für $A \in \mathcal{B}_n$) die Dichte

$$f_{\vartheta}(x_1, \dots, x_n) = \prod_{i=1}^n \vartheta^{x_i} (1 - \vartheta)^{1-x_i} = \vartheta^{\sum_{i=1}^n x_i} (1 - \vartheta)^{n - \sum_{i=1}^n x_i}.$$

Für $0 < \vartheta_0 < \vartheta_1 < 1$ gilt

$$\frac{f_{\vartheta_1}(x_1, \dots, x_n)}{f_{\vartheta_0}(x_1, \dots, x_n)} = \left(\frac{1 - \vartheta_1}{1 - \vartheta_0} \right)^n \cdot \left(\frac{\vartheta_1}{\vartheta_0} \cdot \frac{1 - \vartheta_0}{1 - \vartheta_1} \right)^{\sum_{i=1}^n x_i}.$$

Wegen

$$\frac{\vartheta_1}{\vartheta_0} \cdot \frac{1 - \vartheta_0}{1 - \vartheta_1} > 1 \cdot 1 = 1$$

ist

$$g_{\vartheta_0, \vartheta_1}(u) = \left(\frac{1 - \vartheta_1}{1 - \vartheta_0} \right)^n \cdot \left(\frac{\vartheta_1}{\vartheta_0} \cdot \frac{1 - \vartheta_0}{1 - \vartheta_1} \right)^u$$

streng monoton wachsend (in u), und daher ist $\{w_{\vartheta} = \bigotimes_{i=1}^n b(1, \vartheta) : \vartheta \in \Theta\}$ eine

Klasse mit monotonen Dichtequotienten in $T(x) = \sum_{i=1}^n x_i$, $x = (x_1, \dots, x_n)$.

Beispiel 5.5 Die Zufallsvariablen X_1, \dots, X_n seien u. i. v. mit $P_{X_1} = \mathcal{N}(\mu, \sigma_0^2)$, $\mu \in \mathbb{R} =: \Theta$. Dann hat $X = (X_1, \dots, X_n)$ bzgl. dem LB-Maß die Dichte

$$\begin{aligned} f_{\mu}(x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma_0} \cdot \exp\left(-\frac{(x_i - \mu)^2}{2\sigma_0^2}\right) \\ &= \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{\sigma_0^n} \cdot \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma_0^2}\right). \end{aligned}$$

Für $\mu_0 < \mu_1$ gilt

$$\begin{aligned} \frac{f_{\mu_1}(x_1, \dots, x_n)}{f_{\mu_0}(x_1, \dots, x_n)} &= \exp\left(-\left(\sum_{i=1}^n (x_i - \mu_1)^2 - \sum_{i=1}^n (x_i - \mu_0)^2\right)/(2\sigma_0^2)\right) \\ &= \exp(-n \cdot (\mu_1^2 - \mu_0^2)/(2\sigma_0^2)) \cdot \exp\left(2(\mu_1 - \mu_0) \left(\sum_{i=1}^n x_i\right)/(2\sigma_0^2)\right), \end{aligned}$$

also ist $\{w_\mu = \bigotimes_{i=1}^n \mathcal{N}(\mu, \sigma_0^2) : \mu \in \Theta\}$ Klasse mit monotonen Dichtequotienten in

$$T(x_1, \dots, x_n) = \sum_{i=1}^n x_i.$$

Bemerkung: Ist $\{w_\vartheta : \vartheta \in \Theta\}$ Klasse mit monotonen Dichtequotienten in T , und ist $h : (\overline{\mathbb{R}}, \overline{\mathcal{B}}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ streng monoton wachsend und bijektiv, dann ist $\{w_\vartheta : \vartheta \in \Theta\}$ auch Klasse mit monotonen Dichtequotienten in $h \circ T$. Dies folgt, indem man in Definition 5.3 $g_{\vartheta_0, \vartheta_1}$ durch $g_{\vartheta_0, \vartheta_1} \circ h^{-1}$ ersetzt.

Satz 5.3 (Optimale Tests bei monotonen Dichtequotienten)

Zu testen sei

$$H_0 : \vartheta \leq \vartheta_0 \text{ versus } H_1 : \vartheta > \vartheta_0$$

ausgehend von einer Realisierung einer Zufallsvariable X , für die gilt $P_X = w_\vartheta$ für ein $\vartheta \in \Theta$. Das Niveau sei $\alpha \in (0, 1)$. $\{w_\vartheta : \vartheta \in \Theta\}$ sei Klasse mit monotonen Dichtequotienten in T .

Dann gilt:

Der Test

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

wobei c und γ^* so gewählt sind, dass $P_{\vartheta_0}[T(X) > c] + \gamma^* \cdot P_{\vartheta_0}[T(X) = c] = \alpha$ erfüllt ist, ist ein gleichmäßig bester Test zum Niveau α .

Bemerkung: Satz 5.3 impliziert Satz 5.2. Um dies zu sehen, wähle man $\gamma^* = 0$ und $c = u_\alpha$ und beachte, dass $\bigotimes_{i=1}^n \mathcal{N}(\mu, \sigma_0^2)$ Klasse mit monotonen Dichtequotienten

in $T(x) = \sum_{i=1}^n x_i$ bzw.

$$\overline{T}(x) = \frac{1}{\sqrt{n}\sigma_0} \sum_{i=1}^n (x_i - \mu_0)$$

ist.

Beispiel 5.6: In einem Krankenhaus wurden im Jahr 1999 374 Mädchen und 396 Jungen geboren. Kann man daraus zum Niveau $\alpha = 5\%$ schließen, dass mehr Jungen als Mädchen geboren werden?

Hier ist

$$X = (X_1, \dots, X_n)$$

mit $P_{X_1} = b(1, \vartheta), X_1, \dots, X_n$ u. i. v., ϑ =Wahrscheinlichkeit für "Jungengeburt" und $n = 374 + 396 = 770$.

Ausgehend von einer Realisierung

$$x = (x_1, \dots, x_n)$$

mit $\sum_{i=1}^n x_i = 396$ wollen wir uns zwischen

$$H_0 : \vartheta \leq 0,5 \text{ und } H_1 : \vartheta > 0,5$$

entscheiden. Dabei ist das Niveau als $\alpha = 0,05$ vorgegeben.

Die Klasse $\{w_\vartheta : \vartheta \in \Theta\}$ mit $w_\vartheta = \bigotimes_{i=1}^n b(1, \vartheta)$ ist nach Beispiel 5.4 Klasse mit monotonen Dichtequotienten in

$$T(x_1, \dots, x_n) = \sum_{i=1}^n x_i.$$

Nach Satz 5.3 ist daher ein gleichmäßig bester Test zum Niveau α für dieses Testproblem gegeben durch

$$\varphi^*(x_1, \dots, x_n) = \begin{cases} 1 & \text{falls } x_1 + \dots + x_n > c, \\ \gamma^* & \text{falls } x_1 + \dots + x_n = c, \\ 0 & \text{falls } x_1 + \dots + x_n < c, \end{cases}$$

wobei c und γ^* so gewählt werden, dass gilt:

$$\alpha \stackrel{!}{=} P_{\vartheta=0,5}[X_1 + \dots + X_n > c] + \gamma^* \cdot P_{\vartheta=0,5}[X_1 + \dots + X_n = c].$$

Da $X_1 + \dots + X_n$ $b(n, \vartheta)$ -verteilt ist, ist die rechte Seite (für $c \in \mathbb{N}_0$) gleich

$$\begin{aligned} & \sum_{k=c+1}^n \binom{n}{k} \cdot (0,5)^k \cdot (1 - 0,5)^{n-k} + \gamma^* \cdot \binom{n}{c} \cdot (0,5)^c \cdot (1 - 0,5)^{n-c} \\ &= \sum_{k=c+1}^n \binom{n}{k} \cdot \left(\frac{1}{2}\right)^n + \gamma^* \cdot \binom{n}{c} \left(\frac{1}{2}\right)^n. \end{aligned}$$

Wir wählen nun $c \in \mathbb{N}_0$ minimal mit

$$\sum_{k=c+1}^n \binom{n}{k} \cdot \left(\frac{1}{2}\right)^n = \alpha,$$

und setzen dann

$$\gamma^* = \frac{\alpha - P_{\vartheta=\frac{1}{2}}[X_1 + \dots + X_n > c]}{P_{\vartheta=\frac{1}{2}}[X_1 + \dots + X_n = c]}.$$

Dies ergibt

$$c = 408 \text{ und } \gamma^* = 0.67.$$

Da bei den gegebenen Daten

$$x_1 + \dots + x_n = 396 < c$$

ist, ergibt dann die Anwendung des obigen Tests:

Zum Niveau $\alpha = 0.05$ kann H_0 nicht abgelehnt werden, d. h. man kommt bei dem vorliegenden Datenmaterial zu dem Schluss, dass nicht unbedingt mehr Jungen als Mädchen geboren werden.

Der obige Test wird als **einseitiger Binomialtest** bezeichnet.

Zum Beweis von Satz 5.3 benötigen wir:

Satz 5.4. Voraussetzungen und Bezeichnungen wie in Satz 5.3. Dann gilt:

- a) $E_{\vartheta_0}[\varphi^*(X)] = \alpha$.
- b) φ^* minimiert die Fehlerwahrscheinlichkeiten erster **und** zweiter Art gleichmäßig unter allen Tests mit $E_{\vartheta_0}[\varphi(X)] = \alpha$, d. h. für jeden Test φ mit $E_{\vartheta_0}[\varphi(X)] = \alpha$ gilt:

$$E_{\vartheta}[\varphi(X)] \begin{cases} \geq \\ \leq \end{cases} E_{\vartheta}[\varphi^*(X)] \quad \text{falls} \quad \vartheta \begin{cases} \leq \\ > \end{cases} \vartheta_0.$$

Beweis:

a)

$$\begin{aligned} E_{\vartheta_0}[\varphi^*(X)] &= 1 \cdot P_{\vartheta=\vartheta_0}[T(X) > c] + \gamma^* \cdot P_{\vartheta=\vartheta_0}[T(X) = c] + 0 \\ &= \alpha \end{aligned}$$

nach Wahl von c und γ^* .

b) Sei φ beliebig mit $E_{\vartheta_0}[\varphi(X)] = \alpha$.

b₁) Sei $\vartheta_1 > \vartheta_0$. Zu zeigen:

$$E_{\vartheta_1}[\varphi(X)] \leq E_{\vartheta_1}[\varphi^*(X)].$$

Hilfsproblem: Teste

$$H'_0 : \vartheta = \vartheta_0 \text{ versus } H'_1 : \vartheta = \vartheta_1$$

zum Niveau α .

Mit $k^* = g_{\vartheta_0, \vartheta_1}(c)$ gilt (da $\{w_\vartheta : \vartheta \in \Theta\}$ Klasse mit monotonen Dichtequotienten in T ist):

$$\begin{aligned} f_{\vartheta_1}(x) > k^* \cdot f_{\vartheta_0}(x) &\Rightarrow g_{\vartheta_0, \vartheta_1}(T(x)) > g_{\vartheta_0, \vartheta_1}(c) \\ &\Rightarrow T(x) > c \end{aligned}$$

(da $g_{\vartheta_0, \vartheta_1}$ streng monoton wachsend ist).

Analog folgt

$$f_{\vartheta_1}(x) < k^* \cdot f_{\vartheta_0}(x) \Rightarrow T(x) < c.$$

Also hat φ^* die Bauart (5.1) aus Satz 5.1, und ist folglich gleichmäßig bester Test zum Niveau α für H'_0 gegen H'_1 . Wegen $E_{\vartheta_0}[\varphi(X)] \leq \alpha$ folgt daraus, dass die Fehlerwahrscheinlichkeit 2. Art von φ^* an der Stelle ϑ_1 nicht größer ist als die von φ , d. h.

$$1 - E_{\vartheta_1}[\varphi^*(X)] \leq 1 - E_{\vartheta_1}[\varphi(X)]$$

bzw.

$$E_{\vartheta_1}[\varphi(X)] \leq E_{\vartheta_1}[\varphi^*(X)]. \quad \rightsquigarrow b_1)$$

Bem.: Im Beweis kann oBdA

$$\frac{f_{\vartheta_1, \vartheta_1}(x)}{f_{\vartheta_0, \vartheta_1}(x)} = g_{\vartheta_0, \vartheta_1}(T(x))$$

für μ -fast alle x (statt $(w_{\vartheta_0} + w_{\vartheta_1})$ -fast alle x) vorausgesetzt werden, da man oBdA μ so abändern kann, dass eine $w_{\vartheta_0} + w_{\vartheta_1}$ -Nullmenge auch μ -Nullmenge ist.

b₂) Sei $\vartheta_1 < \vartheta_0$. Zu zeigen:

$$E_{\vartheta_1}\varphi(X) \geq E_{\vartheta_1}\varphi^*(X).$$

Dazu betrachten wir das **Hilfsproblem:** Teste

$$H'_0 : \vartheta = \vartheta_0 \text{ versus } H'_1 : \vartheta = \vartheta_1$$

zum Niveau $1-\alpha$.

Wir betrachten

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

was wegen

$$E_{\vartheta_0} \bar{\varphi}(X) = 1 - E_{\vartheta_0} \varphi^*(X) \stackrel{a)}{=} 1 - \alpha$$

ein Test zum Niveau $1-\alpha$ für das obige Hilfsproblem ist, der das Niveau an der Stelle ϑ_0 voll ausschöpft.

Mit $k^* = \frac{1}{g_{\vartheta_1, \vartheta_0}(c)}$ gilt dann analog zu b_1):

$$\begin{aligned} f_{\vartheta_1}(x) &> k^* \cdot f_{\vartheta_0}(x) \\ \Rightarrow \frac{1}{k^*} &> \frac{f_{\vartheta_0}(x)}{f_{\vartheta_1}(x)} \\ \Rightarrow g_{\vartheta_1, \vartheta_0}(c) &> g_{\vartheta_1, \vartheta_0}(T(x)) \\ \Rightarrow T(x) &< c \end{aligned}$$

da $g_{\vartheta_1, \vartheta_0}$ streng monoton wachsend ist.

Analog sieht man wieder

$$f_{\vartheta_1}(x) < k^* \cdot f_{\vartheta_0}(x) \Rightarrow T(x) > c.$$

Also hat $\bar{\varphi}$ die Bauart (5.1) aus Satz 5.1 und ist daher gleichmäßig bester Test zum Niveau $1-\alpha$ für H'_0 gegen H'_1 .

Wegen

$$E_{\vartheta_0}[1 - \varphi(X)] = 1 - E_{\vartheta_0}[\varphi(X)] = 1 - \alpha$$

ist auch $1 - \varphi$ Test zum Niveau $1-\alpha$ für H'_0 gegen H'_1 . Also ist die Fehlerwahrscheinlichkeit 2. Art von $1 - \varphi^*$ kleiner oder gleich der entsprechenden Fehlerwahrscheinlichkeit von $1 - \varphi$, was impliziert

$$1 - E_{\vartheta_1}[1 - \varphi^*(x)] \leq 1 - E_{\vartheta_1}[1 - \varphi(X)],$$

also

$$E_{\vartheta_1}[\varphi^*(X)] \leq E_{\vartheta_1}[\varphi(X)] \quad \rightsquigarrow b_2)$$

□

Beweis von Satz 5.3

a) **Beh.:** φ^* ist Test zum Niveau α .

Beweis: Sei $\vartheta_1 < \vartheta_0$. Zu zeigen:

$$E_{\vartheta_1}[\varphi^*(X)] \leq \alpha$$

Dazu betrachte

$$\varphi(x) := \alpha.$$

Nach Satz 5.4 sind die Fehlerwahrscheinlichkeiten erster Art von φ^* kleiner oder gleich als die von φ , also gilt

$$E_{\vartheta_1}[\varphi^*(X)] \leq E_{\vartheta_1}[\varphi(X)] = \alpha.$$

b) **Beh.:** φ^* minimiert die Fehlerwahrscheinlichkeiten zweiter Art unter allen Tests zum Niveau α .

Dies folgt wie im Beweis von Teil b_1) von Satz 5.4, da dort nur $E_{\vartheta_0}[\varphi(X)] \leq \alpha$, nicht aber $E_{\vartheta_0}[\varphi(X)] = \alpha$, benötigt wurde. \square

5.4 Tests im Zusammenhang mit der Normalverteilung

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig identisch $\mathcal{N}(\mu, \sigma^2)$ -verteilt. Zu testen sei

$$(5.2) \quad H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

bei bekannter oder unbekannter Varianz σ^2 , sowie

$$(5.3) \quad H_0 : \sigma^2 \leq \sigma_0^2 \text{ versus } H_1 : \sigma^2 > \sigma_0^2.$$

Beim einseitigen Gauß-Test wird zum Test von (5.2) bei bekannter Varianz σ_0^2 die Prüfgröße

$$T(X_1, \dots, X_n) = \frac{1}{\sigma_0} \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu_0)$$

betrachtet. Ist σ^2 unbekannt, so kann σ^2 geschätzt werden durch die empirische Varianz

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \text{ wobei } \bar{X} = \frac{1}{n} \sum_{j=1}^n X_j,$$

und sodann kann die Prüfgröße

$$T(X_1, \dots, X_n) = \frac{1}{S} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$$

(mit $S = \sqrt{S^2}$) betrachtet werden.

Zur Konstruktion eines Tests für (5.3) bietet sich die Betrachtung der Prüfgröße

$$T(X_1, \dots, X_n) = \frac{n-1}{\sigma_0^2} \cdot S^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2$$

an.

Im folgenden verwenden wir jeweils Tests der Bauart

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & , \text{ falls } T(X_1, \dots, X_n) > c, \\ 0 & , \text{ sonst.} \end{cases}$$

Hierbei heißt T **Prüfgröße** und c **kritischer Wert** des Tests φ .

Die Festlegung des kritischen Wertes c erfolgt durch Betrachtung der Fehlerwahrscheinlichkeit erster Art. Hierzu benötigen wir die Verteilung der Prüfgröße bei Gültigkeit der Nullhypothese, die im weiteren hergeleitet wird.

Def. 5.4:

- a) Sind X_1, \dots, X_n unabhängige $\mathcal{N}(0, 1)$ -verteilte Zufallsvariablen, so heißt die Verteilung von

$$\sum_{i=1}^n X_i^2$$

(zentrale) χ_n^2 -Verteilung mit n Freiheitsgraden.

- b) Sind X, Y unabhängig mit X $\mathcal{N}(0, 1)$ -verteilt und Y χ_n^2 -verteilt, so heißt die Verteilung von

$$\frac{X}{\sqrt{\frac{Y}{n}}}$$

(zentrale) t_n -Verteilung.

Bemerkung: Man kann zeigen:

- a) Die χ_n^2 -Verteilung hat die Dichte

$$f(x) = \frac{1}{2^{n/2} \cdot \Gamma(\frac{n}{2})} \cdot x^{\frac{n}{2}-1} \cdot e^{-\frac{x}{2}} \quad (x \in \mathbb{R}_+)$$

(mit

$$\Gamma(x) = \int_0^{\infty} t^{x-1} \cdot e^{-t} dt \quad (x > 0))$$

bzgl. des LB-Maßes und stimmt daher mit der $\Gamma_{\frac{1}{2},n}$ -Verteilung überein (vgl. Wahrscheinlichkeitstheorie).

b) Die t_n -Verteilung hat die Dichte

$$f(x) = \frac{1}{\sqrt{n \cdot \pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (x \in \mathbb{R}).$$

Bemerkung: Werte der Dichte, der Verteilungsfunktion, Fraktile etc. der obigen Verteilungen sind vertafelt.

Satz 5.5

X_1, \dots, X_n seien unabhängige $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariablen.

Setze

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{und} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Dann gilt:

- \bar{X} und S^2 sind unabhängig.
- \bar{X} ist $\mathcal{N}(\mu, \frac{\sigma^2}{n})$ -verteilt.
- $\frac{n-1}{\sigma^2} \cdot S^2$ ist \mathcal{X}_{n-1}^2 -verteilt
- $\sqrt{n} \cdot \frac{\bar{X} - \mu}{S}$ ist t_{n-1} -verteilt.

Im Beweis benötigen wir:

Lemma 5.6

Seien Y_1, \dots, Y_n unabhängige $\mathcal{N}(0, 1)$ -verteilte ZVen, sei $\underline{\underline{A}}$ eine orthogonale $n \times n$ -Matrix (d. h. es gelte $\underline{\underline{A}}^T \underline{\underline{A}} = \underline{\underline{1}}$) und sei

$$\begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = \underline{\underline{A}} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

Dann sind die ZVen Z_1, \dots, Z_n ebenfalls unabhängig $\mathcal{N}(0, 1)$ -verteilt.

Beweis:

Wir zeigen:

$$(*) \quad P[Z_1 \leq z_1, \dots, Z_n \leq z_n] = \int_{(-\infty, z_1] \times \dots \times (-\infty, z_n]} \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \right) d(x_1, \dots, x_n)$$

für alle $z_1, \dots, z_n \in \mathbb{R}$.

Dies impliziert die Behauptung, denn aus der obigen Beziehung folgt, dass die Dichte von (Z_1, \dots, Z_n) das Produkt von n -Dichten von standard-normalverteilten ZVen ist und damit die Komponenten von (Z_1, \dots, Z_n) unabhängig standard-normalverteilt sind.

Zum Nachweis von (*) setzen wir für beliebige $z_1, \dots, z_n \in \mathbb{R}$

$$I = (-\infty, z_1] \times \dots \times (-\infty, z_n]$$

und beachten

$$\begin{aligned} & P[Z_1 \leq z_1, \dots, Z_n \leq z_n] \\ &= P \left[\begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} \in I \right] \\ &\stackrel{\text{Def. } Z_i}{=} P \left[\underline{\underline{A}} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \in I \right] \\ &= P \left[\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \in \underline{\underline{A}}^{-1} I \right] \\ &= \int_{\underline{\underline{A}}^{-1} I} f_{Y_1, \dots, Y_n}(x_1, \dots, x_n) d(x_1, \dots, x_n) \\ &= \int_I f_{Y_1, \dots, Y_n}(\underline{\underline{A}}^{-1}(y_1, \dots, y_n)) \cdot |\det(\underline{\underline{A}}^{-1})| d(y_1, \dots, y_n) \end{aligned}$$

(mit Substitution $(x_1, \dots, x_n) = \underline{\underline{A}}^{-1}(y_1, \dots, y_n)$), wobei

$$\begin{aligned} f_{Y_1, \dots, Y_n}(x_1, \dots, x_n) &= \prod_{i=1}^n f_{Y_i}(x_i) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi}} \cdot e^{-x_i^2/2} \right) \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \cdot \exp \left(- \frac{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} x_1, \dots, x_n \end{pmatrix}}{2} \right) \end{aligned}$$

die Dichte von Y_1, \dots, Y_n ist.

Wegen $\underline{\underline{A}}^T \underline{\underline{A}} = \underline{\underline{1}}$ gilt $\underline{\underline{A}}^{-1} = \underline{\underline{A}}^T$ und $|\det(\underline{\underline{A}}^T)| = 1$. Damit folgt

$$\begin{aligned} & P[Z_1 \leq z_1, \dots, Z_n \leq z_n] \\ &= \int_I f_{Y_1, \dots, Y_n}(\underline{\underline{A}}^T(y_1, \dots, y_n)^T) \cdot 1 \, d(y_1, \dots, y_n) \\ &= \int_I \left(\frac{1}{\sqrt{2\pi}}\right)^n \cdot \exp\left(-\frac{(y_1, \dots, y_n) \underline{\underline{A}} \underline{\underline{A}}^T \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}{2}\right) d(y_1, \dots, y_n). \end{aligned}$$

Unter Beachtung von $\underline{\underline{A}} \underline{\underline{A}}^T = \underline{\underline{A}} \cdot \underline{\underline{A}}^{-1} = \underline{\underline{1}}$ und

$$\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{(y_1, \dots, y_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}{2}\right) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{y_i^2}{2}\right)$$

folgt daraus die Behauptung. □

Beweis von Satz 5.5:

Seien X_1, \dots, X_n unabhängig $\mathcal{N}(\mu, \sigma^2)$ -verteilt.

Dann sind Y_1, \dots, Y_n mit

$$Y_i = \frac{X_i - \mu}{\sigma}$$

unabhängig $\mathcal{N}(0, 1)$ -verteilt.

Wir wählen eine orthogonale Matrix $\underline{\underline{A}}$, deren erste Zeile gerade gleich

$$e^T = (n^{-1/2}, \dots, n^{-1/2})$$

ist, und setzen

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = \underline{\underline{A}} \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}.$$

Nach Lemma 5.6 sind dann die Z_1, \dots, Z_n unabhängig $\mathcal{N}(0, 1)$ -verteilt.

Im Folgenden stellen wir die interessierenden Größen als Funktionen der Z_1, \dots, Z_n dar.

Dazu beachten wir:

$$\sum_{i=1}^n Z_i^2 = \| Z^T Z \|^2 = \| Y^T Y \|^2 = \sum_{i=1}^n Y_i^2$$

(da $\underline{\underline{A}}$ orthogonal ist und $Z = \underline{\underline{A}}Y$), sowie

$$\begin{aligned} \sqrt{n} \cdot \bar{X} &= \sqrt{n} \frac{1}{n} \sum_{i=1}^n (\sigma \cdot Y_i + \mu) = \sqrt{n} \cdot \sigma \cdot \bar{Y} + \mu \cdot \sqrt{n} \\ &= \sigma \cdot e^T Y + \sqrt{n} \mu \\ &= \sigma \cdot Z_1 + \sqrt{n} \mu \end{aligned}$$

und

$$\begin{aligned} (n-1) \cdot S^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \sigma^2 \cdot \sum_{i=1}^n (Y_i - \bar{Y})^2 \\ &= \sigma^2 \cdot \left(\sum_{i=1}^n Y_i^2 - n \cdot (\bar{Y})^2 \right) \\ &= \sigma^2 \cdot \left(\sum_{i=1}^n Y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n Y_i \right)^2 \right) \\ &= \sigma^2 \cdot \left(\sum_{i=1}^n Y_i^2 - \underbrace{\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \right)^2}_{=e^T Y = Z_1} \right) \\ &= \sigma^2 \cdot \left(\sum_{i=1}^n Z_i^2 - Z_1^2 \right) \\ &\quad (\text{da } \| Z \|^2 = \| \underline{\underline{A}}Y \|^2) \\ &= \sigma^2 \cdot \sum_{i=2}^n Z_i^2. \end{aligned}$$

Daraus folgt nun leicht die Behauptung:

a) Da Z_1, \dots, Z_n unabhängig sind, sind auch

$$\sqrt{n} \cdot \bar{X} \stackrel{s.o.}{=} \sigma \cdot Z_1 + \sqrt{n} \cdot \mu$$

und

$$S^2 = \frac{\sigma^2}{n-1} \cdot \sum_{i=2}^n Z_i^2$$

unabhängig.

b) Da $Z_1 \mathcal{N}(0, 1)$ -verteilt ist, ist

$$\bar{X} \stackrel{s.o.}{=} \frac{\sigma}{\sqrt{n}} Z_1 + \mu$$

$\mathcal{N}(\mu, \frac{\sigma^2}{n})$ -verteilt (vgl. Wahrscheinlichkeitstheorie).

c) Da Z_2, \dots, Z_n unabhängig $\mathcal{N}(0, 1)$ -verteilt sind, ist

$$\frac{n-1}{\sigma^2} \cdot S^2 \stackrel{s.o.}{=} \sum_{i=2}^n Z_i^2$$

\mathcal{X}_{n-1}^2 -verteilt.

d)

$$\begin{aligned} \sqrt{n} \cdot \frac{\bar{X} - \mu}{S} &\stackrel{s.o.}{=} \frac{\sigma \cdot Z_1}{\sqrt{\frac{\sigma^2}{n-1} \sum_{i=2}^n Z_i^2}} \\ &= \frac{Z_1}{\sqrt{\frac{1}{n-1} \sum_{i=2}^n Z_i^2}} \end{aligned}$$

ist t_{n-1} -verteilt. □

Die obigen Überlegungen führen auf die folgenden Tests für normalverteilte Zufallsvariablen:

a) Teste bei bekannter Varianz $\sigma^2 = \sigma_0^2$

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

mittels

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & , \text{ falls } \frac{\sqrt{n}}{\sigma_0} (\bar{X} - \mu_0) > u_\alpha \\ 0 & , \text{ sonst} \end{cases}$$

mit $u_\alpha = \alpha$ -Fraktile von $\mathcal{N}(0, 1)$.

(**“Einseitiger Gauß-Test”**)

b) Teste bei unbekannter Varianz σ^2

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

mittels

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{ falls } \sqrt{n} \frac{\bar{X} - \mu_0}{S} > t_{n-1; \alpha}, \\ 0 & \text{ sonst,} \end{cases}$$

mit $t_{n-1; \alpha} = \alpha$ -Fraktile der t_{n-1} -Verteilung.

(**“Einseitiger t -Test von Student”**)

c) Teste bei unbekanntem (μ, σ^2)

$$H_0 : \sigma^2 \leq \sigma_0^2 \text{ versus } H_1 : \sigma^2 > \sigma_0^2$$

mittels

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{falls } \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 > \chi_{n-1; \alpha}^2, \\ 0 & \text{sonst,} \end{cases}$$

mit $\chi_{n-1; \alpha}^2 = \alpha$ -Fraktile der χ^2 -Verteilung mit $(n - 1)$ Freiheitsgraden.
 (“Einseitiger χ^2 -Test für die Varianz”).

Bemerkungen:

a) Der einseitige Gauß-Test ist ein gleichmäßig bester Test zum Niveau α für das obige Testproblem (vgl. Satz 5.2).

b) Der einseitige t -Test von Student ist ein Test zum Niveau α , denn:

Für $\mu_1 \leq \mu_0$ gilt:

$$P_{\mu=\mu_1} \left[\sqrt{n} \cdot \frac{\bar{X} - \mu_0}{S} > t_{n-1; \alpha} \right]$$

$$\leq P_{\mu=\mu_1} \left[\sqrt{n} \frac{\bar{X} - \mu_1}{S} > t_{n-1; \alpha} \right]$$

(da li. S. im Innern der Wahrscheinlichkeit “vergrößert” wurde)

$$= \alpha$$

(da li. S. nach Satz 5.5.d) t_{n-1} - verteilt ist).

c) Der einseitige χ^2 -Test ist ein Test zum Niveau α , denn analog zu b) folgt mit Satz 5.5 c):

Für $\sigma_1 \leq \sigma_0$ gilt:

$$P_{\sigma=\sigma_1} \left[\frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 > \chi_{n-1; \alpha}^2 \right]$$

$$\leq P_{\sigma=\sigma_1} \left[\frac{1}{\sigma_1^2} \sum_{i=1}^n (X_i - \bar{X})^2 > \chi_{n-1; \alpha}^2 \right]$$

$$= \alpha.$$

d) Man kann zeigen:

Der einseitige t -Test von Student und der χ^2 -Test besitzen “ähnliche” Optimalitätseigenschaften wie der einseitige Gauß-Test.

Bemerkung:

a) Möchte man in a)

$$H_0 : \mu \geq \mu_0 \text{ versus } H_1 : \mu > \mu_0$$

testen, so ersetze man in der Definition des Tests

$$\frac{\sqrt{n}}{\sigma_0}(\bar{X} - \mu_0) > u_\alpha$$

durch

$$\frac{\sqrt{n}}{\sigma_0}(\bar{X} - \mu_0) < u_{1-\alpha}.$$

Analog in b) und c).

b) Zweiseitige Tests:

Möchte man in a)

$$H_0 : \mu = \mu_0 \text{ versus } H_1 : \mu \neq \mu_0$$

testen, so ersetze man in der Definition des Tests

$$\frac{\sqrt{n}}{\sigma_0}(\bar{X} - \mu_0) > u_\alpha$$

durch

$$\left| \frac{\sqrt{n}}{\sigma_0}(\bar{X} - \mu_0) \right| > u_{\alpha/2}.$$

Analog in b).

In c) verwende man

$$\chi_{n-1;1-\alpha/2}^2 \leq \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 \leq \chi_{n-1;\alpha/2}^2$$

als "Ablehnungsbereich" von H_1 .

c) **Zweistichprobenprobleme**

Gegeben seien Stichproben

$$X_1, \dots, X_n$$

und

$$Y_1, \dots, Y_m$$

zweier Normalverteilungen mit unbekanntem Erwartungswerten μ_X bzw. μ_Y und gleicher (bekannter oder unbekannter) Varianz σ_0^2 bzw. σ^2 .

Getestet werden soll

$$H_0 : \mu_X = \mu_Y \text{ versus } H_1 : \mu_X \neq \mu_Y.$$

Ist die Varianz σ_0^2 bekannt, so schätzen wir μ_X bzw. μ_Y durch

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ bzw. } \bar{Y} = \frac{1}{m} \sum_{j=1}^m Y_j$$

und betrachten als Testgröße $|Z|$ mit

$$Z = \sqrt{\frac{n \cdot m}{n + m}} \cdot \frac{\bar{X} - \bar{Y}}{\sigma_0}.$$

Bei Gültigkeit von H_0 ist Z $\mathcal{N}(0, 1)$ -verteilt, denn Z ist als Linearkombination unabhängiger normalverteilter Zufallsvariablen normalverteilt und für $\mu_X = \mu_Y$ gilt

$$\begin{aligned} EZ &= \sqrt{\frac{n \cdot m}{n + m}} \cdot \frac{1}{\sigma_0} (E\bar{X} - E\bar{Y}) \\ &= \sqrt{\frac{n \cdot m}{n + m}} \cdot \frac{1}{\sigma_0} \cdot (\mu_X - \mu_Y) \\ &= 0 \end{aligned}$$

sowie

$$\begin{aligned} V(Z) &= \frac{n \cdot m}{n + m} \cdot \frac{1}{\sigma_0^2} (V(\bar{X}) + V(\bar{Y})) \\ &= \frac{n \cdot m}{n + m} \cdot \frac{1}{\sigma_0^2} \left(\frac{\sigma_0^2}{n} + \frac{\sigma_0^2}{m} \right) \\ &= 1. \end{aligned}$$

Also ist es naheliegend, H_0 abzulehnen, falls

$$\left| \sqrt{\frac{n \cdot m}{n + m}} \cdot \frac{\bar{X} - \bar{Y}}{\sigma_0} \right| > u_{\alpha/2}$$

ist (**Zweiseitiger Gauß-Test für zwei Stichproben**).

Ist dagegen die Varianz σ^2 unbekannt, so gehen wir analog zum zweiseitigen t -Test von Student vor. In einem ersten Schritt schätzen wir σ^2 durch die sogenannte gepoolte Stichprobenvarianz

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2}{m + n - 2}.$$

für die gilt

$$E[S^2] = \frac{1}{m + n - 2} \cdot ((n - 1) \cdot V(X_1) + (m - 1) \cdot V(Y_1)) = \sigma^2,$$

und verwenden dann als Testgröße $|Z|$ mit

$$Z = \sqrt{\frac{n \cdot m}{n + m}} \cdot \frac{\bar{X} - \bar{Y}}{\sqrt{S^2}}.$$

Man kann zeigen, dass Z bei Gültigkeit von H_0 t -verteilt ist mit $m + n - 2$ -Freiheitsgraden. Daher lehnt man hier H_0 ab, falls

$$|Z| > t_{\alpha/2; n+m-2}$$

gilt.

(“Zweiseitiger t -Test für zwei Stichproben”)

Anwendung des t -Tests in den Beispielen 5.1 bis 5.3:

In Beispiel 5.1 betrachten wir zunächst das einseitige Testproblem

$$H_0 : \mu \leq 0,7 \text{ versus } H_1 : \mu > 0,7$$

bei normalverteilten Daten mit unbekannter Varianz. Gegeben ist hier

$$n = 100, \bar{x} = 0,71, s^2 = 0,003 \text{ und } \alpha = 0,05.$$

Wegen

$$\sqrt{n} \left(\frac{\bar{x} - \mu_0}{s} \right) = \sqrt{100} \cdot \frac{0,71 - 0,7}{\sqrt{0,003}} \approx 1,83$$

und

$$t_{n-1; \alpha} = t_{99; 0,05} \approx 1,66$$

kann H_0 hier zum Niveau α abgelehnt werden.

Betrachtet man Beispiel 5.1 dagegen als zweiseitiges Testproblem

$$H_0 : \mu = 0,7 \text{ versus } H_1 : \mu \neq 0,7,$$

so muss man

$$\left| \sqrt{n} \frac{\bar{x} - \mu_0}{s} \right| = 1,833$$

mit

$$t_{n-1; \alpha/2} = t_{99; 0,025} \approx 1,98$$

vergleichen und kommt jetzt zum Ergebnis, dass H_0 zum Niveau α nicht abgelehnt werden kann. Dies liegt daran, dass bei zweiseitigen Testproblem der “Ablehnungsbereich” in Richtung $\mu > 0,7$ kleiner wird, da auch Werte mit $\mu < 0,7$ abgelehnt werden.

Als Nächstes wenden wir den einseitigen t -Test in Beispiel 5.2 zum Test von

$$H_0 : \mu \leq 312 \text{ versus } H_1 : \mu > 312$$

an. Hierbei ist

$$n = 40, \bar{x} = 315, s^2 = 120 \text{ und } \alpha = 0,05.$$

Wegen

$$\sqrt{n} \frac{\bar{x} - \mu_0}{s} = \sqrt{40} \cdot \frac{315 - 312}{\sqrt{120}} \approx 1,732$$

und

$$t_{n-1;\alpha} = t_{39;0,05} \approx 1,69$$

kann hier H_0 wieder zum Niveau α abgelehnt werden.

Abschließend wenden wir noch den zweiseitigen t -Test für zwei Stichproben in Beispiel 5.3 an.

Dabei ist

$$H_0 : \mu_X = \mu_Y \text{ und } H_1 : \mu_X \neq \mu_Y$$

und es wird angenommen, dass $\sigma_X^2 = \sigma_Y^2$ gilt und der Wert unbekannt ist.

Gegeben ist

$$n = 10, \bar{x} = 7,885, m = 5, \bar{y} = 8,128$$

und

$$s^2 = \frac{1}{13}(9,65 + 4,15) \approx 1,06.$$

Damit ist

$$\left| \sqrt{\frac{n-m}{n+m}} \cdot \frac{\bar{x} - \bar{y}}{\sqrt{s^2}} \right| \approx 0,4,$$

und wegen

$$t_{n+m-2;\alpha/2} = t_{13;0,025} \approx 2,2$$

kann H_0 anhand dieser Daten zum Niveau $\alpha = 5\%$ nicht abgelehnt werden.

5.5 Robustheit von Tests

Frage: Was passiert, wenn man die Tests aus dem vorigen Abschnitt auf nicht normalverteilte Daten anwendet ?

Wir betrachten dazu exemplarisch den einseitigen t -Test:

X_1, \dots, X_n seien unabhängig identisch verteilt mit unbekanntem Erwartungswert $\mu \in \mathbb{R}$ und unbekannter Varianz $\sigma^2 > 0$. Zu testen sei für festes $\mu_0 \in \mathbb{R}$

$$H_0 : \mu \leq \mu_0 \text{ versus } H_1 : \mu > \mu_0.$$

Wir verwenden dazu den t -Test:

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{falls } \sqrt{n} \frac{\bar{X} - \mu_0}{S} > t_{n-1; \alpha}, \\ 0 & \text{sonst,} \end{cases}$$

wobei $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ und $t_{n-1; \alpha} = \alpha$ -Fraktile der t -Verteilung mit $n - 1$ Freiheitsgraden, d.h., dass α -Fraktile der Verteilung von

$$\frac{Z}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} Z_i^2}}$$

mit Z, Z_1, \dots, Z_{n-1} unabhängig $N(0, 1)$ -verteilt.

Interessieren tun uns die Fehler erster Art

$$\mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) > t_{n-1; \alpha} \right] \quad (1)$$

für Verteilungen mit $\mathbf{E}X_1 = \mu \leq \mu_0$, sowie die Fehler zweiter Art

$$\mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) \leq t_{n-1; \alpha} \right] \quad (2)$$

für Verteilungen mit $\mathbf{E}X_1 = \mu > \mu_0$.

Wir führen dazu bei festgehaltener Verteilung (!) asymptotische Betrachtungen für (1) und (2) (für $n \rightarrow \infty$) durch.

Hilfreich dazu ist:

Lemma 5.7 Mit den obigen Bezeichnungen gilt:

a) Ist X_1 quadratisch integrierbar mit $0 < \mathbf{V}(X_1) < \infty$, so gilt

$$\sqrt{n} \frac{\bar{X} - \mathbf{E}X_1}{S} \rightarrow^{\mathcal{D}} N(0, 1) - \text{verteilte Zufallsvariable.}$$

b) Für beliebiges $\alpha \in (0, 1)$ gilt

$$t_{n; \alpha} \rightarrow u_\alpha \quad (n \rightarrow \infty),$$

wobei u_α das α -Fraktile von $N(0, 1)$ ist.

Beweis: Aus dem Satz von Slutsky folgt

$$\sqrt{n} \frac{\bar{X} - \mathbf{E}X_1}{S} = \sqrt{\frac{\mathbf{V}(X_1)}{S^2}} \cdot \frac{\sqrt{n}}{\sqrt{\mathbf{V}(X_1)}} \cdot (\bar{X} - \mathbf{E}X_1) \rightarrow^{\mathcal{D}} N(0, 1) - \text{verteilte Zufallsvariable,}$$

da nach dem zentralen Grenzwertsatz gilt

$$\frac{\sqrt{n}}{\sqrt{\mathbf{V}(X_1)}} \cdot (\bar{X} - \mathbf{E}X_1) \rightarrow^{\mathcal{D}} N(0, 1) - \text{verteilte Zufallsvariable}$$

und da aus dem starken Gesetz der großen Zahlen folgt

$$\begin{aligned} S^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \cdot \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \right) \\ &\rightarrow 1 \cdot (\mathbf{E}X_1^2 - (\mathbf{E}X_1)^2) = \mathbf{V}(X_1) \quad f.s., \end{aligned}$$

was impliziert

$$\sqrt{\frac{\mathbf{V}(X_1)}{S^2}} \rightarrow 1 \quad f.s.$$

b) Seien Z, Z_1, Z_2, \dots unabhängig identisch $N(0, 1)$ -verteilt. Dann gilt nach dem starken Gesetz der großen Zahlen

$$\frac{1}{n} \sum_{i=1}^n Z_i^2 \rightarrow \mathbf{E}Z_1^2 \quad f.s.,$$

was mit Slutsky impliziert

$$\frac{Z}{\sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}} \rightarrow^{\mathcal{D}} N(0, 1) - \text{verteilte Zufallsvariable.}$$

Mit der Definition von u_α folgt daraus für beliebiges $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbf{P} \left[\frac{Z}{\sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}} \leq u_\alpha - \epsilon \right] < 1 - \alpha$$

und

$$\lim_{n \rightarrow \infty} \mathbf{P} \left[\frac{Z}{\sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}} \leq u_\alpha + \epsilon \right] > 1 - \alpha.$$

Also existiert für beliebiges $\epsilon > 0$ ein $n_0 \in \mathbb{N}$ so, dass für alle $n \geq n_0$ gilt

$$\mathbf{P} \left[\frac{Z}{\sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}} \leq u_\alpha - \epsilon \right] < 1 - \alpha < \mathbf{P} \left[\frac{Z}{\sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}} \leq u_\alpha + \epsilon \right],$$

woraus mit der Definition von

$$t_{n;\alpha} = \min \left\{ z \in \mathbb{R} : \mathbf{P} \left[\frac{Z}{\sqrt{\frac{1}{n} \sum_{i=1}^n Z_i^2}} \leq z \right] \geq 1 - \alpha \right\}$$

folgt

$$u_\alpha - \epsilon \leq t_{n;\alpha} \leq u_\alpha + \epsilon.$$

Da dies äquivalent ist zu

$$|t_{n;\alpha} - u_\alpha| \leq \epsilon,$$

folgt die Behauptung. □

Satz 5.8. Seien X_1, X_2, \dots unabhängig identisch verteilte reelle Zufallsvariablen mit Erwartungswert $\mu = \mathbf{E}X_1 \in \mathbb{R}$ und $0 < \mathbf{V}(X_1) < \infty$. Sei $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, $\alpha \in (0, 1)$ und $\mu_0 \in \mathbb{R}$.

a) Für $\mu \leq \mu_0$ gilt:

$$\limsup_{n \rightarrow \infty} \mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) > t_{n-1;\alpha} \right] \leq \alpha.$$

b) Für $\mu > \mu_0$ gilt:

$$\limsup_{n \rightarrow \infty} \mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) \leq t_{n-1;\alpha} \right] = 0.$$

Beweis.

a) Für $\mu \leq \mu_0$ gilt nach Lemma 5.7 und Slutsky:

$$\begin{aligned} \mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) > t_{n-1;\alpha} \right] &\leq \mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu) > t_{n-1;\alpha} \right] \\ &= \mathbf{P} \left[\frac{u_\alpha}{t_{n-1;\alpha}} \cdot \frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu) > u_\alpha \right] \\ &\stackrel{(n \rightarrow \infty)}{\rightarrow} \mathbf{P} [N(0, 1)\text{-verteilte Zufallsvariable} > u_\alpha] = \alpha. \end{aligned}$$

b) Für $\mu > \mu_0$ gilt für $z \in \mathbb{R}$ beliebig nach Lemma 5.7:

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) \leq t_{n-1;\alpha} \right] \\ &= \limsup_{n \rightarrow \infty} \mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu) \leq t_{n-1;\alpha} - (\mu - \mu_0) \cdot \frac{\sqrt{n}}{S} \right] \\ &= \limsup_{n \rightarrow \infty} \mathbf{P} \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu) \leq z \right] \\ &= \mathbf{P} [N(0, 1)\text{-verteilte Zufallsvariable} \leq z] \rightarrow 0 \quad (z \rightarrow -\infty). \end{aligned}$$

Hierbei haben wir bei der zweiten Gleichheit ausgenutzt, dass nach dem starken Gesetz der großen Zahlen gilt

$$(\mu - \mu_0) \cdot \frac{\sqrt{n}}{S} \rightarrow \infty \quad f.s.,$$

vgl. Beweis von Lemma 5.7. □

Bemerkung: Eigentlich interessant sind Aussagen über

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in \Theta_0} \mathbf{P}_\vartheta \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) > t_{n-1; \alpha} \right] \quad (3)$$

und

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in \Theta_1} \mathbf{P}_\vartheta \left[\frac{\sqrt{n}}{S} \cdot (\bar{X} - \mu_0) \leq t_{n-1; \alpha} \right] \quad (4)$$

für Klassen $\{w_\vartheta : \vartheta \in \Theta_0\}$ und $\{w_\vartheta : \vartheta \in \Theta_1\}$ von Wahrscheinlichkeitsmaßen mit Erwartungswert kleiner oder gleich μ_0 bzw. größer μ_0 , die keine Normalverteilungen sind.

Für Aussagen zu (3) braucht man dann Voraussetzungen, die sicherstellen, dass die Konvergenz im Beweis von Satz 5.8 gleichmäßig bzgl. ϑ ist (z.B. uniforme Beschränktheit geeigneter Momente). Für (4) muss zusätzlich ein "Abstand" größer Null zwischen den beiden Verteilungsklassen bestehen (da i.A. die Gütefunktion stetig von μ abhängen wird, und i.A. (4) größer oder gleich $1 - \alpha$ für n groß sein wird, sofern (3) kleiner oder gleich α ist).

5.6 Zwei nichtparametrische Tests

In diesem Abschnitt behandeln wir zwei Tests, bei denen nicht zwingend eine durch einen endlich-dimensionalen Parameter parametrisierte Klasse von Verteilungen vorgegeben ist.

5.6.1 Der Zeichentest

Im Folgenden seien X_1, \dots, X_n unabhängig identisch verteilte reelle Zufallsvariablen mit **stetiger** Verteilungsfunktion. μ_{med} sei der Median von X_1 , der hier durch

$$\mathbf{P}\{X_1 < \mu_{med}\} = \frac{1}{2} = \mathbf{P}\{X_1 > \mu_{med}\}$$

(aufgrund der Stetigkeit der Verteilungsfunktion) eindeutig definiert ist (ansonsten allgemein:

$$\mu_{med} = \min \left\{ z \in \mathbb{R} : \mathbf{P}\{X_1 \geq z\} \geq \frac{1}{2} \right\}.$$

Zu testen sei für vorgegebenes $\mu_0 \in \mathbb{R}$:

$$H_0 : \mu_{med} \leq \mu_0 \quad \text{vs.} \quad H_1 : \mu_{med} > \mu_0.$$

Bsp. im Zusammenhang mit sogenannten verbundenen Stichproben: n Kunden testen jeweils Produkt 1 und Produkt 2, wobei Kunde i die Qualität von Produkt 1 bzw. 2 durch Y_i bzw. Z_i einschätzt. Wir fassen sodann $(X_1, Y_1), \dots, (X_n, Y_n)$ als eine unabhängig identisch verteilte Stichprobe auf und testen ausgehend von

$$X_i = Y_i - Z_i$$

die obigen Hypothesen mit $\mu_0 = 0$. H_1 besagt dann, dass das neue Produkt besser eingeschätzt wird, während bei Gültigkeit von H_0 das alte Produkt das bessere ist.

Beim **Zeichentest** betrachtet man

$$T(X_1, \dots, X_n) = \sum_{i=1}^n I_{\{X_i > \mu_0\}} = \sum_{i=1}^n (\text{sign}(X_i - \mu_0))_+$$

(mit $\text{sign}(z) = \text{Vorzeichen von } z \in \mathbb{R}$).

Für $\mu_{med} = \mu_0$ gilt

$$\mathbf{P}\{X_1 > \mu_0\} = \frac{1}{2},$$

also ist in diesem Fall $T(X_1, \dots, X_n)$ $b(n, 1/2)$ -verteilt.

Wir lehnen daher H_0 zum Niveau α ab, falls

$$T(X_1, \dots, X_n) > \alpha - \text{Fraktile von } b(n, 1/2)$$

gilt.

Analog für

$$H_0 : \mu_{med} = \mu_0 \quad \text{vs.} \quad H_1 : \mu_{med} \neq \mu_0$$

(vgl. Übungen).

5.6.2 Der Wilcoxon-Rangsummen-Test

$X_1, \dots, X_n, Y_1, \dots, Y_m$ seien unabhängige reelle Zufallsvariablen, wobei X_1, \dots, X_n unabhängig identisch verteilt sind mit Verteilungsfunktion F und Y_1, \dots, Y_m unabhängig identisch verteilt sind mit Verteilungsfunktion G . Zu testen sei

$$H_0 : F = G \quad \text{versus} \quad H_1 : F \geq G \text{ und } F \neq G$$

(also bei $H_1: F(x) \geq G(x)$ für alle $x \in \mathbb{R}$ und $F(y) > G(y)$ für ein $y \in \mathbb{R}$).

F und G seien stetig.

Sind $z_1, \dots, z_N \in \mathbb{R}$ paarweise verschieden, so ist die Rangzahl r_i von z_i bzgl. (z_1, \dots, z_N) der Platz $\in \{1, \dots, N\}$, den z_i einnimmt, wenn man die z_1, \dots, z_N der Größe nach sortiert.

Bsp. Für $(z_1, z_2, z_3, z_4) = (5, 1, 2, 6)$ gilt $(r_1, r_2, r_3, r_4) = (3, 1, 2, 4)$.

Zur Entscheidung zwischen H_0 und H_1 betrachtet man beim Wilcoxon-Rangsummen-Test die Rangzahlen r_1, \dots, r_{n+m} von $X_1, \dots, X_n, Y_1, \dots, Y_m$. Wegen der Unabhängigkeit und der identischen Verteiltheit der X_1, \dots, X_n und der Y_1, \dots, Y_m ist es klar, dass uns eigentlich nur $\{r_1, \dots, r_n\}$ und $\{r_{n+1}, \dots, r_{n+m}\}$ interessieren. Da jede der Mengen die jeweils andere eindeutig bestimmt, betrachten wir sogar nur Rangzahlen

$$1 \leq S_1 < S_2 < \dots < S_m \leq n + m$$

der Y_1, \dots, Y_m im $(n+m)$ -Tupel $(X_1, \dots, X_n, Y_1, \dots, Y_m)$, und zwar der Größe nach geordnet. (Bindungen treten hierbei nur mit Wk. Null auf, ggf. randomisieren).

Bei Gültigkeit von H_0 tritt jede Anordnung der Y_1, \dots, Y_m im $(n+m)$ -Tupel $(X_1, \dots, X_n, Y_1, \dots, Y_m)$ mit der gleichen Wahrscheinlichkeit auf, also gilt unter H_0

$$\mathbf{P}[S_1 = s_1, \dots, S_m = s_m] = \frac{1}{\binom{n+m}{m}}$$

für alle $s_1, \dots, s_m \in \mathbb{N}$ mit $1 \leq s_1 < s_2 < \dots < s_m \leq n + m$.

Jeder **Rangtest**, d.h., jeder auf den obigen Rangzahlen basierende Test, mit

$$\alpha = \frac{k}{\binom{n+m}{m}}$$

als Fehlerwahrscheinlichkeit 1. Art (mit $k \in \mathbb{N}_0$), hat daher einen kritischen Bereich (d.h., einen Bereich in dem H_0 abgelehnt wird), der aus genau k Tupeln

$$(s_1, \dots, s_m) \in \mathbb{N}^m$$

mit $1 \leq s_1 < s_2 < \dots < s_m \leq n + m$ besteht.

Beim **Wilcoxon-Rangsummen-Test** wählt man nun den kritischen Bereich als

$$\{(s_1, \dots, s_m) \in \mathbb{N}^m \quad : \quad s_1 + s_2 + \dots + s_m > c\},$$

wobei s_1, \dots, s_m die Rangzahlen zu y_1, \dots, y_m im $(n+m)$ -Tupel $(x_1, \dots, x_n, y_1, \dots, y_m)$ sind (die hier nicht sortiert werden müssen), und wobei $c > 0$ in Abhängigkeit des Niveaus geeignet gewählt wird.

5.7 Multiples Testen

Oft möchte man nicht nur über eine Eigenschaft eines Datensatzes, sondern über **viele simultan** Aussagen machen.

Beispiel: In einer Studie zur Bildungsforschung werden Tests an Schülern in den 16 verschiedenen Bundesländern gemacht, und man möchte die Ergebnisse der einzelnen Bundesländer miteinander vergleichen.

Formal: X ZV mit $\mathbf{P}_X \in \{w_\vartheta : \vartheta \in \Theta\}$, wobei $\Theta = \Theta_0^{(i)} \cup \Theta_1^{(i)}$ für $i = 1, \dots, s$.

Teste simultan

$$H_0^{(i)} : \vartheta \in \Theta_0^{(i)} \quad \text{versus} \quad H_1^{(i)} : \vartheta \in \Theta_1^{(i)}$$

für $i = 1, \dots, s$ zum Niveau α mittels Tests $\varphi_1, \dots, \varphi_s$.

Beispiel: Ausgehend von einer Stichprobe einer $N(\mu, \sigma^2)$ -Verteilung soll getestet werden

$$H_0 : \mu \leq \mu_0 \text{ und } \sigma \leq \sigma_0 \quad \text{versus} \quad H_1 : \mu > \mu_0 \text{ oder } \sigma > \sigma_0,$$

wobei wir im Falle einer Ablehnung von H_0 auch wissen wollen, welche der beiden Hypothesen

$$H_0^{(1)} : \mu \leq \mu_0$$

bzw.

$$H_0^{(2)} : \sigma \leq \sigma_0$$

abgelehnt wurden.

Definition: Ist $\{\varphi_i\}_{i=1, \dots, s}$ eine Familie von Tests für die obigen Testprobleme, so heißt

$$\sup_{\vartheta \in \Theta} \mathbf{P}_\vartheta \left[\exists i \in \{1, \dots, s\} : \vartheta \in \Theta_0^{(i)}, \varphi_i(X) = 1 \right]$$

die Fehlerrate erster Art von $\{\varphi_i\}_{i=1, \dots, s}$.

Frage: Wie konstruiert man zu vorgegebenen $\alpha \in (0, 1)$ eine Familie von Tests mit Fehlerrate erster Art kleiner oder gleich α (die natürlich im Hinblick auf die Fehler zweiter Art die Nullhypothesen möglichst oft ablehnen soll) ?

Bemerkung: Die Wahl von φ_i als Test zum Niveau α für alle $i = 1, \dots, s$ ist hier i.A. nicht sinnvoll. Denn existiert

$$\vartheta_0 \in \bigcap_{i=1}^s \Theta_0^{(i)}$$

und sind für $\vartheta = \vartheta_0$ die $\varphi_1(X), \dots, \varphi_s(X)$ unabhängig (was z.B. möglich ist, falls X aus s unabhängigen Komponenten besteht), so gilt im Falle, dass alle φ_i das

Niveau für $\vartheta = \vartheta_0$ voll ausschöpfen, aufgrund der Unabhängigkeit der $\varphi_1(X), \dots, \varphi_s(X)$:

$$\begin{aligned}
 \mathbf{P}_{\vartheta_0} \left[\exists i \in \{1, \dots, s\} : \vartheta_0 \in \Theta_0^{(i)}, \varphi_i(X) = 1 \right] &= \mathbf{P}_{\vartheta_0} [\exists i \in \{1, \dots, s\} : \varphi_i(X) = 1] \\
 &= 1 - \mathbf{P}_{\vartheta_0} [\forall i \in \{1, \dots, s\} : \varphi_i(X) \neq 1] \\
 &= 1 - \prod_{i=1}^s \mathbf{P}_{\vartheta_0} [\varphi_i(X) \neq 1] \\
 &= 1 - \prod_{i=1}^s (1 - \mathbf{P}_{\vartheta_0} [\varphi_i(X) = 1]) \\
 &= 1 - (1 - \alpha)^s.
 \end{aligned}$$

Z.B. ergibt sich hier für $\alpha = 0,05$ und $s = 10$: $1 - (1 - \alpha)^{10} \approx 0,4$
 und für $\alpha = 0,05$ und $s = 50$: $1 - (1 - \alpha)^{50} \approx 0,92$

Primitive Vorgehensweise: Wähle φ_i als Test zum Niveau α/s . Denn dann gilt:

$$\begin{aligned}
 &\sup_{\vartheta \in \Theta} \mathbf{P}_{\vartheta} \left[\exists i \in \{1, \dots, s\} : \vartheta \in \Theta_0^{(i)}, \varphi_i(X) = 1 \right] \\
 &\leq \sup_{\vartheta \in \Theta} \sum_{i=1}^s \mathbf{P}_{\vartheta} \left[\vartheta \in \Theta_0^{(i)}, \varphi_i(X) = 1 \right] \\
 &\leq \sum_{i=1}^s \sup_{\vartheta \in \Theta} \mathbf{P}_{\vartheta} \left[\vartheta \in \Theta_0^{(i)}, \varphi_i(X) = 1 \right] \\
 &= \sum_{i=1}^s \sup_{\vartheta \in \Theta_0^{(i)}} \mathbf{P}_{\vartheta} [\varphi_i(X) = 1] \\
 &\leq \sum_{i=1}^s \frac{\alpha}{s} = \alpha,
 \end{aligned}$$

wobei wir bei der letzten Ungleichung ausgenutzt haben, dass φ_i ein Test zum Niveau α/s ist.

Aber: Bei dieser Vorgehensweise wird das Niveau eventuell sehr klein.

Für eine raffiniertere Vorgehensweise benötigen wir den Begriff des p -Wertes:

Ist φ Test der Bauart

$$\varphi(x) = \begin{cases} 1 & \text{falls } T(x) > c \\ 0 & \text{sonst} \end{cases}$$

(mit Teststatistik T und kritischem Wert c) für

$$H_0 : \vartheta \in \Theta_0 \quad \text{versus} \quad H_1 : \vartheta \in \Theta_1,$$

so ist

$$S := \{x : T(x) > c\}$$

der sogenannte kritische Bereich von φ , der alle x enthält, für die der nichtrandomisierte Test φ die Hypothese H_0 ablehnt. Wir nehmen im Folgenden an, dass φ das Niveau α voll ausschöpft, d.h., das gilt

$$\sup_{\vartheta \in \Theta_0} \mathbf{E}_{\vartheta}[\varphi(X)] = \sup_{\vartheta \in \Theta_0} \mathbf{P}_{\vartheta}[T(X) > c] = \alpha. \quad (5)$$

Weiter sei $c = c_{\alpha}$ das minimale (bzw. infimale) c , für das (5) erfüllt ist, und es sei $S_{\alpha} = \{x : T(x) > c_{\alpha}\}$ der zu φ gehörende kritische Bereich. Dann gilt für $0 < \alpha_1 < \alpha_2 < 1$

$$c_{\alpha_1} > c_{\alpha_2} \quad \text{und folglich} \quad S_{\alpha_1} \subseteq S_{\alpha_2}$$

(da $c \mapsto \sup_{\vartheta \in \Theta_0} \mathbf{E}_{\vartheta}[\varphi(X)] = \sup_{\vartheta \in \Theta_0} \mathbf{P}_{\vartheta}[T(X) > c]$ monoton fallend in c ist).

Der p -Wert des Tests φ ist nun dasjenige (vom Werte x von X abhängende) Niveau, das gerade noch zur Ablehnung von H_0 (also zu $\varphi(x) = 1$) führt. Genauer definieren wir:

$$\hat{p} = \hat{p}(x) = \inf\{\alpha \in (0, 1) : x \in S_{\alpha}\} = \inf\{\alpha \in (0, 1) : T(x) > c_{\alpha}\}.$$

Gilt nun $\hat{p} < \alpha_0$, so existiert $\alpha \in (\hat{p}, \alpha_0)$ mit

$$T(x) > c_{\alpha} > c_{\alpha_0},$$

was impliziert $x \in S_{\alpha_0}$, d.h., der Test φ zum Niveau α_0 lehnt H_0 ab.

Für den p -Wert gilt:

Lemma 5.9. Für den Ablehnungsbereich S_{α} des obigen Tests φ_{α} zum Niveau α gelte

$$\sup_{\vartheta \in \Theta_0} \mathbf{P}_{\vartheta}[X \in S_{\alpha}] \leq \alpha$$

für alle $\alpha \in (0, 1)$ (was äquivalent zu

$$\sup_{\vartheta \in \Theta_0} \mathbf{E}_{\vartheta}[\varphi_{\alpha}(X)] \leq \alpha$$

für alle $\alpha \in (0, 1)$ ist). Dann gilt für jedes $\vartheta \in \Theta_0$ und jedes $u \in (0, 1)$:

$$\mathbf{P}_{\vartheta}[\hat{p}(X) \leq u] \leq u.$$

Beweis. Aus $\hat{p}(X) \leq u$ folgt für jedes $0 \leq u < v < 1$ die Beziehung $\hat{p}(X) < v$, was $X \in S_v$ impliziert. Also gilt

$$\mathbf{P}_\vartheta[\hat{p}(X) \leq u] \leq \mathbf{P}_\vartheta[X \in S_v] \leq \sup_{\vartheta \in \Theta_0} \mathbf{P}_\vartheta[X \in S_v] \leq v$$

für alle $0 \leq u < v < 1$, was die Behauptung impliziert. \square

Mit Hilfe des p -Wertes können wir nun wie folgt eine Methode zum Multiplen Testen einführen:

Zu testen sei simultan

$$H_0^{(i)} : \vartheta \in \Theta_0^{(i)} \quad \text{versus} \quad H_1^{(i)} : \vartheta \in \Theta_1^{(i)}$$

für $i = 1, \dots, s$ zum Niveau α . Dazu seien Tests $\varphi_1, \dots, \varphi_s$ der obigen Bauart vorgegeben.

$\hat{p}_1 = \hat{p}_1(X), \dots, \hat{p}_s = \hat{p}_s(X)$ seien die zugehörigen p -Werte, und $\hat{p}_{(1)} \leq \dots \leq \hat{p}_{(s)}$ sei eine aufsteigende Anordnung derselben mit zugehöriger Anordnung $H_0^{((1))}, \dots, H_0^{((s))}$ der Hypothesen $H_0^{(1)}, \dots, H_0^{(s)}$.

Zum Niveau $\alpha \in (0, 1)$ testen wir die Hypothesen $H_0^{(1)}, \dots, H_0^{(s)}$ wie folgt:

Schritt 1: Gilt $\hat{p}_{(1)} \geq \alpha/s$, so akzeptiere $H_0^{((1))}, \dots, H_0^{((s))}$. Gilt dagegen $\hat{p}_{(1)} < \alpha/s$, so verwerfe $H_0^{((1))}$ und fahre mit Schritt 2 fort.

Schritt 2: Gilt $\hat{p}_{(2)} \geq \alpha/(s-1)$, so akzeptiere $H_0^{((2))}, \dots, H_0^{((s))}$. Gilt dagegen $\hat{p}_{(2)} < \alpha/(s-1)$, so verwerfe $H_0^{((2))}$ und fahre mit Schritt 3 fort.

Schritt 3: Gilt $\hat{p}_{(3)} \geq \alpha/(s-2)$, so akzeptiere $H_0^{((3))}, \dots, H_0^{((s))}$. Gilt dagegen $\hat{p}_{(3)} < \alpha/(s-2)$, so ...

Für dieses Verfahren gilt:

Satz 5.10: Für die Tests φ_i ($i = 1, \dots, s$) gelte

$$\sup_{\theta \in \Theta_0^{(i)}} \mathbf{E}_\theta[\varphi_i(X)] \leq \alpha$$

für alle $i \in \{1, \dots, s\}$. Dann ist die Fehlerrate erster Art bei obigem Testverfahren kleiner oder gleich α .

Beweis. Sei $\vartheta \in \Theta$ beliebig und

$$I := \left\{ i \in \{1, \dots, s\} : \vartheta \in \Theta_0^{(i)} \right\} \neq \emptyset.$$

Wir müssen zeigen, dass die Wahrscheinlichkeit, dass der obige Test irgendeine der Hypothesen $H_0^{(i)}$ ($i \in I$) verwirft, kleiner oder gleich α ist.

Sei $j \in I$ der kleinste Index mit

$$\hat{p}_{(j)} = \min_{i \in I} \hat{p}_i.$$

Der obige Test verwirft nur dann eine der Hypothesen $H_0^{(i)}$ ($i \in I$), falls er $H_0^{(j)}$ verwirft, was genau dann eintritt, wenn gilt

$$\hat{p}_{(1)} < \frac{\alpha}{s}, \hat{p}_{(2)} < \frac{\alpha}{s-1}, \dots, \hat{p}_{(j)} < \frac{\alpha}{s-j+1}.$$

Letzteres impliziert

$$\min_{i \in I} \hat{p}_i = \hat{p}_{(j)} < \frac{\alpha}{s-j+1} \leq \frac{\alpha}{|I|},$$

da $I \subseteq \{1, \dots, s\}$ und die zu $\hat{p}_{(1)}, \dots, \hat{p}_{(j-1)}$ gehörenden Indices nicht in I enthalten sind (woraus $|I| \leq s - (j-1)$ folgt).

Daher gilt:

$$\begin{aligned} & \mathbf{P}_{\vartheta} \left[\text{Der Test verwirft eine der Hypothesen } H_0^{(i)} \text{ für } i \in I \right] \\ & \leq \mathbf{P}_{\vartheta} \left[\min_{i \in I} \hat{p}_i < \frac{\alpha}{|I|} \right] \\ & \leq \sum_{i \in I} \mathbf{P}_{\vartheta} \left[\hat{p}_i < \frac{\alpha}{|I|} \right] \\ & \stackrel{\text{Lemma 5.9}}{\leq} \sum_{i \in I} \frac{\alpha}{|I|} = \alpha. \end{aligned}$$

□

Bemerkung: Vergleichen wir die primitive mit der raffinierteren Vorgehensweise, und sagen wir vereinfachend, dass unser Test φ_i zum Niveau α die Hypothese $H_0^{(i)}$ genau dann verwirft, wenn gilt

$$\hat{p}_i < \alpha,$$

so können wir beide Methoden wie folgt miteinander vergleichen:

Bei der primitiven Methode wird $H_0^{(i)}$ genau dann verworfen, wenn gilt

$$\hat{p}_{(i)} < \frac{\alpha}{s}.$$

Dagegen verwirft die raffiniertere Methode $H_0^{(i)}$ genau dann, wenn gilt:

$$\hat{p}_{(1)} < \frac{\alpha}{s}, \hat{p}_{(2)} < \frac{\alpha}{s-1}, \dots, \hat{p}_{(i)} < \frac{\alpha}{s-i+1}.$$

Wegen

$$\hat{p}_{(1)} \leq \hat{p}_{(2)} \leq \dots \hat{p}_{(i)}$$

wird die letzte Bedingung aber von $\hat{p}_{(i)} < \frac{\alpha}{s}$ impliziert, also verwirft die raffinierte Methode in den Fällen, in denen die primitive Methode $H_0^{((i))}$ verwirft, immer ebenso diese Hypothese.

Daher wird bei der raffinierteren Methode $H_0^{((i))}$ öfters abgelehnt, was zu kleineren Fehlern zweiter Art führt (bei denen $H_1^{((i))}$ richtig ist, aber trotzdem $H_0^{((i))}$ nicht abgelehnt wird).

6 Bereichsschätzungen

6.1 Einführung

In Kapitel 3 haben wir unabhängige und identisch verteilte ZVen X_1, \dots, X_n betrachtet mit

$$P_{X_1} = w_{\vartheta_0} \in \{w_{\vartheta} : \vartheta \in \Theta\},$$

und versucht, eine Schätzung $T_n(X_1, \dots, X_n)$ von $g(\vartheta_0)$ für ein gegebenes $g : \Theta \rightarrow \mathbb{R}^k$ zu konstruieren.

Klar: I. A. gilt $T_n(X_1, \dots, X_n) \neq g(\vartheta_0)$.

Daher ist es eventuell realistischer zu versuchen, eine Menge

$$C(X_1, \dots, X_n) \subseteq \mathbb{R}^k$$

zu konstruieren mit

$$g(\vartheta_0) \in C(X_1, \dots, X_n).$$

Wünschenswert dabei ist:

- (1) Die Wahrscheinlichkeit

$$P_{\vartheta}[g(\vartheta) \in C(X_1, \dots, X_n)]$$

soll "möglichst groß" sein für alle $\vartheta \in \Theta$ und alle u. i. v. Zufallsvariablen X_1, \dots, X_n mit $P_{X_1} = w_{\vartheta}$.

- (2) Die Menge

$$C(X_1, \dots, X_n)$$

soll "möglichst klein" sein.

Die folgende Definition formalisiert (1).

Definition 6.1: Sei $\alpha \in (0, 1)$.

- a) $C(X_1, \dots, X_n)$ heißt **Konfidenzbereich zum Konfidenzniveau** $1 - \alpha$ falls für alle $\vartheta \in \Theta$ und alle u. i. v. ZVen X_1, \dots, X_n mit $P_{X_1} = w_{\vartheta}$ gilt:

$$P_{\vartheta}[g(\vartheta) \in C(X_1, \dots, X_n)] \geq 1 - \alpha.$$

Hierbei wird vorausgesetzt, dass die Wahrscheinlichkeit auf der linken Seite existiert.

- b) Ist $C(X_1, \dots, X_n)$ in a) ein Intervall, dann heißt $C(X_1, \dots, X_n)$ **Konfidenzintervall zum Konfidenzniveau** $1 - \alpha$.

6.2 Anwendungsbeispiele

Beispiel 6.1. Bei einer Umfrage ca. 3 Wochen vor der Bundestagswahl 2002 gaben von $n = 2000$ Befragten 912 bzw. 927 an, für Rot-Grün bzw. für Schwarz-Gelb stimmen zu wollen. Wie bestimmt man daraus möglichst kleine Intervalle, die mit Wahrscheinlichkeit größer oder gleich 0,95 den Anteil der entsprechenden Wähler in der Menge aller Wahlberechtigten überdecken?

Beispiel 6.2. Im Jahr 1999 wurden in Deutschland 374.448 Mädchen und 396.296 Jungen geboren. Man gebe ein möglichst kleines Intervall an, das mit Wahrscheinlichkeit größer oder gleich 0,99 die Wahrscheinlichkeit für eine "Jungengeburt" überdeckt.

Beispiel 6.3. Ein Psychologe interessiert sich für die Reaktionszeit im Straßenverkehr von 10-jährigen Schülern. Bei $n = 51$ Schülern wurde eine mittlere Reaktionszeit $\bar{x} = 0,8$ [sec.] mit empirischer Varianz $s^2 = 0,04$ [sec.²] gemessen. Wie bestimmt man daraus ein (möglichst kleines) Intervall, das die mittlere Reaktionszeit mit Wahrscheinlichkeit größer oder gleich 0,95 überdeckt?

6.3 Konstruktion von Bereichsschätzungen mit Hilfe von stochastischen Pivots

Seien wie oben X_1, \dots, X_n u. i. v. ZVen mit

$$P_{X_1} = w_{\vartheta_0} \in \{w_{\vartheta} : \vartheta \in \Theta\}$$

und $g : \Theta \rightarrow \mathbb{R}^k$.

Die Idee bei der Konstruktion von Bereichsschätzungen mit Hilfe von stochastischen Pivots ist die folgende:

Wir konstruieren ein sogenanntes **stochastisches Pivot**

$$Q = Q(X_1, \dots, X_n, g(\vartheta_0))$$

derart, dass die Verteilung von Q unabhängig von $\vartheta_0 \in \Theta$ ist. Wir wählen dann eine Menge B mit

$$P[Q \in B] = 1 - \alpha,$$

und schreiben

$$Q(X_1, \dots, X_n, g(\vartheta_0)) \in B$$

um zu

$$g(\vartheta_0) \in C(X_1, \dots, X_n).$$

Anwendung in Beispiel 6.3:

Wir treffen die vereinfachende Annahme, dass X_1, \dots, X_n unabhängig $\mathcal{N}(\mu, \sigma^2)$ -verteilt sind, wobei $\sigma^2 = \sigma_0^2 = s^2$ bekannt ist.

Dann ist

$$\frac{1}{\sqrt{n}} \frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu) \quad \mathcal{N}(0, 1) \text{ - verteilt,}$$

daher gilt

$$P_\mu \left[\left| \frac{1}{\sqrt{n}} \frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu) \right| \leq u_{\alpha/2} \right] = 1 - \alpha$$

für alle $\mu \in \mathbb{R}$, wobei $u_{\alpha/2}$ das $\alpha/2$ -Fraktil von $\mathcal{N}(0, 1)$ ist.

Mit

$$\begin{aligned} & \left| \frac{1}{\sqrt{n}} \frac{1}{\sigma} \sum_{i=1}^n (X_i - \mu) \right| \leq u_{\alpha/2} \\ \Leftrightarrow & -u_{\alpha/2} \leq \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{i=1}^n X_i - \mu \right) \leq u_{\alpha/2} \\ \Leftrightarrow & \frac{1}{n} \sum_{i=1}^n X_i - \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \leq \mu \leq \frac{1}{n} \sum_{i=1}^n X_i + \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \end{aligned}$$

folgt:

$$C(X_1, \dots, X_n) = \left[\frac{1}{n} \sum_{i=1}^n X_i - \frac{\sigma}{\sqrt{n}} u_{\alpha/2}, \frac{1}{n} \sum_{i=1}^n X_i + \frac{\sigma}{\sqrt{n}} u_{\alpha/2} \right]$$

ist Konfidenzintervall zum Konfidenzniveau $1 - \alpha$.

In Beispiel 6.3 folgt konkret mit $\bar{x} = 0.8, \sigma = s = \sqrt{0.04}$ und $\alpha = 0.05$ (also $u_{\alpha/2} = 1.96$): das gesuchte Konfidenzintervall zum Konfidenzniveau 0.95 ist

$$\begin{aligned} C(x_1, \dots, x_n) &= [0.8 - 1.96 \cdot \frac{0.2}{\sqrt{51}}, 0.8 + 1.96 \cdot \frac{0.2}{\sqrt{51}}] \\ &\approx [0.745 \quad , \quad 0.855]. \end{aligned}$$

Die Annahme von oben, dass die Varianz bekannt ist, ist unrealistisch. Ohne diese Annahme folgt mit

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

aus Satz 6.5:

$$\sqrt{n} \frac{\bar{X} - \mu}{S} \text{ ist } t_{n-1} - \text{verteilt.}$$

Also gilt

$$P_\mu \left[\left| \sqrt{n} \frac{\bar{X} - \mu}{S} \right| \leq t_{n-1; \alpha/2} \right] = 1 - \alpha$$

für alle $\mu \in \mathbb{R}$.

Analoge Rechnung wie oben ergibt:

$$C(X_1, \dots, X_n) = \left[\frac{1}{n} \sum_{i=1}^n X_i - \frac{S}{\sqrt{n}} t_{n-1; \alpha/2}, \frac{1}{n} \sum_{i=1}^n X_i + \frac{S}{\sqrt{n}} t_{n-1; \alpha/2} \right]$$

ist Konfidenzintervall zum Konfidenzniveau $1 - \alpha$.

Konkret folgt daraus mit den Zahlenwerten aus Beispiel 6.3, also mit $\bar{x} = 0.8$, $S = \sqrt{0.04}$, $n = 51$ und $\alpha = 0.05$ bzw. $t_{n-1; \alpha/2} = t_{50; 0.025} \approx 2.01$:

$$C(x_1, \dots, x_n) = [0.743, 0.856]$$

ist Konfidenzintervall zum Konfidenzniveau 0.95. Da die Varianz jetzt als unbekannt vorausgesetzt wird (und damit mehr Unsicherheit über die zugrundliegende Verteilung besteht) ist dieses Konfidenzintervall etwas größer als das obige.

Anwendung in den Beispielen 6.1 und 6.2:

X_1, \dots, X_n unabhängig $b(1, p)$ -verteilt mit $p \in (0, 1)$. Nach dem Zentralen Grenzwertsatz ist dann für große n

$$(*) \quad \frac{1}{\sqrt{n}} \frac{1}{\sqrt{p(1-p)}} \sum_{i=1}^n (X_i - p)$$

annähernd $\mathcal{N}(0, 1)$ -verteilt. Insbesondere gilt:

$$\lim_{n \rightarrow \infty} P_p \left[\left| \frac{1}{\sqrt{n}} \frac{1}{\sqrt{p(1-p)}} \sum_{i=1}^n (X_i - p) \right| \leq u_{\alpha/2} \right] = 1 - \alpha.$$

Man bezeichnet (*) daher als approximatives stochastisches Pivot.

Zur Konstruktion eines approximativen Konfidenzintervalls zum Konfidenzniveau

$1 - \alpha$ verwenden wir nun diejenigen $p \in (0, 1)$ mit

$$\begin{aligned} & \left| \frac{1}{\sqrt{n}} \frac{1}{\sqrt{p(1-p)}} \left(\sum_{i=1}^n X_i - n \cdot p \right) \right| \leq u_{\alpha/2} \\ \Leftrightarrow & \left(\sum_{i=1}^n X_i - n \cdot p \right)^2 \leq n \cdot p \cdot (1-p) \cdot u_{\alpha/2}^2 \\ \Leftrightarrow & \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 - p \cdot 2 \cdot \frac{1}{n} \sum_{i=1}^n X_i + p^2 \leq p(1-p) \cdot \frac{u_{\alpha/2}^2}{n} \\ \Leftrightarrow & \left(1 + \frac{u_{\alpha/2}^2}{n} \right) \cdot p^2 - \left(2 \cdot \frac{1}{n} \sum_{i=1}^n X_i + \frac{u_{\alpha/2}^2}{n} \right) \cdot p + \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \leq 0. \end{aligned}$$

Mit

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

folgt für die Nullstellen des obigen Polynoms:

$$\begin{aligned} p_{1,2} &= \frac{2\bar{X} + \frac{u_{\alpha/2}^2}{n} \pm \sqrt{(2\bar{X} + \frac{u_{\alpha/2}^2}{n})^2 - 4 \cdot \left(1 + \frac{u_{\alpha/2}^2}{n}\right) \cdot (\bar{X})^2}}{2(1 + \frac{u_{\alpha/2}^2}{n})} \\ &= \frac{2\bar{X} + \frac{u_{\alpha/2}^2}{n} \pm \sqrt{4(\bar{X})^2 + 4\bar{X} \cdot \frac{u_{\alpha/2}^2}{n} + \frac{u_{\alpha/2}^4}{n^2} - 4(\bar{X})^2 - 4(\bar{X})^2 \cdot \frac{u_{\alpha/2}^2}{n}}}{2(1 + \frac{u_{\alpha/2}^2}{n})} \\ &= \frac{\left(\bar{X} + \frac{u_{\alpha/2}^2}{2n}\right) \pm u_{\alpha/2} \cdot \sqrt{\bar{X} \cdot (1-\bar{X}) \cdot \frac{1}{n} + \frac{u_{\alpha/2}^2}{4n^2}}}{1 + \frac{u_{\alpha/2}^2}{n}} \end{aligned}$$

Damit erhält man

$$C(X_1, \dots, X_n) = [p_1, p_2]$$

als approximatives Konfidenzintervall zum Konfidenzniveau $1 - \alpha$.

In Beispiel 6.2 erhält man mit $\alpha = 0.05$, also $u_{\alpha/2} = 1.96$, und $n = 2000$ für den Anteil der Wähler, die für Rot-Grün stimmen:

$$\text{Hierbei ist } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{912}{2000} = 0.456,$$

also

$$C(X_1, \dots, X_n) = [0.434, 0.478]$$

Und für den Anteil der Wähler, die für Schwarz-Gelb stimmen, verwendet man

$$\bar{X} = \frac{927}{2000} = 0.4635$$

und erhält:

$$C(X_1, \dots, X_n) = [0.442, 0.485]$$

Bei großem Stichprobenumfang bietet sich wegen

$$\frac{1}{n} \sum_{i=1}^n X_i (1 - \frac{1}{n} \sum_{i=1}^n X_i) \rightarrow p \cdot (1 - p) \quad f.s.$$

an, die Rechnung zu vereinfachen, indem man in (*) $p(1 - p)$ durch $\bar{X}(1 - \bar{X})$ ersetzt. Damit erhält man als approximatives Konfidenzintervall zum Konfidenzniveau $1 - \alpha$:

$$C(X_1, \dots, X_n) = \left[\bar{X} - \frac{\sqrt{\bar{X} \cdot (1 - \bar{X})}}{\sqrt{n}} \cdot u_{\alpha/2}, \bar{X} + \frac{\sqrt{\bar{X} \cdot (1 - \bar{X})}}{\sqrt{n}} \cdot u_{\alpha/2} \right]$$

Anwendung in Beispiel 6.2 ergibt unter Beachtung von $\alpha = 0.05$, also $u_{\alpha/2} = 1.96$, $n = 374448 + 396296 = 770744$ und

$$\bar{x} = \frac{396296}{770744} \approx 0.5142$$

das folgende approximative Konfidenzintervall für die wahrscheinlichkeit einer ‘‘Jungengeburt’’:

$$C(X_1, \dots, X_n) \approx [0.512, 0.516].$$

6.4 Konstruktion von Bereichsschätzungen mit Hilfe von statistischen Tests

Seien X_1, \dots, X_n unabhängig und identisch verteilt mit

$$P_{X_1} = w_{\vartheta_0} \in \{w_{\vartheta} : \vartheta \in \Theta\}$$

und sei $g : \Theta \rightarrow \mathbb{R}$.

Ist dann

$$A(g(\vartheta_0))$$

der “Nicht-Ablehnungsbereich von H_0 ” eines Tests zum Niveau α für

$$H_0 : g(\vartheta) = g(\vartheta_0) \text{ versus } H_1 : g(\vartheta) \neq g(\vartheta_0),$$

so gilt für alle $\vartheta \in \Theta$:

$$P_\vartheta[(X_1, \dots, X_n) \in A(g(\vartheta))] \geq 1 - \alpha.$$

Mittels

$$g(\vartheta) \in C(X_1, \dots, X_n) :\Leftrightarrow (X_1, \dots, X_n) \in A(g(\vartheta))$$

lässt sich daraus ein Konfidenzbereich konstruieren.

Beispiel 6.4:

Die Zufallsvariablen X_1, \dots, X_n seien unabhängig $\mathcal{N}(\mu, \sigma^2)$ -verteilt mit μ und σ^2 unbekannt. Gesucht ist ein Konfidenzbereich für σ^2 zum Konfidenzniveau $1 - \alpha$.

Wir betrachten dazu in der obigen Situation das Testproblem

$$H_0 : \sigma^2 = \sigma_0^2 \text{ versus } H_1 : \sigma^2 \neq \sigma_0^2$$

für $\sigma_0 > 0$ fest. Ein Test zum Niveau α ist dann

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \text{falls } \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 > \chi_{n-1; \alpha/2}^2 \\ & \text{oder } \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 < \chi_{n-1; 1-\alpha/2}^2 \\ 0 & \text{sonst .} \end{cases}$$

H_0 wird hier nicht abgelehnt, falls gilt:

$$\chi_{n-1; 1-\frac{\alpha}{2}}^2 \leq \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 \leq \chi_{n-1; \alpha/2}^2,$$

also falls gilt:

$$\sigma_0^2 \in \left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1; \alpha/2}^2}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1; 1-\frac{\alpha}{2}}^2} \right],$$

und wir erhalten als Konfidenzintervall zum Konfidenzniveau $1 - \alpha$

$$C(X_1, \dots, X_n) = \left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1; \alpha/2}^2}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1; 1-\alpha/2}^2} \right].$$

7 Einige nichtparametrische Testverfahren

Die Schätz- und Testverfahren in den Kapiteln 3 und 4 beruhen auf der Annahme, dass die zugrundeliegende Verteilung bis auf den unbekanntem Wert eines Parameters bekannt ist. Im Folgenden stellen wir einige Tests zur Überprüfung des Wahrheitsgehalts dieser Annahme vor.

Beispiel 7.1. Zufällige Auswahl von 10 Pkw eines festen Typs ergab den folgenden Benzinverbrauch in l/100 km: 10.8, 11.3, 10.4, 9.8, 10.0, 10.6, 11.0, 10.5, 9.5, 11.2. Ist der Benzinverbrauch in l/100 km normalverteilt mit Erwartungswert $\mu = 10$ und Varianz $\sigma^2 = 1$?

Beispiel 7.2. In Beispiel 3.2 wurden die Toten durch Hufschlag in preußischen Kavallerieregimentern beschrieben durch:

# Tote/Jahr	0	1	2	3	4	≥ 5
# Regimenter	109	65	22	3	1	0

Kann man diese Anzahlen sinnvollerweise durch eine Poisson-Verteilung approximieren?

7.1 Der Test von Kolmogoroff-Smirnow

Um festzustellen, ob eine gegebene Stichprobe X_1, \dots, X_n von einer Verteilung mit vorgegebener Verteilungsfunktion $F_0 : \mathbb{R} \rightarrow \mathbb{R}$ stammt, vergleichen wir die empirische Verteilungsfunktion

$$F_n : \mathbb{R} \rightarrow \mathbb{R}, F_n(t) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(X_i)$$

mit F_0 .

Nach dem Satz von Glivenko-Cantelli (Satz 2.1) gilt

$$\sup_{t \in \mathbb{R}} |F_n(t) - F_0(t)| \rightarrow 0 \quad f.s.,$$

sofern die Zufallsvariablen X_1, X_2, \dots unabhängig und identisch verteilt sind mit Verteilungsfunktion F_0 . Dies führt auf die naheliegende Idee,

$$H_0 : F = F_0$$

abzulehnen, falls die Teststatistik

$$T_n(X_1, \dots, X_n) = \sup_{t \in \mathbb{R}} |F_n(t) - F_0(t)|$$

einen kritischen Wert $c \in \mathbb{R}_+$ übersteigt.

Dabei kann das Supremum in der Teststatistik leicht berechnet werden, da $F_n(t)$ stückweise konstant mit Sprungstellen an den X_1, \dots, X_n ist, $F_n(t)$ und $F_0(t)$ monoton wachsend sind, und daher das Supremum entweder an den Punkten X_1, \dots, X_n oder an den “linksseitigen Grenzwerten” dieser Punkte angenommen wird.

Die Festlegung des kritischen Wertes c in Abhängigkeit des Niveaus α benötigt Kenntnisse über die Verteilung von $T_n(X_1, \dots, X_n)$ bei Gültigkeit von H_0 . Dazu ist der folgende Satz hilfreich:

Satz 7.1

Sind X_1, \dots, X_n unabhängig und identisch verteilte reelle Zufallsvariablen mit stetiger Verteilungsfunktion F , so hängt die Verteilung von

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$$

nicht von F ab.

Damit kann bei **stetigem**

F_0 das α -Fraktile $Q_{n;\alpha}$ der Verteilung Q_n der Zufallsvariable

$$\sup_{t \in [0,1]} \left| \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(U_i) - t \right|$$

(mit unabhängigen und auf $[0, 1]$ gleichverteilten Zufallsvariablen U_1, \dots, U_n) als kritischer Wert c des obigen Tests verwendet werden. Dieses ist zum Teil vertafelt und kann auch durch Simulationen erzeugt werden.

Damit testen wir bei gegebenen Werten der Stichprobe x_1, \dots, x_n , gegebenem **stetigem** F_0 und $\alpha \in (0, 1)$

$$H_0 : F = F_0 \text{ versus } H_1 : F \neq F_0$$

zum Niveau α mittels

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1 & , \text{ falls } \sup_{t \in \mathbb{R}} |F_n(t) - F_0(t)| > Q_{n;\alpha} \\ 0 & , \text{ sonst} \end{cases}$$

wobei

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(x_i).$$

Bei Anwendung in Beispiel 7.1 erhalten wir

$$T_n(x_1, \dots, x_n) \approx 0.3555$$

und – mittels Simulationen –

$$Q_{10;0.05} \approx 0.41.$$

Wegen $T_n(x_1, \dots, x_n) < Q_{10;0.05}$ kann hier H_0 zum Niveau $\alpha = 0.05$ nicht abgelehnt werden.

Bemerkung: Wie immer bei Tests zum Niveau α ist dieses Ergebnis eigentlich nicht aussagekräftig, da der Fehler 2. Art hier nicht kontrolliert wird.

Beweis von Satz 7.1:

1. *Schritt:* Wir zeigen: Für eine reelle Zufallsvariable X mit stetiger Verteilungsfunktion F ist $F(X)$ auf $[0, 1]$ gleichverteilt.

Dazu: Setze

$$F^{-1}(u) = \min\{t \in \mathbb{R} | F(t) \geq u\} \quad (u \in (0, 1))$$

Dann gilt:

(i) F^{-1} ist monoton wachsend.

(Denn aus $u \leq v$ folgt $\{t \in \mathbb{R} | F(t) \geq u\} \supseteq \{t \in \mathbb{R} | F(t) \geq v\}$ und daher

$$\min\{t \in \mathbb{R} | F(t) \geq u\} \leq \min\{t \in \mathbb{R} | F(t) \geq v\}.)$$

(ii) Für alle $u \in (0, 1)$ gilt $F(F^{-1}(u)) = u$.

(Denn aus

$$F^{-1}(u) \in \{t \in \mathbb{R} | F(t) \geq u\}$$

folgt

$$F(F^{-1}(u)) \geq u,$$

und wäre

$$F(F^{-1}(u)) > u,$$

so wäre (wegen Stetigkeit von F) für $\varepsilon > 0$ klein auch

$$F(F^{-1}(u) - \varepsilon) \geq u,$$

woraus der Widerspruch

$$F^{-1}(u) \leq F^{-1}(u) - \varepsilon$$

folgen würde.)

$$(iii) \quad \forall u \in (0, 1) \quad \forall x \in \mathbb{R}: F^{-1}(u) \leq x \Leftrightarrow u \leq F(x)$$

(denn “ \Rightarrow ” folgt mit (ii) aus der Monotonie von F , und “ \Leftarrow ” gilt, da $u \leq F(x)$ die Beziehung

$$x \in \{t \in \mathbb{R} | F(t) \geq u\}$$

impliziert, woraus

$$F^{-1}(u) = \min\{t \in \mathbb{R} | F(t) \geq u\} \leq x$$

folgt).

Damit gilt für $u \in (0, 1)$ beliebig:

$$\begin{aligned} P[F(X) \geq u] &= P[X \geq F^{-1}(u)] \quad (\text{nach (iii)}) \\ &= P[X > F^{-1}(u)] \\ &\quad (\text{da } F \text{ stetig, und daher } P[X = x] = 0 \text{ für alle } x \in \mathbb{R}) \\ &= 1 - P[X \leq F^{-1}(u)] \\ &= 1 - F(F^{-1}(u)) \\ &\quad (\text{nach Definition von } F) \\ &= 1 - u \quad (\text{nach (ii)}). \end{aligned}$$

Mit $P[F(X) \leq v] = 0$ für $v < 0$ und $P[F(X) \leq v] = 1$ für $v \geq 1$ (was aus $F(x) \in [0, 1]$ ($x \in \mathbb{R}$) folgt) und der rechtsseitigen Stetigkeit der Verteilungsfunktion folgt daraus

$$P[F(X) \leq v] = \begin{cases} 0 & , \quad v < 0, \\ v & , \quad 0 \leq v \leq 1, \\ 1 & , \quad v > 1, \end{cases}$$

also ist $F(X)$ auf $[0, 1]$ gleichverteilt.

2. Schritt: Wir zeigen:

Mit Wahrscheinlichkeit Eins gilt:

$$X_i \leq t \Leftrightarrow F(X_i) \leq F(t).$$

Dazu:

Wegen der Monotonie von F gilt

$$[X_i \leq t] \subseteq [F(X_i) \leq F(t)].$$

Da nun aber auch

$$F(t) = P[X_i \leq t] \stackrel{s.o.}{=} P[F(X_i) \leq F(t)] \stackrel{\text{Schritt 1}}{=} F(t),$$

also

$$P[X_i \leq t] = P[F(X_i) \leq F(t)]$$

gilt, stimmen die beiden Mengen bis auf eine Menge von Maß Null überein, was zu zeigen war.

3. Schritt: Wir zeigen die Behauptung des Satzes.

Dazu beachten wir, dass mit Wahrscheinlichkeit Eins gilt:

$$\begin{aligned} \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| &= \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(X_i) - F(t) \right| \\ &= \sup_{t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n I_{(-\infty, F(t)]}(F(X_i)) - F(t) \right| \quad (\text{nach Schritt 2}) \\ &= \sup_{u \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n I_{(-\infty, u]}(F(X_i)) - u \right|, \end{aligned}$$

wobei die letzte Gleichheit aus der (aus der Stetigkeit von der Verteilungsfunktion F folgenden) Beziehung

$$(0, 1) \subseteq \{F(t) : t \in \mathbb{R}\} \subseteq [0, 1]$$

folgt.

Die Verteilung von

$$\sup_{u \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^n I_{(-\infty, u]}(F(X_i)) - u \right|$$

hängt nun nur von der Verteilung von

$$F(X_1), \dots, F(X_n)$$

ab, die nach Schritt 1 nicht von F abhängt. □

Satz 7.2

Sind X_1, \dots, X_n unabhängige und identisch verteilte Zufallsvariablen mit stetiger Verteilungsfunktion F , und ist F_n die zu X_1, \dots, X_n gehörende empirische Verteilungsfunktion, so gilt für jedes $\lambda > 0$:

$$\lim_{n \rightarrow \infty} P \left\{ \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \leq \frac{\lambda}{\sqrt{n}} \right\} = Q(\lambda),$$

wobei

$$Q(\lambda) = 1 - 2 \cdot \sum_{j=1}^{\infty} (-1)^{j-1} \cdot e^{-2j^2 \cdot \lambda^2},$$

d. h.

$$\sqrt{n} \cdot \sup_{t \in \mathbb{R}} |F_n(t) - F(t)|$$

konvergiert nach Verteilung gegen eine reelle Zufallsvariable mit Verteilungsfunktion Q .

ohne Beweis.

Anwendung dieses Satzes ergibt den

Test von Kolmogoroff-Smirnow:

Lehne

$$H_0 : F = F_0$$

zum Niveau $\alpha \in (0, 1)$ ab, falls gilt

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| > \frac{\lambda_\alpha}{\sqrt{n}},$$

wobei $\lambda_\alpha \in \mathbb{R}_+$ so gewählt ist, dass gilt: $1 - Q(\lambda_\alpha) = \alpha$.

Die Werte von λ_α sind tabelliert, z. B. gilt

$$\lambda_{0.05} = 1.36 \text{ und } \lambda_{0.01} = 1.63.$$

Gemäß Satz 7.2 ist dieser Test bei stetigem F_0 für große n näherungsweise ein Test zum Niveau α .

Anwendung in Beispiel 7.1 mit $\alpha = 0.05$, $\frac{\lambda_\alpha}{\sqrt{n}} = \frac{1.36}{\sqrt{10}} \approx 0.43$ ergibt wegen

$$\sup_{t \in \mathbb{R}} |F_n(t) - F_0(t)| \approx 0.36 < 0.43 :$$

H_0 kann hier zum Niveau $\alpha = 0.05$ nicht abgelehnt werden.

7.2 Der χ^2 -Anpassungstest

X_1, \dots, X_n seien unabhängige identisch verteilte reelle Zufallsvariablen mit Verteilungsfunktion F . Für eine gegebene Verteilungsfunktion F_0 sei wieder zu testen:

$$(*) \quad H_0 : F = F_0 \text{ versus } H_1 : F \neq F_0.$$

Dazu unterteilen wir den Bildbereich \mathbb{R} von X_1 in messbare disjunkte Mengen C_1, \dots, C_r mit

$$\mathbb{R} = \bigcup_{j=1}^r C_j \text{ und } C_i \cap C_j = \emptyset \text{ für } i \neq j.$$

Wir setzen

$$p_i^0 = P_{F=F_0}(X_1 \in C_i) \quad (i = 1, \dots, r)$$

und

$$p_i = P_F(X_1 \in C_i).$$

Anstelle von $(*)$ testen wir dann die "schwächeren Hypothesen"

$$H_0 : (p_1, \dots, p_r) = (p_1^0, \dots, p_r^0) \text{ versus } H_1 : (p_1, \dots, p_r) \neq (p_1^0, \dots, p_r^0).$$

Dazu setzen wir

$$Y_j = \sum_{i=1}^n 1_{C_j}(X_i) \quad (j = 1, \dots, r).$$

Dann ist Y_j $b(n, p_j)$ -verteilt ($j = 1, \dots, r$), und wegen

$$Y_1 + \dots + Y_r = n \quad f.s.$$

sind die Zufallsvariablen Y_1, \dots, Y_r **nicht** unabhängig (da aus den Werten von Y_1, \dots, Y_{r-1} der Wert von Y_r berechnet werden kann).

Genauer gilt:

$$P\{Y_1 = k_1, \dots, Y_r = k_r\} = \frac{n!}{k_1! \cdot \dots \cdot k_r!} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_r^{k_r}$$

für alle $k_1, \dots, k_r \in \mathbb{N}_0$ mit $k_1 + \dots + k_r = n$.

Man sagt: Der Zufallsvektor (Y_1, \dots, Y_r) ist multinomialverteilt mit Parametern n und p_1, \dots, p_r .

Bei Gültigkeit von H_0 ist

$$E_{F=F_0}\{Y_j\} = n \cdot p_j^0$$

der “erwartete Wert” der $b(n, p_j^0)$ -verteilten Zufallsvariablen Y_j . Zur Entscheidung zwischen H_0 und H_1 betrachten wir die Abweichung zwischen

$$Y_j \text{ und } n \cdot p_j^0 \quad (j = 1, \dots, r).$$

Hierzu gilt:

Satz 7.3 Bei Gültigkeit von $H_0 : (p_1, \dots, p_r) = (p_1^0, \dots, p_r^0)$ gilt:

$$T_n(X_1, \dots, X_n) = \sum_{j=1}^r \frac{(Y_j - n \cdot p_j^0)^2}{n \cdot p_j^0}$$

konvergiert für $n \rightarrow \infty$ nach Verteilung gegen eine χ_{r-1}^2 -verteilte Zufallsvariable.

Beweis: Zur Vereinfachung der Schreibweise schreiben wir im Folgenden

$$(p_1, \dots, p_r)$$

statt

$$(p_1^0, \dots, p_r^0).$$

Setze

$$Z_j = \frac{Y_j - n \cdot p_j}{\sqrt{n \cdot p_j}} \quad (j = 1, \dots, r).$$

Wir zeigen im Folgenden:

Es existieren Zufallsvariablen V_1, \dots, V_r mit $V_r = 0$ f.s. und V_1, \dots, V_{r-1} unabhängig $\mathcal{N}(0, 1)$ -verteilt, und es existiert eine orthogonale Matrix $\underline{\underline{A}}$ so, dass für

$$(U_1, \dots, U_r)^T = \underline{\underline{A}}^T (V_1, \dots, V_r)^T$$

gilt:

$$(Z_1, \dots, Z_r) \rightarrow^{\mathcal{D}} (U_1, \dots, U_r) \quad (n \rightarrow \infty).$$

Daraus folgt die Behauptung, denn nach dem Satz von der stetigen Abbildung impliziert dies

$$\sum_{j=1}^r \frac{(Y_j - n \cdot p_j)^2}{n \cdot p_j} = \sum_{j=1}^r Z_j^2 \rightarrow^{\mathcal{D}} \sum_{j=1}^r U_j^2 \quad (n \rightarrow \infty),$$

und wegen

$$\begin{aligned}
 \sum_{j=1}^r U_j^2 &= (V_1, \dots, V_r) \cdot \underline{\underline{A}} \cdot \underline{\underline{A}}^T \begin{pmatrix} V_1 \\ \vdots \\ V_r \end{pmatrix} \\
 &= (V_1, \dots, V_r) \begin{pmatrix} V_1 \\ \vdots \\ V_r \end{pmatrix} \quad (\text{da } \underline{\underline{A}} \text{ orthogonal ist}) \\
 &= \sum_{j=1}^r V_j^2 \\
 &= \sum_{j=1}^{r-1} V_j^2 \quad (\text{da } V_r = 0 \text{ f.s.})
 \end{aligned}$$

ist die Grenzverteilung die Summe der Quadrate von $(r - 1)$ unabhängigen standardnormalverteilten Zufallsvariablen und damit eine χ_{r-1}^2 -Verteilung.

Also bleibt die Behauptung von oben zu zeigen:

Im ersten Schritt des Beweises bestimmen wir die charakteristische Funktion von (Y_1, \dots, Y_r) .

Dazu beachten wir, dass wegen

$$(Y_1, \dots, Y_r) = \sum_{i=1}^n (1_{C_1}(X_i), \dots, 1_{C_r}(X_i)),$$

(Y_1, \dots, Y_r) die Summe von n unabhängigen multinomialverteilten Zufallsvektoren mit Parametern 1 und p_1, \dots, p_r ist. Jeder einzelne Summand hat die charakteristische Funktion

$$\begin{aligned}
 \Psi^{(1)}(u) &= E \exp \left(i(u_1 \cdot 1_{C_1}(X_1) + u_2 \cdot 1_{C_2}(X_1) + \dots + u_r \cdot 1_{C_r}(X_1)) \right) \\
 &= E \left(\sum_{j=1}^r e^{iu_j} \cdot 1_{C_j}(X_1) \right) \\
 &\quad (\text{da } X_1 \text{ genau in einer der Mengen } C_1, \dots, C_r \text{ enthalten ist}) \\
 &= \sum_{j=1}^r e^{iu_j} \cdot p_j,
 \end{aligned}$$

also hat (Y_1, \dots, Y_r) die charakteristische Funktion

$$\begin{aligned}
 \Psi_n(u_1, \dots, u_r) &= (\Psi^{(1)}(u_1, \dots, u_r))^n \\
 &= \left(\sum_{j=1}^r p_j \cdot e^{iu_j} \right)^n.
 \end{aligned}$$

Im zweiten Schritt des Beweises bestimmen wir die charakteristische Funktion φ_n von (Z_1, \dots, Z_r) .

Wegen

$$Z_j = \frac{Y_j - n \cdot p_j}{\sqrt{n \cdot p_j}}$$

gilt

$$\begin{aligned} \varphi_n(u_1, \dots, u_r) &= E \exp\left(i \cdot \sum_{j=1}^r u_j \cdot Z_j\right) \\ &= E \exp\left(i \sum_{j=1}^r \frac{u_j}{\sqrt{n \cdot p_j}} \cdot Y_j - i \cdot \sum_{j=1}^r u_j \cdot \sqrt{n \cdot p_j}\right) \\ &= \exp\left(-i \cdot \sum_{j=1}^r u_j \cdot \sqrt{n \cdot p_j}\right) \cdot \Psi_n\left(\frac{u_1}{\sqrt{n \cdot p_1}}, \dots, \frac{u_r}{\sqrt{n \cdot p_r}}\right) \end{aligned}$$

und durch Einsetzen des Resultates von Schritt 1 erhält man

$$\varphi_n(u_1, \dots, u_r) = e^{-i \cdot \sum_{j=1}^r u_j \cdot \sqrt{n \cdot p_j}} \cdot \left(\sum_{j=1}^r p_j \cdot e^{i \frac{u_j}{\sqrt{n \cdot p_j}}}\right)^n.$$

Im dritten Schritt des Beweises zeigen wir für alle $u \in \mathbb{R}^r$:

$$\lim_{n \rightarrow \infty} \varphi_n(u) = \varphi^*(u)$$

mit

$$\varphi^*(u_1, \dots, u_r) = \exp\left(-\frac{1}{2} \cdot \left[\sum_{j=1}^r u_j^2 - \left(\sum_{j=1}^r u_j \cdot \sqrt{p_j}\right)^2\right]\right).$$

Dazu beachten wir

$$\begin{aligned} \log \varphi_n(u_1, \dots, u_r) &= -i \cdot \sqrt{n} \sum_{j=1}^r u_j \cdot \sqrt{p_j} + n \cdot \log\left(\sum_{j=1}^r p_j \cdot e^{i \frac{u_j}{\sqrt{n \cdot p_j}}}\right) \\ &= -i \cdot \sqrt{n} \cdot \sum_{j=1}^r u_j \cdot \sqrt{p_j} + n \cdot \log(1 + z_n), \end{aligned}$$

wobei wir

$$e^{i \frac{u_j}{\sqrt{n \cdot p_j}}} = 1 + \frac{i}{\sqrt{n}} \cdot \frac{u_j}{\sqrt{p_j}} - \frac{1}{2n} \cdot \frac{u_j^2}{p_j} + O\left(\frac{1}{n}\right)$$

verwendet haben und

$$z_n = p_j \cdot \left(\frac{i}{\sqrt{n}} \frac{u_j}{\sqrt{p_j}} - \frac{1}{2n} \frac{u_j^2}{p_j}\right) = \frac{i}{\sqrt{n}} \sum_{j=1}^r u_j \cdot \sqrt{p_j} - \frac{1}{2n} \sum_{j=1}^r u_j^2 + O\left(\frac{1}{n}\right)$$

gesetzt haben.

Unter Beachtung von der aus der Taylorentwicklung von $u \mapsto \log(1+u)$ folgenden Beziehung

$$\begin{aligned} n \cdot \log(1+z_n) &= n \cdot (0 + z_n - \frac{1}{2}z_n^2 + \frac{1}{3}z_n^3 - + \dots) \\ &= n \cdot z_n - \frac{1}{2}nz_n^2 + n \cdot z_n^2 \left(\frac{1}{3}z_n - \frac{1}{4}z_n^2 + - \dots \right), \end{aligned}$$

$z_n \rightarrow 0$ ($n \rightarrow \infty$) und $n \cdot z_n^2$ beschränkt (was unmittelbar aus der Definition von z_n folgt), sowie der daraus folgenden Beziehung

$$n \cdot z_n^2 \cdot \left(\frac{1}{3}z_n - \frac{1}{4}z_n^2 + - \dots \right) \rightarrow 0 \quad (n \rightarrow \infty)$$

erhalten wir:

$$\begin{aligned} \log \varphi_n(u_1, \dots, u_r) &= -i \cdot \sqrt{n} \sum_{j=1}^r u_j \cdot \sqrt{p_j} + n \cdot z_n - \frac{1}{2}nz_n^2 + O(1) \\ &= -i \cdot \sqrt{n} \cdot \sum_{j=1}^r u_j \cdot \sqrt{p_j} + i \cdot \sqrt{n} \cdot \sum_{j=1}^r u_j \cdot \sqrt{p_j} - \frac{1}{2} \sum_{j=1}^r u_j^2 \\ &\quad - \frac{1}{2} \cdot n \cdot \frac{i^2}{n} \left(\sum_{j=1}^r u_j \cdot \sqrt{p_j} \right)^2 + O(1) \\ &= -\frac{1}{2} \sum_{j=1}^r u_j^2 + \frac{1}{2} \left(\sum_{j=1}^r u_j \cdot \sqrt{p_j} \right)^2 + O(1), \text{ w.z.z.w.} \end{aligned}$$

Im vierten Schritt des Beweises zeigen wir, dass eine orthogonale $r \times r$ -Matrix $\underline{\underline{A}}$ so existiert, dass für

$$\begin{pmatrix} v_1 \\ \vdots \\ v_r \end{pmatrix} = \underline{\underline{A}} \begin{pmatrix} u_1 \\ \vdots \\ u_r \end{pmatrix}$$

gilt:

$$\varphi^*(u_1, \dots, u_r) = \exp \left(-\frac{1}{2} \sum_{j=1}^{r-1} v_j^2 \right).$$

Dazu wählen wir eine orthogonale Matrix $\underline{\underline{A}} = (a_{ij})_{1 \leq i, j \leq r}$ so, dass die letzte Zeile durch den Einheitsvektor (!)

$$(a_{r1}, \dots, a_{rr}) = (\sqrt{p_1}, \dots, \sqrt{p_r})$$

gegeben ist. Dann gilt

$$\begin{aligned}
\sum_{j=1}^{r-1} v_j^2 &= \sum_{j=1}^r v_j^2 - v_r^2 \\
&= \sum_{j=1}^r u_j^2 - v_r^2 \quad (\text{da } \underline{\underline{A}} \text{ orthogonal ist)} \\
&= \sum_{j=1}^r u_j^2 - \left(\sum_{j=1}^r a_{rj} \cdot u_j \right)^2 \\
&= \sum_{j=1}^r u_j^2 - \left(\sum_{j=1}^r \sqrt{p_j} \cdot u_j \right)^2
\end{aligned}$$

nach Wahl der letzten Zeile von $\underline{\underline{A}}$.

Im fünften Schritt des Beweises zeigen wir die Behauptung. Dazu wählen wir unabhängige $\mathcal{N}(0, 1)$ -verteilte Zufallsvariablen V_1, \dots, V_{r-1} , setzen $V_r = 0$ und definieren

$$\begin{pmatrix} U_1 \\ \vdots \\ U_r \end{pmatrix} = \underline{\underline{A}}^T \begin{pmatrix} V_1 \\ \vdots \\ V_r \end{pmatrix}$$

mit $\underline{\underline{A}}$ wie in Schritt 4 des Beweises. Dann hat $(U_1, \dots, U_r)^T$ die charakteristische Funktion

$$\begin{aligned}
\varphi_{(U_1, \dots, U_r)}(u_1, \dots, u_r) &= E \left\{ \exp \left(i \cdot \sum_{j=1}^r u_j \cdot U_j \right) \right\} \\
&= E \left\{ \exp \left(i \cdot (u_1, \dots, u_r) \begin{pmatrix} U_1 \\ \vdots \\ U_r \end{pmatrix} \right) \right\} \\
&= E \left\{ \exp \left(i \cdot (v_1, \dots, v_r) \underline{\underline{A}} \underline{\underline{A}}^T \begin{pmatrix} V_1 \\ \vdots \\ V_r \end{pmatrix} \right) \right\}
\end{aligned}$$

$$\begin{aligned}
&= E \left\{ \exp \left(i(v_1, \dots, v_r) \begin{pmatrix} V_1 \\ \vdots \\ V_r \end{pmatrix} \right) \right\} \\
&\quad (\text{da } \underline{\underline{A}} \underline{\underline{A}}^T = \underline{\underline{1}}) \\
&= \prod_{j=1}^{r-1} E \exp(i \cdot v_j \cdot V_j) \\
&\quad (\text{da } V_1, \dots, V_{r-1} \text{ unabhängig und } V_r = 0) \\
&= \prod_{j=1}^{r-1} \exp\left(-\frac{1}{2}v_j^2\right) \quad (\text{da } V_j \mathcal{N}(0, 1) \text{ - verteilt ist}) \\
&= \exp\left(-\frac{1}{2} \sum_{j=1}^{r-1} v_j^2\right) \\
&= \varphi^*(u_1, \dots, u_r)
\end{aligned}$$

nach Schritt 4.

Daher ist φ^* die charakteristische Funktion zu $(U_1, \dots, U_r)^T$, und mit dem Stetigkeitssatz von Lévy-Cramér folgt die Behauptung mit Schritt 3. \square

Satz 7.3 führt aus den χ^2 -Anpassungstest:

Lehne H_0 ab, falls

$$T_n(X_1, \dots, X_n) > \chi_{r-1; \alpha}^2,$$

wobei $\chi_{r-1; \alpha}^2$ das α -Fraktile der χ_{r-1}^2 -Verteilung ist.

Nach Satz 7.3 ist dieser Test für $n \rightarrow \infty$ ein Test zum Niveau α .

Bei der Berechnung der Prüfgröße ist hilfreich:

$$\begin{aligned}
T_n(X_1, \dots, X_n) &= \sum_{j=1}^r \frac{Y_j^2 - 2 \cdot n \cdot p_j^0 \cdot Y_j + n^2 \cdot (p_j^0)^2}{n \cdot p_j^0} \\
&= \sum_{j=1}^r \frac{Y_j^2}{n \cdot p_j^0} - 2 \cdot \sum_{j=1}^r Y_j + n \cdot \sum_{j=1}^r p_j^0 \\
&= \sum_{j=1}^r \frac{Y_j^2}{n \cdot p_j^0} - n,
\end{aligned}$$

da

$$\sum_{j=1}^r Y_j = n \text{ und } \sum_{j=1}^r p_j^0 = 1.$$

Bemerkung: Beim χ^2 -Anpassungstest gibt es die folgende Faustregel: C_1, \dots, C_r und n sollten so gewählt sein, dass für

$$p_j^0 = P_{F=F_0}(X_1 \in C_j) \text{ gilt: } n \cdot p_j^0 \geq 5 \quad (j = 1, \dots, r).$$

Oft möchte man wissen, ob eine Verteilung aus einer vorgegebenen Klasse von Verteilung stammt, z. B. ob eine $\pi(\vartheta)$ -Verteilung für ein $\vartheta \in \Theta$ vorliegt. Dann kann man wie folgt vorgehen:

Sei $\{w_\vartheta : \vartheta \in \Theta\}$ mit $\Theta \subseteq \mathbb{R}$ die gegebene Klasse von Verteilungen. Setze

$$p_j^0(\vartheta) = P_\vartheta\{X_1 \in C_j\} = w_\vartheta(C_j) \quad (j = 1, \dots, r).$$

Seien y_1, \dots, y_r die beobachteten Werte von Y_1, \dots, Y_r . Dann kann ϑ mit Hilfe des Maximum-Likelihood-Prinzips geschätzt werden durch

$$\hat{\vartheta} = \operatorname{argmax}_{\vartheta \in \Theta} \frac{n!}{y_1! \dots y_r!} (p_1^0(\vartheta))^{y_1} \cdot \dots \cdot (p_r^0(\vartheta))^{y_r}.$$

Anschließend kann

$$H_0 : P_{X_1} = w_{\hat{\vartheta}} \text{ versus } H_1 : P_{X_1} \neq w_{\hat{\vartheta}}$$

durch Betrachtung von

$$T_n(X_1, \dots, X_n) = \sum_{j=1}^r \frac{(Y_j - n \cdot p_j^0(\hat{\vartheta}))^2}{n \cdot p_j^0(\hat{\vartheta})}$$

getestet werden.

Man kann zeigen:

Für n groß ist bei Gültigkeit von H_0

$$T_n(X_1, \dots, X_n) \text{ annähernd } \chi_{r-1-1}^2$$

verteilt.

Daher lehnt man hier H_0 ab, falls

$$T_n(X_1, \dots, X_n) > \chi_{r-2;\alpha}^2,$$

wobei $\chi_{r-2;\alpha}^2$ das α -Fraktile der χ^2 -Verteilung mit $r - 2$ Freiheitsgraden ist.

Anwendung in Beispiel 7.2

Hier möchten wir wissen, ob die Anzahl der Toten durch Hufschlag wirklich durch eine $\pi(\vartheta)$ -verteilte Zufallsvariable beschrieben werden kann.

Die Darstellung der Daten legt die Klasseneinteilung

$$C_1 = (-\infty, 0], C_2 = (0, 1], C_3 = (1, 2], C_4 = (2, 3], C_5 = (3, 4]$$

und $C_6 = (4, \infty)$ nahe.

Für $j < 6$ gilt hier

$$\begin{aligned} p_j^0(\vartheta) &= P(\text{"Poisson}(\vartheta)\text{-verteilte ZV nimmt Wert } j - 1 \text{ an"}) \\ &= \frac{\vartheta^{j-1}}{(j-1)!} \cdot e^{-\vartheta}. \end{aligned}$$

Unter Beachtung von $y_6 = 0$ ergibt sich als Maximum-Likelihood-Schätzer

$$\begin{aligned} \hat{\vartheta} &= \operatorname{argmax}_{\vartheta \in (0, \infty)} \frac{n!}{y_1! \cdot y_2! \cdot \dots \cdot y_6!} \cdot (p_1^0(\vartheta))^{y_1} \cdot \dots \cdot (p_6^0(\vartheta))^{y_6} \\ &= \operatorname{argmax}_{\vartheta \in (0, \infty)} \frac{200}{109! \cdot 65! \cdot 22! \cdot 3! \cdot 1! \cdot 0!} \cdot \left(\frac{\vartheta^0}{0!} \cdot e^{-\vartheta} \right)^{109} \cdot \left(\frac{\vartheta^1}{1!} \cdot e^{-\vartheta} \right)^{65} \\ &\quad \cdot \left(\frac{\vartheta^2}{2!} \cdot e^{-\vartheta} \right)^2 \cdot \left(\frac{\vartheta^3}{3!} \cdot e^{-\vartheta} \right)^3 \cdot \left(\frac{\vartheta^4}{4!} \cdot e^{-\vartheta} \right)^1 \\ &= \operatorname{argmax}_{\vartheta \in (0, \infty)} \operatorname{const}(n, y_1, \dots, y_r) \cdot \vartheta^{0 \cdot 109 + 1 \cdot 65 + 2 \cdot 22 + 3 \cdot 4 + 4 \cdot 1} \cdot e^{-200 \cdot \vartheta} \\ &= \operatorname{argmax}_{\vartheta \in (0, \infty)} \operatorname{const}(n, y_1, \dots, y_r) \cdot \vartheta^{122} \cdot e^{-200 \cdot \vartheta} \\ &= \frac{122}{200} = 0.61 \end{aligned}$$

(da $f(\vartheta) = \vartheta^k \cdot e^{-n \cdot \vartheta}$ als Maximalstelle in $(0, \infty)$ $\hat{\vartheta} = \frac{k}{n}$ hat).

Damit gilt

$$p_j^0(\hat{\vartheta}) = \frac{0.61^{j-1}}{(j-1)!} \cdot e^{-0.61} \quad (j = 1, \dots, 5)$$

und mit $n = 200$ erhalten wir:

# Toter/Jahr	0	1	2	3	4	≥ 5
# Regimenter	109	65	22	3	1	0
Gehört zur Klasse	C_1	C_2	C_3	C_4	C_5	C_6
$p_j^0(\hat{\vartheta})$	0.543	0.331	0.101	0.02	0.003	0.002
Erwarteter Wert $n \cdot p_j^0(\hat{\vartheta})$	108.7	66.3	20.2	4.1	0.63	0.4

Man sieht, dass bei dieser Klasseneinteilung C_4, C_5 und C_6 nicht die Faustregel $n \cdot p_j^0 \geq 5$ erfüllen.

Daher verwenden wir für den χ^2 -Test die neue Klasseneinteilung.

$$\begin{aligned}\bar{C}_1 &= (-\infty, 0] \\ \bar{C}_2 &= (0, 1] \\ \bar{C}_3 &= (1, 2] \\ \bar{C}_4 &= (2, \infty].\end{aligned}$$

Für diese Klasseneinteilung gilt

$$p_j^0(\vartheta) = \frac{\vartheta^{(j-1)}}{(j-1)!} \cdot e^{-\vartheta} \text{ für } \vartheta \in \{1, 2, 3\}$$

und

$$\begin{aligned}p_4^0(\vartheta) &= 1 - p_1^0(\vartheta) - p_2^0(\vartheta) - p_3^0(\vartheta) \\ &= 1 - e^{-\vartheta} \left(1 + \vartheta + \frac{\vartheta^2}{2}\right)\end{aligned}$$

Damit Maximum-Likelihood-Schätzer bei dieser Klasseneinteilung

$$\begin{aligned}\hat{\vartheta} &= \operatorname{argmax}_{\vartheta \in (0, \infty)} \frac{n!}{y_1! \cdot y_2! \cdot y_3! \cdot (y_4 + y_5 + y_6)!} \cdot (p_1^0(\vartheta))^{y_1} \cdot (p_2^0(\vartheta))^{y_2} \cdot (p_3^0(\vartheta))^{y_3} \\ &\quad \cdot (p_4^0(\vartheta))^{y_4 + y_5 + y_6} \\ &= \operatorname{argmax}_{\vartheta \in (0, \infty)} \operatorname{const} \cdot (e^{-\vartheta})^{109} \cdot (\vartheta \cdot e^{-\vartheta})^{65} \cdot \left(\frac{\vartheta^2}{2} \cdot e^{-\vartheta}\right)^{22} \cdot \left(1 - e^{-\vartheta} \left(1 + \vartheta + \frac{\vartheta^2}{2}\right)\right)^4 \\ &= \operatorname{argmax}_{\vartheta \in (0, \infty)} \operatorname{const}' \cdot \vartheta^{0 \cdot 109 + 1 \cdot 65 + 2 \cdot 22} \cdot e^{-196\vartheta} \left(1 - e^{-\vartheta} \left(1 + \vartheta + \frac{\vartheta^2}{2}\right)\right)^4 \\ &\approx 0.61 \quad (!)\end{aligned}$$

Damit gilt jetzt für die erwarteten Werte

j	1	2	3	4
$n \cdot p_j^0(\hat{\vartheta})$	108.7	66.3	20.2	5

und die Prüfgröße berechnet sich zu

$$\begin{aligned} T(x_1, \dots, x_n) &= \sum_{j=1}^4 \frac{\bar{y}_j^2}{n \cdot p_j^0(\hat{\vartheta})} - n \\ &\approx \frac{109^2}{108.7} + \frac{65^2}{66.3} + \frac{22^2}{20.2} + \frac{4^2}{5} - 200 \\ &\approx 0.187 \end{aligned}$$

Damit ist

$$\begin{aligned} T(x_1, \dots, x_n) &< \chi_{r-2; \alpha}^2 = \chi_{4-2; 0.05}^2 = \chi_{2; 0.05}^2 \\ &\approx 5.99 \end{aligned}$$

und man kommt zu dem Schluss:

H_0 kann bei dem vorliegenden Datenmaterial zum Niveau $\alpha = 5\%$ nicht abgelehnt werden.