

Ingo Bürk

Skript vom Sommersemester 2011

# Mathematische Statistik

*Statistical thinking will one day be as necessary for efficient citizenship as the ability to read or write.*  
H. G. Wells

Universität Stuttgart  
2011

Dieses Skript entstand im Rahmen der Vorlesung „Mathematische Statistik“ bei Hr. Prof. Dr. Ingo Steinwart als Vorlesungsmitschrieb.

Es kann nicht garantiert werden, dass dieses Dokument fehlerfrei ist und der Autor übernimmt für möglicherweise entstandene Schäden jeglicher Art keine Haftung. Dieser Mitschrieb ist kein offizielles Dokument der Universität Stuttgart, Mitarbeiter eben dieser tragen daher ebenfalls keine Verantwortung.

Dieses Werk ist unter dem Lizenzvertrag „Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Germany“ lizenziert. Um die Lizenz anzusehen, gehen Sie bitte auf die Webseite <http://creativecommons.org/licenses/by-nc-sa/3.0/de/> oder schicken Sie einen Brief an:

Creative Commons,  
171 Second Street,  
Suite 300,  
San Francisco,  
California 94105, USA.

Mit freundlichen Grüßen  
Ingo Bürk

# Inhaltsverzeichnis

<b>Vorwort</b>	5
<b>1 Einführung</b>	7
1.1 Motivation . . . . .	7
1.2 Einfache Stichprobenfunktionen . . . . .	10
Exakte Berechnungen statt CLT . . . . .	13
1.3 Die empirische Verteilungsfunktion . . . . .	19
<b>2 Punktschätzungen</b>	23
2.1 Grundlagen . . . . .	23
2.2 Optimalität von Schätzern . . . . .	28
2.3 Konstruktionsmethoden für Punktschätzer . . . . .	35
Momentenmethode . . . . .	35
Maximum-Likelihood-Schätzer . . . . .	37
2.4 Bayes-Schätzer . . . . .	44
2.5 Cramér-Rao-Ungleichung . . . . .	49
2.6 Weitere Eigenschaften von Schätzern . . . . .	53
<b>3 Konfidenzintervalle</b>	57
3.1 Grundlegende Definitionen . . . . .	57
3.2 Einige exakte Berechnungen für Konfidenzintervalle . . . . .	62
3.3 Asymptotische Konfidenzintervalle . . . . .	66
<b>4 Statistische Tests</b>	71
4.1 Qualitätsmerkmale für Tests . . . . .	72
4.2 Existenz gleichmäßig bester Tests . . . . .	77
4.3 Zweistichprobentests . . . . .	94
4.4 Eine Anwendung von Tests . . . . .	96
4.4.1 Informelle Beschreibung des Problems . . . . .	96
4.4.2 Formale Beschreibung des Problems . . . . .	96
4.4.3 Informelle Wahl der Hypothesen . . . . .	97
4.4.4 Test auf einem Datensatz . . . . .	97
4.4.5 Test auf vielen Datensätzen I . . . . .	99
4.4.6 Test auf vielen Datensätzen II . . . . .	100
<b>Abbildungsverzeichnis</b>	103
<b>Stichwortverzeichnis</b>	105

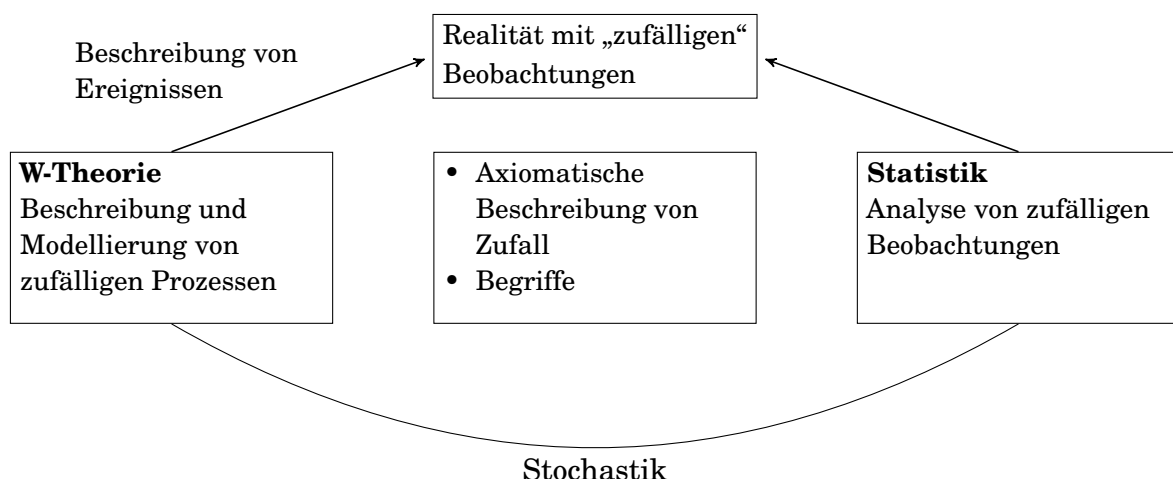


# Vorwort

Die Statistik ist ein Teilgebiet der Stochastik und bildet gewissermaßen den Gegensatz zur Wahrscheinlichkeitstheorie. Die Statistik lässt sich in drei Gebiete gliedern:

- In der *deskriptiven Statistik* werden vorliegende Daten möglichst gut beschrieben und zusammengefasst
- In der *mathematischen Statistik* – auch *induktive Statistik* genannt – versucht man, aus dem Datensatz einer Stichprobe Informationen zu den Eigenschaften der Grundgesamtheit abzuleiten
- In der *explorativen Statistik* werden die beiden anderen Gebiete gewissermaßen vereint; sie kombiniert die Methoden, um Zusammenhänge oder auch Unterschiede zwischen Datensätzen zu ermitteln und diese zu bewerten

Wir werden uns hier mit dem zweiten Gebiet – also der mathematischen Statistik – beschäftigen. Wir werden also annehmen, dass wir einen Satz von Daten zur Verfügung haben, die einem zufälligen, aber unbekanntem Mechanismus unterliegen und werden diese Daten nutzen, um diesen zufälligen Mechanismus zu untersuchen und Eigenschaften in Erfahrung zu bringen.



**Abbildung 0.1:** Übersicht der Zusammenhänge zwischen Stochastik, Wahrscheinlichkeitstheorie und Statistik.



# 1

## Einführung

┌

In diesem ersten Kapitel wollen wir uns mit den Grundlagen der mathematischen Statistik befassen und die wichtigsten Begriffe kennenlernen.

└

### 1.1 Motivation

Wir werden an dieser Stelle einige Grundannahmen und Bezeichnungen einführen, die wir das gesamte Kapitel über hinweg beibehalten werden. Sofern nichts anders vermerkt, beziehen wir uns also immer auf diese Konventionen.

Formal gehen wir von einem Datensatz  $x = (x_1, \dots, x_n) \in \mathcal{X}$  aus, wir können  $x$  also als Beobachtung einer Zufallsvariablen  $X : (\Omega, \mathcal{A}', P') \rightarrow (\mathcal{X}, \mathcal{S}\mathcal{A})$  betrachten, d. h. es existiert ein  $\omega \in \Omega$  mit  $x = X(\omega)$ . Ferner gehen wir davon aus, dass die Verteilung  $P := P'_X$  von  $X$  unbekannt ist.

Unser Ziel ist es nun, Eigenschaften dieser Verteilung  $P$  in Erfahrung zu bringen. Im Allgemeinen haben wir dafür Vorwissen über  $P$  vermöge einer Menge  $\mathcal{P}$  von Wahrscheinlichkeitsmaßen auf  $\mathcal{X}$ , d. h. wir wissen, dass  $P \in \mathcal{P}$  gilt. „Vorwissen“ kann hier aber auch bedeuten, dass wir einfach nur hoffen/annehmen, dass dies tatsächlich zutrifft.

#### Definition 1.1.1 Statistischer Raum

Sei  $(\mathcal{X}, \mathcal{A})$  ein Messraum und  $\mathcal{P}$  eine Menge von Wahrscheinlichkeitsmaßen auf  $\mathcal{X}$ . Dann heißt  $(\mathcal{X}, \mathcal{A}, \mathcal{P})$  statistischer Raum.

#### Bemerkungen:

- Häufig sind die Beobachtungen unabhängig und identisch verteilt generiert, d. h.  $\mathcal{X} = \tilde{\mathcal{X}}^n$ ,  $\mathcal{A} = \bigotimes_1^n \tilde{\mathcal{A}} = \tilde{\mathcal{A}}^n$  und

$$\mathcal{P} \subset \left\{ \tilde{P}^n := \bigotimes_1^n \tilde{P} \text{ mit } \tilde{P} \text{ ist Wahrscheinlichkeitsmaß auf } (\tilde{\mathcal{X}}, \tilde{\mathcal{A}}) \right\}.$$

Zudem werden die „Rohdaten“  $(x_1, \dots, x_n)$  häufig zunächst vorverarbeitet. Wir werden hierauf später mit dem Begriff der Stichprobenfunktion zurückkommen.

- Häufig liegt  $\mathcal{P}$  in *parametrischer* Form vor – d. h.  $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$  –, wobei typischerweise  $\Theta \subset \mathbf{R}^k$  gilt. In diesem Fall heißt  $\mathcal{P}$  **k-dimensional-parametrisch** oder auch nur *k-parametrisch*. Ein Beispiel hierfür wäre

$$\mathcal{P} = \{\mathcal{N}^n(\mu, \sigma^2) : \mu \in \mathbf{R}, \sigma^2 > 0\}.$$

In diesem Fall ist unser Vorwissen vergleichsweise groß und die entsprechenden Fragestellungen entstammen dem Bereich der parametrischen Statistik.

- Liegt keine *einfache* Parametrisierung vor, so spricht man in der Fachliteratur auch von der nicht-parametrischen Statistik. Ein Beispiel hierfür ist

$$\mathcal{P} = \{\tilde{P}^n : \tilde{P} \text{ ist ein Wahrscheinlichkeitsmaß auf } (\tilde{\mathcal{X}}, \tilde{\mathcal{A}})\}.$$

Hier ist unsere einzige Information über die Elemente von  $\mathcal{P}$  also, dass sie i. i. d. sind. Ist  $\mathcal{X}$  groß genug, so liegt nahe, dass keine einfache Parametrisierung existieren wird. Ein zweites Beispiel wäre

$$\mathcal{P} = \{\tilde{P}^n : \tilde{P} \text{ ist ein Wahrscheinlichkeitsmaß auf } (\tilde{\mathcal{X}}, \tilde{\mathcal{A}}) \text{ und } \tilde{P} \ll \mu\},$$

wobei  $\mu$  ein vorgegebenes Referenzmaß ist.

- In der *robusten* Statistik stellt man sich die Frage, was passiert, wenn bei den Methoden der parametrischen Statistik bereits die grundlegende Annahme  $P \in \mathcal{P}$  falsch ist und wie man Methoden findet, die sich robust gegenüber solcher Verletzungen der Modellannahmen verhalten.

In diesem Buch werden wir uns allerdings vor allem mit der klassischen parametrischen Statistik beschäftigen.

### Beispiel 1.1.2

Zur Behandlung einer Krankheit wurde eine neue Heilmethode (I) entwickelt. Bis jetzt wurde Heilmethode (II) benutzt, von welcher wir wissen, dass sie in 75% aller Fälle anschlägt. In einer Kurzzeitstudie wurde Heilmethode (I) bei 100 Patienten getestet, wobei bei 80 Personen ein Behandlungserfolg eintrat. Für die statistische Modellierung können wir verschiedene Ansätze verfolgen:

- *Arbeiten auf den Rohdaten:* Wähle  $\mathcal{X} = \{0, 1\}^{100}$ , wobei 1 einem Behandlungserfolg entspricht, während 0 für das Fehlschlagen der Heilmethode steht. Wir nehmen an, dass die Behandlungsausgänge unabhängig (eine recht realistische Annahme) und identisch verteilt sind (diese Annahme muss im Einzelfall gerechtfertigt werden). Die Annahmen ergeben  $\mathcal{P} = \{B^{100}(p, 1) : p \in [0, 1]\}$ , wobei  $B(p, 1)$  die Binomialverteilung ist, d. h. es gilt  $B(p, 1)(\{1\}) = p$ . Das Problem bei dieser Modellierung ist eine mögliche Informationsflut, die einen sehr hohen Aufwand erzeugt.



- *Vorverarbeitung der Rohdaten:* Bezeichnen  $x_1, \dots, x_{100}$  die einzelnen Ausgänge, so betrachten wir nicht  $(x_1, \dots, x_{100})$ , sondern  $x' := \sum_{i=1}^{100} x_i \in \{0, \dots, 100\} =: \mathcal{X}'$ . Mit  $\mathcal{X}$  und  $\mathcal{P}$  wie oben betrachten wir nun die Abbildung

$$T: \mathcal{X} \rightarrow \mathcal{X}' \quad \text{mit}$$

$$(x_1, \dots, x_{100}) \mapsto \sum_{i=1}^{100} x_i =: x'.$$

Für  $P \in \mathcal{P}$  gilt  $T_P = B(p, 100)$ , daher setzen wir  $\mathcal{P}' := \{B(p, 100) : p \in [0, 1]\}$ , d. h. wir haben unsere Rohdaten in  $x' \in \mathcal{X}'$  zusammengefasst. Die zusammensetzende Funktion  $T$  ist eine Stichprobenfunktion. Zu beachten ist, dass wir – obwohl wir i. i. d. generierte Rohdaten voraussetzen – keine Produktmaße auf  $\mathcal{X}'$  mehr vorfinden. //

Einige typische Fragestellungen in der parametrischen Statistik sind die folgenden Beispiele:

- Schätzproblem:* Versuche den wahren, aber unbekanntem Parameter  $p$  im vorherigen Beispiel 1.1.2 zu schätzen.
- Bereichsschätzung:* Versuche einen Bereich – d. h. typischerweise ein Intervall  $I$  – zu schätzen, so dass  $p \in I$  liegt.
- Testproblem:* Versuche zu entscheiden, ob das wahre  $p$  zum Beispiel größer als 0.75 ist.

## 1.2 Einfache Stichprobenfunktionen

Im Folgenden nehmen wir an, dass  $x = (x_1, \dots, x_n) \in \mathbf{R}^n = \mathcal{X}$  ist. Ferner sei  $\mathcal{P}$  eine Teilmenge aller  $P^n$ , wobei  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathbf{R}$  ist und  $X_i: \mathbf{R}^n \rightarrow \mathbf{R}$  sei die  $i$ -te Projektion, d. h. es gilt  $X = (X_1, \dots, X_n) = \text{id}_{\mathbf{R}^n}$  und  $x$  ist die Realisierung der Zufallsvariablen  $X$  und besteht aus i. i. d. generierten Beobachtungen, da die  $X_1, \dots, X_n$  i. i. d. bezüglich aller  $P^n \in \mathcal{P}$  sind.

### Definition 1.2.1 Arithmetisches Mittel, Stichprobenmittel

Betrachte die Funktion  $T: \mathbf{R}^n \rightarrow \mathbf{R}$ , die durch

$$T(x) := \frac{1}{n} \sum_{i=1}^n x_i \quad \text{mit } x \in \mathbf{R}^n$$

definiert ist. Dann heißt:

- i)  $\bar{x}_n := T(x)$  das **arithmetische Mittel** der Stichprobe  $(x_1, \dots, x_n) =: x$ .
- ii)  $\bar{X}_n := T \circ X$  das **Stichprobenmittel** der **Zufallsstichprobe**  $X := (X_1, \dots, X_n)$ .

### Lemma 1.2.2

Für alle Wahrscheinlichkeitsmaße  $P$  auf  $\mathbf{R}$ , für welche entweder  $X_1 \in \mathcal{L}_1(P^n)$  oder  $X_1 \in \mathcal{L}_2(P^n)$  gilt, gelten die Eigenschaften:

- i)  $\mathbf{E}_{P^n} \bar{X}_n = \mathbf{E}_{P^n} X_1$ .
- ii)  $\text{Var}_{P^n} \bar{X}_n = \frac{1}{n} \text{Var}_{P^n} X_1$ .

**Beweis:** Beide Aussagen lassen sich elementar beweisen. Es gilt

$$\mathbf{E}_{P^n} \bar{X}_n = \mathbf{E}_{P^n} \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \mathbf{E}_{P^n} X_i = \mathbf{E}_{P^n} X_1$$

und

$$\text{Var}_{P^n} \bar{X}_n = \text{Var}_{P^n} \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n^2} \sum_{i=1}^n \text{Var}_{P^n} X_i = \frac{1}{n} \text{Var}_{P^n} X_1. \quad \square$$

### Bemerkungen:

- Der Wert  $\mu := \mathbf{E}_{P^n} X_1$  ist häufig unbekannt. In diesem Fall besagt i), dass bei der Schätzung von  $\mu$  durch  $\bar{X}_n$  kein systematischer Fehler auftritt. Diese Eigenschaft werden wir später **erwartungstreu** nennen.

- Mit Hilfe der Markovschen Ungleichung kann  $\text{Var}_{P^n} \overline{X}_n = \frac{1}{n} \text{Var}_{P^n} X_1 =: \frac{\sigma^2}{n}$  als Maß für die Streuung der Schätzung von  $\mu$  durch  $\overline{X}_n$  angesehen werden, da für alle  $\varepsilon > 0$  gilt:

$$P^n \left( \left\{ \left| \overline{X}_n - \mu \right| \geq \varepsilon \right\} \right) \leq \frac{\sigma^2}{\varepsilon^2 n}$$

- Dert Wert von  $\sigma$  ist im Allgemeinen aber unbekannt. Auf diese Problematik werden wir später noch eingehen.

**Lemma 1.2.3**

Es gelten die beiden folgenden Aussagen:

- i) Gilt  $X_1 \in \mathcal{L}_1(P^\infty)$ , so folgt

$$P^\infty \left( \lim_{n \rightarrow \infty} \overline{X}_n = \mu \right) = 1.$$

- ii) Gilt überdies sogar  $X_1 \in \mathcal{L}_2(P^\infty)$ , so folgt

$$\lim_{n \rightarrow \infty} P^\infty \left( \left\{ \sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \leq t \right\} \right) = \Phi(t),$$

wobei  $\sigma^2 := \text{Var}_{P^\infty} X_1$  und  $\Phi$  die Verteilungsfunktion von  $\mathcal{N}(0, 1)$  ist. Letztere ist gegeben durch

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{s^2}{2}} ds.$$

**Beweis:**

- i) Die Folge  $(X_i)_{i \in \mathbf{N}}$  ist i. i. d. bezüglich  $P^\infty$ , betrachte dazu das kanonische Modell für i. i. d. Zufallsvariablen. Dann folgt die Aussage aus dem SLLN von Kolmogorov.
- ii) Es ist  $X_i \in \mathcal{L}_2(P^\infty)$  und die  $X_i$  sind i. i. d. generiert. Ferner gilt

$$\sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} = \sqrt{n} \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sigma} = \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^n (X_i - \mu).$$

Dann folgt mit dem CLT und der Tatsache, dass Konvergenz in Verteilung äquivalent zur Konvergenz der Verteilungsfunktion ist der Rest der Aussage.  $\square$

**Bemerkungen:**

- Aussage i) besagt, dass der zu schätzende Wert  $\mu$  Grenzwert eines Schätzers  $\overline{X}_n$  ist. Diese Eigenschaft nennt man **Konsistenz**.

- Ist  $\sigma^2 := \text{Var}_{P^\infty} X_1$  bekannt, so kann die Streuung der Schätzung approximiert werden. Es ist

$$\begin{aligned} \left\{ \left| \bar{X}_n - \mu \right| > \varepsilon \right\} &= \left\{ \sqrt{n} \left| \frac{\bar{X}_n - \mu}{\sigma} \right| > \frac{\varepsilon \sqrt{n}}{\sigma} \right\} \\ &= \left\{ \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < -\frac{\varepsilon \sqrt{n}}{\sigma} \right\} \sqcup \left\{ \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} > \frac{\varepsilon \sqrt{n}}{\sigma} \right\}, \end{aligned}$$

woraus mit  $\Phi(-t) = 1 - \Phi(t)$  schließlich folgt:

$$P^\infty \left( \left\{ \left| \bar{X}_n - \mu \right| > \varepsilon \right\} \right) \stackrel{\text{ii)}}{\approx} \Phi \left( -\frac{\varepsilon \sqrt{n}}{\sigma} \right) + \left( 1 - \Phi \left( \frac{\varepsilon \sqrt{n}}{\sigma} \right) \right) = 2 \left( 1 - \Phi \left( \frac{\varepsilon \sqrt{n}}{\sigma} \right) \right)$$

Hier gilt es zu beachten, dass die Approximation eigentlich eine gleichmäßige Konvergenz in ii) benötigt. Dies ist erfüllt, falls  $X_1 \in \mathcal{L}_3(P^\infty)$  gilt (Satz von Berry-Essen).

#### Definition 1.2.4 Stichprobenvarianz

Betrachte die Funktion  $T: \mathbf{R}^n \rightarrow \mathbf{R}$ , die durch

$$T(x) := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$$

definiert ist. Dann heißt:

- i)  $s_n^2 := T(x)$  die **Stichprobenvarianz** der Stichprobe  $x = (x_1, \dots, x_n)$ .
- ii)  $S_n^2 := T \circ X$  die **Stichprobenvarianz** der Zufallsstichprobe  $X = (X_1, \dots, X_n)$ .

#### Lemma 1.2.5

Es sei  $X_1 \in \mathcal{L}_2(P^\infty)$  und  $\sigma^2 := \text{Var}_{P^\infty} X_1$ . Dann folgt:

- i) **Erwartungstreue:**  $\mathbf{E} S_n^2 = \sigma^2$ .
- ii) **Konsistenz:**  $P^\infty \left( \lim_{n \rightarrow \infty} S_n^2 = \sigma^2 \right) = 1$ .

**Beweis:** Für  $X'_i := X_i - \mu$  mit  $\mu := \mathbf{E} X_1$  gilt  $\mathbf{E} X'_i = 0$  und  $\mathbf{E} \bar{X}'_n = 0$ . Wir zeigen als erstes  $S_n'^2 = S_n^2$ . Es gilt

$$\begin{aligned} S_n'^2 &= \frac{1}{n-1} \sum_{i=1}^n (X'_i - \bar{X}'_n)^2 = \frac{1}{n-1} \sum_{i=1}^n \left( (X_i - \mu) - \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \right)^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \\ &= S_n^2. \end{aligned}$$

Wir können also ohne Einschränkung annehmen, dass  $\mathbf{E}X_i = 0$  ist, d. h.  $\mu = 0$  und  $\overline{\mathbf{E}X_n} = 0$ . Nun folgt

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - 2X_i\overline{X_n} + \overline{X_n}^2) = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{2n}{n-1} \overline{X_n} \cdot \frac{1}{n} \sum_{i=1}^n X_i + \frac{n}{n-1} \overline{X_n}^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - \overline{X_n}^2). \end{aligned} \quad (*)$$

i) Aus (\*) folgt nun mit unseren obigen Annahmen

$$\begin{aligned} \mathbf{E}S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{E}X_i^2 - \mathbf{E}\overline{X_n}^2) = \frac{1}{n-1} \sum_{i=1}^n (\text{Var} X_i - \text{Var} \overline{X_n}) \\ &\stackrel{1.2.2}{=} \frac{1}{n-1} \sum_{i=1}^n \left( \text{Var} X_i - \frac{1}{n} \text{Var} X_i \right) = \frac{n}{n-1} \left( 1 - \frac{1}{n} \right) \text{Var} X_1 = \text{Var} X_1 \\ &= \sigma^2. \end{aligned}$$

ii) Ferner gilt

$$\begin{aligned} \lim_{n \rightarrow \infty} S_n^2 &\stackrel{(*)}{=} \lim_{n \rightarrow \infty} \frac{1}{n-1} \sum_{i=1}^n (X_i^2 - \overline{X_n}^2) = \lim_{n \rightarrow \infty} \frac{n}{n-1} \left( \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X_n}^2 \right) \\ &= \mathbf{E}X_1^2 - (\mathbf{E}X_1)^2 \\ &= \text{Var} X_1. \end{aligned} \quad \square$$

### Korollar 1.2.6

Es sei  $X_1 \in \mathcal{L}_2(P^\infty)$ ,  $\mu := \mathbf{E}X_1$  und  $\sigma^2 := \text{Var} X_1$ . Dann folgt für alle  $t \in \mathbf{R}$

$$\lim_{n \rightarrow \infty} P^\infty \left( \sqrt{n} \frac{\overline{X_n} - \mu}{S_n} \leq t \right) = \Phi(t),$$

wobei  $\Phi$  die Verteilungsfunktion von  $\mathcal{N}(0, 1)$  ist.

**Beweis:** Es gilt nach Lemma 1.2.5, dass  $S_n \rightarrow \sigma$   $P^\infty$ -fast sicher konvergiert und damit insbesondere auch in Verteilung. Nun folgt die Aussage aus dem CLT und der folgenden multiplikativen Version des Satzes von Slutsky<sup>1</sup>: Es gelte  $Y_n \rightarrow Y$  und  $Z_n \rightarrow c \in \mathbf{R}$  jeweils in Verteilung, dann folgt  $Y_n Z_n \rightarrow cY$  in Verteilung.  $\square$

## Exakte Berechnungen statt CLT

Wir nehmen nun an, dass die Zufallsvariablen  $X_i$  i. i. d. sind und dass  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  gilt, wobei  $\mu \in \mathbf{R}$  und  $\sigma^2 > 0$  ist. Unser Ziel ist es, die Wahrscheinlichkeiten aus Lemma 1.2.3 und Korollar 1.2.6 exakt auszurechnen.

<sup>1</sup> Zu finden als Satz II.9.9 im Wahrscheinlichkeitstheorie-Skript

**1. Fall:** Der Wert von  $\sigma^2 > 0$  sei bekannt. Dann gilt  $X_i - \mu \sim \mathcal{N}(0, \sigma^2)$  und damit  $\frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ . Wir betrachten nun die Summe, für die dann schließlich  $\sum_{i=1}^n \frac{X_i - \mu}{\sigma} \sim \mathcal{N}(0, n)$  gilt<sup>2</sup>. Dann ist

$$\sqrt{n} \frac{\overline{X_n} - \mu}{\sigma} = \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^n (X_i - \mu) \sim \mathcal{N}(0, 1)$$

und für die Wahrscheinlichkeit gilt damit

$$P(\{|\overline{X_n} - \mu| > \varepsilon\}) = 2 \left( 1 - \Phi \left( \frac{\varepsilon \sqrt{n}}{\sigma} \right) \right),$$

wobei wir diese Gleichung bereits in der Rechnung im Anschluss an Lemma 1.2.3 hergeleitet hatten.

**Lemma 1.2.7**

Sei  $X_i$  i. i. d. und  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  mit  $\mu \in \mathbf{R}$  und  $\sigma^2 > 0$ . Dann gilt für alle  $n \geq 1$  und  $\varepsilon > 0$

$$P(\{|\overline{X_n} - \mu| > \varepsilon\}) = 2 \left( 1 - \Phi \left( \frac{\varepsilon \sqrt{n}}{\sigma} \right) \right),$$

wobei  $\Phi$  die Verteilungsfunktion von  $\mathcal{N}(0, 1)$  ist.

**2. Fall:** Der Wert von  $\sigma^2 > 0$  sei nun unbekannt. Für diesen Fall müssen wir zunächst einige neue Begriffe einführen.

**Definition 1.2.8 Gammafunktion**

Die Funktion  $\Gamma: (0, \infty) \rightarrow (0, \infty)$ , die durch

$$\Gamma(p) = \int_0^{\infty} y^{p-1} e^{-y} dy$$

gegeben ist, heißt **Gammafunktion**.

Für die Gammafunktion gelten folgende Eigenschaften, die an dieser Stelle nicht bewiesen werden:

- i)  $\Gamma(p + 1) = p \cdot \Gamma(p)$
- ii)  $\Gamma(1) = 1$
- iii)  $\Gamma(n + 1) = n!$  für  $n \in \mathbf{N}$
- iv) Für alle  $p > 0$  und  $b > 0$  gilt die Identität

$$\Gamma(p) = b^p \int_0^{\infty} e^{-by} y^{p-1} dy.$$

<sup>2</sup> Vgl. Faltungen im Wahrscheinlichkeitstheorie-Skript.

Aus diesen Eigenschaften folgt leicht, dass die durch

$$y \mapsto \mathbf{1}_{(0,\infty)}(y) \cdot \frac{b^p}{\Gamma(p)} e^{-by} y^{p-1}$$

definierte Funktion Dichte eines Wahrscheinlichkeitsmaßes auf  $\mathbf{R}$  ist.

**Definition 1.2.9 Gammaverteilung**

Das Lebesgue-absolut stetige Wahrscheinlichkeitsmaß mit der Dichte

$$y \mapsto \mathbf{1}_{(0,\infty)}(y) \cdot \frac{b^p}{\Gamma(p)} e^{-by} y^{p-1}$$

heißt **Gammaverteilung** mit den Parametern  $b > 0$  und  $p > 0$ . Wir schreiben hierfür  $\Gamma(b, p)$ .

**Lemma 1.2.10**

Sei  $b > 0$ ,  $p > 0$  und  $X \sim \Gamma(b, p)$ . Dann gelten die folgenden Aussagen:

- i) Die charakteristische Funktion  $\varphi_X(t) := \mathbf{E}e^{itX}$  mit  $t \in \mathbf{R}$  gammaverteilter Zufallsvariablen hat die Gestalt

$$\varphi_X(t) = \frac{1}{\left(1 - \frac{it}{b}\right)^p} \quad (t \in \mathbf{R}).$$

- ii) Für  $k \in \mathbf{N}$  gilt für die  $k$ -ten Momente

$$\mathbf{E}X^k = \frac{p \cdot (p+1) \cdot \dots \cdot (p+k-1)}{b^k}.$$

Daraus folgt, dass für alle  $k \in \mathbf{N}$  auch  $X \in \mathcal{L}_k$  gilt

- iii) Es sei  $X_1 \sim \Gamma(b, p_1)$  und  $X_2 \sim \Gamma(b, p_2)$  für unabhängige Zufallsvariablen  $X_1$  und  $X_2$ . Dann gilt  $X_1 + X_2 \sim \Gamma(b, p_1 + p_2)$ .

**Beweis:**

- i) Der Beweis dieser Eigenschaft erfolgt im Wesentlichen über den Identitätssatz für holomorphe Funktionen. Da dieser Gegenstand der Höheren Analysis ist werden wir den Beweis hier nicht führen.
- ii) Die  $k$ -te Ableitung von  $\varphi_X$  ist

$$\varphi_X^{(k)}(t) = \frac{p \cdot (p+1) \cdot \dots \cdot (p+k-1)}{b^k} \cdot i^k \cdot \frac{1}{\left(1 - \frac{it}{b}\right)^{p+k}}.$$

Nach Satz II.4.7 des Wahrscheinlichkeitstheorie-Skriptes gilt  $\mathbf{E}X^k = i^{-k} \varphi_X^{(k)}(0)$ . Dies entspricht genau der Identität, die wir zeigen wollen.

- iii) Nach Satz II.4.8 des Wahrscheinlichkeitstheorie-Skriptes gilt  $\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)$ . Durch Einsetzen und Anwendung des Eindeutigkeitsatzes für charakteristische Funktionen folgt dann die Aussage.  $\square$

**Korollar 1.2.11**

Für die Exponentialverteilung  $\exp(\lambda)$  mit  $\lambda > 0$ , welche für  $y \in \mathbf{R}$  die Dichte  $f_\lambda(y) = \mathbf{1}_{(0,\infty)}(y) \cdot \lambda e^{-\lambda y}$  besitzt, gelten die beiden folgenden Aussagen:

- i)  $\exp(\lambda) = \Gamma(\lambda, 1)$ .
- ii) Für unabhängige  $Y_i \sim \exp(\lambda)$  gilt  $\sum_{i=1}^n Y_i \sim \Gamma(\lambda, n)$ .

**Beweis:** Der Beweis erfolgt durch simples Nachrechnen.  $\square$

**Satz 1.2.12**

Seien  $X_1, \dots, X_n$  unabhängige Zufallsvariablen mit  $X_i \sim \mathcal{N}(0, 1)$ . Dann gilt

$$\sum_{i=1}^n X_i^2 \sim \Gamma\left(\frac{1}{2}, \frac{n}{2}\right).$$

Die Verteilung  $\Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$  heißt **Chi-Quadrat-Verteilung** mit  $n$  Freiheitsgraden. Wir schreiben hierfür  $\chi_n^2 := \Gamma\left(\frac{1}{2}, \frac{n}{2}\right)$ . Die  $\chi_n^2$ -Verteilung hat die Lebesgue-Dichte

$$y \mapsto \mathbf{1}_{(0,\infty)}(y) \cdot \frac{y^{\frac{n}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}.$$

**Beweis:** Es genügt  $X_1^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$  zu zeigen, da die  $X_1^2, \dots, X_n^2$  unabhängig sind. Sei  $\varphi$  die Dichte von  $\mathcal{N}(0, 1)$ , also  $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$  mit  $t \in \mathbf{R}$ . Dann gilt für  $X := X_1$ :

$$\begin{aligned} P(X^2 \leq a) &= P(-\sqrt{a} \leq X \leq \sqrt{a}) = \int_0^{\sqrt{a}} \varphi(t) dt + \int_{-\sqrt{a}}^0 \varphi(t) dt \\ &\stackrel{t=\pm s}{=} \int_0^a \frac{1}{2\sqrt{s}} \varphi(\sqrt{s}) ds + \int_0^a \frac{1}{2\sqrt{s}} \varphi(-\sqrt{s}) ds \\ &= \frac{1}{\sqrt{2\pi}} \left( \int_0^a \frac{1}{2\sqrt{s}} e^{-\frac{s}{2}} ds + \int_0^a \frac{1}{2\sqrt{s}} e^{-\frac{s}{2}} ds \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_0^a s^{-\frac{1}{2}} e^{-\frac{s}{2}} ds \end{aligned}$$



Die Verteilung von  $X^2$  hat damit die Dichte  $\begin{cases} \frac{1}{\sqrt{2\pi}} s^{-\frac{1}{2}} e^{-\frac{s}{2}} & s > 0 \\ 0 & s \leq 0 \end{cases}$ , da die Verteilungsfunktionen

Maße eindeutig bestimmen. Dies entspricht  $y \mapsto \mathbf{1}_{(0,\infty)}(y) \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}$  für  $y \in \mathbf{R}$ . Da  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$  gilt<sup>3</sup> folgt die Behauptung, denn die Dichte von  $\Gamma(\frac{1}{2}, \frac{1}{2})$  ist

$$y \mapsto \mathbf{1}_{(0,\infty)}(y) \frac{\left(\frac{1}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} e^{-\frac{1}{2}y} y^{-\frac{1}{2}}. \quad \square$$

**Satz 1.2.13**

Es seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$  unabhängige Zufallsvariablen. Dann gelten die folgenden Aussagen:

- i)  $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ .
- ii)  $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2 = \Gamma\left(\frac{1}{2}, \frac{n-1}{2}\right)$ .
- iii)  $\bar{X}_n$  und  $S_n^2$  sind unabhängig.

**Beweis:** Der Beweis des Satzes ist relativ kompliziert und wird daher an dieser Stelle nicht geführt. □

**Definition 1.2.14 Student-t-Verteilung**

Sei  $n \in \mathbf{N}$ . Dann heißt die durch die Lebesguedichte

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\sqrt{n\pi} \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} \quad \text{mit } x \in \mathbf{R}$$

gegebene Verteilung auf  $\mathbf{R}$  die **Student-t-Verteilung** mit  $n$  Freiheitsgraden. Wir schreiben dafür  $t_n$ .

**Satz 1.2.15**

Es seien  $X \sim \mathcal{N}(0, 1)$  und  $Y \sim \chi_n^2$  unabhängige Zufallsvariablen. Dann gilt:

$$\frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$$

<sup>3</sup> Dies ist ein Ergebnis aus der Analysis.

**Kurz:** Formal nicht ganz korrekt lässt sich dies als „ $\frac{\mathcal{N}(0,1)}{\sqrt{\frac{\lambda_n^2}{n}}} = t_n$ “ merken.

**Beweis:** Da der Beweis umfangreich und aufwändig ist wird er hier nicht geführt. □

### Korollar 1.2.16

Es seien  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$  unabhängige Zufallsvariablen. Dann gilt:

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n} \sim t_{n-1}$$

**Beweis:** Wir verwenden Satz 1.2.13, der uns sagt, dass  $\overline{X}_n$  und  $S_n^2$  unabhängig sind. Nun ist  $\sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$ , wie wir im Wesentlichen schon im Beweis von Lemma 1.2.7 gezeigt haben. Satz 1.2.13 sagte uns ferner, dass  $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$  gilt. Mit Satz 1.2.15 folgt nun

$$t_{n-1} \sim \frac{\sqrt{n} \left( \frac{\overline{X}_n - \mu}{\sigma} \right)}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2} \cdot \frac{1}{n-1}}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{S_n}.$$

Die nötige Unabhängigkeit folgt aus den obigen Überlegungen. □

### Bemerkungen:

- Die Aussage des Korollars 1.2.16 kann benutzt werden, um zu berechnen, wie weit  $\overline{X}_n$  von  $\mu$  mit vorgegebener Wahrscheinlichkeit abweicht, falls  $\sigma^2$  unbekannt ist.
- Die in diesem Abschnitt auftretenden Verteilungsfunktionen haben keine geschlossene Formel. Man greift daher auf Tabellen, R, MatLab und andere Software zurück.

## 1.3 Die empirische Verteilungsfunktion

Nachdem wir uns mit einigen Grundbegriffen beschäftigt haben wollen wir uns nun eine komplexere Stichprobenfunktion anschauen.

### Definition 1.3.1 Empirische Verteilungsfunktion

Wir werden den Begriff der empirischen Verteilungsfunktion für Stichproben und für Zufallsvariablen einführen:

- Es seien  $x_1, \dots, x_n \in \mathbf{R}$ . Dann heißt die Funktion

$$F_n : \mathbf{R} \rightarrow [0, 1]$$

$$t \mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, t]}(x_i) \quad \text{für } t \in \mathbf{R}$$

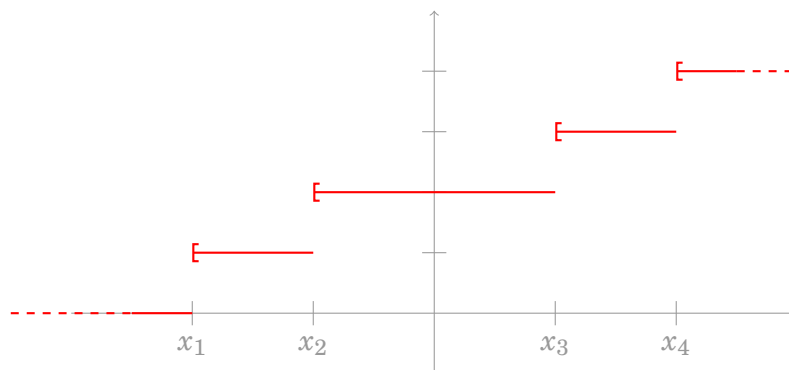
die **empirische Verteilungsfunktion** zu der Stichprobe  $x_1, \dots, x_n$ .

- Es seien  $X_1, \dots, X_n : (\Omega, \mathcal{A}, P) \rightarrow \mathbf{R}$  unabhängig und identisch verteilte Zufallsvariablen. Dann heißt

$$F_n(t, \omega) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, t]}(X_i(\omega)) \quad \text{für } t \in \mathbf{R} \text{ und } \omega \in \Omega$$

die **empirische Verteilungsfunktion** zu der Zufallsstichprobe  $X_1, \dots, X_n$ . Für ein festes  $\omega \in \Omega$  ist dies eine Funktion  $F(\cdot, \omega) : \mathbf{R} \rightarrow [0, 1]$ .

Die Funktionen  $t \mapsto F_n(t)$  und  $t \mapsto F_n(t, \omega)$  für festes  $\omega \in \Omega$  sind Verteilungsfunktionen. Das zu  $t \mapsto F_n(t)$  zugehörige Maß heißt **empirisches Maß** und ist durch  $\frac{1}{n} \sum_{i=1}^n \delta_{\{x_i\}}$  gegeben, wobei  $\delta_{\{x_i\}}$  das Dirac-Maß im Punkt  $x_i$  ist.



**Abbildung 1.1:** Skizzenhafte Darstellung einer empirischen Verteilungsfunktion

**Satz 1.3.2 Glivenko-Cantelli**

Es seien  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $X_i: \Omega \rightarrow \mathbf{R}$  i. i. d. Zufallsvariablen und  $F$  die Verteilungsfunktion von  $X_1$ . Dann gilt

$$P \left( \left\{ \omega \in \Omega : \limsup_{n \rightarrow \infty} \sup_{t \in \mathbf{R}} |F_n(t, \omega) - F(t)| = 0 \right\} \right) = 1.$$

Mit anderen Worten konvergiert  $F_n \rightarrow F$  also  $P$ -fast sicher gleichmäßig.

**Beweis:** Für  $t \in \mathbf{R}$  definieren wir die Zufallsvariablen  $Y_n(t) := \mathbf{1}_{(-\infty, t]}(X_n)$  und  $Z_n(t) := \mathbf{1}_{(-\infty, t)}(X_n)$ . Dann gilt, dass sowohl  $Y_n$  als auch  $Z_n$  jeweils i. i. d. sind. Ferner gilt

$$\begin{aligned} \mathbf{E}Y_n(t) &= \mathbf{E}_{P_{X_n}} \mathbf{1}_{(-\infty, t]} = P_{X_n}((-\infty, t]) = F(t) \quad \text{und} \\ \mathbf{E}Z_n(t) &= \mathbf{E}_{P_{X_n}} \mathbf{1}_{(-\infty, t)} = P_{X_1}((-\infty, t)) = F(t^-) := \lim_{t' \nearrow t} F(t'). \end{aligned}$$

Mit dem SLLN für festes  $t \in \mathbf{R}$  folgt dann

$$\begin{aligned} F_n(t, \omega) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) \longrightarrow \mathbf{E}Y_1(t) = F(t) \quad \text{und} \\ F_n(t^-, \omega) &= \frac{1}{n} \sum_{i=1}^n Z_i(t) \longrightarrow \mathbf{E}Z_1(t) = F(t^-), \end{aligned}$$

wobei die Konvergenz jeweils  $P$ -fast sicher gilt. Unser Ziel ist es nun, die Monotonie und rechtsseitige Stetigkeit zu verwenden, um aus diesen beiden Aussagen gleichmäßige Konvergenz zu folgern. Definiere dazu  $\bar{\mathbf{R}} := \mathbf{R} \cup \{\pm\infty\}$  mit  $F(-\infty) := 0$  und  $F(\infty) := 1$ . Für ein  $m \in \mathbf{N}$  sei ferner  $x_j := \inf \left\{ x \in \bar{\mathbf{R}} : F(x) \geq \frac{j}{m} \right\}$  mit  $j = 0, \dots, m$ . Für eben solche  $j$  gilt dann:

- i) Für  $x \in [x_{j-1}, x_j)$  gilt  $\frac{j-1}{m} \leq F(x) < \frac{j}{m}$ . Dies folgt aus  $x_j > x \geq x_{j-1}$  und der rechtsseitigen Stetigkeit von  $F$ .
- ii) Für  $x \in [x_{j-1}, x_j)$  folgt  $\frac{j-1}{m} \leq F(x^-) < \frac{j}{m}$ , wobei dies aus  $F(x) \leq F(x^-)$  und  $x < x_j$  folgt.

Wir definieren nun ferner

$$R_n(\omega) := \max_{j=1, \dots, m-1} \left\{ |F_n(x_j, \omega) - F(x_j)| + |F_n(x_j^-, \omega) - F(x_j^-)| \right\}.$$

Wegen den obigen Gleichungen gilt dann  $R_n \rightarrow 0$   $P$ -fast sicher. Für  $x \in [x_{j-1}, x_j)$  folgt nun

$$F_n(x, \omega) \leq F(x_j^-, \omega) \leq F(x_j^-) + R_n(\omega) \stackrel{\text{ii)}}{\leq} \frac{j}{m} + R_n(\omega) \stackrel{\text{i)}}{\leq} F(x) + \frac{1}{m} + R_n(\omega)$$

und

$$F_n(x, \omega) \geq F(x_{j-1}, \omega) \geq F(x_{j-1}) - R_n(\omega) \geq \frac{j-1}{m} - R_n(\omega) \geq F(x) - \frac{1}{m} - R_n(\omega).$$

Setzt man dies zusammen, so folgt insgesamt also  $\sup_{x \in \mathbf{R}} |F_n(x, \omega) - F(x)| \leq \frac{1}{m} + R_n(\omega) \rightarrow 0$   $P$ -fast sicher.  $\square$

**Korollar 1.3.3**

Sei  $(\Omega, \mathcal{A}, P)$  ein Wahrscheinlichkeitsraum,  $X_i : \Omega \rightarrow \mathbf{R}$  i. i. d. Zufallsvariablen und  $P_{n,\omega} := \frac{1}{n} \sum_{i=1}^n \delta_{\{X_i(\omega)\}}$  das empirische Maß zu den Beobachtungen  $X_1(\omega), \dots, X_n(\omega)$ . Für  $I := \{(-\infty, a] : a \in \mathbf{R}\}$  gilt dann

$$P \left( \left\{ \omega \in \Omega : \lim_{n \rightarrow \infty} \sup_{A \in I} |P_{n,\omega}(A) - P_{X_1}(A)| = 0 \right\} \right) = 1.$$

**Beweis:** Für ein Intervall  $A \in I$  gilt

$$P_{n,\omega}(A) = P_{n,\omega}((-\infty, a]) = \frac{1}{n} \sum_{i=1}^n \delta_{\{X_i(\omega)\}}((-\infty, a]) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, a]}(X_i(\omega)) = F_n(t, \omega).$$

Da zudem  $P_{X_1}(A) = F(a)$  gilt folgt die Aussage dann aus Satz 1.3.2. □

**Bemerkung:** Es gibt noch andere Familien  $I \subset \mathcal{B}$ , für welche die Aussage des Korollars 1.3.3 gilt. Beispiele hierfür sind  $\{(-\infty, a] : a \in \mathbf{R}\}$ ,  $\{[a, b] : a, b \in \mathbf{R}\}$ , sowie alle anderen Familien von Intervallen,  $\{[a, b] \cup [c, d] : a, b, c, d \in \mathbf{R}\}$  und viele weiterer solcher Konstruktionen. Für  $I = \mathcal{B}$  gilt die Aussage jedoch *nicht*.



Eine genauere Beschreibung, wann dies für allgemeine, nicht notwendigerweise  $\mathbf{R}$ -wertige Zufallsvariablen gilt, kann mit der so genannten VC-Dimension von  $I$  gegeben werden.



# 2

## Punktschätzungen

┌

In diesem Kapitel wollen wir verschiedene Schätzer konstruieren und kennenlernen, um einen unbekanntem Parameter aus der Grundgesamtheit zu schätzen. Mit solchen Punktschätzern werden in der Statistik viele grundlegende Größen geschätzt.

└

### 2.1 Grundlagen

Im Folgenden sei, sofern nicht anders erwähnt, stets das Modell gemeint, das wir nun beschreiben. Es sei wie gewohnt  $\mathcal{X} = \mathbf{R}^n$  und  $\mathcal{P}$  eine Teilmenge aller  $P^n$ , wobei  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathbf{R}$  ist. Ferner sei  $X_i: \mathbf{R}^n \rightarrow \mathbf{R}$  die  $i$ -te Projektion, so dass  $X := (X_1, \dots, X_n)$  aus i. i. d. Zufallsvariablen besteht und  $P_{X_i}^n = P$  gilt. Zusätzlich nehmen wir nun aber an, dass  $\Theta \subset \mathbf{R}^m$  eine messbare Menge ist und wir eine Familie  $(P_\vartheta)_{\vartheta \in \Theta}$  von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$  haben, wobei für  $\vartheta_1 \neq \vartheta_2$  auch  $P_{\vartheta_1} \neq P_{\vartheta_2}$  gelten soll. Wir schränken uns dann auf  $\mathcal{P} = \{P_\vartheta^n : \vartheta \in \Theta\}$  ein.

Unser Ziel ist es nun, den Parameter  $\vartheta^*$  zu schätzen, wobei wir von  $X_i \sim P_{\vartheta^*}$  für ein „wahres“ – aber unbekanntes –  $\vartheta^*$  ausgehen. Dazu suchen wir Verfahren, die gute Schätzungen für jedes  $\vartheta \in \Theta$  liefern. Wir führen an dieser Stelle zudem die Schreibweisen  $\mathbf{E}_\vartheta := \mathbf{E}_{P_\vartheta}$  bzw.  $\mathbf{E}_\vartheta := \mathbf{E}_{P_\vartheta^n}$  und  $\text{Var}_\vartheta := \text{Var}_{P_\vartheta}$  ein, wobei sich aus dem Zusammenhang ergeben wird, was genau gemeint ist.

#### Definition 2.1.1 Schätzer

Eine Borel-messbare Funktion  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}^m$  heißt **Schätzer**.

Die Idee hinter einem Schätzer ist die folgende: Ist  $(x_1, \dots, x_n) \in \mathbf{R}^n$ , so schätzt  $\hat{\Theta}(x_1, \dots, x_n)$  den Parameter. Zu beachten ist hierbei, dass wir nicht  $\hat{\Theta} \in \Theta$  fordern, obwohl dies typischerweise dennoch gilt. In diesem Fall heißt  $\hat{\Theta} \circ X := \hat{\Theta}(X_1, \dots, X_n)$  **Schätzung** unserer Zufallsstichprobe  $X_1, \dots, X_n$ .

Der Begriff des Schätzers ist damit natürlich sehr allgemein. Der nächste Schritt besteht daher darin, Eigenschaften festzulegen, die wir von einem guten Schätzer erwarten. Wir wollen also Eigenschaften festlegen, mit denen wir gute von schlechten Schätzern unterscheiden können.

**Definition 2.1.2 Erwartungstreue**

Es sei  $\hat{\Theta}$  ein integrierbarer Schätzer (d. h. es gilt  $\mathbf{E}_{P_\vartheta^n}|\hat{\Theta}(X_1, \dots, X_n)| < \infty$  für alle  $\vartheta \in \Theta$  bzw. – äquivalent hierzu – es gilt  $\hat{\Theta} \circ X \in \mathcal{L}_1(P_\vartheta^n)$ ). Dann heißt  $\hat{\Theta}$ :

- i) **erwartungstreu** oder **unverzerrt** genau dann, wenn  $\mathbf{E}_\vartheta \hat{\Theta}(X_1, \dots, X_n) = \vartheta$  für alle  $\vartheta \in \Theta$  gilt.
- ii) **asymptotisch erwartungstreu** genau dann, wenn  $\lim_{n \rightarrow \infty} \mathbf{E}_\vartheta \hat{\Theta}(X_1, \dots, X_n) = \vartheta$  für alle  $\vartheta \in \Theta$  gilt.

Es ist klar, dass aus Erwartungstreue auch asymptotische Erwartungstreue folgt.

Einige Beispiele sind:

- Das arithmetische Mittel  $\overline{X}_n$  ist ein erwartungstreuer Schätzer für den Erwartungswert. Sei  $\Theta \subset \mathbf{R}$  und  $\mathbf{E}_\vartheta X_1 = \vartheta$ , dann folgt die Unverzerrtheit aus Lemma 1.2.2.
- Die Stichprobenvarianz  $S_n^2$  ist ein erwartungstreuer Schätzer für die Varianz. Es sei  $\Theta \subset \mathbf{R}$  und  $\text{Var}_\vartheta X_1 = \vartheta$ ; die Erwartungstreue folgt dann aus Lemma 1.2.5.
- Wir definieren  $\hat{\Theta}(x_1, \dots, x_n) := \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x}_n)^2$ , dann ist  $\hat{\Theta} \circ X = \frac{n-1}{n} S_n^2$ . Da  $S_n^2$  unverzerrt ist, aber  $\frac{n-1}{n} \neq 1$  gilt, ist  $\hat{\Theta}$  nicht erwartungstreu. Wegen  $\frac{n-1}{n} \rightarrow 1$  ist der Schätzer jedoch asymptotisch erwartungstreu.
- *Kombination von Schätzern:* Die Kombination  $(\overline{X}_n, S_n^2)$  ist ein erwartungstreuer Schätzer für (Erwartungswert, Varianz). Formal heißt das für  $\Theta \subset \mathbf{R}^2$ , dass  $(\mathbf{E}_\vartheta X_1, \text{Var}_\vartheta X_1) = \vartheta$  für alle  $\vartheta \in \Theta$  gilt.

**Definition 2.1.3 Verzerrung/Bias**

Es sei  $\hat{\Theta}$  ein integrierbarer Schätzer. Dann heißt

$$\text{Bias}_\vartheta \hat{\Theta} := \mathbf{E}_\vartheta (\hat{\Theta} \circ X) - \vartheta$$

die **Verzerrung** oder der **Bias** bezüglich  $\vartheta$ .

**Beispiel**

Wir wollen den Bias für  $\hat{\Theta} \circ X := \frac{n-1}{n} S_n^2$  berechnen. Es gilt  $\text{Bias}_\vartheta \hat{\Theta} = -\frac{1}{n} \text{Var}_\vartheta X_1 = -\frac{1}{n} \sigma^2$  mit  $\sigma^2 := \text{Var}_\vartheta X_1$ , d. h.  $\hat{\Theta}$  unterschätzt den zu schätzenden Parameter  $\sigma^2$  systematisch. Man leitet sich das wie folgt her:

$$\mathbf{E}_\vartheta \frac{n-1}{n} S_n^2 - \sigma^2 = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{1}{n} \sigma^2 \quad //$$



Uns stellt sich nun aber das Problem, dass wir im Allgemeinen nicht erwarten können, dass unser Schätzer exakte Schätzungen liefert, d. h. im Allgemeinen gilt  $\hat{\Theta}(X_1, \dots, X_n) \neq \vartheta$  für  $X_i \sim P_\vartheta$  mit  $\vartheta \in \Theta$ . Die Frage ist nun also, wie wir solche Abweichungen beschreiben können.

**Definition 2.1.4 Verlustfunktion**

Eine messbare Funktion  $L: \Theta \times \mathbf{R}^m \rightarrow [0, \infty)$  heißt **Verlustfunktion**.

Wir können dies so interpretieren, dass  $L(\vartheta, \hat{\Theta})$  die Abweichung der Schätzung  $\hat{\Theta} \in \mathbf{R}^m$  von dem Parameter  $\vartheta$  schätzt. Je kleiner  $L(\vartheta, \hat{\Theta})$  ist, desto besser ist die Schätzung. Oft gilt daher zusätzlich  $L(\vartheta, \hat{\Theta}) = 0$  für  $\vartheta = \hat{\Theta}$ .

Wir wollen uns nun ein Beispiel für eine Verlustfunktion anschauen, die wir immer wieder benötigen werden. Genauer gesagt ist es vorläufig die einzige Verlustfunktion, die uns beschäftigen wird.

**Beispiel Gauß- oder Least-Squares-Verlustfunktion**

Wir betrachten die Funktion  $L: \mathbf{R} \times \mathbf{R} \rightarrow [0, \infty)$  mit  $L(\vartheta, \hat{\Theta}) := (\vartheta - \hat{\Theta})^2$  und nennen sie **Verlustfunktion der kleinsten Quadrate**. Wir können diese Funktion problemlos auf  $m$  Dimensionen verallgemeinern vermöge  $\|\vartheta - \hat{\Theta}\|_2^2 = \sum_{i=1}^m (\vartheta_i - \hat{\Theta}_i)^2$ , wobei  $\vartheta = (\vartheta_1, \dots, \vartheta_m)$  und  $\hat{\Theta} = (\hat{\Theta}_1, \dots, \hat{\Theta}_m)$  ist. //

**Definition 2.1.5 Risiko/Erwarteter Verlust**

Es sei  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}^m$  ein Schätzer und  $L: \Theta \times \mathbf{R}^m \rightarrow [0, \infty)$  eine Verlustfunktion. Dann heißt die Abbildung

$$\mathcal{R}(\vartheta, \hat{\Theta}) := \int L(\vartheta, \hat{\Theta} \circ X) dP_\vartheta^n$$

der **erwartete Verlust** oder das **Risiko** des Schätzers  $\hat{\Theta}$ .

Zu beachten ist, dass  $\mathcal{R}(\vartheta, \hat{\Theta}) = \infty$  durchaus vorkommen kann. Wir können eine Risikofunktion so *interpretieren*, dass sie den durchschnittlichen Verlust angibt, wenn wir  $\hat{\Theta}$  sehr häufig auf Beobachtungen der Länge  $n$  anwenden. Dieser erwartete Verlust sollte natürlich möglichst klein sein.

**Satz 2.1.6**

Es sei  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}^m$  ein integrierbarer Schätzer und  $L$  die Verlustfunktion der kleinsten Quadrate. Dann gilt für alle  $\vartheta \in \Theta$ :

$$\mathcal{R}(\vartheta, \hat{\Theta}) = \text{Var}_\vartheta \hat{\Theta} \circ X + (\text{Bias}_\vartheta \hat{\Theta})^2$$

**Beweis:** Für den Beweis betrachte

$$\begin{aligned}
 \mathcal{R}(\vartheta, \hat{\Theta}) &= \int (\vartheta - \hat{\Theta} \circ X)^2 dP_{\vartheta}^n = \vartheta^2 - 2\vartheta \int \hat{\Theta} \circ X dP_{\vartheta}^n + \int (\hat{\Theta} \circ X)^2 dP_{\vartheta}^n \\
 &= \vartheta^2 - 2\vartheta \int \hat{\Theta} \circ X dP_{\vartheta}^n + \left( \int \hat{\Theta} \circ X dP_{\vartheta}^n \right)^2 - \left( \int \hat{\Theta} \circ X dP_{\vartheta}^n \right)^2 + \int (\hat{\Theta} \circ X)^2 dP_{\vartheta}^n \\
 &= \vartheta^2 - 2\vartheta \int \hat{\Theta} \circ X dP_{\vartheta}^n + \underbrace{\left( \int \hat{\Theta} \circ X dP_{\vartheta}^n \right)^2}_{=(\text{Bias}_{\vartheta} \hat{\Theta})^2} \\
 &\quad - \underbrace{\left( \left( \int \hat{\Theta} \circ X dP_{\vartheta}^n \right)^2 - \int (\hat{\Theta} \circ X)^2 dP_{\vartheta}^n \right)}_{=-\text{Var}_{\vartheta} \hat{\Theta} \circ X}.
 \end{aligned}$$

Damit haben wir genau die gewünschte Identität. □

**Bemerkungen:**

- Die Güte eines Schätzers bezüglich der Verlustfunktion der kleinsten Quadrate wird also durch den Bias und die Varianz beschrieben.
- Ist  $\hat{\Theta}$  unverzerrt, so lässt sich die Güte allein durch die Varianz berechnen.

**Beispiel**

Für den Schätzer  $S_n^2$  gilt  $\mathcal{R}(\vartheta, S_n^2) = \text{Var}_{\vartheta} S_n^2$ , da  $S_n^2$  wie wir wissen unverzerrt ist. Für  $\frac{n-1}{n} S_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2$  gilt hingegen

$$\begin{aligned}
 \mathcal{R}\left(\vartheta, \frac{n-1}{n} S_n^2\right) &= \left(\frac{n-1}{n}\right)^2 \text{Var}_{\vartheta} S_n^2 + \left(\text{Bias}_{\vartheta} \frac{n-1}{n} S_n^2\right)^2 \\
 &= \underbrace{\left(\frac{n-1}{n}\right)^2}_{<1} \text{Var}_{\vartheta} S_n^2 + \underbrace{\frac{1}{n^2} (\text{Var}_{\vartheta} X_1)^2}_{\geq 0 = (\text{Bias}_{\vartheta} S_n^2)^2}.
 \end{aligned}$$

Die Frage ist nun, welche dieser beiden Abschätzungen überwiegt. Sei z. B.  $X_i \sim \mathcal{N}(\mu, \sigma^2)$ , dann kann gezeigt werden, dass  $\text{Var}_{\sigma^2} S_n^2 = \frac{2\sigma^4}{n-1}$  ist. Damit folgt

$$\begin{aligned}
 \mathcal{R}\left(\sigma^2, \frac{n-1}{n} S_n^2\right) &= \left(\frac{n-1}{n}\right)^2 \frac{2\sigma^4}{n-1} + \frac{1}{n^2} \sigma^4 = \frac{(2n-1)\sigma^4}{n^2} = \underbrace{\frac{(2n-1)(n-1)}{2n^2}}_{<1} \cdot \frac{2\sigma^4}{n-1} \\
 &< \text{Var}_{\sigma^2} S_n^2 = \mathcal{R}(\sigma^2, S_n^2).
 \end{aligned}$$

Wir können uns nun fragen, ob diese Abschätzung immer gilt. Dafür definieren wir den **Exzess-Kurtosis** von Verteilungen als

$$\kappa_{\vartheta} := \frac{\mathbf{E}_{\vartheta}(X_1 - \mathbf{E}_{\vartheta} X_1)^4}{\sigma_{\vartheta}^4} - 3.$$

Auf genauere Herleitungen und dergleichen wollen wir hier ebenso verzichten wie auf den Nachweis dieser Gleichung:

$$Z_{\vartheta}^2 := \text{Var}_{\vartheta} S_n^2 = \sigma_{\vartheta}^4 \left( \frac{2}{n-1} + \frac{\kappa_{\vartheta}}{n} \right)$$

Nun gilt  $\mathcal{R}(\vartheta, \frac{n-1}{n} S_n^2) \leq \mathcal{R}(\vartheta, S_n^2)$  genau dann, wenn

$$\begin{aligned} \left( \frac{n-1}{n} \right)^2 Z_{\vartheta}^2 + \frac{1}{n^2} \sigma_{\vartheta}^4 &\leq Z_{\vartheta}^2 \iff \frac{1}{n^2} \sigma_{\vartheta}^4 \leq \left( 1 - \left( \frac{n-1}{n} \right)^2 \right) Z_{\vartheta}^2 = \frac{2n-1}{n^2} Z_{\vartheta}^2 \\ &\iff \sigma_{\vartheta}^4 \leq (2n-1) \sigma_{\vartheta}^4 \left( \frac{2}{n-1} + \frac{\kappa_{\vartheta}}{n} \right) \\ &\iff \kappa_{\vartheta} \frac{2n-1}{n} \geq 1 - \frac{4n-2}{n-1} = \frac{1-3n}{n-1} \\ &\iff \kappa_{\vartheta} \geq \frac{1-3n}{n-1} \cdot \frac{n}{2n-1} \nearrow -\frac{3}{2}. \end{aligned}$$

Damit können wir nun folgern, dass für  $\kappa_{\vartheta} \geq -\frac{3}{2}$  für alle  $n \geq 2$  folgt, dass  $\mathcal{R}(\vartheta, \frac{n-1}{n} S_n^2) < \mathcal{R}(\vartheta, S_n^2)$  gilt. Umgekehrt gilt für  $\kappa_{\vartheta} < -\frac{3}{2}$ , dass es ein  $n_0$  gibt, so dass für alle  $n \geq n_0$  die umgekehrte Abschätzung  $\mathcal{R}(\vartheta, \frac{n-1}{n} S_n^2) > \mathcal{R}(\vartheta, S_n^2)$  gilt. Wir geben hier noch einige Beispiele für den Exzess-Kurtosis:

- Für  $\mathcal{N}(\mu, \sigma^2)$  gilt  $\kappa = 0$ .
- Für die Gleichverteilung auf  $[a, b]$  ist  $\kappa = -1.2$ .
- Für  $B(1, p)$  ist  $\kappa = \frac{1-6p(1-p)}{p(1-p)}$ . Damit gilt  $\kappa < -\frac{3}{2}$  genau für  $p \in (\frac{1}{3}, \frac{2}{3})$ . Insbesondere gilt für  $p = \frac{1}{2}$ , dass  $\kappa = -2$  ist. Man kann zeigen, dass  $\kappa \geq -2$  für *alle* Verteilungen gilt. //

## 2.2 Optimalität von Schätzern

Wir wollen nun davon ausgehen, dass wir 1-dimensionale Schätzprobleme vorliegen haben, es gilt also  $\Theta \subset \mathbf{R}$ . Ferner sei  $L$  stets die Verlustfunktion der kleinsten Quadrate. Wir stellen uns nun die Frage, wann ein Schätzer gleichmäßig besser bezüglich des erwarteten Verlustes ist als alle anderen Schätzer. Die erste Beobachtung ist, dass es im Allgemeinen keinen gleichmäßig besten Schätzer gibt. Im folgenden Satz fassen wir dies genauer:

### Satz 2.2.1

Es sei  $\Theta \subset \mathbf{R}$  messbar mit  $|\Theta| \geq 2$  und  $(P_\vartheta)_{\vartheta \in \Theta}$  sei derart, dass es  $\vartheta_1 \neq \vartheta_2$  aus  $\Theta$  mit  $P_{\vartheta_2} \ll P_{\vartheta_1}$  gibt, d. h. aus  $P_{\vartheta_1}(N) = 0$  folgt auch  $P_{\vartheta_2}(N) = 0$  für alle messbaren  $N \subset \mathbf{R}$ . Dann gilt für die Verlustfunktion der kleinsten Quadrate, dass es für alle Schätzer  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  einen Schätzer  $\hat{\Theta}': \mathbf{R}^n \rightarrow \mathbf{R}$  und ein  $\vartheta \in \Theta$  mit  $\mathcal{R}(\vartheta, \hat{\Theta}') < \mathcal{R}(\vartheta, \hat{\Theta})$  gibt.

**Beweis:** Wir betrachten die Schätzer  $\hat{\Theta}_i: \mathbf{R}^n \rightarrow \mathbf{R}$  mit  $\hat{\Theta}_i(x) := \vartheta_i$  für  $i \in \{1, 2\}$  und alle  $x \in \mathbf{R}^n$ . Dann gilt  $\mathcal{R}(\vartheta_i, \hat{\Theta}_i) = \int (\vartheta_i - \hat{\Theta}_i(x))^2 dP_{\vartheta_i}^n(x) = 0$ . Es sei nun  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  ein beliebiger Schätzer. Wäre  $\mathcal{R}(\vartheta_1, \hat{\Theta}_1) < \mathcal{R}(\vartheta_1, \hat{\Theta})$ , so wären wir bereits fertig. Es sei daher nun  $\mathcal{R}(\vartheta_1, \hat{\Theta}_1) \geq \mathcal{R}(\vartheta_1, \hat{\Theta})$ . Wegen  $\mathcal{R}(\vartheta_1, \hat{\Theta}_1) = 0$ ,  $\mathcal{R}(\vartheta_1, \hat{\Theta}) \geq 0$  und

$$0 = \mathcal{R}(\vartheta_1, \hat{\Theta}_1) = \mathcal{R}(\vartheta_1, \hat{\Theta}) = \int \underbrace{(\vartheta_1 - \hat{\Theta})^2}_{\geq 0} dP_{\vartheta_1}^n$$

folgt  $P_{\vartheta_1}^n$ -fast sicher auch  $\hat{\Theta} = \vartheta_1$ . Da ferner  $P_{\vartheta_2} \ll P_{\vartheta_1}$  ist folgt mit dem Satz von Radon-Nikodym, dass  $P_{\vartheta_2}$  eine Dichte  $f$  bezüglich  $P_{\vartheta_1}$  hat, d. h. es gilt  $P_{\vartheta_2}(A) = \int_A f dP_{\vartheta_1}$ . Mit dem Satz von Fubini folgt dann, dass  $\otimes_1^n f(x) := f(x_1) \cdot \dots \cdot f(x_n)$  eine Dichte von  $P_{\vartheta_2}^n$  bezüglich  $P_{\vartheta_1}^n$  ist. Wenden wir wieder den Satz von Radon-Nikodym an, so folgt  $P_{\vartheta_2}^n \ll P_{\vartheta_1}^n$ . Dann folgt  $P_{\vartheta_2}^n$ -fast sicher  $\hat{\Theta} = \vartheta_1$  und damit schließlich

$$\mathcal{R}(\vartheta_2, \hat{\Theta}) = \int (\vartheta_2 - \hat{\Theta})^2 dP_{\vartheta_2}^n = (\vartheta_2 - \vartheta_1)^2 > 0 = \mathcal{R}(\vartheta_2, \hat{\Theta}_2). \quad \square$$

### Bemerkungen:

- Dieses Ergebnis gilt auch für allgemeinere Verlustfunktionen und mehrdimensionale Parametermengen, wenn ein paar technische Modifikationen vorgenommen werden. Insbesondere sollte angenommen werden, dass  $L(\vartheta, \vartheta) < L(\vartheta, \vartheta')$  für alle  $\vartheta \neq \vartheta'$  ist.
- Der Beweis funktioniert, da die Schätzer  $\hat{\Theta}_i$  den Parameter  $\vartheta_i$  „kennen“. Gegen solche Schätzer hat es keinen Sinn zu vergleichen.

In einem ersten Ansatz können wir konstante Schätzer ausschließen und die Eingangsfrage erneut stellen. Wir werden in Korollar 2.2.3 sehen, dass dieser Ansatz das Problem jedoch nicht löst.

**Satz 2.2.2**

Es sei  $\Theta \subset \mathbf{R}$  und  $\hat{\Theta}_1, \hat{\Theta}_2: \mathbf{R}^n \rightarrow \mathbf{R}$  seien quadrat-integrierbare Schätzer, d. h. es gilt  $\mathbf{E}_\vartheta \hat{\Theta}_i^2 < \infty$  für alle  $\vartheta \in \Theta$  und  $i \in \{1, 2\}$ . Ferner seien  $\hat{\Theta}_1$  und  $\hat{\Theta}_2$  unterschiedlich, d. h.  $\mathbf{E}_\vartheta (\hat{\Theta}_1 - \hat{\Theta}_2)^2 > 0$  für alle  $\vartheta \in \Theta$ . Dann gilt für alle  $\vartheta \in \Theta$  und alle  $\alpha \in (0, 1)$ :

$$\mathcal{R}(\vartheta, \alpha \hat{\Theta}_1 + (1 - \alpha) \hat{\Theta}_2) < \alpha \mathcal{R}(\vartheta, \hat{\Theta}_1) + (1 - \alpha) \mathcal{R}(\vartheta, \hat{\Theta}_2).$$

Mit anderen Worten bedeutet dies, dass  $\mathcal{R}(\vartheta, \cdot)$  eine strikt konvexe Abbildung ist.

**Beweis:** Wir definieren  $h: [0, 1] \rightarrow [0, \infty)$  durch  $h(\alpha) := \mathcal{R}(\vartheta, \alpha \hat{\Theta}_1 + (1 - \alpha) \hat{\Theta}_2)$  und wollen zeigen, dass  $h$  strikt konvex ist, denn dann folgt

$$h(\alpha) = h(\alpha \cdot 1 + (1 - \alpha) \cdot 0) < \alpha h(1) + (1 - \alpha) h(0) = \alpha \mathcal{R}(\vartheta, \hat{\Theta}_1) + (1 - \alpha) \mathcal{R}(\vartheta, \hat{\Theta}_2).$$

Aus der Analysis wissen wir, dass es bereits genügt, zu zeigen, dass  $h$  zweimal stetig differenzierbar mit stetigen Fortsetzungen auf den Rändern  $[0, 1]$  ist und dass  $h''(\alpha) > 0$  für alle  $\alpha \in (0, 1)$  gilt. Mit Hilfe der trinomischen Formel  $(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$  erhalten wir nun

$$\begin{aligned} h(\alpha) &= \int (\vartheta - \alpha \hat{\Theta}_1 - (1 - \alpha) \hat{\Theta}_2)^2 dP_\vartheta^n \\ &= \int \vartheta^2 + \alpha^2 \hat{\Theta}_1^2 + (1 - \alpha)^2 \hat{\Theta}_2^2 - 2\vartheta \alpha \hat{\Theta}_1 - 2\vartheta(1 - \alpha) \hat{\Theta}_2 + 2\alpha(1 - \alpha) \hat{\Theta}_1 \hat{\Theta}_2 dP_\vartheta^n. \end{aligned}$$

Da dies im Wesentlichen ein Polynom ist, folgt nun

$$h'(\alpha) = \int 2\alpha \hat{\Theta}_1^2 - 2(1 - \alpha) \hat{\Theta}_2^2 - 2\vartheta \hat{\Theta}_1 + 2\vartheta \hat{\Theta}_2 + 2(1 - 2\alpha) \hat{\Theta}_1 \hat{\Theta}_2 dP_\vartheta^n.$$

Nochmaliges Ableiten liefert

$$h''(\alpha) = \int 2\hat{\Theta}_1^2 + 2\hat{\Theta}_2^2 - 4\hat{\Theta}_1 \hat{\Theta}_2 dP_\vartheta^n = 2 \int (\hat{\Theta}_1 - \hat{\Theta}_2)^2 dP_\vartheta^n > 0. \quad \square$$

**Korollar 2.2.3**

Es sei  $\Theta \subset \mathbf{R}$ ,  $L$  die Verlustfunktion der kleinsten Quadrate und  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  ein Schätzer mit  $\mathcal{R}(\vartheta, \hat{\Theta}) > 0$  für ein  $\vartheta \in \Theta$ , d. h.  $\hat{\Theta}$  ist  $P_\vartheta^n$ -fast sicher nicht konstant. Dann gibt es einen Schätzer  $\hat{\Theta}': \mathbf{R}^n \rightarrow \mathbf{R}$  mit  $0 < \mathcal{R}(\vartheta, \hat{\Theta}') < \mathcal{R}(\vartheta, \hat{\Theta})$ .

Mit anderen Worten ist  $\hat{\Theta}'$  also ein nicht-konstanter Schätzer, der echt besser als  $\hat{\Theta}$  ist. Ist  $\Theta$  ferner beschränkt und  $\hat{\Theta}$  unverzerrt, so können für  $\hat{\Theta}'$  für beliebiges  $\varepsilon > 0$  so wählen, dass  $|\text{Bias}_{\vartheta'} \hat{\Theta}'| \leq \varepsilon$  für alle  $\vartheta' \in \Theta$  gilt.

**Beweis:** Wir setzen  $\hat{\Theta}_\vartheta(x) := 0$  für alle  $x \in \mathbf{R}^n$  und  $\hat{\Theta}' := (1 - \alpha)\hat{\Theta} + \alpha\hat{\Theta}_\vartheta$  für ein  $\alpha \in (0, 1)$ . Mit Satz 2.2.2, der die Konvexität behandelte, folgt nun

$$\mathcal{R}(\vartheta, \hat{\Theta}') < (1 - \alpha)\mathcal{R}(\vartheta, \hat{\Theta}) + \underbrace{\alpha\mathcal{R}(\vartheta, \hat{\Theta}_\vartheta)}_{=0} < \mathcal{R}(\vartheta, \hat{\Theta}).$$

Wäre  $\mathcal{R}(\vartheta, \hat{\Theta}') = 0$ , so folgt ferner  $P_\vartheta^n$ -fast sicher  $\hat{\Theta}' = 0$  und daraus wiederum  $P_\vartheta^n$ -fast sicher  $\vartheta = (1 - \alpha)\hat{\Theta} + \alpha\vartheta$ . Dann wäre aber  $\vartheta = \hat{\Theta}$  und wir erhielten einen Widerspruch zur vorausgesetzten nicht-Konstanzheit.

Für die zweite Aussage des Korollars betrachten wir

$$\text{Bias}_{\vartheta'} \hat{\Theta}' = \mathbf{E}_{\vartheta'}((1 - \alpha)\hat{\Theta} + \alpha\vartheta) - \vartheta' = \underbrace{\mathbf{E}_{\vartheta'}((1 - \alpha)\hat{\Theta})}_{=\text{Bias}_{\vartheta'} \hat{\Theta} = 0} - (1 - \alpha)\vartheta' + \alpha(\vartheta - \vartheta') = \alpha(\vartheta - \vartheta').$$

Wir können  $\alpha$  nun klein genug wählen und erhalten so die Aussage des Korollars.  $\square$

**Bemerkungen:**

- Aus dem zweiten Teil des Beweises erhalten wir auch die Aussage  $\text{Bias}_\vartheta \hat{\Theta}' = 0$ .
- Das Korollar gilt auch für alle anderen strikt konvexen Verlustfunktionen, da sich die strikte Konvexität von  $L$  auf  $\mathcal{R}(\vartheta, \cdot)$  überträgt. Dies ist technisch ein wenig aufwendiger, konzeptionell aber analog beweisbar.

Das Fazit dieser Arbeit ist nun, dass wir die Klasse der betrachteten Schätzer deutlich weiter einschränken müssen, wenn wir die Existenz gleichmäßig bester Schätzer garantieren wollen. Dies führt uns zur nun folgenden Definition 2.2.4.

**Definition 2.2.4 Bester unverzerrter Schätzer**

Es sei  $\hat{\Theta}^* : \mathbf{R}^n \rightarrow \mathbf{R}$  ein unverzerrter, quadrat-integrierbarer Schätzer. Dann heißt  $\hat{\Theta}^*$  **bester unverzerrter Schätzer** genau dann, wenn  $\mathcal{R}(\vartheta, \hat{\Theta}^*) \leq \mathcal{R}(\vartheta, \hat{\Theta})$  für alle *unverzerrten* Schätzer  $\hat{\Theta} : \mathbf{R}^n \rightarrow \mathbf{R}$  und alle  $\vartheta \in \Theta$  gilt.

Nach Satz 2.1.6 ist dies äquivalent dazu, dass  $\text{Var}_\vartheta \hat{\Theta}^* \leq \text{Var}_\vartheta \hat{\Theta}$  für alle Schätzer  $\hat{\Theta} : \mathbf{R}^n \rightarrow \mathbf{R}$  und alle  $\vartheta \in \Theta$  gilt.

Damit haben wir zunächst den Begriff des besten unverzerrten Schätzers definiert, müssen uns nun also fragen, ob es einen solchen Schätzer  $\hat{\Theta}^*$  überhaupt gibt und falls ja, ob dieser eindeutig ist.

**Satz 2.2.5**

Es sei  $\Theta \subset \mathbf{R}$  messbar und  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen. Dann gibt es *höchstens einen* besten unverzerrten Schätzer  $\hat{\Theta}^*$ .

**Beweis:** Es seien  $\hat{\Theta}^*$  und  $\hat{\Theta}^\times$  zwei beste unverzerrte Schätzer mit  $\mathbf{E}_\vartheta(\hat{\Theta}^* - \hat{\Theta}^\times)^2 > 0$  für ein  $\vartheta \in \Theta$ . Damit sind diese beiden Schätzer insbesondere nicht gleich. Wir definieren nun  $\hat{\Theta} := \frac{1}{2}\hat{\Theta}^* + \frac{1}{2}\hat{\Theta}^\times$  und sehen, dass  $\hat{\Theta}$  ebenfalls quadrat-integrierbar und unverzerrt ist. Aus der Optimalität folgt  $\mathcal{R}(\vartheta, \hat{\Theta}^*) \leq \mathcal{R}(\vartheta, \hat{\Theta}^\times)$  und ebenso die umgekehrte Ungleichung. Mit Satz 2.2.2 folgt nun

$$\mathcal{R}(\vartheta, \hat{\Theta}) < \frac{1}{2}\mathcal{R}(\vartheta, \hat{\Theta}^*) + \frac{1}{2}\mathcal{R}(\vartheta, \hat{\Theta}^\times) = \frac{1}{2}\mathcal{R}(\vartheta, \hat{\Theta}^*) + \frac{1}{2}\mathcal{R}(\vartheta, \hat{\Theta}^*) = \mathcal{R}(\vartheta, \hat{\Theta}^*).$$

Dies steht jedoch im Widerspruch zur Optimalität von  $\hat{\Theta}^*$ . □

### Definition 2.2.6 Nullschätzer

Ein integrierbarer Schätzer  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  heißt **Nullschätzer** für ein  $\vartheta_0 \in \Theta$  genau dann, wenn die folgenden beiden Bedingungen erfüllt sind:

- i)  $\hat{\Theta} \in \mathcal{L}_2(P_{\vartheta_0}^n)$ , das heißt  $\mathbf{E}_{\vartheta_0}\hat{\Theta}^2 < \infty$ .
- ii)  $\mathbf{E}_\vartheta\hat{\Theta} = 0$  für alle  $\vartheta \in \Theta$ .

Ferner heißt  $\hat{\Theta}$  **Nullschätzer** genau dann, wenn  $\hat{\Theta}$  Nullschätzer für alle  $\vartheta \in \Theta$  ist.

### Bemerkungen:

- Die Menge der Nullschätzer bezüglich  $\vartheta_0$  ist ein linearer Teilraum von  $\mathcal{L}_2(P_{\vartheta_0}^n)$ , da  $\mathbf{E}_{\vartheta_0}$  linear ist.
- Sind  $\hat{\Theta}, \hat{\Theta}' \in \mathcal{L}_2(P_{\vartheta_0}^n)$  unverzerrte Schätzer, so ist  $\hat{\Theta} := \hat{\Theta} - \hat{\Theta}'$  ein Nullschätzer für  $\vartheta_0$ . Umgekehrt gilt auch: Ist  $\hat{\Theta} \in \mathcal{L}_2(P_{\vartheta_0}^n)$  unverzerrt und  $\hat{\Theta}$  ein Nullschätzer, so ist  $\hat{\Theta}' := \hat{\Theta} + t\hat{\Theta}$  für alle  $t \in \mathbf{R}$  unverzerrt.

### Satz 2.2.7 Satz von Rao

Es sei  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  ein unverzerrter Schätzer und  $\vartheta_0 \in \Theta$  derart, dass  $\hat{\Theta} \in \mathcal{L}_2(P_{\vartheta_0}^n)$  gilt. Dann sind folgende Aussagen äquivalent:

- i) Es gilt  $\mathcal{R}(\vartheta_0, \hat{\Theta}) \leq \mathcal{R}(\vartheta_0, \hat{\Theta}')$  für alle unverzerrten Schätzer  $\hat{\Theta}' \in \mathcal{L}_2(P_{\vartheta_0}^n)$ .
- ii) Es gilt  $\mathbf{E}_{\vartheta_0}\hat{\Theta}\hat{\Theta} = 0$  für alle Nullschätzer  $\hat{\Theta}$  bezüglich  $\vartheta_0$ .

Insbesondere ist  $\hat{\Theta}$  bester unverzerrter Schätzer genau dann, wenn  $\mathbf{E}_\vartheta\hat{\Theta}\hat{\Theta} = 0$  für alle Nullschätzer  $\hat{\Theta}$  und alle  $\vartheta \in \Theta$  gilt.

**Beweis:** Für die Richtung von i) nach ii) sei  $t \in \mathbf{R}$  und  $\hat{\Theta}$  ein Nullschätzer bezüglich  $\vartheta_0$ . Dann ist  $\hat{\Theta}' := \hat{\Theta} + t\hat{\Theta} \in \mathcal{L}_2(P_{\vartheta_0}^n)$  unverzerrt. Damit gilt  $\mathcal{R}(\vartheta_0, \hat{\Theta}) = \text{Var}_{\vartheta_0} \hat{\Theta}$  und  $\mathcal{R}(\vartheta_0, \hat{\Theta}') = \text{Var}_{\vartheta_0} \hat{\Theta}'$ .

Für  $t > 0$  gilt nach Voraussetzung zudem  $0 \leq t^{-1} (\text{Var}_{\vartheta_0} (\hat{\Theta} + t\hat{\Theta}) - \text{Var}_{\vartheta_0} \hat{\Theta})$ . Überdies gilt auch die allgemeine Formel

$$\begin{aligned} \text{Var}(X + Y) &= \mathbf{E}(X - \mathbf{E}X + Y - \mathbf{E}Y)^2 \\ &= \text{Var} X + \text{Var} Y + \underbrace{2\mathbf{E}((X - \mathbf{E}X)(Y - \mathbf{E}Y))}_{=: \text{Cov}(X, Y)}, \end{aligned}$$

wobei  $\text{Cov}(X, Y)$  die wie hier definierte **Kovarianz** bezeichnet. Damit folgt nun

$$\begin{aligned} 0 &\leq t^{-1} (\text{Var}_{\vartheta_0} \hat{\Theta} + t^2 \text{Var}_{\vartheta_0} \hat{\Theta} + 2t \text{Cov}(\hat{\Theta}, \hat{\Theta}) - \text{Var}_{\vartheta_0} \hat{\Theta}) \\ &= 2\text{Cov}(\hat{\Theta}, \hat{\Theta}) + t \text{Var}_{\vartheta_0} \hat{\Theta}. \end{aligned}$$

Für  $t \searrow 0$  folgt dann  $\text{Cov}(\hat{\Theta}, \hat{\Theta}) \geq 0$ . Analog gilt für  $t < 0$  auch  $0 \geq 2\text{Cov}(\hat{\Theta}, \hat{\Theta}) + t \text{Var}_{\vartheta_0} \hat{\Theta}$  und damit  $\text{Cov}(\hat{\Theta}, \hat{\Theta}) \leq 0$ . Insgesamt erhalten wir also

$$\begin{aligned} 0 &= \text{Cov}(\hat{\Theta}, \hat{\Theta}) = \mathbf{E}_{\vartheta_0} ((\hat{\Theta} - \mathbf{E}_{\vartheta_0} \hat{\Theta})(\hat{\Theta} - \mathbf{E}_{\vartheta_0} \hat{\Theta})) = \mathbf{E}_{\vartheta_0} \hat{\Theta} \hat{\Theta} - \vartheta_0 \mathbf{E}_{\vartheta_0} \hat{\Theta} \quad (*) \\ &= \mathbf{E}_{\vartheta_0} \hat{\Theta} \hat{\Theta}. \end{aligned}$$

Für die Richtung von ii) nach i) sei  $\hat{\Theta}' \in \mathcal{L}_2(P_{\vartheta_0}^n)$  unverzerrt. Dann ist  $\hat{\Theta} := \hat{\Theta}' - \hat{\Theta}$  ein Nullschätzer und wegen  $\text{Var}_{\vartheta_0} \geq 0$  folgt:

$$\begin{aligned} \text{Var}_{\vartheta_0} \hat{\Theta}' &= \text{Var}_{\vartheta_0} (\hat{\Theta} + \hat{\Theta}) = \text{Var}_{\vartheta_0} \hat{\Theta} + \text{Var}_{\vartheta_0} \hat{\Theta} + 2\text{Cov}(\hat{\Theta}, \hat{\Theta}) \\ &\stackrel{(*)}{\geq} \text{Var}_{\vartheta_0} \hat{\Theta} + 2\mathbf{E}_{\vartheta_0} \hat{\Theta} \hat{\Theta} = \text{Var}_{\vartheta_0} \hat{\Theta}. \quad \square \end{aligned}$$

### Beispiel 2.2.8 Anwendung des Satzes von Rao

Es sei  $P_{\vartheta} := B(1, \vartheta)$  für  $\vartheta \in [0, 1]$ . Dann ist  $\overline{X}_n$  bester unverzerrter Schätzer für  $\vartheta$ .

**Beweis:** Zunächst ist nach Lemma 1.2.2 klar, dass  $\overline{X}_n$  quadrat-integrierbar und unverzerrt ist. Unser Ziel ist es, Satz 2.2.7 anzuwenden. Dazu sei  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  ein Nullschätzer und  $h: [0, 1] \rightarrow \mathbf{R}$  mit  $h(\vartheta) := \mathbf{E}_{\vartheta} \hat{\Theta}$  für  $\vartheta \in [0, 1]$ . Offensichtlicherweise folgt daraus  $h(\vartheta) = 0$  und damit auch  $h'(\vartheta) = 0$  für alle  $\vartheta \in [0, 1]$ . Andererseits gilt

$$h(\vartheta) = \sum_{x \in \{0, 1\}^n} \hat{\Theta}(x) \vartheta^{\sum_{i=1}^n x_i} (1 - \vartheta)^{n - \sum_{i=1}^n x_i},$$

woraus wir dann für die Ableitung das Folgende erhalten:

$$\begin{aligned} 0 &= h'(\vartheta) = \sum_{x \in \{0, 1\}^n} \hat{\Theta}(x) \left( \frac{n\overline{x}_n}{\vartheta} - \frac{n - n\overline{x}_n}{1 - \vartheta} \right) \vartheta^{n\overline{x}_n} (1 - \vartheta)^{n - n\overline{x}_n} \\ &= \frac{n}{\vartheta(1 - \vartheta)} \mathbf{E}_{\vartheta} (\overline{X}_n - \vartheta) \hat{\Theta} \\ &= \frac{n}{\vartheta(1 - \vartheta)} \mathbf{E}_{\vartheta} \overline{X}_n \hat{\Theta} \end{aligned}$$

Damit sind wir für  $\vartheta \in (0, 1)$  fertig. Die beiden Randpunkte können separat elementar betrachtet werden, dies führen wir hier jedoch nicht aus.  $\square$



Dies beendet unser Beispiel. //

**Bemerkungen:**

- i) Die entwickelte Theorie benötigt keine reellen Beobachtungen. Alles bleibt unverändert für allgemeine  $x_i \in \mathcal{X}$ .
- ii) Satz 2.2.7 liefert noch für einige andere Schätzer die Optimalität. Der Satz liefert uns jedoch lediglich ein Kriterium zur Verifizierung der Optimalität, *nicht* jedoch zur Konstruktion optimaler Schätzer.
- iii) Die Bedingung ii) des Satzes besagt, dass der beste unverzerrte Schätzer orthogonal zu den Nullschätzern ist.
- iv) Die Existenz bester unverzerrter Schätzer ist eher unüblich als üblich.

**Satz 2.2.9 Satz von Chapman-Robbins**

Es seien  $\vartheta_0 \neq \vartheta \in \Theta$  und  $P_\vartheta \ll P_{\vartheta_0}$ . Ferner sei  $f$  die Dichte von  $P_\vartheta$  bezüglich  $P_{\vartheta_0}$  und  $f^n := \otimes_1^n f$ , das heißt  $f^n(x) = f(x_1) \cdot \dots \cdot f(x_n)$ , sowie  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  ein erwartungstreuer Schätzer. Dann gilt

$$\mathcal{R}(\vartheta_0, \hat{\Theta}) = \text{Var}_{\vartheta_0}(\hat{\Theta}) \geq \frac{(\vartheta - \vartheta_0)^2}{\text{Var}_{\vartheta_0}(f^n)}.$$

**Beweis:** Wir betrachten zunächst

$$\vartheta - \vartheta_0 = \mathbf{E}_\vartheta(\hat{\Theta}) - \mathbf{E}_{\vartheta_0}(\hat{\Theta}) = \int \hat{\Theta} f^n - \hat{\Theta} dP_{\vartheta_0}^n = \int \hat{\Theta}(f^n - 1) dP_{\vartheta_0}^n = \int (\hat{\Theta} - \vartheta_0)(f^n - 1) dP_{\vartheta_0}^n,$$

wobei der letzte Schritt wegen  $\int f^n - 1 dP_{\vartheta_0}^n = 0$  folgt. Durch Anwendung der Cauchy-Schwartz-Ungleichung folgt nun

$$\leq \left( \mathbf{E}_{\vartheta_0}(\hat{\Theta} - \vartheta_0)^2 \right)^{\frac{1}{2}} \left( \mathbf{E}_{\vartheta_0}(f^n - 1)^2 \right)^{\frac{1}{2}} = (\text{Var}_{\vartheta_0} \hat{\Theta})^{\frac{1}{2}} (\text{Var}_{\vartheta_0} f^n)^{\frac{1}{2}}.$$

Durch Quadrieren erhalten wir dann das Gewünschte. □

**Korollar 2.2.10**

Es seien die selben Voraussetzungen wie im vorausgehenden Satz 2.2.9 gegeben. Ferner sei  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  ein unverzerrter Schätzer, für welchen die Gleichheit

$$\text{Var}_{\vartheta_0} \hat{\Theta} = \frac{(\vartheta - \vartheta_0)^2}{\text{Var}_{\vartheta_0} f^n}$$

gelte. Dann ist  $\hat{\Theta}$  bester unverzerrter Schätzer für ein  $\vartheta$ .

**Beweis:** Die Gültigkeit der Aussage des Korollars ist klar, wenn ein Schätzer, wie vorausgesetzt, die untere Schranke des vorausgehenden Satzes annimmt.  $\square$

Im Allgemeinen wird die untere Schranke nicht angenommen. Notwendig zum Annehmen der Schranke ist  $\hat{\Theta} - \vartheta_0 = c(f^n - 1)$  mit  $c \in \mathbf{R}$ , da dann die Cauchy-Schwartz-Ungleichung eine echte Gleichung wird.

## 2.3 Konstruktionsmethoden für Punktschätzer

Wir wollen nun verschiedene Verfahren kennenlernen, um Schätzer zu konstruieren und diese näher untersuchen.

### Momentenmethode

Die erste Methode zur Konstruktion von Punktschätzern, die wir hier einführen wollen, ist die **Momentenmethode** nach Karl Pearson (\*1857 – †1936).

Dazu sei  $\Theta \subset \mathbf{R}^m$  für ein  $m \in \mathbf{R}$  und  $\mathbf{E}_\vartheta |X_i|^r < \infty$  für ein  $r \geq m$  und alle  $\vartheta \in \Theta$ . Ferner seien die Funktionen  $g_k: \Theta \rightarrow \mathbf{R}$  mit  $g_k(\vartheta) := \mathbf{E}_\vartheta X_1^k$  für  $k \in \{1, \dots, r\}$  bekannt. Dann ist unser Ansatz wie folgt:

- Für  $x = (x_1, \dots, x_n) \in \mathbf{R}^n$  betrachten wir die  $k$ -ten empirischen Momente  $\hat{m}_k(x) = \frac{1}{n} \sum_{i=1}^n x_i^k$ .
- Wir setzen  $\hat{m}_k(x) = g_k(\vartheta)$  für  $k = 1, \dots, r$  und versuchen dieses Gleichungssystem zu lösen.
- Wir betrachten nun die Abbildung  $g := (g_1, \dots, g_r): \Theta \rightarrow \mathbf{R}^r$  mit  $g(\Theta) := \{g(\vartheta) : \vartheta \in \Theta\}$  und  $\hat{m} := (\hat{m}_1, \dots, \hat{m}_r): \mathbf{R}^n \rightarrow \mathbf{R}^r$ .

#### Definition 2.3.1 Momentenschätzer

Falls  $\hat{m}(x) \in g(\Theta)$  für alle  $x \in \mathbf{R}^n$  gilt und  $g$  injektiv ist nennen wir

$$\hat{\Theta} = g^{-1} \circ \hat{m}: \mathbf{R}^n \rightarrow \Theta$$

einen **Momentenschätzer**, falls  $\hat{\Theta}$  zudem messbar ist.

#### Bemerkungen:

- Die Voraussetzung  $\hat{m}(x) \in g(\Theta)$  sichert, dass das Gleichungssystem mindestens eine Lösung hat.
- Die Injektivität von  $g$  macht es möglich, das Gleichungssystem sogar eindeutig zu lösen.
- Falls  $g^{-1}$  stetig ist, so ist  $\hat{\Theta}$  ein Momentenschätzer, der überdies konsistent ist, denn es gilt dann  $P_\vartheta^\infty$ -fast sicher

$$\hat{m}_k \circ X = \hat{m}_k(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i^k,$$

woraus  $\mathbf{E}_\vartheta X_1^k = g_k(\vartheta)$  und damit wiederum  $\hat{m} \circ X \rightarrow g(\vartheta)$   $P_\vartheta^\infty$ -fast sicher folgt. Dann folgt  $P_\vartheta^\infty$ -fast sicher aber auch  $\hat{\Theta}(X) = g^{-1} \circ \hat{m}(X) \rightarrow \vartheta$ .

#### Beispiel Normalverteilung

Es sei  $\Theta = \mathbf{R} \times (0, \infty)$  und  $P_\vartheta = \mathcal{N}(\mu, \sigma^2)$  für  $\vartheta = (\mu, \sigma^2)$ . Dann ist  $m = 2$  und  $r = 2$ . Ferner ist

$g_1(\mu, \sigma^2) = \mathbf{E}_{(\mu, \sigma^2)} X_1 = \mu$  und  $g_2(\mu, \sigma^2) = \mathbf{E}_{(\mu, \sigma^2)} X_1^2 = \sigma^2 + \mu^2$ . Damit erhalten wir das folgende Gleichungssystem:

- i)  $\frac{1}{n} \sum_{i=1}^n x_i = \mu$
- ii)  $\frac{1}{n} \sum_{i=1}^n x_i^2 = \mu^2 + \sigma^2$

Daraus folgt nun

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^2 = \frac{1}{n} \left( \sum_{i=1}^n x_i^2 - \bar{x}_n^2 \right) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2,$$

wobei wir die letzte Gleichheit bereits im Beweis von Lemma 1.2.5 gezeigt hatten. Damit ist der Momentenschätzer durch  $\left( \bar{X}_n, \frac{n-1}{n} S_n^2 \right)$  gegeben. Insbesondere ist dieser Schätzer nicht erwartungstreu. //

**Beispiel Binomialverteilung**

Wir betrachten nun  $\Theta = \mathbf{N} \times [0, 1]$  und  $P_\vartheta = B(k, p)$  für  $\vartheta = (k, p)$ . Mit anderen Worten bedeutet dies, dass wir weder  $p$  noch  $k$  kennen. Es ist  $m = r = 2$ ,  $g_1(k, p) = kp$  und  $g_2(k, p) = kp(1-p) + k^2 p^2$ . Damit ist das Gleichungssystem gegeben durch:

- i)  $\frac{1}{n} \sum_{i=1}^n x_i = kp$
- ii)  $\frac{1}{n} \sum_{i=1}^n x_i^2 = kp(1-p) + k^2 p^2$

Das Einsetzen der ersten in die zweite Gleichung liefert  $\frac{1}{n} \sum_{i=1}^n x_i^2 = \bar{x}_n(1-p) + \bar{x}_n^2$ , was äquivalent zu  $\frac{n-1}{n} s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - \bar{x}_n^2) = \bar{x}_n(1-p)$  ist. Damit erhalten wir

$$p = \begin{cases} \frac{\bar{x}_n - \frac{n-1}{n} s_n^2}{\bar{x}_n} & \bar{x}_n > 0 \\ 0 := 0 & \text{sonst} \end{cases}.$$

Ferner ist  $k = \frac{\bar{x}_n}{p}$ . Für  $\bar{X}_n \neq 0$  erhalten wir damit den Momentenschätzer

$$\left( \frac{\bar{X}_n^{-2}}{\bar{X}_n - \frac{n-1}{n} S_n^2}, \frac{\bar{X}_n - \frac{n-1}{n} S_n^2}{\bar{X}_n} \right). //$$

**Bemerkungen:**

- Es gibt Familien  $(P_\vartheta)_{\vartheta \in \Theta}$ , für die das Gleichungssystem für  $r = m$  nicht eindeutig lösbar ist.
- Es kann vorkommen, dass  $m(x) \notin g(\Theta)$  ist. In diesem Fall ergibt die Momentenmethode keinen Sinn und kann nicht angewendet werden.

## Maximum-Likelihood-Schätzer

Der nächste Schätzer, den wir konstruieren wollen, ist der **Maximum-Likelihood-Schätzer**. Dieser stellt eine der wichtigsten Methoden zur Gewinnung von Schätzern dar und wurde erstmals von Gauss erwähnt, später dann von Fisher (\*1890 – †1962) weiterentwickelt.

Wir gehen davon aus, dass  $\Theta \subset \mathbf{R}^m$  ist und die  $P_\vartheta$  Wahrscheinlichkeitsmaße auf  $\mathbf{R}$  sind. Diese zweite Annahme ist nicht notwendig, der Einfachheit halber treffen wir sie dennoch. Ferner existiere ein Maß  $\mu$  auf  $\mathbf{R}$  mit den folgenden Eigenschaften:

- i) Es gilt  $P_\vartheta \ll \mu$ .
- ii) Jedes  $P_\vartheta$  besitze eine „kanonische“ Dichte  $f(\cdot, \vartheta)$  bezüglich  $\mu$ , das heißt zum Beispiel:
  - a)  $\mu$  sei das Zählmaß auf  $C \subset \mathbf{R}$  und  $C$  sei abzählbar. Dann ist

$$f(x, \vartheta) = \begin{cases} P_\vartheta(\{x\}) & x \in C \\ 0 & \text{sonst} \end{cases}.$$

- b) Sei  $\mu$  das Lebesguemaß und die  $P_\vartheta$  haben die stetige Dichte  $f(\cdot, \vartheta)$  bezüglich  $\mu$ . Dann ist dies die kanonische Dichte.

### Definition 2.3.2 Likelihood-Funktion

Es sei  $f(\cdot, \vartheta)$  die kanonische Dichte von  $P_\vartheta$  bezüglich  $\mu$ . Dann heißt für

$$L: \mathbf{R}^n \times \Theta \rightarrow [0, \infty)$$

$$L(x_1, \dots, x_n, \vartheta) := L(x, \vartheta) := \prod_{i=1}^n f(x_i, \vartheta) \quad \text{mit } x \in \mathbf{R}^n, \vartheta \in \Theta$$

die Abbildung  $\vartheta \mapsto L(x, \vartheta)$  die **Likelihood-Funktion** der Stichprobe  $x \in \mathbf{R}^n$ .

Zu beachten ist, dass  $L$  hier für *likelihood* und nicht für *loss* steht. Diese doppelte Namensgebung ist in der Fachliteratur jedoch üblich. Da  $L(\cdot, \vartheta) = \otimes_{i=1}^n f(\cdot, \vartheta)$  gilt ist die i. i. d.-Annahme fest in  $L$  eingebaut.

### Definition 2.3.3 Maximum-Likelihood-Schätzer

Es sei  $f(\cdot, \vartheta)$  die kanonische Dichte von  $P_\vartheta$  und  $L: \mathbf{R}^n \times \Theta \rightarrow [0, \infty)$  wie in Definition 2.3.2 die Likelihood-Funktion. Dann heißt ein Schätzer  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  **Maximum-Likelihood-Schätzer** (MLE, engl. *maximum-likelihood estimator*) genau dann, wenn für alle  $x \in \mathbf{R}^n$  und alle  $\vartheta \in \Theta$  gilt:

$$L(x, \vartheta) \leq L(x, \hat{\Theta}(x)) \tag{MLE 1}$$

Der Maximum-Likelihood-Schätzer nimmt für jede Beobachtung also einen Parameter  $\hat{\Theta}(x)$  an, für welchen der Wert der Dichte in  $x$  maximal wird. Die Idee ist nun: Je größer die Dichte

in  $x$  ist, desto wahrscheinlicher ist  $x$ . In diesem Sinne wählt der MLE den Parameter aus, für welchen die Beobachtung von  $x$  am „wahrscheinlichsten“ ist.

**Bemerkungen:**

- Im Allgemeinen ist der Maximum-Likelihood-Schätzer nicht eindeutig.
- Ferner existiert im Allgemeinen auch nicht unbedingt ein Maximum-Likelihood-Schätzer, zum Beispiel wenn  $L$  stetig und  $\Theta$  offen ist, sowie das Maximum auf dem Rand von  $\Theta$  angenommen wird.
- Das Optimierungsproblem (MLE 1) ist manchmal analytisch lösbar, wofür wir gleich ein Beispiel sehen werden. Häufig ist dies aber auch nicht der Fall. Man kann dann auf numerische Optimierung zurückgreifen, die zum Teil jedoch sehr aufwendig werden kann.
- Zur Vereinfachung von (MLE 1) wird häufig der Logarithmus von  $L$  betrachtet:

$$\log L(x, \vartheta) = \log \prod_{i=1}^n f(x_i, \vartheta) = \sum_{i=1}^n \log f(x_i, \vartheta)$$

Man sieht dann, dass  $\hat{\Theta}$  ein Maximum-Likelihood-Schätzer ist genau dann, wenn für alle  $x \in \mathbf{R}^n$  und  $\vartheta \in \Theta$  die folgende Optimierungsbedingung gilt:

$$\log L(x, \vartheta) \leq \log L(x, \hat{\Theta}(x)) \tag{MLE 2}$$

Wie man sieht wurde aus dem Produkt in (MLE 1) dabei eine Summe in (MLE 2). Dies ist sowohl analytisch als auch numerisch häufig einfacher handhabbar. Wir nennen den Logarithmus  $\vartheta \mapsto \log L(x, \vartheta)$  der Likelihood-Funktion auch die **log-Likelihood-Funktion** der Stichprobe  $x \in \mathbf{R}^n$ .

**Beispiel** *Glühbirnen*

Es gibt drei Hersteller von Glühbirnen, die jeweils einen Ausschussanteil von  $\vartheta_1 := 0.09$ ,  $\vartheta_2 := 0.11$  und  $\vartheta := 0.12$  haben. Wir erhalten eine Lieferung von zehn Glühbirnen, von denen jedoch eine defekt ist. Mit Hilfe des Maximum-Likelihood-Schätzers wollen wir nun schätzen, von welchem Hersteller die Lieferung stammt.

Dazu betrachten wir die  $i$ -te Projektion  $\pi_i : \{0, 1\}^{10} \rightarrow \{0, 1\}$  und die  $B(1, \vartheta)^{10}$ -Verteilung auf  $\{0, 1\}^{10}$ . Unsere Beobachtung ist dann  $\sum_{i=1}^n \pi_i$ , wobei dies eine  $B(10, \vartheta)$ -verteilte Zufallsvariable ist und daher folgende Dichte besitzt:

$$f(k, \vartheta) = \binom{10}{k} \vartheta^k (1 - \vartheta)^{10-k} \quad \text{mit } k \in \{0, \dots, 10\}$$

In unserem Fall haben wir  $n = 1$  Beobachtung, welche den Wert  $k = 1$  annimmt. Wir berechnen für die  $\vartheta_i$  also jeweils die zugehörige Dichte und erhalten folgende Tabelle:

$\theta$	$f(1, \theta)$
0.09	0.3851
0.11	<b>0.3854</b>
0.12	0.3798

Das Maximum wird für  $\vartheta = \vartheta_2 = 0.11$  angenommen, dies ist also unser Schätzer und wir gehen davon aus, dass der zweite Hersteller geliefert hat. //

**Beispiel Binomialverteilung I**

Es sei  $\Theta = [0, 1]$  und  $P_\vartheta = B(1, \vartheta)$  für  $\vartheta \in \Theta$ . Damit haben wir für  $\vartheta \in \Theta$  folgende Dichte gegeben:

$$f(x, \vartheta) = \begin{cases} \prod_{i=1}^n \vartheta^{x_i} (1 - \vartheta)^{1-x_i} & x \in \{0, 1\}^n \\ 0 & \text{sonst} \end{cases}$$

Wir betrachten nun die einzelnen Fälle:

- i) Ist  $x = 0$ , so nimmt  $L(x, \cdot)$  das Maximum bei  $\vartheta = 0$  an.
- ii) Für  $x = (1, \dots, 1)$  nimmt  $L(x, \cdot)$  das Maximum bei  $\vartheta = 1$  an.
- iii) Für  $x \notin \{0, 1\}^n$  erzielt jedes  $\vartheta$  das Maximum.
- iv) Ist hingegen  $x \in \{0, 1\}^n$ , aber sowohl  $x \neq 0$ , als auch  $x \neq (1, \dots, 1)$ , so gilt  $0 < \sum_{i=1}^n x_i < n$ .  
Damit ist

$$\vartheta \mapsto \log L(x, \vartheta) = \left( \sum_{i=1}^n x_i \right) \log \vartheta + \left( n - \sum_{i=1}^n x_i \right) \log(1 - \vartheta)$$

eine stetige Funktion, für welche gilt:

$$\lim_{\vartheta \rightarrow 0} \log L(x, \vartheta) = \lim_{\vartheta \rightarrow 1} \log L(x, \vartheta) = -\infty$$

Damit ist klar, dass  $\log L(x, \cdot)$  ein Maximum besitzt. Ferner gilt nun

$$\frac{\partial}{\partial \vartheta} \log L(x, \vartheta) = \left( \sum_{i=1}^n x_i \right) \frac{1}{\vartheta} - \left( n - \sum_{i=1}^n x_i \right) \frac{1}{1 - \vartheta} \stackrel{!}{=} 0. \quad (*)$$

Eine Gleichung der Form  $\frac{a}{t} - \frac{b}{1-t} = 0$  ist genau für  $t = \frac{a}{a+b}$  erfüllt. Mit  $a = \sum_{i=1}^n x_i$  und  $b = n - \sum_{i=1}^n x_i$  folgt, dass (\*) die eindeutige Lösung  $\vartheta = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}_n$  besitzt. Damit ist  $\bar{X}_n$  der Maximum-Likelihood-Schätzer. //

**Beispiel Binomialverteilung II**

Es sei  $\Theta = [0, 1]$  und  $P_\vartheta = B(m, \vartheta)$  für  $\vartheta \in \Theta$  und ein festes, bekanntes  $m \geq 1$ . Die Dichte ist dann gegeben durch

$$f(x, \vartheta) = \begin{cases} \binom{m}{x} \vartheta^x (1 - \vartheta)^{m-x} & x \in \{0, \dots, m\} \\ 0 & \text{sonst} \end{cases}.$$

Wir betrachten nun die log-Likelihood-Funktion, die wie folgt gegeben ist:

$$\log L(x, \vartheta) = \left( \sum_{i=1}^n x_i \right) \log \vartheta + \left( nm - \sum_{i=1}^n x_i \right) \log(1 - \vartheta) + \log \left( \prod_{i=1}^n \binom{m}{x_i} \right)$$

Die Rechnung verläuft analog zum vorausgehenden Beispiel und liefert den Maximum-Likelihood-Schätzer  $\frac{1}{m} \bar{X}_n$ . //

**Beispiel** Normalverteilung

Es sei  $\Theta = \mathbf{R} \times (0, \infty)$  und  $P_\vartheta = \mathcal{N}(\mu, \sigma^2)$  für  $\vartheta = (\mu, \sigma^2)$ . Für  $\vartheta \in \Theta$  ist die Dichte dann für  $x \in \mathbf{R}^n$  wie folgt gegeben:

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right)$$

Wieder betrachten wir hier die log-Likelihood-Funktion

$$\log L(x, \mu, \sigma^2) = \log \left( \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \prod_{i=1}^n \exp\left(-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}\right) \right) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Differenzieren nach  $\mu$  liefert für die beiden ersten Ableitungen

$$\frac{\partial}{\partial \mu} \log L(x, \mu, \sigma^2) = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = \frac{n(\bar{x}_n - \mu)}{\sigma^2} \stackrel{!}{=} 0,$$

sowie

$$\frac{\partial^2}{\partial \mu^2} \log L(x, \mu, \sigma^2) = -\frac{n}{\sigma^2} < 0.$$

Damit ist für jedes  $\sigma^2 > 0$  und  $x \in \mathbf{R}^n$  die Funktion  $\mu \mapsto \log L(x, \mu, \sigma^2)$  bei  $\mu = \bar{x}_n$  maximal, das heißt die  $\mu$ -Komponente des Maximum-Likelihood-Schätzers ist  $\bar{X}_n$ . Wir betrachten nun die Abbildung  $\sigma^2 \mapsto \log L(x, \bar{x}_n, \sigma^2)$ , für die gilt:

$$\lim_{\sigma^2 \rightarrow 0} \log L(x, \bar{x}_n, \sigma^2) = \lim_{\sigma^2 \rightarrow \infty} \log L(x, \bar{x}_n, \sigma^2) = -\infty$$

Da auch diese Abbildung stetig ist muss sie ein Maximum annehmen. Differenzieren nach der Varianz liefert

$$\frac{\partial}{\partial \sigma^2} \log L(x, \bar{x}_n, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \stackrel{!}{=} 0.$$

Löst man diese Gleichung nach der Varianz, so erhält man  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = \frac{n-1}{n} s_n^2$ . Damit erhalten wir also schließlich den Maximum-Likelihood-Schätzer  $\left(\bar{X}_n, \frac{n-1}{n} S_n^2\right)$ . //

**Beispiel** Gleichverteilung

Es sei  $\Theta = (0, \infty)$  und  $P_\vartheta$  die Gleichverteilung auf  $[0, \vartheta]$ . Für  $\vartheta \in \Theta$  ist die Dichte dann wie folgt gegeben:

$$f(x, \vartheta) = \frac{1}{\vartheta} \cdot \mathbf{1}_{[0, \vartheta]}(x)$$

Die Likelihood-Funktion ist somit  $L(x, \vartheta) = \vartheta^{-n} \cdot \mathbf{1}_{[0, \vartheta]^n}(x)$ . Für  $\vartheta \geq \max_i x_i$  gilt dann  $L(x, \vartheta) = \vartheta^{-n}$ . Diese Funktion ist für steigendes  $\vartheta$  monoton fallend. Für  $\vartheta < \max_i x_i$  gilt hingegen  $L(x, \vartheta) = 0$ . Damit liegt das Maximum bei  $\vartheta = \max_i x_i$  und wir erhalten  $\max_i X_i$  als Maximum-Likelihood-Schätzer. //



**Satz 2.3.4**

Gilt  $\log f(\cdot, \vartheta) \in \mathcal{L}_1(P_{\vartheta'})$  für alle  $\vartheta, \vartheta' \in \Theta$  und  $P_{\vartheta} \neq P_{\vartheta'}$  für  $\vartheta \neq \vartheta'$ , so gilt für alle  $\vartheta, \vartheta' \in \Theta$  mit  $\vartheta \neq \vartheta'$  folgende Abschätzung:

$$\mathbf{E}_{\vartheta} \log f(\cdot, \vartheta) > \mathbf{E}_{\vartheta} \log f(\cdot, \vartheta')$$

**Beweis:** Mit den Logarithmus-Rechenregeln erhalten wir

$$\mathbf{E}_{\vartheta} \log f(\cdot, \vartheta') - \mathbf{E}_{\vartheta} \log f(\cdot, \vartheta) = \mathbf{E}_{\vartheta} \log \frac{f(\cdot, \vartheta')}{f(\cdot, \vartheta)}.$$

Es gilt  $\log t < t - 1$  für  $t \neq 1$ . Wir betrachten nun  $P_{\vartheta}(f(\cdot, \vartheta) = f(\cdot, \vartheta'))$ . Wäre dies 1, so wären die Dichten fast-überall gleich, was unseren Voraussetzungen jedoch widerspräche. Es gibt also eine nicht-Nullmenge, auf denen sie echt verschieden sind und wir erhalten

$$\begin{aligned} &< \mathbf{E}_{\vartheta} \left( \frac{f(\cdot, \vartheta')}{f(\cdot, \vartheta)} - 1 \right) = \int \left( \frac{f(\cdot, \vartheta')}{f(\cdot, \vartheta)} - 1 \right) f(\cdot, \vartheta) \, d\mu \\ &= \int f(\cdot, \vartheta') - f(\cdot, \vartheta) \, d\mu \\ &= 1 - 1 \\ &= 0. \end{aligned}$$

□

**Satz 2.3.5**

Es sei  $\Theta \subset \mathbf{R}^m$  kompakt und die kanonischen Dichten  $f(\cdot, \vartheta)$  erfüllen die folgenden Bedingungen:

- i) Es gilt  $\log f(\cdot, \vartheta) \in \mathcal{L}_1(P_{\vartheta'})$  für alle  $\vartheta, \vartheta' \in \Theta$ .
- ii) Die Funktionen  $\log f(x, \cdot): \Theta \rightarrow \mathbf{R}$  sind für  $x \in \mathbf{R}$  gleichmäßig gleichgradig stetig, das heißt für alle  $\varepsilon > 0$  existiert ein  $\delta > 0$ , so dass für alle  $\vartheta, \vartheta' \in \Theta$  mit  $\|\vartheta - \vartheta'\| < \delta$  und alle  $x \in \mathbf{R}$  gilt:  $|\log f(x, \vartheta) - \log f(x, \vartheta')| < \varepsilon$ .

Dann gilt für jeden Maximum-Likelihood-Schätzer  $\hat{\Theta}_n$

$$\hat{\Theta}_n \longrightarrow \vartheta \quad P_{\vartheta}^{\infty}\text{-fast sicher für alle } \vartheta \in \Theta.$$

Mit anderen Worten bedeutet dies, dass der Maximum-Likelihood-Schätzer konsistent ist. Satz 2.3.4 besagte also, dass  $\vartheta$  das MLE-Optimierungsproblem für „ $n = \infty$ “ löst, während Satz 2.3.5 uns nun sagt, dass die Lösung für „ $n < \infty$ “ gegen die Lösung für „ $n = \infty$ “ konvergiert.

**Beweis:** Wir fixieren  $\vartheta \in \Theta$  und  $x \in \mathbf{R}^\infty$ . Nun definieren wir die folgenden beiden Abbildungen:

$$\begin{aligned} F_{x,n} : \Theta \rightarrow \mathbf{R} & \quad \text{mit} & \quad \vartheta' \mapsto \frac{1}{n} \sum_{i=1}^n \log f(x_i, \vartheta') \\ F : \Theta \rightarrow \mathbf{R} & \quad \text{mit} & \quad \vartheta' \mapsto \mathbf{E}_\vartheta \log f(\cdot, \vartheta') \end{aligned}$$

Wir wollen zunächst zeigen, dass  $F$  und  $F_{x,n}$  gleichmäßig stetig sind. Sei dazu  $\varepsilon > 0$  und  $\delta > 0$  gemäß ii). Dann folgt für alle  $\vartheta', \vartheta'' \in \Theta$  mit  $\|\vartheta' - \vartheta''\| < \delta$

$$|F(\vartheta') - F(\vartheta'')| \leq \mathbf{E}_\vartheta |\log f(\cdot, \vartheta') - \log f(\cdot, \vartheta'')| \leq \varepsilon.$$

Man kann dies analog für  $F_{x,n}$  zeigen. Nun wollen wir die eigentliche Behauptung beweisen. Sei also  $\delta > 0$  gegeben. Da  $\Theta$  kompakt ist existiert ein endliches  $N \subset \Theta$  mit  $\Theta \subset \bigcup_{\vartheta \in N} B(\vartheta, \delta)$ . Dies bedeutet, dass für alle  $\vartheta' \in \Theta$  ein  $\vartheta'' \in N$  mit  $\|\vartheta' - \vartheta''\| < \delta$  existiert. Daraus folgt mit dem ersten Teil, dass  $|F(\vartheta') - F(\vartheta'')| \leq \varepsilon$  und analog  $|F_{x,n}(\vartheta') - F_{x,n}(\vartheta'')| \leq \varepsilon$  gilt. Ferner sichert das starke Gesetz der großen Zahlen (SLLN) für alle  $\vartheta'' \in N$

$$F_{x,n}(\vartheta'') = \frac{1}{n} \sum_{i=1}^n \log f(x_i, \vartheta'').$$

Damit folgt  $P_\vartheta^\infty$ -fast sicher  $\mathbf{E}_\vartheta \log f(\cdot, \vartheta'') = F(\vartheta'')$ . Sei nun  $\vartheta' \in \Theta$  und  $\vartheta'' \in N$  mit  $\|\vartheta' - \vartheta''\| < \delta$ , dann gilt

$$|F_{x,n}(\vartheta') - F(\vartheta')| \leq \underbrace{|F_{x,n}(\vartheta') - F_{x,n}(\vartheta'')|}_{\leq \varepsilon} + |F_{x,n}(\vartheta'') - F(\vartheta'')| + \underbrace{|F(\vartheta'') - F(\vartheta')|}_{\leq \varepsilon},$$

also gilt

$$|F_{x,n}(\vartheta') - F(\vartheta')| \leq 2\varepsilon + |F_{x,n}(\vartheta'') - F(\vartheta'')|.$$

Sei nun  $x \in \mathbf{R}^\infty$  mit  $|F_{x,n}(\vartheta'') - F(\vartheta'')| \rightarrow 0$ . Wir wissen bereits, dass dies  $P_\vartheta^\infty$ -fast alle sind. Für hinreichend große  $n$  und alle  $\vartheta' \in \Theta$  folgt damit

$$|F_{x,n}(\vartheta') - F(\vartheta')| \leq 3\varepsilon.$$

Mit Satz 2.3.4 und dieser Abschätzung folgt dann

$$F(\vartheta) \geq F(\hat{\Theta}(x)) \geq F_{x,n}(\hat{\Theta}(x)) - 3\varepsilon \stackrel{\text{MLE}}{\geq} F_{x,n}(\vartheta) - 3\varepsilon \geq F(\vartheta) - 6\varepsilon.$$

Das heißt für hinreichend große  $n$  gilt  $|F(\vartheta) - F(\hat{\Theta}(x))| \leq 6\varepsilon$ . Für  $\vartheta_n := \hat{\Theta}(x)$  gilt daher  $F(\vartheta_n) \rightarrow F(\vartheta)$ . Wir müssen nun noch zeigen, dass  $\vartheta_n \rightarrow \vartheta$  gilt. Dazu nehmen wir an, dass dies nicht der Fall ist. Dann existiert ein  $\rho > 0$  und eine Teilfolge  $\vartheta_{n_k}$  mit  $|\vartheta_{n_k} - \vartheta| > \rho$ . Da  $\Theta$  kompakt ist, ist  $\Theta$  insbesondere auch folgenkompakt. Es existiert also ein  $\bar{\vartheta} \in \Theta$  und eine Teilfolge  $\vartheta_{n_{k_l}}$  mit  $\vartheta_{n_{k_l}} \rightarrow \bar{\vartheta}$ . Damit gilt  $F(\bar{\vartheta}) \leftarrow F(\vartheta_{n_{k_l}}) \leftarrow F(\vartheta)$ . Also muss  $F(\vartheta) = F(\bar{\vartheta})$  gelten. Mit Satz 2.3.4 folgt  $\vartheta = \bar{\vartheta}$ , was einen Widerspruch zu  $|\vartheta_{n_{k_l}} - \vartheta| > \rho$  darstellt. Damit ist der Satz bewiesen.  $\square$

**Bemerkungen:**

- i) Der Beweis von Satz 2.3.5 benutzt eine gleichmäßige Version des starken Gesetzes der großen Zahlen (SLLN). Diese wurde durch die Kompaktheit von  $\Theta$  und die gleichmäßig gleichgradige Stetigkeit der log-Dichten erreicht. Weder die Kompaktheit noch die gleichmäßig gleichgradige Stetigkeit sind für Satz 2.3.5 jedoch notwendig.
- ii) Die Annahme  $\Theta \subset \mathbf{R}^m$  ist nicht notwendig. Satz 2.3.5 hätte ohne Modifikationen auch für beliebige kompakte metrische Räume  $\Theta$  bewiesen werden können. Der Maximum-Likelihood-Schätzer ist also auch für nicht-klassische parametrische Annahmen konsistent.
- iii) Die Voraussetzung, dass  $P_\theta$  ein Maß auf  $\mathbf{R}$  ist, wurde ebenfalls nicht verwendet, es kann also auch hier verallgemeinert werden.

## 2.4 Bayes-Schätzer

Bis jetzt haben wir bei der Bewertung von Schätzern verlangt, dass gute Schätzer gleichmäßig optimal sind (engl. *worst-case scenario*). Wir wollen dies ändern, indem wir annehmen, dass bestimmte  $\vartheta \in \Theta$  in einem gewissen Sinne „wahrscheinlicher“ oder „wichtiger“ sind.

Es sei  $\Theta \subset \mathbf{R}^m$  und  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$ . Ferner haben  $P_\vartheta$  wie bei den Maximum-Likelihood-Schätzern eine kanonische Dichte bezüglich eines Maßes  $\mu$  auf  $\mathbf{R}$ . Zudem gebe es einen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \nu)$  und eine messbare Abbildung  $\tilde{\Theta}: \Omega \rightarrow \Theta$ , deren Verteilung das Bildmaß  $Q = \tilde{\Theta}_\nu$  sei. Wir nennen  $Q$  die **a-priori-Verteilung** von  $\vartheta$ . Weiter nehmen wir an, dass  $Q$  eine kanonische Dichte  $q$  bezüglich eines Maßes  $\tilde{\mu}$  auf  $\Theta$  besitzt.

Die Idee ist nun, mit zufälligen Parametern  $\vartheta = \tilde{\Theta}(\omega)$  zu arbeiten. Die Verteilung  $Q$  beschreibt die Wahrscheinlichkeit der Parameter  $\vartheta$  *bevor* wir die Daten gesehen haben. Daher modelliert  $Q$  das Vorwissen über die Wahrscheinlichkeit von  $\vartheta \in \Theta$ .

Im Folgenden bezeichne  $L$ , sofern nicht anders angegeben, stets die Likelihood-Funktion vermöge  $L(x, \vartheta) := \prod_{i=1}^n f(x_i, \vartheta)$  für alle  $x \in \mathbf{R}^n$  und  $\vartheta \in \Theta$ . Wir wollen annehmen, dass  $L: \mathbf{R}^n \times \Theta \rightarrow [0, \infty)$  messbar ist.

### Definition 2.4.1 A-posteriori-Wahrscheinlichkeitsdichte

Für  $x \in \mathbf{R}^n$  schreiben wir  $f(x) := \int L(x, \vartheta)q(\vartheta) d\tilde{\mu}(\vartheta)$ . Ist  $f(x) > 0$ , so nennen wir  $q_x: \Theta \rightarrow [0, \infty)$  vermöge  $q_x(\vartheta) := \frac{L(x, \vartheta)q(\vartheta)}{f(x)}$  die **a-posteriori-Wahrscheinlichkeitsdichte** von  $\Theta$  beim Vorliegen der Beobachtung  $x$ .

Falls  $f(x) = 0$  ist, so kann man beispielsweise  $q_x(\vartheta) := 0$  setzen, dieser Fall tritt jedoch fast-sicher nicht ein.

Die Abbildung  $q_x$  ist tatsächlich eine Dichte, wie man durch  $\int q_x(\vartheta) d\tilde{\mu} = \frac{f(x)}{f(x)} = 1$  leicht nachrechnen kann. Das Produkt  $L(x, \vartheta)q(\vartheta)$  beschreibt die infinitesimale Wahrscheinlichkeit für die Beobachtung  $(X, \vartheta)$  und die Abbildung  $f(x)$  ist die Randdichte von  $x$  bezüglich der gemeinsamen Verteilung von  $(x, \tilde{\Theta})$ , sie beschreibt also gewissermaßen die Wahrscheinlichkeit von  $x$ . Damit können wir  $q_x$  so interpretieren, dass diese Abbildung die Wahrscheinlichkeit von „ $\vartheta$  unter der Bedingung  $x$ “ beschreibt.

Sind  $\mu$  und  $\tilde{\mu}$  Zählmaße, so stimmt diese Interpretation mit dem elementaren Begriff der bedingten Wahrscheinlichkeit überein, im anderen Fall stimmt dies zwar prinzipiell auch, benötigt jedoch einige technische Anstrengungen.

**Definition 2.4.2 Bayes-Schätzer**

Es sei  $\theta \in \mathbf{R}$ , das heißt  $m = 1$ . Falls alle auftretenden Integrale existieren, nennen wir die Abbildung  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$ , die durch

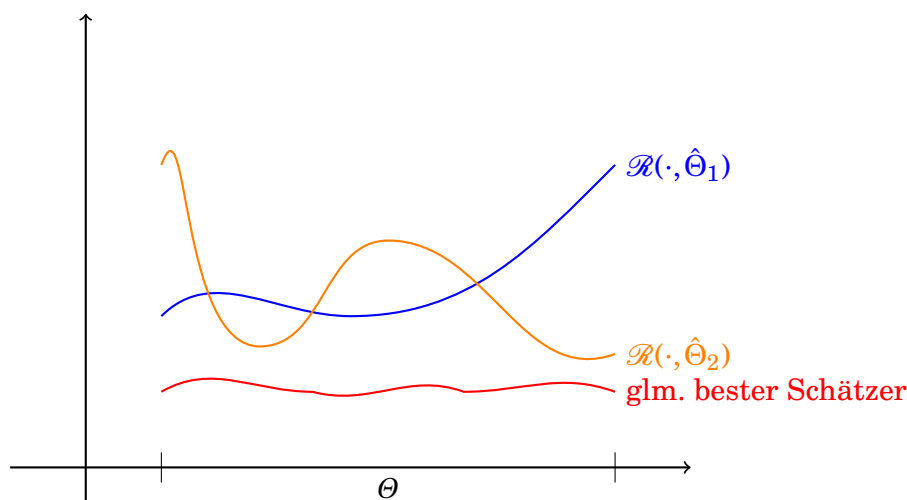
$$\hat{\Theta}(x) := \int_{\theta} \vartheta q_x(\vartheta) d\tilde{\mu}(\vartheta)$$

gegeben ist und gewissermaßen den „Erwartungswert des Maßes  $q_x d\tilde{\mu}$ “ beschreibt, einen Bayes-Schätzer.

Für höhere Dimensionen  $m > 1$  kann man den Begriff analog komponentenweise definieren.

Da  $q_x d\tilde{\mu}$  die Wahrscheinlichkeit von  $\vartheta$  unter der Bedingung  $x$  beschreibt, gibt der Bayes-Schätzer das mittlere  $\vartheta$  bezüglich dieser Verteilung zurück.

Wir können nun die Frage stellen, in welchem Sinne der Bayes-Schätzer optimal ist. Dazu wollen wir unsere Begriffe für das Risiko und die Optimalität von Schätzern verbessern. Bisher vergleichen wir die Schätzer punktweise, Abbildung 2.1 illustriert aber, dass es durchaus passieren kann, dass ein Schätzer an manchen Punkt besser, an anderen Punkten schlechter als ein anderer Schätzer ist.



**Abbildung 2.1:** Schätzer können punktweise besser und schlecht als andere Schätzer sein, gleichmäßig beste Schätzer sind hingegen überall besser. Aufgetragen sind hier die Risikofunktionen.

Die Idee ist daher nun, dass wir die Schätzer nicht mehr punktweise mit Hilfe der Risikofunktionen vergleichen, sondern die Fläche unter eben diesen betrachten. Damit können wir zwei Schätzer *immer* vergleichen.

**Definition 2.4.3 Gemittelter quadratischer Verlust**

Sei  $\Theta \subset \mathbf{R}$  und  $\hat{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R}$  ein quadrat-integrierbarer Schätzer. Dann heißt die Abbildung

$$\mathcal{R}(\hat{\Theta}) := \int_{\Theta} \mathcal{R}(\vartheta, \hat{\Theta}) q(\vartheta) d\tilde{\mu}(\vartheta) = \int_{\Theta} \int_{\mathbf{R}^n} (\vartheta - \hat{\Theta}(x))^2 dP_{\vartheta}^n(x) q(\vartheta) d\tilde{\mu}(\vartheta)$$

der über  $Q = q d\tilde{\mu}$  gemittelte quadratische Verlust des Schätzers  $\hat{\Theta}$ .

Wir wollen nun zeigen, dass der Bayes-Schätzer bezüglich dieses neuen Risikobegriffs optimal ist.

**Satz 2.4.4**

Sei  $\Theta \subset \mathbf{R}$  und  $\int_{\Theta} \vartheta^2 dQ(\vartheta) < \infty$ . Ferner sei der Bayes-Schätzer  $\hat{\Theta}$  quadrat-integrierbar. Dann ist  $\hat{\Theta}$  der *einzigste* quadrat-integrierbare Schätzer, der den über  $Q$  gemittelten quadratischen Verlust minimiert, das heißt für alle quadrat-integrierbaren Schätzer  $\hat{\Theta}'$  gilt  $\mathcal{R}(\hat{\Theta}) \leq \mathcal{R}(\hat{\Theta}')$ , wobei Gleichheit genau dann herrscht, wenn  $\mu^n$ -fast sicher für alle  $x \in \mathbf{R}^n$  mit  $f(x) > 0$  auch  $\hat{\Theta}(x) = \hat{\Theta}'$  gilt.

**Beweis:** Wir betrachten zunächst das Risiko

$$\begin{aligned} \mathcal{R}(\hat{\Theta}') &= \int_{\Theta} \int_{\mathbf{R}^n} (\vartheta - \hat{\Theta}'(x))^2 dP_{\vartheta}^n(x) q(\vartheta) d\tilde{\mu}(\vartheta) \\ &= \int_{\Theta} \int_{\mathbf{R}^n} (\vartheta - \hat{\Theta}'(x))^2 L(x, \vartheta) d\mu^n q(\vartheta) d\tilde{\mu}(\vartheta) \\ &\stackrel{\text{Tonelli}}{=} \int_{\mathbf{R}^n} \underbrace{\int_{\Theta} (\vartheta - \hat{\Theta}'(x))^2 L(x, \vartheta) q(\vartheta) d\tilde{\mu}(\vartheta)}_{h_x(t)} d\mu^n(x), \end{aligned}$$

wobei wir  $t := \hat{\Theta}'(x)$  setzen. Nun wollen wir  $h_x(\cdot)$  für jedes  $x$  minimieren und benötigen dafür die entsprechenden Ableitungen. Im Wesentlichen haben wir dies jedoch bereits im Beweis von Satz 2.2.2 erledigt und es gilt

$$\begin{aligned} \frac{d}{dt} h_x(t) &= \frac{d}{dt} \int_{\Theta} (\vartheta - t)^2 L(x, \vartheta) q(\vartheta) d\tilde{\mu}(\vartheta) = \int_{\Theta} (2t - 2\vartheta) L(x, \vartheta) q(\vartheta) d\tilde{\mu}(\vartheta) \\ &\stackrel{!}{=} 0. \end{aligned}$$

Daraus erhalten wir nun folgende Gleichung:

$$t \underbrace{\int_{\Theta} L(x, \vartheta) q(\vartheta) d\tilde{\mu}(\vartheta)}_{f(x)} = \int_{\Theta} \vartheta L(x, \vartheta) q(\vartheta) d\tilde{\mu}(\vartheta)$$

Für  $f(x) > 0$  können wir durch diesen Ausdruck dividieren und erhalten damit

$$t = \int_{\Theta} \vartheta \underbrace{\frac{L(x, \vartheta)q(\vartheta)}{f(x)}}_{q_x(\vartheta)} d\tilde{\mu}(\vartheta) = \int_{\Theta} \vartheta q_x(\vartheta) d\tilde{\mu}(\vartheta) = \hat{\Theta}(x).$$

Für die zweite Ableitung erhalten wir zudem

$$\frac{d^2}{dt^2} h_x(t) = 2 \int_{\Theta} L(x, \vartheta)q(\vartheta) d\tilde{\mu}(\vartheta) = 2 \cdot f(x) > 0.$$

Damit besitzt  $h_x(\cdot)$  also genau ein globales Maximum, das bei  $\hat{\Theta}(x)$  angenommen wird.  $\square$



Normalerweise wird  $Q$  als Vorwissen über mögliche  $\vartheta$  interpretiert. Typischerweise ist ein Kritikpunkt am Bayes-Schätzer-Ansatz die Frage, woher wir ein derart präzies Vorwissen haben. Man kann hierauf z. B. antworten, dass die Wahl von  $Q$  den Schätzer wenig beeinflusst, dies stimmt jedoch nur zum Teil. Da diesbezüglich ein langer Streit zwischen den Verfechtern existiert, wird in der Fachwelt auch von *Bayesians* und *Frequentists* gesprochen.

In einer alternativen Betrachtungsweise beschreibt  $Q$  unsere Gewichtung, das heißt an welchen Stellen wir gutes Verhalten eher sehen wollen und Stellen, an denen uns gutes Verhalten weniger interessiert. Diese Betrachtungsweise gibt Freiheit bei der Wahl von  $Q$  und appelliert nicht an ein etwaiges Vorwissen. Eine Möglichkeit wäre es so zum Beispiel für  $Q$  die Gleichverteilung auf  $\Theta$  zu wählen, das heißt wir erachten das Verhalten unseres Schätzers für überall gleich wichtig (engl. *uninformed prior*).

### Beispiel Binomialverteilung

Es sei  $P_{\vartheta} := B(m, \vartheta)$  auf  $\mathbf{R}$  mit  $\vartheta \in [0, 1]$  und  $m \geq 1$ . Dann existiert eine kanonische Dichte vermöge

$$f(k, \vartheta) = \begin{cases} \binom{m}{k} \vartheta^k (1 - \vartheta)^{m-k} & k = 0, \dots, m \\ 0 & \text{sonst} \end{cases}.$$

Dies ist die kanonische Dichte bezüglich des Zählmaßes auf  $\mathbf{N}_0$ . Nun sei  $\tilde{\mu}$  die Gleichverteilung auf  $[0, 1]$ , dann ist  $q(\vartheta) := 1$  für alle  $\vartheta \in [0, 1]$  die kanonische Dichte bezüglich des Lebesguemaßes auf  $[0, 1]$ . Ferner sei  $n = 1$ , dann entspricht unsere Situation der Summenbildung nach  $m$ -maligen Werfen einer unfairen Münze und wir erhalten die Likelihood-Funktion  $L(k, \vartheta) = f(k, \vartheta)$ , sowie dementsprechend  $L(k, \vartheta)q(\vartheta) = f(k, \vartheta)$ . Für  $k = 0, \dots, m$  gilt dann

$$\begin{aligned} f(k) &= \int_0^1 L(k, \vartheta)q(\vartheta) d\vartheta = \int_0^1 f(k, \vartheta) d\vartheta = \int_0^1 \binom{m}{k} \vartheta^k (1 - \vartheta)^{m-k} d\vartheta \\ &= \binom{m}{k} \frac{(m-k)!k!}{(m+1)!} = \frac{m!}{(m-k)!k!} \cdot \frac{(m-k)!k!}{(m+1)!} \\ &= \frac{1}{m+1}. \end{aligned}$$

Für den Bayes-Schätzer folgt damit nun

$$\begin{aligned}\hat{\Theta}(k) &= \int_0^1 \vartheta q_k(\vartheta) \, d\vartheta = \int_0^1 \vartheta \cdot \frac{L(k, \vartheta)q(\vartheta)}{f(k)} \, d\vartheta = (m+1) \int_0^1 \binom{m}{k} \vartheta^{k+1} (1-\vartheta)^{m-k} \, d\vartheta \\ &= \frac{k+1}{m+2}.\end{aligned}$$

Mit einfachen algebraischen Umformungen führt dies auch zu

$$= \frac{m}{m+2} \cdot \frac{k}{m} + \frac{1}{2} \cdot \frac{2}{m+2} = \frac{m}{m+2} \cdot \overline{x_m} + \frac{2}{m+2} \cdot \frac{1}{2}.$$

Dabei ist  $\overline{x_m}$  als arithmetisches Mittel ein gleichmäßig bester Schätzer, und da  $\frac{1}{2}$  der Erwartungswert von  $\tilde{\mu}$  ist, ist der Bayes-Schätzer gewissermaßen eine Mischung zwischen diesen beiden Größen. //

Wir wollen an dieser Stelle festhalten, dass die bis jetzt verwendeten Optimalitätsbegriffe wirklich verschieden sind:

- Nach Beispiel 2.2.8 ist  $\overline{X_n}$  gleichmäßig bester unverzerrter Schätzer für  $p \in [0, 1]$  von der Familie  $B(1, p)$ .
- Im vorausgehenden Beispiel war  $\frac{n}{n+2} \overline{X_n} + \frac{2}{n+2} \cdot \frac{1}{2}$  der Bayes-Schätzer.

Ferner ist der Bayes-Schätzer im Allgemeinen nicht erwartungstreu. Die Konsistenz des Bayes-Schätzers hängt stark von der konkreten Situation ab.



## 2.5 Cramér-Rao-Ungleichung

In diesem Kapitel wollen wir uns damit beschäftigen, die Chapman-Robbins-Ungleichung zu verbessern.

### Satz 2.5.1

Es sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Verteilungen auf  $\mathbf{R}$ , wobei  $\Theta \subset \mathbf{R}$  offen sei. Ferner habe  $P_\vartheta$  die kanonische Dichte  $f(\cdot, \vartheta)$  bezüglich des Maßes  $\mu$  und es sei  $L(x, \vartheta) := \prod_{i=1}^n f(x_i, \vartheta)$  mit  $x \in \mathbf{R}^n$  und  $\vartheta \in \Theta$  die Likelihood-Funktion, für welche folgende Eigenschaften gelten:

- i) Die Menge  $B := \{x \in \mathbf{R} : f(x, \vartheta) > 0\}$  sei unabhängig von  $\vartheta \in \Theta$ .
- ii) Die Abbildung  $f(x, \cdot) : \Theta \rightarrow [0, \infty)$  sei für alle  $x \in \mathbf{R}$  differenzierbar.
- iii) Es gelte  $\int_B \frac{\partial}{\partial \vartheta} f(x, \vartheta) \, d\mu(x) = \frac{\partial}{\partial \vartheta} \int_B f(x, \vartheta) \, d\mu(x) = 0$ .

Dann gilt für jeden quadrat-integrierbaren Schätzer  $\hat{\Theta}_n$ , für welchen die Ableitung  $\frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\Theta}_n$  existiert und der die Gleichung

$$\frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\Theta}_n = \int_{B^n} \hat{\Theta}_n(x) \frac{\partial}{\partial \vartheta} L(x, \vartheta) \, d\mu^n(x) \quad (*)$$

erfüllt, die **Cramér-Rao-Ungleichung**:

$$\text{Var}_\vartheta \hat{\Theta}_n \geq \frac{\left(\frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\Theta}_n\right)^2}{n \cdot \mathbf{E}_\vartheta \left(\frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta)\right)^2}$$

Hierbei ist zu beachten, dass  $\mathbf{E}_\vartheta \left(\frac{\partial}{\partial \vartheta} f(\cdot, \vartheta)\right)^2$  von  $n$  unabhängig ist.

**Beweis:** Wir setzen  $\phi_\vartheta(x) := \frac{\partial}{\partial \vartheta} \log L(x, \vartheta) \cdot \mathbf{1}_{B^n}(x)$ . Für  $\vartheta \in \Theta$  gilt dann

$$\begin{aligned} \text{Var}_\vartheta \phi_\vartheta &= \text{Var}_\vartheta \left( \frac{\partial}{\partial \vartheta} \log \prod_{i=1}^n f(X_i, \vartheta) \cdot \mathbf{1}_{B^n}(x) \right) = \text{Var}_\vartheta \left( \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f(X_i, \vartheta) \right) \\ &= \sum_{i=1}^n \text{Var}_\vartheta \frac{\partial}{\partial \vartheta} \log f(X_i, \vartheta) = n \cdot \text{Var}_\vartheta \frac{\partial}{\partial \vartheta} \log f(X_1, \vartheta). \end{aligned}$$

Mit der Regel  $(\log f)' = \frac{f'}{f}$  folgt nun

$$\begin{aligned} \mathbf{E}_\vartheta \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) &= \int_B \frac{\partial}{\partial \vartheta} \log f(x, \vartheta) f(x, \vartheta) \, d\mu(x) = \int_B \frac{\partial}{\partial \vartheta} f(x, \vartheta) \, d\mu(x) \\ &\stackrel{\text{iii)}}{=} 0. \end{aligned}$$

Damit erhalten wir für die Varianz

$$\begin{aligned}\text{Var}_\vartheta \phi_\vartheta &= n \cdot \text{Var}_\vartheta \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) = n \left( \mathbf{E}_\vartheta \left( \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) \right)^2 - \left( \mathbf{E}_\vartheta \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) \right)^2 \right) \\ &= n \cdot \mathbf{E}_\vartheta \left( \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) \right)^2.\end{aligned}$$

Wir verwenden nun die Cauchy-Schwartz-Ungleichung und erhalten

$$\begin{aligned}(\mathbf{E}_\vartheta(\phi_\vartheta - \mathbf{E}_\vartheta \phi_\vartheta)(\hat{\Theta}_n - \mathbf{E}_n \hat{\Theta}_n))^2 &\leq \mathbf{E}_\vartheta(\phi_\vartheta - \mathbf{E}_\vartheta \phi_\vartheta)^2 \cdot \mathbf{E}_\vartheta(\hat{\Theta}_n - \mathbf{E}_\vartheta \hat{\Theta}_n)^2 \\ &= n \cdot \mathbf{E}_\vartheta \left( \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) \right)^2 \cdot \text{Var}_\vartheta \hat{\Theta}_n.\end{aligned}\quad (**)$$

Damit folgt schließlich

$$\begin{aligned}\mathbf{E}_\vartheta(\phi_\vartheta - \mathbf{E}_\vartheta \phi_\vartheta)(\hat{\Theta}_n - \mathbf{E}_\vartheta \hat{\Theta}_n) &= \mathbf{E}_\vartheta \phi_\vartheta \hat{\Theta}_n - \mathbf{E}_\vartheta \phi_\vartheta \mathbf{E}_\vartheta \hat{\Theta}_n = \mathbf{E}_\vartheta \phi_\vartheta \hat{\Theta}_n \\ &= \int_{B^n} \frac{\partial}{\partial \vartheta} \log L(x, \vartheta) \cdot \hat{\Theta}_n(x) L(x, \vartheta) \, d\mu^n(x).\end{aligned}$$

Durch eine vollständige Induktion und mit Hilfe der Kettenregel folgt

$$\begin{aligned}&= \int_{B^n} \hat{\Theta}_n(x) \frac{\partial}{\partial \vartheta} L(x, \vartheta) \, d\mu^n(x) \\ &\stackrel{(*)}{=} \frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\Theta}_n.\end{aligned}$$

Verbindet man dies noch mit (\*\*), so erhalten wir die Aussage und sind fertig. □

### Korollar 2.5.2

Unter den Voraussetzungen des Satzes 2.5.1 gilt für jeden quadrat-integrierbaren und erwartungstreuen Schätzer, der (\*) aus diesem Satz erfüllt, die folgende Abschätzung:

$$\text{Var}_\vartheta \hat{\Theta}_n \geq \frac{1}{n \cdot \mathbf{E}_\vartheta \left( \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) \right)^2}$$

Die Risiken solcher Schätzer können also nicht schneller als  $\frac{1}{n}$  gegen 0 konvergieren.

**Beweis:** Es gilt wegen der Erwartungstreue  $\mathbf{E}_\vartheta \hat{\Theta}_n = \vartheta$ . Daraus folgt  $\frac{\partial}{\partial \vartheta} \mathbf{E}_\vartheta \hat{\Theta}_n = 1$  und wir können Satz 2.5.1 anwenden. □

Ist  $\hat{\Theta}_n$  ein erwartungstreuer Schätzer, sind alle Voraussetzung von Satz 2.5.1 beziehungsweise von Korollar 2.5.2 erfüllt und gilt zusätzlich, dass die Schranke im Korollar 2.5.2 für alle  $\vartheta \in \Theta$  angenommen wird, so folgt, dass  $\hat{\Theta}_n$  gleichmäßig bester erwartungstreuer Schätzer ist.

**Beispiel 2.5.3** Normalverteilung

Es seien  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  Zufallsvariablen mit bekanntem  $\sigma^2 > 0$ , aber unbekanntem  $\mu \in \mathbf{R}$ . Dann ist  $f(x, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$  und wir erhalten für die Ableitung  $\frac{\partial}{\partial \mu} f(x, \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \frac{x-\mu}{\sigma^2} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ . Nun gilt  $B = \mathbf{R}$  und wenn  $\mu$  das Lebesguemaß ist, so erhalten wir

$$\int \frac{\partial}{\partial \mu} f(x, \mu) dx = \frac{1}{\sigma^2} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx}_{\mathbf{E}_\mu X_1 = \mu} - \frac{\mu}{\sigma^2} = \frac{\mu}{\sigma^2} - \frac{\mu}{\sigma^2} = 0.$$

Damit sind die Voraussetzungen i)–iii) von Satz 2.5.1 erfüllt. Ferner betrachten wir  $\overline{X}_n$  als Schätzer. Klar ist, dass  $\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$  gilt. Damit ist  $\overline{X}_n$  quadrat-integrierbar und erwartungstreu und wir erhalten  $\frac{\partial}{\partial \mu} \mathbf{E}_\mu \overline{X}_n = 1$ . Damit fehlt noch die Gleichung (\*) aus Satz 2.5.1. Um diese zu zeigen führen wir eine Induktion über  $n$  durch. Für den Induktionsanfang mit  $n = 1$  gilt

$$\begin{aligned} \int_{\mathbf{R}} x \frac{\partial}{\partial \mu} f(x, \mu) dx &= \int_{\mathbf{R}} x \cdot \frac{x-\mu}{\sigma^2} f(x, \mu) dx \\ &= \frac{1}{\sigma^2} \left( \int_{\mathbf{R}} x^2 f(x, \mu) dx - \mu \int_{\mathbf{R}} x f(x, \mu) dx \right) \\ &= \frac{1}{\sigma^2} \text{Var}_\mu X_1 \\ &= 1. \end{aligned}$$

Damit kommen wir zum Induktionsschritt. Die Induktionsvoraussetzung lautet

$$\int_{\mathbf{R}^n} \left( \sum_{i=1}^n x_i \right) \frac{\partial}{\partial \mu} L_n(x, \mu) d^n x = n. \tag{IV}$$

Wir wollen zunächst die Ableitung im Integranden näher untersuchen. Es gilt

$$\begin{aligned} \frac{\partial}{\partial \mu} L_{n+1}(x, \mu) &= \frac{\partial}{\partial \mu} \prod_{i=1}^{n+1} f(x_i, \mu) = \frac{\partial}{\partial \mu} (L_n(x, \mu) \cdot f(x_{n+1}, \mu)) \\ &= \frac{\partial}{\partial \mu} L_n(x, \mu) f(x_{n+1}, \mu) + L_n(x, \mu) \frac{\partial}{\partial \mu} f(x_{n+1}, \mu). \end{aligned}$$

Ferner gilt nun

$$\left( \sum_{i=1}^n x_i \right) \frac{\partial}{\partial \mu} L_{n+1}(x, \mu) = \left( \sum_{i=1}^n x_i \right) \left( \frac{\partial}{\partial \mu} L_n(x, \mu) f(x_{n+1}, \mu) + L_n(x, \mu) \frac{\partial}{\partial \mu} f(x_{n+1}, \mu) \right).$$

Für das Integral erhalten wir dann mit dem Satz von Fubini

$$\begin{aligned} \int_{\mathbf{R}^{n+1}} \left( \sum_{i=1}^n x_i \right) \frac{\partial}{\partial \mu} L_{n+1}(x, \mu) d^{n+1}x &= \int_{\mathbf{R}} f(x_{n+1}, \mu) \underbrace{\int_{\mathbf{R}^n} \left( \sum_{i=1}^n x_i \right) \frac{\partial}{\partial \mu} L_n(x, \mu) d^n x}_{=n} dx_{n+1} \\ &\quad + \int_{\mathbf{R}^n} \left( \sum_{i=1}^n x_i \right) L_n(x, \mu) \underbrace{\int_{\mathbf{R}} \frac{\partial}{\partial \mu} f(x_{n+1}, \mu) dx_{n+1}}_{=0} d^n x \\ &= n. \end{aligned}$$

Wir betrachten nun noch den anderen Summanden. Es gilt

$$\begin{aligned} \int_{\mathbf{R}^{n+1}} x_{n+1} \frac{\partial}{\partial \mu} L_{n+1}(x, \mu) \, d^{n+1}x &= \int_{\mathbf{R}^{n+1}} x_{n+1} \left( \frac{\partial}{\partial \mu} L_n(x, \mu) \cdot f(x_{n+1}, \mu) + L_n(x, \mu) \frac{\partial}{\partial \mu} f(x_{n+1}, \mu) \right) \, d^{n+1}x \\ &= \int_{\mathbf{R}} x_{n+1} f(x_{n+1}, \mu) \int_{\mathbf{R}^n} \frac{\partial}{\partial \mu} L_n(x, \mu) \, d^n x \, dx_{n+1} \\ &\quad + \underbrace{\int_{\mathbf{R}^n} L_n(x, \mu) \int_{\mathbf{R}} x_{n+1} \frac{\partial}{\partial \mu} f(x_{n+1}, \mu) \, dx_{n+1} \, d^n x}_{=1}. \end{aligned}$$

Wir überprüfen nun den ersten Summanden in diesem letzten Ausdruck. Für das innere Integral gilt für alle  $n \geq 1$

$$\int_{\mathbf{R}^n} \frac{\partial}{\partial \mu} L_n(x, \mu) \, d^n x = 0.$$

Dies lässt sich z. B. mittels einer vollständigen Induktion beweisen. Der Induktionsanfang entspricht hier (\*\*\*) und für den Induktionsschritt gilt skizzenhaft  $\frac{\partial}{\partial \mu} L_{n+1} = \frac{\partial}{\partial \mu} L_n f + L_n \frac{\partial}{\partial \mu} f$ . Integriert man über diesen Ausdruck, so wird  $\int \frac{\partial}{\partial \mu} L_n f = 0$  und  $\int L_n \frac{\partial}{\partial \mu} f = 0$ . Insgesamt erhalten wir damit also

$$\int_{\mathbf{R}^{n+1}} x_{n+1} \frac{\partial}{\partial \mu} L_{n+1}(x, \mu) \, d^{n+1}x = 1.$$

Damit folgt die Gleichung (\*) aus Satz 2.5.1. Wir haben nun also alle Voraussetzungen des Satzes bzw. des folgenden Korollars 2.5.2 erfüllt und erhalten die folgende Abschätzung:

$$\text{Var}_\mu \overline{X}_n \geq \frac{1}{n \cdot \mathbf{E}_\mu \left( \frac{\partial}{\partial \mu} \log f(\cdot, \mu)^2 \right)}$$

Wir wollen nun noch zeigen, dass hier sogar die Gleichheit gilt. Wir betrachten dazu zunächst

$$\frac{\partial}{\partial \mu} \log f(x, \mu) = \frac{\partial}{\partial \mu} \left( -\frac{1}{2} \log 2\pi\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2} \right) = \frac{x-\mu}{\sigma^2}.$$

Damit folgt dann

$$\begin{aligned} \mathbf{E}_\mu \left( \frac{\partial}{\partial \mu} \log f(\cdot, \mu) \right)^2 &= \int_{\mathbf{R}} \left( \frac{x-\mu}{\sigma^2} \right)^2 f(x, \mu) \, dx \\ &= \frac{1}{\sigma^4} \int_{\mathbf{R}} (x-\mu)^2 f(x, \mu) \, dx = \frac{1}{\sigma^4} \text{Var}_\mu X_1 = \frac{1}{\sigma^4} \sigma^2 \\ &= \frac{1}{\sigma^2}. \end{aligned}$$

Damit erhalten wir nun  $\text{Var}_\mu \overline{X}_n \geq \frac{\sigma^2}{n}$ . Es gilt aber auch  $\text{Var}_\mu \overline{X}_n = \frac{1}{n} \text{Var}_\mu X_1 = \frac{\sigma^2}{n}$ , daher ist die Cramér-Rao-Ungleichung exakt und damit ist  $\overline{X}_n$  gleichmäßig bester, erwartungstreuer Schätzer. //

An dieser Stelle wollen wir noch anmerken, dass die Voraussetzungen des Satzes 2.5.1 im Allgemeinen wirklich notwendig sind.

## 2.6 Weitere Eigenschaften von Schätzern

Alle Annahmen und Notationen seien wie im vorausgehenden Kapitel gegeben. Unser erstes Ziel ist es nun, die Verteilung von  $\hat{\Theta}_n - \vartheta$  für  $n \rightarrow \infty$  zu untersuchen.

### Satz 2.6.1

Es sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Verteilungen auf  $\mathbf{R}$  mit kanonischer Dichte  $f(\cdot, \vartheta)$  bezüglich  $\mu$ . Ferner sei  $\Theta \subset \mathbf{R}$  offen. Zudem gelten die folgenden Eigenschaften:

- i) Die Menge  $B := \{x \in \mathbf{R} : f(x, \vartheta) > 0\}$  ist unabhängig von  $\vartheta$ .
- ii) Für  $\vartheta \neq \vartheta'$  gilt  $P_\vartheta \neq P_{\vartheta'}$ .
- iii) Die Abbildung  $f(x, \cdot) : \Theta \rightarrow [0, \infty)$  ist dreimal stetig differenzierbar.
- iv) Für  $k \in \{1, 2\}$  gilt die Gleichung

$$\frac{\partial^k}{\partial \vartheta^k} \int_B f(x, \vartheta) \, d\mu(x) = \int_B \frac{\partial^k}{\partial \vartheta^k} f(x, \vartheta) \, d\mu(x).$$

Für  $k = 1$  soll das Integral dabei zusätzlich den Wert 0 annehmen.

- v) Für alle  $\vartheta_0 \in \Theta$  existiere ein  $c_{\vartheta_0} > 0$  und ein  $g_{\vartheta_0} : B \rightarrow [0, \infty)$  mit  $\mathbf{E}_{\vartheta_0} |g_{\vartheta_0}| < \infty$ , so dass für alle  $x \in B^n$  folgende Ungleichung gilt:

$$\left| \frac{\partial^3}{\partial \vartheta^3} \log f(x, \vartheta) \right| \leq g_{\vartheta_0}(x)$$

Wir definieren nun die **Fisher-Information**  $I(\vartheta) := \mathbf{E}_\vartheta \left( \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) \right)^2$ . Dann gilt:

- i)  $I(\vartheta) = \text{Var}_\vartheta \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta)$  für alle  $\vartheta \in \Theta$ .
- ii) Ist  $I(\vartheta) \in (0, \infty)$  für alle  $\vartheta \in \Theta$ , so folgt für jeden (schwach) konsistenten Maximum-Likelihood-Schätzer  $\hat{\Theta}_n$  und alle  $\vartheta \in \Theta$

$$\sqrt{nI(\vartheta)}(\hat{\Theta}_n - \vartheta) \longrightarrow \mathcal{N}(0, 1) \quad \text{in Verteilung.}$$

Es gilt hierbei folgende Punkte zu beachten:

- i) Die Konsistenz des Maximum-Likelihood-Schätzers wurde in Satz 2.3.4 untersucht.
- ii) Dieses Ergebnis erlaubt es uns, abzuschätzen, wie weit  $\hat{\Theta}_n$  von  $\vartheta$  mit einer gewissen Wahrscheinlichkeit entfernt ist. Die Rechnung verläuft analog zum zentralen Grenzwertsatz (CLT). Für die konkrete Anwendung müsste eigentlich die Fisher-Information  $I(\vartheta)$  bekannt sein, hierauf kann aber verzichtet werden, wenn zum Beispiel  $I(\vartheta) \in [a, b]$  bekannt ist.

iii) Als grobe Interpretation könnte man  $\text{Var}_\vartheta \sqrt{nI(\vartheta)}\hat{\Theta}_n \approx 1$  sagen, das heißt  $\text{Var}_\vartheta \hat{\Theta}_n \approx (nI(\vartheta))^{-1}$ . Dies ist gerade die Cramér-Rao-Schranke für unverzerrte Schätzer. Man spricht hier von asymptotischer Effizienz, das heißt für große  $n$  kommt der Schätzer nahe an den besten Schätzer heran.

**Beweis:** Für die erste Aussage verweisen wir an dieser Stelle auf den Beweis von Satz 2.5.1. Für die zweite Aussage des Satzes setzen wir  $l_n(\vartheta) := l_n(x, \vartheta) := \log \prod_{i=1}^n f(x_i, \vartheta)$  und  $l_n^{(i)}(\vartheta) := l_n^{(i)}(x, \vartheta)$  für die  $i$ -te Ableitung bezüglich  $\vartheta$ . Für die Taylorreihe von  $l_n^{(1)}(\hat{\Theta}_n(x))$  um  $\vartheta \in \Theta$  gilt

$$l_n^{(1)}(\hat{\Theta}_n(x)) = l_n^{(1)}(\vartheta) + (\hat{\Theta}_n(x) - \vartheta)l_n^{(2)}(\vartheta) + \frac{1}{2}(\hat{\Theta}_n(x) - \vartheta)^2 l_n^{(3)}(\vartheta^*). \quad (*)$$

Dabei liegt  $\vartheta^*$  zwischen  $\vartheta$  und  $\hat{\Theta}_n(x)$  und wird geeignet gewählt. Ferner gilt aber auch  $l_n^{(1)}(\hat{\Theta}_n(x)) = 0$  nach Definition des Maximum-Likelihood-Schätzers. Damit erhalten wir

$$\sqrt{n}(\hat{\Theta}_n(x) - \vartheta) = -\frac{\frac{1}{\sqrt{n}}l_n^{(1)}(\vartheta)}{\frac{1}{n}l_n^{(2)}(\vartheta) + \frac{1}{n}(\hat{\Theta}_n(x) - \vartheta)l_n^{(3)}(\vartheta^*)}.$$

Wir wollen nun die folgenden drei Schritte beweisen:

- a)  $\frac{1}{\sqrt{nI(\vartheta)}}l_n^{(1)}(\vartheta) \rightarrow \mathcal{N}(0, 1)$  in Verteilung.
- b)  $l_n^{(2)}(\vartheta) \rightarrow I(\vartheta)$  in Wahrscheinlichkeit.
- c) Es existiert ein  $c > 0$ , so dass  $P_\vartheta^\infty \left( \lim_{n \rightarrow \infty} \left| \frac{1}{n}l_n^{(3)}(\vartheta^*) \right| < c \right) = 1$  gilt.

Für a) betrachten wir

$$\frac{1}{\sqrt{n}}l_n^{(1)}(\vartheta) = \frac{1}{\sqrt{n}} \frac{\partial}{\partial \vartheta} \left( \sum_{i=1}^n \log f(x_i, \vartheta) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \vartheta} \log f(x_i, \vartheta).$$

Ferner gilt  $\mathbf{E}_\vartheta \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta) = 0$ , siehe hierzu den Beweis von Satz 2.5.1, und überdies  $I(\vartheta) = \text{Var}_\vartheta \frac{\partial}{\partial \vartheta} \log f(\cdot, \vartheta)$ . Mit dem zentralen Grenzwertsatz CLT folgt dann  $\frac{1}{\sqrt{nI(\vartheta)}}l_n^{(1)}(\vartheta) \rightarrow \mathcal{N}(0, 1)$  in Verteilung, womit a) bewiesen ist. Für b) betrachten wir

$$\begin{aligned} -\frac{1}{n}l_n^{(2)}(\vartheta) &= -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial \vartheta^2} \log f(x_i, \vartheta) = \frac{1}{n} \sum_{i=1}^n \left( \frac{\frac{\partial}{\partial \vartheta} f(x_i, \vartheta)}{f(x_i, \vartheta)} \right)^2 - \frac{1}{n} \sum_{i=1}^n \frac{\frac{\partial^2}{\partial \vartheta^2} f(x_i, \vartheta)}{f(x_i, \vartheta)} \\ &\xrightarrow{\text{SLLN}} \mathbf{E}_\vartheta \left( \frac{\frac{\partial}{\partial \vartheta} f(\cdot, \vartheta)}{f(\cdot, \vartheta)} \right)^2 - \mathbf{E}_\vartheta \left( \frac{\frac{\partial^2}{\partial \vartheta^2} f(\cdot, \vartheta)}{f(\cdot, \vartheta)} \right)^2 \\ &= I(\vartheta) - \mathbf{E}_\vartheta \left( \frac{\frac{\partial^2}{\partial \vartheta^2} f(\cdot, \vartheta)}{f(\cdot, \vartheta)} \right)^2. \end{aligned}$$

Für den hinteren Term gilt nun aber gerade

$$\mathbf{E}_\vartheta \left( \frac{\frac{\partial^2}{\partial \vartheta^2} f(\cdot, \vartheta)}{f(\cdot, \vartheta)} \right)^2 = \int_B \frac{\partial^2}{\partial \vartheta^2} f(x, \vartheta) \, d\mu(x) \stackrel{\text{iv)}}{=} \frac{\partial^2}{\partial \vartheta^2} \underbrace{\int_B f(x, \vartheta) \, d\mu(x)}_{=1} = 0.$$

Damit kommen wir nun zu c). Wegen v) gilt für  $|\vartheta^* - \vartheta| < c_\vartheta$

$$\left| \frac{1}{n} l_n^{(3)}(x_i, \vartheta^*) \right| \leq \frac{1}{n} \sum_{i=1}^n g_\vartheta(x_i) \longrightarrow \mathbf{E}_\vartheta g_\vartheta =: c < \infty, \quad (**)$$

wobei diese Konvergenz  $P_\vartheta^\infty$ -fast sicher gilt. Ferner liegt  $\vartheta^*$  zwischen  $\hat{\Theta}_n(x)$  und  $\vartheta$  und es gilt  $|\hat{\Theta}_n(x) - \vartheta| \rightarrow 0$  in Wahrscheinlichkeit  $P_\vartheta^\infty$ . Damit folgt, dass (\*\*) mit einer Wahrscheinlichkeit gilt, die gegen 1 konvergiert, woraus Aussage c) folgt. Ferner gilt nun

$$\sqrt{nI(\vartheta)}(\hat{\Theta}_n(x) - \vartheta) = - \frac{\frac{1}{\sqrt{nI(\vartheta)}} l_n^{(1)}(\vartheta)}{\frac{1}{n} l_n^{(2)}(\vartheta) + \frac{1}{n}(\hat{\Theta}_n(x) - \vartheta) l_n^{(3)}(\vartheta^*)} I(\vartheta) \longrightarrow \frac{\mathcal{N}(0, 1)}{I(\vartheta) + 0} I(\vartheta) = \mathcal{N}(0, 1).$$

Dies gilt wegen der multiplikativen Fassung des Satzes von Slutsky<sup>1</sup>. □

---

<sup>1</sup> Siehe hierzu Korollar 1.2.6.





# 3

## Konfidenzintervalle

┌

Bisher haben wir Punktschätzer verwendet, die den wahren und unbekanntem Parameter schätzen, ihn im Allgemeinen jedoch nicht sicher treffen. In diesem Kapitel wollen wir nicht mehr einen einzelnen Punkt, sondern ein Intervall schätzen. ┘

Für die Approximation des wahren, unbekanntem Parameters  $\vartheta$  haben wir im vorherigen Kapitel im Wesentlichen drei Eigenschaften kennengelernt:

- i) *Konsistenz*:  $\hat{\Theta}_n \rightarrow \vartheta$   $P_\vartheta^\infty$ -fast sicher.
- ii) *Erwartungstreue*:  $\mathbf{E}_\vartheta \hat{\Theta}_n = \vartheta$ .
- iii) *Mittlere quadratische Abweichung*:  $\int (\hat{\Theta}_n - \vartheta)^2 dP_\vartheta^n$ .

Wir werden in diesem Kapitel wenn nicht anders angegeben stets davon ausgehen, dass  $\Theta \subset \mathbf{R}$  gilt und  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$  ist. Wie bisher sei  $X_i: \mathbf{R}^n \rightarrow \mathbf{R}$  die  $i$ -te Projektion für  $X = (X_1, \dots, X_n)$  und  $P_\vartheta^n$  die erzeugende Verteilung.

### 3.1 Grundlegende Definitionen

#### Definition 3.1.1 Konfidenzintervall

Es seien  $\underline{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{-\infty\}$  und  $\overline{\Theta}: \mathbf{R}^n \rightarrow \mathbf{R} \cup \{\infty\}$  Stichprobenfunktionen mit  $\underline{\Theta}(x) \leq \overline{\Theta}(x)$  für alle  $x \in \mathbf{R}^n$ . Für  $\gamma \in [0, 1]$  heißt das zufällige Intervall  $[\underline{\Theta}, \overline{\Theta}]$  **Konfidenzintervall** zum Niveau  $\gamma$  genau dann, wenn für alle  $\vartheta \in \Theta$  gilt:

$$P_\vartheta^n \left( \left\{ x \in \mathbf{R}^n : \vartheta \in [\underline{\Theta}(x), \overline{\Theta}(x)] \right\} \right) \geq \gamma$$

Konstruiert man zu einer Beobachtung  $x \in \mathbf{R}^n$  ein Konfidenzintervall  $[\underline{\Theta}(x), \overline{\Theta}(x)]$ , so bedeutet die Ungleichung aus Definition 3.1.1 *nicht*, dass  $\vartheta \in [\underline{\Theta}(x), \overline{\Theta}(x)]$  mit einer Wahrscheinlichkeit größer als  $\gamma$  gilt, da sowohl  $\vartheta$ , als auch  $\underline{\Theta}(x)$  und  $\overline{\Theta}(x)$  keine Zufallsvariablen sind und damit der Wahrscheinlichkeitsbegriff keinen Sinn ergibt. Die Ungleichung bedeutet vielmehr, dass

bei vielen Wiederholungen mit einer Wahrscheinlichkeit von  $100 \cdot \gamma\%$  damit zu rechnen ist, dass  $\vartheta \in [\underline{\Theta}(x), \overline{\Theta}(x)]$  gilt.

Da wir in der Regel möglichst kleine Konfidenzintervalle haben wollen, führen wir noch die folgende Definition ein.

**Definition 3.1.2 Minimales Konfidenzintervall**

Ein Konfidenzintervall  $[\underline{\Theta}, \overline{\Theta}]$  heißt **minimal** zum Niveau  $\gamma \in [0, 1]$  genau dann, wenn gilt:

$$\inf_{\vartheta \in \Theta} P_{\vartheta}^n \left( \left\{ x \in \mathbf{R}^n : \vartheta \in [\underline{\Theta}(x), \overline{\Theta}(x)] \right\} \right) = \gamma$$

**Definition 3.1.3 Asymptotisches Konfidenzintervall**

Für  $\underline{\Theta}_n, \overline{\Theta}_n : \mathbf{R}^n \rightarrow \mathbf{R}$  gelte die Abschätzung

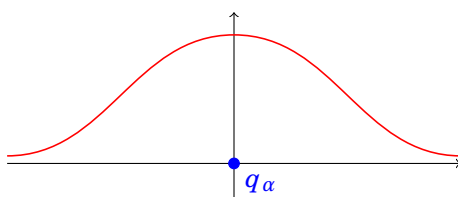
$$\lim_{n \rightarrow \infty} P_{\vartheta}^n \left( \left\{ x \in \mathbf{R}^n : \vartheta \in [\underline{\Theta}_n(x), \overline{\Theta}_n(x)] \right\} \right) \geq \gamma.$$

Dann nennen wir  $[\underline{\Theta}_n, \overline{\Theta}_n]$  **asymptotisches Konfidenzintervall**.

**Definition 3.1.4  $\alpha$ -Quantil**

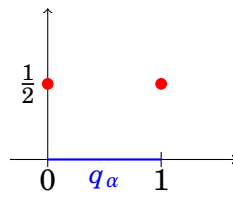
Es sei  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathbf{R}$  und  $\alpha \in [0, 1]$ . Dann definieren wir das  **$\alpha$ -Quantil** von  $P$  wie folgt:

$$q_{\alpha} := \{ t \in \mathbf{R} : P((-\infty, t]) \geq \alpha \text{ und } P([t, \infty)) \geq 1 - \alpha \}$$



**Abbildung 3.1:** Für  $P = \mathcal{N}(0, 1)$  und  $\alpha = 0.5$  ergibt sich das Quantil  $q_{0.5} = 0$ .

Wie wir in Abbildung 3.2 sehen können ist  $q_{\alpha}$  im Allgemeinen nicht einelementig. Ferner ist  $q_{\alpha}$  ein beschränktes und abgeschlossenes Intervall für  $\alpha \in (0, 1)$ . Ist  $q_{\alpha} = [q_{\min, \alpha}^*, q_{\max, \alpha}^*]$ , so gilt  $P\left(\left(q_{\min, \alpha}^*, q_{\max, \alpha}^*\right)\right) = 0$ .



**Abbildung 3.2:** Für  $P = B(1, \frac{1}{2})$  und  $\alpha = \frac{1}{2}$  ergibt sich das nicht einelementige Quantil  $q_{0,5} = [0, 1]$ .

**Lemma 3.1.5**

Ist  $P$  atomlos, d. h. es gilt  $P(\{t\}) = 0$  für alle  $t \in \mathbf{R}$ , so gilt

$$q_\alpha = \{t \in \mathbf{R} : P((-\infty, t]) = \alpha\}.$$

**Beweis:** Die Eigenschaft  $P([t, \infty)) \geq 1 - \alpha$  ist äquivalent zu  $P((-\infty, t)) \leq \alpha$ . Da  $P$  atomlos ist, ist dies äquivalent zu  $P((-\infty, t]) \leq \alpha$ . □

**Lemma 3.1.6**

Sei  $P$  ein Wahrscheinlichkeitsmaße auf  $\mathbf{R}$  mit der Verteilungsfunktion  $F$ . Ferner sei für  $\alpha \in (0, 1)$  das  $\alpha$ -Quantil durch  $q_\alpha = [q_{\min, \alpha}^*, q_{\max, \alpha}^*]$  gegeben. Dann gilt für das sogenannte verallgemeinerte Inverse von  $F$ :

$$\begin{aligned} F^{-1}(\alpha) &:= \inf_{t \in \mathbf{R}} \{F(t) \geq \alpha\} = \min_{t \in \mathbf{R}} \{F(t) \geq \alpha\} \\ &= q_{\min, \alpha}^* \end{aligned}$$

**Beweis:** Es gilt  $F(q_{\min, \alpha}^*) = P((-\infty, q_{\min, \alpha}^*]) \geq \alpha$ . Daraus folgt  $F^{-1}(\alpha) \leq q_{\min, \alpha}^*$ . Nun wollen wir zeigen, dass das Infimum gleich dem Minimum ist. Dazu sei  $t_n \searrow F^{-1}(\alpha)$  mit  $F(t_n) \geq \alpha$ . Als Verteilungsfunktion ist  $F$  stetig und wir erhalten  $F(F^{-1}(\alpha)) \geq \alpha$ .

Wir müssen nun noch die Abschätzung „ $\geq$ “ zeigen. Angenommen, es gilt  $F^{-1}(\alpha) < q_{\min, \alpha}^*$ , dann muss  $F^{-1}(\alpha) \notin q_\alpha$  gelten. Wegen  $F(F^{-1}(\alpha)) = P((-\infty, F^{-1}(\alpha)]) \geq \alpha$  folgt dann, dass  $P((-\infty, F^{-1}(\alpha))) > \alpha$  ist. Da  $(-\infty, F^{-1}(\alpha) - n^{-1}] \nearrow (-\infty, F^{-1}(\alpha))$  gilt, folgt, dass es ein  $n \geq 1$  mit  $P((-\infty, F^{-1}(\alpha) - n^{-1}]) \geq \alpha$  gibt. Wegen  $F^{-1}(\alpha) - n^{-1} < F^{-1}(\alpha)$  erhalten wir dann einen Widerspruch zu unserer Annahme, dass  $F^{-1}(\alpha)$  der kleinste Wert mit  $F(t) \geq \alpha$  ist. □

**Korollar 3.1.7**

Sei  $P$  ein Wahrscheinlichkeitsmaß auf  $\mathbf{R}$ , welches eine Lebesgue-Dichte  $h$  besitzt, für welche die Menge  $\{h > 0\}$  ein Intervall ist. Für  $\alpha \in (0, 1)$  gelten dann folgende Aussagen:

- i) Es ist  $q_{\min, \alpha}^* = q_{\max, \alpha}^*$ .
- ii)  $q_{\min, \alpha}^*$  ist die eindeutige Lösung der Gleichung  $F(s) = \alpha$ .

In diesem Fall schreiben wir  $q_{\alpha}^* := q_{\min, \alpha}^* = q_{\max, \alpha}^*$ .

**Beweis:**

- i) Angenommen, es gilt  $q_{\min, \alpha}^* < q_{\max, \alpha}^*$ , dann folgt

$$0 = P\left((q_{\min, \alpha}^*, q_{\max, \alpha}^*)\right) = \int_{q_{\min, \alpha}^*}^{q_{\max, \alpha}^*} h \, d\lambda.$$

Da  $h \geq 0$  gilt, folgt  $h(x) = 0$  für  $\lambda$ -fast alle  $x \in (q_{\min, \alpha}^*, q_{\max, \alpha}^*)$ . Da  $\{h > 0\}$  n. V. ein Intervall ist, folgt  $\{h > 0\} \subset [q_{\max, \alpha}^*, \infty)$  oder  $\{h > 0\} \subset (-\infty, q_{\min, \alpha}^*]$ . Betrachten wir z. B. den ersten Fall, so folgt

$$\alpha \leq P\left((-\infty, q_{\min, \alpha}^*]\right) = \int_{-\infty}^{q_{\min, \alpha}^*} h \, d\lambda = 0.$$

Damit wäre  $\alpha = 0$ , was im Widerspruch zu  $\alpha \in (0, 1)$  steht. Der andere Fall verläuft analog.

- ii) Diese Aussage folgt direkt aus Lemma 3.1.5, Lemma 3.1.6 und i). □

**Beispiel 3.1.8 Normalverteilung**

Sei  $\varphi(t) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$  die Dichte von  $\mathcal{N}(0, 1)$  und  $\Phi(s) := \int_{-\infty}^s \varphi(t) \, dt$  für  $s \in \mathbf{R}$  die Verteilungsfunktion von  $\mathcal{N}(0, 1)$ . Dann ist mit Korollar 3.1.7  $q_{\alpha}^*$  die eindeutige Lösung von  $\Phi(s) = \alpha$ , das heißt es gilt  $\Phi(q_{\alpha}^*) = \alpha$ , da  $\{\varphi > 0\} = \mathbf{R}$  gilt. Ferner gilt:

- i) Es ist  $q_{\alpha} = -q_{1-\alpha}$  wegen  $\Phi(s) = 1 - \Phi(-s)$  für alle  $s \in \mathbf{R}$ .
- ii) Die Abbildung  $\alpha \mapsto q_{\alpha}$  ist strikt monoton wachsend, da  $\Phi$  strikt monoton wächst.
- iii) Es ist  $q_{0.5}^* = 0$ .

Die Quantile  $q_{\alpha}^*$  können in Tabellen nachgeschlagen oder mit Hilfe von Programmen wie R numerisch berechnet werden. Für  $\alpha \in (0, 1)$  und  $X \sim \mathcal{N}(0, 1)$  gilt

$$\begin{aligned} P\left(q_{\frac{\alpha}{2}}^* \leq X \leq q_{1-\frac{\alpha}{2}}^*\right) &= P\left(X \leq q_{1-\frac{\alpha}{2}}^*\right) - P\left(X \leq q_{\frac{\alpha}{2}}^*\right) = \Phi\left(q_{1-\frac{\alpha}{2}}^*\right) - \Phi\left(q_{\frac{\alpha}{2}}^*\right) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} \\ &= 1 - \alpha. \quad // \end{aligned}$$

**Beispiel 3.1.9** Quantile von  $\chi_n^2$ 

Die  $\chi_n^2$ -Verteilung mit  $n$  Freiheitsgraden besitzt für  $y \in \mathbf{R}$  nach Satz 1.2.12 die folgende Lebesguedichte:

$$h(y) := \mathbf{1}_{[0, \infty)}(y) \cdot \frac{y^{\frac{n}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$$

Im genannten Satz wurde zudem gezeigt, dass für i. i. d. Zufallsvariablen  $X_1, \dots, X_n \sim \mathcal{N}(0, 1)$  die Eigenschaft  $\sum_{i=1}^n X_i^2 \sim \chi_n^2$  gilt. Nun ist Korollar 3.1.7 anwendbar, da  $\{h > 0\} = [0, \infty)$  ein Intervall ist. Die Berechnung der Quantile ist auch hier nur mit Hilfe von Tabellen oder mit numerischen Methoden möglich. //

**Beispiel 3.1.10** Quantile von  $t_n$ 

Die Student- $t$ -Verteilung mit  $n$  Freiheitsgraden besitzt für  $y \in \mathbf{R}$  die folgende Lebesguedichte:

$$h(y) := \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\sqrt{n\pi} \left(1 + \frac{y^2}{n}\right)^{\frac{n+1}{2}}}$$

In diesem Fall gilt  $\{h > 0\} = \mathbf{R}$  und damit ist Korollar 3.1.7 anwendbar. Ferner ist  $h$  symmetrisch um 0 und wir erhalten wie in Beispiel 3.1.8 folgende Aussagen:

- i) Es gilt  $q_\alpha = -q_{1-\alpha}$ .
- ii) Die Abbildung  $\alpha \mapsto q_\alpha$  ist strikt monoton wachsend.
- iii) Es gilt  $q_{0.5} = 0$ . //

## 3.2 Einige exakte Berechnungen für Konfidenzintervalle

Wir wollen für einige Beispiele nun exakte Berechnungen durchführen.

### Konfidenzintervalle für $\mathcal{N}(\mu, \sigma^2)$ – I

Wir gehen davon aus, dass wir i. i. d. Zufallsvariablen  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  haben, wobei  $\mu$  unbekannt und  $\sigma^2$  bekannt ist. Im Beweis von Lemma 1.2.7 haben wir bereits gesehen, dass  $\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$  gilt. Für  $\alpha \in (0, 1)$  und das  $\alpha$ -Quantil von  $\mathcal{N}(0, 1)$ , also  $q_\alpha$ , folgt mit Beispiel 3.1.8:

$$\begin{aligned} P_\mu^n \left( -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq q_{1-\frac{\alpha}{2}} \right) &= 1 - \alpha \\ \Leftrightarrow P_\mu^n \left( \bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \right) &= 1 - \alpha \end{aligned}$$

Damit definieren wir nun:

$$\begin{aligned} \underline{\Theta}_n(X) &:= \bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \\ \bar{\Theta}_n(X) &:= \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}} \end{aligned}$$

Daraus folgt, dass  $[\underline{\Theta}_n, \bar{\Theta}_n]$  ein minimales Konfidenzintervall zum Niveau  $1 - \alpha$  für den Parameter  $\mu$  ist. Die Länge des Konfidenzintervalls ist  $\frac{2\sigma}{\sqrt{n}} q_{1-\frac{\alpha}{2}}$  und wir stellen fest, dass dies keine Zufallsvariable ist. Damit kann a-priori sichergestellt werden, dass die Länge unterhalb eines Schwellwertes  $\varepsilon > 0$  liegt, das heißt  $\bar{\Theta}_n - \underline{\Theta}_n \leq \varepsilon$ , und zwar indem man zum Beispiel das  $n$  wie folgt wählt:

$$n \geq \left( \frac{2\sigma q_{1-\frac{\alpha}{2}}}{\varepsilon} \right)^2$$

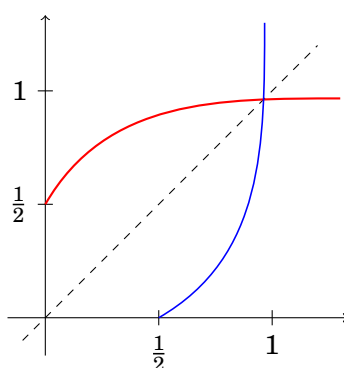
Als nächstes wollen wir asymmetrische Konfidenzintervalle betrachten. Dazu seien  $\alpha_1, \alpha_2 \in [0, \frac{1}{2})$  und  $\alpha := \alpha_1 + \alpha_2 \in (0, 1)$  (insbesondere muss also  $\alpha_1 \alpha_2 > 0$  sein). Die Verteilungsfunktion von  $\mathcal{N}(0, 1)$  sei  $\Phi$ . Dann gilt

$$P_\mu^n \left( q_{\alpha_2} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq q_{1-\alpha_1} \right) = \Phi(q_{1-\alpha_1}) - \Phi(q_{\alpha_2}) = 1 - \alpha_1 - \alpha_2 = 1 - \alpha.$$

Dies ergibt ein asymmetrisches Konfidenzintervall  $[\underline{\Theta}_n, \bar{\Theta}_n]$  zum Niveau  $1 - \alpha$  vermöge

$$\begin{aligned} \underline{\Theta}_n &:= \bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\alpha_1} \\ \bar{\Theta}_n &:= \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_{\alpha_2}. \end{aligned}$$

Für  $\alpha_1 = 0$  ergibt sich insbesondere ein *einseitiges* Konfidenzintervall  $(-\infty, \bar{X}_n + \frac{\sigma}{\sqrt{n}} q_\alpha]$ . Analog ergibt sich für  $\alpha_2 = 0$  das einseitige Konfidenzintervall  $[\bar{X}_n - \frac{\sigma}{\sqrt{n}} q_{1-\alpha}, \infty)$ .



**Abbildung 3.3:** Darstellung von  $s \mapsto \Phi(s)$  und der Umkehrfunktion, die sich anschaulich als Spiegelung an der ersten Winkelhalbierenden ergibt.

Nun wollen wir die Länge der asymmetrischen Konfidenzintervalle betrachten. Diese beträgt  $\bar{\Theta}_n - \underline{\Theta}_n = \frac{\sigma}{\sqrt{n}}(q_{1-\alpha_1} - q_{1-\alpha-\alpha_1})$  und wir wollen untersuchen, wann diese Länge minimal wird.

In Abbildung 3.3 sehen wir, dass  $s \mapsto \Phi(s)$  auf  $[0, \infty)$  strikt konkav und damit  $\alpha \mapsto q_\alpha$  als Umkehrfunktion für  $\alpha \in [\frac{1}{2}, 1)$  strikt konvex ist. Damit ist auch  $\alpha \mapsto q_{1-\alpha}$  auf  $(0, \frac{1}{2}]$  strikt konvex und es folgt

$$\frac{1}{2}q_{1-\alpha+\alpha_1} + \frac{1}{2}q_{1-\alpha_1} \geq q_{\frac{1}{2}(1-\alpha+\alpha_1) + \frac{1}{2}(1-\alpha_1)} = q_{1-\alpha_2}.$$

Die Gleichheit herrscht dabei genau dann, wenn  $1 - \alpha + \alpha_1 = 1 - \alpha_1$ , also  $\alpha_1 = \frac{\alpha}{2}$  ist. Damit folgt, dass  $\alpha_1 \mapsto q_{1-\alpha+\alpha_1} + q_{1-\alpha_1}$  ein eindeutiges Minimum bei  $\alpha_1 = \frac{\alpha}{2}$  annimmt und für  $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$  ist die Länge des Konfidenzintervalls daher minimal. Interessiert man sich also für möglichst kleine Konfidenzintervalle, so sollte man auf symmetrische Konfidenzintervalle zurückgreifen. Ist hingegen nur die obere Schranke des Konfidenzintervalls wichtig, so ist das einseitige Konfidenzintervall für  $\alpha_1 = 0$  besser, da die rechte Intervallgrenze zum Niveau  $1 - \alpha$  kleiner ist als die rechte Intervallgrenze des symmetrischen Konfidenzintervalls.

## Konfidenzintervalle für $\mathcal{N}(\mu, \sigma^2)$ – II

Wieder seien i. i. d. Zufallsvariablen  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  gegeben, diesmal seien jedoch sowohl  $\mu$ , als auch  $\sigma^2$  unbekannt. Unser Ziel ist es ein Konfidenzintervall für  $\mu$  zu finden.

In Korollar 1.2.16 haben wir gesehen, dass  $\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}$  gilt, wobei  $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$  ist. Wir schreiben nun  $t_{n-1, \alpha}$  für das  $\alpha$ -Quantil von  $t_{n-1}$ . Es gilt

$$\begin{aligned} P_{\mu, \sigma^2}^n \left( t_{n-1, \frac{\alpha}{2}} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \leq t_{n-1, 1-\frac{\alpha}{2}} \right) &= 1 - \alpha \\ \Leftrightarrow P_{\mu, \sigma^2}^n \left( \underbrace{\bar{X}_n - t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}}_{\underline{\Theta}_n(X)} \leq \mu \leq \underbrace{\bar{X}_n + t_{n-1, 1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}}_{\bar{\Theta}_n(X)} \right) &= 1 - \alpha. \end{aligned}$$

Damit erhalten wir das minimale symmetrische Konfidenzintervall  $[\underline{\Theta}_n, \bar{\Theta}_n]$  für  $\mu$  zum Niveau  $1 - \alpha$ . Die Länge des Konfidenzintervalls ist in diesem Fall jedoch wirklich eine Zufallsvariable:

$$\bar{\Theta}_n - \underline{\Theta}_n = \frac{2t_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n}} S_n$$

Für das asymmetrische Konfidenzintervall seien wieder  $\alpha_1, \alpha_2 \in [0, \frac{1}{2})$  und  $\alpha := \alpha_1 + \alpha_2 \in (0, 1)$ . Dann folgt mit einer der selben Rechnung wie im ersten Beispiel

$$P_{\mu, \sigma^2}^n \left( t_{n-1, 1-\alpha_2} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \leq t_{n-1, 1-\alpha_1} \right) = 1 - \alpha.$$

Daraus erhalten wir dementsprechend für die Grenzen des Konfidenzintervalls:

$$\begin{aligned} \underline{\Theta}_n(X) &:= \bar{X}_n - t_{n-1, 1-\alpha_1} \frac{S_n}{\sqrt{n}} \\ \bar{\Theta}_n(X) &:= \bar{X}_n + t_{n-1, 1-\alpha_2} \frac{S_n}{\sqrt{n}} \end{aligned}$$

Im Speziellen erhält man für  $\alpha_1 = 0$  das Konfidenzintervall  $(-\infty, \bar{X}_n + t_{n-1, 1-\alpha} \frac{S_n}{\sqrt{n}}]$  und für  $\alpha_2 = 0$  das Konfidenzintervall  $[\bar{X}_n - t_{n-1, 1-\alpha} \frac{S_n}{\sqrt{n}}, \infty)$ . Die Längenbetrachtungen verlaufen analog zum ersten Beispiel.

### Konfidenzintervalle für $\mathcal{N}(\mu, \sigma^2)$ – III

Wieder seien i. i. d. Zufallsvariablen  $X_i \sim \mathcal{N}(\mu, \sigma^2)$  gegeben und beide Parameter seien unbekannt. Nun suchen wir ein Konfidenzintervall für  $\sigma^2$ .

In Satz 1.2.13 haben wir gesehen, dass  $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$  gilt. Für  $\alpha_1, \alpha_2 \in [0, \frac{1}{2})$  und  $\alpha := \alpha_1 + \alpha_2 \in (0, 1)$  folgt damit für das  $\alpha$ -Quantil  $q_\alpha$  von  $\chi_{n-1}^2$

$$\begin{aligned} P_{\mu, \sigma^2}^n \left( q_{\alpha_2} \leq \frac{(n-1)S_n^2}{\sigma^2} \leq q_{1-\alpha_1} \right) &= 1 - \alpha \\ \Leftrightarrow P_{\mu, \sigma^2}^n \left( \frac{(n-1)S_n^2}{q_{1-\alpha_1}} \leq \sigma^2 \leq \frac{(n-1)S_n^2}{q_{\alpha_2}} \right) &= 1 - \alpha. \end{aligned}$$

Dementsprechend setzen wir die Intervallgrenzen wie folgt:

$$\begin{aligned} \underline{\Theta}_n &:= \frac{(n-1)S_n^2}{q_{1-\alpha_1}} \\ \bar{\Theta}_n &:= \frac{(n-1)S_n^2}{q_{\alpha_2}}. \end{aligned}$$

Damit ist  $[\underline{\Theta}_n, \bar{\Theta}_n]$  ein Konfidenzintervall zum Niveau  $1 - \alpha$ . Die Länge berechnet sich dann wie folgt:

$$\bar{\Theta}_n - \underline{\Theta}_n = (n-1) \cdot S_n^2 \left( \frac{1}{q_{\alpha_2}} - \frac{1}{q_{1-\alpha_1}} \right)$$



Damit gilt nun

$$\mathbf{E}_{\mu, \sigma^2}(\bar{\Theta}_n - \underline{\Theta}_n) = (n-1)\sigma^2 \left( \frac{1}{q_{\alpha_2}} - \frac{1}{q_{1-\alpha_1}} \right).$$

Wir wollen nun unsere Annahme ändern und davon ausgehen, dass der Erwartungswert  $\mu$  bekannt ist. Wir betrachten die Stichprobenfunktion  $\tilde{S}_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ . Da  $\frac{1}{\sigma}(X_i - \mu) \sim \mathcal{N}(0, 1)$  gilt, folgt aus der Definition der  $\chi_n^2$ -Verteilung, dass  $\frac{n}{\sigma^2} \tilde{S}_n^2 \sim \chi_n^2$  gilt. Analog zur obigen Rechnung ergeben sich dann

$$\begin{aligned} \tilde{\bar{\Theta}}_n(X) &:= \frac{n\tilde{S}_n^2}{q_{1-\alpha_1}} \\ \tilde{\underline{\Theta}}_n(X) &:= \frac{n\tilde{S}_n^2}{q_{\alpha_2}}. \end{aligned}$$

Dabei ist  $q_\alpha$  das  $\alpha$ -Quantil von  $\chi_n^2$ . Es gilt zu beachten, dass es vorkommen kann, dass die Länge  $\tilde{\bar{\Theta}}_n - \tilde{\underline{\Theta}}_n = n\tilde{S}_n^2 \left( \frac{1}{q_{\alpha_2}} - \frac{1}{q_{1-\alpha_1}} \right)$  größer ist als  $\bar{\Theta}_n - \underline{\Theta}_n$ . Eine Begründung hierfür ist, dass sehr große oder kleine Beobachtungen  $x_i$  von  $\bar{x}_n$  in  $S_n^2$  kompensiert werden. Dieser Effekt tritt bei  $\tilde{S}_n^2$  jedoch nicht auf.

### 3.3 Asymptotische Konfidenzintervalle

Bis jetzt konnten wir die Wahrscheinlichkeiten für die auftretenden Konfidenzintervalle exakt bestimmen. Als nächstes betrachten wir Situationen, in denen dies nicht möglich ist und bestimmen diese Wahrscheinlichkeiten daher asymptotisch. Das Mittel der Wahl für diese Aufgabe ist der zentrale Grenzwertsatz (CLT).

**Satz 3.3.1**

Sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$  mit der Eigenschaft  $\mathbf{E}_\vartheta X_1 = \vartheta$  für alle  $\vartheta \in \Theta$ , d. h. der Erwartungswert parametrisiert die Familie. Für  $\alpha \in (0, 1)$  bezeichne  $q_\alpha$  das  $\alpha$ -Quantil von  $\mathcal{N}(0, 1)$ . Ferner setzen wir

$$\underline{\Theta}_n := \bar{X}_n - \frac{q_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}}$$

$$\bar{\Theta}_n := \bar{X}_n + \frac{q_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}}.$$

Dann ist  $[\underline{\Theta}_n, \bar{\Theta}_n]$  ein asymptotisches Konfidenzintervall für  $\vartheta$  zum Niveau  $1 - \alpha$ .

**Beweis:** Wir betrachten

$$\begin{aligned} \lim_{n \rightarrow \infty} P_\vartheta^n \left( \bar{X}_n - \frac{q_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}} \leq \vartheta \leq \bar{X}_n + \frac{q_{1-\frac{\alpha}{2}} S_n}{\sqrt{n}} \right) &= \lim_{n \rightarrow \infty} P_\vartheta^n \left( -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq q_{1-\frac{\alpha}{2}} \right) \\ &= \lim_{n \rightarrow \infty} P_\vartheta^n \left( \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq q_{1-\frac{\alpha}{2}} \right) \\ &\quad - \lim_{n \rightarrow \infty} P_\vartheta^n \left( \sqrt{n} \frac{\bar{X}_n - \mu}{S_n} \leq -q_{1-\frac{\alpha}{2}} \right). \end{aligned}$$

Mit Korollar 1.2.6 erhalten wir

$$\begin{aligned} &= \Phi(q_{1-\frac{\alpha}{2}}) - \Phi(-q_{1-\frac{\alpha}{2}}) \\ &= 1 - \alpha. \end{aligned}$$

□

Dies funktioniert auch ohne  $S_n$ , falls die Varianz vom Erwartungswert abhängt, das heißt falls  $\text{Var}_\vartheta X_1 = h(\mathbf{E}_\vartheta X_1)$  für alle  $\vartheta \in \Theta$  ist.

**Beispiel 3.3.2** *Poissonverteilung*

Es sei  $\text{Pois}(\lambda)$  die Poissonverteilung zum Parameter  $\lambda > 0$ , das heißt die Verteilung auf  $\mathbf{N}_0$ , die für  $k \in \mathbf{N}_0$  durch die Zähldichte  $h(k) := \frac{\lambda^k}{k!} e^{-\lambda}$  gegeben ist. Es seien nun i. i. d. Zufallsvariablen

$X_i \sim \text{Pois}(\lambda)$  mit unbekanntem  $\lambda$  gegeben. Es gilt  $\mathbf{E}_\lambda X_1 = \text{Var}_\lambda X_1 = \lambda$ . Mit dem CLT folgt für das  $\alpha$ -Quantil  $q_\alpha$  von  $\mathcal{N}(0, 1)$

$$\lim_{n \rightarrow \infty} P_\lambda^n \left( -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \leq q_{1-\frac{\alpha}{2}} \right) = 1 - \alpha. \quad (*)$$

Wir formen den inneren Ausdruck zunächst um und setzen  $q := q_{1-\frac{\alpha}{2}}$ :

$$\begin{aligned} & -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \leq q_{1-\frac{\alpha}{2}} \\ \Leftrightarrow & \left| \sqrt{n} \frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \right| \leq q_{1-\frac{\alpha}{2}} \\ \Leftrightarrow & \left( \sqrt{n} \frac{\overline{X}_n - \lambda}{\sqrt{\lambda}} \right)^2 \leq q_{1-\frac{\alpha}{2}}^2 \\ \Leftrightarrow & (\lambda - \overline{X}_n)^2 \leq \frac{\lambda q_{1-\frac{\alpha}{2}}^2}{n} \\ \Leftrightarrow & \lambda^2 - 2\lambda \overline{X}_n - \frac{\lambda q^2}{n} + \overline{X}_n^2 \leq 0 \\ \Leftrightarrow & \lambda^2 - 2\lambda \left( \overline{X}_n + \frac{q^2}{2n} \right) + \left( \overline{X}_n + \frac{q^2}{2n} \right)^2 \leq \frac{\overline{X}_n q^2}{n} + \frac{q^4}{4n^2} \\ \Leftrightarrow & \left| \lambda - \overline{X}_n - \frac{q^2}{2n} \right| \leq \sqrt{\frac{\overline{X}_n q^2}{n} + \frac{q^4}{4n^2}} \end{aligned}$$

Damit erhalten wir für die Grenzen des Konfidenzintervalls:

$$\underline{\Theta}_n := \overline{X}_n + \frac{q_{1-\frac{\alpha}{2}}^2}{2n} - \sqrt{\frac{\overline{X}_n q^2}{n} + \frac{q^4}{4n^2}} \quad (**)$$

$$\overline{\Theta}_n := \overline{X}_n + \frac{q_{1-\frac{\alpha}{2}}^2}{2n} + \sqrt{\frac{\overline{X}_n q^2}{n} + \frac{q^4}{4n^2}} \quad (**)$$

Dies ist ein asymptotisches Konfidenzintervall für  $\lambda$  zum Niveau  $1 - \alpha$ . Die Länge des Intervalls beträgt

$$\overline{\Theta}_n - \underline{\Theta}_n = 2\sqrt{\frac{\overline{X}_n q_{1-\frac{\alpha}{2}}^2}{n} + \frac{q_{1-\frac{\alpha}{2}}^4}{4n^2}}.$$

Eine Variante wäre es,  $\sqrt{\lambda}$  durch  $\sqrt{\overline{X}_n}$  zu ersetzen. Da das starke Gesetz der großen Zahlen (SLLN) ergibt, dass  $\overline{X}_n \rightarrow \lambda = \mathbf{E}_\lambda X_1$  ist, folgt auch  $\sqrt{\overline{X}_n} \rightarrow \sqrt{\lambda}$   $P_\lambda^\infty$ -fast sicher. Mit dem zentralen Grenzwertsatz CLT und der multiplikativen Fassung des Satzes von Slutsky folgt

$$\lim_{n \rightarrow \infty} P_\lambda^n \left( -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X}_n - \lambda}{\sqrt{\overline{X}_n}} \leq q_{1-\frac{\alpha}{2}} \right) = 1 - \alpha. \quad (***)$$

Jetzt ist die entsprechende Umformung einfacher durchzuführen:

$$\begin{aligned}
 & -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X}_n - \lambda}{\sqrt{\overline{X}_n}} \leq q_{1-\frac{\alpha}{2}} \\
 \Leftrightarrow & -\frac{\sqrt{\overline{X}_n} q_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \overline{X}_n - \lambda \leq \frac{\sqrt{\overline{X}_n} q_{1-\frac{\alpha}{2}}}{\sqrt{n}} \\
 \Leftrightarrow & \underbrace{\overline{X}_n - \frac{\sqrt{\overline{X}_n} q_{1-\frac{\alpha}{2}}}{\sqrt{n}}}_{\underline{\Theta}_n} \leq \lambda \leq \overline{X}_n + \underbrace{\frac{\sqrt{\overline{X}_n} q_{1-\frac{\alpha}{2}}}{\sqrt{n}}}_{\overline{\Theta}_n}
 \end{aligned}$$

Dann ist  $[\underline{\Theta}_n, \overline{\Theta}_n]$  ein asymptotisches Konfidenzintervall für  $\lambda$  zum Niveau  $1 - \alpha$ . Die Länge des Konfidenzintervalls berechnet sich zu

$$\overline{\Theta}_n - \underline{\Theta}_n = 2 \frac{\sqrt{\overline{X}_n} q_{1-\frac{\alpha}{2}}}{\sqrt{n}}.$$

Es gilt zu beachten, dass es vorkommen kann, dass  $\underline{\Theta}_n < 0$  in (\*) gilt. Dies ist zu konservativ geschätzt, da  $\lambda > 0$  ist. Wir setzen daher  $\underline{\Theta}'_n := \max\{0, \underline{\Theta}_n\}$  und erhalten so nach wie vor ein Konfidenzintervall  $[\underline{\Theta}'_n, \overline{\Theta}_n]$  für  $\lambda$  zum Niveau  $1 - \alpha$ . Es kann gezeigt werden, dass die Länge dieses korrigierten Konfidenzintervalls stets kleiner als die des durch (\*\*) definierten Konfidenzintervalls ist. Der Grund hierfür ist, dass Ausreißer in  $X_1, \dots, X_n$  durch  $\sqrt{\overline{X}_n}$  im Nenner von (\*\*\*) besser kompensiert werden als durch den deterministischen Wert  $\lambda$  in (\*). //

### Beispiel 3.3.3 Bernoulli-Verteilung

Es seien i. i. d. Zufallsvariablen  $X_i \sim B(1, p)$  für unbekanntes  $p \in (0, 1)$  gegeben. Dann gilt  $\mathbf{E}_p X_1 = p$  und  $\text{Var}_p X_1 = p(1 - p)$ . Der zentrale Grenzwertsatz CLT mit der wohlbekannten Rechnung ergibt dann

$$\lim_{n \rightarrow \infty} P_p^n \left( -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\overline{X}_n - p}{\sqrt{p(1-p)}} \leq q_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

Dann erhalten wir mit  $q := q_{1-\frac{\alpha}{2}}$ :

$$\begin{aligned}
 & \left| \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \right| \leq q_{1-\frac{\alpha}{2}} \\
 \Leftrightarrow & (\bar{X}_n - p)^2 \leq \frac{p(1-p)}{n} q^2 \\
 \Leftrightarrow & p^2 - 2p\bar{X}_n + \bar{X}_n^2 \leq \frac{p(1-p)}{n} q^2 \\
 \Leftrightarrow & p^2 \left(1 + \frac{q^2}{n}\right) - 2p \left(\bar{X}_n + \frac{q^2}{2n}\right) + \bar{X}_n^2 \leq 0 \\
 \Leftrightarrow & p^2 - 2p \frac{\bar{X}_n + \frac{q^2}{2n}}{1 + \frac{q^2}{n}} + \frac{\bar{X}_n^2}{1 + \frac{q^2}{n}} \leq 0 \\
 \Leftrightarrow & \left( p - \frac{\bar{X}_n + \frac{q^2}{2n}}{1 + \frac{q^2}{n}} \right)^2 \leq \left( \frac{n\bar{X}_n + \frac{q^2}{2}}{n+2} \right)^2 - \frac{n\bar{X}_n^2}{n+q^2} \\
 \Leftrightarrow & \left| p - \frac{n\bar{X}_n + \frac{q^2}{2}}{n+q^2} \right| \leq \sqrt{-\frac{nq^2\bar{X}_n^2}{(n+q^2)^2} + \frac{nq^2\bar{X}_n}{(n+q^2)^2} + \frac{\frac{1}{4}q^2}{(n+q^2)^2}} \\
 \Leftrightarrow & \left| p - \frac{n\bar{X}_n + \frac{q^2}{2}}{n+q^2} \right| \leq \frac{q}{n+q^2} \sqrt{n\bar{X}_n(1-\bar{X}_n) + \frac{q^2}{2}}
 \end{aligned}$$

Damit erhalten wir für die Grenzen des Konfidenzintervalls:

$$\begin{aligned}
 \underline{\Theta}_n &:= \frac{1}{n+q_{1-\frac{\alpha}{2}}^2} \left( \frac{q_{1-\frac{\alpha}{2}}^2}{2} + n\bar{X}_n - q_{1-\frac{\alpha}{2}} \sqrt{n\bar{X}_n(1-\bar{X}_n) + \frac{q_{1-\frac{\alpha}{2}}^2}{4}} \right) \\
 \bar{\Theta}_n &:= \frac{1}{n+q_{1-\frac{\alpha}{2}}^2} \left( \frac{q_{1-\frac{\alpha}{2}}^2}{2} + n\bar{X}_n + q_{1-\frac{\alpha}{2}} \sqrt{n\bar{X}_n(1-\bar{X}_n) + \frac{q_{1-\frac{\alpha}{2}}^2}{4}} \right)
 \end{aligned}$$

Dann ist  $[\underline{\Theta}_n, \bar{\Theta}_n]$  ein asymptotisches Konfidenzintervall für  $p$  zum Niveau  $1 - \alpha$ .

Als Variante betrachten wir nun  $\bar{X}_n \rightarrow p$   $P_p^\infty$ -fast sicher. Mit dem Satz von Slutsky erhalten wir dann

$$\lim_{n \rightarrow \infty} P_p^n \left( -q_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leq q_{1-\frac{\alpha}{2}} \right) = 1 - \alpha.$$

Für das Konfidenzintervall erhalten wir dann die folgenden Grenzen:

$$\begin{aligned}
 \underline{\Theta}_n &:= \bar{X}_n - \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\bar{X}_n(1-\bar{X}_n)} \\
 \bar{\Theta}_n &:= \bar{X}_n + \frac{q_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\bar{X}_n(1-\bar{X}_n)}
 \end{aligned}$$

Dies ist ein asymptotisches Konfidenzintervall für  $p$  zum Niveau  $1 - \alpha$ . Da es vorkommen kann, dass  $\underline{\Theta}_n(x) < 0$  oder  $\overline{\Theta}_n(x) > 1$  gilt, können wir wieder die folgende Korrektur vornehmen:

$$\begin{aligned}\underline{\Theta}'_n &:= \max\{0, \underline{\Theta}_n\} \\ \overline{\Theta}'_n &:= \min\{1, \overline{\Theta}_n\}. \quad //\end{aligned}$$

# 4

## Statistische Tests

┌

Bisher haben wir versucht, den unbekannt Parameter für die Verteilung der Grundgesamtheit zu schätzen. Dabei haben wir sowohl versucht, ihn exakt zu schätzen, als auch lediglich Bereiche anzugeben, in welchen er mit einer gewissen Wahrscheinlichkeit liegt. In diesem Kapitel wollen wir Verfahren entwickeln, um zu entscheiden, in welchem „Gebiet“ der Parameter liegt.

└

Wir gehen davon aus, dass  $\Theta \subset \mathbf{R}^m$  gilt, wobei oft  $m = 1$  sein wird. Ferner sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$ . Dies ist zwar nicht notwendig, aber oft eine hilfreiche zusätzliche Annahme. Wie bisher seien  $X_i: \mathbf{R}^n \rightarrow \mathbf{R}$  die  $i$ -ten Projektionen und  $X = (X_1, \dots, X_n) \sim P_\vartheta^n$ , wobei  $\vartheta$  unbekannt ist. Zudem sei die Parametrisierung der Familie injektiv, das heißt es gilt  $P_\vartheta \neq P_{\vartheta'}$  für alle  $\vartheta \neq \vartheta'$ .

Zusätzlich wollen wir in diesem Kapitel davon ausgehen, dass  $\Theta = \Theta_0 \sqcup \Theta_1$  gilt, das heißt es ist  $\Theta_0 \cap \Theta_1 = \emptyset$ , wobei beide Mengen nicht leer sein sollen. Damit bilden sie eine Partition von  $\Theta$ . Wir wollen nun nicht mehr den Wert des unbekannt Parameters  $\vartheta$  schätzen, sondern entscheiden, ob  $\vartheta \in \Theta_0$  oder  $\vartheta \in \Theta_1$  gilt.

Für dieses Kapitel führen wir folgende Sprechweisen ein:

- *Nullhypothese*  $H_0$ : Der wahre Parameter liegt in  $\Theta_0$ .
- *Alternativhypothese*  $H_1$ : Der wahre Parameter liegt in  $\Theta_1$ .
- *Einfache Hypothese*: Die betrachtete Menge  $\Theta_i$  ist einelementig.
- *Einseitiges Testproblem*: Es gilt  $m = 1$  und  $\Theta, \Theta_0$  und  $\Theta_1$  sind Intervalle.
- *Zweiseitiges Testproblem*: Es gilt  $m = 1$ ,  $\Theta$  ist ein Intervall und  $\Theta_0$  oder  $\Theta_1$  ist ein Intervall, die jeweils andere aber nicht.

## 4.1 Qualitätsmerkmale für Tests

### Definition 4.1.1 Statistischer Test

Ein **statistischer Test** ist eine messbare Funktion  $\varphi: \mathbf{R}^n \rightarrow [0, 1]$ . Gilt  $\varphi(x) \in \{0, 1\}$  für alle  $x \in \mathbf{R}^n$ , so heißt  $\varphi$  **nicht-randomisierter Test**. Andernfalls sprechen wir gelegentlich auch von **randomisierten Tests**.

Das Bild  $\varphi(x)$  eines statistischen Tests beschreibt die Wahrscheinlichkeit, mit der sich der Test  $\varphi$  bei der Beobachtung  $x$  für  $H_1$  entscheidet.

Gilt  $\varphi(x) = 0$ , so bedeutet dies, dass sich  $\varphi$  mit der Wahrscheinlichkeit 0 für  $H_1$ , also mit Wahrscheinlichkeit 1 für die Nullhypothese  $H_0$ , entscheidet. Ist hingegen  $\varphi(x) = 1$ , so entscheidet sich der Test mit Wahrscheinlichkeit 1 für die Alternativhypothese. Für  $\varphi(x) =: p \in (0, 1)$  entscheidet sich der Test *zufällig* mit Wahrscheinlichkeit  $p$  für  $H_1$ .

Ist  $\varphi$  nicht-randomisiert, so heißt  $K := \{\varphi = 1\}$  **kritischer Bereich** von  $\varphi$  und es gilt  $\varphi = \mathbf{1}_K$ .

Beim Anwenden eines Tests können folgende Fehler auftreten:

	Entscheidung für $H_0$	Entscheidung für $H_1$
$H_0$ richtig	–	Fehler erster Art
$H_1$ richtig	Fehler zweiter Art	–

Falsche Modellannahmen werden als Fehler dritter Art bezeichnet, diese liegen jedoch außerhalb der mathematischen Beschreibung unseres Modells.

Es ist offensichtlich, dass wir, falls möglich, an Tests interessiert sind, welche die Wahrscheinlichkeiten für Fehler erster und zweiter Art möglichst gering halten. Im Allgemeinen gibt es jedoch keinen Test, der beide Wahrscheinlichkeiten gleichzeitig minimiert. Betrachte zum Beispiel  $\varphi_i(x) := i$  für alle  $x \in \mathbf{R}^n$  mit  $i \in \{0, 1\}$ . Diese verhindern jeweils den Fehler einer Art, machen im Allgemeinen dafür den Fehler der anderen Art.

Eine mögliche Lösung dieser Problematik wäre eine gewichtete Summe der Fehler als Optimierungsziel. Im Allgemeinen geben Probleme aus der Praxis jedoch keinen Anhaltspunkt, wie die Gewichte zu bestimmen sind.

Die asymmetrische Betrachtung versucht einen der beiden Fehler durch einen Schwellwert  $\alpha$  zu kontrollieren und danach die Wahrscheinlichkeit des anderen Fehlers zu minimieren.



**Definition 4.1.2 Gütefunktion**

Es sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen. Für einen Test  $\varphi: \mathbf{R}^n \rightarrow [0, 1]$  heißt die Abbildung  $\beta_\varphi: \Theta \rightarrow [0, 1]$  vermöge

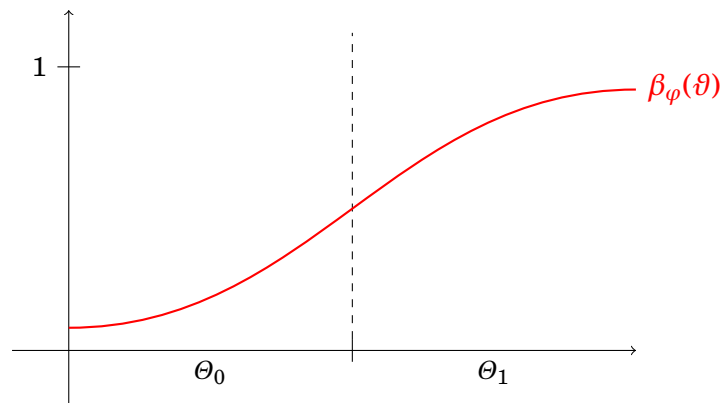
$$\begin{aligned}\beta_\varphi(\vartheta) &:= \mathbf{E}_{P_\vartheta^n} \varphi = \int_{\mathbf{R}^n} \varphi(x) \, dP_\vartheta^n(x) \\ &= \int_{\mathbf{R}^n} \int_{\{0,1\}} \mathbf{1}_{\{1\}}(y) \, dQ_{\varphi(x)}(y) \, dP_\vartheta^n(x)\end{aligned}$$

mit  $Q_{\varphi(x)} := B(1, \varphi(x))$  Gütefunktion von  $\varphi$ .

Sei  $\mu$  das Maß auf  $\mathbf{R}^n \times \{0, 1\}$ , das wie folgt gegeben ist:

$$\mu(A) := \int_{\mathbf{R}^n} \int_{\{0,1\}} \mathbf{1}_A(x, y) \, dQ_{\varphi(x)}(y) \, dP_\vartheta^n(x)$$

Dann beschreibt  $\mu$  die Wahrscheinlichkeit für Kombinationen von Beobachtungen  $x$  und Entscheidungen  $y$  des Tests  $\varphi$ . Die Wahrscheinlichkeiten dafür, welche Fehler  $\varphi$  macht, die wir weiter oben bereits angesprochen haben, werden bezüglich  $\mu$  beschrieben. Die Wahrscheinlichkeit, dass  $\varphi$  Fehler erster Art macht, ist so z. B.  $\mu(\mathbf{R}^n \times \{1\}) = \beta_\varphi(\vartheta)$ . Die eingeschränkte Abbildung  $\beta_\varphi|_{\Theta_1}$  heißt die **Macht** (engl. *power*) von  $\varphi$ .



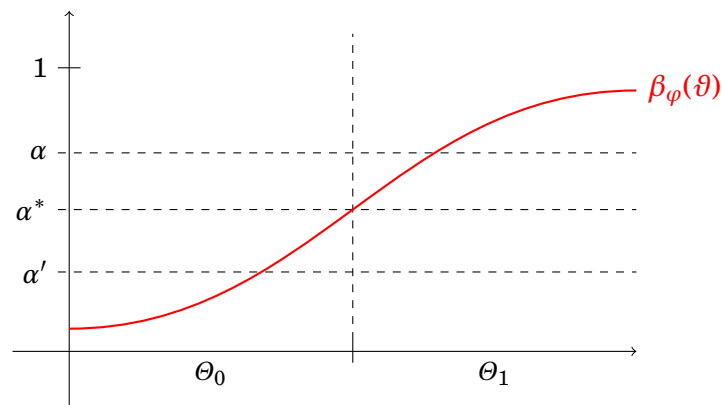
**Abbildung 4.1:** Darstellung einer Gütefunktion.

Wie man in **Abbildung 4.1** sehen kann, sollte die Gütefunktion auf  $\Theta_0$  möglichst klein und auf  $\Theta_1$  möglichst groß sein. Ideal wäre daher ein Test  $\varphi$  mit  $\beta_\varphi = \mathbf{1}_{\Theta_1}$ , dies ist im Allgemeinen aber nicht möglich, zum Beispiel wenn es ein  $B \subset \mathbf{R}$ ,  $\vartheta_0 \in \Theta_0$  und  $\vartheta_1 \in \Theta_1$  mit  $P_{\vartheta_0}(B) > 0$  und  $P_{\vartheta_1}(B) > 0$  gibt, denn für  $x \in B$  muss sich der Test entscheiden.

**Definition 4.1.3 (Unverfälschter) Test zum Niveau  $\alpha$**

Es sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Verteilungen auf  $\mathbf{R}$  und  $\alpha \in [0, 1]$ . Ein Test  $\varphi$  heißt:

- i) **Test zum Niveau  $\alpha$**  genau dann, wenn  $\beta_\varphi(\vartheta) \leq \alpha$  für alle  $\vartheta \in \Theta_0$  gilt, das heißt  $\alpha$  kontrolliert die Wahrscheinlichkeit für Fehler erster Art.
- ii) **unverfälschter Test zum Niveau  $\alpha$**  genau dann, wenn  $\varphi$  ein Test zum Niveau  $\alpha$  ist und zusätzlich  $\beta_\varphi(\vartheta) \geq \alpha$  für alle  $\vartheta \in \Theta_1$  gilt.



**Abbildung 4.2:** Illustration verschiedener  $\alpha$  für den Zusammenhang zwischen Gütefunktionen und (unverfälschter) Tests zum Niveau  $\alpha$ .

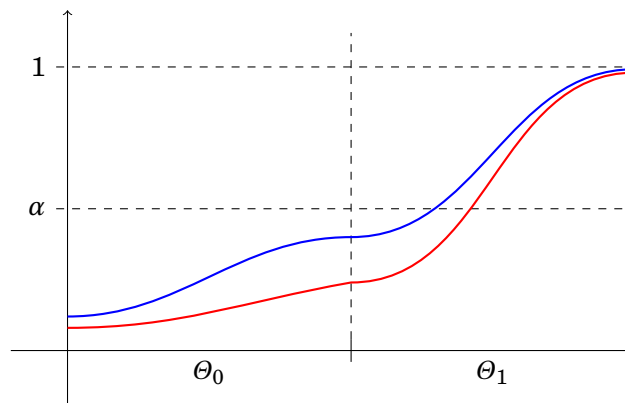
Wir betrachten nun Abbildung 4.2, die eine Erweiterung von Abbildung 4.1 darstellt. Der Test  $\varphi$  ist unverfälscht zum Niveau  $\alpha^*$  und ein nicht unverfälschter Test zum Niveau  $\alpha$ , jedoch kein Test zum Niveau  $\alpha'$ .

**Definition 4.1.4 Konsistenz**

Sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Verteilungen auf  $\mathbf{R}$  und  $\alpha \in [0, 1]$ . Eine Folge von Tests  $\varphi_n : \mathbf{R}^n \rightarrow [0, 1]$  zum Niveau  $\alpha$  heißt **konsistent** genau dann, wenn für alle  $\vartheta \in \Theta_1$  gilt:

$$\lim_{n \rightarrow \infty} \beta_{\varphi_n}(\vartheta) = 1$$

Diese Eigenschaft ist im Allgemeinen schwächer als  $P_\infty^\infty$ -fast sichere Konvergenz von  $\varphi_n(x)$  gegen 1. Eine Illustration findet sich in Abbildung 4.3.



**Abbildung 4.3:** Illustration einer konsistenten Folge statistischer Tests.

#### Definition 4.1.5 Besserer/Gleichmäßig bester Test

Es sei  $(P_{\vartheta})_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$  und  $\alpha \in [0, 1]$ . Ferner sei  $\varphi$  ein Test zum Niveau  $\alpha$ . Dann definieren wir:

- i) Ein Test  $\varphi'$  zum Niveau  $\alpha$  heißt **nicht besser** als  $\varphi$  genau dann, wenn  $\beta_{\varphi}(\vartheta) \geq \beta_{\varphi'}(\vartheta)$  für alle  $\vartheta \in \Theta_1$  gilt.
- ii) Der Test  $\varphi$  heißt **gleichmäßig bester Test** zum Niveau  $\alpha$  genau dann, wenn jeder Test  $\varphi'$  zum Niveau  $\alpha$  nicht besser als  $\varphi$  ist.

Die naheliegende Frage ist, ob es einen gleichmäßig besten Test zum Niveau  $\alpha$  gibt. Diese Frage werden wir im nächsten Abschnitt behandeln.

### Exkursion: $p$ -Werte

Die sogenannten  $p$ -Werte tauchen sehr häufig in der Praxis auf, sind jedoch aus mathematischer Sicht eher problematisch. Aus diesem Grund wollen wir dieses Thema hier nur kurz ansprechen.

Wir nehmen an, die Hypothesen  $H_0$  und  $H_1$  seien einfach, das heißt es gilt  $\Theta = \{\vartheta_0, \vartheta_1\}$ . Weiter sei  $T: \mathbf{R}^n \rightarrow \mathbf{R}$  eine Stichprobenfunktion und  $\varphi: \mathbf{R}^n \rightarrow \{0, 1\}$  ein nicht-randomisierter Test, der für  $k \in \mathbf{R}$  wie folgt gegeben sei:

$$\varphi(x) := \begin{cases} 1 & \text{für } T(x) \geq k \\ 0 & \text{für } T(x) < k \end{cases}$$

Mit dieser Wahl gilt  $\varphi = \mathbf{1}_{[k, \infty)} \circ T$ . Solche Tests werden tatsächlich oft verwendet. Sei nun  $\alpha \in [0, 1]$  mit  $\alpha = P_{\vartheta_0}^n(T \geq k) = P_{\vartheta_0}^n(\varphi = 1) = \beta_{\varphi}(\vartheta_0)$ . In der Praxis wird  $\alpha$  eigentlich vorgegeben und  $k$  dann durch die obige Gleichung bestimmt. Ferner sei  $x \in \mathbf{R}^n$  eine Beobachtung, wir setzen dann  $t := T(x)$ . Dann nennen wir

$$p := P_{\vartheta_0}^n(T \geq t) = P_{\vartheta_0}^n(\{x' \in \mathbf{R}^n : T(x') \geq t\}).$$

den  $p$ -Wert des Tests zu der Beobachtung  $x$ . Insbesondere hängt  $p$  wirklich von der Beobachtung ab und ist daher eigentlich eine Zufallsvariable.

Wir nehmen nun ferner an, dass die Verteilungsfunktion des Bildmaßes von  $P_{\theta_0}^n$  unter  $T$  streng monoton steigend ist. Dann gilt  $p \leq \alpha$  genau dann, wenn

$$\begin{aligned} P_{\theta_0}^n(T \geq t) &\leq P_{\theta_0}^n(T \geq k) \\ \iff t &\geq k \\ \iff T(x) &\geq k \\ \iff \varphi(x) &= 1. \end{aligned}$$

Mit anderen Worten entscheidet sich  $\varphi$  genau dann für  $H_1$ , wenn  $p(x) \leq \alpha$  gilt. Wir wollen nun auf einige *typische Fehler* in Verbindung mit  $p$ -Werten eingehen:

- Man berechnet den  $p$ -Wert anhand der Beobachtung  $x$  und passt das Niveau  $\alpha$  nachträglich an.
- Oft wird der  $p$ -Wert derart fehlinterpretiert, dass man davon ausgeht, er stelle die Wahrscheinlichkeit dafür dar, dass  $H_0$  wahr ist. Dies ist nicht der Fall und es gibt viele Argumente, um dies zu widerlegen. Das einfachste Argument ist jedoch, dass wir für Hypothesen gar keinen Wahrscheinlichkeitsbegriff haben.
- Eine andere gängige Fehlinterpretation ist, dass der  $p$ -Wert die Wahrscheinlichkeit für einen Fehler erster Art angebe. Auch dies ist falsch, da bereits  $\alpha$  diese Wahrscheinlichkeit angibt.

Es gibt noch viele weitere Beispiele für Fehlinterpretationen und Missverständnisse der  $p$ -Werte, auf die wir hier nicht eingehen.

## 4.2 Existenz gleichmäßig bester Tests

In diesem Kapitel wollen wir uns, wie der Titel bereits nahelegt, mit der Frage beschäftigen, wann es gleichmäßig beste Tests zum Niveau  $\alpha$  gibt, die wir in Definition 4.1.5 eingeführt haben. Das Vorgehen ähnelt dabei dem zur Untersuchung der Existenz gleichmäßig bester Schätzer.

Dazu wollen wir zuerst ein fundamentales Lemma behandeln. Da dieses sehr wichtig ist, werden wir es in aller Allgemeinheit formulieren und uns nicht auf  $\mathbf{R}$  beschränken.

### Lemma 4.2.1 Neyman-Pearson-Lemma für einfache Testprobleme

Es sei  $\Theta = \{\vartheta_0, \vartheta_1\}$  und  $\Theta_0 = \{\vartheta_0\}$ . Ferner seien  $P_{\vartheta_0}$  und  $P_{\vartheta_1}$  Wahrscheinlichkeitsmaße auf  $\mathcal{X}$ , sowie  $\mu$  ein  $\sigma$ -endliches Maß auf  $\mathcal{X}$  mit  $P_{\vartheta_i} =: P_i \ll \mu$  für  $i \in \{0, 1\}$ . Ferner seien  $f_0$  und  $f_1$  die Dichten der  $P_i$  bezüglich  $\mu$  und  $\alpha \in [0, 1]$ . Dann gibt es ein  $k \in [0, \infty]$  und ein  $\gamma \in [0, 1]$ , so dass der Test  $\varphi: \mathcal{X} \rightarrow [0, 1]$ , der durch

$$\varphi(x) := \begin{cases} 1 & \text{für } f_1(x) > k f_0(x) \\ \gamma & \text{für } f_1(x) = k f_0(x) \\ 0 & \text{für } f_1(x) < k f_0(x) \end{cases}$$

definiert ist, gleichmäßig bester Test zum Niveau  $\alpha$  ist.

Dabei wollen wir anmerken, dass ein solches Maß  $\mu$ , wie im Lemma gefordert, immer existiert, da man zum Beispiel  $\mu := \frac{1}{2}(P_0 + P_1)$  wählen kann. Ferner haben wir im Lemma für die Definition von  $\varphi$  implizit die Konvention  $0 \cdot \infty = \infty \cdot 0 := 0$  verwendet. In unserer üblichen Vorgehensweise ist  $\mathcal{X} = \mathbf{R}^n$  und  $P_i = P_{\vartheta_i}^n$ . Der Beweis des Lemmas wird uns zudem eine Konstruktionsmethode für die Parameter  $k$  und  $\gamma$  liefern, der Test  $\varphi$  kann also konstruiert werden.

**Beweis:** Wir betrachten zwei Fälle, wobei zunächst  $\alpha = 0$  sein soll. Dann setzen wir  $k := \infty$  und  $\gamma := 0$ , damit folgt  $\varphi(x) > 0$  genau dann, wenn  $\varphi(x) = 1$  ist. Dies ist genau dann der Fall, wenn  $f_1(x) > 0$  und  $f_0(x) = 0$  gilt. Dann gilt

$$\beta_{\varphi}(\vartheta_0) = \int_{\{f_1 > 0\} \cap \{f_0 = 0\}} 1 \, dP_0 = \int_{\{f_1 > 0\} \cap \{f_0 = 0\}} f_0 \, d\mu = 0 = \alpha,$$

sowie

$$\beta_{\varphi}(\vartheta_1) = \int_{\{f_1 > 0\} \cap \{f_0 = 0\}} 1 \, dP_1 = \int_{\{f_1 > 0\} \cap \{f_0 = 0\}} f_1 \, d\mu.$$

Sei nun  $\varphi'$  ein Test zum Niveau  $\alpha = 0$ , dann folgt

$$0 = \beta_{\varphi'}(\vartheta_0) = \int \varphi' \, dP_0 = \int \varphi' f_0 \, d\mu.$$

Da  $\varphi' f_0 \geq 0$  gilt, folgt  $\varphi' f_0 = 0$   $\mu$ -fast überall. Damit folgt

$$\begin{aligned} \beta_{\varphi'}(\vartheta_1) &= \int \varphi' \, dP_1 = \int_{\{f_1 > 0\} \cap \{f_0 = 0\}} \varphi' f_1 \, d\mu + \underbrace{\int_{\{f_1 = 0\} \cap \{f_0 > 0\}} \varphi' f_1 \, d\mu}_{=0} = \int_{\{f_1 > 0\} \cap \{f_0 = 0\}} \varphi' f_1 \, d\mu \\ &\leq \int_{\{f_1 > 0\} \cap \{f_0 = 0\}} f_1 \, d\mu \\ &= \beta_{\varphi}(\vartheta_1). \end{aligned}$$

Damit ist  $\varphi$  wirklich gleichmäßig bester Test. Nun betrachten wir den zweiten Fall mit  $\alpha > 0$ . Dazu definieren wir die Zufallsvariable  $Y : \mathcal{X} \rightarrow [0, \infty)$  vermöge

$$Y(x) := \begin{cases} \frac{f_1(x)}{f_0(x)} & f_0(x) > 0 \\ 0 & \text{sonst} \end{cases}.$$

Ferner sei  $F$  die Verteilungsfunktion von  $Y$  bezüglich  $P_0$ , das heißt  $F(t) := P_0(Y \leq t)$  für  $t \in \mathbf{R}$ . Die erste Möglichkeit ist nun, dass es ein  $k \in [0, \infty)$  mit  $1 - \alpha = F(k)$  gibt. Dann setzen wir  $\gamma := 0$  und es gilt

$$1 - \alpha = P_0(Y \leq k) - \gamma \cdot P_0(Y = k).$$

Die zweite Möglichkeit ist, dass es *kein* solches  $k \in [0, \infty)$  mit  $1 - \alpha = F(k)$  gibt. Da  $F$  monoton steigend und rechtsseitig stetig ist, folgt die Existenz eines  $k \in [0, \infty)$  mit der folgenden Eigenschaft:

$$\sup_{k' < k} F(k') \leq 1 - \alpha < F(k)$$

Da nun  $\sup_{k' < k} F(k') = \sup_{k' < k} P_0(Y \leq k') = P_0(Y < k)$  gilt, folgt  $P_0(Y \leq k) - P_0(Y = k) \leq 1 - \alpha < P_0(Y \leq k)$  und insbesondere  $P_0(Y = k) > 0$ . Nun definieren wir den Parameter  $\gamma$  wie folgt:

$$\gamma := \frac{P_0(Y \leq k) - 1 + \alpha}{P_0(Y = k)}$$

Dies heißt, dass  $1 - \alpha = P_0(Y \leq k) - \gamma P_0(Y = k)$  gilt. Zunächst wollen wir zeigen, dass  $\gamma \in [0, 1]$  gilt. Dazu betrachten wir  $1 - \alpha < P_0(Y \leq k)$ , denn daraus folgt  $\gamma > 0$ . Außerdem ist  $P_0(Y \leq k) - P_0(Y = k) \leq 1 - \alpha$ , woraus wir  $\gamma \leq 1$  erhalten. Damit folgt  $\gamma \in (0, 1] \subset [0, 1]$ . Insgesamt gibt es also immer ein  $k \in [0, \infty)$  und ein  $\gamma \in [0, 1]$ , für die folgende Gleichung gilt:

$$1 - \alpha = P_0(Y \leq k) - \gamma P_0(Y = k) \tag{*}$$

Ein solches Paar  $(k, \gamma)$  benutzen wir nun in der Definition des Tests  $\varphi$ . Es gilt dann

$$\begin{aligned} \beta_{\varphi}(\vartheta_0) &= \int \varphi f_0 \, d\mu = \int_{\{f_0 > 0\}} \varphi f_0 \, d\mu \\ &= \int_{\{f_0 > 0\} \cap \{f_1 > k f_0\}} \varphi f_0 \, d\mu + \int_{\{f_0 > 0\} \cap \{f_1 = k f_0\}} \varphi f_0 \, d\mu \\ &= P_0(Y > k) + \gamma P_0(Y = k) \\ &= 1 - P_0(Y \leq k) + \gamma P_0(Y = k) \\ &\stackrel{(*)}{=} \alpha, \end{aligned}$$

also ist  $\varphi$  gleichmäßig bester Test zum Niveau  $\alpha$ . Sei  $\varphi'$  ein Test zum Niveau  $\alpha$ . Ist  $\varphi(x) > \varphi'(x)$ , so folgt  $\varphi(x) > 0$  und damit  $f_1(x) \geq k f_0(x)$ . Ist hingegen  $\varphi(x) < \varphi'(x)$ , so folgt  $\varphi(x) < 1$  und damit  $f_1(x) \leq k f_0(x)$ . In beiden Fällen, und für  $\varphi(x) = \varphi'(x)$  sowieso, gilt daher

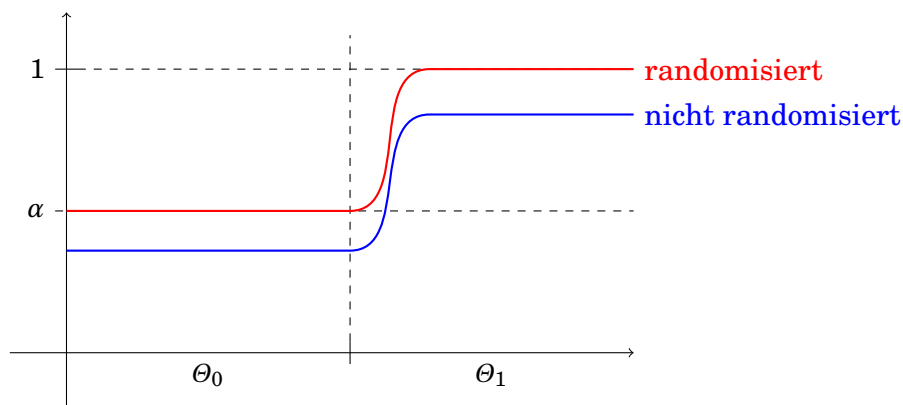
$$(\varphi(x) - \varphi'(x))(f_1(x) - k f_0(x)) \geq 0. \tag{**}$$

Durch Integrieren erhalten wir nun

$$\begin{aligned} 0 &\stackrel{(**)}{\leq} \int (\varphi(x) - \varphi'(x))(f_1(x) - k f_0(x)) \, d\mu(x) \\ &= \int \varphi \, dP_1 - \int \varphi' \, dP_1 - k \underbrace{\int \varphi \, dP_0}_{=\alpha} + k \underbrace{\int \varphi' \, dP_0}_{\leq \alpha} \\ &\leq \beta_\varphi(\vartheta_1) - \beta_{\varphi'}(\vartheta_1). \end{aligned}$$

Damit folgt also  $\beta_{\varphi'}(\vartheta_1) \leq \beta_\varphi(\vartheta_1)$  und wir haben alles gezeigt.  $\square$

Bis jetzt hatten wir die Randomisierung eigentlich nur als mögliche Erweiterung des Test-Begriffs gesehen. Der Beweis des Neyman-Pearson-Lemma 4.2.1 zeigt nun, dass die Randomisierung uns die Möglichkeit gibt, Tests zu konstruieren, die ein beliebiges, vorgegebenes Niveau  $\alpha$  exakt erreichen, wie es in Abbildung 4.4 illustriert wird. Ohne Randomisierung ist dies an den Sprungstellen der Verteilungsfunktion von  $Y$  nicht möglich, wie wir später an Beispielen noch sehen werden.



**Abbildung 4.4:** Mit Hilfe randomisierter Tests kann ein vorgegebenes Niveau exakt erreicht werden.

Als nächstes wollen wir das Neyman-Pearson-Lemma 4.2.1 auf einseitige Tests verallgemeinern.

**Definition 4.2.2 Monotoner Dichtequotient**

Sei  $\Theta \subset \mathbf{R}$ ,  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$  und  $T: \mathbf{R}^n \rightarrow \mathbf{R}$  eine messbare Abbildung. Dann hat die Familie  $(P_\vartheta)_{\vartheta \in \Theta}$  einen **monotonen Dichtequotienten** in  $T$  genau dann, wenn die folgenden Bedingungen erfüllt sind:

- i) Es gilt  $P_\vartheta \neq P_{\vartheta'}$  für alle  $\vartheta \neq \vartheta'$ .
- ii) Es existiert ein  $\sigma$ -endliches Maß  $\mu$  auf  $\mathbf{R}$ , so dass alle  $P_\vartheta$  eine Dichte  $f_\vartheta$  bezüglich  $\mu$  besitzen, es gilt also  $P_\vartheta \ll \mu$ .
- iii) Für alle  $\vartheta_0, \vartheta_1 \in \Theta$  mit  $\vartheta_0 < \vartheta_1$  existiert eine streng monoton wachsende Funktion  $g_{\vartheta_0, \vartheta_1}: \mathbf{R} \rightarrow \mathbf{R}$ , so dass für  $P_{\vartheta_0}^n$ - und  $P_{\vartheta_1}^n$ -fast alle  $x \in \mathbf{R}^n$  gilt:

$$\frac{\otimes_1^n f_{\vartheta_1}(x)}{\otimes_1^n f_{\vartheta_0}(x)} = g_{\vartheta_0, \vartheta_1}(T(x))$$

Wir hatten bereits im Neyman-Pearson-Lemma 4.2.1 gesehen, dass sich optimale Tests für  $H_1 := \{\vartheta_1\}$  und nicht für  $H_0 := \{\vartheta_0\}$  entscheiden, falls

$$\frac{\otimes_1^n f_{\vartheta_1}(x)}{\otimes_1^n f_{\vartheta_0}(x)} \geq f^*$$

für ein geeignetes  $f^* \in \mathbf{R}$  gilt. Diese Ordnung wird durch  $g_{\vartheta_0, \vartheta_1}$  beibehalten. Die Abbildung  $T$  werden wir zur Konstruktion optimaler Tests benutzen, da sie von  $\vartheta_0$  und  $\vartheta_1$  unabhängig ist.

**Satz 4.2.3**

Sei  $\Theta \subset \mathbf{R}$ ,  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathbf{R}$  mit monotonem Dichtequotienten in  $T$  und  $\vartheta_0 \in \Theta$ . Wir setzen  $\Theta_0 := \{\vartheta \in \Theta : \vartheta \leq \vartheta_0\}$  und  $\Theta_1 := \{\vartheta \in \Theta : \vartheta > \vartheta_0\}$ . Für  $\alpha \in (0, 1)$  gibt es dann ein  $k \in [0, \infty)$  und ein  $\gamma \in [0, 1]$ , so dass der Test  $\varphi$  vermöge

$$\varphi(x) := \begin{cases} 1 & \text{für } T(x) > k \\ \gamma & \text{für } T(x) = k \\ 0 & \text{für } T(x) < k \end{cases}$$

ein gleichmäßig bester Test zum Niveau  $\alpha$  ist.

Tatsächlich gilt diese Aussage für jedes  $k \in [0, \infty)$  und jedes  $\gamma \in [0, 1]$ , welche die folgende Gleichung erfüllen:

$$1 - \alpha = P_{\vartheta_0}^n(T \leq k) - \gamma P_{\vartheta_0}^n(T = k) \tag{*}$$

Falls  $P_{\vartheta_0}^n(T = k) = 0$  gilt, so ist jedes  $\gamma$  erlaubt und  $k$  ist dann das  $1 - \alpha$ -Quantil der Verteilung von  $T$  bezüglich  $P_{\vartheta_0}^n$ .



Mit anderen Worten lassen sich optimale Tests häufig mit Hilfe von Quantilen bestimmen.

**Beweis:** Die Existenz von  $k$  und  $\gamma$ , die (\*) erfüllen, lässt sich völlig analog zum Beweis des Neyman-Pearson-Lemmas 4.2.1 führen, wenn man  $Y$  durch  $T$  ersetzt. Nun gilt für das Niveau von  $\varphi$  für  $\vartheta_0$ :

$$\begin{aligned}\beta_\varphi(\vartheta_0) &= \int \varphi \, dP_{\vartheta_0}^n = \int_{\{T>k\}} \varphi \, dP_{\vartheta_0}^n + \int_{\{T=k\}} \varphi \, dP_{\vartheta_0}^n + \int_{\{T<k\}} \varphi \, dP_{\vartheta_0}^n \\ &= P_{\vartheta_0}^n(T > k) + \gamma P_{\vartheta_0}^n(T = k) \\ &= 1 - P_{\vartheta_0}^n(T \leq k) + \gamma P_{\vartheta_0}^n(T = k) \\ &\stackrel{(*)}{=} \alpha\end{aligned}$$

Nun wollen wir das Niveau für  $\Theta_0$  betrachten. Sei  $\vartheta'_0, \vartheta'_1 \in \Theta$  mit  $\vartheta'_0 < \vartheta'_1$ . Ferner sei  $\varphi'$  ein Test mit  $\beta_{\varphi'}(\vartheta'_0) \leq \beta_{\varphi'}(\vartheta'_1)$ . Sei außerdem  $g_{\vartheta'_0, \vartheta'_1}$  gemäß der Definition 4.2.2 gegeben und setze  $f^* := g_{\vartheta'_0, \vartheta'_1}(k)$ . Ist  $\varphi(x) > \varphi'(x)$ , so folgt  $\varphi(x) > 0$  und damit  $T(x) \geq k$ , woraus wiederum  $\frac{\otimes_1^n f_{\vartheta'_1}(x)}{\otimes_1^n f_{\vartheta'_0}(x)} = g_{\vartheta'_0, \vartheta'_1}(T(x)) \geq g_{\vartheta'_0, \vartheta'_1}(k) = f^*$  folgt, wobei die Monotonie von  $g$  verwendet wurde. Ist hingegen  $\varphi(x) < \varphi'(x)$ , so folgt  $\varphi(x) < 1$ , damit  $T(x) \leq k$  und damit  $\frac{\otimes_1^n f_{\vartheta'_1}(x)}{\otimes_1^n f_{\vartheta'_0}(x)} \leq f^*$ . Insgesamt gilt also

$$(\varphi(x) - \varphi'(x)) \left( \otimes_1^n f_{\vartheta'_1}(x) - f^* \otimes_1^n f_{\vartheta'_0}(x) \right) \geq 0,$$

wobei dies auch für  $\varphi(x) = \varphi'(x)$ , also *immer* gilt. Durch Integrieren erhalten wir nun:

$$\begin{aligned}0 &\leq \int (\varphi(x) - \varphi'(x)) \left( \otimes_1^n f_{\vartheta'_1} - f^* \otimes_1^n f_{\vartheta'_0} \right) d\mu^n \\ &= \int \varphi \, dP_{\vartheta'_1}^n - \int \varphi' \, dP_{\vartheta'_1}^n - f^* \int \varphi \, dP_{\vartheta'_0}^n + f^* \int \varphi' \, dP_{\vartheta'_0}^n \\ &= \beta_\varphi(\vartheta'_1) - \beta_{\varphi'}(\vartheta'_1) - \underbrace{f^* \beta_\varphi(\vartheta'_0) + f^* \beta_{\varphi'}(\vartheta'_0)}_{\leq 0} \\ &\leq \beta_\varphi(\vartheta'_1) - \beta_{\varphi'}(\vartheta'_1)\end{aligned}$$

Damit gilt also  $\beta_{\varphi'}(\vartheta'_1) \leq \beta_\varphi(\vartheta'_1)$ . Daraus folgt nun, dass  $\varphi$  ein gleichmäßig bester Test zum Niveau  $\beta_\varphi(\vartheta'_0)$  für die Hypothesen  $H'_0 := \{\vartheta'_0\}$  und  $H'_1 := \{\vartheta'_1\}$  ist. Wir betrachten nun den Test  $\varphi'(x) := \beta_{\varphi'}(\vartheta'_0)$ , der im ersten Augenblick nur bedingt sinnvoll erscheint, da er unabhängig von  $x \in \mathbf{R}^n$  ist. Er erfüllt  $\beta_{\varphi'}(\vartheta'_0) = \beta_\varphi(\vartheta'_0)$  und es gilt wegen (\*):

$$\begin{aligned}\beta_\varphi(\vartheta'_1) &\geq \beta_{\varphi'}(\vartheta'_1) \\ &= \int \varphi' \, dP_{\vartheta'_1}^n = \beta_\varphi(\vartheta'_0)\end{aligned}$$

Damit ist  $\vartheta \mapsto \beta_\varphi(\vartheta)$  eine monoton wachsende Abbildung. Für  $\vartheta'_0 \in \Theta_0$  folgt nun insbesondere  $\beta_\varphi(\vartheta'_0) \leq \beta_\varphi(\vartheta_0) = \alpha$ . Damit ist  $\varphi$  also ein Test zum Niveau  $\alpha$ . Ferner hat die Rechnung um

(\*) gezeigt, dass für  $\vartheta_1 > \vartheta_0$  und für einen Test  $\varphi'$  mit  $\beta_{\varphi'}(\vartheta'_0) \leq \alpha$  für alle  $\vartheta'_0 \in \Theta_0$ , woraus insbesondere  $\beta_{\varphi'}(\vartheta_0) \leq \alpha = \beta_{\varphi}(\vartheta_0)$  folgt, gilt, dass  $\beta_{\varphi}(\vartheta_1) \geq \beta_{\varphi'}(\vartheta_1)$  ist. Nun ist  $\vartheta_1 > \vartheta_0$  genau dann der Fall, wenn  $\vartheta_1 \in \Theta_1$  gilt. Damit ist  $\varphi$  tatsächlich gleichmäßig bester Test zum Niveau  $\alpha$ .  $\square$

Als nächstes beschäftigen wir uns mit der Frage, ob und welche Familien überhaupt einen monotonen Dichtequotienten besitzen.

**Definition 4.2.4 Einparametrische Exponentialfamilie**

Es seien  $T: \mathbf{R}^n \rightarrow \mathbf{R}$ ,  $h: \mathbf{R}^n \rightarrow [0, \infty)$ ,  $C: \Theta \rightarrow (0, \infty)$  und  $\xi: \Theta \rightarrow \mathbf{R}$ , wobei  $T$  und  $h$  zudem messbar und  $\Theta \subset \mathbf{R}$  sein sollen. Ferner sei  $\mu$  ein endliches Maß auf  $\mathbf{R}^n$ . Dann definieren wir für  $\vartheta \in \Theta$  und  $x \in \mathbf{R}^n$

$$f_{\vartheta}(x) := C(\vartheta)h(x) \cdot e^{\xi(\vartheta)T(x)}.$$

Ist  $C$  so gewählt, dass  $\int f_{\vartheta} d\mu = 1$  für alle  $\vartheta \in \Theta$  gilt, so heißt die Familie  $(P_{\vartheta})_{\vartheta \in \Theta}$  mit  $P_{\vartheta} = f_{\vartheta}d\mu$  einparametrische Exponentialfamilie.

Wir werden im Folgenden die Bezeichnungen der Abbildungen in der Definition 4.2.4 verwenden, ohne sie jedesmal wieder explizit zu definieren.

**Lemma 4.2.5**

Sei  $(P_{\vartheta})_{\vartheta \in \Theta}$  eine einparametrische Exponentialfamilie und  $\xi$  streng monoton wachsend. Dann besitzt  $(P_{\vartheta})_{\vartheta \in \Theta}$  einen monotonen Dichtequotienten in  $x \mapsto \frac{1}{n} \sum_{i=1}^n T(x_i)$ .

**Beweis:** Für  $\vartheta_0 < \vartheta_1$  gilt

$$\frac{\otimes_1^n f_{\vartheta_1}(x)}{\otimes_1^n f_{\vartheta_0}(x)} = \frac{C^n(\vartheta_1)}{C^n(\vartheta_0)} \cdot \frac{\otimes_1^n h(x)}{\otimes_1^n h(x)} \cdot \frac{\prod_1^n e^{\xi(\vartheta_1)T(x_i)}}{\prod_1^n e^{\xi(\vartheta_0)T(x_i)}} = \frac{C^n(\vartheta_1)}{C^n(\vartheta_0)} \exp\left(\sum_{i=1}^n (\xi(\vartheta_1) - \xi(\vartheta_0))T(x_i)\right).$$

Wir setzen nun  $g_{\vartheta_0, \vartheta_1}(y) := \frac{C^n(\vartheta_1)}{C^n(\vartheta_0)} \exp(n(\xi(\vartheta_1) - \xi(\vartheta_0)) \cdot y)$ . Für  $y = \frac{1}{n} \sum_{i=1}^n T(x_i)$  gilt dann

$$\frac{\otimes_1^n f_{\vartheta_1}(x)}{\otimes_1^n f_{\vartheta_0}(x)} = g_{\vartheta_0, \vartheta_1}(y).$$

Da  $\xi(\vartheta_1) > \xi(\vartheta_0)$  gilt ist  $g_{\vartheta_0, \vartheta_1}$  auch wirklich streng monoton wachsend.  $\square$

Für einparametrische Exponentialfamilien gibt es viele Beispiele, von denen wir einige später noch kennenlernen werden. Zunächst werden wir uns mit zweiseitigen Tests beschäftigen, für welche wir zwei verschiedene Ansätze verfolgen werden. Der erste Ansatz ist ein theoretischer, der zweite ein pragmatischer Ansatz.

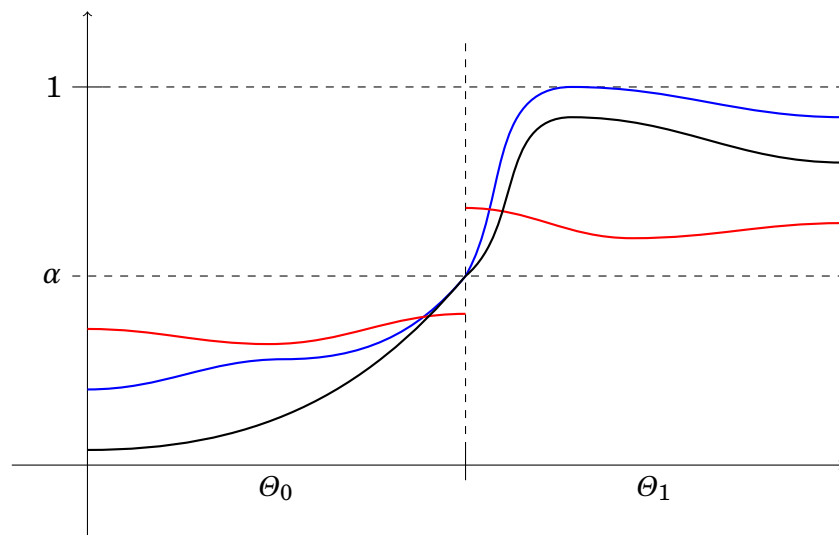
Im Folgenden schreiben wir  $\Phi := \Phi_1 = \{\varphi: \mathcal{X} \rightarrow [0, 1] \mid \varphi \text{ messbar}\}$  für die Menge aller Tests, sowie  $\Phi_\alpha := \{\varphi: \mathcal{X} \rightarrow [0, 1] \mid \varphi \text{ ist Test zum Niveau } \alpha\}$  für  $\alpha \in [0, 1]$ . Wir wollen im Folgenden ferner allgemeine Räume  $\Theta$  und allgemeine Familien  $(P_\vartheta)_{\vartheta \in \Theta}$  zulassen. Zudem sei  $\Theta_0, \Theta_1$  eine beliebige Partition von  $\Theta$ .

**Definition 4.2.6 Maxmin-Test**

Sei  $\alpha \in [0, 1]$ . Ein Test  $\varphi^* \in \Phi_\alpha$  heißt **Maxmin-Test zum Niveau  $\alpha$**  genau dann, wenn gilt:

$$\inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) = \sup_{\varphi \in \Phi_\alpha} \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta)$$

Der Ausdruck  $\inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta)$  beschreibt das schlechtmögliche Verhalten für Fehler zweiter Art eines Tests  $\varphi$ . Eine Interpretation findet sich in Abbildung 4.5. Es gilt zu beachten, dass aus der Tatsache, dass  $\varphi^*$  ein Maxmin-Test zum Niveau  $\alpha$  ist, *nicht* folgt, dass  $\varphi^*$  gleichmäßig bester Test ist.



**Abbildung 4.5:** Illustration für das schlechtmögliche Verhalten Fehler zweiter Art. Der rote Test ist besser als der schwarze Test, der wiederum gleich gut wie der blaue Test ist.

**Lemma 4.2.7**

Für  $\alpha \in [0, 1]$  gelten die folgenden Aussagen:

- i) Ist  $\varphi^* \in \Phi_\alpha$  gleichmäßig bester Test zum Niveau  $\alpha$ , so folgt, dass  $\varphi^*$  unverfälscht ist (vgl. Definition 4.1.3).
- ii) Ist  $\varphi^* \in \Phi_\alpha$  gleichmäßig bester Test zum Niveau  $\alpha$ , so folgt, dass  $\varphi^*$  ein Maxmin-Test zum Niveau  $\alpha$  ist.

iii) Ist  $\varphi^* \in \Phi_\alpha$  ein Maxmin-Test zum Niveau  $\alpha$ , so folgt, dass  $\varphi^*$  unverfälscht ist (vgl. Definition 4.1.3).

**Beweis:** Die erste Aussage i) folgt unmittelbar aus ii) und iii). Für iii) betrachten wir  $\varphi(x) := \alpha$  für alle  $x \in \mathcal{X}$ . Dann gilt

$$\inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) = \sup_{\varphi' \in \Phi_\alpha} \inf_{\vartheta \in \Theta_1} \beta_{\varphi'}(\vartheta) \geq \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta) = \alpha.$$

nach Konstruktion von  $\varphi$ . Für ii) führen wir einen Widerspruchsbeweis. Angenommen,  $\varphi^*$  ist kein Maxmin-Test, dann gilt

$$\inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) < \sup_{\varphi \in \Phi_\alpha} \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta).$$

Dann muss also ein  $\varphi \in \Phi_\alpha$  mit  $\inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) < \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta)$  existieren. Da aber  $\varphi^*$  nach Voraussetzung gleichmäßig bester Test ist, gilt  $\beta_{\varphi^*}(\vartheta) \geq \beta_\varphi(\vartheta)$  für alle  $\vartheta \in \Theta_1$ . Dann folgt  $\beta_{\varphi^*}(\vartheta) \geq \inf_{\vartheta' \in \Theta_1} \beta_\varphi(\vartheta')$  für alle  $\vartheta \in \Theta_1$  und daraus wiederum  $\inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) \geq \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta)$ , dies stellt jedoch einen Widerspruch dar. Damit ist alles gezeigt.  $\square$

#### Satz 4.2.8

Sei  $\mu$  ein  $\sigma$ -endliches Maß auf  $(\mathcal{X}, \mathcal{B})$ . Dann existiert zu jeder Folge  $(\varphi_n) \subset \Phi$  eine Teilfolge  $(\varphi_{n_k}) \subset \Phi$  und ein  $\varphi^* \in \Phi$ , so dass für alle  $f \in \mathcal{L}_1(\mu)$  gilt:

$$\lim_{k \rightarrow \infty} \int \varphi_{n_k} f \, d\mu = \int \varphi^* f \, d\mu$$

Wir sagen daher auch, dass  $\Phi$  schwach folgenkompakt ist. Den Beweis werden wir hier allerdings nicht führen, er lässt sich zum Beispiel in [Witt85, Satz 2.14, S. 205] finden.

Die Menge  $\Phi$  ist eine beschränkte und abgeschlossene Menge in

$$\mathcal{L}_\infty(\mu) := \{f : \mathcal{X} \rightarrow \mathbf{R} \mid f \text{ messbar und } \mu\text{-fast überall beschränkt}\}.$$

Da im Allgemeinen  $\dim \mathcal{L}_\infty(\mu) = \infty$  gilt, können wir den Satz von Heine-Borel nicht anwenden,  $\Phi$  ist also nicht kompakt in  $\mathcal{L}_\infty(\mu)$ . Jedes  $g \in \mathcal{L}_\infty(\mu)$  liefert eine stetige, lineare und  $\mathbf{R}$ -wertige Funktion auf  $\mathcal{L}_1(\mu)$  vermöge  $\mathcal{L}_1(\mu) \ni f \mapsto \int f g \, d\mu$ . Man kann zeigen, dass dies alle stetigen, linearen und  $\mathbf{R}$ -wertigen Funktionen liefert. Mit anderen Worten heißt dies, dass  $\mathcal{L}_\infty(\mu)$  der Dualraum von  $\mathcal{L}_1(\mu)$  ist. Umgekehrt gilt, dass jedes  $f \in \mathcal{L}_1(\mu)$  eine stetige, lineare und  $\mathbf{R}$ -wertige Funktion auf  $\mathcal{L}_\infty(\mu)$  liefert vermöge  $\mathcal{L}_\infty(\mu) \ni g \mapsto \int f g \, d\mu$ . Dies sind im Allgemeinen jedoch nicht alle solche Funktionen.

Sei nun  $w^*$  die kleinste Topologie auf  $\mathcal{L}_\infty(\mu)$  bezüglich welcher alle wie eben definierte Funktionen noch stetig sind, die sogenannte Schwach\*-Topologie (engl. *weak*). Man kann zeigen, dass jede beschränkte,  $w^*$ -abgeschlossene Menge in  $\mathcal{L}_\infty(\mu)$   $w^*$ -kompakt ist. Da  $\Phi$

$w^*$ -abgeschlossen ist, ist  $\Phi$  auch  $w^*$ -kompakt. Die  $w^*$ -Topologie besitzt im Allgemeinen jedoch keine abzählbare Umgebungsbasis und daher ist Satz 4.2.8 eine echte Verschärfung der  $w^*$ -Kompaktheit.

**Satz 4.2.9**

Es sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathcal{X}$  und  $\Theta \neq \emptyset$  eine beliebige Menge. Ferner sei  $\mu$  ein  $\sigma$ -endliches Maß auf  $\mathcal{X}$  mit  $P_\vartheta \ll \mu$  und  $\Theta = \Theta_0 \sqcup \Theta_1$  eine Partition. Dann gibt es zu jedem  $\alpha \in [0, 1]$  einen Maxmin-Test  $\varphi^*$  zum Niveau  $\alpha$ .

**Beweis:** Wir setzen zunächst  $s := \sup_{\varphi \in \Phi_\alpha} \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta)$ . Dann existiert eine Folge  $(\varphi_n) \subset \Phi_\alpha$  mit  $\inf_{\vartheta \in \Theta_1} \beta_{\varphi_n}(\vartheta) \rightarrow s$ . Aus Satz 4.2.8 folgt nun die Existenz einer Teilfolge  $(\varphi_{n_k})$  und eines Tests  $\varphi^* \in \Phi$  mit  $\int \varphi_{n_k} d\mu \rightarrow \int \varphi^* f d\mu$  für alle  $f \in \mathcal{L}_1(\mu)$ . Wir betrachten nun die Dichte  $f_\vartheta$  von  $P_\vartheta$  bezüglich  $\mu$ . Klar ist, dass  $f_\vartheta \in \mathcal{L}_1(\mu)$  gilt, daher gilt

$$\beta_{\varphi^*}(\vartheta) = \int \varphi^* f_\vartheta d\mu = \lim_{k \rightarrow \infty} \int \varphi_{n_k} f_\vartheta d\mu = \lim_{k \rightarrow \infty} \beta_{\varphi_{n_k}}(\vartheta) \quad (*)$$

für alle  $\vartheta \in \Theta$ . Für  $\vartheta \in \Theta_0$  folgt wegen  $\beta_{\varphi_{n_k}}(\vartheta) \leq \alpha$ , dass  $\beta_{\varphi^*}(\vartheta) \leq \alpha$  ist, das heißt  $\varphi^*$  ist ein Test zum Niveau  $\alpha$ . Für  $\vartheta \in \Theta_1$  folgt ferner

$$s = \lim_{n \rightarrow \infty} \inf_{\vartheta' \in \Theta_1} \beta_{\varphi_{n_k}}(\vartheta') \leq \lim_{k \rightarrow \infty} \beta_{\varphi_{n_k}}(\vartheta) \stackrel{(*)}{=} \beta_{\varphi^*}(\vartheta),$$

daher folgt  $\inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) \geq s$ . Da  $\varphi^*$  ein Test zum Niveau  $\alpha$  ist, gilt natürlich auch

$$\inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) \leq \sup_{\varphi \in \Phi_\alpha} \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta) = s. \quad \square$$

**Korollar 4.2.10**

Sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Wahrscheinlichkeitsmaßen auf  $\mathcal{X}$  und  $\mu$  ein  $\sigma$ -endliches Maß auf  $\mathcal{X}$  mit  $P_\vartheta \ll \mu$  für alle  $\vartheta \in \Theta$ . Ferner sei  $\Theta_1 := \{\vartheta_1\}$  und  $\Theta_0 := \Theta \setminus \{\vartheta_1\}$  für ein  $\vartheta_1 \in \Theta$ . Dann gibt es zu jedem  $\alpha \in [0, 1]$  einen gleichmäßig besten Test  $\varphi^*$  zum Niveau  $\alpha$ .

**Beweis:** Sei  $\varphi^*$  ein Maxmin-Test zum Niveau  $\alpha$  nach Satz 4.2.9. Dann gilt für  $\vartheta_1$ :

$$\beta_{\varphi^*}(\vartheta_1) = \inf_{\vartheta \in \Theta_1} \beta_{\varphi^*}(\vartheta) = \sup_{\varphi \in \Phi_\alpha} \inf_{\vartheta \in \Theta_1} \beta_\varphi(\vartheta) = \sup_{\varphi \in \Phi_\alpha} \beta_\varphi(\vartheta_1)$$

Damit folgt, dass  $\varphi^*$  gleichmäßig bester Test zum Niveau  $\alpha$  ist. □

Im Allgemeinen gibt es jedoch keinen gleichmäßig besten Test zum Niveau  $\alpha$  für  $\Theta_0 := \{\vartheta_0\}$  gegen  $\Theta_1 := \Theta \setminus \{\vartheta_0\}$ . Wir werden dies noch näher betrachten.

Als nächstes wollen wir verallgemeinerte zweiseitige Tests für Exponentialfunktionen betrachten.

**Satz 4.2.11**

Sei  $\Theta \subset \mathbf{R}$  ein Intervall  $(P_\vartheta)_{\vartheta \in \Theta}$  eine einparametrische Exponentialfamilie auf  $\mathbf{R}^n$  (vgl. Definition 4.2.4) und  $\Theta_1 := \{\vartheta \in \Theta : \vartheta_1 < \vartheta < \vartheta_2\}$ , sowie  $\Theta_0 := \Theta \setminus \Theta_1$  für beliebige  $\vartheta_1 < \vartheta_2$  aus dem Inneren  $\Theta^\circ$  von  $\Theta$ .

Falls  $\xi$  streng monoton wachsend und stetig ist (womit auch  $\xi^{-1}$  stetig ist), gibt es zu jedem  $\alpha \in (0, 1)$  einen gleichmäßig besten Test  $\varphi^*$  zum Niveau  $\alpha$  vermöge

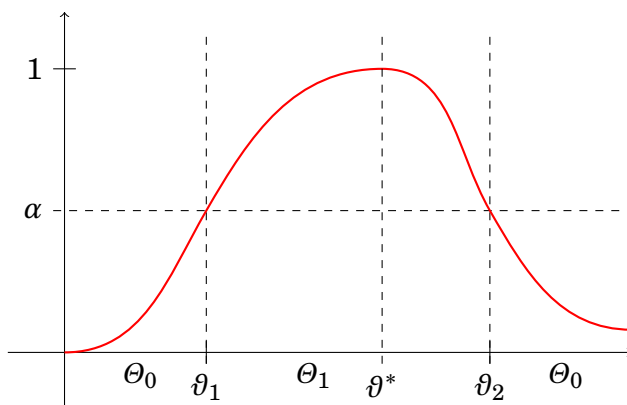
$$\varphi^*(x) := \begin{cases} 1 & \text{falls } T(x) \in (c_1, c_2) \\ \delta_i & \text{falls } T(x) = c_i \text{ für } i = 1, 2, \\ 0 & \text{sonst} \end{cases}$$

wobei  $c_i$  und  $\delta_i$  für  $i = 1, 2$  durch die Gleichungen

$$\beta_{\varphi^*}(\vartheta_i) = P_{\vartheta_i}(T \in (c_1, c_2)) + \delta_1 P_{\vartheta_i}(T = c_1) + \delta_2 P_{\vartheta_i}(T = c_2) = \alpha$$

bestimmt sind.

Der Beweis des Satzes findet sich in [Witt85, Korollar 2.71, S. 261]. Mit einigen wenigen zusätzlichen Annahmen, die nahezu immer erfüllt sind, gibt es zudem ein  $\vartheta^* \in \Theta_1$ , so dass  $\vartheta \mapsto \beta_{\varphi^*}(\vartheta)$  streng monoton wachsend für  $\vartheta < \vartheta^*$  und streng monoton fallend für  $\vartheta > \vartheta^*$  ist. Dies wird in Abbildung 4.6 veranschaulicht.



**Abbildung 4.6:** Strenge Monotonie der Gütefunktion in der Erweiterung von Satz 4.2.11.

Wir wollen uns nun anschauen, was bei Tests passiert, bei denen die Rollen von  $\Theta_0$  und  $\Theta_1$  vertauscht sind. Dazu verweisen zunächst wir auf die Beobachtung in [Witt85, S. 256], die im Wesentlichen folgendes besagt: Selbst bei einparametrischen Exponentialfamilien aus Satz 4.2.11 gibt es keine gleichmäßig besten Tests zum Niveau  $\alpha \in (0, 1)$ , wenn die Hypothesen  $\Theta_0 := [\vartheta_1, \vartheta_2]$  und  $\Theta_1 := \Theta \setminus \Theta_0$  betrachtet werden. Wir müssen uns daher auf eine kleinere Familie einschränken.

**Definition 4.2.12 Gleichmäßig bester Test unter allen unverfälschten Tests**

Sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine Familie von Verteilungen und  $\Phi'_\alpha := \{\varphi \in \Phi_\alpha \mid \varphi \text{ ist unverfälscht}\}$ . Dann heißt ein Test  $\varphi^* \in \Phi'_\alpha$  gleichmäßig bester Test zum Niveau  $\alpha$  **unter allen unverfälschten Tests** zum Niveau  $\alpha$  genau dann, wenn für alle  $\varphi \in \Phi'_\alpha$  und alle  $\vartheta \in \Theta_1$  gilt:

$$\beta_{\varphi^*}(\vartheta) \geq \beta_\varphi(\vartheta)$$



In der Fachliteratur wird  $\varphi^*$  oft auch als **gleichmäßig bester unverfälschter Test** bezeichnet. Hierbei ist zu beachten, dass diese Bezeichnung zwar üblich ist, aber eine andere Bedeutung suggeriert. Es ist also darauf zu achten, dass damit ein Test im Sinne der Definition 4.2.12 gemeint ist. Wir werden diese Bezeichnung ebenfalls verwenden.

Zunächst ist mit Lemma 4.2.7 klar, dass ein gleichmäßig bester Test  $\varphi^*$  zum Niveau  $\alpha$  auch ein gleichmäßig bester unverfälschter Test zum Niveau  $\alpha$  ist. Die Umkehrung gilt im Allgemeinen jedoch nicht. Dies folgt aus den bisherigen Beobachtungen und dem folgenden Satz 4.2.13.

**Satz 4.2.13**

Sei  $(P_\vartheta)_{\vartheta \in \Theta}$  eine einparametrische Exponentialfamilie wie in Satz 4.2.11, insbesondere ist  $\xi$  strikt monoton wachsend und stetig. Für  $\vartheta_1, \vartheta_2 \in \Theta^\circ$  mit  $\vartheta_1 < \vartheta_2$  sei  $\Theta_0 := \{\vartheta \in \Theta : \vartheta \in [\vartheta_1, \vartheta_2]\}$  und  $\Theta_1 := \{\vartheta \in \Theta : \vartheta \notin [\vartheta_1, \vartheta_2]\}$ . Dann gibt es zu jedem  $\alpha \in (0, 1)$  einen gleichmäßig besten unverfälschten Test  $\varphi^*$  zum Niveau  $\alpha$  vermöge

$$\varphi^*(x) := \begin{cases} 1 & \text{falls } T(x) \notin (c_1, c_2) \\ \delta_i & \text{falls } T(x) = c_i \text{ für } i = 1, 2, \\ 0 & \text{sonst} \end{cases}$$

wobei  $c_i$  und  $\delta_i$  für  $i = 1, 2$  durch die Gleichungen

$$\beta_{\varphi^*}(\vartheta_i) = P_{\vartheta_i}(T \notin (c_1, c_2)) + \delta_1 P_{\vartheta_i}(T = c_1) + \delta_2 P_{\vartheta_i}(T = c_2) = \alpha$$

gegeben sind.

Der Beweis findet sich in [Witt85, Satz 2.69, S. 257].

**Beispiel 4.2.14** Normalverteilung

Sei  $P_\mu = \mathcal{N}^n(\mu, \sigma^2)$  mit bekannter Varianz  $\sigma^2 > 0$  und unbekanntem Erwartungswert  $\mu \in \mathbf{R}$ . Für die Dichte von  $P_\mu$  gilt

$$\begin{aligned} f_\mu(x) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i^2 - 2x_i\mu + \mu^2)\right) \\ &= \underbrace{(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{n\mu^2}{2\sigma^2}\right)}_{=:C(\mu)} \underbrace{\exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2\right) \exp\left(\frac{\mu}{\sigma^2} \sum_{i=1}^n x_i\right)}_{=:h(x)}. \end{aligned}$$

Ferner setzen wir  $T(x) := \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n x_i$  und  $\xi(\mu) := \frac{\sqrt{n}\mu}{\sigma}$ . Daraus folgt, dass  $(P_\mu)$  eine einparametrische Exponentialfamilie und  $\xi$  streng monoton wachsend und stetig ist. Damit sind die Voraussetzungen der Sätze 4.2.11 und 4.2.13 erfüllt. Wir wollen exemplarisch Satz 4.2.11 anwenden und betrachten daher das Testproblem  $\Theta_0 := \mathbf{R} \setminus (\mu_1, \mu_2)$  und  $\Theta_1 := (\mu_1, \mu_2)$  für  $\mu_1, \mu_2 \in \mathbf{R}$  mit  $\mu_1 < \mu_2$ . Da  $P_\mu(T = c) = 0$  für alle  $c \in \mathbf{R}$  und  $\mu \in \mathbf{R}$  gilt, vereinfachen sich die zu lösenden Gleichungen aus Satz 4.2.11 zu  $P_{\mu_i}(T \in (c_1, c_2)) = \alpha$  für  $i = 1, 2$ . Der resultierende, gleichmäßig beste Test ist dann für  $x \in \mathbf{R}$  gegeben durch

$$\varphi^*(x) := \mathbf{1}_{(c_1, c_2)}(T(x)).$$

Wir müssen daher nur noch das obige Gleichungssystem lösen. Es gilt, dass  $T$  normalverteilt mit  $\mathbf{E}T = \frac{n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}\mu}{\sigma}$  und  $\text{Var}T = 1$ . Daraus folgt, dass  $T \sim \mathcal{N}\left(\frac{\sqrt{n}\mu}{\sigma}, 1\right)$  ist. Damit gilt

$$\begin{aligned} P_\mu(T \in (c_1, c_2)) &= \frac{1}{\sqrt{2\pi}} \int_{c_1}^{c_2} \exp\left(-\frac{\left(t - \frac{\sqrt{n}\mu}{\sigma}\right)^2}{2}\right) dt = \frac{1}{\sqrt{2\pi}} \int_{c_1 + \frac{\sqrt{n}\mu}{\sigma}}^{c_2 + \frac{\sqrt{n}\mu}{\sigma}} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \Phi\left(c_2 + \frac{\sqrt{n}\mu}{\sigma}\right) - \Phi\left(c_1 + \frac{\sqrt{n}\mu}{\sigma}\right), \end{aligned}$$

wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung  $\mathcal{N}(0, 1)$  ist. Damit ist unser Gleichungssystem für  $i = 1, 2$  gegeben durch

$$\Phi\left(c_2 + \frac{\sqrt{n}\mu}{\sigma}\right) - \Phi\left(c_1 + \frac{\sqrt{n}\mu}{\sigma}\right) = \alpha. \tag{*}$$

Wir betrachten zunächst den Fall  $\mu_1 = -\mu_2 =: \mu$ . Sei  $c > 0$  mit

$$\Phi(c + \tilde{\mu}) - \Phi(-c + \tilde{\mu}) = \alpha, \tag{**}$$

wobei  $\tilde{\mu} =: \frac{\sqrt{n}\mu}{\sigma}$  ist. Dann gilt

$$\Phi(c - \tilde{\mu}) - \Phi(-c - \tilde{\mu}) = 1 - \Phi(-c + \tilde{\mu}) - 1 + \Phi(c + \tilde{\mu}) = \Phi(c + \tilde{\mu}) - \Phi(-c + \tilde{\mu}) = \alpha.$$

Für  $c_2 := c$  und  $c_1 := -c$  wird daher das Gleichungssystem (\*) gelöst. Wir müssen daher lediglich noch (\*\*) lösen. Dies entspricht der Suche nach  $c$  mit  $\mathcal{N}(0, 1)\left(\left[\frac{\sqrt{n}\mu}{\sigma} - c, \frac{\sqrt{n}\mu}{\sigma} + c\right]\right) = \alpha$ . Dies ist ein Problem, das numerisch gelöst werden muss.



Wir kommen nun zum allgemeinen Fall für ein beliebiges Intervall  $(\mu_1, \mu_2)$ . Unser Ansatz ist es, das Intervall so zu verschieben, dass es symmetrisch um die Null liegt, also die Form  $(-\mu, \mu)$  hat. Dazu setzen wir  $\mu_1 - a = -\mu$  und  $\mu_2 - a = \mu$  und erhalten  $a = \frac{1}{2}(\mu_1 + \mu_2)$  und  $\mu = \frac{1}{2}(\mu_2 - \mu_1)$ . Nun wenden wir die oben erarbeitete Lösung für den Spezialfall auf die Daten  $x - \bar{a}$  an, wobei  $\bar{a} := (a, \dots, a) \in \mathbf{R}^n$  sei. Wir erhalten dann den Test

$$\varphi^*(x) = \mathbf{1}_{(-c, c)}(T(x) - \bar{a}),$$

wobei  $c > 0$  wieder wie oben gewählt ist, das heißt  $c$  löst die Gleichung

$$\mathcal{N}(0, 1) \left( \left[ \frac{\sqrt{n} \frac{\mu_2 - \mu_1}{2}}{\sigma} - c, \frac{\sqrt{n} \frac{\mu_2 - \mu_1}{2}}{\sigma} + c \right] \right) = \alpha. \quad //$$

### Beispiel 4.2.15

Es sei  $P_{\mu, \sigma^2} := \mathcal{N}^n(\mu, \sigma^2)$ , sowohl der Erwartungswert  $\mu$ , als auch die Varianz  $\sigma^2$  seien also unbekannt. Wir betrachten  $\Theta := \mathbf{R} \times (0, \infty)$  mit den Hypothesen  $\Theta_0 := \{\mu_0\} \times (0, \infty)$  und  $\Theta_1 := \Theta \setminus \Theta_0$ . Ferner sei  $T: \mathbf{R}^n \rightarrow \mathbf{R}$  vermöge  $T(x) := \sqrt{n} \frac{\bar{x}_n - \mu_0}{s_n}$  und  $q_{1-\frac{\alpha}{2}}$  das  $1 - \frac{\alpha}{2}$ -Quantil der  $t_{n-1}$ -Verteilung. Korollar 1.2.15 und Beispiel 3.1.10 liefern

$$\begin{aligned} P_{\mu, \sigma^2} \left( |T| \geq q_{1-\frac{\alpha}{2}} \right) &= P_{\mu, \sigma^2} \left( -T \geq q_{1-\frac{\alpha}{2}} \right) + P_{\mu, \sigma^2} \left( T \geq q_{1-\frac{\alpha}{2}} \right) \\ &= P_T \left( \left( -\infty, -q_{1-\frac{\alpha}{2}} \right] \right) + P_T \left( \left[ q_{1-\frac{\alpha}{2}}, \infty \right) \right) \\ &= P_T \left( \left( -\infty, q_{\frac{\alpha}{2}} \right] \right) + 1 - P_T \left( \left( -\infty, q_{1-\frac{\alpha}{2}} \right] \right) \\ &= \frac{\alpha}{2} + 1 - \left( 1 - \frac{\alpha}{2} \right) \\ &= \alpha. \end{aligned}$$

Wir führen nun eine Konstruktion analog zu der von einparametrischen Exponentialfamilien durch. Der Test ist dann gegeben vermöge

$$\varphi^*(x) := \begin{cases} 1 & \text{falls } |T(x)| \geq q_{1-\frac{\alpha}{2}} \\ 0 & \text{sonst} \end{cases}.$$

Dieser Test heißt **Student- $t$ -Test** und ist ein gleichmäßig bester Test zum Niveau  $\alpha$ . Der Beweis für diese Eigenschaft findet sich in [Rasch95, S. 338]. Auch andere typische Hypothesen bzw. Testprobleme, die wir bereits kennengelernt haben, sind durch diesen Test behandelbar, wenn der Test entsprechend modifiziert wird. //

Als nächstes wollen wir uns ein paar Beispiele von Familien mit monotonem Dichtequotienten anschauen.

### Beispiel 4.2.16 Binomialverteilung

Die Dichte der Binomialverteilung  $B(n, p)$  ist für  $x \in \{0, \dots, n\}$  und  $p \in (0, 1)$  gegeben durch

$$f_p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} (1-p)^n p^x (1-p)^{-x} = \binom{n}{x} (1-p)^n \exp \left( x \ln \frac{p}{1-p} \right).$$

Daher ist  $B(n, p)$  eine einparametrische Exponentialfamilie vermöge der Abbildungen  $C(p) := (1-p)^n$ ,  $h(x) := \binom{n}{x} \mathbf{1}_{\{0, \dots, n\}}(x)$ ,  $\xi(p) := \ln \frac{p}{1-p}$  und  $T(x) := x$ . Die Voraussetzung  $p \notin \{0, 1\}$  ist wirklich nötig. Betrachten wir nun zum Beispiel das einseitige Testproblem  $\Theta := (0, p_0]$  und  $\Theta := (p_0, 1)$  für ein  $p_0 \in (0, 1)$ , so ist der zugehörige gleichmäßig beste Test zum Niveau  $\alpha$  gegeben durch

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

wobei  $k$  und  $\gamma^*$  durch

$$\begin{aligned} 1 - \alpha &= B(n, p_0)(\{0, \dots, k\}) - \gamma^* B(n, p_0)(\{k\}) \\ &= \sum_{i=0}^k p_0^i \binom{n}{i} (1-p_0)^{n-i} - \gamma^* \binom{n}{k} p_0^k (1-p_0)^{n-k} \end{aligned} \quad (*)$$

bestimmt werden. Dies ist ein numerisches Problem, das man zum Beispiel mit Hilfe von Quantilfunktionen in  $\mathbb{R}$  lösen kann. Ein anderer, einfacher Ansatz ist es, zunächst das kleinste  $k$  mit

$$1 - \alpha \leq \sum_{i=0}^k \binom{n}{i} p_0^i (1-p_0)^{n-i}$$

zu finden und dann  $\gamma^*$  durch die Gleichung (\*) zu ermitteln. Falls  $n$  besonders groß ist, kann man für den ersten Schritt  $B(n, p)$  approximieren. Ist  $np_0^2$  klein, so approximiert man mittels einer Poissonverteilung, ist  $np_0(1-p_0) \geq 0.9$ , so verwendet man die Standardnormalverteilung. Andere Testprobleme werden analog behandelt, sind zum Teil jedoch aufwändiger. //

#### Beispiel 4.2.17 Poissonverteilung

Die Dichte der Poissonverteilung ist für  $\lambda > 0$  und  $x \in \mathbf{N}_0$  gegeben durch

$$f_\lambda(x) = \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \cdot \frac{1}{x!} \cdot e^{x \ln \lambda}.$$

Setzen wir  $C(\lambda) := e^{-\lambda}$ ,  $h(x) := \frac{1}{x!} \mathbf{1}_{\mathbf{N}_0}(x)$ ,  $\xi(\lambda) := \ln \lambda$  und  $T(x) := x$ , so sehen wir, dass auch dies eine einparametrische Exponentialfamilie ist. Wir betrachten wieder die Hypothesen  $\Theta_0 := (0, \lambda_0]$  und  $\Theta_1 := (\lambda_0, \infty)$ , wobei, wenn die Poissonverteilung als Approximation von  $B(n, p)$  verwendet wird,  $\lambda_0 := np_0$  gesetzt wird. Der zugehörige gleichmäßig beste Test zum Niveau  $\alpha$  ist dann gegeben durch

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

wobei  $k$  und  $\gamma^*$  durch

$$\begin{aligned} 1 - \alpha &= \text{Pois}(\lambda)(\{0, \dots, k\}) - \gamma^* \text{Pois}(\lambda)(\{k\}) \\ &= e^{-\lambda_0} \sum_{i=0}^k \frac{\lambda_0^i}{i!} - \gamma^* e^{-\lambda_0} \frac{\lambda_0^k}{k!} \end{aligned} \quad (*)$$

bestimmt werden. Auch dies kann wieder numerisch gelöst werden, allerdings kommt es beim naiven Ansatz über die Implementierung von Schleifen zu numerischen Instabilitäten, auf die man achten muss. In diesem Fall ist die Verwendung von Quantil-Funktionen daher vorzuziehen. //

**Beispiel 4.2.18** *Normalverteilung I - unbekannter Erwartungswert*

Die Dichte der Normalverteilung ist für  $\sigma^2 > 0$  gegeben durch

$$f_\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(\frac{\mu x}{\sigma^2}\right).$$

Den Fall  $n = 1$  haben wir in Beispiel 4.2.14 bereits kennengelernt. Mit  $C(\mu) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right)$ ,  $h(x) := \exp\left(-\frac{x^2}{2\sigma^2}\right)$ ,  $\xi(\mu) := \frac{\mu}{\sigma^2}$  und  $T(x) := x$  ist dies dann eine einparametrische Exponentialfamilie. Für die Hypothesen  $\Theta_0 := (-\infty, \mu_0]$  und  $\Theta_1 := (\mu_0, \infty)$  bei gegebenem  $\mu_0 \in \mathbf{R}$  erhalten wir dann den zugehörigen gleichmäßig besten Test zum Niveau  $\alpha$  vermöge

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

In diesem Test brauchen wir den Fall  $x = k$  nicht zu betrachten, da dies nur auf einer Nullmenge eintritt. Der Parameter  $k$  wird durch

$$1 - \alpha = \mathcal{N}(\mu_0, \sigma^2)((-\infty, k])$$

bestimmt. Mit einer einfachen Transformation<sup>1</sup> erhalten wir

$$\mathcal{N}(\mu_0, \sigma^2)((-\infty, k]) = \Phi\left(\frac{k - \mu_0}{\sigma}\right),$$

wobei  $\Phi$  die Verteilungsfunktion der Standardnormalverteilung ist. Ist nun  $\alpha \mapsto q_\alpha$  die Quantilfunktion von  $\mathcal{N}(0, 1)$ , so folgt  $q_{1-\alpha} = \frac{k - \mu_0}{\sigma}$  genau dann, wenn  $k = \mu_0 + \sigma q_{1-\alpha}$  gilt. //

**Beispiel 4.2.19** *Gauß-Test*

Wir wollen nun den Gauß-Test als Approximation für den Test bei binomialverteilten Beobachtungen betrachten. Es gilt, dass  $B(n, p)$  die Verteilung der Summe  $\sum_{i=1}^n Y_i$  von  $n$  unabhängigen und  $B(1, p)$ -verteilten Zufallsvariablen ist. Mit dem zentralen Grenzwertsatz CLT und  $\text{Var } Y_i = p(1-p)$  folgt

$$\frac{1}{\sqrt{np(1-p)}} \left( \sum_{i=1}^n (Y_i - p) \right) \rightarrow \mathcal{N}(0, 1).$$

<sup>1</sup> Wir verweisen hierfür auf die Wahrscheinlichkeitstheorie-Vorlesung.

Für  $a \in \mathbf{R}$  gilt daher

$$P\left(\frac{1}{\sqrt{np(1-p)}} \sum_{i=1}^n (Y_i - p) \leq a\right) \approx \Phi(a).$$

Wählen wir nun  $k := np_0 + a\sqrt{np_0(1-p_0)}$ , so folgt wegen

$$\frac{1}{\sqrt{np_0(1-p_0)}} \sum_{i=1}^n (Y_i - p_0) \leq a \iff \sum_{i=1}^n Y_i \leq k,$$

dass für den Erwartungswert  $\mu_0 := np_0$  und die Varianz  $\sigma^2 := np_0(1-p_0)$  von  $B(n, p_0)$

$$\begin{aligned} B(n, p_0)(\{0, \dots, k\}) &= P\left(\sum_{i=1}^n Y_i \leq k\right) \\ &\approx \Phi\left(\frac{k - np_0}{\sqrt{np_0(1-p_0)}}\right) \approx \Phi\left(\frac{k - \mu_0}{\sigma}\right) \end{aligned}$$

gilt. Wird im Binomialtest-Problem nun obige Approximation von  $B(n, p)$  verwendet, so muss  $k$  durch

$$\begin{aligned} 1 - \alpha &= B(n, p_0)(\{0, \dots, k\}) - \gamma^* \binom{n}{k} p_0^k (1-p_0)^{n-k} \\ &\approx \Phi\left(\frac{k - \mu_0}{\sigma}\right) \end{aligned}$$

bestimmt werden, das heißt es gilt  $k = \mu_0 + \sigma q_{1-\alpha}$ . Mit anderen Worten bedeutet dies, dass der Gauß-Test für bekannte Varianz eine Approximation des gleichmäßig besten Tests der Binomialverteilung darstellt, falls  $\mu_0 := np_0$  und  $\sigma^2 := np_0(1-p_0)$  verwendet wird. //

**Beispiel 4.2.20** Normalverteilung II - unbekannte Varianz

Diesmal nehmen wir an, dass der Erwartungswert  $\mu$  bekannt, die Varianz  $\sigma^2 > 0$  jedoch unbekannt ist. Die Dichte ist wieder gegeben durch

$$f_\mu(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),$$

wodurch die Normalverteilung für  $\sigma > 0$  mit  $C(\sigma) := \frac{1}{\sqrt{2\pi\sigma^2}}$ ,  $h(x) := 1$ ,  $\xi(\sigma) := -\frac{1}{2\sigma^2}$  und  $T(x) := (x - \mu)^2$  zu einer einparametrischen Exponentialfamilie wird, wobei  $\xi$  offenbar streng monoton wachsend ist. Für die Hypothesen  $\Theta_0 := (0, \sigma_0]$  und  $\Theta_1 := (\sigma_0, \infty)$  ist der zugehörige gleichmäßig beste Test zum Niveau  $\alpha$  gegeben durch

$$\varphi^*(x) = \begin{cases} 1 & \text{falls } (x - \mu)^2 > k \\ 0 & \text{falls } (x - \mu)^2 \leq k \end{cases},$$

wobei der Parameter  $k$  durch

$$1 - \alpha = \mathcal{N}(\mu, \sigma_0^2)(\{T \leq k\})$$

bestimmt werden kann. Nun gilt

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_{\{(x-\mu)^2 \leq k\}} \exp\left(-\frac{(x-\mu)^2}{2\sigma_0^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_{\{x^2 \leq k\}} \exp\left(-\frac{x^2}{2\sigma_0^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\{\sigma_0^2 x^2 \leq k\}} \exp\left(-\frac{x^2}{2}\right) dx \\ &= F_{\chi_1^2}\left(\frac{k}{\sigma_0^2}\right), \end{aligned}$$

wobei  $F_{\chi_1^2}$  die Verteilungsfunktion der  $\chi_n^2$ -Verteilung mit  $n = 1$  ist. Sei nun  $q_{1-\alpha}$  das  $1 - \alpha$ -Quantil der  $\chi_1^2$ -Verteilung. Wir setzen  $\frac{k}{\sigma_0^2} := q_{1-\alpha}$ , was äquivalent zu  $k = \sigma_0^2 q_{1-\alpha}$  ist. Damit erhalten wir, dass  $1 - \alpha = F_{\chi_1^2}\left(\frac{k}{\sigma_0^2}\right)$  erfüllt ist. Der Test ist somit gegeben durch

$$\varphi^*(x) = \begin{cases} 1 & \text{falls } (x - \mu)^2 > \sigma_0^2 q_{1-\alpha} \\ 0 & \text{sonst} \end{cases} . \quad //$$

### 4.3 Zweistichprobentests

Häufig müssen zwei Produkte miteinander verglichen werden, um zum Beispiel zu überprüfen, ob Produkt A die selben Eigenschaften wie Produkt B hat, oder ob Produkt A signifikant besser ist als Produkt B. Ein typisches Beispiel hierfür ist der Vergleich von neuen mit alten Medikamenten in der Medizin. Wir wollen uns jetzt mit entsprechenden statistischen Tests beschäftigen.

Zunächst wollen wir davon ausgehen, dass wir Zufallsvariablen  $X_i \sim \mathcal{N}(\mu_1, \sigma_1^2)$  haben, die wir  $n_1$ -mal beobachten. Außerdem seien Zufallsvariablen  $Y_i \sim \mathcal{N}(\mu_2, \sigma_2^2)$  gegeben, die wir  $n_2$ -mal beobachten. Ferner wollen wir vorerst davon ausgehen, dass  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  unabhängig sind.

Gilt  $n_1 = n_2$ , so sind die Zufallsvariablen  $X_i - Y_i =: Z_i$  unabhängig und normalverteilt. In diesem Fall sind die Testprobleme auf die Beispiele des vorherigen Kapitels zurückführbar, da wir uns darauf beschränken können, nur die  $Z_i$  zu betrachten.

#### Beispiel 4.3.1 Gleichheit des Erwartungswertes

Wir betrachten den Fall, dass  $\sigma_1^2 = \sigma_2^2 =: \sigma^2$  unbekannt sei mit  $\Theta := \mathbf{R}^2 \times (0, \infty)$ . Dann sei  $\Theta_0 := \{(\mu_1, \mu_2, \sigma^2) \in \Theta \mid \mu_1 = \mu_2\}$  und  $\Theta_1 := \Theta \setminus \Theta_0$ . Wir betrachten  $T: \mathbf{R}^n \cong (\mathbf{R}^{n_1}, \mathbf{R}^{n_2}) \rightarrow \mathbf{R}$  mit  $n := n_1 + n_2$  vermöge

$$T(x, y) := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \frac{\bar{x}_{n_1} - \bar{y}_{n_2}}{s_{n_1 + n_2}},$$

wobei  $s_{n_1 + n_2} := \frac{1}{n_1 + n_2 - 2} (\sum_{i=1}^{n_1} (x_i - \bar{x}_n)^2 + \sum_{i=1}^{n_2} (y_i - \bar{y}_n)^2)$  ist. Ferner sei  $q_{1-\frac{\alpha}{2}}$  das  $1 - \frac{\alpha}{2}$ -Quantil der  $t_{n_1 + n_2 - 2}$ -Verteilung. Wir betrachten nun den so genannten **Zweistichproben- $t$ -Test**

$$\varphi^*(x, y) := \begin{cases} 1 & \text{falls } |T(x, y)| > q_{1-\frac{\alpha}{2}} \\ 0 & \text{sonst} \end{cases}.$$

Dieser Test ist ein Test zum Niveau  $\alpha$ . Der Beweis hierfür findet sich in [Rasch95, S. 340 ff.]. //

#### Beispiel 4.3.2 Einseitiger (vergleichender) Test

Der Parameterraum  $\Theta$  sei wie in Beispiel 4.3.1 gegeben. Nun betrachten wir jedoch die Hypothesen  $\Theta_0 := \{(\mu_1, \mu_2, \sigma^2) \in \Theta \mid \mu_1 \geq \mu_2\}$  und  $\Theta_1 := \Theta \setminus \Theta_0$ . Wir betrachten  $T: \mathbf{R}^{n_1 + n_2} \rightarrow \mathbf{R}$  wie in Beispiel 4.3.1 und  $q_{1-\frac{\alpha}{2}}$  sei das  $1 - \frac{\alpha}{2}$ -Quantil der  $t_{n_1 + n_2 - 2}$ -Verteilung. Dann ist der durch

$$\varphi^*(x, y) := \begin{cases} 1 & \text{falls } T(x, y) < -q_{1-\alpha} = q_\alpha \\ 0 & \text{sonst} \end{cases}$$

gegebene Test ein gleichmäßig bester unverfälschter Test zum Niveau  $\alpha$  (vgl. Definition 4.2.12). Auch dies ist ein Zweistichproben- $t$ -Test. Die entsprechenden Beweise finden sich in [Rasch95, S. 340 ff.]. //

Wir wollen nun einige Details der beiden Beispiel 4.3.1 und 4.3.2 diskutieren. Die Einschränkung  $\sigma_1 = \sigma_2$  ist in der Praxis oft schwer zu rechtfertigen, insbesondere, da wir annehmen, die Varianzen seien gleich, ohne sie jedoch zu kennen. Sind die Varianzen  $\sigma_1^2 \neq \sigma_2^2$  jedoch beide bekannt, so kann man den **Zweistichproben-Gauß-Test** anwenden, der durch die Stichprobenfunktion  $T: \mathbf{R}^{n_1+n_2} \rightarrow \mathbf{R}$  vermöge

$$T(x, y) := \frac{\overline{x_{n_1}} - \overline{y_{n_2}}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

beschrieben wird. Betrachtet man zum Beispiel die Hypothese  $H_0 := \{\mu_1 \geq \mu_2\}$ , so ist der zugehörige Test gegeben durch

$$\varphi^*(x, y) := \begin{cases} 1 & \text{falls } T(x, y) < q_\alpha \\ 0 & \text{sonst} \end{cases},$$

wobei  $q_\alpha$  das  $\alpha$ -Quantil der Standardnormalverteilung ist. Auch dies sind gleichmäßig beste unverfälschte Tests zum Niveau  $\alpha$ .

Wir haben bisher den Fall mit unbekanntem Varianzen  $\sigma_1^2 \neq \sigma_2^2$  nicht behandelt. Man kann hier zwar Tests konstruieren, die das Niveau  $\alpha$  erreichen, es ist jedoch deutlich schwieriger einen gleichmäßig besten Tests zu finden, weshalb dies meist nur approximativ lösbar ist. Ein Ansatz findet sich in [Rasch95, S. 343 ff.].

Außerdem ist es oft so, dass  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  nicht unabhängig sind, zum Beispiel für  $n_1 = n_2$  sind dann aber die Zufallsvariablen  $Z_i := X_i - Y_i$  unabhängig. Dies führt zu Tests mit verbundenen Stichproben, beispielsweise wenn in einer Medikamentenstudie die Patienten erst das erste und dann das zweite Medikament einnehmen.

In der Praxis liegen oft Hypothesen der Form  $\{\mu_1 \geq \mu_2 + a\}$  mit einem  $a \in \mathbf{R}$  vor. Diese Situation lässt sich durch eine Translation auf die obigen Beispiele zurückführen.

## 4.4 Eine Anwendung von Tests

In diesem Kapitel wollen wir uns ein ausführliches Beispiel anschauen, bei welchem verschiedene der bisher kennengelernten Techniken zum Einsatz kommen.

### 4.4.1 Informelle Beschreibung des Problems

Es seien  $A$  und  $B$  zwei Algorithmen, die zu Datensätzen  $D = ((x_1, y_1), \dots, (x_n, y_n)) \in (X \times \{-1, 1\})^n$  Funktionen  $f_A, f_B: X \rightarrow \{-1, 1\}$  ausgeben, wobei  $X$  eine Menge ist. Diese Funktionen sollen für einen *neuen* Eingabewert  $x$  den „zugehörigen“ Ausgabewert  $y$  durch  $f_A(x)$  und  $f_B(x)$  richtig beschreiben.

Die Frage, mit der wir uns letztlich beschäftigen wollen, lautet: Welcher Algorithmus produziert die besseren Funktionen? Wir werden dabei jedoch nicht darauf eingehen, wie man diese Algorithmen konkret konstruieren kann.

Die betrachteten Algorithmen sind so genannte **Lernalgorithmen**, welche die unbekannte Eingabe-Ausgabe-Beziehung zwischen  $x$  und  $y$  anhand von  $D$  „lernen“. Solche Algorithmen werden zunehmend interessant, da sehr komplexe Daten an Bedeutung gewinnen. Wir wollen an dieser Stelle einige Beispielanwendungen auflisten:

- Textanalyse, maschinelle Sprachverarbeitung
  - Methoden zur Textkategorisierung (handelt es sich um Politik, Sport, Boulevard, ...?)
  - Spam-Filter
- Bildanalyse
  - *Optical Character Recognition* (Handschrifterkennung)
  - *Image Classification*
- Bioinformatik
- Ingenieurswesen

### 4.4.2 Formale Beschreibung des Problems

Wir haben ein völlig unbekanntes Wahrscheinlichkeitsmaß  $P$  auf  $X \times \{-1, 1\}$  und einen Datensatz  $D \sim P^n$ . Jedes neue Paar  $(x, y)$  ist ebenfalls  $P$ -verteilt und unabhängig, das heißt alle Daten sind i. i. d. Ist  $f: X \rightarrow \{-1, 1\}$ , so heißt  $\mathcal{R}_P(f) := P(\{(x, y) : f(x) \neq y\})$  der **mittlere Klassifikationsfehler** beziehungsweise **Klassifikationsrisiko**. Ist  $(x'_1, y'_1), \dots, (x'_{n'}, y'_{n'})$  eine Folge



von zukünftigen Beobachtungen, so gilt mit dem starken Gesetz der großen Zahlen  $P^\infty$ -fast sicher

$$\frac{1}{n'} \sum_{i=1}^{n'} \mathbf{1}_{\{f(x'_i) \neq y'_i\}} \rightarrow \mathcal{R}_P(f).$$

Die Funktion  $f_B$  ist besser als  $f_A$  genau dann, wenn  $\mathcal{R}_P(f_B) < \mathcal{R}_P(f_A)$  gilt.

Unser Ziel ist es nun, zu entscheiden, ob der Algorithmus  $B$  besser ist als der Algorithmus  $A$ . Es gilt zu beachten, dass „Ist  $B$  besser als  $A$ ?“ und „Ist  $f_B$  besser als  $f_A$ ?“ unterschiedliche Fragen sind und wir oben nur für die Entscheidungsfunktionen definiert haben, was wir unter dem Begriff „besser“ verstehen. Dies wird zu verschiedenen Testproblemen führen.

### 4.4.3 Informelle Wahl der Hypothesen

Wir betrachten  $A$  als einen älteren, bereits bekannten Algorithmus, während  $B$  ein neuer Algorithmus ist, von dem wir testen wollen, ob er besser als  $A$  ist.

Im Sinne des wissenschaftlichen Fortschritts sollte das Testergebnis „ $B$  ist besser als  $A$ “ fundiert sein. Dies muss bei der Wahl der Hypothesen berücksichtigt werden. Dazu wählen folgenden Ansatz:

$$H_0 : B \text{ ist nicht besser als } A$$

$$H_1 : B \text{ ist besser als } A$$

Ist  $\varphi$  ein Test zum Niveau  $\alpha$ , so gilt, dass die Wahrscheinlichkeit für Fehler erster Art (also dafür, dass  $H_0$  wahr ist, während  $\varphi$  das Gegenteil sagt) kleiner oder gleich  $\alpha$  ist. Wir können also durch  $\alpha$  dafür sorgen, dass das Testergebnis „ $B$  ist besser als  $A$ “ fundiert ist, wir haben jedoch im Allgemeinen keine Kontrolle über die Wahrscheinlichkeit für Fehler zweiter Art und daher wissen wir nicht hinreichend sicher, ob das Testergebnis „ $B$  ist nicht besser als  $A$ “ stimmt.

Diese Beobachtung gilt für jedes Testproblem, bei welchem die Fehler zweiter Art nicht kontrolliert werden können. Tests zum Niveau  $\alpha$  können also nur mit hinreichender Sicherheit die Nullhypothese widerlegen, nicht jedoch bestätigen.

### 4.4.4 Test auf einem Datensatz

Wir nehmen zunächst an, dass wir genau einen Trainingsdatensatz  $D$  und einen Testdatensatz  $D' := ((x'_1, y'_1), \dots, (x'_n, y'_n))$  haben. Präziser formuliert lautet unsere Frage nun: Hat  $B$  mit Hilfe von  $D$  die bessere Entscheidungsfunktion erzeugt?

Dazu sei  $p_A := \mathcal{R}_P(f_A)$  und  $p_B := \mathcal{R}_P(f_B)$ , wobei  $p_A, p_B \in [0, 1]$  unbekannt sind. Für das empirische Risiko

$$\mathcal{R}_{D'}(f) := \frac{1}{n'} \sum_{i=1}^{n'} \mathbf{1}_{\{f(x'_i) \neq y'_i\}}$$

gilt nun  $\mathcal{R}_{D'}(f_A) \sim B(n', p_A)$  und  $\mathcal{R}_{D'}(f_B) \sim B(n', p_B)$ . Nach dem zentralen Grenzwertsatz sind beide Zufallsvariablen (in  $D'$ ) approximativ normalverteilt und diese Tatsache werden wir nutzen. Wir können die Varianzen  $n'p_A(1-p_A)$  und  $n'p_B(1-p_B)$  jedoch nicht als gleich voraussetzen und können daher nicht auf die Beispiele in Kapitel 4.3 zurückgreifen.

Nach dem zentralen Grenzwertsatz ist  $D' \mapsto \mathcal{R}_{D'}(f_A) - \mathcal{R}_{D'}(f_B)$  jedoch normalverteilt. Dazu betrachten wir einen Student- $t$ -Test und es sei  $\mu := p_A - p_B$ ,  $\Theta_0 := \{(\mu, \sigma^2) : \mu \leq 0, \sigma^2 > 0\}$ , wobei  $\mu \leq 0$  äquivalent zu  $p_A \leq p_B$  ist und daher auch dazu, dass  $B$  nicht besser als  $A$  ist, diese Hypothese entspricht also der bereits informell eingeführten Version. Entsprechend sei  $\Theta_1 := \{(\mu, \sigma^2) : \mu > 0, \sigma^2 > 0\}$ . Zudem betrachten wir die Zufallsvariablen  $z_i := \mathbf{1}_{\{f_A(x'_i) \neq y'_i\}} - \mathbf{1}_{\{f_B(x'_i) \neq y'_i\}}$  und  $s_n^2 := \frac{1}{n'-1} \sum_{i=1}^{n'} (z_i - \bar{z}_{n'})^2$ . Ferner sei

$$T(z) := \frac{\sqrt{n'} \bar{z}_{n'}}{s_{n'}} = \frac{\sqrt{n'} (\mathcal{R}_{D'}(f_A) - \mathcal{R}_{D'}(f_B))}{s_{n'}}.$$

Ist  $q_{1-\alpha}$  das  $1 - \alpha$ -Quantil der  $t_{n'-1}$ -Verteilung, so ist der Student- $t$ -Test unseres Problems gegeben durch

$$\varphi^*(z) = \begin{cases} 1 & \text{falls } T(z) > q_{1-\alpha} \\ 0 & \text{sonst} \end{cases}.$$

Mit anderen Worten muss  $\mathcal{R}_{D'}(f_A)$  also deutlich größer als  $\mathcal{R}_{D'}(f_B)$  sein um die Nullhypothese zu widerlegen. Wie gut kontrolliert dieser Test nun Fehler zweiter Art? Sei dazu  $\mu_0 := p_A - p_B > 0$ . Mit  $s_n^2 = \frac{n'}{n'-1} \frac{1}{n'} \sum_{i=1}^{n'} (z_i - \bar{z}_{n'})^2 = \frac{n'}{n'-1} \text{Var}_{D'} \text{id} \leq \frac{2n'}{n'-1}$  gilt dann

$$\begin{aligned} P_{\mu_0, \sigma^2} \left( T - \frac{\sqrt{n'} \mu_0}{s_{n'}} \geq x \right) &= P_{\mu_0, \sigma^2} \left( T \geq x + \frac{\sqrt{n'} \mu_0}{s_{n'}} \right) \\ &\leq P_{\mu_0, \sigma^2} \left( T \geq x + \frac{\sqrt{n'-1}}{\sqrt{2}} \mu_0 \right). \end{aligned}$$

Wir setzen nun  $x := q_{1-\alpha} - \frac{\sqrt{n'-1}}{\sqrt{2}} \mu_0$  und erhalten damit

$$\begin{aligned} \beta_{\varphi^*}(\mu_0, \sigma^2) &= P_{\mu_0, \sigma^2}(T \geq q_{1-\alpha}) \\ &\geq P_{\mu_0, \sigma^2} \left( T - \frac{\sqrt{n'} \mu_0}{s_{n'}} \geq x \right) \\ &= P_T([x, \infty)) = 1 - P_T((-\infty, x]). \end{aligned}$$

Die Wahrscheinlichkeit für Fehler zweiter Art, falls  $\mu_0$  wahr ist, ist damit

$$1 - \beta_{\varphi^*}(\mu_0, \sigma^2) \leq P_T((-\infty, x]) = t_{n-1} \left( \left( -\infty, q_{1-\alpha} - \frac{\sqrt{n'-1}}{\sqrt{2}} \mu_0 \right) \right).$$

Für  $n' \rightarrow \infty$  gilt  $\frac{\sqrt{n'-1}}{\sqrt{2}} \mu_0 \rightarrow \infty$  und damit geht die Fehlerwahrscheinlichkeit gegen Null.

Wir haben unsere Eingangsfrage damit jedoch noch nicht beantwortet, da wir bis jetzt lediglich einen Test konstruiert haben, der entscheidet, ob  $B$  auf einem einzelnen Testdatensatz besser

als  $A$  ist. Wie können wir nun entscheiden, ob  $B$  besser als  $A$  für alle von  $P$  generierten Datensätze ist?

Es gilt zu beachten, dass die Verteilungen der Zufallsvariablen  $D \mapsto \mathcal{R}_P(f_{A,D})$  und  $D \mapsto \mathcal{R}_P(f_{B,D})$  völlig unbekannt sind.

### 4.4.5 Test auf vielen Datensätzen I

Bis jetzt haben wir noch nicht definiert, was wir darunter verstehen, dass  $B$  auf einer Verteilung  $P$  besser als  $A$  ist. Eine Möglichkeit hierfür ist es, ein  $n \in \mathbf{N}$  zu fixieren und zu sagen, dass  $B$  für  $P$  und  $n$  besser als  $A$  ist genau dann, wenn  $\mathbf{E}_{D \sim P^n} \mathcal{R}_P(f_{B,D}) < \mathbf{E}_{D \sim P^n} \mathcal{R}_P(f_{A,D})$  gilt. Mit anderen Worten ist das dann der Fall, wenn das mittlere Risiko von  $B$  kleiner als das mittlere Risiko von  $A$  ist.

Wir schreiben  $p_A := \mathbf{E}_{D \sim P^n} \mathcal{R}_P(f_{A,D})$  und analog  $p_B$ . Zudem wählen wir  $m$ -mal einen Trainingsdatensatz  $D_i$  der Länge  $n$  und  $m$ -mal einen Testdatensatz  $D'_i$  der Länge  $n'$ , wobei alles identisch und unabhängig verteilt sei. Wir betrachten die Zufallsvariablen  $z_i := \mathcal{R}_{D'_i}(f_{B,D_i}) - \mathcal{R}_{D'_i}(f_{A,D_i})$  und  $s_m^2 := \frac{1}{m-1} \sum_{i=1}^m (z_i - \bar{z}_m)^2$ . Man sieht, dass  $\frac{1}{m} \sum_{i=1}^m z_i$  approximativ normalverteilt ist, daher wollen wir einen  $t$ -Test auf  $z_i$  konstruieren. Wir müssen zuvor nachprüfen, dass die Erwartungswerte der  $z_i$  den Erwartungswerten unserer Definition von „besser“ entsprechen. Dazu betrachten wir

$$\begin{aligned} \mathbf{E}_{D_i \sim P^n, D'_i \sim P^{n'}}(z_i) &= \mathbf{E}_{D_i \sim P^n} \mathbf{E}_{D'_i \sim P^{n'}} \left( \frac{1}{n'} \sum_{j=1}^{n'} \left( \mathbf{1}_{\{f_{B,D_i}(x'_j) \neq y'_j\}} - \mathbf{1}_{\{f_{A,D_i}(x'_j) \neq y'_j\}} \right) \right) \\ &= \mathbf{E}_{D_i \sim P^n} \frac{1}{n'} \sum_{j=1}^{n'} \underbrace{\mathbf{E}_{(x'_j, y'_j) \sim P} \left( \mathbf{1}_{\{f_{B,D_i}(x'_j) \neq y'_j\}} - \mathbf{1}_{\{f_{A,D_i}(x'_j) \neq y'_j\}} \right)}_{\mathcal{R}_P(f_{B,D_i}) - \mathcal{R}_P(f_{A,D_i})} \\ &= p_B - p_A. \end{aligned}$$

Daher ergibt es Sinn, den  $t$ -Test anzuwenden, um zu entscheiden, ob  $\mathbf{E}z_i \leq 0$  gilt. Dazu wählen wir  $\Theta_0 := \{(\mu, \sigma^2) : \mu \leq 0, \sigma^2 > 0\}$  und  $\Theta_1 := \{(\mu, \sigma^2) : \mu > 0, \sigma^2 > 0\}$ , wobei  $\mu = \mathbf{E}z_i = p_B - p_A$  ist. Unsere Testgröße ist  $T(z) = \sqrt{m} \frac{\bar{z}_m}{s_m}$ , wobei  $z = (z_1, \dots, z_m)$  ist. Dann ist unser Test gegeben durch

$$\varphi^*(z) = \begin{cases} 1 & \text{falls } T(z) > q_{1-\alpha} \\ 0 & \text{sonst} \end{cases},$$

wobei  $q_{1-\alpha}$  das  $1 - \alpha$ -Quantil von  $t_{m-1}$  ist.

Ein Vorteil dieser Methode ist, dass sie immer anwendbar ist, ein Nachteil ist jedoch, dass sehr viele Daten benötigt werden. In der Praxis wird häufig ein *bootstrap approach* (dt. Stiefelschlaufen-Ansatz) durchgeführt: Man hat einen nicht besonders großen Datensatz vorliegen und zerlegt diesen  $m$ -mal zufällig in Trainings- und Testdatensatz, wobei Wiederholungen erlaubt sind. Bootstrapping funktioniert in vielen statistischen Problemen und ist ein oft verwendetes Werkzeug.

### 4.4.6 Test auf vielen Datensätzen II

Die Erwartungswerte  $\mathbf{E}_{D \sim P^n}(\mathcal{R}_P(f_D))$  werden relativ stark durch Extremwertverhalten beeinflusst. Die Frage ist also, ob wir wirklich an mittleren Risiken interessiert sind. Als Alternative sagen wir daher, dass  $B$  besser als  $A$  ist genau dann, wenn  $P^n(D : \mathcal{R}_P(f_{B,D}) < \mathcal{R}_P(f_{A,D})) > \frac{1}{2}$  gilt, das heißt in mindestens der Hälfte aller  $D$  ist  $B$  besser als  $A$ . Etwas pragmatischer ist die Bedingung  $P^{n+n'}((D, D') : \mathcal{R}_{D'}(f_{B,D}) < \mathcal{R}_{D'}(f_{A,D})) > \frac{1}{2}$ , mit der wir uns nun auseinandersetzen wollen.

Dies kann durch den so genannten Vorzeichentest überprüft werden. Dazu sei

$$H_0 : P^{n+n'}(\mathcal{R}_{D'}(f_{B,D}) \geq \mathcal{R}_{D'}(f_{A,D})) \geq \frac{1}{2},$$

$$H_1 : P^{n+n'}(\mathcal{R}_{D'}(f_{B,D}) \geq \mathcal{R}_{D'}(f_{A,D})) < \frac{1}{2} \Leftrightarrow P^{n+n'}(\mathcal{R}_{D'}(f_{B,D}) < \mathcal{R}_{D'}(f_{A,D})) > \frac{1}{2}.$$

Diese Hypothesen sind nicht mehr parametrisch. Wir setzen nun  $z_i := \mathcal{R}_{D'_i}(f_{B,D_i}) - \mathcal{R}_{D'_i}(f_{A,D_i})$  und  $T(z) := \sum_{i=1}^m \mathbf{1}_{\{z_i \geq 0\}} \sim \mathbf{B}(m, p)$ , wobei  $p$  unbekannt ist. Die Nullhypothese besagt nun, dass  $p \geq \frac{1}{2}$  ist. Dies wird nun wie im Beispiel 4.2.19 getestet. Es sei  $p_0 := \frac{1}{2}$  und  $k := mp_0 + \sqrt{mp_0(1-p_0)}q_{1-\alpha} = \frac{m}{2} + \frac{\sqrt{m}}{2}q_{1-\alpha}$ , wobei  $q_{1-\alpha}$  das  $1-\alpha$ -Quantil der Standardnormalverteilung ist. Dann ist der Test gegeben durch

$$\varphi^*(z) := \begin{cases} 1 & \text{falls } T(z) < k \\ 0 & \text{sonst} \end{cases}.$$

Mit anderen Worten bedeutet dies, dass die Nullhypothese abgelehnt wird, wenn nicht genügend Erfolge von  $A$  gezählt werden.

Bis jetzt haben wir nur die Frage beantwortet, ob  $B$  besser als  $A$  für ein festes  $P$  ist. Eigentlich wollen wir dies aber für viele, unbekannte  $P$  wissen. Darunter fallen zum Einen ähnliche Verteilungen, die typisch für anwendungsnahe Betrachtungen sind, andererseits auch für unterschiedliche Verteilungen, die typisch für Mehrzweckalgorithmen sind.

# Literaturverzeichnis

- [Witt85] H. Witting, *Mathematische Statistik I*, 1. Auflage, 1985
- [Leh98] E. L. Lehmann und G. Casella, *Theory of Point Estimation*, 2nd edition, 1998
- [Leh05] E. L. Lehmann und J. P. Romano, *Testing Statistical Hypotheses*, 3rd edition, 2005
- [Rasch95] D. Rasch, *Mathematische Statistik*, 1. Auflage, 1995
- [Schlitt96] R. Schlittgen, *Statistische Referenz*, 1. Auflage, 1996



# Abbildungsverzeichnis

0.1	Übersicht der Zusammenhänge zwischen Stochastik, Wahrscheinlichkeitstheorie und Statistik . . . . .	5
1.1	Skizze einer empirischen Verteilungsfunktion . . . . .	19
2.1	Gegenüberstellung beser und gleichmäßig bester Schätzer . . . . .	45
3.1	$\alpha$ -Quantil für die Normalverteilung . . . . .	58
3.2	$\alpha$ -Quantil für die Binomialverteilung . . . . .	59
3.3	Verteilungsfunktion und Umkehrfunktion der Standardnormalverteilung . . . . .	63
4.1	Darstellung einer Gütefunktion . . . . .	73
4.2	Illustration verschiedener $\alpha$ für den Zusammenhang zwischen Gütefunktionen und (unverfälschter) Tests zum Niveau $\alpha$ . . . . .	74
4.3	Illustration einer konsistenten Folge statistischer Tests . . . . .	75
4.4	Erreichen eines vorgegebenen Niveaus durch Verwendung randomisierter Tests . . . . .	79
4.5	Illustration des schlechtestmöglichen Verhaltens Fehler zweiter Art . . . . .	83
4.6	Strenge Monotonie der Gütefunktion in der Erweiterung von Satz 4.2.11 . . . . .	86





# Stichwortverzeichnis

- a-posteriori-Wahrscheinlichkeitsdichte, 44
- a-priori-Verteilung, 44
- Arithmetisches Mittel, 10
- Atomlosigkeit, 59
- Bayes-Schätzer, 45
- Bias, 24
- Chi-Quadrat-Verteilung, 16
- Cramér-Rao-Ungleichung, 49
- Einparametrische Exponentialfamilie, 82
- Empirische Verteilungsfunktion, 19
- Empirisches Maß, 19
- Erwarteter Verlust, 25
- Erwartungstreue, 10, 12, 24
  - Asymptotische, 24
- Exzess-Kurtosis, 26
- Exponentialfamilie
  - einparametrische, 82
- Fisher-Information, 53
- Gütefunktion, 73
- Gammafunktion, 14
- Gammaverteilung, 15
- Gemittelter quadratischer Verlust, 46
- Klassifikationsrisiko, 96
- Konfidenzintervall, 57
  - asymptotisches, 58
  - minimales, 58
- Konsistenz, 11, 12, 74
- Kovarianz, 32
- Lemma
  - von Neyman-Pearson, 77
- Lernalgorithmus, 96
- Likelihood-Funktion, 37
- log-Likelihood-Funktion, 38
- Maximum-Likelihood-Schätzer, 37
- Maxmin-Test, 83
- mittlerer Klassifikationsfehler, 96
- Momentenmethode, 35
- Momentenschätzer, 35
- Monotoner Dichtequotient, 80
- Nullschätzer, 31
- p-Wert, 76
- Quantil, 58
- Risiko, 25
- Satz
  - von Chapman-Robbins, 33
  - von Glivenko-Cantelli, 20
  - von Rao, 31
- Schätzer, 23
  - Bayes-Schätzer, 45
  - bester unverzerrter, 30
  - Maximum-Likelihood-Schätzer, 37
  - Momentenschätzer, 35
  - Nullschätzer, 31
- Schätzung, 23
- Statistischer Raum, 7
- Statistischer Test, 72
  - besserer, 75
  - gleichmäßig bester, 75
  - gleichmäßig bester unter allen unverfälschten, 87
  - konsistente Folge, 74
  - kritischer Bereich, 72
  - Macht, 73
  - Maxmin-Test, 83
  - nicht-randomisierter, 72
  - randomisierter, 72
  - Student- $t$ -Test, 89

unverfälschter, 74  
zum Niveau  $\alpha$ , 74  
Zweistichproben- $t$ -Test, 94  
Zweistichproben-Gauß-Test, 95  
Stichprobenmittel, 10  
Stichprobenvarianz, 12  
Student- $t$ -Test, 89  
Student- $t$ -Verteilung, 17  
  
Unverzerrtheit, 24  
  
Verlustfunktion, 25  
  der kleinsten Quadrate, 25  
Verzerrung, 24  
  
Zufallsstichprobe, 10  
Zweistichproben- $t$ -Test, 94  
Zweistichproben-Gauß-Test, 95