Mathematische Statistik

Lehrstuhl für Mathematische Statistik

Universität Würzburg

Prof. Dr. Michael Falk

Vorwort

Dieses Skript entstand aus der Vorlesung „Mathematische Statistik I und II“, wie sie Prof. Dr. Michael Falk im Wintersemester 2007/2008 bis Sommersemester 2008 an der Julius-Maximilians-Universität Würzburg gehalten hat.

Basierend auf meinen Aufzeichnungen zu dieser Vorlesung habe ich das vorliegende Skript für Herrn Prof. Dr. Falk erstellt. Ich möchte mich auch bei Johannes Hain bedanken, da er dieses Skript nochmals Korrektur gelesen hat.

Im Folgenden wird eine Einführung in die grundlegenden Begriffe und Werkzeuge der Mathematischen Statistik gegeben. Des Weiteren werden fundamentale Sätze der Mathematischen Statistik besprochen und mit Beispielen erläutert.

Stefan Englert

Würzburg, September 2008

Inhaltsverzeichnis

Problemstellung

Unter *Mathematischer Statistik* versteht man die Untersuchung von Mathematischen Modellen sowie die Herleitung bzw. Begründung von Verfahren zur Auswertung von Beobachtungsdaten.

Ein Beispiel zur Erläuterung der Grundproblematik: Zur Heilung einer bestimmten Krankheit wurde eine neue Behandlungsmethode *M2* entwickelt. Um eine Aussage über ihre Qualität zu erhalten, wurde diese bei 10 Patienten angewendet. Dabei trat in 8 Fällen ein Heilerfolg ein, in 2 Fällen ergab sich ein Mißerfolg. Läßt sich nun aufgrund dieser 10 Überprüfungen bereits sagen, dass die neue Methode *M2* häufiger zum Erfolg führt als die herkömmliche Methoden *M1*, deren Heilungschance erfahrungsgemäß 65% beträgt?

Der für die Statistik spezifische Aspekt ist die Tatsache, dass das Eintreten von Erfolg oder Nichterfolg bei einer einzelnen Überprüfung nicht nur von der Qualität der Heilmethode (dann wäre die Entscheidung klar!), sondern auch von sehr vielen anderen uns unzugänglichen und in der Gesamtentwicklung unübersehbaren Einflüssen abhängt, so dass wir das Ergebnis nicht voraussagen können und daher als zufallsabhängig betrachten. Bei unserer Aussage über die Güte von *M3* müssen wir daher die Zufallsabhängigkeit der 10 Ergebnisse berücksichtigen.

Die Verwendung der Wahrscheinlichkeitstheorie ermöglicht es, solche auch gefühlsmäßig unsicheren Entscheidungen zum Gegenstand mathematischer Überlegungen zu machen. Das geschieht dadurch, dass wir die Beobachtungen (Ergebnisse) als Realisierungen von Zufallsvariablen auffassen und damit unterstellen, dass sich der Vorgang durch eine Wahrscheinlichkeitsverteilung beschreiben lässt (*Grundannahme der Mathematischen Statistik).*
Im obigen Beispiel werden wir Zufallsvariablen $X_1, \ldots, X_{10}$ verwenden, die jeweils nur die beiden Werte 1 (für Heilerfolg) und 0 (für Mißerfolg) mit den Wahrscheinlichkeiten ϑ bzw. $1-\vartheta$ annehmen können. Die X_i sind dann $B(1,\vartheta)$–verteilte Zufallsvariablen, wobei durch den uns unbekannten Parameter ϑ die Güte des neu entwickelten Medikamentes angegeben wird:

$$\textit{M2} \text{ ist besser als } \textit{M1} \Leftrightarrow \vartheta > 0,65.$$

Besonders einfach wird die Behandlung dieses Modells, wenn wir zusätzlich voraussetzen, dass die $X_1, \ldots, X_{10}$ stochastisch unabhängig sind (d.h. die Versuchsausführungen beeinflussen sich nicht gegenseitig). Dann ist die Verteilung von $(X_1, \ldots, X_{10})$ das Produktmaß $B(1,\vartheta)^{10}/\{0,1\}^{10}$ und damit die Anzahl der Erfolge, also $\sum_{i\leq 10} X_i$, $B(10,\vartheta)$–verteilt.
Eine Aussage über die unbekannte Verteilung von $X = (X_1, \ldots, X_{10})$ bzw. den unbekannten Verteilungsparameter aufgrund einer zufallsabhängigen Beobachtung (im obigen Beispiel also aufgrund des beobachteten Tupels $(x_1, \ldots,$

x_{10}) mit $\sum_{i\leq 10} x_i = 8$)) heißt eine *statistische Entscheidung*. Folglich ist eine Vorschrift anzugeben, aus der zu jedem möglichen Versuchsausgang die zu treffende Entscheidung abzulesen ist.

Ein Beispiel für eine derartige Entscheidungsvorschrift in obiger Situation ist die folgende:

> Die Entscheidung *M2 ist besser als M1* (d.h. $\vartheta > 0,65$) wird genau dann getroffen, wenn 8 oder mehr Heilerfolge eintreten.

Durch die Verwendung mathematischer Methoden wird die Unsicherheit statistischer Entscheidungen *nicht aufgehoben*! Man kann sie aber durch die Verwendung wahrscheinlichkeitstheoretischer Hilfsmittel quantitativ erfassen, d.h. Wahrscheinlichkeiten für Fehlentscheidungen können (exakt) angegeben werden.

Im obigen Beispiel ist es etwa durchaus möglich, wenn auch nur mit der kleinen Wahrscheinlichkeit $0,6^{10}$, dass im Fall $\vartheta = 0,6$ bei allen 10 Versuchspersonen ein Heilerfolg eintritt. In diesem Fall liefert aber die oben angegebene Entscheidungsvorschrift die Entscheidung *M2 ist besser als M1*, obwohl sie falsch ist!

Darüber hinaus ermöglicht es die Wahrscheinlichkeitstheorie, unter allen Entscheidungsfunktionen diejenigen zu bestimmen, die ein vorgegebenes *Optimalitätskriterium* erfüllen. Derartige optimale Lösungen sind natürlich für die Praxis von größter Bedeutung. Tatsächlich ist die Bestimmung *optimaler statistischer Entscheidungsverfahren* ein wesentlicher Gegenstand der Mathematischen Statistik.

Kapitel 1

GRUNDLAGEN

1.1 Ausgangssituation statistischer Entscheidungen

Jeder statistischen Entscheidung liegt ein Datenmaterial $x_1, \ldots, x_n$ zugrunde. Dieses denken wir uns zu einer *Beobachtung* $x = (x_1, \ldots, x_n)$ zusammengefaßt, die wir als Realisierung einer Zufallsgröße $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{X}, \mathcal{B})$ auffassen. $(\mathcal{X}, \mathcal{B})$ heißt auch *Stichprobenraum*, x *Stichprobe*. Also:

$(\Omega, \mathcal{A}, \mathbb{P})$ ist ein Wahrscheinlichkeitsraum,
$(\mathcal{X}, \mathcal{B})$ ist ein meßbarer Raum,
$X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{B})$ ist meßbare Abbildung,
$x = X(\omega)$.

Mit der Verteilung $P := \mathbb{P} * X$ von X, d.h.

$$P(B) := (\mathbb{P} * X)(B) = \mathbb{P}(X^{-1}(B)), \quad B \in \mathcal{B},$$

ist $(\mathcal{X}, \mathcal{B}, P)$ ebenfalls ein Wahrscheinlichkeitsraum.
Im Gegensatz zur Wahrscheinlichkeitstheorie ist es ein spezifischer Aspekt der Mathematischen Statistik, dass die zugrundeliegende Verteilung P als unbekannt anzusehen und aufgrund der Beobachtung x eine Aussage über P zu machen ist. Häufig wird man jedoch gewisse Vorinformationen darüber haben, welche Verteilungen überhaupt in Frage kommen.

Definition 1.1.1. Unter einer *Verteilungsannahme* versteht man die Auszeichnung einer Klasse $\mathcal{P}$ von Verteilungen über einem Stichprobenraum $(\mathcal{X}, \mathcal{B})$. Dann heißt $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ ein *statistischer Raum.*

Aus technischen Gründen indiziert man die Elemente $P \in \mathcal{P}$ häufig durch einen *Parameter* ϑ. Die Gesamtheit Θ der zugelassenen Parameterwerte heißt *Parameterraum.* Es gilt also

$$\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}.$$

Ist X eine Zufallsgröße mit Verteilung P_ϑ, so schreiben wir für den Erwartungswert, Varianz, Verteilungsfunktion, Dichte etc. von X

$$E_\vartheta,\ \sigma^2_\vartheta,\ F_\vartheta,\ f_\vartheta \quad \text{etc.}$$

Eine Verteilungsklasse $\mathcal{P} = \{P_\vartheta :\ \vartheta \in \Theta\}$ heißt *k–parametrig*, wenn sie sich „zwanglos" durch einen k–dimensionalen Parameter ϑ parametrisieren läßt. So ist etwa die Familie der eindimensionalen Normalverteilungen

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

eine zwei-parametrige Klasse mit Parameter $\vartheta = (\mu, \sigma^2)$.
Allgemeiner interessiert von einer Verteilungsklasse $\mathcal{P}$ oft nur der Wert $\kappa(P)$ eines *Funktionals*

$$\kappa :\ \mathcal{P} \to \mathcal{K}$$

der unbekannten Verteilung P, etwa der Mittelwert von P.
Ist speziell die Verteilungsklasse parametrisiert, so fassen wir κ als eine Abbildung von Θ nach $\mathcal{K}$ auf. Wir schreiben also

$$\kappa(\vartheta) :\ \Theta \to \mathcal{K}$$

d.h. $\kappa(\vartheta)$ statt $\kappa(P_\vartheta)$.

In den meisten Anwendungen ist Θ eine Teilmenge des $\mathbb{R}^k$.
Erscheint im Rahmen des konkreten Problems eine solche parametrische Verteilungsannahme als zu einschneidend, so wird man z.B. bei Problemen mit einer stetigen Verteilung typischerweise alle bzgl. des n–dimensionalen Lebesgue-Maßes λ^n absolut stetigen Wahrscheinlichkeitsmaße auf $(\mathbb{R}^n, \mathbb{B}^n)$ (also alle Wahrscheinlichkeitsverteilungen auf der Borel–σ–Algebra $\mathbb{B}^n$ des $\mathbb{R}^n$ mit einer Dichte bzgl. λ^n) bei der Verteilungsannahme zulassen müssen. Man spricht in diesem Fall von einer *nichtparametrischen Verteilungsannahme.*
Neben der Verteilungsannahme ist noch die Gesamtheit der Aussagen anzugeben, zwischen denen entschieden werden soll. Die Gesamtheit Δ dieser Aussagen, versehen mit einer σ–Algebra $\mathcal{D}$ heißt der *Entscheidungsraum* $(\Delta, \mathcal{D})$. Die Elemente d von Δ heißen *Entscheidungen.* Aufgabe ist es also, ein statistisches Entscheidungsverfahren anzugeben, d.h. eine Vorschrift, die jeder möglichen Beobachtung $x \in \mathcal{X}$ eindeutig eine Entscheidung

$$d = e(x) \in \Delta$$

zuordnet.

Definition 1.1.2. Eine (nicht-randomisierte) *Entscheidungsfunktion e* ist eine $\mathcal{B}, \mathcal{D}$–meßbare Abbildung des Stichprobenraumes $(\mathcal{X}, \mathcal{B})$ in den Entscheidungsraum $(\Delta, \mathcal{D})$. Deren Gesamtheit bezeichnen wir mit $\mathcal{E}$.

Je nach der Struktur des Entscheidungsraumes unterscheidet man zwischen verschiedenen Grundtypen statistischer Entscheidungsverfahren. Die beiden wichtigsten sind die *Tests* und die *Schätzverfahren*.
In den folgenden Abschnitten wollen wir diese Entscheidungsverfahren näher untersuchen. Dabei gehen wir stets von einer parametrischen Verteilungsannahme $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ aus.
Wir verstehen dann unter einer Hypothese H stets eine Aussage (Annahme) über den Parameter ϑ. Dabei werden wir H mit derjenigen Teilmenge des Parameterraumes Θ, für die H gilt, identifizieren.

Definition 1.1.3. Gegeben sei eine Verteilungsannahme $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ und ein Entscheidungsraum $(\Delta, \mathcal{D})$. Dann heißt eine Funktion

$$L : \Theta \times \Delta \to [0, \infty)$$

Verlustfunktion, falls gilt:

$$\forall \vartheta \in \Theta : \ L(\vartheta, \cdot) \text{ ist } \mathcal{D}, \mathbb{B}\text{–meßbar.}$$

Bemerkung 1.1.4. $L(\vartheta, d)$ drückt den Verlust (Schaden) aus, den man bei Treffen der Entscheidung d und gleichzeitigem Vorliegen von P_ϑ erleidet.

Definition 1.1.5. Es sei L eine Verlustfunktion. Dann heißt die Funktion $R : \Theta \times \mathcal{E} \to [0, \infty]$, definiert durch

$$R(\vartheta, e) := \int_{\mathcal{X}} L(\vartheta, e(x))\, P_\vartheta(dx),$$

Risikofunktion (erwarteter Verlust) bzgl. L und e.

$R(\vartheta, e)$ ist also der erwartete Verlust bei Vorliegen von P_ϑ und Entscheidungsfunktion e.

In der folgenden Definition legen wir *Optimalitätskriterien* für Entscheidungsfunktionen fest.

Definition 1.1.6.

(i) e^* heißt *gleichmäßig beste (n.r.) Entscheidungsfunktion*, falls

$$\forall \vartheta \in \Theta : \ R(\vartheta, e^*) = \min_{e \in \mathcal{E}} R(\vartheta, e).$$

(ii) $\tilde{e}$ heißt eine *Mini-Max Lösung* bzgl. $\mathcal{E}$, falls

$$\sup_{\vartheta \in \Theta} R(\vartheta, \tilde{e}) = \min_{e \in \mathcal{E}} \sup_{\vartheta \in \Theta} R(\vartheta, e).$$

1.2 Tests

Aufgrund einer vorliegenden Beobachtung $x \in \mathcal{X}$ soll zwischen *zwei* Aussagen entschieden werden:

$$\vartheta \in H \text{ oder } \vartheta \in K,$$

wobei

$$\Theta = H \cup K, \quad H \cap K = \emptyset.$$

Bezeichen wir die Entscheidungen für H bzw. K mit d_H bzw. d_K, so definieren wir mit

$$\Delta := \{d_H, d_K\},\ \mathcal{D} := \text{Potenzmenge von } \Delta$$

einen Entscheidungsraum.
Dann ist eine Abbildung $e : \mathcal{X} \to \Delta$ genau dann eine (n.r.) Entscheidungsfunktion, wenn gilt

$$S := \{x \in \mathcal{X} :\ e(x) = d_K\} = e^{-1}(\{d_k\}) \in \mathcal{B},$$

($\Leftrightarrow S^c \in \mathcal{B}$). Dies ist gerade die Meßbarkeit von e.

Definition 1.2.1. Eine Entscheidungsfunktion der Form

$$e(x) = \begin{cases} d_K, & \text{falls } x \in S, \\ d_H, & \text{falls } x \in S^c, \end{cases} \qquad x \in X,$$

mit $S \in \mathcal{B}$, heißt (n.r.) *Test* für das Entscheidungsproblem H gegen K.

Man wird natürlich versuchen, $S \in \mathcal{B}$ bzw. e so zu wählen, dass möglichst wenige Fehlentscheidungen getroffen werden.
Zwei Arten von Fehlern sind dabei möglich:

Fehler 1. Art:	Entscheidung für K, obwohl H richtig ist,
Fehler 2. Art:	Entscheidung für H, obwohl K richtig ist.

Die übliche (unsymmetrische) Vorgehensweise ist die folgende: Man versucht unter allen (n.r.) Tests mit einer *vorgegebenen Irrtumswahrscheinlichkeit* $\alpha \in (0,1)$ für den Fehler 1. Art einen solchen zu bestimmen, der die Wahrscheinlichkeit für den Fehler 2. Art minimiert: Gesucht ist also $S^* \in \mathcal{B}$ mit

$$S^* \in \gamma_\alpha := \{S \in \mathcal{B} : \forall \vartheta \in H :\ P_\vartheta(S) \leq \alpha\} \tag{1.2.2}$$

und

$$\forall \vartheta \in K :\ P_\vartheta(S^{*c}) = \inf_{S \in \gamma_\alpha} P_\vartheta(S^c). \tag{1.2.3}$$

Äquivalent zu (1.2.3) ist

$$\forall \vartheta \in K : P_\vartheta(S^*) = \sup_{S \in \gamma_\alpha} P_\vartheta(S). \tag{1.2.4}$$

Die Elemente der Klasse γ_α, d.h. Tests $S \in \mathcal{B}$ mit der Eigenschaft

$$\forall \vartheta \in H : P_\vartheta(S) \leq \alpha \tag{1.2.5}$$

heißt *n.r. Test zum Niveau* α. $\alpha \in [0,1]$ heißt *Irrtumswahrscheinlichkeit* oder *Wahrscheinlichkeit für den Fehler 1. Art*, $1-\alpha$ heißt *Sicherheitswahrscheinlichkeit*.
Man bezeichnet H auch als *Nullhypothese* oder *Hypothese* und K als *Gegenhypothese* oder *Alternative*. Man nennt

$$S := \{x \in \mathcal{X} \ : \ e(x) = d_K\} =: \{e = d_K\}$$

die *kritische Region* und

$$S^c = \{e = d_H\}$$

den *Annahmebereich* des Tests e.

1.3 Elementare Testverfahren unter Normalverteilungsannahme

Im Folgenden seinen $X_1, \ldots, X_n$ unabhängige und identisch $N(\mu, \sigma^2)$-verteilte Zufallsvariablen, $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^n, \mathbb{B}^n)$. Dabei bezeichnet

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

$B \in \mathbb{B}$, $\mu \in \mathbb{R}$, $\sigma > 0$, die Normalverteilung auf $(\mathbb{R}, \mathbb{B})$ mit Mittelwert μ und Varianz σ^2.

Wir unterscheiden im Folgendem verschiedene Fälle:

(i) Es sei μ unbekannt, σ^2 bekannt. Die parametrische Verteilungsannahme lautet in diesem Fall

$$\mathcal{P} = \{P_\vartheta = N(\mu, \sigma^2)^n \ : \ \vartheta = \mu \in \Theta\}, \Theta = \mathbb{R}$$

Zu vorgegebenen Niveau, d.h. Fehlerwahrscheinlichkeit 1. Art $\alpha \in (0,1)$, ist ein (einseitiger) Test für $H : \vartheta \geq \vartheta_0$ gegen $K : \vartheta \leq \vartheta_0$ zu finden. Dabei ist $\vartheta_0 \in \mathbb{R}$ fest vorgegeben. Aufgrund des starken Gesetzes der großen Zahlen gilt

$$T(X_1, \ldots, X_n) := \frac{1}{n} \sum_{i=1}^n X_i \overset{n\to\infty}{\longrightarrow} E(X_1) = \mu \quad \mathbb{P} - f.s.$$

Es ist daher sinnvoll, H abzulehnen, falls aufgrund vorliegender Beobachtungen

$$\boldsymbol{x} = (x_1, \dots x_n) = (X_1(\omega), \dots X_n(\omega))$$

die Statistik $T(X_1, \dots, X_n)$ „zu klein“ ist, d.h. $T(X_1, \dots, X_n) \leq \gamma < \vartheta_0$. Dabei ist γ so zu wählen, dass die Wahrscheinlichkeit für den Fehler 1. Art nicht größer als α ist, d.h.

$$\sup_{\vartheta \in H} P_\vartheta(T(X_1, \dots X_n) \leq \gamma) \leq \alpha. \tag{1.3.1}$$

Nach dem Faltungstheorem der Normalverteilung gilt für $\vartheta \in \Theta$

$$P_\vartheta\left(\sum_{i=1}^n X_i \in B\right) = N(n\mu, n\sigma^2)(B),\ B \in \mathcal{B}$$

$$\Rightarrow P_\vartheta\left(n^{1/2}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)/\sigma \in B\right) = N(0,1)(B)$$

d.h.

$$P_\vartheta\left(n^{1/2}(T(X_1, \dots X_n) - \mu)/\sigma \in B\right) = N(0,1)(B).$$

Daher gilt für alle $t \in \mathbb{R}$ und $T = T(X_1, \dots, X_n)$ und $\vartheta \in \Theta$

$$P_\vartheta\left(T \leq \vartheta + \frac{t\sigma}{n^{1/2}}\right) = P_\vartheta\left(n^{1/2} \cdot \frac{T - \vartheta}{\sigma} \leq t\right) = \Phi(t), \quad B = (-\infty; t]$$

wobei

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left(-x^2/2\right)\, dx$$

die Verteilungsfunktion der Standardnormalverteilung ist.

Für $\vartheta \in H$ erhalten wir nun

$$\begin{aligned} P_\vartheta\left(T \leq \vartheta_0 + \frac{t\sigma}{n^{1/2}}\right) &= P_\vartheta\left(T \leq \vartheta + \frac{t\sigma}{n^{1/2}} - \underbrace{(\vartheta - \vartheta_0)}_{\geq 0}\right) \\ &\overset{\vartheta \geq \vartheta_0}{\leq} P_\vartheta\left(T \leq \vartheta + \frac{t\sigma}{n^{1/2}}\right) \\ &= \Phi(t), \end{aligned}$$

d.h.

$$\sup_{\vartheta \in H} P_\vartheta\left(T \leq \vartheta_0 + \frac{t\sigma}{n^{1/2}}\right) \leq \Phi(t)$$

Wählen wir also $t \in \mathbb{R}$ so, dass $\Phi(t) = \alpha$, d.h. $t = \Phi^{-1}(\alpha) =: u_\alpha$, so haben wir (1.3.1) mit der Wahl $\gamma := \vartheta_0 + u_\alpha \sigma n^{-1/2}$ erfüllt. Dabei ist u_α das *α-Quantil der Standardnormalverteilung*.

Die kritische Region unseres Tests ist also die Menge

$$C = \left\{ \boldsymbol{x} \in \mathbb{R}^n : T(\boldsymbol{x}) \leq \vartheta_0 + \frac{u_\alpha \sigma}{n^{1/2}} \right\}$$

Dieser Test heißt auch *Gauss-Test.*

(ii) Es sei σ^2 unbekannt, μ hingegeen bekannt. Die Verteilungsannahme lautet in diesem Fall

$$\mathcal{P} = \left\{ P_\vartheta = N(\mu, \sigma^2)^n : \vartheta = \sigma^2 \in \Theta \right\}, \Theta = (0, \infty),$$

Gesucht ist nun ein einseitiger Test zum Niveau α für $H : \vartheta \geq \vartheta_0$ gegen $K : \vartheta < \vartheta_0$. Dabei ist $\vartheta_0 > 0$ vorgegeben.

Das Starke Gesetz der großen Zahlen leifert

$$\hat{T} := \hat{T}(X_1, \dots X_n) := \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \overset{n\to\infty}{\longrightarrow} E\left((X_1 - \mu)^2\right) = \sigma^2 \quad \mathbb{P}-f.s.$$

Wir werden daher H ablehnen, falls aufgrund einer vorliegenden Beobachtung

$$\boldsymbol{x} = (x_1, \dots x_n) = (X_1(\omega), \dots, X_n(\omega))$$

die Testgröße $\hat{T}(\boldsymbol{x})$ „zu klein“ wird, d.h. $\hat{T}(\boldsymbol{x}) \leq \gamma \leq \vartheta_0$. Dabei ist γ so festzulegen, dass

$$\sup_{\vartheta \in H} P_\vartheta \left(\hat{T} \leq \gamma \right) \leq \alpha. \tag{1.3.2}$$

Wir gehen wie folgt vor: Ist X_i $N(\mu, \sigma^2)$-verteilt, so ist $(X_i - \mu)/\sigma$ $N(0,1)$-verteilt. Damit ist $\sum_{i=1}^n (X_i - \mu)^2/\sigma^2$ nach Definition χ^2-verteilt mit n Freiheitsgraden, i. Z. χ^2_n.

Bezeichnen wir mit $F_{\chi^2_n}$ die Verteilungsfunktion der χ^2-Verteilung mit n Freiheitsgraden und setzen wir noch

$$c_{\alpha,n} := F^{-1}_{\chi^2_n}(\alpha),$$

so gilt mit $\gamma := c_{\alpha,n} \cdot \vartheta_0/n$ für alle $\vartheta \in H = [\vartheta_0, \infty)$

$$\begin{aligned}
P_\vartheta \left(\hat{T} \leq \gamma \right) &= P_\vartheta \left(\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \mu)^2 \leq \frac{c_{\alpha,n} \cdot \vartheta_0}{n} \right) \\
&= P_\vartheta \left(\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sqrt{\vartheta}} \right)^2 \leq c_{\alpha,n} \underbrace{\frac{\vartheta_0}{\vartheta}}_{\leq 1} \right) \\
&\leq P_\vartheta \left(\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sqrt{\vartheta}} \right)^2 \leq c_{\alpha,n} \right) \\
&= F_{\chi^2_n}(c_{\alpha,n}) \\
&= \alpha,
\end{aligned}$$

d.h. (1.3.2) ist erfüllt. Die kritische Region unseres Tests ist also die Menge

$$C = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \hat{T}(\boldsymbol{x}) \leq \frac{c_{\alpha,n} \cdot \vartheta_0}{n} \right\}$$

(iii) Es sei nun μ *und* σ^2 unbekannt.
Die Verteilungsannahme lautet nun

$$\mathcal{P} = \left\{ P_\vartheta = N(\mu, \sigma^2)^n : \vartheta = (\mu, \sigma^2) \in \Theta \right\}, \Theta = \mathbb{R} \times (0, \infty).$$

Das starke Gesetz der großen Zahlen liefert mit $\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$

$$\begin{aligned} S^2 &:= S^2(X_1, \ldots X_n) \qquad (1.3.3) \\ &:= \frac{1}{n-1} \sum_{i=1}^n \left(X_i - \overline{X}_n \right)^2 \\ &= \frac{1}{n-1} \sum_{i=1}^n \left((X_i - \mu) + (\mu - \overline{X}_n) \right)^2 \\ &= \underbrace{\frac{n}{n-1}}_{\stackrel{n\to\infty}{\longrightarrow} 1} \left\{ \underbrace{\left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right)}_{\stackrel{n\to\infty}{\longrightarrow} \sigma^2} - \underbrace{\left(\overline{X}_n - \mu \right)^2}_{\stackrel{n\to\infty}{\longrightarrow} 0} \right\} \\ &\stackrel{n\to\infty}{\longrightarrow} \sigma^2 \quad \mathbb{P} - f.s., \end{aligned}$$

falls X_i die Verteilung P_ϑ besitzen mit $\vartheta = (\mu, \sigma^2)$.
Insbesondere gilt

$$E_\vartheta(S^2) = \sigma^2$$

Obige Konvergenzaussage nehmen wir zur Grundlage für die Definition von Tests für die beiden Entscheidungsprobleme.

(a) $H : \sigma^2 \geq \sigma_0^2$ gegen $K : \sigma^2 < \sigma_0^2$, wobei $\sigma_0^2 > 0$ fest vorgegeben ist

(b) $H : \mu \leq \mu_0$ gegen $K : \mu > \mu_0$, wobei $\mu_0 \in \mathbb{R}$ fest vorgegeben ist

Zu (a): Die Konvergenzaussage (1.3.3) legt es nahe H abzulehnen, falls aufgrund einer vorliegenden Beobachtung

$$\boldsymbol{x} = (x_1, \ldots x_n) = (X_1(\omega), \ldots, X_n(\omega))$$

gilt:

$$S^2(\boldsymbol{x}) \leq \gamma < \sigma_0^2,$$

wobei γ so zu wählen ist, dass

$$\sup_{\vartheta \in H} P_\vartheta \left(S^2 \leq \gamma\right) \leq \alpha. \tag{1.3.4}$$

Beachte dabei dass $H = \mathbb{R} \times [\sigma_0^2, \infty)$.

Nun ist $(n-1) \cdot S^2/\sigma^2$ χ^2_{n-1}-verteilt (siehe etwa Theorem 2.2.1 in Falk et. al. (2002)). Setzen wir daher $\gamma := c_{\alpha,n-1}\sigma_0^2/(n-1)$, so gilt für alle $\vartheta \in H$

$$\begin{aligned} P_\vartheta(S^2 \leq \gamma) &= P_\vartheta\left(\frac{n-1}{\sigma^2}S^2 \leq c_{\alpha,n-1}\underbrace{\left(\frac{\sigma_0}{\sigma}\right)^2}_{\leq 1}\right) \\ &\leq P_\vartheta\left(\frac{n-1}{\sigma^2}S^2 \leq c_{\alpha,n-1}\right) \\ &= \alpha \end{aligned}$$

d.h. (1.3.4) ist erfüllt. Die kritische Region unseres Tests ist also die Menge

$$C = \left\{ \boldsymbol{x} \in \mathbb{R}^n : S^2(\boldsymbol{x}) \leq \frac{c_{\alpha,n-1} \cdot \sigma_0^2}{n-1} \right\}$$

Zu (b): Da die Verteilung von $T = \frac{1}{n}\sum_{n-1}^{n}(X_i - \mu)$ von σ^2 abhängt und im vorliegenden Fall σ^2 unbekannt ist liegt es wegen (1.3.3) nahe die Stichprobenfunktion

$$t(x_1, \ldots x_n) := \frac{n^{1/2} \cdot (T(x_1, \ldots, x_n) - \mu)}{\left(S^2(x_1, \ldots, x_n)\right)^{1/2}}$$

zu verwenden. Als Entscheidungsregel verwenden wir:

H wird abgelehnt, falls aufgrund einer vorliegenden Beobachtung $\boldsymbol{x} = (x_1, \ldots x_n)$

$$t_0(\boldsymbol{x}) = \frac{n^{1/2}\left(T(\boldsymbol{x}) - \mu_0\right)}{\left(S^2(\boldsymbol{x})\right)^{1/2}} \geq \gamma > 0.$$

Dabei ist γ so zu wählen, dass

$$\sup_{\vartheta \in H} P_\vartheta(t_0 \geq \gamma) \leq \alpha \tag{1.3.5}$$

Beachte, dass $H = (-\infty, \mu_0] \times (0, \infty)$.

Es gilt

$$t(\boldsymbol{x}) = (n-1)^{1/2} \frac{\frac{T(\boldsymbol{x}) - \mu}{\sigma}}{\left(\frac{n-1}{\sigma^2}S^2(\boldsymbol{x})\right)^{1/2}},$$

wobei $n^{1/2}\,(T(X_1, \dots X_n) - \mu)\,/\sigma$ $N(0,1)$-verteilt ist und $((n-1)/\sigma^2)S^2(X_1, \dots X_n)$ χ^2_{n-1}-verteilt ist, falls $X_1, \dots X_n$ unabhängig und identisch $N(\mu, \sigma^2)$.

Diese beiden Zufallsvariablen sind stochastisch unabhängig und die Verteilung von $t(X_1, \dots X_n)$ ist die *(Studentsche) t-Verteilung mit $n-1$ Freiheitsgraden*, i.Z. t_{n-1} (s. Falk et al. (2002), Theorem 2.2.1).
Ist nun $\gamma := d_{1-\alpha,n-1} := t^{-1}_{n-1}(1-\alpha)$ das $1-\alpha$-Quantil der t_{n-1}-Verteilung, so erhalten wir für alle $\vartheta \in H$

$$\begin{aligned}
P_\vartheta(t_0 \geq \gamma) &= P_\vartheta\left(t + n^{1/2}\underbrace{\frac{\mu - \mu_0}{(S^2)^{1/2}}}_{\leq 0} \geq \gamma\right) \\
&\leq P_\vartheta(t \geq \gamma) \\
&= t_{n-1}([\gamma, \infty)) \\
&= t_{n-1}([d_{1-\alpha,n-1}, \infty)) \\
&= 1 - (1-\alpha) \\
&= \alpha,
\end{aligned}$$

d.h. (1.3.5) ist erfüllt. Die kritische Region dieses (Einstichproben) *t-Tests* ist

$$C := \{\boldsymbol{x} \in \mathbb{R}^n : t_0(\boldsymbol{x}) \geq d_{1-\alpha,n-1}\}.$$

1.4 Punktschätzverfahren

Bei vorgegebenem Stichprobenraum $(\mathcal{X}, \mathcal{B})$ und zugrundegelegter Verteilungsannahme $P = \{P_\vartheta : \vartheta \in \Theta\}$ soll aufgrund einer vorliegenden Beobachtung $x \in \mathcal{X}$ der zugrundeliegende Parameter $\vartheta \in \Theta$ oder allgemein der Wert $\kappa(\vartheta)$ einer (reellen) Funktion κ auf Θ geschätzt werden.

Definition 1.4.1. Eine meßbare Abbildung $\hat{\kappa}$ des Stichprobenraumes $(\mathcal{X}, \mathcal{B})$ in den (meßbaren) Wertebereich der Funktion κ heißt eine *Schätzfunktion*, genauer eine Punktschätzfunktion für $\kappa(\vartheta)$, kurz $\hat{\kappa} : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$.

Bei einer stetig verteilten Schätzfunktion wird jeder spezielle Wert, insbesondere der zugrundeliegende Wert $\kappa(\vartheta)$ mit Wahrscheinlichkeit 0 angenommen

$$P_\vartheta\left(\hat{\kappa} = \kappa(\vartheta)\right) = 0$$

Man trifft in diesem Fall somit fast sicher (= mit Wahrscheinlichkeit 1) eine Fehlentscheidung.

Bei der Wahl einer Schätzfunktion $\hat{\kappa}$ sollte daher die Größe des erwarteten Fehlers berücksichtigt werden. Im Folgenden setzen wir vorraus dass $\kappa : \Theta \to \mathbb{R}$.

Definition 1.4.2. $\hat{\kappa} : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$ heißt *erwartungstreu* oder *unverzerrt* (engl. unbiased), falls

$$\forall \vartheta \in \Theta : \quad E_\vartheta(\hat{\kappa}) = \kappa(\vartheta)$$

Definition 1.4.3. $\hat{\kappa^*}$ heißt erwartungstreue Schätzfunktion mit Minimalvarianz, falls

(i) $\hat{\kappa^*} \in \hat{K} := \{\hat{\kappa} : \forall \vartheta \in \Theta : E_\vartheta(\hat{\kappa}) = \kappa(\vartheta)\}$,

(ii) $\forall \vartheta \in \Theta \quad Var_\vartheta(\hat{\kappa^*}) = E_\vartheta\left((\hat{\kappa^*} - \mu)^2\right) = \inf_{\hat{\kappa} \in \hat{K}} Var_\vartheta(\hat{\kappa})$.

Die Bedeutung der Minimalvarianz lässt sich wie folgt motivieren. Es sei $L : \Theta \times \mathbb{R} \to [0, \infty)$ eine Verlustfunktion mit den beiden Eigenschaften

$\forall \vartheta \in \Theta : L(\vartheta, \cdot)$ ist zweimal stetig diffbar

$\forall \vartheta \in \Theta : L(\vartheta, \kappa(\vartheta)) = 0$

Damit wird für eine beliebige erwartungstreue Schätzfunktion $\hat{\kappa}$ von κ nach der Taylorformel gelten:

$$\begin{aligned}
L(\vartheta, \hat{\kappa}) &= L(\vartheta, \hat{\kappa}) - L(\vartheta, \kappa(\vartheta)) \\
&= \frac{\partial}{\partial \kappa} L(\vartheta, \kappa(\vartheta)) \cdot (\hat{\kappa} - \kappa(\vartheta)) + \frac{\partial^2}{\partial^2 \kappa} L(\vartheta, \xi) \frac{(\hat{\kappa} - \kappa(\vartheta))^2}{2} \\
&\approx \frac{\partial}{\partial \kappa} L(\vartheta, \kappa(\vartheta)) \cdot (\hat{\kappa} - \kappa(\vartheta)) + \frac{\partial^2}{\partial^2 \kappa} L(\vartheta, \kappa(\vartheta)) \frac{(\hat{\kappa} - \kappa(\vartheta))^2}{2} \\
&= \frac{\partial}{\partial \kappa} L(\vartheta, \kappa(\vartheta)) \cdot (\hat{\kappa} - \kappa(\vartheta)) + const(\vartheta) \cdot (\hat{\kappa} - \kappa(\vartheta))^2
\end{aligned}$$

wobei ξ zwischen $\hat{\kappa}$ und κ liegt und $const(\vartheta)$ unabhängig von $\hat{\kappa}$ ist. Es folgt:

$$\begin{aligned}
R_\vartheta(\hat{\kappa}) &= E_\vartheta\left(L(\vartheta, \hat{\kappa})\right) \\
&\approx E_\vartheta\left(\frac{\partial}{\partial \kappa} L(\vartheta, \kappa(\vartheta)) \cdot (\hat{\kappa} - \kappa(\vartheta)) + const(\vartheta) \cdot (\hat{\kappa} - \kappa(\vartheta))^2\right) \\
&= \frac{\partial}{\partial \kappa} L(\vartheta, \kappa(\vartheta)) \underbrace{E_\vartheta\left(\hat{\kappa} - \kappa(\vartheta)\right)}_{=0} + const(\vartheta) \cdot E_\vartheta\left((\hat{\kappa} - \kappa(\vartheta))^2\right) \\
&= const(\vartheta) \cdot Var_\vartheta(\hat{\kappa})
\end{aligned}$$

Ein erwartungstreuer Schätzer mit Minimalvarianz wird also tendenziell jedes Risiko minimieren. Das erklärt die Bedeutung der Minimalvarianz.

Beispiel 1.4.4. Es seinen $X_1, \dots X_n$ unabhängige und identisch $N(\mu, \sigma^2)$-verteilte Zufallsvariablen, wobei μ und σ^2 unbekannt sind.
Es gilt also $(\mathcal{X}, \mathcal{B}, \mathcal{P}) = (\mathbb{R}^n, \mathbb{B}^n, \mathcal{P} = \{P_\vartheta = N(\mu, \sigma^2)^n : \vartheta = (\mu, \sigma^2) \in \Theta\}), \Theta = \mathbb{R} \times (0, \infty)$. Gesucht ist eine Schätung für $\kappa(\vartheta) = \mu$.
Das Gesetz der großen Zahlen legt die Schätzfunktion

$$\hat{\kappa} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n, \text{ für } \kappa(\vartheta) \text{ nahe.}$$

Tatsächlich ist $\hat{\kappa}$ erwartungstreu:

$$\forall \vartheta \in \Theta : E_\vartheta(\hat{\kappa}(X_1, \dots X_n)) = E_\vartheta\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \mu = \kappa(\vartheta).$$

Wir werden in Kapitel 3 zeigen, dass $\hat{\kappa}$ auch 1.4.3 (ii) erfüllt, also ein Schätzer mit Minimalvarianz ist. Zum Nachweis wird dabei wesentlich von der Normalverteilungsannahme Gebrauch gemacht.

Beispiel 1.4.5. Es seien $X_1, \dots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit $\sigma^2 := Var(X_1) < \infty$. Setze

$$\begin{aligned}
\hat{\kappa_1}(\boldsymbol{x}) &:= \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2, \\
\hat{\kappa_2}(\boldsymbol{x}) &:= \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2, \\
\boldsymbol{x} &= (x_1, \dots x_n) \in \mathbb{R}^n, \ \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\end{aligned}$$

Dann liefert $\hat{\kappa_2}(X_1, \dots, X_n)$ eine erwartungstreue Schätzfunktion für σ^2, $\hat{\kappa_1}(X_1, \dots, X_n)$ aber nicht.

1.5 Bereichsschätzungen

Der Vorteil der in 1.4 betrachteten Schätzverfahren präzise Aussagen in Form von Punktschätzungen zu liefern bedingt gleichzeitig, dass in (nahezu) allen Fällen f.s. Fehlentscheidungen getroffen werden.
Eine Alternative zur Punktschätzung besteht in der *Bereichsschätzung* (*Konfidenzbereich*) d.h. in der Angabe einer Abbildung

$$K : \mathcal{X} \to \text{Potenzmenge von } \Theta$$

derart, dass $K(\cdot)(\subset \Theta)$ mit einer vorgegebenen Mindestwahrscheinlichkeit $1-\alpha$ den zugrundeliegenden Parameter enthält

$$\forall \vartheta \in \Theta : P_\vartheta(x \in \mathcal{X} : \vartheta \in K(x)) \geq 1-\alpha$$

Beispiel 1.5.1. Es seien $X_1, \ldots, X_n$ unabhängige und identisch $N(\mu, \sigma_0^2)$-verteilte Zufallsvariablen mit bekanntem $\sigma_0^2 > 0$ und unbekanntem $\mu \in \mathbb{R}$

$$(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^n, \mathbb{B}^n), \quad \mathcal{P} = \{P_\vartheta = N(\mu, \sigma_0^2)^n, \vartheta = \mu \in \Theta = \mathbb{R}\}$$

Setze $T(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^n x_i$, $\boldsymbol{x} = (x_1, \ldots x_n) \in \mathcal{X}$.
Da mit $T = T(X_1, \ldots, X_n)$ unter ϑ die Größe $n^{1/2}(T-\mu)/\sigma_0$ $N(0,1)$-verteilt ist, gilt für ein beliebiges $\alpha \in (0,1)$

$$\forall \vartheta \in \Theta : P_\vartheta\left(-u_{\alpha/2} \leq n^{1/2}\frac{T-\mu}{\sigma_0} \leq u_{\alpha/2}\right) = 1-\alpha$$

wobei $u_{\alpha/2} = \Phi^{-1}\left(1-\frac{\alpha}{2}\right)$
oder

$$\begin{aligned}\forall \vartheta \in \Theta : P_\vartheta\left(T - \frac{\sigma_0\, u_{\alpha/2}}{n^{1/2}} \leq \mu \leq T + \frac{\sigma_0\, u_{\alpha/2}}{n^{1/2}}\right) &= 1-\alpha \\ = P_\vartheta\left(\underbrace{\vartheta}_{=\mu} \in \left[T - \frac{\sigma_0\, u_{\alpha/2}}{n^{1/2}}; T + \frac{\sigma_0\, u_{\alpha/2}}{n^{1/2}}\right]\right).\end{aligned}$$

Dies bedeutet, dass

$$K(\boldsymbol{x}) := \left[T(\boldsymbol{x}) - \frac{\sigma_0\, u_{\alpha/2}}{n^{1/2}}; T(\boldsymbol{x}) + \frac{\sigma_0\, u_{\alpha/2}}{n^{1/2}}\right]$$

$\boldsymbol{x} \in \mathbb{R}^n$, eine Bereichsschätzung ist mit der Eigenschaft

$$P_\vartheta\left(\vartheta \in K(X_1, \ldots, X_n)\right) = 1-\alpha,$$

$\vartheta \in \Theta$. Man nennt $\left[T - \sigma_0 u_{\alpha/2} n^{-1/2}; T + \sigma_0 u_{\alpha/2} n^{-1/2}\right]$ *Konfidenzintervall* (Vertrauensintervall) für den Parameter ϑ zum Niveau $1-\alpha$.

Beispiel 1.5.2. Es seien $X_1, \ldots, X_n$ unabhängige und $N(\mu, \sigma^2)$-verteilte Zufallsvariablen, wobei μ und $\sigma^2 > 0$ unbekannt sind, d.h. $\Theta = \mathbb{R} \times (0, \infty)$.
Die Zufallsvariable

$$t(X_1, \ldots, X_n) := n^{1/2}\,\frac{T(X_1, \ldots, X_n) - \mu}{\sqrt{S^2(X_1, \ldots, X_n)}}$$

ist nach Abschnitt 1.3 t-verteilt mit $n-1$-Freiheitsgraden, i.Z. t_{n-1}.
Für $\alpha \in (0,1)$ sei $t_{\alpha/2} := d_{1-\frac{\alpha}{2},n}$ das $1-\frac{\alpha}{2}$-Quantil der t_{n-1}-Verteilung, d.h.

$$P_\vartheta(t(X_1,\dots,X_n) \le t_{\alpha/2}) = 1 - \frac{\alpha}{2}.$$

Dann gilt $\forall \vartheta = (\mu, \sigma^2) \in \Theta$:

$$\begin{aligned}
&P_\vartheta\left(-t_{\alpha/2} \le t(X_1,\dots,X_n) \le t_{\alpha/2}\right) \\
&= P_\vartheta\left(t(X_1,\dots,X_n) \le t_{\alpha/2}\right) - P_\vartheta\left(t(X_1,\dots,X_n) \le -t_{\alpha/2}\right) \\
&= 1 - \frac{\alpha}{2} - \frac{\alpha}{2} \\
&= 1 - \alpha
\end{aligned}$$

oder $\forall \vartheta = (\mu, \sigma^2) \in \Theta$:

$$\begin{aligned}
&P_\vartheta\left(T - \frac{t_{\alpha/2}(S^2)^{1/2}}{n^{1/2}} \le \mu \le T + \frac{t_{\alpha/2}(S^2)^{1/2}}{n^{1/2}}\right) \\
&= 1 - \alpha \\
&= P_\vartheta\left(\mu \in \left[T - \frac{t_{\alpha/2}(S^2)^{1/2}}{n^{1/2}}; T + \frac{t_{\alpha/2}(S^2)^{1/2}}{n^{1/2}}\right]\right)
\end{aligned}$$

d.h. $\left[T - t_{\alpha/2}(S^2)^{1/2}n^{-1/2}; T + t_{\alpha/2}(S^2)^{1/2}n^{-1/2}\right]$ ist ein Konfidenzintervall für den Parameter $\kappa(\vartheta) = \mu$ zum Niveau $1-\alpha$.

Definition 1.5.3. Eine Abbildung

$$K : \mathcal{X} \to \text{Potenzmenge von } \Theta$$

heißt *Bereichsschätzfunktion* zum Konfidenzniveau $1-\alpha$ oder *Konfidenzbereich* zum Niveau $1-\alpha$:$\Leftrightarrow$

$$\forall \vartheta \in \Theta : P_\vartheta\left(x \in \mathcal{X} : \vartheta \in K(x)\right) \ge 1 - \alpha \tag{1.5.4}$$

Bemerkung 1.5.5. (i) Damit die Wahrscheinlichkeit in (1.5.4) erklärt ist, muss gelten

$$\forall \vartheta \in \Theta : A(\vartheta) := \{x \in \mathcal{X} : \vartheta \in K(x)\} \in \mathcal{B}$$

(ii) Im Fall von Bereichsschätzungen ist der Entscheidungsraum die Potenzmenge von Θ. Er enthält im Gegensatz zu den bisher beobachteten Entscheidungsverfahren i.a. mehrere richtige Entscheidungen.

(iii) Ist Θ ein Intervall in $\mathbb{R}$ und ist für alle $x \in \mathcal{X}$ $K(x)$ ein Intervall in $\mathbb{R}$, so heißt K als Lösung von (1.5.4) ein Konfidenzintervall für ϑ zum Niveau $1-\alpha$.

Satz 1.5.6 (Dualitätsprinzip). *Es sei* $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ *eine beliebige Familie von Wahrscheinlichkeitsmaßen auf dem Stichprobenraum* $(\mathcal{X}, \mathcal{B})$ *und* $\alpha \in (0,1)$. *Dann gilt:*

(i) Ist K *ein Konfidenzbereich zum Niveau* $1-\alpha$, *so ist für jedes* $\vartheta' \in \Theta$ *die Menge* $\mathcal{X} \backslash A(\vartheta') = \{x \in \mathcal{X} : \vartheta' \notin K(x)\}$ *die kritische Region eines Tests zum Niveau* α *für das Problem* $H_{\vartheta'} : \vartheta = \vartheta'$ *gegen* $K_{\vartheta'} : \vartheta \neq \vartheta'$.

(ii) Ist für jedes $\vartheta' \in \Theta$ $C(\vartheta')$ *die kritische Region eines Tests zum Niveau* α *für das Entscheidungsproblem* $H_{\vartheta'} : \vartheta = \vartheta'$ *gegen* $K_{\vartheta'} : \vartheta \neq \vartheta'$, *so wird durch die Festsetzung* $K(x) := \{\vartheta' \in \Theta : x \in C(\vartheta')^c\}$, $x \in \mathcal{X}$, *ein Konfidenzbereich zum Niveau* $1-\alpha$ *definiert.*

Beweisskizze: Zu (i):

$$\begin{aligned} P_{\vartheta'}(\mathcal{X} \backslash A(\vartheta')) &= P_{\vartheta'}(x \in \mathcal{X} : \vartheta' \notin K(x)) \\ &= 1 - \underbrace{P_{\vartheta'}(x \in \mathcal{X} : \vartheta' \in K(x))}_{\geq 1-\alpha} \leq \alpha \end{aligned}$$

Zu (ii): $\vartheta \in K(x) \Leftrightarrow x \in C(\vartheta)^c$; wähle als $K(x)$ alle ϑ mit der Eigenschaft: aufgrund von x wird ϑ nicht verworfen, d.h.

$$\begin{aligned} P_\vartheta(x \in \mathcal{X} : \vartheta \in K(x)) &= P_\vartheta(x \in \mathcal{X} : x \in C(\vartheta)^c) \\ &= 1 - \underbrace{P_\vartheta(x \in \mathcal{X} : x \in C(\vartheta))}_{\leq \alpha} \geq 1-\alpha \end{aligned}$$

□

1.6 Randomisierte Entscheidungsverfahren

Der in Definition 1.2.1 auf Seite 8 eingeführte Begriff der *nichtrandomisierten* Entscheidungsfunktion reicht für die im Folgenden zu entwickelnde Theorie i.a. nicht aus, denn mit zwei Entscheidungsfunktionen e_1, e_2 benötigen wir auch eine solche, bei der aufgrund einer Beobachtung x mit der Wahrscheinlichkeit $\gamma = \gamma(x)$ die Entscheidung $e_1(x)$ und mit der Wahrscheinlichkeit $1-\gamma$ die Entscheidung $e_2(x)$ getroffen wird.
Ob man sich also für $e_1(x)$ oder $e_2(x)$ entscheidet, hängt somit von dem Ausgang eines Hilfsexperimentes ab, nämlich davon, ob bei diesem ein Ereignis, welches die Wahrscheinlichkeit γ besitzt, eingetreten ist oder nicht. Die Ausführung eines solchen (Zusatz-) Experiments heißt *Randomisieren nach einer* $B(1,\gamma)$*-Verteilung.*

Definition 1.6.1. Gegeben seien der Stichprobenraum $(\mathcal{X}, \mathcal{B})$ und der Entscheidungsraum $(\Delta, \mathcal{D})$. Eine *randomisierte Entscheidungsfunktion* ist dann eine *Übergangswahrscheinlichkeit* oder *Markoffscher Kern* von $(\mathcal{X}, \mathcal{B})$ nach $(\Delta, \mathcal{D})$, d.h. eine Abbildung $\delta : \mathcal{X} \times \mathcal{D} \to [0,1]$ mit den Eigenschaften

(i) $\forall x \in \mathcal{X}$: $\delta(x, \cdot)$ ist ein Wahrscheinlichkeitsmaß auf $\mathcal{D}$

(ii) $\forall D \in \mathcal{D}$: $\delta(\cdot, D)$ ist $\mathcal{B}, \mathbb{B}$-meßbar.

Die Zahl $\delta(x, D)$ ist dabei wie folgt zu interpretieren: Bei Vorliegen der Beobachtung $x \in \mathcal{X}$ ist eine Entscheidung zu treffen, welche mit der Wahrscheinlichkeit $\delta(x, D)$ zur Menge $D \in \mathcal{D}$ gehört.

Das bedeutet: Um in einer konkreten Situation zu einer Entscheidung zu gelangen, hat man zunächst die Beobachtung $x \in \mathcal{X}$ zu gewinnen und dann mit diesem Wert x ein Hilfsexperiment mit $(\Delta, \mathcal{D})$ als Stichprobenraum und $\delta(x, \cdot)$ als Wahrscheinlichkeitsverteilung durchzuführen. Der Ausgang dieses Hilfsexperimentes ist dann die tatsächlich zu treffende Entscheidung. Die Ausführung eines Hilfsexperimentes heißt *Randomisieren nach der Verteilung* $\delta(x, \cdot)$.

Eine nicht randomisierte Entscheidungsfunktion kann mit der randomisierten Entscheidungsfunktion $\delta_e(x, D) := 1_D(e(x))$ identifiziert werden: Bei Verwendung dieser Entscheidungsfunktion ist für alle $x \in \mathcal{X}$ mit Wahrscheinlichkeit 1 die Entscheidung $e(x)$ zu treffen sofern $\{e(x)\} \in \mathcal{D}$. $[\delta_e(x, \{e(x)\}) = 1]$.

Im Folgenden verzichten wir daher auf den Zusatz „randomisiert“.

Bei einem Testproblem besteht der Entscheidungsraum Δ nur aus den beiden Elementen d_H und d_K, so dass in diesem Fall eine Entscheidungsfunktion $\delta : \mathcal{X} \times$ Potenzmenge von $\{d_H, d_K\} \to [0,1]$ bereits durch $\varphi(x) := \delta(x, \{d_K\})$ völlig bestimmt ist.
Im Spezialfall eines nicht randomisierten Tests e ergibt sich $\varphi(x) = 1_S(x)$, $x \in \mathcal{X}$ wobei $S = \{x \in \mathcal{X} : e(x) = d_k\}$ die kritische Region von e ist.

Definition 1.6.2. Unter einer Testfunktion oder kurz Test φ versteht man eine $(\mathcal{B}, \mathbb{B})$-meßbare Abbildung $\varphi : \mathcal{X} \to [0,1]$.

Dabei ist $\varphi(x)$ die Wahrscheinlichkeit dafür, dass bei Vorliegen der Beobachtung $x \in \mathcal{X}$ die Entscheidung d_K getroffen wird, d.h. $\varphi(x) =: \delta(x, \{d_K\})$. Der Spezialfall $\varphi = 1_S$ mit $S \in \mathcal{B}$ entspricht dann dem nicht randomisierten Test $e(x) = d_K$, falls $x \in S$ und $e(x) = d_H$, falls $x \in S^c$.
$[\delta(x, \{d_K\}) = \varphi(x) \Rightarrow \delta(x, \{d_H\}) = 1 - \delta(x, \{d_K\}) = 1 - \varphi(x)$, $\delta(\cdot, \{d_K\}) = \varphi(\cdot)$ ist $\mathcal{B}, \mathbb{B}$-meßbar, $\delta(x, \cdot)$ ist Wahrscheinlichkeitsmaß$]$

Bei zugrundeliegender Verteilungsannahme $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ und einer Entscheidungsfunktion $\delta : \mathcal{X} \times \mathcal{D} \to [0,1]$ (Markoffscher Kern) wird durch die Festsetzung

$$\begin{aligned} \forall D \in \mathcal{D} \quad Q_{\vartheta,\delta}(D) &:= (P_\vartheta \otimes \delta)(D) \\ &:= \int_{\mathcal{X}} \delta(x, D) P_\vartheta(dx) \\ &= E_\vartheta(\delta(\cdot, D)) \in [0,1] \end{aligned} \tag{1.6.3}$$

ein Wahrscheinlichkeitsmaß $Q_{\vartheta,\delta}$ auf $\mathcal{D}$ definiert.

Denn es gilt

(i)

$$\begin{aligned} Q_{\vartheta,\delta}(\Delta) &= \int_{\mathcal{X}} \delta(x, \Delta) P_\vartheta(dx) \\ &= P_\vartheta(\mathcal{X}) \\ &= 1. \end{aligned}$$

(ii) Es seinen $D \in \mathcal{D}$, $n \in \mathbb{N}$ paarweise disjunkt. Dann folgt aus dem Satz der monotenen Konvergenz:

$$\begin{aligned} Q_{\vartheta,\delta}\left(\bigcup_{n\in\mathbb{N}} D_n\right) &= \int_{\mathcal{X}} \delta\left(x, \bigcup_{n\in\mathbb{N}} D_n\right) P_\vartheta(dx) \\ &= \int_{\mathcal{X}} \sum_{n\in\mathbb{N}} \delta(x, D_n) P_\vartheta(dx) \\ &= \int_{\mathcal{X}} \lim_{m\to\infty} \underbrace{\sum_{n=1}^{m} \underbrace{\delta(x, D_n)}_{\geq 0}}_{\uparrow} P_\vartheta(dx) \\ &= \lim_{m\to\infty} \int_{\mathcal{X}} \sum_{n=1}^{m} \delta(x, D_n) P_\vartheta(dx) \\ &= \lim_{m\to\infty} \sum_{n=1}^{m} \int_{\mathcal{X}} \delta(x, D_n) P_\vartheta(dx) \\ &= \sum_{n\in\mathbb{N}} Q_{\vartheta,\delta}(D_n). \end{aligned}$$

$Q_{\vartheta,\delta}(D)$ lässt sich als Wahrscheinlichkeit dafür interpretieren, dass bei zugrundeliegender Verteilung P_ϑ eine in der oben beschreibenen *zweistufigen*

Weise (durch Randomisieren nach der Verteilung $\delta(x,\cdot)$ bei vorliegender Beobachtung x) gewonnene Entscheidung d zur Menge D gehört.
Im Spezialfall eines Tests φ ist $Q_{\vartheta,\delta}$ bereits durch

$$\begin{aligned} Q_{\vartheta,\delta}(\{d_K\}) &= \int_{\mathcal{X}} \delta(x,\{d_K\})P_\vartheta(dx) \\ &= \int_{\mathcal{X}} \varphi(x)P_\vartheta(dx) \\ &= E_\vartheta(\varphi) \quad \in [0,1] \end{aligned}$$

eindeutig definiert;

$$\begin{aligned} Q_{\vartheta,\delta}(\{d_H\}) &= 1 - Q_{\vartheta,\delta}(\{d_H\}) \\ &= 1 - E_\vartheta(\varphi) \\ &= E_\vartheta(1-\varphi) \end{aligned}$$

Somit ist $E_\vartheta(\varphi)$ die Wahrscheinlichkeit mit der unter der Verteilung P_ϑ die Entscheidung d_K, d.h. Verwerfen der Hypothese, getroffen wird.
Die Abbildung $\beta : \Theta \to [0,1]$ definiert durch

$$\forall \vartheta \in \Theta : \beta(\vartheta) := E_\vartheta(\varphi)$$

heißt *Gütefunktion* (*power function*) des Tests φ.
Im Spezialfall $\varphi = 1_S$ (eines nichtrandomisierten Tests) gilt:

$$\beta(\vartheta) = P_\vartheta(S).$$

Die Wahrscheinlichkeit für den Fehler erster Art ist (für ein beliebiges φ) $E_\vartheta(\varphi)$, $\vartheta \in H$, die Wahrscheinlichkeit für den Fehler zweiter Art ist $1 - E_\vartheta(\varphi) = E_\vartheta(1-\varphi)$, $\vartheta \in K$.
Ein Test φ für H gegen K mit der Eigenschaft

$$\forall \vartheta \in H : E_\vartheta(\varphi) \le \alpha \tag{1.6.4}$$

heißt *Test zum Niveau α*. Im Folgenden bezeichnen wir mit Φ die Gesamtheit aller Tests, d.h. $\Phi = \{\varphi : \mathcal{X} \to [0,1] : \varphi \text{ ist } \mathcal{B},\mathbb{B}\text{-meßbar } \}$.

Definition 1.6.5. Es sei $\Phi_1 \subset \Phi$. Ein Test φ^* heißt *gleichmäßig bester Test* bzgl. Φ_1 für H gegen K, falls

$$\varphi^* \in \Phi_1 \tag{1.6.6}$$

$$\forall \vartheta \in K : \quad E_\vartheta(\varphi^*) = \sup_{\varphi \in \Phi_1} E_\vartheta(\varphi) \tag{1.6.7}$$

Lemma 1.6.8. *Es sei $\Phi_1 \subset \Phi_2 \subset \Phi$ und $\varphi^* \in \Phi_1$. Ist φ^* ein gleichmäßig bester Test bzgl. Φ_2, dann ist φ^* auch ein gleichmäßig bester Test bzgl. Φ_1.*

Beweis: $\forall \vartheta \in K : E_\vartheta(\varphi^*) = \sup_{\varphi \in \Phi_2} E_\vartheta(\varphi) \geq \sup_{\varphi \in \Phi_1} E_\vartheta(\varphi) \geq E_\vartheta(\varphi^*)$. □

Für $\alpha \in (0,1)$ sei $\Phi_\alpha := \{\varphi \in \Phi : \forall \vartheta \in H : E_\vartheta(\varphi) \leq \alpha\}$ die Gesamtheit aller Tests zum Niveau α.
Ein gleichmäßig bester Test bzgl. Φ_α heißt dann *gleichmäßig bester Test zum Niveau α für H gegen K*.
Wir werden in Kapitel 2 sehen, dass ein gleichmäßig bester Test zum Niveau α bei vielen einseitigen Testproblemen $H : \vartheta \leq \vartheta_0$ gegen $K : \vartheta > \vartheta_0$ existiert, wobei $\Theta \subset \mathbb{R}$.
Bei zweiseitigen Testproblemen $H : \vartheta = \vartheta_0$ gegen $K : \vartheta \neq \vartheta_0$ ist man hingegen häufig gezwungen Φ_α durch eine kleinere Klasse von Testfunktionen zu ersetzen. Dabei nimmt man zumeist die Klasse aller unverfälschten Tests zum Niveau α. Ein Test φ zum Niveau α für H gegen K heißt dabei *unverfälscht*, falls

$$\forall \vartheta \in K : E_\vartheta(\varphi) \geq \alpha \tag{1.6.9}$$

d.h. bei Verwendung von φ ist unter K die Entscheidung für K mindestens so wahrscheinlich wie unter der Hypotese H ($\forall \vartheta \in H : E_\vartheta(\varphi) \leq \alpha$).

Ein gleichmäßig bester Test bzgl.

$$\Phi_\alpha^u := \{\varphi \in \Phi_\alpha : \forall \vartheta \in K : E_\vartheta(\varphi) \geq \alpha\}$$

heißt *gleichmäßig bester unverfälschter Test zum Niveau α für H gegen K.*

Lemma 1.6.10. *Jeder gleichmäßig beste Test φ^* zum Niveau α ist unverfälscht und somit ein gleichmäßig bester unverfälschter Test zum Niveau α.*

Beweis: Wegen $\varphi_\alpha := \alpha \in \Phi_\alpha$ gilt gem. (1.6.7, S. 22)

$$\forall \vartheta \in K : E_\vartheta(\varphi^*) = \alpha.$$

Wegen $\Phi_\alpha^u \subset \Phi_\alpha$ folgt die Behauptung somit aus Lemma 1.6.8. □

Eine weitere Möglichkeit zur Auszeichnung optimaler Tests ist die Folgende: Gibt es keinen gleichmäßig besten (unverfälschten) Test zum Niveau α für H gegen K, so liegt es nahe, unter allen Tests zum Niveau α einen solchen zu bestimmen, der die maximale Wahrscheinlichkeit für den Fehler zweiter Art minimiert. Dies ist gleichbedeutend damit $\inf_{\vartheta \in K} E_\vartheta(\varphi)$ unter allen Tests $\varphi \in \Phi_\alpha$ zu maximieren.

Definition 1.6.11. $\varphi^* \in \Phi_\alpha$ heißt *Maximin-Test zum Niveau α für H gegen K* $:\Leftrightarrow \inf_{\vartheta \in K} E_\vartheta(\varphi^*) = \sup_{\varphi \in \Phi_\alpha} \inf_{\vartheta \in K} E_\vartheta(\varphi)$.

Offenbar ist jeder Maximin-Test zum Niveau α ein unverfälschter Test zum Niveau α.
Im nächsten Kapitel werden wir sehen, dass ein optimaler nicht-randomisierter Test $\varphi = 1_{S^*}$ für das Problem $H : \vartheta \leq \vartheta_0$ gegen $K : \vartheta > \vartheta_0$ mit $\vartheta, \vartheta_0 \in \Theta \subset \mathbb{R}$ häufig von der Form ist

$$1_{S^*} = 1_{\{T > c\}}, \tag{1.6.12}$$

wobei T eine Stichprobenfunktion ist und die Konstante $c \in \mathbb{R}$ durch die Forderungen

$$\begin{aligned} &\forall \vartheta \in H : P_\vartheta(T > c) \leq \alpha, \\ &\forall \vartheta \in K : P_\vartheta(T > c) = \sup_{S \in \gamma_\alpha} P_\vartheta(S) \end{aligned} \tag{1.6.13}$$

bestimmt wird.

Dabei wird c möglichst klein gewählt, ohne dass die Fehlerwahrscheinlichkeit erster Art das Niveau α übersteigt. Diese Konstante c heißt dann *kritischer Wert* und die Stichprobenfunktion $T : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$ heißt *Prüfgröße* bzw. *Teststatistik* für den Parameter ϑ.

Suchen wir hingegen für dasselbe Entscheidungsproblem einen optimalen Test innerhalb der größeren Klasse der randomisierten Tests, d.h. suchen wir für das einseitige Testproblem $H : \vartheta \leq \vartheta_0$ gegen $K : \vartheta > \vartheta_0$ einen gleichmäßig besten Test φ^* zum Niveau α, so wird φ^* häufig von der Form

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } T(x) > c \\ \gamma, & \text{falls } T(x) = c \\ 0, & \text{falls } T(x) < c \end{cases} \tag{1.6.14}$$

mit $\gamma \in [0, 1]$. Randomisierte Entscheidungen werden nur auf der Menge $\{x \in \mathcal{X} : T(x) = c\}$ getroffen. Da diese Menge bei stetig verteiltem T die Wahrscheinlichkeit 0 besitzt, erhält man (optimale) randomisierte Tests vornehmlich im Fall diskreter Verteilungen.

Die Gütefunktion des Tests (1.6.14) ist

$$\beta(\vartheta) = E_\vartheta(\varphi^*) = P_\vartheta(T > c) + \gamma\, P_\vartheta(T = c) \quad \vartheta \in \Theta$$

so dass aufgrund der Optimalitätskriterien (1.6.6) und (1.6.7) (mit $\Phi_1 := \Phi_\alpha$) der kritische Wert c möglichst klein und nach dieser Festsetzung γ möglichst groß zu wählen ist, ohne dass das Niveau α überschritten wird.

Besteht nun unsere Verteilungsannahme aus diskreten Verteilungen P_ϑ, etwa Binomial- oder Hypergeometrischen Verteilungen, so wird bei Verwenden des nicht randomisierten Tests (1.6.12), wobei c gemäß (1.6.13) festgelegt ist, das zugelassene Niveau α i.a. nicht erreicht.
Bei Verwendung des randomisierten Tests (1.6.14) kann hingegen durch geeignete Wahl von c und γ das zugelassene Niveau α erreicht werden. Durch Zulassen randomisierter Tests erzielt man somit einen Gewinn an *Schärfe* (power), d.h. des Wertes der Gütefunktion auf der Alternative.

Kapitel 2

EXISTENZ OPTIMALER TESTS

2.1 Struktureigenschaften des Raumes Φ aller Testfunktionen

Es seien μ, ν Maße auf dem Stichprobenraum $(\mathcal{X}, \mathcal{B})$.

Das Maß ν heißt *absolut stetig bzgl.* μ, i.Z. $\nu \ll \mu :\Leftrightarrow$

$$\mu(B) = 0 \Rightarrow \nu(B) = 0 \quad , B \in \mathcal{B}$$

Wir sagen, dass eine Menge $\mathcal{M}$ von Maßen auf $(\mathcal{X}, \mathcal{B})$ durch μ *dominiert* wird, i.Z. $\mathcal{M} \ll \mu :\Leftrightarrow$

$$\forall \nu \in \mathcal{M} : \nu \ll \mu.$$

Ist μ ein Maß auf $\mathcal{B}$ und

$$f \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, \mu) = \left\{ h : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B}) : \int_{\mathcal{X}} |h(x)| \mu(dx) < \infty \right\}$$

mit $f \geq 0$, so wird durch die Festlegung

$$\nu(B) := \int_B f d\mu := \int_{\mathcal{X}} f \cdot 1_B \, d\mu \quad , B \subset \mathcal{B},$$

ein endliches Maß ν auf $\mathcal{B}$ definiert mit $\nu \ll \mu$.

Definition 2.1.1. Ein Maß μ auf $(\mathcal{X}, \mathcal{B})$ heißt *σ-endliches Maß*, wenn paarweise disjunkte $X_1, X_2, \dots \in \mathcal{B}$ existieren mit $\mathcal{X} = \bigcup_{i \in \mathbb{N}} X_i$, $\mu(X_i) < \infty$, $i \in \mathbb{N}$.

Satz 2.1.2 (Radon-Nikodym). *Es sei μ ein σ-endliches Maß auf $(\mathcal{X}, \mathcal{B})$ und ν/B sei ein endliches Maß mit $\nu \ll \mu$. Dann existiert $f \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, \mu)$, $f \geq 0$, mit*

$$\nu(B) = \int_B f d\mu. \tag{2.1.3}$$

*Die Funktion f ist durch (2.1.3) μ–f.ü. eindeutig bestimmt, d.h. falls $f_1 \geq 0$, $f_2 \geq 0$ Funktionen sind mit (2.1.3), so gilt $\mu(f_1 \neq f_2) = 0$, und heißt (*Radon-Nikodym*) Dichte von ν bzgl. μ, i.Z. $f \in \frac{d\nu}{d\mu}$ oder auch $d\nu = f\ d\mu$*

Beweis: Siehe etwa Bauer, H. (1992): Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie, De Gruyter, Berlin, Satz 17.10. □

Lemma 2.1.4. *Es sei μ, ν endliche Maße auf $(\mathcal{X}, \mathcal{B})$ mit der Eigenschaft $\nu(B) \leq \mu(B)$, $B \in \mathcal{B}$. Dann existiert ein $f \in \frac{d\nu}{d\mu}$ mit $0 \leq f \leq 1$.*

Beweis: Offenbar gilt $\nu \ll \mu$. Also existiert nach dem Satz von Radon-Nikodym $f_0 \in \frac{d\nu}{d\mu}$. Setzte $B_0 := \{f_0 > 1\}$. Dann gilt

$$0 = \nu(B_0) - \nu(B_0) \geq \int_{B_0} f_0\ d\mu - \mu(B_0) = \int_{B_0} f_0 - 1\ d\mu \geq 0$$

$$\Rightarrow \int_{B_0} \underbrace{f_0 - 1}_{>0}\ d\mu = 0$$

$$\Rightarrow \mu(B_0) = 0$$

Damit erfüllt $f := f_0 \cdot 1_{B_0^C}$ die Behauptung. □

Definition 2.1.5. Es sei $\mu/\mathcal{B}$ ein endliches Maß. Eine Menge $B_0 \in \mathcal{B}$ heißt ein *μ-Atom* $:\Leftrightarrow \mu(B_0) > 0$ und $\forall B \in \mathcal{B}$ mit $B \subset B_0$ gilt $\mu(B) = 0$ oder $\mu(B) = \mu(B_0)$. μ heißt *atomlos*, falls kein μ-Atom existiert.

Satz 2.1.6 (Ljapunoff). *Es seinen $\mu_1, \ldots, \mu_n$ endliche atomlose Maße auf $\mathcal{B}$. Dann ist die Menge $C := \{(\mu_1(B), \ldots, \mu_n(B)) : B \in \mathcal{B}\} \subset \mathbb{R}^n$ eine kompakte und konvexe Teilmenge des $\mathbb{R}^n$.*

Beweis: J. Lindenstrauss (1966), Joural of Math. and Mech. *(Indiana University Mathematics Journal)* 15, 971–972. □

Satz 2.1.7. *Es sei $\mu/\mathcal{B}$ ein σ-endliches Maß. Dann existiert zu jeder Folge $(\varphi_n)_{n\in\mathbb{N}} \subset \Phi$ eine Teilfolge $(\varphi_{n_k})_{k\in\mathbb{N}}$ und ein Test $\varphi_0 \in \Phi$ derart, dass*

$$\forall f \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, \mu) : \lim_{k\to\infty} \int \varphi_{n_k} f \; d\mu = \int \varphi_0 f \; d\mu \tag{2.1.8}$$

i.z. $\varphi_{n_k} \rightharpoonup \varphi_0$ bzgl μ.
Man sagt auch, Φ sei schwach folgenkompakt, und man nennt eine Folge $(\varphi_n)_{n\in\mathbb{N}} \subset \Phi$ schwach konvergent gegen $\varphi_0 \in \Phi$, falls $\varphi_n \rightharpoonup \varphi_0$ bzgl. μ gilt.

Beweis: Witting, H. (1985) Mathematische Statistik, Teubner, Satz 2.14 □

Bemerkung 2.1.9. Es sei $\mu/\mathcal{B}$ ein endliches Maß, $\varphi_n \rightharpoonup \varphi_0$ bzgl. μ. Dann gilt

$$\forall g \in \mathcal{L}_b(\mathcal{X}, \mathcal{B}) : \lim_{n\to\infty} \int_{\mathcal{X}} \varphi_n \cdot g \; d\mu = \int_{\mathcal{X}} \varphi_0 \cdot g \; d\mu$$

mit $\mathcal{L}_b(\mathcal{X}, \mathcal{B}) := \{f : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B}) : \text{ f ist beschränkt}\}$.

Im Folgenden bezeichne $\Phi' :=$ Menge aller nicht-randomisierten Test $= \{1_B : B \in \mathcal{B}\}$ und $\Phi'' :=$ Menge aller Tests mit endlichem Wertebereich $= \{\varphi \in \Phi : |\varphi(\mathcal{X})| < \infty\}$. Offenbar sind Φ' und Φ'' konvexe Mengen.

Lemma 2.1.10. *Es seien $P_1, \ldots, P_k$ Wahrscheinlichkeitsmaße auf $\mathcal{B}$. Setze*

$$M := \{(E_1(\varphi), \ldots, E_k(\varphi)) : \varphi \in \Phi\}$$

mit $E_i(\varphi) := E_{P_i}(\varphi) = \int_{\mathcal{X}} \varphi \; dP_i, \varphi \in \Phi, 1 \le i \le k$. Dann gilt

(i) $M \subset [0,1]^k$

(ii) $\alpha \in [0,1] : \quad \underbrace{(\alpha, \ldots, \alpha)}_{k-mal} \in M$

(iii) *M ist konvex*

(iv) $(\alpha_1, \ldots, \alpha_k) \in M \Rightarrow (1-\alpha_1, \ldots, 1-\alpha_k) \in M$

(v) *M ist kompakt*

(vi) *Ist $(\alpha_1, \ldots, \alpha_k)$ ein Extremalpunkt von M, so existiert ein $\varphi' \in \Phi'$ mit $\alpha_i = E_i(\varphi')$ für $i = 1, \ldots, k$*

(vii) $M = \{(E_1(\varphi), \ldots, E_k(\varphi)) : \varphi \in \Phi''\}$

(viii) *Sind $P_1, \dots P_k$ sämtlich atomlos, so gilt $M = \{(E_1(\varphi), \dots, E_k(\varphi)) : \varphi \in \Phi'\}$*

Bemerkung 2.1.11. Es sei L ein linearer Raum, K eine konvexe Teilmenge von L. Ein Punkt $x_0 \in L$ heißt *Extremalpunkt von* K, wenn aus $x_0 = \lambda x + (1-\lambda)y$ und $x, y \in K$ und $0 \leq \lambda \leq 1$ folgt : $x = y = x_0$, d.h. wenn x_0 kein innerer Punkt einer ganzen in K verlaufenden Strecke ist.
Es gilt der Satz von Minkovski: $M \subset \mathbb{R}^n$ sei kompakt und konvex, dann ist M identisch mit der konvexen Hülle der Menge der Extremalpunkte, d.h.

$$\begin{aligned} M &= \left\{ \sum_{i=1}^{n} \lambda_i e_i : \ \lambda_i \geq 0, \quad \sum_{i=1}^{n} \lambda_i = 1, \quad e_i \quad \text{Extremalpunkt} \right\} \\ &= \bigcap_{A \text{ ist konvexe Menge}, \ A \supset \text{Menge der Extremalpunkte von } M} A. \end{aligned}$$

Beweis:[von 2.1.10]

(i) Klar

(ii) Setze $\varphi_\alpha := \alpha \in \Phi$, $\alpha \in [0,1]$

(iii) Es seien $(\alpha_1, \dots, \alpha_k), (\beta_1, \dots, \beta_k) \in M$, $\lambda \in (0,1) \Rightarrow \exists \varphi, \psi \in \Phi$: $\alpha_i = E_i(\varphi), \beta_i = E_i(\psi)$, $1 \leq i \leq k$.
Es gilt: $\lambda\varphi + (1-\lambda)\psi \in \Phi \Rightarrow \lambda(\alpha_1, \dots, \alpha_k) + (1-\lambda)(\beta_1, \dots, \beta_k) = (E_i(\lambda\varphi + (1-\lambda)\psi)_{i=1}^k \in M$

(iv) Ist offensichtlich, da $\varphi \in \Phi \Rightarrow 1 - \varphi \in \Phi$

(v) Wir zeigen dass M folgenkompakt ist, d.h. jede Folge in M besitzt eine konvergente Teilfolge, deren Limes wiederum in M liegt. Dann ist M kompakt.
Für $n \in \mathbb{N}$ sei $(\alpha_1^{(n)}, \dots, \alpha_k^{(n)}) \in M$, d.h. es existiert eine Folge $(\varphi_n)_{n \in \mathbb{N}} \subset \Phi$ mit $\alpha_i^{(n)} = E_i(\varphi_n)$, $1 \leq i \leq k, n \in \mathbb{N}$. Setze

$$\mu := \sum_{i=1}^{k} P_i$$

[d.h. $\mu(B) = P_1(B) + \cdots + P_k(B)$ für $B \in \mathcal{B}$].
$\Rightarrow \mu$ ist endliches Maß auf $\mathcal{B} \overset{2.1.7}{\Rightarrow} \exists$ Teilfolge $(\varphi_{n_j})_{j \in \mathbb{N}}$ von $(\varphi_n)_{n \in \mathbb{N}}$ und $\exists \varphi_0 \in \Phi : \lim_{j \in \mathbb{N}} \int \varphi_{n_j} g \, d\mu = \int \varphi_0 g \, d\mu$, $g \in \mathcal{L}_b(\mathcal{X}, \mathcal{B})$.
Offenbar gilt $P_i \leq \mu$, $1 \leq i \leq k \overset{2.1.4}{\Rightarrow} \exists g_i \in dP_i/d\mu$ mit $0 \leq g_i \leq 1, 1 \leq i \leq k$. Es gilt also

$$\lim_{j \in \mathbb{N}} E_i(\varphi_{n_j}) = \lim_{j \in \mathbb{N}} \int \varphi_{n_j} g_i \, d\mu = \int \varphi_0 g_i \, d\mu = E_i(\varphi_0)$$

$1 \leq i \leq k$, d.h.

$$\lim_{j\in\mathbb{N}}(\alpha_1^{(n_j)}, \dots, \alpha_k^{(n_j)}) = (\alpha_1^{(0)}, \dots, \alpha_k^{(0)}) := (E_1(\varphi^{(0)}), \dots, E_k(\varphi^{(0)})) \in M$$

Also ist $M \subset \mathbb{R}^k$ folgenkompakt.

(vi) Es sei $(\alpha_1, \dots, \alpha_k)$ ein Extremalpunkt von M.
$\Rightarrow \exists \varphi \in \Phi$ mit $\alpha_i = E_i(\varphi)$, $1 \leq i \leq k$. Setze für $\epsilon \in (0, 1/2)$

$$\begin{aligned} A_\epsilon &:= \{x \in \mathcal{X} : \epsilon \leq \varphi(x) \leq 1 - \epsilon\}, \\ \varphi'_\epsilon &:= \varphi - \epsilon \cdot 1_{A_\epsilon} \\ \varphi''_\epsilon &:= \varphi + \epsilon \cdot 1_{A_\epsilon} \end{aligned}$$

Dann gilt $A_\epsilon \in \mathcal{B}$; $\varphi'_\epsilon, \varphi''_\epsilon \in \Phi$ und $\varphi = \frac{1}{2} \cdot \varphi'_\epsilon + \frac{1}{2} \cdot \varphi''_\epsilon$.
$\Rightarrow \alpha_i = E_i(\varphi) = \frac{1}{2}E_i(\varphi'_\epsilon) + \frac{1}{2}E_i(\varphi''_\epsilon)$, $1 \leq i \leq k$, d.h.

$$(\alpha_1, \dots, \alpha_k) = \frac{1}{2}(E_1(\varphi'_\epsilon), \dots, E_k(\varphi'_\epsilon)) + \frac{1}{2}(E_1(\varphi''_\epsilon), \dots, E_k(\varphi''_\epsilon)).$$

Da $(\alpha_1, \dots, \alpha_k)$ Extremalpunkt ist folgt $\alpha_i = E_i(\varphi) = E_i(\varphi'_\epsilon) = E_i(\varphi''_\epsilon)$ $\Rightarrow$ $P_i(A_\epsilon) = 0$ für $1 \leq i \leq k$. Da $\{x \in \mathcal{X} : 0 \leq \varphi(x) \leq 1\} = \bigcup_{\epsilon\in(0,1)\cap\mathbb{Q}} A_\epsilon$ folgt $P_i(0 < \varphi < 1) = 0$ für $1 \leq i \leq k$.
Für $\varphi' := 1_{\{\varphi=1\}}$ gilt daher $\alpha_i = E_i(\varphi) = E_i(\varphi \cdot 1_{\{\varphi>0\}}) = E_i(\varphi \cdot 1_{\{\varphi=1\}}) = E_i(1_{\{\varphi=1\}}) = E_i(\varphi')$ für $1 \leq i \leq k$.

(vii) Da M konvex und kompakt ist, ist M identisch mit der konvexen Hülle seiner Extremalpunkte (Satz von Minkovski). Sei $(\alpha_1, \dots, \alpha_k) \in M \overset{(vi)}{\Rightarrow}$ $\exists c_1, \dots c_n \in [0, 1], \sum_{i=1}^n c_i = 1, \exists B_1, \dots, B_n \in \mathcal{B}$:

$$(\alpha_1, \dots, \alpha_k) = \sum_{j=1}^n c_j(P_1(B_j), \dots P_k(B_j)) = (E_1(\varphi), \dots, E_k(\varphi))$$

mit $\varphi := \sum_{j=1}^n c_j \cdot 1_{B_j} \in \Phi''$.

(viii) Wir setzen den Beweisteil (vii) fort. Sind $P_1, \dots, P_k$ atomlos, so existiert nach 2.1.6 ein $B \in \mathcal{B}$ mit

$$\sum_{j=1}^n c_j(P_1(B_j), \dots, P_k(B_j)) = (P_1(B), \dots, P_k(B))$$

d.h. $(\alpha_1, \dots, \alpha_k) = (E_1(1_B), \dots, E_k(1_B))$.

□

Bemerkung 2.1.12. Ohne die Voraussetzung der Atomlosigkeit von P_i, $1 \leq i \leq k$ ist (viii) i.a. nicht richtig.

Satz 2.1.13. *Es sei $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine beliebige Familie von Wahrscheinlicheitsmaßen auf $(\mathcal{X}, \mathcal{B})$ mit $\mathcal{P} \ll \mu$, μ ein σ-endliches Maß. Dann existiert eine abzählbare Teilfamilie $\{P_{\vartheta_n} : n \in \mathbb{N}\}$ von $\mathcal{P}$, so dass $\mathcal{P} \ll \hat{P} := \sum_{n\in\mathbb{N}} 2^{-n} P_{\vartheta_n}$.*

Beweis: O.B.d.A. sei $|\Theta| = \infty$. Da μ σ-endlich ist, existieren $B_n \in \mathcal{B}$, $n \in \mathbb{N}$, mit $B_n \cap B_m = \emptyset$, $n \neq m$, $\bigcup_{n\in\mathbb{N}} B_n = \mathcal{X}$ und $\mu(B_n) < \infty$.
Setze für $B \in \mathcal{B}$

$$\nu(B) := \sum_{n\in\mathbb{N}_0} \frac{1}{2^n} \frac{\mu(B \cap B_n)}{\mu(B_n)}$$

wobei $\mathbb{N}_0 = \{n \in \mathbb{N} : \mu(B_n) > 0\}$.
Dann ist ν ein endliches Maß auf $(\mathcal{X}, \mathcal{B})$ mit $\mathcal{P} \ll \nu$. Denn:

$$\begin{aligned}
\nu(B) = 0 \quad &\Rightarrow \mu(B \cap B_n) = 0, n \in \mathbb{N} \\
&\Rightarrow \mu(B) = \mu\left(B \cap \underbrace{\bigcup_{n\in\mathbb{N}} B_n}_{=\mathcal{X}}\right) \\
&\qquad\quad = \mu\left(\bigcup_{n\in\mathbb{N}} (B \cap B_n)\right) \\
&\qquad\quad = \sum_{n\in\mathbb{N}} \mu(B \cap B_n) = 0 \\
&\Rightarrow \forall \vartheta \in \Theta : P_\vartheta(B) = 0.
\end{aligned}$$

Es sei für $\vartheta \in \Theta$ nun $f_\vartheta \in dP_\vartheta / d\nu$ (Satz von Radon-Nikodym), ferner sei $\mathcal{C}_1 := \{B \in \mathcal{B} : \exists \vartheta \in \Theta : B \subset \{f_\vartheta > 0\}\}$ sowie $\mathcal{C}_2 :=$ *Gesamtheit aller abzählbaren Vereinigungen von Mengen aus* $\mathcal{C}_1$. Setze

$$\rho := \sup_{C\in\mathcal{C}_2} \nu(C) < \infty$$

und wähle Folge $(C_n)_{n\in\mathbb{N}} \subset \mathcal{C}_2$ mit $\lim_{n\in\mathbb{N}} \nu(C_n) = \rho$. Dann ist $C_0 := \bigcup_{n\in\mathbb{N}} C_n \in \mathcal{C}_2$ mit $\nu(C_0) = \rho$, denn $\nu(C_n) \leq \nu(C_0), n \in \mathbb{N}$. Da $C_0 \in \mathcal{C}_2$, existiert eine Folge $(D_n)_{n\in\mathbb{N}} \subset \mathcal{C}_1$ mit $C_0 = \bigcup_{n\in\mathbb{N}} D_n$. Somit: $\forall n \in \mathbb{N}\ \exists \vartheta_n \in \Theta : D_n \subset \{f_{\vartheta_n} > 0\}$.
Dann gilt:

$$\forall \vartheta \in \Theta, \forall B \in \mathcal{B} : P_\vartheta(B) = P_\vartheta(B \cap C_0 \cap \{f_\vartheta > 0\}) \tag{2.1.14}$$

Denn:

$$\begin{aligned} P_\vartheta(B) &= P_\vartheta(B \cap \{f_\vartheta = 0\}) + P_\vartheta(B \cap \{f_\vartheta > 0\}) \\ &= P_\vartheta(B \cap \{f_\vartheta = 0\}) + P_\vartheta(B \cap \{f_\vartheta > 0\} \cap C_0) \\ &\quad + P_\vartheta(B \cap \{f_\vartheta > 0\} \cap C_0^c), \end{aligned}$$

wobei

$$P_\vartheta(B \cap \{f_\vartheta = 0\}) = \int_{B \cap \{f_\vartheta = 0\}} f_\vartheta \, d\nu = 0.$$

Zu zeigen bleibt also, dass $P_\vartheta(B \cap \{f_\vartheta > 0\} \cap C_0^c) = 0$ ist. Angenommen, dies wäre nicht der Fall, d.h. es gelte

$$P_\vartheta(B \cap \{f_\vartheta > 0\} \cap C_0^c) > 0 \Rightarrow \nu(B \cap \{f_\vartheta > 0\} \cap C_0^c) > 0$$

und damit

$$\nu\Big(\underbrace{\underbrace{C_0}_{\in C_2} \cup \underbrace{(B \cap \{f_\vartheta > 0\} \cap C_0^c)}_{\in C_1}}_{\in C_2}\Big) = \nu(C_0) + \nu(B \cap \{f_\vartheta > 0\} \cap C_0^c) > \nu(C_0) = \rho,$$

d.h. wir haben einen Widerspruch zur Definition von ρ. Also gilt 2.1.14.
Ferner gilt

$$P_{\vartheta_n}(B) = 0 \Rightarrow \forall \vartheta \in \Theta : \nu(B \cap \{f_\vartheta > 0\} \cap \{f_{\vartheta_n} > 0\}) = 0. \qquad (2.1.15)$$

Denn:

$$\begin{aligned} 0 = P_{\vartheta_n}(B) &\geq P_{\vartheta_n}(B \cap \{f_\vartheta > 0\} \cap \{f_{\vartheta_n} > 0\}) \\ &= \int_{B \cap \{f_\vartheta > 0\} \cap \{f_{\vartheta_n} > 0\}} f_{\vartheta_n} \, d\nu \geq 0. \end{aligned}$$

Da auf dem Integrationsbereich der Integrand f_{ϑ_n} strikt positiv ist, das Integral aber gleich 0 ist, muss $\nu(B \cap \{f_\vartheta > 0\} \cap \{f_{\vartheta_n} > 0\}) = 0$ gelten.
Es sei nun $\hat{P}(B) = 0$. Zu zeigen ist

$$\forall \vartheta \in \Theta \; P_\vartheta(B) = 0.$$

Da $\hat{P}(B) = 0 \Rightarrow \forall n \in \mathbb{N} \; P_{\vartheta_n}(B) = 0$

$$\begin{aligned} \Rightarrow \forall \vartheta \in \Theta : \nu(B \cap C_0 \cap \{f_\vartheta > 0\}) &= \nu\left(B \cap \{f_\vartheta > 0\} \cap \bigcup_{n \in \mathbb{N}} D_n\right) \\ &= \nu\left(\bigcup_{n \in \mathbb{N}} (B \cap \{f_\vartheta > 0\} \cap D_n)\right) \\ &\leq \sum_{n \in \mathbb{N}} \nu\left(B \cap \{f_\vartheta > 0\} \cap \underbrace{D_n}_{\subset \{f_{\vartheta_n} > 0\}}\right) \\ &\leq \sum_{n \in \mathbb{N}} \nu\left(B \cap \{f_\vartheta > 0\} \cap \{f_{\vartheta_n} > 0\}\right) \\ &\overset{2.1.15}{=} 0. \end{aligned}$$

Also gilt $\forall\vartheta \in \Theta : \nu(B \cap C_0 \cap \{f_\vartheta > 0\}) = 0$ und damit auch $\forall\vartheta \in \Theta :$ $P_\vartheta(B \cap C_0 \cap \{f_\vartheta > 0\}) = 0$, da $\forall\vartheta \in \Theta\ P_\vartheta \ll \nu$. Die Behauptung folgt nun aus (2.1.14).

□

Satz 2.1.16. *Es sei $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ mit $\mathcal{P} \ll \mu$, μ σ-endliches Maß auf $(\mathcal{X}, \mathcal{B})$. Es sei $\mathcal{P} = H \cup K$, $H \cap K = \emptyset$, $H, K \neq \emptyset$. Dann existiert ein Maximin-Test zum Niveau α für H gegen K.*

Beweis: Setze

$$s := \sup_{\varphi \in \Phi_\alpha} \inf_{\vartheta \in K} E_\vartheta(\varphi).$$

Es existiert eine Folge $(\varphi_n)_{n\in\mathbb{N}} \subset \Phi_\alpha$ mit $\lim_{n\in\mathbb{N}} \inf_{\vartheta\in K} E_\vartheta(\varphi_n) = s$. Aufgrund von Satz 2.1.7 existieren eine Teilfolge $(\varphi_{n_k})_{k\in\mathbb{N}}$ und $\varphi^* \in \Phi$ mit der Eigenschaft

$$\forall f \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, \mu) : \lim_{k\in\mathbb{N}} \int \varphi_{n_k} \cdot f\ d\mu = \int \varphi^* \cdot f\ d\mu.$$

Hieraus folgt speziell für $f = f_\vartheta \in dP_\vartheta/d\mu, \vartheta \in \Theta$:

$$\begin{aligned} \lim_{k\in\mathbb{N}} \int \varphi_{n_k} \cdot f\ d\mu &= \lim_{k\in\mathbb{N}} \int \varphi_{n_k}\ dP_\vartheta \\ &= \lim_{k\in\mathbb{N}} E_\vartheta(\varphi_{n_k}) \\ &= E_\vartheta(\varphi^*). \end{aligned}$$

Für $\vartheta \in H$ folgt hieraus, dass $\varphi^* \in \Phi_\alpha$. Andererseits gilt wegen

$$\lim_{n\in\mathbb{N}} \inf_{\vartheta\in K} E_\vartheta(\varphi_n) = s$$

auch

$$s = \lim_{k\in\mathbb{N}} \inf_{\vartheta\in K} E_\vartheta(\varphi_{n_k}) \leq \lim_{k\in\mathbb{N}} E_\vartheta(\varphi_{n_k}) = E_\vartheta(\varphi^*), \vartheta \in K$$

d.h.

$$\inf_{\vartheta\in K} E_\vartheta(\varphi^*) \geq s.$$

Nach Definition von s und wegen $\varphi^* \in \Phi_\alpha$ gilt aber auch $\inf_{\vartheta\in K} E_\vartheta(\varphi^*) \leq s$. Also gilt

$$\inf_{\vartheta\in K} E_\vartheta(\varphi^*) = \sup_{\varphi\in\Phi_\alpha} \inf_{\vartheta\in K} E_\vartheta(\varphi),\ \varphi^* \in \Phi_\alpha,$$

d.h. φ^* ist ein Maximin-Test zum Niveau α.

□

Im folgenden werden wir mittels Satz 2.1.7 für den Fall einer *einfachen* Alternative K, d.h. $|K| = 1$, die Existenz bester Tests bzgl. gewisser Teilmengen $\tilde{\Phi}$ von Φ nachweisen.

Wir setzen im Folgenden voraus: Die Verteilungsannahme ist $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$, $\vartheta_1 \in \Theta$ ist fest gewählt, $K := \{\vartheta_1\}$, $H := \Theta \backslash \{\vartheta_1\}$. Wir setzen

$$\tilde{\Phi} := \{\varphi \in \Phi : \quad \forall \vartheta \in H : \quad E_\vartheta(\varphi) \in F_\vartheta\}, \tag{2.1.17}$$

wobei F_ϑ für jedes $\varphi \in H$ eine abgeschlossene Teilmenge von $[0, 1]$ ist.

Beispiel 2.1.18. Es sei $\alpha \in [0, 1]$.

(i) $F_\vartheta := [0, \alpha]$, $\vartheta \in H \Rightarrow \tilde{\Phi} = \Phi_\alpha$

(ii) $F_\vartheta := \{\alpha\}$, $\vartheta \in H \Rightarrow \tilde{\Phi} = \{\varphi \in \Phi : E_\vartheta(\varphi) = \alpha, \vartheta \in H\}$.

Satz 2.1.19. *Es gelte $\mathcal{P} \ll \mu$, wobei μ ein σ-endliches Maß auf $(\mathcal{X}, \mathcal{B})$ ist. Dann existiert ein bester Test bzgl. der durch 2.1.17 definierten Klasse $\tilde{\Phi}$ für $H = \Theta \backslash \{\vartheta_1\}$ gegen $K = \{\vartheta_1\}$. Insbesondere gibt es aber einen besten Test zum Niveau α für H gegen K.*

Beweis: Setze

$$s := \sup_{\varphi \in \tilde{\Phi}} E_{\vartheta_1}(\varphi) \Rightarrow \exists \text{ Folge}(\varphi_n)_{n\in\mathbb{N}} \subset \tilde{\Phi} : s = \lim_{n\in\mathbb{N}} E_{\vartheta_1}(\varphi_n).$$

Mit Satz 2.1.7 folgt

$$\exists \text{ Teilfolge}(\varphi_{n_k})_{k\in\mathbb{N}} \subset (\varphi_n)_{n\in\mathbb{N}} \quad \exists \varphi^* \in \Phi : \quad \forall f \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, \mu) :$$

$$\lim_{k\in\mathbb{N}} \int \varphi_{n_k} f \, d\mu = \int \varphi^* f \, d\mu.$$

Speziell für $f = f_\vartheta \in dP_\vartheta/d\mu$, $\vartheta \in \Theta$, folgt:

$$\lim_{k\in\mathbb{N}} E_\vartheta(\varphi_{n_k}) = E_\vartheta(\varphi^*),$$

insbesondere also $E_{\vartheta_1}(\varphi^*) = \lim_{k\in\mathbb{N}} E_{\vartheta_1}(\varphi_{n_k}) = s$. Da F_ϑ abgeschlossen ist für $\vartheta \in H$, folgt auch $E_\vartheta(\varphi^*) \in F_\vartheta$, $\vartheta \in H$, d.h. $\varphi^* \in \tilde{\Phi}$. $\square$

2.2 Das Fundamentallemma von Neyman-Pearson

In diesem Abschnitt betrachten wir die binäre Verteilungsannahme $\mathcal{P} = \{P_{\vartheta_0}, P_{\vartheta_1}\}$, d.h. $\Theta = \{\vartheta_0, \vartheta_1\}$. Nach Satz 2.1.19 existiert für das Testproblem $H = \{\vartheta_0\}$ gegen $K = \{\vartheta_1\}$ stets ein Test zum Niveau α; man beachte, dass $\mathcal{P} \ll \mu = P_{\vartheta_0} + P_{\vartheta_1}$. Ziel dieses Abschnitts ist es, solche besten Tests zu konstruieren.
Im Folgenden sei μ ein $\mathcal{P}$ dominierendes endliches Maß und $p_i \in dP_{\vartheta_i}/d\mu$, $i = 0, 1$.

Definition 2.2.1. (i) $\varphi^* \in \Phi$ heißt *trennscharf* für $(P_{\vartheta_0}, P_{\vartheta_1})$, wenn φ^* bester Test zum Niveau $\alpha^* := E_{\vartheta_0}(\varphi^*)$ für $H = \{\vartheta_0\}$ gegen $K = \{\vartheta_1\}$ ist [most powerful].
Äquivalent: $\forall \varphi \in \Phi: \quad$ Falls $E_{\vartheta_0}(\varphi) \leq E_{\vartheta_0}(\varphi^*) \Rightarrow E_{\vartheta_1}(\varphi) \leq E_{\vartheta_1}(\varphi^*)$

(ii) $\varphi^* \in \Phi$ heißt *eigentlich trennscharf* für $(P_{\vartheta_0}, P_{\vartheta_1})$, wenn für alle $\varphi \in \Phi$ gilt:
Falls $E_{\vartheta_0}(\varphi) \leq E_{\vartheta_0}(\varphi^*)$ und $E_{\vartheta_1}(\varphi) \geq E_{\vartheta_1}(\varphi^*) \Rightarrow E_{\vartheta_i}(\varphi) = E_{\vartheta_i}(\varphi^*)$, $i = 0, 1$.

Bemerkung 2.2.2. φ^* eigentlich trennscharf für $(P_{\vartheta_0}, P_{\vartheta_1}) \Rightarrow \varphi^*$ trennscharf für $(P_{\vartheta_0}, P_{\vartheta_1})$.

Definition 2.2.3. $\varphi^* \in \Phi$ heißt Test vom *Neyman-Pearson Typ* für $(P_{\vartheta_0}, P_{\vartheta_1})$ $:\Leftrightarrow \exists \gamma \in [0, \infty):$

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } p_1(x) > \gamma p_0(x) \\ 0, & \text{falls } p_1(x) < \gamma p_0(x) \end{cases} \qquad x \in \mathcal{X}. \tag{2.2.4}$$

Lemma 2.2.5 (Neyman-Pearson, Teil 1). *$\forall \alpha \in (0, 1]$ $\exists \varphi^* \in \Phi$ vom NP-Typ für $(P_{\vartheta_0}, P_{\vartheta_1})$ mit $E_{\vartheta_0}(\varphi^*) = \alpha$.*

Beweis: Setze für $\gamma \in [0, \infty)$ die Menge $\mathcal{X}_\gamma := \{p_1 > \gamma p_0\}$ und $t(\gamma) := P_{\vartheta_0}(\mathcal{X}_\gamma)$. Die Abbildung $t : [0, \infty) \to [0, 1]$ ist monoton fallend und rechtsseitig stetig, denn:

$$\gamma_1 \leq \gamma_2 \Rightarrow \mathcal{X}_{\gamma_1} \geq \mathcal{X}_{\gamma_2} \Rightarrow t(\gamma_1) \geq t(\gamma_2),$$

$$\gamma_n \overset{n \in \mathbb{N}}{\downarrow} \gamma_0 \Rightarrow \mathcal{X}_{\gamma_0} = \bigcup_{n \in \mathbb{N}} \underbrace{\mathcal{X}_{\gamma_n}}_{\uparrow \text{ in n}} \Rightarrow t(\gamma_0) = \lim_{n \in \mathbb{N}} t(\gamma_n)$$

(aufsteigende Stetigkeit eines Wahrscheinlichkeitsmaßes).

Sei nun $\gamma_\alpha := \inf\{\gamma \in [0,\infty) : t(\gamma) \leq \alpha\}$. Beachte, dass $\{\gamma \in [0,\infty) : t(\gamma) \leq \alpha\} \neq \emptyset$; denn angenommen diese Menge wäre leer. Dann folgte für beliebiges $n \in \mathbb{N}$ mit $\gamma = n$, dass

$$\begin{aligned} 0 &< \alpha < P_{\vartheta_0}(p_1 > np_0) \\ &= P_{\vartheta_0}(p_0 > 0 \text{ und } p_1/p_0 > n) + \underbrace{P_{\vartheta_0}(p_0 = 0 \text{ und } p_1 > np_0)}_{=0} \\ &= P_{\vartheta_0}(p_0 > 0 \text{ und } p_1/p_0 > n). \end{aligned}$$

Andererseits gilt aber $\{p_0 > 0 \text{ und } p_1/p_0 > n\} \overset{n\in\mathbb{N}}{\downarrow} \emptyset$
$\Rightarrow \lim_{n\in\mathbb{N}} P_{\vartheta_0}(p_0 > 0 \text{ und } p_1/p_0 > n) = 0$, (absteigende Stetigkeit eines Wahrscheinlichkeitsmaßes), Widerspruch.
Es folgt $t(\gamma_\alpha) \leq \alpha \leq t(\gamma_\alpha - 0) := \lim_{\epsilon\downarrow 0} t(\gamma_\alpha - \epsilon)$, wobei $t(0-0) := 1$ gesetzt wird. Setzte nun für ein beliebiges $x \in \mathcal{X}$

$$\varphi^*(x) := \begin{cases} 1, & \text{falls } p_1(x) > \gamma_\alpha p_0(x) \\ \delta_\alpha := \frac{\alpha - t(\gamma_\alpha)}{t(\gamma_\alpha - 0) - t(\gamma_\alpha)}, & \text{falls } p_1(x) = \gamma_\alpha p_0(x) \\ 0, & \text{falls } p_1(x) < \gamma_\alpha p_0(x) \end{cases}$$

Wobei $\delta_\alpha := 0$ gesetzt wird falls $t(\gamma_\alpha - 0) = t(\gamma_\alpha)$ $(= \alpha)$. Dann ist φ^* ein Test vom NP[1]-Typ und es gilt

$$\begin{aligned} E_{\vartheta_0}(\varphi^*) &= P_{\vartheta_0}(\mathcal{X}_{\gamma_\alpha}) + \delta_\alpha P_{\vartheta_0}(p_1 = \gamma_\alpha p_0) \\ &= t(\gamma_\alpha) + \delta_\alpha(P_{\vartheta_0}(p_1 \geq \gamma_\alpha p_0) - P_{\vartheta_0}(p_1 > \gamma_\alpha p_0)) \\ &= t(\gamma_\alpha) + \delta_\alpha(t(\gamma_\alpha - 0) - t(\gamma_\alpha)) \\ &= \alpha. \end{aligned}$$

□

Lemma 2.2.6 (Neyman-Pearson, Teil 2). *Es gilt:*

(i) Jeder Test φ^* von NP-Typ für $(P_{\vartheta_0}, P_{\vartheta_1})$ ist trennscharf.

(ii) Ist nun entweder die in der Darstellung 2.2.4 auftretende Konstante γ positiv oder ist im Fall von $\gamma = 0$ φ^* von der speziellen Form

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } p_1(x) > 0 \\ 0, & \text{falls } p_1(x) = 0,\ p_0(x) > 0 \end{cases} \tag{2.2.7}$$

dann ist φ^* sogar eigentlich trennscharf.

[1] NP = Neyman-Pearson

Beweis: $\forall \varphi \in \Phi$:

$$
\begin{aligned}
& (\varphi^* - \varphi) \cdot (p_1 - \gamma p_0) \geq 0 \\
\Rightarrow \quad & \int (\varphi^* - \varphi) \cdot (p_1 - \gamma p_0) \, d\mu \geq 0 \\
\Rightarrow \quad & \int \varphi^* p_1 \, d\mu - \int \varphi p_1 \, d\mu \geq \gamma \cdot \left(\int \varphi^* p_0 \, d\mu - \int \varphi p_0 \, d\mu \right)
\end{aligned}
$$

d.h. aus der Definition von p_0 und p_1 folgt, dass

$$
\underbrace{E_{\vartheta_1}(\varphi^*) - E_{\vartheta_1}(\varphi)}_{\geq 0 \quad \Leftarrow} \geq \underbrace{\gamma}_{\geq 0} \cdot \underbrace{(E_{\vartheta_0}(\varphi^*) - E_{\vartheta_0}(\varphi))}_{\geq 0} \tag{2.2.8}
$$

$\Rightarrow$ (i)

Zu (ii): Es sei nun $\varphi \in \Phi$ gegeben mit $E_{\vartheta_0}(\varphi) \leq E_{\vartheta_0}(\varphi^*)$ und $E_{\vartheta_1}(\varphi) \geq E_{\vartheta_1}(\varphi^*)$. Es folgt wegen 2.2.8

$$
0 \geq E_{\vartheta_1}(\varphi^*) - E_{\vartheta_1}(\varphi) \geq \gamma \cdot (E_{\vartheta_0}(\varphi^*) - E_{\vartheta_0}(\varphi)) \geq 0
$$

$\Rightarrow E_{\vartheta_1}(\varphi^*) = E_{\vartheta_1}(\varphi)$ und außerdem $E_{\vartheta_0}(\varphi^*) = E_{\vartheta_0}(\varphi)$ im Fall $\gamma > 0$. Zu zeigen bleibt also $E_{\vartheta_0}(\varphi^*) = E_{\vartheta_0}(\varphi)$ falls φ^* von der Form 2.2.7 ist.
Wegen $E_{\vartheta_1}(\varphi) = E_{\vartheta_1}(\varphi^*)$ gilt:

$$
\begin{aligned}
0 &= E_{\vartheta_1}(\varphi^*) - E_{\vartheta_1}(\varphi) \\
&= \int (\varphi^* - \varphi) p_1 \, d\mu \\
&= \int_{\{p_1 > 0\}} (\varphi^* - \varphi) p_1 \, d\mu + \int_{\{p_1 = 0\}} (\varphi^* - \varphi) p_1 \, d\mu \\
&= \int_{\{p_1 > 0\}} (1 - \varphi) p_1 \, d\mu \\
&= \int_{\{p_1 > 0\} \cap \{1 - \varphi > 0\}} (1 - \varphi) p_1 \, d\mu
\end{aligned}
$$

$$
\Rightarrow \mu(\{p_1 > 0\} \cap \{1 - \varphi > 0\}) = 0 \overset{P_{\vartheta_0} \ll \mu}{\Rightarrow} P_{\vartheta_0}(\{p_1 > 0\} \cap \{1 - \varphi > 0\}) = 0
$$

Hieraus folgt nun

$$
\begin{aligned}
P_{\vartheta_0}(\varphi^* > \varphi) &= P_{\vartheta_0}(\{p_1 > 0\} \cap \{\varphi^* > \varphi\}) + P_{\vartheta_0}(\{p_1 = 0\} \cap \{\varphi^* > \varphi\}) \\
&= P_{\vartheta_0}(\{p_1 > 0\} \cap \{1 - \varphi > 0\}) \\
&= 0,
\end{aligned}
$$

denn

$$
\begin{aligned}
P_{\vartheta_0}(\{p_1 = 0\} \cap \{\varphi^* > \varphi\}) &= P_{\vartheta_0}(\{p_1 = 0\} \cap \{\varphi^* > \varphi\} \cap \{p_0 = 0\}) \\
&\quad + P_{\vartheta_0}(\{p_1 = 0\} \cap \{\varphi^* > \varphi\} \cap \{p_0 > 0\}) \\
&= \int_{\{p_1 = 0\} \cap \{\varphi^* > \varphi\} \cap \{p_0 = 0\}} p_0 \, d\mu = 0.
\end{aligned}
$$

Wir erhalten also $\varphi^* \leq \varphi \quad P_{\vartheta_0}$-f.ü.

$$\Rightarrow E_{\vartheta_0}(\varphi^*) \leq E_{\vartheta_0}(\varphi) \Rightarrow E_{\vartheta_0}(\varphi^*) = E_{\vartheta_0}(\varphi),$$

denn $E_{\vartheta_0}(\varphi^*) \geq E_{\vartheta_0}(\varphi)$ war vorausgesetzt. □

Lemma 2.2.9 (Neyman-Pearson, Teil 3). *Es gilt:*

(i) Jeder für $(P_{\vartheta_0}, P_{\vartheta_1})$ trennscharfe Test φ_0 mit $E_{\vartheta_0}(\varphi_0) > 0$ ist μ-f.ü. vom NP-Typ für $(P_{\vartheta_0}, P_{\vartheta_1})$.

(ii) Ist φ_0 darüber hinaus eigentlich trennscharf für $(P_{\vartheta_0}, P_{\vartheta_1})$, so ist $\gamma > 0$ oder φ_0 ist μ-f.ü. von der Form 2.2.7 auf Seite 36.

Beweis: Es sei $\varphi_0 \in \Phi$ trennscharf für $(P_{\vartheta_0}, P_{\vartheta_1})$ mit $E_{\vartheta_0}(\varphi_0) > 0$. Nach Lemma 2.2.5 existiert ein Test φ^* vom NP-Typ mit $E_{\vartheta_0}(\varphi^*) = E_{\vartheta_0}(\varphi_0)$. Nach Lemma 2.2.6 (i) ist φ^* ebenfalls trennscharf für $(P_{\vartheta_0}, P_{\vartheta_1})$, d.h. $E_{\vartheta_1}(\varphi^*) = E_{\vartheta_1}(\varphi_0)$.
Da $(\varphi^* - \varphi_0)(p_1 - \gamma p_0) \geq 0$ und

$$\begin{aligned} \int (\varphi^* - \varphi_0)(p_1 - \gamma p_0)\, d\mu &= \int \varphi^* p_1\, d\mu - \int \varphi_0 p_1\, d\mu - \\ &\quad \gamma \left(\int \varphi^* p_0\, d\mu - \int \varphi_0 p_0\, d\mu \right) \\ &= E_{\vartheta_1}(\varphi^*) - E_{\vartheta_1}(\varphi_0) - \gamma \left(E_{\vartheta_0}(\varphi^*) - E_{\vartheta_0}(\varphi_0) \right) \\ &= 0 \end{aligned}$$

$\Rightarrow (\varphi^* - \varphi_0)(p_1 - \gamma p_0) = 0 \quad \mu$-f.ü., d.h. μ-f.ü. gilt

$$\begin{cases} p_1(x) > \gamma \cdot p_0(x) & \Rightarrow \varphi_0(x) = \varphi^*(x) = 1 \\ p_1(x) < \gamma \cdot p_0(x) & \Rightarrow \varphi_0(x) = \varphi^*(x) = 0 \end{cases}$$

Also ist φ_0 μ-f.ü. vom NP-Typ für $(P_{\vartheta_0}, P_{\vartheta_1})$.

Zu (ii): Es ist zu zeigen, dass im Fall $\gamma = 0$ φ_0 μ-f.ü. die Form besitzt:

$$\varphi_0(x) = \begin{cases} 1, & p_1(x) > 0 \\ 0, & p_1(x) = 0,\ p_0(x) > 0 \end{cases}$$

Da φ_0 wegen (i) μ-f.ü. vom NP-Typ mit $\gamma = 0$ ist, gilt für μ-f.a. $x \in \mathcal{X}$:

$$p_1(x) > 0 \Rightarrow \varphi_0(x) = 1.$$

Sei nun $\varphi^* := 1_{\{p_1 > 0\}}$ und $M := \{p_0 > 0, p_1 = 0\}$; es bleibt zu zeigen, dass $\varphi_0 / M = 0$ μ-f.ü.. Angenommen, dies sei nicht der Fall, d.h.

$$\mu(\{\varphi_0 > 0\} \cap M) > 0.$$

Es folgt

$$\begin{aligned}
E_{\vartheta_0}(\varphi^*) &= \int_{\{p_1>0\}} p_0 \, d\mu = \int_{\{p_1>0,p_0>0\}} p_0 \, d\mu \\
&= \int_{\{p_0>0,p_1>0\}} \varphi_0 p_0 \, d\mu \\
&< \int_{\{p_0>0,p_1>0\}} \varphi_0 p_0 \, d\mu + \int_{\{p_0>0,p_1=0\}=M} \varphi_0 p_0 \, d\mu \\
&= \int_{\{p_0>0\}} \varphi_0 p_0 \, d\mu = E_{\vartheta_0}(\varphi_0)
\end{aligned}$$

Andererseits gilt aber

$$\begin{aligned}
E_{\vartheta_1}(\varphi_0) &= \int \varphi_0 p_1 \, d\mu = \int_{\{p_1>0\}} \varphi_0 p_1 \, d\mu = \int_{\{p_1>0\}} p_1 \, d\mu \\
&= E_{\vartheta_1}(\varphi^*)
\end{aligned}$$

Damit erhalten wir einen Widerspruch zur eigentlichen Trennschärfe von φ_0, wonach aus $E_{\vartheta_0}(\varphi^*) \leq E_{\vartheta_0}(\varphi_0)$ und $E_{\vartheta_1}(\varphi^*) \geq E_{\vartheta_1}(\varphi_0)$ eigentlich

$$E_{\vartheta_i}(\varphi^*) = E_{\vartheta_i}(\varphi_0) \quad i = 0, 1$$

folgen müsste. $\square$

Korollar 2.2.10. *Es sei φ_0 ein bester Test zum Niveau $\alpha \in (0,1)$ für $H = \{\vartheta_0\}$ gegen $K = \{\vartheta_1\}$. Falls $P_{\vartheta_0} \neq P_{\vartheta_1}$, so gilt $E_{\vartheta_1}(\varphi_0) > \alpha$.*

2.3 Das verallgemeinerte Fundamentallemma von Neyman-Pearson

In diesem Abschnitt untersuchen wir die Frage nach der Existenz bester Tests im Fall einer zusammengesetzten Hypothese $H = \{\vartheta_1, \ldots, \vartheta_k\}$ gegen eine einfache Alternative $K = \{\vartheta_{k+1}\}$, $\vartheta_{k+1} \neq \vartheta_i$, $1 \leq i \leq k$, $k \in \mathbb{N}$. Ferner sei μ ein endliches Maß auf $(\mathcal{X}, \mathcal{B})$ mit $\{P_{\vartheta_1}, \ldots, P_{\vartheta_{k+1}}\} \ll \mu$, etwa $\mu := P_{\vartheta_1} + \cdots + P_{\vartheta_{k+1}}$, sowie $p_i \in dP_{\vartheta_i}/d\mu$, i = 1,2, …, k+1.
Schließlich seinen $\alpha^{(k)} := (\alpha_1, \ldots, \alpha_k) \in (0,1)^k$.
Wir betrachten folgende zwei Teilklassen von Φ:

$\tilde{\Phi}_{\alpha^{(k)}} := \{\varphi \in \Phi : E_{\vartheta_i}(\varphi) = \alpha_i,\ 1 \leq i \leq k\} \neq \emptyset$ (dies sei vorausgesetzt)
$\Phi_{\alpha^{(k)}} := \{\varphi \in \Phi : E_{\vartheta_i}(\varphi) \leq \alpha_i,\ 1 \leq i \leq k\} \neq \emptyset$

Definition 2.3.1. Ein Test φ^* heißt vom NP-Typ für (H, K), falls $\gamma_1, \ldots, \gamma_k \in \mathbb{R}$ existieren mit

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } p_{k+1}(x) > \sum_{i=1}^k \gamma_i p_i(x) \\ 0, & \text{falls } p_{k+1}(x) < \sum_{i=1}^k \gamma_i p_i(x) \end{cases} \quad x \in \mathcal{X}. \tag{2.3.2}$$

Satz 2.3.3 (Verallgemeinertes Fundamentallemma von Neyman-Peason). *Es gilt:*

(i) Es existiert stets ein bzgl. $\tilde{\Phi}_{\alpha^{(k)}}$ *bester Test für* H *gegen* K.

(ii) Ist $\tilde{\varphi} \in \tilde{\Phi}_{\alpha^{(k)}}$ μ*-f.ü. vom NP-Typ für* (H, K)*, so ist* $\tilde{\varphi}$ *bester Test bzgl.* $\tilde{\Phi}_{\alpha^{(k)}}$ *für* H *gegen* K.

(iii) Ist $\alpha^{(k)}$ *ein innerer Punkt der Menge* $M^{(k)} := \{(E_{\vartheta_1}(\varphi), \ldots, E_{\vartheta_k}(\varphi)) : \varphi \in \Phi\}$*, so ist ein bzgl.* $\tilde{\Phi}_{\alpha^{(k)}}$ *bester Test für* H *gegen* K μ*-f.ü. vom NP-Typ für* (H, K).

Beweis: Behauptung (i) folgt unmittelbar aus Satz 2.1.19 auf Seite 34 mit $F_{\vartheta_i} = \{\alpha_i\}$, $1 \leq i \leq k$.

Zu (ii): Zu zeigen ist

$$E_{\vartheta_{k+1}}(\tilde{\varphi}) = \sup_{\varphi \in \tilde{\Phi}_{\alpha^{(k)}}} E_{\vartheta_{k+1}}(\varphi)$$

Sei dann $\varphi \in \tilde{\Phi}_{\alpha^{(k)}}$ beliebig vorgegeben. Da $\tilde{\varphi}$ μ-f.ü. vom NP-Typ ist, existieren $\gamma_1, \ldots, \gamma_k \in \mathbb{R}$, sodass μ-f.ü. gilt:

$$\begin{aligned}
& (\tilde{\varphi} - \varphi) \cdot \left(p_{k+1} - \sum_{i=1}^k \gamma_i p_i \right) \geq 0 \\
\Rightarrow\quad & \int \tilde{\varphi} p_{k+1} \, d\mu - \int \varphi p_{k+1} \, d\mu \geq \sum_{i=1}^k \gamma_i \left(\int \tilde{\varphi} p_i \, d\mu - \int \varphi p_i \, d\mu \right) \\
\Rightarrow\quad & E_{\vartheta_{k+1}}(\tilde{\varphi}) - E_{\vartheta_{k+1}}(\varphi) \geq \sum_{i=1}^k \gamma_i \left(E_{\vartheta_i}(\tilde{\varphi}) - E_{\vartheta_i}(\varphi) \right) \\
& = \sum_{i=1}^k \gamma_i (\alpha_i - \alpha_i) \\
& = 0 \\
\Rightarrow\quad & E_{\vartheta_{k+1}}(\tilde{\varphi}) \geq E_{\vartheta_{k+1}}(\varphi)
\end{aligned}$$

Zu (iii): Es sei $\tilde{\varphi}$ ein bester Test bzgl. $\tilde{\Phi}_{\alpha^{(k)}}$ für H gegen K. Setze $M^{(k+1)} := \{(E_{\vartheta_1}(\varphi), \ldots, E_{\vartheta_{k+1}}(\varphi)) : \varphi \in \Phi\}$, $I := \{\alpha \in [0,1] : (\alpha_1, \ldots, \alpha_k, \alpha) \in M^{(k+1)}\}$.
Es ist $I \neq \emptyset$, da z.B. $E_{\vartheta_{k+1}}(\tilde{\varphi}) \in I$. Annahme: I enthalte wenigstens zwei Elemente.
Da nach Lemma 2.1.10 $M^{(k+1)}$ kompakt und daher abgeschlossen ist, ist auch $I \subset [0,1]$ abgeschlossen und daher kompakt. Also existiert

$$\alpha_{k+1} := \max\{\alpha : \alpha \in I\} \in I.$$

Da $\tilde{\varphi}$ bester Test bzgl. $\tilde{\Phi}_{\alpha^{(k)}}$ ist, folgt, $E_{\vartheta_{k+1}}(\tilde{\varphi}) = \alpha_{k+1}$ und

$$\alpha^{(k+1)} := (\alpha_1, \ldots, \alpha_k, \alpha_{k+1}) \in \partial M^{(k+1)} = \text{topologischer Rand von } M^{(k+1)}.$$

Da $M^{(k+1)}$ nach Lemma 2.1.10 außerdem konvex ist, existiert eine durch $\alpha^{(k+1)}$ gehende Hyperebene $\mathcal{H}$, so das $M^{(k+1)}$ ganz auf einer Seite dieser Hyperebene liegt, d.h.

$$\alpha^{(k+1)} \in \mathcal{H} := \left\{ (t_1, \ldots, t_{k+1}) \in \mathbb{R}^{k+1} : \sum_{i=1}^{k+1} \gamma_i (t_i - \alpha_i) = 0 \right\}$$

für gewisse reelle Zahlen $\gamma_1, \ldots, \gamma_{k+1}$ und

$$\forall (\beta_1, \ldots, \beta_{k+1}) \in M^{(k+1)} : \sum_{i=1}^{k+1} \gamma_i (\beta_i - \alpha_i) \geq 0$$

(Satz von der Existenz einer Stützhyperebene).
Dann ist $\gamma_{k+1} \neq 0$. Denn andernfalls wäre $\sum_{i=1}^{k} \gamma_i (\beta_i - \alpha_i) \geq 0$ für alle $(\beta_1, \ldots, \beta_k) \in M^{(k)}$, d.h. $M^{(k)}$ liegt ganz auf einer Seite der Hyperebene $\mathcal{H}' := \{(t_1, \ldots, t_k) \in \mathbb{R}^k : \sum_{i=1}^{k} \gamma_i (t_i - \alpha_i) = 0\}$. Andererseits ist $\alpha^{(k)} \in \mathcal{H}'$ und daher $\alpha^{(k)} \in \partial M^{(k)}$ im Widerspruch dazu, dass $\alpha^{(k)}$ immer ein innerer Punkt von $M^{(k)}$ ist. Also folgt $\gamma_{k+1} \neq 0$.
Da nun für $\alpha \in I$ gilt: $(\alpha_1, \ldots, \alpha_k, \alpha) \in M^{(k+1)}$ und somit

$$0 \leq \sum_{i=1}^{k} \gamma_i (\alpha_i - \alpha_i) + \gamma_{k+1} (\alpha - \alpha_{k+1}) = \underbrace{\gamma_{k+1}}_{\neq 0} \underbrace{(\alpha - \alpha_{k+1})}_{<0},$$

folgt $\gamma_{k+1} < 0$ (α_{k+1} ist maximales Element in I).

o.E. sei $\gamma_{k+1} = -1$. Es folgt:

$$\begin{aligned}
&\forall \varphi \in \Phi : (E_{\vartheta_1}(\varphi), \ldots, E_{\vartheta_{k+1}}(\varphi)) \in M^{(k+1)} \\
&\Rightarrow \sum_{i=1}^{k+1} \gamma_i (E_{\vartheta_i}(\varphi) - \alpha_i) \geq 0 \\
&\Rightarrow E_{\vartheta_{k+1}}(\varphi) - \sum_{i=1}^{k} \gamma_i E_{\vartheta_i}(\varphi) \leq \alpha_{k+1} - \sum_{i=1}^{k} \gamma_i \alpha_i \\
&\Rightarrow \int \varphi \cdot \left(p_{k+1} - \sum_{i=1}^{k} \gamma_i p_i \right) \, d\mu \leq \alpha_{k+1} - \sum_{i=1}^{k} \gamma_i \alpha_i \\
&\quad = E_{\vartheta_{k+1}}(\tilde{\varphi}) - \sum_{i=1}^{k} \gamma_i E_{\vartheta_i}(\tilde{\varphi}) \\
&\quad = \int \tilde{\varphi} \left(p_{k+1} - \sum_{i=1}^{k} \gamma_i p_i \right) \, d\mu.
\end{aligned}$$

Setzen wir zur Abkürzung $h := p_{k+1} - \sum_{i=1}^{k} \gamma_i p_i$ und $\varphi^* := 1_{\{h>0\}}$, so haben wir gezeigt, dass für alle $\varphi \in \Phi$ gilt

$$\int \varphi \cdot h \, d\mu \leq \int \tilde{\varphi} \cdot h \, d\mu$$

Da außerdem

$$\int \varphi \cdot h \, d\mu \leq \int \varphi^* \cdot h \, d\mu,$$

folgt

$$\int \tilde{\varphi} \cdot h \, d\mu = \int \varphi^* \cdot h \, d\mu$$

bzw.

$$\int \underbrace{(\varphi^* - \tilde{\varphi}) \cdot h}_{\geq 0} \, d\mu = 0$$

$\Rightarrow (\varphi^* - \tilde{\varphi}) \cdot h = 0 \quad \mu$-f.ü., d.h. $\tilde{\varphi}$ ist μ-f.ü. vom NP-Typ. □

Lemma 2.3.4. *Ist $\varphi^* \in \tilde{\Phi}_{\alpha^{(k)}}$ vom NP-Typ für (H, K) wobei $\gamma_i \geq 0$, $1 \leq i \leq k$, so ist φ^* sogar bester Test bzgl. der größeren Klasse $\Phi_{\alpha^{(k)}}$ für H gegen K.*

Beweis: Zu zeigen ist $E_{\vartheta_{k+1}}(\varphi^*) = \sup_{\varphi \in \Phi_{\alpha^{(k)}}} E_{\vartheta_{k+1}}(\varphi)$. Es sei also $\varphi \in \Phi_{\alpha^{(k)}}$ beliebig. Nach Voraussetzung existieren $\gamma_i \geq 0$, $1 \leq i \leq k$, so dass

$$(\varphi^* - \varphi) \left(p_{k+1} - \sum_{i=1}^{k} \gamma_i p_i \right) \geq 0.$$

Analog zum Beweis von Satz 2.3.3 (ii) folgt

$$E_{\vartheta_{k+1}}(\varphi^*) - E_{\vartheta_{k+1}}(\varphi) \geq \sum_{i=1}^{k} \gamma_i \left(\underbrace{E_{\vartheta_i}(\varphi^*)}_{=\alpha_i} - \underbrace{E_{\vartheta_i}(\varphi)}_{\leq \alpha_i} \right) \geq 0$$

und daraus die Behauptung. □

Definition 2.3.5. Wahrscheinlichkeitsmaße $P_1, \dots, P_k$ auf $(\mathcal{X}, \mathcal{B})$ heißen *linear unabhängig* $:\Leftrightarrow$

$$\forall B \in \mathcal{B} : \sum_{i=1}^{k} \gamma_i P_i(B) = 0 \Rightarrow \gamma_i = 0,\ 1 \leq i \leq k.$$

Bemerkung 2.3.6. Es seien $P_1, \dots, P_k$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{B})$, $\mu/\mathcal{B}$ ein endliches Maß mit $\{P_1, \dots, P_k\} \ll \mu$ und $p_i \in dP_i/d\mu_i$, $1 \leq i \leq k$. Dann gilt:

$$\begin{aligned} &P_1, \dots, P_k \text{ sind linear unabhängig} \\ &\quad \Leftrightarrow \left[\sum_{i=1}^{k} \gamma_i p_i = 0\ \mu\text{-f.ü.} \Rightarrow \gamma_i = 0,\ 1 \leq i \leq k \right] \end{aligned}$$

Lemma 2.3.7. *Es gilt:*

(i) Sind $P_{\vartheta_1}, \dots, P_{\vartheta_k}$ linear unabhängig und ist $\alpha \in (0,1)$, so ist $(\alpha, \dots, \alpha) \in (0,1)^k$ innerer Punkt von $M^{(k)} = \{(E_{\vartheta_1}(\varphi), \dots, E_{\vartheta_k}(\varphi)) : \varphi \in \Phi\}$.

(ii) Sind $P_{\vartheta_1}, \dots, P_{\vartheta_k}, P_{\vartheta_{k+1}}$ linear unabhängig und ist für ein $\alpha \in (0,1)$ der Vektor $(\alpha, \dots, \alpha)$ innerer Punkt von $M^{(k)}$, so existiert $\varphi \in \Phi$ mit $E_{\vartheta_i}(\varphi) = \alpha$, $1 \leq i \leq k$ und $E_{\vartheta_{k+1}}(\varphi) > \alpha$.

Beweis: Zu (ii): Zu zeigen ist $\exists \hat{\varphi} \in \hat{\Phi} := \{\varphi \in \Phi : E_{\vartheta_i}(\varphi) = \alpha,\ 1 \leq i \leq k\}$ mit $E_{\vartheta_{k+1}}(\hat{\varphi}) > \alpha$. Angenommen, dies wäre nicht der Fall, dann wäre $\varphi_\alpha := \alpha\ (\in \hat{\Phi})$ ein bester Test bzgl. $\hat{\Phi}$ für $H = \{\vartheta_1, \dots, \vartheta_k\}$ gegen $K = \{\vartheta_{k+1}\}$. Also ist nach Satz 2.3.3 φ_α μ-f.ü. vom NP-Typ. Wegen $0 < \alpha < 1$ folgt $p_{k+1} - \sum_{i=1}^{k} \gamma_i p_i = 0$ μ-f.ü. für gewisse Zahlen $\gamma_1, \dots, \gamma_k$, im Widerspruch zur linearen Unabhängigkeit von $P_{\vartheta_1}, \dots, P_{\vartheta_{k+1}}$.

Zu (i): Durch vollständige Induktion nach k:
$k = 1$: $M^{(1)} = \{E_{\vartheta_1}(\varphi) : \varphi \in \Phi\} = [0,1]$, d.h. $\alpha \in (0,1)$ ist innerer Punkt von $M^{(1)}$

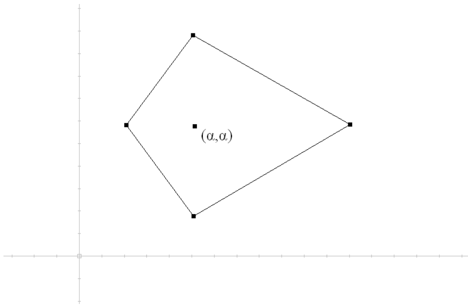
$k-1 \to k$: $P_{\vartheta_1}, \dots, P_{\vartheta_k}$ linear unabhängig.
Dann sind auch $P_{\vartheta_1}, \dots, P_{\vartheta_{i-1}}, P_{\vartheta_{i+1}}, \dots, P_{\vartheta_k}$ linear unabhängig $\Rightarrow (\alpha, \dots, \alpha) \in (0,1)^{k-1}$ ist innerer Punkt von $M^{(k-1)}$ nach Induktionvorraussetzung. Nach (ii) (für k statt k+1 und P_{ϑ_i} statt $P_{\vartheta_{i+1}}$) folgt für $j = 1, \dots, i-1, i+1, \dots, k$:

$$\exists \varphi_i \in \Phi : E_{\vartheta_j}(\varphi_i) = \alpha,\ E_{\vartheta_i}(\varphi_i) > \alpha \tag{2.3.8}$$

$$\exists \varphi_i' \in \Phi : E_{\vartheta_j}(\varphi_i') = \alpha,\ E_{\vartheta_i}(\varphi_i') < \alpha \tag{2.3.9}$$

Denn durch Anwendung von 2.3.8 auf $1-\alpha$ statt α erhält man $\psi_i \in \Phi$, so dass $\varphi_i' := 1 - \psi_i$ 2.3.9 erfüllt. Führt man diesen Schritt nun für $i = 1, \dots, k$ durch, so liegt das k-Tupel $(\alpha, \dots, \alpha)$ im Inneren der konvexen Hülle der $2k$ Punkte

$$(\alpha, \dots, \alpha, E_{\vartheta_i}(\varphi_i), \alpha, \dots, \alpha), (\alpha, \dots, \alpha, E_{\vartheta_i}(\varphi_i'), \alpha, \dots, \alpha),\ 1 \le i \le k.$$



Da jeder dieser $2k$ Punkte zu $M^{(k)}$ gehört und $M^{(k)}$ konvex ist, ist das k–Tupel $(\alpha, \dots, \alpha)$ innerer Punkt von $M^{(k)}$. $\square$

Beispiel 2.3.10. Es sei $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathbb{B})$, $P_{\vartheta_i} = N(\mu_i, \sigma^2)$ $i = 1,2,3$ mit bekanntem $\sigma^2 > 0$ und $\mu_1 < \mu_3 < \mu_2$. Ferner sei $\alpha \in (0,1)$ und φ_0 ein bzgl. $\{\varphi \in \Phi : E_{\vartheta_i}(\varphi) = \alpha,\ i = 1,2\}$ bester Test für $H = \{\vartheta_1, \vartheta_2\}$ gegen $K = \{\vartheta_3\}$. Dann sind $P_{\vartheta_1}, P_{\vartheta_2}$ linear unabhängig, denn:

$$\begin{aligned} &\forall B \in \mathbb{B} : \gamma_1 P_{\vartheta_1}(B) + \gamma_2 P_{\vartheta_2}(B) = 0 \\ &\overset{B=\mathbb{R}}{\Rightarrow} \gamma_2 = -\gamma_1 \\ \Rightarrow\ &\forall B \in \mathbb{B} : \gamma_1 \left(P_{\vartheta_1}(B) - P_{\vartheta_2}(B)\right) = 0 \\ &\Rightarrow \gamma_1 = 0 \end{aligned}$$

Also ist nach Lemma 2.3.7 (i) (α, α) ein innerer Punkt von $M^{(2)}$. Daher folgt aus Satz 2.3.3 (iii): Für λ-fast alle $x \in \mathbb{R}$ besitzt φ_0 die Gestalt

$$\varphi_0(x) = \begin{cases} 1, & \text{falls } \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_3)^2}{2\sigma^2}\right) > \\ & \qquad \frac{\gamma_1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) + \frac{\gamma_2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right) \\ 0, & \text{falls } \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_3)^2}{2\sigma^2}\right) < \\ & \qquad \frac{\gamma_1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right) + \frac{\gamma_2}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right) \end{cases}$$

2.4 Exponentialfamilien

Im folgenden behandeln wir die Frage nach der Existenz und der Gestalt gleichmäßig bester Tests zum Niveau α für H gegen K mit beliebigen disjunkten $H, K \subset \Theta$.
Wir gehen dabei wie folgt vor: Für einen speziellen Wert $\vartheta_1 \in K$ bestimmen wir aufgrund unserer bisherigen Ergebnisse einen besten Test für H gegen $K' := \{\vartheta_1\}$.
Ist dieser Test dann unabhängig von dem speziellen Wert ϑ_1, so ist er ein gleichmäßig bester Test für H gegen K.

Definition 2.4.1. Eine Familie $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ heißt *k-parametrige Exponentialfamilie* $:\Leftrightarrow$

(i) $\exists\, h,\ T_1,\ \ldots,\ T_k : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$

(ii) $\exists\, c,\ q_1,\ \ldots,\ q_k : \Theta \to \mathbb{R}$

(iii) $\exists\, \mu/\mathcal{B}\quad \sigma$-endliches Maß:

$$p_\vartheta(x) := c(\vartheta) \cdot h(x) \cdot \exp\left(\sum_{j=1}^{k} q_j(\vartheta) T_j(x)\right),\ x \in \mathcal{X}\ \vartheta \in \Theta \qquad (2.4.2)$$

ist eine μ-Dichte von P_ϑ. Dabei ist $h \geq 0$ und $c > 0$.

Beispiel 2.4.3. Es sei $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathbb{B})$, $P_\vartheta = N(\mu, \sigma^2)$ mit $\vartheta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)$, $\mu = \lambda =$ Lebesque-Maß auf $(\mathbb{R}, \mathbb{B})$.
P_ϑ hat die Dichte

$$\begin{aligned} p_\vartheta(x) &= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \cdot \exp\left(-\frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2} x\right), \end{aligned}$$

welche die Form (2.4.2) besitzt mit

$$\begin{aligned} c(\vartheta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \\ h &\equiv 1 \\ q_1(\vartheta) &= -\frac{1}{2\sigma^2} \\ q_2(\vartheta) &= \frac{\mu}{\sigma^2} \\ T_1(x) &= x^2 \\ T_2(x) &= x \end{aligned}$$

d.h. $k = 2$, μ und σ unbekannt.

Ist hingegen $\mu = \mu_0$ oder $\sigma^2 = \sigma_0^2$ bekannt, so ist

$$\mathcal{P} = \left\{ N(\mu_0, \sigma^2) : \sigma^2 \in (0, \infty) \right\}$$

bzw.

$$\mathcal{P} = \left\{ N(\mu, \sigma_0^2) : \mu \in \mathbb{R} \right\}$$

eine einparametrige Exponentialfamilie mit

$$c(\vartheta) = \frac{1}{\sqrt{2\pi}\sigma}; \quad h \equiv 1; \quad q_1(\vartheta) = -\frac{1}{2\sigma^2}; \quad T_1(x) = (x - \mu_0)^2$$

bzw.

$$c(\vartheta) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{\mu^2}{2\sigma_0^2} \right); \quad h(x) = \exp\left(-\frac{x^2}{2\sigma_0^2} \right);$$
$$q_1(\vartheta) = \frac{\mu}{\sigma_0^2}; \quad T_1(x) = x.$$

Bemerkung 2.4.4. Durch geeignete Wahl eines dominierenden Maßes μ lässt sich in der Darstellung (2.4.2) o.E. $h \equiv 1$ wählen. Man gehe von μ zu $\tilde{\mu}(B) := \int_B h \, d\mu$, $B \in \mathcal{B}$ über, welches ebenfalls σ-endlich ist:
Es sei $\mathcal{X} = \bigcup_{n \in \mathbb{N}} \mathcal{X}_n$ mit $\mu(\mathcal{X}_n) < \infty$ $n \in \mathbb{N}$.
Setze $A_m := \{m - 1 \leq h < m\}$, $m \in \mathbb{N}$. Dann gilt $\mathcal{X} = \bigcup_{(m,n) \in \mathbb{N}^2} (A_m \cap \mathcal{X}_n)$ und

$$\tilde{\mu}\left(A_m \cap \mathcal{X}_n\right) = \int_{A_m \cap \mathcal{X}_n} h \, d\mu \leq m \cdot \int_{A_m \cap \mathcal{X}_n} 1 \, d\mu \leq m \cdot \mu(\mathcal{X}_n) < \infty.$$

Also ist $\tilde{\mu}$ σ-endlich.
Ferner gilt

$$\begin{aligned} P_\vartheta(B) &= \int_B p_\vartheta(x) \, \mu(dx) \\ &= \int_B h(x) \cdot c(\vartheta) \exp\left(\sum_{j=1}^k q_j(\vartheta) \cdot T_j(x) \right) \mu(dx) \\ &= \int_B c(\vartheta) \cdot \exp\left(\sum_{j=1}^k q_j(\vartheta) \cdot T_j(x) \right) \tilde{\mu}(dx), \end{aligned}$$

siehe Stochastik I, da $d\tilde{\mu} = h \, d\mu$.

Bezeichnen wir mit $\mu * T$ das Bildmaß von T bzgl. μ auf $\mathbb{B}^k$, d.h. $(\mu * T)(B) = \mu(T^{-1}(B))$, $B \in \mathbb{B}^k$ (s. Stochastik I) mit $T = (T_1, \ldots, T_k)$, so ist das Bildmaß $P_\vartheta * T$ von T bzgl. P_ϑ absolut stetig bzgl. $\mu * T$ und besitzt die $\mu * T$-Dichte

$$(t_1, \ldots t_k) \in \mathbb{R}^k \mapsto c(\vartheta) \exp\left(\sum_{j=1}^k q_j(\vartheta) \cdot t_j \right). \tag{2.4.5}$$

Denn für ein beliebiges $B \in \mathbb{B}^k$ gilt nach dem Transformationssatz (s. Stochastik I)

$$\begin{aligned}
&\int_B c(\vartheta) \cdot \exp\left(\sum_{j=1}^{k} q_j(\vartheta) \cdot t_j\right) (\mu * T)\, d(t_1, \ldots, t_k) \\
&= \int_{T^{-1}(B)} c(\vartheta) \cdot \exp\left(\sum_{j=1}^{k} q_j(\vartheta) \cdot T_j\right) \mu(dx) \\
&= P_\vartheta(T^{-1}(B)) = (P_\vartheta * T)(B).
\end{aligned}$$

Sowohl in (2.4.2) also auch in (2.4.5) spielt der Faktor $c(\vartheta)$ nur die Rolle eines Normierungsfaktors, damit jeweils das Integral 1 ergibt.
Die Verteilung P_ϑ bzw. $P_\vartheta * T$ hängt also nur über $q(\vartheta) := (q_1(\vartheta), \ldots, q_k(\vartheta))$ von ϑ ab.
Dies nimmt man zum Anlass, die Menge

$$\tilde{\Theta} := \left\{ (q_1, \ldots, q_k) \in \mathbb{R}^k : \int_{\mathcal{X}} \exp\left(\sum_{j=1}^{k} q_j T_j(x)\right)\ \mu(dx) < \infty \right\}$$

als neuen Parameterraum, den sog. *natürlichen Parameterraum* für die k-parametrige Exponentialfamilie $\mathcal{P}$ einzuführen.
Dabei werden mit

$$K(q) := \frac{1}{\int \exp\left(\sum_{j=1}^{k} q_j T_j\right)\ d\mu}$$

durch

$$p_q(x) := K(q) \cdot \exp\left(\sum_{j=1}^{k} q_j T_j(x)\right),\ x \in \mathcal{X},\ q \in \tilde{\Theta}, \tag{2.4.6}$$

Wahrscheinlichkeitsdichten bzgl. μ definiert.
Anstelle der ursprünglich vorgegebenen Exponentialfamilie $\mathcal{P}$ betrachtet man dann häufig die k-parametrige Exponentialfamilie mit natürlichem Parameterraum.

$$\tilde{\mathcal{P}} := \left\{ P_q : q \in \tilde{\Theta} \right\} \text{ mit } P_q(B) = \int_B p_q\, d\mu,\ B \in \mathcal{B},\ q \in \tilde{\Theta} \tag{2.4.7}$$

Lemma 2.4.8. *Der natürliche Parameterraum $\tilde{\Theta} \subset \mathbb{R}^k$ ist stets eine konvexe Menge.*

Beweis: Es seinen $q', q'' \in \tilde{\Theta}$, $\gamma \in [0,1]$, also

$$\int \exp\left(\sum_{j=1}^{k} q_j' T_j\right) \, d\mu < \infty, \ \int \exp\left(\sum_{j=1}^{k} q_j'' T_j\right) \, d\mu < \infty.$$

Aus der Monotonie und der Konkavität der Logarithmus-Funktion $\log(x)$ mit $x > 0$ folgt:

$$\prod_{i=1}^{n} a_i^{\lambda_i} \leq \sum_{i=1}^{n} a_i \lambda_i \text{ für } a_i > 0, \ \lambda_i \geq 0, \ \sum_{i=1}^{n} \lambda_i = 1$$

und damit

$$\begin{aligned}
&\exp\left(\sum_{j=1}^{k} (\gamma q_j' + (1-\gamma) q_j'') T_j\right) \\
&\quad = \left(\exp\left(\sum_{j=1}^{k} q_j' T_j\right)\right)^{\gamma} \cdot \left(\exp\left(\sum_{j=1}^{k} q_j'' T_j\right)\right)^{1-\gamma} \\
&\quad \leq \gamma \cdot \exp\left(\sum_{j=1}^{k} q_j' T_j\right) + (1-\gamma) \cdot \exp\left(\sum_{j=1}^{k} q_j'' T_j\right)
\end{aligned}$$

Hieraus folgt

$$\begin{aligned}
&\int \exp\left(\sum_{i=1}^{k} (\gamma q_j' + (1-\gamma) q_j'') T_j\right) \, d\mu \\
&\quad \leq \gamma \int \exp\left(\sum_{j=1}^{k} q_j' T_j\right) \, d\mu + (1-\gamma) \int \cdot \exp\left(\sum_{j=1}^{k} q_j'' T_j\right) \, d\mu \\
&\quad < \infty.
\end{aligned}$$

□

Im Folgenden untersuchen wir den Fall $k = 1$ einer einparametrigen Exponentialfamilie mit natürlichem Parameterraum $\tilde{\Theta}$. In diesem Fall ist

$$P_q(B) = \int_B p_q \, d\mu, \quad B \in \mathcal{B}, \quad \text{mit } p_q(x) = K(q) \cdot \exp(qT(x)), \quad q \in \tilde{\Theta}$$

wobei $T : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$ und $\tilde{\Theta}$ ein (möglicherweise entartetes) Intervall in $\mathbb{R}$ ist.

Lemma 2.4.9. *Die Gütefunktion*

$$\tilde{\Theta} \ni q \mapsto E_q(\varphi) := \int \varphi \, dP_q = \int_{\mathcal{X}} \varphi(x) P_p(dx) = \int_{\mathcal{X}} \varphi(x) p_q(x) \; \mu(dx)$$

eines beliebigen Tests $\varphi \in \Phi$ ist in allen inneren Punkten von $\tilde{\Theta}$ beliebig oft differenzierbar.
Speziell gilt:

$$\frac{d \; E_q(\varphi)}{dq} = E_q(\varphi \cdot T) - E_q(\varphi) \cdot E_q(T). \tag{2.4.10}$$

Beweis: Zunächst beweisen wir:
Behauptung: Für alle $\varphi \in \Phi$ ist die Funktion

$$\psi : \tilde{\Theta} \to \mathbb{R}, \quad \psi(q) := \int \varphi(x) \exp(qT(x)) \; \mu(dx)$$

in allen inneren Punkten von $\tilde{\Theta}$ beliebig oft differenzierbar und dort gilt

$$\psi'(q) = \int \varphi(x) T(x) \exp(qT(x)) \, \mu(dx) \tag{2.4.11}$$

Denn: Sei dazu q_0 ein innerer Punkt von $\tilde{\Theta}$

$$\Rightarrow \exists \delta > 0 \; \forall q \in \tilde{\Theta}, \; |q - q_0| < \delta : \psi(q) < \infty$$

Ferner gilt:

$$\begin{aligned}
\frac{\psi(q) - \psi(q_0)}{q - q_0} &= \int \frac{\exp(qT) - \exp(q_0 T)}{q - q_0} \varphi \; d\mu \\
&= \int \varphi \cdot \exp(q_0 T) \cdot \frac{exp((q - q_0)T) - 1}{q - q_0} \; d\mu
\end{aligned}$$

Aus der Reihenentwicklung von $\exp(x)$ folgt zunächst für $|z| \leq \delta, \; t \in \mathbb{R}$

$$\begin{aligned}
\left| \frac{\exp(zt) - 1}{z} \right| &= \left| \frac{\sum_{n=1}^{\infty} \frac{(zt)^n}{n!}}{z} \right| \\
&= \left| \sum_{n=1}^{\infty} \frac{z^{n-1} t^n}{n!} \right| \\
&\leq \sum_{n=1}^{\infty} \frac{|z|^{n-1} \cdot |t|^n}{n!} \\
&\leq \frac{1}{\delta} \sum_{n=1}^{\infty} \frac{\delta^n |t|^n}{n!} \\
&\leq \frac{1}{\delta} \exp(\delta |t|).
\end{aligned}$$

Somit gilt für obigen Integranden:

$$
\begin{aligned}
&\left|\varphi \cdot \exp(q_0 T) \frac{\exp((q-q_o)T)-1}{q-q_0}\right| \\
&\qquad \leq \exp(q_0 T) \cdot \frac{\exp(\delta|t|)}{\delta} \\
&\qquad = \frac{1}{\delta} \exp(q_0 T + \delta|T|) \\
&\qquad \leq \frac{1}{\delta} \left(\exp((q_0+\delta)T) + \exp((q_0-\delta)T)\right)
\end{aligned}
$$

für $|q-q_0| \leq \delta$.
Falls also $|q-q_0| \leq \delta$, so ist $q_0 \pm \delta \in \tilde{\Theta}$, d.h. die rechte Seite in obiger Formel ist μ integrierbar.
Nach dem Satz der dominierten Konvergenz folgt somit für jede Folge q_n, $n \in \mathbb{N}$, in $\tilde{\Theta}$ mit $\lim_{n\to\infty} q_n = q_0$

$$
\lim_{n\to\infty} \frac{\psi(q_n)-\psi(q_0)}{q_n - q_0} = \int \varphi \cdot T \cdot \exp(q_0 T)\, d\mu,
$$

d.h. ψ ist einmal differenzierbar und es gilt (2.4.10). Die Existenz höherer Ableitungen folgt analog durch vollständige Induktion. Also gilt die obige Behauptung.
Für $\varphi \equiv 1$ erhalten wir daraus, dass $\frac{1}{K(q)} = \int \exp(qT)\, d\mu$ und damit auch $K(q)$ in allen inneren Punkten von $\tilde{\Theta}$ beliebig oft differenzierbar ist mit

$$
\left(\frac{1}{K(q)}\right)' = \int T \cdot \exp(qT)\, d\mu.
$$

Da andererseits

$$
\left(\frac{1}{K(q)}\right)' = -\frac{K'(q)}{K^2(q)}
$$

folgt wegen $E_q(\varphi) = \int \varphi\, dP_q = K(q) \cdot \int \varphi \exp(qT)\, d\mu$

$$
\begin{aligned}
\frac{dE_q(\varphi)}{dq} &= K(q) \cdot \psi'(q) + K'(q) \cdot \psi(q) \\
&= K(q) \cdot \int \varphi T \exp(qT)\, d\mu - K^2(q)\psi(q) \int T \exp(qT)\, d\mu \\
&= E_q(\varphi T) - K(q) \cdot E_q(\varphi) \int T \exp(qT)\, d\mu \\
&= E_q(\varphi T) - E_q(\varphi) \cdot E_q(T)
\end{aligned}
$$

für alle inneren Punkte q von $\tilde{\Theta}$. $\square$

2.5 Einseitige Tests bei monotonem Dichtequotienten

Definition 2.5.1. Eine Familie $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ mit $\Theta \subset \mathbb{R}$ und $P_\vartheta \neq P_{\vartheta'}$ für $\vartheta \neq \vartheta'$ heißt *Familie mit monotonem Dichtequotienten in T*, falls gilt: $\exists$ Maß $\mu/\mathcal{B}$, $\mathcal{P} \ll \mu$,

$$\forall\, \vartheta \in \Theta\ \exists\, p_\vartheta \in dP_\vartheta/d\mu\ \exists\, T : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})\ \forall \vartheta',\ \vartheta'' \in \Theta,\ \vartheta' < \vartheta'' :$$

$$\frac{p_{\vartheta''}(x)}{p_{\vartheta'}(x)} = h_{\vartheta',\vartheta''}(T(x)), \tag{2.5.2}$$

für alle $x \in \mathcal{X}_{\vartheta',\vartheta''} := \mathcal{X} \backslash \{x \in \mathcal{X} :\ p_{\vartheta'}(x) = p_{\vartheta''}(x) = 0\}$, (mit $h_{\vartheta',\vartheta''}(T(x)) = \infty$, falls $p_{\vartheta''}(x) > 0$, $p_{\vartheta'}(x) = 0$), wobei $h_{\vartheta',\vartheta''}(t)$ streng monoton in T ist.

Beispiel 2.5.3. (i) Es sei $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine einparametrige Exponentialfamilie, d.h. $p_\vartheta(x) = c(\vartheta) \cdot \exp(q(\vartheta) \cdot T(x))$, $x \in \mathcal{X}$, $\vartheta \in \Theta$. Falls $q(\vartheta)$ streng monoton wachsend in ϑ ist, so ist $\mathcal{P}$ eine Familie mit monotonem Dichtequotienten: Für $\vartheta', \vartheta'' \in \Theta$ mit $\vartheta' < \vartheta''$ gilt

$$\begin{aligned} \frac{p_{\vartheta''(x)}}{p_{\vartheta'}(x)} &= \frac{c(\vartheta'')}{c(\vartheta')} \cdot \exp\left((q(\vartheta'') - q(\vartheta')) \cdot T(x)\right) \\ &=: h_{\vartheta',\vartheta''}(T(x)), \end{aligned}$$

wobei $h_{\vartheta',\vartheta''}(t)$ streng monoton wachsend in t ist.

(ii) Sind $X_1, \ldots, X_n$ unabhängige und identisch $N(0, \sigma^2)$-verteilte Zufallsvariablen, so ist $\sigma^{-2} \sum_{i=1}^n X_i^2$ χ^2-verteilt mit n Freiheitsgraden, i.Z. χ^2_n.
Die Summe $\sum_{i=1}^n X_i^2$ besitzt dann die Lebesgue-Dichte

$$\begin{aligned} p_{\sigma^2}(x) &= \frac{1}{\sigma^2} g_n\left(\frac{x}{\sigma^2}\right) \\ &= \frac{1}{(2\sigma^2)^{n/2}\Gamma(n/2)} x^{\frac{n}{2}-1} \exp\left(-\frac{x}{2\sigma^2}\right),\ x > 0, \end{aligned}$$

wobei g_n die Dichte der χ^2_n-Verteilung ist.

Für $\vartheta = \sigma^2 \in \Theta := (0, \infty)$ sei P_ϑ das Wahrscheinlichkeitsmaß mit der λ-Dichte p_ϑ. Dann besitzt $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ einen monotonen Dichtequotienten in $T(x) = x$. Denn für $\sigma_1^2 < \sigma_2^2$ und $x > 0$ gilt

$$\frac{p_{\sigma_2^2}(x)}{p_{\sigma_1^2(x)}} = \left(\frac{\sigma_1^2}{\sigma_2^2}\right)^{n/2} \cdot \exp\left(\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\right) \cdot x\right) =: h_{\sigma_1^2,\sigma_2^2}(x),$$

wobei $h_{\sigma_1^2,\sigma_2^2}(x)$ streng monoton wachsend in t ist.

(iii) Für $\vartheta > 0$ sei P_ϑ die Poisson-Verteilung zum Parameter $\vartheta > 0$, d.h.

$$P_\vartheta(\{k\}) = \exp(-\vartheta)\frac{\vartheta^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Dann ist $\mathcal{P} = \{P_\vartheta : \vartheta > 0\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B}) = (\mathbb{N} \cup \{0\}, \mathbb{P}(\mathbb{N} \cup \{0\}))$, welche durch das Zählmaß $\mu_B = \sum_{n=0}^{\infty} 1_B(n)$, $B \subset \mathbb{N} \cup \{0\}$ dominiert wird.

$\mathcal{P}$ ist eine Familie mit monotonem Dichtequotienten in $T(x) = x$. Setze $p_\vartheta(x) := e^{-\vartheta}\vartheta^x/x!$, $x \in \mathcal{X}$, $\vartheta \in \Theta$. Dann gilt $p_\vartheta \in dP_\vartheta/d\mu$, denn

$$\begin{aligned} P_\vartheta(B) &= \sum_{k \in B} P_\vartheta(\{k\}) = \sum_{k \in B} e^{-\vartheta}\frac{\vartheta^k}{k!} = \sum_{k=0}^{\infty} 1_B(\{k\}) p_\vartheta(k) \\ &= \int_{\mathcal{X}} 1_B(x) \cdot p_\vartheta(x)\, \mu(dx) = \int_B p_\vartheta(x) \cdot \mu(dx), \quad B \in \mathcal{B} \end{aligned}$$

und für $\vartheta_1 < \vartheta_2$ ist

$$\frac{p_{\vartheta_2}(x)}{p_{\vartheta_1}(x)} = \exp((\vartheta_1 - \vartheta_2)) \left(\frac{\vartheta_2}{\vartheta_1}\right)^x$$

streng monoton wachsend in x.

Mit den in Abschnitt 2.2 bewiesenen Aussagen können wir nun gleichmäßig beste Tests zum Niveau α für das einseitige Testproblem $H : \vartheta \leq \vartheta_0$ gegen $K : \vartheta > \vartheta_0$ (bzw. entsprechend für $H : \vartheta \geq \vartheta_0$ gegen $K : \vartheta < \vartheta_0$) herleiten.

Satz 2.5.4. *$\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta \subset \mathbb{R}\}$ sei eine Familie von Wahrscheinlichkeitsmaßen mit monotonem Dichtequotienten in T. Ferner sei $\alpha \in (0, 1)$ und $\vartheta_0 \in \Theta$. Dann existiert für das Testproblem $H : \vartheta \leq \vartheta_0$ gegen $K : \vartheta > \vartheta_0$ ein gleichmäßig bester Test φ^* zum Niveau α.*
Dieser hat die Gestalt

$$\varphi^* = \begin{cases} 1, & \textit{falls } T(x) > c_\alpha \\ \delta_\alpha & \textit{falls } T(x) = c_\alpha \\ 0, & \textit{falls } T(x) < c_\alpha, \quad x \in \mathcal{X} \end{cases} \tag{2.5.5}$$

wobei c_α möglichst klein und – nach Wahl von c_α – $\delta_\alpha \in [0, 1]$ möglichst groß derart gewählt wird, dass

$$E_{\vartheta_0}(\varphi^*) = P_{\vartheta_0}(T > c_\alpha) + \delta_\alpha P_{\vartheta_0}(T = c_\alpha) = \alpha. \tag{2.5.6}$$

Darüber hinaus besitzt φ^* die Eigenschaft

$$\forall \vartheta \in \Theta : E_\vartheta(\varphi^*) = \inf\{E_\vartheta(\varphi) : \varphi \in \Phi \textit{ mit } E_{\vartheta_0}(\varphi) = \alpha\} \tag{2.5.7}$$

Beweis: Es sei zunächst $\vartheta_1 > \vartheta_0$ beliebig aber fest gewählt. Nach dem Neymann-Pearson Lemma 2.2.5, 2.2.6 existiert für das Testproblem $H_0 : \vartheta = \vartheta_0$ gegen $K_0 : \vartheta = \vartheta_1$ ein bester Test zum Niveau α der Form

$$\hat{\varphi}(x) = \begin{cases} 1, & \text{falls } p_{\vartheta_1}(x) > \gamma_\alpha \, p_{\vartheta_0}(x) \\ \delta_\alpha & \text{falls } p_{\vartheta_1}(x) = \gamma_\alpha \, p_{\vartheta_0}(x) \\ 0, & \text{falls } p_{\vartheta_1}(x) < \gamma_\alpha \, p_{\vartheta_0}(x), \quad x \in \mathcal{X} \end{cases} \tag{2.5.8}$$

mit $E_{\vartheta_0}(\hat{\varphi}) = \alpha$. Da $\mathcal{P}$ einen monotonen Dichtequotienten in T besitzt, gilt

$$\hat{\varphi}(x) = \begin{cases} 1, & \text{falls } h_{\vartheta_0,\vartheta_1}(T(x)) > \gamma_\alpha \\ \delta_\alpha & \text{falls } h_{\vartheta_0,\vartheta_1}(T(x)) = \gamma_\alpha \\ 0, & \text{falls } h_{\vartheta_0,\vartheta_1}(T(x)) < \gamma_\alpha, \quad x \in \mathcal{X} \end{cases} \tag{2.5.9}$$

Da $h_{\vartheta_0,\vartheta_1}(t)$ streng monoton wachsend in t ist, ist (2.5.9) äquivalent zu (2.5.5). Wähle man nun

$$c_\alpha = \inf\{t \in \mathbb{R} : P_{\vartheta_0}(T \leq t) \geq 1 - \alpha\} = \inf\{t \in \mathbb{R} : P_{\vartheta_0}(T > t) < \alpha\}$$

und δ_α so, dass $P_{\vartheta_0}(T > c_\alpha) + \delta_\alpha P_{\vartheta_0}(T = c_\alpha) = \alpha$. Dann ist der zugehörige Test φ^* ein bester Test für H_0 gegen K_0; da c_α und δ_α offenbar nur von ϑ_0 abhängen und nicht von ϑ_1 ist φ^* auch gleichmäßig bester Test zum Niveau α für $H_0 : \vartheta = \vartheta_0$ gegen $K_0 : \vartheta > \vartheta_0$. Durch Vergleich mit dem Test $\varphi_\alpha = \alpha$ ergibt sich aus (2.5.7) (was aber erst noch bewiesen werden muss) $\forall \vartheta < \vartheta_0 : E_\vartheta(\varphi^*) \leq \alpha$, d.h. φ^* ist ein Test zum Niveau α für $H : \vartheta \leq \vartheta_0$ gegen $K : \vartheta > \vartheta_0$.
Daraus folgt, dass φ^* gleichmäßig bester Test zum Niveau α für $H_0 : \vartheta \leq \vartheta_0$ gegen $K_0 : \vartheta > \vartheta_0$ ist. Zum Nachweis von (2.5.7) werden wir zunächst einen Test $\psi \in \tilde{\Phi} := \{\varphi \in \Phi : E_{\vartheta_0}(\varphi) = 1 - \alpha\}$ konstruieren mit der Eigenschaft

$$\forall \vartheta < \vartheta_0 : \; E_\vartheta(\psi^*) = \sup_{\psi \in \tilde{\Phi}} E_\vartheta(\psi).$$

Sei dazu $\vartheta_2 < \vartheta_0$ beliebig. Nach dem Neyman-Pearson-Lemma existiert $\psi^* \in \tilde{\Phi}$ mit $E_{\vartheta_2}(\psi^*) = \sup_{\psi \in \tilde{\Phi}} E_{\vartheta_2}(\psi)$, wobei ψ^* die Gestalt besitzt

$$\psi^*(x) = \begin{cases} 1, & \text{falls } p_{\vartheta_2}(x) > \tilde{\gamma} \, p_{\vartheta_0}(x) \\ \tilde{\delta} & \text{falls } p_{\vartheta_2}(x) = \tilde{\gamma} \, p_{\vartheta_0}(x) \\ 0, & \text{falls } p_{\vartheta_2}(x) < \tilde{\gamma} \, p_{\vartheta_0}(x), \quad x \in \mathcal{X} \end{cases}$$

bzw.

$$\psi^*(x) = \begin{cases} 1, & \text{falls } T(x) > \tilde{c} \\ \tilde{\delta} & \text{falls } T(x) = \tilde{c} \\ 0, & \text{falls } T(x) < \tilde{c}, \quad x \in \mathcal{X} \end{cases}$$

Da andererseits $1 - \varphi^*$ auch diese Gestalt besitzt und $E_\vartheta(1 - \varphi^*) = 1 - \alpha$, folgt $\psi^* := 1 - \varphi^* \in \tilde{\Phi}$ als Lösung von $E_{\vartheta_0}(\psi^*) = \sup_{\psi \in \tilde{\Phi}} E_{\vartheta_2}(\psi)$.

Da $1-\varphi^*$ unabhängig von $\vartheta_2 < \vartheta_0$ ist, gilt also $E_\vartheta(1-\varphi^*) = \sup_{\psi\in\tilde{\Phi}} E_\vartheta(\psi)$, $\vartheta < \vartheta_0$. Dies ist aber äquivalent zu (2.5.7).

$\square$

Bemerkung 2.5.10. (i) Der in Satz 2.5.4 konstruierte gleichmäßig beste Test φ^* zum Niveau α für $H : \vartheta \leq \vartheta_0$ gegen $K : \vartheta > \vartheta_0$ ist auch ein bester Test zum Niveau $\alpha' := E_{\vartheta'}(\varphi^*)$ für $H' : \vartheta = \vartheta'$ gegen $K' : \vartheta = \vartheta''$ für beliebiges $\vartheta', \vartheta'' \in \Theta$ mit $\vartheta' < \vartheta''$. Denn φ^* ist nach Konstruktion ein Test vom NP-Typ für $(P_{\vartheta'}, P_{\vartheta''})$ und nach Lemma 2.2.6 damit trennscharf für $(P_{\vartheta'}, P_{\vartheta''})$ daraus folgt, dass die Gütefunktion $\beta(\vartheta) := E_\vartheta(\varphi^*)$ auf der Menge $\{\vartheta \in \Theta : 0 < \beta < 1\}$ streng monoton wächst:
Seien $\vartheta', \vartheta'' \in \Theta$ mit $\vartheta' < \vartheta''$ und $0 < \beta(\vartheta') < 1$. Dann ist φ^* bester Test zum Niveau $\beta(\vartheta')$ für $H' : \vartheta = \vartheta'$ gegen $K' : \vartheta = \vartheta''$, also gilt nach Korollar 2.2.10 $\beta(\vartheta') < \beta(\vartheta'')$, da nach Voraussetzung stets $P_{\vartheta'} \neq P_{\vartheta''}$ für $\vartheta' \neq \vartheta''$.

(ii) In Analogie zu Satz 2.5.4 ergibt sich sofort, dass es auch für das Testproblem $\tilde{H} : \vartheta \geq \vartheta_0$ gegen $\tilde{K} : \vartheta < \vartheta_0$ einen gleichmäßig besten Test zum Niveau $\alpha \in (0,1)$ gibt, nämlich:

$$\tilde{\varphi}^*(x) = \begin{cases} 1, & \text{falls } T(x) < \tilde{c}_\alpha \\ \tilde{\delta}_\alpha & \text{falls } T(x) = \tilde{c}_\alpha \\ 0, & \text{falls } T(x) > \tilde{c}_\alpha, \quad x \in \mathcal{X}, \end{cases}$$

wobei die Konstanten $\tilde{c}_\alpha$ und $\tilde{\delta}_\alpha \in [0,1]$ bestimmt werden aus der Gleichung $E_{\vartheta_0}(\tilde{\varphi}^*) = P_{\vartheta_0}(T < \tilde{c}_\alpha) + \tilde{\delta}_\alpha P_{\vartheta_0}(T = \tilde{c}_\alpha) = \alpha$

(iii) Ferner folgt aus Lemma 2.2.9, dass die Bereiche strikter Ablehnung der gleichmäßig besten Tests zum Niveau $\alpha \in (0,1)$ für $H_0 : \vartheta = \vartheta_0$ gegen $K : \vartheta > \vartheta_0$ bzw. $\tilde{K} : \vartheta < \vartheta_0$ bis auf μ–Nullmengen von der Form $\{T > c_\alpha\}$ bzw. $\{T < \tilde{c}_\alpha\}$ sind. Folglich gibt es i.A. keinen gleichmäßig besten Test zum Niveau $\alpha \in (0,1)$ für $H_0 : \vartheta = \vartheta_0$ gegen $K_0 : \vartheta \neq \vartheta_0$, da ja ein solcher Test sowohl gegen K, als auch gegen $\tilde{K}$ gleichmäßig bester Test sein müsste.

Beispiel 2.5.11. Im Beispiel 2.4.3 auf Seite 45 haben wir gesehen, dass bei bekanntem Mittelwert μ_0 die Familie $\mathcal{P}_1 = \{N(\mu_0, \sigma^2) : \sigma^2 \in (0,\infty)\}$ eine einparametrige Exponentialfamilie bildet $c_1(\vartheta) = \frac{1}{\sqrt{2\pi}\sigma}$, $q_1(\vartheta) = \frac{-1}{(2\sigma^2)}$ und $T_1(x) = (x-\mu_0)^2$, $x \in \mathbb{R}$ und $\vartheta = \sigma^2 \in \Theta := (0,\infty)$, wenn $\mu := \lambda$ als dominierendes Maß gewählt wird. Dann ist im Stichprobenraum $(\mathcal{X}, \mathcal{B}) := (\mathbb{R}^n, \mathbb{B}^n)$ gemäß Lemma 2.4.5 die Familie $\mathcal{P} := \{N(\mu_0, \sigma^2)^n : \sigma^2 \in (0,\infty)\}$ eine einparametrige Exponentialfamilie mit $T(x) = \sum_{i=1}^n (x_i - \mu_0)^2$, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, und mit streng monoton wachsendem $q(\vartheta) = q_1(\vartheta) =$

$-\frac{1}{(2\vartheta)}$, so dass wir nach Beispiel 2.5.3 (i) und Bemerkung 2.5.10 (ii) folgendes erhalten:
$\mathcal{P}$ ist eine Familie mit monotonem Dichtequotienten in T, und der durch

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } \sum_{i=1}^n (x_i - \mu_0)^2 \leq c_\alpha \\ 0, & \text{falls } \sum_{i=1}^n (x_i - \mu_0)^2 > c_\alpha, \quad x \in \mathbb{R}^n, \end{cases}$$

definierte Test mit c_α so gewählt, dass $E_{\vartheta_0}(\varphi^*) = \alpha$, $\vartheta_0 = \sigma_0^2$, ist ein gleichmäßig bester Test zum Niveau α für das Testproblem

$$H : \sigma^2 \geq \sigma_0^2 \quad \text{gegen} \quad K : \sigma^2 < \sigma_0^2$$

2.6 Gleichmäßig beste Tests in einparametrigen Exponentialfamilien

In diesem Abschnitt sei Θ stets ein reelles Intervall. Wir wollen gleichmäßig beste Test zum Niveau α für das Testproblem

$$\begin{aligned} H &:= \{\vartheta \in \Theta : \vartheta \leq \vartheta_1\} \cup \{\vartheta \in \Theta : \vartheta \geq \vartheta_2\} \quad \text{gegen} \\ K &:= \{\vartheta \in \Theta : \vartheta_1 < \vartheta < \vartheta_2\} \end{aligned} \tag{2.6.1}$$

bestimmen, wobei $\vartheta_1, \vartheta_2 \in \Theta$ mit $\vartheta_1 < \vartheta_2$ vorgegeben sind (vgl. Beispiel 2.3.10).

Satz 2.6.2. *Es sei $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine einparametrige Exponentialfamilie im Stichprobenraum $(\mathcal{X}, \mathcal{B})$, d.h. bzgl. eines geeigneten σ–endlichen Maßes $\mu/\mathcal{B}$ besitzt P_ϑ eine Dichte der Form $p_\vartheta(x) = c(\vartheta)\exp(q(\vartheta)T(x))$, $x \in \mathcal{X}, \vartheta \in \Theta$. Wir setzen voraus, dass $q(\vartheta)$ streng monoton wächst, und T nicht μ-f.s. konstant ist. Dann existiert für das Testproblem (2.6.1) ein gleichmäßig bester Test φ^* zum Niveau $\alpha \in (0,1)$ der Form*

$$\varphi^*(x) = \begin{cases} 1, & \textit{falls } c_1 < T(x) < c_2 \\ \delta_j, & \textit{falls } T(x) = c_j,\ j = 1,2,\ c_1 < c_2 \qquad x \in \mathcal{X}. \\ 0, & \textit{falls } T(x) \in (-\infty, c_1) \cup (c_2, \infty) \end{cases} \tag{2.6.3}$$

Dabei wird man c_1 möglichst klein und c_2 möglichst groß wählen und nach deren Festlegung $\delta_j \in [0,1]$, $j = 1,2$, so bestimmen, dass

$$\begin{aligned} E_{\vartheta_i}(\varphi^*) &= P_{\vartheta_i}(c_1 < T < c_2) + \delta_1 P_{\vartheta_i}(T = c_1) + \delta_2 P_{\vartheta_i}(T = c_2) \\ &= \alpha, \quad i = 1,2 \end{aligned} \tag{2.6.4}$$

gilt. Darüber hinaus gilt

$$E_\vartheta(\varphi^*) = \inf\{E_\vartheta(\varphi) : \varphi \in \Phi \textit{ mit } E_{\vartheta_1}(\varphi) = E_{\vartheta_2}(\varphi) = \alpha\},\ \vartheta \in H. \tag{2.6.5}$$

Beweis: Sei zunächst $\tilde{\vartheta} \in \Theta$ mit $\vartheta_1 < \tilde{\vartheta} < \vartheta_2$ beliebig aber fest gewählt.
Behauptung: P_{ϑ_1} und P_{ϑ_2} sind linear unabhängig.
Angenommen dies wäre nicht der Fall. Dann wäre für ein $\gamma > 0$

$$c(\vartheta_1)\exp(q(\vartheta_1)T(x)) = \gamma\, c(\vartheta_2)\exp(q(\vartheta_2)T(x)) \quad \mu - \text{f.ü.}$$

$\Leftrightarrow$

$$\frac{c(\vartheta_1)}{c(\vartheta_2)} = \gamma\ \exp((q(\vartheta_2) - q(\vartheta_1))T(x)) \quad \mu - \text{f.ü.}$$

d.h. wegen $q(\vartheta_2) - q(\vartheta_1) > 0$ wäre T μ-f.ü. konstant, im Widerspruch zur Voraussetzung.
Also sind P_{ϑ_1} und P_{ϑ_2} linear unabhängig und somit ist (α, α) gemäß Lemma 2.3.7 (i) innerer Punkt der Menge $M := \{(E_{\vartheta_1}(\varphi), E_{\vartheta_2}(\varphi)) : \varphi \in \Phi\}$.
Daher ist nach Satz 2.3.3 der durch

$$\varphi_0(x) = \begin{cases} 1, & \text{falls} \quad c(\tilde{\vartheta}) \cdot \exp(q(\tilde{\vartheta})T(x)) > \\ & \qquad \gamma_1 \cdot c(\vartheta_1) \cdot \exp(q(\vartheta_1) \cdot T(x)) \\ & \qquad\quad +\gamma_2 \cdot c(\vartheta_2) \cdot \exp(q(\vartheta_2)T(x)) \\ 0, & \text{falls} \quad c(\tilde{\vartheta}) \cdot \exp(q(\tilde{\vartheta})T(x)) < \\ & \qquad \gamma_1 \cdot c(\vartheta_1) \cdot \exp(q(\vartheta_1) \cdot T(x)) \\ & \qquad\quad +\gamma_2 \cdot c(\vartheta_2) \cdot \exp(q(\vartheta_2)T(x)), \end{cases} \qquad x \in \mathcal{X},$$

definierte Test ein bzgl. der Klasse $\tilde{\Phi}_\alpha := \{\varphi \in \Phi : \alpha = E_{\vartheta_1}(\varphi) = E_{\vartheta_2}(\varphi)\}$ bester Test für das Problem $H_0 := \{\vartheta_1, \vartheta_2\}$ gegen $K_0 := \{\tilde{\vartheta}\}$. Nach entsprechender Umformung folgt nun

$$\varphi_0(x) = \begin{cases} 1, & \text{falls } a(x) < 1 \\ 0, & \text{falls } a(x) > 1 \end{cases} \qquad x \in \mathcal{X},$$

wobei $a(x) := \gamma_1' \cdot \exp(b_1 T(x)) + \gamma_2' \cdot \exp(b_2 T(x))$, $\gamma_i' := \gamma_i \cdot c(\vartheta_i)/c(\tilde{\vartheta})$, $i = 1, 2$ und $b_1 = q(\vartheta_1) - q(\tilde{\vartheta}) < 0$, $b_2 = q(\vartheta_2) - q(\tilde{\vartheta}) > 0$.
Eine Fallunterscheidung zeigt, dass $\gamma_1' > 0$ und $\gamma_2' > 0$ gelten muss:
1. Fall: Angenommen, $\gamma_1' \le 0$, $\gamma_2' \le 0$. Dann folgt $a(x) \le 0$, $x \in \mathcal{X} \Rightarrow \varphi_0 \equiv 1$ $\Rightarrow \alpha = E_{\vartheta_1}(\varphi_0) = E_{\vartheta_2}(\varphi_0) = 1 \notin (0, 1)$, Widerspruch.
2. Fall: Angenommen, $\gamma_1' > 0$, $\gamma_2' \le 0$. Dann ist $a(x)$ streng monoton fallend in $T(x)$, also $a(x) < 1 \Leftrightarrow T(x) > d_1$. Damit ist der Test φ_0 von der Form

$$\varphi_0(x) = \begin{cases} 1, & \text{falls } T(x) > d_1 \\ 0, & \text{falls } T(x) < d_1 \end{cases} \qquad x \in \mathcal{X},$$

d.h. φ_0 ist ein optimaler Test für ein Testproblem der Form $\vartheta \le \vartheta_0$ gegen $\vartheta > \vartheta_0$ und besitzt daher nach Bemerkung 2.5.10 (i) eine streng monoton wachsende Gütefunktion, also speziell $E_{\vartheta_1}(\varphi_0) < E_{\vartheta_2}(\varphi_0)$, Widerspruch.
3. Fall: $\gamma_1' \le 0$, $\gamma_2' > 0$ analog.

Also gilt $\gamma_1' > 0$ und $\gamma_2' > 0$ und daher auch $\gamma_1 > 0$ und $\gamma_2 > 0$. Es folgt somit

$$\varphi_0(x) = \begin{cases} 1, & \text{falls } c_1 < T(x) < c_2 \\ 0, & \text{falls } T(x) < c_1 \text{ oder } T(x) > c_2 \end{cases} \qquad x \in \mathcal{X}.$$

Aus Lemma 2.3.4 folgt nun, dass der Test φ_0 und damit auch der durch (2.6.3) und (2.6.4) definierte Test φ^* ein gleichmäßig bester Test zum Niveau α für $H_0 = \{\vartheta_1, \vartheta_2\}$ gegen $K_0 = \{\tilde{\vartheta}\}$ ist. Da sich hierbei die Konstanten c_i und δ_i, $i = 1, 2$ aus den von dem gewählten $\tilde{\vartheta}$ unabhängigen Nebenbedingungen (2.6.4) ergeben, ist φ^* sogar ein gleichmäßig bester Test zum Niveau α für $H_0 = \{\vartheta_1, \vartheta_2\}$ gegen K. Durch Vergleich mit dem Test $\varphi \equiv \alpha$ folgt aus (2.6.5), dass $E_\vartheta(\varphi^*) \leq \alpha$, $\vartheta \in H$, d.h. φ^* ist ein gleichmäßig bester Test zum Niveau α für das Testproblem (2.6.1). Zum Nachweis von (2.6.5) sei $\tilde{\vartheta} \in H$, $\tilde{\vartheta} < \vartheta_1$, vorgegeben. Satz 2.3.3 liefert einen Test $\tilde{\psi}^*$, der bester Test ist bzgl. der Klasse

$$\tilde{\Phi}_{1-\alpha} := \{\varphi \in \Phi : E_{\vartheta_1}(\varphi) = E_{\vartheta_2}(\varphi) = 1 - \alpha\}$$

für das Testproblem $H_0 = \{\vartheta_1, \vartheta_2\}$ gegen $K_0 := \{\tilde{\vartheta}\}$. Für den Test $\tilde{\varphi}^* := 1 - \tilde{\psi}^*$ gilt also $\tilde{\varphi}^* \in \tilde{\Phi}_\alpha$ und $E_{\tilde{\vartheta}}(\tilde{\varphi}^*) = \inf\{E_{\tilde{\vartheta}}(\varphi) : \varphi \in \tilde{\Phi}_\alpha\}$. Gemäß 2.3.3 (iii) ist $\tilde{\psi}^*$ μ-f.ü. von der Form

$$\tilde{\psi}^*(x) = \begin{cases} 1, & \text{falls } \gamma_1'' \exp(b_1 T(x)) + \gamma_2'' \exp(b_2 T(x)) < 1 \\ 0, & \text{falls } \gamma_1'' \exp(b_1 T(x)) + \gamma_2'' \exp(b_2 T(x)) > 1 \end{cases} \qquad x \in \mathcal{X},$$

wobei hier $b_1 = q(\vartheta_1) - q(\tilde{\vartheta}) > 0$, $b_2 = q(\vartheta_2) - q(\tilde{\vartheta}) > b_1 > 0$. Hieraus folgt für die Konstanten γ_1'', γ_2'' notwendigerweise $\gamma_1'' > 0$ und $\gamma_2'' > 0$:
1. Fall: Angenommen $\gamma_1'' \leq 0$; $\gamma_2'' \leq 0$. Dann folgt $\tilde{\psi}^* = 1$ μ-f.ü. und damit $1 - \alpha = E_{\vartheta_1}(\tilde{\psi}^*) = E_{\vartheta_2}(\tilde{\psi}^*) = 1$, im Widerspruch zu $\alpha \in (0, 1)$
2. Fall: Angenommen $\gamma_1'' > 0$, $\gamma_2'' \geq 0$. Dann ist $b(x) := \gamma_1'' \exp(b_1 T(x)) + \gamma_2'' \exp(b_2 T(x))$ streng monoton wachsend in $T(x)$, also μ-f.ü.

$$\tilde{\psi}^*(x) = \begin{cases} 1, & \text{falls } T(x) < d_2 \\ 0, & \text{falls } T(x) > d_2 \end{cases} \qquad x \in \mathcal{X},$$

d.h. $\tilde{\psi}^*(x)$ ist optimaler Test für ein einseitiges Problem $\vartheta \geq \vartheta_0$ gegen $\vartheta < \vartheta_0$ und besitzt nach 2.5.10 (i) eine streng monoton fallende Gütefunktion, also speziell $E_{\vartheta_1}(\tilde{\psi}^*) > E_{\vartheta_2}(\tilde{\psi}^*)$, Widerspruch.
3. Fall: Angenommen $\gamma_1'' \leq 0$, $\gamma_2'' > 0$. Wegen $b_1 = b_2 - \epsilon_0$ mit $\epsilon_0 > 0$ ist $b(x) := \gamma_1'' \exp(b_2 T(x)) \exp(-\epsilon_0 T(x)) + \gamma_2'' \exp(b_2 T(x))$ streng monoton wachsend in $T(x)$ und man schließt wie in Fall 2.

Also gilt $\gamma_1'' > 0$, $\gamma_2'' < 0$ und daher

$$\tilde{\psi}^*(x) = \begin{cases} 1, & \text{falls } 1 < \frac{1}{\gamma_1''} \exp(-b_1 T(x)) - \frac{\gamma_2''}{\gamma_1''} \exp((b_2 - b_1) T(x)) \\ 0, & \text{falls } 1 > \frac{1}{\gamma_1''} \exp(-b_1 T(x)) - \frac{\gamma_2''}{\gamma_1''} \exp((b_2 - b_1) T(x)) \end{cases} \qquad x \in \mathcal{X},$$

bzw.

$$\tilde{\varphi}^*(x) = \begin{cases} 1, & \text{falls } c_1 < T(x) < c_2 \\ 0, & \text{falls } T(x) \notin [c_1, c_2] \end{cases} \qquad x \in \mathcal{X},$$

wobei die Konstanten $c_1,\ c_2$ aus $E_{\vartheta_1}(\tilde{\varphi}^*) = E_{\vartheta_2}(\tilde{\varphi}^*) = \alpha$ bestimmt werden. Dies bedeutet $\varphi_0 = \tilde{\varphi}^*$ μ-f.ü.
Hieraus folgt wegen $\tilde{\varphi}^* \in \tilde{\Phi}_\alpha$ die Behauptung (2.6.5). Analog schließt man im Fall $\tilde{\vartheta} > \vartheta_2$ □

Wie wir in 2.5.10 (iii) gesehen haben, existieren im Fall einparametriger Exponentialfamilien i.A. keine gleichmäßig besten Tests zum Niveau α für $H : \vartheta = \vartheta_0$ gegen $K : \vartheta \neq \vartheta_0$. Wir betrachten daher im Folgenden nur unverfälschte Tests zum Niveau α, wobei wir den natürlichen Parameterraum zugrunde legen, d.h. unsere Verteilungsannahme ist $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$, wobei P_ϑ die μ-Dichte $p_\vartheta(x) = c(\vartheta) \cdot \exp(\vartheta T(x))$ für $\vartheta \in \Theta$ und $x \in \mathcal{X}$ besitzt. Ferner nehmen wir an, dass Θ ein offenes Intervall in $\mathbb{R}$ ist. Ist dann φ ein unverfälschter Test zum Niveau α für $H : \vartheta = \vartheta_0$ gegen $K : \vartheta \neq \vartheta_0$, d.h. $E_{\vartheta_0}(\varphi) = \alpha$ und $\inf_{\vartheta \in K} E_\vartheta(\varphi) \geq \alpha$, so folgt aus Lemma 2.4.9

$$\frac{d}{d\vartheta}\, E_\vartheta(\varphi)|_{\vartheta=\vartheta_0} = E_{\vartheta_0}(\varphi\, T) - \alpha\, E_{\vartheta_0}(T) = 0, \tag{2.6.6}$$

da $E_\vartheta(\varphi)$ in ϑ_0 ein Minimum besitzt.

Der abschließende Satz zeigt, dass diese gegenüber der Unverfälschtheit abgeschwächte Bedingung (2.6.6) ausreicht, um die Existenz eines gleichmäßig besten unverfälschten Tests nachzuweisen.

Satz 2.6.7. *Es sei $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine einparametrige Exponentialfamilie mit natürlichem Parameterraum Θ = offenes Intervall in $\mathbb{R}$. Wähle $\vartheta_0 \in \Theta$. Dann existiert für das Testproblem $H : \vartheta = \vartheta_0$ gegen $K : \vartheta \neq \vartheta_0$ ein gleichmäßig bester unverfälschter Test φ^* zum Niveau $\alpha \in (0,1)$ der Gestalt:*

$$\varphi^*(x) = \begin{cases} 1, & \textit{falls } T(x) \notin [c_1, c_2] \\ \delta_j, & \textit{falls } T(x) = c_j,\ j = 1,2, \\ 0, & \textit{falls } T(x) \in (c_1, c_2) \end{cases} \qquad x \in \mathcal{X}. \tag{2.6.8}$$

Dabei werden die Konstanten c_j, $\delta_j \in [0,1]$, $j = 1,2$ so bestimmt, dass

$$E_{\vartheta_0}(\varphi^*) = \alpha \textit{ und } E_{\vartheta_0}(\varphi^* T) = \alpha\, E_{\vartheta_0}(T). \tag{2.6.9}$$

Darüber hinaus gilt

$$\begin{aligned} E_\vartheta(\varphi^*) = \sup\{E_\vartheta(\varphi) \ : \ & \varphi \in \Phi \textit{ mit } E_{\vartheta_0}(\varphi) = \alpha, \\ & E_{\vartheta_0}(\varphi\, T) = \alpha\, E_{\vartheta_0}(T)\},\ \vartheta \in K. \end{aligned} \tag{2.6.10}$$

Beweis: Siehe Satz 2.70 in Witting, (1985) Mathematische Statistik, Teubner, Stuttgart.

□

Beispiel 2.6.11. Es seien $X_1, \ldots, X_n$ unabhängige $B(1, \vartheta)$-verteilte Zufallsvariablen, $\vartheta \in \Theta := (0, 1)$. Gesucht ist ein gleichmäßig bester unverfälschter Test zum Niveau α für das Testproblem $H : \vartheta = \vartheta_0$ gegen $K : \vartheta \neq \vartheta_0$.

Die Familie der Verteilungen $P_\vartheta = B(1, \vartheta)^n$ von $X = (X_1, \ldots, X_n)$ bildet eine einparametrige Exponentialfamilie mit $T(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$. Daher ist (2.6.8), (2.6.9) eine Lösung, wobei die Werte c_j, δ_j wegen $P_\vartheta * T = B(n, \vartheta)$ gem. (2.6.9) aus der $B(n, \vartheta_0)$-Verteilung zu bestimmen sind.

Bei $n = 24$ unabhängigen Wiederholungen eines Experimentes mit zufälligem Ausgang und einer unbekannten Erfolgswahrscheinlichkeit $\vartheta \in (0, 1)$ haben sich $T = 12$ Erfolge eingestellt. Es ist für $\alpha = 0,05$ zwischen $H : \vartheta_0 = \frac{5}{16}$ und $K : \vartheta \neq \frac{5}{16}$ zu unterscheiden. Aus (2.6.9) ergeben sich $c_1 = 3$, $c_2 = 12$, $\delta_1 = 0,757$ und $\delta_2 = 0,398$.

Es erfolgt keine strikte Ablehnung von H, sondern es wird mit der Wahrscheinlichkeit 0,398 verworfen.

Kapitel 3

REDUKTION STATISTISCHER ENTSCHEIDUNGEN

3.1 Problemstellung

In den Abschnitten 2.5 und 2.6 haben wir gesehen, dass alle optimalen Lösungen der dort behandelten Testprobleme stets von der Form $\varphi^* = \psi \circ T$ waren, d.h. die Werte $\varphi^*(x)$ optimaler Tests hingen über $T(x)$ von der Stichprobe x ab.

Wie wir in diesem Kapitel sehen werden, gibt es auch bei anderen statistischen Entscheidungsproblemen häufig Stichprobenfunktionen (Statistiken) T, so dass man sich bei der Suche nach einer optimalen Lösung auf solche Entscheidungsfunktionen beschränken kann, die nur von T abhängen. Um also in solchen Fällen eine Entscheidung zu treffen, benötigt man also nicht die volle Information über $\vartheta \in \Theta$, die in der Stichprobe $x \in X$ enthalten ist, sondern lediglich diejenige, welche durch $T(x)$ gegeben ist.

Statistiken mit einer solchen Eigenschaft werden wir in einem noch zu präzisierenden Sinn *suffizient* (erschöpfend, hinreichend) nennen.

Beispiel 3.1.1. Es seien $x_1, \dots, x_n$ die Realisationen von n unabhängigen und identisch $B(1, \vartheta)$-verteilten Zufallsvariablen $X_1, \dots, X_n$ mit unbekannter Erfolgswahrscheinlichkeit $P(X_1 = 1) = \vartheta \in \Theta := (0, 1)$. Wegen der Unabhängigkeit der Einzelversuche kommt es nicht auf die Reihenfolge der Versuchsergebnisse an, sondern lediglich auf die Gesamtzahl $\sum_{i=1}^n x_i$ der Erfolge. Betrachten wir die zu $X_1, \dots, X_n$ gehörige Verteilungsannahme $\mathcal{P} = \{P_\vartheta = B(1, \vartheta)^n, \vartheta \in \Theta\}$, so hängt

$$P_\vartheta(\{x_1, \dots x_n\}) = \vartheta^{\sum_{i=1}^n x_i} \cdot (1 - \vartheta)^{n - \sum_{i=1}^n x_i}$$

nur von der Gesamtanzahl $T(x_1,\ldots,x_n) = \sum_{i=1}^n x_i$ der Erfolge ab, so dass anschaulich $T(x)$ die selbe Information über den unbekannten Parameter ϑ enthält wie die Stichprobe $\boldsymbol{x} = (x_1,\ldots,x_n)$ selbst.

In diesem Fall wird man vermuten, dass $T(x) = \sum_{i=1}^n x_i$ eine *suffiziente Statistik* ist.

Beispiel 3.1.2. Liegt eine einparametrige Exponentialfamilie $\mathcal{P}$ einem Experiment zugrunde, also Wahrscheinlichkeitsmaße P_ϑ auf $(\mathcal{X},\mathcal{B})$ mit μ-Dichten der Form $c(\vartheta)\exp(q(\vartheta)T)$, wobei $T:\mathcal{X}\to\mathbb{R}\quad \mathcal{B},\mathbb{B}$-messbar ist, so wird man auch hier vermuten, dass die Kenntnis über $\boldsymbol{x}$ keine bessere Information über $\vartheta\in\Theta$ liefert, als diejenige, die man durch $T(x)$ gewinnt. Dadurch reduziert sich das Entscheidungsprolem, etwa die Bestimmung optimaler Lösungen für Testprobleme dahingehend, dass man von der i.A. n-dimensionalen Zufallsgröße $X = (X_1,\ldots,X_n)$ zur eindimensionalen Zufallsgröße $T(x)$ übergeht und optimale Lösungen nur noch unter Funktionen einer Variablen zu bestimmen braucht.

Es wird sich zeigen, dass für die meisten der von uns betrachteten Klassen von Wahrscheinlichkeitsverteilungen Statistiken $T:(\mathcal{X},\mathcal{B})\to(\mathcal{Y},\mathcal{G})$ existieren, die in dem Sinn suffizient sind, dass sie eine „Reduktion statistischer Entscheidungsverfahren" im obigen Sinn ermöglichen.

Betrachten wir etwa die Verteilungsannahme $\mathcal{P} = \{P_\vartheta = B(1,\vartheta)^n : \vartheta\in\Theta = (0,1)\}$ aus Beispiel 3.1.1 auf $(\mathcal{X},\mathcal{B}) = (\{0,1\}^n, \mathbb{P}(\{0,1\})^n)$ und die Statistik $T(x) = \sum_{i=1}^n x_i$, so ist für jedes $A\in\mathcal{B}$ die bezüglich P_ϑ gebildete bedingte Wahrscheinlichkeit von A unter dem Ereignis $\{T=k\}$ für $k = 0,1,\ldots n$ unabhängig von dem Parameter $\vartheta\in\Theta$:

$$\begin{aligned} P_\vartheta(A|\{T=k\}) &= \frac{P_\vartheta(A\cap\{T=k\})}{P_\vartheta(\{T=k\})} \\ &= \frac{\sum_{x\in A\cap\{T=k\}} \vartheta^{\sum_{i=1}^n x_i}(1-\vartheta)^{n-\sum_{i=1}^n x_i}}{\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}} \\ &= \frac{\sum_{x\in A\cap\{T=k\}} \vartheta^k(1-\vartheta)^{n-k}}{\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}} \\ &= \frac{|A\cap\{T=k\}|}{\binom{n}{k}}, \quad k = 0,1,\ldots,n \end{aligned}$$

Somit kann also bei gegebenen $T = k$ keine zusätzliche Information über $\vartheta\in\Theta$ gewonnen werden, d.h. die „gesamte Abhängigkeit dieses Modells $\mathcal{P}$ von dem unbekannten Parameter $\vartheta\in\Theta$ steckt in $T(x)$".

Genauer gilt $\forall \vartheta \in \Theta$ und $\forall B \in \mathcal{B}$:

$$\begin{aligned}
P_\vartheta(B) &= \sum_{k=0}^{n} P_\vartheta(B \cap \{T = k\}) \\
&= \sum_{k=0}^{n} \underbrace{P_\vartheta(B|\{T = k\})}_{=:h(B,k)} \cdot P_\vartheta(T = k) \\
&= \sum_{k=0}^{n} h(B,k) \cdot P_\vartheta(T = k) \\
&= \int_{\{0,1,\dots,n\}} h(B,k)\ (P_\vartheta * T)(dk),
\end{aligned}$$

wobei $h(B,k) = \frac{|B \cap \{T=k\}|}{\binom{n}{k}}$ unabhängig von ϑ ist.

Damit wird für allgemeine statistische Modelle eine Präzisierung des Begriffs „Suffizienz" nahe gelegt: $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ heißt suffizient für eine Verteilungsannahme $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$, falls für alle $B \in \mathcal{B}$ die bedingte Wahrscheinlichkeit (bzgl. P_ϑ) von $B \in \mathcal{B}$ bei gegebenen T unabhängig von $\vartheta \in \Theta$ ist.

3.2 Bedingte Erwartungswerte und bedingte Wahrscheinlichkeiten

Der folgende Satz ist etwa in Gänssler, P. und Stute, W. (1977), Wahrscheinlichkeitstheorie, Springer, Heidelberg, Satz 1.2.24, bewiesen.

Satz 3.2.1. *Es sei $\mathcal{X}$ eine beliebige nicht-leere Menge, $(\mathcal{Y}, \mathcal{G})$ ein messbarer Raum und $T : \mathcal{X} \to \mathcal{Y}$ eine beliebige Abbildung. Dann existiert zu jeder $T^{-1}(\mathcal{G}) = \{T^{-1}(G) : G \in \mathcal{G}\}$–messbaren Abbildung $f : \mathcal{X} \to \mathbb{R}$ eine $\mathcal{G}$-messbare Abbildung $g : \mathcal{Y} \to \mathbb{R}$, so dass $f = g \circ T$.*

Definition 3.2.2. Es sei $(\mathcal{X}, \mathcal{B}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{A}$ eine sub-σ-Algebra von $\mathcal{B}$, d.h. $\mathcal{A} \subset \mathcal{B}$, $\mathcal{A}$ ist σ-Algebra, $B \in \mathcal{B}$ und $X \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$. Dann heißt

$$\begin{aligned}
E_P(X|\mathcal{A}) &:= E(X|\mathcal{A}) \\
&:= \left\{ Z \in \mathcal{L}_1(\mathcal{X}, \mathcal{A}, P) : \forall A \in \mathcal{A} : \int_A X\, dP = \int_A Z\, dP \right\}
\end{aligned}$$

die *bedingte Erwartung* von X bei gegebenen $\mathcal{A}$ (bzgl. P) und

$$P(B|\mathcal{A}) := E_P(1_B|\mathcal{A})$$

die *bedingte Wahrscheinlichkeit* von B bei gegebenen $\mathcal{A}$.

Satz 3.2.3. *Unter den Voraussetzungen von 3.2.2 gilt*

(i) $E(X|\mathcal{A}) \neq \emptyset$,

(ii) $Z_1, Z_2 \in E(X|\mathcal{A}) \Rightarrow Z_1 = Z_2$ *P-f.s.*,

(iii) $Z_1 \in E(X|\mathcal{A})$, $Z_2 : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}, \mathbb{B})$ *mit* $Z_1 = Z_2$ *P-f.ü.* $\Rightarrow Z_2 \in E(X|\mathcal{A})$

Beweis:

(i) Es sei zunächst $X \geq 0$. Dann wird durch

$$Q(A) := \int_A X \, dP, \quad A \in \mathcal{A},$$

ein endliches (wegen $0 \leq E_P(X) < \infty$) Maß auf $\mathcal{A}$ definiert mit $Q \ll P$. Also existiert nach dem Satz von Radon-Nikodym 2.1.3 ein $Z_0 \in \mathcal{L}_1(\mathcal{X}, \mathcal{A}, P)$ mit $Q(A) = \int_A Z_0 \, dP$, $A \in \mathcal{A}$, d.h. $Z_0 \in E(X|\mathcal{A})$. Für ein beliebiges $X \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$ existieren $Z_1 \in E(X^+|\mathcal{A})$ und $Z_2 \in E(X^-|\mathcal{A})$, wobei $X^+ = \max(0, X)$, $X^- = \max(0, -X)$, $X = X^+ - X^-$. Dann ist $Z := Z_1 - Z_2 \in E(X|\mathcal{A})$.

(ii) $\forall A \in \mathcal{A}$:

$$\int_A Z_1 \, dP = \int_A X \, dP = \int_A Z_2 \, dP.$$

Mit $A := \{Z_1 > Z_2\} \in \mathcal{A}$ bzw. $A' := \{Z_1 < Z_2\} \in \mathcal{A}$ folgt die Behauptung.

(iii) Folgt unmittelbar aus obiger Gleichung.

□

Lemma 3.2.4. *Falls*

(i) $\mathcal{A} = \{\mathcal{X}, \emptyset\} \Rightarrow E(X|\mathcal{A}) = \{E(X)\}$

(ii) $\mathcal{A} = \mathcal{B} \Rightarrow X \in E(X|\mathcal{A})$.

Beweis:

(i) Eine Abbildung $f : \mathcal{X} \to \mathbb{R}$ ist messbar bzgl. $\{\emptyset, \mathcal{X}\} \Leftrightarrow$ f ist konstant; also $f \in E(X|\mathcal{A}) \Rightarrow f = \text{const}$.
$\Rightarrow E(f) = \text{ const } = \int_{\mathcal{X}} f \, dP = \int_{\mathcal{X}} X \, dP = E(X)$.

(ii) $\mathcal{A} = \mathcal{B} \Rightarrow X$ ist $\mathcal{A}$-messbar.
$\Rightarrow X \in E(X|\mathcal{A})$

$\square$

Konvention 3.2.5. Es ist üblich jedes $Z \in E(X|\mathcal{A})$ ebenfalls als bedingte Erwartung von X bei gegebenem $\mathcal{A}$ zu bezeichnen und hierfür ebenfalls das Symbol $E(X|\mathcal{A})$ (dann aufgefasst als Zufallsvariable) zu verwenden. Dasselbe gilt für $P(B|\mathcal{A})$.

Satz 3.2.6. *Es sei $(\mathcal{X}, \mathcal{B}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{A}$ eine sub-σ-Algebra von $\mathcal{B}$. Sei $X, X_n \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$, $n \in \mathbb{N}$, sowie $a, b \in \mathbb{R}$. Dann gilt:*

(i) $\forall A \in \mathcal{A} \quad \int_A E(X|\mathcal{A})\, dP = \int_A X\, dP$, *speziell* $E(X) = E(E(X|\mathcal{A}))$

(ii) $X = X_1$ *P-f.s.* $\Rightarrow E(X|\mathcal{A}) = E(X_1|\mathcal{A})$ *P-f.s.*

(iii) $X = a$ *P-f.s.* $\Rightarrow E(X|\mathcal{A}) = a$ *P-f.s.*

(iv) $E(aX_1 + bX_2|\mathcal{A}) = a \cdot E(X_1|\mathcal{A}) + b \cdot E(X_2|\mathcal{A})$ *P-f.s.*

(v) $X \geq 0$ *P-f.s.* $\Rightarrow E(X|\mathcal{A}) \geq 0$ *P-f.s.*

(vi) $X_1 \leq X_2$ *P-f.s.* $\Rightarrow E(X_1|\mathcal{A}) \leq E(X_2|\mathcal{A})$ *P-f.s.*

(vii) $X_n \uparrow_{n\in\mathbb{N}} X$ *P-f.s.* $\Rightarrow \sup_{n\in\mathbb{N}} E(X_n|\mathcal{A}) = E(X|\mathcal{A})$ *P-f.s.*

(viii) X *$\mathcal{A}$-messbar* $\Rightarrow X = E(X|\mathcal{A})$ *P-f.s.*

BEWEIS:

(i) Definition von $E(X|\mathcal{A})$

(ii) Setze $B := \{E(X_1|\mathcal{A}) > E(X|\mathcal{A})\} \in \mathcal{A}$

$$\begin{aligned}
&\Rightarrow \quad \int_B E(X|\mathcal{A})\, dP = \int_B X\, dP = \int_B X_1\, dP = \int_B E(X_1|\mathcal{A})\, dP \\
&\Rightarrow \quad \int_B \underbrace{E(X_1|\mathcal{A}) - E(X|\mathcal{A})}_{>0 \text{ auf B}}\, dP = 0 \\
&\Rightarrow \quad P(B) = 0.
\end{aligned}$$

Vertauschen von X_1 und X liefert die Behauptung.

(iii) folgt aus (ii)

(iv) folgt aus der Linearität des Integrals

(v) Setze $B := \{E(X|\mathcal{A}) < 0\} \in \mathcal{A}$

$$\int_B \underbrace{E(X|\mathcal{A})}_{<0 \text{ auf B}} \, dP = \int_B X \, dP \geq 0$$

$\Rightarrow P(B) = 0.$

(vi) folgt aus (iv) und (v): $0 \leq E(\underbrace{X_2 - X_1}_{\geq 0 \text{ P-f.s.}}|\mathcal{A}) = E(X_2|\mathcal{A}) - E(X_1|\mathcal{A})$ P-f.s.

(vii) Aus (vi) folgt $\sup_{n\in\mathbb{N}} E(X_n|\mathcal{A}) = \lim_{n\in\mathbb{N}} E(X_n|\mathcal{A})$ P-f.s.

$$\begin{aligned} \Rightarrow \forall B \in \mathcal{A} : \int_B \lim_{n\in\mathbb{N}} E(X_n|\mathcal{A}) \, dP &\overset{\textit{mon. Konvergenz}}{=} \lim_{n\in\mathbb{N}} \int_B E(X_n|\mathcal{A}) \, dP \\ &= \lim_{n\in\mathbb{N}} \int_B X_n \, dP \\ &\overset{\textit{mon. Konvergenz}}{=} \int_B \lim_{n\in\mathbb{N}} X_n \, dP \\ &= \int_B X \, dP, \text{ P-f.s.} \end{aligned}$$

(viii) X $\mathcal{A}$-messbar $\Rightarrow X \in E(X|\mathcal{A})$; 3.2.3 (ii) $\Rightarrow X = E(X|\mathcal{A})$ P-f.s.

□

Satz 3.2.7. *Es sei $(\mathcal{X}, \mathcal{B}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{A}$ eine sub-σ-Algebra von $\mathcal{B}$ und $X \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$.*
Falls die σ–Algebren $X^{-1}(\mathbb{B}) := \{X^{-1}(B) : B \in \mathbb{B}\}$ und $\mathcal{A}$ unabhängig sind bzgl. P, d.h. $P(B \cap A) = P(B) \cdot P(A)$, $B \in X^{-1}(\mathbb{B})$, $A \in \mathcal{A}$, so gilt $E(X|\mathcal{A}) = E(X)$ P-f.s.

Beweis: Nach Voraussetzung sind für jedes $A \in \mathcal{A}$ die Zufallsvariablen 1_A und X unabhängig bzgl. P, denn $1_A^{-1}(\mathbb{B}) = \{\emptyset, \mathcal{X}, A, A^c\} \subset \mathcal{A}$. Also gilt für $A \in \mathcal{A}$

$$\int_A X \, dP = E(1_A \cdot X) = E(1_A)E(X) = \int_A E(X) \, dP.$$

Da $E(X)$ als konstante Funktion $\mathcal{A}$-messbar ist, folgt die Behauptung aus Lemma 3.2.3 (ii).

□

Satz 3.2.8. *Es sei $(\mathcal{X}, \mathcal{B}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{A}_1, \mathcal{A}_2$ sub-σ-Algebren von $\mathcal{B}$ mit $\mathcal{A}_1 \subset \mathcal{A}_2$. Dann gilt:*

$$E(E(X|\mathcal{A}_1)|\mathcal{A}_2) \overset{P\text{-f.s.}}{=} E(X|\mathcal{A}_1) \overset{P\text{-f.s.}}{=} E(E(X|\mathcal{A}_2)|\mathcal{A}_1)$$

Beweis: $E(X|\mathcal{A}_1)$ ist $\mathcal{A}_1$-messbar und damit auch $\mathcal{A}_2$-messbar; 3.2.6 (viii) $\Rightarrow E(E(X|\mathcal{A}_1)|\mathcal{A}_2) = E(X|\mathcal{A}_1)$ P-f.s.
Es sei nun $A \in \mathcal{A}_1 \subset \mathcal{A}_2$, dann:

$$\int_A X\ dP \overset{Def.}{=} \int_A E(X|\mathcal{A}_2)\ dP \overset{Def.}{=} \int_A E(E(X|\mathcal{A}_2)|\mathcal{A}_1)\ dP$$

Da dies für alle $A \in \mathcal{A}_1$ gilt, folgt aus der Gleichheit von erstem und letztem Term

$$E(X|\mathcal{A}_1) = E(E(X|\mathcal{A}_2)|\mathcal{A}_1) \text{ P-f.s.}$$

□

Satz 3.2.9. *Es sei $(\mathcal{X}, \mathcal{B}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{A}$ eine sub-σ-Algebra von $\mathcal{B}$ und $X \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$, $Y : \mathcal{X} \to \mathbb{R}$ $\mathcal{A}$-messbar mit der Eigenschaft $XY \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$. Dann gilt*

(i) $\forall A \in \mathcal{A}.\quad \int_A XY\ dP = \int_A Y\ E(X|\mathcal{A})\ dP$

(ii) $E(XY|\mathcal{A}) = Y\ E(X|\mathcal{A})$ *P-f.s.*

Beweis: O.E. sei $X \geq 0$, sonst betrachte man $X = X^+ - X^-$. Wir knüpfen an den Beweis von 3.2.3 (i) an und setzen $Q(A) := \int_A X\ dP$, $A \in \mathcal{A}$. Dann ist Q ein endliches Maß auf $\mathcal{A}$ mit $Q \ll P$ und besitzt eine Dichte $Z_0 = E(X|\mathcal{A})$ bzgl. P. Wir erhalten somit für ein beliebiges $A \in \mathcal{A}$

$$\int_A XY\ dP = \int_A Y\ dQ = \int_A Y Z_0\ dP/\mathcal{A} = \int_A Y\ E(X|\mathcal{A})\ dP.$$

(ii) folgt aus (i) und der Definition von $E(XY|\mathcal{A})$, da $Y\ E(X|Y)$ $\mathcal{A}$-messbar ist.

□

Definition 3.2.10. Es sei $(\mathcal{X}, \mathcal{B}, P)$ ein Wahrscheinlichkeitsraum, $(\mathcal{Y}, \mathcal{G})$ ein messbarer Raum, $X \in \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$, $B \in \mathcal{B}$ und $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$. Dann heißt

(i) $E(X|T) := E(X|\underbrace{T^{-1}(\mathcal{G})}_{:=\{T^{-1}(G):G\in\mathcal{G}\}\subset\mathcal{B}})$ die bedingte Erwartung von X bei gegebenem T,

(ii) $P(B|T) := P(B|T^{-1}(\mathcal{G})) = E(1_B|T^{-1}(\mathcal{G}))$ die bedingte Wahrscheinlichkeit von B bei gegebenem T.

Satz 3.2.1 impliziert die folgende Aussage:

Satz 3.2.11. *Unter den Voraussetzungen von Definition 3.2.10 existiert zu jeden $E(X|T)$ eine $\mathcal{G}$-messbare Funktion $g : \mathcal{Y} \to \mathbb{R}$ mit*

$$E(X|T) = g \circ T$$

Satz 3.2.12. *Unter den Voraussetzungen von Definition 3.2.10 sind für eine $\mathcal{G}$-messbare Funktion $g : \mathcal{Y} \to \mathbb{R}$ die beiden folgenden Aussagen äquivalent:*

(i) $g \circ T = E(X|T)$ *P-f.s.*

(ii) $g \in \mathcal{L}_1(\mathcal{Y}, \mathcal{G}, P * T)$ *und*

$$\int_G g \; d(P * T) = \int_{T^{-1}(G)} X \; dP, \quad G \in \mathcal{G}.$$

Beweis: (i) $\Rightarrow$ (ii): Für $G \in \mathcal{G}$ gilt:

$$\begin{aligned}
\int_{T^{-1}(G)} X \; dP &\overset{3.2.6(i)}{=} \int_{T^{-1}(G)} E(X|T) \; dP \\
&\overset{(i)}{=} \int_{T^{-1}(G)} g \circ T \; dP \\
&= \int (1_{T^{-1}(G)})(g \circ T) \; dP \\
&= \int (1_G \circ T)(g \circ T) \; dP \\
&= \int (1_G \cdot g) \circ T \; dP \\
&= \int 1_G \cdot g \; d(P * T) \\
&= \int_G g \; d(P * T)
\end{aligned}$$

(ii) $\Rightarrow$ (i): $\forall G \in \mathcal{G}$:

$$\begin{aligned}
\int_{T^{-1}(G)} g \circ T \; dP &= \int (1_G \cdot g) \circ T \; dP \\
&= \int (1_G \cdot g) \; d(P * T) \\
&= \int_G g \; d(P * T) \\
&\overset{(ii)}{=} \int_{T^{-1}(G)} X \; dP;
\end{aligned}$$

da $g \circ T$ $T^{-1}(\mathcal{G})$-messbar ist, folgt $g \circ T = E(X|T)$ P-f.s. □

Definition 3.2.13. Es seien die Voraussetzungen von 3.2.10 erfüllt. Dann heißt

$$\begin{aligned}
E_P(X|T=\cdot) &:= E(Y|T=\cdot) \\
&:= \{G \in \mathcal{L}_1(\mathcal{Y},\mathcal{G},P*T): \\
&\qquad \forall G \in \mathcal{G}: \int_G g\ d(P*T) = \int_{T^{-1}(G)} X\ dP\Big\} \\
&= \{g:(\mathcal{Y},\mathcal{G}) \to (\mathbb{R},\mathbb{B}): g \circ T = E(X|T) \text{ P-f.s.}\}
\end{aligned}$$

die bedingte Erwartung von X unter der Hypothese $T = \cdot$ bzw. *Faktorisierung der bedingten Erwartung von X bzgl. T*:

$$E(X|T) = E(X|T=\cdot) \circ T$$

Satz 3.2.14. *Unter den Voraussetzungen von 3.2.10 gilt*

(i) $E(X|T=\cdot) \neq \emptyset$,

(ii) $g_1, g_2 \in E(X|T=\cdot) \Rightarrow g_1 = g_2$ $(P*T)$*-f.ü.*

(iii) $g_1 \in E(X|T=\cdot)$ *und* $g_2 : (\mathcal{Y},\mathcal{G}) \to (\mathbb{R},\mathbb{B})$ *mit* $g_1 = g_2$ $(P*T)$*-f.ü.* $\Rightarrow g_2 \in E(X|T=\cdot)$

Beweis: Teil (i) folgt aus 3.2.3 (i) und 3.2.11.
Teil (ii): $\forall G \in \mathcal{G}$:

$$\int_G g_1\ d(P*T) = \int_{T^{-1}(G)} X\ dP = \int_G g_2\ d(P*T).$$

Mit $G := \{g_1 > g_2\} \in \mathcal{G}$ bzw. $G' := \{g_1 < g_2\} \in \mathcal{G}$ folgt die Behauptung.
Teil (iii) folgt unmittelbar aus obiger Gleichung. □

Hinsichtlich der Verwendung des Symbols $E(X|T=\cdot)$ gilt das Entsprechende wie bei der Konvention 3.2.5.

Für $E(X|T=\cdot)$ gelten die zu 3.2.6 analogen Eigenschaften.

Falls $T(x) = y$, so schreibt man

$$\begin{aligned}
E(X|T)(x) &= (E(X|T=\cdot) \circ T)(x) \\
&= E(X|T=\cdot)(\underbrace{T(x)}_{=y}) \\
&= E(X|T(x)=y) \\
&= E(X|T=y).
\end{aligned}$$

3.3 Suffiziente σ-Algebren und suffiziente Statistiken

Definition 3.3.1. Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$. Eine sub-$\sigma$-Algebra $\mathcal{A}$ von $\mathcal{B}$ heißt *suffizient für* $\mathcal{P}/\mathcal{B}$, falls

$$\forall B \in \mathcal{B} : \bigcap_{P \in \mathcal{P}} P(B|\mathcal{A}) \neq \emptyset,$$

d.h. falls für jedes $B \in \mathcal{B}$ eine von $P \in \mathcal{P}$ unabhängige Version der bedingten Wahrscheinlichkeit von B bei gegebenem $\mathcal{A}$ existiert.
Eine Abbildung $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ heißt *suffizient für* $\mathcal{P}/\mathcal{B}$, falls $\mathcal{A} := T^{-1}(\mathcal{G})$ suffizient für $\mathcal{P}/\mathcal{B}$ ist.

Es sei $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ suffizient für $\mathcal{P}/\mathcal{B} \Rightarrow \forall P \in \mathcal{P} \quad \forall B \in \mathcal{B}$:

$$\begin{aligned}
P(B) &= \int 1_B \, dP \\
&\overset{3.2.6(i)}{=} \int \underbrace{\underbrace{E_P(1_B|T)}_{= P(B|T)}}_{\text{unabhängig von P}} dP \\
&\overset{3.2.1}{=:} g_B \circ T \\
&= \int g_B \circ T \, dP \\
&= \int g_B \, d(P * T)
\end{aligned}$$

Bemerkung 3.3.2. Es sei $\mathcal{P}/\mathcal{B} = \mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$. Dann gilt:

(i) $\mathcal{B}$ ist suffizient für $\mathcal{P}/\mathcal{B}$

(ii) Jede $\mathcal{P}/\mathcal{B}$ suffiziente sub-σ-Algebra $\mathcal{A}$ von $\mathcal{B}$ ist suffizient für jede Teilfamilie $\mathcal{P}_0$ von $\mathcal{P}$.

(iii) Es sei $(\mathcal{Y}, \mathcal{G})$ ein messbarer Raum und $T : \mathcal{X} \to \mathcal{Y}$ eine für $\mathcal{P}/\mathcal{B}$ suffiziente Statistik

$$\Rightarrow \quad \begin{aligned} &\forall B \in \mathcal{B} \; \exists g_B : (\mathcal{Y}, \mathcal{G}) \to (\mathbb{R}, \mathbb{B}) \\ &\forall G \in \mathcal{G} \; \forall P \in \mathcal{P} : P(B \cap T^{-1}(G)) = \int_G g_B \, d(P * T) \end{aligned}$$

(iv) Es sei $\mathcal{Y}$ eine beliebige Menge $T : \mathcal{X} \to \mathcal{Y}$ eine beliebige Abbildung $\Rightarrow$ $\mathcal{G}_T := \{G \subset \mathcal{Y} : T^{-1}(G) \in \mathcal{B}\}$ ist eine σ-Algebra auf $\mathcal{Y}$ und T ist $\mathcal{B}, \mathcal{G}_T$-messbar. Man nennt T suffizient für $\mathcal{P}/\mathcal{B}$, wenn $T^{-1}(\mathcal{G}_T)$ suffizient für $\mathcal{P}/\mathcal{B}$ ist.

(v) Ist $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine parametrische Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$, so nennt man $\mathcal{A}$ bzw. T suffizient für $\vartheta \in \Theta$, wenn $\mathcal{A}$ bzw. T suffizient für $\mathcal{P}/\mathcal{B}$ ist.

Beweis:

(i) folgt aus $1_B \in \bigcap_{P \in \mathcal{P}} \underbrace{P(B|\mathcal{B})}_{E_P(1_B|\mathcal{B})}, \quad B \in \mathcal{B}$

(ii) ist trivial

(iii) Für beliebiges $B \in \mathcal{B}$ sei $h_B \in \bigcap_{P \in \mathcal{P}} P(B|T^{-1}(\mathcal{G}))$ mit 3.2.1 folgt

$$\exists g_B : (\mathcal{Y}, \mathcal{G}) \to (\mathbb{R}, \mathbb{B}) : h_B = g_B \circ T$$

$\Rightarrow \forall G \in \mathcal{G}\ \forall P \in \mathcal{P}$:

$$\begin{aligned}
P(B \cap T^{-1}(G)) &= \int 1_{B \cap T^{-1}(G)}\, dP \\
&= \int 1_B \cdot 1_{T^{-1}(G)}\, dP \\
&= \int_{T^{-1}(G)} 1_B\, dP \\
&= \int_{T^{-1}(G)} h_B\, dP \\
&= \int_{T^{-1}(G)} g_B \circ T\, dP \\
&= \int (g_B \circ T) \cdot 1_{T^{-1}(G)}\, dP \\
&= \int (g_B \circ T) \cdot (1_G \circ T)\, dP \\
&= \int g_B \cdot 1_G\, d(P * T) \\
&= \int_G g_B\, d(P * T).
\end{aligned}$$

□

Beispiel 3.3.3. Es sei G eine Gruppe endlicher Ordnung von bijektiven $\mathcal{B},\mathcal{B}$-messbaren Abbildungen, und

$$g : \mathcal{X} \to \mathcal{X},\ \mathcal{A} := \{B \in \mathcal{B} : \forall g \in G : g(B) = B\}$$

die σ-Algebra der G-invarianten Mengen in $\mathcal{B}$ und $\mathcal{P} = \mathcal{P}/\mathcal{B}$ die Familie aller G-invarianten Wahrscheinlichkeitsmaße P auf $(\mathcal{X}, \mathcal{B})$, d.h. $P * g = P$, $g \in G$. Dann gilt

$$h_B := \frac{1}{|G|} \sum_{g \in G} (1_B \circ g) \in P(B|\mathcal{A}),\ B \in \mathcal{B},\ P \in \mathcal{P}, \tag{3.3.4}$$

d.h. $\mathcal{A}$ ist suffizient für $\mathcal{P}/\mathcal{B}$.

Beweis: Da für ein beliebiges aber festes $g' \in G$ mit g auch $g \circ g'$ die ganze Gruppe G durchläuft, gilt

$$\forall g \in G : h_B \circ g^{-1} = h_B$$

und damit folgt

$$\begin{aligned} g(h_B^{-1}(A)) &= (h_B \circ g^{-1})^{-1}(A) \\ &= h_B^{-1}(A) \quad , A \in \mathcal{A}, B \in \mathcal{B}. \end{aligned}$$

d.h. h_B ist $\mathcal{A}$-messbar.
Für $B \in \mathcal{B}$ gilt

$$P(B \cap A) = \int_A h_B \, dP, \quad A \in \mathcal{A}, P \in \mathcal{P}.$$

Denn aus $1_A = 1_A \circ g$, $A \in \mathcal{A}$, $g \in G$ folgt:

$$\begin{aligned} |G| \cdot \int_A h_B \, dP &= \sum_{g \in G} \int_A 1_B \circ g \, dP \\ &= \sum_{g \in G} \int_{\mathcal{X}} (1_B \circ g)(1_A \circ g) \, dP \\ &= \sum_{g \in G} \int_{\mathcal{X}} 1_B \cdot 1_A \, d(P * g) \\ &= \sum_{g \in G} (P * g)(B \cap A) \\ &= \sum_{g \in G} P(B \cap A) \\ &= |G| \cdot P(B \cap A) \end{aligned}$$

□

Zwei Spezialfälle:

(i) Es sei $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}^n, \mathbb{B}^n)$ und G die Gruppe der Ordnung $n!$ derjenigen Abbildungen $g : \mathbb{R}^n \to \mathbb{R}^n$, die den Permutationen der n Koordinaten entsprechen. Dann ist $\mathcal{A}$ die σ-Algebra derjenigen Borelmengen des $\mathbb{R}^n$, die mit $x \in \mathbb{R}^n$ auch alle Punkte enthalten, die aus x durch Permutation der Koordinaten hervorgeht. Es sei $\mathcal{P}$ die Gesamtheit aller Wahrscheinlichkeitsmaße auf $(\mathbb{R}^n, \mathbb{B}^n)$ und $\mathcal{P}_0 := \{P^n : P \in \mathcal{P}\}$. Dann besteht $\mathcal{P}_0$ aus G-invarianten Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^n, \mathbb{B}^n)$ [denn $(P^n * g)(B_1 \times \cdots \times B_n) = \prod_{i=1}^n P(B_i) = P^n(B_1 \times \cdots \times B_n)$, $B_i \in \mathbb{B}$, $i = 1, \ldots, n$; $g \in G \Rightarrow P^n * g = P$]. Somit ist $\mathcal{A}$ nach 3.3.3 suffizient für $\mathcal{P}_0$.

(ii) Es sei $(\mathcal{X}, \mathcal{B}) = (\mathbb{R}, \mathbb{B})$ und $G := \{\mathrm{id}_\mathbb{R}, -\mathrm{id}_\mathbb{R}\}$ die Spieglungsgruppe (am Nullpunkt). Dann ist $\mathcal{A}$ die Gesamtheit aller um 0 symmetrischen Borel-Mengen von $\mathbb{R}$, d.h. $\mathcal{A} = \{B \in \mathbb{B} : B = -B = \{-x, x \in \mathbb{B}\}\}$. $\mathcal{A}$ ist suffizient für die Familie aller um 0 symmetrischen Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathbb{B})$, d.h. $P(B) = P(-B)$, $B \in \mathbb{B}$.

Da $\mathcal{A} = T^{-1}(\mathbb{B})$ mit $T(x) := |x|$, $x \in \mathbb{R}$, ist also T suffizent für die Familie aller um 0 symmetrischen Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathbb{B})$.

Lemma 3.3.5. *Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$. Eine sub-$\sigma$-Algebra $\mathcal{A}$ von $\mathcal{B}$ ist genau dann suffizient für $\mathcal{P}/\mathcal{B}$, falls*

$$\forall f \in \bigcap_{P \in \mathcal{P}} \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P) : \bigcap_{P \in \mathcal{P}} E_P(f|\mathcal{A}) \neq \emptyset$$

Beweis: „$\Rightarrow$“: Für $f = 1_B$, $B \in \mathcal{B}$ ist $\bigcap_{P \in \mathcal{P}} E_P(1_B|\mathcal{A}) = \bigcap_{P \in \mathcal{P}} P(B|\mathcal{A}) \neq \emptyset$. Für $f \in \bigcap_{P \in \mathcal{P}} \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$ folgt die Behauptung durch algebraische Induktion (d.h. die Behauptung gilt für $f = 1_B \Rightarrow f = \sum_{i=1}^n \alpha_i 1_{B_i} =: e$ einfache Funktion $\Rightarrow f = \lim_{n \to \infty} e_n$ mit $e_1 \leq e_2 \leq \ldots \Rightarrow f = f^+ - f^-$) unter Verwendung von 3.2.6
„$\Leftarrow$“: Offensichtlich

$\square$

Satz 3.3.6. *Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$. Sind dann $\mathcal{A}_0$, $\mathcal{A}_1$ sub-σ-Algebren auf $\mathcal{B}$ mit $\mathcal{A}_0 \subset \mathcal{A}_1 (\subset \mathcal{B})$, so gilt:*

(i) Falls $\mathcal{A}_0$ suffizient ist für $\mathcal{P}/\mathcal{B} \Rightarrow \mathcal{A}_0$ ist suffizient für $\mathcal{P}/\mathcal{A}_1$

(ii) Falls $\mathcal{A}_0$ suffizient ist für $\mathcal{P}/\mathcal{A}_1$ und $\mathcal{A}_1$ ist suffizient für $\mathcal{P}/\mathcal{B} \Rightarrow \mathcal{A}_0$ ist suffizient für $\mathcal{P}/\mathcal{B}$.

Beweis:

(i) ist offensichtlich

(ii) Sei $B \in \mathcal{B}$ und $h_B^1 \in \bigcap_{P \in \mathcal{P}} P(B|\mathcal{A}_1)$. Da $h_B^1 \in \bigcap_{P \in \mathcal{P}} \mathcal{L}_1(\mathcal{X}, \mathcal{A}_1, P)$ und da $\mathcal{A}_0$ suffizient ist für $\mathcal{P}/\mathcal{A}_1$ folgt nach 3.3.5 die Existenz von $h_B \in \bigcap_{P \in \mathcal{P}} E_P(h_B^1|\mathcal{A}_0)$. Nach 3.2.8 gilt für alle $P \in \mathcal{P}$

$$E_P(h_B^1|\mathcal{A}_0) = E_P(E_P(1_B|\mathcal{A}_1)|\mathcal{A}_0) = E_P(1_B|\mathcal{A}_0) = P(B|\mathcal{A}_0) \text{ P-f.s.}$$

also $h_B \in \bigcap_{P \in \mathcal{P}} P(B|\mathcal{A}_0)$.

□

Definition 3.3.7. Für eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ sei

$$\mathcal{P}^L := \left\{ \sum_{k \in \mathbb{N}} c_k P_k : c_k \geq 0, P_k \in \mathcal{P}, k \in \mathbb{N}, \sum_{k \in \mathbb{N}} c_k = 1 \right\}.$$

Beachte dass $\mathcal{P} \subset \mathcal{P}^L$ und dass die Elemente von $\mathcal{P}^L$ Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{B})$ sind.

Satz 3.3.8. *Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ und $\mathcal{A}$ eine sub-σ-Algebra von $\mathcal{B}$. Dann gilt:*

(i) $\forall f \in \bigcap_{P \in \mathcal{P}} \mathcal{L}_1(\mathcal{X}, \mathcal{B}, P)$, $f \geq 0$: $\bigcap_{P \in \mathcal{P}} E_P(f|\mathcal{A}) = \bigcap_{P \in \mathcal{P}^L} E_P(f|\mathcal{A})$

(ii) Ist $\mathcal{A}$ suffizient für $\mathcal{P}/\mathcal{B}$, so auch für $\mathcal{P}^L/\mathcal{B}$.

BEWEIS:

(i) Wegen $\mathcal{P} \subset \mathcal{P}^L$ gilt „$\supset$“. Ist andererseits $f_0 \in \bigcap_{P \in \mathcal{P}} E_P(f|\mathcal{A})$, so folgt $\forall A \in \mathcal{A}$ und $\forall P \in \mathcal{P}^L$:

$$\begin{aligned} \int_A f_0 \, dP &= \int_A f_0 \, d\left(\sum_{k \in \mathbb{N}} c_k P_k\right) \\ &= \sum_{k \in \mathbb{N}} c_k \cdot \int_A f_0 \, dP_k \\ &= \sum_{k \in \mathbb{N}} c_k \cdot \int_A f \, dP_k \\ &= \int_A f \, dP, \end{aligned}$$

d.h. $f_0 \in \bigcap_{P \in \mathcal{P}^L} E_P(f|\mathcal{A})$.

(ii) ist eine unmittelbare Folgerung aus (i).

$\square$

Lemma 3.3.9. *Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ und $\mathcal{A}$ eine sub-σ-Algebra von $\mathcal{B}$. Dann gilt:*

(i) Ist $\mathcal{A}$ suffizient für $\mathcal{P}/\mathcal{B}$, so gilt für jedes $P \in \mathcal{P}$ und jedes $\hat{P} \in \mathcal{P}^L$ mit $\mathcal{P}/\mathcal{B} \ll \hat{P}/\mathcal{B}$:

$$\frac{dP/\mathcal{A}}{d\hat{P}/\mathcal{A}} \subset \frac{dP/\mathcal{B}}{d\hat{P}/\mathcal{B}}$$

d.h. $\forall P \in \mathcal{P}$ $\exists\mathcal{A}$-meßbare Dichte von $P/\mathcal{B}$ bzgl. $\hat{P}/\mathcal{B}$.

(ii) Falls ein $\hat{P} \in \mathcal{P}^L$ mit $\mathcal{P}/\mathcal{B} \ll \hat{P}/\mathcal{B}$ existiert und falls $\forall P \in \mathcal{P}$ gilt

$$\frac{dP/\mathcal{A}}{d\hat{P}/\mathcal{A}} \subset \frac{dP/\mathcal{B}}{d\hat{P}/\mathcal{B}}$$

so gilt $\forall B \in \mathcal{B}$

$$\hat{P}(B|\mathcal{A}) \subset \bigcap_{P \in \mathcal{P}} P(B|\mathcal{A}).$$

Insbesondere ist dann $\mathcal{A}$ suffizient für $\mathcal{P}/\mathcal{B}$.

Beweis:

(i) $\mathcal{A}$ ist nach Voraussetzung suffizient für $\mathcal{P}/\mathcal{B} \Rightarrow \forall B \in \mathcal{B}\ \exists\ h_B \in \bigcap_{P\in\mathcal{P}} P(B|\mathcal{A})$; Nach 3.3.8 (i) folgt $h_B \in \bigcap_{P\in\mathcal{P}^L} P(B|\mathcal{A})$.
Es seien nun $P \in \mathcal{P}$ und $\hat{P} \in \mathcal{P}^L$ mit $\mathcal{P}/\mathcal{B} \ll \hat{P}/\mathcal{B}$ beliebig vorgegeben und es sei $\hat{f} \in \frac{dP/\mathcal{A}}{d\hat{P}/\mathcal{A}}$. Dann gilt $\forall B \in \mathcal{B}$:

$$\begin{aligned}
P(B) \quad &= \quad \int_{\mathcal{X}} h_B \, dP \overset{h_B \text{ ist } \mathcal{A} \text{ messbar}}{=} \int_{\mathcal{X}} h_B \, dP/\mathcal{A} \\
&= \quad \int_{\mathcal{X}} h_B \cdot \hat{f} \, d\hat{P}/\mathcal{A} = \int_{\mathcal{X}} h_B \cdot \hat{f} \, d\hat{P} \\
&\overset{3.2.9(i)}{=} \int_{\mathcal{X}} 1_B \cdot \hat{f} \, d\hat{P} = \int_B \hat{f} \, d\hat{P}
\end{aligned}$$

d.h. $\hat{f} \in \frac{dP/\mathcal{B}}{d\hat{P}/\mathcal{B}}$

(ii) Es sei $B \in \mathcal{B}$ und $P \in \mathcal{P}$. Ferner sei $p \in \frac{dP/\mathcal{B}}{d\hat{P}/\mathcal{B}}$ $\mathcal{A}$-meßbar und $f_B \in$

$\hat{P}(B|\mathcal{A})$. Dann gilt $\forall A \in \mathcal{A}$:

$$\begin{aligned}
\int_A f_B \, dP &= \int_A f_B \, p \, d\hat{P} \\
&\overset{3.2.9(i)}{=} \int_A 1_B \, p \, d\hat{P} \\
&= \int_{A\cap B} p \, d\hat{P} \\
&= P(A \cap B) \\
&= \int_A 1_B \, dP
\end{aligned}$$

d.h. $f_B \in P(B|\mathcal{A})$. Da $B \in \mathcal{B}$ und $P \in \mathcal{P}$ beliebig waren, folgt die Behauptung.

□

Satz 3.3.10 (Faktorisierungssatz von Halmos-Savage, Neyman-Kriterium). *Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B} = \{P_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$, welche durch ein σ-endliches Maß $\mu/\mathcal{B}$ dominiert wird. Dann gilt:*

(i) Eine sub-σ-Algebra $\mathcal{A}$ von $\mathcal{B}$ ist suffizient für $\mathcal{P}/\mathcal{B}$:$\Leftrightarrow$

$$\begin{aligned}
&\exists h : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}^+, \mathbb{B} \cap \mathbb{R}^+) \text{ und } \forall \vartheta \in \Theta \\
&\exists f_\vartheta : (\mathcal{X}, \mathcal{A}) \to (\mathbb{R}^+, \mathbb{B} \cap \mathbb{R}^+) : f_\vartheta \cdot h \in \frac{dP_\vartheta/\mathcal{B}}{d\mu/\mathcal{B}}
\end{aligned} \tag{3.3.11}$$

(ii) Eine Statistik $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ ist suffizient für $\mathcal{P}/\mathcal{B}$:$\Leftrightarrow$

$$\begin{aligned}
&\exists h : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}^+, \mathbb{B} \cap \mathbb{R}^+) \text{ und } \forall \vartheta \in \Theta \\
&\exists g_\vartheta : (\mathcal{Y}, \mathcal{G}) \to (\mathbb{R}^+, \mathbb{B} \cap \mathbb{R}^+) : (g_\vartheta \circ T) \cdot h \in \frac{dP_\vartheta/\mathcal{B}}{d\mu/\mathcal{B}}
\end{aligned} \tag{3.3.12}$$

Beweis:

(i) „$\Rightarrow$“ $\mathcal{A}$ sei suffizient für $\mathcal{P}/\mathcal{B}$. Mit 2.1.13 auf Seite 31 folgt: $\Rightarrow \exists \hat{P} \in \mathcal{P}^L$ mit $\mathcal{P}/\mathcal{B} \ll \hat{P}/\mathcal{B}$. Da auch $\hat{P}/\mathcal{B}$ $\left(= \sum_{k\in\mathbb{N}} 2^{-k} P_k\right) \ll \mu/\mathcal{B}$, existiert also ein $0 \leq h \in \frac{d\hat{P}/\mathcal{B}}{d\mu/\mathcal{B}}$ (Satz von Radon-Nikodym).
3.3.9 (i) $\Rightarrow \forall \vartheta \in \Theta \; \exists f_\vartheta \in \frac{dP_\vartheta/\mathcal{B}}{d\hat{P}/\mathcal{B}}$, f_ϑ $\mathcal{A}$-messbar $\Rightarrow f_\vartheta \cdot h \in \frac{dP_\vartheta/\mathcal{B}}{d\mu/\mathcal{B}}$, denn:

$$P_\vartheta(B) = \int_B f_\vartheta \, d\hat{P} = \int_B f_\vartheta \cdot h \, d\mu.$$

„$\Leftarrow$" Gilt umgekehrt (3.3.11), so wird durch die Festlegung

$$\mu_0(B) := \int_B h \, d\mu$$

ein Maß $\mu_0/\mathcal{B}$ definiert und es gilt $\forall B \in \mathcal{B}; \forall \vartheta \in \Theta$:

$$P_\vartheta(B) = \int_B f_\vartheta \cdot h \, d\mu = \int_B f_\vartheta \, d\mu_0,$$

d.h. $f_\vartheta \in \frac{dP_\vartheta/\mathcal{B}}{d\mu_0/\mathcal{B}}$, $\vartheta \in \Theta$. Mit 2.1.13 folgt: $\exists \hat{P} \in \mathcal{P}^L$ mit $\mathcal{P}/\mathcal{B} \ll \hat{P}/\mathcal{B} \Rightarrow$ $\forall B \in \mathcal{B}:$

$$\begin{aligned} \hat{P}(B) &= \sum_{k\in\mathbb{N}} 2^{-k} P_{\vartheta_k}(B) \\ &= \sum_{k\in\mathbb{N}} 2^{-k} \int_B f_{\vartheta_k} \, d\mu_0 \\ &= \int_B \sum_{k\in\mathbb{N}} 2^{-k} f_{\vartheta_k} \, d\mu_0 \\ &= \int_B \hat{h} \, d\mu_0 \end{aligned}$$

wobei $\hat{h} := \sum_{k\in\mathbb{N}} 2^{-k} f_{\vartheta_k} \in \frac{d\hat{P}/\mathcal{B}}{d\mu_0/\mathcal{B}}$ und h ist $\mathcal{A}$-messbar, als Linearkombination $\mathcal{A}$-messbarer f_ϑ. Wegen

$$\hat{P}(\hat{h} = 0) = \int_{\{\hat{h}=0\}} \hat{h} \, d\mu_0 = 0$$

folgt $P_\vartheta(\hat{h} = 0) = 0, \ \vartheta \in \Theta$ (wegen $\mathcal{P} \ll \hat{P}$). Setzen wir nun

$$p_\vartheta := \frac{f_\vartheta}{\hat{h}} \cdot 1_{\{\hat{h}>0\}}, \ \vartheta \in \Theta,$$

so ist p_ϑ $\mathcal{A}$-messbar und $p_\vartheta \in \frac{dP_\vartheta/\mathcal{B}}{d\hat{P}/\mathcal{B}}$, denn es gilt $\forall B \in \mathcal{B}$:

$$\begin{aligned} P_\vartheta(B) &= P_\vartheta(B \cap \{\hat{h} > 0\}) + \underbrace{P_\vartheta(B \cap \{\hat{h} = 0\})}_{=0} \\ &= \int_B 1_{\{\hat{h}>0\}} \, f_\vartheta \, d\mu_0 \\ &= \int_B \hat{h} \cdot p_\vartheta \, d\mu_0 \\ &= \int_B p_\vartheta \, d\hat{P}. \end{aligned}$$

3.3.9 (ii) $\Rightarrow$ Behauptung.

(ii) folgt aus (i) mit $\mathcal{A} = T^{-1}(\mathcal{G})$ und Satz 3.2.1.

□

Beispiel 3.3.13. (i) Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B} = \{P_\vartheta : \vartheta \in \Theta\}$ eine k-parametrige Exponentialfamilie auf $(\mathcal{X}, \mathcal{B})$. (2.4.2) und Bemerkung 2.4.4 $\Rightarrow$ bzgl. eines geeignet gewählten σ-endlichen Maßes $\mu/\mathcal{B}$ existieren μ-Dichten von P_ϑ, $\vartheta \in \Theta$, der Form

$$c(\vartheta) \cdot \exp\left(\sum_{i=1}^{k} q_i(\vartheta) \cdot T_i(x)\right)$$

mit $\mathcal{B}$-meßbaren $T_i : \mathcal{X} \to \mathbb{R}$, $i = 1, \ldots, k$. Damit folgt gemäß (3.3.12) mit

$$g_\vartheta(y_1, \ldots, y_k) := c(\vartheta) \cdot \exp\left(\sum_{i=1}^{k} q_i(\vartheta) \cdot y_i\right), \ h \equiv 1,$$

dass die durch $T(x) := (T_1(x), \ldots, T_k(x))$ definierte Abbildung von $(\mathcal{X}, \mathcal{B}) \to (\mathbb{R}^k, \mathbb{B}^k)$ eine für $\mathcal{P}/\mathcal{B}$ suffiziente Statistik ist.

(ii) Es sei $\mathcal{P} = \mathcal{P}/\mathbb{B}^n = \{P_\vartheta : N(\mu, \sigma^2)^n : \vartheta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times (0, \infty)\}$. Aus 2.4.3, folgt, dass die durch

$$T(x_1, \ldots, x_n) := \left(\sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} x_i\right)$$

definierte Abbildung von $(\mathbb{R}^n, \mathbb{B}^n) \to (\mathbb{R}^2, \mathbb{B}^2)$ eine für $\mathcal{P}/\mathbb{B}^n$ suffiziente Statistik ist.

(iii) Es sei $\mathcal{P} = \mathcal{P}/\mathbb{P}(\{0,1\}^n) = \{P_\vartheta = B(1, \vartheta)^n : \vartheta \in \Theta = (0,1)\}$. Dann ist $\mathcal{P}$ eine einparametrige Exponentialfamilie in

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

Also ist die Abbildung $T : (\{0,1\}^n, \mathbb{P}(\{0,1\}^n)) \to (\mathbb{R}, \mathbb{B})$ eine für $\mathcal{P}$ suffiziente Statistik, vgl. 3.1.1.

3.4 Einige Anwendungen in der Statistik

Satz 3.4.1 (Rao-Blackwell). *Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B} = \{P_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ und $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ eine suffiziente Statistik für $\vartheta \in \Theta$. Dann gibt es zu jeder erwartungstreuen Schätzfunktion g eines reellen Parameters $\kappa(\vartheta)$ eine erwartungstreue Schätzfunktion h für $\kappa(\vartheta)$, nämlich*

$$h \in \bigcap_{\vartheta \in \Theta} E_\vartheta(g|T)$$

mit gleichmäßig nicht größerer Varianz, d.h.

$$\forall \vartheta \in \Theta : E_\vartheta((h - \kappa(\vartheta)^2)) \leq E_\vartheta((g - \kappa(\vartheta))^2). \tag{3.4.2}$$

Ist $E_\vartheta((g - \kappa(\vartheta))^2) < \infty$, so gilt:

$$E_\vartheta((h - \kappa(\vartheta)^2)) = E_\vartheta((g - \kappa(\vartheta))^2) \Leftrightarrow g = h \quad P_\vartheta\text{-f.ü.} \tag{3.4.3}$$

Beweis: Aus Lemma 3.3.5 folgt: $\exists\ h \in \bigcap_{\vartheta \in \Theta} E_\vartheta(g|T)$. Mit 3.2.6 (i) ergibt sich:

$$\forall \vartheta \in \Theta : E_\vartheta(h) = E_\vartheta(E_\vartheta(g|T)) = E_\vartheta(g) = \kappa(\vartheta),$$

d.h. h ist erwartungstreu für $\kappa(\vartheta)$.
Zum Beweis von (3.4.2) können wir o.E. annehmen, dass $E_\vartheta((g - \kappa(\vartheta))^2) < \infty$. Dann gilt:

$$\begin{aligned} E_\vartheta((g - \kappa(\vartheta))^2) &= E_\vartheta(((g - h) + (h - \kappa(\vartheta)))^2) \\ &= E_\vartheta((g - h)^2) + E_\vartheta((h - \kappa(\vartheta))^2) \\ &\geq E_\vartheta((h - \kappa(\vartheta))^2), \end{aligned}$$

da das beim Ausquadrieren auftretende gemischte Produkt verschwindet:

$$\begin{aligned} E_\vartheta(h \cdot (h - \kappa(\vartheta))) &= E_\vartheta(E_\vartheta(g|T) \cdot (h - \kappa(\vartheta))) \\ &= E_\vartheta(E_\vartheta(g \cdot (h - \kappa(\vartheta))|T)) \\ &\overset{3.2.6(i)}{=} E_\vartheta(g \cdot (h - \kappa(\vartheta))). \end{aligned}$$

Ferner gilt nun

$$E_\vartheta((h - \kappa(\vartheta)^2)) = E_\vartheta((g - \kappa(\vartheta))^2) \Leftrightarrow g = h \quad P_\vartheta\text{-f.ü.}$$

□

Durch Bildung der bedingten Erwartung bei gegebener suffizienter Statistik T kann man also aus einer erwartungstreuen Schätzfunktion g mit $\mathrm{Var}_\vartheta(g) < \infty$, $\vartheta \in \Theta$, eine bessere gewinnen, falls nicht bereits (3.4.3) erfüllt ist.

Durch abermalige Bildung der bedingten Erwartung bzgl. derselben suffizienten Statistik T erreicht man jedoch keine weitere Verbesserung, denn $\forall\vartheta \in \Theta$:

$$E_\vartheta(h|T) = E_\vartheta(E_\vartheta(g|T)|T) = E_\vartheta(g|T) = h \quad P_\vartheta\text{-f.ü.}$$

Satz 3.4.4. *Es sei $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$ und $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ eine für $\mathcal{P}/\mathcal{B}$ suffiziente Statistik. Dann gibt es zu jedem Test φ einen nur von T abhängigen Test $\psi \circ T$, mit $\psi : \mathcal{Y} \to [0, 1]$ $\mathcal{G}$-messbar, mit derselben Gütefunktion.*

Beweis: Sei $\varphi \in \Phi$. Aus Lemma 3.2.5 folgt: $\exists\, h \in \bigcap_{\vartheta\in\Theta} E_\vartheta(\varphi(T))$. Mit Satz 3.2.1 ergibt sich:

$$\exists\, \psi : (\mathcal{Y}, \mathcal{G}) \to (\mathbb{R}, \mathbb{B}) : h = \psi \circ T.$$

Da $\forall\vartheta \in \Theta : 0 \le \psi \circ T \le 1 \quad P_\vartheta$-f.ü. (da $\varphi \in [0, 1]$), kann ψ so festgelegt werden, dass $0 \le \psi \le 1$ gilt: $\psi' := \psi \cdot 1_{\{0\le\psi\le1\}}$. Dann ist ψ' $\mathcal{G}$-messbar, $0 \le \psi' \le 1$ und $\psi = \psi' \quad P_\vartheta * T$-f.ü.. Sei $\vartheta \in \Theta$:

$$E_\vartheta(\psi \circ T) = E_\vartheta(h) = E_\vartheta(E_\vartheta(\varphi|T)) \overset{3.2.6(i)}{=} E_\vartheta(\varphi).$$

□

3.5 Vollständigkeit

Im Zusammenhang mit dem Satz von Rao-Blackwell stellt sich die Frage, ob die mit Hilfe dieses Satzes gewonnene erwartungstreue Schätzfunktion bereits eine gleichmäßig kleinste Varianz besitzt (UMVU-Schätzer: uniformly minimum variance unbiased extimator). Um hierfür eine einfache hinreichende Bedingung angeben zu können, wird der Begriff der Vollständigkeit eingeführt.

Definition 3.5.1. $\mathcal{P} = \mathcal{P}/\mathcal{B} = \{P_\vartheta : \vartheta \in \Theta\}$ sei eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$.

(i) $\mathcal{P}/\mathcal{B}$ heißt *vollständig* $:\Leftrightarrow \forall f : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B}), E_\vartheta(f) = 0, \forall\vartheta \in \Theta \Rightarrow$

$$\forall\, \vartheta \in \Theta : f = 0 \quad P_\vartheta\text{-f.ü.}$$

(ii) $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ ist vollständig für $\vartheta \in \Theta :\Leftrightarrow \mathcal{P}/T^{-1}(\mathcal{G})$ ist vollständig.

Nach 3.2.1 ist $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ genau dann vollständig für $\vartheta \in \Theta$, wenn: $\forall f : (\mathcal{Y}, \mathcal{G}) \to (\mathbb{R}, \mathbb{B})$, $\int_{\mathcal{Y}} f \, d(P_\vartheta * T) = 0$, $\vartheta \in \Theta$, $\Rightarrow f = 0 \quad P_\vartheta * T$-f.ü., $\quad \vartheta \in \Theta$.
$\Leftrightarrow [\forall \vartheta \in \Theta \; E_\vartheta(f(T)) = 0, f : (\mathcal{Y}, \mathcal{G}) \to (\mathbb{R}, \mathbb{B}) \Rightarrow f \circ T = 0 \quad P_\vartheta$-f.ü.]

Die Bedeutung der „Vollständigkeit“ beruht vor allem auf dem folgenden Satz.

Satz 3.5.2 (Lehmann-Scheffé). *Es sei $\mathcal{P} = \mathcal{P}/\mathcal{B} = \{P_\vartheta : \vartheta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{B})$. Ferner sei die Statistik $T : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{G})$ suffizient und vollständig für $\vartheta \in \Theta$. Dann gilt: Falls es überhaupt eine erwartungstreue Schätzfunktion g für den reellen Parameter $\kappa(\vartheta)$ gibt, so existiert auch eine solche erwartungstreue Schätzfunktion mit Minimalvarianz. Diese optimale Schätzfunktion ist gegeben durch $h \in \bigcap_{\vartheta \in \Theta} E_\vartheta(g|T)$.*

Beweis: Sei g eine erwartungstreue Schätzfunktion für $\kappa(\vartheta)$. Nach 3.3.5 existiert

$$h \in \bigcap_{\vartheta \in \Theta} E_\vartheta(g|T)$$

mit h erwartungstreu für $\kappa(\vartheta)$ (siehe Beweis zu 3.4.1). Dieses h besitzt Minimalvarianz: Angenommen dies wäre nicht der Fall $\Rightarrow \exists \; g_1$ erwartungstreue Schätzfunktion für $\kappa(\vartheta)$ und $\exists \; \vartheta_1 \in \Theta$: $\mathrm{Var}_{\vartheta_1}(g_1) < \mathrm{Var}_{\vartheta_1}(h)$. 3.4.1 $\Rightarrow$ für $h_1 \in \bigcap_{\vartheta \in \Theta} E_\vartheta(g_1|T)$: $\mathrm{Var}_{\vartheta_1}(h_1) \le \mathrm{Var}_{\vartheta_1}(g_1) < \mathrm{Var}_{\vartheta_1}(h)$. Da h und h_1 $T^{-1}(\mathcal{G})$-messbar sind und $E_\vartheta(h) = E_\vartheta(h_1)$ $(= \kappa(\vartheta))$, $\vartheta \in \Theta$, $\Rightarrow E_\vartheta(h - h_1) = 0$, $\vartheta \in \Theta$ $\overset{\text{Vollständigkeit}}{\Rightarrow}$ $h - h_1 = 0 \quad P_\vartheta$-f.ü. bzw. $h = h_1 \quad P_\vartheta$-f.ü. $\Rightarrow \mathrm{Var}_{\vartheta_1}(h) = \mathrm{Var}_{\vartheta_1}(h_1)$, Widerspruch. Mit diesem Widerspruch ist der Satz bewiesen. □

Beispiel 3.5.3. Es sei $\mathcal{P} = \mathcal{P}/\mathbb{P}(\{0,1\}^n) := \{P_\vartheta = B(1, \vartheta)^n : \vartheta \in \Theta = (0,1)\}$. Dann ist $T(x_1, \ldots, x_n) := \sum_{i=1}^n x_i$, $(x_1, \ldots, x_n) \in \mathcal{X} = \{0,1\}^n$, vollständig für $\vartheta \in \Theta$: Sei $f \in (\mathbb{R}, \mathbb{B}) \to (\mathbb{R}, \mathbb{B})$ mit der Eigenschaft:

$$\begin{aligned}
0 &= E_\vartheta(f \circ T) \\
&= \int_{\mathcal{X}} f \circ T \, dB(1,\vartheta)^n \\
&= \int_{\{0,1,\dots,n\}} f \, d(B(1,\vartheta)^n * T) \\
&= \int_{\{0,1,\dots,n\}} f \, dB(n,\vartheta) \\
&= \sum_{i=0}^{n} f(i) B(n,\vartheta)(\{i\}) \\
&= \sum_{i=0}^{n} f(i) \binom{n}{i} \vartheta^i (1-\vartheta)^{n-i} \\
&= (1-\vartheta)^n \sum_{i=0}^{n} f(i) \binom{n}{i} \left(\frac{\vartheta}{1-\vartheta}\right)^i, \quad \vartheta \in \Theta
\end{aligned}$$

also

$$\sum_{i=1}^{k} f(i) \binom{n}{i} \left(\underbrace{\frac{\vartheta}{1-\vartheta}}_{=:z} \right)^i = 0, \quad \vartheta \in \Theta$$

d.h. das Polynom $p(z) := f(i)\binom{n}{i} z^i$, $z \in \mathbb{R}$, hat mehr als n Nullstellen $\Rightarrow$ $p(z) = 0$, $z \in \mathbb{R}$. Aus dem Identitätssatz für Polynome folgt $f(i) = 0$, $i = 0, 1, \dots, n$, d.h. $f = 0 \quad B(n,\vartheta)$-f.ü., $\vartheta \in \Theta$. Außerdem ist T suffizient für $\vartheta \in \Theta$ nach 3.3.13(iii).
Da $g(x_1, \dots x_n) := \bar{x} = \frac{1}{n} T(x_1, \dots, x_n)$ ein erwartungstreuer Schätzer für $\kappa(\vartheta) = \vartheta$ ist, der nur von T abhängt, also $g \in \bigcap_{\vartheta \in \Theta} E_\vartheta(g|T)$ folgt aus 3.5.2, dass g ein UMVU-Schätzer für $\kappa(\vartheta) = \vartheta$ ist.
Ferner ist $h(x_1, \dots, x_n) := \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 \right)$ ein erwartungstreuer Schätzer für $\kappa(\vartheta) = \vartheta \cdot (1-\vartheta)$. Da $x_i^2 = x_i$ $(\in \{0,1\})$, hängt $h = \frac{1}{n-1}\left(T - \frac{1}{n}T^2\right)$ nur von T ab, also $h \in \bigcap_{\vartheta \in \Theta} E_\vartheta(h|T)$. Somit ist h aufgrund von 3.5.2 ein erwartungstreuer Schätzer für $\vartheta(1-\vartheta)$ mit Minimalvarianz.
Hingegen ist für $n > 1$ die Familie $\mathcal{P}/\mathbb{P}(\{0,1\}^n)$ nicht vollständig: Bezeichnet $\pi_i \quad i = 1, 2$ die Projektion von $\{0,1\}^n$ auf die i-te Komponente , d.h.

$\pi_i(x_1, \dots, x_n) = x_i$, $i = 1, 2$, so gilt für beliebiges $\vartheta \in \Theta = (0,1)$:

$$
\begin{aligned}
\int \pi_1 - \pi_2 \, dP_\vartheta &= \int \pi_1 \, dP_\vartheta - \int \pi_2 \, dP_\vartheta \\
&= \int_{\{0,1\}^n} \pi_1(x_1, \dots, x_n)(B(1,\vartheta)^n \, d(x_1, \dots, x_n)) \\
&\qquad - \int_{\{0,1\}^n} \pi_2(x_1, \dots, x_n)(B(1,\vartheta)^n \, d(x_1, \dots, x_n)) \\
&= \int_{\{0,1\}} x_1 \, (B(1,\vartheta)^n * \pi_1) \, dx_1 \\
&\qquad - \int_{\{0,1\}} x_2 \, (B(1,\vartheta)^n * \pi_2) \, dx_1 \\
&= \int_{\{0,1\}} x_1 \, B(1,\vartheta)(dx_1) - \int_{\{0,1\}} x_2 \, B(1,\vartheta)(dx_2) \\
&= \vartheta - \vartheta = 0.
\end{aligned}
$$

Es gilt jedoch nicht $\pi_1 = \pi_2 \quad P_\vartheta$-f.ü.

Die Bedeutung des Satzes von Lehmann-Scheffé liegt darin, dass man für die wichtigsten Klassen von Wahrscheinlichkeitsverteilungen vollständige und suffiziente Statistiken angeben kann.

Satz 3.5.4. $\mathcal{P} = \mathcal{P}/\mathcal{B} = \{P_\vartheta : \vartheta \in \Theta\}$ *sei eine k-parametrige Exponentiafamilie derart, dass der zugehörige natürliche Parameterraum* $\tilde{\Theta} \subset \mathbb{R}^k$ *wenigstens einen inneren Punkt besitzt. Dann ist die k–dimensionale Statistik* $T = (T_1, \dots, T_k)$ *suffizient und vollständig für* $\vartheta \in \Theta$*.*

Beweis: Witting, H (1985): Mathematische Statistik, Teubner, Satz 3.39.

□

Beispiel 3.5.5. Sei $\mathcal{P} = \mathcal{P}/\mathbb{B}^n = \{P_\vartheta = N(\mu, \sigma^2)^n : \vartheta = (\mu, \sigma^2) \subset \Theta = \mathbb{R} \times (0, \infty)\}$. Nach 3.3.13(ii) ist $T(x_1, \dots, x_n) = (\sum_{i=1}^n x_i^2, \sum_{i=1}^n x_i)$ suffizient für $\vartheta \in \Theta$ $(n \geq 2)$.
3.5.4 $\Rightarrow T$ ist auch vollständig für $\vartheta \in \Theta$. Folglich sind $g(x_1, \dots, x_n) := \bar{x}$ bzw. $h(x_1, \dots, x_n) := (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2$ erwartungstreue Schätzfunktionen für $\kappa(\vartheta) = \mu$ bzw. $\kappa(\vartheta) = \sigma^2$, d.h. nur von T abhängen, aufgrund von 3.5.2 also solche mit Minimalvarianz.

Lemma 3.5.6. $\mathcal{P}_1, \mathcal{P}_2$ *seien zwei Familien von Wahrscheinlichkeitsmaßen auf* $(\mathcal{X}, \mathcal{B})$ *mit* $\mathcal{P}_1 \subset \mathcal{P}_2$ *ist* $\mathcal{P}_1/\mathcal{B}$ *vollständig und es gilt* $\mathcal{P}_2/\mathcal{B} \ll \mathcal{P}_1/\mathcal{B}$ *(d.h.* $\forall F \in \mathcal{B} : [\forall P_1 \in \mathcal{P}_1 : P_1(F) = 0 \Rightarrow \forall P \in \mathcal{P}_2 : P(F) = 0]$*), so ist auch* $\mathcal{P}_2$ *vollständig.*

Beweis: Es sei $g : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$ mit $\int g \, dP = 0$ für $P \in \mathcal{P}_2 \Rightarrow \int g \, dP = 0$ für $P \in \mathcal{P}_1 \Rightarrow g = 0 \quad P$-f.ü., $P \in \mathcal{P}_1 \Rightarrow \forall \ P \in \mathcal{P}_1 : P(\{g \neq 0\}) = 0 \Rightarrow \forall \ P \in \mathcal{P}_2 : P(\{g \neq 0\}) = 0$, d.h. $g = 0 \quad P$-f.ü., $P \in \mathcal{P}_2$. □

Definition 3.5.7. Es sei μ ein beliebiges Maß auf $(\mathcal{X}, \mathcal{B})$ und $C \in \mathcal{B}$ mit $0 < \mu(C) < \infty$. Dann heißt das Wahrscheinlichkeitsmaß

$$P_C(B) := \frac{1}{\mu(C)} \int_B 1_C \, d\mu = \frac{\mu(B \cap C)}{\mu(C)}, \quad B \in \mathcal{B}$$

μ-Gleichverteilung auf C.

Satz 3.5.8. *Sei μ ein σ–endliches Maß auf $(\mathcal{X}, \mathcal{B})$, dann ist die Familie $\mathcal{P} = \mathcal{P}/\mathcal{B} = \{P_C : C \in \mathcal{B}$ mit $0 < \mu(C) < \infty\}$ aller μ–Gleichverteilungen vollständig.*

Beweis: Sei $g : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$ mit $\frac{1}{\mu(C)} \int g \cdot 1_C \, d\mu = \int g \, dP_C = 0, \quad P_C \in \mathcal{P}$. Dann gilt $\int_C g \, d\mu = 0$, $C \in \mathcal{B}$ mit $0 < \mu(C) < \infty$. Hieraus folgt wegen der σ-Endlichkeit von μ, dass $g = 0$ μ-f.ü.: $\mathcal{X} = \bigcup_{n \in \mathbb{N}} X_n$, $X_n \in \mathcal{B}$, $\mu(X_n) < \infty$ $n \in \mathbb{N}$ und $X_1, X_2, \ldots$ sind paarweise disjunkt; $\int_{C \cap X_n} g \, d\mu = 0$, $C \in \mathcal{B}$, $n \in \mathbb{N}$,

$$C = \bigcup_{n \in \mathbb{N}} (C \cap X_n) \Rightarrow g = 0 \quad \mu\text{-f.ü.}$$

mit $C_1 := \{g > 0\}$ und $C_2 := \{g < 0\}$ folgt $g = 0 \quad P_C$-f.ü., $P_C \in \mathcal{P}$, da $P_C \ll \mu$. □

Satz 3.5.9. *Sei μ ein σ-endliches Maß auf $(\mathcal{X}, \mathcal{B})$. Dann ist die Familie aller Wahrscheinlichkeitsmaße auf $(\mathcal{X}, \mathcal{B})$, die absolut stetig sind bzgl. μ, vollständig.*

Beweis: Wegen $\mathcal{P}_1/\mathcal{B} := \{P_C : C \in \mathcal{B}$ mit $0 < \mu(C) < \infty\} \subset \mathcal{P}/\mathcal{B}$ genügt es wegen 3.5.6 und 3.5.8 zu zeigen, dass $\mathcal{P}/\mathcal{B} \ll \mathcal{P}_1/\mathcal{B}$. Dazu sei $B_0 \in \mathcal{B}$ mit $P_C(B_0) = 0$ für alle $P_C \in \mathcal{P}_1$. Angenommen $\exists P_0 \in \mathcal{P}$ mit $P_0(B_0) > 0 \Rightarrow \mu(B_0) > 0$ und wegen der σ-Endlichkeit von μ ($B_0 = \bigcup_{n \in \mathbb{N}} (\mathcal{X}_n \cap B_0)$) existiert $C_0 = \mathcal{X}_{n_0} \cap B_0$ mit $0 < \mu(C_0) < \infty \Rightarrow P_{C_0}(B_0) = \mu(B_0 \cap C_0)/\mu(C_0) = 1$, im Widerspruch zu $P_{C_0}(B_0) = 0$. □

3.6 Die Ungleichung von Cramér-Rao und die Fisher-Information

In diesem Abschnitt sei $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ und $\Theta \subset \mathbb{R}$. Ist $T : (\mathcal{X}, \mathcal{B}) \to (\mathbb{R}, \mathbb{B})$ eine erwartungstreue Schätzfunktion für $\kappa(\vartheta)$, so ist die $\mathrm{Var}_\vartheta(T)$ ein Gütemaß für den Schätzer T. Wir werden unter gewissen *Regulatitätsvoraussetzungen* eine nur von ϑ abhängige *untere Schranke* für $\mathrm{Var}_\vartheta(T)$ herleiten. Liegt dann $\mathrm{Var}_\vartheta(T)$ in der Nähe dieser unteren Schranke, so ist dies eine Aussage über die Güte von T.

Regularitätsvoraussetzungen 3.6.1. Wir fordern an $\mathcal{P}, T$ und x:

(i) Es existiere ein geeignetes σ-endliches Maß $\mu/\mathcal{B}$ und Dichten $p_\vartheta \in dP_\vartheta/d\mu$ mit $p_\vartheta(x) > 0$, $\vartheta \in \Theta$, $x \in \mathcal{X}$.

(ii) $\Theta \subset \mathbb{R}$ sei offen und die Abbildung $\Theta \ni \vartheta \mapsto p_\vartheta(x)$ sei stetig differenzierbar, $x \in \mathcal{X}$

(iii) $0 < \mathrm{Var}_\vartheta\left(\frac{\partial}{\partial\vartheta}\log(p_\vartheta)\right) < \infty, \quad \vartheta \in \Theta$

(iv) $E\left(\frac{\partial}{\partial\vartheta}\log(p_\vartheta)\right) = 0, \quad \vartheta \in \Theta$

(v) T sei eine erwartungstreue Schätzfunktion für κ. Dabei sei κ differenzierbar und es gelte

$$\kappa'(\vartheta) = \frac{\partial}{\partial\vartheta}\kappa(\vartheta) = \int T(x)\frac{\partial}{\partial\vartheta}\log(p_\vartheta(x))\ P_\vartheta(dx)$$

Motivation von Bedingung (v):

$$\begin{aligned}
\kappa'(\vartheta) &= \left(\int T(x)\ P_\vartheta(dx)\right)' \\
&= \left(\int T(x)p_\vartheta(x)\ \mu(dx)\right)' \\
&\overset{!}{=} \int T(x)p'_\vartheta(x)\ \mu(dx) \\
&= \int T(x)\ \frac{p'_\vartheta(x)}{p_\vartheta(x)}\ p_\vartheta(x)\ \mu(dx) \\
&= \int T(x)\ \frac{\partial}{\partial\vartheta}\log(p_\vartheta(x))\ P_\vartheta(dx).
\end{aligned}$$

Satz 3.6.2 (Ungleichung von Cramér und Rao). *Unter den Regularitätsbedingungen 3.6.1 gilt*

$$Var_\vartheta(T) \geq \frac{(\kappa'(\vartheta))^2}{I(\vartheta)}, \quad \vartheta \in \Theta. \tag{3.6.3}$$

Dabei ist

$$I(\vartheta) = E_\vartheta\left(\left(\frac{\partial}{\partial\vartheta}\log(p_\vartheta)\right)^2\right), \quad \vartheta \in \Theta,$$

die Fisher-Informationsfunktion*.*

Beweis: Setze $l_\vartheta := \frac{\partial}{\partial\vartheta}\log(p_\vartheta)$. Gemäß 3.6.1 gilt dann

$$E_\vartheta((T-\kappa(\vartheta))\, l_\vartheta) \overset{(iv)}{=} E_\vartheta(T\, l_\vartheta) \overset{(v)}{=} \kappa'(\vartheta).$$

Es folgt aus der Cauchy-Schwarz-Ungleichung

$$\begin{aligned}(\kappa'(\vartheta))^2 &= (E_\vartheta((T-\kappa(\vartheta))\cdot l_\vartheta))^2\\ &\leq E_\vartheta((T-\kappa(\vartheta))^2)\cdot E_\vartheta(l_\vartheta^2)\\ &= \mathrm{Var}_\vartheta(T)\cdot I(\vartheta)\end{aligned}$$

□

Beispiel 3.6.4. Sei $\mathcal{X} = (\mathbb{N}\cup\{0\})^n$, $\mathcal{B} = \mathbb{P}(\mathcal{X})$ und $\mathcal{P} = \{P_\vartheta^n : \vartheta \in (0,\infty)\}$, wobei P_ϑ die Poisson-Verteilung zum Parameter $\vartheta > 0$ bezeichne, d.h.

$$P_\vartheta(\{k\}) = e^{-\vartheta}\frac{\vartheta^k}{k!}, \quad k = 0,1,2\ldots.$$

Setze $S(x) = \sum_{i=1}^n x_i$, $x = (x_1,\ldots,x_n) \in \mathcal{X}$. Dann ist

$$\begin{aligned} l_\vartheta &= \frac{\partial}{\partial\vartheta}\log(p_\vartheta)\\ &= \frac{\partial}{\partial\vartheta}\left(-n\vartheta + \sum_{i=1}^n x_i\cdot\log\vartheta - \sum_{i=1}^n \log(x_i!)\right)\\ &= -n + S(x)\cdot\frac{1}{\vartheta}.\end{aligned}$$

Da $T_1(x) := S(x)/n$ eine erwartungstreue Schätzfunktion von ϑ ist, folgt wegen $l_\vartheta = \frac{n}{\vartheta}(T_1-\vartheta)$:

$$\begin{aligned} I(\vartheta) &= E_\vartheta\left(\left(\frac{n}{\vartheta}(T_1-\vartheta)\right)^2\right)\\ &= \frac{n^2}{\vartheta^2}E_\vartheta((T_1-\vartheta)^2)\\ &= \frac{n^2}{\vartheta^2}\int (x_1-\vartheta)^2\, P_\vartheta(dx)\\ &= \frac{\vartheta}{n}.\end{aligned}$$

Also gilt mit $\kappa(\vartheta) = \vartheta$:

$$\begin{aligned}\mathrm{Var}_\vartheta(T_1) &= \frac{n}{\vartheta} \\ &\geq \frac{(\kappa'(\vartheta))^2}{I(\vartheta)} \\ &= \frac{1^2}{\frac{n}{\vartheta}} \\ &= \frac{n}{\vartheta}\end{aligned}$$

d.h. T_1 nimmt die Cramér-Rao-Schranke an, ist also ein erwartungstreuer Schätzer mit Minimalvarianz.
Ferner ist $T_2 := \left(1 - \frac{1}{n}\right)^S$ eine erwartungstreue Schätzung für $\kappa(\vartheta) := e^{-\vartheta}$. Da S suffizient und vollständig ist, ist T_2 nach der Ungleichung von Lehmann-Scheffé eine Schätzung für $\kappa(\vartheta)$ mit Minimalvarianz.
Da

$$\mathrm{Var}_\vartheta(T_2) = e^{-2\vartheta} \cdot \left(e^{\vartheta/n} - 1\right)$$

und die Cramér-Rao-Schranke gleich

$$(\kappa'(\vartheta))^2 / I(\vartheta) = \frac{\vartheta}{n} \cdot e^{-2\vartheta}$$

ist, wird letztere also von keiner für $\kappa(\vartheta)$ erwartungstreuen Schätzung angenommen bzw. erreicht.

Bemerkung 3.6.5. (i) Aus 3.6.1 (iii), (iv) folgt, dass

$$I(\vartheta) = \mathrm{Var}_\vartheta(l_\vartheta) \in (0, \infty)$$

(ii) In (3.6.3) gilt genau dann das Gleichheitszeichen für ein $\vartheta \in \Theta$, wenn $T - \kappa(\vartheta)$ und l_ϑ linear abhängig sind, d.h. $\exists\, c_\vartheta \in \mathbb{R}$ mit $l_\vartheta = c_\vartheta \cdot (T - \kappa(\vartheta))$ P_ϑ-f.ü.. In diesem Fall gilt $I(\vartheta) = |c(\vartheta)| \cdot |\kappa'(\vartheta)|$.

(iii) Die Fischer Information $I(\vartheta)$ lässt sich deuten als eine Maßzahl für die Genauigkeit, mit welcher der unbekannte Parameter ϑ (bzw. $\kappa(\vartheta)$) aufgrund von vorliegenden Beobachtungen geschätzt werden kann. Dementsprechend heißt für eine erwartungstreue Schätzfunktion T von $\kappa(\vartheta)$ unter den Regularitätsvoraussetzungen 3.6.1 die Abbildung

$$\Theta \ni \vartheta \mapsto \frac{\text{Cramér-Rao-Schranke}}{\mathrm{Var}_\vartheta(T)} = \frac{(\kappa'(\vartheta))^2}{I(\vartheta) \cdot \mathrm{Var}_\vartheta(T)}$$

Effizienz von T.

T heißt *effiziente Schätzung* für κ, falls die Effizienz von T konstant gleich 1 ist, d.h., falls $\mathrm{Var}_\vartheta(T)$ die Cramér-Rao-Schranke erreicht für alle $\vartheta \in \Theta$.

Beispiel 3.6.6. (i) Sei $\mathcal{X} = \{0,1\}^n$, $\mathcal{B} = \mathbb{P}(\mathcal{X})$ und $\mathcal{P} = \{B(1,\vartheta)^n : \vartheta \in \Theta\}$. Dann ist $p_\vartheta(x) = \vartheta^{T(x)}(1-\vartheta)^{n-T(x)}$ mit $T(x) = \sum_{i=1}^n x_i$, $x = (x_1, \dots, x_n)$, eine Dichte von $B(1,\vartheta)^n$ bzgl. des Zählmaßes auf $(\mathcal{X}, \mathcal{B})$. Dann gilt $l_\vartheta = \frac{\partial}{\partial\vartheta} \log(p_\vartheta(x)) = \frac{T(x)}{\vartheta} - \frac{n-T(x)}{1-\vartheta} = \frac{n}{\vartheta(1-\vartheta)} \left(\frac{1}{n}T(x) - \vartheta\right)$. Da $T(x)/n$ eine erwartungstreue Schätzfunktion für $\kappa(\vartheta) = \vartheta$ ist, folgt aus Bemerkung 3.6.5 (ii), dass $T(x)/n$ sogar eine effiziente Schätzung für ϑ ist mit $I(\vartheta) = \frac{n}{\vartheta(1-\vartheta)}$.

(ii) Sei $(\mathcal{X}, \mathcal{B}, \mathcal{P})$ wie in Beispiel 3.6.4. Dann folgt aus den dortigen Überlegungen, dass $T_1(x) = \frac{1}{n}\sum_{i=1}^n x_i$ eine effiziente Schätzung für $\kappa(\vartheta) = \vartheta$ ist und andererseits, dass für $\kappa(\vartheta) = e^{-\vartheta}$ keine effiziente Schätzung existiert.

ENDE