

LUDWIG-MAXIMILIANS-UNIVERSITÄT
MÜNCHEN

BACHELORARBEIT

**Robuste Minimax-Tests und eine
Robustifizierung des Bayes-Faktors**

Autor:
Yusuf SALE

Betreuer:
Prof. Dr. Thomas AUGUSTIN
Dr. Georg SCHOLLMAYER
M. Sc. Patrick SCHWAFERTS

24. Oktober 2019



Abstract

Die robuste Statistik befasst sich mit statistischen Verfahren, die mögliche Abweichungen von zugrundeliegenden Modellannahmen berücksichtigen. Insbesondere stellt man fest, dass der Likelihood-Quotienten-Test im Allgemeinen nicht robust gegenüber Abweichungen von zugrundeliegenden Verteilungsannahmen ist. Dahingehend wird in der vorliegenden Arbeit eine robuste Version des Likelihood-Quotienten-Tests mathematisch über Umgebungsmodelle und *Least Favorable Pair* motiviert und vorgestellt. Dabei stellt man fest, dass der *gestutzte Likelihood-Quotienten-Tests* über wünschenswerte minimax Eigenschaften besitzt, weshalb man in diesem Kontext auch von robusten Minimax-Tests spricht. Diese Minimax-Resultate werden anhand des Huber-Strassen-Theorems auf bialternierende Kapazitäten in polnischen Räumen generalisiert. Das Huber-Strassen-Theorem und ebenso die damit verbundene weitreichende Bedeutung für die robuste Statistik werden in der vorliegenden Arbeit thematisiert. Neben mathematischer Ausführlichkeit in der Beweisführung und Veranschaulichung methodologischer Konstrukte, wird zudem versucht die praktische Relevanz solcher robusten Verfahren zu veranschaulichen. Um die Methodik robuster Verfahren ebenso in den bayesianischen Rahmen zu übertragen, beschäftigt sich die Arbeit weiterhin mit dem Bayes-Faktor, der eine Alternative zu klassischen Hypothesentests darstellt. Anhand der erzielten Ergebnisse wird eine robuste Version des Bayes-Faktors vorgeschlagen und ebenso verdeutlicht, wieso eine Robustifizierung sich als besonders schwierig herausstellt.

Inhaltsverzeichnis

1	Robuste Statistik	3
1.1	Robustheitskonzepte	3
1.2	Historisch-statistischer Rahmen	4
1.3	Über Likelihood-Quotienten und Bayes-Faktoren	5
2	Elemente der Testtheorie	8
3	Robuste Minimax-Tests	12
3.1	Umgebungsmodelle	12
3.2	Least Favorable Pair	15
3.3	Gestutzter LQ-Test	17
3.3.1	ϵ -Kontaminationsumgebungen	17
3.3.2	Total-Variations Umgebungen	24
3.4	Generalisierung des Neyman-Pearson-Lemmas	29
3.4.1	Kapazitäten und Umgebungsmodelle	29
3.4.2	Huber-Strassen-Theorem	33
3.5	Bedeutung des Huber-Strassen-Theorems	34
3.6	Simulation	36
4	Robustifizierung des Bayes-Faktors	38
4.1	Eine bayesianische Alternative	38
4.2	Robuste Version des Bayes-Faktors	40
5	Fazit	42
A	Anhang	43
A.1	Metrische und topologische Räume	43
A.2	Maß- und Wahrscheinlichkeitstheorie	45
B	Elektronischer Anhang	47

1 Robuste Statistik

1.1 Robustheitskonzepte

Humans may crave absolute certainty; they may aspire to it; they may pretend ... to have attained it. But the history of science - by far the most successful claim to knowledge accessible to humans—teaches that the most we can hope for is successive improvement in our understanding, learning from our mistakes, an asymptotic approach to the Universe, but with the proviso that absolute certainty will always elude us. We will always be mired in error. The most each generation can hope for is to reduce the error bars a little, and to add to the body of data to which error bars apply. The error bar is a pervasive, visible self-assessment of the reliability of our knowledge...(Sagan, 1995, S.30 - S.31)

Oft fundieren statistische Verfahren nicht nur auf tatsächlichen Beobachtungen, sondern ebenso auf vorab getroffenen *Annahmen* über Modell und Verfahren. Folglich gelten statistische Modelle lediglich als ein Versuch des mathematischen Formalismus die Realität so genau wie möglich abzubilden. Wie bereits der amerikanische Schriftsteller und Astrophysiker *Carl Sagan* erkannte, können Unsicherheiten in der Wissenschaft nicht einfach entgangen werden. Man kann lediglich versuchen, Unsicherheiten zu formalisieren und als mathematisches *Supermodel* aufzufassen. Lange versuchte man in der Statistik, viele Verfahren mit der Normalverteilungsannahme zu begründen, um jedoch festzustellen, dass Abweichungen von eben dieser exakten Verteilungsannahme in vielen Fällen zu verzerrten und instabilen Resultaten führen Huber (1981, S.1). Aufgrund der Kenntnis, dass auch geringfügige Abweichungen von zugrundeliegenden theoretischen Modellen zu enormen Verzerrungen der Resultate führen können, entwickelten sich in der Statistik sogenannte *robuste Verfahren*. Bevor näher die historische Entwicklung der robusten Statistik skizziert wird, soll zuvor diskutiert werden, was unter der Begrifflichkeit *Robustheit* zu verstehen ist und welche wesentlichen Ziele in der robusten Statistik verfolgt werden. Eine mögliche Definition von Robustheit folgt nach Peter J. Huber, einen der wesentlichen Pioniere der robusten Statistik:

» The word robust is loaded with many-sometimes inconsistent connotations. We use it in a relatively narrow sense: for our purposes, robustness signifies insensitivity to small deviations from the assumptions « (Huber, 1981, S.1)

Somit sollen nach Huber, robuste Verfahren stabil gegenüber kleineren Abweichungen von theoretischen Annahmen sein. Weiterhin formuliert Frank R. Hampel Folgendes im Bezug auf die robuste Statistik:

» Robust statistics, in a loose, nontechnical sense, is concerned with the fact that many assumptions commonly made in statistics (such as normality, linearity, in-

dependence) are at most approximations to reality. One reason is the occurrence of gross errors, such as copying or keypunch errors. They usually show up as outliers, which are far away from the bulk of the data, and are dangerous for many classical statistical procedures. The outlier problem is well known and probably as old as statistics, and any method for dealing with it, such as subjective rejection or any formal rejection rule, belongs to robust statistics in this broad sense. « (Hampel, 1986, S.1)

Generell existieren in der robusten Statistik eine Vielzahl von Kriterien und Definitionen für Robustheit. Man unterscheidet hierbei zwischen *quantitativer* und *qualitativer* Robustheit. Das Konzept der qualitativen Robustheit stellt ein *Stetigkeitsargument* dar, trifft jedoch keine weiteren Aussagen über den Grad an Robustheit von statistischen Verfahren. Dahingegen versucht man mittels quantitativen Verfahren Aussagen über das *Ausmaß* von Robustheit zu treffen (Büning, 2012, S.31). Dabei trifft man Entscheidungen über den Grad an Robustheit anhand Kriterien wie das Niveau α oder die Güte β . Zusammenfassend lässt sich angelehnt an (Huber, 1981, S.5) Folgendes an robuste Verfahren postulieren:

1. *Effizienz*: Das angenommene Modell soll im statistischen Sinne effizient sein.
2. *Stabilität*: Kleinere Abweichungen vom zugrundeliegenden Modell sollen zu keiner enormen Verzerrung des Resultates führen.
3. *Breakdown*: Das Verfahren soll unter größeren Abweichungen nicht vollständig unbrauchbar werden.

Um die robuste Statistik auch in den historisch-statistischen Kontext einzuräumen, soll im folgenden Abschnitt auf eine nähere Entwicklung der robusten Statistik und eine Abgrenzung von weiteren statistischen Zweigen eingegangen werden.

1.2 Historisch-statistischer Rahmen

Gehen die Wurzeln der parametrischen Statistik bis auf *Carl Friedrich Gauß* (1777-1855) zurück, bildet die robuste Statistik dagegen ein relativ junges Teilgebiet der statistischen Inferenz (Büning, 2012, S.1). Während sich die sogenannte *non-parametrische* Statistik mit Arbeiten von *Abraham Wald* (1902-1950), *Nikolai Smirnov* (1900-1968), sowie anderen namhaften Statistikern wie *Frank Wilcoxon* (1892-1965) bis Mitte des 20. Jahrhunderts entwickelte, entstanden in den 50er und 60er Jahren erste Arbeiten bezüglich der robusten Statistik. Mit Bedacht, dass angenommene Modelle lediglich ein approximatives Abbild der exakten Situation darstellen, kann die robuste Statistik als Erweiterung der parametrischen Statistik betrachtet werden (s. **Abb. 1**). *George Box* erkannte bereits 1953, dass bei kleineren Abweichungen von der Normalverteilungsannahme, das Niveau des

F-Tests deutlich größer wurde. Maßgeblich beeinflusst wurde die robuste Statistik vor allem durch den sogenannten Minimax-Ansatz 1964 von *Peter J. Huber*. Die Lösung des Minimax-Problems führte letztendlich 1973 zu einer Generalisierung des *Neyman-Pearson-Lemmas* für *bialternierende Kapazitäten* und damit auch zu einer robusten Version der Likelihood-Quotienten-Statistik. Eine infinitesimale Betrachtung wurde 1968 von *Frank R. Hampel* vorgeschlagen.

Im Folgenden sollen nun einige eingangs erwähnte Aspekte der robusten Statistik aufgegriffen und im Bezug auf die Zielsetzung der vorliegenden Abschlussarbeit diskutiert werden.

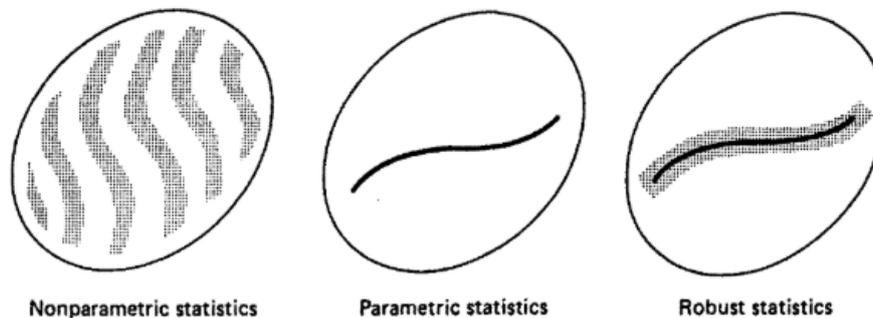


Abbildung 1: Non-parametrische, parametrische und robuste Statistik im Vergleich (Hampel, 1986, S.7)

1.3 Über Likelihood-Quotienten und Bayes-Faktoren

Anhand der bisherigen Überlegungen über robuste Verfahren stellt sich nun die Frage, wie sich klassische Verfahren tatsächlich bei Abweichungen von den prior getroffenen Annahmen über das statistische Modell verhalten. Im Rahmen der vorliegenden Arbeit wird der Fokus auf Abweichungen von zugrundeliegenden *Verteilungsannahmen* liegen. Somit stellt sich konkret die Frage, was passiert, falls die tatsächliche Verteilung, bedingt durch sogenannten *gross-error* (im Form von Ausreißern), von der theoretisch angenommenen Verteilung abweicht? Die **Abb. 2** soll exemplarisch den Kontrast zwischen der tatsächlichen und der angenommenen Verteilung aufzeigen. Dabei erkennt man, dass durch *gross error* die tatsächliche Verteilung an den Rändern eine höhere Wahrscheinlichkeitsmasse als erwartet aufzeigt. Konkret handelt es sich hierbei um die Dichte einer Standardnormalverteilung (angenommene Verteilung) und um eine t-Verteilung mit drei Freiheitsgraden (tatsächliche Verteilung). Welche Konsequenzen kann nun solch eine Abweichung von der Verteilungsannahme auf statistische Verfahren, insbesondere auf statisti-

1. ROBUSTE STATISTIK

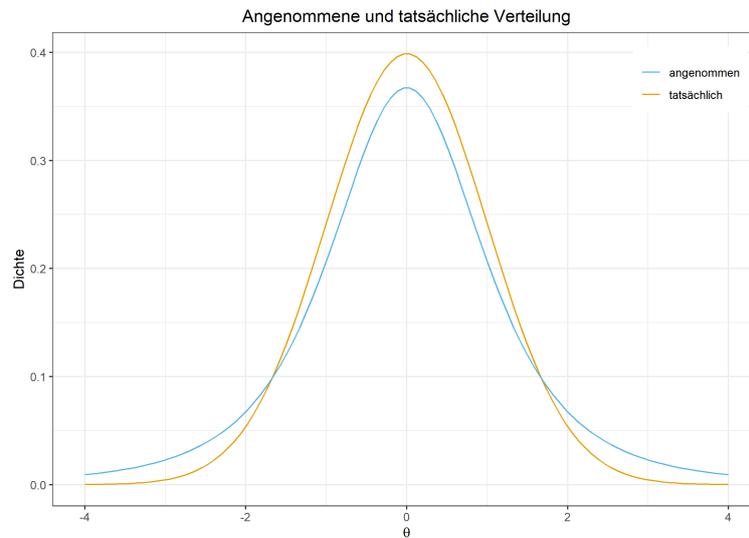


Abbildung 2: Veranschaulichung der angenommenen und tatsächlichen Verteilung (Eigene Skizze)

sche Hypothesentests haben? Dafür soll nun das einfache Testproblem

$$H_0 : \theta = \theta_0 \quad \text{gegen} \quad H_1 : \theta = \theta_1 \quad (1)$$

betrachtet werden. Um eine Entscheidung für bzw. gegen die Nullhypothese zu treffen, zieht man klassisch die *Likelihood-Quotienten-Statistik* (*LQ-Statistik*)

$$\Lambda_{\mathbf{X}}(x) = \prod_{i=1}^n \frac{f(x_i|\theta_1)}{f(x_i|\theta_0)} = \frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} \quad (2)$$

heran. Bei $f(\mathbf{x}|\theta_0)$ und $f(\mathbf{x}|\theta_1)$ handelt es sich jeweils um die Dichte der Verteilung des Zufallsvektors $\mathbf{X} = (X_1, \dots, X_n)$ unter den Hypothesen (1), wobei X_i unabhängig und identisch verteilte (u.i.v.) Zufallsvariablen sind, und $\mathbf{x} = (x_1, \dots, x_n)$ den entsprechenden Realisierungsvektor von \mathbf{X} darstellt. Anhand (2) wird leicht ersichtlich, dass bereits ein Faktor $\frac{f(x_i|\theta_1)}{f(x_i|\theta_0)} \approx 0$ die Teststatistik $\Lambda_{\mathbf{X}}$ hinreichend verzerren kann (Huber, 1965, S. 1753). Somit hält man in Hinblick auf die Robustheit der LQ-Statistik fest, dass es sich hierbei um eine im Allgemeinen nicht robuste Teststatistik handelt. Um dies zu illustrieren, wurden in der **Abb. 3** die p-Werte des Likelihood-Quotienten-Tests (LQ-Test) der Hypothesen

$$H_0 : \mu = 0 \quad \text{gegen} \quad H_1 : \mu \neq 0$$

unter der tatsächlichen und der angenommenen Verteilung dargestellt. Die Vertei-

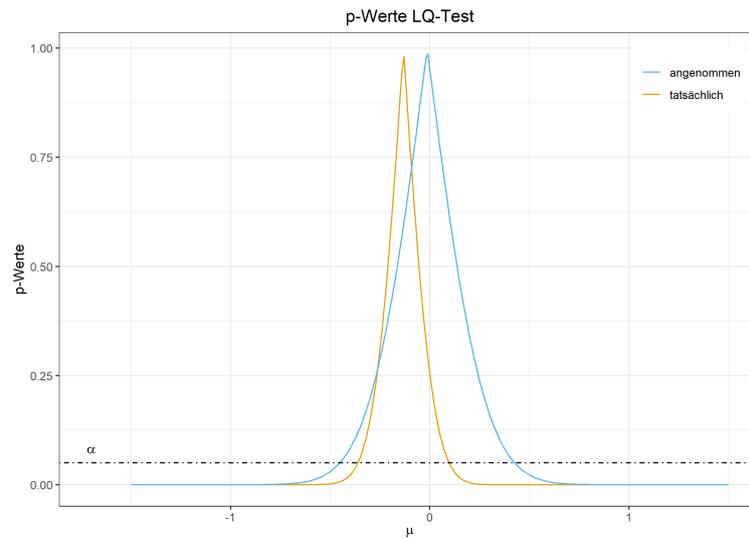


Abbildung 3: p-Werte des LQ-Testes unter tatsächlicher und angenommener Verteilung (Eigene Skizze)

lungen unter den jeweiligen Hypothesen entsprechen denen der **Abb. 2**. Anhand dieses Beispiels erkennt man, dass eine fehlende Robustheit der Teststatistik, deutlich das Resultat des Tests verzerren kann. Um solchen Konsequenzen entgegenzuwirken, wurde von Huber (1965) eine *gestutzte* Version des LQ-Testes vorgeschlagen. Die vorliegende Arbeit soll die Methodik und mathematischen Hintergründe der *gestutzten LQ-Statistik* näher darstellen und sich damit um eine Ausarbeitung von robusten Minimax-Tests und Konstruktionsmöglichkeiten bemühen.

Des Weiteren wird beabsichtigt, den erarbeiteten robusten Rahmen auf den sogenannten *Bayes-Faktor*, der eine bayesianische Alternative zu klassischen Hypothesentests bietet, zu erweitern, um somit eine *robuste Version des Bayes-Faktors* vorzustellen.

Für das Verständnis und die Leserlichkeit der mathematischen Konzepte, die im Rahmen dieser Arbeit vorgestellt werden, wird an entsprechenden Stellen explizit auf die jeweiligen Definitionen im Anhang verwiesen. Hierfür verweist man über hochgestellte Symbolik auf den entsprechenden Teil des Anhanges (bspw. *metrischer Raum*^{A.1} verweist auf die Definition des *metrischen Raumes* im Anhang A.1).

2 Elemente der Testtheorie

In diesem Abschnitt soll nun vorbereitend auf die konzeptionelle Darstellung von robusten Minimax-Tests zunächst in die grundlegenden Elemente der Testtheorie eingeführt werden¹.

Definition 2.1 (Hypothesen)

Eine Hypothese ist eine Aussage über einen Parameter θ aus dem Parameterraum Θ , wobei $\Theta_0 \cap \Theta_1$ eine disjunkte Zerlegung des Parameterraums bildet. Zwei komplementäre Hypothesen werden in der Testtheorie Nullhypothese H_0 und Gegenhypothese H_1 genannt und man schreibt

$$H_0 : \theta \in \Theta_0 \quad \text{gegen} \quad H_1 : \theta \in \Theta_1, \quad (3)$$

wobei $\theta_0 \in \Theta_0$ und $\theta_1 \in \Theta_1$. Falls $\Theta_0 = \{\theta_0\}$ bzw. $\Theta_1 = \{\theta_1\}$ gilt, nennt man die Nullhypothese H_0 bzw. Gegenhypothese H_1 einfach. Ansonsten werden die Hypothesen zusammengesetzt genannt.

Definition 2.2 (Konstruktion eines Tests)

Ein (nicht randomisierter) Test ist durch die Festlegung einer Teststatistik $\mathcal{T}(x)$ und eines kritischen Bereichs \mathcal{A} eindeutig bestimmt:

$$\phi(x) = \begin{cases} 1, & \text{falls } \mathcal{T}(x) \in \mathcal{A}. \\ 0, & \text{sonst.} \end{cases}$$

Vor allem bei der Konstruktion und Optimierung von robusten Testverfahren spielen Gütekriterien eine wichtige Rolle. Um diese näher definieren zu können, müssen vorerst mögliche Fehlerarten von statistischen Tests betrachtet werden.

Bemerkung 2.1 (Gütefunktion)

Bezeichne mit $\mathbb{P}_\theta(H_0)$ die Wahrscheinlichkeit sich für die Nullhypothese zu entscheiden und sei analog $\mathbb{P}_\theta(H_1)$ für die Alternative definiert. Dann gilt mit ϕ beliebiger Test für (3)

$$\mathbb{P}_\theta(H_1) = \int \mathbb{P}(H_1|x)f(x|\theta) dx \stackrel{\text{Def. 2.2}}{=} \int \phi(x)f(x|\theta) dx = \mathbb{E} \{ \phi(X) \}. \quad (4)$$

Mit (4) folgt somit auch

$$\mathbb{P}_\theta(H_0) = \mathbb{E} \{ 1 - \phi(X) \} \quad \forall \theta \in \Theta. \quad (5)$$

Man nennt $\mathcal{G}_\phi(\theta) := \mathbb{E} \{ \phi(X) \}$ Gütefunktion des Tests ϕ .

¹Definitionen und Sätze sind bspw. zu finden in Held (2008), Rüger (1998) und Rüschemdorf (2014).

2. ELEMENTE DER TESTTHEORIE

Aus **Bem. 2.1** folgt die Definition der Fehlerarten:

Definition 2.3 (Fehlerarten)

Die Wahrscheinlichkeit für den Fehler erster Art α ist gegeben durch

$$\mathcal{G}_\phi(\theta) = \mathbb{P}_\theta(H_1) \quad \text{für } \theta \in \Theta_0. \quad (6)$$

Man spricht von der Wahrscheinlichkeit für den Fehler zweiter Art β , falls

$$1 - \mathcal{G}_\phi(\theta) = \mathbb{P}_\theta(H_0) \quad \text{für } \theta \in \Theta_1. \quad (7)$$

Gilt in (6) $\theta \in \Theta_1$, so spricht man von der Macht des Tests ϕ . Weiterhin gilt, dass es sich bei

$$\alpha_t = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(H_1) = \sup_{\theta \in \Theta_0} \mathcal{G}_\phi(\theta) \quad (8)$$

um das tatsächliche Niveau des Testes ϕ handelt. Analog gilt für den Fehler zweiter Art:

$$\beta_t = \sup_{\theta \in \Theta_1} \mathbb{P}_\theta(H_0) = 1 - \inf_{\theta \in \Theta_1} \mathcal{G}_\phi(\theta). \quad (9)$$

Bei der Minimierung von α bzw. β braucht man ein Tradeoff-Argument, denn eine gleichzeitige Minimierung beider Fehlerarten ist nicht realisierbar. Von zentraler Rolle für die Konstruktion von robusten Tests sind sogenannte *Minimax-Tests*.

Definition 2.4 ((Niveau- α)-Minimax-Test)

Sei Φ_α die Klasse aller statistischen Tests zum Niveau $\alpha \in (0, 1)$ für das Hypothesenpaar (3). Dann nennt man ϕ^* einen (Niveau- α) Minimax-Test² zum festen Niveau α , falls

$$\inf_{\theta \in \Theta_1} \mathbb{E} \{ \phi^*(X) \} = \sup_{\phi \in \Phi_\alpha} \inf_{\theta \in \Theta_1} \mathbb{E} \{ \phi(X) \} \quad (10)$$

gilt. Einfachheitshalber bezeichnet man im Folgenden (Niveau- α) Minimax-Tests nur noch als Minimax-Tests. In der Literatur werden Minimax-Tests oft auch Maximin-Tests genannt (Rüschendorf, 2014, S.182). Weiterhin von Bedeutung für die klassische Testtheorie sind Verlust- und Risikofunktion:

Definition 2.5 (Neyman-Pearson Verlustfunktion)

Man nennt für $L_0, L_1 > 0$

$$L_\phi(\theta, H_1) = \begin{cases} 0, & \text{falls } \theta \in \Theta_1. \\ L_0, & \text{falls } \theta \in \Theta_0. \end{cases} \quad (11)$$

²Durch die Namensgebung wird ausgedrückt, dass Minimax-Tests über die Definition des kritischen Bereichs der LQ-Statistik vom Niveau α abhängen.

$$L_\phi(\theta, H_0) = \begin{cases} 0, & \text{falls } \theta \in \Theta_0. \\ L_1, & \text{falls } \theta \in \Theta_1. \end{cases} \quad (12)$$

die (Neyman-Pearson) Verlustfunktion von ϕ .

Somit beschreibt die Neyman-Pearson-Verlustfunktion den Verlust, den man bei Verwendung des statistischen Tests ϕ für das Testproblem (3) erleidet. Um den erwarteten Verlust quantifizieren zu können, braucht man eine Definition für die *Risikofunktion* (Rüschendorf, 2014, S.21).

Definition 2.6 (Risikofunktion)

Für die Funktion über das Risiko des Tests ϕ bzgl. (3) schreibt man:

$$\mathcal{R}(\theta, \phi) = \begin{cases} L_0 \mathbb{E}_\theta \{ \phi(X) \}, & \theta \in \Theta_0 \\ L_1 \mathbb{E}_\theta \{ 1 - \phi(X) \}, & \theta \in \Theta_1 \end{cases} \quad (13)$$

Ein weiteres Grundwerkzeug zur Konstruktion von robusten Minimax-Tests stellt das *Neyman-Pearson-Lemma* für einfach Hypothesen dar. Um das Fundamentallemma einzuführen, wird vorab der Likelihood-Quotienten-Test definiert.

Definition 2.7 (Likelihood-Quotienten-Test)

Betrachtet man nun erneut das einfach Testproblem (3), dann nennt man

$$\Lambda_{\mathbf{X}}(x) = \prod_{i=1}^n \frac{f(x_i|\theta_1)}{f(x_i|\theta_0)} = \frac{f(\mathbf{x}|\theta_1)}{f(\mathbf{x}|\theta_0)} \quad (14)$$

Likelihood-Quotient. Somit gilt für die Form des Likelihood-Quotienten-Tests:

$$\tilde{\phi}(x) = \begin{cases} 1, & \text{falls } \Lambda_{\mathbf{X}}(x) > k \\ \gamma, & \text{falls } \Lambda_{\mathbf{X}}(x) = k \\ 0, & \text{falls } \Lambda_{\mathbf{X}}(x) < k \end{cases} \quad (15)$$

Damit formuliert sich das *Neyman-Pearson-Lemma* zu:

Satz 2.1 (Neyman-Pearson-Lemma)

Liege wiederum das einfache Testproblem (3) vor und sei Φ_α die Klasse aller Niveau- α -Tests für (3). Dann gilt für $\alpha \in (0, 1)$:

1. Für jedes k und $\gamma \in (0, 1)$ besitzt $\tilde{\phi}$ maximale Macht unter allen $\phi \in \Phi_\alpha$.
2. Für $\alpha \in (0, 1)$ existieren k und $\gamma \in (0, 1)$, sodass $\tilde{\phi}$ das Niveau α exakt ausschöpft.

2. ELEMENTE DER TESTTHEORIE

3. $\tilde{\phi}$ ist eindeutig, d.h. falls $\phi \in \Phi_\alpha$ maximale Macht besitzt so gilt $\phi = \tilde{\phi}$.

Auf einen Beweis des Neyman-Pearson-Lemmas wird hier verzichtet, dieser kann jedoch in (Rüschendorf, 2014, S.190 - S.191) nachgelesen werden. Somit kann aus dem sogenannten Fundamentallemma eine Konstruktionsmöglichkeit von optimalen Tests für einfache Hypothesen extrahiert werden (Rüschendorf, 2014, S.190). Bei der Konstruktion von robusten Testverfahren wird das Neyman-Pearson-Lemma auf zusammengesetzte Hypothesen angewandt, um minimax Resultate zu erhalten.

3 Robuste Minimax-Tests

3.1 Umgebungsmodelle

Wie bereits in *Abschnitt 1.3* angesprochen, besteht das Ziel der robusten Statistik darin, Verfahren zu entwickeln, die robust gegenüber Abweichungen von den prior Annahmen sind. In diesem Zusammenhang werden in der robusten Testtheorie statistische Modelle betrachtet, die eventuelle Abweichungen von den vorab getroffenen Annahmen berücksichtigen (Hampel, 1986, S.39). Betrachtet man nun erneut das Hypothesenpaar (3) und formuliert dieses anhand der zugrundeliegenden Verteilungen $P_0 = f_0 \odot \mu$ und $P_1 = f_1 \odot \mu$ mit μ dominierendes Maß^{A.2}, so erhält man

$$H_0 : P = P_0 \quad \text{gegen} \quad H_1 : P = P_1. \quad (16)$$

Um mögliche Abweichungen zu berücksichtigen, geht man von der Idee einer bestimmten Verteilungsannahme P_i ($i = 0, 1$) über zu *Umgebungen*^{A.1} von Verteilungen (Huber, 1965, S.1753). Somit erhält man eine Formulierung von (16) im Kontext von sogenannten *Umgebungsmodellen*:

$$H_0 : P \in \mathcal{P}_0 \quad \text{gegen} \quad H_1 : P \in \mathcal{P}_1 \quad (17)$$

Umgebungen werden mathematisch motiviert über

$$\mathcal{P}_* = \{Q \mid d_*(Q, P) \leq \epsilon\} \quad \text{für} \quad \epsilon \geq 0, \quad (18)$$

wobei d_* beliebige Metrik^{A.1} auf (Ω, \mathcal{X}) mit \mathcal{X} Borelsche σ -Algebra^{A.2} und $Q, P \in \mathcal{W} := \mathcal{W}(\Omega, \mathcal{X})$ ist. Dabei ist \mathcal{W} die Menge aller Wahrscheinlichkeitsmaße^{A.2} auf den Messraum^{A.2} (Ω, \mathcal{X}) . In **Abb. 4** wird die Grundidee, Hypothesen nun nicht mehr über eine feste Verteilungsannahme, sondern über eine Umgebung von Verteilungen zu formulieren, dargestellt.

Nun sollen Umgebungsmodelle, die im Rahmen der robusten Statistik von zentraler Rolle sind, dargestellt und über die jeweiligen zugrundeliegenden Metriken motiviert werden.

ϵ -Kontamination

Die sogenannte ϵ -Kontaminationsumgebung, in der Literatur auch Gross-Error-Model genannt, eignet sich besonders zum Modellieren von Ausreißern und allgemeinen Modellfehlern (Hafner, 1992, S.147):

$$\mathcal{P}_\epsilon = \{Q \in \mathcal{W} \mid Q = (1 - \epsilon)P + \epsilon\mathcal{H}, \mathcal{H} \in \mathcal{W}\} \quad (19)$$

Anschaulich lassen sich somit Abweichungen von der zugrundeliegenden Verteilungsannahme über $0 \leq \epsilon \leq 1$ steuern. Dabei gibt ϵ den Anteil an Ausreißern an.

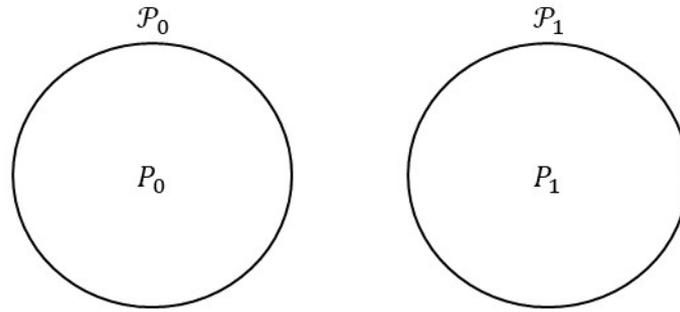


Abbildung 4: Hypothesen in Umgebungen \mathcal{P}_i ($i = 0, 1$) (Eigene Skizze)

Total-Variation

Der Total-Variations Umgebung liegt die Distanz

$$d_{tv}(P, Q) = \sup_{A \in \mathcal{X}} |Q(A) - P(A)| \quad (20)$$

zugrunde. Ist $\Omega = \mathbb{R}$ dann gilt³

$$d_k(P, Q) = \sup_{x \in \mathbb{R}} |Q(x) - P(x)| \quad (21)$$

und man nennt $d_k(P, Q)$ Kolmogorov-Distanz. Mit (20) formuliert man für allgemeine⁴ Räume die Total-Variations-Umgebung

$$\mathcal{P}_{tv} = \{Q \in \mathcal{W} \mid Q(A) \leq P(A) + \epsilon\}, \quad (22)$$

wobei $\epsilon > 0$ und $A \in \mathcal{X}$ ist. Setzt man $\Omega = \mathbb{R}$ und damit $\mathcal{X} = \mathcal{B}(\mathbb{R})$, so folgt die Kolmogorov-Umgebung zu

$$\mathcal{P}_k = \{Q \in \mathcal{W} \mid \forall A \in \Omega : Q(A) \leq P(A) + \epsilon\}. \quad (23)$$

Prohorov-Lévy

Die Prohorov Metrik für allgemeine Räume lautet für $\epsilon > 0$

$$d_p(P, Q) = \inf \{Q(A) \leq P(A_\epsilon) + \epsilon\} \quad (24)$$

wobei

$$A_\epsilon = \{x \in \Omega \mid d(x, A) \leq \epsilon\} \quad \text{mit } A \in \mathcal{X} \quad (25)$$

³Für $\Omega = \mathbb{R}$ handelt es sich trivialerweise bei P bzw. Q um Verteilungsfunktionen, da ansonsten $Q(x) = P(x) = 0$ für $x \in \mathbb{R}$ wäre (Gleiches gilt für die Formulierung der Lévy-Distanz).

⁴Mit allgemeinen Räumen meint man polnische Räume^{A.2}.

3. ROBUSTE MINIMAX-TESTS

und $\epsilon > 0$ ist. Für $\Omega = \mathbb{R}$ und somit für alle $x \in \mathbb{R}$ gilt:

$$d_l(P, Q) = \inf \{P(x - \epsilon) - \epsilon \leq Q(x) \leq P(x + \epsilon)\} \quad (26)$$

Man nennt (26) Lévy-Distanz, die in der **Abb. 5** veranschaulicht wird. Für $\epsilon > 0$ folgt mit (24) bzw. (26) damit die sogenannte Prohorov-

$$\mathcal{P}_p = \{Q \in \mathcal{W} \mid \forall A \in \mathcal{X} : Q(A) \leq P(A_\epsilon) + \epsilon\} \quad (27)$$

bzw. Lévy-Umgebung

$$P_\epsilon = \{Q \in \mathcal{W} \mid P(x - \epsilon) - \epsilon \leq Q(x) \leq P(x + \epsilon) + \epsilon\}. \quad (28)$$

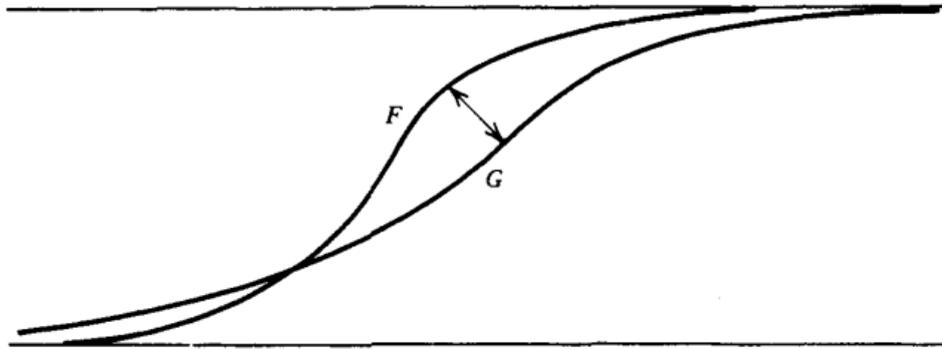


Abbildung 5: Lévy-Distanz zwischen den Verteilungsfunktionen F und G (Huber, 1981, S.28)

3.2 Least Favorable Pair

Neben den vorgestellten Umgebungsmodellen, spielen auch sogenannte *Least Favorable Pairs (LFP)*⁵ in der Konstruktion von robusten Minimax-Tests eine besonders wichtige Rolle. Der Begriff *Least Favorable* geht zurück auf Lehmann (1986) und bietet die Grundlage des robusten Minimax-Ansatzes nach Huber (1965). Die Idee von LFP soll nun anhand eines Minimalbeispiels angelehnt an Österreicher (2006, S.4108) erklärt werden:

Beispiel 3.1 (Unfaire Münze)

Bei einem Münzwurf mit einer unfairen Münze ist man sich nicht sicher, ob Kopf oder Zahl mit der unbekanntem Wahrscheinlichkeit θ auftritt (somit: Kopf oder Zahl mit Wahrscheinlichkeit $1 - \theta$). Damit erhält man zwei mögliche Wahrscheinlichkeitsverteilungen für den Wurf mit einer unfairen Münze:

$$P = (\theta, 1 - \theta) \tag{29}$$

$$Q = (1 - \theta, \theta) \tag{30}$$

Zudem sei bekannt, dass $0.2 \leq \theta \leq 0.4$ gilt. Mit der Kenntnis über Umgebungsmodelle aus dem vorherigen Abschnitt 3.1 lassen sich folgende zusammengesetzte Hypothesen formulieren:

$$\mathcal{P}_0 := \{P : |\theta - 0.3| \leq 0.1\} \quad \text{gegen} \quad \mathcal{P}_1 := \{Q : |\theta - 0.3| \leq 0.1\} \tag{31}$$

*Um bei diesem Testproblem mit Bezug auf (10) beide Fehlerarten zu kontrollieren, muss für den ungünstigsten Fall die Anzahl n an Würfeln bestimmt werden. Dadurch lassen sich konkret Verteilungen $(Q_0, P_0) \in (\mathcal{P}_0 \times \mathcal{P}_1)$ extrahieren, für die sich das Testproblem am schwierigsten gestaltet. Man nennt (Q_0, P_0) *Least Favorable Pair*. In diesem Fall folgt trivialerweise, dass $Q_0 = (0.4, 0.6)$ und $Q_1 = (0.6, 0.4)$.*

Somit ist das *Testproblem am schwierigsten*, wenn sich die Verteilungen unter den jeweiligen Hypothesen am nächsten sind (Huber, 1965, S.1753). Diese Feststellung ist für die Konstruktion von robusten Minimax-Tests essentiell und wird in der **Abb. 6** skizzenhaft veranschaulicht. Weiterhin sollen für das Verständnis der folgenden Abschnitte LFP formal definiert werden:

Least Favorable Pair

Ein Tupel (Q_0, Q_1) nennt man LFP, falls für jeden LQ-Test $\tilde{\phi}$ für das Testproblem

⁵In der Literatur spricht man auch häufig von *Least Favorable Pair of Distributions* (siehe bspw. Österreicher (2006)).

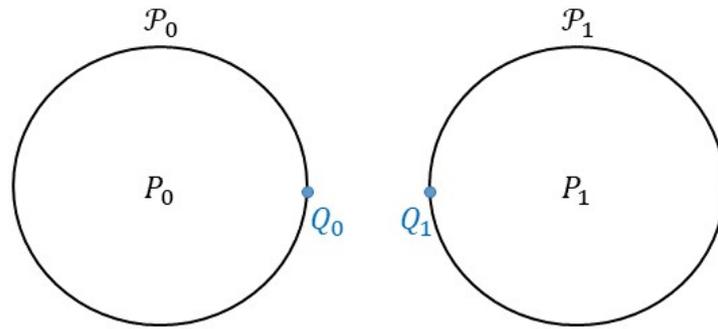


Abbildung 6: Darstellung von (Q_0, Q_1) (Eigene Skizze)

(17) unter Ausnutzung von (13) gilt $\forall (P, Q) \in \mathcal{P}_0 \times \mathcal{P}_1$:

$$\mathcal{R}(P, \tilde{\phi}) \leq \mathcal{R}(Q_0, \tilde{\phi}) \quad (32)$$

$$\mathcal{R}(Q, \tilde{\phi}) \leq \mathcal{R}(Q_1, \tilde{\phi}) \quad (33)$$

Somit werden die Wahrscheinlichkeiten für den Fehler erster und zweiter Art der $(Q, P) \in \mathcal{P}_0 \times \mathcal{P}_1$ über die Fehlerwahrscheinlichkeiten von (Q_0, Q_1) beschränkt (Österreicher, 2006, S.4108).

Das Ziel des nächsten Abschnittes besteht darin, robuste Minimax-Tests für ϵ -Kontaminations- und Total-Variations-Umgebungen über Minimierung der rechten Seiten von (32) und (33) zu motivieren. Dabei wird (Q_0, Q_1) so gewählt, dass das Resultat eine *gestutzte Version* der LQ-Statistik darstellt und minimax für Testprobleme der Form (17) ist.

3.3 Gestutzter LQ-Test

3.3.1 ϵ -Kontaminationsumgebungen

Wie bereits in *Abschnitt 1.3* motiviert wurde, ist der LQ-Test im Allgemeinen nicht robust gegenüber Abweichungen von den Verteilungsannahmen. Mithilfe der in den *Abschnitten 3.2* und *3.3* vorgestellten Konzepte, soll nun für die ϵ -Kontaminations- und Total-Variations-Umgebung eine robuste Version des LQ-Testes erarbeitet und vorgestellt werden. Der sogenannte *gestutzte LQ-Test* geht dabei zurück auf Huber (1965). Sei nun erneut (Ω, \mathcal{X}) ein Messraum und $P_0, P_1 \in \mathcal{W}$, wobei \mathcal{W} die Klasse aller Wahrscheinlichkeitsverteilungen auf (Ω, \mathcal{X}) ist, dann erhält man erneut das einfache Testproblem

$$H_0 : P = P_0 \quad \text{gegen} \quad H_1 : P = P_1. \quad (34)$$

Um eventuelle Abweichungen (im Sinne von *gross error*) zu berücksichtigen, werden analog zum *Abschnitt 3.1* die einfachen Hypothesen (34) als zusammengesetzte Hypothesen formuliert. Mit der ϵ -Kontaminations-Umgebung (19) erhält man für die Nullhypothese

$$\mathcal{P}_{\epsilon_0} = \{Q \in \mathcal{W} \mid Q = (1 - \epsilon_0)P_0 + \epsilon_0 \mathcal{H}_0, \mathcal{H}_0 \in \mathcal{W}\} \quad (0 \leq \epsilon_0 \leq 1) \quad (35)$$

und für die Alternative

$$\mathcal{P}_{\epsilon_1} = \{Q \in \mathcal{W} \mid Q = (1 - \epsilon_1)P_1 + \epsilon_1 \mathcal{H}_1, \mathcal{H}_1 \in \mathcal{W}\} \quad (0 \leq \epsilon_1 \leq 1). \quad (36)$$

Damit erhält man das folgende Testproblem zwischen den zwei Umgebungen \mathcal{P}_{ϵ_0} und \mathcal{P}_{ϵ_1} :

$$H_0 : P \in \mathcal{P}_{\epsilon_0} \quad \text{gegen} \quad H_1 : P \in \mathcal{P}_{\epsilon_1} \quad (37)$$

Für das Hypothesenpaar (37) und ϕ Test der Form (3) formulieren sich dann die Minimax-Testprobleme mit $\alpha \in [0, 1]$ und $Q_i^* \in \mathcal{P}_{\epsilon_i}$ ($i = 0, 1$) zu:

1. $\min_{i \in \{0,1\}} (\max_{Q_i^*} \sup \mathcal{R}(Q_i^*, \phi))$
2. $\min_{Q_1^*} (\sup \mathcal{R}(Q_1^*, \phi))$ für $\sup_{Q_0^*} \mathcal{R}(Q_0^*, \phi) \leq \alpha$

3. ROBUSTE MINIMAX-TESTS

Nun soll gezeigt werden, dass gemäß *Abschnitt 3.2* ein Paar $(Q_0, Q_1) \in \mathcal{P}_{\epsilon_0} \times \mathcal{P}_{\epsilon_1}$ existiert, sodass

$$\forall Q_0^* \in \mathcal{P}_{\epsilon_0} : \mathcal{R}(Q_0^*, \phi) \leq \mathcal{R}(Q_0, \phi) \quad (38)$$

$$\forall Q_1^* \in \mathcal{P}_{\epsilon_1} : \mathcal{R}(Q_1^*, \phi) \leq \mathcal{R}(Q_1, \phi) \quad (39)$$

gilt. Für solche $Q_i \in \mathcal{P}_{\epsilon_i}$ mit $i \in \{0, 1\}$ lassen sich analog zu $Q_i^* \in \mathcal{P}_{\epsilon_i}$ folgende Minimax-Testprobleme formulieren:

$$3. \min_{i \in \{0, 1\}} (\max \mathcal{R}(Q_i, \phi))$$

$$4. \min (\mathcal{R}(Q_1, \phi)) \quad \text{für} \quad \mathcal{R}(Q_0, \phi) \leq \alpha$$

Gemäß dem *Neyman-Pearson-Lemma* (vgl. Satz 2.1) können die Testprobleme 3. und 4. durch gewöhnliche LQ-Tests $\tilde{\phi}$ zwischen Q_0 und Q_1 gelöst werden (Huber, 1965, S.1754). Aufgrund von (38) und (39) werden durch denselben Test $\tilde{\phi}$ ebenso die Minimax-Probleme 1. und 2. gelöst. Dabei entspricht der LQ-Test zwischen Q_0 und Q_1 einer *gestutzten Version* des LQ-Tests zwischen P_0 und P_1 für das Testproblem (34). Die *gestutzte LQ-Statistik* ist robust im Minimax-Sinne und wird wie folgt definiert (Huber, 1965, S.1754):

Gestutzte LQ-Statistik für ϵ -Kontaminationsumgebungen

Für das Testproblem (34) seien gemäß einer möglichen Kontamination durch gross-error die zusammengesetzten Hypothesen (37) formuliert. Mit

$$p_i := \frac{dP_i}{d\mu} \quad \text{und} \quad q_i := \frac{dQ_i}{d\mu} \quad \text{für} \quad (i = 0, 1) \quad (40)$$

heißt

$$\Lambda_X^{ge}(\mathbf{x}) = \prod_{i=1}^n \max \left(s_u, \min \left(s_o, \frac{p_1(x_i)}{p_0(x_i)} \right) \right) = \prod_{i=1}^n \frac{q_1(x_i)}{q_0(x_i)} \quad (41)$$

gestutzte LQ-Statistik. Dabei gilt für $\epsilon_0, \epsilon_1 > 0$ und geeignetes $0 \leq s_u < s_o \leq \infty$:

$$q_0(x_i) = \begin{cases} (1 - \epsilon_0) p_0(x_i), & \text{falls } p_1(x_i)/p_0(x_i) < s_o \\ s_o^{-1} (1 - \epsilon_0) p_1(x_i), & \text{falls } p_1(x_i)/p_0(x_i) \geq s_o \end{cases} \quad (42)$$

$$q_1(x_i) = \begin{cases} (1 - \epsilon_1) p_1(x_i), & \text{falls } p_1(x_i)/p_0(x_i) > s_u \\ s_u (1 - \epsilon_1) p_0(x_i), & \text{falls } p_1(x_i)/p_0(x_i) \leq s_u \end{cases} \quad (43)$$

3. ROBUSTE MINIMAX-TESTS

Zudem folgt wie erwünscht:

$$Q_0 = \int_{\Omega} q_0 d\mu \stackrel{(42)}{=} \int_{p_1/p_0 < s_o} (1 - \epsilon_0) p_0 d\mu + \int_{p_1/p_0 \geq s_o} s_o^{-1} (1 - \epsilon_0) p_1 d\mu \quad (44)$$

$$= (1 - \epsilon_0) \int_{p_1/p_0 < s_o} p_0 d\mu + s_o^{-1} (1 - \epsilon_0) \int_{p_1/p_0 \geq s_o} p_1 d\mu \quad (45)$$

$$= (1 - \epsilon_0) P_{0_{[p_1/p_0 < s_o]}} + s_o^{-1} (1 - \epsilon_0) P_{1_{[p_1/p_0 \geq s_o]}} \quad (46)$$

$$\Rightarrow Q_0 \in \mathcal{P}_{\epsilon_0} \quad (47)$$

und analog

$$Q_1 = \int_{\Omega} q_1 d\mu \stackrel{(43)}{=} \int_{p_1/p_0 > s_u} (1 - \epsilon_1) p_1 d\mu + \int_{p_1/p_0 \leq s_u} s_u (1 - \epsilon_1) p_0 d\mu \quad (48)$$

$$= (1 - \epsilon_1) \int_{p_1/p_0 > s_u} p_1 d\mu + s_u (1 - \epsilon_1) \int_{p_1/p_0 \leq s_u} p_0 d\mu \quad (49)$$

$$= (1 - \epsilon_1) P_{1_{[p_1/p_0 > s_u]}} + s_u (1 - \epsilon_1) P_{0_{[p_1/p_0 \leq s_u]}} \quad (50)$$

$$\Rightarrow Q_1 \in \mathcal{P}_{\epsilon_1}. \quad (51)$$

Somit wird für jeden Faktor $p_1(x_i)/p_0(x_i)$ eine untere bzw. obere Schranke konstruiert, um die eingangs angesprochene Problematik über die fehlende Robustheit der LQ-Statistik zu beheben. Die Existenz solcher LFP $(Q_0, Q_1) \in \mathcal{P}_{\epsilon_0} \times \mathcal{P}_{\epsilon_1}$ mit Dichten der Form (42) und (43) wird durch folgenden Satz gestützt:

Satz 3.1 (Existenz und Eindeutigkeit)

Es existieren $0 \leq s_u < s_o \leq \infty$, sodass q_0, q_1 der Form (42) und (43) eindeutige Wahrscheinlichkeitsdichten sind, d.h.

$$\exists s_o \in \Omega : \int_{\Omega} q_0 d\mu = 1 \quad (52)$$

$$\stackrel{(46)}{\Leftrightarrow} \exists s_o \in \Omega : (1 - \epsilon_0) P_{0_{[p_1/p_0 < s_o]}} + s_o^{-1} (1 - \epsilon_0) P_{1_{[p_1/p_0 \geq s_o]}} = 1 \quad (53)$$

$$\exists s_u \in \Omega : \int_{\Omega} q_1 d\mu = 1 \quad (54)$$

$$\stackrel{(50)}{\Leftrightarrow} \exists s_u \in \Omega : (1 - \epsilon_1) P_{1_{[p_1/p_0 > s_u]}} + s_u (1 - \epsilon_1) P_{0_{[p_1/p_0 \leq s_u]}} = 1. \quad (55)$$

Für den Beweis reicht es, ausschließlich (55) zu zeigen, denn unter Permutation von P_0 und P_1 sowie der Schranken s_u und s_o^{-1} können (53) und (55) vertauscht werden (Huber, 1965, S.1755).

Beweis des Satzes 3.1

Mit (50) gilt

$$(1 - \epsilon_1)P_{1_{[p_1/p_0 > s_u]}} + s_u(1 - \epsilon_1)P_{0_{[p_1/p_0 \leq s_u]}} \stackrel{!}{=} 1 \quad (56)$$

$$\iff (1 - \epsilon_1) \left\{ P_{1_{[p_1/p_0 > s_u]}} + s_u P_{0_{[p_1/p_0 \leq s_u]}} \right\} \stackrel{!}{=} 1 \quad (57)$$

$$\iff P_{1_{[p_1/p_0 > s_u]}} + s_u P_{0_{[p_1/p_0 \leq s_u]}} \stackrel{!}{=} \frac{1}{1 - \epsilon_1} \quad (58)$$

Sei nun

$$f(s) := P_{1_{[p_1/p_0 > s]}} + s P_{0_{[p_1/p_0 \leq s]}} \quad (59)$$

als eine Funktion von s definiert. Und

$$s^* := \operatorname{vrai\,min}_{[\mu]} \frac{p_1(x)}{p_0(x)} = \operatorname{sup} \left\{ s \in \Omega : \mu \left[\frac{p_1(x)}{p_0(x)} < s \right] = 0 \right\} \quad (60)$$

das wesentliche Infimum von $\frac{p_1(x)}{p_0(x)}$ bzgl. μ . Für den Beweis des Satzes 3.1 reicht es, nun zu zeigen:

1. $\forall s \in \Omega, 0 \leq s \leq s^* : f(s) = 1$.
2. $f(s)$ ist stetig auf Ω .
3. $f(s)$ ist streng monoton steigend für $s > s^*$ und es gilt $\lim_{s \rightarrow +\infty} f(s) = +\infty$.

Sei $0 \leq s \leq s^*$, dann folgt

$$f(s) = P_{1_{[p_1/p_0 > s]}} + s P_{0_{[p_1/p_0 \leq s]}} \quad (61)$$

$$\iff f(s) = \int_{[p_1/p_0 > s]} p_1 d\mu + \int_{[p_1/p_0 \leq s]} s p_0 d\mu \quad (62)$$

$$\iff f(s) = 1 - \int_{[p_1/p_0 \leq s]} p_1 d\mu + s \int_{[p_1/p_0 \leq s]} p_0 d\mu \quad (63)$$

$$\iff f(s) = 1 + \int_{[p_1/p_0 \leq s]} (s - p_1/p_0) p_0 d\mu \quad (64)$$

$$\boxed{\stackrel{(60)}{\implies} f(s) = 1} \quad (65)$$

3. ROBUSTE MINIMAX-TESTS

Sei $s \in \Omega$ und $\delta > 0$, dann gilt

$$f(s + \delta) = 1 - \int_{[p_1/p_0 \leq s + \delta]} p_1 d\mu + s \int_{[p_1/p_0 \leq s + \delta]} (s + \delta) p_0 d\mu \quad (66)$$

$$\iff f(s + \delta) = 1 + \int_{[p_1/p_0 \leq s + \delta]} (s + \delta - p_1/p_0) p_0 d\mu \quad (67)$$

Damit folgt

$$f(s + \delta) - f(s) \stackrel{(64),(67)}{=} \left\{ 1 + \int_{[p_1/p_0 \leq s + \delta]} (s + \delta - p_1/p_0) p_0 d\mu \right\} - \quad (68)$$

$$\left\{ 1 + \int_{[p_1/p_0 \leq s]} (s - p_1/p_0) p_0 d\mu \right\} \quad (69)$$

$$\stackrel{(*)}{=} \int_{[s < p_1/p_0 \leq s + \delta]} (s + \delta - p_1/p_0) p_0 d\mu + \int_{p_1/p_0 \leq s} \delta p_0 d\mu \quad (70)$$

wobei (*)

$$\int_{[p_1/p_0 \leq s + \delta]} (s + \delta - p_1/p_0) p_0 d\mu = \int_{[s < p_1/p_0 \leq s + \delta]} (s + \delta - p_1/p_0) p_0 d\mu \quad (71)$$

$$+ \int_{[p_1/p_0 \leq s]} (s + \delta - p_1/p_0) p_0 d\mu \quad (72)$$

Insgesamt gilt dann für $|s + \delta - s| = \delta \leq 2\delta =: \epsilon$:

$$\boxed{0 \leq f(s + \delta) - f(s) \leq \epsilon} \quad (73)$$

Damit folgt 2. und formt man (73) um, so gilt $\forall \delta \geq 0$:

$$s + \delta > s \implies f(s + \delta) \geq f(s) \quad (74)$$

D.h. $f(s)$ ist monoton steigend. Für $s \geq s_1 > s^*$ gilt

$$f(s + \delta) - f(s) \geq \delta \int_{[p_1/p_0] \leq s_1} p_0 d\mu > 0 \quad (75)$$

Aus (73) folgt, dass $f(s)$ streng monoton steigend ist für $s > s^*$ und

$$\boxed{\lim_{s \rightarrow +\infty} f(s) = +\infty} \quad (76)$$

Damit ist der Beweis des Satzes 3.1 vollständig. □

Bemerkung 3.1

Laut (60) gilt in (73) $f(s + \delta) - f(s) = 0$ falls $s < s^*$.

Nach Satz 3.1 gilt somit, dass $\forall \epsilon_i : 0 \leq \epsilon_i < 1$ für (53) und (55) Lösungen existieren, diese jedoch nicht eindeutig sind, denn mit $\epsilon = 0$ gilt für (64):

$$1 + \int_{[p_1/p_0 \leq s]} (s - p_1/p_0) p_0 d\mu \stackrel{!}{=} 1 \quad (77)$$

$$\iff \int_{[p_1/p_0 \leq s]} (s - p_1/p_0) p_0 d\mu \stackrel{!}{=} 0 \quad (78)$$

Gleichung (78) hält aufgrund des wesentlichen Infimums für alle $s < s^*$ und ist somit lösbar, aber nicht eindeutig. Eindeutigkeit folgt, falls $0 < \epsilon_i < 1$ (Huber, 1965, S.1755). Mit dem Beweis über die Eindeutigkeit und Existenz von Wahrscheinlichkeitsdichten q_0, q_1 der Form (42) und (43) folgt die *gestutzte LQ-Statistik* mit $\gamma := \frac{1-\epsilon_1}{1-\epsilon_0}$ und $0 < s_u < s_o$ zu:

$$\Lambda_X^{ge}(x_i) = \frac{q_1(x_i)}{q_0(x_i)} = \begin{cases} \gamma s_u, & \text{falls } p_1(x_i)/p_0(x_i) \leq s_u \\ \gamma p_1(x_i)/p_0(x_i), & \text{falls } s_u < p_1(x_i)/p_0(x_i) < s_o \\ \gamma s_o, & \text{falls } p_1(x_i)/p_0(x_i) \geq s_o \end{cases} \quad (79)$$

Damit ist der LQ-Test zwischen $Q_0 \in \mathcal{P}_{\epsilon_0}$ und $Q_1 \in \mathcal{P}_{\epsilon_1}$ im Minimax-Sinne robust für zusammengesetzte Hypothesen formuliert in ϵ -Kontaminationsumgebungen. Bleibt noch zu zeigen, dass

$$\forall Q_i^* \in \mathcal{P}_{\epsilon_i} : \mathcal{R}(Q_i^*, \tilde{\phi}) \leq \mathcal{R}(Q_i, \tilde{\phi}), \quad (80)$$

wobei $\tilde{\phi}$ gewöhnlicher LQ-Test ist. Um (80) zu zeigen, muss zuvor die Gültigkeit folgenden Lemmas bewiesen werden:

Lemma 3.1

Für alle $Q_i^* \in \mathcal{P}_{\epsilon_i}$ mit $i \in \{0, 1\}$ und $k \in \mathbb{R}_0^+$ gilt für $s_u < s_o$:

$$Q_{0_{[q_1/q_0 < k]}}^* \geq Q_{0_{[q_1/q_0 < k]}} \quad (81)$$

$$Q_{1_{[q_1/q_0 < k]}} \geq Q_{1_{[q_1/q_0 < k]}}^* \quad (82)$$

Formuliert man (81) bzw. (82) um, so erhält man

$$\sup_{Q_0^* \in \mathcal{P}_{\epsilon_0}} Q_{0_{[q_1/q_0 > k]}}^* = Q_{0_{[q_1/q_0 > k]}} \quad (83)$$

$$\inf_{Q_1^* \in \mathcal{P}_{\epsilon_1}} Q_{1_{[q_1/q_0 > k]}}^* = Q_{1_{[q_1/q_0 > k]}} \quad (84)$$

3. ROBUSTE MINIMAX-TESTS

Somit werden die Fehlerwahrscheinlichkeiten für alle $(Q_0^*, Q_1^*) \in \mathcal{P}_{\epsilon_0} \times \mathcal{P}_{\epsilon_1}$ über die Fehlerwahrscheinlichkeiten der LFP (Q_0, Q_1) beschränkt. Damit entspricht der Neyman-Pearson-Test zwischen Q_0 und Q_1 robusten Minimax-Tests zwischen $Q_0^* \in \mathcal{P}_{\epsilon_0}$ und $Q_1^* \in \mathcal{P}_{\epsilon_1}$ (Huber, 1996, S.50).

Beweis Lemma 3.1

Für den Beweis von Lemma 3.1 müssen unterschiedliche Fälle gezeigt werden:

1.Fall: $k \leq \gamma s_u$

$$Q_{0_{[q_1/q_0 < k]}}^* = Q_{0_{[q_1/q_0 < k]}} = Q_{1_{[q_1/q_0 < k]}} = Q_{1_{[q_1/q_0 < k]}}^* = 0 \quad \checkmark \quad (85)$$

2.Fall: $k > \gamma s_o$

$$Q_{0_{[q_1/q_0 < k]}}^* = Q_{0_{[q_1/q_0 < k]}} = Q_{1_{[q_1/q_0 < k]}} = Q_{1_{[q_1/q_0 < k]}}^* = 1 \quad \checkmark \quad (86)$$

3.Fall: $s_u < k \leq \gamma s_o$

$$Q_{0_{[q_1/q_0 < k]}}^* \stackrel{(35)}{=} (1 - \epsilon_0)P_{0_{[q_1/q_0 < k]}} + \epsilon_0 \mathcal{H}_{0_{[q_1/q_0 < k]}} \quad (87)$$

$$\leq (1 - \epsilon_0)P_{0_{[q_1/q_0 < k]}} = Q_{0_{[q_1/q_0 < k]}} \quad \checkmark \quad (88)$$

$$Q_{1_{[q_1/q_0 < k]}}^* \stackrel{(36)}{=} (1 - \epsilon_1)P_{1_{[q_1/q_0 < k]}} + \epsilon_1 \mathcal{H}_{1_{[q_1/q_0 < k]}} \quad (89)$$

$$\leq (1 - \epsilon_1)P_{1_{[q_1/q_0 < k]}} + \epsilon_1 = Q_{1_{[q_1/q_0 < k]}} \quad \checkmark \quad (90)$$

Damit ist der Beweis vollständig. □

Korollar 3.1

Sei nun Z Zufallsvariable, dann gilt mit Lemma 3.1:

$$\exists f_0^*, f_0 : Q_{0_{[q_1/q_0 < k]}}^* = P(f_0^*(Z)) \leq P(f_0(Z)) = Q_{0_{[q_1/q_0 < k]}} \quad (91)$$

$$\exists f_1^*, f_1 : Q_{1_{[q_1/q_0 < k]}}^* = P(f_1^*(Z)) \leq P(f_1(Z)) = Q_{1_{[q_1/q_0 < k]}} \quad (92)$$

Insgesamt folgt nun mit $X = (X_1, \dots, X_n)$ und (Z_i, \dots, Z_n) i.i.d. und

$$\tilde{\phi}(X) = \begin{cases} 1, & \text{falls } q_1(X)/q_0(X) > k \\ \gamma(X), & \text{falls } q_1(X)/q_0(X) = k \\ 0, & \text{falls } q_1(X)/q_0(X) < k \end{cases} \quad (93)$$

LQ-Test zwischen Q_0 und Q_1 der Beweis für (80):

$$\mathcal{R}(Q_0^*, \tilde{\phi}) \stackrel{(13)}{=} \mathcal{L}_0 \mathbb{E}_{Q_0^*} \{ \phi(X) \} \quad (94)$$

$$\stackrel{(93)}{=} \mathcal{L}_0 \left\{ Q_{0^*_{[q_1/q_0 > k]}} + k Q_{0^*_{[q_1/q_0 = k]}} \right\} \quad (95)$$

$$= \mathcal{L}_0 \left\{ k Q_{0^*_{[q_1/q_0 \leq k]}} + (1 - k) Q_{0^*_{[q_1/q_0 > k]}} \right\} \quad (96)$$

$$\stackrel{(91)}{=} \mathcal{L}_0 \left\{ k P \left[\sum_i f_0^*(z_i) \geq k \right] + (1 - k) P \left[\sum_i f_0^*(z_i) > k \right] \right\} \quad (97)$$

$$\stackrel{(91)}{\leq} \mathcal{L}_0 \left\{ k P \left[\sum_i f_0(z_i) \geq k \right] + (1 - k) P \left[\sum_i f_0(z_i) > k \right] \right\} \quad (98)$$

$$= \mathcal{R}(Q_0, \tilde{\phi}) \quad (99)$$

Analog zeigt man die Gültigkeit von

$$\mathcal{R}(Q_1^*, \tilde{\phi}) \leq \mathcal{R}(Q_1, \tilde{\phi}). \quad (100)$$

Damit konnte gezeigt werden, dass eindeutig ein Paar $(Q_0, Q_1) \in \mathcal{P}_{\epsilon_0} \times \mathcal{P}_{\epsilon_1}$ existiert, sodass der gestutzte LQ-Test zwischen Q_0 und Q_1 robuste Minimax-Tests für alle $(Q_0^*, Q_1^*) \in \mathcal{P}_{\epsilon_0} \times \mathcal{P}_{\epsilon_1}$ darstellt. Da hierbei nur Unsicherheiten formalisiert in der ϵ -Kontaminationsumgebung berücksichtigt wurden, bemüht sich der nächste Abschnitt um eine Darstellung des LQ-Testes für Hypothesen, formuliert im Total-Variations Umgebungsmodell.

3.3.2 Total-Variations Umgebungen

Seien nun die einfachen Hypothesen (34) erneut als zusammengesetzte Hypothesen

$$H_0 : P \in \mathcal{P}_{tv_0} \quad \text{gegen} \quad H_1 : P \in \mathcal{P}_{tv_1} \quad (101)$$

formuliert, wobei es sich bei \mathcal{P}_{tv_0} und \mathcal{P}_{tv_1} jeweils um die Total-Variations Umgebung der Null-

$$\mathcal{P}_{tv_0} = \{ Q \in \mathcal{W} \mid \sup(Q - P_0) \leq \epsilon_0 \} \quad (0 \leq \epsilon_0 < 1) \quad (102)$$

bzw. Gegenhypothese

$$\mathcal{P}_{tv_1} = \{ Q \in \mathcal{W} \mid \sup(Q - P_1) \leq \epsilon_1 \} \quad (0 \leq \epsilon_1 < 1) \quad (103)$$

handelt. Analog zum gestutzten LQ-Test für die ϵ -Kontaminationsumgebung soll nun auch für Testprobleme formuliert in Total-Variations Umgebungen, gezeigt

3. ROBUSTE MINIMAX-TESTS

werden, dass ein Paar $(Q_0, Q_1) \in \mathcal{P}_{\epsilon_0} \times \mathcal{P}_{\epsilon_1}$ existiert, sodass es sich beim LQ-Test zwischen Q_0 und Q_1 um eine gestutzte Version des LQ-Testes zwischen P_0 und P_1 handelt, der für alle $(Q_0^*, Q_1^*) \in \mathcal{P}_{tv_0} \times \mathcal{P}_{tv_1}$ im Minimax-Sinne robust ist. Dabei sind Ausgangsproblematik und Minimax-Testprobleme identisch zum Fall der ϵ -Kontamination und sollen hier nicht erneut aufgeführt werden.

Gestutzte LQ-Statistik für Total-Variations Umgebungen

Sei (101) Testproblem gemäß einer möglichen Kontamination durch gross-error und Modellfehler gegeben. Dann heißt

$$\Lambda_X^{ge}(\mathbf{x}) = \prod_{i=1}^n \max \left(s_u, \min \left(s_o, \frac{p_1(x_i)}{p_0(x_i)} \right) \right) = \prod_{i=1}^n \frac{q_1(x_i)}{q_0(x_i)} \quad (104)$$

gestutzte LQ-Statistik für Total-Variations Umgebungen, wobei für $\epsilon_0, \epsilon_1 > 0$ und $0 \leq s_u < s_o \leq \infty$ gilt:

$$q_0(x_i) = \begin{cases} (1 + s_u)^{-1}(p_0(x_i) + p_1(x_i)), & \text{falls } p_1(x_i)/p_0(x_i) \leq s_o \\ p_0(x_i), & \text{falls } s_u < p_1(x_i)/p_0(x_i) < s_o \\ (1 + s_o)^{-1}(p_0(x_i) + p_1(x_i)), & \text{falls } p_1(x_i)/p_0(x_i) \geq s_o \end{cases} \quad (105)$$

$$q_1(x_i) = \begin{cases} [s_u/(1 + s_u)](p_0(x_i) + p_1(x_i)), & \text{falls } p_1(x_i)/p_0(x_i) \leq s_o \\ p_1(x_i), & \text{falls } s_u < p_1(x_i)/p_0(x_i) < s_o \\ [s_o/(1 + s_o)](p_0(x_i) + p_1(x_i)), & \text{falls } p_1(x_i)/p_0(x_i) \geq s_o \end{cases} \quad (106)$$

Nun soll mit Satz 3.2 gezeigt werden, dass Dichten der Form (105) bzw. (106) eindeutig existieren, sodass $Q_0 \in \mathcal{P}_{tv_0}$ und $Q_1 \in \mathcal{P}_{tv_1}$ ist.

Satz 3.2 (Existenz und Eindeutigkeit)

Es existieren $0 \leq s_u < s_o \leq \infty$, sodass q_0, q_1 der Form (105) und (106) eindeutige Wahrscheinlichkeitsdichten sind mit $Q_0 \in \mathcal{P}_{tv_0}$ und $Q_1 \in \mathcal{P}_{tv_1}$, somit fordert man

$$\exists s_u, s_o \in \Omega : \sup_{A \in \mathcal{X}} \{Q_0(A) - P_0(A)\} = \epsilon \quad \text{und} \quad \int q_0 d\mu = 1, \quad (107)$$

und analog

$$\exists s_u, s_o \in \Omega : \sup_{A \in \mathcal{X}} \{Q_1(A) - P_1(A)\} = \epsilon \quad \text{und} \quad \int q_1 d\mu = 1. \quad (108)$$

3. ROBUSTE MINIMAX-TESTS

Nach (Huber, 1965, S.1758) lassen sich die Bedingungen (107) bzw. (108) jeweils zusammenfassend zu

$$\int_{[p_1/p_0 \geq s_o]} (q_0 - p_0) d\mu = \frac{\epsilon}{2} \quad \text{und} \quad \int_{[p_1/p_0 \leq s_u]} (q_1 - p_1) d\mu = \frac{\epsilon}{2} \quad (109)$$

formulieren, denn für die rechte Gleichung in (109) folgt

$$\begin{aligned} d_{tv}(Q_1 - P_1) &\stackrel{(20)}{=} \int |q_1 - p_1| d\mu & (110) \\ &= \underbrace{\int_{[p_1/p_0 \leq s_u]} q_1 - p_1 d\mu}_{\stackrel{(109)}{=} \frac{\epsilon}{2}} + \int_{[s_u < p_1/p_0 < s_o]} 0 d\mu + \underbrace{\int_{[p_1/p_0 \geq s_o]} p_1 - q_1 d\mu}_{\stackrel{(*)}{=} \int_{[p_1/p_0 \geq s_o]} q_0 - p_0 d\mu} \\ & & \stackrel{(109)}{=} \frac{\epsilon}{2} \\ &= \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon & (111) \end{aligned}$$

und

$$\int q_1 d\mu = \int q_1 - p_1 d\mu + \underbrace{\int p_1 d\mu}_{=1} \quad (112)$$

$$= \int_{[p_1/p_0 \leq s_u]} q_1 - p_1 d\mu + \int_{[s_u < p_1/p_0 < s_o]} 0 d\mu + \int_{p_1/p_0 \geq s_o} q_1 - p_1 d\mu + 1 \quad (113)$$

$$= \frac{\epsilon}{2} + 0 - \frac{\epsilon}{2} + 1 \quad (114)$$

$$= 1, \quad (115)$$

wobei für $p_1/p_0 \geq s_o$ mit (*) gilt

$$p_1 - q_1 \stackrel{(106)}{=} p_1 - \frac{s_o}{1 + s_o} (p_0 + p_1) = \frac{s_o}{1 + s_o} p_0 + \frac{1}{1 + s_o} p_1 \quad (116)$$

$$= \frac{1 - (1 + s_o)}{1 + s_o} p_0 + \frac{1}{1 + s_o} p_1 = \frac{1}{1 + s_o} (p_0 + p_1) - p_0 \quad (117)$$

$$= q_0 - p_0. \quad (118)$$

Da beide Gleichungen in (109) analog formuliert sind, reicht es lediglich eins von beiden zu zeigen. Somit konzentriert sich der Beweis des Satzes 3.2 auf die Gültigkeit der rechten Seite von (109):

Beweis des Satzes 3.2

Unter Ausnutzung der Dichte (106) folgt

$$\int_{[p_1/p_0 \leq s_u]} (q_1 - p_1) d\mu = \frac{\epsilon}{2} \quad (119)$$

$$\stackrel{(106)}{\Leftrightarrow} \int_{[p_1/p_0 \leq s_u]} [s_u/(1 + s_u)] (p_0 + p_1) - p_1 d\mu = \frac{\epsilon}{2}. \quad (120)$$

Für den Beweis definiert man die Funktion

$$h(\tau) \stackrel{(*)}{:=} \int_{[p_1 \leq (p_0 + p_1)\tau]} \tau (p_0 + p_1) - p_1 d\mu, \quad (121)$$

mit

$$\tau = \frac{s_u}{1 + s_u}. \quad (122)$$

Sei wiederum $\tau_0 := \text{vrai min}_{[\mu]} \frac{p_1}{p_0 + p_1}$ das wesentliche Infimum von $\frac{p_1}{p_0 + p_1}$ bzgl. μ , dann reicht es Folgendes zu zeigen:

1. $h(\tau)$ ist stetig auf Ω
2. $h(\tau) = 0$ für $0 \leq \tau \leq \tau_0$
3. $h(\tau)$ ist für $\tau_0 < \tau < 1$ streng monoton steigend und es gilt $g(1) = 1$

Sei $\tau \in \Omega$ und $\delta \geq 0$, dann gilt

$$h(\tau + \delta) - h(\tau) = \int_{p_1 \leq (p_0 + p_1)(\tau + \delta)} [(\tau + \delta)(p_0 + p_1) - p_1] d\mu \quad (123)$$

$$\begin{aligned} & - \int_{p_1 \leq (p_0 + p_1)\tau} [\tau(p_0 + p_1) - p_1] d\mu \\ & \stackrel{(*)}{=} \int_{(p_0 + p_1)\tau \leq p_1 \leq (p_0 + p_1)(\tau + \delta)} [(\tau + \delta)(p_0 + p_1) - p_1] d\mu \quad (124) \\ & - \delta \int_{p_1 \leq (p_0 + p_1)\tau} (p_0 + p_1) d\mu, \end{aligned}$$

wobei für (*) folgt

$$\int_{p_1 \leq (p_0 + p_1)(\tau + \delta)} [(\tau + \delta)(p_0 + p_1) - p_1] d\mu \quad (125)$$

$$\begin{aligned} & = \int_{(p_0 + p_1)\tau < p_1 \leq (p_0 + p_1)(\tau + \delta)} [(\tau + \delta)(p_0 + p_1) - p_1] d\mu \quad (126) \\ & + \int_{p_1 \leq (p_0 + p_1)\tau} [(\tau + \delta)(p_0 + p_1) - p_1] d\mu. \end{aligned}$$

3. ROBUSTE MINIMAX-TESTS

Mit (124) gilt dann für $|\tau + \delta - \tau| = \delta \leq \epsilon := 2\delta$:

$$\boxed{0 \leq h(\tau + \delta) - h(\tau) \leq 2\delta = \epsilon} \quad (127)$$

Damit ist 1. erfüllt und somit $g(\tau)$ stetig auf Ω . Mit $0 \leq \tau \leq \tau_0$ gilt trivialerweise

$$h(\tau) = \int_{[p_1/(p_0+p_1) \leq \tau]} \tau (p_0 + p_1) - p_1 d\mu \stackrel{\tau \leq \tau_0}{\geq} 0, \quad (128)$$

und somit ist auch 2. erfüllt. Für $1 > \tau > \tau_0$ folgt die strenge Monotonie durch

$$0 \leq h(\tau + \delta) - h(\tau) \leq \delta \int_{[p_1/(p_0+p_1) \leq \tau]} (p_0 + p_1) > 0 \quad (129)$$

$$\Rightarrow h(\tau + \delta) > h(\tau). \quad (130)$$

Zudem gilt für $\tau = 1$

$$h(1) = \int_{[p_1/(p_0+p_1) \leq 1]} [(p_0 + p_1) - p_1] d\mu \quad (131)$$

$$= \int_{[0 \leq p_0]} p_0 d\mu = 1 \quad (132)$$

und damit ist der Beweis des Satzes 3.2 vollständig. \square

Bemerkung 3.2

Die neue Integralgrenze in (121) ergibt sich direkt durch die Umformung (122). Weiterhin folgt die Ungleichung (127) direkt aus der Gleichung (124), denn liegt die ganze Wahrscheinlichkeitsmasse auf $p_1 \leq (p_0 + p_1)\tau$, so folgt für (124):

$$0 + \delta \left(\int_{[p_1/(p_0+p_1) \leq \tau]} p_0 d\mu + \int_{[p_1/(p_0+p_1) \leq \tau]} p_1 d\mu \right) = 2\delta \quad (133)$$

Mit dem Beweis des Satzes 3.2 konnte nun gezeigt werden, dass für alle $0 \leq \epsilon < 2s_u, s_o$ existieren, sodass es sich bei $(q_0, q_1) \in \mathcal{P}_{tv_0} \times \mathcal{P}_{tv_1}$ um Dichten der Form (105) bzw. (106) handelt. Eindeutigkeit folgt, falls ϵ tatsächlich größer null ist (Huber, 1981, S.1758). Des Weiteren kann durch analoges Vorgehen wie für die ϵ -Kontaminationsumgebung gezeigt werden, dass es sich bei den konstruierten $(Q_0, Q_1) \in \mathcal{P}_{tv_0} \times \mathcal{P}_{tv_1}$ um LFP für das zusammengesetzte Testproblem (101) handelt, sodass der LQ-Test zwischen Q_0 und Q_1 im Minimax-Sinne robust ist und wiederum einer gestutzten Version des LQ-Testes zwischen P_0 und P_1 entspricht.

Nachdem nun robuste Minimax-Tests sowohl für die ϵ -Kontaminations-, als auch für die Total-Variations-Umgebung vorgestellt wurden, soll im Nachfolgenden eine Generalisierung der bisher erzielten Resultate erfolgen. Dabei ist die Arbeit von Huber und Strassen (1973), die anhand des, nach den Autoren benannten *Huber-Strassen-Theorems* das Neyman-Pearson-Lemma auf sogenannte *bialternierende Kapazitäten* in allgemeinen polnischen Räumen generalisiert, fundamental für robuste statistische Testverfahren und kann als einer der wichtigsten Resultate in diesem Gebiet betrachtet werden.

3.4 Generalisierung des Neyman-Pearson-Lemmas

3.4.1 Kapazitäten und Umgebungsmodelle

Bevor das *Huber-Strassen-Theorem* und damit robuste Minimax-Tests für allgemeinere Umgebungen dargestellt werden, sollen zuvor, für das Verständnis der Arbeit von Huber und Strassen (1973) essentielle Begrifflichkeiten näher erläutert werden. Hierbei weicht man von der klassischen Definition der Wahrscheinlichkeit als einzelne feste Größe ab und beschäftigt sich als Ausdruck von *Unsicherheit* mit *Oberen- und Unteren-Wahrscheinlichkeiten*.

Definition 3.1 (Obere- und Untere-Wahrscheinlichkeiten)

Sei (Ω, \mathcal{X}) Messraum, wobei es sich bei Ω um einen beliebigen polnischen Raum und bei \mathcal{X} um die Borelsche- σ -Algebra auf Ω handelt. Weiterhin bezeichne \mathcal{W} die Menge aller Wahrscheinlichkeitsmaße auf (Ω, \mathcal{X}) . Für $\mathcal{P} \subset \mathcal{W}$ heißt

$$o(A) := \sup_{P \in \mathcal{P}} P(A) \quad \text{für alle } A \in \mathcal{X} \quad (134)$$

Obere-Wahrscheinlichkeit, und

$$u(A) := \inf_{P \in \mathcal{P}} P(A) \quad \text{für alle } A \in \mathcal{X} \quad (135)$$

Untere-Wahrscheinlichkeit von \mathcal{P} . Zudem gilt

$$u(A) = 1 - \sup_{P \in \mathcal{P}} P(A^c) = 1 - o(A^c), \quad (136)$$

und somit

$$u(A) + o(A^c) = 1. \quad (137)$$

Damit stellen Obere- bzw. Untere-Wahrscheinlichkeiten im Gegensatz zu der klassischen Wahrscheinlichkeit jeweils supremale bzw. infimale Schranken von \mathcal{P} dar. Über Obere-bzw. Untere-Wahrscheinlichkeiten gelangt man zur Definition von Kapazitäten:

Definition 3.2 (Kapazitäten)

Sei (Ω, \mathcal{X}) erneut Messraum, dann nennt man eine Mengenfunktion $o_k : \mathcal{X} \rightarrow [0, 1]$ Kapazität, falls folgende Bedingungen erfüllt sind:

1. $o_k(\emptyset) = 0, o_k(\Omega) = 1$ (Normiertheit)
2. $\forall A, B \in \mathcal{X} : A \subset B \Rightarrow o_k(A) \leq o_k(B)$ (Monotonie)
3. $(A_n)_{n \in \mathbb{N}} \uparrow A \Rightarrow o_k(A_n) \uparrow o_k(A)$ (Stetigkeit von unten)

3. ROBUSTE MINIMAX-TESTS

Zudem muss für schwach-folgenkompakte^{A.2} \mathcal{P} gelten:

$$4. (B_n)_{n \in \mathbb{N}} \downarrow B, B_n \text{ abgeschlossen}, n \in \mathbb{N} \Rightarrow o_k(B_n) \downarrow o_k(B)$$

Man spricht von bialternierenden Kapazitäten, falls neben 1.-4. zudem

$$5. \forall A, B \in \mathcal{X} : o_k(A \cup B) + o_k(A \cap B) \leq o_k(A) + o_k(B)$$

gilt. Die fünfte Forderung wird auch (starke) Sub-Additivität genannt.

Eine Mengenfunktion $u_k : \mathcal{X} \rightarrow [0, 1]$, die bezüglich o_k gemäß (136) konjugiert ist, nennt man zu u_k duale Kapazität, falls Folgendes gilt:

$$1^* u_k(\emptyset) = 0, u_k(\Omega) = 1 \text{ (Normiertheit)}$$

$$2^* \forall A, B \in \mathcal{X} : A \subset B \Rightarrow u_k(A) \leq u_k(B) \text{ (Monotonie)}$$

$$3^* n \in \mathbb{N} : A_n \downarrow A \rightarrow u_k(A_n) \downarrow u_k(A) \text{ (Stetigkeit von oben)}$$

Für schwach-folgenkompakte \mathcal{P} muss gelten:

$$4^* C_n \text{ offen}, C_n \uparrow C \rightarrow u_k(C_n) \uparrow u_k(C)$$

Man spricht von bimonotonen Kapazitäten, falls neben 1.*-4.* zudem

$$5^* u_k(A \cup B) + u_k(A \cap B) \geq u_k(A) + u_k(B)$$

gilt. Hierbei nennt man die fünfte Forderung (starke) Super-Additivität⁶.

An dieser Stelle sei darauf hingewiesen, dass es sich bei der Definition 3.2 konkret um *Choquet-Kapazitäten* handelt, die erstmals in Choquet (1954) diskutiert werden. Auf eine nähere Darstellung der Eigenschaften von Kapazitäten, die sowohl in Choquet (1954), als auch Huber (1973) diskutiert werden, soll für die Zweckmäßigkeit dieser Arbeit verzichtet werden. Nun soll gezeigt werden, dass alle in *Abschnitt 3.1* dargestellten Umgebungsmodelle sich anhand bialternierenden Kapazitäten formulieren lassen. Für Zwecke der Veranschaulichung wird dies explizit für die ϵ -Kontaminations- und Total-Variations-Umgebung gezeigt. Allgemein definiert man hierfür Umgebungen der Form

$$\boxed{\mathcal{P}_k = \{P \in \mathcal{W} \mid P(A) \leq o_k(A) \forall A \in \mathcal{X}\}} \quad (138)$$

wobei es sich bei \mathcal{X} um die Borelsche- σ -Algebra auf einen kompakten Raum Ω handelt. Wiederum stellt \mathcal{W} die Menge aller möglichen Wahrscheinlichkeitsmaße auf

⁶Bei der Definition der Sub- bzw. Super-Additivität ist zu beachten, dass diese Bezeichnungen in der Literatur nicht einheitlich benutzt werden und ggf. zu Missverständlichkeit führen können.

3. ROBUSTE MINIMAX-TESTS

(Ω, \mathcal{X}) dar. Wie man später feststellen wird, ist die Forderung der Kompaktheit des Raums Ω notwendig und hinreichend, um Umgebungsmodelle über bialternierende Kapazitäten darstellen zu können (Huber, 1973, S.253). Nach Gleichung (138) besteht die Menge \mathcal{P}_k aus allen Wahrscheinlichkeitsmaßen, die durch die bialternierende Kapazität o_k majorisiert werden (Huber, 1973, S.251). Weiterhin gilt, dass o_k im Sinne der Definition 3.1 eine Obere-Wahrscheinlichkeit der Menge \mathcal{P}_k darstellt. Anhand von (138) lassen sich nun ϵ -Kontaminations- und Total-Variation Umgebungen über bialternierende Kapazitäten formulieren.

ϵ -Kontaminations-Kapazitäten

Sei (Ω, \mathcal{X}) Messraum, wobei Ω kompakt ist. Betrachtet man das Testproblem

$$H_0 : P \in \mathcal{P}_{\epsilon_0} \quad \text{gegen} \quad H_1 : P \in \mathcal{P}_{\epsilon_1} \quad (139)$$

so ist

$$o_{\epsilon_i}(A) := \begin{cases} \min((1 - \epsilon_i)P_i(A) + \epsilon_i, 1), & \text{falls } A \neq \emptyset \\ 0, & \text{falls } A = \emptyset \end{cases} \quad (140)$$

eine bialternierende Kapazität für $i \in \{0, 1\}$. Weiterhin gilt mit (136), dass

$$u_{\epsilon_i}(A) := \begin{cases} \max((1 - \epsilon_i)P_i(A), 0), & \text{falls } A \neq \Omega \\ 1, & \text{falls } A = \Omega \end{cases} \quad (141)$$

eine bimonotone Kapazität für $i \in \{0, 1\}$ ist. Anhand der Umgebungen

$$\tilde{\mathcal{P}}_{\epsilon_i} = \{P \in \mathcal{W} | P \leq o_{\epsilon_i}\} = \{P \in \mathcal{W} | P \geq u_{\epsilon_i}\} \quad (142)$$

kann man nun die Hypothesen (139) über bialternierenden Kapazitäten o_{ϵ_0} und o_{ϵ_1} zu

$$\boxed{H_0 : P \in \tilde{\mathcal{P}}_{\epsilon_0} \quad \text{gegen} \quad H_1 : P \in \tilde{\mathcal{P}}_{\epsilon_1}} \quad (143)$$

formulieren.

Es soll nun exemplarisch für o_{ϵ_i} ($i = 0, 1$) gezeigt werden, dass es sich hierbei tatsächlich um eine bialternierende Kapazität handelt:

Da es sich bei o_{ϵ_i} leicht ersichtlich um Obere-Wahrscheinlichkeiten nach Definition 3.1 handelt, sind die Forderungen 1. bis 4. sofort erfüllt. Bleibt noch die Sub-Additivitäts Eigenschaft zu zeigen. Dafür sei die Funktion

$$f(x) := (1 - \epsilon)x + \epsilon \quad (144)$$

3. ROBUSTE MINIMAX-TESTS

definiert. Trivialerweise ist die Funktion f per Definition konkav und für $w \leq x, y \leq z$ und $w + x = y + z$ gilt

$$f(w) + f(x) \leq f(y) + f(z). \quad (145)$$

Über die Definition von o_{ϵ_i} gilt somit, dass es sich um eine bialternierende Kapazität handelt. In analoger Weise zeigt man Super-Additivitäts Eigenschaften von dualen Kapazitäten (Rüschendorf, 2014, S.309).

Abschließend soll auch für die Total-Variations-Umgebung die Majorisierung über bialternierende Kapazitäten veranschaulicht werden.

Total-Variations-Kapazitäten

Sei (Ω, \mathcal{X}) Messraum, wobei Ω kompakt ist. Betrachtet man das Testproblem

$$H_0 : P \in \mathcal{P}_{tv_0} \quad \text{gegen} \quad H_1 : P \in \mathcal{P}_{tv_1} \quad (146)$$

formuliert in der Total-Variations-Umgebung, so sind

$$o_{tv_i}(A) := \begin{cases} \min(P_i(A) + \epsilon_i, 1), & \text{falls } A \neq \emptyset \\ 0, & \text{falls } A = \emptyset \end{cases} \quad (147)$$

und

$$u_{tv_i}(A) := \begin{cases} \max(P_i(A) + \epsilon, 0), & \text{falls } A \neq \Omega \\ 1, & \text{falls } A = \Omega \end{cases} \quad (148)$$

bialternierende bzw. bimonotone Kapazitäten ($i = 0, 1$). Somit lassen sich Umgebungen formalisiert über die Total-Variation über

$$\tilde{\mathcal{P}}_{tv_i} = \{P \in \mathcal{W} | P \leq o_{tv_i}\} = \{P \in \mathcal{W} | P \geq u_{tv_i}\} \quad (149)$$

ausdrücken. Und das Testproblem anhand der bialternierenden Kapazitäten o_{tv_0} und o_{tv_1} ergibt sich zu

$$\boxed{H_0 : P \in \tilde{\mathcal{P}}_{tv_0} \quad \text{gegen} \quad H_1 : P \in \tilde{\mathcal{P}}_{tv_1}.} \quad (150)$$

Obwohl hier nur für die zwei gängigsten Umgebungsmodelle in der robusten Statistik die Darstellung über bialternierenden Kapazitäten veranschaulicht wurde, können alle in *Abschnitt 3.1* vorgestellten Umgebungsmodelle (und weitere die in dieser Arbeit nicht erwähnt wurden) über bialternierende Kapazitäten formuliert werden.

Bemerkung 3.3 (Kompakter Raum Ω)

Wie bereits angesprochen ist die Forderung eines kompakten Raumes Ω für die Darstellung der Umgebungsmodelle über solche speziellen Kapazitäten notwendig, da die Stetigkeitsargumente 4. bzw. 4.* nur dann erfüllt sind, falls Ω kompakt ist (Rieder, 1977, S.912). Im Anschluss auf die Darstellung des Huber-Strassen-Theorems wird ausblickend über den Umgang mit Fällen von nicht-kompakten Ω berichtet.

Nachdem nun in nötige Begrifflichkeiten eingeführt wurde, soll im folgenden Abschnitt angelehnt an Huber (1973, S.251-S.263) das Huber-Strassen-Theorem dargestellt werden. Dabei liegt der Fokus mehr auf die Veranschaulichung und Verständlichkeit des Theorems als auf technische Nachvollziehbarkeit.

3.4.2 Huber-Strassen-Theorem

Sei nun das Testproblem für die Menge aller Wahrscheinlichkeitsmaße $\tilde{\mathcal{P}}_{k_0}$ und $\tilde{\mathcal{P}}_{k_1}$, die gemäß *Abschnitt 3.4.1* über bialternierende Kapazitäten o_{k_0} bzw. o_{k_1} majorisiert werden, formuliert:

$$H_0 : \tilde{\mathcal{P}}_{k_0} = \{P \in \mathcal{W} | P \leq o_{k_0}\} \quad \text{gegen} \quad H_1 : \tilde{\mathcal{P}}_{k_1} = \{P \in \mathcal{W} | P \leq o_{k_1}\} \quad (151)$$

Dabei soll erneut angemerkt werden, dass alle in *Abschnitt 3.1* dargestellten Umgebungsmodelle über bialternierenden Kapazitäten o_{k_0} und o_{k_1} formuliert werden können, sodass (151) nun eine Generalisierung der Testprobleme (37), sowie (101) darstellt. Das Huber-Strassen-Theorem (Huber, 1973, S.257, Theorem 4.1) formuliert sich somit zu:

Huber-Strassen-Theorem

Für das Testproblem (151) existieren LFP $(Q_0, Q_1) \in \tilde{\mathcal{P}}_{k_0} \times \tilde{\mathcal{P}}_{k_1}$ mit $Q_0 \leq o_{k_0}$ und $Q_1 \leq o_{k_1}$, sodass für $k > 0$

$$Q_0 [\pi > k] = o_{k_0} [\pi > k] \quad (152)$$

$$Q_1 [\pi > k] = u_{k_1} [\pi > k] \quad (153)$$

gilt, wobei $\pi = \frac{dQ_1}{dQ_0}$ bzgl. einem dominierendem Maß μ ist.

Bemerkung 3.4

Das Huber-Strassen-Theorem lässt sich aufgrund von (136) ebenso über bimonotone Kapazitäten formulieren. Um sich an Huber (1973) zu halten, wurde das Theorem hier über bialternierende Kapazitäten formuliert. Des Weiteren wird auf den Beweis des Theorems aus Komplexitäts-Gründen nicht weiter eingegangen, dieser kann jedoch in Huber (1973, S.258-S.260) nachgelesen werden.

3. ROBUSTE MINIMAX-TESTS

Nach dem Huber-Strassen-Theorem gilt nun für alle $(Q_0^*, Q_1^*) \in \tilde{\mathcal{P}}_{k_0} \times \tilde{\mathcal{P}}_{k_1}$ mit $\alpha, \alpha^* \in (0, 1)$ und $\beta, \beta^* \in (0, 1)$

$$\alpha = o_{k_0}(\pi > k) = Q_0(\pi > k) \geq Q_0^*(\pi > k) =: \alpha^* \quad (154)$$

$$\beta = u_{k_1}(\pi > k) = Q_1(\pi > k) \leq Q_1^*(\pi > k) =: \beta^* \quad (155)$$

wobei $k > 0$ ist. Damit wird über die Existenz solcher LFP garantiert, dass der LQ-Test zwischen Q_0 und Q_1 ebenso ein Minimax-Test für alle $(Q_0^*, Q_1^*) \in \tilde{\mathcal{P}}_{k_0} \times \tilde{\mathcal{P}}_{k_1}$ ist. Ähnliche Resultate wurden bereits in den *Abschnitten 3.3.1* und *3.3.2* für Hypothesen in ϵ -Kontaminations- und Total-Variations-Umgebungen erzielt. Leicht ersichtlich erhält man durch das Huber-Strassen Theorem dieselbe Minimax-Aussage für eine weitaus größere Klasse von Verteilungen (d.h. Menge von Verteilungen, die über bialternierende Kapazitäten majorisiert werden). Weiterhin wird durch das folgende Lemma die Gleichheit in (152) bzw. (153) garantiert. Der Beweis dazu findet sich in Huber (1973, S.254) und wird hier nicht explizit behandelt.

Lemma 3.2

Sei $\mathcal{P}_k = \{P \in \mathcal{W} | P(A) \leq o_k(A) \forall A \in \mathcal{X}\}$ Umgebung, wobei o_k eine bialternierende Kapazität ist. Dann folgt, dass für alle $A \in \mathcal{X}$ ein $Q \in \mathcal{P}_k$ existiert, sodass

$$Q(A) = o_k(A) \quad (156)$$

gilt.

Es sei angemerkt, dass der Neyman-Pearson Test zwischen Q_0 und Q_1 sowohl vom Stichprobenumfang n , als auch vom Niveau α unabhängig ist (Huber, 1973, S.257).

Welche besondere Bedeutung das Huber-Strassen-Theorem für die weitere Entwicklung der robusten Statistik hatte, soll nun im folgenden Abschnitt näher verdeutlicht werden.

3.5 Bedeutung des Huber-Strassen-Theorems

Mit Pionierarbeiten von Strassen (1964) und Huber (1965) lässt sich die Entwicklung von robusten Minimax-Tests grob in zwei Stränge unterteilen, die letztendlich 1973 zum Huber-Strassen-Theorem führten. Wie bereits in *Abschnitt 3.3* dargestellt, beweist Huber (1965) die Existenz von LFP in ϵ -Kontaminations- und Total-Variations-Umgebungen. Dahingegen konnte die Aussage des Huber-Strassen-Theorems von Strassen (1964) bereits zuvor für sogenannte *totalalternierende* Kapazitäten auf endlichen Räumen gezeigt werden. Diese beiden fundamentalen Arbeiten führten letztendlich 1973 zum Huber-Strassen-Theorem, der über bialternierende Kapazitäten für allgemeine polnische Räume formuliert wurde. Im Gegensatz zu totalalternierenden Kapazitäten, lassen sich alle gängigen

Umgebungsmodelle über bialternierende Kapazitäten darstellen, jedoch mit der Einschränkung, dass es sich um einen kompakten Grundraum handeln muss. Dahingehend beschäftigte sich die weitere Forschung einerseits mit der Erweiterung des Huber-Strassen-Theorems auf nicht-kompakte Räume und andererseits mit Konstruktionsmöglichkeiten von LFP, da das Huber-Strassen-Theorem lediglich einen Existenzbeweis darstellt. Rieder (1977) und hierauf Bednarski (1981) zeigten die Existenz für LFP in Kontaminationsumgebungen über spezielle Kapazitäten, die nicht an die Stetigkeitsargumente 4. bzw. 4.* gebunden sind und somit auch in nicht-kompakten Räumen definiert werden können. Letztendlich gelang es Buja (1986) über das *Kuratowski-Isomorphismus Theorem* die Aussage des Huber-Strassen-Theorems zu generalisieren. Indessen beschäftigte sich Österreicher (1978) mit der Konstruktion von LFP für verschiedene Umgebungsmodelle. Dahingehend konnte Hafner (1982a) und Hafner (1982b) über generalisierte Risikofunktionen LFP für Kontaminations- bzw. Prohorov-Umgebungen konstruieren. Konstruktionsmöglichkeiten über Risikofunktionen werden in Hafner (1992) anschaulich erklärt. Aufbauend auf Konzepte von sogenannten *Intervallwahrscheinlichkeiten* (Weichselberger, 2001), formulierte Augustin (1999) grundlegende Aussagen des Huber-Strassen-Theorems im Rahmen von Intervallwahrscheinlichkeiten.

Es sei besonders darauf hingewiesen, dass weitaus mehr Arbeiten, die im Kontext dieser Arbeit nicht behandelt werden, sich im Zuge des Huber-Strassen-Theorems mit robusten Minimax-Tests beschäftigten. Trotz des sehr theoretischen Charakters der vorliegenden Arbeiten, soll sich der folgende Abschnitt mit einer praktischen Validierung von bisher erzielten Resultaten beschäftigen, sodass auch visuell festgehalten werden kann, wieso solche robusten Resultate eine wichtige Rolle in der Statistik spielen.

3.6 Simulation

Nachdem in vorherigen Abschnitten lediglich theoretische Aspekte von gestutzten LQ-Tests behandelt wurden, sollen nun bisher festgehaltene (theoretische) Resultate anhand von Simulationen veranschaulicht werden. Seien hierzu

$$X_i \stackrel{u.i.v.}{\sim} \mathcal{N}(\mu, \sigma^2) \quad (i = 1, \dots, n) \quad (157)$$

gegeben. Man interessiert sich nun für das folgende Testproblem:

$$H_0 : \mu = 0 \quad \text{gegen} \quad H_1 : \mu \neq 0 \quad (158)$$

Greift man nun den Kerngedanken dieser Arbeit erneut auf, stellt sich trivialerweise die Frage, was passiert, wenn die unter den Hypothesen (158) angenommene Verteilungen von den tatsächlichen Verteilung der Daten abweichen. Eingangs wurde gezeigt, dass in solchen *Kontaminations-Fällen* der klassische LQ-Test im Allgemeinen nicht robust ist (s. **Abb. 3**). Um in diesem Sinne das Auftreten möglicher Ausreißer in den Daten zu simulieren, zieht man aus der ϵ -Kontaminationsumgebung

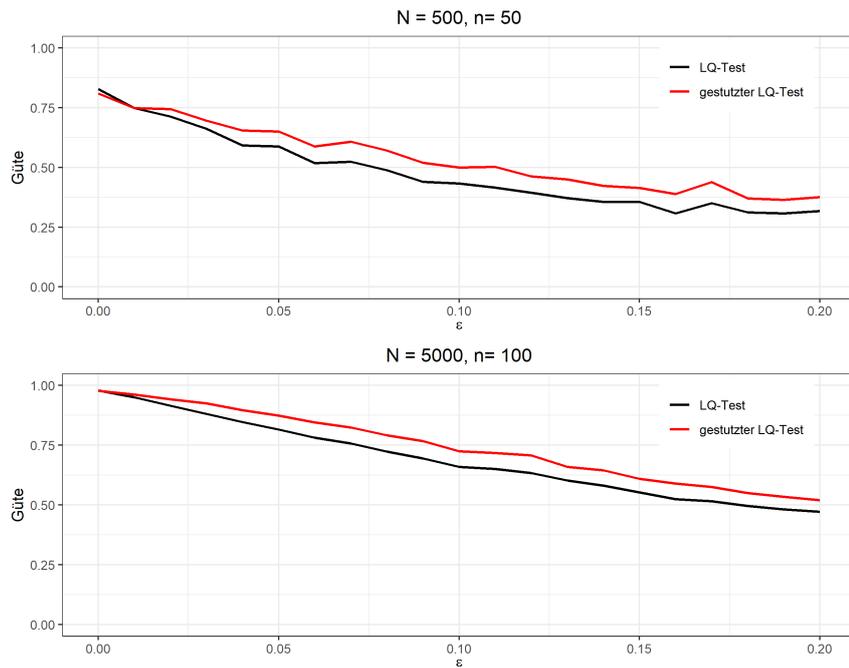
$$\mathcal{P}_\epsilon = (1 - \epsilon)\mathcal{N}(\mu, 1) + \epsilon\mathcal{N}(\mu, 10) \quad (159)$$

N Stichproben vom Umfang n . Um die Güte des klassischen LQ-Tests mit der gestutzten Version nach Huber zu vergleichen, simuliert man zu verschiedenen Niveaus α Daten zu unterschiedlichen Stichprobengrößen. Für die Berechnung der Fehlerwahrscheinlichkeiten werden $\mu = 0$ bzw. $\mu = 0.4$ bei der Datensimulation betrachtet. Da die Simulation lediglich zu Veranschaulichungs Zwecken der theoretisch dargestellten Aspekte dient, wurden die Schranken des gestutzten LQ-Tests empirisch gewählt, sodass für $s_u = 0.3$ und $s_o = 1.3$ gilt. Die **Abb. 7** stellt die Ergebnisse der Simulation dar. Anhand der Abbildung lassen sich die zuvor theoretisch festgehaltenen Resultate validieren: Bis auf $\epsilon = 0$ (d.h. keine kontaminierten Daten liegen vor), weist der gestutzte LQ-Test für verschiedene Niveaus α und unterschiedlichen Stichprobenumfängen eine höhere Güte als der klassische LQ-Test auf. Man bemerke, dass die Veranschaulichungen und der zugrunde liegende R-Code Qin and E.Priebe (2017) berücksichtigt⁷.

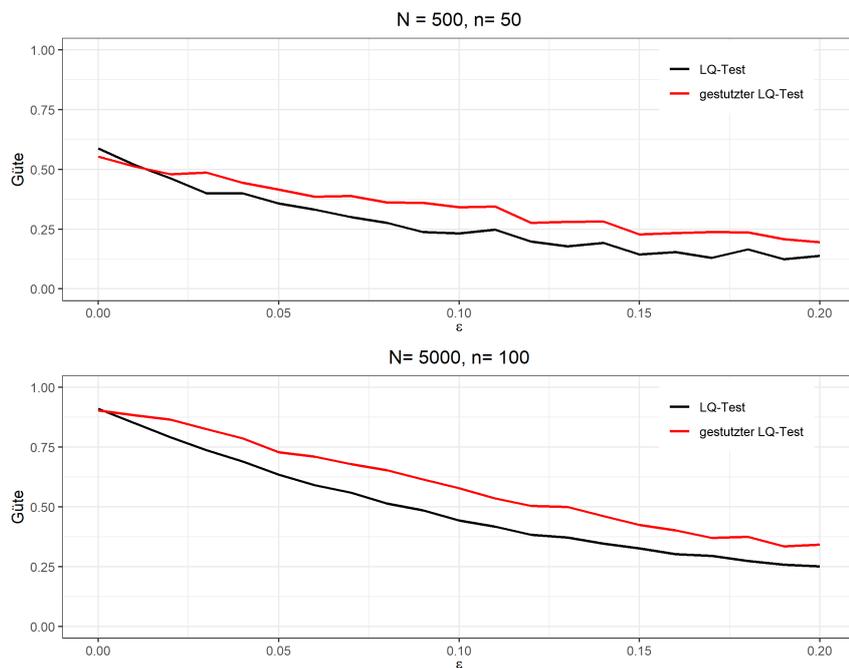
Nun soll im zweiten Teil der vorliegenden Arbeit anhand des Bayes-Faktors eine bayesiansche Alternative zu klassischen Hypothesentests vorgestellt werden. Dabei soll angelehnt an den ersten Teil dieser Arbeit gezeigt werden, dass der Bayes-Faktor im Allgemeinen ebenso nicht robust ist. Um eben dieser fehlenden Robustheit des Bayes-Faktors entgegenzuwirken, wird sich um eine robuste Version bemüht und mögliche Probleme bei der Realisierung einer Robustifizierung aufgezeigt.

⁷Siehe elektronischen Anhang für detaillierte Verweise bzgl. Darstellung und R-Code.

3. ROBUSTE MINIMAX-TESTS



(a) Güte des klassischen bzw. gestutzten LQ-Tests zum Niveau $\alpha = 0.05$



(b) Güte des klassischen bzw. gestutzten LQ-Tests zum Niveau $\alpha = 0.01$

Abbildung 7: Vergleich der Güte des klassischen LQ-Testes mit der gestutzten Version für $\alpha \in \{0.01, 0.05\}$ für $0 \leq \epsilon \leq 0.20$ kontaminierte Daten.

4 Robustifizierung des Bayes-Faktors

4.1 Eine bayesianische Alternative

Modern statisticians have developed extensive mathematical techniques, but for the most part have rejected the notion of the probability of a hypothesis...
(Jeffreys, 1961, S. ix)

Während in der klassischen Inferenz der interessierende Parameter θ als fest, aber unbekannt ist, behandelt man den Parameter θ in der bayesianischen Inferenzschule als eine Zufallsvariable mit einer *Priori-Verteilung* $p(\theta)$, anhand dessen ein mögliches Vorwissen über den Parameter ausgedrückt wird. Das Wissen nach Realisierung der Daten X dahingegen, wird über die sogenannte *Posteriori-Verteilung* $p(\theta|x)$ formuliert:

Definition 4.1 (Posteriori-Verteilung)

Sei \mathbf{X} eine Zufallsvektor mit Realisation \mathbf{x} . Die Posteriori-Verteilung des Parameters θ mit der Priori-Verteilung $p(\theta)$ ist dann definiert als

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta) p(\theta)}{\int p(\mathbf{x}|\theta) p(\theta) d\theta} \propto p(\mathbf{x}|\theta) p(\theta), \quad (160)$$

wobei es sich bei

$$\frac{1}{\int p(\mathbf{x}|\theta) p(\theta) d\theta} \quad (161)$$

um eine Normierungskonstante handelt. Dabei ergibt sich (160) über die Anwendung des Satz von Bayes.

An dieser Stelle sei angemerkt, dass über Konzepte der Bayes-Inferenz bspw. in Rüger (1998, S.73 - S.211) nachgelesen werden kann. Als einer der großen Advokaten der bayesianischen Denkweise und Kritiker der klassischen Interpretation von Hypothesentests über Konzepte wie den p-Werten, bemühte sich *Harold Jeffreys* (1891-1989) um eine intuitivere Methodik statistischer Inferenz (Ly et al., 2016, S.2). Um auch Wahrscheinlichkeiten für Hypothesen methodisch fassbar zu machen, stellte Jeffreys (1961) für verschiedene Szenarien statistische Hypothesentests im Form von sogenannten *Bayes-Faktoren* vor. Betrachtet man hierbei die Modellwahl als zufällig, so können über die Hypothesen H_0 und H_1 *priori Wahrscheinlichkeiten* $p(H_0)$ bzw. $p(H_1)$ mit $p(H_0) + p(H_1) = 1$ definiert werden. Diese, vor der Datenrealisation, festgelegten Wahrscheinlichkeiten, sollen den *Glauben* (*degree of belief*) in die jeweiligen Hypothesen darstellen. Bezeichnet man mit \mathbf{x}

4. ROBUSTIFIZIERUNG DES BAYES-FAKTORS

beobachtete Daten, so folgt die *posteriori Wahrscheinlichkeit* über den Satz von Bayes zu

$$p(H_i|\mathbf{x}) = \frac{p(\mathbf{x}|H_i)p(H_i)}{p(\mathbf{x})} \quad (i = 0, 1). \quad (162)$$

Die posteriori Wahrscheinlichkeiten bringen zum Ausdruck, wie wahrscheinlich die Hypothesen nach Realisierung der Daten sind (a posteriori Situation). Betrachtet man nun den Quotienten aus den posteriori Wahrscheinlichkeiten der Null- und Gegenhypothese, so erhält man folgende Gleichung:

$$\underbrace{\frac{p(H_1|\mathbf{x})}{p(H_0|\mathbf{x})}}_{\text{Posteriori-Chance}} \stackrel{(160)}{=} \frac{\frac{p(\mathbf{x}|H_1)p(H_1)}{p(\mathbf{x})}}{\frac{p(\mathbf{x}|H_0)p(H_0)}{p(\mathbf{x})}} = \underbrace{\frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)}}_{\text{Bayes-Faktor}} \cdot \underbrace{\frac{p(H_1)}{p(H_0)}}_{\text{Priori-Chance}} \quad (163)$$

Die relative Plausibilität der Nullhypothese H_0 und Gegenhypothese H_1 wird durch die *Posteriori-Chance* formalisiert. Anhand der Gleichheit in (163) stellt man fest, dass sich die Posteriori-Chance über das Produkt von dem sogenannten *Bayes-Faktor* und der *Priori-Chance* ergibt. Das *a priori* Verhältnis der beiden Hypothesen wird somit über den Bayes-Faktor zu einer *a posteriori* Betrachtung, d.h. der Bayes-Faktor wird interpretiert als die Evidenz der Daten unter den Hypothesen (Ly et al., 2016, S.4). Betrachtet man nun in (163) explizit den Bayes-Faktor, so lässt sich dieser als Quotient von *marginalen Likelihoods*

$$\frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} = \frac{\int p(\mathbf{x}|\theta, H_1) p(\theta|H_1) d\theta}{\int p(\mathbf{x}|\theta, H_0) p(\theta|H_0) d\theta} \quad (164)$$

schreiben, wobei θ den Parameter unter den Hypothesen H_i ($i = 0, 1$) bezeichnet. Da der Parameter θ als Zufallsvariable betrachtet wird, handelt es sich bei $p(\theta|H_i)$ um die Priori-Verteilung des Parameters θ unter der Hypothese H_i . Handelt es sich um ein einfaches Testproblem, so entspricht der Bayes-Faktor dem klassischen Likelihood-Quotienten (Kass and Raftery, 1995, S.776).

Bemerkung 4.1 (Wahl der Priori)

Obwohl die Wahl der Priori bei bayesianischen Testverfahren eine essentielle Rolle spielt, wird dies hier nicht thematisiert. Analog zum vorhergehenden Teil der Arbeit liegt das Augenmerk hierbei auf der Likelihood-Sensitivität. Jedoch soll hier bemerkt werden, dass die sogenannte Priori-Sensitivität⁸ des Bayes-Faktors im Rahmen von Robustheitsargumenten nicht zu vernachlässigen ist, denn die Priori beeinflusst entscheidend die Berechnung der marginalen Likelihoods.

⁸Siehe hierzu bspw. Rios Insua and Ruggeri (2000), Berger (1990) und Ebner et al. (2019).

Im Folgenden soll nun veranschaulicht werden, wie im Rahmen der robusten Statistik eine mögliche robuste Version des Bayes-Faktors realisiert werden kann. Vor allem die maßgebliche Problematik bei der *Robustifizierung des Bayes-Faktors* spielt hierbei eine essentielle Rolle.

4.2 Robuste Version des Bayes-Faktors

Betrachtet man nun den Bayes-Faktor

$$BF := \frac{p(\mathbf{x}|H_1)}{p(\mathbf{x}|H_0)} = \frac{\int p(\mathbf{x}|\theta, H_1) p(\theta|H_1) d\theta}{\int p(\mathbf{x}|\theta, H_0) p(\theta|H_0) d\theta} \quad (165)$$

für zwei Hypothesen H_0 und H_1 mit Daten $\mathbf{x} = (x_1, \dots, x_n)$, so stellt man anhand ähnlicher Überlegungen wie bei der gestutzten LQ-Statistik (41) fest, dass sich eine robuste Version des Bayes-Faktors über

$$BF_X^{ge} = \prod_{i=1}^n \max \left(s_u, \min \left(s_o, \frac{p(x_i|H_1)}{p(x_i|H_0)} \right) \right) \quad (166)$$

formulieren lässt, sodass wiederum Beobachtungen x_i , die den Bayes-Faktor verzerren können, an der unteren Schranke s_u bzw. oberen Schranke s_o gestutzt werden. Für eine Darstellung gemäß (166) nimmt man stillschweigend an, dass sich der Bayes-Faktor zu

$$BF = \prod_{i=1}^n \frac{p(x_i|H_1)}{p(x_i|H_0)} = \prod_{i=1}^n \frac{\int p(x_i|\theta) p(\theta|H_1) d\theta}{\int p(x_i|\theta) p(\theta|H_0) d\theta} \quad (167)$$

zerlegen lässt. Unter der Annahme, dass die Zufallsvariablen X_i ($i = 1, \dots, n$) unabhängig sind, gilt jedoch nur

$$p(\mathbf{x}|\theta) = \prod_{i=1}^n p(x_i|\theta). \quad (168)$$

Damit kann der Bayes-Faktor lediglich dargestellt werden als:

$$BF = \frac{\int p(\mathbf{x}|\theta) p(\theta|H_1) d\theta}{\int p(\mathbf{x}|\theta) p(\theta|H_0) d\theta} = \frac{\int \prod_{i=1}^n p(x_i|\theta) p(\theta|H_1) d\theta}{\int \prod_{i=1}^n p(x_i|\theta) p(\theta|H_0) d\theta} \quad (169)$$

Die gewünschte Form (166) eines robusten Bayes-Faktors ist aber im Allgemeinen nicht identisch zu der Darstellung (169), sodass sich die Robustifizierung des Bayes-Faktors auf diese Art nicht einfach realisieren lässt. Diese Problematik kann anhand von $\mathbf{x} = (x_1, x_2)$ verdeutlicht werden:

4. ROBUSTIFIZIERUNG DES BAYES-FAKTORS

Mit $p(\theta)$ Priori-Verteilung über den Parameter θ , lassen sich die jeweiligen marginalen Dichten zu

$$p(x_1) = \int p(x_1|\theta) p(\theta) d\theta \quad (170)$$

$$p(x_2) = \int p(x_2|\theta) p(\theta) d\theta \quad (171)$$

schreiben, sodass Folgendes für \mathbf{x} folgt:

$$p(\mathbf{x}) = \int f(\mathbf{x}|\theta) p(\theta) d\theta \quad (172)$$

$$= \int p(x_1|\theta) p(x_2|\theta) p(\theta) d\theta \quad (173)$$

$$\stackrel{i.A.}{\neq} \int p(x_1|\theta) p(\theta) d\theta \cdot \int p(x_2|\theta) p(\theta) d\theta \quad (174)$$

$$= p(x_1) \cdot p(x_2) \quad (175)$$

Man erkennt somit, dass gegeben θ die Zufallsvariablen X_1, X_2 unabhängig sind, jedoch folgt dadurch keine unbedingte Unabhängigkeit. Letztendlich kann man an dieser Stelle festhalten, dass eine Robustifizierung des Bayes-Faktors gemäß vorhergehenden Abschnitten nicht einfach umsetzbar ist.

5 Fazit

Welche Auswirkungen haben Abweichungen von den zugrundeliegenden Annahmen auf statistische Verfahren? Eingangs wurde diese Frage als Leitmotiv der vorliegenden Arbeit formuliert und gezeigt, wieso insbesondere der Likelihood-Quotienten-Test im Allgemeinen nicht robust gegenüber Abweichungen von den angenommenen Verteilungsannahmen ist. Über Umgebungsmodelle und Least Favorable Pair wurde der gestutzte Likelihood-Quotienten-Test nach Huber mathematisch motiviert und als eine robuste Version des klassischen Likelihood-Quotienten-Tests vorgestellt. Hierbei hat man exemplarisch anhand der ϵ -Kontaminations und Total-Variations-Umgebung gesehen, dass der gestutzte LQ-Test einen Minimax-Test darstellt und somit sich als robuste Alternative für den klassischen LQ-Test eignet. Weiterhin hat man festgestellt, dass sich diese Minimax-Resultate auf bialternierende bzw. bimonotone Kapazitäten in polnischen Räumen erweitern lassen, sodass man damit eine Generalisierung des Fundamentallemmas nach Neyman-Pearson erreicht. Dabei wurde gezeigt, dass das Huber-Strassen-Theorem für die weitere Forschung im Bereich der robusten Statistik einen Meilenstein darstellt. Über die schrittweise mathematische Heranführung an robuste Minimax-Test, sowie durch detaillierte Beweisführung wird durch die vorliegende Arbeit erhofft, Konzepte der robusten Statistik einem breiteren Publikum zu eröffnen. Des Weiteren wurde anhand exemplarischen Simulationen verdeutlicht, dass sich theoretische Resultate auch praktisch validieren lassen. Dahingehend lässt sich besonders hervorheben, dass der LQ-Test unter Kontamination (durch bspw. Ausreißer) eine weitaus bessere Güte vorweist als der klassische LQ-Test.

Um die Resultate über robuste Minimax-Tests auch in den bayesianischen Kontext übertragen zu können, hat man den Bayes-Faktor, eine bayesianische Alternative zu klassischen Hypothesentests, betrachtet. Angelehnt an vorherige Resultate wurde eine robuste Version des Bayes-Faktors vorgeschlagen. Man hat jedoch festgestellt, dass die vorgeschlagene Robustifizierung des Bayes-Faktors sich nicht einfach realisieren lässt, sodass nun die Frage offenbleibt, inwiefern sich solch eine robuste Version in der weiterführenden Forschung umsetzen lässt.

Abschließend lässt sich festhalten, dass das Teilgebiet der robusten Statistik eine mathematisch fordernde Disziplin darstellt, die mit vielen Konzepten aus der mathematischen Statistik unterlegt ist. Trotzdem können, wie auch in dieser Arbeit veranschaulicht, robuste Verfahren für die praxisbezogene Anwendung aufgrund von wünschenswerten minimax Eigenschaften von Nutzen sein.

A Anhang

A.1 Metrische und topologische Räume

Inhalt dieses ersten Abschnittes sind elementare Begriffe über metrische und topologische Räume. Definitionen und Sätze können beispielsweise in Werner (2018) und Jänich (2005), sowie Forster (2017) nachgelesen werden.

Definition A.1 (Metrischer Raum)

Sei X eine Menge und

$$d : X \times X \rightarrow \mathbb{R}, \quad (x, y) \rightarrow d(x, y)$$

eine Abbildung mit den Eigenschaften

1. $d(x, y) = 0 \Leftrightarrow x = y$
2. $\forall x, y \in X : d(x, y) = d(y, x)$
3. $\forall x, y, z \in X : d(x, z) \leq d(x, y) + d(y, z)$,

dann wird d eine Metrik und das Paar (X, d) metrischer Raum genannt.

Definition A.2 (Umgebung)

Sei (X, d) ein metrischer Raum. Man nennt

$$B_\epsilon(x) := \{x \in X : d(a, x) < \epsilon\}$$

ϵ -Umgebung von x mit $a \in X$. Damit heißt eine Teilmenge $U \subset X$ Umgebung eines Punktes $x \in X$ falls

$$\exists \epsilon > 0 : B_\epsilon \subset U.$$

Definition A.3 (Offene und abgeschlossene Mengen)

Man nennt $U \subset X$ offen, falls gilt

$$\forall x \in U \exists \epsilon > 0 : B_\epsilon(x) \subset U.$$

Dagegen nennt man $U \subset X$ abgeschlossen, falls U^c offen.

Satz A.1 ($\epsilon - \delta$ -Kriterium der Stetigkeit)

Seien (X, d_1) und (Y, d_2) zwei metrische Räume und $x_0 \in X$. Dann nennt man die Abbildung $f : X \rightarrow Y$ genau dann stetig in x_0 , falls

$$\forall \epsilon > 0 \exists \delta > 0 : \forall x \in X : [(d_1(x, x_0) < \delta) \implies (d_2(f(x), f(x_0)) < \epsilon)]. \quad (176)$$

Man nennt f stetig auf X , falls f in jedem Punkt $x_0 \in X$ stetig ist.

Definition A.4 (Topologie)

Sei X eine Menge. Eine Menge \mathcal{T} von Teilmengen X nennt man Topologie auf X , falls gilt

1. $\emptyset, X \in \mathcal{T}$
2. $A, B \in \mathcal{T} \Rightarrow A \cap B \in \mathcal{T}$
3. $A_i \in \mathcal{T} \Rightarrow \bigcup_{i \in I} A_i \in \mathcal{T}$.

Dann nennt man das Paar (X, \mathcal{T}) topologischen Raum.

Definition A.5 (Kompaktheit)

Eine Teilmenge A des metrischen Raumes (X, d) nennt man kompakt, falls jede offene Überdeckung $(U_i)_{i \in I}$ eine endliche Teilüberdeckung besitzt. Dabei versteht man unter einer offenen Überdeckung von A eine Familie $(U_i)_{i \in I}$ von offenen Teilmengen $U_i \subset X$ mit

$$A \subset \bigcup_{i \in I} U_i, \quad (177)$$

mit I beliebige endliche bzw. unendliche Indexmenge.

Satz A.2 (Heine-Borel)

Eine Teilmenge $A \subset \mathbb{R}^n$ ist genau dann kompakt, wenn sie abgeschlossen und beschränkt ist.

A.2 Maß- und Wahrscheinlichkeitstheorie

Dieser Abschnitt widmet sich Begrifflichkeiten der Maß- und Wahrscheinlichkeitstheorie, die für das Verständnis der vorliegenden Arbeit nötig sind. Definitionen und weiterführende Konzepte findet man bspw. in Elstrodt (2018), Klenke (2013), sowie Schmidt (2011).

Definition A.6 (σ -Algebra)

Ein Mengensystem $\mathcal{F} \subset 2^\Omega$ heißt σ -Algebra über Ω , falls die folgenden drei Bedingungen erfüllt sind:

1. $\Omega \in \mathcal{F}$.
2. $A \in \mathcal{F} \implies A^c := \Omega \setminus A \in \mathcal{F}$.
3. $A_i \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}, i \in \mathbb{N}$

Definition A.7 (Erzeugte σ -Algebra)

Sei $A \subset \Omega$, dann nennt man

$$\sigma(A) = \{\phi, A, A^c, \Omega\} \quad (178)$$

die von A erzeugte σ -Algebra. Zudem ist $\sigma(A)$ die kleinste σ -Algebra, die A enthält.

Definition A.8 (Borelsche σ -Algebra)

Sei \mathcal{T} eine beliebige Topologie auf Ω , dann heißt

$$\mathcal{B}(\Omega, \mathcal{T}) := \sigma(\mathcal{T}) \quad (179)$$

Borelsche σ -Algebra von \mathcal{T} . Speziell für $\Omega = \mathbb{R}$ und $\mathcal{I} = \{]a, b[\mid a, b \in \mathbb{R}, a < b\} \subset \mathbb{R}$ Mengensystem offener Intervalle in \mathbb{R} gilt

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{I}). \quad (180)$$

$B \in \mathcal{B}(\mathbb{R})$ nennt man Borel-Mengen.

Definition A.9 (Messraum)

Falls \mathcal{X} eine σ -Algebra über den Grundraum Ω ist, so nennt man (Ω, \mathcal{X}) Messraum.

Definition A.10 (Maß)

Sei (Ω, \mathcal{X}) Messraum. Eine Mengenfunktion

$$\mu : \mathcal{X} \longrightarrow \bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\} \quad (181)$$

nennt man Maß auf \mathcal{X} , falls gilt

1. $\mu(\emptyset) = 0$
2. $\forall A \in \mathcal{X} : \mu(A) \geq 0$
3. $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$

Definition A.11 (Maßraum)

Sei (Ω, \mathcal{X}) Messraum und $\mu : \mathcal{X} \rightarrow \bar{\mathbb{R}}$ ein Maß, dann nennt man $(\Omega, \mathcal{X}, \mu)$ Maßraum.

Definition A.12 (Wahrscheinlichkeitsmaß, Wahrscheinlichkeitsraum)

Sei (Ω, \mathcal{X}) ein Messraum und $\mathbb{P} : \mathcal{X} \rightarrow [0, 1]$ ein Maß auf (Ω, \mathcal{X}) mit $\mathbb{P} = 1$. Dann nennt man \mathbb{P} Wahrscheinlichkeitsmaß und $(\Omega, \mathcal{X}, \mathbb{P})$ einen Wahrscheinlichkeitsraum.

Definition A.13 (Maß mit Dichte)

Sei $(\Omega, \mathcal{X}, \mu)$ ein Maßraum und f eine nichtnegative, messbare Funktion, dann nennt man $f \odot \mu$ Maß mit der Dichte f bezüglich μ .

Definition A.14 (Dominierendes Maß)

Seien μ und ν zwei Maße auf (Ω, \mathcal{X}) , dann nennt man μ dominierendes Maß bezüglich ν , falls gilt

$$\forall A \in \mathcal{X} : \mu(A) = 0 \Rightarrow \nu(A) = 0. \quad (182)$$

Definition A.15 (Polnischer Raum)

Ein topologischer Raum (X, \mathcal{T}) wird polnischer Raum genannt, falls gilt:

1. (X, \mathcal{T}) ist separabel
2. (X, \mathcal{T}) ist vollständig bzgl. einer durch die Metrik induzierte Topologie \mathcal{T}

Bemerkung A.1 (Bemerkung zur Definition von polnischen Räumen)

Man nennt einen topologischen Raum (X, \mathcal{T}) separabel, falls (X, \mathcal{T}) eine höchstens abzählbare Teilmenge besitzt, die dicht in diesem Raum liegt. Weiterhin bedeutet die zweite Forderung der Definition A.15, dass eine Metrik d auf den Grundraum X existiert, sodass dieser die Topologie \mathcal{T} induziert und diesbezüglich vollständig ist.

B Elektronischer Anhang

Der elektronische Anhang besteht aus dem Ordner *Visualisierungen*, der den Code zum Reproduzieren der Abbildungen 2, 3 und 7 enthält. Die Abbildungen wurden mit der frei zugänglichen statistischen Programmiersprache R erzeugt (R Core Team, 2017). Dabei wurden folgende Pakete verwendet:

- *ggplot2* (Wickham, 2016)
- *grid* (Aguie, 2017)

Weiterhin soll erneut angemerkt werden, dass Teile des Codes bzgl. der im *Abschnitt 3.6* dargestellten Simulation den von Qin and E.Priebe (2017) elektronisch bereitgestellten Code berücksichtigt.

Literatur

- Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid"Graphics*. R package version 2.3.
URL: <https://CRAN.R-project.org/package=gridExtra>
- Augustin, T. (1999). Neyman-Pearson Testing under Interval Probability by Globally Least Favorable Pairs: A Survey of Huber-Strassen Theory and Some Results on its Extension to General Interval Probability.
URL: <http://nbn-resolving.de/urn/resolver.pl?urn=nbn:de:bvb:19-epub-1560-5>
- Bednarski, T. (1981). On Solutions of Minimax Test Problems for Special Capacities, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **58**: 397–405.
- Berger, J. O. (1990). Robust Bayesian Analysis: Sensitivity to the Prior, *Journal of Statistical Planning and Inference* **25**(3): 303 – 328.
- Buja, A. (1986). On the Huber-Strassen Theorem, *Probability Theory and Related Fields* **73**(1): 149–152.
- Büning, H. (2012). *Robuste und adaptive Tests*, Berlin, Boston: De Gruyter.
- Choquet, G. (1954). Theory of Capacities, *Annales de l'Institut Fourier* **5**: 131–295.
- Ebner, L., Schwaferts, P. M. and Augustin, T. (2019). Robust Bayes Factor for Independent Two-Sample Comparisons under Imprecise Prior Information, in J. De Bock, C. P. de Campos, G. de Cooman, E. Quaeghebeur and G. Wheeler (eds), *Proceedings of the Eleventh International Symposium on Imprecise Probabilities: Theories and Applications*, Vol. 103 of *Proceedings of Machine Learning Research*, PMLR, pp. 167 –174.
- Elstrodt, J. (2018). *Maß- und Integrationstheorie*, eight edn, Springer Spektrum.
- Forster, O. (2017). *Analysis 2: Differentialrechnung im \mathbb{R}^n , gewöhnliche Differentialgleichungen*, Grundkurs Mathematik, 11 edn, Vieweg+Teubner Verlag.
- Hafner, R. (1982a). Construction of Least Favourable Pairs of Distributions and of Eobnst Tests for Contamination Neighbourhoods, *Series Statistics* **13**(1): 47–56.
- Hafner, R. (1982b). Simple Construction of Least Favourable Pairs of Distributions and of Robust Tests for Prokhorov-neighbourhoods, *Series Statistics* **13**(1): 33–46.

- Hafner, R. (1992). Konstruktion robuster Teststrategien, in G. Schach, Siegfried und Trenkler (ed.), *Data Analysis and Statistical Inference: Festschrift in Honour of Friedhelm Eicker*, Verlag Josef Eul, Bergisch Gladbach-Köln, pp. 145–160.
- Hampel, F. (1986). *Robust Statistics. The Approach based on Influence Functions*, Wiley Series in Probability and Mathematical Statistics. Wiley.
- Held, L. (2008). *Methoden der statistischen Inferenz, Likelihood und Bayes*, first edn, Spektrum Akademischer-Verlag.
- Huber, P. (1964). Robust Estimation of a Location Parameter, *Ann. Math. Statist.* **35**: 73–101.
- Huber, P. (1965). A Robust Version of the Probability Ratio Test, *Ann. Math. Statist.* **36**: 1753–1758.
- Huber, P. (1981). *Robust Statistics*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series, Wiley.
- Huber, P. (1996). *Robust Statistical Procedures: Second Edition*, CBMS-NSF Regional Conference Series in Applied Mathematics, 2 edn, Society for Industrial and Applied Mathematics.
- Huber, Peter und Strassen, V. (1973). Minimax Tests and the Neyman-Pearson Lemma for Capacities, *Ann. Statistics* **1**: 251–263.
- Jänich, K. (2005). *Topologie*, 8 edn, Springer Berlin Heidelberg.
- Jeffreys, H. (1961). *Theory of Probability*, third edn, Oxford.
- Kass, R. E. and Raftery, A. E. (1995). Bayes Factors, *Journal of the American Statistical Association* **90**(430): 773 – 795.
- Klenke, A. (2013). *Wahrscheinlichkeitstheorie*, third edn, Springer.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, second edn, Springer Verlag, New York.
- Ly, A., Verhagen, J. and Wagenmakers, E. (2016). Harold Jeffreys’s default Bayes factor Hypothesis Tests: Explanation, Extension, and Application in Psychology, *Journal of Mathematical Psychology* **72**: 19 – 32.
- Österreicher, F. (1978). On the Construction of Least Favourable Pairs of Distributions, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **43**(1): 49–55.

- Österreicher, F. (2006). Least Favorable Distributions, *in* S. Kotz, N. Balakrishnan, C. Read and B. Vidakovic (eds), *Encyclopedia of Statistical Sciences*, American Cancer Society, pp. 4108 – 4111.
- Qin, Y. and E.Priebe, C. (2017). Robust Hypothesis Testing via Lq-Likelihood, *Statistica Sinica* **27**(4): 1793 – 1813.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing.
URL: <https://www.R-project.org/>
- Rieder, H. (1977). Least Favorable Pairs for Special Capacities, *Ann. Statist.* **5**(5): 909 – 921.
- Rios Insua, D. and Ruggeri, F. (2000). Robust Bayesian Analysis, *Lecture notes in statistics* **152**.
- Rüger, B. (1998). *Test-und Schätztheorie*, first edn, De Gruyter Oldenbourg.
- Rüschendorf, L. (2014). *Mathematische Statistik*, Springer Verlag, Berlin Heidelberg.
- Sagan, C. (1995). *The Demon-Haunted World: Science As a Candle in the Dark*, New York: Random House.
- Schmidt, K. D. (2011). *Maß und Wahrscheinlichkeit*, second edn, Springer-Verlag Berlin Heidelberg.
- Strassen, V. (1964). Messfehler und Information, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **2**(4): 273 – 305.
- Weichselberger, K. (2001). *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I. Intervallwahrscheinlichkeit als umfassendes Konzept*, first edn, Physica-Verlag Heidelberg.
- Werner, D. (2018). *Funktionalanalysis*, eight edn, Springer Berlin Heidelberg.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.
URL: <https://ggplot2.tidyverse.org>

Abbildungsverzeichnis

1 Non-parametrische, parametrische und robuste Statistik im Vergleich (Hampel, 1986, S.7)	5
2 Veranschaulichung der angenommenen und tatsächlichen Verteilung (Eigene Skizze)	6
3 p-Werte des LQ-Testes unter tatsächlicher und angenommener Verteilung (Eigene Skizze)	7
4 Hypothesen in Umgebungen \mathcal{P}_i ($i = 0, 1$) (Eigene Skizze)	13
5 Lévy-Distanz zwischen den Verteilungsfunktionen F und G (Huber, 1981, S.28)	14
6 Darstellung von (Q_0, Q_1) (Eigene Skizze)	16
7 Vergleich der Güte des klassischen LQ-Testes mit der gestutzten Version für $\alpha \in \{0.01, 0.05\}$ für $0 \leq \epsilon \leq 0.20$ kontaminierte Daten.	37

Eigenständigkeitserklärung:

Hiermit bestätige ich, dass die vorliegende Bachelorarbeit *Robuste Minimax-Tests und eine Robustifizierung des Bayes-Faktors* selbstständig verfasst und keine anderen Publikationen, Vorlagen und Hilfsmittel als die angegebenen benutzt wurden. Alle Teile meiner Arbeit, die wortwörtlich oder dem Sinn nach anderen Werken entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

Ort, Datum

Unterschrift