

Vereinfachte Simplex-Datentiefe in Regressionsmodellen

Diplomarbeit

vorgelegt von

Michael Swora

am 11.06.2015

Betreuerin: Prof. Dr. Christine Müller

Inhaltsverzeichnis

1.	Einl	eitung	3
2.	Datentiefe		
	2.1.	Das Konzept der Datentiefe	5
	2.2.	Simplex- und Tangens-Datentiefe	6
3.	Die	vereinfachte Simplex-Datentiefe	18
	3.1.	Die betrachtete Modellklasse	18
	3.2.	$\label{eq:Vereinfachte} \ensuremath{\operatorname{Vereinfachte}}\xspace \ensuremath{\operatorname{Simplex-Datentiefe}}\xspace \ensuremath{\operatorname{durch}}\xspace \ensuremath{\operatorname{aternierende}}\xspace \ensuremath{\operatorname{Residuenvorzeichen}}\xspace \ensuremath{\operatorname{chen}}\xspace \ensuremath{\{chen}}\xspace \{c$	20
	3.3.	Asymptotische Verteilungen	27
	3.4.	Konsistenz der Tests	35
4.	Simulationsstudien		
	4.1.	Der Vorzeichentest	51
	4.2.	Die drei betrachteten Regressionsmodelle	53
		4.2.1. Vorstellung der Modelle	53
		4.2.2. Anwendbarkeit der vereinfachten Simplex-Datentiefe	55
		4.2.3. Konsistenz der Tests im Rahmen der Modelle	57
	4.3.	Simulation: Zielsetzung und Vorgehensweise	63
	4.4.	Ergebnisse für Modell 1: Die einfache lineare Regression	64
	4.5.	Ergebnisse für Modell 2: Das exponentielle Wachstumsmodell	68
	4.6.	Ergebnisse für Modell 3: Das Power-Modell	71
5.	Zusa	ammenfassung	76

Anhang

A. Programmcode	78
Literaturverzeichnis	85
Erklärung	86

1. Einleitung

Das Konzept der Datentiefe ist ein allgemeiner multivariater, nichtparametrischer und robuster Ansatz in der Statistik, deren früheste Form erstmals im Jahre 1975 von John Tukey formuliert und danach von vielen anderen in unterschiedlichsten Formen weiterentwickelt wurde. Im Rahmen eines Aufsatzes (Kustosz et al., 2015) wurden innerhalb einer großen Klasse von Regressionsmodellen mehrere neue, vereinfachte Varianten einer verallgemeinerten Simplex Datentiefe untersucht. Die Simplex-Datentiefe ist eine wohlbekannte Form der Datentiefe, deren Verallgemeinerung durch eine Verknüpfung mit anderen Formen der Datentiefe realisiert wird.

Die Datentiefe allgemein ordnet jedem Punkt θ eines Parameterraumes Θ eine reelle Zahl zu, die die "Tiefe" von θ innerhalb der Stichprobe misst. Mit ihrer Hilfe können Begriffe wie Ränge und Quantile auf den multivariaten Fall verallgemeinert werden.

Im Aufsatz (Kustosz et al., 2015) werden nun, basierend auf den hergeleiteten vereinfachten Varianten der verallgemeinerten Simplex-Datentiefe, Parametertests für die Regressionsmodelle entwickelt und ihre asymptotische Verteilung hergeleitet. Die Vereinfachung ist weniger Computer-intensiv und ermöglicht gerade die unkomplizierte Herleitung asymptotischer Verteilungen. Weiter wird unter gewissen Bedingungen an die jeweiligen Regressionsmodelle die Konsistenz der Tests bewiesen. Mehrere spezielle Regressionsmodelle werden auf ihre Eignung überprüft, die vorherigen Resultate anzuwenden. Schließlich wird zu zwei Modellen die Gütefunktion durch eine Simulationsstudie geschätzt.

In der vorliegenden Arbeit werden die Resultate des Aufsatzes (Kustosz et al., 2015) dargestellt. Die Beweise werden dabei an vielen Stellen ausführlicher gestaltet. Ein eigenes Kapitel führt in die Thematik der Datentiefe ein und stellt die erforderlichen Grundbegriffe vor. Beispielabbildungen ergänzen im ersten Drittel der Arbeit den Text. Desweiteren wird die Simulationsstudie im Rahmen von drei weiteren Regressionsmodellen fortgeführt.

Kapitel 2 erläutert das Konzept der Datentiefe ausführlicher. Außerdem werden die benötigten Formen der Datentiefe sowie die zum Verständnis erforderlichen mathematischen

1. Einleitung

Begriffe eingeführt und Zusammenhänge dargestellt.

Kapitel 3 stellt zunächst ein zentrales Resultat vor, welches dann (unter gewissen Anforderungen an das Regressionsmodell) die Anwendung der vereinfachten Datentiefe-Formen ermöglicht. Diese werden danach eingeführt und die asymptotischen Verteilungen der Tests, die sich daraus ergeben, werden hergeleitet. Schließlich wird unter bestimmten Bedingungen an das zugrundeliegende Regressionsmodell die Konsistenz der Tests bewiesen.

Kapitel 4 beginnt mit einer kurzen Besprechung des Vorzeichentests, welcher in den Simulationsstudien als Vergleich herangezogen wird. Für drei Regressionsmodelle wird danach die Anwendbarkeit der vereinfachten Simplex-Datentiefe besprochen und die Konsistenz der Tests für diese Modelle diskutiert. Es schließt sich die Simulationsstudie an, die für die drei Modelle durchgeführt wurde, und bei der die Gütefunktionen der Tests für die drei Modelle sowohl für normalverteilte als auch für kontaminiert normalverteilte Fehler simuliert werden.

Abschließend werden die zentralen Ergebnisse in Kapitel 5 zusammengefasst.

In diesem Kapitel wird der Begriff der Datentiefe aufgebaut.

Im ersten Abschnitt werden in kurzer Form allgemeine Informationen zu diesem Konzept gegeben. Der zweite Abschnitt liefert dann zunächst die zum Verständnis der für uns wichtigen Formen der Datentiefe benötigten mathematischen Begriffe. Danach werden diese, zusammen mit einigen ihrer Eigenschaften, eingeführt.

2.1. Das Konzept der Datentiefe

Das Konzept der Datentiefe innerhalb einer Stichprobe ist ein nichtparametrischer statistischer Ansatz. Die erste Version der Datentiefe wurde 1975 in Form der Halbraum-Datentiefe von John Tukey eingeführt. Mit ihrer Hilfe wurde der Tukey-Median definiert, der eine Verallgemeinerung des univariaten Medians auf den multivariaten Fall darstellt.

Allgemein ordnet eine Datentiefe-Funktion jedem Punkt θ eines Parameterraumes Θ eine reelle Maßzahl für die Tiefe von θ innerhalb der Stichprobe zu. Im Beispiel eines Lokationsproblemes im \mathbb{R}^k , bei dem der Parameterraum und der Wertebereich der Stichprobenvariablen als \mathbb{R}^k zusammenfallen, korrespondieren kleine Werte der Datentiefe mit Punkten, die eher außerhalb der durch die Stichprobe generierten Datenwolke liegen, und große Werte mit zentralen Punkten. Die Datentiefe als Abbildung $d : \mathbb{R}^k \to \mathbb{R}$ ist demnach ein Maß dafür, wie weit ein Punkt vom Zentrum der Daten entfernt ist. So entsteht eine Ordnung der Punkte im \mathbb{R}^k . Auf diese Weise werden Begriffe wie Ränge, Quantile, Median und Vorzeichen auf den multivariaten Fall verallgemeinert.

Die Datentiefe hat je nach Situation und Form robuste Eigenschaften unterschiedlicher Ausprägung. Es wurden im Laufe der Zeit viele weitere Datentiefefunktionen vorgeschlagen. Zu nennen sind z.B. sind die Simplex-Datentiefe, die Tangens-Datentiefe, die Oja-Datentiefe oder die Mahalanobis-Datentiefe.

Der Raum der Anwendungen umfasst viele Gebiete der multivariaten Statistik. So findet

das Konzept beispielsweise Anwendung in Gebieten wie Ausreißererkennung, Rangtests, multivariate Dichteschätzer und das wichtige Gebiet der robusten Regression (Serfling, 2006, S.1 ff.).

Im weiteren Verlauf dieser Arbeit sind nur die Simplex- und die Tangens-Datentiefe von Interesse, welche im nächsten Abschnitt auch ausführlicher beschrieben werden. Deren Kombination wird in Kapitel 3 die Basis für die Versionen der vereinfachten Simplex-Datentiefe bereiten.

2.2. Simplex- und Tangens-Datentiefe

Obwohl die im Folgenden einzuführenden Begriffe der affinen Unabhängigkeit, der konvexen Hülle sowie schließlich des Begriffes des Simplex auf allgemeinen Vektorräumen definiert werden können, wird hier (im Hinblick auf die Anwendungen) der Einfachheit halber nur der \mathbb{R} -Vektorraum \mathbb{R}^k , mit $k \in \mathbb{N}$, betrachtet.

Anschaulich erhalten wir ein k-Simplex durch die Verbindung von (k + 1) Punkten x_1, \ldots, x_{k+1} im k-dimensionalen Raum durch Geraden. Beispielsweise ist ein Simplex im zweidimensionalen Raum ein Dreieck.

Damit jeder Punkt des Simplex eindeutig bestimmt ist, benötigen wir die affine Unabhängigkeit der x_i . Dies führt zu folgender Definition:

2.2.1 Definition:

Sei der \mathbb{R} -Vektorraum \mathbb{R}^k und *n* Punkte $x_1, \ldots, x_n \in \mathbb{R}^k$ gegeben.

1. x_1, \ldots, x_n heißen affin unabhängig, falls Folgendes erfüllt ist:

$$\forall \lambda_1, \dots, \lambda_n \in \mathbb{R} \text{ mit } \lambda_1, \dots, \lambda_n \neq 0, \ \sum_{i=1}^n \lambda_i = 0 \text{ gilt: } \sum_{j=1}^n \lambda_j x_j \neq 0.$$

2. Die konvexe Hülle $Conv(x_1, \ldots, x_n)$ von x_1, \ldots, x_n wird definiert als die Menge

$$Conv(x_1,\ldots,x_n) := \left\{ \sum_{j=1}^n \lambda_j x_j \colon \sum_{i=1}^n \lambda_i = 1, \ \lambda_i \ge 0 \right\}.$$

Wir können nun den Begriff des Simplex definieren.





(a) Konvexe Hülle von 5 Punkten in der Ebene

(b) Simplex in der Ebene

Abbildung 2.1.: Beispiel einer konvexen Hülle und eines Simplex

2.2.2 Definition:

Seien $x_1, \ldots, x_{k+1} \in \mathbb{R}^k$ affin unabhängig. Ein *k-Simplex* ist dann definiert als die konvexe Hülle $Conv(x_1, \ldots, x_{k+1})$.

Abbildung 2.1(a) veranschaulicht die konvexe Hülle von fünf Punkten in der Ebene. Der schraffierte Bereich inklusive Rand kennzeichnet diese. Abbildung 2.1(b) zeigt ein Dreieck, das einem Simplex in der Ebene darstellt, welcher in diesem Falle durch k + 1 = 3 Punkte erzeugt wird.

Nun steht uns der begriffliche Apparat für die Einführung der Simplex-Datentiefe zur Verfügung.

2.2.3 Definition (vgl. Liu, 1990, S.407):

Seien *n* Realisationen $\boldsymbol{x} = (x_1, \ldots, x_n)$ von \mathbb{R}^k -wertigen Zufallsvektoren, n > k, gegeben, und bezeichne $\mathcal{X} = \mathbb{R}^{n \times k}$ den zugehörigen Stichprobenraum. Sei $\theta \in \mathbb{R}^k$.

 Die Indexmenge M(d), die alle ⁿ_{d+1} Teilmengen der Mächtigkeit (d+1) von {1,...,n} enthält, definieren wir als

$$M(d) := \{ (z_1, \dots, z_{d+1}) : z_i \in \{1, \dots, n\}, z_1 < z_2 < \dots < z_{d+1} \}.$$
(2.1)

2. Die Simplex-Datentiefe $d_s^*(\theta, \boldsymbol{x}) : \mathbb{R}^k \times \mathcal{X} \to [0, 1]$ von θ bezüglich \boldsymbol{x} definieren wir nun als den Anteil der Simplices, die θ enthalten, an allen möglichen Simplices.

$$d_s^*(\theta, \boldsymbol{x}) := \frac{1}{\binom{n}{k+1}} \sum_{(z_1, \dots, z_{k+1}) \in M(k)} \mathbb{1}_{Conv(x_{z_1}, \dots, x_{z_{k+1}})}(\theta)$$



Abbildung 2.2.: Beispiel: Simplex-Datentiefe in der Ebene

In Abbildung 2.2 sind 4 Datenpunkte $x_1 \ldots, x_4$ in der Ebene gegeben, sowie der rot gekennzeichnete Punkt θ , dessen Simplex-Datentiefe in Bezug auf x_1, \ldots, x_4 wir berechnen wollen. Die Teilabbildungen (a) bis (d) zeigen alle $4 = \binom{4}{3}$ verschiedenen 3-elementigen Punktekombinationen auf, die einen Simplex formen. Zwei dieser Simplices enthalten θ . Die Datentiefe berechnet sich in diesem Beispiel also zu $\frac{2}{4} = \frac{1}{2}$.

Wir wollen eine weitere Form der Datentiefe einführen, die wir später benötigen werden. Es handelt sich um die, historisch gesehen erste, 1975 von John Tukey eingeführte Datentiefe, nämlich die Halbraum-Datentiefe. Zunächst erinnern wir dazu an den Begriff des *Halbraums*.

2.2.4 Definition:

Sei $a \in \mathbb{R}^n$ und $\gamma \in \mathbb{R}$. Sei $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n$ das Standardskalarprodukt im \mathbb{R}^n . Ein *abge-schlossener Halbraum* bezüglich *a* und γ ist definiert als die Menge

$$\{x \in \mathbb{R}^n \mid \langle a, x \rangle \ge \gamma\}$$
 bzw. $\{x \in \mathbb{R}^n \mid \langle a, x \rangle \le \gamma\}.$

Ein abgeschlossener Halbraum ist also ein Teilraum, der dadurch entsteht, dass der ganze Raum durch eine Hyperebene geteilt wird. Die Hyperebene selbst gehört dabei zum abge-

schlossenen Halbraum dazu. Zum Beispiel teilt im zweidimensionalen Raum die x-Achse den Raum in 2 Halbräume.

Nun kommen wir zur Definition der Halbraum-Datentiefe.

2.2.5 Definition (vgl. Dutter et al., 2003, S.195):

Sei $\boldsymbol{x} = (x_1, \dots, x_n)$ eine Stichprobe von \mathbb{R}^k -wertigen Zufallsvektoren, $\boldsymbol{\theta} \in \mathbb{R}^k$ ein Parameter und $\mathcal{X} = \mathbb{R}^{n \times k}$ der zugehörige Stichprobenraum. Die Halbraum-Datentiefe $d_{halb}(\boldsymbol{\theta}, \boldsymbol{x})$: $\boldsymbol{\Theta} \times \mathcal{X} \rightarrow [0, 1]$ von $\boldsymbol{\theta}$ bezüglich \boldsymbol{x} ist folgendermaßen definiert:

$$d_{halb}(\theta, \boldsymbol{x}) := \frac{1}{n} \min_{u \in \mathbb{R}^k} \#\{i \mid \langle u, x_i \rangle \le \langle u, \theta \rangle\} = \min_{u \in \mathbb{R}^k} \#\{i \mid \langle u, x_i - \theta \rangle \le 0\}.$$

Anschaulich wird dabei eine Hyperebene durch θ gelegt und um 360 Grad gedreht, wobei jeweils die Anzahl der Datenpunkte auf einer Seite der Hyperebene gezählt wird. Die minimale Anzahl ist dann der Wert der Halbraum-Datentiefe. Da das Skalarprodukt $\langle a, b \rangle$ bekanntlich geschrieben werden kann als

$$\langle a, b \rangle = ||a|| \cdot ||b|| \cdot \cos \phi(a, b), \quad \phi(a, b)$$
 Winkel zwischen a und b ,
 $||x||$ euklische Norm von x ,

reicht es aufgrund des Vorzeichenwechsels der Kosinusfunktion aus, die Hyperebene um 180 Grad zu rotieren und jeweils das Minimum der Anzahl der Datenpunkte auf beiden Seiten zu notieren. Abbildung 2.3 gibt ein Beispiel für vier Datenpunkte in der Ebene. Jede farbige Gerade teilt die Ebene in (hinsichtlich der Anzahl der Datenpunkte auf beiden Seiten der Geraden) unterschiedliche Halbräume. In der Legende ist für jeden Halbraum die minimale Punkteanzahl notiert. Die kleinste minimale Anzahl beträgt 1, sodass die Halbraumtiefe von θ in diesem Beispiel $\frac{1}{4}$ beträgt.

Die Halbraum-Datentiefe in $\theta \in \mathbb{R}^k$ bezüglich x_1, \ldots, x_{k+1} hat die Eigenschaft, genau dann größer Null zu sein, falls θ in der konvexen Hülle der x_i liegt.

2.2.6 Satz:

Set $\mathbf{x} = (x_1, \dots, x_{k+1})$ eine Stichprobe von \mathbb{R}^k -wertigen Zufallsvektoren und $\theta \in \mathbb{R}^k$ ein Parameter. Dann gilt:

$$\theta \in Conv(x_1, \dots, x_{k+1}) \iff d_{halb}(\theta, \boldsymbol{x}) > 0.$$
(2.2)



Abbildung 2.3.: Beispiel: Halbraum-Datentiefe in der Ebene

Abbildung 2.4 zeigt solch eine Situation in der Ebene. Sobald θ außerhalb der konvexen Hülle der Datenpunkte liegt, lässt sich ein Winkel der durch θ führenden Gerade finden, sodass auf einer Seite keine Datenpunkte vorhanden sind und die Tiefe somit den Wert Null annimmt.

Eine weitere Form der Datentiefe, die wir im späteren Verlauf benötigen werden, ist die *Tangens-Datentiefe*, die von Mizera eingeführt wurde (Mizera, 2002). Mizera definiert hierfür zuerst über den Begriff der *schwachen Optimalität* eines Parameters (hinsichtlich einer Menge von Qualitätsfunktionen) die *globale* und *lokale Datentiefe*.

2.2.7 Definition (vgl. Mizera, 2002, S.1683 f.):

Sei eine Stichprobe $\boldsymbol{x} = (x_1, \dots, x_n)$ mit $x_i \in \mathbb{R}^d$ gegeben sowie der zugehörige Stichprobenraum $\mathcal{X} = \mathbb{R}^{n \times d}$. Sei weiter ein Parameterraum $\Theta \subseteq \mathbb{R}^k$, $k \in \mathbb{N}$ gegeben.

- Eine Funktion Q_x mit Q_x(θ): Θ → ℝⁿ_{≥0} := {y ∈ ℝⁿ | y_i ∈ [0,∞] ∀ i = 1,...,n} ∀ x ∈
 X heißt Qualitätsfunktion. Die Komponente Q_i := (Q_x)_i, i = 1...,n, heißt die zur Beobachtung i gehörende Qualitätsfunktion.
- Sei A ⊆ {1,...,n}. Ein Parameterwert θ* ∈ Θ* ⊆ Θ heißt schwach optimal in Θ* bezüglich A, falls gilt:

$$\forall \ \tilde{\theta} \in \Theta^* \ \exists \ i \in A \ \text{mit} \ Q_i(\theta^*) \le Q_i(\tilde{\theta}).$$



Abbildung 2.4.: Beispiel zu Satz 2.2.6

Die Qualitätsfunktion Q_i misst, wie gut der Fit eines Parameterwertes θ aus einer Teilmenge des Parameterraums \mathbb{R}^d ist. Hierbei soll die Qualitätsfunktion umso kleiner werden, je besser der Parameterwert zur Beobachtung x_i passt. Schwache Optimalität von θ in Θ^* bezüglich A bedeutet dann nur, dass es kein Argument $\tilde{\theta} \in \Theta^*$ der Qualitätsfunktion Q_i gibt, welches Q_i für alle $i \in A$ echt kleiner werden lässt als $Q_i(\theta)$.

Mithilfe des Begriffes der schwachen Optimalität können wir die die globale Datentiefe d_{gl} und ihre lokale Version d_{lok} definieren.

2.2.8 Definition (vgl. Mizera, 2002, S.1684 ff.):

Unter den Voraussetzungen und Bezeichnungen von 2.2.7 sind $d_{gl}(\theta, \boldsymbol{x}) : \Theta \times \mathcal{X} \to [0, 1]$ und $d_{lok}(\theta, \boldsymbol{x}) : \Theta \times \mathcal{X} \to [0, 1]$ definiert als

1. $d_{gl}(\theta; \boldsymbol{x}) := \frac{1}{2^n} \min_{A \subseteq \{1, \dots, n\}} \# \{ A^C : \theta \text{ ist nicht schwach optimal in } \Theta \text{ bezüglich } A \},$

2.
$$d_{lok}(\theta; \boldsymbol{x}) := \frac{1}{2^n} \min_{A \subseteq \{1, \dots, n\}} \# \{ A^C : \forall \Theta^* \text{ offen }, \theta \in \Theta^* \subseteq \Theta, \text{ ist } \theta \text{ nicht schwach}$$
optimal in Θ^* bezüglich $A \}.$

Die globale Datentiefe gibt also die minimale Anzahl an Beobachtungen an, die entfernt werden müssen, sodass θ nicht schwach optimal bezüglich der verbleibenden Beobachtungen ist.

Je nach dem zugrunde liegenden statistischen Modell und der benutzten Qualitätsfunktion können sich bereits bekannte Formen der Datentiefe als Spezialfälle ergeben. In der

Situation von Definition 2.2.5 ergibt sich die Halbraum-Datentiefe, wenn wir als Qualitätsfunktionen Q_i den euklidischen Abstand von x_i zu θ_i verwenden, $Q_i := ||x_i - \theta||$ (Mizera, 2002, S.1685).

Bevor wir nun zu der in dieser Arbeit wichtigen Tangens-Datentiefe kommen, wollen wir auf einen weiteren Spezialfall der globalen Datentiefe eingehen. Es handelt sich um die erstmalig 1999 von Rousseeuw und Hubert eingeführte Regressions-Datentiefe (Rousseeuw und Hubert, 1999). Wir betrachten ein multiples lineares Regressionsmodell mit abhängiger Variable $y_i \in \mathbb{R}$, unabhängigen Variablen $\boldsymbol{x}_i := (x_{i1}, \ldots, x_{i(k-1)}) \in \mathbb{R}^{k-1}$ und der Modellgleichung

$$Y_{i} = \theta_{1} + \theta_{2} x_{i1} + \ldots + \theta_{k} x_{i(k-1)} + \epsilon_{i}, \quad i = 1, \ldots, n,$$
(2.3)

mit den k unbekannten Parametern $\theta_1, \ldots, \theta_k$. Wir setzen $b_i := (y_i, x_i)$. Rousseeuw und Hubert geben nun die folgende Definition:

2.2.9 Definition (vgl. Rousseeuw und Hubert, 1999, S.388):

Die Regressions-Datentiefe eines Parameters (Fits) $\theta \in \mathbb{R}^k$ bezüglich der Beobachtungen b_i , $i = 1, \ldots, n$, ist die kleinste Anzahl an Beobachtungspunkten, die entfernt werden müssen, damit θ ein Nonfit ist.

Hier bedarf der Begriff "Nonfit" einer weiteren Definition.

2.2.10 Definition (Rousseeuw und Hubert, 1999, S.397):

Der Parameter (Fit) $\theta \in \mathbb{R}^k$ heißt ein *Nonfit* bezüglich der Beobachtungen b_i , i = 1, ..., n, falls eine affine Hyperebene H im Raum der \mathbf{x}_i existiert, sodass $\mathbf{x}_i \notin H$ für alle i = 1, ..., nerfüllt ist, und für die Residuen $r_i := y_i - \theta_1 - \theta_2 x_{i1} - \ldots - \theta_k x_{i(k-1)}$ Folgendes gilt:

 $r_i > 0$ für alle Bebobachtungen x_i , die in einem der entstehenden beiden offenen Halbräume des x-Raumes liegen, und $r_i < 0$ für alle Beobachtungen, die in dem anderen offenen Halbraum liegen.

Der Begriff "Nonfit" meint also, dass eine affine Hyperebene (d.h. eine Hyperebene, die nicht notwendigerweise durch den Ursprung geht, sondern um einen Betrag verschoben sein kann) im Raum der \mathbf{x}_i existiert, die diesen in zwei Halbräume teilt. Hierbei darf kein x_i auf der Hyperebene liegen. Die zu den Beobachtungen in einem offenen Halbraum



Abbildung 2.5.: Fit und Nonfit in einer einfachen linearen Regression

gehörenden Residuen müssen entweder strikt größer Null oder strikt kleiner Null sein. Die Regressions-Datentiefe zählt dann die minimale Anzahl an Punkten, die entfernt werden müssen, damit obige Situation eintritt.

Wir veranschaulichen die Situation mit einem Beispiel in der Ebene, d.h. durch eine einfache lineare Regression. Abbildung 2.5 zeigt sechs Datenpunkte in der Ebene. Hier ist ein Nonfit abhängig von der Existenz einer reellen Zahl v, sodass $r_i(\theta) \stackrel{>}{\leq} 0 \forall x_i < v$ und $r_i(\theta) \stackrel{>}{>} 0 \forall x_i > v$. Diese Situation ist bei dem Fit, der durch die blaue Gerade dargestellt wird, offensichtlich nicht gegeben. Bei dem Fit, der durch die rote Gerade dargestellt wird, existiert eine relle Zahl, die in der Abbildung mit "v" markiert ist: Alle Residuen, die zu x-Werten gehören, die kleiner als v sind, sind kleiner als Null. Alle zu x-Werten, die größer als v sind, gehörenden Residuen sind größer Null.

Anhand der Abbildung sehen wir außerdem, dass ein Fit θ , repräsentiert durch eine Gerade, genau dann ein Nonfit ist, falls ein Angelpunkt $z \in \mathbb{R}^2$ existiert, um den die Gerade bis zur Vertikalen rotiert werden kann, ohne dass sie eine Beobachtung b_i passiert (Rousseeuw und Hubert, 1999, S.387). Ist dies jedoch der Fall, so bewirkt eine beliebig kleine Rotation in die andere Richtung, dass die Beträge sämtlicher Residuen sich verringern.

Diese Überlegungen lassen sich problemlos auf den höherdimensionalen Fall der multiplen

linearen Regression übertragen. Wir erhalten daher folgende äquivalente Definition eines Nonfits:

2.2.11 Definition:

Im multiplen Regressionsmodell 2.3 heißt der Fit $\theta \in \mathbb{R}^k$ ein Nonfit bezüglich der Beobachtungen b_i , i = 1, ..., n, falls gilt:

$$\exists \theta' \in \mathbb{R}^k : |y_i - (1, \boldsymbol{x}_i)^T \theta'| < |y_i - (1, \boldsymbol{x}_i)^T \theta| \quad \forall i \in \{1, \dots, n\}.$$

$$(2.4)$$

Wir stellen fest, dass dies aber die Negation der schwachen Optimalität von θ bezüglich $\{1, \ldots, n\}$ ist, wenn wir als Qualitätsfunktionen den Betrag der Residuen verwenden. Die Definition des Nonfits fällt also in diesem Setting mit der der nicht schwachen Optimalität zusammen.

Weiterhin ist dann die Regressions-Datentiefe gleich der globalen Datentiefe, wie aus der Definition selbiger leicht klar wird. Wir sehen also, dass die globale Datentiefe von Mizera als Verallgemeinerung der Regressions-Datentiefe auf beliebige Modelle und Qualitätsfunktionen gesehen werden kann.

Wir kehren nun wieder zum allgemeinen Fall der globalen Datentiefe zurück. Eine einfacher zu handhabende Variante dieser ist die Tangens-Datentiefe. Wir gehen von der Tatsache aus, dass eine notwendige Bedingung für die schwache Optimalität von θ_0 ist (Mizera, 2002, S.1687), dass gilt:

$$\mathbf{0} \in Conv\left(\frac{\partial Q_1}{\partial \theta}(\theta_0), \dots, \frac{\partial Q_n}{\partial \theta}(\theta_0)\right).$$
(2.5)

Hierbei ist $\frac{\partial Q_i}{\partial \theta}(x)$, der Vektor der partiellen Ableitungen von Q_i nach $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$ (an der Stelle $x \in \mathbb{R}^k$), definiert als

$$\frac{\partial Q_i}{\partial \theta}(x) := \begin{pmatrix} \frac{\partial Q_i}{\partial \theta_1}(x) \\ \vdots \\ \frac{\partial Q_i}{\partial \theta_k}(x) \end{pmatrix}.$$

Dann definieren wir die Tangens-Datentiefe d_{tan} folgendermaßen:

2.2.12 Definition (vgl. Mizera, 2002, S.1687):

Unter den Bezeichnungen und Voraussetzungen von 2.2.7 sei Θ offen und $Q_i(\theta, \boldsymbol{x})$ differenzierbar in $\theta_0 \forall \boldsymbol{x} \in \mathcal{X}, i = 1, ..., n$. Die Tangens-Datentiefe $d_{tan} : \Theta \times \mathcal{X} \to [0, 1]$ ist dann gegeben als

$$d_{tan}(\theta_0, \boldsymbol{x}) := \frac{1}{n} \min_{u \in \mathbb{R}^k} \#\{i : u^T \frac{\partial Q_i}{\partial \theta}(\theta_0) \le 0\}$$

Durch Umschreiben erhalten wir $d_{tan}(\theta_0, \boldsymbol{x}) := \frac{1}{n} \min_{u \in \mathbb{R}^k} \#\{i : \langle u, \frac{\partial Q_i}{\partial \theta}(\theta_0) \rangle \leq 0\}$. Dies können wir als die Halbraum-Datentiefe von $\boldsymbol{0} \in \mathbb{R}^k$ interpretieren bezüglich der Datenpunkte $\left(\frac{\partial Q_1}{\partial \theta}(\theta_0), \ldots, \frac{\partial Q_n}{\partial \theta}(\theta_0)\right)$.

Die globale, lokale und Tangens-Datentiefe stehen in Zusammenhang und sind unter bestimmten Bedingungen sogar gleich.

2.2.13 Satz (Mizera, 2002, S.1688): In der Situation von Definition 2.2.12 gilt $\forall \theta \in \Theta, x \in \mathcal{X}$:

- 1. $d_{lok}(\theta, \boldsymbol{x}) \leq d_{gl}(\theta, \boldsymbol{x}) \leq d_{tan}(\theta, \boldsymbol{x}).$
- 2. Ist Θ eine konvexe Menge, und sind alle Qualitätsfunktionen Q_i ebenfalls konvex, dann gilt:

$$d_{lok}(\theta, \boldsymbol{x}) = d_{gl}(\theta, \boldsymbol{x}) = d_{tan}(\theta, \boldsymbol{x}).$$

Wir kommen zurück zur Simplex-Datentiefe. Diese kann aufgrund von 2.2 äquivalent in der Form

$$d_s^*(\theta, \boldsymbol{x}) = \frac{1}{\binom{n}{k+1}} \sum_{(z_1, \dots, z_{k+1}) \in M(k)} \mathbb{1}(d_{halb}(\theta, x_{z_1}, \dots, x_{z_{k+1}}) > 0)$$

geschrieben werden. Hierbei bezeichnet $\mathbb{1}(f(x) > 0)$ die Indikatorfunktion $\mathbb{1}_A(x)$ mit $A := \{x \mid f(x) > 0\}.$

Wird nun die Halbraum-Datentiefe durch eine beliebige Datentiefe-Funktion ersetzt, ergibt sich eine Verallgemeinerung, die wir als *verallgemeinerte Simplex-Datentiefe* bezeichnen.

2.2.14 Definition:

In der Situation von Definition 2.2.7 sei eine Datentiefefunktion $d: \Theta \times \mathbb{R}^{(k+1) \times d} \to [0,1]$ gegeben. Dann ist die *verallgemeinerte Simplex-Datentiefe* $d_s: \Theta \times \mathbb{R}^{n \times d} \to [0,1]$ definiert als

$$d_s(\theta, \boldsymbol{x}) := \frac{1}{\binom{n}{k+1}} \sum_{(z_1, \dots, z_{k+1}) \in M(k)} \mathbb{1}(d(\theta, x_{z_1}, \dots, x_{z_{k+1}}) > 0).$$
(2.6)

Von nun an werden wir mit dem Begriff "Simplex-Datentiefe" die Tiefe aus Definition 2.2.14 bezeichnen, und wir setzen für die Datentiefefunktion d im weiteren Verlauf dieser Arbeit die Tangens-Datentiefe ein.

Wir stellen nun einige Überlegungen an, wie ein Testkonzept auf den Parameter θ im Rahmen der Datentiefe aussehen könnte (Kustosz et al., 2015, S.2). Die Simplex-Datentiefe sollte grob gesprochen groß sein, falls $\theta \in \mathbb{R}^k$ der korrekte Parameter ist, und klein andernfalls. Daraus erhalten wir einen einfachen Ansatz, für $\mathbb{R}^k = \Theta_0 \dot{\cup} \Theta_1$ die Hypothese

$$H_0: \theta \in \Theta_0 \quad \text{vs.} \quad H_1: \theta \in \Theta_1$$

$$(2.7)$$

zum Niveau α zu testen. Sei hierfür q_{α} das α -Quantil der Verteilung bzw. Grenzverteilung der sich aus 2.6 ergebenden Statistik $d_s(\theta, \mathbf{X})$, wobei \mathbf{X} ein n-Tupel \mathbb{R}^d -wertiger Zufallsvektoren ist. Ein Test $\varphi : \mathbb{R}^{n \times d} \to \{0, 1\}$ für das obige Hypothesenpaar hat nun die Form

$$arphi(oldsymbol{x}) = egin{cases} 1 & & \leq q_lpha \ & ext{falls } \sup_{ heta \in \Theta_0} d_s(heta, oldsymbol{x}) & \ & 0 & > q_lpha \ \end{cases}$$

Dieser Test ist wegen $\sup_{\theta \in \Theta_0} d_s(\theta, X) \ge d_s(\theta, X)$ für alle $\theta \in \Theta_0$ und damit

$$P_{\theta}(\sup_{\theta' \in \Theta_0} d_s(\theta', \boldsymbol{x}) < q_{\alpha}) \le P_{\theta}(d_s(\theta, \boldsymbol{x}) < q_{\alpha}) = \alpha \quad \forall \ \theta \in \Theta_0$$

ein (asymptotischer) Test zum Niveau α .

Aufgrund der Schwierigkeiten bei der Herleitung einer zumindest asymptotischen Verteilung der Teststatistik werden wir, unter Beschränkung auf eine ausgewählte Klasse von Regressionsmodellen in Abschnitt 3.1, vereinfachte Simplex-Datentiefen in Abschnitt 3.2 herleiten.

Die Basis dafür ist der Umstand, dass, falls die Regressionsfunktion gewissen Anforderungen genügt, die Tangens-Datentiefe eines Parameters $\theta \in \mathbb{R}^k$ bei k + 1 Beobachtungen genau dann größer Null ist, falls die zugehörigen Residuen alternierende Vorzeichen aufweisen. Da die Berechnung der Simplex-Datentiefe für große n und k sehr zeitintensiv werden kann, da $\binom{n}{k+1}$ Teilmengen auf alternierende Residuenvorzeichen geprüft werden müssen, wird die Simplex-Datentiefe in drei Versionen vereinfacht, indem nur bestimmte Gruppen von Residuen berücksichtigt werden.

3. Die vereinfachte Simplex-Datentiefe

Dieses Kapitel führt die vereinfachte Simplex-Datentiefe in drei Versionen im Rahmen einer bestimmten statistischen Modellklasse ein.

Im ersten Abschnitt wird diese Modellklasse, deren Annahmen in den folgenden Abschnitten immer vorausgesetzt werden, ausführlich vorgestellt. Der zweite Abschnitt bereitet mit einem zentralen Resultat zur Berechnung der Tangens-Datentiefe den Boden für die vereinfachte Simplex-Datentiefe, die sodann eingeführt wird. Der dritte Abschnitt befasst sich mit der Herleitung der asymptotischen Verteilung der (geeignet normierten) drei Versionen der vereinfachten Simplex Datentiefe. Der letzte Abschnitt beschäftigt sich schließlich mit Bedingungen, unter welchen die resultierenden Tests konsistent sind.

3.1. Die betrachtete Modellklasse

Im weiteren Verlauf wird die spezielle Modellklasse der (nicht notwendigerweise linearen) Regressionsmodelle betrachtet. Wir folgen dabei im gesamten Abschnitt (Kustosz et al., 2015). Die Modellklasse ist gegeben durch die Regressionsgleichung

$$Y_i = f(x_i, \theta) + \epsilon_i, \quad \theta \in \Theta = \mathbb{R}^k, \ x_i, y_i \in \mathbb{R}, \ i = 1, \dots, n.$$

$$(3.1)$$

f ist hierbei eine frei wählbare Funktion, deren einzige Voraussetzung ist, dass sie in θ differenzierbar sein muss. Für die Fehler ϵ_i nehmen wir an, dass diese unabhängig identisch verteilte Zufallsvariablen sind, deren Verteilung durch ein geeignetes Maß μ dominiert wird.

Die x_i fassen wir als gegebene Realisationen von Zufallsvariablen X_i auf. Dies bedeutet, dass wir hinsichtlich der Modellgleichung 3.1 immer auf $X_i = x_i$ bedingen, z.B. bei Erwartungswerten oder Varianzen. Dieses notieren wir im weiteren Verlauf dieser Arbeit aus ökonomischen Gründen nicht mehr explizit, behalten es jedoch immer im Hinterkopf. Für die Zufallsvariablen X_i soll überdies die Bedingung

$$X_1 < X_2 < \ldots < X_n \text{ fast sicher bzgl. } P^{X_1,\ldots,X_n}$$
(3.2)

gelten, wobei P^{X_1,\ldots,X_n} die gemeinsame Verteilung der X_i bezeichnet. Das Beobachtungspaar (x_i, y_i) notieren wir zukünftig mit $b_i \in \mathbb{R}^2$ für $i = 1, \ldots, n$, den Zufallsvektor (X_i, Y_i) mit B_i . Weiter definieren wir $\mathbf{b} := (b_1, \ldots, b_n)$ sowie $\mathbf{B} := (B_1, \ldots, B_n)$.

Das *i*-te Residuum erhalten wir als

$$r(b_i, \theta) := y - f(x_i, \theta).$$

Im Rahmen dieses Modells benutzen wir für die Qualitätsfunktion der Tangens-Datentiefe der Beobachtung b_i das quadrierte Residuum:

$$r^2(b_i, \theta) = (y_i - f(x_i, \theta))^2.$$

Es erweist sich im Folgenden als zweckmäßig, die daraus resultierende Form der Tangens-Datentiefe in einer geringfügig anderen, äquivalenten Form zu schreiben:

3.1.1 Bemerkung (Kustosz et al., 2015, S.4): $d_{tan}(\theta, \boldsymbol{b}) := \frac{1}{n} \min_{u \in \mathbb{R}^n} \#\{i : u^T \frac{\partial r^2(b_i, \theta)}{\partial \theta} \le 0\} = \frac{1}{n} \min_{u \in \mathbb{R}^n} \#\{i : u^T w(x_i, \theta) r(b_i, \theta) \le 0\},$

wobei $w(x_i, \theta)$ definiert ist als $w(x_i, \theta) := \frac{\partial f(x_i, \theta)}{\partial \theta}$.

Beweis:

Es ist

$$\frac{\partial r^2(b_i,\theta)}{\partial \theta} = 2r(b_i,\theta)\frac{\partial r(b_i,\theta)}{\partial \theta} = 2r(b_i,\theta)\frac{\partial (y_i - f(x_i,\theta))}{\partial \theta} = -2r(b_i,\theta)\frac{\partial f(b_i,\theta)}{\partial \theta}$$

Weiter gilt

$$-2u^T r(b_i, \theta) \frac{\partial f(b_i, \theta)}{\partial \theta} \le 0 \Leftrightarrow -u^T r(b_i, \theta) \frac{\partial f(b_i, \theta)}{\partial \theta} \le 0$$

und schließlich

$$\min_{u \in \mathbb{R}^n} \#\{i : -u^T r(b_i, \theta) \frac{\partial f(b_i, \theta)}{\partial \theta} \le 0\} = \min_{u \in \mathbb{R}^n} \#\{i : u^T r(b_i, \theta) \frac{\partial f(b_i, \theta)}{\partial \theta} \le 0\},$$

und daraus folgt die Behauptung.

3.2. Vereinfachte Simplex-Datentiefe durch alternierende Residuenvorzeichen

Der folgende Abschnitt folgt dem Aufsatz von Kustosz et al. (2015). Durch die folgende Definition formalisieren wir die Idee von alternierenden Vorzeichen einer Funktion.

3.2.1 Definition (vgl. Kustosz et al., 2015, S.4): Sei $f : [a,b] \to \mathbb{R}, x \in \mathbb{R}, v = (v_1, \dots, v_{k+1}) \in \mathbb{R}^{k+1}$.

1. Die Vorzeichenfunktion $sign: \mathbb{R} \to \{-1, 0, 1\}$ ist definiert als

$$sign(x) := \begin{cases} 1 & x > 0 \\ 0 & falls & x = 0 \\ -1 & x < 0. \end{cases}$$

2. Der Vektor v hat alternierende Vorzeichen, falls gilt:

$$\forall i = 1, \dots, k : sign(v_i) = -sign(v_{i+1}) \neq 0.$$

Falls also v alternierende Vorzeichen hat, besitzt der Vektor k Vorzeichenwechsel.

3. f hat k Vorzeichenwechsel im Intervall [a, b], falls für endliches A mit

$$A := \{ j \mid \exists x_1 < x_2 < \dots < x_{j+1} \in [a, b] : \forall i = 1, \dots, j :$$
$$sign(f(x_i)) = -sign(f(x_{i+1})) \neq 0 \}$$

gilt: $k = \max A$.

Das Maximum einer endlichen Menge von natürlichen Zahlen ist dabei als das größte Element dieser Menge definiert.

Nun kommen wir zu einem zentralen Resultat dieser Arbeit. Wir betrachten den Fall n = k + 1, da wir an der Tangensdatentiefe $d = d_{tan}$ interessiert sind, die in Formel 2.6 mit Teilmengen von genau (k + 1) Beobachtungen eingeht. Der folgende Satz gibt Voraussetzungen an, unter denen die Tangens-Datentiefe genau dann größer Null ist, wenn die Residuen alternierende Vorzeichen aufweisen. Dies wird zu einer Umformulierung der Simplex-Datentiefe führen, die den Boden für die bereits angesprochenen Vereinfachungen bereitet.

3.2.2 Satz (Kustosz et al., 2015, S.4 f.):

Seien zum Modell 3.1 (k+1) Beobachtungspaare $b_i = (x_i, y_i), \mathbf{b} = (b_1, \dots, b_{k+1}), gegeben.$ Sei gemäß 3.2 hierbei $x_1 < x_2 < \dots < x_{k+1}$. Die Funktion $v_u(x) : [x_1, x_{k+1}] \to \mathbb{R}$ sei definiert als $v_u(x) := u^T w(x, \theta)$. Sind dann die Voraussetzungen

- 1. $\forall u \in \mathbb{R}^k$: $v_u(x)$ hat höchstens (k-1) Vorzeichenwechsel im Intervall $[x_1, x_{k+1}]$,
- 2. $\forall t \in S := \{z \mid z \in \{-1, 1\}^{k+1} \land z \text{ hat höchstens } (k-1) \text{ Vorzeichenwechsel} \} :$ $\exists u^* \in \mathbb{R}^k \text{ mit } sign(v_{u^*}(x_i)) = t_i \forall i \in \{1, \dots, k+1\}$

erfüllt, so gilt:

$$d_{tan}(\theta, \mathbf{b}) > 0 \iff (r(b_1, \theta), \dots, r(b_{k+1}, \theta))^T \text{ hat alternierende Vorzeichen,}$$
$$oder \exists i \in \{1, \dots, k+1\}: r(b_i, \theta) = 0.$$
(3.3)

Beweis (vgl. Kustosz et al., 2015, S.25):: Sei θ beliebig, aber fest. Offensichtlich gilt:

$$r(b_j, \theta) = 0$$
 für ein $j \in \{1, \dots, k+1\} \Rightarrow d_{tan}(\theta, \mathbf{b}) > 0.$

Wenn wir nun also annehmen, dass $r(b_i, \theta) > 0 \ \forall i = 1, \dots, k+1$, und dann zeigen können

$$d_{tan}(\theta, \boldsymbol{b}) > 0 \iff (r(b_1, \theta), \dots, r(b_{k+1}, \theta))^T$$
 hat alternierende Vorzeichen,

so ist Satz 3.2.2 bewiesen. Denn für den Fall, dass es unter den (k + 1) Residuen eines gibt, welches gleich Null ist, ist die Rückrichtung von 3.3 klar nach obiger Anmerkung. Die Hinrichtung ist wahr, weil die Oder-Bedingung der rechten Seite der Implikation nach Voraussetzung wahr ist.

Wir setzen nun also $r(b_i, \theta) > 0 \forall i = 1, \dots, k+1$ voraus.

"⇒":

Angenommen, $(r(b_1, \theta), \dots, r(b_{k+1}, \theta))^T$ hat *nicht* alternierende Vorzeichen, d.h. $\exists i_0 \in \{1, \dots, k\}$ mit $sign(r(b_{i_0}, \theta)) = sign(r(b_{i_0+1}, \theta)).$

Definiere $s_i := sign(r(b_i, \theta)) \forall i = 1, ..., k + 1$, sowie $s := (s_1, ..., s_{k+1})$. Dann ist $s \in S$, und nach Voraussetzung 2 gilt:

$$\exists u_* \in \mathbb{R}^k : sign(v_{u^*}(x_i)) = s_i \ \forall \ i = 1, \dots, k+1.$$

Daraus folgt dann

$$sign(v_{u*}(x_i))sign(r(b_i,\theta)) = 1 \ \forall \ i = 1, \dots, k+1,$$

sodass $v_{u^*}(x_i)$ und $r(b_i, \theta)$ für jedes i die gleichen Vorzeichen haben. Daraus folgt aber

$$v_{u^*}(x_i)r(b_i,\theta) = {u^*}^T w(x_i,\theta)r(b_i,\theta) > 0 \ \forall \ i = 1, \dots, k+1.$$

Also ergibt sich, dass $d_{tan}(\theta, \mathbf{b}) = 0$. Die logische Umkehrung liefert:

 $d_{tan}(\theta, \mathbf{b}) > 0 \implies (r(b_1, \theta), \dots, r(b_{k+1}, \theta))$ hat alternierende Vorzeichen.

"⇐":

Angenommen, $d_{tan}(\theta, \mathbf{b}) = 0$. Dann gilt

$$\exists u^* \in \mathbb{R}^k : u^{*^T} w(x_i, \theta) r(b_i, \theta) = v_{u^*}(x_i) r(b_i, \theta) > 0 \ \forall i = 1, \dots, k+1.$$
(3.4)

Da nach Voraussetzung 1 v_{u^*} höchstens (k-1) Vorzeichenwechsel in $[x_1, x_{k+1}]$ hat, gibt

es ein $i_0 \in \{1, \ldots, k\}$ mit

$$sign(v_{u^*}(x_{i_0})) = sign(v_{u^*}(x_{i_0+1})).$$

Zusammen mit 3.4 folgt dann

$$v_{u^*}(x_{i_0})r(b_{i_0},\theta) > 0 \ \land \ v_{u^*}(x_{i_0+1})r(b_{i_0+1},\theta) > 0 \Rightarrow sign(r(b_{i_0},\theta)) = sign(r(b_{i_0+1},\theta)),$$

da sonst auf der linken Seite der Implikation nicht beide Ausdrücke größer Null wären. Also hat $(r(b_1, \theta), \ldots, r(b_{k+1}, \theta))$ nicht alternierende Vorzeichen. Die logische Umkehrung ergibt wieder:

$$(r(b_1, \theta), \ldots, r(b_{k+1}, \theta))$$
 hat alternierende Vorzeichen $\Rightarrow d_{tan}(\theta, \mathbf{b}) > 0.$

Da für die Fehler im Modell 3.1 eine stetige Verteilungsfunktion angenommen wird, besitzen auch die Residuen eine stetige Verteilungsfunktion für alle $\theta \in \mathbb{R}^k$. Dies bedeutet, dass das Ereignis, dass mindestens ein Residuum gleich Null ist, eine Nullmenge bezüglich der gemeinsamen (stetigen) Verteilung der Residuen darstellt. Unter den Voraussetzungen von Satz 3.2.2 gilt daher mit Wahrscheinlichkeit eins die Aussage:

$$d_{tan}(\theta, (b_{z_1}, \dots, b_{z_{k+1}})) > 0 \iff (r(b_{z_1}, \theta), \dots, r(b_{z_{k+1}}, \theta))^T$$
 hat alternierende Vorzeichen,

mit $z_i \in \{1, \ldots, n\} \forall i = 1, \ldots, k+1$, und $z_1 < z_2 < \ldots < z_{k+1}$). Dies impliziert unter $d_{tan}(\theta, (b_{z_1}, \ldots, b_{z_{k+1}})) > 0$ die Identität

$$\prod_{i=1}^{k+1} \mathbb{1}(r(b_{z_i},\theta)(-1)^i > 0) + \prod_{i=1}^{k+1} \mathbb{1}(r(b_{z_i},\theta)(-1)^{i+1} > 0) = 1,$$
(3.5)

da $r(b_{z_i}, \theta)(-1)^i = -r(b_{z_i}, \theta)(-1)^{i+1}$ für alle $i = 1, \ldots, k$. Die Simplex-Datentiefe kann somit auch folgendermaßen formuliert werden:

$$d_{s}(\theta, \boldsymbol{b}) = \frac{1}{\binom{n}{k+1}} \sum_{(z_{1}, \dots, z_{k+1}) \in M(k)} \left(\prod_{i=1}^{k+1} \mathbb{1}(r(b_{z_{i}}, \theta)(-1)^{i} > 0) + \prod_{i=1}^{k+1} \mathbb{1}(r(b_{z_{i}}, \theta)(-1)^{i+1} > 0) \right).$$
(3.6)

Da bei Betrachtung aller $\binom{n}{k+1}$ Teilmengen die Berechnung der Simplex-Datentiefe bei großem n und k einen hohen Rechenaufwand erfordert, werden in der folgenden Definition drei Varianten vereinfachter Simplex-Datentiefe vorgestellt. Diese erfordern dann nicht mehr die Überprüfung aller (k+1)-elementigen Teilmengen der Stichprobe auf alternierende Residuen.

Sei hierzu eine *n*-elementige Stichprobe $b_i = (x_i, y_i)$ gemäß der Modellgleichung 3.1 gegeben. Weiterhin seien die Voraussetzungen von Satz 3.2.2 erfüllt.

3.2.3 Definition (Kustosz et al., 2015, S.6):

Die erste Variante der vereinfachten Simplex-Datentiefe ist gegeben als

$$\begin{aligned} d_s^{(1)}(\theta, \boldsymbol{b}) &:= \frac{1}{\lfloor \frac{n}{k+1} \rfloor} \sum_{j=1}^{\lfloor \frac{n}{k+1} \rfloor} \left(\prod_{i=1}^{k+1} \mathbbm{1} \left((r(b_{(k+1)(j-1)+i}, \theta)(-1)^i > 0) \right. \right. \\ &+ \prod_{i=1}^{k+1} \mathbbm{1} \left(r(b_{(k+1)(j-1)+i}, \theta)(-1)^{i+1} > 0 \right) \right). \end{aligned}$$

Bei dieser Version der vereinfachten Simplex-Datentiefe betrachten wir also nur (gemäß der Anordnung der Indexmenge) nebeneinanderliegende (k + 1)-elementige Residuengruppen. Sind schließlich keine (k+1) Beobachtungen (und damit Residuen) mehr vorhanden, werden diese nicht berücksichtigt.

Abbildung 3.1 gibt ein Beispiel, in dem eine lineare Regression (also $\Theta = \mathbb{R}^2$) an 7 Beobachtungen angepasst wurde. Betrachtet werden also nebeneinanderliegende Dreiergruppen von Residuen, im Beispiel zwei Gruppen, die farblich unterschiedlich gekennzeichnet sind. In der ersten Gruppe finden sich alternierende Vorzeichen, in der zweiten nicht. Die erste Version der vereinfachten Datentiefe hat somit den Wert $\frac{1}{2}$.

Wir kommen nun zur zweiten Version der vereinfachten Simplex-Datentiefe.



Abbildung 3.1.: Beispiel: Bei $d_s^{(1)}$ betrachtete Residuengruppen

3.2.4 Definition (Kustosz et al., 2015, S.6):

Die zweite Variante der vereinfachten Simplex-Datentiefe ist gegeben als

$$d_s^{(2)}(\theta, \mathbf{b}) := \frac{1}{n-k} \sum_{j=1}^{n-k} \left(\prod_{i=1}^{k+1} \mathbb{1} \left((r(b_{j-1+i}, \theta)(-1)^i > 0) + \prod_{i=1}^{k+1} \mathbb{1} \left(r(b_{j-1+i}, \theta)(-1)^{i+1} > 0 \right) \right) \right).$$

Diese zweite Variante berücksichtigt zu jedem j = 1, ..., n - k das Residuum j und die (gemäß der Anordnung der Indexmenge) k folgenden Residuen.

Abbildung 3.2 enthält dieselbe Situation wie Abbildung 3.1., nur sind die letzten beiden Datenpunkte entfernt worden. Wir sehen, dass drei Dreiergruppen von Residuen in $d_s^{(2)}$ eingehen, die hier wiederum farblich unterschiedlich gekennzeichnet sind. Die erste Dreiergruppe umfasst die ersten drei Residuen und verschiebt sich dann zweimal um einen Index nach rechts, bis das Ende der Beobachtungsreihe erreicht ist. Nur einmal sind alternierende Residuen vorhanden, sodass die Datentiefe hier gleich $\frac{1}{3}$ ist.

Wir kommen nun zur dritten und letzten Version der vereinfachten Simplex-Datentiefe. Diese ist nur für den Spezialfall k = 2 definiert.



Abbildung 3.2.: Beispiel: Für $d_s^{(2)}$ betrachtete Residuengruppen in der Ebene

3.2.5 Definition (Kustosz et al., 2015, S.6):

Die dritte Variante der vereinfachten Simplex-Datentiefe ist, für den Fall k = 2, gegeben als

$$\begin{aligned} d_s^{(3)}(\theta, \boldsymbol{b}) &:= \frac{1}{\lfloor \frac{n-1}{2} \rfloor} \sum_{j=1}^{\lfloor \frac{n-1}{2} \rfloor} \left(\mathbbm{1} \left(r(b_j, \theta) > 0 \right) \mathbbm{1} \left(r(b_{\lfloor \frac{n+1}{2} \rfloor}, \theta) < 0 \right) \mathbbm{1} \left(r(b_{n-j+1}, \theta) > 0 \right) \\ &+ \mathbbm{1} \left(r(b_j, \theta) < 0 \right) \mathbbm{1} \left(r(b_{\lfloor \frac{n+1}{2} \rfloor}, \theta) > 0 \right) \mathbbm{1} \left(r(b_{n-j+1}, \theta) < 0 \right) \right). \end{aligned}$$

Diese dritte Variante der vereinfachten Simplex-Datentiefe, welche nur für k = 2 definiert ist und somit nur Dreiergruppen von Residuen erfasst, berücksichtigt immer das "mittlere Residuum"und das *j*-te sowie das (n - j)-te Residuum, $j = 1, \ldots, \lfloor \frac{n-1}{2} \rfloor$.

Abbildung 3.3 zeigt wieder die gleiche Situation wie Abbildung 3.1. Da eine ungerade Anzahl von Beobachtungen vorliegt, ist der mittlere Datenpunkt eindeutig bestimmt. Wieder gehen Dreiergruppen von Residuen in die Datentiefe ein. Das mittlere Residuum wird bei jeder der drei Dreiergruppen betrachtet. Die erste Dreiergruppe betrachtet zusätzlich die beiden äußersten Residuen, die zweite dann die beiden jeweils daneben liegenden Residuen, bis schließlich (im Beispiel schon im nächsten Schritt) die Mitte erreicht ist. In diesem Beispiel finden sich nur einmal alternierende Residuen, sodass die Datentiefe hier $\frac{1}{3}$ beträgt.



Abbildung 3.3.: Beispiel: Für $d_s^{(3)}$ betrachtete Residuengruppen in der Ebene

3.3. Asymptotische Verteilungen

Um unter Voraussetzung der Anwendbarkeit von Satz 3.2.2 die vereinfachten Simplex-Datentiefen als Teststatistiken gemäß nutzbar zu machen, bedarf es zumindest der Kenntnis einer asymptotischen Verteilung. Der folgende Satz gibt uns darüber Auskunft.

3.3.1 Satz (Kustosz et al., 2015, S.6):

Sei eine Folge $(B_i)_{i \in \mathbb{N}}$ von Zufallsvariablen gemäß Modell 3.1 gegeben. Gelte weiter für den wahren Parameter $\theta \in \mathbb{R}^k$, dass $P_{\theta}(\epsilon_i > 0) = P_{\theta}(\epsilon_i < 0) = \frac{1}{2} \forall i = 1, ..., n$. Dann gelten folgende Grenzverteilungen:

1.
$$D_n^{(1)}(\boldsymbol{B}, \theta) := \sqrt{\lfloor \frac{n}{k+1} \rfloor} \frac{d_s^{(1)}(\theta, \boldsymbol{B}) - (\frac{1}{2})^k}{\sqrt{(\frac{1}{2})^k (1 - (\frac{1}{2})^k)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

2.
$$D_n^{(2)}(\boldsymbol{B}, \theta) := \sqrt{n-k} \quad \frac{d_s^{(2)}(\theta, \boldsymbol{B}) - (\frac{1}{2})^k}{\sqrt{(\frac{1}{2})^k \left(3 - (\frac{1}{2})^{k-1} \cdot k - 3 \cdot (\frac{1}{2})^k\right)}} \stackrel{d}{\to} \mathcal{N}(0, 1)$$

3.
$$D_n^{(3)}(\boldsymbol{B}, \theta) := \sqrt{\lfloor \frac{n-1}{2} \rfloor} \quad \frac{d_s^{(3)}(\theta, \boldsymbol{B}) - \frac{1}{4}}{\sqrt{\frac{3}{16}}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Hierbei bezeichnet $\stackrel{d}{\rightarrow}$ die Konvergenz in Verteilung.

Bevor wir den Beweis angeben, müssen wir noch die Definition der *m*-Abhängigkeit und den Grenzwertsatz für *m*-abhängige Zufallsvariablen von Hoeffding und Robbins einführen.

Die *m*-Abhängigkeit bedeutet nichts weiter, als dass für eine Folge von Zufallsvariablen $(X_i)_{i \in \mathbb{N}}$ je zwei Elemente dieser Folge, die bezüglich der Indexmenge einen größeren Abstand als *m* haben, stochastisch unabhängig sind. Daher erhalten wir die folgende Definition:

3.3.2 Definition (Brockwell und Davis, 2006, S.212):

Eine Folge von Zufallsvariablen $(X_i)_{i\in\mathbb{N}}$ heißt *m*-abhängig, falls gilt:

 (X_1,\ldots,X_i) und $(X_{i+m+1},X_{i+m+2},\ldots)$ sind stochastisch unabhängig $\forall i \geq 1$.

Hoeffding und Robbins gaben 1948 einen Grenzwertsatz für Folgen streng stationärer *m*-abhängiger Zufallsvariablen an. Aus der strengen Stationarität folgt bekanntlich, dass $Var(X_i) = \sigma^2$ und $E(X_i) = \mu \ \forall i \in \mathbb{N}.$

3.3.3 Satz (Shang, 2012, S.713):

Sei $(X_1)_{i \in \mathbb{N}}$ eine Folge streng stationärer, m-abhängiger Zufallsvariablen. Sei $E(X_i) = \mu$ und $0 < Var(X_i) < \infty \forall i \in \mathbb{N}$. Dann gilt:

$$\frac{\sqrt{n}}{\tau} \left(\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \mu \right) \right) \xrightarrow{d} \mathcal{N}(0, 1),$$

mit $\tau = Var(X_1) + 2\sum_{j=2}^{m+1} Cov(X_1, X_j).$

Mit diesen Hilfsmitteln können wir jetzt die behaupteten Grenzverteilungen nachweisen.

Beweis zu Satz 3.3.1 (vgl. Kustosz et al., 2015, S.25 ff.): Zu Beginn sei festgehalten, dass, da θ der wahre Parameter ist, $r(B_i, \theta) = \epsilon_i$ gilt, denn:

$$r(B_i, \theta) = Y_i - f(X_i, \theta) = f(X_i, \theta) + \epsilon_i - f(X_i, \theta) = \epsilon_i.$$

Daraus folgt aufgrund der Voraussetzungen

$$P_{\theta}(r(B_i, \theta) > 0) = P_{\theta}(r(B_i, \theta) < 0) = \frac{1}{2},$$

3. Die vereinfachte Simplex-Datentiefe

und die Familie der Residuen $r_i := r(B_i, \theta), i = 1, ..., n$, ist stochastisch unabhängig.

Zu 1.)

Setze

$$W_j^{\theta} := \prod_{i=1}^{k+1} \mathbbm{1} \left((r(B_{(k+1)(j-1)+i}, \theta)(-1)^i > 0) + \prod_{i=1}^{k+1} \mathbbm{1} \left(r(B_{(k+1)(j-1)+i}, \theta)(-1)^{i+1} > 0 \right) \right)$$

Dann sind die W_j nach Gleichung 3.5 Bernoulli-Variablen für alle $j \in 1, \ldots, \lfloor \frac{n}{k+1} \rfloor$ und für alle n > k + 1, $n \in \mathbb{N}$. Wir werden nun den Erwartungswert von W_j^{θ} unter dem wahren Parameter berechnen und erhalten damit, da es sich um Bernoulli-Variablen handelt, gleichzeitig den Wert der Varianz.

Sei $G := \{x \in \{1, \dots, k+1\} \mid \exists a \in \mathbb{N} : x = 2a\}$ die Menge der geraden Zahlen in $\{1, \dots, k+1\}$ und $U := \{x \in \{1, \dots, k+1\} \mid \exists a \in \mathbb{N} : x = 2a-1\}$ die Menge der ungeraden Zahlen in $\{1, \dots, k+1\}$. Sei weiterhin j beliebig, aber fest, und $A_0 := \{r(b_{(k+1)(j-1)+1}, \theta) > 0\}$ das Ereignis, dass das erste der betrachteten (k+1) Residuen größer Null ist.

Es ergibt sich dann, da $P_{\theta}(A_0) = P_{\theta}(\neg A_0) = \frac{1}{2}$,

$$\frac{1}{2} \cdot P_{\theta} \left(W_{j}^{\theta} = 1 \mid A_{0} \right) = P_{\theta} \left(\prod_{i \in G}^{k+1} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta)(-1)^{i+1} > 0 \right) = 1 \right) \right)$$

$$= P_{\theta} \left(\left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta)(-1)^{i+1} > 0 \right) = 1 \right) \right) \right)$$

$$= P_{\theta} \left(\left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta) < 0 \right) = 1 \right) \right) \right)$$

$$= P_{\theta} \left(\left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta) > 0 \right) = 1 \right) \right) \right)$$

$$= P_{\theta} \left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta) > 0 \right) = 1 \right) \right)$$

$$= P_{\theta} \left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta) > 0 \right) = 1 \right)$$

$$= P_{\theta} \left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta) > 0 \right) = 1 \right)$$

$$= P_{\theta} \left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta) > 0 \right) \right)$$

$$= P_{\theta} \left(\prod_{i \in G} \mathbb{1} \left(r(B_{(k+1)(j-1)+i}, \theta) > 0 \right) \right)$$

$$= \left(\frac{1}{2} \right)^{\#G} \cdot \left(\frac{1}{2} \right)^{\#U} = \left(\frac{1}{2} \right)^{\#G+\#U} = \left(\frac{1}{2} \right)^{k+1}.$$
(3.7)

Nach dem gleichen Prinzip erhalten wir

$$\frac{1}{2} \cdot P_{\theta}(W_j = 1 \mid \neg A_{\theta}) = \left(\frac{1}{2}\right)^{k+1}$$

Nach dem Satz von der totalen Wahrscheinlichkeit ergibt sich dann

$$P_{\theta}(W_{j}^{\theta} = 1) = P_{\theta}(W_{j}^{\theta} = 1 \mid A_{0})P_{\theta}(A_{0}) + P_{\theta}(W_{j}^{\theta} = 1 \mid \neg A_{0})P_{\theta}(A_{0}),$$

woraus folgt:

$$P_{\theta}(W_{j}^{\theta} = 1) = 2\left(\frac{1}{2}\right)^{k+1} \frac{1}{2} + 2\left(\frac{1}{2}\right)^{k+1} \frac{1}{2} = \left(\frac{1}{2}\right)^{k+1} + \left(\frac{1}{2}\right)^{k+1} = \left(\frac{1}{2}\right)\left[\left(\frac{1}{2}\right)^{k} + \left(\frac{1}{2}\right)^{k}\right] = \left(\frac{1}{2}\right)^{k}$$

Die Varianz unter dem wahren Parameter berechnet sich mit $p = P_{\theta}(W_j^{\theta} = 1)$ bekanntlich zu $Var_{\theta}(W_j^{\theta}) = p(1-p)$. Da

$$d_s^{(1)}(\boldsymbol{\theta}, \boldsymbol{B}) = \frac{1}{\lfloor \frac{n}{k+1} \rfloor} \sum_{j=1}^{\lfloor \frac{n}{k+1} \rfloor} W_j^{\boldsymbol{\theta}}$$

gilt, ist der zentrale Grenzwertsatz direkt anwendbar, womit die Behauptung bewiesen ist.

Zu 2.)

Wir setzen

$$W_j^{\theta} := \prod_{i=1}^{k+1} \mathbb{1}\left(\left(r(B_{j-1+i}, \theta)(-1)^i > 0 \right) + \prod_{i=1}^{k+1} \mathbb{1}\left(r(B_{j-1+i}, \theta)(-1)^{i+1} > 0 \right) \right)$$

Nach demselben Prinzip wie bei 1.) erhalten wir, dass W_j^{θ} Bernoulli-Variablen sind mit $P_{\theta}(W_j^{\theta} = 1) = \left(\frac{1}{2}\right)^k$, $j = 1, \ldots, n - k$. Da sich die betrachteten Residuen von W_j^{θ} und W_l^{θ} , $j \neq l$, überschneiden für alle n > k + 1, $n \in \mathbb{N}$, ist die stochastische Unabhängigkeit der W_j^{θ} jedoch nicht mehr gegeben und der klassische zentrale Grenzwertsatz nicht mehr anwendbar.

Die Folge $(W_j^{\theta})_{j \in \mathbb{N}}$ ist jedoch k-abhängig, falls θ der wahre Parameter ist, wie wir leicht einsehen können. Sei hierzu $i \in \mathbb{N}$ beliebig, aber fest.

 $W_1^{\theta}, \ldots, W_i^{\theta}$ sind Funktionen der Residuen $r(B_1, \theta), \ldots, r(B_{i+k}, \theta)$, und $W_{i+l}^{\theta}, W_{i+l+1}^{\theta}, \ldots$

mit l > k Funktionen der Residuen $r(B_{i+l}, \theta), r(B_{i+l+1}, \theta), \ldots$ Es gilt nun natürlich i+l > i+k für l > k, sodass aufgrund der stochastischen Unabhängigkeit der Residuen unter θ die k-Abhängigkeit folgt.

Wir zentrieren nun zu $Z_j^{\theta} = W_j^{\theta} - \left(\frac{1}{2}\right)^k$. Die Familie von Zufallsvariablen $(Z_j^{\theta})_{j \in \mathbb{N}}$ ist dann streng stationär mit Erwartungswert Null und endlicher Varianz. Die strenge Stationarität folgt unmittelbar aus der Unabhängigkeit und der identischen Verteilung der Fehler und damit der Residuen unter dem wahren θ .

Somit können wir den Grenzwertsatz 3.3.3 auf die Folge $(Z_j^\theta)_{j\in\mathbb{N}}$ anwenden.

Die Komponente τ^* berechnet sich für $(Z_j^{\theta})_{j \in \mathbb{N}}$ wie folgt:

$$\tau^* = Var_{\theta}\left(W_1^{\theta} - \left(\frac{1}{2}\right)^k\right) + \sum_{j=2}^{k+1} 2 \cdot E_{\theta}(Z_1^{\theta} Z_j^{\theta}) = Var_{\theta}(W_1^{\theta}) + \sum_{j=2}^{k+1} 2 \cdot E_{\theta}(Z_1^{\theta} Z_j^{\theta}).$$
(3.8)

 $E_{\theta}(Z_1^{\theta}Z_j^{\theta})$ ergibt sich für $j = 2, \dots, k+1$ als

$$E_{\theta}(Z_1^{\theta}Z_j^{\theta}) = E_{\theta} \left[\left(W_1^{\theta} - \left(\frac{1}{2}\right)^k \right) \left(W_j^{\theta} - \left(\frac{1}{2}\right)^k \right) \right]$$
$$= E_{\theta} \left[W_1^{\theta} W_j^{\theta} \right] - \left(\frac{1}{2}\right)^k E_{\theta}[W_1^{\theta}] \underbrace{- \left(\frac{1}{2}\right)^k E_{\theta}[W_j^{\theta}] + \left(\frac{1}{2}\right)^{2k}}_{=0}$$
$$= \left(\frac{1}{2}\right)^{k+j-1} - \left(\frac{1}{2}\right)^{2k}.$$

Der letzte Schritt der Gleichungskette folgt aus

$$W_1^{\theta} W_j^{\theta} = \prod_{i=1}^{k+j} \mathbb{1} \left((r(B_i, \theta)(-1)^i > 0) + \prod_{i=1}^{k+j} \mathbb{1} \left(r(B_i, \theta)(-1)^{i+1} > 0 \right) \right)$$

Genauso wie im Beweis zu Teil 1.), Gleichung 3.7, erhalten wir dann $E_{\theta}(W_1^{\theta}W_j^{\theta}) = (\frac{1}{2})^{k+j-1}$, wenn wir die Summationsgrenze (k+1) in 3.7 durch k+j ersetzen.

Einsetzen des soeben Berechneten in Gleichung 3.8 ergibt

$$\begin{aligned} \tau^* &= \sum_{j=2}^{k+1} 2 \cdot \left[\left(\frac{1}{2} \right)^{k+j-1} - \left(\frac{1}{2} \right)^{2k} \right] + \left(\frac{1}{2} \right)^k \left(1 - \left(\frac{1}{2} \right)^k \right) \\ &= \left(\frac{1}{2} \right)^k \left[\sum_{j=2}^{k+1} \left(\frac{1}{2} \right)^{j-2} - k \cdot \left(\frac{1}{2} \right)^{k-1} + 1 - \left(\frac{1}{2} \right)^k \right] \\ &\stackrel{v=j-2}{=} \left(\frac{1}{2} \right)^k \left[\sum_{v=0}^{k-1} \left(\frac{1}{2} \right)^v - k \cdot \left(\frac{1}{2} \right)^{k-1} + 1 - \left(\frac{1}{2} \right)^k \right] \\ &= \left(\frac{1}{2} \right)^k \left[\frac{1 - \left(\frac{1}{2} \right)^k}{\left(\frac{1}{2} \right)^k} - k \cdot \left(\frac{1}{2} \right)^{k-1} + 1 - \left(\frac{1}{2} \right)^k \right] \\ &= \left(\frac{1}{2} \right)^k \left[3 - k \cdot \left(\frac{1}{2} \right)^{k-1} - k \cdot \left(\frac{1}{2} \right)^{k-1} + 1 - \left(\frac{1}{2} \right)^k \right] \\ &= \left(\frac{1}{2} \right)^k \left[3 - k \cdot \left(\frac{1}{2} \right)^{k-1} - \left(\frac{1}{2} \right)^{k-1} \left(1 + \frac{1}{2} \right) \right] \\ &= \left(\frac{1}{2} \right)^k \left[3 - \left(\frac{1}{2} \right)^{k-1} \cdot k - \left(\frac{1}{2} \right)^{k-1} \left(3 \cdot \left(\frac{1}{2} \right) \right) \right] \\ &= \left(\frac{1}{2} \right)^k \left[3 - \left(\frac{1}{2} \right)^{k-1} \cdot k - 3 \cdot \left(\frac{1}{2} \right)^k \right]. \end{aligned}$$

Also ergibt sich die Komponente τ für die Folge $(W_j)_{j \in \mathbb{N}}$ aufgrund der Invarianz von τ^* gegenüber Verschiebungen gerade als τ^* . Mit dem Satz 3.3.3 ist die Aussage somit bewiesen.

Zu 3.)

Wir setzen

$$W_j^{\theta} = \mathbb{1}\left(r(B_j, \theta) > 0\right) \mathbb{1}\left(r(B_{\lfloor \frac{n+1}{2} \rfloor}, \theta) < 0\right) \mathbb{1}\left(r(B_{n-j+1}, \theta) > 0\right) + \mathbb{1}\left(r(B_j, \theta) < 0\right) \mathbb{1}\left(r(B_{\lfloor \frac{n+1}{2} \rfloor}, \theta) > 0\right) \mathbb{1}\left(r(B_{n-j+1}, \theta) < 0\right).$$

Wieder folgt wie in 1.), angewandt auf genau 3 Faktoren, dass W_j^{θ} Bernoulli-Variablen sind mit $P_{\theta}(W_j^{\theta} = 1) = \frac{1}{4}, j = 1, \dots, \lfloor \frac{n+1}{2} \rfloor - 1$. Wir zeigen im Folgenden, dass die Familie von Zufallsvariablen W_j^{θ} stochastisch unabhängig ist. Die Behauptung folgt dann wieder wie in 1.) aus dem zentralen Grenzwertsatz.

Zunächst gilt, für $j = 1, \ldots, \lfloor \frac{n+1}{2} \rfloor - 1$, aufgrund der stochastischen Unabhängigkeit der

Fehler

$$P_{\theta} \left(W_{j}^{\theta} = 0 \mid \epsilon_{\lfloor \frac{n-1}{2} \rfloor} > 0 \right)$$

= $P_{\theta} \left(\{ \epsilon_{j} > 0 \land \epsilon_{n-j+1} > 0 \} \cup \{ \epsilon_{j} > 0 \land \epsilon_{n-j+1} < 0 \} \cup \{ \epsilon_{j} < 0 \land \epsilon_{n-j+1} > 0 \} \right)$
= $P_{\theta}(\epsilon_{j} > 0 \land \epsilon_{n-j+1} > 0) + P_{\theta}(\epsilon_{j} > 0 \land \epsilon_{n-j+1} < 0) + P_{\theta}(\epsilon_{j} < 0 \land \epsilon_{n-j+1} > 0)$
= $\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4} = P_{\theta}(W_{j}^{\theta} = 0).$

Für $P_{\theta}(W_j^{\theta} = 0)$, bedingt auf $\epsilon_{\lfloor \frac{n+1}{2} \rfloor} < 0$, gilt ebenso

$$P_{\theta}\left(W_{j}^{\theta}=0 \mid \epsilon_{\lfloor \frac{n+1}{2} \rfloor} < 0\right)$$

= $P_{\theta}\left(\left\{\epsilon_{j} < 0 \land \epsilon_{n-j+1} < 0\right\} \cup \left\{\epsilon_{j} > 0 \land \epsilon_{n-j+1} < 0\right\} \cup \left\{\epsilon_{j} < 0 \land \epsilon_{n-j-1} > 0\right\}\right)$
= $\frac{3}{4} = P_{\theta}(W_{j}^{\theta}=0).$

Auf analoge Weise erhalten wir

$$P_{\theta}\left(W_{j}^{\theta}=1 \mid \epsilon_{\lfloor \frac{n+1}{2} \rfloor} > 0\right) = P_{\theta}\left(W_{j}^{\theta}=1 \mid \epsilon_{\lfloor \frac{n+1}{2} \rfloor} < 0\right) = \frac{1}{4} = P_{\theta}(W_{j}^{\theta}=1).$$

Zudem sind $W_1^{\theta}, \dots, W_{\lfloor \frac{n+1}{2} \rfloor - 1}^{\theta}$ bedingt stochastisch unabhängig unter der Bernoulli-Variable E mit

$$E := \begin{cases} 1 & \text{falls } \epsilon_{\lfloor \frac{n+1}{2} \rfloor} > 0 \\ 0 & \text{falls } \epsilon_{\lfloor \frac{n+1}{2} \rfloor} < 0. \end{cases}$$

Es gilt also, mit $n^* := \lfloor \frac{n+1}{2} \rfloor - 1$, für jedes Tupel von Ereignissen $\boldsymbol{\omega} := (\omega_1, \dots, \omega_{n^*}) \in \{0, 1\}^{n^*}$

$$P_{\theta}(W_1^{\theta} = \omega_1, \dots, W_{n^*}^{\theta} = \omega_{n^*} \mid E = a) = \prod_{i=1}^{n^*} P_{\theta}(W_i^{\theta} \mid E = a) \quad \text{für } a = 0, 1.$$

Dies können wir leicht folgendermaßen einsehen: Wir definieren die Ereignisse $A_i^{(1)}$ und $A_i^{(0)}$ durch

$$A_i^{(1)} := \{ \epsilon_i < 0 \land \epsilon_{n-i+1} < 0 \}$$

$$A_i^{(0)} := \{ \{ \epsilon_i > 0 \land \epsilon_{n-i+1} > 0 \} \text{ oder } \{ \epsilon_i > 0 \land \epsilon_{n-i+1} < 0 \} \text{ oder } \{ \epsilon_i < 0 \land \epsilon_{n-i+1} > 0 \} \}.$$

3. Die vereinfachte Simplex-Datentiefe

Dann sind $A_1^{(\omega_1)}, \ldots, A_{n^*}^{(\omega_{n^*})}$ stochastisch unabhängig für alle $\boldsymbol{\omega} \in \{0, 1\}^{n^*}$, da die Fehler stochastisch unabhängig sind. Da nach obigen Resultaten

$$P_{\theta}(W_i^{\theta} = \omega_i \mid E = 1) = P_{\theta}(A_i^{(\omega_i)})$$

gilt, erhalten wir das Ergebnis

$$P_{\theta}(W_{1}^{\theta} = \omega_{1}, \dots, W_{n^{*}}^{\theta} = \omega_{n^{*}} \mid E = 1) = P_{\theta}(\bigcap_{i=1}^{n^{*}} A_{i}^{(\omega_{i})}) = \prod_{i=1}^{n^{*}} P_{\theta}(A_{i}^{(\omega_{i})}).$$

Eine analoge Überlegung können wir für die Bedingung E = 0 anstellen. Insgesamt folgt daher die behauptete bedingte Unabhängigkeit.

Zusammen mit dem Satz von der totalen Wahrscheinlichkeit ergibt sich wir abschließend

$$\begin{split} P_{\theta}(W_{1}^{\theta} = \omega_{1}, \dots, W_{n^{*}}^{\theta} = \omega_{n^{*}}) \\ = P_{\theta}(W_{1}^{\theta} = \omega_{1}, \dots, W_{n^{*}}^{\theta} = \omega_{n^{*}} \mid E = 1) P_{\theta}(E = 1) + \\ P_{\theta}(W_{1}^{\theta} = \omega_{1}, \dots, W_{n^{*}}^{\theta} = \omega_{n^{*}} \mid E = 0) P_{\theta}(E = 0) \\ = \prod_{i=1}^{n^{*}} P_{\theta}(W_{i}^{\theta} = \omega_{i} \mid E = 1) \cdot \frac{1}{2} + \prod_{i=1}^{n^{*}} P_{\theta}(W_{i}^{\theta} = \omega_{i} \mid E = 0) \cdot \frac{1}{2} \\ = \frac{1}{2} \prod_{i=1}^{n^{*}} P_{\theta}(W_{i}^{\theta} = \omega_{i}) + \frac{1}{2} \prod_{i=1}^{n^{*}} P_{\theta}(W_{i}^{\theta} = \omega_{i}) \\ = \prod_{i=1}^{n^{*}} P_{\theta}(W_{i}^{\theta} = \omega_{i}). \end{split}$$

An dieser Stelle weisen wir noch einmal auf die in Kapitel 2 gegebene, in (2.2) formulierte allgemeine Form eines auf der Simplex-Datentiefe aufbauenden Hypothesentests hin. Nach den soeben bewiesenen asymptotischen Resultaten ergeben sich die auf $d_{s,n}^{(i)} = d_s^{(i)}$, i =1,2,3, aufbauenden approximativen Tests bei gegebenem Stichprobenumfang n wie folgt:

$$arphi(oldsymbol{b}) = egin{cases} 1 & & \leq q^{(i)}_{lpha,n} \ & & ext{falls } \sup_{ heta \in \Theta_0} d^{(i)}_{s,n}(heta, oldsymbol{b}) \ & & & \ 0 & & > q^{(i)}_{lpha,n} \end{cases}$$

 mit

$$\begin{split} q_{\alpha,n}^{(i)} &= \alpha \text{-} \text{Quantil der } \mathcal{N}\left(\left(\frac{1}{2}\right)^k, \frac{\left(\frac{1}{2}\right)^k \left(1 - \left(\frac{1}{2}\right)^k\right)}{\lfloor\frac{n}{k+1}\rfloor}\right), \\ q_{\alpha,n}^{(i)} &= \alpha \text{-} \text{Quantil der } \mathcal{N}\left(\left(\frac{1}{2}\right)^k, \frac{\left(\frac{1}{2}\right)^k \left(3 - \left(\frac{1}{2}\right)^{k-1} \cdot k - 3 \cdot \left(\frac{1}{2}\right)^k\right)}{n-k}\right), \text{ und} \\ q_{\alpha,n}^{(i)} &= \alpha \text{-} \text{Quantil der } \mathcal{N}\left(\frac{1}{4}, \frac{\frac{3}{16}}{\lfloor\frac{n-1}{2}\rfloor}\right). \end{split}$$

Eine äquivalente Form erhalten wir offensichtlich, wenn für $d_{s,n}^{(i)}$ die Statistik $D_n^{(i)}$ eingesetzt und für $q_{\alpha,n}^{(i)}$ das α -Quantil der Standardnormalverteilung benutzt wird. Diese letzte Darstellung soll im weiteren Verlauf der Arbeit verwendet werden. Den entsprechenden Test bezeichnen wir im Folgenden mit $\varphi_n^{(D_i)}$.

3.4. Konsistenz der Tests

Der folgende Abschnitt folgt wieder dem Aufsatz von Kustosz et al. (2015). Für den wichtigen Fall der einelementigen Nullhypothese $H_0: \theta = \theta_0$ in 2.7, also $\Theta_0 = \{\theta_0\}$, werden wir in diesem Abschnitt die Konsistenz der Tests $\varphi_n^{(D_i)}$, i = 1, 2, 3, herleiten. Hierzu werden wir zwei verschiedene Voraussetzungen, die die Zufallsvariablen X_i aus 3.2 erfüllen sollen, betrachten, und die Konsistenz der drei Teststatistiken unter diesen Voraussetzungen untersuchen.

Unter obiger Nullhypothese wird der Test $\varphi_n^{(D_i)}$ dabei konsistent für $\theta' \in \Theta_1$, also für $\theta' \neq \theta_0$, genannt, falls der Grenzwert gilt:

$$\lim_{n \to \infty} P_{\theta'}(\varphi_n^{(D_i)} = 1) = 1.$$

Wir formulieren nun die zwei verschiedenen Anforderungen an die Zufallsvariablen X_i .

Voraussetzung 1 formuliert ein endliches Szenario und fordert Folgendes:

$$\exists a, b \in \mathbb{R}$$
, sodass gilt: $a \leq X_1, \ldots, X_n \leq b$ fast sicher $\forall n \in \mathbb{N}$.
In **Voraussetzung 2** wird hingegen die Nichtexistenz einer oberen Schranke explizit gefordert:

$$\forall b \in \mathbb{R} \exists n_0 \in \mathbb{N}$$
, sodass gilt: $X_i > b$ fast sicher $\forall n \ge n_0, I = 1, \dots, n$.

Nun geben wir das zentrale Ergebnis über die Konsistenz der Teststatistiken unter den zwei verschiedenen Voraussetzungen an.

3.4.1 Satz (Kustosz et al., 2015, S.8): Seien n Beobachtungen aus Modell 3.1 gegeben. Sei $\theta' \neq \theta_0$.

1. Es gelte die Voraussetzung 1. Die Fehler ϵ_i seien symmetrisch um Null verteilt mit Träger \mathbb{R} . Sei außerdem folgende Bedingung erfüllt:

 $\exists d > 0, \delta > 0$ und eine Zerlegung von [a, b[mit $[a, b[=\bigcup_{i=1}^{I} [a_i, b_i], sodass gilt für alle i = 1, \dots, I:$

 $f(\cdot,\theta_0) - f(\cdot,\theta') \text{ monoton in } [a_i,b_i[, \text{ und } |f(\cdot,\theta_0) - f(\cdot,\theta')| \ge d \text{ in } [a_i + \delta, b_i - \delta[\neq \emptyset.$

Dann gilt: Die Tests $\varphi_n^{D_1}$ und $\varphi_n^{D_2}$ sind konsistent in θ' für die Hypothesen $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$.

2. Es gelte die Voraussetzung 2. Wenn dann für alle $n \in \mathbb{N}$ erfüllt ist

$$\exists a \neq 0: f(X_i, \theta_0) - f(X_i, \theta') = a \ \forall i = 1, \dots, n, \ und \ P_{\theta'}(\epsilon_i > a) \neq \frac{1}{2},$$

dann gilt wieder: Die Tests $\varphi_n^{D_1}$ und $\varphi_n^{D_2}$ sind konsistent in θ' für die Hypothesen $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$.

3. Es gelte die Voraussetzung 2. Wenn erfüllt ist

$$\lim_{c \to \infty} f(c, \theta_0) - f(c, \theta') = \infty \ oder \ \lim_{c \to \infty} f(c, \theta_0) - f(c, \theta') = -\infty,$$

dann sind $\varphi_n^{D_1}$, $\varphi_n^{D_2}$ und $\varphi_n^{D_3}$ konsistent in θ' für die Hypothesen $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0.$ Bevor wir den Beweis zu obigem Satz führen, geben wir zuerst zwei Lemmata an, die uns beim Beweis des Konsistenzsatzes helfen werden.

Das erste Lemma formuliert zwei Bedingungen, bei deren Erfüllung auf Konsistenz geschlossen werden kann.

3.4.2 Lemma (vgl. Kustosz et al., 2015, S.29):

Seien n Beobachtungen gemäß Modell 3.1 gegeben. Seien $\theta_0 \in \mathbb{R}^k$ und $i \in \{1, 2, 3\}$ beliebig. Wenn dann für alle $\epsilon > 0$ Zahlen $N_0 \in \mathbb{N}$, $\delta_0 > 0$ und eine Abbildung $D_n^{*(i)}$ mit $D_n^{*(i)}(\boldsymbol{x}) \geq D_n^{(i)}(\boldsymbol{x}, \theta_0) \ \forall \ \boldsymbol{x} \in \mathbb{R}^{n \times 2}$ existieren, sodass die Ungleichung

$$E_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B})\right) \le -\sqrt{n}\left(\delta_0 - \frac{u_\alpha}{\sqrt{N_0}}\right),\tag{3.9}$$

und

$$Var_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B})\right) \le \epsilon n \delta_0^2$$
 (3.10)

erfüllt ist $\forall n > N_0$, dann gilt für alle i = 1, 2, 3 Folgendes:

 $\varphi_n^{D_i}$ ist konsistent in $\theta' \neq \theta_0$ für die Hypothesen $H_0: \theta = \theta_0$ vs. $H_1: \theta \neq \theta_0$, und das sowohl unter Voraussetzung 1 als auch unter Voraussetzung 2. Hierbei bezeichnet u_{α} das α -Quantil der Standardnormalverteilung.

Beweis (vgl. Kustosz et al., 2015, S.29):

Für den Beweis ziehen wir Teschebyscheffs Ungleichung heran, die bekanntlich besagt:

$$P(|X - E(X)| \ge k) \le \frac{Var(X)}{k^2} \quad \forall \ k > 0 \in \mathbb{R}.$$

Hiermit ergibt sich für alle $N \ge N_0$, unter Beachtung von $\frac{u_{\alpha}}{\sqrt{N}} \ge \frac{u_{\alpha}}{\sqrt{N_0}}$ (vorausgesetzt, das Quantil ist negativ, was bei den üblichen Testniveaus aber erfüllt ist), folgende Abschätzung:

$$\begin{aligned} P_{\theta'}\left(D_n^{(i)}(\boldsymbol{B},\theta_0) \ge u_{\alpha}\right) &\leq P_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B}) \ge u_{\alpha}\right) \\ &\leq P_{\theta'}\left(\left|D_n^{*(i)}(\boldsymbol{B}) - E_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B})\right)\right| \ge u_{\alpha} - E_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B})\right)\right) \\ &\leq P_{\theta'}\left(\left|D_n^{*(i)}(\boldsymbol{B}) - E_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B})\right)\right| \ge \sqrt{n}\frac{u_{\alpha}}{\sqrt{n}} + \sqrt{n}\left(\delta_0 - \frac{u_{\alpha}}{\sqrt{n_0}}\right)\right) \\ &\leq P_{\theta'}\left(\left|D_n^{*(i)}(\boldsymbol{B}) - E_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B})\right)\right| \ge \sqrt{n}\frac{u_{\alpha}}{\sqrt{n}} - \sqrt{n}\frac{u_{\alpha}}{\sqrt{n}} + \sqrt{n}\delta_0\right) \\ &= P_{\theta'}\left(\left|D_n^{*(i)}(\boldsymbol{B}) - E_{\theta'}\left(D_n^{*(i)}(\boldsymbol{B})\right)\right| \ge \sqrt{n}\delta_0\right) \\ &\leq \frac{Var_{\theta'}(D_n^{*(i)}(\boldsymbol{B}))}{n\delta_0^2} \le \frac{\epsilon n\delta_0^2}{n\delta_0^2} = \epsilon. \end{aligned}$$

Durch Grenzwertbildung und Wahl einer Folge ϵ_n mit $\lim_{n\to\infty}\epsilon_n = 0$ erhalten wir $\lim_{n\to\infty} P_{\theta'} \left(\varphi_n^{D_i} = 0\right) = 0$. Wir haben somit die Behauptung bewiesen.

3.4.3 Definition (vgl. Kustosz et al., 2015, S.7):

Wir definieren $I_{n(i)}$ als die Menge der in $d_{s,n}^{(i)}$ betrachteten Indexmengen, i = 1, 2, 3:

1.
$$I_{n(1)} := \left\{ \left(1, \dots, k+1\right), \dots, \left(\left\lfloor \frac{n}{k+1} \rfloor - 1\right) \cdot (k+1) + 1, \dots, \lfloor \frac{n}{k+1} \rfloor \cdot (k+1)\right) \right\},\$$

2. $I_{n(2)} := \left\{ (1, \dots, k+1), (2, \dots, k+2), \dots, (n-k, \dots, n) \right\},\$
3. $I_{n(3)} := \left\{ \left(1, \lfloor \frac{n+1}{2} \rfloor, n\right), \left(2, \lfloor \frac{n+1}{2} \rfloor, n-1\right), \dots, \left(\lfloor \frac{n-1}{2} \rfloor, \lfloor \frac{n+1}{2} \rfloor, n-\lfloor \frac{n-1}{2} \rfloor + 1\right) \right\}.$

Das zweite Lemma hilft beim Beweis von Teil 1 des Satzes 3.4.1, da wir für die Konsistenz der Tests $\varphi_n^{(i)}$, i = 1, 2, 3, für H_0 : $\theta = \theta_0$ vs. H_1 : $\theta \neq \theta_0$, wie wir sehen werden, zeigen müssen, dass die Abschätzung $E_{\theta'}(d_{s,n}^{(i)}(\theta_0, \mathbf{B})) < (\frac{1}{2})^k$ für alle bis auf endlich viele n gilt. Eine hinreichende Bedingung dafür ist, dass der Erwartungswert aller Summanden unter θ' diese Schranke einhält, d.h. dass

$$E_{\theta'}\left(\prod_{l=1}^{k+1} \mathbb{1}\left(r(B_{j_l},\theta_0)(-1)^l > 0\right) + \prod_{l=1}^{k+1} \mathbb{1}\left(r(B_{j_l},\theta_0)(-1)^{l+1} > 0\right)\right) < \left(\frac{1}{2}\right)^k$$
(3.11)

erfüllt ist für alle bis auf endlich viele Tupel $(j_1, \ldots, j_{k+1}) \in I_{n(i)}, i = 1, 2, 3$, wenn wir n gegen unendlich laufen lassen.

Da unter $P_{\theta'}$ die Äquivalenzkette

$$r(B_i, \theta_0) \stackrel{\leq}{>} 0 \Leftrightarrow Y_i - f(X_i, \theta_0) \stackrel{\leq}{>} 0 \Leftrightarrow Y_i - f(X_i, \theta') \stackrel{\leq}{>} f(X_i, \theta_0) - f(X_i, \theta')$$
$$\Leftrightarrow \epsilon_i \stackrel{\leq}{>} f(X_i, \theta_0) - f(X_i, \theta') \tag{3.12}$$

gilt, ist das Verhalten von $f(\cdot, \theta_0) - f(\cdot, \theta')$ von entscheidender Bedeutung. Wir wollen ein Beispiel angeben, bei dem die Abschätzung 3.11 nicht eingehalten wird (vgl. Kustosz et al., 2015, S.8).

Dazu betrachten wir eine quadratische lineare Regression ohne linearen Term, also die Modellgleichung

$$Y_i = f(x_i, \theta) + \epsilon_i = \theta_0 + \theta_2 x_i^2 + \epsilon_i, \quad \theta = (\theta_0, \theta_2) \in \mathbb{R}^2,$$

also k = 2. Wir wählen $\theta_0 = (1, 0)$, und $\theta' = (1, 1)$. Dann erhalten wir $f(x_i, \theta_0) - f(x_i, \theta') = -x_i^2$, sowie für 3.11 den Ausdruck

$$P_{\theta'}(\epsilon_{j_1} < -x_{j_1}^2, \epsilon_{j_2} > -x_{j_2}^2, \epsilon_{j_3} < -x_{j_3}^2) + P_{\theta'}(\epsilon_{j_1} > -x_{j_1}^2, \epsilon_{j_2} < -x_{j_2}^2, \epsilon_{j_3} > -x_{j_3}^2).$$
(3.13)

Dieser Ausdruck wird unter Voraussetzung 2 mit wachsendem n aufgrund der Unbeschränktheit der X_i und der Wachstumsbedingung 3.2 beliebig klein für fast alle (j_1, \ldots, j_3) . Unter Voraussetzung 1 kann jedoch die Situation eintreten, dass $x_{j_2} = 0$ und $x_{j_1} = -x_{j_3} > 0$ für alle bis auf endlich viele Tupel $(j_1, j_2, j_3) \in I_{n(3)}$. Hier ist j_2 fest, sodass kein Konflikt mit der Wachstumsbedingung besteht. Für i = 1, 2 hingegen gibt es fast sicher höchstens ein Tupel (j_1, j_2, j_3) , welches obige Bedingung erfüllt.

Wir ziehen für i = 3 folgende Konsequenz aus dem möglichen Eintreten der oben beschriebenen Situation: Mit der Festsetzung $P_{\theta'}(\epsilon_{j_1} > -x_{j_1}^2) = \frac{1}{2} + \nu$ wird die Gleichung 3.13 zu

$$\begin{pmatrix} \frac{1}{2} - \nu \end{pmatrix} \frac{1}{2} \begin{pmatrix} \frac{1}{2} - \nu \end{pmatrix} + \begin{pmatrix} \frac{1}{2} + \nu \end{pmatrix} \frac{1}{2} \begin{pmatrix} \frac{1}{2} + \nu \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{4} - \frac{1}{2}\nu \end{pmatrix} \begin{pmatrix} \frac{1}{2} - \nu \end{pmatrix} + \begin{pmatrix} \frac{1}{4} + \frac{1}{2}\nu \end{pmatrix} \begin{pmatrix} \frac{1}{2} + \nu \end{pmatrix}$$
$$= \frac{1}{8} - \frac{1}{2}\nu + \frac{1}{2}\nu^2 + \frac{1}{8} + \frac{1}{2}\nu + \frac{1}{2}\nu^2$$
$$= \frac{1}{4} + \nu^2 > \begin{pmatrix} \frac{1}{2} \end{pmatrix}^k = \frac{1}{4}.$$

Hierbei wurden um Null symmetrische Fehler vorausgesetzt.

Mit dem folgenden Lemma führen wir nun eine Monotonitätsbedingung an $f(x_i, \theta_0) - f(x_i, \theta')$, i = 1, ..., n, ein, die 3.11 durch ein $\alpha \leq \left(\frac{1}{2}\right)^k$ beschränkt, und die in einigen Fällen auch gewährleistet, dass $\alpha < \left(\frac{1}{2}\right)^k$ wird.

3.4.4 Lemma (Kustosz et al., 2015, S.8):

Es gelte Voraussetzung A und die Fehler $\epsilon_1, \ldots, \epsilon_n$ seien symmetrisch um Null verteilt. Gegeben sei das (k + 1)-Tupel $(j_1, \ldots, j_{k+1}) \subseteq \{1, \ldots, n\}$, mit $j_1 < j_2 < \ldots < j_{k+1}$. Weiterhin sei a_l definiert als

$$a_l := f(x_{j_l}, \theta_0) - f(x_{j_l}, \theta'), \ l = 1, \dots, k+1.$$

Falls erfüllt ist:

 $a_1 \leq a_2 \leq \ldots \leq a_{k+1} \text{ oder } a_1 \geq a_2 \geq \ldots \geq a_{k+1},$

sowie

 $\exists a_0 \in \mathbb{R} \ge 0: a_l \ge a_0 \quad oder \quad a_l \le -a_0 \quad \forall \ l = 1, \dots, k+1,$

dann gilt, mit $\nu_0 := \frac{1}{2} - P_{\theta'}(\epsilon_l > a_0) \ge 0$, die Beschränktheit von 3.11:

$$E_{\theta'}\left(\prod_{l=1}^{k+1} \mathbb{1}\left(r(B_{j_l},\theta_0)(-1)^l > 0\right) + \prod_{l=1}^{k+1} \mathbb{1}\left(r(B_{j_l},\theta_0)(-1)^{l+1} > 0\right)\right) \le \left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^{k-2}\nu_0^2$$

Beweis (vgl. Kustosz et al., 2015, S.29):

Ohne Beschränkung der Allgemeinheit gelte jeweils die erste der beiden erforderlichen Bedingungen, d.h. $0 \le a_0 \le a_1 \le a_2 \le \ldots \le a_{k+1}$. Aufgrund der stetigen und um Null symmetrischen Verteilung der Fehler gilt:

$$\nu_l := \frac{1}{2} - P_{\theta'}(\epsilon_{j_l} > a_l) = \frac{1}{2} - P_{\theta'}(\epsilon_{j_l} < -a_l) \ge 0 \ \forall \ l = 0, \dots, k+1.$$

Daraus folgt $P_{\theta'}(\epsilon_{j_l} > a_l) = \frac{1}{2} - \nu_l$ und $P_{\theta'}(\epsilon_{j_l} < a_l) = \frac{1}{2} + \nu_l$, letzteres wiederum aufgrund

der Symmetrie der Fehler. Weiter gilt aufgrund der angenommen Ordnung der a_l

$$0 \le \nu_0 \le \nu_1 \le \nu_2 \le \ldots \le \nu_{k+1} \le \frac{1}{2}$$

Mir 3.12 erhalten wir das Ergebnis

$$\begin{split} & E_{\theta'} \left(\prod_{l=1}^{k+1} \mathbbm{1} \left(r(B_{j_l}, \theta_0)(-1)^l > 0 \right) + \prod_{l=1}^{k+1} \mathbbm{1} \left(r(B_{j_l}, \theta_0)(-1)^{l+1} > 0 \right) \right) \\ &= E_{\theta'} \left(\prod_{l=1}^{k+1} \mathbbm{1} \left(\epsilon_{j_l}(-1)^l > a_l(-1)^l \right) + \prod_{l=1}^{k+1} \mathbbm{1} \left(\epsilon_{j_l}(-1)^{l+1} > a_l(-1)^{l+1} \right) \right) \\ &= \prod_{l=1}^{k+1} E_{\theta'} \left(\mathbbm{1} \left(\epsilon_{j_l}(-1)^l > a_l(-1)^{l+1} \right) \right) + \prod_{l=1}^{k+1} E_{\theta'} \left(\mathbbm{1} \left(\epsilon_{j_l}(-1)^{l+1} > a_l(-1)^l \right) \right) \\ &= \prod_{l=1}^{k+1} \left(\frac{1}{2} + (-1)^{l+1} \nu_l \right) + \prod_{l=1}^{k+1} \left(\frac{1}{2} + (-1)^l \nu_l \right). \end{split}$$

Die Behauptung haben wir bewiesen, wenn wir die Ungleichung

$$\prod_{l=1}^{k+1} \left(\frac{1}{2} + (-1)^{l+1}\nu_l\right) + \prod_{l=1}^{k+1} \left(\frac{1}{2} + (-1)^l\nu_l\right) \le \left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^{k-2}\nu_1^2 \le \left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^{k-2}\nu_0^2$$

zeigen, wobei das letzte Ungleichheitszeichen klar ist. Wir zeigen also die erste Ungleichung, welche, mit $y_l := 2 \cdot \nu_l$, äquivalent ist zu

$$\prod_{l=1}^{k+1} \left(1 + (-1)^{l+1} y_l \right) + \prod_{l=1}^{k+1} \left(1 + (-1)^l y_l \right) \le 2 - 2y_1^2, \tag{3.14}$$

was durch

$$\begin{split} &\prod_{l=1}^{k+1} \left(\frac{1}{2} + (-1)^{l+1} \nu_l \right) + \prod_{l=1}^{k+1} \left(\frac{1}{2} + (-1)^l \nu_l \right) \\ &= \prod_{l=1}^{k+1} \left(\frac{1}{2} \left(1 + (-1)^{l+1} y_l \right) \right) + \prod_{l=1}^{k+1} \left(\frac{1}{2} \left(1 + (-1)^l y_l \right) \right) \\ &= \left(\frac{1}{2} \right)^{k+1} \left[\prod_{l=1}^{k+1} \left(1 + (-1)^{l+1} y_l \right) + \prod_{l=1}^{k+1} \left(1 + (-1)^l y_l \right) \right] \end{split}$$

und durch

$$\left(\frac{1}{2}\right)^{-(k+1)} \left[\left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^{k-2} \nu_1^2 \right] = 2 - 8\left(\frac{1}{4}y_1^2\right) = 2 - 2y_1^2$$

leicht einzusehen ist. Die Substitution von $2 \cdot \nu_l$ durch y_l liefert die Ordnung $0 < y_1 \le y_2 \le \dots \le y_{k+1} \le 1$.

Im Folgenden werden wir also die Ungleichung 3.14 durch vollständige Induktion über k beweisen.

Induktionsanfang:

Hier können wir für den eindimensionalen Fall, also für k = 1, mithilfe obiger Ordnung der y_l leicht die Ungleichung

$$(1+y_1)(1-y_2) + (1-y_1)(1+y_2) = 2 - 2y_1y_2 \le 2 - 2y_1^2$$

verifizieren.

Induktionsschluss:

Wegen $y_{k+1} \ge y_k$ gilt

$$(1 - y_k)(1 + y_{k+1}) \ge (1 - y_{k+1})(1 + y_k) = (1 + y_k)(1 - y_{k+1})$$
(3.15)

sowie

$$(1+y_{k-1})(1-y_k)(1+y_{k+1}) \ge (1-y_{k-1})(1+y_k)(1-y_{k+1}).$$
(3.16)

Desweiteren benötigen wir noch eine allgemeine Ungleichung für vier reelle Zahlen a, b, cund d mit $0 < a \le b$ und $0 < c \le d$:

$$ad + bc \le ac + bd,\tag{3.17}$$

deren Gültigkeit wir durch die kurze Äquivalenzkette

$$ad + bc \le ac + bd \Leftrightarrow ad - ac \le bd - bc \Leftrightarrow a(d - c) \le b(d - c) \Leftrightarrow a \le bd$$

nachweisen.

Wir unterscheiden nun zwei Fälle.

1. Fall: Sei (k + 1) gerade. Wiederholte Anwendung von 3.15 liefert uns

$$\prod_{l=1}^{k+1} (1 + (-1)^l y_l) \ge \prod_{l=1}^{k+1} (1 + (-1)^{l+1} y_l)$$

Nun können wir den eigentlichen Induktionsschluss durchführen. Sei also die Induktionsvoraussetzung

$$\prod_{l=1}^{k+1} \left(1 + (-1)^{l+1} y_l \right) + \prod_{l=1}^{k+1} \left(1 + (-1)^l y_l \right) \le 2 - 2y_1^2$$

gegeben. Es ergibt sich

Fall 2: Im Falle, dass (k + 1) ungerade ist, ergibt sich durch wiederholte Multiplikation der beiden Seiten der Ungleichung 3.15 mit den beiden Seiten der Ungleichung 3.16 die Abschätzung

$$\prod_{l=1}^{k+1} (1 + (-1)^{l+1} y_l) \ge \prod_{l=1}^{k+1} (1 + (-1)^l y_l).$$

Die folgende Wahl von a, b, c und d für die Formel 3.17 führt auf einen vollständig analogen Induktionsschluss wie im Fall 1. Die Behauptung ist dann bewiesen.

$$\underbrace{\prod_{l=1}^{k+1} (1+(-1)^{l+1}y_l)}_{=:b} \underbrace{\underbrace{(1+(-1)^{k+3}y_{k+2})}_{=:c} + \underbrace{\prod_{l=1}^{k+1} (1+(-1)^{l}y_l)}_{=:a} \underbrace{(1+(-1)^{k+2}y_{k+2})}_{=:d}}_{=:a}$$

Nun schließlich beweisen wir den Satz 3.4.1.

Beweis zu Satz 3.4.1 (vgl. Kustosz et al., 2015, S.29 ff.): Zu 1.) Wir erinnern uns an die Definition 3.4.3 und definieren folgende Mengen:

$$I_{n(i)}^{1} := \{ (j_{1}, \dots, j_{k+1}) \in I_{n(i)} : \exists z = 1, \dots, I, \text{ sodass } x_{j_{1}}, \dots, x_{j_{k+1}} \in [a_{z} + \delta, b_{z} - \delta[\}, \\ I_{n(i)}^{2} := \{ (j_{1}, \dots, j_{k+1}) \in I_{n(i)} \setminus I_{n(i)}^{1} : \exists z = 1, \dots, I, \text{ sodass } x_{j_{1}}, \dots, x_{j_{k+1}} \in [a_{z}, b_{z}[\}, \\ I_{n(i)}^{3} := I_{n(i)} \setminus \left(I_{n(i)}^{1} \cup I_{n(i)}^{2} \right). \end{cases}$$

 $I_{n(i)}^1$ enthält also alle Indextupel aus $I_{n(i)}$, bei denen die korrespondierenden Realisierungen der unabhängigen Variablen in ein um den Betrag δ verkleinertes Intervall der Zerlegung von [a, b[fallen. $I_{n(i)}^2$ enthält alle Indextupel aus $I_{n(i)}$, bei denen die korrespondierenden Realisierungen der unabhängigen Variablen in eine komplette Partition fallen, aber nicht ausschließlich innerhalb der um den Betrag δ verkleinerten Intervalle aus $I_{n(i)}^1$ liegen.

 $I_{n(i)}^3$ enthält alle Indextupel aus $I_{n(i)}$, bei denen die korrespondierenden Realisierungen der unabhängigen Variablen eine der (I - 1) inneren Klassengrenzen der Zerlegung "schneiden". Die äußeren Intervallgrenzen können aufgrund von Voraussetzung 1 fast sicher nicht geschnitten werden.

Nun folgern wir aus den Voraussetzungen des zu beweisenden Satzes, dass Lemma 3.4.4 gilt, wenn wir $a_0 := d$ setzen, und zwar für alle (j_1, \ldots, j_{k+1}) aus $I^1_{n(i)}$. Daraus ergibt sich, dass 3.11 nach oben beschränkt ist durch $\left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^{k-2}\nu^2$, mit $\nu = \frac{1}{2} - P_{\theta'}(\epsilon_i > d) > 0$

für alle (k+1)-Tupel aus $I_{n(i)}^1$.

Für alle (j_1, \ldots, j_{k+1}) aus $I_{n(i)}^2$ erhält Lemma 3.11 Gültigkeit, wenn wir $a_0 := 0$ setzen. 3.11 ist somit nach oben beschränkt durch $\left(\frac{1}{2}\right)^k$ für alle (k+1)-Tupel aus $I_{n(i)}^2$. Da für die Tupel in $I_{n(i)}$ die Ungleichung $j_1 < j_2 < \ldots < j_{k+1}$ erfüllt ist, sind die korrespondierenden x_{j_l} aufgrund von 3.2 fast sicher streng monoton: $x_{j_1} < x_{j_2} < \ldots < x_{k+1}$. Für i = 1 bedeutet dies, dass $I_{n(i)}^3$ höchstens (I-1) Elemente hat, da aneinander angrenzende Indexgruppen betrachtet werden, und es an jeder der (I-1) inneren Klassengrenzen höchstens einmal passieren kann, dass eine betrachtete Gruppe von x-Werten sowohl in der einen als auch in der anderen Klasse Elemente platziert, dass eine Klassengrenze also geschnitten wird. Für i = 2 haben die Elemente die Form $(N, N + 1, \ldots, N + k)$ mit $N \in \{1, \ldots, n - k\}$. Somit kann es an jeder Klassengrenze höchstens k-mal passieren, dass diese geschnitten

somit kann es an jeder Klassengrenze nochstens k-mai passieren, dass diese geschnitten wird. Insgesamt ist dies also höchstens K(I-1) mal möglich. In jedem Falle ist die Anzahl der Elemente von $I_{n(i)}^3$ also unabhängig von n. Da jedes Indextupel in genau einer der oben definierten Mengen enthalten ist, erhalten wir somit:

$$\forall \ i = 1, 2 \ \exists \ p \in]0, 1[: \ \frac{\#I_{n(i)}^1}{I_{n(i)}} \to p, \ \frac{\#I_{n(i)}^2}{I_{n(i)}} \to 1-p, \ \frac{\#I_{n(i)}^3}{I_{n(i)}} \to 0$$

Mit der Tatsache, dass 3.11 sicher durch 2 nach oben beschränkt ist, erhalten wir schließlich die gewünschte Abschätzung für $E_{\theta'}(d_{s,n}^{(i)}(\theta_0, \boldsymbol{B}))$, und zwar

$$\begin{split} E_{\theta'}(d_{s,n}^{(i)}(\theta_0, \mathbf{B})) &\leq \frac{\#I_{n(i)}^1}{I_{n(i)}} \left(\left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^{k-2} \nu^2 \right) + \frac{\#I_{n(i)}^2}{I_{n(i)}} \left(\frac{1}{2}\right)^k + \frac{\#I_{n(i)}^3}{I_{n(i)}} \cdot 2\\ &\stackrel{n \to \infty}{\longrightarrow} p\left(\left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^{k-2} \nu^2 \right) + (1-p) \left(\frac{1}{2}\right)^k\\ &= \left(\frac{1}{2}\right)^k - p\left(\frac{1}{2}\right)^{k-2} \nu^2 < \left(\frac{1}{2}\right)^k. \end{split}$$

Also gibt es ein $\gamma < 0$ und ein $n_0 \in \mathbb{N}$, sodass für alle $n > n_0$

$$E_{\theta'}\left(d_s^{(i)}(\theta_0, \boldsymbol{B}) - \left(\frac{1}{2}\right)^k\right) \le \gamma < 0$$

gilt.

Falls nun $n \ge 2k$, folgt dadurch mit

$$n \geq 2k \Rightarrow -\frac{n}{2} \leq -k \Rightarrow -\frac{n}{2} + n \leq n - k \Rightarrow n - k \geq \frac{n}{2} \geq \frac{n}{2(k+1)}$$

und

$$\lfloor \frac{n}{k+1} \rfloor \ge \frac{n-k}{k+1} = \frac{2n-2k}{2(k+1)} \ge \frac{n}{2(k+1)}$$

eine Abschätzung des Erwartungswertes der Teststatistik $D_n^{(i)}(\boldsymbol{B}, \theta_0)$ für i = 1, 2:

$$E_{\theta'}\left(D_n^{(i)}(\boldsymbol{B},\theta_0)\right) \le \sqrt{\frac{n}{2(k+1)}} \frac{d_s^{(i)}(\theta_0,\boldsymbol{B}) - \left(\frac{1}{2}\right)^k}{V_i} = \sqrt{\frac{n}{2(k+1)}} \frac{\gamma}{V_i}$$

wobei V_i die Standardabweichung von $d_s^{(i)}$ unter θ_0 , also der Nenner von $D_n^{(i)}(\boldsymbol{B}, \theta_0)$, ist. Mit der Abkürzung $\gamma_i := \frac{\gamma}{\sqrt{2(k+1)V_i}}$ folgern wir dann $E_{\boldsymbol{B},\theta'}\left(D_n^{(i)}(\theta_0)\right) \leq \sqrt{n}\gamma_i$.

Falls $n = n_1$ so groß gewählt wird, dass $\sqrt{n_1} > \frac{u_{\alpha}}{\gamma_1}$ gilt, bzw. $n = n_2$ so groß, dass $\sqrt{n_2} > \frac{u_{\alpha}}{\gamma_2}$, dann wählen wir $\delta_i = \frac{u_{\alpha}}{\sqrt{n_i}} - \gamma_i > \frac{u_{\alpha}}{\frac{u_{\alpha}}{\gamma_i}} - \gamma_i = 0$.

Da unter θ' nun die Familie der $r(B_j, \theta_0)$, j = 1, ..., n, offensichtlich ebenfalls stochastisch unabhängig sind, folgt wie im Beweis von 3.3.1, dass die Summanden von $D_n^{(1)}(\boldsymbol{B}, \theta_0)$ stochastisch unabhängig und die Summanden von $D_n^{(2)}(\boldsymbol{B}, \theta_0)$ k-abhängig sind. Damit ergibt sich $Var_{\theta'}\left(D_n^{(i)}(\boldsymbol{B}, \theta_0)\right) = \sigma_i$, und für alle $n \ge \frac{\sigma_i}{\epsilon \delta_i^2}$:

$$Var_{\theta'}\left(D_n^{(i)}(\theta_0)\right) \le \epsilon \cdot n \cdot \delta_i^2, \quad i=1,2.$$

Damit sowohl die Abschätzung des Erwartungswertes als auch die der Varianz von $D_n^{(i)}(\boldsymbol{B}, \theta_0)$ gilt, wählen wir $n_{0i} := \max\left(\sqrt{n_i}, \frac{Var_{\theta'}\left(D_n^{(i)}(\boldsymbol{B}, \theta_0)\right)}{\epsilon \delta_i^2}\right)$. Zusammen mit $\delta_{0i} := \frac{u_{\alpha}}{\sqrt{n_{0i}}} - \gamma_i \geq \frac{u_{\alpha}}{\sqrt{n_{0i}}} - \gamma_i = \delta_i > 0$ haben wir nun Konstanten konstruiert, damit die erforderlichen Gleichungen von Lemma 3.4.2 erfüllt sind. Tatsächlich gilt für i = 1, 2 für alle $n > n_{0i}$

$$E_{\theta'}\left(D_n^{(i)}(\boldsymbol{B},\theta_0)\right) \le \sqrt{n}\gamma_i = -\sqrt{n}\left(\delta_{0i} - \frac{u_\alpha}{\sqrt{n_{0i}}}\right)$$

und

$$Var_{\theta'}\left(D_n^{(i)}(\boldsymbol{B},\theta_0)\right) \le \epsilon n_{0i}\delta_i^2 \le \epsilon n_{0i}\delta_{0i}^2 \le n\epsilon\delta_{0i}^2;$$

sodass $D_n^{*(i)} = D_n^{(i)}$ gewählt werden kann.

Zu 2.)

Zuerst erinnern wir uns an den Zusammenhang zwischen Residuen und Fehlern unter θ'

in Ungleichung 3.12. Zusammen mit den in diesem Teil des Konsistenzsatzes gegebenen Voraussetzungen lassen sich für i = 1, 2 die vereinfachten Simplex-Datentiefen darstellen als

$$\begin{aligned} d_{s,n}^{i}(\theta_{0},B) &= \\ & \frac{1}{\#I_{n(i)}} \sum_{(i_{1},\dots,i_{k+1})\in I_{n(i)}} \left(\prod_{j=1}^{k+1} \mathbbm{1}\left(\epsilon_{i_{j}}(-1)^{j} > a(-1)^{j}\right) + \prod_{j=1}^{k+1} \mathbbm{1}\left(\epsilon_{i_{j}}(-1)^{j+1} > a(-1)^{j+1}\right) \right). \end{aligned}$$

Wir betrachten $p := P_{\theta'}(\epsilon_i > a)$. Sei ohne Beschränkung der Allgemeinheit $p < \frac{1}{2}$. Es gilt damit, falls k ungerade ist,

$$E_{\theta'}\left(d_{s,n}^{(i)}(\theta, \mathbf{B})\right) = 2p^{\frac{k+1}{2}}(1-p)^{\frac{k+1}{2}} = 2\left(p(1-p)\right)^{\frac{k+1}{2}} < 2\left(\frac{1}{4}\right)^{\frac{k+1}{2}} = 2\left(\frac{1}{2}\right)^{k+1} = \left(\frac{1}{2}\right)^{k},$$

und falls k gerade ist,

$$E_{\theta'}\left(d_{s,n}^{(i)}(\theta,B)\right) = p^{\frac{k}{2}}(1-p)^{\frac{k+2}{2}} + (1-p)^{\frac{k}{2}}p^{\frac{k+2}{2}}$$
$$= p^{\frac{k}{2}}(1-p)^{\frac{k}{2}}\left((1-p)+p\right)$$
$$= p^{\frac{k}{2}}(1-p)^{\frac{k}{2}} = (p(1-p))^{\frac{k}{2}} < \left(\frac{1}{4}\right)^{\frac{k}{2}} = \left(\frac{1}{2}\right)^{k}.$$

Also erhalten wir die Abschätzung

$$E_{\theta'}\left(d_{s,n}^{(i)}(\theta,B) - \left(\frac{1}{2}\right)\right)^k =: \gamma < 0.$$

Komplett analog zum Beweis von 1.) erhält man die Voraussetzungen von Lemma 3.4.2 und damit die Konsistenz der beiden Tests.

Zu 3.)

Sei $\lim_{y\to\infty} f(y,\theta_0) - f(y,\theta') = \infty$. Dann existieren $y_0, \ \gamma > 0$ und $\alpha < \left(\frac{1}{2}\right)^k$, sodass gilt:

$$f(y,\theta_0) - f(y,\theta') > \gamma \quad \forall \ y > y_0 \text{ und } P_{\theta'}(\epsilon_i > \gamma) \le \frac{\alpha}{2} \quad \forall \ i = 1, \dots, n.$$

Da wir hier Voraussetzung 2 betrachten, gibt es ein $n_0 \in \mathbb{N}$ mit $X_i > y_0$ fast sicher für alle $i \ge n_0$. Somit gelangen wir zu folgenden Abschätzungen, zunächst für i = 1 (ohne

$3. \ Die \ vereinfachte \ Simplex-Datentiefe$

Beschränkung der Allgemeinheit nehmen wir an, dass $\lfloor \frac{n}{k+1} \rfloor > n_0)$:

Für i=2 gilt ebenso die Abschätzung

$$d_{s,n}^{(2)}(\theta, \mathbf{B}) \le \frac{1}{n-k} \left(n_0 + \sum_{j=n_0}^{n-k} \left(\mathbb{1} \left(\epsilon_j > \gamma \right) + \mathbb{1} \left(\epsilon_{j+1} > \gamma \right) \right) =: d_s^{*(2)}.$$

Im Falle von i=3erhalten wir schließlich genaus
o die Abschätzung

$$d_{s,n}^{(3)}(\theta, \boldsymbol{B}) \leq \frac{1}{\lfloor \frac{n-1}{2} \rfloor} \left(n_0 + \sum_{j=n_0}^{\lfloor \frac{n-1}{2} \rfloor} \left(\mathbbm{1}\left(\epsilon_{\lfloor \frac{j+1}{2} \rfloor} > \gamma \right) + \mathbbm{1}\left(\epsilon_{n-j+1} > \gamma \right) \right) =: d_s^{*(3)}.$$

3. Die vereinfachte Simplex-Datentiefe

Setzen wir nun $D_n^{*(i)} := \sqrt{n\lambda_i} \left(d_{s,n}^{*(i)} - \left(\frac{1}{2}\right)^k \right)$ für i = 1, 2, 3, so können wir $\lambda_i > 0$ so wählen, dass, wie in Lemma 3.4.2 gefordert, $D_n^{(i)} \le D_n^{*(i)}$ gilt.

Der Erwartungswert von $D_n^{*(i)}$ unter θ' ist, falls n gegen unendlich geht, kleiner als Null. Dies sehen wir, wenn wir $n_1 := \lfloor \frac{n}{k+1} \rfloor$, $n_2 := n - k$, $n_3 := \lfloor \frac{n-1}{2} \rfloor$ setzen, und für den Erwartungswert von $d_{s,n}^{*(i)}$ die Abschätzung

$$E_{\theta'}(d_{s,n}^{*(i)}) \le \frac{n_0}{n_i} + \frac{n_i - n_0}{n_i} 2\frac{\alpha}{2} \xrightarrow{n_i \to \infty} \alpha < \left(\frac{1}{2}\right)^k$$

erhalten. Dies impliziert $E(d_{s,n}^{*(i)}) < \left(\frac{1}{2}\right)^k$ für ein n_{0i} mit $n > n_{0i}$.

Somit ergibt sich die Gültigkeit von Bedingung 3.9 wie im Beweis zu 1.), und zunächst für i = 1, 2 die Gültigkeit von Bedingung 3.10, da die Summanden von $D_n^{*(1)}$ stochastisch unabhängig und die Summanden von $D_n^{*(2)}$ 1-abhängig sind. Um die Gültigkeit von 3.10 auch für i = 3 zeigen zu können, stellen wir folgende Überlegungen an: Wie wählen γ so groß, dass die Abschätzung

$$P_{\theta'}(\epsilon_i > \gamma)^2 \le P_{\theta'}(\epsilon_i > \gamma) \le \frac{\epsilon \delta_0^2}{4\lambda_3^2}$$

erfüllt ist. Aufgrund der unabhängig identisch verteilten Fehler erhalten wir dann

$$Var\left(D_{n}^{*(3)}\right) = n\lambda_{3}^{2}Var\left(d_{s,n}^{*(3)}\right) = n\lambda_{3}^{2}\left[E\left((d_{s,n}^{*(3)})^{2}\right) - E^{2}\left(d_{s,n}^{*(3)}\right)\right]$$

sowie

$$\begin{split} &E\left(\left(d_{s}^{*(3)}\right)^{2}\right) - E^{2}\left(d_{s}^{*(3)}\right) \\ &= E\left[\left(\left(\frac{n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\right) + \frac{\lfloor\frac{n-1}{2}\rfloor - n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\mathbbm{1}\left(\epsilon_{\lfloor\frac{i+1}{2}\rfloor} > \gamma\right) + \mathbbm{1}\left(\epsilon_{n-i+1} > \gamma\right)\right)^{2}\right]\right) \\ &- E^{2}\left[\left(\frac{n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\right) + \frac{\lfloor\frac{n-1}{2}\rfloor - n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\mathbbm{1}\left(\epsilon_{\lfloor\frac{i+1}{2}\rfloor} > \gamma\right) + \mathbbm{1}\left(\epsilon_{n-i+1} > \gamma\right)\right) \\ &= E\left[\left(\frac{n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\right)^{2} + 2\frac{n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\frac{\lfloor\frac{n-1}{2}\rfloor - n_{0}}{\lfloor\frac{n-1}{2}\rfloor}C + \left(\frac{\lfloor\frac{n-1}{2}\rfloor - n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\right)^{2}C^{2}\right] \\ &- E^{2}\left[\left(\frac{n_{0}}{\lfloor\frac{n-1}{2}\rfloor}\right) + \frac{\lfloor\frac{n-1}{2}\rfloor - n_{0}}{\lfloor\frac{n-1}{2}\rfloor}C\right] \end{split}$$

3. Die vereinfachte Simplex-Datentiefe

$$\begin{split} &= \left(\frac{n_0}{\lfloor \frac{n-1}{2} \rfloor}\right)^2 + 2\frac{n_0}{\lfloor \frac{n-1}{2} \rfloor} \frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor} E[C] + \left(\frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor}\right)^2 E[C^2] \\ &- \left(\frac{n_0}{\lfloor \frac{n-1}{2} \rfloor} + \frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor} E[C]\right)^2 \\ &= \left(\frac{n_0}{\lfloor \frac{n-1}{2} \rfloor}\right)^2 + 2\frac{n_0}{\lfloor \frac{n-1}{2} \rfloor} \frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor} E[C] + \left(\frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor}\right)^2 E[C^2] \\ &- \left(\frac{n_0}{\lfloor \frac{n-1}{2} \rfloor}\right)^2 - 2\left(\frac{n_0}{\lfloor \frac{n-1}{2} \rfloor}\right) \frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor} E[C] - \left(\frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor}\right)^2 E^2[C] \\ &= \left(\frac{\lfloor \frac{n-1}{2} \rfloor - n_0}{\lfloor \frac{n-1}{2} \rfloor}\right)^2 \left(E[C^2] - E^2[C]\right) \le E[C^2] - E^2[C] \le E[C^2]. \end{split}$$

Also folgt dann insgesamt die Abschätzung

$$\begin{aligned} \operatorname{Var}\left(D_{n}^{*(3)}\right) &\leq n\lambda_{3}^{2}E_{\theta'}\left[\left(\mathbbm{1}\left(\epsilon_{\lfloor\frac{i+1}{2}\rfloor} > \gamma\right) + \mathbbm{1}\left(\epsilon_{n-i+1} > \gamma\right)\right)^{2}\right] \\ &= n\lambda_{3}^{2}E_{\theta'}\left[\mathbbm{1}\left(\epsilon_{\lfloor\frac{i+1}{2}\rfloor} > \gamma\right) + 2\cdot\mathbbm{1}\left(\epsilon_{\lfloor\frac{i+1}{2}\rfloor} > \gamma\right)\mathbbm{1}\left(\epsilon_{n-i+1} > \gamma\right) \\ &\quad + \mathbbm{1}\left(\epsilon_{\lfloor\frac{i+1}{2}\rfloor} > \gamma\right)\right] \\ &= 2n\lambda_{3}^{2}\left(\left(P_{\theta'}(\epsilon_{n} > \gamma) + P_{\theta'}^{2}(\epsilon_{n} > \gamma)\right)\right) \leq 2n\lambda_{3}^{2}2\frac{\epsilon\delta_{0}^{2}}{\lambda_{3}^{2}4} = \epsilon n\delta_{0}^{2}.\end{aligned}$$

- 1	_	_	٦
- 1			

In diesem Kapitel wird das Verhalten der drei in Kapitel 3 betrachteten Tests mithilfe von computergenerierten (Pseudo-)Zufallszahlen simuliert. Zudem wird als Vergleich der bekannte Vorzeichentest, der im ersten Abschnitt nochmals dargestellt wird, herangezogen. Insgesamt werden drei spezielle Regressionsmodelle betrachtet.

Der zweite Abschnitt stellt die betreffenden Regressionsmodelle vor und illustriert sie grafisch. Desweiteren werden die Anwendbarkeit der vereinfachten Datentiefe und die Konsistenz der entsprechenden Tests für jedes Modell untersucht.

Der dritte Abschnitt beschreibt ausführlich die Vorgehensweise der Simulationsstudie. Die letzten drei Abschnitte schließlich stellen die zentralen Ergebnisse der Simulationsstudie für jedes Modell vor.

4.1. Der Vorzeichentest

Der Vorzeichentest ist ein nichtparametrischer Test auf den Median einer Verteilung, der mit wenigen Voraussetzungen auskommt und einfach zu handhaben ist.

Seien X_1, \ldots, X_n unabhängig identisch verteilte Stichprobenvariablen, deren Verteilungsfunktion ein Element der Familie

$$P_{(F,\theta)} := \{F(x-\theta): F \text{ stetige Verteilungsfunktion und } Median(F) = 0, \ \theta \in \mathbb{R}\}$$

ist. Bei beliebigem, aber festem F enthält die Familie alle Verteilungsfunktionen gleicher Form, die sich nur durch die Lage unterscheiden. Der Parameter θ ist hierbei der Median der Verteilung der jeweiligen X_i . Getestet werden die Hypothesen

$$H_0: \theta = \theta_0 \quad vs. \quad H_1: \theta \neq \theta_0 \tag{4.1}$$

für ein $\theta_0 \in \mathbb{R}$.

Die Idee des Vorzeichentests ist es nun, dass unter der Nullhypothese wegen der Stetigkeit von F die Vorzeichen der Differenzen $D_i := X_i - \theta_0$, i = 1, ..., n, mit Wahrscheinlichkeit $\frac{1}{2}$ positiv und mit Wahrscheinlichkeit $\frac{1}{2}$ negativ sind. Wenn wir also ohne Beschränkung der Allgemeinheit die positiven Vorzeichen von D_i zählen, spricht eine zu kleine oder eine zu große Anzahl gegen die Nullhypothese.

Infolgedessen verwendet der Vorzeichentest die Teststatistik $V(X_1, \ldots, X_n) := \sum_{i=1}^n v(D_i)$, mit v(x) = 1, falls x > 0, und v(x) = 0, falls x < 0. Die Teststatistik folgt unter der Nullhypothese einer Binomialverteilung mit Parametern n und $\frac{1}{2}$.

Insgesamt erhalten wir den Test

$$\varphi(X_1, \dots, X_n) = \begin{cases} 1 & \in [q_{\frac{\alpha}{2}}, q_{\frac{1-\alpha}{2}}] \\ & \text{falls } \sum_{i=1}^n v(D_i) \\ 0 & \text{ sonst.} \end{cases}$$
(4.2)

Hierbei ist q_{α} das α -Quantil der Binomialverteilung mit Parametern n und $\frac{1}{2}$ (vgl. Büning und Trenkler, 1994, S.92 ff.).

Nun betrachten wir den Vorzeichentest im speziellen Modell 3.1 aus Kapitel 3. Wir setzen hier symmetrisch um Null verteilte Fehler voraus.

Wir interessieren uns wieder für die Hypothese 4.1. Dazu betrachten wir die Residuen $r(\boldsymbol{b}, \theta_0)$ als Stichprobe, d.h. als Realisationen der Zufallsvariablen $r(\boldsymbol{B}, \theta_0)$. Im Rahmen des Vorzeichentests testen wir dann die Hypothesen

$$H_0$$
: $Median(r(\boldsymbol{B}, \theta_0)) = 0$ vs. H_1 : $Median(r(\boldsymbol{B}, \theta_0)) \neq 0$.

Die Zufallsvariablen $r(\boldsymbol{B}, \theta_0)$ sind unter Gültigkeit der Nullhyptothese unabhängig identisch verteilt, sodass das Niveau des Vorzeichentests eingehalten wird. Es ist allerdings zu beachten, dass aus $\theta = \theta_0$ zwar $Median(r(\boldsymbol{B}, \theta_0)) = 0$ folgt, jedoch nicht umgekehrt. Es kann also Situationen geben, in denen θ von θ_0 abweicht, der Median von $r(\boldsymbol{B}, \theta_0)$ aber trotzdem gleich Null ist. In diesem Falle läge die Güte des Vorzeichentests nur beim tatsächlichen Testniveau, welches kleiner oder gleich dem vorgegebenen Niveau α ist.

4.2. Die drei betrachteten Regressionsmodelle

4.2.1. Vorstellung der Modelle

Das erste Modell, welches wir betrachten wollen, ist das einfache lineare Regressionsmodell, das bekanntlich die Modellgleichung

$$Y_i = \theta_1 + \theta_2 x_i + \epsilon_i, \quad i = 1, \dots, n \text{ und } \theta = (\theta_1, \theta_2) \in \mathbb{R}^2 = \mathbb{R}^k, \tag{4.3}$$

besitzt.

Es ist klar, dass der Parameter θ_1 den Achsenabschnitt und der Parameter θ_2 die Steigung der Geraden bestimmt. Auf eine grafische Veranschaulichung wird deswegen an dieser Stelle verzichtet.

Das zweite Modell, dass welches wir betrachten, ist das sogenannte *exponentielle Wachstumsmodell*. Es ist gegeben durch die Modellgleichung

$$Y_i = \theta_1 \exp(\theta_2 x_i) + \epsilon_i, \quad i = 1, \dots, n, \ x_i \in \mathbb{R} \text{ und } \theta = (\theta_1, \theta_2) \in \mathbb{R}^2 = \mathbb{R}^k.$$
(4.4)

Im Folgenden betrachten wir das Verhalten von $f_2(x,\theta) := \theta_1 \exp(\theta_2 x_i)$ als Funktion von x bei positivem θ_1, θ_2 . Der Abbildung 4.1 liegen äquidistante Werte von $x_1 = 0.02, x_2 = 0.04, \ldots, x_{100} = 2$ zugrunde. Wir sehen in Abbildung 4.1 (a)-(c), dass sowohl eine Erhöhung von θ_1 als auch von θ_2 eine Verstärkung der Krümmung sowie der Steigung von f_2 nach sich zieht. Die Krümmung von f_2 wächst mit größerem x.

Dieses Verhalten ändert sich bei negativen Werten von θ_1, θ_2 , denn der Wechsel ins Negative beeinflusst die Richtung der Monotonie von f_2 und die Vorzeichen der *y*-Werte, siehe als Beispiel Abbildung 4.1 (d).

Obige Eigenschaften folgen unmittelbar aus denen der Exponentialfunktion.

Das dritte und letzte Modell, welches wir untersuchen wollen, ist das *Power-Modell* mit der Modellgleichung

$$Y_i = \theta_1 x^{\theta_2} + \epsilon_i, \quad i = 1, \dots, n, \ x_i \in \mathbb{R}^{>0} \ \text{und} \ \theta = (\theta_1, \theta_2) \in \mathbb{R}^2 = \mathbb{R}^k.$$
(4.5)

Auch hier betrachten wir das Verhalten von $f_3(x,\theta) := \theta_1 x^{\theta_2}$ als Funktion von x bei positivem θ_1 und $\theta_2 > 2$.



Abbildung 4.1.: Veranschaulichung des exponentiellen Wachstumsmodells 4.4 ohne Fehler für verschiedene Kombinationen von (θ_1, θ_2) . Es liegen hier 100 Werte der unabhängigen Variablen x zugrunde, die äquidistant mit einer Schrittweite von 0.02 Werte von 0.02 bis 2 annehmen.

Der Abbildung 4.2 liegen wieder äquidistante Werte von $x_1 = 0.02, x_2 = 0.04, \dots, x_{100} = 2$ zugrunde. Mithilfe der ersten beiden Ableitungen von $f_3(x, \theta)$ nach x,

$$\begin{split} \frac{\partial f_3(x,\theta)}{\partial x} &= \theta_1 \theta_2 x^{\theta_2 - 1} \quad \forall \ x \in \mathbb{R}^{>0}, \\ \frac{\partial f_3(x,\theta)}{\partial^2 x} &= \theta_1 \theta_2 (\theta_2 - 1) x^{\theta_2 - 2} \quad \forall \ x \in \mathbb{R}^{>0}, \end{split}$$

stellen wir fest, dass sowohl eine Erhöhung von $\theta_1 > 0$ als auch von $\theta_2 > 2$ eine Zunahme der Krümmung von f_3 bewirken. Die Krümmung wächst mit größerem x. Für $1 < \theta_2 < 2$ sinkt die Krümmung mit größerem x. Negative Werte von θ_1 sowie Werte

Fur $1 < \theta_2 < 2$ sinkt die Krummung mit großerem x. Negative Werte von θ_1 sowie Werte von $\theta_2 < 1$ ändern das Krümmungsverhalten, denn dann ist f_3 konkav statt konvex, siehe



Abbildung 4.2.: Veranschaulichung des Power-Modells 4.5 ohne Fehler für verschiedene Kombinationen von (θ_1, θ_2) . Es liegen hier 100 Werte der unabhängigen Variablen x zugrunde, die äquidistant mit einer Schrittweite von 0.02 Werte von 0.02 bis 2 annehmen.

als Beispiel Abbildung 4.2 (d).

Insgesamt verharrt die Funktion f_3 für festes θ bei kleinen x-Werten länger in der Nähe von Null als f_2 , steigt danach jedoch rasch an.

4.2.2. Anwendbarkeit der vereinfachten Simplex-Datentiefe

In diesem Abschnitt beschäftigen wir uns mit dem Nachweis, dass die Voraussetzungen von Satz 3.2.2 erfüllt sind, da sonst die Grundlage für die Definition der vereinfachten Simplex-Datentiefe fehlen würde.

Für alle zu untersuchenden Modelle gilt die Parameterdimension k = 2, also (k - 1) = 1und (k + 1) = 3. Die betrachteten k-Tupel der n gegebenen Regressoren x_1, \ldots, x_n haben

die Form $(x_{j_1}, x_{j_2}, x_{j_3})$, $j_i \in \{1, \ldots, n\}$, $j_1 < j_2 < j_3$. Dies führt auf Intervalle $[x_{j_1}, x_{j_3}]$, für die die Bedingungen aus Satz 3.2.2 erfüllt sein müssen.

Wir betrachten zunächst das einfache lineare Regressionsmodell 4.3. Seien die Regressoren x_i i = 1, ..., n gegeben. Es gelten die Bezeichnungen und Voraussetzungen aus der Modellgleichung 3.1. Daraus ergibt sich zunächst

$$f(x,\theta) = \theta_1 + \theta_2 x \Rightarrow \frac{\partial f(x,\theta)}{\partial \theta} = (1,x)^T,$$

und weiter

$$v_u(x) = u_1 + u_2 x.$$

 $v_u(x)$ ist eine Gerade und hat demnach höchstens einen Vorzeichenwechsel im Intervall $[x_{j_1}, x_{j_3}]$. Somit ist Bedingung (1) aus Satz 3.2.2 erfüllt. In Bedingung 2 aus Satz 3.2.2 hat S die Form

$$S = \{(1,1,1), (-1,1,1), (-1,-1,1), (1,-1,-1), (1,1,-1), (-1,-1,-1)\}.$$

Sei $(x_{j_1}, x_{j_2}, x_{j_3})$ beliebig, aber fest.

Im Falle, dass kein Vorzeichenwechsel in S stattfindet, muss die Nullstelle von $v_u(x)$ kleiner bzw. größer als x_{j_1} bzw. x_{j_3} sein. Es lassen sich offensichtlich problemlos solche u_1, u_2 finden.

Im Falle eines Vorzeichenwechsels müssen $u_1, u_2 \in \mathbb{R}$ existieren, sodass die Gerade $v_u(x)$ monoton wachsend bzw. monoton fallend ist und ihre Nullstelle besitzt im offenen Intervall $]x_{j_i}, x_{j_{i+1}}[, i = 1, 2.$ Dies ist offensichtlich der Fall, und wir können, wenn wir die Nullstelle z.B. in der Mitte des Intervalls wählen, d.h. an der Stelle $\frac{x_{j_i}+x_{j_{i+1}}}{2}$, die Menge der Lösungen,

$$u_1 + u_2 \frac{x_{j_i} + x_{j_{i+1}}}{2} = 0 \Leftrightarrow u_1 = -u_2 \frac{x_{j_i} + x_{j_{i+1}}}{2},$$

angeben. Wir wählen dann u_2 gemäß dem für das betrachtete Element aus S erforderlichen Monotonieverhalten.

Wir betrachten nun das exponentielle Wachstumsmodell 4.4. Hier ist $f(x, \theta)$ gegeben als

$$f(x,\theta) = \theta_1 exp(\theta_2 x) \implies \frac{\partial f(x,\theta)}{\partial \theta} = (exp(\theta_2 x), \theta_1 x \cdot exp(\theta_2 x)) = exp(\theta_2 x)(1,\theta_1 x)^T.$$

Da $exp(\theta_2 x) \in \mathbb{R}^{>0}$ für alle $\theta_2, x \in \mathbb{R}$, betrachten wir nur $v_u(x) = u_1 + u_2 \theta_1 x$. Substituieren wir noch $u_2^* := u_2 \theta_1$, so ergibt sich $v_u^*(x)$ als

$$v_u^*(x) = u_1 + u_2^* x.$$

Der Beweis verläuft nun mit der unbedeutenden Einschränkung $\theta_1 \neq 0$ völlig analog zu dem des einfachen linearen Regressionsmodells.

Wir kommen schließlich zum dritten Modell 4.5, dem Power-Modell. Hier sind die notwendigen Größen gegeben als

$$f(x,\theta) = \theta_1 x^{\theta_2} \ \Rightarrow \ \frac{\partial f(x,\theta)}{\partial \theta} = (x^{\theta_2}, \theta_1 x_2^{\theta} \ln(x)) = x^{\theta_2} (1, \theta_1 \ln(x))^T.$$

Da $x^{\theta_2} > 0$ für alle $x \in \mathbb{R}^{>0}, \theta_2 \in \mathbb{R}$, betrachten wir wieder, wenn wir $u_2\theta_1 =: u_2^*$ setzen, nur

$$v_u^*(x) = u_1 + u_2^* \ln(x).$$

Abermals sei $[x_{j_1}, x_{j_2}, x_{j_3}]$ beliebig, aber fest, sowie $\theta_1 \neq 0$. $v_u^*(x)$ ist streng monoton wachsend für $u_2^* > 0$ und streng monoton fallend für $u_2^* < 0$, sowie und das im gesamten Definitionsbereich $\mathbb{R}^{>0}$. Insbesondere besitzt diese Funktion wieder nur höchstens eine Nullstelle. Dasselbe Vorgehen wie bei dem am Anfang untersuchten Modell 4.3 führt uns im Falle eines Vorzeichenwechsels zu

$$u_1 + u_2^* \ln\left(\frac{x_{j_i} + x_{j_{i+1}}}{2}\right) = 0 \Leftrightarrow u_1 = -u_2^* \ln\left(\frac{x_{j_i} + x_{j_{i+1}}}{2}\right), \ i = 1, 2.$$

Die Parameter u_1, u_2^* und damit u_1, u_2 sind somit in jedem Falle so wählbar, dass das gewünschte Monotonieverhalten und die gewünschte Nullstellenposition erreicht und damit die Bedingung 2 aus Satz 3.2.2 erfüllt ist.

4.2.3. Konsistenz der Tests im Rahmen der Modelle

Im Folgenden kürzen wir die Bezeichnungen der Tests $\varphi_n^{D_i}$ ab durch $\varphi_n^{D_i} =: D^{(i)}, i = 1, 2, 3$. In diesem Abschnitt wollen wir die Konsistenz unserer Tests $D^{(1)}, D^{(2)}$ und $D^{(3)}$, die sich aus Satz 3.3.1 ergeben, unter den drei betrachteten Modellen untersuchen, indem wir den Konsistenzsatz 3.4.1 anwenden.

Für alle Modelle gilt k = 2. Wir betrachten also die Nullhypothese

$$H_0: \ \theta = (\theta_1, \theta_2) = (\theta_1^0, \theta_2^0) = \theta^0.$$

Es gelten die Bezeichnungen aus Satz 3.4.1, insbesondere sei $\theta' = (\theta'_1, \theta'_2)$ der wahre Parameter. Die Fehler seien dabei normalverteilt mit Erwartungswert Null, d.h. es gilt $P_{\theta'}(\epsilon_i > c) \neq \frac{1}{2}$ für alle $c \neq 0$. Wir erinnern an Voraussetzung 1 und 2 aus Kapitel 3.4.

Modell 1: Das einfache lineare Regressionsmodell

Wir betrachten nun das Modell 4.3 mit $f_1(x,\theta) := \theta_1 + \theta_2 x$.

Fall 1: $\theta'_1 \neq \theta^0_1, \ \theta'_2 = \theta^0_2$

Die relevante Größe $f_1(x, \theta^0) - f_1(x, \theta')$ ist gegeben durch

$$f_1(x,\theta^0) - f_1(x,\theta') = \theta_1^0 + \theta_2^0 x - \theta_1' - \theta_2' x = \theta_1^0 - \theta_1' \neq 0.$$
(4.6)

In Teil 1 des Satzes 3.4.1 ist dann I = 1, d.h. die einzige Partition ist das Intervall [a, b], und jedes genügend kleine $\delta > 0$ erfüllt die Bedingung. Da 4.6 konstant ist, also monoton, und betragsmäßig größer Null, sind die Voraussetzungen dort erfüllt. Die Tests $D^{(1)}, D^{(2)}$ sind unter Voraussetzung 1 also konsistent.

Auch die Bedingung im 2. Teil des Satzes ist mit 4.6 offensichtlich erfüllt, woraus die Konsistenz obiger Tests auch unter Voraussetzung 2 folgt.

Die Bedingung im 3. Teil des Satzes ist jedoch aufgrund der Konstanz von 4.6 nicht erfüllt.

Fall 2: $\theta'_1 = \theta^0_1, \ \theta'_2 \neq \theta^0_2$

Die Größe $f_1(x, \theta^0) - f_1(x, \theta')$ ist hier gegeben durch

$$f_1(x,\theta^0) - f_1(x,\theta') = \theta_1^0 + \theta_2^0 x - \theta_1' - \theta_2' x = (\theta_2^0 - \theta_2')x.$$
(4.7)

Die Funktion 4.7 ist eine (natürlich monotone) Gerade, die die Nullstelle 0 besitzt. Also ergibt sich mit I = 2 die Zerlegung

$$[a, b] = [a, 0] \cup [0, b], \tag{4.8}$$

falls $0 \in [a, b[$, und die Partition [a, b[selbst andernfalls. Die Forderungen aus Teil 1 von Satz 3.4.1 sind somit erbracht, denn offensichtlich erfüllt für obige zwei Fälle jedes genügend kleine $\delta > 0$ diese. Dass die Voraussetzungen von Teil 3 von Satz 3.4.1 erfüllt werden, ist leicht zu sehen. Teil 2 des Satzes kann (und muss) jedoch keine Anwendung finden. Die Tests $D^{(1)}$ und $D^{(2)}$ sind somit konsistent unter Voraussetzung 1 und 2, und der Test $D^{(i)}$ ist konsistent unter Voraussetzung 2 für alle $i \in \{1, 2, 3\}$.

Fall 3: $\theta'_1 \neq \theta^0_1, \ \theta'_2 \neq \theta^0_2$

Die Differenz der Funktionen $f_1(x, \theta^0) - f_1(x, \theta')$ ist hier gegeben durch

$$f_1(x,\theta^0) - f_1(x,\theta') = \theta_1^0 + \theta_2^0 x - \theta_1' - \theta_2' x = (\theta_1^0 - \theta_1') + (\theta_2^0 - \theta_2') x.$$
(4.9)

Die Nullstelle $y_{\theta_1^0,\theta_1'}$ dieser monotonen Geraden hängt von θ_1^0 und θ_1' ab. Teil 1 von Satz 3.4.1 ist genau wie in Fall 2 erfüllt, wenn $y_{\theta_1^0,\theta_1'} \notin [a, b]$. Andernfalls ist als Partition in 4.8 die Null durch die Nullstelle zu ersetzen. Teil 2 des Satzes ist wie in Fall 2 nicht erfüllt, jedoch Teil 3 schon, wie leicht zu sehen ist. Also sind im vorliegenden Fall unter Voraussetzung 1 und 2 die Tests $D^{(1)}, D^{(2)}$ konsistent, und unter Voraussetzung 2 alle drei betrachteten Tests.

Wir können somit festhalten, dass unter Voraussetzung 1 und 2 die Tests $D^{(1)}$ und $D^{(2)}$ konsistent sind für die eingangs erwähnte Nullhypothese und jedes $\theta' \neq \theta^0$. Der Test $D^{(3)}$ ist unter Voraussetzung 2 konsistent für alle θ' , die die Bedingung $\theta'_2 \neq \theta^0_2$ erfüllen.

Modell 2: Das exponentielle Wachstumsmodell

Wir kommen jetzt zu Modell 4.4 mit $f_2(x,\theta) := \theta_1 \exp(\theta_2 x)$.

Fall 1: $\theta'_1 \neq \theta^0_1, \ \theta'_2 = \theta^0_2$

Die relevante Differenz von $f_2(x, \theta^0)$ und $f_2(x, \theta')$ ergibt sich hier zu

$$f_2(x,\theta^0) - f_2(x,\theta') = \theta_1^0 \exp(\theta_2^0 x) - \theta_1' \exp(\theta_2' x) = (\theta_1^0 - \theta_1') \exp(\theta_2^0 x) = (\theta_1^0 - \theta_1') \exp(\theta_2' x) \neq 0.$$
(4.10)

Diese Funktion hat keine Nullstelle in x und ist offensichtlich monoton. Teil 1 von Satz 3.4.1 ist also wie im Fall 1 des vorherigen Modells erfüllt. Teil 2 ist offensichtlich nicht erfüllt. Teil 3 gilt mit $+\infty$ für $\theta_1^0 > \theta_1'$ und mit $-\infty$ andernfalls.

Somit ist $D^{(i)}$ konsistent unter Voraussetzung 1 für i = 1, 2, und unter Voraussetzung 2 für i = 1, 2, 3.

Fall 2: $\theta'_1 = \theta^0_1, \ \theta'_2 \neq \theta^0_2$

Wir erhalten für diesen Fall die Funktionendifferenz

$$f_{2}(x,\theta^{0}) - f_{2}(x,\theta') = \theta_{1}^{0} \exp(\theta_{2}^{0}x) - \theta_{1}' \exp(\theta_{2}'x)$$

$$= \theta_{1}^{0} \left(\exp(\theta_{2}^{0}x) - \exp(\theta_{2}'x) \right) = \theta_{1}' \left(\exp(\theta_{2}^{0}x) - \exp(\theta_{2}'x) \right).$$
(4.11)

Diese Funktion hat für $\theta'_1 = \theta_1^0 \neq 0$ abermals keine Nullstelle und ist monoton, sodass Teil 1 von Satz 3.4.1 wieder erfüllt ist. Die Monotonie erkennen wir an der ersten Ableitung von $g(x, \theta) := \exp(\theta_2^0 x) - \exp(\theta_2' x)$, die lautet:

$$\frac{\partial g(x,\theta)}{\partial x} = \theta_2^0 \exp(\theta_2^0 x) - \theta_2' \exp(\theta_2' x).$$

 $\text{ Es ist } g(x,\theta)>0 \text{ für } \theta_2^0>\theta_2' \text{ und } g(x,\theta)<0 \text{ für } \theta_2^0<\theta_2'.$

Teil 2 ist für den Fall $\theta'_1 = \theta^0_1 = 0$ nicht erfüllt (und andernfalls erst recht nicht), da die geforderte Konstante *c* nicht gleich Null sein darf. Für $\theta^0_2 > 0$ oder $\theta'_2 > 0$ strebt mindestens eine der Exponentialfunktionen in 4.11 gegen unendlich, also ist für diese Thetas die Bedingung aus Teil 3 erfüllt. Der Test $D^{(i)}$ ist somit konsistent unter Voraussetzung 2 für i = 1, 2, 3, und unter Voraussetzung 1 für i = 1, 2.

Fall 3: $\theta'_1 \neq \theta^0_1, \ \theta'_2 \neq \theta^0_2$

Hier erhalten wir:

$$g(x,\theta) := f_2(x,\theta^0) - f_2(x,\theta') = \theta_1^0 \exp(\theta_2^0 x) - \theta_1' \exp(\theta_2' x) \neq 0.$$
(4.12)

Wir benutzen folgendes Lemma, um die Bedingung aus Teil 1 von Satz 3.4.1 nachzuweisen.

4.2.1 Lemma:

Sowohl die Funktion $g(x,\theta)$ aus 4.12 als auch ihre erste Ableitung $\frac{\partial g(x,\theta)}{\partial x}$ nach x besitzen höchstens eine Nullstelle in [a,b].

Beweis:

Wir beginnen mit

$$\begin{split} &\frac{\partial g(x,\theta)}{\partial x} = \theta_1^0 \theta_2^0 \exp(\theta_2^0 x) - \theta_1' \theta_2' \exp(\theta_2' x) = 0 \\ \Leftrightarrow exp(\theta_2^0 x) = \frac{\theta_1' \theta_2'}{\theta_1^0 \theta_2^0} \exp(\theta_2' x) \\ \Leftrightarrow \theta_2^0 x = \ln\left(\frac{\theta_1' \theta_2'}{\theta_1^0 \theta_2^0}\right) + \theta_2' x \\ \Leftrightarrow x = \frac{\ln\left(\frac{\theta_1' \theta_2'}{\theta_1^0 \theta_2^0}\right)}{\theta_2^0 - \theta_2'}. \end{split}$$

Dies gilt für alle θ mit

$$\theta_1'\theta_2', \theta_1^0\theta_2^0 > 0 \text{ oder } \theta_1'\theta_2', \theta_1^0\theta_2^0 < 0.$$

Andernfalls ist offensichtlich keine Nullstelle vorhanden. Genauso erhalten wir

$$g(x,\theta) = 0 \Leftrightarrow x = \frac{\ln\left(\frac{\theta_1'}{\theta_1^0}\right)}{\theta_2^0 - \theta_2'} \ \forall \ \theta_1^0, \theta_1' > 0 \ \text{oder} \ \theta_1^0, \theta_1' < 0.$$

Andernfalls ist offensichtlich wieder keine Nullstelle vorhanden.

Da $g(x, \theta)$ also nur höchstens eine Nullstelle in [a, b] besitzt und aus einer endlichen Anzahl an monotonen Teilstücken besteht, ist die Bedingung aus Teil 1 von Satz 3.4.1 erfüllt und die Tests $D^{(1)}$ und $D^{(2)}$ sind konsistent. Teil 2 ist nie erfüllt. Aufgrund von $\theta'_2 \neq \theta_2^0$ wächst für $\theta'_2 > 0$ oder $\theta_2^0 > 0$ eine der beiden Exponentialfunktionen in 4.12 schneller über alle Grenzen als die andere. Damit ist Teil 3 des Satzes erfüllt und die drei Tests $D^{(1)}, D^{(2)}$ und $D^{(3)}$ sind unter Voraussetzung 2 konsistent.

Wir fassen zusammen, dass sicher unter Voraussetzung 2 alle drei Tests konsistent sind für obige Nullhypothese, falls $\theta'_2 > 0$ oder $\theta^0_2 > 0$ gilt. Unter Voraussetzung 1 konnten wir die Konsistenz der beiden Tests $D^{(1)}$ und $D^{(2)}$ nachweisen bis auf den Fall $\theta'_1 = \theta^0_1 = 0, \theta'_1 \neq \theta^0_2$, der aber nur ein Modell umfasst, welches konstant gleich Null ist.

Modell 3: Das Power-Modell

Wir kommen nun zum dritten und letzten Modell, dem Power-Modell, welches gegeben ist durch Modellgleichung 4.5. Hier gilt $f_3(x,\theta) = \theta_1 x^{\theta_2}, x \in \mathbb{R}^{>0}$.

 $\textbf{Fall 1: } \theta_1' \neq \theta_1^0, \ \theta_2' = \theta_2^0$

Die Differenz von f_3 unter θ^0 zu f_3 unter θ' ergibt sich als

$$f_3(x,\theta^0) - f_3(x,\theta') = \theta_1^0 x^{\theta_2^0} - \theta_1' x^{\theta_2'} = x^{\theta_2^0} (\theta_1^0 - \theta_1') = x^{\theta_2'} (\theta_1^0 - \theta_1').$$
(4.13)

Aufgrund der Einschränkung, dass x echt positiv sein soll, hat diese Funktion keine Nullstelle. Sie ist außerdem monoton. Genau wie im Fall 1 des einfachen linearen Regressionsmodells ist also die Bedingung in Teil 1 des Satzes 3.4.1 erfüllt. Die Bedingung in Teil 2 ist offensichtlich nie erfüllt. Teil 3 ist erfüllt, falls $\theta'_2 = \theta_2^0 > 0$ gilt. Somit ist der Test $D^{(i)}$ konsistent unter Voraussetzung 1 für i = 1, 2, und unter Voraussetzung 2 für i = 1, 2, 3, falls $\theta'_2 = \theta_2^0 > 0$ erfüllt ist.

Fall 2:
$$\theta'_1 = \theta^0_1, \ \theta'_2 \neq \theta^0_2$$

Die Differenz $f_3(x,\theta_0) - f_3(x,\theta')$ ergibt sich hier als

$$g(x,\theta) := f_3(x,\theta^0) - f_3(x,\theta') = \theta_1^0 x^{\theta_2^0} - \theta_1' x^{\theta_2'} = \theta_1^0 (x^{\theta_2^0} - x^{\theta_2'}) = \theta_1' (x^{\theta_2^0} - x^{\theta_2'})$$
(4.14)

Die Funktion 4.14 besitzt keine Nullstelle. Die Rechnung

$$\begin{aligned} \frac{\partial g(x,\theta)}{\partial x} &= \theta_1'(\theta_2^0 x^{\theta_2^0 - 1} - \theta_2' x^{\theta_2' - 1}) = 0\\ \Leftrightarrow \theta_2^0 x^{\theta_2^0 - 1} &= \theta_2' x^{\theta_2' - 1}\\ \Leftrightarrow x^{\theta_2^0 - \theta_2'} &= \frac{\theta_2'}{\theta_2^0}\\ \Leftrightarrow x &= \left(\frac{\theta_2'}{\theta_2^0}\right)^{\frac{1}{\theta_2^0 - \theta_2'}} \end{aligned}$$

zeigt, dass die 1. Ableitung von $g(x, \theta)$ nach x genau eine Nullstelle in x besitzt, und somit im Intervall [a,b] höchstens eine. Die Gültigkeit von Teil 1 des Satzes 3.4.1 folgt dann wie in Fall 3 des exponentiellen Wachstumsmodells. Teil 2 ist aufgrund der Nichtkonstanz nie erfüllt. Teil 3 ist offensichtlich erfüllt, solange nicht sowohl θ_2^0 als auch θ'_2 kleiner Null sind. Also ist der Test $D^{(i)}$ konsistent unter Voraussetzung 1 für i = 1, 2, und konsistent unter Voraussetzung 2 für i = 1, 2, 3, wobei unter Voraussetzung 2 die Einschränkung an θ_2^0, θ'_2 zu beachten ist.

Fall 3: $\theta'_1 \neq \theta^0_1, \ \theta'_2 \neq \theta^0_2$

Hier ergibt sich

$$g(x,\theta) := f_3(x,\theta^0) - f_3(x,\theta') = \theta_1^0 x^{\theta_2^0} - \theta_1' x^{\theta_2'}.$$
(4.15)

Mit einer ähnlichen Rechnung wie oben können wir zeigen, dass $g(x,\theta)$ genau eine und damit im Intervall [a, b] höchstens eine Nullstelle besitzt. Dasselbe gilt für die 1. Ableitung von $g(x\theta)$ nach x. Die Gültigkeit von Teil 1 des Satzes 3.4.1 folgt somit wieder wie in Fall 3 des exponentiellen Wachstumsmodells. Die Bedingung aus Teil 2 des Satzes ist nie erfüllt. Teil 3 ist wieder erfüllt, falls wir alle $(\theta_2^0, \theta_2') \in \mathbb{R}^{<0} \times \mathbb{R}^{<0}$ ausschließen.

Wir fassen zusammen, dass der Test $D^{(i)}$ unter Voraussetzung 1 ohne Einschränkung konsistent ist für i = 1, 2, und unter Voraussetzung 2 sicher für alle $\theta' \neq \theta^0$ mit $\theta'_2, \theta^0_2 > 0$ für i = 1, 2, 3.

4.3. Simulation: Zielsetzung und Vorgehensweise

Durch eine Simulation soll das Verhalten der Gütefunktionen der drei verschiedenen Tests $D^{(1)}$, $D^{(2)}$ und $D^{(3)}$ aus Satz 3.3.1 im Rahmen der in Kapitel 4.2.1 vorgestellten drei Modelle dargestellt und verglichen werden. Als weiterer Vergleich dient der in Kapitel 4.1 behandelte Vorzeichentest.

Zu diesem Zweck wird eine Menge von unabhängigen Variablen x_i , $i = 1, ..., n_1$ erzeugt, die der Wachstumsbedingung 3.2 genügen. Weiterhin werden ein Rechteck von Parameterwerten $R := [\theta_1^1, \theta_1^2] \times [\theta_2^1, \theta_2^2]$ vorgegeben und zwei Parameterwerte $\theta^0 = (\theta_1^0, \theta_2^0) \in R$ für die Nullhypothese $H_0: \theta = \theta^0$.

Für jedes 2er-Tupel $\theta' \in R$ werden n_1 Realisationen von Y-Werten der Modelle 4.3-4.5 unter θ' erzeugt, und die Nullhypothese wird durch $D_s^{(i)}$ getestet. Dieses Vorgehen wird n_2 -mal wiederholt, und der Anteil der Ablehnungen der jeweiligen Tests wird errechnet. Wir approximieren so die Erwartungswerte, d.h. die Güte der Tests.

Die Werte n_1 und n_2 wurden im Vergleich zu dem Aufsatz von Kustosz et al. (2015) auf $n_1 = 100, n_2 = 200$ erhöht, um die Asymptotik zu verbessern. Die Schrittweite wurde jeweils so gewählt, dass mindestens 90000 Gitterpunkte berechnet werden. Es wurde für jedes Modell *j* für die Fehler einmal eine Normalverteilung mit Erwartungswert 0 und Varianz σ_j verwendet, sodass die Annahmen von Satz 3.3.1 erfüllt sind und auch die Konsistenzuntersuchungen des letzten Abschnitts gelten. Zum Zweiten wurde eine kontaminierte

Normalverteilung, die definiert ist durch $\epsilon \mathcal{N}(0, \sigma_j) + (1 - \epsilon)\mathcal{N}(\mu_j^k, \sigma_j^k)$, verwendet, um die Robustheit der Tests zu überprüfen. Der Parameter ϵ wurde dabei auf 0.05 gesetzt, sodass durchschnittlich jeder 20ste Wert einen additiven Ausreißer darstellt. Das Testniveau wurde einheitlich auf $\alpha = 0.05$ festgesetzt.

Alle Berechnungen wurden in R (R Core Team, 2015) ausgeführt. Der Programmcode ist im Anhang enthalten. Zur Steigerung der Recheneffizienz wurde das R-Paket "parallel", welches alle Prozessorkerne simultan zur Berechnung nutzt, verwendet. Die Gesamtrechenzeit betrug auf einem 4-Kern Intel i5 2.2 Ghz Computer für alle Tests, Modelle und Fehlerverteilungen rund 24 Stunden.

4.4. Ergebnisse für Modell 1: Die einfache lineare Regression

Wir kommen zu den Simulationsergebnissen für das Modell der einfachen linearen Regression 4.3. Zugrunde liegen hier 100 Werte der unabhängigen Variablen x, die die Ausprägungen $0.1, 0.2, 0.3, \ldots, 10$ besitzen. Die Nullhypothese wurde als $\theta_0 = (1.5, 1.5)$ festgesetzt, und das Rechteck R wurde mit $R = [0.5, 2.5] \times [0.5, 2.5]$ so gewählt, dass bei allen vier Tests das Verhalten der Güte in den Plots adäquat abgebildet wird. Die Fehler folgen einer $\mathcal{N}(0, 0.1)$ -Verteilung, d.h. $\sigma_1 = 0.1$. Die Parameter der im vorherigen Kapitel definierten kontaminierten Verteilung wurden als $\mu_1^k = 5, \sigma_1^k = 1$ festgesetzt. Abbildung 4.3 zeigt eine



Abbildung 4.3.: Simulation des einfachen linearen Regressionsmodells unter der Nullhypothese mit unkontaminierten und kontaminierten Fehlern

Simulation des einfachen linearen Regressionsmodells unter der Nullhypothese mit obi-

gen beiden Fehlerverteilungen. Bei der kontaminierten Fehlerverteilung sind die Ausreißer deutlich zu erkennen.

Die Abbildung 4.4 zeigt das simulierte Verhalten der Gütefunktion auf dem Gitter R bei unkontaminierten Fehlern. Die Parameterwerte der Nullhypothese sind durch die gestrichelten Linien markiert. Hellere Farben korrespondieren mit einer häufigeren Ablehnung der Nullhypothese. Die Gütefunktion steigt bei allen Tests $D^{(i)}$, i = 1, 2, 3, an, je weiter das wahre θ von der Nullhypothese entfernt ist.



Abbildung 4.4.: Simulierte Güte für die Tests $D^{(1)}, D^{(2)}, D^{(3)}$ und den Vorzeichentest unter dem einfachen linearen Regressionsmodell mit $\mathcal{N}(0, 0.1)$ Fehlerverteilung

Wie in Kapitel 4.1 schon angedeutet wurde, hat der Vorzeichentest Probleme, falls der Median der Residuen unter dem wahren Parameterwert θ gleich Null ist. Die Parameterwerte, die also trotz der Entfernung zu dem Parameterwert der Nullhypothese zu einem geringen

Anteil von Ablehnungen führen, betreffen Situationen, in denen der Median der Residuen nicht hinreichend weit von Null entfernt ist.

Die simulierten Gütefunktionen der drei Tests $D^{(1)}, D^{(2)}$ und $D^{(3)}$ im Vergleich sind nicht einfach zu bewerten, jedoch scheint der Test $D^{(1)}$ im Vergleich leichte Nachteile zu haben.



Abbildung 4.5.: Simulierte Güte für die Tests $D^{(1)}, D^{(2)}, D^{(3)}$ und den Vorzeichentest unter dem einfachen linearen Regressionsmodell mit kontaminierter $\mathcal{N}(0, 0.1)$ Fehlerverteilung

Weitere Auskunft gibt die Tabelle 4.1, in der einmal die Anzahl der Gitterpunkte, deren simulierte Güte das gesetzte Signifikanzniveau unterschreiten, verzeichnet ist. Die Spalte mit der Beschriftung "UMP" (Uniformly Most Powerful) gibt die Anzahl der Gitterpunkte an, an denen der Test $D_s^{(i)}$ eine höhere geschätzte Güte besitzt als die beiden anderen auf der vereinfachten Simplex-Datentiefe basierenden Tests.

Es ist zu sehen, dass nur sehr wenige Gitterpunkte eine Güte aufweisen, die den Wert 0.05 unterschreitet, wobei dies beim Test $D_s^{(2)}$ etwa fünfmal so häufig passiert wie bei den beiden anderen. Im Gegenzug hat $D_s^{(2)}$ an 13411 Gitterpunkten eine gleichmäßig höhere Güte, während dies bei $D_s^{(1)}$ nur bei 39 Punkten der Fall ist. Die oben erwähnte Einschätzung scheint sich also zu bestätigen.

Abbildung 4.5 zeigt die Gütefunktionen bei kontaminierten Fehlern, die ein ähnliches Erscheinungsbild wie im Falle unkontaminierter Fehler besitzen. Der Vorzeichentest wird durch die Kontamination kaum berührt, hat aber die gleichen Probleme wie oben beschrieben. Der Test $D^{(2)}$ wird durch die Kontamination ebenfalls kaum berührt, während dies bei $D^{(1)}$ schon deutlicher zu sehen ist. Der Test $D^{(3)}$ verliert über einen größeren Gitterbereich leicht an Güte. Anhand der Abbildung zeichnen sich hier also Vorteile für den Test $D^{(2)}$ ab. Test $D^{(1)}$ schneidet auch hier am schlechtesten ab.

Obige Einschätzungen werden durch einen Blick auf die Tabelle 4.1 bestätigt. Es ist klar zu sehen, dass der Test $D^{(1)}$ weiterhin bei einer sehr geringen Anzahl an Gitterpunkten gleichmäßig beste Güte unter allen drei auf der vereinfachten Simplex-Datentiefe basierenden Tests hat, während der Test $D^{(2)}$ hier deutlich vorne liegt.

Tabelle 4.1.: Modell der einfachen linearen Regression: Anzahl der Gitterpunkte aus 90000, an denen die Güte von $D^{(i)}$ das Niveau 0.05 unterschreitet, und Anzahl, bei der die Güte die gleichmäßig beste ist (UMP), für beide Fehlerverteilungen

Test	unkontaminiert		kontaminiert	
	< 0.05	UMP	< 0.05	UMP
$D^{(1)}$	7	39	7	38
$D^{(2)}$	31	13411	28	48046
$D^{(3)}$	6	5932	10	7966

4.5. Ergebnisse für Modell 2: Das exponentielle Wachstumsmodell



Abbildung 4.6.: Simulation des exponentiellen Wachstumsmodells unter der Nullhypothese mit unkontaminierten und kontaminierten Fehlern

Wir besprechen nun die Simulationsergebnisse für das exponentielle Wachstumsmodell 4.4. Zugrunde liegen 100 Werte der unabhängigen Variablen x, welche die Ausprägungen 0.02, 0.04, 0.06, ..., 2 besitzen, damit die Werte von $f(x, \theta)$ nicht zu stark ansteigen. Die Nullhypothese wurde als $\theta_0 = (1.5, 1)$ festgesetzt, und das Rechteck R wurde mit $R = [1.2, 1.8] \times [0.7, 1.3]$ so gewählt, dass bei allen vier Tests das Verhalten der Güte in den Plots wieder adäquat abgebildet wird. Weiterhin sollten sowohl die Werte x_i , $i = 1, \ldots, 100$ als auch $\theta \in R$ keine negativen Werte enthalten, um die grundsätzliche Gestalt von $f(x, \theta)$ nicht zu verändern (siehe Abschnitt 4.2.1). Die Fehler folgen wieder einer $\mathcal{N}(0, 0.1)$ -Verteilung, und die Parameter der kontaminierten Verteilung wurden auch hier als $\mu_2^k = 5, \sigma_2^k = 1$ festgesetzt.

Für die Extremwerte von $R, \theta = (1.2, 0.7)$ und $\theta = [1.8, 1.3]$, sowie für die Parameterwerte der Nullhypothese ist der Verlauf der Funktion $f(x, \theta)$ in Abbildung 4.1 dargestellt.



Abbildung 4.7.: Simulierte Güte für die Tests $D^{(1)}, D^{(2)}, D^{(3)}$ und den Vorzeichentest unter dem exponentiellen Wachstumsmodell mit $\mathcal{N}(0, 0.1)$ Fehlerverteilung

In Abbildung 4.7 ist die durch Simulation geschätzte Gütefunktion auf dem Gitter R bei unkontaminierten Fehlern zu sehen. Mit zunehmender Entfernung des wahren Parameters zur Nullhypothese steigt die Güte bei allen Tests $D^{(i)}$, i = 1, 2, 3, an. Wieder scheint der Test $D^{(1)}$ leichte Nachteile hinsichtlich seiner Gütefunktion gegenüber den anderen beiden Tests zu haben. Der Vorzeichentest hat dieselben Probleme, wie sie in den Simulationsergebnissen des einfachen linearen Regressionsmodells erklärt wurden.

Tabelle 4.2 zeigt, dass nur bei sehr wenigen Gitterpunkten die Güte unter 0.05 absinkt, bei Test $D^{(2)}$ passiert dies relativ betrachtet etwas häufiger. Ansonsten bestätigt sich die obige Einschätzung.



Abbildung 4.8.: Simulierte Güte für die Tests $D^{(1)}, D^{(2)}, D^{(3)}$ und den Vorzeichentest unter dem exponentiellen Wachstumsmodell mit kontaminierter $\mathcal{N}(0, 0.1)$ Fehlerverteilung

Abbildung 4.8 zeigt die geschätzte Gütefunktion für den Fall kontaminierter Fehler. Grundsätzlich sind die Ergebnisse dem Fall unkontaminierter Fehler recht ähnlich. Der Vorzeichentest wird von der Kontamination kaum beeinflusst, zeigt jedoch weiterhin systematische Probleme. Auch die anderen drei Tests erweisen sich in diesem Falle als robust. Der Test $D^{(3)}$ jedoch zeigt über einen größeren Bereich eine leichte Verringerung der Güte. Der Test $D^{(2)}$ schneidet in diesem Vergleich also besser ab.

Die in Tabelle 4.2 abgegebenen Werte zeigen im Falle der Gitterpunkte mit sehr niedriger Güte auch keine Unterschiede zur Simulation mit unkontaminierten Fehlern. Anhand der Spalte "UMP" sehen wir jedoch, dass im vorliegenden Fall der Test $D^{(2)}$ Vorteile hat. Dieser

besitzt an deutlich mehr Gitterpunkten eine gleichmäßig beste Güte im Vergleich zu den anderen beiden Tests.

Tabelle 4.2.: Das exponentielle Wachstumsmodell: Anzahl der Gitterpunkte aus 160000, an denen die Güte von $D_s^{(i)}$ das Niveau 0.05 unterschreitet, und Anzahl, bei der die Güte die gleichmäßig beste ist (UMP), für beide Fehlerverteilungen

Test	unkont	unkontaminiert		kontaminiert	
	< 0.05	UMP	< 0.05	UMP	
$D^{(1)}$	2	10	4	20	
$D^{(2)}$	14	4781	21	63561	
$D^{(3)}$	3	916	3	1432	

4.6. Ergebnisse für Modell 3: Das Power-Modell



(a) $\theta^0 = (\theta_1^0, \theta_2^0) = (2.5, 3.7)$, unkontaminiert (b) $\theta^0 = (\theta_1^0, \theta_2^0) = (2.5, 3.7)$, kontaminiert

Abbildung 4.9.: Simulation des Power-Modells unter der Nullhypothese mit unkontaminierten und kontaminierten Fehlern

Wir kommen nun zu den Simulationsergebnissen für das exponentielle Wachstumsmodell 4.4. Zugrunde liegen hier wieder 100 Werte der unabhängigen Variablen x. Diese besitzen wie im vorherigen Modell die Ausprägungen $0.02, 0.04, 0.06, \ldots, 2$, damit die Werte von $f(x, \theta)$ auch hier nicht zu stark ansteigen. Die Nullhypothese wurde hier mit $\theta_0 = (2.5, 3.7)$ festgesetzt, und die Größe des Rechteckes R wurde mit $R = [1, 4] \times [2.2, 5.2]$ so gewählt, dass
4. Simulationsstudien

bei allen vier Tests das Verhalten der Gütefunktion gut zu erkennen ist. Außerdem wurden die Werte von θ_2 bewusst größer als zwei gewählt, damit hier ebenfalls die grundsätzliche Gestalt von $f(x,\theta)$ erhalten bleibt (siehe 4.2.1). Die Fehler folgen hier einer $\mathcal{N}(0,0.05)$ -Verteilung, d.h. σ_3 ist nur halb so groß wie σ_1 bzw. σ_2 , da ein großer Teil der Funktion $f(x,\theta)$ in diesem Modell nahe Null verläuft. Die Fehler sollen hier die Funktion nicht zu sehr überlagern. Dementsprechend sind die Parameter der kontaminierten Verteilung als $\mu_3^k = 2.5, \sigma_3^k = 0.5$ festgesetzt.



Abbildung 4.10.: Simulierte Güte für die Tests $D^{(1)}, D^{(2)}, D^{(3)}$ und den Vorzeichentest unter dem Power-Modell mit $\mathcal{N}(0, 0.05)$ -Fehlerverteilung

Für die Extremwerte von R, die Parameterwerte $\theta = (1, 2.2)$ und $\theta = [4, 5.2]$, sowie für die Parameterwerte der Nullhypothese ist der Verlauf der Funktion $f(x, \theta)$ in Abbildung 4.2

4. Simulationsstudien

dargestellt.

Abbildung 4.9 enthält wieder eine Simulation des Power-Modells unter der Nullhypothese mit den obigen beiden Fehlerverteilungen. Bei der kontaminierten Fehlerverteilung sind die Ausreißer auch hier deutlich zu erkennen.



Abbildung 4.11.: Simulierte Güte für die Tests $D^{(1)}, D^{(2)}, D^{(3)}$ und den Vorzeichentest unter dem Power-Modell mit kontaminierter $\mathcal{N}(0, 0.05)$ -Fehlerverteilung

Wir betrachten Abbildung 4.10. Es ist zu sehen, dass der Test $D^{(2)}$ im Vergleich am besten abschneidet. Seine Gütefunktion strebt mit zunehmender Entfernung zur Nullhypothese rasch gegen eins. Die Gütefunktion von $D^{(1)}$ hat grundsätzlich dieselbe Form, strebt aber etwas langsamer gegen eins. Deutlich schlechter schneidet der Test $D^{(3)}$ ab, dessen Gütefunktion in vielen Punkten des Gitters einen Wert besitzt, der deutlich kleiner als eins ist und dabei nur sehr langsam ansteigt.

4. Simulationsstudien

Test	${f unkontaminiert}$		kontaminiert	
	< 0.05	UMP	< 0.05	UMP
$D^{(1)}$	0	5	0	4
$D^{(2)}$	0	17090	1	49544
$D^{(3)}$	186	18766	293	16315

Tabelle 4.3.: Das Power-Modell: Anzahl der Gitterpunkte aus 90000, an denen die Güte von $D_s^{(i)}$ das Niveau 0.05 unterschreitet, und Anzahl, bei der die Güte die gleichmäßig beste ist (UMP), für beide Fehlerverteilungen

Tabelle 4.3 bestätigt den obigen Sachverhalt. $D^{(3)}$ unterschreitet auf relativ vielen Gitterpunkten das Niveau von 0.05. Bei den Tests $D^{(1)}$ und $D^{(2)}$ wäre in der entsprechenden Spalte mindestens der der Nullhypothese entsprechende Gitterpunkt zu erwarten. Aufgrund der durch die Simulation bedingten Diskretheit des Wertebereichs des Anteils der Ablehnungen und der Tatsache, dass es sich um asymptotische Tests handelt, kann es jedoch sein, dass an diesem Gitterpunkt der Anteil der Ablehnungen der Nullhypothese etwas größer oder genau gleich 0.05 ist.

Wir betrachten nun die Anzahl der Gitterpunkte mit gleichmäßig bester Güte bezüglich der Tests $D^{(i)}$, i = 1, ..., 3. Der Test $D^{(1)}$ ist fast niemals gleichmäßig besser als die beiden anderen. Dass die Tests $D^{(2)}$ und $D^{(3)}$ an fast gleich vielen Gitterpunkten gleichmäßig beste sind, ist darauf zurückzuführen, dass die Gütefunktionen grundsätzlich unterschiedliche Formen haben. Anhand der Abbildung 4.10 ist aber zu erkennen, dass $D^{(3)}$ an sehr vielen Punkten gleichmäßig schlechter ist als die anderen beiden Tests.

Der Vorzeichentest offenbart wie in den beiden anderen Modellen auch systematische Probleme seiner Gütefunktion.

Abbildung 4.11 zeigt die Ergebnisse des kontaminierten Falls. Die Rangordnung der Tests hinsichtlich ihrer Gütefunktion ändert sich im Vergleich zum unkontaminierten Fall nicht. Der Test $D^{(2)}$ wird indes am wenigsten von der Kontamination beeinflusst und ist deswegen hier erst recht vorzuziehen. Test $D^{(1)}$ verliert über einen größeren Bereich etwas an Güte. Die Güte in den kritischen Bereichen von $D^{(3)}$ sinkt durch die Kontamination weiter ab. Der Vorzeichentest wird weniger beeinflusst.

Tabelle 4.3 bestätigt die durch die Abbildung gewonnenen Ergebnisse. Die Zahl der Gitterpunkte mit Güte kleiner als 0.05 steigt bei $D^{(1)}$ weiter an. Bei den Punkten mit gleichmäßig

$4. \ Simulations studien$

bester Güte ist $D^{(2)}$ nun eindeutig vorne.

5. Zusammenfassung

Thema dieser Diplomarbeit war die Darstellung der im Rahmen eines Aufsatzes (Kustosz et al., 2015) durchgeführten Entwicklung und Analyse von neuen, basierend auf der Verknüpfung von Simplex- und Tangens-Datentiefe vereinfachten Datentiefen. Dies geschah im Rahmen einer Klasse von Regressionsmodellen. Statistische Parametertests für diese Modellklasse wurden entwickelt und ihre Konsistenz analysiert. Anschließend wurden Simulationsstudien zur Güte der Tests durchgeführt.

Kapitel 2 beschrieb das allgemeine Konzept der Datentiefe genauer und führte wichtige, im späteren Verlauf benötigte Begriffe wie Halbraum-, Simplex- und Tangens-Datentiefe ein.

In Abschnitt 3.1 wurde die betrachtete Klasse von Regressionsmodellen vorgestellt, die nur eine unabhängige Variable einschließt, jedoch auch nichtlineare Regressionsfunktionen zulässt. Abschnitt 3.2 verband die Berechnung der verallgemeinerten Simplex-Datentiefe mit der Analyse der Residuen auf alternierende Vorzeichen und legte so die Grundlagen zur Definition dreier Varianten vereinfachter Simplex-Datentiefen, indem nicht alle Residuen, sondern nur verschiedene Teilmengen dieser betrachtet wurden. In Abschnitt 3.3. wurden die asymptotischen Verteilungen der aus den vereinfachten Simplex-Datentiefen hervorgegangenen Tests untersucht. Diese stellten sich als asymptotisch normalverteilt heraus. Abschnitt 3.4 leitete Voraussetzungen für das Regressionsmodell her, unter denen die drei Tests konsistent sind. Hierbei wurden zwei verschiedene asymptotische Szenarien für die unabhängige Variable betrachtet.

Kapitel 4 war den Simulationsstudien dieser Tests im Rahmen dreier Modelle gewidmet, und zwar einmal dem einfachen Regressionsmodell und zweier nichtlinearer Modelle. Der zum Vergleich herangezogene Vorzeichentest wurde in Abschnitt 4.1 dargestellt. Abschnitt 4.2 stellte die Modelle vor und untersuchte, ob die Voraussetzungen zur Anwendbarkeit der vereinfachten Simplex-Datentiefen gegeben waren, was positiv beantwortet werden konnte. Für große Teile des Parameterraumes konnte bei den Modellen die Konsistenz der Tests

5. Zusammenfassung

unter beiden Szenarien bewiesen werden. Die folgenden Abschnitte stellten die eigentliche Simulation und ihre Ergebnisse dar. Hierbei wurden die drei zuvor hergeleiteten Tests und der Vorzeichentest untersucht. Dies geschah jeweils einmal mit normalverteilten Fehlern, die den theoretischen Anforderungen genügten, und einmal mit kontaminierten normalverteilten Fehlern, um die Robustheit der Tests im Rahmen der Modelle zu untersuchen.

In allen Modellen stieg die Gütefunktion mit zunehmender Entfernung zur Nullhypothese an. Der Vorzeichentest offenbarte systematische Konsistenzprobleme.

Im einfachen linearen Regressionsmodell stellte sich eine geringe Unterlegenheit des auf der ersten Version der vereinfachten Simplex-Datentiefe basierenden Tests im Gegensatz zu den beiden anderen heraus. Im kontaminierten Fall war bei allen drei Tests ein leichter Abfall der Güte zu verzeichnen. Der zweite Test erwies sich in diesem Falle gegenüber den beiden anderen leicht überlegen.

Im ersten nichtlinearen Modell zeigte sich ein ähnliches Bild. Auch hier war insgesamt Test 2 zu bevorzugen, da Test 1 hinsichtlich seiner Gütefunktion leicht unterlegen war, und im kontaminierten Fall Test 3 mehr Güte als Test 2 verlor.

Im zweiten nichtlinearen Regressionsmodell schließlich fiel Test 3 deutlich ab. Unter den beiden anderen Tests war Test 2 überlegen. Im kontaminierten Fall war Test 2 noch deutlicher vorzuziehen, da dieser im Gegensatz zu Test 1 kaum an Güte verlor. Test 3 fiel wieder deutlich zurück.

Insgesamt sind die Simulationsergebnisse sehr vielversprechend. Daher wäre es vorstellbar, noch zu weiteren Modellen Simulationen zu berechnen. Außerdem wäre die Anwendung der drei neuen Tests auf verschiedene reale Datensätze mit gleichzeitigem Vergleich klassischer nichtrobuster Methoden wünschenswert.

A. Programmcode

```
library(parallel)
Tiefe.Calc<-function(Res, Alt=0) # Überprüfung auf alternierende Residuen
 {
 if (all(sign(Res)==c(-1, 1, -1)) | all(sign(Res)==c(1, -1, 1))) {Alt<-1}
 return(Alt)
 }
Sign<-function(Res) # Funktion für Vorzeichentest
 {
 Teststatistik<-length(Res[Res>0])
 Ergebnis<-binom.test(Teststatistik, length(Res), p=0.5, alternative="two.sided")</pre>
 return(Ergebnis$p.value)
 }
Kontamination<-function(n.int, error.1, error.2) # Kontaminierte Fehler erzeugen
{
 welche<-rbinom(n.int, 1, 0.05)</pre>
 Ausgabe<-(1-welche)*rnorm(n.int, 0, error.1)+welche*rnorm(n.int, mu, error.2)</pre>
 return(Ausgabe)
}
Hauptfunktion<-function(Theta.1, Theta.2)</pre>
  #
  # Simulation, Berechnung Teststatistiken und Güte
  #
 {
 if (Übergabe$Kontamination==F)
   {Fehler<-replicate(rep, rnorm(n, 0, Übergabe$Error1))} # Fehler simulieren</pre>
  else
   {
   Fehler<-replicate(rep, Kontamination(n, Übergabe$Error1, Übergabe$Error2))</pre>
```

```
}
if (Modell==1) # lineare Regression
 Wert_f<-Theta.1+x*Theta.2 # Werte von f für jeden Gitterpunkt berechnen
 Y_Werte<-apply(Fehler, 2, function(x){Wert_f+x}) # Y-Werte berechnen für rep
      Fehlerfolgen der Größe n
 Res<-apply(Y_Werte, 2, function(x){x-Übergabe$A[[2]]}) # Residuen berechnen für rep Y_
      Wertefolgen der Größe n
 }
if (Modell==2) # exponentielles Wachstumsmodell
 ſ
 Wert_f<-Theta.1*exp(Theta.2*x)
 Y_Werte<-apply(Fehler, 2, function(x){Wert_f+x})</pre>
 Res<-apply(Y_Werte, 2, function(x){x-Übergabe$A[[3]]})</pre>
 }
if (Modell==3) # Power Model
 {
 Wert_f<-Theta.1*(x^Theta.2)</pre>
 Y_Werte<-apply(Fehler, 2, function(x){Wert_f+x})</pre>
 Res<-apply(Y_Werte, 2, function(x){x-Übergabe$A[[4]]})</pre>
 }
d_s1_roh<-sapply(1:rep, function(y){apply(Übergabe$Index[[1]], 2, function(x){Tiefe.Calc
    (Res[x, y])}) # Datentiefen berechnen
d_s1 <-apply(d_s1_roh, 2, function(x){sum(x)*(1/floor(n/3))})
d_s2_roh<-sapply(1:rep, function(y){apply(Übergabe$Index[[2]], 2, function(x){Tiefe.Calc
    (Res[x, y])}))
d_s^2(apply(d_s^2_roh, 2, function(x){sum(x)*(1/(n-2))})
d_s3_roh<-sapply(1:rep, function(y){apply(Übergabe$Index[[3]], 2, function(x){Tiefe.Calc
    (Res[x, y])}))
d_s3<-apply(d_s3_roh, 2, function(x){sum(x)*(1/floor((n-1)/2))})</pre>
Ablehnen_1<-d_s1<Übergabe$Q1[1] # Güte berechnen
Güte_1<-(1/rep)*(length(Ablehnen_1[Ablehnen_1==TRUE]))</pre>
Ablehnen_2<-d_s2<Übergabe$Q1[2]
Güte_2<-(1/rep)*(length(Ablehnen_2[Ablehnen_2==TRUE]))</pre>
Ablehnen_3<-d_s3<Übergabe$Q1[3]
```

A. Programmcode

```
Güte_3<-(1/rep)*(length(Ablehnen_3[Ablehnen_3==TRUE]))
Teststatistik_4<-apply(Res, 2, Sign) # Vorzeichentest anwenden
Ablehnen_4<-Teststatistik_4<0.05
Güte_4<-(1/rep)*(length(Ablehnen_4[Ablehnen_4==TRUE])) # Güte berechnen
return(list(c(Theta.1, Theta.2), c(Güte_1, Güte_2, Güte_3, Güte_4)))
}</pre>
```

```
Kern<-function(Bereich.Theta.1, Bereich.Theta.2, HO.Theta.1, HO.Theta.2, Schrittweite1,
    Schrittweite2, Modell, rep, x, mu=5, Kont, error1=0.1, error2=1, At1, At2)
 # Aufzurufende Funktion
 # Argumente:
 # Bereich...: Gitterintervalle
 # H0...:
              Thetas für Nullhypothese,
 # Schrittweite: Auflösung Gitter
 # Modell: 1-3, s.u.
          Anzahl der Testauswertungen für Güte
 # rep:
 # 7:
          Werte Kovariable
          Erwartungswert der kontaminierenden Verteilung
 # mu:
            True oder False, für Kontamination
 # Kont:
 # error.1: Varianz Normalverteilung,
 # error.2: Varianz kontaminierende Normalverteilung
 # At1, At2: Orte der Tickmarks
 ſ
 n<-length(x)</pre>
 Grid.x<-seq(Bereich.Theta.1[1], Bereich.Theta.1[2], Schrittweite1)</pre>
 Grid.y<-seq(Bereich.Theta.2[1], Bereich.Theta.2[2], Schrittweite2)</pre>
 Theta.Matrix <- as.matrix (expand.grid (Grid.x, Grid.y)) # Theta-Gitter erzeugen
 Index1<-sapply(seq(1, (n'/)3, 3), function(x){x+c(0, 1, 2)} # Indizes für
      Datentiefen erzeugen
 Index2<-sapply(1:(n-2), function(x){x+c(0, 1, 2)})</pre>
 Index3<-sapply(1:floor((n-1)/2), function(x){c(x, floor((n+1)/2), n-x)})
```

Varianz.1<-(3/16)/floor(n/3) # asymptotische Varianzen der Datentiefen berechnen Varianz.2<-(5/16)/(n-2)

```
Varianz.3<-(3/16)/floor((n-1)/2)
Q<-c()
Q[1]<-qnorm(0.05, 1/4, sqrt(Varianz.1)) # Kritische Werte berechnen
Q[2]<-qnorm(0.05, 1/4, sqrt(Varianz.2))
Q[3]<-qnorm(0.05, 1/4, sqrt(Varianz.3))
A1 < -(H0.Theta.1*x)/(x+H0.Theta.2)
A2<-Anpassung<-H0.Theta.1+x*H0.Theta.2 # Anpassung unter H_0 berechnen
A3<-Anpassung<-H0.Theta.1*exp(H0.Theta.2*x)
A4<-Anpassung<-H0.Theta.1*(x^H0.Theta.2)
Übergabe=list(Index=list(Index1, Index2, Index3), Q1=Q, A=list(A1, A2, A3, A4),
    Kontamination=Kont, Error1=error1, Error2=error2)
cl <- makeCluster(4, type = "PSOCK") # Paralleles Rechnen, hier 4 Kerne</pre>
clusterExport(cl, list("Modell", "n", "rep", "x", "mu", "Übergabe"), envir=environment()
    )
clusterExport(cl, list("Hauptfunktion", "Tiefe.Calc", "Sign", "Kontamination"))
RNGkind("L'Ecuyer-CMRG") # Für unabhängige Zufallszahlen
clusterSetRNGStream(cl)
Ergebnis<-parApply(cl=cl, X=Theta.Matrix, MARGIN=1, FUN=Aufruf)# Für jeden Gitterpunkt
    Güte berechnen
Path_1<-paste("C:\\Dokumente\\Modell", as.character(Modell), "Test1", as.character(Kont)
    , ".png", sep="")
Path_2<-paste("C:\\Dokumente\\Modell", as.character(Modell), "Test2", as.character(Kont)
     , ".png", sep="") # Drucken in Datei
Path_3<-paste("C:\\Dokumente\\Modell", as.character(Modell), "Test3", as.character(Kont)
    , ".png", sep="")
Path_4<-paste("C:\\Dokumente\\Modell", as.character(Modell), "Test4", as.character(Kont)
    , ".png", sep="")
png(Path_1, units="in", width=10, height=8, res=200)
Zeichnen (Ergebnis, Bereich. Theta.1, Bereich. Theta.2, 1, Schrittweite1, Schrittweite2, HO
    .Theta.1, HO.Theta.2, At1, At2)
png(Path_2, units="in", width=10, height=8, res=200)
Zeichnen (Ergebnis, Bereich. Theta. 1, Bereich. Theta. 2, 2, Schrittweite 1, Schrittweite 2, HO
    .Theta.1, HO.Theta.2, At1, At2)
png(Path_3, units="in", width=10, height=8, res=200)
Zeichnen (Ergebnis, Bereich. Theta.1, Bereich. Theta.2, 3, Schrittweite1, Schrittweite2, HO
    .Theta.1, HO.Theta.2, At1, At2)
png(Path_4, units="in", width=10, height=8, res=200)
```

```
Zeichnen(Ergebnis, Bereich.Theta.1, Bereich.Theta.2, 4, Schrittweite1, Schrittweite2, H0
      .Theta.1, HO.Theta.2, At1, At2)
 stopCluster(cl)
 Ausgabe<-Extraktion(Ergebnis)
 return(Ausgabe)
 }
Aufruf<-function(Theta) # Aufruf für apply
 ſ
 Hauptfunktion(Theta[1], Theta[2])
 }
Zeichnen<-function(Ergebnis.int, Bereich.Theta.1.int, Bereich.Theta.2.int, Test,
    Schrittweite1.int, Schrittweite2.int, HO.Theta.1.int, HO.Theta.2.int, At1.int, At2.int
    )
 #
 # Gitter zeichnen
 #
 {
 close.screen(all.screens=T)
 split.screen(rbind(c(0, 0.9, 0, 1), c(0.9, 1, 0, 1)))
 Farben<-gray.colors(20, 0.2, 0.9)</pre>
 screen(1)
 par(mai=c(0.8, 0.96, 0.25, 0.2), yaxs="i", xaxs="i", font=5)
 Sys.setlocale("LC_CTYPE", "greek") # Griechischer Zeichensatz, da sonst expression-
      Ausdruck auf dem verwendeten Rechner nicht funktioniert
 plot(c(), xlim=Bereich.Theta.1.int, ylim=Bereich.Theta.2.int, xlab=expression("\u03b8"
      [1]), ylab=expression("\u03b8"[2]), axes=F, cex.lab=2)
 Sys.setlocale("LC_CTYPE", "german") # Wieder Deutsch
 lapply(Ergebnis.int, Zeichnen_unter, Schrittweite1.int=Schrittweite1.int, Schrittweite2.
      int=Schrittweite2.int, Farben=Farben, Test1=Test, Bereich.Theta.1.int, Bereich.Theta
      .2.int, HO.Theta.1.int, HO.Theta.2.int) # Hier werden Graustufen gezeichnet
 axis(1, at=At1.int, cex.axis=2, lwd=0, lwd.ticks=1) # Achsen zeichnen
 axis(2, at=At2.int, cex.axis=2, lwd=0, lwd.ticks=1)
 Legende() # Legende zeichnen
 dev.off()
 }
```

Zeichnen_unter<-function(Ergebnis.int2, Schrittweite1.int2, Schrittweite2.int2, Farben,

```
Test1, Bereich.Theta.1.int2, Bereich.Theta.2.int2, H0.Theta.1.int2, H0.Theta.2.int2) #
     Unterfunktion für Gitterpunkte
 ſ
 Farbe<-Farben[ceiling(Ergebnis.int2[[2]][Test1]*20)] # Grautöne erzeugen
 if (Ergebnis.int2[[2]][Test1]==0) {Farbe<-Farben[1]} # Töne zuordnen
 s1<-Schrittweite1.int2/2
 s2<-Schrittweite2.int2/2
 polygon(c(Ergebnis.int2[[1]][1]-s1, Ergebnis.int2[[1]][1]-s1, Ergebnis.int2[[1]][1]+s1,
     Ergebnis.int2[[1]][1]+s1), c(Ergebnis.int2[[1]][2]+s2, Ergebnis.int2[[1]][2]-s2,
     Ergebnis.int2[[1]][2]-s2, Ergebnis.int2[[1]][2]+s2), col=Farbe, border=Farbe) #
     Gitterpunkte zeichnen
 lines(c(H0.Theta.1.int2, H0.Theta.1.int2), c(Bereich.Theta.2.int2[1], Bereich.Theta.2.
     int2[2]), lty=2)
 lines(c(Bereich.Theta.1.int2[1], Bereich.Theta.1.int2[2]), c(H0.Theta.2.int2, H0.Theta
      .2.int2), lty=2)
 }
Legende <- function() # Legende zeichnen
 ſ
 screen(2)
 par(mai=c(0.8, 0.3, 0.25, 0.4))
 Farben<-gray.colors(20, 0.2, 0.9)
 plot(c(), xlim=c(0, 1), ylim=c(0, 1), axes=F, ann=F, yaxs='i', xaxs='i')
 for (x in 1:20)
   ł
   polygon(c(0, 1, 1, 0), c(x/20, x/20, x/20-1/20, x/20-1/20), col=Farben[x])
 axis(4, at=c(seq(0, 8, 2)/10, 1), cex.axis=2)
 }
Extraktion <- function (Daten) # Tabellendaten extrahieren
 {
 n1<-n2<-n3<-m1<-m2<-m3<-0
 W1<-W2<-W3<-W4<-W5<-W6<-NULL
 for (i in 1:length(Daten))
   {
   W1<-c(W1, all(c(Daten[[i]][[2]][1]>Daten[[i]][[2]][2], Daten[[i]][[2]][1]>Daten[[i
       ]][[2]][3])))
   W2<-c(W2, all(c(Daten[[i]][[2]]2)>Daten[[i]][[2]][1], Daten[[i]][[2]][2]>Daten[[i
```

A. Programmcode

```
]][[2]][3])))
W3<-c(W3, all(c(Daten[[i]][2]][3]>Daten[[i]][2]][1], Daten[[i]][2]][3]>Daten[[i
]][[2]][2])))
W4<-c(W4, Daten[[i]][2]][1]<0.05)
W5<-c(W5, Daten[[i]][2]][2]<0.05)
W6<-c(W6, Daten[[i]][2]][3]<0.05)
}
n1<-length(W1[W1==T])
n2<-length(W2[W2==T])
n3<-length(W3[W3==T])
m1<-length(W4[W4==T])
m2<-length(W4[W4==T])
m2<-length(W5[W5==T])
m3<-length(W6[W6==T])
return(list(AbsBest1=n1, AbsBest2=n2, AbsBest3=n3, Unter1=m1, Unter2=m2, Unter3=m3))
}</pre>
```

Aufrufe für 3 Modelle, jeweils kontaminiert und unkontaminiert

- Ergebnis1<-Kern(Bereich.Theta.1=c(1.2, 1.8), Bereich.Theta.2=c(0.7, 1.3), H0.Theta.1=1.5, H0.Theta.2=1, Schrittweite1=0.002, Schrittweite2=0.002, Modell=2, rep=200, x=(1:100)/ 50, Kont=F, At1=c(1.3, 1.5, 1.7), At2=c(0.8, 1, 1.2))
- Ergebnis2<-Kern(Bereich.Theta.1=c(1.2, 1.8), Bereich.Theta.2=c(0.7, 1.3), H0.Theta.1=1.5, H0.Theta.2=1, Schrittweite=0.002, Schrittweite2=0.002, Modell=2, rep=200, x=(1:100)/ 50, Kont=T, At1=c(1.3, 1.5, 1.7), At2=c(0.8, 1, 1.2))

Ergebnis3<-Kern(Bereich.Theta.1=c(0.5, 2.5), Bereich.Theta.2=c(0.5, 2.5), H0.Theta.1=1.5, H0.Theta.2=1.5, Schrittweite=0.005, Schrittweite2=0.005, Modell=1, rep=200, x=(1:100)/ 10, Kont=F, At1=c(1, 1.5, 2), At2=c(1, 1.5, 2))

Ergebnis4<-Kern(Bereich.Theta.1=c(0.5, 2.5), Bereich.Theta.2=c(0.5, 2.5), H0.Theta.1=1.5, H0.Theta.2=1.5, Schrittweite=0.005, Schrittweite2=0.005, Modell=1, rep=200, x=(1:100)/ 10, Kont=T, At1=c(1, 1.5, 2), At2=c(1, 1.5, 2))

Ergebnis5<-Kern(Bereich.Theta.1=c(1, 4), Bereich.Theta.2=c(2.2, 5.2), H0.Theta.1=2.5, H0. Theta.2=3.7, Schrittweite1=0.01, Schrittweite2=0.01, Modell=3, rep=200, x=(1:100)/50, error1=0.05, Kont=F, At1=c(1.5, 2.5, 3.5), At2=c(2.7, 3.7, 4.7))

```
Ergebnis6<-Kern(Bereich.Theta.1=c(1, 4), Bereich.Theta.2=c(2.2, 5.2), H0.Theta.1=2.5, H0.
Theta.2=3.7, Schrittweite=0.01, Schrittweite2=0.01, Modell=3, rep=200, x=(1:100)/50,
error1=0.05, error2=0.5, mu=2.5, Kont=T, At1=c(1.5, 2.5, 3.5), At2=c(2.7, 3.7, 4.7))
```

Literaturverzeichnis

- Brockwell P. und Davis R. (2006). *Time Series: Theory and Methods*. Springer, New York.2. Auflage.
- Büning H. und Trenkler G. (1994). Nichtparametrische statistische Methoden. de Gruyter, Berlin.
- Dutter R., Filzmoser P., Gather U. und Rousseeuw P. (2003). Developments in Robust Statistics. Physica-Verlag, Heidelberg; New York.
- Kustosz C., Müller C. und Wendler M. (2015). Simplified Simplicial Depth for Regression and Autoregressive Growth Processes. Zur Veröffentlichung eingereicht.
- Liu R. Y. (1990). On a Notion of Simplicial Depth. The Annals of Statistics, 18(1):405–414.
- Mizera I. (2002). On Depth and Deep Points: a Calculus. The Annals of Statistics, 30(6):1681–1736.
- R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rousseeuw P. und Hubert M. (1999). Regression Depth. Journal of the American Statistical Association, 94:388-433.
- Serfling R. (2006). Depth Functions in Nonparametric Multivariate Inference. Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications, 72:1–16.
- Shang Y. (2012). A Central Limit Theorem for Randomly Indexed m-dependent Random Variables. *Filomat*, 26(6):713–717.

Erklärung

Eidesstattliche Versicherung

Name, Vorname	MatrNr.
Ich versichere hiermit an Eides statt, dass ic dem Titel	h die vorliegende Bachelorarbeit/Masterarbeit* mit
selbstständig und ohne unzulässige fremde angegebenen Quellen und Hilfsmittel benutz gemacht. Die Arbeit hat in gleicher oder ähn vorgelegen.	Hilfe erbracht habe. Ich habe keine anderen als die tt sowie wörtliche und sinngemäße Zitate kenntlich licher Form noch keiner Prüfungsbehörde
Ort, Datum	Unterschrift
	*Nichtzutreffendes bitte streichen
Belehrung:	
Manuara italiah sasan sina dia Tiwashwasi	ikan Dröfungalaistungan kataffanda Dagalung sinas

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -)

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird gfls. elektronische Vergleichswerkzeuge (wie z.B. die Software "turnitin") zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Ort, Datum

Unterschrift