Universität Rostock

Traditio et Innovatio

Unabhängigkeits- und Anpassungstest für doppeltrunkierte Daten

Anwendung auf Unternehmenslebensdauern

Dissertation

zur

Erlangung des akademischen Grades

doctor rerum politicarum (Dr. rer. pol.)

der Universität Rostock

vorgelegt von

Anne-Marie Toparkus

geb. am 18.10.1997 in Halle (Saale)

aus Rostock

Rostock, 15. Mai 2025

https://doi.org/10.18453/rosdok_id00005650

Dissertation am Institut für Volkswirtschaftslehre
Wirtschafts- und Sozialwissenschaftliche Fakultät

Gutachter

Prof. Dr. Rafael Weißbach, Universität Rostock, Wirtschafts- und Sozialwissenschaftliche Fakultät, Institut für Volkswirtschaftslehre

Prof. Dr. Natalie Neumeyer, Universität Hamburg, Fakultät für Mathematik, Informatik und Naturwissenschaften, Fachbereich Mathematik

Jahr der Einreichung: 2025

Jahr der Verteidigung: 2025

Lebenslauf

Berufserfahrung:

2021 – 2025 **Wissenschaftliche Mitarbeiterin**, Universität Rostock
Wirtschafts- und Sozialwissenschaftliche Fakultät,
Lehrstuhl für Statistik und Ökonometrie.
Forschung: im Bereich Verweildaueranalyse mit unvollständigen Beobachtungen
Lehre: Durchführung von Lehrveranstaltungen, Betreuung von
Bachelor- und Masterarbeiten

2019 **Praktikum**, Rostocker Volks- und Raiffeisenbank eG
Tätigkeit im Controlling

Studium:

2019 – 2021 **Masterstudium**, Universität Rostock, Studium der
Wirtschaftsmathematik mit Zweitfach Betriebswirtschaftslehre
Abschlussnote: 1,1
Masterarbeit: Grenzwertsätze für fast instabile Hawkes-Prozesse
und ihre Anwendung bei der Modellierung von Hochfrequenzmärkten

2017 – 2021 **Tätigkeit als studentische Hilfskraft**, Universität Rostock

2016 – 2019 **Bachelorstudium**, Universität Rostock, Studium der
Mathematik mit Zweitfach Betriebswirtschaftslehre
Abschlussnote: 1,0
Bachelorarbeit: Differentiation von Lebesgue-Integralen

Schulausbildung:

2010 – 2016 **Gymnasium**, Friedrich-Franz-Gymnasium, Parchim
allgemeine Hochschulreife, Abschlussnote: 1,0

2008 – 2010 **Regionale Schule**, „Fritz Reuter“, Parchim

2004 – 2008 **Grundschule**, „Adolf Diesterweg“, Parchim

Danksagung

An dieser Stelle möchte ich allen danken, die mich während der vergangenen dreieinhalb Jahre unterstützt und so maßgeblich zum Gelingen dieser Arbeit beigetragen haben.

Mein besonderer Dank gilt Rafael Weißbach für seine engagierte Betreuung. Er stand mir jederzeit als Ansprechpartner zur Verfügung und nahm sich immer wieder viel Zeit für fachliche Diskussionen und das gründliche Lesen meiner Arbeit. Sein konstruktives Feedback war von großem Wert und hat wesentlich zum Fortschritt dieser Arbeit beigetragen.

Mein Kollege Eric Scholz hat mich nicht nur zu Konferenzen begleitet und mich mit zahlreichen fachlichen Gesprächen unterstützt, sondern auch meine Arbeit sorgfältig Korrektur gelesen. Dafür danke ich ihm herzlich.

Weiterhin danke ich Natalie Neumeyer, die sich freundlicherweise als Zweitgutachterin zur Verfügung gestellt hat.

Kristina Pieth übernahm zahlreiche organisatorische Aufgaben, zeigte dabei stets ein offenes Ohr für alle Anliegen und war immer sehr hilfsbereit.

Ebenfalls gilt mein Dank meinen Kolleginnen und Kollegen am Institut, die durch ihre harmonische Zusammenarbeit wesentlich zu einer angenehmen Arbeitsatmosphäre beigetragen haben. Insbesondere möchte ich Philipp Budde hervorheben, der mir bei programmiertechnischen Fragestellungen jederzeit beratend zur Seite stand.

Abschließend sei meiner Familie, meinen Freunden und vor allem meinem Partner Paul gedankt, die mir in den vergangenen Jahren beständigen Rückhalt gegeben, mich unterstützt und an mich geglaubt haben.

Abstract

When analyzing left- or double-truncated durations, the likelihood can be derived from standard results for point processes and the unobserved sample size can be profiled out. Common simplifying assumptions for subsequent analyses include independence between the age at truncation and the duration of interest, or specific distributional assumptions, such as exponential durations and uniformly distributed truncation ages. However, both assumptions can be highly unappealing. We propose a parametric test for the first assumption and a nonparametric test for the second.

The independence assumption becomes questionable when truncation arises from sampling within a restricted time period, making the age at truncation equivalent to the date of birth. In such cases, independence conflicts with any non-stationary progress in time of life expectancy. We model dependence with the Gumbel-Barnett copula, where the null hypothesis parameter lies at the boundary of the parameter space. Consequently, the asymptotic distribution of the test statistic is a mixture of a two- and one-dimensional normal distribution.

For the second assumption, we develop a Kolmogorov–Smirnov-type goodness-of-fit test, which needs to be two-dimensional because of the additional attribute age-at-truncation. In a first step, the asymptotic behavior of the truncated process is analyzed when the true parameter under the null hypothesis and the latent sample size are known. Replacing the unknown true parameter and latent sample size by their estimators further alters the distribution of the test statistic. Nevertheless, due to the compactness of truncation region, the test statistic can be computed using methods for the bivariate uniform case.

Monte Carlo simulations assess the empirical power of the independence test and type I error rates of the goodness-of-fit test. An application to 55,000 German enterprise lifetimes ending between 2014 and 2016 illustrates the methodologies.

Zusammenfassung

Bei der Analyse von links- oder doppeltrunkierten Lebensdauern lässt sich auf Basis der Punktprozesstheorie eine Likelihood herleiten, wobei der Einfluss der unbekannten latenten Stichprobengröße durch Ersetzen mit einem geeigneten Schätzer aufgehoben wird. Für weiterführende Analysen werden häufig vereinfachende Annahmen getroffen, etwa die Unabhängigkeit zwischen der Trunkationsvariablen und der interessierenden Überlebenszeit oder spezifische Verteilungsannahmen, beispielsweise exponentialverteilte Lebensdauern und gleichverteilte Trunkationsgrößen. Beide Annahmen erweisen sich jedoch in vielen Anwendungsfällen als wenig überzeugend. Zur Überprüfung dieser Annahmen werden ein parametrischer Test für die Unabhängigkeit sowie ein nichtparametrischer Test zur Beurteilung der Verteilungsannahmen entwickelt.

Die Annahme der Unabhängigkeit erweist sich insbesondere dann als problematisch, wenn die Trunkation aus einer Erhebung innerhalb eines begrenzten Zeitraums resultiert und somit das Alter bei Studienbeginn äquivalent zum Entstehungszeitpunkt ist. In solchen Fällen widerspricht die Unabhängigkeit jeder nicht-stationären Entwicklung der Lebenserwartung im Zeitverlauf. Zur Modellierung der Abhängigkeit wird die Gumbel-Barnett-Copula herangezogen, wobei der Parameter unter der Nullhypothese am Rand des Parameterraums liegt. Die resultierende asymptotische Verteilung der Teststatistik ist eine Mischung aus einer zwei- und einer eindimensionalen Normalverteilung.

Für die Überprüfung der zweiten Annahme wird ein Kolmogorov–Smirnov-ähnlicher Anpassungstest konstruiert, der aufgrund der zusätzlichen Trunkationsvariablen zweidimensional ausgelegt ist. In einem ersten Schritt wird das asymptotische Verhalten des trunkierten Prozesses unter Bekanntsein des wahren Parameterwerts und des latenten Stichprobenumfangs analysiert. Das Ersetzen dieser Größen durch ihre Schätzwerte führt zu weiteren Veränderungen in der Verteilung der Teststatistik. Aufgrund der Kompaktheit des Beobachtungsbereichs kann die Teststatistik mit Methoden für den bivariat gleichverteilten Fall berechnet werden.

Die Teststärke des Unabhängigkeitstests sowie die Typ-I-Fehlerraten des Anpassungstests werden mithilfe von Monte-Carlo-Simulationen untersucht. Eine Anwendung auf 55.000 Lebensdauern deutscher Unternehmen, die zwischen 2014 und 2016 geschlossen wurden, dient zur Veranschaulichung.

Inhaltsverzeichnis

1. Einführung **1**

2. Grundlagen **5**
2.1. Punktprozesse und ihre Dichtefunktionen 5
2.2. Identifizierbarkeit von Parametern 9
2.3. Copulas 11

3. Modell **20**
3.1. Einführung des Modells 20
3.2. Approximation der Likelihood 26
3.3. Konstruktion eines Z-Schätzers 28
3.4. Weitere Copulas 31

4. Modelleigenschaften **35**
4.1. Identifizierbarkeit 35
4.2. Fisher-Information 38
4.3. Konsistenz 44
4.4. Asymptotische Normalität 49

5. Test auf Unabhängigkeit **52**
5.1. Asymptotische Verteilung 52
5.2. Testdurchführung und Untersuchung der Teststärke 55

6. Kolmogorov-Smirnov-Anpassungstest **58**
6.1. Konvergenzsätze für empirische Prozesse 58
6.2. Anpassungstest für unabhängige Doppeltrunkation 67
6.2.1. Verteilung der Teststatistik bei (θ_0, n) 69
6.2.2. Verteilung der Teststatistik bei $(\hat{\theta}_n, n)$ 72
6.2.3. Verteilung der Teststatistik bei $(\theta_0, \hat{n})$ 74
6.2.4. Verteilung der Teststatistik bei $(\hat{\theta}_n, \hat{n})$ 79
6.2.5. Algorithmus zur Berechnung der Prüfgröße 87

7. Anwendung auf Daten zur Unternehmensdemografie **92**
7.1. Beschreibung der Daten 92
7.2. Test auf Zeitstabilität der Unternehmensinsolvenzen 93
7.2.1. Gumbel-Barnett-Copula 93
7.2.2. Farlie-Gumbel-Morgenstern-Copula 95
7.2.3. Zur Spiegelung assozierte Gumbel-Barnett-Copula 96

7.3. Test auf Altershomogenität der Unternehmensinsolvenzen 98

8. Simulationen **105**
8.1. Parameterschätzung und Test auf Unabhängigkeit 105
8.2. Kolmogorov-Smirnov-Anpassungstest 111

9. Fazit **117**

Literaturverzeichnis **119**

A. Appendix **124**
A.1. Score-Funktion und Schätzer in M^* 124
A.2. Kovarianzen der Grenzprozesse im Kolmogorov-Smirnov-Test 126
A.3. Standardfehler im Farlie-Gumbel-Morgenstern-Modell 130
A.4. Verteilungen der Beobachtungen bei unabhängiger Doppeltrunkation . 131

1. Einführung

In der Verweildaueranalyse stellt die Rechtszensierung ein sehr bekanntes und weit verbreitetes Phänomen dar. Dabei ist der Zeitpunkt des Ereigniseintritts für bestimmte Individuen nicht exakt beobachtbar, da dieser nach dem Beobachtungszeitraum liegt. Einer der am häufigsten verwendeten Ansätze in diesem Bereich geht auf die Arbeit von Kaplan und Meier [28] zurück. Während bei der Zensierung zumindest teilweise Informationen über das Individuum bekannt sind, fehlen beim Phänomen der Trunkierung jegliche Informationen, wenn der Ereigniszeitpunkt außerhalb eines definierten Beobachtungsintervalls liegt. Je nachdem, ob das Beobachtungsintervall nach oben oder unten unbeschränkt ist, spricht man von Links- oder Rechtstrunkierung (vgl. Lawless [34]). Linkstrunkierung und Rechtszensierung treten in medizinischen oder sozialwissenschaftlichen Anwendungen häufig gemeinsam auf. Entsprechend umfangreich ist die Literatur zu ihrer gemeinsamen Modellierung (Klein und Moeschberger [30], Turnbull [60], Gross und Lai [22]). Tritt zusätzlich zur Linkstrunkierung auch eine Rechtstrunkierung auf, so wird dies als Doppeltrunkierung bezeichnet. Solche Fälle sind beispielsweise in epidemiologischen Studien oder in ökonomischen Anwendungen, wie in der vorliegenden Arbeit, gegeben. Auch die statistische Behandlung von Doppeltrunkation wurde in der Literatur bereits untersucht, unter anderem von Efron und Petrosian [12], Shen [52], Moreira und Uña-Álvarez [42].

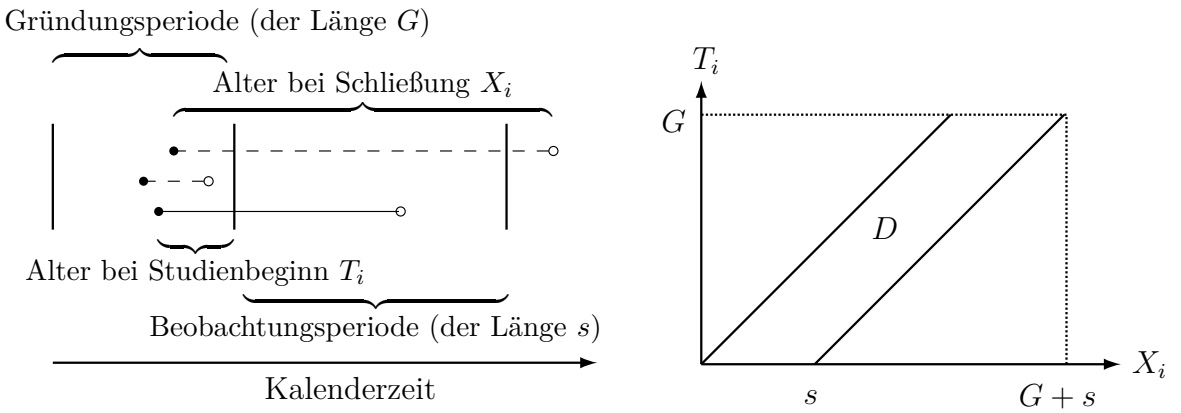

Abbildung 1.1.: Links: Drei Gründungen (schwarzer Punkt) und Schließungen (weißer Kreis): beobachtete (durchgezogene Linie) und trunkierte (gestrichelte Linie) Lebensdauern/ Rechts: Beobachtungsfenster in Abhängigkeit von X_i und T_i als Parallelogramm D (Abbildung aus Weißbach und Wied [65])

Im Gegensatz zu den zuvor genannten Arbeiten verfolgen Weißbach und Wied [65] einen parametrischen Ansatz. In ihrer Arbeit modellieren sie die Lebensdauer X_i von Unternehmen mit einer Exponentialverteilung, während die Gründungszeitpunkte, umkodiert als Alter bei Studienbeginn T_i, als gleichverteilt angenommen werden. Das Alter bei Studienbeginn dient als Trunkationsvariable (siehe Abbildung 1.1). Ein wesentlicher Vorteil dieses Ansatzes besteht darin, dass der Parameter der Exponentialverteilung direkt interpretiert werden kann, da er Informationen über die durchschnittliche Lebensdauer der betrachteten Unternehmen enthält. In dieser sowie in vielen weiteren Arbeiten wird vereinfachend angenommen, dass die interessierende Lebensdauer und der Trunkations- bzw. Gründungszeitpunkt stochastisch unabhängig sind. In der Verweildaueranalyse, insbesondere im Kontext bevölkerungs- oder unternehmensdemografischer Anwendungen, erscheint diese Annahme jedoch nicht realistisch.

In der Literatur wurden Testverfahren entwickelt, die eine abgeschwächte Version der Unabhängigkeit, genauer Unabhängigkeit unter der Bedingung der Beobachtbarkeit, untersuchen (vgl. Martin und Betensky [38], Tsai [59]). Wenn die Annahme dieser Quasi-Unabhängigkeit verletzt ist, liefern die um Trunkation erweiterten Kaplan–Meier- und Turnbull-Schätzer verzerrte Ergebnisse (vgl. Keiding und Moeschberger Abschnitt 5.3 in [29]). Es existieren bereits verschiedene nichtparametrische und semiparametrische Ansätze zur Modellierung von abhängiger Trunkation, beispielsweise von Akritas und LaValley [1], Efron und Petrosian [11] sowie Chiou et al. [7]. Die beiden letztgenannten Arbeiten verwenden einen Transformationsansatz, bei dem eine monotone Transformation der Zielvariablen als Funktion von Kovariaten dargestellt wird. Ist die Verteilung der Ereigniszeit von Interesse, so bieten Copulas einen praktikablen Ansatz zur Modellierung der Abhängigkeit (vgl. Moreira et al. [43], Emura und Wang [16] sowie Chaieb und Rivest [6]). Die Verwendung von Copula-Modellen ermöglicht eine explizite Beschreibung der Abhängigkeitsstrukturen, ohne bereits gewählte Randverteilungen verändern zu müssen. Dadurch bieten sie eine hohe Flexibilität bei der Wahl der Verteilungen und sind somit auch für parametrische Ansätze von besonderem Interesse. Beispielsweise nutzen Emura und Pan [15] den Copula-Ansatz, um Likelihood-Inferenz bei abhängiger Linkstrunkation betreiben zu können. In der Arbeit von Ma und Zhang [37] liegt der Schwerpunkt hingegen weniger auf den asymptotischen Eigenschaften der Schätzer, sondern stärker auf der praktischen Anwendung im Kontext von Ozeandaten.

Die vorliegende Arbeit erweitert den parametrischen Ansatz zur Doppeltrunkation von Weißbach und Wied [65], indem sie mittels einer Copula die Abhängigkeit zwischen Lebensdauer und Trunkationszeit berücksichtigt. Zur Modellierung werden die von Gumbel [23] entwickelten Verteilungen in Form von Copulas herangezogen, die speziell für bivariat exponentialverteilte Zufallsvariablen konzipiert wurden. Die Herleitung der Likelihood erfolgt wie bei Weißbach und Wied [65] auf Grundlage der Punktprozesstheorie nach Reiss [48]. Für die Gumbel-Barnett-Copula ergibt sich bei Unabhängigkeit ein interessanter Spezialfall, bei dem der Parameter auf dem Rand des zulässigen Parameterraums liegt. Asymptotische Eigenschaften wie Konsistenz und Normalität werden insbesondere auch für diesen Grenzfall untersucht, was die Grundlage für die Entwicklung eines Tests auf Unabhängigkeit bildet. Während in den zuvor zitierten Arbeiten die

Trunkation häufig mit bedingten Verteilungen modelliert wird, ermöglicht der Punktprozessansatz hingegen eine Untersuchung der Unabhängigkeit, die sich nicht nur auf die beobachteten, sondern auch auf die latenten Variablen erstreckt. Der in Weißbach und Wied [65] verwendete Datensatz zur Unternehmensdemographie kann im Anschluss mithilfe des entwickelten Tests in Hinblick auf die Zeitstabilität der Unternehmensinsolvenzen analysiert werden. [1]

Neben der Annahme der stochastischen Unabhängigkeit von Lebensdauer und Gründungszeitpunkt treffen Weißbach und Wied [65] auch vereinfachende Verteilungsannahmen, indem sie die Exponential- bzw. Gleichverteilung zugrunde legen.
Konkrete Annahmen über Verteilungen können durch Anpassungstests überprüft werden. Zu den bekanntesten und am häufigsten verwendeten Anpassungstests gehören der Kolomogorov-Smirnov- und der Cramér-von-Mises-Test. Der Kolmogorov-Smirnov-Test geht auf die Arbeiten von Kolmogorov [33] und Smirnov [55] zurück, die sich mit der Teststatistik und ihrer asymptotischen Verteilung befassten. Eine tabellarische Darstellung der Grenzverteilung wurde von Smirnov [56] veröffentlicht. Der Cramér-von-Mises-Test basiert auf den grundlegenden Beiträgen von Cramér [5] und von Mises [63]. Obwohl dem Cramer-von-Mises-Test eine höhere Teststärke zugeschrieben wird (vgl. Genest et al. [19]), findet der Kolmogorov-Smirnov-Test aufgrund seiner einfacheren Berechenbarkeit in der Praxis häufiger Anwendung.
Durch die zusätzliche Trunkationsvariable ist in dieser Arbeit ein zweidimensionaler Anpassungstest erforderlich. Anders als im eindimensionalen Fall hängt die Grenzverteilung der Teststatistik im multivariaten Fall von der zugrunde liegenden Verteilung, zumindest von der enthaltenen Copula, ab. Für den zweidimensionalen Fall zeigt Namaan [44], dass sich die Grenzverteilung bei Unabhängigkeit der einzelnen Zufallsvariablen mit Hilfe der Kolmogorov-Verteilung berechnen lässt. Auch die Berechnung der Teststatistik selbst erweist sich im multivariaten Fall als sehr rechenaufwändig. Justel, Peña und Zamar [27] schlagen eine Prozedur zur Berechnung der Kolmogorov-Smirnov-Statistik im zweidimensionalen Fall vor. Bei Abhängigkeit der Variablen, also bei Vorliegen einer Copula, verwenden Kojadinovic und Yan [32] die sogenannte Multiplier-Methode. Dabei werden zufällige Gewichte, sogenannte Multiplikatoren, generiert, um eine gewichtete Version des empirischen Prozesses zu simulieren. Diese Methode machen sich auch Emura und Konno [14] zunutze, um einen Anpassungstest für abhängige Linkstrunkation durchzuführen. In ihrer Arbeit betonen sie bereits, dass herkömmliche Bootstrap-Verfahren zu rechenintensiv und damit wenig effizient sind. Die Multiplier-Methode basiert dennoch auf der gegebenen Datenstruktur und kann stichprobenabhängig schwanken.

Durch die explizite Bestimmung der Grenzverteilung und ihrer Verwendung zur Simulation wird in der vorliegenden Arbeit sicher gestellt, dass lediglich eine Abhängigkeit von der Verteilung unter der Nullhypothese besteht. Die Schätzung des unbekannten Parameters der Verteilung sowie des latenten Stichprobenumfangs werden dabei stufenweise berücksichtigt. Darüber hinaus ermöglicht die Anwendung der in [27] präsentierten Pro-

[1] Der erste Teil dieser Arbeit zum Unabhängigkeitstest wurde bereits unter [58] veröffentlicht.

zedur eine genauere Berechnung der Teststatistik. Mit dem entwickelten Kolmogorov-Smirnov-ähnlichen Anpassungstest kann die mit der Exponentialverteilung zugrunde gelegte Altershomogenität im Datensatz zur Unternehmensdemographie aus Weißbach und Wied [65] hinterfragt werden.

Auch in anderen Ländern stellen Insolvenzwahrscheinlichkeiten ein aktuelles Forschungsthema dar, und zahlreiche Studien beschäftigen sich mit der Lebensdauer von Unternehmen. Die Studie von Xuezhou et al. [66] mit betriebswirtschaftlicher Fragestellung untersucht den Zusammenhang zwischen Unternehmenswachstum und dem Insolvenzrisiko anhand von Unternehmen aus Pakistan. In der Arbeit von Miao et al. [40] wird eine Methode zur Vorhersage von Insolvenzen chinesischer Unternehmen präsentiert, wobei unter anderem die Rechtszensierung und die hohe Anzahl verfügbarer Kovariablen als Herausforderungen berücksichtigt werden. Der in de Uña-Álvarez et al. [9] entwickelte Ansatz zur Schätzung von Überlebenswahrscheinlichkeiten ist speziell für doppeltrunkierte Daten geeignet und findet Anwendung auf Daten zu spanischen Unternehmen. Konkrete Zahlen zur durchschnittlichen Lebensdauer der Unternehmen werden in den Arbeiten nicht genannt; stattdessen liegt der Fokus auf der Untersuchung von Kovariablen.

Die vorliegende Arbeit gliedert sich folgt: In Kapitel 2 werden zunächst die zentralen mathematischen Konzepte erläutert, die für das Verständnis der weiteren Ausführungen erforderlich sind. Der Schwerpunkt wird dabei auf Punktprozesse, die Identifizierbarkeit von Parametern und Copulas gelegt. Im nächsten Kapitel 3 wird ein Modell zur Beschreibung der abhängigen Doppeltrunkation aufgestellt. Die Likelihood wird approximiert, darauf aufbauend wird ein Schätzer definiert und weitere, für die Anwendung relevante, Copulas werden angeführt. Im Anschluss befasst sich Kapitel 4 mit den theoretischen Eigenschaften wie Identifizierbarkeit, Invertierbarkeit der Fisher-Information, Konsistenz und asymptotischer Normalität. Dem Test auf Unabhängigkeit in Kapitel 5 geht eine Untersuchung der asymptotischen Verteilung des Schätzers auf dem Rand des Parameterraumes voraus. Für den Anpassungstest in Kapitel 6 werden zunächst weitere grundlegende Definitionen und Sätze zur Konvergenz empirischer Prozesse eingeführt. Anschließend wird das asymptotische Verhalten des trunkierten Prozesses analysiert, wobei die Schätzungen des Parameters und der latenten Stichprobengröße schrittweise berücksichtigt werden. Am Ende des Kapitels wird ein Verfahren zur Berechnung der Prüfgröße erläutert. Kapitel 7 widmet sich der Anwendung beider Tests auf Daten zur Unternehmensdemographie, während Kapitel 8 mithilfe von Simulationen die Teststärke des Unabhängigkeitstests sowie die Typ-I-Fehlerraten des Anpassungstests untersucht.

2. Grundlagen

Im ersten Kapitel werden zunächst die für diese Arbeit relevanten mathematischen Grundlagen dargelegt. Die Herleitung der Likelihood in Abschnitt 3.2 basiert auf der Theorie von Punktprozessen. Der erste Abschnitt dieses Grundlagenkapitels bietet eine Einführung in die Thematik der Punktprozesse und die Berechnung ihrer Dichtefunktionen. Der zweite Abschnitt widmet sich der Identifizierbarkeit von Parametern, welche eine wichtige Rolle für die Konsistenz des hergeleiteten Schätzers in Abschnitt 4.3 spielen wird. Für die Abhängigkeitsmodellierung werden in dieser Arbeit Copulas verwendet. Das Konzept der Copulas im Allgemeinen sowie die speziell in dieser Arbeit verwendeten Copulas werden im dritten Abschnitt vorgestellt.
Alle drei Themenbereiche werden für den Test auf Unabhängigkeit in Kapitel 5 und für den Anpassungstest in Kapitel 6 benötigt. Die spezifisch für die jeweiligen Tests benötigten Definitionen und Sätze befinden sich zu Beginn der jeweiligen Kapitel.

2.1. Punktprozesse und ihre Dichtefunktionen

Die Likelihood des Modells wird aus der Dichte eines Poisson-Prozesses bezüglich eines weiteren Poisson-Prozesses gewonnen. Das Modell wird dafür zunächst durch einen trunkierten Punktprozess beschrieben, welcher anschließend durch einen Poisson-Prozess approximiert wird. Die Theorie der Punktprozesse basiert auf den Abschnitten 1.1 bis 1.3 des Buches [48] von Reiss. Dort sind auch die Beweise zu den hier aufgeführten Behauptungen nachzulesen.

Es sei $(S, \mathcal{B})$ ein messbarer Raum. Für eine Menge $B \in \mathcal{B}$ gibt es verschiedene Möglichkeiten die Anzahl der Punkte in B zu beschreiben:

$$|\{x_i \in S : \, i = 1, \ldots, n\} \cap B| = \sum_{i=1}^{n} \mathbb{1}_B(x_i) = \sum_{i=1}^{n} \epsilon_{x_i}(B).$$

Dabei bezeichnet $\epsilon_{x_i}(B) := \mathbb{1}_B(x_i)$ das Diracmaß im Punkt x_i. Nun stellt $\mu = \sum_{i=1}^{n} \epsilon_{x_i}$ ein Maß auf $\mathcal{B}$ dar, welches auch *Punktmaß* genannt wird. Der Raum aller Punktmaße auf der σ-Algebra $\mathcal{B}$ sei mit $\mathbb{M} = \mathbb{M}(S, \mathcal{B})$ notiert. Die σ-Algebra $\mathcal{M} = \mathcal{M}(S, \mathcal{B})$ auf $\mathbb{M}$ ist die kleinste σ-Algebra, bezüglich der die Projektionen $\pi_B : \mathbb{M} \to \overline{\mathbb{N}}_0$ mit $\pi_B(\mu) = \mu(B)$ messbar sind, d.h.

$$\mathcal{M} := \sigma(\{\pi_B^{-1}(C) : \, B \in \mathcal{B}, \, C \subset \overline{\mathbb{N}}_0\}).$$

Es sei weiter $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Abbildung

$$N : \Omega \to \mathbb{M}(S, \mathcal{B})$$

heißt *Punktprozess* auf $(S, \mathcal{B})$, falls diese bezüglich $\mathcal{F}$ und $\mathcal{M}(S, \mathcal{B})$ messbar ist. Somit ist N ein zufälliges Punktmaß auf $(S, \mathcal{B})$. Der folgende Satz gibt ein Kriterium an, welches für den Nachweis der Messbarkeit solcher Abbildungen hilfreich ist.

Satz 2.1.
Die folgenden zwei Behauptungen sind äquivalent:

(*i*) $N : \Omega \to \mathbb{M}$ *ist ein Punktprozess.*

(*ii*) $N(B) : \Omega \to \overline{\mathbb{N}}_0$ *ist* $\mathcal{F} - \mathcal{P}(\overline{\mathbb{N}}_0)$*-messbar für alle* $B \in \mathcal{B}$*, wobei* $\mathcal{P}(\overline{\mathbb{N}}_0)$ *die Potenzmenge von* $\overline{\mathbb{N}}_0$ *bezeichnet.*

Beispiel 2.2.
Es seien $X_i : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{B})$, $i = 1, \ldots, n$ unabhängig identisch verteilte Zufallsvariablen.

(*i*) Diese Zufallsgrößen können auch durch den Punktprozess

$$N_n := \sum_{i=1}^{n} \epsilon_{X_i}$$

repräsentiert werden, wobei jedoch die Information über die Reihenfolge des Auftretens verloren geht. Die Messbarkeit dieses *empirischen Punktprozesses* erhält man mit Satz 2.1 aus der Messbarkeit von $N_n(B) = \sum_{i=1}^{n} \mathbb{1}_B(X_i)$.

(*ii*) Für die Beschreibung der trunkierten Stichprobe in Abschnitt 3.1 ist der *trunkierte Prozess* von zentraler Bedeutung. Der empirische Punktprozess N_n wird auf Ereignisse innerhalb einer Menge $D \in \mathcal{B}$ gestutzt, d.h.

$$N_{n,D}(\cdot) := N_n(\cdot \cap D) = \sum_{i=1}^{n} \epsilon_{X_i}(\cdot \cap D).$$

Wegen $N_{n,D}(B) = N_n(\cdot \cap D)(B) = N_n(B \cap D)$ sowie $B \cap D \in \mathcal{B}$ resultiert die Messbarkeit aus der von N_n.

(*iii*) Ein Punktprozess wird als *gemischt empririscher Prozess* bezeichnet, falls dieser verteilungsgleich zu dem Prozess

$$N := \sum_{i=1}^{\zeta} \epsilon_{X_i}$$

ist, wobei $X_1, \ldots, X_n$ unabhängig identisch verteilte Zufallsgrößen in S sind und

ζ wiederum unabhängig von diesen ist. Aus

$$\{N(B) = k\} = \sum_{j=0}^{\infty} \{N(B) = k,\, \zeta = j\} = \sum_{j=0}^{\infty} \left\{ \sum_{i=1}^{j} \mathbb{1}_B(X_i) = k,\, \zeta = j \right\}$$

folgt die Messbarkeit von $N(B)$, sodass N nach Satz 2.1 ein Punktprozess ist.

An dieser Stelle sei bemerkt, dass die Anwendung des Satzes 2.1 weder die Unabhängigkeit noch die identische Verteilung der zugrunde liegenden Zufallsvariablen erfordert. □

Eine wichtige Rolle bei der Charakterisierung von Punktprozessen spielen die Intensitätsmaße, die wie folgt definiert sind.

Definition 2.3. (Intensitätsmaß)
Das *Intensitätsmaß* $\nu|\mathcal{B}$ eines Punktprozesses N ist definiert durch

$$\nu(B) := \mathbb{E}[N(B)], \quad B \in \mathcal{B}.$$

Somit gibt $\nu(B)$ die erwartete Anzahl von Punkten in der Menge B an. □

Beispiel 2.4.
Die Intensitätsmaße für die drei Punktprozesse aus Beispiel 2.2 lauten

(*i*) $\nu = n \cdot \mathbb{P}(X_1 \in \cdot)$,

(*ii*) $\nu = n \cdot \mathbb{P}(X_1 \in \cdot \cap D)$ und

(*iii*) $\nu = \mathbb{E}[\zeta] \cdot \mathbb{P}(X_1 \in \cdot)$.

□

Ein weiterer Punktprozess von zentraler Bedeutung ist der Poisson-Prozess. Dieser wird später als Annäherung an den trunkierten Prozess dienen. Dafür bezeichne P_s die Wahrscheinlichkeitsfunktion der Poisson-Verteilung zum Parameter $s \geq 0$, d.h.

$$P_s(k) := \frac{s^k}{k!} e^{-s}, \quad k \in \mathbb{N}_0.$$

Definition 2.5. (Poisson-Prozess)
Es sei ν ein endliches Maß auf $\mathcal{B}$. Ein Punktprozess $N : \Omega \to \mathbb{M}(S, \mathcal{B})$ ist ein *Poisson-Prozess*, falls er die folgenden zwei Bedingungen erfüllt:

(*i*) für alle $B \in \mathcal{B}$ ist $N(B)$ poissonverteilt mit $\nu(B)$, d.h. $\mathcal{L}(N(B)) = P_{\nu(B)}$ und

(*ii*) für alle $k \in \mathbb{N}$ und paarweise disjunkte Mengen $B_1, \ldots, B_k \in \mathcal{B}$ sind die Zufallsvariablen $N(B_1), \ldots, N(B_k)$ stochastisch unabhängig.

□

Die Bedingung (i) impliziert $\mathbb{E}[N(B)] = \nu(B)$ für alle $B \in \mathcal{B}$, sodass ν das Intensitätsmaß von N darstellt.
Aus dem nächsten Satz geht hervor, dass Poisson-Prozesse verteilungsgleich zu speziellen gemischten empirischen Punktprozessen sind. Weiterhin besteht eine 1:1-Beziehung zwischen der Verteilung eines Poisson-Prozesses und seines Intensitätsmaßes. Eine explizite Darstellung eines Poisson-Prozesses mit endlichem Intensitätsmaß ν ist gegeben durch

$$N = \sum_{i=1}^{\tau} \epsilon_{X_i}. \tag{2.1}$$

Hierbei sind τ, X_1, X_2, $X_3, \ldots$ stochastisch unabhängig mit $\mathcal{L}(\tau) = P_{\nu(S)}$ und $\mathcal{L}(X_i) = \nu/\nu(S)$, $i \in \mathbb{N}$.

Satz 2.6. (Existenz und Eindeutigkeit von Poisson-Prozessen)
Es sei $\nu|\mathcal{B}$ ein endliches Maß mit $\nu(S) > 0$. Dann gelten die folgenden beiden Aussagen.

- (i) *Der in Gleichung* (2.1) *definierte Punktprozess N ist ein Poisson-Prozess mit Intensitätsmaß ν.*
- (ii) *Poisson-Prozesse mit gleichem Intensitätsmaß besitzen die gleiche Verteilung.*

Für die Approximation des trunkierten Prozesses wird ein Abstandsmaß benötigt. Dafür wird der Hellinger-Abstand wie folgt eingeführt.

Definition 2.7. (Hellinger-Abstand)
Der *Hellinger-Abstand* H zwischen zwei Wahrscheinlichkeitsmaßen Q_1 und Q_2 auf einer σ-Algebra $\mathcal{B}$ mit den ξ-Dichten h_1 und h_2 ist definiert durch

$$H(Q_1, Q_2) := \left(\int (h_1^{1/2} - h_2^{1/2})^2 \, \mathrm{d}\xi \right)^{1/2}.$$

□

Die Genauigkeit der Näherung des trunkierten Prozesses $N_{n,D}$ durch einen Poisson-Prozess N_n^* im Sinne des Hellinger-Abstands hängt von der Wahrscheinlichkeitsmasse ab, die auf der Menge D liegt.

Satz 2.8. (Approximation trunkierter Prozesse)
Es sei $N_{n,D}$ ein trunkierter Prozess wie in Beispiel 2.2(ii), wobei $X_1, \ldots, X_n$ unabhängig identisch verteilte Zufallsvariablen mit der Verteilung Q auf $\mathcal{B}$ sind. Weiter sei N_n^* ein Poisson-Prozess mit dem gleichen Intensitätsmaß ν_n wie $N_{n,D}$. Dann gilt

$$H(\mathcal{L}(N_{n,D}), \mathcal{L}(N_n^*)) \leq 3^{1/2} \cdot Q(D).$$

Der Beweis des Satzes basiert auf der Verteilungsgleichheit des trunkierten Prozesses und einem gemischt empirischen Punktprozess dessen zufällige Summationsgrenze binomialverteilt ist (vgl. Theorem 1.4.1 [48]). Für den approximierenden Prozess kann nun die Dichte- und somit die Likelihood-Funktion bestimmt werden (vgl. Theorem 3.1.1 [48], ähnlich in Abschnitt 7.1 in [8]).

Satz 2.9. (Dichtefunktion von Poisson-Prozessen)
Es seien N_1 und N_0 zwei Poisson-Prozesse mit endlichen Intensitätsmaßen ν_1 bzw. ν_0 auf dem messbaren Raum $(S, \mathcal{B})$. Weiterhin besitze ν_1 bezüglich ν_0 die Dichte h. Dann hat $\mathcal{L}(N_1)$ die $\mathcal{L}(N_0)$-Dichte

$$g(\mu) = \left(\prod_{i=1}^{\mu(S)} h(x_i) \right) \exp\left(\nu_0(S) - \nu_1(S)\right),$$

wobei $\mu = \sum_{i=1}^{\mu(S)} \epsilon_{x_i}$ ein Punktmaß aus $\mathbb{M}$ darstellt.

2.2. Identifizierbarkeit von Parametern

Ein Parameter ist identifizierbar, wenn zu einer gegebenen parameterabhängigen Verteilung verschiedene Parameter zu unterschiedlichen Verteilungen führen. Die Identifizierbarkeit stellt damit eine wichtige Voraussetzung für die Konsistenz von Schätzern dar und wird in diesem Abschnitt auf Grundlage der Kapitel 1.2 und 3.4 aus dem Buch [20] von Gouriéroux und Monfort eingeführt.
Auch im Folgenden bezeichne $(S, \mathcal{B})$ einen messbaren Raum und $(\Omega, \mathcal{F}, \mathbb{P})$ einen Wahrscheinlichkeitsraum.

Definition 2.10. (Identifizierbarkeit)
Es sei $Q_{\boldsymbol{\theta}}$ ein vom Parameter $\boldsymbol{\theta} \in \Theta$ abhängiges Wahrscheinlichkeitsmaß auf $\mathcal{B}$.

(*i*) Ein Parameterwert $\boldsymbol{\theta}_1 \in \Theta$ heißt *identifizierbar*, falls kein anderer Wert $\boldsymbol{\theta}_2 \in \Theta$ des Parameters $\boldsymbol{\theta}$ existiert, sodass $Q_{\boldsymbol{\theta}_2} = Q_{\boldsymbol{\theta}_1}$ gilt.

(*ii*) Ein Parameter heißt *identifizierbar*, wenn jeder Wert des Parameters identifiziert ist, d.h. wenn die Abbildung $\boldsymbol{\theta} \mapsto Q_{\boldsymbol{\theta}}$ injektiv auf Θ ist.

□

Hierbei ist die Gleichheit $Q_{\boldsymbol{\theta}_2} = Q_{\boldsymbol{\theta}_1}$ erfüllt, falls $Q_{\boldsymbol{\theta}_2}(B) = Q_{\boldsymbol{\theta}_1}(B)$ für alle $B \in \mathcal{B}$ gilt. Bei Existenz der Dichtefunktionen, lässt sich die Gleichheit zweier Verteilungen auch mithilfe des folgenden Abstandsmaßes überprüfen.

Definition 2.11. (Kullback-Leibler-Abstand)
Es bezeichnen Q_1 und Q_2 zwei Wahrscheinlichkeitsmaße mit den ξ-Dichten h_1 bzw. h_2. Weiter sei $Y : \Omega \to S$ eine Zufallsvariable, die gemäß Q_2 verteilt ist. Der *Kullback-Leibler-Abstand* zwischen Q_1 und Q_2 ist dann definiert durch

$$KL(Q_1 | Q_2) := \mathbb{E}\left[\log \frac{h_2(Y)}{h_1(Y)}\right] = \int \log \frac{h_2}{h_1} \cdot h_2 \, \mathrm{d}\xi.$$

□

Der Kullback-Leibler-Abstand ist keine Metrik, da er weder symmetrisch ist, noch die Dreiecksungleichung erfüllt. Er besitzt jedoch die folgenden Eigenschaften.

Satz 2.12.
Es seien Q_1 und Q_2 zwei Wahrscheinlichkeitsmaße, dann gelten

(*i*) $KL(Q_1|\,Q_2) \geq 0$,

(*ii*) $KL(Q_1|\,Q_2) = 0$ *genau dann, wenn* $Q_1 = Q_2$.

Beweis.

(*i*) Die Nichtnegativität des Abstands resultiert aus der Anwendung der Jensenschen Ungleichung auf die konvexe Funktion $x \mapsto -\log(x)$, denn

$$\begin{aligned} KL(Q_1|\,Q_2) &= \mathbb{E}\left[-\log\frac{h_1(Y)}{h_2(Y)}\right] \geq -\log\mathbb{E}\left[\frac{h_1(Y)}{h_2(Y)}\right] \\ &= -\log\int h_1\,\mathrm{d}\xi = -\log(1) = 0. \end{aligned}$$

(*ii*) Die Funktion $x \mapsto -\log(x)$ ist strikt konvex, wodurch die Gleichheit in der obigen Abschätzung nur erfüllt ist, wenn $h_1(Y)/h_2(Y)$ konstant ist. Diese Konstante muss eins sein, da $\mathbb{E}[h_1(Y)/h_2(Y)] = 1$ und es folgt $Q_1 = Q_2$.
Es sei umgekehrt $Q_1 = Q_2$ und somit $h_1(Y) = h_2(Y)$. Daraus erkennt man direkt

$$KL(Q_1|\,Q_2) = \mathbb{E}\left[\log\frac{h_2(Y)}{h_1(Y)}\right] = 0.$$

■

Da die Argumentation auf Basis der Likelihood geführt wird, ist die Aussage des Satzes nicht nur für reellwertige, sondern auch für maßwertige Y, wie für den trunkierten Prozess $N_{n,D}$ aus Beispiel 2.2, gültig.
In Hinblick auf die in Satz 2.12 dargestellten Eigenschaften, lässt sich die Identifizierbarkeit auch mittels des Kullback-Leibler-Abstands charakterisieren.

Satz 2.13.
Ein Parameter $\boldsymbol{\theta}$ ist genau dann identifizierbar, wenn

$$\forall\,\boldsymbol{\theta}_1,\,\boldsymbol{\theta}_2 \in \Theta:\; KL(Q_{\boldsymbol{\theta}_1}|Q_{\boldsymbol{\theta}_2}) = 0 \quad\Longrightarrow\quad \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2.$$

Eine weitere Charakterisierung resultiert aus der Tatsache, dass der Abstand $KL(Q_{\boldsymbol{\theta}_1}|\,Q_{\boldsymbol{\theta}_2})$ als Funktion in $\boldsymbol{\theta}_1$ sein Minimum in $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ annimmt. Diese Charakterisierung wird sich als äquivalent zu der Bedingung (*ii*) in Satz 4.10 (vgl. van der Vaart [61] Satz 5.7) herausstellen. Zur Wahrung der Übersicht werden die Erwartungswerte im folgenden Satz sowie in späteren Kapiteln mit dem der Verteilung zugrunde liegenden Parameter gekennzeichnet.

Satz 2.14.
Der Parameter $\boldsymbol{\theta}$ ist genau dann identifizierbar, wenn die Funktion $\boldsymbol{\theta}_1 \mapsto \mathbb{E}_{\boldsymbol{\theta}_2}[\log h_{\boldsymbol{\theta}_1}(Y)]$ für alle Parameterwerte $\boldsymbol{\theta}_2 \in \Theta$ ein eindeutiges Maximum in $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ besitzt.

Beweis. Es sei $\boldsymbol{\theta}$ identifizierbar. Dann folgt aus den Sätzen 2.12 sowie 2.13, dass die Funktion $\boldsymbol{\theta}_1 \mapsto KL(Q_{\boldsymbol{\theta}_1} | Q_{\boldsymbol{\theta}_2})$ für alle Parameterwerte $\boldsymbol{\theta}_2 \in \Theta$ ihr Minimum eindeutig in $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ annimmt. Es sei nun $\boldsymbol{\theta}_2 \in \Theta$ beliebig, aber fest. Dann führt die Definition 2.11 zu

$$\begin{aligned}\boldsymbol{\theta}_2 = \arg\min_{\boldsymbol{\theta}_1 \in \Theta} KL(Q_{\boldsymbol{\theta}_1} | Q_{\boldsymbol{\theta}_2}) &= \arg\min_{\boldsymbol{\theta}_1 \in \Theta} \left(\mathbb{E}_{\boldsymbol{\theta}_2}[\log h_{\boldsymbol{\theta}_2}(Y)] - \mathbb{E}_{\boldsymbol{\theta}_2}[\log h_{\boldsymbol{\theta}_1}(Y)]\right) \\ &= \arg\min_{\boldsymbol{\theta}_1 \in \Theta} -\mathbb{E}_{\boldsymbol{\theta}_2}[\log h_{\boldsymbol{\theta}_1}(Y)] = \arg\max_{\boldsymbol{\theta}_1 \in \Theta} \mathbb{E}_{\boldsymbol{\theta}_2}[\log h_{\boldsymbol{\theta}_1}(Y)],\end{aligned} \tag{2.2}$$

sodass die Funktion $\boldsymbol{\theta}_1 \mapsto \mathbb{E}_{\boldsymbol{\theta}_2}[\log h_{\boldsymbol{\theta}_1}(Y)]$ für alle Parameterwerte $\boldsymbol{\theta}_2 \in \Theta$ ein eindeutiges Maximum in $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ besitzt.
Andersherum führt die Eindeutigkeit des Maximums wegen Gleichung (2.2) auch zur Eindeutigkeit des Minimums der Funktion $\boldsymbol{\theta}_1 \mapsto KL(Q_{\boldsymbol{\theta}_1} | Q_{\boldsymbol{\theta}_2})$ für alle $\boldsymbol{\theta}_2 \in \Theta$. In Konsequenz ergibt sich $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ aus $KL(Q_{\boldsymbol{\theta}_1} | Q_{\boldsymbol{\theta}_2}) = 0$ für alle $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$ aus Satz 2.12, sodass der Parameter $\boldsymbol{\theta}$ wegen Satz 2.13 identifizierbar ist. ■

2.3. Copulas

Der Name *Copula* soll hervorheben, dass Copulas die Randverteilungsfunktionen von Zufallsvariablen und ihre gemeinsame Verteilungsfunktion koppeln (engl. to couple). Zu gegebenen Marginalverteilungen lässt sich mithilfe einer Copula eine gemeinsame Verteilung konstruieren. Dies ermöglicht eine sequentielle Modellierung stochastischer Abhängigkeit. Eine weitere Hauptanwendung von Copulas liegt in der Simulation. Der durch die Copula geschaffene Zusammenhang zwischen der gemeinsamen Verteilung und den Randverteilungsfunktionen lässt sich zum Generieren von Zufallsstichproben bezüglich einer bestimmten Verteilung ausnutzen.
Die folgenden Ausführungen zu den Copulas sind den Abschnitten 2.2, 2.3 sowie 2.9 des Buches [45] von Nelsen entnommen. Auf die Beweise der aufgestellten Behauptungen wird an dieser Stelle verzichtet. Wenn Aussagen über fast alle Elemente einer Menge getroffen werden, so ist dies stets für das Lebesgue-Maß zu verstehen.

Definition 2.15. (Copula)
Eine *Copula* ist eine Funktion $C : [0,1] \times [0,1] \to [0,1]$ mit den folgenden Eigenschaften:

(*i*) Für alle $u, v \in [0,1]$ gelten

$$C(u,0) = C(0,v) = 0, \quad C(u,1) = u \quad \text{sowie} \quad C(1,v) = v.$$

(*ii*) Für alle $u_1, u_2, v_1, v_2 \in [0,1]$ mit $u_1 \leq u_2$ und $v_1 \leq v_2$ gilt

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

□

Harry Joe führt in seinem Buch [26] eine Copula als mehrdimensionale Verteilungsfunktion ein, deren eindimensionale Randverteilungen gleichverteilt auf dem Intervall

$[0,1]$ sind. Die Theorie der Copulas ist auf $p \in \mathbb{N}$ Dimensionen verallgemeinerbar. Eine Verallgemeinerung wird in dieser Arbeit jedoch nicht benötigt.

Beispiel 2.16. (Produktcopula)
Eine der einfachsten und bekanntesten Copulas ist die sogenannte *Produktcopula*. Sie ist für $u, v \in [0,1]$ definiert durch

$$\Pi(u,v) := u \cdot v.$$

Sie wird auch als *Unabhängigkeitscopula* bezeichnet, da mit ihr die stochastische Unabhängigkeit zwischen zwei Zufallsvariablen formuliert werden kann. Es seien X und T stochastisch unabhängige Zufallsvariablen mit den Verteilungsfunktionen F^X und F^T, so gilt für die gemeinsame Verteilungsfunktion H, dass

$$H(x,t) = F^X(x) \cdot F^T(t) = \Pi(F^X(x), F^T(t)).$$

□

Der folgende Satz ist von zentraler Bedeutung für die Theorie der Copulas und bildet die Grundlage für viele Anwendungen im Bereich der Statistik. Er verdeutlicht die Rolle der Copulas im Zusammenhang zwischen den mehrdimensionalen und den eindimensionalen Verteilungen.

Satz 2.17. (Satz von Sklar)
Es bezeichne $H : \overline{\mathbb{R}}^2 \to [0,1]$ eine zweidimensionale Verteilungsfunktion des Vektors (X,T) mit den eindimensionalen Randverteilungen F^X, $F^T : \overline{\mathbb{R}} \to [0,1]$. Dann existiert eine Copula C, sodass

$$H(x,t) = C(F^X(x), F^T(t)) \tag{2.3}$$

für alle $x,t \in \overline{\mathbb{R}}$ erfüllt ist. Sind F^X und F^T stetig, dann ist C eindeutig, sonst ist C nur auf dem Produkt der Bildmengen von F^X und F^T, d.h. auf $\mathrm{Ran}F^X \times \mathrm{Ran}F^T$ eindeutig bestimmt.
Ist andersherum C eine Copula und sind F^X und F^T Verteilungsfunktionen, dann ist die in (2.3) definierte Funktion H eine gemeinsame Verteilungsfunktion mit den Randverteilungen F^X und F^T.

Die folgende Copula wird für die spätere Abhängigkeitsmodellierung verwendet.

Beispiel 2.18. (Gumbel-Barnett-Copula)
Die bivariate Verteilungsfunktion H_ϑ sei für den Parameter $\vartheta \in [0,1]$ gegeben durch

$$H_\vartheta(x,t) = \begin{cases} 1 - e^{-x} - e^{-t} + e^{-(x+t+\vartheta xt)}, & x \geq 0,\ t \geq 0, \\ 0, & \text{sonst.} \end{cases}$$

Dann sind die eindimensionalen Randverteilungen Exponentialverteilungen zum Para-

meter 1, d.h.

$$F^X(x) = \begin{cases} 1 - e^{-x}, & x \geq 0, \\ 0, & \text{sonst}, \end{cases} \quad \text{sowie} \quad F^T(t) = \begin{cases} 1 - e^{-t}, & t \geq 0, \\ 0, & \text{sonst}. \end{cases}$$

Nach dem Satz von Sklar existiert nun eine Copula C_ϑ, sodass für alle $x, t \in \overline{\mathbb{R}}$ der Zusammenhang $H_\vartheta(x,t) = C_\vartheta(F^X(x), F^T(t))$ gilt. Diese ist durch

$$C_\vartheta(u,v) := u + v - 1 + (1-u)(1-v)e^{-\vartheta \log(1-u)\log(1-v)} \tag{2.4}$$

für $u, v \in [0,1]$ gegeben. Man beachte, dass für $\vartheta = 0$ Unabhängigkeit vorliegt und somit $C_\vartheta(F^X(x), F^T(t)) = F^X(x) \cdot F^T(t)$ gilt. □

Aus dem Satz von Sklar geht hervor, dass sich die gemeinsame Verteilungsfunktion mithilfe der Randverteilungen und der entsprechenden Copula darstellen lässt. Andersherum kann mithilfe der gemeinsamen Verteilung sowie der Inversen der Randverteilungen eine Copula konstruiert werden. Sind die Randverteilungen nicht streng monoton wachsend, so besitzen diese lediglich Pseudoinversen.

Definition 2.19. (Pseudeinverse)
Es bezeichne F eine Verteilungsfunktion. Dann ist eine Pseudoinverse von F eine Funktion $F^{(-1)} : [0,1] \to \overline{\mathbb{R}}$, welche die folgenden Eigenschaften erfüllt:

1. Für $t \in \text{Ran}F$ gilt $F(F^{(-1)}(t)) = t$.
2. Für $t \notin \text{Ran}F$ gilt $F^{(-1)}(t) = \inf\{x | F(x) \geq t\} = \sup\{x | F(x) \leq t\}$.

Ist F streng monoton, so besitzt die Funktion eine eindeutige Inverse, die wie üblich mit F^{-1} bezeichnet wird. □

Korollar 2.20.
Es seien H, F^X, F^T und C wie in Satz 2.17. Sind die Funktionen F^X und F^T zusätzlich stetig, dann gilt für alle $u, v \in [0,1]$, dass

$$C(u,v) = H((F^X)^{(-1)}(u), (F^T)^{(-1)}(v)),$$

wobei $(F^X)^{(-1)}$ und $(F^T)^{(-1)}$ die Pseudoinversen von F^X und F^T bezeichnen.

Neben der bivariaten Verteilungsfunktion aus Beispiel 2.18, aus der sich die Gumbel-Barnett-Copula ergibt, hat Gumbel in [23] eine weitere bivariate Exponentialverteilung untersucht, welche zur bekannten Farlie-Gumbel-Morgenstern-Copula führt.

Beispiel 2.21. (Farlie-Gumbel-Morgenstern-Copula)
Mit $\vartheta \in [-1,1]$ sei die bivariate Verteilungsfunktion

$$H_\vartheta(x,t) = \begin{cases} (1 - e^{-x})(1 - e^{-t})[1 + \vartheta e^{-x-t}], & x \geq 0,\ t \geq 0, \\ 0, & \text{sonst}. \end{cases}$$

gegeben. Wie im vorherigen Beispiel sind die Randverteilungen Exponentialverteilungen zum Parameter 1. Für u, $v \in [0,1]$ erhält man somit als Inversen

$$(F^X)^{-1}(u) = -\log(1-u) \quad \text{sowie} \quad (F^T)^{-1}(v) = -\log(1-v).$$

Nach Korollar 2.20 führt das Einsetzen der Inversen der Randverteilungsfunktionen in die gemeinsame Verteilungsfunktion zu der zugrunde liegenden Copula C_ϑ^{FGM}, welche gegeben ist durch

$$C_\vartheta^{FGM}(u,v) := u \cdot v \cdot [1 + \vartheta(1-u)(1-v)]. \tag{2.5}$$

Auch für diese Copula liegt für $\vartheta = 0$ Unabhängigkeit zwischen X und T vor. □

Die Simulation von Zufallsstichproben mithilfe von Copulas erfolgt mit der bedingten Inversionsmethode. Diese setzt sich aus der Inversionsmethode sowie der Methode der bedingten Verteilungen zusammen, die im Folgenden gemäß Devroye [10] (vgl. Abschnitte II.2 und XI.1) kurz vorgestellt werden.
Die bekannte Inversionsmethode basiert auf dem folgenden Zusammenhang:

Es bezeichne $F^{(-1)}$ eine Pseudoinverse einer Verteilungsfunktion F und U eine auf dem Intervall $(0,1)$ gleichverteilte Zufallsvariable. Dann hat $F^{(-1)}(U)$ die Verteilungsfunktion F.

Inversionsmethode

Eine Zufallszahl x bezüglich der Verteilung F wird durch das folgende Vorgehen generiert:

- Erzeuge eine Zufallszahl u aus einer $(0,1)$-Gleichverteilung.
- Setze $x = F^{(-1)}(u)$.

Bei dem Erzeugen von Stichproben mehrdimensionaler Zufallsvektoren müssen die Abhängigkeiten zwischen den Komponenten des Zufallsvektors berücksichtigt werden. Ein möglicher Ansatz funktioniert mithilfe bedingter Verteilungsfunktionen. Dabei wird das multivariate Problem in mehrere eindimensionale Probleme zerlegt. Um das rekursive Vorgehen der Methode zu verdeutlichen, wird diese für $k \in \mathbb{N}$ Dimensionen erläutert.

Methode der bedingten Verteilungen

Eine Realisation $(x_1, \ldots, x_k)$ eines Zufallsvektors $(X_1, \ldots, X_k)$ erhält man wie folgt:

- Erzeuge x_1 gemäß der Verteilung von X_1.
- Erzeuge x_i gemäß $\mathbb{P}(X_i \in \cdot | X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$ für $i = 2, \ldots, k$.

Die bedingte Verteilung berechnet sich dabei durch

$$\mathbb{P}(X_i \in \cdot | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \frac{\mathbb{P}(X_1 = x_1, \ldots, X_{i-1} = x_{i-1}, X_i \in \cdot)}{\mathbb{P}(X_1 = x_1, \ldots, X_{i-1} = x_{i-1})}.$$

Die Anwendung der Methode der bedingten Verteilungen setzt somit die Kenntnis der Randverteilungen voraus.
Es bezeichnen nun X und T zwei Zufallsvariablen mit den stetigen Verteilungen F^X und F^T sowie der Copula C. Nach Korollar 2.20 besitzen $U = F^X(X)$ und $V = F^T(T)$ die gemeinsame Verteilungsfunktion C. Um mithilfe der vorgestellten Methoden eine Realisation des Vektors $(X, T)^{\mathrm{T}}$ mithilfe der Copula C zu generieren, wird zunächst die bedingte Verteilung von V gegeben $U = u$ benötigt. Für $u, v \in [0, 1]$ gilt

$$\begin{aligned}
\mathbb{P}(V \leq v | U = u) &= \lim_{h \searrow 0} \mathbb{P}(V \leq v | u \leq U \leq u + h) \\
&= \lim_{h \searrow 0} \frac{\mathbb{P}(V \leq v, u \leq U \leq u + h)}{\mathbb{P}(u \leq U \leq u + h)} \\
&= \lim_{h \searrow 0} \frac{\mathbb{P}(V \leq v, U \leq u + h) - \mathbb{P}(V \leq v, U \leq u)}{\mathbb{P}(U \leq u + h) - \mathbb{P}(U \leq u)} \\
&= \lim_{h \searrow 0} \frac{C(u + h, v) - C(u, v)}{h} = \frac{\partial C(u, v)}{\partial u}, \qquad (2.6)
\end{aligned}$$

wobei im vorletzten Schritt ausgenutzt wurde, dass U gleichverteilt ist. Es bezeichne

$$c_u(v) := \frac{\partial C(u, v)}{\partial u}. \qquad (2.7)$$

Der nächste Satz sichert die Existenz der Funktion c_u sowie einer zugehörigen Pseudoinversen $c_u^{(-1)}$.

Satz 2.22.
Es sei C eine Copula. Für alle $v \in [0, 1]$ existiert die partielle Ableitung $\partial C(u, v)/\partial u$ für fast alle $u \in [0, 1]$. Im Falle der Existenz ist

$$0 \leq \frac{\partial}{\partial u} C(u, v) \leq 1.$$

Ähnlich existiert die partielle Ableitung $\partial C(u, v)/\partial v$ fast überall für alle $u \in [0, 1]$ und es gilt dann

$$0 \leq \frac{\partial}{\partial v} C(u, v) \leq 1.$$

Weiterhin sind die Funktionen $u \mapsto \partial C(u, v)/\partial v$ und $v \mapsto \partial C(u, v)/\partial u$ auf dem Intervall $[0, 1]$ fast überall monoton wachsend.

2.23. Bedingte Inversionsmethode

Mit den nachstehenden Schritten lassen sich Zufallsstichproben des Vektors $(X, T)^{\mathrm{T}}$ mithilfe der Copula C simulieren.

- Erzeuge zwei unabhängige Realisationen u und v' einer $(0, 1)$-Gleichverteilung.
- Setze $v = c_u^{(-1)}(v')$.
- Erhalte mit $x = (F^X)^{(-1)}(u)$ und $t = (F^T)^{(-1)}(v)$ die gesuchte Realisation von $(X, T)^{\mathrm{T}}$.

Setzt man für die Randverteilungen F^X und F^T jeweils die Gleichverteilung auf $(0, 1)$ ein, so lassen sich mithilfe der bedingten Inversionsmethode Werte gemäß der Copula aus dem Beispiel 2.18 simulieren.

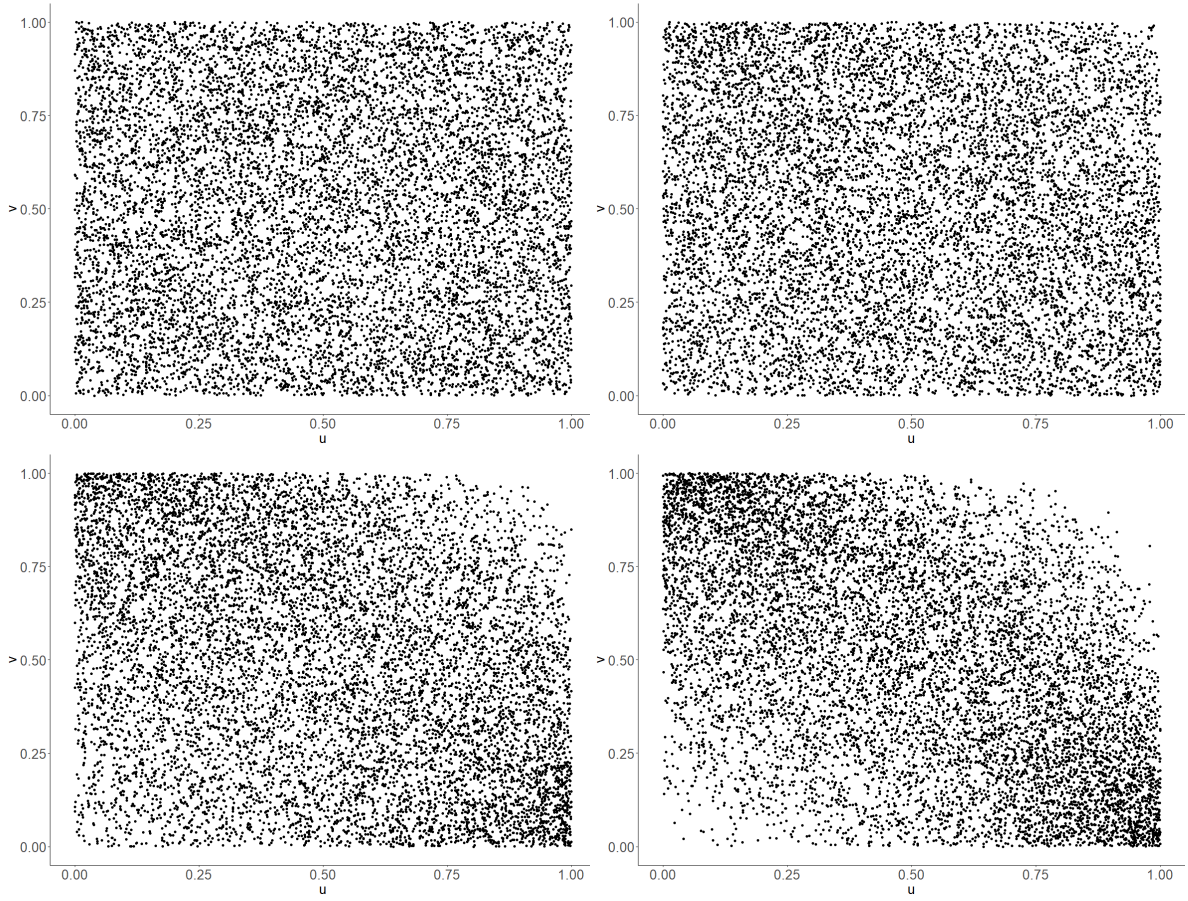

Abbildung 2.1.: Scatterplot von jeweils 10.000 Realisationen der Gumbel-Barnett-Copula (2.4) für die Parameterwerte $\vartheta = 0$; $\vartheta = 0{,}1$; $\vartheta = 0{,}5$ und $\vartheta = 1$

Die Abbildung 2.1 zeigt für verschiedene Parameterwerte von ϑ jeweils 10.000 Punkte, die gemäß der Gumbel-Barnett-Copula (2.4), das heißt für $c_u(v) = 1 + (1 - v)^{1-\vartheta \log(1-u)}[\vartheta \log(1 - v) - 1]$ erzeugt wurden. Im Koordinatensystem links oben wird dabei der Unabhängigkeitsfall für $\vartheta = 0$ dargestellt. Die Realisationen

verteilen sich gleichmäßig über das Quadrat $[0,1] \times [0,1]$. Der größte Kontrast dazu ist in dem Koordinatensystem rechts unten für den Fall $\vartheta = 1$ erkennbar. Zwischen den beiden Koordinaten besteht ein, wenn auch nicht besonders stark ausgeprägter, negativer Zusammenhang. Die weiteren Koordinatensysteme zeigen die Abhängigkeitsverhältnisse für $\vartheta = 0{,}1$ und $\vartheta = 0{,}5$. Je größer der Abhängigkeitsparameter gewählt wird, umso weniger Realisationen befinden sich in der linken unteren und rechten oberen Ecke des Einheitsquadrats.

Die bedingte Inversionsmethode ermöglicht demnach die Generierung von Punkteplots entsprechend der jeweiligen Copula. Zudem gestattet sie die grafische Beurteilung der zugrunde liegenden Abhängigkeitsverhältnisse zwischen den beiden Variablen. Eine rechnerische Ermittlung ist beispielsweise mit dem Rangkorrelationskoeffizienten *Kendalls Tau* möglich. Dieses Zusammenhangsmaß wird nicht von den Randverteilungen der beiden Zufallsvariablen beeinflusst und hängt nur von der entsprechenden Copula ab. Es lässt sich wie folgt nach Satz 5.1.3 in [45] berechnen.

Satz 2.24.
Es seien X und T zwei stetige Zufallsvariablen mit der Copula C. Dann berechnet sich Kendalls Tau τ_C durch

$$\tau_C = 4 \int_0^1 \int_0^1 C(u,v)\, \mathrm{d}C(u,v) - 1.$$

Für das Zusammenhangsmaß gilt $\tau_C \in [-1,1]$, wobei positive Werte einen positiven monotonen Zusammenhang und negative Werte einen negativen monotonen Zusammenhang zwischen den Variablen signalisieren. Bei einem Wert von null liegt kein monotoner Zusammenhang vor. Dies ist beispielsweise bei der Produktcopula, mit deren Hilfe Unabhängigkeit modelliert werden kann, der Fall.

Beispiel 2.25.
Für $\Pi(u,v) = uv$ mit $\partial^2 \Pi(u,v)/(\partial u \partial v) = 1$ ist

$$\tau_\Pi = 4 \int_0^1 \int_0^1 \Pi(u,v)\, \mathrm{d}\Pi(u,v) - 1 = 4 \int_0^1 \int_0^1 uv \cdot 1 \,\mathrm{d}u\,\mathrm{d}v - 1 = 4 \cdot \frac{1}{4} - 1 = 0.$$

□

Für die Gumbel-Barnett-Copula bestätigt sich der in Abbildung 2.1 gewonnene Eindruck über den negativen Zusammenhang zwischen den Zufallsvariablen in den möglichen Werten des Zusammenhangsmaßes. Da die Copula C_ϑ vom Parameter ϑ abhängt, wird die Notation τ_ϑ statt τ_{C_ϑ} verwendet.

Beispiel 2.26.
Für die Gumbel-Barnett-Copula C_ϑ (2.4) mit $\vartheta \in [0,1]$ erhält man

$$\frac{\partial^2 C_\vartheta(u,v)}{\partial u \partial v} = e^{-\vartheta \log(1-u)\log(1-v)}[(1 - \vartheta \log(1-v))(1 - \vartheta \log(1-u)) - \vartheta].$$

Kendalls Tau

$$\tau_\vartheta = 4\int_0^1\int_0^1 C_\vartheta(u,v)\frac{\partial^2 C_\vartheta(u,v)}{\partial u\partial v}\,\mathrm{d}u\,\mathrm{d}v - 1$$

lässt sich numerisch berechnen und man erhält $\tau_\vartheta \in [-0{,}36114; 0]$. Für $\vartheta = 0$ gilt $\tau_\vartheta = 0$ und τ_ϑ ist streng monoton fallend in ϑ. □

Mithilfe der Farlie-Gumbel-Morgenstern-Copula lassen sich hingegen sowohl positive als auch negative Zusammenhänge modellieren.

Beispiel 2.27.
Für C_ϑ^{FGM} wie in (2.5) mit $\vartheta \in [-1, 1]$ gilt

$$\frac{\partial^2 C_\vartheta^{FGM}(u,v)}{\partial u\partial v} = 1 + \vartheta - 2\vartheta v - 2\vartheta u + 4\vartheta uv.$$

Kendalls Tau kann exakt in Abhängigkeit von ϑ berechnet werden und lautet

$$\tau_\vartheta^{FGM} = 4\int_0^1\int_0^1 C_\vartheta^{FGM}(u,v)\frac{\partial^2 C_\vartheta^{FGM}(u,v)}{\partial u\partial v}\,\mathrm{d}u\,\mathrm{d}v - 1 = \frac{2}{9}\vartheta.$$

Somit ist τ_ϑ^{FGM} streng monoton steigend in ϑ und nimmt Werte zwischen $-0{,}22$ und $0{,}22$ an. Bei Unabhängigkeit, d.h. $\vartheta = 0$, ist folglich $\tau_\vartheta^{FGM} = 0$. □

Obwohl mit der Gumbel-Barnett-Copula nur negative Abhängigkeiten modelliert werden können, ist es möglich, die zugrunde liegenden Zufallsvariablen an den Achsen $u = 1/2$ oder $v = 1/2$ zu spiegeln und damit positive Abhängigkeiten zu modellieren. Es seien U und V zwei auf dem Intervall $[0, 1]$ gleichverteilte Zufallsvariablen mit gemeinsamer Verteilungsfunktion und Copula C. Gemäß Abschnitt 1.6 in [26] ergeben sich dann jeweils die folgenden gemeinsamen Verteilungsfunktionen bzw. Copulas:

Tabelle 2.1.: Übersicht der zu den Spiegelungen assoziierten Copulas

Vektor	Spiegelachse	Copula
$(1-U, V)$	$u = 1/2$	$C^{270}(u,v) = v - C(1-u, v)$
$(U, 1-V)$	$v = 1/2$	$C^{90}(u,v) = u - C(u, 1-v)$
$(1-U, 1-V)$	$v = 1-u$	$C^{180}(u,v) = u + v - 1 + C(1-u, 1-v)$

Die Spiegelungen entsprechen Drehungen um jeweils 270, 90 bzw. 180 Grad im Uhrzeigersinn um den Punkt $(1/2, 1/2)$.

Satz 2.28.
Kendalls Tau berechnet sich für die Copulas aus Tabelle 2.1 wie folgt

$$\tau_C^{270} = -\tau_C, \qquad \tau_C^{90} = -\tau_C, \qquad \tau_C^{180} = \tau_C.$$

Beweis. Unter Annahme der Existenz der zweiten partiellen Ableitungen folgt unter Anwendung der Kettenregel $\frac{\partial^2 C^{270}(u,v)}{\partial u \partial v} = \frac{\partial^2 C}{\partial u \partial v}(1-u,v)$ für $C^{270}(u,v) = v - C(1-u,v)$. Mithilfe der Substitutionsregel ergibt sich

$$\begin{aligned}
\tau_C^{270} &= 4\int_0^1\int_0^1 [v - C(1-u,v)]\,\frac{\partial^2 C}{\partial u \partial v}(1-u,v)\,\mathrm{d}u\,\mathrm{d}v - 1 \\
&= 4\int_0^1\int_0^1 [v - C(u,v)]\,\frac{\partial^2 C(u,v)}{\partial u \partial v}\,\mathrm{d}u\,\mathrm{d}v - 1 \\
&= 4\int_0^1\int_0^1 v \cdot \frac{\partial^2 C(u,v)}{\partial u \partial v}\,\mathrm{d}u\,\mathrm{d}v - 4\int_0^1\int_0^1 C(u,v)\cdot\frac{\partial^2 C(u,v)}{\partial u \partial v}\,\mathrm{d}u\,\mathrm{d}v - 1 \\
&= -\tau_C + 4\int_0^1 v\,\mathrm{d}v - 2 = -\tau_C.
\end{aligned}$$

Dabei folgen $\partial C(0,v)/\partial v = 0$ und $\partial C(1,v)/\partial v = 1$ aus $C(0,v) = 0$ bzw. $C(1,v) = v$ für alle $v \in [0,1]$.
Die Berechnungen für τ_C^{90} können analog vorgenommen werden und τ_C^{180} ergibt sich als Kombination aus beiden Spiegelungen. ■

Beispiel 2.29.
Für die Gumbel-Barnett-Copula (2.4) mit $\vartheta \in [0,1]$ gilt

$$C_\vartheta^{270}(u,v) = u - u(1-v)e^{-\vartheta \log(u)\log(1-v)} \tag{2.8}$$

und $\tau_\vartheta^{270} \in [0; 0{,}36114]$. Die folgende Abbildung vergleicht die Scatterplots von C_ϑ und C_ϑ^{270} miteinander.

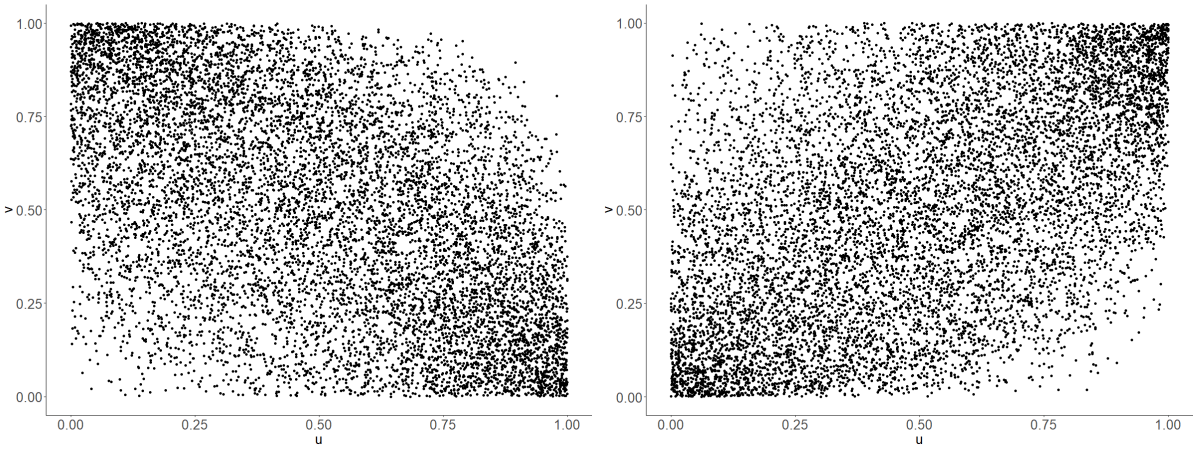

Abbildung 2.2.: Scatterplot von jeweils 10.000 Realisationen der Copulas C_ϑ (2.4) und C_ϑ^{270} (2.8) für den Parameterwert $\vartheta = 1$

□

3. Modell

Nachdem die Grundlagen gelegt wurden, wird mithilfe der Gumbel-Barnett-Copula eine bivariate Verteilungsfunktion erstellt und auf Basis der Theorie der Punktprozesse die Likelihood hergeleitet. Anschließend wird darauf aufbauend ein Z-Schätzer konstruiert. Der Nachweis der Modelleigenschaften erfolgt ausschließlich für die Gumbel-Barnett-Copula. Hinsichtlich der späteren Anwendung auf Daten werden jedoch die Likelihoods für zwei weitere Copulas angeführt.

3.1. Einführung des Modells

Es sei $(\Omega, \mathcal{F}, \mathbb{P}_{\boldsymbol{\theta}})$ für alle $\boldsymbol{\theta} \in \Theta$ ein Wahrscheinlichkeitsraum. Für das Modell wird die einfache Zufallsstichprobe $(X_i, T_i)^{\mathrm{T}}$, $i = 1, \ldots, n$, $n \in \mathbb{N}$ betrachtet, d.h. die Zufallsvektoren $(X_i, T_i)^{\mathrm{T}}$ seien unabhängig identisch verteilt. Dabei bezeichne $X_i \in \mathbb{R}_0^+$ die Lebensdauer eines Individuums oder Unternehmens und $T_i \in [0, G]$ das Alter bei Studienbeginn, wobei nur diejenigen Individuen betrachtet werden, die höchstens $0 < G < \infty$ Zeiteinheiten vor Studienbeginn geboren wurden. Die Studie beginnt für alle Einheiten zur gleichen Kalenderzeit. Es sei weiter $S := \mathbb{R}_0^+ \times [0, G]$ der Wertebereich der betrachteten Zufallsvektoren und $\mathcal{B} := \sigma(\{A \subset S |\, A \text{ offen}\})$ die von den offen Teilmengen von S erzeugte σ-Algebra, sodass $(X_i, T_i)^{\mathrm{T}} : \Omega \to S$ für alle $i = 1, \ldots, n$ eine $(\mathcal{F}, \mathcal{B})$-messbare Funktion ist. Es werden weiterhin die folgenden Notationen und Annahmen benötigt:

(A1) Es sei $\Theta := [\varepsilon_\theta, 1/\varepsilon_\theta] \times [0, 1 - \varepsilon_\vartheta]$, für ε_ϑ, $\varepsilon_\theta \in (0, 1)$ klein, der Parameterraum. Der zu schätzende wahre Parameterwert wird im Folgenden mit $\boldsymbol{\theta}_0 \in \Theta$ bezeichnet.

(A2) Die Lebensdauer X_i sei exponentialverteilt zum Parameter $\theta \in [\varepsilon_\theta, 1/\varepsilon_\theta]$. Das Alter bei Studienbeginn T_i sei gleichverteilt auf dem Intervall $[0, G]$. Die Zufallsgrößen besitzen somit die Verteilungsfunktionen

$$F^X(x) := \begin{cases} 1 - e^{-\theta x}, & x \geq 0, \\ 0, & \text{sonst} \end{cases} \quad \text{sowie} \quad F^T(t) := \begin{cases} 0, & t < 0, \\ \frac{t}{G}, & 0 \leq t < G, \\ 1, & t \geq G. \end{cases}$$

(A3) Zur Modellierung der Abhängigkeit von X_i und T_i wird eine Copula verwendet, sodass die gewählten Randverteilungen erhalten bleiben. Setzt man die eindimensionalen Verteilungen F^X und F^T in die Gumbel-Barnett-Copula (2.4)

$$C_\vartheta(u, v) = u + v - 1 + (1 - u)(1 - v) e^{-\vartheta \log(1-u) \log(1-v)}, \; \vartheta \in [0, 1 - \varepsilon_\vartheta],$$

ein, so sichert der Satz 2.17, dass es sich bei der daraus resultierenden Funktion $F_{\boldsymbol{\theta}}(x,t) := C_\vartheta(F^X(x), F^T(t))$ mit $\boldsymbol{\theta} = (\theta, \vartheta)^{\mathrm{T}} \in \Theta$ wieder um eine zweidimensionale Verteilungsfunktion handelt. Diese hat die Form

$$F_{\boldsymbol{\theta}}(x,t) = \begin{cases} \frac{t}{G} - e^{-\theta x} + e^{-\theta x}\left(1 - \frac{t}{G}\right)^{\vartheta\theta x+1}, & x > 0,\ 0 < t < G, \\ 1 - e^{-\theta x}, & x > 0,\ t \geq G, \\ 0, & \text{sonst.} \end{cases}$$

Daraus lässt sich die gemeinsame Dichtefunktion ableiten, welche für $x > 0$ und $0 < t < G$ durch

$$f_{\boldsymbol{\theta}}(x,t) := -\frac{\theta}{G} e^{-\theta x}\left(1 - \frac{t}{G}\right)^{\vartheta\theta x} [(\vartheta\theta x + 1)(\vartheta \log(1 - t/G) - 1) + \vartheta] \tag{3.1}$$

gegeben ist. Für alle weiteren Werte von x und t nimmt $f_{\boldsymbol{\theta}}$ den Wert null an.

(A4) Zu einer gegebenen Konstante $s > 0$ wird der Vektor $(X_i, T_i)^{\mathrm{T}}$ beobachtet, wenn er in der Menge

$$D := \{(x,t)^{\mathrm{T}} | 0 < t \leq x \leq t + s, t \leq G\}$$

enthalten ist. Die Abbildung 1.1 zeigt eine grafische Darstellung der Menge D. Beobachtungen werden im Folgenden mit $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$ bezeichnet und mit $j = 1, \ldots, M_n \leq n$ umnummeriert. Es gilt $M_n = \sum_{i=1}^n \mathbb{1}_{\{(X_i,T_i)^{\mathrm{T}} \in D\}}$, sodass es sich bei M_n um eine vom Zufall abhängige Größe handelt. Insbesondere ist M_n binomialverteilt.

Die Abbildung 3.1 zeigt jeweils 10.000 Realisationen, die entsprechend der Annahmen (A1)-(A3) mithilfe der bedingten Inversionsmethode 2.23 erzeugt wurden. Dafür wurden die Parameterwerte $\theta = 0{,}08$ und $\vartheta \in \{0;\ 0{,}1;\ 0{,}5;\ 1\}$ gewählt. Die Scatterplots der Gumbel-Barnett-Copula in Abbildung 2.1 veranschaulichen die Abhängigkeitsverhältnisse zwischen den beiden Zufallsgrößen, ohne die Einwirkung von Randdichten. In der Abbildung 3.1 liegen nun die gleichen Abhängigkeitsverhältnisse zwischen den Zufallsgrößen vor, nur dass hierfür die Randdichten F^X und F^T hinzugezogen wurden. Wie in Abbildung 2.1 zeigt das Koordinatensystem links oben den Unabhängigkeitsfall für $\vartheta = 0$, sodass sich die Punkte gleichmäßig über die vertikale Achse verteilen. Der hohe Anteil an Realisationen mit geringen X-Ausprägungen ist durch die Exponentialverteilung begründet. Mit steigendem Abhängigkeitsparameter ϑ ist auch hier erkennbar, dass größere X-Ausprägungen tendenziell mit kleineren T-Ausprägungen einhergehen.

Bemerkung 3.1.
Im Rahmen dieser Arbeit wird das zur Verteilungsfunktion $F_{\boldsymbol{\theta}}$ gehörige Wahrscheinlichkeitsmaß mit $\mathbb{P}_{\boldsymbol{\theta}}$ bezeichnet. Bei allen weiteren Zufallsgrößen ist es nicht erforderlich, in der Notation zwischen den verschiedenen Wahrscheinlichkeitsmaßen zu unterscheiden. Stattdessen wird einheitlich das Symbol $\mathbb{P}$ verwendet. □

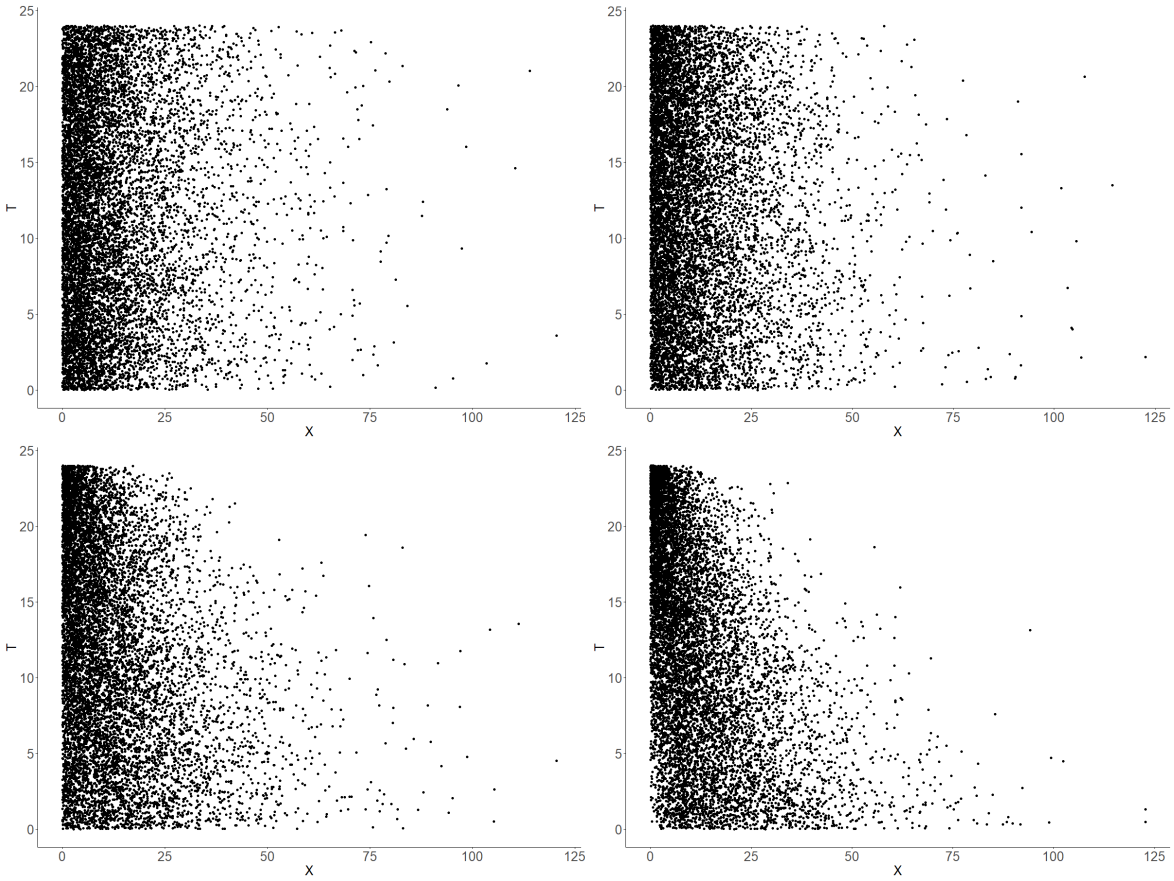

Abbildung 3.1.: Scatterplot von jeweils 10.000 Realisationen, die mithilfe der Dichte (3.1) für die Parameterwerte $\theta = 0{,}08$ sowie $\vartheta = 0$; $\vartheta = 0{,}1$; $\vartheta = 0{,}5$ und $\vartheta = 1$ simuliert wurden

Eine wichtige Kennzahl für die Parameterschätzung in diesem Modell ist die Wahrscheinlichkeit, dass das i-te Individuum ausgewählt beziehungsweise beobachtet wird. Die Auswahlwahrscheinlichkeit $\mathbb{P}_{\boldsymbol{\theta}}(T_i \leq X_i \leq T_i + s)$ wird im Folgenden für $\boldsymbol{\theta} \in \Theta$ mit $\alpha_{\boldsymbol{\theta}}$ bezeichnet und berechnet sich durch die Integration der gemeinsamen Dichtefunktion über dem Parallelogramm D (vgl. Abbildung 1.1), d.h.

$$\alpha_{\boldsymbol{\theta}} = \int_0^G \int_t^{t+s} f_{\boldsymbol{\theta}}(x,t)\,\mathrm{d}x\,\mathrm{d}t.$$

Ähnlich wie in Emura und Pan (vgl. Theorem 1, [15]) lässt sich die Auswahlwahrscheinlichkeit mithilfe der Copula so umschreiben, dass nur noch ein Integral benötigt wird.

Lemma 3.2.
Mit der in (2.7) *eingeführten Notation lässt sich die Auswahlwahrscheinlichkeit als das univariate Integral*

$$\alpha_{\boldsymbol{\theta}} = \int_0^1 \left[c_u\{F^T((F^X)^{-1}(u))\} - c_u\{F^T((F^X)^{-1}(u)) - F^T(s)\} \right] \mathrm{d}u$$

mit

$$c_u(v) = \frac{\partial C_\vartheta(u,v)}{\partial u} = 1 - (1-v)e^{-\vartheta \log(1-u)\log(1-v)}[1 - \vartheta \log(1-v)]$$

sowie

$$(F^X)^{-1}(u) = -\frac{\log(1-u)}{\theta}$$

schreiben.

Beweis. Als monoton wachsende Funktion erhält die Funktion F^T Ungleichungen, sodass

$$\alpha_{\boldsymbol{\theta}} = \mathbb{P}_{\boldsymbol{\theta}}(T_i \leq X_i \leq T_i + s) = \mathbb{P}_{\boldsymbol{\theta}}(F^T(T_i) \leq F^T(X_i) \leq F^T(T_i + s)).$$

Ersetzt man $F^T(T_i)$ durch die auf dem Intervall $[0,1]$ gleichverteilte Zufallsgröße U_1, so lässt sich $F^T(T_i + s)$ durch $U_1 + F^T(s)$ ersetzen. Weiterhin folgt $X_i = (F^X)^{-1}(U_2)$ aus $U_2 = F^X(X_i)$ ebenfalls für eine auf dem Intervall $[0,1]$ gleichverteilte Zufallsgröße U_2. Mithilfe der iterierten Erwartungswertbildung ergibt sich

$$\begin{aligned}
&\mathbb{P}(F^T((F^X)^{-1}(U_2)) \leq U_1 + F^T(s)) - \mathbb{P}(F^T((F^X)^{-1}(U_2)) \leq U_1) \\
&= \mathbb{E}\left[\mathbb{P}(F^T((F^X)^{-1}(U_2)) \leq U_1 + F^T(s)|U_2)\right] - \mathbb{E}\left[\mathbb{P}(F^T((F^X)^{-1}(U_2)) \leq U_1|U_2)\right] \\
&= \mathbb{E}\left[\mathbb{P}(U_1 \leq F^T((F^X)^{-1}(U_2))|U_2)\right] - \mathbb{E}\left[\mathbb{P}(U_1 \leq F^T((F^X)^{-1}(U_2)) - F^T(s)|U_2)\right] \\
&= \int_0^1 \left[c_u(F^T((F^X)^{-1}(u))) - c_u(F^T((F^X)^{-1}(u)) - F^T(s))\right] \mathrm{d}u.
\end{aligned}$$

Im letzten Schritt wurde Gleichung (2.6) verwendet. ■

Diese Darstellung vereinfacht die Berechnungen, vor allem wenn wiederholte Auswertungen für verschiedene Parameterwerte benötigt werden.
In Hinblick auf die noch folgenden Beweise wird die Auswahlwahrscheinlichkeit zunächst auf Eigenschaften wie Stetigkeit und stetige partielle Differenzierbarkeit untersucht.

Lemma 3.3.
Unter den Annahmen (A1)-(A4) *erfüllt die Auswahlwahrscheinlichkeit* $\alpha_{\boldsymbol{\theta}}$ *für alle* $\boldsymbol{\theta} \in \Theta$ *die folgenden Eigenschaften:*

(*i*) $\alpha_{\boldsymbol{\theta}} \in (0,1)$,

(*ii*) $\boldsymbol{\theta} \mapsto \alpha_{\boldsymbol{\theta}}$ *ist stetig,*

(*iii*) $\boldsymbol{\theta} \mapsto \alpha_{\boldsymbol{\theta}}$ *ist dreimal stetig partiell differenzierbar.*

Beweis. Aufgrund der Positivität der Dichtefunktion für $0 < t \leq x \leq t + s$ mit $t < G$ gilt $0 < \alpha_{\boldsymbol{\theta}}$, da auch die Parameter s und G per Definition strikt positiv sind. Weiterhin ist das Parallelogramm D für alle $0 < G < \infty$ und $s > 0$ eine echte Teilmenge von $[0,\infty) \times [0,G]$, sodass mit der Normiertheit der Dichte $\alpha_{\boldsymbol{\theta}} < 1$ folgt.

Die gemeinsame Dichtefunktion $f_{\boldsymbol{\theta}}$ aus Gleichung (3.1) ist nun für alle $\boldsymbol{\theta} \in \Theta$ über D integrierbar. Weiter ist die Funktion $\boldsymbol{\theta} \mapsto f_{\boldsymbol{\theta}}(x,t)$ für alle $(x,t)^{\mathrm{T}} \in D$ als Komposition stetiger Funktionen stetig für alle $\boldsymbol{\theta} \in \Theta$. Darüber hinaus lässt sich $f_{\boldsymbol{\theta}}$ durch eine positive über D integrierbare Funktion abschätzen. Für eine hinreichend große Konstante $K > 0$ führt

$$e^{-\theta x} \cdot \left(1 - \frac{t}{G}\right)^{\vartheta\theta x} \leq 1$$

für alle $\boldsymbol{\theta} \in \Theta$ und $(x,t)^{\mathrm{T}} \in D$ zu der Abschätzung

$$|f_{\boldsymbol{\theta}}(x,t)| \leq \frac{\theta}{G} \cdot [(\vartheta\theta x + 1)(1 - \vartheta \log(1 - t/G)) + \vartheta] \leq K \cdot [1 - \log(1 - t/G)].$$

Das Integral dieser Funktion über dem Parallelogramm D hat den Wert

$$\begin{aligned} K \cdot \int_0^G \int_t^{t+s} (1 - \log(1 - t/G))\,\mathrm{d}x\,\mathrm{d}t &= K \cdot s \cdot \int_0^G (1 - \log(1 - t/G))\,\mathrm{d}t \\ &= K \cdot s \cdot [2t - G(1 - t/G)\log(1 - t/G)]_0^G = 2sKG. \end{aligned}$$

Nach dem Satz 5.6 (Kap. IV § 5, [13]) über die stetige Abhängigkeit des Integrals von einem Parameter ist die Abbildung $\boldsymbol{\theta} \mapsto \alpha_{\boldsymbol{\theta}}$ für alle $\boldsymbol{\theta} \in \Theta$ stetig.
Die partiellen Ableitungen von $f_{\boldsymbol{\theta}}$ nach θ und ϑ für $\boldsymbol{\theta} \in \Theta$, $x > 0$ und $0 < t < G$ lauten

$$\frac{\partial f_{\boldsymbol{\theta}}(x,t)}{\partial \theta} = \frac{f_{\boldsymbol{\theta}}(x,t)}{\theta} - \frac{f_{\boldsymbol{\theta}}(x,t)}{R_{\boldsymbol{\theta}}(x,t)} \cdot x(1 - \vartheta \log(1 - t/G))^2(\vartheta\theta x + 1)$$

sowie

$$\begin{aligned} \frac{\partial f_{\boldsymbol{\theta}}(x,t)}{\partial \vartheta} =& f_{\boldsymbol{\theta}}(x,t) \cdot \theta x \log(1 - t/G) \\ &+ \frac{f_{\boldsymbol{\theta}}(x,t)}{R_{\boldsymbol{\theta}}(x,t)} \cdot [\theta x(1 - 2\vartheta \log(1 - t/G)) + 1 - \log(1 - t/G)] \end{aligned}$$

mit

$$R_{\boldsymbol{\theta}}(x,t) := [(\vartheta\theta x + 1)(1 - \vartheta \log(1 - t/G)) + \vartheta]$$

Somit sind auch die partiellen Ableitungen für alle $\boldsymbol{\theta} \in \Theta$ als Komposition stetiger Funktionen stetig. Diese lassen sich ähnlich wie oben durch eine logarithmische Funktion abschätzen, die über D integrierbar ist. Nach dem Satz 5.6 sowie dem Satz 5.7 (Kap. IV § 5, [13]) über die Differentiation unter dem Integralzeichen ist die Abbildung $\boldsymbol{\theta} \mapsto \alpha_{\boldsymbol{\theta}}$ stetig partiell differenzierbar. Insbesondere dürfen die Integration sowie die Differentiation vertauscht werden.
Dies lässt sich auch für die zweiten und dritten Ableitungen überprüfen. Da die Vorgehensweise sehr ähnlich ist, wird an dieser Stelle auf weitere Ausführungen dazu verzichtet. ■

Zur Herleitung der Likelihood wird die Verteilung eines beobachteten Vektors $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$ benötigt.

Definition 3.4.
Die Verteilung von $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$ ergibt sich als Verteilung von $(X_i, T_i)^{\mathrm{T}}$ bedingt auf das Ereignis $(X_i, T_i)^{\mathrm{T}} \in D$ und wird wie folgt bezeichnet

$$\begin{aligned} F_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}(x,t) &:= \mathbb{P}_{\boldsymbol{\theta}}(X_i \le x, T_i \le t |\, 0 < T_i \le X_i \le T_i + s) \\ &= \mathbb{P}_{\boldsymbol{\theta}}((X_i, T_i)^{\mathrm{T}} \in [0,x] \times [0,t] \cap D)/\alpha_{\boldsymbol{\theta}}. \end{aligned}$$

□

Aus der Definition der Verteilungsfunktion lässt sich direkt die Dichtefunktion herleiten.

Korollar 3.5.
Mit

$$\mathbb{P}_{\boldsymbol{\theta}}((X_i, T_i)^{\mathrm{T}} \in [0,x] \times [0,t] \cap D)/\alpha_{\boldsymbol{\theta}} = \int_0^x \int_0^t \mathbb{1}_D(x',t') \frac{f_{\boldsymbol{\theta}}(x',t')}{\alpha_{\boldsymbol{\theta}}} \,\mathrm{d}t' \,\mathrm{d}x'$$

folgt

$$f_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}(x,t) := \frac{\partial^2 F_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}(x,t)}{\partial t \partial x} = \mathbb{1}_D(x,t) \frac{f_{\boldsymbol{\theta}}(x,t)}{\alpha_{\boldsymbol{\theta}}}.$$

Der Verteilungsfunktion $F_{\boldsymbol{\theta}}$ der Vektoren $(X_i, T_i)^{\mathrm{T}}$ liegt die Copula C_ϑ zugrunde. Es stellt sich die Frage, in welchem Zusammenhang die Copula der Verteilungsfunktion $F_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}$ und C_ϑ stehen. Ma und Zhang führen dazu in [37] die *trunkierte Copula* ein. Die Menge der Beobachtungen D wird dafür zunächst mithilfe der Randverteilungen F^X und F^T transformiert zu

$$D_t := \{(u,v) \in [0,1]^2 :\ 0 \le (F^T)^{-1}(v) \le (F^X)^{-1}(u) \le (F^T)^{-1}(v) + s\}.$$

Bezeichnet c_ϑ die Dichte der Copula C_ϑ, dann ist durch

$$c_\vartheta^D(u,v) := \frac{\mathbb{1}_{D_t}(u,v) \cdot c_\vartheta(u,v)}{\int_{D_t} c_\vartheta(u,v) \,\mathrm{d}(u,v)} = \mathbb{1}_{D_t}(u,v) \cdot \frac{c_\vartheta(u,v)}{\alpha_{\boldsymbol{\theta}}}$$

die Dichte der trunkierten Copula C_ϑ^D gegeben. Die Gleichheit $\alpha_{\boldsymbol{\theta}} = \int_{D_t} c_\vartheta(u,v) \,\mathrm{d}(u,v)$ folgt mithilfe der Substitutionsregel aus

$$\begin{aligned} \int_{D_t} c_\vartheta(u,v) \,\mathrm{d}(u,v) &= \int_{\mathbb{R}^2} \mathbb{1}_{D_t}(F^X(x), F^T(t)) \cdot c_\vartheta(F^X(x), F^T(t)) \cdot f^X(x) f^T(t) \,\mathrm{d}(x,t) \\ &= \int_D f_{\boldsymbol{\theta}}(x,t) \,\mathrm{d}(x,t) = \alpha_{\boldsymbol{\theta}}. \end{aligned}$$

Stellt C_ϑ die gemeinsame Verteilungsfunktion zweier gleichverteilter Zufallsvariablen U und V dar, so ist C_ϑ^D die Verteilungsfunktion von (U,V) bedingt auf $(U,V) \in D_t$. Damit lässt sich $F_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}$ mittels

$$F_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}(x,t) = C_\vartheta^D(F^X(x), F^T(t)) \tag{3.2}$$

darstellen. Auch dieser Zusammenhang kann mithilfe der Substitutionsregel hergeleitet werden:

$$\begin{aligned} C_\vartheta^D(F^X(x), F^T(t)) &= \int_0^{F^X(x)} \int_0^{F^T(t)} \mathbb{1}_{D_t}(u,v) \cdot \frac{c_\vartheta(u,v)}{\alpha_{\boldsymbol{\theta}}} \,\mathrm{d}(v,u) \\ &= \int_0^x \int_0^t \mathbb{1}_D(\tilde{x},\tilde{t}) \cdot \frac{f_{\boldsymbol{\theta}}(\tilde{x},\tilde{t})}{\alpha_{\boldsymbol{\theta}}} \,\mathrm{d}(\tilde{t},\tilde{x}) = F^{\tilde{X},\tilde{T}}(x,t) \end{aligned}$$

Dabei wird die Darstellung $f_{\boldsymbol{\theta}}(x,t) = c_\vartheta(F^X(x), F^T(t)) \cdot f^X(x) f^T(t)$ genutzt. Da in Gleichung (3.2) die Randverteilungen F^X und F^T in die Copula C_ϑ^D eingesetzt werden um $F_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}$ zu erhalten, stellt C_ϑ^D nicht die Copula der Verteilung der Beobachtungen dar. Nach Korollar 2.20 erhalten wir diese jedoch unter Stetigkeit von $F^{\tilde{X}}$ und $F^{\tilde{T}}$ über

$$\tilde{C}(u,v) = F^{\tilde{X},\tilde{T}}((F^{\tilde{X}})^{-1}(u), (F^{\tilde{T}})^{-1}(v)) = C_\vartheta^D(F^X((F^{\tilde{X}})^{-1}(u)), F^T((F^{\tilde{T}})^{-1}(u))).$$

Der Zusammenhang zwischen den beiden Copulas wird also durch den Unterschied zwischen den latenten und den beobachteten Randverteilungen bestimmt.

3.2. Approximation der Likelihood

Die Herleitung der Likelihood der trunkierten Stichprobe soll nun ähnlich wie in dem Artikel [65] von Weißbach und Wied mithilfe der Theorie der Punktprozesse aus dem Buch [48] von Reiss erfolgen.
Die trunkierte Stichprobe wird dafür zunächst durch einen trunkierten empirischen Punktprozess

$$N_{n,D}(\cdot) = N_n(\cdot \cap D) = \sum_{i=1}^n \epsilon_{\binom{X_i}{T_i}}(\cdot \cap D)$$

wie in Beispiel 2.2(ii) beschrieben. Die Anzahl der Beobachtungen in der trunkierten Stichprobe beträgt $M_n = N_{n,D}(S)$ mit $S = \mathbb{R}_0^+ \times [0,G]$ und hängt von $\omega \in \Omega$ ab. Nach Beispiel 2.4(ii) berechnet sich das Intensitätsmaß des Prozesses für $B \in \mathcal{B}$ durch

$$\nu_{n,D}(B) = n \cdot \mathbb{P}_{\boldsymbol{\theta}}((X_1,T_1)^{\mathrm{T}} \in B \cap D),$$

wobei jedoch nur Mengen der Form $B = [0,x] \times [0,t]$, $x \in \mathbb{R}_0^+$, $t \in [0,G]$ betrachtet werden. Aufgrund fehlender Stetigkeitseigenschaften ist die Dichtefunktion der Verteilung des trunkierten Prozesses nicht für das weitere Vorgehen geeignet. Stattdessen wird die Dichtefunktion des folgenden Poisson-Prozesses ermittelt.
Es seien $(X_1',T_1')^{\mathrm{T}}$, $(X_2',T_2')^{\mathrm{T}}, \ldots$ unabhängig identisch verteilte Zufallsgrößen mit

$$\mathbb{P}((X_i',T_i')^{\mathrm{T}} \in B) = \nu_{n,D}(B)/\nu_{n,D}(S) = \mathbb{P}_{\boldsymbol{\theta}}((X_1,T_1)^{\mathrm{T}} \in B | \, (X_1,T_1)^{\mathrm{T}} \in D).$$

Weiter sei τ unabhängig von $(X_i', T_i')^{\mathrm{T}}$, $i \in \mathbb{N}$, und Poisson-verteilt mit dem Parameter $\nu_{n,D}(S) = n\alpha_{\boldsymbol{\theta}}$. Nach Satz 2.6 handelt es sich bei dem Prozess

$$N_n^* := \sum_{i=1}^{\tau} \epsilon_{\binom{X_i'}{T_i'}}$$

um einen Poisson-Prozess mit dem Intensitätsmaß $\nu_{n,D}$. Dieser kann nach Satz 2.8 als Näherungsprozess für den trunkierten Prozess $N_{n,D}$ betrachtet werden. Der Hellinger-Abstand der Verteilungen der Prozesse ist kleiner gleich $3^{1/2}\alpha_{\boldsymbol{\theta}}$. Die Herleitung der Likelihood erfolgt gemäß Satz 2.9 für den Prozess N_n^*. Für die Herleitung wird nun ein zweiter Poisson-Prozess benötigt. Es sei

$$N_0 := \sum_{i=1}^{\tau_0} \epsilon_{\binom{X_i^0}{T_i^0}},$$

wobei τ_0, $(X_1^0, T_1^0)^{\mathrm{T}}$, $(X_2^0, T_2^0)^{\mathrm{T}}, \ldots$ unabhängig sind mit $\mathcal{L}(\tau_0) = P_{(G+s)G}$ sowie $\mathcal{L}(X_i^0, T_i^0) = \lambda_R/\lambda_R(S)$ für $R = [0, G+s] \times [0, G]$. Hierbei steht λ_R für das auf R eingeschränkte Lebesgue-Maß $\lambda_R = \lambda(\cdot \cap R)$. Nach Satz 2.6 ist N_0 ein Poisson-Prozess mit dem Intensitätsmaß $\nu_0 = \lambda_R$. Für die Anwendung des Satzes 2.9 wird die ν_0-Dichte $h_{\boldsymbol{\theta}}$ von $\nu_{n,D}$ benötigt. Somit soll

$$\nu_{n,D}([0,x] \times [0,t]) = \int_0^x \int_0^t h_{\boldsymbol{\theta}}(x', t')\, \lambda(\mathrm{d}t')\, \lambda(\mathrm{d}x')$$

für alle $(x,t)^{\mathrm{T}} \in R$ gelten. Für alle $B \subset S$ gilt wegen $D \subset R$ insbesondere $\nu_0(B) = 0 \implies \nu_{n,D}(B) = 0$. Unter der Annahme der Lebesgueschen Integrierbarkeit von $h_{\boldsymbol{\theta}_0}$ folgt aus dem Hauptsatz der Analysis für Lebesgue-Integrale und Beispiel 2.4(ii)

$$\begin{aligned} h_{\boldsymbol{\theta}}(x,t) &= \frac{\partial^2}{\partial x \partial t} \nu_{n,D}([0,x] \times [0,t]) \\ &= n \cdot \frac{\partial^2}{\partial x \partial t} \mathbb{P}_{\boldsymbol{\theta}} \left((X_i, T_i)^{\mathrm{T}} \in [0,x] \times [0,t] \cap D \right). \end{aligned} \tag{3.3}$$

Aus der Definition 3.4 und dem Korollar 3.5 resultiert schließlich

$$h_{\boldsymbol{\theta}}(x,t) = n \cdot \alpha_{\boldsymbol{\theta}} \cdot f_{\boldsymbol{\theta}}^{\tilde{X},\tilde{T}}(x,t) = n \cdot \mathbb{1}_D(x,t) \cdot f_{\boldsymbol{\theta}}(x,t).$$

Nach Satz 2.9 berechnet sich die Likelihood mit $\nu_0(S) = \lambda_R(S) = G(G+s)$ und $\nu_{n,D}(S) = n\alpha_{\boldsymbol{\theta}}$ durch

$$\begin{aligned} \ell(\boldsymbol{\theta}, n) &= \left(\prod_{i=1}^{N_n^*(S)} h_{\boldsymbol{\theta}}(X_i', T_i') \right) \exp\left((G+s)G - n\alpha_{\boldsymbol{\theta}} \right) \\ &= n^{N_n^*(S)} \left(\prod_{i=1}^{N_n^*(S)} f_{\boldsymbol{\theta}}(X_i', T_i') \right) \exp\left((G+s)G - n\alpha_{\boldsymbol{\theta}} \right). \end{aligned}$$

Diese nimmt für $T_i' < G$ nur positive Werte an. Für $T_i' < G$ hat die logarithmierte Likelihood somit die Form

$$\log \ell\left(\boldsymbol{\theta}, n\right) = \sum_{i=1}^{N_n^*(S)} \log n f_{\boldsymbol{\theta}}(X_i', T_i') + (G+s)G - n\alpha_{\boldsymbol{\theta}}.$$

Bei dieser Darstellung kann die Zufallsgröße $N_n^*(S)$ mit positiver, von $\alpha_{\boldsymbol{\theta}}$ abhängigen Wahrscheinlichkeit Werte annehmen, die echt größer als n sind. Diese Wahrscheinlichkeit ist umso geringer, je besser die Approximation des trunkierten Prozesses durch den Poisson-Prozess im Sinne des Hellinger-Abstandes ist. Um diesen Fall dennoch ausschließen zu können, erfolgt ein weiterer Approximationsschritt, in welchem die Summationsgrenze auf $N_{n,D}(S)$ gesetzt und die Zufallsgrößen $(X_i', T_i')^{\mathrm{T}}$ durch die verteilungsgleichen Zufallsvariablen $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$ ersetzt werden. Es folgt $\log \ell\left(\boldsymbol{\theta}, n\right) \approx \log \ell'\left(\boldsymbol{\theta}, n\right)$ mit

$$\begin{aligned} \log \ell'\left(\boldsymbol{\theta}, n\right) &= \sum_{j=1}^{N_{n,D}(S)} \log n f_{\boldsymbol{\theta}}(\widetilde{X}_j, \widetilde{T}_j) + (G+s)G - n\alpha_{\boldsymbol{\theta}} \\ &= \sum_{i=1}^{n} \mathbb{1}_{\{(X_i,T_i)^{\mathrm{T}} \in D\}} \log n f_{\boldsymbol{\theta}}(X_i, T_i) + (G+s)G - n\alpha_{\boldsymbol{\theta}}, \end{aligned} \tag{3.4}$$

wobei auch hier $\widetilde{T}_j < G$ bzw. $T_i < G$ vorausgesetzt wird. Die Parameterschätzung soll durch Maximierung der Loglikelihood erfolgen. Dabei erfolgt zunächst eine Maximierung in n, da diese Stichprobengröße nicht beobachtbar ist und somit ersetzt werden kann. Die ersten beiden Komponenten von

$$\underset{\boldsymbol{\theta}\in\Theta,\, n\in\mathbb{N}}{\arg\max} \log \ell'\left(\boldsymbol{\theta}, n\right). \tag{3.5}$$

bilden dann den Schätzer $\hat{\boldsymbol{\theta}}_n$. Es wird angenommen, dass mindestens eine Beobachtung vorliegt, d.h. $M_n \geq 1$.
Im folgenden Abschnitt werden unter Zuhilfenahme der Likelihood gewisse Kriteriumsfunktionen aufgestellt, welche für asymptotische Betrachtungen des Schätzers im darauffolgenden Kapitel von Nutzen sein werden.

3.3. Konstruktion eines Z-Schätzers

Eine gängige Methode zur Schätzung eines Parameters $\theta \in \Theta$ der Verteilung von Zufallsgrößen $X_1, \ldots, X_n$ ist die Maximierung einer Kriteriumsfunktion

$$\theta \mapsto \mathcal{M}_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} m_\theta(X_i).$$

Hierbei sind $m_\theta : \mathbb{R} \to \overline{\mathbb{R}}$ bekannte Funktionen. Ein Schätzer, der die Funktion $\mathcal{M}_n$ über Θ maximiert, wird als *M-Schätzer* bezeichnet. Die Bestimmung des Maximums erfolgt häufig durch das Nullsetzen der Ableitung der Kriteriumsfunktion. Ein Schätzer,

der eine Gleichung der Form

$$\Psi_n(\theta) := \frac{1}{n}\sum_{i=1}^{n}\psi_\theta(X_i) = 0$$

erfüllt, wird deshalb als *Z-Schätzer* bezeichnet. Auch hierbei ist ψ_θ eine bekannte Funktion. Ist $\boldsymbol{\theta}$ ein k-dimensionaler Parameter, so ist die Funktion $\psi_{\boldsymbol{\theta}} = (\psi_{\boldsymbol{\theta},1}, \ldots, \psi_{\boldsymbol{\theta},k})$ eine vektorwertige Funktion. Maximum-Likelihood-Schätzer stellen somit eine spezielle Form von M- bzw. Z-Schätzern dar. Besitzen die unabhängig identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$ jeweils die Dichtefunktion $h_{\boldsymbol{\theta}}$, so maximiert der Maximum-Likelihood-Schätzer die Funktion $\boldsymbol{\theta} \mapsto \prod_{i=1}^{n} h_{\boldsymbol{\theta}}(X_i)$ oder äquivalent

$$\boldsymbol{\theta} \mapsto \frac{1}{n}\sum_{i=1}^{n}\log h_{\boldsymbol{\theta}}(X_i).$$

Daher ist der Maximum-Likelihood-Schätzer mit $m_{\boldsymbol{\theta}} = \log h_{\boldsymbol{\theta}}$ ein M-Schätzer. Unter Voraussetzung der Existenz der partiellen Ableitungen der Dichten, ist der Schätzer jedoch auch ein Z-Schätzer mit $\psi_{\theta,j} = \partial/\partial\theta_j \log h_{\boldsymbol{\theta}}$ für $j = 1, \ldots, k$. Die vektorwertige Funktion $\psi_{\boldsymbol{\theta}}$ wird dann auch als *Score-Funktion* bezeichnet.

Im Folgenden wird der Z-Schätzer für das vorliegende Modell hergeleitet und für den Beweis der Konsistenz sowie der asymptotischen Normalität verwendet.
Der Übersicht halber wird im Folgenden $M_n = N_{n,D}(S)$ als Bezeichnung genutzt. Durch das Einsetzen der Definition (3.1) der Dichtefunktion $f_{\boldsymbol{\theta}}$ in die Gleichung (3.4) erhält man

$$\begin{aligned}\log \ell'(\boldsymbol{\theta}, n) = {} & M_n \cdot \log(n) + M_n \cdot \log(\theta) - M_n \cdot \log(G) - \theta \sum_{j=1}^{M_n} \widetilde{X}_j \\ & + \vartheta\theta \sum_{j=1}^{M_n} \widetilde{X}_j \log(1 - \widetilde{T}_j/G) + (G+s)G - n \cdot \alpha_{\boldsymbol{\theta}} \\ & + \sum_{j=1}^{M_n} \log[(\vartheta\theta\widetilde{X}_j + 1)(1 - \vartheta \log(1 - \widetilde{T}_j/G)) - \vartheta].\end{aligned}$$

Die für die Score-Funktion benötigten partiellen Ableitungen nach θ und ϑ lauten

$$\begin{aligned}\frac{\partial}{\partial\theta}\log \ell'(\boldsymbol{\theta}, n) = {} & \frac{M_n}{\theta} + \sum_{j=1}^{M_n} \widetilde{X}_j(\vartheta \log(1 - \widetilde{T}_j/G) - 1) - n \cdot \frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial\theta} \\ & + \sum_{j=1}^{M_n} \frac{\vartheta\widetilde{X}_j(\vartheta \log(1 - \widetilde{T}_j/G) - 1)}{(\vartheta\theta\widetilde{X}_j + 1)(\vartheta \log(1 - \widetilde{T}_j/G) - 1) + \vartheta}\end{aligned}$$

sowie

$$\frac{\partial}{\partial\vartheta}\log\ell'(\boldsymbol{\theta},n) = \theta\sum_{j=1}^{M_n}\widetilde{X}_j\log(1-\widetilde{T}_j/G) - n\cdot\frac{\partial\alpha_{\boldsymbol{\theta}}}{\partial\vartheta} + \sum_{j=1}^{M_n}\frac{(2\vartheta\theta\widetilde{X}_j+1)\log(1-\widetilde{T}_j/G)-\theta\widetilde{X}_j+1}{(\vartheta\theta\widetilde{X}_j+1)(\vartheta\log(1-\widetilde{T}_j/G)-1)+\vartheta}.$$

Da in dem vorliegenden Modell nur M_n, aber nicht n beobachtet wird, sind auch die Ableitungen nach θ und ϑ aufgrund ihrer Abhängigkeit von n nicht sichtbar bzw. bekannt. Um die Score-Funktion aufstellen zu können, soll zunächst n ersetzt werden. Per Definition gilt $\alpha_{\boldsymbol{\theta}} = \mathbb{P}_{\boldsymbol{\theta}}((X_i,T_i)^{\mathrm{T}}\in D)$, sodass $\alpha_{\boldsymbol{\theta}} = \mathbb{E}_{\boldsymbol{\theta}}[\mathbb{1}_{\{(X_i,T_i)^{\mathrm{T}}\in D\}}]$. Nach der Momentenmethode (vgl. Abschnitt 5.5 in [51]) werden die Momente der betrachteten Wahrscheinlichkeitsverteilung durch die Stichprobenmomente geschätzt. Wegen $M_n = \sum_{i=1}^{n}\mathbb{1}_{\{(X_i,T_i)^{\mathrm{T}}\in D\}}$ kann die Auswahlwahrscheinlichkeit demnach über M_n/n geschätzt werden. Dieser Zusammenhang kann genutzt werden, um die latente Stichprobengröße über $n = M_n/\alpha_{\boldsymbol{\theta}}$ zu ersetzen. Momentenschätzer sind typischerweise konsistent und asymptotisch normalverteilt. Für dieses konkrete Beispiel werden diese Eigenschaften erst im späteren Verlauf der Arbeit explizit benötigt und nochmal in Abschnitt 6.2.3 aufgegriffen. Man erhält denselben Schätzwert für n, wenn man die Loglikelihood nach n differenziert und gleich null setzt. Dabei wird M_n als Größe betrachtet, die nicht von n beeinflusst wird und beim Differenzieren als Konstante entfällt. Aufgrund späterer Stetigkeitsvoraussetzungen wird der latente Stichprobenumfang als stetige Größe behandelt.
Durch das Ersetzen von n erhält man so die zweidimensionale Score-Funktion $\psi_{\boldsymbol{\theta}} := (\psi_{\boldsymbol{\theta},1},\psi_{\boldsymbol{\theta},2})$ mit

$$\psi_{\boldsymbol{\theta},1}(X_i,T_i) := \mathbb{1}_{[T_i,T_i+s]}(X_i)\left[\frac{1}{\theta}+X_i[\vartheta\log(1-T_i/G)-1] + \frac{\vartheta X_i(\vartheta\log(1-T_i/G)-1)}{(\vartheta\theta X_i+1)(\vartheta\log(1-T_i/G)-1)+\vartheta} - \frac{\frac{\partial\alpha_{\boldsymbol{\theta}}}{\partial\theta}}{\alpha_{\boldsymbol{\theta}}}\right] \tag{3.6a}$$

und

$$\psi_{\boldsymbol{\theta},2}(X_i,T_i) := \mathbb{1}_{[T_i,T_i+s]}(X_i)\left[\theta X_i\log(1-T_i/G) + \frac{(2\vartheta\theta X_i+1)\log(1-T_i/G)-\theta X_i+1}{(\vartheta\theta X_i+1)(\vartheta\log(1-T_i/G)-1)+\vartheta} - \frac{\frac{\partial\alpha_{\boldsymbol{\theta}}}{\partial\vartheta}}{\alpha_{\boldsymbol{\theta}}}\right]. \tag{3.6b}$$

Die partiellen Ableitungen der Auswahlwahrscheinlichkeit werden nicht explizit angegeben, lassen sich jedoch wegen Lemma 3.3 durch das Vertauschen der Differentiation und Integration berechnen. Streng genommen handelt es sich bei $\psi_{\boldsymbol{\theta}}$ nicht um eine Score-Funktion, da zunächst n ersetzt wird. Weiterhin stimmt $\psi_{\boldsymbol{\theta}}$ nicht mit der Ableitung von $\log f_{\boldsymbol{\theta}}$ überein. Es wird sich jedoch herausstellen, dass $\psi_{\boldsymbol{\theta}}$ in einem anderen Modell M^*

(siehe Abschnitt 4.1) die Definition einer Score-Funktion erfüllt, weshalb $\psi_{\boldsymbol{\theta}}$ in dieser Arbeit als Score-Funktion bezeichnet wird. Die Kriteriumsfunktion $\Psi_n := (\Psi_{n,1}, \Psi_{n,2})$ ist gegeben durch

$$\Psi_{n,1}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{i=1}^{n} \psi_{\boldsymbol{\theta},1}(X_i, T_i) \quad \text{und} \quad \Psi_{n,2}(\boldsymbol{\theta}) := \frac{1}{n}\sum_{i=1}^{n} \psi_{\boldsymbol{\theta},2}(X_i, T_i).$$

Der Schätzer $\hat{\boldsymbol{\theta}}_n$ ergibt sich dann als Nullstelle von Ψ_n, falls diese in Θ liegt. Anderenfalls ist es der nächstgelegene Randwert von Θ.

3.4. Weitere Copulas

Bevor die Eigenschaften des Z-Schätzers im nächsten Kapitel untersucht werden, werden zwei weitere Copulas eingeführt und die entsprechenden Likelihoods aufgestellt. Eigenschaften wie Konsistenz und asymptotische Normalität der zugehörigen Schätzer werden nicht untersucht, jedoch werden die verschiedenen Schätzer bei der Anwendung auf die Daten miteinander verglichen.

Farlie-Gumbel-Morgenstern-Copula

Die in Beispiel 2.21 eingeführte Farlie-Gumbel-Morgenstern-Copula (2.5) modelliert eine schwächere Abhängigkeitsstruktur als die in Annahme (A3) verwendete Gumbel-Barnett-Copula. Dies geht nicht nur aus dem aus dem Vergleich der möglichen Werte von Kendalls Tau in den Beispielen 2.26 und 2.27 hervor, sondern auch aus der nachfolgenden Abbildung, welche die Punkteplots der beiden Copulas miteinander vergleicht. Der zugehörige Parameter der Copula kann jedoch sowohl positive als auch negative Werte annehmen, sodass mit dieser auch positive Abhängigkeit modelliert werden kann. Um die Daten zur Unternehmensdemografie im späteren Verlauf der Arbeit auch auf diese Abhängigkeitsstruktur untersuchen zu können, soll in diesem Abschnitt ein Schätzer auf Basis der Farlie-Gumbel-Morgenstern-Copula hergeleitet werden. Da die Herleitungen ähnlich zu dem Vorgehen in den vorangegangenen Abschnitten sind, werden diese kurz gehalten. Darüber hinaus beschränken sich die Beweise im nachfolgenden Kapitel zu den Eigenschaften des Schätzers auf die Gumbel-Barnett-Copula (2.4), da sich in diesem Fall Besonderheiten für den Unabhängigkeitsfall ergeben werden. Für θ, ϑ und n erfolgt zur Verbesserung der Lesbarkeit zunächst keine Kennzeichnung mit dem Superskript FGM. Diese wird erst in Kapitel 7 vorgenommen, um die Abgrenzung zur Gumbel-Barnett-Copula zu betonen.

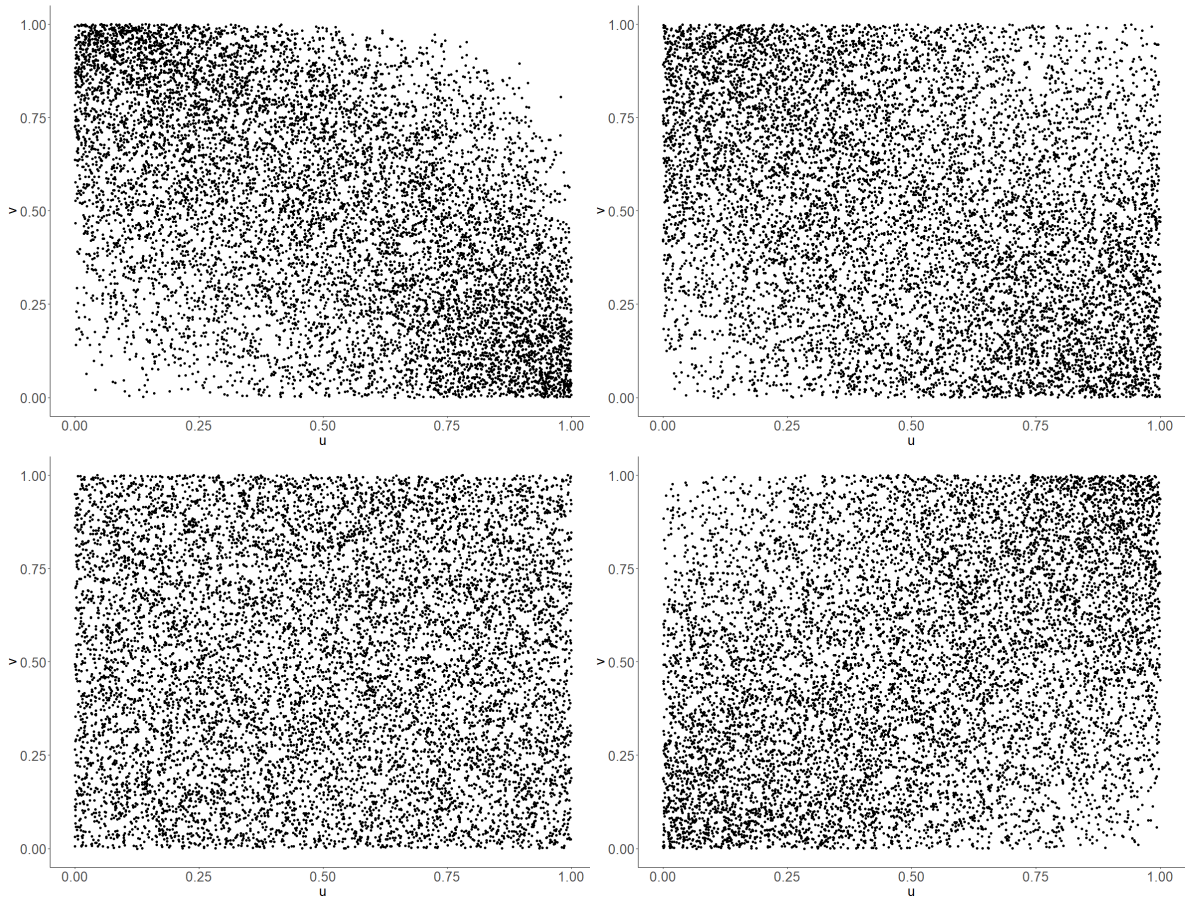

Abbildung 3.2.: Scatterplot von jeweils 10.000 Realisationen der Copula (2.4) für $\vartheta = 1$ (links oben) und von der Copula 2.5 für die Parameterwerte $\vartheta = -1$, $\vartheta = 0$ und $\vartheta = 1$

Die vorangegangene Abbildung zeigt die Abhängigkeitsstruktur der Farlie-Gumbel-Morgenstern-Copula für die Parameterwerte $\vartheta = -1$, $\vartheta = 0$ und $\vartheta = 1$. Mithilfe der bedingten Inversionsmethode 2.23 wurden für $c_u^{FGM}(v) = v + \vartheta(1-2u)v - \vartheta(1-2u)v^2$ jeweils 10.000 Realisationen erzeugt. Man erhält für $\vartheta = -1$ eine negative Abhängigkeit und für $\vartheta = 1$ eine positive Abhängigkeit. Mit $\vartheta = 0$ erhält man auch für diese Copula Unabhängigkeit. Diese Erkenntnisse konnte man in Beispiel 2.27 bereits aus Kendalls Tau $\tau_\vartheta^{FGM} = 2\vartheta/9$ gewinnen. Links oben ist die Abhängigkeitsstruktur der Gumbel-Barnett-Copula mit $\vartheta = 1$ dargestellt. Im direkten Vergleich wird deutlich, dass mit der Farlie-Gumbel-Morgenstern-Copula eine schwächere Abhängigkeit zugrunde liegt.

Es gelten nun die Annahmen aus Abschnitt 3.1. Die Annahme (A3) wird durch die folgende Annahme ersetzt:

(A3') Die Abhängigkeit zwischen X_i und T_i wird mithilfe der Copula (2.5)

$$C_\vartheta^{FGM}(u,v) = u \cdot v \cdot [1 + \vartheta(1-u)(1-v)], \quad \vartheta \in [-1,1]$$

modelliert. Auch in diesem Fall erhält man mithilfe des Satzes von Sklar 2.17 die

gemeinsame Verteilungsfunktion durch das Einsetzen der Randverteilungsfunktionen in die Copula. Dies impliziert

$$F_{\boldsymbol{\theta}}^{FGM}(x,t) = \begin{cases} \frac{t}{G} \cdot (1 - e^{-\theta x}) \left[1 + \vartheta \cdot e^{-\theta x} \cdot \left(1 - \frac{t}{G}\right)\right], & x > 0,\ 0 < t < G, \\ 1 - e^{-\theta x}, & x > 0,\ t \geq G, \\ 0, & \text{sonst.} \end{cases}$$

Die gemeinsame Dichtefunktion lautet

$$f_{\boldsymbol{\theta}}^{FGM}(x,t) = \frac{\theta}{G} e^{-\theta x} \left[1 + \vartheta \cdot (2e^{-\theta x} - 1) \cdot \left(1 - \frac{2t}{G}\right)\right]$$

für $x > 0$ und $0 < t < G$. Für alle weiteren Werte von x und t nimmt $f_{\boldsymbol{\theta}}^{FGM}$ den Wert null an.

Die Auswahlwahrscheinlichkeit, d.h. die Wahrscheinlichkeit dafür, dass das i-te Individuum beobachtet wird, lässt sich in diesem Fall explizit berechnen:

$$\begin{aligned} \alpha_{\boldsymbol{\theta}}^{FGM} &= \int_0^G \int_t^{t+s} \frac{\theta}{G} e^{-\theta x} \left[1 + \vartheta(2e^{-\theta x} - 1)\left(1 - \frac{2t}{G}\right)\right] \mathrm{d}x\, \mathrm{d}t \\ &= \int_0^G -\frac{1}{G} e^{-\theta t}(e^{-\theta s} - 1)\, \mathrm{d}t + \int_0^G \left(1 - \frac{2t}{G}\right) \frac{\vartheta}{G} [e^{-\theta t}(e^{-\theta s} - 1) - e^{-2\theta t}(e^{-2\theta s} - 1)]\, \mathrm{d}t \\ &= \frac{1}{\theta G}(1 - e^{-\theta s})(1 - e^{-\theta G}) - \frac{\vartheta}{G}\left[-\frac{1}{\theta}(e^{-\theta s} - 1)(e^{-\theta G} + 1) + \frac{1}{2\theta}(e^{-2\theta s} - 1)(e^{-2\theta G} + 1)\right] \\ &\quad + \frac{2\vartheta}{G^2}\left[\frac{1}{\theta^2}(1 - e^{-\theta s})(1 - e^{-\theta G}) - \frac{1}{4\theta^2}(1 - e^{-2\theta s})(1 - e^{-2\theta G})\right] \end{aligned}$$

Mit der gleichen Vorgehensweise wie in Abschnitt 3.2 lässt sich damit nun die Loglikelihood aufstellen. Diese lässt sich als Summe der latenten Stichprobe

$$\log \ell^{FGM}(\boldsymbol{\theta}, n) = \sum_{i=1}^{n} \mathbb{1}_{\{(X_i,T_i)^\intercal \in D\}} \log n f_{\boldsymbol{\theta}}^{FGM}(X_i, T_i) + (G + s)G - n\alpha_{\boldsymbol{\theta}}^{FGM}$$

oder als Summe der beobachteten Daten schreiben;

$$\begin{aligned} \log \ell^{FGM}(\boldsymbol{\theta}, n) = \sum_{j=1}^{N_{n,D}^{FGM}(S)} &\log n + \log \frac{\theta}{G} - \theta \widetilde{X_j} + \log \left[1 + \vartheta \left(2e^{-\theta \widetilde{X}_j} - 1\right)\left(1 - \frac{2\widetilde{T}_j}{G}\right)\right] \\ &+ (G + s)G - n\alpha_{\boldsymbol{\theta}}^{FGM} \end{aligned}$$

Die Schätzer der Parameter θ und ϑ erhält man als die ersten beiden Komponenten von

$$\underset{\boldsymbol{\theta} \in \Theta,\, n \in \mathbb{N}}{\arg\max} \log \ell^{FGM}(\boldsymbol{\theta}, n)\,.$$

Zur Spiegelung assozierte Gumbel-Barnett-Copula

Auch mit der zur Spiegelung um die Achse $u = 1/2$ assoziierten Version der Gumbel-Barnett-Copula (2.8) lässt sich eine schwache positive Abhängigkeit modellieren. Im Vergleich zur Farlie-Gumbel-Morgenstern-Copula ist diese nicht symmetrisch um $v = u$.

(A3'') Die Abhängigkeit zwischen X_i und T_i wird mithilfe der Copula (2.8)

$$C_\vartheta^{270}(u,v) = u - u(1-v)e^{-\vartheta \log(u)\log(1-v)}, \quad \vartheta \in [0, 1-\varepsilon_\vartheta]$$

modelliert. Die gemeinsame Verteilungsfunktion lautet dann

$$F_{\boldsymbol{\theta}}^{270}(x,t) = (1-e^{-\theta x}) - (1-e^{-\theta x})\left(1-\frac{t}{G}\right)e^{-\vartheta \log(1-e^{-\theta x})\log\left(1-\frac{t}{G}\right)}$$

für $x > 0$, $0 < t < G$ und $F_{\boldsymbol{\theta}}^{270}(x,t) = 1 - e^{-\theta x}$ für $x > 0$, $t \geq G$. Die gemeinsame Dichtefunktion steht in folgendem Zusammenhang zur Dichte $f_{\boldsymbol{\theta}}$ aus (3.1)

$$f_{\boldsymbol{\theta}}^{270}(x,t) = f_{\boldsymbol{\theta}}\left(\frac{-\log(1-e^{-\theta x})}{\theta}, t\right) \cdot \left(\frac{e^{-\theta x}}{1-e^{-\theta x}}\right).$$

Die Auswahlwahrscheinlichkeit lässt sich nicht explizit, aber mithilfe der Dichte $f_{\boldsymbol{\theta}}$ berechnen

$$\alpha_{\boldsymbol{\theta}}^{270} = \int_0^G \int_t^{t+s} f_{\boldsymbol{\theta}}^{270}(x,t)\,\mathrm{d}x\,\mathrm{d}t = \int_0^G \int_{-\log(1-e^{-\theta(t+s)})/\theta}^{-\log(1-e^{-\theta t})/\theta} f_{\boldsymbol{\theta}}(x,t)\,\mathrm{d}x\,\mathrm{d}t.$$

Die gesuchten Parameter erhält man als Maximalstelle der Loglikelihood, die über die Punktprozesstheorie wie in Abschnitt 3.2 hergeleitet werden kann.

$$\log \ell^{270}(\boldsymbol{\theta}, n) = \sum_{i=1}^{n} \mathbb{1}_{\{(X_i,T_i)^\mathrm{T} \in D\}} \log n f_{\boldsymbol{\theta}}^{270}(X_i,T_i) + (G+s)G - n\alpha_{\boldsymbol{\theta}}^{270}$$

Im folgenden Kapitel werden die Modelleigenschaften auch für diese Copula nicht explizit untersucht. Der aufgrund der Spiegelung bestehende Zusammenhang zwischen den Dichten $f_{\boldsymbol{\theta}}$ und $f_{\boldsymbol{\theta}}^{270}$ kann jedoch genutzt werden, um die folgenden Beweise entsprechend zu übertragen.

4. Modelleigenschaften

Gegenstand dieses Kapitels ist der Nachweis der Konsistenz und der asymptotischen Normalität des Z-Schätzers. Die Identifizierbarkeit der Parameter stellt eine wesentliche Voraussetzung für die Konsistenz dar. In diesem Zusammenhang wird die in dieser Arbeit angewandte, von den gängigen Methoden abweichende Reihenfolge von Stichprobenziehung und Trunkation betont. Weiterhin wird die Invertierbarkeit der Fisher-Information geprüft, welche sowohl in den Konsistenznachweis als auch insbesondere in die Berechnung der asymptotischen Kovarianz eingeht.

4.1. Identifizierbarkeit

Wie bereits in Satz 2.13 dargelegt, lässt sich die Identifizierbarkeit eines Parameters auch mithilfe des Kullback-Leibler-Abstands charakterisieren. Im anschließenden Konsistenzbeweis wird jedoch nicht die Identifizierbarkeit des Parameters des in Abschnitt 3.1 eingeführten Gumbel-Barnett-Modells benötigt, sondern die eines weiteren Modells M^*.

Anders als in der Literatur üblich, wird im Gumbel-Barnett-Modell zunächst eine einfache Zufallsstichprobe gezogen, aus welcher anschließend Datenpunkte außerhalb der Menge D trunkiert werden. In Abbildung 4.1 ist dieser Vorgang den Pfeilen im Uhrzeigersinn folgend beschrieben.

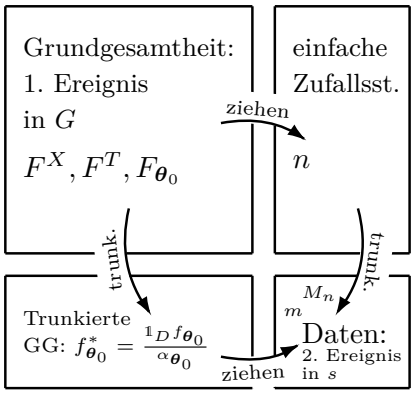

Abbildung 4.1.: Grundgesamtheit mit den Verteilungen von X_i, T_i und $(X_i, T_i)^{\mathrm{T}}$ (oben links), einfache Zufallsstichprobe mit der Stichprobengröße (oben rechts), auf D trunkierte Grundgesamtheit mit der Verteilung von $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$) (unten links), Daten als trunkierte Stichprobe oder als Stichprobe der trunkierten Grundgesamtheit (unten rechts)

Eine ähnliche Vorgehensweise findet sich beispielsweise in dem Aufsatz [24] von Heckman. Den Pfeilen gegen den Uhrzeigersinn folgend wird zuerst der Vorgang der Trunkation auf die Grundgesamtheit angewendet und anschließend daraus eine einfache Zufallsstichprobe gezogen, im Folgenden als M^* bezeichnet. Beispiele für dieses Vorgehen lassen sich in Emura und Pan [15], Moreira, Uña-Álvarez und Braekers [43] sowie Andersen et el. [2] finden.
Bei den beiden Modellansätzen liegen unterschiedliche Unabhängigkeitsannahmen vor, da die Unabhängigkeit jeweils zwischen den einzelnen Ziehungen angenommen wird. Unabhängigkeit zwischen zufällig gezogenen Einheiten aus der Grundgesamtheit scheint dabei plausibler als Unabhängigkeit zwischen zufälligen Ziehungen aus der Menge D. Der Modellansatz gegen den Uhrzeigersinn, M^*, ist dennoch für den Nachweis der Modelleigenschaften in diesem Kapitel hilfreich, da sich die Likelihood als Produkt der (bedingten) Dichten schreiben lässt.

In M^* werde eine einfache Zufallsstichprobe $(X^*, T^*)^{\mathrm{T}}$ gezogen, wobei die einzelnen Zufallsvektoren $(X_i^*, T_i^*)^{\mathrm{T}} \in D$ die gemeinsame Dichte

$$f_{\boldsymbol{\theta}}^*(x,t) := \mathbb{1}_D(x,t)\frac{\theta}{\alpha_{\boldsymbol{\theta}} G} e^{-\theta x}\left(1-\frac{t}{G}\right)^{\vartheta\theta x}\left[(\vartheta\theta x+1)(1-\vartheta\log(1-t/G))+\vartheta\right] \tag{4.1}$$

besitzen. Hierbei sei $\boldsymbol{\theta} \in \Theta$ wie in Annahme (A1). Die Parameter G und s sowie die Menge D seien wie in Abschnitt 3.1 definiert. Vergleicht man $f_{\boldsymbol{\theta}}^*$ mit der Dichte $f_{\boldsymbol{\theta}}$ aus Gleichung (3.1), so erhält man den Zusammenhang $f_{\boldsymbol{\theta}}^*(x,t) = \mathbb{1}_D(x,t) f_{\boldsymbol{\theta}}(x,t)/\alpha_{\boldsymbol{\theta}}$, sodass $f_{\boldsymbol{\theta}}^*$ mit der Dichte der beobachteten Vektoren $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$ übereinstimmt. Für zwei beliebige Parameterwerte $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2 \in \Theta = [\varepsilon_\theta, 1/\varepsilon_\theta] \times [0, 1-\varepsilon_\vartheta]$ für ε_ϑ, $\varepsilon_\theta \in (0,1)$ klein, gilt nach Satz 2.12, dass der Kullback-Leibler-Abstand

$$KL(f_{\boldsymbol{\theta}_1}^* | f_{\boldsymbol{\theta}_2}^*) = \mathbb{E}_{\boldsymbol{\theta}_2}^*\left[\log\frac{f_{\boldsymbol{\theta}_2}^*(X_1^*, T_1^*)}{f_{\boldsymbol{\theta}_1}^*(X_1^*, T_1^*)}\right] := \int_S \log\frac{f_{\boldsymbol{\theta}_2}^*(x,t)}{f_{\boldsymbol{\theta}_1}^*(x,t)} \cdot f_{\boldsymbol{\theta}_2}^*(x,t)\,\mathrm{d}(x,t)$$

genau dann null ist, wenn $f_{\boldsymbol{\theta}_1}^* = f_{\boldsymbol{\theta}_2}^*$ beziehungsweise

$$\frac{f_{\boldsymbol{\theta}_2}(x,t)}{f_{\boldsymbol{\theta}_1}(x,t)} = \frac{\alpha_{\boldsymbol{\theta}_1}}{\alpha_{\boldsymbol{\theta}_2}}$$

für alle $(x,t)^{\mathrm{T}} \in D$ erfüllt ist. Durch das Einsetzen der Definition (3.1) der gemeinsamen Dichtefunktion erhält man

$$\frac{\theta_2}{\theta_1} \cdot e^{-(\theta_2-\theta_1)x} \cdot \left(1-\frac{t}{G}\right)^{(\theta_2\vartheta_2-\theta_1\vartheta_1)x} \cdot \frac{(\vartheta_2\theta_2 x+1)\left(\vartheta_2\log\left(1-\frac{t}{G}\right)-1\right)+\vartheta_2}{(\vartheta_1\theta_1 x+1)\left(\vartheta_1\log\left(1-\frac{t}{G}\right)-1\right)+\vartheta_1} = \frac{\alpha_{\boldsymbol{\theta}_1}}{\alpha_{\boldsymbol{\theta}_2}}.$$

Leitet man beide Seiten nach t ab und setzt $R_{\boldsymbol{\theta}}(x,t) = (\vartheta\theta x+1)(\vartheta\log(1-t/G)-1)+\vartheta$,

so erhält man $F_1 \cdot F_2 = 0$, wobei

$$F_1 = -\frac{1}{G}\frac{\theta_2}{\theta_1} \cdot e^{-(\theta_2-\theta_1)x} \cdot \left(1 - \frac{t}{G}\right)^{(\theta_2\vartheta_2-\theta_1\vartheta_1)x-1} \cdot \frac{1}{R^2_{\boldsymbol{\theta}_1}(x,t)}$$

sowie

$$F_2 = x(\theta_2\vartheta_2 - \theta_1\vartheta_1)R_{\boldsymbol{\theta}_1}(x,t)R_{\boldsymbol{\theta}_2}(x,t) + (\vartheta_2\theta_2 x + 1)\vartheta_2 R_{\boldsymbol{\theta}_1}(x,t) - (\vartheta_1\theta_1 x + 1)\vartheta_1 R_{\boldsymbol{\theta}_2}(x,t).$$

Der erste Faktor nimmt nur für $t = G$ den Wert null an. Setzt man den zweiten Faktor gleich null, so erhält man

$$R_{\boldsymbol{\theta}_2}(x,t)\vartheta_1 \cdot [R_{\boldsymbol{\theta}_1}(x,t)\theta_1 x + (\vartheta_1\theta_1 x + 1)] = R_{\boldsymbol{\theta}_1}(x,t)\vartheta_2 \cdot [R_{\boldsymbol{\theta}_2}(x,t)\theta_2 x + (\vartheta_2\theta_2 x + 1)].$$

Der Parameter t taucht nur über den Term $\log(1-t/G)$ auf. Setzt man $y = \log(1-t/G)$, so stellt man zwei bivariate Polynome gleichen Grades gegenüber. Diese sind genau dann gleich, wenn ihre Koeffizienten übereinstimmen. Ein Vergleich der Koeffizienten der absoluten Glieder liefert

$$-\vartheta_1 + \vartheta_2\vartheta_1 = -\vartheta_2 + \vartheta_2\vartheta_1 \qquad \Longleftrightarrow \qquad \vartheta_1 = \vartheta_2.$$

Setzt man $\vartheta_1 = \vartheta_2$ und stellt die Koeffizienten der Variablen x gegenüber, so gelangt man zu

$$\theta_1(\vartheta_2 - 2\vartheta_2^2 + 2\vartheta_2^3) = \theta_2(\vartheta_2 - 2\vartheta_2^2 + 2\vartheta_2^3)$$

und kann daraus für $\vartheta_2 \neq 0$ auf $\theta_1 = \theta_2$ schließen. Im Falle der Unabhängigkeit, d.h. für $\vartheta_1 = \vartheta_2 = 0$, erhielte man

$$\frac{f_{\boldsymbol{\theta}_2}(x,t)}{f_{\boldsymbol{\theta}_1}(x,t)} = \frac{\theta_2}{\theta_1} \cdot e^{-(\theta_2-\theta_1)x} = \frac{\alpha_{\boldsymbol{\theta}_1}}{\alpha_{\boldsymbol{\theta}_2}}$$

und schließe daraus nach dem Differenzieren beider Seiten nach x auf $\theta_1 = \theta_2$, da $\theta_1, \theta_2 > 0$.

Als Ergebnis erhält man die Äquivalenz von $KL(f^*_{\boldsymbol{\theta}_1}|f^*_{\boldsymbol{\theta}_2}) = 0$ und $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ und somit nach Satz 2.13 die Identifizierbarkeit des Parameters $\boldsymbol{\theta} \in \Theta$.

Satz 4.1.
Im Modell M^* ist der Parameter $\boldsymbol{\theta} \in \Theta$ identifizierbar.

Bemerkung 4.2.
Auch in Abschnitt 4.2 wird das Modell M^* zum Nachweis spezieller Eigenschaften der Fisher-Information genutzt. Die Identifizierbarkeit wird im zweiten Teil des Konsistenzbeweises in Abschnitt 4.3 eingehen. Ein Zusammenhang zwischen den Schätzern, die aus den beiden Modellansätzen in Abbildung 4.1 resultieren, wird im Anhang A.1 aufgegriffen.

□

4.2. Fisher-Information

Mithilfe der Fisher-Information lässt sich beurteilen, wie hoch der Informationsgehalt einer Menge an Daten über den gesuchten Parameter ist. In Kapitel 2.3 in dem Buch [51] von Schervish ist diese für $\boldsymbol{\theta}_0 \in \Theta$ definiert als

$$\mathcal{I}(\boldsymbol{\theta}_0) := \mathrm{Cov}_{\boldsymbol{\theta}_0}\left[\psi_{\boldsymbol{\theta}_0}(X_1, T_1)\right],$$

wobei es sich in der Definition bei $\psi_{\boldsymbol{\theta}}$ um die Score-Funktion, d.h. die Ableitung der logarithmierten Dichte handelt. Wie aus der Gleichung (4.3) und dem Anhang A.1 hervorgeht, handelt es sich bei $\psi_{\boldsymbol{\theta}}$ um die Score-Funktion im Modell M^*, weshalb $\mathcal{I}$ in dieser Arbeit als Fisher-Information bezeichnet wird, auch wenn diese nicht im Kontext von M^* verwendet wird.
Es bezeichne $\Psi(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}}(X_1, T_1)]$. Das folgende Lemma thematisiert die Unverzerrtheit der Score-Funktion $\Psi(\boldsymbol{\theta}_0) = \mathbf{0}$, welche zu einer Vereinfachung der Definition der Fisher-Information zu $\mathcal{I}(\boldsymbol{\theta}_0) = \mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0}(X_1, T_1)\psi_{\boldsymbol{\theta}_0}(X_1, T_1)^{\mathrm{T}}]$ führt.

Lemma 4.3.
Unter den Annahmen (A1)-(A4) *mit* $\boldsymbol{\theta}_0 \in \Theta$ *gilt*

$$\|\Psi(\boldsymbol{\theta}_0)\| = 0.$$

Beweis. Die Aussage kann komponentenweise überprüft werden. In Hinblick auf die Gleichung (3.6a) sei dazu $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0,1}(X_1, T_1)] = s_1 + s_2 - s_3$ mit

$$s_1 := \mathbb{E}_{\boldsymbol{\theta}_0}[\mathbb{1}_{[T_1, T_1+s]}(X_1){\theta_0}^{-1}] = \frac{\alpha_{\boldsymbol{\theta}_0}}{\theta_0},$$

$$s_2 := \mathbb{E}_{\boldsymbol{\theta}_0}\Bigg[\mathbb{1}_{[T_1, T_1+s]}(X_1)\Big(X_1(\vartheta_0 \log(1 - T_1/G) - 1) + \frac{\vartheta_0 X_1(\vartheta_0 \log(1 - T_1/G) - 1)}{(\vartheta_0 \theta X_1 + 1)(\vartheta_0 \log(1 - T_1/G) - 1) + \vartheta_0}\Big)\Bigg]$$

sowie

$$s_3 := \mathbb{E}_{\boldsymbol{\theta}_0}\left[\mathbb{1}_{[T_1, T_1+s]}(X_1)\frac{\frac{\partial \alpha_{\boldsymbol{\theta}_0}}{\partial \theta_0}}{\alpha_{\boldsymbol{\theta}_0}}\right] = \frac{\partial \alpha_{\boldsymbol{\theta}_0}}{\partial \theta_0}.$$

Die einzelnen Erwartungswerte berechnen sich nun durch

$$s_1 = \int_D -\frac{1}{G} e^{-\theta_0 x}\left(1 - \frac{t}{G}\right)^{\vartheta_0 \theta_0 x} \cdot \cdot \Big[(\vartheta_0 \theta_0 x + 1)(\vartheta_0 \log(1 - t/G) - 1) + \vartheta_0\Big]\,\mathrm{d}(x, t),$$

$$s_2 = \int_D -x(\vartheta_0 \log(1-t/G)-1)\frac{\theta_0}{G}e^{-\theta_0 x}\left(1-\frac{t}{G}\right)^{\vartheta_0\theta_0 x} \cdot \\ \cdot \Big[(\vartheta_0\theta_0 x+1)(\vartheta_0\log(1-t/G)-1)+2\vartheta_0\Big]\,\mathrm{d}(x,t)$$

und

$$\begin{aligned} s_3 &= \int_D -\frac{1}{G}e^{-\theta_0 x}\left(1-\frac{t}{G}\right)^{\vartheta_0\theta_0 x}\Big\{\theta_0\vartheta_0 x(\vartheta_0\log(1-t/G)-1)+ \\ &\quad + \Big[1+\theta_0 x(\vartheta_0\log(1-t/G)-1)\Big]\Big[(\vartheta_0\theta_0 x+1)(\vartheta_0\log(1-t/G)-1)+\vartheta_0\Big]\Big\}\,\mathrm{d}(x,t) \\ &= \int_D -\frac{1}{G}e^{-\theta_0 x}\left(1-\frac{t}{G}\right)^{\vartheta_0\theta_0 x}\cdot\Big\{\theta_0\vartheta_0 x(\vartheta_0\log(1-t/G)-1)+\vartheta_0+ \\ &\quad + (\vartheta_0\theta_0 x+1)(\vartheta_0\log(1-t/G)-1)+\vartheta_0\theta_0 x(\vartheta_0\log(1-t/G)-1)+ \\ &\quad + \theta_0 x(\vartheta_0\theta_0 x+1)(\vartheta_0\log(1-t/G)-1)^2\Big\}\,\mathrm{d}(x,t), \end{aligned}$$

wobei für s_3 die Dichtefunktion $f_{\boldsymbol{\theta}_0}$ aus Gleichung 3.1 nach θ_0 abgeleitet wurde. Als Differenz aus s_1 und s_3 erhält man

$$\begin{aligned} s_1 - s_3 = \int_D \frac{1}{G}e^{-\theta_0 x}\left(1-\frac{t}{G}\right)^{\vartheta_0\theta_0 x}\cdot\Big\{&\theta_0\vartheta_0 x(\vartheta_0\log(1-t/G)-1)+ \\ &+\vartheta_0\theta_0 x(\vartheta_0\log(1-t/G)-1)+\theta_0 x(\vartheta_0\theta_0 x+1)(\vartheta_0\log(1-t/G)-1)^2\Big\}\,\mathrm{d}(x,t). \end{aligned}$$

Addiert man dazu nun s_2, zeigt sich, dass $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0,1}(X_1,T_1)]=0$. Für die zweite Komponente aus Gleichung (3.6b) sei $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0,2}(X_1,T_1)] = \tilde{s}_1+\tilde{s}_2-\tilde{s}_3$ mit

$$\tilde{s}_1 = \mathbb{E}_{\boldsymbol{\theta}_0}[\mathbb{1}_{[T_1,T_1+s]}(X_1)\cdot\theta_0 X_1\log(1-T_1/G)],$$

$$\tilde{s}_2 = \mathbb{E}_{\boldsymbol{\theta}_0}\left[\mathbb{1}_{[T_1,T_1+s]}(X_1)\cdot\frac{(2\vartheta\theta X_i+1)\log(1-T_i/G)-\theta X_i+1}{(\vartheta\theta X_i+1)(\vartheta\log(1-T_i/G)-1)+\vartheta}\right]$$

sowie

$$\tilde{s}_3 = \mathbb{E}_{\boldsymbol{\theta}_0}\left[\mathbb{1}_{[T_1,T_1+s]}(X_1)\frac{\frac{\partial\alpha_{\boldsymbol{\theta}_0}}{\partial\vartheta_0}}{\alpha_{\boldsymbol{\theta}_0}}\right] = \frac{\partial\alpha_{\boldsymbol{\theta}_0}}{\partial\vartheta_0}.$$

Diese berechnen sich nun durch

$$\tilde{s}_1 = \int_D -\frac{\theta_0^2}{G}x\log(1-t/G)e^{-\theta_0 x}\left(1-\frac{t}{G}\right)^{\vartheta_0\theta_0 x}\cdot \\ \cdot\Big[(\vartheta_0\theta_0 x+1)(\vartheta_0\log(1-t/G)-1)+\vartheta_0\Big]\,\mathrm{d}(x,t),$$

$$\tilde{s}_2 = \int_D -\frac{\theta_0}{G}e^{-\theta_0 x}\left(1-\frac{t}{G}\right)^{\vartheta_0\theta_0 x}\cdot \\ \cdot\Big[(2\vartheta_0\theta_0 x+1)\log(1-t/G)-\theta_0 x+1\Big]\,\mathrm{d}(x,t)$$

und

$$\begin{aligned}\tilde{s}_3 = \int_D -\frac{\theta_0}{G} e^{-\theta_0 x} \left(1 - \frac{t}{G}\right)^{\vartheta_0\theta_0 x} \cdot \Big\{ & \vartheta_0\theta_0 x \log(1 - t/G) + 1 + \\ & + \theta_0 x(\vartheta_0 \log(1 - t/G) - 1) + (\vartheta_0\theta_0 x + 1)\log(1 - t/G) + \\ & + \theta_0 x \log(1 - t/G)(\theta_0\vartheta_0 x + 1)(\vartheta_0 \log(1 - t/G) - 1) \Big\} \, \mathrm{d}(x, t),\end{aligned}$$

wobei sich $\tilde{s}_3$ durch das Ableiten der Dichtefunktion $f_{\boldsymbol{\theta}_0}$ aus Gleichung 3.1 nach ϑ_0 ergibt. Bildet man die Differenz aus $\tilde{s}_2$ und $\tilde{s}_3$, so erhält man

$$\begin{aligned}\tilde{s}_2 - \tilde{s}_3 = \int_D -\frac{\theta_0}{G} e^{-\theta_0 x} \left(1 - \frac{t}{G}\right)^{\vartheta_0\theta_0 x} \cdot \Big\{ & \vartheta_0\theta_0 x \log(1 - t/G) + \\ & + \theta_0 x \log(1 - t/G)(\theta_0\vartheta_0 x + 1)(\vartheta_0 \log(1 - t/G) - 1) \Big\} \, \mathrm{d}(x, t).\end{aligned}$$

Die Addition von $\tilde{s}_1$ führt zu $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0,2}(X_1, T_1)] = 0$ und somit schließlich zu $\|\Psi(\boldsymbol{\theta}_0)\| = 0$. ■

Ergibt sich die Score-Funktion als Ableitung der logarithmierten Dichtefunktion nach dem gesuchten Parameter, dann beschreibt diese die Änderung der Dichtefunktion in Abhängigkeit des Parameters. Man erhält aus einer großen lokalen Änderung viel Information über den Parameter und einen hohen Wert der Fisher-Information (vgl. Rüschendorf [50]). Intuitiv erwartet man genauere Schätzungen, wenn die Daten mehr Informationen bereitstellen. Dies wird sich im Zusammenhang mit der asymptotischen Normalität in 4.13 bestätigen, da sich herausstellen wird, dass die asymptotische Varianz mit der inversen Fisher-Information übereinstimmt. Dafür wird jedoch eine weitere Darstellungsweise der Fisher-Information benötigt, die in diesem Abschnitt hergeleitet wird.

In diesem sowie in den weiteren Abschnitten wird die folgende Bezeichnung für die Matrix der ersten Ableitungen der Funktion $\psi_{\boldsymbol{\theta}} = (\psi_{\boldsymbol{\theta},1}, \psi_{\boldsymbol{\theta},2})^{\mathrm{T}}$ verwendet:

$$\dot{\psi}_{\boldsymbol{\theta}_0}(x, t) := \begin{pmatrix} \frac{\partial}{\partial\theta}\psi_{\boldsymbol{\theta}_0,1}(x, t) & \frac{\partial}{\partial\theta}\psi_{\boldsymbol{\theta}_0,2}(x, t) \\ \frac{\partial}{\partial\vartheta}\psi_{\boldsymbol{\theta}_0,1}(x, t) & \frac{\partial}{\partial\vartheta}\psi_{\boldsymbol{\theta}_0,2}(x, t) \end{pmatrix}$$

Unter gewissen Regularitätsbedingungen gilt weiterhin die Gleichheit

$$\mathcal{I}(\boldsymbol{\theta}_0) = -\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1, T_1)],$$

welche auch als *Information Matrix Equality* (IME) bekannt ist. Diese lässt sich für das Modell der abhängigen Trunkation aus Kapitel 3.1 aus der IME für das in Kapitel 4.1 eingeführte Modell M^* herleiten. Wegen $f^*_{\boldsymbol{\theta}}(x, t) = \mathbb{1}_D(x, t) f_{\boldsymbol{\theta}}(x, t)/\alpha_{\boldsymbol{\theta}}$ gelten

$$\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\psi_{\boldsymbol{\theta}_0}(X_1^*, T_1^*)\psi_{\boldsymbol{\theta}_0}(X_1^*, T_1^*)^{\mathrm{T}}\right] = \frac{1}{\alpha_{\boldsymbol{\theta}_0}} \cdot \mathbb{E}_{\boldsymbol{\theta}_0}\left[\psi_{\boldsymbol{\theta}_0}(X_1, T_1)\psi_{\boldsymbol{\theta}_0}(X_1, T_1)^{\mathrm{T}}\right]$$

sowie

$$-\mathbb{E}^*_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1^*, T_1^*)] = -\frac{1}{\alpha_{\boldsymbol{\theta}_0}} \cdot \mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1, T_1)].$$

Der Übersicht halber bezeichne $f^*_{\boldsymbol{\theta}_0} = f^*_{\boldsymbol{\theta}_0}(X_1^*, T_1^*)$. In Anlehnung an die Notation in Kapitel 1.6 aus dem Buch [57] von Theil sei weiterhin $\partial f^*_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta} = (\partial f^*_{\boldsymbol{\theta}}/\partial\theta, \partial f^*_{\boldsymbol{\theta}}/\partial\vartheta)^{\mathrm{T}}$ der Gradient der Funktion $f^*_{\boldsymbol{\theta}}$ sowie $\partial f^*_{\boldsymbol{\theta}}/\partial \boldsymbol{\theta}^{\mathrm{T}}$ der transponierte Gradient. Entsprechend bezeichne $\partial^2 f^*_{\boldsymbol{\theta}}/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}})$ die Hesse-Matrix. In diesem sowie in den folgenden Abschnitten wird die abkürzende Notation $\partial f^*_{\boldsymbol{\theta}_0}/\partial\boldsymbol{\theta} = \partial f^*_{\boldsymbol{\theta}}/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ für die Ableitungen der Funktion $f^*_{\boldsymbol{\theta}}$ und in analoger Weise für weitere Funktionen verwendet. Es gilt nun

$$\begin{aligned}
\mathbb{E}^*_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1^*, T_1^*)] &= \mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial^2 \log f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right] = \mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\boldsymbol{\theta}}\left(\frac{1}{f^*_{\boldsymbol{\theta}_0}}\frac{\partial f_{\boldsymbol{\theta}^*_0}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right)\right] \\
&= \mathbb{E}^*_{\boldsymbol{\theta}_0}\left[-\frac{1}{f^{*2}_{\boldsymbol{\theta}_0}}\frac{\partial f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}}\frac{\partial f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}^{\mathrm{T}}} + \frac{1}{f^*_{\boldsymbol{\theta}_0}}\frac{\partial^2 f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right] \\
&= -\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\left(\frac{1}{f^*_{\boldsymbol{\theta}_0}}\frac{\partial f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}}\right)\left(\frac{1}{f^*_{\boldsymbol{\theta}_0}}\frac{\partial f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right)\right] + \mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{1}{f^*_{\boldsymbol{\theta}_0}}\frac{\partial^2 f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right] \\
&= -\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\left(\frac{\partial \log f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial \log f^*_{\boldsymbol{\theta}_0}}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right)\right] + \int_S \frac{\partial^2 f^*_{\boldsymbol{\theta}_0}(x,t)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\,\mathrm{d}(x,t).
\end{aligned}$$

Die IME des Modells M^* erhält man somit, wenn der zweite Summand verschwindet, das heißt wenn alle Einträge der Matrix gleich null sind. Exemplarisch wird dies für die erste Komponente gezeigt. Zur Vereinfachung der Notation bezeichne $\dot{f}_{\boldsymbol{\theta}_0}$ die erste beziehungsweise $\ddot{f}_{\boldsymbol{\theta}_0}$ die zweite Ableitung nach θ. Dies gilt entsprechend auch für die Auswahlwahrscheinlichkeit $\alpha_{\boldsymbol{\theta}_0}$. Auf D gelten

$$\frac{\partial f^*_{\boldsymbol{\theta}_0}}{\partial\theta} = \frac{\dot{f}_{\boldsymbol{\theta}_0}\cdot\alpha_{\boldsymbol{\theta}_0} - \dot{\alpha}_{\boldsymbol{\theta}_0}\cdot f_{\boldsymbol{\theta}_0}}{\alpha^2_{\boldsymbol{\theta}_0}}$$

sowie

$$\frac{\partial^2 f^*_{\boldsymbol{\theta}_0}}{\partial\theta^2} = \frac{\ddot{f}_{\boldsymbol{\theta}_0}\cdot\alpha_{\boldsymbol{\theta}_0} - \ddot{\alpha}_{\boldsymbol{\theta}_0}\cdot f_{\boldsymbol{\theta}_0}}{\alpha^2_{\boldsymbol{\theta}_0}} - \frac{2\dot{\alpha}_{\boldsymbol{\theta}_0}\cdot(\dot{f}_{\boldsymbol{\theta}_0}\cdot\alpha_{\boldsymbol{\theta}_0} - \dot{\alpha}_{\boldsymbol{\theta}_0}\cdot f_{\boldsymbol{\theta}_0})}{\alpha^3_{\boldsymbol{\theta}_0}}.$$

Wegen $\int_D f_{\boldsymbol{\theta}_0}(x,t)\,\mathrm{d}(x,t) = \alpha_{\boldsymbol{\theta}_0}$ erhält man für das Integral

$$\int_D \frac{\partial^2 f^*_{\boldsymbol{\theta}_0}(x,t)}{\partial\theta^2}\,\mathrm{d}(x,t) = \int_D \frac{\ddot{f}_{\boldsymbol{\theta}_0}(x,t)}{\alpha_{\boldsymbol{\theta}_0}}\,\mathrm{d}(x,t) - \frac{\ddot{\alpha}_{\boldsymbol{\theta}_0}}{\alpha_{\boldsymbol{\theta}_0}} - \frac{2\dot{\alpha}_{\boldsymbol{\theta}_0}}{\alpha^2_{\boldsymbol{\theta}_0}}\int_D \dot{f}_{\boldsymbol{\theta}_0}(x,t)\,\mathrm{d}(x,t) + \frac{2\dot{\alpha}^2_{\boldsymbol{\theta}_0}}{\alpha^2_{\boldsymbol{\theta}_0}}.$$

Aus Lemma 3.3 erhält man die Vertauschbarkeit der Differentiation und Integration, das heißt es gelten $\int_D \dot{f}_{\boldsymbol{\theta}_0}(x,t)\,\mathrm{d}(x,t) = \dot{\alpha}_{\boldsymbol{\theta}_0}$ sowie $\int_D \ddot{f}_{\boldsymbol{\theta}_0}(x,t)\,\mathrm{d}(x,t) = \ddot{\alpha}_{\boldsymbol{\theta}_0}$. Der Wert des Integrals der zweiten Ableitung der Dichte $f^*_{\boldsymbol{\theta}_0}$ ist somit gleich null. Dies lässt sich auf ähnliche Weise auch für die weiteren Komponenten der obigen Matrix überprüfen. Auf weitere Berechnungen dazu wird an dieser Stelle jedoch verzichtet.
Als Ergebnis erhält man die IME für das Modell M^* und wie oben beschrieben daraus auch für das Modell der abhängigen Trunkation.

Satz 4.4.
Unter den Annahmen (A1)-(A4) *mit* $\boldsymbol{\theta}_0 \in \Theta$ *gilt für die Fisher-Informationsmatrix die Gleichheit*

$$\mathcal{I}(\boldsymbol{\theta}_0) = -\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1, T_1)].$$

In Lemma 4.3 wurde bereits gezeigt, dass $\boldsymbol{\theta}_0$ eine Nullstelle der Abbildung $\boldsymbol{\theta} \mapsto \Psi(\boldsymbol{\theta})$ darstellt. Die Konsistenz der Schätzfolge erfordert die Eindeutigkeit der Nullstelle dieser Abbildung, die aus der Invertierbarkeit der Fisher-Informationsmatrix abgeleitet werden kann.
Die nachfolgende Abbildung zeigt grafisch, dass die Determinante ausschließlich positive Werte annimmt. G und s sowie der Ausschnitt des Parameterraums wurden entsprechend der Anwendung in Kapitel 7 gewählt.

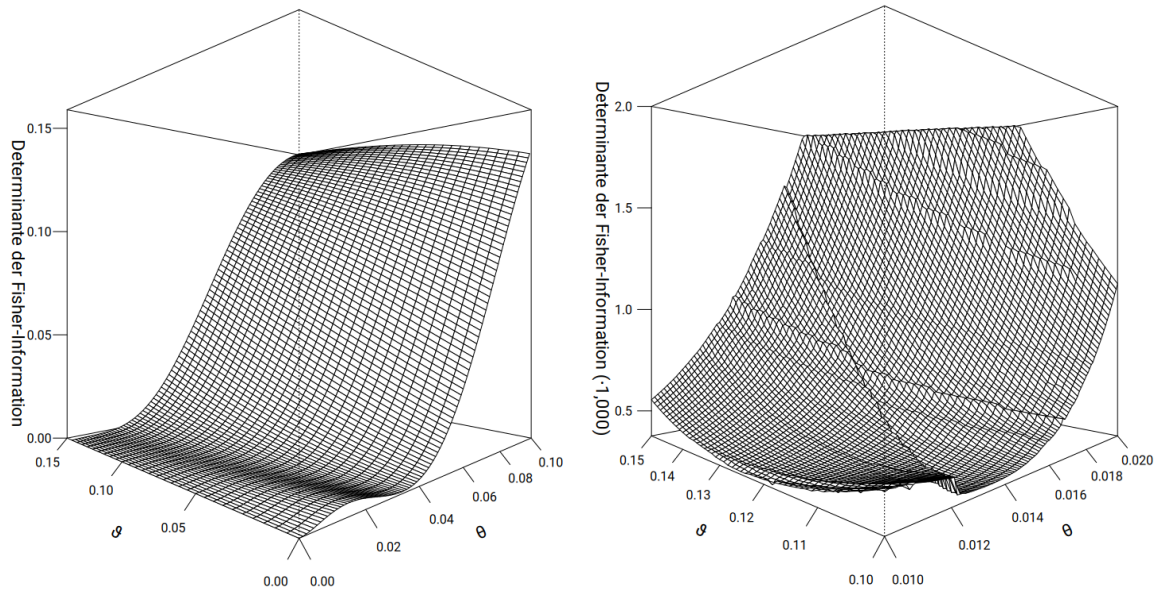

Abbildung 4.2.: Determinante von $\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1, T_1)]$ für einen für die Anwendung 7 relevanten Bereich des Parameterraumes (links) und eine Vergrößerung auf dem Bereich mit den kleinsten Werten der Determinante (rechts) mit $G = 24$ und $s = 3$

Der folgende Satz gibt einen rigorosen Beweis der Invertierbarkeit.

Satz 4.5.
Unter den Annahmen (A1)-(A4) *mit* $\boldsymbol{\theta}_0 \in \Theta$ *ist die Matrix* $\mathcal{I}(\boldsymbol{\theta}_0)$ *invertierbar.*

Beweis. Die Matrix $\mathcal{I}(\boldsymbol{\theta}_0)$ hat die Form

$$\begin{pmatrix} \mathbb{E}_{\boldsymbol{\theta}_0}[\psi^2_{\boldsymbol{\theta}_0,1}(X_1, T_1)] & \mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0,1}(X_1, T_1)\psi_{\boldsymbol{\theta}_0,2}(X_1, T_1)] \\ \mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0,1}(X_1, T_1)\psi_{\boldsymbol{\theta}_0,2}(X_1, T_1)] & \mathbb{E}_{\boldsymbol{\theta}_0}[\psi^2_{\boldsymbol{\theta}_0,2}(X_1, T_1)] \end{pmatrix}.$$

Mithilfe der führenden Hauptminoren lässt sich zeigen, dass die Matrix positiv definit ist. Wegen Lemma 4.3 stellt $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi^2_{\boldsymbol{\theta}_0,1}(X_1, T_1)]$ die Varianz der ersten Komponente der

Score-Funktion dar. Mit Blick auf die Definition 3.6a wird deutlich, dass die Zufallsvariable $\psi_{\boldsymbol{\theta}_0,1}(X_1, T_1)$ für $(X_1, T_1) \in D$ nicht konstant ist und positive Varianz besitzen muss.
Gemäß der Hölder- bzw. Cauchy-Schwarz-Ungleichung gilt

$$\mathbb{E}_{\boldsymbol{\theta}_0}\left[|\psi_{\boldsymbol{\theta}_0,1}(X_1,T_1)\psi_{\boldsymbol{\theta}_0,2}(X_1,T_1)|\right] \leq \left(\mathbb{E}_{\boldsymbol{\theta}_0}[\psi^2_{\boldsymbol{\theta}_0,1}(X_1,T_1)]\right)^{\frac{1}{2}} \left(\mathbb{E}_{\boldsymbol{\theta}_0}[\psi^2_{\boldsymbol{\theta}_0,2}(X_1,T_1)]\right)^{\frac{1}{2}}.$$

Mit der Dreiecksungleichung für Integrale folgt

$$\mathbb{E}_{\boldsymbol{\theta}_0}[\psi^2_{\boldsymbol{\theta}_0,1}(X_1,T_1)] \cdot \mathbb{E}_{\boldsymbol{\theta}_0}[\psi^2_{\boldsymbol{\theta}_0,2}(X_1,T_1)] - \left(\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0,1}(X_1,T_1)\psi_{\boldsymbol{\theta}_0,2}(X_1,T_1)]\right)^2 \geq 0,$$

sodass die Matrix $\mathcal{I}(\boldsymbol{\theta}_0)$ positiv semidefinit ist. In der Cauchy-Schwarz-Ungleichung tritt die Gleichheit genau dann auf, wenn die Funktionen $\psi_{\boldsymbol{\theta}_0,1}$ und $\psi_{\boldsymbol{\theta}_0,2}$ linear abhängig sind. Die Determinante der Matrix ist somit von Null verschieden, wenn $a_1 = a_2 = 0$ aus $a_1 \cdot \psi_{\boldsymbol{\theta}_0,1} + a_2 \cdot \psi_{\boldsymbol{\theta}_0,2} = 0$ folgt. Es bezeichne $\dot{\alpha}_\theta$ die partielle Ableitung von $\alpha_{\boldsymbol{\theta}}$ nach θ, $\dot{\alpha}_\vartheta$ die partielle Ableitung von $\alpha_{\boldsymbol{\theta}}$ nach ϑ und $y = \vartheta \log(1 - t/G) - 1$. In den nachfolgenden Überlegungen wird der Erwartungswert nicht berücksichtigt, weshalb zur besseren Lesbarkeit in $\boldsymbol{\theta}$ statt in $\boldsymbol{\theta}_0$ formuliert wird. Weiterhin wird $(x, t) \in D$ angenommen, da die Score-Funktionen außerhalb von D den Wert Null annehmen (vgl. (3.6)). Die Ausgangsgleichung lautet

$$\begin{aligned}
0 = a_1 &\left\{\frac{\vartheta\theta x + 1}{\theta} y + \frac{\vartheta}{\theta} + xy[(\vartheta\theta x + 1)y + \vartheta] + \vartheta xy - \frac{\dot{\alpha}_\theta}{\alpha_{\boldsymbol{\theta}}}[(\vartheta\theta x + 1)y + \vartheta]\right\} + \\
a_2 &\left\{\theta x \frac{(y+1)}{\vartheta}[(\vartheta\theta x + 1)y + \vartheta] + (2\vartheta\theta x + 1)\frac{(y+1)}{\vartheta} - \theta x + 1 - \frac{\dot{\alpha}_\vartheta}{\alpha_{\boldsymbol{\theta}}}[(\vartheta\theta x + 1)y + \vartheta]\right\}.
\end{aligned}$$

Zusammengefasst ergibt sich

$$\begin{aligned}
0 = a_1 &\left\{3\vartheta xy + \frac{y}{\theta} + \frac{\vartheta}{\theta} + \vartheta\theta x^2 y^2 + xy^2 - \frac{\dot{\alpha}_\theta}{\alpha_{\boldsymbol{\theta}}}[(\vartheta\theta x + 1)y + \vartheta]\right\} + \\
a_2 &\left\{\theta^2 x^2 (y+1)y + \frac{\theta x(y+1)y}{\vartheta} + 3\theta xy + 2\theta x + \frac{(y+1)}{\vartheta} + 1 - \frac{\dot{\alpha}_\vartheta}{\alpha_{\boldsymbol{\theta}}}[(\vartheta\theta x + 1)y + \vartheta]\right\}.
\end{aligned}$$

Differenziert man nun zweimal nach x und einmal nach y, so führt dies zur Gleichung

$$a_1 \cdot \{4\vartheta\theta y\} + a_2\{2\theta^2(y+1)y\} = 0.$$

Nach erneutem Ableiten nach y erhält man unter der Annahme $\vartheta > 0$ den Zusammenhang $a_1 = -\theta/\vartheta a_2$. Eingesetzt in die vorherige Gleichung liefert dies $a_2 = 0$ und somit auch $a_1 = 0$.
Im Fall $\vartheta = 0$ ergibt sich mit $y = \log(1 - t/G)$ die folgende Gleichung

$$a_1\left\{\frac{1}{\theta} - x - \frac{\dot{\alpha}_\theta}{\alpha_{\boldsymbol{\theta}}}\right\} + a_2\left\{(\theta x - 1)y - \frac{\dot{\alpha}_\vartheta}{\alpha_{\boldsymbol{\theta}}}\right\} = 0.$$

Für die Form der Auswahlwahrscheinlichkeit für $\vartheta = 0$ siehe Gleichung 3 in [65]. Für die

Argumentation ist es unerheblich welche Form die zugehörigen Ableitungen für $\vartheta = 0$ annehmen, sodass darauf nicht weiter eingegangen wird. Das Ableiten nach x liefert

$$-a_1 + a_2\theta(y+1) = 0.$$

Differenziert man zusätzlich nach y, so folgt $a_2 = 0$ und mit der vorherigen Gleichung ebenso $a_1 = 0$.
Als Konsequenz sind die beiden Funktionen $\psi_{\boldsymbol{\theta},1}$ und $\psi_{\boldsymbol{\theta},2}$ linear unabhängig und die Matrix $\mathcal{I}(\boldsymbol{\theta}_0)$ ist positiv definit und damit invertierbar. ■

4.3. Konsistenz

Eine wünschenswerte Eigenschaft der Schätzfolge $\hat{\boldsymbol{\theta}}_n$ ist die Konvergenz in Wahrscheinlichkeit gegen den wahren Parameterwert $\boldsymbol{\theta}_0$. Die Folge von Schätzern wird dann als konsistent bezeichnet. Diese wird mithilfe der Theorie über Z-Schätzer von van der Vaart [61] bewiesen.
Mit der Unverzerrtheit der Score-Funktion in Lemma 4.3 und der Invertierbarkeit der Fisher-Information in Satz 4.5 wurden bereits wichtige Voraussetzungen für die Konsistenz gezeigt. Das nächste Lemma thematisiert die Integrierbarkeit der Score-Funktion, welche für die gleichmäßige Konvergenz der Kriteriumsfunktion eine wichtige Rolle spielen wird.

Lemma 4.6.
Unter den Annahmen (A1)-(A4) *existiert eine vom Parameter* $\boldsymbol{\theta}$ *unabhängige Funktion* $g : S \to \mathbb{R}_0^+$ *mit* $\mathbb{E}_{\boldsymbol{\theta}_0}[g(X_1, T_1)] < \infty$, *sodass* $\|\psi_{\boldsymbol{\theta}}(x,t)\| \leq g(x,t)$ *für alle* $\boldsymbol{\theta} \in \Theta$ *und* $(x,t)^T \in S$ *erfüllt. Insbesondere gilt* $\mathbb{E}_{\boldsymbol{\theta}_0}[\|\psi_{\boldsymbol{\theta}_0}(X_1, T_1)\|^2] < \infty$ *für* $\boldsymbol{\theta}_0 \in \Theta$.

Beweis. Mit der umgekehrten Dreiecksungleichung gilt zunächst wegen $\vartheta \in [0, 1 - \varepsilon_\vartheta]$ für alle $(x,t)^{\mathrm{T}} \in D$ die Abschätzung

$$\Big|(\vartheta\theta x + 1)[\vartheta \log(1 - t/G) - 1] + \vartheta\Big| \geq \Big|(\vartheta\theta x + 1)[1 - \vartheta \log(1 - t/G)] - \vartheta\Big| \geq 1 - \vartheta \geq \varepsilon_\vartheta.$$

Damit gilt für die erste Komponente der Funktion (3.6a) nun

$$\begin{aligned}
|\psi_{\boldsymbol{\theta},1}(x,t)| &\leq \mathbb{1}_{[t,t+s]}(x) \cdot \left\{ \frac{1}{\theta} + x[1 - \vartheta \log(1 - t/G)] \right. \\
&\quad \left. + \left| \frac{\vartheta x[\vartheta \log(1 - t/G) - 1]}{(\vartheta\theta x + 1)[\vartheta \log(1 - t/G) - 1] + \vartheta} \right| + \frac{|\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \theta}|}{\alpha_{\boldsymbol{\theta}}} \right\} \\
&\leq \mathbb{1}_{[t,t+s]}(x) \cdot \left\{ \frac{1}{\varepsilon_\theta} + (G+s)[1 - \vartheta \log(1 - t/G)] \right. \\
&\quad \left. + \frac{(G+s)}{\varepsilon_\vartheta}[1 - \vartheta \log(1 - t/G)] + K_1 \right\} \\
&\leq \mathbb{1}_{[t,t+s]}(x) \cdot \left\{ \frac{1}{\varepsilon_\theta} + \frac{2(G+s)}{\varepsilon_\vartheta}[1 - \vartheta \log(1 - t/G)] + K_1 \right\}.
\end{aligned}$$

Hierbei folgt $\theta^{-1} \leq \varepsilon_\theta^{-1}$ aus $\varepsilon_\theta \leq \theta$ und $|\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \theta}| {\alpha_{\boldsymbol{\theta}}}^{-1}$ nimmt aufgrund der in Lemma 3.3 besprochenen Stetigkeit sein Maximum $K_1 \geq 0$ auf Θ an. Ähnlich erhält man für die zweite Komponente (3.6b) die Abschätzung

$$\begin{aligned}
|\psi_{\boldsymbol{\theta},2}(x,t)| &\leq \mathbb{1}_{[t,t+s]}(x) \cdot \left\{ -\theta x \log(1 - t/G) \right. \\
&\quad \left. + \left| \frac{(2\vartheta\theta x + 1)\log(1 - t/G) - \theta x + 1}{(\vartheta\theta x + 1)[\vartheta \log(1 - t/G) - 1] + \vartheta} \right| + \frac{|\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \vartheta}|}{\alpha_{\boldsymbol{\theta}}} \right\} \\
&\leq \mathbb{1}_{[t,t+s]}(x) \cdot \left\{ -\frac{G+s}{\varepsilon_\theta} \log(1 - t/G) + \frac{\theta x[1 - \vartheta \log(1 - t/G)]}{|(\vartheta\theta x + 1)[\vartheta \log(1 - t/G) - 1] + \vartheta|} \right. \\
&\quad \left. + \frac{(\vartheta\theta x + 1)[1 - \log(1 - t/G)]}{|(\vartheta\theta x + 1)[\vartheta \log(1 - t/G) - 1] + \vartheta|} + \frac{1}{\varepsilon_\vartheta} + K_2 \right\} \\
&\leq \mathbb{1}_{[t,t+s]}(x) \cdot \left\{ \left(\frac{3(G+s)}{\varepsilon_\theta \varepsilon_\vartheta} + \frac{1}{\varepsilon_\vartheta} \right) [1 - \log(1 - t/G)] + \frac{1}{\varepsilon_\vartheta} + K_2 \right\},
\end{aligned}$$

wobei auch hier der Satz vom Maximum mit $K_2 \geq 0$ verwendet wurde. Insgesamt erhält man somit für eine hinreichend große Konstante $K_\varepsilon > 0$, welche von ε_θ und ε_ϑ abhängt, dass

$$\psi^2_{\boldsymbol{\theta},1}(x,t) + \psi^2_{\boldsymbol{\theta},2}(x,t) \leq \mathbb{1}_{[t,t+s]}(x) \left\{ K_\varepsilon + K_\varepsilon[1 - \log(1 - t/G)] \right\}^2. \tag{4.2}$$

Es sei $g(x,t) := \mathbb{1}_{[t,t+s]}(x) \left\{ K_\varepsilon + K_\varepsilon[1 - \log(1 - t/G)] \right\}$. Der Erwartungswert der dominierenden Funktion g bezüglich $\boldsymbol{\theta}_0$ ist endlich, denn mithilfe der Randverteilung F^T aus

Annahme (A2) erhält man

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}_0}[g(X_1,T_1)] &\leq \int_0^G \int_0^\infty \{K_\varepsilon + K_\varepsilon[1-\log(1-t/G)]\} \cdot f_{\boldsymbol{\theta}_0}(x,t)\,\mathrm{d}x\,\mathrm{d}t \\ &= K_\varepsilon + \frac{K_\varepsilon}{G}\int_0^G [1-\log(1-t/G)]\,\mathrm{d}t \\ &= K_\varepsilon\left(1+\frac{1}{G}\left[2t - G(1-t/G)\log(1-t/G)\right]_0^G\right) = 3K_\varepsilon < \infty\end{aligned}$$

für eine hinreichend große Konstante $K_\varepsilon > 0$.
Für den Nachweis der zweiten Aussage, wird das Quadrat in Ungleichung (4.2) aufgelöst:

$$\begin{aligned}\psi^2_{\boldsymbol{\theta}_0,1}(x,t) + \psi^2_{\boldsymbol{\theta}_0,2}(x,t) &\leq K_\varepsilon^2\left\{1 + 2[1-\log(1-t/G)] + [1-\log(1-t/G)]^2\right\} \\ &= K_\varepsilon^2\left\{4 - 4\log(1-t/G)] + \log(1-t/G)^2\right\}\end{aligned}$$

Mithilfe der Substitution von $x = 1-t/G$ sowie partieller Integration für $u(x) = \log(x)^2$ und $v(x) = 1$ erhält man

$$\mathbb{E}_{\boldsymbol{\theta}_0}[(\log(1-T_1/G))^2] = \int_0^1 (\log(x))^2\,\mathrm{d}x = -\int_0^1 2\log(x)\,\mathrm{d}x = 2.$$

Mit $\mathbb{E}_{\boldsymbol{\theta}_0}[1-\log(1-T_1/G)] = 2$ folgt daraus $\mathbb{E}_{\boldsymbol{\theta}_0}[\|\psi_{\boldsymbol{\theta}_0}(X_1,T_1)\|^2] \leq 10K_\varepsilon^2 < \infty$. ■

Bevor schließlich die Konsistenz der Schätzfolge bewiesen werden kann, wird ein Satz angeführt, der für den Nachweis der gleichmäßigen Konvergenz der Kriteriumsfunktion hilfreich sein wird (vgl. Lemma 2.4 aus Newey und McFadden [46]).

Satz 4.7. (Gleichmäßiges Gesetz der großen Zahlen)
Es seien $(X_1,T_1),\ldots,(X_n,T_n)$ unabhängig identisch verteilte Zufallsvariablen, Θ eine kompakte Menge und $\boldsymbol{\theta} \mapsto \psi_{\boldsymbol{\theta}}(x,t)$ für fast alle $(x,t)^T \in S$ eine stetige vektorwertige Funktion. Weiterhin existiere eine Funktion $g: S \to \mathbb{R}_0^+$ mit $\mathbb{E}_{\boldsymbol{\theta}_0}[g(X_1,T_1)] < \infty$, sodass $\|\psi_{\boldsymbol{\theta}}(x,t)\| \leq g(x,t)$ für alle $\boldsymbol{\theta} \in \Theta$ und $(x,t)^T \in S$ gilt.
Dann ist die Funktion $\boldsymbol{\theta} \mapsto \Psi(\boldsymbol{\theta}) := \mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}}(X_1,T_1)]$ stetig und es gilt

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n \psi_{\boldsymbol{\theta}}(X_i,T_i) - \Psi(\boldsymbol{\theta})\right\| \xrightarrow{P} 0.$$

Bemerkung 4.8.
Im Kapitel 6 wird der Begriff der Glivenko-Cantelli-Klasse eingeführt. Beispiel 6.9 wird aufzeigen, dass die Klasse $\{\psi_{\boldsymbol{\theta},1}, \psi_{\boldsymbol{\theta},2} : \boldsymbol{\theta} \in \Theta\}$ unter den Voraussetzungen von Satz 4.7 eine Glivenko-Cantelli-Klasse darstellt und somit das gleichmäßige Gesetz der großen Zahlen erfüllt ist. Dies stellt eine alternative Argumentation zu Lemma 2.4 aus Newey und McFadden [46] dar. □

Lemma 4.9.
Unter den Annahmen (A1)-(A4) *gilt* $\Psi_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} \mathbf{0}$.

Beweis. Ziel ist es zu zeigen, dass $\mathbb{P}_{\boldsymbol{\theta}_0}(\|\Psi_n(\hat{\boldsymbol{\theta}}_n)\| > 0) \to 0$. Für $\hat{\boldsymbol{\theta}}_n \in \mathring{\Theta}$ gilt direkt $\Psi_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$, sodass insbesondere die Randwerte betrachtet werden sollen, die keine Nullstelle darstellen. Weißbach und Wied [65] argumentieren in Lemma 2 (v) mit dem Monotonieverhalten der Kriteriumsfunktionen. Die Monotonieeigenschaften können aufgrund der komplexen Struktur der Funktionen in diesem Fall jedoch nicht direkt nachgewiesen werden. Alternativ werden verschiedene Fälle zur Lage der Kriteriumsfunktion relativ zur null betrachtet. Die verschiedenen Fälle führen dann zu den verschiedenen Randpunkten des Rechtecks $[\varepsilon_\theta, 1/\varepsilon_\theta] \times [0, 1-\varepsilon_\vartheta]$. Die vier Eckpunkte werden dabei gesondert betrachtet. Es wird eingangs $\boldsymbol{\theta}_0 \in \mathring{\Theta}$ angenommen.
Es sei zunächst $\hat{\theta}_n = \theta \in (\varepsilon_\theta, 1/\varepsilon_\theta)$. Dann gilt $\Psi_{n,1}(\theta, \hat{\vartheta}_n) = 0$ für $\hat{\vartheta}_n = 0$ und für $\hat{\vartheta}_n = 1 - \varepsilon_\vartheta$. Die Sätze 4.4 und 4.5 liefern $\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0,2}(X_1, T_1)] < 0$. Somit ist die Abbildung $\vartheta \mapsto \Psi_2(\theta, \vartheta)$ aufgrund ihrer Stetigkeit in einer Umgebung um $\boldsymbol{\theta}_0$ streng monoton fallend. Nach Satz 4.3 ist $\boldsymbol{\theta}_0$ eine Nullstelle von $\boldsymbol{\theta} \mapsto \Psi(\boldsymbol{\theta})$. Im Rahmen des Konsistenzbeweises wird sich zeigen, dass die Nullstelle eindeutig ist (Satz 4.10 (ii)). Aus dem Monotonieverhalten, der Eindeutigkeit der Nullstelle und der Stetigkeit müssen auf den Rändern $\Psi_2(\theta, 0) > 0$ sowie $\Psi_2(\theta, 1 - \varepsilon_\vartheta) < 0$ gelten. Falls nun $\Psi_{n,2}(\theta, 0) < 0$, dann folgt direkt mit dem Gesetz der großen Zahlen

$$0 < |\Psi_{n,2}(\theta, 0)| \leq |\Psi_{n,2}(\theta, 0) - \Psi_2(\theta, 0)| \xrightarrow{P} 0.$$

Falls $\Psi_{n,2}(\theta, 0) > 0$ muss auch $\Psi_{n,2}(\theta, 1 - \varepsilon_\vartheta) > 0$ gelten, da sonst $\Psi_{n,2}(\hat{\boldsymbol{\theta}}_n) = 0$. Es folgt ebenfalls mit dem Gesetz der großen Zahlen

$$0 < |\Psi_{n,2}(\theta, 1 - \varepsilon_\vartheta)| \leq |\Psi_{n,2}(\theta, 1 - \varepsilon_\vartheta) - \Psi_2(\theta, 1 - \varepsilon_\vartheta)| \xrightarrow{P} 0.$$

Insbesondere gilt $\mathbb{P}_{\boldsymbol{\theta}_0}(|\Psi_{n,2}(\theta, 0)| > 0) = \mathbb{P}_{\boldsymbol{\theta}_0}(|\Psi_{n,2}(\theta, 1 - \varepsilon_\vartheta)| > 0)$, gegeben, dass sich die Nullstelle von $\Psi_{n,2}$ außerhalb von Θ befindet.
Auf analoge weise kann die Argumentation auch für $\hat{\vartheta}_n = \vartheta \in (0, 1 - \varepsilon_\vartheta)$ und den Punkten $(\varepsilon_\theta, \vartheta)$ sowie $(1/\varepsilon_\theta, \vartheta)$ für die Funktionen $\Psi_{n,1}$ und Ψ_1 geführt werden.
Weiterhin kann die Argumentation ebenso für $\boldsymbol{\theta}_0 \in \Theta$ geführt werden. Die Nullstelle der Funktion $\boldsymbol{\theta} \mapsto \Psi(\boldsymbol{\theta})$ kann sich dann auch auf dem Rand des Parameterraumes befinden. Die anderen Abschätzungen bleiben jedoch erhalten. Für die Eckpunkte kann genauso verfahren werden. Es werden jedoch mehr Fallunterscheidungen benötigt, da beide Komponenten der Funktion $\Psi_n = (\Psi_{n,1}, \Psi_{n,2})$ in den Eckpunkten ungleich null sind. ■

Die Konsistenz des Schätzers lässt sich nun mithilfe des nächsten Satzes (vgl. Satz 5.9 aus [61]) beweisen.

Satz 4.10.
Es seien Ψ_n, $n \in \mathbb{N}$, vektorwertige Zufallsfunktionen und Ψ eine vektorwertige Funktion in $\boldsymbol{\theta} \in \Theta$, sodass für alle $\varepsilon > 0$ die folgenden Bedingungen erfüllt sind

(i) $\sup_{\boldsymbol{\theta} \in \Theta} \|\Psi_n(\boldsymbol{\theta}) - \Psi(\boldsymbol{\theta})\| \xrightarrow{P} 0$,

(ii) $\inf_{\boldsymbol{\theta} \in \Theta:\, d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \geq \varepsilon} \|\Psi(\boldsymbol{\theta})\| > 0 = \|\Psi(\boldsymbol{\theta}_0)\|$.

Dann konvergiert jede Folge von Schätzern $(\hat{\boldsymbol{\theta}}_n)_{n\in\mathbb{N}}$ *mit* $\Psi_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} 0$ *in Wahrscheinlichkeit gegen den wahren Parameterwert* $\boldsymbol{\theta}_0 \in \Theta$.

Satz 4.11. (Konsistenz)
Unter den Annahmen (A1)-(A4) *konvergiert die Schätzfolge* $\hat{\boldsymbol{\theta}}_n$ *aus Gleichung* (3.5) *in Wahrscheinlichkeit gegen den wahren Parameterwert* $\boldsymbol{\theta}_0 \in \Theta$.

Beweis. Zunächst wird die Bedingung (i) des Satzes 4.10 überprüft. Diese liefert der Satz 4.7 über das gleichmäßige Gesetz der großen Zahlen, dessen Bedingungen dafür nachgeprüft werden.
Per Definition sind $(X_1, T_1)^{\mathrm{T}}, \ldots, (X_n, T_n)^{\mathrm{T}}$ unabhängig identisch verteilte Zufallsvariablen und Θ eine kompakte Menge. Die vektorwertige Funktion $\boldsymbol{\theta} \mapsto \psi_{\boldsymbol{\theta}}(x,t)$ ist als Komposition stetiger Funktionen für fast alle $(x,t)^{\mathrm{T}} \in S$ stetig. Die Stetigkeit der Auswahlwahrscheinlichkeit $\alpha_{\boldsymbol{\theta}}$ sowie die Stetigkeit der zugehörigen partiellen Ableitungen wurde in Lemma 3.3 thematisiert. Lemma 4.6 liefert die Abschätzung der Score-Funktion durch eine vom Parameter $\boldsymbol{\theta}$ unabhängige Funktion mit endlichem Erwartungswert. Somit folgt Teil (i) des Satzes 4.10 aus dem Satz 4.7 über das gleichmäßige Gesetz der großen Zahlen.
Nach Problem 5.27 in van der Vaart [61] ist die Bedingung (ii) aus Satz 4.10 äquivalent dazu, dass die Gleichung $\Psi(\boldsymbol{\theta}) = \mathbf{0}$ die eindeutige Lösung $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ besitzt. Aus dem Lemma 4.3 geht hervor, dass $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ eine Lösung dieser Gleichung ist. Die Eindeutigkeit der Lösung kann nun über die in Satz 4.1 bewiesene Identifizierbarkeit des Parameters $\boldsymbol{\theta}$ im Modell M^* hergeleitet werden. Aufgrund des Zusammenhangs $f^*_{\boldsymbol{\theta}_0} = \mathbb{1}_D f_{\boldsymbol{\theta}_0}/\alpha_{\boldsymbol{\theta}_0}$ gilt zunächst

$$\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\psi_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right] = \frac{1}{\alpha_{\boldsymbol{\theta}_0}} \mathbb{E}_{\boldsymbol{\theta}_0}\left[\psi_{\boldsymbol{\theta}}(X_1, T_1)\right].$$

Weiterhin lässt sich mithilfe der Gleichungen in (3.6) überprüfen (siehe Appendix A.1), dass

$$\left(\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\theta}\log f^*_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right], \mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\vartheta}\log f^*_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right]\right) = \mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\psi_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right], \tag{4.3}$$

sodass auch die Gleichung

$$\left(\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\theta}\log f^*_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right], \mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\vartheta}\log f^*_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right]\right) = 0 \tag{4.4}$$

eine eindeutige Lösung in $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ besitzt. Es gelten nun auch

$$\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial^2}{\partial\theta^2}\log f^*_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right] = \frac{1}{\alpha_{\boldsymbol{\theta}_0}} \mathbb{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\theta}\psi_{\boldsymbol{\theta},1}(X_1, T_1)\right],$$
$$\mathbb{E}^*_{\boldsymbol{\theta}_0}\left[\frac{\partial^2}{\partial\vartheta^2}\log f^*_{\boldsymbol{\theta}}(X_1^*, T_1^*)\right] = \frac{1}{\alpha_{\boldsymbol{\theta}_0}} \mathbb{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial}{\partial\vartheta}\psi_{\boldsymbol{\theta},2}(X_1, T_1)\right]$$

sowie

$$\mathbb{E}^*_{\boldsymbol{\theta}_0}\Big[\frac{\partial^2}{\partial\theta\partial\vartheta}\log f^*_{\boldsymbol{\theta}}(X^*_1,T^*_1)\Big] = \frac{1}{\alpha_{\boldsymbol{\theta}_0}}\mathbb{E}_{\boldsymbol{\theta}_0}\Big[\frac{\partial}{\partial\theta}\psi_{\boldsymbol{\theta},2}(X_1,T_1)\Big] = \frac{1}{\alpha_{\boldsymbol{\theta}_0}}\mathbb{E}_{\boldsymbol{\theta}_0}\Big[\frac{\partial}{\partial\vartheta}\psi_{\boldsymbol{\theta},1}(X_1,T_1)\Big].$$

Aufgrund der positiven Definitheit der Matrix $\mathcal{I}(\boldsymbol{\theta}_0)$ in Satz 4.5 und der IME in Satz 4.4 ist die Eindeutigkeit der Lösung von (4.4) äquivalent zur Eindeutigkeit des Maximums der Funktion $\boldsymbol{\theta} \mapsto \mathbb{E}^*_{\boldsymbol{\theta}_0}[\log f^*_{\boldsymbol{\theta}}(X^*_1,T^*_1)]$. Die notwendige Vertauschbarkeit der Differentiation und Integration erhält man durch ähnliches Vorgehen wie in Lemma 3.3 durch das Anwenden von Satz 5.7 (Kap. IV, § 5.) aus dem Buch [13] von Elstrodt. Nach Satz 2.14 ist die Eindeutigkeit des Maximums wiederum äquivalent zur Identifizierbarkeit des Parameterwertes $\boldsymbol{\theta}_0$ im Modell M^*.
Die Bedingungen (i) und (ii) des Satzes 4.10 sind erfüllt. Somit folgt die Konsistenz der Schätzfolge mit Satz 4.10, da nach Lemma 4.9 $\Psi_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} \mathbf{0}$. ■

4.4. Asymptotische Normalität

Nachdem die Konsistenz der Schätzfolge gezeigt wurde, stellt sich die Frage, in welcher Ordnung $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$ gegen null konvergiert. Diese beträgt häufig $n^{-1/2}$ und die Folge $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ konvergiert in Verteilung gegen eine Normalverteilung. Diese sogenannte asymptotische Normalität der Schätzfolge lässt sich mithilfe des nächsten Satzes (vgl. Satz 5.41 aus [61]), der ebenfalls auf der Theorie der Z-Schätzer basiert, beweisen. Im Folgenden sei $\mathring{\Theta} := (\varepsilon_\theta, 1/\varepsilon_\theta) \times (0, 1-\varepsilon_\vartheta)$.

Satz 4.12.
Es sei $\boldsymbol{\theta}_0 \in \mathring{\Theta}$, $\boldsymbol{\theta} \mapsto \psi_{\boldsymbol{\theta}}(x,t)$ für alle $\boldsymbol{\theta} \in \mathring{\Theta}$ und für alle $(x,t)^T \in S$ eine zweimal stetig partiell differenzierbare Abbildung, welche die folgenden Voraussetzungen erfülle:

(i) $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0}(X_1,T_1)] = \mathbf{0}$,

(ii) $\mathbb{E}_{\boldsymbol{\theta}_0}[\|\psi_{\boldsymbol{\theta}_0}(X_1,T_1)\|^2] < \infty$ *sowie*

(iii) $\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1,T_1)]$ *ist invertierbar.*

Darüber hinaus seien die zweiten partiellen Ableitungen von $\psi_{\boldsymbol{\theta}}(x,t)$ für alle $\boldsymbol{\theta}$ aus einer Umgebung von $\boldsymbol{\theta}_0$ durch eine integrierbare Funktion $\ddot{\psi}(x,t)$ dominiert.
Dann erfüllt jede konsistente Schätzfolge $(\hat{\boldsymbol{\theta}}_n)_{n\in\mathbb{N}}$ mit $\Psi_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$ für alle $n \in \mathbb{N}$, dass

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}])^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_{\boldsymbol{\theta}_0}(X_i,T_i) + o_{\mathbb{P}_{\theta_0}}(1),$$

wobei $o_{\mathbb{P}_{\theta_0}}(1)$ für eine in Wahrscheinlichkeit gegen 0 konvergierende Folge steht. Insbesondere ist die Folge $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ asymptotisch normalverteilt mit dem Erwartungswertvektor $\mathbf{0}$ und der Kovarianzmatrix $(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}])^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0}\psi^T_{\boldsymbol{\theta}_0}](\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}])^{-1}$.

Dank der IME aus Satz 4.4 vereinfacht sich die Kovarianzmatrix und die Berechnung der zweiten Ableitungen für die Hesse-Matrix ist nicht mehr notwendig.

Satz 4.13. (Asymptotische Normalität)
Unter den Annahmen (A1)-(A4) *mit* $\boldsymbol{\theta}_0 \in \mathring{\Theta}$ *konvergiert die Folge* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *in Verteilung gegen eine normalverteilte Zufallsgröße mit dem Erwartungswertvektor* $\mathbf{0}$ *und der Kovarianzmatrix* $\mathcal{I}(\boldsymbol{\theta}_0)^{-1}$.

Beweis. Als Funktion in $\boldsymbol{\theta}$ sind die gebrochenrationalen Funktionen in den Gleichungen (3.6a) sowie (3.6b) zweimal stetig partiell differenzierbar, da der Nenner für $\boldsymbol{\theta} \in \Theta$ und $(X_i, T_i)^{\mathrm{T}} \in D$ keine Nullstelle besitzt. Weiterhin ist die Auswahlwahrscheinlichkeit $\alpha_{\boldsymbol{\theta}}$ als Funktion in $\boldsymbol{\theta}$ nach Lemma 3.3 dreimal stetig partiell differenzierbar. Die Funktion $\boldsymbol{\theta} \mapsto \psi_{\boldsymbol{\theta}}(x,t)$ ist somit als Komposition zweimal stetig partiell differenzierbarer Funktionen für alle $(x,t)^{\mathrm{T}} \in S$ ebenfalls zweimal stetig partiell differenzierbar.
Die Bedingung (i) aus Satz 4.12 wurde bereits in Lemma 4.3 bewiesen, (ii) geht aus Lemma 4.6 hervor und Bedingung (iii) resultiert aus der positiven Definitheit der Matrix in Satz 4.5 zusammen mit der IME in Satz 4.4.
Als letztes verbleibt der Nachweis der Dominiertheitsbedingung der zweiten partiellen Ableitungen der Funktion $\psi_{\boldsymbol{\theta}}$. Für

$$\begin{aligned} R_{\boldsymbol{\theta}} &:= (\vartheta\theta x + 1)[\vartheta \log(1 - t/G) - 1] + \vartheta \\ \dot{R}_\theta &:= \vartheta x(\vartheta \log(1 - t/G) - 1) \\ \dot{R}_\vartheta &:= (2\vartheta\theta x + 1)\log(1 - t/G) - \theta x + 1 \end{aligned}$$

lauten die partiellen Ableitungen der ersten Ordnung

$$\begin{aligned} \frac{\partial \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \theta} &= \mathbb{1}_{[t,t+s]}(x,t)\left[-\frac{1}{\theta^2} - \frac{\dot{R}_\theta^2}{R_{\boldsymbol{\theta}}^2} - \frac{\partial}{\partial \theta}\left(\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \theta}\frac{1}{\alpha_{\boldsymbol{\theta}}}\right)\right] \\ \frac{\partial \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \vartheta} &= \mathbb{1}_{[t,t+s]}(x,t)\Bigg[x\log(1 - t/G) + \frac{x(2\vartheta \log(1 - t/G) - 1)}{R_{\boldsymbol{\theta}}} \\ &\qquad - \frac{\dot{R}_\vartheta \dot{R}_\theta}{R_{\boldsymbol{\theta}}^2} - \frac{\partial}{\partial \vartheta}\left(\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \theta}\frac{1}{\alpha_{\boldsymbol{\theta}}}\right)\Bigg] \\ \frac{\partial \psi_{\boldsymbol{\theta},2}(x,t)}{\partial \vartheta} &= \mathbb{1}_{[t,t+s]}(x,t)\left[\frac{2\theta x \log(1 - t/G)}{R_{\boldsymbol{\theta}}} - \frac{\dot{R}_\vartheta^2}{R_{\boldsymbol{\theta}}^2} - \frac{\partial}{\partial \vartheta}\left(\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \vartheta}\frac{1}{\alpha_{\boldsymbol{\theta}}}\right)\right]. \end{aligned}$$

Hierbei gilt $\frac{\partial \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \vartheta} = \frac{\partial \psi_{\boldsymbol{\theta},2}(x,t)}{\partial \theta}$. Daraus lassen sich nun die partiellen Ableitungen der

zweiten Ordnung berechnen

$$\begin{aligned}
\frac{\partial^2 \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \theta^2} &= \mathbb{1}_{[t,t+s]}(x,t)\left[\frac{1}{\theta^3} + \frac{2\dot{R}_\theta^3}{R_{\boldsymbol{\theta}}^3} - \frac{\partial^2}{\partial \theta^2}\left(\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \theta}\frac{1}{\alpha_{\boldsymbol{\theta}}}\right)\right] \\
\frac{\partial^2 \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \vartheta^2} &= \mathbb{1}_{[t,t+s]}(x,t)\Bigg[\frac{2x\log(1-t/G)}{R_{\boldsymbol{\theta}}} - \frac{2x(2\vartheta - \log(1-t/G)-1)\dot{R}_\vartheta}{R_{\boldsymbol{\theta}}^2} \\
&\qquad - \frac{2\theta x \log(1-t/G)\dot{R}_\theta}{R_{\boldsymbol{\theta}}^2} + \frac{2\dot{R}_\theta \dot{R}_\vartheta^2}{R_{\boldsymbol{\theta}}^3} - \frac{\partial^2}{\partial \vartheta^2}\left(\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \theta}\frac{1}{\alpha_{\boldsymbol{\theta}}}\right)\Bigg] \\
\frac{\partial^2 \psi_{\boldsymbol{\theta},2}(x,t)}{\partial \theta^2} &= \mathbb{1}_{[t,t+s]}(x,t)\left[-\frac{2x(\vartheta\log(1-t/G)-1)\dot{R}_\theta}{R_{\boldsymbol{\theta}}^2} + \frac{2\dot{R}_\vartheta \dot{R}_\theta^2}{R_{\boldsymbol{\theta}}^3} - \frac{\partial^2}{\partial \theta^2}\left(\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \vartheta}\frac{1}{\alpha_{\boldsymbol{\theta}}}\right)\right] \\
\frac{\partial^2 \psi_{\boldsymbol{\theta},2}(x,t)}{\partial \vartheta^2} &= \mathbb{1}_{[t,t+s]}(x,t)\left[-\frac{6\theta x \log(1-t/G)\dot{R}_\vartheta}{R_{\boldsymbol{\theta}}^2} + \frac{2\dot{R}_\vartheta^3}{R_{\boldsymbol{\theta}}^3} - \frac{\partial^2}{\partial \vartheta^2}\left(\frac{\partial \alpha_{\boldsymbol{\theta}}}{\partial \vartheta}\frac{1}{\alpha_{\boldsymbol{\theta}}}\right)\right].
\end{aligned}$$

Für alle weiteren Ableitungen gilt

$$\frac{\partial^2 \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \vartheta^2} = \frac{\partial^2 \psi_{\boldsymbol{\theta},2}(x,t)}{\partial \vartheta \partial \theta} = \frac{\partial^2 \psi_{\boldsymbol{\theta},2}(x,t)}{\partial \theta \partial \vartheta}$$

beziehungsweise

$$\frac{\partial^2 \psi_{\boldsymbol{\theta},2}(x,t)}{\partial \theta^2} = \frac{\partial^2 \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \vartheta \partial \theta} = \frac{\partial^2 \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \theta \partial \vartheta}.$$

Die partiellen Ableitungen zweiter Ordnung lassen sich nun ähnlich wie die Score-Funktion selbst im Konsistenzbeweis 4.11 abschätzen. Es lässt sich wieder $\vartheta \leq 1$, $\theta \leq \varepsilon_\theta$ sowie $|R_{\boldsymbol{\theta}}| \geq 1 - \vartheta \geq \varepsilon_\vartheta$ nutzen. Die partiellen Ableitungen der Auswahlwarscheinlichkeit lassen sich durch Lemma 3.3 durch eine positive Konstante abschätzen. Mit $K > 0$ folgt daher beispielsweise für $\partial^2 \psi_{\boldsymbol{\theta},1}/(\partial \theta^2)$, dass

$$\begin{aligned}
\left|\frac{\partial^2 \psi_{\boldsymbol{\theta},1}(x,t)}{\partial \theta^2}\right| &\leq \mathbb{1}_{[t,t+s]}(x,t)\left[\frac{1}{\varepsilon_\theta^3} + \frac{2|\dot{R}_\theta^3|}{|R_{\boldsymbol{\theta}}^3|} + K\right] \\
&\leq \mathbb{1}_{[t,t+s]}(x,t)\left[\frac{1}{\varepsilon_\theta^3} + \frac{2(G+s)^3(1-\log(1-t/G))^3}{\varepsilon_\vartheta^3} + K\right].
\end{aligned}$$

Die Integrierbarkeit der Funktion lässt sich wieder mithilfe partieller Integration nachweisen. Da sich die anderen Ableitungen zweiter Ordnung ähnlich abschätzen lassen, wird auf weitere Ausführungen dazu nicht weiter eingegangen.
Nach Satz 4.12 ist $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ für jede konsistente Schätzfolge $\hat{\boldsymbol{\theta}}_n$ mit $\Psi_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$ asymptotisch normalverteilt mit dem Erwartungswertvektor $\mathbf{0}$ und wegen Satz 4.4 mit der Kovarianzmatrix

$$(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}])^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0}\psi_{\boldsymbol{\theta}_0}^{\mathrm{T}}](\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}])^{-1} = -(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}])^{-1} = \mathcal{I}(\boldsymbol{\theta}_0)^{-1}.$$

■

5. Test auf Unabhängigkeit

In diesem Kapitel wird ein statistischer Test zur Untersuchung der Unabhängigkeit von der Lebensdauer X_i und dem Alter T_i bei Studienbeginn entwickelt. Die Untersuchungen fokussieren sich auf die in Annahme (A3) eingeführte Gumbel-Barnett-Copula C_ϑ, da sich für diese eine Besonderheit für die Unabhängigkeit ergibt. Die Copula vereinfacht sich zur Unabhängigkeitscopula, sofern der Parameter ϑ den Wert null annimmt und sich somit auf dem Rand des Parameterraums befindet. In Folge dessen wird zunächst die asymptotische Verteilung der Schätzfolge auf dem Rand des Parameterraumes untersucht, bevor die Durchführung des Tests besprochen werden kann. Des Weiteren widmet sich der zweite Abschnitt der Untersuchung der Teststärke. Die Anwendung des Tests auf Daten sowie eine Simulation zur Teststärke sind in eigene Kapitel 7 und 8 ausgegliedert.

5.1. Asymptotische Verteilung

Im Abschnitt 4.3 wurde die Konsistenz der Schätzfolge $\hat{\boldsymbol{\theta}}_n$ auf dem abgeschlossenen Parameterraum Θ nachgewiesen (vgl. Satz 4.11). Als Abhängigkeitsstruktur wurde dabei die Gumbel-Barnett-Copula nach Annahme (A3) zugrunde gelegt. Im darauffolgenden Abschnitt 4.4 hält Satz 4.13 die asymptotische Normalität der Folge $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ fest. Im Vergleich zur Konsistenz gilt die asymptotische Normalität lediglich auf dem offenen Parameterraum $\mathring{\Theta}$. Zum einen muss die Differenzierbarkeit der Kriteriumsfunktion Ψ_n erfüllt sein. Zum anderen ist nur für $\hat{\boldsymbol{\theta}}_n \in \mathring{\Theta}$ sichergestellt, dass $\hat{\boldsymbol{\theta}}_n$ eine Nullstelle der Kriteriumsfunktion ist. Da die Gumbel-Barnett-Copula eine Abhängigkeitsmodellierung nur in eine Richtung erlaubt, liegt der Parameterwert für die Unabhängigkeit mit $\vartheta_0 = 0$ auf dem Rand des Parameterraums. Zur Untersuchung des asymptotischen Verhaltens der Folge $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ wird die Idee aus Satz 1 aus der Arbeit von Moran [41] aufgegriffen. Die Unkenntnis darüber, ob die Schätzfolge eine Nullstelle der Kriteriumsfunktion darstellt, wird als Bedingung in die Verteilung aufgenommen.

Satz 5.1. (Asymptotische Verteilung für $\vartheta_0 = 0$)
Unter den Annahmen (A2)-(A4) *sowie* $\vartheta_0 = 0$ *und* $\theta_0 \in (\varepsilon_\theta, 1/\varepsilon_\theta)$ *konvergiert die Folge* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ *für* $n \to \infty$ *in Verteilung gegen*

$$\Phi_{\boldsymbol{\theta}_0}(\mathbf{a}) = \frac{1}{2} F_1^{\boldsymbol{\theta}_0}(\mathbf{a}) + \frac{1}{2} F_2^{\boldsymbol{\theta}_0}(\mathbf{a}).$$

Hierbei ist $F_1^{\boldsymbol{\theta}_0}$ *eine zweidimensionale Verteilungsfunktion, die auf dem Bereich* $\mathbf{a} \in (-\infty, \infty) \times (0, \infty)$ *definiert ist und als Dichte zweimal die Dichtefunktion von* $\mathcal{N}_2(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1})$ *besitzt. Die Funktion* $F_2^{\boldsymbol{\theta}_0}$ *ist für* $\mathbf{a} \in (-\infty, \infty) \times \{0\}$ *definiert und stellt*

die eindimensionale Verteilungsfunktion von

$$\sigma^{(2)}\tilde{Y}_1$$

mit $\sigma^{(2)} = (\mathcal{I}(\boldsymbol{\theta}_0)_{11})^{-1}$ *dar. Die Verteilung von* $\tilde{Y}_1$ *ist die Verteilung von* Y_1 *bedingt auf die Ungleichung*

$$Y_2 - \mathcal{I}(\boldsymbol{\theta}_0)_{12} \cdot \sigma^{(2)} Y_1 \leq 0,$$

wobei $(Y_1, Y_2)^T \sim \mathcal{N}_2(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0))$.

Beweis. Um die stetige Differenzierbarkeit der Funktion $\psi_{\boldsymbol{\theta}}$ aus Gleichung (3.6) in $\boldsymbol{\theta}_0$ nachzuweisen, kann der Definitionsbereich der Funktion um eine (kleine) offene Kugel um $\vartheta = 0$ erweitert werden. Die Eigenschaften der Funktion $f_{\boldsymbol{\theta}}$ als Dichte kann nicht mehr garantiert werden. Allerdings lässt sich zeigen, dass die Auswahlwahrscheinlichkeit weiterhin strikt positiv ist und der Nenner in der Funktion $\psi_{\boldsymbol{\theta}}$ für $(x,t) \in D$ keine Nullstellen besitzt. Die im Beweis von Satz 4.13 berechneten Ableitungen existieren somit auch für diese offene Menge und sind stetig.
Der Beweis des Satzes folgt zunächst dem Beweis von Satz 5.41 aus van der Vaart [61] (vgl. Satz 4.12) und greift dann die Idee der Fallunterscheidung für $\hat{\vartheta}_n > 0$ und $\hat{\vartheta}_n = 0$ aus dem Satz 1 von Moran [41] auf.

Nach dem Satz von Taylor existiert ein Vektor $\tilde{\boldsymbol{\theta}}_n \in \{\boldsymbol{\theta}_0 + t(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) | t \in (0,1)\}$, der für beide Komponenten $\Psi_{n,1}$ und $\Psi_{n,2}$ unterschiedlich sein kann, sodass

$$\Psi_n(\hat{\boldsymbol{\theta}}_n) = \Psi_n(\boldsymbol{\theta}_0) + \dot{\Psi}_n(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T \ddot{\Psi}_n(\tilde{\boldsymbol{\theta}}_n)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \tag{5.1}$$

Hierbei gilt $\Psi_{n,1}(\hat{\boldsymbol{\theta}}_n) = 0$ und $\Psi_{n,2}(\hat{\boldsymbol{\theta}}_n) = 0$, falls $\hat{\vartheta}_n > 0$, und $\Psi_{n,2}(\hat{\boldsymbol{\theta}}_n) \leq 0$, falls $\hat{\vartheta}_n = 0$. Der Term $\Psi_n(\boldsymbol{\theta}_0)$ ergibt sich als Mittelwert aus den unabhängig identisch verteilten Zufallsvektoren $\psi_{\boldsymbol{\theta}_0}(X_i, T_i)$, $i = 1, \ldots, n$, mit $\mathbb{E}_{\boldsymbol{\theta}_0}[\psi_{\boldsymbol{\theta}_0}(X_i, T_i)] = \mathbf{0}$ (vgl. Satz 4.3). Nach Lemma 4.6 gilt $\mathbb{E}_{\boldsymbol{\theta}_0}[\|\psi_{\boldsymbol{\theta}_0}(X_1, T_1)\|^2] < \infty$ für $\boldsymbol{\theta}_0 \in \Theta$, sodass die Einträge der Matrix $\mathcal{I}(\boldsymbol{\theta}_0)$ endlich sind. Nach dem multivariaten zentralen Grenzwertsatz konvergiert die Folge $\sqrt{n}\Psi_n(\boldsymbol{\theta}_0)$ in Verteilung gegen $\mathcal{N}_2(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0))$. Die erste Ableitung $\dot{\Psi}_n(\boldsymbol{\theta}_0)$ im zweiten Term konvergiert nach dem Gesetz der großen Zahlen in Wahrscheinlichkeit gegen $\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1, T_1)]$. Der Notation in [61] folgend, ist die zweite Ableitung $\ddot{\Psi}_n(\tilde{\boldsymbol{\theta}}_n)$ ein zweidimensionaler Vektor bestehend aus (2×2)-Matrizen. Im Beweis von Satz 4.13 wurde bereits gezeigt, dass ein $\delta > 0$ existiert, sodass die Funktion $\ddot{\psi}_{\boldsymbol{\theta}}(x,t)$ für alle $\boldsymbol{\theta} \in (\theta_0 - \delta, \theta_0 + \delta) \times [\vartheta_0, \vartheta_0 + \delta)$ von einer integrierbaren Funktion $\ddot{\psi}(x,t)$ dominiert wird. Der Spezialfall $\vartheta_0 = 0$ wird in Lemma 3 in [65] behandelt. Aufgrund der in Satz 4.11 gezeigten Konsistenz geht die Wahrscheinlichkeit von $\{\hat{\boldsymbol{\theta}}_n \in (\theta_0 - \delta, \theta_0 + \delta) \times [\vartheta_0, \vartheta_0 + \delta)\}$ gegen eins. Auf dieser Menge gilt nun

$$\|\ddot{\Psi}_n(\tilde{\boldsymbol{\theta}}_n)\| = \left\| \frac{1}{n} \sum_{i=1}^{n} \ddot{\psi}_{\tilde{\boldsymbol{\theta}}_n}(X_i, T_i) \right\| \leq \frac{1}{n} \sum_{i=1}^{n} \|\ddot{\psi}(X_i, T_i)\|.$$

Letzteres ist aufgrund der Integrierbarkeit der Funktion $\ddot{\psi}(x,t)$ und dem Gesetz der großen Zahlen in Wahrscheinlichkeit beschränkt. Der zweite und dritte Term aus (5.1)

lassen sich damit zu

$$\begin{aligned}&\left(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1,T_1)] + o_{\mathbb{P}_{\theta_0}}(1) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T O_{\mathbb{P}_{\theta_0}}(1)\right)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\ &= \left(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1,T_1)] + o_{\mathbb{P}_{\theta_0}}(1)\right)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\end{aligned}$$

umschreiben. In diesem Fall stehen $o_{\mathbb{P}_{\theta_0}}(1)$ und $O_{\mathbb{P}_{\theta_0}}(1)$ für eine (2×2)-Matrix und einen zweidimensionalen Vektor aus (2×2)-Matrizen, deren Einträge aus Folgen bestehen, die in Wahrscheinlichkeit gegen null gehen bzw. beschränkt sind. Die Dimension von $o_{\mathbb{P}_{\theta_0}}(1)$ variiert in den folgenden Verwendungen, worauf jedoch nicht genauer eingegangen wird. Die rechte Seite der Gleichung folgt mit der Konsistenz, sodass $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T O_{\mathbb{P}_{\theta_0}}(1) = o_{\mathbb{P}_{\theta_0}}(1)^{\mathrm{T}} O_{\mathbb{P}_{\theta_0}}(1) = o_{\mathbb{P}_{\theta_0}}(1)$. Durch Kombination der Aussagen in Satz 4.5 und Satz 4.4 wird die Matrix $\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1,T_1)] + o_{\mathbb{P}_{\theta_0}}(1)$ für große n invertierbar.
Für den Fall $\hat{\vartheta}_n > 0$ erhält man somit

$$\begin{aligned}0 &= \Psi_n(\boldsymbol{\theta}_0) + \left(\mathbb{E}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1,T_1)] + o_{\mathbb{P}_{\theta_0}}(1)\right)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) &&\Leftrightarrow\\ \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 &= -\left(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1,T_1)] + o_{\mathbb{P}_{\theta_0}}(1)\right)^{-1}\Psi_n(\boldsymbol{\theta}_0) &&\Leftrightarrow\\ \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) &= -(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}_0}(X_1,T_1)])^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\boldsymbol{\theta}_0}(X_i,T_i) + o_{\mathbb{P}_{\theta_0}}(1),\end{aligned}$$

sodass $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ bedingt auf $\hat{\vartheta}_n > 0$ gegen eine zweidimensionale Verteilung $F_1^{\boldsymbol{\theta}_0}(a_\theta, a_\vartheta)$ konvergiert, deren Dichte für $a_\vartheta \leq 0$ gleich null ist und für $a_\vartheta > 0$ zweimal der Dichtefunktion der Normalverteilung $\mathcal{N}_2(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1})$ entspricht. Der Faktor zwei ergibt sich dabei wie folgt: Ist $(Z_1, Z_2)^{\mathrm{T}}$ ein normalverteilter Zufallsvektor mit dem Erwartungswert $\mathbf{0}$ und der Dichtefunktion φ, dann gilt

$$\mathbb{P}(Z_1 \leq z_1, Z_2 \leq z_2 | Z_2 > 0) = \frac{\mathbb{P}(Z_1 \leq z_1, 0 < Z_2 \leq z_2)}{\mathbb{P}(Z_2 > 0)} = \int_{-\infty}^{z_1}\int_{0}^{z_2} 2\varphi(\tilde{z}_1, \tilde{z}_2)\,\mathrm{d}\tilde{z}_2\,\mathrm{d}\tilde{z}_2$$

wegen $\mathbb{P}(Z_2 > 0) = 1/2$.
Wie in Satz 1 in [41] erhält man für den Fall $\hat{\vartheta}_n = 0$ die Vereinfachung

$$\begin{aligned}\Psi_n(\hat{\boldsymbol{\theta}}_n) &= \Psi_n(\boldsymbol{\theta}_0) + \left(\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\psi}_{\boldsymbol{\theta}}(X_1,T_1)] + o_{\mathbb{P}_{\theta_0}}(1)\right)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\\ &= \begin{pmatrix}\Psi_{n,1}(\boldsymbol{\theta}_0)\\ \Psi_{n,2}(\boldsymbol{\theta}_0)\end{pmatrix} - \begin{pmatrix}\mathcal{I}(\boldsymbol{\theta}_0)_{11} & \mathcal{I}(\boldsymbol{\theta}_0)_{12}\\ \mathcal{I}(\boldsymbol{\theta}_0)_{21} & \mathcal{I}(\boldsymbol{\theta}_0)_{22}\end{pmatrix}\begin{pmatrix}\hat{\theta}_n - \theta_0\\ 0\end{pmatrix} + \begin{pmatrix}o_{\mathbb{P}_{\theta_0}}(1)\\ o_{\mathbb{P}_{\theta_0}}(1)\end{pmatrix}\\ &= \begin{pmatrix}\Psi_{n,1}(\boldsymbol{\theta}_0)\\ \Psi_{n,2}(\boldsymbol{\theta}_0)\end{pmatrix} - \begin{pmatrix}\mathcal{I}(\boldsymbol{\theta}_0)_{11}(\hat{\theta}_n - \theta_0)\\ \mathcal{I}(\boldsymbol{\theta}_0)_{21}(\hat{\theta}_n - \theta_0)\end{pmatrix} + \begin{pmatrix}o_{\mathbb{P}_{\theta_0}}(1)\\ o_{\mathbb{P}_{\theta_0}}(1)\end{pmatrix}.\end{aligned}$$

Mit $\Psi_{n,1}(\hat{\boldsymbol{\theta}}_n) = 0$ und $\Psi_{n,2}(\hat{\boldsymbol{\theta}}_n) \leq 0$ folgt somit

$$\begin{aligned} 0 &= \Psi_{n,1}(\boldsymbol{\theta}_0) - \mathcal{I}(\boldsymbol{\theta}_0)_{11}(\hat{\theta}_n - \theta_0) + o_{\mathbb{P}_{\theta_0}}(1) \qquad \Leftrightarrow \\ \hat{\theta}_n - \theta_0 &= (\mathcal{I}(\boldsymbol{\theta}_0)_{11})^{-1}\Psi_{n,1}(\boldsymbol{\theta}_0) + o_{\mathbb{P}_{\theta_0}}(1) \end{aligned} \tag{5.2}$$

und

$$0 \geq \Psi_{n,2}(\boldsymbol{\theta}_0) - \mathcal{I}(\boldsymbol{\theta}_0)_{21}(\hat{\theta}_n - \theta_0) + o_{\mathbb{P}_{\theta_0}}(1).$$

Das Einsetzen von Gleichung (5.2) führt weiterhin zu

$$0 \geq \Psi_{n,2}(\boldsymbol{\theta}_0) - \mathcal{I}(\boldsymbol{\theta}_0)_{21}(\mathcal{I}(\boldsymbol{\theta}_0)_{11})^{-1}\Psi_{n,1}(\boldsymbol{\theta}_0) + o_{\mathbb{P}_{\theta_0}}(1).$$

Somit ist $\sqrt{n}(\hat{\theta}_n - \theta_0)$ unter der Bedingung $Y_2 - \mathcal{I}(\boldsymbol{\theta}_0)_{21}(\mathcal{I}(\boldsymbol{\theta}_0)_{11})^{-1}Y_1 \leq 0$ asymptotisch normalverteilt mit dem Erwartungswert 0 und der Varianz $(\mathcal{I}(\boldsymbol{\theta}_0)_{11})^{-1}$, wobei $(Y_1, Y_2) \sim \mathcal{N}_2(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0))$.
Es bezeichne nun A die Menge aller $\omega \in \Omega$, auf der die Ungleichung $Y_2 - \mathcal{I}(\boldsymbol{\theta}_0)_{21}(\mathcal{I}(\boldsymbol{\theta}_0)_{11})^{-1}Y_1 \leq 0$ erfüllt ist. Die Linearkombination aus den zwei normalverteilten Zufallsvariablen Y_1 und Y_2 ist ebenfalls normalverteilt mit dem Erwartungswert null. Aus Symmetriegründen gilt daher $\mathbb{P}(A) = 1/2$. Mit dem Gesetz der totalen Wahrscheinlichkeit folgt somit für die asymptotische Verteilung $\Phi_{\boldsymbol{\theta}_0}$ von $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$, dass

$$\Phi_{\boldsymbol{\theta}_0} = F_1^{\boldsymbol{\theta}_0} \cdot (1 - \mathbb{P}(A)) + F_2^{\boldsymbol{\theta}_0} \cdot \mathbb{P}(A) = \frac{1}{2}F_1^{\boldsymbol{\theta}_0} + \frac{1}{2}F_2^{\boldsymbol{\theta}_0}.$$

■

5.2. Testdurchführung und Untersuchung der Teststärke

Unter Verwendung von Satz 5.1 kann der Test zur Überprüfung der Unabhängigkeit nun formuliert werden. Die Null- und Alternativhypothese lauten

$$\begin{aligned} H_0 &: \vartheta_0 = 0 \\ H_1 &: \vartheta_0 > 0. \end{aligned}$$

Nach Satz 5.1 besitzt die Folge $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \sqrt{n}(\hat{\theta}_n - \theta_0, \hat{\vartheta}_n)^{\mathrm{T}}$ unter H_0 die asymptotische Verteilung $\Phi_{\boldsymbol{\theta}_0}(\mathbf{a}) = \frac{1}{2}F_1^{\boldsymbol{\theta}_0}(\mathbf{a}) + \frac{1}{2}F_2^{\boldsymbol{\theta}_0}(\mathbf{a})$. Als asymptotische Randverteilung für $\sqrt{n}\hat{\vartheta}_n$ erhält man

$$\Phi_{\boldsymbol{\theta}_0}(\infty, a_\vartheta) = \frac{1}{2} + \int_0^{a_\vartheta} \frac{1}{\sqrt{2\pi\sigma_\vartheta^2}} e^{-\frac{1}{2}\left(\frac{x}{\sigma_\vartheta}\right)^2} \,\mathrm{d}x$$

für $a_\vartheta \geq 0$ mit $\sigma_\vartheta^2 = (\mathcal{I}(\boldsymbol{\theta}_0)^{-1})_{22}$, da $F_1^{\boldsymbol{\theta}_0}$ auf $(-\infty, \infty) \times (0, \infty)$ als Dichte zweimal die Dichtefunktion der zweidimensionalen Normalverteilung $\mathcal{N}_2(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1})$ besitzt und

$F_2^{\boldsymbol{\theta}_0}(\infty, 0) = 1$ gilt. Die durch Standardisierung erhaltene Prüfgröße

$$T_n := \frac{\sqrt{n}\hat{\vartheta}_n}{\sigma_\vartheta} \tag{5.3}$$

besitzt damit unter H_0 die asymptotische Verteilung

$$\Phi_{\boldsymbol{\theta}_0}(\infty, a_\vartheta \sigma_\vartheta) = \frac{1}{2} + \int_0^{a_\vartheta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, \mathrm{d}x, \quad a_\vartheta \geq 0.$$

Mit einer Wahrscheinlichkeit von 50 Prozent wird somit der Wert null angenommen. Bei einem Fehler der 1. Art von α beträgt die Wahrscheinlichkeit dafür, dass die Nullhypothese abgelehnt wird, obwohl diese wahr ist, $\alpha \cdot 100$ Prozent. Es bezeichne $z_{1-\alpha}$ das $(1-\alpha)$-Quantil der Standardnormalverteilung, dann wird die Nullhypothese für $T_n > z_{1-\alpha}$ abgelehnt. Der Fehler 2. Art, β, gibt die Wahrscheinlichkeit dafür an, dass die Nullhypothese nicht abgelehnt wird, wobei die Alternativhypothese zutrifft. Dieser und somit die Güte des Tests, $1-\beta$, kann für eine einfache Alternativhypothese der Form $H_1 : \vartheta_0 = \vartheta^1 > 0$ berechnet werden. Unter H_1 gilt

$$1 - \beta = \mathbb{P}(T_n > z_{1-\alpha}) = 1 - \mathbb{P}\left(\frac{\hat{\vartheta}_n - \vartheta^1}{\sigma_\vartheta / \sqrt{n}} < z_{1-\alpha} - \frac{\vartheta^1}{\sigma_\vartheta / \sqrt{n}} \right).$$

Nach Satz 4.13 ist $\sqrt{n}(\hat{\vartheta}_n - \vartheta^1)/\sigma$ unter H_1 asymptotisch standardnormalverteilt und β kann durch

$$\Phi\left(z_{1-\alpha} - \frac{\sqrt{n}\vartheta^1}{\sigma_\vartheta} \right)$$

mit Φ als Verteilungsfunktion der Standardnormalverteilung approximiert werden. Wegen $-\sqrt{n}\vartheta^1/\sigma_\vartheta \to -\infty$ für $n \to \infty$ und Satz 4.13 folgt unter H_1 die Konsistenz der Testfolge

$$\mathbb{P}(T_n > z_{1-\alpha}) \to 1 \quad (n \to \infty).$$

Mit größer werdendem Stichprobenumfang konvergiert die Güte der Testfolgen gegen eins, sodass die Wahrscheinlichkeit für einen Fehler 2. Art verschwindet. Am Ende von Kapitel 8 wird die Testgüte mithilfe von Simulationen näherungsweise ermittelt. Die Ergebnisse sind in Abbildung 8.3 dargestellt.

Um den Test auf empirische Daten anwenden zu können, wird eine Schätzung von $\mathcal{I}(\boldsymbol{\theta}_0)$ unter H_0 benötigt. Dafür bezeichne $\hat{\boldsymbol{\theta}}_n^0 = (\hat{\theta}_n^0, 0)^{\mathrm{T}}$ den unter H_0 eingeschränkten Schätzer. Ähnlich wie in Bemerkung 17.4 in [21] kann $\mathcal{I}(\boldsymbol{\theta}_0)$ wegen des Gesetzes der großen Zahlen und wegen des Satzes über stetige Abbildungen (vgl. Satz 2.3 in [61]) durch

$$\frac{1}{n}\sum_{i=1}^{n} \psi_{\hat{\boldsymbol{\theta}}_n^0}(X_i, T_i)\psi_{\hat{\boldsymbol{\theta}}_n^0}(X_i, T_i)^{\mathrm{T}} = \frac{1}{n}\sum_{j=1}^{M_n} \psi_{\hat{\boldsymbol{\theta}}_n^0}(\widetilde{X}_j, \widetilde{T}_j)\psi_{\hat{\boldsymbol{\theta}}_n^0}(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}} \tag{5.4}$$

angenähert werden. Die Konsistenz von $\hat{\boldsymbol{\theta}}_n^0$ folgt dabei aus Satz 1 aus dem Aufsatz von

Weißbach und Wied [65]. Damit lässt sich die Varianz $\sigma_\vartheta^2 = (\mathcal{I}(\boldsymbol{\theta}_0)^{-1})_{22}$ näherungsweise bestimmen. Durch Multiplikation mit n^{-1} ist diese nicht mehr von n abhängig. Das Lemma von Slutsky (vgl. Lemma 2.8 in [61]) stellt sicher, dass die asymptotische Verteilung der Folge $\sqrt{n}\hat{\vartheta}_n/\sigma_\vartheta$ durch die beschriebene Annäherung von σ_ϑ nicht beeinflusst wird. Eine Schätzung des latenten Stichprobenumfangs ist nicht notwendig, da die Teststatistik nach der Annäherung von σ_ϑ nicht mehr von n abhängt.

Auch die Varianz σ_θ^2 lässt sich mithilfe von (5.4) näherungsweise bestimmen. Die asymptotische Verteilung von $\sqrt{n}(\hat{\theta}_n - \theta_0)$ unter H_0 erhält man als Randverteilung von Φ_{θ_0} aus Satz 5.1. Es bezeichne $\varphi_{\mathcal{I}(\boldsymbol{\theta}_0)^{-1}}$ die Dichtefunktion einer normalverteilten Zufallsvariable mit dem Erwartungswertvektor $\mathbf{0}$ und der Kovarianzmatrix $\mathcal{I}(\boldsymbol{\theta}_0)^{-1}$, dann berechnet sich $\Phi_{\theta_0}(a_\theta, \infty)$ durch

$$\int_{-\infty}^{a_\theta} \int_0^\infty \varphi_{\mathcal{I}(\boldsymbol{\theta}_0)^{-1}}(y_1, y_2)\,\mathrm{d}y_2\,\mathrm{d}y_1 + \frac{1}{2}\mathbb{P}\left(Y_1 \leq a_\theta/\sigma^{(2)} \middle| Y_2 - \mathcal{I}(\boldsymbol{\theta}_0)_{12} \cdot \sigma^{(2)} Y_1 \leq 0\right).$$

Aus der Reproduktionseigenschaft der Normalverteilung folgt

$$\mathbb{P}\left(Y_2 - \mathcal{I}(\boldsymbol{\theta}_0)_{12} \cdot \sigma^{(2)} Y_1 \leq 0\right) = 1/2.$$

Der zweite Summand lässt sich somit zu $\mathbb{P}\left(Y_1 \leq a_\theta/\sigma^{(2)}, Y_2 \leq \mathcal{I}(\boldsymbol{\theta}_0)_{12} \cdot \sigma^{(2)} Y_1\right)$ vereinfachen. Mit $(Y_1, Y_2) \sim \mathcal{N}_2(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0))$ und mithilfe von Substitution erhält man schließlich

$$\int_{-\infty}^{a_\theta} \left[\int_0^\infty \varphi_{\mathcal{I}(\boldsymbol{\theta}_0)^{-1}}(y_1, y_2)\,\mathrm{d}y_2 + \int_{-\infty}^{\mathcal{I}(\boldsymbol{\theta}_0)_{12} \cdot y_1} \frac{1}{\sigma^{(2)}} \varphi_{\mathcal{I}(\boldsymbol{\theta}_0)}(y_1/\sigma^{(2)}, y_2)\,\mathrm{d}y_2 \right] \mathrm{d}y_1 \tag{5.5}$$

als asymptotische Verteilung von $\sqrt{n}(\hat{\theta}_n - \theta_0)$, welche zur Berechnung des Standardfehlers verwendet werden kann. Bei der Schätzung von $\mathcal{I}(\boldsymbol{\theta}_0)$ ist jedoch zu beachten, dass auch n durch $M_n/\alpha_{\hat{\theta}_n^0}$ geschätzt werden muss. Die Approximation der Verteilung kann dabei mit dem Lemma von Slutsky und dem Satz über stetige Abbildungen begründet werden.

Bemerkung 5.2.
Der Unabhängigkeitstest kann in dieser Form auch für die zur Spiegelung assoziierten Version der Gumbel-Barnett-Copula C_ϑ^{270} aus Annahme (A3'') durchgeführt werden. Lediglich die Score-Funktion zur Berechnung der Fisher-Information muss im Vergleich zum oben beschriebenen Unabhängigkeitstest entsprechend angepasst werden.
Der Parameter der Farlie-Gumbel-Morgenstern-Copula C_ϑ^{FGM} aus Annahme (A3') kann Werte aus dem Intervall $[-1, 1]$ annehmen. Bei Unabhängigkeit liegt der Parameter mit $\vartheta = 0$ somit im Inneren des Intervalls und nicht auf dem Rand. Zur Ermittlung der asymptotischen Verteilung kann Satz 4.13 verwendet werden. Die Fisher-Information kann wie in Gleichung (5.4) beschrieben angenähert werden. Der wahre Parameter der Exponentialverteilung wird durch dessen Schätzung unter H_0 ersetzt. Durch die Berechnung des Erwartungswertes als arithmetisches Mittel sind die durch $\sqrt{n}$ dividierten Standardfehler nicht mehr vom latenten Stichprobenumfang n abhängig. □

6. Kolmogorov-Smirnov-Anpassungstest

Während sich der erste Teil der Arbeit vor allem mit der Abhängigkeitsstruktur zwischen der interessierenden und der Trunkationsvariablen beschäftigt, soll im zweiten Teil auch die Wahl der Randverteilungen kritisch hinterfragt werden. Ein Anpassungstest ermöglicht die Beurteilung, ob die empirische Verteilungsfunktion einer gegebenen Stichprobe mit einer theoretischen Verteilung übereinstimmt. Unter Verwendung des trunkierten Punktprozesses kann ein zweidimensionaler Kolmogorov-Smirnov-ähnlicher Test für doppeltrunkierte Daten konstruiert werden. Ziel dieses Kapitels ist es, die theoretischen Grundlagen des Tests zu entwickeln. Die Verteilung der Teststatistik sowie die Berechnung der Prüfgröße werden in separaten Abschnitten behandelt. Eine Auseinandersetzung mit der asymptotischen Verteilung der Teststatistik erfordert zunächst eine umfassende Betrachtung der bestehenden Konvergenzsätze empirischer Prozesse. Die praktische Anwendbarkeit anhand von Daten und Simulationen wird in eigenen Kapiteln 7 und 8 evaluiert.

6.1. Konvergenzsätze für empirische Prozesse

Bevor der trunkierte Punktprozess betrachtet wird, erfolgt zunächst eine Einleitung mit bekannten Resultaten über empirische Prozesse, welche den Kapiteln 19.1 und 19.3 aus van der Vaart [61] entnommen worden sind.
Es seien $Y_1, \ldots, Y_n$ unabhängig identisch verteilte Zufallsvariablen mit der Verteilungsfunktion G. Dann gilt nach dem *starken Gesetz der großen Zahlen*, dass die empirische Verteilungsfunktion punktweise fast sicher gegen G konvergiert, d.h. für alle $y \in \mathbb{R}$ gilt

$$G_n(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, y]}(Y_i) \to G(y) \quad \text{für} \quad n \to \infty.$$

Nach dem *zentralen Grenzwertsatz* erhält man mit $\mathbb{E}[G_n(y)] = \mathbb{P}(Y_1 \leq y) = G(y)$ sowie mit $\mathrm{Var}[G_n(y)] = G(y)(1 - G(y))/n$ die asymptotische Normalität

$$\frac{\sqrt{n}(G_n(y) - G(y))}{\sqrt{G(y)\,(1 - G(y))}} \to \mathcal{N}(0, 1) \quad \text{für} \quad n \to \infty.$$

Der *Satz von Glivenko-Cantelli* erweitert das Gesetz der großen Zahlen auf gleichmäßige Konvergenz, sodass für $n \to \infty$

$$\|G_n(y) - G(y)\|_\infty := \sup_{y \in \mathbb{R}} |G_n(y) - G(y)| \to 0 \quad \text{fast sicher.}$$

Auch für den zentralen Grenzwertsatz gibt es eine gleichmäßige bzw. funktionale Version. Diese ist als funktionaler zentraler Grenzwertsatz oder auch als *Satz von Donsker* bekannt. Der stochastische Prozess wird dabei als Element des Skorohod-Raums $D[-\infty, \infty]$ betrachtet. Dieser enthält alle rechtsseitig stetigen Funktionen mit linksseitigen Limiten (auch càdlàg als Akronym für continue à droite, limites à gauche), d.h.

$$D[-\infty, \infty] := \{g : \mathbb{R} \to \mathbb{R} | \lim_{s \to t^+} g(s) = g(t) \quad \text{und} \\ \lim_{s \to t^-} g(s) \text{ existiert für alle } t \in \mathbb{R}\}.$$

Versieht man den Raum mit der von der Supremumsnorm $\|\cdot\|_\infty$ induzierten Metrik, so ist $(D[-\infty, \infty], \|\cdot\|_\infty)$ nicht separabel (vergleiche Kapitel 15 in Billingsley [3]). Als Folge daraus ist die σ-Algebra, die von den, bezüglich $\|\cdot\|_\infty$, offenen Kugeln erzeugt wird, eine echte Teilmenge der Borel-σ-Algebra. Zwar ist $\sqrt{n}(G_n(y) - G(y))$ nicht messbar bezüglich der Borel-σ-Algebra, jedoch liegt bei Stetigkeit der Verteilungsfunktion G Messbarkeit bezüglich der von den offenen Kugeln erzeugten σ-Algebra vor, sodass die Konvergenz im Satz von Donsker auch in $(D[-\infty, \infty], \|\cdot\|_\infty)$ formuliert werden kann: Es sei G stetig, dann konvergiert die Folge der empirischen Prozesse $\sqrt{n}(G_n - G)$ in Verteilung in $(D[-\infty, \infty], \|\cdot\|_\infty)$ gegen einen Gaußprozess $\mathbb{G}$ mit dem Erwartungswert $\mathbb{E}[\mathbb{G}(x)] = 0$ und der Kovarianz

$$\text{Cov}[\mathbb{G}(x), \mathbb{G}(y)] = \mathbb{E}[\mathbb{G}(x)\mathbb{G}(y)] = \min\{G(x), G(y)\} - G(x)G(y).$$

Der Prozess $\mathbb{G}(y)$ kann auch als $\mathbb{B}(G(y))$ geschrieben werden, wobei $\mathbb{B}(t)$ eine *Brownsche Brücke* auf dem Intervall $[0, 1]$ darstellt. Diese ist nach Shreve [53] (Definition 4.7.4) wie folgt definiert wird, wobei der Begriff *Wiener Prozess* als Synonym für *Brownsche Bewegung* verwendet wird:

Definition 6.1. (Brownsche Brücke)
Es sei $\mathbb{W}(t)$ ein Wiener Prozess. Die *Brownsche Brücke* wird für $0 \leq t \leq 1$ definiert als

$$\mathbb{B}(t) := \mathbb{W}(t) - t\mathbb{W}(1).$$

□

Bemerkung 6.2.
Die Pfade der Brownschen Brücke beginnen und enden bei null, d.h. $\mathbb{B}(0) = \mathbb{B}(1) = 0$, woraus sich der Begriff „Brücke“ herleitet. Aus den Eigenschaften des Wiener Prozesses lassen sich die Erwartungswert- und Kovarianzfunktion ermitteln. Diese lauten

$$\mathbb{E}[\mathbb{B}(t)] = 0 \qquad \text{sowie} \quad \text{Cov}[\mathbb{B}(s), \mathbb{B}(t)] = \min\{s, t\} - s \cdot t, \quad \text{für} \quad 0 \leq s, t \leq 1.$$

In Definition 4.7.4 in [53] wird die Brownsche Brücke auf einem abgeschlossenen Intervall mit beliebiger, aber fester Länge definiert. In dieser Arbeit wird jedoch lediglich die Version auf dem Einheitsintervall benötigt. □

Eine wichtige Anwendung finden die empirischen Verteilungsfunktionen bei den Anpassungstests. Ein bekannter Anpassungstest ist der Kolmogorov-Smirnov-Test, mit

$$\sqrt{n}\|G_n - G\|_\infty$$

als Prüfgröße. Die Abbildung $(D[-\infty,\infty], \|\cdot\|_\infty) \to (\mathbb{R}, |\cdot|)$ mit $g \mapsto \|g\|_\infty$ ist stetig. Mit dem Satz über stetige Abbildungen (vgl. Satz 18.11 in [61] als Version für metrische Räume) sowie dem Satz von Donsker konvergiert daher $\sqrt{n}\|G_n - G\|_\infty$ in Verteilung gegen $\|\mathbb{B}(G)\|_\infty$. Die Verteilung von $\|\mathbb{B}(G)\|_\infty$ kann aus den Eigenschaften der Brownschen Brücke hergeleitet werden und hängt, unter Voraussetzung der Stetigkeit von G, nicht von der zugrunde liegenden Verteilung G ab (vgl. Korollar 19.21 van der Vaart [61]). Sie ist auch als Kolmogorov-Verteilung bekannt und lautet

$$\Pr(\|\mathbb{B}\|_\infty > t)^1 = 2\sum_{k=1}^{\infty}(-1)^{k-1}e^{-2k^2t^2}. \tag{6.1}$$

Das starke Gesetz der großen Zahlen sowie der zentrale Grenzwertsatz können direkt auf den in Abschnitt 3.2 definierten trunkierten Prozess angewendet werden. Die weiteren drei Sätze sind zunächst nur für empirische Verteilungsfunktionen formuliert und müssen entsprechend verallgemeinert werden. Die folgenden Definitionen und Sätze sind sowohl in den Abschnitten 2.1 und 2.4 in van der Vaart und Wellner [62] als auch im Abschnitt 19.2 in van der Vaart [61] zu finden.
Es sei $Y_1, \ldots, Y_n$ eine Zufallsstichprobe der Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf einem messbaren Raum $(\mathcal{Y}, \mathcal{A})$. Unter Messbarkeit ist die *empirische Verteilung* das zufällige Maß $\mathbb{P}_n := n^{-1}\sum_{i=1}^n \epsilon_{Y_i}$. Für eine Klasse $\mathcal{G}$ messbarer Funktionen $g : \mathcal{Y} \to \mathbb{R}$ bezeichne $\mathbb{P}_n(g)$ den Erwartungswert von g unter $\mathbb{P}_n$, d.h.

$$g \mapsto \mathbb{P}_n(g) := \frac{1}{n}\sum_{i=1}^n g(Y_i).$$

Weiterhin sei $\mathbb{E}(g) := \mathbb{E}(g(Y_1)) = \int g(Y_1)\,\mathrm{d}\mathbb{P}$. Die zentrierte und skalierte Version der Abbildung $g \mapsto \mathbb{P}_n(g)$ ist der in g ausgewertete *empirische Prozess*

$$g \mapsto \mathbb{G}_n(g) := \sqrt{n}(\mathbb{P}_n - \mathbb{E})(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (g(Y_i) - \mathbb{E}(g)).$$

Für Funktionen aus der Funktionenklasse $\mathcal{G} = \{g_t = \mathbb{1}_{(-\infty,t]},\, t \in \mathbb{R}\}$ stimmt der empirische Prozess mit der skalierten Differenz aus der empirischen und der tatsächlichen

[1] Zur Abgrenzung des Wahrscheinlichkeitsmaßes $\mathbb{P}$ von Y_i wird an dieser Stelle einmalig das Symbol Pr verwendet.

Verteilungsfunktion überein:

$$\mathbb{G}_n(g_t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\mathbb{1}_{(-\infty,t]}(Y_i) - \mathbb{E}(\mathbb{1}_{(-\infty,t]}(Y_1)) \Big) = \sqrt{n}(G_n(t) - G(t))$$

Wie bereits oben erwähnt, lässt sich fehlende Messbarkeit durch die Wahl einer geeigneten Metrik oder σ-Algebra wiederherstellen. Eine andere Möglichkeit ist es, die Supremumsnorm sowie die Borel-σ-Algebra beizubehalten und die Erwartungswerte durch sogenannte *äußere Erwartungswerte* zu ersetzen. Die Definitionen der Konvergenzarten können dementsprechend angepasst werden und auch für nicht messbare Abbildungen formuliert werden. Borel-Messbarkeit wird dann nur noch für den jeweiligen Grenzprozess gefordert. Die folgenden beiden Sätze machen von dieser Vorgehensweise Gebrauch, wobei darauf in der Notation nicht weiter eingegangen wird. Für Details siehe Abschnitt 18.2 und 19.2 in van der Vaart [61] sowie Seite 4 ff. in van der Vaart und Wellner [62].

Definition 6.3. (Glivenko-Cantelli-Klasse)
Eine Klasse $\mathcal{G}$ messbarer Funktionen $g : \mathcal{Y} \to \mathbb{R}$ wird als *Glivenko-Cantelli-Klasse* bezeichnet, falls

$$\|\mathbb{P}_n(g) - \mathbb{E}(g)\|_{\mathcal{G}} := \sup_{g \in \mathcal{G}} |\mathbb{P}_n(g) - \mathbb{E}(g)| \to 0 \quad \text{fast sicher für } n \to \infty.$$

□

Um die funktionale Version des zentrales Grenzwertsatzes zu diskutieren, wird

$$\sup_{g \in \mathcal{G}} |g(y) - \mathbb{E}(g)| < \infty$$

für alle $y \in \mathcal{Y}$ angenommen. Unter dieser Bedingung kann der Prozess $\{\mathbb{G}_n(g) : g \in \mathcal{G}\}$ als Abbildung nach $\ell^\infty(\mathcal{G})$ betrachtet werden. Hierbei bezeichnet $\ell^\infty(\mathcal{G})$ den Raum aller bezüglich der Norm $\|\cdot\|_{\mathcal{G}}$ gleichmäßig beschränkten Funktionen von $\mathcal{G}$ nach $\mathbb{R}$. Darüber hinaus ist eine Erweiterung der Definition 6.1 der Brownschen Brücke auf Funktionen erforderlich. Van der Vaart [61] (Abschnitt 19.2) und van der Vaart und Wellner [62] (Abschnitt 2.1) führen dazu die sogenannte $\mathbb{P}$-Brownsche Brücke ein.

Definition 6.4. ($\mathbb{P}$-Brownsche Brücke)
Ein Gauß-Prozess $\{\mathbb{G}(g) : g \in \mathcal{G}\}$ wird als $\mathbb{P}$-*Brownsche Brücke* bezeichnet, falls für die Erwartungswert- und Kovarianzfunktion für $g, h \in \mathcal{G}$ folgendes gilt:

$$\mathbb{E}[\mathbb{G}(g)] = 0 \quad \text{und} \quad \mathrm{Cov}[\mathbb{G}(g), \mathbb{G}(h)] = \mathbb{E}[\mathbb{G}(g)\mathbb{G}(h)] = \mathbb{E}(gh) - \mathbb{E}(g)\mathbb{E}(h)$$

Das Wahrscheinlichkeitsmaß $\mathbb{P}$ bezieht sich dabei auf den Erwartungswert, d.h. $\mathbb{E}(g) = \int g(Y_1)\,\mathrm{d}\mathbb{P}$. Der Prozess $\mathbb{G}(g)$ geht gemäß der Transformation $\mathbb{B}(\mathbb{E}(g))$[2] aus der Brownschen Brücke aus Definition 6.1 hervor. Zur besseren Lesbarkeit wird in dieser Arbeit die Notation $\mathbb{B}_{\mathbb{P}}(g) := \mathbb{B}(\mathbb{E}(g))$ verwendet. □

[2] In dieser Arbeit werden vor allem Funktionen g mit $\mathbb{E}(g) \in [0, 1]$ verwendet.

Beispiel 6.5. (Brownsche Brücke auf $[0,1]$)
Für die Funktionenklasse $\mathcal{G} = \{g_t = \mathbb{1}_{[0,t]},\, t \in [0,1]\}$ und $\mathbb{P}$ als Lebesgue-Maß (Gleichverteilung) auf dem Intervall $[0,1]$ stimmt die $\mathbb{P}$-Brownsche Brücke mit der Brownschen Brücke aus Definition 6.1 überein. Für die Kovarianzfunktion gilt

$$\begin{aligned}\mathrm{Cov}[\mathbb{G}(g_s), \mathbb{G}(g_t)] &= \mathbb{E}(g_s \cdot g_t) - \mathbb{E}(g_s)\mathbb{E}(g_t) \\ &= \mathbb{P}(Y_1 \in [0, \min\{s,t\}]) - \mathbb{P}(Y_1 \in [0,s])\mathbb{P}(Y_1 \in [0,t]) \\ &= \min\{s,t\} - s \cdot t.\end{aligned}$$ □

Definition 6.6. (Donsker-Klasse)
Eine Klasse $\mathcal{G}$ messbarer Funktionen $g : \mathcal{Y} \to \mathbb{R}$ wird als *Donsker-Klasse* bezeichnet, falls die Folge $\{\mathbb{G}_n(g) : g \in \mathcal{G}\}$ in Verteilung gegen einen straffen Borel-messbaren Grenzprozess $\mathbb{G}$ in $\ell^\infty(\mathcal{G})$ konvergiert. Der Grenzprozess ist dann ein Gauß-Prozess mit dem Erwartungswert $\mathbb{E}[\mathbb{G}(g)] = 0$ und der Kovarianzfunktion

$$\mathrm{Cov}[\mathbb{G}(g), \mathbb{G}(h)] = \mathbb{E}[\mathbb{G}(g)\mathbb{G}(h)] = \mathbb{E}(gh) - \mathbb{E}(g)\mathbb{E}(h)$$

und damit insbesondere eine $\mathbb{P}$-Brownsche Brücke $\mathbb{B}_{\mathbb{P}}(g)$ (vgl. Definition 6.4). □

Die im Satz verwendeten Erwartungswerte sind jeweils im Sinne der zugehörigen Wahrscheinlichkeitsmaße zu verstehen. Zur Vereinfachung der Notation wird für den Erwartungswert der Gauß-Prozesse kein separates Symbol eingeführt.
Ein Borel-Wahrscheinlichkeitsmaß $\mathbb{P}$ heißt *straff*, falls für jedes $\varepsilon > 0$ eine kompakte Menge K existiert, sodass $\mathbb{P}(K) \geq 1 - \varepsilon$. Eine Borel-messbare Abbildung heißt *straff*, falls das zugehörige Wahrscheinlichkeitsmaß straff ist (vgl. dazu Abschnitt 1.3 in [62]).

Ob eine Klasse von Funktionen nun eine Glivenko-Cantelli- oder Donsker-Klasse ist, hängt von ihrer Mächtigkeit ab, welche mithilfe verschiedener Entropien untersucht werden kann.

Definition 6.7. (Klammerungszahl)
Es seien l und u zwei Funktionen mit endlicher Norm. Dann bezeichnet die *Klammer* $[l,u]$ die Menge aller Funktionen $g \in \mathcal{G}$ mit $l \leq g \leq u$. Eine *ε-Klammer* ist eine Klammer $[l,u]$ mit $\|u - l\| < \varepsilon$. Die Funktionen u und l müssen dabei nicht in $\mathcal{G}$ liegen. Die *Klammerungszahl* $N_{[]}(\varepsilon, \mathcal{G}, \|\cdot\|)$ ist die kleinste Anzahl von ε-Klammern bezüglich der Norm $\|\cdot\|$, die benötigt wird, um $\mathcal{G}$ zu überdecken. Die *Entropie mit Klammerung* ist die logarithmierte Klammerungszahl. □

Die Klammerungszahl wird im Folgenden in erster Linie bezüglich der $L_r(\mathbb{P})$-Norm benötigt. Diese ist für eine Funktion g bezüglich des Wahrscheinlichkeitsmaßes $\mathbb{P}$ für $1 \leq r < \infty$ definiert durch

$$\|g\|_{\mathbb{P},r} := \left(\int |g(Y_1)|^r \, \mathrm{d}\mathbb{P} \right)^{1/r}.$$

Satz 6.8. (Glivenko-Cantelli)
Es sei $\mathcal{G}$ eine Klasse messbarer Funktionen, sodass $N_{[]}(\varepsilon, \mathcal{G}, L_1(\mathbb{P})) < \infty$ für alle $\varepsilon > 0$. Dann ist $\mathcal{G}$ eine Glivenko-Cantelli-Klasse.

Das folgende Beispiel ist in Bezug auf den Konsistenznachweis eines Z-Schätzers in Abschnitt 4.3 von Interesse. Unter Berücksichtigung der gleichen Voraussetzungen des Satzes 4.7 bietet es eine alternative Argumentation, um das gleichmäßige Gesetz der großen Zahlen für die Kriteriumsfunktion zu erhalten.

Beispiel 6.9. (Punktweise kompakte Klasse)
Es sei $\mathcal{G} = \{g_\theta : \theta \in \Theta\}$ eine Klasse messbarer Funktionen auf einer kompakten Menge Θ. Die Abbildung $\theta \mapsto g_\theta(y)$ sei für alle $y \in \mathcal{Y}$ stetig. Weiterhin existiere eine Funktion g mit $\|g\|_{G,1} < \infty$, sodass $|g_\theta(y)| \leq g(y)$ für alle $\theta \in \Theta$ und für alle $y \in \mathcal{Y}$ gilt. Im Folgenden wird gezeigt, dass die Klasse $\mathcal{G}$ bezüglich der $L_1(G)$-Norm eine endliche Klammerungszahl besitzt, sodass $\mathcal{G}$ eine Glivenko-Cantelli-Klasse ist.
Es bezeichne $B \subset \Theta$ eine offene Kugeln und g_B sowie g^B das Infimum bzw. Supremum von g_θ für $\theta \in B$. Als ε-Klammern werden dann die Klammern $[g_B, g^B]$ betrachtet.
Für ein festes $\theta \in \Theta$ bezeichne $\{B_i\}_{i\in\mathbb{N}}$ eine Folge offener Kugeln mit Mittelpunkt θ. Die Folge der Radien gehe gegen null für $i \to \infty$. Aufgrund der Stetigkeit gelten $g_{B_i}(y) \to g_\theta(y)$ und $g^{B_i}(y) \to g_\theta(y)$ punktweise in $y \in \mathcal{Y}$. Somit geht auch der Abstand $g^{B_i}(y) - g_{B_i}(y)$ punktweise gegen null. Wegen $|g^{B_i}(y) - g_{B_i}(y)| \leq 2g(y)$ und $\|g\|_{G,1} < \infty$ folgt mit dem Satz von der majorisierten Konvergenz (vgl. Kap. IV, §5. Satz 5.2 in Elstrodt [13]) auch $\|g^{B_i} - g_{B_i}\|_{G,1} \to 0$.
Für gegebenes $\varepsilon > 0$ existiert demnach für jedes $\theta \in \Theta$ eine offene Kugel B um θ, sodass die Klammer $[g_B, g^B]$ bezüglich der $L_1(G)$-Norm eine Breite kleiner ε aufweist, d.h. $\|g^B - g_B\|_{G,1} < \varepsilon$. Aufgrund der Kompaktheit der Menge Θ besitzt die so konstruierte offene Überdeckung der Menge Θ nach der Überdeckungseigenschaft von Heine und Borel (vgl. Teil 1, §3. Definition 2 und Satz 3 in Forster [17]) eine endliche Teilüberdeckung. Die zugehörigen ε-Klammern überdecken die Klasse $\mathcal{G}$ und die Klammerungszahl $N_{[]}(\varepsilon, \mathcal{G}, L_1(G))$ ist endlich. □

Für viele Funktionenklassen geht die Klammerungszahl für $\varepsilon \to 0$ gegen unendlich. Eine hinreichende Bedingung für eine Funktionenklasse die Donsker-Eigenschaft zu erfüllen ist, dass die Klammerungszahl nicht „zu schnell“ gegen unendlich geht. Die Divergenzgeschwindigkeit kann mithilfe des *Klammerungsintegrals*

$$J_{[]}(\delta, \mathcal{G}, L_2(\mathbb{P})) := \int_0^\delta \sqrt{\log N_{[]}(\varepsilon, \mathcal{G}, L_2(\mathbb{P}))}\, \mathrm{d}\varepsilon$$

gemessen werden. Ist dieses Integral endlich, so ist die Funktionenklasse $\mathcal{G}$ eine Donsker-Klasse (vgl. van der Vaart [61], Satz 19.5 sowie Ossiander [47], Satz 3.1).

Satz 6.10. (Donsker)
Jede Klasse $\mathcal{G}$ von messbaren Funktionen mit $J_{[]}(1, \mathcal{G}, L_2(\mathbb{P})) < \infty$ ist eine Donsker-Klasse.

Beispiel 6.11. (Empirische Verteilungsfunktionen)
Für die Funktionenklasse $\mathcal{G} = \{g_t = \mathbb{1}_{(-\infty,t]}, t \in \mathbb{R}\}$ lautet der empirische Prozess $\mathbb{G}_n(g_t) = \sqrt{n}(G_n(t) - G(t))$. Es lässt sich zeigen, dass diese Funktionenklasse eine endliche Klammerungszahl und ein endliches Klammerungsintegral besitzt, sodass sich die Sätze 6.8 und 6.10 zu den Sätzen von Glivenko-Canteli und Donsker für empirische Verteilungsfunktionen vereinfachen. Es werden dazu die ε-Klammern der Form

$[\mathbb{1}_{(-\infty,t_{i-1}]}, \mathbb{1}_{(-\infty,t_i)}]$ betrachtet und es sei $G(-\infty) = 0$ sowie $G(\infty) = 1$. Es lässt sich zeigen, dass eine endliche Partition $-\infty = t_0 < t_1 < \ldots < t_k = \infty$ existiert, sodass $\varepsilon/2 \leq G(t_i-) - G(t_{i-1}) \leq \varepsilon$ (vgl. van der Vaart [61], Satz 19.1). $L_1(G)$ bezeichne im Folgenden die L_1-Norm bezüglich des zur Verteilungsfunktion G gehörigen Wahrscheinlichkeitsmaßes. Dies ist über den Zusammenhang $G(y) = \mathbb{P}(Y_1 \leq y)$ gegeben. Die $L_1(G)$-Norm der ε-Klammern liegt nun ebenfalls zwischen $\varepsilon/2$ und ε, denn

$$\|\mathbb{1}_{(-\infty,t_i)} - \mathbb{1}_{(-\infty,t_{i-1}]}\|_{G,1} = \int_{\mathbb{R}} |\mathbb{1}_{(-\infty,t_i)}(y) - \mathbb{1}_{(-\infty,t_{i-1}]}(y)| \,\mathrm{d}G(y) = G(t_i-) - G(t_{i-1}).$$

Anschaulich betrachtet wird das Intervall $[0,1]$ dabei in Intervalle der Form $[G(t_{i-1}), G(t_i-)]$ zerlegt, von welchen nun $k \leq 2/\varepsilon$ zur vollständigen Überdeckung benötigt werden. Damit gilt für die Klammerungszahl $N_{[]}(\varepsilon, \mathcal{G}, L_1(G)) \leq 2/\varepsilon$ und $\mathcal{G}$ ist eine Glivenko-Cantelli-Klasse.
Für die Donsker-Eigenschaft muss nun gezeigt werden, dass das Klammerungsintegral mit der $L_2(G)$-Norm endlich ist. Dies lässt sich mit der Endlichkeit von $N_{[]}(\varepsilon, \mathcal{G}, L_1(G))$ über die Beziehung der $L_1(G)$- und der $L_2(G)$-Norm herleiten. Die Differenz der Indikatorfunktionen der ε-Klammern lässt sich wiederum als Indikatorfunktion darstellen, welche durch das Quadrieren nicht verändert wird. Somit stimmen die $L_1(G)$-Norm und die quadrierte $L_2(G)$-Norm der ε-Klammern überein:

$$\begin{aligned}
&\|\mathbb{1}_{(-\infty,t_i)} - \mathbb{1}_{(-\infty,t_{i-1}]}\|_{G,1} = \int_{\mathbb{R}} |\mathbb{1}_{(-\infty,t_i)}(y) - \mathbb{1}_{(-\infty,t_{i-1}]}(y)| \,\mathrm{d}G(y) \\
&= \int_{\mathbb{R}} |\mathbb{1}_{(-\infty,t_i)}(y) - \mathbb{1}_{(-\infty,t_{i-1}]}(y)|^2 \,\mathrm{d}G(y) = \|\mathbb{1}_{(-\infty,t_i)} - \mathbb{1}_{(-\infty,t_{i-1}]}\|_{G,2}^2
\end{aligned}$$

Ist also die $L_1(G)$-Norm der obigen Klammern durch ε beschränkt, so ist die entsprechende $L_2(G)$-Norm durch $\sqrt{\varepsilon}$ beschränkt. Es gelten demzufolge auch $N_{[]}(\sqrt{\varepsilon}, \mathcal{G}, L_2(G)) \leq N_{[]}(\varepsilon, \mathcal{G}, L_1(G)) \leq 2/\varepsilon$ und $N_{[]}(\varepsilon, \mathcal{G}, L_2(G)) \leq 2/\varepsilon^2$. Wegen $\log(3/\varepsilon^2) \geq 1$ für $0 < \varepsilon \leq 1 < \sqrt{3/\exp(1)}$ folgt schließlich

$$\begin{aligned}
J_{[]}(1, \mathcal{G}, L_2(G)) &= \int_0^1 \sqrt{\log N_{[]}(\varepsilon, \mathcal{G}, L_2(G))} \,\mathrm{d}\varepsilon \\
&\leq \int_0^1 \sqrt{\log(2/\varepsilon^2)} \,\mathrm{d}\varepsilon \leq \int_0^1 \sqrt{\log(3/\varepsilon^2)} \,\mathrm{d}\varepsilon \\
&\leq \int_0^1 \log(3/\varepsilon^2) \,\mathrm{d}\varepsilon = 2 + \log 3 < \infty.
\end{aligned}$$

□

Wie aus Beispiel 2.5.4 in [62] hervorgeht, kann für höhere Dimensionen eine ähnliche Abschätzung vorgenommen werden, jedoch mit einer größeren Potenz von $1/\varepsilon$. Die Klasse von Indikatorfunktionen mit Intervallen in $\mathbb{R}^k$ ist somit für jede Dimension eine Donsker-Klasse.
Van der Vaart und Wellner [62] betrachten in Abschnitt 2.10 etliche Operationen, welche die Donsker-Eigenschaft erhalten und es erlauben, neue Donsker-Klassen zu bilden. Das Hauptresultat bilden dabei Lipschitz-Transformationen.

Satz 6.12.
Es seien $\mathcal{G}_1$ und $\mathcal{G}_2$ zwei Donsker-Klassen mit $\|\mathbb{E}\|_{\mathcal{G}_i} < \infty$. Weiter erfülle die Abbildung $\phi : \mathbb{R}^2 \to \mathbb{R}$ die Eigenschaft

$$|\phi(g(y)) - \phi(h(y))|^2 \leq (g_1(y) - h_1(y))^2 + (g_2(y) - h_2(y))^2$$

für alle $g(y) = (g_1(y), g_2(y))$, $h(y) = (h_1(y), h_2(y)) \in \mathcal{G}_1 \times \mathcal{G}_2$ und für alle $y \in \mathcal{Y}$.
Ist $\phi \circ (g_1, g_2)$ für mindestens ein Element $(g_1, g_2) \in \mathcal{G}_1 \times \mathcal{G}_1$ quadratintegrierbar, dann ist $\phi \circ (\mathcal{G}_1, \mathcal{G}_2) = \{\phi \circ (g_1, g_2) : (g_1, g_2) \in \mathcal{G}_1 \times \mathcal{G}_2\}$ ebenso eine Donsker-Klasse.

Beispiel 6.13.
Ist $\mathcal{G}$ eine Donsker-Klasse mit $\|\mathbb{E}\|_{\mathcal{G}} < \infty$ und h eine gleichmäßig beschränkte, messbare Funktion mit $\|h\|_\infty \leq 1$, dann ist auch $\mathcal{G} \cdot h$ eine Donsker-Klasse, denn mit $\phi(g, h) = g \cdot h$ gilt

$$|\phi(g_1(y), h(y)) - \phi(g_2(y), h(y))| = |h(y)| \cdot |g_1(y) - g_2(y)| \leq |g_1(y) - g_2(y)|.$$

□

Hängt das zugrunde liegende Wahrscheinlichkeitsmaß von einem Parameter $\theta \in \mathbb{R}$ ab, bedarf es einer Schätzung des Parameters mithilfe der vorliegenden Daten, welche mit $\hat{\theta}_n$ bezeichnet wird. Der Grenzprozess von $\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\hat{\theta}_n})$ ist dann zwar weiterhin ein Gauß-Prozess, jedoch keine $\mathbb{P}_{\theta_0}$-Brownsche Brücke mehr. Van der Vaart [61] präzisiert diese Aussage in Satz 19.23, welcher nachfolgend in angepasster Notation wiedergegeben wird.

Satz 6.14.
Es sei $Y_1, \ldots, Y_n$ eine Zufallsstichprobe einer Verteilung $\mathbb{P}_{\theta_0}$ mit $\theta_0 \in \mathbb{R}$, welche die folgenden Voraussetzungen erfüllt:

(*i*) *Die Klasse $\mathcal{G}$ sei eine $\mathbb{P}_{\theta_0}$-Donsker-Klasse messbarer Funktionen.*

(*ii*) *Es existiere eine Funktion ϕ_{θ_0} mit $\mathbb{E}_{\theta_0}(\phi_{\theta_0}) = 0$ und $\mathbb{E}_{\theta_0}(\phi_{\theta_0}^2) < \infty$, sodass die Schätzfolge $\hat{\theta}_n$ asymptotisch linear sei, d.h.*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{\theta_0}(Y_i) + o_{\mathbb{P}_{\theta_0}}(1).$$

(*iii*) *Die Abbildung $\theta \in \mathbb{R} \mapsto \mathbb{E}_\theta \in \ell^\infty(\mathcal{G})$ sei Fréchet-differenzierbar in θ_0, d.h. es existiere eine Abbildung $\dot{\mathbb{E}}_{\theta_0} : \mathcal{G} \to \mathbb{R}$, sodass $\|\mathbb{E}_{\theta_0 + h} - \mathbb{E}_{\theta_0} - h \cdot \dot{\mathbb{E}}_{\theta_0}\|_{\mathcal{G}} = o_{\mathbb{P}_{\theta_0}}(|h|)$ für $h \to 0$.*

Dann konvergiert die Folge $\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\hat{\theta}_n})$ für den wahren Parameterwert θ_0 in Verteilung in $\ell^\infty(\mathcal{G})$ gegen den Prozess $g \mapsto \mathbb{B}_{\mathbb{P}_{\theta_0}}(g) - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0})\dot{\mathbb{E}}_{\theta_0}(g)$.

Ist das asymptotische Verhalten von $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{E})$ bekannt, so kann mithilfe der Delta-Methode auf die Verteilungskonvergenz von $\sqrt{n}(\phi_\Delta(\mathbb{P}_n) - \phi_\Delta(\mathbb{E}))$ für eine bekannte Funktion ϕ_Δ geschlossen werden. Für die Anwendung auf den empirischen Prozess

wird die Delta-Methode in einer funktionalen Variante benötigt. Die Umsetzung dieser Variante erfordert jedoch zunächst eine Anpassung des zugrunde liegenden Differenzierbarkeitskonzepts. Die folgenden Inhalte sind dem Kapitel 20 aus van der Vaart [61] entnommen.

Definition 6.15. (Hadamard-Differenzierbarkeit)
Es sei $\mathbb{D}$ ein normierter Raum. Eine Abbildung $\phi_\Delta : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{D}$ heißt *Hadamard-differenzierbar* in $y \in \mathbb{D}_\phi$, falls eine stetige, lineare Abbildung $\phi'_\Delta(y) : \mathbb{D} \to \mathbb{D}$ existiert, sodass

$$\left\| \frac{\phi_\Delta(y + th_t) - \phi_\Delta(y)}{t} - \phi'_\Delta(y)h \right\|_{\mathbb{D}} \to 0, \quad \text{für } t \to 0 \quad \text{und für alle } h_t \to h.$$

□

Bemerkung 6.16.
In den nachfolgenden Abschnitten wird der Raum $\ell^\infty(\mathcal{G})^k$, $k \in \mathbb{N}$ die Rolle des normierten Raumes $\mathbb{D}$ übernehmen, sodass es sich bei $\phi'_\Delta(y)$ um eine Matrix handeln wird. In der obigen Notation wird deswegen $\phi'_\Delta(y)h$ bereits als Produkt dargestellt.
Existiert die Ableitung ϕ'_Δ nicht auf dem gesamten Raum $\mathbb{D}$, sondern nur auf einer Teilmenge $\mathbb{D}_0$, und konvergiert die Folge $h_t \to h$ gegen einen Grenzwert $h \in \mathbb{D}_0$, dann wird die Abbildung ϕ_Δ auch *Hadamard-differenzierbar tangential* an $\mathbb{D}_0$ genannt. □

Im vorangegangenen Satz wurde bereits die Fréchet-Differenzierbarkeit für Abbildungen speziell auf den reellen Zahlen verwendet, welche jedoch ebenfalls für allgemeine normierte Räume definiert werden kann. Im Allgemeinen kann die Fréchet-Differenzierbarkeit als eine stärkere Form der Differenzierbarkeit im Vergleich zur Differenzierbarkeit nach Hadamard betrachtet werden. Für $\mathbb{D} = \mathbb{R}^k$ stimmen die Konzepte jedoch überein.
Ist über die Funktion ϕ_Δ bereits bekannt, dass sie stetig differenzierbar im gewöhnlichen Sinne ist, so folgt daraus unmittelbar die Hadamard-Differenzierbarkeit für bestimmte Räume.

Lemma 6.17.
Es sei $\phi_\Delta : [0,1] \to \mathbb{R}$ eine stetig differenzierbare Funktion. Dann ist die Abbildung $z \mapsto \phi_\Delta \circ z$ für Funktionen $z : \mathcal{G} \to [0,1]$ aus $\ell^\infty(\mathcal{G})$ Hadamard-differenzierbar.

Mithilfe der Maximumsnorm kann das Lemma auf mehrere Dimensionen erweitert werden. Für die nächsten Abschnitte wird insbesondere der folgende Spezialfall benötigt.

Lemma 6.18.
Es sei $\phi^1_\Delta : [0,1] \to \mathbb{R}$ eine stetig differenzierbare Funktion mit der Ableitung $(\phi^1_\Delta)'$. Weiter bezeichne $\phi_\Delta(y_1, y_2) = (y_1, \phi^1_\Delta(y_2))$ für y_1, $y_2 \in [0,1]$. Der Raum $(\ell^\infty(\mathcal{G})^2, \|\cdot\|)$ sei mit der Maximumsnorm versehen. Dann ist die Abbildung $z = (z_1, z_2) \mapsto \phi_\Delta \circ z$ für Funktionen z_1, $z_2 : \mathcal{G} \to [0,1]$ aus $\ell^\infty(\mathcal{G})$ Hadamard-differenzierbar mit

$$\phi'_\Delta(z_1, z_2) = \begin{pmatrix} 1 & 0 \\ 0 & (\phi^1_\Delta)'(z_2) \end{pmatrix}.$$

Beweis. Für $t \to 0$ und $h_t = (h_t^1, h_t^2) \to h = (h^1, h^2)$ gilt mit Lemma 6.17

$$\left\| \frac{\phi_\Delta(z + th_t) - \phi_\Delta(z)}{t} - \phi'_\Delta(z)h \right\| = \left\| \begin{pmatrix} (z_1 + th_t^1 - z_1)/t \\ (\phi_\Delta^1(z_2 + th_t^2) - \phi_\Delta^1(z_2))/t \end{pmatrix} - \begin{pmatrix} h^1 \\ (\phi_\Delta^1)'(z_2)h^2 \end{pmatrix} \right\|$$
$$= \max \left\{ \|h_t^1 - h^1\|_\mathcal{G}, \left\| (\phi_\Delta^1(z_2 + th_t^2) - \phi_\Delta^1(z_2))/t - (\phi_\Delta^1)'(z_2)h^2 \right\|_\mathcal{G} \right\} \to 0.$$

■

Die folgende, funktionale Delta-Methode kann auf den empirischen Prozess angewendet werden.

Satz 6.19. (Funktionale Delta-Methode)
Es seien $\mathbb{D}$ und $\mathbb{E}$ zwei normierte Vektorräume und $\phi_\Delta : \mathbb{D}_\phi \subset \mathbb{D} \to \mathbb{E}$ Hadamard-differenzierbar in y tangential an $\mathbb{D}_0$. Weiter seien $R_n : \Omega_n \to \mathbb{D}_\phi$ Abbildungen, sodass $\sqrt{n}(R_n - y)$ in Verteilung gegen R mit Werten in $\mathbb{D}_0$ konvergiert. Dann konvergiert auch $\sqrt{n}(\phi_\Delta(R_n) - \phi_\Delta(y))$ in Verteilung gegen $\phi'_\Delta(y)R$.

Die Sätze können nun auf den trunkierten Prozess übertragen werden. Zur Vereinfachung wird stochastische Unabhängigkeit zwischen den Variablen X_i und T_i angenommen und der wahre Parameterwert θ_0 sowie der latente Stichprobenumfang n als bekannt vorausgesetzt. In den darauffolgenden Schritten wird die Schätzung der beiden Größen berücksichtigt.

6.2. Anpassungstest für unabhängige Doppeltrunkation

Im Folgenden gelten (A2) sowie (A4) und es sei $\theta_0 \in [\varepsilon_\theta, 1/\varepsilon_\theta]$ bekannt. Die Zufallsvariablen X_i und T_i seien stochastisch unabhängig, sodass sich die gemeinsame Verteilung nicht über die Copulas wie in (A3), (A3') oder (A3''), sondern wie bei Weißbach und Wied [65] als Produkt der beiden Randverteilungen ergibt:

(A3''') Die Unabhängigkeit zwischen X_i und T_i wird mithilfe der Produktcopula (2.16)

$$\Pi(u, v) = u \cdot v$$

modelliert.

Es werden keine neuen Symbole und Bezeichnungen eingeführt. Die verwendeten Symbole und Bezeichnungen aus Abschnitt 3.1 sind im Sinne der vorangegangenen Annahmen zu verstehen.
Die Auswahlwahrscheinlichkeit α_{θ_0} kann in diesem Fall explizit berechnet werden. Nach Weißbach und Wied [65] (dort Gleichung (1)) gilt

$$\alpha_{\theta_0} = \int_0^G \int_t^{t+s} \frac{\theta_0}{G} e^{-\theta_0 x} \, \mathrm{d}x \, \mathrm{d}t = \frac{(1 - e^{-\theta_0 s})(1 - e^{-\theta_0 G})}{G\theta_0}. \tag{6.2}$$

Der in Abschnitt 3.2 definierte zweidimensionale trunkierte Prozess $N_{n,D}$ ist durch

$$N_{n,D}(\cdot) = N_n(\cdot \cap D) = \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}(\cdot \cap D)$$

gegeben. In Analogie zu dem Beispiel 2.4 ist $\nu_{n,D} = n \cdot \mathbb{P}_{\theta_0}((X_1, T_1)^{\mathrm{T}} \in \cdot \cap D)$ das zugehörige Intensitätsmaß. Für eine Menge $B \in \mathcal{B}$ folgt die Zufallsvariable $N_{n,D}(B)$ aufgrund der Unabhängigkeit von $(X_1, T_1)^{\mathrm{T}}, \ldots, (X_n, T_n)^{\mathrm{T}}$ der Binomialverteilung. Für deren Erwartungswert und Varianz gelten folglich

$$\begin{aligned}
\mathbb{E}[N_{n,D}(B)] &= n \cdot \mathbb{P}_{\theta_0}\Big((X_1, T_1)^{\mathrm{T}} \in B \cap D\Big) = \nu_{n,D}(B) \quad \text{sowie} \\
\mathrm{Var}[N_{n,D}(B)] &= n \cdot \mathbb{P}_{\theta_0}\Big((X_1, T_1)^{\mathrm{T}} \in B \cap D\Big)\Big[1 - \mathbb{P}_{\theta_0}\Big((X_1, T_1)^{\mathrm{T}} \in B \cap D\Big)\Big] \\
&= \nu_{n,D}(B)\Big(n - \nu_{n,D}(B)\Big)/n.
\end{aligned}$$

Das starke Gesetz der großen Zahlen sowie der zentrale Grenzwertsatz können direkt auf den trunkierte Prozess angewendet werden. Es gelten

$$n^{-1} N_{n,D}(B) \to n^{-1} \nu_{n,D}(B) = \mathbb{P}_{\theta_0}((X_1, T_1)^{\mathrm{T}} \in B \cap D) \quad \text{für} \quad n \to \infty$$

fast sicher sowie

$$\frac{n^{-1/2}\Big(N_{n,D}(B) - \nu_{n,D}(B)\Big)}{\sqrt{\nu_{n,D}(B)\Big(n - \nu_{n,D}(B)\Big)/n}} \to \mathcal{N}(0,1) \quad \text{für} \quad n \to \infty.$$

Es sollen nun die folgenden Hypothesen gegeneinander getestet werden:

$$\begin{aligned}
H_0 &: \text{Der Zufallsvektor } (X_i, T_i)^{\mathrm{T}} \text{ besitzt die Verteilung } F_{\theta_0} = F^X \cdot F^T \text{ mit} \\
&\quad \text{dem Parameter } \theta_0 \in \Theta. \\
H_1 &: \text{Der Zufallsvektor } (X_i, T_i)^{\mathrm{T}} \text{ besitzt nicht die Verteilung } F_{\theta_0} = F^X \cdot F^T \\
&\quad \text{mit dem Parameter } \theta_0 \in \Theta.
\end{aligned} \tag{6.3}$$

Wie im vorherigen Abschnitt bereits angedeutet, soll der Kolmogorov-Smirnov-Anpassungstest zur Überprüfung der Hypothese herangezogen werden. Die zugehörige Teststatistik lautet $\sqrt{n}\, d_n(\theta_0, n)$ mit

$$d_n(\theta_0, n) := \sup_{(x,t) \in S} \frac{1}{n}\Big|N_{n,D}\Big([0,x] \times [0,t]\Big) - \nu_{n,D}\Big([0,x] \times [0,t]\Big)\Big|. \tag{6.4}$$

Im Folgenden wird zunächst die asymptotische Verteilung der Prüfgröße anhand der im vorangegangenen Abschnitt zusammengefassten Konvergenzsätze für empirische Verteilungsfunktionen und unter Gültigkeit der Nullhypothese hergeleitet. Es wird dabei unterschieden, ob der wahre Parameterwert θ_0 bekannt ist oder geschätzt werden muss und ob die latente Stichprobegröße n bekannt ist oder geschätzt werden muss. Es sei angemerkt, dass die Prämisse der Bekanntheit der latenten Stichprobengröße dabei nicht

die Beobachtbarkeit der latenten Stichprobe selbst impliziert. Die Konvergenz wird immer für $n \to \infty$ betrachtet, auch wenn n nicht beobachtet werden kann.
Für die spätere Anwendung auf die Daten ist insbesondere die Konstellation von Relevanz, bei der sowohl θ_0 als auch n unbekannt sind. Algorithmus 4 ist folglich der zentrale Algorithmus zur Bestimmung des Quantils. Die Algorithmen 1 bis 3 und die jeweils zugehörigen Sätze legen die Grundlagen, in dem sie grundlegende Teilprobleme lösen. Im Anschluss wird eine Prozedur zur Berechnung der Prüfgröße erläutert, welche in Algorithmus 5 festgehalten wird.

6.2.1. Verteilung der Teststatistik bei (θ_0, n)

Für den Anpassungstest ist es ausreichend, Mengen der Form $B = [0, x] \times [0, t]$ mit $x \in \mathbb{R}_0^+$, $t \in [0, G]$ zu betrachten. Wie bereits angedeutet, sollen die Konvergenzsätze über empirische Prozesse aus van der Vaart [61] und van der Vaart und Wellner [62] auf das in Weißbach und Wied [65] vorgestellte Modell zur unabhängigen Doppeltrunkation in der Punktprozess-Notation von Reiss [48] angewendet werden. Die Notation $\mathbb{G}_n(g)$ entspricht somit nun $n^{-1/2}(N_{n,D}([0,x]\times[0,t]) - \nu_{n,D}([0,x]\times[0,t]))$ mit g als $\mathbb{1}_{[0,x]\times[0,t]\cap D}$.

Nach Beispiel 6.11 besitzt die Funktionenklasse $\mathcal{G} = \{g_{x,t} = \mathbb{1}_{[0,x]\times[0,t]}, (x,t) \in S\}$ eine endliche Klammerungszahl und ein endliches Klammerungsintegral. Die unteren und oberen Grenzen der Partitionen müssen dabei der Menge S angepasst werden. Nach Satz 6.8 ist $\mathcal{G}$ eine Glivenko-Cantelli-Klasse, d.h. für den empirischen Punktprozess $N_n(B) = \sum_{i=1}^n \epsilon_{(X_i,T_i)^\mathrm{T}}(B)$ und sein Intensitätsmaß $\nu_n(B) = n \cdot \mathbb{P}_{\theta_0}((X_1, T_1)^\mathrm{T} \in B)$ gilt

$$\sup_{(x,t)\in S} n^{-1}\Big|N_n([0,x]\times[0,t]) - \nu_n([0,x]\times[0,t])\Big| \to 0$$

fast sicher für $n \to \infty$. Nach Satz 6.10 ist $\mathcal{G}$ eine Donsker-Klasse, sodass die Folge

$$\Big\{n^{-1/2}\Big(N_n([0,x]\times[0,t]) - \nu_n([0,x]\times[0,t])\Big) : (x,t) \in S\Big\} \tag{6.5}$$

in Verteilung gegen eine $\mathbb{P}_{\theta_0}$-Brownsche Brücke $\mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t})$ in $\ell^\infty(\mathcal{G})$ konvergiert. Der Grenzprozess hat den Erwartungswert $\mathbb{E}[\mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t})] = 0$ und die Kovarianzfunktion

$$\begin{aligned}\mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x_1,t_1}), \mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x_2,t_2})] = {}& \mathbb{P}_{\theta_0}\Big((X_1,T_1)^\mathrm{T} \in [0, \min\{x_1,x_2\}] \times [0, \min\{t_1,t_2\}]\Big) \\ & - \mathbb{P}_{\theta_0}\Big((X_1,T_1)^\mathrm{T} \in B_1\Big) \cdot \mathbb{P}_{\theta_0}\Big((X_1,T_1)^\mathrm{T} \in B_2\Big)\end{aligned}$$

mit $g_{x_i,t_i} \in \mathcal{G}$ und $B_i = [0, x_i] \times [0, t_i]$ für $i = 1, 2$ (vgl. Beispiel 6.5).
Die Folge (6.5) ist nur für Vektoren $(x,t) \in S$ und nicht für Funktionen $g_{x,t} \in \mathcal{G}$ definiert, sodass streng genommen nicht von Konvergenz in $\ell^\infty(\mathcal{G})$ gesprochen werden kann. Im gesamten Kapitel werden die Ausdrücke (6.5) und $\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0})(g_{x,t})$ sowie deren Abwandlungen synonym verwendet, sodass sprachlich nicht bei jeder Verwendung zwischen den zugehörigen Normen unterschieden wird.

Wählt man in Beispiel 6.13 $h = \mathbb{1}_D$, dann resultiert aus $\|\mathbb{1}_D\|_\infty = 1$ die Donsker-Eigenschaft für die Funktionenklasse $\mathcal{G}_D = \{g_{x,t,D} = \mathbb{1}_{[0,x]\times[0,t]}\mathbb{1}_D,\ (x,t) \in S\}$.

Satz 6.20.
Es bezeichne $\mathbb{B}_{\mathbb{P}_{\theta_0}}$ eine $\mathbb{P}_{\theta_0}$-Brownsche Brücke auf $\mathcal{G}_D$. Die Teststatistik $\sqrt{n}d_n(\theta_0, n)$ konvergiert unter H_0 in Verteilung gegen $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$.

Beweis. Nach Beispiel 6.13 konvergiert die Folge

$$\Big\{n^{-1/2}\Big(N_{n,D}([0,x]\times[0,t]) - \nu_{n,D}([0,x]\times[0,t])\Big) : (x,t) \in S\Big\}$$

unter H_0 in Verteilung gegen eine $\mathbb{P}_{\theta_0}$-Brownsche Brücke $\mathbb{B}_{\mathbb{P}_{\theta_0}}$ in $\ell^\infty(\mathcal{G}_D)$. Der Erwartungswert des Prozesses ist null. Die Kovarianzfunktion $\mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x_1,t_1,D}), \mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x_2,t_2,D})]$ lautet

$$\begin{aligned}&\mathbb{P}_{\theta_0}\Big((X_1,T_1)^{\mathrm{T}} \in [0,\min\{x_1,x_2\}]\times[0,\min\{t_1,t_2\}]\cap D\Big)\\ &-\mathbb{P}_{\theta_0}\Big((X_1,T_1)^{\mathrm{T}} \in B_1\cap D\Big)\cdot\mathbb{P}_{\theta_0}\Big((X_1,T_1)^{\mathrm{T}} \in B_2\cap D\Big)\end{aligned} \tag{6.6}$$

mit $g_{x_i,t_i,D} \in \mathcal{G}_D$ und $B_i = [0,x_i]\times[0,t_i]$ für $i = 1,2$ (vgl. Beispiel 6.5). Die Abbildung $(\ell^\infty(\mathcal{G}_D), \|\cdot\|_{\mathcal{G}_D}) \to (\mathbb{R}, |\cdot|)$ mit $y \mapsto \|y\|_{\mathcal{G}_D}$ ist stetig und der Grenzprozess $\mathbb{B}_{\mathbb{P}_{\theta_0}}$ besitzt fast sicher stetige Pfade. Mit dem Satz über stetige Abbildungen konvergiert daher

$$\sup_{(x,t)\in S} n^{-1/2}\Big|N_{n,D}([0,x]\times[0,t]) - \nu_{n,D}([0,x]\times[0,t]\Big|$$

in Verteilung gegen $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$. ■

Es stellt sich nun die Frage, welche Verteilung $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ besitzt.

Im eindimensionalen Fall, d. h. ohne Trunkation, hängt die Grenzverteilung der Kolmogorov-Smirnov-Teststatistik bekanntlich nicht von der zugrundeliegenden (stetigen) Verteilung ab (vgl. Korollar 19.21 in van der Vaart [61]). Dies kann durch eine Transformation der Zufallsvariablen mithilfe ihrer Verteilung gezeigt werden. Im zweidimensionalen Fall üben zumindest die Randverteilungen keinen Einfluss auf die Grenzverteilung aus. Diese ist nach Naaman [44] Satz 3.1 identisch mit der von $\max\{K_1, K_2\}$. Hierbei haben K_1 und K_2 die Kolmogorov-Verteilung (6.1). Bei stochastischer Unabhängigkeit der zugrundeliegenden Zufallsvariablen berechnet sich diese als Produkt der Verteilungen von K_1 und K_2. Welchen Einfluss die zugrundeliegende Copula im Falle stochastischer Abhängigkeit hat, bleibt Gegenstand zukünftiger Forschung.

Bei Trunkation hingegen führt eine Transformation der Ränder nicht zum Erfolg, da der Einfluss der Randverteilungen in der Menge D nicht aufgelöst werden kann. Die Verteilung von $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ hängt von den Verteilungen von X_i und T_i ab und lässt sich nicht mithilfe der Kolmogorov-Verteilung berechnen. Es seien dafür $U_{1,i} = F^X(X_i)$ und $U_{2,i} = F^T(T_i)$ auf dem Intervall $[0,1]$ gleichverteilte Zufallsvariablen. Dann folgt $X_i = (F^X)^{-1}(U_{1,i})$ sowie $T_i = (F^T)^{-1}(U_{2,i})$. Für die gemeinsame Verteilungsfunktion

gilt dann

$$\mathbb{P}_{\theta_0}\left((X_i, T_i)^{\mathrm{T}} \in [0,x] \times [0,t] \cap D\right) = \mathbb{P}_{\theta_0}\left((U_{1,i}, U_{2,i})^{\mathrm{T}} \in [0,u_1] \times [0,u_2] \cap D_{tr}\right)$$

mit $u_1 = F^X(x)$ und $u_2 = F^T(t)$. Aufgrund der Monotonie der Verteilungsfunktion sind beispielsweise $0 \leq X_i \leq x$ und $0 \leq F^X(X_i) = U_{1,i} \leq F^X(x) = u_1$ äquivalent. Für die Menge D gilt nun

$$\begin{aligned}
&(X,T)^{\mathrm{T}} \in D && \Leftrightarrow \\
&0 < T \leq X \leq T+s,\, T \leq G && \Leftrightarrow \\
&0 < U_{2,i} \leq F^T((F^X)^{-1}(U_{1,i})) \leq F^T((F^T)^{-1}(U_{2,i}) + s),\, U_{2,i} \leq F^T(G) = 1 && \Leftrightarrow \\
&(U_{1,i}, U_{2,i})^{\mathrm{T}} \in D_{tr}.
\end{aligned} \tag{6.7}$$

Folglich hängt die Menge D_{tr} auch nach der Transformation von F^X und F^T ab und die Verteilung von $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ kann nicht exakt ermittelt werden.
Zur Bestimmung des Quantils können jedoch Werte von $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ wie folgt simuliert werden.

Algorithmus 1: Simulation der Verteilung von $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$

1. Erzeuge ein Gitter auf dem Rechteck $[0, G+s] \times [0, G]$.
2. Berechne für alle Gitterpunkte die Kovarianzmatrix nach (6.6).
3. Wiederhole für eine festgelegte Anzahl von Iterationen:
   i) Erzeuge einen normalverteilten Vektor mit Erwartungswert null und der berechneten Kovarianzmatrix.
   ii) Ermittle das betragsmäßige Maximum des Vektors.
4. Wähle aus den sortierten Maxima das entsprechende Quantil aus.

Bemerkung 6.21.
Die Berechnung der Kovarianzmatrix unter Verwendung der Kovarianzfunktion gemäß Gleichung (6.6) erfordert keine numerische Integration. Die Schnittmengen aus den Rechtecken und dem Beobachtungsparallelogramm D können aus Dreiecken und Rechtecken konstruiert werden. Zu diesem Zweck ist eine Fallunterscheidung zwischen den verschiedenen Lagen des Rechtecks zum Parallelogramm erforderlich. Aufgrund der stochastischen Unabhängigkeit zwischen den Variablen X und T, ist es möglich, die Integrale direkt zu berechnen.
Für $(x,t)^{\mathrm{T}} \in D$ beispielsweise berechnet sich $\mathbb{P}_{\theta_0}\left((X_1, T_1)^{\mathrm{T}} \in [0,x] \times [0,t] \cap D\right)$ durch

$$\frac{e^{-\theta x}}{G} \cdot (\max\{x-s, 0\} - t) + \frac{1 - e^{-\theta t}}{G\theta} + \frac{e^{-\theta x} - e^{-\theta s}}{G\theta} \cdot \mathbb{1}_{\{x>s\}}(x).$$

Das Einführen einer Copula kann die Notwendigkeit zur numerischen Integration bedingen, wodurch sich die Komplexität der Berechnungen der Matrix erheblich erhöht. □

6.2.2. Verteilung der Teststatistik bei $(\hat{\theta}_n, n)$

Sofern der wahre Parameterwert nicht bekannt ist, kann dieser unter Zuhilfenahme der vorliegenden Daten geschätzt werden. Dies führt zu folgender zweiseitiger Fragestellung:

H_0 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt aus der Verteilungsfamilie $\{F_\theta = F^X \cdot F^T \,|\, \theta \in \Theta\}$. (6.8)

H_1 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt nicht aus der Verteilungsfamilie $\{F_\theta = F^X \cdot F^T \,|\, \theta \in \Theta\}$.

Wie bereits aus Satz 6.14 hervorgeht, weicht die Grenzverteilung von der einer Brownschen Brücke ab, wenn der Wert des Parameters geschätzt wird. Die Idee des Satzes ist es, eine Zerlegung der Art

$$\begin{aligned}\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\hat{\theta}_n}) &= \sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0}) - \sqrt{n}(\mathbb{E}_{\hat{\theta}_n} - \mathbb{E}_{\theta_0}) \\ &\approx \sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0}) - \sqrt{n}(\hat{\theta}_n - \theta_0)\dot{\mathbb{E}}_{\theta_0} \end{aligned} \tag{6.9}$$

vorzunehmen. Die Approximation wird dabei durch die folgende Fréchet-Differenzierbarkeit (*iii*) 6.14 gerechtfertigt.

Lemma 6.22.
Die Abbildung $\theta \in [\varepsilon_\theta, 1/\varepsilon_\theta] \mapsto \mathbb{E}_\theta \in \ell^\infty(\mathcal{G}_D)$ ist Fréchet-differenzierbar in θ_0.

Beweis. Da $\mathcal{G}_D$ nur Funktionen der Form $\mathbb{1}_{[0,x]\times[0,t]}\mathbb{1}_D$ für $(x,t) \in S$ enthält, kann die Fréchet-Ableitung explizit angegeben werden. Mithilfe der Dichtefunktion lässt sich die Abbildung $g_{x,t,D} \in \mathcal{G}_D \mapsto \mathbb{E}_{\theta_0}(g_{x,t,D}) \in \mathbb{R}$ schreiben als

$$\mathbb{E}_{\theta_0}(g_{x,t,D}) = \int_{[0,x]\times[0,t]\cap D} \frac{\theta_0}{G} e^{-\theta_0 \tilde{x}} \,\mathrm{d}(\tilde{x}, \tilde{t}).$$

Die Ableitung der Dichte nach θ ist auf D beschränkt:

$$\left| \frac{e^{-\theta_0 x}}{G}(1 - \theta_0 x) \right| \leq \frac{1}{G}\left(\frac{G+s}{\varepsilon_\theta} - 1\right)$$

Nach Satz 5.7 (Kap. IV, § 5.) aus dem Buch [13] von Elstrodt ist $\theta \mapsto \mathbb{E}_\theta(g_{x,t,D})$ für eine Funktion $g_{x,t,D} \in \mathcal{G}_D$ in θ_0 differenzierbar mit

$$\dot{\mathbb{E}}_{\theta_0}(g_{x,t,D}) := \int_{[0,x]\times[0,t]\cap D} \frac{e^{-\theta_0 \tilde{x}}}{G}(1 - \theta_0 \tilde{x}) \,\mathrm{d}(\tilde{x}, \tilde{t}).$$

Mit der Differenzierbarkeit von $f_\theta(x,t) = \theta e^{-\theta x}/G$ in θ_0 folgt zusammen mit dem Satz von Scheffé (vgl. Satz 5.9, Kap. IV, § 5. in [13]) daraus die Fréchet-Differenzierbarkeit,

denn

$$\lim_{h\to 0}\sup_{(x,t)\in S}\left|\int_{[0,x]\times[0,t]\cap D}\frac{f_{\theta_0+h}(\tilde{x},\tilde{t})-f_{\theta_0}(\tilde{x},\tilde{t})}{h}-\frac{\mathrm{d}f_{\theta_0}(\tilde{x},\tilde{t})}{\mathrm{d}\theta}\,\mathrm{d}(\tilde{x},\tilde{t})\right|$$
$$\leq\lim_{h\to 0}\int_D\left|\frac{f_{\theta_0+h}(\tilde{x},\tilde{t})-f_{\theta_0}(\tilde{x},\tilde{t})}{h}-\frac{\mathrm{d}f_{\theta_0}(\tilde{x},\tilde{t})}{\mathrm{d}\theta}\right|\mathrm{d}(\tilde{x},\tilde{t})=0.$$

■

Die Grenzverteilung der Approximation in Gleichung (6.9) erhält man mit Hilfe des Satzes über stetige Abbildungen aus der Grenzverteilung der Folge $\sqrt{n}(\mathbb{P}_n-\mathbb{E}_{\theta_0},\hat{\theta}_n-\theta_0)$. Zur Verteilungskonvergenz der ersten Komponente gelangt man mit der Donsker-Eigenschaft (i) 6.14 der Klasse $\mathcal{G}_D$. Für die Konvergenz des Vektors wird die asymptotische Normalität aus Satz 2 in [65] benötigt. Weißbach und Wied nehmen dabei Bezug auf Satz 5.41 in [61], wonach die Schätzfolge asymptotisch linear (ii) 6.14 ist, d.h.

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}(X_1,T_1)])^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i,T_i)+o_{\mathbb{P}_{\theta_0}}(1) \tag{6.10}$$

erfüllt. In Gleichung (4) wird die Funktion ψ_θ definiert und gemäß Lemma 2 und 3 in [65] erfüllt diese $\mathbb{E}_{\theta_0}[\psi_{\theta_0}(X_1,T_1)]=0$, $\mathbb{E}_{\theta_0}[\psi^2_{\theta_0}(X_1,T_1)]<\infty$ sowie $\dot{\psi}_{\theta_0}(X_1,T_1)>0$. Für $\phi_{\theta_0}:=-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1}\psi_{\theta_0}$ gelten damit ebenfalls

$$\mathbb{E}_{\theta_0}[\phi_{\theta_0}(X_1,T_1)]=0\quad\text{und}\quad\mathbb{E}_{\theta_0}[\phi^2_{\theta_0}(X_1,T_1)]<\infty. \tag{6.11}$$

Die Bedingungen $(i)-(iii)$ sind erfüllt und Satz 6.14 liefert damit die folgende Aussage.

Satz 6.23.
Es bezeichne $\hat{\nu}_{n,D}:=n\cdot\mathbb{P}_{\hat{\theta}_n}((X_i,T_i)^{\mathrm{T}}\in\cdot\cap D)$ sowie $\phi_{\theta_0}:=-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1}\psi_{\theta_0}$. Dann konvergiert

$$\Big\{n^{-1/2}\Big(N_{n,D}([0,x]\times[0,t])-\hat{\nu}_{n,D}([0,x]\times[0,t])\Big):(x,t)\in S\Big\}$$

in Verteilung in $\ell^\infty(\mathcal{G}_D)$ unter H_0 gegen den Gauß-Prozess

$$g_{x,t,D}\mapsto\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}):=\mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t,D})-\mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0})\dot{\mathbb{E}}_{\theta_0}(g_{x,t,D}). \tag{6.12}$$

Bemerkung 6.24.
Der Gauß-Prozess $\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}$ hat den Erwartungswert $\mathbb{E}_{\theta_0}[\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_{x,t,D})]=0$ und die Kovarianz von $\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_{x_1,t_1,D})$ und $\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_{x_2,t_2,D})$ lautet

$$\mathbb{E}_{\theta_0}(g_{\min(x_1,x_2),\min(t_1,t_2),D})-\mathbb{E}_{\theta_0}(g_{x_1,t_1,D})\mathbb{E}_{\theta_0}(g_{x_2,t_2,D})+\dot{\mathbb{E}}_{\theta_0}(g_{x_1,t_1,D})\dot{\mathbb{E}}_{\theta_0}(g_{x_2,t_2,D})(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1}$$
$$-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1}\frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}}\left(\dot{\mathbb{E}}_{\theta_0}(g_{x_2,t_2,D})\mathbb{E}_{\theta_0}(g_{x_1,t_1,D})+\dot{\mathbb{E}}_{\theta_0}(g_{x_1,t_1,D})\mathbb{E}_{\theta_0}(g_{x_2,t_2,D})\right). \tag{6.13}$$

Die Berechnung der Kovarianz kann im Abschnitt A.2 nachvollzogen werden. □

Der Satz über stetige Abbildungen impliziert die entsprechende Konvergenz der Folge in der Supremumsnorm $\|\cdot\|_{\mathcal{G}_D}$, d.h. die Konvergenz der modifizierten Kolmogorov-Smirnov-Teststatistik $\sqrt{n}d_n(\hat{\theta}_n, n)$. Die Verteilung von $\|\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ kann wie in Algorithmus 1 simuliert werden, wobei die Kovarianz aus Gleichung (6.13) verwendet wird.

Algorithmus 2: Simulation der Verteilung von $\|\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$

1. Erzeuge ein Gitter auf dem Rechteck $[0, G+s] \times [0, G]$.
2. Berechne für alle Gitterpunkte die Kovarianzmatrix nach (6.13).
3. Wiederhole für eine festgelegte Anzahl von Iterationen:
   i) Erzeuge einen normalverteilten Vektor mit Erwartungswert null und der berechneten Kovarianzmatrix.
   ii) Ermittle das betragsmäßige Maximum des Vektors.
4. Wähle aus den sortierten Maxima das entsprechende Quantil aus.

Bemerkung 6.25.
Die Kovarianzfunktion (6.13) ergänzt die Kovarianzfunktion (6.6) um die Fréchet-Ableitung der Funktion $\theta \mapsto \mathbb{E}_\theta$. Für deren Berechnung wird die Ableitung der gemeinsamen Dichtefunktion von X und T benötigt. Eine unmittelbare Berechnung der Integrale ist weiterhin möglich. Für die Ableitung der Auswahlwahrscheinlichkeit α_{θ_0} sowie den Erwartungswert der Ableitung der Funktion ψ_{θ_0} geben Weißbach und Wied [65] in Korollar 1 und Lemma 2 geschlossene Formeln an. □

6.2.3. Verteilung der Teststatistik bei $(\theta_0, \hat{n})$

Im vorherigen Abschnitt wurde der latente Stichprobenumfang als bekannt angenommen, was bei Trunkation jedoch nicht der Fall ist. In diesem Abschnitt soll deshalb die Schätzung von n in den Kolmogorov-Smirnov-Anpassungstest einbezogen werden, zunächst unter der Annahme, dass θ_0 bekannt. Da es sich bei n nicht um einen Parameter der Grundgesamtheit handelt, werden für diesen Fall ebenfalls die Hypothesen (6.3) untersucht.
Der nicht beobachtbare latente Stichprobenumfang n kann mittels $\hat{n} = M_n/\alpha_{\theta_0}$ (vgl. Abschnitt 2.2 in [65]) mit $\alpha_{\theta_0} \in (0,1)$ aus (6.2) und $M_n = \sum_{i=1}^{n} \mathbb{1}_{\{(X_i,T_i)^\intercal \in D\}}$ geschätzt werden. Mit dem schwachen Gesetz der großen Zahlen gilt

$$\frac{M_n}{n} \to \alpha_{\theta_0} \quad \text{in Wahrscheinlichkeit für} \quad n \to \infty. \tag{6.14}$$

Die Zufallsvariable M_n ist binomialverteilt mit dem Erwartungswert $n\alpha_{\theta_0}$ und der Varianz $n\alpha_{\theta_0}(1-\alpha_{\theta_0}) > 0$. Nach dem zentralen Grenzwertsatz gilt

$$\frac{M_n - n\alpha_{\theta_0}}{\sqrt{n\alpha_{\theta_0}(1-\alpha_{\theta_0})}} = \frac{\frac{1}{n}M_n - \alpha_{\theta_0}}{\frac{\sqrt{\alpha_{\theta_0}(1-\alpha_{\theta_0})}}{\sqrt{n}}} \to \mathcal{N}(0,1)$$

bzw.

$$\sqrt{n}\left(\frac{1}{n}M_n - \alpha_{\theta_0}\right) \to \mathcal{N}\left(0, \alpha_{\theta_0}(1-\alpha_{\theta_0})\right).$$

Für die Herleitung der asymptotischen Verteilung der Prüfgröße wird weiterhin die Verteilungskonvergenz von $\sqrt{n}(\sqrt{M_n/n} - \sqrt{\alpha_{\theta_0}})$ eine zentrale Rolle spielen. Diese kann mithilfe der Delta-Methode (vgl. Abschnitt 3.1 in [61]) für die Funktion $\phi_\Delta(x) = \sqrt{x}$ untersucht werden. Die Funktion ϕ_Δ ist differenzierbar mit der Ableitung $\phi'_\Delta(x) = (2\sqrt{x})^{-1}$, woraus sich die Konvergenz

$$\begin{aligned}\sqrt{n}\left(\phi_\Delta(M_n/n) - \phi_\Delta(\alpha_{\theta_0})\right) &= \sqrt{n}\left(\sqrt{M_n/n} - \sqrt{\alpha_{\theta_0}}\right) \\ \to \mathcal{N}(0, \alpha_{\theta_0}(1-\alpha_{\theta_0})(\phi'_\Delta)^2(\alpha_{\theta_0})) &= \mathcal{N}(0, (1-\alpha_{\theta_0})/4)\end{aligned} \tag{6.15}$$

ergibt. Es sei darauf hingewiesen, dass die Auswahlwahrscheinlichkeit wegen $s, G, \theta_0 > 0$ strikt positiv ist (vgl. Gleichung (6.2)).

Mit $F_{\theta_0}^{\tilde{X},\tilde{T}}$ ähnlich wie in Definition 3.4 lautet die Teststatistik $\sqrt{\hat{n}}d_n(\theta_0, \hat{n})$ nun

$$\begin{aligned}&\sqrt{\hat{n}} \sup_{(x,t)\in S} \left| \frac{\alpha_{\theta_0}}{M_n} \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}\Big([0,x]\times[0,t]\cap D\Big) - \mathbb{P}_{\theta_0}\Big((X_i,T_i)^{\mathrm{T}} \in [0,x]\times[0,t]\cap D\Big)\right| \\ &= \sup_{(x,t)\in S} \left| \sqrt{\hat{n}}\frac{\alpha_{\theta_0}}{M_n} \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}\Big([0,x]\times[0,t]\cap D\Big) - \sqrt{\hat{n}}\alpha_{\theta_0} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right| \\ &= \sup_{(x,t)\in S} \left| \frac{\sqrt{\alpha_{\theta_0}}}{\sqrt{M_n}} \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}\Big([0,x]\times[0,t]\cap D\Big) - \sqrt{M_n}\sqrt{\alpha_{\theta_0}} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right|.\end{aligned}$$

Um zunächst die asymptotische Verteilung des absolut zu maximierenden Arguments

$$\frac{\sqrt{\alpha_{\theta_0}}}{\sqrt{M_n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{M_n}\sqrt{\alpha_{\theta_0}} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t) \tag{6.16}$$

zu bestimmen, wird die Idee der Zerlegung aus Gleichung (6.9) aufgegriffen. Dafür wird der Term

$$\pm\left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\theta_0} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right)$$

eingefügt. Das Ausklammern von $N_{n,D}$ und $F_{\theta_0}^{\tilde{X},\tilde{T}}$ führt zu

$$\begin{aligned}&\left(\frac{\sqrt{\alpha_{\theta_0}}}{\sqrt{M_n}} - \frac{1}{\sqrt{n}}\right) N_{n,D}\Big([0,x]\times[0,t]\Big) - \left(\sqrt{M_n}\sqrt{\alpha_{\theta_0}} - \sqrt{n}\alpha_{\theta_0}\right) F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t) \\ &\qquad + \left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\theta_0} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right).\end{aligned}$$

Anschließend wird der Faktor $\sqrt{M_n}n/(\sqrt{M_n}n)$ ergänzt

$$\begin{aligned}\Big(\sqrt{n}\sqrt{\alpha_{\theta_0}}-\sqrt{M_n}\Big)\cdot\frac{\sqrt{n}}{\sqrt{M_n}}\cdot\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)-\Big(\sqrt{M_n}\sqrt{\alpha_{\theta_0}}-\sqrt{n}\alpha_{\theta_0}\Big)F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\\+\left(\frac{1}{\sqrt{n}}N_{n,D}\Big([0,x]\times[0,t]\Big)-\sqrt{n}\alpha_{\theta_0}F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right)\end{aligned}$$

und der Term $(\sqrt{M_n}-\sqrt{n}\sqrt{\alpha_{\theta_0}})$ kann ausgeklammert werden

$$\begin{aligned}&\underbrace{\Big(\sqrt{M_n}-\sqrt{n}\sqrt{\alpha_{\theta_0}}\Big)}_{1}\cdot\underbrace{\left[-\frac{\sqrt{n}}{\sqrt{M_n}}\cdot\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)-\sqrt{\alpha_{\theta_0}}F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right]}_{2}\\&\qquad+\underbrace{\left(\frac{1}{\sqrt{n}}N_{n,D}\Big([0,x]\times[0,t]\Big)-\sqrt{n}\alpha_{\theta_0}F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right)}_{3}.\end{aligned}\tag{6.17}$$

Die Verteilungskonvergenz von Teil 1 in der Darstellung (6.17) wurde bereits in Gleichung (6.15) mithilfe der Delta-Methode hergeleitet. Die Verteilungskonvergenz von Teil 3 in $\ell^\infty(\mathcal{G}_D)$ wurde bereits kurz vor und im Beweis von Satz 6.20 behandelt. Für Teil 2 in Darstellung (6.17) soll ähnlich wie in Gleichung (6.9) eine Approximation in der $\mathcal{G}_D$-Norm erfolgen. Anstelle der Fréchet-Differenzierbarkeit wird hier jedoch der Satz von Glivenko-Cantelli Verwendung finden.

Lemma 6.26.
Für $n\to\infty$ gilt die Konvergenz in Wahrscheinlichkeit

$$\sup_{(x,t)\in D}\left|\frac{\sqrt{n}}{\sqrt{M_n}}\cdot\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)-\frac{1}{\sqrt{\alpha_{\theta_0}}}\mathbb{P}_{\theta_0}\Big((X_i,T_i)^{\mathrm{T}}\in[0,x]\times[0,t]\cap D\Big)\right|\to 0.$$

Beweis. Das Einfügen von $\pm 1/\sqrt{\alpha_{\theta_0}}$ führt zunächst zu

$$\begin{aligned}&\sup_{(x,t)\in D}\left|\left(\frac{\sqrt{n}}{\sqrt{M_n}}\pm\frac{1}{\sqrt{\alpha_{\theta_0}}}\right)\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)-\frac{1}{\sqrt{\alpha_{\theta_0}}}\mathbb{P}_{\theta_0}\Big((X_i,T_i)^{\mathrm{T}}\in[0,x]\times[0,t]\cap D\Big)\right|\\&\quad\le\left|\frac{\sqrt{n}}{\sqrt{M_n}}-\frac{1}{\sqrt{\alpha_{\theta_0}}}\right|\cdot\sup_{(x,t)\in D}\left|\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)\right|\\&\quad+\frac{1}{\sqrt{\alpha_{\theta_0}}}\sup_{(x,t)\in D}\left|\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)-\mathbb{P}_{\theta_0}\Big((X_i,T_i)^{\mathrm{T}}\in[0,x]\times[0,t]\cap D\Big)\right|.\end{aligned}$$

Wegen (6.14) folgen

$$\left|\frac{\sqrt{n}}{\sqrt{M_n}}-\frac{1}{\sqrt{\alpha_{\theta_0}}}\right|\to 0\quad\text{sowie}\quad\sup_{(x,t)\in D}\left|\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)\right|=\frac{M_n}{n}\to\alpha_{\theta_0}$$

in Wahrscheinlichkeit. Kurz vor der Ausführung von Satz 6.20 wird dargelegt, dass es sich bei $\mathcal{G}_D$ um eine Donsker-Klasse handelt. Jede Donsker-Klasse ist auch eine

Glivenko-Cantelli-Klasse (vgl. Abschnitt 2.1 in [62]). Es lässt sich schließen, dass

$$\sup_{(x,t)\in D}\left|\frac{1}{n}N_{n,D}\Big([0,x]\times[0,t]\Big)-\mathbb{P}_{\theta_0}\Big((X_i,T_i)^{\mathrm{T}}\in[0,x]\times[0,t]\cap D\Big)\right|\to 0 \quad \text{fast sicher.}$$

■

Der folgende Satz führt die Erkenntnisse zu den Termen 1 bis 3 aus der Darstellung (6.17) formal zusammen.

Satz 6.27.
Die Folge

$$\left\{\frac{\sqrt{\alpha_{\theta_0}}}{\sqrt{M_n}}N_{n,D}\Big([0,x]\times[0,t]\Big)-\sqrt{M_n}\sqrt{\alpha_{\theta_0}}F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t):(x,t)\in S\right\}$$

konvergiert in Verteilung in $\ell^\infty(\mathcal{G}_D)$ unter H_0 gegen den Gauß-Prozess

$$g_{x,t,D}\mapsto \mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}):=\mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t,D})-\mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D)\mathbb{E}_{\theta_0}(g_{x,t,D})/\alpha_{\theta_0},\quad g_{x,t,D}\in\mathcal{G}_D.$$

Beweis. Wie bereits ausgeführt, kann der Ausdruck (6.16) in (6.17) umformuliert werden. Die folgende Approximation kann im Sinne der $\|\cdot\|_\infty$-Norm mit Lemma 6.26 gerechtfertigt werden:

$$\begin{aligned}(6.16)=(6.17)=&\left(\frac{1}{\sqrt{n}}N_{n,D}\Big([0,x]\times[0,t]\Big)-\sqrt{n}\alpha_{\theta_0}F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right)\\&+\left(\sqrt{M_n}-\sqrt{n}\sqrt{\alpha_{\theta_0}}\right)\cdot\left[-2\sqrt{\alpha_{\theta_0}}F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right]+o_{\mathbb{P}_{\theta_0}}(1)\\=&\sqrt{n}(\mathbb{P}_n(g_{x,t,D})-\mathbb{E}_{\theta_0}(g_{x,t,D}))\\&+\sqrt{n}\left(\sqrt{\mathbb{P}_n(\mathbb{1}_D)}-\sqrt{\mathbb{E}_{\theta_0}(\mathbb{1}_D)}\right)\cdot\left(-2\,\mathbb{E}_{\theta_0}(g_{x,t,D})/\sqrt{\alpha_{\theta_0}}\right)+o_{\mathbb{P}_{\theta_0}}(1)\end{aligned}$$

Zur Formulierung der Konvergenz im Raum $\ell^\infty(\mathcal{G}_D)$ wird auf die Notation aus Abschnitt 6.1 zurückgegriffen. Dabei gilt $\mathbb{E}_{\theta_0}(\mathbb{1}_D)=\alpha_{\theta_0}$. Die Approximation gilt dann entsprechend in der $\|\cdot\|_{\mathcal{G}_D}$-Norm.
Nach den Ausführungen kurz vor Satz 6.20 ist die Klasse $\mathcal{G}_D$ eine Donsker-Klasse, sodass die Folge $\{\sqrt{n}(\mathbb{P}_n-\mathbb{E}_{\theta_0})(g_{x,t,D}):\ g_{x,t,D}\in\mathcal{G}_D\}$ in Verteilung in $\ell^\infty(\mathcal{G}_D)$ gegen eine $\mathbb{P}_{\theta_0}$-Brownsche Brücke $\mathbb{B}_{\mathbb{P}_{\theta_0}}$ konvergiert. Es sei darauf hingewiesen, dass auch $\mathbb{1}_D=g_{G+s,G,D}$ in der Klasse $\mathcal{G}_D$ enthalten ist. Die Abbildung

$$f:\Big(\sqrt{n}(\mathbb{P}_n-\mathbb{E}_{\theta_0})\Big)\mapsto\Big(\sqrt{n}(\mathbb{P}_n-\mathbb{E}_{\theta_0}),\sqrt{n}(\mathbb{P}_n(\mathbb{1}_D)-\mathbb{E}_{\theta_0}(\mathbb{1}_D))\Big)^{\mathrm{T}},$$

mit $(\sqrt{n}(\mathbb{P}_n-\mathbb{E}_{\theta_0}))\in\ell^\infty(\mathcal{G}_D)$ ist stetig, sodass der Satz über stetige Abbildungen Anwendung finden kann. Demnach konvergiert $\{f(\sqrt{n}(\mathbb{P}_n-\mathbb{E}_{\theta_0}))(g_{x,t,D}):\ g_{x,t,D}\in\mathcal{G}_D\}$

in Verteilung in $\ell^\infty(\mathcal{G}_D)$ gegen $f(\mathbb{B}_{\mathbb{P}_{\theta_0}})$, d.h.

$$\begin{pmatrix} \sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0}) \\ \sqrt{n}(\mathbb{P}_n(\mathbb{1}_D) - \mathbb{E}_{\theta_0}(\mathbb{1}_D)) \end{pmatrix} \to \begin{pmatrix} \mathbb{B}_{\mathbb{P}_{\theta_0}} \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D) \end{pmatrix}$$

in Verteilung in $\ell^\infty(\mathcal{G}_D)$. Wie bereits in Gleichung (6.15) angedeutet, wird nun die Delta-Methode zum Einsatz gebracht. Es wird dafür die Funktion

$$\phi_\Delta : \ell^\infty(\mathcal{G}_D) \times \ell^\infty(\mathcal{G}_D) \to \ell^\infty(\mathcal{G}_D) \times \ell^\infty(\mathcal{G}_D), \quad \phi_\Delta(x, y) = (x, \sqrt{y})^{\mathrm{T}}$$

betrachtet. Die Identität sowie die Wurzelfunktion sind (auf der positiven reellen Achse) stetig differenzierbar. Die einzelnen Komponenten der Funktion ϕ_Δ sind somit nach Lemma 6.17 Hadamard-differenzierbar. Mit Lemma 6.18 folgt nun daraus auch die Hadamard-Differenzierbarkeit der Funktion ϕ_Δ mit

$$\phi'_\Delta(x, y) = \begin{pmatrix} 1 & 0 \\ 0 & (2\sqrt{y})^{-1} \end{pmatrix}.$$

Nach Satz 6.19 (funktionale Delta-Methode) konvergiert nun

$$\begin{aligned} &\sqrt{n}\Big(\phi_\Delta(\mathbb{P}_n, \mathbb{P}_n(\mathbb{1}_D)) - \phi_\Delta(\mathbb{E}_{\theta_0}, \mathbb{E}_{\theta_0}(\mathbb{1}_D))\Big) = \begin{pmatrix} \sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0}) \\ \sqrt{n}\left(\sqrt{\mathbb{P}_n(\mathbb{1}_D)} - \sqrt{\mathbb{E}_{\theta_0}(\mathbb{1}_D)}\right) \end{pmatrix} \\ &\to \phi'_\Delta(\mathbb{E}_{\theta_0}, \mathbb{E}_{\theta_0}(\mathbb{1}_D)) \cdot \begin{pmatrix} \mathbb{B}_{\mathbb{P}_{\theta_0}} \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D) \end{pmatrix} = \begin{pmatrix} \mathbb{B}_{\mathbb{P}_{\theta_0}} \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D)/(2\sqrt{\alpha_{\theta_0}}) \end{pmatrix} \end{aligned}$$

in Verteilung in $\ell^\infty(\mathcal{G}_D)$. Schließlich führt der Satz über stetige Abbildungen (vgl. Satz 18.11 in van der Vaart [61]) zum gewünschten Resultat. Die Folge

$$\left\{ \left[\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0}) + \sqrt{n}\left(\sqrt{\mathbb{P}_n(\mathbb{1}_D)} - \sqrt{\mathbb{E}_{\theta_0}(\mathbb{1}_D)}\right) \cdot \left(-2\,\mathbb{E}_{\theta_0}/\sqrt{\alpha_{\theta_0}}\right) \right] (g_{x,t,D}) : g_{x,t,D} \in \mathcal{G}_D \right\}$$

konvergiert unter H_0 in Verteilung in $\ell^\infty(\mathcal{G}_D)$ gegen den Gauß-Prozess

$$g_{x,t,D} \mapsto \mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) = \mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D)\mathbb{E}_{\theta_0}(g_{x,t,D})/\alpha_{\theta_0}, \quad g_{x,t,D} \in \mathcal{G}_D.$$

■

Bemerkung 6.28.
Der Gauß-Prozess $\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}$ hat den Erwartungswert $\mathbb{E}_{\theta_0}[\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x,t,D})] = 0$ und die Kovarianz von $\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x_1,t_1,D})$ und $\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x_2,t_2,D})$ vereinfacht sich zu

$$\begin{aligned}
&\mathbb{E}_{\theta_0}(g_{\min(x_1,x_2),\min(t_1,t_2),D}) - \mathbb{E}_{\theta_0}(g_{x_1,t_1,D})\mathbb{E}_{\theta_0}(g_{x_2,t_2,D})/\alpha_{\theta_0} \\
&= \mathbb{P}_{\theta_0}\Big((X_1,T_1)^{\mathrm{T}} \in [0,\min\{x_1,x_2\}] \times [0,\min\{t_1,t_2\}] \cap D\Big) \\
&- \mathbb{P}_{\theta_0}\Big((X_1,T_1)^{\mathrm{T}} \in B_1 \cap D\Big) \cdot \mathbb{P}_{\theta_0}\Big((X_1,T_1)^{\mathrm{T}} \in B_2 \cap D\Big) \cdot \frac{1}{\alpha_{\theta_0}}
\end{aligned} \tag{6.18}$$

mit $B_i = [0,x_i] \times [0,t_i]$ für $i = 1,2$. Der Abschnitt A.2 beinhaltet die detaillierte Berechnung der Kovarianz. □

Der Satz über stetige Abbildungen impliziert die entsprechende Konvergenz der Folge in der Supremumsnorm $\|\cdot\|_{\mathcal{G}_D}$, d.h. die Konvergenz der modifizierten Kolmogorov-Smirnov-Teststatistik $\sqrt{\hat{n}}d_n(\theta_0,\hat{n})$. Die Verteilung von $\|\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ kann wie folgt simuliert werden.

Algorithmus 3: Simulation der Verteilung von $\|\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$

1. Erzeuge ein Gitter auf dem Rechteck $[0,G+s] \times [0,G]$.
2. Berechne für alle Gitterpunkte die Kovarianzmatrix nach (6.18).
3. Wiederhole für eine festgelegte Anzahl von Iterationen:
   i) Erzeuge einen normalverteilten Vektor mit Erwartungswert null und der berechneten Kovarianzmatrix.
   ii) Ermittle das betragsmäßige Maximum des Vektors.
4. Wähle aus den sortierten Maxima das entsprechende Quantil aus.

6.2.4. Verteilung der Teststatistik bei $(\hat{\theta}_n, \hat{n})$

Schließlich kann eine Verknüpfung der beiden vorangegangenen Stufen hergestellt werden, sodass nun sowohl die Schätzung des Parameters der Exponentialverteilung als auch des latenten Stichprobenumfangs berücksichtigt werden. Dies hat zur Folge, dass nun auch die Auswahlwahrscheinlichkeit in $\hat{n} := M_n/\alpha_{\hat{\theta}_n}$ von $\hat{\theta}_n$ abhängt und $\hat{n}$ im Vergleich zum vorherigen Abschnitt neu definiert wird. Die zu untersuchenden Hypothesen entsprechen denen aus Gleichung (6.8).
Die Teststatistik $\sqrt{\hat{n}}d_n(\hat{\theta}_n,\hat{n})$ für diese vierte Stufe lautet

$$\begin{aligned}
&\sqrt{\hat{n}} \sup_{(x,t)\in S} \left| \frac{\alpha_{\hat{\theta}_n}}{M_n} \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}\Big([0,x] \times [0,t] \cap D\Big) - \mathbb{P}_{\hat{\theta}_n}\Big((X_i,T_i)^{\mathrm{T}} \in [0,x] \times [0,t] \cap D\Big) \right| \\
&= \sup_{(x,t)\in S} \left| \sqrt{\hat{n}} \frac{\alpha_{\hat{\theta}_n}}{M_n} \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}\Big([0,x] \times [0,t] \cap D\Big) - \sqrt{\hat{n}}\alpha_{\hat{\theta}_n} F^{\tilde{X},\tilde{T}}_{\hat{\theta}_n}(x,t) \right| \\
&= \sup_{(x,t)\in S} \left| \frac{\sqrt{\alpha_{\hat{\theta}_n}}}{\sqrt{M_n}} \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}\Big([0,x] \times [0,t] \cap D\Big) - \sqrt{M_n}\sqrt{\alpha_{\hat{\theta}_n}} F^{\tilde{X},\tilde{T}}_{\hat{\theta}_n}(x,t) \right|.
\end{aligned}$$

Zur Bestimmung der asymptotischen Verteilung des absolut zu maximierenden Arguments

$$\frac{\sqrt{\alpha_{\hat{\theta}_n}}}{\sqrt{M_n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{M_n}\sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t), \tag{6.19}$$

wird ebenfalls eine Zerlegung vorgenommen. Diese wird so gewählt, dass der resultierende Ausdruck möglichst kurz gehalten ist. In Bemerkung 6.35 wird eine intuitivere Zerlegung vorgestellt, die sich an den vorherigen Stufen orientiert. Beide Varianten führen jedoch zum selben Grenzprozess. Die einzufügenden Terme werden nachfolgend aufgeführt:

$$\begin{aligned}
&\pm \left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\hat{\theta}_n} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) \right) \\
&\pm \sqrt{\alpha_{\theta_0}} \left(\frac{1}{\sqrt{M_n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) \right)
\end{aligned}$$

Anschließendes Ausklammern von $N_{n,D}$ und $F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}$ führt zu

$$\begin{aligned}
&\left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\hat{\theta}_n} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) \right) \\
&+ \frac{1}{\sqrt{M_n}} N_{n,D}\Big([0,x]\times[0,t]\Big) \left[\Big(\sqrt{\alpha_{\hat{\theta}_n}} - \sqrt{\alpha_{\theta_0}}\Big) - \left(\frac{\sqrt{M_n}}{\sqrt{n}} - \sqrt{\alpha_{\theta_0}} \right) \right] \\
&+ \sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) \left[\Big(\sqrt{n}\sqrt{\alpha_{\hat{\theta}_n}} - \sqrt{n}\sqrt{\alpha_{\theta_0}}\Big) - \Big(\sqrt{M_n} - \sqrt{n}\sqrt{\alpha_{\theta_0}}\Big) \right].
\end{aligned}$$

Es erfolgt die Erweiterung des zweiten Summanden mit $\sqrt{n}\sqrt{n}/n$, woraufhin sich der vorherige Ausdruck zu

$$\begin{aligned}
&\underbrace{\left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\hat{\theta}_n} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) \right)}_{1} \\
&+ \left[\underbrace{\sqrt{n}\left(\sqrt{\alpha_{\hat{\theta}_n}} - \sqrt{\alpha_{\theta_0}}\right)}_{2} - \underbrace{\left(\sqrt{M_n} - \sqrt{n}\sqrt{\alpha_{\theta_0}}\right)}_{3} \right] \\
&\quad \cdot \underbrace{\left[\frac{\sqrt{n}}{\sqrt{M_n}} \frac{1}{n} N_{n,D}\Big([0,x]\times[0,t]\Big) + \sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) \right]}_{4}
\end{aligned} \tag{6.20}$$

umschreiben lässt. Die Verteilungskonvergenz von Teil 1 in der Darstellung (6.20) ist Inhalt von Satz 6.23. Im vorangegangenen Abschnitt wurde der dritte Teil unter Anwendung der Delta-Methode in Gleichung (6.15) analysiert. Die Untersuchung der übrigen Elemente der Darstellung erfolgt in den nachfolgenden Lemmata.

Lemma 6.29.
Die Folge $\sqrt{n}(\sqrt{\alpha_{\hat{\theta}_n}} - \sqrt{\alpha_{\theta_0}})$ *ist asymptotisch linear mit der Darstellung*

$$-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}(X_1, T_1)])^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\theta_0}(X_i, T_i) \cdot \frac{\dot{\alpha}_{\theta_0}}{2\sqrt{\alpha_{\theta_0}}} + o_{\mathbb{P}_{\theta_0}}(1) \tag{6.21}$$

und asymptotisch normalverteilt mit Erwartungswert null und Varianz $-\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}(X_1, T_1)]^{-1} \dot{\alpha}_{\theta_0}^2 / (4\alpha_{\theta_0})$.

Beweis. Nach Korollar 1 in Weißbach und Wied [65] ist die Abbildung $\theta \mapsto \alpha_\theta$ differenzierbar in θ_0. Wegen $\alpha_\theta \in (0,1)$ ist auch $\theta \mapsto \sqrt{\alpha_\theta}$ differenzierbar in θ_0 und es folgt

$$\sqrt{\alpha_\theta} - \sqrt{\alpha_{\theta_0}} - (\theta - \theta_0) \cdot \frac{\dot{\alpha}_{\theta_0}}{2\sqrt{\alpha_{\theta_0}}} = o(|\theta - \theta_0|).$$

In Satz 1 zeigen Weißbach und Wied [65] die Konsistenz $\hat{\theta}_n \to \theta_0$. Mit Lemma 2.12 aus van der Vaart [61] wird auch

$$\sqrt{\alpha_{\hat{\theta}_n}} - \sqrt{\alpha_{\theta_0}} = (\hat{\theta}_n - \theta_0) \cdot \frac{\dot{\alpha}_{\theta_0}}{2\sqrt{\alpha_{\theta_0}}} + o_{\mathbb{P}_{\theta_0}}(|\hat{\theta}_n - \theta_0|)$$

ersichtlich. Aus der asymptotischen Linearität von $\sqrt{n}(\hat{\theta}_n - \theta_0)$ aus Gleichung (6.10) folgt die Darstellung (6.21) und die Folge ist asymptotisch normalverteilt mit dem oben genannten Erwartungswert und der Varianz. Dabei wurde verwendet, dass $o_{\mathbb{P}_{\theta_0}}(1) + o_{\mathbb{P}_{\theta_0}}(|\hat{\theta}_n - \theta_0|) = o_{\mathbb{P}_{\theta_0}}(1)$. ■

Bemerkung 6.30.
Die Herleitung der asymptotischen Normalität könnte auch mittels der Delta-Methode erreicht werden. Für spätere Zwecke ist jedoch die asymptotische Linearität der Folge von entscheidender Bedeutung. □

Das nächste Lemma liefert eine Approximation des vierten Terms in Gleichung (6.20).

Lemma 6.31.
Es gilt

$$\sup_{(x,t)\in D} \left| \frac{\sqrt{n}}{\sqrt{M_n}} \frac{1}{n} N_{n,D}\Big([0,x] \times [0,t]\Big) + \sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) - 2\sqrt{\alpha_{\theta_0}} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t) \right| \to 0$$

in Wahrscheinlichkeit für $n \to \infty$.

Beweis. Die Konvergenz kann in zwei Schritten gezeigt werden:

$$\begin{aligned} &\sup_{(x,t)\in D} \left| \frac{\sqrt{n}}{\sqrt{M_n}} \frac{1}{n} N_{n,D}\Big([0,x] \times [0,t]\Big) - \sqrt{\alpha_{\theta_0}} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t) \right| \\ &\quad + \sup_{(x,t)\in D} \left| \sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) - \sqrt{\alpha_{\theta_0}} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t) \right| \end{aligned}$$

Der erste Summand konvergiert nach Lemma 6.26 in Wahrscheinlichkeit für $n \to \infty$ gegen null. Im zweiten Schritt wird der zweite Summand betrachtet. Zunächst ist festzuhalten, dass $\sqrt{\alpha_\theta} F_\theta^{\tilde{X},\tilde{T}}(x,t) = \mathbb{E}_\theta(g_{x,t,D})/\sqrt{\alpha_\theta}$ gilt. Nach Lemma 6.22 ist die Abbildung $\theta \in [\varepsilon_\theta, 1/\varepsilon_\theta] \mapsto \mathbb{E}_\theta \in \ell^\infty(\mathcal{G}_D)$ Fréchet-differenzierbar in θ_0 und somit auch stetig in θ_0, d.h.

$$\|\mathbb{E}_\theta - \mathbb{E}_{\theta_0}\|_{\mathcal{G}_D} \to 0, \quad \text{für } \theta \to \theta_0.$$

Die Abbildung $\theta \mapsto 1/\sqrt{\alpha_\theta}$ ist stetig in θ_0 und als Abbildung nach $\ell^\infty(\mathcal{G}_D)$ konstant. Die Komposition $\theta \mapsto \mathbb{E}_\theta/\sqrt{\alpha_\theta}$ ist demnach ebenfalls stetig in θ_0. Weißbach und Wied weisen in Satz 1 in [65] die Konsistenz der Folge $\hat{\theta}_n \to \theta_0$ nach. Der Satz über stetige Abbildungen (vgl. Satz 18.11 in van der Vaart [61]) führt schließlich zu

$$\sup_{(x,t)\in D} \left| \sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) - \sqrt{\alpha_{\theta_0}} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t) \right| \to 0$$

in Wahrscheinlichkeit für $n \to \infty$. ■

Nach dieser ersten Annäherung an die Idee der Zerlegung soll nun der Beweis der Konvergenz des Prozesses (6.19) bzw. (6.20) gegen einen Gauß-Prozess einer detaillierten Betrachtung unterzogen werden.

Satz 6.32.
Es sei auf die bereits angeführte Definition $\phi_{\theta_0} = -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1}\psi_{\theta_0}$ verwiesen. Die Folge

$$\left\{ \frac{\sqrt{\alpha_{\hat{\theta}_n}}}{\sqrt{M_n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{M_n}\sqrt{\alpha_{\hat{\theta}_n}} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) : (x,t) \in S \right\}$$

konvergiert in Verteilung in $\ell^\infty(\mathcal{G}_D)$ unter H_0 gegen den Gauß-Prozess

$$\begin{aligned} g_{x,t,D} \mapsto \mathbb{H}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) := {} & \mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0})\dot{\mathbb{E}}_{\theta_0}(g_{x,t,D}) \\ & + \Big[\mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0})\cdot\dot{\alpha}_{\theta_0} - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D)\Big]\cdot\frac{\mathbb{E}_{\theta_0}(g_{x,t,D})}{\alpha_{\theta_0}} \end{aligned}$$

für $g_{x,t,D} \in \mathcal{G}_D$.

Beweis. Der Ausdruck (6.19) kann wie oben beschrieben in (6.20) überführt werden. Der nächste Schritt umfasst verschiedene Approximationen im Sinne der $\|\cdot\|_\infty$-Norm bzw. nach Wechsel in die funktionenwertige Notation im Sinne der $\ell^\infty(\mathcal{G}_\mathcal{D})$-Norm. Dafür findet zunächst das Lemma 6.31 Anwendung auf Teil 4 in Ausdruck (6.20):

$$\begin{aligned} (6.19) = (6.20) = {} & \left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\hat{\theta}_n} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t) \right) \\ & + \left[\sqrt{n}\left(\sqrt{\alpha_{\hat{\theta}_n}} - \sqrt{\alpha_{\theta_0}}\right) - \left(\sqrt{M_n} - \sqrt{n}\sqrt{\alpha_{\theta_0}}\right) \right] \cdot \left[2\sqrt{\alpha_{\theta_0}} F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t) \right] + o_{\mathbb{P}_{\theta_0}}(1) \end{aligned}$$

Die zweite Approximation bezieht sich auf den zweiten Teil in Ausdruck (6.20) und

nutzt das Lemma 6.29:

$$\begin{aligned}(6.19) = &\left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\hat{\theta}_n} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t)\right)\\ &+\Big[-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}(X_1,T_1)])^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i,T_i)\cdot\frac{\dot{\alpha}_{\theta_0}}{2\sqrt{\alpha_{\theta_0}}} - \left(\sqrt{M_n}-\sqrt{n}\sqrt{\alpha_{\theta_0}}\right)\Big]\\ &\cdot\Big[2\sqrt{\alpha_{\theta_0}}F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\Big] + o_{\mathbb{P}_{\theta_0}}(1)\end{aligned}$$

Die dritte und letzte Approximation ist Teil 1 des Ausdrucks (6.20) zuzuordnen. Wie bereits erwähnt, wurde die Verteilungskonvergenz dieses Ausdrucks in Satz 6.23 bewiesen. Der Beweis des Satzes basiert auf der in Gleichung (6.9) dargestellten Zerlegung. Anschließend wird die Fréchet-Differenzierbarkeit der Funktion $\theta \mapsto \mathbb{E}_\theta$ aus Lemma 6.22 sowie die asymptotische Linearität der Schätzfolge $\sqrt{n}(\hat{\theta}_n - \theta_0)$ aus Gleichung (6.10) verwendet. Zur Wahrung der Übersicht wird für diesen Approximationsschritt in die funktionenwertige Notation gewechselt. In diesem Sinne gilt nun

$$\begin{aligned}(6.19) =&\sqrt{n}\Big(\mathbb{P}_n(g_{x,t,D}) - \mathbb{E}_{\theta_0}(g_{x,t,D})\Big) - \sqrt{n}\Big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\Big)\cdot\dot{\mathbb{E}}_{\theta_0}\\ &+\left[\sqrt{n}\Big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\Big)\cdot\frac{\dot{\alpha}_{\theta_0}}{2\sqrt{\alpha_{\theta_0}}} - \sqrt{n}\left(\sqrt{\mathbb{P}_n(\mathbb{1}_D)} - \sqrt{\mathbb{E}_{\theta_0}(\mathbb{1}_D)}\right)\right] \qquad (6.22)\\ &\cdot\Big(2\,\mathbb{E}_{\theta_0}(g_{x,t,D})/\sqrt{\alpha_{\theta_0}}\Big) + o_{\mathbb{P}_{\theta_0}}(1).\end{aligned}$$

Für die Umformung in die funktionenwertige Notation sei angemerkt, dass

$$-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}(X_1,T_1)])^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i,T_i) = \sqrt{n}\Big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\Big),$$

wobei $\mathbb{E}_{\theta_0}(\phi_{\theta_0}) = 0$ und $\phi_{\theta_0} = -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1}\psi_{\theta_0}$ genutzt wird.
Nach den Ausführungen, die kurz vor Satz 6.20 vorgenommen wurden, ist die Klasse $\mathcal{G}_D$ eine Donsker-Klasse. Weiterhin gilt mit der Definition von ψ_{θ_0} in Gleichung (4) in Weißbach und Wied [65], dass $\|\phi_{\theta_0}\|_\infty < \infty$. Damit ist auch $\{\phi_{\theta_0}\}$ als endliche Funktionenklasse eine Donsker-Klasse. Nach Beispiel 2.10.7 in van der Vaart und Wellner [62] überträgt sich die Donsker-Eigenschaft zweier Klassen auf deren Vereinigung, sodass auch $\mathcal{G}_D^\phi := \mathcal{G}_D \cup \{\phi_{\theta_0}\}$ diese besitzt. Demnach konvergiert die Folge $\{\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0})(g_{x,t,D}) : g_{x,t,D} \in \mathcal{G}_D^\phi\}$ in Verteilung in $\ell^\infty(\mathcal{G}_D^\phi)$ gegen eine $\mathbb{P}_{\theta_0}$-Brownsche Brücke $\mathbb{B}_{\mathbb{P}_{\theta_0}}$. Die Abbildung

$$f : \Big(\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0})\Big) \mapsto \begin{pmatrix}\sqrt{n}\Big(\mathbb{P}_n - \mathbb{E}_{\theta_0}\Big)\\ \sqrt{n}\Big(\mathbb{P}_n(\mathbb{1}_D) - \mathbb{E}_{\theta_0}(\mathbb{1}_D)\Big)\\ \sqrt{n}\Big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\Big)\end{pmatrix}$$

mit $(\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0})) \in \ell^\infty(\mathcal{G}_D^\phi)$ ist stetig. Daher kann der Satz über stetige Abbildungen angewendet werden, sodass $\{f\Big(\sqrt{n}(\mathbb{P}_n - \mathbb{E}_{\theta_0})\Big)(g_{x,t,D}) : g_{x,t,D} \in \mathcal{G}_D^\phi\}$ in Verteilung in

$\ell^\infty(\mathcal{G}_D^\phi)$ gegen $f(\mathbb{B}_{\mathbb{P}_{\theta_0}})$ konvergiert, d.h.

$$\begin{pmatrix} \sqrt{n}\big(\mathbb{P}_n - \mathbb{E}_{\theta_0}\big) \\ \sqrt{n}\big(\mathbb{P}_n(\mathbb{1}_D) - \mathbb{E}_{\theta_0}(\mathbb{1}_D)\big) \\ \sqrt{n}\big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\big) \end{pmatrix} \to \begin{pmatrix} \mathbb{B}_{\mathbb{P}_{\theta_0}} \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D) \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \end{pmatrix}$$

in Verteilung in $\ell^\infty(\mathcal{G}_D^\phi)$ für $n \to \infty$. Mit Blick auf Gleichung (6.22) wird deutlich, dass auch in diesem Fall die Ergänzung der fehlenden Transformationen durch die Wurzelfunktion die Anwendung der Delta-Methode erforderlich macht. Es wird dafür die Funktion

$$\phi_\Delta : (\ell^\infty(\mathcal{G}_D^\phi))^3 \to (\ell^\infty(\mathcal{G}_D^\phi))^3, \quad \phi_\Delta(x, y, z) = (x, \sqrt{y}, z)^{\mathrm{T}}$$

betrachtet, welche nach Lemma 6.18, entsprechend auf drei Dimensionen erweitert, Hadamard-differenzierbar ist mit

$$\phi'_\Delta(x, y, z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (2\sqrt{y})^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Nach Satz 6.19 (funktionale Delta-Methode) konvergiert nun

$$\begin{aligned} &\sqrt{n}\left(\phi_\Delta\begin{pmatrix} \mathbb{P}_n \\ \mathbb{P}_n(\mathbb{1}_D) \\ \mathbb{P}_n(\phi_{\theta_0}) \end{pmatrix} - \phi_\Delta\begin{pmatrix} \mathbb{E}_{\theta_0} \\ \mathbb{E}_{\theta_0}(\mathbb{1}_D) \\ \mathbb{E}_{\theta_0}(\phi_{\theta_0}) \end{pmatrix}\right) = \begin{pmatrix} \sqrt{n}\big(\mathbb{P}_n - \mathbb{E}_{\theta_0}\big) \\ \sqrt{n}\left(\sqrt{\mathbb{P}_n(\mathbb{1}_D)} - \sqrt{\mathbb{E}_{\theta_0}(\mathbb{1}_D)}\right) \\ \sqrt{n}\big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\big) \end{pmatrix} \\ &\to \phi'_\Delta\begin{pmatrix} \mathbb{E}_{\theta_0} \\ \mathbb{E}_{\theta_0}(\mathbb{1}_D) \\ \mathbb{E}_{\theta_0}(\phi_{\theta_0}) \end{pmatrix} \cdot \begin{pmatrix} \mathbb{B}_{\mathbb{P}_{\theta_0}} \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D) \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \end{pmatrix} = \begin{pmatrix} \mathbb{B}_{\mathbb{P}_{\theta_0}} \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D)/(2\sqrt{\alpha_{\theta_0}}) \\ \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \end{pmatrix} \end{aligned}$$

in Verteilung in $\ell^\infty(\mathcal{G}_D^\phi)$ für $n \to \infty$. Die drei Komponenten des Vektors können schließlich über den Satz über stetige Abbildungen (vgl. Satz 18.11 in van der Vaart [61]) kombiniert werden, sodass der Ausdruck in Gleichung (6.22) entsteht. Die Folge

$$\begin{aligned} \Bigg\{\Bigg[&\sqrt{n}\big(\mathbb{P}_n - \mathbb{E}_{\theta_0}\big) - \sqrt{n}\big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\big) \cdot \dot{\mathbb{E}}_{\theta_0} \\ &+ \left[\sqrt{n}\big(\mathbb{P}_n(\phi_{\theta_0}) - \mathbb{E}_{\theta_0}(\phi_{\theta_0})\big) \cdot \frac{\dot{\alpha}_{\theta_0}}{2\sqrt{\alpha_{\theta_0}}} - \sqrt{n}\left(\sqrt{\mathbb{P}_n(\mathbb{1}_D)} - \sqrt{\mathbb{E}_{\theta_0}(\mathbb{1}_D)}\right)\right] \\ &\cdot \left(2\,\mathbb{E}_{\theta_0}/\sqrt{\alpha_{\theta_0}}\right)\Bigg](g_{x,t,D}) : g_{x,t,D} \in \mathcal{G}_D^\phi\Bigg\} \end{aligned}$$

konvergiert unter H_0 in Verteilung in $\ell^\infty(\mathcal{G}_D^\phi)$ und somit auch in $\ell^\infty(\mathcal{G}_D)$ (ähnlich wie

in Satz 19.23 in [61]) gegen den Gauß-Prozess

$$g_{x,t,D} \mapsto \mathbb{H}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) := \mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0})\dot{\mathbb{E}}_{\theta_0}(g_{x,t,D}) + \left[\mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \dot{\alpha}_{\theta_0} - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D)\right] \cdot \frac{\mathbb{E}_{\theta_0}(g_{x,t,D})}{\alpha_{\theta_0}}$$

für $g_{x,t,D} \in \mathcal{G}_D$. ■

Bemerkung 6.33.
Der Gauß-Prozess $\mathbb{H}_{\mathbb{P}_{\theta_0}}$ hat den Erwartungswert $\mathbb{E}_{\theta_0}[\mathbb{H}_{\mathbb{P}_{\theta_0}}(g_{x,t,D})] = 0$ und die Kovarianz von $\mathbb{H}_{\mathbb{P}_{\theta_0}}(g_{x_1,t_1,D})$ und $\mathbb{H}_{\mathbb{P}_{\theta_0}}(g_{x_2,t_2,D})$ lautet

$$\begin{aligned}
&\mathbb{E}_{\theta_0}(g_{\min(x_1,x_2),\min(t_1,t_2),D}) - \mathbb{E}_{\theta_0}(g_{x_1,t_1,D})\mathbb{E}_{\theta_0}(g_{x_2,t_2,D})/\alpha_{\theta_0} \\
&+ (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \cdot \dot{\mathbb{E}}_{\theta_0}(g_{x_1,t_1,D})\dot{\mathbb{E}}_{\theta_0}(g_{x_2,t_2,D}) \\
&- (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \cdot \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \cdot \left[\dot{\mathbb{E}}_{\theta_0}(g_{x_2,t_2,D})\mathbb{E}_{\theta_0}(g_{x_1,t_1,D}) + \dot{\mathbb{E}}_{\theta_0}(g_{x_1,t_1,D})\mathbb{E}_{\theta_0}(g_{x_2,t_2,D})\right] \\
&+ (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \cdot \frac{\dot{\alpha}_{\theta_0}^2}{\alpha_{\theta_0}^2} \cdot \mathbb{E}_{\theta_0}(g_{x_1,t_1,D})\mathbb{E}_{\theta_0}(g_{x_2,t_2,D})
\end{aligned} \tag{6.23}$$

Ausführliche Erläuterungen zur Berechnung der Kovarianz befinden sich in Abschnitt A.2. □

Aus dem Satz über stetige Abbildungen folgt die entsprechende Konvergenz der Folge in der Supremumsnorm $\|\cdot\|_{\mathcal{G}_D}$, d.h. die Konvergenz der modifizierten Kolmogorov-Smirnov-Teststatistik $\sqrt{\hat{n}}d_n(\hat{\theta}_n, \hat{n})$. Die Simulation der Verteilung von $\|\mathbb{H}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ kann wie folgt durchgeführt werden.

Algorithmus 4: Simulation der Verteilung von $\|\mathbb{H}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$

1. Erzeuge ein Gitter auf dem Rechteck $[0, G+s] \times [0, G]$.
2. Berechne für alle Gitterpunkte die Kovarianzmatrix nach (6.23).
3. Wiederhole für eine festgelegte Anzahl von Iterationen:
   i) Erzeuge einen normalverteilten Vektor mit Erwartungswert null und der berechneten Kovarianzmatrix.
   ii) Ermittle das betragsmäßige Maximum des Vektors.
4. Wähle aus den sortierten Maxima das entsprechende Quantil aus.

Bemerkung 6.34.
Der Prozess $\mathbb{H}_{\mathbb{P}_{\theta_0}}$ steht in folgendem Zusammenhang zu den vorangegangenen drei Grenzprozessen:

$$\begin{aligned}\mathbb{H}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) &= \mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) + \mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) \\ &\quad - \left(\mathbb{B}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \mathbb{E}_{\theta_0}(g_{x,t,D}) \cdot \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}}\right) \\ &= \mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) + \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \left[\frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}}\mathbb{E}_{\theta_0}(g_{x,t,D}) - \dot{\mathbb{E}}(g_{x,t,D})\right] \\ &= \mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}(g_{x,t,D}) + \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \mathbb{E}_{\theta_0}[g_{x,t,D}\psi_{\theta_0}]\end{aligned} \tag{6.24}$$

Für den letzten Schritt wurde $\psi_{\theta_0} = \dot{f}_{\theta_0}/f_{\theta_0} - \dot{\alpha}_{\theta_0}/\alpha_{\theta_0}$ auf D (ähnlich wie in Gleichung (A.1)) verwendet. □

Bemerkung 6.35.
Zur Herleitung der asymptotischen Verteilung könnte der Ausdrucks (6.19) umfangreicher zerlegt werden. Für eine alternative Zerlegung könnten die Prozesse der drei vorherigen Stufen herangezogen werden, d.h.

$$\begin{aligned}&\pm \left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\theta_0} F^{\tilde{X},\tilde{T}}_{\theta_0}(x,t)\right) \\ &\pm \left(\frac{1}{\sqrt{n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{n}\alpha_{\hat{\theta}_n} F^{\tilde{X},\tilde{T}}_{\hat{\theta}_n}(x,t)\right) \\ &\pm \sqrt{\alpha_{\theta_0}}\left(\frac{1}{\sqrt{M_n}} N_{n,D}\Big([0,x]\times[0,t]\Big) - \sqrt{M_n}\sqrt{\alpha_{\theta_0}} F^{\tilde{X},\tilde{T}}_{\theta_0}(x,t)\right).\end{aligned}$$

Im Ergebnis würde man direkt zu der ersten Darstellung in Gleichung (6.24) gelangen. □

Die nachfolgende Tabelle fasst die Resultate der Untersuchungen zu den Grenzprozessen der vier Stufen zusammen.

Tabelle 6.1.: Übersicht über die vier Stufen und die zugehörigen Grenzprozesse

Teststatistik	Satz	Gauß-Prozess	
		Symbol	Darstellung
$\sqrt{n}d(\theta_0, n)$	6.20	$\mathbb{B}_{\mathbb{P}_{\theta_0}}$	$\mathbb{B}_{\mathbb{P}_{\theta_0}}$
$\sqrt{n}d(\hat{\theta}_n, n)$	6.23	$\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}$	$\mathbb{B}_{\mathbb{P}_{\theta_0}} - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \dot{\mathbb{E}}_{\theta_0}$
$\sqrt{\hat{n}}d(\theta_0, \hat{n})$	6.27	$\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}$	$\mathbb{B}_{\mathbb{P}_{\theta_0}} - \mathbb{B}_{\mathbb{P}_{\theta_0}}(\mathbb{1}_D) \cdot \mathbb{E}_{\theta_0}/\alpha_{\theta_0}$
$\sqrt{\hat{n}}d(\hat{\theta}_n, \hat{n})$	6.32	$\mathbb{H}_{\mathbb{P}_{\theta_0}}$	$\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}} + \mathbb{B}_{\mathbb{P}_{\theta_0}}(\phi_{\theta_0}) \cdot \left[\frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}}\mathbb{E}_{\theta_0} - \dot{\mathbb{E}}_{\theta_0}\right]$

6.2.5. Algorithmus zur Berechnung der Prüfgröße

Im Falle einer eindimensionalen Stichprobe erfolgen die Sprünge der empirischen Verteilungsfunktion lediglich in den beobachteten Punkten, sodass sich die Kolmogorov-Smirnov-Teststatistik durch das Auswerten des Abstands der empirischen und der theoretischen Verteilung in diesen Punkten ergibt (vgl. Comment 27, S. 572 in [25] sowie Abschnitt 4.2.1 in [4]). Bereits im bivariaten Fall liegen jedoch unendlich viele Sprünge vor. In ihrer Arbeit [27] stellen Justel, Peña und Zamar eine Prozedur zur Berechnung der Kolmogorov-Smirnov-Statistik im zweidimensionalen Fall vor, bei welcher die Auswertung nur in endlich vielen Punkten erforderlich ist. Die wesentlichen Aussagen aus Lemma 1 und 2 sowie Satz 1 aus [27] werden in angepasster Notation wiedergegeben, deren Herleitungen skizziert und diese um eigene Abbildungen ergänzt. Zur Vereinfachung wird der Anpassungstest als Test auf die Hypothese zweier unabhängiger Gleichverteilungen vorgestellt. Eine der zugrundeliegenden Annahmen ist, dass der Träger der zwei Merkmale das Einheitsquadrat $[0,1]^2$ ist. Wie die Methode auf den Fall der unabhängigen Doppeltrunkation übertragen werden kann, wird im Anschluss erläutert.

Es sei $\mathbf{u}_1 = (u_{1,1}, u_{2,1}), \ldots, \mathbf{u}_{\text{no}} = (u_{1,\text{no}}, u_{2,\text{no}})$ die Realisation einer Zufallsstichprobe einer zweidimensionalen Gleichverteilung auf dem Einheitsquadrat mit unabhängigen Komponenten. Zur Abgrenzung vom latenten Stichprobenumfang n und der Anzahl der Beobachtungen M_n, wird der Stichprobenumfang hier mit no bezeichnet. Für die Berechnung der Prüfgröße werden gewisse Knotenpunkte benötigt. Ein Punkt $(u_{1,j}, u_{2,i})$ wird in diesem Zusammenhang als Knotenpunkt bezeichnet, falls $u_{1,i} < u_{1,j}$ und $u_{2,i} > u_{2,j}$, womit die Punkte $(u_{1,i}, u_{2,i})$ und $(u_{1,j}, u_{2,j})$ diskordant sind. In Abbildung 6.1 wurden beispielhaft die Knotenpunkte für sechs Beobachtungen ermittelt.

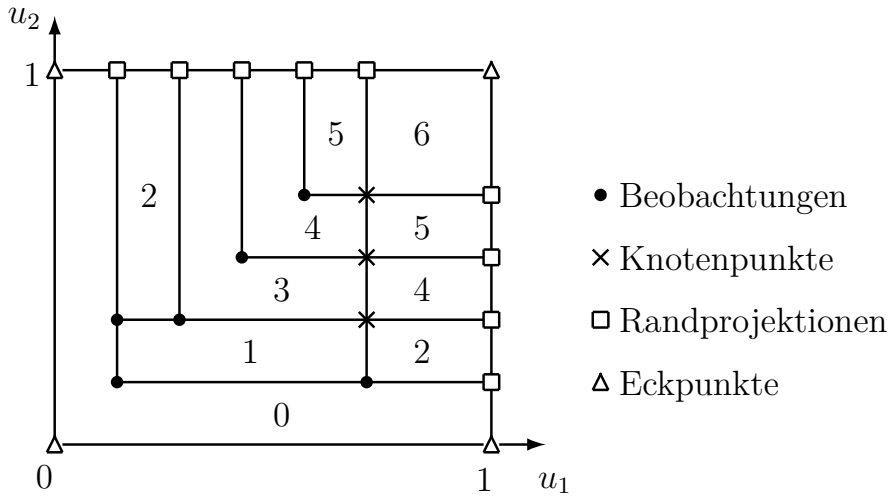

Abbildung 6.1.: Werte einer zweidimensionalen empirischen Verteilungsfunktion mit sechs Beobachtungen (·6) in Abhängigkeit der Beobachtungen, Knotenpunkte, Eckpunkte und Projektionen auf den oberen und rechten Rand

Für $\mathbf{u} = (u_1, u_2)$ sei $D^+_{\text{no}}(\mathbf{u}) := (G_{\text{no}}(\mathbf{u}) - G(\mathbf{u}))$ und $D^-_{\text{no}}(\mathbf{u}) := (G(\mathbf{u}) - G_{\text{no}}(\mathbf{u}^-))$. Hierbei bezeichnet G die Verteilungsfunktion zweier unabhängiger, gleichverteilter Zu-

fallsvariablen auf dem Intervall $[0,1]$ und G_{no} die empirische Verteilungsfunktion. Es sei bereits an dieser Stelle angedeutet, dass die Rolle von G als Verteilungsfunktion von $\mathbf{u}_i$ nicht in die Argumentation eingeht. Weiter sei $G_{\text{no}}(\mathbf{u}^-) := \lim_{\delta\to 0} G_{\text{no}}(u_1-\delta, u_2-\delta)$. Die Kolmogorov-Smirnov-Statistik berechnet sich dann als Maximum über die beiden Größen $D_{\text{no}}^+ := \sup_{\mathbf{u}} D_{\text{no}}^+(\mathbf{u})$ und $D_{\text{no}}^- := \sup_{\mathbf{u}} D_{\text{no}}^-(\mathbf{u})$. Aus den folgenden Lemmata geht nun hervor, in welchen Punkten D_{no}^+ und D_{no}^- zur Berechnung der Teststatistik ausgewertet werden müssen.
Zunächst kann D_{no}^+ durch das Maximum über die Beobachtungen und die oben eingeführten Knotenpunkte abgeschätzt werden. Dafür bezeichne

$$I := \{(u_{1,j}, u_{2,i}) |\, u_{1,i} \le u_{2,j},\ u_{2,i} \ge u_{2,j}; i,j = 0,1,\ldots,\text{no}\}.$$

Die Menge I enthält den Punkt $(0,0)$, die beobachteten Punkte sowie alle Knotenpunkte. Zu gegebenem $\mathbf{u} = (u_1,u_2) \in [0,1]^2$ wird der Punkt $(u_1^u, u_2^u) \in I$ durch $u_1^u = \max\{u_{1,i} |\, u_{1,i} \le u_1,\ u_{2,i} \le u_2\}$ und $u_2^u = \max\{u_{2,i} |\, u_{2,i} \le u_2,\ u_{1,i} \le u_1\}$ definiert. Aus der Definition von u_1^u folgt $G_{\text{no}}(u_1,u_2) = G_{\text{no}}(u_1^u, u_2)$. Ebenso ergibt sich aus der Definition von u_2^u die Gleichheit $G_{\text{no}}(u_1,u_2) = G_{\text{no}}(u_1, u_2^u)$. In der Menge $(u_1^u, u_1] \times (u_2^u, u_2]$ befinden sich keine Beobachtungen, da dies den Definitionen von u_1^u und u_2^u widersprechen würde. Es gilt $G_{\text{no}}(u_1^u, u_2^u) = G_{\text{no}}(u_1^u, u_2) + G_{\text{no}}(u_1, u_2^u) - G_{\text{no}}(u_1,u_2)$ und somit insgesamt $G_{\text{no}}(u_1^u, u_2^u) = G_{\text{no}}(u_1,u_2)$. Die nachfolgende Abbildung 6.2 (links) veranschaulicht diesen Sachverhalt beispielhaft.

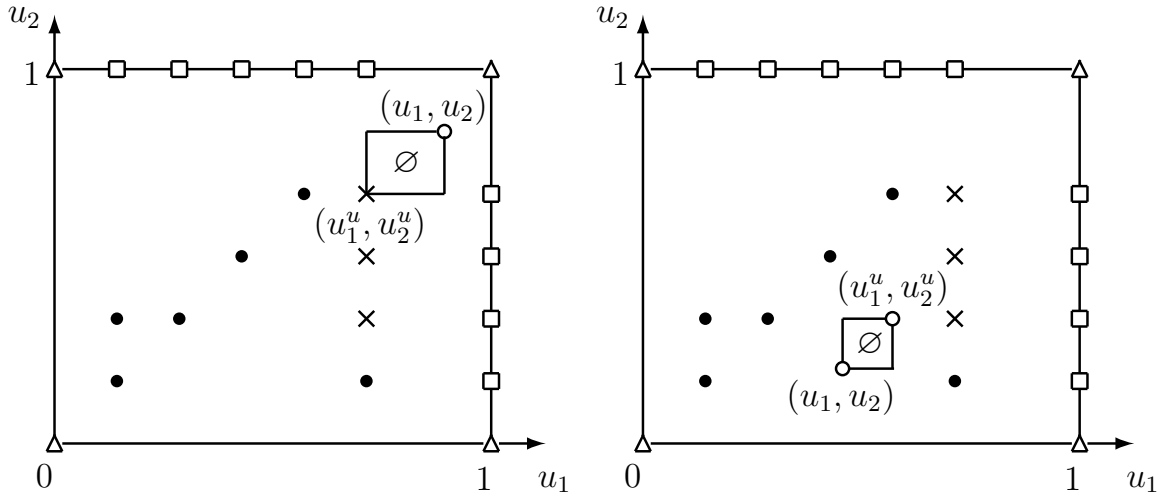

Abbildung 6.2.: Veranschaulichung der Herleitung von Lemma 6.36 (links) und Lemma 6.37 (rechts)

Daraus kann anschließend die Ungleichung

$$D_{\text{no}}^+(\mathbf{u}) = G_{\text{no}}(u_1,u_2) - G(u_1,u_2) \le G_{\text{no}}(u_1^u,u_2^u) - G(u_1^u,u_2^u) \le \max_{\mathbf{v}\in I} D_{\text{no}}^+(\mathbf{v})$$

und somit das Lemma hergeleitet werden.

Lemma 6.36.
Ist $u_{1,0} = u_{2,0} = 0$*, dann berechnet sich* D_{no}^+ *durch* $D_{\text{no}}^+ = \max_{\mathbf{v}\in I} D_{\text{no}}^+(\mathbf{v})$*.*

In einem zweiten Schritt wird nun D_{no}^- durch das Maximum über die Knotenpunkte und die Randprojektionen der Beobachtungen abgeschätzt. Es bezeichne

$$P := \{(u_{1,j}, u_{2,i}) |\, u_{1,j} > u_{1,i},\, u_{2,j} < u_{2,i};\, i,j = 0,1,\ldots,\text{no}+1\}.$$

Die Menge P enthält den Punkt $(1,1)$, alle Knotenpunkte und die Projektionen der beobachteten Punkte auf den oberen und rechten Rand des Einheitsquadrats.
Das Supremum D_{no}^- wird zunächst durch das Maximum über eine Menge Q, definiert durch $Q = P \cup \{(u_{1,j}, u_{2,i}) |\, u_{1,j} > u_{1,i},\, u_{2,j} > u_{2,i};\, i,j = 1,\ldots,\text{no}\}$, abgeschätzt. Mit $u_1^u = \min\{u_{1,i} |\, u_{1,i} > u_1\}$ und $u_2^u = \min\{u_{2,i} |\, u_{2,i} > u_2, u_{1,i} \le u_1\}$ wird dafür der Punkt $(u_1^u, u_2^u) \in Q$ definiert. Es gilt

$$\begin{aligned} D_{\text{no}}^-(\mathbf{u}) &= G(\mathbf{u}) - G_{\text{no}}(\mathbf{u}) < G(u_1^u, u_2^u) - G_{\text{no}}(\mathbf{u}) \\ &= G(u_1^u, u_2^u) - \lim_{\delta\to 0} G_{\text{no}}(u_1^u - \delta, u_2^u - \delta) + \lim_{\delta\to 0} G_{\text{no}}(u_1^u - \delta, u_2^u - \delta) - G_{\text{no}}(\mathbf{u}) \\ &\le \max_{\mathbf{v}\in Q}(G(\mathbf{v}) - G_{\text{no}}(\mathbf{v}^-)) + \lim_{\delta\to 0} G_{\text{no}}(u_1^u - \delta, u_2^u - \delta) - G_{\text{no}}(\mathbf{u}). \end{aligned}$$

Aus der Definition von (u_1^u, u_2^u) folgt $G_{\text{no}}(\mathbf{u}) = \lim_{\delta\to 0} G_{\text{no}}(u_1^u - \delta, u_2)$ sowie $G_{\text{no}}(\mathbf{u}) = \lim_{\delta\to 0} G_{\text{no}}(u_1, u_2^u - \delta)$. Aus der Abbildung 6.2 (rechts) geht hervor, dass sich in der Menge $(u_1, u_1^u) \times (u_2, u_2^u)$ keine Beobachtungen befinden. Dies impliziert auch $\lim_{\delta\to 0} G_{\text{no}}(u_1^u - \delta, u_2^u - \delta) = G_{\text{no}}(\mathbf{u})$. Weiterhin lässt sich zeigen, dass die Differenz $G(\mathbf{u}) - G_{\text{no}}(\mathbf{u}^-)$ für jeden Punkt $\mathbf{u} \in Q \backslash P$ durch das Maximum $\max_{\mathbf{v}\in P}(G(\mathbf{v}) - G_{\text{no}}(\mathbf{v}^-))$ abgeschätzt werden kann und, dass dieses Maximum auch angenommen wird, sodass $D_{\text{no}}^- = \max_{\mathbf{v}\in P}(G(\mathbf{v}) - G_{\text{no}}(\mathbf{v}^-))$.

Lemma 6.37.
Für $u_{1,0} = u_{2,\text{no}+1} = 0$ *und* $u_{2,0} = u_{1,\text{no}+1} = 1$*, ist* $D_{\text{no}}^- = \max_{\mathbf{v}\in P}(G(\mathbf{v}) - G_{\text{no}}(\mathbf{v}^-))$*.*

Die Kombination der vorangegangenen Lemmata resultiert schließlich in dem folgenden Satz.

Satz 6.38.
Im bivariaten Fall berechnet sich die Kolmogorov-Smirnov-Statistik (6.4) *durch*

$$\sqrt{\text{no}} \max_{\mathbf{u}\in I, \mathbf{v}\in P} \{G_{\text{no}}(\mathbf{u}) - G(\mathbf{u}), G(\mathbf{v}) - G_{\text{no}}(\mathbf{v}^-)\}.$$

Je nach Struktur der Stichprobe werden 3no bis $3\text{no} + \binom{\text{no}}{2}$ Auswertungen benötigt. Die Auswertungen für D_{no}^- lassen sich auf D_{no}^+ zurückführen. Für die Menge I der

Knotenpunkte gilt beispielsweise

$$\begin{aligned}\max_{\mathbf{v}\in I}(G(\mathbf{v}) - G_{\text{no}}(\mathbf{v}^-)) &= -\min_{\mathbf{v}\in I}(G_{\text{no}}(\mathbf{v}^-) - G(\mathbf{v})) \\ &= -\min_{\mathbf{v}\in I}\left[\frac{1}{\text{no}}\sum_{j=1}^{\text{no}}\epsilon_{\binom{u_{1,j}}{u_{2,j}}}\Big([0,v_1)\times[0,v_2)\Big) - G(\mathbf{v})\right] \\ &= \frac{2}{\text{no}} - \min_{\mathbf{v}\in I}\left[\frac{1}{\text{no}}\sum_{j=1}^{\text{no}}\epsilon_{\binom{u_{1,j}}{u_{2,j}}}\Big([0,v_1]\times[0,v_2]\Big) - G(\mathbf{v})\right] \\ &= \frac{2}{\text{no}} - \min_{\mathbf{v}\in I}(G_{\text{no}}(\mathbf{v}) - G_{\text{no}}(\mathbf{v})).\end{aligned}$$

Wegen $\mathbf{v} = (v_1, v_2) \in I$ gibt es zwei Beobachtungen $(u_{1,i}, u_{2,i})$ und $(u_{1,j}, u_{2,j})$ mit $u_{1,i} > u_{1,j}$, $u_{2,i} < u_{2,j}$, $u_{1,i} = v_1$ sowie $u_{2,j} = v_2$. Daher gilt sowohl für i als auch für j, dass

$$\epsilon_{\binom{u_{1,i}}{u_{2,i}}}\Big([0,v_1)\times[0,v_2)\Big) = \epsilon_{\binom{u_{1,i}}{u_{2,i}}}\Big([0,v_1]\times[0,v_2]\Big) - 1.$$

Dies kann auch auf die Randprojektionen übertragen werden, wobei dort jeweils nur eine Komponente betroffen ist.
Die Prüfgröße erhält man somit über die Berechnung der folgenden Werte.

Algorithmus 5: Prozedur zur Berechnung der Kolmogorov-Smirnov-Teststatistik

1. Berechne die den Beobachtungen zugehörigen Rand- sowie Knotenpunkte und daraus die folgenden Abstände:
2. $D^1_{\text{no}} := \max_{i=1,\dots,\text{no}} D^+_{\text{no}}(\mathbf{u}_i)$, der maximale Abstand in den beobachteten Punkten
3. $D^2_{\text{no}} := \max_{i,j=1,\dots,\text{no}}\{D^+_{\text{no}}(u_{1,j}, u_{2,i}) |\, u_{1,j} > u_{1,i},\, u_{2,j} < u_{2,i}\}$, der maximale Abstand in den Knotenpunkten,
4. $D^3_{\text{no}} := 2/\text{no} - \min_{i,j=1,\dots,\text{no}}\{D^+_{\text{no}}(u_{1,j}, u_{2,i}) |\, u_{1,j} > u_{1,i},\, u_{2,j} < u_{2,i}\}$, der minimale Abstand in den Knotenpunkten
5. $D^4_{\text{no}} := 1/\text{no} - \min_{i=1,\dots,\text{no}} D^+_{\text{no}}(1, u_{2,i})$, der maximale Abstand in den Projektionen der beobachteten Punkte auf den rechten Rand des Einheitsquadrats
6. $D^5_{\text{no}} := 1/\text{no} - \min_{i=1,\dots,\text{no}} D^+_{\text{no}}(u_{1,i}, 1)$, der maximale Abstand in den Projektionen der beobachteten Punkte auf den oberen Rand des Einheitsquadrats,
7. $D_{\text{no}} = \max\{D^1_{\text{no}}, D^2_{\text{no}}, D^3_{\text{no}}, D^4_{\text{no}}, D^5_{\text{no}}\}$, das Maximum all dieser Werte

Die Abbildung 6.1 zeigt beispielhaft sechs Beobachtungen und die mit sechs multiplizierten Werte der empirischen Verteilungsfunktion. Zusätzlich wurden alle weiteren zur Auswertung der Prüfgröße benötigten Punkte gekennzeichnet: die Knotenpunkten, die Randprojektionen sowie die Eckpunkte.

Die Prozedur kann nun direkt auf die abhängige Doppeltrunkation übertragen werden, da auch hier ein beschränkter Träger vorliegt. Bei Bekanntheit von θ_0 und n lautet die

Prüfgröße

$$\sqrt{n}\, d_n(\theta_0, n) = \sup_{(x,t)\in S} \frac{1}{\sqrt{n}} \Big| N_{n,D}\Big([0,x]\times[0,t]\Big) - \nu_{n,D}\Big([0,x]\times[0,t]\Big)\Big|.$$

Es lässt sich $d_n(\theta_0, n)$ auch schreiben als

$$d_n(\theta_0, n) = \sup_{(x,t)\in D} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_{\binom{X_i}{T_i}}\Big([0,x]\times[0,t]\cap D\Big) - \mathbb{P}_{\theta_0}\Big((X_i,T_i)^{\mathrm{T}} \in [0,x]\times[0,t]\cap D\Big)\right|.$$

Das Trunkieren der Werte außerhalb der Menge D hat zur Folge, dass das Supremum nur über D gebildet werden muss. Für die Auswertung sind demnach lediglich die Beobachtungen und Knotenpunkte in D sowie die Randpunkte auf dem Rand von D von Relevanz. Über $(\widetilde{T}_j + s, \widetilde{T}_j)$ und $(\widetilde{X}_j, \min(\widetilde{X}_j, G))$ gelangt man zu den Projektionen der Beobachtungen $(\widetilde{X}_j, \widetilde{T}_j)$ auf die jeweiligen Ränder des Parallelogramms.

Für den Fall, dass der latente Stichprobenumfang n unbekannt ist, wird dieser durch M_n/α_{θ_0} geschätzt. Die Prüfgröße lautet dann $\sqrt{\hat{n}} d_n(\theta_0, \hat{n})$ mit

$$d_n(\theta_0, \hat{n}) = \alpha_{\theta_0} \cdot \sup_{(x,t)\in D} \left| \frac{1}{M_n} \sum_{j=1}^{M_n} \epsilon_{\binom{\widetilde{X}_j}{\widetilde{T}_j}}\Big([0,x]\times[0,t]\Big) - F_{\theta_0}^{\tilde{X},\tilde{T}}(x,t)\right|.$$

Die Zufallsgrößen $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$ nehmen nur Werte in D an, sodass der Schnitt mit der Menge D vernachlässigt und das Supremum nur über D gesucht werden kann (vgl. Definition 3.4). Nachdem der maximale Wert über Algorithmus 5 ermittelt wurde, muss dieser mit der Auswahlwahrscheinlichkeit α_{θ_0} multipliziert werden, um $d_n(\theta_0, \hat{n})$ zu erhalten. Um den Wert der Prüfgröße als skalierte Version von $d_n(\theta_0, \hat{n})$ zu berechnen, muss weiterhin der Faktor $\sqrt{M_n/\alpha_{\theta_0}}$ berücksichtigt werden.

Wie bereits angedeutet, findet die Rolle von G als Verteilungsfunktion der der empirischen Verteilungsfunktion zugrunde liegenden Zufallsvariablen in der Herleitung keine Verwendung. Für den Fall, dass der Parameter θ geschätzt wird, kann somit problemlos $\mathbb{P}_{\hat{\theta}_n}\Big((X_i,T_i)^{\mathrm{T}} \in [0,x]\times[0,t]\cap D\Big)$ bzw. $\alpha_{\hat{\theta}_n} F_{\hat{\theta}_n}^{\tilde{X},\tilde{T}}(x,t)$ aufgrund von Monotonie- und Stetigkeitseigenschaften eingesetzt werden.

7. Anwendung auf Daten zur Unternehmensdemografie

In diesem Kapitel wird zunächst das in Abschnitt 3.1 eingeführte Modell auf einen Datensatz über Unternehmensinsolvenzen angewendet. Die Parameter des Modells werden geschätzt und die in Abschnitt 3.4 eingeführten Copulas werden im Vergleich betrachtet. Ziel ist es, die Unternehmensinsolvenzen hinsichtlich ihrer Zeitstabilität und Altershomogenität zu untersuchen. Zur Analyse werden der Test auf Unabhängigkeit aus Kapitel 5 sowie der Kolmogorov-Smirnov-Anpassungstest aus Kapitel 6 verwendet.

7.1. Beschreibung der Daten

Die Grundgesamtheit der empirischen Daten besteht aus Unternehmen mit Sitz in Deutschland, die zwischen dem 01.01.1990 und dem 31.12.2013 gegründet wurden. Die Länge der Gründungsperiode beträgt somit $G = 24$ Jahre. Der Beobachtungszeitraum mit einer Länge von $s = 3$ Jahren beginnt am 01.01.2014 und endet am 31.12.2016. Es bezeichne $T_i \in [0, 24]$ das Alter des i-ten Unternehmens zu Beginn der Beobachtungsperiode. Es wird angenommen, dass eine Gründung zu jedem Zeitpunkt innerhalb der Gründungsperiode gleich wahrscheinlich ist, sodass die Zufallsgrößen T_i gleichverteilt auf dem Intervall $[0, 24]$ sind (vgl. Annahme (A2)). Von Interesse ist nun das Ereignis der Insolvenz eines Unternehmens. Dafür bezeichne X_i das Alter des i-ten Unternehmens bei Insolvenz. Somit ergibt sich X_i als Länge der Zeitspanne zwischen dem Tag der beurkundeten Gründung eines Unternehmens und dem Tag der Beantragung der Insolvenz. Die Lebensdauer X_i eines Unternehmens sei exponentialverteilt zum unbekannten Parameter θ_0 (vgl. Annahme (A2)). Der Erwartungswert $1/\theta_0$ gibt somit Auskunft über die durchschnittliche Lebensdauer. Beobachtet werden nur diejenigen Unternehmen, welche innerhalb der Beobachtungsperiode Insolvenz beantragt haben (vgl. Annahme (A4)). Die somit doppelt trunkierte Stichprobe umfasst $m_n = 55.279$ Unternehmen. Die zugrunde liegenden Daten wurden durch die Insolvenzgerichte der Bundesrepublik Deutschland unter [39] veröffentlicht. Eben jener Datensatz wurde sowohl von Weißbach und Wied in [65] als auch von Weißbach und Dörre [64] verwendet. Für detailliertere Informationen bezüglich des Datensatzes wird auf die genannten Quellen verwiesen. Die Modellierung der Abhängigkeit zwischen X_i und T_i wird zu Beginn jedes Unterabschnitts beschrieben.

7.2. Test auf Zeitstabilität der Unternehmensinsolvenzen

Die Zeitstabilität der Unternehmensinsolvenzen soll durch die Untersuchung der Entwicklung der durchschnittlichen Unternehmenslebensdauer in Abhängigkeit vom Gründungsjahr analysiert werden. Mittels des in Kapitel 5 eingeführten Tests kann die Signifikanz der Abhängigkeit der Lebensdauern und des Gründungsjahres bzw. des Alters bei Studienbeginn geprüft werden. Zunächst soll der Frage nachgegangen werden, ob ein Anstieg der Lebensdauern der Unternehmen zu erwarten ist.

7.2.1. Gumbel-Barnett-Copula

Für diesen Abschnitt wird angenommen, dass der Zusammenhang zwischen X_i und T_i durch die Gumbel-Barnett-Copula C_{ϑ_0} aus Annahme (A3) beschrieben wird, wobei auch ϑ_0 ein unbekannter zu schätzender Parameter ist. Die Gumbel-Barnett-Copula modelliert eine schwache negative Abhängigkeit zwischen X_i und T_i, was einer steigenden Lebenserwartung entspricht.
Mittels der in (3.4) und (3.5) beschriebenen Methode, können sowohl die Parameter θ_0 und ϑ_0 als auch n geschätzt werden. Als Schätzung erhält man

$$\hat{\theta}_n = 0{,}08256 \text{ und } \hat{\vartheta}_n = 0.$$

Daraus ergeben sich die Auswahlwahrscheinlichkeit, der Stichprobenumfang sowie Kendalls Tau mit

$$\alpha_{\hat{\boldsymbol{\theta}}_n} = 0{,}09545,\ \hat{n} = m/\alpha_{\hat{\boldsymbol{\theta}}_n} \approx 579.089 \text{ und } \tau_{\hat{\vartheta}_n} = 0.$$

Auf Basis der Beobachtung von 55.279 Unternehmensinsolvenzen zwischen 2014 und 2016 kann darauf geschlossen werden, dass zwischen 1990 und 2013 insgesamt 579.089 Firmen in Deutschland gegründet wurden. Die Schätzung des Parameters θ_0 ist hierbei ähnlich wie bei Weißbach und Wied [65] mit $\hat{\theta}_n = 0{,}08261$ und entspricht einer durchschnittlichen Lebensdauer insolventer Unternehmen von rund 12,11 Jahren.

Unter der Nullhypothese $H_0 : \vartheta_0 = 0$ können die Standardfehler mithilfe der Schätzung der Fisher-Informationsmatrix in (5.4) ermittelt werden. Die folgende Tabelle zeigt die Standardfehler von $\hat{\vartheta}_n$ und $(\hat{\theta}_n - \theta_0)$. Für die Schätzung des Standardfehlers von $(\hat{\theta}_n - \theta_0)$ wird zusätzlich die Randverteilung (5.5) benötigt. Weiterhin kürzt sich der Stichprobenumfang n nicht raus und muss ebenfalls geschätzt werden.

Tabelle 7.1.: Vergleich der Standardfehler von $\hat{\vartheta}_n$ und $(\hat{\theta}_n - \theta_0)$ unter H_0, für $\boldsymbol{\theta}_0 \in \mathring{\Theta}$ und nach Weißbach und Wied [65]

$\hat{\sigma}_\vartheta/\sqrt{n}$	$\hat{\sigma}_\theta/\sqrt{\hat{n}}$	$\hat{\sigma}_\theta/\sqrt{n}$	
(5.4)	(5.5)	4.13	[65]
$5{,}19 \cdot 10^{-3}$	$6{,}58 \cdot 10^{-4}$	$6{,}63 \cdot 10^{-4}$	$6{,}7 \cdot 10^{-4}$

Zum Vergleich wurden zusätzlich der Standardfehler von $(\hat{\theta}_n - \theta_0)$ unter der Annahme $\boldsymbol{\theta}_0 \in \mathring{\Theta}$ nach Satz 4.13 sowie nach Weißbach und Wied [65] angegeben. Die Abbildung 7.1 zeigt die negative logarithmierte Profillikelihood der Daten der Unternehmensinsolvenzen, die aus (3.4) mit $n = M/\alpha_{\boldsymbol{\theta}}$ hervorgeht. Aufgrund der starken Krümmung in θ-Richtung lässt sich gut erkennen, dass das Minimum in $\hat{\theta}_n \approx 0{,}08$ angenommen wird. Die Krümmung in ϑ-Richtung ist im Vergleich deutlich geringer, was sich in dem größeren Standardfehler widerspiegelt. Dennoch ist ein leichter Anstieg erkennbar, sodass das Minimum in $\hat{\vartheta}_n \approx 0$ lokalisiert werden kann.

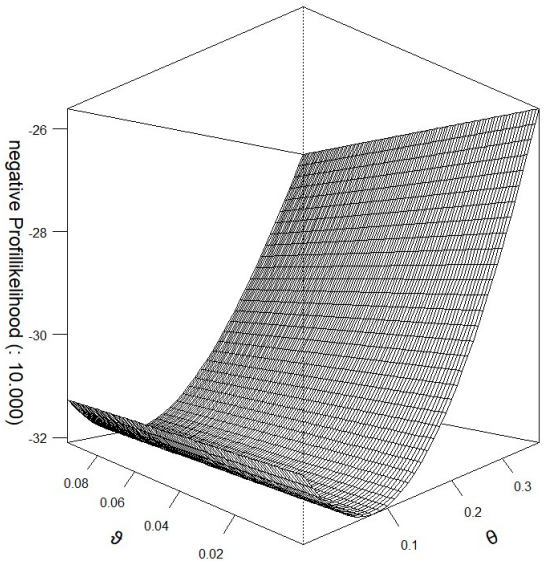

Abbildung 7.1.: Plot der negativen Profillikelihood für die Daten der Unternehmensinsolvenzen

Für einen Fehler der 1. Art von $\alpha = 0{,}05$ kann die Nullhypothese wegen $T_n = 0 < z_{0,95} = 1{,}645$ nicht abgelehnt werden. Die Unabhängigkeit der Lebensdauer einer Firma und des Alters bei Studienbeginn beziehungsweise des Gründungsjahres des Unternehmens kann nicht widerlegt werden. Auf Basis der vorliegenden Daten ist ein Anstieg der Lebensdauer bei später gegründeten Firmen, ähnlich wie es bei der Entwicklung der menschlichen Lebensdauer zu beobachten ist, nicht zu erwarten.

7.2.2. Farlie-Gumbel-Morgenstern-Copula

Es stellt sich nun die Frage, ob die Lebenserwartung der deutschen Unternehmen mit der Zeit sogar geringer wird. Um diese Frage zu beantworten, wird angenommen, dass der Zusammenhang zwischen X_i und T_i durch die Farlie-Gumbel-Morgenstern-Copula C_ϑ^{FGM} aus Annahme (A3') beschrieben wird. Der Parameter der Copula kann Werte zwischen -1 und 1 annehmen. Positive Parameterwerte stehen hierbei auch für eine positive Abhängigkeit. Das Maximieren der in Abschnitt 3.4 angegebenen Loglikelihood führt zu den Schätzungen

$$\hat{\theta}_n^{FGM} = 0{,}08172 \text{ sowie } \hat{\vartheta}_n^{FGM} = 0{,}10256.$$

Daraus lassen sich die Auswahlwahrscheinlichkeit, der Stichprobenumfang und Kendalls Tau berechnen. Diese lauten

$$\alpha_{\hat{\boldsymbol{\theta}}_n}^{FGM} = 0{,}09753,\ \hat{n}^{FGM} = m/\alpha_{\hat{\boldsymbol{\theta}}_n}^{FGM} \approx 566.816 \text{ und } \tau_{\hat{\vartheta}_n}^{FGM} = 0{,}02279.$$

Für dieses Modell bildet Satz 4.13 die Grundlage zur Berechnung der Standardfehler unter der Nullhypothese $H_0 : \vartheta_0^{FGM} = 0$. Diese können wie in Bemerkung 5.2 beschrieben berechnet werden und sind in der folgenden Tabelle dargestellt.

Tabelle 7.2.: Vergleich der Standardfehler von $(\hat{\vartheta}_n^{FGM} - \vartheta_0^{FGM})$ und $(\hat{\theta}_n^{FGM} - \theta_0^{FGM})$ unter der Nullhypothese

$\hat{\sigma}_\vartheta^{FGM}/\sqrt{n}$	$\hat{\sigma}_\theta^{FGM}/\sqrt{n}$
$1{,}80 \cdot 10^{-2}$	$6{,}58 \cdot 10^{-4}$

Im Vergleich zu den Standardfehlern im Gumbel-Barnett-Modell hat sich vor allem der Standardfehler für den Parameter der Farlie-Gumbel-Morgenstern-Copula erhöht. Genauere Ausführungen zur Berechnung der Standardfehler können im Appendix A.3 nachgelesen werden. Der Quotient aus der Schätzung des Copula-Parameters und dem Standardfehler beträgt $T_n^{FGM} = 5{,}709$, sodass die Nullhypothese zu den gängigen Signifikanzniveaus abgelehnt werden kann.

Aus diesem Abhängigkeitsmodell resultiert somit eine ähnliche durchschnittliche Lebensdauer von rund 12,24 Jahren. Auf Basis der Schätzung kann auf eine schwache positive Abhängigkeit zwischen dem Alter bei Studienbeginn und der Lebensdauer geschlossen werden. Dies ist nun äquivalent zu einer schwachen negativen Abhängigkeit zwischen dem Gründungsdatum und der Lebensdauer eines deutschen Unternehmens. Mithilfe der folgenden Abbildung erhält man einen Eindruck über die Entwicklung der Lebensdauer der deutschen Unternehmen mit dem Gründungsjahr. Sie zeigt den bedingten Erwartungswert der Lebensdauer X_i gegeben eines konkreten Gründungsjahres bzw. Alters bei Studienbeginn $T_i = t$.

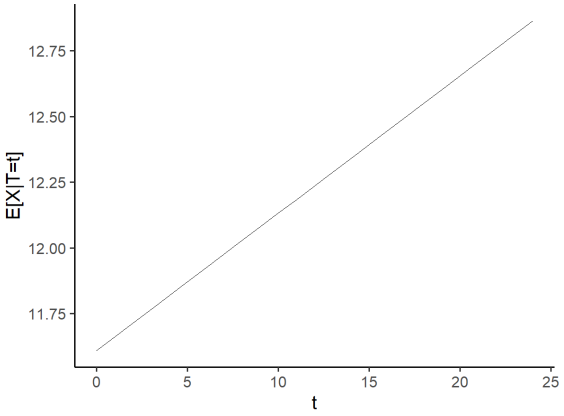

Abbildung 7.2.: Plot der erwarteten Lebensdauer deutscher Unternehmen bedingt auf das Alter bei Studienbeginn

Dabei berechnet sich der bedingte Erwartungswert durch

$$\begin{aligned}\mathbb{E}[X|T=t] &= \int_0^\infty x \cdot \frac{f_{\boldsymbol{\theta}}^{FGM}(x,t)}{f^T(t)}\,\mathrm{d}x \\ &= \int_0^\infty x\theta e^{-\theta x}\left[1+\vartheta(2e^{-\theta x}-1)\left(1-\frac{2t}{G}\right)\right]\,\mathrm{d}x \\ &= \frac{1}{\theta}\left[1-\frac{1}{2}\vartheta\left(1-\frac{2t}{G}\right)\right].\end{aligned}$$

Die durchschnittliche Lebensdauer eines im Jahr 1990 ($t = 24$) gegründeten Unternehmens beträgt rund 12,86 Jahre und fällt auf 11,61 Jahre für ein in 2014 ($t = 0$) gegründetes Unternehmen. Pro Jahr verringert sich die erwartete bedingte Lebensdauer um rund 18,82 Tage.

7.2.3. Zur Spiegelung assozierte Gumbel-Barnett-Copula

Im Vergleich soll nun die zur Spiegelung assoziierte Version der Gumbel-Barnett-Copula C_ϑ^{270} betrachtet werden. Es wird angenommen, dass der Zusammenhang zwischen X_i und T_i gemäß Annahme (A3'') beschrieben wird. Der Parameter der Copula kann nach wie vor Werte zwischen 0 und 1 annehmen. Die Parameterwerte stehen nach der Spiegelung jedoch für eine positive Abhängigkeit. Durch das Maximieren der in Abschnitt 3.4 angegebenen Loglikelihood ergeben sich die Schätzungen

$$\hat{\theta}_n^{270} = 0{,}08302 \;\text{ sowie }\; \hat{\vartheta}_n^{270} = 0{,}060005.$$

Daraus lassen sich die Auswahlwahrscheinlichkeit, der Stichprobenumfang und Kendalls Tau berechnen

$$\alpha_{\hat{\boldsymbol{\theta}}_n}^{270} = 0{,}09681,\;\; \hat{n}^{270} = m/\alpha_{\hat{\boldsymbol{\theta}}_n}^{270} \approx 571.018 \;\text{ und }\; \tau_{\hat{\vartheta}_n}^{270} = 0{,}02909.$$

Durch Schätzung der Fisher-Informationsmatrix in Gleichung (5.4) können die Standardfehler unter der Nullhypothese $H_0 : \vartheta_0^{270} = 0$ bestimmt werden. Diese sind in der folgenden Tabelle angegeben. Um den Standardfehler von $(\hat{\theta}_n^{270} - \theta_0^{270})$ zu schätzen, wird zusätzlich die Randverteilung (5.5) benötigt. Außerdem kürzt sich der Stichprobenumfang n nicht raus und muss ebenfalls geschätzt werden. Als Vergleich wird der Standardfehler von $(\hat{\theta}_n^{270} - \theta_0^{270})$ nach Satz 4.13, d.h. bei asymptotischer Normalität angegeben.

Tabelle 7.3.: Vergleich der Standardfehler von $\hat{\vartheta}_n^{270}$ und $(\hat{\theta}_n^{270} - \theta_0^{270})$ unter H_0 nach Satz 5.1 und 4.13

$\hat{\sigma}_{\vartheta}^{270}/\sqrt{n}$	$\hat{\sigma}_{\theta}^{270}/\sqrt{\hat{n}}$	$\hat{\sigma}_{\theta}^{270}/\sqrt{n}$
(5.4)	(5.5)	4.13
$5{,}82 \cdot 10^{-3}$	$6{,}60 \cdot 10^{-4}$	$6{,}71 \cdot 10^{-4}$

Für die Prüfgröße berechnet sich ein Wert von $T_n^{270} = 10{,}315$, sodass die Nullhypothese für alle üblichen Signifikanzniveaus abgelehnt werden kann.

Auch mit dieser Copula konnte eine schwache negative Abhängigkeit zwischen dem Gründungsdatum und der Lebensdauer eines deutschen Unternehmens gefunden werden. Der bedingte Erwartungswert gibt auch in diesem Fall Auskunft über die Entwicklung der Lebensdauer der deutschen Unternehmen mit dem Gründungsjahr. Dieser kann zwar nicht explizit berechnet werden, sein Verlauf wird jedoch in der folgenden Abbildung grafisch dargestellt.

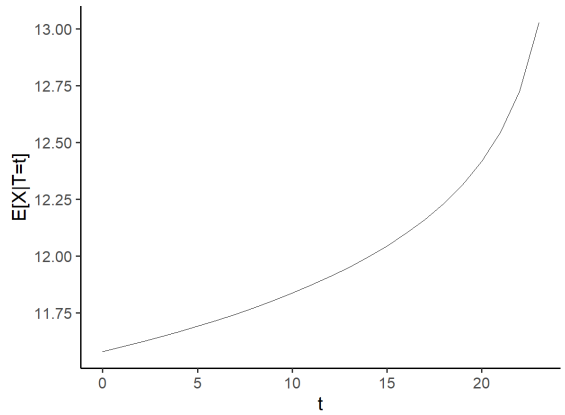

Abbildung 7.3.: Plot der erwarteten Lebensdauer deutscher Unternehmen bedingt auf das Alter bei Studienbeginn

Die durchschnittliche Lebensdauer eines im Jahr 1990 ($t = 24$) gegründeten Unternehmens beträgt rund 13,03 Jahre und fällt (in diesem Fall nicht linear) auf 11,58 Jahre

für ein in 2014 ($t = 0$) gegründetes Unternehmen.
Die folgende Tabelle gibt einen Überblick über die Ergebnisse der drei Abhängigkeitsmodelle.

Tabelle 7.4.: Vergleich der Ergebnisse des Unabhängigkeitstests für die drei Copulas C_ϑ, C_ϑ^{FGM} und C_ϑ^{270}

Copula	$\hat{\theta}_n$	$\hat{\vartheta}_n$	$1/\hat{\theta}_n$	τ	$\alpha_{\hat{\theta}_n}$	$\hat{n}$	$\hat{\sigma}_\vartheta/\sqrt{n}$	T_n
C_ϑ	0,083	0	12,11	0	0,095	579.089	$5{,}19 \cdot 10^{-3}$	0
C_ϑ^{FGM}	0,082	0,10	12,24	0,023	0,098	566.816	$1{,}80 \cdot 10^{-2}$	5,709
C_ϑ^{270}	0,083	0,06	12,05	0,029	0,097	571.018	$5{,}82 \cdot 10^{-3}$	10,315

Ein Vergleich der Ergebnisse zeigt, dass sich für den Parameter der Exponentialverteilung ähnliche Schätzwerte ergeben. Dies spiegelt sich auch in der mittleren Lebensdauer der Unternehmen wider. Die Schätzungen für den Parameter der Copula unterscheiden sich auf den ersten Blick. Unter Berücksichtigung von Kendalls Tau wird jedoch deutlich, dass mit beiden Copulas eine ähnlich schwache positive Abhängigkeit in den Daten gefunden wurde. Während die Standardfehler für den Parameter θ bei allen drei Modellen sehr ähnliche Werte aufweisen, ist der Standardfehler für den Parameter ϑ bei der Copula C_ϑ^{270} etwas geringer als bei der Copula C_ϑ^{FGM}.
Wichtig für die Wahl der Copula in Hinblick auf die Unternehmensdaten scheint zu sein, dass die Modellierung positiver Abhängigkeit möglich ist und diese insgesamt eher schwach ausgeprägt ist. Weiterhin sollte die Copula symmetrisch um $v = 1 - u$ sein, da keine Auffälligkeiten am unteren oder oberen Ende festgestellt werden können.

7.3. Test auf Altershomogenität der Unternehmensinsolvenzen

Die Copula-Modelle basieren auf konkreten Annahmen (vgl. (A2)) über die Verteilung der Lebensdauer der Unternehmen auf der einen Seite und des Alters bei Studienbeginn auf der anderen Seite. Die Frage, ob die zugrunde gelegte theoretische Verteilung die empirische Verteilung hinreichend gut beschreibt, kann mithilfe eines Anpassungstest überprüft werden. Damit wird auch die Gültigkeit der mit der Exponentialverteilung assoziierten Prämisse der Altershomogenität, das heißt der Abhängigkeit der Insolvenzwahrscheinlichkeit vom Alter des Unternehmens, in Frage gestellt. Es soll vereinfacht angenommen werden, dass die Lebensdauer X_i und das Alter bei Studienbeginn T_i stochastisch unabhängig sind (vgl. Annahme (A3’’’)). Konkret wird anhand der in Kapitel 6 zusammengefassten Theorie ein Kolmogorov-Smirnov-Anpassungstest für die vorliegenden Daten und das von Weißbach und Wied [65] eingeführte Modell durchgeführt. Der wahre Parameterwert θ_0 der Grundgesamtheit ist unbekannt, sodass die folgenden

Hypothesen (vgl. Gleichung (6.8)) überprüft werden sollen:

H_0 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt aus der Verteilungsfamilie $\{F_\theta = F^X \cdot F^T \,|\, \theta \in \Theta\}$.

H_1 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt nicht aus der Verteilungsfamilie $\{F_\theta = F^X \cdot F^T \,|\, \theta \in \Theta\}$.

Für den Parameter der Exponentialverteilung wurde von Weißbach und Wied ein Schätzwert von $\hat{\theta}_n = 0{,}08261$ ermittelt. Infolge dessen beträgt die Auswahlwahrscheinlichkeit $\alpha_{\hat{\theta}_n} = 0{,}0955$ (vgl. Gleichung (6.2)). Es liegen $m_n = 55.279$ Beobachtungen vor, wobei aufgrund der Doppeltrunkation der latente Stichprobenumfang n nicht beobachtbar ist. Dieser kann jedoch mittels $\hat{n} = m_n/\alpha_{\hat{\theta}_n} \approx 578.837$ geschätzt werden.

Berechnung der Prüfgröße

Zunächst wird die Berechnung der Prüfgröße vorgenommen. Dafür wird die in Algorithmus 5 beschriebene Schrittfolge durchgeführt. Für die Beobachtungen ($m_n = 55.279$) werden die Projektionen auf den oberen (15.268) und rechten Rand (6.768) sowie die Knotenpunkte (5.971.513) ermittelt. Insgesamt sind somit 6.048.849 Auswertungen zur Berechnung der Prüfgröße erforderlich. Daraus ergeben sich die folgenden Hilfswerte:

Tabelle 7.5.: Zwischenwerte zur Ermittlung der Kolmogrov-Smirnov-Statistik nach Algorithmus 5

D_m^1	D_m^2	D_m^3	D_m^4	D_m^5
$2{,}2725 \cdot 10^{-2}$	$2{,}2733 \cdot 10^{-2}$	$3{,}6173 \cdot 10^{-5}$	$1{,}7721 \cdot 10^{-5}$	$1{,}7728 \cdot 10^{-5}$

Die größte Abweichung tritt demnach in D_m^2 und folglich in den Knotenpunkten auf. Durch Multiplikation des Maximums mit $\sqrt{m}\sqrt{\alpha_{\hat{\theta}_n}}$, lässt sich schließlich der Wert der Prüfgröße mit $\sqrt{\hat{n}}\, d_n(\hat{\theta}_n, \hat{n}) = 1{,}6515$ ermitteln.

Berechnung des kritischen Wertes

Unter der Nullhypothese kann das Quantil mithilfe der in Algorithmus 4 beschriebenen Methode berechnet werden. Es wurde eine Gitterbreite von $1/100 = 0{,}01$ gewählt. Die Variablen und Parameter wurden den Daten entsprechend gewählt, d.h. $G = 24$ und $s = 3$. In diesem Fall beträgt die Anzahl der Einträge der Kovarianzmatrix (6.23) bereits $(100G \cdot 100(G+s))^2 = 4{,}199 \cdot 10^{13}$. Die Prozedur wurde 1.000 mal wiederholt und ein 99%-Quantil von 0,3558 konnte ermittelt werden. Der Wert der Prüfgröße übersteigt den kritischen Wert und die Nullhypothese wird verworfen.
Die nachfolgende Tabelle vergleicht die Quantile bei Schätzung von θ_0 und n mit denen bei Bekanntheit von θ_0 oder n. Die Berechnungen erfolgten unter den gleichen

Simulationsbedingungen wie oben beschrieben.

Tabelle 7.6.: Vergleich der Quantile für vier Stufen basierend auf der Schätzung von θ und n (vgl. Algorithmen 1-4)

	Signifikanzniveau α		
Schätzgrößen	0,10	0,05	0,01
(θ_0, n)	0,5745	0,6446	0,7805
$(\hat{\theta}_n, n)$	0,5719	0,6558	0,8298
$(\theta_0, \hat{n})$	0,3820	0,4235	0,5034
$(\hat{\theta}_n, \hat{n})$	0,2869	0,3051	0,3558

Die Quantile der oberen drei Stufen liegen zwar über dem Quantil der letzten, jedoch weiterhin unter dem Wert der Prüfgröße, sodass die Nullhypothese auch für die oberen drei Stufen verworfen werden muss.

Im eindimensionalen Fall ohne Trunkation konnte Lilliefors für die Normalverteilung [35] und die Exponentialverteilung [36] nachweisen, dass sich die kritischen Werte, unabhängig vom wahren Parameterwert, verringern, wenn die Parameter der Verteilung geschätzt werden. Ein Vergleich der ersten beiden Zeilen der Tabelle 7.9 zeigt, dass dies in dieser Arbeit nicht der Fall ist. Mit Blick auf die Varianzen der Gauß-Prozesse ((6.6) und (6.13)) wird ersichtlich, dass die Grenzverteilung bei Trunkation unter H_0 vom zugrunde liegenden Parameter abhängt. Die folgende Abbildung vergleicht die Varianzen der asymptotischen Verteilung bei Bekanntheit des Parameters (dunkelgrau) und bei Schätzung des Parameters (hellgrau) miteinander. Der latente Stichprobenumfang n wird als bekannt vorausgesetzt. Der Parameter θ nimmt dabei Werte aus der Menge $\{0{,}001; 0{,}08261; 0{,}5; 1\}$ an.

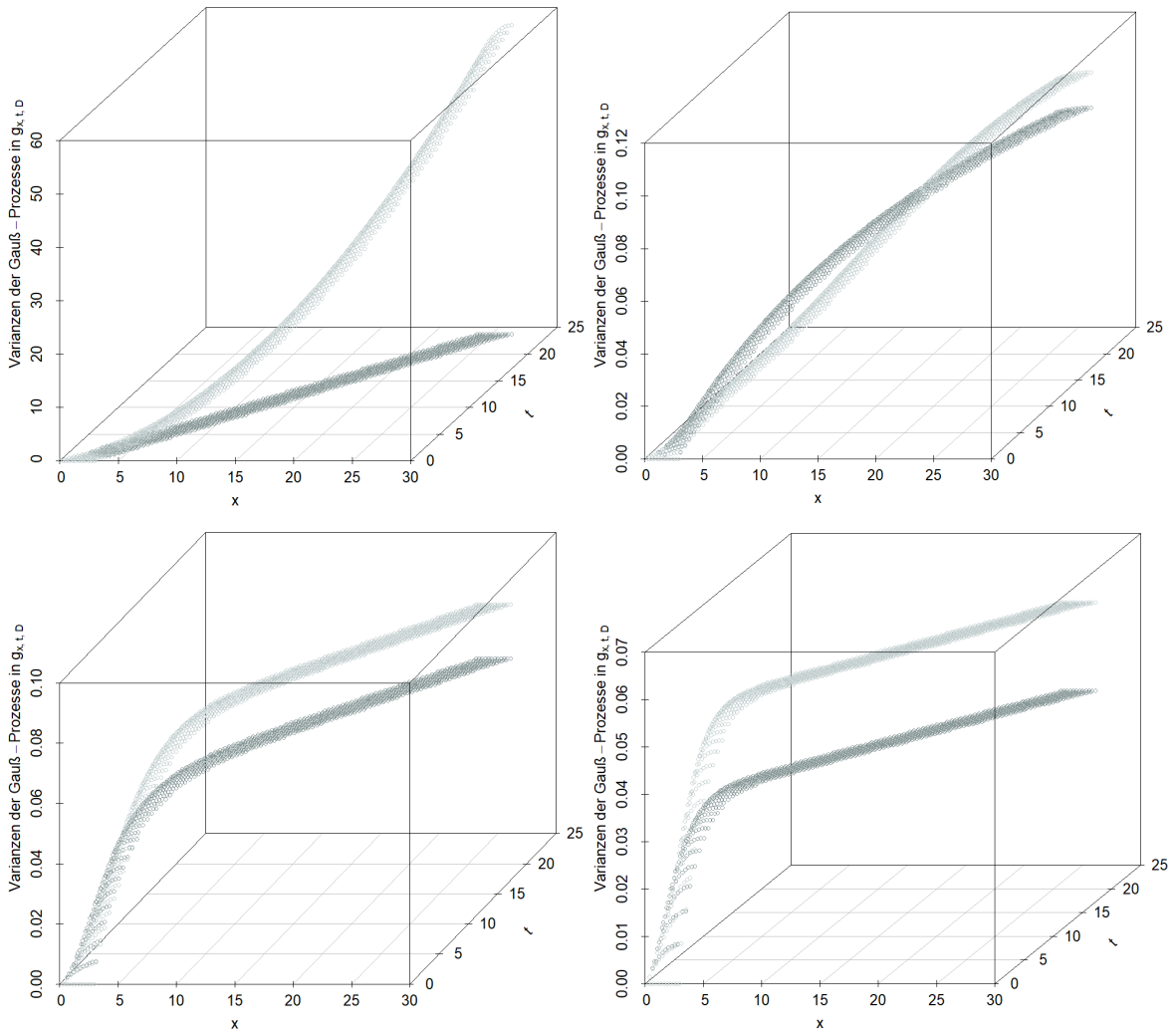

Abbildung 7.4.: Plot der Varianzen (6.6) für θ_0 (dunkelgrau) und (6.13) für $\hat{\theta}_n$ (hellgrau) für die Parameterwerte $\theta \in \{0{,}001; 0{,}08261; 0{,}5; 1\}$ auf D

Für die Stufen $(\theta_0, \hat{n})$ und $(\hat{\theta}_n, \hat{n})$ hingegen lässt sich eine deutliche Verringerung der Quantile im Vergleich zu (θ_0, n) in Tabelle 7.9 beobachten. Im Falle der Unbekanntheit von θ_0 und n würde die Nichtablehnung einer richtigen Nullhypothese begünstigt, wenn die Quantile aus den drei weiteren Stufen zum Vergleich herangezogen würden. Dies ließe den Test in dem Sinne als konservativ erscheinen.

Nach Modifikation der Daten

Bei genauerer Betrachtung der Daten (vgl. linke Abbildung in 7.5) werden horizontale Linien im Parallelogramm ersichtlich. Die Auffälligkeiten in den Daten werden durch die Transformation der Beobachtungen über die (bedingten) Verteilungen deutlicher sichtbar. Es erfolgt die Transformation

$$U_{1,j} = F^{\widetilde{X}}(\widetilde{X}_j) \qquad \text{und} \qquad U_{2,j} = F^{\widetilde{T}}(\widetilde{T}_j|\widetilde{X}_j) \tag{7.1}$$

und die transformierten Zufallsvariablen sind gemäß Rosenblatt [49] unabhängig und gleichverteilt auf $[0, 1]$. (Details zur Abhängigkeit von $\widetilde{X}_j$ und $\widetilde{T}_j$ sowie zur Berechnung der Transformation können im Anhang A.4 nachgelesen werden.) Zusätzlich zu den nun eher vertikalen schwarzen Linien, erscheinen im rechten Teil der Abbildung 7.5 nun auch weiße horizontale Linien.

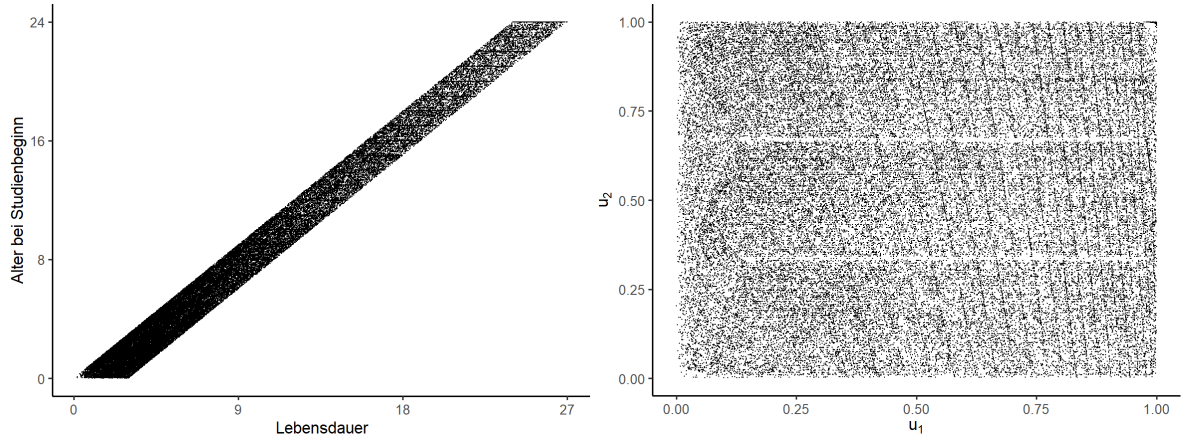

Abbildung 7.5.: (links) Beobachtungen in D, (rechts) mit (7.1) transformierte Beobachtungen

Die erkennbaren schwarzen bzw. weißen Linien sind darauf zurückzuführen, dass zwischen Weihnachten und Neujahr keine Gründungen bzw. Insolvenzen registriert wurden. Die Gründungen zwischen dem 01.01.1990 und dem 31.12.2013 häufen sich dann zum 01.01 des jeweiligen Folgejahres, was sich in den schwarzen Linien widerspiegelt. In der linken Abbildung entspricht dies den horizontalen schwarzen Linien. Die Insolvenzen werden für den Zeitraum vom 01.01.2014 bis zum 31.12.2016 beobachtet. Zwischen dem 24. und dem 27. Dezember der Jahre 2014 und 2015 wurden keine Insolvenzen registriert, was die beiden weißen Streifen erklärt. In den nicht transformierten Daten führt dies zu weißen Linien um die Geraden $T = X - 1$ und $T = X - 2$, die jedoch kaum sichtbar sind.

Wie aus der Abbildung 7.5 hervorgeht, verteilen sich die transformierten Beobachtungen nicht gleichmäßig über das Einheitsquadrat. In diesem Zusammenhang stellt sich die Frage, inwiefern die Häufung von Gründungen zu Monats- und Jahresbeginn das Ergebnis des Anpassungstests beeinflusst. Dazu werden die Gründungen zu Monats- und Jahresbeginn gleichmäßig über den Monat bzw. das Jahr aufgeteilt. Erwartungsgemäß erhöht sich der Schätzwert für den Parameter der Exponentialverteilung, was wiederum zu einer Verringerung der erwarteten Lebensdauer führt. Im Folgenden werden die Ergebnisse einander gegenübergestellt.

Tabelle 7.7.: Vergleich der Prüfgröße des Kolmogorov-Smirnov-Anpassungstest mit und ohne Modifizierung der Gründungen

Modifizierung der Gründungen	$\hat{\theta}_n$	$1/\hat{\theta}_n$	$\sqrt{\hat{n}}d_n(\hat{\theta}_n, \hat{n})$
ohne	0,08261	12,1051	1,6515
mit	0,08397	11,910	1,5412

Der Wert der Prüfgröße hat sich zwar verringert, die Nullhypothese wird aber nach wie vor abgelehnt. Es kann nicht davon ausgegangen werden, dass die latenten Daten die angegebene Verteilung besitzen.

Farlie-Gumbel-Morgenstern-Copula

An dieser Stelle wird der entwickelte Anpassungstest ergänzend für die Farlie-Gumbel-Morgenstern-Copula C_ϑ^{FGM} aus Annahme (A3') durchgeführt. Die Wahl dieser Copula begründet sich in ihrer einfacheren Form, da auftretende Integrale, wie bei der Auswahlwahrscheinlichkeit, sich explizit berechnen lassen. Die zugehörigen Hypothesen lauten:

H_0 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt aus der Verteilungsfamilie $\{C_\vartheta^{FGM}(F^X, F^T) \,|\, \boldsymbol{\theta} \in \Theta\}$.

H_1 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt nicht aus der Verteilungsfamilie $\{C_\vartheta^{FGM}(F^X, F^T) \,|\, \boldsymbol{\theta} \in \Theta\}$.

Die Parameter wurden bereits in Abschnitt 7.2.2 geschätzt mit $\hat{\theta}_n^{FGM} = 0{,}08172$ und $\hat{\vartheta}_n^{FGM} = 0{,}10256$. Daraus erhält man $\hat{n}^{FGM} = m/\alpha_{\hat{\boldsymbol{\theta}}_n}^{FGM} \approx 566.816$ mit $\alpha_{\hat{\boldsymbol{\theta}}_n}^{FGM} = 0{,}09753$.

Für die Berechnung der Prüfgröße werden dieselben Beobachtungen, Rand- und Knotenpunkte herangezogen, die bereits bei der Unabhängigkeitscopula ermittelt wurden. Die folgende Tabelle enthält die Zwischenwerte nach Algorithmus 5.

Tabelle 7.8.: Zwischenwerte zur Ermittlung der Kolmogrov-Smirnov-Statistik nach Algorithmus 5, angewendet auf die Farlie-Gumbel-Morgenstern-Copula aus Beispiel 2.21

D_m^1	D_m^2	D_m^3	D_m^4	D_m^5
$2{,}3040 \cdot 10^{-2}$	$2{,}3039 \cdot 10^{-2}$	$3{,}6173 \cdot 10^{-5}$	$1{,}7697 \cdot 10^{-5}$	$1{,}7702 \cdot 10^{-5}$

Die größte Abweichung tritt in den Beobachtungen auf und der Wert der Prüfgröße beträgt $\sqrt{\hat{n}}\, d_n(\hat{\boldsymbol{\theta}}_n, \hat{n})$ =1,6917.

Die Berechnung der Quantile erfolgt ohne explizite Herleitung der Konvergenz des trunkierten Prozesses, da die im Abschnitt 6 verwendete Argumentation direkt auf die Copula übertragen werden kann. Hervorzuheben ist, dass die in Stufe zwei und vier zur Approximation genutzten Fréchet-Ableitungen für beide Parameter, θ und ϑ, bestimmt werden müssen. Auch die Funktion $\phi_{\boldsymbol{\theta}_0} = (\phi_{\boldsymbol{\theta}_0,1}, \phi_{\boldsymbol{\theta}_0,2})$ ist zweidimensional. Die Grenzprozesse für Stufe zwei und vier lauten dann:

$$\mathbb{G}^{\phi}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(g_{x,t,D}) := \mathbb{B}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(g_{x,t,D}) - \left(\mathbb{B}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(\phi_{\boldsymbol{\theta}_0,1}), \mathbb{B}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(\phi_{\boldsymbol{\theta}_0,2})\right) \cdot \begin{pmatrix} \dot{\mathbb{E}}_{\theta}(g_{x,t,D}) \\ \dot{\mathbb{E}}_{\vartheta}(g_{x,t,D}) \end{pmatrix} \qquad \text{sowie}$$

$$\mathbb{H}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(g_{x,t,D}) := \mathbb{G}^{\phi}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(g_{x,t,D}) + \left[\left(\mathbb{B}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(\phi_{\boldsymbol{\theta}_0,1}), \mathbb{B}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(\phi_{\boldsymbol{\theta}_0,2})\right) \cdot \begin{pmatrix} \dot{\alpha}_{\theta} \\ \dot{\alpha}_{\vartheta} \end{pmatrix} - \mathbb{B}_{\mathbb{P}_{\boldsymbol{\theta}_0}}(\mathbb{1}_D)\right] \cdot \frac{\mathbb{E}_{\boldsymbol{\theta}_0}(g_{x,t,D})}{\alpha_{\boldsymbol{\theta}_0}}$$

Die folgende Tabelle enthält die berechneten Quantile. Zur Berechnung wurden $G = 24$, $s = 3$ und eine Gitterbreite von 0,01 herangezogen.

Tabelle 7.9.: Vergleich der Quantile für vier Stufen basierend auf der Schätzung von $\boldsymbol{\theta}$ und n (vgl. Algorithmen 1-4, angewendet auf die Farlie-Gumbel-Morgenstern-Copula aus Beispiel 2.21)

	Signifikanzniveau α		
Schätzgrößen	0,10	0,05	0,01
$(\boldsymbol{\theta}_0, n)$	0,5799	0,6518	0,7862
$(\hat{\boldsymbol{\theta}}_n, n)$	0,7320	0,8510	1,0816
$(\boldsymbol{\theta}_0, \hat{n})$	0,3868	0,4275	0,5092
$(\hat{\boldsymbol{\theta}}_n, \hat{n})$	0,2378	0,2535	0,2914

Die Nullhypothese wird auch für diese Copula verworfen. In Kombination mit dem Unabhängigkeitstest ist dies ein Indiz dafür, dass die Ablehnung der Nullhypothese beim Anpassungstest mit der Unabhängigkeitscopula nicht allein aufgrund der zugrunde gelegten Unabhängigkeit erfolgt. Die Altershomogenität kann für die vorliegenden Daten nicht gestützt und die Wahl der Randverteilungen sollte weitergehend geprüft werden.

8. Simulationen

Nachdem die beiden Tests Anwendung auf Daten zur Unternehmensdemografie gefunden haben, sollen diese nun mittels Simulationen evaluiert werden. Der erste Abschnitt enthält zunächst umfangreiche Simulationen zur Konsistenz der Schätzfolge aus Satz 4.11 sowie Untersuchungen zur Güte des Unabhängigkeitstests aus Kapitel 5. Der zweite Abschnitt untersucht die Eintrittswahrscheinlichkeit von Fehlern der ersten Art im Kolmogorov-Smirnov-Anpassungstest aus Kapitel 6.

8.1. Parameterschätzung und Test auf Unabhängigkeit

In diesem Abschnitt werden die Parameter θ_0 und ϑ_0 für simulierte Daten geschätzt. Anhand der Verzerrung sowie der Varianz der Schätzer soll die Konsistenz aus Satz 4.11 veranschaulicht werden. Weiterhin können die mithilfe der Simulation erhaltenen Daten mit denen aus der Arbeit [65] von Weißbach und Wied verglichen werden. Im Anschluss erfolgen Simulationen zur Güte des in Kapitel 5 vorgestellten Tests auf Unabhängigkeit.

Simulation basierend auf der Gumbel-Barnett-Copula

Gemäß der bedingten Inversionsmethode 2.23 wurden jeweils $n_{sim} = 1.000$ Zufallsstichproben der Größen $n \in \{10^3, 10^4, 10^5\}$ generiert. Dafür wurden die Gumbel-Barnett-Copula (2.4) (vgl. Annahme (A3)) sowie die Randdichten aus der Annahme (A2) zugrunde gelegt.

- Erzeuge zwei unabhängige Realisationen u und v' einer $(0,1)$-Gleichverteilung.
- Setze $v = c_u^{(-1)}(v')$ mit $c_u(v) = 1 + (1-v)^{1-\vartheta_0 \log(1-u)}[\vartheta_0 \log(1-v) - 1]$.
- Erhalte mit $x = -\log(1-u)/\theta_0$ und $t = v \cdot G$ die gesuchte Realisation von (X, T).

Die Parameter $G \in \{24, 48\}$ und $s \in \{2, 3, 24, 48\}$ werden ähnlich wie bei Weißbach und Wied [65] gewählt. Zusätzlich zu dem Parameter der Exponentialverteilung $\theta_0 \in \{0{,}05;\ 0{,}1\}$ wird der Parameter $\vartheta_0 \in \{0{,}001;\ 0{,}01;\ 0{,}2;\ 0{,}5\}$ der Copula berücksichtigt. Dies entspricht einem Wert von Kendalls Tau von jeweils $\tau_{\vartheta_0} \in \{-0{,}0005;\ -0{,}005;\ -0{,}0912;\ -0{,}206\}$. Zu den Beobachtungen gelangt man durch das Trunkieren der Werte $(X, T) \notin D$ gemäß Annahme (A4).
Die Schätzer $\hat{\theta}_n$ und $\hat{\vartheta}_n$ berechnen sich nun als Nullstelle der Score-Funktion (3.6). In den folgenden Tabellen 8.1 sind jeweils der Bias sowie die Varianz der Schätzer aufgeführt.

Tabelle 8.1.: Simulationsergebnisse für Bias und Varianz von $\hat{\theta}_n$ sowie $\hat{\vartheta}_n$

G=24, s=3

$\theta_0 = 0{,}05$; $\vartheta_0 = 0{,}001$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}0648$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,00403	0,001456	0,000446
$\text{Bias}(\hat{\vartheta}_n)$	0,12373	0,047852	0,017451
$\text{Var}(\hat{\theta}_n)$	0,00026	0,000026	0,000003
$\text{Var}(\hat{\vartheta}_n)$	0,02513	0,003987	0,000610

$\theta_0 = 0{,}05$; $\vartheta_0 = 0{,}01$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}0808$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,00241	0,001411	0,001110
$\text{Bias}(\hat{\vartheta}_n)$	0,13245	0,057141	0,038426
$\text{Var}(\hat{\theta}_n)$	0,00025	0,000027	0,000003
$\text{Var}(\hat{\vartheta}_n)$	0,02961	0,004975	0,001038

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}001$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}0982$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,00075	-0,00010	-0,000192
$\text{Bias}(\hat{\vartheta}_n)$	0,07880	0,02142	0,005680
$\text{Var}(\hat{\theta}_n)$	0,00028	0,00003	0,000003
$\text{Var}(\hat{\vartheta}_n)$	0,01147	0,00099	0,000087

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}01$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}0978$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,00015	-0,00095	-0,000884
$\text{Bias}(\hat{\vartheta}_n)$	0,07874	0,01768	0,006281
$\text{Var}(\hat{\theta}_n)$	0,00025	0,00003	0,000003
$\text{Var}(\hat{\vartheta}_n)$	0,01340	0,00117	0,000195

G=24, s=48

$\theta_0 = 0{,}05$; $\vartheta_0 = 0{,}001$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}5294$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,000136	-0,000011	0,0000044
$\text{Bias}(\hat{\vartheta}_n)$	0,024781	0,005959	0,0013721
$\text{Var}(\hat{\theta}_n)$	0,000010	0,000001	0,0000001
$\text{Var}(\hat{\vartheta}_n)$	0,001227	0,000102	0,0000108

$\theta_0 = 0{,}05$; $\vartheta_0 = 0{,}01$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}5286$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,00002	-0,000106	-0,0000881
$\text{Bias}(\hat{\vartheta}_n)$	0,02023	0,003510	-0,0000125
$\text{Var}(\hat{\theta}_n)$	0,00001	0,000001	0,0000001
$\text{Var}(\hat{\vartheta}_n)$	0,00150	0,000198	0,0000312

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}001$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}37575$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,00035	0,000011	-0,0000253
$\text{Bias}(\hat{\vartheta}_n)$	0,01883	0,004558	0,0007957
$\text{Var}(\hat{\theta}_n)$	0,00002	0,000002	0,0000002
$\text{Var}(\hat{\vartheta}_n)$	0,00083	0,000064	0,0000063

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}01$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}37574$

n	10^3	10^4	10^5
$\text{Bias}(\hat{\theta}_n)$	0,00011	-0,00018	-0,0001896
$\text{Bias}(\hat{\vartheta}_n)$	0,01086	-0,002902	-0,0061084
$\text{Var}(\hat{\theta}_n)$	0,00002	0,000002	0,0000002
$\text{Var}(\hat{\vartheta}_n)$	0,00078	0,000080	0,0000122

G=48, s=3

θ_0 =0,05; ϑ_0 =0,001; $\alpha_{\boldsymbol{\theta}_0}$ =0,0528

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00104	-0,00016	-0,000031
Bias($\hat{\vartheta}_n$)	0,12574	0,03096	0,008380
Var($\hat{\theta}_n$)	0,00014	0,00001	0,000001
Var($\hat{\vartheta}_n$)	0,02701	0,00200	0,000167

$\theta_0 = 0{,}05$; $\vartheta_0 = 0{,}01$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}0525$

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00055	-0,00035	-0,000379
Bias($\hat{\vartheta}_n$)	0,11406	0,02914	0,007752
Var($\hat{\theta}_n$)	0,00012	0,00001	0,000001
Var($\hat{\vartheta}_n$)	0,02528	0,00244	0,000298

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}001$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}0535$

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00229	0,00018	-0,000015
Bias($\hat{\vartheta}_n$)	0,11244	0,02525	0,006045
Var($\hat{\theta}_n$)	0,00030	0,00002	0,000002
Var($\hat{\vartheta}_n$)	0,02188	0,00141	0,000104

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}01$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}0534$

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00275	-0,00023	-0,000580
Bias($\hat{\vartheta}_n$)	0,11327	0,01970	-0,000619
Var($\hat{\theta}_n$)	0,00030	0,00002	0,000002
Var($\hat{\vartheta}_n$)	0,02576	0,00154	0,000123

G=24, s=24

$\theta_0 = 0{,}05$; $\vartheta_0 = 0{,}2$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}3884$

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00006	0,000083	0,0000155
Bias($\hat{\vartheta}_n$)	0,00775	0,000939	-0,0003613
Var($\hat{\theta}_n$)	0,00002	0,000002	0,0000002
Var($\hat{\vartheta}_n$)	0,00830	0,000860	0,0000843

$\theta_0 = 0{,}05$; $\vartheta_0 = 0{,}5$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}3672$

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00004	0,000029	0,0000132
Bias($\hat{\vartheta}_n$)	0,00752	0,000412	0,0000543
Var($\hat{\theta}_n$)	0,00001	0,000001	0,0000001
Var($\hat{\vartheta}_n$)	0,00984	0,001018	0,0000959

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}2$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}3397$

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00058	-0,000031	-0,0000185
Bias($\hat{\vartheta}_n$)	0,00799	0,001248	0,0004030
Var($\hat{\theta}_n$)	0,00004	0,000004	0,0000004
Var($\hat{\vartheta}_n$)	0,00473	0,000482	0,0000481

$\theta_0 = 0{,}1$; $\vartheta_0 = 0{,}5$; $\alpha_{\boldsymbol{\theta}_0} = 0{,}3367$

n	10^3	10^4	10^5
Bias($\hat{\theta}_n$)	0,00036	0,000056	-0,0000023
Bias($\hat{\vartheta}_n$)	0,00815	0,000794	0,0001562
Var($\hat{\theta}_n$)	0,00003	0,000003	0,0000003
Var($\hat{\vartheta}_n$)	0,00820	0,000853	0,0000848

Die Verzerrungen berechnen sich durch

$$Bias(\hat{\theta}_n) = \frac{1}{n_{sim}} \sum_{\nu=1}^{n_{sim}} \hat{\theta}_n^{(\nu)} - \theta_0 \quad \text{und} \quad Bias(\hat{\vartheta}_n) = \frac{1}{n_{sim}} \sum_{\nu=1}^{n_{sim}} \hat{\vartheta}_n^{(\nu)} - \vartheta_0.$$

Für die Simulationsdaten lässt sich gut erkennen, dass der Bias von $\hat{\theta}_n$ als Funktion in n für alle Szenarien gegen null geht. Für die Parameterwerte ϑ_0, die nah am Rand des Parameterraumes liegen, ist der Bias von $\hat{\vartheta}_n$ jeweils im Vergleich zum wahren Parameterwert wesentlicher größer und tendiert, wenngleich wesentlicher langsamer, ebenso gegen null. Für $\vartheta_0 \in \{0{,}2;\ 0{,}5\}$ ist ein kleinerer Bias zu beobachten, der mit ähnlicher Konvergenzgeschwindigkeit wie der von Bias$(\hat{\theta}_n)$ gegen null tendiert.
Ähnliches lässt sich für die Varianzen beobachten, die berechnet werden durch

$$Var(\hat{\theta}_n) = \frac{1}{n_{sim}} \sum_{\nu=1}^{n_{sim}} (\hat{\theta}_n^{(\nu)} - \theta_0)^2 \quad \text{und} \quad Var(\hat{\vartheta}_n) = \frac{1}{n_{sim}} \sum_{\nu=1}^{n_{sim}} (\hat{\vartheta}_n^{(\nu)} - \vartheta_0)^2.$$

Die Varianz von $\hat{\theta}_n$ ist jeweils im Vergleich zum Parameter θ_0 sehr gering und fällt schnell als Funktion in n. Dagegen ist die Streuung von $\hat{\vartheta}_n$ in Relation zu ϑ_0 recht groß und geht langsamer gegen null. Die mittlere quadratische Abweichung als Summe aus der Varianz und der quadrierten Verzerrung des Schätzers tendiert somit ebenfalls für große n gegen null und deutet auf Konsistenz im quadratischen Mittel hin. Die in Satz 4.11 gezeigte Konsistenz kann dadurch veranschaulicht werden.
Für $G = 24$ und $s = 3$ erhält man Auswahlwahrscheinlichkeiten zwischen sechs und zehn Prozent. Bei einer Stichprobengröße von $n = 10^5$ und einer Auswahlwahrscheinlichkeit von $\alpha_{\boldsymbol{\theta}_0} = 0{,}0808$ verbleiben nach Trunkation $m_n = 8.080$ Beobachtungen. Die Konsistenz ist somit bereits für realistische Stichprobengrößen sichtbar.

Grafische Darstellung der asymptotischen Varianz

Weiterhin ist die Wahl der Parameter G und s von Bedeutung. In den Szenarien $s = 24$ sowie $s = 48$ wird deutlich, dass die Varianzen der Schätzer bei einem großen Beobachtungszeitraum deutlich geringer sind. Die folgende Abbildung zeigt exemplarisch für $G = 24$; $\theta_0 = 0{,}1$ und $\vartheta_0 = 0{,}01$ den Verlauf der asymptotischen Varianzen aus dem Satz 4.13. Sowohl für die θ-Komponente in der linken Abbildung als auch für die ϑ-Komponente der Matrix $\mathcal{I}(\boldsymbol{\theta}_0)^{-1}$ in der rechten Abbildung ist deutlich erkennbar, dass die Varianz für einen größeren Beobachtungszeitraum geringer wird. Dabei führt ein kleiner Beobachtungszeitraum vor allem für den Schätzer von ϑ zu einer großen Varianz im Vergleich zum wahren Parameterwert. Bei einer langen Beobachtungsperiode werden weniger Beobachtungen trunkiert, sodass am Ende mehr Daten für die Schätzung zur Verfügung stehen, was zu einer geringeren Varianz führt.

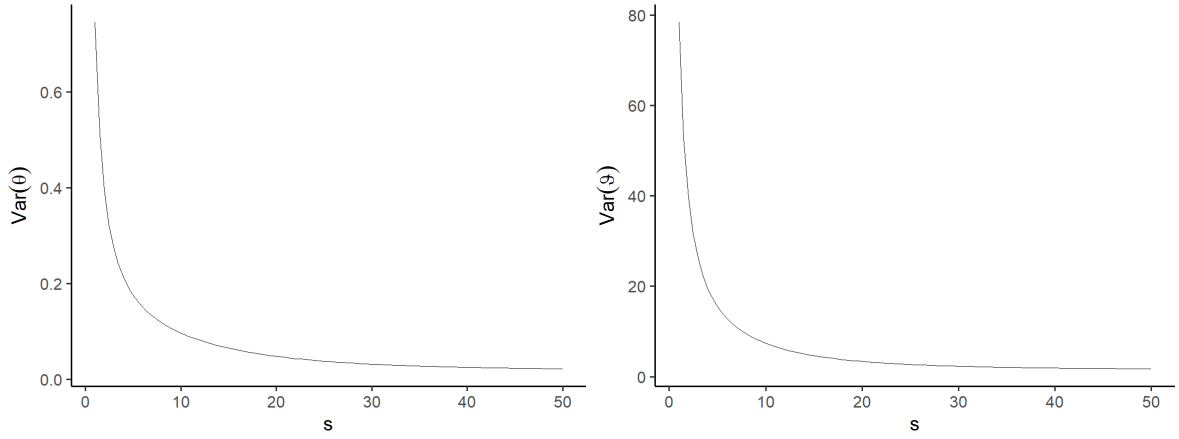

Abbildung 8.1.: Asymptotische Varianz von $\sqrt{n}(\hat{\theta}_n - \theta_0)$ (links) und $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)$ (rechts) aus 4.13 in Abhängigkeit von s für $G = 24$ und $\theta_0 = 0{,}1$ sowie $\vartheta_0 = 0{,}01$

Der nicht monotone Zusammenhang zwischen den Varianzen und der Länge der Geburtenperiode G lässt sich aus den hier dargestellten Szenarien nicht direkt erfassen. Auch hier soll die folgende Abbildung zur Veranschaulichung dienen. In dieser wird die Entwicklung der θ- und ϑ-Komponente der Matrix $\mathcal{I}(\boldsymbol{\theta}_0)^{-1}$ für $s = 3$; $\theta_0 = 0{,}1$ und $\vartheta_0 = 0{,}01$ in Abhängigkeit von G dargestellt. Darauf lässt sich zunächst ein steiler Anstieg, gefolgt von einem Abfall der Varianzen erkennen. Die hohen Werte der ϑ-Komponente der Matrix $\mathcal{I}(\boldsymbol{\theta}_0)^{-1}$ in der rechten Abbildung lassen sich durch die kleine Beobachtungsperiode $s = 3$ sowie den geringen Parameterwert $\vartheta_0 = 0{,}01$ erklären. Für größere Beobachtungsperioden verringert sich der Wert des lokalen Maximums. Für große Geburtenperioden steigen die Varianzen. Dies lässt sich durch die größere Spanne an möglichen Lebensdauern erklären. Bei früherer Geburt beziehungsweise Gründung und gleicher oder geringerer Lebensdauer werden somit mehr Beobachtungen trunkiert.

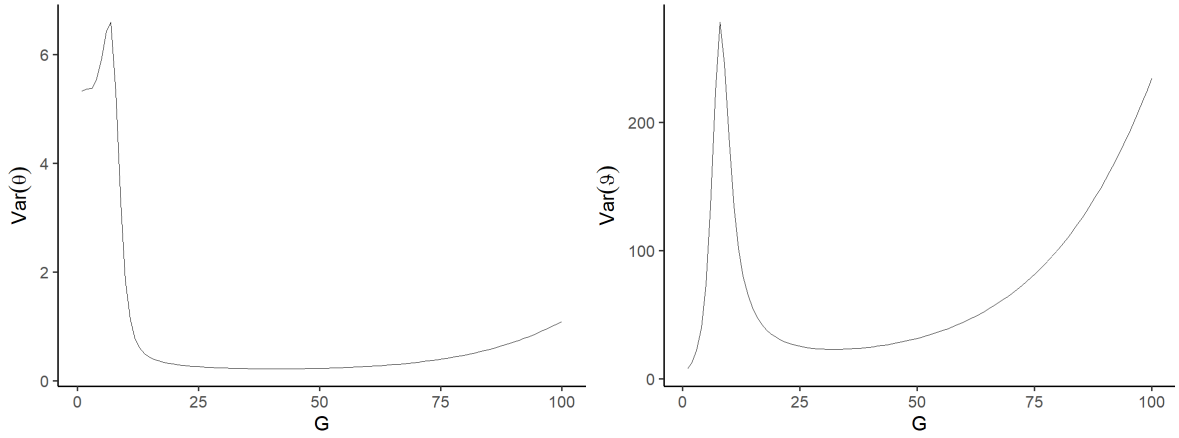

Abbildung 8.2.: Asymptotische Varianz von $\sqrt{n}(\hat{\theta}_n - \theta_0)$ (links) und $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)$ (rechts) aus 4.13 in Abhängigkeit von G für $s = 3$ und $\theta_0 = 0{,}1$ sowie $\vartheta_0 = 0{,}01$

Vergleich der Simulationsergebnisse mit dem Unabhängigkeitsfall

Im direkten Vergleich sollen nun die Simulationsergebnisse aus der Arbeit [65] von Weißbach und Wied betrachtet werden. Die dortige Unabhängigkeitsannahme zwischen dem Alter bei Studienbeginn und der Lebensdauer entspricht einem Parameterwert von $\vartheta_0 = 0$, wodurch die Gumbel-Barnett-Copula (2.4) zur Produktcopula (vgl. Annahme (A3''')) wird. Im Szenario $G = 24$, $s = 3$ gehen aus den dortigen Simulationen Biaswerte von 0,003 und $-0{,}00004$ für $\theta_0 = 0{,}05$ und $\theta_0 = 0{,}1$ bei $n = 10^4$ hervor. Die Berücksichtigung des Abhängigkeitsparameters führt in diesem Fall zu einer wesentlichen Vergrößerung der Verzerrung. Bei einer längeren Beobachtungsperiode und somit einer größeren Auswahlwahrscheinlichkeit zum Beispiel für $G = 24$ und $s = 48$ ist dieser starke Unterschied nicht zu beobachten. Bei einem Stichprobenumfang von $n = 10^4$ beträgt der Bias bei Unabhängigkeit $-0{,}00002$ für $\theta_0 = 0{,}05$ und $-0{,}000026$ bei $\theta_0 = 0{,}1$. Für $\vartheta_0 = 0{,}001$ sind ähnliche Größenordnungen zu beobachten. Die Varianzen von $\hat{\theta}_n$ stimmen in beiden Szenarien jeweils mit denen aus den Simulationen von [65] Weißbach und Wied überein.

Simulationen zur Güte des Tests auf Unabhängigkeit

Schließlich werden die Simulationen zur Testgüte betrachtet. In Kapitel 5 wurde bereits gezeigt, wie die Güte des Tests mit $H_0 : \vartheta_0 = 0$ und $H_1 : \vartheta_0 = \vartheta^1 > 0$ näherungsweise mithilfe der Standardnormalverteilung berechnet werden kann. Eine weitere Möglichkeit besteht in der Berechnung mithilfe einer Simulation. Dafür wurden $n = 500.000$ Werte unter den Bedingungen von H_1 für $G = 24$, $s = 3$ und $\theta_0 = 0{,}08$ simuliert und der Schätzer $\hat{\vartheta}_n$ bestimmt. Anschließend wurde die Prüfgröße $T_n = \sqrt{n}\hat{\vartheta}_n/\sigma_\vartheta$ berechnet und für einen Fehler der 1. Art von $\alpha = 0{,}05$ mit dem Wert $z_{0,95} = 1{,}645$ verglichen. Diese Vorgehensweise wurde jeweils 1.000 mal wiederholt und die Güte des Tests, $1-\beta$, somit als Anteil der Ablehnungen der Nullhypothese unter Gültigkeit der Alternativhypothese näherungsweise bestimmt. Diese Ablehnungswahrscheinlichkeiten sind in der Abbildung 8.3 dargestellt. Für $\vartheta_0 = 0$ liegt die Ablehnungswahrscheinlichkeit mit rund 3,1 Prozent etwas unter dem Nominallevel von $\alpha = 0{,}05$. Die Rate steigt jedoch schnell an und liegt bereits für $\vartheta_0 = 0{,}01$ bei 24,8 Prozent, was für eine hohe Testgüte spricht.

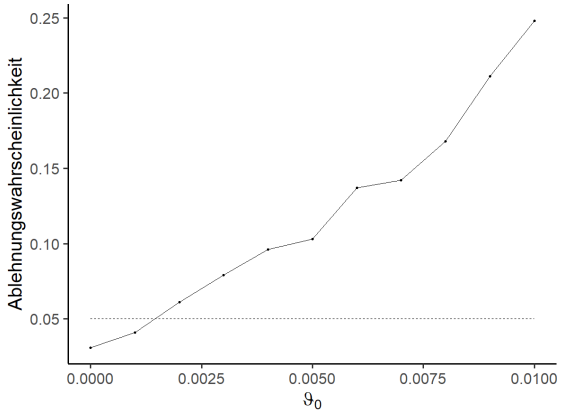

Abbildung 8.3.: Anteil der Ablehnungen von $H_0 : \vartheta_0 = 0$ für verschiedene ϑ^1-Werte unter $H_1 : \vartheta_0 = \vartheta^1$

8.2. Kolmogorov-Smirnov-Anpassungstest

Für den Fall der unabhängigen Doppeltrunkation wurde in Kapitel 6 ein Kolmogorov-Smirnov-ähnlicher Anpassungstest entwickelt. Im Rahmen dessen erfolgte die Definition der Teststatistik, deren Berechnung sowie die Untersuchung ihres asymptotischen Verhaltens. Der vorliegende Abschnitt befasst sich mit der Durchführung von Simulationen, um die Korrektheit der Fehlerwahrscheinlichkeiten erster Art zu demonstrieren. Für die Simulationen wird dafür zwischen den vier betrachteten Stufen unterschieden, je nachdem ob der Parameter der Exponentialverteilung θ_0 oder die latente Stichprobengröße n bekannt sind, oder geschätzt werden müssen. In Anbetracht des erheblichen Rechenaufwands wird im Vergleich zum vorherigen Abschnitt lediglich ein Szenario für die Simulationen herangezogen. Die Auswahl der einzelnen Größen orientiert sich dabei an dem Anwendungsbeispiel aus Kapitel 7.

Vor der Durchführung der Simulationen des gesamten Anpassungstests soll zunächst die den Algorithmen 1-4 zugrunde liegende Methode zur Approximation der asymptotischen Verteilung der Teststatistik untersucht werden. Die Untersuchungen werden am eindimensionalen Fall vorgenommen. Nach Gleichung (6.1) ist die Verteilung von $\|\mathbb{B}\|_\infty$, mit $\mathbb{B}$ als Brownsche Brücke, die Kolmogorov-Verteilung. Mit dem folgenden Algorithmus werden Quantile für verschiedene Gitterbreiten simuliert und mit den Quantilen der Kolmogorov-Verteilung verglichen.
Die Ergebnisse können der darauf folgenden Abbildung entnommen werden. Die Feinheit des Gitters wird dabei so variiert, dass das Intervall $[0, 1]$ in 25 bis 5000 Abschnitte zerlegt wird, im Folgenden als Gitterfrequenz bezeichnet. Es werden jeweils 1000 Realisationen der Zufallsvariable $\|\mathbb{B}\|_\infty$ generiert, aus denen das Quantil ausgewählt wird. Im Diagramm sind die ermittelten $(1-\alpha)$-Quantile für $\alpha = 0{,}1;\ 0{,}05;\ 0{,}01$ abgetragen. Die gestrichelte Linie kennzeichnet das entsprechende Quantil der Kolmogorov-Verteilung. Für alle drei α-Niveaus ist eine deutliche Annäherung zu erkennen.

Algorithmus 6: Simulation der Verteilung von $\|\mathbb{B}\|_\infty$

1. Erzeuge ein Gitter auf dem Intervall $[0, 1]$.
2. Berechne für alle Gitterpunkte (s, t) die Kovarianzmatrix mittels $\min\{s, t\} - st$.
3. Wiederhole für eine festgelegte Anzahl von Iterationen:
   i) Erzeuge einen normalverteilten Vektor mit Erwartungswert null und der berechneten Kovarianzmatrix.
   ii) Ermittle das betragsmäßige Maximum des Vektors.
4. Wähle aus den sortierten Maxima das entsprechende Quantil aus.

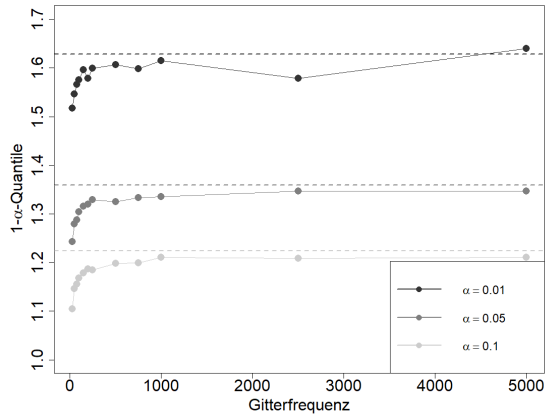

Abbildung 8.4.: Darstellung der in Algorithmus 6 simulierten Quantile in Abhängigkeit von α und der Gitterfrequenz im Vergleich zu den Quantilen der Kolmogorov-Verteilung (6.1) (gestrichelt dargestellt)

Die Ermittlung der Teststatistik nach Algorithmus 5 setzt für alle vier Stufen die Simulation der trunkierten Stichprobe voraus. Aufgrund der Unabhängigkeit zwischen X und T erübrigt sich die im vorherigen Abschnitt zur Simulation genutzte bedingte Inversionsmethode 2.23. In diesem Fall ist die gewöhnliche Inversionsmethode ausreichend. Die beiden Randverteilungen werden gemäß Annahme (A2) und die Copula als Produktcopula gemäß Annahme (A3''') gewählt.

- Erzeuge zwei unabhängige Realisationen u und v einer $(0, 1)$-Gleichverteilung.
- Erhalte mit $x = -\log(1-u)/\theta_0$ und $t = v \cdot G$ die gesuchte Realisation von (X, T).

Bei der Wahl der Gründungsperiode wird eine Länge von $G = 24$ und bei der Beobachtungsperiode eine Länge von $s = 3$ festgelegt, wie es im vorliegenden Anwendungsbeispiel geschieht. So verbleibt nach dem Trunkieren der Ziehungen außerhalb des Beobachtungsparallelogramms D (vgl. Annahme (A4)) bei einem Parameterwert

von $\theta_0 = 0{,}08261$ ein Anteil von rund $\alpha_\theta = 0{,}095$ der latenten Stichprobengröße. Letztere wird für die Simulation einmal auf $n = 20.000$ und einmal auf 50.000 gesetzt.

Simulationen des Anpassungstests bei (θ_0, n)

In der ersten Stufe werden sowohl die Bekanntheit von θ_0 als auch die Bekanntheit von n vorausgesetzt. Dies führt zu den Hypothesen (vgl. Gleichung (6.3)):

H_0 : Der Zufallsvektor $(X_i, T_i)^{\mathrm{T}}$ besitzt die Verteilung $F_{\theta_0} = F^X \cdot F^T$ mit dem Parameter $\theta_0 \in \Theta$.

H_1 : Der Zufallsvektor $(X_i, T_i)^{\mathrm{T}}$ besitzt nicht die Verteilung $F_{\theta_0} = F^X \cdot F^T$ mit dem Parameter $\theta_0 \in \Theta$.

Für die verbleibenden Beobachtungen nach Trunkation kann die Teststatistik unter Anwendung von Algorithmus 5 berechnet werden. Im Anschluss werden die Quantile der asymptotischen Verteilung der Teststatistik gemäß Algorithmus 1 berechnet. Zu diesem Zweck wird ein Gitter auf das Rechteck $[0, G] \times [0, G + s]$ gelegt, wodurch die beiden Intervalle in jeweils 50 äquidistante Teile zerlegt werden. Zur Quantilsermittlung werden 1000 Realisationen der Zufallsvariable $\|\mathbb{B}_{\mathbb{P}_{\theta_0}}\|_{\mathcal{G}_D}$ generiert. Überschreitet der Wert der Prüfgröße das jeweilige Quantil, wird die Nullhypothese abgelehnt. Die Ablehnungsraten für die verschiedenen Quantile und Simulationsdurchläufe werden in der folgenden Tabelle festgehalten.

Tabelle 8.2.: Ablehnungsraten der Teststatistik (θ_0 und n bekannt) für verschiedene $(1 - \alpha)$-Quantile und Simulationsdurchläufe (n_{sim}) mit $G = 24$, $s = 3$, $\theta_0 = 0{,}08261$ und einer Gitterfrequenz von 50

	$n = 20.000$			$n = 50.000$		
$n_{sim} \setminus \alpha$	0,10	0,05	0,01	0,10	0,05	0,01
10	10,0%	0,0%	0,0%	0,0%	0,0%	0,0%
100	12,0%	7,0%	1,0%	11,0%	7,0%	2,0%
1000	9,9%	5,4%	1,2%	12,3%	5,7%	1,8%

Die ermittelten Ablehnungsraten zeigen für beide latente Stichprobengrößen eine gute Übereinstimmung mit den ausgewählten Nominalraten ($\alpha = 0{,}1; 0{,}05; 0{,}01$). Ähnlich wie in Abbildung 8.4 verläuft die Annäherung nicht monoton in n.

Simulationen des Anpassungstests bei $(\hat{\theta}_n, n)$

Wird auf die Annahme der Bekanntheit des Parameters verzichtet, lauten die Hypothesen (vgl. Gleichung (6.8))wie folgt:

H_0 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt aus der Verteilungsfamilie $\{F_\theta = F^X \cdot F^T \,|\, \theta \in \Theta\}$.

H_1 : Die Verteilung des Zufallsvektors $(X_i, T_i)^{\mathrm{T}}$ stammt nicht aus der Verteilungsfamilie $\{F_\theta = F^X \cdot F^T \,|\, \theta \in \Theta\}$.

Nach Trunkation kann der Parameter anhand der verbleibenden Beobachtungen mithilfe der Likelihood aus Gleichung (2) in [65] geschätzt werden. Die Teststatistik kann wie in Algorithmus 5 beschrieben berechnet werden, indem der wahre Parameter durch den geschätzten Parameter ersetzt wird. Im Anschluss werden die Quantile der asymptotischen Verteilung der Teststatistik gemäß Algorithmus 2 berechnet. Die Zerlegung des Gitters und die Berechnung der Ablehnungsraten erfolgt wie bereits im obigen Fall beschrieben.

Tabelle 8.3.: Ablehnungsraten der Teststatistik (θ_0 geschätzt und n bekannt) für verschiedene $(1-\alpha)$-Quantile und Simulationsdurchläufe (n_{sim}) mit $G = 24$, $s = 3$, $\theta_0 = 0{,}08261$ und einer Gitterfrequenz von 50

	$n = 20.000$			$n = 50.000$		
$n_{sim} \setminus \alpha$	0,10	0,05	0,01	0,10	0,05	0,01
10	20,0%	10,0%	0,0%	10,0%	0,0%	0,0%
100	9,0%	3,0%	0,0%	11,0%	6,0%	0,0%
1000	11,2%	5,2%	1,3%	12,0%	6,5%	1,5%

Entgegen der Erwartung zeigen sich für diese Stufe ähnlich zufriedenstellende Ergebnisse. Größere Unterschiede in den Ablehnungsraten bei der Berücksichtigung der Schätzung von θ werden vermutlich bei anderen Größen von n_{sim} und n beobachtbar.

Simulationen des Anpassungstests bei $(\theta_0, \hat{n})$

In der dritten Stufe wird die Kenntnis des wahren Parameterwerts θ_0 vorausgesetzt, während der latente Stichprobenumfang n geschätzt werden muss. Die Unbekanntheit von n hat keinen Einfluss auf die Hypothesen, sodass auch hier die Hypothesen (6.3) der ersten Stufe untersucht werden.
Die Schätzung des latenten Stichprobenumfangs ergibt sich als Quotient aus der Anzahl der Beobachtungen und der Auswahlwahrscheinlichkeit. Die Teststatistik kann

wie in Algorithmus 5 beschrieben berechnet werden, indem der Stichprobenumfang n durch den geschätzten Wert ersetzt wird. Im Anschluss werden die Quantile der asymptotischen Verteilung der Teststatistik gemäß Algorithmus 3 berechnet. Die ermittelten Ablehnungsraten für die verschiedenen Quantile und Simulationsdurchläufe werden nachfolgend in tabellarischer Form dargelegt.

Tabelle 8.4.: Ablehnungsraten der Teststatistik (θ_0 bekannt und n geschätzt) für verschiedene $(1-\alpha)$-Quantile und Simulationsdurchläufe (n_{sim}) mit $G = 24$, $s = 3$, $\theta_0 = 0{,}08261$ und einer Gitterfrequenz von 50

	$n = 20.000$			$n = 50.000$		
$n_{sim} \setminus \alpha$	0,1	0,05	0,01	0,1	0,05	0,01
10	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
100	16,0%	12,0%	4,0%	15,0%	8,0%	0,0%
1000	14,8%	7,1%	1,3%	14,6%	7,0%	1,6%

Bei der Berücksichtigung der Schätzung des latenten Stichprobenumfangs ist eine höhere Ungenauigkeit in den Ablehnungsraten zu beobachten als in den vorangegangenen Stufen. Die beobachteten Ablehnungsraten zeigen eine leichte Überschreitung der gewählten Nominalniveaus. Eine Annäherung wird in dieser Stufe sichtbar, wenn der latente Stichprobenumfang und die Anzahl der Simulationsdurchläufe erhöht werden.

Simulationen des Anpassungstests bei $(\hat{\theta}_n, \hat{n})$

Schließlich lassen sich die beiden zuvor betrachteten Stufen miteinander verknüpfen, sodass sowohl die Schätzung des Parameters der Exponentialverteilung als auch die des latenten Stichprobenumfangs einbezogen werden. Die zu testenden Hypothesen entsprechen denen in Gleichung (6.8).
Wie bereits in der Stufe $(\hat{\theta}_n, n)$ beschrieben, kann der Parameter der Exponentialverteilung nach Trunkation anhand der verbleibenden Beobachtungen mithilfe der Likelihood aus Gleichung (2) in [65] geschätzt werden. Die Schätzung von n berechnet sich wie für die Stufe $(\theta_0, \hat{n})$: als Quotient aus der Anzahl der Beobachtungen und der Auswahlwahrscheinlichkeit, wobei auch die Auswahlwahrscheinlichkeit von $\hat{\theta}_n$ abhängt. Für die Berechnung der Teststatistik nach Algorithmus 5 muss also zunächst n durch $\hat{n}$ und anschließend jedes Vorkommen von θ_0 durch $\hat{\theta}_n$ ersetzt werden. Die Vorgehensweise zur Berechnung der Quantile wird in Algorithmus 4 beschrieben. Die erhaltenen Ablehnungsraten sind in der folgenden Tabelle dargestellt.

Tabelle 8.5.: Ablehnungsraten der Teststatistik (θ_0 geschätzt und n geschätzt) für verschiedene $(1-\alpha)$-Quantile und Simulationsdurchläufe (n_{sim}) mit $G = 24$, $s = 3$, $\theta_0 = 0{,}08261$ und einer Gitterfrequenz von 50

	$n = 20.000$			$n = 50.000$		
$n_{sim} \setminus \alpha$	0,1	0,05	0,01	0,1	0,05	0,01
10	20,0%	10,0%	0,0%	40,0%	40,0%	20,0%
100	17,0%	11,0%	0,0%	22,0%	15,0%	2,0%
1000	23,9%	14,3%	3,5%	23,8%	13,8%	4,2%

Hinsichtlich der praktischen Umsetzung des Anpassungstests ist diese Stufe von entscheidender Relevanz, da in der Praxis weder der wahre Parameterwert θ noch der latente Stichprobenumfang n bekannt oder direkt beobachtbar sind. Die Simulationen verdeutlichen, dass die Konstruktion eines exakten Tests zum Niveau α nicht nur analytisch komplex, sondern auch numerisch anspruchsvoll ist. Die in den Simulationen beobachteten Abweichungen zwischen Prüfgröße und Quantil, die die jeweiligen Ablehnungsraten bestimmen, fallen jedoch sehr gering aus. Die Testergebnisse für das vorliegende Anwendungsbeispiel erscheinen daher glaubhaft, da hier die Differenz zwischen Prüfgröße und Quantil im Vergleich zu den Simulationen größer ist. Zudem ist der latente Stichprobenumfang im Anwendungsfall etwa zehnmal so groß wie in den Simulationen, was eine noch stärkere Annäherung der beobachteten Ablehnungsraten an das Nominalniveau erwarten lässt.

9. Fazit

Die bei der Analyse von links- und doppeltrunkierten Daten häufig verwendete Annahme der Unabhängigkeit zwischen der interessierenden Lebensdauer und der Trunkationszeit wurde in dieser Arbeit durch eine konkrete Modellierung ihrer Abhängigkeitsstruktur ersetzt. Mithilfe einer Copula konnte der Ansatz zur Doppeltrunkation in Weißbach und Wied [65] so erweitert werden, dass die bisherigen Annahmen über die Randverteilungen erhalten bleiben. Für die konstruierte Schätzfolge konnten sowohl Konsistenz als auch asymptotische Normalität nachgewiesen werden. Umfangreiche Simulationen unter verschiedenen Szenarien stützen die theoretisch hergeleiteten Eigenschaften. Der Spezialfall der Unabhängigkeit für die Gumbel-Barnett-Copula wurde gesondert betrachtet und ein entsprechender Parametertest konnte hergeleitet werden. Die Hypothese der Unabhängigkeit konnte für unternehmensdemographische Daten verworfen werden. Für den vorliegenden Datensatz beträgt die durchschnittliche Lebensdauer etwa 12 Jahre. Analysen mit verschiedenen Copulas zeigen, dass die Lebenserwartung deutscher Unternehmen im Zeitverlauf abnimmt. Bei Verwendung der Farlie-Gumbel-Morgenstern-Copula ergibt sich eine Verringerung der bedingten Lebenserwartung um etwa 19 Tage pro Jahr. Die Abhängigkeit, gemessen mit Kendalls Tau, ist nur schwach ausgeprägt, ähnlich wie im Fall der spanischen Unternehmensdaten in de Uña-Álvarez et al. [9]. Die Schließung von Unternehmen kann durch vielfältige Ursachen bedingt sein. Wie in den in der Einleitung zitierten Studien (Xuezhou et al. [66], Miao et al. [40] sowie de Uña-Álvarez et al. [9]) dargelegt, lassen sich die Ursachen für die abnehmende Lebensdauer genauer mithilfe von Kovariablen analysieren, die die modellierte Abhängigkeit beeinflussen oder möglicherweise sogar ersetzen. Sieg, Toparkus und Weißbach [54] erweitern den Ansatz zur Doppeltrunkation aus Weibach und Wied [65] um den Aspekt der Zensierung. Auch hier kann eine Copula verwendet werden, um mögliche Abhängigkeiten zwischen der Lebensdauer und den Trunkations- und Zensierungsvariablen zu modellieren.

Der zweite Teil der Arbeit widmete sich der in der Exponentialverteilung implizierten Annahme der Altershomogenität in den Unternehmensdaten. Ziel war die Entwicklung eines Kolmogorov-Smirnov-Anpassungstests zur Überprüfung spezifischer Verteilungsannahmen bei doppeltrunkierten Daten. Zur Vereinfachung wird zunächst angenommen, dass die Variablen unabhängig sind und sowohl der zugrunde liegende Parameter als auch der latente Stichprobenumfang bekannt sind. Die entsprechenden Schätzer werden schrittweise berücksichtigt und für jede Stufe wird die asymptotische Verteilung des trunkierten Prozesses bestimmt. Die erzielten Resultate bleiben auch im Fall einer zugrunde liegenden Copula gültig. Es ist lediglich zu beachten, dass dadurch ein zusätzlicher Parameter eingeführt wird. Die für die bivariate Schätzfolge erforderlichen Eigenschaften wurden im ersten Teil der Arbeit nachgewiesen. Der zusätzliche

Rechenaufwand, der durch die numerische Berechnung der Integrale entsteht, lässt sich verringern, indem die bivariaten Integrale, wie in Emura und Pan [15] gezeigt, geschickt in eindimensionale Integrale überführt werden. Die Prüfgröße kann aufgrund der Kompaktheit des Beobachtungsparallelogramms mithilfe der Methode von Justel, Peña und Zamar [27] berechnet werden. Dabei verursacht die Bestimmung der Knotenpunkte den größten Rechenaufwand, trägt jedoch zur höheren Genauigkeit bei. Die Anwendung des Kolmogorov-Smirnov-ähnlichen Tests auf die Daten deutscher Unternehmenslebensdauern weist die in Weißbach und Wied [65] zugrunde gelegte bivariate Verteilung zurück. Dies steht im Einklang mit den Ergebnissen des Unabhängigkeitstests. Der Test wurde ebenfalls für die Farlie-Gumbel-Morgenstern-Copula durchgeführt und führte ebenfalls zum Verwerfen der Nullhypothese. Zusammen mit den Ergebnissen des Unabhängigkeitstests könnte dies auf eine ungeeignete Wahl der Randverteilungen hindeuten. Somit kann auch die Altershomogenität nicht gestützt werden, sodass andere Lebensdauerverteilungen herangezogen werden sollten. Ein aktueller Befund von Weißbach und Dörre [64] widerspricht der Annahme einer gleichverteilten Trunkierungsvariable. Sie zeigen, dass Unternehmensgründungen in den Jahren 1990 bis 2013 zunehmend seltener wurden, allerdings unter der Annahme unabhängiger Trunkierung. Alternativ ließe sich ein semiparametrischer Ansatz wie in Genest et al. [18] in Betracht ziehen, bei dem Copula-Parameter ohne Annahmen über Randverteilungen geschätzt werden.
Auch zum Anpassungstest wurden ergänzend Simulationen durchgeführt. Aufgrund des hohen Rechenaufwands beschränken sich diese auf die Typ-I-Fehlerraten und ein einzelnes, an der Anwendung orientiertes Szenario. Analog zu Emura und Konno [14] könnten die Simulationen künftig auf weitere Randverteilungen oder Copulas ausgeweitet und um Analysen zur Teststärke ergänzt werden. Zudem wäre ein Vergleich mit der Multiplikator-Methode aus Kojadinovic und Yan [32] von Interesse.

Zum betrachteten Datensatz ist anzumerken, dass ausschließlich Insolvenz als Schließungsgrund berücksichtigt wird. Dies stellt insofern einen Modellfehler dar, als dass Firmen auch aus anderen Gründen, etwa durch freiwillige Geschäftsaufgabe, Fusion oder behördliche Anordnung, geschlossen werden können. Die Einbeziehung solcher Alternativen würde ein Modell konkurrierender Ereignisse erfordern. Streng genommen handelt es sich bei den beobachteten Lebensdauern nicht um eine Zufallsstichprobe, sondern um die gesamte (trunkierte) Grundgesamtheit. Die entwickelte Schätzmethode kann dennoch angewendet werden, jedoch ist die Interpretation der Standardfehler problematisch, da die Zufälligkeit der Stichprobenziehung nicht gegeben ist.
Die Lebensdauer von Unternehmen stellt nicht nur für Banken und Versicherungen eine wichtige Information zur Einschätzung von Ausfallrisiken dar, sondern dient auch Investoren sowie wirtschaftspolitischen Entscheidungsträgern als wertvolle Orientierungshilfe. Untersuchungen zu Gründungen, Schließungen und Ausfallraten von Unternehmen bilden daher ein bedeutendes Feld der wirtschaftswissenschaftlichen Forschung. Gleichwohl beschränkt sich die Anwendung der entwickelten Tests nicht ausschließlich auf unternehmensdemografische Daten. Beliebige doppeltrunkierte Dauern lassen sich damit analysieren, insbesondere wenn die Annahme der Unabhängigkeit fraglich ist oder spezielle Verteilungsannahmen geprüft werden sollen.

Literaturverzeichnis

[1] M. G. Akritas, M. P. LaValley: *A generalized product-limit estimator for truncated data*. Journal of Nonparametric Statistics 17 (6), pp. 643-663 (2005)

[2] P. K. Andersen, Ø. Borgan, R. Gill, N. Keiding: *Censoring, Truncation and Filtering in Statistical Models Based on Counting Processes*. Statistical inference from stochastic processes 80, pp. 19-60. Centre for Mathematics and Computer Science, Amsterdam (1988)

[3] P. Billingsley: *Convergence of Probability Measures*. 2nd Edition, John Wiley & Sons, Inc., New York (1999)

[4] H. Büning, G. Trenkler: *Nichtparametrische statistische Methoden*. Walter de Gruyter & Co., Berlin (1978)

[5] H. Cramér: *On the Composition of Elementary Errors*. Scandinavian Actuarial Journal 1, pp. 13-27 (1928)

[6] L. L. Chaieb, L. Rivest: *Estimating survival under a dependent truncation*. Biometrika 93 (3), pp. 655-669 (2006)

[7] S. H. Chiou, M. D. Austin, J. Qian, R. A. Betensky: *Transformation model estimation of survival under dependent truncation and independent censoring*. Statistical Methods in Medical Research 28 (12), pp. 3785-3798 (2019)

[8] D. J. Daley, D. Vere-Jones: An Introduction to the Theory of Point Processes. 2nd Edition, Vol. I, Springer-Verlag, New York (2008)

[9] J. de Uña-Álvarez, A. I. Martínez-Senra, M. S. Otero-Giráldez, M. A. Quintás: *Cox regression with doubly truncated responses and time-dependent covariates: the impact of innovation on firm survival*. Journal of Applied Statistics 51 (4), pp. 780-792 (2024)

[10] L. Devroye: *Non-Uniform Random Variate Generation*. Springer-Verlag, New York (1986)

[11] B. Efron, V. Petrosian: *Survival Analysis of the Gamma-Ray Burst Data*. Journal of the American Statistical Association 89 (426), pp. 452-462 (1994)

[12] B. Efron, V. Petrosian: *Nonparametric Methods for Doubly Truncated Data*. Journal of the American Statistical Association 94 (447), pp. 824-834 (1999)

[13] J. Elstrodt: *Maß- und Integrationstheorie*. 8. Aufl., Springer-Verlag, Berlin (2018)

[14] T. Emura, Y. Konno: *A goodness-of-fit test for parametric models based on dependently truncated data*. Computational Statistics & Data Analysis 56 (7), pp. 2237-2250 (2012)

[15] T. Emura, C.-H. Pan: *Parametric Likelihood inference and goodness-of-fit for dependently left-truncated data, a copula-based approach*. Statistical Papers 61, pp. 479-501 (2020)

[16] T. Emura, W. Wang: *Nonparametric maximum likelihood estimation for dependent truncation data based on copulas*. Journal of Multivariate Analysis 110, pp. 171-188 (2012)

[17] O. Forster: *Analysis 2*. 11. Aufl., Springer Spektrum, Wiesbaden (2017)

[18] C. Genest, K. Ghoudi, L. Rivest: *A semiparametric estimation procedure of dependence parameters in multivariate families of distributions*. Biometrika 82 (3), pp. 543-552 (1995)

[19] C. Genest, B. Rémillard, D. Beaudoin: *Goodness-of-fit tests for copulas: A review and a power study*. Insurance: Mathematics and Economics 44 (2), pp. 199-213 (2009)

[20] C. Gouriéroux, A. Monfort: *Statistics and Econometric Models*. Vol. 1, Cambridge University Press, New York (1995)

[21] C. Gouriéroux, A. Monfort: *Statistics and Econometric Models*. Vol. 2, Cambridge University Press, New York (1995)

[22] S. T. Gross, T. L. Lai: *Nonparametric Estimation and Regression Analysis with Left-Truncated and Right-Censored Data*. Journal of the American Statistical Association 91 (435), pp. 1166-1180 (1996)

[23] E. J. Gumbel: *Bivariate Exponential Distributions*. Journal of the American Statistical Association 55 (292), pp. 698-707 (1960)

[24] J. J. Heckman: *The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models*. Annals of Economic and Social Measurement 5 (4), pp. 475-492 (1976)

[25] M. Hollander, D. A. Wolfe, E. Chicken: *Nonparametric Statistical Methods*. 3rd Edition, John Wiley & Sons, Inc., Hoboken, New Jersey (2014)

[26] H. Joe: *Multivariate Models and Dependence Concepts*. CRC Press, Boca Raton (1997)

[27] A. Justel, D. Peña, R. Zamar: *A multivariate Kolmogorov-Smirnov test of goodness of fit*. Statistics and Probability Letters 35, pp. 251-259 (1997)

[28] E. L. Kaplan, P. Meier: *Nonparametric Estimation from Incomplete Observations*. Journal of the American Statistical Association 53 (282), pp. 457-481 (1958)

[29] J. P. Klein, P. K. Goel: *Survival Analysis: State of the Art*. Kluwer Academic Publishers, Dordrecht (1992)

[30] J. P. Klein, M. L. Moeschberger: *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd Edition, Springer-Verlag, New York (2003)

[31] A. Klenke: *Wahrscheinlichkeitstheorie*. 3. Aufl., Springer Spektrum, Berlin (2013)

[32] I. Kojadinovic, J. Yan: *A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems*. Statistics and Computing 21, pp. 17-30 (2011)

[33] A. Kolmogorov: *Sulla Determinazione Empirica di Una Legge di Distribuzione*. Giornale dell'Istituto Italiano degli Attuari 4, pp. 89-91 (1933)

[34] J. F. Lawless: *Statistical Models and Methods for Lifetime Data*. 2nd Edition, John Wiley & Sons, Inc., Hoboken, New Jersey (2003)

[35] H. W. Lilliefors: *On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown*. Journal of the American Statistical Association 62 (318), pp. 399–402 (1967)

[36] H. W. Lilliefors: *On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown*. Journal of the American Statistical Association 64 (325), pp. 387–389 (1969)

[37] P. Ma, Y. Zhang: *Modeling asymmetrically dependent multivariate ocean data using truncated copulas*. Ocean Engineering 244, 110226 (2022)

[38] E. C. Martin, R. A. Betensky: *Testing Quasi-Independence of Failure and Truncation Times via Conditional Kendall's Tau*. Journal of the American Statistical Association 100 (470), pp. 484-492 (2005)

[39] Ministerium der Justiz Nordrhein-Westfalen: *Insolvenzbekanntmachungen*. (2023) https://neu.insolvenzbekanntmachungen.de/ap/index.jsf

[40] W. Miao, J. Beyhum, J. Striaukas, I. van Keilegom: *High-dimensional censored MIDAS logistic regression for corporate survival forecasting*. arXiv:2502.09740 (2025)

[41] P. A. P. Moran: *Maximum-Likelihood estimation in non-standard conditions*. Mathematical Proceedings of the Cambridge Philosophical Society 70, pp. 441-450 (1971)

[42] C. Moreira, J. de Uña-Álvarez: *A semiparametric estimator of survival for doubly truncated data*. Statistics in Medicine 29 (30), pp. 3147-3159 (2010)

[43] C. Moreira, J. de Uña-Álvarez, R. Braekers: *Nonparametric estimation of a distribution function from doubly truncated data under dependence*. Computational Statistics 36, pp. 1693-1720 (2021)

[44] M. Naaman: *On the tight constant in the multivariate Dvoretzky–Kiefer–Wolfowitz inequality*. Statistics and Probability Letters 173, 109088 (2021)

[45] R. B. Nelsen: *An Introduction to Copulas*. 2nd Edition, Springer Science+Business Media, New York (2006)

[46] W. K. Newey, D. McFadden: *Large Sample Estimation and Hypothesis Testing, Handbook of Econometrics*. Vol. 4, ch. 36, p. 2129, Elsevier Science B.V. (1994)

[47] M. Ossiander: *A Central Limit Theorem Under Metric Entropy with L_2 Bracketing*. The Annals of Probability 15 (3), pp. 897-919 (1987)

[48] R.-D. Reiss: *A Course an Point Processes*. Springer-Verlag, New York (1993)

[49] M. Rosenblatt: *Remarks on a Multivariate Transformation*. The Annals of Mathematical Statistics 23 (3), pp. 470-472 (1952)

[50] L. Rüschendorf: *Mathematische Statistik*. Springer-Verlag, Berlin (2014)

[51] M. J. Schervish: *Theory of Statistics*. Springer-Verlag, New York (1995)

[52] P. S. Shen: *Nonparametric analysis of doubly truncated data*. Annals of the Institute of Statistical Mathematics 62, pp. 835-853 (2010)

[53] S. E. Shreve: *Stochastic Calculus for Finance II*. Springer-Verlag, New York (2010)

[54] F. Sieg, A. Toparkus, R. Weißbach: *Censored lifespans in a double-truncated sample: Maximum likelihood inference for the exponential distribution*. arXiv:2504.16623 (2025)

[55] N. Smirnov: *Sur les écarts de la courbe de distribution empirique*. Matematicheskii Sbornik 48 (1), pp. 3-26 (1939)

[56] N. Smirnov: *Table for Estimating the Goodness of Fit of Empirical Distributions*. The Annals of Mathematical Statistics 19 (2), pp. 279-281 (1948)

[57] H. Theil: *Principles of Econometrics*. John Wiley & Sons, Inc., Amsterdam (1971)

[58] A. Toparkus, R. Weißbach: *Testing truncation dependence: The Gumbel-Barnett copula*. Journal of Statistical Planning and Inference 234, 106194 (2025)

[59] W. Tsai: *Testing the assumption of independence of truncation time and failure time*. Biometrika 77 (1), pp. 169-177 (1990)

[60] B. W. Turnbull: *The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data*. Journal of the Royal Statistical Society 38 (3), pp. 290-295 (1976)

[61] A. W. van der Vaart: *Asymptotic Statistics*. Cambridge University Press, New York (2000)

[62] A. W. van der Vaart, J. A. Wellner: *Weak Convergence and Empirical Processes*. Springer-Verlag, New York (1996)

[63] R. von Mises: *Wahrscheinlichkeit, Statistik und Wahrheit*. Springer-Verlag, Wien (1928)

[64] R. Weißbach, A. Dörre: *Retrospective sampling of survival data based on a Poisson birth process: conditional maximum likelihood*. Statistics 56 (4), pp. 844–866 (2022)

[65] R. Weißbach, D. Wied: *Truncating the exponential with a uniform distribution*. Statistical Papers 63, pp. 1247-1270 (2022)

[66] W. Xuezhou, R. Y. Hussain, A. A. Salameh, H. Hussain, A. B. Khan, M. Fareed: *Does Firm Growth Impede or Expedite Insolvency Risk? A Mediated Moderation Model of Leverage Maturity and Potential Fixed Collaterals*. Frontiers in Environmental Science 10, 841380 (2022)

A. Appendix

A.1. Score-Funktion und Schätzer in M^*

Um die Gültigkeit von (4.3) zu überprüfen, werden zunächst die Ableitungen der logarithmierten Dichte $f^*_{\boldsymbol{\theta}}$ aus (4.1), d.h. von

$$\begin{aligned}\log f^*_{\boldsymbol{\theta}}(x,t) =& \log(\theta) - \log(G) - \log(\alpha_{\boldsymbol{\theta}}) - \theta x + \vartheta\theta x \log(1-t/G) + \\ &+ \log\left[(\vartheta\theta x+1)(1-\vartheta\log(1-x/G)) - \vartheta\right], \quad (x,t) \in D\end{aligned}$$

benötigt. Diese lauten

$$\begin{aligned}\frac{\partial \log f^*_{\boldsymbol{\theta}}(x,t)}{\partial\theta} =& \frac{1}{\theta} + x(\vartheta\log(1-t/G)-1) - \frac{1}{\alpha_{\boldsymbol{\theta}}}\cdot\frac{\partial\alpha_{\boldsymbol{\theta}}}{\partial\theta} + \\ &+ \frac{\vartheta x(\vartheta\log(1-t/G)-1)}{(\vartheta\theta x+1)(\vartheta\log(1-t/G)-1)+\vartheta}\end{aligned}$$

sowie

$$\begin{aligned}\frac{\partial \log f^*_{\boldsymbol{\theta}}(x,t)}{\partial\vartheta} =& \theta x\log(1-t/G) - \frac{1}{\alpha_{\boldsymbol{\theta}}}\cdot\frac{\partial\alpha_{\boldsymbol{\theta}}}{\partial\vartheta} + \\ &+ \frac{(2\vartheta\theta x+1)\log(1-t/G) - \theta x + 1}{(\vartheta\theta x+1)(\vartheta\log(1-t/G)-1)+\vartheta}.\end{aligned}$$

In Hinblick auf die Gleichungen (3.6a) und (3.6b) folgt

$$\frac{\partial \log f^*_{\boldsymbol{\theta}}(X_i^*,T_i^*)}{\partial\theta} = \psi_{\boldsymbol{\theta},1}(X_i^*,T_i^*) \quad \text{und} \quad \frac{\partial \log f^*_{\boldsymbol{\theta}}(X_i^*,T_i^*)}{\partial\vartheta} = \psi_{\boldsymbol{\theta},2}(X_i^*,T_i^*) \tag{A.1}$$

wegen $(X_i^*,T_i^*)^{\mathrm{T}} \in D$ und somit Gleichung (4.3). Insbesondere erhält man daraus die Übereinstimmung der Score-Funktionen von M^* und dem Gumbel-Barnett-Modell nach dem Ersetzen von n. Aus der Likelihood

$$\begin{aligned}\ell^*(\boldsymbol{\theta}) = \prod_{i=1}^m \frac{f_{\boldsymbol{\theta}}(X_i^*,T_i^*)}{\alpha_{\boldsymbol{\theta}}} =& \frac{\theta^m}{G^m\alpha_\theta^m}\exp\left(-\theta\sum_{i=1}^m X_i^*\right)\cdot \\ &\cdot\prod_{i=1}^m\left(1-\frac{T_i^*}{G}\right)^{\vartheta\theta X_i^*}\left[(\vartheta\theta X_i^*+1)(1-\vartheta\log(1-T_i^*/G))-\vartheta\right]\end{aligned}$$

lässt sich die logarithmierte Likelihood

$$\begin{aligned}\log \ell^*(\boldsymbol{\theta}) =& m\log(\theta) - m\log(G) - m\log(\alpha_{\boldsymbol{\theta}}) - \\ &- \theta\sum_{i=1}^{m} X_i^* + \sum_{i=1}^{m} \vartheta\theta X_i^* \log(1 - T_i^*/G) + \\ &+ \sum_{i=1}^{m} \log\left[(\vartheta\theta X_i^* + 1)(1 - \vartheta\log(1 - T_i^*/G)) - \vartheta\right]\end{aligned}$$

mit den partiellen Ableitungen nach θ

$$\begin{aligned}\frac{\partial \log \ell^*(\boldsymbol{\theta})}{\partial\theta} =& \frac{m}{\theta} + \sum_{i=1}^{m} X_i^*(\vartheta\log(1 - T_i^*/G) - 1) - \frac{m}{\alpha_{\boldsymbol{\theta}}} \cdot \frac{\partial\alpha_{\boldsymbol{\theta}}}{\partial\theta} + \\ &+ \sum_{i=1}^{m} \frac{\vartheta X_i^*(\vartheta\log(1 - T_i^*/G) - 1)}{(\vartheta\theta X_i^* + 1)(\vartheta\log(1 - T_i^*/G) - 1) + \vartheta}\end{aligned}$$

und nach ϑ

$$\begin{aligned}\frac{\partial \log \ell^*(\boldsymbol{\theta})}{\partial\vartheta} =& \theta\sum_{i=1}^{m} X_i^* \log(1 - T_i^*/G) - \frac{m}{\alpha_{\boldsymbol{\theta}}} \cdot \frac{\partial\alpha_{\boldsymbol{\theta}}}{\partial\vartheta} + \\ &+ \sum_{i=1}^{m} \frac{(2\vartheta\theta X_i^* + 1)\log(1 - T_i^*/G) - \theta X_i^* + 1}{(\vartheta\theta X_i^* + 1)(\vartheta\log(1 - T_i^*/G) - 1) + \vartheta}\end{aligned}$$

berechnen. Diese stimmen mit den Ableitungen der logarithmierten Likelihood in Abschnitt 3.3 überein, nachdem n durch $M_n/\alpha_{\boldsymbol{\theta}}$ ersetzt wurde. Dabei kann m als Realisation von M_n aufgefasst werden. Dies folgt ebenso für die Score-Funktionen, sodass die Schätzer aus den beiden Modellansätzen in Abbildung 4.1 übereinstimmen.

Ein ähnliches Resultat befindet sich im (fortgesetzten) Beispiel 7.1(a) in Daley und Vere-Jones [8]. Die Aussage gilt jedoch nur aufgrund der Approximation des trunkierten Prozesses durch einen Poisson-Prozess in Abschnitt 3.2.
In dem genannten Beispiel wird, ausgedrückt in der dort verwendeten Notation, ein Punktprozess N in $A \subset \mathbb{R}^d$ mit dem Intensitätsmaß Λ_N betrachtet. Das Intensitätsmaß besitze bezüglich des Lebesgue-Maßes λ die Dichte λ_N. Somit ist die Anzahl der Punkte $N(A)$, auch kurz N geschrieben, Poisson-verteilt mit dem Parameter $\Lambda(A)$. Es wird nun die logarithmierte Dichte des Poisson-Prozesses N bezüglich eines homogenen Poisson-Prozess mit dem Parameter 1 betrachtet. Diese lautet

$$\log(L_A/L_A^{\#}) = \sum_{i=1}^{N} \log\lambda_N(x_i) - \int_A [\lambda_N(x) - 1]\,\mathrm{d}x.$$

Unter der Annahme $\lambda_N(x) = C \cdot \phi(x)$ mit $C > 0$ und $\int_A \phi(x)\,\mathrm{d}x = 1$ folgt

$$\log(L_A/L_A^{\#}) = N \cdot \log C + \sum_{i=1}^{N} \log\phi(x_i) - C + \lambda(A).$$

Leitet man die logarithmierte Dichte nun nach C ab, so erhält man als Nullstelle der Ableitung den Maximum-Likelihood-Schätzer $\hat{C} = N$. Ersetzt man diesen Wert in der logarithmierten Likelihood, so folgt

$$\log(\hat{L}_A/L_A^{\#}) = N \cdot \log N - N + \lambda(A) + \sum_{i=1}^{N} \log \phi(x_i).$$

Beim vorherigen Bedingen auf N, reduziert sich die Likelihood auf das Produkt der Dichten $\phi(x_i)$ von N unabhängigen Beobachtungen. Die bedingte Likelihood sowie die unbedingte, aber substituierte Likelihood $\log(\hat{L}_A/L_A^{\#})$ stimmen bis auf einen konstanten Term überein und liefern übereinstimmende Schätzer. Nach Daley und Vere-Jones trifft diese Aussage nur für Poisson-Prozesse zu.
In Abbildung 4.1 ergibt sich die bedingte Likelihood aus den bedingten Dichten $f_{\boldsymbol{\theta}}^*$, den Pfeilen gegen dem Uhrzeigersinn folgend. Damit man nun den Pfeilen dem Uhrzeigersinn nach zum gleichen Schätzer gelangt, ist die Approximation des trunkierten Prozesses $N_{n,D}$ durch den Poisson-Prozess N_n^* in Abschnitt 3.2 notwendig. Da das Lebesgue-Maß der Menge S nicht endlich ist, wurde als Intensitätsmaß des zweiten Poisson-Prozesses N_0 das eingeschränkte Lebesgue-Maß λ_R gewählt. In der Likelihood ergeben sich dadurch andere Konstanten als bei Daley und Vere-Jones, welche auf den Schätzer keinen Einfluss haben. Die Produktgestalt der Dichte des Intensitätsmaßes ist durch $n \cdot f_{\boldsymbol{\theta}} \mathbb{1}_D$ gegeben. Wegen $\int_D f_{\boldsymbol{\theta}}(x,t)\,\mathrm{d}(x,t) = \alpha_{\boldsymbol{\theta}}$ ergeben sich auch hier Unterschiede zu $\log(L_A/L_A^{\#})$. Dies führt dazu, dass n nicht nur durch τ bzw., aufgrund einer weiteren Approximation am Ende des Abschnitts 3.2, durch M_n, sondern durch $M_n/\alpha_{\boldsymbol{\theta}}$ geschätzt wird. Wie bereits am Anfang des Abschnitts veranschaulicht wurde, erhält man, nach dem Ersetzen von n, übereinstimmende Schätzer.

A.2. Kovarianzen der Grenzprozesse im Kolmogorov-Smirnov-Test

Berechnung der Kovarianz bei $(\hat{\theta}_n, n)$

Ziel dieses Abschnitts ist die Herleitung der Kovarianz (6.13) aus Bemerkung 6.24. Zur Wahrung der Übersicht wird die verkürzte Notation $g_i := g_{x_i,t_i,D} = \mathbb{1}_{[0,x_i]\times[0,t_i]\cap D} \in \mathcal{G}_D$ für $i = 1, 2$ verwendet.
Zunächst lässt sich $\mathrm{Cov}[\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_1), \mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_2)]$ mithilfe der Definition des Prozesses $\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}$ in Gleichung (6.12) umschreiben zu

$$\begin{aligned}
&\mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1 - \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0} \dot{\mathbb{E}}_{\theta_0} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} g_2 - \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0} \dot{\mathbb{E}}_{\theta_0} g_2] \\
&= \mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} g_2] - \dot{\mathbb{E}}_{\theta_0} g_2 \cdot \mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}] - \dot{\mathbb{E}}_{\theta_0} g_1 \cdot \mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_2, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}] \\
&\quad + \dot{\mathbb{E}}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \cdot \mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}].
\end{aligned}$$

Zur weiteren Umformung kann die Kovarianz der $\mathbb{P}_{\theta_0}$-Brownschen Brücke aus Definition 6.4 genutzt werden. Es gilt $\mathrm{Cov}[\mathbb{B}_{\mathbb{P}_{\theta_0}} f, \mathbb{B}_{\mathbb{P}_{\theta_0}} g] = \mathbb{E}_{\theta_0}[fg] - \mathbb{E}_{\theta_0}[f]\mathbb{E}_{\theta_0}[g]$ für Funktionen

$f, g \in \mathcal{G}_D \cup \{\phi_{\theta_0}\}$. Zusammen mit $\mathbb{E}_{\theta_0}[\phi_{\theta_0}] = 0$ (vgl. Gleichung (6.11)) kann die Kovarianz $\mathrm{Cov}[\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_1), \mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_2)]$ in die folgende Form gebracht werden:

$$\begin{aligned}
&\mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 - \dot{\mathbb{E}}_{\theta_0} g_2 \cdot [\mathbb{E}_{\theta_0} g_1 \phi_{\theta_0} - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} \phi_{\theta_0}] \\
&\quad - \dot{\mathbb{E}}_{\theta_0} g_1 \cdot [\mathbb{E}_{\theta_0} g_2 \phi_{\theta_0} - \mathbb{E}_{\theta_0} g_2 \mathbb{E}_{\theta_0} \phi_{\theta_0}] + \dot{\mathbb{E}}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \cdot [\mathbb{E}_{\theta_0} \phi_{\theta_0}^2 - \mathbb{E}_{\theta_0} \phi_{\theta_0} \mathbb{E}_{\theta_0} \phi_{\theta_0}] \\
&= \mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 - \dot{\mathbb{E}}_{\theta_0} g_2 \cdot \mathbb{E}_{\theta_0} g_1 \phi_{\theta_0} - \dot{\mathbb{E}}_{\theta_0} g_1 \cdot \mathbb{E}_{\theta_0} g_2 \phi_{\theta_0} + \dot{\mathbb{E}}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \cdot \mathbb{E}_{\theta_0} \phi_{\theta_0}^2
\end{aligned}$$

Nun gilt es, die Erwartungswerte $\mathbb{E}_{\theta_0} g_i \phi_{\theta_0}$ und $\mathbb{E}_{\theta_0} \phi_{\theta_0}^2$ genauer zu bestimmen. Den nachfolgenden Umformungsschritten liegen die Definition $\phi_{\theta_0} := -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \psi_{\theta_0}$ aus Satz 6.23 sowie der auf dem Beobachtungsparallelogramm D geltende Zusammenhang $\psi_{\theta_0} = \dot{f}_{\theta_0}/f_{\theta_0} - \dot{\alpha}_{\theta_0}/\alpha_{\theta_0}$ zu Grunde. Letzterer wurde in Gleichung (A.1) unter der Annahme (A3), d.h. für die Gumbel-Barnett-Copula, gezeigt.

$$\begin{aligned}
\mathbb{E}_{\theta_0} g_i \phi_{\theta_0} &= -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \mathbb{E}_{\theta_0} g_i \psi_{\theta_0} \\
&= -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \cdot \int_D \mathbb{1}_{[0,x_i]\times[0,t_i]}(x',t') \cdot \psi_{\theta_0}(x',t') \cdot f_{\theta_0}(x',t') \,\mathrm{d}(x',t') \\
&= -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \cdot \int_D \mathbb{1}_{[0,x_i]\times[0,t_i]}(x',t') \cdot \left(\dot{f}_{\theta_0}(x',t') - \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} f_{\theta_0}(x',t') \right) \mathrm{d}(x',t') \\
&= -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \left[\dot{\mathbb{E}}_{\theta_0} g_i - \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \mathbb{E}_{\theta_0} g_i \right] \qquad \text{(A.2)}
\end{aligned}$$

Für die letzte Gleichheit wird die Notation der Fréchet-Ableitung aus Lemma 6.22 genutzt. Für die Berechnung von $\mathbb{E}_{\theta_0} \phi_{\theta_0}^2$ kann die in Satz 4.4 gezeigte Information Matrix Equality der Fisher-Information genutzt werden:

$$\begin{aligned}
\mathbb{E}_{\theta_0} \phi_{\theta_0}^2 &= (-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1})^2 \cdot \mathbb{E}_{\theta_0} \psi_{\theta_0}^2 \\
&= (-(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1})^2 \cdot (-\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}]) = -(\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \qquad \text{(A.3)}
\end{aligned}$$

Unter Verwendung der umformulierten Erwartungswerte kann die Kovarianz $\mathrm{Cov}[\mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_1), \mathbb{G}^{\phi}_{\mathbb{P}_{\theta_0}}(g_2)]$ abschließend auf die Darstellung (6.13) gebracht werden:

$$\begin{aligned}
&\mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 + \dot{\mathbb{E}}_{\theta_0} g_2 \cdot (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \left[\dot{\mathbb{E}}_{\theta_0} g_1 - \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \mathbb{E}_{\theta_0} g_1 \right] \\
&\quad + \dot{\mathbb{E}}_{\theta_0} g_1 \cdot (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \left[\dot{\mathbb{E}}_{\theta_0} g_2 - \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \mathbb{E}_{\theta_0} g_2 \right] - \dot{\mathbb{E}}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \cdot (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \\
&= \mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 + \dot{\mathbb{E}}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \, (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \\
&\quad - (\mathbb{E}_{\theta_0}[\dot{\psi}_{\theta_0}])^{-1} \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \cdot \left[\mathbb{E}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 + \mathbb{E}_{\theta_0} g_2 \dot{\mathbb{E}}_{\theta_0} g_1 \right]
\end{aligned}$$

Berechnung der Kovarianz bei $(\theta_n, \hat{n})$

In diesem Abschnitt soll die Kovarianz (6.18) errechnet werden. Zur Vereinfachung wird die Kurzschreibweise $g_i := g_{x_i,t_i,D} = \mathbb{1}_{[0,x_i]\times[0,t_i]\cap D} \in \mathcal{G}_D$ für $i = 1, 2$ verwendet. In die Berechnung von $\mathrm{Cov}[\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}} g_2]$ geht zunächst die Definition des Prozesses $\mathbb{G}^{\mathbb{1}}_{\mathbb{P}_{\theta_0}}$ aus Satz 6.27 ein:

$$
\begin{aligned}
&\mathrm{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1 - \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D \cdot \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}}, \mathbb{B}_{\mathbb{P}_{\theta_0}} g_2 - \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D \cdot \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}}\right] \\
&= \mathrm{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} g_2\right] - \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}} \cdot \mathrm{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] \\
&\quad - \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}} \cdot \mathrm{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_2, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] + \frac{\mathbb{E}_{\theta_0} g_1 \cdot \mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}^2} \cdot \mathrm{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] \\
&= \mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 - \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}} \cdot [\mathbb{E}_{\theta_0} g_1 \mathbb{1}_D - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} \mathbb{1}_D] \\
&\quad - \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}} \cdot [\mathbb{E}_{\theta_0} g_2 \mathbb{1}_D - \mathbb{E}_{\theta_0} g_2 \mathbb{E}_{\theta_0} \mathbb{1}_D] + \frac{\mathbb{E}_{\theta_0} g_1 \cdot \mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}^2} \cdot \left[\mathbb{E}_{\theta_0} \mathbb{1}_D^2 - \mathbb{E}_{\theta_0} \mathbb{1}_D \mathbb{E}_{\theta_0} \mathbb{1}_D\right]
\end{aligned}
$$

Im zweiten Schritt wurde die Kovarianz der $\mathbb{P}_{\theta_0}$-Brownschen Brücke aus Definition 6.4 verwendet. Abschließend können die Erwartungswerte $\mathbb{E}_{\theta_0} g_i \mathbb{1}_D$ und $\mathbb{E}_{\theta_0} \mathbb{1}_D^2$ vereinfacht werden. Für Funktionen $g_i \in \mathcal{G}_{\mathcal{D}}$, wobei auch $\mathbb{1}_D \in \mathcal{G}_D$, gilt, dass $\mathbb{E}_{\theta_0}[\mathbb{1}_D g_i] = \mathbb{E}_{\theta_0}[g_i]$ und $\mathbb{E}_{\theta_0}[\mathbb{1}_D] = \alpha_{\theta_0}$. Einsetzen und Zusammenfassen liefert

$$
\begin{aligned}
&\mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 - \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}} \cdot \mathbb{E}_{\theta_0} g_1 \left[1 - \alpha_{\theta_0}\right] - \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}} \cdot \mathbb{E}_{\theta_0} g_2 \left[1 - \alpha_{\theta_0}\right] \\
&\quad + \frac{\mathbb{E}_{\theta_0} g_1 \cdot \mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}^2} \cdot \alpha_{\theta_0}[1 - \alpha_{\theta_0}] \\
&= \mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 \cdot \frac{1 - \alpha_{\theta_0}}{\alpha_{\theta_0}} \\
&= \mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 \cdot \frac{1}{\alpha_{\theta_0}}.
\end{aligned}
$$

Berechnung der Kovarianz bei $(\hat{\theta}_n, \hat{n})$

Als drittes folgt die Berechnung der Kovarianz aus Gleichung 6.23. Wie bereits in den vorhergehenden Berechnungen, wird auch hier $g_i := g_{x_i,t_i,D} = \mathbb{1}_{[0,x_i]\times[0,t_i]\cap D} \in \mathcal{G}_D$ für $i = 1, 2$ gesetzt. Im ersten Schritt wird in $\mathrm{Cov}[\mathbb{H}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{H}_{\mathbb{P}_{\theta_0}} g_2]$ die Definition des Prozesses $\mathbb{H}_{\mathbb{P}_{\theta_0}}$ aus Satz 6.32, d.h.

$$
\begin{aligned}
\mathbb{H}_{\mathbb{P}_{\theta_0}} g_i &= \mathbb{B}_{\mathbb{P}_{\theta_0}} g_i - \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0} \dot{\mathbb{E}}_{\theta_0} g_i + \left(\mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0} \cdot \dot{\alpha}_{\theta_0} - \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right) \cdot \frac{\mathbb{E}_{\theta_0} g_i}{\alpha_{\theta_0}} \\
&= \mathbb{B}_{\mathbb{P}_{\theta_0}} g_i + \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0} \left(\mathbb{E}_{\theta_0} g_i \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} - \dot{\mathbb{E}}_{\theta_0} g_i\right) - \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D \frac{\mathbb{E}_{\theta_0} g_i}{\alpha_{\theta_0}}
\end{aligned}
$$

eingesetzt. Für die bessere Lesbarkeit sei dafür $K_i := \left(\mathbb{E}_{\theta_0} g_i \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} - \dot{\mathbb{E}}_{\theta_0} g_i\right)$, für $i = 1, 2$. Damit gilt:

$$
\begin{aligned}
&\operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1 + \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0} K_1 - \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}}, \mathbb{B}_{\mathbb{P}_{\theta_0}} g_2 + \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0} K_2 - \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}}\right] \\
&= \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} g_2\right] + K_2 \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}\right] - \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}} \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] \\
&\quad + K_1 \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_2, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}\right] + K_1 K_2 \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}\right] \\
&\quad - \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}} K_1 \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] - \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}} \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_2, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] \\
&\quad - \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}} K_2 \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] + \frac{\mathbb{E}_{\theta_0} g_1 \cdot \mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}^2} \cdot \operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right]
\end{aligned}
$$

Die darin enthaltenen Kovarianzen können mithilfe der Kovarianz der $\mathbb{P}_{\theta_0}$-Brownschen Brücke aus Definition 6.4 wie folgt für $i = 1, 2$ vereinfacht werden:

$$
\begin{aligned}
\operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_1, \mathbb{B}_{\mathbb{P}_{\theta_0}} g_2\right] &= \mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 \\
\operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_i, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}\right] &= \mathbb{E}_{\theta_0} g_i \phi_{\theta_0} - \mathbb{E}_{\theta_0} g_i \mathbb{E}_{\theta_0} \phi_{\theta_0} &&= \mathbb{E}_{\theta_0} g_i \phi_{\theta_0} \\
\operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} g_i, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] &= \mathbb{E}_{\theta_0} g_i \mathbb{1}_D - \mathbb{E}_{\theta_0} g_i \mathbb{E}_{\theta_0} \mathbb{1}_D &&= (1 - \alpha_\theta)\, \mathbb{E}_{\theta_0} g_i \\
\operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}, \mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}\right] &= \mathbb{E}_{\theta_0} \phi_{\theta_0}^2 - \mathbb{E}_{\theta_0} \phi_{\theta_0} \mathbb{E}_{\theta_0} \phi_{\theta_0} &&= \mathbb{E}_{\theta_0} \phi_{\theta_0}^2 \\
\operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \phi_{\theta_0}, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] &= \mathbb{E}_{\theta_0} \phi_{\theta_0} \mathbb{1}_D - \mathbb{E}_{\theta_0} \phi_{\theta_0} \mathbb{E}_{\theta_0} \mathbb{1}_D &&= \mathbb{E}_{\theta_0} \phi_{\theta_0} \mathbb{1}_D = 0 \\
\operatorname{Cov}\left[\mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D, \mathbb{B}_{\mathbb{P}_{\theta_0}} \mathbb{1}_D\right] &= \mathbb{E}_{\theta_0} \mathbb{1}_D^2 - \mathbb{E}_{\theta_0} \mathbb{1}_D \mathbb{E}_{\theta_0} \mathbb{1}_D &&= \alpha_{\theta_0}(1 - \alpha_{\theta_0})
\end{aligned}
$$

In die Vereinfachungen gingen weiterhin $\mathbb{E}_{\theta_0} \phi_{\theta_0} = 0$ (vgl. (6.11)) und $\mathbb{E}_{\theta_0} \mathbb{1}_D = \alpha_{\theta_0}$ ein. Das Einsetzen der Werte der Kovarianzen führt zu:

$$
\begin{aligned}
&\mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 + K_2 \cdot \mathbb{E}_{\theta_0} g_1 \phi_{\theta_0} - \frac{\mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}} \cdot (1 - \alpha_\theta)\, \mathbb{E}_{\theta_0} g_1 + K_1 \cdot \mathbb{E}_{\theta_0} g_2 \phi_{\theta_0} \\
&\quad + K_1 K_2 \cdot \mathbb{E}_{\theta_0} \phi_{\theta_0}^2 - \frac{\mathbb{E}_{\theta_0} g_1}{\alpha_{\theta_0}} \cdot (1 - \alpha_\theta)\, \mathbb{E}_{\theta_0} g_2 + \frac{\mathbb{E}_{\theta_0} g_1 \cdot \mathbb{E}_{\theta_0} g_2}{\alpha_{\theta_0}^2} \cdot \alpha_{\theta_0}(1 - \alpha_{\theta_0}) \\
&= \mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 \cdot \frac{1}{\alpha_{\theta_0}} + K_2 \cdot \mathbb{E}_{\theta_0} g_1 \phi_{\theta_0} + K_1 \cdot \mathbb{E}_{\theta_0} g_2 \phi_{\theta_0} + K_1 K_2 \cdot \mathbb{E}_{\theta_0} \phi_{\theta_0}^2
\end{aligned}
$$

Wie bereits in den Gleichungen (A.3) und (A.2) gezeigt wurde, gelten $\mathbb{E}_{\theta_0} \phi_{\theta_0}^2 = -\mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1}$ sowie

$$
\mathbb{E}_{\theta_0} g_i \phi_{\theta_0} = \left(-\mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1}\right) \left[\dot{\mathbb{E}}_{\theta_0} g_i - \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \mathbb{E}_{\theta_0} g_i\right] = \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \cdot K_i.
$$

Das Ersetzen von $\mathbb{E}_{\theta_0}\phi_{\theta_0}^2$, $\mathbb{E}_{\theta_0}g_i\phi_{\theta_0}$ und K_i ergibt

$$\begin{aligned}
&K_2 \cdot \mathbb{E}_{\theta_0} g_1 \phi_{\theta_0} + K_1 \cdot \mathbb{E}_{\theta_0} g_2 \phi_{\theta_0} + K_1 K_2 \cdot \mathbb{E}_{\theta_0} \phi_{\theta_0}^2 \\
&= K_2 \cdot \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \cdot K_1 + K_1 \cdot \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \cdot K_2 + K_1 K_2 \cdot (-\mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1}) \\
&= \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \cdot \left(\mathbb{E}_{\theta_0} g_1 \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} - \dot{\mathbb{E}}_{\theta_0} g_1 \right) \left(\mathbb{E}_{\theta_0} g_2 \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} - \dot{\mathbb{E}}_{\theta_0} g_2 \right) \\
&= \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \cdot \left[\frac{\dot{\alpha}_{\theta_0}^2}{\alpha_{\theta_0}^2} \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 - \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \mathbb{E}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 - \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \dot{\mathbb{E}}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 + \dot{\mathbb{E}}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \right].
\end{aligned}$$

Addiert man dazu wieder $\mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 / \alpha_{\theta_0}$ ergibt sich schließlich die Darstellung

$$\begin{aligned}
&\mathbb{E}_{\theta_0} g_1 g_2 - \mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 \cdot \frac{1}{\alpha_{\theta_0}} + \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \cdot \dot{\mathbb{E}}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \\
&\quad - \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \frac{\dot{\alpha}_{\theta_0}}{\alpha_{\theta_0}} \left[\dot{\mathbb{E}}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 + \mathbb{E}_{\theta_0} g_1 \dot{\mathbb{E}}_{\theta_0} g_2 \right] + \mathbb{E}_{\theta_0}(\dot{\psi}_{\theta_0})^{-1} \frac{\dot{\alpha}_{\theta_0}^2}{\alpha_{\theta_0}^2} \left[\mathbb{E}_{\theta_0} g_1 \mathbb{E}_{\theta_0} g_2 \right].
\end{aligned}$$

A.3. Standardfehler im Farlie-Gumbel-Morgenstern-Modell

Auch im Farlie-Gumbel-Morgenstern-Modell wird zur Berechnung der Standardfehler die Score-Funktion benötigt. Es gelten dafür im Folgenden die Annahmen (A1)-(A4), wobei die Annahme (A3) durch (A3') ersetzt wird. Es bezeichne $M_n^{FGM} = N_{n,D}^{FGM}(S)$. Die Loglikelihood lautet

$$\begin{aligned}
\log \ell^{FGM}(\boldsymbol{\theta}, n) &= \sum_{i=1}^{n} \mathbb{1}_{\{(X_i,T_i)^{\mathrm{T}} \in D\}} \log n f_{\boldsymbol{\theta}}^{FGM}(X_i, T_i) + (G+s)G - n\alpha_{\boldsymbol{\theta}}^{FGM} \\
&= \sum_{j=1}^{M_n^{FGM}} \log n + \log \frac{\theta}{G} - \theta \widetilde{X}_j + \log \left[1 + \vartheta \left(2e^{-\theta \widetilde{X}_j} - 1 \right) \left(1 - \frac{2\widetilde{T}_j}{G} \right) \right] \\
&\quad + (G+s)G - n\alpha_{\boldsymbol{\theta}}^{FGM}.
\end{aligned}$$

Zum Aufstellen der Score-Funktion werden die partiellen Ableitungen nach θ und ϑ benötigt. Diese lauten

$$\begin{aligned}
\frac{\partial \log \ell^{FGM}(\boldsymbol{\theta}, n)}{\partial \theta} = \sum_{j=1}^{M_n^{FGM}} &\frac{1}{\theta} - \widetilde{X}_j + \left[1 + \vartheta \left(2e^{-\theta \widetilde{X}_j} - 1 \right) \left(1 - \frac{2\widetilde{T}_j}{G} \right) \right]^{-1} \cdot \\
&\cdot \left(-\vartheta \widetilde{X}_j \right) \left(1 - \frac{2\widetilde{T}_j}{G} \right) 2e^{-\theta \widetilde{X}_j} - n \cdot \frac{\partial \alpha_{\boldsymbol{\theta}}^{FGM}}{\partial \theta}
\end{aligned}$$

und

$$\frac{\partial \log \ell^{FGM}(\boldsymbol{\theta}, n)}{\partial \vartheta} = \sum_{j=1}^{M_n^{FGM}} \left[1 + \vartheta \left(2e^{-\theta \widetilde{X}_j} - 1\right)\left(1 - \frac{2\widetilde{T}_j}{G}\right)\right]^{-1} \cdot \\ \cdot \left(2e^{-\theta \widetilde{X}_j} - 1\right)\left(1 - \frac{2\widetilde{T}_j}{G}\right) - n \cdot \frac{\partial \alpha_{\boldsymbol{\theta}}^{FGM}}{\partial \vartheta}.$$

Da n nicht sichtbar ist, soll auch hier wie im Abschnitt 3.3 als erster Teilschritt die Loglikelihood in n maximiert werden. Interpretiert man n als stetige Variable, so ist auch hier $n = M_n^{FGM}/\alpha_{\boldsymbol{\theta}}^{FGM}$ die Nullstelle der Ableitung der Loglikelihood. Dies führt zu der zweidimensionalen Score-Funktion $\psi_{\boldsymbol{\theta}}^{FGM} = (\psi_{\boldsymbol{\theta},1}^{FGM}, \psi_{\boldsymbol{\theta},2}^{FGM})$ mit

$$\psi_{\boldsymbol{\theta},1}^{FGM}(X_i, T_i) = \mathbb{1}_{[T_i, T_i+s]}(X_i)\left[\frac{1}{\theta} - X_i - \frac{2\vartheta X_i e^{-\theta X_i}\left(1 - \frac{2T_i}{G}\right)}{1 + \vartheta\left(2e^{-\theta X_i} - 1\right)\left(1 - \frac{2T_i}{G}\right)} - \frac{\partial \alpha_{\boldsymbol{\theta}}^{FGM}}{\partial \theta}\frac{1}{\alpha_{\boldsymbol{\theta}}^{FGM}}\right]$$

sowie

$$\psi_{\boldsymbol{\theta},2}^{FGM}(X_i, T_i) = \mathbb{1}_{[T_i, T_i+s]}(X_i)\left[\frac{\left(2e^{-\theta X_i} - 1\right)\left(1 - \frac{2T_i}{G}\right)}{1 + \vartheta\left(2e^{-\theta X_i} - 1\right)\left(1 - \frac{2T_i}{G}\right)} - \frac{\partial \alpha_{\boldsymbol{\theta}}^{FGM}}{\partial \vartheta}\frac{1}{\alpha_{\boldsymbol{\theta}}^{FGM}}\right].$$

Die partiellen Ableitungen der Auswahlwahrscheinlichkeit werden nicht explizit angegeben. Als Kriteriumsfunktion $\Psi_n^{FGM} = (\Psi_{n,1}^{FGM}, \Psi_{n,2}^{FGM})$ ergibt sich

$$\Psi_{n,1}^{FGM}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \psi_{\boldsymbol{\theta},1}^{FGM}(X_i, T_i) \quad \text{und} \quad \Psi_{n,2}^{FGM}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \psi_{\boldsymbol{\theta},2}^{FGM}(X_i, T_i).$$

Der Schätzer berechnet sich als Nullstelle der Kriteriumsfunktion. Auf den Nachweis der Eigenschaften der Konsistenz und asymptotischen Normalität soll an dieser Stelle verzichtet werden. Unter Annahme der asymptotischen Normalität für $\theta_0^{FGM} \in (0, 1 - \varepsilon_\theta)$ und $\vartheta_0^{FGM} \in (-1, 1)$ besitzt die Folge $\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{FGM} - \boldsymbol{\theta}_0^{FGM})$ den Erwartungswertvektor $\mathbf{0}$ und die Kovarianzmatrix $\mathcal{I}(\boldsymbol{\theta}_0^{FGM})^{-1}$. Unter Gültigkeit der Information Matrix Equality (vgl. Abschnitt 4.2) berechnet sich die Fisher-Information als negativer Erwartungswert der Ableitung der Score-Funktion. Die Standardfehler lassen sich somit schätzen, in dem man die geschätzten Parameter in die Fisher-Information einsetzt.

A.4. Verteilungen der Beobachtungen bei unabhängiger Doppeltrunkation

Der Kolmogorov-Smirnov-Anpassungstest wird in Abschnitt 6.2 für den Fall der stochastischen Unabhängigkeit der latenten Variablen X_i und T_i, $i = 1, \ldots, n$, betrachtet. Falls $(X_i, T_i)^{\mathrm{T}} \in D$, so wird der Vektor beobachtet. Die Beobachtungen werden umnummeriert und mit $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$, $j = 1, \ldots, M_n$, bezeichnet. Mithilfe der gemeinsamen sowie der Randdichten von $(\widetilde{X}_j, \widetilde{T}_j)^{\mathrm{T}}$ lässt sich zeigen, dass $\widetilde{X}_j$ und $\widetilde{T}_j$ nicht mehr stochas-

tisch unabhängig sind. Die gemeinsame Verteilung der Beobachtungen ergibt sich als gemeinsame Verteilung der latenten Variablen bedingt darauf, beobachtet zu werden. Für α_{θ_0} aus (6.2) lautet die gemeinsame Dichtefunktion

$$f^{\widetilde{X},\widetilde{T}}(x,t) = \mathbb{1}_D(x,t)\frac{\theta_0 e^{-\theta_0 x}}{G\alpha_{\theta_0}}.$$

Die Dichte von $\widetilde{X}_j$ ergibt sich für $x \in [0, G+s]$ durch

$$f^{\widetilde{X}}(x) = \int_0^G \mathbb{1}_D(x,t)\frac{\theta_0 e^{-\theta_0 x}}{G\alpha_{\theta_0}}\,\mathrm{d}t.$$

Die Indikatorfunktion lässt sich durch Fallunterscheidungen über die Integralgrenzen berücksichtigen. Wie bereits in Korollar 2 bei Weißbach und Wied [65] zu lesen ist, erhält man unter der Annahme $s \leq G$, dass

$$f^{\widetilde{X}}(x) = \frac{\theta_0}{G\alpha_{\theta_0}} e^{-\theta_0 x}\Big(\mathbb{1}_{[0,s]}(x)x + \mathbb{1}_{(s,G]}(x)s + \mathbb{1}_{(G,G+s]}(x)(G-x+s)\Big)$$

Die Dichte von $\widetilde{T}_j$ berechnet sich für $t \in (0, G]$ ähnlich über

$$f^{\widetilde{T}}(t) = \int_0^\infty \mathbb{1}_D(x,t)\frac{\theta_0 e^{-\theta_0 x}}{G\alpha_{\theta_0}}\,\mathrm{d}x = \int_t^{t+s} \frac{\theta_0 e^{-\theta_0 x}}{G\alpha_{\theta_0}}\,\mathrm{d}x = \frac{e^{-\theta_0 t}}{G\alpha_{\theta_0}}(1 - e^{-\theta_0 s})$$

Die gemeinsame Dichte ergibt sich nicht als Produkt der Randdichten, da die Beobachtungen über die Menge D voneinander abhängen.
Für die in Gleichung (7.1) durchzuführende Transformation der Daten, werden nun die Verteilung von $\widetilde{X}_j$ und die auf das Ereignis $\widetilde{X}_j = x$ bedingte Verteilung von $\widetilde{T}_j$ benötigt. Ersteres ergibt sich direkt durch Integration über die Dichte $f^{\widetilde{X}}$:

$$\begin{aligned} F^{\widetilde{X}}(x) =& \frac{1}{G\alpha_{\theta_0}\theta_0}\Big(\mathbb{1}_{[0,s]}(x)\Big[1 - e^{-\theta_0 x}(\theta_0 x + 1)\Big] + \mathbb{1}_{(s,G]}(x)\Big[1 - e^{-\theta_0 s} - s\theta_0 e^{-\theta_0 x}\Big] \\ &+ \mathbb{1}_{(G,G+s]}(x)\Big[1 - e^{-\theta_0 s} - e^{-\theta_0 G} + e^{-\theta_0 x}(1 - \theta_0[G - x + s])\Big]\Big) \end{aligned}$$

Die auf $\widetilde{X}_j = x$ bedingte Verteilung von $\widetilde{T}_j$ ergibt sich für $t \in (0, G]$ und $x \in [0, G+s]$ durch

$$F^{\widetilde{T}}(t|x) = \int_0^t \frac{f^{\widetilde{X},\widetilde{T}}(x,t')}{f^{\widetilde{X}}(x)}\,\mathrm{d}t'.$$

Die gleiche Fallunterscheidung wie oben und das Berücksichtigen der Restriktion der Dichte auf die Menge D führt zu

$$\begin{aligned} F^{\widetilde{T}}(t|x) =& \mathbb{1}_{[0,s]}(x)\frac{\min(t,x)}{x} + \mathbb{1}_{(s,G]}(x)\left(\mathbb{1}_{[x-s,x]}(t)\frac{t-x+s}{s} + \mathbb{1}_{(x,G]}(t)\right) \\ &+ \mathbb{1}_{(G,G+s]}(x)\frac{\max(t, x-s) - x + s}{G - x + s}. \end{aligned}$$

Eidesstattliche Versicherung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.
Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde zur Erlangung eines akademischen Grades vorgelegt.

Rostock, den 15. Mai 2025

Anne-Marie Toparkus