

Stochastische Modelle mit zeitlich gekoppelten Ereignissen: Rekonstruktion der Zuordnung und Schätzen der Verzögerungsverteilung

Von der Fakultät für Mathematik und Physik
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften

Dr. rer. nat.

genehmigte Dissertation

von

Dipl.-Math. Hendrik Wegener

geboren am 18.10.1979 in Lübbecke

2010

Referent: Prof. Dr. R. Grübel
Korreferentin: Prof. Dr. N. Bäuerle

Tag der Promotion: 1. Juli 2010

Zusammenfassung

Ein beispielhafter Ausgangspunkt für die Untersuchungen in dieser Arbeit ist die folgende Situation. Es liegen die jeweils in chronologischer Reihenfolge geordneten Geburts- und Todesdaten von n Komponisten vor, allerdings besteht keine Kenntnis darüber, welche Daten zusammengehören. Dieser Informationsverlust tritt z.B. auf, wenn nur die Anzahl der lebenden Komponisten als Funktion der Zeit beobachtbar ist, und erscheint generell im Zusammenhang mit Warteschlangen vom Typ $M/G/\infty$. Im ersten Fall entspricht die Verzögerung der zeitlich gekoppelten Ereignisse „Geburt“ und „Tod“ der Lebensdauer; bei $M/G/\infty$ geht es um Ankünfte und Abgänge, und die Verzögerung entspricht der Bedienzeit.

Nach Einführung der grundlegenden Begriffe und Konzepte sowie Formalisierung der beschriebenen Ausgangssituation durch ein stochastisches Modell in Kapitel 1 gliedert sich die Betrachtung dieses Modells in zwei Teile. In Kapitel 2 nehmen wir die Verzögerungsverteilung als bekannt an und es ist die Zuordnung zwischen den beobachteten Daten zu rekonstruieren. Dazu behandeln wir zwei Verfahren, die Maximum a posteriori Methode und Minimierung des mittleren Abstands. Für das zuerst genannte Verfahren wird ein Zusammenhang zu dem im Operations Research untersuchten linearen Assignment-Problem hergestellt. Weiter bestimmen wir in beiden Fällen eine explizite Lösung, wenn die Dichte der Verzögerungsverteilung logkonkav ist. Die Untersuchung der asymptotischen Fehlerverteilung bei Verwendung der durch eines der Verfahren ermittelten Zuordnung im Spezialfall exponentialverteilter Verzögerungen sowie ein Test auf Exponentialverteilung in einer bestimmten Situation bilden den Abschluss des Kapitels.

In Kapitel 3 ist die hier als unbekannt angenommene Verzögerungsverteilung zu bestimmen, dabei verwenden wir nur parametrische Verfahren. Während ein Maximum-Likelihood-Ansatz zu keiner praktikablen Lösung führt, können wir asymptotisch normale Momentenschätzer für die Verteilungsparameter angeben, falls die Geburts- bzw. Ankunftszeitpunkte unabhängig und identisch verteilt sind. Stammen letztere dagegen aus einem Poisson-Prozess, so sind die resultierenden Schätzer nicht konsistent im quadratischen Mittel. Wir verwenden deswegen den Zusammenhang zu den sog. „busy cycles“ eines $M/G/\infty$ -Warteschlangensystems und bestimmen weitere, asymptotisch normale und konsistente Momentenschätzer. Abschließend wird die Korrelationsstruktur zwischen der Anzahl der Kunden im System zu verschiedenen Zeitpunkten ausgenutzt, um eine dritte Klasse von Momentenschätzer zu betrachten und Konsistenz für diese zu zeigen. Im Verlauf der gesamten Arbeit illustrieren wir unsere Resultate wiederholt mit den historischen Daten von 22 Komponisten.

Schlagwörter: Maximum a posteriori Methode, Momentenschätzer, $M/G/\infty$ -Warteschlangensystem

Abstract

As a motivation for the research conducted in this dissertation, consider the following situation. We are given two chronologically ordered lists, containing the dates of birth and death of n composers, but there is no information about the original link between individual dates. This loss of information occurs if, for instance, we only observe the number of living composers as a function of time, and typically arises in the context of $M/G/\infty$ queues. In the first case, the delay between the two coupled events 'birth' and 'death' corresponds to the lifetime, while, in the $M/G/\infty$ setting, arrivals and departures are studied, and the service time parallels the delay.

After the introduction, in Chapter 1, of basic notation and concepts, as well as formalization of the situation described above using a stochastic model, analysis of this model is organized in two parts. In Chapter 2, we assume the delay distribution to be known, and we try to reconstruct the original matching between the observed data. Two approaches are considered, the maximum a posteriori method and minimization of the mean distance. For the former, we establish a link to the so called linear assignment problem studied in operations research. Furthermore, in both cases, we obtain an explicit solution if the density of the delay distribution is log-concave. We close this chapter with an examination of the asymptotic error distribution as $n \rightarrow \infty$ for exponentially distributed delays if the matching is determined by one of the above methods, and a test for exponential distribution in a certain situation.

In Chapter 3, we consider estimation of the delay distribution, which we now assume to be unknown. While a maximum likelihood approach yields no viable result, we develop asymptotically normal moment type estimators for the distribution parameters if dates of birth and times of arrival, respectively, are independent and identically distributed. If, in contrast, the latter stem from a Poisson process, the corresponding estimators are not consistent in the quadratic mean. Therefore, we adopt the connection to the busy cycles of a $M/G/\infty$ queue and determine other moment type estimators that are asymptotically normal and consistent as $n \rightarrow \infty$. Finally, we exploit the correlation structure between the number of customers in the queue at different points in time to obtain a third class of moment type estimators, and we show that these are consistent. We repeatedly make use of historical data from 22 composers to illustrate our results.

Keywords: Maximum a posteriori method, moment type estimators, $M/G/\infty$ queue

Inhaltsverzeichnis

| | |
|--|-----------|
| 1. Einführung | 1 |
| 1.1. Beispiele | 1 |
| 1.2. Stochastische Modellbildung und Problemstellung | 5 |
| 1.2.1. Rekonstruktion der Zuordnung | 5 |
| 1.2.2. Schätzen der Verzögerungsverteilung | 6 |
| 1.2.3. Das TaS-Modell | 7 |
| 1.3. Annahmen und Notation | 9 |
| 1.4. Verwandte Problemstellungen | 9 |
| 1.4.1. Deconvolution | 9 |
| 1.4.2. Broken Sample Problem | 10 |
| 1.5. Ein erstes Resultat | 11 |
| 1.5.1. Momentenschätzer | 12 |
| 1.5.2. Maximum-Likelihood-Schätzer | 13 |
| 1.5.3. Vergleich der asymptotischen Varianzen | 16 |
| | |
| 2. Rekonstruktion der Zuordnung | 19 |
| 2.1. Grundlagen | 19 |
| 2.2. Rekonstruktion mit der MAP-Methode | 24 |
| 2.2.1. Problemstellung | 24 |
| 2.2.2. Explizite Lösungen | 29 |
| 2.3. Rekonstruktion mit der MMD-Methode | 37 |
| 2.3.1. Die Struktur der Menge der zulässigen Permutationen | 37 |
| 2.3.2. Problemstellung | 41 |
| 2.3.3. Explizite Lösungen | 45 |
| 2.3.4. Asymptotik bei Exponentialverteilungen | 54 |
| 2.4. Ein Test auf Exponentialverteilung | 61 |
| | |
| 3. Schätzer für die Verzögerungsverteilung | 67 |
| 3.1. Die Likelihood-Methode | 67 |
| 3.2. Die Momentenmethode | 72 |
| 3.2.1. Grundlagen | 72 |
| 3.2.2. Aus einem Poisson-Prozess stammende Geburtszeitpunkte | 74 |

| | |
|---|------------|
| 3.2.3. I.i.d. Geburtszeitpunkte | 81 |
| 3.2.4. Deterministische Geburtszeitpunkte | 94 |
| 3.3. Modellierung als $M/G/\infty$ -Warteschlange | 96 |
| 3.3.1. Grundlagen | 96 |
| 3.3.2. Auf den Betriebsdauern basierende Schätzer | 101 |
| 3.3.3. Auf der Schlängellänge basierende Schätzer | 118 |
| A. Verschiedene Berechnungsformen von Kendalls Tau-Abstand | 127 |
| B. Hilfsaussagen zu Maximum-Likelihood-Schätzern | 129 |
| C. R-Quellcode | 131 |
| Literaturverzeichnis | 137 |

1. Einführung

Das erste Kapitel der vorliegenden Arbeit dient dazu, die untersuchte Problemstellung und die verwendeten Methoden vorzustellen. Wir beginnen in Abschnitt 1.1 mit drei motivierenden Beispielen. Dabei arbeitet das erste dieser Beispiele mit konkreten Daten, die im weiteren Verlauf wiederholt zur Illustration der verschiedenen Verfahren herangezogen werden. Im Abschnitt 1.2 beschreiben wir die grundlegenden Probleme und diskutieren die stochastischen Modelle, für die wir diese Probleme näher untersuchen werden. Nach Einführung und Erläuterung erster Notation in Abschnitt 1.3 werden in Abschnitt 1.4 zwei verwandte Fragestellungen beschrieben und gegen die uns interessierende Situation abgegrenzt. Zum Abschluss dieses einführenden Kapitels beweisen wir in Abschnitt 1.5 ein beispielhaftes Resultat.

1.1. Beispiele

Wir beginnen damit, die in dieser Arbeit untersuchten Problemstellungen exemplarisch darzustellen. Im folgenden Abschnitt werden wir diese Beispiele dann formalisieren.

Beispiel 1.1.1. In Tabelle 1.1 sind die Geburts- und Todesdaten von 22 Komponisten nach dem Gregorianischen Kalender angegeben, jeweils in chronologischer Reihenfolge und ohne Information darüber, welche Daten zusammengehören. (Im Fall von Frédéric Chopin ist der Geburtstag nicht eindeutig gesichert; wir verwenden das in der Geburtsurkunde genannte Datum.) Wie kann nun die Zuordnung von Geburten zu Todesfällen rekonstruiert werden?

Bei 22 Paaren gibt es zunächst einmal $22! \approx 1.12 \cdot 10^{21}$ verschiedene Möglichkeiten, Geburt-Tod-Paare zu bilden, so dass jeder Geburt genau ein und nicht zwei Geburten derselbe Todeszeitpunkt zugeordnet wird. Aus jeder dieser Zuordnungen folgt für alle Komponisten eine bestimmte Lebensdauer, dabei verwenden wir zur Vereinfachung die bei der Berechnung von Zinsen übliche Unterteilung eines Jahres in zwölf Monate zu je 30 Tagen. Dann ergibt z.B. die Zuordnung der ersten Geburt 23.2.1685 zum ersten Todesfall 28.7.1750 eine Lebensdauer von 23555 Tagen für Händel, den ersten in der Liste. Offensichtlich sind nicht alle Zuordnungen sinnvoll, z.B. würde die Zuteilung einer

| Geburten | Todesfälle |
|------------|------------|
| 23.2.1685 | 28.7.1750 |
| 21.3.1685 | 23.7.1757 |
| 26.10.1685 | 14.4.1759 |
| 8.3.1714 | 14.12.1788 |
| 31.3.1732 | 5.12.1791 |
| 27.1.1756 | 31.5.1809 |
| 16.12.1770 | 5.6.1826 |
| 18.11.1786 | 26.3.1827 |
| 31.1.1797 | 19.11.1828 |
| 3.2.1809 | 4.11.1847 |
| 22.2.1810 | 17.4.1849 |
| 8.6.1810 | 29.7.1856 |
| 22.10.1811 | 31.7.1886 |
| 13.9.1819 | 6.11.1893 |
| 10.12.1823 | 20.5.1896 |
| 7.5.1833 | 3.4.1897 |
| 7.5.1840 | 18.9.1903 |
| 8.9.1841 | 1.5.1904 |
| 15.6.1843 | 4.9.1907 |
| 22.8.1862 | 27.4.1915 |
| 6.1.1872 | 11.5.1916 |
| 19.3.1873 | 25.3.1918 |

Tabelle 1.1. Chronologisch geordnete Geburts- und Todesdaten

Geburt zu einem *früher* liegenden Todesdatum auf eine negative Lebensdauer führen. Andererseits kann zusätzlich davon ausgegangen werden, dass die Komponisten ein gewisses Alter nicht überschritten haben, z.B. 150 Jahre.

In Abb. 1.1 ist die tatsächliche Zuordnung grafisch dargestellt. Es gibt „Überkreuzungen“ der Lebensdauer in dem Sinne, dass z.B. Mozart (deutlich) nach Haydn geboren wurde, aber trotzdem vor ihm starb. Den aufsteigend geordneten Geburtsdaten der Reihe nach die aufsteigend geordneten Todesdaten zuzuteilen, führt also nicht zur richtigen Lösung. Andererseits tritt z.B. bei C. Bach und Haydn diese „Überkreuzung“ nicht auf, so dass auch nicht immer die Lebensdauer früher Geborener maximiert werden darf (und damit die Anzahl der beschriebenen „Überkreuzungen“; letztere werden später unter der Bezeichnung „Inversionen“ formalisiert). ◁

Beispiel 1.1.2. Bei einem Museum (Postamt, Kaufhaus, ...) kommen zu zufälligen Zeitpunkten „Kunden“ an. Diese halten sich eine zufällige Zeit in der jeweiligen Institu-

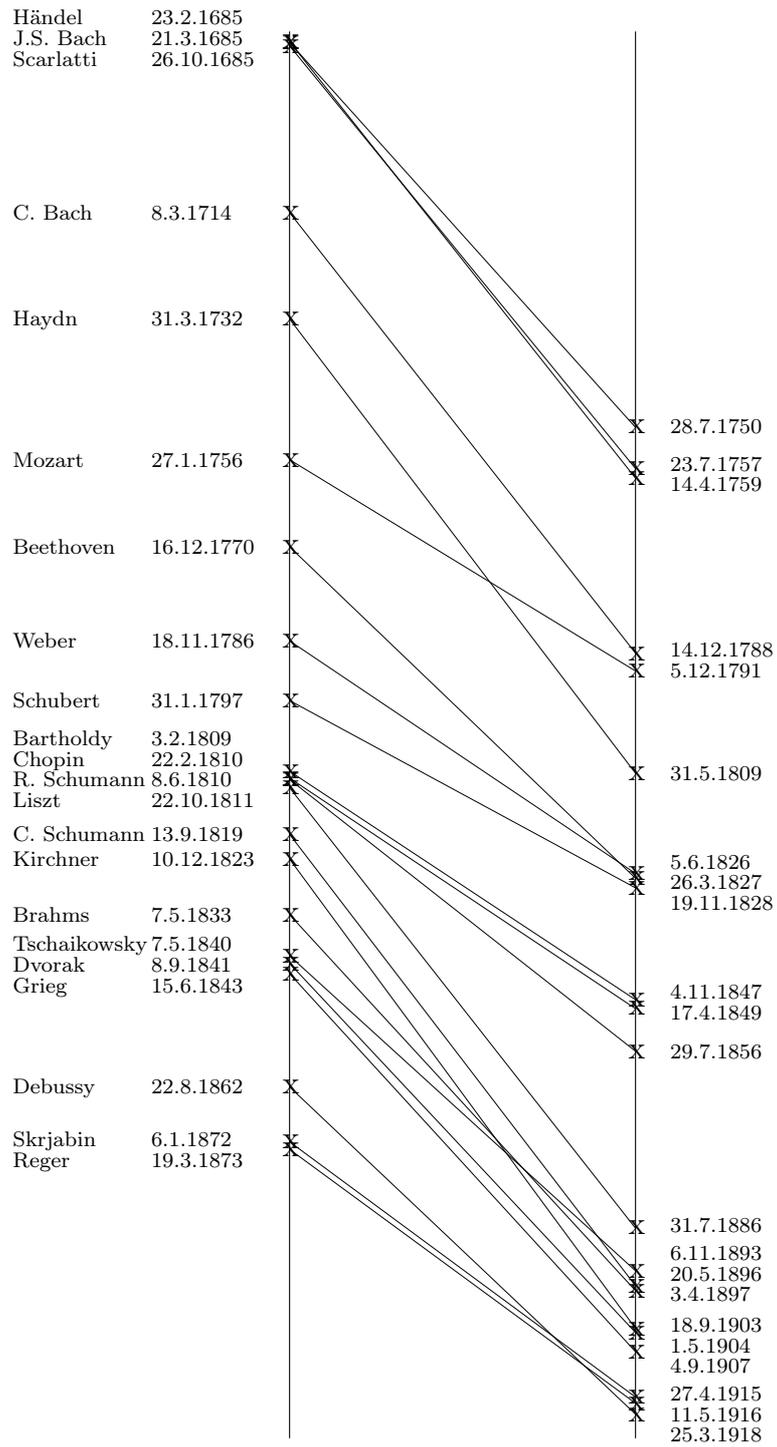


Abbildung 1.1. Korrekte Zuordnung von Geburten zu Todesfällen für 22 Komponisten

tion auf, bevor sie sie wieder verlassen. Die Länge des Aufenthalts sei jeweils unabhängig vom Ankunftszeitpunkt. Analog zu Beispiel 1.1.1 sollen wieder nur geordnete Daten vorliegen, dieses Mal von Ankünften und Abgängen, wie sie z.B. von Messeinrichtungen an Ein- und Ausgang aufgezeichnet würden. Hierbei beschränken wir uns auf Daten aus einem bestimmten Zeitraum, z.B. während eines Tages. Dies führt dazu, dass bei Beginn oder Ende der Aufzeichnung schon bzw. noch Kunden anwesend sein können (z.B. bei einem Kaufhaus, das rund um die Uhr geöffnet hat). Hieraus ergeben sich „Randeffekte“ bei der Beobachtung, da im untersuchten Zeitraum Ankunfts- bzw. Abgangszeiten ohne entsprechendes Gegenstück vorliegen können; insbesondere muss die Anzahl von Ankünften und Abgängen nicht gleich sein. Ziel ist deshalb auch nicht mehr eine Zuordnung der Zeiten untereinander, sondern Aussagen über die zufällige Aufenthaltszeit, genauer deren Verteilung. Weiterhin soll untersucht werden, welche Schlüsse noch möglich sind, wenn anstatt der Ankunfts- und Abgangsdaten nur die Information vorliegt, in welchen Zeiträumen sich Besucher in der jeweiligen Einrichtung aufgehalten haben und in welchen nicht. \triangleleft

Beispiel 1.1.3. Das dritte Beispiel schließlich hat einen versicherungsmathematischen Hintergrund. Bei einer Versicherung gehen zu zufälligen Zeitpunkten Schadensmeldungen ein, wobei jeder dieser Schäden eine schwankende Anzahl weiterer Schäden nach sich zieht. Bei den anfänglichen Schäden könnte es sich etwa um durch einen Hurrikan verursachte Gebäudeschäden handeln. Später folgen dann Forderungen z.B. aus Kranken-, Kfz-, Rechtsschutz- und Lebensversicherung, die sich in der Folge des Hurrikans ergeben, aber je nach Versichertem nicht unbedingt alle Bereiche umfassen. Wie in Beispiel 1.1.2 werden geordnete Daten aus einem bestimmten Zeitraum, in diesem Fall Schadenszeitpunkte während z.B. eines Quartals, betrachtet, und es gibt zwei Klassen von Daten, nämlich die der „auslösenden Schäden“ und die der Folgeschäden. Im allgemeinen, hier nicht weiter untersuchten Fall können zu einem Zeitpunkt aus der ersten Klasse mehrere Zeitpunkte der zweiten Klasse gehören. Auch kann es wieder zu „Randeffekten“ derart kommen, dass Folgeschäden auftauchen, deren „auslösende Schäden“ in einen vorhergehenden Zeitraum fielen. \triangleleft

Allen vorangegangenen Beispielen ist gemein, dass jeweils separat die Zeitpunkte festgehalten werden, an denen zuerst ein Ereignis A und dann ein Ereignis B für bestimmte Objekte eintreten (Geburt - Tod, Ankunft - Abgang, „auslösender“ Schaden und Folgeschaden). Beziehen sich A und B auf dasselbe Objekt, sind sie in diesem Sinne verbunden (oder auch „gekoppelt“) und lassen sich entsprechend paarweise zuordnen. Im ersten Beispiel soll die bestehende, aber unbekannte Zuordnung zwischen Eintrittszeitpunkten von A und B rekonstruiert werden. Bei den anderen beiden Fällen liegen die jeweiligen Daten aus einem zusammenhängenden Zeitraum vor, es beziehen sich aber nicht notwendigerweise immer genau zwei Daten auf ein Objekt. Hier ist die Verteilung der Zeit zu schätzen, um die sich das Eintreten von B nach dem Eintritt von A verzögert.

1.2. Stochastische Modellbildung und Problemstellung

Wir führen für die im vorhergehenden Abschnitt betrachteten Beispiele stochastische Modelle ein und formulieren in diesem Rahmen die in dieser Arbeit untersuchten Problemstellungen. Sind Zufallsvariablen unabhängig und identisch verteilt, so sagen wir stattdessen im Folgenden häufig nur, sie seien „i.i.d.“, nach der geläufigen Abkürzung für „independent and identically distributed“. Weiter wird für festes $n \in \mathbb{N}$ und Zufallsvariablen X_1, \dots, X_n die Anordnung letzterer in aufsteigender Reihenfolge als Ordnungsstatistiken $X_{(n:1)}, \dots, X_{(n:n)}$ bezeichnet.

1.2.1. Rekonstruktion der Zuordnung

Es sei $\tilde{X}_1, \tilde{X}_2, \dots$ eine Folge von absolutstetig verteilten Zufallsvariablen. Für die reellen Datenpaare $(\tilde{X}_1, \tilde{Y}_1), \dots, (\tilde{X}_n, \tilde{Y}_n)$ und $n \in \mathbb{N}$ gelte $\tilde{X}_i < \tilde{Y}_i$ und $\tilde{Y}_i = \tilde{X}_i + Z_i$ für $i = 1, \dots, n$, wobei Z_1, \dots, Z_n i.i.d. mit bekannter Dichte f_Z bzgl. des Lebesgue-Maßes und unabhängig von $\tilde{X}_1, \dots, \tilde{X}_n$ seien. Wir interpretieren (\tilde{X}_i) und (\tilde{Y}_i) als Geburts- bzw. Todeszeitpunkte (einer Person, eines Testobjekts, ...) und damit die (Z_i) als Lebensdauern. Für die (\tilde{X}_i) treffen wir zunächst außer den bereits genannten keine weiteren Annahmen. Alle Größen sind als stetige Zufallsvariablen mit Wahrscheinlichkeit 1 paarweise verschieden, insbesondere auch \tilde{X}_i und \tilde{Y}_i für alle i . Im Folgenden gemachte Aussagen, die dies voraussetzen, gelten entsprechend nur mit Wahrscheinlichkeit 1, ohne dass wir jedes Mal explizit darauf hinweisen. Neben der offensichtlichen Nichtnegativität einer Lebensdauer, mithin $f_Z(z) = 0$ für $z < 0$, beschäftigen wir uns ausschließlich mit Verteilungen, für die $f_Z(z) > 0$ für alle $z > 0$ ist; somit gibt es in unserem Modell keine Obergrenze für die Lebensdauer.

Es seien $X_i := \tilde{X}_{(n:i)}$ und $Y_i := \tilde{Y}_{(n:i)}$, jeweils für $i = 1, \dots, n$. Wir untersuchen den Fall, in dem nicht mehr die Paare (und damit die Zuordnung untereinander) bekannt sind, sondern nur noch die Ordnungsstatistiken $(X_i)_{i=1, \dots, n}$ und $(Y_i)_{i=1, \dots, n}$ vorliegen. Gesucht ist die zufällige Permutation Π („wahre“ Permutation), so dass $((X_i, Y_{\Pi(i)}))_{i=1, \dots, n}$ wieder die ursprünglichen Paare bilden. Bei der Suche können von vornherein alle Permutationen ausgeschlossen werden, die nicht zu positiven Lebensdauern führen. Dies konkretisieren wir in der folgenden Definition.

Definition 1.2.1. Für $n \in \mathbb{N}$ seien $x, x_1, \dots, x_n, y, y_1, \dots, y_n \in \mathbb{R}$.

- (a) Gilt $x < y$, so heißt y *zulässig für x* , und sonst *unzulässig*.
- (b) Für $i, j \in \{1, \dots, n\}$ heißt i *zulässig für j* , wenn gilt $x_i < y_j$, und sonst *unzulässig*.

- (c) Eine Permutation $\pi \in S_n$ heißt *zulässig*, wenn $y_{\pi(i)}$ für jedes $i \in \{1, \dots, n\}$ zulässig für x_i ist, und sonst *unzulässig*.
- (d) Wir definieren $S'_n := \{\pi \in S_n : x_i < y_{\pi(i)}, i = 1, \dots, n\}$ als die *Menge der zulässigen Permutationen*.

Verwendung von S'_n impliziert im Folgenden also immer das Vorliegen von Daten $(x_i)_{i=1}^n$ und $(y_i)_{i=1}^n$, ohne dass diese explizit genannt sein müssen. Die Rekonstruktion von Π ist Gegenstand von Kapitel 2. Wir behandeln zwei Verfahren, die maximum a posteriori Methode und die Minimierung des mittleren Abstands. Beide lassen sich über Verlustfunktionen formalisieren.

1.2.2. Schätzen der Verzögerungsverteilung

Wir modellieren die in Beispiel 1.1.2 beschriebene Situation durch ein sog. Warteschlangensystem. Dies besteht allgemein aus dem Prozess $(X_i)_{i \in \mathbb{N}}$ der Ankünfte und dem Prozess $(Z_i)_{i \in \mathbb{N}}$ der Bedienzeiten von bestimmten, als Kunden bezeichneten Individuen sowie der Anzahl $c \in \mathbb{N} \cup \{\infty\}$ der Bediener. Im Rahmen dieses Modells soll die untersuchte Verzögerung der individuellen Bedienzeit entsprechen. Die (X_i) mögen einen Poisson-Prozess mit Rate θ bilden, d.h. die Zwischenankunftszeiten sind exponentialverteilt mit Parameter θ . Die Bedienzeiten (Z_i) seien i.i.d. mit einer Dichte f_Z bzgl. des Lebesgue-Maßes und unabhängig von den Ankünften. Wir fassen die Zeit, die sich die einzelnen Kunden im System aufhalten, als reine Bedienzeit auf, jeder Kunde wird demnach bei Ankunft sofort bedient. Damit dies möglich ist, muss es unendlich viele Bediener geben, also $c = \infty$ gelten. Die für das System wesentlichen Größen können auch in der sogenannten Kendall-Notation (siehe z.B. [Asm03, S. 61-63]) als Tripel angegeben werden. An der ersten Position steht die Art des Ankunftsprozesses, im betrachteten Fall ist dies M für „Markovian“. Für die Bedienzeiten fordern wir zunächst keine bestimmte Verteilung, was durch G („General Distribution“) an der zweiten Stelle ausgedrückt wird. An der dritten Stelle schließlich steht die Anzahl der Bediener. Wir modellieren die Situation aus Beispiel 1.1.2 also als $M/G/\infty$ -Warteschlangensystem.

Die Verteilung der Bedienzeit wird auf Grundlage von verschiedenen Informationen bestimmt, dabei sei $(Q_t)_{t \in \mathbb{R}^+}$ die Anzahl der Kunden im System zur Zeit t . Im ersten Fall liegt nur $1_{Q_t > 0}$ für $0 \leq t \leq T$ und festes $T \in \mathbb{R}^+$ vor, d.h. für jedes t aus dem genannten Bereich die Information, ob zu diesem Zeitpunkt mindestens ein Kunde im System ist oder nicht. Dies führt auf die Analyse der sog. idle-busy-cycles eines Bediensystems. Im zweiten Fall betrachten wir ausschließlich Realisierungen von $(Q_t)_{0 \leq t \leq T}$. Auf diese Weise können zwar die Zeitpunkte von Ankünften und Abgängen abgelesen werden, aber nicht,

welche Zeiten jeweils zusammengehören. In beiden Fällen ist es möglich, dass zur Zeit $t = 0$ schon bzw. zur Zeit $t = T$ noch Kunden im System sind.

Die als zweites beschriebene Situation findet sich auch in Beispiel 1.1.1 wieder: Dort entspricht eine Geburt einer Ankunft, ein Todesfall einem Abgang und die Lebensdauer der Bedienzeit. Der zugehörige Prozess (Q_t) mit den Daten aus Tabelle 1.1 ist in Abb. 1.2 dargestellt. Man beachte, dass sich umgekehrt auch die Daten aus der Abbildung ablesen ließen, beide Darstellungen somit äquivalent sind. Insbesondere führen Situationen, bei denen nur die Anzahl der „Kunden“ im System als Funktion der Zeit beobachtbar ist, in natürlicher Weise zu dem eingangs geschilderten Datenverlust, also dem Übergang zu den Ordnungsstatistiken.

Zur Vereinheitlichung benutzen wir im Großteil dieser Arbeit die Bezeichnungen „Geburten“, „Todesfälle“ und „Lebensdauer“ und erwähnen den Bezug zu anderen Modellen nicht jedes Mal explizit. In Abschnitt 3.3 kommen wir dann aber auf die Modellierung als Warteschlangensystem zurück und verwenden dort die übliche Nomenklatur, sprechen also von „Ankünften“, „Abgängen“ und „Bedienzeiten“.

1.2.3. Das TaS-Modell

Formale Grundlage der Darstellung von auslösenden und einer schwankenden Zahl von später folgenden Schäden aus Beispiel 1.1.3 ist das sog. TaS-Modell aus [BG05], wobei TaS für „thin and shift“ steht. Dieses Modell besteht aus drei voneinander unabhängigen Komponenten: Einem Poisson-Prozess auf \mathbb{R} sowie zwei Verfahren zu Ausdünnung („thinning“) und Verschiebung („shifting“). Die Zeitpunkte $(X_l)_{l \in \mathbb{Z}}$, zu denen die auslösenden Schäden („Primärschäden“) eintreten, stammen aus dem Poisson-Prozess. In der Modellierung gibt es nun zunächst zu *jedem* Primärschaden $d \in \mathbb{N}$ Folgeschäden („Sekundärschäden“), dabei wird die Verzögerung durch eine Folge $(Z_l)_{l \in \mathbb{Z}}$ von unabhängig und identisch verteilten d -dimensionalen Zufallsvariablen dargestellt. Man erhält also d Folgen $(X_l + Z_{l,1})_{l \in \mathbb{Z}}, \dots, (X_l + Z_{l,d})_{l \in \mathbb{Z}}$ als Zeitpunkte der Sekundärschäden. Nun werden einige (ggf. auch alle oder keiner) der Sekundärschäden durch die mittels einer Verteilung $(p_D)_{D \subseteq \{1, \dots, d\}}$ gesteuerten Ausdünnung eliminiert. Für jedes $X_l, l \in \mathbb{Z}$, gibt dann ein D_l mit Verteilung (p_D) an, welche auf X_l folgenden Sekundärschäden tatsächlich eintreten, nämlich die mit den Eintrittszeitpunkten $(X_l + Z_{l,i})_{i \in D_l}$. Die bisher geschilderten Fälle lassen sich durch den einfachen Spezialfall $d = 1$ und $p_{\{1\}} = 1$ des Modells darstellen, auf diesen beschränken wir uns auch im Rahmen der vorliegenden Arbeit.

1.3. Annahmen und Notation

Für den Rest dieser Arbeit bezeichne (\tilde{X}_i) die Folge der Geburten (Ankünfte, ...), (Z_i) die i.i.d. und von (\tilde{X}_i) unabhängige Folge der Lebensdauern (Bedienzeiten, ...) sowie (\tilde{Y}_i) , $\tilde{Y}_i := \tilde{X}_i + Z_i$, die Folge der Todeszeitpunkte (Abgänge, ...). Für festes $n \in \mathbb{N}$ und $1 \leq i \leq n$ sei ferner $X_i := \tilde{X}_{(n:i)}$ und $Y_i := \tilde{Y}_{(n:i)}$, X_1, \dots, X_n und Y_1, \dots, Y_n sind also die Ordnungsstatistiken von $\tilde{X}_1, \dots, \tilde{X}_n$ bzw. $\tilde{Y}_1, \dots, \tilde{Y}_n$. Wir untersuchen die Situation, in der nur $(X_i)_{i=1, \dots, n}$ und $(Y_i)_{i=1, \dots, n}$ für ein festes $n \in \mathbb{N}$ vorliegen und keine Information über die Zusammengehörigkeit vorhanden ist.

Als abkürzende Schreibweise definieren wir $\bar{V}_{n,k} := \frac{1}{n} \sum_{i=1}^n V_i^k$ für die empirischen Momente zu den Zufallsvariablen V_1, \dots, V_n , $k, n \in \mathbb{N}$. Wird ein unbekannter Parameter ϑ mittels der Momentenmethode oder der Maximum-Likelihood-Methode aus einer Stichprobe der Größe n geschätzt, bezeichnen wir den jeweiligen Schätzer mit $\hat{\vartheta}_n^{\text{MM}}$ bzw. $\hat{\vartheta}_n^{\text{ML}}$. Die Indizes n , MM und ML werden manchmal zur Vereinfachung weggelassen, wenn dies nicht zu Unklarheiten führt. Für die Grenzvarianzen dieser Schätzer verwenden wir gelegentlich die Bezeichnungen $\text{var}(\hat{\vartheta}^{\text{MM}})$ bzw. $\text{var}(\hat{\vartheta}^{\text{ML}})$, weiter ist $\widehat{\text{var}}(\hat{\vartheta}_n)$ der bekannte erwartungstreue Schätzer für die Varianz bei einer an entsprechender Stelle spezifizierte Anzahl von Realisierungen von ϑ_n . Wir schreiben $E_{\vartheta}(\hat{\vartheta})$ und $\text{var}_{\vartheta}(\hat{\vartheta})$ für Erwartungswert bzw. Varianz des Schätzers $\hat{\vartheta}$, wenn ϑ der wahre Parameter ist. Auch hier lassen wir in eindeutigen Fällen den Index weg.

Für $n \in \mathbb{N}$ bezeichnet \mathfrak{B}^n die Borelsche σ -Algebra und ℓ^n das Lebesgue-Maß auf \mathbb{R}^n ; δ_x das Dirac-Maß in einem gewissen Punkt x . Mit „log“ ist immer der natürliche Logarithmus gemeint, weiter sind Ψ und Ψ' die Digamma- bzw. Trigammafunktion. Ein häufig verwendetes Hilfsmittel ist die sog. Deltamethode, siehe dazu z.B. [CB90, S. 329f.] oder [Bil95, S. 358f.]. Generell finden sich in der Arbeit nicht erläuterte Konzepte in diesen Quellen.

1.4. Verwandte Problemstellungen

In diesem Abschnitt zeigen wir Gemeinsamkeiten und Unterschiede unserer Problemstellung in Bezug auf zwei andere Themengebiete auf.

1.4.1. Deconvolution

Beim sog. Deconvolution-Problem (siehe z.B. [Jon98]) liegt für $n \in \mathbb{N}$ eine Stichprobe Y_1, \dots, Y_n vor. Dabei gilt $Y_i = X_i + Z_i$ für alle i mit zwei unabhängigen Zufallsvariablen X_i und Z_i , ferner sind (X_i) und (Z_i) jeweils i.i.d. Die Dichte f_X von X_1 ist bekannt,

die Verteilungsfunktion F_Z von Z_1 dagegen nicht, genauso wenig wie die tatsächlichen Werte von X_1, \dots, X_n . In dieser Situation sind auch die (Y_i) i.i.d. und für ihre Dichte f_Y gilt

$$f_Y(y) = \int f_X(y-x) dF_Z(x).$$

Ziel ist es nun, F_Z aus der Stichprobe zu bestimmen.

Unsere Untersuchungen unterscheiden sich von diesem Problem in zwei Punkten. Zum einen liegen die geordneten Werte von X_1, \dots, X_n vor und können somit in die Betrachtung einbezogen werden. Zum anderen nehmen wir im entsprechenden Kapitel eine konkrete parametrische Familie an, aus der die (Z_i) stammen, so dass die Anwendung von parametrischen Verfahren möglich ist.

1.4.2. Broken Sample Problem

Das Broken Sample Problem in seiner allgemeinen Form wurde nach unserer Kenntnis in [DFG71] eingeführt. Es liegt eine Stichprobe $(X_1, Y_1), \dots, (X_n, Y_n)$ vor, die aus der bekannten gemeinsamen Verteilung zweier Zufallsvariablen bzw. -vektoren auf \mathbb{R}^d , $d \geq 2$, stammt. Deren Dichte besitzt eine gewisse parametrische Form, die insbesondere eine Anwendbarkeit der Ergebnisse auf eine multivariate Normalverteilung sicherstellt. Die beobachteten Werte liegen aber nur in der Form X_1, \dots, X_n und $Y_{\pi(1)}, \dots, Y_{\pi(n)}$ mit einer unbekanntem Permutation π (z.B. der zur Ordnungsstatistik führenden) vor, also ohne Kenntnis darüber, welche Werte jeweils zusammengehören. Nun sollen die ursprünglichen Paare bestimmt werden. Ferner bietet sich ein Vergleich mit dem in [DG80] betrachteten Spezialfall an, dass die Paare aus einer bivariaten Normalverteilung mit unbekanntem Parametern stammen. Hier ist der Korrelationskoeffizient ρ zu schätzen, während die Rekonstruktion der ursprünglichen Paare zweitrangig ist. Da die Erwartungswerte und Varianzen auch aus den „getrennten“ Werten bestimmt werden können, ermöglicht die Kenntnis von ρ , die Normalverteilung komplett zu charakterisieren.

Die geschilderten Situationen haben mit unserer Fragestellung gemein, dass nur die einzelnen Komponenten von Elementen des \mathbb{R}^2 (bzw. in zwei Teile zerlegte Vektoren des \mathbb{R}^d) vorliegen, nicht aber die individuellen Zuordnungen. Bei der Rekonstruktion der Zuordnung (Kap. 2) ist die Verteilung der Geburtszeitpunkte irrelevant, so dass wir sie als Konstanten annehmen. In Kapitel 3 stammen die (X_i) aus einem Poisson-Prozess, lediglich in Teilen von Abschnitt 3.2 wird diese Annahme aufgegeben. Die Paare $((X_i, Y_i))_{i=1, \dots, n}$ sind also in den meisten Fällen nicht identisch verteilt. Außerdem gehen wir von einer positiven Lebensdauer aus, wodurch die X -Werte immer echt kleiner als die Y -Werte sind (Da diese Annahme in Abschnitt 3.2 keine Rolle spielt und dort auch der i.i.d.-Fall betrachtet wird, kann das dort beschriebene Verfahren u.U. auch für einige

Varianten des Broken Sample Problems verwendet werden.). Bei Verteilungen der in [DFG71] betrachteten Form dagegen ist $P(X_1 < Y_1) = 1$ nicht möglich. Ferner benötigen wir bei der Rekonstruktion keine parametrische Annahme an irgendeine der beteiligten Verteilungen.

Wenn wir bei der Unabhängigkeit der Lebensdauern (Z_i) von den Geburtszeitpunkten (X_i) bleiben, aber wie im obigen Spezialfall des Broken Sample Problems jeweils dieselbe Normalverteilung (zur Vereinfachung mit Erwartungswert 0) für X_1, X_2, \dots bzw. Z_1, Z_2, \dots unterstellen, dann erhalten wir

$$\begin{pmatrix} X_1 \\ X_1 + Z_1 \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_2^2 \end{pmatrix} \right),$$

also eine zweidimensionale Normalverteilung eines ganz bestimmten Typs. Diese Verteilungen werden offensichtlich durch die Marginalverteilungen determiniert, da die Kovarianz mit der Varianz der vorderen Komponente übereinstimmt. Nach den Ausführungen oben können in diesem Fall also alle relevanten Größen aus den einander nicht zugeordneten Komponenten geschätzt werden.

1.5. Ein erstes Resultat

Wir demonstrieren anhand einer beispielhaften Situation, dass die Berücksichtigung der X -Werte im betrachteten Fall sinnvoll sein kann. Seien dazu $(X_i)_{i \in \mathbb{N}}$ i.i.d. $\text{unif}(0,1)$ -verteilte und $(Z_i)_{i \in \mathbb{N}}$ i.i.d. $\text{Exp}(\theta)$ -verteilte, voneinander unabhängige Folgen von Zufallsvariablen sowie f_X und f_Z Dichten von X_1 bzw. Z_1 bzgl. des Lebesgue-Maßes. Im Folgenden wird der Momenten- und der Maximum-Likelihood-Schätzer für θ bestimmt, jeweils zum einen nur aus Realisierungen von $(Y_i)_{i=1, \dots, n}$, $Y_i := X_i + Z_i$ für alle i , zum anderen bei zusätzlicher Kenntnis der Werte von $(X_i)_{i=1, \dots, n}$. Da nur ein unbekannter (skalärer) Parameter geschätzt werden soll, ist es in diesem Rahmen irrelevant, ob die Zuordnung der X - und Y -Werte untereinander bekannt ist oder nicht. Dies liegt an der trivialen Tatsache, dass

$$\sum_{i=1}^n (Y_{\pi(i)} - X_i) = \sum_{i=1}^n Y_i - \sum_{i=1}^n X_i = \sum_{i=1}^n Z_i$$

für jedes $\pi \in S_n$ ist; entsprechend gilt auch

$$\prod_{i=1}^n \exp(Y_{\pi(i)} - X_i) = \exp\left(\sum_{i=1}^n Y_i\right) \cdot \exp\left(-\sum_{i=1}^n X_i\right).$$

Wir werden im weiteren Verlauf der Arbeit noch häufiger auf diese Beobachtungen zurückkommen. Im aktuellen Abschnitt dienen sie als Begründung dafür, nicht zwischen unsortierten Werten und Ordnungsstatistiken zu unterscheiden.

Die (Y_i) sind in der vorliegenden Situation ebenfalls i.i.d. und für ihre Dichte f_Y gilt

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_X(x)g_Z(y-x) dx = \int_0^{y \wedge 1} \theta e^{-\theta(y-x)} dx \\ &= e^{-\theta(y-x)} \Big|_0^{y \wedge 1} = \begin{cases} 1 - e^{-\theta y}, & 0 < y \leq 1, \\ (e^\theta - 1)e^{-\theta y}, & y > 1, \end{cases} \end{aligned} \quad (1.1)$$

für $y > 0$ und 0 sonst, $y \wedge 1 := \min\{y, 1\}$. Als Erwartungswert und Varianz ergeben sich unter Ausnutzung der Unabhängigkeit von X_1 und Z_1

$$EY_1 = EX_1 + EZ_1 = \frac{1}{2} + \frac{1}{\theta}, \quad \text{var}(Y_1) = \text{var}(X_1) + \text{var}(Z_1) = \frac{1}{12} + \frac{1}{\theta^2}.$$

1.5.1. Momentenschätzer

Nun soll $EZ_1 = \frac{1}{\theta}$ durch die beiden Momentenschätzer $(\widehat{EZ}_o)_n^{\text{MM}}$ („ohne (X_i) “) und $(\widehat{EZ}_m)_n^{\text{MM}}$ („mit (X_i) “) geschätzt werden; wir definieren diese durch

$$(\widehat{EZ}_o)_n^{\text{MM}} := \bar{Y}_{n,1} - \frac{1}{2}, \quad (\widehat{EZ}_m)_n^{\text{MM}} := \bar{Y}_{n,1} - \bar{X}_{n,1}.$$

Wie angekündigt lassen wir einen oder auch beide der Indizes n und MM weg, wenn dies nicht zu Unklarheiten führt.

Lemma 1.5.1. *Es gilt*

$$\begin{aligned} \sqrt{n} \left((\widehat{EZ}_o)_n^{\text{MM}} - \frac{1}{\theta} \right) &\xrightarrow{d} U, U \sim N \left(0, \frac{1}{\theta^2} + \frac{1}{12} \right), \\ \sqrt{n} \left((\widehat{EZ}_m)_n^{\text{MM}} - \frac{1}{\theta} \right) &\xrightarrow{d} U, U \sim N \left(0, \frac{1}{\theta^2} \right). \end{aligned}$$

Beweis. Die asymptotische Normalität folgt aus dem zentralen Grenzwertsatz. Weiter gilt

$$E\widehat{EZ}_o = E\bar{Y}_{n,1} - \frac{1}{2} = \frac{1}{\theta} = E\widehat{EZ}_m$$

und für die Varianzen ergibt sich

$$\text{var}(\widehat{EZ}_o) = \text{var}(\bar{Y}_{n,1}) = \frac{1}{n} \left(\frac{1}{\theta^2} + \frac{1}{12} \right), \quad \text{var}(\widehat{EZ}_m) = \text{var}(\bar{Z}_{n,1}) = \frac{1}{\theta^2 n}. \quad \square$$

Mit $f(x) := \frac{1}{x}$ erhalten wir aus $\widehat{\text{EZ}}_o$ und $\widehat{\text{EZ}}_m$ für θ die beiden Schätzer

$$(\hat{\theta}_o)_n^{\text{MM}} := f((\widehat{\text{EZ}}_o)_n^{\text{MM}}) = \frac{1}{(\widehat{\text{EZ}}_o)_n^{\text{MM}}}, \quad (\hat{\theta}_m)_n^{\text{MM}} := f((\widehat{\text{EZ}}_m)_n^{\text{MM}}) = \frac{1}{(\widehat{\text{EZ}}_m)_n^{\text{MM}}},$$

auch hier verzichten wir gelegentlich auf die Indizes n und MM.

Satz 1.5.2. *Es gilt*

$$\sqrt{n}((\hat{\theta}_o)_n^{\text{MM}} - \theta) \xrightarrow{d} U, U \sim N\left(0, \theta^2 + \frac{\theta^4}{12}\right), \quad \sqrt{n}((\hat{\theta}_m)_n^{\text{MM}} - \theta) \xrightarrow{d} U, U \sim N(0, \theta^2).$$

Beweis. Mit $f'(x) = \frac{1}{x^2}$ folgt die Behauptung aus Lemma 1.5.1, der Delta-Methode und

$$f'(x) \Big|_{x=\frac{1}{\theta}} = \frac{-1}{x^2} \Big|_{x=\frac{1}{\theta}} = -\theta^2. \quad \square$$

1.5.2. Maximum-Likelihood-Schätzer

Wir beginnen mit dem ML-Schätzer $(\hat{\theta}_o)_n^{\text{ML}}$ für θ in der Situation, in der nur die (Y_i) berücksichtigt werden. Zunächst geben wir eine Gleichung an, aus der $(\hat{\theta}_o)_n^{\text{ML}}$ bestimmt werden kann, auch wenn dies zur Berechnung der eigentlich interessierenden Grenzwartianz unnötig ist.

Lemma 1.5.3. *Wurden die Werte y_1, \dots, y_n als Realisierungen von Y_1, \dots, Y_n beobachtet, ergibt sich der ML-Schätzer $(\hat{\theta}_o)_n^{\text{ML}}$ als Nullstelle der Funktion*

$$f(\theta) = (n - k) \left(1 + \frac{1}{e^\theta - 1}\right) - \sum_{i=1}^n y_i + \sum_{i=1}^k \frac{y_i}{1 - e^{-\theta y_i}}.$$

Dabei ist $k := k(y_1, \dots, y_n) := \#\{1 \leq i \leq n : y_i \leq 1\}$ die Anzahl der Werte in der Stichprobe, die kleiner oder gleich 1 sind. Besitzt f eine Nullstelle, so ist diese eindeutig.

Beweis. Mit (1.1) folgt für die Likelihoodfunktion

$$l(\theta|y_1, \dots, y_n) = \left(\prod_{i=1}^k (1 - e^{-\theta y_i}) \right) \cdot \left(\prod_{i=k+1}^n e^{-\theta y_i} (e^\theta - 1) \right),$$

Logarithmieren führt auf

$$\log(l(\theta|y_1, \dots, y_n)) = \sum_{i=1}^k \log(1 - e^{-\theta y_i}) + (n - k) \log(e^\theta - 1) - \theta \sum_{i=k+1}^n y_i$$

und als Ableitung erhalten wir

$$\begin{aligned} \frac{d}{d\theta} \log(l(\theta|y_1, \dots, y_n)) &= \sum_{i=1}^k \frac{y_i e^{-\theta y_i}}{1 - e^{-\theta y_i}} + (n - k) \frac{e^\theta}{e^\theta - 1} - \sum_{i=k+1}^n y_i \\ &= (n - k) \left(1 + \frac{1}{e^\theta - 1} \right) - \sum_{i=1}^n y_i + \sum_{i=1}^k \frac{y_i}{1 - e^{-\theta y_i}}. \end{aligned} \quad (1.2)$$

Monotonieüberlegungen (oder die zweite Ableitung) zeigen, dass dieser Ausdruck höchstens eine Nullstelle besitzt und die Likelihoodfunktion im Fall der Existenz dort ein Maximum annimmt. \square

Satz 1.5.4. *Es gilt*

$$\sqrt{n}((\hat{\theta}_n)^{\text{ML}} - \theta) \xrightarrow{d} U, U \sim N(0, \sigma^2)$$

mit

$$\sigma^2 := \left(\frac{2\zeta(3) + \theta^2 \log(e^\theta - 1) - \theta^3 - 2\theta \text{Li}_2(e^{-\theta}) - 2 \text{Li}_3(e^{-\theta})}{\theta^3} + \frac{1}{\theta(e^\theta - 1)} \right)^{-1},$$

wobei $\text{Li}_s(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^s}$ den Polylogarithmus und $\zeta(3)$ die Apéry-Konstante bezeichnet.

Beweis. Wir benutzen, dass im vorliegenden Fall die Grenzvarianz von $(\hat{\theta}_n)^{\text{ML}}$ gleich dem Inversen der Fisher-Information I ist, und weiter, dass unter bestimmten, hier erfüllten Bedingungen nach [LC98, S. 125] für eine Zufallsvariable X mit Dichte $f(x|\theta)$ gilt

$$I = -E \left(\frac{\partial^2}{\partial \theta^2} \log(f(X|\theta)) \right).$$

Man erhält hier für $t > 0$

$$\log(f(t)) = \begin{cases} \log(1 - e^{-\theta t}), & 0 < t \leq 1, \\ \log(e^\theta - 1) - \theta t, & t > 1, \end{cases}$$

und es ergibt sich

$$\frac{\partial}{\partial \theta} \log(f(t)) = \begin{cases} \frac{te^{-\theta t}}{1 - e^{-\theta t}}, & 0 < t \leq 1, \\ \frac{e^\theta}{e^\theta - 1} - t, & t > 1, \end{cases} \quad \frac{\partial^2}{\partial \theta^2} \log(f(t)) = \begin{cases} \frac{-t^2 e^{-\theta t}}{(1 - e^{-\theta t})^2}, & 0 < t \leq 1, \\ \frac{-e^\theta}{(e^\theta - 1)^2}, & t > 1. \end{cases}$$

Die Fisher-Information I berechnet sich jetzt zu

$$\begin{aligned} I &= \int_0^1 \frac{x^2 e^{-\theta x}}{(1 - e^{-\theta x})^2} \cdot (1 - e^{-\theta x}) dx + \int_1^\infty \frac{e^\theta}{(e^\theta - 1)^2} \cdot (e^\theta - 1) e^{-\theta x} dx \\ &= \int_0^1 \frac{x^2 e^{-\theta x}}{1 - e^{-\theta x}} dx + \frac{e^\theta}{e^\theta - 1} \int_1^\infty e^{-\theta x} dx \\ &= \frac{2\zeta(3) + \theta^2 \log(e^\theta - 1) - \theta^3 - 2\theta \text{Li}_2(e^{-\theta}) - 2 \text{Li}_3(e^{-\theta})}{\theta^3} + \frac{1}{\theta(e^\theta - 1)}, \end{aligned}$$

dann ergibt sich die Behauptung durch Invertieren dieses Ausdrucks. \square

Als nächstes betrachten wir den ML-Schätzer $(\hat{\theta}_m)_n^{\text{ML}}$, der sowohl die (X_i) als auch die (Y_i) berücksichtigt.

Lemma 1.5.5. *Es ist*

$$(\hat{\theta}_m)_n^{\text{ML}} := (\bar{Y}_{n,1} - \bar{X}_{n,1})^{-1}$$

der aus der gemeinsamen Dichte der (X_i) und (Y_i) bestimmte ML-Schätzer für θ .

Beweis. Die gemeinsame Dichte von X_1 und Y_1 ist

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) = f_Z(y-x)f_X(x) = \mathbf{1}_{(0,1)}(x) \cdot \theta e^{-\theta(y-x)} = \mathbf{1}_{(0,1)}(x) \cdot \theta e^{-\theta y} e^{\theta x}.$$

Dann ergeben sich die Likelihood-Funktion

$$l(\theta|x_1, \dots, x_n, y_1, \dots, y_n) = \theta^n e^{-\theta \sum_{i=1}^n y_i} e^{\theta \sum_{i=1}^n x_i},$$

die Loglikelihoodfunktion

$$\log(l(\theta|x_1, \dots, x_n, y_1, \dots, y_n)) = n \log(\theta) - \theta \sum_{i=1}^n y_i + \theta \sum_{i=1}^n x_i$$

und die erste Ableitung

$$\frac{d}{d\theta} \log(l(\theta|x_1, \dots, x_n, y_1, \dots, y_n)) = \frac{n}{\theta} - \sum_{i=1}^n y_i + \sum_{i=1}^n x_i.$$

Letztere hat im Punkt

$$\theta_0 := \frac{1}{\frac{1}{n}(\sum_{i=1}^n y_i - \sum_{i=1}^n x_i)} = \frac{1}{\bar{y}_{n,1} - \bar{x}_{n,1}}$$

die einzige Nullstelle. Wieder folgt durch Monotonieüberlegungen bzw. Betrachtung der zweiten Ableitung, dass die Likelihoodfunktion in θ_0 ihr Maximum annimmt. \square

Die ML-Methode führt in diesem Fall also auf denselben Schätzer wie die Momentenmethode, somit ergibt sich mit Satz 1.5.2 die analoge asymptotische Verteilungsaussage.

Korollar 1.5.6. *Es gilt*

$$\sqrt{n}((\hat{\theta}_m)_n^{\text{ML}} - \theta) \xrightarrow{d} U, U \sim N(0, \theta^2).$$

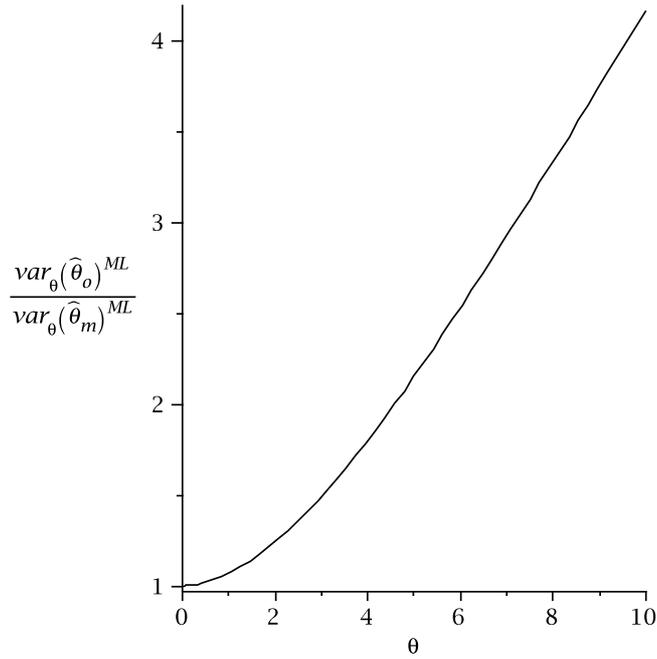


Abbildung 1.3. Vergleich der asymptotischen Varianzen von $(\hat{\theta}_o)_n^{ML}$ und $(\hat{\theta}_m)_n^{ML}$

1.5.3. Vergleich der asymptotischen Varianzen

Nach den Ergebnissen in Satz 1.5.2 wird die asymptotische Varianz bei den Momentenschätzern durch Berücksichtigung der (X_i) verringert; in diesem Sinne ist $(\hat{\theta}_m)_n^{MM}$ also besser als $(\hat{\theta}_o)_n^{MM}$. Wir präzisieren dies im folgenden Korollar.

Korollar 1.5.7. *Es ist*

$$v_{MM}(\theta) := \frac{\text{var}_\theta((\hat{\theta}_o)^{MM})}{\text{var}_\theta((\hat{\theta}_m)^{MM})} = 1 + \frac{\theta^2}{12}$$

sowie $\lim_{\theta \rightarrow 0} v_{MM}(\theta) = 1$ und $\lim_{\theta \rightarrow \infty} v_{MM}(\theta)/\theta^2 = 12^{-1}$.

Bei den ML-Schätzern ist der Vergleich der asymptotischen Varianzen aus Satz 1.5.4 und Korollar 1.5.6 wegen der komplizierteren Ausdrücke nicht so einfach. Deshalb findet sich in Abb. 1.3 eine graphische Auswertung von $v_{ML}(\theta) := \text{var}_\theta((\hat{\theta}_o)^{ML}) / \text{var}_\theta((\hat{\theta}_m)^{ML})$, dem Quotienten der asymptotischen Varianzen von $(\hat{\theta}_o)_n^{ML}$ und $(\hat{\theta}_m)_n^{ML}$, für $\theta \in (0, 10)$. Mit wachsendem θ wird $\text{var}_\theta((\hat{\theta}_o)^{ML})$ deutlich (ungefähr Faktor 4 bei $\theta = 10$) größer als $\text{var}_\theta((\hat{\theta}_m)^{ML})$, während die Varianzen bei kleinen Werten von θ nahe beieinander liegen. Das folgende Lemma zeigt, dass sich die asymptotischen Varianzen bei den ML-Schätzern für $\theta \rightarrow 0, \theta \rightarrow \infty$ ähnlich verhalten wie bei den Momentenschätzern. Allerdings geht

$v_{\text{MM}}(\theta)$ mit $\theta \rightarrow \infty$ von der Größenordnung θ^2 gegen ∞ , während man bei $v_{\text{ML}}(\theta)$ die Größenordnung θ hat.

Lemma 1.5.8. *Für*

$$v_{\text{ML}}(\theta) = \frac{\text{var}_{\theta}((\hat{\theta}_o)^{\text{ML}})}{\text{var}_{\theta}((\hat{\theta}_m)^{\text{ML}})}$$

gilt $\lim_{\theta \rightarrow 0} v_{\text{ML}}(\theta) = 1$ und $\lim_{\theta \rightarrow \infty} v_{\text{ML}}(\theta)/\theta = (2\zeta(3))^{-1}$.

Beweis. Im Verlauf des Beweises verwenden wir mehrfach die Regel von L'Hospital, ohne jedes Mal explizit darauf hinzuweisen. Es gilt

$$\begin{aligned} \frac{\text{var}_{\theta}((\hat{\theta}_o)^{\text{ML}})}{\text{var}_{\theta}((\hat{\theta}_m)^{\text{ML}})} &= \frac{1}{\theta^2 \frac{2\zeta(3) + \theta^2 \log(e^\theta - 1) - \theta^3 - 2\theta \text{Li}_2(e^{-\theta}) - 2\text{Li}_3(e^{-\theta})}{\theta^3} + \frac{1}{\theta(e^\theta - 1)}} \\ &= \left(\frac{2\zeta(3) + \theta^2 \log(e^\theta - 1) - \theta^3 - 2\theta \text{Li}_2(e^{-\theta}) - 2\text{Li}_3(e^{-\theta})}{\theta} + \frac{\theta}{e^\theta - 1} \right)^{-1}. \end{aligned}$$

Wir betrachten zunächst den Fall $\theta \rightarrow 0$. Für den zweiten Term in der Klammer kann

$$\lim_{\theta \rightarrow 0} \frac{\theta}{e^\theta - 1} = \lim_{\theta \rightarrow 0} \frac{1}{e^\theta} = 1$$

gezeigt werden. Beim ersten Term konvergieren sowohl Zähler als auch Nenner für $\theta \rightarrow 0$ gegen 0, u.a. wegen $\lim_{\theta \rightarrow 0} \text{Li}_s(e^{-\theta}) = \zeta(s)$, $s = 2, 3$. Dann ergibt sich

$$\begin{aligned} &\lim_{\theta \rightarrow 0} \frac{2\zeta(3) + \theta^2 \log(e^\theta - 1) - \theta^3 - 2\theta \text{Li}_2(e^{-\theta}) - 2\text{Li}_3(e^{-\theta})}{\theta} \\ &= \lim_{\theta \rightarrow 0} \frac{2\theta \log(e^\theta - 1) + \theta^2 \frac{e^\theta}{e^\theta - 1} - 3\theta^2 - 2\text{Li}_2(e^{-\theta}) - 2\theta \frac{d}{d\theta} \text{Li}_2(e^{-\theta}) - 2\frac{d}{d\theta} \text{Li}_3(e^{-\theta})}{1} \\ &= \lim_{\theta \rightarrow 0} \left(2\theta \log(e^\theta - 1) + \theta^2 \frac{e^\theta}{e^\theta - 1} - 3\theta^2 + 2\theta \sum_{k=1}^{\infty} \frac{e^{-k\theta}}{k} \right), \end{aligned}$$

dabei wurde im letzten Schritt $\frac{d}{d\theta} \text{Li}_s(e^{-\theta}) = -\sum_{k=1}^{\infty} \frac{e^{-k\theta}}{k^{s-1}} = -\text{Li}_{s-1}(e^{-\theta})$, $s = 2, 3$, benutzt. Die ersten drei Terme haben alle den Grenzwert 0. Beim vierten Term schließlich ist wegen

$$\lim_{\theta \rightarrow 0} \sum_{k=1}^{\infty} \frac{e^{-k\theta}}{k} = \sum_{k=1}^{\infty} \frac{1}{k} = \infty$$

mehr Aufwand nötig und es gilt

$$\lim_{\theta \rightarrow 0} \theta \sum_{k=1}^{\infty} \frac{e^{-k\theta}}{k} = \lim_{\theta \rightarrow 0} \frac{\sum_{k=1}^{\infty} \frac{e^{-k\theta}}{k}}{\frac{1}{\theta}} = \lim_{\theta \rightarrow 0} \theta^2 \sum_{k=1}^{\infty} e^{-k\theta} = \lim_{\theta \rightarrow 0} \left(\frac{\theta^2}{1 - e^{-\theta}} - \theta^2 \right) = \lim_{\theta \rightarrow 0} \frac{2\theta}{e^{-\theta}} = 0.$$

Insgesamt folgt dann die erste Aussage. Der zweite Grenzwert ist identisch mit

$$\lim_{\theta \rightarrow \infty} \left(2\zeta(3) + \theta^2 \log(e^\theta - 1) - \theta^3 - 2\theta \text{Li}_2(e^{-\theta}) - 2\text{Li}_3(e^{-\theta}) + \frac{\theta^2}{e^\theta - 1} \right)^{-1}.$$

Es kann

$$\lim_{\theta \rightarrow \infty} (\theta^2 \log(e^\theta - 1) - \theta^3) = 0 = \lim_{\theta \rightarrow \infty} \frac{\theta^2}{e^\theta - 1}$$

gezeigt werden, außerdem gilt $\lim_{\theta \rightarrow \infty} \text{Li}_3(e^{-\theta}) = 0$ wegen der Stetigkeit der beteiligten Funktionen. Abschließend ergibt sich

$$\lim_{\theta \rightarrow \infty} \frac{\text{Li}_2(e^{-\theta})}{\frac{1}{\theta}} = \lim_{\theta \rightarrow \infty} \frac{-\sum_{k=1}^{\infty} \frac{e^{-k\theta}}{k}}{-\frac{1}{\theta^2}} = \lim_{\theta \rightarrow \infty} \frac{-\sum_{k=1}^{\infty} e^{-k\theta}}{-\frac{2}{\theta^3}} = \lim_{\theta \rightarrow \infty} \frac{\frac{\theta^3}{2} e^{-\theta}}{1 - e^{-\theta}} = 0,$$

so dass insgesamt der gesuchte Grenzwert identisch mit $\lim_{\theta \rightarrow \infty} (2\zeta(3))^{-1} = (2\zeta(3))^{-1}$ ist, wie behauptet. \square

Die Ergebnisse lassen sich durch folgende Überlegungen anschaulich erklären. Es liegen die Todesdaten von mehreren Hundertjährigen (große mittlere Lebensdauer bei kleinen Werten von θ) vor und man weiß, dass alle am selben Tag zu verschiedenen Uhrzeiten geboren wurden. Dann ist es für die Bestimmung der Lebensdauerverteilung unwesentlich, ob die genauen Uhrzeiten bekannt sind oder nicht. Dagegen ist diese Information von sehr großer Bedeutung, wenn die einzelnen Lebensdauern sehr kurz sind (kleine mittlere Lebensdauer bei großen Werten von θ).

2. Rekonstruktion der Zuordnung

2.1. Grundlagen

Ziel von Kapitel 2 ist die Behandlung des in Beispiel 1.1.1 vorgestellten und in Abschnitt 1.2.1 formalisierten Problems, die durch eine unbekannte Permutation gegebene Zuordnung der beobachteten Daten $(x_i)_{i=1}^n$ und $(y_i)_{i=1}^n$ mittels eines festen Elements aus S'_n zu rekonstruieren. Ganz allgemein führt die Verwendung einer Konstanten a anstelle des durch die Zufallsvariable X beschriebenen exakten Wertes zu einem Fehler (oder auch Verlust) $L(X,a)$. Ist die Verteilung von X bekannt, kann a so gewählt werden, dass der resultierende mittlere Fehler $EL(X,a)$ minimiert wird. Typischerweise sind Fehlerfunktionen die Potenz einer Metrik, also z.B. $L(X,a) = d(x,a)$ oder $L(X,a) = (d(x,a))^2$ für eine Metrik d auf dem zugrundeliegenden Zustandsraum S . Es seien

$$a_1 := \arg \min_{a \in S} E d(X,a), \quad a_2 := \arg \min_{a \in S} E (d(X,a))^2,$$

dann nennen wir a_1 den *MMD-Wert* (*minimum mean distance*) und a_2 den *MMSD-Wert* (*minimum mean squared distance*). Die Bestimmung dieser Werte wird im Folgenden gelegentlich auch als Verwendung von MMD bzw. MMSD bezeichnet. Zusätzliche Informationen durch beobachtete Daten Y können in dieses Schema einbezogen werden, indem a durch eine (messbare) Funktion ϕ von Y ersetzt und dann $EL(X,\phi(Y))$ über alle ϕ minimiert wird. Wir nennen entsprechend der obigen Definition auch die so gefundene Funktion der Daten MMD- bzw. MMSD-Wert. Bekanntlich ergibt sich bei einer reellen Zufallsvariable X und Verwendung des euklidischen Abstands jeder Median der Verteilung von X als MMD-Wert und der Erwartungswert von X als MMSD-Wert. Bei Kenntnis des Wertes y einer weiteren Zufallsgröße Y ist der bedingte Erwartungswert $E[X|Y = y]$ der MMSD-Wert.

Die hier skizzierten allgemeinen Überlegungen werden im weiteren Verlauf für das Rekonstruktionsproblem konkretisiert. Dazu benutzen wir die beiden im Folgenden vorgestellten Metriken auf dem Raum der Permutationen in Verbindung mit MMD und MMSD. Der erste dieser Abstandsbegriffe wird gegeben durch die diskrete Metrik

$$d_1(x,a) := \begin{cases} 0, & x = a, \\ 1, & x \neq a. \end{cases}$$

Dann ist $d_1(\Pi, a)$ Indikator dafür, ob $a \in S_n$ die wahre Permutation ist oder nicht, und es ergibt sich bei Berücksichtigung der Daten

$$\begin{aligned} E[d_1(\Pi, a) | X_1, \dots, X_n, Y_1, \dots, Y_n] &= P(\Pi \neq a | X_1, \dots, X_n, Y_1, \dots, Y_n) \\ & (= E[(d_1(\Pi, a))^2 | X_1, \dots, X_n, Y_1, \dots, Y_n]); \end{aligned}$$

für diese Metrik führen MMD und MMSD also auf dasselbe Ziel. Der Ausdruck wird offenbar für das $a \in S_n$ minimal, welches $a \mapsto P(\Pi = a | X_1, \dots, X_n, Y_1, \dots, Y_n)$ maximiert. Als vermuteter Wert von Π wird somit der Modalwert dessen bedingter Verteilung bei Vorliegen der beobachteten Daten angegeben, also der Wert, den Π unter Berücksichtigung der vorhandenen Informationen am wahrscheinlichsten annimmt. Dasselbe Ergebnis liefert auch die sog. MAP-Methode, die wir im Folgenden kurz allgemein vorstellen.

Bei der Maximum a posteriori Methode (MAP-Methode) soll aufgrund von vorliegenden Daten, die Realisierung einer oder mehrerer Zufallsgrößen sind, eine Aussage über den Wert einer anderen, nicht direkt beobachtbaren Zufallsgröße T getroffen werden. Dabei besitze T bzgl. eines Maßes μ_1 die Dichte g und die Daten haben bei gegebenem Wert von T bzgl. des Maßes μ_2 die Dichte $f(\cdot | T)$.

Im vorliegenden Fall handelt es sich nicht wie gewohnt um eine auf die Zukunft gerichtete Aussage über eine Zufallsgröße („Prädiktion“), sondern um eine sog. Retrodiktion, die sich auf ein schon zurückliegendes Ereignis bezieht. Wir werden diese Bezeichnung auch im Rest der Arbeit verwenden, wenn es um die Lösung des Rekonstruktionsproblems geht.

Die MAP-Retrodiktion für T ist der Wert, den T bei gegebener Realisierung x des Datenvektors am wahrscheinlichsten annimmt, also der Modalwert der a posteriori Dichte $f(t|x)$. Diese Dichte ergibt sich mittels des Satzes von Bayes aus $f(x|t)$ und $g(t)$ zu

$$f(t|x) = \frac{f(x|t)g(t)}{\int f(x|t')P^T(dt')} = \frac{f(x|t)g(t)}{\int f(x|t')g(t')\mu_1(dt')}. \quad (2.1)$$

Da der Nenner nicht von t abhängt, setzen wir

$$\hat{T}_{\text{MAP}} := \arg \max_t f(x|t)g(t)$$

bzw.

$$\hat{T}_{\text{MAP}} := \arg \max_t f(x, t), \quad (2.2)$$

falls die gemeinsame Dichte von X und T bzgl. $\mu_1 \otimes \mu_2$ vorliegt. Wie sich im Verlauf dieses Kapitels zeigen wird, muss die MAP-Retrodiktion nicht unbedingt eindeutig sein.

Das beschriebene Vorgehen bzw. die auftretenden Rechnungen haben gewisse Ähnlichkeiten zur Maximum-Likelihood-Methode (ML-Methode). Bei dieser handelt es sich allerdings um ein Verfahren zum Schätzen eines unbekanntem Parameters einer Verteilung, während sich die MAP-Methode auf die Werte von Zufallsvariablen an sich bezieht. Für eine detailliertere Untersuchung der Gemeinsamkeiten und Unterschiede von MAP- und ML-Methode, insbesondere auch für einen Fall, in dem beide Methoden dasselbe Ergebnis liefern, siehe man z.B. [Sor80, Kap. 5]. Dort findet sich auch eine umfangreichere als die hier gegebene Einführung in die MAP-Methode.

Da die Permutation Π eine diskrete Zufallsgröße ist, entspricht $f(t|x)$ aus (2.1) beim Rekonstruktionsproblem der bedingten Wahrscheinlichkeit $P(\Pi = \cdot | X_1, \dots, X_n, Y_1, \dots, Y_n)$. Weiter ergibt sich aus den der genannten Gleichung folgenden Ausführungen

$$\begin{aligned} P(\Pi = \pi | X_1 = x_1, \dots, X_n = x_n, Y_1 = y_1, \dots, Y_n = y_n) \\ = c \cdot P(\Pi = \pi, X_1 = x_1, \dots, X_n = x_n, Y_1 = y_1, \dots, Y_n = y_n), \end{aligned}$$

wobei c eine von π unabhängige Konstante ist (die aber durchaus von x_1, \dots, x_n und y_1, \dots, y_n abhängen kann) und somit bei der Maximierung keine Rolle spielt. MMD und MMSD in Verbindung mit d_1 liefern also, wie oben angekündigt, dasselbe Ergebnis wie die MAP-Methode. Deshalb werden im Folgenden alle auf den beiden Kombinationen MMD bzw. MMSD mit d_1 basierenden Operationen zur Abkürzung unter dem Oberbegriff MAP zusammengefasst. Benutzung der Bezeichnungen MMD und MMSD impliziert daher von nun an die Verwendung der zweiten in dieser Arbeit benutzten Metrik, welche nachfolgend eingeführt wird. Für eine genauere Beschreibung sowie eine ausführlichere Einführung in Metriken auf Gruppen und deren Anwendung in der Stochastik siehe [Dia88, Kap. 6]. Dort findet sich insbesondere auf Seite 129f. eine Aufstellung von wünschenswerten Eigenschaften, die mit einer Empfehlung der in der nächsten Definition genannten Metrik abschließt.

Definition 2.1.1. Es seien $\sigma, \pi \in S_n$ zwei Permutationen. Wir definieren *Kendalls Tau-Abstand* $d_\tau(\sigma, \pi)$ durch

$$d_\tau(\sigma, \pi) := \sum_{i=1}^{n-1} \sum_{j=i+1}^n D_{ij}$$

mit

$$D_{ij} := \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\pi(i) < \pi(j)\}} + \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\pi(i) > \pi(j)\}},$$

d.h. man prüft für alle Paare (i, j) mit $1 \leq i < j \leq n$, ob $\sigma(i) - \sigma(j)$ und $\pi(i) - \pi(j)$ dasselbe Vorzeichen haben, und zählt die Fälle, in denen dies nicht der Fall ist.

Allgemein sind in der Literatur mehrere Definitionen bzw. Berechnungsformen von Kendalls Tau-Abstand verbreitet; siehe dazu den Anhang A.

| i | j | $\sigma(i)$ | $\sigma(j)$ | $\pi(i)$ | $\pi(j)$ | zu zählen |
|-----|-----|-------------|-------------|----------|----------|-----------|
| 1 | 2 | 2 | 4 | 1 | 3 | × |
| 1 | 3 | 2 | 3 | 1 | 4 | × |
| 1 | 4 | 2 | 1 | 1 | 2 | ✓ |
| 2 | 3 | 4 | 3 | 3 | 4 | ✓ |
| 2 | 4 | 4 | 1 | 3 | 2 | × |
| 3 | 4 | 3 | 1 | 4 | 2 | × |

Tabelle 2.1. Ausführliche Darstellung der Bestimmung von Kendalls Tau-Abstand in Teil (a) von Beispiel 2.1.3

Bemerkung 2.1.2. Unter der Bezeichnung „Kendall’s tau“ findet sich in der Literatur sowohl ein Rangkorrelationskoeffizient mit Werten in $[-1,1]$ als auch ein Abstandsmaß auf dem Raum der Permutationen. Kendall beschrieb den Koeffizienten ausführlich in [Ken70] und wies dabei auf die zweite Verwendungsmöglichkeit hin, die dann auch eigenständig Verwendung fand. In dieser Arbeit wird ausschließlich das Abstandsmaß verwendet und, um Verwechslungen zu vermeiden, sprechen wir von Kendalls Tau-Abstand. \triangleleft

Wir geben nun zwei Möglichkeiten an, wie man die äquivalente Formulierung aus Definition A.1 (b) in Anhang A für einen Algorithmus verwenden kann. Zum einen können natürlich alle Paare (i,j) mit $1 \leq i < j \leq n$ einzeln betrachtet werden (z.B. systematisch nach dem Wert von j) und man zählt, wie oft die im Satz genannte Bedingung erfüllt ist. Eine zweite, graphisch sehr anschauliche Möglichkeit wird in [Gri58] beschrieben. Dazu schreibt man π und σ übereinander, ordnet dann die Spalten so um, dass die erste Zeile in aufsteigender Reihenfolge steht, und verbindet gleiche Zahlen jeweils durch eine Strecke. Kendalls Tau-Abstand entspricht nun der Anzahl von Überkreuzungen von Strecken, da für jedes Paar (a_i, a_j) von Einträgen in der zweiten Zeile mit $i < j$ und $a_i > a_j$ genau eine Überkreuzung entsteht.

Beispiel 2.1.3. Es sei $n = 4$, $\sigma := \begin{pmatrix} 1234 \\ 2431 \end{pmatrix}$ und $\pi := \begin{pmatrix} 1234 \\ 1342 \end{pmatrix}$. Wir betrachten nun die verschiedenen in Anhang A beschriebenen Berechnungsformen von Kendalls Tau-Abstand.

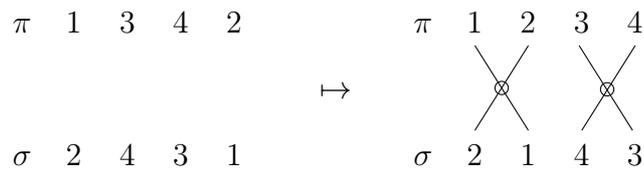
- (a) Wir müssen für alle Paare (i,j) mit $1 \leq i < j \leq n$ prüfen, ob $\sigma(i) - \sigma(j)$ und $\pi(i) - \pi(j)$ das gleiche Vorzeichen haben oder nicht. Für $i = 1, j = 2$ erhält man $\sigma(i) = 2, \sigma(j) = 4$ und $\pi(i) = 1, \pi(j) = 3$, dies führt in beiden Fällen auf die Differenz -2 , so dass dieses Paar nicht gezählt werden muss. Eine detaillierte Betrachtung findet sich in Tabelle 2.1, wir erhalten also $d_{\tau,1}(\sigma, \pi) = 2$.
- (b) Für die erste oben genannte algorithmische Vorgehensweise zur Bestimmung von $d_{\tau,2}(\sigma, \pi)$ müssen wir für alle möglichen Paare (i,j) mit $1 \leq i < j \leq n$ prüfen, ob

| j | i | $\pi^{-1}(i)$ | $\pi^{-1}(j)$ | $\sigma(\pi^{-1}(i))$ | $\sigma(\pi^{-1}(j))$ | zu zählen |
|-----|-----|---------------|---------------|-----------------------|-----------------------|-----------|
| 1 | — | — | — | — | — | — |
| 2 | 1 | 1 | 4 | 2 | 1 | ✓ |
| 3 | 1 | 1 | 2 | 2 | 4 | × |
| 3 | 2 | 4 | 2 | 1 | 4 | × |
| 4 | 1 | 1 | 3 | 2 | 3 | × |
| 4 | 2 | 4 | 3 | 1 | 3 | × |
| 4 | 3 | 2 | 3 | 4 | 3 | ✓ |

Tabelle 2.2. Ausführliche Darstellung der Bestimmung von Kendalls Tau-Abstand in Teil (b) von Beispiel 2.1.3

$\sigma(\pi^{-1}(i)) > \sigma(\pi^{-1}(j))$ gilt. Motiviert durch eine Darstellung von d_τ im Beweis von Satz 2.3.23 weiter unten gehen wir die Paare systematisch nach dem Wert von j durch, an dieser Stelle ist dies aber völlig willkürlich. Für $j = 1$ gibt es keinen zugelassenen Wert von i , also beginnen wir mit $j = 2$ und $i = 1$. Dann ist $\pi^{-1}(i) = 1$ und $\pi^{-1}(j) = 4$ sowie $\sigma(\pi^{-1}(i)) = 2$ und $\sigma(\pi^{-1}(j)) = 1$. Wegen $2 > 1$ wird das Paar gezählt. So verfährt man mit den übrigen Paaren, die Ergebnisse sind in Tabelle 2.2 zusammengefasst. Es gibt also zwei Paare, die zu zählen sind, dies entspricht Kendalls Tau-Abstand.

- (c) Durch die zweite algorithmische Vorgehensweise können wir Kendalls Tau-Abstand auch aus der folgenden Zeichnung ablesen (Überkreuzungen sind mit einem Kreis markiert). Zunächst werden die beiden Permutationen übereinander geschrieben, dann nach der ersten Zeile geordnet und gleiche Zahlen verbunden.



Man sieht an diesem Beispiel auch, dass die Anordnung nach der ersten Zeile wichtig ist, denn in der linken Abbildung gäbe es mehr Überkreuzungen als in der rechten. In der ersten Zeile steht nun die Identität, in der zweiten $\sigma \circ \pi^{-1}$.

- (d) Nach der dritten Berechnungsweise folgt wegen $\pi^{-1} = \begin{pmatrix} 1234 \\ 1423 \end{pmatrix}$ und $\sigma^{-1} = \begin{pmatrix} 1234 \\ 4132 \end{pmatrix}$ schließlich $d_{\tau,3}(\sigma, \pi) = 2$ (um die eine Inverse in die andere umzustellen, müssen 1 und 4 sowie 2 und 3 vertauscht werden, man benötigt also zwei Nachbartranspositionen). \triangleleft

2.2. Rekonstruktion mit der MAP-Methode

2.2.1. Problemstellung

In diesem Abschnitt wird die korrekte Zuordnung der Geburts- und Todesdaten mittels MAP betrachtet. Dazu kann nach den einführenden Bemerkungen die gemeinsame Dichte von geordneten Daten und den einzelnen Permutationen bzgl. eines dominierenden Maßes verwendet werden, genauer ergibt sich aus (2.2)

$$\hat{\Pi}_{\text{MAP}} := \arg \max_{\pi \in S'_n} f_{X,Y,\Pi}(x,y,\pi).$$

Unter Benutzung von $f_{X,Y,\Pi}(x,y,\pi) = f_{Y,\Pi|X}(y,\pi|x) \cdot f_X(x)$ formulieren wir dies so um, dass die Dichte der (X_i) bei der Bestimmung des Argmax keine Rolle spielt. Die Verteilung der Geburtszeitpunkte hat also keinen Einfluss auf die Wahl, deshalb nehmen wir diese im Folgenden als gegebene, feste Werte x_1, \dots, x_n an und schreiben $f_{Y,\Pi}(y,\pi)$ für die bedingte Dichte $f_{Y,\Pi|X}(y,\pi|x)$.

Satz 2.2.1. *Bei Vorliegen von Geburtsdaten x_1, \dots, x_n ist*

$$f_{Y,\Pi}(y_1, \dots, y_n, \pi) = \begin{cases} \prod_{i=1}^n f_Z(y_{\pi(i)} - x_i), & y_1 < \dots < y_n, \pi \in S'_n, \\ 0, & \text{sonst} \end{cases}$$

eine gemeinsame Dichte bezüglich $\ell^n \otimes \sum_{\pi \in S_n} \delta_\pi$ für den Vektor der Ordnungsstatistiken $Y = (Y_1, \dots, Y_n)$ und die korrekte Zuordnung Π .

Beweis. Wir erinnern daran, dass $Y = (Y_1, \dots, Y_n)$ der Vektor der Ordnungsstatistiken und $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_n)$ der Vektor der ursprünglichen, ungeordneten Werte ist. Hat Π den Wert π , besteht zwischen diesen der Zusammenhang

$$(Y_{\pi(1)}, \dots, Y_{\pi(n)}) = (\tilde{Y}_1, \dots, \tilde{Y}_n) \quad \text{bzw.} \quad (\tilde{Y}_{\pi^{-1}(1)}, \dots, \tilde{Y}_{\pi^{-1}(n)}) = (Y_1, \dots, Y_n).$$

Bezeichnen weiter f_U und P^U Lebesgue-Dichte bzw. Verteilung eines Zufallsvektors U , dann gilt für jedes $A = A_1 \times \dots \times A_n \in \mathfrak{B}^n$ und $\pi \in S'_n$

$$\begin{aligned} & P(\Pi = \pi, Y \in A) \\ &= P(\tilde{Y}_{\pi^{-1}(1)} < \dots < \tilde{Y}_{\pi^{-1}(n)}, \tilde{Y}_{\pi^{-1}(1)} \in A_1, \dots, \tilde{Y}_{\pi^{-1}(n)} \in A_n) \\ &= \int_{\substack{\{(y_1, \dots, y_n) \in \mathbb{R}^n: \\ y_1 < \dots < y_n, y_1 \in A_1, \dots, y_n \in A_n\}}} dP^{(\tilde{Y}_{\pi^{-1}(1)}, \dots, \tilde{Y}_{\pi^{-1}(n)})} \\ &= \int_{-\infty}^{\infty} \int_{y_1}^{\infty} \dots \int_{y_{n-1}}^{\infty} f_{(\tilde{Y}_{\pi^{-1}(1)}, \dots, \tilde{Y}_{\pi^{-1}(n)})}(y_1, \dots, y_n) \cdot \mathbf{1}_A(y_1, \dots, y_n) dy_n \dots dy_1. \quad (2.3) \end{aligned}$$

Da die \tilde{Y} -Werte unabhängig und die Lebensdauern i.i.d. sind, folgt weiter

$$f_{(\tilde{Y}_{\pi^{-1}(1)}, \dots, \tilde{Y}_{\pi^{-1}(n)})}(y_1, \dots, y_n) = \prod_{i=1}^n f_{\tilde{Y}_{\pi^{-1}(i)}}(y_i) = \prod_{i=1}^n f_Z(y_i - x_{\pi^{-1}(i)}).$$

Es sei nun $\tilde{f}_{Y,\Pi}(y_1, \dots, y_n, \pi) = \prod_{i=1}^n f_Z(y_i - x_{\pi^{-1}(i)})$ für $y_1 < \dots < y_n, \pi \in S'_n$ und 0 sonst. Dann ist der Ausdruck in (2.3) für $y = (y_1, \dots, y_n)$ identisch mit

$$\int_{\mathbb{R}^n} \tilde{f}_{Y,\Pi}(y, \pi) \cdot 1_A(y) dy = \int_A \tilde{f}_{Y,\Pi}(y, \pi) dy.$$

Da sich $\tilde{f}_{Y,\Pi}$ und das oben genannte $f_{Y,\Pi}$ nur in der Reihenfolge beim Bilden des Produkts unterscheiden, folgt nun die Behauptung. \square

Insgesamt ergibt sich also

$$\hat{\Pi}_{\text{MAP}} = \arg \max_{\pi \in S'_n} \prod_{i=1}^n f_Z(y_{\pi(i)} - x_i) \quad (2.4)$$

als MAP-Retrodiktion der gesuchten Permutation. Zur Lösung dieses Problems stellen wir nun einen Zusammenhang zu einer anderen, in der Literatur schon umfangreich behandelten Fragestellung her. Logarithmieren des Produkts in (2.4) führt auf

$$\log \left(\prod_{i=1}^n f_Z(y_{\pi(i)} - x_i) \right) = \sum_{i=1}^n \log(f_Z(y_{\pi(i)} - x_i)) \quad (2.5)$$

und lässt die Lage des Maximums unverändert. Nach dieser Umformung kann die gesuchte Permutation mit Mitteln der kombinatorischen Optimierung bestimmt werden, genauer besteht große Ähnlichkeit zu einem sog. linearen Assignment-Problem (LAP) in der Maximierungs-Variante mit Gewichtungsmatrix $A = (a_{ij})_{i,j=1,\dots,n}$. Bei unserem Problem gilt $a_{ij} = \log(f_Z(y_j - x_i))$ für eine zulässige Zuordnung i zu j und zur Berücksichtigung der unzulässigen Zuordnungen müssen Abwandlungen gegenüber dem üblichen Vorgehen vorgenommen werden.

Im klassischen LAP geht es für $n \in \mathbb{N}$ darum, n Aufgaben auf n Personen zu verteilen. Übernimmt Person i die Aufgabe j , $i, j = 1, \dots, n$, führt dies zu den Kosten $a_{ij} \in \mathbb{R}^+$. Nun ist jeder Aufgabe bijektiv eine Person zuzuteilen, so dass die Summe der entstehenden Kosten minimal wird, d.h. es ist das Argmin von $S_n \ni \pi \mapsto \sum_{i=1}^n a_{i\pi(i)}$ zu bestimmen. In der Maximierungs-Variante betrachtet man entsprechend Gewinne, deren Summe dann zu maximieren ist. Für eine detailliertere Einführung in kombinatorische Optimierung allgemein und speziell das Assignment-Problem siehe [Jun99] bzw. [Bur02]. Mit obiger Notation entspricht das LAP dem linearen Programm

$$\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_{ij} \rightarrow \max / \min$$

unter den Bedingungen

$$\sum_{k=1}^n x_{ik} = 1, \quad \sum_{k=1}^n x_{kj} = 1, \quad x_{ij} \in \{0,1\}, \quad i,j = 1, \dots, n,$$

dabei soll Person i genau dann die Aufgabe j übernehmen, wenn $x_{ij} = 1$ ist. Die ersten beiden Bedingungen sorgen dafür, dass es sich bei $(x_{ij})_{i,j=1,\dots,n}$ um eine Permutationsmatrix handelt, die Lösung also Element von S_n ist.

Das vorliegende Zuordnungsproblem führt wie oben schon teilweise ausgeführt auf eine Formulierung in der Maximierungs-Variante, allerdings sind im Gegensatz zum allgemeinen LAP nur Permutationen aus S'_n als Lösung zugelassen. Eine Übertragung dieser Einschränkung in die klassische Fragestellung ist möglich, indem $a_{ij} := -\infty$ (ungünstigster Wert bei einem Maximierungsproblem) für Zuordnungen i zu j mit $y_j - x_i \leq 0$ definiert wird. Dabei ist ∞ ggf. durch eine entsprechend große Konstante M zu ersetzen, so dass Kombinationen mit Gewichtung $-M$ nicht Bestandteil einer optimalen Lösung sein können. Dies leistet z.B. $M = n \cdot \max_{i,j} |a_{ij}| + 1$, wobei das Maximum über alle zulässigen Paare (i,j) genommen wird. Das entsprechende Problem ist nach [Jun99, S. 389f.] mit Zeitkomplexität $O(n^3)$ lösbar. Im statistischen Programmpaket R ([R D09]) gibt es für das LAP die Funktion `lp.assign` aus dem Paket `lpSolve`; ferner sind dort Algorithmen für weitere Spezialisierungen des klassischen Problems verfügbar, bei denen Zuordnungen von vorneherein ausgeschlossen werden können, hierzu siehe z.B. [Han04] und das darin genannte R-Paket `optmatch` mit der Funktion `fullmatch`. Sind negative Gewichte, die beim Logarithmieren der Dichten entstehen können, nicht zugelassen, muss ggf. eine hinreichend große Konstante zu allen Gewichten addiert werden.

Im Folgenden führen wir die beschriebenen Schritte für einige der Komponistendaten und unter der Annahme vor, dass die Lebensdauern aus einer inversen Gauß-Verteilung stammen. Für diese ist durch $f_Z(z) = (\lambda/2\pi z^3)^{\frac{1}{2}} \exp(-\lambda(z - \mu)^2/(2\mu^2 z))$ für $z > 0$ und 0 sonst eine Dichte mit den Parametern $\mu, \lambda > 0$ gegeben. Abkürzend schreiben wir auch $Z_1 \sim \text{IG}(\mu, \lambda)$.

Beispiel 2.2.2. Um einerseits das beschriebene Vorgehen zu illustrieren, andererseits aber auch den Platzbedarf in einem angemessenen Rahmen zu halten, verwenden wir in diesem Beispiel elf der 22 Geburtsdaten aus Abb. 1.1 und die zugehörigen Todesdaten; genauer handelt es sich um die Daten an den Positionen 5, 6, 7, 8, 14, 15, 16, 17, 20, 21 und 22. Zunächst wird ohne Berücksichtigung von Zulässigkeit für alle Paare $(i,j) \in \{1, \dots, 11\}^2$ die theoretische Lebensdauer $y_j - x_i$ bei Zuordnung von Geburtstag x_i zu Todesdatum y_j bestimmt (siehe Tabelle 2.3). Da bei der Rekonstruktion f_Z als bekannt vorausgesetzt wird, unterstellen wir nun eine inverse Gauß-Verteilung für die Lebensdauern der Komponisten, und bestimmen die zugehörigen Parameter durch Anwendung der entsprechenden ML-Schätzer aus Anhang B auf die wahren Lebensdauern.

| | j | | | | | | | | | | |
|-------|-------|-------|-------|-------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 0.86 | 1.11 | 1.36 | 1.37 | 2.33 | 2.36 | 2.38 | 2.47 | 2.64 | 2.65 | 2.68 |
| 2 | 0.52 | 0.77 | 1.01 | 1.02 | 1.98 | 2.02 | 2.03 | 2.13 | 2.29 | 2.31 | 2.34 |
| 3 | 0.30 | 0.55 | 0.80 | 0.81 | 1.77 | 1.81 | 1.82 | 1.91 | 2.08 | 2.09 | 2.12 |
| 4 | 0.07 | 0.32 | 0.57 | 0.58 | 1.54 | 1.58 | 1.59 | 1.68 | 1.85 | 1.86 | 1.89 |
| 5 | -0.40 | -0.15 | 0.10 | 0.11 | 1.07 | 1.10 | 1.12 | 1.21 | 1.38 | 1.39 | 1.42 |
| i 6 | -0.46 | -0.21 | 0.04 | 0.05 | 1.01 | 1.04 | 1.06 | 1.15 | 1.32 | 1.33 | 1.36 |
| 7 | -0.60 | -0.34 | -0.10 | -0.09 | 0.87 | 0.91 | 0.92 | 1.01 | 1.18 | 1.20 | 1.22 |
| 8 | -0.70 | -0.45 | -0.20 | -0.19 | 0.77 | 0.81 | 0.82 | 0.91 | 1.08 | 1.09 | 1.12 |
| 9 | -1.02 | -0.77 | -0.52 | -0.51 | 0.45 | 0.49 | 0.50 | 0.59 | 0.76 | 0.77 | 0.80 |
| 10 | -1.15 | -0.90 | -0.66 | -0.64 | 0.31 | 0.35 | 0.36 | 0.46 | 0.62 | 0.64 | 0.67 |
| 11 | -1.17 | -0.92 | -0.67 | -0.66 | 0.30 | 0.33 | 0.35 | 0.44 | 0.61 | 0.62 | 0.65 |

Tabelle 2.3. (Theoretische) Lebensdauern im Komponistenbeispiel in $25 \cdot 10^3$ Tagen bei Paarung von Todesdatum y_j mit Geburtstag x_i , $(i,j) \in \{1, \dots, 11\}^2$, wobei ein Jahr aus 12 Monaten zu je 30 Tagen besteht

Dies führte zu den Werten $\mu \approx 0.82$ und $\lambda \approx 11$. Die Einträge der Gewichtungsmatrix A (siehe Tabelle 2.4) ergeben sich als Bilder der Lebensdauern in Tabelle 2.3 unter der Dichtefunktion einer $IG(0.82, 11)$ -Verteilung und anschließendes Logarithmieren aller positiven Werte, während Einträge 0 durch $-M$ ersetzt werden. Wählt man noch M wie oben beschrieben, kann das LAP in der Maximierungs-Variante z.B. mit einem der oben genannten Pakete gelöst werden und wir erhalten $\hat{\Pi}_{\text{MAP}} = \text{id}$. Ein zeilenweiser Vergleich des jeweiligen Eintrags auf der Hauptdiagonale von A mit den anderen Werten macht diese Lösung zumindest plausibel, obwohl $a_{ii}, i = 1, \dots, 11$, nicht immer das Zeilenmaximum ist. Die Identität führt auf die Lebensdauern (in Jahren zu 360 Tagen)

59.68, 53.34, 55.47, 40.36, 74.15, 72.44, 63.91, 63.36, 52.68, 44.35, 45.02,

dagegen ergibt die korrekte Zuordnung $\Pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 2 & 1 & 4 & 3 & 6 & 8 & 7 & 5 & 11 & 9 & 10 \end{pmatrix}$ die Lebensdauern

77.17, 35.86, 56.28, 39.55, 76.69, 79.77, 63.91, 53.45, 55.59, 43.31, 43.14.

In acht von elf Fällen liegt der relative Fehler der durch MAP angegebenen Lebensdauer bei unter 10%, bei sechs Komponisten sind es sogar weniger als 5%. Weiter ist $d_\tau(\Pi, \hat{\Pi}_{\text{MAP}}) = 8$, während der maximal mögliche Kendalls Tau-Abstand 31 beträgt. \triangleleft

Zur Motivation des weiteren Vorgehens geben wir im folgenden Lemma eine alternative Formulierung des Optimierungsproblems bei einer inversen Gauß-Verteilung an.

| | j | | | | | | | | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1 | 0.49 | -0.51 | -1.93 | -2.01 | -9.03 | -9.32 | -9.42 | -10.2 | -11.5 | -11.6 | -11.8 |
| 2 | -0.18 | 0.65 | -0.05 | -0.01 | -6.38 | -6.66 | -6.76 | -7.47 | -8.77 | -8.89 | -9.10 |
| 3 | -5.17 | 0.13 | 0.61 | 0.59 | -4.78 | -5.05 | -5.15 | -5.84 | -7.11 | -7.22 | -7.43 |
| 4 | -58.6 | -4.20 | 0.23 | 0.30 | -3.15 | -3.41 | -3.49 | -4.15 | -5.37 | -5.49 | -5.69 |
| 5 | $-M$ | $-M$ | -40.3 | -34.5 | -0.30 | -0.48 | -0.54 | -1.05 | -2.07 | -2.16 | -2.34 |
| i 6 | $-M$ | $-M$ | -135 | -98.0 | -0.02 | -0.18 | -0.24 | -0.71 | -1.68 | -1.77 | -1.94 |
| 7 | $-M$ | $-M$ | $-M$ | $-M$ | 0.46 | 0.35 | 0.31 | -0.05 | -0.88 | -0.97 | -1.12 |
| 8 | $-M$ | $-M$ | $-M$ | $-M$ | 0.65 | 0.60 | 0.58 | 0.34 | -0.36 | -0.43 | -0.57 |
| 9 | $-M$ | $-M$ | $-M$ | $-M$ | -1.00 | -0.50 | -0.36 | 0.36 | 0.66 | 0.64 | 0.61 |
| 10 | $-M$ | $-M$ | $-M$ | $-M$ | -4.61 | -3.26 | -2.87 | -0.90 | 0.49 | 0.54 | 0.60 |
| 11 | $-M$ | $-M$ | $-M$ | $-M$ | -5.41 | -3.85 | -3.41 | -1.17 | 0.42 | 0.48 | 0.57 |

Tabelle 2.4. Gewichtungsmatrix A im Komponistenbeispiel

Lemma 2.2.3. Sind die Lebensdauern $IG(\mu, \lambda)$ -verteilt, dann ist

$$S'_n \ni \pi \mapsto \sum_{i=1}^n \log(f_Z(y_{\pi(i)} - x_i)) \rightarrow \max$$

äquivalent zu

$$S'_n \ni \pi \mapsto \sum_{i=1}^n \left(3 \log(y_{\pi(i)} - x_i) + \frac{\lambda}{y_{\pi(i)} - x_i} \right) \rightarrow \min \quad (2.6)$$

in dem Sinn, dass beide Ausdrücke für dasselbe π ihr Optimum annehmen.

Beweis. Einsetzen der entsprechenden Dichte f_Z in (2.5) ergibt für festes $\pi \in S'_n$

$$\begin{aligned} & \sum_{i=1}^n \log(f_Z(y_{\pi(i)} - x_i)) \\ &= \sum_{i=1}^n \log \left(\left(\frac{\lambda}{2\pi(y_{\pi(i)} - x_i)^3} \right)^{\frac{1}{2}} \exp \left(\frac{-\lambda(y_{\pi(i)} - x_i - \mu)^2}{2\mu^2(y_{\pi(i)} - x_i)} \right) \right) \\ &= \frac{n}{2} \log \left(\frac{\lambda}{2\pi} \right) - \frac{3}{2} \sum_{i=1}^n \log(y_{\pi(i)} - x_i) - \frac{\lambda}{2\mu^2} \sum_{i=1}^n \frac{(y_{\pi(i)} - x_i - \mu)^2}{y_{\pi(i)} - x_i}. \end{aligned}$$

Der erste Term ist unabhängig von der gewählten Permutation und für den dritten gilt

$$\sum_{i=1}^n \frac{(y_{\pi(i)} - x_i - \mu)^2}{y_{\pi(i)} - x_i} = \sum_{i=1}^n \left((y_{\pi(i)} - x_i) - 2\mu + \frac{\mu^2}{y_{\pi(i)} - x_i} \right).$$

Damit hat die Wahl der Permutation nur Einfluss auf

$$-\frac{1}{2} \sum_{i=1}^n \left(3 \log(y_{\pi(i)} - x_i) + \frac{\lambda}{y_{\pi(i)} - x_i} \right),$$

dieser Ausdruck ist zu maximieren. Wegen des negativen Vorfaktors ist dies äquivalent zu dem in der Behauptung genannten Problem. \square

Der zu optimierende Ausdruck ist also unabhängig von einem der beiden Parameter, nämlich dem Erwartungswert μ . Außerdem treten sowohl eine (konkave) Funktion mit positiver Ableitung als auch eine (konvexe) Funktion mit negativer Ableitung auf, dadurch führen große Werte der einen Funktion zu kleinen Werten der anderen Funktion und umgekehrt. Würde dagegen nur eine der beiden Funktionen vorkommen, ergäbe sich dieser Effekt nicht. Wir werden am Ende des folgenden Abschnitts noch einmal an diese Beobachtung erinnern; dort spielen auch Konvexität und Konkavität eine wichtige Rolle.

2.2.2. Explizite Lösungen

Auch wenn es sich bei MAP unter komplexitätstheoretischen Gesichtspunkten um ein einfaches Problem handelt, kann es nützlich sein, eine explizite Lösung vorliegen zu haben. Um dies unter bestimmten Umständen zu erreichen, benötigen wir zunächst einige Hilfsmittel. Dabei bezeichnen wir für eine Permutation $\sigma \in S_n$ ein Paar (i, j) mit $1 \leq i < j \leq n$ und $\sigma(i) > \sigma(j)$ als *Inversion*. Wir sprechen beim Übergang von σ zu einer Permutation σ' mit $\sigma'(i) := \sigma(j)$, $\sigma'(j) := \sigma(i)$ und $\sigma'(l) := \sigma(l)$ für alle $l \neq i, j$ davon, dass die Inversion (i, j) aufgelöst wird. Der umgekehrte Vorgang wird Konstruktion dieser Inversion genannt.

Lemma 2.2.4. *Es sei $\sigma \in S'_n$ und σ besitze eine Inversion (i, j) . Dann ist auch die Permutation zulässig, die durch Auflösen dieser Inversion entsteht. Eine entsprechende Aussage bei Konstruktion einer beliebigen Inversion gilt im Allgemeinen nicht.*

Beweis. Wegen $\sigma \in S'_n$ ist in der beschriebenen Situation $x_i \leq x_j < y_{\sigma(j)} \leq y_{\sigma(i)}$. Damit sind auch die umgekehrten Zuordnungen $(i, \sigma(j))$ und $(j, \sigma(i))$ zulässig und die erste Behauptung folgt. Gegenbeispiel für den Beweis der zweiten Behauptung ist z.B. der Fall $S'_n = \{\text{id}\}$. \square

Lemma 2.2.5. *Für die Daten $(x_t)_{t=1}^n$ und $(y_t)_{t=1}^n$ und jedes Quadrupel (x_i, x_j, y_k, y_l) mit $x_i \leq x_j < y_k \leq y_l$ sei das Tripel $(a_{ijkl}, b_{ijkl}, c_{ijkl})$ jeweils definiert durch $a_{ijkl} := x_j - x_i$, $b_{ijkl} := y_k - x_j$ und $c_{ijkl} := y_l - y_k$. Weiter sei*

$$g(x, y, z) := \frac{f_Z(x + y) \cdot f_Z(y + z)}{f_Z(y) \cdot f_Z(x + y + z)}.$$

Ist $g(a_{ijkl}, b_{ijkl}, c_{ijkl}) \geq 1$ für alle diese Tripel, wird der Wert von $\sum_{i=1}^n \log(f(y_{\sigma(i)} - x_i))$ für jedes $\sigma \in S'_n$ durch Auflösen einer Inversion nicht vermindert. Gilt die umgekehrte Ungleichung, wird der Wert nicht vergrößert. Ist die jeweilige Ungleichung strikt und sind die Daten paarweise verschieden, wird der Wert sogar echt vergrößert bzw. verkleinert.

Beweis. Es ergebe sich σ' aus σ durch Auflösen der Inversion (i, j) , dann ist σ' nach Lemma 2.2.4 zulässig. Weiter seien $l := \sigma(i) = \sigma'(j)$, $k := \sigma(j) = \sigma'(i)$ sowie $a := x_j - x_i$, $0 < b := y_k - x_j$, $c := y_l - y_k$. Nun gilt

$$\sum_{i=1}^n \log(f_Z(y_{\sigma'(i)} - x_i)) - \sum_{i=1}^n \log(f_Z(y_{\sigma(i)} - x_i)) = \log\left(\frac{f_Z(a+b) \cdot f_Z(b+c)}{f_Z(b) \cdot f_Z(a+b+c)}\right).$$

und es folgt die Behauptung. \square

Lemma 2.2.6. *Es seien σ eine Permutationen aus S_n , bei der eine Inversion (k, l) vorliegt, und σ' die Permutation, die aus σ durch Auflösen dieser Inversion entsteht. Dann gilt*

$$d_\tau(\text{id}, \sigma) > d_\tau(\text{id}, \sigma').$$

Beweis. Nach Anhang A ist

$$d_\tau(\text{id}, \sigma) = \#\{(i, j) : 1 \leq i < j \leq n : \sigma(i) > \sigma(j)\} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}}$$

und entsprechend

$$d_\tau(\text{id}, \sigma') = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}}.$$

Um nun die Differenz $d_\tau(\text{id}, \sigma) - d_\tau(\text{id}, \sigma')$ berechnen zu können, zerlegen wir diese Summen in Hinblick auf k und l . Wegen

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n = \sum_{i=1}^{k-1} \sum_{j=i+1}^n + \sum_{i=k}^k \sum_{j=i+1}^n + \sum_{i=k+1}^{l-1} \sum_{j=i+1}^n + \sum_{i=l}^l \sum_{j=i+1}^n + \sum_{i=l+1}^n \sum_{j=i+1}^n$$

unterscheiden wir die folgenden fünf Fälle, wobei wie üblich eine Summe als 0 definiert wird, wenn die obere Grenze kleiner als die untere ist. Für σ erhält man

$$\sum_{i=1}^{k-1} \sum_{j=i+1}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} = \sum_{i=1}^{k-1} \left(\sum_{\substack{j=i+1 \\ j \neq k, l}}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} + \mathbf{1}_{\{\sigma(i) > \sigma(k)\}} + \mathbf{1}_{\{\sigma(i) > \sigma(l)\}} \right), \quad (2.7)$$

$$\sum_{i=k}^k \sum_{j=i+1}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} = \sum_{\substack{j=k+1 \\ j \neq l}}^n \mathbf{1}_{\{\sigma(k) > \sigma(j)\}} + \mathbf{1}_{\{\sigma(k) > \sigma(l)\}}, \quad (2.8)$$

$$\sum_{i=k+1}^{l-1} \sum_{j=i+1}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} = \sum_{i=k+1}^{l-1} \left(\sum_{\substack{j=i+1 \\ j \neq l}}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} + \mathbf{1}_{\{\sigma(i) > \sigma(l)\}} \right), \quad (2.9)$$

$$\sum_{i=l}^l \sum_{j=i+1}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} = \sum_{j=l+1}^n \mathbf{1}_{\{\sigma(l) > \sigma(j)\}}, \quad (2.10)$$

und für σ' ergibt sich

$$\begin{aligned} \sum_{i=1}^{k-1} \sum_{j=i+1}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}} &= \sum_{i=1}^{k-1} \left(\sum_{\substack{j=i+1 \\ j \neq k, l}}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}} + \mathbf{1}_{\{\sigma'(i) > \sigma'(k)\}} + \mathbf{1}_{\{\sigma'(i) > \sigma'(l)\}} \right) \\ &= \sum_{i=1}^{k-1} \left(\sum_{\substack{j=i+1 \\ j \neq k, l}}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} + \mathbf{1}_{\{\sigma(i) > \sigma(l)\}} + \mathbf{1}_{\{\sigma(i) > \sigma(k)\}} \right), \end{aligned} \quad (2.11)$$

$$\begin{aligned} \sum_{i=k}^k \sum_{j=i+1}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}} &= \sum_{\substack{j=k+1 \\ j \neq l}}^n \mathbf{1}_{\{\sigma'(k) > \sigma'(j)\}} + \mathbf{1}_{\{\sigma'(k) > \sigma'(l)\}} \\ &= \sum_{\substack{j=k+1 \\ j \neq l}}^n \mathbf{1}_{\{\sigma(l) > \sigma(j)\}} + \mathbf{1}_{\{\sigma(l) > \sigma(k)\}}, \end{aligned} \quad (2.12)$$

$$\begin{aligned} \sum_{i=k+1}^{l-1} \sum_{j=i+1}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}} &= \sum_{i=k+1}^{l-1} \left(\sum_{\substack{j=i+1 \\ j \neq l}}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}} + \mathbf{1}_{\{\sigma'(i) > \sigma'(l)\}} \right) \\ &= \sum_{i=k+1}^{l-1} \left(\sum_{\substack{j=i+1 \\ j \neq l}}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} + \mathbf{1}_{\{\sigma(i) > \sigma(k)\}} \right), \end{aligned} \quad (2.13)$$

$$\sum_{i=l}^l \sum_{j=i+1}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}} = \sum_{j=l+1}^n \mathbf{1}_{\{\sigma'(l) > \sigma'(j)\}} = \sum_{j=l+1}^n \mathbf{1}_{\{\sigma(k) > \sigma(j)\}}, \quad (2.14)$$

$$\sum_{i=l+1}^n \sum_{j=i+1}^n \mathbf{1}_{\{\sigma'(i) > \sigma'(j)\}} = \sum_{i=l+1}^n \sum_{j=i+1}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}}. \quad (2.15)$$

Offenbar verschwinden die Differenzen von $\sum_{i=l+1}^n \sum_{j=i+1}^n \mathbf{1}_{\{\sigma(i) > \sigma(j)\}}$ und dem Ausdruck in (2.15) sowie der Ausdrücke in (2.7) und (2.11), außerdem folgt unter Verwendung von

$$\sigma(k) > \sigma(l) \quad \text{bzw.} \quad \mathbf{1}_{\{\sigma(k) > \sigma(l)\}} = 1 = 1 - \mathbf{1}_{\{\sigma(l) > \sigma(k)\}},$$

dass

$$(2.8) - (2.12) = \sum_{\substack{j=k+1 \\ j \neq l}}^n (\mathbf{1}_{\{\sigma(k) > \sigma(j)\}} - \mathbf{1}_{\{\sigma(l) > \sigma(j)\}}) + (1 - 0) \geq 1$$

$$(2.9) - (2.13) = \sum_{i=k+1}^{l-1} (\mathbf{1}_{\{\sigma(i) > \sigma(l)\}} - \mathbf{1}_{\{\sigma(i) > \sigma(k)\}}) \geq 0$$

$$(2.10) - (2.14) = \sum_{j=l+1}^n (\mathbf{1}_{\{\sigma(l) > \sigma(j)\}} - \mathbf{1}_{\{\sigma(k) > \sigma(j)\}}) \leq 0$$

ist. Insgesamt erhält man also

$$\begin{aligned} & d_\tau(\text{id}, \sigma) - d_\tau(\text{id}, \sigma') \\ &= \sum_{\substack{j=k+1 \\ j \neq l}}^n (\mathbf{1}_{\{\sigma(k) > \sigma(j)\}} - \mathbf{1}_{\{\sigma(l) > \sigma(j)\}}) + (1 - 0) \\ &+ \sum_{i=k+1}^{l-1} (\mathbf{1}_{\{\sigma(i) > \sigma(l)\}} - \mathbf{1}_{\{\sigma(i) > \sigma(k)\}}) - \sum_{j=l+1}^n (\mathbf{1}_{\{\sigma(k) > \sigma(j)\}} - \mathbf{1}_{\{\sigma(l) > \sigma(j)\}}) \\ &= \sum_{j=k+1}^{l-1} (\mathbf{1}_{\{\sigma(k) > \sigma(j)\}} - \mathbf{1}_{\{\sigma(l) > \sigma(j)\}}) + \sum_{i=k+1}^{l-1} (\mathbf{1}_{\{\sigma(i) > \sigma(l)\}} - \mathbf{1}_{\{\sigma(i) > \sigma(k)\}}) + 1 \\ &= 2 \sum_{i=k+1}^{l-1} \mathbf{1}_{\{\sigma(l) < \sigma(i) < \sigma(k)\}} + 1 \geq 1, \end{aligned} \tag{2.16}$$

wie behauptet. □

Bevor wir fortfahren, halten wir die folgende Beobachtung fest: Gleichung (2.16) formalisiert die z.B. mit einer Skizze schnell einzusehende Tatsache, dass für die Änderungen (=Verminderung) bei der Anzahl der Inversionen in einer Permutation π insgesamt bei Auflösen einer einzelnen Inversion (k, l) nur Zuordnungen i zu $\pi(i)$ mit $k < i < l$ und $\pi(l) < \pi(i) < \pi(k)$ eine Rolle spielen.

Lemma 2.2.7. *Für $n \in \mathbb{N}$ sowie feste Daten x_1, \dots, x_n und y_1, \dots, y_n sei S'_n der zugehörige Raum der zulässigen Permutationen. Dann gilt $\text{id} \in S'_n$ und jedes Element $\sigma_1 \in S'_n$ kann nur durch Auflösung bzw. Konstruktion von Inversionen in endlich vielen Schritten in jedes $\sigma_2 \in S'_n$ umgestellt werden, ohne in den Zwischenschritten S'_n zu verlassen.*

Beweis. Nach Lemma 2.2.4 wird durch Auflösen einer Inversion der Raum der zulässigen Permutationen nicht verlassen. Da nach den Ausführungen in Anhang A die Anzahl von Inversionen in einem $\pi \in S'_n$ gerade dem Tau-Abstand $d_\tau(\pi, \text{id})$ entspricht, ist nach Lemma 2.2.6 beim dortigen σ' die Anzahl der Inversionen insgesamt echt kleiner als bei σ . Ferner kann es maximal $\binom{n}{2}$ Inversionen geben, so dass nach endlich vielen solcher Umstellungen die (inversionsfreie) Identität erreicht ist. Man kann also nur durch Verwendung von Inversionen jede beliebige zulässige Permutation in die Identität umstellen, ohne in den einzelnen Schritten S'_n zu verlassen, somit muss $\text{id} \in S'_n$ gelten. Sei

Gegeben sind (geordnete) Geburtszeitpunkte $(x_i)_{i=1}^n$ und Todesdaten $(y_j)_{j=1}^n$.

- (1) markiere alle y_j als frei
- (2) FOR (i in $n : 1$) {
- (3) bestimme $l := \min\{j : y_j \text{ ist frei und zulässig für } x_i\}$
- (4) markiere y_l als nicht frei
- (5) speichere Paarung (x_i, y_l)
- (6) }

Abbildung 2.1. Der Algorithmus *MaxCross*

nun $(i_{l1}, j_{l1}), \dots, (i_{lk_l}, j_{lk_l})$ für $l = 1, 2$ eine mögliche Folge von aufzulösenden Inversionen, um ein beliebiges σ_l in id umzustellen. Beginnend bei σ_1 führt dann zunächst Auflösen der Inversionen $(i_{11}, j_{11}), \dots, (i_{1k_1}, j_{1k_1})$ und anschließendes Konstruieren der Inversionen $(i_{2k_2}, j_{2k_2}), \dots, (i_{21}, j_{21})$ zu σ_2 , ohne in den Zwischenschritten S'_n zu verlassen. \square

Mithilfe der vorhergehenden Ergebnisse kann unter bestimmten Umständen die MAP-Retrodiktio für die wahre Permutation direkt angegeben werden.

Satz 2.2.8. *Es seien die Tripel $(a_{ijkl}, b_{ijkl}, c_{ijkl})$ und die Funktion g definiert wie in Lemma 2.2.5.*

- (a) *Die Identität ist genau dann eine MAP-Retrodiktio, wenn $g(a_{ijkl}, b_{ijkl}, c_{ijkl}) \geq 1$ für alle der obigen Tripel gilt.*
- (b) *Die sich aus dem in Abb. 2.1 beschriebenen Algorithmus *MaxCross* ergebende Permutation π_{MC} ist genau dann eine MAP-Retrodiktio, wenn $g(a_{ijkl}, b_{ijkl}, c_{ijkl}) \leq 1$ für alle der obigen Tripel ist.*

Sind die Daten paarweise verschieden und ist in (a) bzw. (b) die Ungleichung strikt, ist das Ergebnis eindeutig.

Beweis. Wir zeigen zunächst die Rückrichtung für (a), dazu ist

$$S'_n \ni \pi \mapsto \sum_{i=1}^n \log(f_Z(y_{\pi(i)} - x_i)) \rightarrow \max$$

in der gegebenen Situation zu lösen. Der Wert obiger Summe wird hier nach Lemma 2.2.5 ausgehend von einem beliebigen $\pi \in S'_n$ nicht vermindert, wenn eine bestehende Inversion (i, j) aufgelöst wird. Nach Lemma 2.2.7 kann so in endlich vielen Schritten die (noch unbekannte) zulässige Permutation erreicht werden, die möglichst wenige Inversionen besitzt. Da es in der Identität keine Inversionen gibt und diese nach dem gleichen

Lemma zulässig ist, folgt nun diese Teilbehauptung. Für die Rückrichtung von (b) ist entsprechend die zulässige Permutation mit den meisten Inversionen zu bestimmen. Für $\pi \in S_n$ ist dies π_{Max} , $\pi_{\text{Max}}(i) := n - i + 1$, die allerdings nicht immer zulässig ist. Es gilt $\pi_{\text{Max}}(j) < \pi_{\text{Max}}(i)$ für alle i echt kleiner als j , also ist hier die Anzahl der Inversionen

$$m_{\text{Max}} = (n - 1) + (n - 2) + \dots + 1 = \frac{n(n - 1)}{2}.$$

Im Fall, dass nicht alle Permutationen zulässig sind, bezeichne r_i für $i = 1, \dots, n$ den kleinsten für i zulässigen Wert und $\pi'_{\text{Max}} = \pi'_{\text{Max}}(r_1, \dots, r_n)$ die (unbekannte) zulässige Permutation mit den meisten Inversionen. Bei Beschränkung auf S'_n vermindert sich für alle j die Anzahl der Inversionen mindestens um $r_j - 1$ und für deren maximale Anzahl m_{zul} ergibt sich dann

$$m_{\text{zul}} \leq m_{\text{max}} - \sum_{j=1}^n (r_j - 1) = \frac{n(n + 1)}{2} - \sum_{j=1}^n r_j.$$

Wir zeigen nun, dass die Anzahl m_{MC} der Inversionen in der durch *MaxCross* bestimmten Permutation π_{MC} diese Obergrenze erreicht. Nach Definition des Algorithmus' gilt für jedes $2 \leq j \leq n$

$$\#\{i : 1 \leq i < j, \pi_{\text{MC}}(i) > \pi_{\text{MC}}(j)\} = n - r_j - (n - j) = j - r_j,$$

also folgt

$$m_{\text{MC}} = \#\{(i, j) : n \geq j > i \geq 1, \pi_{\text{MC}}(i) > \pi_{\text{MC}}(j)\} = \sum_{j=1}^n (j - r_j) = \frac{n(n + 1)}{2} - \sum_{j=1}^n r_j.$$

Ferner ist π_{MC} nach Konstruktion zulässig, so dass auch diese Teilbehauptung folgt. Zum Beweis der anderen Richtung von (a) bzw. (b) nehmen wir an, dass es ein Tripel $(a_{ijkl}, b_{ijkl}, c_{ijkl})$ gibt, für das die jeweilige Ungleichung nicht erfüllt ist. Dann lässt sich wie im Beweis von Lemma 2.2.5 zeigen, dass die zu maximierende Summe durch Auflösen bzw. Konstruktion der Inversion (i, j) einen größeren Wert annimmt als bei der Identität bzw. π_{MC} . Die Eindeutigkeitsbehauptung ist sofort einzusehen, da unter der genannten Bedingung die Ungleichungen am Anfang des Beweises nach Lemma 2.2.5 strikt sind. \square

Bemerkung 2.2.9. Wie bereits erwähnt entspricht die Anzahl von Inversionen in einem $\sigma \in S'_n$ nach den Ausführungen in Anhang A gerade dem Tau-Abstand $d_\tau(\sigma, \text{id})$. Mit der zweiten oben genannten algorithmischen Verwendungsmöglichkeit von Definition A.1 kann das Vorgehen im Algorithmus *MaxCross* dann folgendermaßen veranschaulicht werden: Für eine Permutation σ werden in zwei übereinanderstehende Zeilen die Zahlen $1, \dots, n$ bzw. die Werte von $\sigma(1), \dots, \sigma(n)$ geschrieben und gleiche Zahlen in der ersten und zweiten Zeile durch Strecken verbunden. Jede Überkreuzung von zwei Linien entspricht einer Inversion (i, j) . Der Algorithmus bestimmt nun die Permutation,

für die die Anzahl der Überkreuzungen maximal ist, dabei darf der (durch vorliegende Geburts- und Todesdaten gegebene) Raum S'_n der zulässigen Permutationen nicht verlassen werden. Dies motiviert auch den Namen für den Algorithmus *MaxCross*. \triangleleft

Die Überprüfung der Voraussetzungen von Satz 2.2.8 ist gerade bei größeren Werten von n sehr aufwändig, deshalb geben wir im Folgenden ein hinreichendes Kriterium an, das diesen Nachteil vermeidet. Dazu erinnern wir daran, dass eine Funktion f (*streng*) *logkonkav* heißt, wenn $x \mapsto \log(f(x))$ (*streng*) konkav ist, und entsprechend (*streng*) *logkonvex*, wenn $x \mapsto \log(f(x))$ (*streng*) konvex ist.

Lemma 2.2.10. *Für jede konkave (konvexe) Funktion f wird*

$$S'_n \ni \pi \mapsto \sum_{i=1}^n f(y_{\pi(i)} - x_i)$$

durch Auflösen einer Inversion nicht vermindert (nicht erhöht). Ist f streng konkav (streng konvex) und sind die Daten paarweise verschieden, wird der entsprechende Ausdruck dann sogar echt erhöht (vermindert).

Beweis. Analog zum Beweis von Lemma 2.2.5 kann gezeigt werden, dass sich der Wert obiger Summe beim Auflösen einer Inversion um $f(a+b) + f(b+c) - f(b) - f(a+b+c)$ für gewisse $a, c \geq 0$ und $b > 0$ verändert. Ist $a = 0$ oder $c = 0$, verschwindet die Differenz. Seien nun $a \neq c$ und o.E. $0 < a < c$. Bekanntlich gilt $g(x) \geq \lambda g(x_1) + (1 - \lambda)g(x_2)$ für eine konkave Funktion g , $\lambda \in (0,1)$ und $x_1 < x := \lambda x_1 + (1 - \lambda)x_2 < x_2$. Die Ungleichung ist äquivalent zu

$$\frac{g(x) - g(x_1)}{x - x_1} \geq \frac{g(x_2) - g(x)}{x_2 - x}.$$

In der betrachteten Situation ergibt sich daraus

$$\frac{f(a+b) - f(b)}{(a+b) - b} \geq \frac{f(b+c) - f(a+b)}{(b+c) - (a+b)} \geq \frac{f(a+b+c) - f(b+c)}{(a+b+c) - (b+c)}$$

und damit die Nichtnegativität obiger Differenz. Entsprechend erhält man auch die benötigte Aussage für $a = c$. Bei einer konvexen Funktion gelten die umgekehrten Ungleichungen, weiter sind bei strenger Konkavität bzw. strenger Konvexität und paarweise verschiedenen Daten die jeweiligen Ungleichungen strikt. \square

Satz 2.2.11. *Ist f_Z logkonkav, so ist die Identität eine MAP-Retrodiktion. Im Fall, dass f_Z logkonvex ist, ergibt sich π_{MC} aus dem Algorithmus *MaxCross* als eine MAP-Retrodiktion. Bei paarweise verschiedenen Daten und Vorliegen von strenger Logkonkavität bzw. Logkonvexität ist das jeweilige Ergebnis eindeutig.*

Beweis. Da für MAP $\sum_{i=1}^n \log(f_Z(y_{\pi(i)} - x_i))$ über alle $\pi \in S'_n$ zu maximieren ist, folgt die Behauptung analog zum Beweis von Satz 2.2.8 unter Verwendung von Lemma 2.2.10. \square

In der nächsten Bemerkung sammeln wir einige Folgerungen, die sich aus dem vorhergehenden Satz ergeben. Dazu erinnern wir daran, dass die Verteilung mit Dichte $f_Z(z) = \frac{k}{\lambda} (\frac{z}{\lambda})^{k-1} e^{-(z/\lambda)^k}$ für $\lambda, k > 0$ Weibull-Verteilung, die Verteilung mit Dichte $f_Z(z) = \lambda^\alpha z^{\alpha-1} \exp(-\lambda z) / \Gamma(\alpha)$ für $\alpha, \lambda > 0$ Gamma-Verteilung sowie die Verteilung mit Dichte $f_Z(z) = \alpha \lambda^\alpha / (\lambda + z)^{\alpha+1}$ für $\alpha, \lambda > 0$ Pareto-Verteilung heißt, jeweils für $z > 0$. Im als zweites genannten Fall schreiben wir auch $X \sim \Gamma(\alpha, \lambda)$, wenn X dieser Verteilung genügt.

Bemerkung 2.2.12. Bekanntlich sind die Dichten einer Weibull-Verteilung mit $k > 1$ und einer Gamma-Verteilung bei $\alpha > 1$ logkonkav. Hier ist also die Identität eine MAP-Retrodiktion. Dagegen sind die Dichten einer Weibull-Verteilung bei $k < 1$ und einer Gamma-Verteilung für $\alpha < 1$ sowie einer Pareto-Verteilung logkonvex, so dass sich die durch den Algorithmus *MaxCross* bestimmte Permutation π_{MC} als eine Retrodiktion ergibt. In allen Fällen erfolgt die Bestimmung des Ergebnisses unabhängig vom Skalenparameter der jeweiligen Verteilung (und bei der Pareto-Verteilung auch unabhängig vom anderen Parameter).

Für $\alpha = 1$ und $k = 1$ entspricht die Gamma- bzw. Weibull-Verteilung einer Exponential-Verteilung, deren Dichte logkonvex *und* logkonkav ist, somit sind sowohl die Identität als auch π_{MC} eine MAP-Retrodiktion. Die obige Voraussetzung strenger Logkonvexität bzw. strenger Logkonkavität ist also tatsächlich für die Eindeutigkeit notwendig. Bei exponentialverteilten Lebensdauern lässt sich sogar

$$\sum_{i=1}^n \log(f_Z(y_{\pi(i)} - x_i)) = \sum_{i=1}^n \log(f_Z(y_i - x_i)) \quad \text{für alle } \pi \in S'_n$$

zeigen, d.h. die Summe nimmt für alle zulässigen Permutationen denselben Wert an und damit ist die durch die Daten bedingte Verteilung der wahren Permutation eine Gleichverteilung. Damit ergibt *jede* zulässige Permutation eine andere MAP-Retrodiktion für die wahre Permutation im Sinne der Definition in (2.2). Durch die MAP-Methode erhält man in diesem Fall somit weder eine Empfehlung, welche Permutation zu wählen ist, noch, welche *nicht* zu wählen ist. Eine Möglichkeit für den Anwender besteht darin, beliebig ein Element von S'_n zu verwenden, dazu geben wir in Abschnitt 2.3.1 weiter unten den Algorithmus *UnifPerm* an. Welche Ergebnisse MMD- und MMSD-Methode hier liefern, wird in Abschnitt 2.3.3 untersucht. \triangleleft

Beispiel 2.2.13. In Beispiel 2.2.2 war auch bei Lebensdauern aus einer inversen Gauß-Verteilung die Identität eine MAP-Retrodiktion. Wären allerdings dieselben Daten für $\lambda = 0.5$ beobachtet worden (der Wert von μ ist nach Lemma 2.2.3 irrelevant), hätte

sich $(\frac{1}{11} \frac{2}{3} \frac{3}{1} \frac{4}{2} \frac{5}{4} \frac{6}{10} \frac{7}{9} \frac{8}{8} \frac{9}{7} \frac{10}{5} \frac{11}{6})$ als Retrodiktioen ergeben. Diese Permutation ist offensichtlich sowohl von der Identität als auch vom hier durch den Algorithmus *MaxCross* bestimmten $(\frac{1}{11} \frac{2}{10} \frac{3}{2} \frac{4}{1} \frac{5}{4} \frac{6}{3} \frac{7}{9} \frac{8}{8} \frac{9}{7} \frac{10}{5} \frac{11}{6})$ verschieden. Die beiden bei logkonvexen bzw. logkonkaven Dichten vorkommenden Permutationen treten allerdings im nachfolgend beschriebenen Sinn als „Grenzlösung“ auf. Für sehr kleine Werte von λ verliert der zweite Term in (2.6) an Einfluss und verschwindet bei $\lambda \rightarrow 0$ ganz, so dass der verbleibende Ausdruck (log ist eine konkave Funktion) durch Einsetzen des durch den Algorithmus *MaxCross* gelieferten Ergebnisses minimiert(!) wird. Für hinreichend große Werte von λ dagegen kann der erste Term vernachlässigt werden, und der übrig gebliebene Ausdruck ($\frac{1}{x}$ ist konvex) nimmt für die Identität seinen kleinsten Wert an. Dies macht das obige Ergebnis für $\lambda \approx 11$ plausibel. \triangleleft

2.3. Rekonstruktion mit der MMD-Methode

2.3.1. Die Struktur der Menge der zulässigen Permutationen

Wir beginnen diesen Abschnitt damit, einige allgemeine Resultate zur Struktur von S'_n bereitzustellen, auf die nicht nur in diesem, sondern auch im nächsten Kapitel als Hilfsmittel zurückgegriffen wird. Zunächst geben wir durch die nachfolgende Definition eine Möglichkeit an, wie das Zusammenspiel von zulässigen und unzulässigen Todeszeitpunkten für verschiedene Geburtsdaten veranschaulicht werden kann.

Definition 2.3.1. Es seien x_1, \dots, x_n und y_1, \dots, y_n die aufsteigend geordneten Geburts- und Todeszeitpunkte. Dann heißt die Matrix $A = (a_{ij})_{i,j=1,\dots,n}$ mit

$$a_{ij} = \begin{cases} 1, & x_i < y_j, \\ 0, & x_i \geq y_j, \end{cases}$$

die *Zulässigkeitsmatrix* von $(x_i)_{i=1}^n$ und $(y_i)_{i=1}^n$. Ist ersichtlich, welche Daten gemeint sind, spricht man auch nur von der Zulässigkeitsmatrix. Weiter sei

$$r_i := \min\{j = 1, \dots, n : x_i < y_j\} \quad (= \min\{j = 1, \dots, n : a_{ij} = 1\})$$

der *kleinste für x_i zulässige Index*, $i = 1, \dots, n$.

Die Gestalt der Zulässigkeitsmatrix und die für die kleinsten Indizes möglichen Werte werden durch das folgende Lemma genauer charakterisiert.

Lemma 2.3.2. *Es sei $\mathbb{N}_n := \{1, \dots, n\}$ und r_1, \dots, r_n kleinste zulässige Indizes.*

(a) *Die Funktion $\mathbb{N}_n \rightarrow \mathbb{N}_n : i \mapsto r_i$ ist (schwach) monoton in i .*

(b) Für $i = 1, \dots, n$ gilt $r_i \in \{1, \dots, i\}$.

Beweis. (a) Es sei $i \in \mathbb{N}_{n-1}$. Wegen $x_i \leq x_{i+1}$ ist jeder für x_{i+1} zulässige Todeszeitpunkt auch für x_i zulässig, dementsprechend gilt $r_i \leq r_{i+1}$.

(b) Für $i \leq j \leq n$ ist $x_i \leq x_j$, außerdem gibt es zu jedem x_i mindestens ein y_j , das größer ist. Entsprechend hat man für jedes $i \in \mathbb{N}_n$ mindestens $n - i + 1$ der y -Werte, die größer als x_i sind, nämlich $y_n, y_{n-1}, \dots, y_{n-(n-i+1)+1}$, und damit gilt $x_i < y_i$. \square

Jede Zulässigkeitsmatrix A besteht also aus zwei jeweils zusammenhängenden Teilen mit allen Einträgen 1 bzw. 0, dabei umfasst der Bereich der Einsen mindestens die Hauptdiagonale und den darüber liegenden Teil. Dies ist eine direkte Folge unserer Annahme, dass die Lebensdauern nach oben unbeschränkt sind. Die Isotonie (Lemma 2.3.2 (a)) bedeutet anschaulich, dass beim Vergleich einer Zeile der Zulässigkeitsmatrix mit einer beliebigen darunter liegenden der Bereich mit Einträgen 0 bei der letztgenannten Zeile mindestens die gleichen Positionen umfasst wie bei der anderen. Anders ausgedrückt: Über Einträgen 1 stehen nur weitere Einsen, unter Einträgen 0 nur weitere Nullen.

Bemerkung 2.3.3. (a) Sind für $1 \leq i \leq n$ genau $n - i + 1$ der Todesdaten größer als x_i , ergibt sich aus dem Beweis von Lemma 2.3.2 (b), dass es sich dabei um y_i, \dots, y_n handeln muss und folglich jede zulässige Permutation diese den Geburten x_i, \dots, x_n zuordnet. Damit kommen Zuordnungen zwischen Elementen aus $\{y_i, \dots, y_n\}$ und $\{x_1, \dots, x_{i-1}\}$ nicht mehr in Frage. Man erhält also eine „Blockbildung“ in dem Sinn, dass das Rekonstruktionsproblem für die Daten $(x_i)_{i=1}^n, (y_i)_{i=1}^n$ in die beiden Teilprobleme („Blöcke“) mit Daten $x_1, \dots, x_{i-1}, y_1, \dots, y_{i-1}$ bzw. $x_i, \dots, x_n, y_i, \dots, y_n$ zerfällt. Die „Blöcke“ werden uns in Kapitel 3 unter der Bezeichnung „Betriebsphasen“ bei $M/G/\infty$ -Warteschlangen wiederbegegnen.

(b) Durch die Zulässigkeitsmatrix lässt sich das (weiter unten ausführlicher betrachtete) Problem, eine beliebige zulässige Permutation zu finden, anschaulich dadurch beschreiben, dass man n Einträge „1“ finden muss, so dass in jeder Zeile und Spalte genau eine „1“ steht. Aus der Position (Nummer von Zeile und Spalte) der so ausgewählten Einträge in der Matrix lässt sich dann die resultierende Zuordnung ablesen. Lemma 2.3.2 (b) liefert dann, dass die Identität immer zulässig ist, da sie der Auswahl der Hauptdiagonalen in der Zulässigkeitsmatrix entspricht. In Lemma 2.2.7 haben wir $\text{id} \in S'_n$ für jedes S'_n auf andere Weise bewiesen. \triangleleft

Beispiel 2.3.4. Es seien

$$A := \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad B := \left(\begin{array}{cc|ccc} 1 & \boxed{1} & 1 & 1 & 1 \\ \boxed{1} & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 1 & 1 & \boxed{1} \\ 0 & 0 & \boxed{1} & 1 & 1 \\ 0 & 0 & 0 & \boxed{1} & 1 \end{array} \right).$$

Liegen Daten $x = (1,2,3,4,5)$ und $y = (2.25,2.75,4.5,5.5,6.5)$ vor, so ist A die zugehörige Zulässigkeitsmatrix. Dann ergibt sich B , wenn man durch Linien die in der vorhergehenden Bemerkung beschriebenen „Blöcke“ links oben und rechts unten markiert. Durch die eingerahmten Kästchen wird die Permutation $(\frac{1}{2} \frac{2}{1} \frac{3}{5} \frac{4}{3} \frac{5}{4})$ dargestellt. \triangleleft

Es soll nun ein Algorithmus angegeben werden, der bei Vorliegen von Daten eine Gleichverteilung auf den zulässigen Permutationen realisiert. Durch diesen kann z.B. bei der MAP-Methode im Fall exponentialverteilter Lebensdauern (s.o.) eine beliebige Permutation als Retrodiktio ausgewählt werden bzw. bei anderen Lebensdauervertelungen eine Vergleichsgröße für die gefundene Retrodiktio bestimmt werden; zusätzliche Anwendungsmöglichkeiten kommen im weiteren Verlauf der Arbeit noch hinzu. Das folgende Lemma gibt einen Hinweis darauf, wie der gewünschte Algorithmus aufgebaut sein kann, außerdem enthält es später benötigte Hilfsresultate.

Lemma 2.3.5. *Es seien $\Pi \sim \text{unif}(S'_n)$ und $\{r_i, \dots, n\}$ die Menge der für x_i zulässigen Indizes, $1 \leq i \leq n$. Dann gilt:*

- (a) $\#S'_n = \prod_{i=1}^n (i - r_i + 1)$.
- (b) $\Pi(n) \sim \text{unif}\{r_n, \dots, n\}$.
- (c) *Die bedingte Verteilung $\mathcal{L}(\Pi(k) | \Pi(k+1) = \pi_{k+1}, \dots, \Pi(n) = \pi_n)$ ist eine Gleichverteilung auf der Menge $\{r_k, \dots, n\} \setminus \{\pi_{k+1}, \dots, \pi_n\}$ für $k = 1, \dots, n-1$.*

Beweis. (a) Es gilt für $(i_1, \dots, i_n) \in \{1, \dots, n\}^n$

$$\begin{aligned} \#S'_n &= \#\{(i_1, \dots, i_n) : x_j < y_{i_j} \text{ für } j = 1, \dots, n, i_k \neq i_l \text{ für } k \neq l\} \\ &= \sum_{i_n=r_n}^n \sum_{\substack{i_{n-1}=r_{n-1} \\ i_{n-1} \neq i_n}}^n \cdots \sum_{\substack{i_1=r_1 \\ i_1 \neq i_n, \dots, i_2}}^n \#\{(i_1, \dots, i_n)\} \\ &= \prod_{i=1}^n (i - r_i + 1). \end{aligned}$$

Gegeben sind (geordnete) Geburtszeitpunkte $(x_i)_{i=1}^n$ und Todesdaten $(y_j)_{j=1}^n$.

- (1) markiere alle y_j als verfügbar
- (2) FOR (i in $n : 1$) {
- (3) wähle aus $\{y_j : y_j > x_i, y_j \text{ verfügbar}\}$ gleichverteilt ein y_l aus
- (4) speichere Paarung (x_i, y_l)
- (5) markiere y_l als nicht verfügbar
- (6) }

Abbildung 2.2. Der Algorithmus *UnifPerm*

(b) Es sei $k \in \{r_n, \dots, n\}$ beliebig. Dann ist wegen $\Pi \sim \text{unif}(S'_n)$

$$\begin{aligned}
P(\Pi(n) = k) &= \sum_{(i_1, \dots, i_{n-1}) \text{ zulässig}} P(\Pi(n) = k, \Pi(1) = i_1, \dots, \Pi(n-1) = i_{n-1}) \\
&= \sum_{\substack{i_{n-1}=r_{n-1} \\ i_{n-1} \neq k}}^n \sum_{\substack{i_{n-2}=r_{n-2} \\ i_{n-2} \neq k, i_{n-1}}}^n \cdots \sum_{\substack{i_1=r_1 \\ i_1 \neq k, i_{n-1}, \dots, i_2}}^n \frac{1}{\#S'_n} \\
&= \frac{\prod_{i=1}^{n-1} (i - r_i + 1)}{\#S'_n} = \frac{1}{n - r_n + 1},
\end{aligned}$$

beim letzten Schritt wurde Teil (a) verwendet.

(c) Mit der Argumentation aus dem Beweis von Teil (b) ergibt sich für $1 \leq k \leq n-1$

$$\begin{aligned}
P(\Pi(k) = \pi_k | \Pi(k+1) = \pi_{k+1}, \dots, \Pi(n) = \pi_n) \\
&= \frac{P(\Pi(k) = \pi_k, \Pi(k+1) = \pi_{k+1}, \dots, \Pi(n) = \pi_n)}{P(\Pi(k+1) = \pi_{k+1}, \dots, \Pi(n) = \pi_n)} = \frac{\sum_{(i_1, \dots, i_{k-1}) \text{ zul.}} \frac{1}{\#S'_n}}{\sum_{(i_1, \dots, i_k) \text{ zul.}} \frac{1}{\#S'_n}} \\
&= \frac{\prod_{i=1}^{k-1} (i - r_i + 1)}{\prod_{i=1}^k (i - r_i + 1)} = \frac{1}{k - r_k + 1}. \quad \square
\end{aligned}$$

Nach diesen Vorüberlegungen geben wir in Abb. 2.2 den Algorithmus *UnifPerm* an, der wie im folgenden Satz bewiesen die gewünschte Gleichverteilung realisiert.

Satz 2.3.6. *Durch den Algorithmus UnifPerm erhält man eine Gleichverteilung auf S'_n , d.h. es wird gleichverteilt eine zulässige Permutation bestimmt.*

Beweis. Der Beweis erfolgt über vollständige Induktion, dabei ist der Fall $n = 1$ trivial, da dort nur x_1 und y_1 vorliegen. Gibt es $n+1$ Geburt-Tod-Paare, wird laut Algorithmus zuerst das zur Geburt x_{n+1} gehörige Sterbedatum, bezeichnet mit $y_{\pi_{n+1}}$, gleichverteilt ausgewählt. Dieses Vorgehen ist nach Lemma 2.3.5 (b) notwendig für die gewünschte Gleichverteilung und gegeben die Paarung $(x_{n+1}, y_{\pi_{n+1}})$ müssen nun noch jeweils n

Geburts- und Todesdaten einander zugeordnet werden. Für letzteres liefert der Algorithmus nach Induktionsannahme gleichverteilt eine zulässige Zuordnung. Ergänzt man diese um die Paarung $(x_{n+1}, y_{\pi_{n+1}})$, ergibt sich wegen

$$\begin{aligned} P(\Pi(1) = \pi_1, \dots, \Pi(n+1) = \pi_{n+1}) \\ = P(\Pi(1) = \pi_1, \dots, \Pi(n) = \pi_n | \Pi(n+1) = \pi_{n+1}) \cdot P(\Pi(n+1) = \pi_{n+1}) \end{aligned}$$

eine Gleichverteilung auf S'_{n+1} , da in der zweiten Zeile zwei Gleichverteilungen stehen. \square

Bemerkung 2.3.7. (a) Mit dem in Bemerkung 2.3.3 (b) beschriebenen Vorgehen bei einer Zulässigkeitsmatrix kann die Funktionsweise von *UnifPerm* folgendermaßen veranschaulicht werden:

Zu Beginn wird in der n -ten Zeile ein Eintrag „1“ gleichverteilt ausgewählt und anschließend die komplette Spalte, in der der entsprechende Eintrag steht, als nicht verfügbar markiert. Nun ist in der $(n-1)$ -ten Zeile ein verfügbarer Eintrag „1“ gleichverteilt zu bestimmen und danach auch die zugehörige Spalte mit „nicht verfügbar“ zu bezeichnen. Analog verfährt man mit der $(n-2)$ -ten, $(n-3)$ -ten, ..., 2-ten Zeile, wobei sich in jedem Schritt die Anzahl der verfügbaren Einträge in jeder Zeile um einen vermindert. Die erste Zeile besitzt dann nur noch einen verwendbaren Eintrag.

(b) Der vorliegende Algorithmus teilt den Todesfällen nach fallendem Index Geburtsdaten zu. Nach dem gleichen Prinzip ließe sich ein zweiter Algorithmus entwickeln, der den Geburtsdaten nach steigendem Index Todesfälle zuordnet und ebenfalls gleichverteilt eine zulässige Permutation liefert. \triangleleft

2.3.2. Problemstellung

In diesem Abschnitt wird die MMD-Methode zur Rekonstruktion der korrekten Zuordnung angewendet, d.h. der mittlere Abstand bezüglich d_τ soll minimiert werden. Auch diese Aufgabenstellung kann als lineares Programm formuliert werden, allerdings lässt ein Blick auf die Definition der Metrik vermuten, dass das resultierende Optimierungsproblem schwieriger zu handhaben ist als bei der MAP-Methode. Genauer ergibt sich die Zielfunktion

$$Ed_\tau(\Pi, \sigma) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n (P(\Pi(i) < \Pi(j)) \cdot \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} + P(\Pi(i) > \Pi(j)) \cdot \mathbf{1}_{\{\sigma(i) < \sigma(j)\}}),$$

die über alle $\sigma \in S'_n$ zu minimieren ist.

Zur Darstellung als lineares Programm muss die Zulässigkeit einer Permutation durch Ungleichungen überprüfbar sein, welche linear in den verwendeten Entscheidungsvariablen sind. Wie dies realisiert werden kann, demonstrieren wir im Folgenden. Zunächst führen wir binäre Entscheidungsvariablen $(x_{ij})_{i,j=1,\dots,n}$ mit der Interpretation $x_{ij} = 1 \Leftrightarrow \sigma(i) = j$ ein. Entsprechend muss sichergestellt werden, dass die Matrix (x_{ij}) genau einen Eintrag 1 pro Zeile und Spalte besitzt. Aus jedem Wert dieser Matrix kann die zugehörige Permutation $\sigma \in S_n$ abgelesen werden, wobei $\sigma(i)$ für jedes i durch $\sum_{j=1}^n j \cdot x_{ij}$ bestimmt ist. Soll σ außerdem Element von S'_n sein, darf die letztgenannte Summe für alle i nicht kleiner als der entsprechende kleinste zulässige Index r_i sein. (Eine andere Möglichkeit, die Zulässigkeit von Permutationen ohne Verwendung der kleinsten Indizes sicherzustellen, wurde oben bei dem Optimierungsproblem für die MAP-Methode angegeben. Natürlich können beide Möglichkeiten auch für beide Probleme verwendet werden.) Definieren wir weiter $a_{ij} := P(\Pi(i) < \Pi(j))$ und führen anstelle von $\mathbf{1}_{\{\sigma(i) > \sigma(j)\}}$ die Entscheidungsvariable $y_{ij} \in \{0,1\}$ ein, jeweils für alle $1 \leq i < j \leq n$, so lässt sich obige Zielfunktion in der Form $\sum_{i=1}^{n-1} \sum_{j=i+1}^n (a_{ij} \cdot y_{ij} + (1 - a_{ij}) \cdot (1 - y_{ij}))$ schreiben. Man prüft schnell nach, dass der einzige Unterschied zu $\sum_{i=1}^{n-1} \sum_{j=i+1}^n (a_{ij} - \frac{1}{2}) \cdot y_{ij}$ in einer von den y -Variablen unabhängigen Konstante besteht, so dass beide Ausdrücke für dieselben dieser Variablen ihr Minimum annehmen.

Das Problem

$$S'_n \ni \sigma \mapsto \text{Ed}_\tau(\Pi, \sigma) \rightarrow \min$$

entspricht nach obigen Bemerkungen dem linearen Programm

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n (a_{ij} - \frac{1}{2}) \cdot y_{ij} \rightarrow \min \quad (2.17)$$

unter den Bedingungen

$$\begin{aligned} \sum_{j=1}^n x_{ij} &= 1, & \sum_{i=1}^n x_{ij} &= 1, & x_{ij} &\in \{0,1\}, & i, j &= 1, \dots, n, \\ \sum_{j=1}^n j \cdot x_{ij} &\geq r_i, & i &= 1, \dots, n, & y_{ij} &\in \{0,1\}, & 1 \leq i < j \leq n, \\ 0 \leq \frac{1}{n} \sum_{j=1}^n j \cdot x_{ij} - \frac{1}{n} \sum_{l=1}^n l \cdot x_{kl} + (1 - y_{ik}) &\leq 1, & 1 \leq i < k \leq n. \end{aligned}$$

Die letzte Bedingung stellt einen Zusammenhang zwischen den x - und y -Variablen her und sichert, dass jedes y_{ij} in genannter Weise als Indikator fungiert. Bevor nun dieses lineare Programm mit den bekannten Paketen gelöst werden kann, besteht die verbleibende Schwierigkeit in der Bestimmung der Wahrscheinlichkeiten a_{ij} . Wir geben dazu zunächst das folgende Lemma an.

Lemma 2.3.8. Für $\pi \in S'_n$ gilt

$$P(\Pi = \pi | X_1 = x_1, \dots, X_n = x_n, Y_1 = y_1, \dots, Y_n = y_n) = \frac{\prod_{i=1}^n f_Z(y_{\pi(i)} - x_i)}{\sum_{\pi' \in S'_n} \prod_{i=1}^n f_Z(y_{\pi'(i)} - x_i)},$$

dabei ist f_Z eine Dichte der Lebensdauervertelung bzgl. des Lebesgue-Maßes.

Beweis. Aus Gleichung (2.1) ergibt sich mit μ_1 als dem abzählenden Maß auf S'_n , ferner f als gemeinsamer Dichte von X, Y und Π bzgl. $\ell^{2n} \otimes \sum_{\pi \in S'_n} \delta_\pi$ sowie f_X als Lebesgue-Dichte von X für $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n)$

$$\begin{aligned} P(\Pi = \pi | X_1 = x_1, \dots, X_n = x_n, Y_1 = y_1, \dots, Y_n = y_n) \\ = \frac{f(x, y, \pi)}{\sum_{\pi' \in S'_n} f(x, y, \pi')} = \frac{f(y, \pi | x) \cdot f_X(x)}{f_X(x) \cdot \sum_{\pi' \in S'_n} f(y, \pi' | x)} = \frac{f(y, \pi | x)}{\sum_{\pi' \in S'_n} f(y, \pi' | x)}. \end{aligned}$$

Dann ergibt sich die Behauptung aus Satz 2.2.1, dabei beachte man die einleitenden Anmerkungen vor jenem Satz. \square

Offenbar geht die Verteilung von X nur über S'_n in die bedingte Verteilung von Π ein, nämlich über die Anzahl der Summanden der Summe im Nenner. Es kommt bei den X -Werten also nur darauf an, zu wie vielen zulässigen Permutationen sie führen, nicht, aus welcher Verteilung sie stammen. Je größer diese Anzahl ist, desto kleiner wird der Wert des Bruchs im obigen Ausdruck. Dieser Zusammenhang ist plausibel: Die Wahrscheinlichkeitsmasse wird stärker verteilt, so dass die einzelnen Wahrscheinlichkeiten kleiner werden müssen.

Aus Platz- und Übersichtlichkeitsgründen verwenden wir im Folgenden anstelle der durch die beobachteten Daten bedingten Ausdrücke den Index y , schreiben also $P_y(\Pi = \pi)$ und E_y für die bedingte Wahrscheinlichkeit bzw. den bedingten Erwartungswert. Offenbar können nun die benötigten Konstanten zumindest theoretisch mittels

$$a_{ij} = P_y(\Pi(i) < \Pi(j)) = \sum_{\pi \in S'_n, \pi(i) < \pi(j)} P_y(\Pi = \pi)$$

für alle $1 \leq i < j \leq n$ bestimmt werden. In vielen Fällen dürfte allerdings die Mächtigkeit der Menge S'_n dieses Vorgehen unpraktikabel machen. Wir beschränken uns deshalb darauf, eine Möglichkeit zur Berechnung der Werte nur zu skizzieren, und anschließend eine hinreichende Bedingung anzugeben, unter der die optimale Lösung des linearen Programms direkt ersichtlich ist. Für eine Matrix $B = (b_{ij})_{i,j=1,\dots,n}$ heißt

$$\text{perm}(B) := \sum_{\sigma \in S_n} \prod_{i=1}^n b_{i\sigma(i)}$$

bekanntlich die *Permanente* von B . Diese Größe tritt in den folgenden Formeln mehrfach auf, dabei ist zu beachten, dass dort jedes $\sigma \in S_n \setminus S'_n$ den Summanden 0 zur Permanente beiträgt. Es sei $p := \sum_{\pi \in S'_n} \prod_{i=1}^n f_Z(y_{\pi(i)} - x_i)$, dann gilt für $1 \leq i < j \leq n$

$$\begin{aligned} P_y(\Pi(i) < \Pi(j)) &= p^{-1} \cdot \sum_{k < l} \sum_{\pi \in S'_n, \pi(i)=k, \pi(j)=l} \prod_{r=1}^n f_Z(y_{\pi(r)} - x_r) \\ &= p^{-1} \cdot \sum_{k < l} f_Z(y_k - x_i) \cdot f_Z(y_l - x_j) \cdot \sum_{\pi \in S'_{n-2}} \prod_{r=1}^{n-2} f_Z(\check{y}_{\pi(r)} - \check{x}_r), \end{aligned}$$

wobei die geordneten Elemente von $\{x_1, \dots, x_n\} \setminus \{x_i, x_j\}$ und $\{y_1, \dots, y_n\} \setminus \{y_k, y_l\}$ mit $(\check{x}_i)_{i=1}^{n-2}$ bzw. $(\check{y}_i)_{i=1}^{n-2}$ sowie die Menge der für die \check{x} - und \check{y} -Werte zulässigen Permutationen mit \check{S}'_{n-2} bezeichnet sind. Die benötigten Permanenten können entweder direkt nach obiger Definition mit der Zeitkomplexität $O((\#S'_n) \cdot n)$ oder nach dem Algorithmus von Ryser in $O(2^n \cdot n)$ berechnet werden ([Wei99, S. 1348 bzw. S. 1588]), ohne eventuelle Vorteile durch die besondere Struktur der Matrix auszunutzen. Welcher Ansatz schneller ist, hängt offensichtlich von der Gestalt von S'_n ab. Um die Zeitkomplexität für alle Wahrscheinlichkeiten $(a_{ij})_{1 \leq i < j \leq n}$ zu erhalten, muss der jeweilige Wert noch mit n^4 multipliziert werden.

Da zusätzlich die Zeit zur Lösung des linearen Programms zu berücksichtigen ist, bietet der folgende brute force Ansatz möglicherweise für manche Instanzen des Problems eine Alternative zu dem beschriebenen Vorgehen. Dazu wertet man

$$\pi \mapsto E_y d_\tau(\pi, \Pi) = \sum_{\sigma \in S'_n} d_\tau(\pi, \sigma) \cdot \prod_{i=1}^n f_Z(y_{\sigma(i)} - x_i)$$

für alle $\pi \in S'_n$ aus, dabei wurde $P_y(\Pi = \sigma) = p^{-1} \cdot \prod_{i=1}^n f_Z(y_{\sigma(i)} - x_i)$ verwendet; dieses Vorgehen führt offenbar zu der Zeitkomplexität $O((\#S'_n)^2)$. Vergleicht man die hier erhaltenen Komplexitäten mit dem entsprechenden Wert für MAP auf Seite 26, so wird deutlich, dass MMD unter diesem Gesichtspunkt ein schwerer zu lösendes Problem ist. Wie das folgende Lemma zeigt, ist es aber unter bestimmten Umständen möglich, direkt eine optimale Lösung anzugeben.

Lemma 2.3.9. *Gilt $P_y(\Pi(i) > \Pi(j)) \leq P_y(\Pi(i) < \Pi(j))$ für alle $1 \leq i < j \leq n$ bei dem in (2.17) genannten Problem, so ist die Identität eine optimale Lösung. Ist die Ungleichung strikt, ist das Minimum sogar eindeutig.*

Beweis. Die genannte Bedingung ist äquivalent zu $P(\Pi(i) < \Pi(j)) \geq \frac{1}{2}$. In diesem Fall wird die Zielfunktion in (2.17) minimal, wenn $y_{ij} = 0$ für alle $1 \leq i < j \leq n$ ist. Dies führt auf die Identität, die immer zulässig ist. Die zweite Behauptung ist nun unmittelbar einzusehen. \square

2.3.3. Explizite Lösungen

MMD bei logkonkaven Dichten

Unter Verwendung einiger Resultate aus Abschnitt 2.2.2 können wir direkt das zentrale Resultat für die Anwendung der MMD-Methode bei Lebensdauern mit logkonkaven Dichten angeben. Diese Voraussetzung ist beispielsweise bei einer Gamma-Verteilung mit $\alpha \geq 1$ oder einer Weibull-Verteilung mit $k \geq 1$ erfüllt.

Satz 2.3.10. *Die Lebensdauern mögen aus einer Verteilung mit logkonkaver Dichte stammen und Π gebe die korrekte Zuordnung zwischen den vorliegenden Geburts- und Todesdaten an. Dann gilt für festes $\sigma \in S'_n$*

$$E_y d_\tau(\text{id}, \Pi) \leq E_y d_\tau(\sigma, \Pi),$$

d.h. die Identität ist im vorliegenden Fall eine MMD-Retrodiktion. Ist die Dichte streng logkonkav und sind die Größen (a_{ij}) , $a_{ij} := y_j - x_i$ für alle zulässigen (i, j) , paarweise verschieden, so ist das Minimum eindeutig.

Beweis. Für $1 \leq i < j \leq n$ gilt

$$P_y(\Pi(i) > \Pi(j)) = \sum_{\sigma \in S'_n, \sigma(i) > \sigma(j)} P_y(\Pi = \sigma), \quad (2.18)$$

$$P_y(\Pi(i) < \Pi(j)) = \sum_{\sigma' \in S'_n, \sigma'(i) < \sigma'(j)} P_y(\Pi = \sigma'). \quad (2.19)$$

Nach Lemma 2.2.4 gibt es für jede Zählvariable $\sigma_>$ der ersten Summe eine andere Zählvariable $\sigma_<$ der zweiten Summe, so dass sich $\sigma_>$ und $\sigma_<$ nur durch das Vorliegen bzw. Fehlen der Inversion (i, j) unterscheiden. Dann folgt

$$\frac{P_y(\Pi = \sigma_<)}{P_y(\Pi = \sigma_>)} = \frac{\prod_{t=1}^n f_Z(y_{\sigma_<(t)} - x_t)}{\prod_{t=1}^n f_Z(y_{\sigma_>(t)} - x_t)}$$

mit Lemma 2.3.8. Durch Logarithmieren und Verwendung von Lemma 2.2.10 ist zu erkennen, dass der Zähler nicht kleiner als der Nenner und damit der Bruch größer oder gleich 1 ist. Demnach gibt es für jeden Summanden in der obigen ersten Summe jeweils einen mindestens genauso großen Summanden in der zweiten Summe. Da außerdem alle Größen nicht-negativ sind, folgt nun die Behauptung mit Lemma 2.3.9. Sind die Daten paarweise verschieden und liegt strenge Logkonkavität vor, so ist der obige Quotient sogar echt größer als 1 und auch die zweite Behauptung folgt mit Lemma 2.3.9. \square

Bemerkung 2.3.11. (a) Die Argumentation im Beweis von Satz 2.3.10 kann nur unter Einschränkungen zum Nachweis einer analogen Aussage für Lebensdauern mit einer logkonvexen Dichte verwendet werden. In diesem Fall ist der obige Quotient nicht größer als 1, daraus kann $P(\Pi(i) < \Pi(j)) \leq P(\Pi(i) > \Pi(j))$ aber nur bei Vorliegen einer Bijektion zwischen den Mengen $\{\sigma \in S'_n : \sigma(i) < \sigma(j)\}$ und $\{\sigma' \in S'_n : \sigma'(i) > \sigma'(j)\}$ gefolgert werden (vgl. zweite Aussage von Lemma 2.2.4). Beispielsweise in der schon häufiger erwähnten Situation, dass alle Permutationen zulässig sind, ist Letzteres erfüllt und die durch den Algorithmus *MaxCross* aus Abb. 2.1 bestimmte Permutation führt dann zu einem kleinsten mittleren Abstand. (Bei paarweise verschiedenen (im Satz definierten) (a_{ij}) und strenger Logkonvexität ist das Ergebnis eindeutig.) Im Allgemeinen aber ist so keine Aussage möglich.

(b) Da für Lebensdauerverteilungen mit logkonkaver Dichte und alle $1 \leq i < j \leq n$ gilt $P_y(\Pi(i) < \Pi(j)) \geq \frac{1}{2}$, wird offensichtlich der mittlere Abstand unter den Voraussetzungen des Satzes *maximiert*, wenn unter Einhaltung der Nebenbedingungen im linearen Programm (2.17) möglichst viele der $(y_{ij})_{1 \leq i < j \leq n}$ den Wert 1 haben. Nach dem Beweis von Satz 2.2.8 führt dies auf die durch den Algorithmus *MaxCross* in Abb. 2.1 erstellte Permutation. Wird diese verwendet, macht man im Mittel also am meisten Fehler bei der Rekonstruktion. \triangleleft

Da die Dichte der Exponentialverteilung nur schwach logkonkav ist, lässt sich das Eindeutigkeitsresultat aus dem vorherigen Satz hier nicht anwenden. Wir untersuchen deshalb im Folgenden, wann die MMD-Retrodiktion in diesem Fall eindeutig ist.

Satz 2.3.12. *Bei exponentialverteilten Lebensdauern ist die MMD-Retrodiktion genau dann eindeutig, wenn nur die Identität zulässig ist.*

Beweis. Es sei wieder $N_n := \{1, \dots, n\}$ und r_j bezeichne für $j \in \mathbb{N}_n$ den jeweiligen kleinsten Index. Der Fall $S'_n = \{\text{id}\}$ ist trivial. Im Fall $\#S'_n > 1$ definieren wir $i := \min\{j = 2, \dots, n : r_j = j - 1\}$, daraus ergibt sich sofort $r_{i-1} = i - 1 = r_i$. Man beachte, dass das Minimum wegen $r_j = j \forall j \Leftrightarrow \#S'_n = 1$ existieren muss, dabei gilt die nicht-triviale Richtung der Äquivalenz wegen Lemma 2.3.2 (a). Weiter sei die Permutation σ gegeben durch $\sigma(i) := i - 1, \sigma(i - 1) := i$ und $\sigma(k) := k$ für alle $k \in \mathbb{N}_n \setminus \{i, i - 1\}$. Offenbar liegt σ in S'_n und es gilt

$$E_y d_\tau(\sigma, \Pi) - E_y d_\tau(\text{id}, \Pi) = P_y(\Pi(i - 1) < \Pi(i)) - P_y(\Pi(i - 1) > \Pi(i)).$$

Wegen Teil (a) und $\alpha = 1$ ist $P_y(\Pi = \sigma_>) = P_y(\Pi = \sigma_<)$ für alle $\sigma_>$ und $\sigma_<$, die sich nur durch die Inversion $(i - 1, i)$ unterscheiden. Außerdem führt bei jedem Element von S'_n mit dieser Inversion das Auflösen derselben nicht aus dem Raum der zulässigen Permutationen hinaus, so dass in diesem Fall die Anzahl der Summanden in (2.18) und (2.19) gleich ist. Damit verschwindet die obige Differenz. \square

Bemerkung 2.3.13. Nach dem vorhergehenden Satz ist das Minimum für den mittleren Abstand bei exponentialverteilten Lebensdauern niemals eindeutig, sobald mehr als nur die Identität zulässig ist. Für wie viele Permutationen aus S'_n wird es dann aber angenommen?

Die Antwort kann aus dem Beweis des Satzes abgelesen werden: Die Anzahl hängt davon ab, wie viele Daten x_i, x_{i+1}, \dots es gibt, so dass man sie beliebig y_i, y_{i+1}, \dots zuteilen kann und alle entstehenden Permutationen trotzdem noch in S'_n liegen. In diesem Sinne sind diese Daten dann äquivalent zueinander, da es nicht darauf ankommt, welche von ihnen an welcher Stelle eingesetzt werden, sondern nur, einen beliebigen Vertreter der entsprechenden „Klasse“ zu verwenden. Dementsprechend verwundert es nicht, dass die Identität, die diese Vorgabe durch die Zuordnung x_i zu y_i für alle i einhält, in gewisser Hinsicht äquivalent zu den anderen Permutationen ist, die dies ebenfalls tun.

Nach diesen Überlegungen kann nun die Anzahl m_{\min} der „minimierenden“ Permutationen leicht berechnet werden. Die Rolle von Daten lässt sich genau dann beliebig vertauschen, wenn ihre kleinsten Indizes gleich sind. Ist $d_i := \#\{j = 1, \dots, n : r_j = i\}$ für $i = 1, \dots, n$ und $(r_j)_{j=1}^n$ als kleinsten zulässigen Indizes, dann gilt $m_{\min} = \prod_{i=1}^n d_i!$. Im Fall $S'_n = \{\text{id}\}$ ist $d_i = 1$ für alle i , also $m_{\min} = 1$ (die Menge, über die im obigen Beweis das Minimum genommen wird, ist dann leer), für $S'_n = S_n$ ergibt sich $d_1 = n, d_i = 0$ für $i > 1$, also $m_{\min} = n!$. \triangleleft

Durch das folgende Beispiel zeigen wir, dass MMD in einem bestimmten Sinn bei Exponentialverteilungen bessere Ergebnisse liefert als MAP.

Beispiel 2.3.14. Nach Satz 2.3.12 ist die MMD-Retrodiktion für exponentialverteilte Lebensdauern in den meisten Fällen nicht eindeutig. Bei $S'_n = \{\text{id}\}$ und $S'_n = S_n$ gilt $d_\tau(\text{id}, \Pi) = d_\tau(\sigma, \Pi)$ für alle $\sigma \in S'_n$, so dass jede zulässige Permutation auf denselben mittleren Abstand führt und somit eine von ihnen gleichverteilt ausgewählt werden kann. Das gleiche Resultat ergab sich weiter oben für die MAP-Methode, dort allerdings unabhängig von S'_n . Wir demonstrieren nun, dass es bei MMD tatsächlich Fälle gibt, in denen die Identität in einem bestimmten Sinn echt besser ist als eine gleichverteilt aus S'_n ausgewählte Permutation. Dazu sollen für $n = 3$ die Daten $x_1 < x_2 < y_1 < x_3 < y_2 < y_3$ beobachtet worden sein. Dies führt auf

$$S'_3 = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \right\} =: \{\text{id}, \pi_2, \pi_3, \pi_4\}$$

und man erhält die in Tabelle 2.5 genannten Tau-Abstände. Sind Π und Σ jeweils gleichverteilt auf S'_3 und unabhängig voneinander, ergibt sich $E_y d_\tau(\text{id}, \Pi) = 1$ und $E_y d_\tau(\Sigma, \Pi) = \frac{5}{4}$. In diesem Beispiel führt die Identität also wirklich auf einen kleineren mittleren Abstand als eine gleichverteilt ausgewählte zulässige Permutation. Wie oben

| $d_\tau(\cdot, \cdot)$ | id | π_2 | π_3 | π_4 |
|------------------------|----|---------|---------|---------|
| id | 0 | 1 | 2 | 1 |
| π_2 | 1 | 0 | 1 | 2 |
| π_3 | 2 | 1 | 0 | 3 |
| π_4 | 1 | 2 | 3 | 0 |

Tabelle 2.5. *Tau-Abstände der Permutationen in der jeweiligen Zeile und Spalte*

bewiesen ist das Minimum aber nicht eindeutig, hier wegen $E_y d_\tau(\pi_2, \Pi) = 1$. Bei Exponentialverteilungen ist MMD somit besser als MAP in dem Sinne, dass eine schärfere Trennung erfolgt, d.h. nicht mehr (und in vielen Fällen echt weniger) optimale Lösungen ermittelt werden. \triangleleft

Die Ergebnisse in Satz 2.3.10 hängen nur von allgemeinen Eigenschaften des Rekonstruktionsproblems ab, nicht aber von konkreten Datenwerten. Dies benutzen wir für den Beweis des folgenden Satzes, der sich auf unbedingte Erwartungswerte bezieht.

Satz 2.3.15. *Stammen die Lebensdauern aus einer Verteilung mit logkonkaver Dichte, so gilt für alle $\sigma \in S'_n$*

$$Ed_\tau(\text{id}, \Pi) \leq Ed_\tau(\sigma, \Pi).$$

Beweis. Geht man analog zum Beweis von Satz 2.3.10 vor, stehen anstelle der dortigen bedingten nun die unbedingten Wahrscheinlichkeiten $P(\Pi = \pi)$. Aus den folgenden Ausführungen ergibt sich, dass auch für diese die im Beweis benötigte Ungleichung zutrifft. Ist Y der (unbekannte) Vektor der geordneten Todesdaten mit Verteilung P^Y und nehmen wir die Geburtsdaten o.E. als Konstanten an, so gilt mit den obigen Bezeichnungen und der vorher verwendeten Argumentation

$$P(\Pi = \sigma_<) = \int P(\Pi = \sigma_< | Y = y) dP^Y \geq \int P(\Pi = \sigma_> | Y = y) dP^Y = P(\Pi = \sigma_>). \quad \square$$

Es erscheint bemerkenswert, dass bei logkonkaven Dichten die im Sinn von Satz 2.2.11 und Satz 2.3.10 optimale Rekonstruktion nicht von den Daten abhängt.

MMSD bei der Exponentialverteilung

Eine Formulierung von MMSD als lineares Programm ist analog zu MMD möglich, beinhaltet aber auch die gleichen Schwierigkeiten bei der Komplexität. Wir verzichten deshalb darauf und geben unsere Ergebnisse für exponentialverteilte Lebensdauern direkt an. Nach Bemerkung 2.2.12 ist in dieser Situation die bedingte Verteilung von Π

eine Gleichverteilung auf den zulässigen Permutationen und die folgenden Ergebnisse hängen nur von dieser Voraussetzung ab, so dass wir in diesem Abschnitt allgemeiner $\Pi \sim \text{unif}(S'_n)$ verwenden. Unser Ziel ist es, das Argmin von $S'_n \ni \sigma \mapsto E_y(d_\tau(\sigma, \Pi))^2$ zu bestimmen. Für obiges Π und $1 \leq i < j \leq n$ sei im Folgenden

$$D_{ij}^\sigma := \mathbf{1}_{\{\Pi(i) > \Pi(j)\}} \cdot \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} + \mathbf{1}_{\{\Pi(i) < \Pi(j)\}} \cdot \mathbf{1}_{\{\sigma(i) > \sigma(j)\}},$$

dann ergibt sich

$$(d_\tau(\sigma, \Pi))^2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^n (D_{ij}^\sigma)^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\substack{l=i+1 \\ l \neq j}}^n D_{ij}^\sigma D_{il}^\sigma + \sum_{i=1}^{n-1} \sum_{\substack{k=1 \\ k \neq i}}^{n-1} \sum_{j=i+1}^n \sum_{l=k+1}^n D_{ij}^\sigma D_{kl}^\sigma.$$

Der Erwartungswert dieses Ausdrucks besteht offenbar aus Summen von Funktionen gewisser Wahrscheinlichkeiten und Indikatoren. Bei MMD im vorherigen Abschnitt lag (im Rahmen eines linearen Programms) eine ähnliche Situation vor und dort wurde jeder Summand durch die Identität minimiert. Das folgende Beispiel zeigt, dass dies hier nicht der Fall ist.

Beispiel 2.3.16. Für die Situation aus Beispiel 2.3.14 ist

$$\begin{aligned} (d_\tau(\sigma, \Pi))^2 &= (D_{12}^\sigma)^2 + (D_{13}^\sigma)^2 + (D_{23}^\sigma)^2 + D_{12}^\sigma \cdot D_{13}^\sigma + D_{13}^\sigma \cdot D_{12}^\sigma \\ &\quad + D_{12}^\sigma \cdot D_{23}^\sigma + D_{13}^\sigma \cdot D_{23}^\sigma + D_{23}^\sigma \cdot D_{12}^\sigma + D_{23}^\sigma \cdot D_{13}^\sigma. \end{aligned}$$

In Tabelle 2.6 sind die Erwartungswerte der einzelnen Summanden angegeben, wenn $\Pi \sim \text{unif}\{\text{id}, \binom{123}{132}, \binom{123}{213}, \binom{123}{312}\}$ ist und id bzw. $\binom{123}{213}$ als Permutation σ gewählt wurde. Bei MMSD minimiert die Identität also nicht unbedingt jeden einzelnen Summanden von $E(d_\tau(\sigma, \Pi))^2$, dafür wird zumindest in der hier betrachteten Situation der größere Wert an anderer Stelle wieder ausgeglichen. In einem der folgenden Lemmata zeigen wir, dass dies bei $\Pi \sim \text{unif}(S'_n)$ auch im Allgemeinen geschieht und somit die Identität in diesem Fall eine MMSD-Retrodiktion ist. Das Beispiel demonstriert allerdings auch, dass das Minimum nicht eindeutig sein muss (ein anderes einfaches Beispiel hierfür ist natürlich der Fall $S'_n = S_n$, dann ist der mittlere quadratische Abstand aufgrund der Rechtsinvarianz von Kendalls Tau-Abstand für alle Permutationen gleich). \triangleleft

Wir erarbeiten nun mehrere technische Hilfsmittel bezüglich später auftretender Wahrscheinlichkeiten, wobei die folgende Idee zugrunde liegt. Wegen $\Pi \sim \text{unif}(S'_n)$ können Aussagen über das Verhältnis von Wahrscheinlichkeiten auf das Verhältnis der Mächtigkeit von Mengen zurückgeführt werden. Genauer ist $P(\Pi \in A)$ nicht größer als $P(\Pi \in B)$, wenn A injektiv auf B abgebildet werden kann. Wir beschränken uns deshalb in den folgenden Beweisen darauf, die benötigten Abbildungen anzugeben.

| σ | $E_y(D_{12}^\sigma)^2$ | $E_y(D_{13}^\sigma)^2$ | $E_y(D_{23}^\sigma)^2$ | $E_y D_{12}^\sigma \cdot D_{13}^\sigma$ | $E_y D_{12}^\sigma \cdot D_{23}^\sigma$ | $E_y D_{13}^\sigma \cdot D_{23}^\sigma$ |
|--|------------------------|------------------------|------------------------|---|---|---|
| id | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | 0 | 0 |
| $\begin{pmatrix} 123 \\ 213 \end{pmatrix}$ | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ | 0 |

Tabelle 2.6. Werte der Summanden von $E_y(d_\tau(\sigma, \Pi))^2$ in Abhängigkeit von σ

Lemma 2.3.17. Es seien $\Pi \sim \text{unif}(S'_n)$ sowie $1 \leq i < j, l \leq n$ und $l \neq j$. Dann gilt

$$\begin{aligned} P_y(\Pi(i) > \Pi(j), \Pi(i) > \Pi(l)) &\leq P_y(\Pi(i) < \Pi(j), \Pi(i) < \Pi(l)), \\ P_y(\Pi(i) > \Pi(j), \Pi(i) > \Pi(l)) &\leq P_y(\Pi(i) < \Pi(j), \Pi(j) > \Pi(l)), \\ P_y(\Pi(i) > \Pi(j), \Pi(j) > \Pi(l)) &\leq P_y(\Pi(i) < \Pi(j), \Pi(i) > \Pi(l)). \end{aligned}$$

Beweis. Die benötigten Abbildungen sind

$$f_1(\pi) = \pi' \text{ mit } \pi'(k) = \begin{cases} \min\{\pi(j), \pi(l)\}, & \text{falls } k = i, \\ \pi(i), & \text{falls } k = \arg \min\{\pi(j), \pi(l)\}, \\ \pi(k), & \text{sonst,} \end{cases}$$

im ersten und

$$f_2(\pi) = \pi' \text{ mit } \pi'(k) = \begin{cases} \pi(j), & \text{falls } k = i, \\ \pi(i), & \text{falls } k = j, \\ \pi(k), & \text{sonst,} \end{cases}$$

im zweiten und dritten Fall. Wegen Lemma 2.3.2 (a) ergeben sich so in der Situation in den Mengen auf der linken Seite der Ungleichung wieder verschiedene zulässige Permutationen, die jeweils Elemente der Mengen auf der rechten Seite sind, und es folgt die Behauptung. \square

Lemma 2.3.18. Es sei $\Pi \sim \text{unif}(S'_n)$. Dann gilt für $1 \leq i, k < j \leq n$ und $k \neq i$

$$\begin{aligned} P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(j)) &\leq P_y(\Pi(i) > \Pi(j), \Pi(k) < \Pi(j)), \\ P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(j)) &\leq P_y(\Pi(i) < \Pi(j), \Pi(k) > \Pi(j)), \\ P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(j)) &\leq P_y(\Pi(i) < \Pi(j), \Pi(k) < \Pi(j)). \end{aligned}$$

Beweis. Die Beweisidee ist die gleiche wie beim vorhergehenden Lemma und stützt sich darauf, dass bei $i, k < j$ und $\Pi(j) < \Pi(i), \Pi(k)$, also der Situation in der Menge auf der linken Seite der Ungleichungen oben, immer $\min\{y_i, y_k, y_j\} \geq \max\{x_i, x_k, x_j\}$ gilt. Damit

ist jede Zuordnung dieser Geburts- und Todesfälle untereinander zulässig, insbesondere also auch die, die auf die Situationen in den Mengen auf der rechten Seite führen. Wir gehen die Behauptungen nacheinander durch und unterscheiden dabei u.U. noch einmal jeweils zwei Fälle, dabei schreiben wir kurz „BxFy“ für Behauptung x Fall y und geben nur informell an, was die Abbildung leisten muss. In B1F1 sei $\Pi(i) > \Pi(k)$, dann vertausche die Rollen von $\Pi(k)$ und $\Pi(j)$, und beim dazu komplementären B1F2 setze $\pi'(i) := \pi(k)$, $\pi'(k) := \pi(j)$, $\pi'(j) := \pi(i)$. Bei B2F1 sei $\Pi(i) > \Pi(k)$, dann definieren wir $\pi'(i) := \pi(j)$, $\pi'(k) := \pi(i)$, $\pi'(j) := \pi(k)$, während im dazu komplementären B2F2 die Rollen von $\Pi(i)$ und $\Pi(j)$ vertauscht werden. Bei B3 schließlich vertausche die Rollen von $\Pi(j)$ und $\arg \max\{\Pi(i), \Pi(k)\}$. \square

Lemma 2.3.19. *Es sei $\Pi \sim \text{unif}(S'_n)$. Dann gilt für $1 \leq i < j \leq n$, $1 \leq k < l \leq n$ und $i \neq k, j \neq l$*

$$\begin{aligned} P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(l)) &\leq P_y(\Pi(i) > \Pi(j), \Pi(k) < \Pi(l)), \\ P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(l)) &\leq P_y(\Pi(i) < \Pi(j), \Pi(k) > \Pi(l)), \\ P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(l)) &\leq P_y(\Pi(i) < \Pi(j), \Pi(k) < \Pi(l)). \end{aligned}$$

Beweis. Gilt $j \neq k$, kann man die verwendete Beweisidee separat auf $i, j, \Pi(i), \Pi(j)$ bzw. $k, l, \Pi(k), \Pi(l)$ anwenden. Ist $j = k$, erhält man nach Lemma 2.3.2 (a) durch Abbildungen, die die Rollen von $\Pi(j)$ und $\Pi(l)$ im ersten, von $\Pi(i)$ und $\Pi(j)$ im zweiten und von $\Pi(i)$ und $\Pi(l)$ im dritten Fall vertauschen, aus den Elementen der linken Menge jeweils verschiedene Elemente der rechten Mengen. Analog zur Argumentation in den beiden vorhergehenden Beweisen ergibt sich nun die Behauptung. \square

Nach diesen Vorbereitungen können wir das Hauptresultat für MMSD bei exponentialverteilten Lebensdauern beweisen.

Satz 2.3.20. *Es sei $\Pi \sim \text{unif}(S'_n)$. Dann gilt für festes $\sigma \in S'_n$*

$$\mathbb{E}_y(d_\tau(\text{id}, \Pi))^2 \leq \mathbb{E}_y(d_\tau(\sigma, \Pi))^2,$$

d.h. die Identität ist in der betrachteten Situation die beste MMSD-Retrodiktion, also insbesondere bei exponentialverteilten Lebensdauern.

Beweis. Wir untersuchen die Erwartungswerte der einzelnen Summanden in der eingangs erwähnten Darstellung des mittleren quadrierten Abstands. Für die Erwartungswerte in der ersten Summe gilt

$$\begin{aligned} \mathbb{E}_y(D_{ij}^\sigma)^2 &= \mathbb{E}_y \left((\mathbf{1}_{\{\Pi(i) > \Pi(j)\}} \cdot \mathbf{1}_{\{\sigma(i) < \sigma(j)\}})^2 + (\mathbf{1}_{\{\Pi(i) < \Pi(j)\}} \cdot \mathbf{1}_{\{\sigma(i) > \sigma(j)\}})^2 \right. \\ &\quad \left. + 2 \cdot \mathbf{1}_{\{\Pi(i) > \Pi(j)\}} \cdot \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\Pi(i) < \Pi(j)\}} \cdot \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \right) \end{aligned}$$

$$= \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot P_y(\Pi(i) > \Pi(j)) + \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot P_y(\Pi(i) < \Pi(j))$$

und nach den Ausführungen im Beweis von Satz 2.3.10 wird dies durch die zulässige Permutation σ minimiert, bei der für möglichst viele $1 \leq i < j \leq n$ gilt $\sigma(i) < \sigma(j)$; dies ist die Identität. In der dritten Summe sind $j = l$ und $j \neq l$ unterschiedlich zu behandeln. Im ersten Fall ist $k < j$ für alle Summanden und es ergibt sich

$$\begin{aligned} E_y D_{ij}^\sigma D_{kj}^\sigma &= \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) < \sigma(j)\}} \cdot P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(j)) \\ &\quad + \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) > \sigma(j)\}} \cdot P_y(\Pi(i) > \Pi(j), \Pi(k) < \Pi(j)) \\ &\quad + \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) < \sigma(j)\}} \cdot P_y(\Pi(i) < \Pi(j), \Pi(k) > \Pi(j)) \\ &\quad + \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) > \sigma(j)\}} \cdot P_y(\Pi(i) < \Pi(j), \Pi(k) < \Pi(j)), \end{aligned}$$

nach Lemma 2.3.18 wird dies für die Identität minimal. Im anderen Fall $j \neq l$ gilt

$$\begin{aligned} E_y D_{ij}^\sigma D_{kl}^\sigma &= \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) < \sigma(l)\}} \cdot P_y(\Pi(i) > \Pi(j), \Pi(k) > \Pi(l)) \\ &\quad + \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) > \sigma(l)\}} \cdot P_y(\Pi(i) > \Pi(j), \Pi(k) < \Pi(l)) \\ &\quad + \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) < \sigma(l)\}} \cdot P_y(\Pi(i) < \Pi(j), \Pi(k) > \Pi(l)) \\ &\quad + \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(k) > \sigma(l)\}} \cdot P_y(\Pi(i) < \Pi(j), \Pi(k) < \Pi(l)) \end{aligned}$$

und mit Lemma 2.3.19 ist auch hier die Identität das Argmin. Bei den Summanden der Form $E_y D_{ij}^\sigma D_{il}^\sigma$ in der zweiten Summe ergibt sich

$$\begin{aligned} E_y D_{ij}^\sigma D_{il}^\sigma &= \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(i) < \sigma(l)\}} \cdot P_y(\Pi(i) > \Pi(j), \Pi(i) > \Pi(l)) \\ &\quad + \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(i) < \sigma(l)\}} \cdot P_y(\Pi(i) < \Pi(j), \Pi(i) > \Pi(l)) \\ &\quad + \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(i) > \sigma(l)\}} \cdot P_y(\Pi(i) > \Pi(j), \Pi(i) < \Pi(l)) \\ &\quad + \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\sigma(i) > \sigma(l)\}} \cdot P_y(\Pi(i) < \Pi(j), \Pi(i) < \Pi(l)). \end{aligned}$$

Hier ist eine pauschale Aussage wie bei den vorherigen Summanden nur beim Vergleich von $P_y(\Pi(i) > \Pi(j), \Pi(i) > \Pi(l))$ und $P_y(\Pi(i) < \Pi(j), \Pi(i) < \Pi(l))$ möglich (erste Aussage in Lemma 2.3.17). Beim Vergleich von $P_y(\Pi(i) > \Pi(j), \Pi(i) > \Pi(l))$ mit den beiden anderen Wahrscheinlichkeiten betrachten wir nur $P_y(\Pi(i) < \Pi(j), \Pi(i) > \Pi(l))$, die Argumentation für den anderen Fall ist analog.

Falls also $P_y(\Pi(i) > \Pi(j), \Pi(i) > \Pi(l)) > P_y(\Pi(i) < \Pi(j), \Pi(i) > \Pi(l))$ gilt, ergibt sich mit Lemma 2.3.17

$$\begin{aligned} &P_y(\Pi(i) > \Pi(j), \Pi(i) > \Pi(l)) + P_y(\Pi(i) > \Pi(j), \Pi(j) > \Pi(l)) \\ &\leq P_y(\Pi(i) < \Pi(j), \Pi(j) > \Pi(l)) + P_y(\Pi(i) < \Pi(j), \Pi(i) > \Pi(l)), \end{aligned}$$

der größere Wert für die Identität in der zweiten Summe wird also durch einen kleineren Wert in der dritten Summe ausgeglichen. Insgesamt folgt nun die Behauptung. \square

| Permutation | \bar{d}_τ | \bar{d}_τ^2 |
|--------------------|----------------|------------------|
| id | 57.18 | 3344.18 |
| π_{Ref} | 76.24 | 5956.42 |
| π_{MC} | 104.68 | 11215.3 |

Tabelle 2.7. *Mittlere Tau-Abstände und mittlere quadrierte Tau-Abstände zu 100 gleichverteilt ausgewählten Permutationen*

Bemerkung 2.3.21. Mit dem gleichen Vorgehen wie im Beweis von Satz 2.3.12 kann auch bei MMSD gezeigt werden, dass das Minimum für exponentialverteilte Lebensdauern nicht eindeutig ist. Die nach Bemerkung 2.3.13 für MMD berechnete Größe ist hier zumindest eine Unterschranke für die Anzahl der „minimierenden“ Permutationen. Auch eine analoge Aussage zu Satz 2.3.15 ergibt sich entsprechend. \triangleleft

Abschließendes Beispiel

Wir illustrieren auch an dieser Stelle wieder die bewiesenen Resultate anhand einer auf den Komponistendaten basierenden Simulation. Während in den anderen Abschnitten meist die Annäherung an Grenzwertaussagen für endliches $n \in \mathbb{N}$ untersucht wird, verwenden wir hier den Mittelwert einer Stichprobe anstelle des Erwartungswertes. Die ursprünglichen Aussagen über letzteren sind dabei keine Grenzwertaussagen, gelten also insbesondere für alle $n \in \mathbb{N}$.

Beispiel 2.3.22. Nach Lemma 2.3.5 (a) in Verbindung mit den in Beispiel 2.3.31 weiter unten genannten Werten gibt es für die Komponisten $\#S'_{22} \approx 1.316819 \cdot 10^{16}$ zulässige Permutationen. Von diesen wurden durch den Algorithmus *UnifPerm* aus Abb. 2.2 gleichverteilt $k = 100$ Permutationen ausgewählt und ihr durchschnittlicher Abstand \bar{d}_τ sowie ihr durchschnittlicher quadrierter Abstand \bar{d}_τ^2 zu drei festen Permutationen bestimmt. Bei diesen handelt es sich um die Identität, eine vorher gleichverteilt ausgewählte Referenzpermutation π_{Ref} und die sich durch den Algorithmus *MaxCross* in Abb. 2.1 ergebende Permutation π_{MC} . Verwendung der letztgenannten Permutation führt nach Bemerkung 2.3.11 (b) zu einer Oberschranke für den erwarteten (nicht quadrierten) Abstand; dies sollte sich auch bei der Simulation bemerkbar machen. Die Ergebnisse sind in Tabelle 2.7 aufgeführt, offenbar sind sowohl \bar{d}_τ als auch \bar{d}_τ^2 bei der Identität am kleinsten und bei π_{MC} am größten. Diese Beobachtungen passen gut zu den oben bewiesenen Optimalitätsresultaten. Außerdem sind die großen Werte für π_{MC} bei \bar{d}_τ^2 ein Hinweis darauf, dass auch der mittlere quadrierte Abstand zu einer gleichverteilt ausgewählten Permutation durch π_{MC} maximiert werden könnte. Da ein solches Resultat für unsere

Untersuchungen aber uninteressant wäre, geben wir uns hier mit dieser Beobachtung zufrieden, ohne sie weiter zu untersuchen. \triangleleft

2.3.4. Asymptotik bei Exponentialverteilungen

Bisher war die Anzahl n der Geburts- und Todesdaten konstant. Wir geben nun diese Annahme auf und untersuchen für exponentialverteilte Lebensdauern die asymptotische Verteilung des resultierenden Fehlers $d_\tau(\text{id}, \Pi_n)$, wenn anstelle der korrekten Zuordnung $\Pi_n \in S'_n$ die Identität verwendet wird, die nach Satz 2.3.10 eine MMD-Retrodiktion in diesem Fall ist. Wie weiter oben bei MMSD hängen die Ergebnisse nur von $\Pi_n \sim \text{unif}(S'_n)$ für alle $n \in \mathbb{N}$ ab, so dass wir allgemeiner diese Situation betrachten. Wir beginnen mit einem Resultat für die exakte Fehler-Verteilung.

Satz 2.3.23. *Für $n \in \mathbb{N}$ seien $\Pi_n \sim \text{unif}(S'_n)$ und $(r_i)_{i=1}^n$ die zugehörigen kleinsten zulässigen Indizes. Dann gilt*

$$d_\tau(\text{id}, \Pi_n) \sim *_{j=1}^n \text{unif}\{0, \dots, j - r_j\}.$$

Beweis. Für $D_j := \#\{i : 1 \leq i < j, \Pi(i) > \Pi(j)\}$ folgt $d_\tau(\text{id}, \Pi) = \sum_{j=1}^n D_j$ mit einer der alternativen Berechnungsformen aus Anhang A. Wir zeigen nun, dass D_1, \dots, D_n unabhängig sind und $D_j \sim \text{unif}\{0, \dots, j - r_j\}$ für alle j ist, dann ergibt sich die Behauptung.

Zunächst einmal gilt

$$D_n = i \Leftrightarrow \Pi(n) = n - i \quad \text{für } i \in \{0, \dots, n - r_n\},$$

da für $D_n = i$ von $\Pi(1), \dots, \Pi(n - 1)$ genau i größer als $\Pi(n)$ sein müssen, und dies genau dann passiert, wenn $\Pi(n)$ den $(i + 1)$ -größten Wert aus seinem Wertebereich annimmt. Nach Lemma 2.3.5 (b) hat $\Pi(n)$ eine $\text{unif}\{r_n, \dots, n\}$ -Verteilung, also ist D_n $\text{unif}\{0, \dots, r_n - n\}$ -verteilt.

Im Folgenden sei nun $r_i \leq \pi_i \leq n$ für alle i , $\pi_i \neq \pi_j$ für $i \neq j$ und $1 \leq k < n$. Ferner bezeichnen wir bei gegebenen Werten $\Pi(n) = \pi_n, \dots, \Pi(k + 1) = \pi_{k+1}$ die aufsteigend geordneten Elemente von $\{r_k, \dots, n\} \setminus \{\pi_{k+1}, \dots, \pi_n\}$ mit $j_0^{(k)} < \dots < j_{k-r_k}^{(k)}$ und setzen $j_i^{(n)} := n - i$ für alle $i \in \{0, \dots, n - r_n\}$. Dann ergibt sich mit der analogen Argumentation wie oben

$$D_k = i \Leftrightarrow \Pi(k) = j_i^{(k)} \quad \text{für } i \in \{0, \dots, k - r_k\}.$$

Nach Lemma 2.3.5 (c) gilt nun für $(u_k, \dots, u_n) \in \times_{l=k}^n \{0, \dots, l - r_l\}$

$$\begin{aligned} &P(D_k = u_k, \dots, D_n = u_n) \\ &= P(D_k = u_k | D_{k+1} = u_{k+1}, \dots, D_n = u_n) \end{aligned}$$

$$\begin{aligned}
& \cdot P(D_{k+1} = u_{k+1} | D_{k+2} = u_{k+2}, \dots, D_n = u_n) \cdot \dots \cdot P(D_n = u_n) \\
&= P(\Pi(k) = j_{u_k}^{(k)} | \Pi(k+1) = j_{u_{k+1}}^{(k+1)}, \dots, \Pi(n) = j_{u_n}^{(n)}) \\
& \quad \cdot P(\Pi(k+1) = j_{u_{k+1}}^{(k+1)} | \Pi(k+2) = j_{u_{k+2}}^{(k+2)}, \dots, \Pi(n) = j_{u_n}^{(n)}) \cdot \dots \cdot P(\Pi(n) = j_{u_n}^{(n)}) \\
&= \frac{1}{k - r_k + 1} \cdot \frac{1}{k + 1 - r_{k+1} + 1} \cdot \dots \cdot \frac{1}{n - r_n + 1} = \prod_{l=k}^n \frac{1}{l - r_l + 1}. \tag{2.20}
\end{aligned}$$

Nutzt man im Folgenden obiges Zwischenergebnis beim Übergang von der ersten zur zweiten Zeile, so erhält man für $u_k \in \{0, \dots, k - r_k\}$

$$\begin{aligned}
P(D_k = u_k) &= \sum_{u_{k+1}, \dots, u_n} P(D_k = u_k, \dots, D_n = u_n) \\
&= \sum_{u_{k+1}=0}^{k-r_k} \dots \sum_{u_n=0}^{n-r_n} \prod_{l=k}^n \frac{1}{l - r_l + 1} \\
&= \prod_{l=k}^n \frac{1}{l - r_l + 1} \cdot \prod_{l=k+1}^n (l - r_l + 1) = \frac{1}{k - r_k + 1},
\end{aligned}$$

also die behauptete Gleichverteilung. Setzt man dieses Ergebnis nun für die Faktoren in (2.20) ein und betrachtet den Fall $k = 1$, dann ergibt sich

$$P(D_1 = u_1, \dots, D_n = u_n) = \prod_{l=1}^n P(D_l = u_l)$$

und damit auch die Unabhängigkeit. \square

Bemerkung 2.3.24. Welche anschauliche Interpretation gibt es für die im vorhergehenden Satz (und auch später) vorkommenden Größen $(j - r_j)_{j \in \mathbb{N}}$?

Für festes $j \in \mathbb{N}$ sind vor der Geburt der j -ten Person zur Zeit x_j natürlich $j - 1$ andere Personen geboren worden. Da r_j der kleinste für x_j zulässige Index ist, müssen zu diesem Zeitpunkt genau $r_j - 1$ von diesen bereits wieder gestorben sein. Damit gibt $j - 1 - (r_j - 1) = j - r_j$ die Anzahl der Personen an, die vor Person j geboren wurden und gleichzeitig zu deren Geburt noch leben. (In der Warteschlangeninterpretation ist dies die Anzahl der Kunden im System, die der j -te Kunde vorfindet.) Diese Anzahl ist genau um 1 kleiner als die Anzahl der für den j -ten Geburtstag noch verfügbaren Todesdaten, wenn man den Geburtstagen in absteigender Reihenfolge zulässige Todesdaten zuteilt (zulässige Daten für die „Zeitgenossen“ von j zuzüglich eines Datums für j selbst), so dass ihr Auftauchen in Satz 2.3.23 plausibel wird: Jedes der möglichen Daten führt auf einen anderen Wert des entsprechenden Summanden in Kendalls Tau-Abstand, dieser Wert richtet sich nur nach der relativen Position des gewählten Datums innerhalb der noch verfügbaren und die Gleichverteilung auf den Permutationen induziert eine Gleichverteilung des genannten Summanden auf den möglichen Werten. Tatsächlich ist dies die Beweisidee von Satz 2.3.23. \triangleleft

Wir können also für jedes $n \in \mathbb{N}$ die exakte Fehler-Verteilung $d_\tau(\text{id}, \Pi_n)$ angeben, allerdings ist die Berechnung von konkreten Wahrscheinlichkeiten aufwändig, wie man z.B. an Beispiel 2.3.31 weiter unten sehen kann. Auf der anderen Seite deutet die Darstellung als Summe unabhängiger Zufallsvariablen auf die Anwendbarkeit des zentralen Grenzwertsatzes hin. Bevor wir nun unter bestimmten Voraussetzungen dessen Gültigkeit nachweisen können, machen wir in einer Bemerkung auf mögliche Schwierigkeiten aufmerksam und treffen Annahmen, um diese zu vermeiden.

Bemerkung 2.3.25. Für den folgenden Satz gelten zwei Einschränkungen; diese sollen den Aufwand bei der Notation verringern und damit die Lesbarkeit erhöhen. Zum einen gehen wir davon aus, dass beim Übergang von S'_n nach S'_{n+1} Daten x_{n+1} und y_{n+1} mit $x_n \leq x_{n+1} < y_{n+1}$ zu den vorliegenden $x_1, \dots, x_n, y_1, \dots, y_n$ hinzukommen. Für S'_n und S'_{n+1} treten also dieselben Werte r_1, \dots, r_n auf, daher hängen die kleinsten zulässigen Indizes in diesem Sinne nicht von n ab. Ein Einfügen einer „neuen“ Geburt zwischen zwei schon beobachtete wird folglich explizit ausgeschlossen. Zum anderen wird ein Ausdruck der Art $(d_\tau(\text{id}, \Pi_n) - \text{Ed}_\tau(\text{id}, \Pi_n)) / \sqrt{\text{var}(d_\tau(\text{id}, \Pi_n))}$ in einem Grenzwertresultat verwendet, wobei der Nenner für einzelne n den Wert 0 annehmen kann. Da die Verteilung in diesem Fall aber auf ihren Erwartungswert konzentriert ist, verschwindet dann auch der Zähler, so z.B. bei $n = 1$ ($S_1 = \{\text{id}\}$). Deswegen sei dieser Ausdruck so aufzufassen, dass wir die Betrachtung für den ersten Wert von n beginnen, bei dem der Nenner positiv ist. Nach dem folgenden Lemma gilt dies dann auch für alle folgenden n . \triangleleft

Lemma 2.3.26. *Unter dem im ersten Teil der vorhergehenden Bemerkung beschriebenen Vorgehen sei $n_0 := \min\{n \in \mathbb{N} : \text{var}(d_\tau(\text{id}, \Pi_n)) > 0\}$. Dann gilt*

$$\text{var}(d_\tau(\text{id}, \Pi_n)) > 0 \quad \text{für alle } n \geq n_0.$$

Beweis. Aus dem Beweis von Satz 2.3.23 ergibt sich $d_\tau(\text{id}, \Pi_n) = \sum_{j=1}^n D_j$ mit unabhängigen, $\text{unif}\{0, \dots, j - r_j\}$ -verteilten Summanden D_j . Geht man nun wie in der Bemerkung beschrieben vor, ändern sich die kleinsten Indizes r_1, \dots, r_{n_0} beim Übergang von n_0 zu einem größeren n nicht, und damit auch nicht die Varianzen von D_1, \dots, D_{n_0} . Ferner sind Varianzen immer nichtnegativ, also ergibt sich wegen der Unabhängigkeit der D_j für $n > n_0$

$$\text{var}(d_\tau(\text{id}, \Pi_n)) = \sum_{j=1}^{n_0} \text{var}(D_j) + \sum_{n_0+1}^n \text{var}(D_j) \geq \sum_{j=1}^{n_0} \text{var}(D_j) > 0$$

und es folgt die Behauptung. \square

Nach diesen Vorarbeiten beweisen wir nun das angekündigte Grenzwertresultat.

Satz 2.3.27. Für jedes $n \in \mathbb{N}$ sei unter den Einschränkungen der vorherigen Bemerkung $\Pi_n \sim \text{unif}(S'_n)$ und $(r_j)_{j=1}^n$ bezeichne die zugehörigen kleinsten zulässigen Indizes. Gilt dann für $a_j := j - r_j$

$$\lim_{n \rightarrow \infty} \frac{\sum_{j=1}^n a_j^4}{(\sum_{j=1}^n a_j^2)^2} = 0$$

so folgt

$$\frac{d_\tau(\text{id}, \Pi_n) - \text{Ed}_\tau(\text{id}, \Pi_n)}{\sqrt{\text{var}(d_\tau(\text{id}, \Pi_n))}} \xrightarrow{d} U, U \sim N(0,1),$$

d.h. der Fehler bei Ersetzen der wahren Zuordnung durch die Identität ist in diesem Fall asymptotisch normalverteilt.

Beweis. Wir benutzen $d_\tau(\text{id}, \Pi_n) = \sum_{j=1}^n D_j$ mit $D_j := \#\{i : 1 \leq i < j, \Pi_n(i) > \Pi_n(j)\}$, dabei sind die (D_j) nach dem Beweis von Satz 2.3.23 unabhängig und haben endliche Varianz. Für die Gültigkeit des zentralen Grenzwertsatzes nach [Bil95, S. 362] muss noch die dort genannte Lyapunov-Bedingung

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^n \text{E}|D_j - \text{ED}_j|^{2+\delta} = 0 \quad \text{für ein } \delta > 0$$

mit $s_n^2 := \text{var}\left(\sum_{j=1}^n D_j\right)$ erfüllt sein. Wir zeigen dies im Folgenden für den Fall $\delta = 2$, dann folgt die Behauptung.

Wegen $D_j \sim \text{unif}\{0, \dots, a_j\}$ mit $a_j := j - r_j$ (vgl. Beweis von Satz 2.3.23) gilt $\text{ED}_j = a_j/2$ und $\text{var}(D_j) = a_j(a_j + 2)/12$. Wegen der Unabhängigkeit der (D_j) und $a_j \geq 0$ folgt dann für alle j

$$s_n^2 = \sum_{j=1}^n \text{var}(D_j) = \frac{1}{12} \sum_{j=1}^n (a_j^2 + 2a_j) \geq \frac{1}{12} \sum_{j=1}^n a_j^2.$$

Für $W_j := D_j - \text{ED}_j$ ist $W_j \sim \text{unif}\{-\frac{a_j}{2}, -\frac{a_j}{2} + 1, \dots, \frac{a_j}{2}\}$ und damit

$$\begin{aligned} \text{E}|D_j - \text{ED}_j|^4 &= \text{EW}_j^4 = \frac{1}{a_j + 1} \sum_{l=0}^{a_j} \left(-\frac{a_j}{2} + l\right)^4 \\ &= \frac{-a_j}{30(a_j + 1)} + \frac{a_j^5}{80(a_j + 1)} + \frac{a_j^4}{16(a_j + 1)} + \frac{a_j^3}{12(a_j + 1)} \\ &\leq \frac{13}{80} a_j^4. \end{aligned}$$

Daraus ergibt sich

$$0 \leq \frac{1}{s_n^4} \sum_{j=1}^n \text{E}|D_j - \text{ED}_j|^4 \leq \frac{\frac{13}{80} \sum_{j=1}^n a_j^4}{\left(\frac{1}{12} \sum_{j=1}^n a_j^2\right)^2},$$

nach Voraussetzung konvergiert die rechte Seite für wachsendes n gegen 0. \square

Bemerkung 2.3.28. Bei $S'_n = S_n$ gilt $r_j = 1$ und folglich $a_j = j - 1$ für alle j , dann kann der vorhergehende Satz wegen

$$\frac{\sum_{j=1}^n a_j^4}{(\sum_{j=1}^n a_j^2)^2} \sim c \cdot \frac{n^5}{(n^3)^2} \rightarrow 0$$

für wachsendes n und eine Konstante $c > 0$ angewendet werden. Damit ergibt sich die bei [Dia88, S. 117] erwähnte asymptotische Normalität von Kendalls Tau-Abstand bei Zulässigkeit aller Permutationen als Spezialfall des Satzes. \triangleleft

Ein weiteres Beispiel für eine Situation, in der die Voraussetzungen für die asymptotische Normalität erfüllt sind, liefert das folgende Lemma.

Lemma 2.3.29. *Es seien $a_1, a_2, \dots \in \mathbb{N}_0$. Gibt es dann ein $K \in \mathbb{N}$, so dass $a_i \leq K$ für alle $i \in \mathbb{N}$ ist, und gilt $\lim_{n \rightarrow \infty} \sum_{i=1}^n a_i = \infty$, so folgt*

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n a_i^4}{(\sum_{i=1}^n a_i^2)^2} = 0.$$

Beweis. Wir definieren $b_i := a_i^2$ für alle i . Dann ist $b_i > 0 \Leftrightarrow a_i > 0$ und unter den obigen Voraussetzungen gilt $b_i \leq K^2$ für alle i und $\lim_{n \rightarrow \infty} \sum_{i=1}^n b_i = \infty$. Weiter seien $M_n := \{i = 1, \dots, n : b_i > 0\}$ und $m_n := \#M_n$, dann folgt

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n b_i^2}{(\sum_{i=1}^n b_i)^2} = \lim_{n \rightarrow \infty} \frac{\sum_{i \in M_n} b_i^2}{(\sum_{i \in M_n} b_i)^2} \leq \lim_{n \rightarrow \infty} \frac{\sum_{i \in M_n} K^4}{(\sum_{i \in M_n} 1)^2} = \lim_{n \rightarrow \infty} \frac{K^4 m_n}{m_n^2} = 0. \quad \square$$

Bemerkung 2.3.30. Die Voraussetzungen des vorhergehenden Lemmas sind recht restriktiv. Wir machen mit den folgenden Überlegungen plausibel, dass aber zumindest der Teil $\lim_{n \rightarrow \infty} \sum_{j=1}^n (j - r_j) = \infty$ auch notwendig für die Gültigkeit des bewiesenen Grenzwertresultats ist. Da es sich für alle $j \in \mathbb{N}$ bei $j - r_j$ um Elemente aus \mathbb{N}_0 handelt, ist die genannte Forderung äquivalent zu $j - r_j > 0$ für unendlich viele Werte von j . Nehmen wir im Gegensatz dazu an, dass es ein $n_0 \in \mathbb{N}$ mit $j - r_j = 0$ für alle $j \geq n_0$ gibt, so finden ab dem Zeitpunkt x_{n_0} Geburten und Todesfälle im Wechsel statt. Nach der anschaulichen Interpretation von $j - r_j$ aus Bemerkung 2.3.24 trifft also jeder neu Geborene bei seiner Geburt keine „Zeitgenossen“ an und stirbt, bevor der nächste geboren wird. Dementsprechend ist ab diesem n_0 nur noch die Identität zulässig und der Prozess, dass sich die Verteilung von Kendalls Tau-Abstand der Normalverteilung annähert, wird abgebrochen. Oben genannte Bedingung ist also tatsächlich notwendig. \triangleleft

Auch diesen Abschnitt beschließen wir mit einer Veranschaulichung der Resultate.

Beispiel 2.3.31. Für das Komponistenbeispiel wird untersucht, wie gut die Annäherung durch die Normalverteilung in diesem Fall ist. Es sind $n = 22$ sowie

$$(r_1, \dots, r_{22}) = (1, 1, 1, 1, 1, 2, 4, 4, 6, 6, 7, 7, 7, 7, 10, 10, 10, 10, 13, 13, 13)$$

und mit $a_j := j - r_j$ ergibt sich für $j = 1, \dots, 22$

$$(a_1, \dots, a_{22}) = (0, 1, 2, 3, 4, 4, 3, 4, 3, 4, 3, 4, 4, 5, 6, 7, 8, 6, 7, 8, 9, 7, 8, 9).$$

Nun ist für $\Pi_{22} \sim \text{unif}(S'_{22})$ der Abstand $d_\tau(\text{id}, \Pi_{22}) = \sum_{j=1}^{22} D_j$ mit $D_j \sim \text{unif}\{0, \dots, a_j\}$ und man erhält die wahrscheinlichkeitserzeugende Funktion

$$\hat{p}_{d_\tau(\text{id}, \Pi_{22})}(z) = \prod_{j=1}^{22} \frac{1}{a_j + 1} (1 + z^1 + \dots + z^{a_j}).$$

Aus dieser kann für $k = 0, 1, \dots, \sum_{j=1}^{22} a_j$ die exakte Wahrscheinlichkeit $P(d_\tau(\text{id}, \Pi_{22}) = k)$ als Koeffizient von z^k abgelesen werden, dabei ist u.U. die Verwendung von Maple hilfreich. Wie schon oben bemerkt, ist dies jedoch einigermaßen aufwändig. Andererseits erhält man mit $ED_j = \frac{a_j}{2}$ und $\text{var}(D_j) = \frac{a_j(a_j+2)}{12}$, dass $E(d_\tau(\text{id}, \Pi_{22})) = 56$ und $\sqrt{\text{var}(d_\tau(\text{id}, \Pi_{22}))} = 8.82232$ ist. In Abb. 2.3 sind die Dichte einer Normalverteilung mit diesen Parametern und die obigen exakten Wahrscheinlichkeiten dargestellt. Offenbar ist die Approximation durch die Normalverteilung sehr gut und die Kurve verläuft nahe bei den Werten der exakten Wahrscheinlichkeiten.

Wir erklären im Folgenden das Zustandekommen dieser (nicht selbstverständlichen) Beobachtung. Nach Satz 2.3.27 ist Kendalls Tau-Abstand unter bestimmten Bedingungen asymptotisch normalverteilt. Anscheinend ist im Beispiel die Annäherung schon recht genau und mit den oben bestimmten Werten $\mu = 56$ und $\sigma = 8.82232$ ergibt sich

$$\begin{aligned} P(d_\tau(\text{id}, \Pi_{22}) = k) &\approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \int_{(k-\frac{1}{2}-\mu)/\sigma}^{(k+\frac{1}{2}-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \\ &\approx \frac{1}{\sigma} \phi_{0,1}\left(\frac{k-\mu}{\sigma}\right) = \phi_{\mu, \sigma^2}(k) \end{aligned}$$

mit ϕ_{μ, σ^2} als Dichte einer $N(\mu, \sigma^2)$ -Verteilung. Bei der zweiten Approximation wurde ausgenutzt, dass über ein Intervall der Länge $\frac{1}{\sigma}$ integriert wird und σ einigermaßen groß ist, der Integrand also durch eine Gerade angenähert werden kann. Anschaulich gesprochen zerteilt man die wegen der recht großen Standardabweichung sehr flache Kurve der Normalverteilung in viele (genauer $\sum_{j=1}^{22} a_j = 112$) Abschnitte, die dann für sich betrachtet jeweils fast eine Gerade sind. Für kleinere Werte von σ würde die obige Argumentation nicht mehr funktionieren, so dass die Dichte nicht zwangsläufig durch die Werte der exakten Wahrscheinlichkeiten verlaufen muss. Es gilt z.B. $\phi_{56, 0.1}(56) > 1$, damit ist die Dichte an dieser Stelle verschieden von jeder Wahrscheinlichkeit. \triangleleft

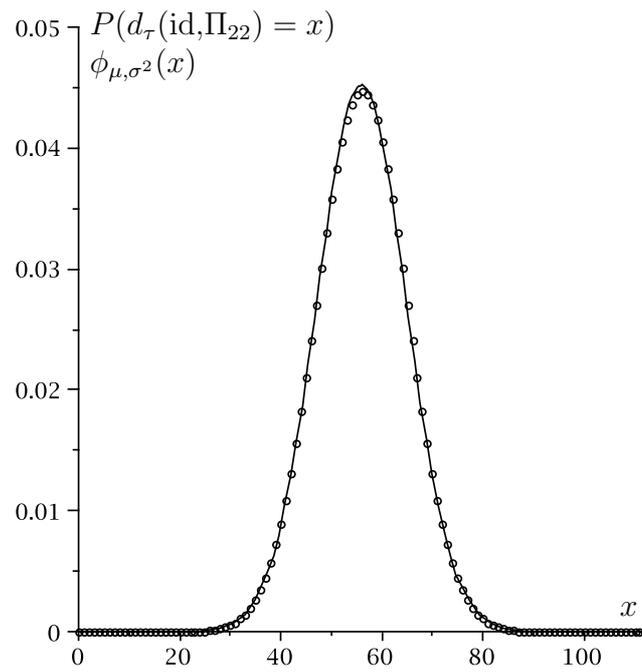


Abbildung 2.3. Vergleich der Dichte ϕ_{μ, σ^2} einer $N(\mu, \sigma^2)$ -Verteilung für $\mu = 56$ und $\sigma = 8.82232$ (durchgezogene Linie) mit der exakten Wahrscheinlichkeitsmassenfunktion (Kreise) im Komponistenbeispiel

2.4. Ein Test auf Exponentialverteilung

Wir schließen dieses Kapitel mit der Entwicklung eines Tests, der in einem bestimmten Modell die Hypothese überprüft, dass die Lebensdauern zu vorliegenden Daten aus einer Exponentialverteilung stammen. Dazu weisen wir zunächst auf folgendes technisches Detail hin. Bei Tests ist die Verwendung von α für das Signifikanzniveau üblich; dies kollidiert mit unserer bisherigen Bezeichnung für den Gestaltparameter einer Gamma-Verteilung. Deshalb verwenden wir in diesem Abschnitt für letzteren die Bezeichnung η , sprechen also von einer $\Gamma(\eta, \lambda)$ -Verteilung. Unsere Untersuchungen beginnen mit der Situation im bisher betrachteten Problem, diese wird im Verlauf aber abgewandelt. Nach Bemerkung 2.2.12 gilt für exponentialverteilte Lebensdauern $\Pi \sim \text{unif}(S'_n)$ und durch die Sätze 2.3.23 und 2.3.27 kennen wir dann die exakte und unter gewissen Umständen auch die asymptotische Verteilung von $d_\tau(\text{id}, \Pi)$. Zur Motivation der weiteren Überlegungen betrachten wir das folgende Beispiel.

Beispiel 2.4.1. Eine Supermarktkette betreibt mehrere Filialen. In einer davon wurde zu Marktforschungszwecken der Eingangsbereich so umgebaut, dass wie in Beispiel 1.1.2 beschrieben Ankunfts- und Abgangszeiten der Kunden ohne Zuordnung untereinander gespeichert werden können. Aus diesen Daten wurde die mittlere Aufenthaltszeit berechnet (siehe dazu Gleichung (3.5) weiter unten) und es ergab sich der Wert λ^{-1} . Das Management möchte nun aus Kostengründen vermeiden, auch in allen anderen Filialen den erwähnten Umbau vorzunehmen. Deswegen wurden die Einkaufskörbe gut sichtbar durchnummeriert, so dass ein Mitarbeiter ohne Beeinträchtigung der Kunden die Reihenfolge, in der die Körbe aufgenommen und wieder abgestellt werden, als eine einzige Zahlenfolge aufschreiben kann. Zur Vereinfachung nehmen wir an, dass die Körbe in aufsteigender Nummerierung verwendet werden und jeder Kunde einen Korb benutzt. Es ergibt sich also eine Zahlenfolge der Form $1, 2, 3, 4, 2, 1, 5, 3, 5, 4, \dots$. Das erste Auftreten einer Zahl bedeutet jeweils, dass der entsprechende Korb aufgenommen wurde, und das zweite Auftauchen dessen Rückgabe. Ein genaues Vermerken von Ankunfts- und Abgangszeiten durch einen einzelnen Mitarbeiter wäre dagegen bei entsprechend großem Kundenaufkommen nicht ohne weiteres möglich. Nun soll überprüft werden, ob die beobachtete Zahlenfolge mit der Hypothese vereinbar ist, dass die (nicht vorliegenden) Aufenthaltszeiten der Kunden aus einer Exponentialverteilung mit Parameter λ stammen. Ist dies der Fall, wird kein Umbau vorgenommen. ◁

Bei Unterstellung des gleichen Zusammenhangs von Geburts- und Todesdaten wie bisher soll nun ein Test für die Situation entwickelt werden, dass nur die wahre Permutation und der Raum der zulässigen Permutationen bekannt sind, nicht aber die Daten selbst. Diese Fragestellung ist äquivalent zu der im Beispiel genannten Formulierung, da sich die wahre Permutation und S'_n sofort aus der dort beschriebenen Zahlenfolge bestimmen

lassen und umgekehrt. Wir beginnen damit, einen Test für die Hypothese

H : „Die Lebensdauern sind exponentialverteilt mit Parameter λ “

gegen eine der Alternativen

K_1 : „Die Lebensdauern stammen aus einer Verteilung mit streng logkonkaver Dichte“

K_2 : „Die Lebensdauern stammen aus einer Verteilung mit streng logkonvexer Dichte“

bei bekanntem $\lambda > 0$ zu konstruieren. Wie oben lassen wir in $P(\Pi = \pi | \cdot)$ die Bedingung weg und schreiben hier sogar nur $P(\Pi = \pi)$. Dann ergibt sich nach Lemma 2.3.8 mit $K \in \{K_1, K_2\}$

$$\frac{P_K(\Pi = \pi)}{P_H(\Pi = \pi)} = \frac{\#S'_n}{\sum_{\pi' \in S'_n} \prod_{i=1}^n f_Z(y_{\pi'(i)} - x_i)} \prod_{i=1}^n f_Z(y_{\pi(i)} - x_i)$$

als Testgröße für einen Likelihoodquotienten-Test. In diese geht π nur über das Produkt $\prod_{i=1}^n f_Z(y_{\pi(i)} - x_i)$ ein. Durch Logarithmieren und Verwendung von Lemma 2.2.10 wird deutlich, dass die Testgröße mit Wahrscheinlichkeit 1 eine entweder streng monoton steigende (bei strenger Logkonvexität) oder streng monoton fallende (bei strenger Logkonkavität) Funktion in der Anzahl der Inversionen von Π ist. Diese Anzahl wiederum entspricht nach Anhang A dem Abstand $d_\tau(\text{id}, \Pi)$ und ist nach Satz 2.3.27 unter bestimmten Bedingungen asymptotisch normalverteilt. Damit kann ein Test angegeben werden, der nur auf der wahren Permutation Π und auf S'_n basiert, und dessen asymptotische Testgrößenverteilung wir recht einfach bestimmen können. Aus diesen Überlegungen ergibt sich sofort der folgende Satz.

Satz 2.4.2. *Soll die Hypothese*

H : „Die Lebensdauern sind exponentialverteilt mit Parameter λ “

gegen eine der Alternativen

K_1 : „Die Lebensdauern stammen aus einer Verteilung mit streng logkonkaver Dichte“

K_2 : „Die Lebensdauern stammen aus einer Verteilung mit streng logkonvexer Dichte“

für unbekanntes $\lambda > 0$ getestet werden, so sind für K_1

$$\phi(d_\tau(\text{id}, \Pi)) = \begin{cases} 0, & d_\tau(\text{id}, \Pi) > c_1, \\ \gamma_1, & d_\tau(\text{id}, \Pi) = c_1, \\ 1, & d_\tau(\text{id}, \Pi) < c_1 \end{cases} \quad \text{und} \quad \phi(d_\tau(\text{id}, \Pi)) = \begin{cases} 1, & d_\tau(\text{id}, \Pi) > c_2, \\ \gamma_2, & d_\tau(\text{id}, \Pi) = c_2, \\ 0, & d_\tau(\text{id}, \Pi) < c_2 \end{cases}$$

für K_2 Tests zum Niveau $\alpha \in (0,1)$ mit Randomisierungskonstante γ_1 bzw. $\gamma_2, \gamma_i \in [0,1)$, und kritischem Wert c_1 bzw. c_2 . Letztere werden aus

$$P_H(d_\tau(\text{id}, \Pi) < c_1) + \gamma_1 P_H(d_\tau(\text{id}, \Pi) = c_1) = \alpha$$

bzw.

$$P_H(d_\tau(\text{id}, \Pi) > c_2) + \gamma_2 P_H(d_\tau(\text{id}, \Pi) = c_2) = \alpha$$

bestimmt. Bei Gültigkeit der jeweiligen Alternative ist die Ablehnungswahrscheinlichkeit echt größer als α .

Bemerkung 2.4.3. Durch entsprechende Anpassungen kann in der beschriebenen Situation mit derselben Testgröße wie im vorherigen Satz auch ein Test zum Niveau $\alpha \in (0,1)$ für das Problem mit

H : „Die Lebensdauern sind exponentialverteilt mit Parameter λ “

gegen

K : „Die Lebensdauern stammen aus einer Verteilung mit streng logkonkaver oder streng logkonvexer Dichte“

entwickelt werden. ◁

Die Verteilung der Testgröße in Satz 2.4.2 kann offenbar entweder exakt aus Satz 2.3.23 oder unter den entsprechenden Voraussetzungen asymptotisch aus Satz 2.3.27 bestimmt werden, im zweiten Fall ist $\gamma \approx 0$. Wir verwenden den entwickelten Test nun in dem Spezialfall, dass für die Lebensdauern die Alternative einer $\Gamma(\eta, \lambda)$ -Verteilung mit $\eta > 1$ getestet werden soll, und vergleichen ihn mit einer Abwandlung des bekannten, im Sinne des Neyman-Pearson-Lemmas besten Tests aus [Sho72] in der Situation, dass Realisierungen z_1, \dots, z_n der Lebensdauern Z_1, \dots, Z_n explizit vorliegen. Dieser hat die Testgröße $T_n := \prod_{i=1}^n (Z_i / \bar{Z}_{n,1})$ und lehnt ab, wenn ein gewisser kritischer Wert überschritten wird.

Bemerkung 2.4.4. Für die angekündigte Gegenüberstellung benutzen wir die asymptotische Verteilung von $d_\tau(\text{id}, \Pi)$ bei Gültigkeit der Hypothese, so dass auch beim „Vergleichstest“ eine asymptotische Aussage für die Verteilung der Testgröße ausreichend ist. Als weitere Vereinfachung nehmen wir λ als bekannt an. Ausgehend von T_n ergibt sich durch monoton wachsende Transformationen

$$\log(T_n) + n \log(\lambda) = \sum_{i=1}^n \log(\lambda Z_i) - n \log(\bar{Z}_{n,1}).$$

Für hinreichend großes n ist $\bar{Z}_{n,1}$ näherungsweise eine Konstante, deshalb verwenden wir $\tilde{T}_n := \sum_{i=1}^n \log(\lambda Z_i)$ als Testgröße und verwerfen die Hypothese, wenn diese einen gewissen Wert übersteigt. (Man kann dasselbe Ergebnis auch direkt durch monotone Umformungen des logarithmierten Likelihoodquotienten

$$\log \left(\frac{\prod_{i=1}^n f(z_i|\eta, \lambda)}{\prod_{i=1}^n f(z_i|1, \lambda)} \right)$$

für $f(z|\eta, \lambda) := \lambda^\eta z^{\eta-1} \exp(-\lambda z)/\Gamma(\eta)$ bei $z > 0$ herleiten.) Zur Bestimmung der asymptotischen Verteilung verwenden wir, dass für $X \sim \text{Exp}(\lambda)$ folgt $\lambda X \sim \text{Exp}(1)$. Somit ist \tilde{T}_n als Summe der Logarithmen von unabhängigen, $\text{Exp}(1)$ -verteilten Zufallsvariablen nach dem zentralen Grenzwertsatz asymptotisch normalverteilt. Für $X \sim \Gamma(\eta, 1)$ ist

$$\begin{aligned} \text{E} \log(X) &= \int_0^\infty \log(x) \frac{1}{\Gamma(\eta)} x^{\eta-1} e^{-x} dx = \Psi(\eta), \\ \text{var}(\log(X)) &= \int_0^\infty (\log(x))^2 \frac{1}{\Gamma(\eta)} x^{\eta-1} e^{-x} dx - (\text{E} \log(X))^2 = \Psi'(\eta), \end{aligned}$$

so dass sich bei Gültigkeit der Hypothese, d.h. $\eta = 1$, für Erwartungswert und Varianz

$$\text{E} \log(Z_1) = \Psi(1) = -\gamma \approx -0.5772, \quad \text{var}(\log(Z_1)) = \Psi'(1) = \frac{1}{6}\pi^2$$

ergeben; dabei ist γ die Euler-Mascheroni-Konstante.

Im anschließenden Beispiel erfolgt nun ein Vergleich der beiden konstruierten Tests.

Beispiel 2.4.5. Die beiden beschriebenen Tests wurden im Rahmen einer Simulationsstudie verglichen. Der Skalenparameter λ hatte immer den Wert 1 und die Geburten stammten aus einem Poisson-Prozess mit Rate η^{-1} , d.h. die Rate wurde so angepasst, dass die mittlere Lebensdauer der mittleren Zeit zwischen zwei Geburten entsprach. Die kritischen Werte wurden in beiden Fällen über eine Approximation der jeweiligen Verteilung durch eine Normalverteilung bestimmt.

In Tabelle 2.8 finden sich die simulativ ermittelten Werte der Fehler erster (zur Kontrolle) und zweiter Art $\hat{\alpha}$ und $\hat{\beta}$ für verschiedene Werte von η und verschiedene Signifikanzniveaus α bei $n = 200$ und jeweils 1000 Wiederholungen. Dabei gehören Ergebnisse mit dem tiefgestellten Index K zum Test aus Satz 2.4.2 und solche mit Index A zum anderen Test. Im Vergleich sind die Ergebnisse des zweiten Tests beim Fehler 2. Art deutlich besser. Dies ist nicht überraschend, da der erste Test keine Kenntnis von λ voraussetzt und mit reduzierten Daten arbeitet; genauer wird nur die Rangfolge der realisierten Werte berücksichtigt. Diese Rangfolge ändert sich anscheinend bei Werten von η in der Nähe von 1 noch nicht signifikant gegenüber Werten aus einer Exponentialverteilung, deshalb wird nicht erkannt, dass die Hypothese falsch ist. Eine Absenkung des Signifikanzniveaus

| η | α | $\hat{\alpha}_K$ | $\hat{\alpha}_A$ | $\hat{\beta}_K$ | $\hat{\beta}_A$ |
|--------|----------|------------------|------------------|-----------------|-----------------|
| 1 | 0.1 | 0.097 | 0.109 | N/A | N/A |
| 1 | 0.05 | 0.051 | 0.054 | N/A | N/A |
| 1 | 0.01 | 0.013 | 0.013 | N/A | N/A |
| 1.1 | 0.05 | N/A | N/A | 0.896 | 0.464 |
| 1.2 | 0.05 | N/A | N/A | 0.787 | 0.003 |
| 1.5 | 0.05 | N/A | N/A | 0.433 | 0 |
| 1.5 | 0.1 | N/A | N/A | 0.299 | 0 |
| 2 | 0.05 | N/A | N/A | 0.085 | 0 |
| 3 | 0.05 | N/A | N/A | 0 | 0 |

Tabelle 2.8. *Vergleich von Fehler erster und zweiter Art der beiden Tests mittels mehrerer Simulationen*

auf 10% scheint dieses Problem zumindest etwas zu verringern. Ausgehend von diesen Ergebnissen kann der Test aus Satz 2.4.2 nicht für Fälle empfohlen werden, in denen es auf eine „scharfe“ Trennung zwischen der Exponentialverteilung und Gammaverteilungen mit einem Gestaltparameter in der Nähe von 1 ankommt. ◁

3. Schätzer für die Verzögerungsverteilung

In diesem Kapitel wandeln wir die bisherigen Annahmen ab und nehmen die Verteilung der „Verzögerungen“ (Z_i), die unter Beibehaltung der Benennung als Lebensdauern bezeichnet werden, als unbekannt an. Allerdings wird unterstellt, dass letztere aus einer bestimmten parametrischen Familie stammt. In dieser Situation werden die unbekannt-ten Werte der zugrunde liegenden Parameter anhand der beobachteten Daten geschätzt, während die Wiederherstellung der ursprünglichen Paare nicht untersucht wird. Wir bezeichnen diese Ausgangslage abkürzend auch als Geburten-und-Todesfälle-Problem. Dabei werden die Lebensdauern weiter als i.i.d. angenommen, während wir für die Geburtsdaten die drei Fälle betrachten, dass diese i.i.d. oder deterministisch sind oder aus einem Poisson-Prozess stammen. Wir beginnen mit einem Maximum-Likelihood-Verfahren und einem einfachen Momentenschätzer. Für den dritten der genannten Fälle wird anschließend der in der Einleitung erwähnte Zusammenhang zwischen dem betrachteten Problem und einer $M/G/\infty$ -Warteschlange ausgenutzt und es werden weitere Momentenschätzer angegeben. Bei allen im Folgenden genannten Dichten handelt es sich um Dichten bzgl. des Lebesgue-Maßes.

3.1. Die Likelihood-Methode

Im Verlauf wird schnell deutlich, dass die parametrische Likelihood-Methode für das vorliegende Problem in vielen Fällen zwar funktioniert und eine eindeutige Lösung liefert, aber ein sehr hoher Zeitaufwand für die Auswertung der (Log-)Likelihood-Funktion nötig werden kann. Wir stellen deshalb exemplarisch nur das Vorgehen bei gammaverteilten Lebensdauern vor, in anderen Fällen ist die Behandlung ähnlich. Die beiden unbekannt-ten Verteilungen sind die Verteilung der Geburten (X_i) und der Lebensdauern (Z_i) bei Vorliegen von Realisierungen der (X_i) und der Todeszeitpunkte (Y_i). Da die gemeinsame Dichte $f_{X,Y}$ auch durch das Produkt $f_{Y|X} \cdot f_X$ dargestellt werden kann, zerfällt die Bestimmung der Parameter in zwei Teilprobleme. Den Teil für die (X_i) behandeln wir nicht

weiter, da uns in erster Linie die Lebensdauern interessieren. In den nächsten Abschnitten sind die Geburtszeitpunkte entweder i.i.d. oder stammen aus einem Poisson-Prozess; in beiden Fällen können bekannte Verfahren verwendet werden. Stattdessen wird das Teilproblem zur Bestimmung des ML-Schätzers für die Parameter der Lebensdauer-Verteilung unter Verwendung von $f_{Y|X}$ untersucht. Wie auch schon weiter oben verzichten wir im Folgenden darauf, die Bedingung durch die (X_i) zu kennzeichnen, und nehmen die Geburtszeitpunkte als gegebene, feste Werte x_1, \dots, x_n an. Somit kann dieses Teilproblem auch dann bearbeitet werden, wenn die Verteilung der (X_i) nicht bekannt ist. Die gemeinsame Dichte der Ordnungsstatistiken (Y_i) erhält man sofort als Folgerung aus Satz 2.2.1.

Korollar 3.1.1. *Eine gemeinsame Dichte $f_Y(y) := f_{(Y_1, \dots, Y_n)}(y_1, \dots, y_n)$ für die Todeszeitpunkte Y_1, \dots, Y_n bei Vorliegen von Geburtszeitpunkten x_1, \dots, x_n ist gegeben durch*

$$f_Y(y) = \begin{cases} \sum_{\pi \in S'_n} \prod_{i=1}^n f_Z(y_{\pi(i)} - x_i), & y_1 < \dots < y_n, \\ 0, & \text{sonst.} \end{cases}$$

Für den untersuchten Spezialfall gammaverteilter Lebensdauern kann die entsprechende Dichte mithilfe des Korollars direkt angegeben werden.

Korollar 3.1.2. *Sind im Geburten-und-Todesfälle-Problem die Lebensdauern gammaverteilt mit den Parametern α und λ , so ergibt sich für die geordneten Todeszeitpunkte bei Vorliegen von x_1, \dots, x_n und mit der obigen Notation die Dichte*

$$f_Y(y) = \left(\frac{\lambda^\alpha}{\Gamma(\alpha)} \right)^n \cdot \exp \left(-\lambda \sum_{i=1}^n (y_i - x_i) \right) \cdot \sum_{\pi \in S'_n} \prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1}.$$

Beweis. Unter Verwendung von Korollar 3.1.1 ergibt sich

$$\begin{aligned} f_Y(y) &= \sum_{\pi \in S'_n} \prod_{i=1}^n f_Z(y_{\pi(i)} - x_i) \\ &= \sum_{\pi \in S'_n} \prod_{i=1}^n \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot (y_{\pi(i)} - x_i)^{\alpha-1} \cdot \exp(-\lambda(y_{\pi(i)} - x_i)) \\ &= \left(\frac{\lambda^\alpha}{\Gamma(\alpha)} \right)^n \cdot \exp \left(-\lambda \sum_{i=1}^n (y_i - x_i) \right) \cdot \sum_{\pi \in S'_n} \prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1}, \end{aligned}$$

dabei wurde $\sum_{i=1}^n y_{\pi(i)} = \sum_{i=1}^n y_i$ für alle $\pi \in S_n$ benutzt. □

Wir betrachten zunächst die Exponentialverteilung als Spezialfall der Gamma-Verteilung, da sich hier ein geschlossener Ausdruck für den ML-Schätzer angeben lässt. Anschließend wird der allgemeine Fall behandelt.

Satz 3.1.3. *Liegen im Geburten-und-Todesfälle-Problem $\text{Exp}(\lambda)$ -verteilte Lebensdauern und geordnete Geburts- und Sterbedaten x_1, \dots, x_n bzw. y_1, \dots, y_n vor, so ist*

$$\hat{\lambda}_{\text{ML}} := \frac{1}{\bar{y}_{n,1} - \bar{x}_{n,1}}$$

der Maximum-Likelihood-Schätzer für $\lambda \in \mathbb{R}^+$.

Beweis. Unter Verwendung von Korollar 3.1.2 und $\alpha = 1$ ergibt sich für die Log-Likelihood-Funktion

$$l(\lambda|y_1, \dots, y_n) = n \log(\lambda) - \lambda \sum_{i=1}^n (y_i - x_i) + \log(S'_n).$$

„Ableiten und Nullsetzen“ führt auf

$$l'(\lambda|y_1, \dots, y_n) = \frac{n}{\lambda} - \sum_{i=1}^n (y_i - x_i) \stackrel{!}{=} 0 \Leftrightarrow \hat{\lambda}_{\text{ML}} = \frac{n}{\sum (y_i - x_i)} = \frac{1}{\bar{y}_{n,1} - \bar{x}_{n,1}}.$$

Wegen $l''(\lambda|y_1, \dots, y_n) < 0$ für alle $\lambda \in \mathbb{R}^+$ ist dies das eindeutige Maximum. □

Für den Rest dieses Abschnitts untersuchen wir den allgemeinen Fall einer Gamma-Verteilung. Wie sich zeigen wird, konnten wir dort keinen geschlossenen Ausdruck für die ML-Schätzer der Parameter ermitteln. Das folgende Lemma ergibt sich direkt aus Korollar 3.1.2 durch Logarithmieren des dortigen Ausdrucks und anschließende Division durch n .

Lemma 3.1.4. *Liegen im Geburten-und-Todesfälle-Problem $\Gamma(\alpha, \lambda)$ -verteilte Lebensdauern und geordnete Geburts- und Sterbedaten x_1, \dots, x_n bzw. y_1, \dots, y_n vor, so ergibt sich die normierte Log-Likelihood-Funktion*

$$\frac{1}{n} l(\alpha, \lambda|y_1, \dots, y_n) = \alpha \log(\lambda) - \log(\Gamma(\alpha)) - \lambda \frac{1}{n} \sum_{i=1}^n (y_i - x_i) + \frac{1}{n} \log(g(\alpha|x, y)) \quad (3.1)$$

mit $g(\alpha|x, y) := \sum_{\pi \in S'_n} \prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1}$.

Man beachte, dass es sich bei $g(\alpha|x, y)$ um eine Permanente handelt. Deren Auftreten deutet auf Komplexitätstheoretische Schwierigkeiten hin, denn offensichtlich ist es zur Auswertung der Log-Likelihood-Funktion nötig, über alle Permutationen aus S'_n zu summieren. Dies setzt sich auch bei der Bestimmung des Maximums mittels der partiellen Ableitungen fort, denn es gilt

$$\frac{\partial}{\partial \alpha} \frac{1}{n} l(\alpha, \lambda|y_1, \dots, y_n) = \log(\lambda) - \Psi(\alpha) + \frac{1}{n} \frac{g'(\alpha|x, y)}{g(\alpha|x, y)}, \quad (3.2)$$

$$\frac{\partial}{\partial \lambda} \frac{1}{n} l(\alpha, \lambda | y_1, \dots, y_n) = \frac{\alpha}{\lambda} - \frac{1}{n} \sum_{i=1}^n (y_i - x_i). \quad (3.3)$$

Nullsetzen der zweiten Gleichung führt auf

$$\frac{\alpha}{\lambda} \stackrel{!}{=} \frac{1}{n} \sum_{i=1}^n (y_i - x_i); \quad (3.4)$$

dieser Ausdruck wird uns in (3.7) weiter unten wieder begegnen.

Die Summation über alle zulässigen Permutationen verursacht in vielen Fällen schon für mittelgroßes n einen erheblichen Zeitaufwand. Zwar konnten wir die Ausdrücke für $g(\alpha|x,y)$ oder $g'(\alpha|x,y)/g(\alpha|x,y)$ in einigen Spezialfällen vereinfachen, im Allgemeinen ist uns dies aber nicht gelungen. Deshalb beschränken wir uns darauf, in einer Bemerkung einige Erkenntnisse und Ideen zusammenzufassen, die neben der Interpretation als Permanente bei der praktischen Behandlung des Problems hilfreich sein können.

Bemerkung 3.1.5. (a) Der Ausdruck $\prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1}$ kann schon für z.B. $n = 100$ zu numerischen Problemen führen. Liegen die x - und y -Werte nahe beieinander, ist das Produkt sehr klein und wird u.U. vom Rechner als 0 angegeben. Umgekehrt führen weit auseinander liegende Werte ggf. zur Ausgabe ∞ . Hier ist möglicherweise hilfreich, dass wegen $g(\alpha|cx, cy) = c^{n(\alpha-1)} g(\alpha|x, y)$ für ein $c > 0$ eine Umskalierung vorgenommen werden kann. Außerdem ist in manchen Fällen der Übergang zum Logarithmus oder zur Exponentialfunktion nützlich.

(b) Es gilt

$$\frac{g'(\alpha|x,y)}{g(\alpha|x,y)} = \frac{\sum_{\pi \in S'_n} \prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1} \sum_{j=1}^n \log(y_{\pi(j)} - x_j)}{\sum_{\pi \in S'_n} \prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1}}.$$

Für jedes $\alpha \in \mathbb{R}^+$ und festes $n \in \mathbb{N}$ ist dies ein Ausdruck der Form $\sum_{\pi \in S'_n} \beta_\pi \gamma_\pi$ mit

$$\beta_\pi := \frac{\prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1}}{\sum_{\pi \in S'_n} \prod_{i=1}^n (y_{\pi(i)} - x_i)^{\alpha-1}}, \quad \gamma_\pi := \sum_{j=1}^n \log(y_{\pi(j)} - x_j) = \log \left(\prod_{i=1}^n (y_{\pi(i)} - x_i) \right).$$

Wegen $\sum \beta_\pi = 1$ kann $g'(\alpha|x,y)/g(\alpha|x,y)$ als Erwartungswert von $f(\Pi)$ aufgefasst werden, wobei die Permutation Π die Wahrscheinlichkeitsmassenfunktion $(\beta_\pi)_{\pi \in S'_n}$ besitzt und $f(\pi) = \sum_{k=1}^n \log(y_{\pi(k)} - x_k)$ für die gegebenen x - und y -Werte ist. Weiter gilt mit diesen Bezeichnungen

$$\frac{\partial}{\partial \alpha} \frac{g'(\alpha|x,y)}{g(\alpha|x,y)} = \mathbb{E}(f(\Pi))^2 - (\mathbb{E}f(\Pi))^2 = \text{var}(f(\Pi)) \geq 0.$$

- (c) Durch Lemma 2.2.10 sind die Permutationen bestimmbar, für die γ_π maximal bzw. minimal wird. Dies kann dazu benutzt werden, den Wert von $g'(\alpha|x,y)/g(\alpha|x,y)$ nach oben und unten abzuschätzen. Liegen die Werte nahe beieinander, kann der Ausdruck möglicherweise als näherungsweise konstant aufgefasst werden.
- (d) Auch die in Bemerkung 2.3.3 erwähnte Blockbildung trägt zur Vereinfachung der Berechnung von $g'(\alpha|x,y)/g(\alpha|x,y)$ bei. Dazu betrachtet man jeweils die für die Elemente der einzelnen Blöcke zulässigen Permutationen und berechnet damit den Beitrag des „Blocks“ zu $g'(\alpha|x,y)/g(\alpha|x,y)$. Natürlich kann sowohl diese als auch die nachfolgend in (e) genannte Idee ebenso zur Berechnung der Log-Likelihood-Funktion an sich benutzt und das Maximum aus z.B. einem Plot abgelesen werden. Dieses Vorgehen führte in unseren Simulationen allerdings immer zu ähnlichen Ergebnissen wie bei Annäherung der Ableitung.
- (e) Auch wenn nicht in akzeptabler Zeit über alle zulässigen Permutationen summiert werden kann, ist es möglich, den Wert von $g'(\alpha|x,y)/g(\alpha|x,y)$ durch Monte-Carlo-Simulation zumindest anzunähern. Dabei kann der aus Abb. 2.2 auf Seite 40 bekannte Algorithmus *UnifPerm* verwendet werden, um sukzessive eine beliebige Anzahl von zulässigen Permutationen zu erzeugen. Wie viele zulässigen Permutationen für ein hinreichend genaues Ergebnis benötigt werden, hängt nicht von der absoluten Anzahl an x - und y -Werten ab. Stattdessen ist die Anzahl der Geburt-Tod-Paare im in diesem Sinne „größten“ in (d) genannten „Block“ entscheidend. Weiter können auch die genannten Ideen kombiniert werden, indem man das Problem „in Blöcke aufteilt“, über alle Permutationen in „kleinen“ Blöcken summiert und den Wert von „großen“ Blöcken durch Monte-Carlo-Simulation bestimmt. \triangleleft

Im folgenden Beispiel betrachten wir die einfachste der oben vorgestellten Ideen, Annäherung des Wertes von $g'(\alpha|x,y)/g(\alpha|x,y)$ durch Monte-Carlo-Simulation ohne Ausnutzung der „Blöcke“. Um uns nicht in algorithmischen Details zu verlieren, verzichten wir auf die Umsetzung und nähere Untersuchung der komplexeren Ideen.

Beispiel 3.1.6. Es wurden $n = 100$ Geburtsdaten durch einen Poisson-Prozess mit Rate 1 erzeugt, mittels Verschiebung um i.i.d. $\Gamma(2,2)$ -verteilte Zufallsvariablen ergaben sich die Todeszeitpunkte. Nach der Formel aus Lemma 2.3.5 (a) betrug die Anzahl der zulässigen Permutationen ungefähr $1.07 \cdot 10^{29}$ und es gab 32 „Blöcke“ mit durchschnittlicher Länge 3.125, darunter je einen mit 13 und 14 Paaren. Nun wurden $k \in \{10^l, l = 2, \dots, 6\}$ zulässige Permutationen erzeugt und gespeichert. Anschließend bestimmte die Funktion `BBsolve` aus dem R-Paket `BB` simultan die Nullstellen der Ausdrücke in (3.2) und (3.3), dabei lief die Summe in $g'(\alpha|x,y)/g(\alpha|x,y)$ anstatt über alle Permutationen nur über die gespeicherten. Die Ergebnisse für die so erhaltenen ML-Schätzer finden sich in Tabelle 3.1. Anscheinend stabilisieren sich die Schätzer mit wachsendem Stichprobenumfang

| k | 10^2 | 10^3 | 10^4 | 10^5 | 10^6 |
|-----------------------------|----------|----------|----------|----------|----------|
| $\hat{\alpha}_{\text{ML}}$ | 1.913053 | 1.758679 | 1.942602 | 2.086917 | 1.960948 |
| $\hat{\lambda}_{\text{ML}}$ | 1.892249 | 1.739554 | 1.921477 | 2.064222 | 1.939623 |

Tabelle 3.1. ML-Schätzer bei Annäherung von $g'(\alpha|x,y)/g(\alpha|x,y)$ durch eine Monte-Carlo-Simulation mit verschiedenen Anzahlen k von gleichverteilt ausgewählten Permutationen bei $\Gamma(2,2)$ -verteilten Lebensdauern und $n = 100$

k in der Nähe der wahren Werte, wobei verglichen mit diesen selbst bei $k = 10^6$ noch ein relativer Fehler von mehr als 1.5% verbleibt. Eine geringere Abweichung ist bei der im Verhältnis recht kleinen Stichprobe nicht zu erwarten, eventuell würde die Berücksichtigung von mehr Permutationen zu einem anderen Ergebnis führen. Hier ergibt sich allerdings ein neues Problem: Wie viele Permutationen sind genug und woran erkennt man dies, wenn die wahren Werte nicht zum Vergleich vorliegen? \triangleleft

Es ist in unserer Problemstellung somit bei entsprechendem Zeitaufwand beliebig genau bestimmbar, an welcher Stelle die Likelihood-Funktion maximal wird (wie nah dieses Argmax dann an den wahren Werten liegt, hängt natürlich auch vom Wert von n ab). In der Praxis treten dabei aber möglicherweise numerische Schwierigkeiten oder ein Anwachsen der Zeit- und ggf. auch der Speicherkomplexität auf (beispielsweise beim Zwischenspeichern der für die Monte-Carlo-Simulation ausgewählten Permutationen). Dies wurde anhand der Gammaverteilung demonstriert, aber die Notwendigkeit, über alle zulässigen Permutationen summieren zu müssen, besteht auch bei anderen Verteilungen mit mehr als einem Parameter. Vom praktischen Standpunkt ist die ML-Methode in der vorliegenden Situation also nur bedingt geeignet, so dass Anlass für die Betrachtung anderer Verfahren besteht.

3.2. Die Momentenmethode

3.2.1. Grundlagen

Bei Vorliegen einer Stichprobe X_1, \dots, X_n aus einer von einem, möglicherweise mehrdimensionalen, Parameter θ abhängenden Verteilung F_θ ist eine Möglichkeit, einen Schätzer für θ zu bestimmen, die sog. Momentenmethode. Dabei werden zunächst die theoretischen Momente $E_\theta X_1^k$ der Verteilung durch die Stichprobenmomente $\frac{1}{n} \sum_{i=1}^n X_i^k$ für hinreichend viele $k = 1, 2, \dots$ geschätzt. Besteht nun zusätzlich ein funktionaler Zusammenhang zwischen dem Parameter und den Momenten, ist es es nahe liegend, die

Stichprobenmomente in diesen einzusetzen und so einen Schätzer $\hat{\theta}$ zu erhalten (sog. Einsetzprinzip).

In diesem Abschnitt werden Momentenschätzer für die Parameter der Lebensdauerverteilung bestimmt. Wir benötigen also zunächst Ausdrücke für die Stichprobenmomente der $(Z_i)_{i=1}^n$, die nur von $(X_i)_{i=1}^n$ und $(Y_i)_{i=1}^n$ abhängen. Wegen

$$\sum_{i=1}^n Y_i = \sum_{i=1}^n \tilde{Y}_i = \sum_{i=1}^n X_i + \sum_{i=1}^n Z_i$$

ergibt sich $\bar{Z}_{n,1} = \bar{Y}_{n,1} - \bar{X}_{n,1}$. Wir benutzen deswegen den Ansatz

$$\bar{Y}_{n,1} - \bar{X}_{n,1} \stackrel{!}{=} \text{E}Z_1; \quad (3.5)$$

dies ist nach Gleichung (3.4) auch der ML-Schätzer für den Erwartungswert. Das Zeichen $\stackrel{!}{=}$ bedeutet also, dass die Gleichheit als gegeben angenommen wird. Im nächsten Schritt benutzen wir dann die Abhängigkeit der rechten Seite von den interessierenden Parametern, um einen Schätzer für diese zu erhalten. Weiter gilt

$$\bar{Y}_{n,2} = \frac{1}{n} \sum_{i=1}^n (X_i + Z_i)^2 = \bar{X}_{n,2} + \frac{2}{n} \sum_{i=1}^n X_i Z_i + \bar{Z}_{n,2},$$

oder äquivalent $\bar{Y}_{n,2} - \bar{X}_{n,2} - \frac{2}{n} \sum_{i=1}^n X_i Z_i = \bar{Z}_{n,2}$. Aus den vorliegenden Daten können $\bar{X}_{n,2}$ und $\bar{Y}_{n,2}$ sofort berechnet werden, deshalb fehlt zur Bestimmung des zweiten Moments der Lebensdauerverteilung nur $\frac{2}{n} \sum_{i=1}^n X_i Z_i$. Hierfür wäre es aber notwendig, die Zusammengehörigkeit von Geburts- und Lebensdauern zu kennen. Weil die Geburtsdaten vorliegen und die Lebensdauern nicht, ersetzen wir jeweils $X_i Z_i$ durch $\text{E}[X_i Z_i | X_1, \dots, X_n]$ und erhalten unter Ausnutzung der Unabhängigkeit von (X_i) und (Z_i)

$$\frac{2}{n} \sum_{i=1}^n X_i Z_i \stackrel{!}{=} \frac{2}{n} \sum_{i=1}^n \text{E}[X_i Z_i | X_1, \dots, X_n] = \frac{2}{n} \sum_{i=1}^n X_i \text{E}[Z_i | X_1, \dots, X_n] = 2(\text{E}Z_1)\bar{X}_{n,1}.$$

Ersetzen von $\text{E}Z_1$ nach Gleichung (3.5) ergibt als zweite Bestimmungsgleichung

$$\bar{Y}_{n,2} - \bar{X}_{n,2} - 2(\bar{Y}_{n,1} - \bar{X}_{n,1})\bar{X}_{n,1} \stackrel{!}{=} \text{E}Z_1^2. \quad (3.6)$$

Aus den Gleichungen (3.5) und (3.6) resultieren nun die beiden Momentenschätzer

$$(\widehat{\text{E}Z_1})_n^{\text{MM}} := \bar{Y}_{n,1} - \bar{X}_{n,1}, \quad (3.7)$$

$$(\widehat{\text{E}Z_1^2})_n^{\text{MM}} := \bar{Y}_{n,2} - \bar{X}_{n,2} - 2(\bar{Y}_{n,1} - \bar{X}_{n,1})\bar{X}_{n,1}. \quad (3.8)$$

Wir schreiben $\widehat{\text{EZ}}_1$ für $(\widehat{\text{EZ}}_1)_n^{\text{MM}}$ und $\widehat{\text{EZ}}_1^2$ für $(\widehat{\text{EZ}}_1^2)_n^{\text{MM}}$, wenn Verwechslungen, z.B. mit anderen Schätzern, ausgeschlossen sind. Für einige der folgenden theoretischen Überlegungen ist ein anderer Ausdruck für $\widehat{\text{EZ}}_1^2$ als der in (3.8) angegebene sinnvoll. Unter Ausnutzung von

$$\bar{Y}_{n,2} - \bar{X}_{n,2} - 2(\bar{Y}_{n,1} - \bar{X}_{n,1})\bar{X}_{n,1} = \bar{Z}_{n,2} + 2 \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right)$$

erhalten wir

$$\widehat{\text{EZ}}_1^2 = \bar{Z}_{n,2} + 2 \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right). \quad (3.8b)$$

Man beachte, dass dieser Ausdruck für die tatsächliche Berechnung der Schätzer ungeeignet ist, da er im Gegensatz zu (3.8) nicht unmittelbar von den vorliegenden Daten $(X_i)_{i=1}^n$ und $(Y_i)_{i=1}^n$ abhängt.

3.2.2. Aus einem Poisson-Prozess stammende Geburtszeitpunkte

Für die Resultate in diesem Abschnitt ist es notwendig, das zugrundeliegende Modell weiter zu konkretisieren und zu spezialisieren. Wir stellen die benötigten Voraussetzungen hier zusammen und verweisen weiter unten darauf.

- (A1) Es ist $X_i = \sum_{j=1}^i U_j$ für alle i , $U_j \sim \text{Exp}(\theta)$ für alle j und die (U_j) sind unabhängig.
- (A2) Die (Z_i) sind i.i.d. mit $\text{EZ}_1^4 < \infty$ und $\text{var}(Z_1) > 0$.
- (A3) Die (X_i) und die (Z_i) sind unabhängig.
- (A4) Es ist $\tilde{Y}_i = X_i + Z_i$ für alle i .
- (A5) Es liegen nur $(X_i)_{i=1,\dots,n}$ und $(Y_i)_{i=1,\dots,n}$ vor mit $Y_i := \tilde{Y}_{(n:i)}$ für $i = 1, \dots, n$.

Offenbar hängt jedes Y_i in dem Sinn von n ab, dass sich einige oder sogar alle vorher bestehenden Ordnungsstatistiken ändern können, wenn (wie später untersucht) zusätzliche Paare von Geburts- und Todeszeitpunkten hinzugenommen werden. Dies spielt bei unseren Untersuchungen aber keine Rolle, so dass wir die Abhängigkeit nicht weiter kenntlich machen. Aus (A2) kann die Existenz von $E|Z_1|$ gefolgert werden, außerdem haben wir in Hinblick auf möglichst allgemeingültige Ergebnisse keine Nichtnegativität für die (Z_i) gefordert. Nimmt ein Z_i tatsächlich negative Werte an, kann man formell nicht mehr von „Lebensdauern“ sprechen, was wir im Folgenden aber ignorieren. Da für einige Resultate nur die oben vorausgesetzte Struktur der (X_i) notwendig ist und nicht die konkrete Verteilung der Zwischenankunftszeiten, benutzen wir aus dem gleichen Bestreben nach Allgemeingültigkeit anstelle von (A1) gelegentlich die folgende Bedingung.

(A1') Es ist $X_i = \sum_{j=1}^i U_j$ für alle i , (U_j) i.i.d. mit $EU_1 \neq 0$ und $\text{var}(U_1) < \infty$.

Wir werden bei Voraussetzung von (A1) bis (A5) vom Standardmodell sprechen und die Bezeichnung verallgemeinertes Standardmodell verwenden, wenn (A1) durch (A1') ersetzt wird. Ist eine Aussage im verallgemeinerten Standardmodell gültig, folgt sie natürlich sofort auch für das (spezialisierte) Standardmodell.

Bevor wir fortfahren, sei an die folgenden Definitionen erinnert. Ist θ_0 ein unbekannter Parameter und $\hat{\theta}_n$ eine Folge von Schätzern für θ_0 , so heißt $\hat{\theta}_n$ bekanntlich (*schwach*) *konsistent* genau dann, wenn $\hat{\theta}_n \xrightarrow{P} \theta_0$ gilt, *stark konsistent* genau dann, wenn $\hat{\theta}_n \xrightarrow{\text{f.s.}} \theta_0$ gilt und *konsistent im quadratischen Mittel* genau dann, wenn $\lim_{n \rightarrow \infty} E(\hat{\theta}_n - \theta_0)^2 = 0$ ist. Da bei erwartungstreuen Schätzern die mittlere quadratische Abweichung (MSE) mit der Varianz übereinstimmt, ist dort die Konsistenz im quadratischen Mittel äquivalent zu $\lim_{n \rightarrow \infty} \text{var}(\hat{\theta}_n) = 0$. Es soll nun im verallgemeinerten Standardmodell überprüft werden, ob die oben definierten Schätzer einige dieser Eigenschaften besitzen. Wir beginnen mit \widehat{EZ}_1 und formulieren die Voraussetzungen des folgenden Satzes etwas allgemeiner, um das Resultat später auch in Abwandlungen des Standardmodells verwenden zu können.

Satz 3.2.1. *Unter den Bedingungen (A2), (A4) und (A5) ist \widehat{EZ}_1 stark konsistent und erwartungstreu.*

Beweis. Es ist

$$\widehat{EZ}_1 = \bar{Y}_{n,1} - \bar{X}_{n,1} = \bar{Z}_{n,1}.$$

Die Erwartungstreue ist jetzt offensichtlich. Nach dem starken Gesetz der großen Zahlen konvergiert $\bar{Z}_{n,1}$ unter obigen Voraussetzungen fast sicher gegen EZ_1 . \square

Offensichtlich kann Satz 3.2.1 im verallgemeinerten Standardmodell angewendet werden, damit ergibt sich in diesem Rahmen sofort eine Güteaussage für \widehat{EZ}_1 . Als nächstes untersuchen wir \widehat{EZ}_1^2 und zeigen in einem weiter unten folgenden Satz, dass dieser Schätzer hier nicht konsistent im quadratischen Mittel ist. Grundlage für den Beweis dieser Behauptung ist das nächste, sehr allgemeine Resultat.

Satz 3.2.2. *Es seien $(X_i)_{i \in \mathbb{N}}$ und $(Z_i)_{i \in \mathbb{N}}$ zwei Folgen von Zufallsvariablen. Für jedes $i \in \mathbb{N}$ sei $EX_i < \infty$ und die (Z_i) seien unabhängig von den (X_i) und i.i.d. mit $EZ_1 < \infty$.*

(a) *Dann gilt*

$$E \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right) = 0$$

und

$$\text{var} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right) = E \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right)^2.$$

(b) Ist zusätzlich $X_i = \sum_{k=1}^i U_k$ für alle $i \in \mathbb{N}$ mit (U_k) i.i.d., $\mathbb{E}U_1 \neq 0$ und $\text{var}(U_1) < \infty$ sowie $0 < \text{var}(Z_1) < \infty$, dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right)^2 = \infty.$$

Beweis. (a) Unter Verwendung der Unabhängigkeit von (X_i) und (Z_i) erhalten wir

$$\begin{aligned} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right) &= \frac{1}{n} \sum_{i=1}^n (\mathbb{E}X_i)(\mathbb{E}Z_i) - \mathbb{E}Z_1 \mathbb{E}\bar{X}_{n,1} \\ &= \mathbb{E}Z_1 \mathbb{E}\bar{X}_{n,1} - \mathbb{E}Z_1 \mathbb{E}\bar{X}_{n,1} = 0, \end{aligned}$$

da mit $\mathbb{E}X_i, i = 1, \dots, n$, auch $\mathbb{E}\bar{X}_{n,1}$ endlich ist. Nun folgt direkt

$$\text{var} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right) = \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right)^2.$$

(b) Es gilt

$$\begin{aligned} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right)^2 &= \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i (Z_i - \bar{Z}_{n,1}) \right)^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}X_i X_j (Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1}) \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n (\mathbb{E}X_i^2)(\mathbb{E}(Z_i - \bar{Z}_{n,1})^2) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n (\mathbb{E}X_i X_j)(\mathbb{E}(Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1})) \right), \end{aligned} \tag{3.9}$$

dabei haben wir die Unabhängigkeit der X - und Z -Werte ausgenutzt.

1. Bestimmung von $(\mathbb{E}X_i X_j)(\mathbb{E}(Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1}))$

Wir benutzen im Folgenden, dass wegen $i < j$ gilt $X_j = X_i + \sum_{k=i+1}^j U_k$, außerdem sind Z_i und Z_j deshalb unabhängig. Damit ergibt sich

$$\mathbb{E}((Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1})) = \mathbb{E}Z_i Z_j - \mathbb{E}Z_i \bar{Z}_{n,1} - \mathbb{E}Z_j \bar{Z}_{n,1} + \mathbb{E}(\bar{Z}_{n,1})^2.$$

Es ist $\mathbb{E}Z_i Z_j = (\mathbb{E}Z_1)^2$, $\mathbb{E}Z_i \bar{Z}_{n,1} = \mathbb{E}Z_j \bar{Z}_{n,1} = \frac{n-1}{n}(\mathbb{E}Z_1)^2 + \frac{1}{n}\mathbb{E}Z_1^2$ und

$$\mathbb{E}(\bar{Z}_{n,1})^2 = \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}Z_k Z_l = \frac{1}{n} (\mathbb{E}Z_1^2 + (n-1)(\mathbb{E}Z_1)^2),$$

so dass wir insgesamt erhalten

$$\begin{aligned}
& \mathbb{E}((Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1})) \\
&= (\mathbb{E}Z_1)^2 - 2(\mathbb{E}Z_1)^2 + \frac{2}{n}(\mathbb{E}Z_1)^2 - \frac{2}{n}\mathbb{E}Z_1^2 + \frac{1}{n}\mathbb{E}Z_1^2 + (\mathbb{E}Z_1)^2 - \frac{1}{n}(\mathbb{E}Z_1)^2 \\
&= \frac{1}{n}((\mathbb{E}Z_1)^2 - \mathbb{E}Z_1^2) = -\frac{1}{n}\text{var}(Z_1).
\end{aligned} \tag{3.10}$$

Weiter gilt

$$\begin{aligned}
\mathbb{E}X_iX_j &= \mathbb{E}\left(\left(\sum_{k=1}^i U_k\right)^2 + \left(\sum_{k=1}^i U_k\right)\left(\sum_{k=i+1}^j U_k\right)\right) \\
&= i\mathbb{E}U_1^2 + i(i-1)(\mathbb{E}U_1)^2 + i(j-i)(\mathbb{E}U_1)^2 \\
&= i(\mathbb{E}U_1^2 - (\mathbb{E}U_1)^2) + ij(\mathbb{E}U_1)^2 = i\text{var}(U_1) + ij(\mathbb{E}U_1)^2
\end{aligned}$$

und es folgt

$$(\mathbb{E}X_iX_j)(\mathbb{E}(Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1})) = -\frac{1}{n}\text{var}(Z_1)(i\text{var}(U_1) + ij(\mathbb{E}U_1)^2).$$

2. Asymptotik von $2\sum_{i=1}^{n-1}\sum_{j=i+1}^n(\mathbb{E}X_iX_j)(\mathbb{E}(Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1}))$

Mit 1. erhalten wir

$$\begin{aligned}
& 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n(\mathbb{E}X_iX_j)(\mathbb{E}(Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1})) \\
&= -\frac{2}{n}\text{var}(Z_1)\sum_{i=1}^{n-1}\sum_{j=i+1}^n(i\text{var}(U_1) + ij(\mathbb{E}U_1)^2) \\
&= -\frac{2}{n}\text{var}(Z_1)\left(\text{var}(U_1)\sum_{i=1}^{n-1}i(n-i) + (\mathbb{E}U_1)^2\sum_{i=1}^{n-1}\sum_{j=i+1}^n ij\right).
\end{aligned} \tag{3.11}$$

Wegen $\sum_{i=1}^{n-1}i(n-i) = n\frac{n(n-1)}{2} - \frac{n(n-1)(2n-1)}{6}$ verhält sich die erste Summe für große n wie $\frac{1}{6}n^3$ und

$$\begin{aligned}
\sum_{i=1}^{n-1}\sum_{j=i+1}^n ij &= \frac{n(n+1)}{2}\frac{n(n-1)}{2} - \frac{1}{2}\sum_{i=1}^{n-1}(i^3 + i^2) \\
&= \frac{n(n+1)}{2}\frac{n(n-1)}{2} - \frac{1}{2}\left(\left(\frac{n(n-1)}{2}\right)^2 + \frac{n(n-1)(2n-1)}{6}\right)
\end{aligned}$$

liefert, dass die zweite Summe sich für große n wie $\frac{1}{8}n^4$ verhält. Insgesamt folgt

$$2\sum_{i=1}^{n-1}\sum_{j=i+1}^n(\mathbb{E}X_iX_j)(\mathbb{E}(Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1})) \sim -\frac{1}{4}n^3(\mathbb{E}U_1)^2\text{var}(Z_1).$$

3. Bestimmung von $(EX_i^2)(E(Z_i - \bar{Z}_{n,1})^2)$

Es gilt

$$EX_i^2 = E\left(\sum_{k=1}^i U_k\right)^2 = iEU_1^2 + i(i-1)(EU_1)^2$$

und

$$\begin{aligned} E(Z_i - \bar{Z}_{n,1})^2 &= EZ_i^2 - 2EZ_i\bar{Z}_{n,1} + E(\bar{Z}_{n,1})^2 \\ &= EZ_1^2 - 2\left(\frac{n-1}{n}(EZ_1)^2 + \frac{1}{n}EZ_1^2\right) + \left(\frac{1}{n}(EZ_1^2 + (n-1)(EZ_1)^2)\right) \\ &= \frac{n-1}{n}(EZ_1^2 - (EZ_1)^2) = \left(1 - \frac{1}{n}\right) \text{var}(Z_1), \end{aligned} \quad (3.12)$$

wobei wir im drittletzten Schritt zwei Teilergebnisse aus 1. verwendet haben. Damit ergibt sich insgesamt

$$(EX_i^2)(E(Z_i - \bar{Z}_{n,1})^2) = \left(1 - \frac{1}{n}\right) \text{var}(Z_1) ((iEU_1)^2 + i \text{var}(U_1)).$$

4. Asymptotik von $\sum_{i=1}^n (EX_i^2)(E(Z_i - \bar{Z}_{n,1})^2)$

Mit den Ergebnissen aus 3. folgt

$$\begin{aligned} &\sum_{i=1}^n (EX_i^2)(E(Z_i - \bar{Z}_{n,1})^2) \\ &= \left(1 - \frac{1}{n}\right) \text{var}(Z_1) \sum_{i=1}^n ((iEU_1)^2 + i \text{var}(U_1)) \\ &= \left(1 - \frac{1}{n}\right) \text{var}(Z_1) \left((EU_1)^2 \frac{n(n+1)(2n+1)}{6} + \text{var}(U_1) \frac{n(n+1)}{2} \right). \end{aligned} \quad (3.13)$$

Offenbar gilt also für große n

$$\sum_{i=1}^n (EX_i^2)(E(Z_i - \bar{Z}_{n,1})^2) \sim \frac{1}{3}n^3(EU_1)^2 \text{var}(Z_1).$$

5. Zusammengefasst führen die Ergebnisse aus 2. und 4. schließlich zu

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i(Z_i - \bar{Z}_{n,1})\right)^2 \sim \frac{(EU_1)^2 \text{var}(Z_1)}{12} \cdot n.$$

Dies läuft mit wachsendem n gegen ∞ und auch die zweite Behauptung folgt. □

Bemerkung 3.2.3. Für den Beweis von Satz 3.2.2 ist es nicht nötig, die Varianz explizit auszurechnen, da die Asymptotik genügt. Durch Auswertung von (3.11) und (3.13) mit einem Computeralgebrasystem wie Maple folgt

$$\text{var} \left(\frac{1}{n} \sum_{i=1}^n X_i (Z_i - \bar{Z}_{n,1}) \right) = \frac{\text{var}(Z_1)}{12} \left((\text{EU}_1)^2 \left(n - 2 - \frac{1}{n} + \frac{2}{n^2} \right) + \text{EU}_1^2 \left(2 - \frac{2}{n^2} \right) \right).$$

Falls $\text{EU}_1 = 0$ und $\text{EU}_1^2 > 0$ gelten, wenn also die (U_i) positive und negative Werte annehmen können und sich diese im Mittel aufheben, ist die Varianz asymptotisch konstant. \triangleleft

Bevor wir Satz 3.2.2 in Bezug auf $\widehat{\text{EZ}}_1^2$ anwenden, benötigen wir noch die beiden folgenden Resultate.

Satz 3.2.4. *Im verallgemeinerten Standardmodell ist $\widehat{\text{EZ}}_1^2$ erwartungstreu.*

Beweis. Unter Verwendung von Satz 3.2.2 (a) und der Umformung (3.8b) ergibt sich

$$\begin{aligned} \text{E} \left(\widehat{\text{EZ}}_1^2 \right) &= \text{E} \left(\bar{Y}_{n,2} - \bar{X}_{n,2} - 2(\bar{Y}_{n,1} - \bar{X}_{n,1})\bar{X}_{n,1} \right) \\ &= \text{E} \left(\bar{Z}_{n,2} \right) + 2\text{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) = \text{EZ}_1^2. \quad \square \end{aligned}$$

Lemma 3.2.5. *Es seien $(X_i)_{i \in \mathbb{N}}$ und $(Z_i)_{i \in \mathbb{N}}$ zwei Folgen von Zufallsvariablen. Für jedes $i \in \mathbb{N}$ sei $\text{EX}_i < \infty$ und die (Z_i) seien unabhängig von den (X_i) sowie i.i.d. mit $\text{EZ}_1^3 < \infty$. Dann gilt*

$$\text{E} \left((\bar{Z}_{n,2} - \text{EZ}_1^2) \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) \right) = 0.$$

Beweis. Anwendung von Satz 3.2.2 (a) führt auf

$$\begin{aligned} &\text{E} \left((\bar{Z}_{n,2} - \text{EZ}_1^2) \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) \right) \\ &= \text{E} \bar{Z}_{n,2} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) - (\text{EZ}_1^2) \text{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{EZ}_i^2 X_j (Z_j - \bar{Z}_{n,1}). \end{aligned}$$

Unter Verwendung der Unabhängigkeit von (X_i) und (Z_i) folgt

$$\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{EZ}_i^2 X_j (Z_j - \bar{Z}_{n,1})$$

$$\begin{aligned}
&= \frac{1}{n^2} \left(\sum_{i=1}^n (\mathbb{E}X_i)(\mathbb{E}Z_i^2(Z_i - \bar{Z}_{n,1})) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (\mathbb{E}X_j)(\mathbb{E}Z_i^2(Z_j - \bar{Z}_{n,1})) \right) \\
&= \frac{1}{n^2} \left(\sum_{i=1}^n (\mathbb{E}X_i)(\mathbb{E}Z_1^3 - \frac{1}{n}(\mathbb{E}Z_1^3 + (n-1)(\mathbb{E}Z_1^2)(\mathbb{E}Z_1))) \right. \\
&\quad \left. + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (\mathbb{E}X_j) \left((\mathbb{E}Z_1^2)(\mathbb{E}Z_1) - \frac{1}{n}(\mathbb{E}Z_1^3 + (n-1)(\mathbb{E}Z_1^2)(\mathbb{E}Z_1)) \right) \right) \\
&= \frac{1}{n^2} \left(\frac{n-1}{n} (\mathbb{E}Z_1^3 - (\mathbb{E}Z_1^2)(\mathbb{E}Z_1)) \sum_{i=1}^n \mathbb{E}X_i - \frac{1}{n} (\mathbb{E}Z_1^3 - (\mathbb{E}Z_1^2)(\mathbb{E}Z_1)) \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}X_j \right) \\
&= \frac{1}{n^3} (\mathbb{E}Z_1^3 - (\mathbb{E}Z_1^2)(\mathbb{E}Z_1)) \left((n-1) \sum_{i=1}^n \mathbb{E}X_i - \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}X_j \right) = 0,
\end{aligned}$$

weil sich die Terme in der letzten Klammer aufheben. \square

Satz 3.2.6. *Im verallgemeinerten Standardmodell ist $\widehat{\mathbb{E}Z_1^2}$ nicht konsistent im quadratischen Mittel.*

Beweis. Nach Satz 3.2.4 ist $\widehat{\mathbb{E}Z_1^2}$ erwartungstreu und somit ist die Konsistenz im quadratischen Mittel äquivalent zu $\lim_{n \rightarrow \infty} \text{var}(\widehat{\mathbb{E}Z_1^2}) = 0$. Außerdem ergibt sich

$$\begin{aligned}
\text{var}(\widehat{\mathbb{E}Z_1^2}) &= \mathbb{E}(\widehat{\mathbb{E}Z_1^2} - \mathbb{E}Z_1^2)^2 \\
&= \mathbb{E} \left(\bar{Z}_{n,2} + 2 \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) - \mathbb{E}Z_1^2 \right)^2 \\
&= \mathbb{E}(\bar{Z}_{n,2} - \mathbb{E}Z_1^2)^2 + 4\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right)^2 \\
&\quad + 4\mathbb{E} \left((\bar{Z}_{n,2} - \mathbb{E}Z_1^2) \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) \right) \\
&\stackrel{(*)}{=} \mathbb{E}(\bar{Z}_{n,2} - \mathbb{E}Z_1^2)^2 + 4\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right)^2 \\
&\geq \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right)^2,
\end{aligned}$$

dabei gilt $(*)$ wegen Lemma 3.2.5. Im verallgemeinerten Standardmodell sind die Voraussetzungen von Satz 3.2.2 (b) erfüllt. Damit geht in obiger Ungleichung die rechte Seite mit wachsendem n gegen ∞ und folglich auch $\text{var}(\widehat{\mathbb{E}Z_1^2})$. \square

Wären in (3.7) und (3.8) alle Werte X_i gleich 0, erhielte man die Bestimmungsgleichungen der klassischen Momentenschätzer der Lebensdauerverteilung. Da diese bekanntermaßen erwartungstreu für die theoretischen Momente sind und ihre Varianzen für $n \rightarrow \infty$ verschwinden, könnten nach den vorhergehenden Ausführungen die Parameter einer zweiparametrischen Verteilung geschätzt werden. In manchen Modellen funktioniert also die Idee, $X_i Z_i$ durch $E[X_i Z_i | X_1, \dots, X_n]$ zu ersetzen, und in anderen nicht. Wir untersuchen dies im nächsten Abschnitt unter anderen Voraussetzungen als im Standardmodell näher.

Bemerkung 3.2.7. Da alle in Bezug auf das Ausgangsproblem untersuchten Verteilungen zweiparametrisch sind, haben wir uns stillschweigend auf diesen Fall konzentriert. Im allgemeinen Fall erhält man für $k = 1, 2, \dots$ wegen

$$\bar{Y}_{n,k} = \frac{1}{n} \sum_{i=1}^n (X_i + Z_i)^k = \frac{1}{n} \sum_{i=1}^n \sum_{l=0}^k \binom{k}{l} X_i^l Z_i^{k-l} = \bar{X}_{n,k} + \bar{Z}_{n,k} + \sum_{l=1}^{k-1} \binom{k}{l} \frac{1}{n} \sum_{i=1}^n X_i^l Z_i^{k-l}$$

und durch das von oben bekannte Ersetzen von $X_i^l Z_i^{k-l}$ mit $E[X_i^l Z_i^{k-l} | X_1, \dots, X_n]$ für $l = 1, \dots, k-1$ analog zum obigen Vorgehen einen Schätzer für das k -te Moment durch

$$\bar{Y}_{n,k} - \bar{X}_{n,k} - \sum_{l=1}^{k-1} \binom{k}{l} \bar{X}_{n,l} E Z_1^{k-l} \stackrel{!}{=} E Z_1^k.$$

In der letzten Zeile müssen dann noch $E Z_1, E Z_1^2, \dots, E Z_1^{k-1}$ durch die jeweiligen Schätzer ersetzt werden, die man sukzessive für $k = 1, 2, \dots, k-1$ bestimmen kann. \triangleleft

3.2.3. I.i.d. Geburtszeitpunkte

Allgemeine Ergebnisse

Das Standardmodell wird durch Ersetzen von **(A1)** mit der folgenden Annahme **(A1'')** abgewandelt.

(A1'') Die (X_i) sind i.i.d. mit $E X_1^k < \infty$ für $k = 1, 2$.

In diesem Abschnitt sprechen wir vom abgewandelten Standardmodell, wenn die Annahmen **(A1'')** und **(A2)** bis **(A5)** erfüllt sind. Wir zeigen zunächst, dass die Aussage von Satz 3.2.2 (b) bei i.i.d. Geburtszeitpunkten nicht gilt.

Satz 3.2.8. *Es seien $(X_i)_{i \in \mathbb{N}}$ und $(Z_i)_{i \in \mathbb{N}}$ zwei voneinander unabhängige Folgen von je i.i.d. Zufallsvariablen mit endlichen ersten und zweiten Momenten. Dann gilt*

$$\lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{Z}_{n,1} \bar{X}_{n,1} \right) = 0.$$

Beweis. Wegen Satz 3.2.2 (a) können wir von (3.9) ausgehen und erhalten unter Verwendung von (3.12) bzw. (3.10)

$$\begin{aligned}
\sum_{i=1}^n (\mathbb{E}X_i^2)(\mathbb{E}(Z_i - \bar{Z}_{n,1})^2) &= \left(1 - \frac{1}{n}\right) \text{var}(Z_1) \sum_{i=1}^n \mathbb{E}X_i^2 = (n-1) \text{var}(Z_1) \mathbb{E}X_1^2, \\
2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n (\mathbb{E}X_i X_j) (\mathbb{E}(Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1})) \\
&= -\frac{2}{n} \text{var}(Z_1) \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}X_i X_j \\
&= -\frac{2}{n} \text{var}(Z_1) (\mathbb{E}X_1)^2 \cdot \frac{n(n-1)}{2} = -(n-1) \text{var}(Z_1) (\mathbb{E}X_1)^2.
\end{aligned}$$

In diesem Fall ergibt sich also insgesamt

$$\text{var} \left(\frac{1}{n} \sum_{i=1}^n X_i (Z_i - \bar{Z}_{n,1}) \right) = \frac{n-1}{n^2} \text{var}(Z_1) (\mathbb{E}X_1^2 - (\mathbb{E}X_1)^2) = \frac{n-1}{n^2} \text{var}(Z_1) \text{var}(X_1).$$

Dies konvergiert für wachsendes n gegen 0. □

Unter Verwendung des vorhergehenden Resultats können wir nun den folgenden Satz beweisen.

Satz 3.2.9. *Im abgewandelten Standardmodell ist der Schätzer $\widehat{\mathbb{E}Z_1^2}$ konsistent im quadratischen Mittel.*

Beweis. Nach Satz 3.2.4 ist $\widehat{\mathbb{E}Z_1^2}$ im untersuchten Fall erwartungstreu, so dass wir für die Konsistenz im quadratischen Mittel wieder die Varianz betrachten können. Da wegen Satz 3.2.2 (a)

$$\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right)^2 = \text{var} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right)$$

gilt, folgt

$$\begin{aligned}
\text{var}(\widehat{\mathbb{E}Z_1^2}) &= \mathbb{E} \left(\bar{Z}_{n,2} + 2 \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) - \mathbb{E}Z_1^2 \right)^2 \\
&= \text{var}(\bar{Z}_{n,2}) + 4 \text{var} \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) \\
&\quad + 4 \mathbb{E} \left((\bar{Z}_{n,2} - \mathbb{E}Z_1^2) \left(\frac{1}{n} \sum_{i=1}^n X_i Z_i - \bar{X}_{n,1} \bar{Z}_{n,1} \right) \right).
\end{aligned}$$

Der erste Summand ist identisch mit $\text{var}(Z_1^2)/n$ und konvergiert deshalb genauso gegen 0 wie laut Satz 3.2.8 der zweite Summand. Der dritte Summand schließlich verschwindet nach Lemma 3.2.5 sogar für jedes $n \in \mathbb{N}$ und es folgt die Behauptung. \square

Aus der Konsistenz im quadratischen Mittel folgt wegen

$$0 \leq P(|\hat{\theta}_n - \theta| > \epsilon) \leq \frac{1}{\epsilon^2} \text{E}(\hat{\theta}_n - \theta)^2 \rightarrow 0 \quad \text{für } n \rightarrow \infty \text{ und jedes feste } \epsilon > 0$$

auch die (schwache) Konsistenz. Da im abgewandelten Modell Satz 3.2.1 verwendet werden kann, ist dort neben $\widehat{\text{E}Z_1^2}$ auch $\widehat{\text{E}Z_1}$ konsistent (im klassischen Sinn), deshalb besteht ein möglicher nächster Schritt darin, die Schätzer in diesem Rahmen auf asymptotische Normalität zu überprüfen. Wir weisen diese Eigenschaft im Folgenden mithilfe der Delta-Methode nach; dazu benötigen wir zunächst asymptotisch normale Schätzer, die in einem funktionalen Zusammenhang mit den unseren stehen.

Satz 3.2.10. *Es gilt*

$$\sqrt{n} \left(\begin{pmatrix} \bar{Z}_{n,1} \\ \bar{Z}_{n,2} \\ \frac{1}{n} \sum_{i=1}^n X_i Z_i \\ \bar{X}_{n,1} \end{pmatrix} - \begin{pmatrix} \text{E}Z_1 \\ \text{E}Z_1^2 \\ (\text{E}X_1)(\text{E}Z_1) \\ \text{E}X_1 \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_4(0, \Sigma)$$

mit $\Sigma = (\sigma_{ij})_{i,j=1,\dots,4}$, $\sigma_{ij} = \sigma_{ji}$ und

$$\begin{aligned} \sigma_{11} &= \text{var}(Z_1), & \sigma_{12} &= \text{E}Z_1^3 - (\text{E}Z_1)(\text{E}Z_1^2), & \sigma_{13} &= \text{E}X_1 \text{var}(Z_1), & \sigma_{14} &= 0, \\ \sigma_{22} &= \text{var}(Z_1^2), & \sigma_{23} &= \text{E}X_1(\text{E}Z_1^3 - (\text{E}Z_1)(\text{E}Z_1^2)), & \sigma_{24} &= 0, \\ \sigma_{33} &= \text{var}(X_1)(\text{E}Z_1)^2 + \text{var}(Z_1)(\text{E}X_1)^2 + \text{var}(X_1) \text{var}(Z_1), & \sigma_{34} &= \text{E}Z_1 \text{var}(X_1), \\ \sigma_{44} &= \text{var}(X_1). \end{aligned}$$

Beweis. Die asymptotische Normalität folgt sofort aus dem mehrdimensionalen zentralen Grenzwertsatz, dabei ist Σ die mit n multiplizierte Varianzmatrix des Vektors

$$\left(\bar{Z}_{n,1}, \bar{Z}_{n,2}, \frac{1}{n} \sum_{i=1}^n X_i Z_i, \bar{X}_{n,1} \right)'$$

Die Varianzen σ_{ii} , $i = 1, \dots, 4$, erhält man direkt durch Nachrechnen bzw. durch Anwendung von

$$\text{var}(UV) = \text{var}(U)(\text{E}V)^2 + \text{var}(V)(\text{E}U)^2 + \text{var}(U) \text{var}(V)$$

für unabhängige Zufallsvariablen U und V . Da die (X_i) und (Z_i) unabhängig sind, folgt sofort $\sigma_{14} = 0 = \sigma_{24}$. Zur Bestimmung der Kovarianzen benutzen wir, dass $\text{cov}(X_i Z_i, Z_j)$,

$\text{cov}(X_i, X_j Z_j)$ etc. nur für $i = j$ nicht verschwinden, weil ansonsten alle beteiligten Zufallsvariablen unabhängig sind. Dann ergibt sich mit der jeweiligen identischen Verteilung

$$\begin{aligned}\sigma_{12} &= \text{cov}(\bar{Z}_{n,1}, \bar{Z}_{n,2}) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov}(Z_i, Z_j^2) = \frac{1}{n} (\mathbb{E}Z_1^3 - (\mathbb{E}Z_1)(\mathbb{E}Z_1^2)), \\ \sigma_{13} &= \text{cov}\left(\bar{Z}_{n,1}, \frac{1}{n} \sum_{i=1}^n X_i Z_i\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov}(Z_i, X_j Z_j) = \frac{1}{n} \mathbb{E}X_1 (\mathbb{E}Z_1^2 - (\mathbb{E}Z_1)^2), \\ \sigma_{23} &= \text{cov}\left(\bar{Z}_{n,2}, \frac{1}{n} \sum_{i=1}^n X_i Z_i\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov}(Z_i^2, X_j Z_j) = \frac{1}{n} \mathbb{E}X_1 (\mathbb{E}Z_1^3 - \mathbb{E}Z_1 \mathbb{E}Z_1^2), \\ \sigma_{34} &= \text{cov}\left(\frac{1}{n} \sum_{i=1}^n X_i Z_i, \bar{X}_{n,1}\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i Z_i, X_j) = \frac{1}{n} \mathbb{E}Z_1 (\mathbb{E}X_1^2 - (\mathbb{E}X_1)^2)\end{aligned}$$

und somit folgt, dass Σ die angegebenen Einträge hat. \square

Nun kann mittels der Delta-Methode die asymptotische Normalität der Schätzer $\widehat{\mathbb{E}Z_1}$ und $\widehat{\mathbb{E}Z_1^2}$ hergeleitet werden. Dazu sei $F_1((a,b,c,d)') := (a, b + 2(c - da))'$, dann ist

$$\begin{pmatrix} (\widehat{\mathbb{E}Z_1}_n^{\text{MM}}) \\ (\widehat{\mathbb{E}Z_1^2}_n^{\text{MM}}) \end{pmatrix} = F_1 \left(\begin{pmatrix} \bar{Z}_{n,1} \\ \bar{Z}_{n,2} \\ \frac{1}{n} \sum_{i=1}^n X_i Z_i \\ \bar{X}_{n,1} \end{pmatrix} \right).$$

Satz 3.2.11. *Es gilt*

$$\sqrt{n} \left(\begin{pmatrix} (\widehat{\mathbb{E}Z_1}_n^{\text{MM}}) \\ (\widehat{\mathbb{E}Z_1^2}_n^{\text{MM}}) \end{pmatrix} - \begin{pmatrix} \mathbb{E}Z_1 \\ \mathbb{E}Z_1^2 \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_2(0, A\Sigma A')$$

mit $A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2\mathbb{E}X_1 & 1 & 2 & -2\mathbb{E}Z_1 \end{pmatrix}$ und Σ aus Satz 3.2.10.

Beweis. Der Schätzer $(\widehat{\mathbb{E}Z_1}, \widehat{\mathbb{E}Z_1^2})'$ geht unter Verwendung der Abbildung F_1 aus dem nach Satz 3.2.10 asymptotisch normalen Vektor $(\bar{Z}_{n,1}, \bar{Z}_{n,2}, \frac{1}{n} \sum_{i=1}^n X_i Z_i, \bar{X}_{n,1})'$ hervor. Diese Abbildung ist differenzierbar mit der im Satz angegebenen Jacobi-Matrix A . Damit ist nach der Delta-Methode auch $(\widehat{\mathbb{E}Z_1}, \widehat{\mathbb{E}Z_1^2})'$ asymptotisch normal mit der genannten Varianzmatrix. \square

Nachdem das asymptotische Verhalten der Schätzer bei wachsendem Stichprobenumfang bekannt ist, interessieren wir uns für ihr Verhalten bei bestimmten Änderungen der Daten. Im untersuchten Modell macht es zum einen keinen Unterschied für die Lebensdauer, wann jemand geboren wurde, zum anderen bewirkt eine Umskalierung der Daten

z.B. von Jahren auf Monate die gleiche Umskalierung der Lebensdauern. In der folgenden Definition werden die entsprechenden Eigenschaften von Schätzern eingeführt.

Definition 3.2.12. Es sei $\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n)$ ein Schätzer.

(a) $\hat{\theta}_n$ heißt *homogen vom Grad β* , $\beta \in \mathbb{R}$, genau dann, wenn

$$\hat{\theta}(aX_1, \dots, aX_n) = a^\beta \cdot \hat{\theta}(X_1, \dots, X_n)$$

für alle $a \in \mathbb{R}^+$ gilt. Ist $\hat{\theta}_n$ homogen vom Grad 1, heißt $\hat{\theta}_n$ auch *skalierungsäquivalent*, und ist $\hat{\theta}_n$ homogen vom Grad 0, wird dieser Schätzer *skalierungsinvariant* genannt.

(b) $\hat{\theta}_n$ heißt *verschiebungsinvariant* genau dann, wenn

$$\hat{\theta}(X_1 + c, \dots, X_n + c) = \hat{\theta}(X_1, \dots, X_n)$$

für alle $c \in \mathbb{R}$ gilt.

In unserem Modell basieren die Schätzer auf den Vektoren (X_1, \dots, X_n) und (Y_1, \dots, Y_n) , so dass sich eine Umskalierung und/oder Verschiebung immer simultan auf beide Vektoren bezieht. Der folgende Satz zeigt, dass die entwickelten Schätzer einige der oben angesprochenen Eigenschaften besitzen.

Satz 3.2.13. $\widehat{\text{EZ}}_1$ und $\widehat{\text{EZ}}_1^2$ sind verschiebungsinvariant, außerdem ist $\widehat{\text{EZ}}_1$ skalierungsäquivalent und $\widehat{\text{EZ}}_1^2$ homogen vom Grad 2.

Beweis. Es gilt

$$\begin{aligned} \widehat{\text{EZ}}_1(aX_1 + c, \dots, aX_n + c, aY_1 + c, \dots, aY_n + c) &= a(\bar{Y}_{n,1} + c) - a(\bar{X}_{n,1} + c) \\ &= a(\bar{Y}_{n,1} - \bar{X}_{n,1}) = a \cdot \widehat{\text{EZ}}_1 \end{aligned}$$

und

$$\begin{aligned} \widehat{\text{EZ}}_1^2(aX_1 + c, \dots, aX_n + c, aY_1 + c, \dots, aY_n + c) &= \frac{1}{n} \sum_{i=1}^n (aY_i + c)^2 - \frac{1}{n} \sum_{i=1}^n (aX_i + c)^2 - 2((a\bar{Y}_{n,1} + c) - (a\bar{X}_{n,1} + c))(a\bar{X}_{n,1} + c) \\ &= a^2 \bar{Z}_{n,2} + \frac{2a^2}{n} \sum_{i=1}^n X_i Z_i + 2ac \bar{Z}_{n,1} - 2a^2 \bar{Z}_{n,1} \bar{X}_{n,1} - 2ac \bar{Z}_{n,1} \\ &= a^2 \cdot \widehat{\text{EZ}}_1^2(X_1, \dots, X_n, Y_1, \dots, Y_n). \quad \square \end{aligned}$$

Bemerkung 3.2.14. Die bisherigen Resultate kommen ohne parametrische Annahmen an die Verteilungen von (X_i) und (Z_i) aus. Deshalb erhält man im abgewandelten Standardmodell durch $\widehat{\text{EZ}}_1$ und $\widehat{\text{EZ}}_1^2$ nicht-parametrische Schätzer für das erste und zweite Moment der Lebensdauervertelung. Mit diesen kann natürlich auch die Varianz nicht-parametrisch geschätzt werden.

Anwendung I: Die Gammaverteilung

Wir demonstrieren im Folgenden anhand der Gamma- und der inversen Gauß-Verteilung, wie aufbauend auf den bisherigen Ergebnissen Momentenschätzer für die Verteilungsparameter bestimmt werden können. Bei unabhängig und identisch $\Gamma(\alpha, \lambda)$ -verteilten Lebensdauern ist $EZ_1 = \frac{\alpha}{\lambda}$ und $EZ_1^2 = \frac{\alpha(\alpha+1)}{\lambda^2}$. Für die Abbildung

$$F_2 \left(\begin{pmatrix} a \\ b \end{pmatrix} \right) := \begin{pmatrix} \frac{a^2}{b-a^2} \\ \frac{a}{b-a^2} \end{pmatrix}$$

folgt $(\alpha, \lambda)' = F_2((E_{(\alpha, \lambda)'}, Z_1, E_{(\alpha, \lambda)'}, Z_1^2)')$, dabei bezeichnet wie in der Einleitung erwähnt $E_{(\alpha, \lambda)'}$ den Erwartungswert bezüglich des wahren Parametervektors $(\alpha, \lambda)'$. Das Einsetzprinzip führt dann auf die beiden durch

$$\begin{pmatrix} \hat{\alpha}_n^{\text{MM}} \\ \hat{\lambda}_n^{\text{MM}} \end{pmatrix} := F_2 \left(\begin{pmatrix} \widehat{EZ_1}^{\text{MM}} \\ \widehat{EZ_1^2}^{\text{MM}} \end{pmatrix} \right)$$

definierten Schätzer $\hat{\alpha}_n^{\text{MM}}$ und $\hat{\lambda}_n^{\text{MM}}$, wobei wir $\hat{\alpha}_n$ für $\hat{\alpha}_n^{\text{MM}}$ und $\hat{\lambda}_n$ für $\hat{\lambda}_n^{\text{MM}}$ schreiben, wenn Verwechslungen ausgeschlossen sind. Auch für die Schätzer $\hat{\alpha}_n$ und $\hat{\lambda}_n$ ergibt sich asymptotische Normalität.

Satz 3.2.15. *Es gilt*

$$\sqrt{n} \left(\begin{pmatrix} \hat{\alpha}_n^{\text{MM}} \\ \hat{\lambda}_n^{\text{MM}} \end{pmatrix} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \right) \xrightarrow{d} U, U \sim N(0, \Sigma_\Gamma)$$

mit

$$\Sigma_\Gamma = \begin{pmatrix} 2\alpha(\alpha + 1 + 2\lambda^2 \text{var}(X_1)) & 2\lambda(\alpha + 1 + 2\lambda^2 \text{var}(X_1)) \\ 2\lambda(\alpha + 1 + 2\lambda^2 \text{var}(X_1)) & \frac{\lambda^2}{\alpha}(2\alpha + 3 + 4\lambda^2 \text{var}(X_1)) \end{pmatrix}.$$

Beweis. Die Abbildung F_2 ist differenzierbar in $(a, b)'$ mit der Jacobi-Matrix

$$J \left(\begin{pmatrix} a \\ b \end{pmatrix} \right) = \frac{1}{(b-a^2)^2} \begin{pmatrix} 2ab & -a^2 \\ b+a^2 & -a \end{pmatrix}.$$

Dies ergibt für $a = \frac{\alpha}{\lambda}$ und $b = \frac{\alpha(\alpha+1)}{\lambda^2}$

$$J := J \left(\begin{pmatrix} E_{(\alpha, \lambda)'}, Z_1 \\ E_{(\alpha, \lambda)'}, Z_1^2 \end{pmatrix} \right) = \frac{\lambda^3}{\alpha} \begin{pmatrix} \frac{2\alpha(\alpha+1)}{\lambda^2} & -\frac{\alpha}{\lambda} \\ \frac{2\alpha+1}{\lambda} & -1 \end{pmatrix}.$$

Ferner sind $\widehat{EZ_1}$ und $\widehat{EZ_1^2}$ nach Satz 3.2.11 asymptotisch normal mit einer Varianzmatrix der Form $A\Sigma A'$. Für $Z_1 \sim \Gamma(\alpha, \lambda)$ gilt

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2EX_1 & 1 & 2 & -2\frac{\alpha}{\lambda} \end{pmatrix}$$

und Σ ist symmetrisch mit den Einträgen

$$\begin{aligned}\sigma_{11} &= \frac{\alpha}{\lambda^2}, & \sigma_{12} &= \frac{2\alpha(\alpha+1)}{\lambda^3}, & \sigma_{13} &= \frac{\alpha}{\lambda^2} \mathbb{E}X_1, & \sigma_{14} &= 0, \\ \sigma_{22} &= \frac{2\alpha(\alpha+1)(2\alpha+3)}{\lambda^4}, & \sigma_{23} &= \frac{2\alpha(\alpha+1)}{\lambda^3} \mathbb{E}X_1, & \sigma_{24} &= 0, \\ \sigma_{33} &= \frac{\alpha}{\lambda^2} (\alpha \operatorname{var}(X_1) + \mathbb{E}X_1^2), & \sigma_{34} &= \frac{\alpha}{\lambda} \operatorname{var}(X_1), & \sigma_{44} &= \operatorname{var}(X_1).\end{aligned}$$

Da $(\hat{\alpha}_n, \hat{\lambda}_n)'$ aus $(\widehat{\mathbb{E}Z_1}, \widehat{\mathbb{E}Z_1^2})'$ unter Verwendung von F_2 hervorgeht, ist auch der erstgenannte Vektor asymptotisch normal mit der Varianzmatrix $J A \Sigma A' J'$. Rechnet man Letzteres aus, ergibt sich die behauptete Form. \square

Man beachte, dass die Verteilung der Geburtszeitpunkte in Σ_Γ nur über ihre Varianz eingeht. Einige der im Zusammenhang mit Definition 3.2.12 angesprochenen Eigenschaften finden sich auch bei diesen Schätzern, wie der nächste Satz zeigt.

Satz 3.2.16. *Es gilt: $\hat{\alpha}_n$ ist skalierungsinvariant und $\hat{\lambda}_n$ homogen von Grad -1.*

Beweis. Mit Satz 3.2.13 erhalten wir

$$\begin{aligned}\hat{\alpha}_n(aX_1, \dots, aX_n, aY_1, \dots, aY_n) &= \frac{(\widehat{\mathbb{E}Z_1}(aX_1, \dots, aX_n, aY_1, \dots, aY_n))^2}{\widehat{\mathbb{E}Z_1^2}(aX_1, \dots, aX_n, aY_1, \dots, aY_n) - (\widehat{\mathbb{E}Z_1}(aX_1, \dots, aX_n, aY_1, \dots, aY_n))^2} \\ &= \frac{(a \cdot \widehat{\mathbb{E}Z_1}(X_1, \dots, X_n, Y_1, \dots, Y_n))^2}{a^2 (\widehat{\mathbb{E}Z_1^2}(X_1, \dots, X_n, Y_1, \dots, Y_n) - (\widehat{\mathbb{E}Z_1}(X_1, \dots, X_n, Y_1, \dots, Y_n))^2)} \\ &= \hat{\alpha}_n(X_1, \dots, X_n, Y_1, \dots, Y_n)\end{aligned}$$

und

$$\begin{aligned}\hat{\lambda}_n(aX_1, \dots, aX_n, aY_1, \dots, aY_n) &= \frac{\widehat{\mathbb{E}Z_1}(aX_1, \dots, aX_n, aY_1, \dots, aY_n)}{\widehat{\mathbb{E}Z_1^2}(aX_1, \dots, aX_n, aY_1, \dots, aY_n) - (\widehat{\mathbb{E}Z_1}(aX_1, \dots, aX_n, aY_1, \dots, aY_n))^2} \\ &= \frac{a \cdot \widehat{\mathbb{E}Z_1}(X_1, \dots, X_n, Y_1, \dots, Y_n)}{a^2 (\widehat{\mathbb{E}Z_1^2}(X_1, \dots, X_n, Y_1, \dots, Y_n) - (\widehat{\mathbb{E}Z_1}(X_1, \dots, X_n, Y_1, \dots, Y_n))^2)} \\ &= \frac{1}{a} \hat{\lambda}_n(X_1, \dots, X_n, Y_1, \dots, Y_n).\end{aligned} \quad \square$$

Wir veranschaulichen die vorhergehenden theoretischen Überlegungen durch Simulationen für einen Spezialfall. Dabei sind die gerade bewiesenen Eigenschaften von $\hat{\alpha}_n$ und $\hat{\lambda}_n$ in Verbindung mit einer Eigenschaft der Gammaverteilung nützlich. Sind $\alpha, \lambda, \theta \in \mathbb{R}^+$, so ist bekanntlich $\frac{\lambda}{\theta} X \sim \Gamma(\alpha, \theta)$ für $X \sim \Gamma(\alpha, \lambda)$. Bei Vorliegen einer Stichprobe aus

einer $\Gamma(\alpha, \lambda)$ -Verteilung erhält man dementsprechend eine Stichprobe aus einer $\Gamma(\alpha, \theta)$ -Verteilung, indem man alle Werte mit $\frac{\lambda}{\theta}$ multipliziert. Wurde für feste Werte α und λ sowie für Daten aus einer $\Gamma(\alpha, \lambda)$ -Verteilung die Varianz von $\hat{\lambda}_n$ bestimmt, dann folgt zusammen mit Satz 3.2.16 bei Daten aus einer $\Gamma(\alpha, \theta)$ -Verteilung für den Schätzer $\hat{\theta}_n$

$$\text{var}_{(\alpha, \theta)'}(\hat{\theta}_n) = \text{var}_{(\alpha, \lambda)'}\left(\frac{\theta}{\lambda}\hat{\lambda}_n\right) = \frac{\theta^2}{\lambda^2} \text{var}_{(\alpha, \lambda)'}(\hat{\lambda}_n).$$

Bei einer Simulation mit festem α muss also nur $\text{var}_{(\alpha, \lambda)'}(\hat{\lambda}_n)$ für einen Wert von λ bestimmt werden, die Varianzen für andere λ -Werte (und demselben α) lassen sich dann sofort berechnen. Dass dies auch im Grenzfall, also für $\text{var}(\hat{\lambda}^{\text{MM}})$, gilt, kann sofort aus dem Ausdruck für die zugehörige Varianz in Satz 3.2.15 abgelesen werden. Wir untersuchen nun in einem Beispiel, wie sich die asymptotische Varianz der Schätzer $\hat{\alpha}_n$ und $\hat{\lambda}_n$ im Vergleich zu der bestimmter Maximum-Likelihood-Schätzer verhält. Diese ML-Schätzer haben in Bezug auf Multiplikation mit einer positiven Konstanten die gleiche Eigenschaft wie die Momentenschätzer (siehe Lemma B.1 bzw. Gleichung (B.1) in Anhang B). Dies ermöglicht, dass bei der Simulation nur ein Wert von λ pro α -Wert verwendet werden braucht.

Beispiel 3.2.17. Die Geburtszeitpunkte (X_i) seien unabhängig und identisch $\text{Exp}(\lambda)$ - und die Lebensdauern (Z_i) unabhängig und identisch $\Gamma(\alpha, \lambda)$ -verteilt. Bestimmen wir im abgewandelten Standardmodell die Schätzer $\hat{\alpha}_n^{\text{MM}}$ und $\hat{\lambda}_n^{\text{MM}}$, so gilt nach Satz 3.2.15 für die asymptotischen Varianzen $\text{var}_{(\alpha, \lambda)'}(\hat{\alpha}^{\text{MM}}) = 2\alpha(\alpha+3)$ und $\text{var}_{(\alpha, \lambda)'}(\hat{\lambda}^{\text{MM}}) = 2\lambda^2 + \frac{7\lambda^2}{\alpha}$. Insbesondere ist $\text{var}_{(\alpha, \lambda)'}(\hat{\alpha}^{\text{MM}})$ unabhängig von λ . Dies passt zu den bisherigen Ergebnissen: Da man bei Vorliegen einer Stichprobe aus einer $\Gamma(\alpha, \lambda)$ -Verteilung eine Stichprobe aus einer $\Gamma(\alpha, \theta)$ -Verteilung erhält, indem man alle Werte der ursprünglichen Stichprobe mit der positiven Konstanten $\frac{\lambda}{\theta}$ multipliziert, $\hat{\alpha}_n^{\text{MM}}$ aber skalierungsinvariant ist, hat diese Veränderung keinen Einfluss auf den Schätzer $\hat{\alpha}_n^{\text{MM}}$, also insbesondere auch nicht auf seine Varianz. Dies gilt dann auch für die Grenzvarianz $\text{var}(\hat{\alpha}^{\text{MM}})$. Die (Y_i) bilden nun eine Stichprobe aus einer $\text{Exp}(\lambda) \star \Gamma(\alpha, \lambda) = \Gamma(\alpha + 1, \lambda)$ -Verteilung. Werden daraus die Maximum-Likelihood-Schätzer $\hat{\alpha}_n^{\text{ML}}$ und $\hat{\lambda}_n^{\text{ML}}$ ohne Berücksichtigung der (X_i) bestimmt, so gilt nach Gleichung (B.1) (in Anhang B) für die asymptotischen Varianzen

$$\text{var}_{(\alpha, \lambda)'}(\hat{\alpha}^{\text{ML}}) = \frac{\alpha + 1}{\Psi'(\alpha + 1)(\alpha + 1) - 1}, \quad \text{var}_{(\alpha, \lambda)'}(\hat{\lambda}^{\text{ML}}) = \frac{\lambda^2 \Psi'(\alpha + 1)}{\Psi'(\alpha + 1)(\alpha + 1) - 1}.$$

In den Abbildungen 3.1 und 3.3 sind die asymptotischen Varianzen der Momenten- und Maximum-Likelihood-Schätzer gegeneinander aufgetragen. Dabei haben wir uns aus oben genannten Gründen in Abb. 3.3 auf den Fall $\lambda = 1$ beschränkt. Abgesehen von kleinen Werten von α ist die Grenzvarianz von $\hat{\alpha}_n^{\text{ML}}$ immer kleiner als die von $\hat{\alpha}_n^{\text{MM}}$. Allerdings sind beide Varianzen wachsend in α mit einer ähnlichen Steigung und man kann $\lim_{\alpha \rightarrow \infty} \frac{\text{var}_{(\alpha, \lambda)'}(\hat{\alpha}^{\text{MM}})}{\text{var}_{(\alpha, \lambda)'}(\hat{\alpha}^{\text{ML}})} = 1$ zeigen. Bei $\alpha_0 = 0.5864\dots$ sind beide Varianzen gleich, für Werte

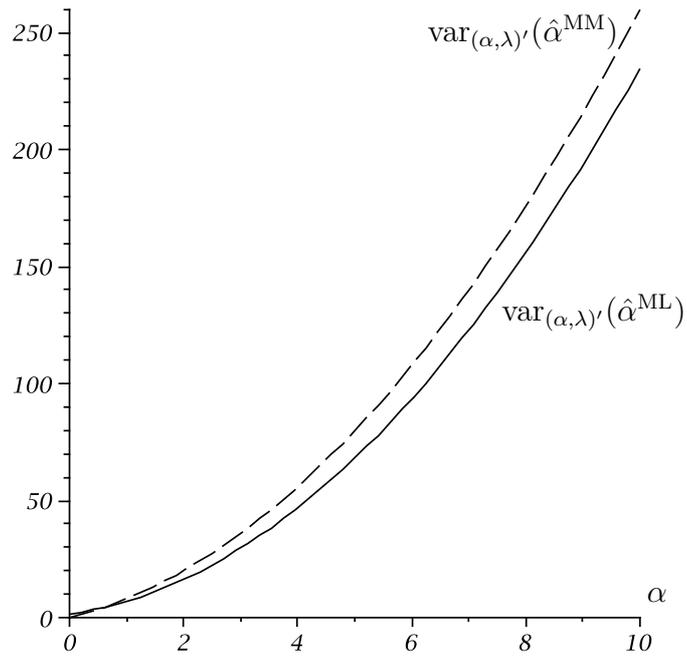


Abbildung 3.1. Vergleich der Grenzvarianzen von $\hat{\alpha}_n^{\text{ML}}$ und $\hat{\alpha}_n^{\text{MM}}$

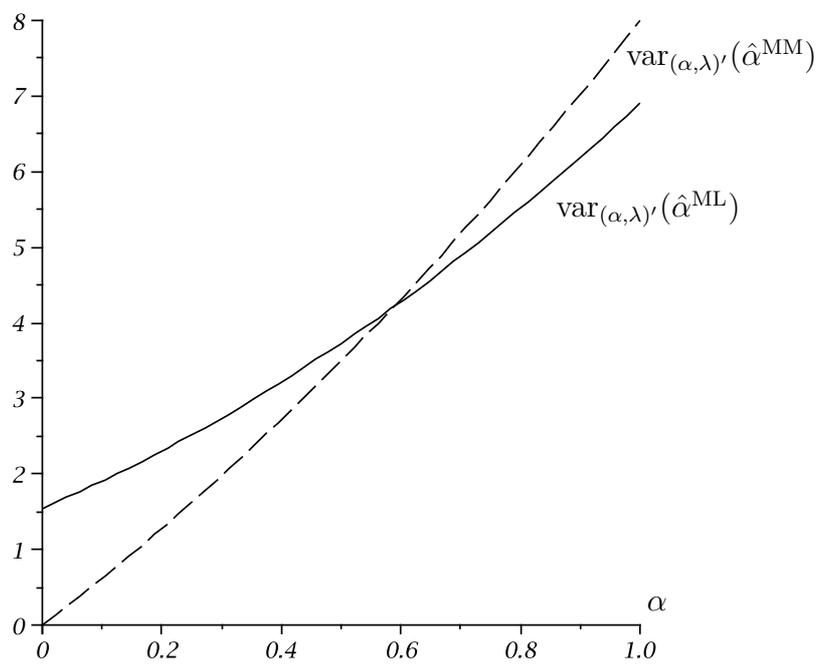


Abbildung 3.2. Vergleich der Grenzvarianzen von $\hat{\alpha}_n^{\text{ML}}$ und $\hat{\alpha}_n^{\text{MM}}$ für kleine Werte von α

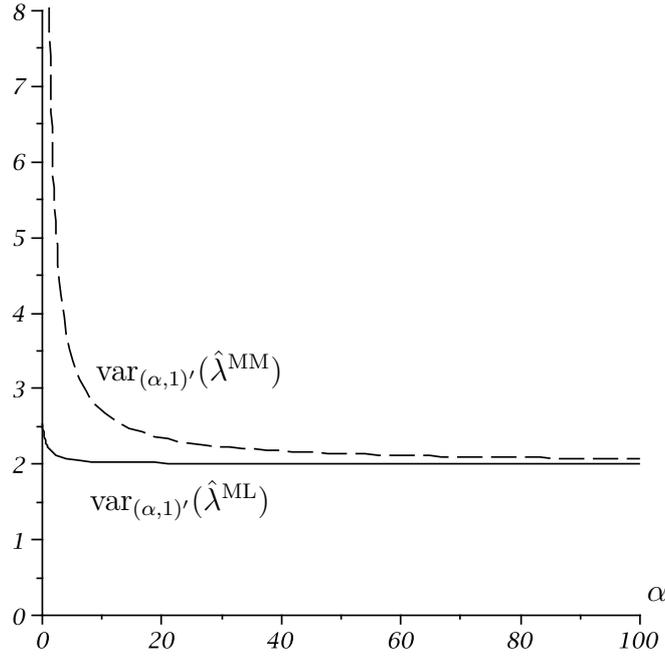


Abbildung 3.3. Vergleich der Grenzvarianzen von $\hat{\lambda}_n^{\text{ML}}$ und $\hat{\lambda}_n^{\text{MM}}$ bei $\lambda = 1$

links davon ist $\text{var}(\hat{\alpha}^{\text{MM}})$ kleiner. Für $\alpha \rightarrow 0$ konvergiert $\text{var}_{(\alpha,\lambda)'}(\hat{\alpha}^{\text{MM}})$ gegen 0, während die Grenzvarianz des ML-Schätzers gegen $(\frac{1}{6}\pi^2 - 1)^{-1} \approx 1.55$ geht. Zur Veranschaulichung des Verhaltens bei kleinem α dient Abb. 3.2. Im Vergleich ist $\text{var}(\hat{\lambda}^{\text{MM}})$ bei kleinen Werten von α deutlich größer als $\text{var}(\hat{\lambda}^{\text{ML}})$: Während die Grenzvarianz von $\hat{\lambda}_n^{\text{ML}}$ für $\alpha \rightarrow 0$ gegen den Wert $\frac{\pi^2}{\pi^2-6} \approx 2.55$ konvergiert, wird sie bei $\hat{\lambda}_n^{\text{MM}}$ unendlich groß. Mit wachsendem α nähern sich beide Varianzen von oben dem Wert 2 an, allerdings geschieht dies bei $\text{var}(\hat{\lambda}^{\text{ML}})$ deutlich schneller. Auch hier ist es nicht schwer, $\lim_{\alpha \rightarrow \infty} \frac{\text{var}_{(\alpha,1)'}(\hat{\lambda}^{\text{MM}})}{\text{var}_{(\alpha,1)'}(\hat{\lambda}^{\text{ML}})} = 1$ zu zeigen. Die ML-Schätzer besitzen für fast alle Parameter-Werte kleinere asymptotische Varianzen, obwohl sie die Geburtszeitpunkte nicht berücksichtigen. Nur bei kleinen Werten von α , wenn die Lebensdauern sehr kurz sind ($\mathbb{E}Z_1 = \frac{\alpha}{\lambda}$), ergibt sich durch die zusätzliche Berücksichtigung ein Vorteil bei der Grenzvarianz für den Momentenschätzer von α . Hier setzt sich der am Ende von Abschnitt 1.5.3 beschriebene Effekt fort.

Da es sich bei den hier bestimmten Varianzen um Grenzwertresultate handelt, ist für die Anwendung von Interesse, wie gut die Annäherung an diese Werte für endliches n ist. In Tabelle 3.2 sind die mit $\widehat{\text{var}}(\cdot)$ bezeichneten Stichprobenvarianzen von Simulationen für $n = 10^2, 10^3, 10^4, \alpha \in \{0.2, 8\}$ und 1000 Wiederholungen pro Kombination von α und n aufgeführt. Bei der Wahl der Werte wurde der nach dem Beispiel zu erwartenden Gestalt der Kurven Rechnung getragen. Um die Grenzvarianzen mit den jeweiligen empirischen Varianzen vergleichen zu können, wurden letztgenannte jeweils mit n multipliziert. Für

| | | | | | |
|-----------------------------|--|--|---|---|-------|
| $\alpha = 0.2, \lambda = 1$ | $\text{var}(\hat{\alpha}^{\text{MM}})$ | $\text{var}(\hat{\alpha}^{\text{ML}})$ | $\text{var}(\hat{\lambda}^{\text{MM}})$ | $\text{var}(\hat{\lambda}^{\text{ML}})$ | |
| | 1.28 | 2.304 | 37 | 2.433 | |
| | $\widehat{\text{var}}(\hat{\alpha}_n^{\text{MM}}) \cdot n$ | $\widehat{\text{var}}(\hat{\alpha}_n^{\text{ML}}) \cdot n$ | $\widehat{\text{var}}(\hat{\lambda}_n^{\text{MM}}) \cdot n$ | $\widehat{\text{var}}(\hat{\lambda}_n^{\text{ML}}) \cdot n$ | |
| | 10^2 | 385.508 | 2.401 | 18205.47 | 2.582 |
| | 10^3 | 1.491 | 2.245 | 44.457 | 2.378 |
| 10^4 | 1.259 | 2.325 | 36.422 | 2.305 | |
| $\alpha = 8, \lambda = 1$ | $\text{var}(\hat{\alpha}^{\text{MM}})$ | $\text{var}(\hat{\alpha}^{\text{ML}})$ | $\text{var}(\hat{\lambda}^{\text{MM}})$ | $\text{var}(\hat{\lambda}^{\text{ML}})$ | |
| | 176 | 156.228 | 2.875 | 2.040 | |
| | $\widehat{\text{var}}(\hat{\alpha}_n^{\text{MM}}) \cdot n$ | $\widehat{\text{var}}(\hat{\alpha}_n^{\text{ML}}) \cdot n$ | $\widehat{\text{var}}(\hat{\lambda}_n^{\text{MM}}) \cdot n$ | $\widehat{\text{var}}(\hat{\lambda}_n^{\text{ML}}) \cdot n$ | |
| | 10^2 | 189.750 | 165.497 | 3.058 | 2.132 |
| | 10^3 | 184.734 | 167.394 | 3.010 | 2.185 |
| 10^4 | 180.914 | 159.346 | 2.961 | 2.066 | |

Tabelle 3.2. Vergleich von Grenz- und Stichprobenvarianz bei einer Gammaverteilung

$\alpha = 0.2$ und $n = 100$ ist bei den Momentenschätzern eine starke Abweichung der aus der Stichprobe bestimmten Varianzen von den Grenzvarianzen zu beobachten. Dies liegt zum einen daran, dass für dieses n die Varianzen der Ausgangsschätzer $(\widehat{\text{E}Z_1})_n$ und $(\widehat{\text{E}Z_1^2})_n$ noch deutlich von 0 verschieden sind. Zum anderen sind die auftretenden Lebensdauern im Vergleich zu den Geburtszeitpunkten im Mittel kleiner, deswegen können bei der Berechnung des Schätzers kleine und negative Werte im Nenner vorkommen. Bei $\alpha = 8$ tritt der zweite Effekt wegen der größeren Lebensdauern nicht auf. Insgesamt ist sowohl bei den Momenten- als auch bei den ML-Schätzern die Annäherung der empirischen Varianzen an die Grenzwerte recht gut. Insbesondere stützt dies die These, dass auch für $n \in \mathbb{N}$ der Schätzer $\hat{\lambda}_n^{\text{ML}}$ gegenüber $\hat{\lambda}_n^{\text{MM}}$ in Bezug auf die Variabilität vorzuziehen ist, während beim Schätzen von α keine pauschale Aussage möglich ist. \triangleleft

Anwendung II: Die inverse Gauß-Verteilung

Sind die (Z_i) unabhängig und identisch $\text{IG}(\mu, \lambda)$ -verteilt, ist $\text{E}Z_1 = \mu$ und $\text{E}Z_1^2 = \frac{\mu^3}{\lambda} + \mu^2$. Für die Abbildung

$$F_3 \left(\begin{pmatrix} a \\ b \end{pmatrix} \right) := \begin{pmatrix} a \\ \frac{a^3}{b-a^2} \end{pmatrix}$$

folgt $(\mu, \lambda)' = F_3((\text{E}_{(\mu, \lambda)'} Z_1, \text{E}_{(\mu, \lambda)'} Z_1^2)')$. Das Einsetzprinzip führt hier auf die beiden durch

$$\begin{pmatrix} \hat{\mu}_n^{\text{MM}} \\ \hat{\lambda}_n^{\text{MM}} \end{pmatrix} := F_3 \left(\begin{pmatrix} (\widehat{\text{E}Z_1}_n)^{\text{MM}} \\ (\widehat{\text{E}Z_1^2}_n)^{\text{MM}} \end{pmatrix} \right)$$

definierten Schätzer $\hat{\mu}_n^{\text{MM}}$ und $\hat{\lambda}_n^{\text{MM}}$. Wie zuvor wird der Index MM weggelassen, wenn keine Verwechslungsgefahr besteht. Auch für $\hat{\mu}_n$ und $\hat{\lambda}_n$ ergibt sich asymptotische Normalität.

Satz 3.2.18. *Es gilt*

$$\sqrt{n} \left(\begin{pmatrix} \hat{\mu}_n^{\text{MM}} \\ \hat{\lambda}_n^{\text{MM}} \end{pmatrix} - \begin{pmatrix} \mu \\ \lambda \end{pmatrix} \right) \xrightarrow{d} U, U \sim N(0, \Sigma_{\text{IG}})$$

mit

$$\Sigma_{\text{IG}} = \begin{pmatrix} \frac{\mu^3}{\lambda} & 0 \\ 0 & \frac{2\lambda(\mu^3\lambda + 3\mu^4 + 2\lambda^2 \text{var}(X_1))}{\mu^3} \end{pmatrix}.$$

Beweis. Die Abbildung F_3 ist differenzierbar in $(a, b)'$ mit der Jacobi-Matrix

$$J \left(\begin{pmatrix} a \\ b \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 \\ \frac{a^2(3b-a^2)}{(b-a^2)^2} & \frac{-a^3}{(b-a^2)^2} \end{pmatrix},$$

so dass sich für $a = \mu$ und $b = \frac{\mu^3}{\lambda} + \mu^2$

$$J := J \left(\begin{pmatrix} \text{E}_{(\mu, \lambda)'} Z_1 \\ \text{E}_{(\mu, \lambda)'} Z_1^2 \end{pmatrix} \right) = \begin{pmatrix} 1 & 0 \\ \frac{\lambda}{\mu}(3 + 2\frac{\lambda}{\mu}) & \frac{-\lambda^2}{\mu^3} \end{pmatrix}$$

ergibt. Ferner ist $(\widehat{\text{E}Z_1}, \widehat{\text{E}Z_1^2})'$ nach Satz 3.2.11 asymptotisch normal mit einer Varianzmatrix der Form $A\Sigma A'$. Für $Z_1 \sim \text{IG}(\mu, \lambda)$ ist

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2\text{E}X_1 & 1 & 2 & -2\mu \end{pmatrix}$$

und Σ ergibt sich, indem man die nach Satz 3.2.10 benötigten Größen ausrechnet und dabei benutzt, dass laut [CF89, S. 11]

$$\text{E}Z_1 = \mu, \text{E}Z_1^2 = \mu^2 + \frac{\mu^3}{\lambda}, \text{E}Z_1^3 = \mu^3 + \frac{3\mu^4}{\lambda} + \frac{3\mu^5}{\lambda^2}, \text{E}Z_1^4 = \mu^4 + \frac{6\mu^5}{\lambda} + \frac{15\mu^6}{\lambda^2} + \frac{15\mu^7}{\lambda^3}$$

gilt. Da $(\hat{\mu}_n, \hat{\lambda}_n)'$ aus $(\widehat{\text{EZ}}_1, \widehat{\text{EZ}}_1^2)'$ unter Verwendung von F_3 hervorgeht, besitzt auch dieser Vektor asymptotisch eine Normalverteilung mit der Varianzmatrix $JA\Sigma A'J'$. Berechnung des Matrizenprodukts ergibt die genannte Form. \square

Die Matrix Σ_{IG} hat zwei bemerkenswerte Eigenschaften. Zum einen geht die Verteilung der Geburtszeitpunkte nur über die Varianz ein, zum anderen hat sie Diagonalform. Letzteres bedeutet, dass die Schätzer $\hat{\mu}_n^{\text{ML}}$ und $\hat{\lambda}_n^{\text{ML}}$ im Grenzfall $n \rightarrow \infty$ unkorreliert und damit wegen der gemeinsamen Normalverteilung auch unabhängig sind.

Auch bei diesen Schätzern findet sich eine der im Zusammenhang mit Definition 3.2.12 angesprochenen Eigenschaften, wie der nächste Satz zeigt. Der Beweis verläuft analog zum Beweis von Satz 3.2.16 durch Nachrechnen.

Satz 3.2.19. *Die Schätzer $\hat{\mu}_n$ und $\hat{\lambda}_n$ sind skalierungsäquivalent.*

Nutzt man aus, dass für $X \sim \text{IG}(\mu, \lambda)$ und $t > 0$ folgt $tX \sim \text{IG}(t\mu, t\lambda)$, so kann Satz 3.2.19 dazu benutzt werden, den Aufwand bei Simulationen zu verringern. Wurden für feste μ und λ sowie für Daten aus einer $\text{IG}(\mu, \lambda)$ -Verteilung die asymptotischen Varianzen von $\hat{\mu}_n$ und $\hat{\lambda}_n$ bestimmt, so gilt

$$\text{var}_{(t\mu, t\lambda)'}(\hat{\mu}_n) = t^2 \cdot \text{var}_{(\mu, \lambda)'}(\hat{\mu}_n), \quad \text{var}_{(t\mu, t\lambda)'}(\hat{\lambda}_n) = t^2 \cdot \text{var}_{(\mu, \lambda)'}(\hat{\lambda}_n).$$

Dass dies auch im Grenzfall, also für $\text{var}(\hat{\mu}^{\text{MM}})$ und $\text{var}(\hat{\lambda}^{\text{MM}})$, erfüllt ist, lässt sich sofort aus Satz 3.2.18 ablesen.

Den Abschluss des Abschnitts bildet die folgende simulative Gegenüberstellung der entwickelten Momentenschätzer $\hat{\mu}_n$ und $\hat{\lambda}_n$ mit bestimmten Maximum-Likelihood-Schätzern. Der Aufwand wird dabei dadurch verringert, dass sich die ML-Schätzer bei einer inversen Gauß-Verteilung unter Multiplikation der Daten mit einer positiven Konstanten genauso verhalten wie die Momentenschätzer (siehe Lemma B.2 bzw. Gleichung (B.2) in Anhang B).

Beispiel 3.2.20. Die Geburtszeitpunkte (X_i) und die Lebensdauern (Z_i) seien jeweils unabhängig und identisch $\text{IG}(\mu, \lambda)$ -verteilt. Dann ergeben sich im abgewandelten Standardmodell nach Satz 3.2.18 die asymptotischen Varianzen $\text{var}_{(\mu, \lambda)'}(\hat{\mu}^{\text{MM}}) = \frac{\mu^3}{\lambda}$ und $\text{var}_{(\mu, \lambda)'}(\hat{\lambda}^{\text{MM}}) = 6\lambda^2 + 6\lambda\mu$ für die Schätzer $\hat{\mu}_n^{\text{MM}}$ bzw. $\hat{\lambda}_n^{\text{MM}}$. Die (Y_i) bilden eine Stichprobe aus einer $\text{IG}(\mu, \lambda) \star \text{IG}(\mu, \lambda) = \text{IG}(2\mu, 4\lambda)$ -Verteilung. Bestimmung der Maximum-Likelihood-Schätzer $\hat{\mu}_n^{\text{ML}}$ und $\hat{\lambda}_n^{\text{ML}}$ ohne Berücksichtigung der (X_i) führt nach (B.2) (in Anhang B) zu den Grenzvarianzen

$$\text{var}_{(2\mu, 4\lambda)'}(\hat{\mu}^{\text{ML}}) = \frac{(2\mu)^3}{4\lambda} = \frac{2\mu^3}{\lambda}, \quad \text{var}_{(2\mu, 4\lambda)'}(\hat{\lambda}^{\text{ML}}) = 2 \cdot (4\lambda)^2 = 32\lambda^2.$$

| | | | | |
|------------------------|---|---|---|---|
| $\mu = 1, \lambda = 1$ | $\text{var}(\hat{\mu}^{\text{MM}})$ 1 | $\text{var}(\tilde{\mu}^{\text{ML}})$ 0.5 | $\text{var}(\hat{\lambda}^{\text{MM}})$ 12 | $\text{var}(\tilde{\lambda}^{\text{ML}})$ 2 |
| n | $\widehat{\text{var}}(\hat{\mu}_n^{\text{MM}}) \cdot n$ | $\widehat{\text{var}}(\tilde{\mu}_n^{\text{ML}}) \cdot n$ | $\widehat{\text{var}}(\hat{\lambda}_n^{\text{MM}}) \cdot n$ | $\widehat{\text{var}}(\tilde{\lambda}_n^{\text{ML}}) \cdot n$ |
| 10^2 | 1.001 | 0.511 | 17.422 | 2.257 |
| 10^3 | 0.977 | 0.493 | 11.904 | 2.105 |
| 10^4 | 0.952 | 0.508 | 11.929 | 1.990 |

Tabelle 3.3. Vergleich von Grenz- und Stichprobenvarianz bei einer inversen Gauß-Verteilung

Da bei der ML-Methode die Parameter 2μ und 4λ geschätzt werden, definieren wir die beiden Schätzer $\tilde{\mu}_n^{\text{ML}} := \frac{\hat{\mu}_n^{\text{ML}}}{2}$ und $\tilde{\lambda}_n^{\text{ML}} := \frac{\hat{\lambda}_n^{\text{ML}}}{4}$, und erhalten für diese die asymptotischen Varianzen

$$\text{var}_{(\mu,\lambda)}(\tilde{\mu}^{\text{ML}}) = \frac{\text{var}(\hat{\mu}^{\text{ML}})}{4} = \frac{\mu^3}{2\lambda}, \quad \text{var}_{(\mu,\lambda)}(\tilde{\lambda}^{\text{ML}}) = \frac{\text{var}(\hat{\lambda}^{\text{ML}})}{16} = 2\lambda^2.$$

Ein Vergleich mit den entsprechenden Werten der Momentenschätzer zeigt, dass diese Größen bei den Maximum-Likelihood-Schätzern immer kleiner sind. Genauer ist die Grenzvarianz von $\tilde{\mu}_n^{\text{ML}}$ immer halb so groß wie die von $\hat{\mu}_n^{\text{MM}}$ und $\text{var}(\hat{\lambda}^{\text{MM}})$ ist mehr als dreimal so groß wie $\text{var}(\tilde{\lambda}^{\text{ML}})$. Das theoretische Verhältnis der asymptotischen Varianzen fand sich auch näherungsweise in Simulationen für endliches n wieder. Exemplarisch ist das in Tabelle 3.3 aufgeführte Ergebnis bei 1000 Wiederholungen, dabei wurde die gleich Notation wie im vorhergehenden Beispiel verwendet. \triangleleft

3.2.4. Deterministische Geburtszeitpunkte

Bei (hier nicht aufgeführten) Simulationen lieferten unsere Momentenschätzer für deterministische Geburtszeitpunkte in manchen Fällen brauchbare Ergebnisse und in manchen nicht. In einem kurzen Abschnitt erklären wir diese Beobachtung theoretisch, ohne darauf so ausführlich wie auf das abgewandelte Standardmodell einzugehen. Wir zeigen zunächst, dass die Aussage von Satz 3.2.2 in dieser Situation nicht immer richtig ist.

Satz 3.2.21. *Gilt neben A2 bis A5 noch, dass $x_i := ci^\kappa$ für $i = 1, \dots, n$ und $\kappa, c \in \mathbb{R}^+$ ist, also deterministische Werte x_i vorliegen, dann folgt für ein $\gamma \in (0, \infty)$*

$$\lim_{n \rightarrow \infty} \text{var} \left(\frac{1}{n} \sum_{i=1}^n x_i (Z_i - \bar{Z}_{n,1}) \right) = \begin{cases} 0, & \kappa < 0.5, \\ \gamma, & \kappa = 0.5, \\ \infty, & \kappa > 0.5. \end{cases}$$

Beweis. Bei $x_i = ci^\kappa$ für $i = 1, \dots, n$ und ein $\kappa \in \mathbb{R}^+$ ergibt sich mit (3.10) und (3.12)

$$\begin{aligned} \text{var} \left(\frac{1}{n} \sum_{i=1}^n x_i (Z_i - \bar{Z}_{n,1}) \right) &= \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n x_i (Z_i - \bar{Z}_{n,1}) \right)^2 \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n x_i^2 \mathbb{E} (Z_i - \bar{Z}_{n,1})^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n x_i x_j \mathbb{E} (Z_i - \bar{Z}_{n,1})(Z_j - \bar{Z}_{n,1}) \right) \\ &= \frac{c^2}{n^2} \left(\left(1 - \frac{1}{n}\right) \text{var}(Z_1) \sum_{i=1}^n i^{2\kappa} - \frac{1}{n} \text{var}(Z_1) \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n i^\kappa j^\kappa \right). \end{aligned}$$

Wegen

$$\begin{aligned} \sum_{i=1}^n i^{2\kappa} &\sim \int_1^n x^{2\kappa} dx \sim \frac{1}{2\kappa + 1} n^{2\kappa+1}, \\ \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n i^\kappa j^\kappa &\sim \int_1^n \int_1^n x^\kappa y^\kappa dy dx \sim \frac{1}{\kappa + 1} \int_1^n x^\kappa n^{\kappa+1} dx \\ &\sim \frac{1}{\kappa + 1} n^{\kappa+1} \frac{1}{\kappa + 1} n^{\kappa+1} = \frac{1}{(\kappa + 1)^2} n^{2\kappa+2} \end{aligned}$$

erhalten wir

$$\text{var} \left(\frac{1}{n} \sum_{i=1}^n x_i (Z_i - \bar{Z}_{n,1}) \right) \sim \left(\frac{1}{2\kappa + 1} - \frac{1}{(\kappa + 1)^2} \right) \text{var}(Z_1) c^2 n^{2\kappa-1}.$$

Da der Term in der Klammer positiv ist, folgt daraus die Behauptung. \square

Für $k > 0.5$ ließe sich nun eine Aussage ähnlich der von Satz 3.2.6 herleiten. Wir konzentrieren uns deshalb auf den Fall $k < 0.5$.

Satz 3.2.22. *In der Situation aus Satz 3.2.21 mit $k < 0.5$ ist $\widehat{\mathbb{E}Z_1^2}$ konsistent im quadratischen Mittel, d.h. es gilt*

$$\lim_{n \rightarrow \infty} \mathbb{E}(\widehat{\mathbb{E}Z_1^2} - \mathbb{E}Z_1^2)^2 = 0.$$

Beweis. Wegen

$$\mathbb{E}\widehat{\mathbb{E}Z_1^2} = \mathbb{E} \left(\bar{Z}_{n,2} + 2 \left(\frac{1}{n} \sum_{i=1}^n x_i (Z_i - \bar{Z}_{n,1}) \right) \right) = \mathbb{E}Z_1^2$$

ist $\widehat{\text{E}Z_1^2}$ erwartungstreu und damit die Konsistenz im quadratischen Mittel äquivalent zu $\lim_{n \rightarrow \infty} \text{var}(\widehat{\text{E}Z_1^2}) = 0$. Mit dem bekannten Vorgehen erhalten wir

$$\begin{aligned} \text{var}(\widehat{\text{E}Z_1^2}) &= \text{E} \left(\bar{Z}_{n,2} + 2 \left(\frac{1}{n} \sum_{i=1}^n x_i Z_i - \bar{x}_{n,1} \bar{Z}_{n,1} \right) - \text{E}Z_1^2 \right)^2 \\ &= \text{var}(\bar{Z}_{n,2}) + 4 \text{var} \left(\frac{1}{n} \sum_{i=1}^n x_i Z_i - \bar{x}_{n,1} \bar{Z}_{n,1} \right) \\ &\quad + 4 \text{E} \left((\bar{Z}_{n,2} - \text{E}Z_1^2) \left(\frac{1}{n} \sum_{i=1}^n x_i Z_i - \bar{x}_{n,1} \bar{Z}_{n,1} \right) \right). \end{aligned}$$

Der erste Term konvergiert für wachsendes n gegen 0, für $k < \frac{1}{2}$ tut dies nach Satz 3.2.21 auch der zweite Term und der dritte Term verschwindet nach Lemma 3.2.5, wenn man die (x_i) als Zufallsvariablen auffasst, die mit Wahrscheinlichkeit 1 jeweils einen bestimmten Wert annehmen. \square

Wir schließen den Abschnitt mit der Beobachtung, dass es bei $k = 0.5$ eine Art Phasenübergang gibt: Für $k < 0.5$ ist $\widehat{\text{E}Z_1^2}$ konsistent im quadratischen Mittel, für $k > 0.5$ nicht.

3.3. Modellierung als $M/G/\infty$ -Warteschlange

3.3.1. Grundlagen

Aus Satz 3.2.6 ergibt sich, dass der in Gleichung (3.8) definierte Momentenschätzer im (spezialisierten) Standardmodell nicht konsistent im quadratischen Mittel ist. In diesem Abschnitt entwickeln wir einen asymptotisch normalen Momentenschätzer, der in dieser Situation anwendbar ist. Wir benutzen dazu die in Beispiel 1.1.2 eingeführte und dann in Abschnitt 1.2.2 formalisierte Interpretation der Problemstellung als $M/G/\infty$ -Warteschlange.

Unsere Überlegungen setzen an der gleichen Stelle an wie [BD97, S. 673f., 4.C], beschränken sich dabei aber nicht auf das Schätzen der mittleren Bedienzeit. Wir nehmen stattdessen an, dass deren Verteilungsfamilie bereits bekannt ist, und schätzen direkt die zugrunde liegenden Parameter. Dieses Vorgehen steht im Kontrast zu [BP99] (Probleme II und III in den Abschnitten 5 und 6), wo die Verteilungsfunktion der Bedienzeit nicht-parametrisch geschätzt wird, bzw. kann als nächster Schritt nach Bestimmung des nicht-parametrischen Schätzers gesehen werden. Als weitere Annahmen gehen wir davon aus, dass die Rate θ des Ankunftsprozesses bekannt ist und eine „große“ Anzahl an Daten

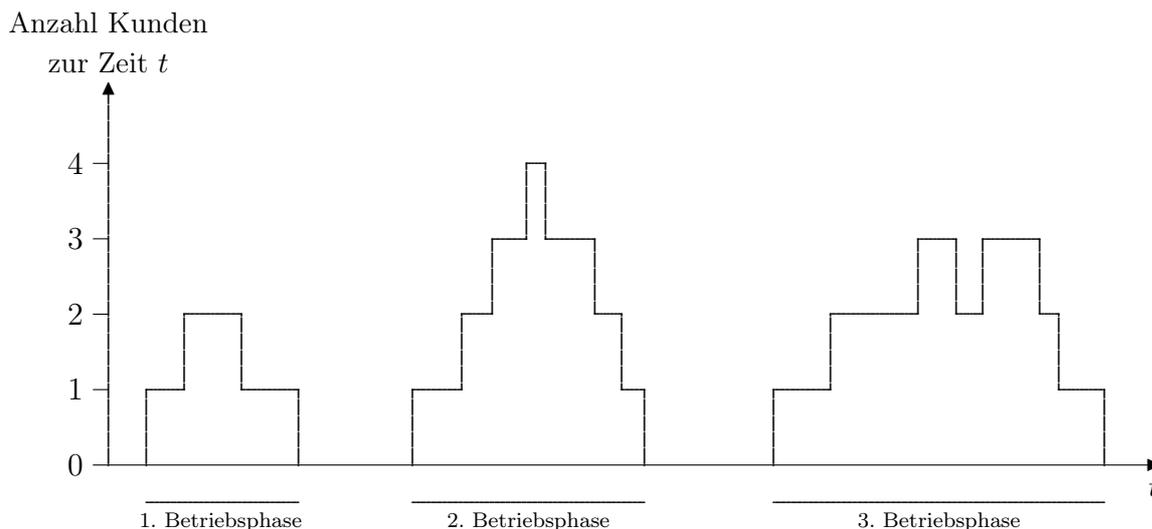


Abbildung 3.4. Veranschaulichung der Betriebsphasen eines $M/G/\infty$ -Systems

vorliegt; was Letzteres genau bedeutet, erläutern wir an geeigneter Stelle. Die Annahme an θ dient im Wesentlichen der Vereinfachung von Notation und hergeleiteten Formeln. Wir untersuchen in Bemerkung 3.3.27 am Ende dieses Abschnitts den Fall, in dem auch die Rate geschätzt wird, bei gammaverteilten Bedienzeiten.

Für den Rest des Abschnitts soll die Bedienzeitverteilung von genau zwei unbekannt Parametern abhängen. Die hier beschriebenen Methoden lassen sich auch auf andere Fälle erweitern, allerdings verzichten wir darauf, dies explizit vorzuführen. Wir betrachten die in Abschnitt 1.2.2 beschriebenen sog. „idle-busy-periods“ einer $M/G/\infty$ -Warteschlange, genauer beschreibe C_i für $i = 1, 2, \dots$ die Länge der i -ten „busy period“ (verwendete deutsche Übersetzung: Betriebsphase), d.h. der Länge des i -ten Zeitraums, während dessen mindestens ein Kunde im System ist. Eine Veranschaulichung findet sich in Abb. 3.4. Die Länge einer Betriebsphase bezeichnen wir kurz als Betriebsdauer.

Da die Betriebsdauern unabhängig und identisch verteilt sind (Übungsaufgabe in [Hal88, S. 81]), können die bekannten Momentenschätzer $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$ als Schätzer für das erste und zweite Moment verwendet werden; diese sind nach dem mehrdimensionalen zentralen Grenzwertsatz asymptotisch normal. Die für die Varianzmatrix benötigten ersten vier Momente der Betriebsdauerverteilung lassen sich durch ein Resultat von Hall in [Hal88] mit etwas Rechenaufwand bestimmen. Aus diesem Resultat geht genauer hervor, dass die Momente Funktionen der Bedienzeitverteilungsparameter (und von θ) sind; folglich erhält man bei Vorliegen von Schätzern für die Momente zumindest implizit auch Schätzer für die Parameter und kann dann wie in Abschnitt 3.2.3 vorgehen. Wir leiten nun zunächst Ausdrücke für die ersten vier Momente her und benötigen dazu noch etwas Notation.

Definition 3.3.1. Es sei X eine nichtnegative Zufallsvariable mit Dichte f_X . Dann heißt

$$\gamma(s) := \int_0^\infty e^{-st} f_X(t) dt$$

die *Laplace-Stieltjes-Transformierte* (LST) der Verteilung von X . Wir sprechen abkürzend auch nur von der LST von X .

Im folgenden Lemma geben wir zwei bekannte (und leicht nachzuprüfende) Eigenschaften der LST ohne Beweis an.

Lemma 3.3.2. *Es sei γ die Laplace-Stieltjes-Transformierte der Verteilung von X . Dann gilt für $k \in \mathbb{N}$*

$$\gamma(0) = 1, \quad \gamma^{(k)}(0) = (-1)^k EX^k,$$

insofern die beteiligten Größen existieren.

Unter Voraussetzung der Existenz ist also eine Möglichkeit, die Momente der Betriebsdauerverteilung zu bestimmen, das Differenzieren der entsprechenden LST. Diese ist bei [Hal88, S. 91] in Theorem 2.2 gegeben.

Satz 3.3.3 (Hall). *Es sei G die Verteilungsfunktion der Bedienzeit. Dann ist die Laplace-Stieltjes-Transformierte γ der Betriebsdauerverteilung gegeben durch*

$$\gamma(s) = 1 + \frac{s}{\theta} - \left(\theta \int_0^\infty e^{-st - \theta \int_0^t 1 - G(x) dx} dt \right)^{-1}.$$

Bevor wir fortfahren, muss noch das im Folgenden beschriebene Problem gelöst werden. Aus Lemma 3.3.2 ist bekannt, dass jede LST in 0 den Wert 1 hat. Damit dies auch beim durch den vorhergehenden Satz gegebenen Ausdruck der Fall ist, muss der dritte Term an dieser Stelle den Kehrwert ∞ haben und deshalb können Probleme bei der Differentiation auftreten. Um diese zu vermeiden, geben wir im folgenden Lemma eine andere Darstellung von γ an.

Lemma 3.3.4. *Die Bedienzeit habe Verteilungsfunktion G und Erwartungswert $\mu < \infty$. Dann gilt für die Funktion γ aus dem vorhergehenden Satz*

$$\gamma(s) = 1 + s \frac{1}{\theta} \left(1 - e^{\theta\mu} \left(1 + \int_0^\infty se^{-st} (e^{\theta \int_t^\infty 1 - G(x) dx} - 1) dt \right)^{-1} \right).$$

Beweis. Es gilt

$$\begin{aligned} \int_0^\infty e^{-st - \theta \int_0^t 1 - G(x) dx} dt &= \frac{e^{-\theta\mu}}{s} \int_0^\infty se^{-st} (e^{\theta \int_t^\infty 1 - G(x) dx} - 1) + se^{-st} dt \\ &= \frac{e^{-\theta\mu}}{s} \left(1 + \int_0^\infty se^{-st} (e^{\theta \int_t^\infty 1 - G(x) dx} - 1) dt \right), \end{aligned}$$

damit ergibt sich die Behauptung nach einigen offensichtlichen Umformungen. □

Auch bei den beiden nachfolgenden Lemmas verzichten wir auf einen Beweis, da die Aussagen durch Anwendung elementarer Ableitungsregeln bzw. ein Computeralgebrasystem sofort nachgewiesen werden können.

Lemma 3.3.5. Für Konstanten $c_1, c_2 \in \mathbb{R}^+$ sei

$$g(s) := 1 + c_1 s (1 - c_2 (f(s))^{-1})$$

mit einer viermal differenzierbaren Funktion f . Dann gilt

$$\begin{aligned} g'(s) &= \frac{c_1}{(f(s))^2} ((f(s))^2 - f(s)c_2 + s f'(s)c_2), \\ g''(s) &= \frac{c_1 c_2}{(f(s))^3} (2f(s)f'(s) + s(f(s)f''(s) - 2(f'(s))^2)), \\ g^{(3)}(s) &= \frac{c_1 c_2}{(f(s))^4} (3(f(s))^2 f''(s) - 6f(s)(f'(s))^2 \\ &\quad + s((f(s))^2 f^{(3)}(s) - 6f(s)f'(s)f''(s) + 6(f'(s))^3)), \\ g^{(4)}(s) &= \frac{c_1 c_2}{(f(s))^5} (4(f(s))^3 f^{(3)}(s) - 24(f(s))^2 f'(s)f''(s) + 24f(s)(f'(s))^3 \\ &\quad + s(36f(s)(f'(s))^2 f''(s) - 6(f(s))^2 (f''(s))^2 - 8(f(s))^2 f'(s)f^{(3)}(s) \\ &\quad + (f(s))^3 f^{(4)}(s) - 24(f'(s))^4)). \end{aligned}$$

Lemma 3.3.6. Für die Funktion $f(s) := 1 + \int_0^\infty s e^{-st} (e^{\theta \int_t^\infty 1-G(x) dx} - 1) dt$ gilt

$$\begin{aligned} f'(s) &= \int_0^\infty (e^{\theta \int_t^\infty 1-G(x) dx} - 1) e^{-st} (1 - st) dt, \\ f''(s) &= -2 \int_0^\infty (e^{\theta \int_t^\infty 1-G(x) dx} - 1) e^{-st} t (1 - \frac{1}{2} st) dt, \\ f^{(3)}(s) &= 3 \int_0^\infty (e^{\theta \int_t^\infty 1-G(x) dx} - 1) e^{-st} t^2 (1 - \frac{1}{3} st) dt. \end{aligned}$$

Nach diesen Vorbereitungen können wir nun das wesentliche Resultat dieses Abschnitts angeben.

Satz 3.3.7. Die Zufallsvariable C gebe die Länge einer Betriebsperiode an. Dann gilt mit den Bezeichnungen von oben und für den Fall, dass die jeweiligen Ausdrücke existieren,

$$\begin{aligned} EC &= \frac{1}{\theta} (e^{\theta\mu} - 1), \quad EC^2 = \frac{2}{\theta} e^{\theta\mu} \int_0^\infty e^{\theta \int_t^\infty 1-G(x) dx} - 1 dt, \\ EC^3 &= \frac{3}{\theta} e^{\theta\mu} \left(2 \left(\int_0^\infty e^{\theta \int_t^\infty 1-G(x) dx} - 1 dt \right)^2 + 2 \int_0^\infty t (e^{\theta \int_t^\infty 1-G(x) dx} - 1) dt \right), \\ EC^4 &= \frac{4}{\theta} e^{\theta\mu} \left(3 \int_0^\infty t^2 (e^{\theta \int_t^\infty 1-G(x) dx} - 1) dt + 6 \left(\int_0^\infty e^{\theta \int_t^\infty 1-G(x) dx} - 1 dt \right)^3 \right) \end{aligned}$$

$$+ 12 \left(\int_0^\infty e^{\theta \int_t^\infty 1-G(x) dx} - 1 dt \right) \left(\int_0^\infty t (e^{\theta \int_t^\infty 1-G(x) dx} - 1) dt \right).$$

Für $k = 1, \dots, 4$ existiert EC^k genau dann, wenn dies auch für das k -te Moment der Bedienzeitverteilung gilt.

Beweis. Nach den Lemmas 3.3.2 und 3.3.4 gilt

$$EC = -g'(0), \quad EC^2 = g''(0), \quad EC^3 = -g^{(3)}(0), \quad EC^4 = g^{(4)}(0).$$

Anwendung von Lemma 3.3.5 mit $c_1 = \frac{1}{\theta}$, $c_2 = e^{\theta\mu}$ und $f(s)$ wie in Lemma 3.3.6 auf g liefert jetzt die Ausdrücke für die Momente, die dabei benötigten Ableitungen von f erhält man aus dem zweitgenannten Lemma. Für die Integrale im Satz (d.h. $k = 0, 1, 2$) ergibt sich unter Verwendung partieller Integration

$$\int_0^\infty t^k (e^{\theta \int_t^\infty 1-G(x) dx} - 1) dt = \frac{\theta}{k+1} \int_0^\infty t^{k+1} (1-G(t)) e^{\theta \int_t^\infty 1-G(x) dx} dt.$$

Mit $\mu = \int_0^\infty 1-G(x) dx$ erhält man wegen

$$\int_0^\infty t^{k+1} (1-G(t)) dt \leq \int_0^\infty t^{k+1} (1-G(t)) e^{\theta \int_t^\infty 1-G(x) dx} dt \leq e^{\theta\mu} \int_0^\infty t^{k+1} (1-G(t)) dt$$

auch die zweite Behauptung. Dabei wurde die bekannte Tatsache verwendet, dass für eine nichtnegative Zufallsvariable Z mit Verteilungsfunktion G und für $n \in \mathbb{N}$ gilt $EZ^n = n \int_0^\infty x^{n-1} (1-G(x)) dx$. \square

Wir haben die Aussage zur Existenz der Momente im vorherigen Satz bewiesen, obwohl nach [Gha86] schon ein allgemeineres Resultat vorliegt. Dies geschieht aus einem Bestreben nach Vollständigkeit, da wir erwähnte Aussage explizit bei der unten betrachteten Pareto-Verteilung benötigen.

Mithilfe des folgenden Satzes erhält man sofort die asymptotische Varianz der Momentenschätzer $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$ für das erste und zweite Moment der Betriebsdauerverteilung. Sein Beweis ergibt sich direkt aus dem Beweis von Satz 3.2.10 und die benötigten Erwartungswerte haben wir im vorhergehenden Satz bestimmt.

Satz 3.3.8. *Es seien C_1, C_2, \dots i.i.d. Zufallsvariablen aus einer Verteilung mit den beiden Parametern κ_1 und κ_2 und endlichem i -ten Moment μ_i für $i = 1, \dots, 4$. Dann gilt*

$$\sqrt{m} \left(\begin{pmatrix} \bar{C}_{m,1} \\ \bar{C}_{m,2} \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} v_{\kappa_1, \kappa_2} & u_{\kappa_1, \kappa_2} \\ u_{\kappa_1, \kappa_2} & w_{\kappa_1, \kappa_2} \end{pmatrix} \right)$$

mit $u_{\kappa_1, \kappa_2} := \mu_3 - \mu_1 \mu_2$, $v_{\kappa_1, \kappa_2} := \mu_2 - \mu_1^2$ und $w_{\kappa_1, \kappa_2} := \mu_4 - \mu_2^2$.

3.3.2. Auf den Betriebsdauern basierende Schätzer

Hängt die Verteilung der Bedienzeit von genau zwei unbekanntem Parametern ab, erhält man durch den Ansatz

$$\bar{C}_{m,1} \stackrel{!}{=} EC_1, \quad \bar{C}_{m,2} \stackrel{!}{=} EC_1^2$$

zwei Gleichungen, in denen die unbekanntem Parameter vorkommen, und kann diese u.U. daraus bestimmen. Ferner ergibt sich durch Satz 3.3.8 und die Deltamethode unter bestimmten Voraussetzungen auch eine Aussage über die asymptotische Verteilung der so gewonnenen Schätzer. Für die Anwendung der Delta-Methode wird die Jacobi-Matrix der Funktion benötigt, die die Momente auf die Parameter abbildet; ein Blick auf die entsprechenden Formeln verdeutlicht allerdings, dass diese nicht ohne weiteres anzugeben ist. Wir verwenden deshalb den aus Grundvorlesungen in Analysis bekannten Satz von der Umkehrabbildung, um die Jacobi-Matrix zu bestimmen. Diese ergibt sich als Inverse der Jacobi-Matrix der Abbildung, die die Parameter auf die Momente abbildet, falls letztere nicht singular ist. Durch dieses Vorgehen erhält man noch eine weitere Eigenschaft der bestimmten Schätzer, nämlich starke Konsistenz. Das erste und zweite Stichprobenmoment konvergieren bekanntlich fast sicher gegen das erste bzw. zweite Moment der zugrunde liegenden Verteilung. Sind die partiellen Ableitungen in der Jacobi-Matrix stetig und ist diese wie oben benötigt invertierbar, so ist die Umkehrabbildung stetig differenzierbar, also insbesondere stetig. Damit wird die fast sichere Konvergenz der Stichprobenmomente durch die Abbildung auch auf die bestimmten Schätzer übertragen.

In Abhängigkeit von der Bedienzeitverteilung ist es mehr oder weniger aufwändig, den Nachweis der Regularität für die entsprechende Jacobi-Matrix zu führen. Um die breite Anwendbarkeit unseres Verfahrens zu demonstrieren bzw. Vergleichsmöglichkeiten für weiter unten entwickelte, alternative Schätzer bereitzustellen, betrachten wir im Folgenden mehrere verschiedene Verteilungen. Dabei wird nur bei der am Ende untersuchten Gammaverteilung ein ausführliches Beispiel für die Güte der hergeleiteten Resultate bei endlichem Stichprobenumfang präsentiert, um Redundanzen zu vermeiden. Im Folgenden bezeichne Σ die Kovarianzmatrix der Momentenschätzer $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$. Wir beschränken uns bei den Beweisen weiter unten darauf, die Invertierbarkeit der jeweiligen Jacobi-Matrix in einem Lemma zu zeigen und ansonsten auf die hier gemachten Ausführungen zu verweisen.

Die Pareto-Verteilung

Nach [JKB94, Kap. 20] ist die Benennung Pareto-Verteilung nicht eindeutig. Wir untersuchen hier die schon in Abschnitt 2.2.2 verwendete und in der Quelle als Pareto-

Verteilung der 2. Art bezeichnete Verteilung, wobei im Folgenden der Zusatz weggelassen wird. Die zugehörige Verteilungsfunktion ist für die Parameter $\alpha > 4$ und $\lambda > 0$ und für $x \geq 0$ gegeben durch $F(x) = 1 - \frac{\lambda^\alpha}{(\lambda+x)^\alpha}$ ([JKB94, S. 575]) und man berechnet schnell $\lambda/(\alpha-1)$ als Erwartungswert. Dabei wurde die Annahme an α getroffen, damit das vierte Moment der Betriebsdauerverteilung existiert. Wir schreiben im Folgenden kurz $X \sim \text{ParII}(\alpha, \lambda)$, wenn X diese Verteilung besitzt. Es sei nun

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\theta} (e^{\frac{\theta\lambda}{\alpha-1}} - 1) \\ \frac{2}{\theta} e^{\frac{\theta\lambda}{\alpha-1}} \int_0^\infty e^{\theta \int_t^\infty \lambda^\alpha / (\lambda+x)^\alpha dx} - 1 dt \end{pmatrix} =: \begin{pmatrix} F_{P1}(\alpha, \lambda) \\ F_{P2}(\alpha, \lambda) \end{pmatrix}.$$

Lemma 3.3.9. *Die Funktion*

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \mapsto \begin{pmatrix} F_{P1}(\alpha, \lambda) \\ F_{P2}(\alpha, \lambda) \end{pmatrix}$$

hat eine invertierbare Jacobi-Matrix für alle $\alpha > 2$ und $\lambda > 0$.

Beweis. Wir bestimmen die Determinante der entsprechenden Jacobi-Matrix. Unter Ausnutzung von $\int_t^\infty \frac{\lambda^\alpha}{(\lambda+x)^\alpha} dx = \frac{\lambda^\alpha (\lambda+t)^{-\alpha+1}}{\alpha-1}$ und mit

$$\begin{aligned} \frac{\partial}{\partial \alpha} F_{P1}(\alpha, \lambda) &= \frac{-\lambda e^{\theta\lambda/(\alpha-1)}}{(\alpha-1)^2}, & \frac{\partial}{\partial \lambda} F_{P1}(\alpha, \lambda) &= \frac{e^{\theta\lambda/(\alpha-1)}}{\alpha-1}, \\ \frac{\partial}{\partial \alpha} F_{P2}(\alpha, \lambda) &= \frac{2e^{\theta\lambda/(\alpha-1)}}{(\alpha-1)^2} \left(-\lambda \int_0^\infty e^{\theta\lambda^\alpha (\lambda+t)^{-\alpha+1}/(\alpha-1)} - 1 dt \right. \\ &\quad \left. + \int_0^\infty \frac{\lambda^\alpha ((\alpha-1) \log(\frac{\lambda}{\lambda+t}) - 1) e^{\theta\lambda^\alpha (\lambda+t)^{-\alpha+1}/(\alpha-1)}}{(\lambda+t)^{\alpha-1}} dt \right), \\ \frac{\partial}{\partial \lambda} F_{P2}(\alpha, \lambda) &= \frac{2e^{\theta\lambda/(\alpha-1)}}{\alpha-1} \left(\int_0^\infty e^{\theta\lambda^\alpha (\lambda+t)^{-\alpha+1}/(\alpha-1)} - 1 dt \right. \\ &\quad \left. + \int_0^\infty \frac{\lambda^{\alpha-1} (\alpha t + \lambda) e^{\theta\lambda^\alpha (\lambda+t)^{-\alpha+1}/(\alpha-1)}}{(\lambda+t)^\alpha} dt \right) \end{aligned}$$

ergibt sich

$$\begin{aligned} &\frac{\partial}{\partial \alpha} F_{P1}(\alpha, \lambda) \cdot \frac{\partial}{\partial \lambda} F_{P2}(\alpha, \lambda) - \frac{\partial}{\partial \lambda} F_{P1}(\alpha, \lambda) \cdot \frac{\partial}{\partial \alpha} F_{P2}(\alpha, \lambda) \\ &= \frac{-2\lambda^\alpha e^{2\theta\lambda/(\alpha-1)}}{(\alpha-1)^2} \int_0^\infty \frac{e^{\theta\lambda^\alpha (\lambda+t)^{-\alpha+1}/(\alpha-1)}}{(\lambda+t)^\alpha} (t + (\lambda+t) \log(\frac{\lambda}{\lambda+t})) dt. \end{aligned}$$

Der Bruch vor dem Integral ist offenbar immer negativ und der Bruch im Integral immer positiv. Definieren wir den verbleibenden geklammerten Term als $g(t)$, so gilt $g(0) = 0$ und $g'(t) = -\log(1 + \frac{t}{\lambda}) < 0$, also ist $g(t) < 0$ für alle $t > 0$. Damit ist die Determinante für alle zulässigen Parameterwerte positiv. \square

Zur besseren Übersichtlichkeit vereinbaren wir für die im Beweis des vorhergehenden Lemmas berechneten Größen die Schreibweisen

$$a_{\alpha,\lambda} := \frac{\partial}{\partial \alpha} F_{P1}(\alpha,\lambda), \quad b_{\alpha,\lambda} := \frac{\partial}{\partial \lambda} F_{P1}(\alpha,\lambda), \quad c_{\alpha,\lambda} := \frac{\partial}{\partial \alpha} F_{P2}(\alpha,\lambda), \quad d_{\alpha,\lambda} := \frac{\partial}{\partial \lambda} F_{P2}(\alpha,\lambda)$$

und bezeichnen die entsprechende Jacobi-Matrix mit J_P . Es ist uns nicht gelungen, die partiellen Ableitungen substanziell noch weiter als oben angegeben zu vereinfachen. Weiter seien $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$ die Momentenschätzer für die ersten beiden Momente der Betriebsdauerverteilung, wenn die Bedienzeit Pareto-verteilt mit den Parametern α und λ ist und die Längen von m Betriebsphasen vorliegen. Dann sind implizit zwei Momentenschätzer für α und λ gegeben, diese bezeichnen wir mit $\hat{\alpha}_m$ und $\hat{\lambda}_m$. Der Beweis des folgenden Satzes ergibt sich nun direkt aus den obigen Ausführungen, dabei entsteht die Matrix Σ_P durch das Matrizenprodukt $J_P^{-1} \Sigma (J_P^{-1})'$.

Satz 3.3.10. *Sind die Bedienzeiten $\text{ParII}(\alpha,\lambda)$ -verteilt, dann gilt*

$$\sqrt{m} \left(\begin{pmatrix} \hat{\alpha}_m \\ \hat{\lambda}_m \end{pmatrix} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_2(0, \Sigma_P)$$

mit $\Sigma_P = (\sigma_{ij})_{i,j=1,2}$, $\sigma_{12} = \sigma_{21}$ und

$$\sigma_{11} = \frac{d_{\alpha,\lambda}^2 v_{\alpha,\lambda} - 2b_{\alpha,\lambda} d_{\alpha,\lambda} u_{\alpha,\lambda} + b_{\alpha,\lambda}^2 w_{\alpha,\lambda}}{(a_{\alpha,\lambda} d_{\alpha,\lambda} - b_{\alpha,\lambda} c_{\alpha,\lambda})^2}, \quad \sigma_{22} = \frac{c_{\alpha,\lambda}^2 v_{\alpha,\lambda} - 2a_{\alpha,\lambda} c_{\alpha,\lambda} u_{\alpha,\lambda} + a_{\alpha,\lambda}^2 w_{\alpha,\lambda}}{(a_{\alpha,\lambda} d_{\alpha,\lambda} - b_{\alpha,\lambda} c_{\alpha,\lambda})^2},$$

$$\sigma_{12} = \frac{(b_{\alpha,\lambda} c_{\alpha,\lambda} + a_{\alpha,\lambda} d_{\alpha,\lambda}) u_{\alpha,\lambda} - c_{\alpha,\lambda} d_{\alpha,\lambda} v_{\alpha,\lambda} - a_{\alpha,\lambda} b_{\alpha,\lambda} w_{\alpha,\lambda}}{(a_{\alpha,\lambda} d_{\alpha,\lambda} - b_{\alpha,\lambda} c_{\alpha,\lambda})^2}$$

sowie $u_{\alpha,\lambda}, v_{\alpha,\lambda}, w_{\alpha,\lambda}$ aus Satz 3.3.8. Weiter sind die Schätzer $\hat{\alpha}_m$ und $\hat{\lambda}_m$ stark konsistent.

Beispiel 3.3.11. Bei $\theta = 0.5$, $\alpha = 5$ und $\lambda = 5$ gilt für die asymptotischen Varianzen $\text{var}(\hat{\alpha}) = 1801.284$ und $\text{var}(\hat{\lambda}) = 2731.233$. Dies sind erhebliche größere Werte als bei allen nachfolgend betrachteten Beispielen, wo Verteilungen mit ähnlichem Erwartungswert betrachtet werden, und lässt sich möglicherweise durch die heavy-tail-Eigenschaft der Pareto-Verteilung erklären. Um die Annäherung an diese Werte bei endlichem m zu prüfen, haben wir in 1000 Simulationsdurchläufen die Länge von jeweils $m = 8000$ Betriebsdauern bestimmt, dabei ergab sich $m \cdot \text{var}(\hat{\alpha}_m) = 1298.151$ und $m \cdot \text{var}(\hat{\lambda}_m) = 1957.285$. Offensichtlich treten große Abweichungen von den asymptotischen Werten auf, diese sind bei allen nachfolgend betrachteten Verteilungen deutlich geringer. Da die asymptotische Varianz unterschritten wird, ist davon auszugehen, dass 1000 Wiederholungen zu wenig waren, um die bei der Pareto-Verteilung wegen oben genannter heavy-tail-Eigenschaft gelegentlich vorkommenden extrem großen Werte repräsentativ auftreten zu lassen. \triangleleft

Bemerkung 3.3.12. Auch für die bei [JKB94, S. 574] als Pareto-Verteilung der 1. Art bezeichnete Verteilung mit Dichte $f(x) = \alpha k x^{-\alpha-1}$ für $x \geq k, k, \alpha > 0$ ließe sich ein ähnliches Resultat wie im obigen Satz herleiten. Da diese Verteilung allerdings nur Werte auf einer echten Teilmenge von \mathbb{R}^+ annimmt, verfolgen wir dies hier nicht weiter. \triangleleft

Die Lognormal-Verteilung

Der Logarithmus einer lognormalverteilten Zufallsvariable X mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ ist normalverteilt mit denselben Parametern, dementsprechend ergibt sich mit Φ als der Verteilungsfunktion der Standardnormalverteilung für X die Verteilungsfunktion $F_X(x) = \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)$. Weiter gilt $EX = \exp(\mu + \sigma^2/2)$ [JKB94, S. 212] und wir schreiben kurz $X \sim \text{LogN}(\mu, \sigma^2)$, wenn X mit den Parametern μ und σ lognormalverteilt ist. Es sei nun

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\theta} (e^{\theta \exp(\mu + \sigma^2/2)} - 1) \\ \frac{2}{\theta} e^{\theta \exp(\mu + \sigma^2/2)} \int_0^\infty e^{\theta \int_t^\infty 1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)} dx - 1 dt \end{pmatrix} =: \begin{pmatrix} F_{L1}(\mu, \sigma) \\ F_{L2}(\mu, \sigma) \end{pmatrix}.$$

Lemma 3.3.13. *Die Funktion*

$$\begin{pmatrix} \mu \\ \sigma \end{pmatrix} \mapsto \begin{pmatrix} F_{L1}(\mu, \sigma) \\ F_{L2}(\mu, \sigma) \end{pmatrix}$$

hat eine invertierbare Jacobi-Matrix für alle $\mu \in \mathbb{R}$ und $\sigma > 0$.

Beweis. Wir bestimmen die Determinante der entsprechenden Jacobi-Matrix. Mit

$$\begin{aligned} \frac{\partial}{\partial \mu} F_{L1}(\mu, \sigma) &= e^{\theta \exp(\mu + \sigma^2/2) + \mu + \sigma^2/2}, & \frac{\partial}{\partial \sigma} F_{L1}(\mu, \sigma) &= \sigma e^{\theta \exp(\mu + \sigma^2/2) + \mu + \sigma^2/2}, \\ \frac{\partial}{\partial \mu} F_{L2}(\mu, \sigma) &= 2e^{\theta \exp(\mu + \sigma^2/2)} \left(e^{\mu + \sigma^2/2} \int_0^\infty e^{\theta \int_t^\infty 1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)} dx - 1 dt \right. \\ &\quad \left. + \int_0^\infty e^{\theta \int_t^\infty 1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)} dx \int_t^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right) \cdot \frac{1}{\sigma} dx dt \right), \\ \frac{\partial}{\partial \sigma} F_{L2}(\mu, \sigma) &= 2e^{\theta \exp(\mu + \sigma^2/2)} \left(\sigma e^{\mu + \sigma^2/2} \int_0^\infty e^{\theta \int_t^\infty 1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)} dx - 1 dt \right. \\ &\quad \left. + \int_0^\infty e^{\theta \int_t^\infty 1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)} dx \cdot \int_t^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right) \cdot \frac{\log(x)-\mu}{\sigma^2} dx dt \right) \end{aligned}$$

ergibt sich

$$\begin{aligned} &\frac{\partial}{\partial \mu} F_{L1}(\mu, \sigma) \cdot \frac{\partial}{\partial \sigma} F_{L2}(\mu, \sigma) - \frac{\partial}{\partial \sigma} F_{L1}(\mu, \sigma) \cdot \frac{\partial}{\partial \mu} F_{L2}(\mu, \sigma) \\ &= 2e^{2\theta \exp(\mu + \sigma^2/2) + \mu + \sigma^2/2} \int_0^\infty e^{\theta \int_t^\infty 1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)} dx \\ &\quad \cdot \int_t^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right) \cdot \left(\frac{\log(x)-\mu}{\sigma^2} - 1\right) dx dt \\ &= 2e^{2\theta \exp(\mu + \sigma^2/2) + \mu + \sigma^2/2} \int_0^\infty \frac{t}{\sqrt{2\pi}} e^{\theta \int_t^\infty 1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)} dx - \frac{1}{2} \left(\frac{\log(t)-\mu}{\sigma}\right)^2 dt, \end{aligned}$$

dabei wurde benutzt, dass $\frac{-x}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$ eine Stammfunktion für das Integral in der vorletzten Zeile ist. Der Integrand in der letzten Zeile hat in 0 den Grenzwert 0 und ist ansonsten immer positiv, also ist auch der gesamte Ausdruck positiv. \square

Analog zu unserem Vorgehen bei Pareto-Verteilungen vereinbaren wir für die im Beweis des vorhergehenden Lemmas berechneten Größen die Schreibweisen

$$a_{\mu,\sigma} := \frac{\partial}{\partial \mu} F_{L1}(\mu,\sigma), \quad b_{\mu,\sigma} := \frac{\partial}{\partial \sigma} F_{L1}(\mu,\sigma), \quad c_{\mu,\sigma} := \frac{\partial}{\partial \mu} F_{L2}(\mu,\sigma), \quad d_{\mu,\sigma} := \frac{\partial}{\partial \sigma} F_{L2}(\mu,\sigma)$$

und bezeichnen die entsprechende Jacobi-Matrix mit J_{LN} . Weiter seien $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$ die Momentenschätzer für die ersten beiden Momente der Betriebsdauerverteilung, wenn die Bedienzeit lognormalverteilt mit den Parametern μ und σ ist und die Längen von m Betriebsphasen vorliegen. Dann sind implizit zwei Momentenschätzer für μ und σ gegeben, diese bezeichnen wir mit $\hat{\mu}_m$ und $\hat{\sigma}_m$. Der Beweis des folgenden Satzes ergibt sich nun direkt aus den obigen Ausführungen, dabei entsteht die Matrix Σ_{LN} durch das Matrizenprodukt $J_{LN}^{-1} \Sigma (J_{LN}^{-1})'$.

Satz 3.3.14. *Sind die Bedienzeiten $\text{LogN}(\mu,\sigma)$ -verteilt, dann gilt*

$$\sqrt{m} \left(\begin{pmatrix} \hat{\mu}_m \\ \hat{\sigma}_m \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_2(0, \Sigma_{LN})$$

mit $\Sigma_{LN} = (\sigma_{ij})_{i,j=1,2}$, $\sigma_{12} = \sigma_{21}$ und

$$\sigma_{11} = \frac{d_{\mu,\sigma}^2 v_{\mu,\sigma} - 2b_{\mu,\sigma} d_{\mu,\sigma} u_{\mu,\sigma} + b_{\mu,\sigma}^2 w_{\mu,\sigma}}{(a_{\mu,\sigma} d_{\mu,\sigma} - b_{\mu,\sigma} c_{\mu,\sigma})^2}, \quad \sigma_{22} = \frac{c_{\mu,\sigma}^2 v_{\mu,\sigma} - 2a_{\mu,\sigma} c_{\mu,\sigma} u_{\mu,\sigma} + a_{\mu,\sigma}^2 w_{\mu,\sigma}}{(a_{\mu,\sigma} d_{\mu,\sigma} - b_{\mu,\sigma} c_{\mu,\sigma})^2},$$

$$\sigma_{12} = \frac{(b_{\mu,\sigma} c_{\mu,\sigma} + a_{\mu,\sigma} d_{\mu,\sigma}) u_{\mu,\sigma} - c_{\mu,\sigma} d_{\mu,\sigma} v_{\mu,\sigma} - a_{\mu,\sigma} b_{\mu,\sigma} w_{\mu,\sigma}}{(a_{\mu,\sigma} d_{\mu,\sigma} - b_{\mu,\sigma} c_{\mu,\sigma})^2}$$

sowie $u_{\mu,\sigma}, v_{\mu,\sigma}, w_{\mu,\sigma}$ aus Satz 3.3.8. Weiter sind die Schätzer $\hat{\mu}_m$ und $\hat{\sigma}_m$ stark konsistent.

Beispiel 3.3.15. Bei $\theta = 0.5$, $\mu = 0$ und $\sigma = 1$ gilt für die asymptotischen Varianzen $\text{var}(\hat{\mu}) = 3.711780$ und $\text{var}(\hat{\sigma}) = 3.854515$. Um die Annäherung an diese Werte bei endlichem m zu prüfen, haben wir in 1000 Simulationsdurchläufen die Länge von jeweils $m = 8000$ Betriebsdauern bestimmt, dabei ergab sich $m \cdot \text{var}(\hat{\mu}_m) = 3.385111$ und $m \cdot \text{var}(\hat{\sigma}_m) = 3.295206$. Die simulativ ermittelten Werte liegen in der Nähe der Grenzwerte, allerdings ist die Abweichung deutlich größer als bei den weiter unten untersuchten Verteilungen, nur bei der im vorherigen Abschnitt betrachteten Pareto-Verteilung ist sie noch stärker. ◁

Die Weibull-Verteilung

Die Weibull-Verteilung hat für $x \geq 0$ die Verteilungsfunktion $F(x) = 1 - e^{-(\frac{x}{\lambda})^k}$ und den Erwartungswert $\lambda \Gamma(1 + \frac{1}{k})$ ([JKB94, S. 629 bzw. S. 632], λ ist lt. der dortigen S. 631 ein Skalenparameter). Wir definieren nun

$$\begin{pmatrix} \lambda \\ k \end{pmatrix} \mapsto \left(\frac{1}{\theta} (e^{\theta \lambda \Gamma(1 + \frac{1}{k})} - 1) \int_0^\infty e^{\theta \lambda \Gamma(1 + \frac{1}{k})} \int_0^\infty e^{\theta \int_t^\infty \exp(-(\frac{x}{\lambda})^k) dx} - 1 dt \right) =: \begin{pmatrix} F_{W1}(\lambda, k) \\ F_{W2}(\lambda, k) \end{pmatrix}.$$

Um zu zeigen, dass diese Funktion umkehrbar ist, benötigen wir noch ein Hilfsresultat.

Lemma 3.3.16. *Für festes $k > 0$ und alle $u \geq 0$ gilt*

$$\Psi(1 + 1/k) - \frac{\int_u^\infty e^{-y} y^{\frac{1}{k}} \log(y) dy}{\int_u^\infty e^{-y} y^{\frac{1}{k}} dy} \leq 0,$$

dabei ist Ψ die Digamma-Funktion. Ferner gibt es ein $u_0 > 0$, so dass die Ungleichung für alle $u \geq u_0$ strikt ist.

Beweis. Offenbar ist der Bruch für $u = 0$ gleich $\Psi(1 + 1/k)$, der gesamte Ausdruck somit also 0. Sei nun $u > 0$, dann gilt

$$\begin{aligned} & \frac{\partial}{\partial u} \left(\Psi(1 + 1/k) - \frac{\int_u^\infty e^{-y} y^{\frac{1}{k}} \log(y) dy}{\int_u^\infty e^{-y} y^{\frac{1}{k}} dy} \right) \\ &= - \frac{-e^{-u} u^{\frac{1}{k}} \log(u) \cdot \int_u^\infty e^{-y} y^{\frac{1}{k}} dy + \int_u^\infty e^{-y} y^{\frac{1}{k}} \log(y) dy \cdot e^{-u} u^{\frac{1}{k}}}{\left(\int_u^\infty e^{-y} y^{\frac{1}{k}} dy \right)^2}. \end{aligned}$$

Unter Verwendung von $\int_u^\infty e^{-y} y^{\frac{1}{k}} \log(y) dy \geq \log(u) \int_u^\infty e^{-y} y^{\frac{1}{k}} dy$ ergibt sich, dass der Zähler des obigen Bruchs nichtnegativ, die untersuchte Funktion wegen des negativen Vorzeichens also monoton fallend ist. Für den Beweis der zweiten Behauptung nehmen wir an, dass der untersuchte Ausdruck immer konstant 0 ist. Wegen

$$\begin{aligned} \Psi(1 + 1/k) \cdot \int_u^\infty e^{-y} y^{\frac{1}{k}} dy &= \int_u^\infty e^{-y} y^{\frac{1}{k}} \log(y) dy \\ \Rightarrow \Psi(1 + 1/k) \cdot e^{-u} u^{\frac{1}{k}} &= e^{-u} u^{\frac{1}{k}} \log(u) \Leftrightarrow \Psi(1 + 1/k) = \log(u), \end{aligned}$$

jeweils für alle $u > 0$, führt dies auf einen Widerspruch, zusammen mit der Monotonie ergibt sich nun auch die zweite Behauptung. \square

Lemma 3.3.17. *Die Funktion*

$$\begin{pmatrix} \lambda \\ k \end{pmatrix} \mapsto \begin{pmatrix} F_{W_1}(\lambda, k) \\ F_{W_2}(\lambda, k) \end{pmatrix}$$

hat eine invertierbare Jacobi-Matrix für alle $k > 0$ und $\lambda > 0$.

Beweis. Wir bestimmen die Determinante der entsprechenden Jacobi-Matrix. Mit

$$\begin{aligned} \frac{\partial}{\partial \lambda} F_{W_1}(\lambda, k) &= e^{\theta \lambda \Gamma(1 + \frac{1}{k})} \Gamma(1 + \frac{1}{k}), & \frac{\partial}{\partial k} F_{W_1}(\lambda, k) &= \frac{-1}{k^2} e^{\theta \lambda \Gamma(1 + \frac{1}{k})} \lambda \Gamma'(1 + \frac{1}{k}), \\ \frac{\partial}{\partial \lambda} F_{W_2}(\lambda, k) &= 2e^{\theta \lambda \Gamma(1 + \frac{1}{k})} \cdot \left(\Gamma(1 + \frac{1}{k}) \cdot \int_0^\infty e^{\theta \int_t^\infty \exp(-(\frac{x}{\lambda})^k) dx} - 1 dt \right) \end{aligned}$$

$$\begin{aligned}
& + \int_0^\infty e^{\theta \int_t^\infty \exp(-(\frac{x}{\lambda})^k) dx} \int_t^\infty e^{-(\frac{x}{\lambda})^k} \left(\frac{x}{\lambda}\right)^k \cdot \frac{k}{\lambda} dx dt, \\
\frac{\partial}{\partial k} F_{W_2}(\lambda, k) & = -2e^{\theta \lambda \Gamma(1+\frac{1}{k})} \cdot \left(\lambda \Gamma'(1 + \frac{1}{k}) \cdot \frac{1}{k^2} \cdot \int_0^\infty e^{\theta \int_t^\infty \exp(-(\frac{x}{\lambda})^k) dx} - 1 dt \right. \\
& \left. + \int_0^\infty e^{\theta \int_t^\infty \exp(-(\frac{x}{\lambda})^k) dx} \int_t^\infty e^{-(\frac{x}{\lambda})^k} \left(\frac{x}{\lambda}\right)^k \log\left(\frac{x}{\lambda}\right) dx dt \right)
\end{aligned}$$

ergibt sich wegen $\Gamma'(1 + \frac{1}{k}) = \Gamma(1 + \frac{1}{k}) \cdot \Psi(1 + \frac{1}{k})$

$$\begin{aligned}
& \frac{\partial}{\partial \lambda} F_{W_1}(\lambda, k) \cdot \frac{\partial}{\partial k} F_{W_2}(\lambda, k) - \frac{\partial}{\partial k} F_{W_1}(\lambda, k) \cdot \frac{\partial}{\partial \lambda} F_{W_2}(\lambda, k) \\
& = 2e^{2\theta \lambda \Gamma(1+\frac{1}{k})} \Gamma(1 + \frac{1}{k}) \\
& \cdot \int_0^\infty e^{\theta \int_t^\infty \exp(-(\frac{x}{\lambda})^k) dx} \cdot \left(\frac{\Psi(1 + \frac{1}{k})}{k} \int_t^\infty e^{-(\frac{x}{\lambda})^k} \left(\frac{x}{\lambda}\right)^k dx - \int_t^\infty e^{-(\frac{x}{\lambda})^k} \left(\frac{x}{\lambda}\right)^k \log\left(\frac{x}{\lambda}\right) dx \right) dt.
\end{aligned}$$

Durch die Substitution $y := (\frac{x}{\lambda})^k$ folgt nun für den geklammerten Term innerhalb des Integrals

$$\begin{aligned}
& \frac{\Psi(1 + \frac{1}{k})}{k} \int_t^\infty e^{-(\frac{x}{\lambda})^k} \left(\frac{x}{\lambda}\right)^k dx - \int_t^\infty e^{-(\frac{x}{\lambda})^k} \left(\frac{x}{\lambda}\right)^k \log\left(\frac{x}{\lambda}\right) dx \\
& = \frac{\lambda}{k^2} \left(\Psi(1 + \frac{1}{k}) \int_{(t/\lambda)^k}^\infty e^{-y} y^{1/k} dy - \int_{(t/\lambda)^k}^\infty e^{-y} y^{1/k} \log(y) dy \right) \\
& = \frac{\lambda}{k^2} \int_{(t/\lambda)^k}^\infty e^{-y} y^{1/k} dy \left(\Psi(1 + \frac{1}{k}) - \frac{\int_{(t/\lambda)^k}^\infty e^{-y} y^{1/k} \log(y) dy}{\int_{(t/\lambda)^k}^\infty e^{-y} y^{1/k} dy} \right).
\end{aligned}$$

Wir zeigen nun, dass der geklammerte Term immer nichtpositiv und für bestimmte Werte sogar echt negativ ist, dann ist das obige Integral und damit auch die gesamte Funktionaldeterminante ungleich 0 und es folgt die Behauptung. Definieren wir die Abbildung aus Lemma 3.3.16 als $f(u)$ und $g(t) := (t/\lambda)^k$, dann lässt sich der geklammerte Term als $f(g(t))$ schreiben und die Behauptung folgt mit Lemma 3.3.16, da $g(t) \geq 0$ für $t \geq 0$ wegen $k > 0$ und $\lambda > 0$ ist. \square

Wir vereinbaren wieder für die im Beweis des vorhergehenden Lemmas berechneten Größen die Schreibweisen

$$a_{\lambda,k} := \frac{\partial}{\partial \lambda} F_{W_1}(\lambda, k), \quad b_{\lambda,k} := \frac{\partial}{\partial k} F_{W_1}(\lambda, k), \quad c_{\lambda,k} := \frac{\partial}{\partial \lambda} F_{W_2}(\lambda, k), \quad d_{\lambda,k} := \frac{\partial}{\partial k} F_{W_2}(\lambda, k),$$

und bezeichnen die entsprechende Jacobi-Matrix mit J_W . Weiter seien $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$ die Momentenschätzer für die ersten beiden Momente der Betriebsdauerverteilung, wenn die Bedienzeit Weibull-verteilt mit den Parametern λ und k ist und die Längen von m Betriebsphasen vorliegen. Dann sind implizit zwei Momentenschätzer für λ und k

gegeben, diese bezeichnen wir mit $\hat{\lambda}_m$ und \hat{k}_m . Der Beweis des folgenden Satzes ergibt sich nun direkt aus den obigen Ausführungen, dabei entsteht die Matrix Σ_W durch das Matrizenprodukt $J_W^{-1}\Sigma(J_W^{-1})'$.

Satz 3.3.18. *Sind die Bedienzeiten Weibull-verteilt mit den Parametern λ und k , dann gilt*

$$\sqrt{m} \left(\begin{pmatrix} \hat{\lambda}_m \\ \hat{k}_m \end{pmatrix} - \begin{pmatrix} \lambda \\ k \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_2(0, \Sigma_W)$$

mit $\Sigma_W = (\sigma_{ij})_{i,j=1,2}$, $\sigma_{12} = \sigma_{21}$ und

$$\sigma_{11} = \frac{d_{\lambda,k}^2 v_{\lambda,k} - 2b_{\lambda,k} d_{\lambda,k} u_{\lambda,k} + b_{\lambda,k}^2 w_{\lambda,k}}{(a_{\lambda,k} d_{\lambda,k} - b_{\lambda,k} c_{\lambda,k})^2}, \quad \sigma_{22} = \frac{c_{\lambda,k}^2 v_{\lambda,k} - 2a_{\lambda,k} c_{\lambda,k} u_{\lambda,k} + a_{\lambda,k}^2 w_{\lambda,k}}{(a_{\lambda,k} d_{\lambda,k} - b_{\lambda,k} c_{\lambda,k})^2},$$

$$\sigma_{12} = \frac{(c_{\lambda,k} b_{\lambda,k} + a_{\lambda,k} d_{\lambda,k}) u_{\lambda,k} - c_{\lambda,k} d_{\lambda,k} v_{\lambda,k} - a_{\lambda,k} b_{\lambda,k} w_{\lambda,k}}{(a_{\lambda,k} d_{\lambda,k} - b_{\lambda,k} c_{\lambda,k})^2}$$

sowie $u_{\lambda,k}, v_{\lambda,k}, w_{\lambda,k}$ aus Satz 3.3.8. Weiter sind die Schätzer $\hat{\lambda}_m$ und \hat{k}_m stark konsistent.

Beispiel 3.3.19. Bei $\theta = 0.5$, $k = 2$ und $\lambda = 2$ gilt für die asymptotischen Varianzen $\text{var}(\hat{k}) = 17.8198607$ und $\text{var}(\hat{\lambda}) = 1.029155$. Um die Annäherung an diese Werte bei endlichem m zu prüfen, haben wir in 1000 Simulationsdurchläufen die Länge von jeweils $m = 8000$ Betriebsdauern bestimmt, dabei ergab sich $m \cdot \text{var}(\hat{k}_m) = 17.55501$ und $m \cdot \text{var}(\hat{\lambda}_m) = 1.076982$. Die beobachteten Werte liegen nahe bei den Grenzwerten, bei \hat{k}_m noch etwas deutlicher als bei $\hat{\lambda}_m$. \triangleleft

Die inverse Gauß-Verteilung I

Anhand der als nächstes betrachteten Verteilung wird deutlich, dass wir die Invertierbarkeit der Jacobi-Matrix nicht immer mit dem bisherigen Vorgehen zeigen können. Bekanntlich besitzt die inverse Gauß-Verteilung für Parameter $\mu > 0, \lambda > 0$ die Verteilungsfunktion

$$F_{\mu,\lambda}(x) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right)$$

für $x > 0$ und den Erwartungswert μ ([CF89, S. 15 bzw. S. 8]). Definieren wir

$$\begin{pmatrix} \mu \\ \lambda \end{pmatrix} \mapsto \left(\frac{1}{\theta} e^{\theta\mu} \int_0^\infty e^{\theta \int_t^\infty 1 - F_{\mu,\lambda}(x) dx} - 1 dt \right) =: \begin{pmatrix} F_{IG1}(\mu, \lambda) \\ F_{IG2}(\mu, \lambda) \end{pmatrix},$$

so ist $\frac{\partial}{\partial \lambda} F_{IG1}(\mu, \lambda) = 0$. Dies bringt die Vereinfachung mit sich, dass bei einem Vorgehen wie bei den bisher untersuchten Verteilungen wegen $\frac{\partial}{\partial \mu} F_{IG1}(\mu, \lambda) = e^{\theta\mu} > 0$ nur zu prüfen ist, ob $\frac{\partial}{\partial \lambda} F_{IG2}(\mu, \lambda)$ für alle Werte von μ und λ nicht verschwindet. Im Beweis zu

Lemma 3.3.17 trat an dieser Stelle ein Ausdruck der Form $\int_0^\infty \exp(f(t)) \cdot \int_t^\infty g(x) dx dt$ auf und wir konnten zeigen, dass $g(x) \geq 0$ für alle $x \geq 0$ gilt. Geht man hier analog vor, führt dies auf die Untersuchung von

$$\begin{aligned} \frac{\partial}{\partial \lambda}(1 - F_{\mu,\lambda}(x)) &= -\Phi'\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) \cdot \left(\frac{x}{\mu} - 1\right) \frac{1}{2\sqrt{\lambda x}} - \exp\left(\frac{2\lambda}{\mu}\right) \frac{2}{\mu} \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right) \\ &\quad + \exp\left(\frac{2\lambda}{\mu}\right) \Phi'\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right) \left(\frac{x}{\mu} + 1\right) \frac{1}{2\sqrt{\lambda x}} \end{aligned}$$

für alle $x > 0$ anstelle des obigen $g(x)$ (als Grenzwert in 0 ergibt sich der Wert 0). Wie sich weiter unten zeigt, ließe sich dies noch erheblich vereinfachen. An dieser Stelle ist es allerdings ausreichend darauf hinzuweisen, dass der Ausdruck bei $\lambda = 1 = \mu$ für $x = 1$ einen positiven und für $x = 2$ einen negativen Wert annimmt. Dementsprechend kann hier nicht wie bei der Weibull-Verteilung argumentiert werden. Auch ein Vorgehen wie bei der Lognormal-Verteilung konnten wir wegen des auch nach allen uns möglichen Vereinfachungen komplizierteren Integranden nicht realisieren. Wir benötigen hier also einen anderen Ansatz und beginnen mit einigen allgemeinen Vorüberlegungen.

Allgemeine Überlegungen

Bei der inversen Gauß-Verteilung entspricht der Parameter μ dem Erwartungswert. Deshalb ist es bei der Überprüfung, ob die Funktionaldeterminante der interessierenden Abbildung nicht verschwindet, nach obigen Ausführungen ausreichend, nur die partielle Ableitung des Ausdrucks für das zweite Moment nach dem anderen Parameter zu betrachten. An dieser Stelle setzen wir an und befassen uns in diesem Abschnitt allgemein mit zweiparametrischen Verteilungen, die diese Anforderung erfüllen. Dabei nehmen wir an, dass die im Folgenden durchgeführten Vertauschungen von Integration und Differentiation erlaubt sind, wie es z.B. bei der inversen Gauß-Verteilung der Fall ist. Es bezeichne χ den Parameter, der gleichzeitig der Erwartungswert ist, η den zweiten Parameter und $F_{\chi,\eta}$ die von den Parametern abhängende Verteilungsfunktion. Wir untersuchen also die Jacobi-Matrix der Abbildung

$$\begin{pmatrix} F_1(\chi,\eta) \\ F_2(\chi,\eta) \end{pmatrix} := \begin{pmatrix} \theta^{-1}(\exp(\theta\chi) - 1) \\ 2\theta^{-1} \exp(\theta\chi) \int_0^\infty (\exp(\theta \int_t^\infty 1 - F_{\chi,\eta}(x) dx) - 1 dt) \end{pmatrix}.$$

Diese hat die Gestalt

$$\begin{pmatrix} \exp(\theta\chi) & 0 \\ * & \frac{\partial F_2(\chi,\eta)}{\partial \eta} \end{pmatrix}$$

und wir betrachten den Eintrag rechts unten näher. Es gilt

$$\frac{\partial F_2(\chi,\eta)}{\partial \eta} = \frac{\partial}{\partial \eta} \left(2\theta^{-1} \exp(\chi\theta) \int_0^\infty \left(\exp\left(\theta \int_t^\infty 1 - F_{\chi,\eta}(x) dx\right) - 1 dt \right) \right)$$

$$= 2\theta^{-1} \exp(\chi\theta) \int_0^\infty \exp\left(\theta \int_t^\infty 1 - F_{\chi,\eta}(x) dx\right) \cdot \theta \cdot \frac{\partial}{\partial \eta} \int_t^\infty 1 - F_{\chi,\eta}(x) dx dt.$$

Da die Exponentialausdrücke und θ positiv sind, konzentrieren wir uns auf die partielle Ableitung $\frac{\partial}{\partial \eta} \int_t^\infty 1 - F_{\chi,\eta}(x) dx$. Es sei \aleph die Menge der $t \geq 0$, für die diese nicht verschwindet. Kann für alle $\eta, \chi > 0$ gezeigt werden, dass jeweils \aleph keine Nullmenge ist und die partielle Ableitung auf ganz \aleph positiv oder auf ganz \aleph negativ ist, so folgt daraus, dass die Funktionaldeterminante nicht verschwindet. Ganz allgemein gilt für eine nichtnegative Zufallsvariable X mit Dichte f und Verteilungsfunktion F wegen $1 - F(x) = P(X \geq x)$ für $t \geq 0$

$$\begin{aligned} \int_t^\infty P(X \geq x) dx &= \int_t^\infty \int_x^\infty f(y) dy dx = \iint_{\{(x,y) \in \mathbb{R}^2 : x \geq t, y \geq x\}} f(y) dy dx \\ &= \int_t^\infty f(y) \int_t^y dx = \int_t^\infty (y - t) f(y) dy \\ &= \int_0^\infty (y - t)^+ dF(y) = E(h_1(X)) \text{ mit } h_1(x) := (x - t)^+. \end{aligned} \quad (3.14)$$

Dieser letzte Ausdruck ist als Stop-Loss-Transformierte (stop-loss transform) bzw. Stop-Loss-Prämie (stop-loss premium) in der Versicherungsmathematik und als integrierte Überlebensfunktion (integrated survival function) in der Zuverlässigkeitstheorie bekannt (siehe z.B. die Ausführungen bei [MS02, S. 19 und S. 278ff.]). Die obige Abbildung ist also umkehrbar, wenn die Ableitung des Erwartungswertes (3.14) nach η für alle Werte von χ und η sowie für $t \in \aleph$ entweder immer negativ oder immer positiv ist. Bei der inversen Gauß-Verteilung konnten wir dies durch direktes Ausrechnen der Ableitung von $1 - F_{\mu,\lambda}$ nicht sicherstellen. Da h_1 konvex ist, in (3.14) also der Erwartungswert einer konvexen Funktion einer Zufallsvariable betrachtet wird, verfolgen wir als nächstes einen Ansatz über konvexe stochastische Ordnungen.

Definition 3.3.20. Eine Zufallsvariable X heißt *kleiner als Y in der konvexen stochastischen Ordnung*, in Zeichen $X \leq_{cx} Y$, genau dann, wenn gilt

$$E(h(X)) \leq E(h(Y)) \text{ für alle konvexen Funktionen } h.$$

Stochastische Ordnungen werden z.B. eingesetzt, um Risiken zu ordnen. Für andere Anwendungen und Informationen siehe z.B. [MS02], insbesondere dort S. 16ff.

Die inverse Gauß-Verteilung II

Für die Stop-Loss-Transformierte einer inversen Gauß-Verteilung gilt laut [Tad81] mit der Notation von oben

$$E(X - t)^+ = (\mu - t)\Phi\left((1 - t/\mu)\sqrt{\lambda/t}\right) + (\mu + t)e^{2\lambda/\mu}\Phi\left(- (1 + t/\mu)\sqrt{\lambda/t}\right). \quad (3.15)$$

Dieses Resultat benutzen wir für den Beweis des folgenden Lemmas.

Lemma 3.3.21. *Die oben definierte Abbildung*

$$\begin{pmatrix} \mu \\ \lambda \end{pmatrix} \mapsto \begin{pmatrix} F_{\text{IG1}}(\mu, \lambda) \\ F_{\text{IG2}}(\mu, \lambda) \end{pmatrix}$$

hat eine invertierbare Jacobi-Matrix für alle $\mu > 0$ und $\lambda > 0$.

Beweis. Es sei $X \sim \text{IG}(\mu, \lambda)$. Nach den obigen Ausführungen ist nur noch zu überprüfen, ob $\frac{\partial}{\partial \lambda} F_{\text{IG2}}(\mu, \lambda)$ für alle Werte von μ und λ nicht verschwindet. Dies ist der Fall, wenn

$$\frac{\partial}{\partial \lambda} \int_0^\infty (y-t)^+ dF_{\mu, \lambda}(y) = \frac{\partial}{\partial \lambda} E_{\mu, \lambda}(X-t)^+$$

für alle $t \in \mathbb{R}^+$ dasselbe Vorzeichen hat und auf einer Nicht-Nullmenge von 0 verschieden ist. Mit (3.15) gilt

$$\begin{aligned} \frac{\partial}{\partial \lambda} E_{\mu, \lambda}(X-t)^+ &= \frac{2(\mu+t)}{\mu} e^{2\lambda/\mu} (1 - \Phi((1+t/\mu)\sqrt{\lambda/t}) - 2e^{-\lambda(1-t/\mu)^2/(2t)} \sqrt{t/(2\pi\lambda)}) \\ &< \frac{2}{\sqrt{2\pi}} e^{-\lambda(1-t/\mu)^2/(2t)} \sqrt{t/\lambda} \left(\frac{\mu+t}{\mu} \cdot \frac{\mu}{\mu+t} - 1 \right) = 0, \end{aligned}$$

dabei wurde die bekannte Ungleichung $1 - \Phi(x) < \phi(x)/x$ für Verteilungsfunktion bzw. Dichte der Standardnormalverteilung und $x > 0$ sowie

$$e^{2\lambda/\mu} \cdot e^{-\lambda(1+t/\mu)^2/(2t)} = e^{-\lambda(1-t/\mu)^2/(2t)}$$

verwendet. Daraus folgt sofort das Gewünschte. \square

Aus dem Beweis des Lemmas ergibt sich sofort, dass inverse Gauß-Verteilungen mit demselben Erwartungswert im Sinn von \leq_{cx} geordnet sind. Das restliche Vorgehen ist nun wie bei den bisher in diesem Abschnitt betrachteten Verteilungen. Dabei verzichten wir auf die Verwendung von (3.15) und benutzen stattdessen die (vereinfachte) oben angegebene Ableitung von $1 - F_{\mu, \lambda}$. Es seien

$$\begin{aligned} a_{\mu, \lambda} &:= \frac{\partial}{\partial \mu} F_{\text{IG1}}(\mu, \lambda) = e^{\theta\mu}, \\ c_{\mu, \lambda} &:= \frac{\partial}{\partial \mu} F_{\text{IG2}}(\mu, \lambda) = 2e^{\theta\mu} \left(\int_0^\infty e^{\theta \int_t^\infty 1 - F_{\mu, \lambda}(x) dx} - 1 dt \right. \\ &\quad \left. + \int_0^\infty e^{\theta \int_t^\infty 1 - F_{\mu, \lambda}(x) dx} \cdot \int_t^\infty \frac{2\lambda}{\mu^2} e^{2\lambda/\mu} \cdot \Phi\left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1\right)\right) dx dt \right), \\ d_{\mu, \lambda} &:= \frac{\partial}{\partial \lambda} F_{\text{IG2}}(\mu, \lambda) = 2e^{\theta\mu} \int_0^\infty e^{\theta \int_t^\infty 1 - F_{\mu, \lambda}(x) dx} \\ &\quad \cdot \int_t^\infty \frac{1}{\sqrt{2\pi\lambda x}} \exp\left(-\frac{\lambda}{2x} \left(\frac{x}{\mu} - 1\right)^2\right) - \frac{2}{\mu} e^{2\lambda/\mu} \cdot \Phi\left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1\right)\right) dx dt \end{aligned}$$

und J_{IG} die entsprechende Jacobi-Matrix. Weiter seien $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$ die Momentenschätzer für die ersten beiden Momente der Betriebsdauerverteilung, wenn die Bedienzeit $\text{IG}(\mu, \lambda)$ -verteilt ist und die Längen von m Betriebsphasen vorliegen. Dann sind implizit zwei Momentenschätzer für μ und λ gegeben, diese bezeichnen wir mit $\hat{\mu}_m$ und $\hat{\lambda}_m$. Der Beweis des folgenden Satzes ergibt sich nun direkt aus den obigen Ausführungen, dabei entsteht die Matrix Σ_{IG} durch das Matrizenprodukt $J_{\text{IG}}^{-1} \Sigma (J_{\text{IG}}^{-1})'$.

Satz 3.3.22. *Sind die Bedienzeiten $\text{IG}(\mu, \lambda)$ -verteilt, dann gilt*

$$\sqrt{m} \left(\begin{pmatrix} \hat{\mu}_m \\ \hat{\lambda}_m \end{pmatrix} - \begin{pmatrix} \mu \\ \lambda \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_2(0, \Sigma_{\text{IG}})$$

mit $\Sigma_{\text{IG}} = (\sigma_{ij})_{i,j=1,2}$, $\sigma_{12} = \sigma_{21}$ und

$$\sigma_{11} = \frac{v_{\mu,\lambda}}{a_{\mu,\lambda}^2}, \quad \sigma_{12} = \frac{a_{\mu,\lambda} u_{\mu,\lambda} - c_{\mu,\lambda} v_{\mu,\lambda}}{a_{\mu,\lambda}^2 d_{\mu,\lambda}}, \quad \sigma_{22} = \frac{c_{\mu,\lambda}^2 v_{\mu,\lambda} - 2a_{\mu,\lambda} c_{\mu,\lambda} u_{\mu,\lambda} + a_{\mu,\lambda}^2 w_{\mu,\lambda}}{a_{\mu,\lambda}^2 d_{\mu,\lambda}^2}$$

sowie $u_{\mu,\lambda}, v_{\mu,\lambda}, w_{\mu,\lambda}$ aus Satz 3.3.8. Weiter sind die Schätzer $\hat{\mu}_m$ und $\hat{\lambda}_m$ stark konsistent.

Beispiel 3.3.23. Bei $\theta = 0.5$, $\mu = 1$ und $\lambda = 1$ gilt für die asymptotischen Varianzen $\text{var}(\hat{\mu}) = 0.7512185$ und $\text{var}(\hat{\lambda}) = 8.4851334$. Um die Annäherung an diese Werte bei endlichem m zu prüfen, wurden in 1000 Simulationsdurchläufen die Länge von jeweils $m = 8000$ Betriebsdauern bestimmt, dabei ergab sich $m \cdot \text{var}(\hat{\mu}_m) = 0.7600927$ und $m \cdot \text{var}(\hat{\lambda}_m) = 9.120996$. Auch in diesem Beispiel liegen also die beobachteten Werte recht nah an den Grenzwerten. \triangleleft

Die Gammaverteilung

Wir verwenden aus dem gleichen Grund wie oben auch bei der Gammaverteilung den im vorherigen Abschnitt verwendeten Ansatz. Dazu muss in einem Zwischenschritt die Verteilung so umparametrisiert werden, dass einer der beiden Parameter der Erwartungswert ist. Sei also

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \mapsto \begin{pmatrix} \alpha \\ \frac{\alpha}{\lambda} \end{pmatrix} =: \begin{pmatrix} F_{Z1}(\alpha, \lambda) \\ F_{Z2}(\alpha, \lambda) \end{pmatrix} =: \begin{pmatrix} \eta \\ \chi \end{pmatrix},$$

wodurch eine $\Gamma(\alpha, \lambda)$ -Verteilung zu einer $\Gamma(\eta, \frac{\eta}{\chi})$ -Verteilung wird. Diese Abbildung ist umkehrbar und die Umkehrabbildung besitzt die Jacobi-Matrix

$$B := B(\eta, \chi) := \begin{pmatrix} 1 & 0 \\ \frac{1}{\chi} & \frac{-\eta}{\chi^2} \end{pmatrix}.$$

Seien $\Gamma_o(x,y) := \int_0^y e^{-t} t^{x-1} dt$ und $\Gamma_u(x,y) = \int_y^\infty e^{-t} t^{x-1} dt$ die unvollständige Gammafunktion der oberen bzw. unteren Grenze. Bekannterweise lässt sich die Verteilungsfunktion F einer Gammaverteilung mit Parametern a und b durch $F(x) = \Gamma_o(a,bx)/\Gamma(a)$ darstellen. Dann definieren wir

$$\begin{pmatrix} \eta \\ \chi \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\theta}(e^{\theta\chi} - 1) \\ \frac{2}{\theta}e^{\theta\chi} \int_0^\infty e^{\theta \int_t^\infty \Gamma_u(\eta, \frac{\eta}{\chi}x)/\Gamma(\eta) dx} - 1 dt \end{pmatrix} := \begin{pmatrix} F_{\Gamma_1}(\eta, \chi) \\ F_{\Gamma_2}(\eta, \chi) \end{pmatrix}.$$

Nach diesen Vorarbeiten gehen wir im Rest des Abschnitts wie bei den anderen Verteilungen vor.

Lemma 3.3.24. *Die Abbildung*

$$\begin{pmatrix} \eta \\ \chi \end{pmatrix} \mapsto \begin{pmatrix} F_{\Gamma_1}(\eta, \chi) \\ F_{\Gamma_2}(\eta, \chi) \end{pmatrix}$$

hat eine invertierbare Jacobi-Matrix für alle $\eta > 0$ und $\chi > 0$.

Beweis. Wir benutzen, dass für unabhängige Zufallsvariablen $X \sim \Gamma(\eta_1, \eta_1/\chi)$ und $Y \sim \Gamma(\eta_2, \eta_2/\chi)$ sowie $0 < \eta_1 < \eta_2$ nach [Sto83, S. 14] gilt $X \geq_{cx} Y$, also insbesondere $E(X - t)^+ \geq E(Y - t)^+$ für alle $t \geq 0$. Ausgehend von den Ausführungen in Abschnitt 3.3.2 ist jetzt zum Beweis der Behauptung nur noch zu zeigen, dass die letzte Ungleichung für alle t aus einer Nicht-Nullmenge strikt ist. Da X und Y unterschiedliche Varianzen haben, müssen sich ihre Verteilungen unterscheiden. Damit muss es ein t_0 geben, für das die Stop-Loss-Transformierten verschieden sind und für das folglich die Ungleichung strikt ist. Schließlich ist die Abbildung $t \mapsto E(X - t)^+$ stetig in t , so dass die Ungleichung dann auch in einer Umgebung von t_0 strikt sein muss. \square

Seien nun wieder

$$\begin{aligned} b_{\eta, \chi} &:= \frac{\partial}{\partial \chi} F_{\Gamma_1}(\eta, \chi) = e^{\theta\chi}, \\ c_{\eta, \chi} &:= \frac{\partial}{\partial \eta} F_{\Gamma_2}(\eta, \chi) = 2e^{\theta\chi} \int_0^\infty e^{\theta \int_t^\infty \Gamma_u(\eta, \frac{\eta}{\chi}x)/\Gamma(\eta) dx} \cdot \int_t^\infty \frac{1}{\Gamma(\eta)} \\ &\quad \left(\int_{\eta x/\chi}^\infty y^{\eta-1} e^{-y} \log(y) dy - \eta^{\eta-1} (x/\chi)^\eta e^{-\eta x/\chi} - \int_{\eta x/\chi}^\infty y^{\eta-1} e^{-y} dy \cdot \Psi(\eta) \right) dx dt, \\ d_{\eta, \chi} &:= \frac{\partial}{\partial \chi} F_{\Gamma_2}(\eta, \chi) = 2e^{\theta\chi} \left(\int_0^\infty e^{\theta \int_t^\infty \Gamma_u(\eta, \frac{\eta}{\chi}x)/\Gamma(\eta) dx} - 1 dt \right. \\ &\quad \left. + \int_0^\infty e^{\theta \int_t^\infty \Gamma_u(\eta, \frac{\eta}{\chi}x)/\Gamma(\eta) dx} \int_t^\infty \frac{1}{\chi \Gamma(\eta)} (\eta x/\chi)^\eta e^{-\eta x/\chi} dx dt \right) \end{aligned}$$

und es bezeichne J_Γ die entsprechende Jacobi-Matrix. Weiter seien $\bar{C}_{m,1}$ und $\bar{C}_{m,2}$ die Momentenschätzer für die ersten beiden Momente der Betriebsdauerverteilung, wenn

die Bedienzeit $\Gamma(\alpha, \lambda)$ -verteilt ist und die Längen von m Betriebsphasen vorliegen. Dann sind implizit zwei Momentenschätzer für α und λ gegeben, diese bezeichnen wir mit $\hat{\alpha}_m$ und $\hat{\lambda}_m$. Der Beweis des folgenden Satzes ergibt sich nun direkt aus den obigen Ausführungen, dabei entsteht Σ_Γ durch das Matrizenprodukt $B J_\Gamma^{-1} \Sigma (J_\Gamma^{-1})' B'$.

Satz 3.3.25. *Sind die Bedienzeiten $\Gamma(\alpha, \lambda)$ -verteilt, dann gilt*

$$\sqrt{m} \left(\begin{pmatrix} \hat{\alpha}_m \\ \hat{\lambda}_m \end{pmatrix} - \begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_2(0, \Sigma_\Gamma)$$

mit $\Sigma_\Gamma = (\sigma_{ij})_{i,j=1,2}$, $\sigma_{12} = \sigma_{21}$ und

$$\begin{aligned} \sigma_{11} &= \frac{d_{\alpha, \frac{\alpha}{\lambda}}^2 v_{\alpha, \lambda} - 2b_{\alpha, \frac{\alpha}{\lambda}} d_{\alpha, \frac{\alpha}{\lambda}} u_{\alpha, \lambda} + b_{\alpha, \frac{\alpha}{\lambda}}^2 w_{\alpha, \lambda}}{b_{\alpha, \frac{\alpha}{\lambda}}^2 c_{\alpha, \frac{\alpha}{\lambda}}^2}, \\ \sigma_{12} &= \frac{\lambda \left((d_{\alpha, \frac{\alpha}{\lambda}} + \lambda c_{\alpha, \frac{\alpha}{\lambda}}) d_{\alpha, \frac{\alpha}{\lambda}} v_{\alpha, \lambda} - (2d_{\alpha, \frac{\alpha}{\lambda}} + \lambda c_{\alpha, \frac{\alpha}{\lambda}}) b_{\alpha, \frac{\alpha}{\lambda}} u_{\alpha, \lambda} + b_{\alpha, \frac{\alpha}{\lambda}}^2 w_{\alpha, \lambda} \right)}{\alpha b_{\alpha, \frac{\alpha}{\lambda}}^2 c_{\alpha, \frac{\alpha}{\lambda}}^2}, \\ \sigma_{22} &= \frac{\lambda^2 \left((\lambda c_{\alpha, \frac{\alpha}{\lambda}} + d_{\alpha, \frac{\alpha}{\lambda}})^2 v_{\alpha, \lambda} - 2(d_{\alpha, \frac{\alpha}{\lambda}} + \lambda c_{\alpha, \frac{\alpha}{\lambda}}) b_{\alpha, \frac{\alpha}{\lambda}} u_{\alpha, \lambda} + b_{\alpha, \frac{\alpha}{\lambda}}^2 w_{\alpha, \lambda} \right)}{\alpha^2 b_{\alpha, \frac{\alpha}{\lambda}}^2 c_{\alpha, \frac{\alpha}{\lambda}}^2}, \end{aligned}$$

sowie $u_{\alpha, \lambda}, v_{\alpha, \lambda}, w_{\alpha, \lambda}$ aus Satz 3.3.8. Weiter sind die Schätzer $\hat{\alpha}_m$ und $\hat{\lambda}_m$ stark konsistent.

Beispiel 3.3.26. Wie schon weiter oben überprüfen wir, wie gut die Annäherung an die Grenzverteilung bei Berücksichtigung von endlich vielen Betriebsperioden ist. Dazu haben wir die zur Bestimmung der Grenzvarianzmatrix benötigten Größen berechnet. Bei den meisten Integralen waren hierzu numerische Verfahren notwendig; verwendet wurde die R-Funktion *integrate*. Die empirischen Werte ergaben sich durch die Realisierungen von m Betriebsdauern für $l = 1000$ Wiederholungen und verschiedene m . Wir bezeichnen die Grenzvarianzen wieder mit $\text{var}(\hat{\alpha})$ bzw. $\text{var}(\hat{\lambda})$ und die Stichprobenvarianz mit $\widehat{\text{var}}(\cdot)$. Die Ergebnisse finden sich in der Doppelspalte mit der Überschrift „ θ bekannt“ in Tabelle 3.4. Offenbar ist die Annäherung an die Grenzvarianzen im betrachteten Fall schon bei einer Stichprobe vom Umfang $m = 500$ recht gut und wird wie zu erwarten bei wachsendem m noch besser. Bei den kleineren Stichproben dagegen sind größere Abweichungen festzustellen. \triangleleft

Das Beispiel konkretisiert unsere Forderung vom Anfang des Abschnitts, es müsse eine „große“ Menge von Daten vorliegen: Genauer wird die Länge von hinreichend vielen Betriebsphasen benötigt. Um diese zu erhalten, muss die Verkehrsintensität, d.h. das Verhältnis von Ankünften und Abgängen, so niedrig sein, dass immer wieder Leerlaufphasen auftreten, und über einen entsprechend langen Zeitraum beobachtet werden. Bei aus der Praxis stammenden Fragestellungen mit begrenzten Daten kann diese Voraussetzung allerdings unerfüllbar sein - und führt somit zu einem neuen Problem, das den Ausgangspunkt für die Überlegungen im nächsten Abschnitt bildet.

| | θ bekannt | | θ unbekannt | |
|---------------------------|---|---|---|---|
| $\alpha = 2, \lambda = 1$ | $\text{var}(\hat{\alpha})$ | $\text{var}(\hat{\lambda})$ | $\text{var}(\hat{\alpha})$ | $\text{var}(\hat{\lambda})$ |
| $\theta = 0.5$ | 37.91733 | 9.692327 | 38.39347 | 10.20055 |
| m | $\widehat{\text{var}}(\hat{\alpha}_m - \alpha) \cdot m$ | $\widehat{\text{var}}(\hat{\lambda}_m - \lambda) \cdot m$ | $\widehat{\text{var}}(\hat{\alpha}_m - \alpha) \cdot m$ | $\widehat{\text{var}}(\hat{\lambda}_m - \lambda) \cdot m$ |
| 100 | 58.35693 | 14.88781 | 62.51598 | 17.01517 |
| 300 | 41.55276 | 10.68295 | 43.24978 | 11.84091 |
| 500 | 38.44524 | 9.871477 | 39.65101 | 10.70695 |
| 1000 | 38.33407 | 9.864015 | 39.31151 | 10.60942 |
| 8000 | 37.96668 | 9.745942 | 37.86334 | 9.873017 |

Tabelle 3.4. Vergleich von Grenz- und empirischer Varianz bei 1000 Wiederholungen und verschiedenen Stichprobengrößen für eine $\Gamma(2,1)$ -Verteilung

Wir schließen diesen Abschnitt nun mit der eingangs angekündigten Bemerkung zu den sich ergebenden Änderungen, wenn auch die Ankunftsrate θ geschätzt wird.

Bemerkung 3.3.27. Bislang wurde θ als bekannt vorausgesetzt. Wie aber kann man auch die Ankunftsrate aus den Ankunfts- und Abgangsdaten schätzen und welche Auswirkungen hat dies auf die Varianz der Schätzer für die Bedienzeitparameter?

Eine elegante Möglichkeit zur Beantwortung dieser Frage ergibt sich, wenn man ausnutzt, dass neben den Betriebsdauern auch die Längen der Leerlaufphasen („idle period“) unabhängig sowohl von diesen als auch voneinander und identisch $\text{Exp}(\theta)$ -verteilt sind (Übungsaufgabe bei [Hal88, S. 81]). Wir bezeichnen die Länge einer Leerlaufphase kurz als Leerlaufdauer und bleiben beim Fall einer Gammaverteilung; in anderen Fällen verläuft die Argumentation analog. Es seien D_1, D_2, \dots, D_m die Werte von m Leerlaufdauern, dann ist $\bar{D}_{m,1}$ der klassische Momentenschätzer für $ED_1 = 1/\theta$. Mit den von oben bekannten Bezeichnungen ergibt sich

$$n \cdot \text{cov}((\bar{D}_{n,1}, \bar{C}_{n,1}, \bar{C}_{n,2})') = \begin{pmatrix} \frac{1}{\theta^2} & 0 & 0 \\ 0 & v_{\alpha,\lambda} & u_{\alpha,\lambda} \\ 0 & u_{\alpha,\lambda} & w_{\alpha,\lambda} \end{pmatrix} =: \Sigma'$$

und für die zur Abbildung $G = (G_{\Gamma_0}, G_{\Gamma_1}, G_{\Gamma_2})' : (\theta, \eta, \chi)' \mapsto (ED_1, EC_1, EC_1^2)'$ gehörige Jacobi-Matrix $J_{\Gamma'}$ gilt

$$J_{\Gamma'} = \begin{pmatrix} \frac{-1}{\theta^2} & 0 & 0 \\ k_{\alpha,\lambda} & 0 & b_{\alpha,\lambda} \\ l_{\alpha,\lambda} & c_{\alpha,\lambda} & d_{\alpha,\lambda} \end{pmatrix}, \quad J_{\Gamma'}^{-1} = \frac{1}{b_{\alpha,\lambda} c_{\alpha,\lambda}} \begin{pmatrix} -\theta^2 b_{\alpha,\lambda} c_{\alpha,\lambda} & 0 & 0 \\ \theta^2 (b_{\alpha,\lambda} l_{\alpha,\lambda} - k_{\alpha,\lambda} d_{\alpha,\lambda}) & -d_{\alpha,\lambda} & b_{\alpha,\lambda} \\ \theta^2 k_{\alpha,\lambda} c_{\alpha,\lambda} & c_{\alpha,\lambda} & 0 \end{pmatrix}$$

($J_{\Gamma'}^{-1}$ existiert wegen Lemma 3.3.24) mit

$$\begin{aligned}
k_{\eta,\chi} &:= \frac{\partial}{\partial \theta} G_{\Gamma_1}(\theta, \eta, \chi) = \theta^{-1}(\chi e^{\theta\chi} - \theta^{-1}(e^{\theta\chi} - 1)) \\
l_{\eta,\chi} &:= \frac{\partial}{\partial \theta} G_{\Gamma_2}(\theta, \eta, \chi) = 2\theta^{-1}e^{\theta\chi} \left((\chi - \theta^{-1}) \cdot \int_0^\infty e^{\theta \int_t^\infty \Gamma_u(\eta, \frac{\eta}{\chi}x)/\Gamma(\eta)} dx - 1 dt \right. \\
&\quad \left. + \int_0^\infty e^{\theta \int_t^\infty \Gamma_u(\eta, \frac{\eta}{\chi}x)/\Gamma(\eta)} dx \cdot \int_t^\infty \frac{\Gamma_u(\eta, \frac{\eta}{\chi}x)}{\Gamma(\eta)} dx dt \right).
\end{aligned}$$

Bezeichnen wir die mithilfe von $\bar{D}_{m,1}, \bar{C}_{m,1}, \bar{C}_{m,2}$ bestimmten Schätzer für θ, α, λ mit $\hat{\theta}'_m, \hat{\alpha}'_m, \hat{\lambda}'_m$, so gilt für diese analog zu oben

$$\sqrt{m} \left(\begin{pmatrix} \hat{\theta}'_m \\ \hat{\alpha}'_m \\ \hat{\lambda}'_m \end{pmatrix} - \begin{pmatrix} \theta \\ \alpha \\ \lambda \end{pmatrix} \right) \xrightarrow{d} U, U \sim N_3(0, B J_{\Gamma'}^{-1} \Sigma' (J_{\Gamma'}^{-1})' B')$$

Berechnet man das Matrizenprodukt in der Varianz der Normalverteilung, ergibt sich für die Grenzvarianzen

$$\begin{aligned}
\text{var}(\hat{\theta}') &= \theta^2, \quad \sigma'_{11} := \text{var}(\hat{\alpha}') = \sigma_{11} + \frac{\theta^2 (b_{\alpha, \frac{\alpha}{\lambda}} l_{\alpha, \frac{\alpha}{\lambda}} - k_{\alpha, \frac{\alpha}{\lambda}} d_{\alpha, \frac{\alpha}{\lambda}})^2}{b_{\alpha, \frac{\alpha}{\lambda}}^2 c_{\alpha, \frac{\alpha}{\lambda}}^2} \geq \sigma_{11}, \\
\sigma'_{22} := \text{var}(\hat{\lambda}') &= \sigma_{22} + \frac{\lambda^2 \theta^2 (b_{\alpha, \frac{\alpha}{\lambda}} l_{\alpha, \frac{\alpha}{\lambda}} - d_{\alpha, \frac{\alpha}{\lambda}} k_{\alpha, \frac{\alpha}{\lambda}} - \lambda c_{\alpha, \frac{\alpha}{\lambda}} k_{\alpha, \frac{\alpha}{\lambda}})^2}{\alpha^2 b_{\alpha, \frac{\alpha}{\lambda}}^2 c_{\alpha, \frac{\alpha}{\lambda}}^2} \geq \sigma_{22},
\end{aligned}$$

dabei sind σ_{11} und σ_{22} die entsprechenden Varianzen bei bekanntem θ aus Satz 3.3.25. Auch die Kovarianzen ließen sich bestimmen und vergleichen, wir verzichten aber darauf.

In der Doppelspalte mit der Überschrift „ θ unbekannt“ von Tabelle 3.4 finden sich die Varianzen der auf denselben Daten wie im vorherigen Beispiel basierenden Schätzer, wenn auch θ aus der Stichprobe geschätzt wird, und verschiedene Werte von m . Diese Ergebnisse stützen unser obiges Vorgehen, θ als bekannt anzunehmen, da dies (zumindest im betrachteten Beispiel) die Varianzen nicht wesentlich verändert. Werden nur die Varianzen von auf Stichproben der Mindestgröße 500 basierenden Schätzern berücksichtigt, beträgt die maximale Abweichung bei den in der Tabelle aufgeführten Ergebnissen gegenüber bekanntem θ (meist sogar deutlich) weniger als 10%.

Schließlich findet sich in Abb. 3.5 ein Vergleich der relativen Abweichung der asymptotischen Varianz von $\hat{\lambda}'_m$ in θ bei aus der Stichprobe geschätztem θ gegenüber dem Fall, dass θ bekannt ist, für $\alpha = 2$ und verschiedene Werte von λ . Das Verhalten für verschiedene Werte von α und festem λ ist ähnlich und auch für die asymptotische Varianz von $\hat{\alpha}'_m$ ergibt sich, allerdings bei geringerer Abweichung, ein vergleichbares Bild. Wir verzichten deswegen auf eine eigene Abbildung für diese Fälle. Die relative Abweichung steigt von 0 beginnend in θ kontinuierlich bis zu ihrem Maximalwert an und fällt

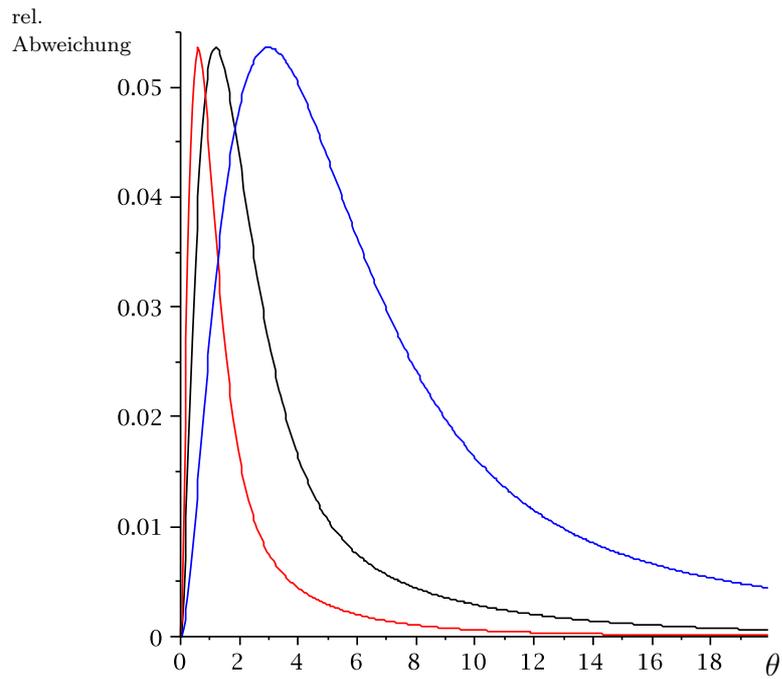


Abbildung 3.5. Relative Abweichung der asymptotischen Varianz von $\hat{\lambda}_m$ in θ bei aus der Stichprobe geschätztem θ gegenüber dem Fall, dass θ bekannt ist, für $\alpha = 2$ und $\lambda = 1$ (rot), $\lambda = 2$ (schwarz) und $\lambda = 5$ (blau)

dann mit weiter wachsendem θ , bis sie schließlich verschwindet. Die Erklärung für dieses Verhalten ist naheliegend. Solange $1/\theta$ deutlich kleiner als der Erwartungswert EZ_1 der Bedienzeit ist, wird in einer Betriebsperiode in den meisten Fällen nur ein Kunde bedient. Die Ankunftsrate spielt also nur insofern eine Rolle, als dass sie so klein ist, dass meist kein weiterer Kunde ankommt, während jemand bedient wird. Je größer die Ankunftsrate wird, desto häufiger sind gleichzeitig mehrere Kunden im System, damit wird es „schwieriger“, von der Betriebsdauer auf die Bedienzeit zu schließen. Wächst die Rate dann über einen bestimmten Wert (der so groß ist, dass $1/\theta$ kleiner als EZ_1 ist, also die mittlere Zeit zwischen zwei Ankünften kleiner als die mittlere Bedienzeit; in dieser Situation tritt der Fall, dass das System leer ist, seltener auf und es finden eher „viele“ Bedienungen während einer Betriebsperiode statt), erhöht sich wegen der genannten Schwierigkeit die Varianz des Schätzers generell, so dass die zusätzlichen Schwankungen durch Schätzen von θ asymptotisch vernachlässigbar werden. Dies lässt sich auch in der obigen Abbildung wiederfinden: Der Scheitelpunkt liegt jeweils rechts von dem θ -Wert, für den $1/\theta$ kleiner als 2 (rote Kurve), 1 (schwarze Kurve) bzw. $2/5$ (blaue Kurve) ist. Eine Veränderung von λ hat also die Verschiebung der Lage des Maximums zur Folge, da ein zunehmender Wert von λ die mittlere Bedienzeit verringert. Bei einer Abbildung für verschiedene Werte von α und festem λ wäre dies ähnlich, lediglich vergrößert hier ein höherer Wert die mittlere Bedienzeit, deshalb erfolgt die Verschiebung in die jeweils andere Richtung. ◀

3.3.3. Auf der Schlängellänge basierende Schätzer

Großer Nachteil der im vorherigen Abschnitt ausgearbeiteten Schätzer ist die Voraussetzung, dass hinreichend viele Betriebsdauern vorliegen. Aus diesem Grund entwickeln wir nun stark konsistente Schätzer, die nur auf der Anzahl der Kunden im System $(Q_t)_{0 \leq t \leq T}$ während eines hinreichend langen Zeitraums $[0, T]$ basieren, $T \in \mathbb{R}^+$ fest. Gegenüber der vorher untersuchten Situation, in der nicht bekannt war, wie viele Kunden während der jeweiligen Betriebsphasen im System waren, stehen also mehr Informationen zur Verfügung. Es werden lediglich die Pareto- und die Gammaverteilung betrachtet, wobei eine Abwandlung für andere Verteilungen in gewissen Grenzen möglich ist. Grundlage und Ausgangspunkt unserer Untersuchung ist [BP99], insbesondere der Abschnitt 5. Auch Beweise bzw. Quellenangaben für die im Nachfolgenden aufgestellten Behauptungen finden sich dort.

In einem $M/G/\infty$ -System, das sich im Gleichgewichtszustand befindet, ist die Schlängellänge zur Zeit $t > 0$ Poisson-verteilt mit einem für alle t konstanten (und natürlich positiven) Parameter. In dieser Situation sei

$$\rho(h) := \frac{\text{cov}(Q_{t+h}, Q_t)}{\text{var}(Q_t)}$$

der Korrelationskoeffizient zwischen den Schlangenlängen zu den Zeiten t und $t + h$ für ein festes $h > 0$. Weiter seien G und $\mu < \infty$ Verteilungsfunktion bzw. Erwartungswert der Bedienzeitverteilung. Dann gilt nach einem Resultat von Reynolds

$$\rho(h) = 1 - \frac{1}{\mu} \int_0^h 1 - G(u) du.$$

Der Wert von $\rho(h)$ soll nun aus einer Realisierung von $(Q_t)_{0 \leq t \leq T}$ geschätzt werden, dazu benötigen wir das folgende Lemma. Im weiteren Verlauf wird angenommen, dass sich das System im Gleichgewichtszustand befindet.

Lemma 3.3.28. *Es gilt*

$$\frac{1}{T} \int_0^T Q_{t+h} Q_t dt \xrightarrow{f.s.} \mathbb{E} Q_{s+h} Q_s, \quad \frac{1}{T} \int_0^T (Q_t)^k dt \xrightarrow{f.s.} \mathbb{E} (Q_s)^k, \quad k = 1, 2$$

und

$$\frac{\frac{1}{T} \int_0^T Q_{t+h} Q_t dt - (\frac{1}{T} \int_0^T Q_t dt)^2}{\frac{1}{T} \int_0^T (Q_t)^2 dt - (\frac{1}{T} \int_0^T Q_t dt)^2} \xrightarrow{f.s.} \rho(h).$$

Beweis. Der erste Teil ist wohlbekannt, siehe [BP99, S. 78]. Es sei $f(r, s, t) := \frac{t-r^2}{s-r^2}$. Der zweite Teil ergibt sich nun wegen $\mathbb{E}(Q_{t+h})^k = \mathbb{E}(Q_t)^k$ für $k = 1, 2$ aus dem bekannten Zusammenspiel zwischen fast sicherer Konvergenz und stetigen Funktionen, da $f(\frac{1}{T} \int_0^T Q_t dt, \frac{1}{T} \int_0^T (Q_t)^2 dt, \frac{1}{T} \int_0^T Q_{t+h} Q_t dt)$ wegen $\mathbb{E}(Q_t)^2 - (\mathbb{E} Q_t)^2 = \text{var}(Q_t) > 0$ für hinreichend große T nur auf der Menge der Stetigkeitspunkte von f verwendet wird. \square

Sei nun für feste $T, h > 0$

$$\hat{\rho}_{T,h} := \frac{\frac{1}{T-h} \int_0^{T-h} Q_{t+h} Q_t dt - (\frac{1}{T-h} \int_0^{T-h} Q_t dt)^2}{\frac{1}{T-h} \int_0^{T-h} (Q_t)^2 dt - (\frac{1}{T-h} \int_0^{T-h} Q_t dt)^2}.$$

Ist wieder eine zweiparametrische Bedienzeitverteilung vorausgesetzt, können deren Parameter mithilfe des Ansatzes

$$\begin{pmatrix} \hat{\rho}_{T,h_1} \\ \hat{\rho}_{T,h_2} \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} \rho(h_1) \\ \rho(h_2) \end{pmatrix} = \begin{pmatrix} \frac{1}{\mu} \int_{h_1}^{\infty} 1 - G(x) dx \\ \frac{1}{\mu} \int_{h_2}^{\infty} 1 - G(x) dx \end{pmatrix}$$

für feste $T, h_1, h_2 > 0$ und $h_1 \neq h_2$ bestimmt werden. Eine Verallgemeinerung auf mehr als zwei Parameter ergibt sich offensichtlich. Es bezeichnen sowohl bei der Pareto- als auch bei der Gammaverteilung $\hat{\alpha}_T$ und $\hat{\lambda}_T$ die daraus ggf. numerisch bestimmten Schätzer, wobei wir als Obergrenze T in der Regel die Realisierung des letzten Abgangs Y_n verwenden. Man bemerke, dass in der hier gegebenen Darstellung von $\rho(\cdot)$ die aus (3.14) bekannte Stop-Loss-Transformierte vorkommt. Für den Beweis des Konsistenz-Resultats können die Werte von h_1 und h_2 beliebig aus \mathbb{R}^+ gewählt sein. Wir belassen es dabei und untersuchen nicht weiter, welche Werte z.B. zu einer minimalen Varianz führen.

Die Pareto-Verteilung

Für die auf S. 36 und S. 101 beschriebene Pareto-Verteilung und beliebige $h_1, h_2 > 0$, $h_1 \neq h_2$, betrachten wir die Abbildung

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \mapsto \begin{pmatrix} \frac{\alpha-1}{\lambda} \int_{h_1}^{\infty} \frac{\lambda^\alpha}{(x+\lambda)^\alpha} dx \\ \frac{\alpha-1}{\lambda} \int_{h_2}^{\infty} \frac{\lambda^\alpha}{(x+\lambda)^\alpha} dx \end{pmatrix} = \begin{pmatrix} \frac{\lambda^{\alpha-1}}{(\lambda+h_1)^{\alpha-1}} \\ \frac{\lambda^{\alpha-1}}{(\lambda+h_2)^{\alpha-1}} \end{pmatrix} =: \begin{pmatrix} G_{P1}(\alpha, \lambda) \\ G_{P2}(\alpha, \lambda) \end{pmatrix}.$$

Im Gegensatz zu oben ist es ausreichend, $\alpha > 2$ anzunehmen, da wir [BP99] folgend nur die Existenz des zweiten Moments benötigen.

Lemma 3.3.29. *Die Funktion*

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \mapsto \begin{pmatrix} G_{P1}(\alpha, \lambda) \\ G_{P2}(\alpha, \lambda) \end{pmatrix}$$

hat eine invertierbare Jacobi-Matrix für alle $\alpha > 2$ und $\lambda > 0$.

Beweis. Wir bestimmen die Determinante der entsprechenden Jacobi-Matrix. Wegen

$$\frac{\partial}{\partial \alpha} \frac{\lambda^{\alpha-1}}{(\lambda+t)^{\alpha-1}} = \frac{\lambda^{\alpha-1} \log\left(\frac{\lambda}{\lambda+t}\right)}{(\lambda+t)^{\alpha-1}}, \quad \frac{\partial}{\partial \lambda} \frac{\lambda^{\alpha-1}}{(\lambda+t)^{\alpha-1}} = \frac{(\alpha-1)t\lambda^{\alpha-2}}{(\lambda+t)^\alpha}$$

ergibt sich

$$\begin{aligned} & \frac{\partial}{\partial \alpha} G_{P1}(\alpha, \lambda) \cdot \frac{\partial}{\partial \lambda} G_{P2}(\alpha, \lambda) - \frac{\partial}{\partial \lambda} G_{P1}(\alpha, \lambda) \cdot \frac{\partial}{\partial \alpha} G_{P2}(\alpha, \lambda) \\ &= \frac{\lambda^{\alpha-1} \log\left(\frac{\lambda}{\lambda+h_1}\right)}{(\lambda+h_1)^{\alpha-1}} \cdot \frac{(\alpha-1)h_2\lambda^{\alpha-2}}{(\lambda+h_2)^\alpha} - \frac{\lambda^{\alpha-1} \log\left(\frac{\lambda}{\lambda+h_2}\right)}{(\lambda+h_2)^{\alpha-1}} \cdot \frac{(\alpha-1)h_1\lambda^{\alpha-2}}{(\lambda+h_1)^\alpha} \\ &= \frac{\lambda^{2\alpha-3}(\alpha-1)}{(\lambda+h_1)^{\alpha-1}(\lambda+h_2)^{\alpha-1}} \left(\frac{h_2}{\lambda+h_2} \cdot \log\left(\frac{\lambda}{\lambda+h_1}\right) - \frac{h_1}{\lambda+h_1} \cdot \log\left(\frac{\lambda}{\lambda+h_2}\right) \right). \end{aligned}$$

Der Ausdruck vor der Klammer ist offensichtlich positiv. Wegen

$$\frac{h_2}{\lambda+h_2} - \frac{h_1}{\lambda+h_1} = \frac{\lambda(h_2-h_1)}{(\lambda+h_1)(\lambda+h_2)}, \quad \log\left(\frac{\lambda}{\lambda+h_1}\right) - \log\left(\frac{\lambda}{\lambda+h_2}\right) = \log\left(\frac{\lambda+h_2}{\lambda+h_1}\right)$$

ist die verbleibende Differenz positiv für $h_1 < h_2$ und negativ im Fall $h_1 > h_2$, damit verschwindet die Determinante nicht. \square

Satz 3.3.30. *Sind die Bedienzeiten $\text{ParII}(\alpha, \lambda)$ -verteilt, ist der Schätzer $(\hat{\alpha}_{Y_n}, \hat{\lambda}_{Y_n})'$ stark konsistent.*

Beweis. Zu zeigen ist $(\hat{\alpha}_{Y_n}, \hat{\lambda}_{Y_n})' \xrightarrow{\text{f.s.}} (\alpha, \lambda)'$ und es gilt $(\hat{\rho}_{Y_n, h_1}, \hat{\rho}_{Y_n, h_2})' \xrightarrow{\text{f.s.}} (\rho(h_1), \rho(h_2))'$ nach Lemma 3.3.28. Die Jacobi-Matrix der Abbildung

$$\begin{pmatrix} \alpha \\ \lambda \end{pmatrix} \mapsto \begin{pmatrix} \hat{\rho}_{Y_n, h_1} \\ \hat{\rho}_{Y_n, h_2} \end{pmatrix}$$

ist nach Lemma 3.3.29 invertierbar. Damit ist nach dem Satz von der Umkehrabbildung auch die Umkehrfunktion dieser Funktion stetig differenzierbar, also insbesondere stetig. Insgesamt ergibt sich nun die Behauptung aus dem Zusammenspiel von fast sicherer Konvergenz und stetigen Funktionen. \square

Bemerkung 3.3.31. Theorem 5.1 auf S. 78 in [BP99] kombiniert mit Lemma 3.3.29 ließe zusätzlich auf asymptotische Normalität der betrachteten Schätzer schließen. In unseren Simulationen ergab sich allerdings ein großer Unterschied zwischen den Ergebnissen bei endlichem Zeithorizont und der nach der Theorie zu erwartenden Grenzvarianzmatrix, wobei näherungsweise eine Normalverteilung vorzuliegen schien, nur nicht mit der nach dem Theorem zu erwartenden Varianz. Da zusätzlich der Beweis zu erwähntem Theorem unvollständig zu sein scheint, verfolgen wir den Nachweis asymptotischer Normalität hier nicht weiter. \triangleleft

Die Gammaverteilung

Auf den ersten Blick bietet sich hier ein ähnliches Vorgehen wie bei der Pareto-Verteilung im Abschnitt zuvor an. Leider ist es uns aber wegen der vorkommenden komplizierteren Ausdrücke nicht gelungen, ein Resultat analog zu Lemma 3.3.29 nachzuweisen. Ein weiterer Ansatz wäre die Verwendung des aus (3.7) bekannten (ML-)Schätzers $(\widehat{EZ}_1)_n^{\text{MM}} = \bar{Z}_{n,1} = \bar{Y}_{n,1} - \bar{X}_{n,1}$ für den Erwartungswert in Verbindung mit $\hat{\rho}_{Y_n, h}$ für ein festes $h > 0$, um die benötigten zwei Gleichungen zu erhalten. In diesem Fall wird bei einem analogen Vorgehen für den Nachweis von Konsistenz allerdings benötigt, dass $\frac{\partial}{\partial \eta} E_{\eta, \eta/\chi}(X - h)^+$ für jedes h von 0 verschieden ist. Auch dies konnten wir leider nicht zeigen, sondern nur in Spezialfällen durch Simulationen nachprüfen.

Beispiel 3.3.32. Auch zum Abschluss dieses Abschnitts sollen unsere Ergebnisse mittels einer Simulationsstudie illustriert werden. Darin vergleichen wir drei Schätzer bei gammaverteilter Bedienzeit: Den asymptotisch normalen Schätzer aus Satz 3.3.25, den soeben beschriebenen Schätzer (den wir zur Unterscheidung von den anderen beiden als „konsistenten“ Schätzer bezeichnen) und den Fall, dass die Parameter über Schätzer für $\rho(h_1)$ und $\rho(h_2)$ mit positiven $h_1 \neq h_2$ bestimmt werden. In letzterer Situation konnten wir zwar keine Eigenschaften für die Schätzer nachweisen, wollen aber trotzdem das Verhalten der resultierenden Schätzer vergleichen. Der verwendete Quellcode ist zur besseren Nachvollziehbarkeit der Ergebnisse in Anhang C aufgeführt.

Die theoretischen Grundlagen für den „konsistenten“ Schätzer setzen voraus, dass sich das System im Gleichgewichtszustand befindet. Da es in unseren Probeläufen die Ergebnisse der Simulationen nur vernachlässigbar beeinflusste, ob eine gewisse „Einschwingphase“ vor der zur Berechnung dieses Schätzers verwendeten Realisierung von (Q_t) gelöscht wurde oder nicht, haben wir der Einfachheit halber darauf verzichtet, und auch den Teil der Realisierung benutzt, in der das System noch nicht im Gleichgewicht war.

Wir beginnen damit, anhand eines Beispiels aufzuzeigen, dass die Wahl von h einen Einfluss auf die Güte des „konsistenten“ Schätzers hat. Dazu wurde dieser Schätzer in einem bestimmten Modell für verschiedene Werte von n jeweils $l = 1000$ Mal bestimmt, einmal für $h = 0.5$ und einmal für $h = 2.5$. Die Ergebnisse finden sich in Tabelle 3.5. Wir erläutern im Folgenden die Beschriftung der Spalten dieser Tabelle. In der ersten Spalte mit Überschrift n steht die Anzahl der berücksichtigten Ankunfts- und Abgangszeitpunkte. Die Größen \tilde{l}_{K1} bzw. \tilde{l}_{K5} geben an, in wie vielen der 1000 Wiederholungen tatsächlich ein Schätzer bestimmt wurde, in den übrigen Fällen konnten die Bestimmungsgleichungen nicht mit positiven Werten für α und λ aufgelöst werden, so dass das Schätzverfahren versagte. Die Spalten mit Kopf $\overline{(Y_n)}_{\tilde{l}_{K1}}$ bzw. $\overline{(Y_n)}_{\tilde{l}_{K5}}$ geben die durchschnittliche Länge der jeweils verwendeten Intervalle $[0, T]$ an. In der Mitte der Tabelle schließlich werden die Mittelwerte und Stichprobenvarianzen der bestimmten Schätzer für α und λ gegenübergestellt, wieder in Bezug auf die Fälle, in denen das jeweilige Verfahren funktioniert hat.

Offenbar sind die Ergebnisse für $h = 2.5$ sowohl bei der Anzahl der Fälle, in denen das Verfahren funktioniert hat, als auch bei den Werten der Schätzer deutlich besser. Nur für den kleinsten Wert von n wurde dort in sieben Fällen keine Lösung gefunden, während bei $h = 0.5$ bei jedem untersuchten Wert von n mindestens einmal keine Lösung bestimmt werden konnte, für $n = 1000$ sogar in mehr als einem Drittel der Fälle nicht. Auch bei Mittelwerten und Stichprobenvarianzen ergeben sich bei $h = 2.5$ deutlich bessere Werte: Für $n = 1000$ liegen die Mittelwerte in beiden Fällen deutlich vom wahren Wert entfernt, aber schon ab $n = 5000$ liefert das Verfahren mit $h = 2.5$ akzeptable Werte und trifft für $n = 5 \cdot 10^4$ die wahren Werte im Durchschnitt fast exakt, während bei $h = 0.5$ immer ein höherer, wenn auch für $n = 5 \cdot 10^4$ mit 2% relativer Abweichung annehmbarer Fehler auftritt. Wie zu erwarten, nimmt die Güte in beiden Fällen mit wachsender Datenmenge zu. Ausgehend von diesen Beobachtungen verwenden wir im folgenden Vergleich den konsistenten Schätzer mit $h = 2.5$.

Die Ergebnisse des angekündigten Vergleichs für Daten aus einer $\Gamma(2,1)$ -Verteilung und $\theta \in \{0.5, 1.5\}$ finden sich in den Tabellen 3.6 und 3.7; dabei wird die gleiche Notation wie oben verwendet. Der Index H steht für den asymptotisch normalen Schätzer, der auf Arbeiten von Peter Hall basiert. Zu dem „konsistenten“ Schätzer mit $h = 2.5$ gehörige Größen erhalten den Index K, während die vom verbleibenden Schätzer, für den $h_1 = 0.5$

| $\alpha = 2, \lambda = 1$ | | | | | | |
|---------------------------|------------------|-------------------------------------|--|--|-------------------------------------|------------------|
| $\theta = 1.5$ | | | | | | |
| n | \tilde{l}_{K1} | $\overline{(Y_n)}_{\tilde{l}_{K1}}$ | $\overline{(\hat{\alpha}_m)}_{\tilde{l}_{K1}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_{K1}}$ $\overline{(\hat{\lambda}_m)}_{\tilde{l}_{K1}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_{K1}}$ | $\overline{(\hat{\alpha}_m)}_{\tilde{l}_{K5}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_{K5}}$ $\overline{(\hat{\lambda}_m)}_{\tilde{l}_{K5}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_{K5}}$ | $\overline{(Y_n)}_{\tilde{l}_{K5}}$ | \tilde{l}_{K5} |
| 10^3 | 659 | 671.6 | 1.8084 0.8067679 0.9042001 0.201503 | 2.33659 0.812547 1.16903 0.204864 | 671.4 | 993 |
| $5 \cdot 10^3$ | 867 | 3338.6 | 2.14590 0.627543 1.07353 0.157930 | 2.07046 0.112103 1.03554 0.0282190 | 3339.4 | 10^3 |
| 10^4 | 949 | 6670.8 | 2.12514 0.498934 1.06295 0.124882 | 2.02683 0.0521873 1.01381 0.0131433 | 6670.9 | 10^3 |
| $5 \cdot 10^4$ | 999 | 33339.8 | 2.04238 0.081440 1.02144 0.0203731 | 2.00701 0.0101333 1.00376 0.00255485 | 33340.0 | 10^3 |

Tabelle 3.5. Vergleich des konsistenten Schätzers für $h = 0.5$ (Spalten mit K1) und $h = 2.5$ (Spalten mit K5)

und $h_2 = 1$ gewählt wurde, stammende Werte in Erinnerung an die Resultate von John Reynolds mit R markiert werden. Wir sprechen deshalb im Folgenden auch vom Hall- bzw. Reynolds-Schätzer. Die bisher nicht aufgetretene Größe $\bar{m}_{\bar{h}}$ gibt an, auf wie vielen Betriebsdauern der asymptotisch normale Schätzer durchschnittlich beruht.

In den Tabellen wird die den Ausgangspunkt für die Untersuchungen in diesem Abschnitt bildende Schwäche des Hall-Schätzers deutlich, die darin besteht, dass eine ausreichend große Anzahl von Betriebsdauern benötigt wird. Liegen diese nicht vor, z.B. bei den kleineren Werten von n oder generell bei dem größeren Wert von θ , sind die Ergebnisse des Schätzers bei Mittelwert und Stichprobenvarianz entsprechend schlecht. Bei $\theta = 1.5$ und $n = 1000$ versagt das Verfahren in ungefähr einem Viertel der Fälle. Mit wachsender Datenmenge verbessern sich aber seine Kenngrößen, bis er für $\theta = 0.5$ und $n = 5 \cdot 10^4$ den beiden anderen ebenbürtig bzw. bei der Varianz sogar leicht überlegen ist. Die beiden anderen Schätzer liefern dagegen auch schon ab $n = 5000$ gute Ergebnisse. Zunächst ist der Reynolds-Schätzer besser, bei $n = 5 \cdot 10^4$ übertrifft ihn dann aber der „konsistente“ Schätzer für beide θ -Werte leicht. Letzteres könnte darauf hindeuten, dass $\bar{Z}_{n,1}$ für große Werte von n den Erwartungswert der Bedienzeitverteilung genauer schätzt als $\hat{\rho}_{Y_n, h_1}$ bzw. $\hat{\rho}_{Y_n, h_2}$ den entsprechenden Korrelationskoeffizienten. Es ist nicht auszuschließen, dass für andere Werte von h bzw. h_1 und h_2 umgekehrte oder allgemein bessere Ergebnisse erzielt werden können. Hier sind weitere Untersuchungen nötig.

Wir bemerken noch, dass sich das Vorschalten einer „Einschwingphase“ nicht auf den Hall-Schätzer auswirken würde, weil dieser keinen Gleichgewichtszustand voraussetzt. Da durch dieses Vorgehen eine Reduzierung der für die beiden anderen Schätzer vorliegenden Daten erfolgen würde, ist davon auszugehen, dass sich deren Güte (je nach Länge des Einschwingens stärker oder schwächer) vermindern würde. \triangleleft

Bemerkung 3.3.33. Die Genauigkeit des Hall-Schätzers verschlechtert sich bei zunehmendem θ , da mit diesem die mittlere Länge einer Betriebsperiode wächst und somit deren Anzahl insgesamt geringer wird. Ein Vergleich der Kenngrößen des Hall-Schätzers für $\theta = 0.5$ und $\theta = 1.5$ sowie jeweils gleiches n im vorherigen Beispiel illustriert diesen Effekt. Bei konstanten Parameterwerten der Bedienzeitverteilung besteht für hinreichend großes θ jeder beobachtete Zeitraum aus nur einer Betriebsperiode und das Verfahren versagt. Ein Blick in die Ausführungen nach Lemma 2.3.8 auf Seite 43 zeigt, dass die bei großem θ auftretende hohe Anzahl an zulässigen Permutationen dazu führt, dass sich die bedingte Verteilung der wahren Permutation einer Gleichverteilung annähert. Für Verteilungen ohne theoretisches Ergebnis wie Satz 2.2.8 führt dies in der praktischen Anwendung dazu, dass bei der MAP-Methode numerische Probleme auftreten können, weil sich die einzelnen Gewichte im linearen Assignment-Problem kaum noch unterscheiden. In beiden Fällen führt ein hinreichend großer Wert von θ also dazu, dass in den jeweiligen Fällen die Lösung des Problems erschwert werden kann. \triangleleft

| $\alpha = 2, \lambda = 1, h = 2.5$ | | | | | | | | |
|------------------------------------|---------------|-------------------------|--|--|--|----------------------------------|---------------|---------------|
| $\theta = 1.5$ | | | | | | | | |
| n | \tilde{l}_H | $\bar{m}_{\tilde{l}_H}$ | $\overline{(\hat{\alpha}_m)_{\tilde{l}_H}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_H}$ $\overline{(\hat{\lambda}_m)_{\tilde{l}_H}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_H}$ | $\overline{(\hat{\alpha}_m)_{\tilde{l}_K}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_K}$ $\overline{(\hat{\lambda}_m)_{\tilde{l}_K}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_K}$ | $\overline{(\hat{\alpha}_m)_{\tilde{l}_R}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_R}$ $\overline{(\hat{\lambda}_m)_{\tilde{l}_R}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_R}$ | $\overline{(Y_n)_{\tilde{l}_K}}$ | \tilde{l}_K | \tilde{l}_R |
| 10^3 | 755 | 51.9 | 2.61867 1.49640 1.31758 0.369319 | 2.33659 0.812547 1.16903 0.204864 | 2.16742 0.534916 1.08719 0.114074 | 671.4 | 993 | 998 |
| $5 \cdot 10^3$ | 985 | 250.435 | 2.30477 0.660042 1.15264 0.162459 | 2.07046 0.112103 1.03554 0.0282190 | 2.02896 0.0707333 1.01678 0.0161644 | 3339.4 | 10^3 | 10^3 |
| 10^4 | 10^3 | 499.6 | 2.09991 0.232028 1.05044 0.0574423 | 2.02683 0.0521873 1.01381 0.0131433 | 2.00730 0.0336412 1.00361 0.00786151 | 6670.9 | 10^3 | 10^3 |
| $5 \cdot 10^4$ | 10^3 | 2494.0 | 2.03335 0.043494 1.01717 0.0108098 | 2.00701 0.0101333 1.00376 0.00255485 | 2.00979 0.00672590 1.00480 0.0015795 | 33340.0 | 10^3 | 10^3 |

Tabelle 3.6. Vergleich der Kenngrößen dreier Schätzer aus einer Simulationsstudie I

| $\alpha = 2, \lambda = 1, h = 2.5$ | | | | | | | | |
|------------------------------------|---------------|-------------------------|--|--|--|----------------------------------|---------------|---------------|
| $\theta = 0.5$ | | | | | | | | |
| n | \tilde{l}_H | $\bar{m}_{\tilde{l}_H}$ | $\overline{(\hat{\alpha}_m)_{\tilde{l}_H}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_H}$ $\overline{(\hat{\lambda}_m)_{\tilde{l}_H}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_H}$ | $\overline{(\hat{\alpha}_m)_{\tilde{l}_K}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_K}$ $\overline{(\hat{\lambda}_m)_{\tilde{l}_K}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_K}$ | $\overline{(\hat{\alpha}_m)_{\tilde{l}_R}}$ $\widehat{\text{var}}(\hat{\alpha}_m)_{\tilde{l}_R}$ $\overline{(\hat{\lambda}_m)_{\tilde{l}_R}}$ $\widehat{\text{var}}(\hat{\lambda}_m)_{\tilde{l}_R}$ | $\overline{(Y_n)_{\tilde{l}_K}}$ | \tilde{l}_K | \tilde{l}_R |
| 10^3 | 10^3 | 369.9 | 2.0576 0.114706 1.03089 0.0288762 | 2.10601 0.220305 1.05357 0.0557853 | 2.04394 0.141123 1.02573 0.0327979 | 2007.8 | 10^3 | 10^3 |
| $5 \cdot 10^3$ | 10^3 | 1842.4 | 2.01634 0.0208759 1.00877 0.0053072 | 2.02419 0.0365460 1.01239 0.0092793 | 2.00317 0.0256839 1.00363 0.00591647 | 10011.4 | 10^3 | 10^3 |
| 10^4 | 10^3 | 3682.8 | 1.99859 0.0102823 1.00011 0.00266336 | 1.99835 0.0168499 0.999515 0.00420871 | 1.99775 0.0128930 0.99921 0.00294203 | 20006.0 | 10^3 | 10^3 |
| $5 \cdot 10^4$ | 10^3 | 18397.7 | 2.00015 0.00222124 1.00031 0.000572202 | 2.00003 0.00374050 1.00026 0.00094074 | 1.99819 0.0025924 0.99942 0.00059894 | 100012.9 | 10^3 | 10^3 |

Tabelle 3.7. Vergleich der Kenngrößen dreier Schätzer aus einer Simulationsstudie II

A. Verschiedene Berechnungsformen von Kendalls Tau-Abstand

Bei der Sichtung von Literatur, die sich mit Metriken auf Permutationen beschäftigt, stößt man schnell auf verschiedene, allerdings äquivalente Definitionen von Kendalls Tau-Abstand. Beweise für die Äquivalenzen dagegen finden sich nicht so einfach. Deswegen geben wir im Folgenden drei dieser Definitionen an und nennen Quellen, mit deren Hilfe die Gleichwertigkeit gezeigt werden kann. Wir bezeichnen dabei eine Transposition, bei der zwei benachbarte Elemente vertauscht werden, als *Nachbartransposition* und schreiben $\text{Nb}(S_n)$ für die Teilmenge der Nachbartranspositionen von S_n .

Definition A.1. Es seien $\sigma, \pi \in S_n$ zwei Permutationen.

(a) Wir setzen $d_{\tau,1}(\sigma, \pi) := \sum_{i=1}^{n-1} \sum_{j=i+1}^n D_{ij}$ mit

$$D_{ij} := \mathbf{1}_{\{\sigma(i) > \sigma(j)\}} \cdot \mathbf{1}_{\{\pi(i) < \pi(j)\}} + \mathbf{1}_{\{\sigma(i) < \sigma(j)\}} \cdot \mathbf{1}_{\{\pi(i) > \pi(j)\}}.$$

(b) Es sei $d_{\tau,2}(\sigma, \pi) := \#\{(i, j) : 1 \leq i < j \leq n, \sigma(\pi^{-1}(i)) > \sigma(\pi^{-1}(j))\}$.

(c) Wir definieren $d_{\tau,3}(\sigma, \pi) := \min\{k \in \mathbb{N}_0 : \sigma^{-1} \circ \tau_1 \circ \dots \circ \tau_k = \pi^{-1}, \tau_i \in \text{Nb}(S_n) \forall i\}$, dabei wird die leere Komposition, d.h. der Fall $k = 0$, als Identität aufgefasst.

Bei [HŠS99, S. 126] findet sich, übertragen in unsere Notation, der Zusammenhang

$$T = \frac{n(n-1)}{4}(\tau + 1),$$

wobei $T = \binom{n}{2} - d_{\tau,1}(\sigma, \pi)$ ist und für ein dort gegebenes τ . Dagegen wird bei [Háj69, S. 122] für dasselbe τ (der Unterschied beim Fall $i = j$ ist irrelevant)

$$K = \frac{n(n-1)}{4}(\tau + 1)$$

angegeben, mit $K = \binom{n}{2} - d_{\tau,2}(\sigma, \pi)$, so dass sich sofort $d_{\tau,1}(\sigma, \pi) = d_{\tau,2}(\sigma, \pi)$ ergibt. Die Gültigkeit von $d_{\tau,1}(\sigma, \pi) = d_{\tau,3}(\sigma, \pi)$ wurde von Kendall in [Ken70] auf den Seiten 5–8

formuliert und auf den Seiten 24–26 bewiesen. Schließlich haben sich die Autoren in [DG77] auf S. 264 mit der Gleichheit $d_{\tau,2}(\sigma,\pi) = d_{\tau,3}(\sigma,\pi)$ beschäftigt.

Wir verwenden deshalb in dieser Arbeit für

$$d_{\tau}(\sigma,\pi) := d_{\tau,1}(\sigma,\pi) = d_{\tau,2}(\sigma,\pi) = d_{\tau,3}(\sigma,\pi)$$

die Bezeichnung Kendalls Tau-Abstand und benutzen die für die jeweilige Verwendung angenehmste der obigen drei Berechnungsformen.

B. Hilfsaussagen zu Maximum-Likelihood-Schätzern

Bekanntlich ist die Grenzvarianz der Maximum-Likelihood-Schätzer unter bestimmten Regularitätsbedingungen gleich dem Inversen der Fisher-Informations-Matrix $I = (I_{ij})$. Da letztere für einige Verteilungen in den Standardquellen gar nicht oder fehlerhaft angegeben ist, erfolgt die Herleitung hier. Weiterhin stellen wir Eigenschaften der ML-Schätzer bereit, die bei Simulationen nützlich sind. Unter gewissen, bei Gamma- und inverser Gauß-Verteilung erfüllten Bedingungen gilt nach [LC98, S. 125] für eine Zufallsvariable X mit vom Parameter θ abhängender Dichte $f(x|\theta)$

$$I_{ij} = -\mathbb{E} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log(f(X|\theta)) \right).$$

Im Fall der Gammaverteilung erhalten wir mit $\theta = (\alpha, \lambda)'$

$$\log(f(x|\alpha, \lambda)) = \alpha \log(\lambda) - \log(\Gamma(\alpha)) + (\alpha - 1) \log(x) - \lambda x,$$

so dass sich

$$\begin{aligned} \frac{\partial}{\partial \alpha} \log(f(x|\alpha, \lambda)) &= \log(\lambda) - \Psi(\alpha) + \log(x), & \frac{\partial}{\partial \lambda} \log(f(x|\alpha, \lambda)) &= \frac{\alpha}{\lambda} - x, \\ \frac{\partial^2}{\partial \alpha^2} \log(f(x|\alpha, \lambda)) &= -\Psi'(\alpha), & \frac{\partial^2}{\partial \alpha \partial \lambda} \log(f(x|\alpha, \lambda)) &= \frac{1}{\lambda}, & \frac{\partial^2}{\partial \lambda^2} \log(f(x|\alpha, \lambda)) &= \frac{-\alpha}{\lambda^2} \end{aligned}$$

ergibt. Damit folgt

$$I_{\Gamma}(\alpha, \lambda) = \begin{pmatrix} \Psi'(\alpha) & \frac{-1}{\lambda} \\ \frac{-1}{\lambda} & \frac{\alpha}{\lambda^2} \end{pmatrix}, \quad I_{\Gamma}^{-1}(\alpha, \lambda) = \frac{\lambda^2}{\alpha \cdot \Psi'(\alpha) - 1} \begin{pmatrix} \frac{\alpha}{\lambda^2} & \frac{1}{\lambda} \\ \frac{1}{\lambda} & \Psi'(\alpha) \end{pmatrix}. \quad (\text{B.1})$$

Bekanntermaßen gibt es keinen exakten geschlossenen Ausdruck für die ML-Schätzer $\hat{\alpha}_n^{\text{ML}}$ und $\hat{\lambda}_n^{\text{ML}}$ einer Gammaverteilung. Trotzdem kann das folgende Resultat aus der Likelihood-Funktion abgeleitet werden.

Lemma B.1. *Es gilt: $\hat{\alpha}_n^{\text{ML}}$ ist skalierungsinvariant und $\hat{\lambda}_n^{\text{ML}}$ ist homogen vom Grad -1 .*

Beweis. Wir benutzen, dass für die Dichte $f_{\alpha,\lambda}$ einer Gammaverteilung mit Parametern α und λ für $a \in \mathbb{R}^+$ gilt $f_{\alpha,a\lambda}(x) = a \cdot f_{\alpha,\lambda}(ax)$. Damit erhalten wir für die Likelihoodfunktion l

$$\begin{aligned} l((\alpha,\lambda)|(ax_1, \dots, ax_n)) &= \prod_{i=1}^n \frac{1}{a} \cdot a \cdot f_{\alpha,\lambda}(ax_i) = \left(\frac{1}{a}\right)^n \prod_{i=1}^n f_{\alpha,a\lambda}(x_i) \\ &= \left(\frac{1}{a}\right)^n l((\alpha,a\lambda)|(x_1, \dots, x_n)). \end{aligned}$$

Beide betrachteten Likelihood-Funktionen haben ihr Maximum an derselben Stelle und die Behauptung folgt. \square

Bei der inversen Gauß-Verteilung gilt

$$\log(f(x|\mu,\lambda)) = \frac{1}{2} \log(\lambda) - \frac{1}{2} \log(2\pi x^3) - \frac{\lambda(x - \mu)^2}{2\mu^2 x}$$

und durch die entsprechenden Ableitungen oder aus den Formeln bei [CF89, S. 103] ergibt sich

$$I_{\text{IG}}(\mu,\lambda) = \begin{pmatrix} \frac{\lambda}{\mu^3} & 0 \\ 0 & \frac{1}{2\lambda^2} \end{pmatrix}, \quad I_{\text{IG}}^{-1}(\mu,\lambda) = \begin{pmatrix} \frac{\mu^3}{\lambda} & 0 \\ 0 & 2\lambda^2 \end{pmatrix}. \quad (\text{B.2})$$

Weiter ist nach [CF89, S. 55]

$$\hat{\mu}_n^{\text{ML}}(X_1, \dots, X_n) = \bar{X}_{n,1}, \quad \hat{\lambda}_n^{\text{ML}}(X_1, \dots, X_n) = \frac{1}{\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{X_i} - \frac{1}{\hat{\mu}_{\text{ML}}}\right)}.$$

Das folgende Lemma kann sofort durch Einsetzen in die Gleichungen der Schätzer bewiesen werden.

Lemma B.2. *Die Schätzer $\hat{\mu}_n^{\text{ML}}$ und $\hat{\lambda}_n^{\text{ML}}$ sind skalierungsäquivalent.*

C. R-Quellcode

Wir geben in diesem Anhang beispielhaft den R-Code an, mit dem die für Beispiel 3.3.32 benötigten Größen bestimmt wurden. Aus Platzgründen verzichten wir darauf, dies auch bei den anderen Simulationen zu tun.

In Zeile 1 wird das Paket `BB` geladen, das die zur Lösung eines Systems von nicht-linearen Gleichungen benötigte Funktion `BBsolve` enthält. Anschließend werden in den Zeilen 2–7 die Anzahl `n` der Geburts- und Todesdaten, die Anzahl `wdhs` der Wiederholungen und die Parameter `theta`, `alpha` und `lambda` des zugrundeliegenden Poisson-Prozesses bzw. der Gammaverteilung sowie der Startwert („seed“) des Zufallszahlengenerators definiert.

```
1 library("BB")
2 n <- 1*10^4
3 wdhs <- 10^3
4 theta <- 0.5
5 alpha <- 2
6 lambda <- 1
7 set.seed(18101979)
```

Die in den Zeilen 8–13 definierte Funktion `vectorize` ist ein Hilfsmittel, das in bestimmten Fällen die Anwendbarkeit der Funktion `integrate` ermöglicht.

```
8 vectorize <- function (f){
9   function(x,...) {
10     result <- rep(f(x[1],...),length(x))
11     for (i in 2:length(x)) result[i] <- f(x[i],...)
12     result
13   } }
```

Die Funktion `alFkt` stellt `alpha` in Abhängigkeit von `lambda` und dem Erwartungswert $\exp(\text{theta}*\text{alpha}/\text{lambda})/\text{theta}$ der Betriebsdauer dar. Dies ermöglicht es, weiter unten in einigen Fällen nur die Nullstelle bezüglich eines Parameters zu bestimmen, was den Zeitaufwand verringert. Die Funktionen in den Zeilen 15–22 dienen der Berechnung des für den Hall-Schätzer benötigten ersten und zweiten Moments der Betriebsdauer. In den Arrays in den Zeilen 23 und 24 werden die Schätzer für `alpha` und `lambda` aus den einzelnen Durchläufen gespeichert. Die Objekte in den Zeilen 25 und 26 schließlich sollen Kenngrößen des Hall-Schätzers aufnehmen; nämlich, wie viele Betriebsperioden („Blöcke“) es im jeweiligen Durchgang gab, und ob das Verfahren funktioniert hat.

```

14 alFkt <- function(x,l){ l*log(theta*x+1)/theta }
15 innerIntegrand <- function(u,para){
16     1-pgamma(u,para[1],rate=para[2]) }
17 outerIntegrand <- function(u,para){
18     exp(theta*integrate(innerIntegrand,u,Inf,para=para)$value)-1}
19 Mom1Fkt <- function(para){ (exp(theta*para[1]/para[2])-1)/theta }
20 Mom2Fkt <- function(para){
21     2/theta*exp(theta*para[1]/para[2])
22     *integrate(vectorize(outerIntegrand),0,Inf,para=para)$value }
23 als_H <- array(dim=wdhs)
24 las_H <- array(dim=wdhs)
25 anzBls_H <- array(dim=wdhs)
26 hatFkt_H <- 0

```

In den Zeilen 27–33 werden die entsprechenden Definitionen für den Reynolds-Schätzer vorgenommen. Der Array `anzPkts_Rey` wird später die Anzahl der Punkte speichern, an denen die Schlangenlänge in den einzelnen Durchläufen ausgewertet wird.

```

27 integrand <- function(u,para){ 1-pgamma(u,para[1],rate=para[2]) }
28 gSternQuer <- function(h,para){
29     1-para[2]/para[1] * integrate(integrand,0,h,para=para)$value }
30 h <- 0.5
31 als_Rey <- array(dim=wdhs)
32 las_Rey <- array(dim=wdhs)
33 anzPkts_Rey <- array(dim=wdhs)

```

Die noch notwendigen Definitionen für den konsistenten Schätzer erfolgen in den Zeilen 34–43. Bei unseren Simulationen wurde der in Tabelle 3.5 dargestellte Vergleich simultan mit den Untersuchungen durchgeführt, die zu den Ergebnissen in den Tabellen 3.6 und 3.7 führten. Deswegen treten alle Kenngrößen einmal mit Index K und einmal mit Index K2 auf.

```

34 alFkt_K <- function(x,l){ x*l }
35 als_K <- array(dim=wdhs)
36 las_K <- array(dim=wdhs)
37 anzPkts_K <- array(dim=wdhs)
38 hatFkt_K <- 0
39
40 als_K2 <- array(dim=wdhs)
41 las_K2 <- array(dim=wdhs)
42 anzPkts_K2 <- array(dim=wdhs)
43 hatFkt_K2 <- 0

```

Es folgt die FOR-Schleife im Hauptprogramm. In jedem Durchlauf werden zufällige Geburtstage `x1` und Todesdaten `x2_un` sowie die Ordnungsstatistiken `x2` erzeugt.

```

44 for (maini in 1:wdhs){
45     x1 <- cumsum(rexp(n) / theta)

```

```

46  shift1 <- rgamma(n, alpha , rate=lambda)
47  x2_un <- x1 + shift1
48  x2 <- sort(x2_un)

```

In den Zeilen 49–67 wird der Hall-Schätzer berechnet. Zunächst ermittelt das Hilfsprogramm „Blocklaenge.R“, wie viele Geburts- und Todesdaten die einzelnen Betriebsperioden enthalten, und deren Anzahl `anzBl`. Basierend auf diesen Information werden in den folgenden Zeilen die zugehörigen Betriebsdauern bestimmt und im Array `dauer_Bl` gespeichert. Nun können mithilfe der Funktion `uniroot` die Momentenschätzer für `alpha` und `lambda` berechnet werden, falls die Zielfunktion eine Nullstelle hat.

```

49  source("Blocklaenge.R")
50  dauer_Bl <- array(dim=anzBl)
51  unten <- 0
52  for (i in 1:anzBl){
53    dauer_Bl[i] <- x2[unten+laenge_Bl[i]]-x1[unten+1]
54    unten <- unten+laenge_Bl[i] }
55
56  hatMom1 <- mean(dauer_Bl)
57  hatMom2 <- mean(dauer_Bl^2)
58  f <- function(x) {
59    if (x>0) { Mom2Fkt(c(alFkt(hatMom1,x),x))-hatMom2 } else { Inf } }
60
61  if (f(0.5)*f(3)<0) {
62    hatFkt_H <- hatFkt_H+1
63    ausg <- uniroot(f, c(0.5,3))
64
65    als_H[maini] <- alFkt(hatMom1, ausg$root)
66    las_H[maini] <- ausg$root
67    anzBl_H[maini] <- anzBl }

```

Für den Reynolds-Schätzer werten wir die Schlangenlänge an einer bestimmten, von `h` und dem letzten Todeszeitpunkt abhängenden Anzahl `anzPunkte` an Punkten aus und speichern sie im Array `que` (Z. 68–76). Anhand dieser Werte werden nun die Korrelationskoeffizienten $\rho(h)$ und $\rho(2h)$ wie bei [BP99] beschrieben geschätzt (Z. 78-85) und schließlich in den Zeilen 87–94 unter Anwendung der Funktion `BBsolve` zum Schätzen von `alpha` und `lambda` benutzt, danach erfolgt die Speicherung der Resultate.

```

68  anzPunkte <- floor(x2[n]*0.95)/h
69  que <- array(dim=anzPunkte)
70
71  lauf_x1 <- 1
72  lauf_x2a <- 1
73  for (i in 1:anzPunkte){
74    while (x1[lauf_x1]<=i*h) { lauf_x1 <- lauf_x1+1 }
75    while (x2[lauf_x2a]<=i*h) { lauf_x2a <- lauf_x2a+1 }
76    que[i] <- lauf_x1 - lauf_x2a }

```

```

77
78 r_m <- function(j) {
79   queQuer <- mean(que)
80   zaehler <- sum((que[1:(anzPunkte-j)]-queQuer)
81                 *(que[(j+1):anzPunkte]-queQuer))/anzPunkte
82   nenner <- sum((que-queQuer)^2)/anzPunkte
83   zaehler/nenner }
84 erg1_m <- r_m(1)
85 erg2_m <- r_m(2)
86
87 f_m <- function(x) {
88   if ((x[1]>0) && (x[2]>0)) {
89     c(gSternQuer(h,x)-erg1_m,gSternQuer(2*h,x)-erg2_m)
90   } else { c(Inf,Inf) } }
91 ausg_m <- BBSolve(par=c(2,1),fn=f_m)
92 als_Rey[maini] <- ausg_m$par[1]
93 las_Rey[maini] <- ausg_m$par[2]
94 anzPkts_Rey[maini] <- anzPunkte

```

Die Vorgehensweise bei dem konsistenten Schätzer mit zwei verschiedenen Werten von h ist nun analog wie bei den beiden anderen Verfahren.

```

95 rho1 <- erg1_m
96 hatErw <- mean(x2)-mean(x1)
97
98 f <- function(x) {
99   if (x>0) { gSternQuer(h,c(alFkt_K(hatErw,x),x))-rho1 } else { Inf } }
100
101 if (f(0.25)*f(3)<0){
102   hatFkt_K <- hatFkt_K+1
103   ausg <- uniroot(f,c(0.25,3))
104   als_K[maini] <- alFkt_K(hatErw,ausg$root)
105   las_K[maini] <- ausg$root
106   anzPkts_K[maini] <- anzPunkte }
107
108 rho5 <- r_m(5)
109 f <- function(x){
110   if (x>0) { gSternQuer(5*h,c(alFkt_K(hatErw,x),x))-rho5 } else { Inf } }
111
112 if (f(0.25)*f(3)<0){
113   hatFkt_K2 <- hatFkt_K2+1
114   ausg <- uniroot(f,c(0.25,3))
115   als_K2[maini] <- alFkt_K(hatErw,ausg$root)
116   las_K2[maini] <- ausg$root
117   anzPkts_K2[maini] <- anzPunkte }
118 }

```

Damit ist das Hauptprogramm beendet. Abschließend geben wir noch den Code des Hilfsprogramms „Blocklaenge.R“ an. Zunächst werden der Array `wo_Ende`, in dem die Indizes der Todesdaten gespeichert werden, nach denen eine Betriebsperiode beendet ist, sowie die Zähler `anzB1` (für die Anzahl der Betriebsperioden) und `anzKlei` initialisiert (Z. 1–3). Anschließend ist für jeden Todeszeitpunkt der nächstkleinere Geburtstag zu bestimmen (Z. 6). Hat dieser den gleichen Index, ist nach Bemerkung 2.3.3 (a) das Ende einer Betriebsperiode erreicht; der Zähler wird erhöht (Z. 8) und der Index gespeichert (Z. 9). In den Zeilen 12–14 schließlich wird basierend auf den gespeicherten Indizes ermittelt, wie viele Geburts- und Todesdaten die jeweiligen Betriebsperioden enthalten, und dies im Array `laenge_B1` vermerkt.

```

1 wo_Ende <- array(dim=n)
2 anzB1 <- 0
3 anzKlei <- 1
4
5 for (i in 1:n){
6   while ( (x2[i]>x1[anzKlei]) && (anzKlei<=n) ) { anzKlei <- anzKlei+1 }
7   if (anzKlei-1==i) {
8     anzB1 <- anzB1+1
9     wo_Ende[anzB1] <- i
10  } }
11
12 length(wo_Ende) <- anzB1
13 laenge_B1 <- array(dim=anzB1)
14 laenge_B1 <- wo_Ende - c(0,wo_Ende[1:(anzB1-1)])

```


Literaturverzeichnis

- [Asm03] Søren Asmussen. *Applied Probability and Queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [BD97] Nick H. Bingham and Bruce Dunham. Estimating diffusion coefficients from count data: Einstein-Smoluchowski theory revisited. *Ann. Inst. Statist. Math.*, 49(4):667–679, 1997.
- [BG05] Nicole Bäuerle and Rudolf Grübel. Multivariate counting processes: copulas and beyond. *Astin Bull.*, 35(2):379–408, 2005.
- [Bil95] Patrick Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
- [BP99] Nick H. Bingham and Susan M. Pitts. Non-parametric estimation for the $M/G/\infty$ queue. *Ann. Inst. Statist. Math.*, 51(1):71–97, 1999.
- [Bur02] Rainer E. Burkard. Selected topics on assignment problems. *Discrete Appl. Math.*, 123(1-3):257–302, 2002.
- [CB90] George Casella and Roger L. Berger. *Statistical inference*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.
- [CF89] Raj S. Chhikara and J. Leroy Folks. *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Marcel Dekker, Inc., New York, NY, USA, 1989.
- [DFG71] Morris H. DeGroot, Paul I. Feder, and Prem K. Goel. Matchmaking. *Ann. Math. Statist.*, 42:578–593, 1971.
- [DG77] Persi Diaconis and Ronald L. Graham. Spearman’s footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B*, 39(2):262–268, 1977.
- [DG80] Morris H. DeGroot and Prem K. Goel. Estimation of the correlation coefficient from a broken random sample. *Ann. Statist.*, 8(2):264–278, 1980.

- [Dia88] Persi Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
- [Gha86] Saeed Ghahramani. Finiteness of moments of partial busy periods for $M/G/C$ queues. *J. Appl. Probab.*, 23(1):261–264, 1986.
- [Gri58] Harold D. Griffin. Graphic computation of tau as a coefficient of disarray. *J. Amer. Statist. Assoc.*, 53:441–447, 1958.
- [Háj69] Jaroslav Hájek. *A Course in Nonparametric Statistics*. Holden-Day, San Francisco, Calif., 1969.
- [Hal88] Peter Hall. *Introduction to the Theory of Coverage Processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1988.
- [Han04] Ben B. Hansen. Full matching in an observational study of coaching for the SAT. *J. Amer. Statist. Assoc.*, 99(467):609–618, 2004.
- [HŠS99] Jaroslav Hájek, Zbyněk Šidák, and Pranab K. Sen. *Theory of Rank Tests*. Probability and Mathematical Statistics. Academic Press Inc., San Diego, CA, second edition, 1999.
- [JKB94] Norman L. Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous Univariate Distributions. Vol. 1*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1994. A Wiley-Interscience Publication.
- [Jon98] Geurt Jongbloed. Exponential deconvolution: two asymptotically equivalent estimators. *Statist. Neerlandica*, 52(1):6–17, 1998.
- [Jun99] Dieter Jungnickel. *Graphs, Networks and Algorithms*, volume 5 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 1999. Translated from the third German (1994) edition by Tilla Schade and revised by the author.
- [Ken70] Maurice G. Kendall. *Rank Correlation Methods*. A Charles Griffin Title. London, forth edition, 1970.
- [LC98] Erich L. Lehmann and George Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.
- [MS02] Alfred Müller and Dietrich Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2002.

- [R D09] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. ISBN 3-900051-07-0.
- [Sho72] Galen R. Shorack. The best test of exponentiality against gamma alternatives. *J. Amer. Statist. Assoc.*, 67:213–214, 1972.
- [Sor80] Harold W. Sorenson. *Parameter Estimation*, volume 9 of *Control and Systems Theory*. Marcel Dekker Inc., New York, 1980. Principles and problems.
- [Sto83] Dietrich Stoyan. *Comparison methods for queues and other stochastic models*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1983. Translation from the German edited by Daryl J. Daley.
- [Tad81] Pandu R. Tadikamalla. The inverse Gaussian approximation to the lead time demand. *International Journal of Production Research*, 19(2):213–219, 1981.
- [Wei99] Eric W. Weisstein. *CRC Concise Encyclopedia of Mathematics*. CRC Press, Boca Raton, FL, 1999.

Danksagung

Mein besonderer Dank gilt meinem Doktorvater Herrn Prof. Dr. Grübel für die intensive Betreuung und sein konstant hohes Interesse an meiner Arbeit.

Auch allen anderen Mitgliedern des Instituts für Mathematische Stochastik, insbesondere Herrn Prof. Dr. Baringhaus, sei an dieser Stelle für ihre Hilfsbereitschaft und die angenehme Arbeitsatmosphäre gedankt.

Außerdem möchte ich Frau Prof. Dr. Bäuerle für die Übernahme des Koreferats sehr herzlich danken.

Lebenslauf

Persönliche Daten

| | |
|---------------------|------------------|
| Name | Wegener |
| Vorname | Hendrik |
| Geburtstag | 18. Oktober 1979 |
| Geburtsort | Lübbecke |
| Familienstand | ledig |
| Staatsangehörigkeit | deutsch |

Bildungsweg

| | |
|-------------|---|
| 1986 – 1990 | Grundschule Petershagen |
| 1990 – 1999 | Städt. Gymnasium Petershagen |
| 2000 – 2006 | TU Clausthal Studium der Wirtschaftsmathematik |

Praktika

| | |
|------|--|
| 2003 | Egyptian Iron&Steel Co., Kairo (Ägypten) |
| 2005 | Volvo Car Corp., Skövde (Schweden) |

Berufliche Tätigkeiten

| | |
|-------------|---|
| 1999 – 2000 | Zivildienst im Diakoniezentrum Bethesda, Lahde |
| 2002 – 2005 | Wiss. Hilfskraft, TU Clausthal |
| seit 2006 | Wiss. Mitarbeiter, Leibniz Universität Hannover |
| 2007 – 2008 | Lehrauftrag „Statistik“, FH Hannover |