



**h\_da**

HOCHSCHULE DARMSTADT  
UNIVERSITY OF APPLIED SCIENCES

**fbmn**

FACHBEREICH MATHEMATIK  
UND NATURWISSENSCHAFTEN



**h\_da**

HOCHSCHULE DARMSTADT  
UNIVERSITY OF APPLIED SCIENCES

**fbi**

FACHBEREICH INFORMATIK

## **Hochschule Darmstadt**

Fachbereiche Mathematik und Naturwissenschaften  
&  
Informatik

---

# **Progressionsfreies Überleben in onkologischen Studien: Verzerrung durch Intervallzensurierung in Ereigniszeitdaten und Methodiken zur Biaskorrektur in unkontrollierten Studien**

---

Abschlussarbeit zur Erlangung des akademischen Grades  
Master of Science (M. Sc.)  
im Studiengang Data Science

vorgelegt von  
Robert Miltenberger

Referentin: Prof. Dr. Antje Jahn (Hochschule Darmstadt)  
Korreferent: Dr. Heiko Götte (Merck KGaA)  
Ausgabedatum: 15.05.2019  
Abgabedatum: 16.08.2019

## **Selbstständigkeitserklärung**

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht. Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen. Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

---

Ort, Datum

---

Unterschrift

# Zusammenfassung

Für die Zulassung neuer Medikamente ist es unerlässlich deren Wirksamkeit im Vergleich zu aktuell verwendeten Behandlungsmethoden zu testen, denn sowohl für die Pharmaunternehmen als auch für die Erkrankten sind neue Medikamente nur dann sinnvoll, wenn eine Wirkungssteigerung vorliegt.

Ein häufig verwendeter Endpunkt für onkologische Studien ist das *Progressionsfreie Überleben*. Dabei können bei Patienten eine Progression oder der Tod als Ereignisse beobachtet werden. Progressionen lassen sich allerdings nur an Visiterterminen festgestellt und dadurch unterschieden sich die wahre und die beobachtete Ereigniszeit.

Der Standard für den Vergleich eines neuen mit einem etablierten Medikament ist eine randomisierte klinische Studie. Häufig werden allerdings aus ethischen oder ökonomischen Gründen unkontrollierte Studien durchgeführt. Eine unkontrollierte Studie kann sich bezüglich des Studienaufbaus von Studien unterscheiden, die in der Vergangenheit mit dem etablierten Medikament durchgeführt wurden. In beiden Fällen entsteht eine Verzerrung (Bias), allerdings wurden kontrollierte Studien nach dem gleichen Visitenplan aufgebaut und unterliegen daher derselben systematischen Verzerrung. Dies gilt für unkontrollierte Studien nicht, denn diese sind nicht zwingend nach dem selben Visitenplan aufgebaut. Daher werden in der vorliegenden Arbeit statistische Methoden entwickelt, mit denen auch in unkontrollierten Studien ein Vergleich der Ergebnisse ermöglicht wird.

Zu Beginn dieser Arbeit wird der grundsätzliche Aufbau klinischer Studien beschrieben. Anschließend werden die benötigten Grundbegriffe und Funktionen der Ereigniszeitanalyse hergeleitet und erläutert.

Alle Methoden werden an Simulationsstudien evaluiert. Deren Aufbau wird in Kapitel 4 erläutert. Des Weiteren wird in diesem Kapitel genauer darauf eingegangen, warum die Ereigniszeiten unter Verwendung des *Progressionsfreien Überlebens* fehlerhaft beobachtet werden. Die Ereigniszeiten benötigt man für die Schätzung mit Hilfe der *Maximum-Likelihood*-Methode und aufgrund der Fehlbeobachtung resultiert der Bias.

In den Kapiteln 5 und 6 werden zwei Methoden vorgestellt, welche den Bias korrigieren sollen, wobei der in Kapitel 5 erklärte Ansatz in zwei verschiedenen Varianten angewandt wurde.

Bei diesen handelt es sich um Modifizierungen der bekannten *Maximum a posteriori*-Methode und damit um einen klassischen bayesianischen Ansatz. Das herkömmliche Verfahren konnte aufgrund der fehlerhaft beobachteten Ereigniszeiten nicht verwendet werden, da diese Methode der gleichen Verzerrung wie die *Maximum-Likelihood*-Methode unterliegen würde.

Simulationsstudien zeigen, dass beide Varianten des Verfahrens zwar zu einer Reduktion, aber noch nicht zu einer vollständigen Korrektur des Bias führen.

Daher wird in Kapitel 6 mit der *Approximate Bayesian Computation*-Methode eine sogenannte *likelihood freie*-Methodik verwendet. Dabei handelt es sich ebenfalls um eine bayesianische Methode, diese umgeht allerdings eine analytische Auswertung der Likelihoodfunktion und schätzt den Parameter mit Hilfe von Simulationen.

Mit Hilfe dieser Methode konnte der Bias weitgehend korrigiert werden, sodass ein sinnvoller Vergleich von unkontrollierten Studien ermöglicht wird.

**Schlagwörter:** Biaskorrektur, bayesianische Statistik, approximate bayesian computation, klinische Studien, Ereigniszeitanalyse

## Abstract

In clinical trials the comparison of the results of different studies is indispensable. It is used to find out which medicine has a higher impact on helping the illness.

*Progression-free Survival* is often used as an endpoint of oncological trials. In this case, the events of progression and death can be observed in patients. Progressions can only be identified on visit appointments and thus the true and the observed event time differ. The standard for comparing a new with an established drug is a randomized clinical trial. Often, however, uncontrolled studies are conducted for ethical or economic reasons. An uncontrolled study may differ in study design from studies that have been conducted with the established drug in the past. In both cases this results in a bias but controlled studies were planned with the same assessment schedule and therefore the bias is identical in both studies. Uncontrolled trials are usually not planned with the same assessment schedule. Therefore, the thesis develops statistical methods that allow a comparison of uncontrolled studies.

At the beginning the general construction of a clinical trial will be explained. Afterwards the required basic concepts and functions of survival analysis are derived and explained. All methods will be evaluated with simulated studies. The construction of these will be shown in chapter 4. Furthermore it is explained why using the *Progression-free Survival* as endpoint leads to misobserved event times. These event times are needed for the *Maximum-Likelihood*-estimation and due the misobservation this results in a bias.

In chapter 5 and 6 methods are presented, which shall correct the bias. The first one is a classical bayesian approach and is used in two different implementations. Both are modifications of the well known *Maximum a posteriori*-method. This basic approach could not been used, because it would lead to the same bias as the *Maximum-Likelihood*-method.

The evaluation with simulated studies shows, both implementations of this method lead to a reduction but not to a fully correction of the bias.

Therefore, in chapter 6 the *Approximate Bayesian Computation*-method is introduced. This is a so called *likelihood free*-method, which means this approach does not need an analytical evaluation of the likelihood. The estimation is based on simulations. This approach results in a nearly perfect correction of the bias and so a comparison of uncontrolled studies is possible.

**Keywords:** bias correction, bayesian statistic, approximate bayesian computation, clinical trials, survival analysis

## **Danksagung**

Zuerst möchte ich meiner Familie danken, dass sie mich trotz aller Widrigkeiten immer so gut wie möglich unterstützt hat.

Ein weiterer Dank geht an meine Lebensgefährtin Daliah Dieckmann, die mir während meines gesamten Masterstudiums mit Rat und Tat zur Seite stand, auch wenn ich eventuell mal wieder etwas zu ehrgeizig war.

An dieser Stelle möchte ich mich auch bei Cyril Martin Czernin, Björn Severitt, Christoph Wies, Hasan Kutlu, Sven Bergmann, Christina Habermehl, Till Körner und Ott Vahtrik nicht nur für das Korrekturlesen bedanken, sondern auch für jedwede Hilfe während des Erstellens der Arbeit.

Ebenfalls nicht zu vergessen sind Frau Prof. Dr. Antje Jahn und Herr Dr. Heiko Götte die mir mit ihrer außerordentlich guten Betreuung das erfolgreiche absolvieren ermöglicht haben.

**Vielen Dank!**

# Inhaltsverzeichnis

<b>1</b>	<b>Aufbau klinischer Studien</b>	<b>3</b>
<b>2</b>	<b>Grundlagen</b>	<b>5</b>
2.1	Die Gammaverteilung . . . . .	5
2.2	Maximum-Likelihood-Schätzer (ML-Schätzer) . . . . .	6
2.3	Bedingte Wahrscheinlichkeit . . . . .	7
2.4	Der Satz von Bayes . . . . .	7
2.5	Bayessche Statistik . . . . .	8
<b>3</b>	<b>Grundbegriffe der Überlebenszeitanalyse</b>	<b>9</b>
3.1	Survivalfunktion . . . . .	9
3.1.1	Survivalfunktion im stetigen Fall . . . . .	9
3.1.2	Survivalfunktion im diskreten Fall . . . . .	10
3.2	Hazardfunktion . . . . .	10
3.2.1	Hazardfunktion im stetigen Fall . . . . .	11
3.2.2	Hazardfunktion im diskreten Fall . . . . .	12
3.3	Konkurrierende Ereignisse . . . . .	13
3.4	Likelihoodfunktion für zensierte Daten . . . . .	14
3.5	Kaplan-Meier-Schätzer . . . . .	16
3.6	Mediane Überlebenszeit . . . . .	17
<b>4</b>	<b>Erläuterung und Umsetzung der Problematik</b>	<b>18</b>
4.1	Simulation einer klinischen Studie . . . . .	18
4.2	Entstehung der Verzerrung . . . . .	23
<b>5</b>	<b>Biaskorrektur durch Maximum a posteriori-Ansatz</b>	<b>25</b>
5.1	Herleitung . . . . .	25
5.2	Sensitivitätsanalyse . . . . .	28
<b>6</b>	<b>Biaskorrektur durch ABC-Ansatz</b>	<b>37</b>
6.1	Approximate Bayesian Computation (ABC) . . . . .	37
6.2	Anwendung auf eigentliche Problemstellung . . . . .	41
6.3	Sensitivitätsanalyse . . . . .	43

## Einführung

Für die Zulassung neuer Medikamente geht es in klinischen Studien darum, eine signifikante Verbesserung des neuen Wirkstoffes im Vergleich zur aktuell verwendeten Behandlungsmethode nachzuweisen. Dafür werden gewisse Kennzahlen der Probanden, wie beispielsweise die *Gesamtüberlebenszeit* (engl. *Overall Survival (OS)*) seit Behandlungsbeginn benötigt. Da die Ergebnisse neuer Studien so schnell wie möglich ausgewertet werden sollen, um effektivere Medikamente zuzulassen, wurde der Fokus in den letzten Jahren immer mehr auf das *Progressionsfreie Überleben* (engl. *Progression-free Survival (PFS)*) gelegt. Bei einer Progression handelt es sich um eine Verschlechterung des Gesundheitszustands des Patienten um ein bestimmtes Maß. Bei einer Krebserkrankung kann dies z.B. das Wachsen des Tumors oder ein Auftreten von Metastasen sein. Das PFS dient als Ersatz für das OS und beschreibt den Zeitraum seit Beitritt in die Studie bis zum Eintreten eines Ereignisses. Bei einem Ereignis handelt es sich hierbei entweder um eine Progression oder das Eintreten des Todes. Diese beiden Risiken werden als *Konkurrierende Ereignisse* (engl. *Competing Risks*) bezeichnet.

In den meisten Fällen führt die Verwendung des PFS allerdings zu einer Verzerrung der Ereigniszeiten, da der genaue Zeitpunkt der Progression im Vergleich zum Todeszeitpunkt nicht bekannt ist. Die Verschlechterung wird erst zur Kenntnis genommen, wenn der Patient zum nächsten Kontrollbesuch beim Arzt ist.

Diese Problematik wird in Abbildung 1 verdeutlicht.

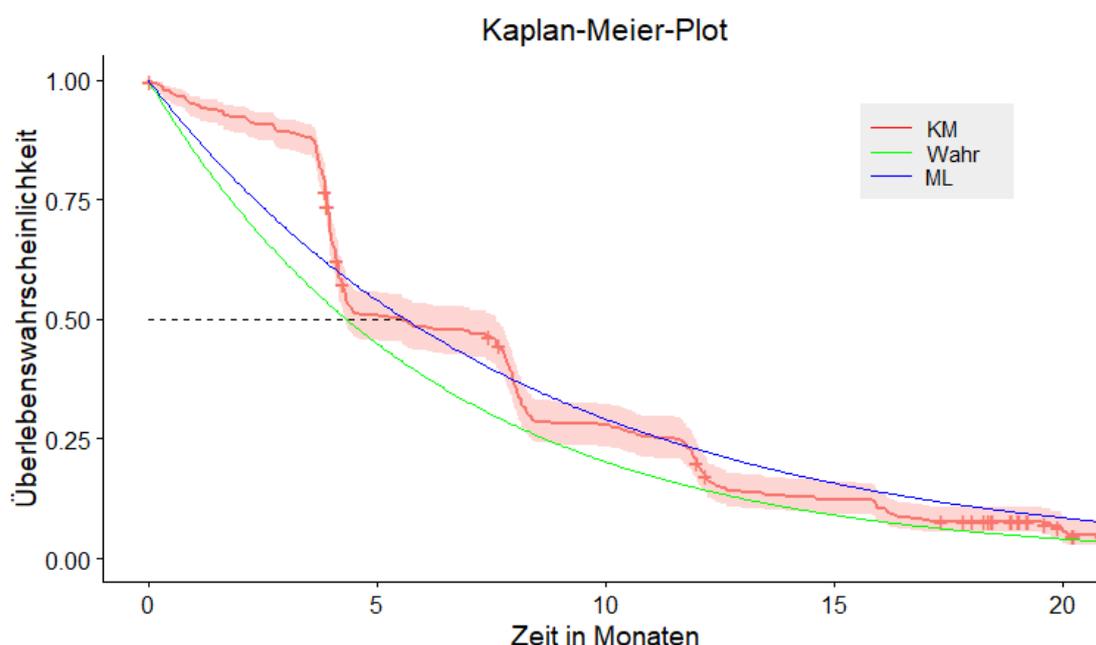


Abbildung 1: Verdeutlichung des Bias durch Intervallzensurierung mit Hilfe des Kaplan-Meier- (KM) und des Maximum-Likelihood-Schätzers (ML)

In diesem Beispiel wurde eine klinische Studie mit konkurrierenden Ereignissen simuliert.

Für diese Studie wurde ein Kontrollintervall von 4 Monaten angenommen. Das bedeutet, wenn ein Patient zwischen zwei Visiten eine Progression erleidet, so wird diese mit bis zu 4 Monate Verspätung beobachtet.

An den Sprungstellen der roten Kurve befinden sich die Visitenzeitpunkte der Patienten, da zu diesen Zeitpunkten die Progressionen beobachtet werden. Kleinere Abfälle der Kurve zwischen den Visitenzeitpunkten sind durch Todesfälle bedingt, denn zu diesen sind die exakten Ereigniszeit bekannt.

Die rote Kurve in Abbildung 1 spiegelt den Kaplan-Meier-Plot samt Konfidenzintervall wider, wobei die Kreuze Zensierungen kennzeichnen. Der Kaplan-Meier-Schätzer ist eine nichtparametrische Schätzung der Survivalfunktion. Darauf wird genauer in Kapitel 3.5 eingegangen.

Die hellgrüne Kurve ist die wahre Überlebensfunktion und die blaue eine Schätzung dieser mit Hilfe der *Maximum-Likelihood*-Methode, welche in Kapitel 2.2 genauer erläutert wird. Die eingezeichnete Survivalfunktion ergibt sich daraus, dass das mit Hilfe der ML-Methode berechnete Schätzer in die Survivalfunktion der Exponentialverteilung eingesetzt wurde.

Die horizontale gestrichelte Linie, bei einer Überlebenswahrscheinlichkeit von 50%, kennzeichnet die mediane Überlebenszeit. Diese wird im Folgenden als Vergleichsstatistik zwischen den Ergebnissen verwendet und wird mit *mPFS* bezeichnet. Für dieses einleitende Beispiel ergibt sich eine Differenz von ca. 1.3 Monaten. Während die wahre mediane Überlebenszeit bei ungefähr 4.3 Monaten liegt, hat die Schätzung einen Wert von  $\approx 5.6$  ergeben.

Damit lässt sich der, bei der Schätzung der blauen Kurve entstehende, systematische Bias deutlich erkennen. Dieser ist durch die verzerrten Beobachtungszeitpunkte begründet, da die Schätzung die Verlagerung der Progressionszeitpunkte nicht berücksichtigt.

Möchte man nun zwei Studien, welche nicht die gleiche Zeitspanne zwischen den Visiten hatten, vergleichen, so ist das nicht möglich, da beide verschiedenen Bias unterliegen. Würden beide derselben systematischen Verzerrung unterliegen, so wären sie vergleichbar.

Werden gleichzeitig zwei Studien angelegt, wobei die eine Patientengruppe das neue und die andere das aktuell verwendete Präparat erhält, so nennt man diesen Versuchsaufbau *kontrollierte Studien*. Dabei werden beide Studien identisch und parallel durchgeführt.

Möchte man nun die Ergebnisse eines neuen Medikaments mit den Resultaten einer bereits vergangenen Studie vergleichen, so spricht man dabei von unkontrollierten Studien. Da der Ablauf bei unkontrollierten Studien in der Regel nicht identisch ist, können die Ergebnisse nicht miteinander verglichen werden.

Diese Arbeit behandelt verschiedene Methodiken um diese systematische Verzerrung zu korrigieren und damit einen sinnvollen Vergleich unkontrollierter Studien zu ermöglichen.

Sämtliche Programmcodes befinden sich auf der beigelegten CD. Das in Kapitel 6.3 verwendete Programm ist zudem im Anhang dieser Arbeit zu finden.

# 1 Aufbau klinischer Studien

Der Ablauf klinischer Studien folgt einem strikten Schema, dem sogenannten Studienprotokoll. In diesem werden vor Beginn der Studie die genauen Bedingungen zur Studienteilnahme und deren exakte Durchführung festgehalten. Die Teilnahmebedingungen werden benötigt, damit es sich bei den Patienten um eine homogene Gruppe handelt. Dadurch wird die Gefahr verringert, mögliche Unterschiede in den Behandlungseffekten nur aufgrund gesundheitlicher Differenzen der Probanden zu beobachten. Die exakte Durchführung der Studie enthält Details, wie die Anzahl der Teilnehmer, die maximale Studiendauer, den Zeitraum in welchem nach Beginn der Studie noch Probanden aufgenommen werden, die Zeitspanne zwischen den Visiten, die maximale Anzahl an Visiten, eine Maximalanzahl an beobachteten Ereignissen usw..

Der grundlegende Aufbau einer klinischen Studie ist in Abbildung 2 dargestellt.

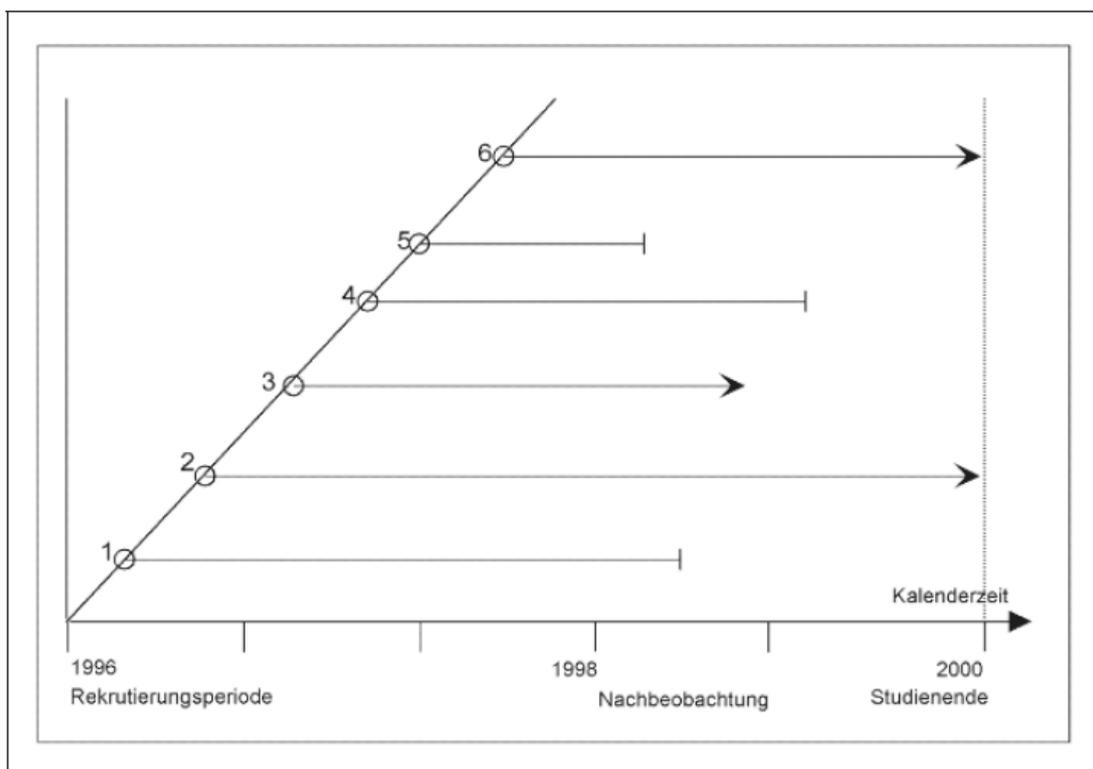


Abbildung 2: Schematische Darstellung eines typischen Ablaufs einer klinischen Studie: (o) Zeitpunkt der Aufnahme in die Studie, (-|) Todeszeitpunkt, (->) Zeitpunkt des letzten Kontakts (zensierte Beobachtung). [SCH-08]

Es wird eine klinische Studie dargestellt, welche am 1.1.1996 begann und bis zum 31.12.1998 Probanden aufnahm (*recruitment time*). Im Anschluss an die Rekrutierungsperiode standen alle Patienten zwei weitere Jahre unter Beobachtung (*follow-up time*), bis die Studie am 31.12.2000 beendet wurde.

Patient 1 ist beispielsweise direkt zu Beginn der Studie beigetreten und ca. Mitte 1999, aufgrund tödlicher Folgen seiner Krankheit, ausgeschieden. Deshalb endet seine Linie mit „-|“. Patient 2 hingegen hat sogar das Ende der Nachbeobachtungszeit erreicht und darüber hinaus noch überlebt. Darum endet seine Linie mit „->“ zum 31.12.2000. Somit ist die Überlebenszeit von Patient 2 die gesamte Zeitspanne, welche er der Studie

angehört hat und sein Status ist eine Zensierung.

Probanden werden allerdings nicht nur zensiert, wenn das Studienende erreicht wird. Eine Zensierung kann etliche Gründe haben und, wie Patient 3 verdeutlicht, jeder Zeit eintreten. Beispielsweise kann die Behandlung abgebrochen oder der Arzt gewechselt werden. In solchen Fällen betrachtet man den letzten Zeitpunkt (*last known alive*), zu dem der Patient sicher noch lebte. Im Fall von Progressionen ist dieser Begriff unpassend. Gibt es keine weiteren Informationen über den Progressionsstatus, so wird der Proband als *lost to follow-up* bezeichnet. Solche Zensierungen nennt man *Drop out*.

Ein Proband mit Zensierung ist bis zum Zensierungszeitpunkt ebenso repräsentativ für die restlichen Studienmitglieder unter Risiko wie ein Proband ohne Zensierung, falls die Überlebenszeit und Zensierung unabhängig voneinander sind. Ebenso werden Individuen mit Zensierung nach dieser durch alle noch unter Risiko befindlichen Individuen repräsentiert. Gilt die Unabhängigkeit, so wird die Zensierung als nicht informativ für das Eintreten des Ereignisses beschrieben.

Anders als in Abbildung 2 muss eine Studie nicht zwingend zu einem vor Studienbeginn vordefinierten Datum enden. Alternativ kann eine bestimmte Anzahl an Ereignissen vorgegeben werden. Die Studie endet, sobald diese Anzahl an Events beobachtet wurde und alle noch unter Risiko stehenden Teilnehmer werden zensiert.

Dieses Vorgehen wird verwendet, da man die Ergebnisse so schnell wie möglich auswerten möchte, sobald genügend Informationen vorliegen. Würde man das eigentliche Ende der Studie abwarten, könnten zwar weitere Ereignisse beobachtet werden, welche allerdings, im Verhältnis zur Wartezeit und den zusätzlichen Kosten, nur wenig neue Informationen liefern.

## 2 Grundlagen

Zu Beginn werden einige, bereits aus dem Studium bekannte, Methoden, welche im Laufe dieser Arbeit ihre Anwendung finden, wiederholt und erläutert. Diese Inhalte sind Bestandteil von Grundlagenmodulen, wie Wahrscheinlichkeitsrechnung oder Statistik.

### 2.1 Die Gammaverteilung

Eine Zufallsvariable  $X$  mit Formparameter  $\alpha > 0$  und Skalenparameter  $\beta > 0$  folgt genau dann einer Gammaverteilung, wenn ihre Dichte durch

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

gegeben ist. Dabei handelt es sich bei  $\Gamma(\alpha)$  um die Gammafunktion. Die Notation einer solchen Zufallsvariablen lautet  $X \sim \text{Gamma}(\alpha, \beta)$ .

Der Erwartungswert und die Varianz der Gammaverteilung lauten

$$E(X) = \frac{\alpha}{\beta} \quad \text{und} \quad \text{Var}(X) = \frac{\alpha}{\beta^2}. \quad (1)$$

*Herleitung:*

$$\begin{aligned} E(X) &= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\beta x} dx \\ &= \int_0^\infty \frac{\alpha \beta^{\alpha+1}}{\alpha \beta \Gamma(\alpha)} x^\alpha e^{-\beta x} dx \\ &= \frac{\alpha}{\beta} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\beta x} dx, \quad \text{da } \alpha \cdot \Gamma(\alpha) = \Gamma(\alpha+1) \\ &= \frac{\alpha}{\beta} \cdot 1 = \frac{\alpha}{\beta}, \end{aligned}$$

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\alpha(\alpha+1)}{\beta^2} \int_0^\infty \frac{\beta^{\alpha+2}}{\Gamma(\alpha+2)} x^{\alpha+1} e^{-\beta x} dx \\ &= \frac{\alpha(\alpha+1)}{\beta^2} \cdot 1 = \frac{\alpha(\alpha+1)}{\beta^2}, \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - E(X)^2 \\ &= \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 \\ &= \frac{\alpha}{\beta^2} \end{aligned}$$

## 2.2 Maximum-Likelihood-Schätzer (ML-Schätzer)

Bei der *Maximum-Likelihood-Methode* wird von einer Zufallsvariablen  $X$  ausgegangen, deren Dichte- bzw. Wahrscheinlichkeitsfunktion  $f_X(x)$  bzw.  $P(X=x)$  zwar bekannt ist, der dazugehörige Parametervektor  $\vec{\beta} = (\beta_1, \dots, \beta_m)^t \in \mathbb{R}^m$  allerdings nicht.

Liegt eine einfache Stichprobe  $\vec{x} = (x_1, \dots, x_n)^t \in \mathbb{R}^n$  von  $n$  unabhängig identisch verteilten Zufallsvariablen  $X_1, \dots, X_n$  vor, so kann  $\vec{\beta}$  mit Hilfe der ML-Methode geschätzt werden. Diese hat das Ziel, unbekannte Parameter so zu schätzen, dass die Wahrscheinlichkeit für die vorliegende Stichprobenrealisation größtmöglich ist, wobei stets  $\vec{\beta} \in \Theta \subseteq \mathbb{R}^m$  gilt. Hierbei stellt  $\Theta$  den Parameterraum der in Frage kommenden Parameter der Verteilung dar.

Im Folgenden wird die Konstruktion der ML-Methode für den stetigen und den diskreten Fall erläutert.

Je nach Verteilungstyp lässt sich die Likelihoodfunktion schreiben als

$$\begin{aligned} \text{diskret: } L(\vec{\beta}; \vec{x}) &= P(X_1 = x_1, \dots, X_n = x_n | \vec{\beta}), \\ \text{stetig: } L(\vec{\beta}; \vec{x}) &= f_X(x_1, \dots, x_n | \vec{\beta}). \end{aligned}$$

Es wird vorausgesetzt, dass es sich um eine einfache Stichprobe handelt. Ist dies der Fall, stimmt für den diskreten Fall die gemeinsame Wahrscheinlichkeit mit dem Produkt der Einzelwahrscheinlichkeiten und im stetigen Fall die Dichtefunktion mit dem Produkt der Randdichtefunktionen überein.

Es gilt dementsprechend:

$$\begin{aligned} \text{diskret: } L(\vec{\beta}; \vec{x}) = P(X_1 = x_1, \dots, X_n = x_n | \vec{\beta}) &= \prod_{i=1}^n P(X_i = x_i | \vec{\beta}), \\ \text{stetig: } L(\vec{\beta}; \vec{x}) = f_X(x_1, \dots, x_n | \vec{\beta}) &= \prod_{i=1}^n f_{X_i}(x_i | \vec{\beta}). \end{aligned}$$

Nun gilt es die Likelihoodfunktion zu maximieren. Um die Funktion leichter differenzieren zu können, verwendet man die Eigenschaften der Logarithmusfunktion. Zum einen wird durch dessen Monotonieeigenschaft die Abzisse des Maximums nicht verändert und zum anderen erlaubt der Logarithmus das Umformen eines Produkts zu einer Summe.

$$\begin{aligned} \text{diskret: } \ln(L(\vec{\beta}; \vec{x})) = \ln\left(\prod_{i=1}^n P(X_i = x_i | \vec{\beta})\right) &= \sum_{i=1}^n \ln(P(X_i = x_i | \vec{\beta})), \\ \text{stetig: } \ln(L(\vec{\beta}; \vec{x})) = \ln\left(\prod_{i=1}^n f_{X_i}(x_i | \vec{\beta})\right) &= \sum_{i=1}^n \ln(f_{X_i}(x_i | \vec{\beta})). \end{aligned}$$

Die daraus resultierende Funktion wird als Log-Likelihoodfunktion bezeichnet und es ist der Parametervektor  $\hat{\beta}$  gesucht, für den gilt:

$$\begin{aligned} \nabla \ln(L(\hat{\beta}, \vec{x})) &= \vec{0} \\ y^t \cdot H_{\ln(L)}(\hat{\beta}, \vec{x}) \cdot y &< 0 \quad \forall y \in \mathbb{R}^m. \end{aligned}$$

Dabei bezeichnet  $\nabla \ln(L(\hat{\beta}, \vec{x}))$  den Gradienten und  $H_{\ln(L)}(\hat{\beta}, \vec{x})$  die Hesse-Matrix der Log-Likelihoodfunktion.

Die erste Ableitung der Log-Likelihood-Funktion wird als Scorefunktion bezeichnet. Möchte man das Nullstellenproblem lösen, so ist dies aufgrund der Komplexität in den meisten Fällen nicht analytisch möglich. Aus diesem Grund behilft man sich mit numerischen Verfahren, wie beispielsweise dem *Newton-Raphson-Verfahren* vgl. *Deuffhard (2004)* [DEU-04].

Mit Hilfe der *Maximum-Likelihood-Methode* erhält man, unter geeigneten Regularitätsbedingungen, einen stark konsistenten Schätzer, das bedeutet mit wachsendem Stichprobenumfang  $n$  konvergiert der ML-Schätzer  $\hat{\beta}_{ML}$  fast sicher gegen den wahren Parametervektor  $\beta$ .

## 2.3 Bedingte Wahrscheinlichkeit

Im Folgenden sei ein Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, P)$  gegeben.

Unter einer bedingten Wahrscheinlichkeit versteht man den Eintritt eines gewissen Ereignisses  $A \in \mathcal{A}$ , unter dem Vorwissen, dass ein anderes Ereignis  $B \in \mathcal{A}$  bereits eingetreten ist. Dies lässt sich wie folgt darstellen und berechnen:

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Hierbei sei vorausgesetzt, dass die Wahrscheinlichkeit des Eintritts von  $B$  echt größer als 0 ist.

Sei  $A_i, i = 1, \dots, N$ , eine Zerlegung der Ereignismenge in paarweise disjunkte Ereignisse, so folgt für die A-posteriori-Wahrscheinlichkeit

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)}. \quad (2)$$

## 2.4 Der Satz von Bayes

Der Satz von Bayes besagt:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (3)$$

*Beweis.*

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{\frac{P(A \cap B)}{P(A)} P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B)} \end{aligned}$$

□

## 2.5 Bayessche Statistik

Die Aussage des *Satzes von Bayes* kann zur Parameterschätzung in der Statistik verwendet werden. Ersetzt man die beiden Ereignisse  $A$  und  $B$  durch einen zu schätzenden Parameter  $\theta$  und einen beobachteten Datensatz  $D$ , so ergibt sich

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}.$$

Dabei stellen die einzelnen Wahrscheinlichkeiten folgende Eigenschaften dar:

$P(\theta|D)$ : A-posteriori-Wahrscheinlichkeit, Wahrscheinlichkeit des Parameters  $\theta$  unter beobachtetem Datensatz  $D$

$P(D|\theta)$ : Likelihood, Wahrscheinlichkeit die Daten  $D$  bei gegebenem Parameter  $\theta$  zu beobachten

$P(\theta)$ : A-priori-Wahrscheinlichkeit, Wahrscheinlichkeit von  $\theta$  basierend auf Vorwissen

$P(D)$ : Evidenz, Normierungsfaktor

Da die Evidenz nicht vom zu schätzenden Parameter abhängt, kann sie vernachlässigt werden und man erhält die *Maximum a posteriori*-Schätzung

$$\begin{aligned} \theta_{MAP} &= \underset{\theta \in \Theta}{\operatorname{argmax}} P(\theta|D) \\ &= \underset{\theta \in \Theta}{\operatorname{argmax}} \frac{P(D|\theta)P(\theta)}{P(D)} \\ &\propto \underset{\theta \in \Theta}{\operatorname{argmax}} P(D|\theta)P(\theta). \end{aligned}$$

Unterliegt die Likelihood nun einer Verteilung aus der Verteilungsklasse der Exponentialfamilie, so existiert ein sogenannter konjugierter Prior, welcher in den meisten Fällen wieder aus der Exponentialfamilie stammt. Ein konjugierter Prior hat die Eigenschaft, dass die aus der *Maximum a posteriori*-Schätzung resultierende Posteriorverteilung dem selben Verteilungstypen wie die Priorverteilung unterliegt.

Folgt die Likelihoodfunktion beispielsweise einer Exponentialverteilung mit Parameter  $\theta$  und die Prior einer Gammaverteilung mit den Parametern  $\alpha$  und  $\beta$ , so unterliegt die Posterior einer Gammaverteilung mit den Parametern  $\alpha + n$  und  $\beta + \sum_{i=1}^n x_i$ . Dabei stellt  $n$  die Anzahl und  $\sum_{i=1}^n x_i$  die Summe der Stichprobenrealisationen dar.

*Beweis.*

$$\begin{aligned} \underset{\theta \in \Theta}{\operatorname{argmax}} P(\theta|D) &\propto \underset{\theta \in \Theta}{\operatorname{argmax}} P(D|\theta)P(\theta) \\ &= \prod_{i=1}^n \theta e^{-\theta x_i} \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} \\ &\propto \prod_{i=1}^n \theta e^{-\theta x_i} \theta^{\alpha-1} e^{-\beta\theta} \\ &= \theta^n e^{-\theta \sum_{i=1}^n x_i} \theta^{\alpha-1} e^{-\beta\theta} \\ &= \theta^{\alpha+n-1} e^{-\theta(\beta + \sum_{i=1}^n x_i)} \end{aligned}$$

□

### 3 Grundbegriffe der Überlebenszeitanalyse

Im Wesentlichen muss unterschieden werden, ob die Zeitvariable  $T$  einer stetigen oder einer diskreten Verteilung folgt. Sollte eine diskrete Verteilung vorliegen, so bezeichnet  $p(t_i)$  die Wahrscheinlichkeit eines Probanden exakt zum Zeitpunkt  $t_i$  ein Ereignis zu erleiden. Wohingegen  $P(T \leq t)$  aussagt, wie wahrscheinlich es ist bis zum Zeitpunkt  $t$  ausgeschieden zu sein.

Analog dazu bezeichnen  $f(t)$  und  $F(t)$  die Dichte- bzw. Verteilungsfunktion für den Fall einer stetigen Verteilung von  $T$ .

In diesem Kapitel werden mit Survival- und Hazardfunktion zwei elementare Funktionen der Überlebenszeitanalyse hergeleitet und erläutert. Dabei wird stets zwischen stetigen und diskreten Funktionen unterschieden. Der Fall einer Mischung aus stetiger und diskreter Verteilung wird vernachlässigt.

Sämtliche Herleitungen dieses Kapitels beruhen auf *Klein & Moeschberger (2003)* [KLE-03].

#### 3.1 Survivalfunktion

$P(T \leq t)$  gibt die Wahrscheinlichkeit an, dass bei einem Proband, bis zu einem Zeitpunkt  $t$ , ein Ereignis eingetreten ist. Da man sich aber im Allgemeinen eher für die Wahrscheinlichkeit interessiert, ob ein Patient zu diesem Zeitpunkt noch ohne Ereignis ist, wird eine Survivalfunktion  $S(t)$  definiert.

Dementsprechend gilt

$$S(t) := P(T > t) \quad , t \geq 0$$

mit  $S(0) = 1$  und  $\lim_{t \rightarrow \infty} S(t) = 0$ . Die Startbedingung wird getätigt, da ein Proband nur Teil einer Studie sein kann, falls er zu Beginn dieser noch lebt. Als Endbedingung eine Grenzwahrscheinlichkeit von 0 festzulegen ist sinnvoll, da, egal ob mit oder ohne Krankheit, eine unendliche Lebenszeit ausgeschlossen ist.

##### 3.1.1 Survivalfunktion im stetigen Fall

Da die Verteilungsfunktion  $F(t)$  die Sterbewahrscheinlichkeit bis zum Zeitpunkt  $t$  angibt, folgt im stetigen Fall für  $S(t)$ :

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - \int_0^t f(u) du \\ &= \int_t^\infty f(u) du. \end{aligned}$$

Diese Umformung ist aufgrund der Eigenschaft  $\int_0^\infty f(t) dt = 1$  möglich.

Durch diesen Zusammenhang lassen sich, sobald eine der drei Funktionen  $F(t)$ ,  $f(t)$  oder  $S(t)$  eindeutig gegeben ist, die beiden anderen problemlos bestimmen.

Da  $f(t)$  die Rate angibt, in einem infinitesimalen Abstand um den Zeitpunkt  $t$  auszu-

scheiden, gilt:

$$\begin{aligned}
 f(t) &= \lim_{h \searrow 0} \frac{P(t \leq T < t+h)}{h} \\
 &= \lim_{h \searrow 0} \frac{P(T < t+h) - P(T < t)}{h} \\
 &= \lim_{h \searrow 0} \frac{(1 - P(T \geq t+h)) - (1 - P(T \geq t))}{h} \\
 &\stackrel{\text{stetig}}{=} \lim_{h \searrow 0} \frac{-P(T > t+h) + P(T > t)}{h} \\
 &= \lim_{h \searrow 0} \frac{-S(t+h) + S(t)}{h} \\
 &= -\frac{\partial S(t)}{\partial t}.
 \end{aligned}$$

### 3.1.2 Survivalfunktion im diskreten Fall

Für den diskreten Fall lässt sich die Wahrscheinlichkeit, bis zu einem Zeitpunkt  $t$  ein Ereignis zu erleiden, als die Summe aller Ausscheidewahrscheinlichkeiten bis zu diesem Zeitpunkt darstellen.

Dementsprechend gilt:

$$P(T \leq t) = \sum_{t_i \leq t} p(t_i).$$

Nun lässt sich, analog zum stetigen Fall, die Survivalfunktion  $S(t)$  durch die Gegenwahrscheinlichkeit ausdrücken:

$$\begin{aligned}
 S(t) &= 1 - P(T \leq t) \\
 &= 1 - \sum_{t_i \leq t} p(t_i) \\
 &= 1 - \left(1 - \sum_{t_i > t} p(t_i)\right), \quad \text{da } \sum_{t_i} p(t_i) = 1 \\
 &= \sum_{t_i > t} p(t_i).
 \end{aligned}$$

Im diskreten Fall entspricht die Survivalfunktion der Summe aller bis zum Zeitpunkt  $t$  noch nicht eingetroffenen Ausfallwahrscheinlichkeiten.

## 3.2 Hazardfunktion

Die Hazardfunktion gibt die Rate an, dass der Ereigniseintritt zu einem bestimmten Zeitpunkt erfolgt. Mathematisch formuliert lässt sich diese Funktion als die Rate ausdrücken, bei einem Proband genau zum Zeitpunkt  $t$  ein Ereignis zu beobachten, falls er diesen ohne Event erreicht hat.

Im Folgenden wird die Hazardfunktion mit  $h(t)$  bezeichnet.

### 3.2.1 Hazardfunktion im stetigen Fall

Für diesen Fall sei die Hazardfunktion definiert durch

$$h(t) = \lim_{h \searrow 0} \frac{P(t \leq T < t+h | T \geq t)}{h}, t \geq 0$$

und deren Zusammenhang zur Survivalfunktion ergibt sich als

$$h(t) = -\frac{\partial \ln(S(t))}{\partial t}. \quad (4)$$

*Beweis.*

Laut Definition gilt:

$$\begin{aligned} h(t) &= \lim_{h \searrow 0} \frac{P(t \leq T < t+h | T \geq t)}{h} \\ &= \lim_{h \searrow 0} \frac{P(t \leq T < t+h \cap T \geq t)}{h} \cdot \frac{1}{P(T \geq t)} \\ &= \lim_{h \searrow 0} \frac{P(t \leq T < t+h)}{h} \cdot \frac{1}{P(T \geq t)}. \end{aligned}$$

Aus 3.1.1 ist die Beziehung

$$\lim_{h \searrow 0} \frac{P(t \leq T < t+h)}{h} = -\frac{\partial S(t)}{\partial t}$$

bekannt und aufgrund der Stetigkeit gilt  $P(T \geq t) = P(T > t) = S(t)$ .

Somit lässt sich  $h(t)$  wie folgt darstellen:

$$\begin{aligned} h(t) &= \lim_{h \searrow 0} \frac{P(t \leq T < t+h | T \geq t)}{h} \\ &= -\frac{\partial S(t)}{\partial t} \cdot \frac{1}{P(T \geq t)} \\ &= -\frac{S'(t)}{S(t)} \\ &= -\frac{\partial \ln(S(t))}{\partial t}. \end{aligned}$$

□

Durch dieses Verhältnis lässt sich die Survivalfunktion  $S(t)$  auch schreiben als:

$$S(t) = e^{-\int_0^t h(u) du}. \quad (5)$$

### 3.2.2 Hazardfunktion im diskreten Fall

Für diskrete Dichten ist die Hazardfunktion definiert durch

$$h(t_i) = P(T = t_i | T \geq t_i).$$

Aus Kapitel 2.3 und Formel (2) ist bekannt, dass sich  $h(t_i)$  umformen lässt als

$$\begin{aligned} h(t_i) &= \frac{P(T = t_i)}{P(T \geq t_i)} \\ &= \frac{p(t_i)}{P(T > t_{i-1})} \\ &= \frac{p(t_i)}{S(t_{i-1})}. \end{aligned} \tag{6}$$

Im diskreten Fall lässt sich die Wahrscheinlichkeit, dass bei einem Patient genau zum Zeitpunkt  $t_i$  das Ereignis eintritt, auch als die Wahrscheinlichkeit des nicht Eintretens bis zum Zeitpunkt  $t_{i-1}$ , also  $S(t_{i-1})$ , subtrahiert um  $S(t_i)$  ausdrücken.

Der Zusammenhang lässt sich demnach schreiben als

$$p(t_i) = S(t_{i-1}) - S(t_i).$$

Ersetzt man  $p(t_i)$  nun in Gleichung (6) durch diesen Term, so ergibt sich

$$\begin{aligned} h(t_i) &= \frac{p(t_i)}{S(t_{i-1})} \\ &= \frac{S(t_{i-1}) - S(t_i)}{S(t_{i-1})} \\ &= 1 - \frac{S(t_i)}{S(t_{i-1})}. \end{aligned}$$

Seien zwischen der Zeitspanne von Beginn der Studie bis zum gewünschten Zeitpunkt  $t$  genau  $k$  Zwischenzeiten enthalten, wobei  $t_k = t$  gilt, so lässt sich die Survivalfunktion für diskrete Dichten darstellen als

$$S(t) = \prod_{t_i < t} (1 - h(t_i)). \tag{7}$$

*Beweis.*

$$\begin{aligned} \prod_{t_i < t} (1 - h(t_i)) &= \prod_{t_i < t} \left( 1 - \left( 1 - \frac{S(t_i)}{S(t_{i-1})} \right) \right) \\ &= \prod_{t_i < t} \frac{S(t_i)}{S(t_{i-1})} \\ &= \frac{S(t_1)S(t_2) \dots S(t_k)}{S(t_0)S(t_1) \dots S(t_{k-1})} \quad , \text{ da } S(t_0) = 1 \\ &= S(t_k) = S(t) \end{aligned}$$

□

### 3.3 Konkurrierende Ereignisse

Bis jetzt wurden ausschließlich Fälle betrachtet, in denen nur ein mögliches Ereignis zu beobachten war. Sind während einer klinischen Studie nun allerdings mehrere Events von Interesse, so werden diese als *konkurrierende Ereignisse* (engl. *competing risks*) bezeichnet.

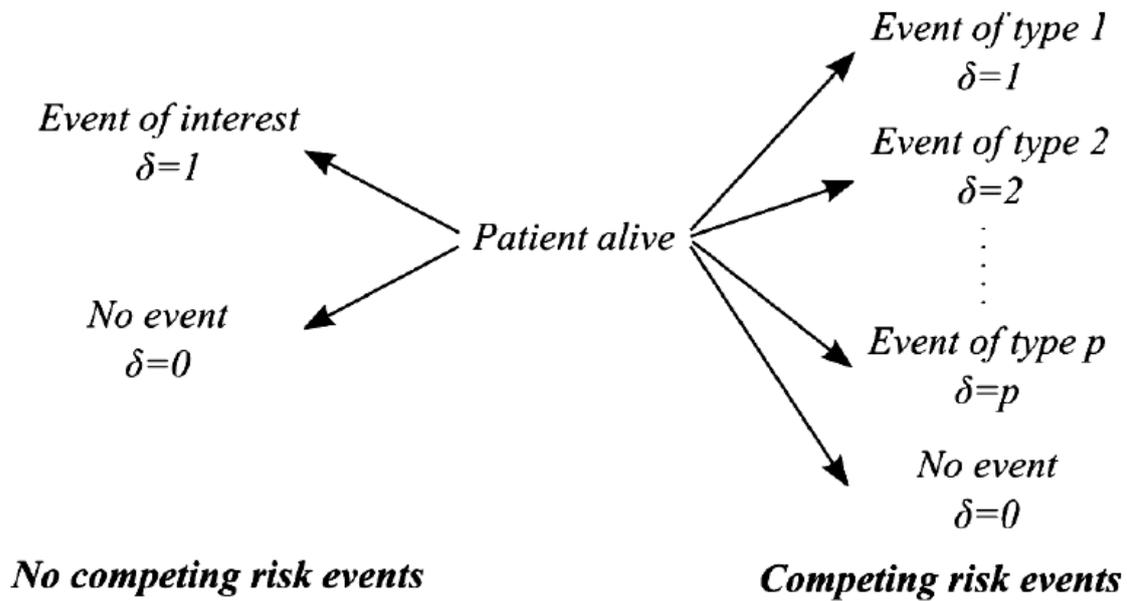


Abbildung 3: Darstellung konkurrierender Ereignisse [KRE-13]

Wie in Kapitel 3.2 die Hazardfunktion  $h(t)$  als Eintrittsrate des Ereignisses zum Zeitpunkt  $t$  definiert wurde, lässt sich dies analog für konkurrierende Ereignisse einführen. Dabei gilt

$$h_k(t) = \lim_{h \searrow 0} \frac{P(t \leq T < t+h, \delta = k | T \geq t)}{h}, \quad \forall k = 1, \dots, p.$$

Die Gesamtausfallrate lässt sich als Summe der ereignisspezifischen Ausfallraten

$$\begin{aligned} h(t) &= \lim_{h \searrow 0} \frac{P(t < T < t+h | T > t)}{h} \\ &= \lim_{h \searrow 0} \frac{\sum_{k=1}^p P(t \leq T < t+h, \delta = k | T \geq t)}{h} \\ &= \lim_{h \searrow 0} \sum_{k=1}^p \frac{P(t \leq T < t+h, \delta = k | T \geq t)}{h} \\ &= \sum_{k=1}^p \lim_{h \searrow 0} \frac{P(t \leq T < t+h, \delta = k | T \geq t)}{h} \\ &= \sum_{k=1}^p h_k(t) \end{aligned}$$

darstellen. Der Grenzwert und die Summe dürfen aufgrund der Absolutkonvergenz vertauscht werden.

Daraus lässt sich sofort die Survivalfunktion ableiten. Im Fall von konkurrierenden Risiken beschreibt diese die Wahrscheinlichkeit, dass keines der Ereignisse bis zum Zeitpunkt  $t$  eingetreten ist und lautet

$$S(t) = e^{-\int_0^t \sum_{k=1}^p h_k(u) du}.$$

Diese setzt sich also als Produkt der ereignisspezifischen Survivalfunktionen  $S_k(t)$  zusammen.

$$\begin{aligned} \prod_{k=1}^p S_k(t) &= \prod_{k=1}^p e^{-\int_0^t h_k(u) du} \\ &= e^{-\int_0^t \sum_{k=1}^p h_k(u) du} \\ &= S(t). \end{aligned}$$

### 3.4 Likelihoodfunktion für zensierte Daten

Da im späteren Verlauf der Arbeit exponentialverteilte Ereigniszeiten vorausgesetzt werden, bezieht sich dieses Kapitel ebenfalls ausschließlich auf die Exponentialverteilung.

Analog zu Kapitel 2.2, ergibt sich für die Exponentialverteilung ohne Zensierungen unter einer beobachteten Stichprobenrealisation  $\vec{t}$  die Likelihoodfunktion

$$L(\lambda, \vec{t}) = \prod_{i=1}^n \lambda e^{-\lambda t_i}.$$

Die Survivalfunktion der Exponentialverteilung ist nach Kapitel 3.1 durch

$$S(t) = 1 - F(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}$$

gegeben.

Für die Hazardfunktion  $h(t)$  gilt

$$\begin{aligned} h(t) &= -\frac{\partial \ln(S(t))}{\partial t} \\ &= -\frac{\partial \ln(\exp(-\lambda t))}{\partial t} \\ &= \frac{\partial \lambda t}{\partial t} \\ &= \lambda \end{aligned}$$

und damit stimmt die Ausfallrate für jeden Zeitpunkt  $t$  genau mit der Rate der Exponentialverteilung  $\lambda$  überein.

Sei  $\delta_i = 1$  wenn der Proband  $i$  zum Zeitpunkt  $t_i$  ein Ereignis erlitten hat und  $\delta_i = 0$  falls er zensiert wurde.

Damit ergibt sich für die Likelihoodfunktion der Exponentialverteilung mit Zensierungen

$$\begin{aligned}
L(\lambda, \vec{t}) &= \prod_{i=1}^n (\lambda e^{-\lambda t_i})^{\delta_i} (e^{-\lambda t_i})^{1-\delta_i} \\
&= \prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda t_i \delta_i} e^{-\lambda t_i} e^{\lambda t_i \delta_i} \\
&= \prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda t_i} \\
&= \prod_{i=1}^n h(t_i)^{\delta_i} S(t_i).
\end{aligned} \tag{8}$$

Optimiert man diese Funktion nun bezüglich des Parameters  $\lambda$ , so folgt:

$$\hat{\lambda}_{ML} = \underset{\lambda \in \Lambda}{\operatorname{argmax}} \prod_{i=1}^n \lambda^{\delta_i} e^{-\lambda t_i} = \frac{\sum_{i=1}^n \delta_i}{\sum_{i=1}^n t_i}$$

Der *Maximum-Likelihood*-Schätzer ist demnach der Quotient aus der Anzahl der beobachteten Ereignisse und der Gesamtüberlebenszeit aller Probanden.

Analog lässt sich Gleichung (8) für konkurrierende Ereignisse herleiten. Dazu muss zusätzlich über das Produkt aller  $p$  Parameter optimiert werden.

Sei nun  $\delta_{ki} = 1$ , falls der Proband  $i$  zum Zeitpunkt  $t_i$  das Ereignis  $k$  erlitt und ansonsten 0. Damit folgt

$$\begin{aligned}
L(\vec{\lambda}, \vec{t}) &= \prod_{k=1}^p \left( \prod_{i=1}^n h_k(t_i)^{\delta_{ki}} S_k(t_i) \right) \\
&= \prod_{k=1}^p \left( \prod_{i=1}^n \lambda_k^{\delta_{ki}} e^{-\lambda_k t_i} \right) \\
&= \prod_{i=1}^n \left( \prod_{k=1}^p \lambda_k^{\delta_{ki}} e^{-\lambda_k t_i} \right) \\
&= \prod_{i=1}^n e^{-t_i \sum_{k=1}^p \lambda_k} \prod_{k=1}^p \lambda_k^{\delta_{ki}} \\
&= \prod_{i=1}^n S(t_i) \prod_{k=1}^p h_k(t_i)^{\delta_{ki}},
\end{aligned} \tag{9}$$

wobei  $S_k(t)$  die Survivalfunktionen bzgl. jedes der  $k$  Events und  $S(t)$  die Gesamtsurvivalfunktion ist.

Optimiert man diese Funktion, so ergibt sich

$$\hat{\lambda}_k^{ML} = \underset{\vec{\lambda} \in \Lambda}{\operatorname{argmax}} \prod_{i=1}^n e^{-t_i \sum_{k=1}^p \lambda_k} \prod_{k=1}^p \lambda_k^{\delta_{ki}} = \frac{\sum_{i=1}^n \delta_{ki}}{\sum_{i=1}^n t_i}, \quad \forall k = 1, \dots, p$$

und damit folgt, in Analogie zum Fall ohne konkurrierende Ereignisse, für den Gesamthazard:

$$\hat{\lambda}_{ML} = \sum_{k=1}^p \hat{\lambda}_k^{ML} = \frac{\sum_{k=1}^p \sum_{i=1}^n \delta_{ki}}{\sum_{i=1}^n t_i}$$

Ist nur das Gesamtüberleben und nicht die Art des Ereignisses von Interesse, so wie es bei dem *PFS* der Fall ist, lässt sich die Survivalfunktion einer Studie mit Hilfe der ML-Methode durch

$$\hat{S}_{ML}(t) = e^{-\hat{\lambda}_{ML}t}$$

schätzen.

Der Nachteil dieser Schätzung ist, dass eine Verteilungsannahme getroffen werden muss. Sobald diese verletzt ist, liefert die Methode möglicherweise einen unpassenden Schätzer.

### 3.5 Kaplan-Meier-Schätzer

Der Kaplan-Meier-Schätzer (KM-Schätzer), auch Kaplan-Meier-Methode genannt, nach *Kaplan und Meier (1958)* [KAP-58], stellt eine nichtparametrische Möglichkeit dar, die Survivalfunktion zu schätzen.

Es gilt also

$$\hat{S}(t) = \hat{P}(T > t),$$

wobei  $T$  die Zufallsvariable der Überlebenszeit darstellt.

Im Allgemeinen lassen sich alle Überlebenszeitdaten wie in folgender Tabelle zusammenfassen:

$i$	$t_i$	$\delta_i$	$n_i$	$d_i$	
1	$t_1$	$\delta_1$	$n_1$	$d_1$	<b>i:</b> Nummer des Probanden
2	$t_2$	$\delta_2$	$n_2$	$d_2$	<b>t<sub>i</sub>:</b> Sortierter Zeitpunkt des Events
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	<b>δ<sub>i</sub>:</b> 1 ist eingetreten, 0 ist zensiert
$m-1$	$t_{m-1}$	$\delta_{m-1}$	$n_{m-1}$	$d_{m-1}$	<b>n<sub>i</sub>:</b> Anzahl der Versuchsobjekte, welche unmittelbar vor dem Zeitpunkt $t_i$ unter Risiko stehen
$m$	$t_m$	$\delta_m$	$n_m$	$d_m$	<b>d<sub>i</sub>:</b> Versuchsobjekte, bei denen das Ereignis zum Zeitpunkt $t_i$ eingetreten ist

Der Kaplan-Meier-Schätzer beruht auf Formel (7). Das einzige Problem ist, dass die Hazardfunktion noch geschätzt werden muss. Dazu geht man von der Annahme aus, dass es für alle Probanden, welche zum Zeitpunkt  $t$  noch ohne Ereignis sind, gleich wahrscheinlich ist als nächstes aus der Studie auszuschneiden. Dementsprechend ist nur aufgrund der vorliegenden Daten  $t_i$  bekannt, welche Probanden als nächstes ausscheiden. Die Wahrscheinlichkeit eines einzelnen Patienten zu diesem Zeitpunkt ein Ereignis erleidet, lässt sich als der Quotient aus der Anzahl der Probanden die zu diesem Zeitpunkt ein Ereignis erleiden und der Anzahl noch unter Risiko stehenden Studienteilnehmer formulieren:

$$\hat{h}_i = \frac{d_i}{n_i}.$$

Wendet man diese Schätzung nun auf die Survivalfunktion an, so folgt:

$$\hat{S}_{KM}(t) = \prod_{t_i \leq t} (1 - \hat{h}_i) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right). \quad (10)$$

Betrachtet man nun  $\hat{h}_i$ , so fällt auf, dass die zensierten Teilnehmer der Studie nur einen indirekten Einfluss auf den Schätzer haben, da zu einem Zeitpunkt, an dem ausschließlich Zensierungen vorliegen, der Quotient  $\hat{h}_i = 0$  ist und somit der gesamte Faktor gleich 1 wird. Deshalb ist es unerlässlich, dass die oben gezeigte Tabelle vor Berechnung des Schätzers nach den Ausscheidezeiten sortiert ist.

### 3.6 Mediane Überlebenszeit

Diese Informationen sind *Zwiener, Blettner & Hommel (2011)* [ZWI-11] entnommen. Eine der interessantesten Schätzgrößen des Kaplan-Meier-Schätzers ist die mediane Überlebenszeit. Diese ist definiert als der Zeitpunkt, zu dem die Überlebenswahrscheinlichkeit genau 50% beträgt.

Sollte der Fall eintreten, dass der Schätzer genau bei 50% landet, so bildet man für die geschätzte mediane Überlebenszeit den Mittelwert der Grenzen des Zeitintervalls zwischen denen der Wert genau bei 0.5 liegt.

Handelt es sich bei der Krankheit um eine gut zu behandelnde, wie beispielsweise Haut- oder Brustkrebs, so kann es sein, dass während des gesamten Beobachtungszeitraums nie der Wert 0.5 angenommen wird. In diesem Fall lässt sich keine mediane Überlebenszeit bestimmen.

Nicht nur für den Kaplan-Meier-Plot lässt sich eine mediane Überlebenszeit bestimmen, sondern auch für beliebige Survivalfunktionen. Diese stimmt genau mit dem Median der Verteilung überein.

Für den Median  $m$  der Exponentialverteilung gilt beispielsweise

$$m = \frac{\ln(2)}{\lambda}.$$

*Herleitung:*

$$\frac{1}{2} = e^{-\lambda m} \quad \Leftrightarrow \quad m = \frac{\ln(2)}{\lambda}$$

## 4 Erläuterung und Umsetzung der Problematik

Zu Beginn dieses Kapitels wird auf die genaue Simulation einer klinischen Studie eingegangen und anschließend wird erläutert, wie bei der Schätzung mittels ML-Methode die systematische Verzerrung entsteht.

### 4.1 Simulation einer klinischen Studie

Für die Simulation einer klinischen Studie werden die, in Kapitel 1 aufgezählten, Bestandteile des Studienprotokolls benötigt.

Zu Beginn wird das Visitenraster der Probanden anhand der Zeitspanne zwischen den Visiten und der maximalen Anzahl an Kontrollbesuchen erstellt. Dieses würde beispielsweise für eine Studie mit 10 Patienten, einem Kontrollintervall von zwei Monaten und einer maximalen Visitenanzahl von 10 folgendermaßen aussehen:

Proband \ Visite	1	2	3	4	5	6	7	8	9	10
1	2	4	6	8	10	12	14	16	18	20
2	2	4	6	8	10	12	14	16	18	20
3	2	4	6	8	10	12	14	16	18	20
4	2	4	6	8	10	12	14	16	18	20
5	2	4	6	8	10	12	14	16	18	20
6	2	4	6	8	10	12	14	16	18	20
7	2	4	6	8	10	12	14	16	18	20
8	2	4	6	8	10	12	14	16	18	20
9	2	4	6	8	10	12	14	16	18	20
10	2	4	6	8	10	12	14	16	18	20

Tabelle 1: Beispiel für ein Visitenraster

Da die simulierte klinische Studie so realitätsgetreu wie möglich sein soll, werden auf die eigentlichen Zeitpunkte Zufallsschocks addiert. Denn aufgrund etlicher Gründe, wie beispielsweise Feiertagen, kommt es immer zu leichten Abweichungen.

Diese Zufallsschocks werden als normalverteilte Zufallszahlen mit Erwartungswert 0 und, im Bezug auf das Kontrollintervall, geeigneter Standardabweichung erzeugt. Dies stellt sicher, dass die Wahrscheinlichkeit für sich überlappende Visitenzeitpunkte hinreichend gering ist. Sollte dieser Fall dennoch auftreten, so wird programmtechnisch nichtsdestotrotz der korrekte Visitenzeitpunkt ermittelt.

Für dieses Beispiel wurde eine Standardabweichung von 0.25 gewählt und damit ergeben sich die tatsächlichen Visitenzeitpunkte:

P \ V	V									
	1	2	3	4	5	6	7	8	9	10
1	2.32	3.92	6.11	8.05	10.08	11.90	13.73	16.37	18.34	20.30
2	2.57	3.55	6.17	7.91	9.80	12.04	13.97	16.06	17.88	20.26
3	1.65	3.95	6.25	8.19	10.39	12.14	14.15	16.02	18.16	19.74
4	1.93	4.30	5.84	7.81	10.16	12.35	13.76	15.97	18.34	20.46
5	1.96	4.47	6.12	7.65	10.02	11.81	13.86	15.70	17.72	19.83
6	2.14	3.89	5.57	8.10	10.06	12.32	14.14	16.15	17.78	20.02
7	1.92	3.93	5.80	7.79	10.17	12.08	14.19	15.94	17.71	19.89
8	1.33	3.55	5.78	8.36	10.02	12.26	14.11	15.95	17.63	19.96
9	1.39	4.11	5.39	7.89	9.25	12.23	13.77	16.23	18.02	20.04
10	2.33	3.84	6.00	8.16	10.07	12.18	13.72	16.20	18.16	20.03

Tabelle 2: Tatsächliche Visitenzeitpunkte

Anschließend werden für jeden Patienten jeweils eine Ereigniszeit für Zensierung, Progression und Tod generiert. Diese Zeiten folgen der Exponentialverteilung. Die Parameter der drei Verteilungen werden aufgrund von Vorwissen über die *mPFS* Zeit gewählt. Für das Beispiel wurde die *mPFS* Zeit für Zensierung, Progression bzw. Tod gleich 15, 5 bzw. 10 Monate gewählt. Für die Hazards gilt nach Kapitel 3.6 damit  $\lambda_C = \frac{\ln(2)}{15}$ ,  $\lambda_P = \frac{\ln(2)}{5}$  und  $\lambda_D = \frac{\ln(2)}{10}$ .

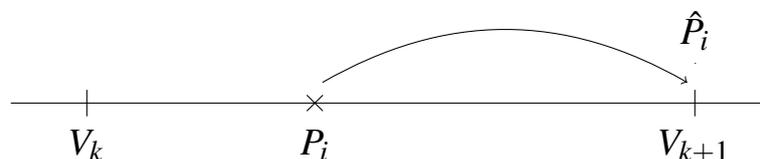
Die daraus resultierten Ereigniszeiten sind in nachstehender Tabelle zu finden:

Proband	Ereigniszeiten		
	Zensierung	Progression	Tod
1	25.28	6.64	3.58
2	19.19	17.48	6.41
3	8.73	18.52	7.23
4	4.31	11.48	0.02
5	54.77	2.21	18.61
6	21.23	2.15	5.05
7	11.17	9.14	13.16
8	59.51	10.97	9.08
9	9.10	5.86	12.43
10	1.63	0.93	19.18

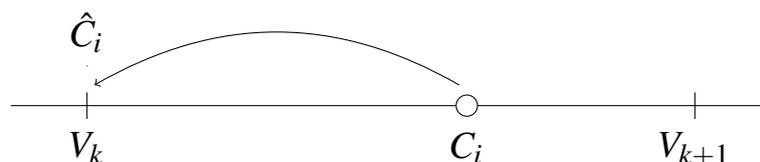
Tabelle 3: Tatsächliche Ereigniszeiten

Für jede der Zensierungs- und Progressionszeiten wird anschließend der zugehörige Visitenzeitpunkt ermittelt, zu welchem die Statusänderung des Patienten beobachtet werden konnte.

Dabei wird eine Progression immer auf jenen Visitenzeitpunkt datiert, zu welchem sie festgestellt wurde,



wohingegen bei einer Zensur der letzte wahrgenommene progressionsfreie Visitertermin ermittelt wird.



Dabei bezeichnen  $C_i$  bzw.  $P_i$  die tatsächlichen und  $\hat{C}_i$  bzw.  $\hat{P}_i$  die beobachteten Ereigniszeiten.

Nach der Anpassung an die Visiten ergeben sich folgende beobachteten Zeiten.

Proband	Ereigniszeiten		
	Zensierung	Progression	Tod
1	20.30	8.05	3.58
2	17.88	17.88	6.41
3	8.19	19.74	7.23
4	4.30	12.35	0.02
5	19.83	4.47	18.61
6	20.02	3.89	5.05
7	10.17	10.17	13.16
8	19.96	10.02	9.08
9	7.89	7.89	12.43
10	0.03	2.33	19.18

Tabelle 4: Beobachtete Ereigniszeiten

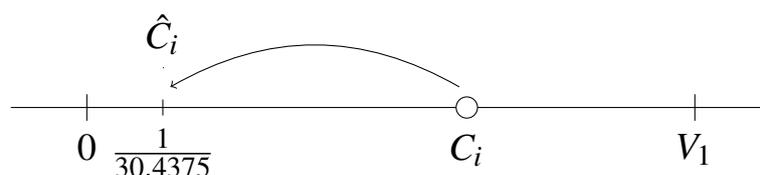
Nun wird der Status jedes Probanden über das Minimum der beobachteten Ereigniszeiten aus Tabelle 4 bestimmt.

Bei der Programmierung mussten dabei einige Spezialfälle beachtet werden.

Zwischen Progressions- und Zensierungszeiten kann es zu folgenden 3 kritischen Szenarien kommen:

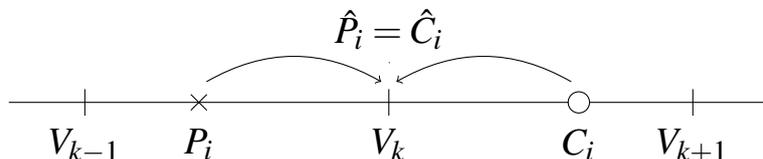
#### 1. Fall

Sollte der Zensierungszeitpunkt vor den ersten Visitertermin fallen, so kann er nicht auf den vorhergehenden Termin datiert werden, da die Studie zu diesem Zeitpunkt noch nicht begonnen hat. Da jeder Proband zu Beginn der Studie auch ein Teil dieser sein muss, ist das also nicht möglich. Dem wird abgeholfen, indem man den angepassten Zensierungszeitpunkt auf  $\hat{C}_i = \frac{1}{30.4375} \approx 0.03$  datiert. Dieser Wert kommt zustande, da, unter Berücksichtigung von Schaltjahren, ein Monat im Schnitt 30.4375 Tage besitzt. Dieser Fall trifft auf Proband 10 der Tabelle 4 zu.



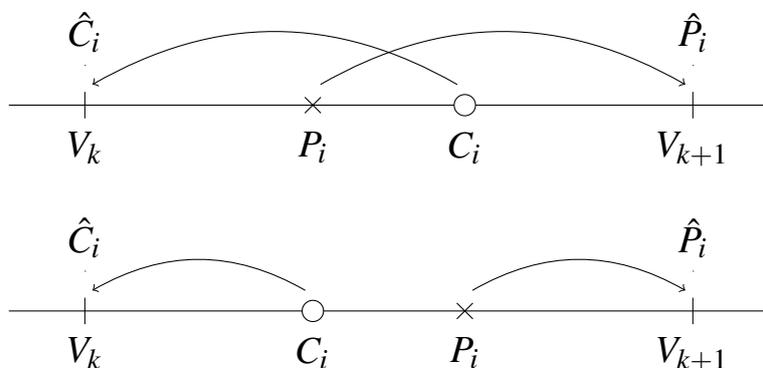
### 2.Fall

Da Progressionen immer auf den nächsten und Zensierungen auf den vorhergehenden Visitenzeitpunkt datiert werden, kann es passieren, dass  $\hat{C}_i$  und  $\hat{P}_i$  zusammenfallen. Bei Betrachtung der nachstehenden Grafik wird allerdings sofort klar, Probanden mit diesem Szenario haben auf jeden Fall eine Progression erlitten.



### 3.Fall

Fallen  $C_i$  und  $P_i$  in das gleiche Intervall, dann gilt  $\hat{C}_i < \hat{P}_i$ . Allerdings lässt sich daran noch nicht ermitteln, welche tatsächliche Ereigniszeit nun zuerst eingetreten ist. Denn wie die nachstehenden zwei Grafiken zeigen, sind beide Ereignisse möglich.



In der Realität wird in beiden Fällen die Zensierung als Ereignis ermittelt, denn egal ob der tatsächliche Progressionszeitpunkt vor oder nach dem tatsächlichen Zensierungszeitpunkt liegt, erscheint der Proband nie wieder zu einer Visite und damit wird in jedem Fall eine Zensierung beobachtet.

Unter Berücksichtigung aller möglichen Probleme ergeben sich für die zehn Probanden des Beispiels folgende Ereigniszeiten samt dazugehörigen Status.

Proband	Ereigniszeit	Status
1	3.58	1
2	6.41	1
3	7.23	1
4	0.02	1
5	4.47	1
6	3.89	1
7	10.17	1
8	9.08	1
9	7.89	1
10	0.03	0

Tabelle 5: Beobachtete Events mit dazugehörigen Ereigniszeiten

Ein Status gleich 1 bedeutet dabei, dass der Proband entweder eine Progression erlitt oder verstorben ist, wohingegen 0 eine Zensierung bedeutet.

Da aber nicht alle Patienten gleichzeitig in die Studie aufgenommen werden, wird zu jeder Zeit ein gleichverteilter Zufallsschock addiert. Dieser spiegelt den Rekrutierungszeitraum wider. Der Zeitpunkt 0 ist damit der tatsächliche Start der Studie. Bei einer Rekrutierungsperiode von einem Monat lauten die an den Studienbeginn angepassten Ereigniszeiten wie folgt.

Proband	Ereigniszeit	Studienzeitpunkt	Status
1	3.58	3.91	1
2	6.41	7.37	1
3	7.23	8.06	1
4	0.02	0.14	1
5	4.47	5.31	1
6	3.89	4.83	1
7	10.17	11.08	1
8	9.08	10.03	1
9	7.89	8.81	1
10	0.03	0.74	0

Tabelle 6: An Studienstart angepasste Ereigniszeitpunkte

Wie in Kapitel 1 beschrieben, wird nicht immer das Ende einer Studie abgewartet. Im Studienprotokoll wird eine maximale Ereignisanzahl und eine *maximale follow up*-Zeit festgehalten. Würde man für dieses Beispiel eine maximale Anzahl von acht Ereignissen festlegen, so bezeichnet der Zeitpunkt 10.03 Monate den sogenannten *cutoff*. Alle Ereignisse nach diesem Zeitpunkt werden als Zensurierungen betrachtet und als Ereigniszeit wird die tatsächliche Zeitspanne der Probanden in der Studie verwendet. Diese berechnet sich aus der Differenz von *cutoff*- und Rekrutierungszeitpunkt. Für das Beispiel würden demnach folgende Ergebnisse resultieren:

Proband	Ereigniszeit	Status
1	3.58	1
2	6.41	1
3	7.23	1
4	0.02	1
5	4.47	1
6	3.89	1
7	9.12	0
8	9.08	1
9	7.89	1
10	0.03	0

Tabelle 7: Ereigniszeiten nach *cutoff*

Analog verhält es sich mit der *maximalen follow up*-Zeit. Ist bis zu einem gewünschten Studienzeitpunkt nicht die vorgegebene Anzahl an Events eingetreten, so bildet die *maximale follow up*-Zeit den *cutoff*.

## 4.2 Entstehung der Verzerrung

Das in Kapitel 4.1 geschilderte Prozedere zur Simulation einer klinischen Studie spiegelt einen realistischen Ablauf wider und durch die Anpassung an die Visitenzeitpunkte entsteht die systematische Verzerrung des ML-Schätzers.

Da in einer klinischen Studie, im Normalfall, die Progression das am häufigsten beobachtete Ereignis ist, wird die Summe der Ereigniszeiten überschätzt und daraus ergibt sich eine systematische Überschätzung der *mPFS* Zeit.

Im Folgenden wird der Bestandteil der Likelihoodfunktion ermittelt, welcher zu der Verzerrung führt.

Bevor auf den genauen Aufbau der Likelihoodfunktion eingegangen wird, müssen erst einige Mengen definiert werden.

Es gilt :

**A:** Menge der Probanden mit Ereignis Tod

**C:** Menge der Probanden mit Zensierung durch den *cutoff*

**P:** Menge der Probanden mit Progression

**I:** Menge der Probanden mit Zensierung während des Visitenplans

Im Folgenden sei  $t_i$  die wahre Ereigniszeit des Probanden  $i$ , egal ob diese bekannt ist oder nicht.

Analog zur Gleichung (9) lässt sich für jede Menge  $M$ , mit  $M \in \{A, I, P, C\}$ , der Anteil der Likelihoodfunktion

$$\prod_{i \in M} S(t_i) \prod_{k=1}^p h_k(t_i)^{\delta_{ki}}$$

bestimmen.

Im Folgenden bezeichnen  $\lambda_P$  bzw.  $\lambda_D$  die Parameter der Exponentialverteilung für die Progressions- bzw Todeszeiten. Wie bereits gezeigt, gilt für beliebige exponentialverteilte Ereigniszeiten  $h(t) \equiv \lambda$ . Dabei ist es irrelevant, ob die tatsächliche Überlebenszeit bekannt ist und somit gilt, sowohl für die Zensierungs- als auch für die Progressionszeiten,  $h(t^*) = \lambda, \forall t^* \in [v_{i-1}, v_i)$  bzw.  $t^* \in [v_i, v_{i+1})$ .

Es bezeichne  $v_i$  den Visitermin, an welchem bei Proband  $i$  die Progression bzw. die Zensierung beobachtet wurde.

Somit gilt nach Formel (9) für die einzelnen Bestandteile der Likelihoodfunktion:

$$\begin{aligned} \text{A: } & \prod_{i \in A} \lambda_P^{\Delta_i} \cdot \lambda_D^{\delta_i} \cdot e^{-(\lambda_P + \lambda_D)t_i} \\ \text{C: } & \prod_{i \in C} \lambda_P^{\Delta_i} \cdot \lambda_D^{\delta_i} \cdot e^{-(\lambda_P + \lambda_D)t_i} \\ \text{P: } & \prod_{i \in P} \lambda_P^{\Delta_i} \cdot \lambda_D^{\delta_i} \cdot e^{-(\lambda_P + \lambda_D)t_i} \\ \text{I: } & \prod_{i \in I} \lambda_P^{\Delta_i} \cdot \lambda_D^{\delta_i} \cdot e^{-(\lambda_P + \lambda_D)t_i} \end{aligned} \quad \begin{aligned} & \text{mit } \delta_i = \begin{cases} 1 & t_i = D_i \\ 0 & \text{sonst} \end{cases} \\ & \text{und } \Delta_i = \begin{cases} 1 & \hat{P}_i = \min\{\hat{C}_i, \hat{P}_i, D_i\} \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Diese Bestandteile können noch etwas vereinfacht werden. Über Probanden der Menge  $A$  ist bekannt, dass sie einen Tod erlitten haben. Somit kann bei ihnen keine Progression aufgetreten sein. Analoge Aussagen lassen sich für die Mengen  $C$ ,  $P$  bzw.  $I$  tätigen und damit ergeben sich die Bestandteile der Likelihoodfunktion durch:

$$\begin{aligned}
\text{A: } & \prod_{i \in A} \lambda_D^{\delta_i} \cdot e^{-(\lambda_P + \lambda_D)t_i} \\
\text{C: } & \prod_{i \in C} e^{-(\lambda_P + \lambda_D)t_i} \\
\text{P: } & \prod_{i \in P} \lambda_P^{\Delta_i} \cdot e^{-(\lambda_P + \lambda_D)t_i} \\
\text{I: } & \prod_{i \in I} e^{-(\lambda_P + \lambda_D)t_i}
\end{aligned}
\quad \text{mit } \delta_i = \begin{cases} 1 & t_i = D_i \\ 0 & \text{sonst} \end{cases}$$

$$\text{und } \Delta_i = \begin{cases} 1 & \hat{P}_i = \min\{\hat{C}_i, \hat{P}_i, D_i\} \\ 0 & \text{sonst} \end{cases}$$

Da für die Mengen  $P$  und  $I$  die wahren Zeiten  $t_i$  nicht bekannt sind, können diese in zwei Summanden aufgesplittet werden.

Die wahren  $t_i$  lassen sich, abhängig vom Status, durch

$$\begin{aligned}
\text{P: } & t_i = v_{i-1} + x_i \\
\text{I: } & t_i = v_i + x_i
\end{aligned}$$

ausdrücken.

Über den Term  $x_i$  ist dabei nur  $x_i \in [0, k)$  bekannt. Bei  $k$  handelt es sich um die Zeitspanne zwischen den Visiten.

Die Anteile an der Likelihoodfunktion lassen sich wie folgt umformulieren:

$$\begin{aligned}
\text{P: } & \prod_{i \in P} \lambda_P^{\Delta_i} \cdot e^{-(\lambda_P + \lambda_D)t_i} = \prod_{i \in P} \lambda_P^{\Delta_i} \cdot e^{-(\lambda_P + \lambda_D)v_{i-1}} \cdot e^{-(\lambda_P + \lambda_D)x_i} \\
\text{I: } & \prod_{i \in I} e^{-(\lambda_P + \lambda_D)t_i} = \prod_{i \in I} e^{-(\lambda_P + \lambda_D)v_i} \cdot e^{-(\lambda_P + \lambda_D)x_i}
\end{aligned} \tag{11}$$

Dabei handelt es sich bei  $e^{-(\lambda_P + \lambda_D)x_i}$  um den Term, welcher, durch seine Ungewissheit, den Bias in der Schätzung verursacht.

In den Kapiteln 5 und 6 werden zwei Methodiken vorgestellt, mit deren Hilfe versucht wird diese Ungewissheit zu reduzieren.

## 5 Biaskorrektur durch Maximum a posteriori-Ansatz

Zu Beginn dieses Kapitels wird die Idee hinter dem *Maximum a posteriori*-Ansatz erläutert und dieser wird mathematisch hergeleitet. Anschließend folgt eine Sensitivitätsanalyse zur Bewertung der Ergebnisse.

### 5.1 Herleitung

Wie in Kapitel 2.5 bereits erläutert wurde, kann mit Hilfe der *Maximum a posteriori*-Methode

$$\theta_{MAP} = \underset{\theta \in \Theta}{\operatorname{argmax}} P(D|\theta)P(\theta)$$

eine Schätzung des Hazards vorgenommen werden.

Dabei wird allerdings die Likelihoodfunktion  $P(D|\theta)$  benötigt und diese führt, wie in Kapitel 4.2 gezeigt, zu einer Verzerrung des Schätzers.

Man betrachte erneut den gesamten Term aus Gleichung (11), welcher diese Verzerrung beinhaltet

$$e^{-(\lambda_P + \lambda_D)t_i}$$

Sei o.B.d.A das erlittene Ereignis eine Progression und es gelte  $\lambda := \lambda_P + \lambda_D$ .

Obiger Term gibt für eine exponentialverteilte Zufallsvariable  $T \sim \operatorname{Exp}(\lambda)$  die Wahrscheinlichkeit  $P(T \geq t_i)$  an. Des Weiteren ist  $t_i \in [v_{i-1}, v_i)$  bekannt und aufgrund der äquidistanten Zeitabstände zwischen den Visiten gilt  $v_i = v_{i-1} + k$ .

Um den Bias in der Schätzung zu korrigieren war nun die Idee  $t_i$  durch den bedingten Erwartungswert  $E(T|v_{i-1} \leq T < v_i)$  zu ersetzen. Dies führt zu einer Minimierung der fehlerhaft beobachteten Zeit.

Für den bedingten Erwartungswert gilt:

$$\begin{aligned} E(T|v_{i-1} \leq T < v_i) &= \frac{E(T \cdot \mathbb{1}_{\{v_{i-1} \leq T < v_i\}})}{P(v_{i-1} \leq T < v_i)} \\ &= \frac{\int_{v_{i-1}}^{v_i} x \lambda e^{-\lambda x} dx}{\int_{v_{i-1}}^{v_i} \lambda e^{-\lambda x} dx} \\ &= \frac{-v_i e^{-\lambda v_i} + v_{i-1} e^{-\lambda v_{i-1}} - \frac{1}{\lambda} e^{-\lambda v_i} + \frac{1}{\lambda} e^{-\lambda v_{i-1}}}{-e^{-\lambda v_i} + e^{-\lambda v_{i-1}}} \\ &= \frac{1}{\lambda} + \frac{-v_i e^{-\lambda v_i} + v_{i-1} e^{-\lambda v_{i-1}}}{-e^{-\lambda v_i} + e^{-\lambda v_{i-1}}} \\ &= \frac{1}{\lambda} + \frac{-(v_{i-1} + k) e^{-\lambda v_i} + v_{i-1} e^{-\lambda v_{i-1}}}{-e^{-\lambda v_i} + e^{-\lambda v_{i-1}}} \\ &= \frac{1}{\lambda} + \frac{-k e^{-\lambda v_i} - v_{i-1} e^{-\lambda v_i} + v_{i-1} e^{-\lambda v_{i-1}}}{-e^{-\lambda v_i} + e^{-\lambda v_{i-1}}} \\ &= \frac{1}{\lambda} - \frac{e^{-\lambda v_i}}{e^{-\lambda v_i}} \cdot \frac{k}{e^{\lambda k} - 1} + v_{i-1} \\ &= \frac{1}{\lambda} - \frac{k}{e^{\lambda k} - 1} + v_{i-1} \end{aligned}$$

Für eine Zensur verläuft die Berechnung analog, nur gilt dabei  $t_i \in [v_i, v_{i+1})$  und damit folgt

$$E(T|v_i \leq T < v_{i+1}) = \frac{1}{\lambda} - \frac{k}{e^{\lambda k} - 1} + v_i.$$

Betrachtet man erneut Gleichung (11), so fällt auf, dass mit Hilfe dieser Methode  $v_{i-1}$  bzw.  $v_i$  unberührt bleiben und nur eine Substitution von  $x_i$  durch  $\frac{1}{\lambda} - \frac{k}{e^{\lambda k} - 1}$  stattfindet. Zur besseren Darstellung wird dieser Term im Folgenden mit  $t(\lambda, k)$  bezeichnet. Mit Hilfe der Substitution ergibt sich für die Bestandteile der Likelihoodfunktion:

$$\begin{aligned} \text{A: } & \lambda \sum_{i \in A} \delta_i \cdot e^{-(\lambda_P + \lambda_D) \sum_{i \in A} t_i} \\ \text{C: } & e^{-(\lambda_P + \lambda_D) \sum_{i \in C} t_i} \\ \text{P: } & \lambda_P^{\sum_{i \in P} \Delta_i} \cdot e^{-(\lambda_P + \lambda_D) \sum_{i \in P} v_{i-1}} \cdot e^{-(\lambda_P + \lambda_D) \cdot t(\lambda_P + \lambda_D, k) \cdot p} \quad \text{mit } |P| = p \\ \text{I: } & e^{-(\lambda_P + \lambda_D) \sum_{i \in I} v_i} \cdot e^{-(\lambda_P + \lambda_D) \cdot t(\lambda_P + \lambda_D, k) \cdot j} \quad \text{mit } |I| = j \end{aligned}$$

Das Produkt all dieser Wahrscheinlichkeiten und der Priorinformationen über die Parameter  $\lambda_P$  und  $\lambda_D$  ergibt die zu maximierende Funktion. Aufgrund der Komplexität des Problems müssen für die Optimierung numerische Verfahren verwendet werden.

Da  $t(\lambda, k)$  allerdings ebenfalls von der Zeitspanne zwischen den Visiten  $k$  abhängt, muss erst überprüft werden, ob die Schätzung mit Hilfe dieser Methode keiner systematischen Verzerrung unterliegt.

Jenes Verhalten und die Verbesserung im Vergleich zur ML-Schätzung werden an folgendem, deutlich vereinfachten, Beispiel dargestellt. Vereinfacht ist es dadurch, dass zur besseren grafischen Darstellung nur Progressionen und Zensierungen betrachtet werden.

Für alle in den nächsten Kapiteln folgenden Simulationen, wurden die Parameter der Gammaverteilung  $\alpha_i = 5$  und  $\beta_i = 5/\lambda_i^{\text{prior}}$ , mit  $i \in \{C, P, D\}$ , gewählt, wobei die  $\lambda_i^{\text{prior}}$  stets explizit angegeben werden. Dadurch wird eine hinreichende Varianz garantiert ohne eine uninformative Priorverteilung zu erhalten. Alle zeitlichen Angaben werden in Monaten getätigt.

Die in den Kapiteln 1 und 4.1 eingeführten Parameter zur Simulation einer klinischen Studie seien für dieses Beispiel:

- Stichprobenumfang = 200
- Rekrutierungszeitraum = 5 Monate
- Anzahl Visiten = 50
- Follow up = 120 Monate
- Visiten SD = 0.25
- $\lambda_P^{\text{prior}} = \frac{\ln(2)}{7}$
- $\lambda_C^{\text{prior}} = \frac{\ln(2)}{50}$

Um die Verbesserung aufzuzeigen, wurden die vier nachstehenden Plots mit einer Zeitspanne von 3, 4, 5 und 6 Monaten simuliert. Aufgrund einer Umskalierung der Dichten auf die gleiche Größe, werden keine Werte auf der y-Achse angegeben.

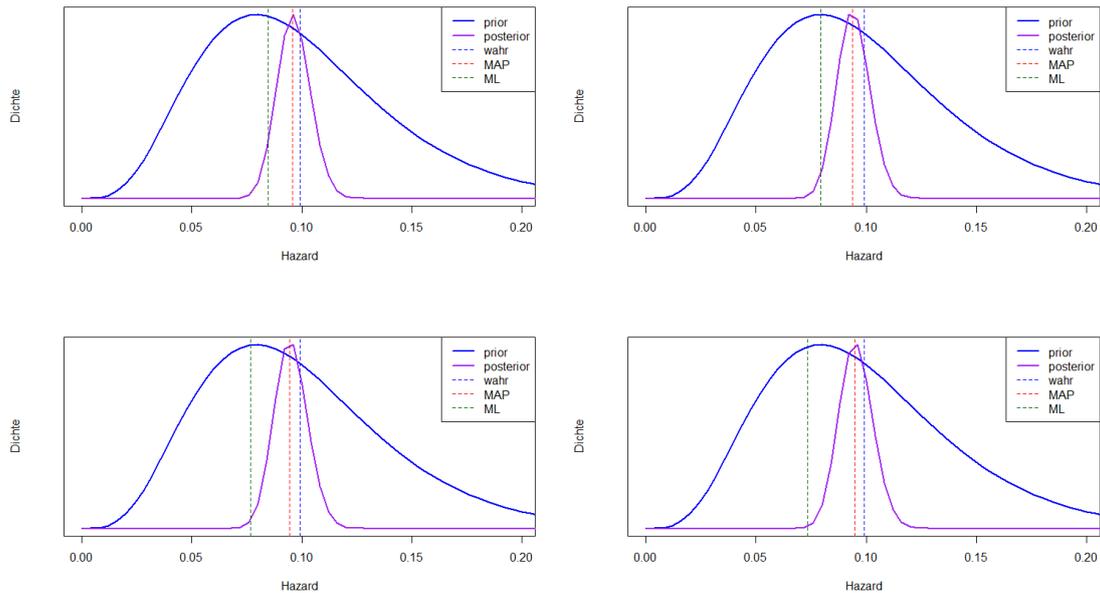


Abbildung 4: Verdeutlichung der Intervallunabhängigkeit der MAP-Methode für eine zunehmende Zeitspanne zwischen den Visiten von links nach rechts bei 3, 4, 5 und 6 Monaten

Die blaue Kurve stellt die Prior- und die violette die Posteriordichte dar. Die grüne vertikale Linie ist die *Maximum-Likelihood*-Schätzung, die rote der *Maximum a posteriori*-Schätzer und bei der blauen handelt es sich um den wahren Hazard.

Wie zu sehen ist, nimmt der Abstand zwischen ML-Schätzung und wahren Wert mit größerer Zeitspanne zwischen den Visiten zu, wohingegen sich der Unterschied zwischen MAP-Schätzung und wahren Wert annähernd konstant verhält.

Dies wird in nachstehender Abbildung noch einmal verdeutlicht.

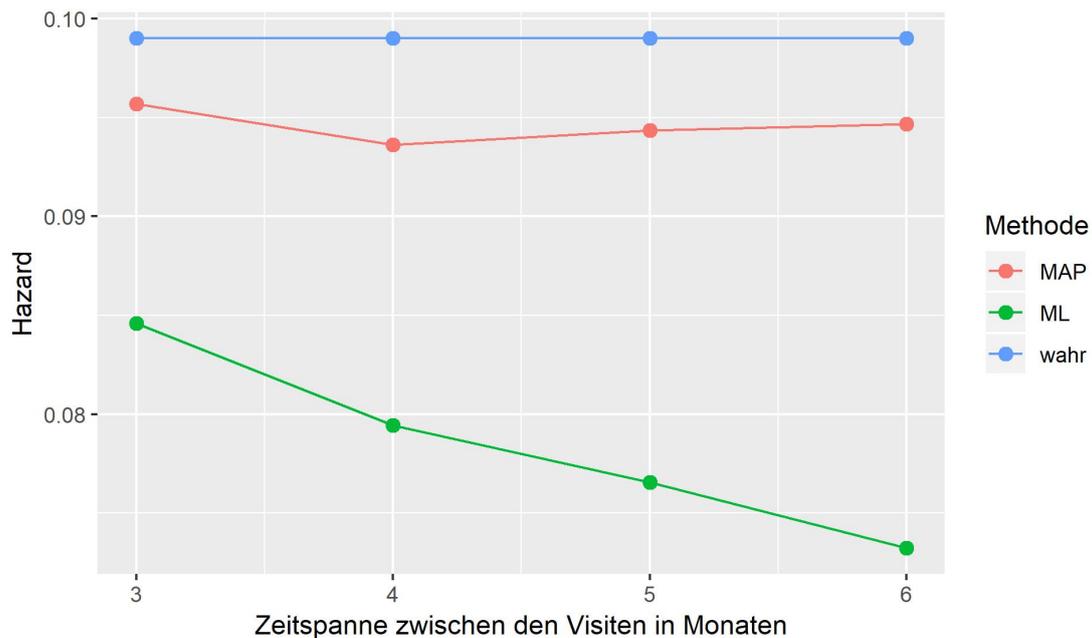


Abbildung 5: Verdeutlichung der Intervallunabhängigkeit der MAP-Methode (2)

Wie zu erkennen ist, wächst die Verzerrung durch die Berechnung mittels ML-Schätzer mit nahezu linearem Zusammenhang zur Zeitspanne zwischen den Visiten. Im Gegensatz dazu kann solch ein Zusammenhang für die Schätzung mit Hilfe der *Maximim a posteriori*-Methode nicht festgestellt werden.

Um eine genauere Analyse der Methodik durchzuführen wird eine Sensitivitätsanalyse durchgeführt.

## 5.2 Sensitivitätsanalyse

Um ein möglichst großes Spektrum an Variation in den Parametern abzudecken, werden für alle zu prüfenden Parameterkonstellationen 5000 Studien simuliert, deren Parameter aus der Priorverteilung gezogen werden. Anschließend wird der Hazard sowohl mit Hilfe der ML- als auch mit der MAP-Methode geschätzt. Die Darstellung der Ergebnisse geschieht mit Hilfe von Boxplots, welche die Differenz der *mPFS* Zeiten zwischen geschätztem und wahren Wert aufzeigen.

Die verschiedenen Parameter der Konstellationen werden dabei stets rechts neben den Boxplots aufgelistet.

Die Abbildungen 6 bis 9 stellen prinzipiell das gleiche Beispiel dar, nur wird jeweils ein Übergabeparameter geändert. Damit soll überprüft werden, ob der Bias lediglich durch die Zeitspanne zwischen den Visiten verursacht wird oder ob es weitere Einflussfaktoren gibt.

In den Abbildungen 10 und 11 werden Extrembeispiele dargestellt.

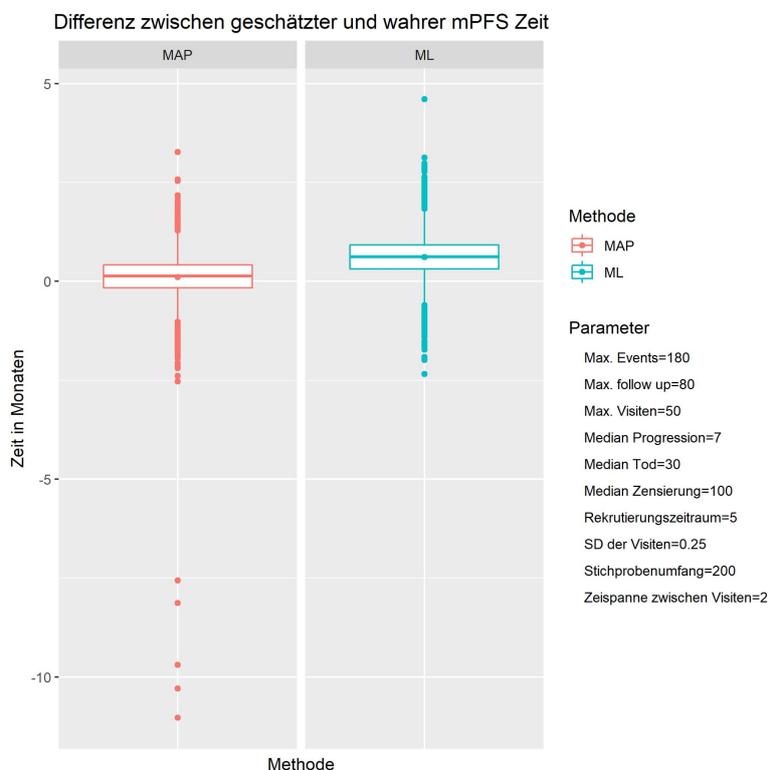


Abbildung 6: Referenzbeispiel zur Überprüfung der Ergebnisse der MAP-Methode

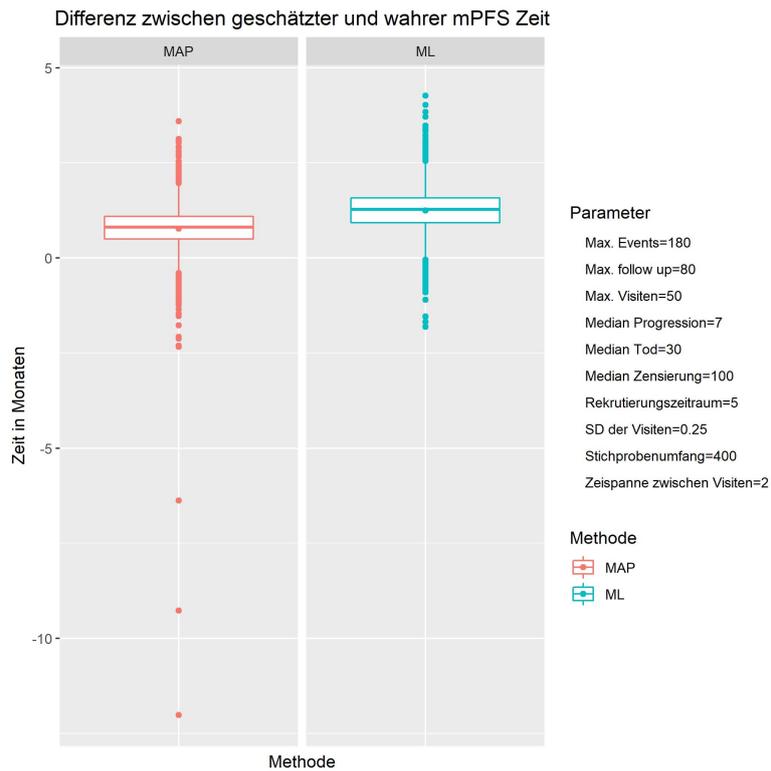


Abbildung 7: Referenzbeispiel zur Überprüfung der Ergebnisse der MAP-Methode mit einem Stichprobenumfang von 400 anstatt 200 Probanden

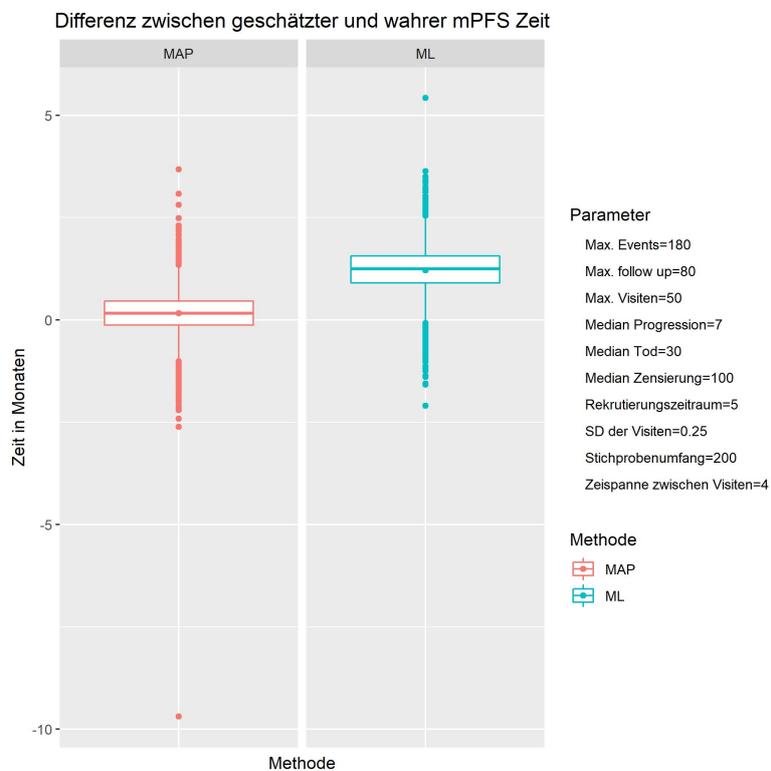


Abbildung 8: Referenzbeispiel zur Überprüfung der Ergebnisse der MAP-Methode mit einer Zeitspanne zwischen den Visiten von 4 anstatt 2 Monaten

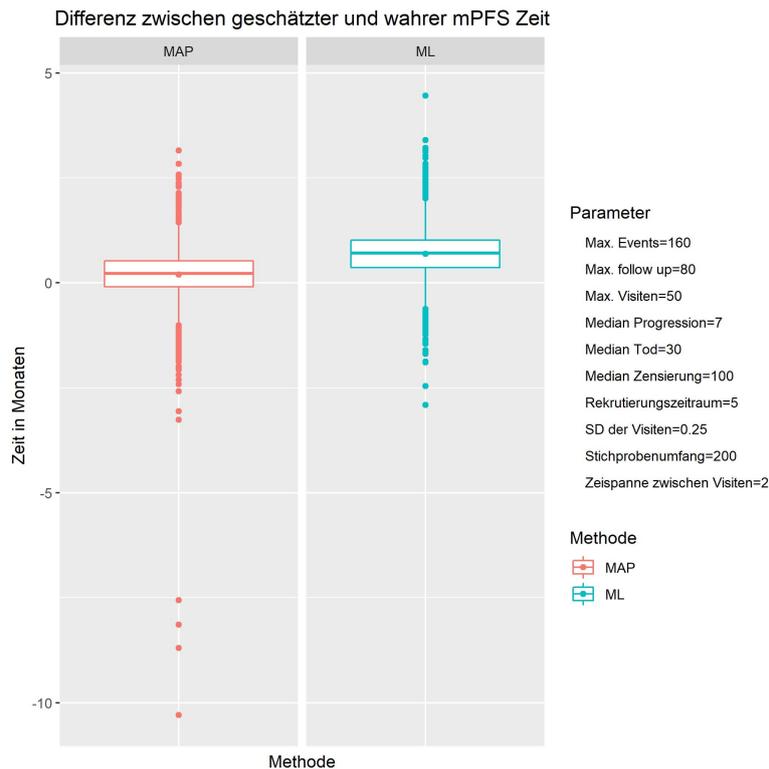


Abbildung 9: Referenzbeispiel zur Überprüfung der Ergebnisse der MAP-Methode mit 160 anstatt 180 als maximale Anzahl an beobachteten Ereignissen

Analog zum vereinfachten Beispiel des vorigen Kapitels lässt sich im Mittel eine Verbesserung in der Schätzung des *mPFS* mit Hilfe der MAP- im Vergleich zur ML-Methode beobachten. Einzig in Abbildung 7 ist die 0 nicht zwischen dem 25%- und 75%-Quantil enthalten.

Dies ist auf den erhöhten Stichprobenumfang, im Vergleich zu den anderen Beispielen, zurückzuführen, da die Substitution von  $x_i$  durch  $t(\lambda, k)$  öfter eingesetzt werden muss. Die Substitution mit Hilfe des bedingten Erwartungswertes führt alleinig zu einer Minimierung der Ungewissheit des unbekanntes Terms und nicht zu einer vollständigen Eliminierung.

Im Mittel wird der kleinstmögliche Fehler begangen, allerdings kommt es für die tatsächlichen Ausprägungen der  $x_i$  zu Abweichungen, welche wiederum zu einer verzerrten Schätzung führen.

Für die Schätzung mit Hilfe der MAP-Methode zeigen die Boxplots zwar eine relativ große Varianz und einige Ausreißer auf, da dies allerdings auch für den ML-Schätzer gilt, handelt es sich dabei um keinen Nachteil der Methode.

Wie bereits erwähnt folgen nun zwei Extrembeispiele. Extrem sind sie deshalb, weil solche Parameterkonstellationen unrealistisch für echte klinische Studien sind.

In Abbildung 10 liegt für alle drei Ereignisse eine mediane Eintrittszeit von zehn Monaten vor. Es ist also für jedes Ereignis gleich wahrscheinlich einzutreten. Im Beispiel der Abbildung 11 gelten für die medianen Ereigniszeiten für Progression drei bzw. für Tod sieben Monate, wohingegen eine Zeitspanne von acht Monaten zwischen den Visiten liegt. Dies bedeutet, der Großteil der Ereignisse tritt bereits ein bevor die Patienten überhaupt ihren ersten Visitermin wahrnehmen konnten.

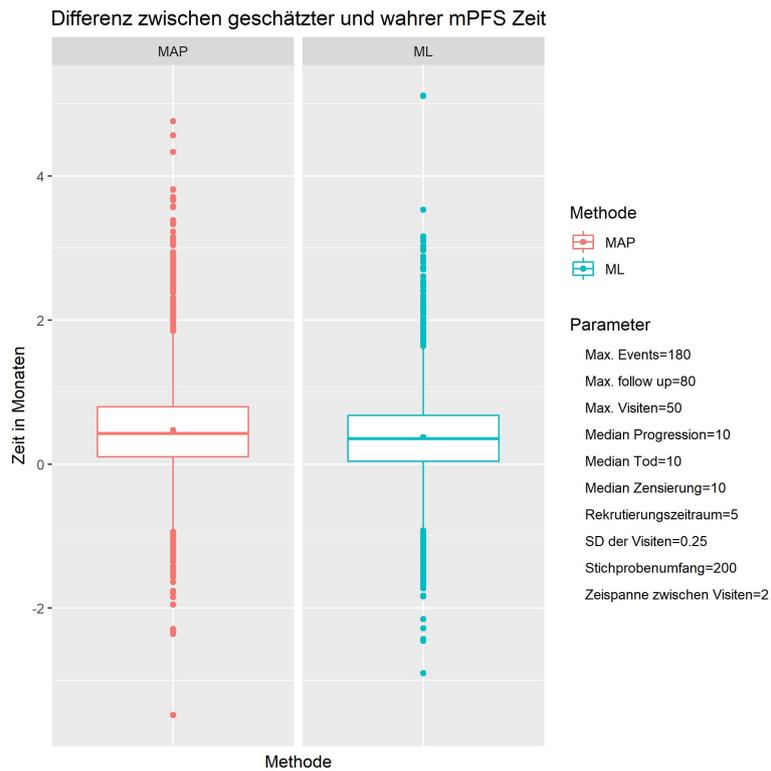


Abbildung 10: Aufgrund von identischen Raten der drei konkurrierenden Ereignissen 1. Extrembeispiel für die Ergebnisse der MAP-Methode

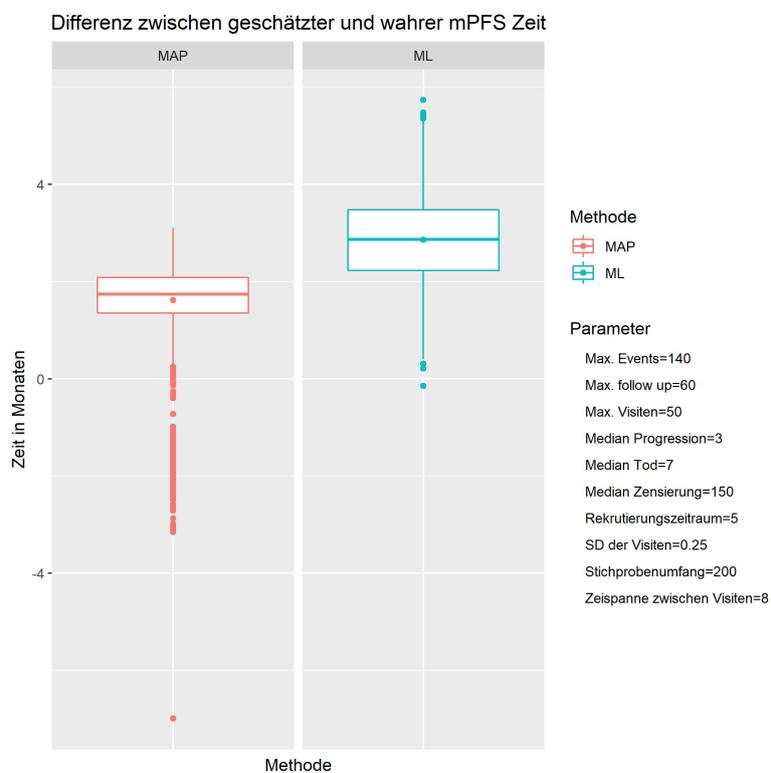


Abbildung 11: Aufgrund von einer Zeitspanne von 8 Monaten zwischen den Visiten bei geringeren Ausfallraten 2. Extrembeispiel für die Ergebnisse der MAP-Methode

In Abbildung 10 ähneln sich die Ergebnisse der beiden Schätzungen stark. Dies ist darin begründet, dass während die MAP-Methode jeweils für Zensierung und Progression eine Minimierung des Fehlers durchführt, wird dieser bei der ML-Methode selbst minimiert. Da es in diesem Beispiel für alle Ereignisse gleich wahrscheinlich ist einzu- treffen, heben sich die Unterschätzung der Ereigniszeiten durch die Zensierungen und deren Überschätzung aufgrund der Progressionen gegenseitig auf.

Im zweiten Extrembeispiel liefert der MAP-Schätzer zwar ein besseres Ergebnis, wenn- gleich dieses auch nicht zufriedenstellend ist. Während der ML-Schätzer sich um un- gefähr 3 Monate in der Differenz des *medianen progressionsfreien Überlebens* über- schätzt, konnte das Ergebnis auf ca. 1.9 Monate reduziert werden.

Da die Substitution der  $x_i$  durch  $t(\lambda, k)$  nicht immer zu akzeptablen Ergebnissen führt, wurde nach einer Möglichkeit gesucht die Methode zu verbessern. Anstelle die  $x_i$  durch ihren bedingten Erwartungswert zu ersetzen, können exponentialverteilte Zufallszah- len mit Hilfe von Modulorechnung an das Intervall  $[0, k)$  angepasst werden. Anschlie- ßend ersetzt man die  $x_i$  durch diese angepassten Zufallszahlen und berechnet den MAP- Schätzer. Die Resultate der Modulorechnung stehen in Analogie zu den Fehlern  $x_i$  der beobachteten Ereigniszeiten. Da die Parameter  $\lambda_D$  und  $\lambda_P$  gesucht sind, werden für die Generierung der Zufallszahlen die Priorparameter  $\lambda_D^{prior}$  und  $\lambda_P^{prior}$  verwendet. Dieses Verfahren wird wiederholt durchgeführt und anschließend wird der Mittelwert aller be- rechneten Schätzer als Schätzung für den Hazard verwendet. Die Anzahl der Durchfüh- rungen wird mit Hilfe der Variable *Anzahl* der Boxplots angegeben.

Zum besseren Vergleich der Ergebnisse wird dieses Verfahren auf die identischen Test- fälle wie der eigentliche MAP-Ansatz angewandt.

Die Ergebnisse sind in den nachstehenden Abbildungen 12 bis 17 dargestellt.

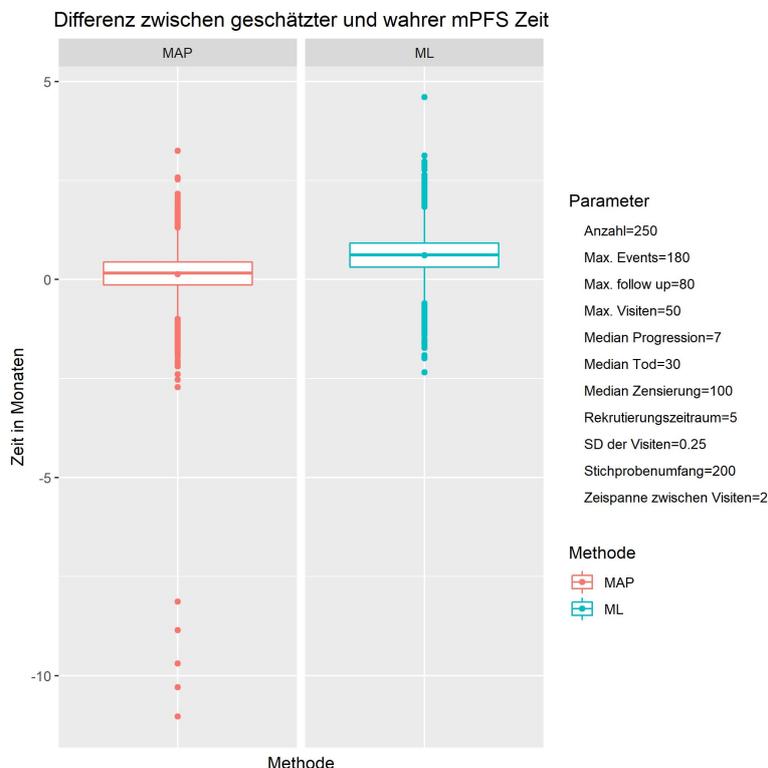


Abbildung 12: Referenzbeispiel zur Überprüfung der Ergebnisse der modifizierten MAP-Methode

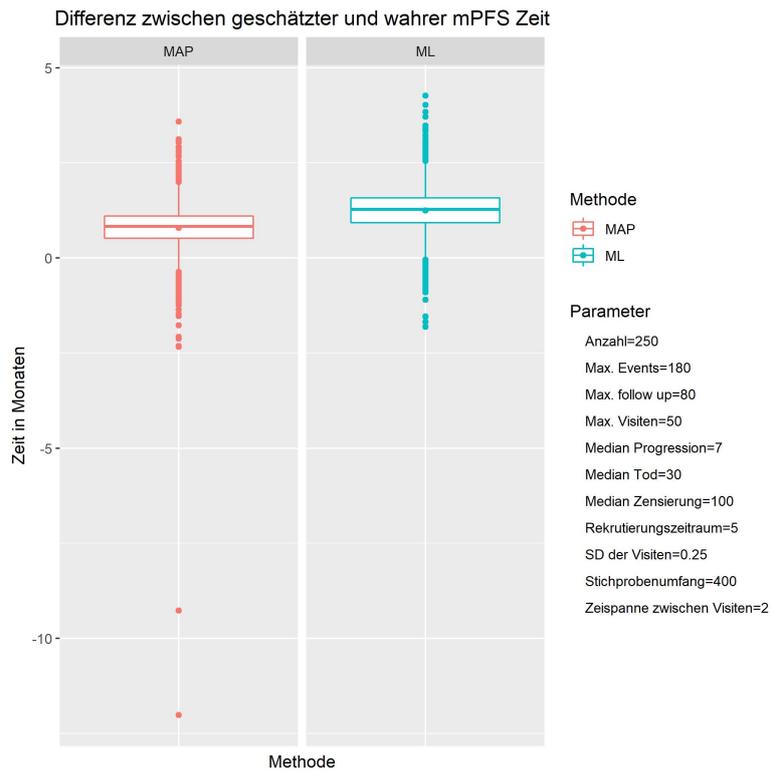


Abbildung 13: Referenzbeispiel zur Überprüfung der Ergebnisse der modifizierten MAP-Methode mit einem Stichprobenumfang von 400 anstatt 200 Probanden

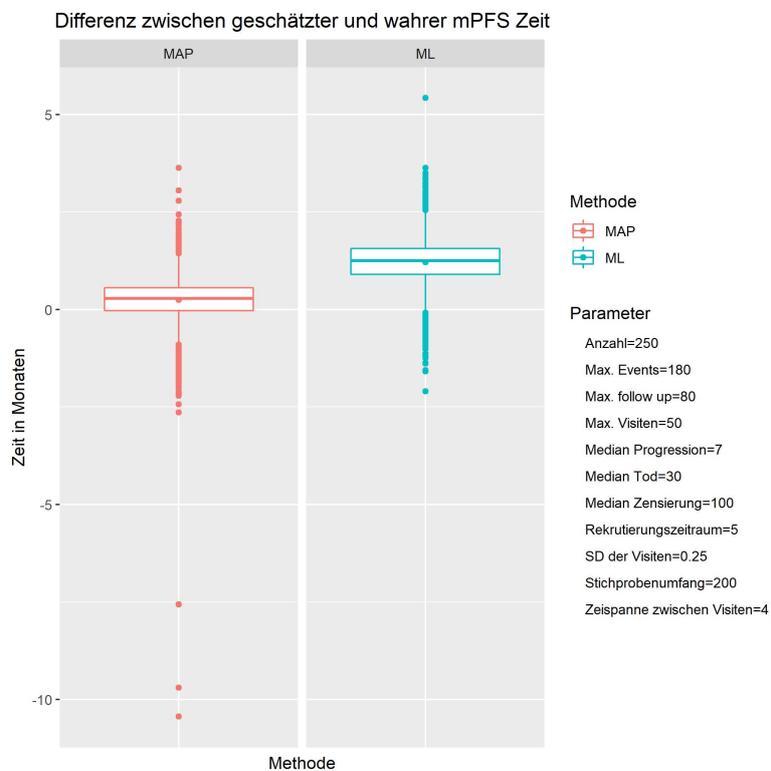


Abbildung 14: Referenzbeispiel zur Überprüfung der Ergebnisse der modifizierten MAP-Methode mit einer Zeitspanne zwischen den Visiten von 4 anstatt 2 Monaten

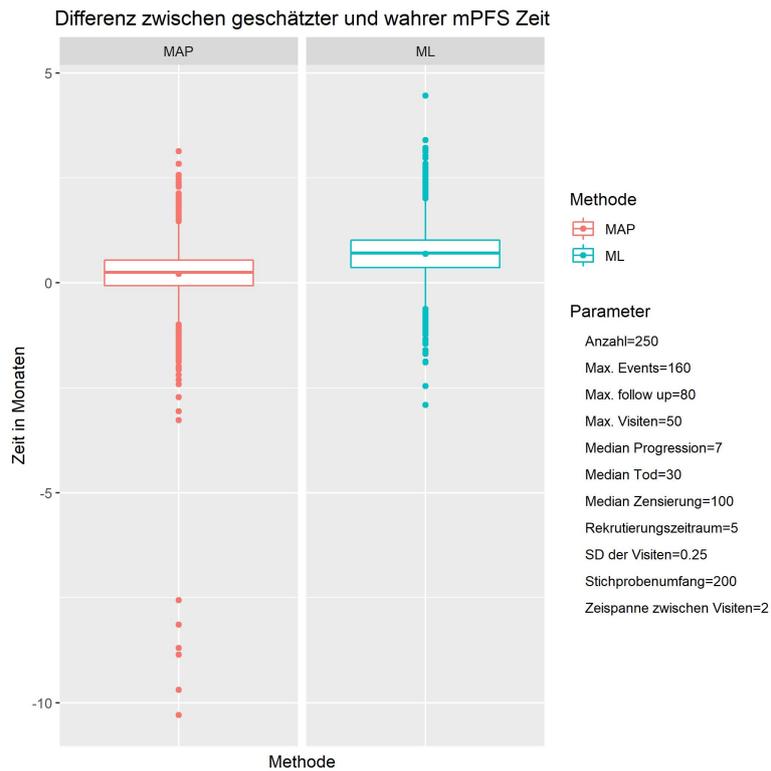


Abbildung 15: Referenzbeispiel zur Überprüfung der Ergebnisse der modifizierten MAP-Methode mit 160 anstatt 180 als maximale Anzahl an beobachteten Ereignissen

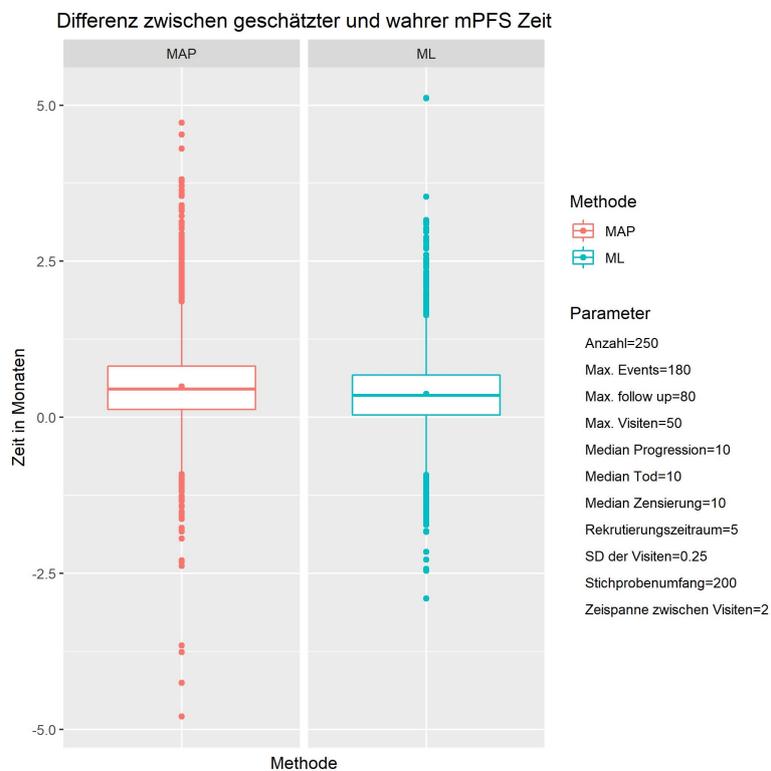


Abbildung 16: Aufgrund von identischen Raten der drei konkurrierenden Ereignissen 1. Extrembeispiel für die Ergebnisse der modifizierten MAP-Methode

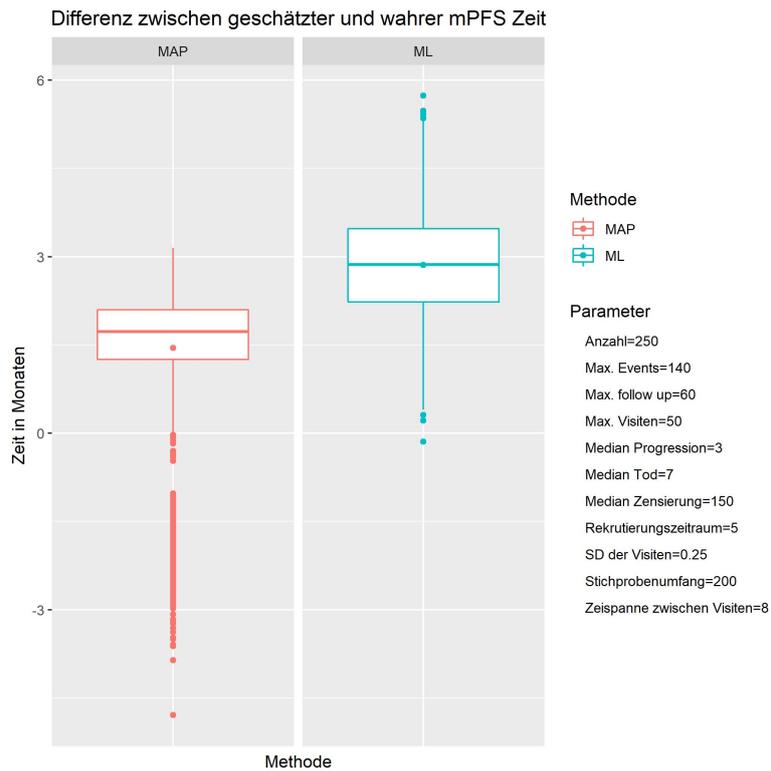


Abbildung 17: Aufgrund von einer Zeitspanne von 8 Monaten zwischen den Visiten bei geringeren Ausfallraten 2. Extrembeispiel für die Ergebnisse der modifizierten MAP-Methode

Auch bei dieser Methode lässt sich zwar eine Verbesserung im Vergleich zur ML-Schätzung erkennen, es kann jedoch nicht mehr Ungewissheit ausgeräumt werden als unter Verwendung des bedingten Erwartungswertes.

Wie Abbildung 18 aufzeigt, verbessert auch eine höhere Anzahl an Wiederholungen die Schätzung nicht.

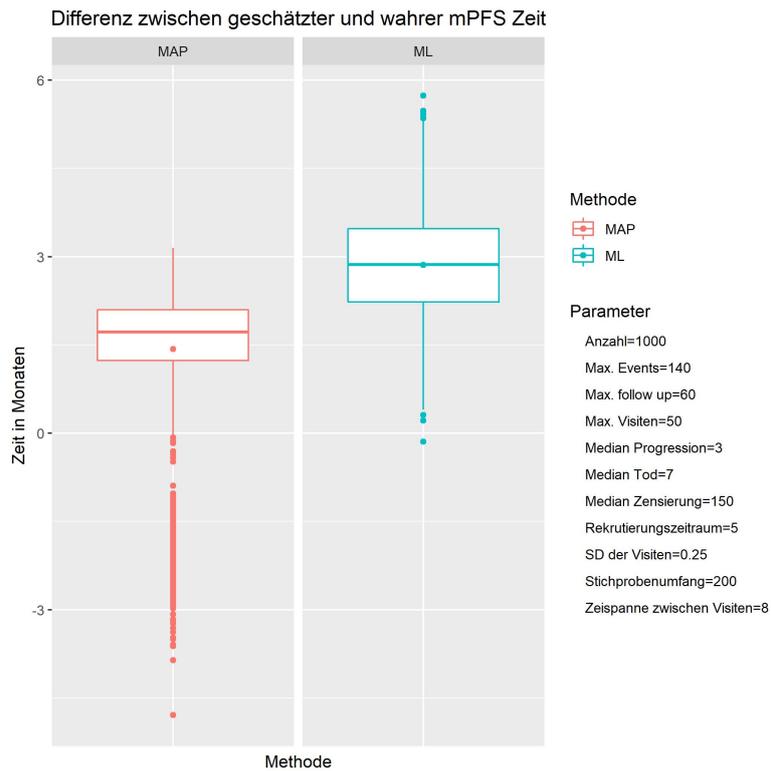


Abbildung 18: Darstellung der Ergebnisse der modifizierten MAP-Methode für das zweite Extrembeispiel mit 1000 anstelle von 250 Durchführungen

Eine Methode die bessere Ergebnisse erzielen könnte ist die *Approximate Bayesian Computation*-Methode (ABC-Methode).

## 6 Biaskorrektur durch ABC-Ansatz

In diesem Kapitel wird zu Beginn die grundlegende Idee hinter dem *Approximate Bayesian Computation*-Ansatz erläutert. Anschließend wird der Bezug und die Umsetzung zur Problemstellung erklärt und eine Analyse der Ergebnisse durchgeführt.

### 6.1 Approximate Bayesian Computation (ABC)

Die Grundzüge dieses Kapitels beruhen auf *Beaumont (2019)* [BEA-19]. Weitere Literatur stammt von *Wilkinson (2013)* [WIL-13] oder *Beaumont (2010)* [BEA-10].

Der ABC-Ansatz ist eine sogenannte *likelihood freie*-Schätzmethode. Dies bedeutet allerdings nicht, dass keinerlei Informationen der Likelihoodfunktion verwendet werden. Es wird lediglich eine explizite Auswertung umgangen. Daher gewinnt die Methodik in den letzten Jahren immer mehr Aufmerksamkeit, denn es ist, wie beispielsweise im vorigen Kapitel gezeigt wurde, für viele Problemstellungen nur schwer bis gar nicht möglich die Likelihoodfunktion auszuwerten. Bei höherdimensionalen Problemen nimmt zudem die Genauigkeit der numerischen Lösungsmethoden signifikant ab und es kann zu verzerrten Ergebnissen kommen.

Der ABC-Ansatz hat viele verschiedene Umsetzungsmöglichkeiten, welche im Folgenden chronologisch aufbauend eingeführt werden.

Die allgemeine Idee beruht auf der *Verwerfungsmethode* (engl. *rejection sampling algorithm*) zur Erzeugung von Zufallszahlen.

#### Algorithmus A: Verwerfungsmethode für Zufallszahlen

*Inputs:*

- eine Dichte  $f(\theta)$  zu der Zufallszahlen erzeugt werden sollen
- eine Dichte  $g(\theta)$  zu der sich leicht Zufallszahlen erzeugen lassen, mit  $g(\theta) > 0$  wenn  $f(\theta) > 0$
- gewünschte Anzahl an Zufallszahlen  $N$

*Sampling:*

Für  $i = 1, \dots, N$

1. Erzeuge  $\theta^{(i)} \sim g(\theta)$
2. Berechne  $p^{(i)} = \frac{f(\theta^{(i)})}{K \cdot g(\theta^{(i)})}$  mit  $K \geq \max_{\theta \in \Theta} \frac{f(\theta)}{g(\theta)}$
3. Erzeuge  $u^{(i)} \sim U(0, 1)$
4. Akzeptiere  $\theta^{(i)}$ , wenn  $u^{(i)} < p^{(i)}$ , ansonsten gehe zurück zu 1.

*Output:*

Ein Vektor  $\vec{\theta} = (\theta^{(1)}, \dots, \theta^{(N)})$  mit Zufallszahlen, welche der Dichte  $f(\theta)$  entstammen

Um diesen Algorithmus nun auf die Problemstellung der Parametersuche anzupassen, müssen ein paar kleine Änderungen vorgenommen werden.

### Algorithmus B: Verwerfungsmethode für Parameter

*Inputs:*

- eine Posterioridichte  $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ , bestehend aus der Priordichte des Parameters  $\pi(\theta)$  und einem Modell zur Datengenerierung  $p(y_{obs}|\theta)$
- eine Dichte  $g(\theta)$ , mit  $g(\theta) > 0$  wenn  $\pi(\theta|y_{obs}) > 0$
- gewünschte Anzahl an Parametern  $N$

*Sampling:*

Für  $i = 1, \dots, N$

1. Erzeuge  $\theta^{(i)} \sim g(\theta)$
2. Generiere  $y^{(i)} \sim p(y|\theta^{(i)})$
3. Erzeuge  $u^{(i)} \sim U(0, 1)$
4. Wenn  $y^{(i)} = y_{obs}$ , berechne  $p^{(i)} = \frac{\pi(\theta^{(i)})}{K \cdot g(\theta^{(i)})}$  mit  $K \geq \max_{\theta \in \Theta} \frac{\pi(\theta)}{g(\theta)}$
5. Akzeptiere  $\theta^{(i)}$ , wenn  $u^{(i)} < p^{(i)}$ , ansonsten gehe zurück zu 1.

*Output:*

Ein Vektor  $\vec{\theta} = (\theta^{(1)}, \dots, \theta^{(N)})$  mit Parametern, welche der Posterioridichte  $\pi(\theta|y_{obs})$  entstammen

Stimmen die Dichten  $\pi(\theta)$  und  $g(\theta)$  überein, so wird jeder Parameter für den  $y^{(i)} = y_{obs}$  folgt akzeptiert. Dieser Algorithmus ist allerdings nur sinnvoll, wenn die generierten Daten  $y^{(i)} \sim p(y|\theta)$  diskret sind. Sobald diese stetig sind, ist es unmöglich  $y^{(i)} = y_{obs}$  zu beobachten und damit würde nie ein Parameter akzeptiert werden.

Das Problem kann durch eine kleine Änderung in Schritt 4 des Algorithmus B angepasst werden.

### Algorithmus B': Verwerfungsmethode für Parameter

- 4'. Wenn  $\|y^{(i)} - y_{obs}\| < \varepsilon$ , berechne  $p^{(i)} = \frac{\pi(\theta^{(i)})}{K \cdot g(\theta^{(i)})}$  mit  $K \geq \max_{\theta \in \Theta} \frac{\pi(\theta)}{g(\theta)}$

Dabei sei  $\varepsilon > 0$  das Akzeptanzniveau und bei  $\|\cdot\|$  handelt es sich um ein beliebiges Distanzmaß.

Durch diese Abänderung ist klar, dass die akzeptierten Parameter  $\theta^{(i)}$  nicht länger Realisierungen der wahren Posterioridichte  $\pi(\theta|y_{obs})$ , sondern lediglich einer Approximation dieser, sind. Wie gut diese Approximation ist, lässt sich über das Akzeptanzniveau  $\varepsilon$  beeinflussen.

Die Durchführung des Algorithmus B' ist äquivalent zu einer Ziehung von Realisationspaaren  $(\theta^{(i)}, y^{(i)})$  aus der gemeinsamen Dichte

$$\hat{\pi}(\theta, y|y_{obs}) \propto \mathbb{1}_{\{\|y - y_{obs}\| < \varepsilon\}} p(y|\theta) g(\theta).$$

Wird die Realisierung durch den Algorithmus akzeptiert, so geschieht dies mit einer zu  $\frac{\pi(\theta)}{g(\theta)}$  proportionalen Wahrscheinlichkeit und damit folgt für die Dichte eines akzeptierten Realisationspaares

$$\mathbb{1}_{\{\|y-y_{obs}\|<\varepsilon\}} p(y|\theta) g(\theta) \frac{\pi(\theta)}{g(\theta)} = \mathbb{1}_{\{\|y-y_{obs}\|<\varepsilon\}} p(y|\theta) \pi(\theta).$$

Für  $\varepsilon \rightarrow 0$  entspricht die Randdichte der gemeinsamen Dichte bzgl.  $\theta$  der wahren Posterioridichte  $\pi(\theta|y_{obs})$

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \hat{\pi}_{\theta}(\theta|y_{obs}) &= \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}} \mathbb{1}_{\{\|y-y_{obs}\|<\varepsilon\}} p(y|\theta) \pi(\theta) dy \\ &= \int_{\mathbb{R}} \delta_{y_{obs}}(y) p(y|\theta) \pi(\theta) dy \\ &= p(y_{obs}|\theta) \pi(\theta) \\ &= \pi(\theta|y_{obs}) \end{aligned}$$

Wobei  $\delta_{y_{obs}}(y)$  das Diracmaß bezeichnet.

In Fällen, in denen die Daten  $y^{(i)}$  und  $y_{obs}$  nicht so einfach verglichen werden können, ist es trotz der Abschwächung durch  $\|y^{(i)} - y_{obs}\| < \varepsilon$  oft unmöglich Parameter mit einem hinnehmbaren Akzeptanzniveau  $\varepsilon$  zu beobachten. Solche Fälle können schon durch eine höhere Dimension der beobachteten Daten entstehen oder, wie in der vorliegenden Problemstellung, durch das Zusammenspiel mehrerer wichtiger Beobachtungen. Hierbei handelt es sich um die gekoppelten Ergebnisse der Ereigniszeit und des beobachteten Status jedes Patienten.

Um dieses Problem zu beseitigen, können die Informationen in einen Vektor von Statistiken  $s^{(i)} = S(y^{(i)})$  zusammengefasst werden. Dadurch ändert sich erneut Schritt 4 des Algorithmus B.

#### Algorithmus B'': Verwerfungsmethode für Parameter

4''. Wenn  $\|s^{(i)} - s_{obs}\| < \varepsilon$ , berechne  $p^{(i)} = \frac{\pi(\theta^{(i)})}{K \cdot g(\theta^{(i)})}$  mit  $K \geq \max_{\theta \in \Theta} \frac{\pi(\theta)}{g(\theta)}$

Die Substitution von  $\|y^{(i)} - y_{obs}\|$  durch  $\|s^{(i)} - s_{obs}\|$  ergibt allerdings nur Sinn, wenn es sich bei  $s$  um eine suffiziente, oder zumindest eine hochinformativ, Statistik des Parameters  $\theta$  handelt. Ist dies nicht der Fall, so wird die Approximation der Posterioridichte  $\pi(\theta|y_{obs})$  zu ungenau und die resultierenden Ergebnisse haben keinerlei Aussagekraft. Aktuell werden alle Parameter  $\theta^{(i)}$ , welche das Kriterium  $\|s^{(i)} - s_{obs}\| < \varepsilon$  erfüllen, gleich stark gewichtet. Dadurch wird nicht der maximale Informationsgehalt aus den Daten gewonnen. Der Informationsgewinn kann erhöht werden, indem die Indikatorfunktion  $\mathbb{1}_{\{\|s-s_{obs}\|<\varepsilon\}}$  durch eine Kerndichtefunktion

$$K_{\varepsilon}(u) = \frac{1}{\varepsilon} K\left(\frac{u}{\varepsilon}\right)$$

ersetzt wird. Mit Hilfe dieser werden die Anteile der akzeptierten Parameter an der Schätzung immer größer je näher  $s^{(i)}$  an  $s_{obs}$  liegt. Die Indikatorfunktion ist ein Spezialfall dieser Kerndichtefunktionen.

Die Funktion  $K(u)$  muss einige Kriterien erfüllen.

1.  $K(u)$  ist symmetrisch
2.  $K(u) \geq 0, \forall u \in \mathbb{R}$
3.  $\int_{\mathbb{R}} K(u) du = 1$
4.  $\int_{\mathbb{R}} uK(u) du = 0$
5.  $\int_{\mathbb{R}} u^2 K(u) du < \infty$

Handelt es sich bei  $s \in \mathbb{R}^m$  um einen Vektor von Statistiken, so sei  $u := s - s_{obs}$  und damit  $u_j = s_j - s_{obs,j}$ . Für die Kerndichtefunktion folgt damit

$$K_{\varepsilon}(u) = \prod_{j=1}^m K_{\varepsilon}(u_j).$$

Diese Definition muss getätigt werden, da der Wert der Distanzfunktion, aufgrund der Symmetrie der Kerndichtefunktion um 0, für jeden Wert  $u_j$  ebenfalls um 0 liegen muss. Ändert man diese Verbesserungen in Algorithmus B ab, so erhält man den ABC-Verwerfungsalgorithmus für Parameter.

### Algorithmus C: ABC-Verwerfungsmethode für Parameter

*Inputs:*

- eine Posterioridichte  $\pi(\theta|y_{obs}) \propto p(y_{obs}|\theta)\pi(\theta)$ , bestehend aus der Priordichte des Parameters  $\pi(\theta)$  und einem Modell zur Datengenerierung  $p(y_{obs}|\theta)$
- eine Dichte  $g(\theta)$ , mit  $g(\theta) > 0$  wenn  $\pi(\theta|y_{obs}) > 0$
- eine Kerndichtefunktion  $K_{\varepsilon}(u)$  mit Akzeptanzniveau  $\varepsilon > 0$
- suffiziente Statistik  $s = S(y)$
- gewünschte Anzahl an Parametern  $N$

*Sampling:*

Für  $i = 1, \dots, N$

1. Erzeuge  $\theta^{(i)} \sim g(\theta)$
2. Generiere  $y^{(i)} \sim p(y|\theta^{(i)})$
3. Berechne  $s^{(i)} = S(y^{(i)})$
4. Erzeuge  $u^{(i)} \sim U(0, 1)$
5. Berechne  $p^{(i)} = \frac{K_{\varepsilon}(\|s^{(i)} - s_{obs}\|)\pi(\theta^{(i)})}{K \cdot g(\theta^{(i)})}$  mit  $K \geq K_{\varepsilon}(0) \max_{\theta \in \Theta} \frac{\pi(\theta)}{g(\theta)}$
6. Akzeptiere  $\theta^{(i)}$ , wenn  $u^{(i)} < p^{(i)}$ , ansonsten gehe zurück zu 1.

*Output:*

Ein Vektor  $\vec{\theta} = (\theta^{(1)}, \dots, \theta^{(N)})$  mit Parametern, welche der ABC-Posterioridichte  $\pi_{ABC}(\theta|s_{obs})$  entstammen

Analog zu den Argumentationen bei Algorithmus B' lässt sich auch hier die Posterior-dichte durch

$$\pi_{ABC}(\theta, s|s_{obs}) \propto K_{\varepsilon}(\|s - s_{obs}\|)p(s|\theta)\pi(\theta)$$

und damit die Randdichte mit Hilfe von

$$\pi_{ABC_{\theta}}(\theta|s_{obs}) = \int_{\mathbb{R}} \pi_{ABC}(\theta, s|s_{obs})ds$$

ausdrücken.

Für  $\varepsilon \rightarrow 0$  folgt ebenfalls analog:

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \pi_{ABC_{\theta}}(\theta|s_{obs}) &= \lim_{\varepsilon \rightarrow 0} \int_{\mathbb{R}} K_{\varepsilon}(\|s - s_{obs}\|)p(s|\theta)\pi(\theta)ds \\ &= \int_{\mathbb{R}} \delta_{s_{obs}}(s)p(s|\theta)\pi(\theta)ds \\ &= p(s_{obs}|\theta)\pi(\theta) \\ &= \pi(\theta|s_{obs}) \end{aligned}$$

Somit geht die Approximation  $\pi_{ABC_{\theta}}(\theta|s_{obs})$  in die wahre Posterioridichte  $\pi(\theta|s_{obs})$  über und wenn es sich bei  $s$  um eine suffiziente Statistik handelt, welche damit alle Informationen über  $y_{obs}$  enthält, so gilt sogar  $\pi(\theta|s_{obs}) \equiv \pi(\theta|y_{obs})$ .

Es sei allerdings erwähnt, dass es in den meisten Fällen keine Statistik  $s$  gibt, welche sämtliche Informationen über die Daten  $y$  enthält.

Der Schätzer des Parameters errechnet sich anschließend aus dem arithmetischen Mittel der akzeptierten Parameter.

## 6.2 Anwendung auf eigentliche Problemstellung

Die Durchführung in dieser Arbeit weicht etwas von den zuvor erklärten Algorithmen ab. Da es bei ungünstiger Wahl von  $\varepsilon$  und  $g(\theta)$  eventuell sehr viele Wiederholungen benötigt bis  $N$  Parameter akzeptiert werden können, wurde eine Anzahl  $\tilde{N}$  von Simulationen durchgeführt und anschließend überprüft wie viele Parameter akzeptiert wurden. Dies führt dazu, dass nicht immer gleich viele Realisationen der Posterioridichte resultieren. Daraus ergibt sich jedoch keine Verschlechterung des Algorithmus. Darauf wird in Kapitel 6.3 genauer eingegangen.

Die  $\tilde{N}$  Simulationen werden im Folgenden als *ABC-Studien* und Simulationen zu denen der Hazard ermittelt werden soll als *Vergleichsstudien* bezeichnet.

Als Prozedur  $p(y_{obs}|\theta)$ , unter der die Daten generiert wurden, wird das, in Kapitel 4.1 erläuterte Schema, verwendet. Als Statistik  $s$  wurde der logarithmierte ML-Schätzer des Hazards verwendet. Im Laufe dieser Arbeit wurde gezeigt, dass der ML-Schätzer einer Verzerrung unterliegt, aber da die *ABC-Simulationen* genau nach dem Schema der *Vergleichsstudie* simuliert werden, unterliegen alle Schätzer der gleichen Verzerrung und sind damit vergleichbar.

Ob es sich bei der Verwendung des logarithmierten ML-Schätzers für den Hazard als  $s$  um eine suffiziente Statistik handelt ist unter der Datengenerierung mit  $p(y_{obs}|\theta)$  nur schwer zu sagen. Mit Hilfe des nachstehenden Beispiels soll veranschaulicht werden, dass es sich um eine geeignete Wahl handelt.

Analog zum Beispiel aus Kapitel 5 werden auch hier nur die Ereignisse Progression und Zensur betrachtet. Außerdem wurde nicht die in Kapitel 4.1 erläuterte Likelihood zur Datengenerierung verwendet.

Für jeden Probanden wurden Progressions- und Zensurzeiten simuliert und das Minimum der beiden bestimmt den Status. In diesem Beispiel liegt also keinerlei Verzerrung vor und der ML-Schätzer stellt einen konsistenten Schätzer für den Hazard dar. Der MAP-Schätzer stimmt für diesen Fall nahezu mit dem ML-Schätzer überein, nur lässt sich damit, neben der Parameterschätzung, die gesamte Posterioridichte veranschaulichen.

Da im Folgenden stets die Generierung des Hazards  $\lambda$  im Fokus steht, wird die Dichte  $g(\theta)$  nun mit  $g(\lambda)$  bezeichnet.

Die Parameter stimmen mit denen des Beispiels aus Kapitel 5.1 überein, wobei die Visiten vernachlässigt wurden. Als zusätzliche Parameter galt für die Priorverteilung  $\pi(\lambda) \sim \text{Gamma}(5, \frac{5.7}{\ln(2)})$ , für die Dichte aus der gezogen wurde galt  $g(\lambda) \sim U(0, 0.3)$ , da diese das notwendige Kriterium  $g(\theta) > 0$  wenn  $\pi(\theta|y_{obs}) > 0$  erfüllt, und als Kern-dichtefunktion wurde der Gaußkern

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$$

mit Akzeptanzniveau  $\varepsilon = 0.03$  verwendet.

Abbildung 19 stellt die Prior- und die MAP-Posterioridichte in blau bzw. violett dar. Zudem sind der wahre Hazard samt Schätzung durch die *Maximum-Likelihood*- und *Maximum a posteriori*-Methode eingezeichnet.

Auch hier wurde aufgrund einer Umskalierung auf die Werte der y-Achse verzichtet.

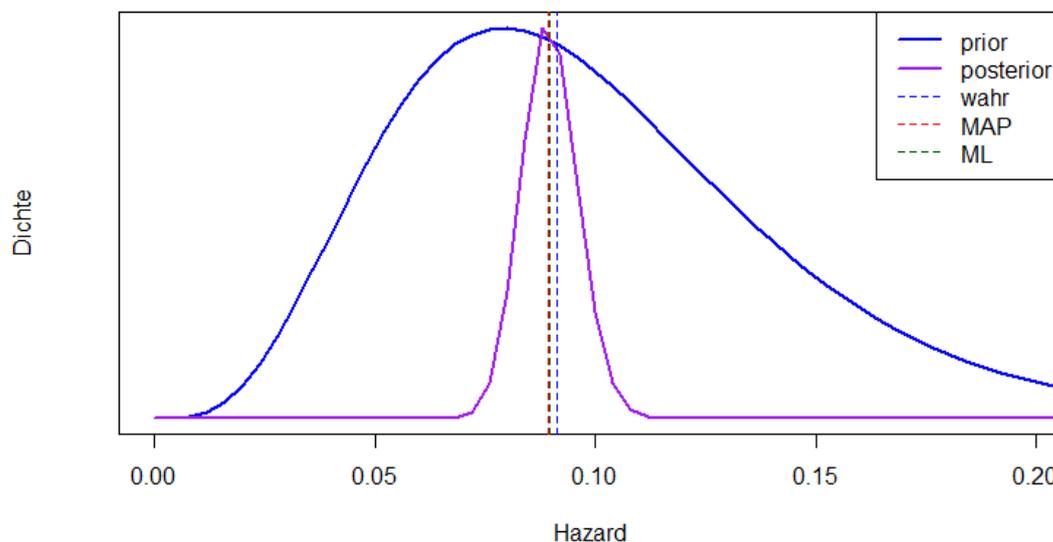


Abbildung 19: Darstellung der Ergebnisse der ML- und MAP-Methode

Wie zu erkennen ist, liegen beide Schätzungen nah am wahren Hazard. Mit diesem Ergebnis war aufgrund der Konsistenz beider Schätzer auch zu rechnen.

Abbildung 20 stellt ebenfalls die Prior- und die Posterioridichte dar, jedoch wurde letztere nun mit Hilfe des ABC-Algorithmus ermittelt.

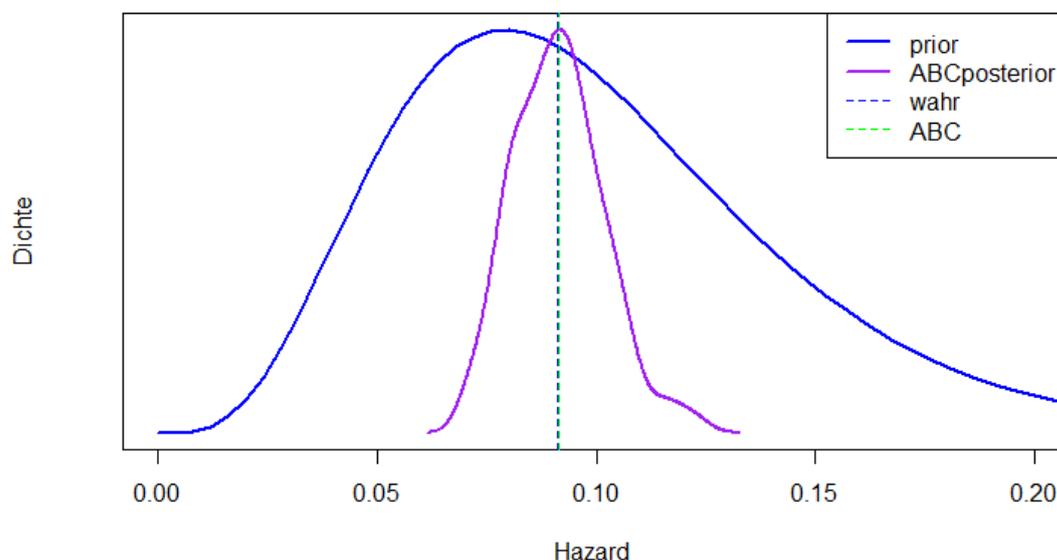


Abbildung 20: Darstellung der Ergebnisse der ABC-Methode

Es ist zu erkennen, dass der wahre Wert und die Schätzung der ABC-Methode nahezu aufeinander liegen und das, obwohl ein sehr konservatives Akzeptanzniveau von  $\varepsilon = 0.03$  verwendet wurde. Insgesamt wurden 10.000 *ABC-Simulationen* durchgeführt, von denen lediglich 178 das Kriterium  $\|s - s_{obs}\| < \varepsilon$  erfüllt haben. Wie bereits erwähnt, wurde der ABC-Schätzer als arithmetisches Mittel der akzeptierten Parameter berechnet.

Damit ist in diesem unverzerrten Beispiel die Schätzung mittels ABC-Methode sogar besser als die beiden konsistenten Schätzungen. Dies ist zwar kein Beweis für die Suffizienz des logarithmierten Hazards aber es untermauert die Wahl als Vergleichsstatistik  $s$ .

### 6.3 Sensitivitätsanalyse

In Analogie zur Sensitivitätsanalyse des Kapitels 5.2 werden die Ergebnisse dieser Methodik erst mit den gleichen sechs Beispielen veranschaulicht und anschließend werden genauere Analysen vorgenommen. Der dazugehörige Programmcode befindet sich im Anhang.

Zur genaueren Verdeutlichung der Ergebnisse beinhalten die Boxplots ihren Median als Zahlenwert auf drei Nachkommastellen gerundet. Diese Kenngröße wurde dem Mittelwert vorgezogen, da das *mediane progressionsfreie Überleben* betrachtet wird.

Für *ABCGaussian* bzw. *ABCUniform* wurden der Gauß- bzw. Gleichverteilungskern verwendet und  $g(\lambda)$  gibt an, aus welcher Verteilung die Parameter der *ABC-Simulationen* gezogen wurden.

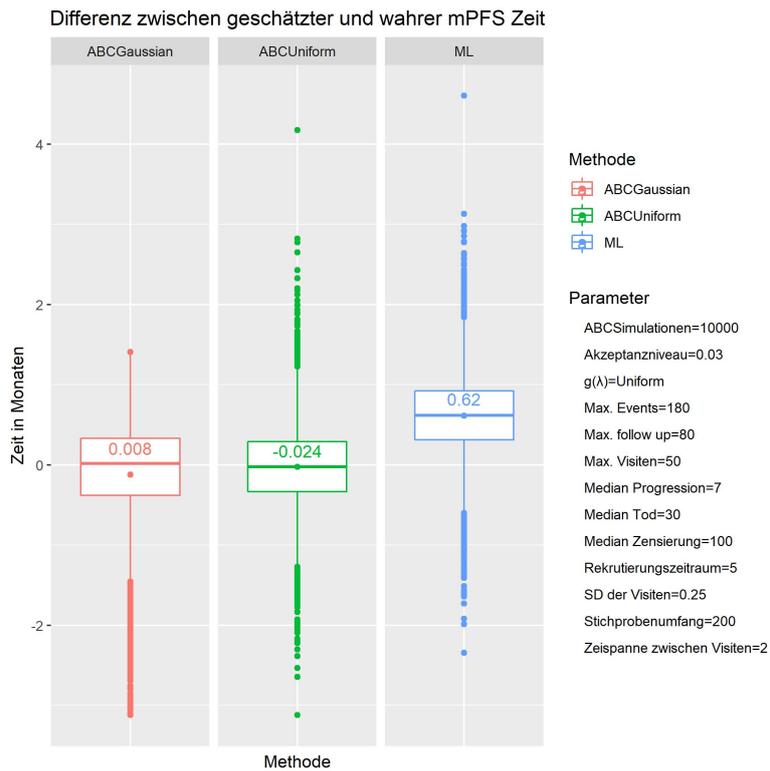


Abbildung 21: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode

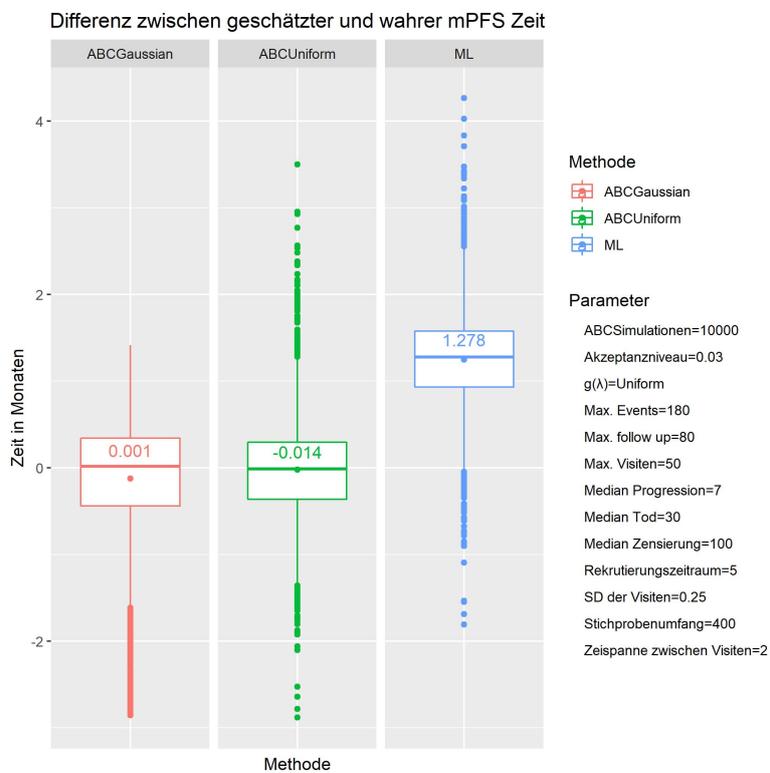


Abbildung 22: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode mit einem Stichprobenumfang von 400 anstatt 200 Probanden

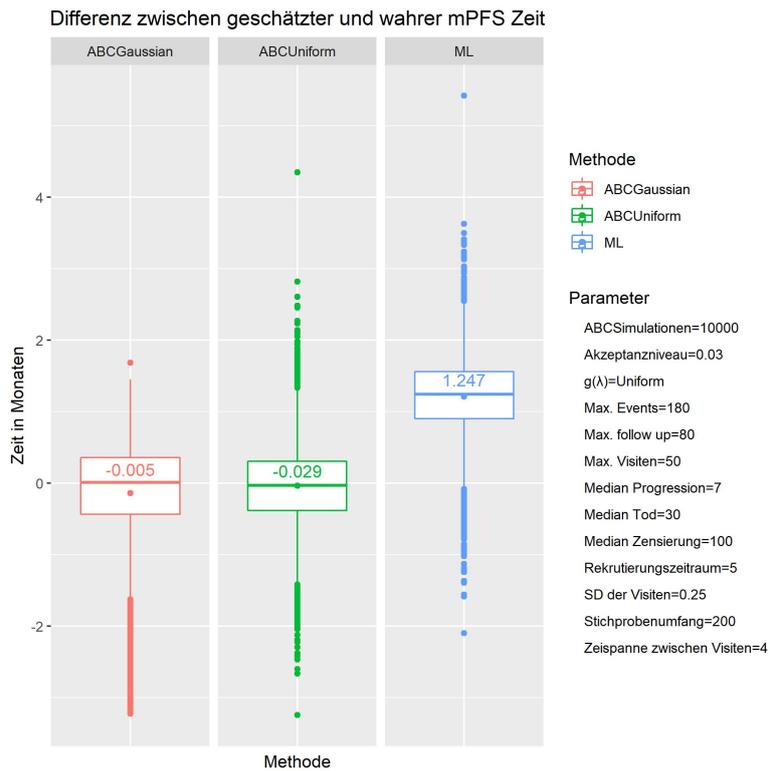


Abbildung 23: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode mit einer Zeitspanne zwischen den Visiten von 4 anstatt 2 Monaten

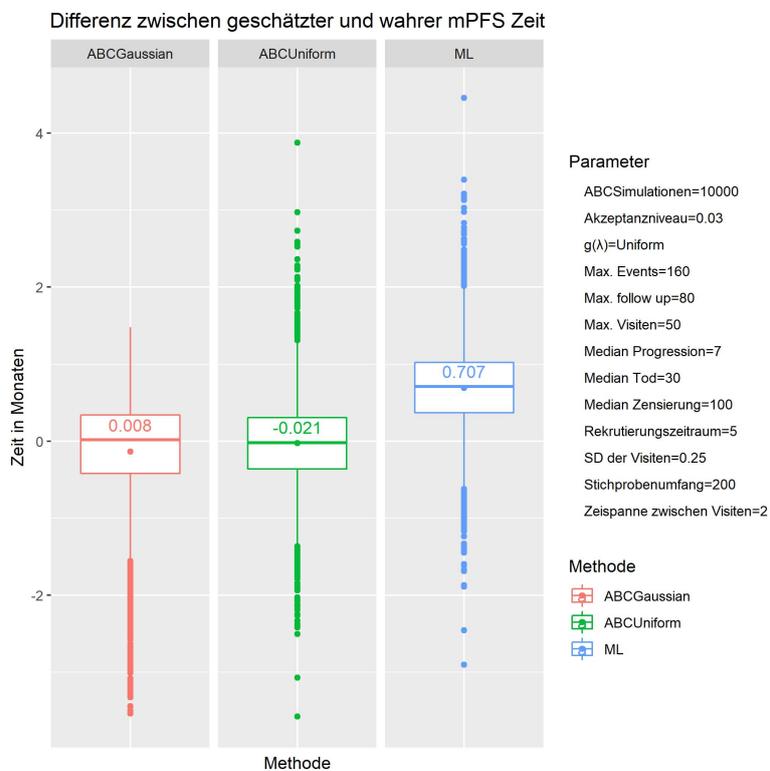


Abbildung 24: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode mit 160 anstatt 180 als maximale Anzahl an beobachteten Ereignissen

Die Boxplots des ML-Schätzers entsprechen genau jenen der Abbildungen 6 bis 9 aus Kapitel 5.2.

In den Beispielen 2 und 4 ergibt sich für den ML-Schätzer eine Verzerrung von ungefähr 1.25 Monaten, wohingegen es sich bei den Beispielen 1 und 3 „nur“ um einen Unterschied von 0.6 bis 0.7 Monaten handelt. Anhand dieser Beispiele lässt sich gut erkennen, durch welche Kenngrößen einer Studie die Verzerrung des ML-Schätzers resultiert.

Beispiel 1 bildet das Basismodell und in den übrigen drei ändert sich jeweils ein Parameter. Für das zweite Beispiel wurde der Stichprobenumfang pro Studie von 200 Probanden auf 400 erhöht und damit hat sich auch die Differenz von wahrer und geschätzter *mPFS* Zeit ungefähr verdoppelt. Selbiges gilt für das dritte Beispiel, bei welchem die Zeitspanne zwischen den Visiten von zwei auf vier Monate erhöht wurde.

Für das vierte Beispiel wurde die gewünschte Anzahl an beobachteten Events von 180 auf 160 verringert. Daraus resultierte kein signifikanter Unterschied in der Verzerrung. Betrachtet man nun die Ergebnisse der ABC-Schätzung, so fällt auf, dass diese nicht durch dieselben Einflussgrößen beeinflusst werden. Sowohl für den Gauß- als auch für den Gleichverteilungskern ergibt sich im Median ein Unterschied in der Differenz der *mPFS* Zeiten von 1 bis 14 Stunden.

Vergleicht man diese Ergebnisse mit der MAP-Methode, so stellt sich eine weitere deutliche Verbesserung ein. Gerade bei Beispiel 2 ergibt sich nun eine unverzerrte Schätzung aber auch bei allen anderen Beispielen konnten die Ergebnisse noch einmal verbessert werden.

Vergleicht man die Resultate der beiden Kerne, so lässt sich eine leichte Verbesserung unter Verwendung des Gaußkerns bemerken. Diese ergibt sich durch die unterschiedliche Gewichtung, je nachdem wie nahe die ML-Schätzer der logarithmierten Hazards beieinander liegen. Da aber schon die Schätzung mit Hilfe des Gleichverteilungskerns sehr gute Korrekturen liefert, ist dieser Unterschied nur marginal.

Auch hier sollte die resultierende Varianz durch die Schätzung nicht unterschlagen werden. Da sich durch den ML-Schätzer allerdings eine ähnliche Varianz ergibt, ist dies erneut kein Nachteil.

Im Gegensatz zur ML-Schätzung lässt sich die Varianz durch Verwendung des ABC-Schätzers sogar erklären, denn, wie bereits erwähnt, wurde auf eine Mindestanzahl an akzeptierten Parametern verzichtet. Dies kann für extreme Parameterkonstellationen eine zu kleine Menge an akzeptierten *ABC-Simulationen* zur Folge haben. Sollte diese Anzahl zu gering sein, so lässt sich keine sinnvolle Schätzung des Hazards durchführen. Extreme Parameterkonstellationen entstehen durch das Ziehen von Parametern der 5000 *Vergleichsstudien* aus der Priordichte. Dabei kann es zu Ziehungen von Parametern aus dem Rand der Verteilung kommen. Selbiges gilt auch für die Parameter der *ABC-Simulationen*, allerdings ist es dabei deutlich unkritischer, denn diese Parameter werden im schlimmsten Fall bei keiner *Vergleichsstudie* berücksichtigt.

Die Abbildungen 25 und 26 stellen die Ergebnisse der ABC-Methode für die beiden Extrembeispiele aus Kapitel 5.2 dar.

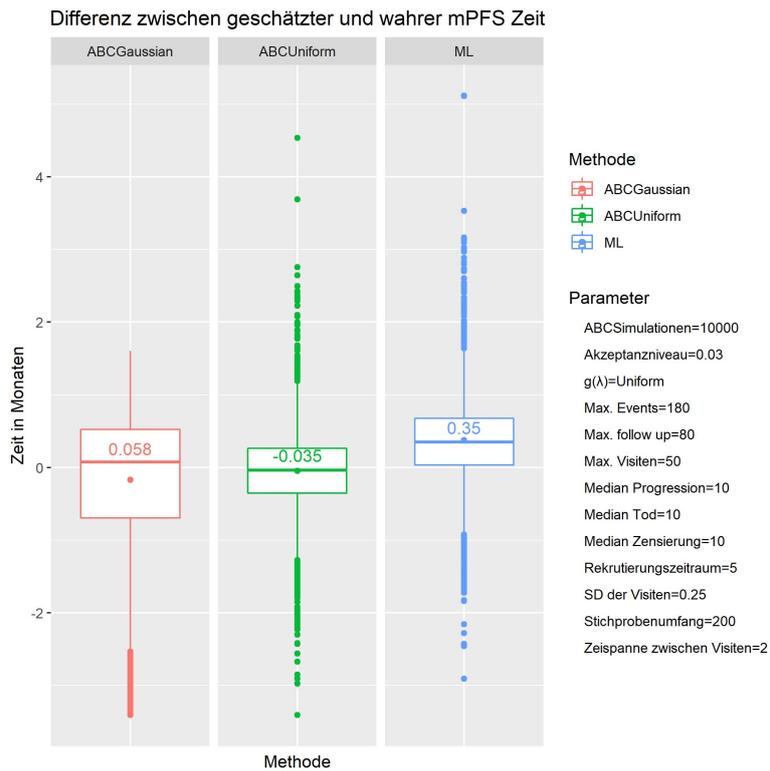


Abbildung 25: Aufgrund von identischen Raten der drei konkurrierenden Ereignissen 1. Extrembeispiel für die Ergebnisse der ABC-Methode

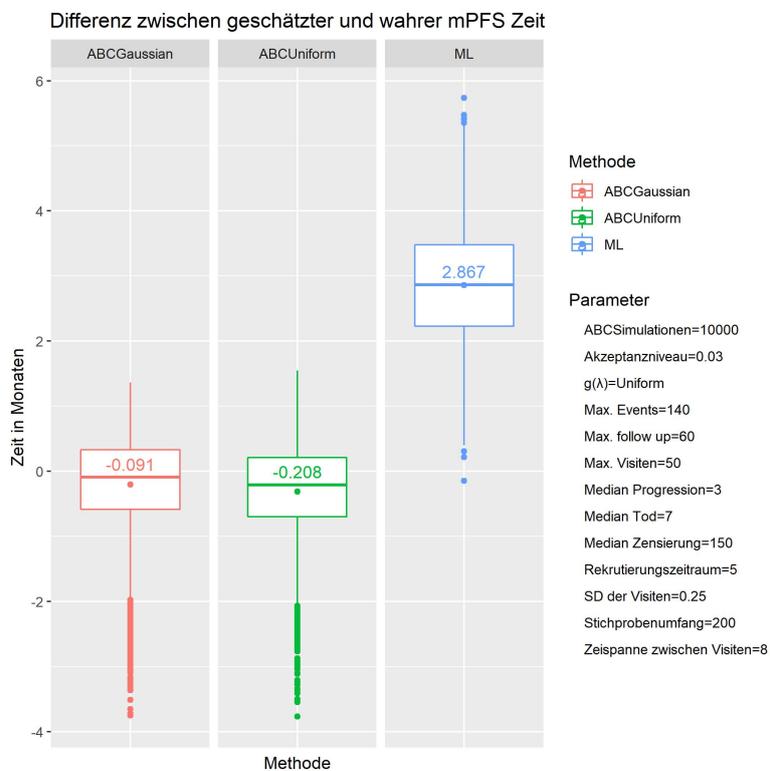


Abbildung 26: Aufgrund von einer Zeitspanne von 8 Monaten zwischen den Visiten bei geringeren Ausfallraten 2. Extrembeispiel für die Ergebnisse der ABC-Methode

Es wird ersichtlich, dass sich auch in diesen Fällen eine deutliche Verbesserung der Ergebnisse einstellt, wobei die Resultate in Beispiel 5 relativ ähnlich sind. Wie bereits erläutert wurde, liegt dies für dieses Beispiel allerdings an der geringen Verzerrung des ML-Schätzers. Jedoch konnte die Differenz von ungefähr 10 auf einen Tag verringert werden.

Der prägnanteste Unterschied tritt in Beispiel 6 auf. Dort liegt für den Median der Differenzen der *mPFS* Zeiten ein Wert von -0.208 für die Berechnung mit Gleichverteilungskerns und -0.091 unter Verwendung des Gaußkerns vor. Da die Ergebnisse in Monaten angegeben werden, handelt es sich dabei um eine Abweichung von ca. 2.8 bzw. 6.3 Tagen. Der *Maximum-Likelihood*-Schätzer dagegen überschätzt sich um etwa 2.9 Monate. Diese Verzerrung wurde bereits in Kapitel 5.2 erläutert.

Trotz der unrealistischen Konstruktion der letzten beiden Studien liefert die ABC-Methode eine klare Korrektur des Bias.

Bei den Beispielen der Abbildungen 21 bis 26 wurde die Gleichverteilung zum ziehen der Parameter verwendet. Wählt man für  $g(\lambda)$  eine Verteilung, welche ähnlicher zur Priordichte der Parameter ist, so lässt sich die Parameterschätzung nochmals verbessern.

Die Abbildungen 27 bis 32 geben die Ergebnisse unter Verwendung der Priordichte als  $g(\lambda)$  wieder.

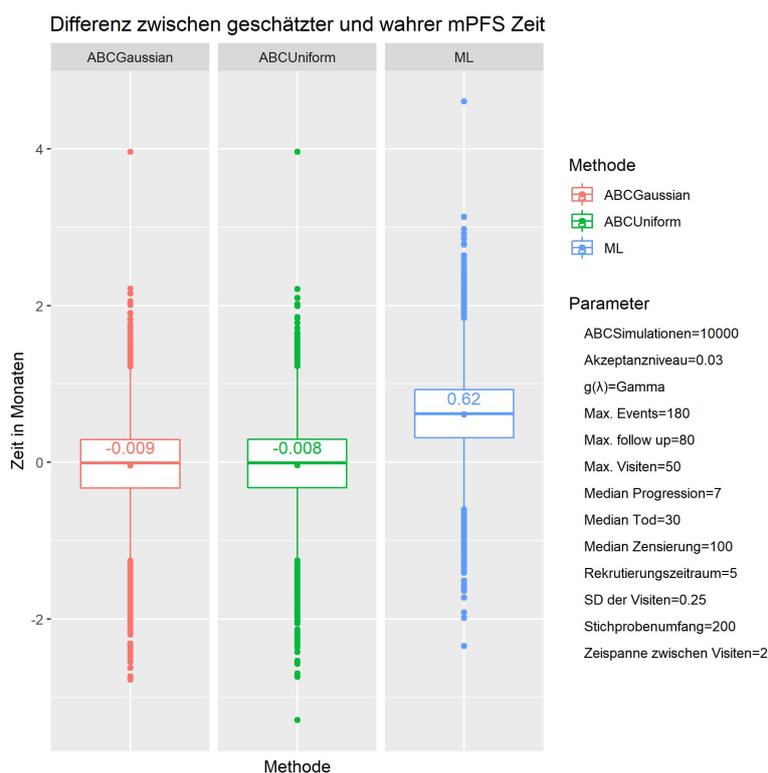


Abbildung 27: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung

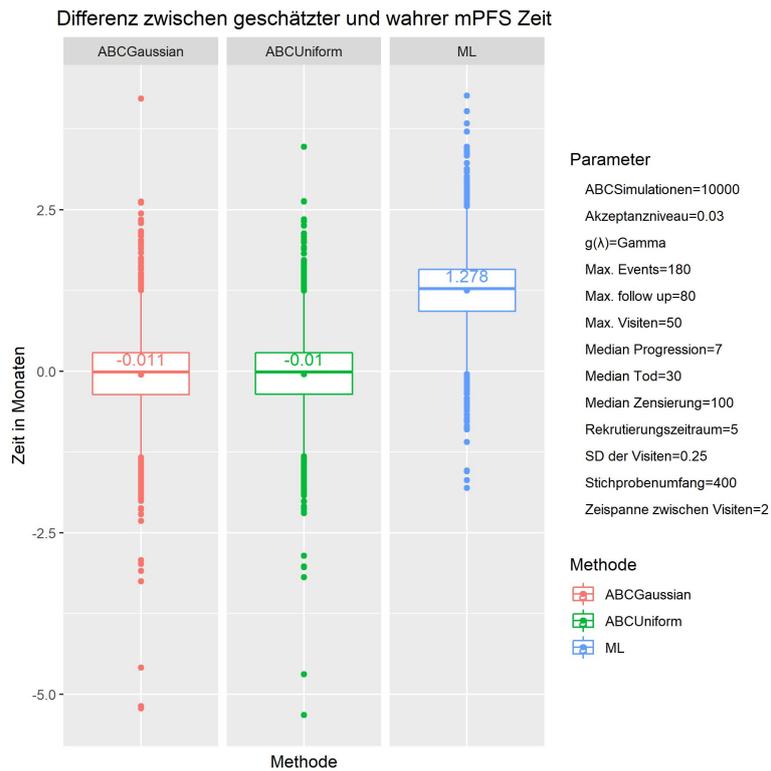


Abbildung 28: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung mit einer Zeitspanne zwischen den Visiten von 4 anstatt 2 Monaten

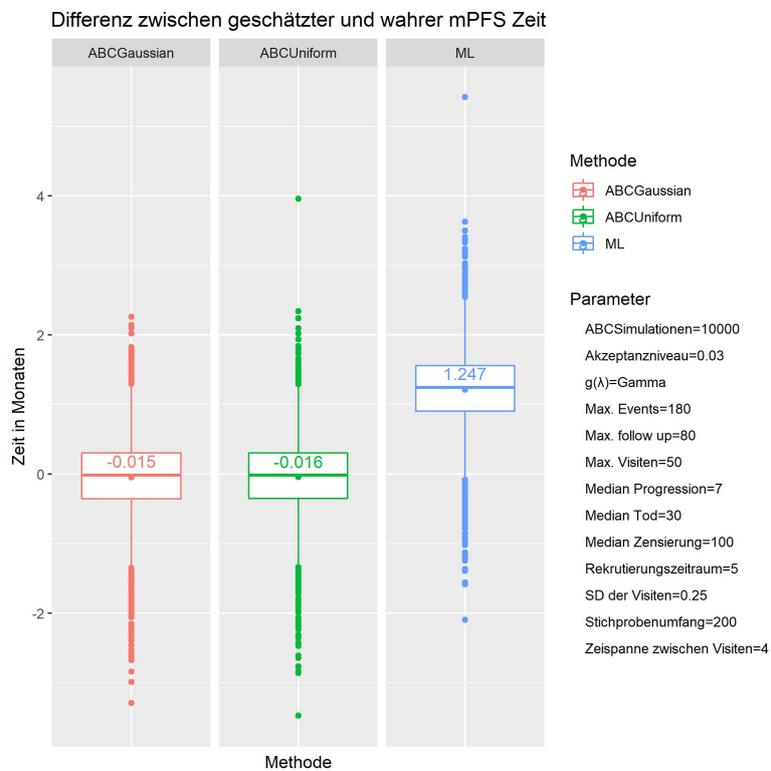


Abbildung 29: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung mit einem Stichprobenumfang von 400 anstatt 200 Probanden

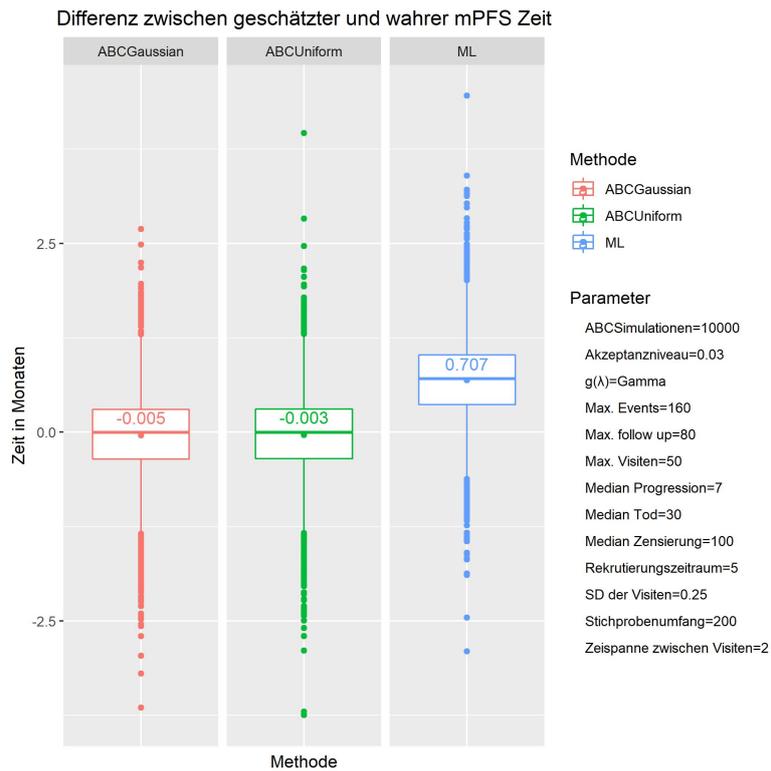


Abbildung 30: Referenzbeispiel zur Überprüfung der Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung mit 160 anstatt 180 als maximale Anzahl an beobachteten Ereignissen

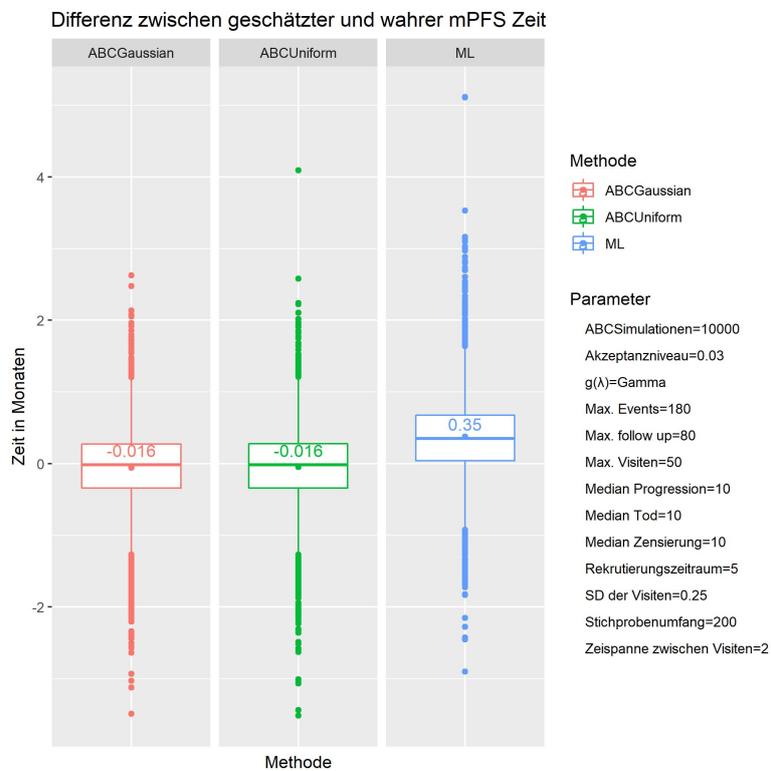


Abbildung 31: Aufgrund von identischen Raten der drei konkurrierenden Ereignissen 1. Extrembeispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung

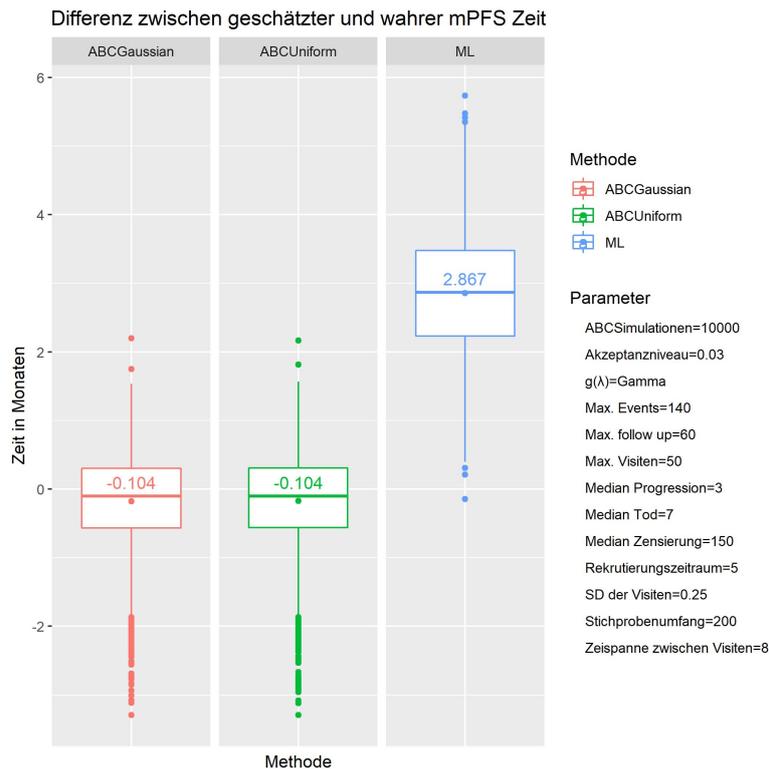


Abbildung 32: Aufgrund von einer Zeitspanne von 8 Monaten zwischen den Visiten bei geringeren Ausfallraten 2. Extrembeispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung

Verwendet man die Priordichte zum ziehen, so werden deutlich weniger Parameter durch den ABC-Algorithmus verworfen und damit resultiert eine genauere Schätzung des Hazards.

Gerade unter Verwendung des Gleichverteilungskerns wird die Schätzung nochmals sichtbar genauer.

Um die Effektivität der ABC-Methode darzustellen, werden im Folgenden weitere Beispiele betrachtet. Dabei wird, für eine kompaktere Darstellung, jede Studie vier mal unter den gleichen Parametern aber mit wachsender Zeitspanne zwischen den Visiten simuliert.

Diese Darstellung bringt gleich mehrere Vorteile mit sich, denn es wird ersichtlich,

- wie sich der Bias durch die Berechnung mit Hilfe des ML-Schätzers mit wachsendem Abstand zwischen den Visiten verhält
- dass der ABC-Schätzer keiner Verzerrung unterliegt
- wie präzise die Schätzungen sind.

Im Vergleich zu den Boxplots beinhalten die Abbildungen 33 bis 38 nur den Median der Ergebnisse. Dies führt zu einer übersichtlicheren Darstellung.

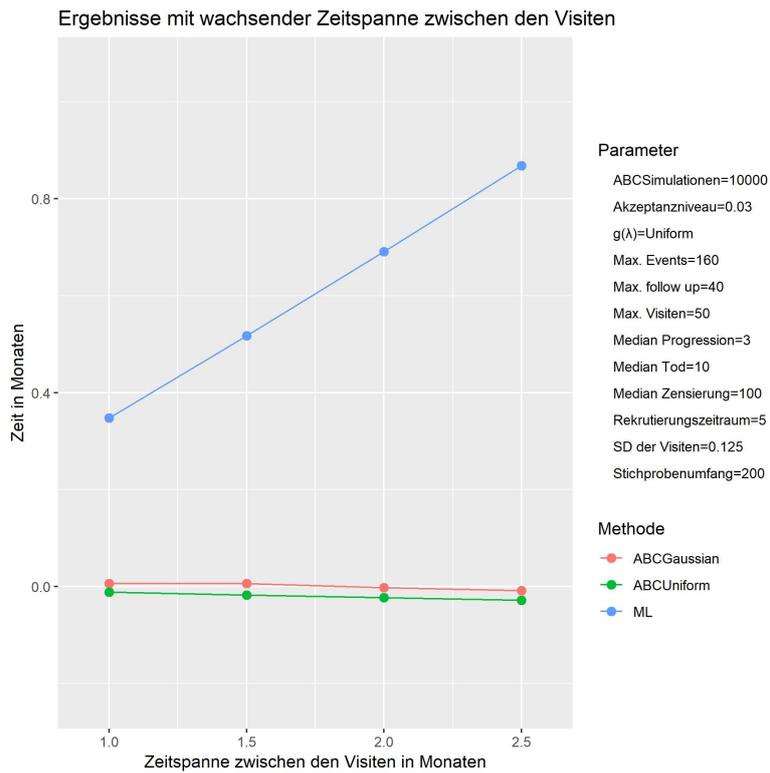


Abbildung 33: 7. Beispiel für die Ergebnisse der ABC-Methode

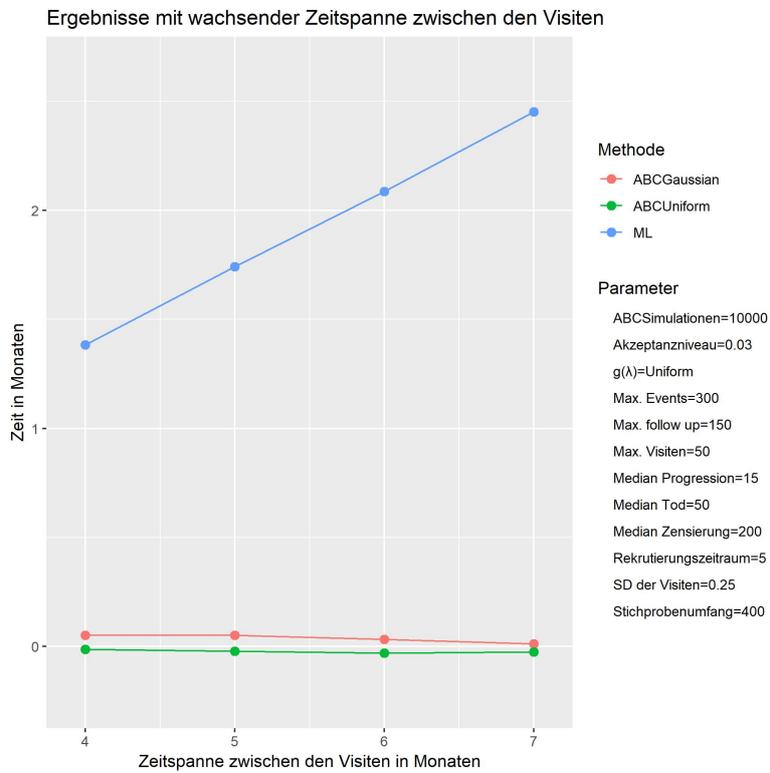


Abbildung 34: 8. Beispiel für die Ergebnisse der ABC-Methode

Der Graph aus Abbildung 33 stellt die Ergebnisse für relativ geringe Abstände zwischen den Visiten dar. Dafür wurden für die Risiken auch geringe mediane Ereigniszeiten vorgegeben.

Während sich für den ML-Schätzer ein linearer Anstieg der Verzerrung mit steigender Zeitspanne zwischen den Visiten herausstellt, korrigiert die ABC-Methode diesen nahezu vollkommen.

Der lineare Anstieg des Bias lässt sich ebenso im Beispiel der Abbildung 34 erkennen. Für dieses wurden etwas größere Visitenabstände vorgegeben, dafür aber auch die Risiken gesenkt. Zudem wurde die Probandenanzahl der Studien auf 400 Personen festgelegt. Durch diese Voraussetzungen ergibt sich für den ML-Schätzer schon bei einer Zeitspanne von 4 Monaten zwischen den Visiten eine Unterschätzung von über einem Monat in den Differenzen der *mPFS* Zeiten. Das bekräftigt die Vermutung aufgrund der vorigen Ergebnisse.

Die ABC-Methode dagegen korrigiert diese Verzerrung erneut nahezu perfekt, obwohl sich leichte Unterschiede durch die Wahl der Kerndichtefunktion ergeben.

Betrachtet man die Unterschiede genauer, so fällt auf, dass diese marginal sind. Durch die Darstellung in Monaten handelt es sich bei Verwendung des Gaußkerns um eine Abweichung von 0.064, dies entspricht ungefähr 2 Tagen. Mit Hilfe des Gleichverteilungskerns ergibt sich indes ein noch geringerer Unterschied von -0.016.

Wie das Beispiel aus Abbildung 26 gezeigt hat, erzielt aber auch die ABC-Methode nicht immer eine perfekte Biaskorrektur.

Ein weiteres Beispiel wird in Abbildung 35 aufgezeigt.

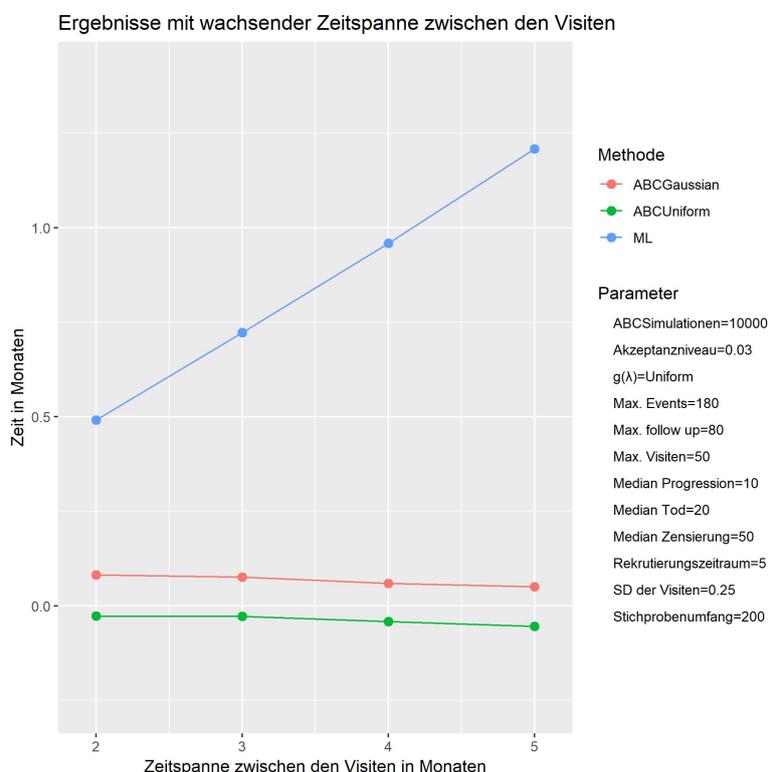


Abbildung 35: 9. Beispiel für die Ergebnisse der ABC-Methode

Die errechnete Abweichung entsteht allerdings nur bei der Wahl des Gaußkerns. Bei genauerer Betrachtung fällt auf, dass es sich dabei um eine Abweichung von ungefähr 3 Tagen handelt. Dieser Unterschied wirkt allerdings größer, da auch die Schätzung mit Hilfe der ML-Methode „nur“ eine Abweichung von 13 Tagen aufweist.

Verwendet man für selbiges Beispiel für  $g(\lambda)$  die Priorverteilung, so wird die Abweichung erneut komplett beseitigt.

Diese Korrektur ist in Abbildung 36 zu erkennen.

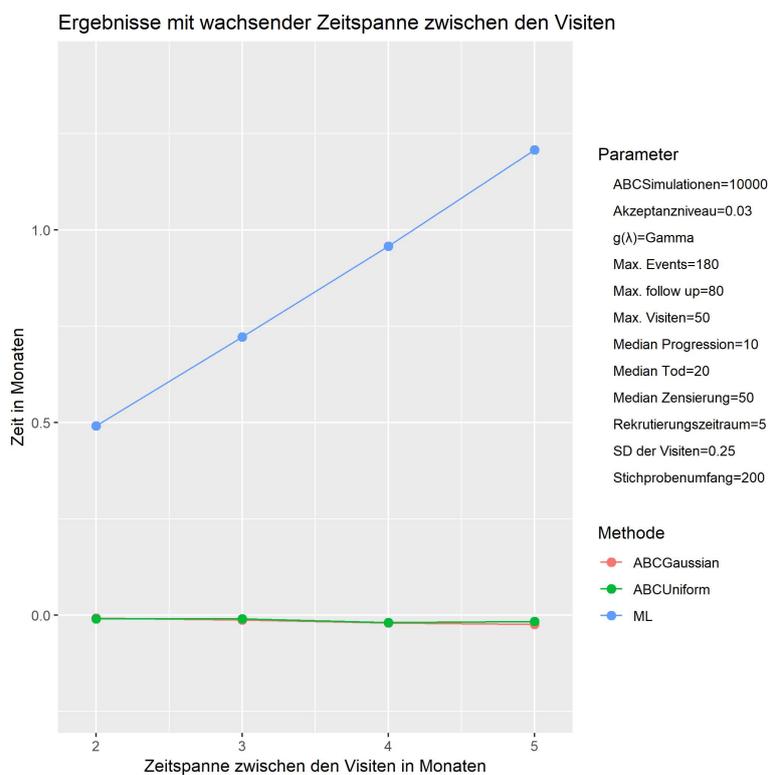


Abbildung 36: 9. Beispiel mit Parameterziehung aus der Priorverteilung

Die Abbildungen 37 und 38 zeigen zwei weitere Beispiele für die Biaskorrektur auf.

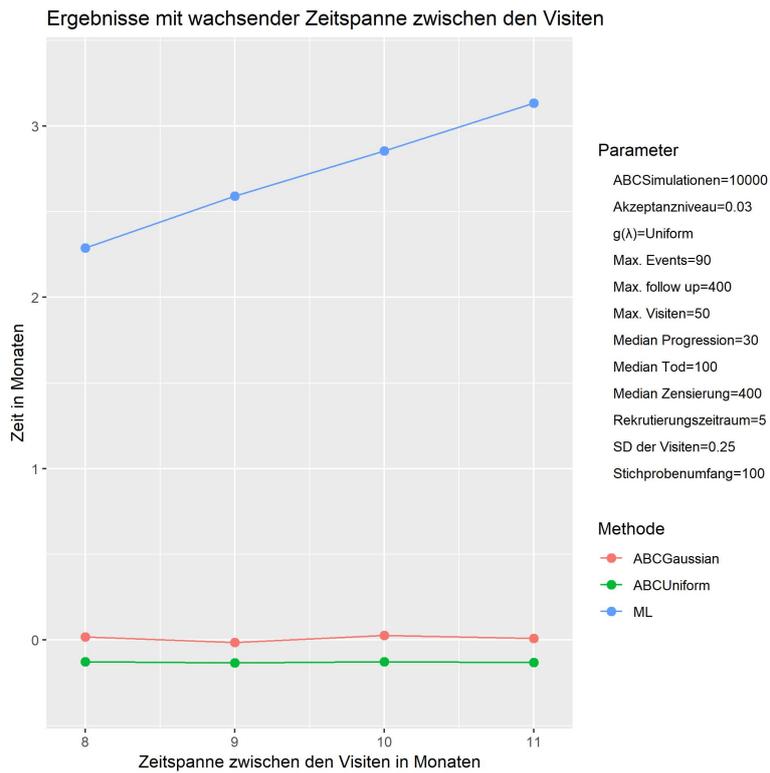


Abbildung 37: 10. Beispiel für die Ergebnisse der ABC-Methode

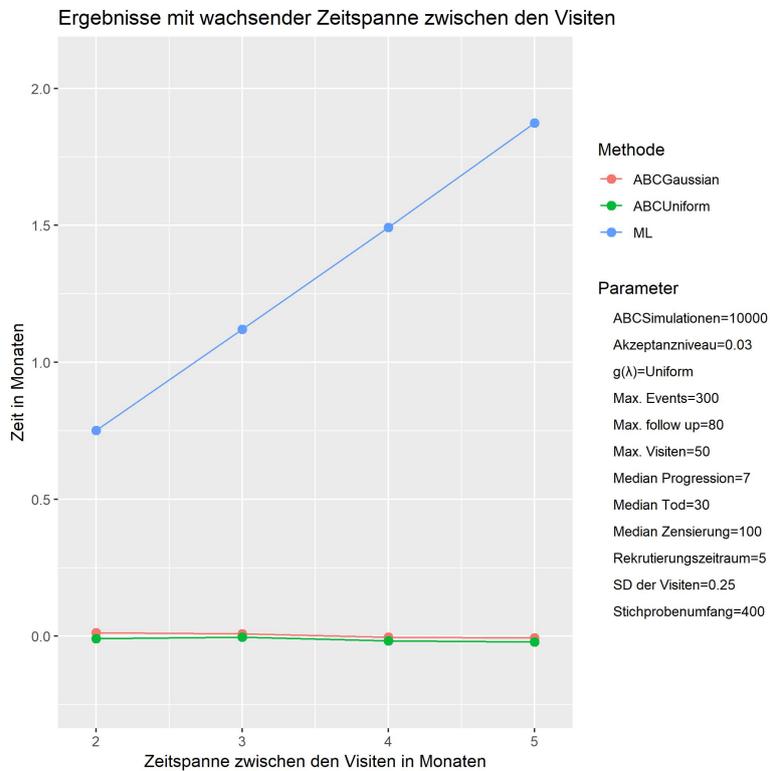


Abbildung 38: 11. Beispiel für die Ergebnisse der ABC-Methode

Auch in diesen beiden Szenarien liegt eine deutliche Korrektur des Bias vor.

## Fazit

Alles in allem lässt sich sagen, dass zwar beide, in dieser Arbeit vorgestellten, Methodiken zu einer Reduzierung des Bias geführt haben, die ABC-Methode allerdings deutlich bessere Ergebnisse hervorbrachte.

Betrachtet man erneut das Beispiel der Einleitung und ergänzt hierbei nun die mit Hilfe der ABC-Methode geschätzte Survivalfunktion, so ergibt sich der folgende Plot.

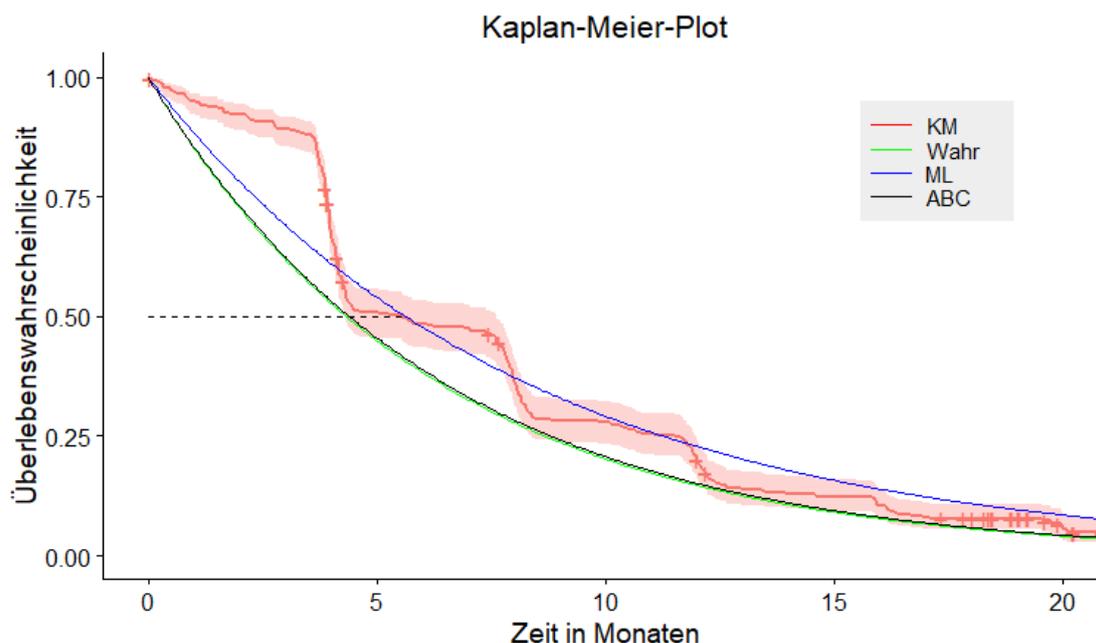


Abbildung 39: Verdeutlichung der Biaskorrektur durch die ABC-Methode

Analog zur Einleitung ergeben sich die eingezeichneten Survivalfunktionen daraus, dass die mit den unterschiedlichen Methodiken berechneten Parameter in die Survivalfunktion der Exponentialverteilung eingesetzt werden. Die Ergebnisse der MAP-Methode sind nicht dargestellt, da deren Survivalfunktion die Ergebnisse der ABC-Methode überdecken würde.

Die Priorinformationen für dieses Beispiel lauteten  $\lambda_C^{prior} = \ln(2)/130$ ,  $\lambda_P^{prior} = \ln(2)/10$  und  $\lambda_D^{prior} = \ln(2)/30$ . Aus den selben Verteilungen wurden die Parameter der *ABC-Simulationen* gezogen. Als Kerndichtefunktion wurde der Gleichverteilungskern mit Akzeptanzniveau  $\varepsilon = 0.03$  verwendet. Als wahre *mPFS* Zeit ergab sich ein Wert von  $\approx 4.34$ , wobei die Schätzung mit Hilfe der ABC-Methode eine *mPFS* Zeit von  $\approx 4.38$  ergab. Damit stellt sich auch hier eine Verbesserung in der Schätzung im Vergleich zur ML-Methode ein. Die Differenz von ungefähr 1.3 Monaten in der *mPFS* Zeit konnte auf etwas mehr als einen Tag verringert werden.

Neben der Korrektur des Bias hat die Verwendung der ABC-Methode den Vorteil, dass zum Vergleich zweier unkontrollierter Studien nur der ML-Schätzer des Hazards und das Studienprotokoll benötigt werden. Damit können die Ergebnisse neuer Studien auch mit Studien verglichen werden, bei denen nicht die gesamten Daten veröffentlicht wurden.

Als Ausblick für weitere Arbeiten auf diesem Gebiet könnten anstelle der Exponentialverteilung allgemeinere Verteilungen für die Ausfallzeiten, wie beispielsweise die Weibull-, Gamma- oder eine Gompertzverteilung, betrachtet werden. Für exponentialverteilte Ereigniszeitdaten hat die ABC-Methodik zu einer enormen Biaskorrektur geführt. Dieses Verhalten ist für die anderen Verteilungen allerdings noch nicht untersucht worden und da Ereigniszeitdaten in der Praxis nicht immer einer Exponentialverteilung folgen, können die drei anderen Verteilungen, durch ihre höhere Flexibilität, Abweichungen von der Verteilungsannahme besser korrigieren.

## Anhang

Listing 1: Code zur Durchführung der ABC-Methode

```
1 library(doParallel)
2 library(parallel)
3 library(MASS)
4 library(survival)
5 library(survminer)
6 library(profvis)
7 library(tidyr)
8 library(Rfast)
9 kern<-detectCores()-1
10 cl<-makeCluster(kern)
11 registerDoParallel(cl)
12 #y=Anzahl der Vergleichsstudien
13 #x=Anzahl der ABC-Studien
14 #n=Stichprobenumfang der Studien
15 #maxvis=Maximale Anzahl an Visiten
16 #maxfollowup=Maximale Follow-up Zeit
17 #recruit=Laenge der Rekrutierungsperiode
18 #obsevents=Beobachtete Anzahl an Events fuer cutoff
19 #lambdacensored=Priorinformationen ueber Hazard der
    Zensierung
20 #lambdaprogession=Priorinformationen ueber Hazard der
    Progression
21 #lambdadeath=Priorinformationen ueber Hazard des Todes
22 #sdprog=Standardabweichung um die eigentlichen
    Visitentermine
23 #kontrollintervall=Zeitspanne zwischen den Visiten
24 #delta=Akzeptanzniveau
25 #gamma=Wenn gamma=T werden die Parameter der ABC-
    Simulationen aus der Priorverteilung gezogen
26 ABCrejection<-function(y, x, n, maxvis,
27 maxfollowup, recruit,
28 obsevents, prioranzcensored=5,
29 prioranzprog=5, prioranzdeath=5,
30 lambdacensored, lambdaprogession,
31 lambdadeath, sdprog, kontrollintervall,
32 delta, gamma)
33 {
34 is.integer0 <- function(x)
35 {
36 is.integer(x) && length(x) == 0L
37 }
38 viscen<-function(cen, vis) #Funktion, um tatsaechlichen
    Zensierungszeitpunkt zu ermitteln
39 {
40 vis[!(vis<cen)]<-1/30.4375 #wenn Visitenzeitpunkt gleich
```

```

0 ware setze diesen auf 1/30.4375
41 rowMaxs(vis, value = T)
42 }
43 visprog<-function(prog,vis) #Funktion, um tatsaechlichen
Progressionszeitpunkt zu ermitteln
44 {
45 vis[!(vis>prog)]<-1000
46 rowMins(vis, value = T)
47 }
48 set.seed(42)
49 visit<-kontrollintervall*matrix(rep(1:maxvis,each=n),n,
maxvis) #Matrix fuer n Probanden mit allen
eigentlichen Visitenzeitpunkten
50 varvi<-matrix(rnorm(n*maxvis,sd=sdprog),n,maxvis) #
Erzeugung der Zufallsschocks fuer Visitenzeitpunkte
51 actualvis<-visit+varvi #Wahre Visitenzeitpunkte
52
53 lambdacensoredgamma<-rgamma(y,shape=prioranzcensored,
rate=prioranzcensored/lambdacensored) #Ziehung der
Parameter f???r Zensierung
54 lambdaprogressiongamma<-rgamma(y,shape=prioranzprog,rate
=prioranzprog/lambdaprogression) #Ziehung der
Parameter f???r Progression
55 lambdadeathgamma<-rgamma(y,shape=prioranzdeath,rate=
prioranzdeath/lambdadeath) #Ziehung der Parameter f
???r Tod
56
57 censored<-matrix(rexp(n*y,lambdacensoredgamma),ncol=y,
byrow=T) #erzeugung der Zensierungszeiten
58 progression<-matrix(rexp(n*y,lambdaprogressiongamma),
ncol=y, byrow=T) #erzeugung der Progressionszeiten
59 death<-matrix(rexp(n*y,lambdadeathgamma),ncol=y,byrow=T)
#erzeugung der Todeszeiten
60
61 random<-runif(n,0,recruit) #Zufallsschocks der
Rekrutierung
62
63 censored <- foreach(j=1:y, .combine=cbind, .packages="
Rfast") %dopar% {
64 viscen(cen=censored[,j],vis=actualvis)
65 }
66 progression <- foreach(j=1:y, .combine=cbind, .packages="
Rfast") %dopar% {
67 visprog(prog=progression[,j],vis=actualvis)
68 }
69
70 simMinzeiten<-pmin(censored,progression,death) #
Tatsaechlichen Status des Probanden ermitteln
71

```

```

72 vali2<-(simMinzeiten==progression)|(simMinzeiten==death)
73 simMinzeiten2<-simMinzeiten+random
74
75 tel<-simMinzeiten2*vali2
76 te.cutoff<-apply (tel,2,function(x,obsdata, follow){a<-
  which(rank(x[x!=0])==obsdata);if(is.integer0(a)==T)
  follow else (x[x!=0])[a]},obsdata=obsevents, follow=
  maxfollowup)
77 te.cutoff<-pmin(te.cutoff,colMaxs(simMinzeiten2,value =
  T))
78 te.vali2<-vali2*t(t(simMinzeiten2)<=te.cutoff)
79 simMinzeiten[(t(t(simMinzeiten2)>=te.cutoff))==1]<- (
  matrix(rep(te.cutoff,times=n),nrow = n,byrow=T)-
  matrix(rep(random,times=y), ncol=y,byrow=F))[(t(t(
  simMinzeiten2)>=te.cutoff))==1]
80
81
82 hazdata<-(colSums(te.vali2)/colSums(simMinzeiten)) #
  Hazard fuer Progression oder Tod
83 MLcen<-(colSums(!te.vali2)/colSums(simMinzeiten)) #ML-
  Schaetzer fuer den Hazard der Zensierung
84 MLprog<-(colSums((simMinzeiten==(progression)))/colSums(
  simMinzeiten)) #ML-Schaetzer fuer den Hazard der
  Progression
85 MLdeath<-(colSums((simMinzeiten==(death)))/colSums(
  simMinzeiten)) #ML-Schaetzer fuer den Hazard des
  Todes
86
87 simdata<-cbind(lambdacensoredgamma,
  lambdaprogressiongamma, lambdadeathgamma, MLcen,
  MLprog, MLdeath)
88
89 if(gamma==T)#Wenn ABC-Parameter aus der Priorverteilung
  stammen
90 {
91 approxlambdacensored<-rgamma(x,shape=prioranzcensored,
  rate=prioranzcensored/lambdacensored)#Ziehung der
  Zensierungsparameter fuer ABC-Simulation
92 approxlambdaprog<-rgamma(x,shape=prioranzprog,rate=
  prioranzprog/lambdaprogression)#Ziehung der
  Progressionsparameter fuer ABC-Simulation
93 approxlambdadeath<-rgamma(x,shape=prioranzdeath,rate=
  prioranzdeath/lambdadeath)#Ziehung der
  Todesparameter fuer ABC-Simulation
94 } else {#Wenn ABC-Parameter nicht aus der
  Priorverteilung stammen
95 approxlambdacensored<-runif(x,0,2*lambdacensored)#
  Ziehung der Zensierungsparameter fuer ABC-Simulation
96 approxlambdaprog<-runif(x,0,2*lambdaprogression)#

```

```

    Ziehung der Progressionsparameter fuer ABC-Simulation
97 approxlambdadeath<-runif(x,0,2*lambdadeath) #Ziehung der
    Todesparameter fuer ABC-Simulation
98 }
99
100 censoredzeiten<-matrix(rexp(n*x,approxlambdaacensored),
    ncol=x,byrow=T) #Simulation aller Zensierungszeiten
101 progzeiten<-matrix(rexp(n*x,approxlambdaaprog),ncol=x,
    byrow=T) #Simulation aller Progressionszeiten
102 deathzeiten<-matrix(rexp(n*x,approxlambdaadeath),ncol=x,
    byrow=T) #Simulation aller Todeszeiten
103
104 censoredzeiten <- foreach(j=1:x, .combine=cbind, .
    packages="Rfast") %dopar% { #Anpassung der
    Zensierungszeiten an Visitenschema
105 viscen(cen=censoredzeiten[,j],vis=actualvis)
106 }
107
108 progzeiten <- foreach(j=1:x, .combine=cbind, .packages="
    Rfast") %dopar% { #Anpassung der Progressionszeiten
    an Visitenschema
109 visprog(prog=progzeiten[,j],vis=actualvis)
110 }
111
112 minzeiten<-pmin(censoredzeiten,progzeiten,deathzeiten) #
    Ermittlung des eigentlichen Status
113
114 vali1<-(minzeiten==deathzeiten)|(minzeiten==progzeiten)
115 minzeiten2<-minzeiten+random
116
117 te2<-minzeiten2*vali1
118 te.cutoff2<-apply(te2,2,function(x,obsdata, follow){a<-
    which(rank(x[x!=0])==obsdata);if(is.integer0(a)==T)
    follow else (x[x!=0])[a]},obsdata=obsevents, follow=
    maxfollowup)
119 te.cutoff2<-pmin(te.cutoff2,colMaxs(minzeiten2,value = T
    ))
120 te.vali1<-vali1*t(t(minzeiten2)<=te.cutoff2)
121 minzeiten[(t(t(minzeiten2)>=te.cutoff2))==1]<-(matrix(
    rep(te.cutoff2,times=n),nrow = n,byrow=T)-matrix(rep(
    random,times=x), ncol=x,byrow=F))[(t(t(minzeiten2)>=
    te.cutoff2))==1]
122
123 haz<-colSums(te.vali1)/colSums(minzeiten) #Hazard fuer
    Progression oder Tod der ABC-Simulationen
124
125 ABCdata<-cbind(approxlambdaacensored, approxlambdaaprog,
    approxlambdaadeath, haz)
126 #Funktion um den Schaetzer mit Hilfe des

```

```

    Gleichverteilungskerns zu scheatzen
127 post<-function(hazard, ABCdat, grenze)
128 {
129 a1<-mean(ABCdat[(-grenze<(log(ABCdat[,4]/hazard))&((log
    (ABCdat[,4]/hazard))<(grenze)),1])#Mittel aller
    akzeptierten Zensurparameter
130 a2<-mean(ABCdat[(-grenze<(log(ABCdat[,4]/hazard))&((log
    (ABCdat[,4]/hazard))<(grenze)),2])#Mittel aller
    akzeptierten Progressionsparameter
131 a3<-mean(ABCdat[(-grenze<(log(ABCdat[,4]/hazard))&((log
    (ABCdat[,4]/hazard))<(grenze)),3])#Mittel aller
    akzeptierten Todesparameter
132 a4<- length(ABCdat[(-grenze<(log(ABCdat[,4]/hazard))&((
    log(ABCdat[,4]/hazard))<grenze),3]) #Anzahl
    akzeptierter Parameter
133 c(a1,a2,a3,a4) #Wenn ja wird der Mittelwert jeder
    einzelnen zurueckgegeben
134 }
135 d<-matrix()
136 d <- foreach(i=1:length(hazdata), .combine=rbind) %dopar
    % { #Anwendung der Funktion "post" auf alle der y
    eigentlichen Simulationen
137 post(hazard=hazdata[i], ABCdat = ABCdata, grenze = delta
    )
138 }
139 #Schaetzung des Parameters mit Hilfe des Gausskerns
140 if(gamma==F)#Wenn die Parameter der ABC-Simulationen
    nicht aus der Priorverteilung gezogen wurden
141 {
142 xmax<-(prioranzprog-1)/(prioranzprog/lambdaproggression)
    #ist Max wegen Gleichverteilung
143 k<-1/delta*dnorm(0)*dgamma(xmax,shape=prioranzprog,rate=
    prioranzprog/lambdaproggression)/dunif(xmax,0,2*
    lambdaproggression)
144
145 versuch<-matrix()
146 versuch<-foreach(i=1:y,.combine=rbind) %dopar% {
147 p<-1/delta*dnorm(log(hazdata[i]/haz)/delta)*dgamma(
    approxlambdaprog,shape=prioranzprog,rate=prioranzprog
    /lambdaproggression)/(k*dunif(approxlambdaprog,0,2*
    lambdaproggression))
148 zz<-runif(length(p),0,1)
149 wahr<-approxlambdaprog[zz<p]
150 mean(wahr)
151 }
152
153 xmax1<-(prioranzdeath-1)/(prioranzdeath/lambdadeath) #
    ist Max wegen Gleichverteilung
154 k1<-1/delta*dnorm(0)*dgamma(xmax1,shape=prioranzdeath,

```

```

    rate=prioranzdeath/lambdadeath)/dunif(xmax1,0,2*
    lambdadeath)
155
156 versuch1<-matrix()
157 versuch1<-foreach(i=1:y,.combine=rbind) %dopar% {
158 p<-1/delta*dnorm(log(hazdata[i]/haz)/delta)*dgamma(
    approxlambdadeath,shape=prioranzdeath,rate=
    prioranzdeath/lambdadeath)/(k1*dunif(
    approxlambdadeath,0,2*lambdadeath))
159 zz<-runif(length(p),0,1)
160 wahr<-approxlambdadeath[zz<p]
161 mean(wahr)
162 }
163 } else {
164 xmax<-(prioranzprog-1)/(prioranzprog/lambdaprogession)
    #ist Max wegen Gleichverteilung
165 k<-1/delta*dnorm(0)
166
167 versuch<-matrix()
168 versuch<-foreach(i=1:y,.combine=rbind) %dopar% {
169 p<-1/delta*dnorm(log(hazdata[i]/haz)/delta)/(k)
170 zz<-runif(length(p),0,1)
171 wahr<-approxlambdaprog[zz<p]
172 mean(wahr)
173 }
174
175 xmax1<-(prioranzdeath-1)/(prioranzdeath/lambdadeath) #
    ist Max wegen Gleichverteilung
176 k1<-1/delta*dnorm(0)
177
178 versuch1<-matrix()
179 versuch1<-foreach(i=1:y,.combine=rbind) %dopar% {
180 p<-1/delta*dnorm(log(hazdata[i]/haz)/delta)/(k1)
181 zz<-runif(length(p),0,1)
182 wahr<-approxlambdadeath[zz<p]
183 mean(wahr)
184 }
185 }
186 erg<-cbind(simdata,d,versuch[,1],versuch1[,1])
187 }

```

## Literatur

- [BEA-10] Mark Beaumont  
*Approximate Bayesian Computation in Evolution and Ecology, Annual Review of Ecology, Evolution, and Systematics, Vol. 41 (2010), Seiten 379-406*
- [BEA-19] Scott Sisson, Yanan Fan, Mark Beaumont  
*Handbook of Modern Statistical Methods - Handbook of Approximate Bayesian Computation, Broken Sound Parkway NW: Chapman & Hall/CRC, 2019*
- [DEU-04] Peter Deuffhard  
*Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms., Springer, Berlin 2004*
- [KAP-58] E. L. Kaplan, Paul Meier  
*Nonparametric estimation from incomplete observations, Journal of the American Statistical Association, Vol. 53, No. 282 (Jun., 1958), Seiten 457-481*
- [KLE-03] John P. Klein, Melvin L. Moeschberger  
*Survival Analysis - Techniques for Censored and Truncated Data, 2. Edition, New York: Springer 2003, Seite 21ff*
- [KRE-13] Małgorzata Krętowska  
*Dipolar Tree Ensemble With and Without Adjustment to Competing Risks: Application to Medical Data, Studies in logic, grammar and rhetoric 35 (2013), Seite 28*
- [SCH-08] Martin Schumacher, Gabi Schulgen  
*Methodik klinischer Studien. 3rd edition. Berlin, Heidelberg, New York: Springer 2008, Seite 79, Abbildung 2*
- [WIL-13] Richard David Wilkinson  
*Approximate Bayesian computation (ABC) gives exact results under the assumption of model error; De Gruyter, Statistical Applications in Genetics and Molecular Biology 2013; 12(2): Seiten 129–141*
- [ZWI-11] Isabella Zwiener, Maria Blettner, Gerhard Hommel  
*Überlebenszeitanalyse: Teil 15 der Serie zur Bewertung wissenschaftlicher Publikationen, Deutsches Ärzteblatt, Jg. 108, Heft 10, 11. März 2011, Seite 165f.*

## Abbildungsverzeichnis

1	Verdeutlichung des Bias durch Intervallzensurierung . . . . .	1
2	Ablauf einer klinischen Studie . . . . .	3
3	Darstellung konkurrierender Risiken . . . . .	13
4	Verdeutlichung der Intervallunabhängigkeit der MAP-Methode (1) . . . . .	27
5	Verdeutlichung der Intervallunabhängigkeit der MAP-Methode (2) . . . . .	27
6	1. Beispiel für die Ergebnisse der MAP-Methode . . . . .	28
7	2. Beispiel für die Ergebnisse der MAP-Methode . . . . .	29
8	3. Beispiel für die Ergebnisse der MAP-Methode . . . . .	29
9	4. Beispiel für die Ergebnisse der MAP-Methode . . . . .	30
10	1. Extrembeispiel für die Ergebnisse der MAP-Methode . . . . .	31
11	2. Extrembeispiel für die Ergebnisse der MAP-Methode . . . . .	31
12	1. Beispiel für die Ergebnisse der modifizierten MAP-Methode . . . . .	32
13	2. Beispiel für die Ergebnisse der modifizierten MAP-Methode . . . . .	33
14	3. Beispiel für die Ergebnisse der modifizierten MAP-Methode . . . . .	33
15	4. Beispiel für die Ergebnisse der modifizierten MAP-Methode . . . . .	34
16	1. Extrembeispiel für die Ergebnisse der modifizierten MAP-Methode . . . . .	34
17	2. Extrembeispiel für die Ergebnisse der modifizierten MAP-Methode . . . . .	35
18	Beispiel für die Ergebnisse der modifizierten MAP-Methode mit höherer Anzahl an Wiederholungen . . . . .	36
19	Darstellung der Ergebnisse der ML- und MAP-Methode . . . . .	42
20	Darstellung der Ergebnisse der ABC-Methode . . . . .	43
21	1. Beispiel für die Ergebnisse der ABC-Methode . . . . .	44
22	2. Beispiel für die Ergebnisse der ABC-Methode . . . . .	44
23	3. Beispiel für die Ergebnisse der ABC-Methode . . . . .	45
24	4. Beispiel für die Ergebnisse der ABC-Methode . . . . .	45
25	1. Extrembeispiel für die Ergebnisse der ABC-Methode . . . . .	47
26	2. Extrembeispiel für die Ergebnisse der ABC-Methode . . . . .	47
27	1. Beispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung . . . . .	48
28	2. Beispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung . . . . .	49
29	3. Beispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung . . . . .	49
30	4. Beispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung . . . . .	50
31	1. Extrembeispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung . . . . .	50
32	2. Extrembeispiel für die Ergebnisse der ABC-Methode mit Parameterziehung aus der Priorverteilung . . . . .	51
33	7. Beispiel für die Ergebnisse der ABC-Methode . . . . .	52
34	8. Beispiel für die Ergebnisse der ABC-Methode . . . . .	52
35	9. Beispiel für die Ergebnisse der ABC-Methode . . . . .	53
36	9. Beispiel mit Parameterziehung aus der Priorverteilung . . . . .	54
37	10. Beispiel für die Ergebnisse der ABC-Methode . . . . .	55
38	11. Beispiel für die Ergebnisse der ABC-Methode . . . . .	55
39	Verdeutlichung der Biaskorrektur durch die ABC-Methode . . . . .	56