

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN  
INSTITUT FÜR STATISTIK

---

# **Change-Point-Schätzung bei Überlebensdauern von Intensivpatienten**

---

## **Bachelorarbeit**

zur Erlangung des akademischen Grades Bachelor of Science (B.Sc.)

von  
Stefanie Krügel

Betreuer:  
Prof. Dr. Helmut Küchenhoff

München, 02. November 2015

## Abstract

In dieser Arbeit wird ein von [Küchenhoff u. a., 2015] entwickeltes Verfahren zur Schätzung des Change-Points bei Intensivpatienten vorgestellt. Dieser Zeitpunkt trennt die akute kritische Phase eines Patienten nach Einlieferung auf eine Intensivstation, welche durch eine anfänglich sehr hohe und im Laufe der Zeit abfallende Sterberate geprägt ist, vom postakuten Stadium in welchem die Hazardrate als konstant angenommen wird. Dieses nichtparametrische Schätzverfahren beruht auf p-Werten, welche sich aus der Durchführung von Binomialtests für festgelegte Intervalle des Beobachtungszeitraumes ergeben. Um die Güte der resultierenden Schätzung beurteilen zu können, wird eine Simulationsstudie anhand mehrerer geeigneter Szenarien durchgeführt. Dabei zeigt sich, dass der wahre Change-Point tendenziell unterschätzt wird, wobei der Bias insbesondere für Datensätze mit einer stetigen Hazardfunktion nicht zu vernachlässigen ist. Weiterhin wurde ein Paket für die Statistik-Software *R* programmiert, welches eine auf dem Verfahren von [Küchenhoff u. a., 2015] aufbauende Funktion zur Change-Point-Schätzung implementiert. Ein Kapitel widmet sich dem Aufbau und der Funktionsweise dieses Paketes. Abschließend werden Schätzungen zur Dauer der akuten Phase für zwei Datensätze ermittelt, welche die Überlebenszeiten von Intensivpatienten des Klinikums der Universität München enthalten, und im Bezug auf bereits veröffentlichte Studien zu diesen Daten diskutiert.

# Inhaltsverzeichnis

<b>1 Einleitung</b>	<b>1</b>
1.1 Hintergrund . . . . .	1
1.2 Problemstellung . . . . .	2
<b>2 Verfahren zur Change-Point-Schätzung</b>	<b>4</b>
2.1 Schätzung der konstanten Hazardrate . . . . .	6
2.2 Ermittlung der p-Werte und Regression . . . . .	7
2.3 Korrektur um zensierte Beobachtungen . . . . .	10
2.4 Diskussion . . . . .	10
<b>3 Simulationsstudie</b>	<b>12</b>
3.1 Datenerzeugung . . . . .	12
3.2 Szenarien . . . . .	14
3.3 Simulationsergebnisse . . . . .	17
3.4 Weiterführende Simulationen . . . . .	30
3.4.1 Schätzungen mit tatsächlichem $\lambda$ . . . . .	30
3.4.2 Untersuchung der Güte der Konfidenzintervalle . . . . .	32
3.5 Fazit zu den Simulationsergebnissen . . . . .	33
<b>4 R-Paket CPest</b>	<b>35</b>
4.1 Aufbau und Funktionsweise . . . . .	35
4.1.1 Funktion tau.est . . . . .	36
4.1.2 Funktion int.shift . . . . .	37
4.1.3 Methode CPest.default . . . . .	38
4.1.4 Methode plot.CP . . . . .	39
4.1.5 print- und summary-Methoden . . . . .	42
<b>5 Anwendungsbeispiele</b>	<b>43</b>
5.1 Datensatz 1 - Überlebenszeiten nach Leberresektion . . . . .	43
5.2 Datensatz 2 - Langzeitüberleben von Intensivpatienten nach chirurgischem Eingriff . . . . .	49
<b>6 Zusammenfassung und Ausblick</b>	<b>54</b>
<b>Anhang</b>	<b>55</b>
<b>A Plots zur Simulationsstudie</b>	<b>55</b>
<b>B Plots zu den Datensätzen</b>	<b>67</b>
<b>Literaturverzeichnis</b>	<b>74</b>

# Abbildungsverzeichnis

1	theoretische Hazardfunktion für $\tau = 100$	15
2	theoretische Hazardfunktion für $\tau = 50$	16
3	Boxplots der Change-Point-Schätzer für $\tau = 90$	21
4	Boxplots der Change-Point-Schätzer für $\tau = 360$	29
5	Plot der Hazardfunktion für simulierte Daten	40
6	Plot der p-Werte für simulierte Daten	41
7	Boxplot der Schätzungen aus Bootstrapping für simulierten Datensatz	41
8	p-Werte für Schätzung aus Überlebenszeiten nach Leberresektion	46
9	Histogramm für Todesfälle nach Leberresektion	47
10	Geschätzte Hazardrate für Patienten nach Leberresektion	49
11	Geschätzte Hazardrate für Intensivpatienten	52
12	Geschätzte Hazardrate für Intensivpatienten mit neuer Skalierung	53
13	theoretische Hazardfunktion	55

## Tabellenverzeichnis

1	Simulationsergebnisse für Szenario 1 . . . . .	17
2	Simulationsergebnisse für Szenario 2 . . . . .	22
3	Simulationsergebnisse für Szenario 3 . . . . .	27
4	Simulationsergebnisse für Schätzung mit $\hat{\lambda}$ vs. $\lambda$ . . . . .	31
5	Simulation mit Konfidenzintervallen . . . . .	33
6	Schätzergebnisse für Daten mit Überlebenszeiten nach Leberresektion . . .	45
7	Schätzergebnisse für Überlebenszeiten von Intensivpatienten . . . . .	51

# 1 Einleitung

Im Zusammenhang mit Intensivpatienten wird als Change-Point jener Zeitpunkt bezeichnet, welcher die akute Phase nach Einlieferung auf eine Intensivstation, insbesondere nach einem chirurgischen Eingriff, von der postakuten Phase trennt. Direkt nach einer OP befindet sich ein Patient meist in einem kritischen Stadium, weshalb die Sterberate anfänglich am höchsten ist und mit der Zeit abfällt, bis sie schließlich annähernd konstant bleibt. Der Change-Point kennzeichnet diese Änderung im Verlauf der Hazardrate, stellt also einen Wendepunkt dar, an dem die Phase der konstanten Sterberate beginnt. In medizinischen Studien ist es oft von Interesse, statistische Analysen nur für eine dieser beiden Stadien vorzunehmen. So werden die langfristigen Auswirkungen auf die Sterberate häufig nur für medizinische Maßnahmen während der akuten Phase durchgeführt (vgl. [Schiergens u. a., 2015]). Andererseits können jedoch auch die Langzeit-Effekte bestimmter Therapien in der postakuten Phase von vorrangigem Interesse sein (vgl. [Li u. a., 2013]).

Die postoperative Sterberate ergibt sich häufig aus der Anzahl der Todesfälle innerhalb der ersten 30 Tage nach einer Operation oder während des Krankenhausaufenthaltes (30-Tage/“In-Hospital“ Mortalität). Nach [Schiergens u. a., 2015] ist diese Definition anhand eines festgelegten Zeitraumes in vielen Fällen zu allgemein, was die Schätzung der tatsächlichen Dauer der akuten postoperativen Phase anhand von beobachteten Überlebenszeiten bestimmter Patienten notwendig macht. Daraus können beispielsweise individuelle Risikofaktoren ermittelt werden, was einerseits eine bessere Selektion von Patienten für eine Operation und andererseits einen effektiveren Einsatz von medizinischen Maßnahmen ermöglicht, um somit auf lange Sicht die postoperative Sterberate zu verringern (vgl. [Schiergens u. a., 2015] & [Schneider u. a., 2010]). Die vorliegende Arbeit beschäftigt sich mit einem nichtparametrischen Verfahren von [Küchenhoff u. a., 2015], um den Change-Point und damit die Dauer der akuten kritischen Phase nach einem medizinischen Eingriff zu schätzen.

## 1.1 Hintergrund

In der Statistik gibt es viele Bereiche, in denen die Change-Point Analyse eine Rolle spielt. So ist es nicht nur in medizinischen Studien von Interesse, einen Strukturbruch in der Verteilung der erhobenen Daten zu untersuchen, sondern auch in vielen anderen Fachgebieten, wie zum Beispiel der Produktionstechnik oder Klimaforschung, um nur zwei weitere Anwendungsbereiche zu nennen. Wie auch in [Küchenhoff u. a., 2015] anhand einiger Literaturbeispiele angeführt, sind in den letzten Jahrzehnten bereits diverse wissenschaftliche Arbeiten entstanden, die sich mit dem Problem der Change-Point-Schätzung befassen.

In [Li u. a., 2013] wird beispielsweise ein Verfahren beschrieben, bei welchem im Zusammenhang mit klinischen Studien der Change-Point auf Basis der Maximum-Likelihood-

Methode ermittelt wird. Neben den beobachteten Überlebenszeiten werden dabei außerdem Kovariablen-Effekte berücksichtigt, um den Zeitpunkt zu finden, ab dem die Hazardrate konstant ist.

Ein nichtparametrisches Verfahren zur Change-Point-Schätzung wird in [Yang u. a., 2012] vorgestellt. Dabei soll der Übergang von einer anfänglich abfallenden zu einer konstanten Ausfallrate mithilfe von Techniken aus der Bayes-Statistik identifiziert werden.

Eine weitere Möglichkeit, den Change-Point ohne eine Verteilungsannahme zu finden, bietet eine nichtparametrische Schätzung der Hazardfunktion, aus der anschließend ein Zeitpunkt bestimmt wird, ab dem diese Funktion einen bestimmten Schwellwert überschreitet (vgl. [Küchenhoff u. a., 2015]). Ein Kernel-basiertes Verfahren, um diese Hazardfunktion zu schätzen, wird in [Muller u. Wang, 1994] beschrieben. Dieses findet u.a. im R-Paket *muhaz* Anwendung, welches in späteren Kapiteln für die grafische Darstellung der Hazardrate verwendet wird. Einige weitere Literaturbeispiele, welche Verfahren zur Change-Point-Schätzung beinhalten, sind in [Küchenhoff u. a., 2015] zu finden.

Das Thema der Change-Point-Schätzung bei Überlebensdauern von Intensivpatienten wurde unter anderem bereits in [Schneider u. a., 2010] aufgegriffen, bei dem für einen Datensatz von [Hartl u. a., 2007] der Zeitpunkt geschätzt wurde, ab welchem die Hazardrate konstant ist. Grundlage bildet dabei ein Likelihood-Quotienten-Test, welcher in Kapitel 5.2 im Zusammenhang mit den Anwendungsbeispielen etwas ausführlicher beschrieben wird. Die Verteilungsannahme der Überlebensdauern, welche für solche parametrische Verfahren getroffen wird, ist jedoch oft problematisch. Deshalb ist das Finden eines nichtparametrischen Change-Point-Schätzers Gegenstand weiterer Forschungsarbeit, woraus auch das in dieser Arbeit vorgestellte Verfahren von [Küchenhoff u. a., 2015] entstanden ist. Dieses baut auf der Idee von [Mallik u. a., 2011] auf, den Wendepunkt in der Hazardrate durch eine Regression über die p-Werte eines statistischen Test zu ermitteln, wie im Kapitel 2 ausführlich beschrieben wird.

## 1.2 Problemstellung

Weil diese Arbeit auf den Ausführungen zur Change-Point-Schätzung von [Küchenhoff u. a., 2015] aufbaut, wird die verwendete Notation und die im Folgenden getroffenen Annahmen größtenteils daraus übernommen.

Gegeben seien unabhängige und identisch verteilte Überlebensdauern  $t_i$ ,  $i = 1, \dots, n$ , mit Verteilungsfunktion  $F$ , d.h.  $t_i \sim F$ . Da es in dieser Arbeit um das Sterberisiko von Intensivpatienten geht, beschreiben die Beobachtungen  $t_i$  die Dauer vom Beginn der Beobachtung, d.h. dem ersten Tag auf einer Intensivstation bis zum Todeszeitpunkt. Das Szenario lässt sich jedoch auch auf andere Bereiche übertragen, in denen die Lebensdaueranalyse Anwendung findet, z.B. für Ausfallzeiten bei technischen Geräten.

Da nicht jeder Patient innerhalb der Beobachtungsdauer verstirbt, kann der tatsäch-

liche Todeszeitpunkt für einen Teil der Personen unter Risiko unbekannt sein. Für diese sogenannten rechts-zensierten Daten wird eine Zufallsvariable  $C$  definiert, deren Ausprägungen  $c_i$  unabhängig von den Lebensdauern  $t_i$  sind und die maximale Beobachtungsdauer beschreiben. Unter Berücksichtigung dieser Zensierungszeiten ergeben sich für jedes Individuum  $i$  zwei Beobachtungen: ein Zeitraum  $T_i$ , welcher als  $T_i := \min(c_i, t_i)$  definiert ist und ein Indikator  $\delta_i$  für den Zensierungsstatus. Es gilt

$$\delta_i = I_{\{c_i < t_i\}} = \begin{cases} 1, & \text{falls } c_i < t_i \text{ (d.h. Beobachtung } i \text{ ist zensiert)} \\ 0 & \text{sonst} \end{cases} \quad (1)$$

Bei einer rechts-zensierten Beobachtung liegt demnach das Ende des Beobachtungszeitraumes vor dem Todeszeitpunkt.

Das Risiko für einen Intensivpatienten, zu einem bestimmten Zeitpunkt  $t$  zu sterben, wenn er bis dahin überlebt hat, wird durch die Hazardrate  $h(t)$  beschrieben, welche nach [Küchenhoff u. a., 2015] wie folgt definiert sei:

$$h(t) = h_1(t) * I_{\{t \leq \tau\}}(t) + \lambda * I_{\{t > \tau\}}(t) \quad \text{mit } I_{\{x \in A\}}(x) = \begin{cases} 1 & \text{falls } x \in A \\ 0 & \text{sonst} \end{cases} \quad (2)$$

Mit  $\tau$  wird dabei der Change-Point bezeichnet, d.h. für alle  $t > \tau$  ist die Hazardrate konstant mit Wert  $\lambda$ . Über  $h_1(t)$  werden keine weiteren Annahmen getroffen. Für die Verteilungsfunktion muss ebenfalls zwischen den Zeitpunkten vor und nach dem Change-Point unterschieden werden, womit sich für  $F$  folgende Form ergibt:

$$F(t) = F_1(t) * I_{\{t \leq \tau\}}(t) + F_1(\tau) * (1 - \exp(-\lambda(t - \tau))) * I_{\{t > \tau\}}(t)$$

mit  $F_1(t) := 1 - \exp(-\int_0^t h(s)ds)$

Ziel ist es nun, aus den gegebenen Daten einen Schätzer für  $\tau$  zu finden.

Das Verfahren der nichtparametrischen Change-Point-Schätzung mittels p-Werten nach [Küchenhoff u. a., 2015] wird zunächst im Kapitel 2 detailliert beschrieben. Ein Ziel dieser Ausarbeitung ist es, die Güte des daraus resultierenden Schätzers zu untersuchen. Die Ergebnisse der dafür durchgeführten Simulationsstudie werden in Kapitel 3 diskutiert. Weiterhin wurde eine erste Version des *R*-Paketes “CPest” erstellt, welches eine Funktion zur Change-Point-Schätzung nach der Methode von [Küchenhoff u. a., 2015] implementiert, wobei durch Anwendung eines nichparametrischen Bootstrap-Verfahrens zusätzlich die Schätzung der Varianz möglich ist. Details zum Aufbau und Ausschnitte aus dem Programmcode sind im vierten Kapitel zu finden. Als Anwendungsbeispiele dienen abschließend zwei Datensätze mit Überlebenszeiten von Intensivpatienten des Klinikums der Universität München. Bereits veröffentlichte Studien zu diesen Daten enthalten unter anderem eine geschätzte Dauer der akuten postoperativen Phase, welche im letzten Kapitel mit den Ergebnissen aus dem hier vorgestellten Verfahren verglichen werden.

## 2 Verfahren zur Change-Point-Schätzung

Ein Verfahren, bei dem der Change-Point mittels p-Werten bestimmt wird, ist in [Mallik u. a., 2011] anhand von verschiedenen Szenarien beschrieben. Dabei ist das Ziel, jeweils „einen Schwellenwert zu finden, ab dem eine Regressionsfunktion von ihrem Basiswert abweicht“, welcher im Folgenden mit  $\tau$  bezeichnet ist. Formal wird dabei für ein mögliches „Dose-Response-Setting“ ein Regressionsmodell  $Y = \mu(X) + \epsilon$  angenommen, bei dem  $\mu(x)$  eine Funktion auf  $[0, 1]$  ist. Die Zufallsvariable  $X$  beschreibt dabei die verabreichte Dosis und  $Y$  die zugehörige Response. Diese ist für  $x \leq \tau$  konstant mit  $\mu(x) = c_0$  und steigt für Werte  $x > \tau$  an, d.h.  $\mu(x) > c_0$  für  $x > \tau$ , wobei  $\tau \in (0, 1)$  gelte. Zu beachten ist, dass die Notation im Hinblick auf das Verfahren nach [Küchenhoff u. a., 2015] von der in [Mallik u. a., 2011] ursprünglich verwendeten abweicht. Um den Change-Point  $\tau$  zu finden, wird zunächst ein geeigneter Schätzer  $\hat{c}_0$  für die Response bestimmt und anschließend ein einseitiger Gauss-Test mit  $H_0 : \mu(x) = \hat{c}_0$  und  $H_1 : \mu(x) > \hat{c}_0$  für die Dosen  $X_i = x$  durchgeführt.

Unter den Annahmen  $E(\epsilon | X = x) = 0$  und  $\sigma^2(x) = var(\epsilon | X = x) > 0$  für alle  $x \in [0, 1]$  ergeben sich die p-Werte nach [Mallik u. a., 2011] für jeweils  $m$  Beobachtungen zu den  $n$  untersuchten Kovariablen durch

$$p_{m,n}(X_i, \hat{\tau}_{m,n}) = 1 - \Phi \left( \sqrt{m} \frac{\bar{Y}_i - \hat{\tau}}{\hat{\sigma}(X_i)} \right),$$

wobei  $\bar{Y}_i$  durch  $\bar{Y}_i = \sum_{j=1}^m Y_{ij}/m$  definiert ist und  $\hat{\sigma}$  den Schätzer für die Standardabweichung bezeichnet.

Für alle  $X_i \leq \tau$  ist die Nullhypothese erfüllt und weil die zugehörigen p-Werte auf  $[0, 1]$  gleichverteilt sind, nehmen diese im Mittel den Wert  $1/2$  an. Für x-Werte oberhalb von  $\tau$ , für welche die Alternative  $H_1$  gilt, konvergieren die p-Werte gegen 0. Daraus ergibt sich eine stückweise konstante Funktion, deren Sprungstelle  $\tau$  durch eine einfache Regression geschätzt werden kann.

Nach [Mallik u. a., 2011] kann dieses Vorgehen leicht auf andere Situationen übertragen werden, bei denen beispielsweise die Response ab dem Change-Point  $\tau$  abfällt und es sich somit bei  $c_0$  um ein globales Maximum statt eines Minimums handelt. Ebenso ist es möglich, dass  $\mu(x)$  für  $x \leq \tau$  ansteigt bzw. abfällt und erst danach einen konstanten Wert annimmt. Genau dieses Szenario ist für unseren Fall der Intensivpatienten von Interesse. Im Unterschied zu [Mallik u. a., 2011] gibt es hier jedoch keine Regressionsfunktion, sondern es wird die Hazardrate  $h$  in Abhängigkeit von der Zeit  $t$  betrachtet. Wie bereits in (2) definiert, fällt diese zu Beginn des Beobachtungszeitraumes ab und ist ab einem bestimmten Zeitpunkt  $\tau$  konstant. Aus  $h(t)$  lässt sich die Wahrscheinlichkeit für ein Ereignis, d.h. für den Tod eines Patienten, innerhalb eines bestimmten Zeitraumes berechnen, welche nach [Küchenhoff u. a., 2015] als Grundlage für einen Binomialtest dienen soll. Ein weiterer Unterschied zu dem Verfahren der Change-Point-Schätzung nach [Mallik u. a., 2011] besteht also darin, dass hier kein Gauss-Test durchgeführt, sondern ein einseitiger

exakter Binomialtest, unter Verwendung der aus der Hazardrate ermittelten Wahrscheinlichkeit und der beobachteten Ereignisse innerhalb eines Zeitraumes. Im Folgenden wird die Methode nach [Küchenhoff u. a., 2015] ausführlicher beschrieben, wobei sich viele der getroffenen Annahmen auf den Fall der Intensivpatienten beziehen, jedoch ohne weiteres auf andere geeignete Anwendungsgebiete übertragen werden können, bei denen die Ermittlung eines Change-Points von Interesse ist.

Es sei  $T$  eine nichtnegative, stetige Zufallsvariable, welche die Überlebensdauer eines Individuums bezeichnet. Gegeben sind also unabhängige und identisch verteilte Überlebenszeiten  $t_i, i = 1, \dots, n$ , mit Dichte  $f(t)$  und Verteilungsfunktion  $F(t)$ .

Weil die Hazardrate  $h(t)$  nach Gleichung (2) für Zeitpunkte nach dem Change-Point als konstant angenommen wird, gilt  $\{t_i - \tau \mid t_i > \tau\} \sim \text{Exp}(\lambda)$ . Für die Zeitpunkte vor dem Change-Point wird keine Verteilungsannahme getroffen. Für das Beispiel der Intensivpatienten bedeutet dies, dass das Risiko, im nächsten Moment zu sterben, nach Einlieferung auf die Intensivstation am größten ist und immer weiter abfällt, bis es ab dem Zeitpunkt  $\tau$  konstant bleibt.

Oft beinhalten die in diesem Zusammenhang untersuchten Datensätze rechts-zensierte Beobachtungen, welche durch das Schätzverfahren berücksichtigt werden sollen. Für den Zensierungsstatus wird im Folgenden wieder die im Kapitel 1.2 unter (1) eingeführte Notation verwendet. D.h.  $\delta$  bezeichnet den Indikator für zensierte Beobachtungen mit  $\delta_i = I_{\{c_i < t_i\}}$ , wobei  $c_i$  die Zensierungszeit für Individuum  $i$  darstellt. Die beobachteten Zeiten  $T_i$  wurden mit  $T_i = \min(c_i, t_i)$  definiert, bezeichnen demnach entweder die Überlebensdauer oder die Beobachtungsdauer bis zum Zensierungszeitpunkt.

Die Idee des p-Wert-Verfahrens nach [Küchenhoff u. a., 2015] ist nun, dass der Beobachtungszeitraum in Intervalle einer vordefinierten Länge aufgeteilt wird und für jedes Intervall ein Binomialtest mit den beobachteten Ereignissen innerhalb eines Intervalls durchgeführt wird. Unter der Nullhypothese ist die Wahrscheinlichkeit  $p$  für ein Ereignis innerhalb eines Intervalls gleich jener, die sich aus der konstanten Hazardrate  $\lambda$  ergibt. Für Intervalle, die nach dem Change-Point liegen, sollte  $H_0$  somit erfüllt sein, das heißt die p-Werte sind für große Beobachtungszahlen auf  $[0,1]$  gleichverteilt und nehmen im Mittel den Wert  $1/2$  an. Für  $T_i \leq \tau$  ist die Nullhypothese wegen  $h_1(t) > \lambda$  nicht erfüllt und die p-Werte der zugehörigen Intervalle konvergieren für große  $n$  gegen 0. Abschließend wird, wie in Kapitel 2.2 ausführlich beschrieben, eine Regression über die ermittelten p-Werte durchgeführt und  $\hat{\tau}$  ergibt sich als KQ-Schätzer durch Anpassen der Funktion  $f(T_i) = 0.5 * I_{\{T_i > \tau\}}(T_i)$  an die Daten.

Zuerst benötigen wir also die konstante Hazardrate für  $T > \tau$ . Nachfolgend wird kurz hergeleitet, wie sich diese aus den gegebenen Daten ermitteln lässt.

## 2.1 Schätzung der konstanten Hazardrate

Für eine exponentialverteilte Zufallsvariable  $T$  mit Parameter  $\lambda > 0$  gelten im Allgemeinen folgende Zusammenhänge:

- $h(t) = \lambda$  (mit  $\lambda > 0$ ), d.h. die Hazardrate ist für alle Zeitpunkte  $t_i$  konstant
- Dichte:  $f(t) = \lambda \exp(-\lambda t)$
- Survivalfunktion:  $S(t) = 1 - F(t) = \exp(-\lambda t)$

Unter der Annahme, dass  $\tau$  vorerst gegeben sei, werden die Zeitpunkte nach dem Change-Point definiert als  $T_j^* := \{T_j - \tau | T_j > \tau\}$  mit  $T_j = \min(c_j, t_j)$  und  $j = 1, \dots, n^*$ . Dann gilt  $T_j^* \sim \text{Exp}(\lambda)$  mit der konstanten Hazardrate  $\lambda$ , welche aus den gegebenen Daten geschätzt werden soll.

In Anlehnung an die Herleitung eines ML-Schätzers für zensierte, exponentialverteilte Daten nach [Keppler u. Müller, 2013, S.20] lässt sich die Likelihoodfunktion wie folgt definieren und umformen:

$$\begin{aligned} L(\lambda, T_j^*) &\propto \prod_{j=1}^{n^*} f(T_j^*)^{1-\delta_j} S(T_j^*)^{\delta_j} \\ &= \prod_{j=1}^{n^*} f(T_j^*)^{1-\delta_j} S(T_j^*)^{-(1-\delta_j)} S(T_j^*)^{1-\delta_j} S(T_j^*)^{\delta_j} \\ &= \prod_{j=1}^{n^*} \left( \frac{f(T_j^*)}{S(T_j^*)} \right)^{1-\delta_j} S(T_j^*) \\ &= \prod_{j=1}^{n^*} \lambda^{1-\delta_j} \exp(-\lambda T_j^*) \end{aligned}$$

Daraus ergibt sich die log-Likelihood

$$l(\lambda, T_j^*) = \log L(\lambda, T_j^*) = \sum_{j=1}^{n^*} (1 - \delta_j) \log(\lambda) - \lambda \sum_{j=1}^{n^*} T_j^*$$

Durch Ableiten und Nullsetzen erhält man einen Schätzer für  $\lambda$ :

$$\begin{aligned} \frac{\partial l(\lambda, T_j^*)}{\partial \lambda} &= \frac{1}{\lambda} \sum_{j=1}^{n^*} (1 - \delta_j) - \sum_{j=1}^{n^*} T_j^* \stackrel{!}{=} 0 \\ \Leftrightarrow \hat{\lambda}^* &= \frac{\sum_{j=1}^{n^*} (1 - \delta_j)}{\sum_{j=1}^{n^*} T_j^*} = \frac{n_{unzens.}^*}{\sum_{j=1}^{n^*} T_j^*} \end{aligned}$$

Die Formel muss nun noch dahingehend angepasst werden, dass  $\lambda$  nicht nur aus den  $T_j^*$ , sondern aus allen Beobachtungen  $T_i$  ( $i = 1, \dots, n$ ) mit zugehörigem Zensierungsstatus geschätzt werden kann. Da jedoch der Zeitpunkt  $\tau$ , ab dem die Hazardrate als konstant

angenommen wird, nicht gegeben ist, wird ein Zeitpunkt  $\tau_{max}$  vorgegeben, welcher ein oberes Limit für den Change-Point darstellt. Das heißt wir nehmen an, dass die Hazardrate für alle  $T_i > \tau_{max}$  konstant ist und der wahre Change-Point  $\tau$  darunter liegt. Aus den beobachteten Überlebenszeiten  $T_i$  lässt sich der Schätzer für  $\lambda$  dann wie folgt berechnen:

$$\hat{\lambda} = \frac{\sum_{i=1}^n (1 - \delta_i) * I_{\{T_i > \tau_{max}\}}(T_i)}{\sum_{i=1}^n (T_i - \tau_{max}) * I_{\{T_i > \tau_{max}\}}(T_i)} \quad (3)$$

## 2.2 Ermittlung der p-Werte und Regression

### Methode 1

Die nachfolgend beschriebene Methode zur Ermittlung eines Schätzers  $\hat{\tau}$  aus den p-Werten entspricht den Ausführungen zum Schätzverfahren in [Küchenhoff u. a., 2015]. Unter der Annahme, dass der wahre Change-Point  $\tau$  kleiner oder gleich dem definierten Maximum ist, d.h.  $0 \leq \tau \leq \tau_{max}$ , wird ein Gitter  $0 \leq a_0 < a_1 < \dots < a_K = \tau_{max}$  definiert, welches den Beobachtungszeitraum bis  $\tau_{max}$  in disjunkte Intervalle unterteilt. Die Anzahl der Ereignisse innerhalb eines Intervalls  $(a_{k-1}; a_k]$  wird mit  $X_k$  bezeichnet. Weiterhin ist mit  $N_k = \sum_{i=1}^n I_{\{T_i > a_{k-1}\}}(T_i)$  die Zahl der Personen definiert, welche zu Beginn des k-ten Intervalls, d.h. zu  $t = a_{k-1}$  unter Beobachtung und damit unter Risiko standen.

Für Intervalle nach dem Change-Point (d.h. für  $a_{k-1} > \tau$ ) sind die  $X_k$  binomialverteilt mit den Parametern  $N_k$  und  $pr_k$ , wobei  $pr_k$  der Wahrscheinlichkeit für ein Ereignis innerhalb eines Intervalls  $k$  entspricht. Diese ergibt sich aus der konstanten Hazardrate  $\lambda$  zu  $pr_k = 1 - \exp(-\lambda(a_k - a_{k-1}))$ . Für den Fall, dass nur Intervalle der selben Länge betrachtet werden, ist diese Wahrscheinlichkeit für jedes Intervall gleich und es gilt  $pr = pr_k$  für beliebiges  $k = 1, \dots, K$ .

Daraus lässt sich ein (einseitiger) exakter Binomialtest mit den Hypothesen  $H_0 : p = pr_k$  gegen  $H_1 : p > pr_k$  durchführen. Unter der Nullhypothese gilt also  $X_i \stackrel{H_0}{\sim} B(N_k, pr_k)$ . Die zugehörigen p-Werte ergeben sich dann aus der Verteilungsfunktion der Binomialverteilung:

$$pv_k = 1 - F_{Bin}(X_k, N_k, pr_k)$$

Diese folgen bei einer großen Beobachtungszahl  $n$  für Intervalle nach dem Change-Point einer (0,1)-Gleichverteilung mit Erwartungswert  $E(pv) = 0.5$ , da für  $T_i > \tau$  die Nullhypothese erfüllt ist. Für Intervalle vor dem Change-Point konvergieren die p-Werte für große  $n$  gegen 0.

Um nun einen Schätzer  $\hat{\tau}$  zu finden, kann mittels der KQ-Methode eine Funktion der Form  $f(T_i) = 0.5 * I_{\{T_i > \tau\}}(T_i)$  an die ermittelten p-Werte angepasst werden:

$$\hat{\tau} = \arg \min_{\tau} \sum_{k=1}^n (pv_k - 0.5 * I_{\{\tau \leq a_{k-1}\}})^2$$

Analog zum Vorgehen in [Mallik u. a., 2011] wird die zu minimierende Summe noch wie folgt vereinfacht:

$$\begin{aligned}
S(\tau) &= \sum_{k=1}^n (pv_k - 0.5 * I_{\{\tau \leq a_{k-1}\}})^2 \\
&= \sum_{k|a_{k-1} < \tau} pv_k^2 + \sum_{k|a_{k-1} \geq \tau} (pv_k - 0.5)^2 \\
&= \sum_{k=1}^n pv_k^2 + \sum_{k|a_{k-1} \geq \tau} (-pv_k + 0.25)
\end{aligned}$$

Da der erste Summand nicht von  $\tau$  abhängig ist, kann  $S(\tau)$  durch maximieren von

$$S^*(\tau) = \sum_{k|a_{k-1} \geq \tau} (pv_k - 0.25) \quad (4)$$

minimiert werden. Der Schätzer für  $\tau$  hat sich demnach vereinfacht zu

$$\hat{\tau} = \arg \max_{\tau} \sum_{k|a_{k-1} \geq \tau} (pv_k - 0.25). \quad (5)$$

Ein Nachteil der Schätzung mittels disjunkter Intervalle ist die Tatsache, dass sich für  $\hat{\tau}$  nur die definierten Intervallgrenzen  $a_0, \dots, a_K$  ergeben können.

## Methode 2

Eine alternative Möglichkeit besteht darin, fortlaufende Intervalle der selben Länge zu definieren. Dazu bezeichnen  $0 \leq l_1 < \dots < l_K = \tau_{max}$  die unteren Intervallgrenzen. Aus  $u_k := l_k + wd$  mit  $wd$  als festgelegte Intervallbreite ergeben sich die oberen Grenzen zu  $u_1 < \dots < u_K = (\tau_{max} + wd)$ .

Die Wahrscheinlichkeit  $pr$  für ein Ereignis innerhalb eines Intervalls  $k$  ist in diesem Fall definiert als  $pr = 1 - \exp(-\lambda wd)$  und ist somit für alle  $k = 1, \dots, K$  gleich. Damit lassen sich die p-Werte durch  $pv_k = 1 - F_{Bin}(X_k, N_k, pr)$  ermitteln. Da sich die Intervalle nun überlappen, ergeben sich je nach definierter Breite und Intervallgrenzen entsprechend mehr p-Werte als nach Methode 1. Analog zur Berechnung von  $\hat{\tau}$  für disjunkte Intervalle nach (5) ergibt sich der Schätzer für den Change-Point für fortlaufende Intervalle zu:

$$\hat{\tau} = \arg \max_{\tau} \sum_{k|l_k \geq \tau} (pv_k - 0.25)$$

Somit ist es nun möglich, die Intervalle so festzulegen, dass alle Zeitpunkte zwischen  $0$  und  $\tau_{max}$  als Schätzer für den Change-Point in Frage kommen. Dazu müssen die unteren Intervallgrenzen als Sequenz  $\{0, 1, \dots, \tau_{max}\}$  definiert werden, womit sich die oberen Grenzen zu  $\{(0 + wd), (1 + wd), \dots, (\tau_{max} + wd)\}$  ergeben. Anders ausgedrückt bedeutet dies, dass die Intervalle  $(l_k, u_k]$  der Länge  $wd$  immer um einen Zeitpunkt nach rechts verschoben werden.

Zu beachten ist bei diesem Schätzverfahren, dass die p-Werte durch die sich überlappenden Intervalle nicht unabhängig sind. Durch die Verletzung der Unabhängigkeitssannahme ist der Schätzer unter Umständen nicht effizient und die Anwendung dieser Methode problematisch.

### Methode 3

Nachfolgend möchte ich eine weitere Möglichkeit der Change-Point-Schätzung vorstellen, wobei die Idee der fortlaufenden Intervalle aufgegriffen, jedoch das Problem der Abhängigkeit der p-Werte umgangen wird.

Dazu sei  $a_{0_1}, a_{0_2}, \dots, a_{0_l}$  eine Sequenz von “Startpunkten“, sodass sich für jeden Zeitpunkt  $a_{0_j}$  ein Gitter  $a_{0_j} < a_{1_j} < \dots < a_{K_j}$  ergibt, welches den Beobachtungszeitraum in disjunkte Intervalle der vorgegebenen Breite  $wd$  unterteilt. Es gilt also  $wd := a_{k_j} - a_{(k-1)_j}$  für alle  $j = 1, \dots, l$ . Dabei soll die Anzahl der definierten Gitter  $l$  der Intervallbreite  $wd$  entsprechen, d.h. es gelte  $l = wd$ . Beispielsweise würden sich bei einer festgelegten Breite von  $wd = 10$  und  $a_{0_1} = 0$  die Intervallgrenzen  $\{0, 10, 20, \dots\}$  für  $j = 1$  bis  $\{9, 19, 29, \dots\}$  für  $j = 10$  ergeben.

Für jedes dieser Gitter werden nun entsprechend dem Vorgehen in Methode 1 die p-Werte  $pv_{k_j}$  für  $k = 1, \dots, K$  berechnet. Daraus ergibt sich analog zu (4) für jedes  $j = 1, \dots, l$  die zu maximierende Summe

$$S_j^*(\tau) = \sum_{k|a_{(k-1)_j} \geq \tau} (pv_{k_j} - 0.25)$$

und damit jeweils ein Schätzer  $\hat{\tau}_j$ .

Durch Optimieren über diese endliche und vergleichsweise geringe Anzahl an Summen  $S_j^*(\tau)$  wird jenes  $\hat{\tau}_j$  bestimmt, für welches die Anpassung der Funktion

$$f(T_i) = 0.5 * I_{\{T_i > \hat{\tau}_j\}}(T_i)$$

an die zugehörigen p-Werte am besten ist. Dazu ermittelt man das Maximum über alle  $S^*(\hat{\tau}_j)$  und der Change-Point-Schätzer  $\hat{\tau}$  ergibt sich zu

$$\hat{\tau} = \arg \max_j S^*(\hat{\tau}_j).$$

Im Gegensatz zu Methode 2 werden bei diesem Verfahren  $l$  separate Regressionsmodelle mit diskreten Intervallen aufgestellt, sodass die p-Werte für jedes  $j = 1, \dots, l$  voneinander unabhängig sind. Problematisch an dieser Methode ist jedoch, dass dadurch über mehrere Modelle hinweg optimiert wird.

### 2.3 Korrektur um zensierte Beobachtungen

Für den Fall, dass sich innerhalb eines Intervalls  $(a_{k-1}; a_k]$  zensierte Beobachtungen befinden, ist die tatsächliche Anzahl der Ereignisse möglicherweise höher als beobachtet, wodurch sich aus dem Binomialtest ein geringerer p-Wert ergeben würde.

Eine Möglichkeit, die zensierten Beobachtungen innerhalb des k-ten Intervalls zu berücksichtigen, besteht darin, die Wahrscheinlichkeit für ein Ereignis innerhalb des “Restintervalls“  $(c_i, a_k]$  zu berechnen und die Anzahl der beobachteten Ereignisse  $X_k$  entsprechend zu erhöhen. Problematisch ist dabei, dass sich für einen relativ geringen Anteil an zensierten Beobachtungen der Wert für  $X_k$  nur unwesentlich ändern würde. Da für einen Binomialtest die Parameter ganzzahlig sein müssen, hätte die Korrektur somit keinen Effekt.

Stattdessen kann jedoch auch die Anzahl der Personen unter Risiko  $N_k$ , welche im Allgemeinen deutlich größer ist als  $X_k$ , nach unten korrigiert werden. Ein solches Vorgehen ist in [Hamerle u. Tutz, 1989, S.23] vorgeschlagen. Dabei wird  $N_k$  “willkürlich“ um die Hälfte der Anzahl an Zensierungen im k-ten Intervall verringert. Für das hier vorgestellte Verfahren soll diese Idee erweitert werden, sodass eine präzisere Korrektur der Risikomenge  $N_k$  stattfindet.

Dazu wird  $N_k$  für jede zensierte Beobachtung  $i$  innerhalb des k-ten Intervalls um den Anteil der Zeitpunkte reduziert, an denen Person  $i$  nicht mehr unter Beobachtung stand. Das korrigierte  $N_k$  ergibt sich demnach zu

$$N_k^* = N_k - \sum_{i \mid T_i \in (a_{k-1}; a_k]} \left( \frac{a_k - T_i}{a_k - a_{k-1}} \right) * \delta_i \quad (6)$$

### 2.4 Diskussion

Bei diesem nichtparametrischen Verfahren wird der Change-Point-Schätzer in allen drei vorgestellten Methoden durch Optimieren über eine kleine endliche Menge bestimmt, was mathematisch einfach zu berechnen ist. Somit ist auch die praktische Anwendung, beispielsweise in Form einer in die Statistik-Software *R* implementierten Funktion, mit vergleichsweise wenig Aufwand verbunden. Eine mögliche Umsetzung einer solchen Funktion wird in Kapitel 4 ausführlich beschrieben, bei dem es um die Erstellung eines zugehörigen *R*-Paketes geht. Da für die Zeitpunkte  $\{t \mid t < \tau_{max}\}$  keine parametrischen Verteilungsannahmen getroffen werden, wird die Standardabweichung mittels eines nichtparametrischen Bootstrap-Verfahrens geschätzt, womit sich auch die zugehörigen Konfidenzintervalle berechnen lassen. Darauf wird ebenfalls im Kapitel 4 detaillierter eingegangen.

Ein Nachteil dieses Verfahrens ist die Abhängigkeit von der Intervallbreite, welche angemessen gewählt werden muss. Bei zu kleinen Intervallen besteht die Gefahr, dass sich innerhalb eines Intervalls nur sehr wenige oder gar keine Beobachtungen befinden, womit die p-Werte des Binomialtests gegen 1 konvergieren würden. Wird die Intervallbreite dagegen zu groß gewählt, kommen bei Methode 1 zu wenige Zeitpunkte als Schätzung

für  $\tau$  in Frage, da sich bei Verwendung von diskreten Intervallen nur die Intervallgrenzen  $a_0, \dots, a_K$  als Change-Point-Schätzer ergeben können. Durch das Verwenden von fortlaufenden Intervallen (Methode 2), oder das Ermitteln des “besten Fits“ aus mehreren Modellen mit jeweils um den Wert 1 verschobenen Intervallen (Methode 3), ist jeder Zeitpunkt  $t = 1, \dots, \tau_{max}$  als Schätzer für  $\tau$  möglich. Allerdings sind diese Methoden durch die Abhängigkeit der p-Werte bzw. die Optimierung über mehrere Regressionsmodelle nicht ganz unproblematisch. Um Aussagen über die Güte des Schätzverfahrens treffen zu können, vor allem im Hinblick auf den Vergleich dieser drei Methoden, wurde eine Simulationsstudie durchgeführt, in der die drei Verfahren mit verschiedenen Intervallbreiten auf diverse Szenarien angewandt werden. Die Ergebnisse der Simulationen und weitere Details sind im Kapitel 3 zu finden.

Als weitere Besonderheit dieses Verfahrens ist zu beachten, dass im Vorhinein ein Wert für  $\tau_{max}$  festgelegt werden muss, und zwar derart, dass noch ausreichend viele Beobachtungen mit  $T_i > \tau_{max}$  vorhanden sind, um einen möglichst genauen Schätzwert für  $\lambda$  zu finden. Gleichzeitig soll für  $\tau_{max}$  jedoch ein hinreichend großer Wert gewählt werden, sodass das wahre  $\tau$  mit hoher Wahrscheinlichkeit darunter liegt. Dazu kann es sinnvoll sein, den Datensatz im Vorfeld mithilfe geeigneter deskriptiver Methoden zu untersuchen, beispielsweise durch eine nichtparametrische Schätzung der Hazardrate.

In Kapitel 3 werden anhand der Simulationsergebnisse noch einige Probleme bezüglich der Anwendbarkeit des Verfahrens diskutiert.

### 3 Simulationsstudie

Um die Güte des Schätzverfahrens nach [Küchenhoff u. a., 2015] zu beurteilen, wird in diesem Kapitel anhand verschiedener Szenarien untersucht, ob es einerseits eine systematische Abweichung des Schätzers  $\hat{\tau}$  vom wahren Parameterwert gibt und andererseits, wie groß die Streuung der geschätzten Werte ist. Dazu wurden die in Kapitel 2.2 vorgestellten Methoden zur Schätzung von  $\tau$  auf mit verschiedenen Parameterkombinationen erzeugte Datensätze angewendet. Dabei geht es insbesondere um den Vergleich der drei Methoden, aber auch um die grundsätzliche Anwendbarkeit des Verfahrens auf realitätsnahe Datensätze. Nachfolgend wird zunächst die Erzeugung der Datensätze sowie der Aufbau der Simulationsstudie erklärt. In Kapitel 3.3 werden dann die Ergebnisse der Schätzungen diskutiert. Für die Simulationen sowie die Umsetzung des Schätzverfahrens wurde die Statistik-Software *R* (Version 3.1.0) verwendet.

#### 3.1 Datenerzeugung

Vor dem Hintergrund, dass die Überlebensdauern von Intensivpatienten untersucht werden sollen, müssen die Daten derart generiert werden, dass die Sterberate zu Beginn der Beobachtungsdauer am höchsten ist, bis zu einem vorgegebenen Zeitpunkt  $\tau$  abfällt und ab diesem Change-Point konstant bleibt. Eine geeignete Verteilung, welche oft im Zusammenhang mit Lebensdauern verwendet wird, ist die Weibullverteilung. Im Gegensatz zur Exponentialverteilung besitzt diese nicht die Eigenschaft der Gedächtnislosigkeit, sodass es mithilfe der Weibullverteilung möglich ist, neben einer konstanten auch eine ansteigende bzw. abfallende Sterberate zu modellieren. Die Dichte einer Weibull-verteilten Zufallsvariablen  $T$  hat die Form

$$f(t) = \frac{\gamma}{\eta} \left( \frac{t}{\eta} \right)^{\gamma-1} \exp\left(-\left(\frac{t}{\eta}\right)^\gamma\right)$$

mit Shape-Parameter  $\eta$  und Scale-Parameter  $\gamma$ . Daraus ergibt sich die Hazardrate zu:

$$h(t) = \frac{\gamma}{\eta} \left( \frac{t}{\eta} \right)^{\gamma-1}. \quad (7)$$

Für den Fall  $\gamma = 1$  ist  $T$  exponentialverteilt mit Parameter  $\lambda = \frac{1}{\eta}$ , wobei die Hazardrate konstant ist.

Für einen Datensatz mit  $n$  Beobachtungen werden zunächst mithilfe der *R*-Funktion `rweibull`, mit vorgegebenem Shape- und Scale-Parameter  $n$  Weibull-verteilte (Pseudo-) Zufallszahlen  $t_i$  erzeugt. Basis für die Generierung dieser Todeszeitpunkte bildet die Inversionsmethode, welche mittels der inversen Weibullverteilung aus auf  $[0,1]$  gleichverteilten Zufallszahlen  $n$  Realisationen einer Weibull-verteilten Zufallsvariablen simuliert.

Um zu erreichen, dass sich zum Zeitpunkt  $\tau$  ein Change-Point befindet, ab welchem  $h(t)$  konstant ist, werden alle generierten Zeiten  $\{t_i \mid t_i \geq \tau\}$  für vorgegebenes  $\tau$  durch

exponentialverteilte Zufallszahlen ersetzt. Hierfür ist zunächst der Parameter  $\lambda$  der Exponentialverteilung zu bestimmen. Soll der Übergang zwischen den Weibull- und exponentialverteilten Zeiten bei  $t = \tau$  stetig sein, so lässt sich  $\lambda$  nach (7) aus der Hazardfunktion  $h(t)$  berechnen:

$$\lambda = h(\tau) = \frac{\gamma}{\eta} \left( \frac{\tau}{\eta} \right)^{\gamma-1}.$$

Mithilfe des Parameters  $\lambda$  werden nun  $n_{exp}$  exponentialverteilte Zufallszahlen  $t_i^*$  erzeugt, d.h.  $t_i^* \sim Exp(\lambda)$ , wobei sich die Anzahl  $n_{exp}$  dieser Datenpunkte aus  $n_{exp} = \sum_{i=1}^n I_{\{t_i \geq \tau\}}(t_i)$  ergibt.

Die zuvor aus der Weibullverteilung generierten Zeiten  $t_i$ , für welche  $t_i \geq \tau$  gilt, werden nun wie folgt durch die exponentialverteilten Daten ersetzt:  $\{t_i | t_i \geq \tau\} = t_i^* + \tau$ .

Für die Simulationsstudie werden sowohl Daten mit einem stetigen Übergang als auch mit einem Sprung bei  $t = \tau$  erzeugt, was in den Ergebnissen durch  $jump=0$  bzw.  $jump=1$  gekennzeichnet ist. Für die Datensätze mit Sprung in der Hazardrate wird der Parameter der Exponentialverteilung durch  $\lambda_{jump} := \frac{\lambda}{2}$  ersetzt.

Da in der Praxis ein Datensatz oft zensierte Beobachtungen beinhaltet, welche durch das Schätzverfahren ebenfalls berücksichtigt werden können, soll auch ein Teil der simulierten Daten so erzeugt werden, dass nicht zu jeder Beobachtung eine Überlebenszeit vorliegt, sondern ein bestimmter Anteil durch Zensierungszeiten ersetzt wird. Dafür wird für jedes  $i = 1, \dots, n$  unabhängig von den Überlebenszeiten  $t_i$  eine maximale Beobachtungszeit  $c_i$  generiert, wobei die  $c_i$  ebenfalls exponentialverteilt sind mit  $c_i \sim Exp(\lambda_{cens})$ . Der Wert für  $\lambda_{cens}$  kann dabei so gewählt werden, dass sich ein gewünschter Anteil an zensierten Daten ergibt. Im Fall, dass die generierten Daten zensierte Beobachtungen enthalten, sind die zugehörigen Ergebnisse mit  $cens=1$  gekennzeichnet, andernfalls mit  $cens=0$ . Der simulierte Datensatz enthält letztendlich die Zeiten  $T_i$ , welche sich für jedes  $i = 1, \dots, n$  aus  $T_i = \min(t_i, c_i)$  ergeben. Zusätzlich wird eine binäre Variable für den Zensierungsstatus erzeugt. Zu beachten ist, dass für die Simulation gerundete Zeiten verwendet werden, weil Überlebensdauern im medizinischen Kontext in der Praxis meist in Tagen gemessen werden.

Nachfolgend ist der R-Code der Funktion dargestellt, welche für die Datengenerierung verwendet wurde. Dabei wird durch das Argument `cens = TRUE/FALSE` vorgegeben, ob zensierte Daten erzeugt werden sollen, wobei `lambda.cens` dem vorher definierten  $\lambda_{cens}$  entspricht, mit dem der Anteil der zensierten Beobachtungen beeinflusst werden kann. Durch `jump = TRUE/FALSE` wird bestimmt, ob Daten mit Sprung in der Hazardrate erzeugt werden sollen. Die Funktion gibt einen Datensatz zurück, welcher die Vektoren `time` für die generierten Zeiten  $T_i$  und `event` für den zugehörigen Zensierungsstatus enthält.

```
sim.survdata <- function(n, tau, shape, scale, cens, jump, lambda.cens){
  lambda <- shape/scale * (tau/scale)^(shape-1)
```

```

if(jump) lambda <- 0.5*lambda

t.weib <- ceiling(rweibull(n, shape=shape, scale=scale))
n.exp <- sum(t.weib>=tau)
t.exp <- ceiling(rexp(n.exp, lambda)) + tau
t <- t.weib
t[t.weib >= tau] <- t.exp

#censoring
if(cens){
  t.cens <- ceiling(rexp(n, lambda.cens))
  event <- as.numeric(t.cens > t)
  t <- pmin(t, t.cens)
  data <- data.frame(time=t, event)
}
else data <- data.frame(time=as.integer(t), event=rep(1,length(t)))
return(data)
}

```

## 3.2 Szenarien

Für die Simulationsstudie werden verschiedene Settings betrachtet, welche sich aus einer Kombination der Parameter für die Datengenerierung und jener für die Change-Point-Schätzung ergeben. Untersucht werden drei Szenarien, welche sich bezüglich des für die Datenerzeugung vorgegebenen Change-Points und des für die Schätzung relevanten Wertes  $\tau_{max}$  unterscheiden ( $\tau_{max} = 360$  in Szenarien 1 und 3;  $\tau_{max} = 200$  in Szenario 2). Weiterhin werden folgende variierende und feste Werte für die Simulationen verwendet:

- variierende Parameter für die Erzeugung der Datensätze:
  - Anzahl der Beobachtungen:  $n = \{1000; 5000\}$
  - Zensierung:  $cens = \{1; 0\}$  (1  $\hat{=}$  Datensatz mit Zensierung)
  - Sprung in der Hazardrate:  $jump = \{1; 0\}$  (1  $\hat{=}$  Sprung in der Hazardrate)
  - Change-Point  $\tau$ :
    - $\tau = \{90; 100\}$  für Szenario 1
    - $\tau = \{50; 55\}$  für Szenario 2
    - $\tau = \{360\}$  für Szenario 3
- variierende Parameter für die Schätzung:
  - Intervallbreite:  $wd = \{10; 20\}$
  - Schätzmethode:
    - *dis.int* (disjunkte Intervalle ohne Verschiebung)
    - *cont.int* (fortlaufende Intervalle)
    - *shift.int* (Intervallverschiebung)

Insgesamt resultieren daraus in den ersten beiden Szenarien 16 und im dritten 8 Parameterkombinationen für die Datengenerierung mit jeweils 6 Fällen für die Schätzung. Eine solche Kombination wird im Folgenden als “Setting“ bezeichnet. Die für die Schätzungen verwendete *R*-Funktion, welche die Grundlage für das erstellte Paket **CPest** darstellt, wird in Kapitel 4 ausführlich erklärt.

Die Parameter der Weibullverteilung, welche für die Datengenerierung benötigt werden, sind für alle Settings mit den Werten  $shape=0,2$  und  $scale=100$  vorbelegt. Die zugehörige theoretische Hazardfunktion ist für zwei beispielhafte Settings mit  $\tau = 100$  und  $\tau = 50$  in den Abbildungen 1 und 2 dargestellt, wobei jeweils die zwei Fälle mit und ohne Sprung in der Hazardrate betrachtet werden.

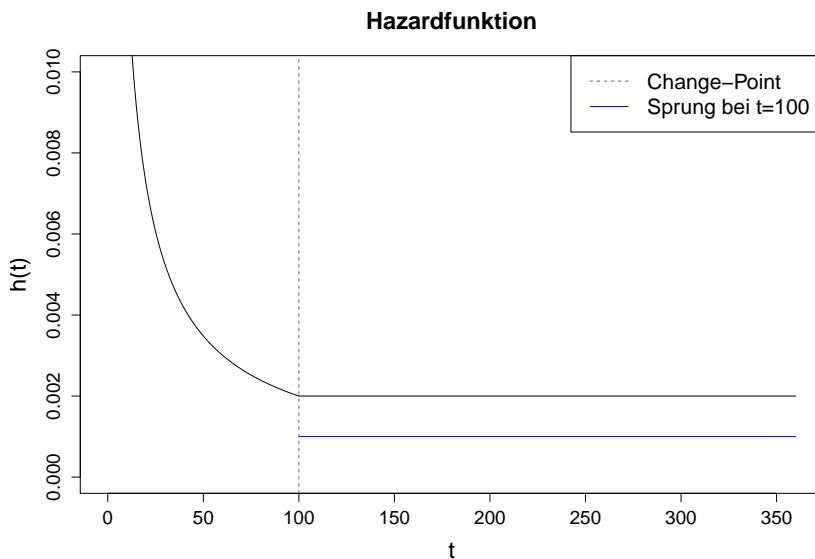


Abbildung 1: Hazardfunktion für  $\tau = 100$  mit und ohne Sprung;  
für die  $t_i$  gilt:  $(t_i|t_i < 100) \sim WB(\eta = 100, \gamma = 0,2)$  und  $(t_i|t_i \geq 100) \sim Exp(\lambda)$

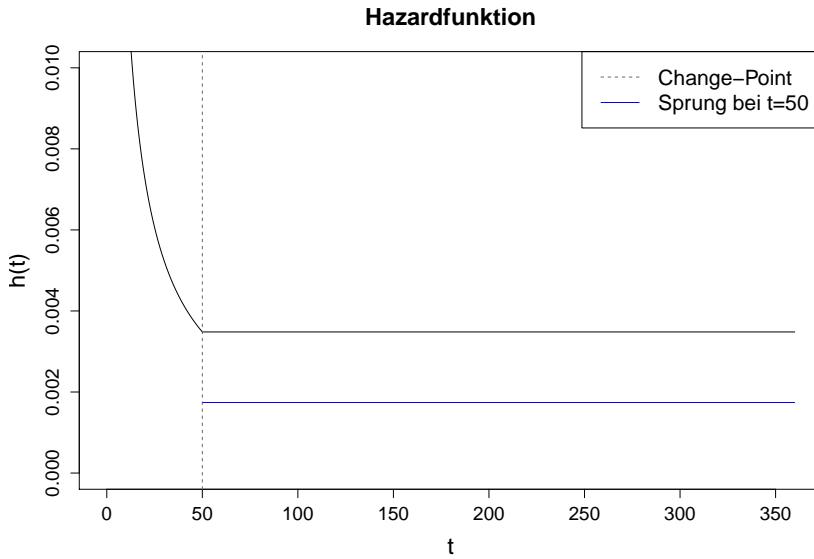


Abbildung 2: Hazardfunktion für  $\tau = 50$  mit und ohne Sprung;  
für die  $t_i$  gilt:  $(t_i|t_i < 50) \sim WB(\eta = 100, \gamma = 0, 2)$  und  $(t_i|t_i \geq 50) \sim Exp(\lambda)$

Weiterhin wird der Parameter  $\lambda_{cens}$  für die Erzeugung der Zensierungszeiten so vorgegeben, dass sich ein Anteil an zensierten Daten von ca. 30% ergibt.

Die Simulationen werden für jedes Setting mit 500 Wiederholungen durchgeführt. D.h. für jede Parameterkombination werden 500 Datensätze erzeugt und jeweils 6 Schätzer für  $\tau$  ermittelt. Die Simulationsergebnisse in Kapitel 3.3 beinhalten für jeden dieser Fälle den Median sowie das arithmetische Mittel aus allen 500 Schätzungen. Als weitere Maße für die Güte der Schätzfunktion wird der RMSE (root mean square error) und der MAD (mean absolute error) betrachtet, welche sich wie folgt berechnen lassen:

$$\text{RMSE} = \sqrt{\frac{1}{r} \sum_{i=1}^r (\hat{\tau}_i - \tau)^2}$$

$$\text{MAD} = \frac{1}{r} \sum_{i=1}^r |\hat{\tau}_i - \tau|$$

Dabei entspricht  $r$  der Anzahl an Wiederholungen, d.h. in diesem Fall  $r = 500$ .

Da die Funktion für die Change-Point-Schätzung so konstruiert ist, dass sich für  $\hat{\tau}$  maximal der Wert  $\tau_{max}$  ergeben kann, ist für die Simulationsergebnisse im Fall  $\tau = 360$  der Anteil der Schätzungen mit  $\hat{\tau} = \tau_{max}$  interessant, weil  $\tau_{max}$  in diesem Szenario ebenfalls mit 360 vorgegeben ist (siehe Kapitel 4 für Details zu der in R verwendeten Schätz-Funktion). Die zugehörigen Ergebnisse in Tabelle 4 enthalten deshalb eine zusätzliche Spalte "prop.tau max" für eben diesen Anteil, wobei der Übersichtlichkeit wegen nur jene Werte dargestellt sind, welche von 0 verschieden sind.

### 3.3 Simulationsergebnisse

Ein Ziel der Simulationsstudie ist es, festzustellen wie nahe der geschätzte Change-Point am tatsächlichen Wert ist und ob sich die Genauigkeit des Schätzers für große Stichprobenumfänge verbessert. Deshalb werden für alle Settings jeweils die Beobachtungszahlen  $n = 1000$  und  $n = 5000$  untersucht. Die Ergebnisse enthalten für jeden betrachteten Fall den Median und das arithmetische Mittel aus den jeweils 500 Schätzungen. Da in der Überlebenszeitanalyse der Median wegen seiner Robustheit gegen Ausreißer eine größere Rolle spielt, wird dieser im Folgenden gegenüber dem Mittelwert (“mean”) vorrangig betrachtet. Weiterhin ist für jeden Fall der RMSE und der MAD angegeben, wobei auch hier der MAD wegen seiner geringeren Ausreißerempfindlichkeit von größerem Interesse ist.

Zunächst soll analysiert werden, wie gut das Schätzverfahren theoretisch funktioniert, d.h. wie genau der Change-Point in einem Fall geschätzt wird, bei dem der Übergang zur konstanten Hazardrate besonders eindeutig ist. Dafür wurden die Settings mit  $jump=1$  erstellt, welche bei  $t = \tau$  einen Sprung in der Hazardrate besitzen. Nachfolgend wird das erste Szenario mit  $\tau = 90/100$  genauer untersucht, dessen Ergebnisse aus den Schätzungen in Tabelle 1 zu finden sind.

Tabelle 1: Ergebnisse für Szenario 1 mit  $\tau = 90 / 100$ :

Simulation mit festen Weibull-Parametern  $shape=0.2$  und  $scale=100$ ; je 500 Wiederholungen für jede Kombination der Parameter  $cens=\{1,0\}$ ,  $jump=\{1,0\}$ ,  $\tau=\{90,100\}$ ,  $n=\{1000,5000\}$ ; pro Kombination 6 Schätzungen für  $\tau$ : mit disjunkten Intervallen ohne und mit Verschiebung (dis.int, shift.int) und mit kontinuierlichen Intervallen (cont.int) mit einer Breite von je 10 und 20 sowie  $\tau_{max} = 360$ ; für jedes Setting sind die Ergebnisse für Median, Mittelwert, RMSE und MAD mit der jeweils besten Güte unterstrichen

	median	mean	RMSE	MAD
<i>cens=1; jump=1; τ=90; n=1000</i>				
dis.int10	<u>90.00</u>	88.34	13.81	<u>8.50</u>
dis.int20	100.00	102.20	36.43	20.84
cont.int10	83.00	83.88	15.30	11.10
cont.int20	82.00	<u>90.68</u>	31.59	17.25
shift.int10	87.00	86.48	<u>13.30</u>	8.60
shift.int20	92.00	97.96	30.61	15.51

	median	mean	RMSE	MAD
<i>cens=1; jump=1; tau=90; n=5000</i>				
dis.int10	<u>90.00</u>	95.26	12.74	<u>5.54</u>
dis.int20	100.00	114.92	43.92	25.88
cont.int10	87.00	<u>90.70</u>	11.03	6.82
cont.int20	87.00	101.62	40.07	18.95
shift.int10	89.00	93.37	<u>10.96</u>	5.66
shift.int20	96.00	106.54	36.57	17.80
<i>cens=1; jump=1; tau=100; n=1000</i>				
dis.int10	<u>100.00</u>	96.52	17.99	11.04
dis.int20	<u>100.00</u>	108.00	30.09	13.76
cont.int10	92.00	91.66	17.90	13.12
cont.int20	92.00	<u>97.96</u>	26.17	16.72
shift.int10	97.00	94.51	<u>15.60</u>	<u>10.84</u>
shift.int20	99.00	104.69	26.34	14.32
<i>cens=1; jump=1; tau=100; n=5000</i>				
dis.int10	<u>100.00</u>	104.58	11.77	<u>4.66</u>
dis.int20	<u>100.00</u>	112.80	31.92	12.80
cont.int10	97.00	<u>99.78</u>	<u>9.93</u>	6.49
cont.int20	96.00	107.84	34.71	15.88
shift.int10	99.00	103.22	11.17	5.64
shift.int20	99.00	111.57	31.13	13.81
<i>cens=1; jump=0; tau=90; n=1000</i>				
dis.int10	60.00	64.60	<u>35.22</u>	29.96
dis.int20	<u>80.00</u>	<u>80.60</u>	43.22	29.72
cont.int10	57.00	59.78	38.23	33.60
cont.int20	62.50	72.36	47.27	36.10
shift.int10	59.00	62.72	36.10	30.96
shift.int20	70.50	77.80	40.44	<u>29.18</u>
<i>cens=1; jump=0; tau=90; n=5000</i>				
dis.int10	<u>80.00</u>	78.30	26.23	18.90
dis.int20	<u>80.00</u>	<u>91.56</u>	44.23	24.12
cont.int10	71.00	72.79	25.15	21.56
cont.int20	70.00	80.05	40.26	28.16
shift.int10	76.00	77.13	<u>21.45</u>	<u>17.25</u>
shift.int20	78.00	86.92	36.69	21.93

	median	mean	RMSE	MAD
<i>cens=1; jump=0; tau=100; n=1000</i>				
dis.int10	70.00	69.32	37.82	34.08
dis.int20	<u>80.00</u>	<u>89.24</u>	49.08	35.96
cont.int10	64.00	65.51	40.18	36.91
cont.int20	67.00	81.38	51.59	41.20
shift.int10	66.00	68.36	<u>37.76</u>	34.34
shift.int20	77.00	86.28	42.63	<u>32.86</u>
<i>cens=1; jump=0; tau=100; n=5000</i>				
dis.int10	80.00	85.78	23.46	19.34
dis.int20	80.00	102.04	44.21	26.84
cont.int10	80.00	82.08	25.91	22.48
cont.int20	80.00	92.35	42.81	30.23
shift.int10	84.00	85.47	<u>23.39</u>	<u>19.27</u>
shift.int20	<u>91.00</u>	<u>99.24</u>	39.42	23.58
<i>cens=0; jump=1; tau=90; n=1000</i>				
dis.int10	<u>90.00</u>	<u>88.92</u>	10.94	<u>6.72</u>
dis.int20	100.00	99.24	23.68	16.36
cont.int10	84.00	83.71	12.26	9.24
cont.int20	83.00	87.96	19.85	12.75
shift.int10	88.00	87.54	<u>10.32</u>	7.22
shift.int20	92.00	96.42	19.88	11.59
<i>cens=0; jump=1; tau=90; n=5000</i>				
dis.int10	<u>90.00</u>	93.86	10.11	<u>3.98</u>
dis.int20	100.00	107.96	28.64	18.52
cont.int10	87.00	<u>89.51</u>	<u>9.44</u>	5.62
cont.int20	87.00	93.08	20.21	9.98
shift.int10	89.00	92.62	9.82	4.89
shift.int20	94.00	99.90	22.77	11.16
<i>cens=0; jump=1; tau=100; n=1000</i>				
dis.int10	<u>100.00</u>	<u>98.60</u>	<u>12.00</u>	7.08
dis.int20	<u>100.00</u>	105.12	17.20	<u>6.96</u>
cont.int10	93.00	93.32	12.39	9.64
cont.int20	92.00	95.94	17.28	12.12
shift.int10	98.00	96.45	12.38	8.22
shift.int20	99.00	103.05	14.75	9.01

	median	mean	RMSE	MAD
<i>cens=0; jump=1; tau=100; n=5000</i>				
dis.int10	<u>100.00</u>	104.50	10.85	<u>4.62</u>
dis.int20	<u>100.00</u>	109.84	22.87	9.84
cont.int10	97.00	<u>99.91</u>	<u>9.37</u>	5.83
cont.int20	97.00	103.58	19.40	10.44
shift.int10	99.00	103.04	10.76	5.07
shift.int20	99.00	108.85	20.04	10.67
<i>cens=0; jump=0; tau=90; n=1000</i>				
dis.int10	60.00	64.04	<u>29.95</u>	27.16
dis.int20	<u>80.00</u>	<u>75.88</u>	34.26	24.32
cont.int10	58.00	59.32	33.77	31.24
cont.int20	62.00	66.40	39.06	32.18
shift.int10	62.00	62.76	30.51	27.66
shift.int20	71.00	73.53	32.85	<u>24.27</u>
<i>cens=0; jump=0; tau=90; n=5000</i>				
dis.int10	<u>80.00</u>	77.40	<u>17.75</u>	15.52
dis.int20	<u>80.00</u>	<u>86.72</u>	27.90	17.00
cont.int10	73.00	73.18	20.11	18.46
cont.int20	70.00	75.78	28.47	22.35
shift.int10	77.00	77.78	17.88	<u>15.15</u>
shift.int20	79.00	84.43	25.43	16.65
<i>cens=0; jump=0; tau=100; n=1000</i>				
dis.int10	70.00	69.24	34.01	31.20
dis.int20	<u>80.00</u>	<u>82.28</u>	32.61	26.52
cont.int10	64.00	65.20	37.75	35.08
cont.int20	68.00	71.92	36.12	32.80
shift.int10	68.00	68.87	34.22	31.35
shift.int20	78.00	80.49	<u>29.15</u>	<u>24.69</u>
<i>cens=0; jump=0; tau=100; n=5000</i>				
dis.int10	<u>90.00</u>	85.16	<u>21.31</u>	17.88
dis.int20	80.00	<u>96.56</u>	31.92	20.96
cont.int10	80.00	80.53	24.26	21.56
cont.int20	80.00	86.40	31.47	25.55
shift.int10	85.00	84.85	21.32	<u>17.57</u>
shift.int20	90.00	94.86	28.71	18.82

Im Allgemeinen enthalten die simulierten Daten in den Fällen mit  $cens=0$ ;  $jump=1$  und  $n=5000$  die meiste Information bezüglich des Change-Points. Wir betrachten aus Szenario 1 deshalb zunächst die zwei Settings  $\{cens=0; jump=1; \tau=90/100; n=5000\}$ . Aus den zugehörigen Ergebnissen in Tabelle 1 ist ersichtlich, dass für diese beiden Settings sowohl der MAD als auch der RMSE die geringsten Werte im Vergleich zu allen anderen Settings aus Szenario 1 annehmen. Weiterhin ist der Median für alle drei Schätzverfahren nahe dem tatsächlichen Change-Point, wobei trotz des hohen Stichprobenumfangs von  $n=5000$  noch ein Bias (Verzerrung) vorhanden und der Schätzer somit nicht erwartungstreu ist. Es ist auffällig, dass der Median für die Schätzmethoden *cont.int* und *shift.int* meist unter dem wahren  $\tau$  liegt, was bereits an dieser Stelle darauf schließen lässt, dass der Change-Point tendenziell unterschätzt wird. Im Mittel liegen die geschätzten Werte zwar häufig über dem tatsächlichen  $\tau$ , allerdings wird das arithmetische Mittel durch einzelne hohe Schätzwerte nach oben verzerrt. Dass solche hohen Werte in der Simulation tatsächlich vorkommen, ist z.B. aus den Boxplots für den Fall  $\tau = 90$  in Abb. 3 ersichtlich.

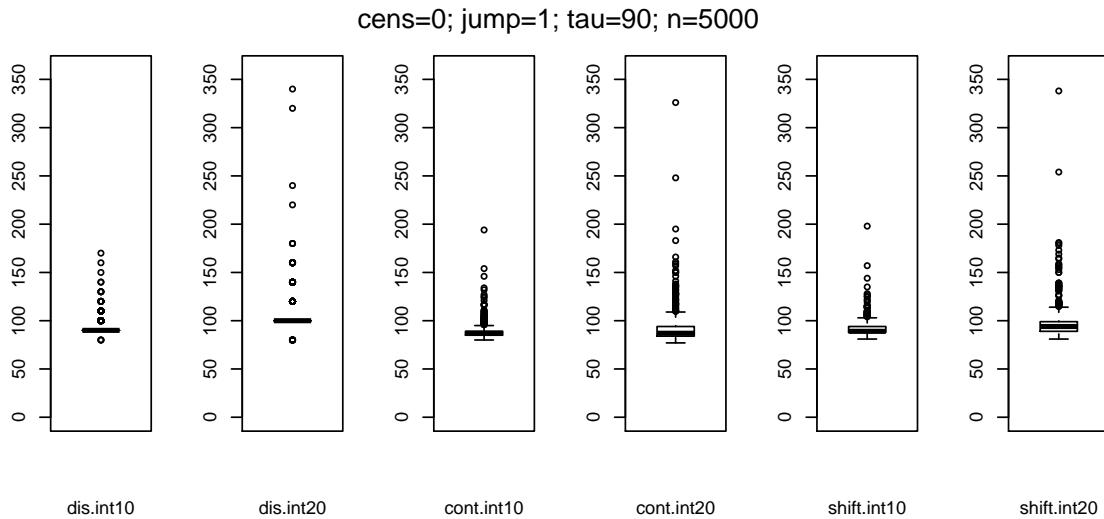


Abbildung 3: Boxplots der Change-Point-Schätzer für die drei Schätzverfahren mit Intervallbreiten 10 und 20 im Setting  $\{cens=0; jump=1; \tau=90; n=5000\}$

Für diese beiden Settings hat der Schätzwert, welcher ohne Intervallverschiebung berechnet wurde (*dis.int*), jeweils den kleinsten MAD. Eine mögliche Begründung könnte darin liegen, dass für eine Intervallbreite von 10 bei dieser Schätzmethode mit diskreten Intervallen die beiden tatsächlichen Werte von  $\tau$ , also 90 und 100, exakt auf den Intervallgrenzen liegen.

Um den Fall zu untersuchen, dass sich der wahre Change-Point an einem Zeitpunkt befindet, welcher kein Vielfaches von 10 ist, betrachten wir nun das Setting  $\{cens=0; jump=1; \tau=55; n=5000\}$  des zweiten Szenarios, dessen Ergebnisse in Tabelle 2 aufgelistet sind. Hierbei ist der MAD für die Verfahren mit kontinuierlichen Intervallen (*cont.int*) und mit Intervallverschiebung (*shift.int*) geringer als für die erste Schätzmethode, weil

mittels diskreter Intervalle ohne Verschiebung nur Vielfache von 10 bzw. 20 für  $\hat{\tau}$  möglich sind, die geschätzten Werte also mindestens den Abstand 5 zum wahren  $\tau = 55$  haben. Außerdem liegt der Median für *shift.int10* mit 56 sehr nahe am tatsächlichen Change-Point.

Tabelle 2: Ergebnisse für Szenario 2 mit  $\tau = 50 / 55$ :

Simulation mit festen Weibull-Parametern  $\text{shape}=0.2$  und  $\text{scale}=100$ ; je 500 Wiederholungen für jede Kombination der Parameter  $\text{cens}=\{1,0\}$ ,  $\text{jump}=\{1,0\}$ ,  $\tau=\{50,55\}$ ,  $n=\{1000,5000\}$ ; pro Kombination 6 Schätzungen für  $\tau$ : mit disjunkten Intervallen ohne und mit Verschiebung (dis.int, shift.int) und mit kontinuierlichen Intervallen (cont.int) mit einer Breite von je 10 und 20 sowie  $\tau_{\max} = 200$ ; für jedes Setting sind die Ergebnisse für Median, Mittelwert, RMSE und MAD mit der jeweils besten Güte unterstrichen

	median	mean	RMSE	MAD
<i>cens=1; jump=1; τ=50; n=1000</i>				
dis.int10	<u>50.00</u>	53.20	11.31	<u>5.52</u>
dis.int20	60.00	64.40	28.10	18.80
cont.int10	46.00	47.99	10.81	7.60
cont.int20	45.00	52.60	23.01	13.42
shift.int10	49.00	<u>51.13</u>	<u>9.81</u>	5.78
shift.int20	54.00	60.11	23.88	13.41
<i>cens=1; jump=1; τ=50; n=5000</i>				
dis.int10	<u>50.00</u>	56.16	16.60	6.16
dis.int20	60.00	72.60	35.16	22.76
cont.int10	47.00	<u>51.87</u>	<u>14.21</u>	6.97
cont.int20	47.00	58.17	28.44	13.93
shift.int10	49.00	54.19	<u>14.21</u>	<u>6.03</u>
shift.int20	55.00	64.31	30.03	15.31
<i>cens=1; jump=1; τ=55; n=1000</i>				
dis.int10	50.00	57.76	13.12	8.50
dis.int20	60.00	69.64	27.69	16.26
cont.int10	50.00	53.08	12.29	7.83
cont.int20	50.00	57.65	22.78	13.47
shift.int10	<u>54.00</u>	<u>56.72</u>	<u>11.07</u>	<u>6.51</u>
shift.int20	57.00	64.15	22.64	11.93

	median	mean	RMSE	MAD
<i>cens=1; jump=1; τ=55; n=5000</i>				
dis.int10	60.00	62.62	14.28	9.30
dis.int20	60.00	70.60	27.14	15.60
cont.int10	52.00	<u>55.61</u>	<u>10.08</u>	5.80
cont.int20	52.00	60.11	21.28	11.25
shift.int10	<u>56.00</u>	59.19	10.42	<u>5.61</u>
shift.int20	57.00	65.04	22.01	11.26
<i>cens=1; jump=0; τ=50; n=1000</i>				
dis.int10	<u>40.00</u>	42.40	19.90	14.68
dis.int20	<u>40.00</u>	54.52	32.39	20.32
cont.int10	35.00	38.14	21.60	17.84
cont.int20	34.00	44.36	31.94	23.82
shift.int10	39.00	41.40	<u>17.59</u>	<u>14.14</u>
shift.int20	39.00	<u>51.42</u>	29.08	19.20
<i>cens=1; jump=0; τ=50; n=5000</i>				
dis.int10	40.00	47.78	20.12	12.22
dis.int20	40.00	55.24	31.58	19.04
cont.int10	40.00	42.57	18.35	13.15
cont.int20	36.00	<u>48.80</u>	32.80	22.03
shift.int10	<u>45.00</u>	46.89	<u>15.75</u>	<u>9.29</u>
shift.int20	<u>45.00</u>	54.29	29.79	16.52
<i>cens=1; jump=0; τ=55; n=1000</i>				
dis.int10	40.00	45.22	19.70	<u>15.90</u>
dis.int20	40.00	<u>55.00</u>	29.25	20.04
cont.int10	38.00	39.93	22.18	19.38
cont.int20	36.00	44.74	28.99	23.46
shift.int10	41.00	43.55	<u>19.49</u>	15.96
shift.int20	<u>45.00</u>	52.36	25.88	17.54
<i>cens=1; jump=0; τ=55; n=5000</i>				
dis.int10	50.00	51.82	16.94	10.88
dis.int20	60.00	62.84	34.89	22.42
cont.int10	44.00	47.06	17.64	14.12
cont.int20	41.00	<u>52.05</u>	29.86	21.29
shift.int10	48.00	50.74	<u>15.08</u>	<u>10.52</u>
shift.int20	<u>51.00</u>	58.13	26.68	15.65

	median	mean	RMSE	MAD
<i>cens=0; jump=1; τ=50; n=1000</i>				
dis.int10	<u>50.00</u>	52.68	<u>8.56</u>	<u>4.20</u>
dis.int20	60.00	64.80	27.03	18.24
cont.int10	46.00	47.56	8.90	6.36
cont.int20	45.00	<u>51.16</u>	20.17	11.70
shift.int10	49.00	51.37	8.90	5.04
shift.int20	55.00	59.23	20.47	11.97
<i>cens=0; jump=1; τ=50; n=5000</i>				
dis.int10	<u>50.00</u>	54.84	11.66	4.84
dis.int20	60.00	70.56	30.76	20.60
cont.int10	48.00	<u>51.14</u>	9.81	5.73
cont.int20	47.00	55.22	19.29	10.44
shift.int10	49.00	52.88	<u>9.41</u>	<u>4.78</u>
shift.int20	55.00	60.57	20.44	11.58
<i>cens=0; jump=1; τ=55; n=1000</i>				
dis.int10	60.00	58.06	12.35	8.32
dis.int20	60.00	68.76	24.82	14.72
cont.int10	51.00	52.32	<u>9.23</u>	6.95
cont.int20	50.00	<u>56.09</u>	19.57	11.80
shift.int10	<u>54.00</u>	56.34	9.73	<u>6.23</u>
shift.int20	57.00	62.67	19.88	10.54
<i>cens=0; jump=1; τ=55; n=5000</i>				
dis.int10	60.00	62.92	14.23	9.26
dis.int20	60.00	69.76	26.79	14.76
cont.int10	52.00	<u>54.80</u>	<u>8.12</u>	<u>4.83</u>
cont.int20	52.00	59.59	21.90	10.68
shift.int10	<u>56.00</u>	58.55	9.58	4.93
shift.int20	58.00	65.51	24.06	11.54
<i>cens=0; jump=0; τ=50; n=1000</i>				
dis.int10	<u>40.00</u>	42.42	17.33	13.66
dis.int20	<u>40.00</u>	52.88	26.46	17.28
cont.int10	35.00	36.93	18.05	16.18
cont.int20	33.00	41.90	27.05	21.35
shift.int10	39.00	40.45	<u>14.98</u>	<u>12.55</u>
shift.int20	<u>40.00</u>	<u>49.23</u>	23.52	15.99

	median	mean	RMSE	MAD
<i>cens=0; jump=0; τ=50; n=5000</i>				
dis.int10	40.00	44.40	14.44	10.56
dis.int20	40.00	52.32	27.58	16.92
cont.int10	40.00	40.57	15.33	12.51
cont.int20	36.00	43.45	26.14	18.94
shift.int10	<u>44.00</u>	45.15	<u>12.81</u>	<u>8.15</u>
shift.int20	43.00	<u>49.58</u>	21.34	13.14
<i>cens=0; jump=0; τ=55; n=1000</i>				
dis.int10	40.00	43.42	<u>16.51</u>	<u>13.90</u>
dis.int20	40.00	<u>52.64</u>	25.00	17.86
cont.int10	37.00	38.34	19.96	17.98
cont.int20	36.00	42.35	25.91	21.70
shift.int10	<u>40.50</u>	42.26	16.54	14.43
shift.int20	43.00	49.05	20.94	15.43
<i>cens=0; jump=0; τ=55; n=5000</i>				
dis.int10	50.00	50.12	12.98	<u>8.72</u>
dis.int20	60.00	58.68	28.34	18.14
cont.int10	45.00	45.00	14.40	12.48
cont.int20	41.00	47.01	22.76	17.65
shift.int10	48.00	49.30	<u>12.00</u>	8.76
shift.int20	<u>51.00</u>	<u>54.95</u>	20.98	12.20

Insgesamt fallen in beiden Szenarien die Schätzergebnisse für  $cens=1$ , d.h. für jene Settings in denen die Personen unter Risiko um die censierten Beobachtungen korrigiert wurden (vgl. Kap. 2.3), ähnlich zu denen mit  $cens=0$  aus. Bereits aus diesen, im Bezug auf die Lage des Change-Points, eindeutigen Daten ist erkennbar, dass bei allen drei Schätzmethoden im Allgemeinen eine Unterschätzung stattfindet. Wesentlich deutlicher wird dies aus den Ergebnissen der Settings ohne Sprung in der Hazardrate ( $jump=0$ ), was eher der Realität entspricht, da bei tatsächlich beobachteten Daten der Übergang von der anfänglich abfallenden zu einer konstanten Sterberate im Allgemeinen stetig ist.

Als direkten Vergleich zu den bisher untersuchten Settings mit  $jump=1$  betrachten wir nun den Fall  $\{cens=0; jump=0; \tau=90; n=5000\}$  ohne Sprung aus Szenario 1. Hier ist der Bias deutlich größer und es findet eine noch stärkere Unterschätzung statt, insbesondere weicht der Median über 10 Tage vom tatsächlichen  $\tau$  ab, für  $cont.int20$  sogar um 20 Tage. Auch der MAD ist hier wesentlich höher als in den Settings mit Sprung. Für  $n=1000$  ist

die Differenz zwischen dem wahren und dem geschätzten Change-Point noch größer, was sowohl am Median als auch am arithmetischen Mittel der geschätzten Werte erkennbar ist. In den übrigen Settings ohne Sprung zum Zeitpunkt  $\tau$  sind die Abweichungen nach unten ähnlich bzw. teilweise noch stärker ausgeprägt.

Im 2. Szenario ist der Bias für  $jump=0$  im Allgemeinen geringer als in Szenario 1, jedoch wird der wahre Change-Point auch hier tendenziell unterschätzt. Beispielsweise ist das arithmetische Mittel im Setting  $\{cens=0; jump=0; \tau=55; n=5000\}$  bei 54,95, d.h. im Mittel wurde der wahre Change-Point fast exakt geschätzt. Allerdings ist auch hier zu beachten, dass es viele Schätzungen mit großen Werten gibt, welche den Mittelwert nach oben verzerren (vgl. zugehörigen Boxplot im Anhang A.2).

Möglicherweise sind die Abweichungen zwischen den geschätzten Werten und dem tatsächlichen  $\tau$  geringer, weil sich der Change-Point mit  $\tau = 50$  bzw.  $\tau = 55$  an einer Stelle befindet, an welcher die Hazardfunktion der Weibullverteilung relativ steil ist und somit der Übergang zur konstanten Hazardrate auch ohne Sprung eindeutiger ist als in Szenario 1 (vgl. Abb. 2 zur theoretischen Hazardfunktion).

Um zu untersuchen, welche Schätzungen sich im Fall  $\tau = \tau_{max}$  ergeben, d.h. wenn der wahre Change-Point dem maximal in Frage kommenden  $\hat{\tau}$  entspricht, wurde ein weiteres Szenario mit  $\tau = 360$  erstellt. Die entsprechenden Simulationsergebnisse sind in Tabelle 4 zu finden, welche analog zu denen der ersten beiden Szenarien dargestellt sind. Zusätzlich gibt es nun jedoch eine Spalte “prop.taumax“ für den Anteil der Schätzungen mit  $\hat{\tau} = \tau_{max}$ .

Besonders auffällig ist in den Ergebnissen der Fall  $\{cens=0; jump=1; \tau=360; n=5000\}$ , welcher die Schätzung mit dem geringsten MAD (=1.04) aus allen Settings der drei betrachteten Szenarien enthält. Dabei ist jedoch zu beachten, dass der Anteil für  $\hat{\tau} = \tau_{max}$  für eben dieses Ergebnis, also den Fall *dis.int20*, 95% entspricht, wobei nicht bekannt ist, wie oft dabei eine Überschätzung des wahren Change-Points stattgefunden hat. Da die Schätzungen ausschließlich nach unten vom tatsächlichen Wert für  $\tau$  abweichen können, ist der MAD in diesem Setting entsprechend gering. Im Fall des kleineren Stichprobenumfangs von  $n=1000$  weichen die Schätzungen generell sehr weit nach unten ab, sodass der Anteil für  $\hat{\tau} = \tau_{max}$  dann kaum mehr eine Rolle spielt.

Tabelle 3: Ergebnisse für Szenario 3 mit  $\tau = 360$ :

Simulation mit festen Weibull-Parametern shape=0.2 und scale=100; je 500 Wiederholungen für jede Kombination der Parameter cens={1,0}, jump={1,0},  $\tau=\{360\}$ , n={1000,5000}; pro Kombination 6 Schätzungen für  $\tau$ : mit disjunkten Intervallen ohne und mit Verschiebung (dis.int, shift.int) und mit kontinuierlichen Intervallen (cont.int) mit einer Breite von je 10 und 20 sowie  $\tau_{max} = 360$ ; für jedes Setting sind die Ergebnisse für Median, Mittelwert, RMSE und MAD mit der jeweils besten Güte unterstrichen; “prop.taumax“ gibt den Anteil der Schätzungen mit  $\hat{\tau} = \tau_{max}$  an

	median	mean	RMSE	MAD	prop.taumax
<i>cens=1; jump=1; <math>\tau=360</math>; n=1000</i>					
dis.int10	200.00	205.56	160.61	154.44	
dis.int20	<u>260.00</u>	<u>269.96</u>	<u>105.73</u>	<u>90.04</u>	0.11
cont.int10	197.00	200.13	165.60	159.87	
cont.int20	258.00	262.78	111.20	97.22	0.01
shift.int10	189.00	194.22	171.29	165.78	
shift.int20	255.00	255.85	117.29	104.15	0.02
<i>cens=1; jump=1; <math>\tau=360</math>; n=5000</i>					
dis.int10	340.00	332.00	41.12	28.00	0.31
dis.int20	<u>360.00</u>	<u>350.00</u>	<u>21.61</u>	<u>10.00</u>	0.72
cont.int10	336.00	328.44	43.26	31.56	0.04
cont.int20	348.00	343.13	24.95	16.87	0.11
shift.int10	333.50	325.95	46.27	34.05	0.07
shift.int20	356.00	345.41	26.04	14.59	0.22
<i>cens=1; jump=0; <math>\tau=360</math>; n=1000</i>					
dis.int10	140.00	148.84	214.73	211.16	
dis.int20	<u>180.00</u>	<u>190.60</u>	<u>177.23</u>	<u>169.40</u>	
cont.int10	143.00	146.02	217.24	213.98	
cont.int20	177.50	182.80	184.34	177.20	
shift.int10	139.00	145.00	218.24	215.00	
shift.int20	178.00	184.37	182.64	175.63	
<i>cens=1; jump=0; <math>\tau=360</math>; n=5000</i>					
dis.int10	230.00	231.18	134.17	128.82	
dis.int20	<u>260.00</u>	<u>263.68</u>	<u>106.01</u>	<u>96.32</u>	0.03
cont.int10	222.00	226.12	138.73	133.88	
cont.int20	253.00	255.90	112.22	104.10	0.01
shift.int10	221.50	224.78	140.14	135.22	
shift.int20	258.00	260.15	108.67	99.85	0.01

	median	mean	RMSE	MAD	prop.taumax
<i>cens=0; jump=1; <math>\tau=360</math>; n=1000</i>					
dis.int10	260.00	259.74	113.22	100.26	0.03
dis.int20	320.00	<u>320.48</u>	<u>55.61</u>	<u>39.52</u>	0.32
cont.int10	259.00	257.66	113.40	102.34	
cont.int20	<u>323.00</u>	313.72	59.72	46.28	0.03
shift.int10	249.00	249.19	122.02	110.81	0.01
shift.int20	315.50	309.00	66.23	51.00	0.07
<i>cens=0; jump=1; <math>\tau=360</math>; n=5000</i>					
dis.int10	<u>360.00</u>	355.04	12.00	4.96	0.72
dis.int20	<u>360.00</u>	<u>358.96</u>	<u>4.56</u>	<u>1.04</u>	0.95
cont.int10	353.00	351.13	13.04	8.87	0.09
cont.int20	353.00	351.83	10.65	8.17	0.12
shift.int10	358.00	353.27	12.42	6.73	0.19
shift.int20	359.00	356.78	6.62	3.22	0.33
<i>cens=0; jump=0; <math>\tau=360</math>; n=1000</i>					
dis.int10	160.00	167.70	195.65	192.30	
dis.int20	200.00	<u>210.76</u>	<u>156.57</u>	<u>149.24</u>	0.01
cont.int10	162.00	164.36	198.78	195.64	
cont.int20	<u>201.50</u>	202.97	163.70	157.03	
shift.int10	159.00	161.64	201.42	198.36	
shift.int20	199.00	203.90	162.38	156.10	
<i>cens=0; jump=0; <math>\tau=360</math>; n=5000</i>					
dis.int10	260.00	256.28	109.64	103.72	
dis.int20	<u>280.00</u>	<u>286.96</u>	<u>81.13</u>	<u>73.04</u>	0.03
cont.int10	252.50	252.13	113.22	107.87	
cont.int20	278.00	277.22	89.71	82.78	0.01
shift.int10	249.00	251.10	114.31	108.90	
shift.int20	279.00	280.46	87.16	79.54	0.01

Somit geht auch aus den Ergebnissen dieses dritten Szenarios hervor, dass der wahre Change-Point tendenziell unterschätzt wird. Allerdings ist für  $\tau = 360$  der Unterschied zwischen den Schätzungen mit  $jump=1$  und  $jump=0$  bzgl. der Abweichung vom tatsächlichen  $\tau$  noch extremer als in den beiden zuvor betrachteten Szenarien. Dabei sollte jedoch beachtet werden, dass die Aussagekraft der Ergebnisse mit  $jump=0$  in diesem Szenario aufgrund der sehr hohen Werte des RMSE und des MAD eher gering ist. Auch aus den

Boxplots, wie beispielhaft für das Setting  $\{cens=0; jump=0; \tau=360; n=5000\}$  in Abb. 4 dargestellt, ist erkennbar, dass für  $\tau = 360$  sowohl der Interquartilsabstand als auch der Abstand zwischen den geschätzten Minima und Maxima deutlich größer ist, als in den Settings der beiden anderen Szenarien. Eine mögliche Begründung liegt darin, dass die Hazardrate der Weibullverteilung für die in den Simulationen gewählten Parameter zum Zeitpunkt  $t = 360$  bereits vergleichsweise flach ist, sodass sich aus der Change-Point-Schätzung in vielen Fällen ein deutlich geringerer Wert ergibt (vgl. Plot der theoretischen Hazardfunktion in Abb.13 im Anhang).

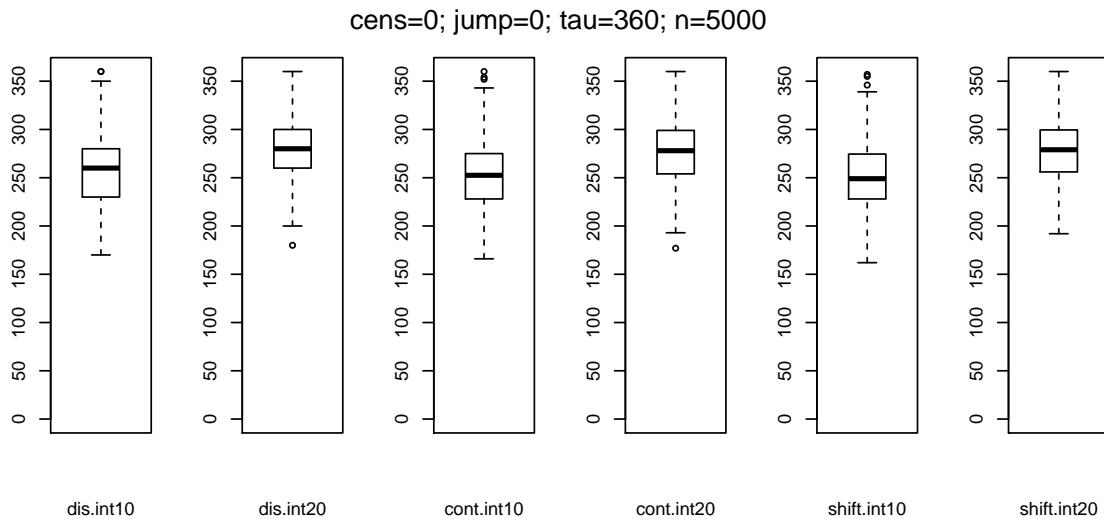


Abbildung 4: Boxplots der Change-Point-Schätzer für die drei Schätzverfahren mit Intervallbreiten 10 und 20 im Setting  $cens=0; jump=0; \tau=360; n=5000$

Die Boxplots für alle übrigen Settings der beiden Szenarien sind im Anhang A.2 zu finden.

Um einen Überblick zu erhalten, für welche der drei Methoden die Schätzungen die größte Genauigkeit haben, sind in jedem Setting jeweils für RMSE und MAD die Ergebnisse mit dem geringsten Wert unterstrichen, und für Median und Mittelwert jene, welche am nächsten am vorgegebenen  $\tau$  liegen. Zum einen ist insgesamt auffällig, dass die Schätzungen, welche mit einer Intervallbreite von 10 Tagen durchgeführt wurden, in der Mehrheit der Fälle eine höhere Güte aufweisen. Bezuglich des MAD trifft dies im zweiten Szenario mit  $\tau = 50/55$  sogar auf alle untersuchten Settings zu. Auffällig ist, dass meist dann die Schätzung mit Intervallbreite 20 die besseren Ergebnisse liefert, wenn  $\tau$  ein Vielfaches von 20 ist, insbesondere ist die Güte dieser Schätzungen in allen Settings mit  $\tau = 360$  höher.

Weiterhin ist der RMSE für *cont.int* nur in 7 der 40 Settings am geringsten, der zugehörige MAD sogar nur in einem Fall, nämlich für  $\{cens=0; jump=1; \tau=55; n=5000\}$ . Aus diesem Grund und weil bei Verwendung dieser Schätzmethode das Problem der Abhängigkeit zwischen den p-Werten verschiedener Intervalle besteht (vgl. Kap. 2.2), sollten die

beiden Verfahren mit diskreten Intervallen gegenüber jenem mit fortlaufenden Intervallen bevorzugt werden. Deshalb wird die Funktion für die Schätzung mittels kontinuierlicher Intervalle auch bei der Implementierung des *R*-Paketes nicht mit berücksichtigt. Aus den Ergebnissen für *dis.int* und *shift.int* ist jedoch nicht eindeutig feststellbar, welche Methode eine höhere Genauigkeit aufweist, wobei zum Teil trotzdem eine gewisse Systematik erkennbar ist. Beispielsweise liefert in Szenario 1 für fast alle Settings mit Sprung die Methode *dis.int* einen geringeren MAD und für die Settings ohne Sprung die Methode *shift.int*. Man könnte an dieser Stelle vermuten, dass das Verfahren mit Intervallverschiebung für Daten mit einem stetigen Übergang zur konstanten Hazardrate besser geeignet ist, allerdings wird dies im zweiten Szenario nicht bestätigt, da die Ergebnisse dort nicht diese Systematik aufweisen. Hierfür sind gegebenenfalls noch weitere Tests sinnvoll.

## 3.4 Weiterführende Simulationen

### 3.4.1 Schätzungen mit tatsächlichem $\lambda$

Um zu untersuchen, ob die großen Abweichungen vom wahren Change-Point in den Settings ohne Sprung ihre Ursache in der Schätzung der konstanten Hazardrate  $\lambda$  haben, wurde eine zusätzliche Simulation durchgeführt. Hierbei wurde der Change-Point für einen simulierten Datensatz jeweils einmal wie bisher mit  $\hat{\lambda}$  und zusätzlich mit dem wahren  $\lambda$  geschätzt, welches sich aus der Weibull-Hazardfunktion berechnen lässt. Die verwendete konstante Hazardrate ergibt sich nach (7) aus Kapitel 3.1 zu  $\lambda = \frac{\gamma}{\eta} \left( \frac{\tau}{\eta} \right)^{\gamma-1} \approx 0,002$ . Betrachtet werden nur die Settings mit *cens=1/0*, *jump=0* und *n=5000*, da im ersten und zweiten Szenario in den Fällen ohne Sprung zum Zeitpunkt  $t = \tau$  die Abweichung vom wahren Change-Point wesentlich größer war als mit Sprung in der Hazardrate. Beispielsweise soll nur der Fall  $\tau = 90$  betrachtet werden, wobei  $\hat{\tau}$  jeweils mit  $\hat{\lambda}$  und mit wahren  $\lambda$  geschätzt wurde.

Aus den Ergebnissen in Tabelle 4 ist ersichtlich, dass im Fall *cens=1* sich der Median für die Schätzungen mit wahrem  $\lambda$  und jene mit  $\hat{\lambda}$  nur wenig unterscheidet. Sowohl der RMSE und der MAD sind zwar mit tatsächlichem  $\lambda$  im Allgemeinen geringer, da es anscheinend weniger Ausreißer in den Schätzungen gibt, was auch aus den Mittelwerten ersichtlich ist, jedoch wird der Change-Point nach wie vor deutlich unterschätzt. Noch geringer sind die Unterschiede, wenn in der Schätzung keine Korrektur um zensierte Beobachtungen stattfindet. Hier weichen sämtliche Ergebnisse für die Schätzung mit  $\hat{\lambda}$  noch weniger von denen mit wahrem  $\lambda$  ab. Die Ursache für die tendenzielle Unterschätzung des wahren Change-Points scheint somit nicht an der Schätzung der konstanten Hazardrate zu liegen.

Tabelle 4: Vergleich der Schätzungen mit geschätzter Hazardrate  $\hat{\lambda}$  und wahrem  $\lambda$ :  
 Simulation mit festen Weibull-Parametern shape=0.2 und scale=100; je 500 Wiederholungen für jede Kombination der Parameter cens={1,0}, jump=0,  $\tau=90$ , n=5000; pro Kombination 6 Schätzungen für  $\tau$ : mit disjunkten Intervallen ohne und mit Verschiebung (dis.int, shift.int) und mit kontinuierlichen Intervallen (cont.int) mit einer Breite von je 10 und 20 sowie  $\tau_{max} = 360$

	median	mean	RMSE	MAD
<i>cens=1; jump=0; <math>\tau=90</math>; n=5000; mit <math>\hat{\lambda}</math></i>				
dis.int10	70.00	76.20	21.17	17.60
dis.int20	80.00	97.40	54.59	29.60
cont.int10	73.50	74.28	23.90	20.52
cont.int20	70.50	89.69	54.16	34.01
shift.int10	76.00	76.50	21.58	17.46
shift.int20	79.00	93.59	46.63	26.03
<i>cens=1; jump=0; <math>\tau=90</math>; n=5000; mit wahrem <math>\lambda</math></i>				
dis.int10	80.00	76.00	18.60	16.20
dis.int20	80.00	83.20	17.55	14.80
cont.int10	73.00	73.14	20.22	18.60
cont.int20	70.00	73.93	22.90	21.19
shift.int10	76.00	75.39	18.46	15.85
shift.int20	79.00	82.90	17.04	14.04
<i>cens=0; jump=0; <math>\tau=90</math>; n=5000; mit <math>\hat{\lambda}</math></i>				
dis.int10	70.00	74.60	18.33	16.20
dis.int20	80.00	85.20	26.91	16.80
cont.int10	69.00	70.56	21.16	19.58
cont.int20	70.00	74.12	28.12	23.68
shift.int10	74.50	74.93	17.51	15.39
shift.int20	79.00	82.39	21.33	14.69
<i>cens=0; jump=0; <math>\tau=90</math>; n=5000; mit wahrem <math>\lambda</math></i>				
dis.int10	70.00	74.50	18.14	16.10
dis.int20	80.00	81.40	16.85	14.20
cont.int10	70.00	70.60	21.13	19.52
cont.int20	70.00	71.19	22.35	21.01
shift.int10	75.00	74.74	17.71	15.58
shift.int20	79.00	79.98	15.13	12.82

### 3.4.2 Untersuchung der Güte der Konfidenzintervalle

Für die ersten beiden Schätzmethoden, d.h. mit diskreten und kontinuierlichen Intervallen, wurde für die Fälle  $\{cens=1; jump=0/1; \tau=90; n=1000/5000\}$  zusätzlich eine Simulation mit Bootstrapping durchgeführt. Im Gegensatz zu den bisherigen Simulationen wurden die vier Schätzungen hierbei nicht für ein und denselben Datensatz eines Settings berechnet, sondern für jedes der Verfahren *dis.int* und *cont.int* mit Intervallbreiten 10 und 20 jeweils 500 Datensätze erzeugt und die Change-Point-Schätzungen ermittelt. Der Grund liegt in der hohen Rechenzeit bei Verwendung des Bootstrap-Verfahrens. Die daraus resultierenden Ergebnisse für Median, Mittelwert, RMSE und MAD sind in bekannter Form in Tabelle 5 aufgelistet. Zu beachten ist dabei, dass die Ergebnisse aufgrund des gleichen Simulationsaufbaus mit denen aus Szenario 1 qualitativ übereinstimmen, die genauen Werte jedoch leicht abweichen, was auf die Streuung der mit Zufallszahlen erzeugten Daten zurückzuführen ist.

Zusätzlich wurden nun für jeden dieser 500 Datensätze 1000 Bootstrap-Samplings mit zugehöriger Change-Point-Schätzung durchgeführt und die zugehörigen Bootstrap-Perzentil-Intervalle zum Niveau  $\alpha=5\%$  berechnet. Details zu diesen Konfidenzintervallen, welche nicht wie üblich auf einer Normal-Approximation beruhen, sondern die tatsächlich beobachtete Verteilung der  $\hat{\tau}$  verwenden, sind in Kapitel 4.1.3 zu finden. Aus den resultierenden Intervallen wurde berechnet, wie hoch der Anteil jener Schätzungen ist, für die der wahre Change-Point  $\tau = 90$  vom zugehörigen Intervall überdeckt wird. Diese Überdeckungsrate ist in Tabelle 5 in der Spalte “cover” zu finden.

Vergleicht man für jedes Ergebnis die Fälle  $jump=1$  und  $jump=0$ , so ist die Überdeckungsrate für die Daten mit Sprung fast immer höher, wobei der wahre Change-Point dabei auch deutlich weniger unterschätzt wird, wie sich auch bereits aus den Simulationen in Kapitel 3.3 ergeben hat. Im Setting  $\{jump=1; \tau=90; n=5000\}$ , in welchem die Daten wieder die meiste Information bezüglich des Change-Points enthalten, überdecken die Konfidenzintervalle für die Schätzung *dis.int10* den Zeitpunkt  $\tau$  sogar zu 100%. Insgesamt liegt die Überdeckungsrate jedoch in den meisten der betrachteten Fälle unter 90%, für  $jump=0$  sogar oft deutlich darunter.

Allerdings wurden in dieser Simulation zu wenige Szenarien betrachtet, um aus den Ergebnissen allgemeine Aussagen treffen zu können. Ein Grund, weshalb nur diese vier Settings untersucht wurden, liegt in der hohen Rechenzeit. Insbesondere dauert die Durchführung der Simulation für eines dieser 16 betrachteten Fälle mit den hier verwendeten Parametern bis zu 10 Tage. Deshalb ist die Simulation mit Bootstrapping Gegenstand weiterer Untersuchungen außerhalb dieser Arbeit, wofür die in *R* geschriebene Funktion zur Change-Point-Schätzung zunächst hinsichtlich der Laufzeit optimiert werden sollte.

Tabelle 5: Simulation mit Bootstrap-Samplings:

je 500 Wiederholungen für jede Kombination der Parameter  $\text{cens}=1$ ,  $\text{jump}=\{1,0\}$ ,  $\tau=90$ ,  $n=\{1000,5000\}$ ; pro Kombination 4 Schätzungen für  $\tau$ : mit disjunkten Intervallen ohne Verschiebung (dis.int) und fortlaufenden Intervallen (cont.int) einer Breite von je 10 und 20 sowie  $\tau_{max} = 360$ ; 'cover' bezeichnet die Überdeckungsrate der Perzentil-Intervalle

	median	mean	RMSE	MAD	cover
<i>jump=1; <math>\tau=90</math>; n=1000</i>					
dis.int10	90.00	87.02	13.45	8.34	0.96
dis.int20	100.00	98.44	28.24	17.72	0.97
cont.int10	84.00	83.19	13.68	10.08	0.73
cont.int20	83.00	89.35	25.27	15.03	0.89
<i>jump=1; <math>\tau=90</math>; n=5000</i>					
dis.int10	90.00	95.30	16.63	5.62	1.00
dis.int20	100.00	108.84	34.58	20.12	0.70
cont.int10	87.00	90.83	14.79	6.99	0.87
cont.int20	87.00	98.95	35.07	16.21	0.95
<i>jump=0; <math>\tau=90</math>; n=1000</i>					
dis.int10	60.00	63.88	33.40	29.32	0.64
dis.int20	80.00	82.56	41.98	29.20	0.85
cont.int10	57.00	60.14	36.37	33.01	0.46
cont.int20	62.00	73.29	47.89	37.16	0.76
<i>jump=0; <math>\tau=90</math>; n=5000</i>					
dis.int10	80.00	79.30	23.46	18.38	0.84
dis.int20	80.00	97.20	51.20	27.88	0.91
cont.int10	72.00	74.35	28.00	21.58	0.64
cont.int20	70.00	84.21	47.77	31.04	0.83

### 3.5 Fazit zu den Simulationsergebnissen

Zusammenfassend lässt sich aus den Simulationsergebnissen feststellen, dass der wahre Change-Point bei Schätzung mithilfe des Verfahrens nach [Küchenhoff u. a., 2015] tendenziell unterschätzt wird. Besonders stark sind die Abweichungen des geschätzten Wertes  $\hat{\tau}$  vom wahren  $\tau$  im Fall eines stetigen Übergangs in der Hazardrate zum Zeitpunkt  $t = \tau$ . Dies wirkt sich auch auf die Überdeckungsrate der Konfidenzintervalle aus, welche in den Ergebnissen aus Kapitel 3.4.2 für die meisten Fälle nicht zufriedenstellend ist, was weitere Tests erforderlich macht.

Bei Verwendung der Schätz-Methoden mit Intervallverschiebung oder fortlaufenden Intervallen ergeben sich zum Teil noch etwas größere Abweichungen nach unten als für das Verfahren ohne Intervallverschiebung, was jedoch auch daran liegt, dass für letzteres ausschließlich die Intervallgrenzen als Schätzergebnis in Frage kommen können. Aus diesem Grund liefert das Verfahren mit Intervallverschiebung in den Fällen, bei denen der Change-Point nicht als Vielfaches von 10 definiert wurde (d.h. für  $\tau=55$ ), sowohl im Bezug auf den Median als auch auf den MAD bessere Ergebnisse. Weiterhin ist aus den Ergebnissen ersichtlich, dass der MAD für die Methode mit fortlaufenden Intervallen (*cont.int*) in fast allen untersuchten Fällen am größten ist. Für die übrigen zwei Methoden ist jedoch keine eindeutige Tendenz zu erkennen. Deshalb wird für das R-Paket auf die Implementierung der Funktion für fortlaufende Intervalle verzichtet und nur die Methoden mit diskreten Intervallen (mit und ohne Verschiebung) berücksichtigt.

Aus den Ergebnissen für die Settings mit Sprung in der Hazardrate wird deutlich, dass die Schätzung mithilfe der aus einem Binomialtest ermittelten p-Werte nach [Küchenhoff u. a., 2015] theoretisch funktioniert. D.h. in einem Fall mit eindeutigem Übergang zwischen abfallender und konstanter Sterberate ist der Bias für große Stichprobenumfänge relativ gering. Gegebenenfalls sind weitere Nachforschungen notwendig, um die Ursache für die vergleichsweise großen Abweichungen vom wahren Change-Point im Fall mit stetigem Übergang in der Hazardrate zu finden.

## 4 **R**-Paket CPest

Im Zuge dieser Arbeit wurde eine erste Version des *R*-Paketes *CPest* geschrieben, welches eine Funktion zur Change-Point-Schätzung auf Grundlage des Verfahrens von [Küchenhoff u.a., 2015] beinhaltet. Zusätzlich ist eine Varianz-Schätzung mittels eines nichtparametrischen Bootstrap-Verfahrens möglich, wobei auch zwei Konfidenzintervalle für ein gewünschtes  $\alpha$ -Niveau ausgegeben werden. Für die grafische Darstellung der Schätzungen ist außerdem eine plot-Funktion eingearbeitet.

### 4.1 Aufbau und Funktionsweise

Das Paket enthält die zentrale Funktion `CPest` für die Schätzung von  $\tau$ , welche durch den Benutzer aufgerufen werden kann. Nachfolgende Aufzählung gibt eine Übersicht über die Argumente, welche der Funktion übergeben werden und zum Teil bereits mit default-Werten vorbelegt sind.

- `time`: Vektor mit beobachteten Überlebens-/Zensierungszeiten  $T_i$
- `event`: Vektor mit zugehörigem Status für jede Beobachtung  
(0=zensiert / 1=unzensiert)
- `intwd`: gewünschte Intervallbreite
- `taumax`: maximal für  $\tau$  angenommener Wert ( $\tau_{max}$ )
- `taumin = 0`: optionales Minimum für den Change-Point-Schätzer
- `shift = TRUE`: Indikator für die Schätzung mit (default) oder ohne Intervallverschiebung
- `cens = TRUE`: Berücksichtigung der zensierten Daten
- `boot = FALSE`: Erzeugung von Bootstrap-Samples
- `alpha = 0.05`: Konfidenzniveau
- `B = 1000`: Anzahl der Wiederholungen beim Bootstrapping

Um die Anwendung der Funktion möglichst einfach zu halten, sind vom Benutzer nur die ersten vier Argumente zu übergeben. Da die Generierung von Bootstrap-Stichproben mit zugehörigen Change-Point-Schätzungen bei einer großen Wiederholungszahl  $B$  eine nicht unerhebliche Rechenzeit in Anspruch nimmt, wird der Bootstrap-Algorithmus standardmäßig nicht ausgeführt und das Argument `boot` muss vom Benutzer explizit auf `TRUE` gesetzt werden.

#### 4.1.1 Funktion tau.est

Durch die Funktion `tau.est` wird der Change-Point mittels der Methode mit diskreten Intervallen geschätzt. `tau.est` bildet damit die Basis, auf welche andere Funktionen innerhalb des Paketes zugreifen. Der nachfolgende Code beinhaltet die für die Schätzung notwendigen Teile des Funktionsrumpfes.

```
tau.est <- function(time, event, intwd, taumax, taumin, cens, move=0,...){
...
lambda.est <- sum(event[time>taumax]==1) / sum(time[time>taumax] - taumax)
pr = pexp(intwd, lambda.est) #probability for event within an interval
if(cens==TRUE) tcens <- time[which(event == 0)] #times of censored data
lim <- seq(taumin+move, (taumax+intwd+move), by=intwd) #interval limits
pv <- vector("numeric", length=length(lim)-1) #vector for p-values

for(k in 1:length(pv)){
  x <- sum(event[time > lim[k] & time <= lim[k+1]])
  n <- sum(time > lim[k])

  if(cens==TRUE){
    #n is corrected by time of right-censored data within interval k
    n <- n - round(sum( (lim[k+1]
      - tcens[tcens > lim[k] & tcens <= lim[k+1]] ) / (lim[k+1] - lim[k])))
  }

  pv[k] <- binom.test(x, n, pr, alternative="greater")$p.value
}

S <- c(rev(cumsum(rev(pv-0.25))), 0)
tau <- lim[which.max(S)] #estimatet tau
if(tau > taumax) tau <- taumax
list(CP.est=tau, ...)
```

Im oberen Teil wird aus den Beobachtungen mit `time>taumax` zunächst der Parameter der Exponentialverteilung  $\hat{\lambda}$  (`lambda.est`) nach Formel (3) aus Kapitel 2.1 geschätzt, wobei angenommen wird, dass für diese Beobachtungen die Hazardrate konstant ist. Daraus ergibt sich mithilfe der Funktion `pexp()` die Wahrscheinlichkeit `pr` für ein Ereignis innerhalb eines Intervalls der Breite `intwd`. Weiterhin wird eine Variable für die Zensierungszeiten (`tcens`) sowie ein Vektor mit den Intervallgrenzen (`lim`) definiert. Der Befehl für die Erzeugung dieser Sequenz enthält einen Parameter `move`, welcher einen Wert für die Intervallverschiebung angibt. Da dies nur im Fall `shift=TRUE`, d.h. bei der Schätzme-

thode mit verschobenen Intervallen, eine Rolle spielt, ist `move` per default mit 0 vorbelegt und kann vom Anwender nicht verändert werden. Für die p-Werte wird ein “leerer“ Vektor `pv` definiert, dessen Länge der Anzahl der aus den Funktionsargumenten ermittelten Intervalle entspricht.

Innerhalb der for-Schleife werden nun für jedes Intervall zunächst die Anzahl der darin vorkommenden Ereignisse `x` sowie die Zahl der Personen unter Risiko zu Beginn des Intervalls `n` bestimmt. Für den Fall `cens=TRUE` wird die Variable `n`, entsprechend Formel (6) aus Kapitel 2.3, nach unten korrigiert. Aus diesen Werten wird für jedes Intervall mittels dem Befehl `binom.test()` ein exakter Binomialtest durchgeführt und der daraus resultierende p-Wert an der entsprechenden Stelle des Vektors `pv` gespeichert.

Wie in Kapitel 2.2 beschrieben, entspricht der Change-Point  $\tau$  jenem Wert, welcher die Summe aus (4), also  $S^*(\tau) = \sum_{k|a_{k-1} \geq \tau} (pv_k - 0.25)$ , maximiert.  $S^*$  lässt sich in *R* mittels des Ausdrucks `rev(cumsum(rev(pv-0.25)))` darstellen, woraus im letzten Teil der Funktion der Schätzer `tau` ermittelt wird.

Die if-Anweisung am Ende soll sicherstellen, dass die Funktion maximal das vorgegebene `taumax` als Schätzer zurückgibt, was der Annahme  $\tau \leq \tau_{max}$  entspricht.

#### 4.1.2 Funktion `int.shift`

Diese Funktion dient der Schätzung des Change-Points im Falle `shift=TRUE`, also für die Methode mit Intervallverschiebung. Wie in Kapitel 2.2 erläutert, lässt sich dieses Verfahren so vorstellen, dass die Intervalle in jedem Iterationsschritt um  $t=1$  verschoben und jeweils ein Schätzer für  $\tau$  ermittelt wird. Wie im nachfolgenden Auszug aus dem zugehörigen *R*-Code ersichtlich, besteht die Funktion im Kern aus einer for-Schleife, welche in jedem Durchlauf auf die Schätzfunktion `tau.est` zugreift. Die daraus resultierenden Change-Point-Schätzer und das Maximum der zugehörigen Summe `S` werden in einer Ergebnismatrix `res` gespeichert. Diejenige Spalte der Matrix, für welche der Eintrag `max.sum` maximal ist, enthält den Schätzer  $\hat{\tau}$ , welcher letztendlich durch die Funktion zurückgegeben wird.

```
int.shift <- function(time, event, intwd, taumax, taumin, cens){
  res <- matrix(NA, intwd, 2)  #results
  colnames(res) <- c("CPest", "max.sum")
  ...
  for(s in 0:(intwd-1)){
    est <- tau.est(time, event, intwd, taumax, taumin, cens, move=s)
    res[s+1,1] <- est$CP.est
    res[s+1,2] <- est$max.sum
    ...
  }
}
```

---

```
bestfit <- which.max(res[, "max.sum"])    #row with best fit
tau <- res[bestfit, "CPest"]
list(CP.est=tau, ...)
}
```

Innerhalb des *R*-Paketes wurde **CPest** als generische Funktion für S3-Klassen erstellt, für welche noch geeignete Methoden angelegt werden müssen. Da das Paket nur eine einzige Klasse **CP** für das Ergebnis der Change-Point-Schätzung enthält, ist eine default-Methode ausreichend, auf die beim Aufruf der Funktion **CPest** zugegriffen wird. Weiterhin soll für die Klasse **CP** eine eigene plot-Methode, sowie eine print- und eine summary-Methode für die Ausgabe der Ergebnisse definiert werden. Nachfolgend werden die einzelnen Funktionen mit Beispiel-Outputs genauer betrachtet.

#### 4.1.3 Methode **CPest.default**

Abhängig davon, welche Methode für die Schätzung verwendet werden soll, wird innerhalb von **CPest.default** zunächst die Funktion **tau.est** im Fall diskreter Intervalle oder **int.shift** bei Intervallverschiebung aufgerufen. In beiden Fällen wird eine Liste zurückgegeben, welche den geschätzten Change-Point, die zugehörigen p-Werte und die Intervallgrenzen enthält und als Variable **tau** gespeichert wird.

Im Fall **boot=TRUE** wird ein Bootstrap-Sampling mit  $B$  Wiederholungen durchgeführt. D.h. aus den übergebenen Vektoren **time** und **event** wird ein Datensatz erzeugt, aus welchem  $B \cdot n$  mal mit Zurücklegen gezogen wird, woraus wiederum  $B$  neue Datensätze der Länge  $n$  entstehen. Die Bootstrap-Samples werden durch folgenden Code generiert:

```
data <- data.frame(time, event)
idx <- sample(x=1:nrow(data), size=B * nrow(data), replace=T)
bdata <- data[idx,]
bdata$sample <- rep(1:B, each=nrow(data))
```

Anschließend wird für jeden dieser Datensätze der Change-Point geschätzt, wobei wiederum die vom Benutzer gewünschte Schätz-Methode verwendet wird. Die sich daraus ergebenden Schätzer für  $\tau$  werden im Vektor **b.tau** gespeichert, wie aus nachfolgendem *R*-Code ersichtlich ist.

```
b.tau <- vector("numeric", length=B)
if(shift){
  for(b in 1:B){
    b.tau[b] <- int.shift(bdata[bdata["sample"]==b,"time"],
                           bdata[bdata["sample"]==b,"event"],
                           intwd, taumax, taumin, cens)$CP.est
  }
}
```

```

else{
  for(b in 1:B){
    b.tau[b] <- tau.est(bdata[bdata["sample"]==b,"time"] ,
                           bdata[bdata["sample"]==b,"event"] ,
                           intwd, taumax, taumin, cens)$CP.est
  }
}

```

Aus den Ergebnissen des Bootstrappings wird mittels `sd()` ein Schätzer  $\hat{s}e$  für die Standardabweichung berechnet.

Weiterhin sollen zwei Konfidenzintervalle (KI) ausgegeben werden. Zum einen das übliche KI mit Normal-Approximation, was durch

$$[\hat{\tau} - z_{\alpha/2} \cdot \hat{s}e ; \hat{\tau} + z_{1-\alpha/2} \cdot \hat{s}e] \stackrel{\alpha=0.05}{=} [\hat{\tau} - 1.96 \cdot \hat{s}e ; \hat{\tau} + 1.96 \cdot \hat{s}e]$$

definiert ist. Zum anderen wird ein Bootstrap-Perzentil-Intervall ermittelt. Dazu werden die unter `b.tau` gespeicherten Werte der Größe nach geordnet, womit man eine “Bootstrap-Verteilung“ erhält, welche mit  $\hat{F}_B(\tau)$  bezeichnet sei. Das Perzentil-Intervall ergibt sich aus den geschätzten Werten an den Positionen  $B \cdot \frac{\alpha}{2}$  und  $B \cdot (1 - \frac{\alpha}{2})$  zu

$$[\hat{F}_B^{-1}(\alpha/2) ; \hat{F}_B^{-1}(1 - \alpha/2)]$$

was den 2,5%- und 97,5%-Quantilen dieser Verteilung entspricht (vgl. [Holling u. Gediga, 2013, S.277f]). Für die Intervalle kann der Funktion mit `alpha` ein gewünschtes  $\alpha$ -Niveau übergeben werden.

Abschließend wird dem Objekt `tau`, d.h. der Liste mit sämtlichen Ergebnissen aus der Schätzung, noch die Klasse `CP` zugewiesen, sodass für Objekte, welche mit `CPest` erzeugt wurden, auch eine `plot-` und eine `summary-`Funktion bereitgestellt werden kann.

#### 4.1.4 Methode `plot.CP`

Für die grafische Darstellung stehen drei mögliche Plots zur Verfügung, deren Ausgabe vom Benutzer durch entsprechende Argumente in der `plot`-Funktion gesteuert werden kann. Der erste Plot, welcher standardmäßig beim Aufruf von `plot(CP-Objekt)` ausgegeben wird, stellt die aus den Daten geschätzte Hazardrate dar. Dafür wird mithilfe der Funktion `muhaz()` aus dem Paket von [Hess u. Gentleman, 2010] eine geglättete Hazardfunktion geschätzt und diese mittels der zugehörigen `plot`-Funktion grafisch dargestellt. Zusätzlich wird in dem durch `plot(muhaz(times, delta))` erstellten Plot der mit `CPest` geschätzte Change-Point, der Schätzer  $\hat{\lambda}$  für die konstante Hazardrate sowie das Bootstrap-Perzentil-Intervall ergänzt.

Um die Darstellung der Plots beim Anwenden von `plot()` auf ein CP-Objekt demonstrieren zu können, wurde mithilfe der Funktion zur Datensimulation `sim.survdata` aus

Kapitel 3 ein Beispieldatensatz `simdat` erzeugt und auf diesen die Funktion `CPest` mit Intervallverschiebung angewendet. Der *R*-Code für die Simulation sowie der Aufruf der `plot`-Funktion für das resultierende Objekt der Klasse `CP` ist nachfolgend dargestellt.

```
simdat <- sim.survdata(n=1000,tau=100,shape=0.2,scale=100,cens=T,jump=T)

cp_shift <- CPest(simdat$time, simdat$event, taumax=360, intwd=10,
                    shift=T, boot=T, B=1000)

plot(cp_shift, plot.pval=T, plot.boot=T)
```

Die Abbildungen 5 bis 7 zeigen die zugehörigen Grafiken, welche bei Aufruf des Programmcodes ausgegeben wurden.

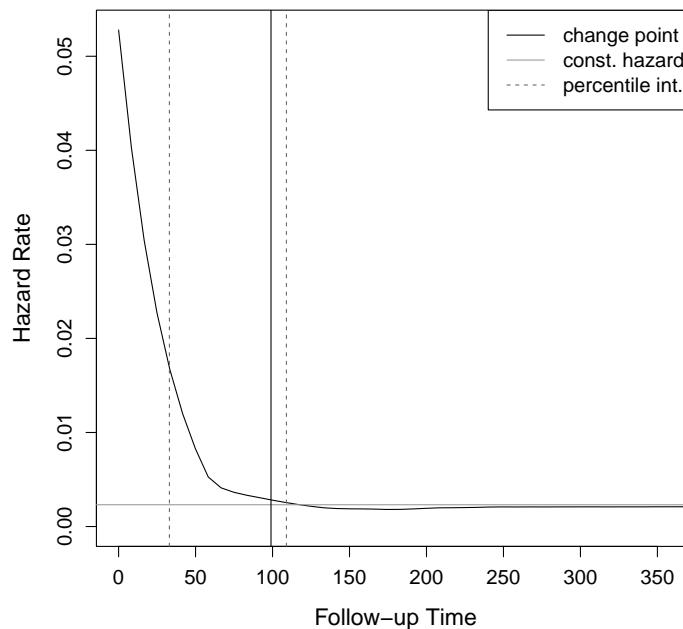


Abbildung 5: Für einen simulierten Datensatz mittels `plot.CP` erstellter Plot der geglätteten Hazardfunktion, welche mithilfe der `muhaz`-Funktion geschätzt wurde. Die Grafik enthält zusätzlich die Schätzungen des Change-Points und der konstanten Hazardrate sowie das aus den Bootstrap-Samples ermittelte Perzentil-Intervall.

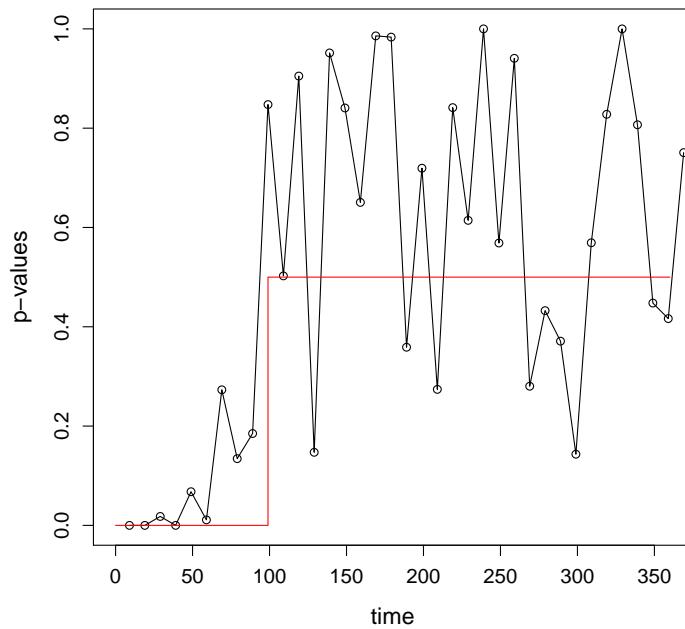


Abbildung 6: Mit `plot.CP` erstellter Plot der p-Werte aus einer Schätzung mit diskreten Intervallen der Länge 10 (mit Intervallverschiebung) für einen mit  $\tau = 100$  simulierten Datensatz mit Sprung in der Hazardrate, welcher zensierte Beobachtungen enthält.

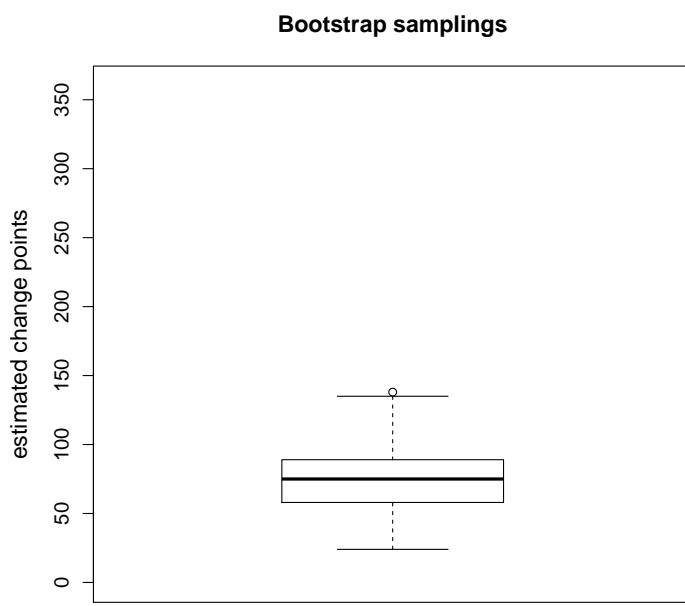


Abbildung 7: Mit `plot.CP` erstellter Boxplot der geschätzten Werte für  $\tau$  aus einem Bootstrapping mit 1000 Wiederholungen für einen simulierten Datensatz mit  $\tau = 100$ .

#### 4.1.5 print- und summary-Methoden

Weil diese Methoden nur dazu dienen, die Ergebnisse der Schätzung in einer ansprechenden Form auszugeben, soll an dieser Stelle auf Details zum zugehörigen *R*-Code verzichtet werden. Auf die print-Methode wird bei Aufruf eines Objektes der Klasse CP zugegriffen, wobei unabhängig davon, ob eine Varianzschätzung per Bootstrap-Verfahren durchgeführt wurde, nur der geschätzte Change-Point ausgegeben wird. Durch Aufruf der summary-Funktion für ein CP-Objekt, enthält der Output hingegen zusätzlich die geschätzte Varianz, sowie die beiden Konfidenzintervalle (vgl. Kap. 4.1.3), sofern diese berechnet wurden.

Nachfolgend sind die Outputs der print- und summary-Methoden für den in Kapitel 4.1.4 erstellten Beispieldatensatz dargestellt.

```
cp_shift

##
## Call:
## CPest.default(time = simdat$time, event = simdat$event, intwd = 10,
##     taumax = 360, shift = T, boot = T, B = 1000)
##
## 
## estimated change point:  99

summary(cp_shift)

##
## Call:
## CPest.default(time = simdat$time, event = simdat$event, intwd = 10,
##     taumax = 360, shift = T, boot = T, B = 1000)
##
## 
## change point:
## Estimate Std.Error
##      99     18.975
##
## confidence intervals (alpha=0.05):
## percentile interval: [33, 109]
## normal approximation: [61.809, 136.191]
```

## 5 Anwendungsbeispiele

In diesem Kapitel soll das in *R* geschriebene Paket und somit das Verfahren zur Change-Point-Schätzung nach [Küchenhoff u. a., 2015] auf zwei Datensätze angewandt werden, welche empirisch erhobene Überlebenszeiten von Intensivpatienten enthalten. Die Daten stammen in beiden Fällen aus Studien, welche in Zusammenarbeit mit dem Klinikum der Universität München durchgeführt wurden. Ein gemeinsames Ziel beider Forschungsarbeiten war das Finden eines Wendepunkts, ab welchem die Sterberate in Bezug auf einen chirurgischen Eingriff (Datensatz 1) bzw. nach Einlieferung auf eine Intensivstation (Datensatz 2) konstant bleibt. Wie in [Küchenhoff u. a., 2015] beschrieben, lässt sich der Beobachtungszeitraum dabei in verschiedene Phasen einteilen, welche sich hinsichtlich des Sterberisikos unterscheiden. Demnach befinden sich die Patienten anfänglich in einer akuten Phase, mit sehr hohem Sterberisiko, welches in der darauffolgenden post-akuten Phase allmählich sinkt. Der sogenannte Change-Point trennt nun diese zwei Zeiträume von der Langzeitmortalität, in der die Sterberate als konstant angenommen wird. Durch eine genaue Definition dieses Wendepunkts anhand von empirisch erhobenen Daten soll eine bessere Analyse der Risikofaktoren, getrennt nach den genannten Phasen, möglich sein und damit sowohl die Sterberate für die akute Phase als auch die Langzeitmortalität gesenkt werden. Im Folgenden werden die aus diesen Studien resultierenden Ergebnisse mit denen verglichen, welche sich aus dem in dieser Arbeit vorgestellten Verfahren ergeben.

### 5.1 Datensatz 1 - Überlebenszeiten nach Leberresektion

#### Details zur Studie

Dieser Datensatz ist in Zusammenarbeit zwischen dem Klinikum der Universität München und dem Institut für Statistik der Ludwig-Maximilians-Universität entstanden und war Gegenstand einer Studie, welche die Dauer der postoperativen Phase nach partieller Hepatektomie (Leberresektion) untersucht (vgl. [Schiergens u. a., 2015]). Die Forschungsarbeit fand vor dem Hintergrund statt, dass die Sterberate nach einem solchen operativen Eingriff in den letzten Jahrzehnten signifikant gestiegen ist. Als Ursache werden Fortschritte in der Diagnose, verbesserte Operationstechniken und eine bessere perioperative Betreuung angeführt, was zur Folge hat, dass immer mehr Patienten für eine Resektion von Tumoren und Metastasen in Frage kommen. Somit erhöht sich auch die Rate der durch einen solchen chirurgischen Eingriff verursachten Todesfälle. Ein Ziel der Studie ist es, durch bessere Schätzung der Prädiktoren, die Patienten in Risikogruppen einzuteilen und somit eine bessere Selektion zu ermöglichen, was wiederum die postoperative Sterberate verringern soll.

Hierfür wurden innerhalb einer prospektiven Datenerhebung zwischen 2003 und 2013 Patienten erfasst, bei denen eine Resektion von bösartigem Lebergewebe durchgeführt wurde. Nach Bereinigung um Patienten mit unvollständigem Erhebungsbogen ergab sich

daraus ein Datensatz mit 784 Beobachtungen.

Ein Problem, welches in dieser Studie untersucht wurde, ist nach [Schiergens u. a., 2015] die Definition der postoperativen Mortalität. Demnach werden dieser Sterberate für gewöhnlich nur diejenigen Tode zugerechnet, welche innerhalb von 30 Tagen nach einer Operation bzw. während des Krankenhausaufenthalts stattfinden (30-Tage und “in Hospital“ Mortalität). Durch medizinische Fortschritte könnte der durch Komplikationen nach einer Operation verursachte Todeszeitpunkt inzwischen jedoch solange hinausgezögert werden, dass dieser erst mehr als 30 Tage nach dem Eingriff bzw. nach Verlassen des Krankenhauses eintritt. Somit werden diese Fälle der Langzeitüberlebensrate zugerechnet und die postoperative Mortalität im Allgemeinen unterschätzt. Um die Risikofaktoren richtig analysieren und somit die Sterberate nach einem chirurgischen Eingriff verringern zu können, ist es nach [Schiergens u. a., 2015] allerdings wichtig, die akute postoperative Phase (auch APP oder “acute postoperative period“) anhand der erhobenen Daten genau zu definieren. Das vorrangige Ziel der Studie von [Schiergens u. a., 2015] ist deshalb das Schätzen der APP nach dem operativen Entfernen von bösartigen Leber-Tumoren. Dafür soll ein Zeitpunkt  $\tau$  gefunden werden, welcher die akute von der postakuten Phase nach einem chirurgischen Eingriff trennt, wobei für letztere eine konstante Hazardrate angenommen wird. Durch die individuelle Schätzung der APP für Patienten mit Leberresektion soll es nach [Schiergens u. a., 2015] möglich sein, Risikofaktoren zu identifizieren, wodurch eine bessere Selektion der Patienten und damit eine Verringerung der postoperativen Sterberate ermöglicht würde.

Die Ergebnisse dieser Studie sind in [Schiergens u. a., 2015] ausführlich dargelegt, insbesondere die ermittelte Dauer der APP und die sich aus der darauf aufbauenden Regression ergebenden Risikofaktoren für postoperative Mortalität. Für diese Arbeit ist jedoch hauptsächlich die Schätzung des Change-Points von Interesse, welcher in [Schiergens u. a., 2015] mit “transition point  $t$ “ bezeichnet ist. Der Zeitpunkt, welcher die akute von der postakuten Phase nach einer Operation trennt, wird dabei ähnlich dem Vorgehen nach [Küchenhoff u. a., 2015] bestimmt, welches im Kapitel 2 beschrieben ist. Dabei wurde die Sterberate ab einem Jahr nach einem chirurgischen Eingriff als konstant angenommen und aus den entsprechenden Beobachtungen des vorliegenden Datensatzes die konstante Hazardrate  $h_1$  geschätzt. Der Zeitraum zwischen 20 und 360 Tagen nach einer OP wurde in Intervalle einer Länge von 20 Tagen eingeteilt und mittels der sich aus  $h_1$  ergebenden Wahrscheinlichkeit, innerhalb von 20 Tagen zu sterben, ein Binomialtest durchgeführt. Unter der Annahme, dass die Hazardrate innerhalb der akuten Phase signifikant höher als  $h_1$  ist und sich nach dem Change-Point nicht signifikant von  $h_1$  unterscheidet, wurde aus den resultierenden p-Werten eine Regression durchgeführt. Daraus ergab sich letztendlich, dass die APP nach einer Resektion 80 Tage lang andauert und erst danach die postakute Phase eintritt. Inwiefern dies von den Ergebnissen abweicht, welche aus dem

in dieser Arbeit vorgestellten Verfahren nach [Küchenhoff u. a., 2015] resultieren, wird im Folgenden diskutiert.

## Ergebnisse

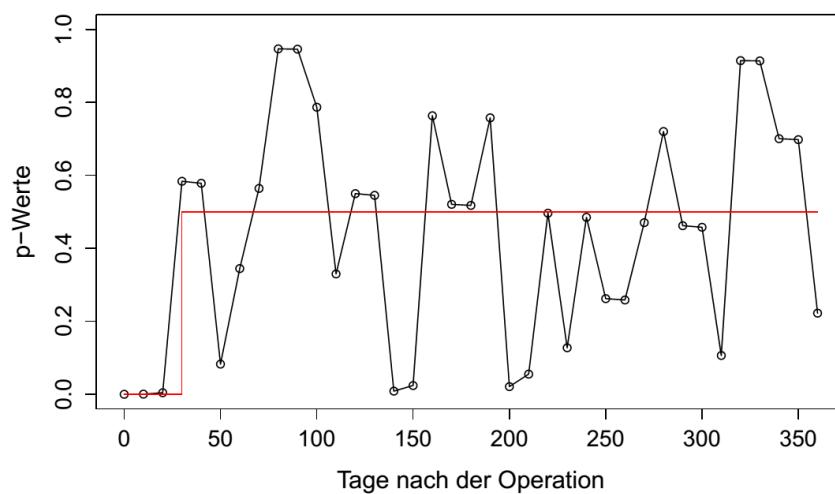
Um aus den vorliegenden Daten die Dauer der akuten postoperativen Phase nach Leberresektion, d.h. insbesondere den zugehörigen Change-Point schätzen zu können, wird aus den erhobenen Daten ausschließlich die jeweilige Beobachtungsdauer eines Patienten sowie der zugehörige Indikator für das Eintreten des Todes benötigt. Die entsprechenden Variablen wurden aus dem Datensatz extrahiert und die nach dem Verfahren von [Küchenhoff u. a., 2015] in *R* programmierte Funktion darauf angewandt. Für die Schätzung der in Tabelle 6 präsentierten Ergebnisse wurden die drei in Kapitel 2.2 vorgestellten Methoden mit einer Intervallbreite von jeweils 10 und 20 Tagen verwendet. Dabei wurde angenommen, dass sich der Change-Point innerhalb des ersten Jahres nach Resektion befindet, d.h. dass die Sterberate ab Tag 360 nach dem Eingriff als konstant angenommen wird. Diese konstante Hazardrate wurde gemäß Formel (2) in Kapitel 1.2 aus der Anzahl der Patienten geschätzt, bei denen der Todeszeitpunkt erst nach mehr als einem Jahr nach der OP beobachtet wurde. Für jedes der sechs Ergebnisse wurde zusätzlich ein nicht-parametrisches Bootstrapping mit 1000 Wiederholungen durchgeführt, woraus jeweils eine Schätzung für den Standardfehler sowie ein Perzentil-Intervall mit  $\alpha = 5\%$  ermittelt wurde, welche ebenfalls in Tabelle 6 zu finden sind.

Tabelle 6: Change-Point-Schätzung für den in [Schiergens u. a., 2015] verwendeten Datensatz mittels der in Kapitel 2.2 vorgestellten Verfahren mit jeweiliger Intervallbreite von 10 und 20. Der Wert für  $\tau_{max}$  wurde mit 360 vorgegeben. Die Standardfehler und Perzentil-Intervalle ergeben sich aus einem nichtparametrischen Bootstrapping. *dis.int* und *shift.int* bezeichnen die Ergebnisse für diskrete Intervalle ohne und mit Verschiebung, die Zeilen für *cont.int* beinhalten die Schätzungen für fortlaufende Intervalle.

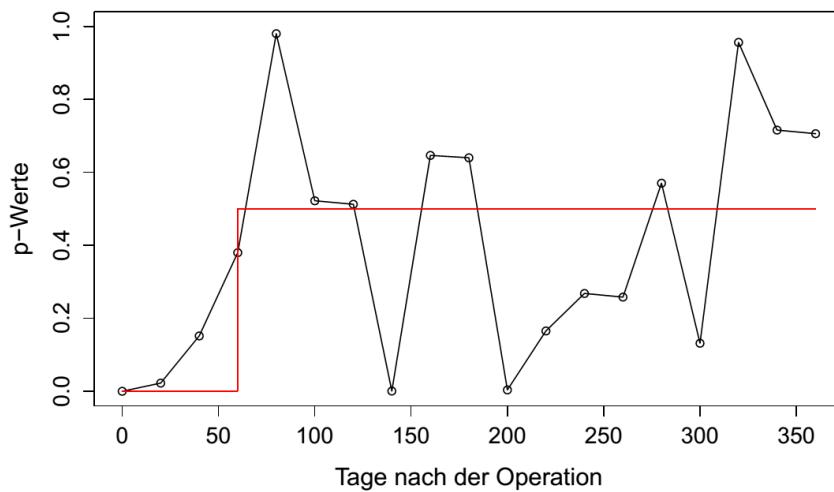
	estimate	std.Error	percentile interval
dis.int10	30	12.055	[20; 70]
dis.int20	60	32.317	[20; 80]
cont.int10	28	14.594	[12; 67]
cont.int20	60	31.900	[13; 71]
shift.int10	35	15.766	[14; 74]
shift.int20	63	23.599	[20; 79]

Bei Betrachtung der Ergebnisse ist auffällig, dass sich die Bereiche, in denen die Change-Point-Schätzungen liegen, für die gewählten Intervallbreiten stark unterscheiden. So dauert die akute Phase bei einer Intervallbreite von 20 Tagen zwischen 60 und 63 Ta-

ge an, bei einer Einteilung in 10-Tages-Intervalle jedoch höchstens bis Tag 35 nach dem operativen Eingriff. Anhand der sich aus dem Binomialtest ergebenden p-Werte, welche in Abbildung 8 für die zwei vorgegebenen Intervallbreiten grafisch dargestellt sind, lässt sich nachvollziehen, wie es zu diesen Unterschieden in der Schätzung kommt. Bei einer Intervallbreite von 10 Tagen sind die p-Werte bereits für die Intervalle  $(30;40]$  und  $(40;50]$  nicht mehr signifikant, d.h. innerhalb dieser Zeiträume wurden so wenige Ereignisse beobachtet, dass die Abweichung der Sterberate von der unter  $H_0$  geltenden konstanten Hazardrate nicht signifikant ist. Wird der gleiche Zeitraum jedoch in Intervalle von je 20 Tagen eingeteilt, ist die Anzahl der Ereignisse und damit die Sterberate erst im Intervall  $(60;80]$  so gering, dass die Nullhypothese des Binomialtests erfüllt ist (vgl. Kapitel 2.2 für Details zum Schätzverfahren). Die Plots der p-Werte für sämtliche Schätzungen aus Tabelle 6 sind im Anhang B.1 zu finden, welcher zusätzlich für jedes Ergebnis einen Boxplot der zugehörigen Bootstrap-Schätzungen enthält.



(a) Schätzung mit einer Intervallbreite von 10 Tagen



(b) Schätzung mit einer Intervallbreite von 20 Tagen

Abbildung 8: Aus der Schätzung mit diskreten Intervallen ohne Verschiebung ermittelte p-Werte.  $\tau_{max}$  wurde mit 360 Tagen vorgegeben.

Ergänzend zur Darstellung der p-Werte soll die Verteilung der Todesfälle innerhalb der Intervalle mithilfe des Histogramms in Abbildung 9 veranschaulicht werden. Daraus ist ersichtlich, dass im Zeitraum zwischen 30 und 40 Tagen nur drei Ereignisse stattgefunden haben, wodurch sich aus dem Binomialtest für dieses Intervall ein entsprechend hoher p-Wert ergibt. Vor dem Hintergrund, dass der zugehörige Datensatz insgesamt 784 Beobachtungen enthält, ist aus dieser Darstellung der absoluten Häufigkeiten außerdem zu erkennen, dass die Sterberate selbst in den ersten 30 Tagen relativ gering ist, was auch aus den bereits veröffentlichten Ergebnissen in [Schiergens u.a., 2015] hervorgeht. Dies hat den Nachteil, dass die Intervalle mit einer Breite von 10 Tagen im Allgemeinen nur wenige Ereignisse enthalten, wodurch die Schätzung des Change-Points bereits von einzelnen Beobachtungen beeinflusst wird. Deshalb werden im Folgenden nur noch die Ergebnisse für eine Intervallbreite von 20 Tagen betrachtet.

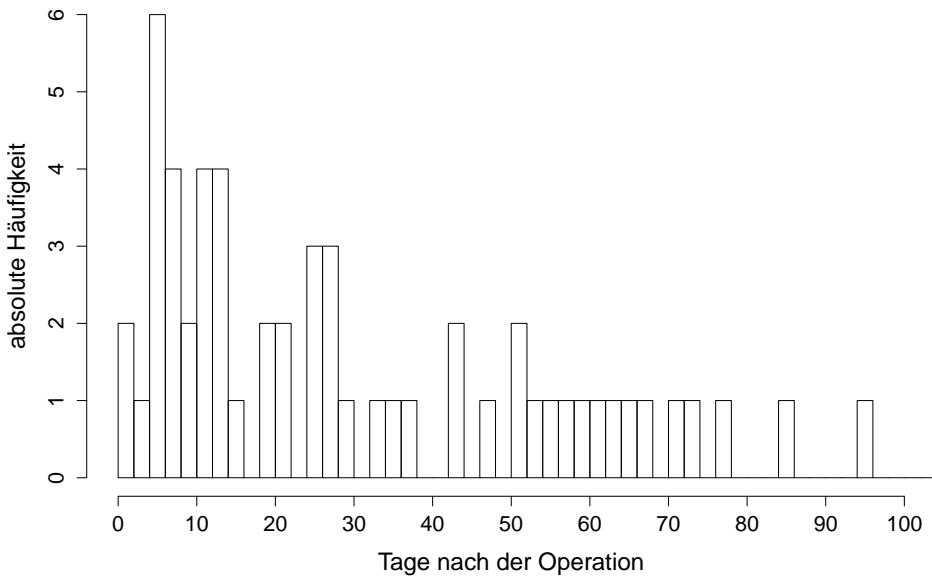


Abbildung 9: Verteilung der absoluten Häufigkeiten der beobachteten Todesfälle nach Leberresektion für den Datensatz von [Schiergens u.a., 2015]

Aus den Ergebnissen in Tabelle 6 ergibt sich der geschätzte Change-Point also zu 60 bzw. 63, was davon abhängig ist, welches Verfahren angewandt wurde. Bei der Auswertung der Simulationsergebnisse in Kapitel 3 und dem dabei angestellten Vergleich der drei Methoden zur Schätzung hinsichtlich ihrer Güte ging hervor, dass das Schätzverfahren mit diskreten Intervallen (mit oder ohne Intervallverschiebung) gegenüber der Methode mit fortlaufenden Intervallen (*cont.int*) zu bevorzugen ist. Weil hier der geschätzte Change-Point für *dis.int* und *cont.int* gleich ist, spielt dies jedoch für die Ergebnisse zu diesem Datensatz nur bezüglich der zugehörigen Standardabweichung bzw. des Perzentil-Intervalls eine Rolle. Da im Allgemeinen nicht davon ausgegangen werden kann, dass der Change-Point auf einer der Intervallgrenzen  $\{20, 40, 60, \dots\}$  liegt, ist es sinnvoll, die Schätzung für *shift.int20*, d.h. für das Verfahren mit Intervallverschiebung, als endgültiges

Ergebnis zu betrachten. Nach dem in dieser Arbeit vorgestellten Verfahren nach [Küchenhoff u. a., 2015] beträgt die geschätzte Dauer der akuten postoperativen Phase nach einer Leberresektion also 63 Tage.

Dieses Ergebnis deckt sich jedoch nicht mit der in [Schiergens u. a., 2015] angegebenen Schätzung des Change-Points. Für diese bereits veröffentlichte Studie wurde die Dauer der akuten Phase nach einem chirurgischen Eingriff auf 80 Tage geschätzt. Die Ursache für diese Abweichung liegt in diversen Simulationsstudien, welche im Zuge dieser Forschungsarbeit bereits durchgeführt wurden. Dabei wurden nur Datensätze mit stetigem Übergang in der Hazardrate zwischen der akuten und der postakuten Phase untersucht, wobei sich, ähnlich zu den im Kapitel 3 vorgestellten Simulationsergebnissen, eine Unterschätzung des tatsächlichen Change-Points um ca. 20 Tage ergab. Daraufhin wurde der Schätzer als Obergrenze desjenigen Intervalls definiert, welches nach Regression über die p-Werte den Change-Point enthält. Nach dem Verfahren, worauf diese Arbeit aufbaut, entspricht der Schätzer jedoch der Untergrenze dieses Intervalls. Dies ist aus theoretischer Sicht richtig und wurde auch durch die Simulationsergebnisse für die mit Sprung in der Hazardrate generierten Daten bestätigt. Ist der Übergang zur konstanten Sterberate jedoch stetig, wie auch im vorliegenden Datensatz mit den Überlebenszeiten nach Leberresektion, hat sich aus den Simulationsergebnissen eine Abweichung zum tatsächlichen Change-Point nach unten ergeben. Auch aus der Darstellung der geschätzten glatten Hazardfunktion in Abbildung 10, welche mithilfe der Funktion *muhaz* aus dem Paket von [Hess u. Gentleman, 2010] erstellt wurde, ist erkennbar, dass eine Dauer der akuten Phase von 63 Tagen nicht zu den beobachteten Daten passt. In der Grafik ist zu sehen, dass die Sterberate auch oberhalb von 63 Tagen nach einer OP noch weiter abfällt und es erst nach etwa 80 Tagen einen Wendepunkt gibt, ab welchem die Hazardfunktion annähernd konstant verläuft.

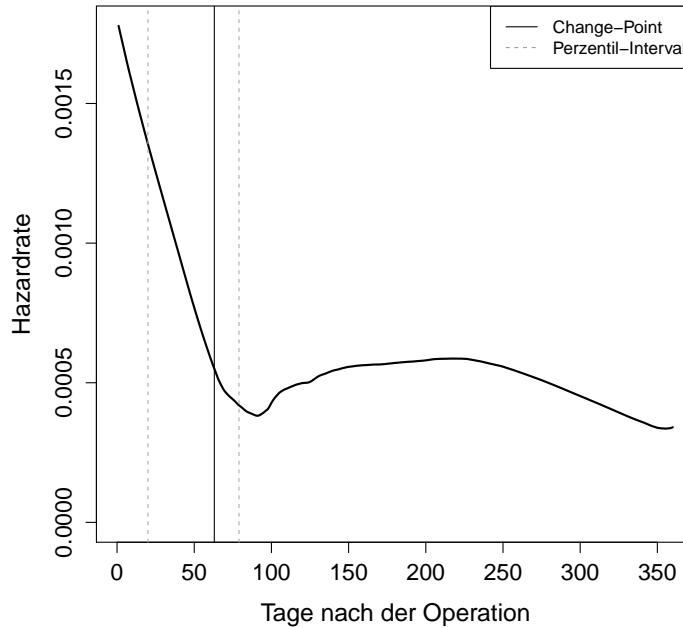


Abbildung 10: Geschätzte Hazardrate pro Tag (geglättet) für die ersten 360 Tage nach Durchführung einer Leberresektion. Der dargestellte Change-Point bei Tag 63 wurde nach dem Verfahren von [Küchenhoff u. a., 2015] mit diskreten Intervallen einer Breite von 20 und Intervallverschiebung sowie  $\tau_{max} = 360$  geschätzt. Das Perzentil-Intervall wurde mittels Bootstrapping erstellt. Zugrunde liegt der unbereinigte Datensatz aus [Schiergens u. a., 2015].

## 5.2 Datensatz 2 - Langzeitüberleben von Intensivpatienten nach chirurgischem Eingriff

### Details zur Studie

Die Hintergründe zur Studie, Details zu den statistischen Methoden sowie die daraus resultierenden Ergebnisse sind in [Schneider u. a., 2010] ausführlich beschrieben. Demnach war ein Ziel dieser Forschungsarbeit, zu untersuchen, wie Fortschritte in der Akutversorgung nach einem chirurgischen Eingriff die Sterberate auf lange Sicht beeinflussen, insbesondere wie die Auswirkungen auf bestimmte Gruppen von Patienten sind, welche sich nach einer Operation in einem kritischen Zustand befinden. Neben dem Überstehen dieser akuten postoperativen Phase sei es außerdem wichtig, erneute Fehlfunktionen oder gar Ausfälle der Organe zu verhindern. Dafür seien spezielle Organ-Unterstützungs-Therapien (z.B. invasive Beatmung oder Nierenersatztherapie) notwendig, deren dauerhafter Einsatz zum einen sehr kostspielig ist und außerdem regelmäßige lebenserhaltende Maßnahmen erfordert. Deshalb würde oft diskutiert, ob dies auf lange Sicht einem effizienten Einsatz der verfügbaren Ressourcen eines Krankenhauses gerecht wird. Nach [Schneider u. a., 2010] gibt es bisher keine ausreichend informativen Studien über Langzeit-Effekte solcher spezieller Therapien. Für diese Forschungsarbeit sollten deshalb nur diejenigen Patienten

untersucht werden, welche die akute Phase nach einem chirurgischen Eingriff überstanden haben, um daraus Einflussfaktoren auf das Langzeitüberlegen ermitteln zu können.

Zunächst wurde dafür am Klinikum der Universität München in Großhadern eine prospektive Datenerhebung durchgeführt. Dafür wurden zwischen den Jahren 1993 und 2005 Patienten mit einem Aufenthalt auf der Intensivstation (ICU) von mehr als 4 Tagen beobachtet und sämtliche Informationen, welche für die statistischen Analysen der untersuchten Therapien notwendig sind, erfasst. Nach Ausschluss einiger Patientengruppen, u.a. jener ohne einen chirurgischen Eingriff, ergab sich ein Datensatz mit insgesamt 1462 Beobachtungen. Der Beobachtungszeitraum wurde auf 2 Jahre beschränkt, weil sich die daraus ergebenden Einflussfaktoren für die Langzeit-Sterberate nicht signifikant von denen unterscheiden, welche aus einer längeren Beobachtungszeit resultieren würden. Gegenstand der statistischen Auswertungen waren letztendlich nur jene Patienten, welche die akute kritische Phase nach einer Operation überlebt haben und somit erst für die Untersuchung der Langzeiteffekte einer bestimmten Therapie in Frage kamen.

Um diese Personen identifizieren zu können, war zunächst die Schätzung des Zeitpunktes notwendig, bis zu welchem die akute postoperative Phase anhält. Dieser Change-Point trennt also das akute vom postakuten Stadium nach einem chirurgischen Eingriff. Wie in [Schneider u. a., 2010] im Abschnitt “Statistical Analysis” beschrieben, basiert dessen Schätzung auf einem Likelihood-Quotienten-Test (LQ-Test). Für die Durchführung wurde für jeden Tag  $t$  zwischen 0 und 730 ein Weibull-Modell an jene Daten mit einer Überlebensdauer größer als  $t$  angepasst. Unter der Nullhypothese des LQ-Tests ist die Hazardrate und folglich der Shape-Parameter der Weibull-Verteilung konstant mit  $H_0 : \gamma = 1$ . Als Change-Point wurde der kleinste Zeitpunkt  $t$  definiert, für den die Nullhypothese bei einem Niveau von  $\alpha = 0.05$  abgelehnt wurde, d.h. für den der Test keine signifikante Abweichung von der Annahme einer konstanten Hazardrate ergab. Daraus ergab sich in der Studie eine Dauer der akuten Phase von 196 Tagen ab dem Zeitpunkt der Einlieferung auf die Intensivstation. Darauf aufbauend wurden für die Intensivpatienten mit einer Überlebenszeit von mehr als 196 Tagen weitere statistische Analysen durchgeführt, um die Langzeit-Wirkung von bestimmten medizinischen Maßnahmen zu untersuchen. Da für diese Arbeit jedoch die Change-Point-Schätzung anhand der vorliegenden Daten von vorrangigem Interesse ist, wird auf die übrigen Ergebnisse der Forschungsarbeit, welche ebenfalls in [Schneider u. a., 2010] ausführlich diskutiert werden, nicht mehr explizit eingegangen.

## Ergebnisse

Für die Ermittlung der akuten postoperativen Phase für Intensivpatienten nach dem Verfahren von [Küchenhoff u. a., 2015] wurden die Überlebenszeiten mit zugehörigem Zensierungsstatus aus dem unbereinigten Datensatz verwendet, welcher auch der Studie von [Schneider u. a., 2010] zugrunde lag. Die Schätzung wurde für die drei in Kapitel 2.2 vorgestellten Schätzmethoden mit disjunkten Intervallen (mit und ohne Intervallverschiebung) sowie mit fortlaufenden Intervallen durchgeführt. Für die Tage  $t > 540$  wurde eine kon-

stante Hazardrate angenommen, welche aus den beobachteten Überlebenszeiten größer als 540 geschätzt wurde. Die Einteilung des Zeitraumes zwischen 0 und 540 Tagen erfolgte in Intervalle der Breite 10 und 20, für welche gemäß dem in Kapitel 2 erläuterten Verfahren ein exakter Binomialtest durchgeführt und aus den resultierenden p-Werten jeweils ein Change-Point geschätzt wurde.

Die Ergebnisse der Schätzung für jedes der drei Methoden mit jeweils zwei Intervallbreiten sind in Tabelle 7 aufgeführt. Zusätzlich sind für jeden geschätzten Change-Point die zugehörige Standardabweichung sowie das Perzentil-Intervall mit  $\alpha = 5\%$  angegeben, welche aus der Durchführung eines nichtparametrischen Bootstrap-Verfahrens mit je 1000 Wiederholungen resultieren. Die zugehörigen Plots der p-Werte sowie die Boxplots der Schätzungen aus den Bootstrap-Samplings sind im Anhang B.1 zu finden.

Weil sich aus den Simulationsergebnissen in Kapitel 3 für eine Intervallbreite von 10 Tagen im Allgemeinen eine höhere Güte der Schätzung bezüglich des MAD ergibt, sollen die Ergebnisse für diese Fälle vorrangig betrachtet werden. Dabei spielt das für den ersten Datensatz (vgl. Kapitel 5.1) auftretende Problem, dass sich innerhalb eines Zeitraumes von 10 Tagen oft nur sehr wenige Ereignisse befinden, aufgrund der hohen Beobachtungszahl und der insgesamt wesentlich höheren Anzahl an Todesfällen in diesem Datensatz keine Rolle. Weiterhin resultiert aus den Simulationen, dass die beiden Methoden mit diskreten Intervallen in den meisten Fällen zu einem geringeren MAD führen, als das Verfahren mit fortlaufenden Intervallen (*cont.int*). Wird für die Schätzung mit diskreten Intervallen zusätzlich eine Intervallverschiebung durchgeführt, können sich als Wert für den Change-Point nicht nur Vielfache von 10, sondern jeder Tag zwischen 0 und 540 ergeben. Deshalb soll der geschätzte Change-Point für *shift.int10* in Tabelle 7 als endgültiges Ergebnis betrachtet werden. D.h. nach dem Verfahren von [Küchenhoff u. a., 2015], auf welchem diese Arbeit basiert, ergibt sich für die untersuchten Intensivpatienten nach einem chirurgischen Eingriff eine akute Phase von 348 Tagen.

Tabelle 7: Change-Point-Schätzung für den in [Schneider u. a., 2010] verwendeten Datensatz mittels der in Kapitel 2.2 vorgestellten Verfahren mit jeweiliger Intervallbreite von 10 und 20. Der Wert für  $\tau_{max}$  wurde mit 540 vorgegeben. Die Standardfehler und Perzentil-Intervalle ergeben sich aus einem nichtparametrischen Bootstrapping. *dis.int* und *shift.int* bezeichnen die Ergebnisse für diskrete Intervalle ohne und mit Verschiebung, die Zeilen für *cont.int* beinhalten die Schätzungen für fortlaufende Intervalle.

	estimate	std.Error	percentile interval
dis.int10	350	62.611	[180; 380]
dis.int20	380	56.463	[240; 460]
cont.int10	348	64.061	[177; 351]
cont.int20	343	49.568	[222; 452]
shift.int10	348	61.524	[177; 358]
shift.int20	386	53.292	[226; 452]

Eine grafische Darstellung der (geglätteten) geschätzten Hazardrate für die ersten zwei Jahre nach ICU-Einlieferung ist in Abbildung 11 zu finden. Der Plot wurde mit der R-Funktion *muhaz* (aus dem Paket von [Hess u. Gentleman, 2010]) erstellt und enthält zusätzlich den geschätzten Change-Point für *shift.int10* aus Tabelle 7 mit dem zugehörigen Percentil-Intervall für  $\alpha = 5\%$ .

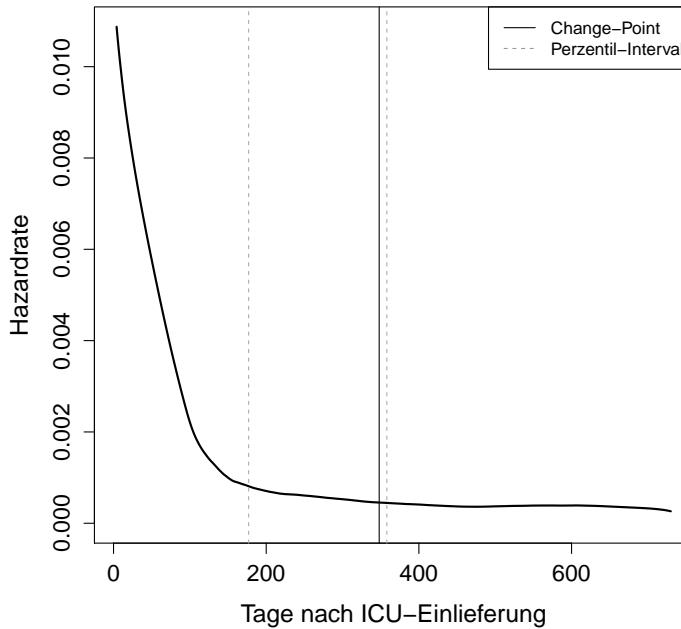


Abbildung 11: Geschätzte Hazardrate pro Tag (geglättet) für die ersten 730 Tage nach Einlieferung auf die Intensivstation. Der dargestellte Change-Point bei Tag 348 wurde nach dem Verfahren von [Küchenhoff u. a., 2015] mit diskreten Intervallen einer Breite von 10 und Intervallverschiebung sowie  $\tau_{max} = 540$  geschätzt. Das Perzentil-Intervall wurde mittels Bootstrapping erstellt. Zugrunde liegt der unbereinigte Datensatz aus [Hartl u. a., 2007].

Es ist zu beachten, dass der wahre Change-Point in den Simulationsergebnissen bei einer stetigen Hazardrate, d.h. ohne Sprung zum Zeitpunkt des Übergangs zur konstanten Sterberate, tendenziell unterschätzt wird. Weiterhin scheint in den Simulationen die Abweichung des Schätzers vom wahren Wert davon abzuhängen, an welchem Zeitpunkt  $t$  sich der vorgegebene tatsächliche Change-Point  $\tau$  befindet, insbesondere scheint diese Differenz größer zu werden, je später die Phase mit konstanter Hazardrate beginnt. Für das Ergebnis zum vorliegenden Datensatz bedeutet dies, dass die akute Phase nach ICU-Einlieferung tatsächlich noch länger als die geschätzten 348 Tage andauern würde.

Aus Abbildung 11 lässt sich erahnen, dass sich die geschätzte Hazardfunktion in drei Bereiche einteilen lässt, nämlich zu Beginn eine Phase mit stark abfallender Hazardrate, dann ein Zeitraum mit einem leicht abfallenden Verlauf, bis sie schließlich annähernd konstant bleibt. Um dies zu verdeutlichen, ist in Abbildung 12 ein vergrößerter Ausschnitt dieser Hazardfunktion dargestellt, wobei deutlich wird, dass erst im Bereich zwischen 450 und 500 Tagen ein Wendepunkt in der Hazardrate zu erkennen ist. Dies deckt sich auch

mit der Annahme, dass der geschätzte Change-Point bei 348 Tagen vom tatsächlichen Wert abweicht und die postakute Phase tatsächlich erst später beginnt.

Um eine präzisere Aussage über die geschätzte Dauer des akuten Stadiums nach ICU-Einlieferung anhand des betrachteten Datensatzes nach dem Verfahren von [Küchenhoff u. a., 2015] treffen zu können, sind noch weitere Untersuchungen bzgl. der beobachteten systematischen Unterschätzung des wahren Change-Points, ggf. mit weiterführenden Simulationen, notwendig.

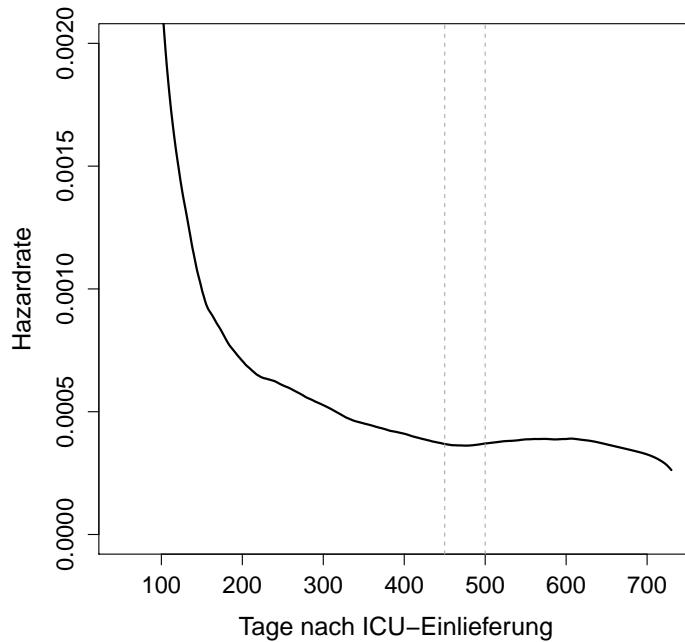


Abbildung 12: Vergrößerte Darstellung der geschätzten Hazardfunktion aus Abbildung 11 zwischen 50 und 730 Tagen mit Begrenzung der y-Achse auf 0.002. Der abfallende Verlauf der Hazardrate bis zu dem markierten Bereich zwischen 450 und 500 Tagen ist hier deutlicher zu erkennen.

## 6 Zusammenfassung und Ausblick

Ziel dieser Studie war es, das von [Küchenhoff u. a., 2015] entwickelte nichtparametrische Schätzverfahren hinsichtlich der Güte des daraus resultierenden Change-Point-Schätzers zu untersuchen. Die Ergebnisse der zu diesem Zweck durchgeführten Simulationsstudie haben gezeigt, dass der für die Datenerzeugung vorgegebene Change-Point tendenziell unterschätzt wird. Besonders deutlich sind die Abweichungen des Schätzers vom tatsächlichen Wert bei realistisch simulierten Daten mit einer stetigen Hazardfunktion. Wesentlich kleiner ist der beobachtete Bias dagegen für jene Daten, bei denen die zugrundeliegende theoretische Hazardfunktion einen Sprung aufweist, sodass der Übergang zur konstanten Sterberate besonders eindeutig ist. Weiterhin scheint aus den Simulationen hervorzugehen, dass die Verzerrung davon abhängig ist, an welchem Zeitpunkt sich der Change-Point befindet, insbesondere war die Abweichungen in den Schätzergebnissen größer, je später die Phase mit konstanter Hazardrate beginnt. Auch die Überdeckungsrate der in Kapitel 3.4.2 untersuchten Konfidenzintervalle war, insbesondere für die Szenarien ohne Sprung in der Hazardrate, teilweise sehr gering und damit nicht zufriedenstellend.

Aus dem Vergleich der drei in Kapitel 2.2 vorgestellten Methoden war ersichtlich, dass die Schätzungen mit diskreten Intervallen - mit und ohne Intervallverschiebung - gegenüber dem Verfahren mit fortlaufenden Intervallen bessere Ergebnisse hinsichtlich des MAD liefern, sodass nur die Funktionen zu diesen beiden Methoden für das im Zuge dieser Arbeit erstellte *R*-Paket **CPest** verwendet wurden.

Die Anwendung des Schätzverfahrens auf die beiden Datensätze mit Überlebenszeiten von Intensivpatienten hat ergeben, dass sich die Ergebnisse deutlich von denen aus den bereits veröffentlichten Studien zu diesen Datensätzen unterscheiden. Jedoch war bei der grafischen Auswertung der geschätzten Hazardfunktion dieser Daten ersichtlich, dass die Schätzergebnisse nach dem hier vorgestellten Verfahren nicht plausibel sind, was wiederum mit der systematischen Abweichung des geschätzten vom tatsächlichen Change-Point zu begründen ist.

Mit dem Ziel, dass das Schätzverfahren auch für empirisch erhobene Daten zufriedenstellende Ergebnisse liefert, sind weitere Untersuchungen, ggf. mit zusätzlichen Simulationen, notwendig. Auch hinsichtlich des erstellten Programmpaketes sind noch Verbesserungen möglich, insbesondere bezüglich der Laufzeit der implementierten Funktionen, was auch für die Durchführung weiterer Simulationen von Vorteil wäre.

Abschließend kann festgehalten werden, dass das untersuchte Schätzverfahren theoretisch, d.h. für einen hinsichtlich des Change-Points eindeutigen Datensatz, funktioniert, jedoch in einem praxisnahen Fall mit stetigem Übergang zur konstanten Hazardrate noch keine ausreichend zufriedenstellenden Ergebnisse liefert, weshalb dieses Verfahren Inhalt weiterer Forschungsarbeit ist.

# Anhang

## A Plots zur Simulationsstudie

### A.1 Theoretische Hazardfunktion

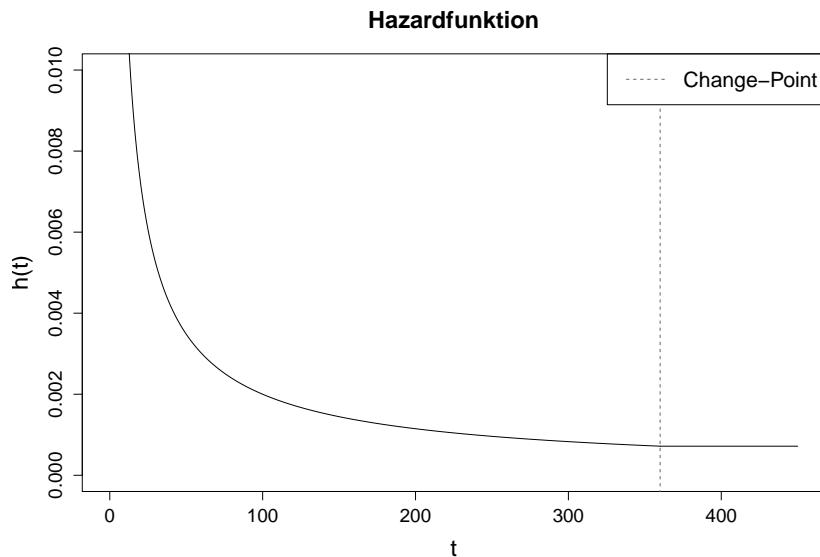
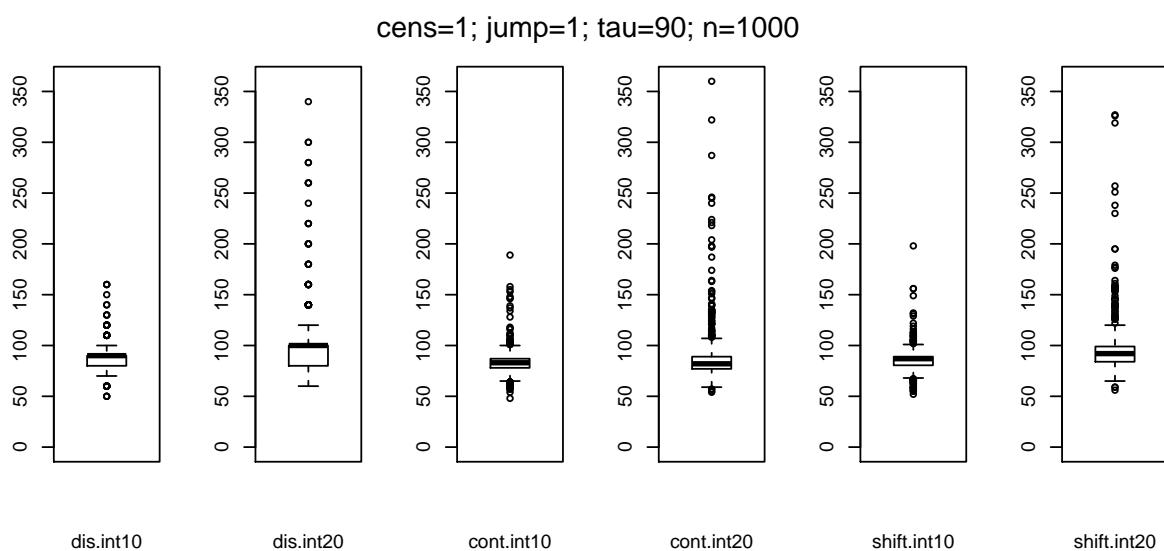


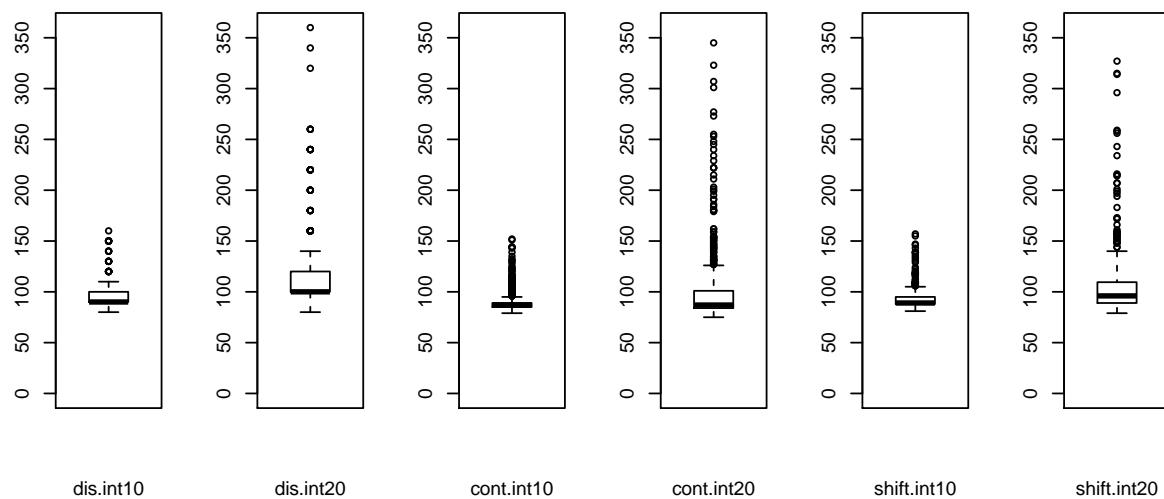
Abbildung 13: theoretische Hazardfunktion für  $\tau = 360$ ;  
für die  $t_i$  gilt:  $(t_i|t_i < 360) \sim WB(\eta = 100, \gamma = 0, 2)$  und  $(t_i|t_i \geq 360) \sim Exp(\lambda)$

### A.2 Boxplots zu simulierten Szenarien

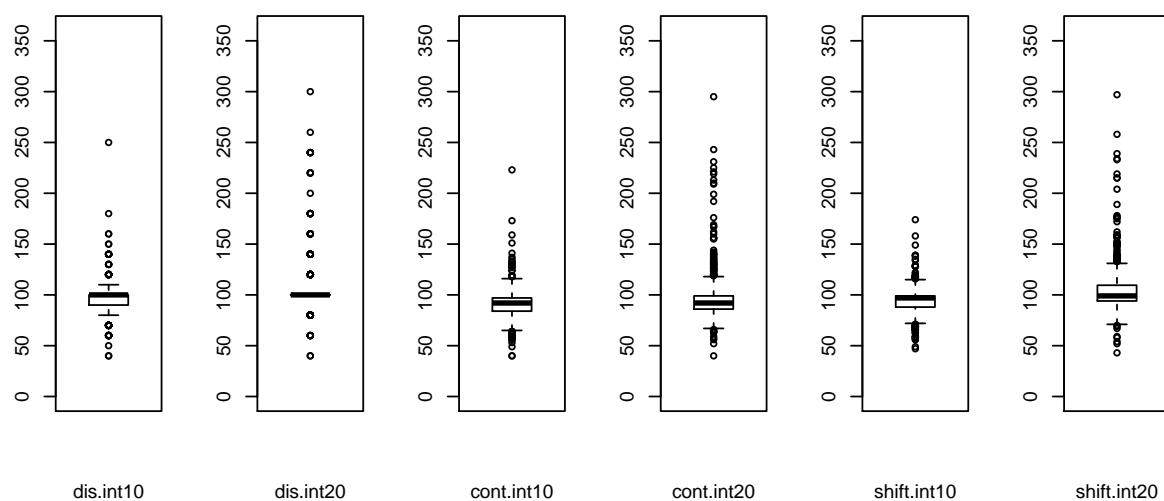
Boxplots der Schätzungen für Szenario 1 mit  $\tau = \{90, 100\}$ , cens= $\{1,0\}$ , jump= $\{1,0\}$ , n= $\{1000, 5000\}$  und  $\tau_{max} = 360$ :



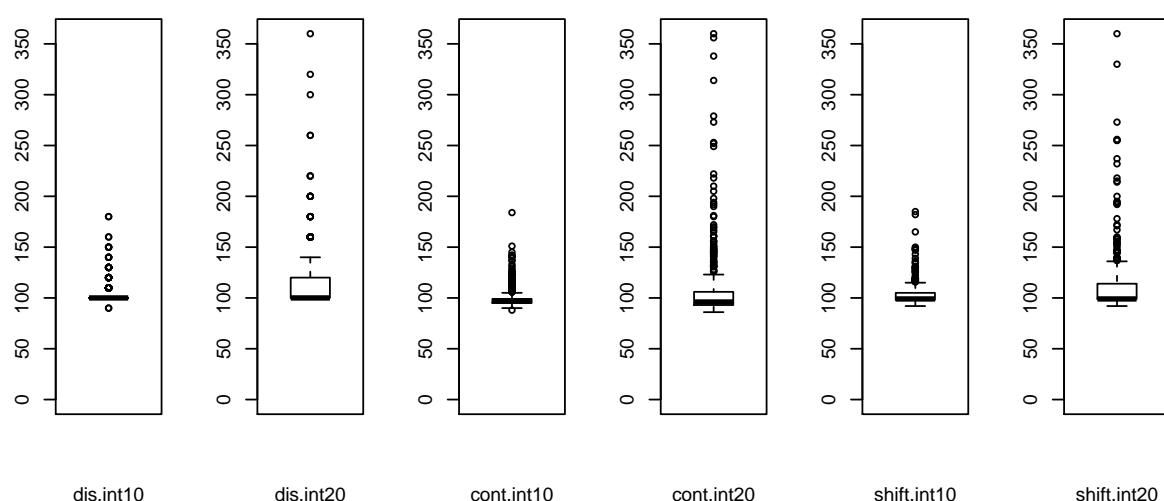
cens=1; jump=1; tau=90; n=5000



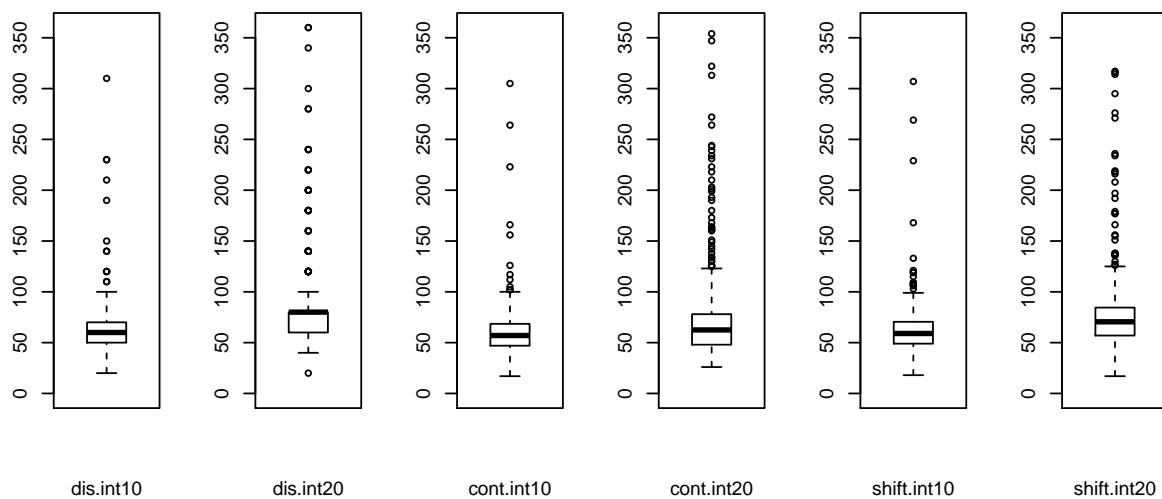
cens=1; jump=1; tau=100; n=1000



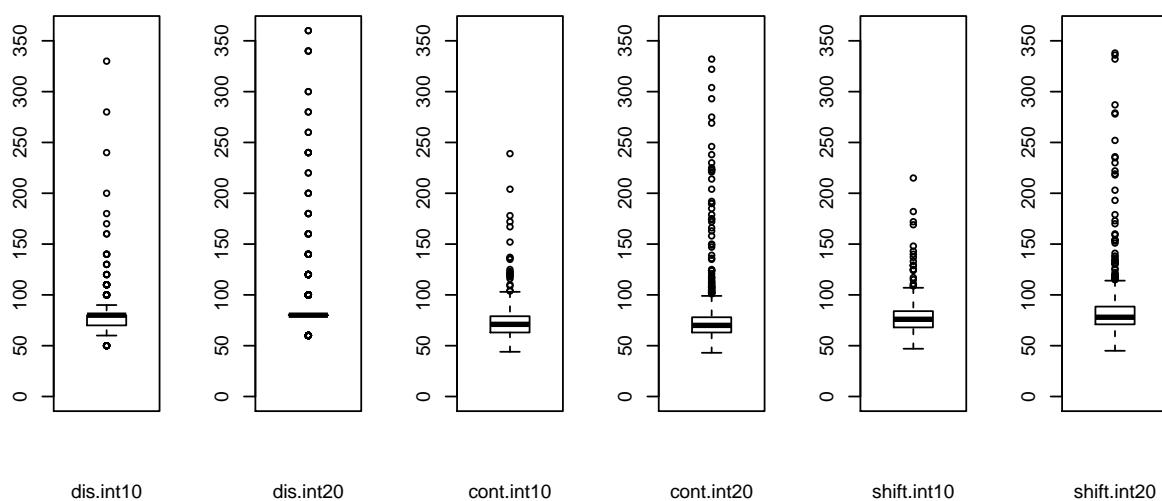
cens=1; jump=1; tau=100; n=5000



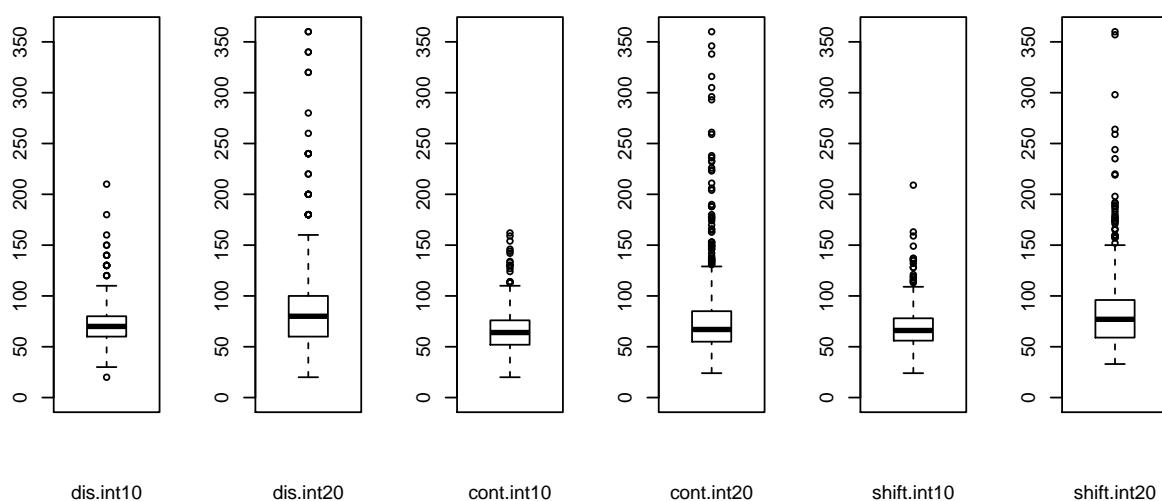
cens=1; jump=0; tau=90; n=1000



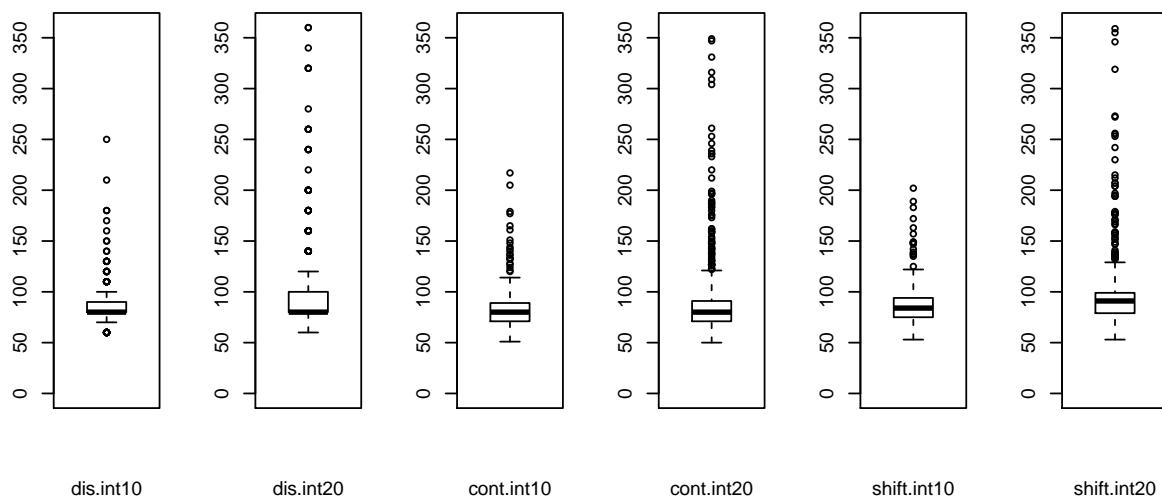
cens=1; jump=0; tau=90; n=5000



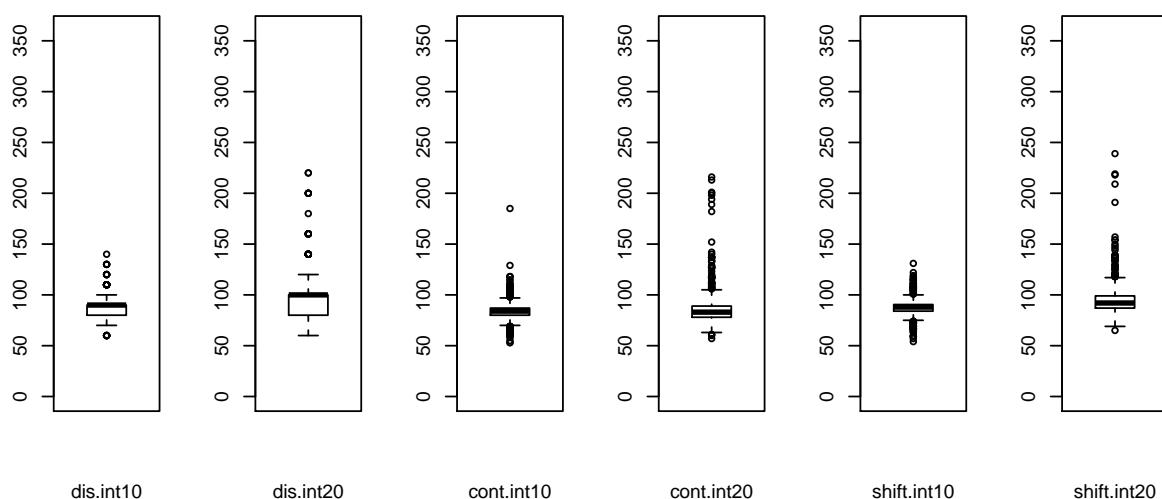
cens=1; jump=0; tau=100; n=1000



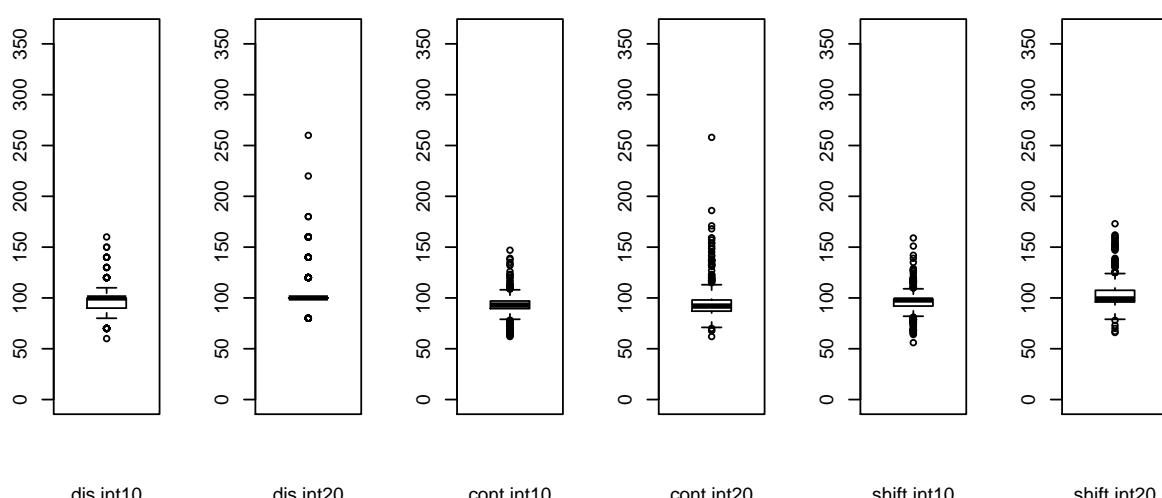
cens=1; jump=0; tau=100; n=5000



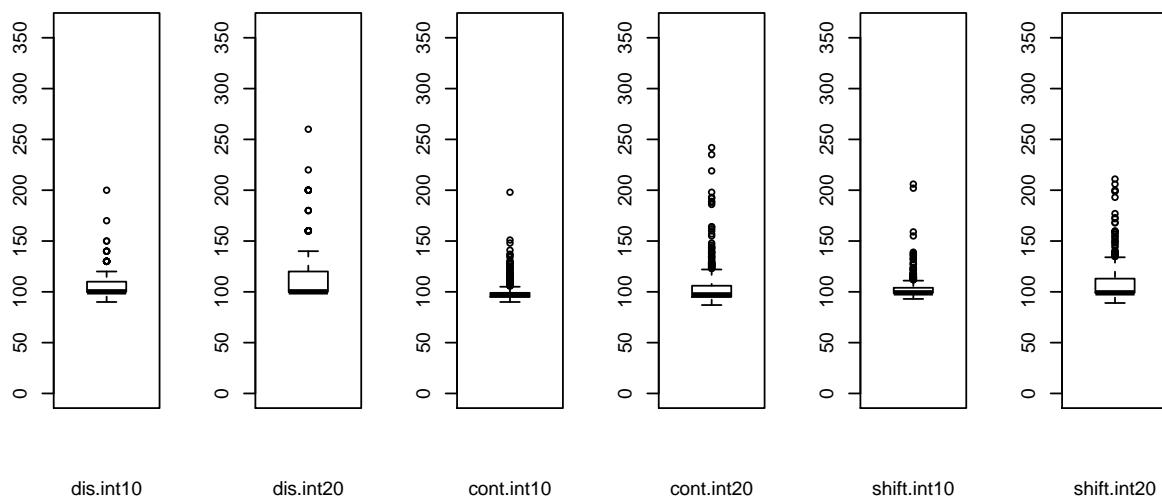
cens=0; jump=1; tau=90; n=1000



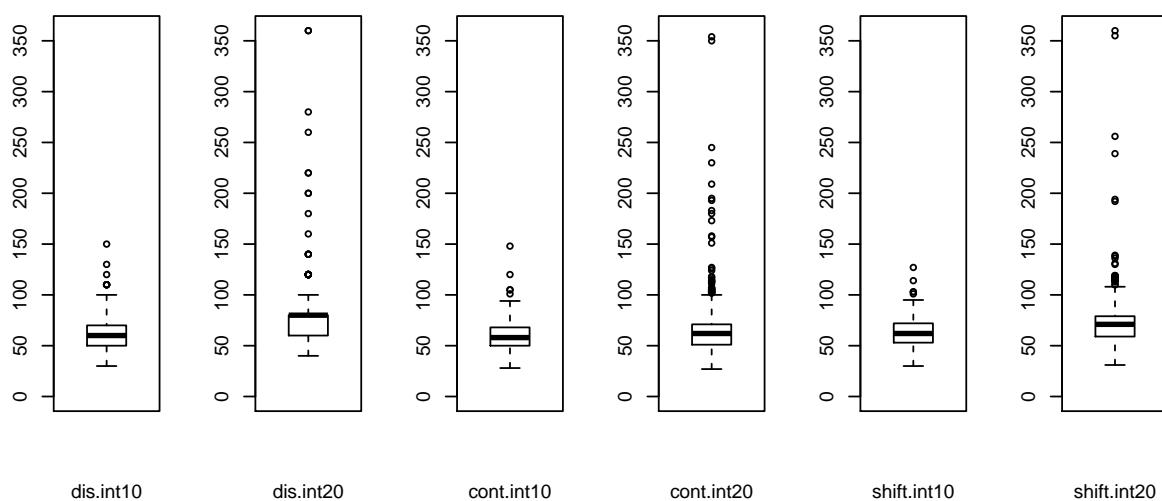
cens=0; jump=1; tau=100; n=1000



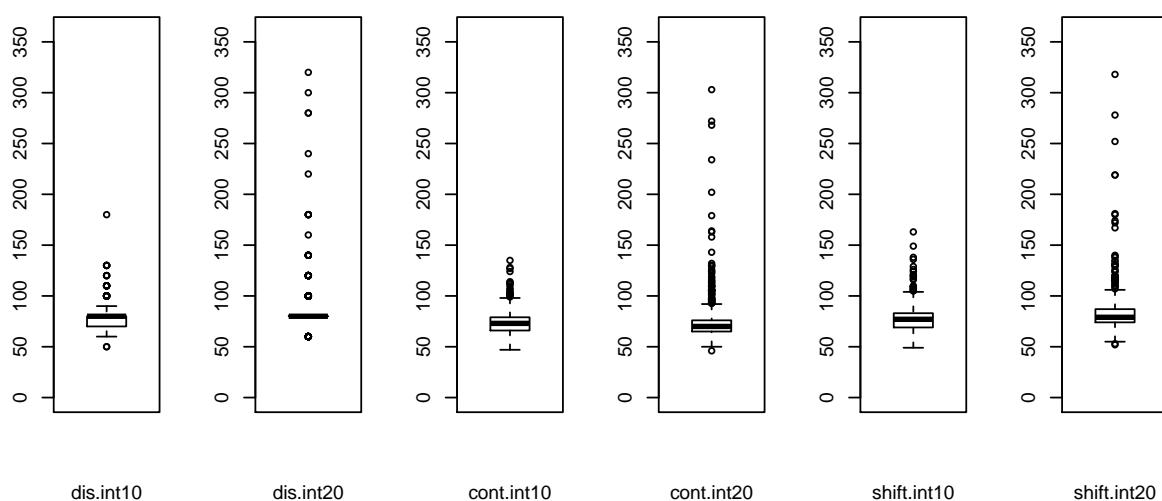
cens=0; jump=1; tau=100; n=5000



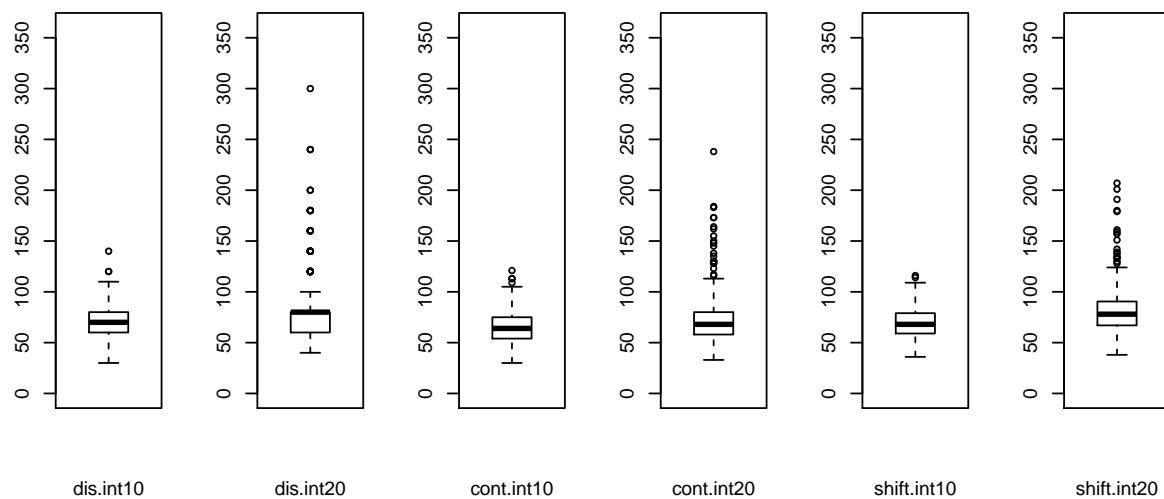
cens=0; jump=0; tau=90; n=1000



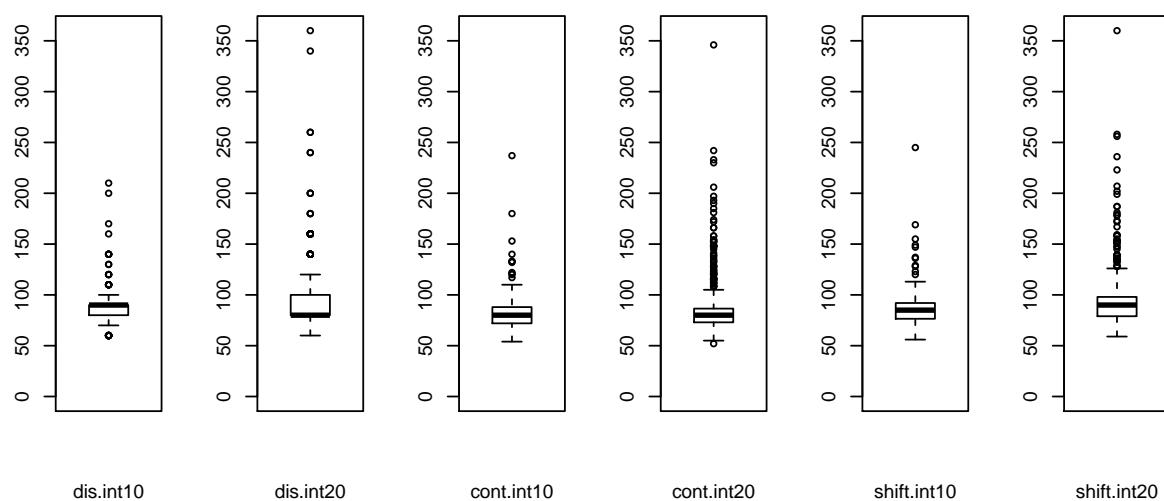
cens=0; jump=0; tau=90; n=5000



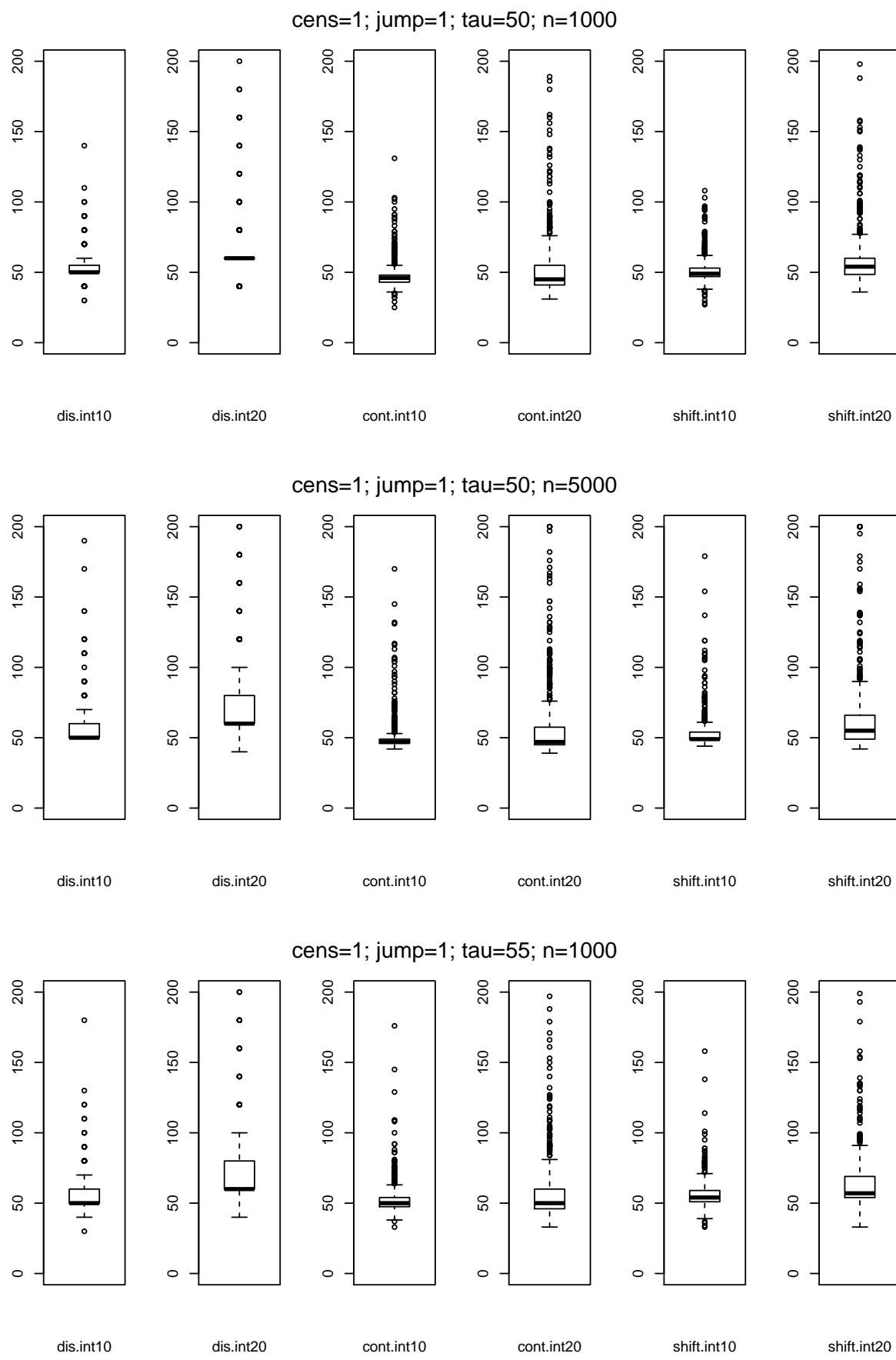
cens=0; jump=0; tau=100; n=1000



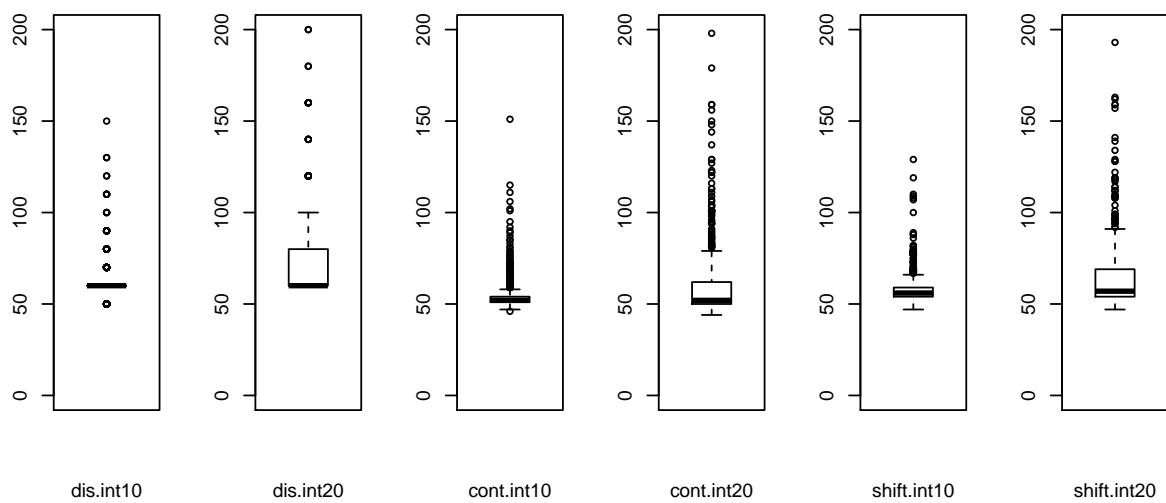
cens=0; jump=0; tau=100; n=5000



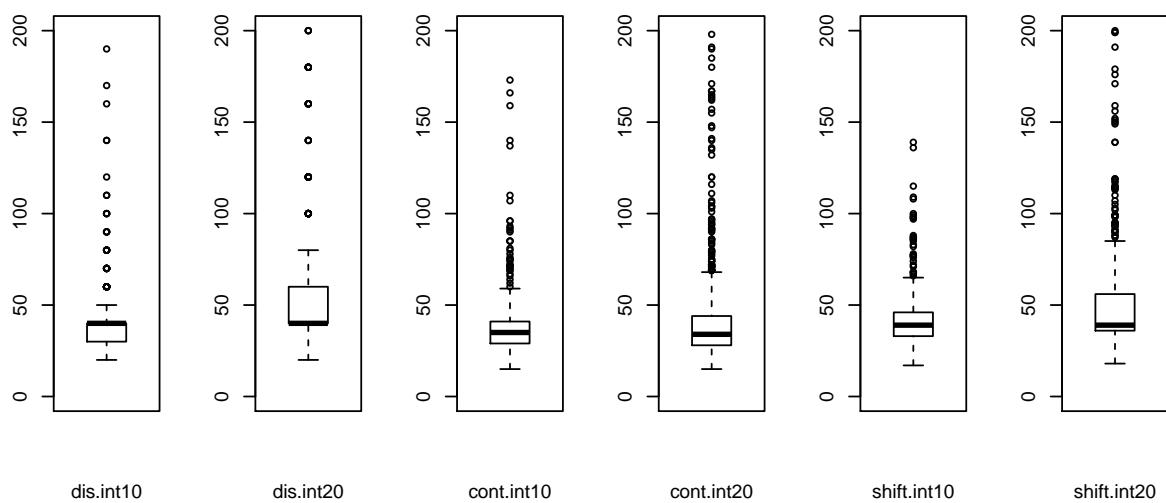
Boxplots für Szenario 2 mit  $\tau = \{50, 55\}$  und  $\tau_{max} = 200$ :



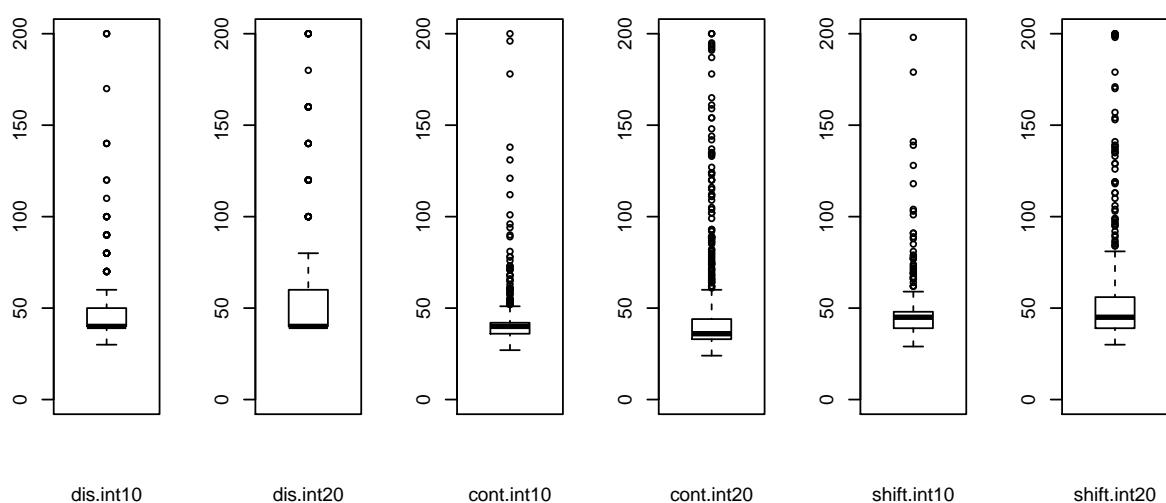
cens=1; jump=1; tau=55; n=5000

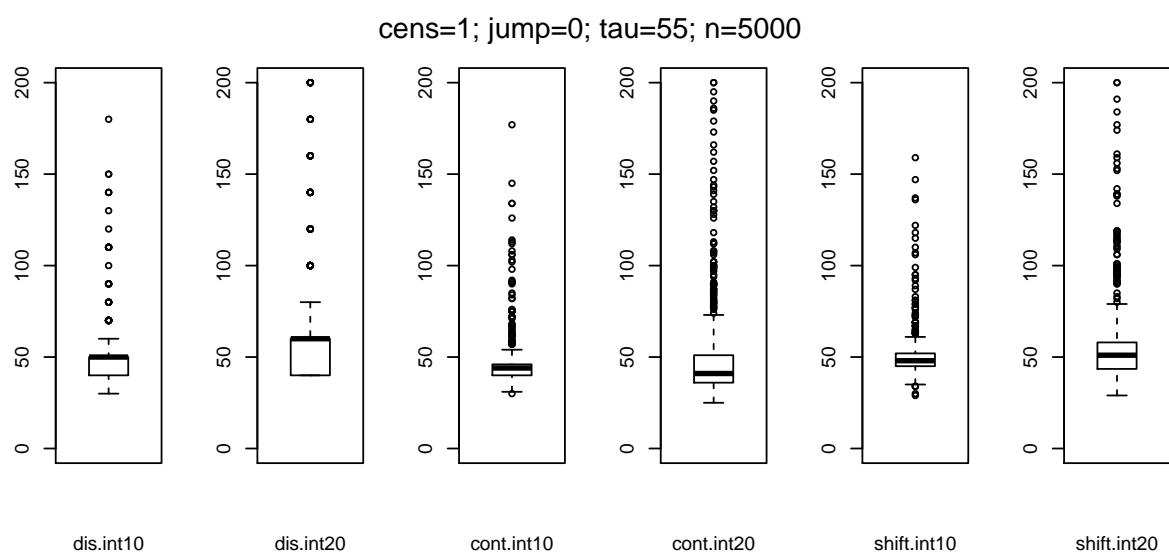
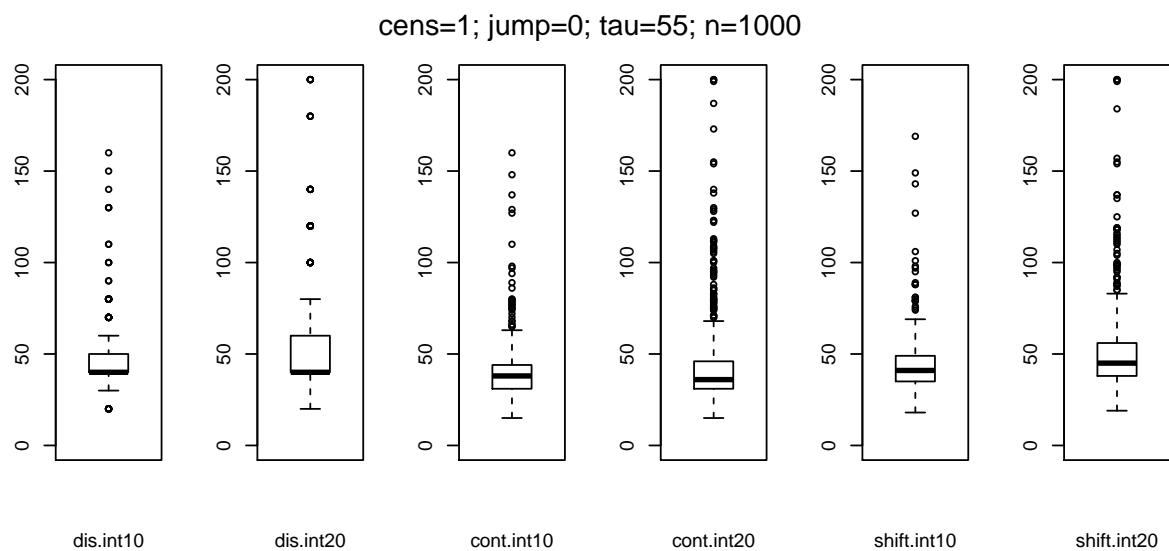


cens=1; jump=0; tau=50; n=1000

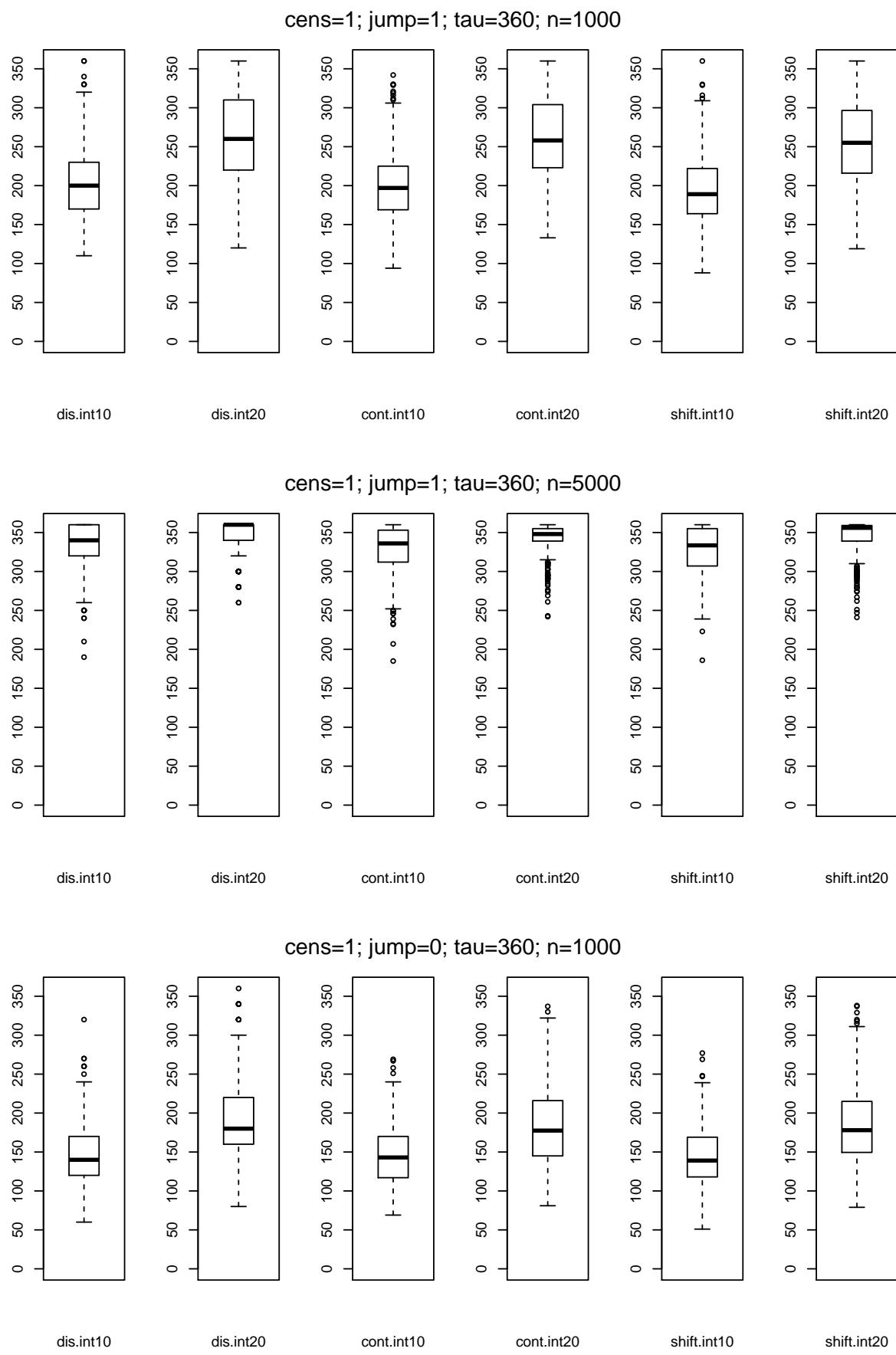


cens=1; jump=0; tau=50; n=5000

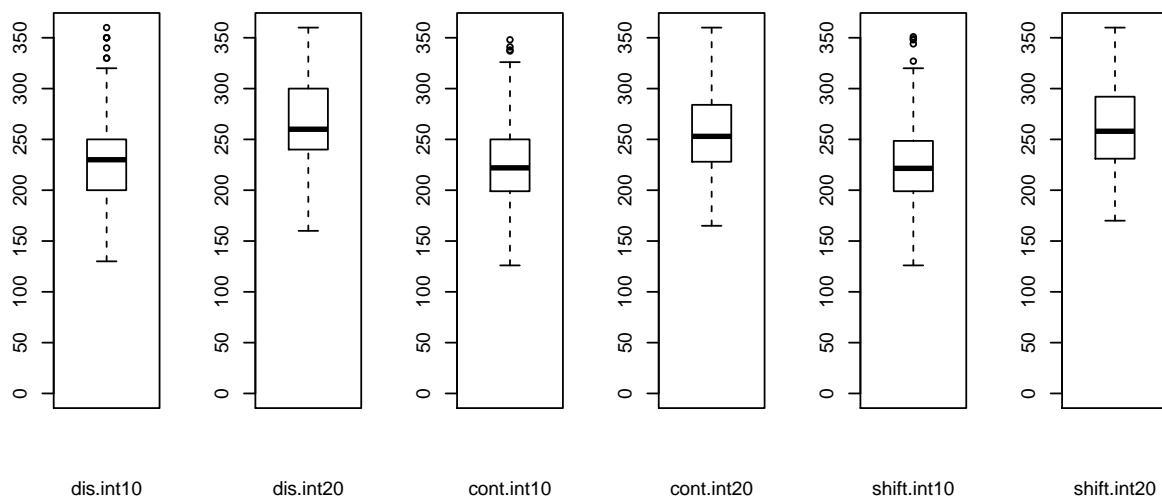




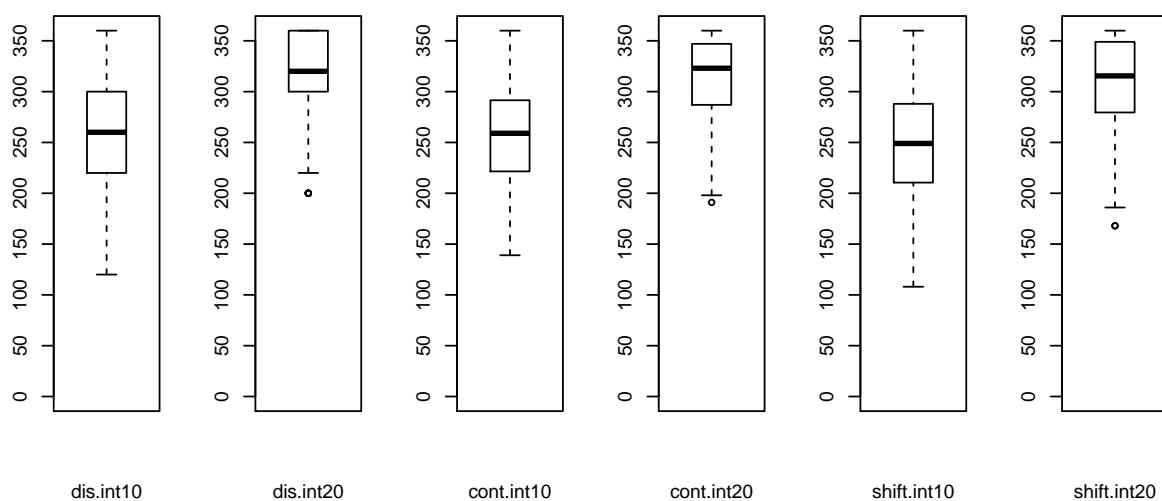
Boxplots für Szenario 3 mit  $\tau = 360$  und  $\tau_{max} = 200$ :



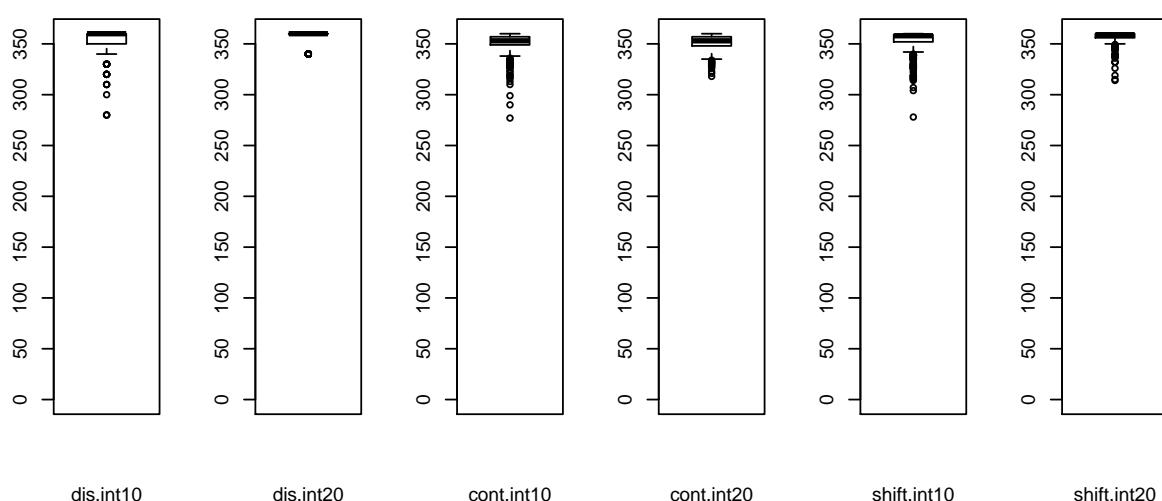
cens=1; jump=0; tau=360; n=5000

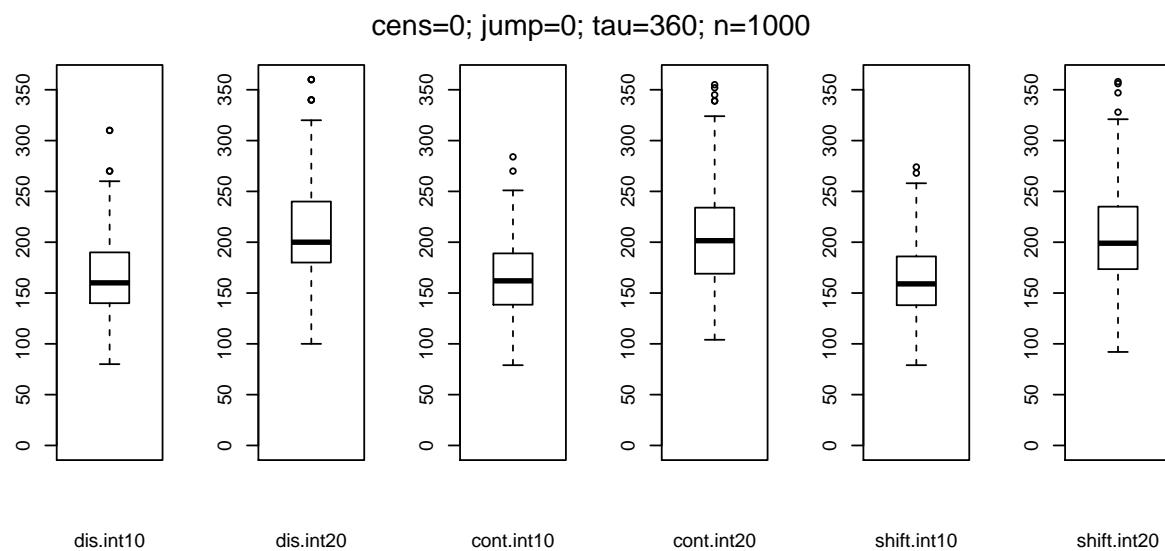


cens=0; jump=1; tau=360; n=1000



cens=0; jump=1; tau=360; n=5000

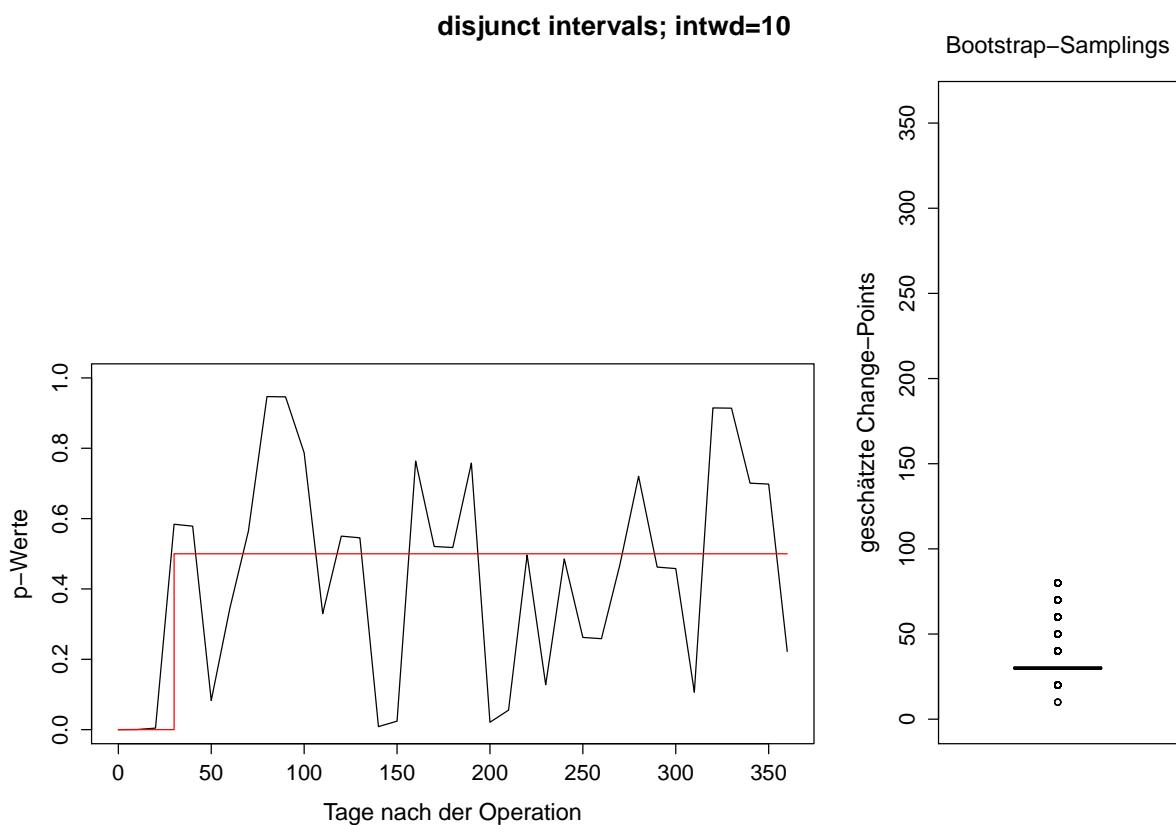


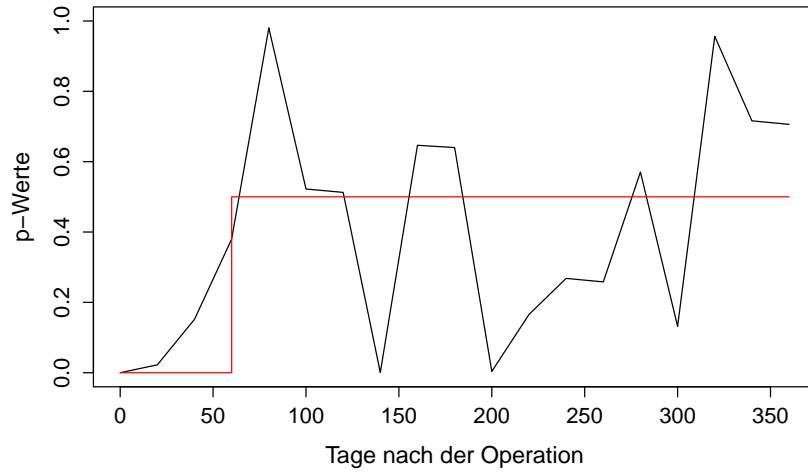


## B Plots zu den Datensätzen

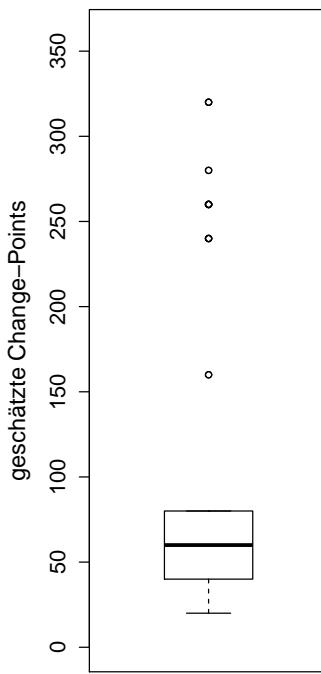
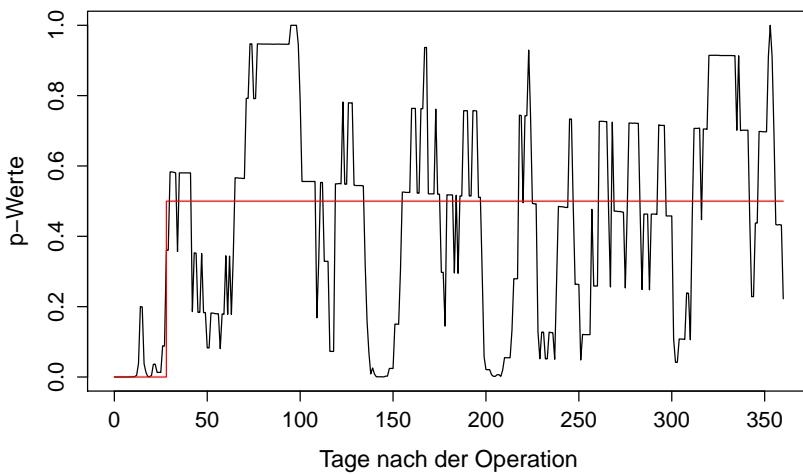
### B.1 Datensatz mit Überlebensdauern nach Leberresektion

Nachfolgend sind die Plots der p-Werte zu den Ergebnissen aus Kapitel 5.1 für die Daten nach [Schiergens u. a., 2015] mit zugehörigen Boxplots aus den Bootstrap-Samplings dargestellt. Die Schätzungen wurden für die drei vorgestellten Methoden mit einer Intervallbreite (“intwd”) von jeweils 10 und 20 Tagen durchgeführt. Die Bezeichnung “disjunct intervals” steht dabei für disjunkte Intervalle ohne Verschiebung, “continuous intervals” für fortlaufende Intervalle und “shifted intervals” für die Schätzung mit Intervallverschiebung. Der Wert für  $\tau_{max}$  wurde jeweils mit 360 vorgegeben.



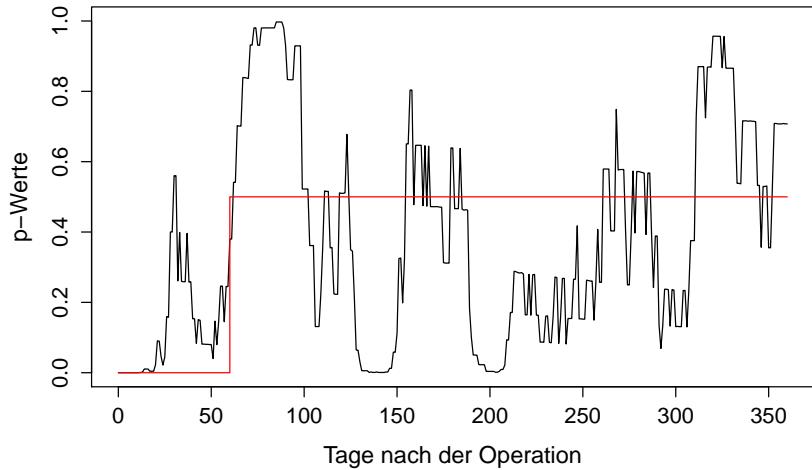
**disjunct intervals; intwd=20**

Bootstrap-Samplings

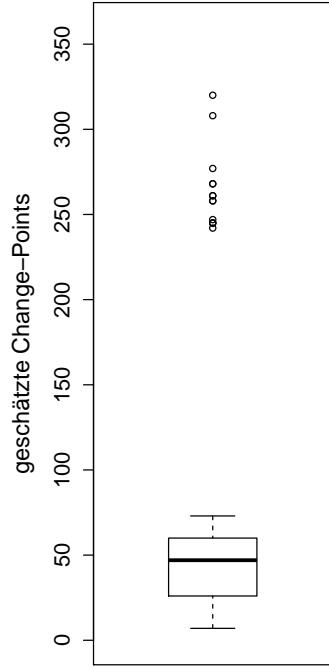
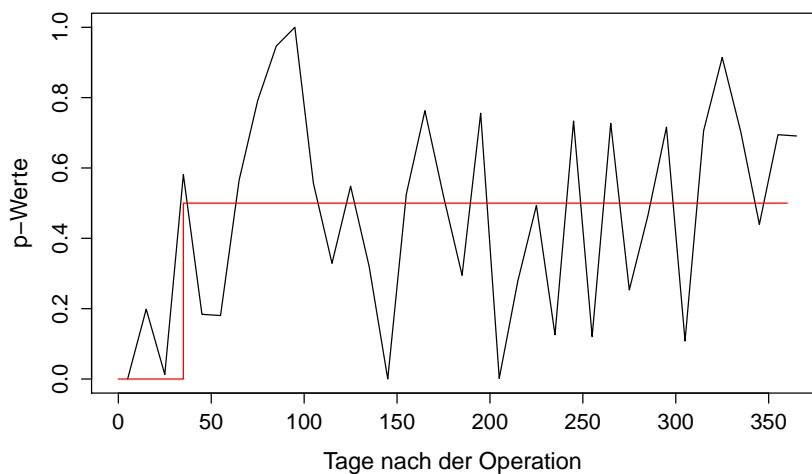
**continuous intervals; intwd=10**

Bootstrap-Samplings

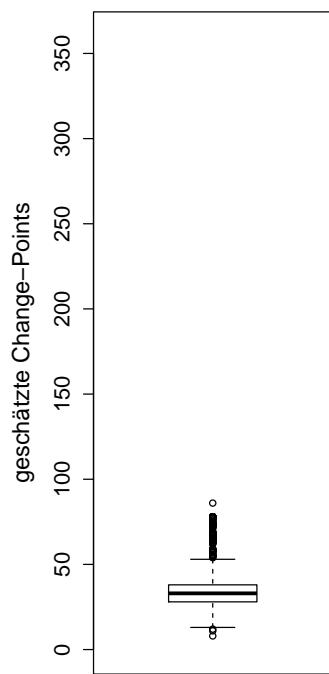


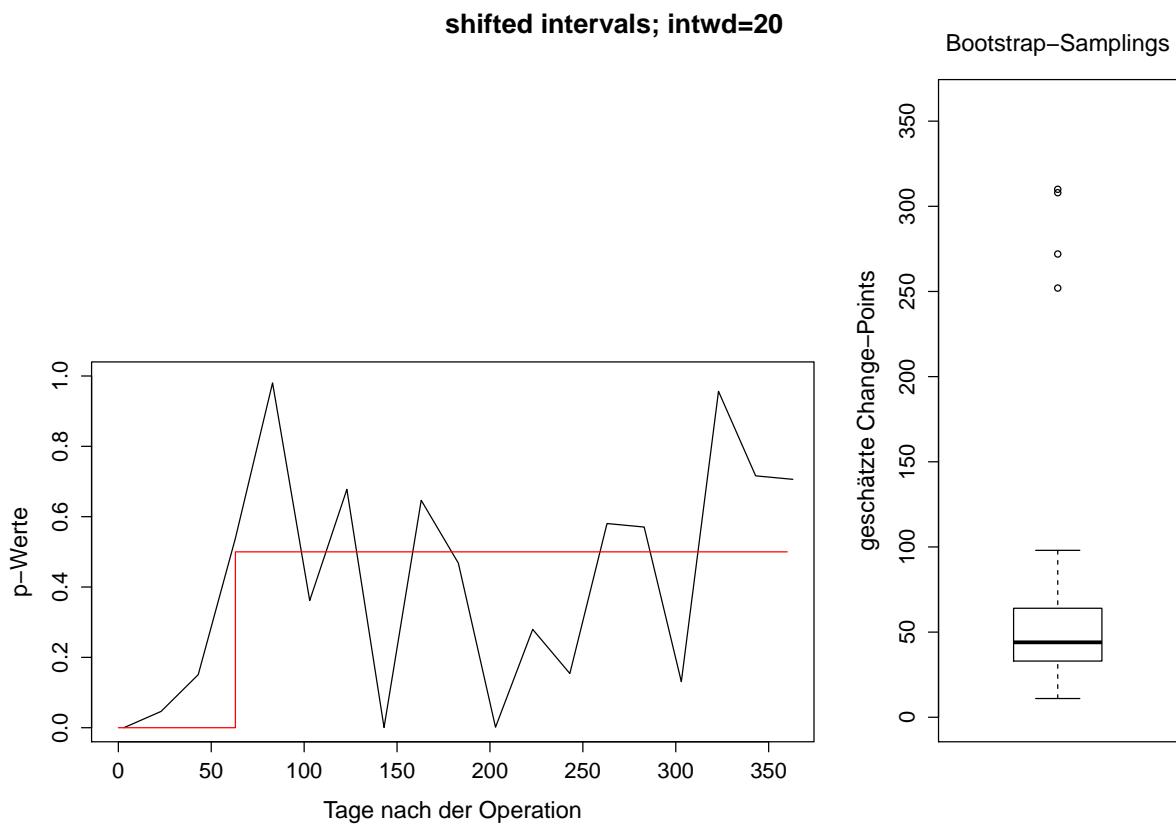
**continuous intervals; intwd=20**

Bootstrap-Samplings

**shifted intervals; intwd=10**

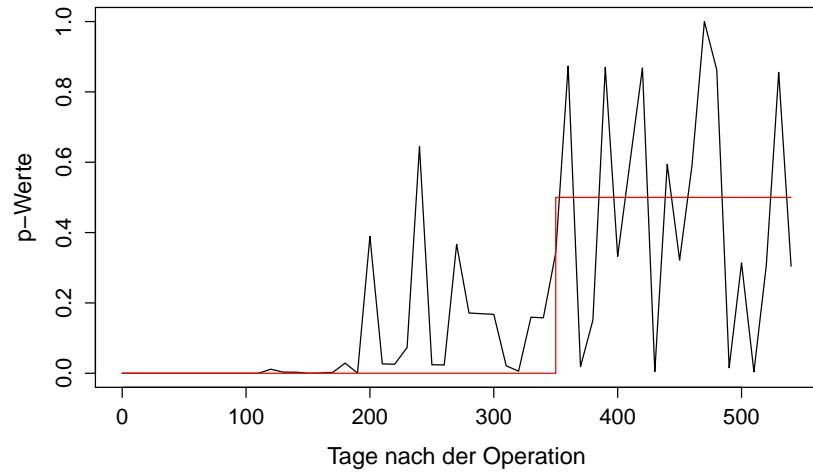
Bootstrap-Samplings



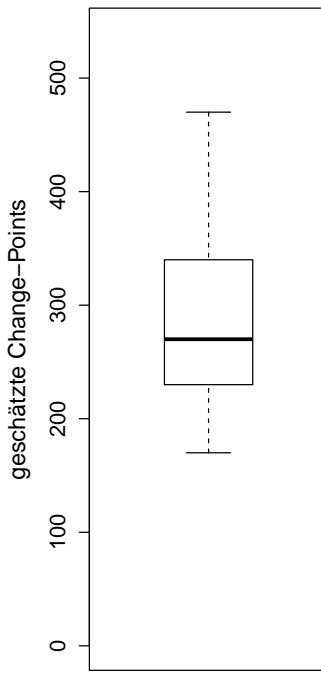
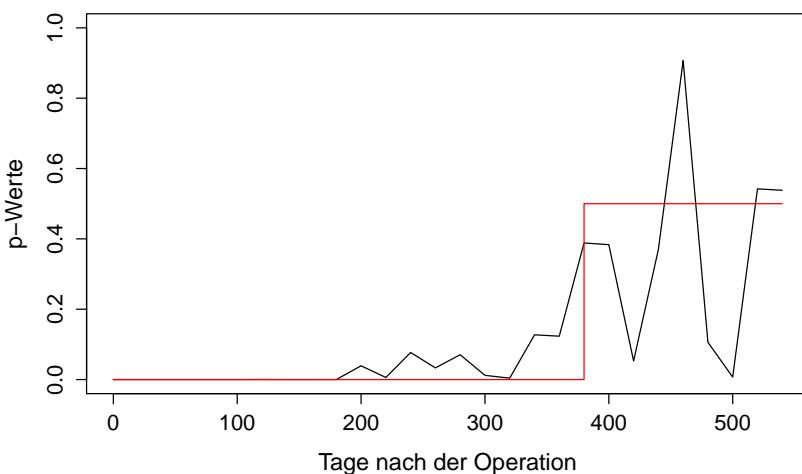


## B.2 Hartl-Datensatz

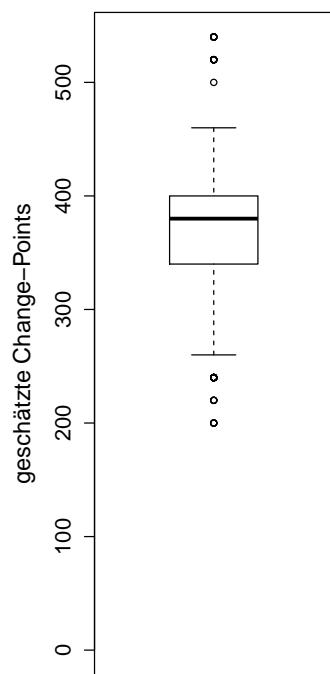
Nachfolgend sind die Plots der p-Werte zu den Ergebnissen aus Kapitel 5.2 für die Daten nach [Hartl u. a., 2007] mit zugehörigen Boxplots aus den Bootstrap-Samplings dargestellt. Die Schätzungen wurden für die drei vorgestellten Methoden mit einer Intervallbreite ("intwd") von jeweils 10 und 20 Tagen durchgeführt. Die Bezeichnung "disjunct intervals" steht dabei für disjunkte Intervalle ohne Verschiebung, "continuous intervals" für fortlaufende Intervalle und "shifted intervals" für die Schätzung mit Intervallverschiebung. Der Wert für  $\tau_{max}$  wurde jeweils mit 540 vorgegeben.

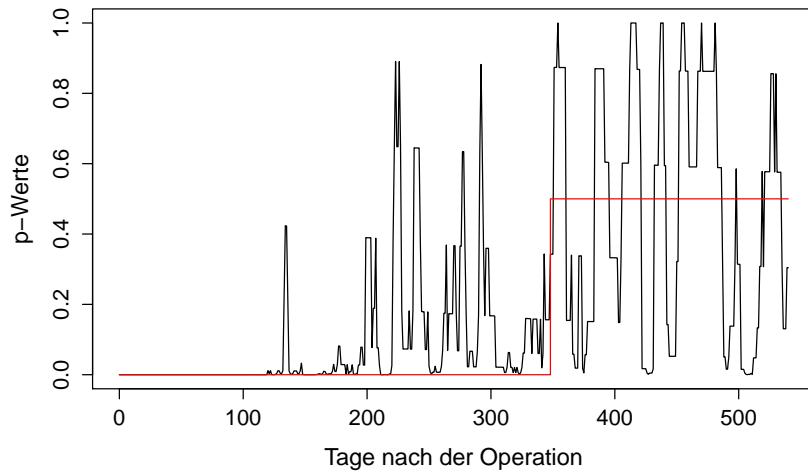
**disjunct intervals; intwd=10**

Bootstrap-Samplings

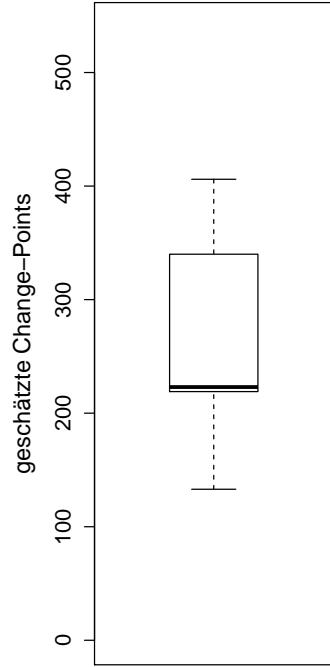
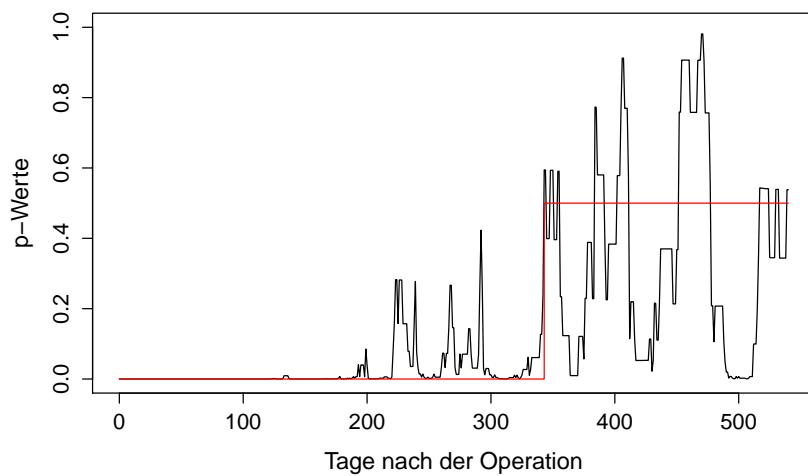
**disjunct intervals; intwd=20**

Bootstrap-Samplings

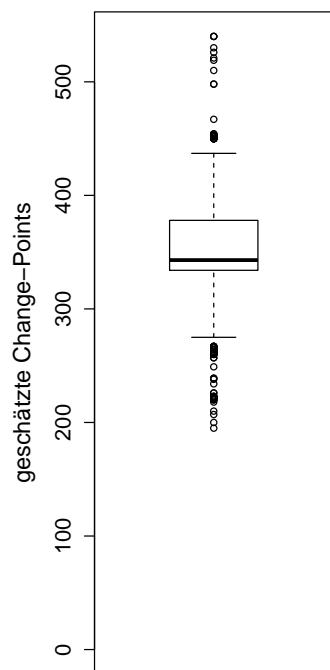


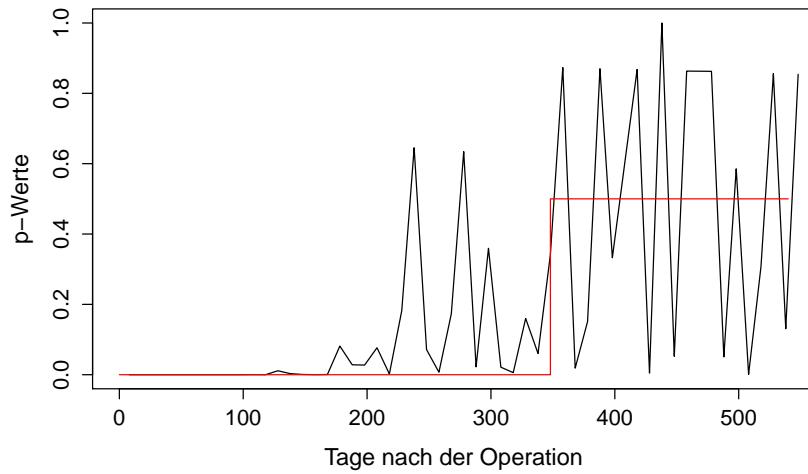
**continuous intervals; intwd=10**

Bootstrap-Samplings

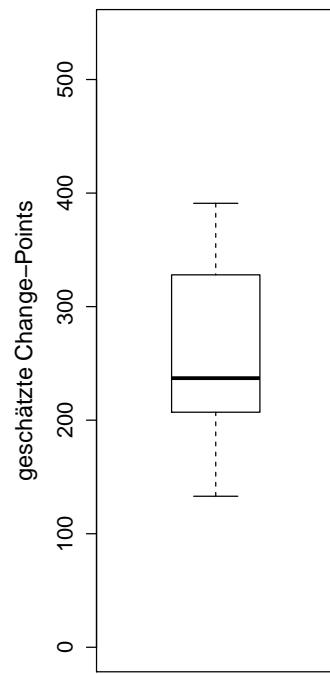
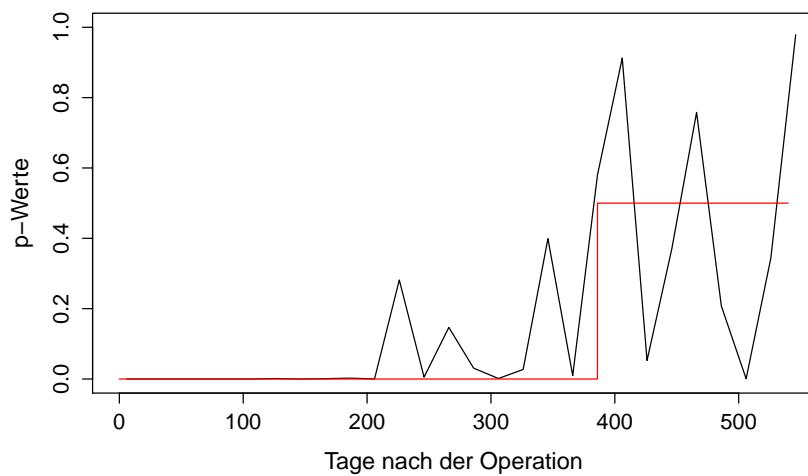
**continuous intervals; intwd=20**

Bootstrap-Samplings

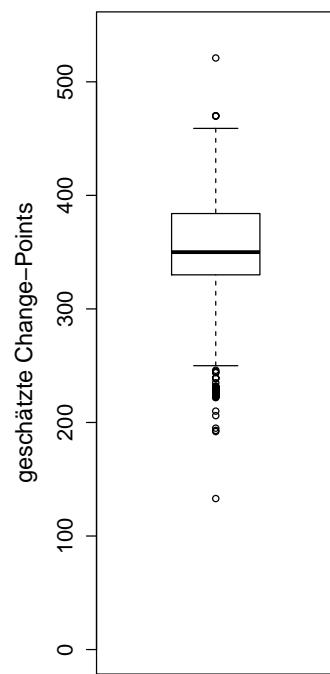


**shifted intervals; intwd=10**

Bootstrap-Samplings

**shifted intervals; intwd=20**

Bootstrap-Samplings



## Literaturverzeichnis

[Hamerle u. Tutz 1989] HAMERLE, Alfred ; TUTZ, Gerhard: *Diskrete Modelle zur Analyse von Verweildauer und Lebenszeiten*. Bd. 568. Campus, 1989

[Hartl u. a. 2007] HARTL, Wolfgang H. ; WOLF, Hilde ; SCHNEIDER, Christian P. ; KÜCHENHOFF, Helmut ; JAUCH, Karl-Walter: Acute and long-term survival in chronically critically ill surgical patients: a retrospective observational study. In: *Critical Care* 11 (2007), Nr. 3, S. R55

[Hess u. Gentleman 2010] HESS, K ; GENTLEMAN, R: Muhaz: hazard function estimation in survival analysis. In: *R package version 1* (2010), Nr. 5

[Holling u. Gediga 2013] HOLLING, Heinz ; GEDIGA, Günther: *Statistik-Wahrscheinlichkeitstheorie und Schätzverfahren*. Hogrefe Verlag, 2013

[Keppler u. Müller 2013] KEPPLER, Janine ; MÜLLER, Christine: Maximum-Likelihood-Schätzung für zensierte und getrimmte Daten. (2013)

[Küchenhoff u. a. 2015] KÜCHENHOFF, Helmut ; BRAZZALE, Alessandra R. ; HARTL, Wolfgang: Nonparametric changepoint estimation for survival distributions with a partially constant hazard rate. (2015)

[Li u. a. 2013] LI, Yunxia ; QIAN, Lianfen ; ZHANG, Wei: Estimation in a change-point hazard regression model with long-term survivors. In: *Statistics & Probability Letters* 83 (2013), Nr. 7, S. 1683–1691

[Mallik u. a. 2011] MALLIK, A. ; SEN, B. ; BANERJEE, M. ; MICHAELIDIS, G.: Threshold estimation based on a p-value framework in dose-response and regression setting. In: *arXiv preprint arXiv:1008.4316* (2011)

[Muller u. Wang 1994] MULLER, Hans-Georg ; WANG, Jane-Ling: Hazard rate estimation under random censoring with varying kernels and bandwidths. In: *Biometrics* (1994), S. 61–76

[Schiergens u. a. 2015] SCHIERGENS, Tobias S. ; DÖRSCH, Maximilian ; MITTERMEIER, Laura ; BRAND, Katharina ; KÜCHENHOFF, Helmut ; LEE, Serene M. L. ; FENG, Hao ; JAUCH, Karl-Walter ; WERNER, Jens ; THASLER, Wolfgang E.: Thirty-day mortality leads to underestimation of postoperative death after liver resection: A novel method to define the acute postoperative period. (2015)

[Schneider u. a. 2010] SCHNEIDER, Christian P. ; FERTMANN, Jan ; GEIGER, Simon ; WOLF, Hilde ; BIERMAIER, Helga ; HOFNER, Benjamin ; KÜCHENHOFF, Helmut ; JAUCH, Karl-Walter ; HARTL, Wolfgang H.: Long-term survival after surgical critical illness: the impact of prolonged preceding organ support therapy. In: *Annals of surgery* 251 (2010), Nr. 6, S. 1145–1153

[Yang u. a. 2012] YANG, Chia-Han ; YUAN, Tao ; KUO, Way ; KUO, Yue: Non-parametric Bayesian modeling of hazard rate with a change point for nanoelectronic devices. In: *IIE Transactions* 44 (2012), Nr. 7, S. 496–506

## **Eidesstattliche Erklärung**

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

München, den 02.11.2015

Stefanie Krügel