

Bachelorarbeit

**Korrigierte Scorefunktionen unter
nichtnormalverteiltem Messfehler**

Jana Gauß

Betreuer: Prof. Dr. Thomas Augustin

8. September 2020

Ludwig-Maximilians-Universität München
Institut für Statistik

Abstract

Diese Arbeit beschäftigt sich mit der Methode der korrigierten Scorefunktionen für allgemeine Messfehlerverteilungen sowie unter der Annahme von konkreten, von der Normalverteilung verschiedenen Messfehlerverteilungen. Im linearen Regressionsmodell ist die Verteilung der Messfehler bei der Korrektur irrelevant. Für die Poisson- und Gamma-Regression werden exakte korrigierte Scorefunktionen für allgemeine Messfehlerverteilungen hergeleitet. Dabei spielt die Momenterzeugende Funktion des Messfehlers eine wichtige Rolle.

Nicht jede Methode zu korrigierten Scorefunktionen unter der Annahme von normalverteilten Messfehlern lässt sich ohne Weiteres verallgemeinern. Bei der Monte Carlo Corrected Score Methode ist beispielsweise keine direkte Übertragbarkeit auf nichtnormalverteilte Messfehler gewährleistet.

Die Ergebnisse von Simulationen mit Poisson-verteilter Zielvariable sowie Messfehlern aus verschiedenen Verteilungen (schiefe Normal-, Laplace- sowie Mischverteilungen) zeigen, dass die Korrektur teilweise schon für kleine Stichprobenumfänge ($n \leq 100$) sinnvolle Ergebnisse liefert. Die Güte und die Robustheit der Schätzung variieren dabei zwischen den betrachteten Messfehlerverteilungen. Die Resultate der Simulation lassen zudem vermuten, dass eine Korrektur unter der Annahme von normalverteilten Messfehlern in vielen Fällen auch dann eine gute Schätzung liefert, wenn diese Verteilungsannahme verletzt ist. Hier ergeben sich ebenfalls Unterschiede zwischen den verschiedenen (wahren) Messfehlerverteilungen.

Inhaltsverzeichnis

Notation	1
1. Einleitung	3
2. Korrigierte Scorefunktionen	5
2.1. Methodik	5
2.1.1. Maximum-Likelihood- und M-Schätzer	5
2.1.2. Korrigierte log-Likelihood und korrigierte Scorefunktion	6
2.1.3. Varianzschätzung	8
2.2. Beispiele für korrigierte Scorefunktionen	9
2.2.1. Annahmen über den Messfehler & Vorüberlegungen	9
2.2.2. Lineare Regression	11
2.2.3. Poisson-Regression	12
2.2.4. Gamma-Regression	14
2.2.5. Logistische Regression	15
3. Mögliche Verteilungen der Messfehler	18
3.1. Schiefe Normalverteilung	18
3.1.1. Definition, Momente und Momenterzeugende Funktion	18
3.1.2. Korrigierte Scorefunktion für die Simulation	20
3.2. Laplace-Verteilung	22
3.2.1. Definition, Momente und Momenterzeugende Funktion	22
3.2.2. Korrigierte Scorefunktion für die Simulation	24
3.2.3. Trunkierte Verteilung	25
3.3. Mischverteilungen	28
3.3.1. Definition, Momente und Momenterzeugende Funktion	28
3.3.2. Bei der Simulation betrachtete Fälle	29
3.3.3. Korrigierte Scorefunktion für die Simulation	30
4. Praktische Umsetzung & Simulation	32
4.1. Bestimmung der Nullstelle	32
4.1.1. Newton-Raphson-Algorithmus	32
4.1.2. Mögliche Probleme	33
4.2. Allgemeines Vorgehen bei der Simulation	35

4.3. Schiefe Normalverteilung	36
4.3.1. Unterschiedliche Messfehlervarianzen	36
4.3.2. Unterschiedliche Schiefeparameter	39
4.3.3. Unterschiedliche Stichprobenumfänge	40
4.4. Laplace-Verteilung	42
4.4.1. Unterschiedliche Messfehlervarianzen	42
4.4.2. Unterschiedliche Stichprobenumfänge	44
4.4.3. Trunkierte Laplace-Verteilung	45
4.5. Mischverteilungen	47
4.5.1. Unterschiedliche Messfehlervarianzen	47
4.5.2. Unterschiedliche Stichprobenumfänge	49
4.6. Zusammenfassung	52
5. Fazit & Ausblick	53
A. Anhang	55
A.1. Monte Carlo Corrected Score	55
A.2. Korrekturmöglichkeiten logistische Regression	58
A.3. Momenterzeugende Funktion Laplace-Verteilung	60
A.4. Momenterzeugende Funktion trunkierte Laplace-Verteilung	62
A.5. Umgang mit Warnmeldungen bei der Simulation	64
A.6. Weitere Abbildungen Simulation	66
Abbildungs- & Tabellenverzeichnis	71
Literatur	73

Notation

$Y = (Y_1, Y_2, \dots, Y_n)^t$	Vektor der Zielgrößen, $\in \mathbb{R}^n$
$X = (X_1, X_2, \dots, X_n)^t$	Designmatrix der wahren, nicht beobachtbaren Kovariablen, $X_i \in \mathbb{R}^p$, $X \in \mathbb{R}^{n \times p}$
$U = (U_1, U_2, \dots, U_n)^t$	Additiver Messfehler in den X -Variablen, $U_i \in \mathbb{R}^p$, $U \in \mathbb{R}^{n \times p}$
$W = (W_1, W_2, \dots, W_n)^t$	Matrix der fehlerhaft gemessenen Kovariablen, $W_i = X_i + U_i$, $W_i \in \mathbb{R}^p$, $W \in \mathbb{R}^{n \times p}$
Σ_{U_i}	Kovarianzmatrix von U_i , $\in \mathbb{R}^{p \times p}$
$M_{U_i}(\beta)$	Momenterzeugende Funktion von U_i , $M_{U_i}(\beta) = \mathbb{E}(\exp(\beta^t U_i))$
$\ell_{X_i}(\beta, Y_i, X_i)$, $\ell_X(\beta, Y, X)$	wahre, ideale log-Likelihood, $\ell_{X_i}(\beta, Y_i, X_i) = \log(f_i(Y_i X_i, \beta))$, $\ell_X(\beta, Y, X) = \sum_{i=1}^n \ell_{X_i}(\beta, Y_i, X_i)$
$s_{X_i}(\beta, Y_i, X_i)$, $s_X(\beta, Y, X)$	wahre, ideale Scorefunktion, $\in \mathbb{R}^p$, $s_{X_i}(\beta, Y_i, X_i) = \frac{\delta}{\delta \beta} \ell_{X_i}(\beta, Y_i, X_i)$, $s_X(\beta, Y, X) = \sum_{i=1}^n s_{X_i}(\beta, Y_i, X_i)$
$\ell_{X_i}(\beta, Y_i, W_i)$, $\ell_X(\beta, Y, W)$	naive log-Likelihood (Ersetzen von X_i bzw. X durch W_i bzw. W in $\ell_{X_i}(\beta, Y_i, X_i)$ bzw. $\ell_X(\beta, Y, X)$)
$s_{X_i}(\beta, Y_i, W_i)$, $s_X(\beta, Y, W)$	naive Scorefunktion (analog zur naiven log-Likelihood)
$\ell_{W_i}(\beta, Y_i, W_i)$, $\ell_W(\beta, Y, W)$	korrigierte log-Likelihood, $\ell_W(\beta, Y, W) = \sum_{i=1}^n \ell_{W_i}(\beta, Y_i, W_i)$
$s_{W_i}(\beta, Y_i, W_i)$, $s_W(\beta, Y, W)$	korrigierte Scorefunktion, $s_W(\beta, Y, W) = \sum_{i=1}^n s_{W_i}(\beta, Y_i, W_i)$

β_t	wahrer Parameter, $\in \mathbb{R}^p$
$\hat{\beta}_{ML}$	Maximum-Likelihood-Schätzer, Lösung von $s_X(\beta, Y, X) = 0$
$\hat{\beta}_{naiv}$	naiver Schätzer, Lösung von $s_X(\beta, Y, W) = 0$
$\hat{\beta}_{cor}$	korrigierter Schätzer, Lösung von $s_W(\beta, Y, W) = 0$
$\hat{\beta}_{normal}$	korrigierter Schätzer unter der (falschen) Annahme von normalverteilten Messfehlern (Simulation)
$\hat{\beta}_{tr}$	korrigierter Schätzer bei Messfehlern aus einer trunkierten Laplace-Verteilung (Simulation)
$\hat{\beta}_{lapl}$	korrigierter Schätzer unter der (falschen) Annahme einer untrunkierten Laplace-Verteilung (Simulation)
$\tilde{Y}_i, \quad i = 1, \dots, n$	Vektor, der alle beobachteten Größen enthält (Zielvariable, Kovariablen o. mit Messfehler behafteter Ersatz), nur relevant für M-Schätzer (Abschnitt 2.1.1)
$\psi_i(\beta, \tilde{Y}_i)$	Schätzfunktion bei der M-Schätzung
$A(\beta)$	Komponente der Kovarianzmatrix bei M-Schätzern, siehe Abschnitt 2.1.3
$B(\beta)$	Komponente der Kovarianzmatrix bei M-Schätzern, siehe Abschnitt 2.1.3

Im Anhang werden teilweise weitere Symbole verwendet. Da diese für die restliche Arbeit von untergeordneter Relevanz sind, werden sie erst in den jeweiligen Abschnitten des Anhangs erläutert.

1. Einleitung

Häufig können in Regressionsmodellen in der Praxis die wahren Kovariablen X_i ($X_i \in \mathbb{R}^p, i = 1, \dots, n$) nicht beobachtet werden und es liegt lediglich ein messfehlerbehafteter Ersatz W_i vor. In vielen Fällen handelt es sich dabei um einen sogenannten *klassischen Messfehler*: Der Messfehler U_i ist additiv und es gilt $W_i = X_i + U_i$. W_i ist unverzerrt, d.h. der Erwartungswert des Messfehlers ist 0 (Carroll et al. 2006, Kapitel 1.2)¹.

Die Verwendung der üblichen Methoden zur Parameterschätzung durch Ersetzen von X_i durch W_i resultiert meist in einer Verzerrung (siehe z.B. Stefanski 1985). So führt beispielsweise die normale kleinste-Quadrate-Methode im einfachen linearen Modell zu einer Abschwächung („Attenuation“) des Steigungsparameters (Carroll et al. 2006, Kapitel 3.2). Das Gebiet der Messfehlerkorrektur beschäftigt sich damit, den durch die Messfehler induzierten Bias zu reduzieren oder gar ganz zu beseitigen. Einen umfassenden Überblick zu diesem Thema bietet etwa das Buch *Measurement Error in Nonlinear Models: A Modern Perspective* (Carroll et al. 2006).

Auch den *korrigierten Scorefunktionen* ist dort ein Kapitel gewidmet. Diese von Nakamura (1990) und Stefanski (1989) entwickelte Methode liefert einen allgemeinen Ansatz zur Herleitung von konsistenten Schätzern. Sie basiert auf dem Prinzip, eine Schätzfunktion $s_W(\beta, Y, W)$ (korrigierte Scorefunktion) für die wahre Scorefunktion $s_X(\beta, Y, X)$ herzuleiten. Von Vorteil ist, dass dabei keine Annahmen über die Verteilung der X_i getroffen werden.

In der Literatur zur korrigierten Scorefunktion werden häufig additive, normalverteilte Messfehler angenommen (siehe z.B. Nakamura (1990, Herleitungen von exakten korrigierten Scorefunktionen in Abschnitt 4), die Monte Carlo Corrected Score Methode von Novick und Stefanski (2002) oder die Überlegungen von Buzas (2009) zur Korrektur im logistischen Regressionsmodell).

In dieser Arbeit soll untersucht werden, wie mit Abweichungen von der Normalverteilungsannahme umgegangen werden kann. Dabei werden zum einen die existierenden Ergebnisse (beispielsweise die exakten korrigierten Scorefunktionen von Nakamura) auf allgemeine Messfehlerverteilungen übertragen beziehungsweise ihre Anwendbarkeit bei nichtnormalverteilten Messfehlern überprüft. Zum anderen

¹Beim sogenannten *Berkson-Fehler* gilt stattdessen: $X_i = W_i + U_i$. Während beim klassischen Messfehler die Varianz der gemessenen W_i größer ist als die der wahren X_i , ist es beim Berkson-Fehler umgekehrt (Carroll et al. 2006, Kapitel 1.4).

wird die praktische Umsetzung mithilfe von Simulationen betrachtet: Liefern die hergeleiteten korrigierten Scorefunktionen für endliche Stichprobenumfänge stabile Schätzungen? Dabei werden auch die Konsequenzen untersucht, wenn bei der Korrektur fälschlicherweise von normalverteilten Messfehlern ausgegangen wird.

Diese Arbeit ist wie folgt gegliedert: In Kapitel 2 wird zunächst die Methodik der korrigierten Scorefunktionen genauer erläutert. Anschließend werden exakte korrigierte Scorefunktionen für verschiedene generalisierte Regressionsmodelle hergeleitet. Über die Verteilung der Messfehler werden dabei nur wenige Annahmen getroffen. In Kapitel 3 wird dargelegt, welche Verteilungen als Messfehlerverteilung plausibel sein können. Im abschließenden Kapitel 4 werden Messfehler aus den vorgestellten Verteilungen generiert und die Resultate der naiven, der korrigierten sowie der Schätzung unter Annahme von normalverteilten Messfehlern in einem Poisson-Regressionsmodell präsentiert. Bei der Simulation werden der Stichprobenumfang, die Messfehlervarianz sowie ggf. weitere Parameter der Messfehlerverteilung systematisch variiert, um die Grenzen der Methoden sowie das Verhalten unter unterschiedlichen Bedingungen auszutesten.

2. Korrigierte Scorefunktionen

In diesem Kapitel wird das Konzept der *korrigierten Scorefunktionen* erläutert. Zur besseren Verständlichkeit wird zunächst in das Prinzip der Maximum-Likelihood- und der M-Schätzung eingeführt. Zuletzt werden exakte korrigierte Scorefunktionen für einige generalisierte lineare Regressionsmodelle hergeleitet sowie Möglichkeiten zur Korrektur bei der logistischen Regression aufgezeigt.

2.1. Methodik

2.1.1. Maximum-Likelihood- und M-Schätzer

Sei $Y = (Y_1, Y_2, \dots, Y_n)^t \in \mathbb{R}^n$ ein Vektor mit voneinander unabhängigen Zufallsvariablen, deren Verteilung von den Kovariablen $X = (X_1, X_2, \dots, X_n)^t \in \mathbb{R}^{n \times p}$ und einem unbekanntem Parametervektor $\beta \in \mathbb{R}^p$ abhängt. Eine weitverbreitete Methode zur konsistenten Schätzung von β ist die *Maximum-Likelihood-Schätzung* (siehe z.B. Carroll et al. 2006, A.5; Fahrmeir et al. 2013, B.4).

Im Folgenden bezeichnen $\ell_{X_i}(\beta, Y_i, X_i) = \log(f_i(Y_i|X_i, \beta))$ die log-Likelihood und $s_{X_i}(\beta, Y_i, X_i) = \frac{\delta}{\delta\beta} \ell_{X_i}(\beta, Y_i, X_i)$ die Scorefunktion² von β für eine gegebene Beobachtung (Y_i, X_i) , $i = 1, \dots, n$. Der Maximum-Likelihood-Schätzer $\hat{\beta}_{ML}$ ist der Wert von β , der $\ell_X(\beta, Y, X) = \sum_{i=1}^n \ell_{X_i}(\beta, Y_i, X_i)$ maximiert, er ergibt sich somit durch das Bestimmen der Nullstelle von $\sum_{i=1}^n s_{X_i}(\beta, Y_i, X_i)$.

Unter bestimmten Regularitätsbedingungen ist der ML-Schätzer asymptotisch (multivariat) normalverteilt mit Erwartungswert β_t (wahrer Parameter) und der inversen Fisherinformationsmatrix $I_n^{-1}(\beta_t)$ als Varianz³. Die Konsistenz des ML-Schätzers beruht darauf, dass der Erwartungswert der Scorefunktion am wahren Parameter β_t 0 ist ($\mathbb{E}(s_{X_i}(\beta_t, Y_i, X_i)) = 0$, $i = 1, \dots, n$). Damit gehört der Maximum-Likelihood-Schätzer zu den von Huber (1964) vorgestellten *M-Schätzern* (siehe auch Carroll et al. 2006, A.6).

² $\frac{\delta}{\delta\beta} \ell_{X_i}(\beta, Y_i, X_i) = (\frac{\delta}{\delta\beta_1} \ell_{X_i}(\beta, Y_i, X_i), \dots, \frac{\delta}{\delta\beta_p} \ell_{X_i}(\beta, Y_i, X_i))^t$, die Scorefunktion bildet somit nach \mathbb{R}^p ab. Mit einer Nullstelle der Scorefunktion ist ein Nullvektor in \mathbb{R}^p gemeint; selbiges gilt auch für die später vorgestellte korrigierte Scorefunktion.

³ $I_n(\beta_t) = -\sum_{i=1}^n \mathbb{E}(\frac{\delta}{\delta\beta^t} s_{X_i}(\beta_t, Y_i, X_i)) = \sum_{i=1}^n \mathbb{E}(s_{X_i}(\beta_t, Y_i, X_i) s_{X_i}(\beta_t, Y_i, X_i)^t)$

Die Einträge der inversen Fisherinformationsmatrix $I_n^{-1}(\beta_t)$ konvergieren gegen 0 für $n \rightarrow \infty$ (Carroll et al. 2006, A.5)

Ein M -Schätzer ist die Lösung einer Gleichung der Form ($\psi_i \in \mathbb{R}^p$)

$$\frac{1}{n} \sum_{i=1}^n \psi_i(\beta, \tilde{Y}_i) = 0 \quad (2.1)$$

wobei die Schätzfunktion ψ_i am wahren Parameter β_t Erwartungswert 0 hat:

$$\mathbb{E}(\psi_i(\beta_t, \tilde{Y}_i)) = 0 \quad \text{für } i = 1, \dots, n \quad (2.2)$$

(Carroll et al. 2006, A.6). Der Vektor \tilde{Y}_i enthält alle beobachteten Variablen, d.h. Zielvariable Y_i , Kovariablen X_i oder mit Messfehler behafteter Ersatz W_i . Wenn $\psi_i(\beta, \tilde{Y}_i)$ die Scorefunktion $s_{X_i}(\beta, Y_i, X_i)$ ist, ergibt sich der Maximum-Likelihood-Schätzer.

Die Lösung von (2.1), $\hat{\beta}$, ist unter gewissen Regularitätsbedingungen (Huber 1967) ein konsistenter Schätzer für β : Für jeden Wert von β konvergiert $\frac{1}{n} \sum_{i=1}^n \psi_i(\beta, Y_i, X_i)$ nach dem Gesetz der großen Zahlen gegen seinen Erwartungswert. Der wahre Parameter β_t ist eine Nullstelle dieses Erwartungswerts (und β_t muss die einzige Nullstelle sein; dies ist eine der Regularitätsbedingungen), wodurch $\hat{\beta}$ gegen β_t konvergiert. Darüber hinaus ist $\hat{\beta}$ asymptotisch (multivariat) normalverteilt mit Erwartungswert β_t (Carroll et al. 2006, A.6). Zur Formel der Kovarianzmatrix und ihrer Schätzung siehe Abschnitt 2.1.3.

2.1.2. Korrigierte log-Likelihood und korrigierte Scorefunktion

Nun können die wahren X_i -Werte nicht beobachtet werden, sondern lediglich die mit einem Messfehler⁴ behafteten W_i . Das Einsetzen von W_i statt X_i in die log-Likelihood bzw. die Scorefunktion liefert die naive log-Likelihood $\ell_{X_i}(\beta, Y_i, W_i)$ sowie die naive Scorefunktion $s_{X_i}(\beta, Y_i, W_i)$. Die Nullstelle von $\sum_{i=1}^n s_{X_i}(\beta, Y_i, W_i)$, der naive Schätzer $\hat{\beta}_{naiv}$, ist im Allgemeinen kein konsistenter Schätzer für β , da der Erwartungswert der naiven Scorefunktion am wahren Parameter β_t in der Regel ungleich 0 ist ($\mathbb{E}(s_{X_i}(\beta_t, Y_i, W_i)) \neq 0$) (Nakamura 1990).

Die Idee der *korrigierten Scorefunktion* ist es, eine Schätzfunktion $s_{W_i}(\beta, Y_i, W_i)$ für die ideale Scorefunktion $s_{X_i}(\beta, Y_i, X_i)$ zu konstruieren. Dieses Prinzip wurde

⁴Hier werden nur additive Messfehler betrachtet; das Verfahren der korrigierten Scorefunktion setzt dies allerdings nicht zwingend voraus. Siehe z. B. Nakamura (1990, Abschnitt 4.2) zur Korrektur bei einem multiplikativen Messfehler in einem linearen Regressionsmodell.

unabhängig voneinander von Nakamura (1990) und Stefanski (1989) entwickelt.

Sei F eine offene, konvexe Teilmenge der Parameterraums und $\beta_t \in F$. Eine *korrigierte Scorefunktion* $s_{W_i}(\beta, Y_i, W_i)$ ist eine Funktion mit der Eigenschaft

$$\mathbb{E}(s_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) = s_{X_i}(\beta, Y_i, X_i) \quad \text{für } i = 1, \dots, n \quad (2.3)$$

für alle β aus F (siehe z.B. Carroll et al. 2006, Kapitel 7.4 oder Nakamura 1990). $s_{W_i}(\beta, Y_i, W_i)$ ist unverzerrt, d.h. hat am wahren Parameter β_t Erwartungswert 0:

$$\begin{aligned} \mathbb{E}(s_{W_i}(\beta_t, Y_i, W_i)) &= \mathbb{E}\left(\mathbb{E}(s_{W_i}(\beta_t, Y_i, W_i)|Y_i, X_i)\right) \\ &\stackrel{(2.3)}{=} \mathbb{E}(s_{X_i}(\beta_t, Y_i, X_i)) = 0 \quad \text{für } i = 1, \dots, n \end{aligned} \quad (2.4)$$

Die Lösung von $\sum_{i=1}^n s_{W_i}(\beta, Y_i, W_i) = 0$, der korrigierte Schätzer $\hat{\beta}_{cor}$, ist unter gewissen Regularitätsbedingungen (Nakamura 1990, Proposition 1 & 2) konsistent und als M-Schätzer asymptotisch normalverteilt (Varianz siehe Abschnitt 2.1.3).

Die Schwierigkeit liegt darin, eine Funktion mit der in (2.3) geforderten Eigenschaft zu finden. Häufig ist es von Vorteil, zunächst die log-Likelihood zu korrigieren: Die *korrigierte log-Likelihood* $\ell_{W_i}(\beta, Y_i, W_i)$ ist definiert durch

$$\mathbb{E}(\ell_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) = \ell_{X_i}(\beta, Y_i, X_i) \quad \text{für } i = 1, \dots, n \quad (2.5)$$

für alle $\beta \in F$ (Nakamura 1990). Wenn die Erwartungswertbildung und das Differenzieren nach β vertauschbar sind, ist die Ableitung der korrigierten log-Likelihood $\frac{\delta}{\delta\beta}\ell_{W_i}(\beta, Y_i, W_i) = s_{W_i}(\beta, Y_i, W_i)$ eine korrigierte Scorefunktion:

$$\begin{aligned} \mathbb{E}(s_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) &= \mathbb{E}\left(\frac{\delta}{\delta\beta}\ell_{W_i}(\beta, Y_i, W_i)|Y_i, X_i\right) \\ &= \frac{\delta}{\delta\beta}\mathbb{E}(\ell_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) \\ &\stackrel{(2.5)}{=} \frac{\delta}{\delta\beta}\ell_{X_i}(\beta, Y_i, X_i) = s_{X_i}(\beta, Y_i, X_i) \end{aligned} \quad (2.6)$$

d.h. Gleichung (2.3) ist erfüllt für $i = 1, \dots, n$. In Abschnitt 2.2 werden korrigierte log-Likelihoods und korrigierte Scorefunktionen für verschiedene generalisierte Regressionsmodelle hergeleitet.

Ein Vorteil der Methode der korrigierten Scorefunktionen ist es, dass keine Annahmen über die Verteilung von X getroffen werden müssen (Carroll et al. 2006,

Kapitel 7.4). Dieses Vorgehen gehört somit zu den *funktionalen Methoden*, im Gegensatz zu den *strukturellen Methoden*, die auf Annahmen über die als Zufallsvariablen aufgefassten X -Variablen basieren (Carroll et al. 2006, Kapitel 2.1).

Korrigierte Scorefunktionen sind unverzerrt, d.h. $\mathbb{E}(s_{W_i}(\beta_t, Y_i, W_i)) = 0$ (siehe Gleichung 2.4). Andere Messfehlerkorrekturverfahren auf Basis der Scorefunktion, wie die von Stefanski und Carroll (1987) vorgestellte Conditional-Score-Methode, verwenden ebenfalls unverzerrte Scorefunktionen. Während korrigierte Scorefunktionen stets unverzerrt sind, ist aber umgekehrt nicht jede unverzerrte Scorefunktion auch eine korrigierte (Nakamura 1990).

2.1.3. Varianzschätzung

Die Varianz des korrigierten Schätzers $\hat{\beta}_{cor}$ kann über die Methoden der M-Schätzung bestimmt werden: M-Schätzer $\hat{\beta}$ sind asymptotisch normalverteilt mit Mittelwert β_t und Kovarianzmatrix $\frac{1}{n}A(\beta_t)^{-1}B(\beta_t)(A(\beta_t)^{-1})^t$, wobei

$$A(\beta) = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left(\frac{\delta}{\delta \beta^t} \psi_i(\beta, \tilde{Y}_i) \right) \quad (2.7)$$

$$B(\beta) = \frac{1}{n} \sum_{i=1}^n \text{Cov}(\psi_i(\beta, \tilde{Y}_i)) \quad (2.8)$$

(siehe z.B. Carroll et al. 2006, A.6, dies kann über eine Taylorreihenapproximation hergeleitet werden). Die Kovarianzmatrix kann über den sogenannten *Sandwich-Schätzer* konsistent geschätzt werden:

$$\hat{A}(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \frac{\delta}{\delta \beta^t} \psi_i(\hat{\beta}, \tilde{Y}_i) \quad (2.9)$$

$$\hat{B}(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \psi_i(\hat{\beta}, \tilde{Y}_i) \psi_i(\hat{\beta}, \tilde{Y}_i)^t \quad (2.10)$$

(ebenda, A.6)⁵. Bei der Schätzung der Varianz von $\hat{\beta}_{cor}$ ist $\psi_i(\hat{\beta}, \tilde{Y}_i)$ in (2.9) und (2.10) gleich $s_{W_i}(\hat{\beta}_{cor}, Y_i, W_i)$ (ebenda, Kapitel 7.5). Nakamura (1990) stellt weitere mögliche Varianten zur Schätzung der Kovarianzmatrix vor.

⁵Eine weitere Möglichkeit zur Schätzung ist es, die Formeln (2.7) und (2.8) exakt auszuwerten und dann β durch $\hat{\beta}$ zu ersetzen.

Wenn die Varianz der Messfehler unbekannt ist und geschätzt wird, muss die Berechnung der Kovarianz von $\hat{\beta}_{cor}$ angepasst werden, indem die Varianz der Schätzung der Messfehlervarianz miteinbezogen wird (Carroll et al. 2006, Kapitel 7.5). Siehe z.B. Buzas und Stefanski (1996) zur Schätzung der Kovarianzmatrix von $\hat{\beta}_{cor}$ bei unabhängigen Wiederholungsmessungen und normalverteilten Messfehlern.

2.2. Beispiele für korrigierte Scorefunktionen

2.2.1. Annahmen über den Messfehler & Vorüberlegungen

In diesem Kapitel werden korrigierte log-Likelihoods und korrigierte Scorefunktionen für verschiedene generalisierte Regressionsmodelle hergeleitet. Anders als z.B. bei Nakamura (1990) wird dabei nicht zwingend von normalverteilten Messfehlern ausgegangen. Einige Annahmen werden aber dennoch getroffen (jeweils für $i = 1, \dots, n$):

- $W_i = X_i + U_i$, d.h. es liegt ein additiver Messfehler U_i vor. Statt den wahren X_i wird W_i beobachtet ($W_i, X_i, U_i \in \mathbb{R}^p$).
- U_i ist unabhängig von U_j ($j \neq i$) und von X_j und Y_j ($j = 1, \dots, n$), d.h. die Messfehler sind voneinander und von allen Kovariablen und Zielvariablen unabhängig.
- $\mathbb{E}(U_i) = 0$ (Daraus folgt, dass $\mathbb{E}(W_i) = \mathbb{E}(X_i) + \mathbb{E}(U_i) = X_i$ gilt.)
- Die Verteilung der U_i ist bekannt (wobei die U_i nicht zwingend identisch verteilt sein müssen, Heteroskedastie ist beispielsweise möglich).
- Die Momenterzeugende Funktion⁶ $M_{U_i}(\beta)$ existiert auf einem realistischen Bereich von β und ist zwei mal nach β differenzierbar.

In Kapitel 3 werden verschiedene Verteilungen vorgestellt, denen die Messfehler anstelle einer Normalverteilung folgen könnten und für die diese Annahmen erfüllt sind.

⁶ $M_{U_i}(\beta) = \mathbb{E}(\exp(\beta^t U_i))$ (siehe z.B. Gut 2009, Kapitel 3.3), wenn dieser Erwartungswert existiert und endlich ist.

Folgende Zusammenhänge werden bei den Herleitungen verwendet:

$$\begin{aligned}\mathbb{E}(W_i W_i^t) &= \mathbb{E}((X_i + U_i)(X_i + U_i)^t) = \mathbb{E}(X_i X_i^t + X_i U_i^t + U_i X_i^t + U_i U_i^t) \\ &\stackrel{(*)}{=} X_i X_i^t + \mathbb{E}(U_i U_i^t) = X_i X_i^t + \Sigma_{U_i}\end{aligned}\quad (2.11)$$

mit $\Sigma_{U_i} = \text{Cov}(U_i) = \mathbb{E}(U_i U_i^t)$, wobei in (*) verwendet wurde, dass $\mathbb{E}(X_i U_i^t) = \mathbb{E}(U_i X_i^t) = 0$ gilt (da $\mathbb{E}(U_i) = 0$).

$$\begin{aligned}\mathbb{E}(\exp(\beta^t W_i)) &= \mathbb{E}(\exp(\beta^t X_i + \beta^t U_i)) = \mathbb{E}(\exp(\beta^t X_i) * \exp(\beta^t U_i)) \\ &= \exp(\beta^t X_i) * \mathbb{E}(\exp(\beta^t U_i)) = \exp(\beta^t X_i) * M_{U_i}(\beta)\end{aligned}\quad (2.12)$$

$$\mathbb{E}\left(\frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)}\right) \stackrel{(2.12)}{=} \frac{\exp(\beta^t X_i) * M_{U_i}(\beta)}{M_{U_i}(\beta)} = \exp(\beta^t X_i)\quad (2.13)$$

Gleichung (2.13) spielt für die Korrektur der log-Likelihood eine wichtige Rolle. Die Ableitung von $\frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)}$ kann (unter Verwendung von $\frac{\delta}{\delta x} \log(f(x)) = \frac{\frac{\delta}{\delta x} f(x)}{f(x)}$) wie folgt dargestellt werden:

$$\begin{aligned}\frac{\delta}{\delta \beta} \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} &= \frac{W_i \exp(\beta^t W_i) M_{U_i}(\beta) - \exp(\beta^t W_i) \frac{\delta}{\delta \beta} M_{U_i}(\beta)}{(M_{U_i}(\beta))^2} \\ &= \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} W_i - \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \frac{\frac{\delta}{\delta \beta} M_{U_i}(\beta)}{M_{U_i}(\beta)} \\ &= \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \left(W_i - \frac{\delta}{\delta \beta} \log(M_{U_i}(\beta)) \right)\end{aligned}\quad (2.14)$$

Je nach Gestalt der Momenterzeugenden Funktion kann diese Darstellung mit dem natürlichen Logarithmus sehr sinnvoll sein. Für multivariat normalverteilte Messfehler ($U_i \sim N_p(0, \Sigma)$) gilt zum Beispiel⁷:

$$M_{U_i}(\beta) = \exp\left(\frac{1}{2} \beta^t \Sigma \beta\right)\quad (2.15)$$

⁷Allgemein gilt für die Momenterzeugende Funktion einer multivariat normalverteilten Zufallsvariable $Z \sim N_p(\mu, \Sigma)$: $M_Z(\beta) = \exp(\beta^t \mu + \frac{1}{2} \beta^t \Sigma \beta)$ (siehe z.B. Gut 2009, Kapitel 5.4)

wodurch sich in (2.14) folgender Ausdruck ergibt:

$$\frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \left(W_i - \frac{\delta}{\delta\beta} \log(M_{U_i}(\beta)) \right) = \exp\left(\beta^t W_i - \frac{1}{2}\beta^t \Sigma \beta\right) (W_i - \Sigma\beta) \quad (2.16)$$

2.2.2. Lineare Regression

Seien die Y_i unabhängig normalverteilt mit Erwartungswert $\beta^t X_i$ und Varianz σ^2 , also $Y_i|X_i \sim N(\beta^t X_i, \sigma^2)$ ($i = 1, \dots, n$). Die Dichte von Y_i ist (Fahrmeir et al. 2013, B.1):

$$f(Y_i|X_i, \beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta^t X_i)^2}{2\sigma^2}\right) \quad (2.17)$$

Die log-Likelihood und die Scorefunktion ergeben sich somit zu

$$\begin{aligned} \ell_{X_i}(\beta, Y_i, X_i) &= \log(f(Y_i|X_i, \beta)) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(Y_i - \beta^t X_i)^2}{2\sigma^2} \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_i^2 - 2Y_i\beta^t X_i + \beta^t X_i X_i^t \beta) \end{aligned} \quad (2.18)$$

$$s_{X_i}(\beta, Y_i, X_i) = \frac{\delta}{\delta\beta} \ell_{X_i}(\beta, Y_i, X_i) = -\frac{1}{2\sigma^2}(-2Y_i X_i + 2X_i X_i^t \beta) \quad (2.19)$$

Der Maximum-Likelihood-Schätzer $\hat{\beta}_{ML}$ ist daher die Lösung der Gleichung $\sum_{i=1}^n (Y_i X_i - X_i X_i^t \beta) = 0$ (Umformen in Matrizen und Auflösen nach β liefert den bekannten kleinste-Quadrate-Schätzer $\hat{\beta}_{ML} = (X^t X)^{-1} X^t Y$, siehe z.B. ebenda, Kapitel 3.2).

Zur Herleitung der korrigierten log-Likelihood wird zunächst der Erwartungswert der naiven log-Likelihood bestimmt:

$$\begin{aligned} \mathbb{E}(\ell_{X_i}(\beta, Y_i, W_i)|Y_i, X_i) &= \mathbb{E}\left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(Y_i - \beta^t W_i)^2}{2\sigma^2} | Y_i, X_i\right) \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_i^2 - 2Y_i\beta^t \mathbb{E}(W_i) + \beta^t \mathbb{E}(W_i W_i^t) \beta) \\ &\stackrel{(2.11)}{=} -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_i^2 - 2Y_i\beta^t X_i + \beta^t (X_i X_i^t + \Sigma_{U_i}) \beta) \\ &= -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_i^2 - 2Y_i\beta^t X_i + \beta^t X_i X_i^t \beta + \beta^t \Sigma_{U_i} \beta) \end{aligned} \quad (2.20)$$

Die korrigierte log-Likelihood ($\mathbb{E}(\ell_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) = \ell_{X_i}(\beta, Y_i, X_i)$) ist somit

$$\ell_{W_i}(\beta, Y_i, W_i) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y_i^2 - 2Y_i\beta^t W_i + \beta^t W_i W_i^t \beta - \beta^t \Sigma_{U_i} \beta) \quad (2.21)$$

und die korrigierte Scorefunktion ($\mathbb{E}(s_{W_i}(\beta, Y_i, W_i)|Y_i, X_i) = s_{X_i}(\beta, Y_i, X_i)$) lautet

$$\begin{aligned} s_{W_i}(\beta, Y_i, W_i) &= \frac{\delta}{\delta\beta} \ell_{W_i}(\beta, Y_i, W_i) = -\frac{1}{2\sigma^2} (-2Y_i W_i + 2W_i W_i^t \beta - 2\Sigma_{U_i} \beta) \\ &= \frac{1}{\sigma^2} (Y_i W_i - W_i W_i^t \beta + \Sigma_{U_i} \beta) \end{aligned} \quad (2.22)$$

(siehe auch Nakamura (1990)). Der korrigierte Schätzer $\hat{\beta}_{cor}$ ist somit die Lösung der Gleichung $\sum_{i=1}^n (Y_i W_i - W_i W_i^t \beta + \Sigma_{U_i} \beta) = 0$. Es kann auch eine geschlossene Form für $\hat{\beta}_{cor}$ hergeleitet werden (siehe so ähnlich ebenda):

$$\begin{aligned} \sum_{i=1}^n (Y_i W_i - W_i W_i^t \beta + \Sigma_{U_i} \beta) &= \sum_{i=1}^n Y_i W_i - \left(\sum_{i=1}^n W_i W_i^t - \sum_{i=1}^n \Sigma_{U_i} \right) \beta \\ &= W^t Y - (W^t W - \sum_{i=1}^n \Sigma_{U_i}) \beta = 0 \\ \Leftrightarrow W^t Y &= (W^t W - \sum_{i=1}^n \Sigma_{U_i}) \beta \\ \Leftrightarrow \hat{\beta}_{cor} &= (W^t W - \sum_{i=1}^n \Sigma_{U_i})^{-1} W^t Y = (W^t W - \sum_{i=1}^n \Sigma_{U_i})^{-1} W^t W \hat{\beta}_{naiv} \end{aligned} \quad (2.23)$$

mit $\hat{\beta}_{naiv} = (W^t W)^{-1} W^t Y$ und unter der Voraussetzung, dass die Matrix $(W^t W - \sum_{i=1}^n \Sigma_{U_i})$ invertierbar ist. Bei der Korrektur wird lediglich Σ_{U_i} , die Kovarianz der Messfehler, verwendet. Die Verteilung der U_i ist im linearen Modell somit (anders als bei den Modellen in den folgenden Abschnitten) irrelevant.

2.2.3. Poisson-Regression

Seien die Y_i unabhängig Poisson-verteilt mit Erwartungswert und Varianz $\lambda_i = \exp(\beta^t X_i)$ (d.h. $Y_i|X_i \sim \text{Po}(\exp(\beta^t X_i))$, $i = 1, \dots, n$). Die Wahrscheinlichkeitsfunktion von Y_i ist (Fahrmeir et al. 2013, B.1):

$$f(Y_i|X_i, \beta) = \frac{\exp(\beta^t X_i)^{Y_i}}{Y_i!} \exp(-\exp(\beta^t X_i)) \quad (2.24)$$

wodurch die log-Likelihood und die Scorefunktion wie folgt lauten:

$$\ell_{X_i}(\beta, Y_i, X_i) = \log(f(Y_i|X_i, \beta)) = Y_i \beta^t X_i - \log(Y_i!) - \exp(\beta^t X_i) \quad (2.25)$$

$$s_{X_i}(\beta, Y_i, X_i) = \frac{\delta}{\delta\beta} \ell_{X_i}(\beta, Y_i, X_i) = Y_i X_i - \exp(\beta^t X_i) X_i \quad (2.26)$$

Der Maximum-Likelihood-Schätzer ist folglich die Lösung der Gleichung $\sum_{i=1}^n (Y_i X_i - \exp(\beta^t X_i) X_i) = \sum_{i=1}^n X_i (Y_i - \exp(\beta^t X_i)) = 0$

Zur Bestimmung der korrigierten Scorefunktion wird zunächst die korrigierte log-Likelihood aufgestellt:

$$\ell_{W_i}(\beta, Y_i, W_i) = Y_i \beta^t W_i - \log(Y_i!) - \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \quad (2.27)$$

Unter Verwendung von Gleichung (2.13) kann die Erwartungstreue gezeigt werden:

$$\begin{aligned} \mathbb{E}(\ell_{W_i}(\beta, Y_i, W_i) | Y_i, X_i) &= Y_i \beta^t X_i - \log(Y_i!) - \frac{\exp(\beta^t X_i) * M_{U_i}(\beta)}{M_{U_i}(\beta)} \\ &= Y_i \beta^t X_i - \log(Y_i!) - \exp(\beta^t X_i) = \ell_{X_i}(\beta, Y_i, X_i) \end{aligned} \quad (2.28)$$

Die korrigierte Scorefunktion lautet dann

$$\begin{aligned} s_{W_i}(\beta, Y_i, W_i) &= \frac{\delta}{\delta\beta} \ell_{W_i}(\beta, Y_i, W_i) = Y_i W_i - \frac{\delta}{\delta\beta} \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \\ &\stackrel{(2.14)}{=} Y_i W_i - \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \left(W_i - \frac{\delta}{\delta\beta} \log(M_{U_i}(\beta)) \right) \end{aligned} \quad (2.29)$$

Die Nullstelle von $\sum_{i=1}^n s_{W_i}(\beta, Y_i, W_i) = 0$ ist der korrigierte Schätzer $\hat{\beta}_{cor}$. Bei normalverteilten Messfehlern ergeben sich durch Einsetzen der Momenterzeugenden Funktion der multivariaten Normalverteilung (siehe (2.15)) die selben Formeln für die korrigierte log-Likelihood und Scorefunktion wie bei Nakamura (1990). Im Gegensatz zum linearen Modell existiert keine geschlossene Form für $\hat{\beta}_{cor}$. In Abschnitt 4.1 wird die *Newton-Raphson-Methode* zur Bestimmung der Nullstelle von $\sum_{i=1}^n s_{W_i}(\beta, Y_i, W_i)$ kurz vorgestellt.

Bei der Korrektur wird hier die Momenterzeugende Funktion benötigt. Anders als im linearen Regressionsmodell ergeben sich somit unterschiedliche korrigierte Scorefunktionen, wenn die Messfehler anstatt einer Normalverteilung z.B. einer Laplace-Verteilung folgen. In Kapitel 4 werden die Ergebnisse von Simulationen mit Poisson-verteilter Zielvariable und Messfehlern aus verschiedenen Verteilungen vorgestellt. Dabei wird auch untersucht, welche Auswirkungen es hat, wenn für die Messfehler fälschlicherweise eine Normalverteilung anstelle der wahren Verteilung angenommen wird.

Bei der Simulation wird ein Modell mit einer Kovariable und Intercept betrachtet, d.h. $Y_i|X_i \sim \text{Po}(\exp(\beta_0 + \beta_1 X_i))$; $X_i, U_i, W_i, \beta_0, \beta_1 \in \mathbb{R}$; $W_i = X_i + U_i$. In diesem Fall lauten die wahren und die korrigierten Scorefunktionen von β_0 und β_1 wie folgt:

$$s_{X_i}(\beta_0, Y_i, X_i) = Y_i - \exp(\beta_0 + \beta_1 X_i) \quad (2.30)$$

$$s_{X_i}(\beta_1, Y_i, X_i) = Y_i X_i - \exp(\beta_0 + \beta_1 X_i) X_i \quad (2.31)$$

$$s_{W_i}(\beta_0, Y_i, W_i) = Y_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} \quad (2.32)$$

$$\begin{aligned} s_{W_i}(\beta_1, Y_i, W_i) &= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} \left(W_i - \frac{\delta}{\delta \beta_1} \log(M_{U_i}(\beta_1)) \right) \\ &= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} \left(W_i - \frac{\frac{\delta}{\delta \beta_1} M_{U_i}(\beta_1)}{M_{U_i}(\beta_1)} \right) \end{aligned} \quad (2.33)$$

Dies ergibt sich durch Auswerten von $s_{\tilde{X}_i}(\beta, Y_i, \tilde{X}_i)$ (2.26) und $s_{\tilde{W}_i}(\beta, Y_i, \tilde{W}_i)$ (2.29) mit $\beta = (\beta_0, \beta_1)^t$, $\tilde{X}_i = (1, X_i)^t$, $\tilde{U}_i = (0, U_i)^t$ und $\tilde{W}_i = \tilde{X}_i + \tilde{U}_i = (1, W_i)^t$.

Der korrigierte Schätzer $\hat{\beta}_{cor} = (\hat{\beta}_{cor,0}, \hat{\beta}_{cor,1})^t$ ist dann die Lösung von

$$s_W(\beta, Y, W) = \begin{pmatrix} \sum_{i=1}^n s_{W_i}(\beta_0, Y_i, W_i) \\ \sum_{i=1}^n s_{W_i}(\beta_1, Y_i, W_i) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.34)$$

2.2.4. Gamma-Regression

Nun wird ein Regressionsmodell mit Gamma-verteilter Zielvariable betrachtet. Als Responsefunktion wird $\mu_i = \exp(-\beta^t X_i)$ verwendet⁸. Die Dichte der Y_i ist dann

$$\begin{aligned} f(Y_i|X_i, \beta) &= \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\exp(-\beta^t X_i)} \right)^\nu Y_i^{\nu-1} \exp \left(-\frac{\nu}{\exp(-\beta^t X_i)} Y_i \right) \\ &= \frac{1}{\Gamma(\nu)} (\nu * \exp(\beta^t X_i))^\nu Y_i^{\nu-1} \exp(-\nu * \exp(\beta^t X_i) * Y_i) \end{aligned} \quad (2.35)$$

⁸Das Minus vor dem linearen Prädiktor $\eta_i = \beta^t X_i$ erleichtert das Vorgehen bei der Herleitung der korrigierten log-Likelihood und Scorefunktion, weswegen nicht die typischere Responsefunktion $\mu_i = \exp(\beta^t X_i)$ verwendet wird.

mit Skalenparameter $\nu > 0$ (Fahrmeir et al. 2013, B.1). Für die Varianz von Y_i gilt $\text{Var}(Y_i) = \frac{\exp(-\beta^t X_i)^2}{\nu}$ (ebenda, Kapitel 5.3). Die log-Likelihood und die Scorefunktion dieses Modells lauten

$$\begin{aligned} \ell_{X_i}(\beta, Y_i, X_i) &= \log(f(Y_i|X_i, \beta)) \\ &= -\log(\Gamma(\nu)) + \nu(\log(\nu) + \beta^t X_i) + (\nu - 1)\log(Y_i) - \nu * \exp(\beta^t X_i) * Y_i \end{aligned} \quad (2.36)$$

$$s_{X_i}(\beta, Y_i, X_i) = \frac{\delta}{\delta\beta} \ell_{X_i}(\beta, Y_i, X_i) = \nu * X_i - \nu * Y_i * \exp(\beta^t X_i) * X_i \quad (2.37)$$

Der ML-Schätzer $\hat{\beta}_{ML}$ ist somit die Nullstelle von $\sum_{i=1}^n X_i - Y_i * \exp(\beta^t X_i) * X_i$. Die korrigierte log-Likelihood ist (unter Verwendung von (2.13))

$$\begin{aligned} \ell_{W_i}(\beta, Y_i, W_i) \\ &= -\log(\Gamma(\nu)) + \nu(\log(\nu) + \beta^t W_i) + (\nu - 1)\log(Y_i) - \nu \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} Y_i \end{aligned} \quad (2.38)$$

und die korrigierte Scorefunktion ergibt sich zu

$$\begin{aligned} s_{W_i}(\beta, Y_i, W_i) &= \frac{\delta}{\delta\beta} \ell_{W_i}(\beta, Y_i, W_i) = \nu * W_i - \nu * Y_i \frac{\delta}{\delta\beta} \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \\ &\stackrel{(2.14)}{=} \nu * W_i - \nu * Y_i \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \left(W_i - \frac{\delta}{\delta\beta} \log(M_{U_i}(\beta)) \right) \\ &= \nu * \left(W_i - Y_i \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \left(W_i - \frac{\delta}{\delta\beta} \log(M_{U_i}(\beta)) \right) \right) \end{aligned} \quad (2.39)$$

Der korrigierte Schätzer $\hat{\beta}_{cor}$ ist somit die Lösung der Gleichung $\sum_{i=1}^n \left(W_i - Y_i \frac{\exp(\beta^t W_i)}{M_{U_i}(\beta)} \left(W_i - \frac{\delta}{\delta\beta} \log(M_{U_i}(\beta)) \right) \right) = 0$. Wie bei der Poisson-Verteilung ist die Korrektur von der Messfehlerverteilung abhängig und es kann keine geschlossene Form für $\hat{\beta}_{cor}$ hergeleitet werden.

2.2.5. Logistische Regression

Im logistischen Regressionsmodell ist die Zielvariable Bernoulli-verteilt, d.h. $Y_i|X_i \sim B(1, \pi_i)$, wobei für den Erwartungswert π_i folgendes gilt:

$$\pi_i = \mathbb{E}(Y_i|X_i) = P(Y_i = 1|X_i) = \frac{1}{1 + \exp(-\beta^t X_i)} = \frac{\exp(\beta^t X_i)}{1 + \exp(\beta^t X_i)} \quad (2.40)$$

(Fahrmeir et al. 2013, Kapitel 5.1.1). Die Scorefunktion lautet

$$s_{X_i}(\beta, Y_i, X_i) = X_i (Y_i - \pi_i) = X_i \left(Y_i - \frac{1}{1 + \exp(-\beta^t X_i)} \right) \quad (2.41)$$

(siehe z.B. ebenda, Kapitel 5.1.2).

Anders als bei den vorangegangenen Modellen existiert bei der logistischen Regression im Allgemeinen keine exakte korrigierte Scorefunktion (Stefanski 1989, Abschnitte 2.1 und 3.1): Die zu schätzende Funktion (d.h. hier die wahre Scorefunktion) muss, aufgefasst als Funktion einer komplexen Variable, eine ganze Funktion sein. Ist dies nicht der Fall, kann es keine Schätzfunktion geben, die für jedes β erwartungstreu ist (d.h. es existiert keine korrigierte Scorefunktion)⁹. Da die Verteilungsfunktion der logistischen Verteilung ($F(z) = (1 + \exp(-z))^{-1}$) in der komplexen Ebene Singularitäten an Vielfachen von $i\pi$ hat, ist die Scorefunktion im Logit-Modell keine ganze Funktion (ebenda). Selbiges gilt auch für die Scorefunktion der Probit-Regression (Buzas 2009).

Für den Spezialfall, dass die Zielvariable Y ein seltenes Ereignis darstellt (d.h. $\sup(P(Y_i = 1|X_i, \beta)) < 0.5$), kann nach Buzas und Stefanski (1996) eine korrigierte Scorefunktion hergeleitet werden. Aus der Annahme $\sup(P(Y_i = 1|X_i, \beta)) < 0.5$ folgt $\exp(\beta^t X_i) < 1$ für alle X_i , wodurch sich die logistische Verteilungsfunktion als Reihenentwicklung darstellen lässt, auf deren Basis eine exakte Korrektur möglich ist. Nähere Details dazu befinden sich im Anhang A.2.

Im allgemeinen Fall gibt es trotz Nichtexistenz einer exakten korrigierten Scorefunktion Möglichkeiten zur (approximativen) Korrektur:

Buzas (2009) konstruiert zunächst durch Approximation der logistischen Verteilungsfunktion durch die Verteilungsfunktion der Normalverteilung eine approximative Scorefunktion, die in Abwesenheit von Messfehlern konsistente Schätzungen liefert. Dann wird diese approximative Scorefunktion korrigiert, d.h. es wird eine Funktion hergeleitet, die erwartungstreu für die approximative Scorefunktion ist (unter der Annahme von normalverteilten Messfehlern). Dieses Vorgehen wird ebenfalls im Anhang A.2 ausführlicher erläutert.

Novick und Stefanski (2002) untersuchen die Anwendung der Monte Carlo Corrected Score Methode bei der logistischen Regression. Die Lösung ist nur approximativ, liefert in der Praxis aber häufig gute Ergebnisse (Carroll et al.

⁹Stefanski bezieht sich dabei auf additive, normalverteilte Messfehler.

2006, Kapitel 7.2).

Mit der *Monte Carlo Corrected Score Methode* können allgemein korrigierte Scorefunktionen konstruiert werden, was vor allem dann hilfreich ist, wenn sich exakte Resultate nicht so direkt herleiten lassen wie in 2.2.2, 2.2.3 und 2.2.4. Das Prinzip basiert auf komplexen Zufallsvariablen und auf der Annahme von additiven, normalverteilten Messfehlern. Es wird im Anhang A.1 genauer beschrieben, wobei auch dargelegt wird, weshalb sich das Vorgehen nicht ohne Weiteres auf nichtnormalverteilte Messfehler übertragen lässt.

3. Mögliche Verteilungen der Messfehler

In Abschnitt 2.2.1 wurden bestimmte Annahmen¹⁰ über die Messfehler getroffen. In diesem Kapitel werden verschiedene Verteilungen vorgestellt, denen die Messfehler folgen könnten und für die diese Annahmen erfüllt sind. Zunächst wird in der schiefen Normalverteilung eine Modifikation der Normalverteilung betrachtet, durch die geneigte Verteilungen modelliert werden können. Im zweiten Abschnitt wird mit der Laplace-Verteilung ein Fall untersucht, bei dem die Momenterzeugende Funktion nur auf einem gewissen Intervall existiert. Hierbei wird auch auf trunkierte Verteilungen eingegangen. Zuletzt werden die sehr flexiblen Mischverteilungen genauer betrachtet.

In Kapitel werden Messfehler aus den hier vorgestellten Verteilungen in einem Modell mit Poisson-verteilter Zielvariable und einer Kovariable simuliert. Die exakten korrigierten Scorefunktionen für dieses Modell und die verschiedenen Messfehlerverteilungen werden bereits in diesem Kapitel im jeweiligen Abschnitt aufgestellt.

3.1. Schiefe Normalverteilung

3.1.1. Definition, Momente und Momenterzeugende Funktion

Eine Normalverteilung ist stets symmetrisch um ihren Erwartungswert μ . Die *schiefe Normalverteilung* ist eine von Azzalini (2013) vorgestellte Erweiterung der Normalverteilung, mit der sich Abweichungen von der Symmetrie modellieren lassen. Zunächst wird die univariate schiefe Normalverteilung betrachtet.

Seien $\phi(x)$ und $\Phi(x)$ die Dichte- und die Verteilungsfunktion der Standardnormalverteilung und $\alpha \in \mathbb{R}$. Dann ist folgende Funktion eine Dichtefunktion:

$$f(x; \alpha) = 2\phi(x)\Phi(\alpha x) \quad (x \in \mathbb{R}) \quad (3.1)$$

(ebenda, Kapitel 2.1.1). α ist der Neigungsparameter (*slant*-Parameter bei Azzalini). Für positive Werte von α ergibt sich eine rechtsschiefe, für negative Werte eine linksschiefe Verteilung. Je größer der Betrag von α , umso stärker die Neigung (siehe Abb. 1. Für negative α ergeben sich die selben Kurven, gespiegelt an $x = 0$). Für $\alpha = 0$ ist $f(x)$ die Dichtefunktion der Standardnormalverteilung.

¹⁰Relevant ist hier die Voraussetzung $\mathbb{E}(U_i) = 0$ sowie die Existenz der Momenterzeugenden Funktion.

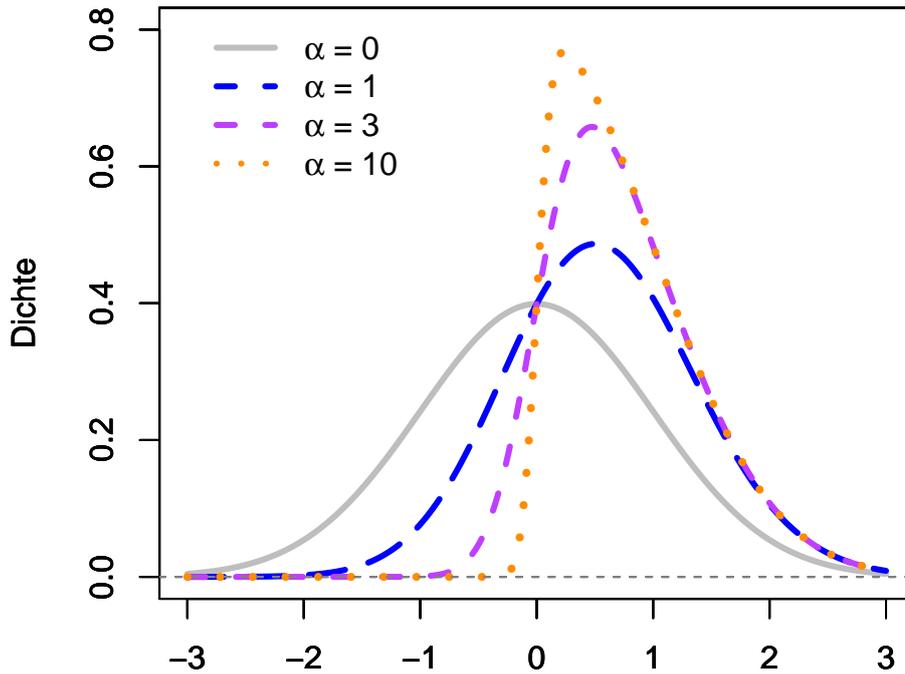


Abb. 1: Schiefe Normalverteilung: Dichte $f(x; \alpha)$ für verschiedene Werte von α

Sei nun \tilde{U} eine Zufallsvariable mit Dichtefunktion $f(x; \alpha)$. Dann folgt die Zufallsvariable U mit

$$U = \xi + \omega \tilde{U} \quad \text{mit } \xi \in \mathbb{R}, \omega \in \mathbb{R}^+ \quad (3.2)$$

einer schiefen Normalverteilung mit Neigungsparameter α (*slant*), Lageparameter ξ (*location*) und Skalenparameter ω (*scale*), kurz $U \sim SN(\xi, \omega^2, \alpha)$ ¹¹ (Azzalini 2013, Kapitel 2.1.1). Die Dichte von U lautet:

$$f_U(u; \alpha, \xi, \omega) = \frac{2}{\omega} \phi\left(\frac{u - \xi}{\omega}\right) \Phi\left(\alpha \frac{u - \xi}{\omega}\right) \quad (u \in \mathbb{R}) \quad (3.3)$$

Für den Erwartungswert und die Varianz von U gilt (ebenda, Kapitel 2.1.4):

$$\mathbb{E}(U) = \xi + \omega \mu_{\tilde{U}} = \xi + \omega \sqrt{2/\pi} \delta \quad (3.4)$$

$$\text{Var}(U) = \omega^2 \sigma_{\tilde{U}}^2 = \omega^2 (1 - \mu_{\tilde{U}}^2) = \omega^2 (1 - 2/\pi \delta^2) \quad (3.5)$$

$$\text{mit } \delta = \frac{\alpha}{\sqrt{1 + \alpha^2}} \quad (\delta \in (-1, 1))$$

¹¹ ω^2 statt ω in Analogie zur Normalverteilung $N(\mu, \sigma^2)$. Für $\alpha = 0$ ist $U \sim N(\xi, \omega^2)$.

Die Momenterzeugende Funktion von U lautet (Azzalini 2013, Kapitel 2.1.2):

$$M_U(\beta) = \mathbb{E}(\exp(\beta U)) = 2\exp(\xi\beta + \frac{1}{2}\omega^2\beta^2)\Phi(\delta\omega\beta) \quad (3.6)$$

Sie existiert für alle $\beta \in \mathbb{R}$. Wenn $\alpha = 0$, ist $U \sim N(\xi, \omega^2)$; für $M_U(\beta)$ ergibt sich dementsprechend die Momenterzeugende Funktion einer Normalverteilung ($\exp(\xi\beta + \frac{1}{2}\omega^2\beta^2)$), da $\delta = 0$ und $\Phi(0) = 0.5$.

Analog zum univariaten Fall ist die *multivariate schiefe Normalverteilung* eine Erweiterung der multivariaten Normalverteilung. Im p -dimensionalen Fall ist die Dichte von U :

$$f_U(u; \alpha) = 2\phi_p(u; \tilde{\Omega})\Phi(\alpha^t u), \quad u, \alpha \in \mathbb{R}^p \quad (3.7)$$

(ebenda, Kapitel 5.1.1); $\phi_p(u; \tilde{\Omega})$ bezeichnet die Dichte einer $N_p(0, \tilde{\Omega})$ -Verteilung. Dann ist $U = \xi + \omega\tilde{U}$ multivariat schief normalverteilt ($U \sim SN_p(\xi, \Omega, \alpha)$), wobei $\xi \in \mathbb{R}^p$, $\omega = \text{diag}(\omega_1, \dots, \omega_p) > 0$ und $\Omega = \omega\tilde{\Omega}\omega$.

Die Momenterzeugende Funktion existiert auch im mehrdimensionalen Fall ohne Beschränkung des Definitionsbereichs (Formel siehe ebenda, Kapitel 5.1.2, Momente siehe Kapitel 5.1.5). Anders als bei der Normalverteilung gilt bei der schiefen Normalverteilung nicht, dass die gemeinsame Verteilung von unabhängigen, univariaten schief normalverteilten Zufallsvariablen (jeweils mit $\alpha \neq 0$) eine multivariate schiefe Normalverteilung ist (ebenda, Kapitel 5.1.6).

3.1.2. Korrigierte Scorefunktion für die Simulation

Bei der Simulation in Abschnitt 4.3 wird eine Poisson-verteilte Zielvariable mit Erwartungswert $\exp(\beta_0 + \beta_1 X_i)$ betrachtet (siehe Abschnitt 2.2.3). Wenn die eindimensionalen Messfehler U_i einer schiefen Normalverteilung folgen ($U_i \sim SN(\xi, \omega^2, \alpha)$ für $i = 1, \dots, n$), lauten die korrigierten Scorefunktionen für β_0 und β_1 (vgl. Gleichungen 2.32 und 2.33):

$$\begin{aligned} s_{W_i}(\beta_0, Y_i, W_i) &= Y_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} = Y_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{2\exp(\xi\beta_1 + \frac{1}{2}\omega^2\beta_1^2)\Phi(\delta\omega\beta_1)} \\ &= Y_i - \frac{\exp(\beta_0 + \beta_1 W_i - \xi\beta_1 - \frac{1}{2}\omega^2\beta_1^2)}{2\Phi(\delta\omega\beta_1)} \end{aligned} \quad (3.8)$$

$$\begin{aligned}
s_{W_i}(\beta_1, Y_i, W_i) &= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} \left(W_i - \frac{\frac{\delta}{\delta \beta_1} M_{U_i}(\beta_1)}{M_{U_i}(\beta_1)} \right) \\
&= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i - \xi \beta_1 - \frac{1}{2} \omega^2 \beta_1^2)}{2\Phi(\delta \omega \beta_1)} \left(W_i - \frac{\frac{\delta}{\delta \beta_1} 2 \exp(\xi \beta_1 + \frac{1}{2} \omega^2 \beta_1^2) \Phi(\delta \omega \beta_1)}{2 \exp(\xi \beta_1 + \frac{1}{2} \omega^2 \beta_1^2) \Phi(\delta \omega \beta_1)} \right) \\
&= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i - \xi \beta_1 - \frac{1}{2} \omega^2 \beta_1^2)}{2\Phi(\delta \omega \beta_1)} * \\
&\quad \left(W_i - \frac{(\xi + \omega^2 \beta_1) 2 \exp(\xi \beta_1 + \frac{1}{2} \omega^2 \beta_1^2) \Phi(\delta \omega \beta_1) + 2 \exp(\xi \beta_1 + \frac{1}{2} \omega^2 \beta_1^2) \delta \omega * \phi(\delta \omega \beta_1)}{2 \exp(\xi \beta_1 + \frac{1}{2} \omega^2 \beta_1^2) \Phi(\delta \omega \beta_1)} \right) \\
&= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i - \xi \beta_1 - \frac{1}{2} \omega^2 \beta_1^2)}{2\Phi(\delta \omega \beta_1)} \left(W_i - (\xi + \omega^2 \beta_1) - \frac{\delta \omega * \phi(\delta \omega \beta_1)}{\Phi(\delta \omega \beta_1)} \right)
\end{aligned} \tag{3.9}$$

Bei der Simulation wird α und die Varianz der U_i übergeben und daraus ω und ξ bestimmt. Umformen von Gleichung (3.5) nach ω liefert:

$$\omega = \sqrt{\frac{\text{Var}(U_i)}{1 - 2/\pi \delta^2}} \quad \left(\text{mit } \delta = \frac{\alpha}{\sqrt{1 + \alpha^2}} \right) \tag{3.10}$$

Aus der Annahme $\mathbb{E}(U_i) = 0$ folgt dann für ξ (Umstellen von (3.4)):

$$\xi = -\omega \sqrt{2/\pi} \delta \tag{3.11}$$

Abbildung 2 zeigt die daraus resultierenden Dichten für $\alpha = 3$ und $\alpha = 10$ bei einer Messfehlervarianz von 1. Für $\alpha = 3$ ergibt sich beispielsweise $\delta \approx 0,95$, $\omega \approx 1,53$ und $\xi \approx -1,16$. Für $\alpha = 10$ gilt $\delta \approx 0,995$, $\omega \approx 1,64$ und $\xi \approx -1,31$. Zum Vergleich ist in grau die Dichte der Standardnormalverteilung dargestellt, diese ergibt sich für $\alpha = 0$.

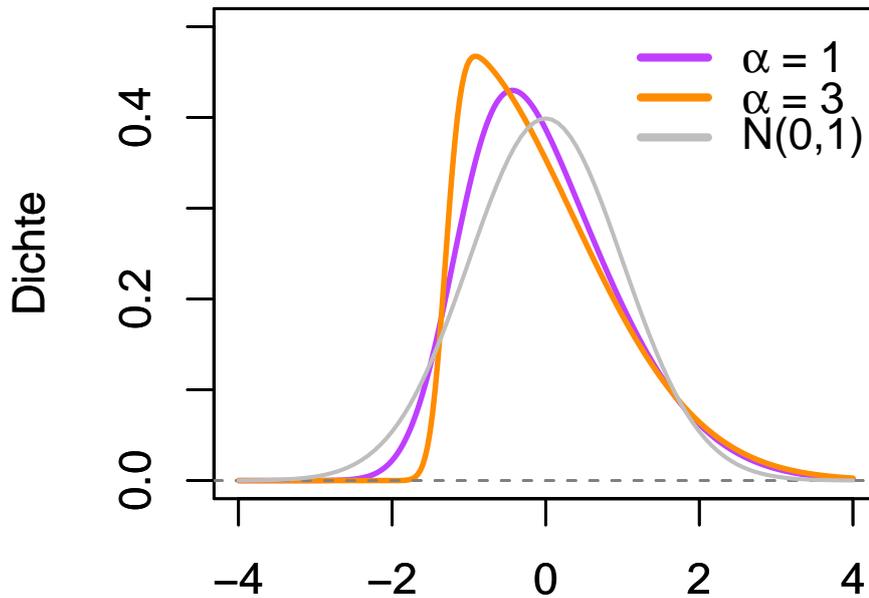


Abb. 2: Dichte der schiefen Normalverteilung mit Varianz 1 und Erwartungswert 0 für verschiedene Werte von α

3.2. Laplace-Verteilung

3.2.1. Definition, Momente und Momenterzeugende Funktion

Die Laplace-Verteilung (auch Doppelsexponential- oder zweiseitige Exponentialverteilung genannt) ist eine weitere mögliche Verteilung, aus der Messfehler stammen könnten. Die Dichte einer Laplace-verteilten Zufallsvariable U lautet

$$f_U(u; \mu, s) = \frac{1}{2s} \exp\left(-\frac{|u - \mu|}{s}\right) \quad (u \in \mathbb{R}) \quad (3.12)$$

(siehe z.B. Fahrmeir et al. 2013, B.1) mit Lageparameter $\mu \in \mathbb{R}$ und Skalenparameter $s \in \mathbb{R}^+$. Für Erwartungswert und Varianz gelten

$$\mathbb{E}(U) = \mu \quad (3.13)$$

$$\text{Var}(U) = 2s^2 \quad \left(\Leftrightarrow s = \sqrt{\frac{\text{Var}(U)}{2}} \right) \quad (3.14)$$

(siehe z.B. ebenda, B.1). Bei der Simulation in Abschnitt 4.4 wird μ immer auf 0 gesetzt (da $\mathbb{E}(U_i) = 0$). Für eine gegebene Varianz $\text{Var}(U_i)$ lässt sich der zugehörige Skalenparameter durch die Formel in (3.14) berechnen. Abbildung 3 zeigt

die Dichten von zwei Laplace-Verteilungen mit $\mu = 0$ und Varianz 2, d.h. $s = 1$ (Standard-Laplace-Verteilung) und Varianz 1, d.h. $s = \sqrt{0,5} \approx 0,71$. Zum Vergleich ist in grau die Dichte der Standardnormalverteilung dargestellt.

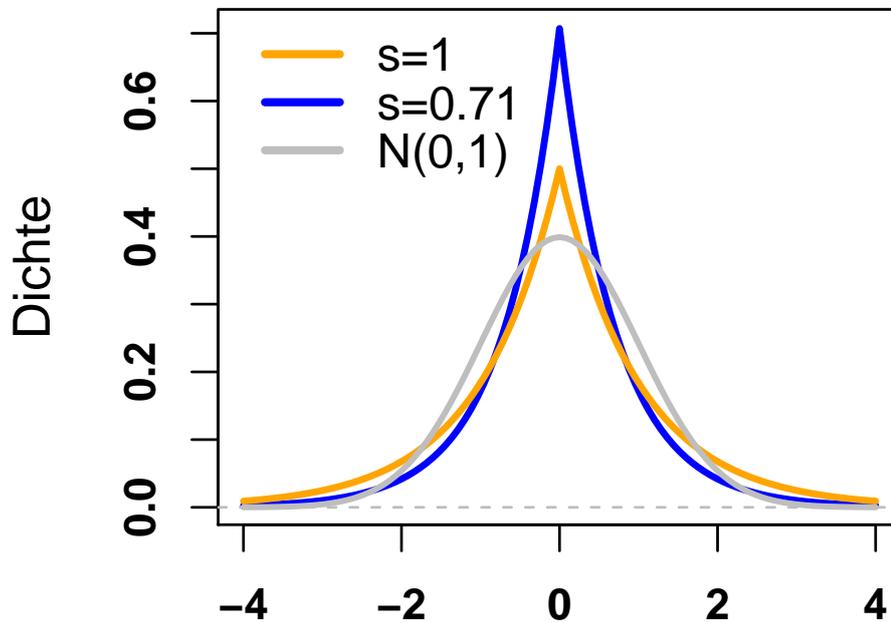


Abb. 3: Laplace-Verteilungen mit $\mu = 0$ und Varianz = 1 (d.h. $s \approx 0,71$) sowie Varianz = 2 ($s = 1$)

Die Standard-Laplace-Verteilung hat im Vergleich zur Standardnormalverteilung mehr Wahrscheinlichkeitsmasse bei sehr kleinen und sehr großen Werten (deswegen eignet sie sich gut für Verteilungen mit Ausreißern) und als Ausgleich in den mittleren Bereichen eine geringere Dichte. Die Kurtosis¹² der Laplace-Verteilung beträgt 6 (d.h. Exzess-Wölbung 3), sie hat also einen steileren Gipfel als die Normalverteilung (Kotz et al. 2001, Kapitel 2.1.3.).

Die Momenterzeugende Funktion der Laplace-Verteilung lautet

$$M_U(\beta) = \mathbb{E}(\exp(\beta U)) = \frac{\exp(\beta\mu)}{1 - s^2\beta^2} \quad \text{für } -\frac{1}{s} < \beta < \frac{1}{s} \quad (3.15)$$

(ebenda, Kapitel 2.1.2, Herleitung auch im Anhang A.3). Im hier betrachteten Fall mit $\mu = 0$ ergibt sich $M_U(\beta) = (1 - s^2\beta^2)^{-1}$ für $|\beta| < 1/s$. $M_U(\beta)$

¹²Kurtosis bzw. Wölbung = $\frac{\mathbb{E}((U - \mathbb{E}(U))^4)}{\mathbb{E}((U - \mathbb{E}(U))^2)^2}$. Die Kurtosis der Normalverteilung beträgt 3, die Exzess-Kurtosis ist definiert als Kurtosis - 3 (siehe z.B. Boos und Stefanski 2013, Kapitel 1.5.2).

existiert somit nur auf einem Intervall, dessen Größe vom Skalenparameter s abhängig ist: Je größer s , d.h. je größer die Messfehlervarianz, umso kleiner ist der Definitionsbereich von $M_U(\beta)$. Für $|\beta| \geq 1/s$ ist $\mathbb{E}(\exp(\beta U))$ unendlich (siehe Anhang A.3). Für die korrigierte Scorefunktion muss somit (falls der Term $\exp(\beta W_i) = \exp(\beta X_i)\exp(\beta U_i)$ in der naiven Scorefunktion auftaucht) eine Restriktion an den Parameterraum gestellt werden: Nur für $|\beta| < 1/s$ kann die naive Scorefunktion durch Division mit $M_{U_i}(\beta)$ korrigiert werden. In der Praxis ist diese Einschränkung womöglich verkraftbar, wenn die Größenordnung von β bekannt ist.

Mögliche Erweiterungen bzw. Generalisierungen der Laplace-Verteilung sind die asymmetrische sowie die multivariate Laplace-Verteilung (Kotz et al. 2001, Kapitel 3 und 5). Die asymmetrische Laplace-Verteilung lässt Schiefe zu. Auch hier existiert die Momenterzeugende Funktion nur auf einem (hier asymmetrischen) Intervall (dessen Grenzen neben dem Skalenparameter auch von der Schiefe abhängen) (ebenda, Kapitel 3.1.4). Die multivariate Laplace-Verteilung wird über eine Generalisierung der charakteristischen Funktion hergeleitet (ebenda, Kapitel 5.2). Wie bei der multivariaten schiefen Normalverteilung gilt auch hier nicht, dass die gemeinsame Verteilung von unabhängigen, Laplace-verteilten Zufallsvariablen einer multivariaten Laplace-Verteilung entspricht (ebenda, Kapitel 5.1).

3.2.2. Korrigierte Scorefunktion für die Simulation

Die korrigierten Scorefunktionen für β_0 und β_1 bei der Simulation mit $Y_i|X_i \sim \text{Po}(\exp(\beta_0 + \beta_1 X_i))$, $W_i = X_i + U_i$ (siehe Abschnitt 2.2.3) und $U_i \sim \text{Laplace}(0, s)$ lauten (vgl. Gleichungen 2.32 und 2.33) mit der Restriktion $|\beta_1| < \frac{1}{s}$:

$$s_{W_i}(\beta_0, Y_i, W_i) = Y_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} = Y_i - \exp(\beta_0 + \beta_1 W_i)(1 - s^2 \beta_1^2) \quad (3.16)$$

$$\begin{aligned} s_{W_i}(\beta_1, Y_i, W_i) &= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} \left(W_i - \frac{\frac{\delta}{\delta \beta_1} M_{U_i}(\beta_1)}{M_{U_i}(\beta_1)} \right) \\ &= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{M_{U_i}(\beta_1)} \left(W_i - (M_{U_i}(\beta_1))^{-1} * \frac{\delta}{\delta \beta_1} M_{U_i}(\beta_1) \right) \\ &= Y_i W_i - \frac{\exp(\beta_0 + \beta_1 W_i)}{(1 - s^2 \beta_1^2)^{-1}} \left(W_i - (1 - s^2 \beta_1^2) \left(-\frac{-2s^2 \beta_1}{(1 - s^2 \beta_1^2)^2} \right) \right) \end{aligned}$$

$$\begin{aligned}
&= Y_i W_i - \exp(\beta_0 + \beta_1 W_i) (1 - s^2 \beta_1^2) \left(W_i - \frac{2s^2 \beta_1}{1 - s^2 \beta_1^2} \right) \\
&= Y_i W_i - \exp(\beta_0 + \beta_1 W_i) \left((1 - s^2 \beta_1^2) W_i - 2s^2 \beta_1 \right) \quad (3.17)
\end{aligned}$$

3.2.3. Trunkierte Verteilung

Um eine Momenterzeugende Funktion ohne Restriktion des Parameterraums zu erhalten, kann eine trunkierte Laplace-Verteilung betrachtet werden: Unter der Annahme, dass die Messfehler aus einer trunkierten Laplace-Verteilung stammen, ist $\mathbb{E}(\exp(\beta U))$ für alle $\beta \in \mathbb{R}$ endlich. Auf der anderen Seite ist die entstehende Momenterzeugende Funktion allerdings deutlich komplexer als die einer untrunkierten Laplace-Verteilung.

Seien f_X bzw. F_X die Dichte bzw. die Verteilungsfunktion einer Zufallsvariablen X . Wenn von den Realisierungen von X all diejenigen aussortiert werden, die $\leq a$ oder $> b$ sind ($a, b \in \mathbb{R}, a < b, a = -\infty$ oder $b = \infty$ möglich), folgen die übrig gebliebenen Beobachtungen einer sogenannten *trunkierten Verteilung* (siehe z.B. Cramer 1946, Kapitel 19.3). Die Verteilungsfunktion und Dichte einer solchen trunkierten Variable X_T lauten (ebenda, Kapitel 19.3):

$$F_{X_T}(x) = F_X(x|a < X \leq b) = \begin{cases} 0 & \text{für } x \leq a \\ \frac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & \text{für } x \in (a, b] \\ 1 & \text{für } x > b \end{cases} \quad (3.18)$$

$$f_{X_T}(x) = f_X(x|a < X \leq b) = \frac{f_X(x)}{\int_a^b f_X(t) dt} = \frac{f_X(x)}{F_X(b) - F_X(a)} \quad (3.19)$$

für $x \in (a, b]$, 0 sonst.

Bei Messfehlern ist eine Trunkierung womöglich inhaltlich durchaus sinnvoll: Während beispielsweise bei einer Laplace-Verteilung in der Theorie jeder noch so (betragsmäßig) große Wert eine positive Dichte hat, können die Fehler in der Praxis nicht beliebig groß sein, wenn die messfehlerbehaftete Variable X bzw. die gemessenen W z.B. nur positive Werte annehmen können¹³.

Nun soll die Momenterzeugende Funktion für eine trunkierte Laplace-Verteilung

¹³Allerdings darf die Verteilung von U_i und somit auch die Trunkierung nicht vom wahren X_i abhängen, siehe Annahmen in Abschnitt 2.2.1

(mit $\mu = 0$) hergeleitet werden. Es wird eine symmetrische Trunkierung betrachtet¹⁴, d.h. $a = -b$ (mit $b > 0$). Zunächst wird die Dichte $f_{U_T}(u)$ aufgestellt. Im Nenner ($F_U(b) - F_U(-b)$) wird die Verteilungsfunktion der Laplace-Verteilung benötigt (Gleichung (3.19)), die wie folgt lautet:

$$F_U(u; \mu, s) = \begin{cases} \frac{1}{2} \exp\left(-\frac{|u - \mu|}{s}\right) & \text{für } u \leq \mu \\ 1 - \frac{1}{2} \exp\left(-\frac{|u - \mu|}{s}\right) & \text{für } u \geq \mu \end{cases}$$

$$\stackrel{\mu=0}{=} \begin{cases} \frac{1}{2} \exp\left(-\frac{|u|}{s}\right) & \text{für } u \leq 0 \\ 1 - \frac{1}{2} \exp\left(-\frac{|u|}{s}\right) & \text{für } u \geq 0 \end{cases} \quad (3.20)$$

(Kotz et al. 2001, Abschnitt 2.1.1). Für $F_U(b) - F_U(-b)$ ergibt sich somit (mit $b > 0$):

$$\begin{aligned} F_U(b) - F_U(-b) &= 1 - \frac{1}{2} \exp\left(-\frac{|b|}{s}\right) - \frac{1}{2} \exp\left(-\frac{|-b|}{s}\right) \\ &= 1 - \frac{1}{2} \exp\left(-\frac{b}{s}\right) - \frac{1}{2} \exp\left(-\frac{b}{s}\right) = 1 - \exp\left(-\frac{b}{s}\right) \end{aligned} \quad (3.21)$$

Die trunkierte Dichte $f_{U_T}(u)$ lautet dann (siehe Gleichung (3.19)):

$$f_{U_T}(u) = \frac{f_U(u)}{F_U(b) - F_U(-b)} = \frac{1}{2s} \exp\left(-\frac{|u|}{s}\right) * \left(1 - \exp\left(-\frac{b}{s}\right)\right)^{-1} \quad (3.22)$$

für $u \in (-b, b]$, 0 sonst. Nun kann die Momenterzeugende Funktion der trunkierten Variable U_T aufgestellt werden:

$$\begin{aligned} M_{U_T}(\beta) &= \mathbb{E}(\exp(\beta U_T)) = \int_{-\infty}^{\infty} \exp(\beta u) f_{U_T}(u) du \quad (3.23) \\ &\stackrel{f_{U_T}(u)=0 \text{ für } u \notin (-b, b]}{=} \int_{-b}^b \exp(\beta u) * 0,5/s * \exp(-|u|/s) * (1 - \exp(-b/s))^{-1} du \\ &= 0,5/s * (1 - \exp(-b/s))^{-1} \int_{-b}^b \exp(\beta u - |u|/s) du \end{aligned}$$

¹⁴Dadurch wird die Symmetrie beibehalten, wodurch der Erwartungswert unverändert bei 0 bleibt.

$$\begin{aligned}
&= \frac{1}{2s * (1 - \exp(-b/s))} \left(\int_{-b}^0 \exp(\beta u - |u|/s) du + \int_0^b \exp(\beta u - |u|/s) du \right) \\
&= \frac{1}{2s * (1 - \exp(-b/s))} \left(\int_{-b}^0 \exp(u(\beta + 1/s)) du + \int_0^b \exp(u(\beta - 1/s)) du \right)
\end{aligned}$$

Für $\beta \neq 1/s$ und $\beta \neq -1/s$ gilt dann:

$$\begin{aligned}
M_{U_T}(\beta) &= \frac{1}{2s * (1 - \exp(-b/s))} \left([(\beta + 1/s)^{-1} \exp(u(\beta + 1/s))]_{u=-b}^{u=0} \right. \\
&\quad \left. + [(\beta - 1/s)^{-1} \exp(u(\beta - 1/s))]_{u=0}^{u=b} \right) \\
&= \frac{1}{2s * (1 - \exp(-b/s))} \left(\frac{1}{\beta + 1/s} * (1 - \exp(-b(\beta + 1/s))) \right. \\
&\quad \left. + \frac{1}{\beta - 1/s} * (\exp(b(\beta - 1/s)) - 1) \right) \tag{3.24}
\end{aligned}$$

Durch Herleiten von $M_{U_T}(\beta)$ an den Stellen $\beta = 1/s$ und $\beta = -1/s$ und Berechnung der Grenzwerte von $M_{U_T}(\beta)$ aus (3.24) für $\beta \rightarrow 1/s$ und $\beta \rightarrow -1/s$ kann gezeigt werden, dass die Momenterzeugende Funktion stetig ist (siehe Anhang A.4). Die Ableitung von $M_{U_T}(\beta)$ lautet (für $\beta \neq 1/s$ und $\beta \neq -1/s$):

$$\begin{aligned}
\frac{\delta}{\delta\beta} M_{U_T}(\beta) &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\frac{\delta}{\delta\beta} \frac{1 - \exp(-b(\beta + 1/s))}{\beta + 1/s} \right. \\
&\quad \left. + \frac{\delta}{\delta\beta} \frac{\exp(b(\beta - 1/s)) - 1}{\beta - 1/s} \right) \\
&= \frac{1}{2s * (1 - \exp(-b/s))} \left(\frac{b * \exp(-b(\beta + 1/s)) * (\beta + 1/s) - (1 - \exp(-b(\beta + 1/s)))}{(\beta + 1/s)^2} \right. \\
&\quad \left. + \frac{b * \exp(b(\beta - 1/s)) * (\beta - 1/s) - (\exp(b(\beta - 1/s)) - 1)}{(\beta - 1/s)^2} \right) \tag{3.25}
\end{aligned}$$

Die Differenzierbarkeit an den Stellen $\beta = 1/s$ und $\beta = -1/s$ wird hier nicht gezeigt. Die korrigierten Scorefunktionen für die Simulation ergeben sich dann durch Einsetzen von $M_{U_T}(\beta_1)$ und deren Ableitung in die Formeln für die korrigierten Scorefunktionen der Poissonverteilung (Gleichungen 2.32 und 2.33).

3.3. Mischverteilungen

3.3.1. Definition, Momente und Momenterzeugende Funktion

Im letzten Abschnitt dieses Kapitels wird die Familie der Mischverteilungen vorgestellt. Sie erlauben eine sehr flexible Modellierung von Messfehlerverteilungen. Die Dichte einer endlichen, d -dimensionalen Mischverteilung ist eine gewichtete Summe von Dichten, d.h. sie hat die Form (siehe z.B. Everitt 1996, Abschnitt 1):

$$f_U(u; p, \theta) = \sum_{j=1}^c p_j * g_j(u; \theta_j) \quad (3.26)$$

mit $u \in \mathbb{R}^d$, $\theta = (\theta_1, \theta_2, \dots, \theta_c)$, $p = (p_1, p_2, \dots, p_c)$ (Mischanteile), $p_j \in (0, 1)$ für alle $j = 1, \dots, c$ und $\sum_{j=1}^c p_j = 1$. Die g_j sind jeweils Dichten mit einem Parameter(vektor) θ_j (Komponentendichten. Im Folgenden ist mit $U^{(j)}$ jeweils eine Zufallsvariable mit Dichte $g_j(u; \theta_j)$ gemeint).

Eine Mischverteilung tritt beispielsweise dann auf, wenn die Population in Subgruppen zerfällt. In der j -ten Gruppe folgt U einer Verteilung mit Dichte $g_j(x; \theta_j)$. p_j gibt den Anteil der j -ten Gruppe an der Grundgesamtheit an. Wenn die Gruppenzugehörigkeit unbekannt ist, ergibt sich für U die obige Dichtefunktion¹⁵.

Die Komponentendichten g_j können alle aus unterschiedlichen Verteilungen stammen, in der Praxis treten allerdings am häufigsten Mischungen aus Verteilungen der selben Form auf, z.B. Mischungen aus Normalverteilungen (ebenda, Abschnitt 1). Der Simulationsteil dieser Arbeit beschränkt sich auf Mischungen aus zwei univariaten Normalverteilungen.

Der Erwartungswert einer Zufallsvariable U (für $d = 1$) mit Dichte $f_U(u; p, \theta)$ ist die gewichtete Summe der Erwartungswerte von $U^{(j)}$, d.h. $\mathbb{E}(U) = \sum_{j=1}^c p_j \mathbb{E}(U^{(j)})$, da

$$\begin{aligned} \mathbb{E}(U) &= \int_{-\infty}^{\infty} u * f_U(u; p, \theta) du = \int_{-\infty}^{\infty} u * \sum_{j=1}^c p_j * g_j(u; \theta_j) du \\ &= \sum_{j=1}^c p_j * \int_{-\infty}^{\infty} u * g_j(u; \theta_j) du = \sum_{j=1}^c p_j \mathbb{E}(U^{(j)}) \end{aligned} \quad (3.27)$$

¹⁵Da in dieser Arbeit angenommen wird, dass der Messfehler U_i unabhängig von den Kovariablen X_i ist, muss die Gruppenzugehörigkeit unabhängig von den Kovariablen sein.

Die Varianz von U lautet (mit $\mu_j = \mathbb{E}(U^{(j)})$, $\sigma_j = \text{Var}(U^{(j)})$ und $\mu = \mathbb{E}(U)$)

$$\begin{aligned} \text{Var}(U) &= \mathbb{E}(U^2) - \mathbb{E}(U)^2 = \left(\int_{-\infty}^{\infty} u^2 * \sum_{j=1}^c p_j * g_j(u; \theta_j) du \right) - \mu^2 \\ &= \left(\sum_{j=1}^c p_j * \int_{-\infty}^{\infty} u^2 g_j(u; \theta_j) du \right) - \mu^2 \\ &= \left(\sum_{j=1}^c p_j * \mathbb{E}(U^{(j)^2}) \right) - \mu^2 = \left(\sum_{j=1}^c p_j * (\sigma_j^2 + \mu_j^2) \right) - \mu^2 \quad (3.28) \end{aligned}$$

Die Momenterzeugende Funktion von U ist (analog zum Erwartungswert) die gewichtete Summe der Momenterzeugenden Funktionen der $U^{(j)}$:

$$\begin{aligned} M_U(\beta) &= \mathbb{E}(\exp(\beta U)) = \int_{-\infty}^{\infty} \exp(\beta U) * \sum_{j=1}^c p_j * g_j(u; \theta_j) du \\ &= \sum_{j=1}^c p_j * \int_{-\infty}^{\infty} \exp(\beta U) * g_j(u; \theta_j) du = \sum_{j=1}^c p_j * M_{U^{(j)}}(\beta) \quad (3.29) \end{aligned}$$

3.3.2. Bei der Simulation betrachtete Fälle

Wie oben erwähnt, beschränkt sich die Simulation in Abschnitt 4.5 auf Mischungen von zwei univariaten Normalverteilungen, d.h. $d = 1$, $c = 2$, $p_2 = 1 - p_1$ und g_1 und g_2 sind Dichten von Normalverteilungen $N(\mu_1, \sigma_1^2)$ sowie $N(\mu_2, \sigma_2^2)$. Dabei werden drei Fälle betrachtet, die jeweils wie folgt definiert sind (grafische Darstellung siehe Ende des Abschnitts):

In *Fall 1* folgt der Messfehler U einer symmetrischen, unimodalen Verteilung. Damit Symmetrie gilt, ist $\mu_1 = -\mu_2$ und $\sigma_1 = \sigma_2$. Aus der Bedingung $\mathbb{E}(U) = 0$ folgt $p_1 = p_2 = 1/2$. Nach Lindsay (1995, Abschnitt 1.1.1) ist beispielsweise eine Mischung aus zwei Normalverteilungen mit $p_1 = p_2 = 1/2$, gleicher Varianz (d.h. $\sigma_1^2 = \sigma_2^2$) und Mittelwerten, die zwei Standardabweichungen σ_1 auseinander liegen, unimodal. Aus $|\mu_1 - \mu_2| = 2\sigma_1$ und $\mu_1 > 0$ ergibt sich $\mu_1 = \sigma_1$, da

$$2\sigma_1 = |\mu_1 - \mu_2| = |\mu_1 - (-\mu_1)| = 2\mu_1 \quad \Leftrightarrow \quad \mu_1 = \sigma_1 \quad (3.30)$$

Bei der Simulation wird die Varianz von U vorgegeben und daraus der Parameter

$\mu_1 (= \sigma_1 = \sigma_2 = -\mu_2)$ berechnet: Aus obiger Formel für die Varianz von U folgt

$$\begin{aligned}\text{Var}(U) &= p_1 * (\sigma_1^2 + \mu_1^2) + p_2 * (\sigma_2^2 + \mu_2^2) - \mu^2 = \frac{1}{2}(\mu_1^2 + \mu_1^2) + \frac{1}{2}(\mu_1^2 + \mu_1^2) - 0 \\ &= 2\mu_1^2 \quad \Leftrightarrow \quad \mu_1 = \sqrt{\frac{\text{Var}(U)}{2}}\end{aligned}\quad (3.31)$$

Fall 2 ist eine symmetrische, bimodale Verteilung. Es gilt weiterhin $\mu_1 = -\mu_2$, $\sigma_1 = \sigma_2$ und $p_1 = p_2 = 1/2$. Wenn die Mittelwerte vier Standardabweichungen σ_1 auseinander sind, ergibt sich nach Lindsay (1995, Abschnitt 1.1.1) eine bimodale Verteilung. Aus $|\mu_1 - \mu_2| = 4\sigma_1$ folgt für μ_1 und σ_1 (mit $\mu_1 > 0$):

$$4\sigma_1 = |\mu_1 - \mu_2| = |\mu_1 - (-\mu_1)| = 2\mu_1 \quad \Leftrightarrow \quad \mu_1 = 2\sigma_1 \quad (3.32)$$

$$\begin{aligned}\text{Var}(U) &= \frac{1}{2}(\sigma_1^2 + \mu_1^2) + \frac{1}{2}(\sigma_2^2 + \mu_2^2) = \frac{1}{2}(\sigma_1^2 + (2\sigma_1)^2) + \frac{1}{2}(\sigma_1^2 + (2\sigma_1)^2) \\ &= 5\sigma_1^2 \quad \Leftrightarrow \quad \sigma_1 = \sqrt{\frac{\text{Var}(U)}{5}}, \quad \mu_1 = 2\sqrt{\frac{\text{Var}(U)}{5}}\end{aligned}\quad (3.33)$$

In *Fall 3* ist die Verteilung von U asymmetrisch. Sei $\sigma_1 = \sigma_2$, $\sigma_1 = \mu_1$ und $\mu_2 = -2\mu_1$. Aus dem Verhältnis der Mittelwerte und der Bedingung $\mathbb{E}(U) = 0$ folgt $p_1 = 2/3$ (denn $p_1\mu_1 + p_2\mu_2 = 2/3 * \mu_1 + 1/3 * (-2\mu_1) = 0$). μ_1 kann für eine gegebene Varianz dann wie folgt bestimmt werden:

$$\begin{aligned}\text{Var}(U) &= \frac{2}{3}(\sigma_1^2 + \mu_1^2) + \frac{1}{3}(\sigma_2^2 + \mu_2^2) = \frac{2}{3}(\mu_1^2 + \mu_1^2) + \frac{1}{3}(\mu_1^2 + (-2\mu_1)^2) \\ &= \frac{4}{3}\mu_1^2 + \frac{5}{3}\mu_1^2 = \frac{9}{3}\mu_1^2 = 3\mu_1^2 \quad \Leftrightarrow \quad \mu_1 = \sqrt{\frac{\text{Var}(U)}{3}}\end{aligned}\quad (3.34)$$

In Abbildung 4 (folgende Seite) sind die Dichten, die sich jeweils für eine Varianz von 1 ergeben, für die drei Fälle dargestellt. In Fall 1 gilt $\mu_1 = -\mu_2 = \sigma_1 = \sigma_2 = \sqrt{0,5} \approx 0,71$, in Fall 2 $\sigma_1 = \sigma_2 = \sqrt{0,2} \approx 0,45$; $\mu_1 = -\mu_2 \approx 0,89$ und in Fall 3 $\mu_1 = \sigma_1 = \sigma_2 = \sqrt{1/3} \approx 0,58$; $\mu_2 \approx -1,15$. Zum Vergleich ist die Dichte der Standardnormalverteilung in grau dargestellt.

3.3.3. Korrigierte Scorefunktion für die Simulation

Wenn die Messfehler U_i einer Mischverteilung aus zwei Normalverteilungen ($N(\mu_1, \sigma_1^2)$ und $N(\mu_2, \sigma_2^2)$) mit Mischanteilen p_1 und $p_2 = 1 - p_1$ folgen, lautet

die Momenterzeugende Funktion von U_i nach Gleichung (3.29)

$$M_{U_i}(\beta_1) = p_1 * M_{U_i^{(1)}}(\beta_1) + (1 - p_1) * M_{U_i^{(2)}}(\beta_1) \quad (3.35)$$

$$\stackrel{(*)}{=} p_1 * \exp\left(\beta_1 \mu_1 + \frac{1}{2} \sigma_1^2 \beta_1^2\right) + (1 - p_1) * \exp\left(\beta_1 \mu_2 + \frac{1}{2} \sigma_2^2 \beta_1^2\right)$$

wobei in (*) verwendet wurde, dass für $X \sim N(\mu, \sigma^2)$ $M_X(\beta) = \exp(\beta\mu + 1/2\sigma^2\beta^2)$ gilt (siehe z.B. Gut 2009, Kapitel 5.4). Die Ableitung von $M_{U_i}(\beta_1)$ lautet

$$\frac{\delta}{\delta\beta_1} M_{U_i}(\beta_1) = (\mu_1 + \sigma_1^2 \beta_1) * p_1 * \exp\left(\beta_1 \mu_1 + \frac{1}{2} \sigma_1^2 \beta_1^2\right)$$

$$+ (\mu_2 + \sigma_2^2 \beta_1) * (1 - p_1) * \exp\left(\beta_1 \mu_2 + \frac{1}{2} \sigma_2^2 \beta_1^2\right) \quad (3.36)$$

Die korrigierten Scorefunktionen für die Simulation ergeben sich dann durch Einsetzen von $M_{U_i}(\beta_1)$ und deren Ableitung in die Formeln der korrigierten Scorefunktionen der Poissonverteilung (Gleichungen 2.32 und 2.33).

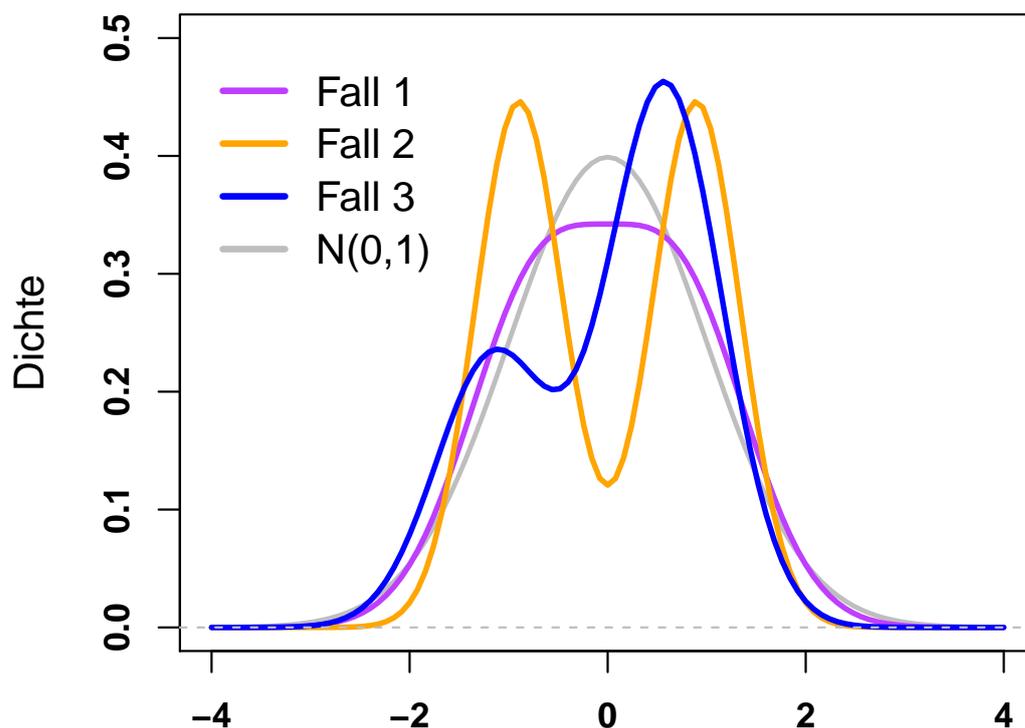


Abb. 4: Mischverteilungen aus zwei Normalverteilungen; Fall 1: symmetrisch, unimodal; Fall 2: symmetrisch, bimodal; Fall 3: asymmetrisch; jeweils mit Varianz 1

4. Praktische Umsetzung & Simulation

In diesem Kapitel wird zunächst die Idee des *Newton-Raphson-Algorithmus*, einer Methode zur Bestimmung der Nullstelle der korrigierten Scorefunktion, d.h. des korrigierten Schätzers, skizziert. Dabei wird auch auf mögliche Probleme, die dabei auftreten können, eingegangen.

Die korrigierten Scorefunktionen beruhen auf Asymptotik, d.h. auf einem Stichprobenumfang, der gegen unendlich geht. Um die Qualität der Methode bei endlichen Stichprobenumfängen zu untersuchen, werden anschließend die Ergebnisse von Simulationen mit einer Poisson-verteilten Zielvariable und Messfehlern aus den im vorherigen Kapitel vorgestellten Verteilungen betrachtet. Dabei wird der Stichprobenumfang, die Varianz der Messfehler sowie mögliche weitere Parameter der Messfehlerverteilung systematisch variiert, um die Möglichkeiten und Grenzen des Verfahrens auszutesten.

4.1. Bestimmung der Nullstelle

4.1.1. Newton-Raphson-Algorithmus

In den meisten Fällen lässt sich keine geschlossene Form für $\hat{\beta}_{cor}$, d.h. die Lösung von $\sum_{i=1}^n s_{W_i}(\beta, Y_i, W_i) = 0$, herleiten (eine Ausnahme ist der korrigierte Schätzer bei der linearen Regression, siehe Abschnitt 2.2.2). Zur Bestimmung der Nullstelle kann die *Newton-Raphson-Methode* verwendet werden (siehe z.B. Fahrmeir et al. 2013, B.4.2).

Mit diesem Algorithmus lassen sich allgemein nichtlineare Gleichungssysteme lösen, weshalb er auch bei der Maximum-Likelihood-Schätzung eingesetzt werden kann. Gesucht ist die Lösung $\hat{\beta} \in \mathbb{R}^p$ einer Gleichung $s(\beta) = 0_{p \times 1}$ (bei der korrigierten Scorefunktion ist $s(\beta)$ somit $\sum_{i=1}^n s_{W_i}(\beta, Y_i, W_i)$). Ausgehend von einem Startwert $\hat{\beta}^{(0)}$ wird in jeder Iteration eine verbesserte approximative Lösung für $s(\beta) = 0$ bestimmt. Dabei ergibt sich $\hat{\beta}^{(k+1)}$, der Schätzer im $(k+1)$ -ten Schritt, aus der aktuellen Schätzung $\hat{\beta}^{(k)}$ durch:

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} - \left(\frac{\delta}{\delta \beta^t} s(\hat{\beta}^{(k)}) \right)^{-1} * s(\hat{\beta}^{(k)}) \quad (4.1)$$

(Fahrmeir et al. 2013, B.4.2, so ähnlich auch in Carroll et al. 2006, Kapitel 7.5.)¹⁶ Dieser Schritt wird so lange wiederholt, bis sich der Schätzer bei aufeinanderfolgenden Iterationen praktisch nicht mehr ändert (Fahrmeir et al. 2013, B.4.2).

Für eindimensionale β lässt sich die Idee des Algorithmus recht einfach veranschaulichen: $s(\beta)$ wird an der Stelle $\hat{\beta}^{(0)}$ durch eine Tangente (d.h. eine lineare Funktion) $g(\beta)$ approximiert. Die Nullstelle von $g(\beta)$ ist die verbesserte Schätzung $\hat{\beta}^{(1)}$. Die Formel für die Tangente $g(\beta)$ lautet (mit $s'(\beta) = \frac{\delta}{\delta\beta}s(\beta)$):

$$g(\beta) = s(\hat{\beta}^{(0)}) + s'(\hat{\beta}^{(0)}) * (\beta - \hat{\beta}^{(0)}) \quad (4.2)$$

(ebenda, B.4.2). Nullsetzen und Umformen liefert :

$$\hat{\beta}^{(1)} = \hat{\beta}^{(0)} - \left(\frac{1}{s'(\hat{\beta}^{(0)})} \right) * s(\hat{\beta}^{(0)}) \quad (4.3)$$

Im nächsten Schritt wird $s(\beta)$ an der Stelle $\hat{\beta}^{(1)}$ erneut durch eine Tangente approximiert und deren Nullstelle bestimmt usw (siehe Gleichung 4.1).

Wenn $\sum_{i=1}^n s_{W_i}(\beta, Y_i, W_i) = 0$ mehrere Nullstellen hat, beeinflusst die Wahl des Startwerts $\hat{\beta}^{(0)}$, gegen welche der Nullstellen der Newton-Raphson-Algorithmus konvergiert (Carroll et al. 2006, Kapitel 7.5). Nach Stefanski (1989, Abschnitt 3.1) und Nakamura (1990, Abschnitt 6) ist der naive Schätzer $\hat{\beta}_{naiv}$ ein sinnvoller Startwert.

In der Statistiksoftware R (R Core Team 2019) ist das Newton-Raphson-Verfahren zum Beispiel im *rootSolve*-Package (Soetaert 2009) in der Funktion *multiroot* implementiert. Diese wurde für den Simulationsteil dieser Arbeit zur Berechnung der korrigierten Schätzer verwendet. Als Startwert $\hat{\beta}^{(0)}$ wurde das Ergebnis der naiven Schätzung übergeben¹⁷.

4.1.2. Mögliche Probleme

Nicht immer führt der Newton-Raphson-Algorithmus zu einem Ergebnis. Folgende Probleme können beispielsweise bei der Schätzung auftreten:

¹⁶Bei der ML-Schätzung ist $-\frac{\delta s(\hat{\beta}^{(k)})}{\delta\beta^t}$ die beobachtete Fisherinformationsmatrix. Den häufiger verwendeten *Fisher-Scoring-Algorithmus* erhält man, indem die beobachtete durch die erwartete Fisherinfo von $\hat{\beta}^{(k)}$ ersetzt wird (Fahrmeir et al. 2013, B.4.2).

¹⁷Eine Ausnahme befindet sich im Anhang A.5, dort wird der im Folgenden definierte Schätzer $\hat{\beta}_{normal}$ als Startwert verwendet.

- Die Matrix $\frac{\delta}{\delta\beta^i} s(\hat{\beta}^{(k)})$ (Jacobi-Matrix der Scorefunktion) in Gleichung (4.1) ist nicht invertierbar (d.h. hat nicht vollen Rang), weshalb der Algorithmus nicht fortfahren kann. Bei der *multiroot*-Funktion macht sich dies durch die Warnmeldung „error during factorisation of matrix; singular matrix“ bemerkbar (siehe auch Dokumentation des *rootSolve*-Packages, Soetaert 2009).
- Der Schätzer ist auch nach der maximalen Anzahl an Iterationen (hier wurde 100, der Default-Wert der *multiroot*-Funktion, verwendet) nicht konvergiert. Die zugehörige Warnmeldung lautet „steady-state not reached“.
- Die für das Resultat $\hat{\beta}_{cor}$ ausgegeben Funktionswerte der korrigierten Scorefunktion lauten *NaN* (keine Warnmeldung, im Folgenden Fall „NaN“ genannt). Dies scheint nur bei der schiefen Normalverteilung, den Mischverteilungen und der trunkierten Laplace-Verteilung aufzutreten und könnte daran liegen, dass bei diesen Verteilungen in der korrigierten Scorefunktion bei extrem großen oder kleinen Werten für β 0 geteilt durch 0 gerechnet wird (siehe z.B. Gleichungen 3.8 und 3.9).

Beispiele zu den Warnmeldungen bzw. zum dritten Punkt befinden sich in *Warnmeldungen.R* im elektronischen Anhang. In allen drei Fällen sind die Lösungen, die von der Funktion übergeben werden, häufig weit von realistischen Schätzungen entfernt (z.B. betragsmäßig $> 10^6$ oder größer). Die Ergebnisse aus Simulationen, bei denen einer dieser Fälle eingetreten ist, werden nicht gespeichert, sondern durch *NA* ersetzt. Bei der Vorstellung der Ergebnisse der Simulationen in den nachfolgenden Abschnitten wird teilweise auf die Häufigkeit dieser Probleme eingegangen. Da bei der Laplace-Verteilung die Momenterzeugende Funktion nur auf einem Intervall existiert, werden dort zusätzlich die Schätzwerte, die außerhalb dieses Intervalls liegen, entfernt.

In *Warnmeldungen.R* wird zudem stichprobenhaft untersucht, ob sich die Mediane der Schätzungen von $\hat{\beta}_{cor,1}$ stark verändern, wenn die Schätzwerte in den oben genannten Fällen nicht verworfen, sondern behalten werden. Die Unterschiede, die sich dadurch ergeben, sind sehr gering. Eine erneute Schätzung mit einer höheren Anzahl an maximalen Iterationen und einem anderen Startwert, wenn der Schätzer im ersten Versuch nicht konvergiert ist, führt zu keiner Verbesserung der Ergebnisse. Die Resultate des Vergleichs der unterschiedlichen Vorgehensweisen befinden sich in Anhang A.5.

4.2. Allgemeines Vorgehen bei der Simulation

Simulationen in der Literatur zu korrigierten Scorefunktionen gibt es beispielsweise bei Nakamura (1990) und Novick und Stefanski (2002). Das Vorgehen in dieser Arbeit orientiert sich größtenteils an diesen Quellen.

Es wird eine Poisson-verteilten Zielvariablen in einem Modell mit einer Kovariable und Intercept (d.h. $Y_i|X_i \sim \text{Po}(\exp(\beta_0 + \beta_1 X_i), X_i \in \mathbb{R})$) betrachtet. Die wahren Parameter werden dabei stets auf $\beta_{t,0} = 1$ und $\beta_{t,1} = 0,25$ festgelegt. Die X_i -Werte werden aus einer $N(0,1)$ -Verteilung gezogen.

Jeder Simulationsdurchgang läuft wie folgt ab: Zunächst werden die wahren X_i (z.B. mit $n = 100$) gezogen und daraus die Y_i simuliert. Aus diesen Daten wird der Maximum-Likelihood-Schätzer $\hat{\beta}_{ML}$ berechnet. Anschließend wird ein messfehlerbehafteter Ersatz für die X_i simuliert: W_i mit $W_i = X_i + U_i$. Die Messfehler U_i werden dabei aus einer vorgegeben Verteilung gezogen (z.B. Laplace-Verteilung mit Varianz 1). Das ganze wird 500 Mal wiederholt, es werden somit $500 * n$ Messfehler simuliert. Für jeden der 500 messfehlerbehafteten Datensätze werden drei Schätzer berechnet:

- Der naive Schätzer $\hat{\beta}_{naiv}$, d.h. die Nullstelle der naiven Scorefunktion $s_X(\beta, Y, W)$ (siehe Abschnitt 2.2.3, Gleichungen 2.30 und 2.31).
- Der Schätzer $\hat{\beta}_{normal}$, den man unter der (hier falschen) Annahme, dass die Messfehler normalverteilt seien, erhält. Er ist somit die Lösung der korrigierten Scorefunktion (siehe Gleichungen 2.32 und 2.33), wenn man für $M_{U_i}(\beta_1)$ die Momenterzeugende Funktion der Normalverteilung einsetzt¹⁸.
- Der korrigierte Schätzer $\hat{\beta}_{cor}$, d.h. Korrektur unter der Annahme der wahren Messfehlerverteilung. Die korrigierten Scorefunktionen wurden im vorangegangenen Kapitel aufgestellt.

Bei Messfehlern aus einer trunkierten Laplace-Verteilung wird anstelle von $\hat{\beta}_{normal}$ ein Schätzer unter der Annahme einer untrunkierten Laplace-Verteilung berechnet. Dies wird im dazugehörigen Abschnitt 4.4.3 genauer erläutert.

In jeden Simulationsdurchgang werden somit (im Optimalfall) 500 $\hat{\beta}_{naiv}$, 500 $\hat{\beta}_{normal}$ und 500 $\hat{\beta}_{cor}$ bestimmt. Wie in Abschnitt 4.1.2 beschrieben, führt die

¹⁸Mit $M_{U_i}(\beta_1) = \exp(1/2\sigma^2\beta_1^2)$ (Momenterzeugende Funktion von $N(0, \sigma^2)$) ergibt sich $s_{W_i}(\beta_0, Y_i, W_i) = Y_i - \exp(\beta_0 + \beta_1 W_i - 1/2\sigma^2\beta_1^2)$ und $s_{W_i}(\beta_1, Y_i, W_i) = Y_i W_i - \exp(\beta_0 + \beta_1 W_i - 1/2\sigma^2\beta_1^2) * (W_i - \sigma^2\beta_1^2)$.

korrigierte Schätzung nicht immer zu einem Ergebnis. Die Anzahl der Wiederholungen, in denen ein Schätzwert erhalten wurde, ist in den Grafiken jeweils über dem Boxplot angegeben (N=...). N=500 bedeutet somit, dass bei keiner Iteration ein Problem (z.B. keine Invertierbarkeit, Schätzer betragsmäßig zu groß (Laplace-Verteilung) etc.) aufgetreten ist.

Um die Güte der Schätzer zu bestimmen, werden sie mit dem Maximum-Likelihood-Schätzer $\hat{\beta}_{ML}$ verglichen¹⁹. Dafür wird der relative Bias bestimmt (analog zu Nakamura 1990, Abschnitt 5): $(\hat{\beta}_1 - \hat{\beta}_{ML,1})/\hat{\beta}_{ML,1}$; analog für $\hat{\beta}_0$ (wobei im Folgenden lediglich die Ergebnisse für $\hat{\beta}_1$ dargestellt sind). Die Verteilung des relativen Bias bei den jeweils 500 Wiederholungen wird durch Boxplots veranschaulicht. Dadurch wird neben dem Median auch die Varianz der Schätzung deutlich. Obwohl alle Schätzer, die nicht oder nur unter Warnmeldungen konvergiert sind, aussortiert werden (siehe Abschnitt 4.1.2), ergeben sich zum Teil trotzdem unrealistische Werte (bis hin zu betragsmäßig $> 10^6$ oder noch größer). Deswegen ist die Skala der Boxplots teilweise eingeschränkt, um eine gute Sichtbarkeit der Box zu gewährleisten. Aus demselben Grund eignet sich der Mittelwert hier weniger zum Vergleich der Schätzer, da dieser nicht robust gegenüber solchen Ausreißern ist.

Der R-Code zu den Simulationen befindet sich im elektronischen Anhang dieser Arbeit. Durch *set.seed* sind die Ergebnisse reproduzierbar.

4.3. Schiefe Normalverteilung

4.3.1. Unterschiedliche Messfehlervarianzen

Zunächst wird der Stichprobenumfang auf $n = 100$ gesetzt und für feste Schiefeparameter α die Varianz der Messfehler zwischen $0,25^2$ und $1,9^2$ variiert (man beachte, dass die Varianz der wahren X_i bei 1 liegt). Für jeden Wert der Varianz werden je 500 Mal 100 Messfehler aus einer schiefen Normalverteilung simuliert (zur Bestimmung von ξ und ω für gegebene α und $\text{Var}(U_i)$ siehe Abschnitt 3.1.2).

Abbildung 5 zeigt die Ergebnisse für $\alpha = 3$. Die wahren X_i und die Y_i werden dabei nur einmal simuliert, d.h. in jeder 10 betrachteten Simulationen ist der wahre Datensatz der selbe. Die Grafik zeigt Boxplots zur relativen Abweichung von $\hat{\beta}_1$ zu $\hat{\beta}_{ML,1}$ für die drei berechneten Schätzer in Abhängigkeit der Standardabweichung der Messfehler U_i .

¹⁹Da der korrigierte Schätzer gegen den ML-Schätzer konvergiert, wird dieser als Bezugsgröße genommen und nicht die wahren Parameter.

Je größer die Varianz, desto mehr steigt der Bias der drei Schätzer. $\hat{\beta}_{normal}$ und $\hat{\beta}_{cor}$ verhalten sich dabei ähnlich und liefern bis zu einer Standardabweichung von 1 sehr genaue Schätzwerte. Die Schätzungen von $\hat{\beta}_{normal}$ sind trotz der falschen Annahme nicht schlechter. Auffällig ist, dass bei der korrigierten Schätzung bei größeren Varianzen in mehr als 20 % aller Fälle eines der oben genannten Probleme aufgetreten ist. Dies lag meistens an der fehlenden Invertierbarkeit der Jacobi-Matrix oder an dem Fall „NaN“. Die naive Schätzung ist stark verzerrt, weist aber vor allem bei großen Messfehlervarianzen eine deutlich geringere Streuung auf.

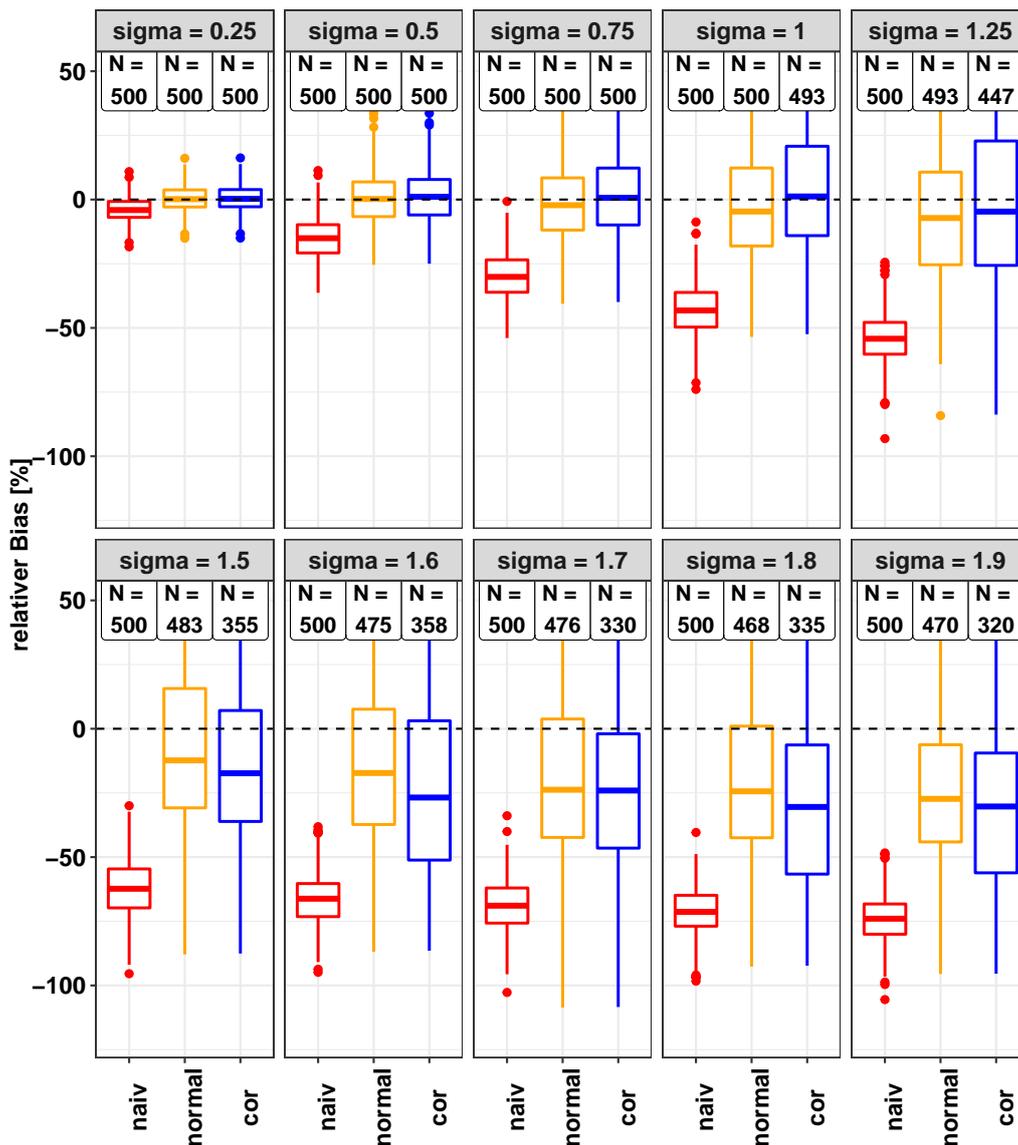


Abb. 5: Schiefe Normalverteilung: verschiedene Standardabweichungen von U_i mit $\alpha = 3$, $n = 100$. Relativer Bias von $\hat{\beta}_1$

Abbildung 6 zeigt die Ergebnisse für $\alpha = 10$ ²⁰. Die Boxplots ähneln sehr stark denen von $\alpha = 3$, der Bias bei großen Varianzen ist hier nur geringfügig höher. Im folgenden Abschnitt wird genauer untersucht, ob die Korrektur invariant gegenüber Veränderungen von α ist.

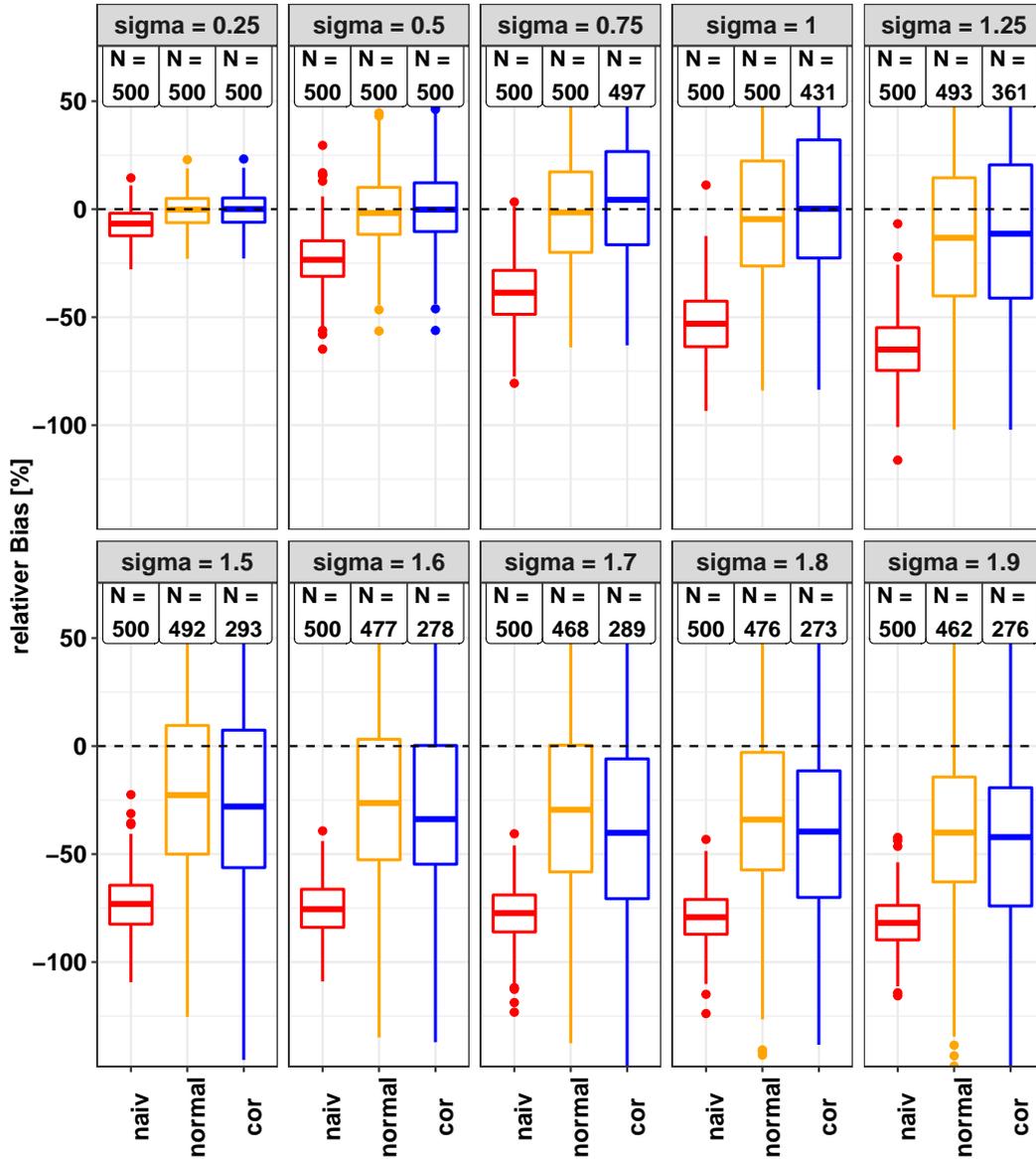


Abb. 6: Schiefe Normalverteilung: verschiedene Standardabweichungen von U_i mit $\alpha = 10$, $n = 100$

²⁰Auch hier wurde insgesamt nur ein wahrer Datensatz generiert, allerdings ein anderer als für $\alpha = 3$. Dies ist auch bei den folgenden Simulationen (außer für unterschiedliche n) der Fall.

4.3.2. Unterschiedliche Schiefe Parameter

Nun werden für eine feste Messfehlervarianz und $n = 100$ unterschiedliche Schiefe Parameter α betrachtet ($\alpha = 1$ bis $\alpha = 50$). Zunächst wird die Varianz von U_i auf $0,75^2$ gesetzt. Abbildung 7 zeigt die Ergebnisse für sechs Werte von α .

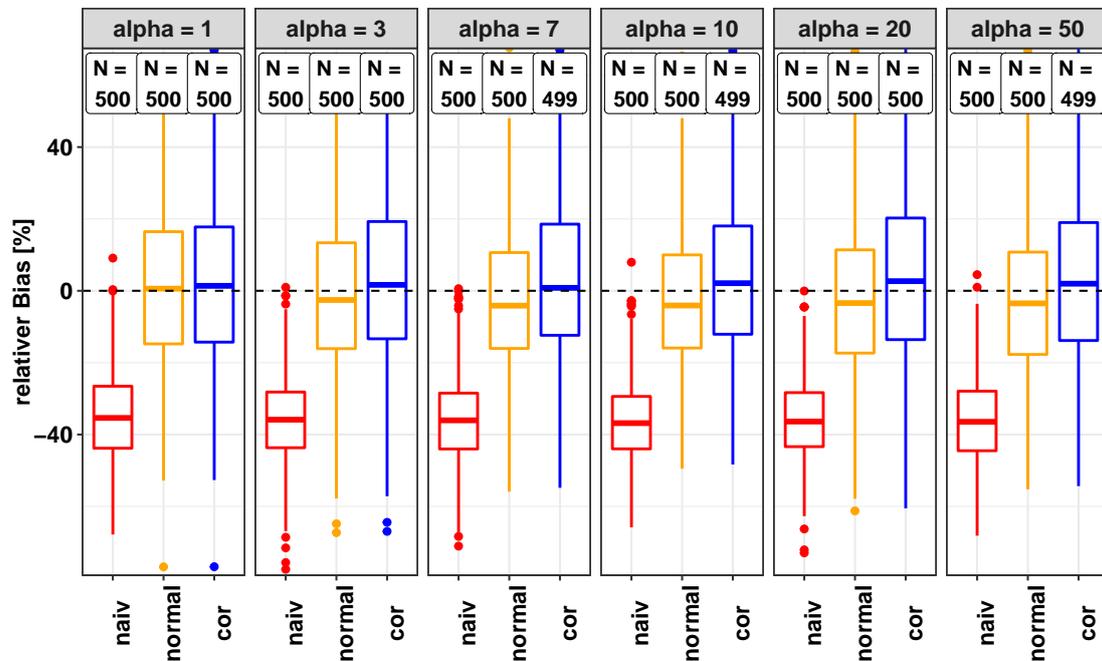


Abb. 7: Schiefe Normalverteilung: verschiedene Werte von α mit $\text{Var}(U_i) = 0,75^2$, $n = 100$

Der Eindruck aus dem vorangegangenen Abschnitt bestätigt sich: eine Erhöhung von α führt nur zu einer geringen Zunahme des Bias. Dass $\hat{\beta}_{normal}$ für geringe α gute Ergebnisse liefert, ist wenig überraschend ($\alpha = 0$ ist eine Normalverteilung; je kleiner α , desto geringer sind die Abweichungen der schiefen zur „normalen“ Normalverteilung), aber selbst für sehr schiefe Verteilungen sind die Schätzungen von $\hat{\beta}_{normal}$ ziemlich genau.

In Abbildung 8 ist die Varianz auf $1,7^2$ festgelegt. Auch hier scheint α keinen großen Einfluss auf Median und Varianz der drei Schätzer zu haben. Allerdings nimmt mit steigendem α die Anzahl der „Problemfälle“ bei der korrigierten Schätzung zu. Für $\alpha = 50$ wird nur in etwa 50% aller Fälle ein Schätzwert erhalten. Bei $\hat{\beta}_{normal}$ ist keine solche Zunahme zu beobachten.

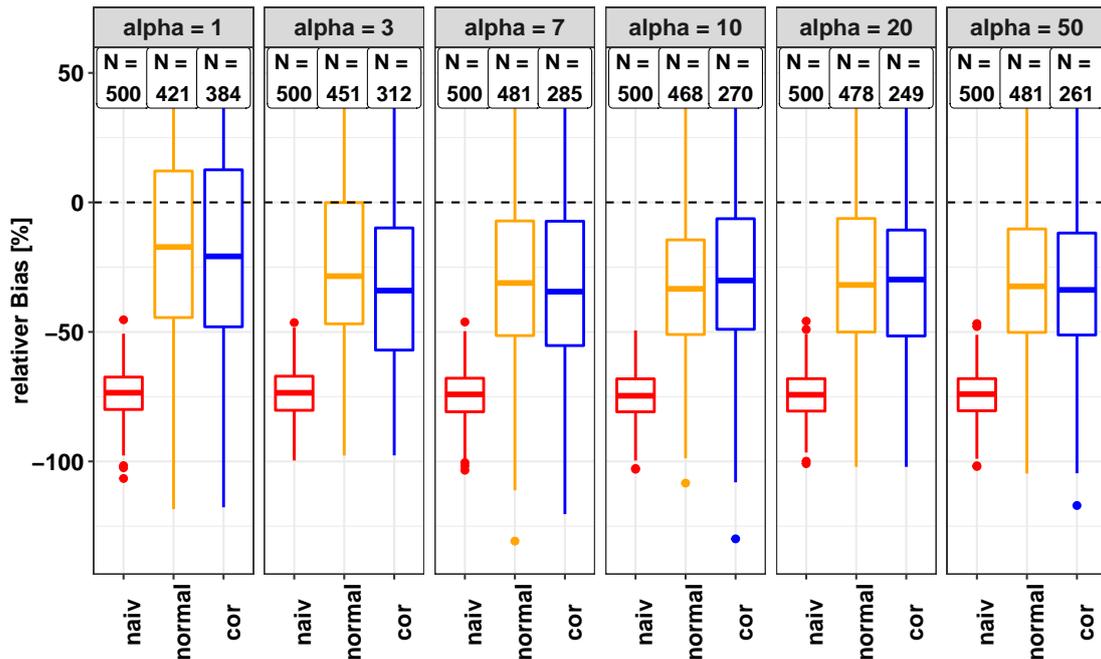


Abb. 8: Schiefe Normalverteilung: verschiedene Werte von α mit $\text{Var}(U_i) = 1, 7^2$, $n = 100$

4.3.3. Unterschiedliche Stichprobenumfänge

Abschließend werden die Resultate für verschiedene Stichprobenumfänge n untersucht. α wird dabei auf 3 gesetzt. Bei Abbildung 9 (folgende Seite) liegt die Varianz stets bei $0, 75^2$. Da für $n = 100$ die Ergebnisse bei dieser Varianz bislang stabil waren, wird nun der Stichprobenumfang sukzessive bis $n = 20$ verkleinert. Die Mediane der drei Schätzer bleiben von der Verkleinerung weitestgehend unberührt, die Varianz der Schätzungen nimmt wie erwartet zu. Die Unterschiede zwischen $\hat{\beta}_{normal}$ und $\hat{\beta}_{cor}$ sind weiterhin sehr gering. Selbst bei kleinen Stichprobenumfängen wird bei fast jeder der 500 Wiederholungen ein Schätzwert erhalten.

In Abbildung 10 ist die Varianz $1, 7^2$. Für $n = 100$ war der korrigierte Schätzer in den vorangegangenen Simulationen bei dieser Varianz leicht verzerrt und führte nur in etwa 60% aller Fälle zu einem Ergebnis (vgl. Abb. 5 und 8 mit $\alpha = 3$). Deswegen wird der Stichprobenumfang nun schrittweise bis $n = 5000$ erhöht. Die Anzahl der Fälle, in denen kein Schätzer $\hat{\beta}_{cor}$ erhalten wird, nimmt dadurch ab. Hier ergeben sich erstmals größere Unterschiede zwischen $\hat{\beta}_{normal}$ und $\hat{\beta}_{cor}$ bezüglich der Güte: während der Bias von $\hat{\beta}_{normal}$ über alle Stichprobenumfänge konstant bleibt, nähern sich die Schätzungen $\hat{\beta}_{cor}$ für große n dem Maximum-Likelihood-Schätzer

immer weiter an. Dies belegt die asymptotische Erwartungstreue des korrigierten Schätzers, während $\hat{\beta}_{normal}$ durch die falsche Annahme verzerrt ist.

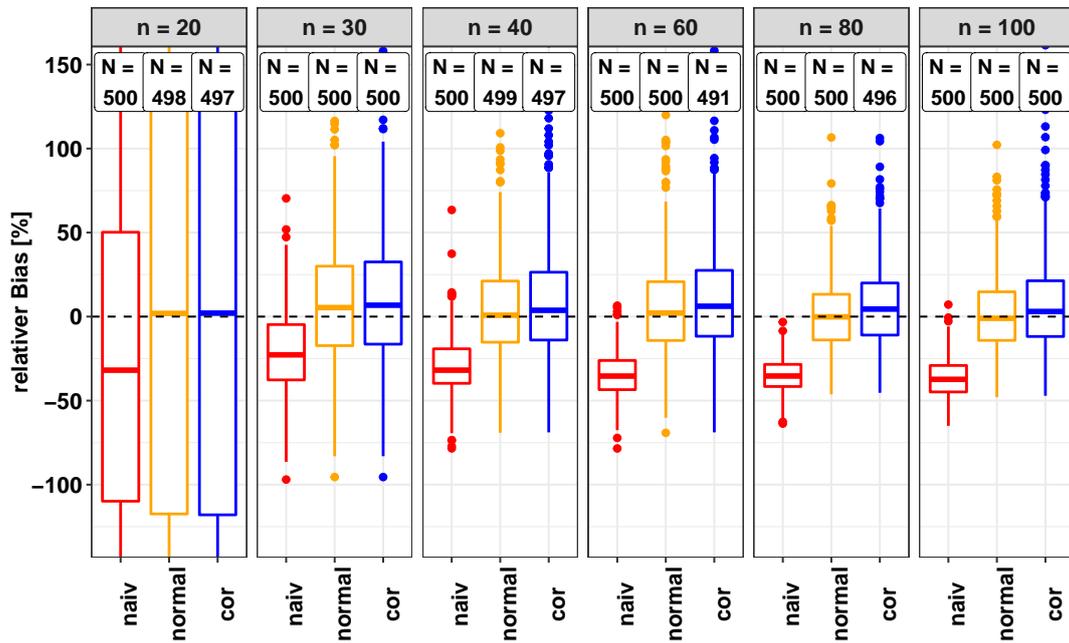


Abb. 9: Schiefe Normalverteilung: verschiedene Stichprobenumfänge n mit $\alpha = 3$ und $\text{Var}(U_i) = 0,75^2$

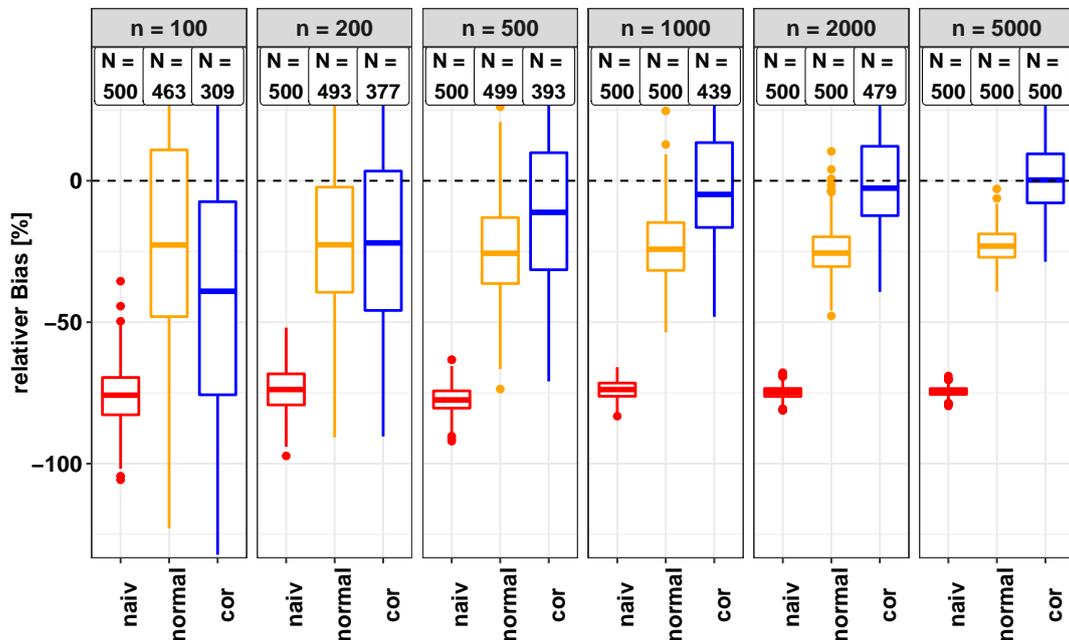


Abb. 10: Schiefe Normalverteilung: verschiedene Stichprobenumfänge n mit $\alpha = 3$ und $\text{Var}(U_i) = 1,7^2$

4.4. Laplace-Verteilung

4.4.1. Unterschiedliche Messfehlervarianzen

Nun werden die Messfehler aus einer Laplace-Verteilung gezogen. Zunächst wird für ein festes n die Varianz der U_i variiert zwischen $0,25^2$ und $1,9^2$. Im extremsten Fall mit Varianz $= 1,9^2$ ist der Skalenparameter $s = \sqrt{1,9^2/2} \approx 1,34$. Die Momenterzeugende Funktion von β_1 existiert dann auf dem Intervall $[-1/1,34; 1/1,34] \approx [-0,74; 0,74]$, das den wahren Parameter $\beta_{t,1} = 0,25$ somit enthält. Schätzungen $\hat{\beta}_{cor}$, die außerhalb des von s abhängigen Intervalls liegen, wird NA zugewiesen.

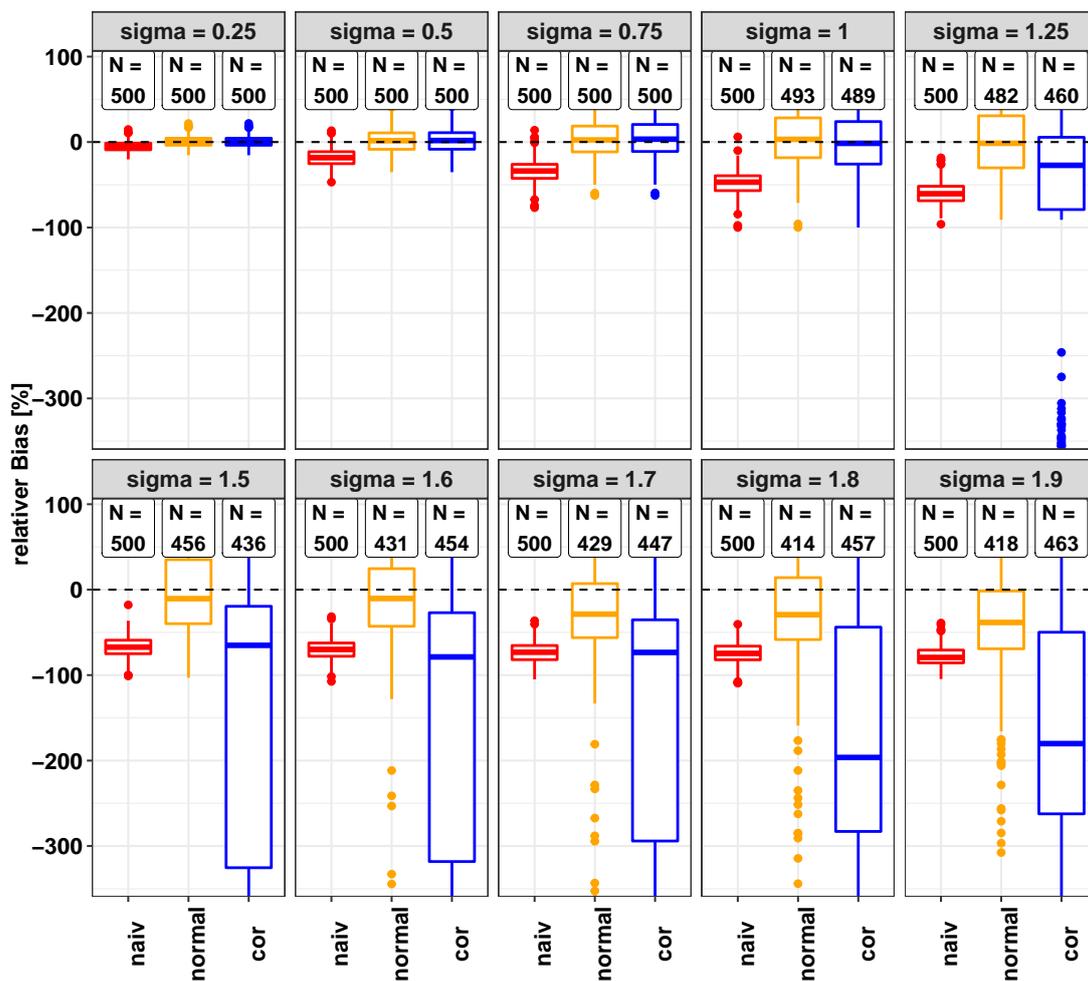


Abb. 11: Laplace-Verteilung: verschiedene Standardabweichungen von U_i mit $n = 100$

In Abbildung 11 ist $n = 100$. Bis zu einer Varianz von 1 liefern sowohl $\hat{\beta}_{normal}$ als auch $\hat{\beta}_{cor}$ gute Schätzungen, bei größeren Messfehlervarianzen nimmt die Güte schrittweise ab. Die Varianz von $\hat{\beta}_{cor}$ ist dabei deutlich größer und die Ergebnisse

schlechter als bei $\hat{\beta}_{normal}$. Dafür liefert die korrigierte Schätzung häufiger ein Ergebnis. Bei $\hat{\beta}_{normal}$ ist die Invertierbarkeit der Jacobi-Matrix öfters nicht gegeben. Bei beiden Schätzern ist außerdem in mehreren Fällen auch nach 100 Iterationen keine Konvergenz erreicht. Verglichen mit Messfehlern aus einer schiefen Normalverteilung (vgl. Abb. 5 und 6), liefert der korrigierte Schätzer hier stärker verzerrte Schätzungen, allerdings treten seltener Probleme auf (nur in ca. 10% aller Fälle).

In Abb. 12 wurden die selben Varianzen bei einer Stichprobengröße von 5000 simuliert. Der Bias verkleinert sich deutlich (siehe Skala der y-Achse). Bei diesem Stichprobenumfang liefert $\hat{\beta}_{cor}$ bessere Ergebnisse als der auf einer falschen Annahme basierende Schätzer $\hat{\beta}_{normal}$, hat allerdings eine größere Varianz. Des Weiteren ist auch der korrigierte Schätzer bei Varianzen größer als 1,7² leicht verzerrt.

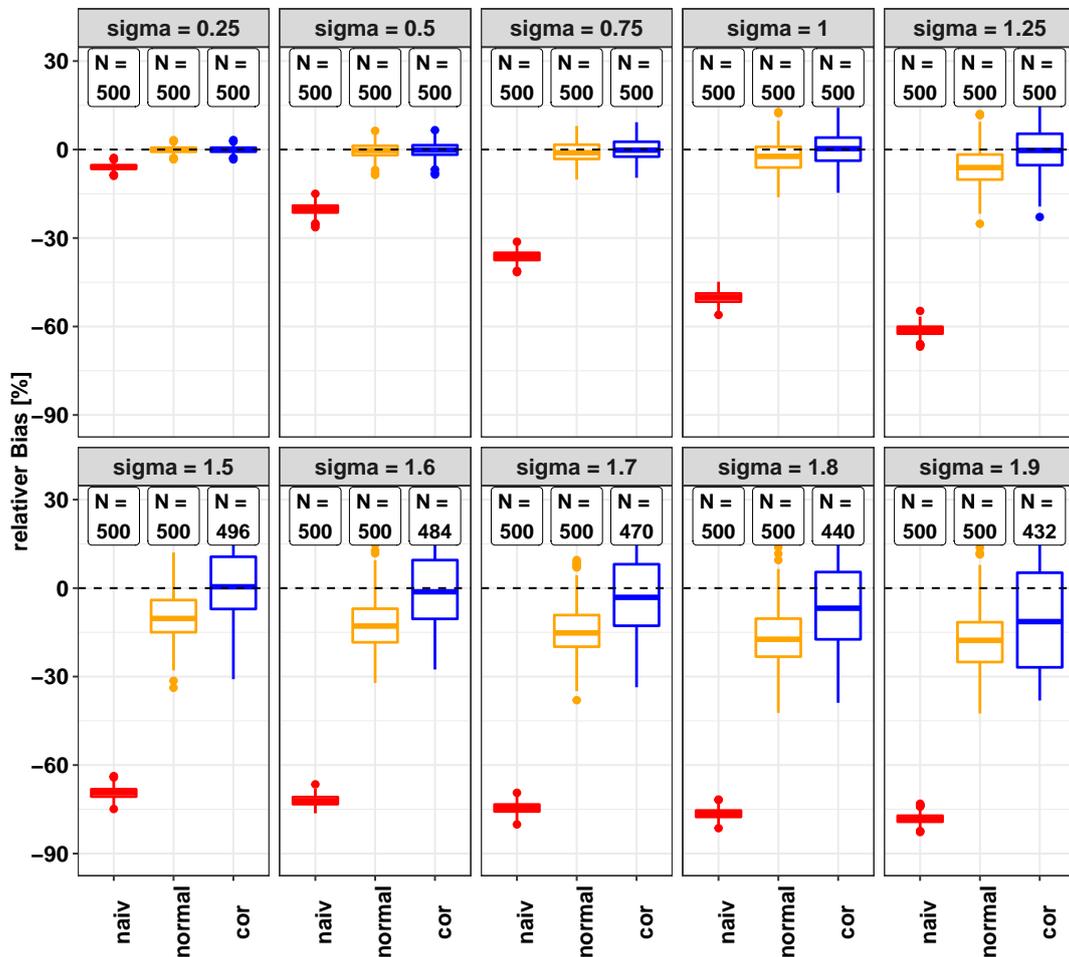


Abb. 12: Laplace-Verteilung: verschiedene Standardabweichungen von U_i mit $n = 5000$

4.4.2. Unterschiedliche Stichprobenumfänge

Im vorherigen Abschnitt wurden bereits zwei verschiedene Stichprobenumfänge betrachtet. Jetzt wird n für eine feste Messfehlervarianzen variiert. In Abbildung 13 sind die Ergebnisse einer Verkleinerung des Stichprobenumfangs bei einer Varianz von $0,75^2$ dargestellt. Die Varianz nimmt mit kleinerem n nur leicht zu, allerdings treten häufiger Probleme bei der Schätzung auf.

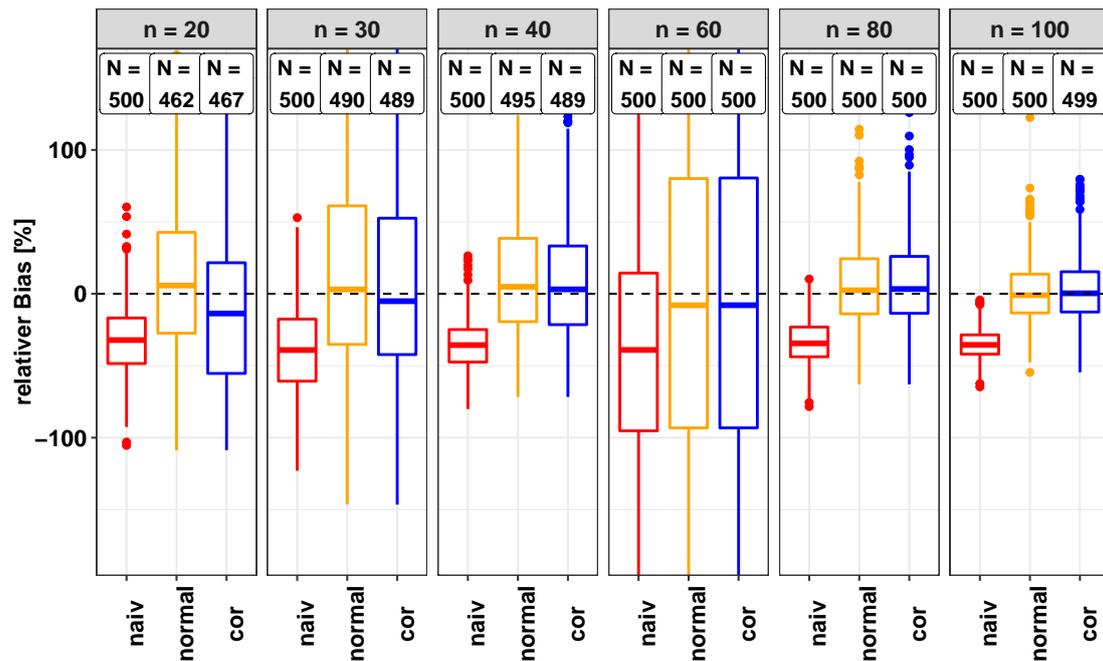


Abb. 13: Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$

Abbildung 14 (nächste Seite) zeigt die Ergebnisse einer Vergrößerung des Stichprobenumfangs bei einer Varianz von $1,7^2$. Für große n nähert sich der Bias der 0 an, allerdings gibt es immer noch teilweise (betragsmäßig) große Ausreißer. Insgesamt ist die Schätzung deutlich ungenauer als bei der schiefen Normalverteilung (Abb. 10). Erst bei einem Stichprobenumfang von $n = 5000$ liefert der korrigierte Schätzer bessere Ergebnisse als $\hat{\beta}_{normal}$.

Insgesamt scheint die Korrektur somit bei Messfehlern aus einer schiefen Normalverteilung besser zu funktionieren als bei Laplace-verteilten U_i , wenn man die Genauigkeit (Median des Bias) und die Varianz der Schätzung betrachtet. Auf der anderen Seite treten bei der Schätzung mit Laplace-verteilten Messfehlern seltener Probleme (z.B. mit der Invertierbarkeit) auf.

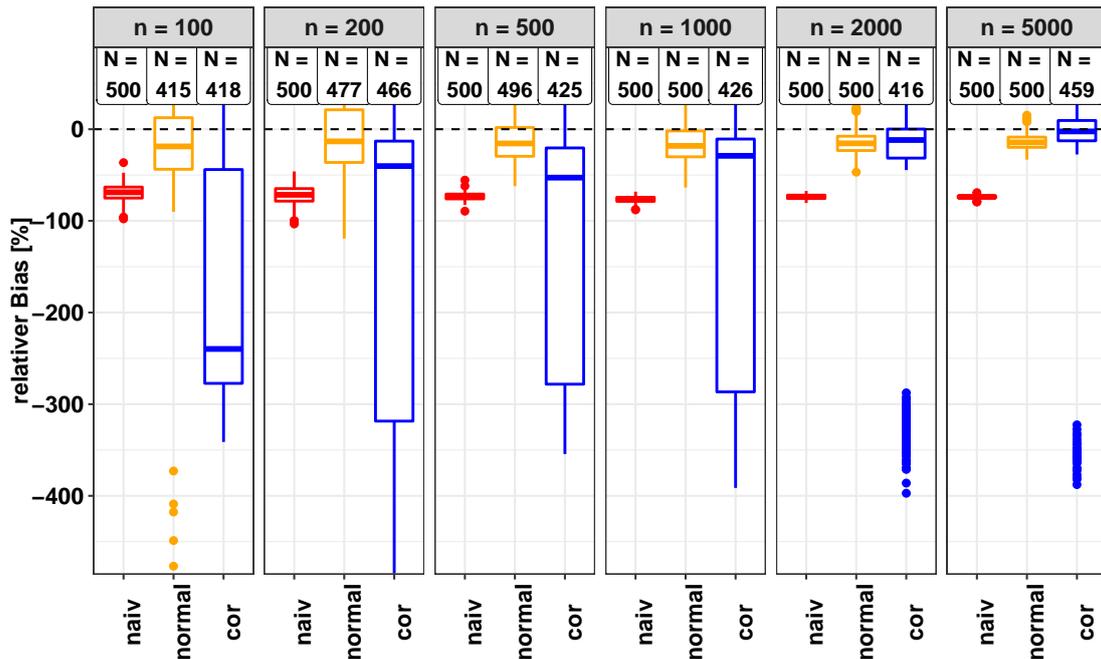


Abb. 14: Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$

4.4.3. Trunkierte Laplace-Verteilung

Bei der trunkierten Laplace-Verteilung existiert die Momenterzeugende Funktion auf ganz \mathbb{R} . Nun soll untersucht werden, ob die Ergebnisse bei trunkierten Messfehlern stabiler sind als im vorangegangenen Abschnitt. Die trunkierten Messfehler werden aus einer normalen Laplace-Verteilung gezogen und zurückgelegt, wenn sie nicht im Intervall $[-b; b]$ liegen. b ist dabei der Trunkierungsparameter, er wird im Folgenden variiert. In diesem Abschnitt wird $\hat{\beta}_{normal}$ nicht berechnet. Stattdessen wird der richtige korrigierte Schätzer $\hat{\beta}_{tr}$ mit dem Schätzer $\hat{\beta}_{lapl}$, der fälschlicherweise eine untrunkierte Verteilung annimmt, verglichen.

Nun wird b für eine feste Varianz (der untrunkierten Verteilung!) und $n = 100$ variiert. b wird dabei in Vielfachen der Standardabweichung ausgedrückt. Ein kleines b führt zu einer starken Trunkierung, während sich für große b die trunkierte und die untrunkierte Verteilung nahezu nicht unterscheiden. Abb. 15 (folgende Seite) zeigt die Resultate für eine Varianz von $0,75^2$ (untrunkierte Verteilung). Der Schätzer $\hat{\beta}_{tr}$ liefert durchweg gute Ergebnisse. Da für kleine b die Verteilung der Messfehler stärker von einer normalen Laplace-Verteilung abweicht, schneidet $\hat{\beta}_{lapl}$ nicht gut ab. Bei geringer Trunkierung nähern sich die beiden Schätzer an.

Die Varianz von $\hat{\beta}_{tr}$ steigt bei größerem b geringfügig an.

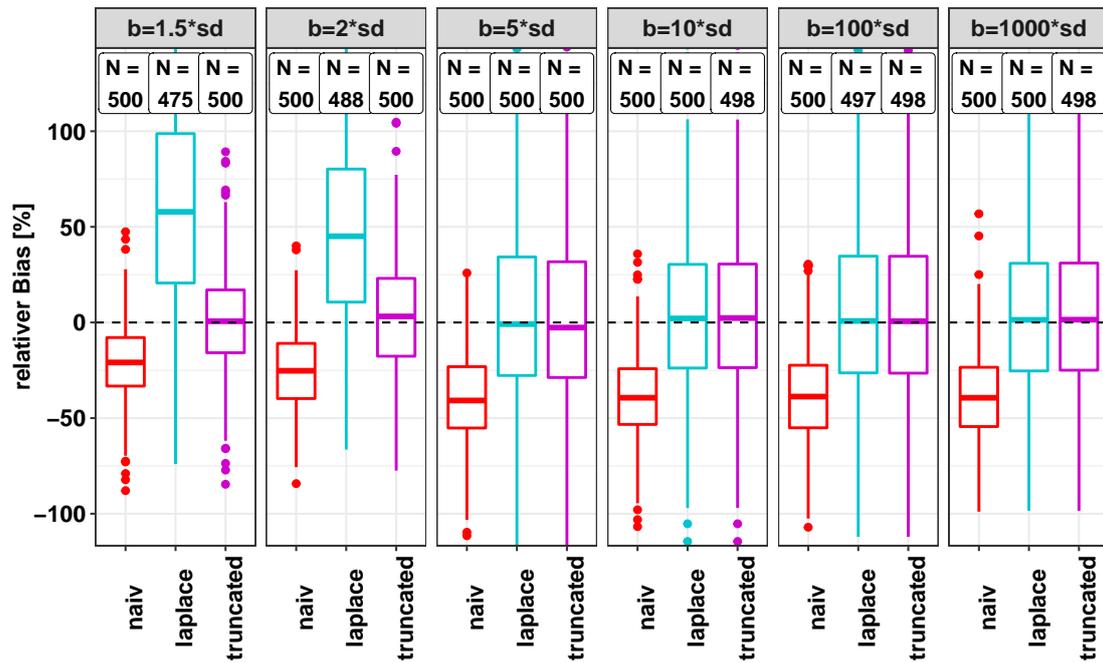


Abb. 15: Trunkierte Laplace-Verteilung: unterschiedlich starke Trunkierung mit $n = 100$ und $\text{Var}(U_i) = 0,75^2$ (untrunkierte Verteilung)

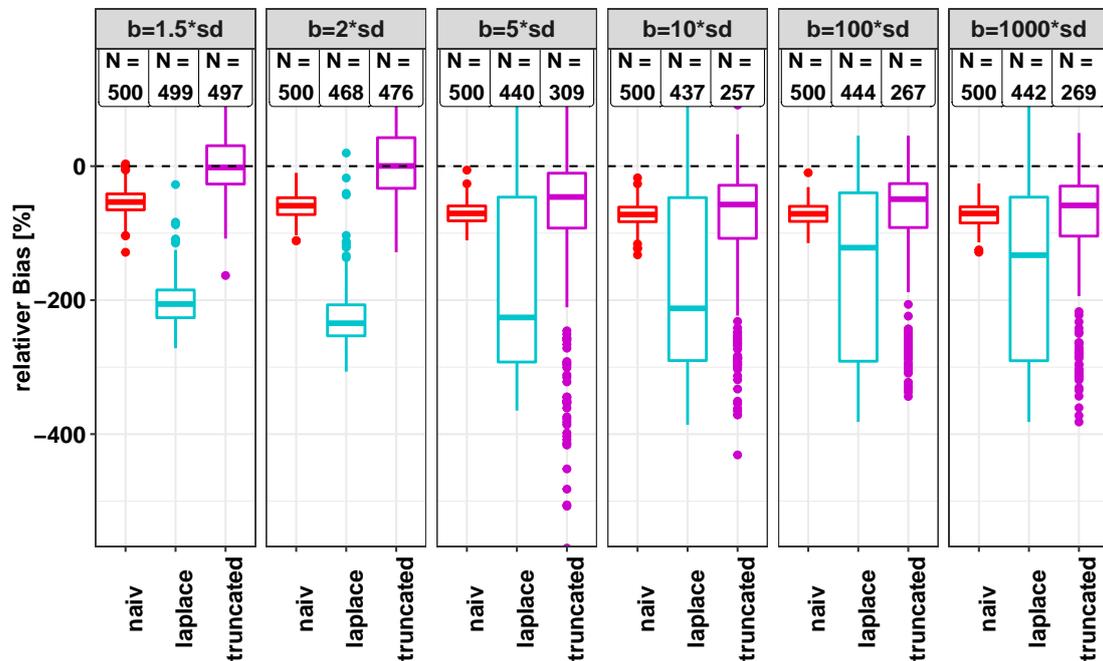


Abb. 16: Trunkierte Laplace-Verteilung: unterschiedlich starke Trunkierung mit $n = 100$ und $\text{Var}(U_i) = 1,7^2$ (untrunkierte Verteilung)

In Abbildung 16 ist die Varianz auf $1,7^2$ festgesetzt. Bei starker Trunkierung haben die Schätzungen $\hat{\beta}_{tr}$ eine geringe Varianz und sind unverzerrt. Mit steigendem b nimmt die Genauigkeit ab, wobei die Varianz gering bleibt (bezogen auf die Länge der Box; es gibt starke Ausreißer nach unten). Die Streuung von $\hat{\beta}_{lapl}$ nimmt mit steigendem b zu, die Verzerrtheit nimmt aber ab. Die Schätzungen $\hat{\beta}_{tr}$ sind für große b sehr instabil, teilweise wird nur in knapp über der Hälfte aller Wiederholungen ein Ergebnis erzielt. Dabei treten alle drei der in Abschnitt 4.1.2 genannten Probleme auf.

Variation der Varianz für festes n und fester Trunkierung²¹ sowie Variation des Stichprobenumfangs für eine feste Trunkierung und feste Varianz liefert ähnliche Ergebnisse (weitere Grafiken siehe Anhang A.6): Die Streuung von $\hat{\beta}_{tr}$ ist häufig deutlich geringer als die von $\hat{\beta}_{lapl}$. Vor allem bei einer starken Trunkierung ist $\hat{\beta}_{lapl}$ stark verzerrt, während $\hat{\beta}_{tr}$ meistens recht genaue Schätzungen liefert. Allerdings ist $\hat{\beta}_{tr}$ ziemlich instabil, vor allem bei steigender Varianz und geringer Trunkierung.

4.5. Mischverteilungen

4.5.1. Unterschiedliche Messfehlervarianzen

Im letzten Abschnitt werden die Messfehler aus Mischungen zweier Normalverteilungen gezogen. Die drei betrachteten Fälle sind in Abschnitt 3.3.2 definiert. Zunächst wird für $n = 100$ die Messfehlervarianz variiert. Die folgenden drei Abbildungen zeigen die Ergebnisse für die unterschiedlichen Fälle (1: symmetrisch, unimodal, 2: symmetrisch, bimodal, 3: asymmetrisch). Es zeigen sich mehrere Unterschiede zwischen den drei Fällen: Bei (2) ist der Median weitestgehend invariant gegenüber einer Vergrößerung der Messfehlervarianz. Auch $\hat{\beta}_{normal}$ liefert hier sehr ordentliche Resultate. In den anderen beiden Fällen nimmt die Genauigkeit von $\hat{\beta}_{cor}$ und $\hat{\beta}_{normal}$ mit zunehmendem σ ab. Bei (3) schneidet $\hat{\beta}_{normal}$ dann sogar schlechter ab als der naive Schätzer. Auch die Anzahl an Wiederholungen, in denen kein Problem auftrat, variiert zwischen den drei Fällen: Bei (3) wird für $\hat{\beta}_{normal}$ bei großer Varianz in mehr als der Hälfte aller Wiederholungen kein Ergebnis erhalten, was fast immer auf eine nicht invertierbare Jacobi-Matrix zurückzuführen ist. Die Schätzung von $\hat{\beta}_{cor}$ ist in (3) allerdings stabiler als in (1). In (2) wird sowohl für $\hat{\beta}_{cor}$ als auch für $\hat{\beta}_{normal}$ am häufigsten ein Ergebnis erhalten.

²¹Damit ist gemeint, dass b unabhängig von der Varianz der untrunkierten Verteilung z.B. immer der fünffachen Standardabweichung entspricht.

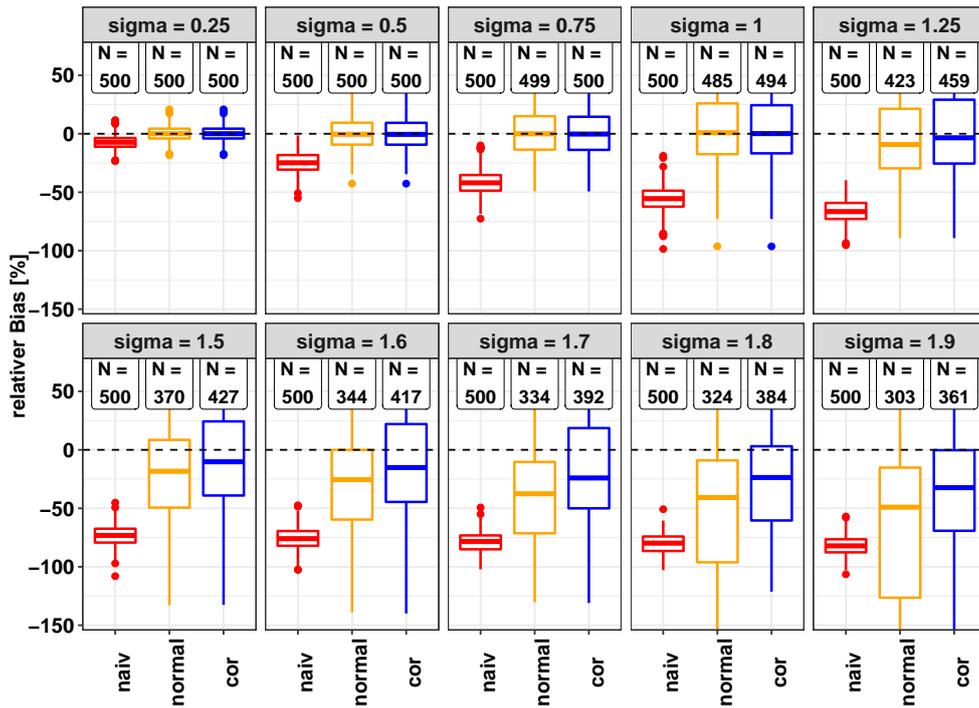


Abb. 17: Mischverteilung, Fall 1 (symm. unimodal): verschiedene Standardabweichungen von U_i mit $n = 100$

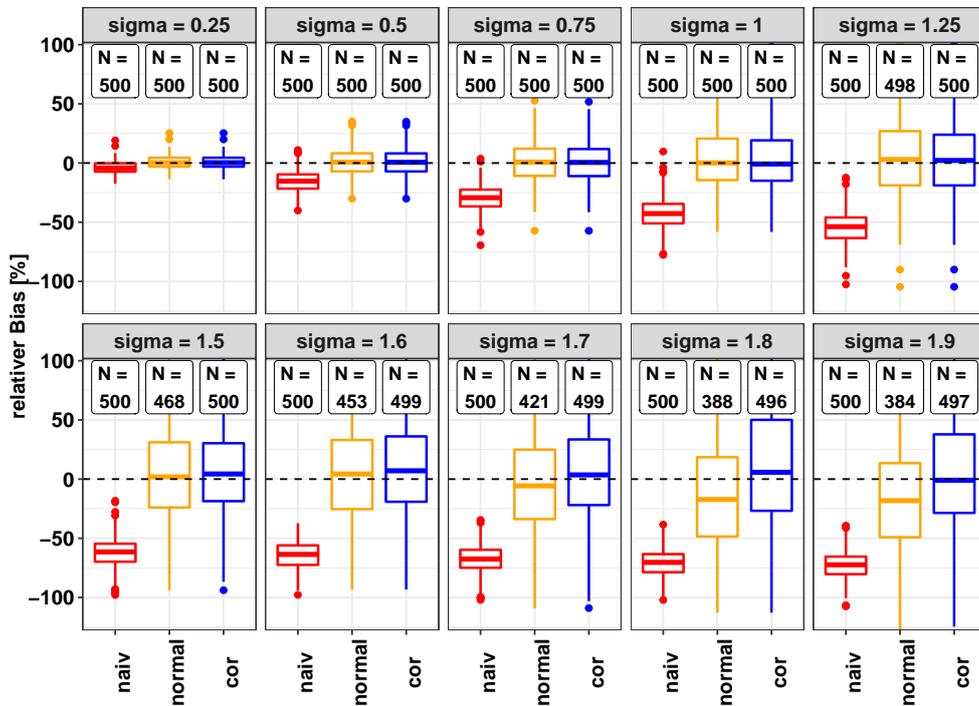


Abb. 18: Mischverteilung, Fall 2 (symm. bimodal): verschiedene Standardabweichungen von U_i mit $n = 100$

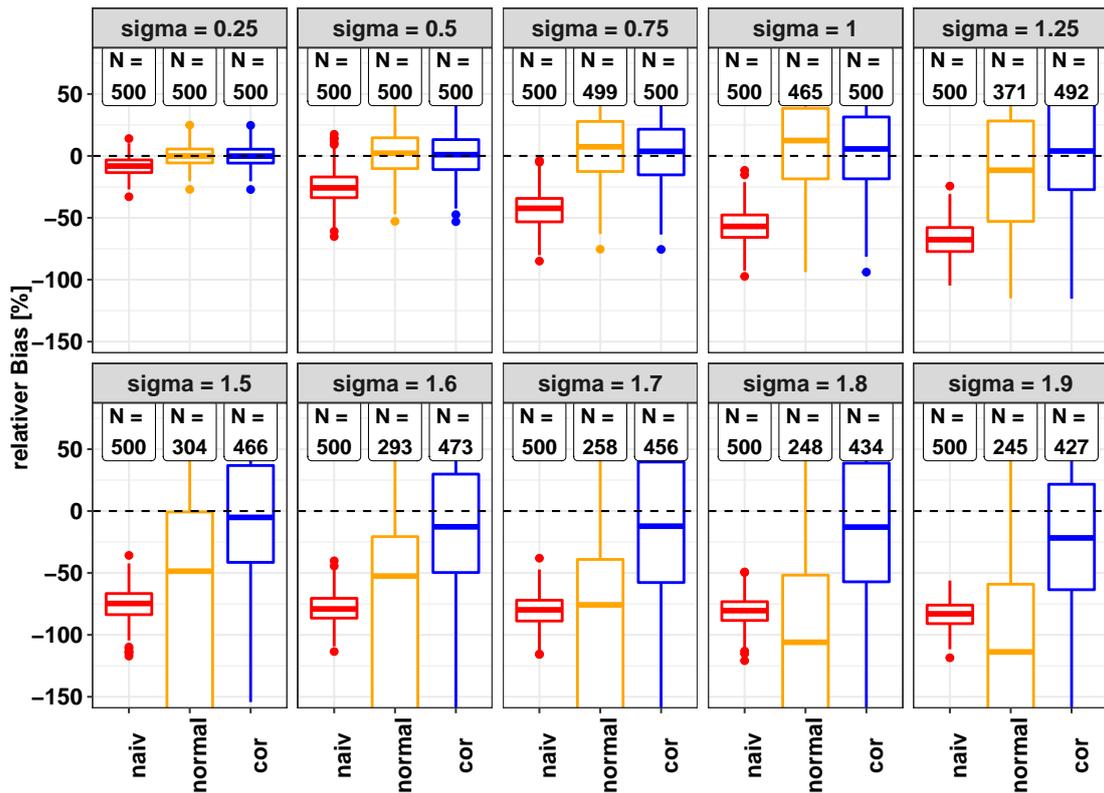


Abb. 19: Mischverteilung, Fall 3 (asymm.): verschiedene Standardabweichungen von U_i mit $n = 100$

Die Streuung der Schätzungen $\hat{\beta}_{cor}$ unterscheidet sich nur geringfügig, während die Varianz von $\hat{\beta}_{normal}$ in (3) größer ist als bei den anderen Fällen. Verglichen mit der Laplace-Verteilung (Abb. 11) ist die Varianz von $\hat{\beta}_{cor}$ bei den Mischverteilungen kleiner und der Bias deutlich geringer. Bei der schiefen Normalverteilung wird eine ähnliche Genauigkeit erzielt, allerdings treten dort häufiger Probleme bei der Bestimmung von $\hat{\beta}_{cor}$ auf als hier (vgl. Abb. 5 und 6).

4.5.2. Unterschiedliche Stichprobenumfänge

Nun wird der Stichprobenumfang n variiert. Da eine Varianz von $0,75^2$ und $n = 100$ im vorangegangenen Abschnitt zu stabilen Resultaten geführt hat, wird nun der Stichprobenumfang schrittweise bis $n = 20$ verkleinert. Die dabei beobachteten Unterschiede sind geringer als bei Variation der Streuung der Messfehler, weshalb hier nur das Ergebnis für Fall 1 dargestellt ist (Abb. 20, Resultate von (2) und (3) im Anhang A.6, Abb. 27, 28). Es ergeben sich auch keine Auffälligkeiten im Vergleich mit den Ergebnissen der schiefen Normalverteilung (Abb. 9) und der

Laplace-Verteilung (Abb. 13).

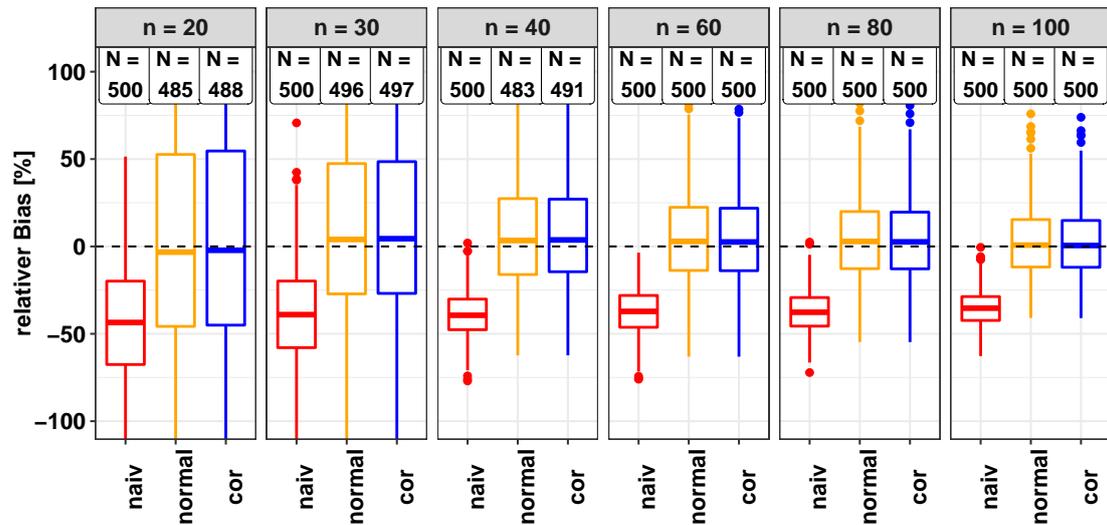


Abb. 20: Mischverteilung, Fall 1 (symm. unimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$

Abschließend wird noch das Verhalten für große Stichprobenumfänge untersucht. Die Varianz wird auf $1,7^2$ gesetzt und n bis 5000 erhöht. Fall 1 und 2 zeigen dabei nur sehr geringe Unterschiede, weshalb hier (Abb. 21, nächste Seite) nur die Ergebnisse von (1) dargestellt sind ((2) siehe Anhang A.6, Abb. 29). Abbildung 22 zeigt das Resultat von Fall 3. Während sich $\hat{\beta}_{cor}$ wie bei den anderen Fällen verhält, ergeben sich bei der Bestimmung von $\hat{\beta}_{normal}$ extrem häufig Probleme, für große n sogar noch öfter als für kleine. In nahezu allen Fällen, bei denen kein Schätzer erhalten wird, ist die Jacobi-Matrix nicht invertierbar. Der Median des relativen Bias von $\hat{\beta}_{normal}$ ist trotz der vielen NAs und den großen Ausreißern nach unten nahe bei 0.

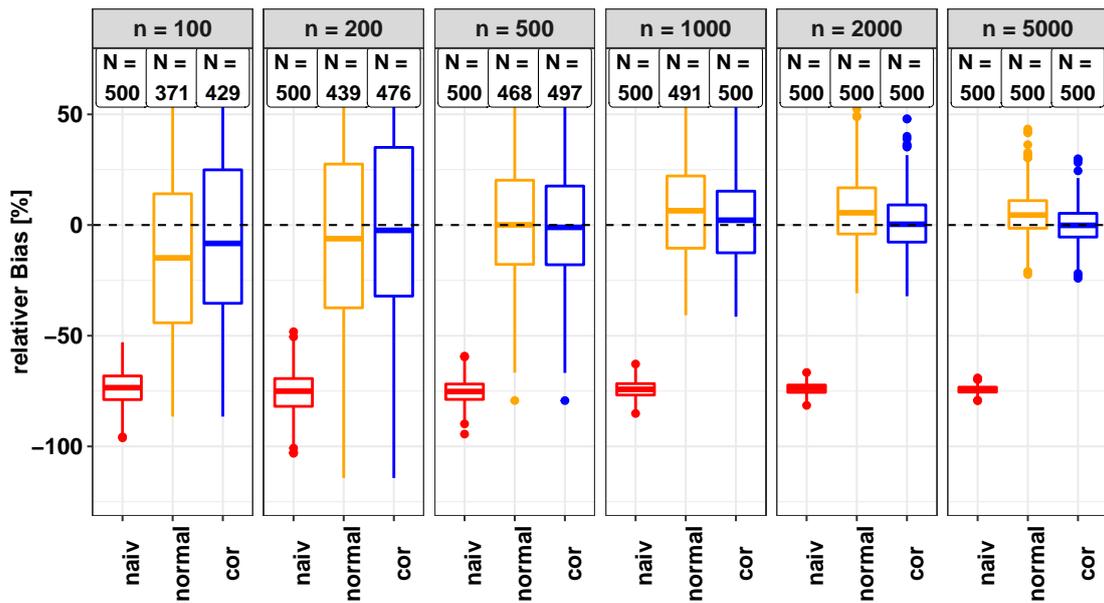


Abb. 21: Mischverteilung, Fall 1 (symm. unimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$

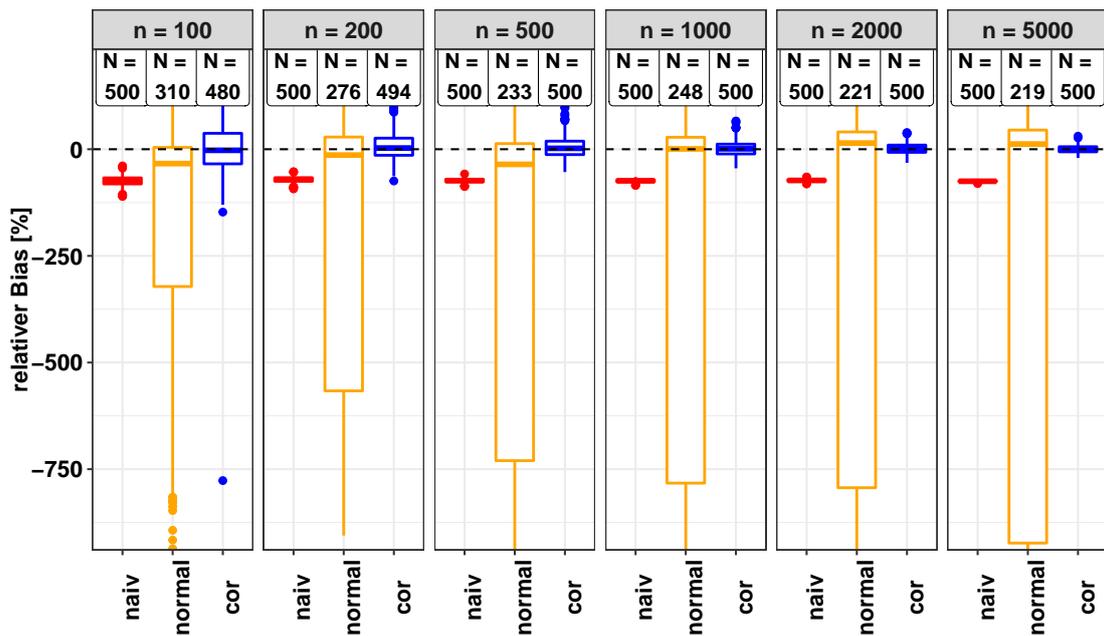


Abb. 22: Mischverteilung, Fall 3 (asymm.): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$

4.6. Zusammenfassung

Die Ergebnisse der Simulation geben Aufschluss über die Qualität der Methode der korrigierten Scorefunktionen für endliche Stichprobenumfänge: Selbst mit $n = 100$ oder kleiner lassen sich bei moderaten Stichprobenvarianzen meist schon gute Ergebnisse erzielen. Für $n = 5000$ geht der Median der Abweichung des korrigierten Schätzers vom ML-Schätzer in allen hier betrachteten Fällen selbst für eine große Messfehlervarianz gegen 0, was für die Unverzerrtheit der Schätzer spricht.

Die Korrektur funktioniert je nach Messfehlerverteilung unterschiedlich gut: Bei Laplace-verteilten Messfehlern hat die korrigierte Schätzung $\hat{\beta}_{cor}$ eine hohe Varianz und ist häufig verzerrt. Werden die Messfehler stattdessen aus einer trunkierten Verteilung gezogen, ergeben sich genauere Schätzungen. Allerdings treten hier bei geringer Trunkierung durch Nicht-Invertierbarkeit der Jacobi-Matrix häufig Probleme bei der Schätzung auf. Selbiges gilt für schief normalverteilte Messfehler mit hoher Varianz und hohem Schiefeparameter α . Die Genauigkeit der Schätzung (d.h. der Median der relativen Abweichung vom ML-Schätzer) ist dagegen weitestgehend invariant gegenüber Veränderungen von α .

Bei der Schätzung von $\hat{\beta}_{normal}$ werden anstatt der wahren Messfehlerverteilung normalverteilte Messfehler angenommen. Die Ergebnisse zeigen, dass sich trotz der Annahmenverletzung teilweise gute Schätzungen ergeben. Auch hier hängt die Qualität der Resultate und die Häufigkeit des Auftretens von Problemen stark von der wahren Messfehlerverteilung ab. Auffällig sind die großen Unterschiede zwischen den betrachteten Mischverteilungen: Die Ergebnisse bei Messfehlern aus einer asymmetrischen Mischverteilung sind deutlich schlechter als bei den betrachteten symmetrischen Verteilungen und die Nicht-Invertierbarkeit der Jacobi-Matrix führt dazu, dass teilweise in nicht einmal der Hälfte aller Fälle ein Ergebnis erhalten wird. Dennoch führt die Schätzung von $\hat{\beta}_{normal}$ hier und bei den anderen Messfehlerverteilungen meistens zu einer (häufig sogar großen) Reduktion des Bias der naiven Schätzung. Der naive Schätzer ist teilweise stark verzerrt, hat aber eine deutlich geringere Varianz als die korrigierten Schätzer.

5. Fazit & Ausblick

Das Ziel dieser Arbeit war es zunächst zu untersuchen, inwieweit sich existierende Ergebnisse und Methoden zu korrigierten Scorefunktionen, bei denen normalverteilte Messfehler angenommen werden, verallgemeinern lassen. Die Herleitungen der exakten korrigierten Scorefunktionen in Kapitel 2 haben gezeigt, dass die Korrektur bei den dort betrachteten generalisierten Regressionsmodellen für allgemeine Messfehlerverteilungen analog zur Korrektur unter normalverteilten Messfehlern funktioniert. Eine Bedingung, die dabei (außer im linearen Regressionsmodell) an die Messfehlerverteilung gestellt werden muss, ist die Existenz der Momenterzeugenden Funktion $M_{U_i}(\beta) = \mathbb{E}(\exp(\beta^t U_i))$. Dadurch kann der Term $\exp(\beta^t W_i) = \exp(\beta^t X_i) * \exp(\beta^t U_i)$ in der naiven log-Likelihood und Scorefunktion erwartungstreu korrigiert werden. Die korrigierte Scorefunktion bietet dann ein breites Anwendungsfeld, da keinerlei Annahmen über die Verteilung der wahren X_i getroffen werden müssen.

Andererseits gibt es für den Fall, dass sich die korrigierte Scorefunktion nicht ohne Weiteres herleiten lässt, keine allgemeine Methode zur Konstruktion der korrigierten Scorefunktion bei nichtnormalverteilten Messfehlern, da sich beispielsweise die Monte Carlo Corrected Score Methode nicht direkt auf allgemeine Messfehlerverteilungen übertragen lässt (siehe Anhang A.1).

Die Ergebnisse der Simulation in Kapitel 4 zeigen, dass die Methode schon für kleine Stichprobenumfänge in vielen Fällen gut funktioniert, wobei die Performance sich zwischen den verschiedenen Messfehlerverteilungen unterscheidet. Zudem lassen die Resultate vermuten, dass auch bei einer Verletzung der Verteilungsannahme beziehungsweise mangelndem Wissen über die wahre Verteilung der Messfehler die Schätzung unter der Annahme von normalverteilten Messfehlern ein sinnvoller Ansatz zur Reduktion des Bias der naiven Schätzung sein kann.

Interessant wäre es sicherlich, die Simulation noch auszuweiten. Vor allem die Mischverteilungen bieten viele Möglichkeiten zur Variation: Beispielsweise könnte die Asymmetrie systematisch verändert und die Auswirkungen auf die Schätzung untersucht werden. Bei der schiefen Normalverteilung hat eine Veränderung von α nahezu keine Folgen auf die Güte der Schätzung gehabt. Offen ist, ob dies bei asymmetrischen Mischverteilungen ebenfalls so ist.

Auch die trunkeerte Laplace-Verteilung könnte genauer untersucht werden. Hier

könnte z.B. β oder die Messfehlervarianz so groß gewählt werden, dass die Momenterzeugende Funktion der untrunkierten Verteilung für das wahre β gar nicht existiert. Von Interesse ist, ob dann (ggf. je nach Stärke der Trunkierung) mit der trunkierten Verteilung dennoch gute Ergebnisse erzielt werden können, zumindest für sehr große Stichprobenumfänge.

Ebenso könnte der wahre Parameter β_t verändert werden, um Y_i in einer anderen Größenordnung zu generieren. Auch die Verteilung, aus denen die X_i gezogen werden, kann variiert werden. Bei Nakamura (1990) werden die X_i beispielsweise aus einer Gleichverteilung gezogen. Außerdem könnte die Simulation auch für eine Gamma-verteilte Zielvariable umgesetzt werden.

Die Ergebnisse der Simulation bei Laplace-verteilten Messfehlern waren am schlechtesten und auch bei der trunkierten Laplace-Verteilung war die Performance umso schlechter, je geringer die Trunkierung. Interessant wäre es, Messfehler aus trunkierten *heavy-tail*-Verteilungen zu simulieren. Womöglich ergeben sich Parallelen zu den Resultaten der (trunkierten) Laplace-Verteilung.

Die Schätzung der Varianz wurde hier nur am Rande betrachtet. Man könnte die Berechnung, wenn gleichzeitig die Varianz der Messfehler geschätzt wird, genauer untersuchen. Um die Güte der Schätzung der Varianz zu überprüfen, kann wie bei Nakamura (ebenda) ebenfalls eine Simulation durchgeführt werden.

Zuletzt bietet die logistische Regression noch viele offene Fragen: Grundsätzlich müsste sich das Verfahren zur approximativen Korrektur von Buzas (2009, siehe Anhang A.2) auf nichtnormalverteilte Messfehler übertragen lassen. Die Anwendung der Monte Carlo Corrected Score Methode ist selbst bei normalverteilten Messfehlern nur approximativ, liefert aber in der Praxis sehr gute Ergebnisse. Womöglich ist dies auch bei der Verwendung bei nichtnormalverteilten Messfehlern der Fall, was durch eine Simulation überprüft werden könnte.

A. Anhang

A.1. Monte Carlo Corrected Score

Mit der von Novick und Stefanski 2002 (siehe auch Carroll et al. 2006, Kapitel 7.2, 7.4) vorgestellten Monte Carlo (MC) Corrected Score Methode lassen sich allgemein korrigierte Scorefunktionen konstruieren. Das Vorgehen soll hier genauer beschrieben werden.

Bei diesem Verfahren werden normalverteilte, additive Messfehler angenommen. Die beobachteten Variablen W_j ergeben sich somit aus den wahren X_j wie folgt²²:

$$W_j = X_j + \Sigma^{\frac{1}{2}}U_j, \quad U_j \sim N_p(0, I_p), \quad W_j, X_j, U_j \in \mathbb{R}^p \quad (\text{A.1})$$

Nun werden komplexe Zufallsvariablen $\tilde{W}_{k,j}$ generiert, wobei $Z_{k,j}$ normalverteilt ist mit der Einheitsmatrix I_p als Kovarianz und i die imaginäre Einheit darstellt:

$$\tilde{W}_{k,j} = W_j + i\Sigma^{\frac{1}{2}}Z_{k,j}, \quad Z_{k,j} \sim N_p(0, I_p), \quad i = \sqrt{-1} \quad (\text{A.2})$$

Unter der Voraussetzung, dass die Scorefunktion $s_{X_j}(\beta, Y_j, X_j)$ im Komplexen eine ganze Funktion der Komponente X_j ist, gilt, dass der Realteil von $s_{X_j}(\beta, Y_j, \tilde{W}_{k,j})$ eine korrigierte Scorefunktion ist, d.h.

$$\mathbb{E}\left(\text{Re}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j}))|Y_j, X_j\right) = s_{X_j}(\beta, Y_j, X_j) \quad (\text{A.3})$$

(Novick und Stefanski 2002, Stefanski 1989). Für den Fall $p = 1$ kann dies wie folgt verifiziert werden (mit $\Sigma^{\frac{1}{2}} = \sigma$, d.h. $W_j = X_j + \sigma U_j$, analog für $\tilde{W}_{k,j}$):

Da $s_{X_j}(\beta, Y_j, X_j)$ eine ganze Funktion ist, kann $s_{X_j}(\beta, Y_j, \tilde{W}_{k,j})$ als Taylorreihenentwicklung dargestellt werden:

$$s_{X_j}(\beta, Y_j, \tilde{W}_{k,j}) = \sum_{m=0}^{\infty} \frac{s_{X_j}^{(m)}(\beta, Y_j, X_j)}{m!} (\tilde{W}_{k,j} - X_j)^m \quad (\text{A.4})$$

wobei $s_{X_j}^{(m)}(\beta, Y_j, X_j)$ die m -te Ableitung von s_{X_j} nach X_j ist. Für $\mathbb{E}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j})|Y_j, X_j)$ gilt somit (unter der Annahme, dass die Erwartungsbildung und die unendliche Summe vertauschbar sind):

$$\mathbb{E}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j})|Y_j, X_j) = \sum_{m=0}^{\infty} \frac{s_{X_j}^{(m)}(\beta, Y_j, X_j)}{m!} \mathbb{E}((\tilde{W}_{k,j} - X_j)^m|Y_j, X_j) \quad (\text{A.5})$$

²²Um Verwechslung mit der imaginären Einheit i zu vermeiden, wird in diesem Abschnitt j statt i als Index der X -, W - und Y -Variablen benutzt.

Da $\tilde{W}_{k,j} = X_j + \sigma U_j + i\sigma Z_{k,j}$, gilt für $\mathbb{E}((\tilde{W}_{k,j} - X_j)^m | Y_j, X_j)$:

$$\mathbb{E}((\tilde{W}_{k,j} - X_j)^m | Y_j, X_j) = \mathbb{E}((\sigma U_j + i\sigma Z_{k,j})^m) = \sigma^m \mathbb{E}((U_j + iZ_{k,j})^m) \quad (\text{A.6})$$

(Bedingen auf Y_j und X_j ist nicht nötig, da U_j und $Z_{k,j}$ unabhängig von Y und X sind.) U_j und $Z_{k,j}$ sind zwei voneinander unabhängige, standardnormalverteilte Zufallsvariablen, weshalb für die komplexe Variable $U_j + iZ_{k,j}$ gilt:

$$\mathbb{E}((U_j + iZ_{k,j})^m) = 0 \quad \text{für } m = 1, 2, \dots \quad (\text{A.7})$$

(Stefanski 1989). In Gleichung (A.5) sind somit alle Terme mit $m \geq 1$ gleich 0 und für $m = 0$ bleibt $s_{X_j}(\beta, Y_j, X_j)$ stehen. Daraus folgt $\mathbb{E}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j}) | Y_j, X_j) = s_{X_j}(\beta, Y_j, X_j)$ und da $s_{X_j}(\beta, Y_j, X_j)$ nicht komplex ist (d.h. der Imaginärteil ist 0) gilt $\mathbb{E}(\text{Re}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j})) | Y_j, X_j) = s_{X_j}(\beta, Y_j, X_j)$.

Die korrigierte Scorefunktion $\text{Re}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j}))$ hängt vom Zufallsvektor $Z_{k,j}$ ab. Die durch $Z_{k,j}$ induzierte Variabilität sollte eliminiert werden, weshalb $\mathbb{E}(s_{X_j}(\beta, Y_j, W_j + i\Sigma^{\frac{1}{2}} Z_{k,j}) | Y_j, W_j)$ als korrigierte Scorefunktion gegenüber $\text{Re}(s_{X_j}(\beta, Y_j, W_j + i\Sigma^{\frac{1}{2}} Z_{k,j}))$ zu bevorzugen wäre. Da dieser Erwartungswert allerdings teilweise schwer zu bestimmen ist, kann er durch eine Monte Carlo Simulation approximiert werden: Der sogenannte *Monte Carlo Corrected Score* lautet

$$s_{K,W_j}(\beta, Y_j, W_j) = \frac{1}{K} \sum_{k=1}^K \text{Re} \left(s_{X_j}(\beta, Y_j, W_j + i\Sigma^{\frac{1}{2}} Z_{k,j}) \right) \quad (\text{A.8})$$

$s_{K,W_j}(\beta, Y_j, W_j)$ ist eine korrigierte Scorefunktion und konvergiert für $K \rightarrow \infty$ gegen die präferierte korrigierte Scorefunktion $\mathbb{E}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j}) | Y_j, W_j)$ (Novick und Stefanski 2002):

$$s_{\infty,W_j}(\beta, Y_j, W_j) = \lim_{K \rightarrow \infty} s_{K,W_j}(\beta, Y_j, W_j) = \mathbb{E} \left(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j}) | Y_j, W_j \right) \quad (\text{A.9})$$

Die in Novick und Stefanski (ebenda) hergeleiteten $s_{\infty,W_j}(\beta, Y_j, W_j)$ für die lineare und die Poisson-Regression stimmen mit den exakten korrigierten Scorefunktionen aus Kapitel 2.2 (unter Annahme von normalverteilten Messfehlern) überein.

Im logistischen Regressionsmodell dagegen ist die Annahme der MC Corrected Score Methode, dass die Scorefunktion eine ganze Funktion im Komplexen ist,

nicht erfüllt (dies ist auch der Grund, weshalb i.A. keine exakte korrigierte Scorefunktion für dieses Modell existiert, siehe Abschnitt 2.2.5). Im Logit-Modell ist weder $s_{K,W_j}(\beta, Y_j, W_j)$ noch $s_{\infty,W_j}(\beta, Y_j, W_j)$ eine korrigierte Scorefunktion (Novick und Stefanski 2002). Das Anwenden der MC Corrected Score Methode führt dennoch zu einer beinahe unverzerrten Schätzung (Carroll et al. 2006, Kapitel 7.2). Folglich liefert die MC Corrected Score Methode hier trotz einer Verletzung der theoretischen Annahmen in der Praxis gute Ergebnisse (Novick und Stefanski 2002).

Wie oben erwähnt, werden bei der MC Corrected Score Methode normalverteilte Messfehler angenommen. Wenn U_j und $Z_{k,j}$ standardnormalverteilt sind, ist $\mathbb{E}((U_j + iZ_{k,j})^m) = 0$ für $m \geq 1$. Für nichtnormalverteilte Messfehler gilt dies im Allgemeinen nicht: Seien U und Z zwei identisch verteilte, unabhängige Zufallsvariablen mit Erwartungswert 0 und Varianz σ^2 und sei $\mathcal{Z} = U + iZ$ eine komplexe Zufallsvariable. Unter diesen Annahmen gilt beispielsweise für $\mathbb{E}(\mathcal{Z}^4)$:

$$\begin{aligned} \mathbb{E}(\mathcal{Z}^4) &= \mathbb{E}(U^4 + 4 * U^3 iZ + 6 * U^2 i^2 Z^2 + 4 * U i^3 Z^3 + i^4 Z^4) \\ &= \mathbb{E}(U^4) + 6 * \mathbb{E}(U^2 i^2 Z^2) + \mathbb{E}(i^4 Z^4) \\ &= \mathbb{E}(U^4) - 6 * \mathbb{E}(U^2) * \mathbb{E}(Z^2) + \mathbb{E}(Z^4) = 2 * \mathbb{E}(U^4) - 6 * (\sigma^2)^2 \quad (\text{A.10}) \end{aligned}$$

Während für die Normalverteilung mit Kurtosis 3 somit $\mathbb{E}(\mathcal{Z}^4) = 0$ gilt²³, ist dies für Verteilungen mit anderer Wölbung nicht der Fall²⁴.

Dadurch, dass für nichtnormalverteilte Messfehler die Eigenschaft $\mathbb{E}((U_j + iZ_{k,j})^m) = 0$ i.A. nicht für alle $m \geq 1$ erfüllt ist, fallen (anders als bei normalverteilten Messfehlern) in Gleichung (A.5) nicht alle Summanden mit $m \geq 1$ weg. Folglich ist die Eigenschaft $\mathbb{E}(s_{X_j}(\beta, Y_j, \tilde{W}_{k,j}) | Y_j, X_j) = s_{X_j}(\beta, Y_j, X_j)$ nicht erfüllt, weshalb $s_{X_j}(\beta, Y_j, \tilde{W}_{k,j})$ und somit auch der Monte Carlo Corrected Score $s_{K,W_j}(\beta, Y_j, W_j)$ (mit $\tilde{W}_{k,j} = W_j + Z_{k,j}$; $Z_{k,j}$ identisch verteilt wie U_j) i.A. keine korrigierten Scorefunktion sind.

²³Kurtosis = $\frac{\mathbb{E}((U - \mathbb{E}(U))^4)}{\mathbb{E}((U - \mathbb{E}(U))^2)^2}$ (siehe z.B. Boos und Stefanski 2013, Kapitel 1.5.2) d.h. für $U \sim N(0, \sigma^2)$ gilt (da Kurtosis=3): $\mathbb{E}(U^4) = 3 * (\mathbb{E}(U^2))^2 = 3 * (\sigma^2)^2$

²⁴Analog lässt sich zum Beispiel verifizieren, dass $\mathbb{E}(\mathcal{Z}^3) = 0$ nur für symmetrische Verteilungen (d.h. mit $\mathbb{E}(U^3) = 0$) gilt.

A.2. Korrekturmöglichkeiten logistische Regression

Wie in Abschnitt 2.2.5 dargelegt, existiert bei der Logit-Regression nicht in jedem Fall eine exakte korrigierte Scorefunktion. Hier werden zwei der in 2.2.5 skizzierten Methoden etwas ausführlicher erläutert:

Wenn Y ein seltenes Ereignis ist ($\sup(P(Y_i = 1|X_i)) < 0,5$), lässt sich wie in 2.2.5 erwähnt eine korrigierte Scorefunktion herleiten. Buzas und Stefanski (1996) präsentieren in ihrem Artikel eine Methode zur allgemeinen Konstruktion einer korrigierten Scorefunktion für den Fall, dass sich die Responsefunktion $G(\beta^t X_i) = \mathbb{E}(Y_i|X_i)$ als Reihenentwicklung von $\exp(\beta^t X_i)$ darstellen lässt und die wahre Scorefunktion die Form $s_{X_i}(\beta, Y_i, X_i) = X_i(Y_i - G(\beta^t X_i))$ hat²⁵.

Im Logit-Modell ist die Responsefunktion die Verteilungsfunktion der logistischen Verteilung (hier mit F bezeichnet, $F(\beta^t X_i) = (1 + \exp(-\beta^t X_i))^{-1}$). Aus der Annahme $\sup(P(Y_i = 1|X_i)) < 0,5$ folgt $P(Y_i = 1|X_i) = F(\beta^t X_i) < 0,5$ für alle X_i , was äquivalent ist zu $\exp(\beta^t X_i) < 1$ (bzw. $\beta^t X_i < 0$) für alle X_i . Durch $\exp(\beta^t X_i) < 1$ ist folgende Reihendarstellung von F möglich ($\beta^t X_i = \eta_i$):

$$F(\eta_i) = \frac{1}{1 + \exp(-\eta_i)} = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \sum_{k=1}^{\infty} (-1)^{k-1} (\exp(\eta_i))^k \quad (\text{A.11})$$

(ebenda). Dies kann über eine geometrische Reihe hergeleitet werden:

$$\begin{aligned} \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} &= \exp(\eta_i) * \left(\frac{1}{1 - (-\exp(\eta_i))} \right)^{|-\exp(\eta_i)| < 1} = \exp(\eta_i) * \sum_{k=0}^{\infty} (-\exp(\eta_i))^k \\ &= \sum_{k=0}^{\infty} (-1)^k (\exp(\eta_i))^{k+1} = \sum_{k=1}^{\infty} (-1)^{k-1} (\exp(\eta_i))^k \end{aligned} \quad (\text{A.12})$$

Mit den Formeln von Buzas und Stefanski kann dann die korrigierte Scorefunktion aufgestellt werden. Die Autoren konzentrieren sich dabei auf den Fall von normalverteilten Messfehlern, das Vorgehen lässt sich aber auf nichtnormalverteilte Messfehler übertragen (ebenda, Abschnitt 6).

Das Herleiten einer exakten korrigierten Scorefunktion ist hier nur durch die Restriktion, dass Y ein seltenes Ereignis ist, möglich. Ohne diese Einschränkung ist es dennoch möglich, eine konsistente Schätzung zu erhalten:

Bei der Methode von Buzas (2009) wird nicht die wahre, sondern eine appro-

²⁵Dies ist bei der logistischen Regression der Fall, siehe Fußnote Seite 15.

ximative Scorefunktion korrigiert. Die Scorefunktion in einem Modell mit binärer Zielvariable kann allgemein wie folgt dargestellt werden (Buzas 2009):

$$s_{X_i}(\beta, Y_i, X_i) = h(\beta^t X_i) * (Y_i - G(\beta^t X_i)) * X_i \quad (\text{A.13})$$

mit $G(\beta^t X_i) = \mathbb{E}(Y_i | X_i)$, $h(\eta) = \frac{G'(\eta)}{G(\eta)(1 - G(\eta))}$ und $G'(\eta) = \frac{\delta}{\delta \eta} G(\eta)$. Im Logit-Modell mit $G(\eta) = F(\eta) = (1 + \exp(-\eta))^{-1}$ ist $h(\eta) = 1$. Zum Aufstellen einer approximativen Scorefunktion im logistischen Regressionsmodell wird folgende Approximation benutzt: $F(\eta) \approx \Phi(\lambda \eta)$, wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist²⁶. Die approximative Scorefunktion $s_{X_i}^A(\beta, Y_i, X_i)$ ergibt sich, indem $F'(\eta)$ in $h(\eta)$ durch $\Phi'(\lambda \eta) = \lambda \phi(\lambda \eta)$ ersetzt wird (ϕ ist die Dichte der Standardnormalverteilung):

$$s_{X_i}^A(\beta, Y_i, X_i) = \frac{\lambda \phi(\lambda \beta^t X_i)}{F(\beta^t X_i)(1 - F(\beta^t X_i))} * (Y_i - F(\beta^t X_i)) * X_i \quad (\text{A.14})$$

Für diese Funktion gilt $\mathbb{E}(s_{X_i}^A(\beta, Y_i, X_i)) = 0$. In Abwesenheit von Messfehlern ist die Nullstelle von $\sum_{i=1}^n s_{X_i}^A(\beta, Y_i, X_i)$ somit ein konsistenter Schätzer für β , die Schätzung ist allerdings weniger effizient (d.h. hat eine höhere Varianz) als die Schätzung unter Verwendung der wahren Scorefunktion (ebenda). Diese approximative Scorefunktion kann korrigiert werden, d.h. es kann eine Funktion $s_{W_i}^A(\beta, Y_i, W_i)$ mit der Eigenschaft $\mathbb{E}(s_{W_i}^A(\beta, Y_i, W_i) | Y_i, X_i) = s_{X_i}^A(\beta, Y_i, X_i)$ aufgestellt werden. Die Nullstelle der korrigierten Funktion ist (als M-Schätzer) ein konsistenter, asymptotisch normalverteilter Schätzer für β . Die konkrete Formel für $s_{W_i}^A(\beta, Y_i, W_i)$ unter normalverteilten Messfehlern ist in Buzas (ebenda) zu finden²⁷. Eine approximative Scorefunktion und deren Korrektur für den Fall der Probit-Regression ($G(\eta) = \Phi(\eta)$) wird dort ebenfalls hergeleitet. Auch hier wird auf die Approximation zwischen den Verteilungsfunktionen der logistischen und der Normalverteilung zurückgegriffen.

²⁶Für $\lambda = 1, 7^{-1}$ wird der maximale Abstand zwischen den beiden Funktionen minimiert. λ kann für gegebene Daten aber auch so bestimmt werden, dass die Varianz (genauer: die Determinante der asymptotischen Kovarianzmatrix) bei der Schätzung von β mit $s_{X_i}^A(\beta, Y_i, X_i)$ minimiert wird (Buzas 2009).

²⁷Dabei ist eine Restriktion nötig: $|\beta_X| < \sqrt{1/(\lambda^2 \sigma^2)}$ (für den Fall von genau einer messfehlerbehafteten Variable $X \in \mathbb{R}$, β_X ist der zugehörige Koeffizient und σ^2 die Varianz der (normalverteilten) Messfehler).

A.3. Momenterzeugende Funktion Laplace-Verteilung

In diesem Abschnitt wird die Momenterzeugende Funktion der Laplace-Verteilung (siehe Gleichung (3.15)) hergeleitet und gezeigt, dass der Erwartungswert $\mathbb{E}(\exp(\beta U))$ für $|\beta| \geq \frac{1}{s}$ unendlich ist (d.h. die Momenterzeugende Funktion existiert nicht für $|\beta| \geq \frac{1}{s}$). Außerdem wird die Stetigkeit der Momenterzeugenden Funktion einer trunkeierten Laplace-Verteilung an den Stellen $\beta = 1/s$ und $\beta = -1/s$ gezeigt.

Zunächst wird eine Standard-Laplace-verteilte Zufallsvariable U betrachtet (d.h. $\mu = 0, s = 1$) und die Ergebnisse anschließend verallgemeinert.

$$\begin{aligned} \mathbb{E}(\exp(\beta U)) &= \int_{-\infty}^{\infty} \exp(\beta u) f_U(u) du = \int_{-\infty}^{\infty} \exp(\beta u) * 0,5 * \exp(-|u|) du \\ &= 0,5 * \left(\int_{-\infty}^0 \exp(\beta u) * \exp(u) du + \int_0^{\infty} \exp(\beta u) * \exp(-u) du \right) \\ &= 0,5 * \left(\int_{-\infty}^0 \exp(u(\beta + 1)) du + \int_0^{\infty} \exp(u(\beta - 1)) du \right) \quad (\text{A.15}) \end{aligned}$$

Für $\beta \neq 1$ und $\beta \neq -1$ gilt dann:

$$\mathbb{E}(\exp(\beta U)) = 0,5 * \left(\left[\frac{1}{\beta + 1} \exp(u(\beta + 1)) \right]_{u=-\infty}^{u=0} + \left[\frac{1}{\beta - 1} \exp(u(\beta - 1)) \right]_{u=0}^{u=\infty} \right) \quad (\text{A.16})$$

Für $\beta \in (-1, 1)$ ist $\beta + 1$ stets größer als 0, wodurch $\lim_{u \rightarrow -\infty} \exp(u(\beta + 1)) = 0$. Analog ist $\beta - 1$ immer kleiner als 0. Dadurch ist $\lim_{u \rightarrow \infty} \exp(u(\beta - 1)) = 0$. Für $|\beta| < 1$ ergibt sich somit:

$$\begin{aligned} M_U(\beta) &= 0,5 * \left(\frac{1}{\beta + 1} * (\exp(0) - 0) + \frac{1}{\beta - 1} * (0 - \exp(0)) \right) \\ &= 0,5 * \left(\frac{1}{\beta + 1} - \frac{1}{\beta - 1} \right) = 0,5 * \left(\frac{\beta - 1}{\beta^2 - 1^2} - \frac{\beta + 1}{\beta^2 - 1^2} \right) \\ &= 0,5 * \left(\frac{-2}{\beta^2 - 1} \right) = 0,5 * \left(\frac{2}{1 - \beta^2} \right) = \frac{1}{1 - \beta^2} \quad (\text{A.17}) \end{aligned}$$

Dies entspricht der Formel (3.15) mit $\mu = 0$ und $s = 1$.

Für den Fall $|\beta| > 1$ wird (A.16) erneut genauer betrachtet: Für $\beta < -1$ ist

$$\lim_{u \rightarrow -\infty} \frac{1}{\beta + 1} * (\exp(u(\beta + 1))) = -\infty \quad (\text{A.18})$$

da $(\beta + 1) < 0$ und $\lim_{u \rightarrow -\infty} \exp(u(\beta + 1)) = \infty$ (da $\lim_{u \rightarrow -\infty} u(\beta + 1) = \infty$). Der Term $[\frac{1}{\beta + 1} \exp(u(\beta + 1))]_{u=-\infty}^{u=0}$ geht somit gegen unendlich. Der zweite Term, $[\frac{1}{\beta - 1} \exp(u(\beta - 1))]_{u=0}^{u=\infty}$, ist wie in (A.17) gleich $\frac{1}{\beta - 1} * (0 - \exp(0)) \cdot \mathbb{E}(\exp(\beta U))$ ist im Fall $\beta < -1$ folglich unendlich. Das selbe gilt für $\beta > 1$ (Verifikation analog, $\lim_{u \rightarrow \infty} \exp(u(\beta - 1)) = \infty$, da $(\beta - 1) > 0$). Für den Fall $|\beta| > 1$ existiert $M_U(\beta)$ somit nicht.

Für $\beta = -1$ ergibt sich Gleichung (A.15) zu:

$$\begin{aligned} \mathbb{E}(\exp(\beta U)) &\stackrel{(\text{A.15})}{=} 0,5 * \left(\int_{-\infty}^0 \exp(u * 0) du + \int_0^{\infty} \exp(-2u) du \right) \\ &= 0,5 * \left([u]_{u=-\infty}^{u=0} + \left[\frac{1}{-2} \exp(-2u) \right]_{u=0}^{u=\infty} \right) \\ &= 0,5 * \left(0 - \lim_{u \rightarrow -\infty} u - \frac{1}{2} (0 - \exp(0)) \right) \end{aligned} \quad (\text{A.19})$$

Für $\beta = -1$ ist $\mathbb{E}(\exp(\beta U))$ somit ebenfalls unendlich, d.h. $M_U(\beta)$ existiert nicht. Selbiges gilt für $\beta = 1$ (Herleitung analog). Die Momenterzeugende Funktion existiert folglich nur für $\beta \in (-1, 1)$.

Die Momenterzeugende Funktion im allgemeinen Fall lässt sich über Transformation von U bestimmen: Wenn $U \sim \text{Laplace}(0, 1)$ dann ist $Z = \mu + sU$ Laplace(μ, s)-verteilt (Kotz et al. 2001, Kapitel 2.1.2). Für $M_Z(\beta)$ gilt dann:

$$\begin{aligned} M_Z(\beta) &= \mathbb{E}(\exp(\beta Z)) = \mathbb{E}(\exp(\beta \mu + \beta s U)) = \exp(\beta \mu) \mathbb{E}(\exp(\beta s U)) \\ &= \exp(\beta \mu) M_U(\beta s) = \frac{\exp(\beta \mu)}{1 - \beta^2 s^2} \end{aligned} \quad (\text{A.20})$$

mit der Restriktion $|\beta s| < 1 \Leftrightarrow |\beta| s < 1 \Leftrightarrow |\beta| < 1/s$ (siehe Gleichung (3.15)).

A.4. Momenterzeugende Funktion trunkierte Laplace-Verteilung

Die Momenterzeugende Funktion einer trunkierten Laplace-Verteilung (mit $\mu = 0$) wurde in Abschnitt 3.2.3 (Gleichung 3.24) für $\beta \neq 1/s$ und $\beta \neq -1/s$ hergeleitet. Hier wird $M_{U_T}(\beta)$ an den Stellen $\beta = 1/s$ und $\beta = -1/s$ bestimmt und gezeigt, dass $M_{U_T}(\beta)$ auch an diesen Stellen stetig ist. In Gleichung (3.23) ergab sich für $M_{U_T}(\beta)$ Folgendes:

$$M_{U_T}(\beta) = \frac{1}{2s * (1 - \exp(-b/s))} \left(\int_{-b}^0 \exp(u(\beta + 1/s)) du + \int_0^b \exp(u(\beta - 1/s)) du \right) \quad (\text{A.21})$$

Für $\beta = 1/s$ und $\beta = -1/s$ ergibt sich dann:

$$\begin{aligned} M_{U_T}(1/s) &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\int_{-b}^0 \exp(u(1/s + 1/s)) du + \int_0^b \exp(u(1/s - 1/s)) du \right) \quad (\text{A.22}) \\ &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\int_{-b}^0 \exp(u * 2/s) du + \int_0^b 1 \, du \right) \\ &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\left[\frac{s}{2} \exp(u * 2/s) \right]_{u=-b}^{u=0} + [u]_{u=0}^{u=b} \right) \\ &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\frac{s}{2} * (1 - \exp(-b * 2/s)) + b \right) \end{aligned}$$

$$\begin{aligned} M_{U_T}(-1/s) &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\int_{-b}^0 \exp(u(-1/s + 1/s)) du + \int_0^b \exp(u(-1/s - 1/s)) du \right) \quad (\text{A.23}) \\ &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\int_{-b}^0 1 \, du + \int_0^b \exp(-u * 2/s) du \right) \\ &= \frac{1}{2s * (1 - \exp(-b/s))} \left([u]_{u=-b}^{u=0} + \left[-\frac{s}{2} \exp(-u * 2/s) \right]_{u=0}^{u=b} \right) \\ &= \frac{1}{2s * (1 - \exp(-b/s))} \left(b - \frac{s}{2} * (\exp(-b * 2/s) - 1) \right) \end{aligned}$$

Nun wird gezeigt, dass die Grenzwerte von $M_{U_T}(\beta)$ aus Gleichung (3.24) für $\beta \rightarrow 1/s$ und $\beta \rightarrow -1/s$ mit den soeben hergeleiteten Werten $M_{U_T}(1/s)$ und $M_{U_T}(-1/s)$ übereinstimmen, d.h. $M_{U_T}(\beta)$ ist in diesen Punkten stetig.

$$\begin{aligned}
\lim_{\beta \rightarrow 1/s} M_{U_T}(\beta) &= \lim_{\beta \rightarrow 1/s} \frac{1}{2s * (1 - \exp(-b/s))} \left(\frac{1}{\beta + 1/s} * (1 - \exp(-b(\beta + 1/s))) \right. \\
&\quad \left. + \frac{1}{\beta - 1/s} * (\exp(b(\beta - 1/s)) - 1) \right) \quad (\text{A.24}) \\
&= \frac{1}{2s * (1 - \exp(-b/s))} * \left(\frac{1}{1/s + 1/s} * (1 - \exp(-b(1/s + 1/s))) \right) \\
&\quad + \lim_{\beta \rightarrow 1/s} \frac{1}{\beta - 1/s} * (\exp(b(\beta - 1/s)) - 1) \\
&\stackrel{(*)}{=} \frac{1}{2s * (1 - \exp(-b/s))} * \left(\frac{s}{2} * (1 - \exp(-b * 2/s)) + b \right) = M_{U_T}(1/s)
\end{aligned}$$

$$\begin{aligned}
\lim_{\beta \rightarrow -1/s} M_{U_T}(\beta) &= \frac{1}{2s * (1 - \exp(-b/s))} \left(\lim_{\beta \rightarrow -1/s} \frac{1}{\beta + 1/s} * \right. \quad (\text{A.25}) \\
&\quad \left. (1 - \exp(-b(\beta + 1/s))) + \frac{1}{-1/s - 1/s} * (\exp(b(-1/s - 1/s)) - 1) \right) \\
&\stackrel{(*)}{=} \frac{1}{2s * (1 - \exp(-b/s))} \left(b - \frac{2}{s} * (\exp(-b * 2/s) - 1) \right) = M_{U_T}(-1/s)
\end{aligned}$$

Somit ist $M_{U_T}(\beta)$ auch in den Punkten $\beta = 1/s$ und $\beta = -1/s$ stetig. Bei der Herleitung wurde in (*) jeweils die Regel von de l'Hospital (siehe z.B. Deitmar 2014, Kapitel 5.3) zur Bestimmung folgender Grenzwerte verwendet:

$$\begin{aligned}
\lim_{\beta \rightarrow 1/s} \frac{\exp(b(\beta - 1/s)) - 1}{\beta - 1/s} &= \lim_{\beta \rightarrow 1/s} \frac{\frac{\delta}{\delta\beta} \exp(b(\beta - 1/s)) - 1}{\frac{\delta}{\delta\beta} \beta - 1/s} \\
&= \lim_{\beta \rightarrow 1/s} \frac{b * \exp(b(\beta - 1/s))}{1} = b \quad (\text{A.26})
\end{aligned}$$

$$\lim_{\beta \rightarrow -1/s} \frac{1 - \exp(-b(\beta + 1/s))}{\beta + 1/s} = \lim_{\beta \rightarrow -1/s} \frac{-(-b) * \exp(-b(\beta + 1/s))}{1} = b \quad (\text{A.27})$$

A.5. Umgang mit Warnmeldungen bei der Simulation

In diesem Abschnitt werden die Resultate von unterschiedlichen Vorgehensweisen beim Auftreten von Warnmeldungen oder Problemen (siehe Abschnitt 4.1.2) bei der Simulation betrachtet. Drei verschiedene Methoden werden verglichen:

- Simulation1: Beim Auftreten von Warnmeldungen (Jacobi-Matrix nicht invertierbar, keine Konvergenz) oder bei *NaN* als Funktionswert der korrigierten Scorefunktion an der Stelle $\hat{\beta}_{cor}$ oder (bei der Laplace-Verteilung) bei einem betragsmäßig zu großem Schätzer wird dem Schätzwert *NA* zugewiesen. Bei den Simulationen in Kapitel 4 wurde stets dieses Vorgehen gewählt.
- Simulation2: Alle Schätzwerte werden behalten. (In vereinzelt Fällen treten bei der Lösung der *multiroot*-Funktion dennoch *NAs* auf.)
- Simulation3: Bei Auftreten einer Warnmeldung oder bei *NaN* als Funktionswert der korrigierten Scorefunktion wird $\hat{\beta}_{cor}$ erneut geschätzt, falls $\hat{\beta}_{normal}$, d.h. die korrigierte Schätzung unter der Annahme von normalverteilten Messfehlern, existiert. Dann wird $\hat{\beta}_{normal}$ als Startwert übergeben und die maximale Anzahl an Iterationen von 100 auf 500 erhöht.

In Tabelle 1 sind die Ergebnisse von Simulation1 und Simulation2 gegenübergestellt. Die erste Spalte gibt den Maximum-Likelihood-Schätzer für β_1 an, die beiden folgenden Spalten den Median von $\hat{\beta}_{cor,1}$. *NA_1* und *NA_2* geben an, wie viele *NAs* unter den jeweils 500 Wiederholungen aufgetreten sind. Die weiteren Spalten beziehen sich auf die Messfehlerverteilung, -standardabweichung sowie ggf. Parameter dieser Verteilung. n ist stets 100, die X_i und die Y_i wurden wie in Abschnitt 4.2 generiert. Es zeigt sich, dass die Mediane der beiden Vorgehensweisen nur gering voneinander abweichen. Selbst beispielsweise bei der letzten Simulation (Zeile 18) ergeben sich keine großen Unterschiede, obwohl bei Simulation1 in beinahe der Hälfte der Fälle kein Schätzwert erhalten wurde.

Tabelle 2 zeigt den Vergleich von Simulation1 und Simulation3. Bei der schiefen Normalverteilung und den Mischverteilungen ergeben sich keine größeren Unterschiede bei der Anzahl *NAs*²⁸. Deswegen werden hier nur die Ergebnisse von Laplace-verteilten Messfehlern gezeigt. Da $\hat{\beta}_{normal}$ meist deutlich näher am ML-Schätzer liegt als die naive Schätzung (vgl. z.B. Abb. 11), könnte man erwarten,

²⁸d.h. wenn z.B. beim ersten Versuch ein Problem auftrat, dann war dies häufig auch beim zweiten Versuch mit anderem Startwert und mehr Iterationen der Fall. Außerdem wird immer dann nicht erneut geschätzt, wenn $\hat{\beta}_{normal}$ nicht existiert.

dass $\hat{\beta}_{normal}$ als Startwert zu verbesserten Schätzungen führt. Dies ist hier allerdings nicht der Fall, die Ergebnisse von Simulation3 sind sogar ein wenig schlechter.

	ML	Sim1	Sim2	NA_1	NA_2	Distr.	σ	Fall	α
1	0.2530	0.1614	0.1637	44	3	laplace	1.25		
2	0.2825	-0.3314	-0.3108	59	4	laplace	1.70		
3	0.2640	-0.2936	-0.2918	67	4	laplace	1.90		
4	0.2831	0.2820	0.2820	31	12	mixture	1.25	1	
5	0.3362	0.2803	0.2804	71	26	mixture	1.70	1	
6	0.2286	0.1691	0.1621	101	36	mixture	1.90	1	
7	0.3488	0.3634	0.3676	1	0	mixture	1.25	2	
8	0.2156	0.2291	0.2265	5	3	mixture	1.70	2	
9	0.2844	0.2917	0.2909	18	5	mixture	1.90	2	
10	0.3278	0.3322	0.3322	3	1	mixture	1.25	3	
11	0.1698	0.1754	0.1743	4	1	mixture	1.70	3	
12	0.2205	0.1997	0.2018	37	16	mixture	1.90	3	
13	0.2316	0.2006	0.2039	91	0	skew_norm	1.25		3
14	0.2697	0.1665	0.1649	177	0	skew_norm	1.70		3
15	0.1671	0.1024	0.1027	140	0	skew_norm	1.90		3
16	0.2477	0.2339	0.2319	104	0	skew_norm	1.25		10
17	0.1207	0.0956	0.0915	138	0	skew_norm	1.70		10
18	0.1988	0.1302	0.1218	232	0	skew_norm	1.90		10

Tab. 1: Umgang mit Warnmeldungen: Vergleich Simulation1 und Simulation2

	ML	Sim1	Sim3	NA_1	NA_3	Distr.	σ
1	0.2530	0.1614	0.1467	44	8	laplace	1.25
2	0.2825	-0.3314	-0.3709	59	16	laplace	1.70
3	0.2640	-0.2936	-0.3244	67	18	laplace	1.90

Tab. 2: Umgang mit Warnmeldungen: Vergleich Simulation1 und Simulation3 für die Laplace-Verteilung

A.6. Weitere Abbildungen Simulation

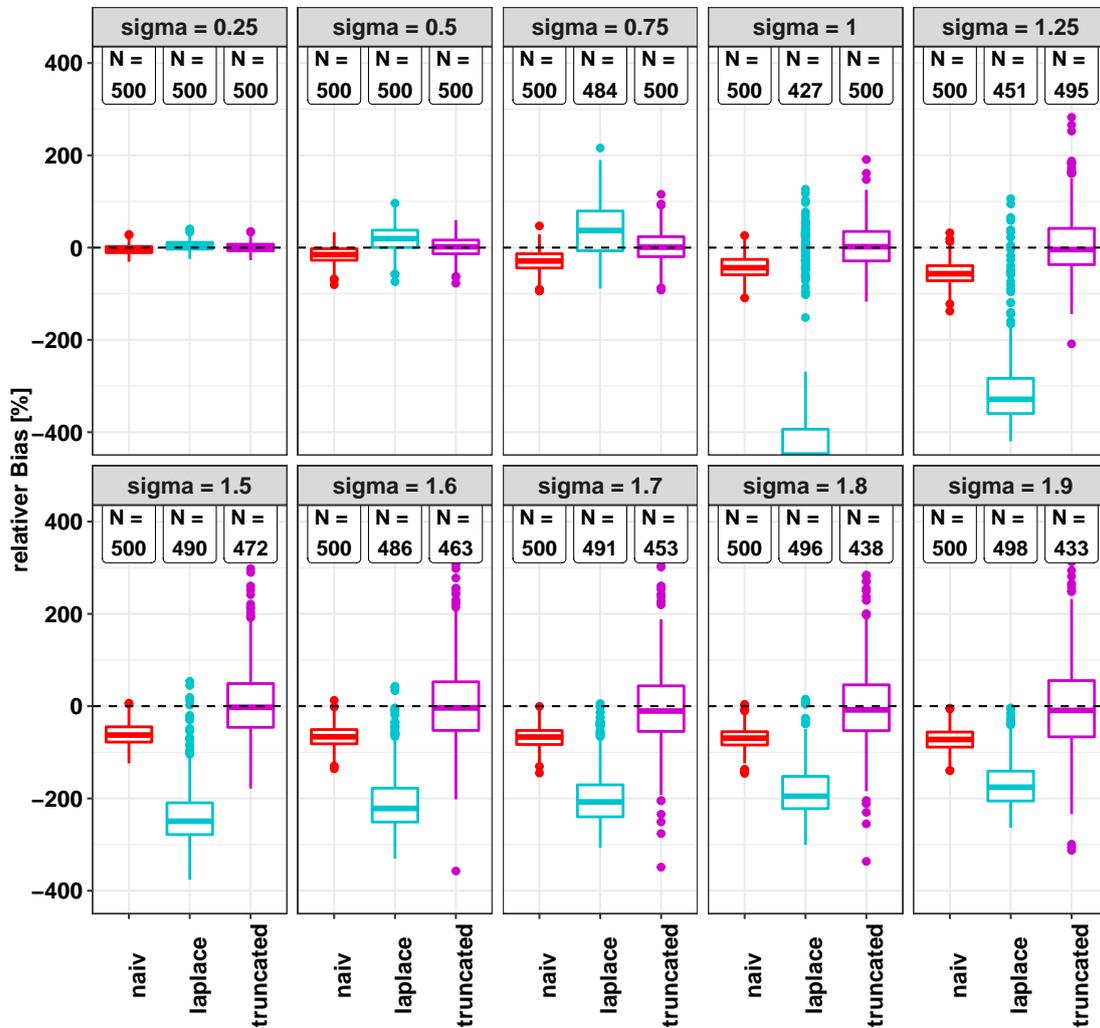


Abb. 23: Trunkierte Laplace-Verteilung: verschiedene Standardabweichungen von U_i (untrunkierte Verteilung) mit $n = 100$ und $b = 2 * \sigma$

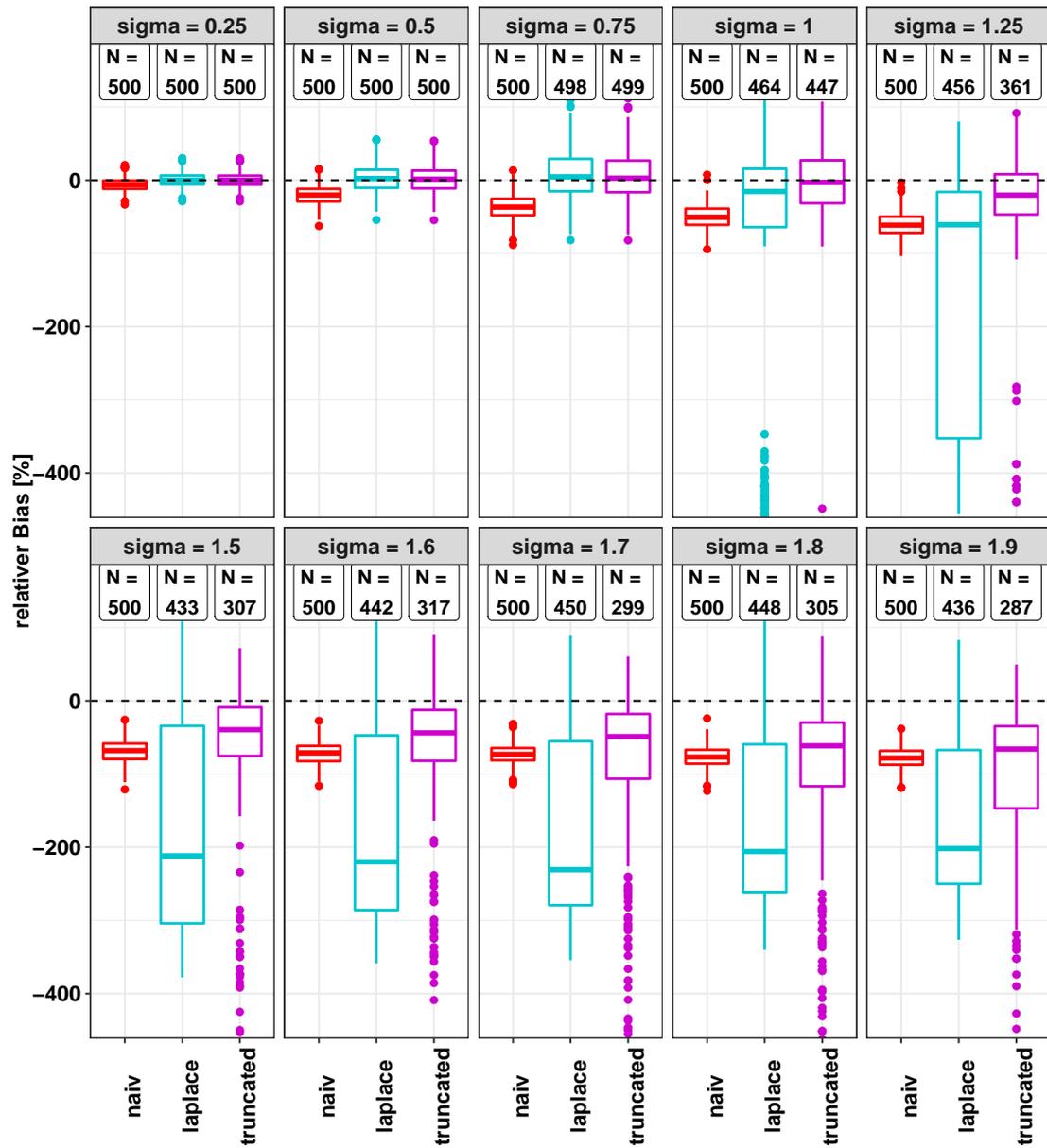


Abb. 24: Trunkierte Laplace-Verteilung: verschiedene Standardabweichungen von U_i (untrunkierte Verteilung) mit $n = 100$ und $b = 5 * \sigma$

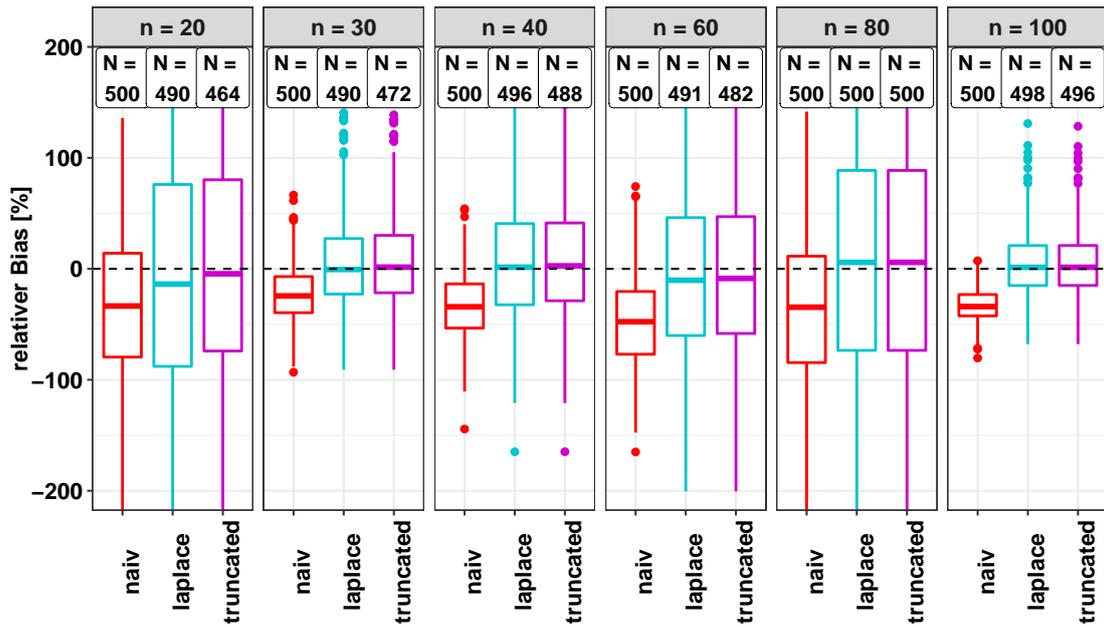


Abb. 25: Trunkierte Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$ (untrunkierte Verteilung) und $b = 10 * 0,75$

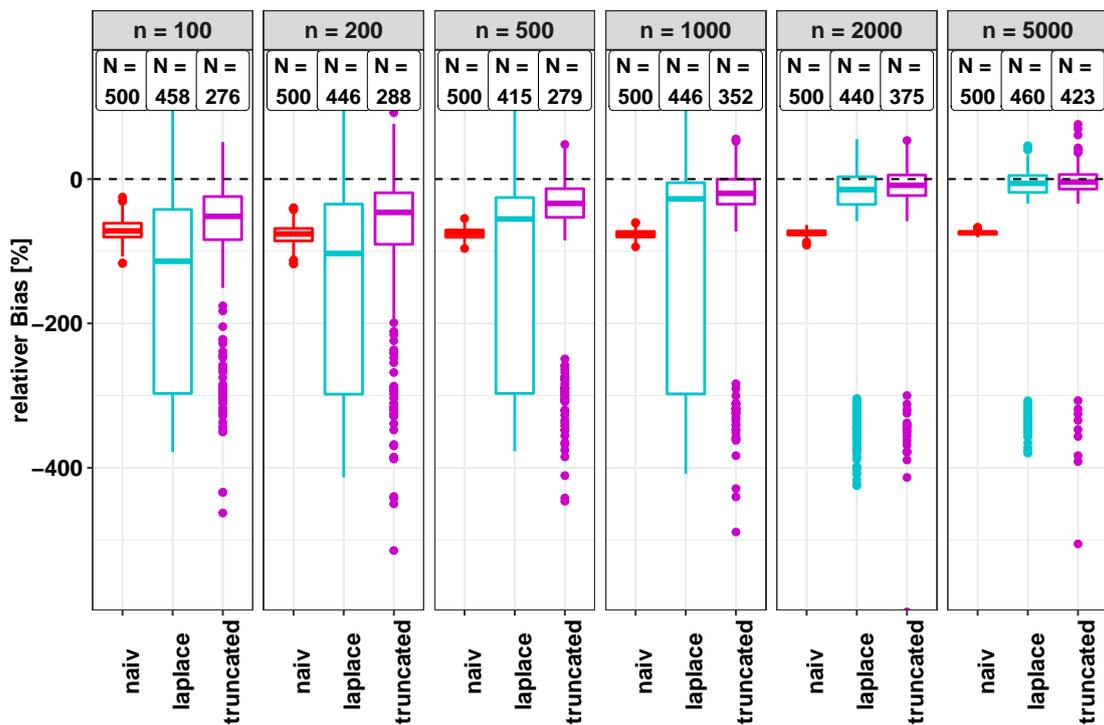


Abb. 26: Trunkierte Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$ (untrunkierte Verteilung) und $b = 10 * 1,7$

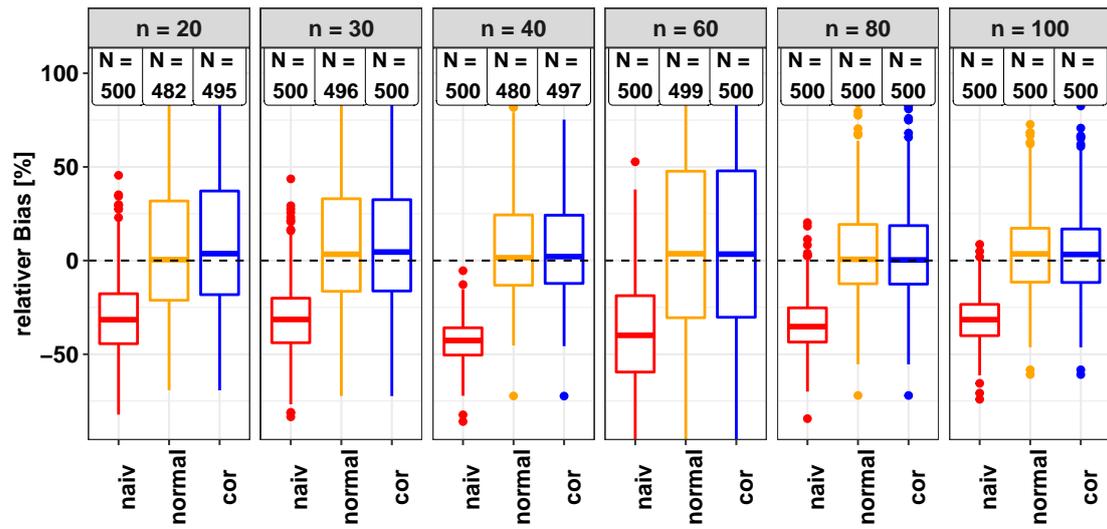


Abb. 27: Mischverteilung, Fall 2 (symm. bimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$

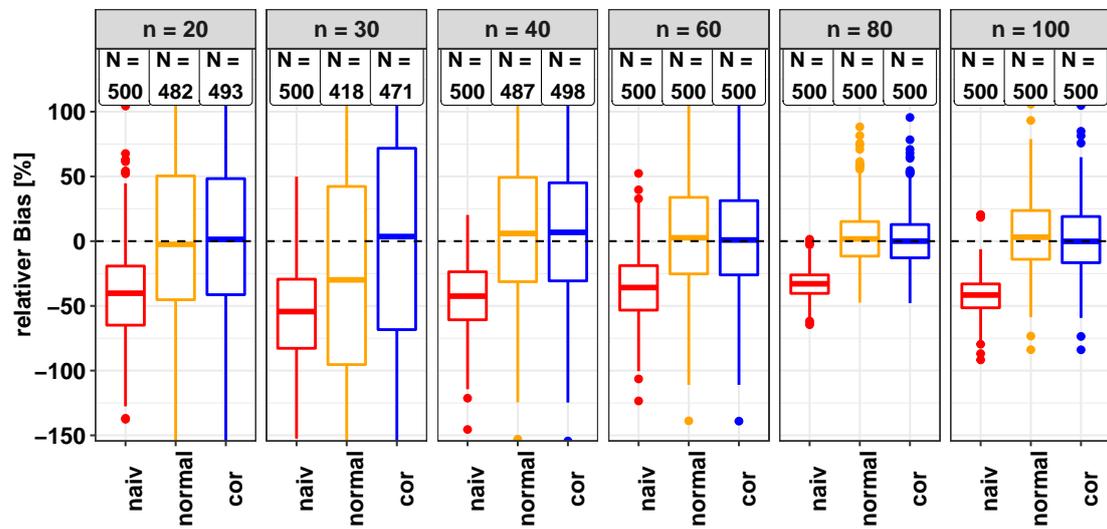


Abb. 28: Mischverteilung, Fall 3 (asymm.): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$

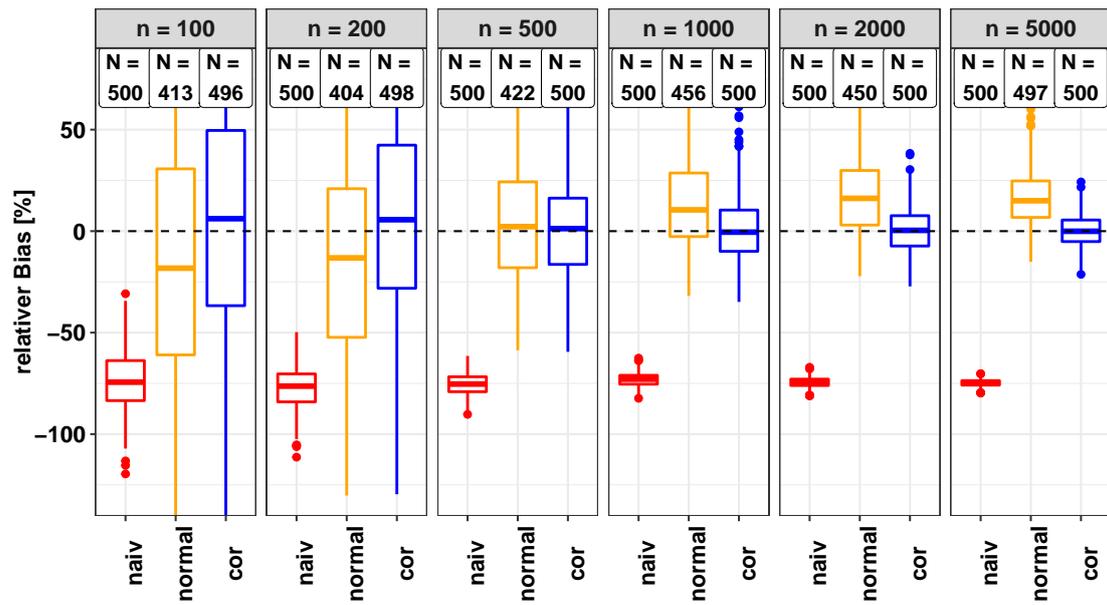


Abb. 29: Mischverteilung, Fall 2 (symm. unimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$

Abbildungsverzeichnis

1.	Schiefe Normalverteilung: Dichte $f(x; \alpha)$ für verschiedene Werte von α	19
2.	Dichte der schiefen Normalverteilung mit Varianz 1 und Erwartungswert 0 für verschiedene Werte von α	22
3.	Laplace-Verteilungen mit $\mu = 0$ und Varianz = 1 (d.h. $s \approx 0,71$) sowie Varianz = 2 ($s = 1$)	23
4.	Mischverteilungen aus zwei Normalverteilungen; Fall 1: symmetrisch, unimodal; Fall 2: symmetrisch, bimodal; Fall 3: asymmetrisch; jeweils mit Varianz 1	31
5.	Schiefe Normalverteilung: verschiedene Standardabweichungen von U_i mit $\alpha = 3, n = 100$. Relativer Bias von $\hat{\beta}_1$	37
6.	Schiefe Normalverteilung: verschiedene Standardabweichungen von U_i mit $\alpha = 10, n = 100$	38
7.	Schiefe Normalverteilung: verschiedene Werte von α mit $\text{Var}(U_i) = 0,75^2, n = 100$	39
8.	Schiefe Normalverteilung: verschiedene Werte von α mit $\text{Var}(U_i) = 1,7^2, n = 100$	40
9.	Schiefe Normalverteilung: verschiedene Stichprobenumfänge n mit $\alpha = 3$ und $\text{Var}(U_i) = 0,75^2$	41
10.	Schiefe Normalverteilung: verschiedene Stichprobenumfänge n mit $\alpha = 3$ und $\text{Var}(U_i) = 1,7^2$	41
11.	Laplace-Verteilung: verschiedene Standardabweichungen von U_i mit $n = 100$	42
12.	Laplace-Verteilung: verschiedene Standardabweichungen von U_i mit $n = 5000$	43
13.	Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$	44
14.	Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$	45
15.	Trunkierte Laplace-Verteilung: unterschiedlich starke Trunkierung mit $n = 100$ und $\text{Var}(U_i) = 0,75^2$ (untrunkierte Verteilung)	46
16.	Trunkierte Laplace-Verteilung: unterschiedlich starke Trunkierung mit $n = 100$ und $\text{Var}(U_i) = 1,7^2$ (untrunkierte Verteilung)	46

17.	Mischverteilung, Fall 1 (symm. unimodal): verschiedene Standardabweichungen von U_i mit $n = 100$	48
18.	Mischverteilung, Fall 2 (symm. bimodal): verschiedene Standardabweichungen von U_i mit $n = 100$	48
19.	Mischverteilung, Fall 3 (asymm.): verschiedene Standardabweichungen von U_i mit $n = 100$	49
20.	Mischverteilung, Fall 1 (symm. unimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$	50
21.	Mischverteilung, Fall 1 (symm. unimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$	51
22.	Mischverteilung, Fall 3 (asymm.): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$	51
23.	Trunkierte Laplace-Verteilung: verschiedene Standardabweichungen von U_i (untrunkierte Verteilung) mit $n = 100$ und $b = 2 * \sigma$	66
24.	Trunkierte Laplace-Verteilung: verschiedene Standardabweichungen von U_i (untrunkierte Verteilung) mit $n = 100$ und $b = 5 * \sigma$	67
25.	Trunkierte Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$ (untrunkierte Verteilung) und $b = 10 * 0,75$	68
26.	Trunkierte Laplace-Verteilung: verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$ (untrunkierte Verteilung) und $b = 10 * 1,7$	68
27.	Mischverteilung, Fall 2 (symm. bimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$	69
28.	Mischverteilung, Fall 3 (asymm.): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 0,75^2$	69
29.	Mischverteilung, Fall 2 (symm. unimodal): verschiedene Stichprobenumfänge n mit $\text{Var}(U_i) = 1,7^2$	70

Tabellenverzeichnis

1.	Umgang mit Warnmeldungen: Vergleich Simulation1 und Simulation2	65
2.	Umgang mit Warnmeldungen: Vergleich Simulation1 und Simulation3 für die Laplace-Verteilung	66

Literatur

- Azzalini, Adelchi (2013). *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press. DOI: 10.1017/CB09781139248891.
- Boos, Dennis D. und Leonard A. Stefanski (2013). *Essential Statistical Inference: Theory and Methods*. Springer Texts in Statistics. New York: Springer. DOI: 10.1007/978-1-4614-4818-1.
- Buzas, Jeffrey S. (2009). „A note on corrected scores for logistic regression“. In: *Statistics & Probability Letters* 79, S. 2351–2358. DOI: 10.1016/j.spl.2009.08.005.
- Buzas, Jeffrey S. und Leonard A. Stefanski (1996). „A note on corrected-score estimation“. In: *Statistics & Probability Letters* 28.1, S. 1–8. DOI: 10.1016/0167-7152(95)00074-7.
- Carroll, Raymond J., David Ruppert, Leonard A. Stefanski und Ciprian M. Crainiceanu (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. New York: Chapman & Hall/CRC. DOI: 10.1201/9781420010138.
- Cramer, Harald (1946). *Mathematical Methods of Statistics (PMS-9), Volume 9*. Princeton: Princeton University Press. DOI: <https://doi.org/10.1515/9781400883868>.
- Deitmar, Anton (2014). *Analysis*. Springer-Lehrbuch. Berlin: Springer. DOI: 10.1007/978-3-642-54810-9.
- Everitt, B. S. (1996). „An introduction to finite mixture distributions“. In: *Statistical Methods in Medical Research* 5.2, S. 107–127. DOI: 10.1177/096228029600500202.
- Fahrmeir, Ludwig, Thomas Kneib, Stefan Lang und Brian Marx (2013). *Regression: Models, Methods and Applications*. Berlin: Springer. DOI: 10.1007/978-3-642-34333-9.
- Gut, Allan (2009). *An Intermediate Course in Probability*. New York: Springer-Verlag. DOI: 10.1007/978-1-4419-0162-0.
- Huber, Peter J. (1964). „Robust Estimation of a Location Parameter“. In: *Annals of Mathematical Statistics* 35.1, S. 73–101. DOI: 10.1214/aoms/1177703732.

- Huber, Peter J. (1967). „The behavior of maximum likelihood estimates under nonstandard conditions“. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, S. 221–233. URL: <https://projecteuclid.org/euclid.bsmsp/1200512988>.
- Kotz, Samuel, Tomaz J. Kozubowski und Krzysztof Podgórski (2001). *The Laplace Distribution and Generalizations*. Birkhäuser Boston. DOI: 10.1007/978-1-4612-0173-1.
- Lindsay, Bruce G. (1995). „Mixture Models: Theory, Geometry and Applications“. In: *NSF-CBMS Regional Conference Series in Probability and Statistics* 5. DOI: 10.2307/4153184.
- Nakamura, Tsuyoshi (1990). „Corrected score function for errors-in-variables models: Methodology and application to generalized linear models“. In: *Biometrika* 77.1, S. 127–137. DOI: 10.1093/biomet/77.1.127.
- Novick, Steven J. und Leonard A. Stefanski (2002). „Corrected Score Estimation via Complex Variable Simulation Extrapolation“. In: *Journal of the American Statistical Association* 97.458, S. 472–481. URL: <http://www.jstor.org/stable/3085663>.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Soetaert, Karline (2009). *rootSolve: Nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations*. R package 1.6. URL: <https://cran.r-project.org/web/packages/rootSolve/index.html>.
- Stefanski, Leonard A. (1985). „The Effects of Measurement Error on Parameter Estimation“. In: *Biometrika* 72.3, S. 583–592. DOI: 10.2307/2336730.
- (1989). „Unbiased estimation of a nonlinear function a normal mean with application to measurement error models“. In: *Communications in Statistics - Theory and Methods* 18.12, S. 4335–4358. DOI: 10.1080/03610928908830159.
- Stefanski, Leonard A. und Raymond J. Carroll (1987). „Conditional Scores and Optimal Scores for Generalized Linear Measurement- Error Models“. In: *Biometrika* 74.4, S. 703–716. DOI: 10.2307/2336464.

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Gedanken, die aus fremden Quellen direkt oder indirekt übernommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und nicht veröffentlicht.

München, den 8. September 2020