

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN  
INSTITUT FÜR STATISTIK

---

# Bayesianische Ansätze zur Integration von externen Daten und Szenarien in die OpRisk-Messung mit Anwendungen

---

Master-Thesis  
Master Statistik mit wirtschafts- und sozialwissenschaftlicher  
Ausrichtung

**Autor:** Nataliia Semenenko  
**E-Mail:** [nataly\\_semenenko@gmx.de](mailto:nataly_semenenko@gmx.de)

**Betreuer:** Prof. Dr. Christian Heumann  
**Projektpartner:** BayernLB, Risk Office  
**Interne Betreuer:** Dr. Matthias Fischer,  
Dr. Anna-Lena Kißlinger-Schuderer

München, 17. Oktober 2018

## Zusammenfassung

Nach den Basel-II-Kriterien ist die Verwendung externer Daten und der Szenarioanalyse zur Messung operationeller Risiken zwingend erforderlich, da wenig interne Verlustbeobachtungen speziell im Randbereich vorliegen und die externen Daten sowie die Szenarien dazu benutzt werden können, um die Schätzung von Verlustverteilungen zu verbessern. Bei der Messung operationeller Risiken muss eine Bank bei der Integration externer Daten bzw. Szenarien jedoch Vorsicht walten lassen. Ohne passende Skalierung bzw. Gewichtung der spezifischen Art des Risikos der Bank kann es zu relevanten systematischen Risiken kommen (Vgl. [Torresetti and Nordio 2014](#)). Neben den unmittelbaren Skalierungs- bzw. Gewichtungsmethoden existieren einige Kombinationsmethoden, die es ermöglichen, externe Daten und Szenarien in geeigneter Weise in die OpRisk-Messung zu integrieren und die Vielfältigkeit der Szenarien sowie Unterschiede der internen und externen Daten zu berücksichtigen.

Zur Modellierung der Verlustverteilung aus operationellen Risiken greift die BayernLB neben intern gesammelten Schadensfalldaten auf externe Daten zurück, die von DakOR (Datenkonsortium für Operationelle Risiken) bereitgestellt werden. Die Hinzunahme externer Daten ist notwendig, um eine verlässliche Schätzung der Verlustverteilung durchzuführen. Neben der Modellierung der Verlustverteilung aus operationellen Risiken auf Basis von gesammelten historischen Daten führt die BayernLB Szenarioanalysen durch, um die nicht beobachteten, hypothetische Schadensfälle aus operationellen Risiken zu berücksichtigen.

Ziel dieser Arbeit ist die Darstellung verschiedener Bayes-Ansätze zum Kombinieren der internen und externen Daten sowie der Szenarien und die Illustration der Methoden am Beispiel der simulierten Daten wie auch die Anwendungsmöglichkeiten der Methoden auf die realen Schadensfalldaten zum operationellen Risiko. Insbesondere wird beim reinen Bayes-Ansatz auf Kombination von zwei Datenquellen fokussiert. Zwei Möglichkeiten der Umsetzung der Bayes-Ansätze werden in dieser Arbeit präsentiert: die Umsetzung auf Basis der analytischen Herleitungen sowie die numerische Äquivalenzlösung. Um das Problem von hochdimensionalen Integralen zu lösen, braucht man Verfahren zur praktischen Ermittlung von *a posteriori*. Die meisten Verfahren drehen sich hierbei um MCMC (*Markov-Chain-Monte-Carlo*). Die Hauptidee von MCMC ist, approximative *a posteriori* zu generieren, wenn diese nicht direkt simuliert werden kann. Diverse Methoden lösen approximativ das Problem von Integralen unter der Verwendung der Markov-Kette (Kap. 4.2, Def. 3), die gegen die wahre Verteilung konvergiert (Vgl. [Robert 2007](#): S. 301-303; [Gelman et al. 2013](#): S. 275-276). *Metropolis-Hastings* und *Gibbs-Sampler* sind die bekanntesten Methoden dafür (Kap. 4). Desweiteren wird aufgezeigt, wie die drei Datenquellen kombiniert werden können. Zum Schluss werden die Vor- und Nachteile der Methoden diskutiert.

## Danksagung

Für die Unterstützung bei meiner Masterarbeit möchte ich mich bei folgenden Personen herzlich bedanken:

Mein Dank gilt Herrn Prof. Dr. Christian Heumann für die Betreuung dieser Masterarbeit und die freundliche Hilfsbereitschaft, die er mir entgegenbrachte.

Ebenso danke ich Herrn Dr. Matthias Fischer und Frau Dr. Anna-Lena Kißlinger-Schuderer für das Bereitstellen dieses interessanten Themas der Masterarbeit sowie ihre Unterstützung. Ohne deren Hilfe und Bemühungen wäre diese Arbeit nicht zustande gekommen. Ich danke ihnen für Korrekturen, Verbesserungen und Kritik.

Herzlichen Dank auch dem Familienfreund Michael Hummel für seine Hilfsbereitschaft, Korrekturen und Verbesserungen. Dadurch ist diese Arbeit wesentlich lesenswerter geworden.

Ich möchte mich herzlich bedanken bei meinem Freund, der mich immer ermutigte und mit vielen nützlichen Tipps einen wesentlichen Teil zur Masterarbeit beigetragen hat.

Mein ganz besonderer Dank gilt abschließend meiner Familie, insbesondere meinen Eltern, die mir mein Studium ermöglicht und mich in all meinen Entscheidungen unterstützt haben.

## **Eidesstattliche Erklärung**

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit zum Thema:

**„Bayesianische Ansätze zur Integration von externen Daten und Szenarien  
in die OpRisk-Messung mit Anwendungen“**

ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

München, 17. Oktober 2018

Nataliia Semenenko

# Notation

## Abkürzungsverzeichnis

ACF - Autokorrelationsfunktion (*Autocorrelation Function*)

AMISE - asymptotischer mittlerer integrierter quadratischer Fehler (*Asymptotic Mean Integrated Squared Error*)

BayernLB - Bayerische Landesbank

BCBS - Basler Ausschuss für Bankenaufsicht (*Basel Committee on Banking Supervision*)

DakOR - Datenkonsortium für Operationelle Risiken

*iid* - unabhängig identisch verteilt (*independent identically distributed*)

KDE - Kerndichte-Schätzer (*Kernel Density Estimator*)

LDA - Verlustverteilung-Ansatz (*Loss Distribution Approach*)

MCMC - Markov-Ketten-Monte-Carlo (*Markov-Chain-Monte-Carlo*)

MISE - mittlerer integrierter quadratischer Fehler (*Mean Integrated Squared Error*)

MLE - Maximum Likelihood Schätzer (*Maximum Likelihood Estimator*)

MSEP - mittlerer quadratischer Fehler der Vorhersage (*Mean Squared Error of Prediction*)

OpRisk - Operationelle Risiken (*Operational Risk*)

SA - Simulierte Abkühlung (*Simulated Annealing*)

SCV - Geglättete Kreuzvalidierung (*Smoothed Cross Validation*)

VaR - Wert im Risiko (*Value at Risk*)

## Symbolverzeichnis

$X, Y$  - Zufallsvariablen

$\underline{X}, \underline{Y}$  - Vektor der Zufallsvariablen

$x, y$  - Beobachtungen

$\underline{x}, \underline{y}$  - Vektor der Beobachtungen

$f(\cdot)$  - Dichtefunktion

$\hat{f}(\cdot)$  - geschätzte Dichtefunktion

$F(\cdot)$  - Verteilungsfunktion

$\hat{F}(\cdot)$  - geschätzte Verteilungsfunktion

$E(\cdot) = \mu$  - Erwartungswert

$Var(\cdot) = \sigma^2$  - Varianz

$\sigma$  - Standardabweichung der Daten

$\tau$  - Standardabweichung der Parameter

$\theta$  - der wahre Parameter

$\underline{\theta}$  - Vektor der wahren Parameter

$\hat{\theta}$  - der geschätzte Parameter

$n$  - Anzahl der Daten

*int* - Bezeichnung von internen Daten, Schätzern, etc.

*ext* - Bezeichnung von externen Daten, Schätzern, etc.

*sz* - Bezeichnung von Daten, Schätzern, etc. aus Szenarien

*total* - Bezeichnung von gemeinsamen Daten, Schätzern, etc.

*Loss* - Verlusthöhe

$S$  - Höhe der Einzel-Schaden

$H$  - Anzahl der eingetretenen Schäden

$P(\cdot)$  - Wahrscheinlichkeit

$\underline{P}$  - Wahrscheinlichkeitsmatrix

$p(\underline{\theta})$  - *a priori*, Dichtefunktion der Parameter (Wahrscheinlichkeitsverteilung vor Beobachtung der Daten)

$p(\underline{\theta}|\underline{x})$  - *a posteriori*, Dichtefunktion der Parameter gegeben die Daten (Wahrscheinlichkeitsverteilung nach Beobachtung der Daten)

$f(\underline{x}|\underline{\theta})$  - *Likelihood*, Dichtefunktion der Daten gegeben die Parameter

$f(\underline{x}, \underline{\theta})$  - gemeinsame Dichtefunktion der Daten und Parameter

$f(\underline{x}) = \int f(\underline{x}|\underline{\theta})p(\underline{\theta})d\underline{\theta}$  - marginale Dichte der Daten (der stetige Fall)

$DP(\alpha, H)$  - Dirichlet-Prozess (mit  $H(x)$  - Basisverteilung  $\alpha$  - Konzentrationsparameter)

$\frac{1}{n} \sum_i \mathbb{1}_{x_i < x}$  - empirische Verteilungsfunktion

*seed* - Startwert für Simulation

# Inhaltsverzeichnis

<b>1. Einführung</b>	<b>1</b>
<b>2. Datengrundlage</b>	<b>5</b>
<b>3. Überblick von Bayes-Ansätzen</b>	<b>8</b>
3.1. Der reine Bayes-Ansatz . . . . .	9
3.2. Der empirische Bayes-Ansatz . . . . .	10
3.3. Der nicht parametrische Bayes-Ansatz . . . . .	11
<b>4. Methoden zur praktischen Umsetzung der Bayes-Ansätze</b>	<b>13</b>
4.1. Monte-Carlo-Methoden . . . . .	13
4.2. MCMC . . . . .	13
4.3. Metropolis-Hastings-Algorithmus (MH-Algorithmus) . . . . .	14
4.4. Gibbs-Sampler . . . . .	16
4.5. Simulated Annealing (SA) . . . . .	17
4.6. Klassische Kerndichte-Schätzung (KDE) . . . . .	18
4.7. Erweiterungen des MH-Algorithmus . . . . .	19
4.7.1. Wahl der Vorschlagsdichte und deren Parameter . . . . .	20
4.7.2. MH- und MH-SA-Algorithmus im Vergleich . . . . .	24
<b>5. Der reine Bayes-Ansatz in der OpRisk-Messung</b>	<b>29</b>
5.1. Kombination von zwei Datenquellen . . . . .	30
5.1.1. Schätzung der Schadenhäufigkeitsverteilung . . . . .	30
5.1.2. Schätzung der Schadenhöhenverteilung unter Verwendung der analytischen Herleitung . . . . .	36
5.1.3. Schätzung der Schadenhöhenverteilung unter Verwendung des MH-Algorithmus (Verallgemeinerung) . . . . .	43
5.1.4. Auswirkung der Schätzung auf VaR . . . . .	51
5.2. Kombination von drei Datenquellen . . . . .	55
<b>6. Der empirische Bayes-Ansatz in der OpRisk-Messung</b>	<b>56</b>
<b>7. Der nicht parametrische Bayes-Ansatz in der OpRisk-Messung</b>	<b>57</b>
<b>8. Zusammenfassung und Ausblick</b>	<b>58</b>
<b>A. Anhang</b>	<b>67</b>
A.1. Kernfunktionen . . . . .	67

A.2. Bestimmung der Parameter für Normal- und Lognormal-Verteilung in Abhängigkeit von Modus und Krümmung . . . . .	68
A.2.1. Bestimmung der Normal-Verteilung Parameter . . . . .	68
A.2.2. Bestimmung der Lognormal-Verteilung Parameter . . . . .	68
A.3. MH- und MH-SA-Algorithmus im Vergleich . . . . .	70
A.3.1. Markov-Kette . . . . .	70
A.3.2. Autokorrelation . . . . .	94
A.4. Konjugierte Verteilungen . . . . .	106
A.5. Verteilungen . . . . .	107
A.5.1. Lognormal-Verteilung . . . . .	107
A.5.2. Normal-Verteilung . . . . .	107
A.5.3. Inverse Gamma-Verteilung . . . . .	108
A.5.4. Skalierte inverse Chi-Quadrat-Verteilung . . . . .	108
A.5.5. Poisson-Verteilung . . . . .	109
A.5.6. Gamma-Verteilung . . . . .	109
A.5.7. Negative Binomial-Verteilung . . . . .	110
A.5.8. Beta-Verteilung . . . . .	110
A.6. Bestimmung der Jeffreys <i>a priori</i> für Poisson-Verteilung und der dazugehörigen <i>a posteriori</i> Dichte . . . . .	111
A.7. Bestimmung der bedingten Dichte der zukünftigen Beobachtung und des dazugehörigen Erwartungswertes . . . . .	112
A.8. Bestimmung der gemeinsamen <i>a priori</i> und der dazugehörigen <i>a posteriori</i> Dichte bei Lognormal-Verteilung mit zwei unbekanntem abhängigen Parametern . . . . .	114
A.9. Bestimmung der Jeffreys <i>a priori</i> für Lognormal-Verteilung und der dazugehörigen <i>a posteriori</i> Dichte mit zwei unbekanntem abhängigen Parametern	115
A.10. Kombination von zwei Datenquellen. Ergebnisse zur Kap. 5.1 . . . . .	119
A.10.1. Interne Daten: 100, externe Daten 1350, Simulationen: 1500 . . . . .	119
A.10.2. Interne Daten: 300, externe Daten 3000, Simulationen: 1500 . . . . .	129
A.10.3. Interne Daten: 100, externe Daten 300, Simulationen: 1500 . . . . .	139
A.10.4. Interne Daten: 200, externe Daten 300, Simulationen: 1500 . . . . .	149
A.10.5. Interne Daten: 300, externe Daten 300, Simulationen: 1500 . . . . .	159

## B. Datendokumentation

169

# 1. Einführung

Der Basler Ausschuss für Bankenaufsicht (BCBS) wurde 1974 von den Zentralbanken und Bankaufsichtsbehörden der G10-Staaten gegründet. Operationelle Risiken (OpRisk), wie z. B. Betrug, Diebstahl, Systemausfälle, etc., wurden lange Zeit in den Aufsichtsrichtlinien ohne eindeutige Definition den „sonstigen Risiken“ zugeordnet. Das erste Arbeitspapier zum Thema „*Operational Risk Management*“ wurde im Jahr 1998 publiziert und erste verbindliche Standards zu diesem Thema sowie die eindeutige Definition wurden erst im Jahr 2004 mit Basel II eingeführt (Inkrafttreten 2007). Der Basler Ausschuss für Bankenaufsicht definiert operationelles Risiko als „*Die Gefahr von unmittelbaren oder mittelbaren Verlusten, die infolge der Unangemessenheit oder des Versagens von internen Verfahren, Menschen und Systemen oder von externen Ereignissen eintreten*“ ([Bank für Internationalen Zahlungsausgleich 2003](#): S. 2). In der BayernLB wird ein mathematisches Modell zur Analyse und Schätzung der operationellen Risiken seit 2016 verwendet. OpRisk-Datengrundlage wird aus zwei Perspektiven betrachtet: Analyse historisch beobachteter Schadensfälle (ex post) und Abbildung der zukünftigen Schadensfälle unter der Berücksichtigung von Schadensszenarien (ex ante). Dabei werden in der ex post Perspektive nicht nur interne Schadensfälle betrachtet, sondern auch externe Schadensfälle aus DakOR<sup>1</sup> herangezogen. In der ex ante Perspektive werden die nicht beobachteten, hypothetischen Schadensfälle aus operationellen Risiken, deren Eintreten in der Zukunft möglich ist, durch Hinzunahme der Szenarien abgebildet.

Der gesamte Verlust  $Loss_{total}$  aus OpRisk für einen gegebenen Betrachtungszeitraum lässt sich als Summe der Verluste in einzelnen Segmenten  $Loss_m$  darstellen. Ein Segment wird als Kombination aus dem Geschäftsfeld und dem Verlustereignis definiert (siehe Tabelle 1.1). Der Verlust in einzelnen Segmenten setzt sich zusammen aus den eingetretenen Einzel-Schäden im Betrachtungszeitraum. Der gesamte Verlust aus OpRisk für einen Betrachtungszeitraum lässt sich somit als doppelte Summe der Höhen der Einzel-Schäden  $S_{m,h}$  darstellen<sup>2</sup>:

$$Loss_{total} = \sum_{m=1}^M Loss_m = \sum_{m=1}^M \sum_{h=1}^{H_m} S_{mh}, \quad (1.1)$$

mit  $S_{m,h}$  - Höhe der Einzel-Schäden im Segment  $m$ ,  $H_m$  - Anzahl der aufgetretenen Schäden im Segment  $m$  und  $M$  - Anzahl der Segmente.

---

<sup>1</sup>DakOR wurde 2006 von Finanzinstituten zum Austausch von OpRisk-Schadensfalldaten gegründet.

<sup>2</sup>Nach Verlustverteilungs-Ansatz (LDA - *Loss Distribution Approach*).

**Geschäftsfelder**

		Gesamtbank	Handel	Firmenkundengeschäft	Depot- & Treuhandgeschäfte	Privatkundengeschäft	Unternehmensfinanzierung/-beratung	Vermögensverwaltung	Wertpapierprovisionsgeschäft	Zahlungsverkehr & Abwicklung
<b>Verlustereignisse</b>	Interner Betrug	$Seg_1$	$Seg_2$	...	...	...	...	...	...	...
	Externer Betrug	...	...	...	...	...	...	...	...	...
	Beschäftigungspraxis & Arbeitsplatzsicherheit	...	...	...	...	...	...	...	...	...
	Kunden-, Produkte & Geschäftsgepflogenheiten	...	...	...	...	...	...	...	...	...
	Sachschäden	...	...	...	...	...	...	...	...	...
	Geschäftsunterbrechungen & Systemausfälle	...	...	...	...	...	...	...	...	...
	Ausführung-, Lieferung- & Prozessmanagement	...	...	...	...	...	...	...	...	$Seg_M$

Tabelle 1.1.: Beispiel Segmentbildung.  $Seg_1 =$  Segment 1, ...,  $Seg_M =$  Segment M.

Ziel der OpRisk-Messung ist die Prognose der unbekanntenen Verlustverteilung und der daraus berechneten Risikomaße, wie z. B. VaR (*Value at Risk*) - einer der standardmäßigen Risikomaße. Der VaR zum Signifikanzniveau  $\alpha$  ist als  $\alpha$ -Quantil der prognostizierten Verlustverteilung definiert:

$$\begin{aligned}
 Var_\alpha(Loss_{total}) &= \inf\{loss \in \mathbb{R}^+ | F_{Loss_{total}}(loss) \geq \alpha\} \\
 &= \inf\{loss \in \mathbb{R}^+ | P(Loss_{total,T+1} > loss) \leq 1 - \alpha\}.
 \end{aligned}
 \tag{1.2}$$

Der VaR gibt den Verlust für die nächste Periode  $T + 1$  an, der mit einer Wahrscheinlichkeit von  $\alpha$  nicht überschritten wird. Dabei wird  $\alpha$  typischerweise nahe 1 gewählt. In der BayernLB wird VaR zum Niveau 99.95% ( $\alpha = 0.9995$ ) bestimmt. Um eine möglichst gute Schätzgenauigkeit zu gewährleisten, wird die rechtsschiefe *Heavy-tailed*-Verteilung<sup>3</sup>

<sup>3</sup>Eine Wahrscheinlichkeitsverteilung mit relativ viel Masse im *tail*-Bereich. Die Dichte-Funktion fällt langsamer als exponentiell.

des gesamten Verlusts in 2 Bereiche aufgeteilt: *body* und *tail* (siehe Abb. 1.1). Dabei

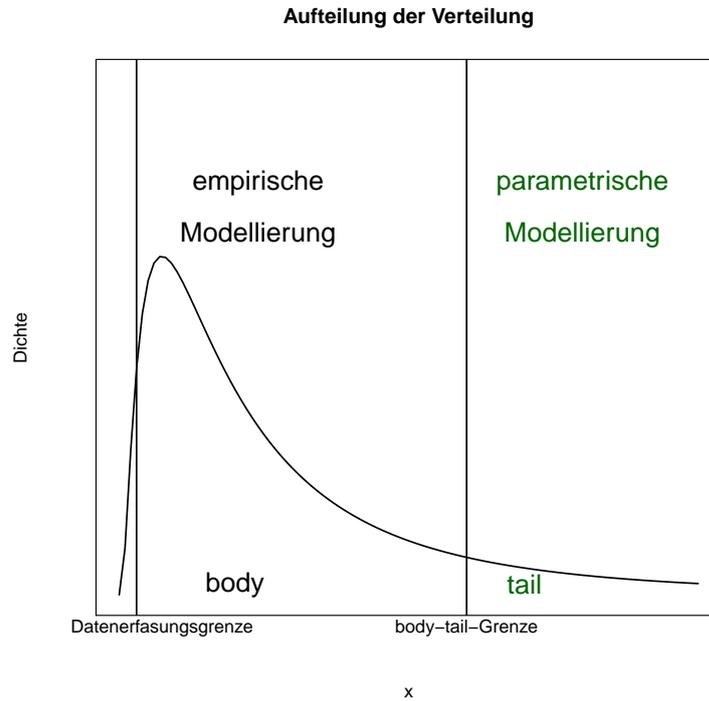


Abbildung 1.1.: Aufteilung der Verteilung von Gesamtschaden.

werden die Schäden nur ab einer bestimmten Höhe (*Datenerfassungsgrenze*) erfasst. Die Verteilung des gesamten Verlusts wird für den *body* empirisch und für den *tail* parametrisch bestimmt. Die Grenze zwischen *body* und *tail* wird dabei so bestimmt, dass möglichst viele Beobachtungen im *body*-Bereich und möglichst wenige, aber immer noch für die parametrische Schätzung ausreichend viele, Beobachtungen im *tail*-Bereich liegen. Es handelt sich im *tail*-Bereich um die Verluste mit großen Verlusthöhen und niedriger Häufigkeit. Die hohen Quantile sind hierbei besonders schwierig genau zu schätzen, da nur wenige interne Beobachtungen vorliegen, sodass die externe Daten und Szenarien zur Schätzung herangezogen werden sollen. Erste Analysen historisch beobachteter Schadensfälle zeigten, dass die externen Schadensfalldaten sich von der internen Schadensfalldaten unterscheiden. Um die Möglichkeit zu gewährleisten, externe Schadensfälle sowie die Szenarien in der Modellierung der Verlustverteilung aus OpRisk sinnvoll nutzen zu können, müssen diese in geeigneter Weise integriert werden. Dies ist der Ausgangspunkt dieser Arbeit.

Neben den expliziten Skalierungs- bzw. Gewichtungsmethoden, die es ermöglichen, externe Verluste an die internen so anzupassen, dass sich externe und interne Daten ähnlich verhalten und es möglich und sinnvoll wird, sowohl interne als auch externe Daten für Analysen und Vorhersagen zu nutzen, existieren Methoden ohne explizite Skalierung bzw. Gewichtung, die eine sinnvolle Integration der externen Daten sowie der Szenarien

in die OpRisk-Messung bieten und den Unterschied der internen und externen Verlustdaten sowie die Vielfältigkeit der Szenarien berücksichtigen. Die in den letzten Jahren in der Literatur meistbeschriebenen Methoden sind u. a. die Bayes-Ansätze. Diese stellen die grundlegenden Methoden dieser Arbeit dar.

Diese Arbeit gliedert sich wie folgt: Im Kapitel 2 wird die Datengrundlage für die Analysen präsentiert. Kapitel 3 beinhaltet einen Überblick von Bayes-Ansätzen, die in den letzten Jahren in der Literatur am häufigsten beschrieben und für die Analysen verwendet wurden. Dies sind der reine Bayes-Ansatz (Kap. 3.1), der empirische Bayes-Ansatz (Kap. 3.2) und der nicht parametrische Bayes-Ansatz (Kap. 3.3). Kapitel 4 stellt die Methoden zur praktischen Umsetzung der Bayes-Ansätze dar. Dies sind Monte-Carlo-Methoden (Kap. 4.1), die es ermöglichen, das Berechnungsproblem von hochdimensionalen Integralen, welches eines der größten Probleme bei Bayes-Ansätzen ist, auf Basis statistischer Stichproben approximativ zu lösen. Das Markov-Ketten-Monte-Carlo-Verfahren (Kap. 4.2) ist eines davon, welches unter anderem mithilfe der Metropolis-Hastings-Algorithmus (Kap. 4.3) und Gibbs-Sampler (Kap. 4.4) praktisch realisierbar ist. Kapitel 4.5 und 4.6 stellen die Hilfs-Methoden dar, die bei der praktischen Umsetzung der Bayes-Ansätze verwendet werden können. Kapitel 4.7 präsentiert die Erweiterungen des Metropolis-Hastings-Algorithmus und deren Vergleich. Danach werden drei ausgewählte Bayes-Ansätze detailliert mit Hinsicht auf die OpRisk-Messung beschrieben: der reine Bayes-Ansatz in der OpRisk-Messung (Kap. 5), der empirische Bayes-Ansatz in der OpRisk-Messung (Kap. 6) und der nicht parametrische Bayes-Ansatz in der OpRisk-Messung (Kap. 7). Außerdem werden in Kapitel 5.1 die Simulationsstudien beschrieben, sowie die Funktionalität der ausgewählten Methoden anhand dieser Simulationsstudien dargestellt, die zum Schluss im Kapitel 8 zusammengefasst und diskutiert werden.

## 2. Datengrundlage

Die Datengrundlage besteht aus internen (Bank) und externen (DakOR) Schadensfalldaten für ein Segment bis März 2017. Für diese Arbeit sind vor allem Schadensfälle aus dem *tail*-Bereich relevant, weil genau dort zu wenig interne Beobachtungen vorliegen, um eine verlässliche parametrische Schätzung der Verlustverteilung durchführen zu können. Aus diesem Grund beschränkt sich die Präsentation der Daten auf den *tail*-Bereich. In der Abb. 2.1 sind relative Häufigkeiten der Schadensfälle der zwei Mandanten abgebildet. Es liegen insgesamt 34 Beobachtungen (10.66%) bei internen (intSF) und 285 Beobachtungen (89.34%) bei externen (extSF) Schadensfalldaten vor.

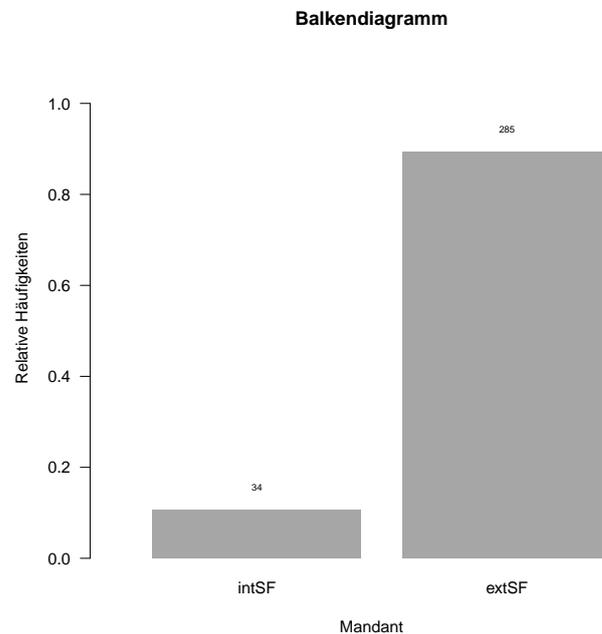


Abbildung 2.1.: Balkendiagramm *tail*-Bereich. Schadensfälle der zwei Mandanten.

In der Abb. 2.2 sind die Histogramme von internen (oben) und externen (unten) Verlusten dargestellt<sup>1</sup>. Aus den Grafiken ist der Unterschied der Verteilungen zu erkennen. Verlust-Intervalle sind durch externe Daten gleichmäßiger abgedeckt als durch interne (besonders das erste Intervall auf der x-Achse), was unter anderem durch die Anzahl der Schäden erklärbar ist. Außerdem sind die zwei größten Beobachtungen bei den internen Schäden auffällig, da diese erheblich größer sind als bei den externen Schäden (das letzte Intervall auf der x-Achse).

<sup>1</sup>Die Skala der x-Achse ist aus den Datenschutzgründen ausgelassen.

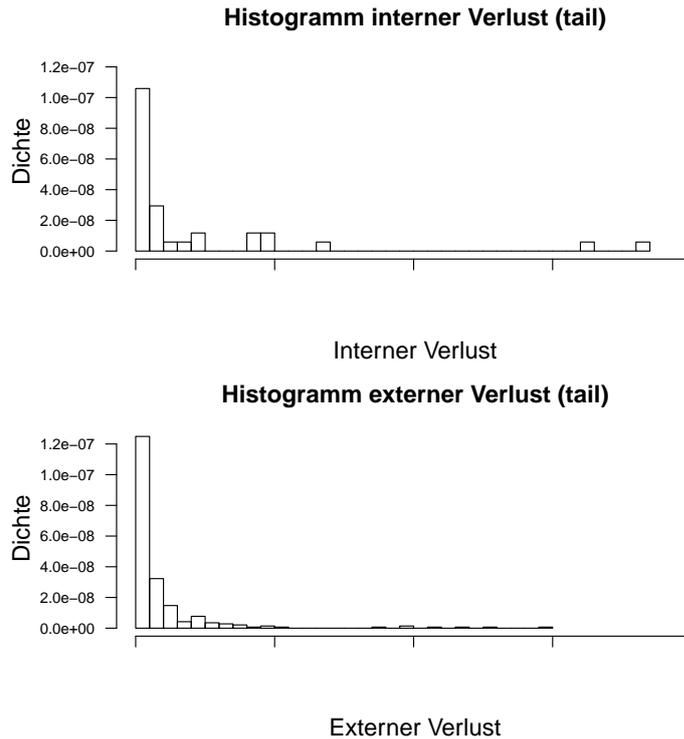


Abbildung 2.2.: Histogramme *tail*-Bereich. Verlust intern (oben) vs. Verlust extern (unten).

Die Unterschiede der beiden Verteilungen sind im QQ-Plot (Abb. 2.3) noch besser zu erkennen, wobei die empirischen Quantile der internen gegen die empirischen Quantile der externen Schadensfalldaten geplottet werden. Die Punkte liegen auf der Diagonale, wenn beide Verteilungen gleich sind, was bei internen und externen Schadensfalldaten im *tail*-Bereich nicht der Fall ist. Es ist deutlich zu erkennen, dass die Punkte von der Diagonale abweichen. Die höheren Quantile der internen Verluste sind deutlich größer als die höheren Quantile der externen Verluste.

Eine weitere Möglichkeit den Unterschied (bzw. Ähnlichkeit) der *heavy-tailed*-Verteilungen darzustellen ist der Log-log-Plot (Abb. 2.4). Dabei werden die logarithmierten Beobachtungen  $\log(x)$  (x-Achse) gegen die logarithmierten Werte  $1 - F(x)$  (y-Achse) geplottet. Wenn beide Verteilungen gleich sind, werden die Punkte auf einer Linie liegen. Aus der Abbildung ist zu erkennen, dass sich die Verteilung der internen Daten im *tail*-Bereich (schwarze Punkte) von der Verteilung der externen Daten im *tail*-Bereich (rote Punkte) unterscheidet. Die Unterschiede liegen besonders für höhere Werte (x-Achse: [16; 20], y-Achse: [-6; -1]) vor. Die Ungleichmäßigkeit und Sprünge bei höheren Werten weisen auf eine nicht genügende Anzahl der Beobachtungen hin. Größere Abstände zwischen den Punkten einer Verteilung weisen auf die weiter auseinander liegende Werte hin. Diese Art von Grafiken wird in dieser Arbeit für die Darstellung der Ergebnisse verwendet.

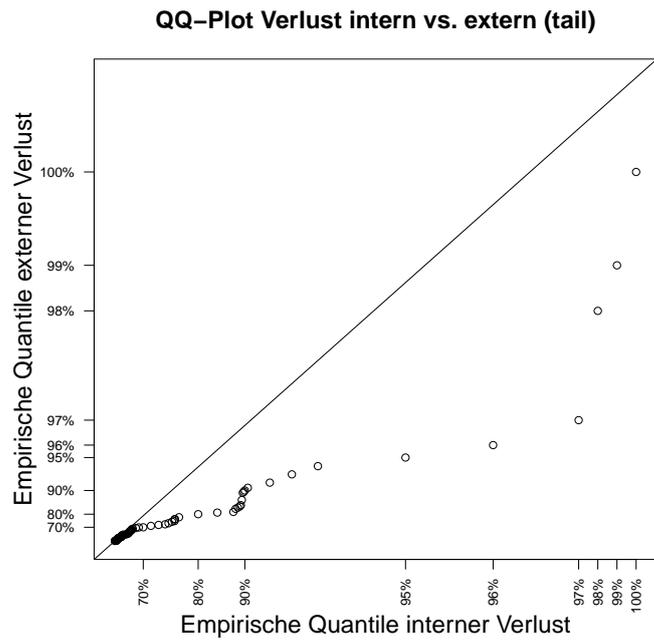


Abbildung 2.3.: QQ-Plot *tail*-Bereich. Verlust intern vs. extern.

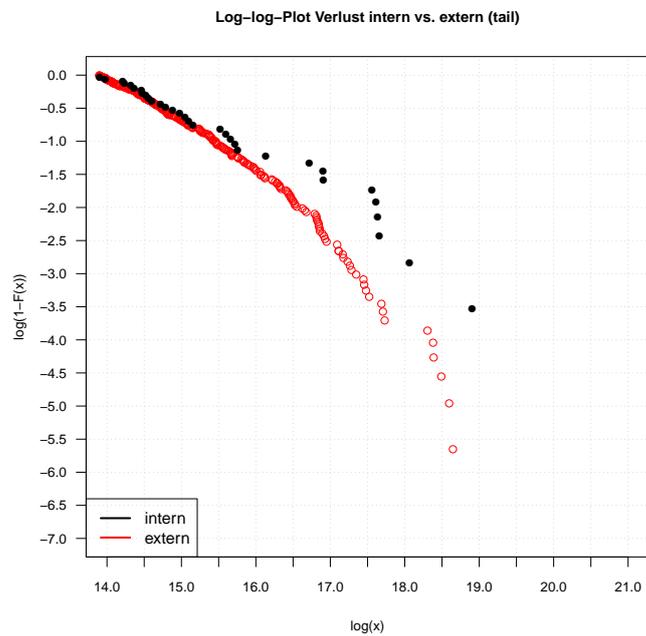


Abbildung 2.4.: Log-log-Plot *tail*-Bereich. Verlust intern vs. extern.

### 3. Überblick von Bayes-Ansätzen

Die Bayes-Ansätze basieren auf dem Satz von Bayes:

**Satz 1 (Satz von Bayes).** Sei  $A_1, \dots, A_k$  eine disjunkte Zerlegung von  $\Omega$ , wobei für mindestens ein  $i$ ,  $i = 1, \dots, k$ ,  $P(A_i) > 0$  und  $P(B|A_i) > 0$  erfüllt ist. Dann gilt:

$$P(A_j|B) = \frac{P(B, A_j)}{\sum_{i=1}^k P(B|A_i) \cdot P(A_i)} = \frac{P(B, A_j)}{P(B)} = \frac{P(B|A_j) \cdot P(A_j)}{P(B)}, \quad j = 1, \dots, k \quad (3.1)$$

(Vgl. [Fahrmeir et al. 2016](#): S. 199).

Diese Bayes-Regel für Wahrscheinlichkeitsrechnung lässt sich auf die Analyse von parametrischen Modellen übertragen. Integriert man die Parameter aus der gemeinsamen Dichte der Daten und Parameter heraus, so erhält man die marginale Dichte der Daten. Teilt man die gemeinsame Dichte der Daten und Parameter durch diese marginale Dichte, bekommt man die *a posteriori* Dichte:

$$p(\underline{\theta}|\underline{x}) = \frac{f(\underline{x}|\underline{\theta})p(\underline{\theta})}{f(\underline{x})} = \frac{f(\underline{x}|\underline{\theta})p(\underline{\theta})}{\int f(\underline{x}|\underline{\theta})p(\underline{\theta})d\underline{\theta}}, \quad (3.2)$$

mit  $p(\underline{\theta}|\underline{x})$  - *a posteriori* (Parameterverteilung nach Beobachtung der Daten),  $p(\underline{\theta})$  - *a priori* (Parameterverteilung vor Beobachtung der Daten),  $f(\underline{x}|\underline{\theta})$  - *Likelihood* (Dichtefunktion der Daten gegeben die Parameter),  $f(\underline{x}) = \int f(\underline{x}|\underline{\theta})p(\underline{\theta})d\underline{\theta}$  - marginale Dichte der Daten (Vgl. [Cruz et al. 2015](#): S. 561-562). Die marginale Dichte der Daten spielt hierbei die Rolle der Normierungskonstante, die nicht von Parametern abhängt. D. h. die *a posteriori* kann als Proportionalität zu dem Produkt aus der *Likelihood* und *a priori* erfasst werden:

$$p(\underline{\theta}|\underline{x}) \propto f(\underline{x}|\underline{\theta})p(\underline{\theta}). \quad (3.3)$$

Der Unterschied der Bayes-Ansätze zu den anderen Schätzmethoden liegt darin, dass die Parameter  $\Theta$  als Zufallsvariablen betrachtet werden.

Oft ist nicht die *a posteriori* von Interesse, sondern die Vorhersageverteilung der zukünftigen Beobachtung  $X_{n+1}$  bedingt auf alle beobachtete Daten  $\underline{X} = (X_1, \dots, X_n)$ . Unter den Annahmen, dass die zukünftige Beobachtung  $X_{n+1}$  und alle beobachtete Daten  $\underline{X}$  gegeben Parameter  $\underline{\Theta} = (\Theta_1, \dots, \Theta_k)$  unabhängig - bedingt unabhängig - und  $X_1, \dots, X_{n+1}$  gegeben Parameter  $\underline{\Theta}$  *iid* sind, ist die bedingte Dichte der zukünftigen Beobachtung folgendermaßen definiert:

$$f(x_{n+1}|\underline{x}) = \int f(x_{n+1}|\underline{\theta})p(\underline{\theta}|\underline{x})d\underline{\theta}. \quad (3.4)$$

Falls  $X_{n+1}$  und  $\underline{X}$  nicht unabhängig sind, ist die bedingte Dichte der zukünftigen Beobachtung folgendermaßen definiert:

$$f(x_{n+1}|\underline{x}) = \int f(x_{n+1}|\underline{\theta}, \underline{x})p(\underline{\theta}|\underline{x})d\underline{\theta} \quad (3.5)$$

(Vgl. [Cruz et al. 2015](#): S. 589).

Um die Vorhersageverteilung zu schätzen, müssen die folgenden Schritte ausgeführt werden:

1. Schätzung der *a priori*
2. Schätzung der *a posteriori* (Formel [3.2](#))
3. Berechnung der Vorhersageverteilung gegeben die Daten (Formel [3.4](#) oder [3.5](#))

Sowohl in Schritt 2 als auch in Schritt 3 müssen die hochdimensionalen Integrale berechnet werden. Die Monte-Carlo-Methoden (Kap. [4.2](#)) können hierbei benutzt werden, um das Problem numerisch zu lösen und praktisch umzusetzen.

Der Unterschied der Bayes-Ansätze liegt in der Schätzung der *a priori*. Die meist verwendete Möglichkeiten, die *a priori* zu schätzen, wurden unter anderen von [Cruz et al. 2015](#): Kap. 15 zusammengefasst. Das sind unter anderen: der reine Bayes-Ansatz, der empirische Bayes-Ansatz und der nicht parametrische Bayes-Ansatz.

### 3.1. Der reine Bayes-Ansatz

Der reine Bayes-Ansatz basiert auf unterschiedlichen Expertenmeinungen und Annahmen. Die Experten äußern die Meinungen z. B. über die Verteilungsfamilie, Verteilungsparameter, etc. Die Expertenmeinungen können auf theoretischen Wissen oder auch auf den Ergebnissen früherer Studien oder Forschungen aufgebaut werden. In diesem Fall können die Expertenmeinungen für die Schätzung der Parameterverteilung vor der Beobachtung der Daten (*a priori*) genutzt werden. Diese werden mit der *Likelihood* der beobachteten Daten kombiniert und daraus die Parameterverteilung nach der Beobachtung der Daten (*a posteriori*) geschätzt und anschließend die Vorhersageverteilung bestimmt.

Oft wird als *a priori* die konjugierte Verteilung angenommen:

**Definition 1 (Konjugierte Verteilung).** Eine Familie  $\mathcal{F}$  von Verteilungen auf dem Parameterraum  $\Theta$  heißt konjugiert zu einer Dichte  $f(\underline{x}|\theta)$ , wenn für jede *a priori*  $p(\theta)$  auf  $\mathcal{F}$  die *a posteriori*  $p(\theta|\underline{x})$  ebenfalls zu  $\mathcal{F}$  gehört (Vgl. [Schmid and Happ 2017](#): S. 14-15).

D. h. die *a priori* ist zu einer Dichte konjugiert, wenn *a priori* und *a posteriori* aus gleicher Verteilungstyp stammen<sup>1</sup>.

[Berger 1980](#): S. 63-68 beschreibt einige Möglichkeiten, die *a priori* unter Verwendung von Expertenmeinungen zu schätzen:

- **Histogramm-Methode:** Der Parameterraum wird in Intervalle geteilt und die subjektive Wahrscheinlichkeit für jedes Intervall spezifiziert. Das Histogramm dieser Wahrscheinlichkeiten kann geplottet werden. Daraus kann die *a priori* z. B. als Kerndichte-Schätzer bestimmt werden.

---

<sup>1</sup>Einige konjugierte Verteilungen sind in Anh. [A.4](#) dargestellt.

- **Relative Likelihood-Methode:** *Likelihoods* für verschiedene Parameterwerte werden verglichen. Daraus kann die *a priori* bestimmt werden.
- **Bestimmung der Verteilungsfunktion:** subjektive Konstruktion der Verteilungsfunktion für *a priori* und Skizzieren der Dichte.
- **Anpassung der gegebenen Funktionsform:** *a priori* Verteilungsparameter werden so bestimmt, dass eine Funktionsform für die *a priori* Verteilung so nahe wie möglich an das *a priori* Wissen über Momente, Quantile, etc. angepasst wird.

Die Anpassung der gegebenen Funktionsform wird bevorzugt, weil die Expertenmeinungen öfters über einige Größen wie Quantile oder andere Risikomerkmale anstatt direkt über die Parameter angegeben werden. [Cruz et al. 2015](#) und [Shevchenko and Peters 2013](#) beschreiben die Möglichkeit der Anpassung der gegebenen Funktionsform genauer. Experten definieren die Risikoeigenschaften, die von den Parametern abhängen  $d_i = g_i(\underline{\theta}), i = 1, 2, \dots, n$  (z.B. Quantile, Erwartungswerte, erwartete Dauer zwischen Verlusten, die hohe Schwellenwerte überschreiten, etc.<sup>2</sup>). Wenn die gemeinsame *a priori* von den definierten Risikoeigenschaften abhängt  $p(d_1, \dots, d_n)$ , kann die *a priori* für die Parameter folgendermaßen definiert werden:

$$p(\underline{\theta}) = p(g_1(\underline{\theta}), \dots, g_n(\underline{\theta})) \cdot \left| \frac{\partial(g_1(\underline{\theta}), \dots, g_n(\underline{\theta}))}{\partial(\theta_1, \dots, \theta_k)} \right|, \quad (3.6)$$

mit  $|\partial(g_1(\underline{\theta}), \dots, g_n(\underline{\theta}))/\partial(\theta_1, \dots, \theta_k)|$  - Determinante der Jacobi-Matrix. Die Risikoeigenschaften müssen hierbei so gewählt werden, dass die Unabhängigkeit der Parameter angenommen werden kann.

## 3.2. Der empirische Bayes-Ansatz

Der empirische Bayes-Ansatz basiert auf den empirischen Analysen der historisch beobachteten Daten aus den zur Verfügung stehenden externen Quellen vor der Beobachtung der aktuellen Daten. In diesem Fall wird die *a priori* aus den Daten ermittelt, also die Parameterverteilung aus den vorher gesammelten Daten geschätzt. Die Grundidee hierbei ist, die Parameter aus den gelieferten identischen Daten der unabhängigen Quellen oder Datengruppen abzuleiten und daraus die gemeinsame Verteilung der Parameter zu beobachten, die folglich als *a priori* für die weiteren Berechnungen dienen soll.

[Shevchenko and Peters 2013](#): S. 14-15 beschreiben dazu die theoretische Grundlage. Zu betrachten sind die Risikodaten  $\underline{X}^{(j)} = (X_1^{(j)}, \dots, X_{n_j}^{(j)})$ ,  $j = 1, \dots, J$  mit  $n_j$  - Anzahl der Beobachtungen im Risikoprofil<sup>3</sup>  $j$ .  $X_1^{(j)}, \dots, X_{n_j}^{(j)}$  seien für gegebene  $\underline{\Theta}^{(j)} = \underline{\theta}^{(j)}$  bedingt unabhängig und *iid* mit der Dichte  $f(\cdot|\underline{\theta}^{(j)})$ . Für jede Risikoart<sup>4</sup> bezeichnet  $\underline{\Theta}^{(j)}, j = 1, \dots, J$

<sup>2</sup>Beliebige Risikoeigenschaften, die ausreichend sind, um die Parameter der *a priori* zu bestimmen.

Die Verteilungstyp der *a priori* muss hierbei ebenfalls von Experten festgelegt werden.

<sup>3</sup>Ein Profil kann z. B. für eine Bank bzw. Bankgruppe definiert werden ([Shevchenko and Peters 2013](#): S. 14-15).

<sup>4</sup>Im Rahmen der OpRisk kann eine Risikoart als ein Segment (siehe Kap. 1) definiert werden.

den Parametervektor des  $j$ -ten Risikoprofils. Die Risiken sind aus der bedingten Verteilung  $F(x|\underline{\theta}^{(j)})$  generiert. Die Risiken sind unterschiedlich und haben unterschiedliche Profile. Die Parameter für alle ähnliche Risikoprofile  $\underline{\Theta}^{(1)}, \dots, \underline{\Theta}^{(J)}$  stammen allerdings aus der gleichen Verteilung  $p(\underline{\theta})$ , also *iid* (siehe Abb. 3.1).

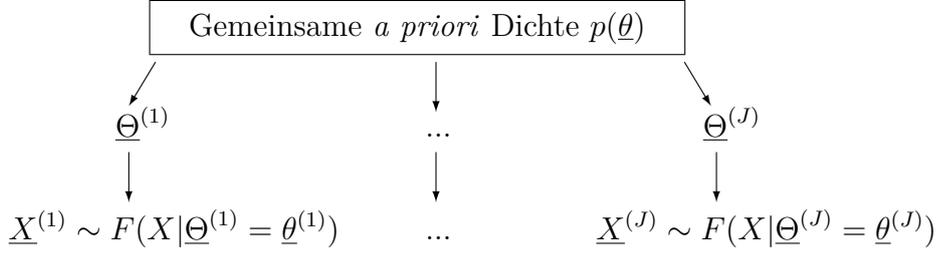


Abbildung 3.1.: Der empirische Bayes-Ansatz. Interpretation der *a priori* Dichte  $p(\underline{\theta})$ .  $\underline{\Theta}^{(j)}$  - Parameter des  $j$ -ten Risikoprofils. Gegeben  $\underline{\Theta}^{(j)} = \underline{\theta}^{(j)}$  sind Risikodaten  $\underline{X}^{(j)}$  aus der bedingten Verteilung  $F(x|\underline{\theta}^{(j)})$  generiert. Die Risiken sind unterschiedlich und haben unterschiedliche Profile  $\underline{\theta}^{(j)}$ , die Parameter  $\underline{\Theta}^{(1)}, \dots, \underline{\Theta}^{(J)}$  haben allerdings die gleiche gemeinsame *a priori* Dichte  $p(\underline{\theta})$ <sup>5</sup>.

Die marginale Dichte aller Beobachtungen ist folgendermaßen definiert:

$$f(\underline{x}^{(1)}, \dots, \underline{x}^{(J)}) = \prod_{j=1}^J \int \left[ \prod_{i=1}^{n_j} f(x_i^{(j)}|\underline{\theta}^{(j)}) \right] p(\underline{\theta}^{(j)}) d\underline{\theta}^{(j)}. \quad (3.7)$$

Daraus können die Parameter der *a priori* Dichte  $p(\underline{\theta})$  ermittelt werden, beispielsweise unter Verwendung des Maximum Likelihood Schätzers (MLE - *Maximum Likelihood Estimator*), indem die marginale Dichte (Formel 3.7) maximiert wird. Die *a posteriori* Formel 3.3 kann für ein Risikoprofil für die jeweilige Risikoart folgendermaßen umgeschrieben werden:

$$p(\underline{\theta}^{(j)}|\underline{x}^{(j)}) \propto \prod_{i=1}^{n_j} f(x_i|\underline{\theta}^{(j)})p(\underline{\theta}^{(j)}) = \prod_{i=1}^{n_j} f(x_i|\underline{\theta}^{(j)})p(\underline{\theta}), \quad (3.8)$$

wobei  $p(\underline{\theta}^{(j)}) = p(\underline{\theta})$ . Diese *a posteriori* kann weiterhin bei der Schätzung der Vorhersageverteilung benutzt werden.

### 3.3. Der nicht parametrische Bayes-Ansatz

Im Unterschied zu den parametrischen Bayes-Ansätzen wird bei dem nicht parametrischen Bayes-Ansatz keine Annahme getroffen, dass die zugrundeliegende das Verlustprozess generierende Verteilung parametrisch ist. Die *a priori* wird direkt in die Verteilung

<sup>5</sup>Eigene Darstellung in Anlehnung an Cruz et al. 2015: S. 592

versteckt und die *a posteriori* (bedingte Verteilung der Parameter gegeben die Daten) wird als Kombination der *a priori* und der empirischen Verteilung der Daten bestimmt. Der nicht parametrische Bayes-Ansatz basiert auf dem Dirichlet-Prozess (der Wahrscheinlichkeitsverteilung von den Wahrscheinlichkeitsverteilungen) (Shevchenko and Peters 2013: S. 30). Der Dirichlet-Prozess  $DP(\alpha, H(x))$  kann mithilfe einer Basisverteilung  $H(x)$  und einem Konzentrationsparameter  $\alpha > 0$  spezifiziert werden. Eine unbekannte Verteilung  $F(x)$ , die zufällig in jedem Punkt  $x$  mithilfe des Dirichlet-Prozesses  $DP(\alpha, H(x))$  modelliert wurde, hat Mittelwert  $H(x)$  und Varianz  $H(x)(1 - H(x))/(\alpha + 1)$ . Wenn der Konzentrationsparameter  $\alpha$  sinkt, nähert sich die wahre Verteilung der Basisverteilung  $H(x)$  an. Die Ziehung aus dem Dirichlet-Prozess ist eine Verteilungsfunktion und für geordnete  $x$ -Werte  $x_1 < x_2 < \dots < x_n$  ist die Verteilungsfunktion von  $F(x_1), F(x_2) - F(x_1), \dots, 1 - F(x_n)$  definiert als  $n + 1$  multivariate Dirichlet-Verteilung  $Dir(\alpha H(x_1), \alpha H(x_2), \dots, \alpha H(x_k))$  (Def. 2).

**Definition 2 (Dirichlet-Verteilung).** Die  $d$ -variante Dirichlet-Verteilung wird bezeichnet als  $Dir(\alpha_1, \alpha_2, \dots, \alpha_d)$  mit  $\alpha > 0$ . Der Zufallsvektor  $(Q_1, Q_2, \dots, Q_d)$  folgt einer Dirichlet-Verteilung, wenn seine Dichte folgendermaßen ist:

$$f(q_1, q_2, \dots, q_{d-1}) = \frac{\Gamma(\alpha_1, \alpha_2, \dots, \alpha_d)}{\prod_{i=1}^d \Gamma(\alpha_i)} \prod_{i=1}^d q_i^{\alpha_i - 1}, \quad (3.9)$$

wobei  $q_i > 0$ ,  $\sum_{i=1}^d q_i = 1$ ,  $\Gamma(\cdot)$  - Gammafunktion ist (Vgl. Shevchenko and Peters 2013: S. 31).

Die  $i$ -te marginale Verteilungsfunktion von  $Dir(\alpha_i)$  ist die Beta-Verteilung  $Be(\alpha_i, \sum_{j=1}^d \alpha_j - \alpha_i)$ . Die marginale Verteilung des Dirichlet-Prozesses  $DP(\alpha, H(x))$  ist somit die Beta-Verteilung:

$$F(x) \sim Be(\alpha H(x), \alpha(1 - H(x))). \quad (3.10)$$

Wenn die *a priori* für die unbekannte Verteilung  $F(x)$  der Dirichlet-Prozess  $DP(\alpha, H(x))$  ist, dann ist die *a posteriori* Verteilung für  $F(x)$ :

$$DP\left(\alpha + n, \frac{\alpha}{\alpha + n} H(x) + \frac{n}{\alpha + n} \frac{1}{n} \sum_i \mathbb{1}_{x_i < x}\right), \quad (3.11)$$

mit  $\frac{1}{n} \sum_i \mathbb{1}_{x_i < x}$  - empirische Verteilungsfunktion. Anders gesagt, der Dirichlet-Prozess ist die konjugierte<sup>6</sup> *a priori* für die empirische Stichprobenverteilung. Die *a posteriori* Verteilung konvergiert gegen die empirischen Verteilung, die ihrerseits gegen die wahre Verteilung  $F(x)$  konvergiert. Die Basisverteilung  $H(x)$  kann hierbei als die Expertenmeinung für  $F(x)$  festgelegt werden.

---

<sup>6</sup>Siehe Def. 1.

## 4. Methoden zur praktischen Umsetzung der Bayes-Ansätze

### 4.1. Monte-Carlo-Methoden

Zu den Monte-Carlo-Methoden gehören die Methoden, die mithilfe statistischer Stichproben ein mathematisches Problem annäherungsweise lösen. Dies sind z. B. die Berechnung eines Flächeninhaltes, die numerische Lösung der Differentialgleichungen oder die (hochdimensionale) numerische Integration (Uhlmann 1965: S. 116; Neumann 2013: S. 193-194). Die Berechnung der hochdimensionalen Integrale ist bei den Bayes-Ansätzen ein großes Problem, das sich unter anderem mithilfe der Monte-Carlo-Methoden lösen lässt. Die Simulation von Zufallsprozessen kann zur numerischen Berechnung von Integralen eingesetzt werden.

Das gesuchte Integral lässt sich numerisch berechnen, wenn es sich auf folgende Weise faktorisieren lässt:

$$I = \int f(x)g(x)dx^1, \quad (4.1)$$

mit  $f(x)$  - Wahrscheinlichkeitsdichte:  $f(x) \geq 0$  und  $\int f(x)dx = 1$ . Das gesuchte Integral ist in diesem Fall der Erwartungswert der Funktion  $g(x)$ :

$$\int f(x)g(x)dx = E(g(x)). \quad (4.2)$$

Zieht man eine Stichprobe  $\xi_1, \dots, \xi_n$  aus  $f(x)$ , kann das gesuchte Integral durch den Stichprobenmittelwert  $\bar{g}$  geschätzt werden:

$$\int f(x)g(x)dx = E(g(x)) \approx \bar{g} = \frac{1}{n} \sum_{i=1}^n g(\xi_i) \quad (4.3)$$

(Neumann 2013: S. 210-211).

### 4.2. MCMC

Das Markov-Ketten-Monte-Carlo-Verfahren (MCMC - *Markov-Chain-Monte-Carlo*) ist eine der Monte-Carlo-Methoden, welches auf den Ziehungen aus den approximativen Verteilungen und deren Korrektur basiert (Gelman et al. 2013: S. 275). Diese Methode eignet sich unter anderem für die Schätzung der *a posteriori*  $p(\underline{\theta}|\underline{x})$ . MCMC ist dann anzuwenden, wenn es nicht möglich ist, die Parameter  $\underline{\theta}$  direkt aus der *a posteriori*  $p(\underline{\theta}|\underline{x})$  zu ziehen. Die Parameter  $\underline{\theta}$  werden aus einer Verteilung bestimmt, die gegen die

---

<sup>1</sup>Hier kann sowohl ein ein- oder ein hochdimensionales Integral gesucht werden. Ebenso kann  $x$  eine einzelne Variable oder ein Vektor sein.

$a$  posteriori  $p(\underline{\theta}|\underline{x})$  konvergiert. Dieses Verfahren hat gewisse Vorteile, die die Verwendung vor allem der komplexen Bayes-Ansätze praktisch erleichtern. Die interessierende Dichte muss nur bis auf eine Normierungskonstante bekannt sein, d. h. die Berechnung der marginalen Dichte der Daten  $f(\underline{x})$  (Formel 3.2) kann ausgelassen werden und die Proportionalität der  $a$  posteriori zum Produkt aus der *Likelihood* und  $a$  priori (Formel 3.3) kann benutzt werden. Hierbei können auch Zufallszahlen aus hochdimensionalen Dichten gezogen werden. Der zugrunde liegende Prozess in diesem Verfahren ist die Markov-Kette (auch Markov-Prozess genannt) (Def. 3).

**Definition 3 (Markov-Kette).**  $\underline{P}$  sei eine  $k \times k$ -Matrix mit Einträgen  $(P_{i,j})_{i,j=1,\dots,k}$ . Ein (zeitdiskreter) stochastischer Prozess  $(X_0, X_1, \dots)$ <sup>2</sup> mit endlichem Zustandsraum  $\{s_1, \dots, s_k\}$  heißt genau dann (homogene<sup>3</sup>) Markov-Kette mit Übergangsmatrix  $\underline{P}$ , wenn für alle  $n \in \mathbb{N}$ , alle  $i, j \in \{1, \dots, k\}$  und alle  $i_0, \dots, i_{n-1} \in \{1, \dots, k\}$  gilt:

$$P(X_{n+1} = s_j | X_n = s_i, X_{n-1} = s_{i_{n-1}}, \dots, X_0 = s_{i_0}) = P(X_{n+1} = s_j | X_n = s_i) = P_{i,j} \quad (4.4)$$

mit  $P_{i,j}$  - die Übergangswahrscheinlichkeit (die bedingte Wahrscheinlichkeit, sich zum Zeitpunkt  $n + 1$  im Zustand  $s_j$  zu befinden, gegeben, dass man zum Zeitpunkt  $n$  im Zustand  $s_i$  ist) (Vgl. Schomaker 2012: S. 3).

Das Grundprinzip des MCMC-Verfahrens ist es, eine Markov-Kette zu erzeugen, deren stationäre<sup>4</sup> Verteilung die spezifizierte  $a$  posteriori  $p(\underline{\theta}|\underline{x})$  ist und die Simulation dann so lange laufen zu lassen, dass die Verteilung der aktuellen Ziehungen gegen diese stationäre Verteilung konvergiert (Robert 2007: S. 301-303; Gelman et al. 2013: S. 275-276). Der Erwartungswert lässt sich hierbei durch den Mittelwert der Markov-Kette ohne *burn-in* (die ersten Ziehungen vor dem die Konvergenz erreicht wurde) schätzen. Es gibt verschiedene Algorithmen, die es ermöglichen, MCMC-Verfahren praktisch umzusetzen. Der *Metropolis-Hastings-Algorithmus* (Kap. 4.3) und der *Gibbs-Sampler* (Kap. 4.4) sind die grundlegende Algorithmen.

### 4.3. Metropolis-Hastings-Algorithmus (MH-Algorithmus)

Der Metropolis-Hastings-Algorithmus wurde von Metropolis et al. 1953: S. 1087–1092 und Hastings 1970: S. 97–109 für die Integration komplexer Funktionen mithilfe der Zufallsziehung eingeführt. Diese Methode ist dann anzuwenden, wenn eine direkte Ziehung aus der Dichte schwierig oder unmöglich ist. Ziel des MH-Algorithmus ist, die Markov-Kette zu erzeugen, sodass die Dichte vom aktuellen Zustand abhängt. Der Algorithmus basiert auf der Vorschlagsdichte, aus der die Vorschlagswerte generiert werden

<sup>2</sup>Ein (zeitdiskreter) stochastischer Prozess  $(X_0, X_1, \dots)$  ist eine Folge von Zufallsvariablen  $X_n$  mit demselben Wertebereich. Ist dieser Wertebereich gleich der Menge  $S$ , so spricht man von einem  $S$ -wertigen stochastischen Prozess (Kersting and Wakolbinger 2014: S. 33).

<sup>3</sup>Unter Homogenität der Markov-Kette ist die Zeitunabhängigkeit zu verstehen: zu allen Zeitpunkten  $n$  liegen dieselben Übergangswahrscheinlichkeiten  $P_{i,j}$  vor. Siehe dazu auch Stationarität<sup>4</sup>.

<sup>4</sup>Ein stochastischer Prozess heißt stationär, wenn die Verteilungen der Zufallsvariablen und darüber hinaus alle endlichdimensionalen Verteilungen des Prozesses zeitlich konstant bleiben, also derselben Verteilung genügen (Heller et al. 1978: S. 131). Siehe dazu auch Homogenität<sup>3</sup>.

und entsprechend der bestimmten Wahrscheinlichkeit (Akzeptanzwahrscheinlichkeit) angenommen bzw. abgelehnt werden. Im Falle der Ablehnung wird die letzte Ziehung (der aktuelle Zustand) der Kette als nächste Ziehung (der neue Zustand) beibehalten. Es wird die Existenz einer komplexen stetigen Zieldichte  $\pi(\theta) = f(\theta)/K$ , mit  $f(\theta)$  - nicht normierte Dichte und  $K$  - die Normierungskonstante (kann unbekannt sein) angenommen. Weiterhin existiert eine Vorschlagsdichte  $q(\theta)$ , aus der die Vorschlagswerte  $\vartheta_i$ ,  $i = 1, \dots, B$  gezogen werden können, die mit einer Akzeptanzwahrscheinlichkeit

$$\alpha_i(\theta_i, \vartheta) = \min \left( 1, \frac{\pi(\vartheta|\underline{x})q(\theta_i|\vartheta)}{\pi(\theta_i|\underline{x})q(\vartheta|\theta_i)} \right), \quad i = 1, \dots, B \quad (4.5)$$

angenommen werden (Vgl. [Chib and Greenberg 1995](#): S. 329). Da die Akzeptanzwahrscheinlichkeit den Quotient der zwei *a posteriori* Dichten enthält, wird sich die Normierungskonstante wegekürzen (d. h. die Normierungskonstante  $K$  kann unbekannt sein). Ist die Vorschlagsdichte unabhängig von der letzten Ziehung (der aktuelle Zustand)  $q(\vartheta_i|\theta_{i-1}) = q(\vartheta_i)$ , wird die Akzeptanzwahrscheinlichkeit folgendermaßen modifiziert:

$$\alpha_i(\theta_i, \vartheta) = \min \left( 1, \frac{\pi(\vartheta|\underline{x})q(\theta_i)}{\pi(\theta_i|\underline{x})q(\vartheta)} \right), \quad i = 1, \dots, B. \quad (4.6)$$

Wenn die Vorschlagsdichte eine symmetrische Dichte<sup>5</sup> ist ( $q(\theta_i|\vartheta) = q(\vartheta|\theta_i)$ ), reduziert sich die Akzeptanzwahrscheinlichkeit zu

$$\alpha_i(\theta_i, \vartheta) = \min \left( 1, \frac{\pi(\vartheta|\underline{x})}{\pi(\theta_i|\underline{x})} \right), \quad i = 1, \dots, B. \quad (4.7)$$

Die Vorschlagsdichte spielt hierbei wichtige Rolle. Wie diese auszuwählen ist, wird genauer im [Kap. 4.7.1](#) beschrieben. Der MH-Algorithmus ist in der [Tabelle 4.1](#) dargestellt. Der MH-Algorithmus lässt sich für Verteilungen mit zwei (bzw. mehr) unbekannt Parametern erweitern<sup>6</sup>. In Kombination mit anderen Methoden, z. B. Gibbs-Sampler ([Kap. 4.4](#)), Simulated Annealing (SA) ([Kap. 4.5](#)), etc., ermöglicht der MH-Algorithmus, die Bayes-Ansätze für beliebig komplexe Verteilungen sowohl von Daten (*Likelihood*) als auch von *a priori* praktisch anzuwenden. Das ist besonders wichtig, wenn die analytische Herleitung komplex oder unmöglich ist.

---

<sup>5</sup>Der Algorithmus mit symmetrischen Vorschlagsdichten ist als *Metropolis-Algorithmus* bekannt, der von [Hastings 1970](#): S. 97–109 für nicht symmetrische Vorschlagsdichten verallgemeinert wurde.

<sup>6</sup>Erweiterung des MH-Algorithmus für zwei unbekannt Parameter siehe [Kap. 4.7](#).

---

**Algorithmus:** MH-Algorithmus

---

1. Initialisiere Startwert  $\theta_0$

Für  $i = 1, \dots, B$ :

2. Generiere Vorschlagswert  $\vartheta_i$  aus Vorschlagsdichte  $q(\vartheta_i|\theta_{i-1})$  und Zufallszahl  $u_i$  aus Gleichverteilung  $U(0, 1)$

4. Akzeptiere  $\vartheta_i$  mit der Wahrscheinlichkeit  $\alpha_i$ :

$$\theta_i = \begin{cases} \vartheta_i, & \text{wenn } u_i \leq \alpha_i(\theta_{i-1}, \vartheta_i) = \min \left( 1, \frac{\pi(\vartheta_i|\underline{x})q(\theta_{i-1}|\vartheta_i)}{\pi(\theta_{i-1}|\underline{x})q(\vartheta_i|\theta_{i-1})} \right) \\ \theta_{i-1}, & \text{sonst} \end{cases}$$

5. Gebe die erhaltenen Werte  $\theta_1, \dots, \theta_B$  zurück

---

Tabelle 4.1.: MH-Algorithmus<sup>7</sup>.

#### 4.4. Gibbs-Sampler

Der Gibbs-Sampler ist ein Spezialfall des Ein-Schritt-MH-Algorithmus, der nach J. W. Gibbs benannt und von [Geman and Geman 1984](#): S. 721-741 entwickelt wurde. Diese Methode eignet sich besonders dann, wenn die gemeinsame Verteilung des Zufallsvektors  $q(\underline{\theta}) = q(\theta_1, \dots, \theta_k)$  unbekannt, jedoch die bedingte Verteilung einer jeden Zufallsvariable  $q_j(\theta_j|\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k) = q_j(\theta_j|\underline{\theta}_{-j})$  (*full conditional*) bekannt ist. Der Gibbs-Sampler-Algorithmus ist in der Tabelle 4.2 dargestellt.

---

**Algorithmus:** Gibbs-Sampler

---

1. Initialisiere Startwerte  $\theta_2^0, \dots, \theta_k^0$

2. Für  $i = 1, \dots, B$ :

2.1. Generiere  $\theta_1^i$  aus  $q_1(\theta_1|\theta_2^{i-1}, \dots, \theta_k^{i-1})$

2.2. Generiere  $\theta_2^i$  aus  $q_2(\theta_2|\theta_1^i, \theta_3^{i-1}, \dots, \theta_k^{i-1})$

.

.

.

2.k. Generiere  $\theta_k^i$  aus  $q_k(\theta_k|\theta_1^i, \dots, \theta_{k-1}^i)$

3. Gebe die erhaltenen Werte  $\underline{\theta}^1, \dots, \underline{\theta}^B$  zurück

---

Tabelle 4.2.: Gibbs-Sampler<sup>8</sup>.

Da der Gibbs-Sampler ein Spezialfall des Ein-Schritt-MH-Algorithmus ist, können die beiden Algorithmen kombiniert werden. In der Praxis wird öfters der nicht kombinierte MH-Algorithmus bevorzugt, weil die *full conditionals* meistens unbekannt sind und (bzw. oder) die Implementation des nicht kombinierten MH-Algorithmus leichter ist.

---

<sup>7</sup>Der Algorithmus wurde aus [Chib and Greenberg 1995](#): S. 329 übernommen.

<sup>8</sup>Der Algorithmus wurde aus [Dahlberg 2015](#): S. 17-18 übernommen.

## 4.5. Simulated Annealing (SA)

Simulated Annealing (SA) ist ein approximatives Optimierungsverfahren, das hauptsächlich in der Physik verwendet wird (Kirkpatrick et al. 1983: S. 671-680). Die Grundidee der Methode ist die Nachbildung des Abkühlungsprozesses in der Metallurgie. Im Sinne von Optimierungsverfahren kann die Temperatur als die Wahrscheinlichkeit interpretiert werden, mit der sich der Zwischenstand des gesuchten Optimums auch verschlechtern kann, d. h. es kann das lokale Optimum verlassen werden und somit wird nach einem globalen Optimum gesucht. Diese Methode eignet sich zum Optimieren des Prozesses mit mehreren Einflussfaktoren (Parametern) und wird oft in der Statistik mit unterschiedlichen Schätzmethoden kombiniert, beispielsweise MLE, etc. (Ergashev 2008: S. 7-10) . Der SA-Algorithmus ist in der Tabelle 4.3 dargestellt.  $g(\underline{\theta})$  ist hierbei zu minimierende Funktion.

---

### Algorithmus: Simulated Annealing

---

1. Initialisiere Startwerte  $\underline{\theta}_{old} = \underline{\theta}_0, t = t_0 > 0, \Delta, g_{old} = g(\underline{\theta}_{old})$

2. Für  $t = t_0, \dots, 0$ :

2.1. Berechne die neuen Vorschlagswerte

$$\underline{\theta}_{new} = \underline{\theta}_{old} + c\underline{e}z$$

2.2. Berechne den neuen Funktionswert:

$$g_{new} = g(\underline{\theta}_{new})$$

2.3. Akzeptiere und speichere den Funktionswert:

$$g_{old} = \begin{cases} g_{new}, & \text{wenn } g_{old} - g_{new} > 0 \\ g_{new}, & \text{sonst wenn } u \leq \exp((g_{old} - g_{new})/t) \\ g_{old}, & \text{sonst} \end{cases}$$

2.4. Speichere entsprechende Vorschlagswerte:

$$\underline{\theta}_{old} = \begin{cases} \underline{\theta}_{new}, & \text{wenn } g_{old} = g_{new} \\ \underline{\theta}_{old}, & \text{sonst} \end{cases}$$

2.5. Reduziere Temperatur  $t = \Delta t$

3. Gebe die erhaltenen Werte  $\underline{\theta}_{old}$  und  $g_{old}$  zurück

---

**Notation:**  $g(\underline{\theta})$  - zu minimierende Funktion,  $g$  - Funktionswert,  $t$  - die Temperatur,  $\Delta \in (0, 1)$  - Reduktionsfaktor der Temperatur,  $c > 0$  - der Skalierungsfaktor,  $\underline{e}$  - Einheitsvektor mit 1 in zufällig gewählter Dimension,  $z \sim N(0, 1)$  - Zufallszahl,  $u \sim U(0, 1)$  - Zufallszahl.

---

Tabelle 4.3.: Simulated Annealing<sup>9</sup>.

Da der Metropolis-Algorithmus ein SA mit fixer Temperatur ist (Wegener 2004: S. 1), kann der MH-Algorithmus gut mit SA kombiniert werden.

---

<sup>9</sup>Der Algorithmus wurde aus Ergashev 2008: S. 12 übernommen.

Bei der praktischen Anwendung der SA kann das existierende R-Paket “*optimization*” (Husmann et al. 2017: S. 1-16) mit der implementierten R-Funktion “*optim\_sa()*” verwendet werden.

## 4.6. Klassische Kerndichte-Schätzung (KDE)

Für den klassischen Kerndichte-Schätzer (KDE - *Kernel Density Estimator*) muss die Annahme getroffen werden, dass  $X_1, \dots, X_n \sim iid$  sind. Unter dieser Annahme ist die geschätzte univariate Dichte im Punkt  $x$  gegeben als

$$\hat{k}_b(x) = \frac{1}{bn} \sum_{i=1}^n K\left(\frac{x - X_i}{b}\right), \quad (4.8)$$

mit  $b > 0$  - Bandbreite (Glättungsparameter), sowie einer  $K(\cdot)$  - Kernfunktion. Die geschätzte  $d$ -variate Dichte im  $d$ -variaten Datenpunkt ist gegeben als

$$\hat{k}_{\underline{B}}(\underline{x}) = \frac{1}{|\underline{B}|^{1/2}n} \sum_{i=1}^n K\left(\frac{\underline{x} - \underline{X}_i}{\underline{B}^{1/2}}\right), \quad (4.9)$$

mit  $\underline{B} > 0$  -  $d \times d$  symmetrische Bandbreitenmatrix<sup>10</sup> (Glättungsparameter),  $|\cdot|$  - Determinante einer Matrix,  $\underline{x} = (x_1, x_2, \dots, x_d)^T$  - der  $d$ -variate Datenpunkt,  $\underline{X}_i = (X_{i,1}, X_{i,2}, \dots, X_{i,d})^T, i = 1, \dots, n$  - die  $d$ -variate Daten und  $K(\cdot)$  - multivariate Kernfunktion (Wang 2009: 3, 17). Die Auswahl der Kernfunktion und der Bandbreite spielt dabei eine wichtige Rolle. Kernfunktionen müssen positiv, symmetrisch und stetig sein, desweiteren muss die Fläche unter der Kurve eins ergeben (Nedden 2012: S. 119-129). Es gibt unterschiedliche Kernfunktionen, die diese Eigenschaften erfüllen<sup>11</sup>. Allerdings wird im univariaten Fall der Epanechnikov-Kern

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2), & \text{wenn } |x| < 1 \\ 0, & \text{sonst} \end{cases} \quad (4.10)$$

bevorzugt, weil er die mittlere quadratische Abweichung des zugehörigen Kerndichte-Schätzers minimiert. Im  $d$ -variaten Fall kann der  $d$ -variate Gauß-Kern benutzt werden<sup>12</sup>:

$$K(\underline{x}) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}\underline{x}^T \underline{x}\right). \quad (4.11)$$

Die Wahl der Bandbreite spielt bei der Kerndichte-Schätzung ebenfalls eine wichtige Rolle. Je mehr Daten vorhanden sind, desto kleiner ist die Bandbreite  $b$  auszuwählen und desto besser ist die Anpassung der geschätzten Kerndichte an die Daten. Es gibt viele Faustregeln, wie die Bandbreite ausgewählt werden kann. Für eindimensionale Daten wird oft die Regel von Silverman benutzt (Bolance et al. 2012: S. 39):

$$b = 2.34\hat{\sigma}n^{-\frac{1}{5}}, \quad (4.12)$$

<sup>10</sup>Im univariaten Fall ist die Bandbreitenmatrix  $\underline{B} = b^2$ .

<sup>11</sup>Siehe dazu Anhang A.1.

<sup>12</sup>Kernfunktion übernommen aus Duong 2018: S. 1.

mit 2.34 - Konstante<sup>13</sup> für Epanechnikov-Kern,  $\hat{\sigma}$  - geschätzte Standardabweichung der Daten,  $n$  - Anzahl der Beobachtungen. Viele Methoden wurden in den letzten Jahren entwickelt, die zeigen, wie die Glättungsparameter im  $d$ -variaten Fall ausgewählt werden können. Einige davon wurden unter anderen von [Duong 2004](#): S. 1-161 und [Duong and Hazelton 2005](#): S. 485–506 zusammengefasst und analysiert. SCV (*Smoothed Cross Validation*) ist eine der Methoden für die Auswahl der Glättungsparameter, die von [Jones et al. 1991](#): S. 1919–1932 vorgeschlagen wurde. Diese Methode kombiniert MISE (*Mean Integrated Squared Error*) und AMISE (*Asymptotic Mean Integrated Squared Error*)<sup>14</sup> in der zu minimierenden Zielfunktion und ist auch für die Schätzung der Daten mit Duplikaten geeignet ([Duong and Hazelton 2005](#): S. 486-489). Der nach SCV geschätzte Glättungsparameter für den KDE (Formel 4.9) ist folgendermaßen definiert:

$$\hat{\underline{B}} = \operatorname{argmin}(SCV(\underline{B})),$$

wobei

$$SCV(\underline{B}) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (K_{\underline{B}} * K_{\underline{B}} * L_G * L_G - 2K_{\underline{B}} * L_G * L_G + L_G * L_G)(X_i - X_j) \\ + \frac{1}{n|\underline{B}|^{1/2}} \int_{R^d} K(\underline{x})^2 d\underline{x}$$

die Zielfunktion der Methode ist, mit  $K_{\underline{B}}(x) = \frac{1}{|\underline{B}|^{1/2}} K\left(\frac{x}{\underline{B}^{1/2}}\right)$  - die Kernfunktion zu bestimmten Glättungsparameter  $\underline{B}$ , \* - Zeichen für die Faltung und  $L_G(\cdot)$  - Pilotkern mit Pilotbandbreitenmatrix  $G$ <sup>15</sup> ([Duong and Hazelton 2005](#): S. 489). KDE wurde praktisch in R umgesetzt. Die univariate KDE wurde im R-Paket “*stats*” ([R Core Team 2017](#)) R-Funktion “*density()*” implementiert. Die  $d$ -variante KDE wurde im R-Paket “*ks*” ([R Core Team 2017](#)) zusammengefasst: R-Funktion “*kde()*” - für Berechnung der  $d$ -variaten KDE, R-Funktion “*Hscv()*” - für die Berechnung der Bandbreitenmatrix. Diese kann zur Dichteschätzung der mit MH simulierten Parameter genutzt werden.

## 4.7. Erweiterungen des MH-Algorithmus

In dieser Arbeit wird der MH-Algorithmus für zwei unbekannt Parameter (Kap. 4.3) und die Kombination<sup>16</sup> von MH-Algorithmus und SA (Kap. 4.5) verglichen. Der auf Basis der Analysen ausgewählte Algorithmus wird im Zwei-Schritt-Bayes-Ansatz (Kap. 5) verwendet.

<sup>13</sup>Näheres zu der Konstante bei der Auswahl der Bandbreite unter der Benutzung von unterschiedlichen Kernfunktionen siehe [Hansen 2009](#): S. 1-24.

<sup>14</sup>Das Ziel der meisten Methoden für den Auswahl der Glättungsparameter ist, MISE oder AMISE zu minimieren.

<sup>15</sup>Der Pilotkern sorgt für die Betrachtung der Duplikate in den Daten.

<sup>16</sup>Die Kombination von MH-Algorithmus und SA wurde vom [Dahlberg 2015](#): S. 1-60 beschrieben.

### 4.7.1. Wahl der Vorschlagsdichte und deren Parameter

Die Vorschlagsdichte spielt im MH-Algorithmus eine sehr wichtige Rolle. Zum einen muss diese eine möglichst einfache Dichte sein, um die Anwendung einfach und leicht zu machen. Zum anderen müssen die simulierten Werte eine gute Streuung (Vermischung) aufweisen und die Akzeptanzrate muss hoch genug sein. In dieser Arbeit wird die Dichte der Normal-Verteilung als Vorschlagsdichte für die Simulation der Lokations- und die Dichte der Lognormal-Verteilung als Vorschlagsdichte für die Simulation der Skalenparameter der Lognormal-Verteilung benutzt<sup>17</sup>. Eine Möglichkeit ist, die auf die letzte Ziehung (der aktuelle Zustand) bedingte Vorschlagsdichte zu benutzen. In diesem Fall ist der Lokationsparameter der Normal- bzw. Lognormal-Verteilung gleich dem Wert der letzten Ziehung  $q(\vartheta_i|\theta_{i-1}) = N(\theta_{i-1}, \sigma)$  bzw.  $q(\vartheta_i|\theta_{i-1}) = LN(\theta_{i-1}, \sigma)$ . Der Skalenparameter  $\sigma$  muss hierbei so gewählt werden, dass die Akzeptanzwahrscheinlichkeit nicht zu niedrig ist<sup>18</sup>. Der erweiterte MH-Algorithmus für zwei unbekanntem Parameter und bedingter Vorschlagsdichte ist in der Tabelle 4.4 dargestellt.

Eine weitere Möglichkeit ist, die Vorschlagsdichte unabhängig von der letzten Ziehung (dem aktuellen Zustand) zu machen. In diesem Fall muss die Vorschlagsdichte im Modus und Standardabweichung ähnlich der Zieldichte sein (Dahlberg 2015: S. 19). Die Auswahl der Lokations- und Skalenparameter geschieht auf Basis von SA (Kap. 4.5) - Bestimmung des Modus - und Berechnung der Krümmung in diesem Modus. Somit wird die Vorschlagsdichte benutzt, die nicht vom vorherigen Wert abhängt  $q(\vartheta_i|\theta_{i-1}) = q(\vartheta_i) = N(\mu, \sigma)$  bzw.  $q(\vartheta_i|\theta_{i-1}) = q(\vartheta_i) = LN(\mu, \sigma)$ .

Die Krümmung der Zieldichte<sup>19</sup> kann durch die negative Hesse-Matrix im gesuchten Parameter (Modus) approximiert werden:

$$k = -\frac{\partial^2 p(\underline{\theta}|\underline{x})}{\partial^2 \theta_{Modus}}, \quad (4.13)$$

die für die Berechnung der Skalenparameter für die Vorschlagsdichte verwendet wird. Bei der praktischen Berechnung der Hesse-Matrix kann das existierende R-Paket “*num-Deriv*” (Gilbert, Paul and Varadhan, Ravi 2016: S. 1-12) mit der implementierten R-Funktion “*hessian()*” verwendet werden, das die numerische Lösung beliebiger komplexer Funktionen ermöglicht.

---

<sup>17</sup>Die Vorschlagsdichte für die Simulation der Skalenparameter *sdlog* der Lognormal-Verteilung muss auf dem positiven Bereich definiert werden, da der Skalenparameter nicht negativ werden darf. Für andere Verteilungen, falls der Lokationsparameter auch nicht negativ werden darf (z.B. Weibull-Verteilung), kann auch die Lognormal-Verteilung als Vorschlagsdichte benutzt werden. Der Algorithmus muss entsprechend angepasst werden.

<sup>18</sup>Eine ausreichende Akzeptanzwahrscheinlichkeit liegt bei 10% – 40%, eine gute Akzeptanzwahrscheinlichkeit liegt bei 40% – 80%.

<sup>19</sup>Zur Vereinfachung der Berechnung und numerische Stabilität muss die *Log-Likelihood* der Parameter statt der *Likelihood* der Parameter benutzt werden.

---

**Algorithmus:** MH-Algorithmus für zwei unbekannte Parameter

---

1. Initialisiere Startwerte  $\theta_{1,(i=0)}$  und  $\theta_{2,(i=0)}$   
entsprechend dem Definitionsbereich der Dichten  $\pi(\theta_1|\underline{x})$  und  $\pi(\theta_2|\underline{x})$

Für  $i = 1, \dots, B$ :

2. Generiere Vorschlagswert  $\vartheta_{1,(i)}$  aus Vorschlagsdichte  $q(\vartheta_1|\theta_{1,(i-1)})$   
und Zufallszahl  $u_{1,(i)}$  aus Gleichverteilung  $U(0, 1)$
3. Akzeptiere  $\vartheta_{1,(i)}$  mit der Wahrscheinlichkeit  $\alpha_{1,(i)}$ :

$$\theta_{1,(i)} = \begin{cases} \vartheta_{1,(i)}, & \text{wenn } u_{1,(i)} \leq \alpha_{1,(i)}(\theta_{1,(i-1)}, \vartheta_{1,(i)}) \\ \theta_{1,(i-1)}, & \text{sonst} \end{cases}$$

$$\text{mit } \alpha_{1,(i)}(\theta_{1,(i-1)}, \vartheta_{1,(i)}) = \min \left( 1, \frac{\pi(\vartheta_{1,(i)}|\underline{x}, \theta_{2,(i-1)})q(\theta_{1,(i-1)}|\vartheta_{1,(i)})}{\pi(\theta_{1,(i-1)}|\underline{x}, \theta_{2,(i-1)})q(\vartheta_{1,(i)}|\theta_{1,(i-1)})} \right)$$

4. Generiere Vorschlagswert  $\vartheta_{2,(i)}$  aus Vorschlagsdichte  $q(\vartheta_2|\theta_{2,(i-1)})$   
und Zufallszahl  $u_{2,(i)}$  aus Gleichverteilung  $U(0, 1)$
5. Akzeptiere  $\vartheta_{2,(i)}$  mit der Wahrscheinlichkeit  $\alpha_{2,(i)}$ :

$$\theta_{2,(i)} = \begin{cases} \vartheta_{2,(i)}, & \text{wenn } u_{2,(i)} \leq \alpha_{2,(i)}(\theta_{2,(i-1)}, \vartheta_{2,(i)}) \\ \theta_{2,(i-1)}, & \text{sonst} \end{cases}$$

$$\text{mit } \alpha_{2,(i)}(\theta_{2,(i-1)}, \vartheta_{2,(i)}) = \min \left( 1, \frac{\pi(\vartheta_{2,(i)}|\underline{x}, \theta_{1,(i)})q(\theta_{2,(i-1)}|\vartheta_{2,(i)})}{\pi(\theta_{2,(i-1)}|\underline{x}, \theta_{1,(i)})q(\vartheta_{2,(i)}|\theta_{2,(i-1)})} \right)$$

6. Gebe die erhaltenen Werte  $\underline{\theta}_1$  und  $\underline{\theta}_2$  zurück

---

**Bemerkung:**  $\theta_{2,(i-1)}$  im Schritt 3. und  $\theta_{1,(i)}$  im Schritt 5. werden als feste bekannte Größen angenommen.

---

Tabelle 4.4.: Tab. MH-Algorithmus für zwei unbekannte Parameter<sup>20</sup>.

Die Krümmung des Graphen einer flachen Funktion im Punkt  $x$  ist gegeben durch:

$$k(x) = \frac{|f''(x)|}{(1 + (f'(x))^2)^{\frac{3}{2}}}, \quad (4.14)$$

mit  $f'(x)$  und  $f''(x)$  - die erste und die zweite Ableitung im Punkt  $x$ . Die laut Formel 4.14 ausgerechnete Krümmung der Normal-Verteilung im Modus-Punkt ( $x = \mu$ ) ist gegeben als<sup>21</sup>:

$$k_N(\mu) = \frac{1}{\sigma^3 \sqrt{2\pi}}. \quad (4.15)$$

Die laut Formel 4.14 ausgerechnete Krümmung der Lognormal-Verteilung im Modus-

---

<sup>20</sup>Der erweiterte Algorithmus wurde auf Basis [Dahlberg 2015](#): S. 19 ermittelt.

<sup>21</sup>Herleitung siehe Anh. [A.2.1](#).

Punkt ( $x = \exp(\mu - \sigma^2)$ ) ist gegeben als<sup>22</sup>:

$$k_{LN}(\exp(\mu - \sigma^2)) = \frac{\exp(\frac{1}{2})}{Modus^3 \sigma^3 \sqrt{2\pi}}. \quad (4.16)$$

Die Auflösung der Formeln 4.15 und 4.16 nach  $\sigma$  ergibt die gesuchten Skalenparameter der Normal- bzw. Lognormal-Verteilung in Abhängigkeit von der Krümmung<sup>23</sup>:

$$\sigma_N = \left( \frac{1}{k\sqrt{2\pi}} \right)^{\frac{1}{3}} \quad (4.17)$$

und

$$\sigma_{LN} = \left( \frac{\exp(\frac{1}{2})}{kModus^3 \sqrt{2\pi}} \right)^{\frac{1}{3}}. \quad (4.18)$$

Der um die SA und die Berechnung der Krümmung erweiterte MH-Algorithmus<sup>24</sup> ist in der Tabelle 4.5 dargestellt.

---

<sup>22</sup>Herleitung siehe Anh. A.2.2.

<sup>23</sup>Die Subskripte  $N$  und  $LN$  stehen für Normal- und Lognormal-Verteilung.

<sup>24</sup>Hierbei wird der MH-Algorithmus für zwei unbekannte Parameter 4.4 benutzt.

---

**Algorithmus:** MH-SA-Algorithmus für zwei unbekannte Parameter

---

1. Initialisiere Startwerte  $\theta_{1,(i=0)}$  und  $\theta_{2,(i=0)}$

Für  $i = 1, \dots, B$ :

2. Optimierte die Zielfunktion  $\pi(\theta_{1,(i)}|\underline{x}, \theta_{2,(i-1)})$  bzgl. ersten Parameter  $\theta_{1,(i)}$  mit SA-Algorithmus (Tabelle 4.3) und nutze den optimalen Wert  $\hat{m}_{\theta_{1,(i)}}$  als Modus für die Vorschlagsdichte

3. Berechne die Krümmung der Zieldichte (Formel 4.13) im Punkt  $\hat{m}_{\theta_{1,(i)}}$  und die dazugehörige Standardabweichung  $\hat{sd}_{\theta_{1,(i)}}$  für die Vorschlagsdichte

4. Generiere den Vorschlagswert  $\vartheta_{1,(i)}$  aus Vorschlagsdichte  $q(\vartheta_1|\hat{m}_{\theta_{1,(i)}}, \hat{sd}_{\theta_{1,(i)}})$  und Zufallszahl  $u_{1,(i)}$  aus Gleichverteilung  $U(0, 1)$

5. Akzeptiere  $\vartheta_{1,(i)}$  mit der Wahrscheinlichkeit  $\alpha_{1,(i)}$ :

$$\theta_{1,(i)} = \begin{cases} \vartheta_{1,(i)}, & \text{wenn } u_{1,(i)} \leq \alpha_{1,(i)}(\theta_{1,(i-1)}, \vartheta_{1,(i)}) \\ \theta_{1,(i-1)}, & \text{sonst} \end{cases}$$

$$\text{mit } \alpha_{1,(i)}(\theta_{1,(i-1)}, \vartheta_{1,(i)}) = \min \left( 1, \frac{\pi(\vartheta_{1,(i)}|\underline{x}, \theta_{2,(i-1)})q(\theta_{1,(i-1)}|\hat{m}_{\theta_{1,(i)}}, \hat{sd}_{\theta_{1,(i)}})}{\pi(\theta_{1,(i-1)}|\underline{x}, \theta_{2,(i-1)})q(\vartheta_{1,(i)}|\hat{m}_{\theta_{1,(i)}}, \hat{sd}_{\theta_{1,(i)}})} \right)$$

6. Optimierte die Zielfunktion  $\pi(\theta_{2,(i)}|\underline{x}, \theta_{1,(i)})$  bzgl. zweiten Parameter  $\theta_{2,(i)}$  mit SA-Algorithmus (Tabelle 4.3) und nutze den optimalen Wert  $\hat{m}_{\theta_{2,(i)}}$  als Modus für die Vorschlagsdichte

7. Berechne die Krümmung der Zieldichte (Formel 4.13) im Punkt  $\hat{m}_{\theta_{2,(i)}}$  und die dazugehörige Standardabweichung  $\hat{sd}_{\theta_{2,(i)}}$  für die Vorschlagsdichte

8. Generiere den Vorschlagswert  $\vartheta_{2,(i)}$  aus Vorschlagsdichte  $q(\vartheta_2|\hat{m}_{\theta_{2,(i)}}, \hat{sd}_{\theta_{2,(i)}})$  und Zufallszahl  $u_{2,(i)}$  aus Gleichverteilung  $U(0, 1)$

9. Akzeptiere  $\vartheta_{2,(i)}$  mit der Wahrscheinlichkeit  $\alpha_{2,(i)}$ :

$$\theta_{2,(i)} = \begin{cases} \vartheta_{2,(i)}, & \text{wenn } u_{2,(i)} \leq \alpha_{2,(i)}(\theta_{2,(i-1)}, \vartheta_{2,(i)}) \\ \theta_{2,(i-1)}, & \text{sonst} \end{cases}$$

$$\text{mit } \alpha_{2,(i)}(\theta_{2,(i-1)}, \vartheta_{2,(i)}) = \min \left( 1, \frac{\pi(\vartheta_{2,(i)}|\underline{x}, \theta_{1,(i)})q(\theta_{2,(i-1)}|\hat{m}_{\theta_{2,(i)}}, \hat{sd}_{\theta_{2,(i)}})}{\pi(\theta_{2,(i-1)}|\underline{x}, \theta_{1,(i)})q(\vartheta_{2,(i)}|\hat{m}_{\theta_{2,(i)}}, \hat{sd}_{\theta_{2,(i)}})} \right)$$

10. Gebe die erhaltenen Werte  $\underline{\theta}_1$  und  $\underline{\theta}_2$  zurück

---

**Bemerkung:**  $\theta_{2,(i-1)}$  im Schritt 2. und 5. und  $\theta_{1,(i)}$  im Schritt 6. und 9. werden als feste bekannte Größen angenommen.

---

$$\pi(\theta_1, \theta_2|\underline{x}) = \pi(\theta_1|\underline{x}, \theta_{2,est}) = \pi(\theta_2|\underline{x}, \theta_{1,est})$$

---

Tabelle 4.5.: MH-SA-Algorithmus für zwei unbekannte Parameter<sup>25</sup>.

---

<sup>25</sup>Der MH-Algorithmus für zwei unbekannte Parameter (Tabelle 4.4) wurde hierbei als Basis-Algorithmus genommen.

### 4.7.2. MH- und MH-SA-Algorithmus im Vergleich

Um beide Algorithmen (MH- und MH-SA-Algorithmus) zu vergleichen, wurden die Daten aus der Lognormal-Verteilung mit den wahren Parametern  $meanlog = 16.2$  und  $sdlog = 0.96$  simuliert ( $seed = 5728$ ). Die empirischen Parameter der Stichprobe sind  $meanlog = 16.14274$  und  $sdlog = 0.91307$ . Für die Übersichtlichkeit wurde eine konstante  $a priori$  gleich Eins als gemeinsame  $a priori$  für beide Parameter angenommen. Das Ziel war, die Parameter aus der gemeinsamen  $a posteriori$  zu ziehen. In der Abb. 4.1 und Abb. 4.2 sind Simulationen des Parameters  $meanlog$  der Lognormal-Verteilung mit MH- (Abb. 4.1) und MH-SA-Algorithmus (Abb. 4.2) dargestellt. Jeweils oben links ist die simulierte Markov-Kette für 1250 Iterationen abgebildet. In beiden Fällen konvergiert der Mittelwert der Markov-Kette gegen den empirischen Parameter. Jeweils in der unteren Zeile der Abb. 4.1 und Abb. 4.2 sind die Histogramme für alle (unten links) und letzte 1000 (unten rechts) Iterationen dargestellt.

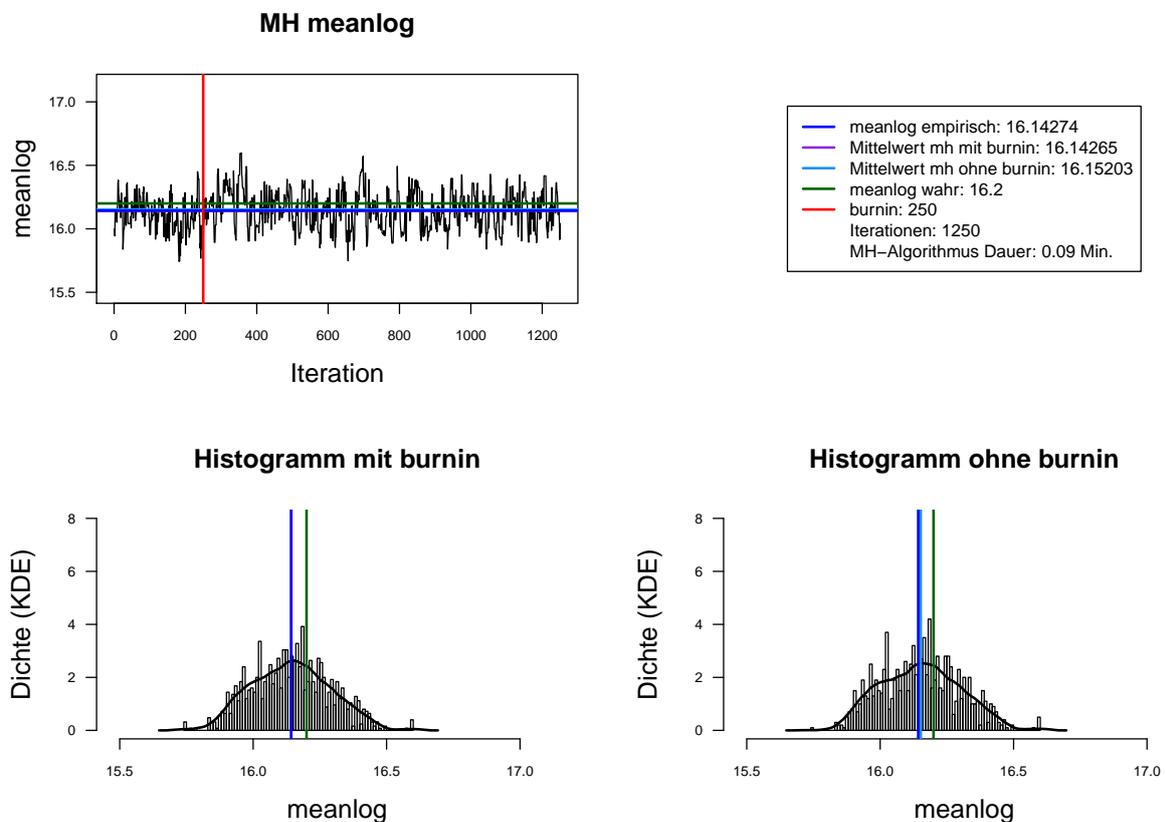


Abbildung 4.1.: Der mit MH-Algorithmus simulierte  $meanlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 1250 Iterationen beträgt 0.09 Minuten. Markov-Kette (oben links), Histogramm für 1250 Iterationen mit KDE (unten links) und Histogramm für letzte 1000 Iterationen mit KDE (unten rechts).

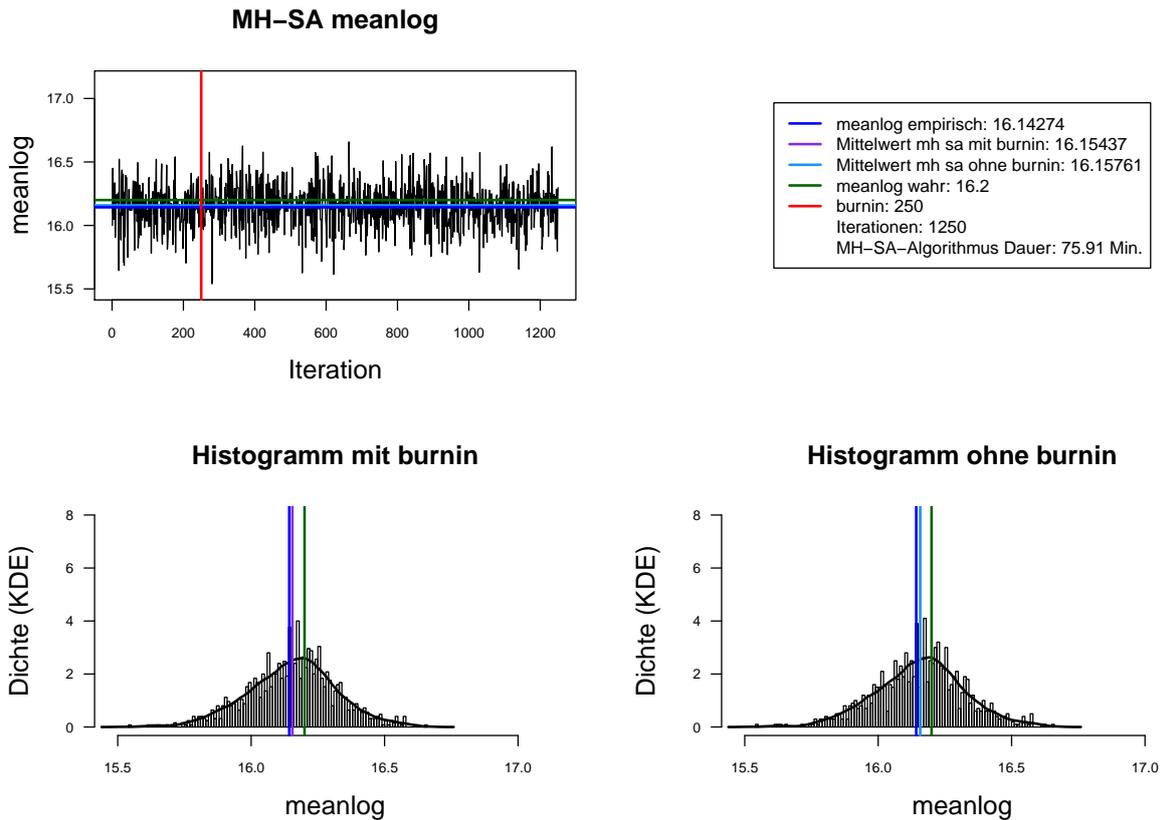
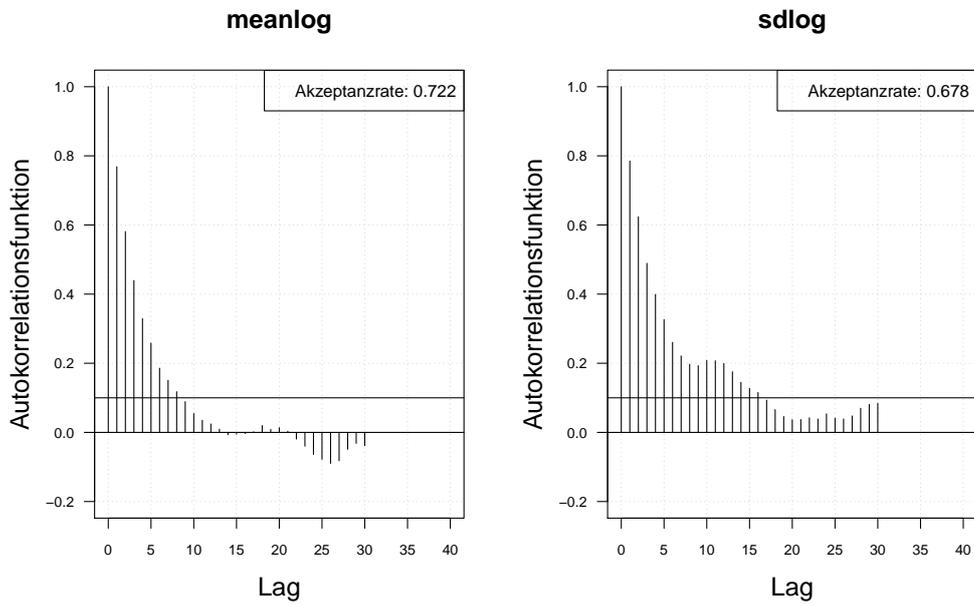


Abbildung 4.2.: Der mit MH-SA-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1250 Iterationen beträgt 75.91 Minuten = 1.27 Stunden. Markov-Kette (oben links), Histogramm für 1250 Iterationen mit KDE (unten links) und Histogramm für letzte 1000 Iterationen mit KDE (unten rechts).

Aus den Histogrammen ist die bessere Stabilität und Approximation des MH-SA- im Vergleich zu MH-Algorithmus erkennbar. Allerdings dauert der MH-SA- bei gleicher Anzahl der Iterationen ca. 850 Mal länger als der MH-Algorithmus (Iterationen: 1250, MH: 0.09 Minuten, MH-SA: 75.91 Minuten). Dies ist durch die Optimierung (SA) in jeder Iteration für jeden Parameter erklärbar. Mit einer wachsenden Anzahl der Iterationen werden simulierte Verteilungen mit beiden Algorithmen immer ähnlicher und der Unterschied in den Laufzeiten immer größer. Für die ausreichende Anzahl der Iterationen<sup>26</sup> (ca. 3000 Iterationen) liefern beide Algorithmen ähnliche Ergebnisse, d. h. der MH-Algorithmus ist aus Sicht der Laufzeiten für die praktische Umsetzung attraktiver.

<sup>26</sup>Für Ergebnisse beider Algorithmen für beide Parameter der Lognormal-Verteilung mit unterschiedlicher Anzahl der Iterationen siehe Anh. A.3.

Autokorrelation MH ohne burnin, Iterationen: 1000



Autokorrelation MH-SA ohne burnin, Iterationen: 1000

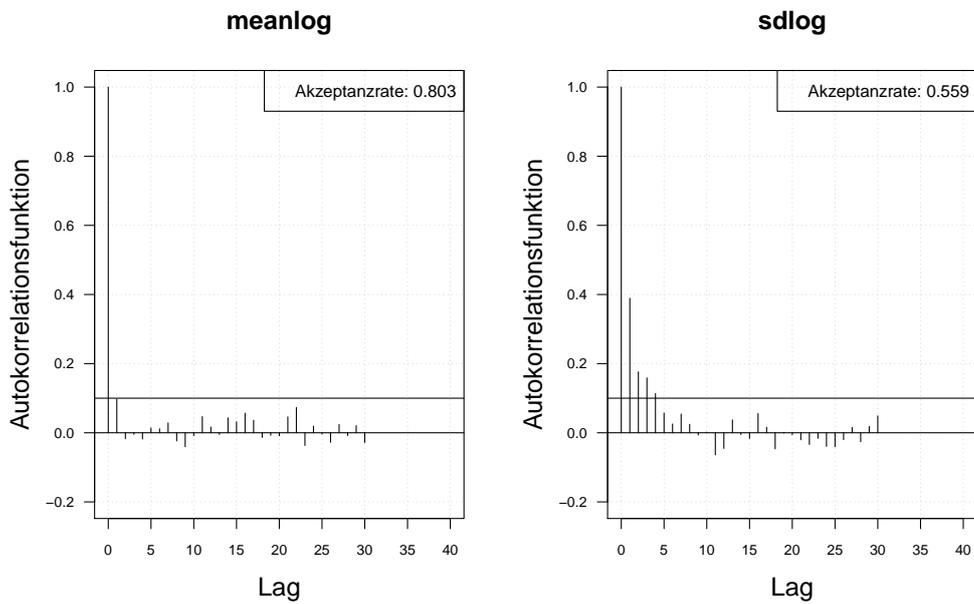


Abbildung 4.3.: Die Autokorrelation der mit MH- (obere Zeile) und mit MH-SA-Algorithmus (untere Zeile) simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 1000.

Ein weiterer wichtiger Punkt beim Vergleich der beiden Algorithmen ist die Autokorrelation der simulierten Markov-Kette. Da die Parameter in Bayes-Ansätzen als Zufallsvariablen betrachtet werden und die Annahme einer unabhängigen identischen Verteilung ( $\Theta \text{ iid}$ ) getroffen wird, muss bei der Simulation idealerweise keine Autokorrelation auftreten.

In Abb. 4.3 ist die Autokorrelation der zwei mit beiden Algorithmen simulierten Parametern abgebildet. Eine langsam abnehmende Autokorrelationsfunktion bedeutet, dass der Algorithmus eine niedrige Streuung (Vermischung) hat und dass nicht nur die nacheinander gezogenen Werte sondern auch die weiter in der Vergangenheit (Lag größer Eins) gezogenen Werte miteinander korrelieren. Die Autokorrelationsfunktion des MH-SA-Algorithmus für beide Parameter (Abb. 4.3, untere Zeile) nimmt viel schneller ab als die Autokorrelationsfunktion des MH-Algorithmus (Abb. 4.3, obere Zeile). Die Akzeptanzrate des MH-SA-Algorithmus ( $meanlog : 0.803$  und  $sdlog : 0.559$ ) für beide Parameter liegt etwa im gleichen Bereich wie die Akzeptanzrate des MH-Algorithmus ( $meanlog : 0.722$  und  $sdlog : 0.678$ ). Das bedeutet, dass der MH-SA-Algorithmus aus Sicht der Unabhängigkeit der Ziehungen und Vermischung besser ist als der MH-Algorithmus.

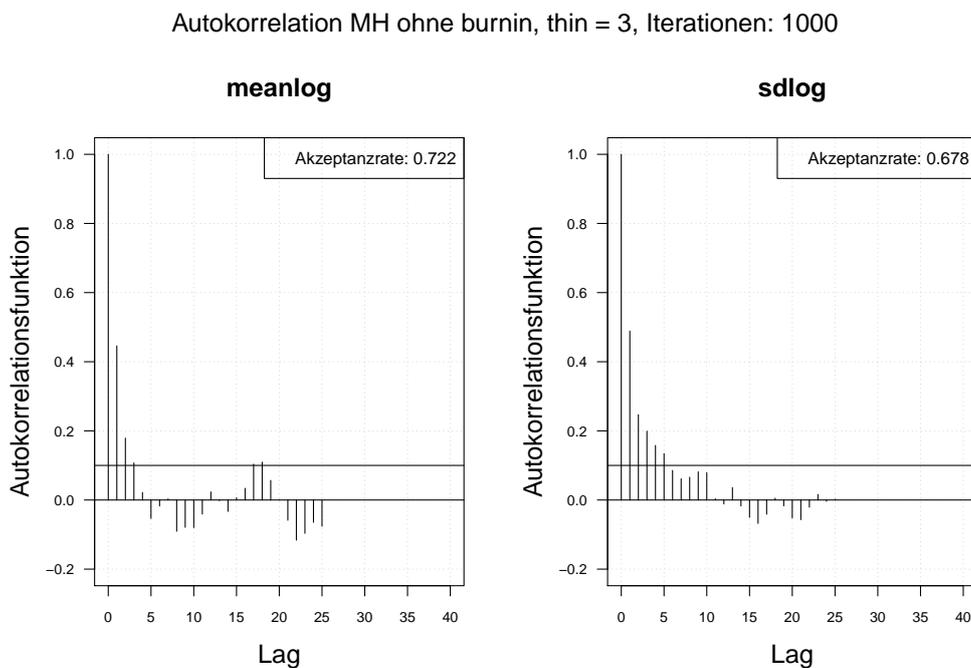


Abbildung 4.4.: Die Autokorrelation der mit MH-Algorithmus simulierten und verdünnten  $meanlog$  und  $sdlog$  der Lognormal-Verteilung. Iterationen: 1000. Verdünnungsfaktor ist gleich drei ( $thin = 3$ ).

Es gibt die Möglichkeit, das Problem der Abhängigkeit der simulierten Werten zu verbessern. Die simulierte Markov-Kette kann hierzu verdünnt werden. D.h. für weitere Analysen werden nicht alle gezogene Werte sondern z. B. jeder zweite, dritte, usw. Wert

betrachtet. Je langsamer die Autokorrelationsfunktion abnimmt, desto größer muss der Verdünnungsfaktor gewählt werden. Das Verhalten der Autokorrelationsfunktion hängt von der Stichprobe ab, was auch nachteilig bei der praktischen Anwendung ist. Somit muss jedes Mal untersucht und analysiert werden, wie die Markov-Kette zu verdünnen ist. In der Abb. 4.4 ist die Autokorrelation der verdünnten Markov-Kette mit MH-Algorithmus simulierten Parameter dargestellt. Hierbei wird jeder dritte Wert ( $thin = 3$ ) betrachtet. Aus der Grafik ist zu erkennen, dass die Autokorrelationsfunktion in diesem Fall viel schneller abfällt als bei der nicht verdünnten Markov-Kette.

In dieser Arbeit wird der MH-Algorithmus für die praktische Umsetzung des reinen Bayes-Ansatzes in der OpRisk-Messung (Kap. 5.1.3) verwendet. Hierbei wird in jedem weiteren Schritt die verdünnte Markov-Kette ( $thin = 3$ ,  $thin = 5$  und  $thin = 7$ ) verwendet.

## 5. Der reine Bayes-Ansatz in der OpRisk-Messung

Der reine Bayes-Ansatz basiert auf der Szenario-Analyse und der Einbeziehung von externen Daten. Szenarien sind die hypothetischen Realisationen der spezifischen Risiken in einer Bank unter bestimmten Umständen. Szenarien werden von Experten festgelegt. Die Experten äußern die Meinungen über potentielle Verluste und dazu korrespondierende Wahrscheinlichkeiten im jeweiligen Segment, z. B. die Meinungen über Verteilungsfamilie, Verteilungsparameter, Anzahl der Verluste im bestimmten Höhenbereich, Quantile der Verlustverteilung und gesamte Häufigkeit, etc. Der reine Bayes-Ansatz kann verwendet werden, um die Parameter der Häufigkeits- und Höhenverteilung der Verluste zu schätzen. In diesem Fall können die Expertenmeinungen für die Schätzung der Parameterverteilung vor der Beobachtung der Daten (*a priori*) genutzt werden. Diese werden mit der *Likelihood* der beobachteten Daten kombiniert, daraus die Parameterverteilung nach der Beobachtung der Daten (*a posteriori*) geschätzt und anschließend die Vorhersageverteilung bestimmt.

Der reine Bayes-Ansatz ermöglicht es, auf drei unterschiedliche Weisen die Daten zu kombinieren: die Kombination der internen mit den externen Verlustdaten, die Kombination der internen Verlustdaten mit den Szenarien und die Kombination aller drei Quellen. Dabei ist es möglich, sowohl die Schadenhöhen- als auch Schadenhäufigkeitsverteilung auf Basis des reinen Bayes-Ansatzes zu schätzen. Da das Ziel der OpRisk-Messung ist, die Vorhersage für die interne Verlustverteilung zu modellieren, kann es fachlich sinnvoller sein, die Häufigkeitsverteilung nur auf Basis der internen Daten zu schätzen.

Das mathematische Konstrukt zur Kombination der internen mit den externen Verlustdaten und der internen Verlustdaten mit den Szenarien ist identisch. Dabei wird die *Likelihood* der internen Verluste mit der *a priori* kombiniert, die auf Basis der externen Daten oder Szenarien geschätzt werden kann. Hierbei kann die Parameterverteilung von Experten festgelegt werden (z. B. die konjugierte<sup>1</sup> Verteilung) und die Hyperparameter<sup>2</sup> können aus externen Daten bzw. Szenarien ermittelt werden.

Sobald die externen Verlustdaten zur Analyse herangezogen werden, kann der reine Bayes-Ansatz im Zwei-Schritt-Verfahren realisiert werden (Tabelle 5.1). Dieses Verfahren wurde vom [Hassani and Renaudin 2013](#): S. 7-8 beschrieben und ist in der Literatur als *Two-Step Bayes Approach* oder *Cascade Bayesian Approach* bekannt. Im ersten Schritt wird die *a posteriori* aus externen Verlustdaten ermittelt: *Likelihood* der externen Daten mit *a priori* kombiniert. Diese *a posteriori* aus Schritt 1 wird als *a priori* in Schritt 2 verwendet.

---

<sup>1</sup>Siehe Def. 1.

<sup>2</sup>Parameter der *a priori* Verteilung.

## Kombination von mehreren Datenquellen in zwei Schritten

**Schritt 1:**  $p_1(\underline{\theta}|loss_{ext}) \propto f(loss_{ext}|\underline{\theta})p_0(\underline{\theta})$

**Schritt 2:**  $p_2(\underline{\theta}|loss_{int}) \propto f(loss_{int}|\underline{\theta})p_1(\underline{\theta}|loss_{ext}) = f(loss_{int}|\underline{\theta})p_1(\underline{\theta})$

Tabelle 5.1.: Zwei-Schritt-Bayes-Verfahren.

Wenn keine Vorinformationen über die Verteilung der Hyperparameter der externen Daten vorliegen, kann die Jeffreys *a priori* (uninformative *a priori*)  $p_0(\underline{\theta}) = p_J(\underline{\theta})$  verwendet werden. Diese Art von *a priori* Information wurde von [Jeffreys 1946](#): S. 453–461 vorgeschlagen. Jeffreys *a priori* ist proportional zu der Wurzel aus der Determinante der Fisher-Information<sup>3</sup>:

$$p_J(\underline{\theta}) = \sqrt{\det(I(\underline{\theta}))} = \sqrt{\det\left(-E\left(\frac{\partial^2}{\partial^2\underline{\theta}}\log(f(\underline{x}|\underline{\theta}))\right)\right)}, \quad (5.1)$$

mit  $\det(I(\underline{\theta})) = \det\left(-E\left(\frac{\partial^2}{\partial^2\underline{\theta}}\log(f(\underline{x}|\underline{\theta}))\right)\right)$  - Determinante aus Fisher-Information (Vgl. [Schmid and Happ 2017](#): S. 16-17).

An diesem Punkt können alternativ die Szenarien  $p_0(\underline{\theta}) = p_{sz}(\underline{\theta})$  herangezogen werden. Somit kann das Zwei-Schritt-Bayes-Verfahren zur Kombination von zwei (interne und externe Verlustdaten) als auch von drei (interne und externe Verlustdaten sowie Szenarien) Datenquellen verwendet werden.

## 5.1. Kombination von zwei Datenquellen

In diesem Abschnitt wird aufgezeigt, wie zwei Datenquellen kombiniert werden können. Als Beispiel wird hier die Kombination der internen und externen Verlustdaten dargestellt. Ersetzt man die externen Daten durch Szenarien, bekommt man den Ansatz zur Kombination der internen Verlustdaten mit den Szenarien.

### 5.1.1. Schätzung der Schadenhäufigkeitsverteilung

Das Ziel der OpRisk-Messung ist es, die Vorhersage für interne Verlustverteilung zu modellieren. Deswegen kann es fachlich viel sinnvoller sein, die Häufigkeitsverteilung nur auf Basis der internen Daten zu schätzen. Das Eintreten der Schäden hängt von unterschiedlichen Faktoren (z. B. der Größe der Bank, etc.) ab, so dass der Einfluss bzw. Nicht-Einfluss dieser Faktoren dabei berücksichtigt werden muss. Aus diesem Grund muss vorher klargestellt und bestimmt werden, ob die externen Informationen über die Häufigkeitsverteilung mitberücksichtigt werden sollen und damit die Verteilung basierend auf der gesamten Situation in der Bankenbranche geschätzt oder die Verteilung für interne Zwecke vorhergesagt und die interne Situation beurteilt werden soll.

---

<sup>3</sup>Die Fisher-Information beschreibt, wie viel Information über den Parameter in der Stichprobe zu erwarten ist. Benannt nach dem Statistiker Ronald Fisher.

Öfters werden für die Modellierung des Schadenseintritts die Poisson-<sup>4</sup> oder die negative Binomial-Verteilung<sup>5</sup> angenommen. Im Folgendem wird anhand des Beispiels der Poisson-Verteilung präsentiert, wie die internen und externen Daten bzw. interne Verlustdaten und Szenarien kombiniert werden können. Solche Anwendung zeigen unter anderen [Cruz et al. 2015](#): S. 593-597.

### Poisson-, Gamma-Verteilung

Zu betrachten sind die Anzahl der eingetretenen Schäden  $N = (N_1, \dots, N_T)$  über die letzten  $T$  Jahre. Die Anzahl der eingetretenen Schäden  $N_{T+1}$  für das nächste Jahr  $T + 1$  muss modelliert werden. Es wird angenommen, dass die Anzahl der eingetretenen Schäden einer Poisson-Verteilung  $N \sim Pois(\lambda)$ ,  $\lambda > 0$  folgt mit der Dichte:

$$f(n|\lambda) = P(N_t = n|\Theta_\lambda = \lambda) = \frac{\lambda^n}{n!}e^{-\lambda}, \quad n \in \mathbb{N}^0. \quad (5.2)$$

Der einzige Parameter  $\theta = \lambda$  ist unbekannt und nach Bayes-Ansatz ist die Zufallsvariable  $\Theta_\lambda$  zu modellieren. Bei der Modellierung müssen folgende Annahmen getroffen werden:

- Die Daten  $N$  und  $N_{T+1}$  sind bedingt unabhängig (gegeben Parameter  $\Theta$ ).
- Die Daten  $N_1, \dots, N_T, N_{T+1}$  gegeben Parameter  $\Theta_\lambda = \lambda$  sind *iid*.
- Die *a priori* Verteilung von  $\Theta_\lambda$  ist die Gamma-Verteilung<sup>6</sup>  $\Theta_\lambda \sim G(\alpha, \beta)$ , mit  $\alpha > 0$  - der Formparameter,  $\beta > 0$  - der inverse<sup>7</sup> Skalenparameter.

Die *a priori* Dichte ist folgendermaßen definiert:

$$p(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda), \quad \lambda > 0 \quad (5.3)$$

mit  $\Gamma(\cdot)$  - Gammafunktion. Die *a posteriori* (Formel 3.3) ist für die Poisson-Verteilung (*Likelihood*) und die konjugierten Gamma-Verteilung (*a priori*) folgendermaßen definiert:

$$\begin{aligned} p(\lambda|\underline{n}) &\propto \frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}\exp(-\beta\lambda) \prod_{t=1}^T \frac{\lambda^{n_t}}{n_t!}\exp(-\lambda) \\ &\stackrel{\text{bzgl. } \lambda}{\propto} \lambda^{\alpha-1}\exp(-\beta\lambda) \prod_{t=1}^T \lambda^{n_t}\exp(-\lambda) \\ &= \lambda^{\alpha-1}\exp(-\beta\lambda)\lambda^{\sum_{t=1}^T n_t}\exp(-T\lambda) \end{aligned}$$

<sup>4</sup>Informationen zur Poisson-Verteilung siehe Anh. A.5.5.

<sup>5</sup>Informationen zur negativen Binomial-Verteilung siehe Anh. A.5.7.

<sup>6</sup>Die Gamma-Verteilung ist die zur Poisson-Verteilung konjugierte *a priori* Verteilung. Informationen zu konjugierten *a priori* Verteilungen siehe Anh. A.4. Informationen zur Gamma-Verteilung siehe Anh. A.5.6.

<sup>7</sup>Der inverse Skalenparameter ist definiert als  $\frac{1}{\beta}$ .

$$= \lambda^{\alpha + \sum_{t=1}^T n_t - 1} \exp(-\lambda(\beta + T)), \quad (5.4)$$

mit  $\Gamma(\cdot)$  - Gammafunktion. Die *a posteriori* ist proportional zu dem Kern einer Gamma-Verteilung (Formel 5.4):  $\Theta_{\lambda, a \text{ posteriori}} \sim G(\tilde{\alpha}, \tilde{\beta})$ , wobei  $\tilde{\alpha} = \alpha + \sum_{t=1}^T n_t$  und  $\tilde{\beta} = \beta + T$  die *a posteriori* Hyperparameter sind ( $\tilde{\alpha}$  - der Formparameter,  $\tilde{\beta}$  - der inverse Skalenparameter<sup>8</sup>).

Die bedingte Dichte der zukünftigen Beobachtung (Formel 3.4) ist demzufolge die Dichte der negativen Binomial-Verteilung<sup>9</sup>:

$$\begin{aligned} f(m|\underline{n}) &= P(N_{T+1} = m | \underline{N} = \underline{n}) = \int f(m|\lambda) p(\lambda|\underline{n}) d\lambda \\ &= \int \frac{\lambda^m}{m!} \exp(-\lambda) \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \lambda^{\tilde{\alpha}-1} \exp(-\tilde{\beta}\lambda) d\lambda \\ &= \binom{m + \tilde{\alpha} - 1}{m} \left( \frac{\tilde{\beta}}{1 + \tilde{\beta}} \right)^{\tilde{\alpha}} \left( 1 - \frac{\tilde{\beta}}{1 + \tilde{\beta}} \right)^m \\ &= \binom{m + \tilde{\alpha}_{neu} - 1}{m} (\tilde{\beta}_{neu})^{\tilde{\alpha}_{neu}} (1 - \tilde{\beta}_{neu})^m, \end{aligned} \quad (5.5)$$

mit  $\tilde{\alpha}_{neu} = \tilde{\alpha}$  - der Formparameter und  $\tilde{\beta}_{neu} = \frac{\tilde{\beta}}{1 + \tilde{\beta}}$  der Skalenparameter der negativen Binomial-Verteilung.

Die erwartete Anzahl der Verluste im folgenden Jahr ergibt sich als gewichtete Summe der MLE für Parameter  $\lambda$  aus den internen Daten und dem Schätzer aus der *a priori* Verteilung<sup>10</sup>:

$$E(N_{T+1}) = \omega \lambda_{MLE} + (1 - \omega) \lambda_{a \text{ priori}}, \quad (5.6)$$

mit  $\lambda_{MLE} = \frac{1}{T} \sum_{t=1}^T n_{int,t}$  - der MLE für Parameter  $\lambda$  aus den beobachteten internen Daten,  $\lambda_{a \text{ priori}} = \frac{\alpha}{\beta} = E(X)$ ,  $X \sim G(\alpha, \beta)$  - die Schätzung für Parameter  $\lambda$  aus der *a priori* auf Basis von externen Daten bzw. Szenarien,  $\omega = \frac{T}{\beta + T}$ ,  $\omega \in [0, 1]$  - das Gewicht zur Kombination der beiden Schätzer,  $\alpha$  und  $\beta$  - die Hyperparameter der *a priori* Gamma-Verteilung,  $T$  - Anzahl der Jahre (Zeitperioden) in historisch beobachteten Daten.

Die Hyperparameter  $\alpha$  und  $\beta$  können von Experten festgelegt (Szenarien) oder aus externen Daten geschätzt werden. [Shevchenko and Wüthrich 2006](#): S. 12-14 beschreiben, wie die Hyperparameter  $\alpha$  und  $\beta$  auf Basis von Expertenmeinungen geschätzt werden können. Die beste Schätzung für die erwartete Anzahl der Verluste ist gegeben als  $E(E(N|\lambda))$ . Die erwartete Anzahl der Verluste für die Poisson-verteilte Zufallsvariable

<sup>8</sup>Der *a posteriori* inverse Skalenparameter ist definiert als  $\frac{1}{\tilde{\beta}}$ . Der *a posteriori* nicht inverse Skalenparameter ist damit  $\frac{1}{\tilde{\beta} + T} = \frac{1}{\frac{1+T\tilde{\beta}}{\tilde{\beta}}} = \frac{\tilde{\beta}}{1+T\tilde{\beta}}$ , mit  $\tilde{\beta}$  - der nicht inverse Skalenparameter.

<sup>9</sup>Zur Herleitung der Formel siehe Anh. A.7 Formel A.3

<sup>10</sup>Zur Herleitung der Formel siehe Anh. A.7 Formel A.4.

ist  $\lambda$  und der Erwartungswert der Gamma-verteilen Zufallsvariable ist  $\frac{\alpha}{\beta}$ <sup>11</sup>. Folglich gilt  $E(E(N|\lambda)) = E(\lambda) = \frac{\alpha}{\beta}$ . Der Experte kann  $E(\lambda)$  (z. B.  $E(\lambda) = 0.5$ ) und die Unsicherheit über den wahren Parameter  $\lambda$  spezifizieren, indem die Wahrscheinlichkeit  $p$  (z. B.  $p = \frac{2}{3}$ ) angegeben wird, mit der der wahre Wert in einem Intervall zwischen 2 Werten  $a$  und  $b$  (z. B.  $a = 0.25, b = 0.75$ ) liegt:

$$P(a \leq \lambda \leq b) = p = \int_a^b p(\lambda|\alpha, \beta) d\lambda = F_{G(\alpha, \beta)}(b) - F_{G(\alpha, \beta)}(a), \quad (5.7)$$

wobei

$$F_{G(\alpha, \beta)}(y) = \int_0^y \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) dx \quad (5.8)$$

die Verteilungsfunktion der Gamma-Verteilung ist. Die Formel 5.7 kann unter Verwendung von Formel 5.8 und der Nebenbedingung  $E(\lambda) = \frac{\alpha}{\beta}$  numerisch gelöst (optimiert) und die Parameter  $\alpha$  und  $\beta$  können folglich bestimmt werden. Für die numerische Lösung des nicht linearen Gleichungssystems kann das R-Paket “*nleqslv*” (Hasselmann 2018: S. 1-19) mit der implementierten R-Funktion “*nleqslv()*” verwendet werden. Neben der Optimierung besteht noch eine Möglichkeit, die Parameter zu bestimmen, wenn neben dem  $E(\lambda)$  der Variationskoeffizient  $Vco(\lambda)$  von Experten geschätzt wurde:

$$Vco(\lambda) = \frac{\sqrt{Var(\lambda)}}{E(\lambda)}. \quad (5.9)$$

Unter den Nebenbedingungen

$$E(\lambda) = \frac{\alpha}{\beta} \text{ und } Vco(\lambda) = \frac{1}{\sqrt{\alpha}} \quad (5.10)$$

lassen sich die Parameter leicht berechnen.

Alternativ können die Hyperparameter aus externen Daten geschätzt werden. Die Parameterverteilung kann wiederum anhand des Bayes-Ansatzes ermittelt werden. Da wir über keine andere *a priori* Informationen verfügen, kann hierbei Jeffreys *a priori* verwendet werden. Unter Verwendung der Formel 5.1 kann Jeffreys *a priori* für die Poisson-Verteilung berechnet werden<sup>12</sup>:

$$p_J(\lambda) \propto \frac{1}{\sqrt{\lambda}}. \quad (5.11)$$

Die *a posteriori* Verteilung aus externen Daten (**Schritt 1**, Tabelle 5.1) unter Verwendung der Formel 5.11 ist die Gamma-Verteilung mit neuen Parametern<sup>13</sup>:  $\Theta_{\lambda, ext} \sim G(\alpha_{ext}, \beta_{ext})$ , wobei  $\alpha_{ext} = \frac{1}{2} + \sum_{t=1}^{T_{ext}} n_{ext,t} = \frac{1}{2} + n_{ext}$  und  $\beta_{ext} = T_{ext}$  - die *a posteriori* Hyperparameter der Gamma-Verteilung sind. Diese Verteilung bzw. die Hyperparameter können als *a priori* Information bei der Schätzung aus internen Daten genutzt werden. Somit ist die *a posteriori* Verteilung aus internen Verlustdaten (**Schritt 2**, Tabelle

<sup>11</sup> $\beta$  ist der inverse Skalenparameter.

<sup>12</sup>Zur Herleitung der Formel siehe Anh. A.6 Formel A.1.

<sup>13</sup>Die Herleitung siehe Anh. A.6 Formel A.2.

5.1) unter Verwendung von Formel 5.4 die Gamma-Verteilung mit neuen Hyperparametern:  $\Theta_{\lambda,int} \sim G(\alpha_{int}, \beta_{int})$ , wobei  $\alpha_{int} = \alpha_{ext} + \sum_{t=1}^{T_{int}} n_{int,t} = \frac{1}{2} + n_{ext} + n_{int}$  und  $\beta_{int} = \beta_{ext} + T_{int} = T_{ext} + T_{int}$ <sup>14</sup> - die *a posteriori* Hyperparameter der internen Daten sind.

## Vergleich der Schätzungen

Um zu entscheiden, wie der Parameter  $\lambda$  der Poisson-Verteilung zu schätzen ist, wurden die Analysen auf simulierten Daten durchgeführt, die die Realität und Problemstellen der OpRisk-Messung möglichst genau abbilden. Die wahren Parameter sind in der Tabelle 5.2 dargestellt.

Die wahren Parameter für die Simulation der Daten	
Poisson-Verteilung ( <i>lambda</i> )	
$\lambda_{int}$	0.86
$\lambda_{ext}$	6
Lognormal-Verteilung ( <i>meanlog, sdlog</i> )	
$\mu_{int}$	16.5
$\sigma_{int}$	1.41
$\mu_{ext}$	16.2
$\sigma_{ext}$	0.96

Tabelle 5.2.: Die wahren Parameter für die Simulation der Daten.

Die Häufigkeiten wurden jeweils aus der Poisson-Verteilung gezogen. Zu den existierenden Häufigkeiten wurden die Verluste aus der Lognormal-Verteilung simuliert<sup>15</sup>. Neben dem MLE aus internen Daten wurden Schätzer nach Bayes-Ansatz aus internen (Einschritt-Verfahren<sup>16</sup>), aus externen (Zwei-Schritt-Verfahren) Daten und aus Expertenmeinungen (Zwei-Schritt-Verfahren) ermittelt. Hierbei wurden Parameter jeweils aus der *a posteriori* Verteilung und aus der Verteilung für die zukünftige Beobachtung (zukünftig) ermittelt. Die Ergebnisse sind in der Tabelle 5.3 dargestellt.

Der MLE aus den internen Daten ( $\hat{\lambda}_{int}$  MLE: *a posteriori* = 0.66, zukünftig = 0.66) und der Schätzer nach Bayes-Ansatz nur aus internen Daten ( $\hat{\lambda}_{int}$  Bayes-Ansatz: *a posteriori* = 0.674, zukünftig = 0.66) sind identisch und nähern sich eher dem wahren Parameter der internen ( $\lambda_{int}$  wahr = 0.86) als der externen Daten ( $\lambda_{ext}$  wahr = 6). Die Schätzer nach Bayes-Ansatz aus den externen Daten unterscheiden sich grundlegend ( $\hat{\lambda}_{ext}$  Bayes-Ansatz: *a posteriori* = 6.199, zukünftig = 3.165). Beide Schätzer sind deutlich größer als der wahre Wert der internen Daten. Der Schätzer aus der *a posteriori* Verteilung nähert sich dem wahren Parameter der externen Daten und ist größer als der Schätzer aus der Verteilung für die zukünftige Beobachtung. Es ist durch die Anzahl der externen Schadensfälle erklärbar, die im ersten Fall in den Schätzer direkt einfließen und im

<sup>14</sup> $\beta_{int}$  ist der inverse Skalenparameter. Der nicht inverse Skalenparameter ist somit  $\frac{1}{T_{ext}+T_{int}}$ .

<sup>15</sup>Um die Simulation nachvollziehen zu können, wurde *seed* = 5728 gewählt.

<sup>16</sup>Hierbei werden die *Likelihood* der internen Daten und die Jeffrey's *a priori* benutzt.

**Vergleich der Schätzungen:** Parameter  $\lambda$  der Poisson-Verteilung

$\lambda_{int}$ wahr	0.86
$\lambda_{ext}$ wahr	6
$\hat{\lambda}_{int}$ MLE <i>a posteriori</i>	0.66
$\hat{\lambda}_{int}$ MLE zukünftig	0.66
$\hat{\lambda}_{int}$ Bayes-Ansatz <i>a posteriori</i>	0.674
$\hat{\lambda}_{int}$ Bayes-Ansatz zukünftig	0.66
$\hat{\lambda}_{ext}$ Bayes-Ansatz <i>a posteriori</i>	6.199
$\hat{\lambda}_{ext}$ Bayes-Ansatz zukünftig	3.165
$\hat{\lambda}_{Experten}$ erwartet 1	0.86
Wahrscheinlichkeit für erwartetes $\hat{\lambda}_{Experten}$ 1	1
untere Grenze für erwartetes $\hat{\lambda}_{Experten}$ 1	0.795
obere Grenze für erwartetes $\hat{\lambda}_{Experten}$ 1	0.925
$\hat{\lambda}_{Experten}$ Bayes-Ansatz <i>a posteriori</i> 1	1.314
$\hat{\lambda}_{Experten}$ Bayes-Ansatz zukünftig 1	0.797
$\hat{\lambda}_{Experten}$ erwartet 2	1.2
Wahrscheinlichkeit für erwartetes $\hat{\lambda}_{Experten}$ 2	0.65
untere Grenze für erwartetes $\hat{\lambda}_{Experten}$ 2	0.655
obere Grenze für erwartetes $\hat{\lambda}_{Experten}$ 2	1.745
$\hat{\lambda}_{Experten}$ Bayes-Ansatz <i>a posteriori</i> 2	9.12
$\hat{\lambda}_{Experten}$ Bayes-Ansatz zukünftig 2	0.692

Tabelle 5.3.: Poisson-, Gamma-Verteilung. Vergleich der Schätzer.

zweiten Fall wegen der Gewichtung einen kleineren Einfluss auf die Schätzung haben. Um die externen Daten für die Bestimmung der Schadenhäufigkeitsverteilung sinnvoll heranziehen zu können, muss hierbei z. B. die mittlere Anzahl der Schadensfälle in einer Bank ermittelt oder der geschätzte Parameter entsprechend skaliert werden. Expertenmeinungen verbessern die Schätzung ( $\hat{\lambda}_{Experten}$  Bayes-Ansatz: *a posteriori* 1 = 1.314 zukünftig 1 = 0.797), sobald die Genauigkeit der erwarteten Parameter hoch ist: der erwartete Parameter ( $\hat{\lambda}_{Experten}$  erwartet 1 = 0.86) liegt mit hoher Wahrscheinlichkeit (Wahrscheinlichkeit für erwartetes  $\hat{\lambda}_{Experten}$  1 = 1) im sehr engen Bereich ([untere Grenze 1, obere Grenze 1] = [0.795, 0.925]). Je schlechter die Vorhersage der Experten ist, desto schlechter wird der Schätzer.

Das Ziel der OpRisk-Messung ist, vor allem die interne Situation abzuschätzen und die internen Verluste vorherzusagen. Aus diesem Grund ist es sinnvoller, die Schadenhäufigkeitsverteilung aus internen Daten zu schätzen. Weiterhin wird in dieser Arbeit der MLE aus internen Daten für die Schadenhäufigkeitsverteilung benutzt.

### 5.1.2. Schätzung der Schadenhöhenverteilung unter Verwendung der analytischen Herleitung

Wie schon in Kap. 1 erwähnt wurde, werden rechtsschiefe *Heavy-tailed*-Verteilungen für die Schätzung der Schadenhöhen im *tail*-Bereich verwendet. Solche sind beispielsweise unter anderen die Lognormal-, Loggamma- und Weibull-Verteilung. Die Datenverteilung kann auf Basis des statistischen Anpassungstests<sup>17</sup> ausgewählt werden. Im Folgenden wird am Beispiel der Lognormal-Verteilung dargestellt, wie die internen Verlustdaten mit den externen Daten bzw. mit den Szenarien kombiniert werden können.

#### Lognormal-, Normal-, skalierte inverse Chi-Quadrat-Verteilung

Die Modellierung der Verlustverteilung mit der Lognormal-Verteilung kann unterschiedlich ausfallen. Dies hängt von den Modellannahmen ab, die getroffen werden: ob die beiden Parameter  $\mu$  und  $\sigma^2$  unbekannt sind oder einer der beiden Parametern bekannt ist und als eine Konstante betrachtet wird. Ebenso spielt die Unabhängigkeit bzw. Abhängigkeit der beiden Parameter eine große Rolle.

Für die Ausgangssituation können die folgenden Annahmen über die Parameter  $\mu$  und  $\sigma^2$  der Lognormal-Verteilung getroffen werden:

- Die Daten  $Loss$  und  $Loss_{n+1}$  sind bedingt unabhängig (gegeben Parameter  $\Theta$ ).
- Die Daten  $Loss_1, \dots, Loss_n, Loss_{n+1}$  gegeben Parameter  $\Theta_\mu = \mu$  und  $\Theta_{\sigma^2} = \sigma^2$  sind *iid*.
- Die bedingte *a priori* Verteilung von  $\Theta_{\mu|\sigma^2}$  ist die Normal-Verteilung<sup>18</sup>  $\Theta_{\mu|\sigma^2} \sim N(\mu_0, \frac{\sigma^2}{\phi})$ , mit  $\mu_0$  - der Erwartungswert,  $\frac{\sigma^2}{\phi}$  - die Varianz.
- Die marginale *a priori* Verteilung von  $\Theta_{\sigma^2}$  ist die skalierte inverse Chi-Quadrat-Verteilung<sup>19</sup>  $\Theta_{\sigma^2} \sim scaledI\chi^2(\nu, \tau^2)$ , mit  $\nu > 0$  - die Freiheitsgrade,  $\tau^2 > 0$  - der Skalenparameter.
- Die Parameter sind abhängig und es gilt für die *a priori* :  $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$ .

Shevchenko and Wüthrich 2006: S. 24-25 beschreiben die Modellierung der Verlustverteilung mithilfe der Lognormal-Verteilung mit zwei unbekanntem abhängigen Parametern. Die bedingte und marginale<sup>20</sup> *a priori* Dichten der Parameter sind entsprechend:

$$p(\mu|\sigma^2) = p(\mu|\mu_0, \frac{\sigma^2}{\phi}) = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{\phi}}} \exp\left(-\frac{[\mu - \mu_0]^2}{2\frac{\sigma^2}{\phi}}\right), \quad \mu \in \mathbb{R}, \quad (5.12)$$

<sup>17</sup>Z. B. Anderson-Darling-, Kolmogorov-Smirnov-Test, etc.

<sup>18</sup>Für Informationen zur Normal-Verteilung siehe Anh. A.5.2.

<sup>19</sup>Für Informationen zur skalierten inversen Chi-Quadrat-Verteilung siehe Anh. A.5.4.

<sup>20</sup>Die Formel 5.13 ist die umgeschriebene Formel der Dichte der skalierten inversen Chi-Quadrat-Verteilung A.5.4.

$$p(\sigma^2) = p(\sigma^2|\nu, \tau^2) = \frac{2^{-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})\nu\tau^2} \left(\frac{\sigma^2}{\nu\tau^2}\right)^{-(1+\frac{\nu}{2})} \exp\left(-\frac{\nu\tau^2}{2\sigma^2}\right), \quad \sigma^2 > 0. \quad (5.13)$$

Die gemeinsame *a priori* laut Bayes-Regel  $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$  ist die Mischverteilung aus der Normal- und der skalierten inversen Chi-Quadrat-Verteilung mit dem Kern<sup>21</sup>:

$$p(\mu, \sigma^2) \propto (\sigma^2)^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}(\phi[\mu - \mu_0]^2 + \nu\tau^2)\right). \quad (5.14)$$

Diese ist die zur Lognormal-Verteilung konjugierte *a priori* Verteilung. Die *a posteriori* (Formel 3.3) ist für diesen Fall folgendermaßen definiert<sup>22</sup>:

$$p(\mu, \sigma^2|\underline{Loss}) \propto (\sigma^2)^{-\frac{\nu_{neu}+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}(\phi_{neu}[\mu - \mu_{0,neu}]^2 + (\nu\tau^2)_{neu})\right), \quad (5.15)$$

mit neuen *a posteriori* Parametern der Mischverteilung:

$$\nu_{neu} = \nu + n,$$

$$(\nu\tau^2)_{neu} = \nu_{neu}\tau_{neu}^2 = \nu\tau^2 + \phi\mu_0^2 + n\overline{loss^2} - \frac{(n\overline{loss} + \phi\mu_0)^2}{\phi + n},$$

$$\tau_{neu}^2 = \frac{(\nu\tau^2)_{neu}}{\nu_{neu}},$$

$$\mu_{0,neu} = \frac{n\overline{loss} + \phi\mu_0}{\phi + n},$$

$$\phi_{neu} = \phi + n,$$

$$\overline{loss} = \frac{1}{n} \sum_{i=1}^n \log(loss_i),$$

$$\overline{loss^2} = \frac{1}{n} \sum_{i=1}^n \log^2(loss_i).$$

Die Hyperparameter  $\mu_0, \phi, \nu, \tau^2$  können auf Basis von Expertenmeinungen (Szenarien) oder aus externen Daten ermittelt werden. Cruz et al. 2015: S. 599-601 beschreiben die Schätzung der Hyperparameter auf Basis von Expertenmeinungen. Dieses Vorgehen ist der in Kap. 5.1.1 beschriebenen Möglichkeit ähnlich.

Das  $q$ -Quantil der Lognormal-verteiltern Zufallsvariable ist folgendermaßen definiert:

$$Q_{Loss \sim LN(\mu, \sigma^2)}(q) = \exp(\Theta_\mu + \Theta_\sigma Z_q), \quad (5.16)$$

mit  $Z_q = \Phi^{-1}(q)$  - das  $q$ -Quantil der Standard-Normal-Verteilung. Unter Verwendung von Formel 5.16 kann die Zufallsvariable  $\Theta_\sigma$  durch das Verhältnis von zwei Quantilen dargestellt werden:

$$\Theta_\sigma = \frac{\log(Q_{Loss}(q_2)/Q_{Loss}(q_1))}{Z_{q_2} - Z_{q_1}}. \quad (5.17)$$

Unter Verwendung von Expertenmeinungen über den  $E(\log(Q_{Loss}(q_2)/Q_{Loss}(q_1)))$  und  $P(a \leq (Q_{Loss}(q_2)/Q_{Loss}(q_1)) \leq b)$  kann die *a priori* für  $\Theta_\sigma$  geschätzt werden<sup>23</sup>. Wenn die

<sup>21</sup>Die Herleitung der Formel siehe Anh. A.8 Formel A.5.

<sup>22</sup>Die Herleitung der Formel siehe Anh. A.8 Formel A.6

<sup>23</sup>Die Hyperparameter  $\nu$  und  $\tau^2$  können hierbei numerisch bestimmt werden.

ermittelte Varianz  $\sigma^2$  als bekannte Konstante betrachtet wird, kann die *a priori* für  $\Theta_\mu$  auf Basis von Expertenmeinungen wie folgt geschätzt werden. Es wird die Lognormal-Verteilung für die Verluste  $Loss \sim LN(\mu_0 + \frac{\sigma^2}{2}, \sigma_0^2)$  und die Normal-Verteilung für die unbekannt Parameter  $\Theta_\mu \sim N(\mu_0, \sigma_0^2)$  angenommen. Der erwartete Verlust  $E(Loss) = exp(\mu_0 + \frac{\sigma^2}{2} + \frac{\sigma_0^2}{2})$  und die Wahrscheinlichkeit, dass der wahre Verlust in einem bestimmten Intervall liegt (Formel 5.18) werden von Experten angegeben.

$$P(a \leq Loss \leq b) = p = \Phi\left(\frac{\log(b) - \frac{\sigma^2}{2} - \mu_0}{\sigma_0^2}\right) - \Phi\left(\frac{\log(a) - \frac{\sigma^2}{2} - \mu_0}{\sigma_0^2}\right), \quad (5.18)$$

mit  $\Phi(\cdot)$  - die Verteilungsfunktion der Standardnormal-Verteilung.

Die Hyperparameter können alternativ aus externen Daten geschätzt werden. Genauso wie bei der in Kap. 5.1.1 beschriebener Vorgehensweise wird die *a posteriori* Verteilung für die Parameter aus externen Daten ermittelt. Hierbei wird die Jeffreys<sup>24</sup> *a priori* für die Lognormal-Verteilung<sup>25</sup> benutzt:

$$p_J(\mu, \sigma^2) \propto (\sigma^2)^{-\frac{3}{2}}. \quad (5.19)$$

Die *a posteriori* aus externen Verlustdaten (**Schritt 1**, Tabelle 5.1) ist die Mischverteilung aus der Normal-  $\Theta_{\mu|\sigma^2, ext} \sim N(\mu_{0, ext}, \frac{\sigma^2}{\phi_{ext}})$  und der skalierten inversen Chi-Quadrat-Verteilung  $\Theta_{\sigma^2, ext} \sim scaledI\chi^2(\nu_{ext}, \tau_{ext}^2)$  mit folgenden Hyperparametern aus externen Verlustdaten<sup>26</sup>:

$$\begin{aligned} \nu_{ext} &= n_{ext}, \\ (\nu\tau^2)_{ext} &= \nu_{ext}\tau_{ext}^2 = n_{ext}(\overline{loss_{ext}^2} - \overline{loss_{ext}}^2), \\ \tau_{ext}^2 &= \overline{loss_{ext}^2} - \overline{loss_{ext}}^2, \\ \mu_{0, ext} &= \overline{loss_{ext}}, \\ \phi_{ext} &= n_{ext}, \\ \overline{loss_{ext}} &= \frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} \log(loss_{ext, i}), \\ \overline{loss_{ext}^2} &= \frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} \log^2(loss_{ext, i}), \\ \overline{loss_{ext}}^2 &= \left(\frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} \log(loss_{ext, i})\right)^2. \end{aligned}$$

Diese Parameterverteilungen können anschließend als *a priori* Verteilung bei der Schätzung aus internen Daten genutzt werden. Die *a posteriori* aus internen Verlustdaten (**Schritt 2**, Tabelle 5.1) ist demzufolge wiederum die Mischverteilung aus der Normal- und der skalierten inversen Chi-Quadrat-Verteilung mit neuen *a posteriori* Parametern:

<sup>24</sup>Informationen zur Jeffreys *a priori* siehe Kap. 5.1.1 Formel 5.1.

<sup>25</sup>Zur Herleitung der Formel siehe Anh. A.9 Formel A.7.

<sup>26</sup>Die Herleitung der *a posteriori* aus externen Daten sowie die Bestimmung der Hyperparameter siehe Anh. A.9 Formel A.15

$$\nu_{int} = n_{ext} + n_{int},$$

$$(\nu\tau^2)_{int} = n_{ext}(\overline{loss_{ext}^2} - \overline{loss_{ext}}^2) + n_{ext}\overline{loss_{ext}}^2 + n_{int}\overline{loss_{int}^2} - \frac{(n_{int}\overline{loss_{int}} + n_{ext}\overline{loss_{ext}})^2}{n_{ext} + n_{int}},$$

$$\mu_{0,int} = \frac{n_{int}\overline{loss_{int}} + n_{ext}\overline{loss_{ext}}}{n_{ext} + n_{int}},$$

$$\phi_{int} = n_{ext} + n_{int},$$

$$\overline{loss_{int}} = \frac{1}{n_{int}} \sum_{i=1}^{n_{int}} \log(loss_{int,i}),$$

$$\overline{loss_{int}^2} = \frac{1}{n_{int}} \sum_{i=1}^{n_{int}} \log^2(loss_{int,i}),$$

$$\overline{loss_{int}}^2 = \left( \frac{1}{n_{int}} \sum_{i=1}^{n_{int}} \log(loss_{int,i}) \right)^2,$$

$$\overline{loss_{ext}} = \frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} \log(loss_{ext,i}),$$

$$\overline{loss_{ext}^2} = \frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} \log^2(loss_{ext,i}),$$

$$\overline{loss_{ext}}^2 = \left( \frac{1}{n_{ext}} \sum_{i=1}^{n_{ext}} \log(loss_{ext,i}) \right)^2.$$

Die geschätzte *a posteriori* Verteilung (bzw. die Hyperparameter) kann im Weiteren zur Schätzung der Vorhersageverteilung benutzt werden.

## Ergebnisse der Schätzung

Für die Analysen der Schätzung der Schadenshöhenverteilung wurden dieselben Verlustdaten benutzt (siehe Tabelle 5.2) und noch fünf zusätzliche Datensätze mit verschiedener Anzahl von internen und externen Verlustdaten erzeugt. Die wahren Parameter aller sechs Datensätze sind in der Tabelle 5.4 dargestellt.

<b>Die wahren Parameter für die Simulation der sechs Datensätze</b>						
	Daten 1	Daten 2	Daten 3	Daten 4	Daten 5	Daten 6
<b>Poisson-Verteilung (<i>lambda</i>)</b>						
$\lambda_{int}$	0.86	2.112	6	2.112	4.06	6
$\lambda_{ext}$	6	25.52	56.3	6	6	6
<b>Lognormal-Verteilung (<i>meanlog</i>, <i>sdlog</i>)</b>						
$\mu_{int}$	16.5	16.5	16.5	16.5	16.5	16.5
$\sigma_{int}$	1.41	1.41	1.41	1.41	1.41	1.41
$\mu_{ext}$	16.2	16.2	16.2	16.2	16.2	16.2
$\sigma_{ext}$	0.96	0.96	0.96	0.96	0.96	0.96

Tabelle 5.4.: Die wahren Parameter für Simulation der sechs Datensätze.

Parameter  $\lambda$  der Poisson-Verteilung reguliert hierbei die Anzahl der internen und externen Verlustdaten. Die Häufigkeiten wurden wie vorher jeweils aus der Poisson-Verteilung gezogen. Zu den existierenden Häufigkeiten wurden die Verluste aus der Lognormal-Verteilung simuliert<sup>27</sup>. Unter der Verwendung der sechs simulierten Datensätze und analytisch hergeleiteten Hyperparametern der Normal- und der skalierten inversen Chi-Quadrat-Verteilung für Schritt 2 wurden jeweils 1000 Parameter  $\mu$  und  $\sigma$  aus der *a posteriori* gezogen. In der Abb. 5.1 und Abb. 5.2 sind die Ziehungen (jeweils oben links), die dazugehörige Histogramme (jeweils unten links) und die Autokorrelationsfunktionen (jeweils unten rechts) dargestellt, die unter Verwendung des ersten Datensatzes ( $\lambda_{int} = 0.86$  und  $\lambda_{ext} = 6$ ) erzeugt wurden.

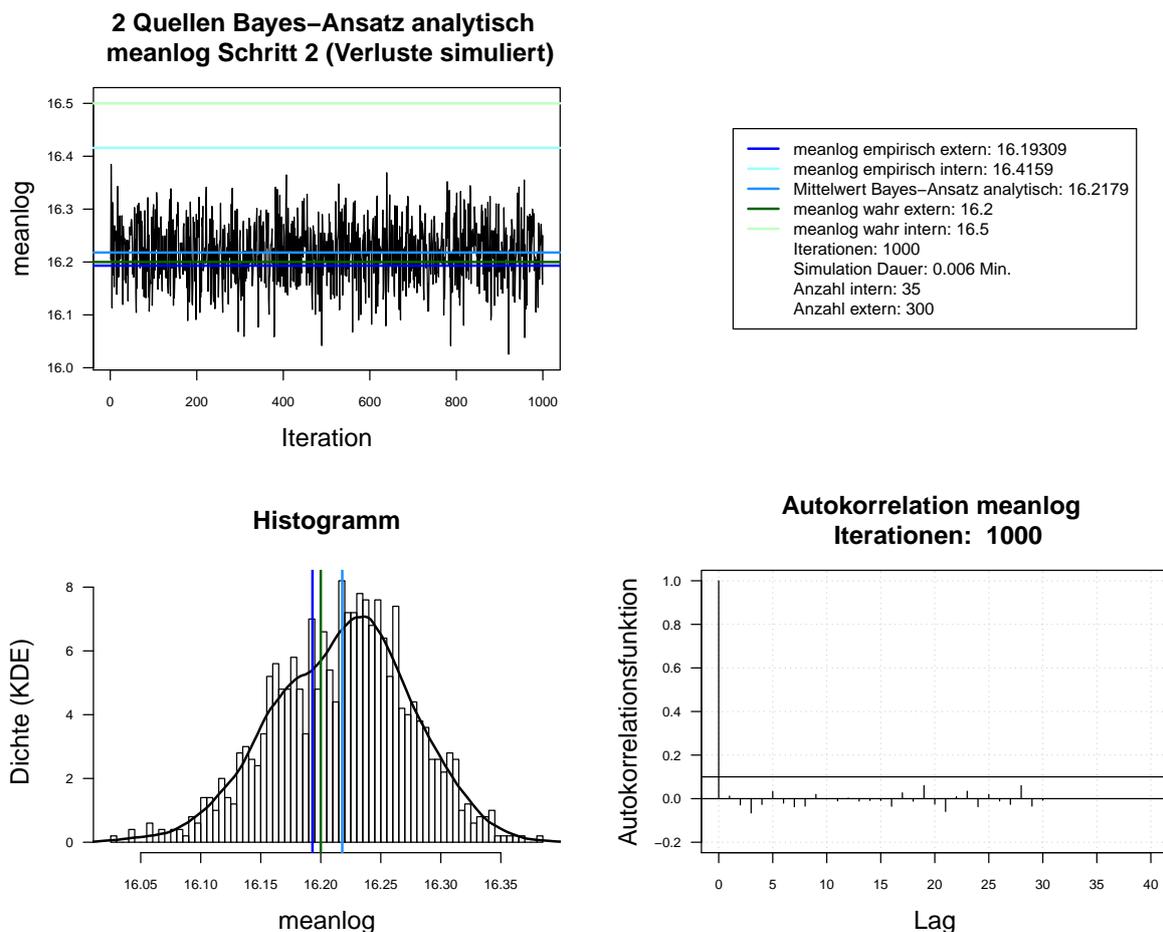


Abbildung 5.1.: Der mit analytischer Herleitung simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. simulierte Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links). Schritt 2. Interne Daten: 35. Externe Daten: 300.

<sup>27</sup>Um die Simulation nachvollziehen zu können, wurde  $seed = 5728$  gewählt.

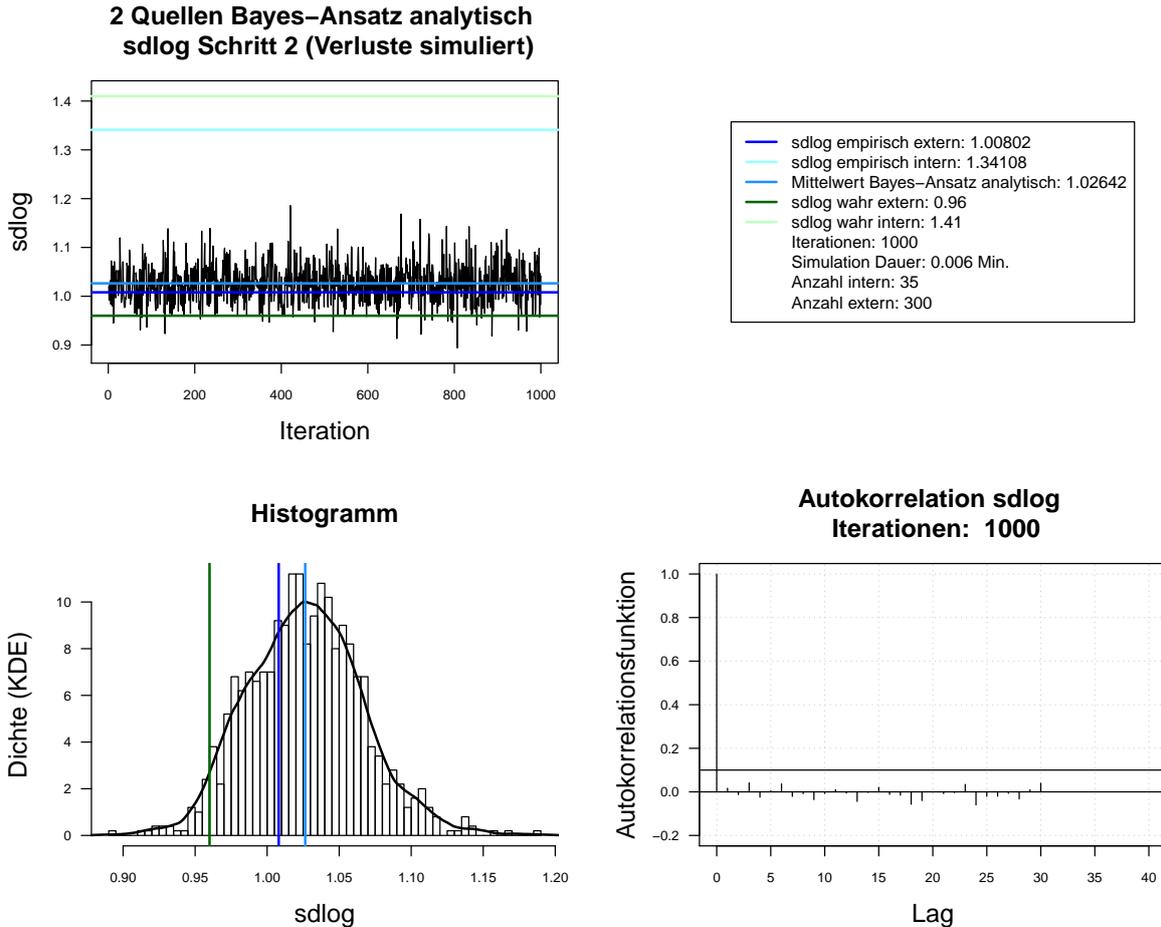


Abbildung 5.2.: Der mit analytischer Herleitung simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. simulierte Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links). Schritt 2. Interne Daten: 35. Externe Daten: 300.

Aus der Grafiken ist zu erkennen, dass die Mittelwerte der beiden *a posteriori* Verteilungen ( $meanlog = 16.2179$  und  $sdlog = 1.02642$ ) ähnlich den Parametern der externen Daten sind (empirisch:  $meanlog = 16.19309$ ,  $sdlog = 1.00802$  und wahr:  $meanlog = 16.2$ ,  $sdlog = 0.96$ ), was durch den Einfluss der wesentlich größeren Anzahl externer Daten erklärbar ist. Außerdem ist zu bemerken, dass keine Autokorrelation für beide Ketten vorliegt, weil hierbei keine Markov-Kette, sondern gleich die *iid* Werte aus der entsprechenden Verteilungen gezogen wurden. Aus diesem Grund müssen die Ziehungen für die weitere Verwendung nicht verdünnt werden.

Das Ergebnis der simulierten Vorhersageverteilung der Schadenshöhen im Vergleich zu den Verteilungen der internen und externen Daten ist im Log-log-Plot (Abb. 5.3) zusammengefasst. Die schwarzen und roten Punkte bilden die internen und externen Verteilungen entsprechend ab. Zu jeder Kombination der aus *a posteriori* Verteilungen gezogenen

Parameter (1000) wurden Verluste für die folgenden 1500 Jahre (Simulation: 1500) simuliert (dunkelgrüne Punkte). Die hellgrünen Punkte stellen die simulierten Verluste mit Bayes-Schätzer (Erwartungswert der *a posteriori* Verteilung<sup>28</sup>) dar. Die Vorhersageverteilung neigt sich bei kleineren Werten (x-Achse: ca. 13 bis 18, y-Achse: ca. -2.5 bis 0) zur externen Verteilung und weicht im Randbereich (x-Achse: ca. ab 18, y-Achse: ca. bis -2.5) in Richtung der internen Daten ab. Die Ergebnisse auf Basis der anderen fünf Datensätze sind im Anh. A.10 aufgeführt.

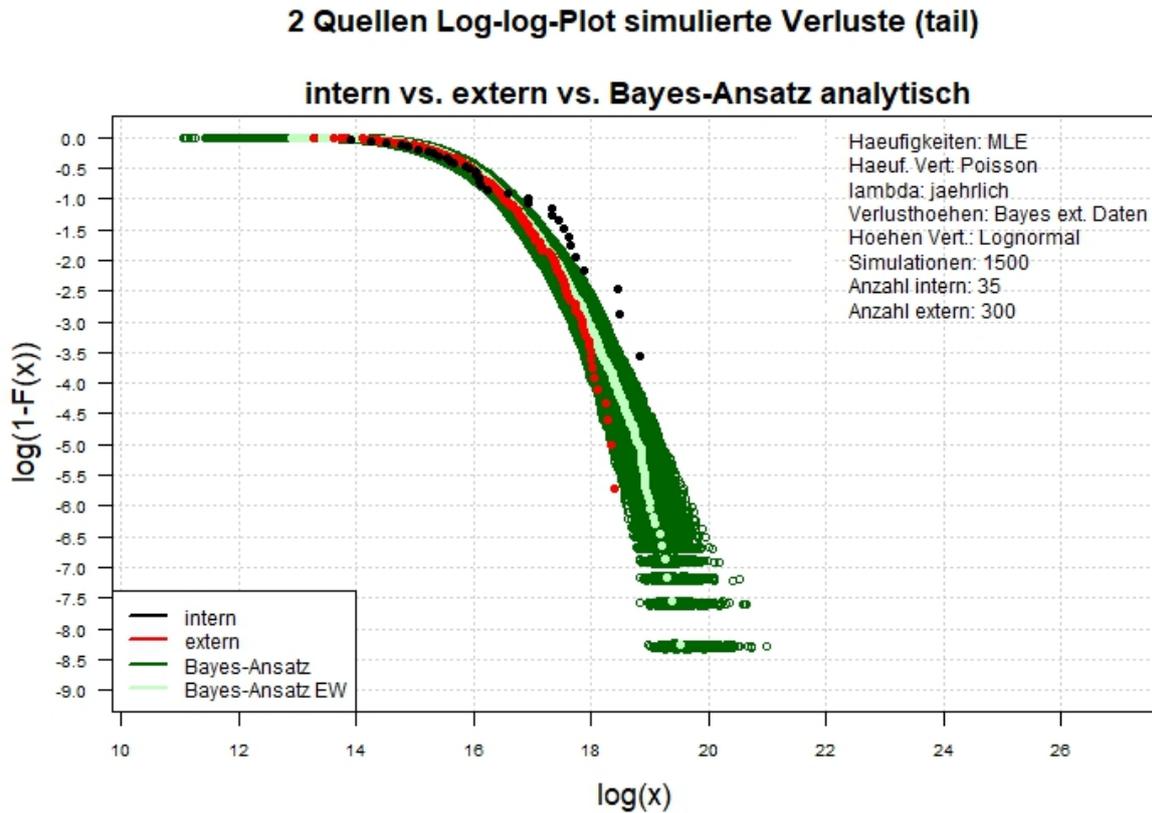


Abbildung 5.3.: Der Log-log-Plot. Bayes-Ansatz analytisch. Zwei-Schritt-Verfahren. Interne Daten: 35. Externe Daten: 300.

<sup>28</sup>Als Bayes-Schätzer kann hierbei auch *a posteriori* Modus genommen werden.

### 5.1.3. Schätzung der Schadenhöhenverteilung unter Verwendung des MH-Algorithmus (Verallgemeinerung)

Die im Kap. 5.1.2 beschriebene Vorgehensweise bei der Schätzung der Schadenshöhenverteilung eignet sich für Fälle, in denen die analytische Herleitung der *a posteriori* bzw. Hyperparameter unkompliziert ist, sodass die direkte Ziehung aus der *a posteriori* Verteilung möglich ist. Dies ist unter vielen Annahmen und unter Verwendung von bestimmten Verteilungen realisierbar<sup>29</sup>. In der Realität ist die analytische Herleitung der *a posteriori* Verteilung meistens schwierig und komplex. Allerdings ermöglichen die Monte-Carlo-Methoden (Kap. 4.1) die praktische Umsetzung der Bayes-Ansätze ohne großen Aufwand und lange Herleitungen. Wie die Analysen gezeigt haben (Kap. 4.7.2), ist der MH-Algorithmus wegen der Rechenzeiten für die praktische Anwendungen besser als der MH-SA-Algorithmus geeignet und wird aus diesem Grund in dieser Arbeit verwendet.

Um die Schätzung der Schadenhöhenverteilung unter Verwendung des MH-Algorithmus und unter Verwendung der analytischen Herleitung gegenüberstellen zu können, wird die gleiche Ausgangssituation (Lognormal-verteilte Verluste) mit folgender Annahmen betrachtet:

- Die Daten  $Loss$  und  $Loss_{n+1}$  sind bedingt unabhängig (gegeben Parameter  $\Theta$ ).
- Die Daten  $Loss_1, \dots, Loss_n, Loss_{n+1}$  gegeben Parameter  $\Theta_\mu = \mu$  und  $\Theta_{\sigma^2} = \sigma^2$  sind *iid.*
- Die *a priori* Verteilung im Schritt 1. des Zwei-Schritt-Bayes-Ansatzes (Tabelle 5.1) ist die Jeffreys *a priori* für die Lognormal-Verteilung (Formel 5.19).
- Die *a priori* Dichte im Schritt 2. des Zwei-Schritt-Bayes-Ansatzes (Tabelle 5.1) wird mithilfe der Kerndichte-Schätzung (Kap. 4.6) aus den simulierten Parametern bestimmt.

### Ergebnisse der Schätzung

Für die Analysen der Schätzung der Schadenshöhenverteilung wurden die gleichen sechs Datensätze benutzt (siehe Tabelle 5.4). Zunächst werden die Ergebnisse auf Basis des ersten Datensatzes ( $\lambda_{int} = 0.86$  und  $\lambda_{ext} = 6$ ) dargestellt. Die Ergebnisse auf Basis der anderen fünf Datensätzen sind im Anh. A.10 zu finden.

In der Abb. 5.4 und Abb. 5.5 sind Simulationen der zwei Parameter *meanlog* und *sdlog* aus dem Schritt 1 des Zwei-Schritt-Bayes-Verfahrens dargestellt. Jeweils oben links sind die Markov-Ketten mit *burnin* (rote Linie) für 8750 Iterationen abgebildet und unten links die dazugehörigen Histogramme. Jeweils unten rechts sind die Histogramme für die letzten 7000 Iterationen (ohne *burnin*) dargestellt. Aus den Grafiken ist zu erkennen, dass die Mittelwerte der Markov-Kette ohne *burnin* (*meanlog* = 16.19308 und *sdlog* = 1.00833) gegen die empirischen Parameter der externen Daten (*meanlog* = 16.19309,

---

<sup>29</sup>Z. B. nicht komplizierte Daten-, konjugierte *a priori* Verteilungen (Siehe dazu Kap. 3.1 Def. 1).

$sdlog = 1.00802$ ) konvergieren, was auch durch die Verwendung der Jeffrey's *a priori* erklärbar ist.

In der Abb. 5.6 sind die Autokorrelationsfunktionen der Markov-Kette ohne *burnin* (obere Zeile) und der verdünnten Markov-Kette ohne *burnin* (untere Zeile) abgebildet. Für die KDE werden die verdünnten Markov-Ketten ( $thin = 7$ ) verwendet, weil die Autokorrelationen zwischen den weiter in der Vergangenheit liegenden Werten hierbei kleiner ist.

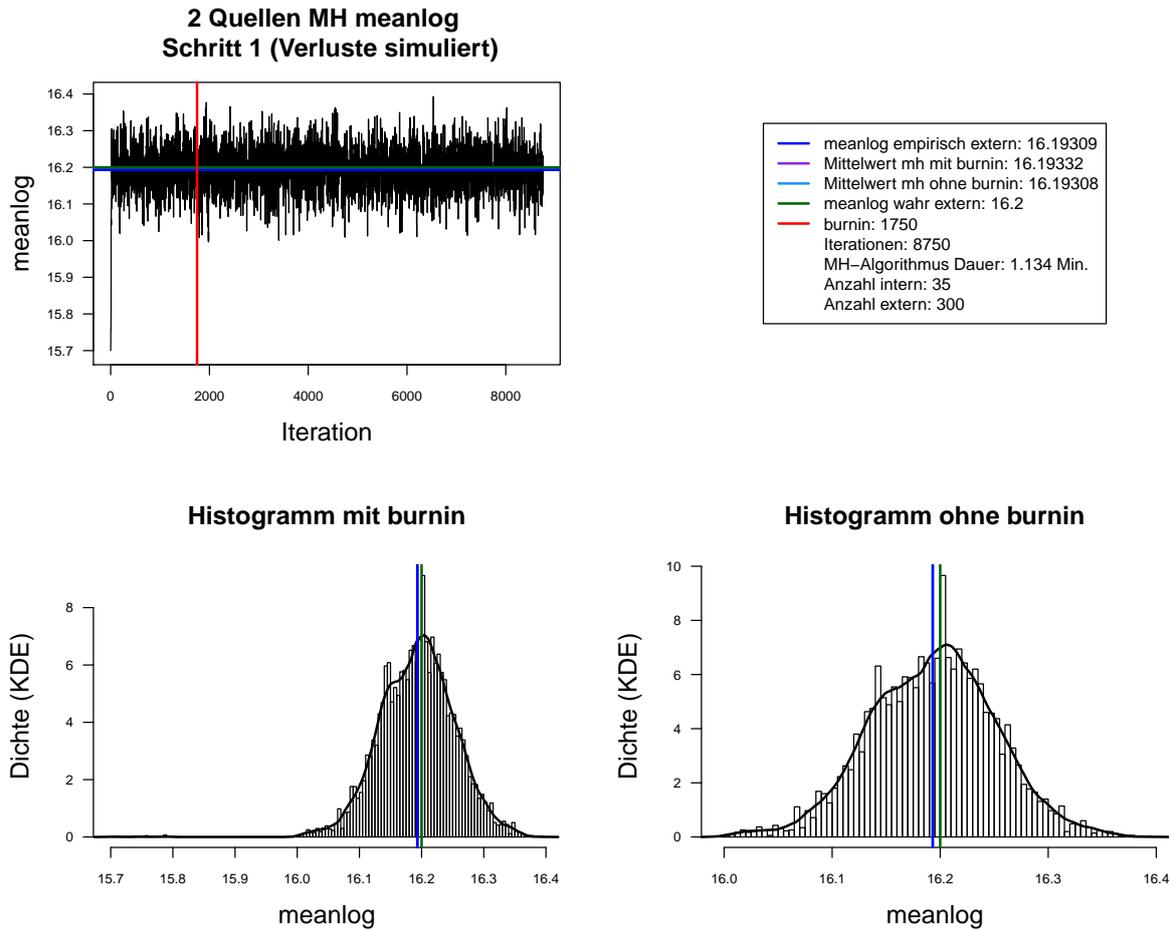


Abbildung 5.4.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 8750 Iterationen beträgt 1.134 Minuten. Markov-Kette (oben links), Histogramm für 8750 Iterationen mit KDE (unten links) und Histogramm für letzte 7000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 35. Externe Daten: 300.

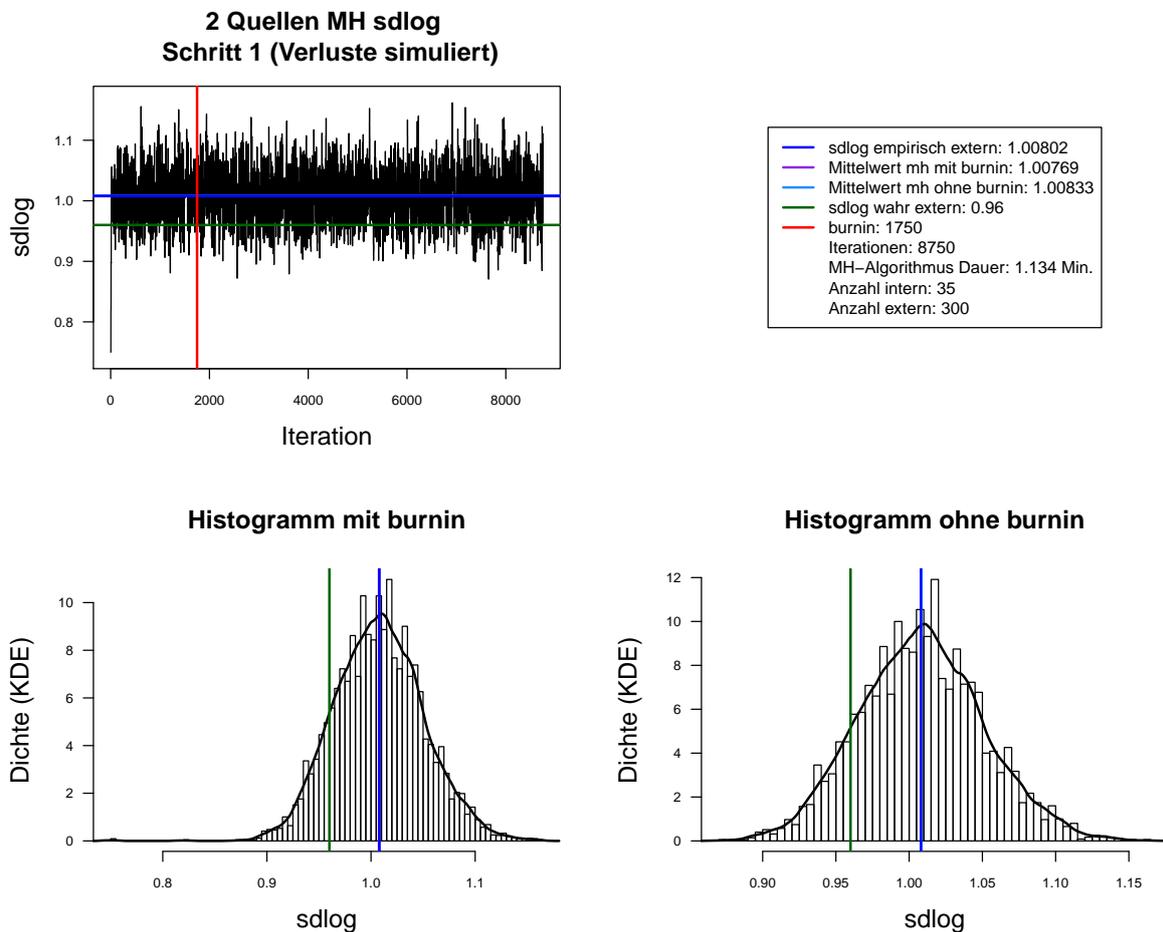
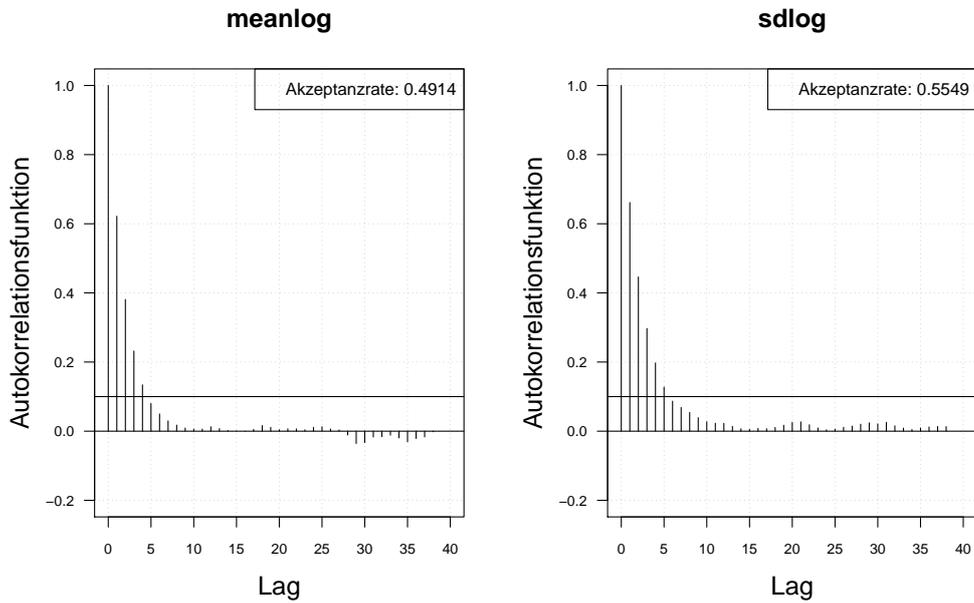


Abbildung 5.5.: Der mit MH-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 8750 Iterationen beträgt 1.134 Minuten. Markov-Kette (oben links), Histogramm für 8750 Iterationen mit KDE (unten links) und Histogramm für letzte 7000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 35. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
Iterationen: 7000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 7, Iterationen: 7000

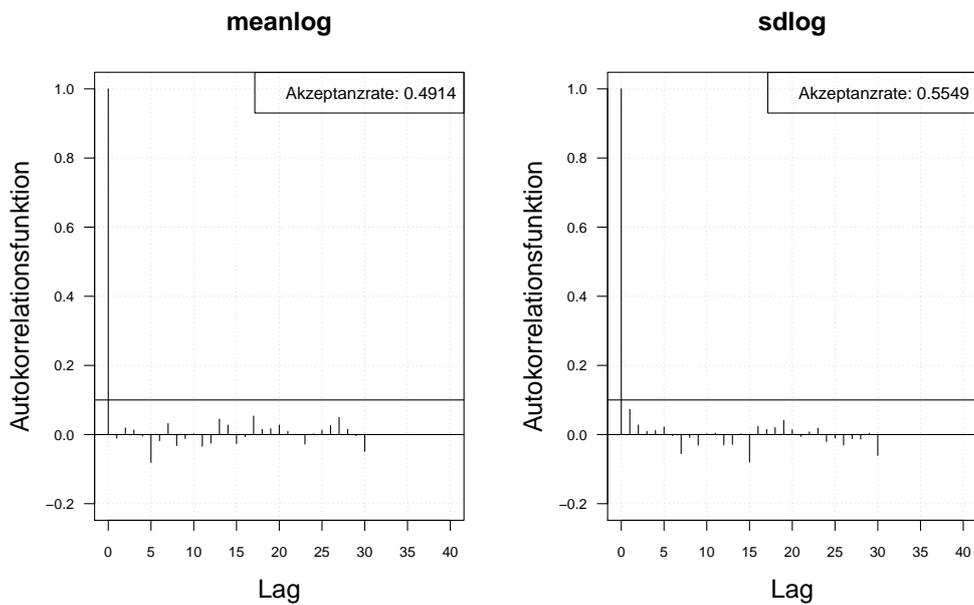


Abbildung 5.6.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 7000. Schritt 1. Interne Daten: 35. Externe Daten: 300.

In der Abb. 5.7 und Abb. 5.8 sind Simulationen der zwei Parameter  $meanlog$  und  $sdlog$  aus dem Schritt 2 des Zwei-Schritt-Bayes-Verfahrens dargestellt. Jeweils oben links sind die Markov-Ketten mit  $burnin$  (rote Linie) für 8750 Iterationen abgebildet und unten links die dazugehörigen Histogramme. Jeweils unten rechts sind die Histogramme für letzte 7000 Iterationen (ohne  $burnin$ ) dargestellt. Aus den Grafiken ist zu erkennen, dass die Mittelwerte der Markov-Ketten ohne  $burnin$  ( $meanlog = 16.21458$  und  $sdlog = 1.05366$ ) etwas höher sind als in Schritt 1 ( $meanlog = 16.19308$ : Abb. 5.4 und  $sdlog = 1.00833$ : Abb. 5.5). Allerdings genau so, wie im analytischen Fall neigen sie sich zu den Parametern der externen Daten (empirisch:  $meanlog = 16.19309$ ,  $sdlog = 1.00802$  und wahr:  $meanlog = 16.2$ ,  $sdlog = 0.96$ ).

In der Abb. 5.9 sind die Autokorrelationsfunktionen der Markov-Kette ohne  $burnin$  (obere Zeile) und der verdünnten Markov-Kette ohne  $burnin$  (untere Zeile) dargestellt. Für weitere Simulationen werden die verdünnten Markov-Ketten ( $thin = 7$ ) verwendet, weil die Autokorrelationen zwischen den weiter in der Vergangenheit liegenden Werten hierbei kleiner ist.

Das Ergebnis der simulierten Vorhersageverteilung der Schadenshöhen im Vergleich zu den Verteilungen der internen und externen Daten ist wie im analytischen Fall (Vgl. Abb. 5.3) im Log-log-Plot (Abb. 5.10) zusammengefasst. Die schwarzen und roten Punkte bilden die internen und externen Verteilungen entsprechend ab. Zu jeder Kombination der aus *a posteriori* Verteilungen gezogenen Parameter (die verdünnte Markov-Kette beträgt Länge 1000) wurden wie zuvor Verluste für die folgenden 1500 Jahre (Simulation: 1500) simuliert (dunkelgrüne Punkte). Die hellgrünen Punkte zeigen wiederum die simulierten Verluste mit Bayes-Schätzer (Erwartungswert der *a posteriori* Verteilung<sup>30</sup>). Wie im analytischen Fall neigt sich die Vorhersageverteilung bei kleineren Werten (x-Achse: ca. 13 bis 18, y-Achse: ca. -2.5 bis 0) zur externen Verteilung und weicht im Randbereich (x-Achse: ca. ab 18, y-Achse: ca. bis -2.5) in Richtung der internen Daten ab.

Die Ergebnisse der Schätzungen der Schadenshöhenverteilung auf Basis aller sechs Datensätze (siehe Tabelle 5.4) unter der Verwendung der analytischen Herleitung und unter der Verwendung der MH-Algorithmus sind identisch<sup>31</sup>. Für kleine Anteile der internen Daten ähneln die geschätzten Parameter und folglich auch die simulierten Verluste den externen Daten. Je größer der Anteil der internen Daten ist, desto mehr neigt die Vorhersageverteilung zu den internen Daten. Die Identität der Ergebnisse bestätigt, dass die Verwendung des MH-Algorithmus die numerische Äquivalenzlösung der analytischen Herleitung ist. Die Verwendung des MH-Algorithmus kann somit als generische Lösung (Verallgemeinerung) bei Bayes-Ansätzen verwendet werden, insbesondere dann, wenn die analytische Herleitung kompliziert oder nicht möglich ist.

---

<sup>30</sup>Als Bayes-Schätzer kann hierbei auch *a posteriori* Modus genommen werden.

<sup>31</sup>Ergebnisse siehe Anh. A.10

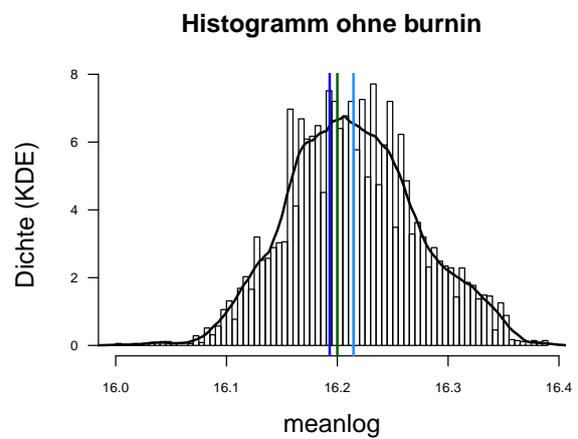
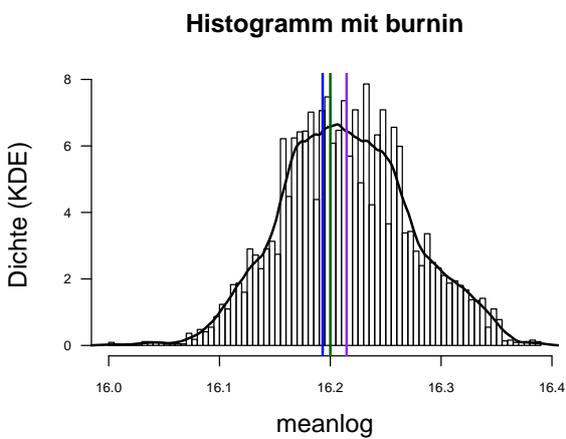
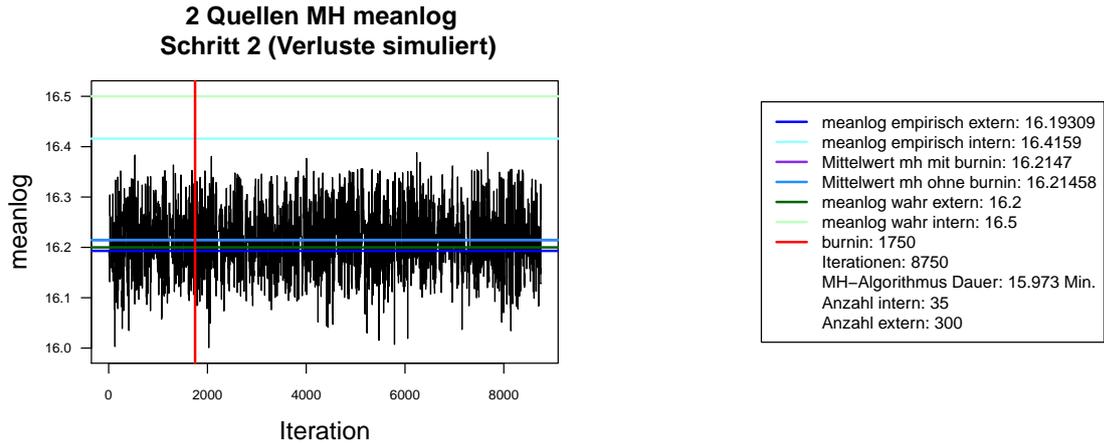


Abbildung 5.7.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 8750 Iterationen beträgt 15.973 Minuten. Markov-Kette (oben links), Histogramm für 8750 Iterationen mit KDE (unten links) und Histogramm für letzte 7000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 35. Externe Daten: 300.

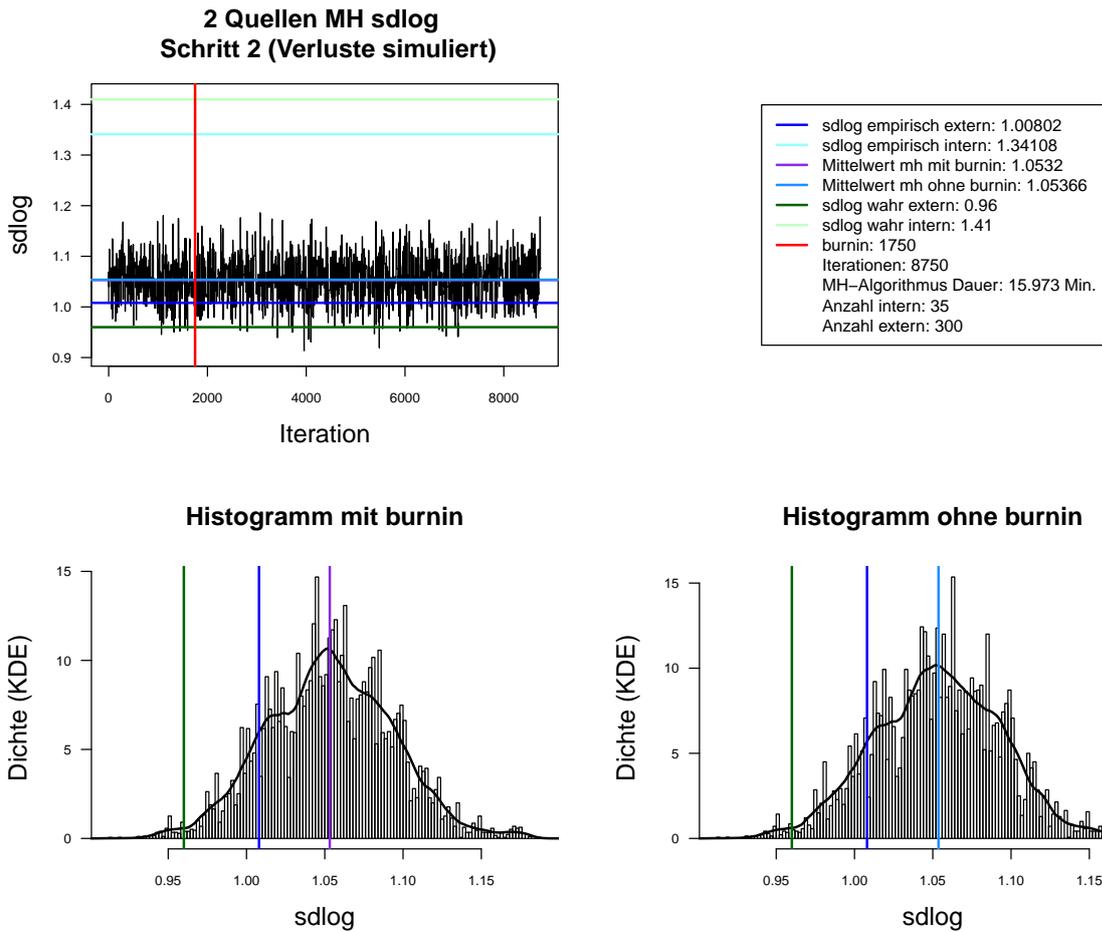
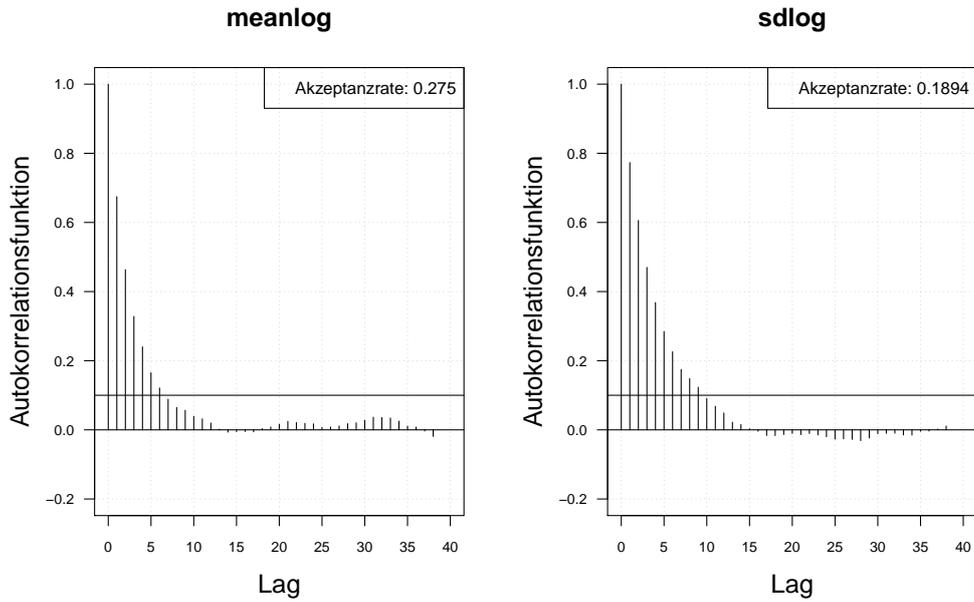


Abbildung 5.8.: Der mit MH-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 8750 Iterationen beträgt 15.973 Minuten. Markov-Kette (oben links), Histogramm für 8750 Iterationen mit KDE (unten links) und Histogramm für letzte 7000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 35. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 2. Ohne burnin (Verluste simuliert),  
Iterationen: 7000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 7, Iterationen: 7000

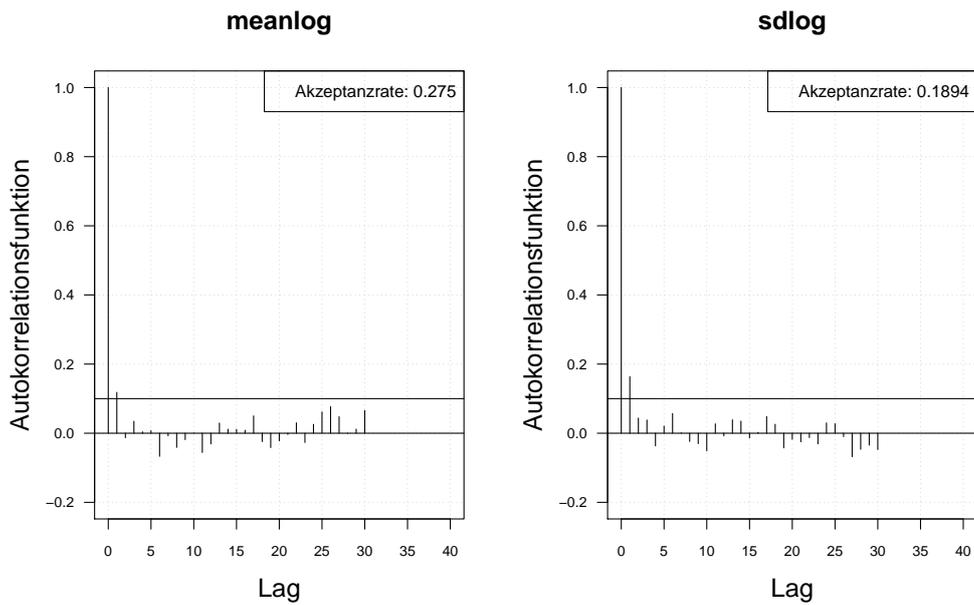


Abbildung 5.9.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 7000. Schritt 2. Interne Daten: 35. Externe Daten: 300.

## 2 Quellen Log-log-Plot simulierte Verluste (tail)

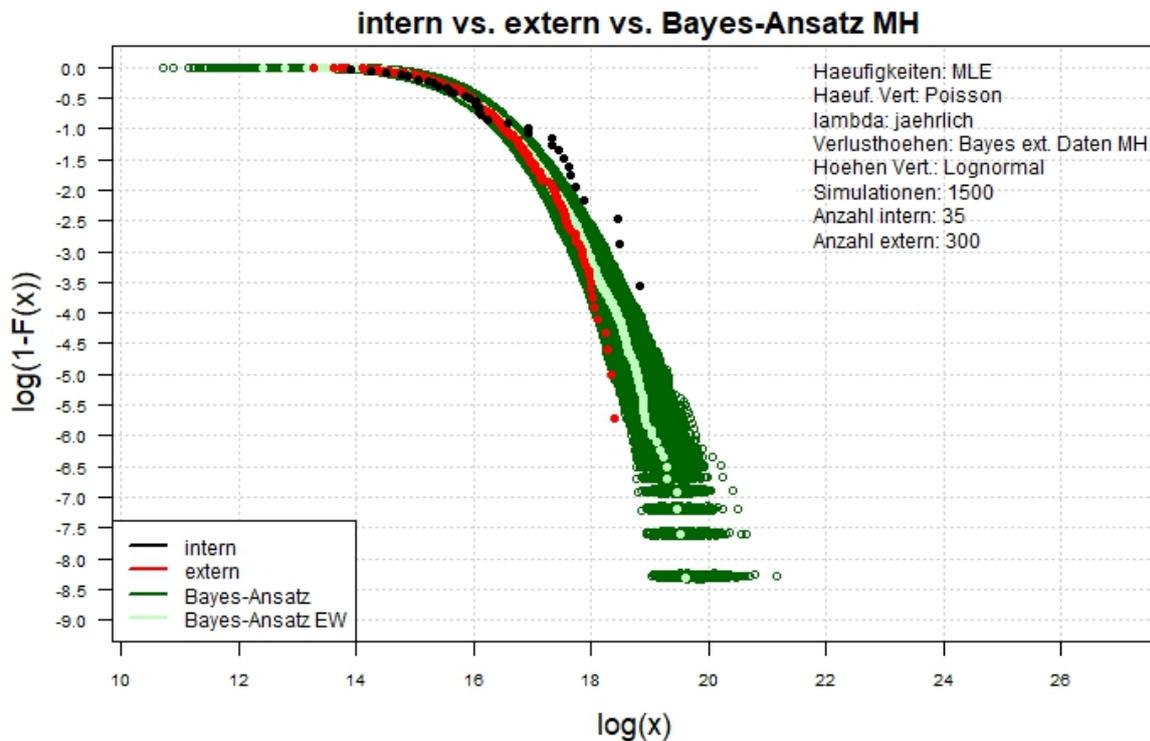


Abbildung 5.10.: Der Log-log-Plot. Bayes-Ansatz MH. Zwei-Schritt-Verfahren. Interne Daten: 35. Externe Daten: 300.

### 5.1.4. Auswirkung der Schätzung auf VaR

Da das Ziel der OpRisk-Messung unter anderem ist, VaR zu bestimmen, ist es sinnvoll zu untersuchen, welche Auswirkungen die Umsetzung des Zwei-Schritt-Bayes-Verfahrens unter Verwendung von unterschiedlichen Datensätzen auf VaR hat. Da die Verwendung des MH-Algorithmus die generische Lösung (Verallgemeinerung) bei Bayes-Ansätzen ist, wurde diese zur Ermittlung des Bayes-Schätzers verwendet. Neben dem Bayes-Schätzer nach dem Zwei-Schritt-Verfahren wurden zum Vergleich MLE und Bayes-Schätzer nach dem Ein-Schritt-Verfahren ermittelt. Die Schätzer wurden hierbei jeweils aus internen, externen und kombinierten (zusammengefügten internen und externen) Daten ermittelt. Die geschätzten Parameter und deren Standardabweichung sind in der Tabelle 5.5 dargestellt. Es ist zu erkennen, dass der MLE und Ein-Schritt-Bayes-Schätzer jeweils für interne, externe und kombinierte Daten identisch sind, was durch die Verwendung der uninformativen *a priori* erklärbar ist. Die geschätzten Parameter nach Zwei-Schritt-Bayes-Verfahren liegen im gleichen Bereich mit MLE und Ein-Schritt-Bayes-Schätzer für kombinierte Daten und nähern sich bei allen sechs Datensätzen eher den wahren Parametern der externen als internen Daten. Im Unterschied zum Ein-Schritt-Bayes-

Schätzer für kombinierte Daten liegen der MLE und der Zwei-Schritt-Bayes-Schätzer zwischen den wahren Parametern den internen und externen Daten für alle sechs Datensätze. Ebenso ist es ersichtlich, dass die Schätzer für die Datensätze mit dem kleineren Anteil von internen Daten (z. B. Daten 1) weniger von den wahren Parametern der externen Daten abweichen als die Schätzer für die Datensätze mit dem größeren Anteil der internen Daten (z. B. Daten 4, 5 und 6). Je größer der Anteil der internen Daten in der Datenbank ist, desto mehr ähnelt der Schätzer den wahren Parametern der internen Daten. Die Standardabweichung des Zwei-Schritt-Bayes-Schätzers für kombinierte Daten ist für alle sechs Datensätze kleiner als die Standardabweichung des Ein-Schritt-Bayes-Schätzers und fast immer kleiner als die Standardabweichung des MLE (außer Standardabweichung für *meanlog* im ersten Datensatz).

Zu den oben geschätzten Parametern wurde die VaR zum Niveau von  $\alpha = 99.95$  ermittelt. Die Werte sind in der Tabelle 5.6 zusammengefasst. Es ist zu erkennen, dass VaR ermittelt auf Basis des Zwei-Schritt-Bayes-Schätzers deutlich kleiner ist als VaR ermittelt auf Basis des Ein-Schritt-Bayes-Schätzers und MLE.

### Die geschätzten Parameter für sechs Datensätze

	Daten 1		Daten 2		Daten 3		Daten 4		Daten 5		Daten 6	
	meanlog	sdlog										
int. wahr	16.500	1.410	16.500	1.410	16.500	1.410	16.500	1.410	16.500	1.410	16.500	1.410
ext. wahr	16.200	0.960	<b>16.200</b>	0.960	16.200	0.960	16.200	0.960	16.200	0.960	16.200	0.960
int. MLE	16.416	1.322	16.360	1.515	16.490	1.478	16.360	1.515	16.406	1.481	16.490	1.478
sd. int. MLE	0.223	0.158	0.152	0.107	0.085	0.060	0.152	0.107	0.105	0.074	0.085	0.060
ext. MLE	16.193	1.006	16.188	0.990	16.210	0.977	16.193	1.006	16.193	1.006	16.193	1.006
sd. ext. MLE	0.058	0.041	0.027	0.019	0.018	0.013	0.058	0.041	0.058	0.041	0.058	0.041
<b>total. MLE</b>	<b>16.216</b>	<b>1.046</b>	<b>16.200</b>	<b>1.036</b>	<b>16.236</b>	<b>1.035</b>	<b>16.235</b>	<b>1.157</b>	<b>16.278</b>	<b>1.223</b>	<b>16.341</b>	<b>1.273</b>
sd. total. MLE	<b>0.057</b>	0.040	0.027	0.019	0.018	0.013	0.058	0.041	0.055	0.039	0.052	0.037
int. Bayes 1-Schritt	16.417	1.334	16.364	1.524	16.490	1.476	16.381	1.522	16.397	1.482	16.490	1.476
sd. int. Bayes 1-Schritt	0.216	0.157	0.154	0.110	0.089	0.058	0.168	0.108	0.106	0.071	0.089	0.058
ext. Bayes 1-Schritt	16.191	1.006	16.187	0.990	16.209	0.976	16.196	1.008	16.191	1.006	16.196	1.009
sd. ext. Bayes 1-Schritt	0.059	0.042	0.026	0.019	0.017	0.013	0.057	0.042	0.055	0.041	0.055	0.043
<b>total. Bayes 1-Schritt</b>	<b>16.218</b>	<b>1.049</b>	<b>16.198</b>	<b>1.035</b>	<b>16.236</b>	<b>1.037</b>	<b>16.235</b>	<b>1.158</b>	<b>16.276</b>	<b>1.224</b>	<b>16.341</b>	<b>1.273</b>
sd. total. Bayes 1-Schritt	0.059	0.041	0.028	0.019	0.019	0.013	0.060	0.043	0.054	0.039	0.052	0.036
<b>Bayes EW 2-Schritt</b>	<b>16.215</b>	<b>1.053</b>	<b>16.201</b>	<b>1.040</b>	<b>16.223</b>	<b>1.008</b>	<b>16.233</b>	<b>1.123</b>	<b>16.258</b>	<b>1.141</b>	<b>16.286</b>	<b>1.151</b>
sd. Bayes EW 2-Schritt	<b>0.058</b>	0.040	0.026	0.019	0.013	0.006	0.047	0.021	0.034	0.015	0.030	0.013
Bayes min. 2-Schritt	16.037	0.931	16.123	0.969	16.176	0.986	16.056	1.039	16.094	1.087	16.204	1.102
Bayes max. 2-Schritt	16.389	1.178	16.282	1.075	16.271	1.023	16.369	1.172	16.351	1.179	16.363	1.179

53

Tabelle 5.5.: Die geschätzten Parameter für sechs Datensätze. Verschiedene Schätzer und deren Standardabweichung (sd.) für interne Daten (int.), externe Daten (ext.), zusammengefügte/kombinierte Daten (total.). Standardabweichung von Zwei-Schritt-Bayes-Schätzer für *meanlog* der kombinierten Daten beim ersten Datensatz (0.058) ist größer als die Standardabweichung von MLE für kombinierte Daten (0.057). Zwei-Schritt-Bayes-Schätzer für *meanlog* der kombinierten Daten beim zweiten Datensatz (16.198) ist kleiner als der wahre Parameter der externen Daten (16.2).

**VaR<sub>99,95</sub> für sechs Datensätze auf Basis der verschiedenen Schätzer**

	Daten 1	Daten 2	Daten 3	Daten 4	Daten 5	Daten 6
99.95-Quantil int. wahr	1516457245	1516457245	1516457245	1516457245	1516457245	1516457245
99.95-Quantil ext. wahr	255544127	255544127	255544127	255544127	255544127	255544127
VaR <sub>99,95</sub> int. MLE	1420452279	3118814352	54445380512	9406421634	6885534844	54445380512
VaR <sub>99,95</sub> ext. MLE	1323095725	2748602603	6298927111	1276584256	1409048939	1187512130
<b>VaR<sub>99,95</sub> total. MLE</b>	<b>1265488245</b>	<b>3452591330</b>	<b>6710821324</b>	<b>3104353685</b>	<b>3303930512</b>	<b>4862503155</b>
VaR <sub>99,95</sub> int. Bayes 1-Schritt	1914184368	7664747982	6563472091	5673327307	12203764580	4933800691
VaR <sub>99,95</sub> ext. Bayes 1-Schritt	1043915677	3020298291	5618656328	1764259774	1475099280	1819132455
<b>VaR<sub>99,95</sub> total. Bayes 1-Schritt</b>	<b>1436943316</b>	<b>3306614580</b>	<b>6615851022</b>	<b>2502210361</b>	<b>2381608780</b>	<b>5992902618</b>
<b>VaR<sub>99,95</sub> Bayes EW 2-Schritt</b>	<b>471435535</b>	<b>1142208684</b>	<b>1177315500</b>	<b>889122093</b>	<b>2208109831</b>	<b>2102041426</b>
VaR <sub>99,95</sub> Bayes min. 2-Schritt	270518318	473928857	943457373	559394769	882021759	1318866659
VaR <sub>99,95</sub> Bayes max. 2-Schritt	2463005923	3964036312	4972781105	6076981729	8749679908	12096490840

Tabelle 5.6.: VaR<sub>99,95</sub> für sechs Datensätze. VaR<sub>99,95</sub> ist auf Basis der verschiedenen Schätzer für interne Daten (int.), externe Daten (ext.), zusammengefügte/kombinierte Daten (total.) ermittelt.

## 5.2. Kombination von drei Datenquellen

Das Zwei-Schritt-Bayes-Verfahren kann zur Kombination von drei Datenquellen benutzt werden, indem die *a priori* aus den Szenarien im Schritt 1.  $p_0(\underline{\theta}) = p_{sz}(\underline{\theta})$  herangezogen wird. Bei der Schätzung der Schadenshöhenverteilung unter Verwendung der analytischen Herleitung können die Hyperparameter der *a priori* Verteilung beispielsweise wie im Kap. 5.1.2 beschrieben bestimmt werden. Bei der Schätzung der Schadenshöhenverteilung unter Verwendung des MH-Algorithmus (Verallgemeinerung) können die Verluste aus Szenarien mehrmals simuliert und daraus die Vektoren der empirischen Parameter erzeugt werden. Die daraus geschätzte Dichte (KDE) kann als *a priori* im Schritt 1. benutzt werden.

## 6. Der empirische Bayes-Ansatz in der OpRisk-Messung

Die Idee des empirischen Bayes-Ansatzes ist weiterhin, die *a priori* aus den externen Verlustdaten zu bestimmen und diese mit der *Likelihood* der internen Verlustdaten zu kombinieren, um die *a posteriori* für die internen Verlustdaten unter Berücksichtigung der externen Verlustdaten zu bekommen. Die marginale Dichte aller Beobachtungen (Formel 3.7) kann für externe Verluste folgendermaßen umgeschrieben werden:

$$f(\underline{loss}_{ext}^{(1)}, \dots, \underline{loss}_{ext}^{(J)}) = \prod_{j=1}^J \int \left[ \prod_{i=1}^{n_{ext}^{(j)}} f(\underline{loss}_{ext,i}^{(j)} | \underline{\theta}_{ext}^{(j)}) \right] p(\underline{\theta}_{ext}^{(j)}) d\underline{\theta}_{ext}^{(j)}, \quad (6.1)$$

mit  $J$  - Anzahl der externen Banken bzw. Bankgruppen. Die daraus geschätzte *a priori*  $p(\underline{\theta}_{ext}^{(j)}) = p(\underline{\theta}_{ext})$  wird für die Ermittlung der *a posteriori* benutzt, die dann für die Berechnung der Vorhersageverteilung genutzt werden kann:

$$p(\underline{\theta}_{int} | \underline{loss}_{int}) \propto \prod_{i=1}^{n_{int}} f(\underline{loss}_{int,i} | \underline{\theta}_{int}) p(\underline{\theta}_{ext}). \quad (6.2)$$

Da das DakOR aus Datenschutzgründen nur anonymisierte Daten liefert, liegen die Informationen über die Verteilung der Verlustdaten zwischen den externen Banken bzw. Bankgruppen nicht vor. Die praktische Umsetzung des empirischen Bayes-Ansatzes ist somit nur unter Verwendung der Bootstrap-Methode möglich, indem die  $J$  Stichproben mit Zurücklegen der Länge  $n_{ext}$  aus externen Daten gezogen und für jede Stichprobe die Parameter geschätzt (z. B. MLE) werden und daraus die *a priori* ermittelt wird.

## 7. Der nicht parametrische Bayes-Ansatz in der OpRisk-Messung

Die Basisverteilung  $H(x)$  kann hierbei als die Expertenmeinung für  $F(x)$  festgelegt werden, die auf Informationen aus Szenarien basiert.

Setzt man den Dirichlet-Prozess  $DP(\alpha, H(loss_{sz}))$  als *a priori* mit marginaler Verteilung

$$F(loss_{sz}) \sim Be(\alpha H(loss_{sz}), \alpha(1 - H(loss_{sz}))), \quad (7.1)$$

hat man als *a posteriori* wiederum einen Dirichlet-Prozess:

$$DP(\alpha + n_{int}, \frac{\alpha}{\alpha + n_{int}} H(loss_{sz}) + \frac{n_{int}}{\alpha + n_{int}} \frac{1}{n_{int}} \sum_i \mathbb{1}_{loss_{int,i} < loss_{int}}), \quad (7.2)$$

mit dem *a posteriori* Erwartungswert:

$$E_{a \text{ posteriori}} (F(loss_{int})) = \frac{\alpha}{\alpha + n_{int}} H(loss_{sz}) + \frac{n_{int}}{\alpha + n_{int}} \frac{1}{n_{int}} \sum_i \mathbb{1}_{loss_{int,i} < loss_{int}}. \quad (7.3)$$

Der *a posteriori* Erwartungswert ist hier eine gewichtete Summe der Verteilungsfunktion aus Szenarien  $H(loss_{sz})$  und der empirischen Verteilungsfunktion der internen Verluste  $\frac{1}{n} \sum_i \mathbb{1}_{loss_{int,i} < loss_{int}}$ . Die *a posteriori* Verteilung hängt stark von der Anzahl der internen Daten  $n_{int}$  ab. Liegen im Datenbestand nur wenige interne Beobachtungen vor, so haben diese kaum einen Einfluss auf die Schätzung, was eher gegen die Verwendung dieser Methode in der OpRisk-Messung spricht.

## 8. Zusammenfassung und Ausblick

In dieser Arbeit wurden drei Bayes-Ansätze zur Integration von externen Daten und Szenarien in die OpRisk-Messung dargestellt: der reine, der empirische und der nicht parametrische Bayes-Ansatz. Hierbei wurde insbesondere auf den reinen Bayes-Ansatz eingegangen. Zum einen liegen im Vergleich zum empirischen Bayes-Ansatz alle hierfür nötigen Informationen vor. Zum anderen ist es im Vergleich zum nicht parametrischen Bayes-Ansatz möglich, relativ gute Schätzung der Parameter zu bekommen, auch im Falle, dass im Datenbestand nur wenige interne Daten vorliegen.

Neben der Umsetzung auf Basis der analytischen Herleitungen wurde die numerische Äquivalenzlösung unter Verwendung der Monte-Carlo-Methoden (MH-Algorithmus) sowie anderer Hilfsmethoden (KDE und SA-Algorithmus) dargestellt. Die Anwendung des MH-Algorithmus hat gewisse Vorteile. Zum einen ist es so möglich, Bayes-Ansätze auch dann anzuwenden, wenn die analytische Herleitung komplex oder unmöglich ist. Zum anderen sind die Rechenzeiten relativ klein für häufige Durchläufe. Der MH-Algorithmus für zwei unbekannte Parameter und der um SA-Algorithmus und Berechnung der Krümmung erweiterte MH-Algorithmus wurden in dieser Arbeit verglichen. Die Analysen haben gezeigt, dass der MH-SA-Algorithmus aus Sicht der Unabhängigkeit der Ziehungen und der Vermischung viel besser ist als der MH-Algorithmus. Es gibt die Möglichkeit, die Unabhängigkeit der Ziehungen beim MH-Algorithmus zu steuern, indem die verdünnte Markov-Kette verwendet wird. Ebenso können mit der Angabe des Skalenparameters für die Vorschlagsdichten die Vermischung und die Akzeptanzwahrscheinlichkeit gesteuert werden. Dies muss für jeden neuen Datensatz und für jeden Schritt des Zwei-Schritt-Bayes-Verfahrens erledigt werden, was auch den MH- im Vergleich zum MH-SA-Algorithmus nachteilig macht. Außerdem leistet der MH-SA-Algorithmus eine bessere Stabilität der Ergebnisse und Approximation der Parameterverteilung. Allerdings sind die Rechenzeiten des MH-SA-Algorithmus wegen der zweifachen Optimierung in jeder Iteration viel größer: z. B. dauert der MH-SA-Algorithmus (75.91 Minuten) für 1250 Iterationen ca. 850 Mal länger als der MH-Algorithmus (0.09 Minuten). An diesem Punkt wäre es sinnvoll, für weitere über diese Arbeit hinausgehende Analysen oder produktive Umsetzung den MH-SA-Algorithmus in den Rechenzeiten zu optimieren (Parallelisierung, etc.).

Basierend auf simulierten Schadensfalldaten wurde die Umsetzung des reinen Bayes-Ansatzes auf Basis der analytischen Herleitung bei der Schätzung der Schadenhäufigkeitsverteilung mit MLE verglichen und deren sinnvolle Anwendung diskutiert. Abgesehen vom Ziel der OpRisk-Messung, vor allem die interne Situation abzuschätzen und die internen Verluste vorherzusagen, ist es sinnvoller, die Schadenhäufigkeitsverteilung nur aus internen Daten zu schätzen.

Basierend auf simulierten Schadensfalldaten wurden zwei Möglichkeiten der Umsetzung des reinen Bayes-Ansatzes zum Kombinieren der internen und externen Daten

untersucht: die Schätzung der Schadenshöhenverteilung unter Verwendung der analytischen Herleitung und unter Verwendung des MH-Algorithmus. Die Ergebnisse der Schätzungen der Schadenshöhenverteilung in beiden Fällen sind identisch. Die Identität der Ergebnisse bestätigt, dass die Verwendung der MH-Algorithmus die numerische Äquivalenzlösung der analytischen Herleitung ist, d. h. die Verwendung des MH-Algorithmus kann als generische Lösung (Verallgemeinerung) bei Bayes-Ansätzen verwendet werden, insbesondere dann, wenn die analytische Herleitung kompliziert oder nicht möglich ist. Die Analysen der simulierten Daten zeigen, dass die geschätzten Parameter und folglich auch die simulierte Vorhersageverteilung für kleine Anteile der internen Daten den externen Daten ähneln. Je größer der Anteil der internen Daten ist, desto mehr neigt die Vorhersageverteilung zu den internen Daten.

Die Schwachstelle des Zwei-Schritt-Bayes-Verfahrens unter Verwendung des MH-Algorithmus ist nicht nur die kleine Anzahl der internen Beobachtungen, sondern auch der Unterschied der Parameter der Verlustverteilungen. Je unterschiedlicher die Parameter der Verlustverteilungen (intern, extern sowie Szenarien) sind, desto kleiner wird die Akzeptanzwahrscheinlichkeit der MH-Algorithmus. Weitere Schwachstelle des MH-Algorithmus ist die Auswahl der Startparameter. Wenn diese sehr stark von der wahren Parametern abweichen, ergibt sich eine Akzeptanzwahrscheinlichkeit von Null. Die optimalen Startwerte wären hier die empirischen Parameter der Daten.

Über die in dieser Arbeit beschriebenen Analysen hinaus gibt es noch viele weitere Möglichkeiten, die Untersuchungen zu erweitern. Ein Beispiel hierfür wäre die Anwendung des Zwei-Schritt-Bayes-Verfahrens unter der Verwendung des MH-SA-Algorithmus mit Erweiterung für andere Schadenshöhenverteilungen (z. B. Weibull-, generalisierte Pareto-Verteilung oder beliebige andere rechtsschiefe *heavy-tailed*-Verteilungen). Vorab müssten die Verteilungs-Anpassungstests sowohl für die internen als auch für die externen Verluste durchgeführt werden.

Außerdem sollte die Kombination der drei Datenquellen in der Praxis untersucht und analysiert werden. Der Unterschied der Parameter der verschiedenen Szenarien kann Schwierigkeiten bei der praktischen Umsetzung verursachen. Folglich könnte es sinnvoller sein, die anderen Methoden zur Integration von Szenarien zu untersuchen und zu analysieren.

## Abbildungsverzeichnis

1.1. Aufteilung der Verteilung von Gesamtschaden. . . . .	3
2.1. Balkendiagramm Schadensfälle <i>tail</i> -Bereich. . . . .	5
2.2. Histogramme <i>tail</i> -Bereich. Interne und externe Verluste. . . . .	6
2.3. QQ-Plot <i>tail</i> -Bereich. Verlust intern vs. extern. . . . .	7
2.4. Log-log-Plot <i>tail</i> -Bereich. Verlust intern vs. extern. . . . .	7
3.1. Der empirische Bayes-Ansatz. Interpretation der <i>a priori</i> . . . . .	11
4.1. MH <i>meanlog</i> der Lognormal-Verteilung. . . . .	24
4.2. MH-SA <i>meanlog</i> der Lognormal-Verteilung. . . . .	25
4.3. Autokorrelation <i>meanlog</i> und <i>sdlog</i> . . . . .	26
4.4. MH Autokorrelation <i>meanlog</i> und <i>sdlog</i> verdünnt. . . . .	27
5.1. <i>meanlog</i> analytisch. Schritt 2. Int.: 35. Ext.: 300. . . . .	40
5.2. <i>sdlog</i> analytisch. Schritt 2. Int.: 35. Ext.: 300. . . . .	41
5.3. Log-log-Plot analytisch. Int.: 35. Ext.: 300. . . . .	42
5.4. <i>meanlog</i> MH. Schritt 1. Int.: 35. Ext.: 300. . . . .	44
5.5. <i>sdlog</i> MH. Schritt 1. Int.: 35. Ext.: 300. . . . .	45
5.6. ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 1. Int.: 35. Ext.: 300. . . . .	46
5.7. <i>meanlog</i> MH. Schritt 2. Int.: 35. Ext.: 300. . . . .	48
5.8. <i>sdlog</i> MH. Schritt 2. Int.: 35. Ext.: 300. . . . .	49
5.9. ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 2. Int.: 35. Ext.: 300. . . . .	50
5.10. Log-log-Plot MH. Int.: 35. Ext.: 300. . . . .	51
A.1. MH <i>meanlog</i> der Lognormal-Verteilung. 625 Iterationen. . . . .	70
A.2. MH-SA <i>meanlog</i> der Lognormal-Verteilung. 625 Iterationen. . . . .	71
A.3. MH <i>sdlog</i> der Lognormal-Verteilung. 625 Iterationen. . . . .	72
A.4. MH-SA <i>sdlog</i> der Lognormal-Verteilung. 625 Iterationen. . . . .	73
A.5. MH <i>meanlog</i> der Lognormal-Verteilung. 1250 Iterationen. . . . .	74
A.6. MH-SA <i>meanlog</i> der Lognormal-Verteilung. 1250 Iterationen. . . . .	75
A.7. MH <i>sdlog</i> der Lognormal-Verteilung. 1250 Iterationen. . . . .	76
A.8. MH-SA <i>sdlog</i> der Lognormal-Verteilung. 1250 Iterationen. . . . .	77
A.9. MH <i>meanlog</i> der Lognormal-Verteilung. 3125 Iterationen. . . . .	78
A.10. MH-SA <i>meanlog</i> der Lognormal-Verteilung. 3125 Iterationen. . . . .	79
A.11. MH <i>sdlog</i> der Lognormal-Verteilung. 3125 Iterationen. . . . .	80
A.12. MH-SA <i>sdlog</i> der Lognormal-Verteilung. 3125 Iterationen. . . . .	81
A.13. MH <i>meanlog</i> der Lognormal-Verteilung. 6250 Iterationen. . . . .	82
A.14. MH-SA <i>meanlog</i> der Lognormal-Verteilung. 6250 Iterationen. . . . .	83
A.15. MH <i>sdlog</i> der Lognormal-Verteilung. 6250 Iterationen. . . . .	84

A.16.MH-SA <i>sdlog</i> der Lognormal-Verteilung. 6250 Iterationen. . . . .	85
A.17.MH <i>meanlog</i> der Lognormal-Verteilung. 9375 Iterationen. . . . .	86
A.18.MH-SA <i>meanlog</i> der Lognormal-Verteilung. 9375 Iterationen. . . . .	87
A.19.MH <i>sdlog</i> der Lognormal-Verteilung. 9375 Iterationen. . . . .	88
A.20.MH-SA <i>sdlog</i> der Lognormal-Verteilung. 9375 Iterationen. . . . .	89
A.21.MH <i>meanlog</i> der Lognormal-Verteilung. 12500 Iterationen. . . . .	90
A.22.MH-SA <i>meanlog</i> der Lognormal-Verteilung. 12500 Iterationen. . . . .	91
A.23.MH <i>sdlog</i> der Lognormal-Verteilung. 12500 Iterationen. . . . .	92
A.24.MH-SA <i>sdlog</i> der Lognormal-Verteilung. 12500 Iterationen. . . . .	93
A.25.MH Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 500 Iterationen. . . . .	94
A.26.MH-SA Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 500 Iterationen. . . . .	95
A.27.MH Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 1000 Iterationen. . . . .	96
A.28.MH-SA Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 1000 Iterationen. . . . .	97
A.29.MH Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 2500 Iterationen. . . . .	98
A.30.MH-SA Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 2500 Iterationen. . . . .	99
A.31.MH Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 5000 Iterationen. . . . .	100
A.32.MH-SA Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 5000 Iterationen. . . . .	101
A.33.MH Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 7500 Iterationen. . . . .	102
A.34.MH-SA Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 7500 Iterationen. . . . .	103
A.35.MH Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 10000 Iterationen. . . . .	104
A.36.MH-SA Autokorrelation <i>meanlog</i> und <i>sdlog</i> . 10000 Iterationen. . . . .	105
A.37. <i>meanlog</i> MH. Schritt 1. Int.: 100. Ext.: 1350. . . . .	119
A.38. <i>sdlog</i> MH. Schritt 1. Int.: 100. Ext.: 1350. . . . .	120
A.39.ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 1. Int.: 100. Ext.: 1350. . . . .	121
A.40. <i>meanlog</i> MH. Schritt 2. Int.: 100. Ext.: 1350. . . . .	122
A.41. <i>sdlog</i> MH. Schritt 2. Int.: 100. Ext.: 1350. . . . .	123
A.42.ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 2. Int.: 100. Ext.: 1350. . . . .	124
A.43.Log-log-Plot MH. Int.: 100. Ext.: 1350. . . . .	125
A.44. <i>meanlog</i> analytisch. Schritt 2. Int.: 100. Ext.: 1350. . . . .	126
A.45. <i>sdlog</i> analytisch. Schritt 2. Int.: 100. Ext.: 1350. . . . .	127
A.46.Log-log-Plot analytisch. Int.: 100. Ext.: 1350. . . . .	128
A.47. <i>meanlog</i> MH. Schritt 1. Int.: 300. Ext.: 3000. . . . .	129
A.48. <i>sdlog</i> MH. Schritt 1. Int.: 300. Ext.: 3000. . . . .	130
A.49.ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 1. Int.: 300. Ext.: 3000. . . . .	131
A.50. <i>meanlog</i> MH. Schritt 2. Int.: 300. Ext.: 3000. . . . .	132
A.51. <i>sdlog</i> MH. Schritt 2. Int.: 300. Ext.: 3000. . . . .	133
A.52.ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 2. Int.: 300. Ext.: 3000. . . . .	134
A.53.Log-log-Plot MH. Int.: 300. Ext.: 3000. . . . .	135
A.54. <i>meanlog</i> analytisch. Schritt 2. Int.: 300. Ext.: 3000. . . . .	136
A.55. <i>sdlog</i> analytisch. Schritt 2. Int.: 300. Ext.: 3000. . . . .	137
A.56.Log-log-Plot analytisch. Int.: 300. Ext.: 3000. . . . .	138
A.57. <i>meanlog</i> MH. Schritt 1. Int.: 100. Ext.: 300. . . . .	139
A.58. <i>sdlog</i> MH. Schritt 1. Int.: 100. Ext.: 300. . . . .	140
A.59.ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 1. Int.: 100. Ext.: 300. . . . .	141

A.60.	<i>meanlog</i> MH. Schritt 2. Int.: 100. Ext.: 300.	142
A.61.	<i>sdlog</i> MH. Schritt 2. Int.: 100. Ext.: 300.	143
A.62.	ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 2. Int.: 100. Ext.: 300.	144
A.63.	Log-log-Plot MH. Int.: 100. Ext.: 300.	145
A.64.	<i>meanlog</i> analytisch. Schritt 2. Int.: 100. Ext.: 300.	146
A.65.	<i>sdlog</i> analytisch. Schritt 2. Int.: 100. Ext.: 300.	147
A.66.	Log-log-Plot analytisch. Int.: 100. Ext.: 300.	148
A.67.	<i>meanlog</i> MH. Schritt 1. Int.: 200. Ext.: 300.	149
A.68.	<i>sdlog</i> MH. Schritt 1. Int.: 200. Ext.: 300.	150
A.69.	ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 1. Int.: 200. Ext.: 300.	151
A.70.	<i>meanlog</i> MH. Schritt 2. Int.: 200. Ext.: 300.	152
A.71.	<i>sdlog</i> MH. Schritt 2. Int.: 200. Ext.: 300.	153
A.72.	ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 2. Int.: 200. Ext.: 300.	154
A.73.	Log-log-Plot MH. Int.: 200. Ext.: 300.	155
A.74.	<i>meanlog</i> analytisch. Schritt 2. Int.: 200. Ext.: 300.	156
A.75.	<i>sdlog</i> analytisch. Schritt 2. Int.: 200. Ext.: 300.	157
A.76.	Log-log-Plot analytisch. Int.: 200. Ext.: 300.	158
A.77.	<i>meanlog</i> MH. Schritt 1. Int.: 300. Ext.: 300.	159
A.78.	<i>sdlog</i> MH. Schritt 1. Int.: 300. Ext.: 300.	160
A.79.	ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 1. Int.: 300. Ext.: 300.	161
A.80.	<i>meanlog</i> MH. Schritt 2. Int.: 300. Ext.: 300.	162
A.81.	<i>sdlog</i> MH. Schritt 2. Int.: 300. Ext.: 300.	163
A.82.	ACF <i>meanlog</i> und <i>sdlog</i> MH. Schritt 2. Int.: 300. Ext.: 300.	164
A.83.	Log-log-Plot MH. Int.: 300. Ext.: 300.	165
A.84.	<i>meanlog</i> analytisch. Schritt 2. Int.: 300. Ext.: 300.	166
A.85.	<i>sdlog</i> analytisch. Schritt 2. Int.: 300. Ext.: 300.	167
A.86.	Log-log-Plot analytisch. Int.: 300. Ext.: 300.	168

## Tabellenverzeichnis

1.1. Beispiel Segmentbildung. . . . .	2
4.1. MH-Algorithmus. . . . .	16
4.2. Gibbs-Sampler. . . . .	16
4.3. Simulated Annealing. . . . .	17
4.4. Tab. MH-Algorithmus 2 Parameter. . . . .	21
4.5. Tab. MH-SA-Algorithmus 2 Parameter. . . . .	23
5.1. Zwei-Schritt-Bayes-Verfahren. . . . .	30
5.2. Die wahren Parameter für Simulation der Daten. . . . .	34
5.3. Schätzervergleich Poisson-, Gamma-Verteilung. . . . .	35
5.4. Die wahren Parameter für die Simulation der Daten. . . . .	39
5.5. Die geschätzten Parameter für sechs Datensätze. . . . .	53
5.6. VaR <sub>99,95</sub> für sechs Datensätze. . . . .	54
A.1. Konjugierte <i>a priori</i> . . . . .	106
B.1. Benennung des Outputs. . . . .	171

## Literaturverzeichnis

- Bank für Internationalen Zahlungsausgleich 2003. Basler Ausschuss für Bankenaufsicht: Management operationeller Risiken - Praxisempfehlungen für Banken und Bankenaufsicht.  
**URL:** <https://www.bis.org/publ/bcbs96de.pdf>
- Berger, J. O. 1980. *Statistical decision theory and Bayesian analysis.*, Springer series in statistics, 1. ed. edn, Springer, New York.
- Bolance, C., Nielsen, J. P., Guillen, M. and Gustafsson, J. 2012. *Quantitative operational risk models*, Chapman & Hall/CRC finance series, Taylor & Francis, Boca Raton.
- Chib, S. and Greenberg, E. 1995. Understanding the metropolis-hastings algorithm., *The American Statistician* **49**(4): 327–335.
- Cruz, M. G., Peters, G. W. and Shevchenko, P. V. 2015. *Fundamental aspects of operational risk and insurance analytics: A handbook of operational risk.*, Wiley handbooks in financial engineering and econometrics, Wiley, Hoboken, NJ.
- Dahlberg, E. A. 2015. *Bayesian Inference Methods in Operational Risk.*, Master's thesis in mathematical statistics, School of Engineering Sciences, Stockholm.
- Duong, T. 2004. *Bandwidth selectors for multivariate kernel density estimation.*, PhD thesis, University of Western Australia, Australia.
- Duong, T. 2018. The r package ks: Kernel smoothing.  
**URL:** <https://CRAN.R-project.org/package=ks>
- Duong, T. and Hazelton, M. L. 2005. Cross-validation bandwidth matrices for multivariate kernel density estimation., *Scandinavian Journal of Statistics* **32**(3): 485–506.
- Ergashev, B. 2008. Should risk managers rely on maximum likelihood estimation method while quantifying operational risk?
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I. and Tutz, G. 2016. *Statistik: Der Weg zur Datenanalyse.*, Springer-Lehrbuch, 8., überarb. u. erg. Aufl. 2017 edn, Springer Berlin, Berlin and S.l. and Berlin and S.l.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. 2013. *Bayesian Data Analysis.*, Chapman & Hall / CRC Texts in Statistical Science, 3rd ed. edn, CRC Press, Hoboken.

- Geman, S. and Geman, D. 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images., *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(6): 721–741.
- Gilbert, Paul and Varadhan, Ravi 2016. The r package numderiv: Accurate numerical derivatives.  
**URL:** <http://optimizer.r-forge.r-project.org/>
- Hansen, B. E. 2009. *Lecture Notes on Nonparametrics: Density Estimation.*, Lecture notes, University of Wisconsin, Wisconsin-Madison.
- Hassani, B. and Renaudin, A. 2013. *The Cascade Bayesian Approach for a controlled integration of internal data, external data and scenarios.*, PhD thesis, Panthéon-Sorbonne University, Paris.
- Hasselmann, B. 2018. The r package nleqslv: Solve systems of nonlinear equations.  
**URL:** <https://CRAN.R-project.org/package=nleqslv>
- Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications., *Biometrika* **57**(1): 97–109.
- Heller, W.-D., Lindenberg, H. and Nuske, M. 1978. *Stochastische Systeme: Markoffketten - Stochastische Prozesse - Warteschlangen.*, De Gruyter Lehrbuch, WALTER DE GRUYTER, Hawthorne.
- Husmann, K., Lange, A. and Spiegel, E. 2017. The r package optimization: Flexible global optimization with simulated-annealing.
- Jeffreys, H. 1946. An invariant form for the prior probability in estimation problems., *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **186**(1007): 453–461.
- Jones, M. C., Marron, J. S. and Park, B. U. 1991. A simple root n bandwidth selector., *The Annals of Statistics* **19**(4): 1919–1932.
- Kersting, G. and Wakolbinger, A. 2014. *Stochastische Prozesse.*, Mathematik Kompakt, Springer, Basel [u.a.].
- Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. 1983. Optimization by simulated annealing., *Science (New York, N.Y.)* **220**(4598): 671–680.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. 1953. Equation of state calculations by fast computing machines., *The Journal of Chemical Physics* **21**(6): 1087–1092.
- Nedden, M. 2012. *Neue Methoden zur Charakterisierung der QSAR-Anwendungsdomäne: Modifizierte Kerndichteschätzung der Vorhersagegüte empirisch abgeleiteter Modelle in der Chemie.*, Springer Research, Springer, Berlin [u.a.].

- Neumann, M. 2013. Monte Carlo-Verfahren. Vorlesung: Computational Physics I: Grundlagen.
- R Core Team 2017. R: A language and environment for statistical computing.  
**URL:** <https://www.R-project.org/>
- Robert, C. P. 2007. *The Bayesian choice: From decision-theoretic foundations to computational implementation.*, 2nd ed. edn, Springer, New York.
- Schmid, V. and Happ, K. 2017. Einführung in die Bayes-Statistik. Vorlesung.
- Schomaker, J. 2012. Einführung in die Theorie der Markov-Ketten. Seminar: Anwendungen der Wahrscheinlichkeitstheorie.
- Shevchenko, P. V. and Peters, G. W. 2013. Loss Distribution Approach for Operational Risk Capital Modelling Under Basel II: Combining Different Data Sources for Risk Estimation., *SSRN Electronic Journal* .
- Shevchenko, P. and Wüthrich, M. 2006. The structural modeling of operational risk via bayesian inference: Combining loss data with expert opinions., *The Journal of Operational Risk* 1(3): 3–26.
- Torresetti, R. and Nordio, C. 2014. Scaling operational loss data and its systemic risk implications., *SSRN Electronic Journal* .
- Uhlmann, W. 1965. Einführung in die Monte-Carlo-Methoden., *Biometrische Zeitschrift* 7(2): 116–119.
- Wang, v. H. 2009. *Multivariate Dichteschätzung in der explorativen Datenanalyse.*, Dissertation, Universität Augsburg, Augsburg.
- Wegener, I. 2004. Simulated annealing beats metropolis in combinatorial optimization., *Electronic Colloquium on Computational Complexity* **26. October**(89).

## A. Anhang

### A.1. Kernfunktionen

Beispiele für univariate Kernfunktionen für die klassische Kerndichte-Schätzung<sup>1</sup>.

- **Rechtseckkern:**

$$K(x) = \begin{cases} \frac{1}{2}, & \text{wenn } |x| < 1 \\ 0, & \text{sonst} \end{cases}$$

- **Dreieckskern:**

$$K(x) = \begin{cases} 1 - |x|, & \text{wenn } |x| < 1 \\ 0, & \text{sonst} \end{cases}$$

- **Gauß-Kern:**

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-0.5x^2)$$

- **Epanechnikov-Kern:**

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2), & \text{wenn } |x| < 1 \\ 0, & \text{sonst} \end{cases}$$

Äquivalenzform (wird oft in R implementierten Funktionen verwendet)

$$K(x) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{1}{5}x^2), & \text{wenn } |x| < \sqrt{5} \\ 0, & \text{sonst} \end{cases}$$

Beispiel für  $d$ -variate Kernfunktion für klassische Kerndichte-Schätzung<sup>2</sup>

- **Gauß-Kern  $d$ -variater:**

$$K(\underline{x}) = \frac{1}{2\pi} \exp(-\frac{1}{2}\underline{x}^T \underline{x})$$

---

<sup>1</sup>Beispiele übernommen aus [Nedden 2012](#): S. 119

<sup>2</sup>Kernfunktion übernommen aus [Duong 2018](#): S. 1.

## A.2. Bestimmung der Parameter für Normal- und Lognormal-Verteilung in Abhängigkeit von Modus und Krümmung

Herleitungen zur Kap. 4.7.1

### A.2.1. Bestimmung der Normal-Verteilung Parameter

Dichte-Funktion:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[x - \mu]^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

Modus:  $Modus = \mu$

Erste Ableitung nach  $x$ :

$$f'(x) = -\frac{(x - \mu)}{\sigma^3\sqrt{2\pi}} \exp\left(-\frac{[x - \mu]^2}{2\sigma^2}\right) \stackrel{x=\mu}{=} 0$$

Zweite Ableitung nach  $x$ :

$$f''(x) = -\frac{(\mu + \sigma - x)(-\mu + \sigma + x)}{\sigma^5\sqrt{2\pi}} \exp\left(-\frac{[x - \mu]^2}{2\sigma^2}\right) \stackrel{x=\mu}{=} -\frac{1}{\sigma^3\sqrt{2\pi}}$$

Krümmung im Modus-Punkt  $x = \mu$ :

$$k(\mu) = \frac{|f''(\mu)|}{(1 + (f'(\mu))^2)^{\frac{3}{2}}} = \frac{1}{\sigma^3\sqrt{2\pi}}$$

$$\Leftrightarrow$$

$$\sigma = \left(\frac{1}{k\sqrt{2\pi}}\right)^{\frac{1}{3}}$$

### A.2.2. Bestimmung der Lognormal-Verteilung Parameter

Dichte-Funktion:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{[\log(x) - \mu]^2}{2\sigma^2}\right), \quad x > 0$$

Modus:  $Modus = \exp(\mu - \sigma^2) \Leftrightarrow \mu = \log(Modus) + \sigma^2$

Erste Ableitung nach  $x$ :

$$f'(x) = -\frac{(\sigma^2 + \log(x) - \mu)}{x^2\sigma^3\sqrt{2\pi}} \exp\left(-\frac{[\log(x) - \mu]^2}{2\sigma^2}\right) \stackrel{x=\exp(\mu-\sigma^2)}{=} 0$$

Zweite Ableitung nach  $x$ :

$$f''(x) = \frac{(\mu^2 + \log(x)[3\sigma^2 + \log(x) - 2\mu] - \sigma^2[3\mu + 1] + 2\sigma^4)}{x^3\sigma^5\sqrt{2\pi}} \exp\left(-\frac{[\log(x) - \mu]^2}{2\sigma^2}\right)$$

$$x = \exp(\mu - \sigma^2) = \frac{\exp(\frac{1}{2})}{[\exp(\mu - \sigma^2)]^3 \sigma^3 \sqrt{2\pi}}$$

**Krümmung im Modus-Punkt**  $x = \exp(\mu - \sigma^2)$ :

$$k(\exp(\mu - \sigma^2)) = \frac{|f''(\exp(\mu - \sigma^2))|}{(1 + (f'(\exp(\mu - \sigma^2)))^2)^{\frac{3}{2}}} = \frac{\exp(\frac{1}{2})}{[\exp(\mu - \sigma^2)]^3 \sigma^3 \sqrt{2\pi}}$$

$$\mu = \log(\text{Modus}) + \sigma^2 \quad \frac{\exp(\frac{1}{2})}{\text{Modus}^3 \sigma^3 \sqrt{2\pi}}$$

$\Leftrightarrow$

$$\sigma = \left( \frac{\exp(\frac{1}{2})}{k \text{Modus}^3 \sqrt{2\pi}} \right)^{\frac{1}{3}}$$

## A.3. MH- und MH-SA-Algorithmus im Vergleich

### A.3.1. Markov-Kette

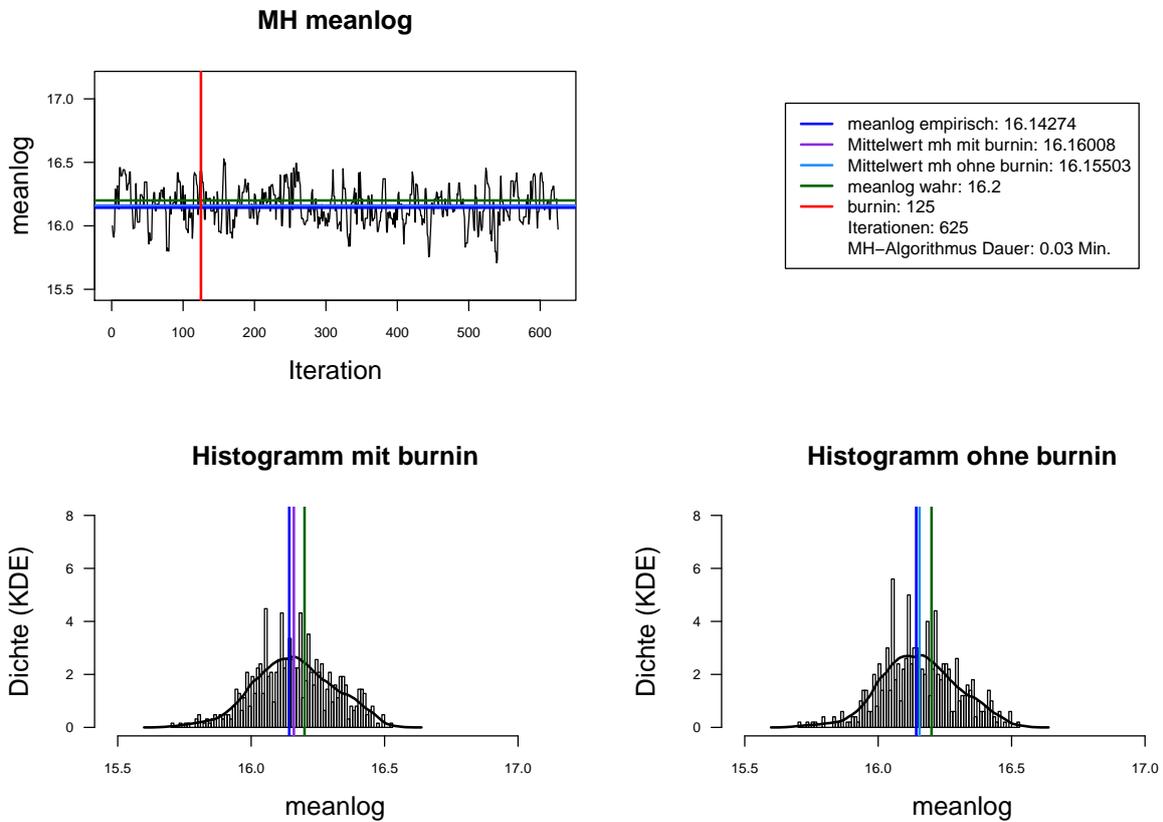


Abbildung A.1.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 625 Iterationen beträgt 0.03 Minuten. Markov-Kette (oben links), Histogramm für 625 Iterationen mit KDE (unten links) und Histogramm für letzte 500 Iterationen mit KDE (unten rechts).

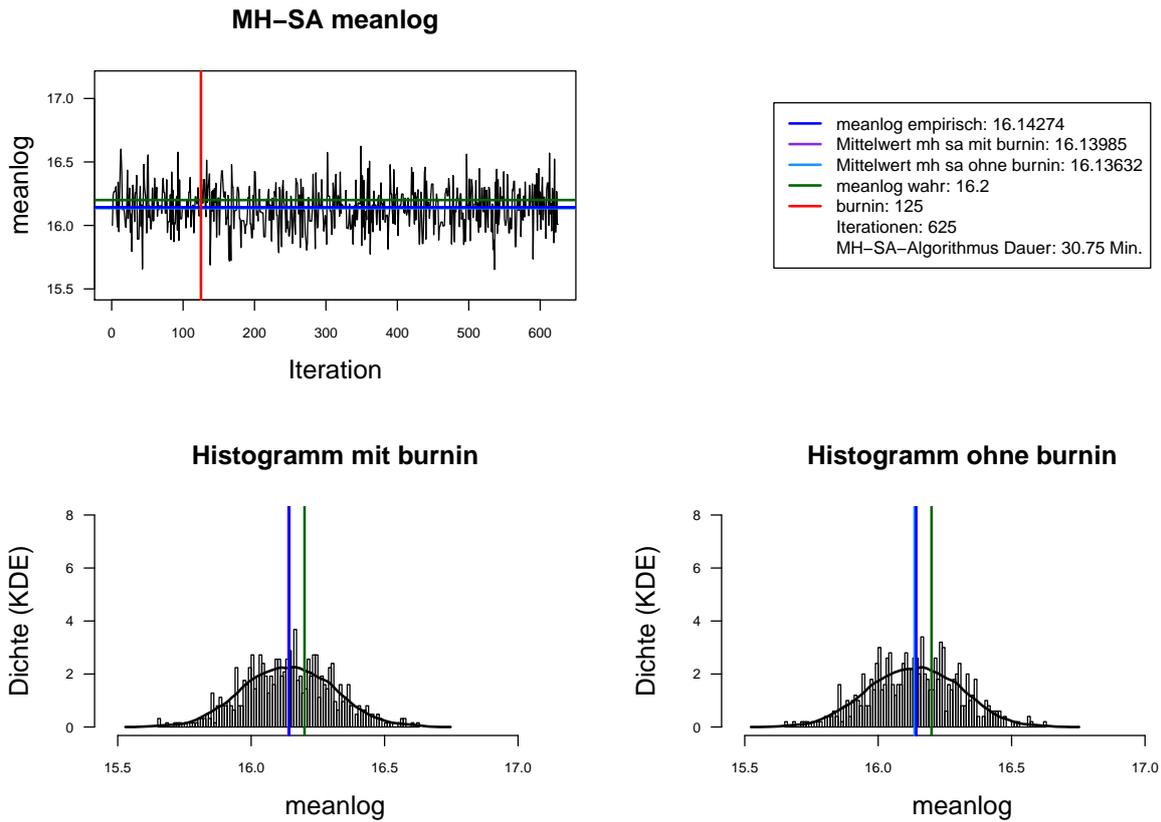


Abbildung A.2.: Der mit MH-SA-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 625 Iterationen beträgt 30.75 Minuten. Markov-Kette (oben links), Histogramm für 625 Iterationen mit KDE (unten links) und Histogramm für letzte 500 Iterationen mit KDE (unten rechts).

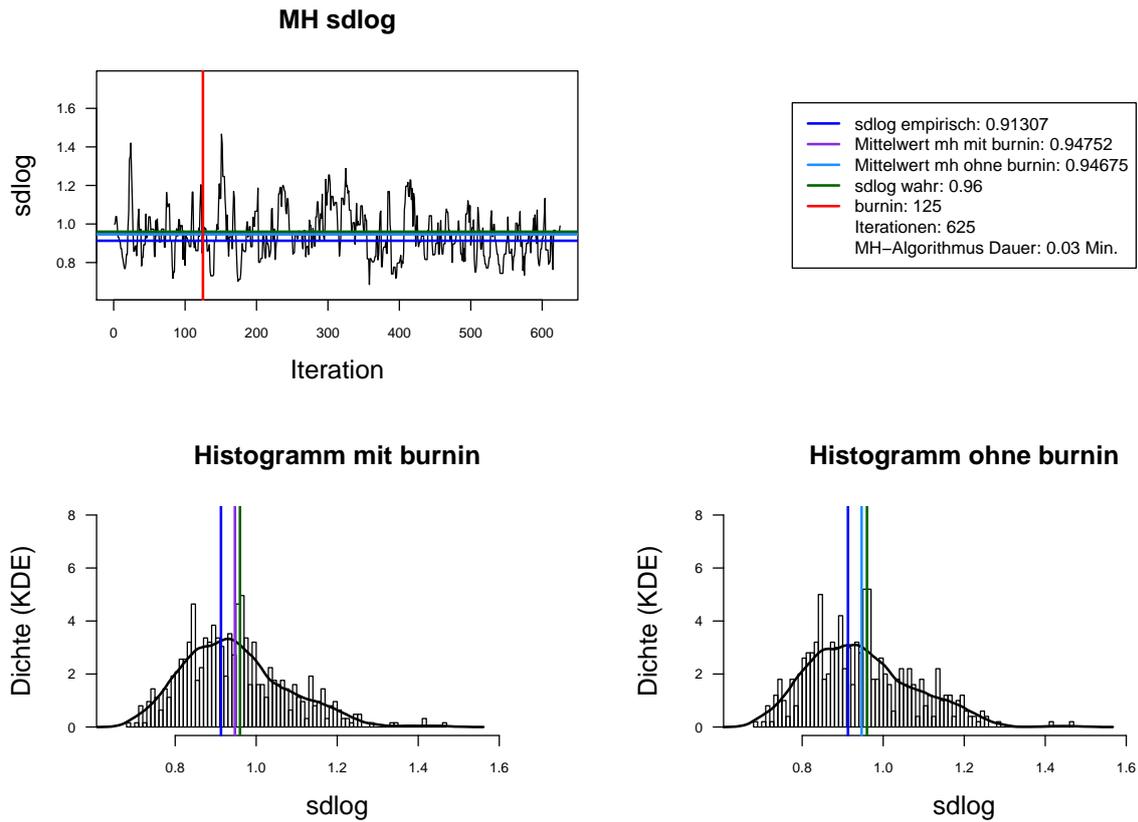


Abbildung A.3.: Der mit MH-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 625 Iterationen beträgt 0.03 Minuten. Markov-Kette (oben links), Histogramm für 625 Iterationen mit KDE (unten links) und Histogramm für letzte 500 Iterationen mit KDE (unten rechts).

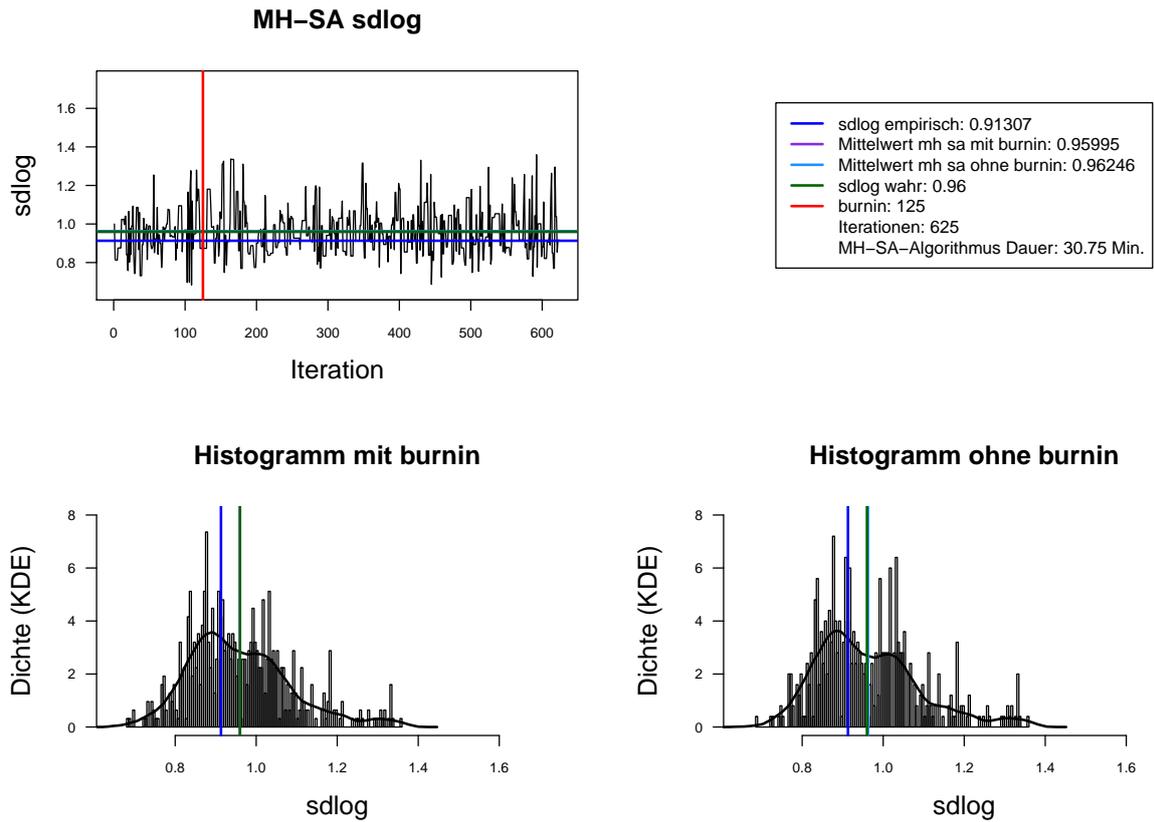


Abbildung A.4.: Der mit MH-SA-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 625 Iterationen beträgt 30.75 Minuten. Markov-Kette (oben links), Histogramm für 625 Iterationen mit KDE (unten links) und Histogramm für letzte 500 Iterationen mit KDE (unten rechts).

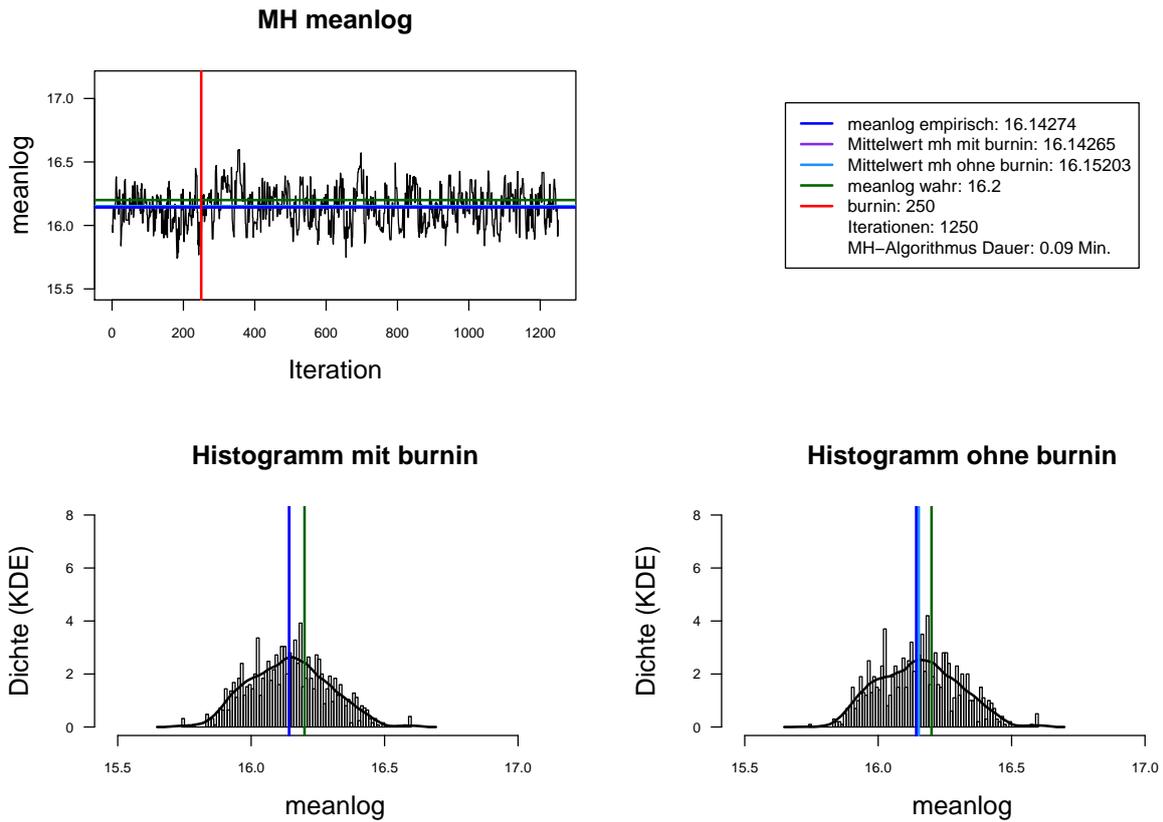


Abbildung A.5.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1250 Iterationen beträgt 0.09 Minuten. Markov-Kette (oben links), Histogramm für 1250 Iterationen mit KDE (unten links) und Histogramm für letzte 1000 Iterationen mit KDE (unten rechts).

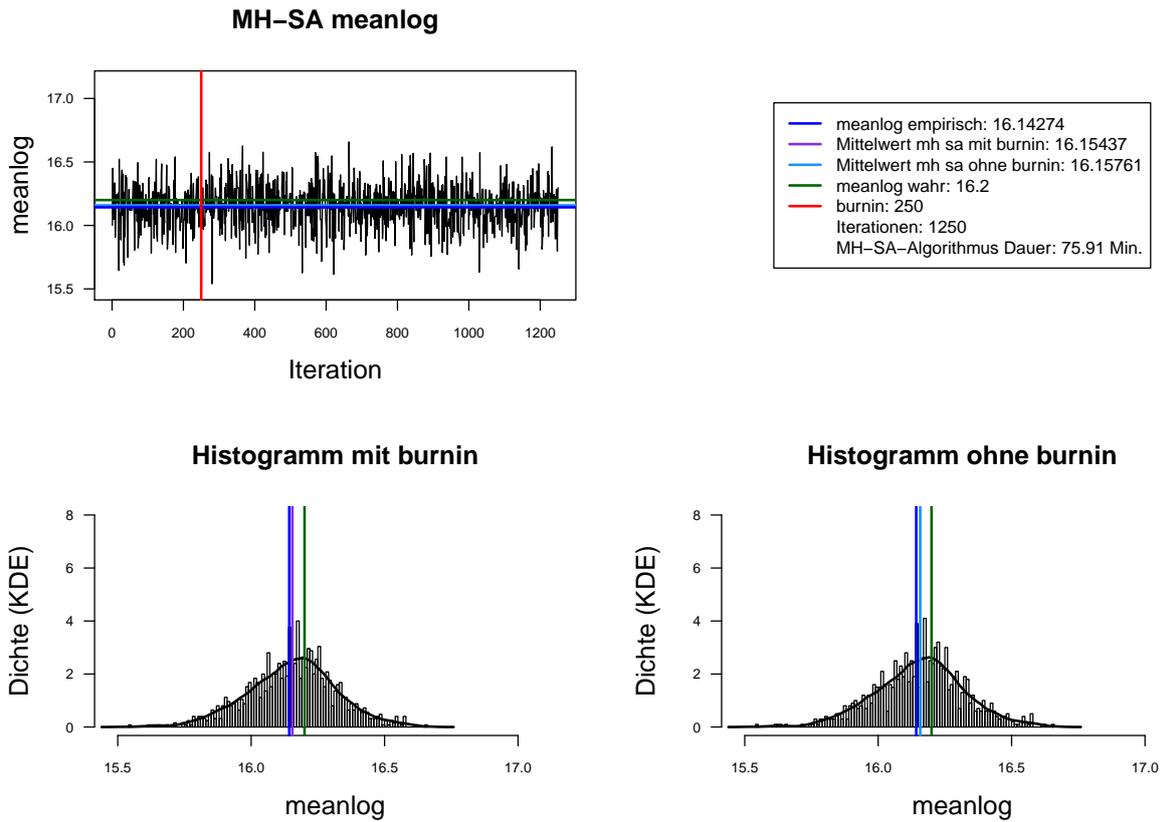


Abbildung A.6.: Der mit MH-SA-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1250 Iterationen beträgt 75.91 Minuten = 1.27 Stunden. Markov-Kette (oben links), Histogramm für 1250 Iterationen mit KDE (unten links) und Histogramm für letzte 1000 Iterationen mit KDE (unten rechts).

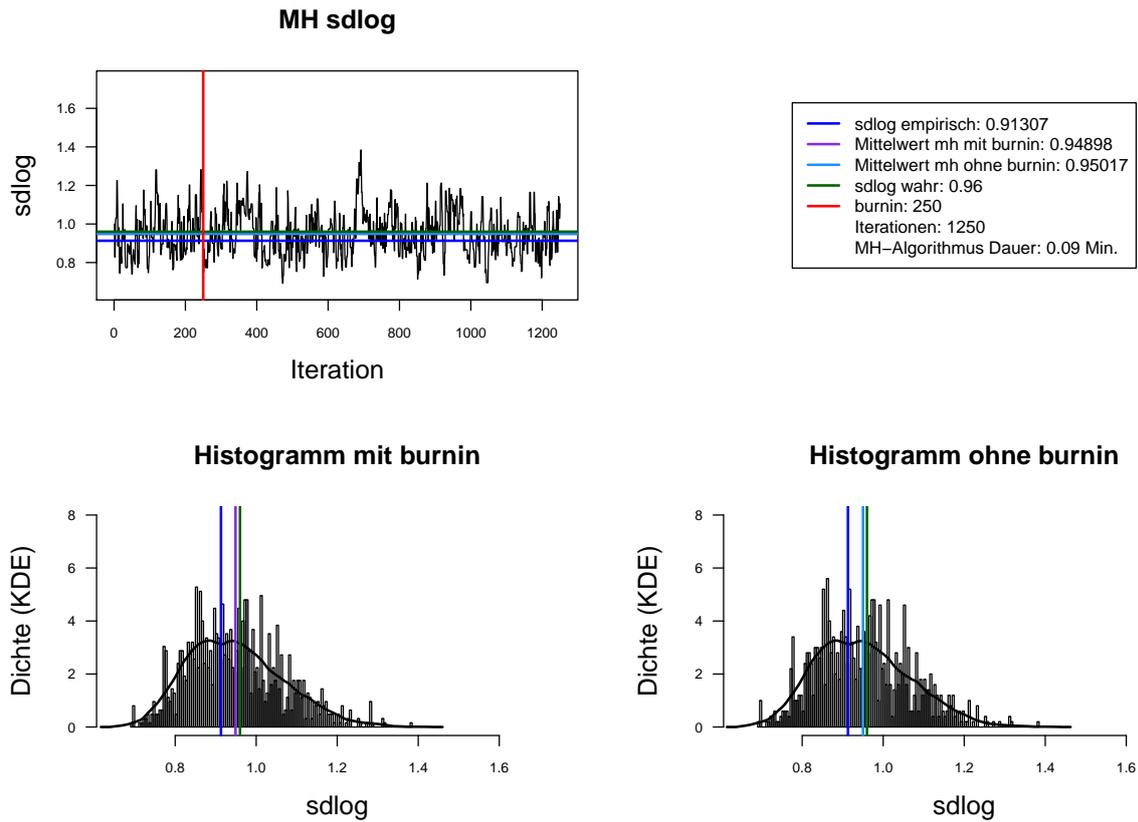


Abbildung A.7.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 1250 Iterationen beträgt 0.09 Minuten. Markov-Kette (oben links), Histogramm für 1250 Iterationen mit KDE (unten links) und Histogramm für letzte 1000 Iterationen mit KDE (unten rechts).

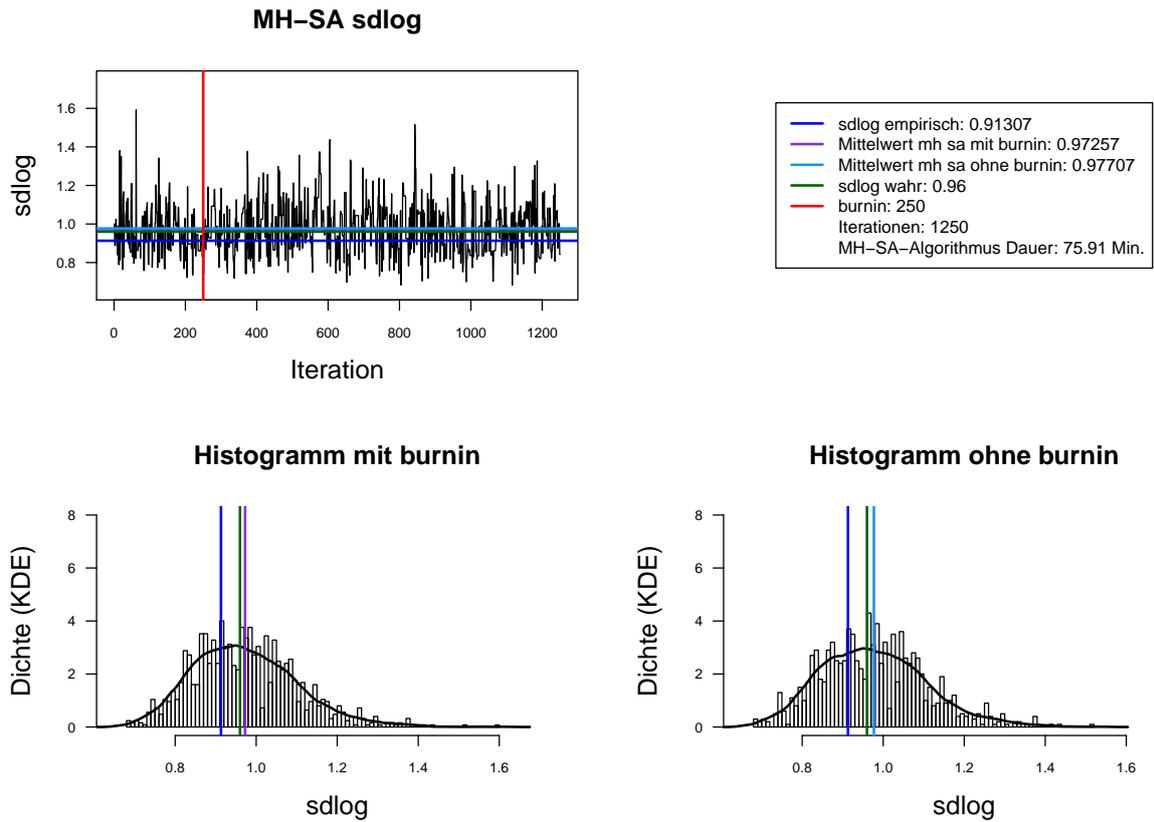


Abbildung A.8.: Der mit MH-SA-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 1250 Iterationen beträgt 75.91 Minuten = 1.27 Stunden. Markov-Kette (oben links), Histogramm für 1250 Iterationen mit KDE (unten links) und Histogramm für letzte 1000 Iterationen mit KDE (unten rechts).

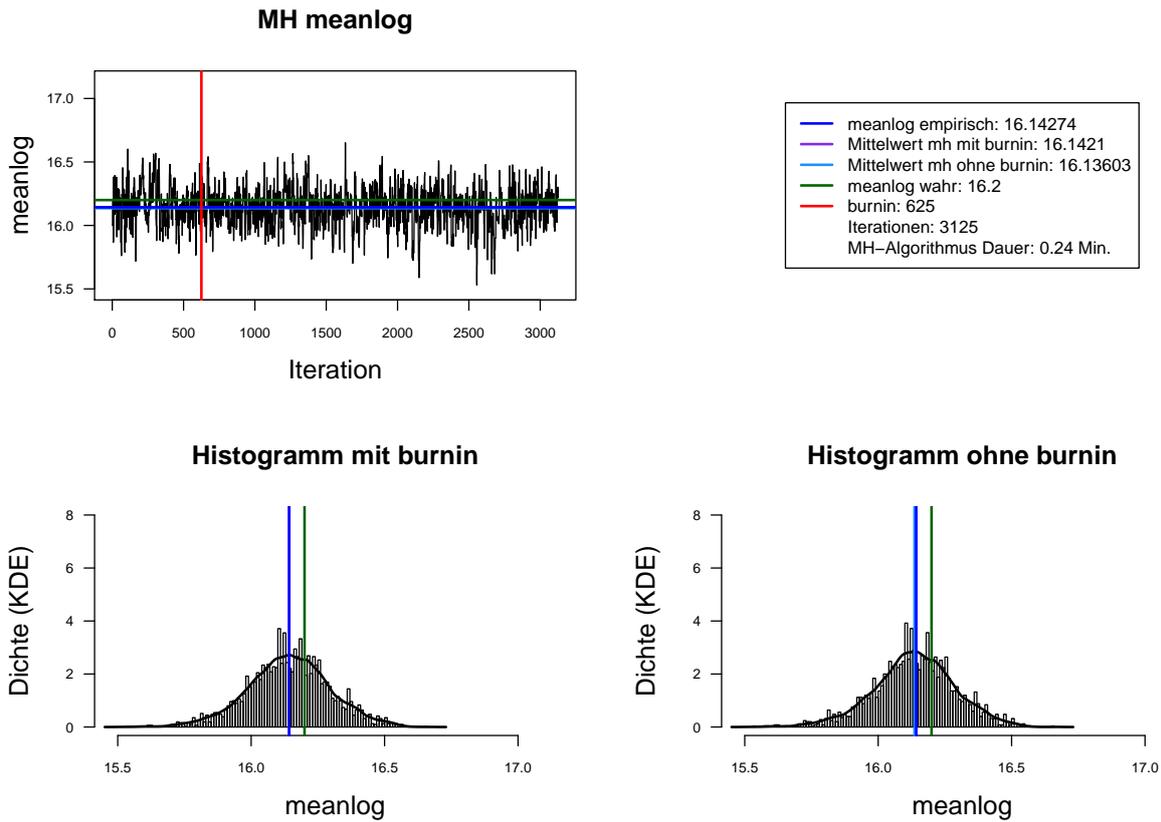


Abbildung A.9.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3125 Iterationen beträgt 0.24 Minuten. Markov-Kette (oben links), Histogramm für 3125 Iterationen mit KDE (unten links) und Histogramm für letzte 2500 Iterationen mit KDE (unten rechts).

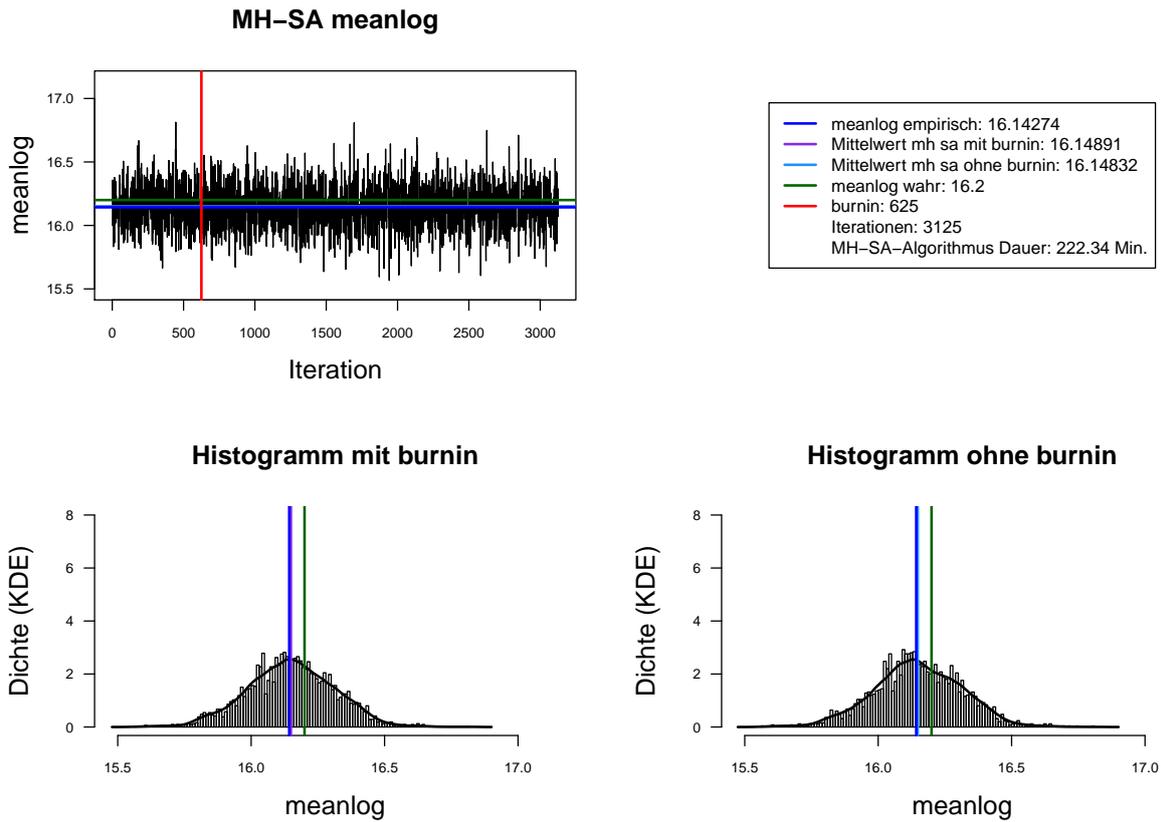


Abbildung A.10.: Der mit MH-SA-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3125 Iterationen beträgt 222.34 Minuten = 3.71 Stunden. Markov-Kette (oben links), Histogramm für 3125 Iterationen mit KDE (unten links) und Histogramm für letzte 2500 Iterationen mit KDE (unten rechts).

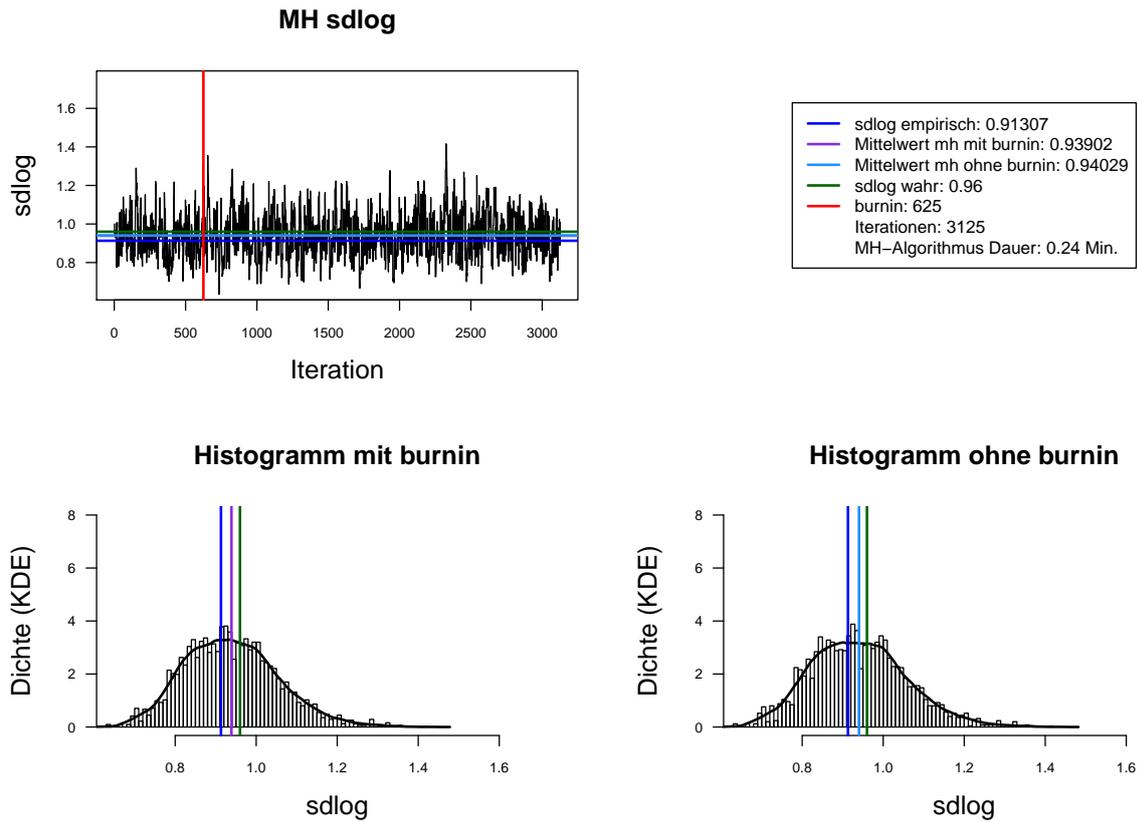


Abbildung A.11.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 3125 Iterationen beträgt 0.24 Minuten. Markov-Kette (oben links), Histogramm für 3125 Iterationen mit KDE (unten links) und Histogramm für letzte 2500 Iterationen mit KDE (unten rechts).

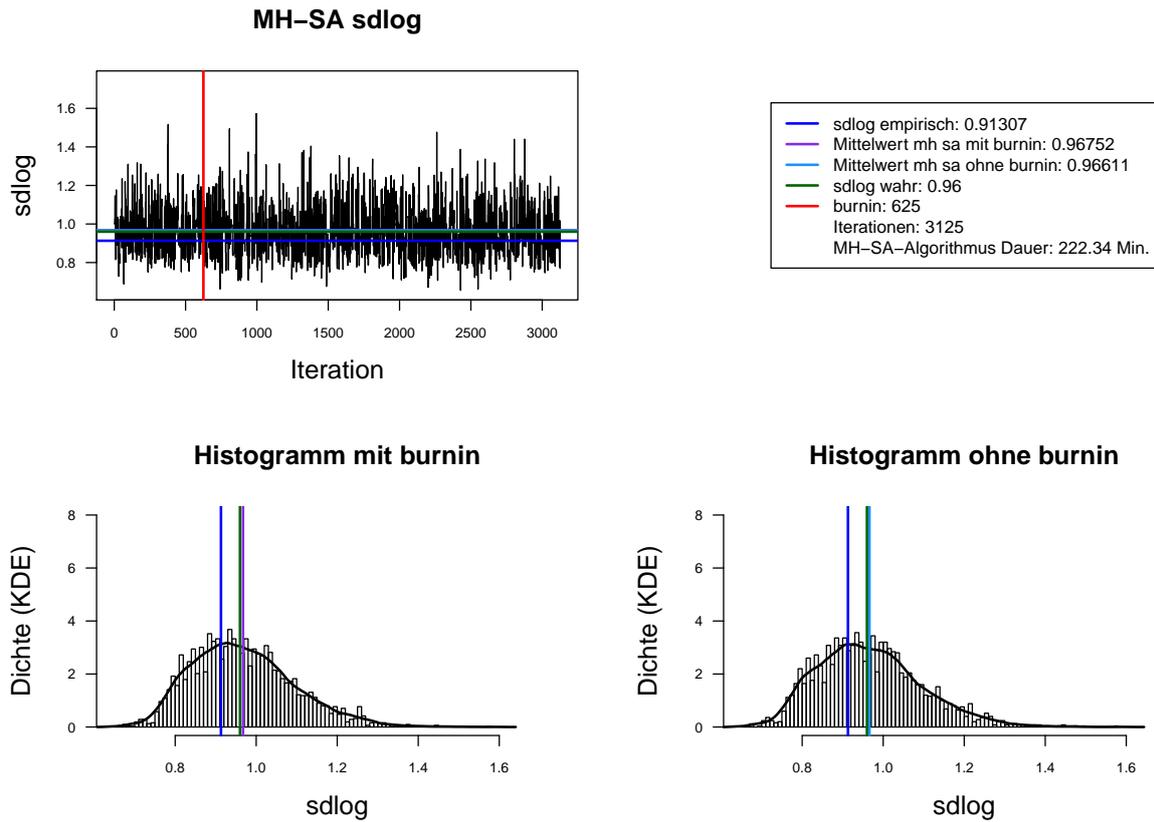


Abbildung A.12.: Der mit MH-SA-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 3125 Iterationen beträgt 222.34 Minuten = 3.71 Stunden. Markov-Kette (oben links), Histogramm für 3125 Iterationen mit KDE (unten links) und Histogramm für letzte 2500 Iterationen mit KDE (unten rechts).

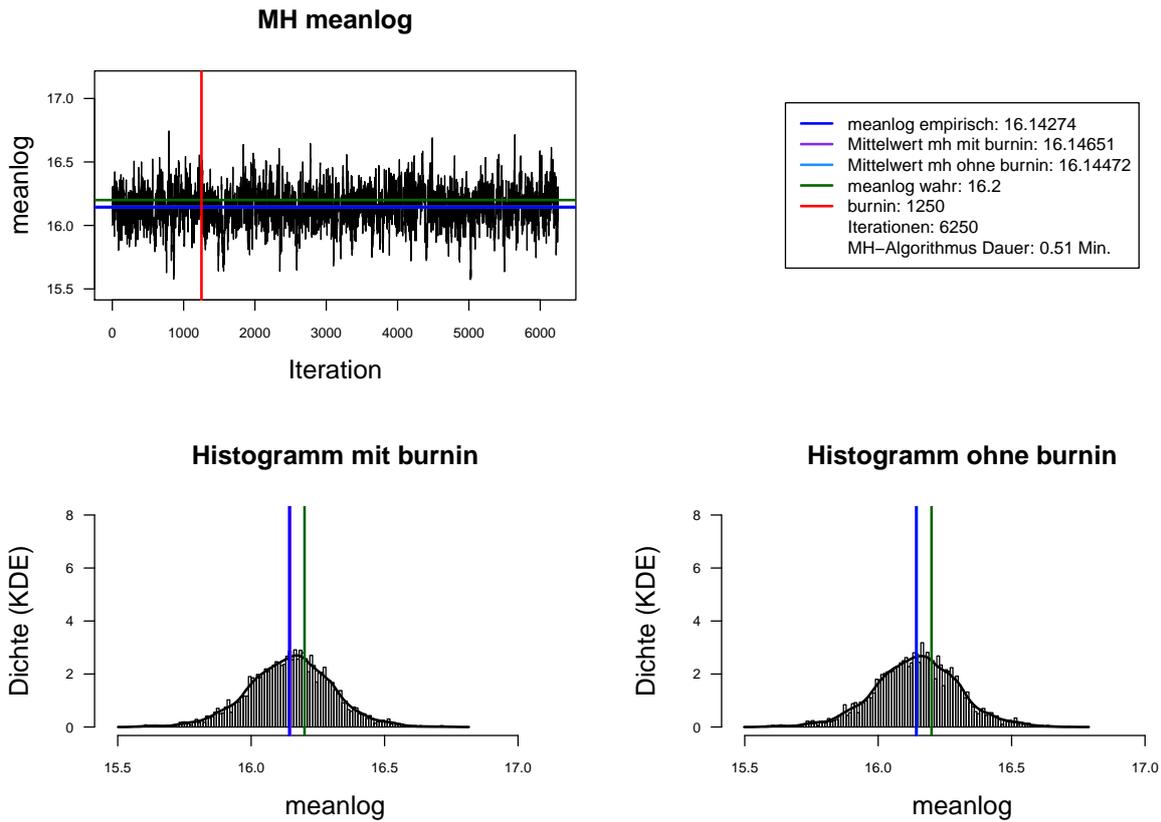


Abbildung A.13.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 0.51 Minuten. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts).

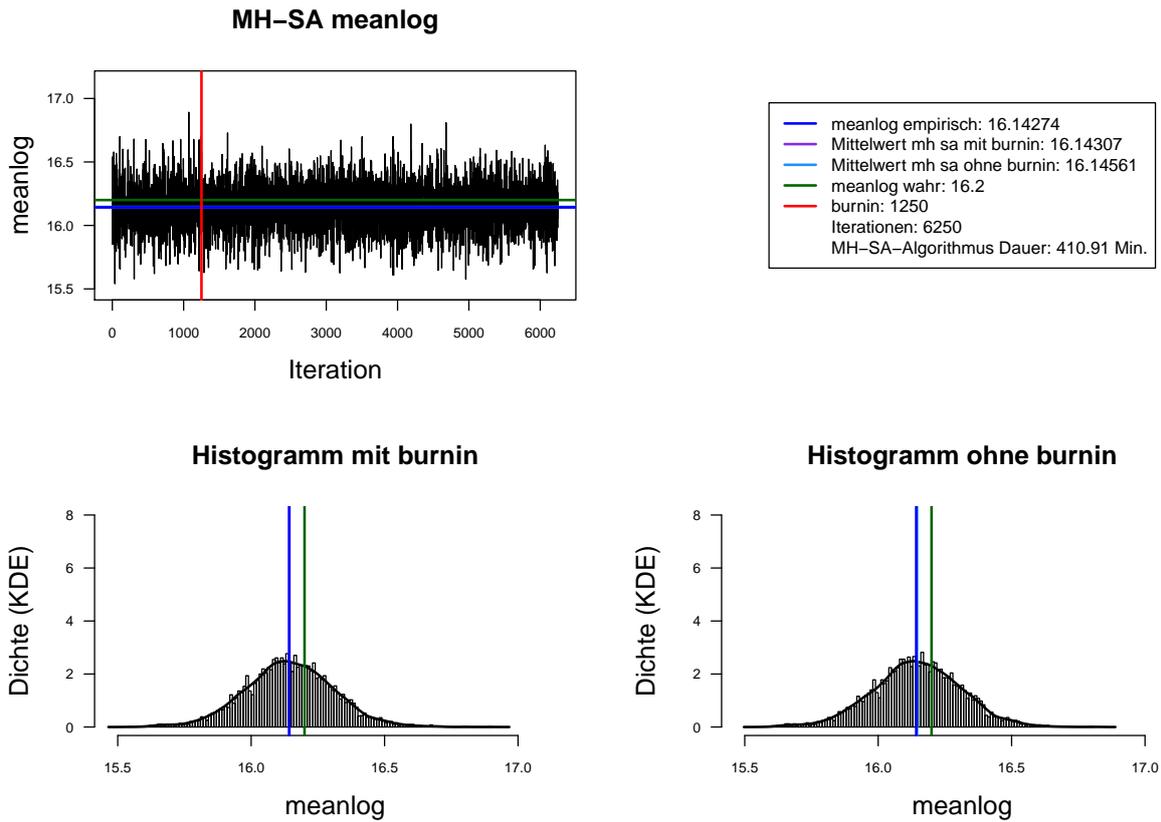


Abbildung A.14.: Der mit MH-SA-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 410.91 Minuten = 6.85 Stunden. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts).

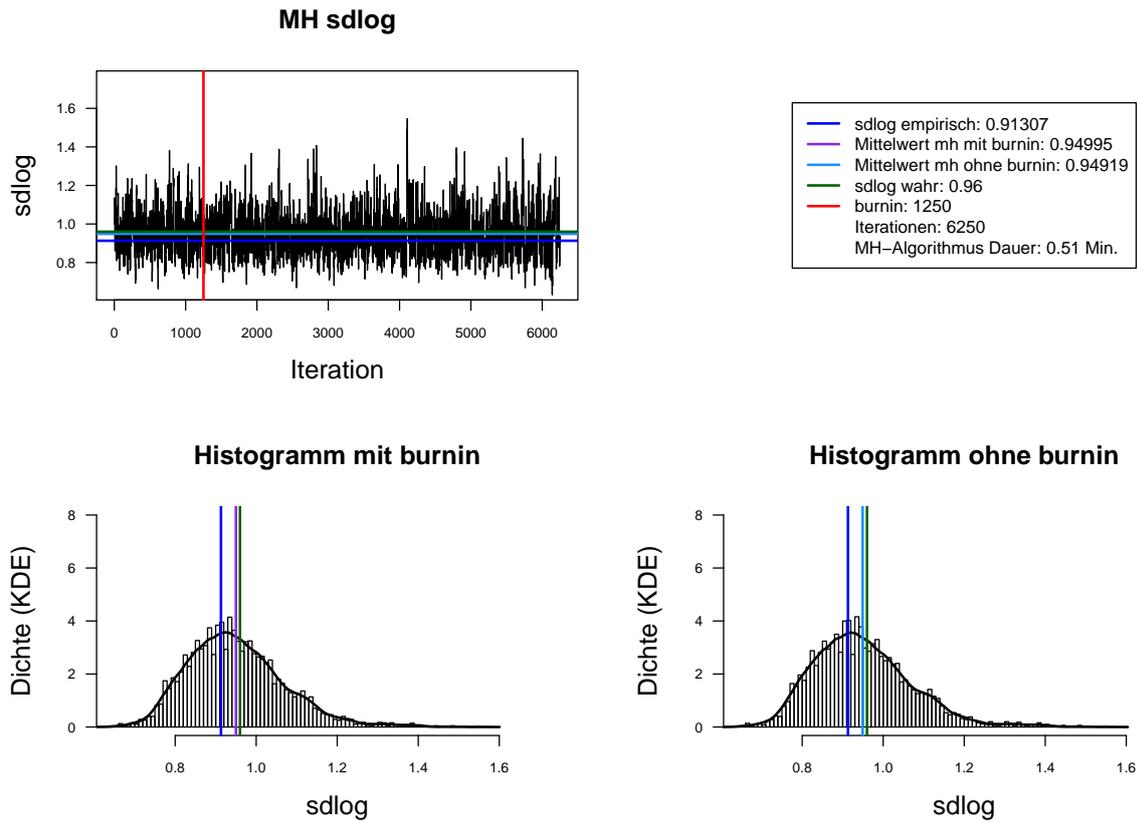


Abbildung A.15.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 0.51 Minuten. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts).

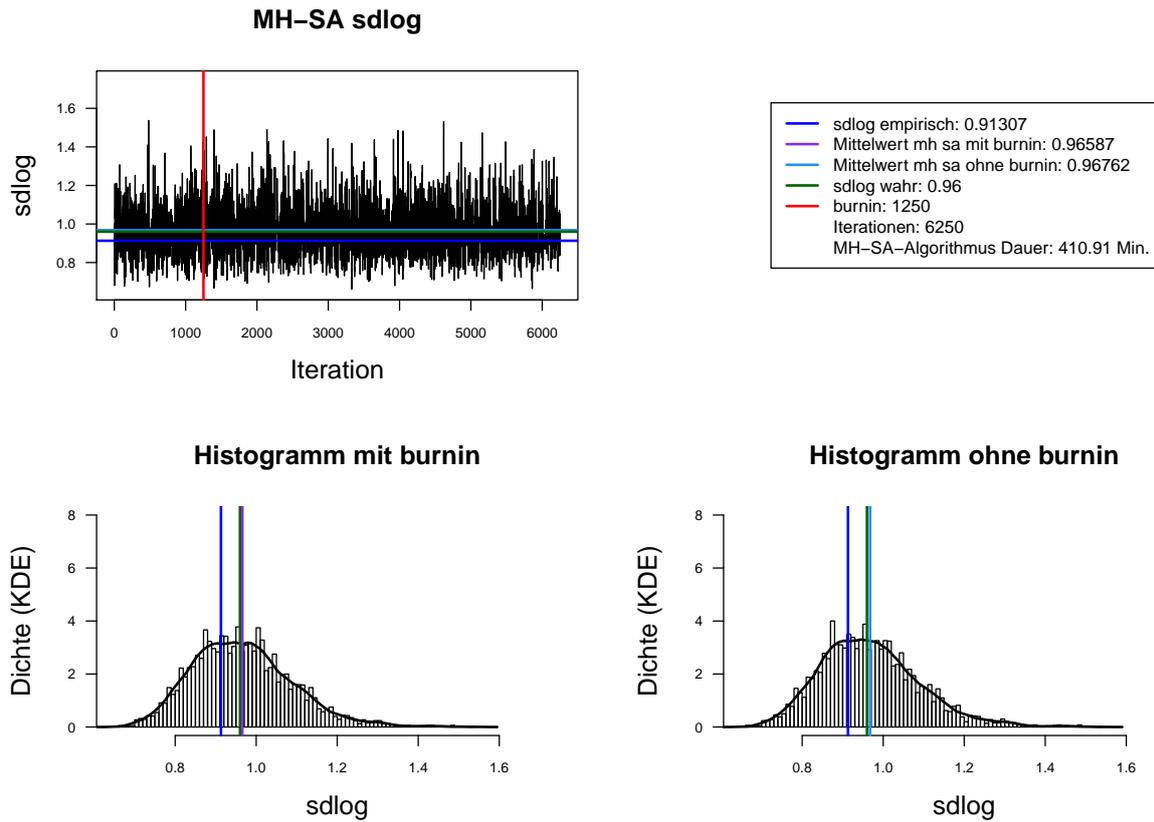


Abbildung A.16.: Der mit MH-SA-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 410.91 Minuten = 6.85 Stunden. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts).

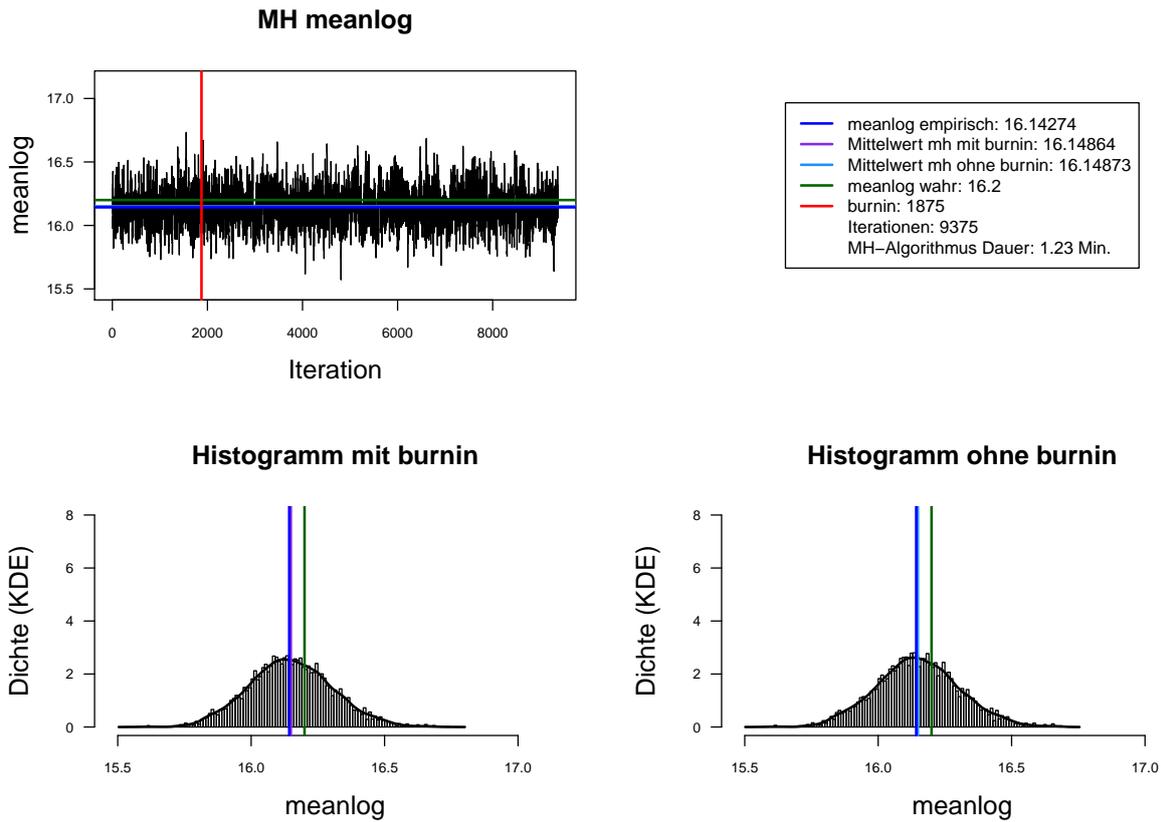


Abbildung A.17.: Der mit MH-Algorithmus simulierte  $\text{meanlog}$  der Lognormal-Verteilung. Dauer des Algorithmus mit 9375 Iterationen beträgt 1.23 Minuten. Markov-Kette (oben links), Histogramm für 9375 Iterationen mit KDE (unten links) und Histogramm für letzte 7500 Iterationen mit KDE (unten rechts).

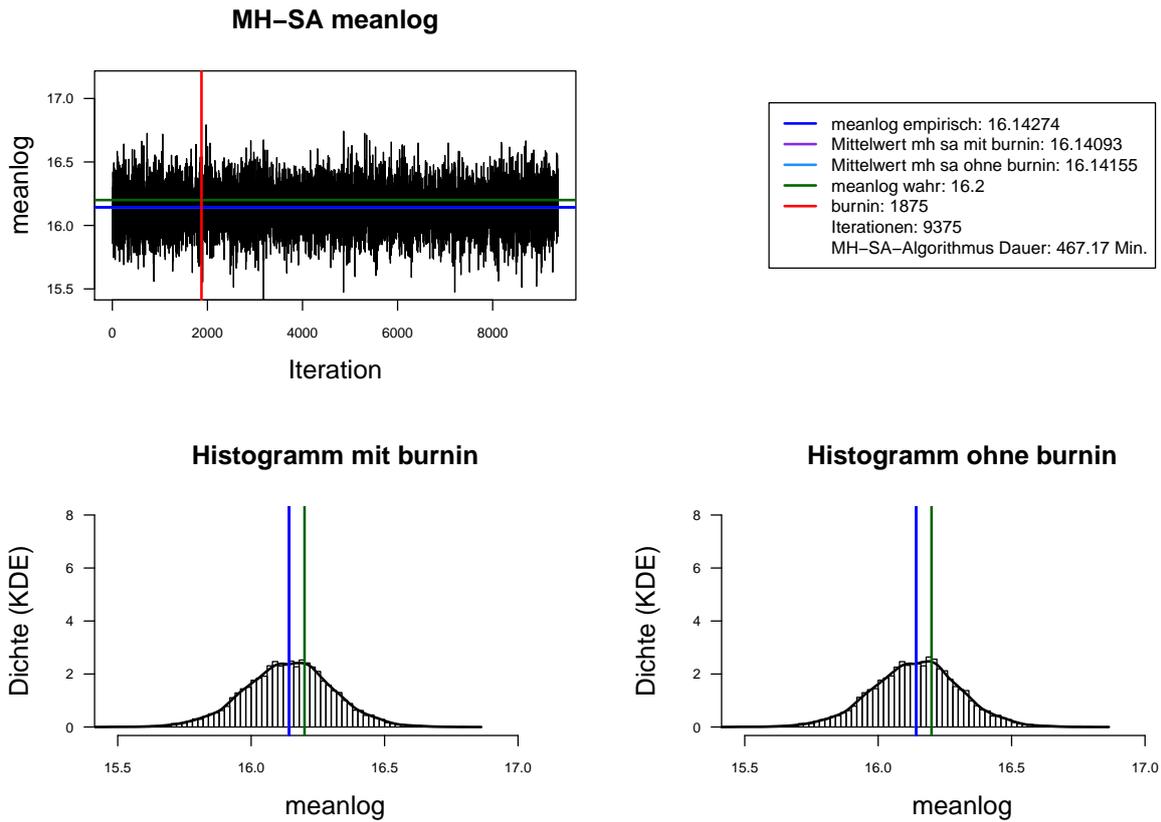


Abbildung A.18.: Der mit MH-SA-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 9375 Iterationen beträgt 467.17 Minuten = 7.79 Stunden. Markov-Kette (oben links), Histogramm für 9375 Iterationen mit KDE (unten links) und Histogramm für letzte 7500 Iterationen mit KDE (unten rechts).

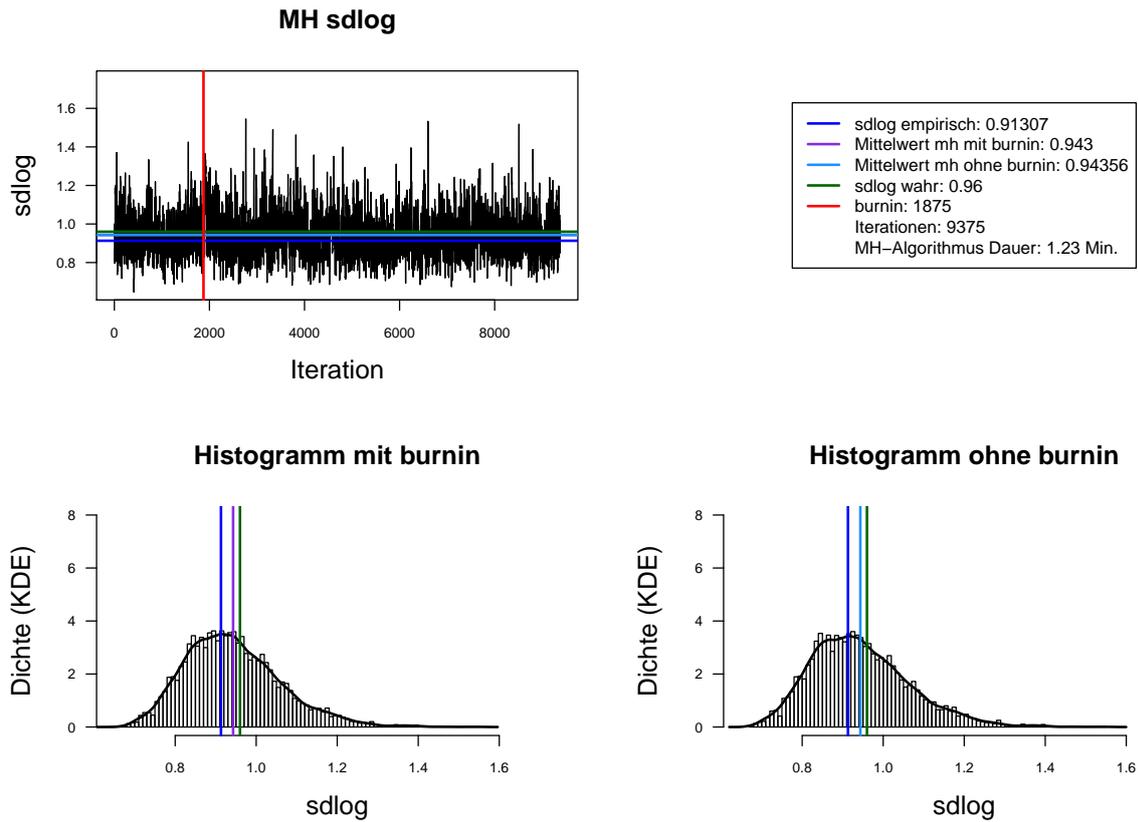


Abbildung A.19.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 9375 Iterationen beträgt 1.23 Minuten. Markov-Kette (oben links), Histogramm für 9375 Iterationen mit KDE (unten links) und Histogramm für letzte 7500 Iterationen mit KDE (unten rechts).

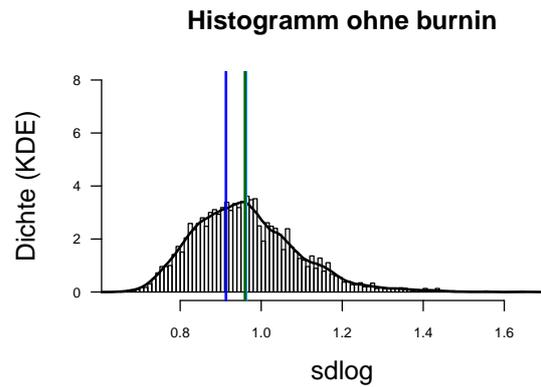
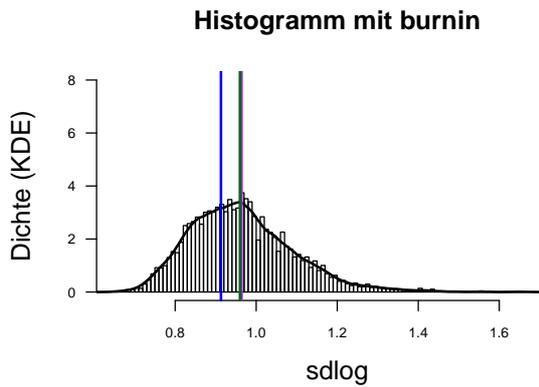
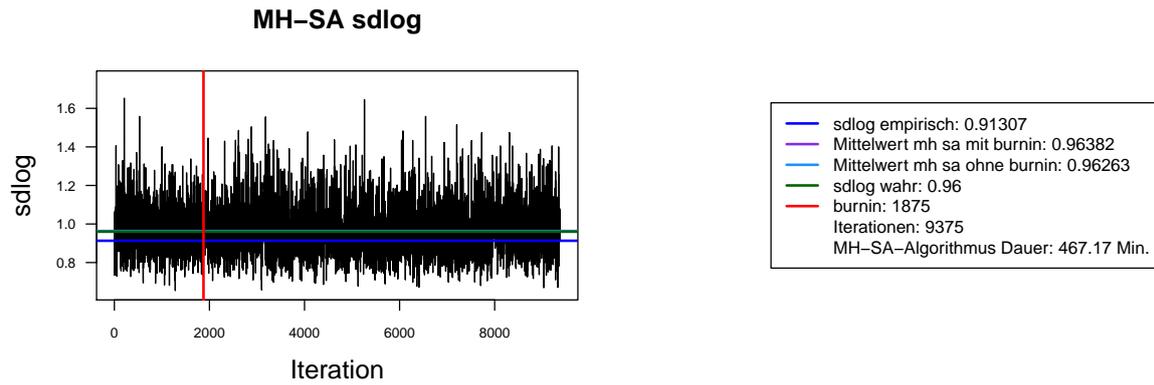


Abbildung A.20.: Der mit MH-SA-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 9375 Iterationen beträgt 467.17 Minuten = 7.79 Stunden. Markov-Kette (oben links), Histogramm für 9375 Iterationen mit KDE (unten links) und Histogramm für letzte 7500 Iterationen mit KDE (unten rechts).

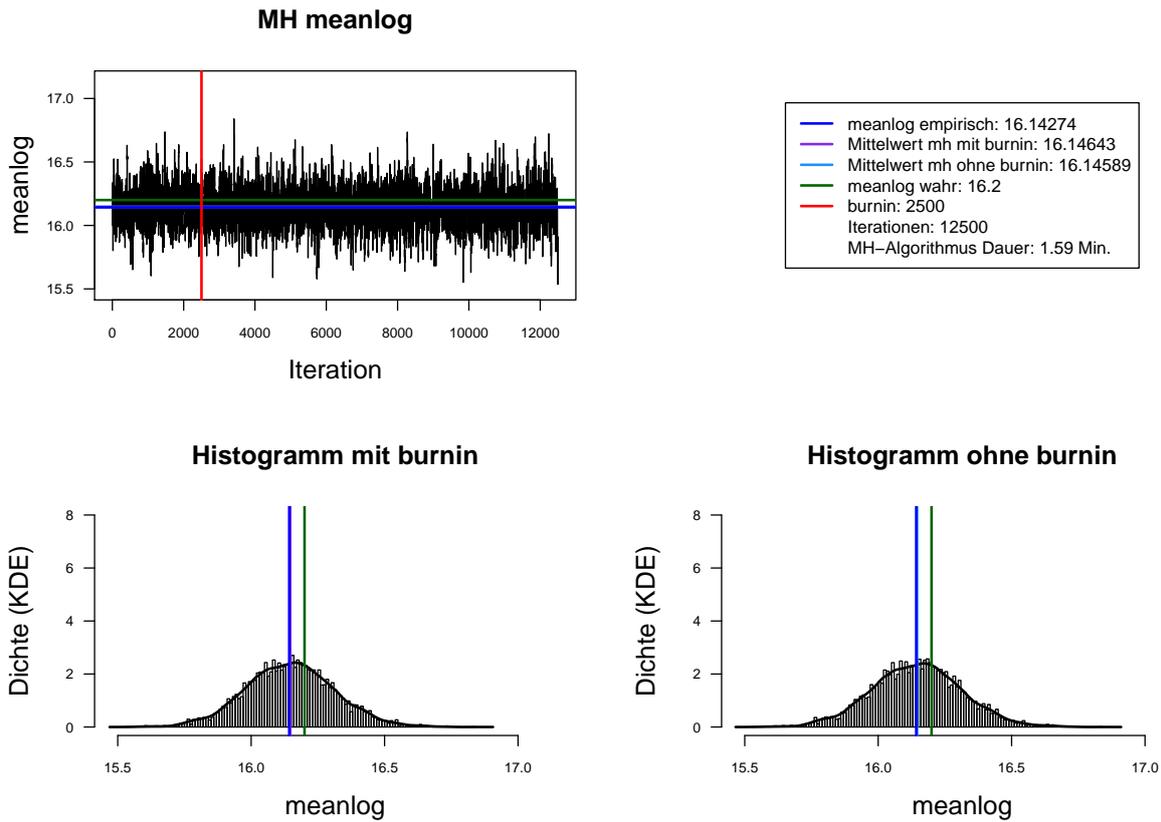


Abbildung A.21.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 12500 Iterationen beträgt 1.59 Minuten. Markov-Kette (oben links), Histogramm für 12500 Iterationen mit KDE (unten links) und Histogramm für letzte 10000 Iterationen mit KDE (unten rechts).

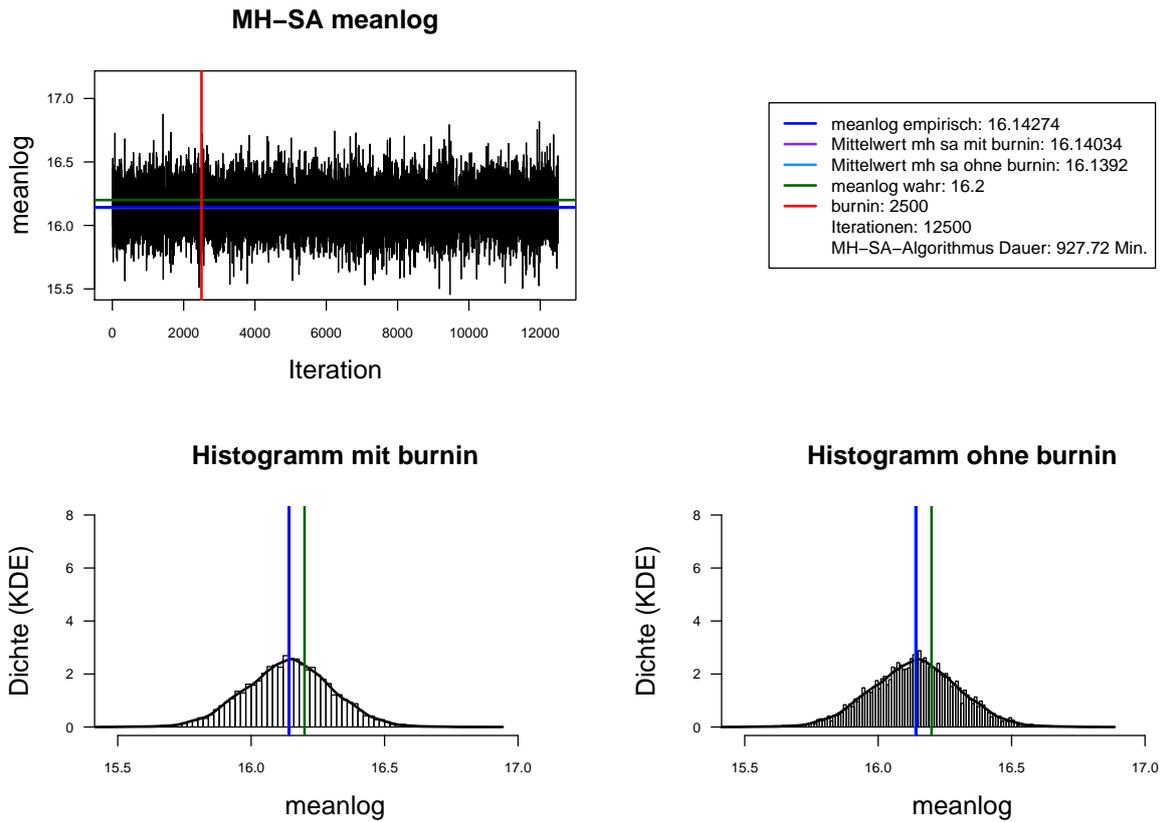


Abbildung A.22.: Der mit MH-SA-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 12500 Iterationen beträgt 927.72 Minuten = 15.46 Stunden. Markov-Kette (oben links), Histogramm für 12500 Iterationen mit KDE (unten links) und Histogramm für letzte 10000 Iterationen mit KDE (unten rechts).

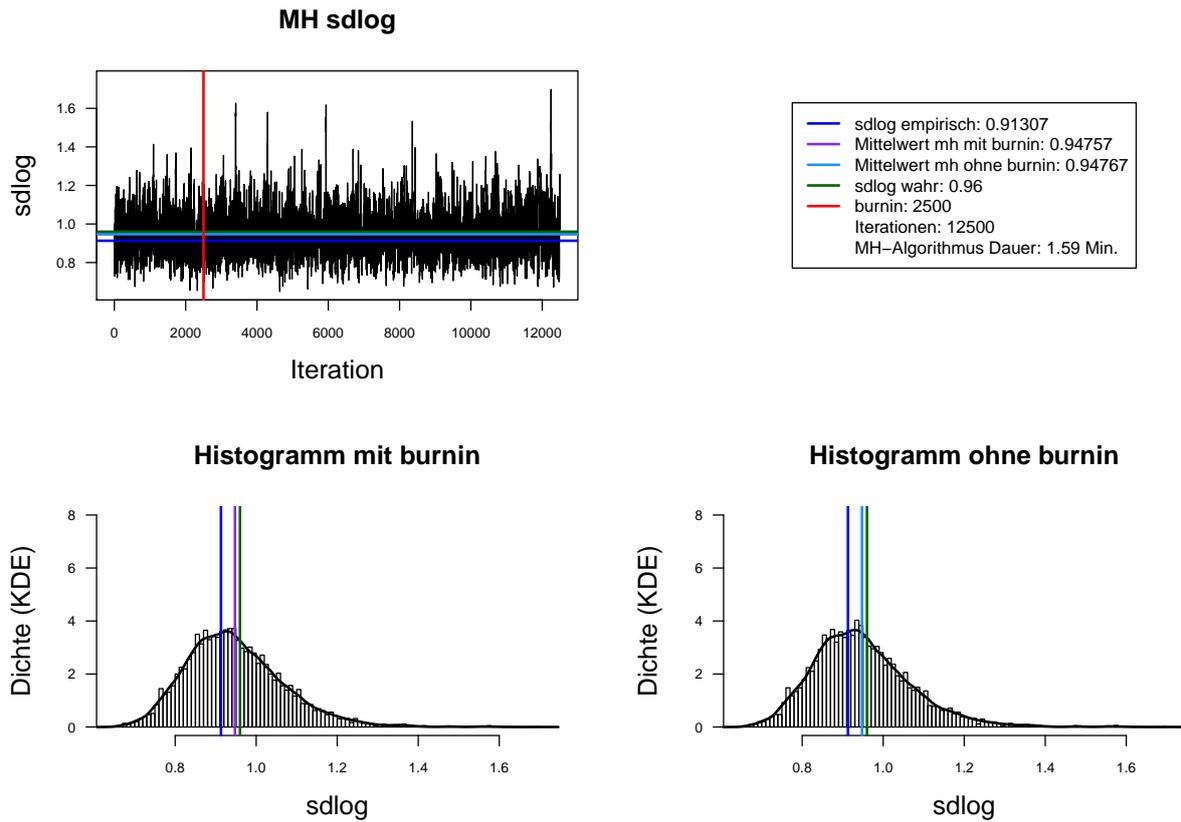


Abbildung A.23.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 12500 Iterationen beträgt 1.59 Minuten. Markov-Kette (oben links), Histogramm für 12500 Iterationen mit KDE (unten links) und Histogramm für letzte 10000 Iterationen mit KDE (unten rechts).

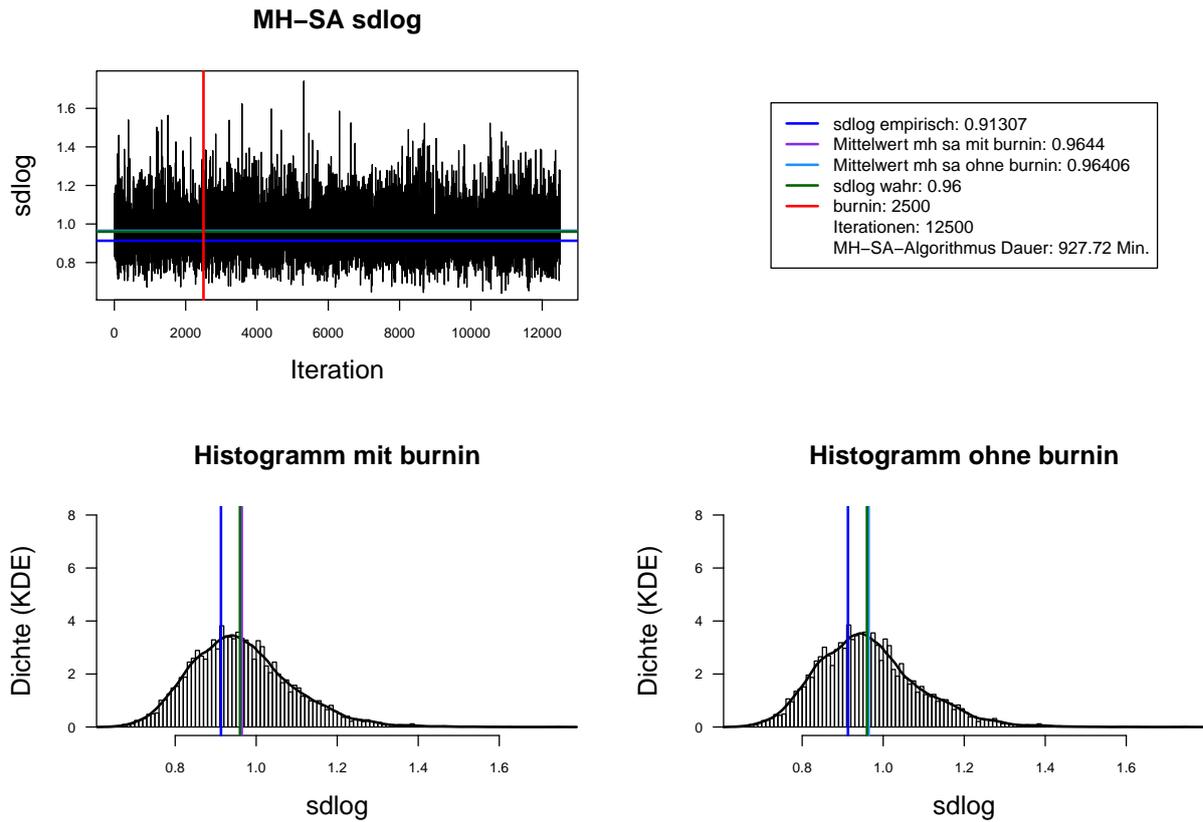
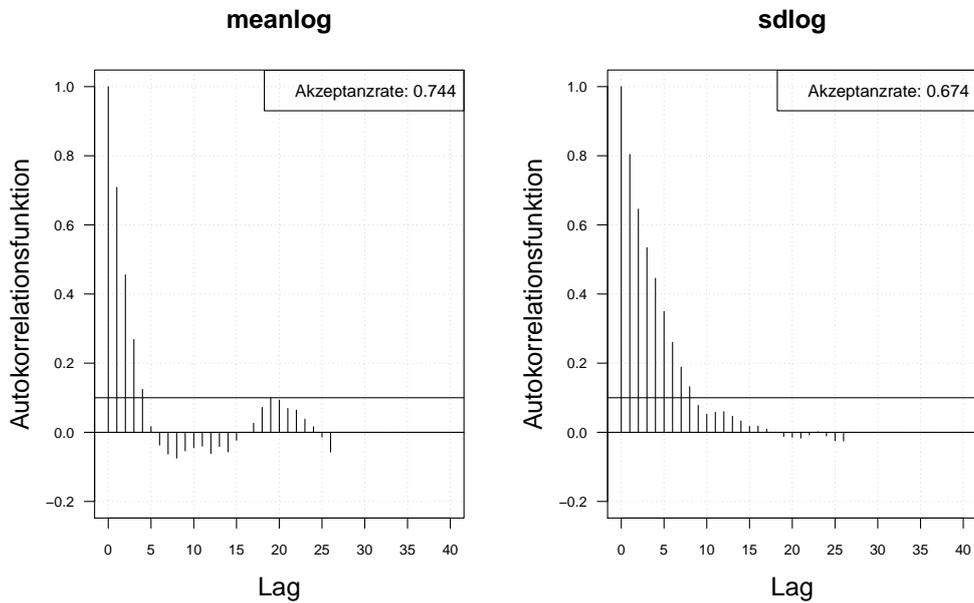


Abbildung A.24.: Der mit MH-SA-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 12500 Iterationen beträgt 927.72 Minuten = 15.46 Stunden. Markov-Kette (oben links), Histogramm für 12500 Iterationen mit KDE (unten links) und Histogramm für letzte 10000 Iterationen mit KDE (unten rechts).

### A.3.2. Autokorrelation

Autokorrelation MH ohne burnin, Iterationen: 500



Autokorrelation MH ohne burnin, thin = 3, Iterationen: 500

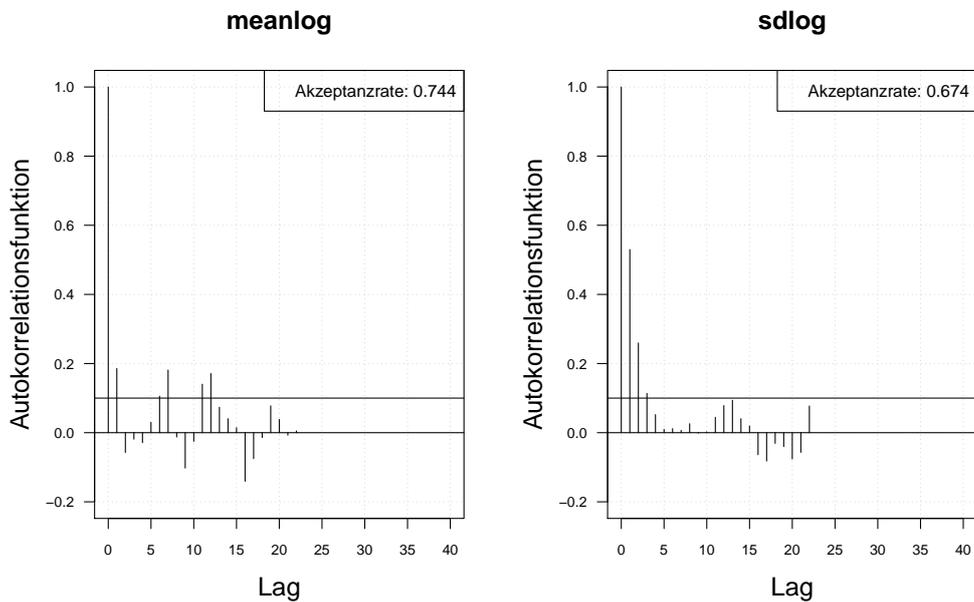
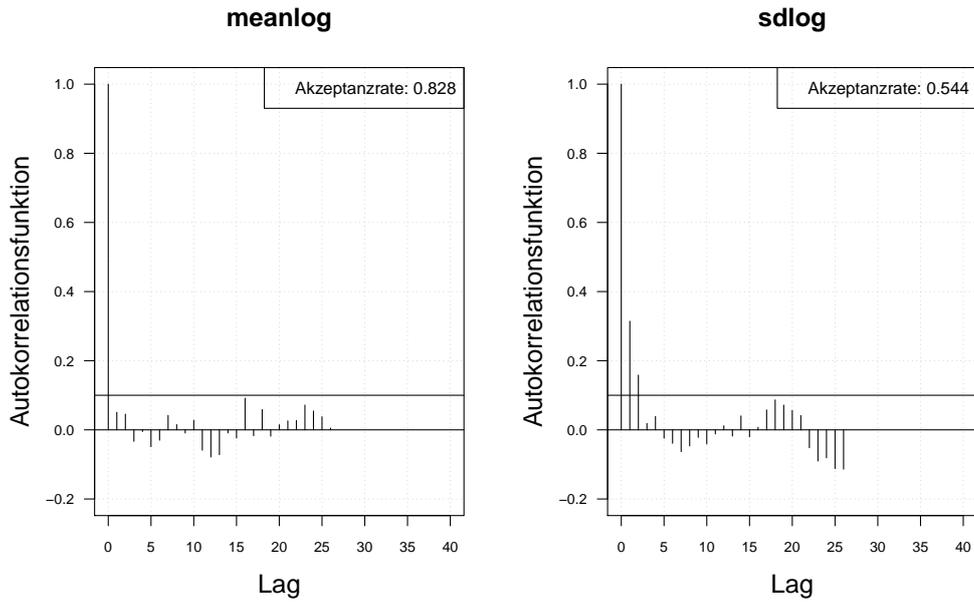


Abbildung A.25.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 500.

Autokorrelation MH-SA ohne burnin, Iterationen: 500



Autokorrelation MH-SA ohne burnin, thin = 3, Iterationen: 500

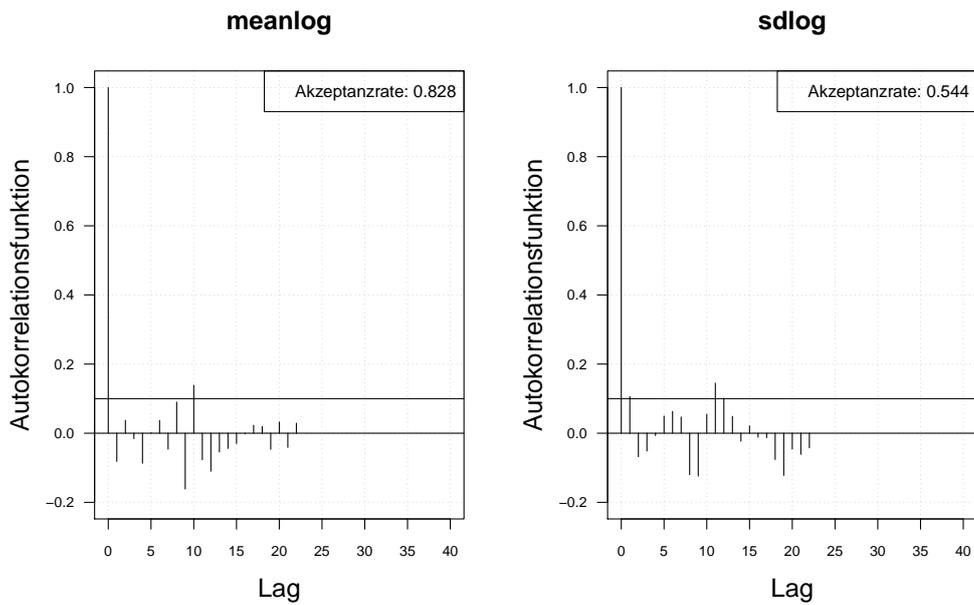
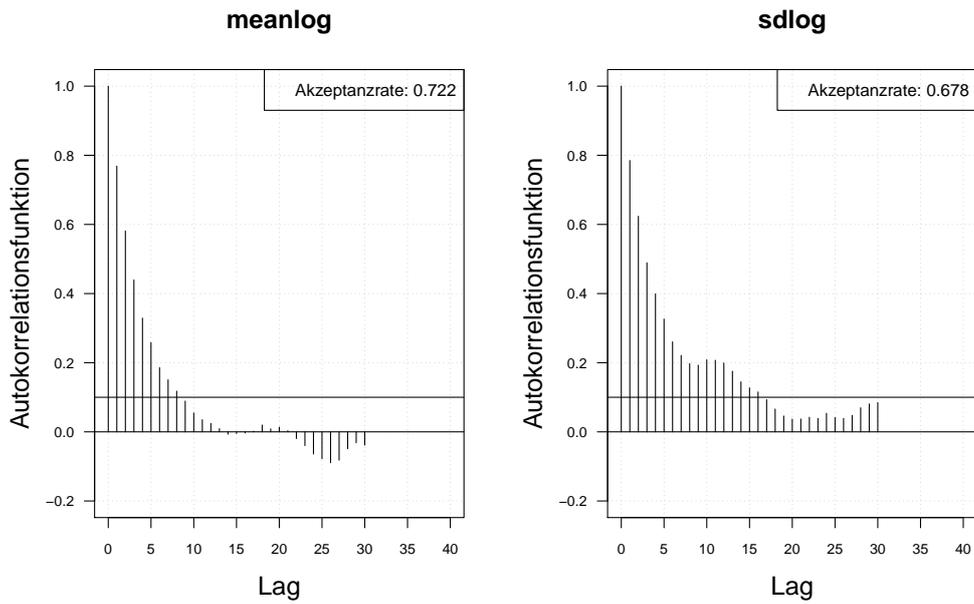


Abbildung A.26.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-SA-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 500.

Autokorrelation MH ohne burnin, Iterationen: 1000



Autokorrelation MH ohne burnin, thin = 3, Iterationen: 1000

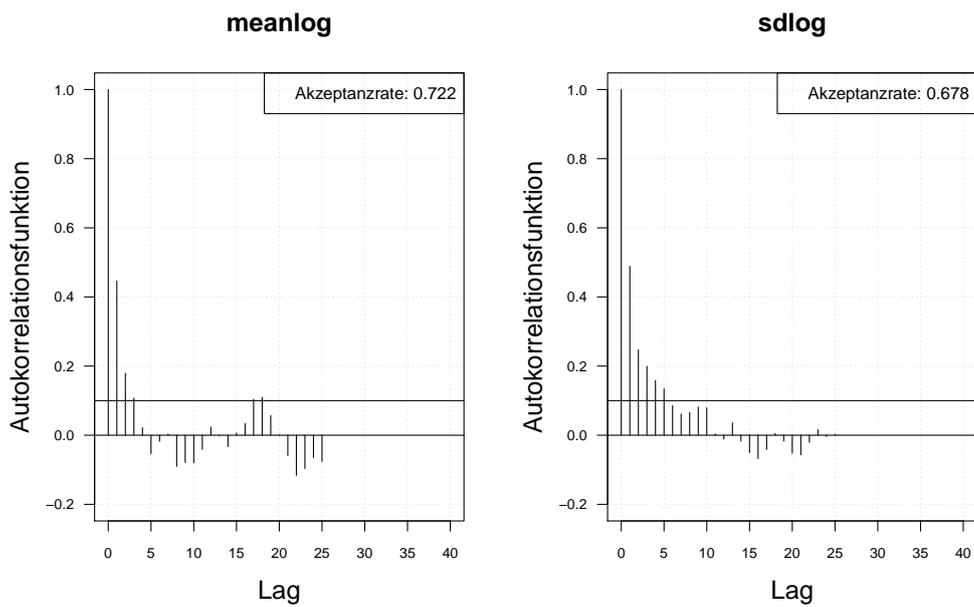
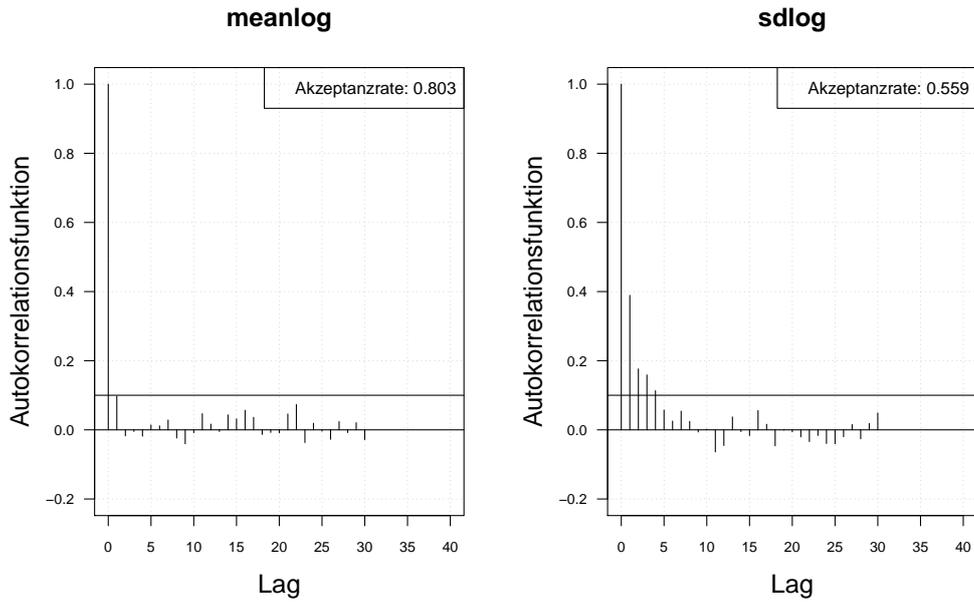


Abbildung A.27.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 1000.

Autokorrelation MH-SA ohne burnin, Iterationen: 1000



Autokorrelation MH-SA ohne burnin, thin = 3, Iterationen: 1000

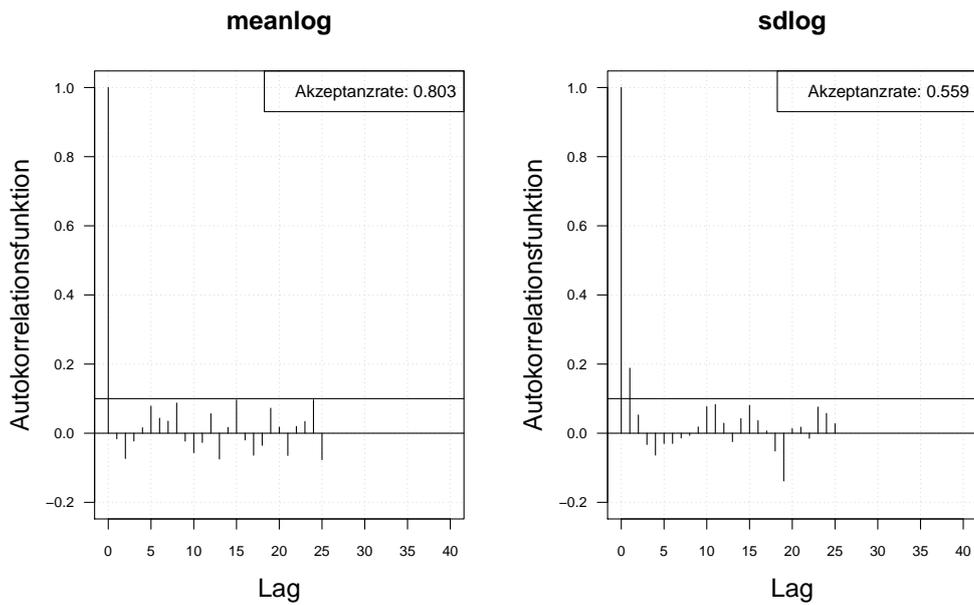
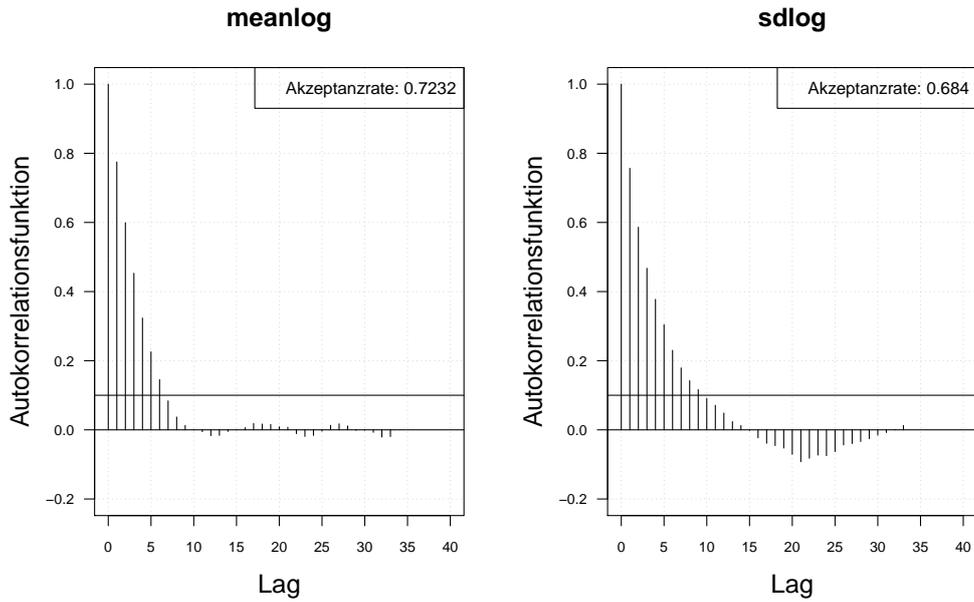


Abbildung A.28.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-SA-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 1000.

Autokorrelation MH ohne burnin, Iterationen: 2500



Autokorrelation MH ohne burnin, thin = 3, Iterationen: 2500

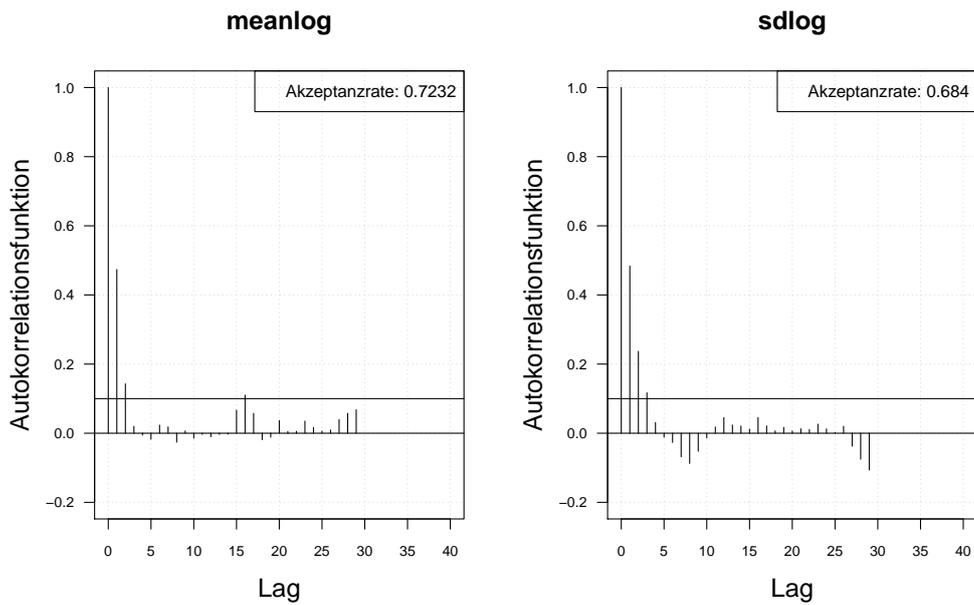
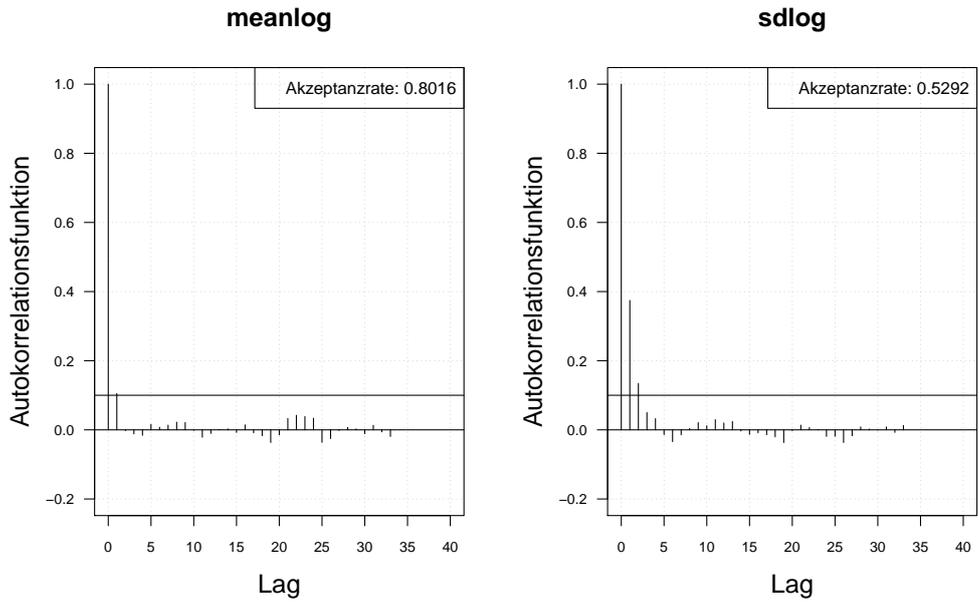


Abbildung A.29.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 2500.

Autokorrelation MH-SA ohne burnin, Iterationen: 2500



Autokorrelation MH-SA ohne burnin, thin = 3, Iterationen: 2500

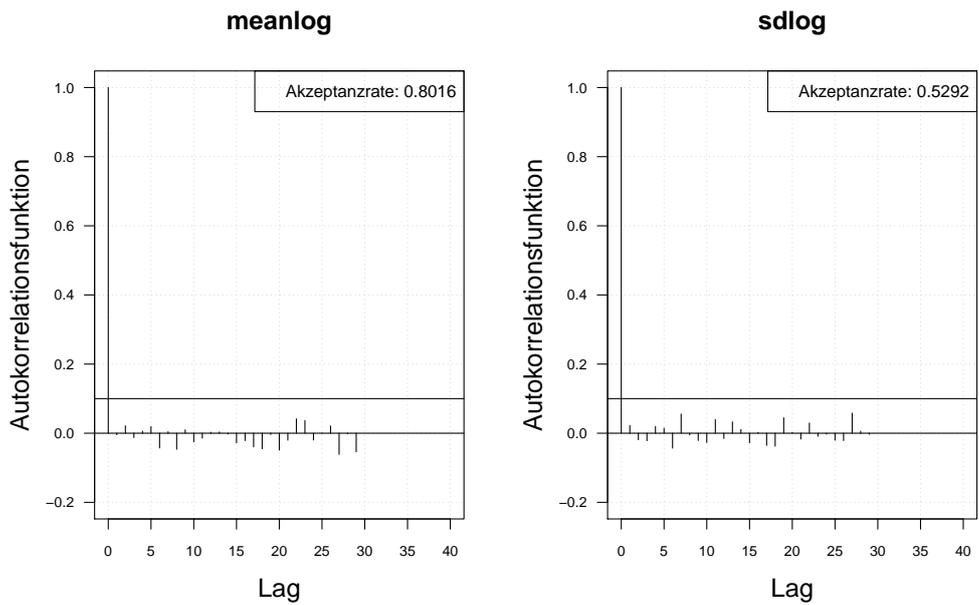
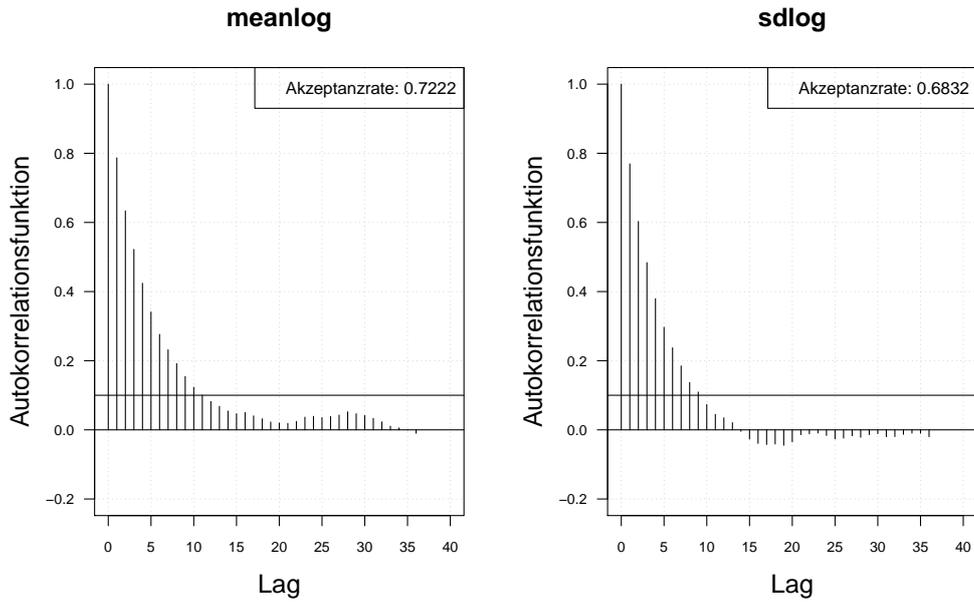


Abbildung A.30.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-SA-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 2500.

Autokorrelation MH ohne burnin, Iterationen: 5000



Autokorrelation MH ohne burnin, thin = 3, Iterationen: 5000

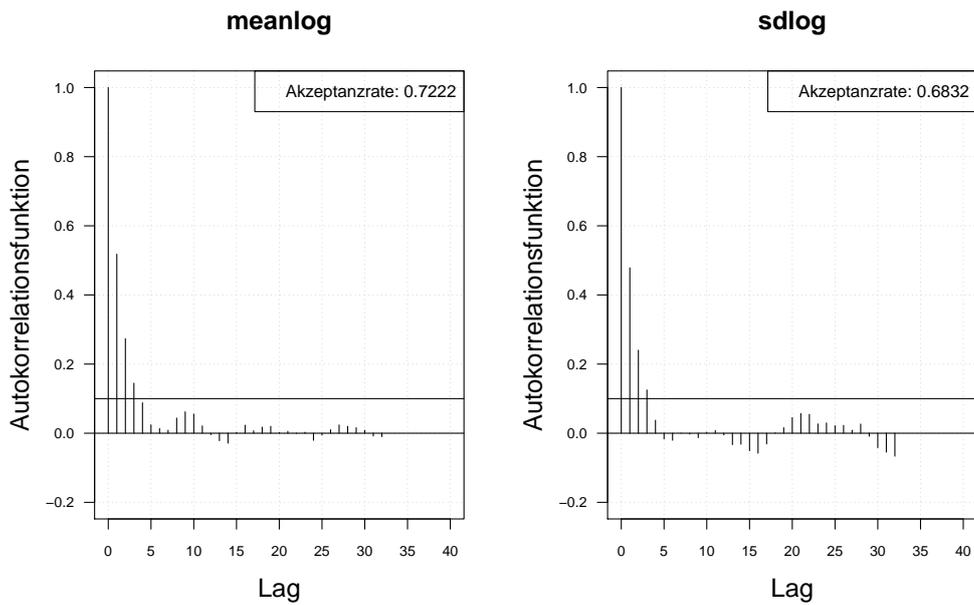
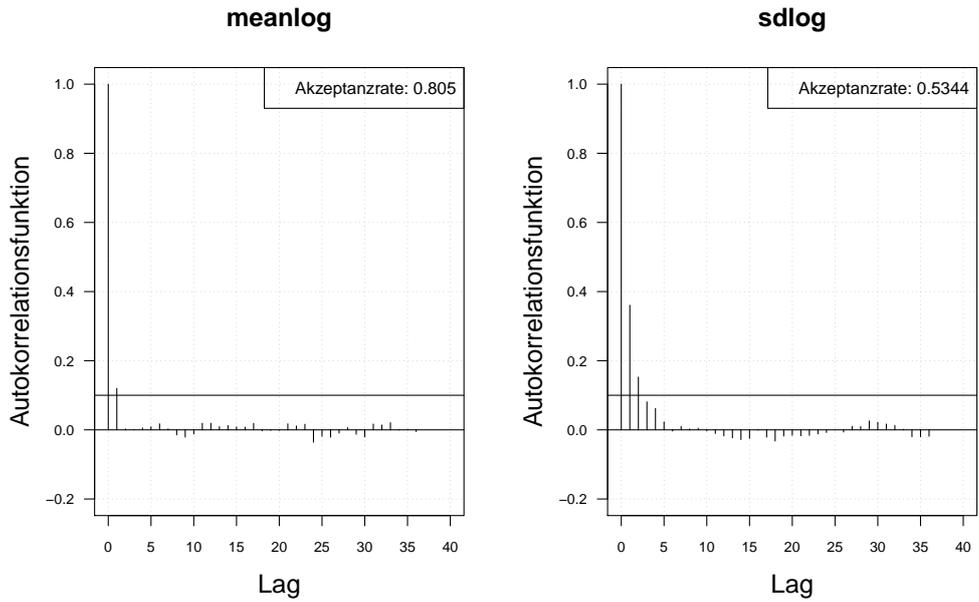


Abbildung A.31.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 5000.

Autokorrelation MH-SA ohne burnin, Iterationen: 5000



Autokorrelation MH-SA ohne burnin, thin = 3, Iterationen: 5000

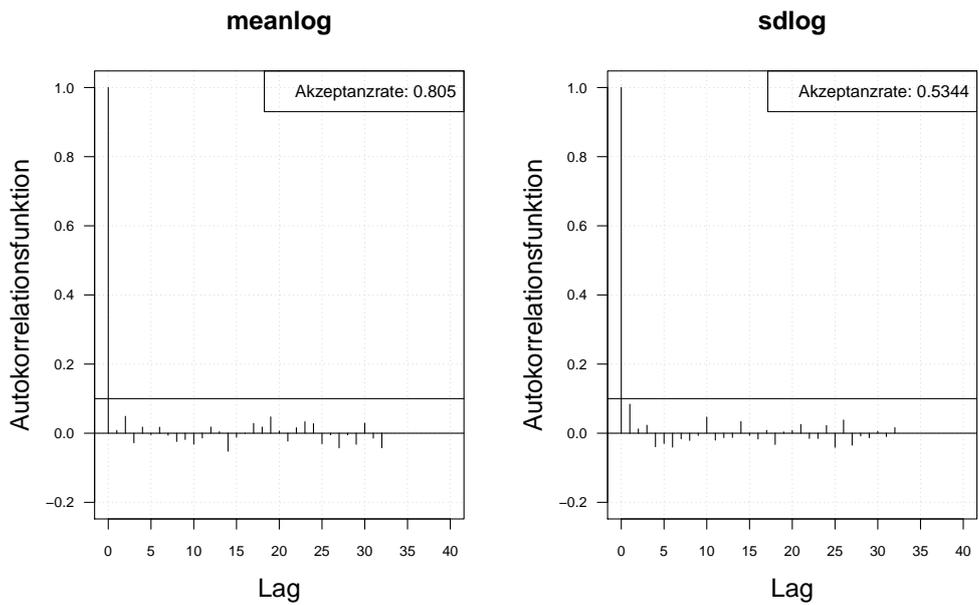
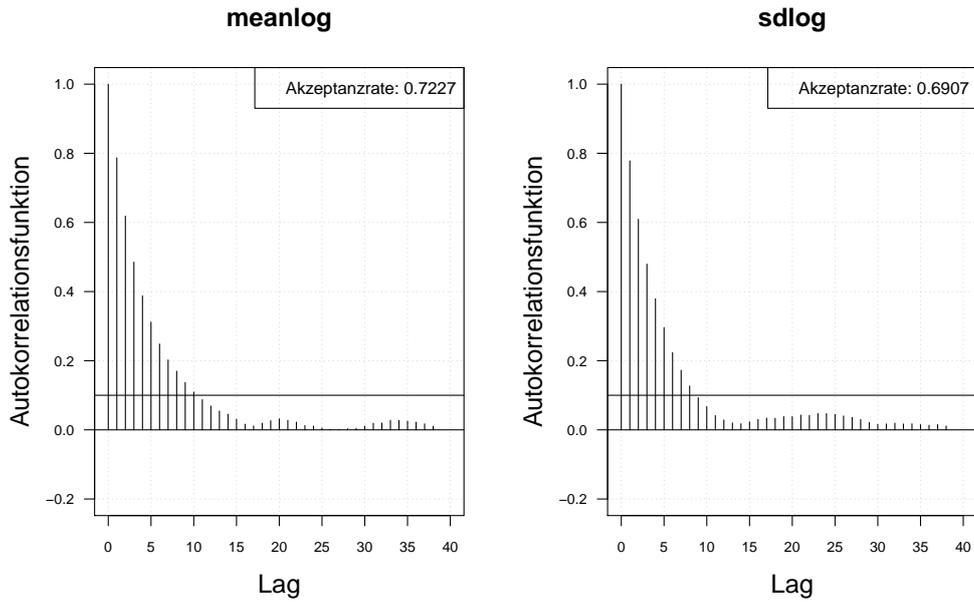


Abbildung A.32.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-SA-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 5000.

Autokorrelation MH ohne burnin, Iterationen: 7500



Autokorrelation MH ohne burnin, thin = 3, Iterationen: 7500

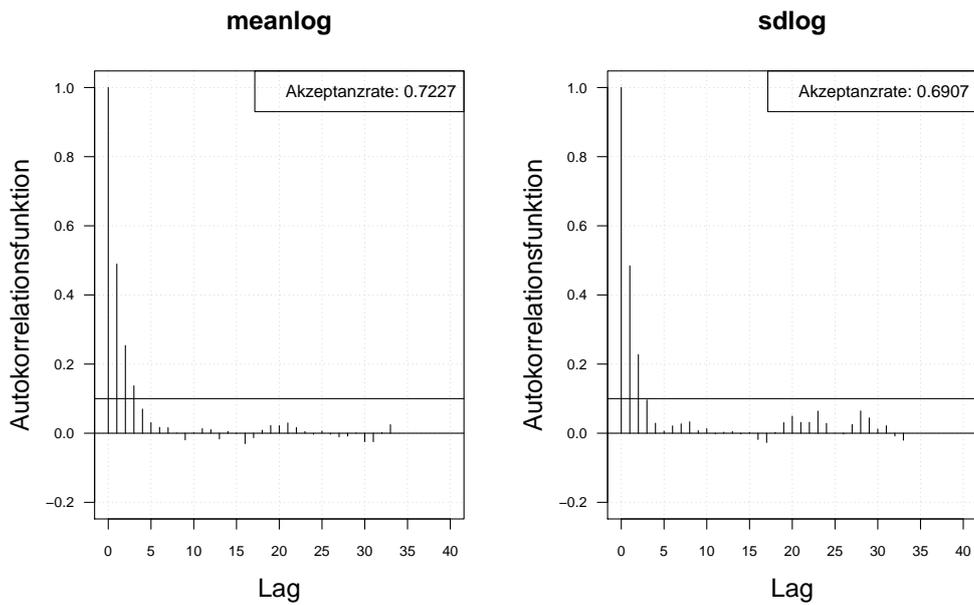
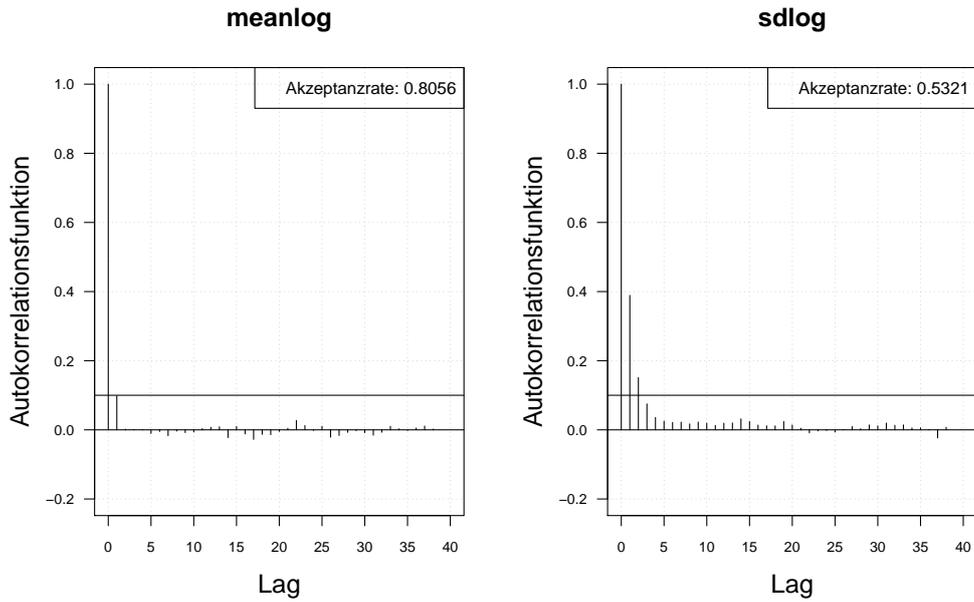


Abbildung A.33.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 7500.

Autokorrelation MH-SA ohne burnin, Iterationen: 7500



Autokorrelation MH-SA ohne burnin, thin = 3, Iterationen: 7500

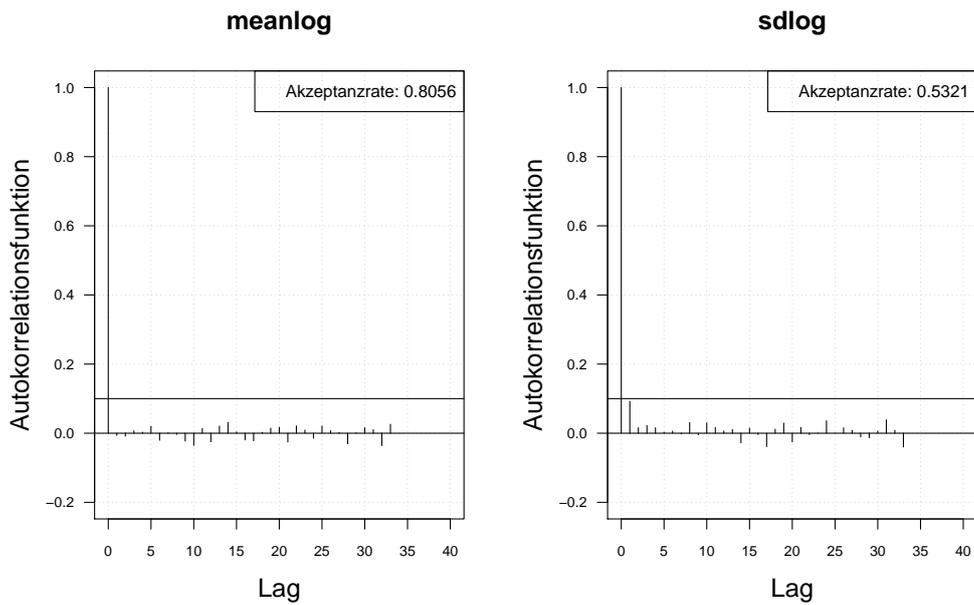
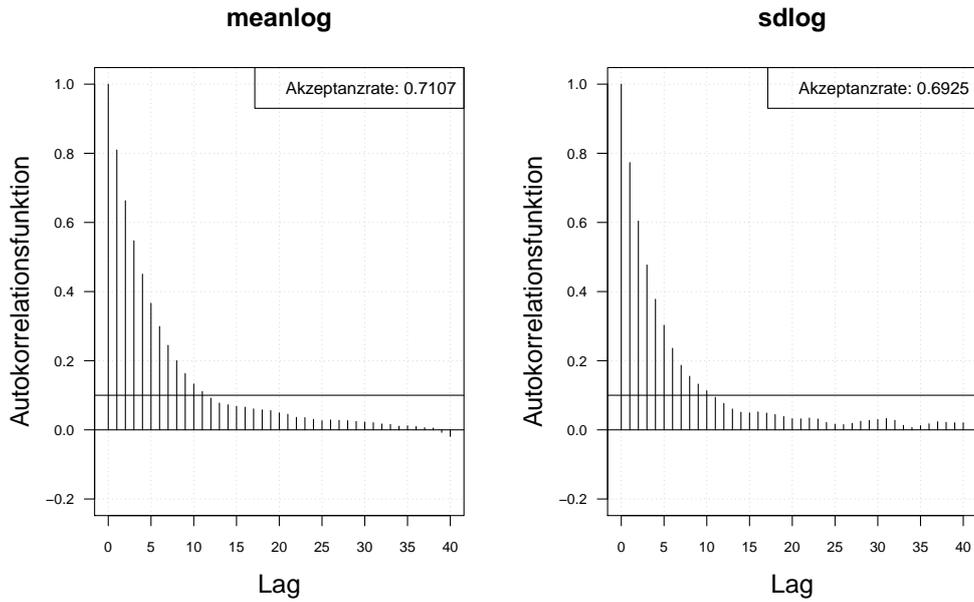


Abbildung A.34.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-SA-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 7500.

Autokorrelation MH ohne burnin, Iterationen: 10000



Autokorrelation MH ohne burnin, thin = 3, Iterationen: 10000

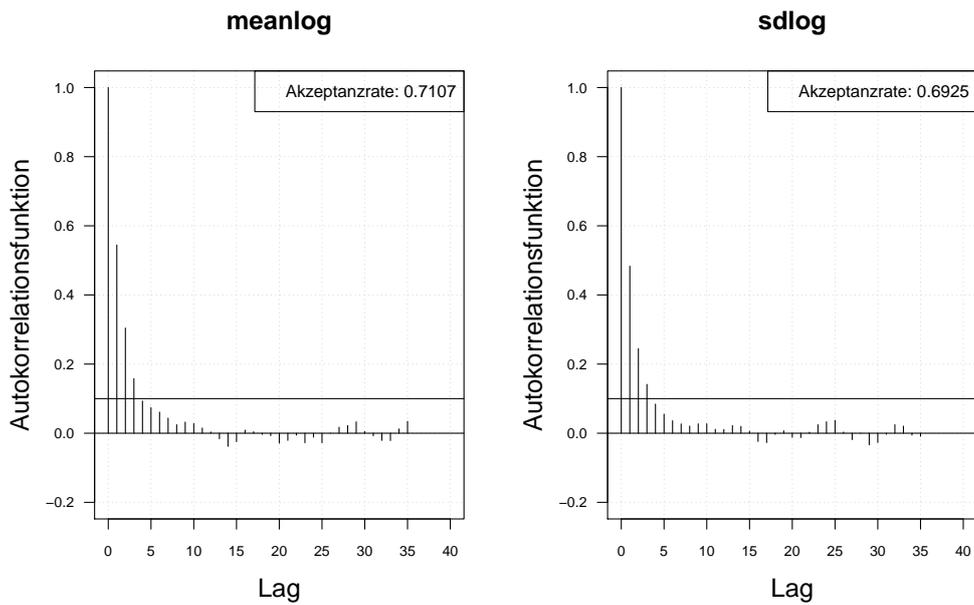
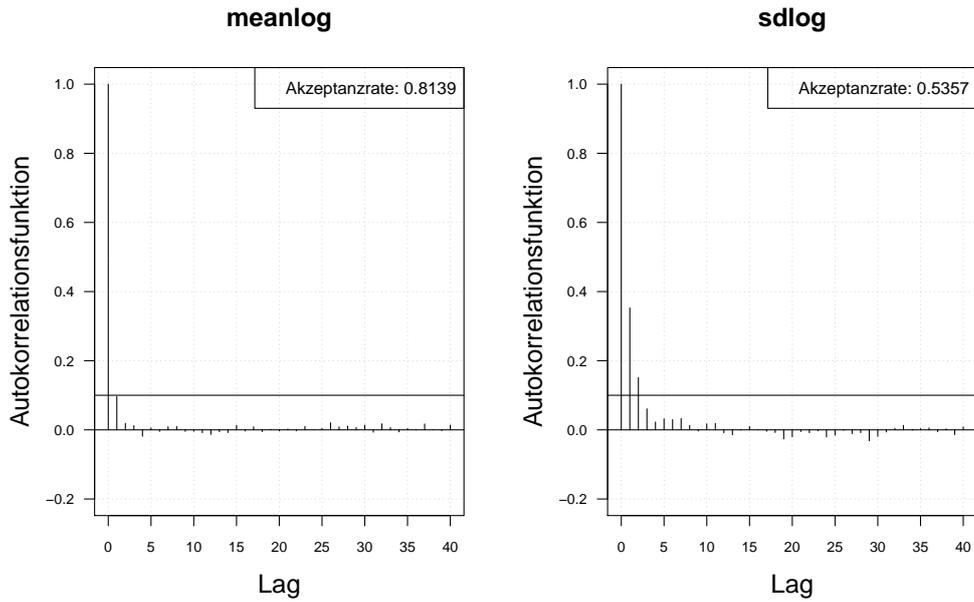


Abbildung A.35.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 10000.

Autokorrelation MH-SA ohne burnin, Iterationen: 10000



Autokorrelation MH-SA ohne burnin, thin = 3, Iterationen: 10000

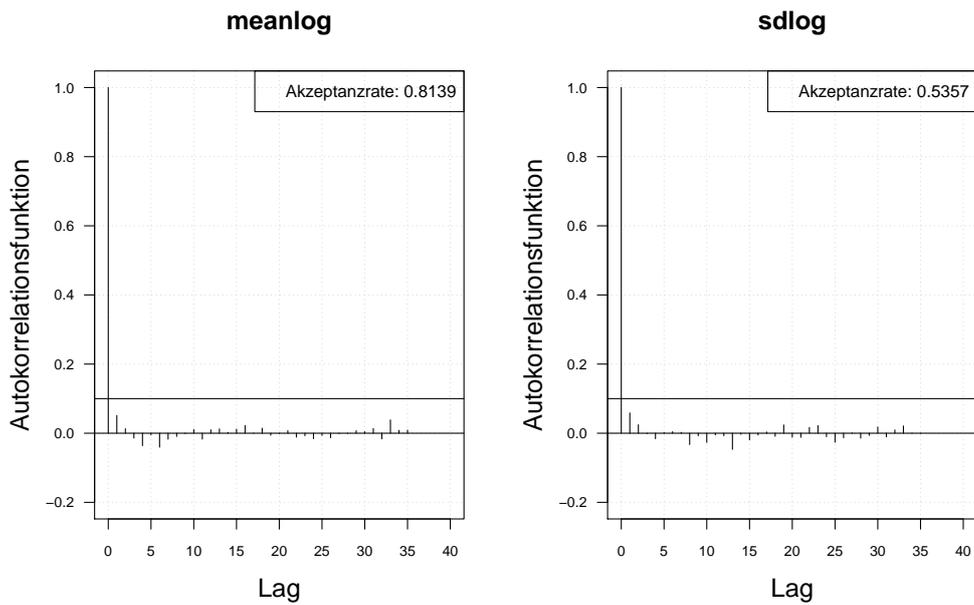


Abbildung A.36.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-SA-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 10000.

## A.4. Konjugierte Verteilungen

Datenverteilung	<i>a priori</i> Verteilung	<i>a posteriori</i> Verteilung
<b>Lognormal-:</b> $X \sim LN(\mu, \sigma^2)$ , $\mu$ unbekannt	<b>Normal-:</b> $\Theta_\mu \sim N(\mu_0, \tau^2)$	<b>Normal-:</b> $\Theta_\mu \sim N(\tilde{\mu}, \tilde{\tau}^2)$ , mit $\tilde{\mu} = \tilde{\tau}^2 \left( \frac{1}{\tau^2} \mu_0 + \frac{1}{\sigma^2} \sum_{i=1}^n \log(x_i) \right)$ $\tilde{\tau}^2 = \frac{\sigma^2 \tau^2}{\sigma^2 + n \tau^2}$
<b>Lognormal-:</b> $X \sim LN(\mu, \sigma^2)$ , $\sigma^2$ unbekannt	<b>Inverse Gamma-:</b> $\Theta_{\sigma^2} \sim IG(\alpha, \beta)$	<b>Inverse Gamma-:</b> $\Theta_{\sigma^2} \sim IG(\tilde{\alpha}, \tilde{\beta})$ , mit $\tilde{\alpha} = \alpha + 0.5n$ $\tilde{\beta} = \beta + 0.5 \sum_{i=1}^n (x_i - \mu)^2$
<b>Lognormal-:</b> $X \sim LN(\mu, \sigma^2)$ , $\mu, \sigma^2$ unbekannt, $\mu, \sigma^2$ abhängig	<b>Normal-:</b> $\Theta_{\mu \sigma^2} \sim N(\mu_0, \frac{\sigma^2}{\phi})$ <b>Skalierte inverse</b> <b>Chi-Quadrat-:</b> $\Theta_{\sigma^2} \sim I\chi^2(\nu, \tau^2)$	<b>Mischverteilung<sup>3</sup>:</b> mit neuen Parametern $\mu_{0,neu} = \frac{n\bar{x} + \phi\mu_0}{\phi + n}$ $\phi_{neu} = \phi + n$ $\nu_{neu} = \nu + n$ $(\nu\tau^2)_{neu} = \nu\tau^2 + \phi\mu_0^2 + n\bar{x}^2 - \frac{(n\bar{x} + \phi\mu_0)^2}{\phi + n}$ $\bar{x} = \frac{1}{n} \sum_{i=1}^n \log(x_i)$ $\bar{x}^2 = \frac{1}{n} \sum_{i=1}^n \log^2(x_i)$
<b>Poisson-:</b> $X \sim Pois(\lambda)$ , $\lambda$ unbekannt	<b>Gamma-:</b> $\Theta_\lambda \sim G(\alpha, \beta)$	<b>Gamma-:</b> $\Theta_\lambda \sim G(\alpha + \sum_{i=1}^n x_i, \beta + n)$
<b>Negative Binomial-:</b> $X \sim NB(r, p)$ , $p$ unbekannt	<b>Beta-:</b> $\Theta_p \sim Beta(\alpha, \beta)$	<b>Beta-:</b> $\Theta_p \sim Beta(\tilde{\alpha}, \tilde{\beta})$ , mit $\tilde{\alpha} = \alpha + \sum_{i=1}^n x_i$ $\tilde{\beta} = \beta + rn$

Tabelle A.1.: Konjugierte *a priori*.

<sup>3</sup>Die Informationen zur Mischverteilung und die Herleitung siehe Anh. A.8.

## A.5. Verteilungen

### A.5.1. Lognormal-Verteilung

$X \sim LN(\mu, \sigma^2)$ , mit  $\mu \in \mathbb{R}$  - Lokationsparameter,  $\sigma > 0$  - Skalenparameter.

- **Dichte-Funktion:**

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{[\log(x) - \mu]^2}{2\sigma^2}\right), \quad x > 0$$

- **Verteilungsfunktion:**

$$F(x) = \Phi\left(\frac{\log(x) - \mu}{\sigma}\right), \quad x > 0.$$

mit  $\Phi(\cdot)$  - die Verteilungsfunktion der Standardnormal-Verteilung

- **Erwartungswert:**  $E(X) = \exp(\mu + \frac{\sigma^2}{2})$
- **Varianz:**  $Var(X) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$

### A.5.2. Normal-Verteilung

$X \sim N(\mu, \sigma^2)$ , mit  $\mu \in \mathbb{R}$  - Mittelwert (Lokationsparameter),  $\sigma^2 > 0$  - Varianz (Skalenparameter).

- **Dichte-Funktion:**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{[x - \mu]^2}{2\sigma^2}\right), \quad x \in \mathbb{R}$$

- **Verteilungsfunktion:**

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}$$

mit  $\Phi(\cdot)$  - die Verteilungsfunktion der Standardnormal-Verteilung

- **Erwartungswert:**  $E(X) = \mu$
- **Varianz:**  $Var(X) = \sigma^2$

### A.5.3. Inverse Gamma-Verteilung

$X \sim \text{IG}(\alpha, \beta)$ , mit  $\alpha > 0$  - der Formparameter,  $\beta > 0$  - der Skalenparameter.  
Inverse Chi-Quadrat-Verteilung für  $\alpha = \frac{\nu}{2}$  und  $\beta = \frac{1}{2}$ .  
Skalierte inverse Chi-Quadrat-Verteilung für  $\alpha = \frac{\nu}{2}$  und  $\beta = \frac{\nu}{2}\tau^2$ .

- **Dichte-Funktion:**

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{\beta}{x}\right), \quad x > 0$$

mit  $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Verteilungsfunktion:**

$$F(x) = \frac{\Gamma(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}, \quad x > 0$$

mit

$\Gamma(s, z) = \int_z^\infty t^{s-1} \exp(-t) dt$  - die obere unvollständige Gammafunktion.

$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Erwartungswert:**  $E(X) = \frac{\beta}{\alpha - 1}$ , für  $\alpha > 1$
- **Varianz:**  $\text{Var}(X) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}$ , für  $\alpha > 2$

### A.5.4. Skalierte inverse Chi-Quadrat-Verteilung

$X \sim \text{scaledI}\chi^2(\nu, \tau^2)$ , mit  $\nu > 0$  - die Freiheitsgrade,  $\tau^2 > 0$  - der Skalenparameter.  
Sonderfall von inversen Gamma-Verteilung mit  $\alpha = \frac{\nu}{2}$  und  $\beta = \frac{\nu}{2}\tau^2$ .

- **Dichte-Funktion:**

$$f(x) = \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \tau^\nu x^{-(1+\frac{\nu}{2})} \exp\left(-\frac{\nu\tau^2}{2x}\right), \quad x > 0$$

mit  $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Verteilungsfunktion:**

$$F(x) = \frac{\Gamma\left(\frac{\nu}{2}, \frac{\nu\tau^2}{2x}\right)}{\Gamma\left(\frac{\nu}{2}\right)}$$

mit

$\Gamma(s, z) = \int_z^\infty t^{s-1} \exp(-t) dt$  - die obere unvollständige Gammafunktion.

$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Erwartungswert:**  $E(X) = \frac{\nu\tau^2}{\nu-2}$ , für  $\nu > 2$
- **Varianz:**  $Var(X) = \frac{2\nu^2\tau^4}{(\nu-2)^2(\nu-4)}$ , für  $\nu > 4$

### A.5.5. Poisson-Verteilung

$X \sim Poiss(\lambda)$ , mit  $\lambda > 0$  - Rate.

- **Dichte-Funktion:**

$$f(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x \in \mathbb{N}^0$$

- **Verteilungsfunktion:**

$$F(x) = \sum_{k=0}^x P(X = k) = e^{-\lambda} \sum_{k=0}^x \frac{\lambda^k}{k!}, \quad x \in \mathbb{N}^0$$

- **Erwartungswert:**  $E(X) = \lambda$
- **Varianz:**  $Var(X) = \lambda$

### A.5.6. Gamma-Verteilung

$X \sim G(\alpha, \beta)$ , mit  $\alpha > 0$  - der Formparameter,  $\beta > 0$  - der inverse<sup>4</sup> Skalenparameter.  
Chi-Quadrat-Verteilung für  $\alpha = \frac{\nu}{2}$  und  $\beta = \frac{1}{2}$ .

- **Dichte-Funktion:**

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad x > 0$$

mit  $\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Verteilungsfunktion:**

$$F(x) = \frac{\Gamma(\alpha, \beta x)}{\Gamma(\alpha)}, \quad x > 0$$

mit

$\Gamma(s, z) = \int_0^z t^{s-1} \exp(-t) dt$  - die untere unvollständige Gammafunktion.

$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Erwartungswert:**  $E(X) = \frac{\alpha}{\beta}$
- **Varianz:**  $Var(X) = \frac{\alpha}{\beta^2}$

---

<sup>4</sup>Der inverse Skalenparameter ist definiert als  $\frac{1}{\beta}$ .

### A.5.7. Negative Binomial-Verteilung

$X \sim NB(r, p)$ , mit  $r > 0$  - Formparameter (Anzahl Erfolge bis zum Abbruch),  $p \in (0, 1)$  - Skalenparameter (Erfolgswahrscheinlichkeit).

- **Dichte-Funktion:**

$$f(x) = \binom{x+r-1}{x} p^r (1-p)^x, \quad x \in \mathbb{N}^0$$

- **Verteilungsfunktion:**

$$F(x) = 1 - I_{1-p}(x+1, r), \quad x \in \mathbb{N}^0$$

mit  $I(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$  - die Eulersche Betafunktion, auch Eulersches Integral 1. Art.

- **Erwartungswert:**  $E(X) = \frac{r(1-p)}{p}$
- **Varianz:**  $Var(X) = \frac{r(1-p)}{p^2}$

### A.5.8. Beta-Verteilung

$X \sim Beta(\alpha, \beta)$ , mit  $\alpha > 0$  - Formparameter,  $\beta > 0$  - Skalenparameter.

- **Dichte-Funktion:**

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad x \in [0, 1]$$

mit

$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Verteilungsfunktion:**

$$F(x) = I_x(\alpha, \beta), \quad x \in [0, 1]$$

mit  $I_x(a, b) = \frac{B(x; a, b)}{B(a, b)}$  - die regulierte unvollständige Beta-Funktion, wobei

$B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$  - die unvollständige Beta-Funktion

$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$  - die vollständige Beta-Funktion

$\Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt$  - die Gammafunktion.

- **Erwartungswert:**  $E(X) = \frac{\alpha}{\alpha+\beta}$
- **Varianz:**  $Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

## A.6. Bestimmung der Jeffreys *a priori* für Poisson-Verteilung und der dazugehörigen *a posteriori* Dichte

Herleitung zur Kap. 5.1.1 Formel 5.11

$$\begin{aligned}
p_J(\lambda) &= \sqrt{\det(I(\lambda))} = \sqrt{\det\left(-E\left(\frac{\partial^2}{\partial^2\lambda}\log(f(\underline{n}|\lambda))\right)\right)} \\
&= \sqrt{-E\left(\frac{\partial^2}{\partial^2\lambda}\log\left(\prod_{t=1}^T \frac{\lambda^{n_t}}{n_t!} e^{-\lambda}\right)\right)} \\
&= \sqrt{-E\left(\frac{\partial^2}{\partial^2\lambda}\log(\exp(-T\lambda)) + \log\left(\prod_{t=1}^T \lambda^{n_t}\right) - \log\left(\prod_{t=1}^T n_t!\right)\right)} \\
&= \sqrt{-E\left(\frac{\partial^2}{\partial^2\lambda} - T\lambda + \log\left(\lambda^{\sum_{t=1}^T n_t}\right) - \sum_{t=1}^T \log(n_t!)\right)} \\
&= \sqrt{-E\left(\frac{\partial^2}{\partial^2\lambda} - T\lambda + \log(\lambda) \sum_{t=1}^T n_t - \sum_{t=1}^T \log(n_t!)\right)} \\
&= \sqrt{-E\left(\frac{\partial}{\partial\lambda} - T + \frac{1}{\lambda} \sum_{t=1}^T n_t\right)} = \sqrt{-E\left(-\frac{1}{\lambda^2} \sum_{t=1}^T n_t\right)} \\
&= \sqrt{\frac{1}{\lambda^2} \sum_{t=1}^T E(n_t)} = \sqrt{\frac{1}{\lambda^2} \sum_{t=1}^T E(n)} = \sqrt{\frac{1}{\lambda^2} \sum_{t=1}^T \lambda} \\
&= \sqrt{\frac{1}{\lambda^2} T\lambda} = \sqrt{\frac{T}{\lambda}} \propto \sqrt{\frac{1}{\lambda}} = \frac{1}{\sqrt{\lambda}}
\end{aligned} \tag{A.1}$$

Herleitung zur Kap. 5.1.1. Bestimmung der *a posteriori* für Poisson-Verteilung mit Jeffreys *a priori*

Die *a posteriori* Verteilung für Poisson-Verteilung mit Jeffreys *a priori* ist die Gamma-Verteilung mit neuen Hyperparametern  $\Theta_\lambda \sim G\left(\frac{1}{2} + \sum_{t=1}^T n_t, T\right)$

$$\begin{aligned}
p(\lambda|\underline{n}) &\propto \text{Jeffreys } a \text{ priori} \times \text{Likelihood} = \frac{1}{\sqrt{\lambda}} \times \prod_{t=1}^T \frac{\lambda^{n_t}}{n_t!} e^{-\lambda} \stackrel{\text{bzgl. } \lambda}{\propto} \lambda^{-\frac{1}{2}} \prod_{t=1}^T \lambda^{n_t} e^{-\lambda} \\
&= \lambda^{-\frac{1}{2}} \lambda^{\sum_{t=1}^T n_t} \exp\left(-\sum_{t=1}^T \lambda\right) = \lambda^{-\frac{1}{2} + \sum_{t=1}^T n_t} \exp(-T\lambda) \\
&\quad \lambda^{\tilde{\alpha}-1} \exp(-\tilde{\beta}\lambda)
\end{aligned} \tag{A.2}$$

mit  $\tilde{\alpha} = \frac{1}{2} + \sum_{t=1}^T n_t$  - der Formparameter der Gamma-Verteilung,  $\tilde{\beta} = T$  der inverse Skalenparameter<sup>5</sup> der Gamma-Verteilung.

<sup>5</sup>Der inverse Skalenparameter ist definiert als  $\frac{1}{\beta}$ . Der nicht inverse Skalenparameter ist somit  $\frac{1}{T}$ .

## A.7. Bestimmung der bedingten Dichte der zukünftigen Beobachtung und des dazugehörigen Erwartungswertes

### Herleitung zur Kap. 5.1.1 Formel 5.5

Die bedingte Dichte der zukünftigen Beobachtung für gemischte Poisson-Gamma-Verteilung ist eine Dichte der negativen Binomial-Verteilung<sup>6</sup> ( $N_{T+1} \sim NB(\tilde{\alpha}_{neu}, \tilde{\beta}_{neu})$ ):

$$\begin{aligned}
 f(m|\underline{n}) &= P(N_{T+1} = m | N = \underline{n}) = \int f(m|\lambda)p(\lambda|\underline{n})d\lambda \\
 &= \int \frac{\lambda^m}{m!} \exp(-\lambda) \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \lambda^{\tilde{\alpha}-1} \exp(-\tilde{\beta}\lambda) d\lambda \\
 &= \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \frac{1}{m!} \int \lambda^{m+\tilde{\alpha}-1} \exp(-\lambda(1+\tilde{\beta})) d\lambda \\
 &= \frac{\tilde{\beta}^{\tilde{\alpha}}}{\Gamma(\tilde{\alpha})} \frac{1}{m!} \Gamma(m+\tilde{\alpha}) \left( \frac{1}{1+\tilde{\beta}} \right)^{\tilde{\alpha}+m} \\
 &= \frac{\Gamma(m+\tilde{\alpha})}{\Gamma(\tilde{\alpha})m!} \left( \frac{\tilde{\beta}}{1+\tilde{\beta}} \right)^{\tilde{\alpha}} \left( \frac{1}{1+\tilde{\beta}} \right)^m \\
 &= \frac{\Gamma(\tilde{\alpha}+m)}{\Gamma(\tilde{\alpha})m!} \left( \frac{\tilde{\beta}}{1+\tilde{\beta}} \right)^{\tilde{\alpha}} \left( \frac{1}{1+\tilde{\beta}} \right)^m \\
 &= \frac{\Gamma(m+\tilde{\alpha})}{\Gamma(\tilde{\alpha})\Gamma(m+1)} \left( \frac{\tilde{\beta}}{1+\tilde{\beta}} \right)^{\tilde{\alpha}} \left( \frac{1}{1+\tilde{\beta}} \right)^m \\
 &= \frac{\Gamma(m+\tilde{\alpha}-1+1)}{\Gamma(m+\tilde{\alpha}-1-m+1)\Gamma(m+1)} \left( \frac{\tilde{\beta}}{1+\tilde{\beta}} \right)^{\tilde{\alpha}} \left( \frac{1}{1+\tilde{\beta}} \right)^m \\
 &= \binom{m+\tilde{\alpha}-1}{m} \left( \frac{\tilde{\beta}}{1+\tilde{\beta}} \right)^{\tilde{\alpha}} \left( 1 - \frac{\tilde{\beta}}{1+\tilde{\beta}} \right)^m \\
 &= \binom{m+\tilde{\alpha}_{neu}-1}{m} (\tilde{\beta}_{neu})^{\tilde{\alpha}_{neu}} (1-\tilde{\beta}_{neu})^m \tag{A.3}
 \end{aligned}$$

mit  $\tilde{\alpha}_{neu} = \tilde{\alpha}$  - der Formparameter und  $\tilde{\beta}_{neu} = \frac{\tilde{\beta}}{1+\tilde{\beta}}$  der Skalenparameter der negativen Binomial-Verteilung.

<sup>6</sup>Die Informationen zur negativen Binomialverteilung siehe Anh. A.5.7.

### Herleitung zur Kap. 5.1.1 Formel 5.6

Die erwartete Anzahl der Verluste im folgenden Jahr:

$$\begin{aligned}
 E(N_{T+1}) &= \frac{\tilde{\alpha}_{neu}(1 - \tilde{\beta}_{neu})}{\tilde{\beta}_{neu}} = \frac{\tilde{\alpha} \left(1 - \frac{\tilde{\beta}}{1+\tilde{\beta}}\right)}{\frac{\tilde{\beta}}{1+\tilde{\beta}}} = \tilde{\alpha} \left(\frac{1}{1+\tilde{\beta}}\right) \left(\frac{1+\tilde{\beta}}{\tilde{\beta}}\right) = \frac{\tilde{\alpha}}{\tilde{\beta}} \\
 &= \frac{\alpha + \sum_{t=1}^T n_t}{\beta + T} = \frac{T\beta \left(\alpha + \sum_{t=1}^T n_t\right)}{T\beta(\beta + T)} = \frac{T\beta\alpha + T\beta \sum_{t=1}^T n_t}{T\beta(\beta + T)} = \frac{\beta\alpha}{\beta(\beta + T)} + \frac{T \sum_{t=1}^T n_t}{T(\beta + T)} \\
 &= \frac{\beta}{\beta + T} \frac{\alpha}{\beta} + \frac{T}{\beta + T} \frac{1}{T} \sum_{t=1}^T n_t = \frac{\beta}{\beta + T} \lambda_{a \text{ priori}} + \frac{T}{\beta + T} \lambda_{MLE} \\
 &= \omega \lambda_{MLE} + (1 - \omega) \lambda_{a \text{ priori}} \tag{A.4}
 \end{aligned}$$

mit

- $\lambda_{MLE} = \frac{1}{T} \sum_{t=1}^T n_t$  - der MLE für Parameter  $\lambda$  aus der beobachteten internen Daten.
- $\lambda_{a \text{ priori}} = \frac{\alpha}{\beta} = E(X)$ ,  $X \sim G(\alpha, \beta)$  - die Schätzung für Parameter  $\lambda$  aus der *a priori* auf Basis von externen Daten bzw. Szenarien.
- $\omega = \frac{T}{\beta + T}$ ,  $\omega \in [0, 1]$  - das Gewicht zur Kombination der beiden Schätzer.
- $\alpha$  - der Hyperparameter der *a priori* Gamma-Verteilung.
- $\beta$  - der Hyperparameter der *a priori* Gamma-Verteilung.
- $T$  - Anzahl der Jahre (Zeitperioden) in historisch beobachteten internen Daten.

## A.8. Bestimmung der gemeinsamen *a priori* und der dazugehörigen *a posteriori* Dichte bei Lognormal-Verteilung mit zwei unbekanntem abhängigen Parametern

### Herleitung zur Kap. 5.1.2 Formel 5.14

Die gemeinsame *a priori* der Parameter  $\mu$  und  $\sigma^2$  der Lognormal-Verteilung für  $\Theta_{\mu|\sigma^2} \sim N(\mu_0, \frac{\sigma^2}{\phi})$  und  $\Theta_{\sigma^2} \sim scaledI\chi^2(\nu, \tau^2)$  ist eine Mischverteilung aus der Normal- und der skalierten inversen Chi-Quadrat-Verteilung:

$$\begin{aligned}
 p(\mu, \sigma^2) &= p(\mu|\sigma^2)p(\sigma^2) = p(\mu|\mu_0, \frac{\sigma^2}{\phi}) \times p(\sigma^2|\nu, \tau^2) \\
 &= \frac{1}{\sqrt{2\pi\frac{\sigma^2}{\phi}}} \exp\left(-\frac{[\mu - \mu_0]^2}{2\frac{\sigma^2}{\phi}}\right) \times \frac{2^{-\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})\nu\tau^2} \left(\frac{\sigma^2}{\nu\tau^2}\right)^{-(1+\frac{\nu}{2})} \exp\left(-\frac{\nu\tau^2}{2\sigma^2}\right) \\
 &\stackrel{bzgl.\mu, \sigma^2}{\propto} (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{[\mu - \mu_0]^2}{2\sigma^2\frac{1}{\phi}}\right) \times (\sigma^2)^{-(1+\frac{\nu}{2})} \exp\left(-\frac{\nu\tau^2}{2\sigma^2}\right) \\
 &= (\sigma^2)^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{[\mu - \mu_0]^2}{2\sigma^2\frac{1}{\phi}} - \frac{\nu\tau^2}{2\sigma^2}\right) \\
 &= (\sigma^2)^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}(\phi[\mu - \mu_0]^2 + \nu\tau^2)\right) \tag{A.5}
 \end{aligned}$$

### Herleitung zur Kap. 5.1.2 Formel 5.15

Die *a posteriori* der Parameter  $\mu$  und  $\sigma^2$  der Lognormal-Verteilung für  $\Theta_{\mu|\sigma^2} \sim N(\mu_0, \frac{\sigma^2}{\phi})$  und  $\Theta_{\sigma^2} \sim scaledI\chi^2(\nu, \tau^2)$  ist eine Mischverteilung aus der Normal- und der skalierten inversen Chi-Quadrat-Verteilung mit neuen Hyperparametern:

$$\begin{aligned}
 p(\mu, \sigma^2 | \underline{loss}) &\propto \prod_{i=1}^n \frac{1}{loss_i \sqrt{2\pi\sigma^2}} \exp\left(-\frac{[\log(loss_i) - \mu]^2}{2\sigma^2}\right) \times \\
 &\quad \times (\sigma^2)^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}(\phi[\mu - \mu_0]^2 + \nu\tau^2)\right) \\
 &\stackrel{bzgl.\mu, \sigma^2}{\propto} (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n [\log(loss_i) - \mu]^2\right) \times \\
 &\quad \times (\sigma^2)^{-\frac{\nu+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2}(\phi[\mu - \mu_0]^2 + \nu\tau^2)\right) \\
 &= (\sigma^2)^{-\frac{\nu+n+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n [\log(loss_i) - \mu]^2 + \phi[\mu - \mu_0]^2 + \nu\tau^2\right)\right) \\
 &= (\sigma^2)^{-\frac{\nu+n+1}{2}-1} \times
 \end{aligned}$$

$$\begin{aligned}
& \times \exp \left( -\frac{1}{2\sigma^2} \left( \sum_{i=1}^n \log^2(\text{loss}_i) - 2\mu \sum_{i=1}^n \log(\text{loss}_i) + n\mu^2 + \phi\mu^2 - 2\phi\mu\mu_0 + \phi\mu_0^2 + \nu\tau^2 \right) \right) \\
& = (\sigma^2)^{-\frac{\nu+n+1}{2}-1} \times \\
& \times \exp \left( -\frac{1}{2\sigma^2} \left( \nu\tau^2 + \phi\mu_0^2 + n\overline{\text{loss}^2} - \frac{(n\overline{\text{loss}} + \phi\mu_0)^2}{\phi+n} + (\phi+n) \left[ \mu - \frac{n\overline{\text{loss}} + \phi\mu_0}{\phi+n} \right]^2 \right) \right) \\
& = (\sigma^2)^{-\frac{\nu_{neu}+1}{2}-1} \exp \left( -\frac{1}{2\sigma^2} \left( \phi_{neu} [\mu - \mu_{0,neu}]^2 + (\nu\tau^2)_{neu} \right) \right) \quad (\text{A.6})
\end{aligned}$$

mit

- $\nu_{neu} = \nu + n$
- $(\nu\tau^2)_{neu} = \nu_{neu}\tau_{neu}^2 = \nu\tau^2 + \phi\mu_0^2 + n\overline{\text{loss}^2} - \frac{(n\overline{\text{loss}} + \phi\mu_0)^2}{\phi+n}$
- $\tau_{neu}^2 = \frac{(\nu\tau^2)_{neu}}{\nu_{neu}}$
- $\mu_{0,neu} = \frac{n\overline{\text{loss}} + \phi\mu_0}{\phi+n}$
- $\phi_{neu} = \phi + n$
- $\overline{\text{loss}} = \frac{1}{n} \sum_{i=1}^n \log(\text{loss}_i)$
- $\overline{\text{loss}^2} = \frac{1}{n} \sum_{i=1}^n \log^2(\text{loss}_i)$

## A.9. Bestimmung der Jeffreys *a priori* für Lognormal-Verteilung und der dazugehörigen *a posteriori* Dichte mit zwei unbekanntem abhängigen Parametern

Herleitung zur Kap. 5.1.2 Formel 5.19

Jeffreys *a priori*

$$\begin{aligned}
p_J(\mu, \sigma^2) &= \sqrt{\det(I(\mu, \sigma^2))} = \sqrt{\det \left( -E \left( \frac{\partial^2}{\partial^2(\mu, \sigma^2)} \log(f(\underline{x}|\mu, \sigma^2)) \right) \right)} \\
&= \sqrt{\det \left( -E \left[ \begin{array}{cc} \frac{\partial^2}{\partial^2\mu} \log(f(\underline{x}|\mu, \sigma^2)) & \frac{\partial^2}{\partial\mu\partial\sigma^2} \log(f(\underline{x}|\mu, \sigma^2)) \\ \frac{\partial^2}{\partial\sigma^2\partial\mu} \log(f(\underline{x}|\mu, \sigma^2)) & \frac{\partial^2}{\partial^2\sigma^2} \log(f(\underline{x}|\mu, \sigma^2)) \end{array} \right] \right)}
\end{aligned}$$

$$\begin{aligned}
& \stackrel{\substack{\text{A.10} \\ \text{A.11} \\ \text{A.13} \\ \text{A.14}}}{=} \sqrt{\det \left( -E \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{(\sigma^2)^2} \sum_{i=1}^n \log(x_i) + \frac{n\mu}{(\sigma^2)^2} \\ -\frac{1}{(\sigma^2)^2} \sum_{i=1}^n \log(x_i) + \frac{n\mu}{(\sigma^2)^2} & \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^n [\log(x_i) - \mu]^2 \end{bmatrix} \right)} \\
&= \sqrt{\det \begin{bmatrix} \frac{n}{\sigma^2} & \frac{1}{(\sigma^2)^2} \sum_{i=1}^n E(\log(x_i)) - \frac{n\mu}{(\sigma^2)^2} \\ \frac{1}{(\sigma^2)^2} \sum_{i=1}^n E(\log(x_i)) - \frac{n\mu}{(\sigma^2)^2} & -\frac{n}{2(\sigma^2)^2} + \frac{n}{(\sigma^2)^3} E\left(\frac{1}{n} \sum_{i=1}^n [\log(x_i) - \mu]^2\right) \end{bmatrix}} \\
&= \sqrt{\det \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n\mu}{(\sigma^2)^2} - \frac{n\mu}{(\sigma^2)^2} \\ \frac{n\mu}{(\sigma^2)^2} - \frac{n\mu}{(\sigma^2)^2} & -\frac{n}{2(\sigma^2)^2} + \frac{n\sigma^2}{(\sigma^2)^3} \end{bmatrix}} = \sqrt{\det \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2(\sigma^2)^2} \end{bmatrix}} \\
&= \sqrt{\frac{n}{\sigma^2} \frac{n}{2(\sigma^2)^2}} = \sqrt{\frac{n^2}{2(\sigma^2)^3}} \propto \frac{1}{\sqrt{(\sigma^2)^3}} = (\sigma^2)^{-\frac{3}{2}} \tag{A.7}
\end{aligned}$$

### Die Log-Likelihood

$$\begin{aligned}
\log(f(\underline{x}|\mu, \sigma^2)) &= \log \left( \prod_{i=1}^n \frac{1}{x_i \sqrt{2\pi\sigma^2}} \exp \left( -\frac{[\log(x_i) - \mu]^2}{2\sigma^2} \right) \right) \\
&= \log \left( (2\pi\sigma^2)^{-\frac{n}{2}} \prod_{i=1}^n (x_i)^{-1} \exp \left( -\frac{[\log(x_i) - \mu]^2}{2\sigma^2} \right) \right) \\
&= \log((2\pi\sigma^2)^{-\frac{n}{2}}) + \sum_{i=1}^n \log \left( (x_i)^{-1} \exp \left( -\frac{[\log(x_i) - \mu]^2}{2\sigma^2} \right) \right) \\
&= \log((2\pi\sigma^2)^{-\frac{n}{2}}) + \sum_{i=1}^n \log \left( (x_i)^{-1} \right) + \sum_{i=1}^n \log \left( \exp \left( -\frac{[\log(x_i) - \mu]^2}{2\sigma^2} \right) \right) \\
&= -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \log(x_i) - \frac{1}{2\sigma^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \tag{A.8}
\end{aligned}$$

### Die 1. Ableitung nach $\mu$

$$\begin{aligned}
\frac{\partial}{\partial \mu} \log(f(\underline{x}|\mu, \sigma^2)) &\stackrel{\text{A.8}}{=} \frac{\partial}{\partial \mu} \left( -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \log(x_i) - \frac{1}{2\sigma^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \right) \\
&= \frac{\partial}{\partial \mu} \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \right) = \frac{\partial}{\partial \mu} \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n \log^2(x_i) + \frac{2\mu}{2\sigma^2} \sum_{i=1}^n \log(x_i) - \frac{n\mu^2}{2\sigma^2} \right) \\
&= \frac{\partial}{\partial \mu} \left( \frac{\mu}{\sigma^2} \sum_{i=1}^n \log(x_i) - \frac{n\mu^2}{2\sigma^2} \right) = \frac{1}{\sigma^2} \sum_{i=1}^n \log(x_i) - \frac{n\mu}{\sigma^2} \tag{A.9}
\end{aligned}$$

**Die 2. Ableitung nach  $\mu$**

$$\begin{aligned}\frac{\partial^2}{\partial^2 \mu} \log(f(\underline{x}|\mu, \sigma^2)) &\stackrel{\text{A.9}}{=} \frac{\partial}{\partial \mu} \left( \frac{1}{\sigma^2} \sum_{i=1}^n \log(x_i) - \frac{n\mu}{\sigma^2} \right) \\ &= \frac{\partial}{\partial \mu} \left( -\frac{n\mu}{\sigma^2} \right) = -\frac{n}{\sigma^2}\end{aligned}\tag{A.10}$$

**Die 2. Ableitung nach  $\mu$  und  $\sigma^2$**

$$\begin{aligned}\frac{\partial^2}{\partial \mu \partial \sigma^2} \log(f(\underline{x}|\mu, \sigma^2)) &\stackrel{\text{A.9}}{=} \frac{\partial}{\partial \sigma^2} \left( \frac{1}{\sigma^2} \sum_{i=1}^n \log(x_i) - \frac{n\mu}{\sigma^2} \right) \\ &= -\frac{1}{(\sigma^2)^2} \sum_{i=1}^n \log(x_i) + \frac{n\mu}{(\sigma^2)^2}\end{aligned}\tag{A.11}$$

**Die 1. Ableitung nach  $\sigma^2$**

$$\begin{aligned}\frac{\partial}{\partial \sigma^2} \log(f(\underline{x}|\mu, \sigma^2)) &\stackrel{\text{A.8}}{=} \frac{\partial}{\partial \sigma^2} \left( -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \log(x_i) - \frac{1}{2\sigma^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \right) \\ &= \frac{\partial}{\partial \sigma^2} \left( -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \right) = -\frac{n}{2} \frac{1}{2\pi\sigma^2} 2\pi + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n [\log(x_i) - \mu]^2\end{aligned}\tag{A.12}$$

**Die 2. Ableitung nach  $\sigma^2$**

$$\begin{aligned}\frac{\partial}{\partial^2 \sigma^2} \log(f(\underline{x}|\mu, \sigma^2)) &\stackrel{\text{A.12}}{=} \frac{\partial}{\partial \sigma^2} \left( -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \right) \\ &= \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^n [\log(x_i) - \mu]^2\end{aligned}\tag{A.13}$$

**Die 2. Ableitung nach  $\sigma^2$  und  $\mu$**

$$\begin{aligned}\frac{\partial^2}{\partial \sigma^2 \partial \mu} \log(f(\underline{x}|\mu, \sigma^2)) &\stackrel{\text{A.12}}{=} \frac{\partial}{\partial \mu} \left( -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n [\log(x_i) - \mu]^2 \right) \\ &= \frac{\partial}{\partial \mu} \left( \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n \log^2(x_i) - \frac{2\mu}{2(\sigma^2)^2} \sum_{i=1}^n \log(x_i) + \frac{n\mu^2}{2(\sigma^2)^2} \right) \\ &= \frac{\partial}{\partial \mu} \left( -\frac{\mu}{(\sigma^2)^2} \sum_{i=1}^n \log(x_i) + \frac{n\mu^2}{2(\sigma^2)^2} \right) \\ &= -\frac{1}{(\sigma^2)^2} \sum_{i=1}^n \log(x_i) + \frac{n\mu}{(\sigma^2)^2}\end{aligned}\tag{A.14}$$

## Herleitung zur Kap. 5.1.2

Die *a posteriori* der Parameter  $\mu$  und  $\sigma^2$  der Lognormal-Verteilung und Jeffreys *a priori* ist eine Mischverteilung aus der Normal-  $\Theta_{\mu|\sigma^2} \sim N(\mu_{0,neu}, \frac{\sigma^2}{\phi_{neu}})$  und der skalierten inversen Chi-Quadrat-Verteilung  $\Theta_{\sigma^2} \sim scaledI\chi^2(\nu_{neu}, \tau_{neu}^2)$  mit neuen Hyperparametern<sup>7</sup>:

$$\begin{aligned}
 p(\mu, \sigma^2 | \underline{x}) &\propto \text{Jeffreys } a \text{ priori} \times \text{Likelihood} = (\sigma^2)^{-\frac{3}{2}} \times \prod_{i=1}^n \frac{1}{x_i \sqrt{2\pi\sigma^2}} \exp\left(-\frac{[\log(x_i) - \mu]^2}{2\sigma^2}\right) \\
 &\stackrel{\text{bzgl. } \mu}{\propto} (\sigma^2)^{-\frac{3}{2}} \prod_{i=1}^n (\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{[\log(x_i) - \mu]^2}{2\sigma^2}\right) = (\sigma^2)^{-\frac{3}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n [\log(x_i) - \mu]^2\right) \\
 &= (\sigma^2)^{-\frac{n+3}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n [\log^2(x_i) - 2\mu \log(x_i) + \mu^2]\right) \\
 &= (\sigma^2)^{-\frac{n+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2} (n\bar{x}^2 - 2\mu n\bar{x} + n\mu^2)\right) \\
 &= (\sigma^2)^{-\frac{n+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2} (n(\bar{x}^2 - 2\mu\bar{x} + \bar{x}^2 - \bar{x}^2 + \mu^2))\right) \\
 &= (\sigma^2)^{-\frac{n+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2} (n(\mu^2 - 2\mu\bar{x} + \bar{x}^2) + n\bar{x}^2 - n\bar{x}^2)\right) \\
 &= (\sigma^2)^{-\frac{n+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2} (n(\mu - \bar{x})^2 + n\bar{x}^2 - n\bar{x}^2)\right) \\
 &= (\sigma^2)^{-\frac{n+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2} (n(\mu - \bar{x})^2 + n(\bar{x}^2 - \bar{x}^2))\right) \\
 &= (\sigma^2)^{-\frac{\nu_{neu}+1}{2}-1} \exp\left(-\frac{1}{2\sigma^2} (\phi_{neu}(\mu - \mu_{0,neu})^2 + (\nu\tau^2)_{neu})\right) \tag{A.15}
 \end{aligned}$$

mit

- $\nu_{neu} = n$
- $(\nu\tau^2)_{neu} = \nu_{neu}\tau_{neu}^2 = n(\bar{x}^2 - \bar{x}^2)$
- $\tau_{neu}^2 = \bar{x}^2 - \bar{x}^2$
- $\mu_{0,neu} = \bar{x}$
- $\phi_{neu} = n$
- $\bar{x} = \frac{1}{n} \sum_{i=1}^n \log(x_i)$
- $\bar{x}^2 = \frac{1}{n} \sum_{i=1}^n \log^2(x_i)$
- $\bar{x}^2 = \left(\frac{1}{n} \sum_{i=1}^n \log(x_i)\right)^2$

---

<sup>7</sup>Vgl. Anh. A.8.

## A.10. Kombination von zwei Datenquellen. Ergebnisse zur Kap. 5.1

### A.10.1. Interne Daten: 100, externe Daten 1350, Simulationen: 1500

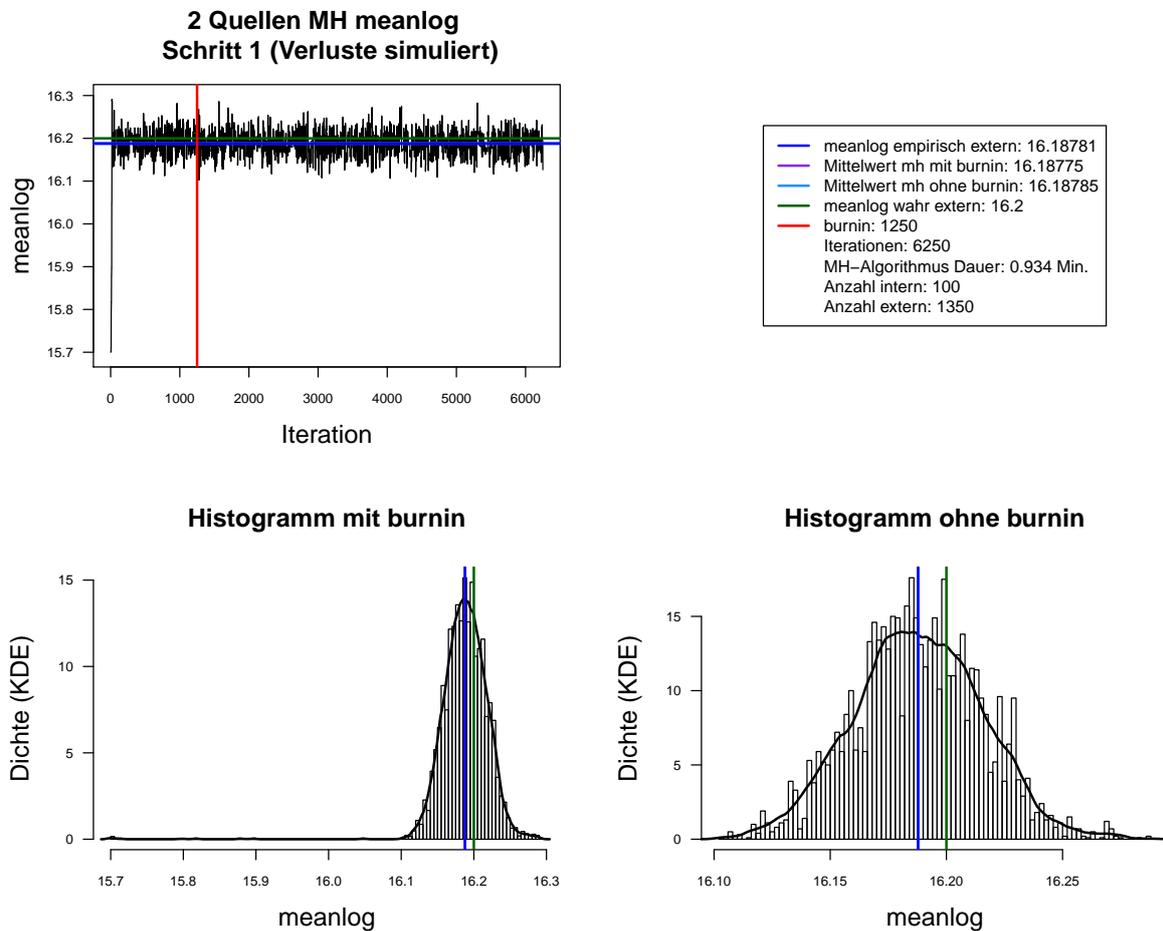


Abbildung A.37.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 0.934 Minuten. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 100. Externe Daten: 1350.

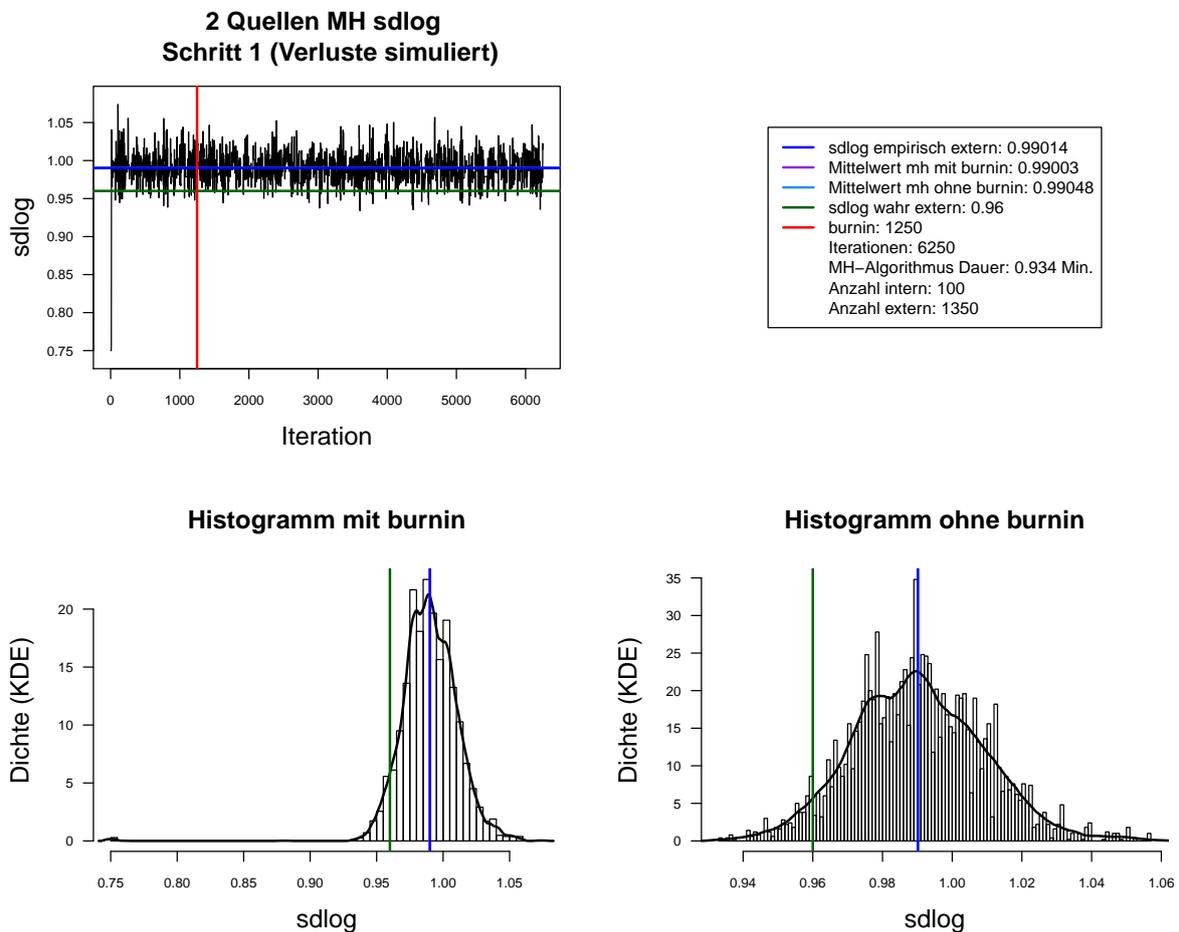
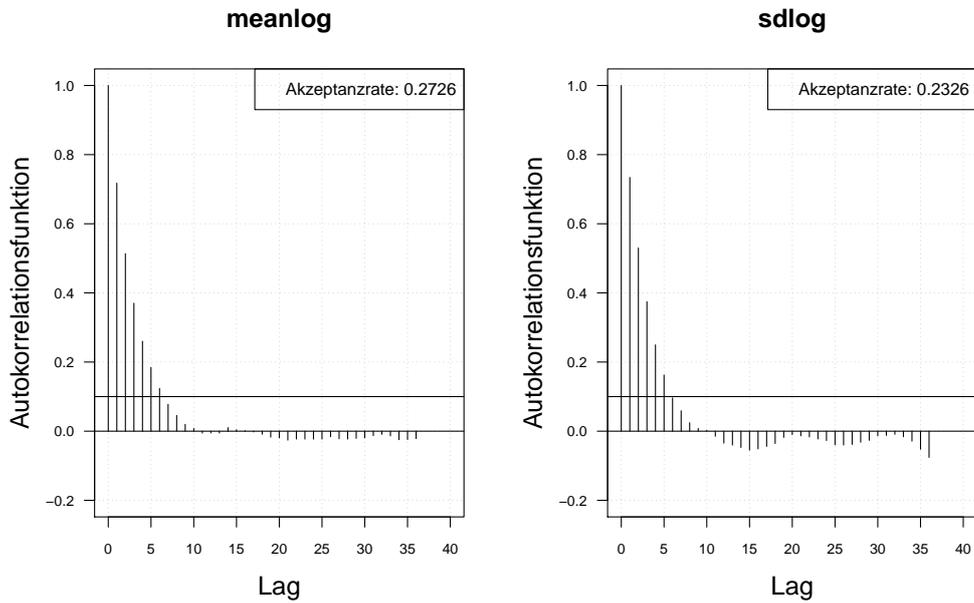


Abbildung A.38.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 0.943 Minuten. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 100. Externe Daten: 1350.

Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
Iterationen: 5000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 5, Iterationen: 5000

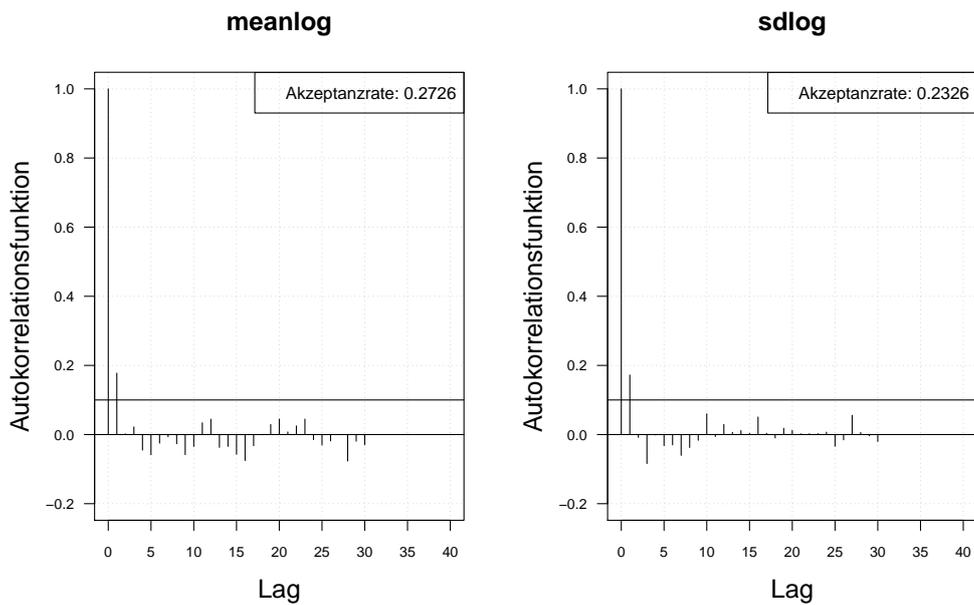


Abbildung A.39.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 5000. Schritt 1. Interne Daten: 100. Externe Daten: 1350.

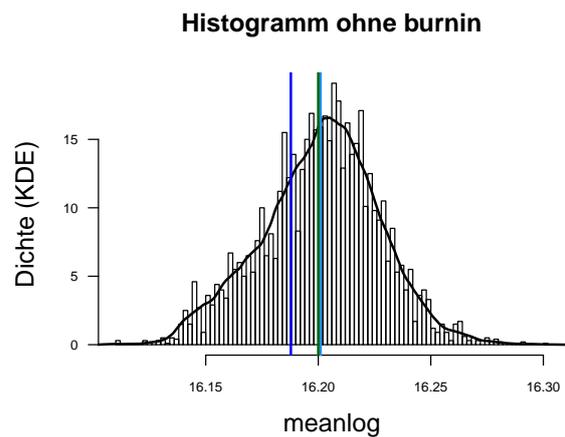
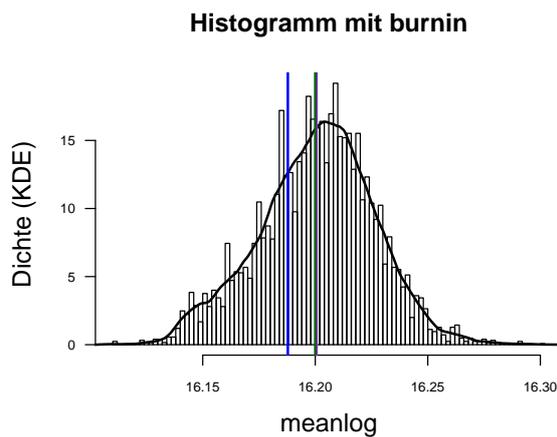
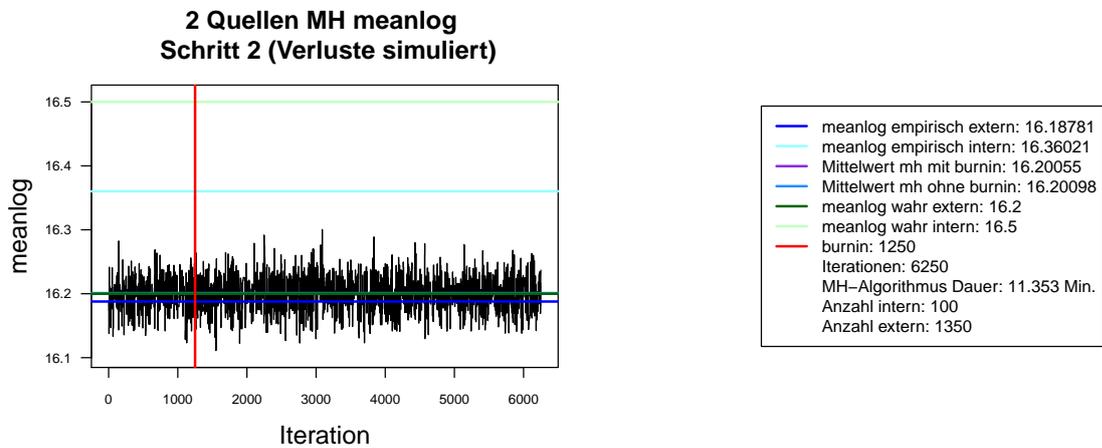


Abbildung A.40.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 11.353 Minuten. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 1350.

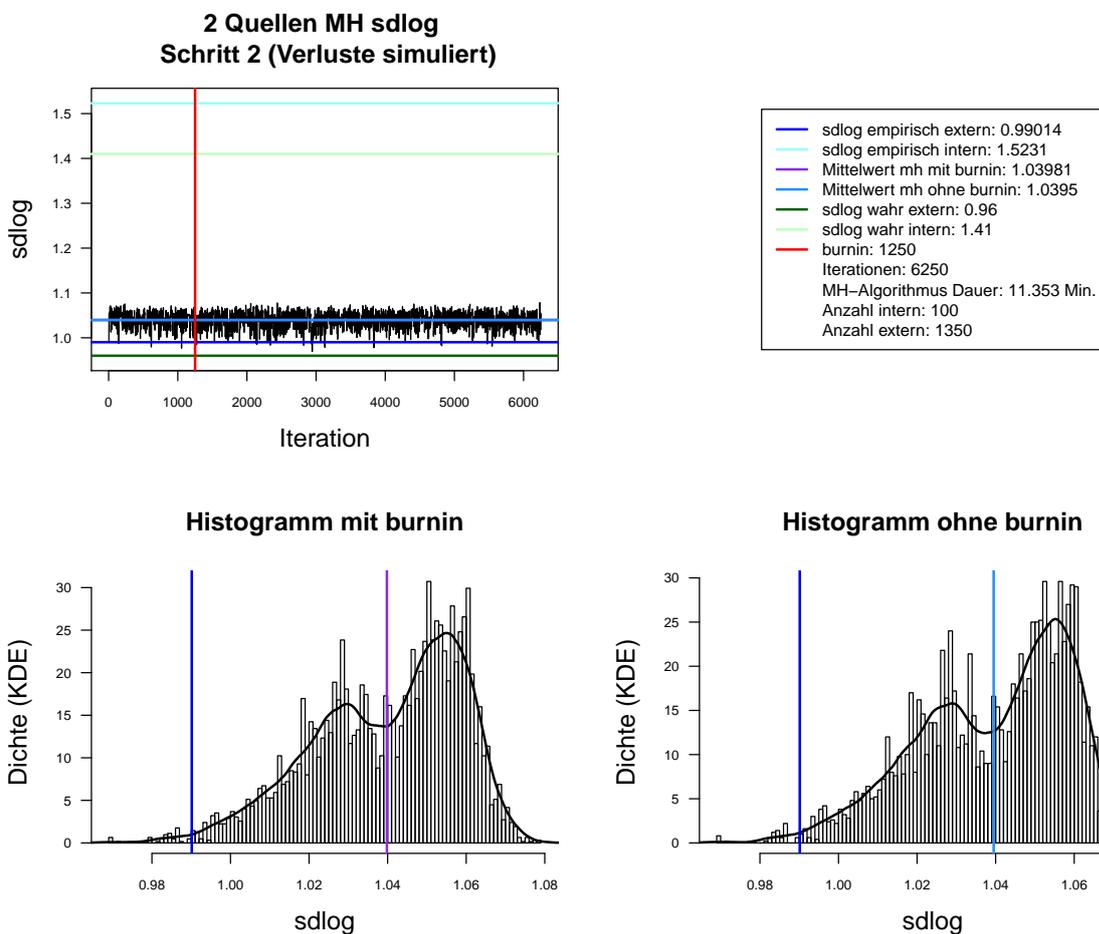
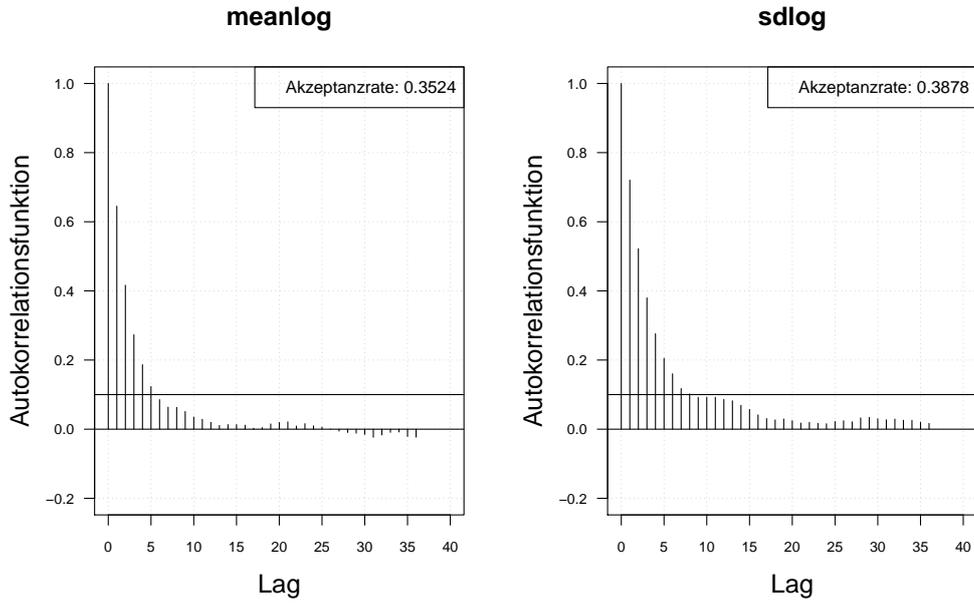


Abbildung A.41.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 6250 Iterationen beträgt 11.353 Minuten. Markov-Kette (oben links), Histogramm für 6250 Iterationen mit KDE (unten links) und Histogramm für letzte 5000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 1350.

Autokorrelation 2 Quellen Schritt 2. Ohne burnin (Verluste simuliert),  
Iterationen: 5000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 5, Iterationen: 5000

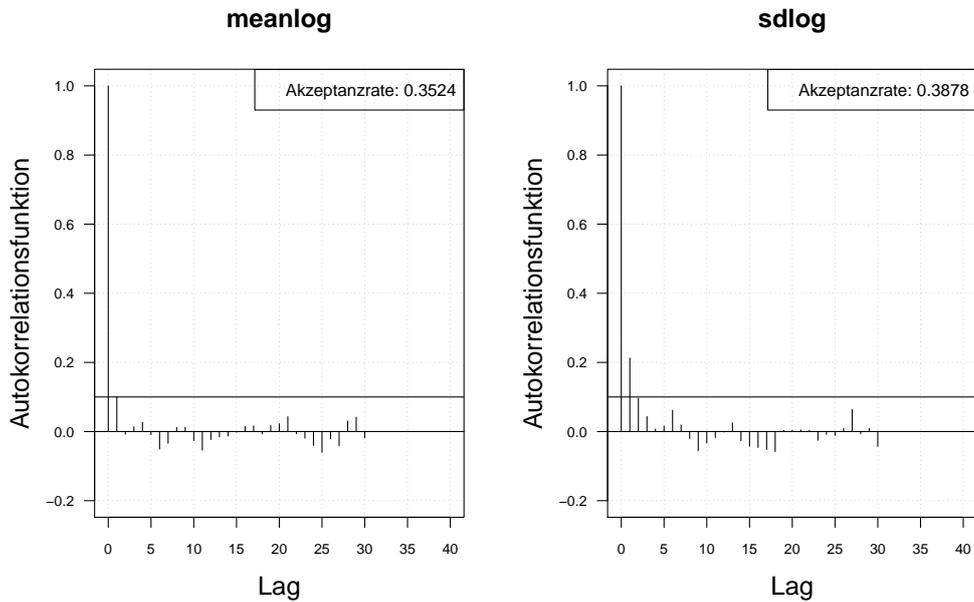


Abbildung A.42.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 5000. Schritt 2. Interne Daten: 100. Externe Daten: 1350.

## 2 Quellen Log-log-Plot simulierte Verluste (tail)

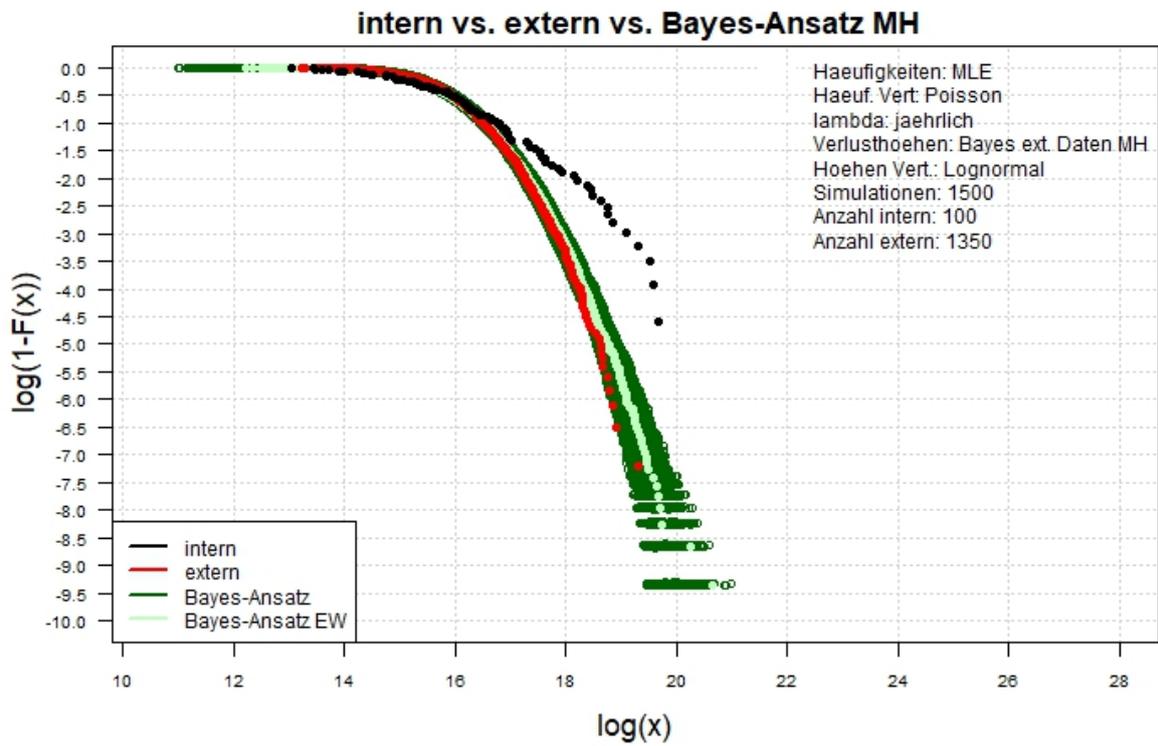


Abbildung A.43.: Der Log-log-Plot. Bayes-Ansatz MH. Zwei-Schritt-Verfahren. Interne Daten: 100. Externe Daten: 1350.

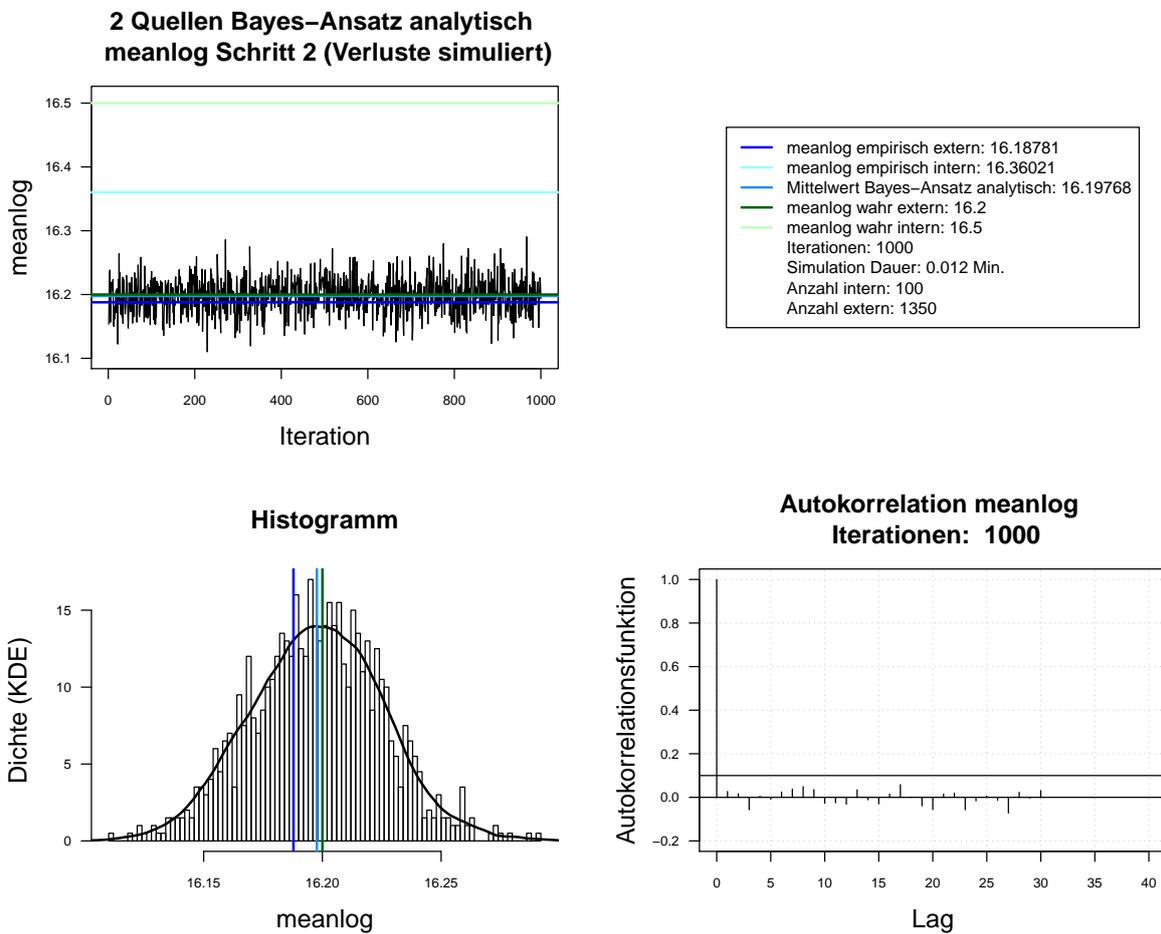


Abbildung A.44.: Der mit analytischen Herleitung simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.012 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 1350.

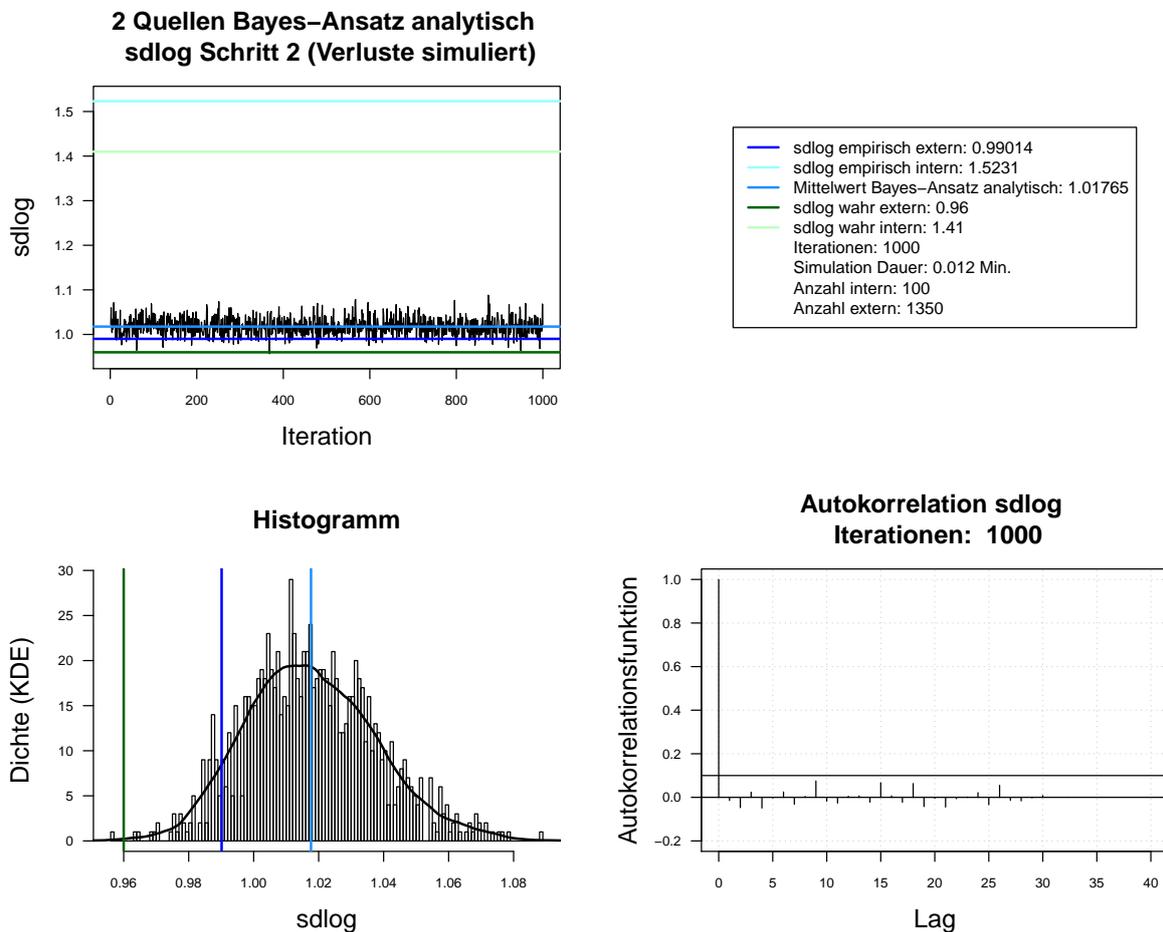


Abbildung A.45.: Der mit analytischen Herleitung simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.012 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 1350.

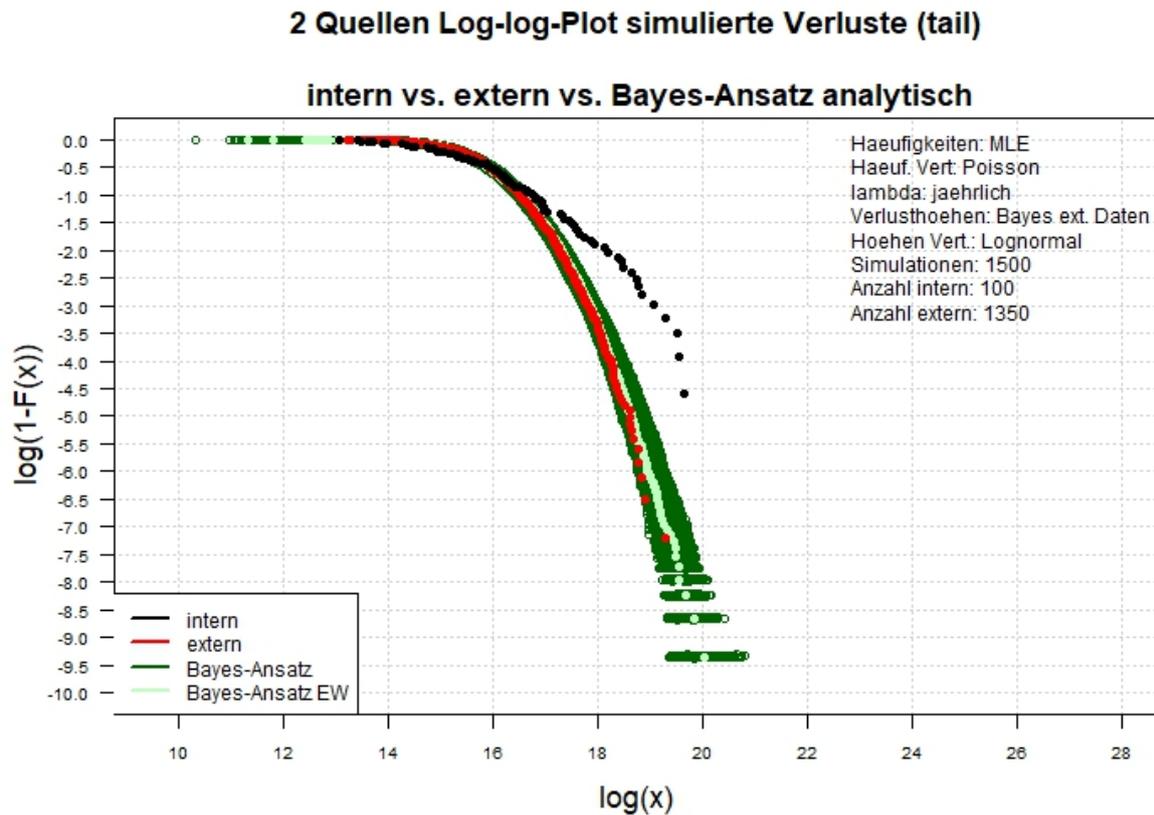


Abbildung A.46.: Der Log-log-Plot. Bayes-Ansatz analytisch. Zwei-Schritt-Verfahren. Interne Daten: 100. Externe Daten: 1350.

## A.10.2. Interne Daten: 300, externe Daten 3000, Simulationen: 1500

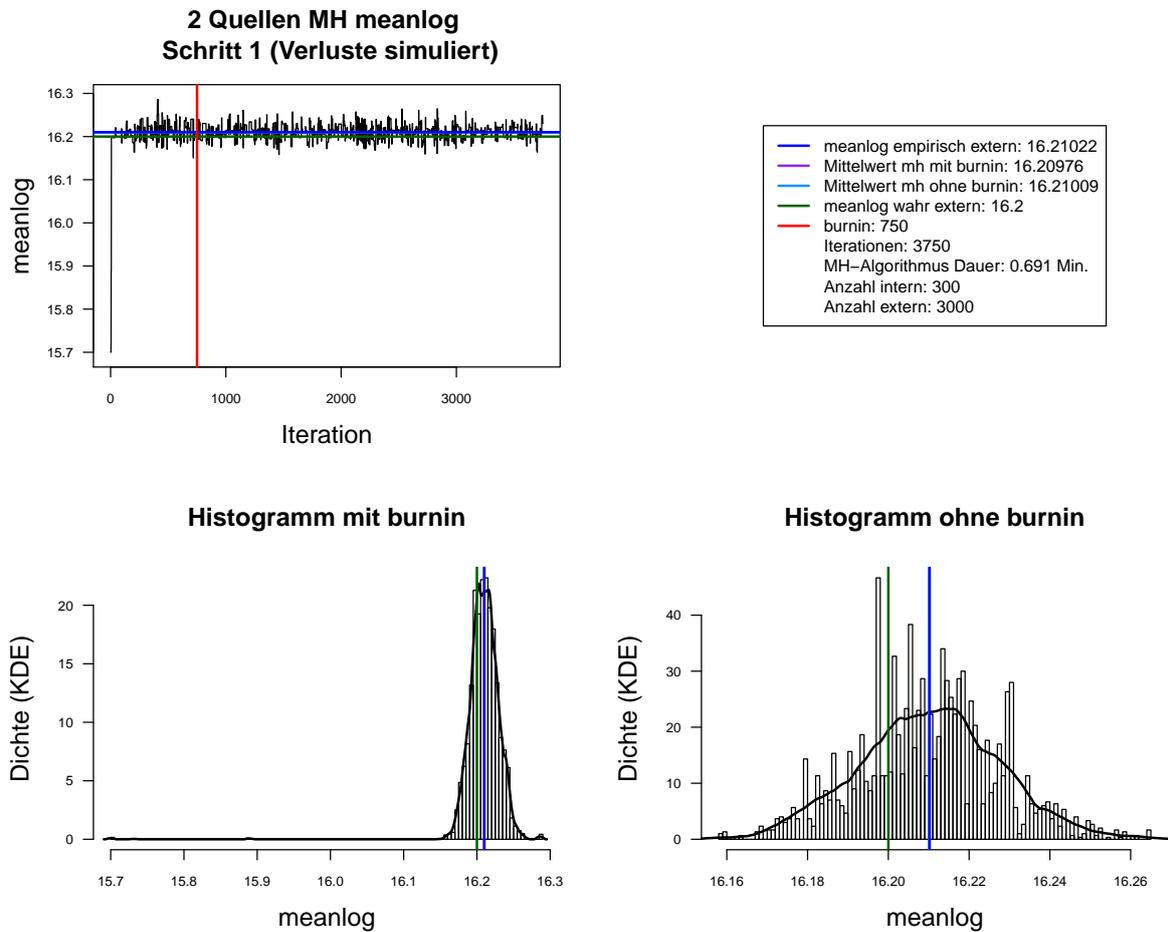


Abbildung A.47.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.691 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 300. Externe Daten: 3000.

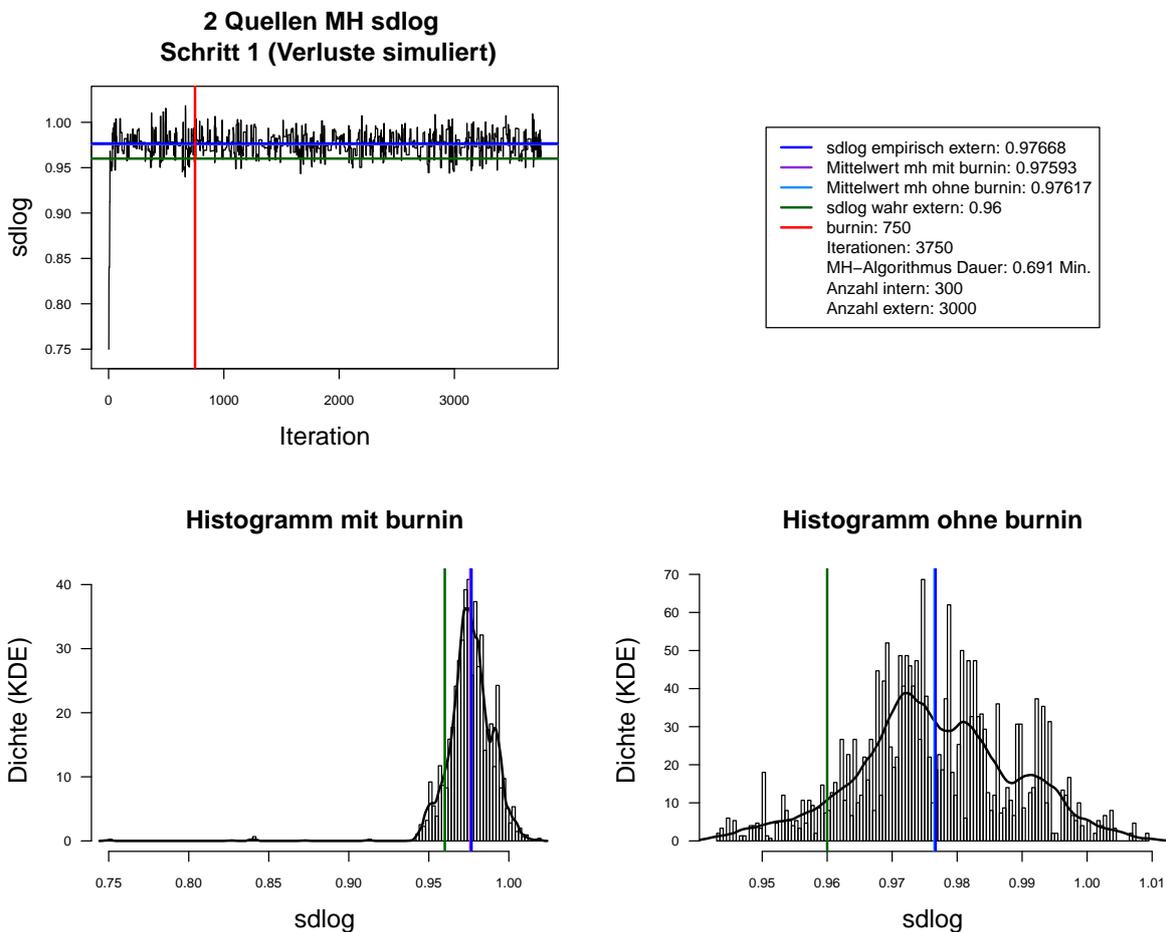
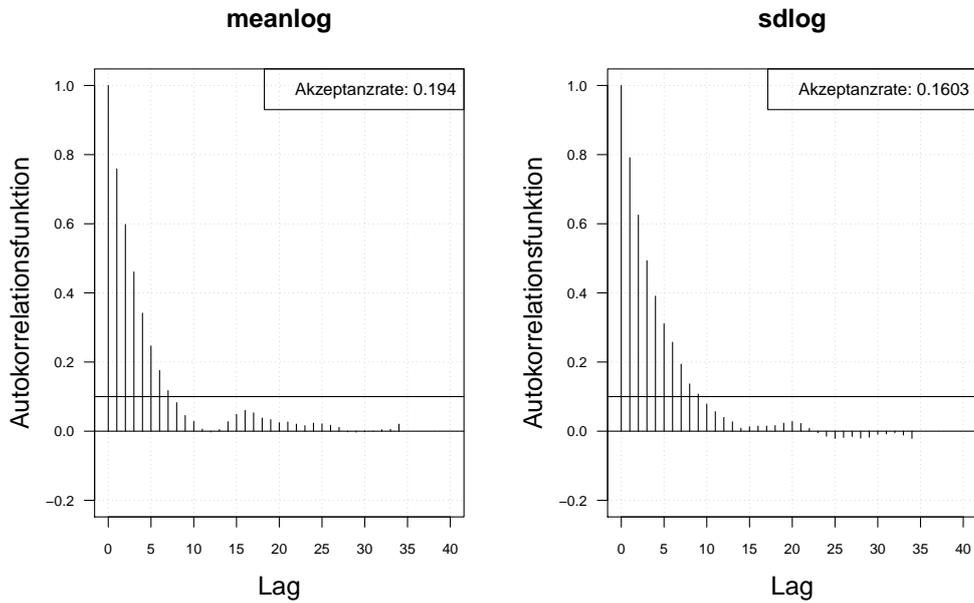


Abbildung A.48.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.691 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 300. Externe Daten: 3000.

Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

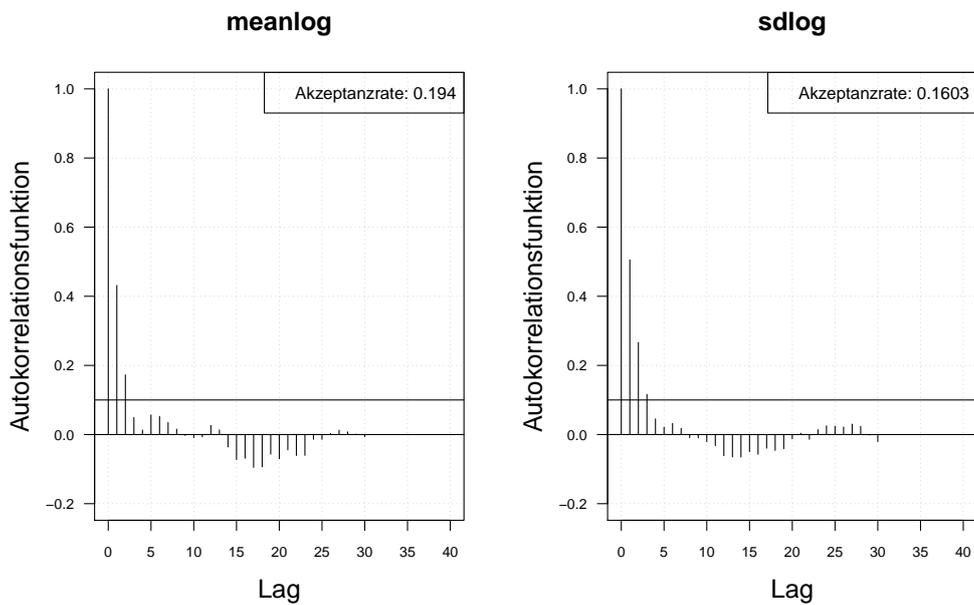


Abbildung A.49.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 1. Interne Daten: 300. Externe Daten: 3000.

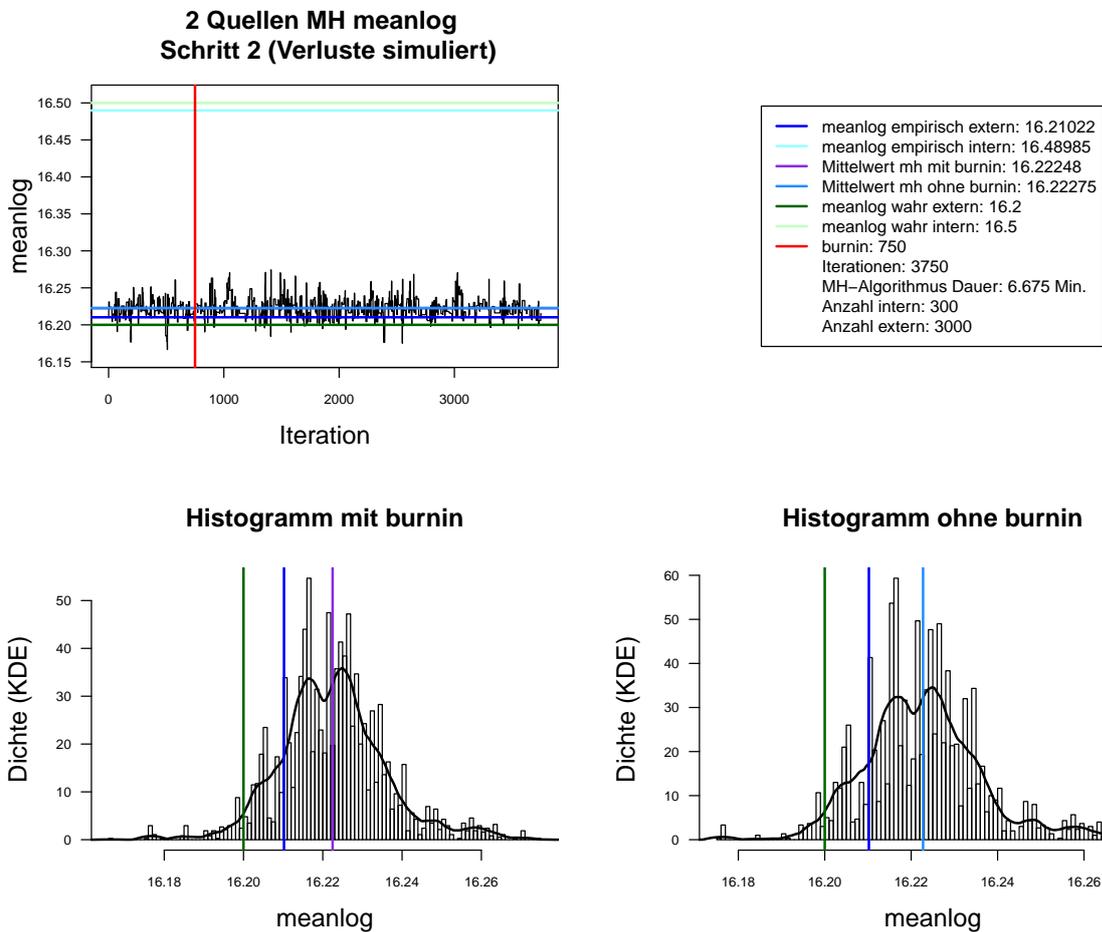


Abbildung A.50.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 6.675 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 3000.

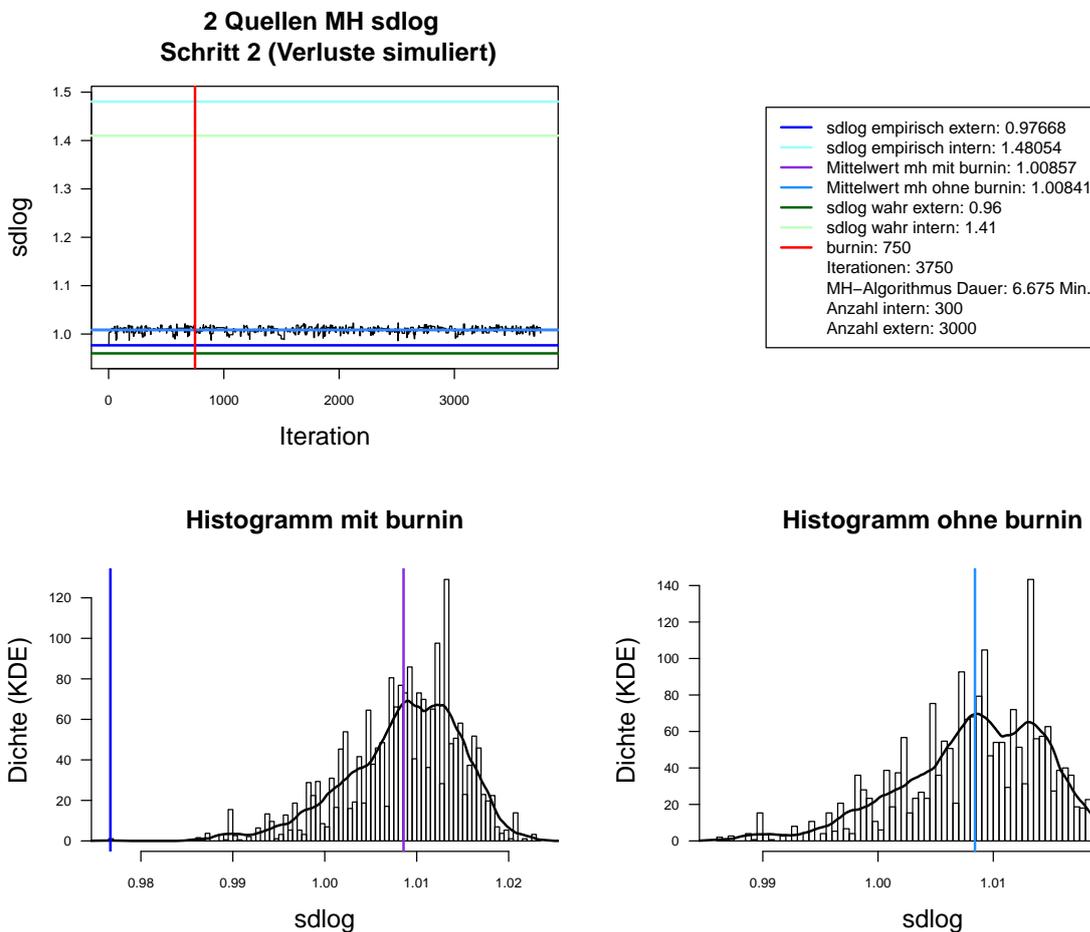
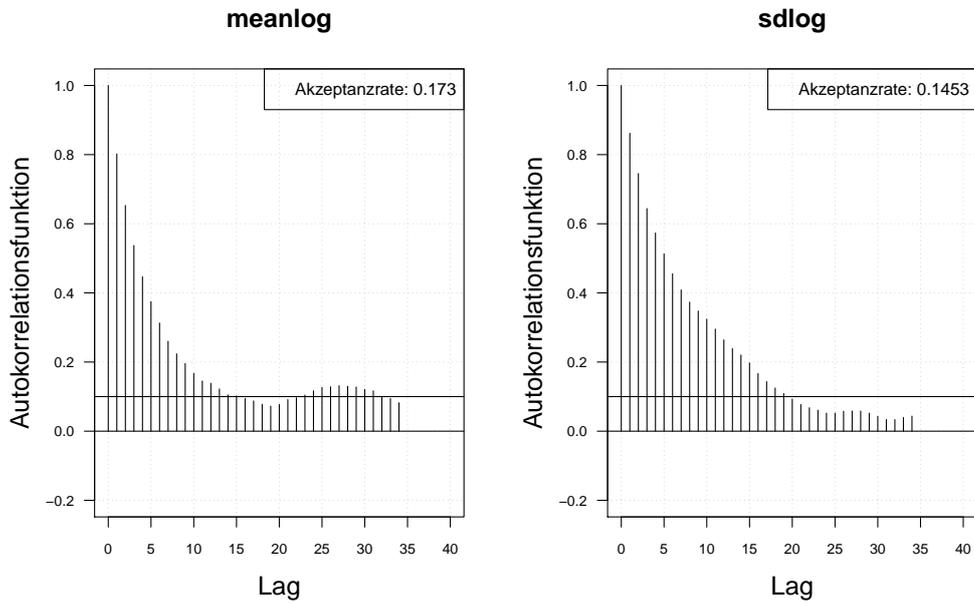


Abbildung A.51.: Der mit MH-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 6.675 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 3000.

Autokorrelation 2 Quellen Schritt 2. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

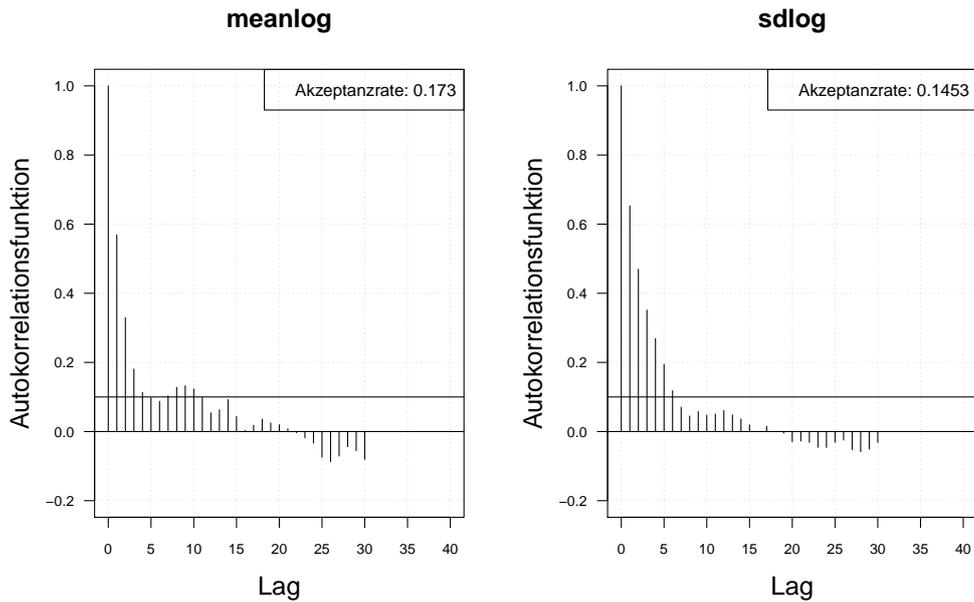


Abbildung A.52.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 2. Interne Daten: 300. Externe Daten: 3000.

## 2 Quellen Log-log-Plot simulierte Verluste (tail)

### intern vs. extern vs. Bayes-Ansatz MH

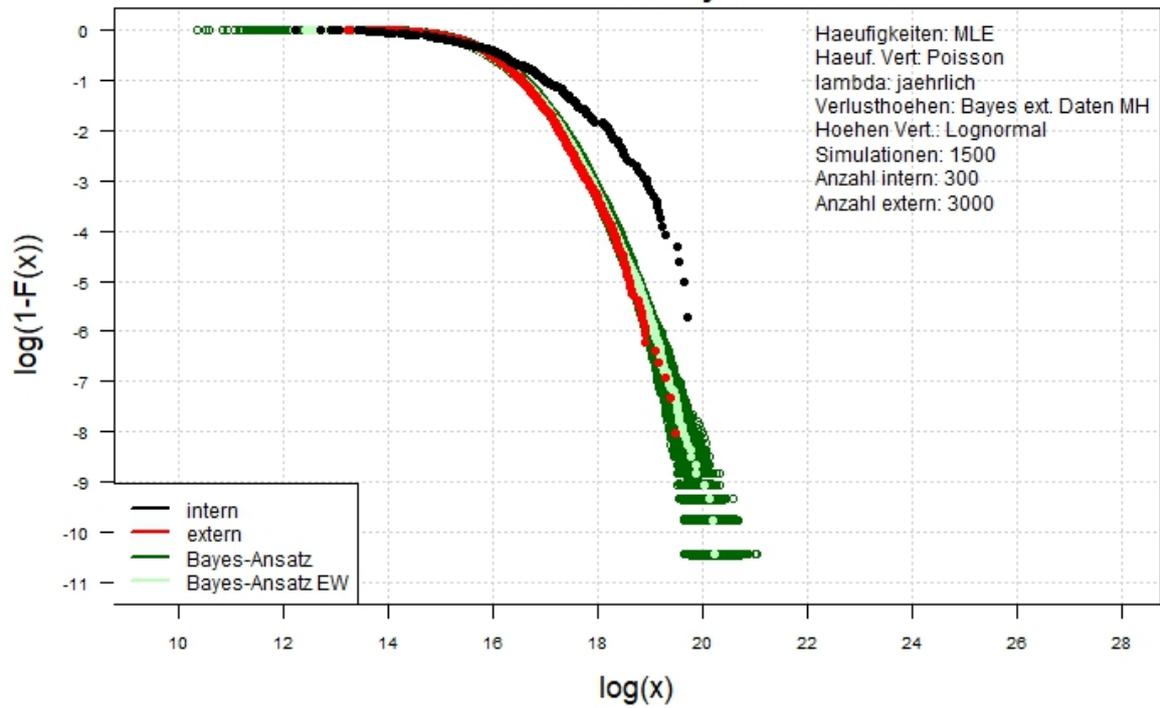


Abbildung A.53.: Der Log-log-Plot. Bayes-Ansatz MH. Zwei-Schritt-Verfahren. Interne Daten: 300. Externe Daten: 3000.

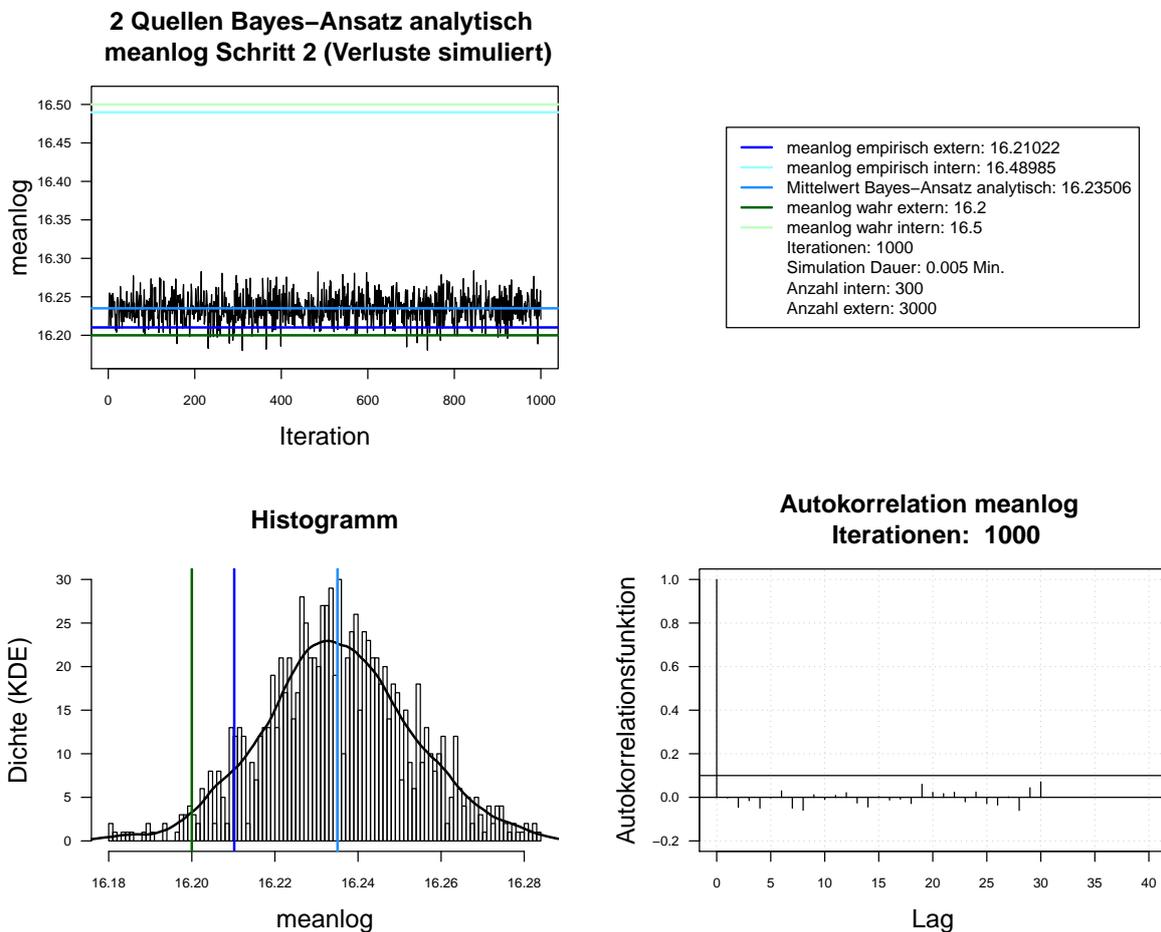


Abbildung A.54.: Der mit analytischen Herleitung simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 3000.

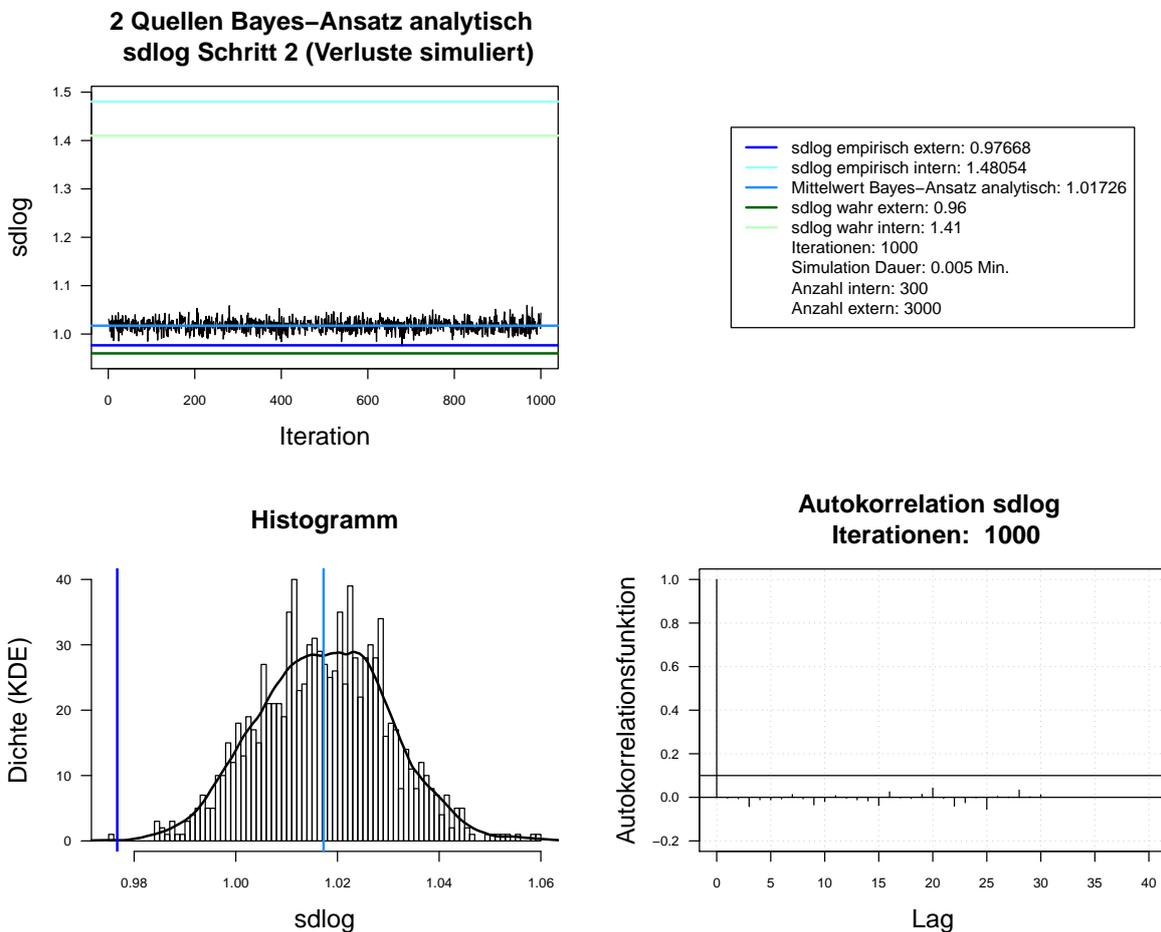


Abbildung A.55.: Der mit analytischen Herleitung simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 3000.

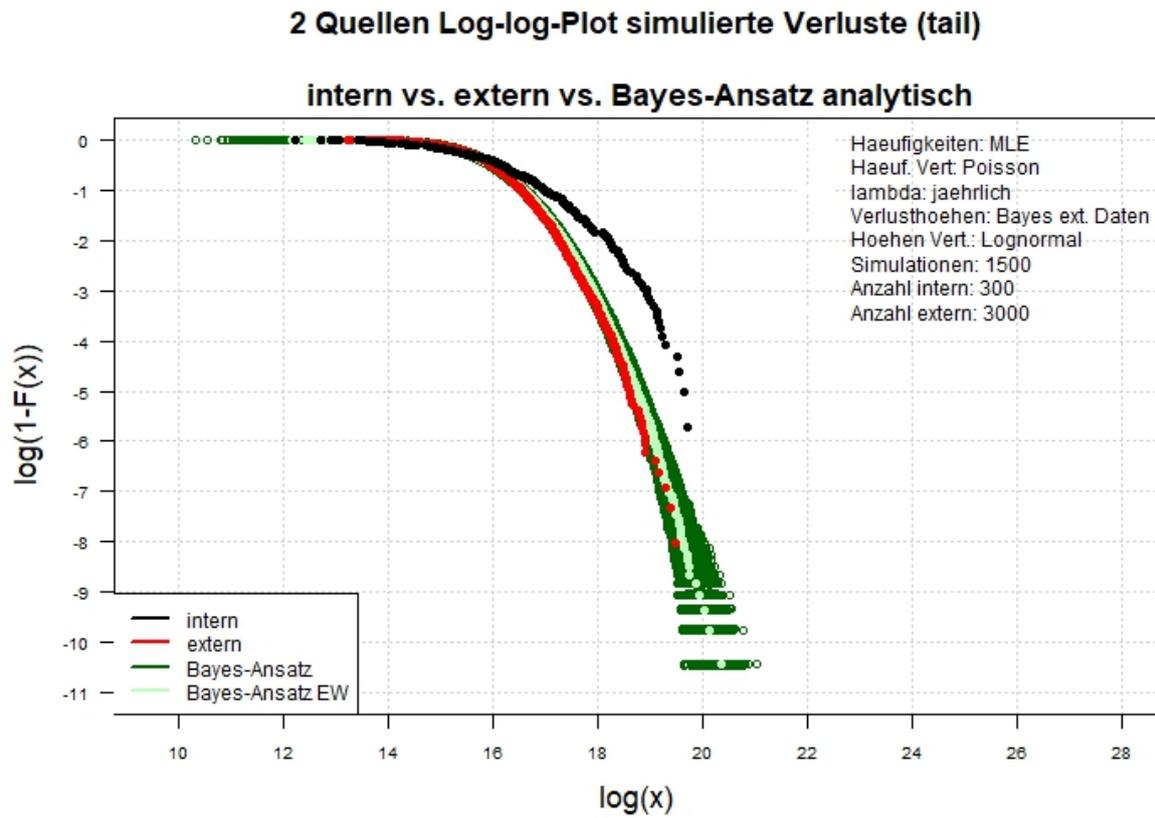


Abbildung A.56.: Der Log-log-Plot. Bayes-Ansatz analytisch. Zwei-Schritt-Verfahren. Interne Daten: 300. Externe Daten: 3000.

### A.10.3. Interne Daten: 100, externe Daten 300, Simulationen: 1500

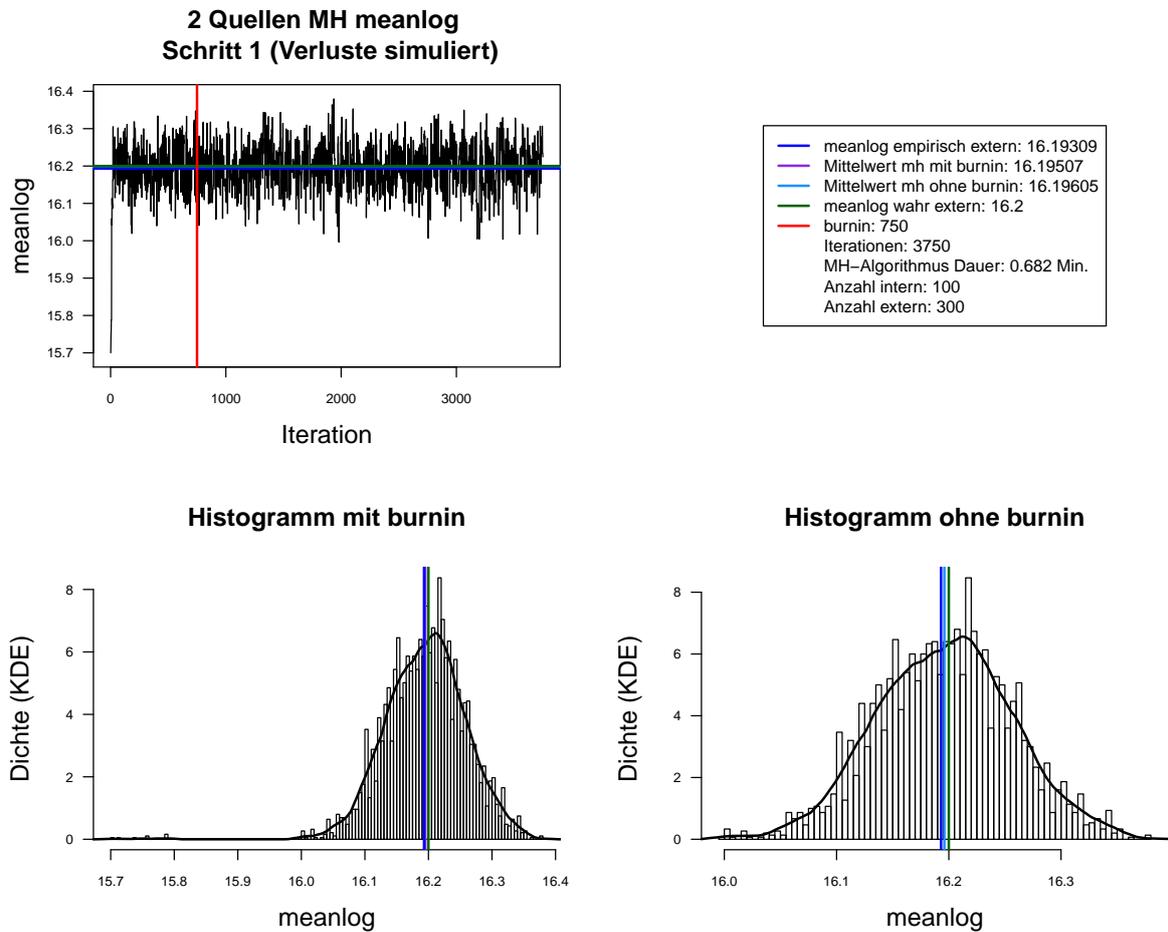


Abbildung A.57.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.682 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 100. Externe Daten: 300.

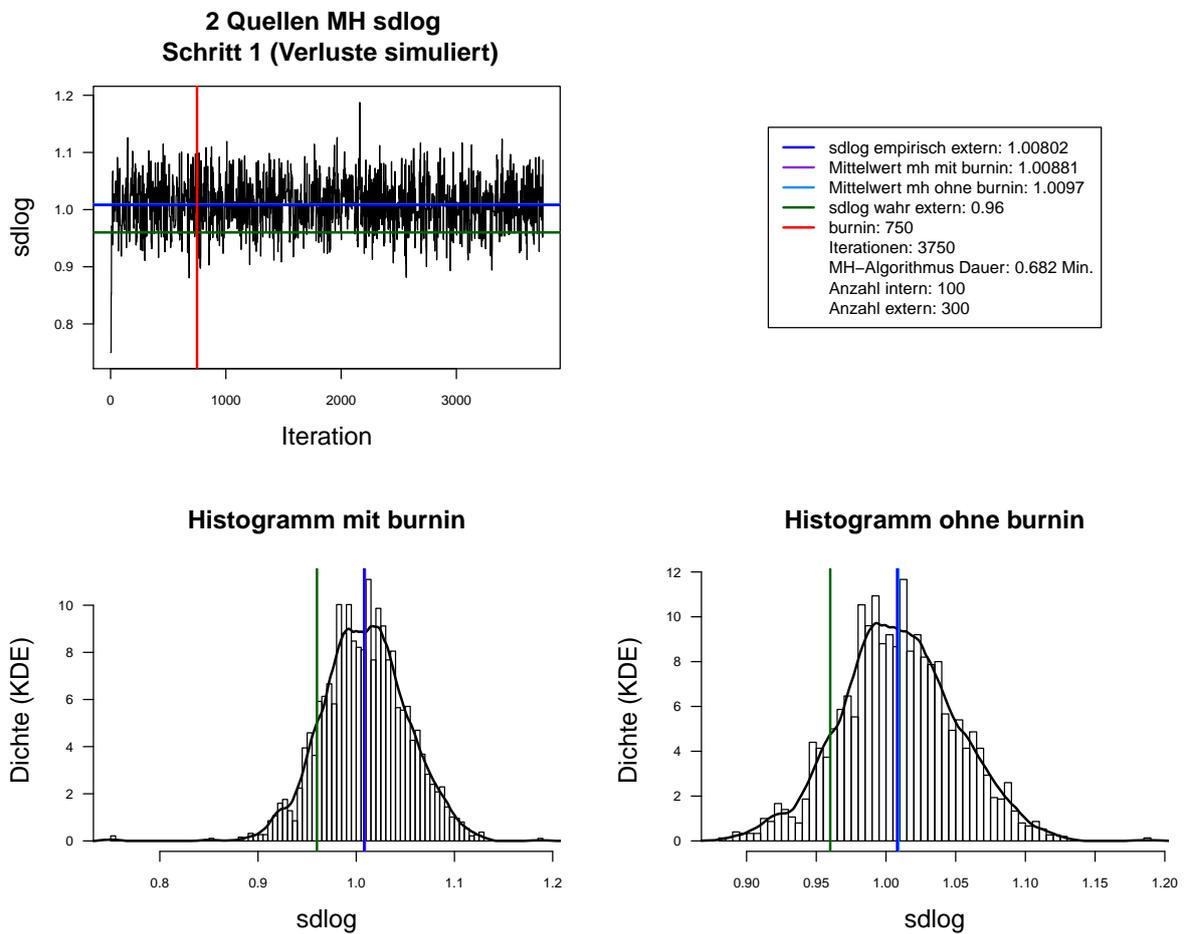
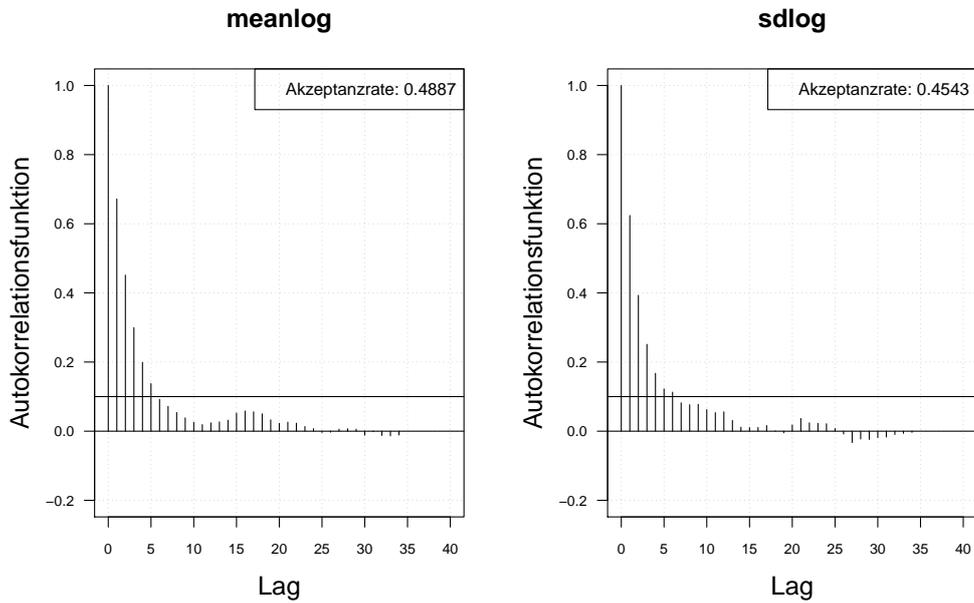


Abbildung A.58.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.682 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 100. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

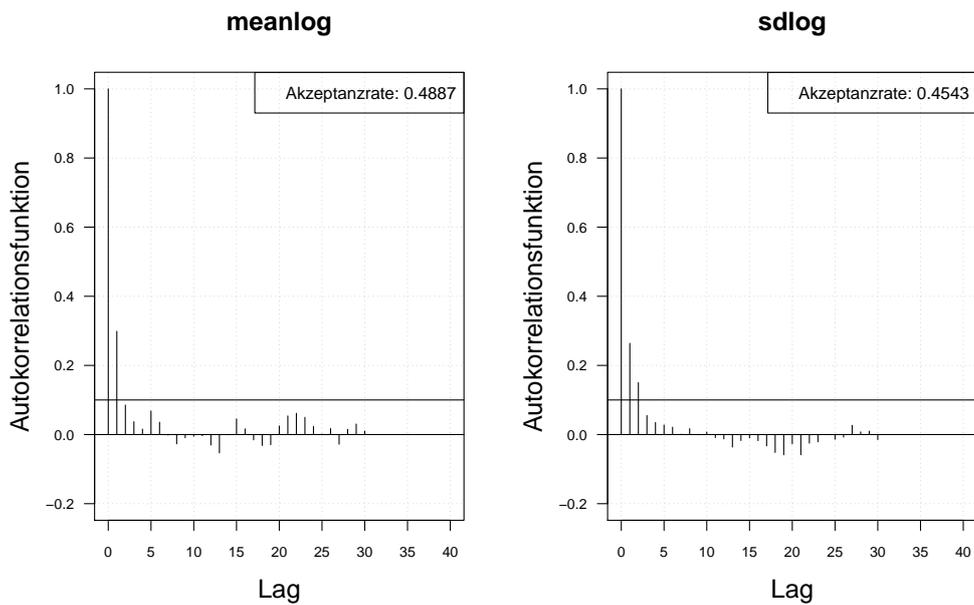


Abbildung A.59.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 1. Interne Daten: 100. Externe Daten: 300.

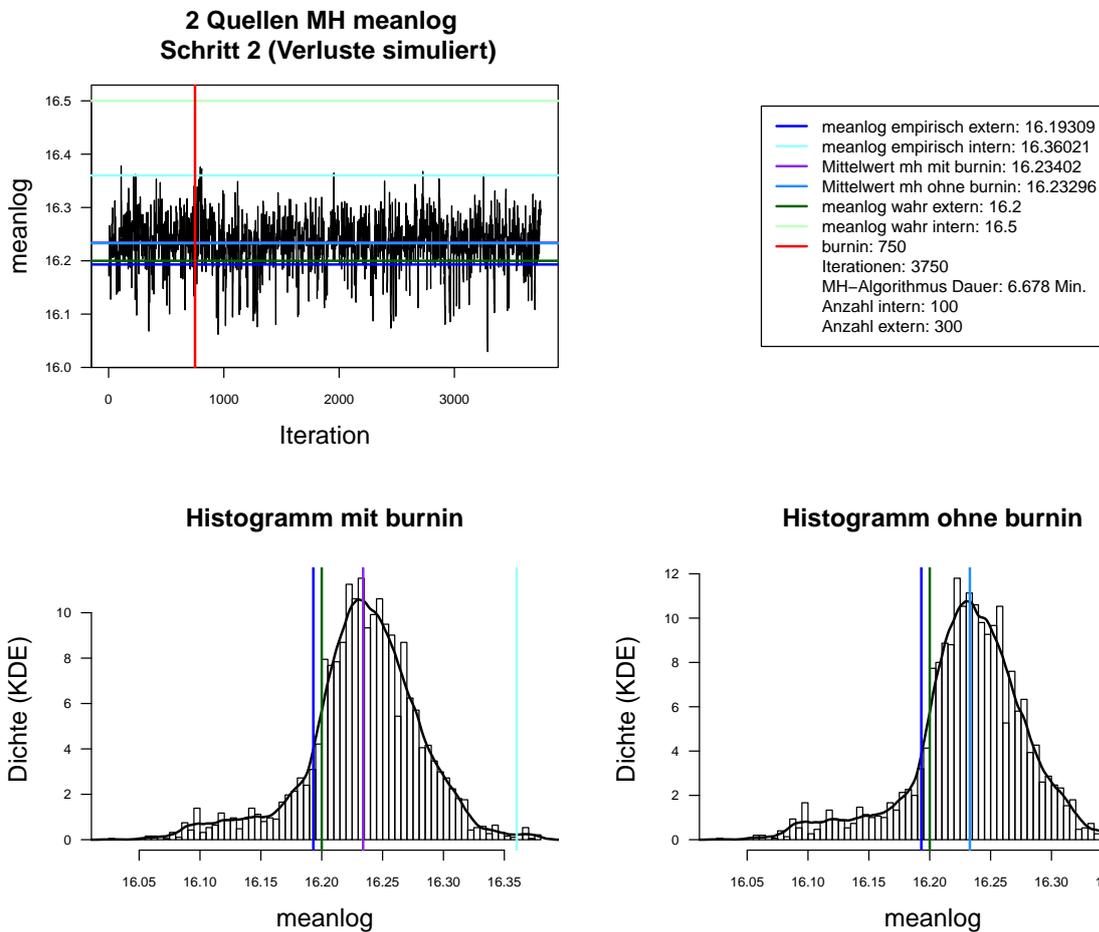


Abbildung A.60.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 6.678 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 300.

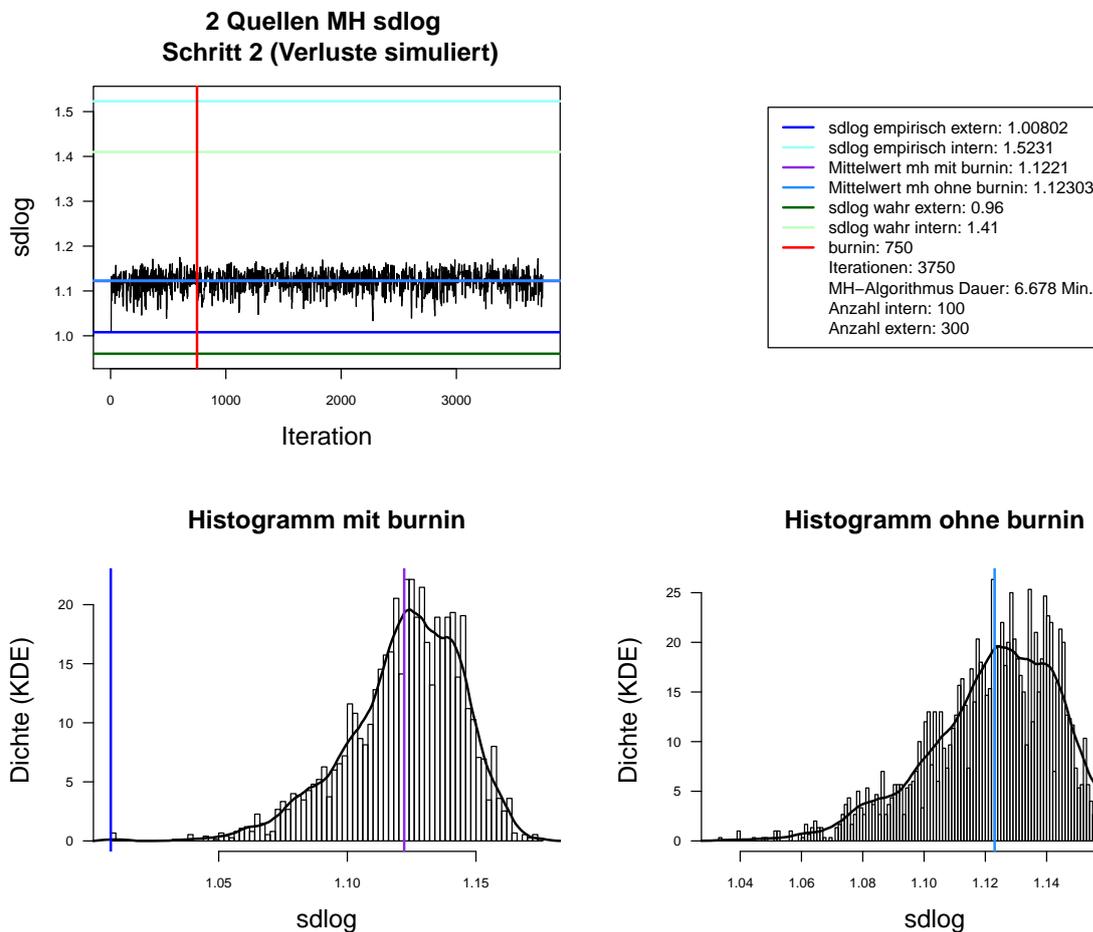
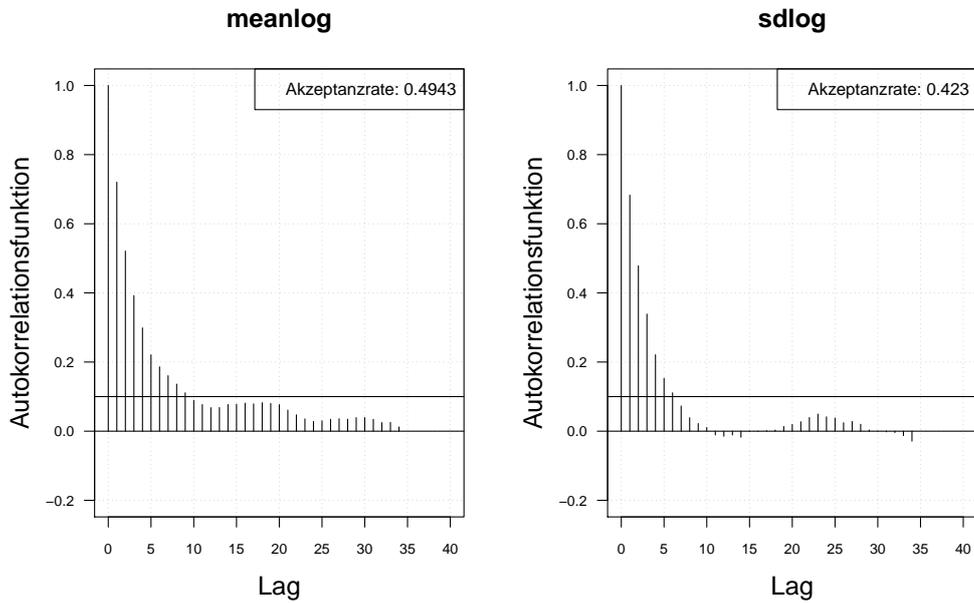


Abbildung A.61.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 6.678 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 2. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

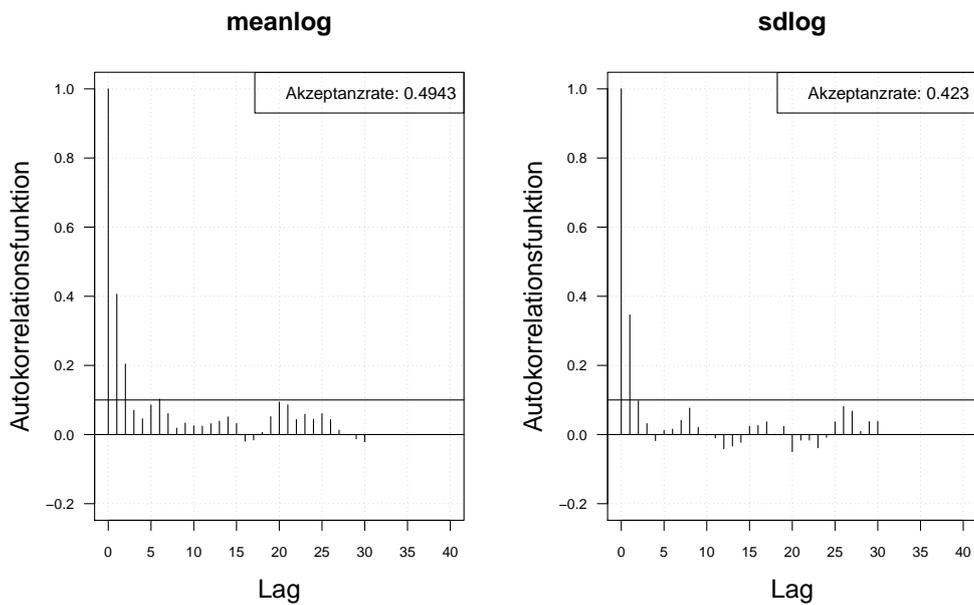


Abbildung A.62.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 2. Interne Daten: 100. Externe Daten: 300.

## 2 Quellen Log-log-Plot simulierte Verluste (tail)

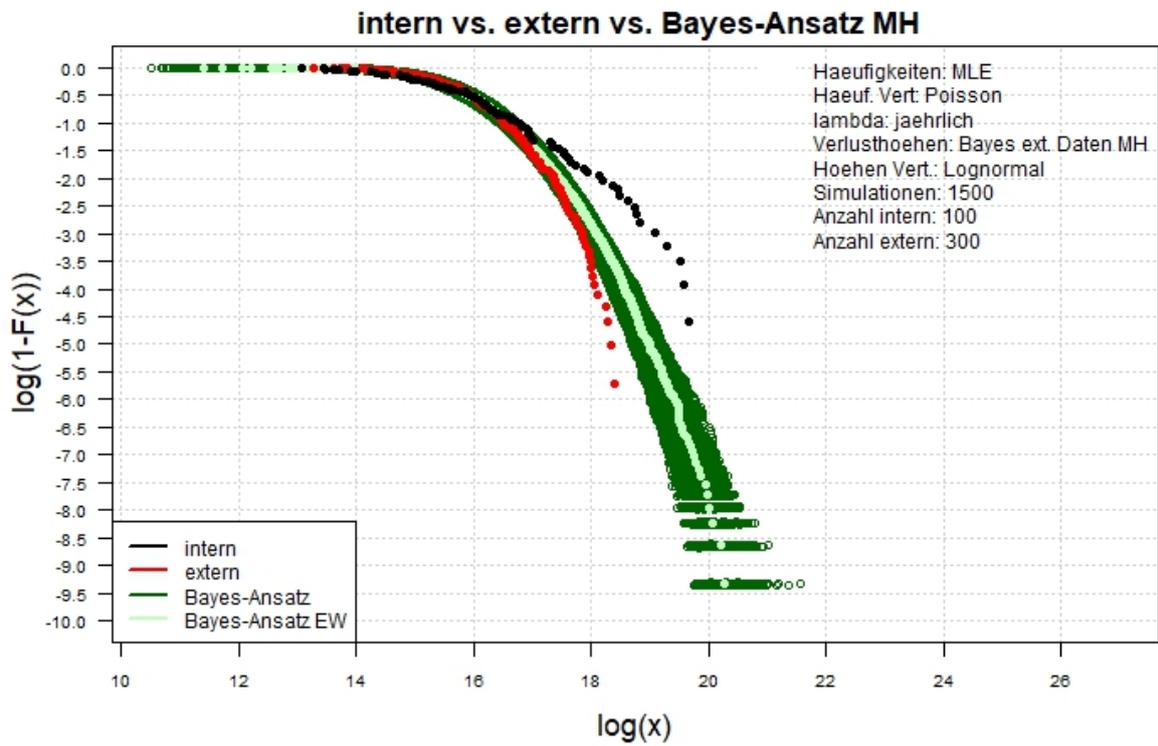


Abbildung A.63.: Der Log-log-Plot. Bayes-Ansatz MH. Zwei-Schritt-Verfahren. Interne Daten: 100. Externe Daten: 300.

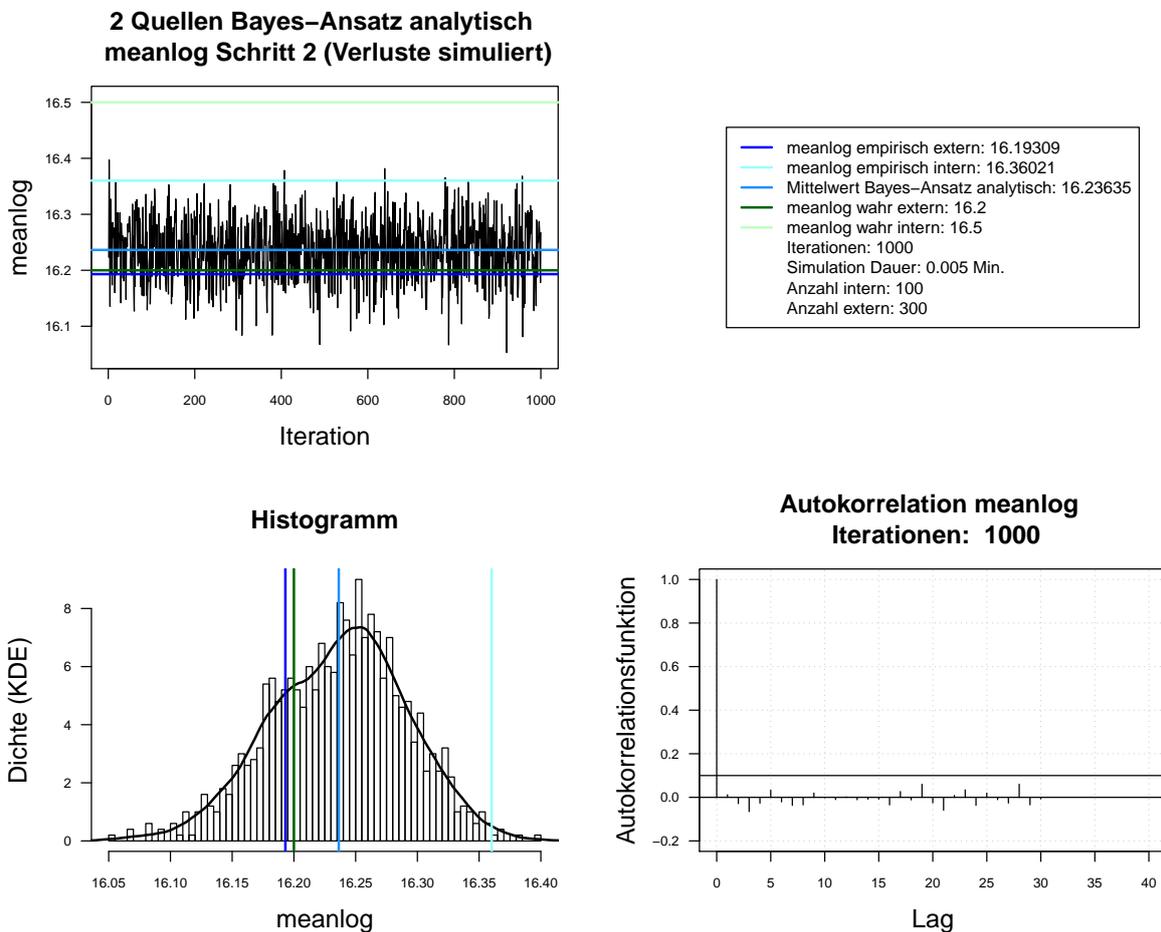


Abbildung A.64.: Der mit analytischen Herleitung simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 300.

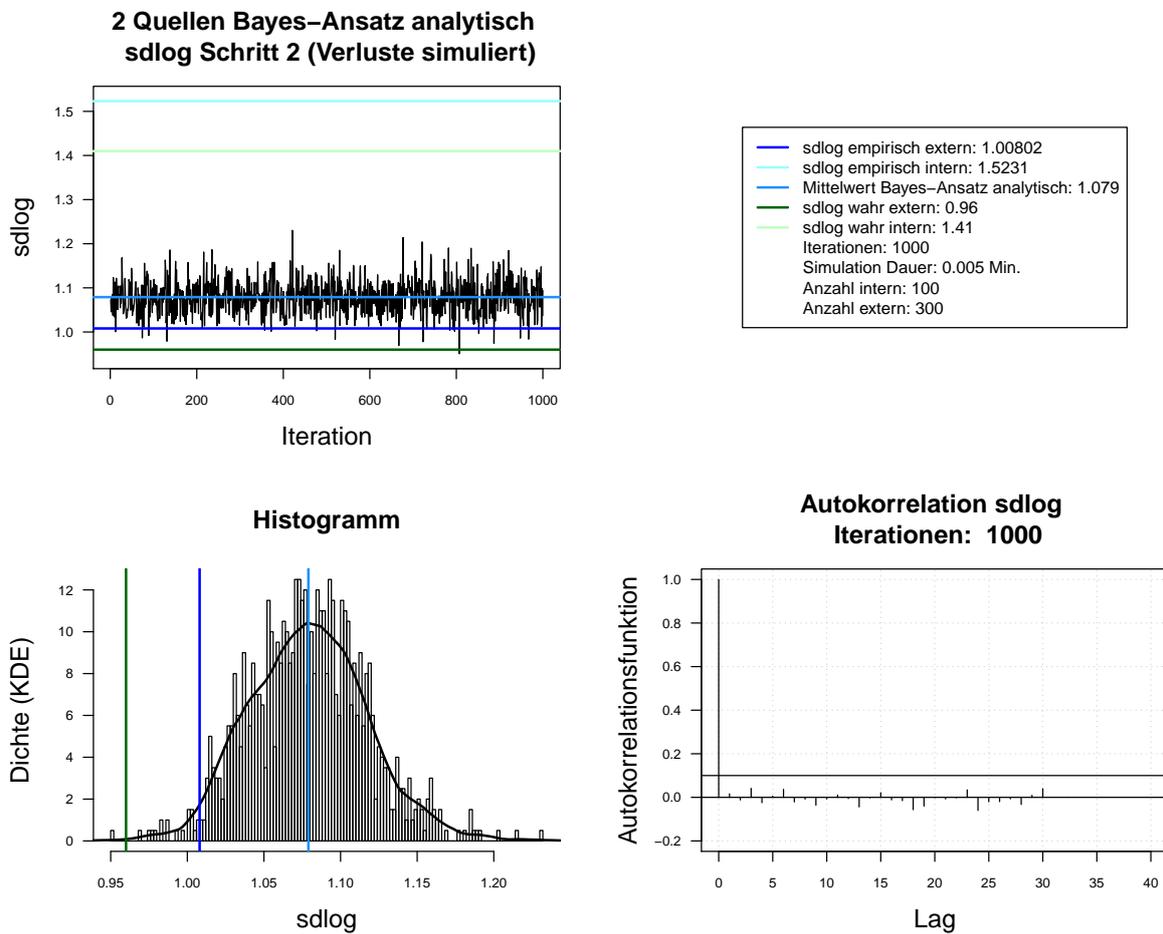


Abbildung A.65.: Der mit analytischen Herleitung simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 100. Externe Daten: 300.

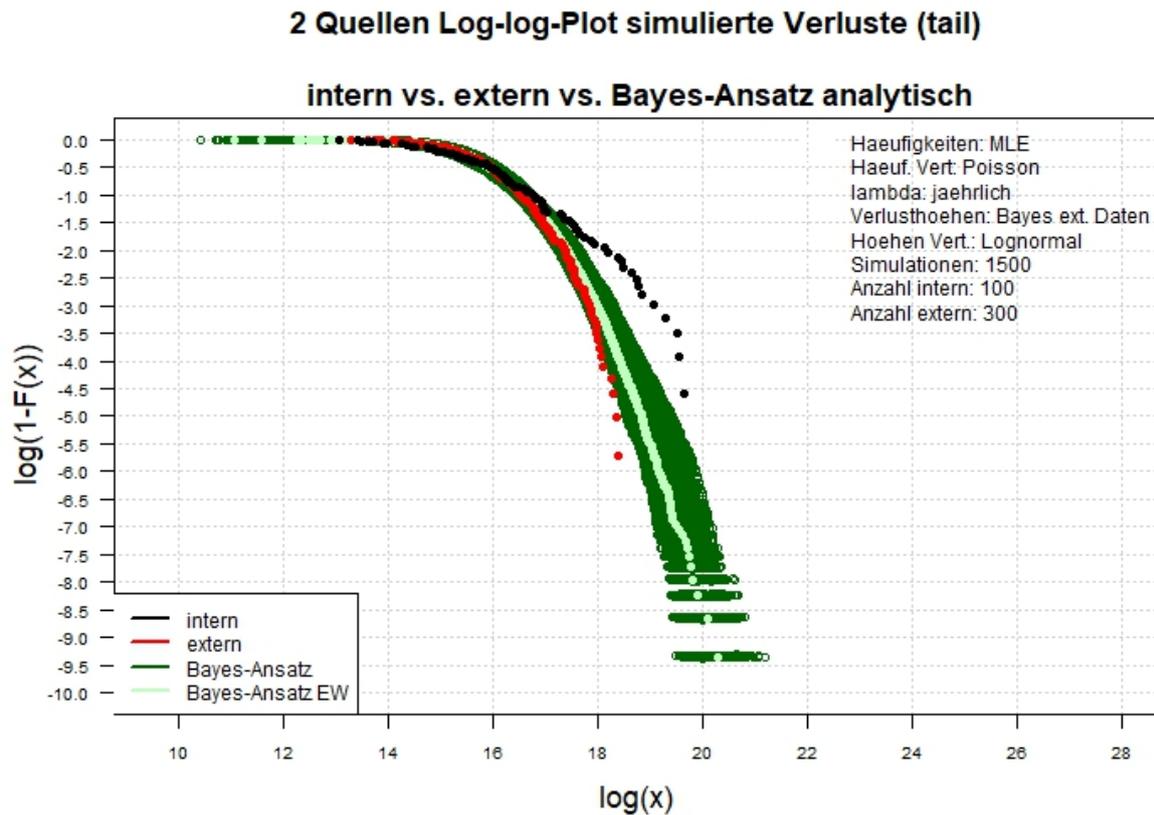


Abbildung A.66.: Der Log-log-Plot. Bayes-Ansatz analytisch. Zwei-Schritt-Verfahren.  
 Interne Daten: 100. Externe Daten: 300.

#### A.10.4. Interne Daten: 200, externe Daten 300, Simulationen: 1500

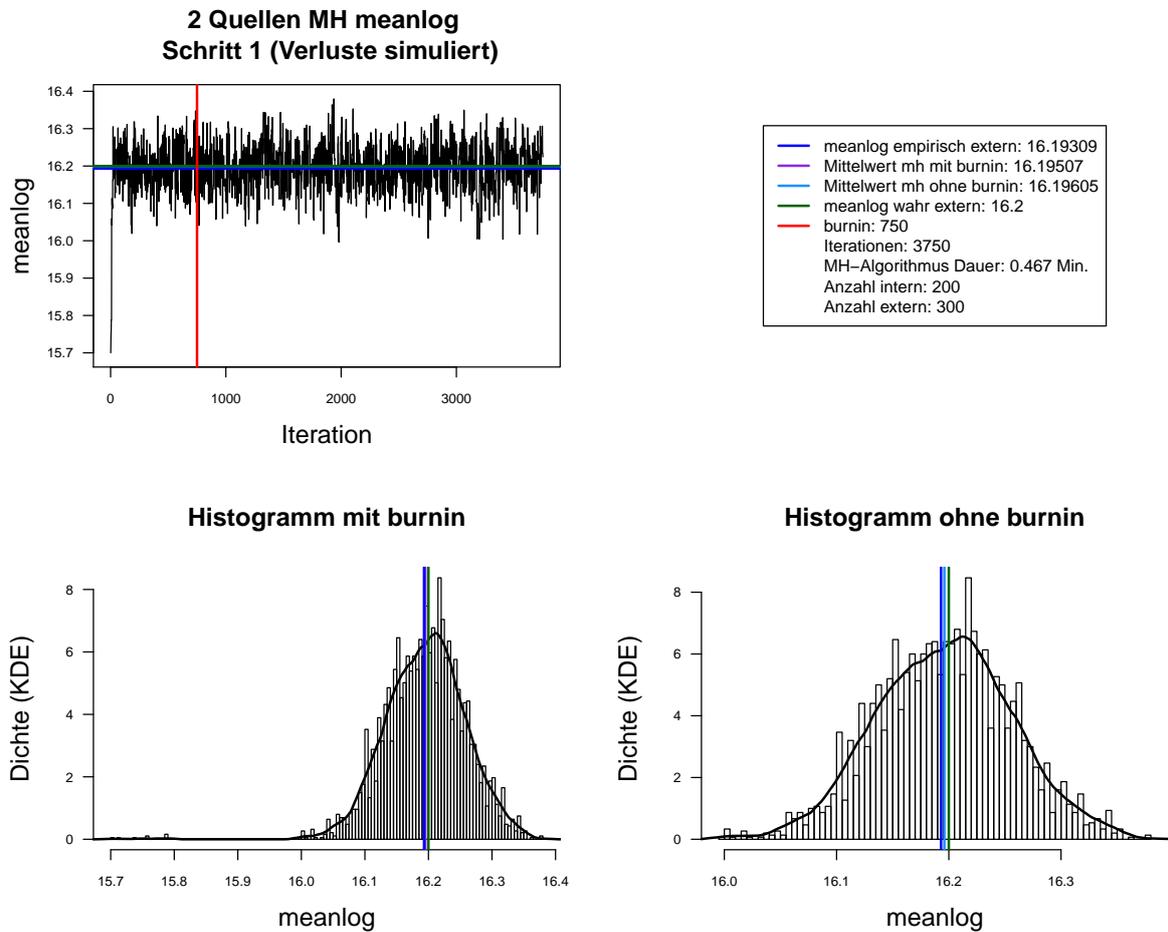


Abbildung A.67.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.467 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 200. Externe Daten: 300.

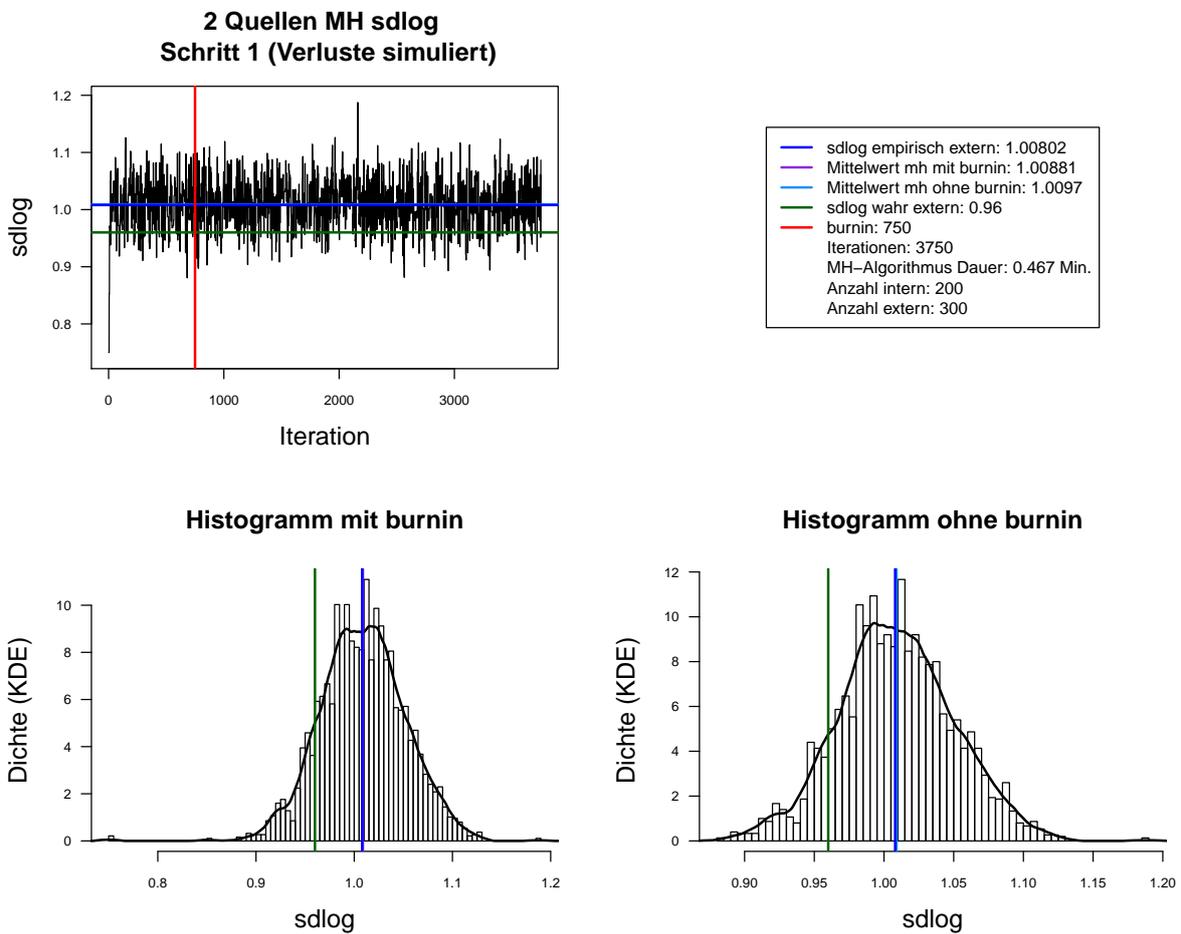
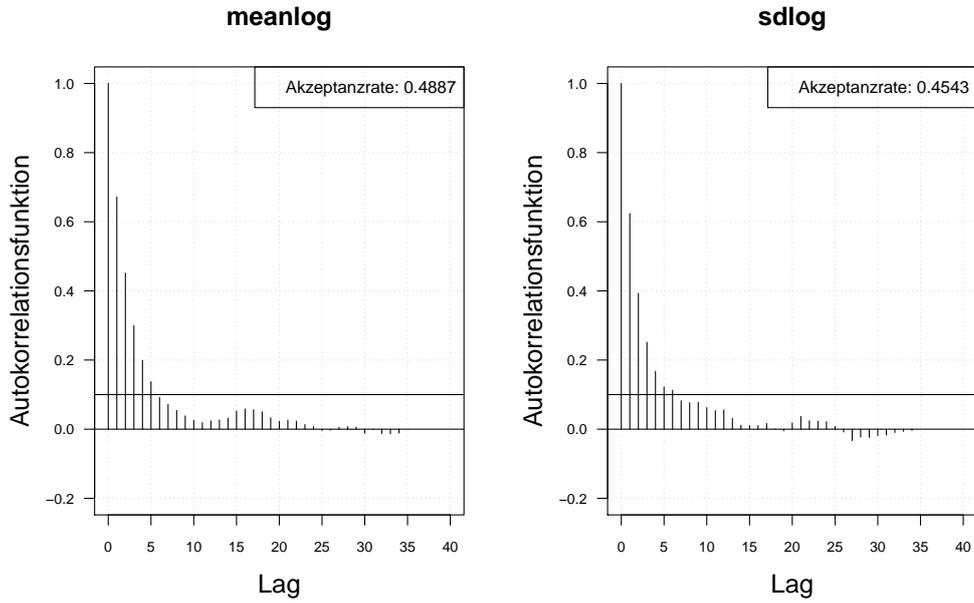


Abbildung A.68.: Der mit MH-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.467 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 200. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

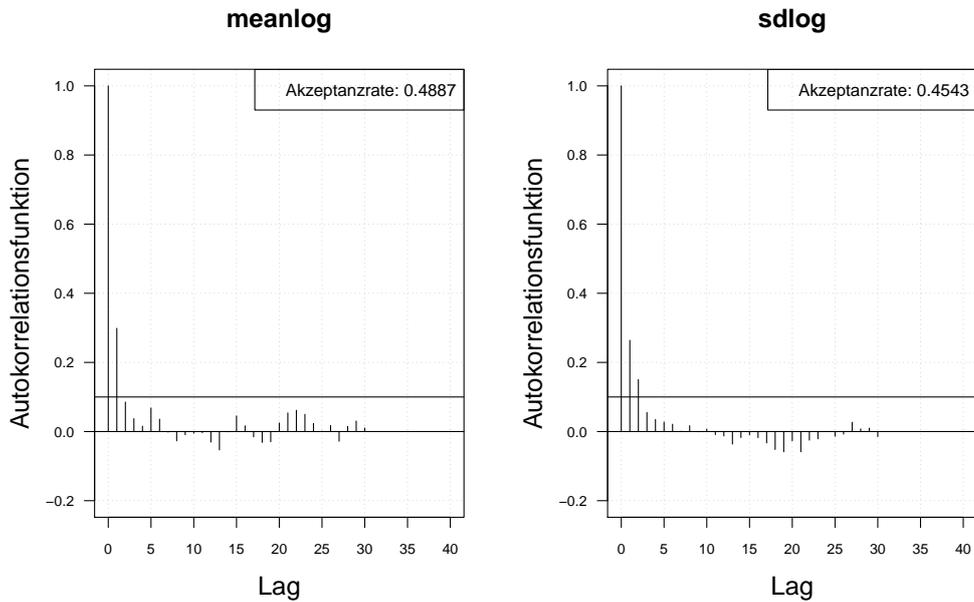


Abbildung A.69.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 1. Interne Daten: 200. Externe Daten: 300.

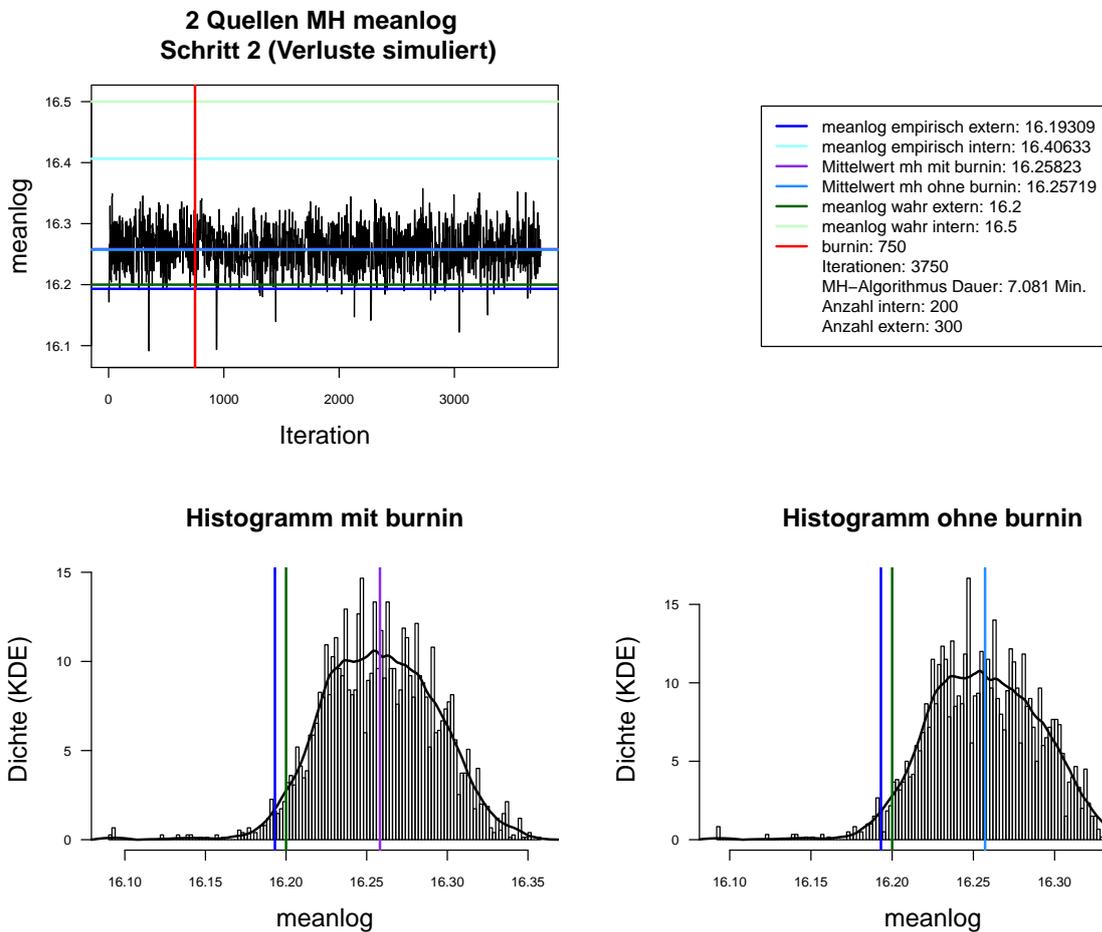


Abbildung A.70.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 7.081 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 200. Externe Daten: 300.

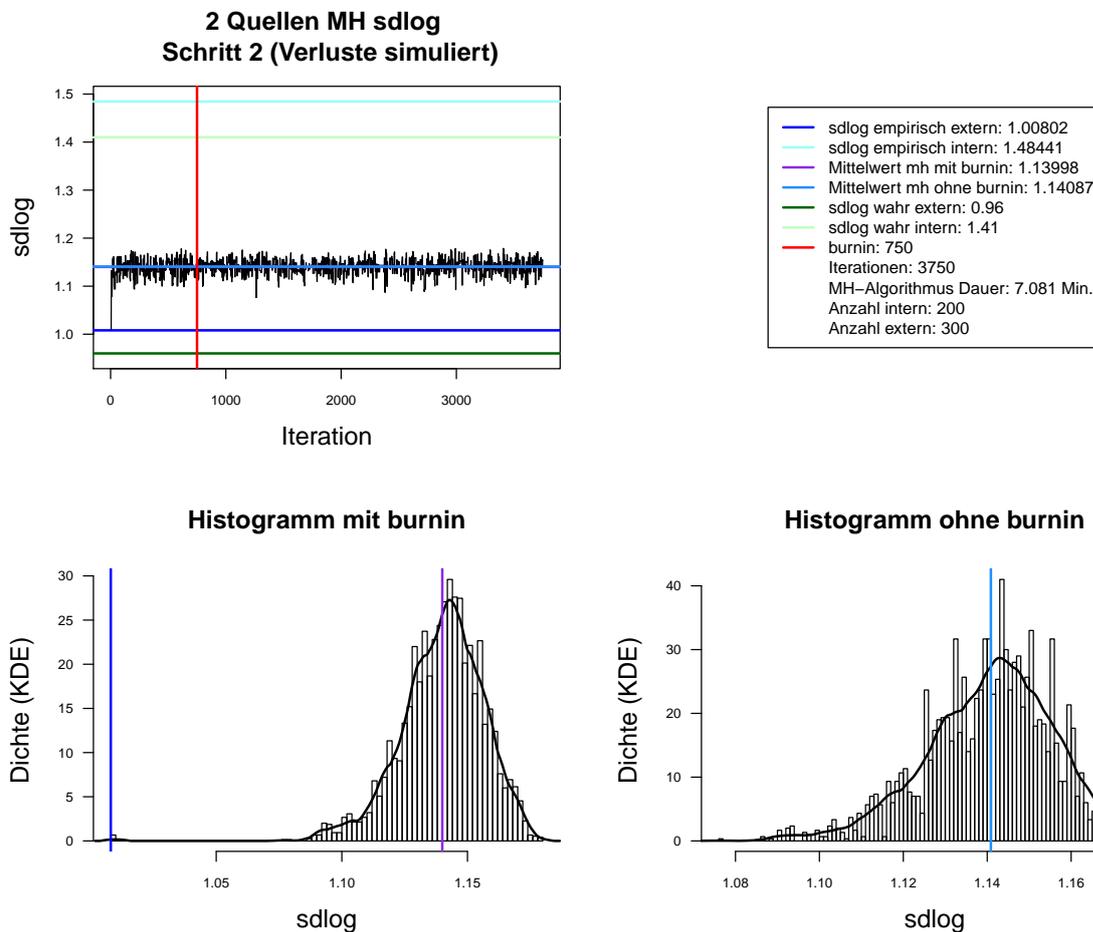
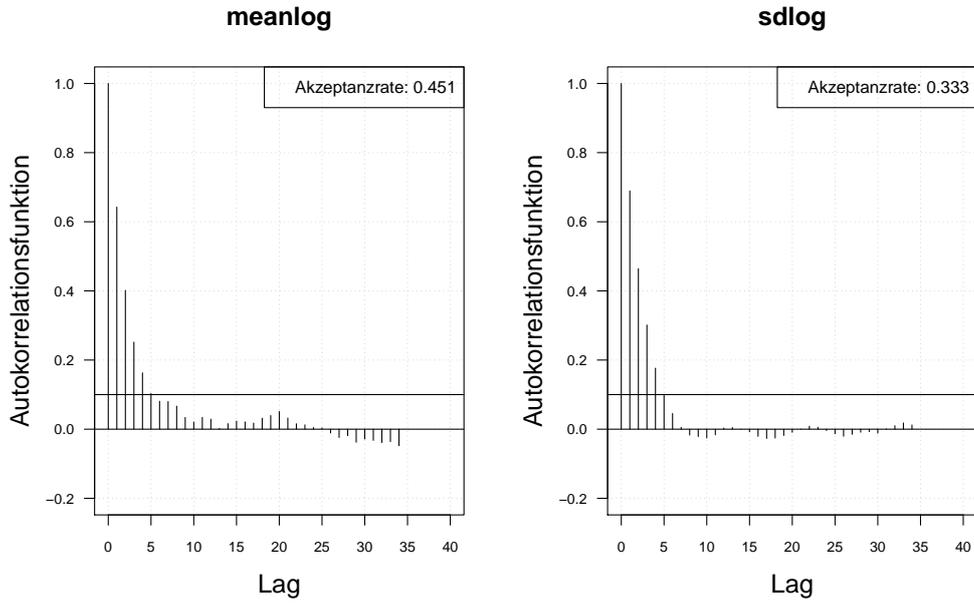


Abbildung A.71.: Der mit MH-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 7.081 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 200. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 2. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

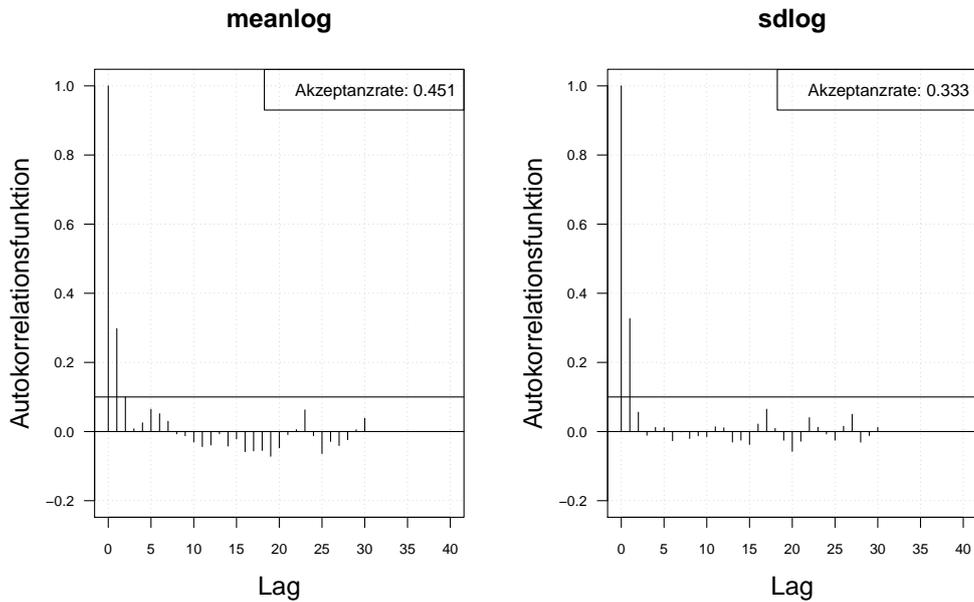


Abbildung A.72.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 2. Interne Daten: 200. Externe Daten: 300.

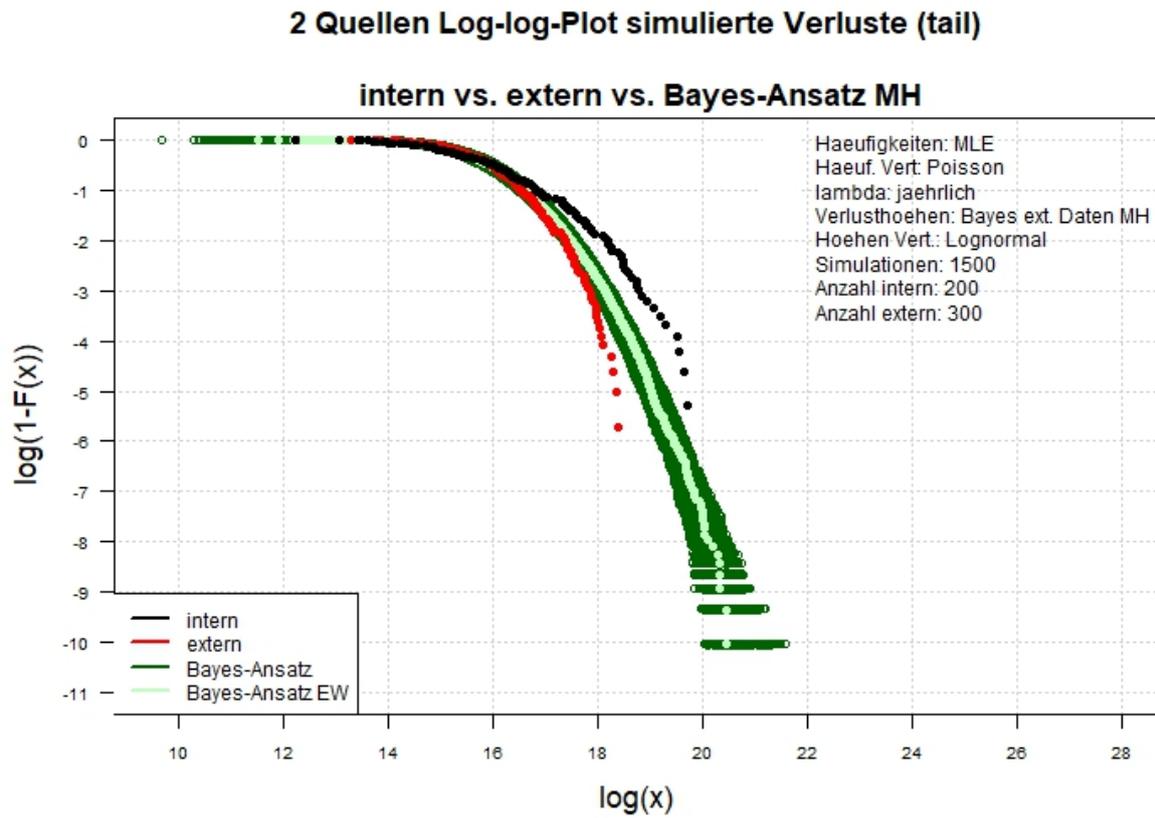


Abbildung A.73.: Der Log-log-Plot. Bayes-Ansatz MH. Zwei-Schritt-Verfahren. Interne Daten: 200. Externe Daten: 300.

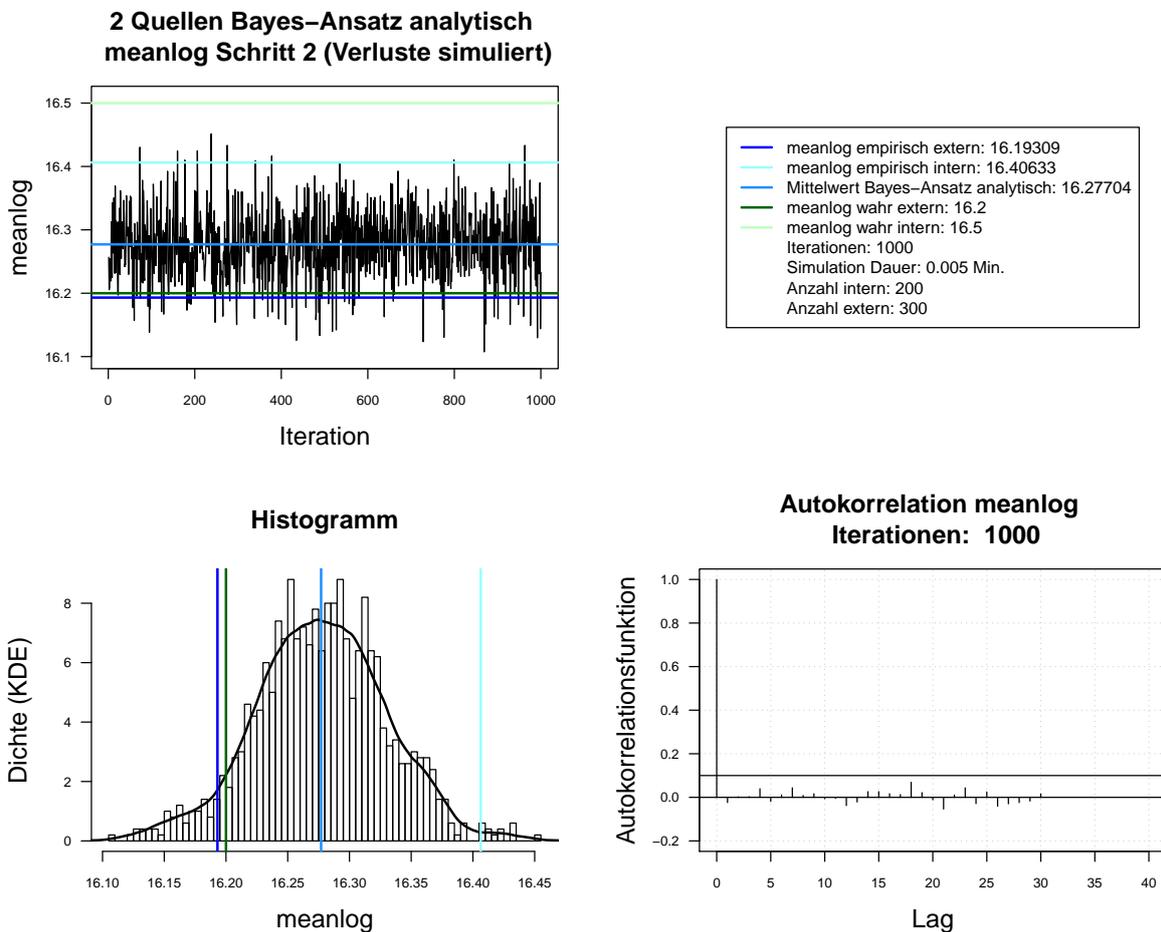


Abbildung A.74.: Der mit analytischen Herleitung simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 200. Externe Daten: 300.

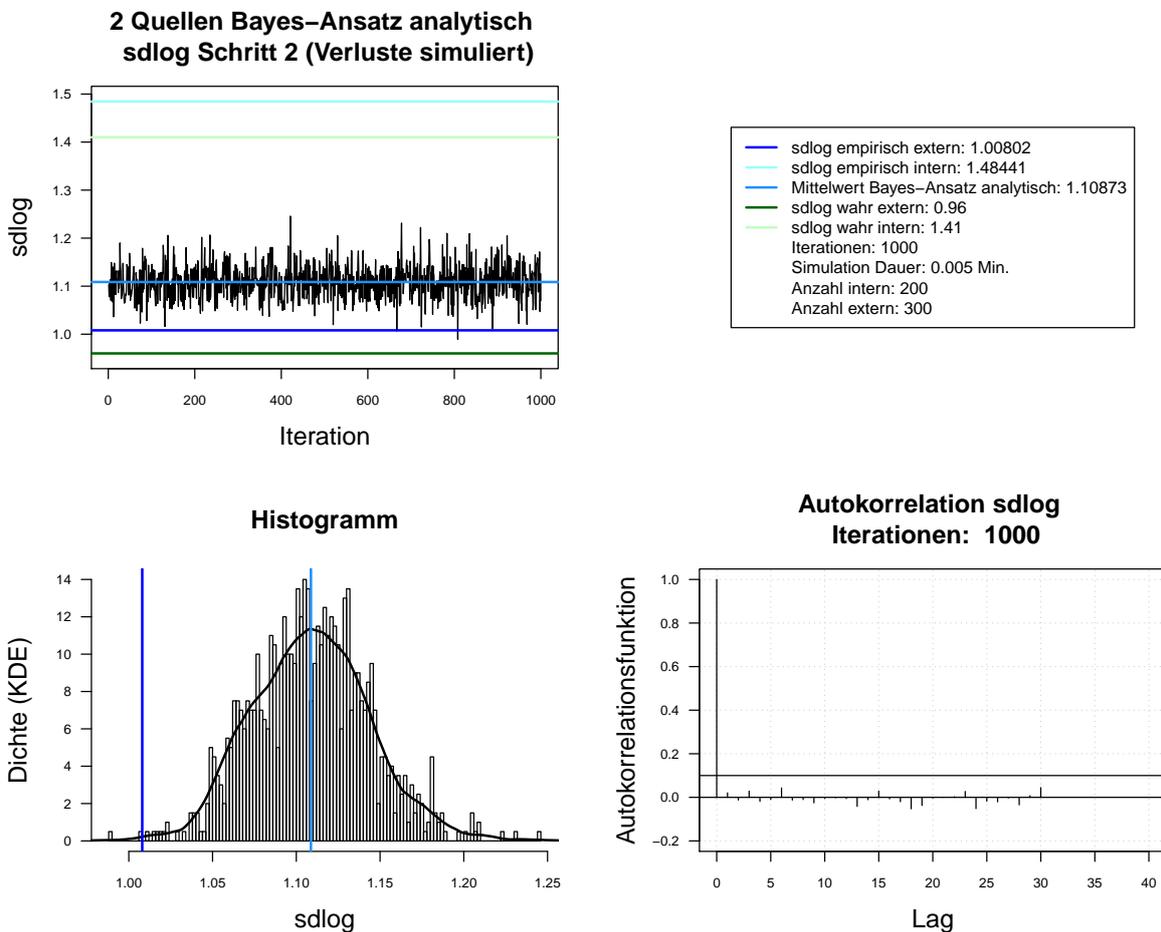


Abbildung A.75.: Der mit analytischen Herleitung simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.005 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 200. Externe Daten: 300.

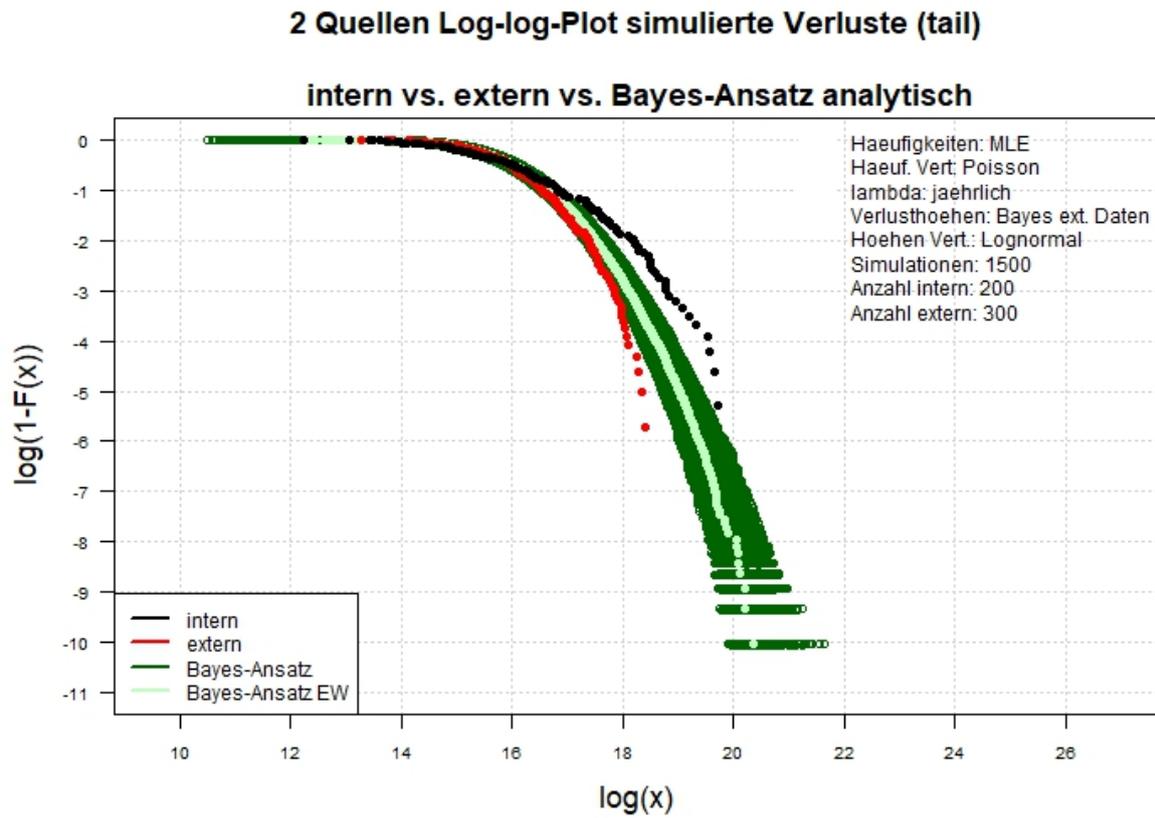


Abbildung A.76.: Der Log-log-Plot. Bayes-Ansatz analytisch. Zwei-Schritt-Verfahren. Interne Daten: 200. Externe Daten: 300.

### A.10.5. Interne Daten: 300, externe Daten 300, Simulationen: 1500

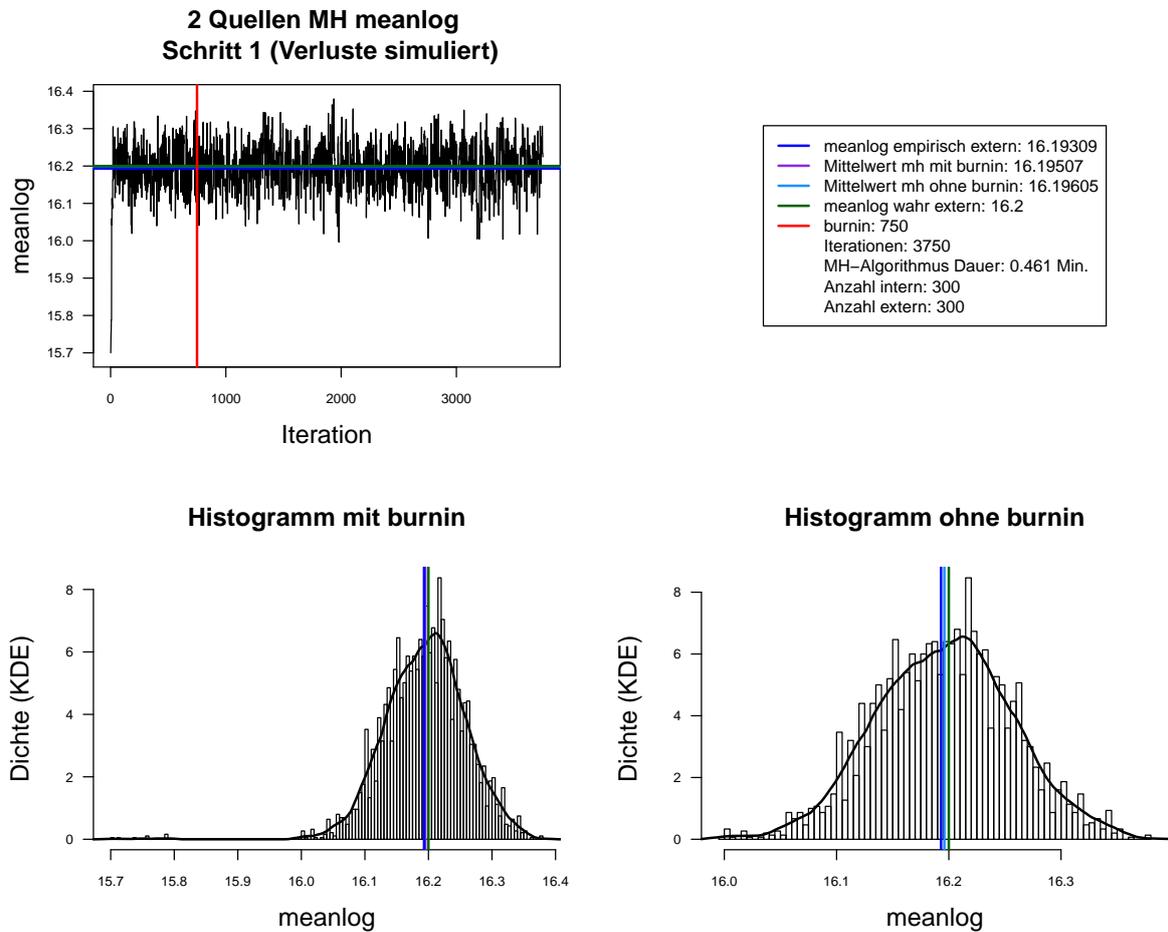


Abbildung A.77.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.461 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 300. Externe Daten: 300.

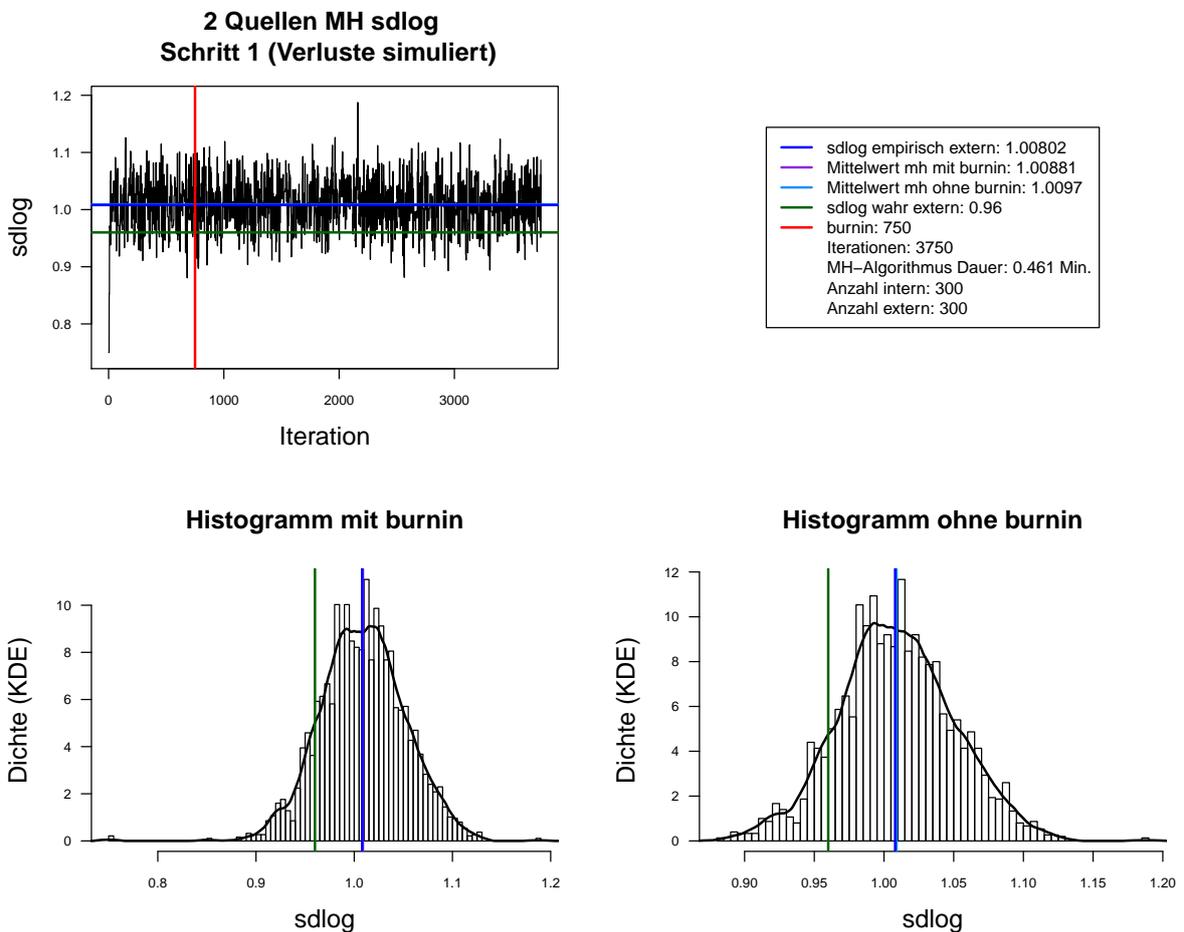
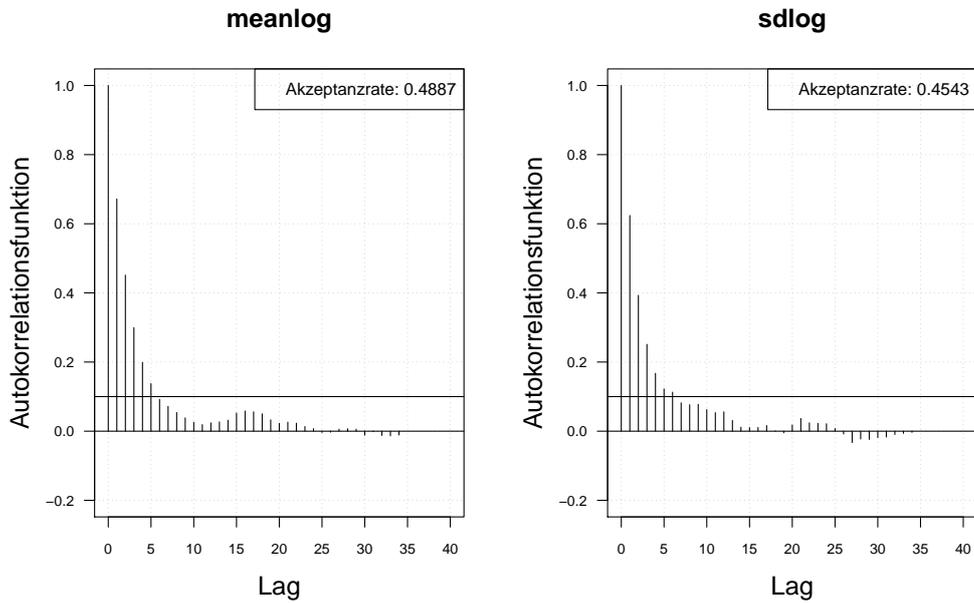


Abbildung A.78.: Der mit MH-Algorithmus simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 0.461 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 1. Interne Daten: 300. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

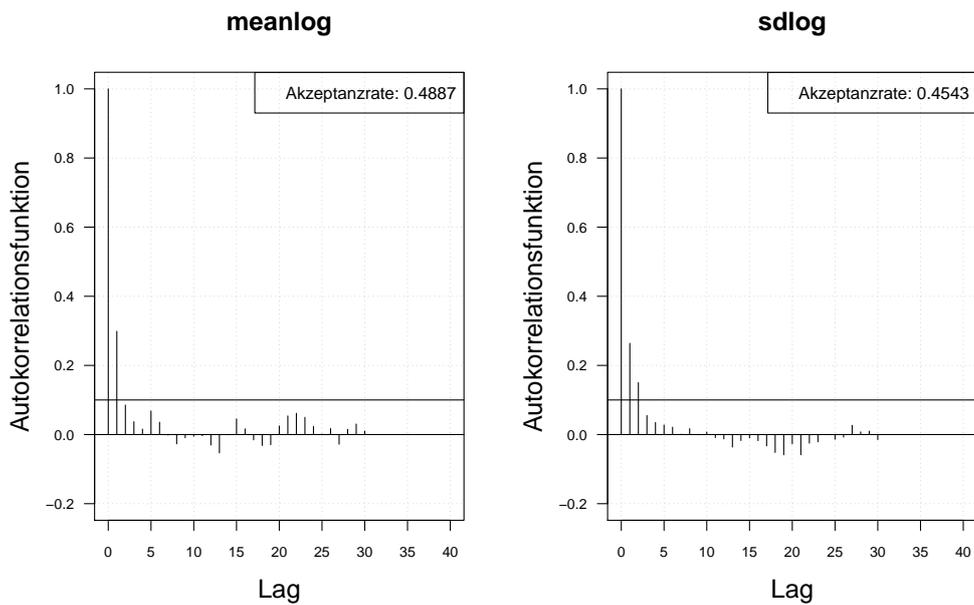


Abbildung A.79.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 1. Interne Daten: 300. Externe Daten: 300.

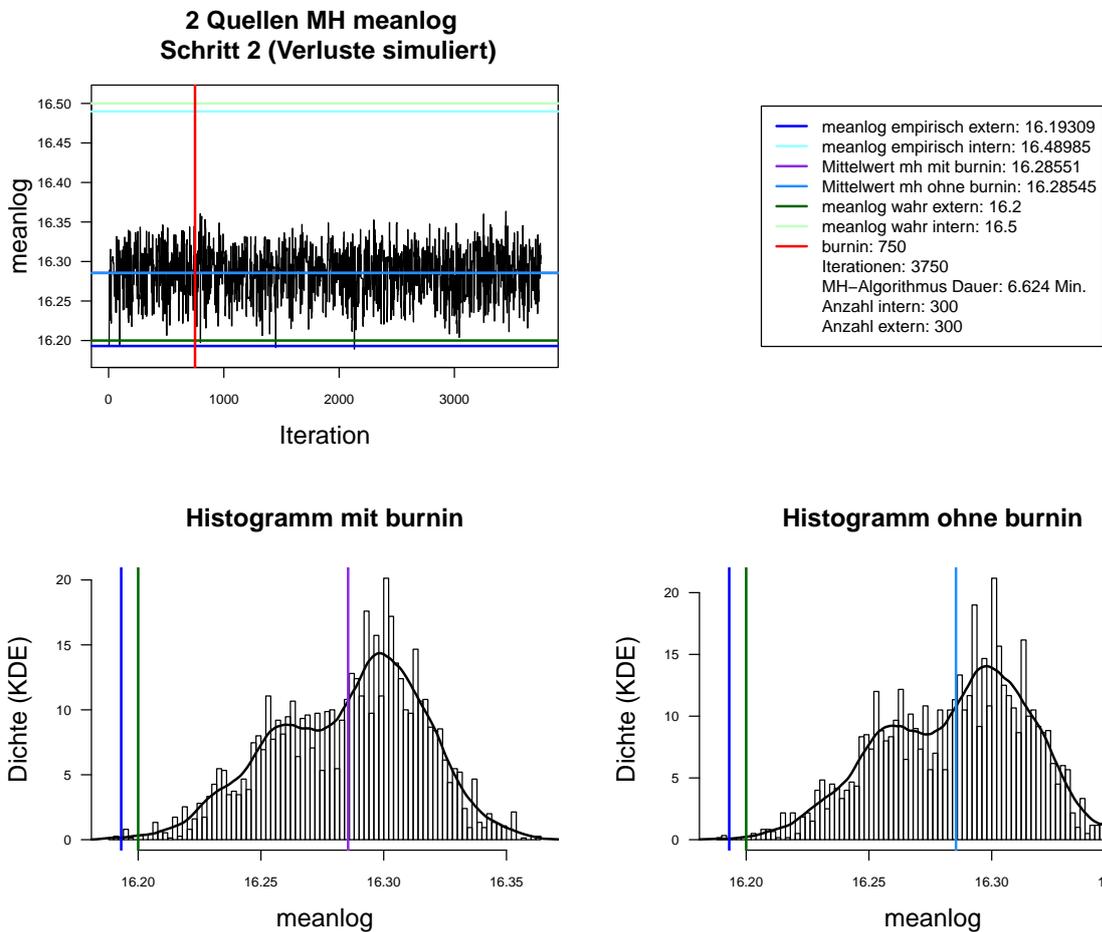


Abbildung A.80.: Der mit MH-Algorithmus simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 6.624 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 300.

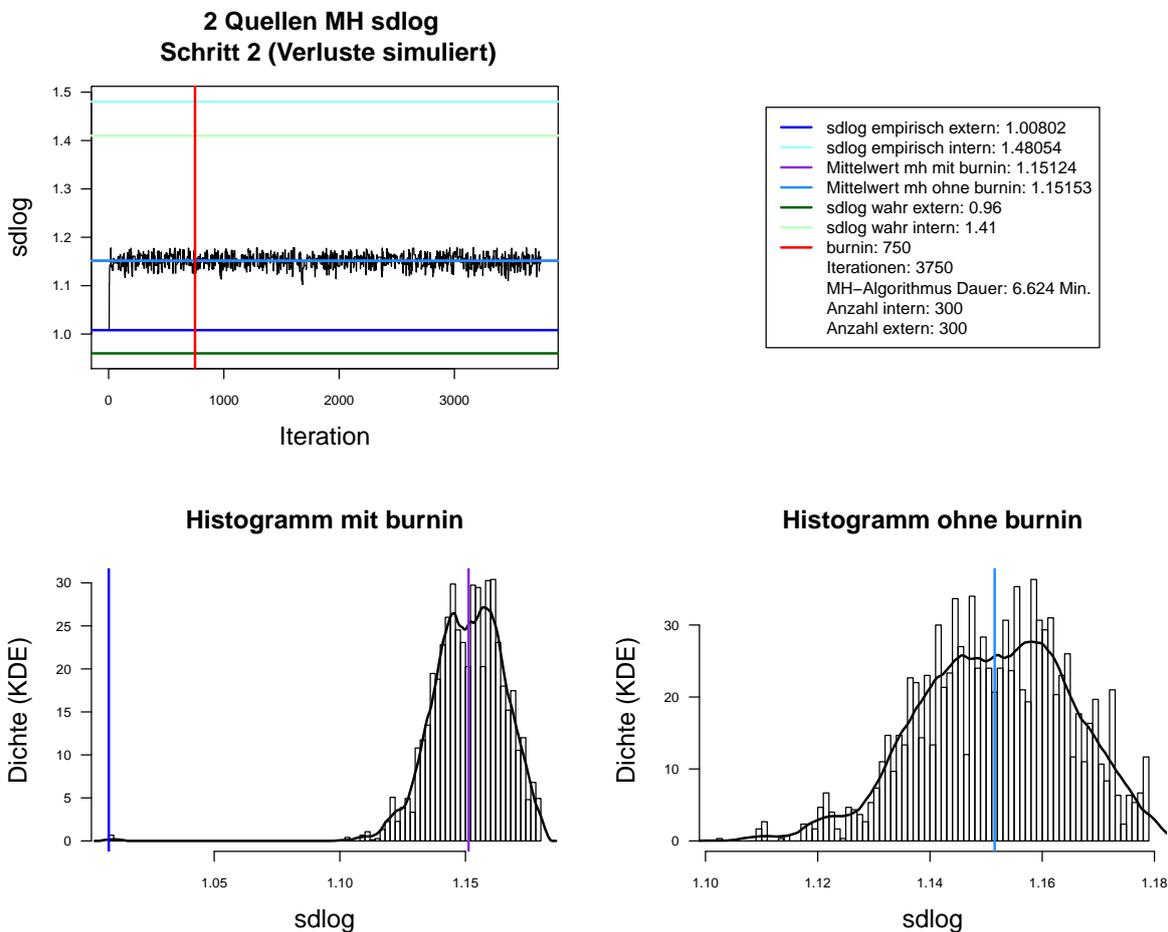
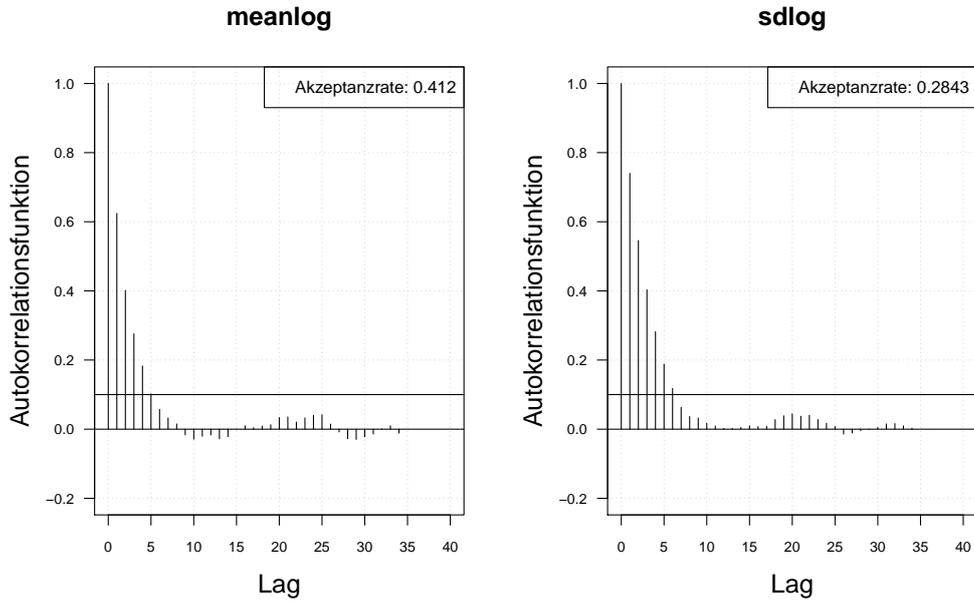


Abbildung A.81.: Der mit MH-Algorithmus simulierte  $sdlog$  der Lognormal-Verteilung. Dauer des Algorithmus mit 3750 Iterationen beträgt 6.624 Minuten. Markov-Kette (oben links), Histogramm für 3750 Iterationen mit KDE (unten links) und Histogramm für letzte 3000 Iterationen mit KDE (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 300.

Autokorrelation 2 Quellen Schritt 2. Ohne burnin (Verluste simuliert),  
Iterationen: 3000



Autokorrelation 2 Quellen Schritt 1. Ohne burnin (Verluste simuliert),  
thin = 3, Iterationen: 3000

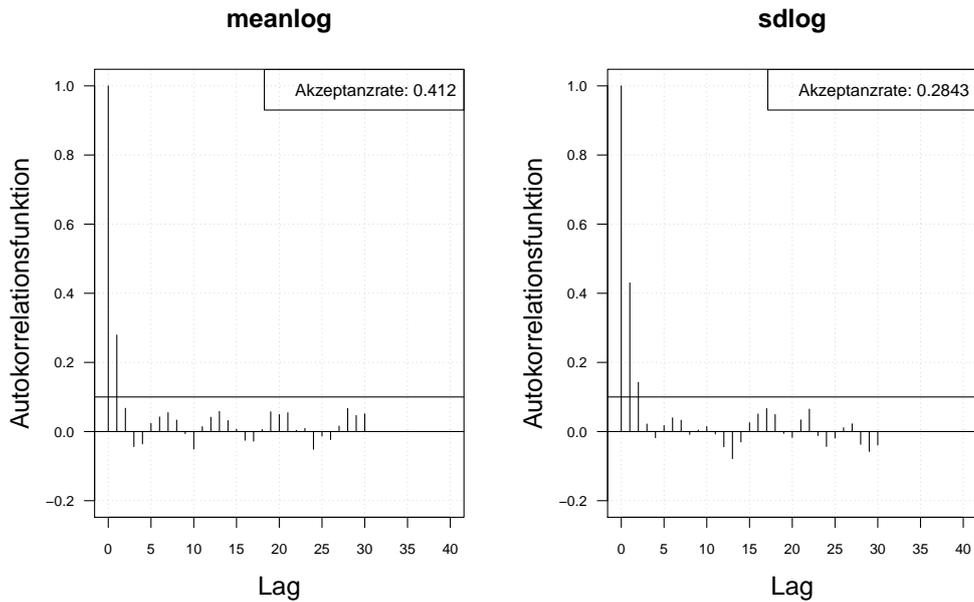


Abbildung A.82.: Die Autokorrelation der Markov-Kette (obere Zeile) und der verdünnten Markov-Kette (untere Zeile) mit MH-Algorithmus simulierten *meanlog* und *sdlog* der Lognormal-Verteilung. Iterationen: 3000. Schritt 2. Interne Daten: 300. Externe Daten: 300.

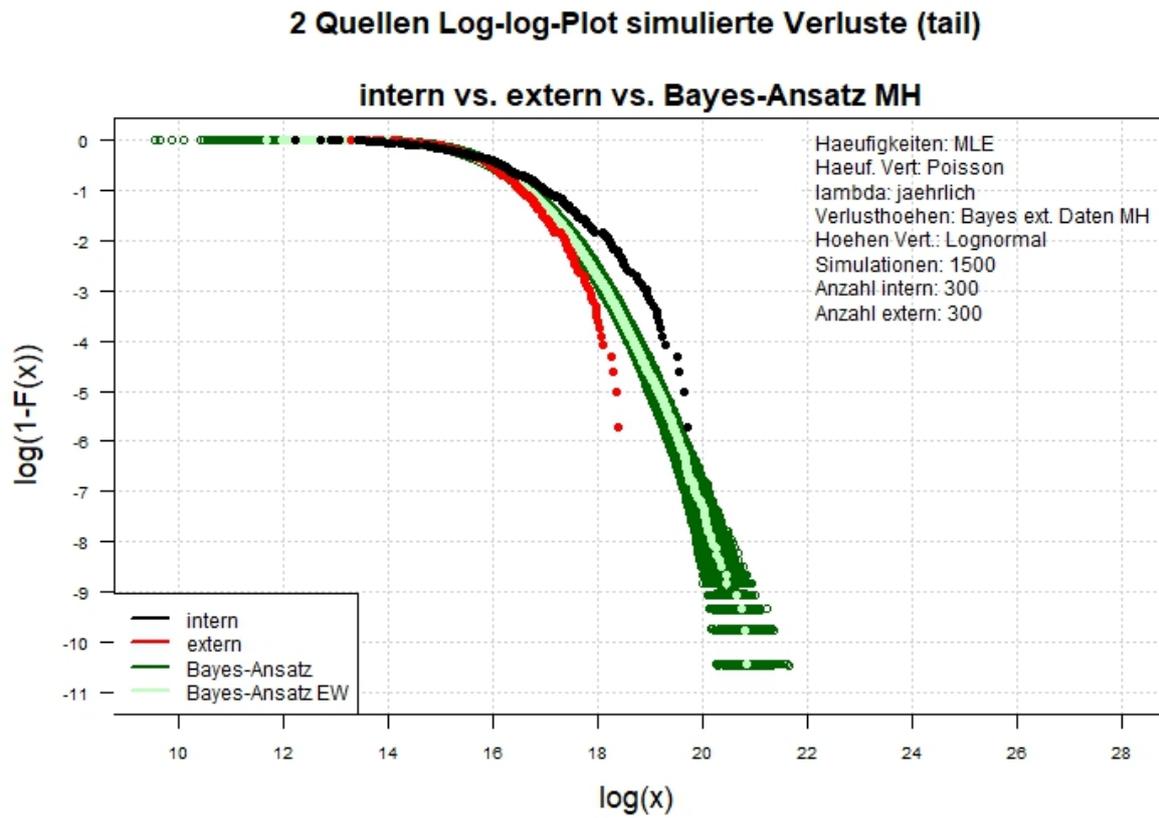


Abbildung A.83.: Der Log-log-Plot. Bayes-Ansatz MH. Zwei-Schritt-Verfahren. Interne Daten: 300. Externe Daten: 300.

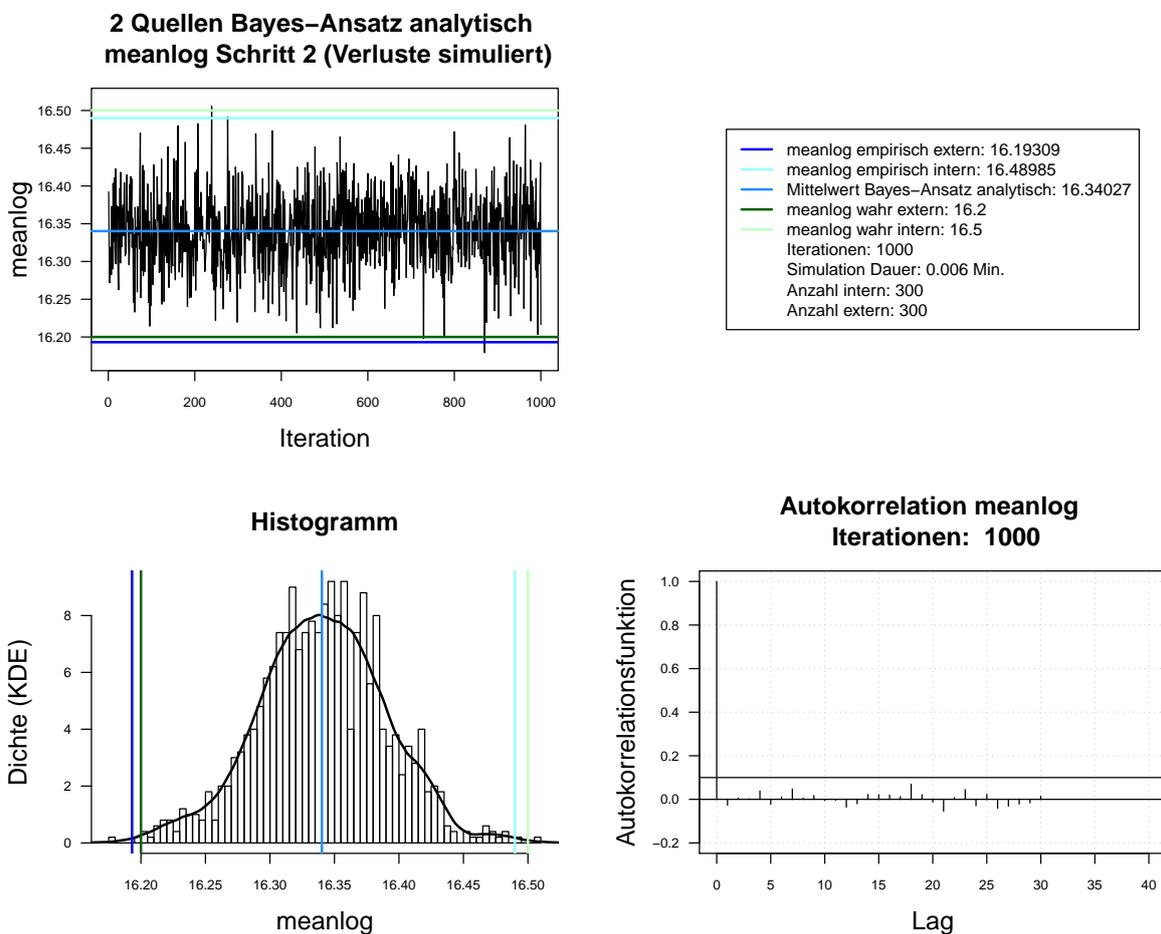


Abbildung A.84.: Der mit analytischen Herleitung simulierte *meanlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.006 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 300.

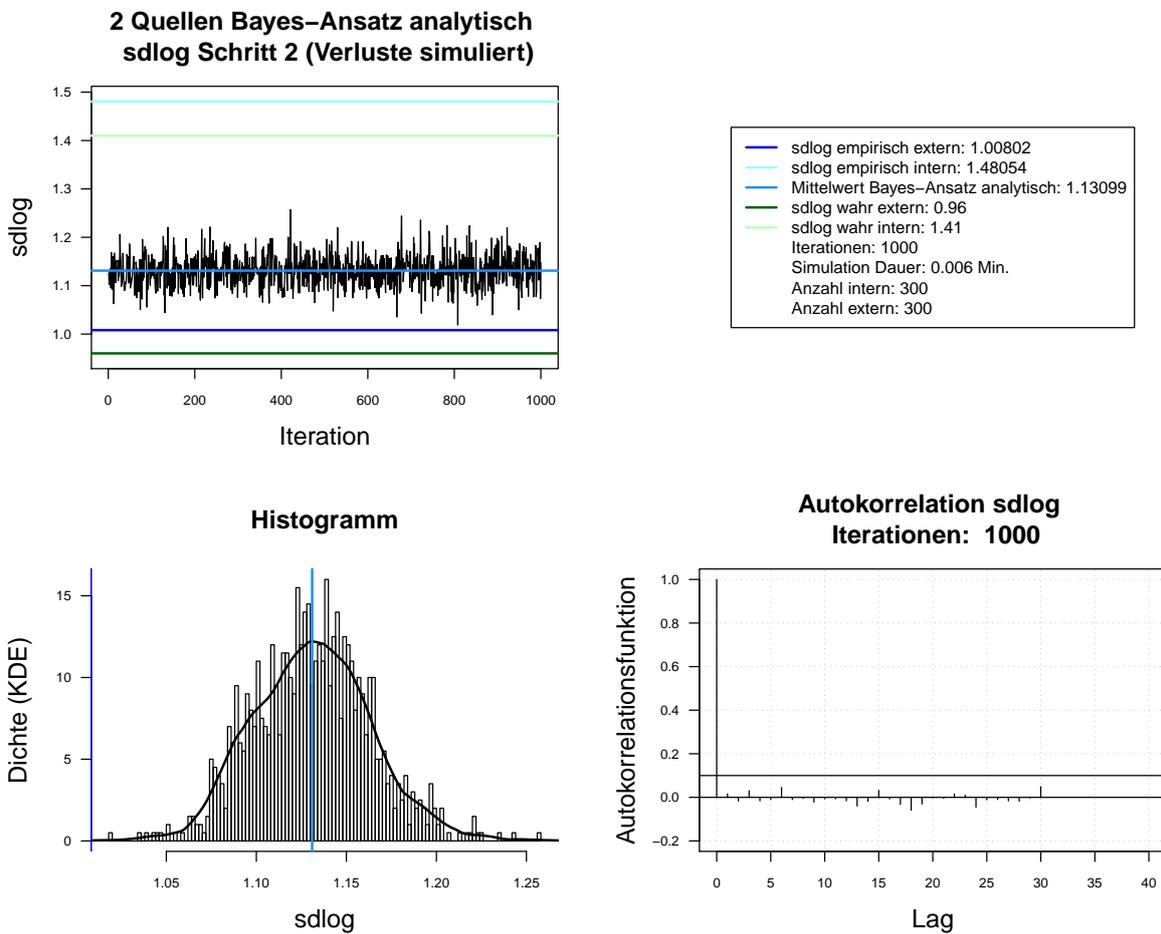


Abbildung A.85.: Der mit analytischen Herleitung simulierte *sdlog* der Lognormal-Verteilung. Dauer des Algorithmus mit 1000 Iterationen beträgt 0.006 Minuten. Markov-Kette (oben links), Histogramm für 1000 Iterationen mit KDE (unten links), ACF für 1000 Iterationen (unten rechts). Schritt 2. Interne Daten: 300. Externe Daten: 300.

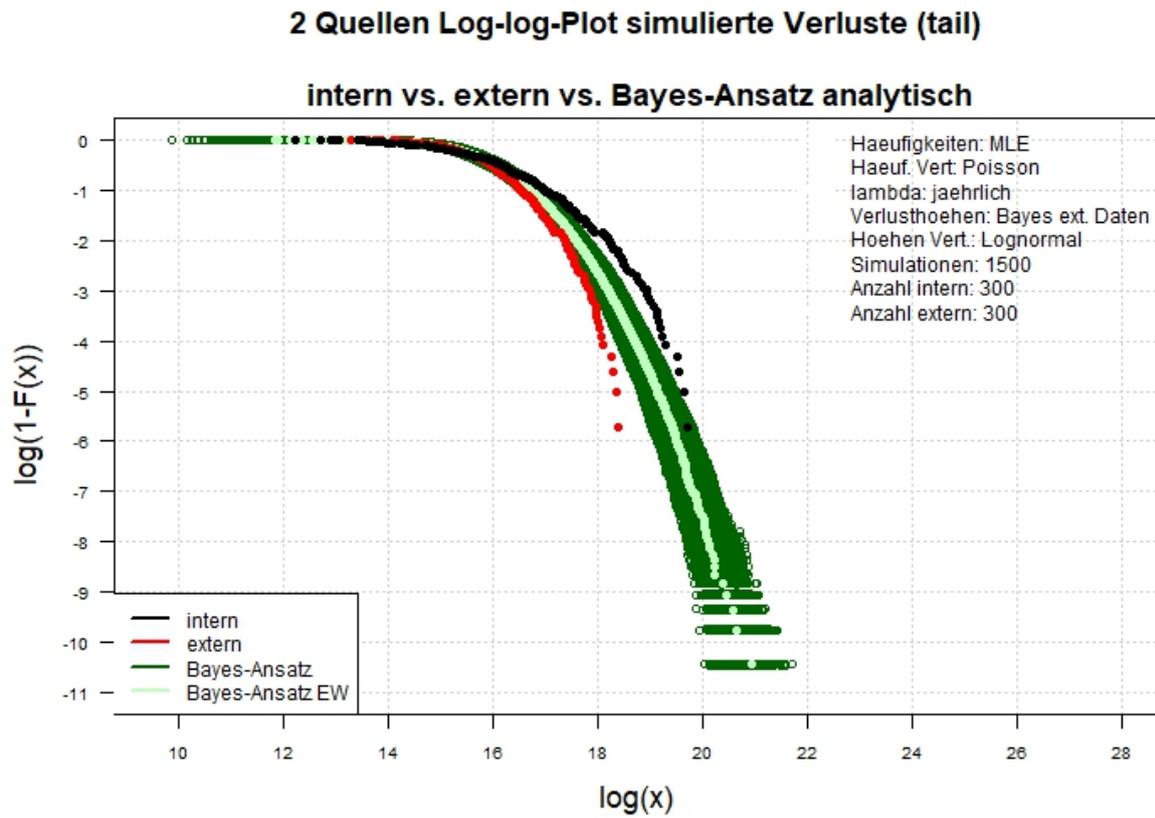


Abbildung A.86.: Der Log-log-Plot. Bayes-Ansatz analytisch. Zwei-Schritt-Verfahren. Interne Daten: 300. Externe Daten: 300.

## B. Datendokumentation

In elektronischer Form sind dieser Arbeit beigelegt:

- Literatur
- R-Code (mit der dazugehörigen Dokumentation)
- R-Data
- R-Output
- Master-Thesis.pdf

Im Ordner “Literatur” befinden sich alle in dieser Arbeit verwendete Literaturquellen. Im Ordner “R-Code” befindet sich der auskommentierte R-Code mit Umsetzung des Zwei-Schritt-Bayes-Ansatzes (Hauptdatei “*bayes\_estimate.R*”) sowie die Analysen der MH- und MH-SA-Algorithmen (Hauptdatei “*mh\_vs\_mhsa.R*”). Alle andere R-Code und R-Data werden von den Hauptdateien aus eingelesen und durchgeführt. Die entsprechende Dokumentation zum R-Code sowie zur Ablagestruktur sind als .docx Dateien abgespeichert: “*0. Dokumentation zum Code (Analyse MH vs. MH-SA).docx*” und “*0. Dokumentation zum Code (Zwei-Schritt-Bayes-Verfahren).docx*”.

Im Ordner “R-Output” befinden sich die .csv Tabellen mit geschätzten Parametern und VaR sowie 2 Unterordner mit Grafiken zum Zwei-Schritt-Bayes-Ansatz und zu den Analysen der MH- und MH-SA-Algorithmen:

- “Analysen”
- “Simulation Loss”

Der Ordner “R-Ddata” besteht aus zwei Unterordnern:

- “Analysen”
- “Simulation Loss”

In beiden Unterordnern befinden sich die abgespeicherte Workspace zur Umsetzung des Zwei-Schritt-Bayes-Ansatzes sowie Analysen zum Vergleich der zwei Algorithmen: MH- und MH-SA-Algorithmus.

Die Benennung von Tabellen .csv, Grafiken .pdf und .jpg und Workspace .RData für den Zwei-Schritt-Bayes-Ansatz besteht aus folgenden Bausteinen (getrennt mit dem Unterstrich):

- Name (für Plot, Tabelle bzw. Workspace)

- Datentyp
- Verteilung der Schadenshöhen
- Modell für Schadenshöhen
- Verteilung der Schadenshäufigkeiten
- Modell für Schadenshäufigkeiten
- Anzahl der internen Verluste
- Anzahl der externen Verluste
- Anzahl Iterationen (für Markov-Kette oder direkte Ziehung aus *a posteriori* )
- Anzahl Simulationen (für Vorhersageverteilung)

Die möglichen Ausprägungen der Bausteine sind in der Tabelle [B.1](#) dargestellt. Die Benennung von Grafiken `.pdf` und Workspace `.RData` für Analysen besteht aus folgenden Bausteinen (getrennt mit dem Unterstrich):

- Name (für Analyse bzw. Plot)
- Anzahl der Iterationen

## Benennung des Outputs

Baustein	Ausprägungen
Name Plot	<i>“histogramme_int_vs_ext”</i> <i>“qq_plot_int_vs_ext”</i> <i>“log_log_plot_int_vs_ext”</i> <i>“log_log_plot_int_vs_ext_vs_geschaetzt”</i> <i>“acf_step1”</i> <i>“acf_step2”</i> <i>“meanlog_step1”</i> <i>“meanlog_step2”</i> <i>“sdlog_step1”</i> <i>“sdlog_step2”</i>
Name Tabelle	<i>“1_predicted_parameters0.9995”</i> <i>“1_predicted_VaR0.9995”</i>
Name Workspace	<i>“bayes_estimate_two_step_approach”</i>
Datentyp	<i>“sim”</i> <i>“true”</i>
Verteilung der Schadenshöhen	<i>“Lognormal”</i>
Modell für Schadenshöhen	<i>“ext_bayes”</i> <i>“ext_bayes_mh”</i> <i>“int_onestepbayes_mh”</i> <i>“ext_onestepbayes_mh”</i> <i>“together_onestepbayes_mh”</i>
Verteilung der Schadenshäufigkeiten	<i>“Poisson”</i>
Modell für Schadenshäufigkeiten	<i>“int”</i> <i>“int_bayes”</i> <i>“ext_bayes”</i> <i>“expert_bayes”</i>
Anzahl der internen Verluste	z. B. <i>“int35”</i>
Anzahl der externen Verluste	z. B. <i>“ext300”</i>
Anzahl Iterationen	z. B. <i>“it1000”</i>
Anzahl Simulationen	z. B. <i>“sim1500”</i>

---

**Bemerkung:** Genauer dazu siehe Dokumentation und Kommentaren in R-Code.

---

Tabelle B.1.: Benennung des Outputs.