

Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

Christian Langesberg^{*†} Uwe Ligges^{*} Claus Weihs^{*}

2. Mai 2018

Zusammenfassung Die Bildung von Mittelwerten bei ausreichend großen Stichproben führt gemäß dem Zentralen Grenzwertsatz (ZGWS) zu normalverteilten Größen. Bei gleitenden Mittelwerten bildet die Fensterbreite die bei der Konvergenz relevante Größe, und auch der Grad der Überlappung aufeinander folgender Fenster ist als Parameter zu berücksichtigen. Gemeinsam mit einem vorbereitenden Schritt der Datenverarbeitung und auf der Basis einer Simulationsstudie kann eine allgemeine Empfehlung für die Kombination von Fensterbreite und Schrittweite gegeben werden. Verschiedene Formen von Verletzungen der Annahme unabhängig identisch verteilter Beobachtungen können die für Anwendung des ZGWS toleriert werden.

Stichworte Zentraler Grenzwertsatz, Normalisierung, Zeitreihen, gleitender Mittelwert, Simulationsstudie

^{*}TU Dortmund

[†]E-Mail: clangesberg@statistik.tu-dortmund.de

Inhaltsverzeichnis

1. Einleitung	3
2. Grundlegendes	3
2.1. Voraussetzungen	3
2.2. Vorgehen	6
3. Simulationsstudie	10
3.1. Simulationsdesign	10
3.2. Ergebnisse	14
3.2.1. Die Potenzen d und k	14
3.2.2. Die Breite w und die Weite s	19
3.3. Folgerungen	23
4. Allgemeinere Szenarien	23
4.1. Konvergenzgeschwindigkeit	23
4.2. Unabhängig, nicht identisch verteilt	26
4.3. Nicht unabhängig, nicht identisch verteilt	27
4.4. Datenbeispiel	30
5. Fazit	33
A. Ergänzende Grafiken	34
Literatur	38

1. Einleitung

Das Adjektiv *zentral* im Namen des Zentralen Grenzwertsatzes (ZGWS) kann auf zwei Weisen verstanden werden: Zum einen handelt es sich um den *Satz vom zentralen Grenzwert*, was auch im Englischen bei der Verwendung des Namens *central limit theorem* deutlicher wird. Zum anderen ist der ZGWS aber auch ein *zentraler Satz* in der Statistik. Beide Lesarten verdeutlichen die Wichtigkeit dieses Satzes.

In zahlreichen Veröffentlichungen werden der Satz selbst, sowie dessen Eigenschaften, Folgerungen und Erweiterungen diskutiert. Eine der vielförmigen möglichen Anwendungen soll auch in der vorliegenden Ausarbeitung näher beleuchtet werden: Die Anwendung des ZGWS für gleitende Mittelwerte einer Zeitreihe. Damit können beispielsweise Glättungen durchgeführt werden, deren Verteilung dann bekannt ist. Auch wenn die Verteilung der Zeitreihenelemente selbst nicht bekannt ist, stehen für die Mittelwerte verschiedenste statistische Werkzeuge zur Verfügung, welche bekannte Verteilungen voraussetzen. So können beispielsweise Methoden aus der statistischen Prozesskontrolle besser justiert werden, wenn auch die theoretischen Eigenschaften eines Prozesses bekannt sind.

Die Anwendung des ZGWS als gleitendes Fenster führt insbesondere zu der Frage, wie stark sich die Überlappung aufeinander folgender Fenster auswirkt. Für disjunkte Fenster kann sich eine sehr starke Ausdünnung der Werte ergeben, bei starken Überlappungen hingegen sind die resultierenden gleitenden Mittelwerte autokorreliert, was zu unerwünschten Effekten führen kann. Vor diesem Hintergrund soll eine praktikable Fensterbreite ermittelt werden. Dabei sollen die Ergebnisse der Untersuchungen möglichst unabhängig von der Verteilung der Zeitreihenwerte sein. Zudem kommen Transformationen der Daten zur schnelleren „Normalisierung“ in Frage, welche hier ebenfalls betrachtet werden sollen.

Im folgenden Kapitel werden zunächst einzelne Begriffe und Notationen zusammengestellt, und es wird insbesondere das Prozedere zur gleitenden Normalisierung einer Stichprobe vorgestellt. Dieses Vorgehen wird dann mit einer Simulationsstudie genauer untersucht, welche im Kapitel 3 einschließlich Design, Ergebnissen und Schlussfolgerungen vorgestellt wird. Dabei wird stets von stochastisch unabhängigen Größen mit identischer Verteilung ausgegangen. Es folgt ein Blick auf Verletzungen dieser Annahme, welche in vielen Fällen *nicht* zu Problemen bei der Anwendung führen. Eine abschließende Beurteilung erfolgt dann im Kapitel 5.

2. Grundlegendes

2.1. Voraussetzungen

Bereits in der Einleitung, aber insbesondere auch in den folgenden Teilen werden verschiedene Begrifflichkeiten, Notationen und Zusammenhänge verwendet. Diese sollen hier kurz zusammengefasst werden. Die Ausgangslage besteht immer aus Zufallsvariablen, welche

Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen der Forschergruppe FOR 1511 gefördert.

durch Majuskel X, Y oder Z notiert werden, oder aus Realisationen, welche durch Minuskel x, y oder z notiert werden. Für Stichproben der Größe $n \in \mathbb{N}$ wird ein Index $i \in \mathbb{N}$, $1 \leq i \leq n$, verwendet; bei Verwendung des gleichen Buchstabens X liegt den X_i oder x_i jeweils eine identische Verteilung zugrunde. Wenn nicht anders notiert, seien diese Größen stochastisch unabhängig.

Der Zentrale Grenzwertsatz (ZGWS) liegt in mehreren Formulierungen vor. Eine kurze Übersicht findet sich bei [Patel und Read, 1996] im Kapitel 6, wobei auch die historische Entwicklung berücksichtigt wird. So beginnen die Erkenntnisse über Summen von standardisierten Zufallsvariablen mit dem Theorem von de Moivre und Laplace, welches Aussagen über den Limes der Verteilungsfunktion von standardisierten binomial verteilten Zufallsvariablen ermöglicht (ebenda, Seite 146). Pierre-Simon de Laplace bewies dann zu napoleonischen Zeiten Anfang des 19. Jahrhunderts „den ersten allgemeinen Zentralen Grenzwertsatz“ (ebenda, Seite 2).

Im Folgenden wird in erster Linie die Variante des ZGWS von Lindeberg und Lévy verwendet. Diese findet sich beispielsweise bei [Witting und Müller-Funk, 1995] auf Seite drei, wobei hier als „Bewertungsfunktion“ h die Identität gewählt wird. Mit dem Erwartungswert $\mu := E(X_1)$ ($= E(X_2) = E(X_3) = \dots$), der Varianz $\sigma^2 := V(X_1) > 0$ und dem arithmetischen Mittel $\bar{X} = 1/n \sum_{i=1}^n X_i$ gilt demnach

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} Y \sim N(0, \sigma^2)$$

$$\Leftrightarrow F_Z(z) := P(Z \leq z) := P\left(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq z\right) \xrightarrow{n \rightarrow \infty} \Phi(z) \quad \forall z \in \mathbb{R}.$$

Dabei ist F_Z die Verteilungsfunktion der angegebenen Transformation der X_i , Φ ist die Verteilungsfunktion der Standardnormalverteilung. Da die Normalverteilung invariant gegenüber linearen Transformationen ist, bedeutet der ZGWS auch, dass die Summe der X_i gegen eine Normalverteilung mit entsprechenden Parametern konvergiert.

Soll der Abstand von F_Z und Φ untersucht werden, so kann die Kolmogorov-Metrik verwendet werden. Diese ist für Verteilungsfunktionen F und G definiert durch

$$K(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|,$$

und findet sich unter anderem bei [Rüger, 2002], siehe Seite 50. Der ZGWS bedeutet als auch:

$$K(F_Z, \Phi) \xrightarrow{n \rightarrow \infty} 0.$$

Es gibt zahlreiche Erweiterungen und Übertragungen dieses prominenten Grenzwertsatzes auf komplexere Anwendungen, als das hier angenommene Szenario unabhängig und identisch verteilter Beobachtungen. Ein Teil dieser veränderten Annahmen an die Datensituation wird unten im Kapitel 4 weiter beleuchtet.

Als Zeitreihe wird eine Folge von Zufallsvariablen $\{X_{t_i}\}_{t_i \in \mathbb{R}}$ oder von Realisierungen $\{x_{t_i}\}_{t_i \in \mathbb{R}}$ bezeichnet, wobei $t_i \in \mathbb{R}$ der Index der Zeitpunkte ist und die Menge der

$i \in \mathbb{Z}$ zweiseitig, einseitig oder nicht beschränkt sein kann. Im Folgenden wird stets von äquidistanten Abständen $t_i - t_{i-1} \equiv c$ für eine echt positive reelle Konstante c ausgegangen, womit die t_i auch geordnet sind. Es kann dann kurz $\{x_i\}$ für $\{x_{t_i}\}_{t_i \in \mathbb{R}}$ geschrieben werden.

Eine auf eine Zeitreihe mittels eines gleitenden Fensters der Breite $w \in \mathbb{N}$ angewandte Funktion $f : \mathbb{R}^w \rightarrow \mathbb{R}$ wird hier stets als „online“ angewandt verstanden, womit nur Werte für den jeweils aktuellen und vergangene Zeitpunkte vorliegen. Wird also mittels f eine Zeitreihe $\{y_i\}$ aus der Zeitreihe $\{x_i\}$ gebildet und dazu ein gleitendes Fenster verwendet, so gilt für das i -te Element

$$y_i = f(x_i, x_{i-1}, \dots, x_{i-w+1}).$$

Für eine gegebene, endliche Zeitreihe der Länge $n \in \mathbb{N}$ muss dazu offenbar $w \leq n$ sein. Ist die Indexmenge nach unten begrenzt, was insbesondere bei Realisierungen immer der Fall ist, sind die ersten Elemente y_1 bis y_{w-1} nicht definiert. Es bestehen zwar Methoden, dem zu begegnen, so kann beispielsweise der Wert y_w fortgeschrieben werden, oder auch ein an die jeweilige Funktion f angepasstes spezielles Verfahren Anwendung finden. Für die folgenden Analysen wird darauf aber verzichtet: Entweder sind die Resultate sonst auch durch die Wahl des gewählten Vorgehens beeinflusst, oder es ist $w \ll n$, so dass der Einfluss des Auslassens nicht relevant ist. Es folgt dann, dass Stichproben der Größe n zu Stichproben der Größe $n - w$ verkleinert werden.

Die Korrelation zweier aufeinander folgender y_i wird groß, wenn w groß gewählt wird: Die Mengen der Elemente zweier aufeinander folgender Fenster, das bedeutet die Mengen $\{x_i, x_{i-1}, \dots, x_{i-w+1}\}$ und $\{x_{i+1}, x_i, \dots, x_{i-w+2}\}$, unterscheiden sich nur um den Anteil $1/w$. Es kann daher sinnvoll sein, die Zeitreihe $\{y_i\}$ auszudünnen und nur eine Teilmenge bestehend aus $y_{1+0s}, y_{1+1s}, y_{1+2s}, \dots$ zu betrachten. Dabei wird $s \in \mathbb{N}$ als Schrittweite bezeichnet. Ist $s = 1$, so hat die Einführung dieser Größe offenbar keine Auswirkungen, für alle $s \geq w$ entstehen y_i aus disjunkten Mengen. Für unkorrelierte x_i sind dann auch die y_i unkorreliert. Eine Illustration der Parameter w und s findet sich als Abbildung 1.

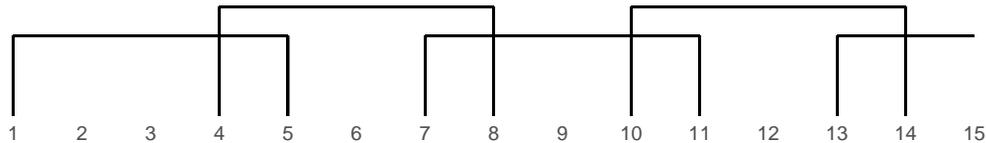


Abbildung 1: Mit einer Fensterbreite $w = 5$ und einer Schrittweite $s = 3$ ergeben sich die dargestellten Indexmengen für die Anwendung der gleitenden Funktion. Statt 15 Werten liegen hier nur noch vier Gruppen vor.

Durch dieses Prozedere verkleinert sich die Stichprobengröße für die Anwendung gleitender Funktionen je nach Fensterbreite und Schrittweite auf

$$m = \left\lceil \frac{n - w}{s} \right\rceil.$$

2.2. Vorgehen

Das Ziel des hier untersuchten Vorgehens besteht darin, aus einer gegebenen Zeitreihe identisch verteilter Elemente x_i mit sonst unbekanntem Eigenschaften eine Zeitreihe y_i zu erzeugen, welche einer bekannten Verteilung unterliegt. Die Verwendung von Grenzwertsätzen erscheint dabei als praktikabler Ansatz, da diese sehr wenige Voraussetzungen an die x_i stellen - insbesondere keine bestimmte Verteilung. Mit dem ZGWS kann dabei eine Normalverteilung als Limes verwendet werden. Dann folgen Beobachtungen aus einer ursprünglich unbekanntem Verteilung nach einer Transformation nur nicht nur einer definierten, sondern auch einer gut untersuchten Verteilung.

Mit der Wahl dieses Grenzwertsatzes wird dann der Mittelwert verwendet, welcher als weiterer Vorteil eine einfach zu berechnende und gut interpretierbare Größe darstellt. Da den x_i eine Zeitreihe zugrunde liegt, wird dieser in einem gleitenden Fenster angewandt - wobei dann folglich $f(x_i, x_{i-1}, \dots, x_{i-w+1}) = 1/w \sum_{j=0}^{w-1} x_{i-j}$ ist. Die Wahl der Fensterbreite w wird im Folgenden thematisiert, für die Sicherstellung kleiner Abstände der (Verteilung der) transformierten Beobachtungen zur Normalverteilung ist es aber offenbar hilfreich, große w zu wählen.

Soll für die Transformationen y_i nicht nur eine bekannte und identische Verteilung erreicht werden, sondern zusätzlich auch die Unabhängigkeit der Beobachtungen erhalten bleiben, so ist nach den vorigen Überlegungen auch die Schrittweite s als zweiter Parameter zu berücksichtigen. In diesem Sinne ist es offenbar sinnvoll, $s = w$ zu wählen. Dem gegenüber steht aber der Informationsverlust, der sich durch die Aggregation ergibt: Soll möglichst viel von der Struktur der x_i in den daraus abgeleiteten y_i erhalten bleiben, stehen diesem Ziel große Breiten und Schrittweiten offenbar entgegen.

Bei der Betrachtung von endlich vielen Beobachtungen, also insbesondere bei der Analyse von Stichproben, kann die Schrittweite über den Umweg der Autokorrelation auch einen Einfluss auf die Verteilung der y_i haben, wenn die Stichprobe verhältnismäßig klein ist. Als Beispiel stelle man sich eine zufällige Zeitreihe identisch verteilter Beobachtungen vor, welche der Größe nach sortiert werden. Die Sortierung ändert die Verteilung der Daten offenbar nicht, führt aber zu einer starken Autokorrelation der Reihe. Wird nun nur ein kurzer Ausschnitt der sortierten Reihe alleinstehend untersucht, wird die Streuung viel geringer ausfallen, als in der gesamten Datenreihe. Die eigentliche Verteilung wird erst deutlich, wenn ein ausreichend langer Ausschnitt der Daten herangezogen wird, oder die Stärke der Autokorrelation diesen Effekt nicht zu stark forciert.

Neben sinnvollen Einstellungen dieser beiden Parameter können für eine möglichst hohe Konvergenzgeschwindigkeit aber auch bestimmte Schritte der Vorverarbeitung der Daten gewinnbringend sein: So sind beispielsweise gerade für kleine Stichprobengrößen von einer symmetrischen Verteilung „stabiler“ Mittelwertreihen zu erwarten, als von ei-

ner stark asymmetrischen, welche viele Ausreißer produziert. Als vorbereitender Schritt wird daher überprüft, inwiefern die Daten „symmetrisiert“ werden können. Dazu können unter anderem Differenzen betrachtet werden: Diese folgen für identisch verteilte Zufallsvariablen grundsätzlich einer symmetrischen Verteilung, da Minuend und Subtrahend den gleichen Verteilungen entstammen. Wenn auch hier nicht weiter betrachtet, ist die Differenzenbildung zudem auch ein oft genutztes Werkzeug, um eventuelle Autokorrelationen der Ausgangsdaten zu filtern.

Bei der Betrachtung von Zeitreihen fällt eine einfache Summe von Differenzen allerdings als Teleskopsumme in eine einzelne Differenz zusammen:

$$(x_2 - x_1) + (x_3 - x_2) + \dots + (x_w - x_{w-1}) = x_w - x_1$$

Aus diesem Grund werden hier absolute Abstände verwendet. Weiter kommt anschließend eine zweite Variante zum Einsatz: So können schief verteilte Beobachtungen auch durch die Betrachtung von Potenzen in eine symmetrische oder zumindest *symmetrischere* Form gebracht werden. Die Idee findet sich etwa bei der Box-Cox-Transformation als bekanntem Vertreter der „power transforms“ wieder. Die Box-Cox-Transformation selbst ist etwas komplexer als hier angewandt², hat aber ebenfalls das Ziel, nicht-normale Daten in eine Normalverteilung zu überführen. Beispiele geben [Judge et al., 1985] auf den Seiten 839 ff. Vor der Bildung der gleitenden Mittelwerte kann also eine Datenvorverarbeitung in Form von $\tilde{x}_i = |x_{i+1} - x_i|^d$ die Konvergenzgeschwindigkeit der Mittelwert-Verteilung erhöhen.

Ebenfalls eine power transform wird *nach* der Mittelwertbildung in Betracht gezogen. Damit kann für eine bestimmte Wahl der Parameter (s. u.) die gut untersuchte Minkowski-Metrik nachgebildet werden. Es wird aber zunächst keine Einschränkung für diese zweite Potenz eingebracht: Auch hier kann eine gegebenenfalls schiefe Verteilung durch Ansetzen eines entsprechenden Hebels näher die Normalverteilung „gedrückt“ werden. Neben der Funktion $f(\cdot) = \bar{x}$ als gleitendem Mittelwert wird also auch die Variante $f(\cdot) = \bar{x}^k$ berücksichtigt.

Um aus einer gegebenen Zeitreihe normalverteilte Werte zu erzeugen, werden also die beschriebenen Schritte unternommen. Dabei ist die Notation so gewählt, dass neue Beobachtungen in Form der Reihe y_1, y_2, \dots, y_m gebildet werden, die $n - m$ leeren Elemente also gelöscht werden. Eine Zeitreihe bestehend aus Differenzen ist um eine Beobachtung kürzer als die Ausgangsreihe. Der Einfachheit halber wird im Folgenden auch für den Fall ohne Differenzenbildung eine Beobachtung ausgelassen, um stets neue Reihen mit gleicher Länge zu produzieren.

Da die Potenzierung der Ausgangsdaten für eine Potenz $d = 0$ nicht von Interesse ist, wird der Fall „keine Vorverarbeitung vornehmen“ durch $d = 0$ notiert. Das soll aber keine Fortsetzung (und insbesondere keine stetige Fortsetzung) andeuten.

²Eine „komplette“ Box-Cox-Transformation bestünde nicht nur aus der Potenz x^d , sondern aus dem Term $d^{-1}(x^d - 1)$, s. a. a. O.

Die drei Schritte bilden sich zusammenfassend durch

1. Symmetrische Differenzen: $\tilde{x}_i = \begin{cases} |x_{i+1} - x_i|^d & , d > 0 \\ x_{i+1} & , d = 0 \end{cases}, i = 1, 2, \dots, n - 1$
2. Gleitende Mittelwertbildung: $\tilde{y}_j = \frac{1}{w} \sum_{i=(j-1)s+1}^{(j-1)s+w} \tilde{x}_i, j = 1, 2, \dots, m$
3. „Hebeln“ der Mittelwerte durch $y_j = \tilde{y}_j^k$ für $k \geq 1$

Es werden nur Potenzen $d \in \mathbb{R}_{\geq 0}$ und $k \in \mathbb{R}_{\geq 1}$ berücksichtigt. Dabei soll die Beschränkung für k sicherstellen, dass die Potenz auch für Mittelwerte kleiner oder gleich der Null definiert ist. Da eine diskrete Verteilung der x_i nicht ausgeschlossen wird, hat auch ein Mittelwert gleich Null eine echt positive Wahrscheinlichkeit. Gleiches gilt für Differenzen mit dem Wert 0, so dass ein nicht-negatives d vorausgesetzt wird.

Ob statt der Mittelwertbildung im zweiten Schritt Summen verwendet werden, ist wie oben bereits erwähnt unerheblich für die Normalverteilungseigenschaft (Invarianz gegenüber linearen Transformationen). Je nach numerischer Größe der Differenzen kann daher die passendere Skalierung gewählt werden. Für Summenbildung, $d \geq 1$ und $k = 1/d$ ergibt sich die Minkowski-Metrik.

Die Abbildung 2 zeigt eine beispielhafte Durchführung des Verfahrens. Dabei wurden Werte aus der χ_5^2 -Verteilung gezogen und für das angegebene Parameter-Setting normalisiert. Es ist zu erkennen, dass mit den Transformationen eine Glättung und eine Ausdünnung einhergehen. Währenddessen wird die Normalverteilung immer weiter angenähert. Zu untersuchen ist nun, für welche Parameter dieses Vorgehen zum Ziel führt, und wie gleichzeitig die Ausdünnung möglichst gering gehalten werden kann.

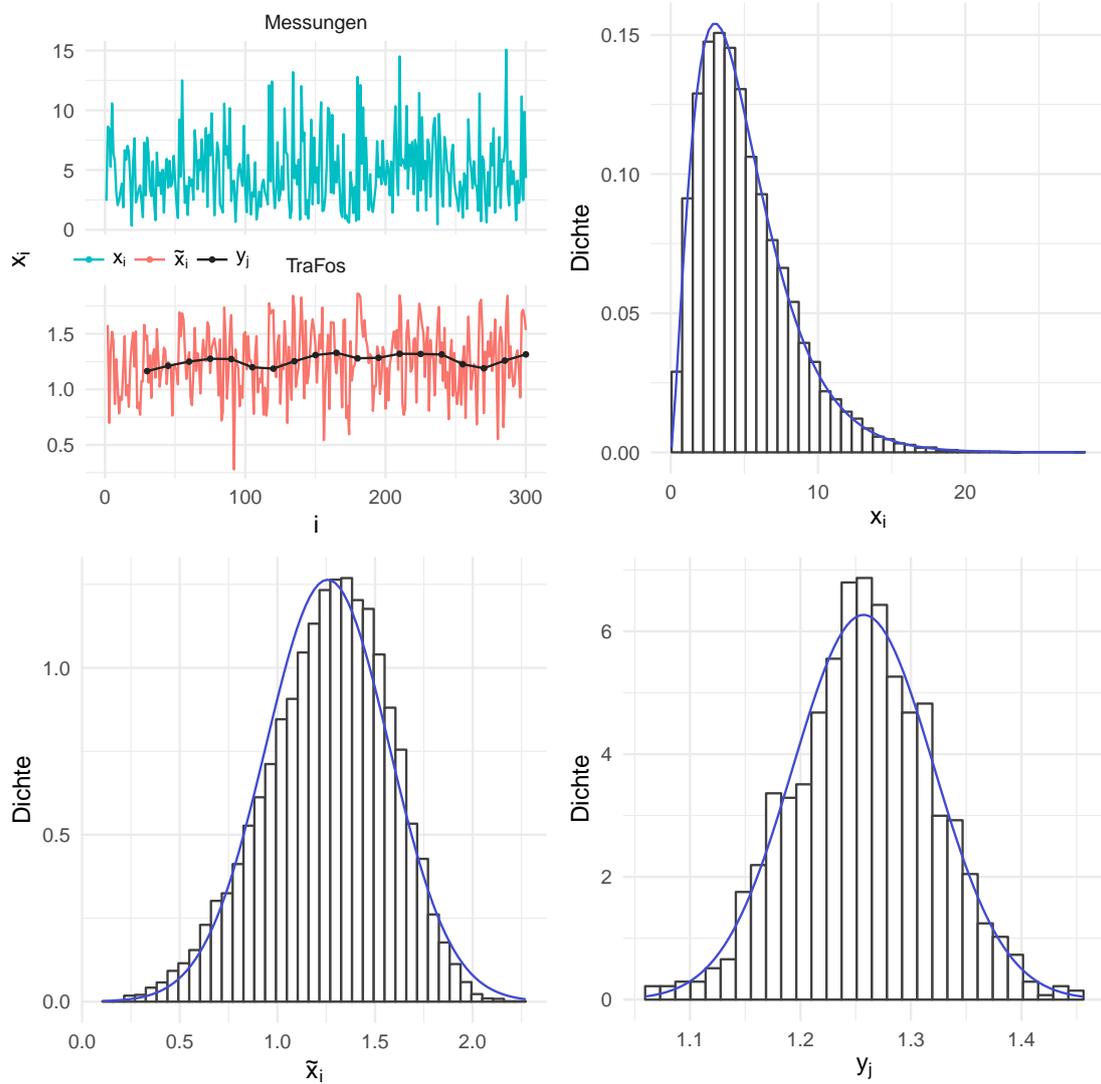


Abbildung 2: Das beschriebene Vorgehen beispielhaft für $w = 30$, $s = w/2 = 15$, $d = 0.25$, $k = 1$ und $n = 15\,016$, was $m = 1\,000$ impliziert. Dargestellt ist die Glättung, welche mit jedem Schritt stärker ausfällt. Dabei ist für den Verlauf oben links nur ein kleiner Ausschnitt der Daten dargestellt. Die Histogramme sind ergänzt mit der Dichte der χ^2_5 -Verteilung (oben rechts, x_i), beziehungsweise der Normalverteilung mit den jeweiligen Parameterschätzern.

3. Simulationsstudie

Es sollen nun die Eigenschaften und optimalen Einstellungen des beschriebenen Verfahrens untersucht werden, wozu eine Simulationsstudie durchgeführt wird. Dazu sind sechs „Stellschrauben“ zu berücksichtigen: Neben der Verteilung der Ausgangswerte sind die Fensterbreite, die Schrittweite und die beiden Potenzen variabel, weiter werden mögliche Einflüsse der Stichprobengröße $n \in \mathbb{N}$ berücksichtigt.

3.1. Simulationsdesign

Für die Verteilung der x_i soll eine möglichst breite Auswahl häufig auftretender statistischer Verteilungen verwendet werden, ohne dass eine tatsächlich umfassende und repräsentative Auswahl ohne Weiteres möglich ist. Es werden an dieser Stelle stellvertretend 13 Verteilungen berücksichtigt: Die Standardnormalverteilung sowie jeweils zwei Gleich-, χ^2 -, t -, Weibull-, Poisson- und Binomialverteilungen. Die Wahl der Parameter kann der unten folgenden Liste auf der Seite 13 oder der Abbildung 3 auf der folgenden Seite entnommen werden. In dieser Grafik sind alle verwendeten Dichtefunktionen dargestellt. Die Parameter sind so gewählt, dass in allen Fällen das dritte und vierte Moment existieren, somit Schiefe und Kurtosis berechnet werden können.

Aus diesen Verteilungen werden nun Zufallsstichproben gezogen und die beschriebenen Transformationen angewandt. Anschließend wird die Distanz zur Normalverteilung als logarithmierte Jarque-Bera-Statistik sowie nach Kolmogorov mit Berücksichtigung der Stichprobengröße berechnet. Im zweiten Fall wird dazu als Vertreter der Normalverteilungsfamilie eine Verteilungsfunktion G mit Parameterschätzungen durch Mittelwert und Stichprobenvarianz verwendet, die empirische Verteilungsfunktion wird mit F_n bezeichnet. Der Abstand der gegebenenfalls transformierten Daten zur Normalverteilung wird folglich beurteilt durch

$$K^* = \sqrt{n} K = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - G(x)|$$

und

$$LJB = \log(JB + 1) = \log\left(\frac{n}{6} \left(M_3^2 + \frac{(M_4 - 3)^2}{4}\right) + 1\right)$$

mit M_3 und M_4 als drittem und viertem empirischen Moment. Die Auswahl gerade dieser Maße wird in [Langesberg et al., 2018] neben deren Eigenschaften ausführlich diskutiert.

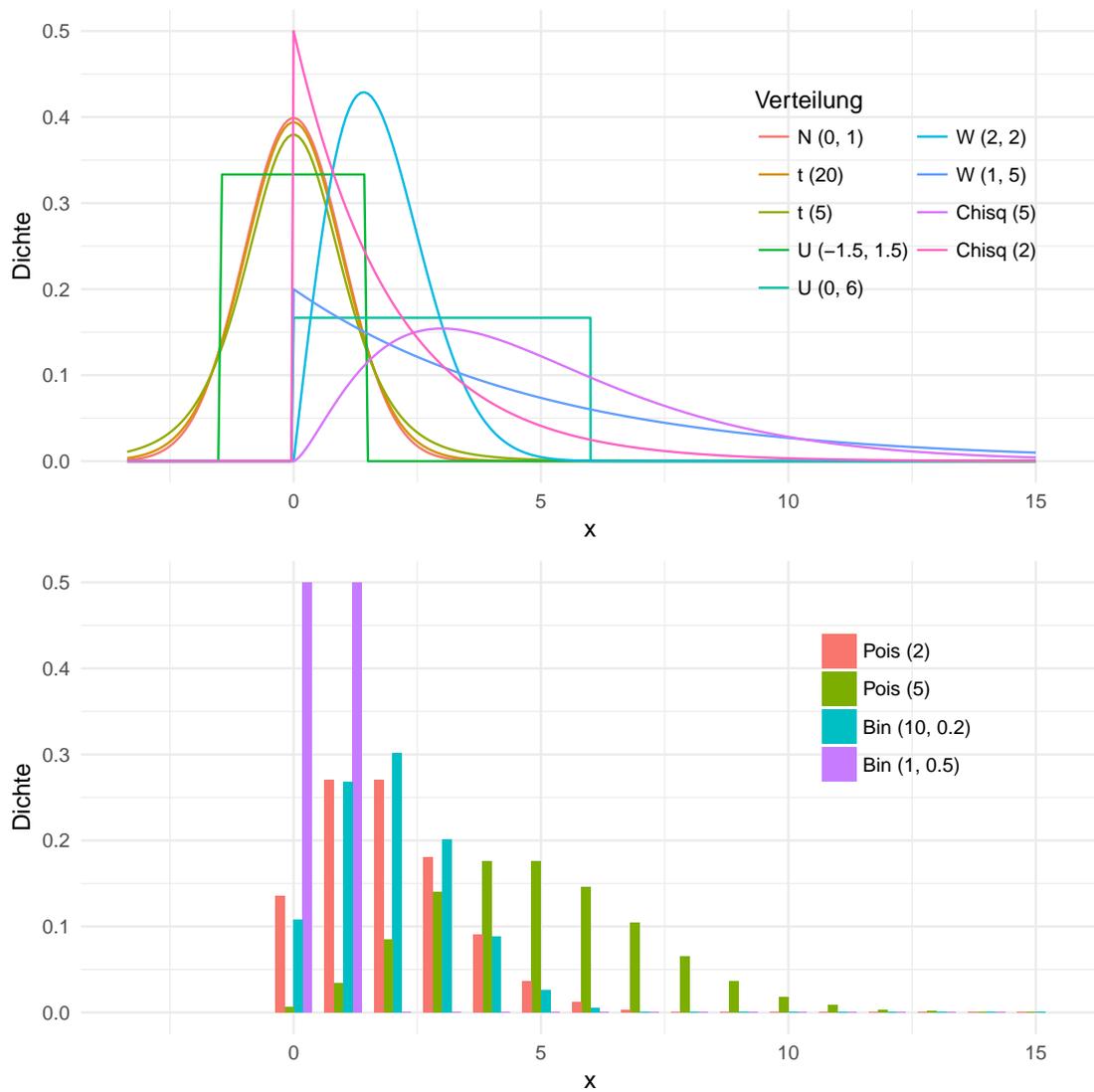


Abbildung 3: Die dreizehn im Folgenden betrachteten Verteilungen: Acht stetige und vier diskrete Verteilungen sollen durch das gleitende Mittel „normalisiert“ werden. Außerdem wird das Verhalten für Werte betrachtet, welche schon aus einer (Standard-)Normalverteilung stammen.

Als Ankerpunkt für Vergleiche soll die Abbildung 4 dienen. Für diese wurden die beiden verwendeten Differenzmaße bei normalverteilten Zufallsdaten in den später verwendeten Stichprobengrößen für $m = 100$ und $m = 1\,000$ berechnet (m entspricht wie oben der Stichprobengröße *nach* der Transformation). Bei beiden Maßen ist für perfekt normalverteilte Daten die Null zu erwarten, diese ist aber gleichzeitig der kleinstmögliche Wert, so dass die Null nicht der mittleren Lage der streuenden Realisierungen entspricht.

Für den Kolmogorov-Abstand ergeben sich größtenteils Werte zwischen 0.42 und 0.90 (5- und 95-Prozent-Quantil), im Mittel 0.62. Für die Jarque-Bera-Variante betragen diese Größen 0.10, 0.87 und 1.97; in beiden Fällen sind Unterschiede zwischen den Stichprobengrößen nicht als systematisch einzustufen. Die Maße für die beispielhaften Daten der Abbildung 2 oben betragen $K^* = 10.5$ und $LJB = 8.7$ für die Ausgangsdaten, und sinken dann ab: Für die \tilde{x}_i ist $K^* = 3.8$ und $LJB = 5.7$, für die y_j ist $K^* = 0.63$ und $LJB = 0.39$.

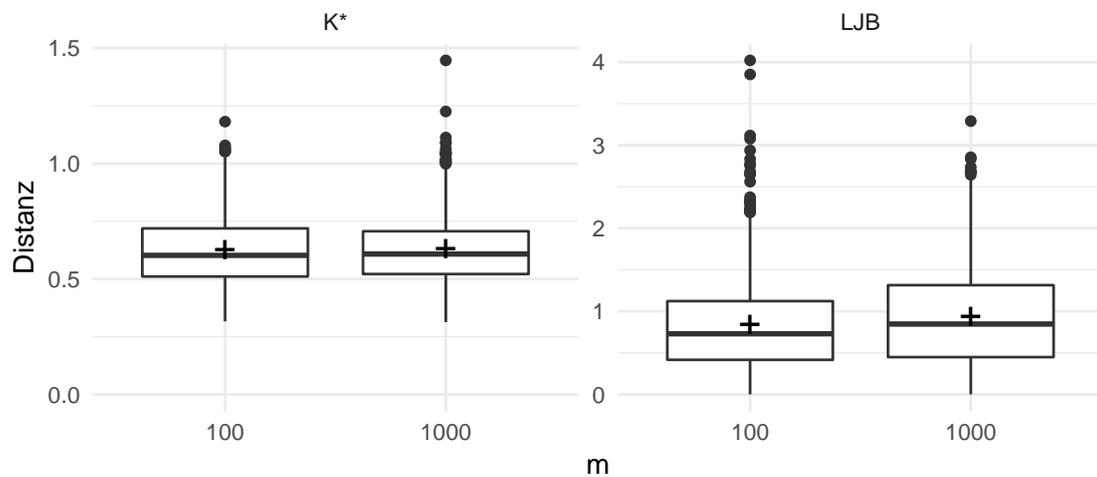


Abbildung 4: Wie nah eine Stichprobe an der Normalverteilung ist, wird durch die oben erläuterten Methoden K^* und LJB „gemessen“. Stammen die Daten aus einer Normalverteilung der Größe m , so ergeben sich die hier dargestellten Werte (enthalten sind jeweils Ergebnisse zu 1 000 Zufallsstichproben).

Für die Fensterbreite w werden in der Simulation absolute Werte wie unten angeführt gewählt. Die Schrittweite s bestimmt, wie stark sich zwei aufeinander folgende Fenster überdecken. Der Grad dieser Überdeckung ist aber auch abhängig von der Fensterbreite, so führt beispielsweise eine Schrittweite von 20 Beobachtungen bei einer Fensterbreite von 10 Beobachtungen zu übersprungenen Bereichen zwischen zwei aufeinander folgenden Fenstern. Bei einer Fensterbreite von 20 ergeben sich für $s = 20$ disjunkte Intervalle ohne Zwischenräume, und bei einer Fensterbreite von 100 ergeben sich stark überdeckte, und damit autokorrelierte Bereiche. Daher wird statt dem Wert s selbst ein relativer Wert festgelegt, welcher abhängig von der jeweiligen Fensterbreite definiert wird: $s_{\text{rel}} = s/w$.

An dieser Stelle werden stochastisch unabhängige X_i angenommen, es liegt also keine

Autokorrelation vor. Damit besteht kein Unterschied zwischen allen Schrittweiten, die größer als die oder gleich der Fensterbreite sind: Alle $s_{\text{rel}} \geq 1$ führen zu disjunkten Intervallen. Es genügt hier somit, ausschließlich $s_{\text{rel}} \leq 1$ zu betrachten, das heißt $s \leq w$.

Die Ergebnisse werden *nach* den Transformationen verglichen, somit ist es plausibel, statt der Ausgangsgröße n die Anzahl der resultierenden y -Werte zu betrachten. Diese ist dann auch unabhängig von der Fensterbreite und der relativen Schrittweite, die benötigte Ausgangsgröße kann mittels $n = \lceil w + (m - 1)s \rceil + 1 = \lceil w + (m - 1)ws_{\text{rel}} \rceil + 1$ bestimmt werden. Durch die Addition der Eins wird die Bildung von Differenzen - beziehungsweise das Auslassen der ersten Beobachtung - im ersten Schritt des in Teil 2.2 beschriebenen Vorgehens berücksichtigt.

Für die Simulation wird nun ein vollständiges sechs-dimensionales Gitter aus den folgenden Punkten aufgespannt:

- Verteilungen³:
 - Normalverteilung: $N(0,1)$
 - Chi-Quadrat-Verteilung: χ_2^2, χ_5^2
 - t -Verteilungen: t_5, t_{20}
 - Gleichverteilung: $U(-1.5,1.5), U(0,6)$
 - Weibull-Verteilung: $W(1, 5), W(2,2)$
 - Poisson-Verteilung: $P(2), P(5)$
 - Binomialverteilung: $B(1,0.5), B(10,0.2)$
- w : 1, 4, 8, 12, 16, 20, 24, 28, 36, 48, 72 und 108
- s_{rel} : $1/w, 1/4, 2/4, 3/4, 1$
- m : 10^2 und 10^3
- d : 0, $1/4, 1/2, 1, 2$
- k : 1, 2, 3

Das Gitter umfasst demnach $13 \cdot 12 \cdot 5 \cdot 2 \cdot 5 \cdot 3 = 23\,400$ Kombinationen. Die Einstellung $s_{\text{rel}} = 1/w$ führt jeweils zur Schrittweite $s = 1$. Die Gitterpunkte für Fensterbreite und die relative Schrittweite wurden so gewählt, dass jeweils ganzzahlige $s = w \cdot s_{\text{rel}}$ erzeugt werden. Für $w = 1$ ergeben sich dennoch Schrittweiten kleiner als 1, wenn $s_{\text{rel}} < 1$ ist. Für $w = 1$ und $w = 4$ ist der Wert $s_{\text{rel}} = 1/w$ bereits durch die Einstellung $s_{\text{rel}} = 1$ beziehungsweise $s_{\text{rel}} = 1/4$ abgedeckt. Werden diese unzulässigen und doppelten Punkte entfernt, so verbleiben $23400 - 3 \cdot 390 - 2 \cdot 390 = 21\,450$ Gitterpunkte.

³Es wird hier stets von der „englischen Schreibweise“ der Weibull-Verteilung ausgegangen, das heißt von der Dichtefunktion $f(x; \lambda, k) = \frac{k}{\lambda} (x/\lambda)^{k-1} e^{-(x/\lambda)^k}$. Diese Form ist nicht zu verwechseln mit der mitunter, und oft in der deutschsprachigen Literatur, genutzten Notation mittels reziprok angegebene Skalenparameter, das heißt $f(x; \lambda^*, k) = k\lambda^* (x\lambda^*)^{k-1} e^{-(x\lambda^*)^k}$. Durch die Aufnahme einer Weibullverteilung mit Formparameter $k = 1$ ist implizit auch eine Exponentialverteilung vertreten.

Auch mit Blick auf die Potenzen führt ein Teil der Gitterpunkte trotz unterschiedlicher Einstellungen zu gleichen Transformationen, diese werden dann auf eine dieser Einstellungen reduziert. Betroffen sind drei Konstellationen: Zum einen betragen alle Differenzen der Bernoulli-Reihe nur Eins oder Null, also Werte, welche sich durch Potenzierung nicht ändern. Ebenfalls aus diesem Grund hat die Potenz k keinen Einfluss, wenn zusätzlich die Fensterbreite 1 angesetzt wird, also de facto keine Summen gebildet werden. Drittens ist, unabhängig von der Verteilung der x_i , für $w = 1$

$$y_j = \begin{cases} \left(\frac{1}{w} \sum_{i=(j-1)s+1}^{(j-1)s+w} |x_{i+1} - x_i|^d \right)^k = (|x_{(j-1)s+2} - x_{(j-1)s+1}|^d)^k & d \neq 0 \\ \left(\frac{1}{w} \sum_{i=(j-1)s+1}^{(j-1)s+w} x_{i+1} \right)^k = (x_{(j-1)s+1})^k & d = 0 \end{cases}$$

Damit ergeben sich die gleichen Werte für verschiedene Kombinationen von d und k . Für $k = 2$ wird $0 < d \leq 1$ hier herausgenommen, alle Produkte lassen sich bereits mit $k = 1$ realisieren.

Es verbleiben 20 380 Gitterpunkte. Für jede dieser Einstellungskombinationen werden nun 100 Wiederholungen simuliert. Für die indirekt festgesetzten Parameter s und n ergeben sich - nach der Bereinigung und ohne Berücksichtigung der Wiederholungen - 26 beziehungsweise 218 verschiedene Einstellungen. Die Schrittweite ergibt sich dabei zu Werten zwischen 1 und 108, die Ausgangsstichproben werden in Größen zwischen 100 und 108 001 benötigt. Es muss für jede Kombination von Verteilung, w , s_{rel} und m eine neue Stichprobe gezogen werden. Die Vergleiche für die Parameter d und k hingegen sind für jeweils die gleichen Daten möglich (was durch die unabhängig von d gleiche Länge der \tilde{x}_i sichergestellt ist, s. o.).

Die Simulationen sowie alle weiteren Berechnungen und Abbildungen dieser Ausarbeitung wurden mit der statistischen Programmiersprache R [R Core Team, 2017] durchgeführt.

3.2. Ergebnisse

3.2.1. Die Potenzen d und k

Aufgrund der Vielzahl an Einstellungen (sechs Dimensionen mit jeweils zwei bis dreizehn Variationen) ist eine übersichtliche Darstellung der Simulationsergebnisse nur sukzessive möglich. Im ersten Schritt werden an dieser Stelle die Parameter d und k betrachtet. Diese stehen zwar nur in mittelbarem Zusammenhang, da sie vor und nach der Mittelwertbildung zum Tragen kommen. Allerdings können sie für die gleichen Stichproben verglichen werden, so dass eine gemeinsame Analyse sinnvoll erscheint. Die folgende Tabelle 1 zeigt, wie häufig welche Kombination zum besten Ergebnis im Sinne der kleinsten Distanz geführt hat. Dargestellt sind die Werte für K^* , für LJB ergeben sich qualitativ keine Unterschiede.

Offenbar führt das Bilden von Differenzen (d. h. $d \neq 0$) mehrheitlich zu den besseren Ergebnissen, ist aber bei weitem nicht immer das beste Vorgehen. Klar ist jedoch,

	d=0	0.25	0.5	1	2	Insgesamt
k=1	21.1	28.5	19.2	5.2	2.0	76.1
2	1.6	10.3	3.3	0.9	0.2	16.4
3	1.0	4.4	1.6	0.5	0.2	7.5
Insges.	23.7	43.3	24.1	6.6	2.4	n=143 000

Tabelle 1: Über alle Einstellungen von w , s und m , sowie über alle Verteilungen hinweg, führt eine Kombination von d und k zur jeweils besten Annäherung an die Normalverteilung. Angegeben sind die Anteile in Prozent, zu denen eine bestimmte Kombination das beste Resultat liefert.

dass die Differenzen mit einer Potenz kleiner als Eins zu versehen, sprich zu radizieren sind. Das ist plausibel: Da absolute Differenzen betrachtet werden, sind die Zwischenergebnisse $|x_{i+1} - x_i|$ tendenziell rechtsschief verteilt, mit Potenzen $0 < d < 1$ wird die Schiefe dann gesenkt. Dass sich die Verteilung einer Summe symmetrisch verteilter Werte schneller einer Normalverteilung annähert, erscheint wie oben bereits vermutet, plausibel: Bei kleiner Stichprobengröße können einzelne Ausreißer die Verteilung der Mittelwerte noch recht stark beeinflussen, und die Wahrscheinlichkeit für das Auftreten eben dieser Ausreißer steigt mit der Schiefe einer Verteilung.

Die (hier nicht aufgeführten) Gegenüberstellungen der Einstellung von d mit den Größen w und s zeigt nur schwache Zusammenhänge auf. Stärker fällt hingegen der Einfluss der Verteilungen aus. Die Abbildung 5 gibt einen Eindruck der Systematik. Darin sind, wie in der vorangegangenen Tabelle, die relativen Häufigkeiten dargestellt, zu denen ein bestimmtes d zu den besten Resultaten führt.

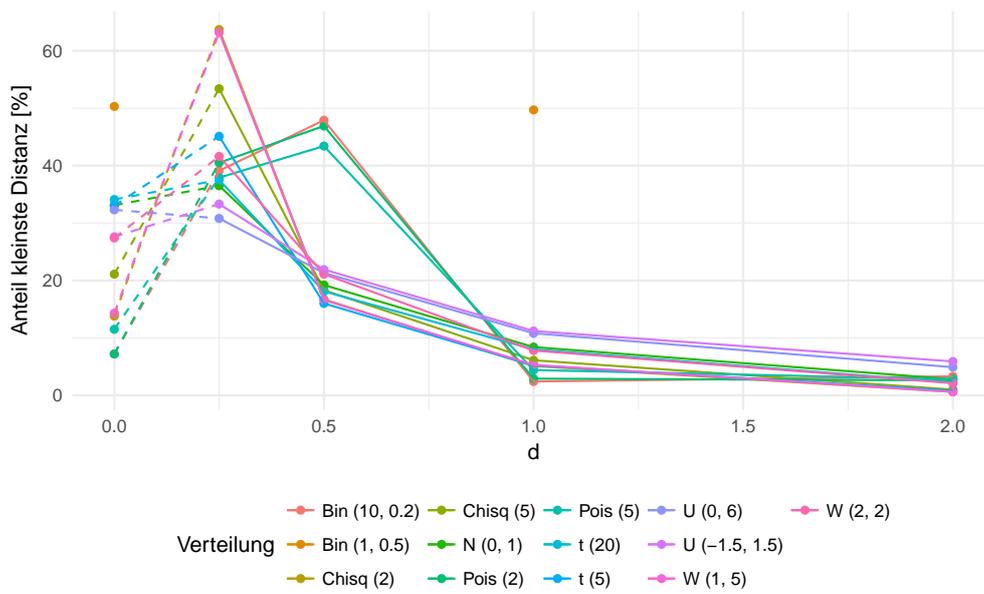


Abbildung 5: Je nach Verteilung führt ein bestimmtes d zum kleinsten Distanzmaß. Die Anteile für jedes d sind hier dargestellt. Zu beachten ist, dass der Wert $d = 0$ keine stetige Fortführung des Kontinuums ist. Diese Einstellung symbolisiert die Verwendung der Werte selbst, statt der Differenzen.

Es lassen sich vier Gruppen ausmachen: Für einen Teil der stetigen Verteilungen ergibt sich die klare Aussage „ $d = 0.25$ ist in den meisten Fällen die beste Wahl“. Das ist für die χ^2 - und die Weibull-Verteilungen der Fall, welche auch recht große Schiefen aufweisen. Die übrigen stetigen Verteilungen weisen eher geringe Unterschiede zwischen der Variante ohne Differenzen und $d = 0.25$ auf. Es handelt sich dabei um die symmetrischen Verteilungen (N, t und U). Das stützt die These, dass vor der Mittelwertbildung eine möglichst symmetrische Verteilung anzustreben ist.

Eine dritte Gruppe bilden die beiden Poisson- und die Bin(10, 0.2)-Verteilungen: hier ist $d = 0.5$ am häufigsten die beste Wahl. Es gelten vermutlich die gleichen Effekte bezüglich der der Schiefe, jedoch kommt hier die diskrete Struktur zum Tragen: Differenzen ungleicher Zahlen betragen dann absolut mindestens Eins, während sie bei stetigen Verteilung auch Werte nahe der Null annehmen können. Als vierte Klasse ist die Bernoulli-Verteilung zu nennen: Für diese sind wiederum $d = 0$ und $d = 1$ ähnlich gut, aus den oben genannten Gründen liegen keine Werte für die anderen Einstellungen von d vor. Die Gleichheit der beiden angeführten Datenpunkte begründet sich ebenfalls in der speziellen Struktur dieser Verteilung: Die Werte stammen aus der Menge $\{0, 1\}$, was dann auch für die absoluten Differenzen gilt. Die einzelne Reihe ändert sich zwar, die Verteilung aber bleibt für absolute Differenzen die gleiche.

Insbesondere ist an der Abbildung auch die systematische Abweichung aller Verläufe von einer horizontalen Geraden zu erkennen: Diese wäre zu erwarten, wenn alle Werte für d gleich gut wären und die Lage des besten Parameters damit aus einer Gleichverteilung stammen würde. Insgesamt ergibt sich also der Eindruck, dass die optimale Lage von d insbesondere durch die Schiefe beeinflusst wird und in Abhängigkeit von dieser zu optimieren ist. Da für die stetigen Verteilungen $d = 0.25$ sowie für die diskreten Verteilungen $d = 0.5$ beziehungsweise $d = 1$ (Bernoulli) als durchweg zumindest akzeptable Wahl erscheinen, werden diese für die folgenden Auswertungen angesetzt, die Ergebnisse auf diese Bereiche reduziert.

Recht eindeutig erscheint das Bild für den Parameter k : Gemäß der obigen Tabelle 1 könnte der Parameter in vielen Fällen ganz ausgelassen werden, da in den meisten Fällen ohnehin die Eins die beste Wahl ist. Dieser Wert stellt im Sinne der Potenz die Identitätsabbildung dar. Allerdings ist bei k zu berücksichtigen, ob die beiden vorangegangenen Schritte zur Distanzen- und Mittelwertbildung bereits zum Erfolg geführt haben: Ist für \tilde{y}_j bereits die Normalverteilung „erreicht“, so kann $y_j = \tilde{y}_j^k$ nicht noch näher heran rücken. Hier könnte folglich eine Fallunterscheidung zielführend sein. Die Abbildung 6 zeigt die „Zielgröße“ LJB in Abhängigkeit von k und der Ausgangsverteilung.

Es ergibt sich auch hier der Eindruck, dass im Allgemeinen die Wahl von $k = 1$ die beste ist. Für die beiden Gleichverteilungen scheinen zwar leichte Vorteile für $k = 2$ zu bestehen, der Zugewinn an Nähe zur Normalverteilung ist aber eher gering. Insbesondere rechtfertigt er allein nicht die Zunahme der Komplexität des Vorgehens durch Einführung eines zusätzlichen Parameters. Es wird im Folgenden stets von $k = 1$ ausgegangen; der Schritt 3 („Hebeln“ der Mittelwerte) des auf Seite 8 beschriebenen Vorgehens entfällt damit.

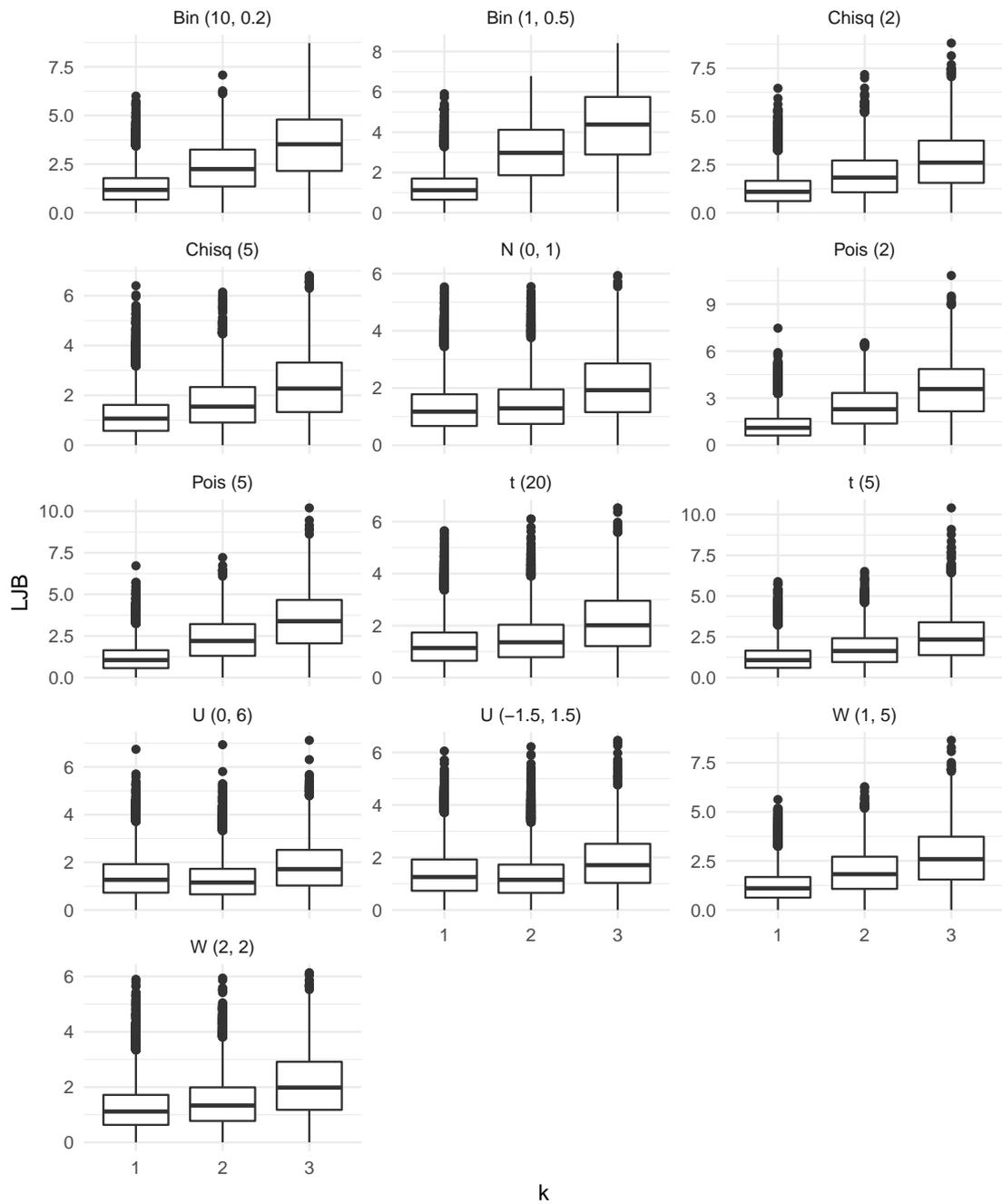


Abbildung 6: Für verschiedene w , s und m , und wie oben angegeben ausgewählte d , ergeben sich die dargestellten LJB -Maße, je nach Parameter k und Verteilung. Die Ergebnisse für K^* sind qualitativ gleich, numerisch fallen die Unterschiede etwas geringer aus.

3.2.2. Die Breite w und die Weite s

Vier weitere Parameter wurden bei der Simulation variiert. Dabei stellen die Verteilung der Ausgangsstichprobe sowie die Stichprobengröße keine Dimensionen dar, welche später in der Anwendung optimal eingestellt werden sollen, vielmehr nehmen diese Größen die Rolle von Kovariaten ein. Das Hauptinteresse der vorliegenden Ausarbeitung besteht hingegen in der Untersuchung der Fensterbreite w und der Schrittweite s . Es soll dabei geklärt werden, wie groß diese Parameter zu wählen sind, um das Ziel einer normalverteilten Größe ausreichend genau approximieren zu können. Dabei sollen möglichst kleine Werte für die Schrittweite verwendet werden, um die durch die gleitenden Mittel implizite Glättung der jeweiligen Zeitreihe möglichst gering zu halten.

Im ersten Schritt zur Bestimmung „guter“ Werte werden nun die Resultate der Standardnormalverteilung als Ausgangsgröße herangezogen - wie aus dem vorigen Teil abgeleitet für $d = 0.25$ und $k = 1$. Dass aus normalverteilten Daten durch die Transformationen weiterhin eine Normalverteilung erzeugt wird, stellt offenbar ein Minimal-Kriterium für ein gutes Verfahren dar. Die (Nachher)-Stichprobengröße wird auf den Wert 1 000 festgesetzt, um eventuelle Verzerrungen durch zu kleine Fallzahlen und Konvergenzprobleme auszuschließen. Die folgende Abbildung 7 zeigt die beiden Distanzmaße in Abhängigkeit von der Fensterbreite w und der Schrittweite s , welche relativ als Anteil von w angegeben wird. Für $w \leq 4$ „fehlen“ einzelne Boxplots, da hier nicht alle Schrittweiten $s = s_{\text{rel}}w$ ganzzahlig und eindeutig sind. Analoge Darstellungen finden sich beispielhaft für die Binomial (10, 0.2)-, die t_5 und die χ_5^2 -Verteilung im Anhang als Abbildungen 13, 14 und 15 auf den Seiten 34 bis 36.

Zunächst fallen die Boxplots für kleine s_{rel} auf, insbesondere bei großem w . Das weist darauf hin, dass die oben angestellten Überlegungen zur Unabhängigkeit auch ihren Grund haben: Werden die Überdeckungsraten aufeinander folgender Fenster beim gleitenden Mittelwert zu groß, kann das Ergebnis nicht mehr als aus einer Verteilung stammend angesehen werden. Dass sich dieser Effekt für $s_{\text{rel}} = 1/w$, das heißt für $s = 1$, gerade bei großen Fensterbreiten zeigt, erklärt sich durch den wachsenden relativen Anteil der Überdeckung: Für $w = 4$ unterscheiden sich $s = 1$ Beobachtungen aufeinander folgender Abschnitte, also 25 %. Für $w = 24$ oder $w = 108$ unterscheiden sich wiederum nur $s = 1$ Beobachtungen, was aber nur noch 4.2 beziehungsweise 0.9 % bedeutet. Dies wirkt sich zusätzlich auf die Schnittmengen dreier, vierer und weiter folgender Intervalle aus, die Autokorrelation der gleitenden Summen steigt dann erwartungsgemäß. Für $s_{\text{rel}} = 1/4$ besteht dieser Effekt nur noch leicht und wird vor allem von der logarithmierten JB-Statistik sensiert. Die K^* -Werte lassen hier auf kleinere Unterschiede schließen.

Werden die Werte mit Blick auf die Annäherung an eine (gedachte) Asymptote bei steigender Fensterbreite beurteilt und zudem die Hilfslinien der Quantile als Orientierungshilfe hinzugezogen, so scheinen die Unterschiede zwischen den Gruppen im Verlauf der Abszissenachse ab etwa $w = 12$ nur noch von geringer Größenordnung sein. Für die Betrachtung der anderen zwölf Verteilungen werden die kleinsten beiden Fensterbreiten eins und vier mit dieser Überlegung nicht weiter berücksichtigt, ebenso die Ergebnisse für die Schrittweite $s = 1$ ($\Leftrightarrow s_{\text{rel}} = 1/w$). Hervorzuheben ist das steile Absinken der Distanzmaße für kleine w , während das Gefälle mit steigendem w geringer wird. Die

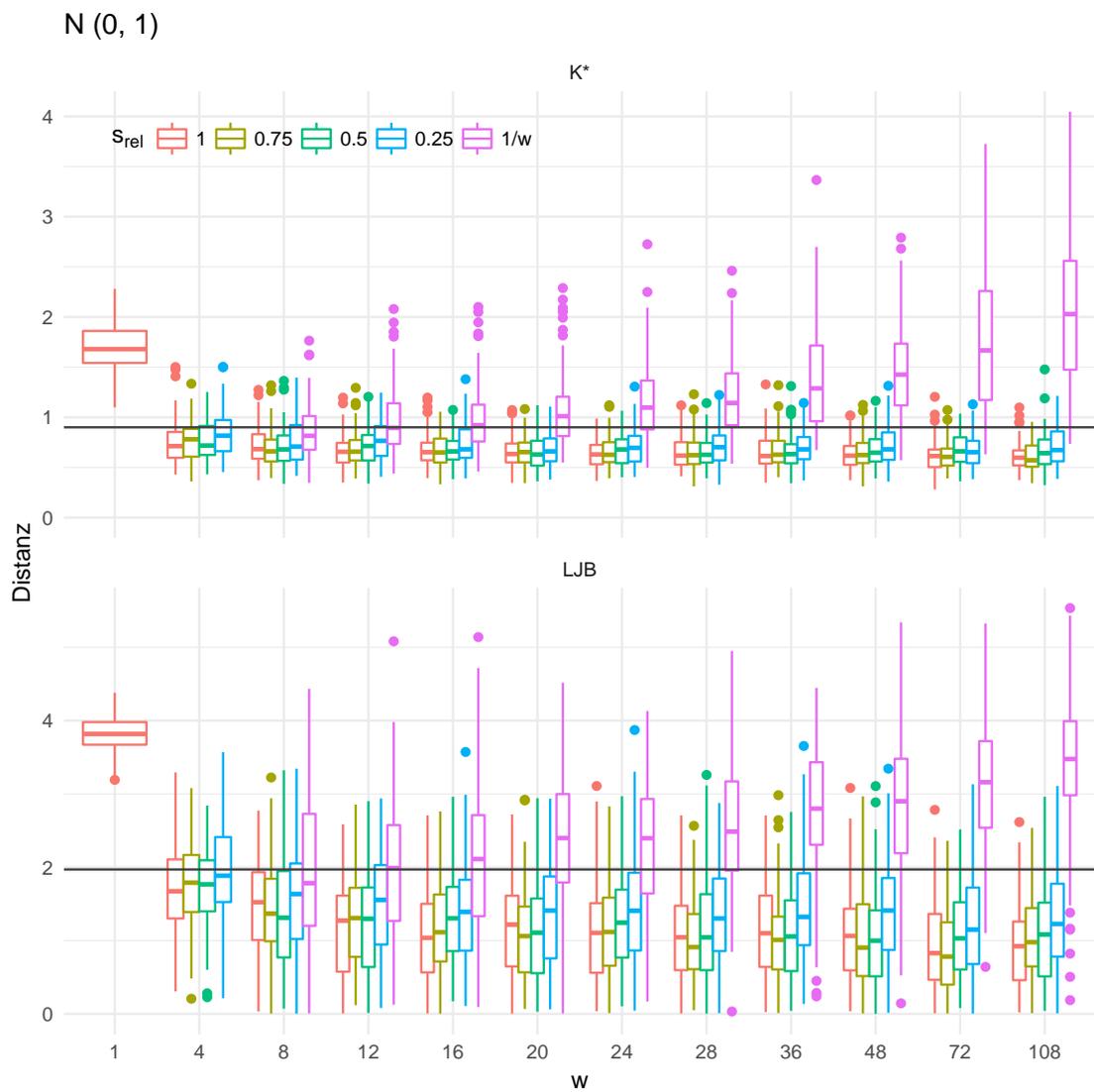


Abbildung 7: Die beiden Distanzmaße für a priori normalverteilte Daten, in Abhängigkeit von der Fensterbreite w und der Schrittweite s . Die Ergebnisse wurden auf solche für a posteriori-Stichproben der Größe $m = 1\,000$ reduziert. Die horizontalen Geraden bilden die 95 %-Quantile der Maße für normalverteilte Daten ab und sollen lediglich die Einordnung der numerischen Resultate vereinfachen. Siehe dazu auch den Ankerpunkt in Form von Abbildung 4.

Ergebnisse für $m = 100$ werden nicht gesondert gezeigt, führen aber insgesamt zu den gleichen Schlussfolgerungen.

Die Abbildung 8, sowie die Abbildung 16 im Anhang auf Seite 37, zeigt ähnlich wie die vorige Grafik zusammengefasste Verläufe der Distanzmaße in Abhängigkeit der Parameter. Die Aggregation der Resultate wird nun aber noch stärker als bei den Boxplots vorgenommen, dafür werden mehr Verteilungen aufgenommen. Durch die Änderung der Vorauswahl bezüglich w und s_{rel} resultieren andere Skalierungen als in der vorigen Abbildung 7, was bei Vergleichen zu beachten ist.

Die beiden Abbildungen umfassen verschiedene Verteilungen und verschiedene Stichprobengrößen, und verwenden zudem unterschiedliche Distanzmaße. Dennoch ergibt sich in beiden Fällen ein ähnliches Bild, wobei die kleinere Stichprobe zu stärker streuenden Ergebnissen führt.

Als erstes Teilergebnis ist zu notieren, dass eine Schrittweite von $w/4$ zu gering ist, um eine gute Annäherung an die Normalverteilung zu erreichen: Alle drei Kennwerte liegen auch für große Fensterbreiten noch recht deutlich oberhalb der „Soll“-Werte. Hier kommt die beschriebene, implizit erzeugte Abhängigkeit der Werte recht deutlich zum Tragen. Beim Vergleich der übrigen drei Abbildungen erscheint der Unterschied zwischen den Schrittweiten recht gering, so dass eine s_{rel} von 0.5 als gut verwendbar erscheint. Da kleine Schrittweiten angestrebt werden, wird diese Wahl getroffen.

Auch wenn - zumindest für $s = w$ - durch den Zentralen Grenzwertsatz klar ist, dass die Abstände zur Normalverteilung mit steigender Fensterbreite geringer werden, so sind diese Näherungen hier ab etwa $w = 30$ optisch nicht mehr zu erkennen: Die Unterschiede zu Werten für größere Breiten stammen offenbar eher aus zufälligen Streuungen der Simulation, als aus der Annäherung an die Verteilung. Für einzelne Verteilungen genügen auch kleinere Fensterbreiten, ohne dass deren Gemeinsamkeiten und Abgrenzungen von den anderen Verteilungen hier offensichtlich sind.

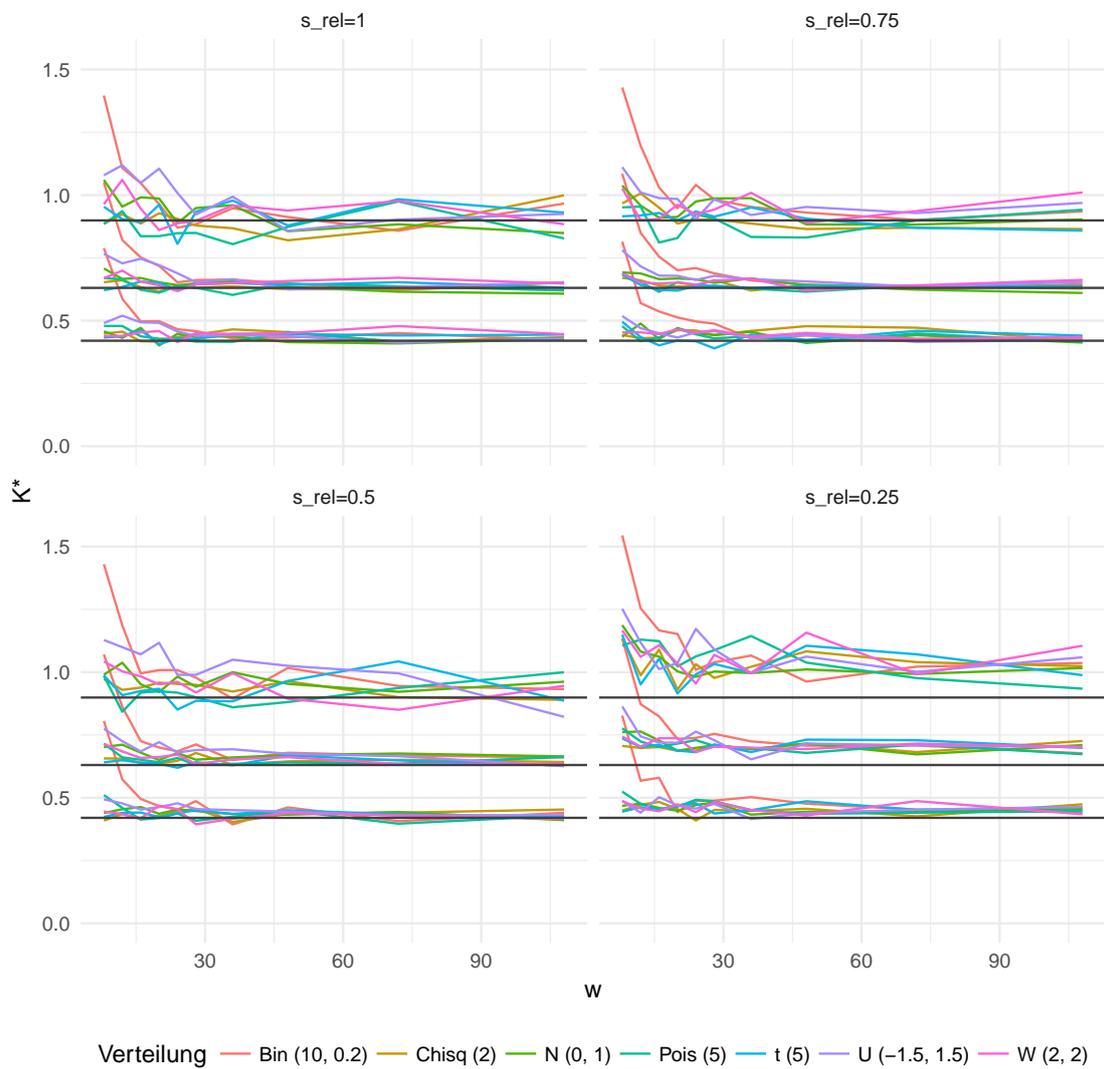


Abbildung 8: Je nach Einstellung der Schrittweite s_{rel} ergeben sich andere Zusammenhänge von der Fensterbreite w und dem Abstand zur Normalverteilung. Dargestellt sind hier die Quantile zum 5- und 95-Prozent-Niveau sowie das arithmetische Mittel aus jeweils 100 Wiederholungen. Es werden Ergebnisse für $m = 1\,000$ gezeigt. Die horizontalen Linien geben diese Kennzahlen als für eine Normalverteilung simulierte Werte an.

3.3. Folgerungen

Das Vorgehen ist den vorigen Ausführungen entsprechend auch für $d \neq 0$ und $s < w$ geeignet, die Daten „nahe“ an eine Normalverteilung zu transformieren. Insgesamt werden die folgenden Punkte aus der Simulationsstudie abgeleitet:

- Die Bildung von absoluten Differenzen verbessert die Annäherung an die Normalverteilung, wobei eine Schiefe-abhängige Potenz um $d = 0.25$ zu ergänzen ist.
- Die Verwendung einer Potenz k nach der Mittelwertbildung ist hingegen nicht förderlich.
- Als Fensterbreite kann $w = 30$ angesetzt werden.
- Als Schrittweite für nicht korrelierte Daten genügt $s = w/2$, mit dem vorigen Punkt also $s = 15$.

Die hier angegebenen Werte wurden bei der Abbildung 2 verwendet, welche auf der Seite 9 einen beispielhaften Durchlauf des Verfahrens zeigt.

4. Allgemeinere Szenarien

Für das im Vorangegangenen untersuchte Verfahren wurden durch die verschiedenen simulierten Zeitreihen indirekt auch Beispiele gegeben. Dabei erfüllten diese durchweg die Forderung nach unabhängigen und identisch verteilten Beobachtungen. Es werden nun weitere Situationen betrachtet, welche durch ihr Vorkommen in der Anwendung relevant sind. Neben einer Zusammenstellung der Methoden werden jeweils auch kurze Beispiele gegeben. Zunächst wird auch die Konvergenzgeschwindigkeit betrachtet, welche durch das Berry-Esseen-Theorem abgeschätzt werden kann. Abschließend wird ein Datenbeispiel gegeben. Die dazu verwendeten Werte sind hochgradig selbstabhängig, zudem ist die zugrunde liegende Verteilung als unbekannt einzustufen.

4.1. Konvergenzgeschwindigkeit

Die Ermittlung passender Einstellungen für den Parameter der Fensterbreite w beinhaltet implizit auch Aussagen über die Konvergenzgeschwindigkeit bei der vorgestellten Anwendung des zentralen Grenzwertsatzes. Neben simulativ hergeleiteten Werten können dabei auch theoretische Überlegungen zum Zuge kommen. Bei den vorigen Analysen wurden bewusst keine Annahmen über die Verteilung der X_i gemacht, bei den Simulationen daher verschiedene und beispielsweise stetige sowie diskrete Verteilungen einbezogen. Die Konvergenzgeschwindigkeit der transformierten Größe Y_j hängt jedoch direkt von Eigenschaften der X_i ab. Werden diese genauer spezifiziert, so ist es auch denkbar, optimale Werte für den Parameter d direkt aus der Schiefe der Beobachtungen abzuleiten, oder die Parameter w und s direkt in Abhängigkeit eines zulässigen maximalen Abstandes zur Normalverteilung zu bestimmen. Die Parameter könnten dabei unter Ausnutzen des Satzes von Andrew C. Berry und Carl-Gustav Esseen bestimmt werden, welcher sich in leicht

verschiedenen Varianten beispielsweise bei [Witting und Müller-Funk, 1995], Seite 123, oder bei [Patel und Read, 1996], Seite 160, findet: Sind die X_i , $1 \leq i \leq n$, unabhängig mit Erwartungswerten μ_i und Varianzen σ_i^2 verteilt, und ist $\beta_i := E(|X_i - \mu_i|^3) < \infty \forall i$, so gilt demnach für die Verteilungsfunktion F_Y von $Y = \sum_{i=1}^n (X_i - \mu_i) / \sqrt{\sum_i \sigma_i^2}$ und für alle $y \in \mathbb{R}$ und eine unbekannte Konstante $C > 0$:

$$|F_Y(y) - \Phi(x)| \leq K(F_Y, \Phi) \leq C \frac{\sum_{i=1}^n \beta_i}{(\sum_{i=1}^n \sigma_i^2)^{3/2}} =: C \rho_n \leq 0.4748 \rho_n =: b.$$

Sind die X_i zudem identisch verteilt, so vereinfacht sich der Term ρ_n , die Ungleichung ist dann

$$|F_Y(y) - \Phi(x)| \leq C \frac{E(|X_1 - \mu_1|^3)}{\sqrt{n} \sigma^3}.$$

Die Abschätzung der Konstanten C ist seit den 1940er-Jahren immer wieder Gegenstand der Forschung gewesen, der hier verwendete Wert ist [Shevtsova, 2011] entnommen. Auf der anderen Seite kann $C \geq (\sqrt{10}+3)/(6\sqrt{2\pi}) \approx 0.4097$ gezeigt werden, wie [Patel und Read, 1996] notieren.

Die angegebene Ungleichung stellt zwar eine prominente Möglichkeit zur Abschätzung der Konvergenzgeschwindigkeit dar, ist aber nicht die einzige: So können beispielsweise auch die Entwicklungen („expansions“) nach Edgeworth oder Gram und Charlier verwendet werden, um den Abstand von F_Y und Φ abzuschätzen (ebenda, Teil 6.6). Für Variablen mit Abhängigkeiten in bestimmten Formen wie Mischungen oder m-Abhängigkeiten (s. u.) stellt [Tikhomirov, 1981] Ansätze zur Abschätzung vor.

Zur genaueren Betrachtung wird an dieser Stelle die Ungleichung von Berry und Esseen herangezogen. Dabei werden Resultate für eine Verletzung der Annahme unabhängiger Beobachtungen berücksichtigt, welche sich bei „zu kleinen“ Schrittweiten s ergibt. Dazu wird beispielhaft eine Zeitreihe bestehend aus 100 000 unabhängig identisch χ_2^2 -verteilten Beobachtungen gezogen. Für verschiedene Fensterbreiten wird nun der Verlauf der Verteilungsfunktion der y_i betrachtet, wobei diese Werte zur besseren Vergleichbarkeit zentriert und skaliert werden, also einen Mittelwert von 0 und eine Varianz von 1 aufweisen. Es wird hier jeweils eine Schrittweite von $s = 1$ verwendet, das im Sinne der Forderung nach Unabhängigkeit der größten Annahmeverletzung gleichkommt.

Dennoch umfassen die durch $\Phi(x) \pm b$ gegebenen Grenzen die empirische Verteilungsfunktion F_Y komplett, wie an der Abbildung 9 zu erkennen ist. Dabei ist zu berücksichtigen, dass der benötigte Wert von ρ_n hier aus den simulierten Daten geschätzt wird. Weiter ist nicht klar, ob dieser Sachverhalt bestehen bleibt, falls eine genauere Abschätzung der Konstanten C gefunden wird.

Aus der Betrachtung dieser einzelnen Fälle kann offenbar keine Systematik abgeleitet werden, sondern nur ein beispielhafter erster Eindruck gegeben werden. Ein Punkt wird aber deutlich: Die Abstände zwischen oberer und unterer Grenze, also gerade die obere Schranke für $K(F_Y, \Phi)$, sind hier erst für große Schrittweiten von eher kleinem Maßstab. Kleinere Abstände ergeben sich, wie an der Ungleichung ob zu erkennen ist, für eine größere Varianz oder einen kleineren Erwartungswert von $|X_1 - \mu_1|^3$. Beide Größen

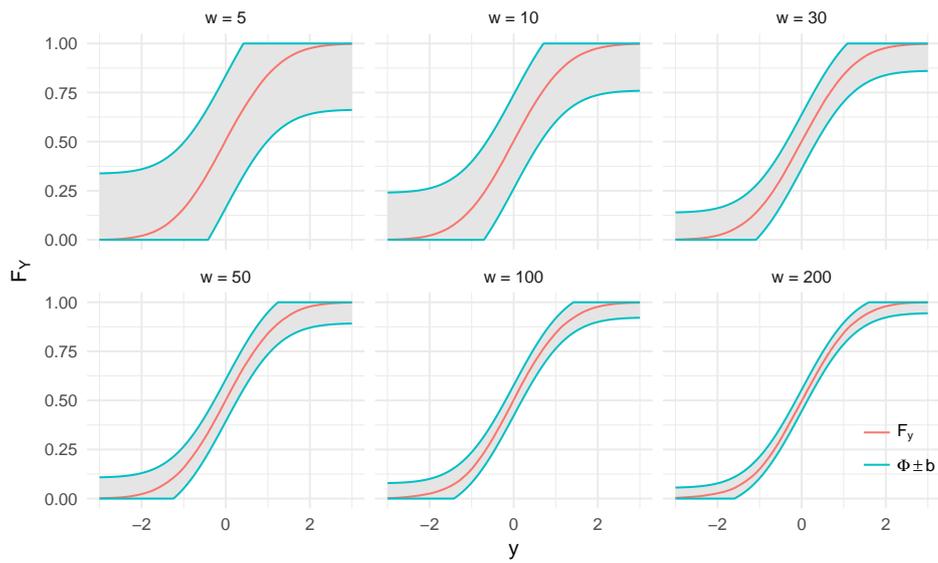


Abbildung 9: Für simulierte Daten mit $X_1 \sim \chi_2^2$, $d = 0.25$, $s = 1$ und verschiedene Fensterbreiten sind die empirischen Verteilungsfunktionen, sowie die geschätzten Berry-Esseen-Grenzen dargestellt. Die Grenzen sind hier auf Bereiche zwischen 0 und 1 begrenzt worden. Trotz großer Abhängigkeiten der y_i werden die Grenzen von den empirischen Verteilungsfunktionen eingehalten.

können über den Parameter d direkt beeinflusst werden. Im gegebenen Beispiel wäre der Abstand der Grenzen beispielsweise für $d = 0$, also beim Auslassen dieses Schritts der Vorverarbeitung, größer als in der Abbildung ($\hat{\rho}_n = 2.44/\sqrt{n}$ statt $1.54/\sqrt{n}$).

4.2. Unabhängig, nicht identisch verteilt

Die identische Verteilung der in den gleitenden Mittelwert eingehenden Größen ist eine Annahme, welche die Anwendung der Methodik stark eingrenzt. Als praktisches Beispiel sei die wiederholte Prüfung eines Bauteils genannt, wobei verschiedenen Messinstrumente zwar alle einen erwarteten Messfehler von Null, jedoch verschiedene Varianzen aufweisen.

Wie [Klenke, 2006] auf Seite 305 erläutert, ist für die Geltung des ZGWS bei der Summe zentrierter Zufallsvariablen „entscheidend [..], dass die Varianz der normierten Summe 1 ist, und dass jeder einzelne [der] Summanden nur einen kleinen Beitrag liefert“. Bei [Bauer, 2001] findet sich gleichermaßen die Forderung, dass „keine der Variablen [..] einen beherrschenden Einfluss auf die Verteilung der Summe [..] ausübt“ (Seite 238).

Um für unabhängige und quadratisch integrierbare $X_i \in \mathbb{R}$ die Verteilungskonvergenz der Summe

$$S_n := \frac{\sum_{i=1}^n (X_i - E(X_i))}{\sqrt{\sum_{i=1}^n \sigma_i^2}} =: \frac{\sum_{i=1}^n (X_i - E(X_i))}{s_n}$$

gegen die Standardnormalverteilung sicherzustellen, genügt die Erfüllung einer der Bedingungen von J. Lindeberg und A. Lyapunov. Die erstgenannte besteht mit I als Indikatorfunktion in der Gleichung

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n E \left((X_i - E(X_i))^2 I_{|X_i - E(X_i)| \geq \varepsilon s_n} \right) = 0 \quad \forall \varepsilon > 0,$$

die zweitgenannte fordert die Existenz eines $\delta > 0$, so dass

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E \left(|X_i - E(X_i)|^{2+\delta} \right) = 0$$

gilt. Dabei liegen an dieser Stelle die Formulierungen von [Bauer, 2001] zugrunde, siehe Seite 238 f. Dort ist auch vermerkt, dass die Lyapunov-Bedingung die Lindeberg-Bedingung impliziert.

Für identisch verteilte Größen ist $1/s_n^2 = 1/n\sigma^2 \rightarrow 0$ ($n \rightarrow \infty$) und

$$\sum_{i=1}^n E \left((X_i - E(X_i))^2 I_{|X_i - E(X_i)| \geq \varepsilon s_n} \right) = n \cdot E \left((X_1 - E(X_1))^2 I_{|X_1 - E(X_1)| \geq \varepsilon s_n} \right).$$

Dabei nimmt die Indikatorfunktion für alle $\varepsilon > 0$ wegen $s_n \rightarrow \infty$ für ausreichend große n den Wert Null an, daraus folgt die Erfüllung der Lindeberg-Bedingung.

Für ein zweites Beispiel seien X_i , $1 \leq i \leq n$, unabhängige reelle Zufallsvariablen so, dass ihr Erwartungswert $E(X_i) = 0$ und ihre Varianz $V(X_i) = 2^{(i \bmod 2)}$, also abwechselnd 2

und 1, betragen. Zudem sie die Wölbung endlich und, der Einfachheit halber, n gerade. Dann ist beispielsweise für $\delta = 2$ die Lyapunov-Bedingung erfüllt, wobei der Wert des Terms nur von n und der Wölbung der X_i abhängt:

$$\begin{aligned} & \frac{1}{s_n^{2+2}} \sum_{i=1}^n E(|X_i - E(X_i)|^{2+2}) \\ &= \frac{1}{(n/2 \cdot 1 + n/2 \cdot 2)^2} \left(\frac{n}{2} E((X_1 - E(X_1))^4) + \frac{n}{2} E((X_2 - E(X_2))^4) \right) \\ &= \frac{4}{9n^2} \frac{n}{2} (E(X_1^4) + E(X_2^4)) \\ &= \frac{2}{9n} (E(X_1^4) + E(X_2^4)) \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Die oben vorgestellte Normalisierung von Zeitreihen lässt sich also auch auf Fälle übertragen, in denen nicht alle Elemente einer Zeitreihe aus der gleichen Verteilung stammen. Die Vorschläge zur Einstellung von Fensterbreite und weiteren Parametern müssen dabei jedoch nicht ohne Weiteres eine gute Wahl oder gar optimal sein. Auch hier sei auf die Invarianz der Normalverteilung gegenüber linearen Transformationen verwiesen: Wenn die Verteilung S_n der gegen eine Normalverteilung konvergiert, dann gilt das auch für die Verteilung der Summe der X_i . Die Parameter der Normalverteilung müssen bei der Anwendung nicht bekannt sein damit der ZGWS gilt, und können auch nach der Summenbildung geschätzt werden.

Simulativ der folgende Fall betrachtet: Die Elemente eines Prozesses $\{X_i\}$ stammen abwechselnd aus zwei Verteilungen, so dass:

$$X_i = \begin{cases} Y_i - 1, & \text{mit } Y_i \sim \chi_1^2, & \text{falls } i \text{ ungerade} \\ Z_i, & \text{mit } Z_i \sim N(0, 1), & \text{falls } i \text{ gerade} \end{cases}$$

Damit ist gerade die Situation des vorigen Beispiels erfüllt. Für verschiedene Fensterbreiten w sowie für Schrittweiten $s = w$ werden Zufallsvektoren aus den obigen Verteilungen gezogen. Dabei wird entgegen den obigen Resultaten $d = 0$ gesetzt, also keine „Symmetrisierung“ vorgenommen, um den ZGWS hier für bekannte Ausgangs-Verteilungen zu demonstrieren. Anhand der logarithmierten Jarque-Bera-Statistik wird dann die Nähe zur Normalverteilung beurteilt. Die Abbildung 10 zeigt das Ergebnis: Sowohl für die X_i , als auch für die Y_i und die Z_i als Ausgangswerte ergibt sich eine Konvergenz gegen die Normalverteilung. Während diese für die normalverteilten Z_i nicht in *eigentlicher* Konvergenz, sondern in Gleichheit besteht, liegt der *LJB*-Wert für die Mix-Verteilung sogar unter dem der identisch verteilten verschobenen Y_i .

4.3. Nicht unabhängig, nicht identisch verteilt

Neben der (Nicht-)Identität wurde im Vorigen bisher stets ein weiterer Sachverhalt vorausgesetzt, welcher als allgemeine Voraussetzung in der Praxis an Grenzen stoßen kann:

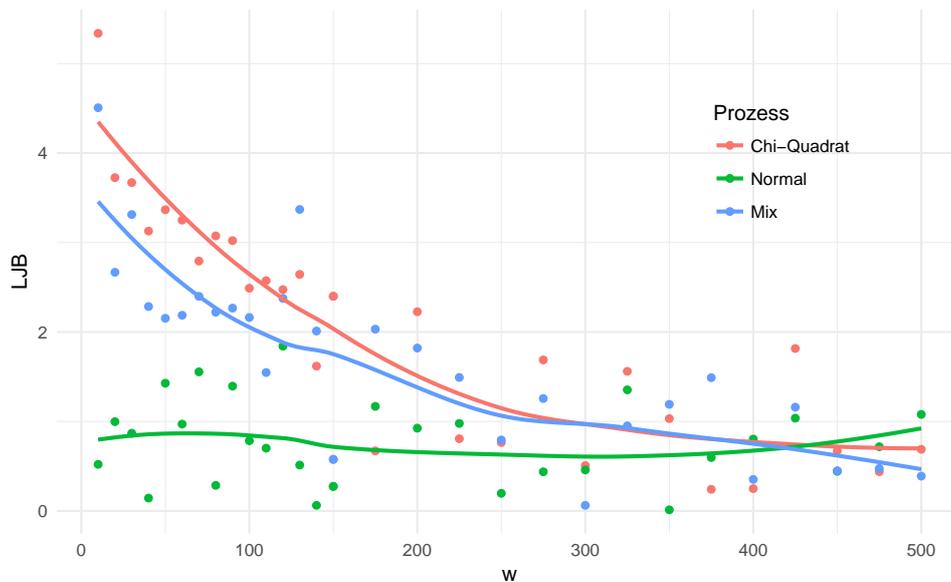


Abbildung 10: Wird aus Prozessen der zwei genannten Verteilungen der Wert S_n gebildet, und anschließend der Abstand zur Normalverteilung durch LJB „gemessen“, zeigt sich die durch den ZGWS beschriebene Konvergenz. Dass diese auch für nicht identisch verteilte Größen bestehen kann, zeigt diese beispielhafte Mix-Verteilung. Die Linien stellen polynomielle Regressionen der Ergebnisse dar.

Es wurde stets davon ausgegangen, dass die Beobachtungen x_i stochastisch unabhängig sind. Gerade bei Zeitreihendaten muss diese Annahme nicht immer realistisch sein, je nach Abtastrate kann für einen stetigen Prozess theoretisch sogar jede beliebige Autokorrelation erreicht werden (vgl. das $\varepsilon - \delta$ -Kriterium der Stetigkeit). Durch die Betrachtung von Differenzen wird durch das oben beschriebene Vorgehen gegebenenfalls und je nach Struktur ein großer Teil dieser Abhängigkeiten behandelt. Dennoch soll ein kurzer Blick auf das Verhalten des Verfahrens bei Selbstabhängigkeiten einer Zeitreihe gegeben werden.

Bei [Angus, 1993] finden sich Herangehensweisen für bestimmte ARIMA-Prozesse (autoregressive integrated moving average). Dabei wird jedoch auch deutlich, dass sich diese Abweichung von den Annahmen zwar auf die Gestalt von Erwartungswert und Varianz des Limes-Verteilung, nicht jedoch auf die Konvergenz gegen eine Normalverteilung an sich auswirkt (vgl. bspw. ebenda, Korollar 1).

Eine Übersicht allgemeinerer Resultate für den ZGWS bei abhängigen Variablen findet sich etwa bei [Serfling, 1968]. Das dort unter der Nummer 1.1 vorgestellte Theorem von Hoeffding und Robbins stellt eine grundlegende Erkenntnis dar und wird hier in der Formulierung nach [Hoeffding und Robbins, 1948] wiedergegeben (vgl. Theorem 2). Dazu wird der Begriff der m -Unabhängigkeit benötigt: Zwei Abschnitte einer Reihe von Zufallsvariablen, $\{X_1, X_2, \dots, X_r\}$ und $\{X_s, X_{s+1}, \dots, X_n\}$, sind m -unabhängig, falls sie unabhängig für $s - r > m$ sind. Für $m = 0$ muss die Unabhängigkeit auch für $r = 1$ sowie $s = n = 2$ erfüllt sein, so dass die gewöhnliche Unabhängigkeit der Zufallsvariablen folgt. Eine Reihe, für welche die gemeinsame Verteilung eines Abschnitts $X_i, X_{i+1}, \dots, X_{i+r}$ nicht von i abhängt, wird stationär genannt.

Ist nun die Folge X_1, X_2, \dots stationär sowie m -unabhängig, und ist X_1 zentriert (d. h. $E(X_1) = 0$) sowie $E|X_1|^3 < \infty$, so gilt

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \xrightarrow{D} Z \sim N(0, A).$$

Dabei ist

$$A = E(X_1^2) + 2E(X_1X_2) + 2E(X_1X_3) + \dots + 2E(X_1X_{m+1}),$$

für identisch verteilte X_i also $A = E(X_1^2) + 2mE(X_1X_2)$. Auch hier ist die Form der Varianz für die Anwendung nicht unbedingt von Interesse, wichtig ist vor allem die Existenz eines Wertes, der dann aus den Daten geschätzt werden kann. Insbesondere ist außerdem zu bemerken, dass der Term, welcher konvergiert, im Kern gegenüber der Darstellung des ZGWS oben auf Seite 4 nicht verändert ist. Dort wurde $\sqrt{n}(\bar{X} - \mu)$ betrachtet, was aber gerade $1/\sqrt{n} \sum_{i=1}^n X_i - \sqrt{n}\mu$ entspricht. Der Unterschied besteht also nur im Subtrahenden, welcher sich bei identisch verteilten Variablen („klassischer“ ZGWS) und gleichzeitig $E(X_1) = 0$ (Hoeffding) zu 0 berechnet.

In der Praxis werden viele autokorrelierte Zeitreihen als m -unabhängige Prozesse aufgefasst werden können, so dass dann das oben beschriebene Verfahren auch bei der Datensituation des Hoeffding-Robbins-Theorems angewandt werden kann. Eine „gute“ Wahl

von w und insbesondere von s wird dann wiederum auch direkt von der Größe m abhängen.

4.4. Datenbeispiel

Ein Beispiel zu autokorrelierten Reihen wird anhand von Messdaten der Stromnetzfrequenz gegeben. Dazu wird eine Zeitreihe von Mittelwerten bestehend aus Messwerten zu fünf Orten im kontinental-europäischen Verbundnetz verwendet. Diese Reihe ist mit einem Autokorrelationskoeffizienten von 0.9998 hochgradig selbstabhängig. Das begründet sich insbesondere durch die hohe Abtastrate: Bei 50 Messungen pro Sekunde kann sich der Betriebszustand eines Generators zwischen zwei Datenpunkten nicht sehr stark ändern. Die Abbildung 11 zeigt einen beispielhaften Verlauf der Daten. Darin sind auch die Transformationen \tilde{x}_i (für $d = 0.25$) und y_i (für $w = 30$ und $s = 15$) bereits enthalten.

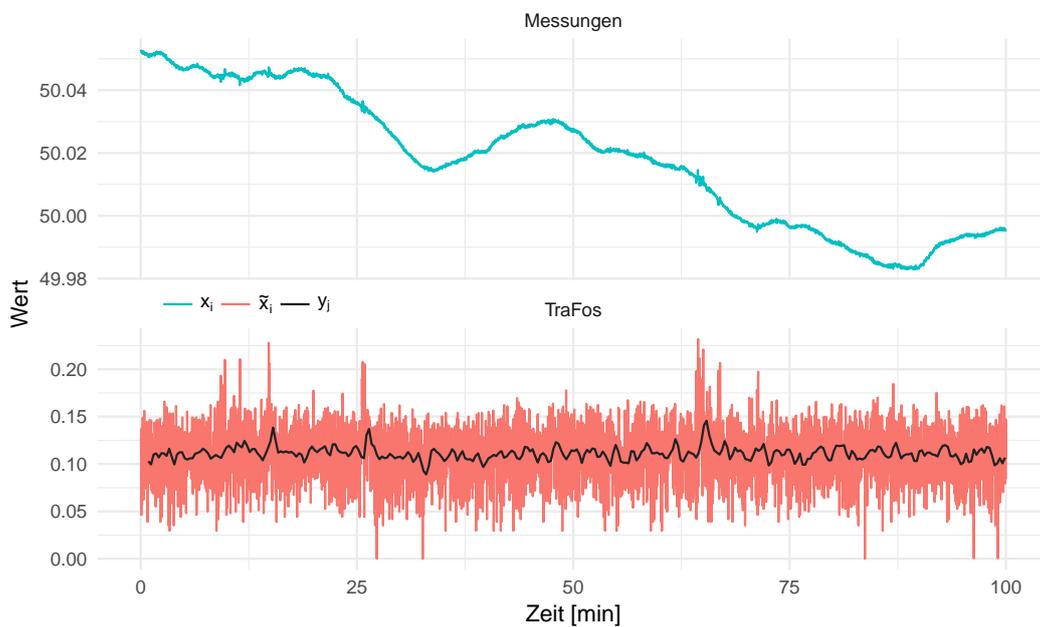


Abbildung 11: Ein kurzer Ausschnitt der als Beispiel verwendeten Messungen der Netzfrequenz und der abgeleiteten Transformationen. Gezeigt werden nur 100 Minuten, insgesamt liegen Werte für mehrere Tage vor.

Das Ziel der Transformationen besteht in der Verschiebung der Verteilung der Daten hin zu einer Normalverteilung. Hier kommt hinzu, dass der erste Schritt „Differenzbildung“ die Autokorrelation deutlich senken kann - numerisch von 0.9998 auf unter 0.12, wenn die Variante potenziert absoluter Differenzen wie oben verwendet wird.

Optisch sind die Ausgangswerte hier zwar symmetrisch, nicht aber normal verteilt, wie das Histogramm oben links in der Abbildung 12 erkennen lässt. Werden die beiden Schritte zur Normalisierung angewandt, so ergeben sich die beiden oben rechts und un-

ten links gezeigten Histogramme. Auch hier lassen sich die Verteilungen allenfalls mit Einschränkungen durch die jeweiligen Normalverteilungsdichten annähern.

Dieser Sachverhalt widerspricht den obigen Ergebnissen nicht, da die zentrale Forderung nach Unabhängigkeit der Beobachtungen nicht erfüllt ist: Die Wahl der empfohlenen Parameter beruht auf dieser Annahme. Um die Unabhängigkeit hier zu erreichen, wurden die Daten nach der Differenzen- und vor der Mittelwertbildung in zufälliger Reihenfolge neu angeordnet. Damit bleiben die Verteilungs-Eigenschaften gleich, lediglich die Autokorrelation verschwindet. Dann führt das Verfahren auch hier zu einer guten Approximation der Normalverteilung, wie auch die Abbildung 12 unten rechts mit der Kennung y_j^* zeigt.

Zusammen mit dem vorigen Beispiel ergibt sich also die Erkenntnis, dass das beschriebene Vorgehen für die gewählten Parametereinstellungen nicht tolerant gegenüber starken Autokorrelationen der Daten ist. Entsprechend wird dann eine vorausgehende Behandlung erforderlich. Diese kann aus verschiedensten Methoden bestehen, etwa aus einer Modellierung der Daten und der Betrachtung der Modellfehler, einer zusätzlichen Differenzenbildung schon vor der hier vorgenommenen, oder auch dem Ausdünnen der Daten vor der Mittelwertbildung. Die beste Wahl ergibt sich dabei durch die spezifischen Charakteristiken der jeweiligen Datensituation. Möglich ist es offenbar auch, die Parameter w , s und d passend für die jeweils vorliegenden Daten zu optimieren - was offenbar auch im Allgemeinen sowie auch ohne Autokorrelationen gewinnbringend sein kann, aber andererseits den Aufwand gegenüber einer Verwendung von Standard-Werten deutlich erhöht.

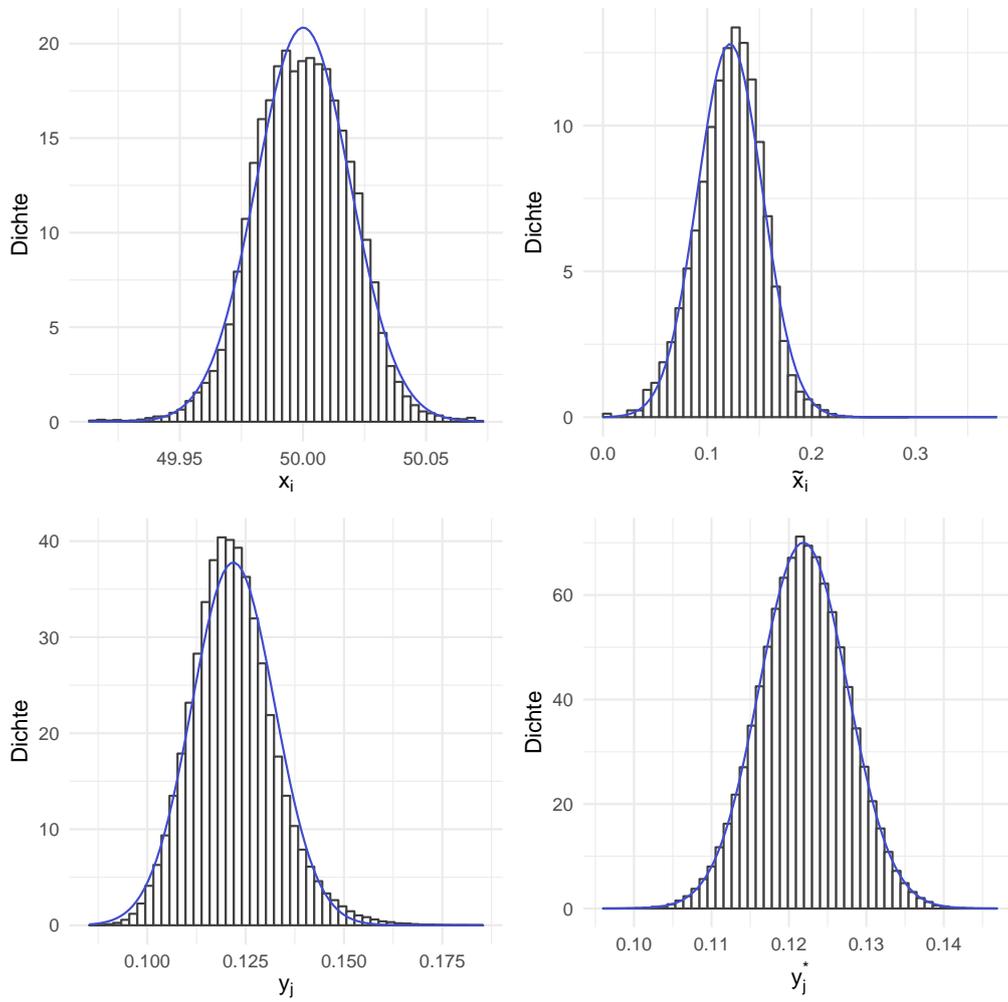


Abbildung 12: Die Frequenzdaten eines Tages, das heißt 4 320 000 Datenpunkte, sind oben links in ihrer Roh-Form dargestellt. Die anderen drei Histogramme zeigen die im Text angeführten Transformationen, die Linien geben die Dichte der Normalverteilung mit arithmetischem Mittel und Stichprobenvarianz als Parameterschätzern an.

5. Fazit

Die vorliegende Ausarbeitung stellte die Konvergenzeigenschaften der Verteilung eines gleitenden Mittelwerts in den Vordergrund. Es wurde eine Methodik vorgeschlagen, welche gemäß der Beschreibung auf Seite 8 einem Dreischritt folgt: Zunächst findet eine Vorbehandlung der Daten statt, welche aus potenzierten absoluten Differenzen, aber auch aus der Identitätsfunktion bestehen kann. Anschließend werden gleitende Mittelwerte gebildet und durch das Prinzip des Zentralen Grenzwertsatzes somit eine Normalverteilung erzeugt. Dabei sind die Wahl von Fensterbreite und Schrittweite zu beachten. Durch anschließende Potenzierung könnte sich die Verteilung der Mittelwerte der Normalverteilungsfamilie schneller annähern.

Die optimale Wahl der Parameter wurde durch eine Simulationsstudie untersucht. Optimiert wurden dabei Varianten der Kolmogorov-Metrik und der Teststatistik von Jarque und Bera. Dabei konnten, für verschiedene Ausgangsverteilungen und Stichprobengrößen, allgemeine Empfehlungen für die Potenz der absoluten Differenzen ($d = 0.25$ oder so, dass \tilde{X}_i möglichst symmetrisch verteilt ist), für die Fensterbreite ($w = 30$), für die Schrittweite ($s = 15$) und für die Nachher-Potenz ($k = 1$, entspricht dem Auslassen des dritten Schrittes) abgeleitet werden.

Zwei Forderungen für die Gültigkeit des *klassischen* ZGWS sind die Unabhängigkeit und die identische Verteilung der verwendeten Datenpunkte. Wie im Kapitel 4 mit einem Datenbeispiel aufgezeigt, kann das Vorgehen mit den genannten Parametrierungen keine starken Autokorrelationen verarbeiten, so dass Anpassungen nötig sind. Diese werden direkt von Eigenschaften der jeweiligen Datenstrukturen abhängen und bilden einen Ansatz für weitergehende Analysen.

Grundsätzlich gilt der Zentrale Grenzwertsatz aber auch für bestimmte Formen abhängig und/oder nicht identisch verteilter Zufallsvariablen, so dass das Verfahren auch dann angewandt werden kann. Insbesondere sind hier die Bedingungen von Lindeberg und Lyapunov zu nennen, welche den ZGWS erweitern. Außerdem gilt dieser auch bei m -Unabhängigkeit der Daten. Mit dem Satz von Berry-Esseen können Abschätzungen für den Abstand der Verteilung einer Summe zur Normalverteilung vorgenommen werden. Diese fallen, wie gezeigt, mitunter aber recht groß aus.

A. Ergänzende Grafiken

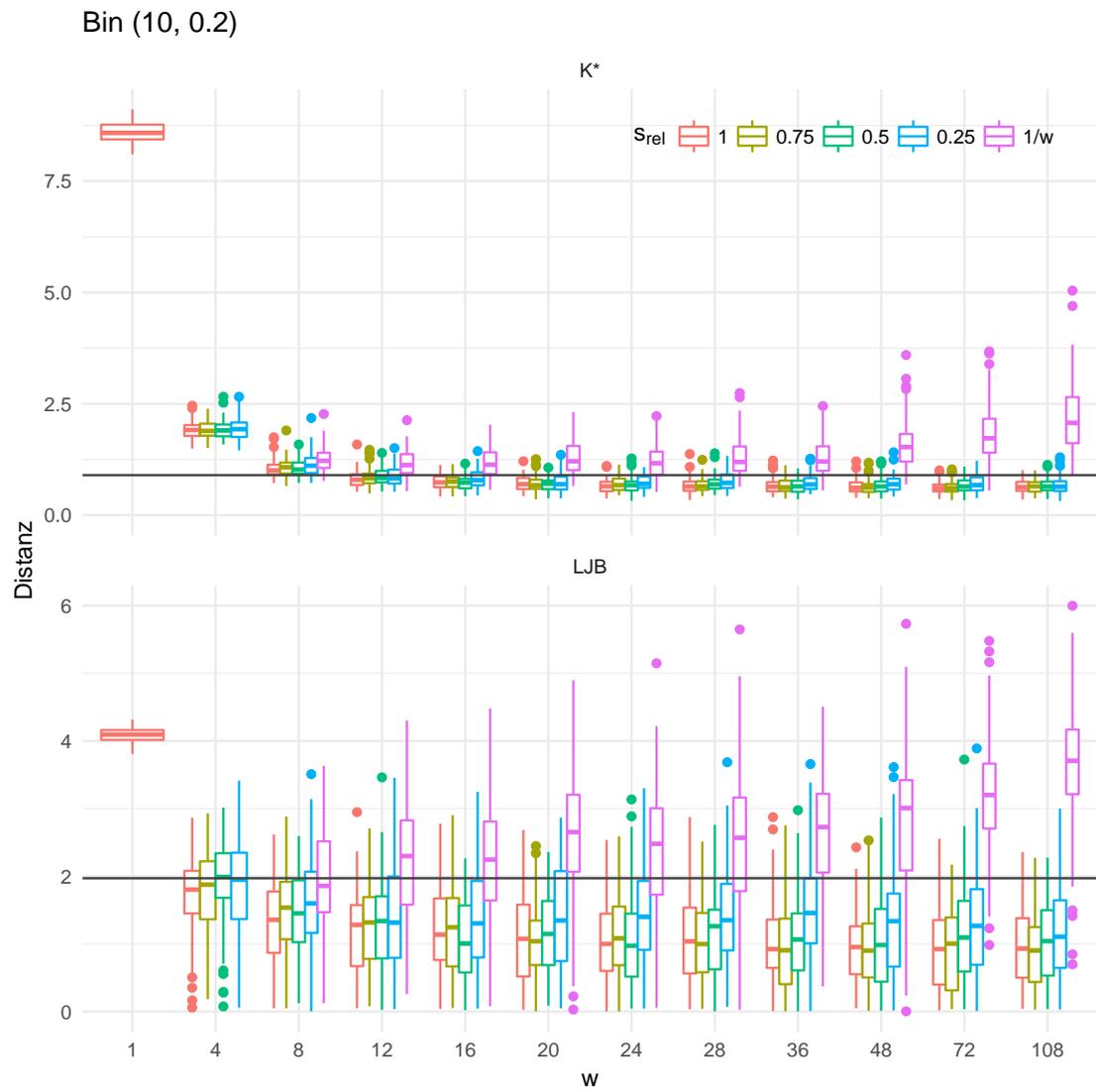


Abbildung 13: Ergänzend zu Abbildung 7: Die beiden Distanzmaße für a priori Bin (10, 0.2)-verteilte Daten und $m = 1\,000$.

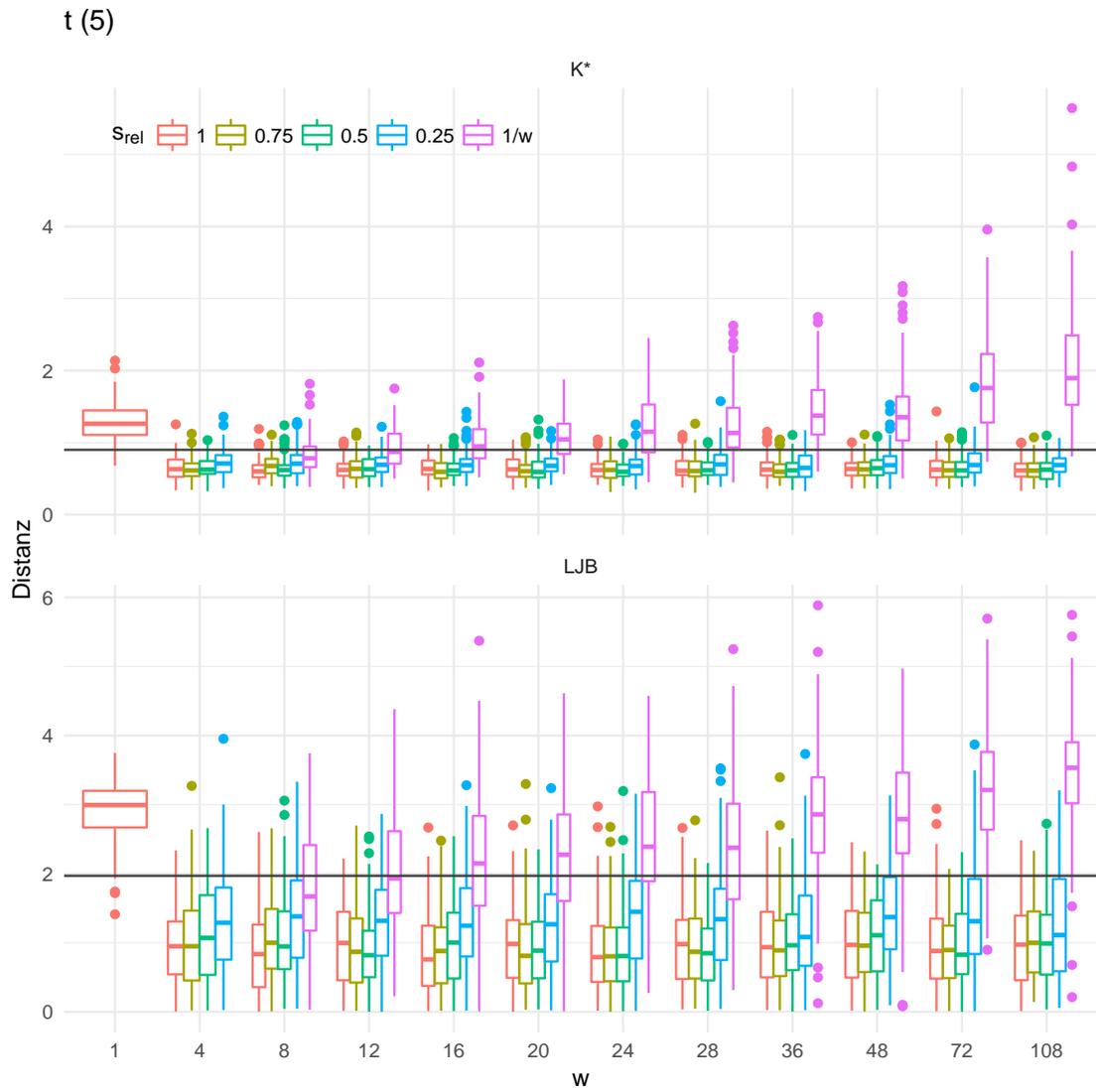


Abbildung 14: Ergänzend zu Abbildung 7: Die beiden Distanzmaße für a priori t_5 -verteilte Daten und $m = 1\,000$.

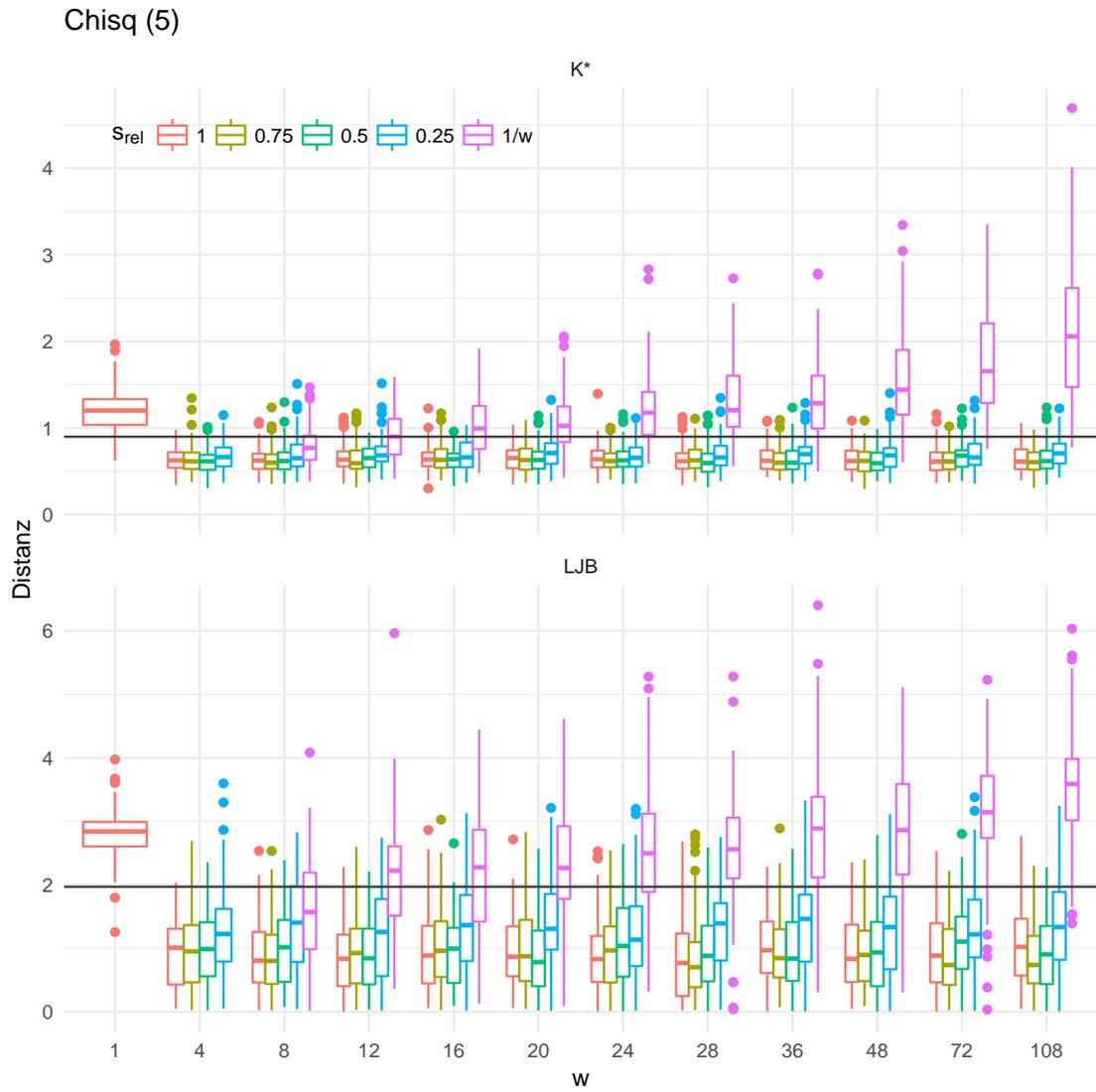


Abbildung 15: Ergänzend zu Abbildung 7: Die beiden Distanzmaße für a priori χ^2_5 -verteilte Daten und $m = 1\,000$.

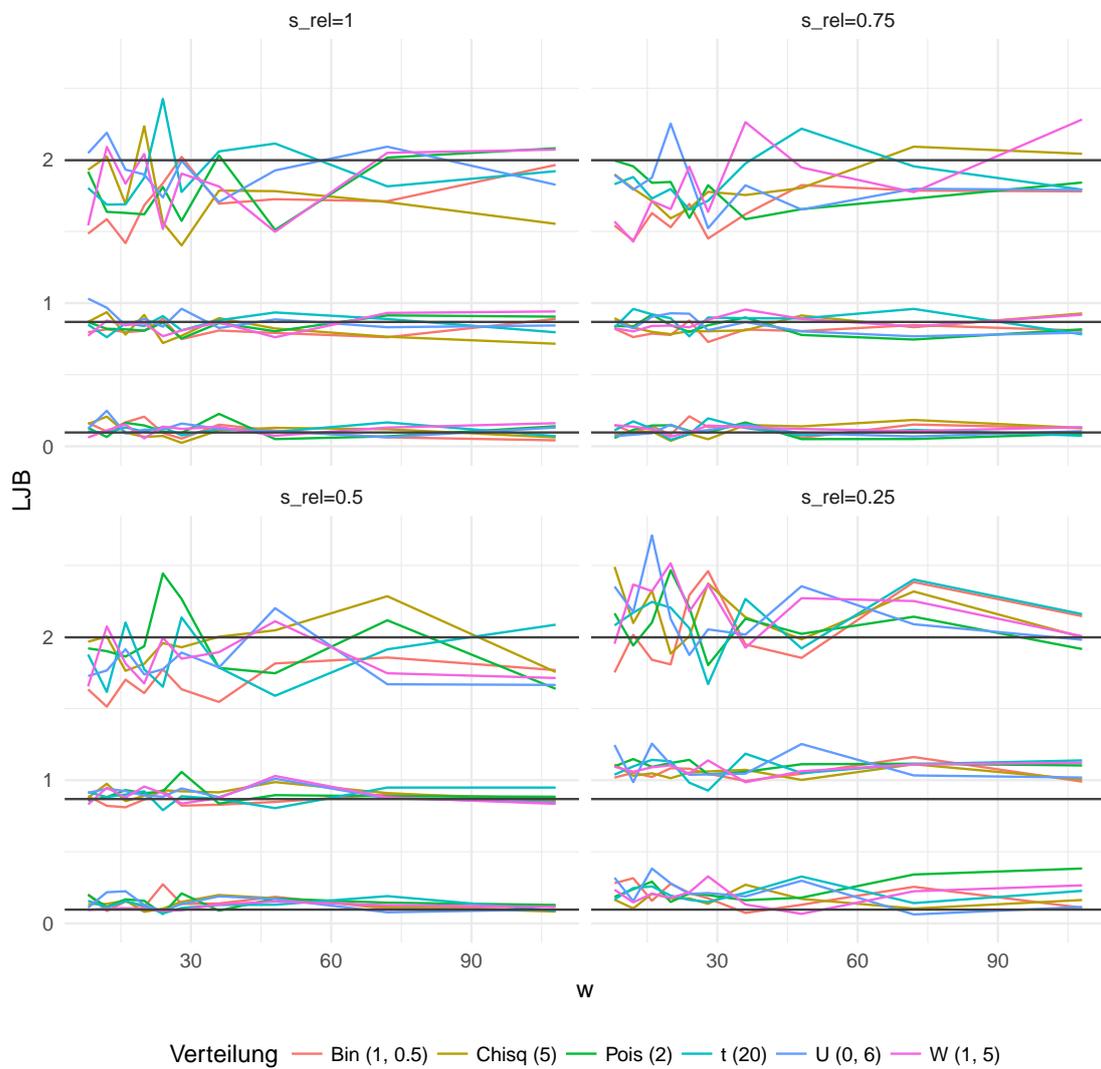


Abbildung 16: Ergänzend zu Abbildung 8 zeigt diese Grafik Verläufe von Quantilen und Mittelwert. Dabei werden hier die übrigen Verteilungen gezeigt, eine Stichprobengröße von $m = 100$ angesetzt und die LJB -Statistik als Distanzmaß verwendet.

Literatur

- [Angus, 1993] John E. Angus, 1993: *A Central limit theorem for autoregressive integrated moving average processes*. Mathematical and Computer Modelling, Jahrgang 17, Nummer 10, Seite 3-9. Elsevier Ltd.
- [Bauer, 2001] Heinz Bauer, 2001: *Wahrscheinlichkeitstheorie*. Walter de Gruyter Verlag, Berlin. Fünfte Auflage.
- [Hoeffding und Robbins, 1948] Wassily Hoeffding und Herbert Robbins, 1948: *The central limit theorem for dependent random variables*. Duke Mathematical Journal, Jahrgang 15, Nummer 3, Seite 773-780. Duke University Press. Erneut veröffentlicht von N. Fisher und P. Sen (beide Hrsg.), 1994: *The Collected Works of Wassily Hoeffding*. Springer-Verlag, New York.
- [Judge et al., 1985] George G. Judge, W. E. Griffiths, R. Carter Hill, Helmut Lütkepohl und Tsoung-Chao Lee, 1985: *The Theory and practice of econometrics*. John Wiley & Sons, New York. Zweite Auflage.
- [Klenke, 2006] Achim Klenke, 2006: *Wahrscheinlichkeitstheorie*. Springer-Verlag, Heidelberg.
- [Langesberg et al., 2018] Christian Langesberg, Uwe Ligges und Claus Weihs, 2018: *Zur Quantifizierung des Normalverteilungsgrades*. Eldorado, Sammlung der DFG-Forschergruppe: Schutz- und Leittechnik zur zuverlässigen und sicheren Energieübertragung, Online: <http://hdl.handle.net/2003/36856> (letzter Abruf am 2. Mai 2018).
- [Patel und Read, 1996] Jagdish K. Patel und Campbell B. Read, 1996 *Handbook of the normal distribution*. Marcel Dekker, New York. Zweite Auflage.
- [R Core Team, 2017] R Core Team, 2017: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien.
- [Rüger, 2002] Bernhard Rüger, 2002: *Test- und Schätztheorie, Band II: Statistische Tests*. Oldenbourg Wissenschaftsverlag, München.

- [Serfling, 1968] R. J. Serfling, 1968: *Contributions to central limit theory for dependent variables*. The Annals of Mathematical Statistics, Jg. 39, Nr. 4, S. 1158-1175. Institute of Mathematical Statistics.
- [Shevtsova, 2011] Irina Shevtsova, 2011: *On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands*. arXiv:1111.6554 [math.PR].
- [Tikhomirov, 1981] A. N. Tikhomirov, Übersetzung von B. Seckler, 1981: *On the convergence rate in the central limit theorem for weakly dependent random variables*. Theory of probability and its applications, Jahrgang 25, Nummer 4, Seite 790-809. Society for Industrial and Applied Mathematics.
- [Witting und Müller-Funk, 1995] Hermann Witting und Ulrich Müller-Funk, 1995: *Mathematische Statistik II. Asymptotische Statistik: Parametrische Modelle und nicht-parametrische Funktionale*. B. G. Teubner, Stuttgart.