
Kontrollkarten zur Alarmgebung in Stromnetzen

Dissertation

zur Erlangung des akademischen Grades eines Doktors der
Naturwissenschaften (Dr. rer. nat.) der Technischen Universität
Dortmund

Der Fakultät Statistik
der Technischen Universität Dortmund
im August 2018
vorgelegt von

Christian Langesberg

Erstgutachter: Dr. Uwe Ligges
Zweitgutachter: Prof. Dr. Claus Weihs

Mündliche Prüfung am 23. November 2018

Bringing a production process into a state of 'statistical control' where there is only chance-cause variation, and keeping it in control, is necessary to predict future output and to manage a process economically.

(Walter Andrew Shewhart)

Diese Witzbolde – wenn der Kern explodiert, kriegt doch das Warnsystem keinen Strom mehr!

(Homer Jay Simpson)

Inhalt

1	Einleitung	11
2	Problemstellung und Grundlagen	15
2.1	Inhaltliche Einführung	15
2.1.1	Netzfrequenz	15
2.1.2	Regelung	16
2.1.3	Grenzwerte	17
2.1.4	Netzzeit	19
2.2	Datenlage	21
2.3	ARIMA-Prozesse	26
2.4	Kontrollkarten	27
3	Alarmsysteme mit Standard-Methodik	35
3.1	Vorgehen	35
3.2	Ansätze mit klassischen Karten	36
3.2.1	Mittelwert-Karte	36
3.2.2	Aufspreizen der Grenzen	41
3.2.3	EWMA	44
3.2.4	Subsample Charts	49
3.2.5	T2-Karten	50
3.3	Ansätze aus der Zeitreihenanalyse	54
3.3.1	Zur Idee	54
3.3.2	ARIMA	56
3.3.3	VAR-Karten	60
3.4	Neue Größen	61
4	Zur Quantifizierung des Normalverteilungsgrades	69
4.1	Ausgewählte Methoden	71
4.1.1	Anforderungen	71
4.1.2	Distanzen	73
4.1.3	Teststatistiken	79
4.1.4	Umsetzung	84
4.1.5	Ausgewählte Verteilungen	86
4.2	Allgemeine Vergleiche	88
4.2.1	Vergleich der Konzepte	88
4.2.2	Vergleiche für theoretische Verteilungen	91
4.2.3	Folgerungen	94

4.3	Simulationsstudie	96
4.3.1	Simulationsdesign	97
4.3.2	Ergebnisse	97
4.3.3	Vergleich mit theoretischen Größen	105
4.3.4	Folgerungen	108
4.4	Fazit	110
4.5	Ergänzungen	112
4.5.1	Zur Stichprobengröße	112
4.5.2	Zur Optimierung	115
4.6	Zusammenfassung	119
5	Zur Verteilungs-Konvergenz von gleitenden Mittelwerten	123
5.1	Simulationsstudie	124
5.1.1	Simulationsdesign	124
5.1.2	Ergebnisse	129
5.1.3	Folgerungen	137
5.2	Allgemeinere Szenarien	137
5.2.1	Konvergenzgeschwindigkeit	138
5.2.2	Unabhängig, nicht identisch verteilt	140
5.2.3	Nicht unabhängig, nicht identisch verteilt	143
6	Alarmsysteme mit neuer Methodik	145
6.1	Zur Fehlalarmrate	145
6.2	Zur Power	146
6.2.1	Reale Beispiele	149
6.2.2	Struktur der Störungen	153
6.2.3	Simulation	157
6.2.4	Erweiterungen	160
7	Zusammenfassung und Ausblick	165
7.1	Zusammenfassung	165
7.2	Fazit	167
7.3	Ausblick	168
8	Anhang	171
8.1	Ergänzende Abbildungen	171
8.2	Ergänzende Tabellen	180
	Literaturverzeichnis	183

Abbildungen

2.1	Grenzwerte für die Netzfrequenz	19
2.2	Beispiele zum Frequenzverlauf	23
2.3	Beispiel zum Fahrplanwechsel	24
2.4	ACF und PACF der Daten	25
2.5	Idee der Kontrollkarten	29
3.1	Einfache \bar{x} -Karte	37
3.2	Einfache \bar{x} -Karte, Ausschnitt	39
3.3	Aufspreizen der Kontrollgrenzen	42
3.4	EWMA-Glättung	46
3.5	\bar{x} -Karte für für OSAPes	47
3.6	EWMA, OSAPes und Aufspreizen	48
3.7	Eigenschaften von Subsample Charts	51
3.8	T^2 -Karte	53
3.9	ACF und PACF der Differenzen	57
3.10	ARIMA-Ordnungen nach BIC	59
3.11	Verteilung der Differenzen	62
3.12	Zu Fensterbreite und Schrittweite	64
3.13	Verteilung der Transformationen	67
4.1	Idee der Lévy-Distanz	75
4.2	Beispiele der Abstandsmaße	78
4.3	Verteilungen zur Beurteilung der Abstandsmaße	87
4.4	Abstands-Optimierung als Parameterschätzer	99
4.5	Ergebnis der Abstands-Simulation (n=10)	100
4.6	Ergebnis der Abstands-Simulation (n=100)	102
4.7	Ergebnis der Abstands-Simulation (Vergrößerung, Teststatistiken)	104
4.8	Ergebnis der Abstands-Simulation (n=1 000, mit Erwartungen)	107
4.9	K und JB nach Verteilung und StiPro-Größe	113
4.10	K^* und LJB nach Verteilung und StiPro-Größe	115

4.11	K^* mit und ohne Parameteroptimierung	116
4.12	K^* mit und ohne Parameteroptimierung, alternative Darstellung	118
5.1	Verteilungen für die Konvergenz-Simulation	126
5.2	Beispiele zu K^* und LJB	127
5.3	Ergebnis der Konvergenz-Simulation (Parameter d)	130
5.4	Ergebnis der Konvergenz-Simulation (Parameter w, s und k)	132
5.5	Ergebnis der Konvergenz-Simulation (Parameter w und s_{rel} , NV)	134
5.6	Ergebnis der Konvergenz-Simulation (K^* nach s_{rel} und w)	136
5.7	Beispiel zu Berry-Esseen-Grenzen	139
5.8	Zur Konvergenz von Misch-Prozessen	142
6.1	Anwendung des Verfahrens	147
6.2	Der Stromausfall von 2006	149
6.3	Der Stromausfall von 2003	150
6.4	Nachbildung der Fehlermuster	152
6.5	Kontrollkarte für die Fehlermuster	153
6.6	Verlauf von $a(\cos(b(t + 1)) - \cos(bt))$	157
6.7	Erkennen von Störungen (summiertes Fehlermuster)	159
6.8	Erkennen von Störungen (multiplikatives Fehlermuster)	160
6.9	Erkennen von Störungen (schwingendes Fehlermuster)	161
8.1	ACF und PACF für zweifache Differenzenbildung	171
8.2	Ergebnis der Abstands-Simulation (n=1 000)	172
8.3	Ergebnis der Abstands-Simulation (Vergrößerung, Metriken)	173
8.4	Ergebnis der Abstands-Simulation (n=10, mit Erwartungen)	174
8.5	Ergebnis der Abstands-Simulation (n=100, mit Erwartungen)	175
8.6	Ergebnis der Konvergenz-Simulation (w und s_{rel} , Bin(10,0.2)-Verteilung) .	176
8.7	Ergebnis der Konvergenz-Simulation (w und s_{rel} , t_5 -Verteilung)	177
8.8	Ergebnis der Konvergenz-Simulation (w und s_{rel} , χ_5^2 -Verteilung)	178
8.9	Ergebnis der Konvergenz-Simulation (LJB nach s_{rel} und w)	179

Tabellen

2.1	Deskriptive Zusammenfassung der Daten	22
3.1	Mittelwert- und Urwertkarten	40
3.2	Korrelation der Zeitreihen	52
3.3	Eigenschaften der VAR-Karte	61
4.1	Abstandsmaße im Beispiel zur $N(0,1)$	92
4.2	Eigenschaften der Maße (theoretisch)	96
4.3	Eigenschaften der Maße (simulativ)	109
4.4	Eigenschaften der Maße (zusammenfassend)	111
5.1	Zur Parameterwahl von d und k	129
6.1	Laufängen, ohne und mit run rules, schwingendes Fehlermuster	163
8.1	Laufängen, ohne und mit run rules, multipliziertes Fehlermuster	180
8.2	Laufängen, ohne und mit run rules, summiertes Fehlermuster	181

1 Einleitung

Seit Walter Andrew Shewhart 1931 sein Standardwerk *Economic Control of Quality of Manufactured Product* (Shewhart, 1931) der Öffentlichkeit vorlegte, gehört die statistische Prozesskontrolle in Form von Kontrollkarten zu den wichtigsten Werkzeugen der Statistik in den Anwendungsbereichen des Ingenieurwesens. Zahlreiche Weiterentwicklungen und Übertragungen der Kernidee machen die Methodik flexibel und für zahlreiche Problemstellungen verfügbar. Wie üblich in der Statistik und der Mathematik müssen für die korrekte Anwendung dieses Verfahrens bestimmte Voraussetzungen – mehr oder weniger exakt – erfüllt sein. Die vorliegende Arbeit zeigt einen Fall, in dem die Abweichungen von diesen Annahmen zu groß sind und eine sinnvolle Anwendung somit nicht möglich ist. Es handelt sich dabei um den Verlauf der Stromnetzfrequenz, welche den Status eines elektrischen Netzes anzeigen kann. Im Weiteren wird ein Verfahren vorgeschlagen, welches allgemein formuliert wird und sich für die konkrete Anwendung eignet.

Zum Ende des 19. Jahrhunderts begann die Elektrifizierung Deutschlands. Im Zuge des Hafenbaus wurde beispielsweise in Dortmund 1897 das Städtische Elektrizitätswerk eröffnet, welches die zweite öffentliche Stromversorgung in Westfalen bereitstellte (DEW, 2017). Seitdem ist die Versorgung mit Strom aus dem privaten und beruflichen Leben praktisch aller Menschen der Industrie- und vieler weiterer Länder nicht mehr wegzudenken: Beleuchtung, Ampelanlagen, Produktionsstraßen, Unterhaltungselektronik und vieles mehr schaltet sich bei einem Ausfall der Stromversorgung direkt ab. Die Folgen eines länger andauernden Stromausfalls werden im Science-Thriller *Blackout* des Österreichers Marc Elsberg mit viel Fachwissen und Phantasie beschrieben: In Krankenhäusern geht der Treibstoff von Notstromaggregaten zur Neige, Tankstellen können kein Benzin mehr pumpen, Lebensmittel in Kühllhäusern verderben und auf großen Bauernhöfen können die zahlreichen Kühe ohne maschinelle Unterstützung nicht mehr gemolken werden (Elsberg, 2013). Aus diesen Gründen ist eine durchgängige Überwachung des Stromnetzes eine zentrale Aufgabe der Betreiber.

Das deutsche Stromnetz, Teil des kontinentaleuropäischen Netzes, befindet sich zur Zeit der Erstellung der vorliegenden Arbeit in zwei großen Umbrüchen: Zum einen wird die Stromerzeugung im Rahmen der Energiewende weg von der Kernkraft und fossilen

1 Einleitung

Brennstoffen hin zu erneuerbaren Energiequellen verschoben. Auch auf die Energieübertragung hat das Auswirkungen, wie die online-Ausgabe der Tagesschau bereits vor Verabschiedung der entsprechenden Gesetze unter dem Titel *Ein Netz für „Dinosaurier“ statt für „Kleinvieh“* erläuterte (Tagesschau, 2011). Zum anderen bietet die technologische Entwicklung nicht mehr nur die Möglichkeit der Übertragung des Stroms selbst, auch Daten zum Zustand der einzelnen Netzkomponenten können inzwischen in großen Mengen transferiert, gespeichert und umgehend ausgewertet werden. Damit kann schnell und dezentral auf Änderungen in Nachfrage und Angebot auf dem Energiemarkt, sowie beim Übertragungs- und Sicherheitsmanagement reagiert werden. Diese Möglichkeiten werden unter dem Begriff des intelligenten Stromnetzes zusammengefasst, wobei auch im Deutschen der kürzere englischsprachige Begriff *smart grid* gebräuchlich ist. Die Dimension dieser Netzanpassungen verdeutlichen unter anderem die von Aichele und Doleski (2014) gesammelten Beiträge (wobei beispielsweise die dort vorgenommenen Abgrenzungen zwischen Smart Grid und Smart Market für den Rahmen dieser Arbeit nicht weiter berücksichtigt werden müssen).

Als Sinnbild dieser Entwicklungen kann das Elektroauto gelten: Zum einen findet hier eine Verschiebung vom fossilen Kraftstoff Öl zu regenerativen Energien statt – die Einspeisung des Stroms durch entsprechende Kraftwerke vorausgesetzt. Zum anderen können die Akkus von Elektroautos als Zwischenspeicher verwendet werden, was wiederum nur durch ein smart grid effizient gesteuert werden kann. Auch der Zeitpunkt des Akkuladens kann in Rahmenzeiten wie etwa der Arbeitszeit des Nutzers durch die Stromversorger optimal gewählt werden. Bereits seit 2012 untersucht das Unternehmen IBM in Zusammenarbeit mit einem Stromversorger und einem Autohersteller in einem Big-Data-Projekt, wie sich unter anderem das allgemeine Stromverbrauchsmuster, die Wettervorhersage und auch spezifische Daten der einzelnen Autos wie Standort, Ladestand des Akkus und so weiter zur optimalen Planung der Ladezeitpunkte und -orte bestimmen lassen. Auch die günstigsten Standorte der Ladepunkte können aus den resultierenden Modellen bestimmt werden (vgl. Mayer-Schönberger und Cukier, 2013, S. 130 f.). Allein bei der Betrachtung des Automobils lassen sich somit immense Änderungen der Netze erkennen.

Mit all diesen Entwicklungen gehen offenbar große Änderungen für die Anforderungen an die Sicherheitssysteme einher, wobei gerade ein intelligentes Stromnetz mit moderner Informations- und Kommunikationstechnologie auch viele Möglichkeiten bietet, die bisher nicht zur Verfügung standen. Die Erforschung solcher Möglichkeiten stand unter dem Titel *Schutz- und Leitsysteme zur zuverlässigen und sicheren elektrischen Energieübertragung* im Fokus der von der Deutschen Forschungsgemeinschaft geförderten Forschergruppe FOR 1511 an der Technischen Universität Dortmund, dessen Mitglied auch

der Verfasser dieser Arbeit war. Ein – jedoch keineswegs vollständiger – Überblick über die Arbeiten der Gruppe findet sich bei Kubis et al. (2015).

Im Rahmen der Arbeit dieser Gruppe ergab sich die Frage nach den Möglichkeiten einer online-Überwachung des Stromnetzes. Als mögliche Kennzahl dazu wurde die Netzfrequenz ausgewählt, als prinzipiell mögliche Methodik die statistische Prozesskontrolle herangezogen. Wie eingangs bereits angeführt, ist eine Bearbeitung mit Standard-Methodik jedoch fehlgeschlagen und Forschungsbedarf nach weiteren Verfahren identifiziert worden. Auf den folgenden Seiten sind die Ergebnisse dokumentiert.

Im folgenden Kapitel wird zunächst der inhaltliche Rahmen kurz erläutert. Dazu wird ein grober Überblick über das europäische Stromnetz gegeben und insbesondere die Netzfrequenz als Kenngröße für den Zustand eines elektrischen Energienetzes erläutert. Darauf folgend werden die Messdaten vorgestellt, welche für die weiteren Arbeiten zur Verfügung stehen: An fünf europäischen Orten wurde für zwei Wochen zum Ende des Jahres 2014 sowie zu Anfang des Folgejahres die Netzfrequenz mitgeschrieben. Auf der Basis dieser Stichproben sollen dann Konzepte für ein Alarmsystem erstellt werden. Das grundsätzliche Vorgehen zur Erstellung solcher Standard-Kontrollkarten wird anschließend zum Ende des zweiten Kapitels erläutert.

Die Zusammenführung von Daten und Methodik wird im dritten Kapitel beschrieben. Dabei wird deutlich, dass das einfache Standard-Vorgehen nicht zielführend ist, vielmehr entsteht ein Alarmierungssystem mit indiskutablen Raten falscher Alarme. Eine Literaturrecherche zeigt, dass Methoden für die vorliegenden Annahmeverletzungen bereitstehen. Gleichzeitig sind jedoch die vorliegenden Kombinationen und die teils extremen Formen dieser Verletzungen oft nicht berücksichtigt, was sich in schlechtem bis mäßigem Erfolg der Verfahren niederschlägt. Zum Ende des Kapitels wird ein neues Verfahren zur Bildung einer Kontrollkarte vorgeschlagen. Es handelt sich dabei um die Betrachtung von Differenzen, welche absolut und anschließend potenziert Verwendung finden. Durch gleitende Mittelwerte wird zudem eine Normalverteilung der Karten-Statistik im Sinne des Zentralen Grenzwertsatzes erreicht, beziehungsweise angenähert.

Für die Anwendung dieser Methodik müssen Parametereinstellungen vorgenommen werden. Um eine Einstellung wiederum bewerten zu können, wird auch der Grad der Normalverteilung einer Stichprobe benötigt. Hierzu eignen sich prinzipiell mehrere „Messmittel“, wobei zunächst nicht klar ist, unter welchen Kriterien sich welche Vorgehensweise am besten eignet. Dieser Frage wird im vierten Kapitel nachgegangen. Darauf aufbauend erfolgt dann im fünften Kapitel die Optimierung der Parametereinstellungen, wobei eine möglichst allgemein gültige Empfehlung angestrebt wird.

1 Einleitung

Im sechsten Kapitel wird die Anwendung des vorgeschlagenen Verfahrens auf die Daten aus der Frequenzmessung dargestellt. Dabei wird dann nicht nur die Rate falscher Alarme im Zustand eines Netzes *unter Kontrolle* untersucht, auch die Dauer bis zum Erkennen eines Fehlerfalls wird ermittelt. Dazu werden zwei reale Stromausfälle aus den Jahren 2003 und 2006 betrachtet, im Weiteren dann durch Simulationen Fehlerfälle erzeugt. Abschließend erfolgt eine Zusammenfassung der vorangegangenen Kapitel, sowie die Diskussion möglicher weiterer Schritte (Kapitel 7).

Teile dieser Arbeit, insbesondere die Kapitel 4 und 5 sowie Teile von Abschnitt 3.4, wurden bereits vorab veröffentlicht und finden sich somit auch bei Langesberg et al. (2018a) und Langesberg et al. (2018b). Die Inhalte dieser Ausarbeitungen sind hier weitestgehend wortgleich übernommen. Bei der Übernahme durchgeführte Änderungen und Kürzungen berühren in keinem Fall die getätigten Aussagen, sondern tragen neben sprachlichen Überarbeitungen jeweils dem geänderten Kontext und der geänderten Typografie Rechnung. Sie begründen sich also jeweils durch die Einbindung in diese Arbeit statt der Verwendung als alleinstehende Texte.

2 Problemstellung und Grundlagen

2.1 Inhaltliche Einführung

2.1.1 Netzfrequenz

An dieser Stelle erfolgt zunächst eine kurze Einführung in das Thema *Stromnetz und Netzfrequenz*. Es sei betont, dass dabei lediglich ein grober Überblick gewährleistet werden soll, eine detaillierte Darstellung also nicht angestrebt wird. So ist auch die Netzfrequenz nur eines von vielen Merkmalen des Netzes und seines jeweiligen Zustands, bleibt aber hier das einzige Betrachtete.

Das deutsche Stromnetz ist Teil des zentraleuropäischen Verbundnetzes. Dessen Betrieb wird gemeinsam mit dem weiteren Netze durch den Verband der europäischen Übertragungsnetzbetreiber, kurz ENTSO-E für European Network of Transmission System Operators, abgestimmt. Diesem Verband gehören 43 Mitglieder aus 36 Ländern an, aus Deutschland sind vier Unternehmen vertreten: Die TransnetBW GmbH mit Sitz in Stuttgart, die TenneT TSO GmbH aus Bayreuth, die Dortmunder Amprion GmbH und die Berliner 50Hertz Transmission GmbH. Das gesamte Gebiet der ENTSO-E erstreckt sich im Westen von Portugal bis nach Island und im Osten von Finnland über Estland und die Slowakei bis in die Türkei und nach Zypern (ENTSO-E, 2018a). Eine Darstellung des Leitungsnetzes kann im Internet abgerufen werden (ENTSO-E, 2018b).

In Europa wird Wechselstrom übertragen. Dieser wechselt per definitionem periodisch die Richtung oder Polung. Die Dauer einer Periode beträgt im europäischen Verbundsystem 20 Millisekunden, womit sich eine Frequenz von 50 Hertz (Hz) ergibt (Amprion, 2018). Die Festlegung gerade dieses Normwertes hat technische wie historische Gründe, welche von Neidhöfer (2008) dargestellt werden. Beispielsweise die USA und Kanada verwendet 60 Hz, wiederum andere Frequenzen werden im Netz der Deutschen Bahn genutzt.

Der Normwert wird idealerweise konstant gehalten: „Eine einheitliche Periodenzahl bildet die Voraussetzung dafür, dass alle Kraftwerksgeneratoren im Verbund zusammenarbeiten können“ (ebenda). Wird mehr Energie benötigt als eingespeist wird, so drehen sich die Generatoren langsamer und die Frequenz sinkt. Werden die Generatoren hin-

2 Problemstellung und Grundlagen

gegen zu stark für den aktuellen Verbrauch angetrieben, drehen sie sich schneller, die Frequenz steigt also. Nur eine Frequenz in Höhe des Sollwertes bildet damit ein Gleichgewicht der Verbraucher und Erzeuger ab, und die Netzfrequenz ist somit ein direkter Indikator für den Status des Systems. Zu starke Abweichungen hingegen beeinflussen elektrische Geräte, „und auch die Generatoren der angeschlossenen Kraftwerke können beschädigt werden“, wie die Amprion GmbH am angegebenen Ort erläutert.

Auch wenn variable Einflüsse wie das Verhalten der Verbraucher, Übertragungsverluste und das schwankende Angebot beispielsweise von wetterabhängiger Stromerzeugung mit über einem Jahr Vorlauf möglichst berücksichtigt werden: Im realen Betrieb wird die Forderung nach 50 Hz nicht exakt eingehalten, es ergibt sich ein Rauschen um diesen Wert. Zudem werden die Turbinen durch ständige Eingriffe übermäßig belastet, so dass Abweichungen von der Zielfrequenz in einem als *normal* definierten Ausmaß toleriert werden. Hinzu können größere Faktoren wie der Ausfall eines Kraftwerks kommen, welche durch die Mess- und Regelsysteme zu berücksichtigen sind. Die Regelung erfolgt im Allgemeinen durch eine Anpassung der Turbinenleistung, teilweise auch durch eine (bezahlte) Anpassung des Verbrauchs von größeren industriellen Unternehmen.

2.1.2 Regelung

Um die Frequenz zu steuern beziehungsweise in die gewünschte Richtung zu beeinflussen, können verschiedene Maßnahmen unternommen werden. Diese werden nach ihrem zeitlichen Horizont unterteilt. Die direkte „Reaktion“ auf Änderungen ist die Trägheit des Netzes: Durch die Bewegungsenergie der rotierenden Massen der Generatoren wird ein Leistungsdefizit zunächst ausgeglichen. Dieser Speicher in Form von kinetischer Energie wird Momentanreserve genannt, und ist im Rahmen der Energiewende einer der Blickpunkte: Mit einem sinkenden Anteil konventioneller Kraftwerke sinkt auch die Momentanreserve. Im Auftrag unter anderem der vier deutschen ENTSO-E-Mitglieder und des Bundesministeriums für Wirtschaft und Energie stellt die Studie *Momentanreserve 2030* der dena (Deutsche Energie-Agentur GmbH) Probleme und Schlussfolgerungen vor (dena, 2016).

Nach Nutzung dieser ersten Reserve sinken die Drehzahlen der Generatoren und folglich auch die Netzfrequenz aber merklich. Die Reaktionen hierauf sind im Teil P1 des Continental Europe Operation (CEO-)Handbook (ENTSO-E, 2009) detailliert aufgeführt, hier wird nur ein kurzer Überblick gegeben. Das Handbuch betrifft das kontinentaleuropäische Netz, welches mitunter auch noch als UCTE-Netz bezeichnet wird. Dieser Name geht auf die *Union for the Coordination of Transmission of Electricity* zurück, wie die

ENTSO-E ein Verband von Übertragungsnetzbetreibern. Die Aufgaben der UCTE wurden inzwischen von der größeren ENTSO-E übernommen.

Es wird hier beispielhaft von einem Rückgang der Leistung ausgegangen, wie er etwa auf einen Kraftwerksausfall folgt. Zunächst greift in diesem Fall die *Primärregelung*: Innerhalb weniger Sekunden gleichen möglichst viele Kraftwerke das Defizit durch eine jeweils gering erhöhte Bereitstellung aus. Die Kraftwerke halten stets bestimmte Kapazitäten hierfür bereit, welche in 15 Sekunden zur Hälfte und in 30 Sekunden zur Gänze verfügbar sein müssen. Hierdurch wird das Gleichgewicht zwar wieder hergestellt, die in Reserve gehaltene Leistung wird aber offenbar geringer. Aus diesem Grund wird dann die *Sekundärregelung* eingesetzt: Es erfolgt eine Änderung des Betriebsplans. Innerhalb weniger Minuten werden entsprechend schnell zu startende Kraftwerke wie beispielsweise Wasserspeicher- oder Gaskraftwerke in Betrieb genommen, um die Reserven für die Primärregelung wieder freizugeben. Auch hier sind Reaktionszeiten und Verantwortlichkeiten bis hin zur erforderlichen Messgenauigkeit klar geregelt.

Mit der Sekundärregelung wird zwar wieder ein stabiler Netzzustand erreicht, was aber nicht zu einer wirtschaftlichen Verteilung der Energiebereitstellung führt. Ist absehbar, dass die zusätzliche Energiemenge für einen längeren Zeitraum benötigt wird, so wird die benötigte Mehr-Leistung neu auf die vorhandenen Generatoren verteilt, dabei können auch langsamer startende Kraftwerke hinzugenommen werden. Diese als *Tertiärregelung* bezeichneten Schritte können eigenständig, oder auch zur Unterstützung einer länger andauernden Sekundärregelung durchgeführt werden. Die Maßnahmen führen dazu, dass das Einspeisen von mehr Leistung möglichst gleichmäßig auf das Netz verteilt wird, und wiederum eine sichere Sekundärregelung zur Verfügung steht.

Die Verantwortlichkeiten dieser einzelnen Schritte ergeben sich durch die Zeit, in der eine entsprechende Regelung greifen muss. So sind für die Primärregelung als Sekundenreserve alle Übertragungsnetzbetreiber verantwortlich, bei der Sekundär- und gegebenenfalls der Tertiärregelung (Minutenreserve) nur die betroffenen.

2.1.3 Grenzwerte

Eine starke Abweichung der Netzfrequenz vom Sollwert ist ein klares Indiz für ein Ungleichgewicht, welches zu verschiedenen Problemen bis hin zur Schädigung der Generatoren etwa durch Resonanzfrequenzen führen kann. Zur Behandlung von Störungen gibt es daher klare Vorgaben der ENTSO-E. Diese sind im oben bereits genannten „UCTE-Handbuch“ beschrieben, wobei hier der Teil P5 *Emergency Operations* von Interesse ist (ENTSO-E, 2010). Die Vorgaben beinhalten nicht zuletzt auch vom Energienetz unab-

2 Problemstellung und Grundlagen

hängige Kommunikations-Strukturen und unterscheiden vier Status: Neben dem normalen, ausbalancierten Zustand und dem (nahezu) vollständigen *Blackout* als Extrema wird auch zwischen einem Risiko-Zustand und dem eingetretenen Notfall unterschieden.

Den Betreibern von Generatoren ist die automatische Trennung vom Netz zwischen 47.5 und 51.5 Hz verboten. Im Umkehrschluss muss davon ausgegangen werden, dass ab diesen Werten zumindest einzelne Kraftwerke in den Inselbetrieb gehen, sich also vom Netz trennen. Im Fall einer Überfrequenz kann das bei einem geordneten Ablauf gerade zur Lösung des Problems führen, während die Trennung eine zu geringe Energiebereitstellung (Unterfrequenz) noch verstärkt. Unterhalb von 47.5 Hz wird damit de facto keine Stromübertragung stattfinden.

Unter der Berücksichtigung von Regelungen wie im vorigen Teil beschrieben werden *normale Abweichungen* von der Zielfrequenz toleriert. Als normal werden bis zu 200 Millihertz (mHz) betrachtet, in diesem Bereich werden keine ungeplanten Regelungen vorgenommen. Ab 49.8 beziehungsweise 50.2 Hz erfolgen dann manuelle Eingriffe. Diese können in der Schaltung von Pumpspeicherkraftwerken in beide Richtungen oder anderen Anpassungen des Niveaus der bereitgestellten Energie bestehen. Mitunter ist auch die Dauer der Abweichung von Interesse.

Eine Unterfrequenz bis zu 49 Hz stellt einen Alarm dar, führt jedoch noch nicht automatisch zu einer Versorgungsunterbrechung. Erst ab 49.2 Hz *dürfen* und ab 49.0 Hz *müssen* die Netzbetreiber Lastabwürfe vornehmen. Dabei werden dann einzelne Regionen oder Großverbraucher und mindestens 5 % der Last vom Netz getrennt, um einem Blackout vorzubeugen. An dieser Stelle käme auch einer der Vorteile eines intelligenten Stromnetzes zum Tragen: Statt beispielsweise Gelsenkirchen vom Stromnetz zu trennen, um die Versorgung des restlichen Ruhrgebiets zu sichern, könnte ein Lastabwurf Unterhaltungselektronik und viele Haushaltsgeräte trennen, um Beleuchtung, Kommunikation und Ampeln abzusichern.

Ein weiteres Absinken der Netzfrequenz führt zu weiteren Lastabwürfen, wobei die Pläne hierfür vom jeweiligen Netzbetreiber erstellt werden. Beim Erreichen von 48 Hz sind idealerweise 50 % der Last abgeworfen. Im Allgemeinen sollten die Abwurfpunkte dabei 200 mHz nicht überschreiten, und an jedem Punkt nicht mehr als 10 % der Last abgeworfen werden. Weitere Maßnahmen bis zum Wert von 47.5 Hz liegen ebenfalls in der Verantwortung der Betreiber. An dieser Stelle werden sich dann die Kraftwerke zum Selbstschutz vom Netz trennen und dazu entsprechende Maßnahmen durchführen, beispielsweise wird etwa der Dampf von Dampfkraftwerken (also u. a. Kohle- und Kernkraftwerken) an der Turbine vorbei geleitet („Turbinenschnellschluss“) oder die Rotorblätter von Windkraftanlagen werden so eingestellt, dass sie dem Wind nur die

Kanten als Angriffsfläche bieten („Fahnenstellung“). Die angeführten Bereiche werden in der Abbildung 2.1 noch einmal zusammengefasst. Sofern nicht eindeutig definiert, sind dort beispielhafte Lastabwurf-Punkte dargestellt.

Das Ziel der folgenden Arbeiten ist die frühzeitige Erkennung einer Störung. So könnten beispielsweise auffällige Muster, wie im einfachsten Fall ein *unnormales schnelles* Ansteigen der Frequenz-Messwerte, schon vor dem Erreichen fester Schwellwerte eine Störung detektieren. Des Weiteren ist nicht ohne Weiteres klar, dass die festgesetzten Grenzen ideal sind: Als simples Beispiel könnte etwas das Erreichen eines Zustandes von 50.15 Hz mit hoher Wahrscheinlichkeit auch das spätere Überschreiten der Grenze von 50.20 Hz bedeuten. Es könnte dann bei der Regelung des Netzes Zeit gewonnen werden, indem bereits bei 50.15 Hz direkt die für die höhere Grenze vorgesehenen Eingriffe vorgenommen werden.

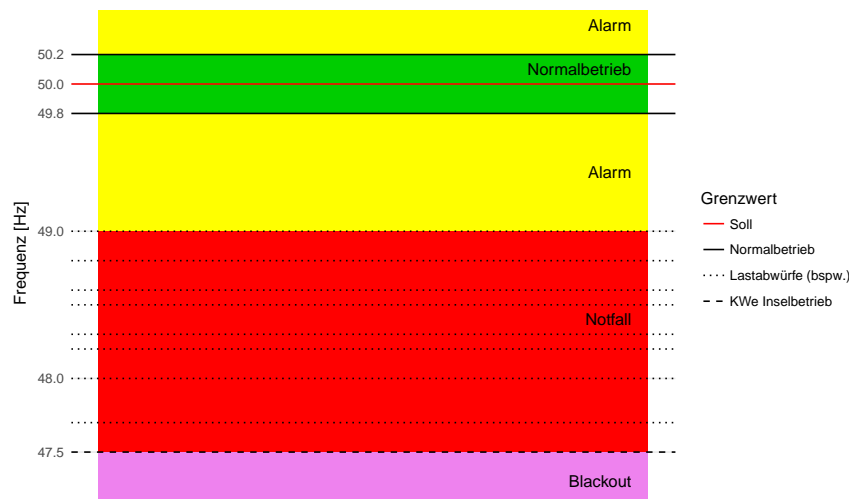


Abbildung 2.1: Das europäische Kontinentalnetz soll bei 50 Hertz betrieben werden, zu Versorgungsunterbrechungen kommt es ab 49 Hertz. Bis auf diese Grenze sind die Lastabwurf-Punkte nicht exakt vorgeschrieben und hier beispielhaft eingetragen. Die Reaktion auf eine Überfrequenz ist durch das Handbuch weniger stark reglementiert und hier nicht bis zum möglichen Inselbetrieb von Kraftwerken bei 51.5 Hertz enthalten.

2.1.4 Netzzeit

Die Frequenz dient nicht nur als Indikator für den Zustand des Energienetzes, auch die Verwendung zur Zeitmessung wird vielfach angewandt: Ein an das Stromnetz angeschlos-

2 Problemstellung und Grundlagen

senes Gerät kann die Wechsel der Polung zählen, und alle 100 Wechsel beziehungsweise alle 50 Perioden auf das Vergehen einer Sekunde schließen. Ist zudem die Uhrzeit zu einem bestimmten Zeitpunkt bekannt, kann einfach die aktuelle Zeit berechnet werden.

Da die Frequenz selten exakt 50 Hz beträgt, sind auch diese sogenannten Synchronuhren nicht exakt. Können aber kleinere Ungenauigkeiten in Kauf genommen werden, so sind sie vielfach eine vorteilhafte, weil bequeme Variante. Verwendet wird diese Technik etwa in Backofenuhren, Radioweckern oder Zeitschaltuhren. Der bereits zitierte Teil P1 des UCTE-Handbuchs legt fest, dass die Abweichungen der *Netzzeit* stets weniger als 30 Sekunden betragen sollen, in Ausnahmefällen weniger als 60. Um diese Grenzen zu gewährleisten, kann die Ziel-Frequenz von 50 Hz zeitweise um 10 mHz verändert werden.

Wie stark diese Abweichungen in Ausnahmefällen dennoch sein können, zeigte sich Anfang des Jahres 2018: Als Folge von politischen Unstimmigkeiten zwischen Serbien und dem Kosovo wurde dauerhaft etwas zu wenig Energie in das Netz eingespeist. Die durchschnittliche Frequenz sank auf unter 50 Hz, die Synchronuhren gingen um bis zu sechs Minuten nach. „Weil es irgendwo zwischen Serbien und dem Kosovo im Stromnetz knirscht, kommen Kinder bei uns zu spät zur Schule“, fasst die Tagesschau die Konsequenzen im Internet-Artikel zusammen (Tagesschau, 2018).

Von einem zweiten Beispiel der jüngeren Vergangenheit berichtet ebenfalls die Tagesschau (2017): Für Arbeiten am Stromnetz musste sichergestellt werden, dass die Solaranlagen des belgischen Dorfes Waanrode nicht anspringen. Da diese, entsprechenden Abschaltregeln folgend, bei 51 Hz keinen Strom in das Netz einspeisen, wurde die lokale Frequenz auf eben diesen Wert gesetzt. In der Folge sammelten die Waanroder Synchronuhren alle 50 Sekunden eine Sekunde Vorsprung gegenüber der realen Zeit, innerhalb eines Tages somit 28.8 Minuten. Die Verwirrung in einigen Haushalten ist gut vorstellbar, zumal sich der Fehler über mehrere Tage kumulierte – und es vorab keine Informationen des Elektrizitätswerks gab.

Eine weitere Eigenschaft der Netzfrequenz versucht die bayerische Polizei zu nutzen: Die Schwankungen in einer gewissen Zeitspanne sind so spezifisch, dass sie als Zeitstempel verwendet werden können. Beispielsweise bei einer Videoaufnahme enthält die Tonspur „ein leises Brummen“, welches auf die Frequenz zurückgeht. Wie die Süddeutsche Zeitung berichtet, kann so unter anderem überprüft werden, ob ein als Alibi verwendetes Video tatsächlich zur behaupteten Zeit aufgenommen wurde (SZ, 2011).

2.2 Datenlage

Das Ziel der vorliegenden Arbeit besteht in der Konstruktion eines statistischen Alarmsystems, welches den Zustand eines Stromnetzes *live* oder *online* beurteilen kann. Wie im vorigen Kapitel beschrieben, sind dazu Messwerte der Netzfrequenz eine geeignete Basis. Weiter ist die statistische Prozesskontrolle mittels Kontrollkarten (s. u.) ein prinzipiell geeignetes Werkzeug. Zur Durchführung dieses Vorhabens stehen zwei Datensätze zur Verfügung, welche Frequenzmessungen von fünf Orten des zentraleuropäischen Verbundnetzes über acht beziehungsweise über mehr als fünf Tage enthalten. Die Daten werden an dieser Stelle kurz vorgestellt, wobei insbesondere jene Aspekte berücksichtigt werden, welche im Verlauf der weiteren Analyse noch von Interesse sein werden.

Die Messwerte wurden durch das Institut für Energiesysteme, Energieeffizienz und Energiewirtschaft der TU Dortmund durch ein eigenes verteiltes Frequenzmesssystem niederspannungsseitig (380 V) aufgezeichnet und dem Verfasser dieser Arbeit zur Verfügung gestellt. Enthalten sind Zeitabschnitte aus dem Dezember 2014 und dem Januar 2015. Gemessen wurden die Frequenzen jeweils in Dortmund, Magdeburg, Ljubljana (Slowenien), Bukarest (Rumänien) und Almelo (Niederlande). Durch eine Aufzeichnungsrate von 50 Hz ergeben sich insgesamt 34 559 952 beziehungsweise 23 160 483 Werte für jeden der fünf Messpunkte.

Kennzahlen der Messwerte sowie des Mittels über alle fünf Messorte können der Tabelle 2.1 entnommen werden. Dabei werden die Reihen vor der Berechnung auf ihre 0.0001- und 99.9999-Prozent-Quantile begrenzt: Durch einzelne extreme Ausreißer werden die Werte stark verzerrt. Diese Ausreißer sind vermutlich durch sogenannte Fahrplanwechsel zu erklären, das heißt durch das planmäßige und möglichst gleichzeitige Zu- und Abschalten verschiedener Kraftwerke (s. u.). Die Angabe der Extrema wie beschrieben, notiert als $Q_{0.000001}$ und $Q_{0.999999}$, bildet die *eigentliche* Spannweite damit besser ab. Auch die anderen Schätzer sind in ihren getrimmten Varianten robuster. Angegeben als Streuungsschätzer ist an dieser Stelle jeweils das klassische Maß der Standardabweichung. Im folgenden Teil 2.4 wird dann eine Unterscheidung von globaler und lokaler Variabilität berücksichtigt.

Fehlende Werte sind nur für relativ kleine Zeitbereiche zu notieren. Insgesamt fehlen über alle zehn Reihen (fünf Orte, zwei Zeitpunkte) in 218 Abschnitten Werte, bezogen auf die einzelne Reihe sind es 2 bis 81. Diese Abschnitte weisen eine Länge von durchschnittlich 32 bis 5 5583.75 Beobachtungen auf, was 0.64 bis 111.7 Sekunden entspricht. Die Werte aller fünf Orte zugleich fehlen nur für drei Abschnitte mit Längen von 56, 64 und 1 470 Beobachtungen (1.12 bis 29.40 Sekunden). Für alle anderen Bereiche kann der

2 Problemstellung und Grundlagen

Messpunkt	Mittelw.	Std.-A.	Schiefe	Wölbung	$Q_{0.000001}$	$Q_{0.999999}$
Zeitraum 1					08.12.2014 00:00:00	15.12.2014 23:59:59
Ljubljana	49.99984	0.02108	0.13136	3.84035	49.90200	50.10848
Dortmund	50.00033	0.02107	0.13226	3.84481	49.90300	50.10900
Magdeburg	49.99992	0.02104	0.13287	3.84797	49.90300	50.10800
Bukarest	49.99985	0.02114	0.13246	3.83335	49.90016	50.10934
Almelo	49.99983	0.02107	0.13233	3.84272	49.90228	50.10818
Mittel	49.99995	0.02103	0.13315	3.84991	49.90318	50.10810
Zeitraum 2					01.01.2015 15:02:25	06.01.2015 23:42:34
Ljubljana	49.99941	0.01972	-0.03880	3.64795	49.90580	50.10004
Dortmund	49.99988	0.01969	-0.03850	3.65002	49.90600	50.10100
Magdeburg	49.99948	0.01968	-0.03897	3.65106	49.90600	50.10000
Bukarest	49.99941	0.02184	-0.03845	3.64948	49.87478	50.10049
Almelo	49.99941	0.01970	-0.03841	3.64747	49.90561	50.10064
Mittel	49.99951	0.01981	-0.03851	3.65187	49.90617	50.10021

Tabelle 2.1: Deskriptive Zusammenfassung der beiden Datenreihen. Die Randpunkte der Zeiträume sind nicht als Quantile angegeben.

Mittelwert aller Orte berechnet werden, wenn auch mitunter für verschiedene Stichprobengrößen.

Nach Rücksprache mit dem Bereitsteller der Daten werden die betroffenen Messwerte als *missing at random* angesehen – das Fehlen wird auf die Messtechnik, nicht auf die Messgröße oder andere Systematiken zurückgeführt. Das Fehlen eines Wertes ist damit nicht interpretierbar und es sind keine Verzerrungen durch fehlende Werte zu erwarten.

Die Abbildung 2.2 zeigt beispielhaft einen Ausschnitt der Daten von zwei Orts-Reihen sowie dem Mittelwert aller fünf Messpunkte in verschiedenen Auflösungsgraden. Gut zu erkennen ist die starke Abhängigkeit zwischen zwei Messpunkten zur gleichen Zeit, obwohl mit Almelo und Bukarest zwei geografisch recht weit entfernte Orte betrachtet werden (die „Luftlinie“ als direkte Verbindung ist 1 650 Kilometer lang und zieht sich durch sechs Staaten). Weiter ist auch die hohe zeitliche Abhängigkeit zu erkennen, welche im Weiteren noch näher in den Fokus gestellt wird. Beide Arten der Abhängigkeit sind physikalisch bedingt und sogar ein Ziel der regeltechnischen Eingriffe in den Prozess.

Eine nennenswerte Auffälligkeit der Daten sind die bereits genannten „Fahrplanwechsel“, welche das Muster der Abbildung 2.3 aufweisen. Kleinste zeitliche Unterschiede beim Abschaltzeitpunkt eines Kraftwerks und dem Zuschaltzeitpunkt eines anderen Kraft-



Abbildung 2.2: Verlauf der Netzfrequenz am 1. Dezember 2014, in Betrachtung verschieden langer Ausschnitte und jeweils beginnend um 0:00:00 Uhr. Der Mittelwert wurde aus allen fünf Messorten berechnet, nicht nur aus den zwei dargestellten.

2 Problemstellung und Grundlagen

werks können zu kurzzeitigen Schwankungen führen. In den Daten äußert sich dieses Verhalten durch eine kurze Abfolge von Sinken und Anstieg der Frequenz, was dann aber nicht auf einen etwaigen kritischen Netzzustand schließen lässt. Die Datenpunkte mit Fahrplanwechseln werden in ihrer Form belassen: Zum einen handelt sich hier um ein charakteristisches Merkmal der Daten, zum anderen müsste eine automatisierte Erkennung erst eingeführt werden. Grundlage dieser Entscheidung ist auch die fachkundige Einschätzung des Bereitstellers der Daten, nach der die Verläufe nicht auf einen kritischen Zustand schließen lassen.

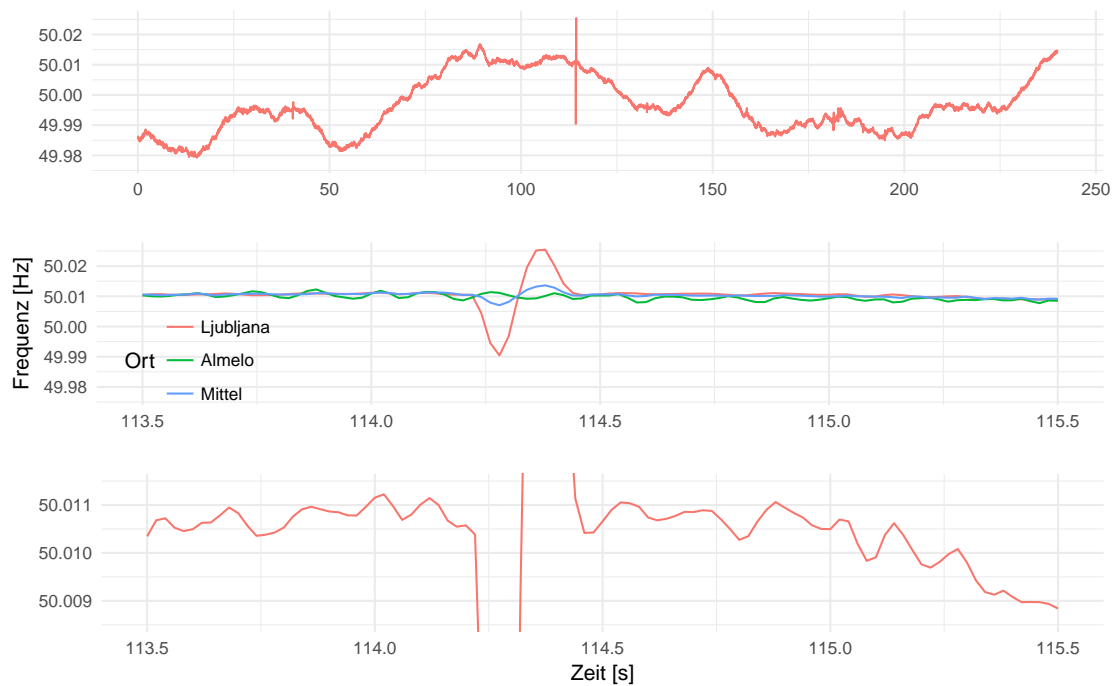


Abbildung 2.3: Ein beispielhafter Fahrplanwechsel, aufgenommen in Ljubljana am 4. Januar 2015 um 18:26:35 Uhr: Ein Kraftwerk wird vom Netz genommen oder reduziert die Leistung, und die Frequenz beginnt als direkte Folge schnell zu sinken. Kurz danach wird ein anderes Kraftwerk angeschlossen oder fährt die Leistung weiter hoch. Auf die anderen Messstellen wirkt sich das nicht aus, auf den Mittelwert entsprechend gedämpft. Die mittlere Grafik stellt einen Ausschnitt der oberen Grafiken dar, die untere wiederum einen Ausschnitt der mittleren. Aus Gründen der besseren Lesbarkeit sind die anderen beiden Reihen nur in der mittleren Grafik enthalten.

Wie bereits angedeutet, besteht eine sehr große zeitliche Abhängigkeit in den Daten. Diese stellt eine dem Stromnetz immanente Eigenschaft dar, wobei die Trägheit wie

oben beschrieben sogar im positiven Sinne aufgefasst und als Momentanreserve verwendet wird. Mit dem auf der Seite 390 von Venables und Ripley (2002) beschriebenen Verfahren zur Schätzung der Autokorrelation ergeben sich für die einzelnen Zeitreihen Werte zwischen 0.99782 und 0.99985. Bei der Reihe der Mittelwerte beträgt diese Kennzahl 0.99978.

Die Abbildung 2.4 zeigt die Autokorrelationsfunktion für zwei der Orte und die Reihe der Mittelwerte. Auch für einen Lag 100, was zwei Sekunden entspricht, ist das Selbstabhängigkeits-Maß noch größer als 0.99

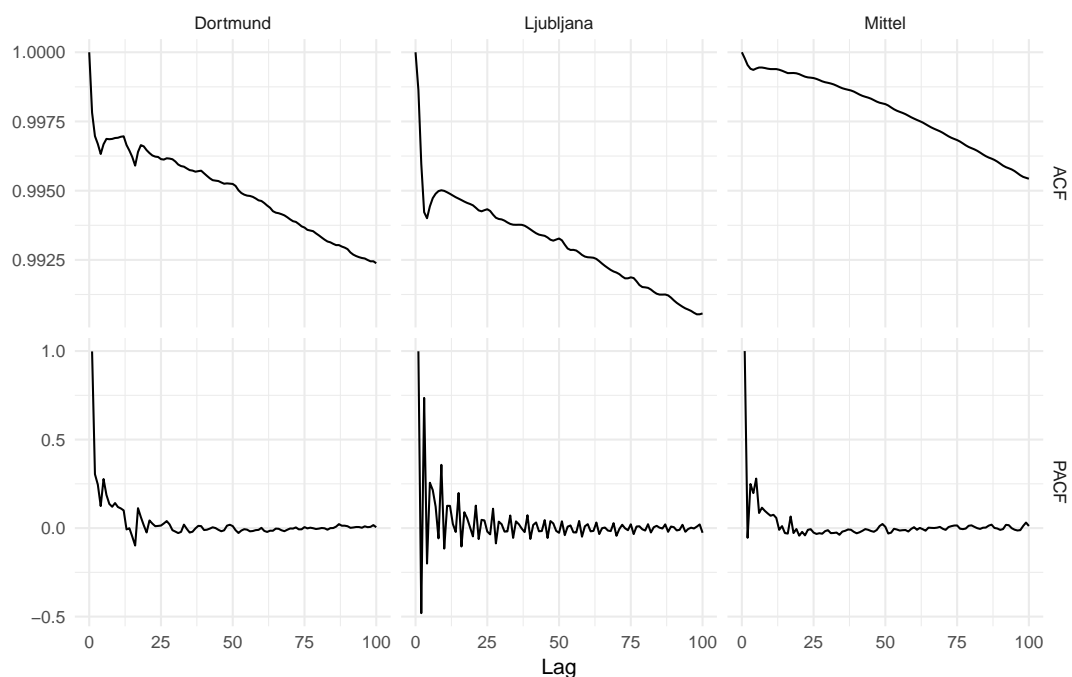


Abbildung 2.4: Die Autokorrelationsfunktionen (ACF) sowie deren partielle Pendanten (PACF) für zwei Messpunkte und den Mittelwert aller fünf Messorte eines Zeitraums, dargestellt für Lags bis 100.

Auch bei größeren Lags fallen die Autokorrelationen noch sehr groß aus. Für die Messstelle Dortmund beispielsweise beträgt die Autokorrelation auch bei Lag 10 000 (umgerechnet also nach weit über 3 Minuten) noch über 0.58, ohne vorher die Null zu erreichen. Eine genauere Analyse der Daten mit Methoden der Zeitreihenanalyse wird im Verlauf der Auswertungen folgen (siehe Teil 3.3). An dieser Stelle sei aber bereits erwähnt, dass die Autokorrelationen der Differenzen zum Lag 1 zwischen -0.3855 und 0.4795 betragen (0.0548 für die Mittelwertreihe).

2.3 ARIMA-Prozesse

Im Folgenden werden an mehreren Stellen verschiedene, recht allgemeine Begrifflichkeiten und das Konzept von sogenannten ARIMA-Prozessen verwendet. Diese werden hier zusammenfassend erläutert.

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. Offenbar sind Zeitreihen damit Spezialfälle von stochastischen Prozessen. 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 $\{X_t\}$ für $\{X_{t_i}\}_{t_i \in \mathbb{R}}$ geschrieben werden. Weiter kann stets eine eindeutige Zuordnung der Zeitpunkte erstellt werden, so dass $t \in \mathbb{Z}$ angenommen werden kann. Die Elemente des Prozesses werden entsprechend kurz durch X_t notiert, die Realisierungen durch x_t .

Sei der stochastische Prozess $\{\varepsilon_t\}$ ein weißes Rauschen, das bedeutet seien $E(\varepsilon_t) = 0$ und $V(\varepsilon_t) = \sigma_\varepsilon^2$ für alle t , und sei $E(\varepsilon_{t_i} \varepsilon_{t_j}) = 0$ für alle $t_i \neq t_j$. Lässt sich ein stochastischer Prozess $\{Y_t\}$ mit von Null verschiedenen Parametern $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ dann durch eine Gleichung der Art

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

beschreiben, so wird dieser als *ARMA(p,q)-Prozess* bezeichnet (vgl. dazu neben vielen anderen Hartung et al. (2005), S. 682). Dieser Name setzt sich aus Abkürzungen der beiden Modellteile zusammen: Wird die obige Gleichung nur bis zum Summanden ε_t betrachtet, wird von einem Autoregressiven (AR-)Prozess gesprochen. Nur der hintere Teil der Summe (wiederum inklusive dem Summanden ε_t) bildet einen moving average (MA-)Prozess. Die Realisationen von ε_t werden für alle Fälle häufig als Innovationen bezeichnet. Ein oft verwendeter Sonderfall von ARMA-Prozess ist der AR(1)-Prozess, das heißt ein Prozess mit der Eigenschaft

$$Y_t = \rho Y_{t-1} + \varepsilon_t$$

Die Notation mit ρ statt ϕ_1 deutet keine Unterschiede an, sondern gibt nur verschiedene gebräuchliche Schreibweisen wieder. Statt des Prozesses selbst können auch die Differenzen

$$\nabla^d Y_t = \nabla^{d-1}(\nabla Y_t) = \nabla^{d-1}(Y_t - Y_{t-1})$$

einem ARMA(p,q)-Prozess entstammen. Dann wird Y_t als *ARIMA(p,d,q)-Prozess* bezeichnet (a. a. O., S. 668 und 684). Das *I* in der Abkürzung steht dabei für *integrated*. Der einfache ARMA(p,q)-Prozess wird als ARIMA(p,0,q)-Prozess aufgefasst, was $\nabla^0 Y_t = Y_t$ impliziert.

ARIMA-Prozesse sind weit verbreitete und gut untersuchte Strukturen. Weitergehende Einführungen in die Eigenschaften, Parameterschätzer, Erweiterungen und Anwendungen finden sich – neben vielen anderen – bei Hartung et al. (2005) und bei Judge et al. (1985).

2.4 Kontrollkarten

Die Netzfrequenz kann also wie oben beschrieben aus technischer Sicht gut für eine online-Beurteilung des Netzstatus verwendet werden. Aus statistischer Sichtweise handelt es sich hier um ein Klassifikationsproblem: Zu schätzen ist der ordinal skalierte oder binär klassierte Netzzustand zu einem bestimmten Zeitpunkt t_0 , wozu die Zeitreihen der Frequenzen x_{t_i} für alle t_i , $i \leq 0$, zur Verfügung stehen. Diese Prognose ist ständig neu durchzuführen. Das Werkzeug zur Erstellung dieser Prognosen sollte es offenbar ermöglichen, möglichst wenige Fehlalarme zu produzieren, im Falle einer Netzstörung diese aber möglichst schnell anzeigen. Zudem sollte eine Prognose-Methode möglichst wenig Rechenzeit benötigen, um eine Klassifikation des Status möglichst aktuell bereitzustellen. Außerdem muss die Methodik an zahlreichen Messpunkten zugleich eingesetzt werden können. Eine weitere Anforderung stellt die Automatisierbarkeit der Modellanpassung dar: Änderungen der Netzarchitektur werden laufend vorgenommen, so dass etwa Schwellenwerte der Kennzahlen nicht manuell vom Bediener festzulegen sein sollten.

Gegenüber vielen anderen Anwendungen von Klassifikationsverfahren besteht im Fall der Beurteilung des Stromnetzes eine Besonderheit: Alle vorliegenden Beobachtungen stammen aus einer Klasse, nämlich der eines intakten Netzes. Daher wird im Folgenden zum einen nur zwischen den Status „unter Kontrolle“ und „nicht unter Kontrolle“ unterschieden. Zum anderen sind aus dem gleichen Grund eine Vielzahl von Klassifikationsverfahren nicht anwendbar, da diese für die Modellanpassung Beobachtungen aus allen Gruppen benötigen.

Eine Bearbeitung dieser Aufgabenstellung findet sich bei Kontrollkarten wieder. Dieses Werkzeug wurde 1931 von Walter Shewhart, einem Mitarbeiter der Bell Laboratories, im Grundlagenwerk *Economic Control of Quality of Manufactured Product* veröffentlicht (Shewhart, 1931). Für die Beurteilung eines (Fertigungs-)Prozesses wird das Prinzip eines Systems *unter (statistischer) Kontrolle* verwendet. Der grundlegende Gedanke ist dabei:

2 Problemstellung und Grundlagen

„[A] phenomenon will be said to be controlled when, through the use of past experience, we can predict, at least within limits, how the phenomenon may be expected to vary in the future. Here it is understood that prediction within limits means that we can state, at least approximately, the probability that the observed phenomenon will fall within the given limits.“ (Seite 6)

Für die Übertragung dieser Idee in eine reale Anwendung ist zunächst zu klären, welches Phänomen, also welche Prozessgröße, beobachtet wird. Möglich sind direkt interpretierbare Messwerte, aber auch Transformationen hiervon. Je nachdem kann sich dann die Suche nach den Gründen für eine Beobachtung außerhalb der gut durch Zufall erklärbaren Bereiche einfacher oder schwieriger gestalten. Eine Kontrollkarte kann konstruiert werden, indem die Grenzen bestimmt werden, welche die beobachtete Größe im Fall unter Kontrolle nur sehr unwahrscheinlich überschreitet. Vergleichsweise einfach ist dann die Möglichkeit, die Daten durch die Annahme einer passenden statistischen Verteilung zu modellieren. Auch der Begriff *unwahrscheinlich* muss für die Anwendung spezifiziert werden. Mit Schätzern für Erwartungswert und Standardabweichung des Prozesses, notiert als $\hat{\mu}$ und $\hat{\sigma}$, werden die sogenannten Kontrollgrenzen (engl. upper/lower control limit) dann festgelegt durch

$$\text{UCL/LCL} = \hat{\mu} \pm c \hat{\sigma}$$

wobei $c \in \mathbb{R}_{>0}$ vom Anwender oder von der Anwenderin zu wählen ist, und die zulässige Rate an Fehlalarmen steuert sowie indirekt auch die Rate korrekt erkannter Alarme beeinflusst.

Die Abbildung 2.5 auf der folgenden Seite verdeutlicht die Systematik: Für eine Prozessgröße x_t sei bekannt, dass sie einer Standardnormalverteilung folgt. Die Kontrollgrenzen -3 und 3 folgen der obigen Konstruktionsweise. Dabei ist $c = 3$ gewählt, was einer gängigen Vorgehensweise entspricht. Der Begriff *unwahrscheinlich* wird damit durch $\Phi(-3) + 1 - \Phi(3) = 2\Phi(-3) \approx 0.0027$ definiert, wobei Φ die Verteilungsfunktion der Standardnormalverteilung bezeichnet. Werte innerhalb der Grenzen gelten als unauffällig, bei Werten außerhalb dieser Grenzen wird ein Alarm ausgelöst. Das Prinzip weist eine große Verwandtschaft zu dem der statistischen Testtheorie und dem von Konfidenzintervallen auf: Etwa die Idee der sogenannten Mittelwertkarten entspricht prinzipiell der ständigen Wiederholung eines Gauß-Tests.

Eben diese Mittelwertkarten geben auch ein Beispiel dafür, statt der einfachen Beobachtung der jeweiligen Messwerte neue Kenngrößen zu bilden. Dies geschieht dann etwa durch die Zusammenfassung von aufeinander folgenden Messungen durch Mittelwert,

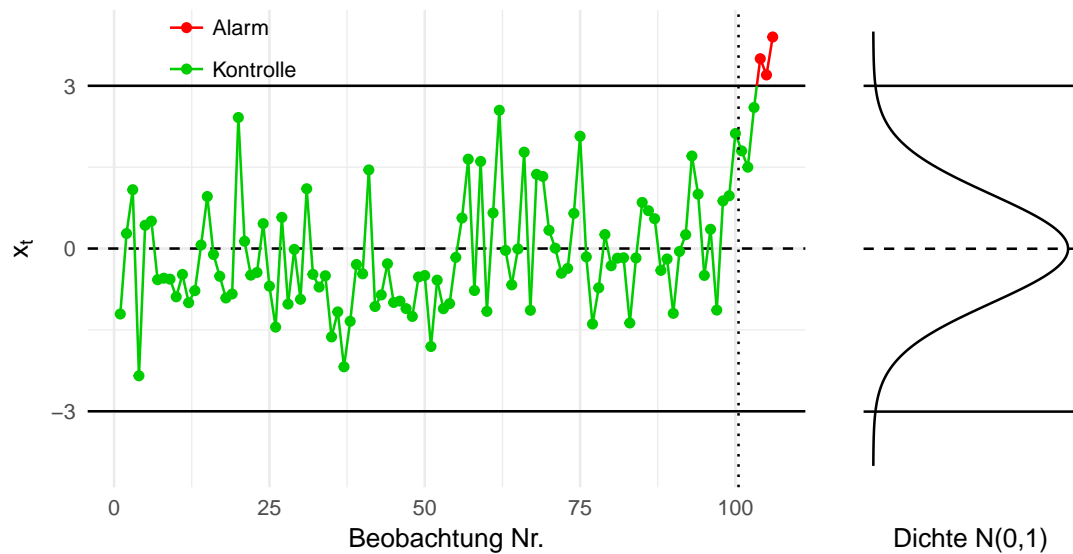


Abbildung 2.5: Ein simples Beispiel der Funktionsweise einer Kontrollkarte: 100 aus einer $N(0,1)$ -Verteilung gezogene Werte zeigen hier keine Auffälligkeiten. Eine Prozessstörung zwischen den Beobachtungen Nummer 100 und 101 verändert die Verteilung der Werte, was hier nach vier Beobachtungen aufgedeckt wird.

Spannweite oder dergleichen. Auch diskrete Varianten sind vielfach bekannt. Weitere Forderungen, wie beispielsweise die Unabhängigkeit der Werte, können erfüllt oder nicht erfüllt sein, was wiederum eine weitere Variantenbildung nach sich zieht.

In der Praxis und gerade traditionell werden die Verfahren der statistischen Prozesskontrolle in der Industriestatistik eingesetzt, was sich auch im Großteil der Auswahl an Beispielen in der einschlägigen Literatur niederschlägt. Beobachtet werden können Eigenschaften der Produkte selbst, aber auch Zustände der Maschinen und anderer Prozessparameter. Anwendungen in anderen Gebieten finden sich aber ebenfalls reichlich, wobei hier mit Anderson und Diaz (1996) nur eines der exotischsten angeführt werden soll: Die Effektivität bestimmter struktureller Änderungen bei der Polizei von Houston (Texas) wurde mit Kontrollkarten untersucht. So finden sich am angegebenen Ort dann unter anderem auch eine „Motor vehicle theft control chart“ und eine „Rape control chart“.

Die Vielfalt der Varianten von Kontrollkarten wird in zahlreichen Lehr- und Handbüchern zur Statistik im Bereich industrieller Anwendungen behandelt. So beispielsweise auch in den entsprechenden Kapiteln von Weihs und Jessenberger (1999) oder von Dietrich und Schulze (2014), wobei mitunter auch der Begriff der *Qualitätsregelkarte* verwendet wird. Ebenso widmen sich ganze Monografien dem Thema, hier sind insbesondere

2 Problemstellung und Grundlagen

Wheeler (1995) und Montgomery (2009) zu nennen. Auch bei einer Einschränkung auf Spezialthemen wie multivariate Kontrollkarten findet sich eine große Auswahl an Methoden, wie zum Beispiel bei Trautmann (2004) nachgelesen werden kann.

In der folgenden Anwendung der Kontrollkarten-Methodik werden Messwerte der Netzfrequenz beobachtet, wozu ohne Weiteres keine sinnvolle Verteilungsannahme getroffen werden kann. Zumindest versuchsweise kann zunächst aber dennoch eine Normalverteilung unterstellt werden: Wie Wheeler (1995) in den Teilen 5.1 und 5.3 mit Verweis auf Untersuchungen von Burr (1967) ausführt, begründet sich die häufige Verwendung der 3σ -Regel, also der Verwendung von Kontrollgrenzen der Art $\mu \pm 3\sigma$ wie oben, gerade in ihrer Robustheit gegenüber Abweichungen von der Verteilungsannahme.

Die Vielzahl der Möglichkeiten, eine Kontrollkarte zu erstellen, kann an dieser Stelle nicht abgebildet werden. Stellvertretend wird die bereits angesprochene und nach Weihs und Jessenberger „beliebteste“ (1999, S. 291) Variante vorgestellt: Die Mittelwert-Karte, häufig notiert als \bar{x} -Karte.

Einer Mittelwertkarte liegt die Realisierung einer univariaten Zeitreihe x_1, x_2, \dots, x_m zugrunde. Im Rahmen dieser Arbeit wird dies stets die über die Orte gemittelte Netzfrequenz sein, das Vorgehen könnte aber auch für jede der Einzelreihen genau so angewandt werden. Liegt nun eine solche Reihe vor, werden Gruppen der Größe $g \in \mathbb{N}$ gebildet, wobei $g \geq 2$ vorausgesetzt wird (zu $g = 1$ siehe unten). Lässt sich m nicht ganzzahlig durch g teilen, werden die übrigen Beobachtungen nicht berücksichtigt. Innerhalb der Gruppen werden dann Mittelwerte gebildet, deren Verteilung bei ausreichend großem g approximativ einer Normalverteilung folgt (ebenda). Weiteres zur Gruppenbildung durch *rational subgrouping* beschreibt Montgomery (2009) auf den Seiten 193 ff. Nach der Aufteilung auf n Gruppen können die Werte in Matrixform angeordnet werden, dabei wird die Notation durch zwei Indizes angepasst:

$$Y := \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1g} \\ y_{21} & y_{22} & \dots & y_{2g} \\ \vdots & & & \vdots \\ y_{n1} & y_{n2} & \dots & y_{ng} \end{pmatrix} := \begin{pmatrix} x_1 & x_2 & \dots & x_g \\ x_{g+1} & x_{g+2} & \dots & x_{2g} \\ \vdots & & & \vdots \\ x_{(n-1)g+1} & x_{(n-1)g+2} & \dots & x_{ng} \end{pmatrix}$$

Die durch eine Mittelwertkarte beobachtete Statistik ergibt sich dann aus den Mittelwerten der Gruppen

$$\bar{y}_i = \frac{1}{g} \sum_{j=1}^g y_{ij}, \quad i = 1, \dots, n$$

also durch die Reihenmittel der Matrix Y . Wie oben bereits erwähnt, werden Grenzen der Art $\mu \pm 3\sigma$ verwendet, diese beiden Parameter sind bezüglich der \bar{y}_i zu schätzen. Für die Lage wird dabei im Allgemeinen der Mittelwert der Mittelwerte verwendet:

$$\hat{\mu} = \bar{\bar{y}} = \frac{1}{n} \sum_{i=1}^n \bar{y}_i \left(= \frac{1}{n} \sum_{i=1}^n \frac{1}{g} \sum_{j=1}^g y_{ij} = \frac{1}{m} \sum_{l=1}^m x_l \right)$$

Nicht eindeutig ist hingegen das Vorgehen zur Schätzung der Streuung. Einerseits steht mit der empirischen Varianz ein gebräuchlicher Schätzer zur Verfügung, andererseits ist dieser bei der Erstellung von Kontrollkarten häufig nicht ideal: Weihs und Jessenberger (1999) weisen auf der Seite 292 auf die Unterscheidung von globaler und lokaler Varianz hin. Dabei wird die Streuung der Werte auf der Ebene aller Werte von der Streuung unterschieden, welche sich nur innerhalb der Gruppen ergibt. Für einen Schätzer auf der Basis lokaler Variabilität hat dann „[e]in etwaiger Unterschied zwischen den Gruppenmittelwerten [...] keinen Einfluss.“ Auch da für unabhängige Datenpunkte die Schätzvarianten beider Vorgehensweisen zu gleichen Werten führen, wird zumeist ein Schätzer verwendet werden, welcher auf der lokalen Varianz beruht. Dieses Konzept wird auch von Wheeler (1995) bevorzugt, demzufolge es „basic to Shewhart’s control charts“ ist (S. 70).

Weihs und Jessenberger (1999) geben mit diesen Überlegungen als Schätzer für die Standardabweichung eine Methode auf Basis der durchschnittlichen Standardabweichung der Gruppen an (S. 291 ff.):

$$\hat{\sigma}_{\bar{s}} = \frac{1}{a_g \sqrt{g}} \bar{s} = \frac{1}{a_g \sqrt{g}} \frac{1}{n} \sum_{i=1}^n s_i = \frac{1}{a_g \sqrt{g}} \frac{1}{n} \sum_{i=1}^n \frac{1}{g-1} \sum_{j=1}^g (y_{ij} - \bar{y}_i)^2$$

Dabei ist \bar{s}/a_g ein Schätzer für die Standardabweichung der y_{ij} , der Faktor $1/\sqrt{g}$ macht daraus einen Schätzer für die Standardabweichung der \bar{y}_i . Es wird die Unabhängigkeit der Beobachtungen vorausgesetzt (vgl. bekannte Rechenregeln der Varianz). Die von der Gruppengröße abhängige Konstante a_g sichert die Erwartungstreue der Schätzung. Wie Montgomery (2009) auf der Seite 111 angibt, ist zwar $E(s^2) = \sigma^2$, aber $E(s) \neq \sigma$:

$$E(s) = \sqrt{\frac{2}{g-1}} \frac{\Gamma(n/2)}{\Gamma(n-1)/2} \sigma =: a_g \sigma.$$

Mit Γ wird die Eulersche Gammafunktion bezeichnet. Je nach verwendeter Literatur wird a_g auch als c_4 bezeichnet, zudem werden Näherungen $a_g \approx 1 - 1/4(g-1)$ (Weihs und Jessenberger, 1999, Seite 292) und $a_g \approx 4(g-1)/4g-3$ (Montgomery, 2009, Seite 702) angegeben. An den beiden Näherungen lässt sich einfach $a_g \rightarrow 1$ für $g \rightarrow \infty$ ablesen.

2 Problemstellung und Grundlagen

Unter Unabhängigkeit der Beobachtungen innerhalb einer Gruppe sind die obere und die untere Alarmgrenze UCL und LCL dann also zu bilden durch

$$\text{UCL}_{\bar{y}}/\text{LCL}_{\bar{y}} = \bar{y} \pm 3 \frac{\bar{s}}{a_g \sqrt{g}}$$

An vielen Stellen der Literatur findet sich ein ebenfalls lokaler Streuungsschätzer auf Basis der Spannweiten innerhalb der Gruppen. Montgomery (2009) betont, dass dieses Vorgehen dem beschriebenen Schätzer $\hat{\sigma}_{\bar{y}}$ unterlegen ist und sich die Verwendung nur in Zeiten vor modernen Rechnern und Computern durch die Einfachheit des Vorgehens gerechtfertigt hat („Using the range to estimate σ dates from the earliest days of statistical quality control, and it was popular because it is very simple to calculate.“, Seite 112).

Zudem stehen viele weitere Ansätze zur Schätzung der Streuung bereit, entsprechend widmet beispielsweise Wheeler (1995) dem Thema „Estimating Dispersion Parameters“ ein ganzes Kapitel. Der Autor erklärt aber auch, dass die Methoden für Prozesse unter statistischer Kontrolle ähnliche Schätzer der Standardabweichung liefern (S. 65). Mit Blick auf die im Rahmen dieser Arbeit verwendeten großen Fallzahlen gilt diese Überlegung umso deutlicher. Alternative Schätzer für die Streuung wurden auch von Chen (1997) untersucht. Dort wird auch berücksichtigt, dass die Grenzen geschätzt werden und nicht, wie beim beschriebenen Vorgehen implizit unterstellt, bekannte Konstanten sind. Bei den hier betrachteten Stichprobengrößen fällt dies aber nicht ins Gewicht (s. S. 797).

Bisher wurde für die Mittelwertkarte eine Gruppengröße $g \geq 2$ vorausgesetzt, der Sonderfall *jede Beobachtung bildet ihre eigene Gruppe* also nicht zugelassen. Im Prinzip ist das Vorgehen zur Bildung einer Mittelwertkarte auf diesen Fall übertragbar, es wird dann von *Urwert- oder Einzelwertkarten* gesprochen. Allerdings ist die Schätzung der Streuung dann nicht wie beschrieben möglich.

Eine Möglichkeit, die lokale Variabilität zu beschreiben, ist in diesem Fall nach Weihs und Jessenberger (1999, S. 298 ff.) die Verwendung der mittleren gleitenden Spannweite mit Fenstern der Breite $q \in \mathbb{N}$, $q \geq 2$. Zwar wäre auch hier die Nutzung der empirischen Standardabweichung ohne Weiteres möglich, im Folgenden verwendete Methoden bauen aber auf dem Schätzer $\hat{\sigma}_{mR}$ auf (vgl. Kapitel 3). Für Urwertkarten wird daher als Streuungsschätzer verwendet

$$\hat{\sigma}_{\overline{mR}} = \frac{\overline{mR}}{d_q} = \frac{1}{d_q} \frac{1}{m-q+1} \sum_{i=q}^m mR_q = \frac{1}{d_q} \frac{1}{m-q+1} \sum_{i=q}^m (\max X_i - \min X_i),$$

dabei ist $X_i = \{x_{i-q+1}, x_{i-q+2}, \dots, x_i\}$

zudem gilt $g = 1 \Rightarrow n = m$

Auch hier wird ein Korrekturfaktor benötigt, um die Erwartungstreue des Schätzers zu sichern. Dieser Faktor d_q ist in der Literatur vertafelt, teils auch für alle q notiert als d_2 . Aus dem Zusammenhang $E(\overline{mR}) = d_q \sigma$ für unabhängig standardnormalverteilte Zufallsvariablen kann der Wert mit numerischen Methoden und der Verteilungsfunktion Φ der $N(0,1)$ auch direkt berechnet werden durch (vgl. Tippett, 1925)

$$d_q = \int_{-\infty}^{\infty} (1 - (1 - \Phi(x))^q - \Phi(x)^q) dx$$

Da als Lageparameter auch hier der Mittelwert aller Beobachtungen $\bar{x} = 1/m \sum_{i=1}^m x_i$ verwendet werden kann, werden die Kontrollgrenzen einer Urwertkarte gebildet durch

$$\text{UCL}_{\overline{mR}}/\text{LCL}_{\overline{mR}} = \bar{x} \pm 3 \frac{\overline{mR}}{d_q}$$

Für die in den folgenden Teilen dieses Kapitels beschriebenen Auswertungen bildet die Mittelwertkarte den Ausgangspunkt, im Weiteren wird die Eignung verschiedener Varianten untersucht. Die entsprechenden Vorgehensweisen werden dann jeweils in Kombination mit den Ergebnissen beschrieben.

Dabei ist von vornherein klar, dass es weder sinnvoll, noch realistisch ist, *alle* bekannten Typen von Kontrollkarten zu berücksichtigen. So könnten beispielsweise immer noch andere Schätzer für Lage oder Dispersion ausprobiert werden. Auch sehr bekannte Verfahren wie die CUSUM-Karte werden keine Berücksichtigung finden, so lange keine Eigenschaften der Verfahren eine Eignung vermuten lassen. Etwa beim Beispiel der CUSUM-Karte, welche die Abstände der Zeitreihe zum Zentrum kumuliert, wird die bestehende Autokorrelation dafür sorgen, dass auch große kumulierte Abstände als *normal* eingeordnet werden. Damit ist nicht anzunehmen, dass schnelle Änderungen der Frequenz auch schnell erkannt werden können.

Die Berechnungen für die Erstellung von Kontroll-, und insbesondere der Mittelwert-, Urwert- und EMWA-Karten, beruhen auf einer durch den Verfasser überarbeiteten und angepassten Version des R-Pakets `qcc` (vgl. dazu R Core Team, 2018, sowie Scrucca,

2 Problemstellung und Grundlagen

2004). Die Modifikationen beziehen sich dabei in keinem Fall auf inhaltliche Punkte, sondern stets auf die Art und Weise der Implementierung: Beispielsweise erscheint für die vorliegenden Daten wegen der Größe des hier verwendeten Datensatzes das erneute Ablegen des `data.frames` im `qcc`-Objekt nicht sinnvoll, die Einführung von parallelem Rechnen ist für größere Datensätze hingegen gewinnbringend. Nicht zuletzt wurde die `plot`-Funktion auch unter Berücksichtigung der Anzahl an Datenpunkten neu gestaltet.

3 Alarmsysteme mit Standard-Methodik

3.1 Vorgehen

Zur Erstellung eines Alarmsystems auf Basis der Frequenzdaten wurden die folgenden Verfahren auf ihre Eignung überprüft. Dabei handelt es sich um verschiedene Varianten der Kontrollkartentechnik, deren Kernideen im vorigen Teil erläutert wurden. Als erstes Qualitätsmerkmal wird der Anteil falscher Alarme in den Fokus gestellt. Da die ganze vorliegende Stichprobe als *unter Kontrolle* anzusehen ist, sollte dieser Anteil natürlich möglichst klein ausfallen. Ein Wert von exakt null ist dabei aber der Erfahrung nach nicht unbedingt von Nutzen, da derartige Grenzen meist mit sehr langen Zeiten bis zum Erkennen eines tatsächlichen Alarmfalls einhergehen.

Im CEO-Handbuch der UCTE wird von vier möglichen Zuständen des Netzes ausgegangen: Normal, Alert, Emergency und Blackout (ENTSO-E, 2009, S. P5-3; siehe auch Teil 2.1.3 dieser Arbeit). Wie bereits erwähnt wird diese Vierteilung im Folgenden auf eine Zweiteilung *Normal oder Alarm* reduziert. Zum einen soll eine Alarmierung ausdrücken, dass das Risiko besteht, dass ohne Eingriffe in das System die Energieübertragung unterbrochen wird und/oder dass das Netz Schaden nimmt. Dies fasst die Stufen Alert und Teile der Stufe Emergency zusammen. Der bereits eingetretene Fall eines Blackouts hingegen („Characterised by the almost or total absence of voltage in the transmission power system“, ebenda) muss nicht mehr durch ein Warnsystem angezeigt werden. Neben dieser inhaltlichen Überlegung für eine Reduzierung der Klassen besteht auch ein rein verfahrenstechnischer Grund auf Seiten der statistischen Methodik: Die Daten liegen nur für eine Klasse vor, nämlich für einen *normalen* Netzzustand. Gerade diese Tatsache führt auch zur Verwendung von Kontrollkarten statt anderer Klassifikationsverfahren. Diese können dann aber ohne Weiteres nur zwischen *normal* und *nicht normal* unterscheiden. Konzepte für Kontrollkarten mit einer Ampel nachempfundenen Warngrenzen bestehen zwar (vgl. bspw. Page, 1955), werden aber im Rahmen des hier vorgenommenen ersten Schrittes zu Erstellung eines Alarmsystems nicht berücksichtigt.

3.2 Ansätze mit klassischen Karten

3.2.1 Mittelwert-Karte

Den ersten Versuch zur Erstellung eines Alarmgebungs-Verfahrens bildet die einfache Anwendung einer Mittelwert-Karte. Die Abbildung 3.1 zeigt den Beginn der ersten Zeitreihe mit den wie beschrieben ermittelten Kontrollgrenzen. Dargestellt werden die ersten zwanzig Minuten, was $20 \cdot 60 \cdot 50 = 60\,000$ Messzeitpunkten entspricht. Durch eine Gruppenbildung der Größe $g = 5$ werde diese dann auf 12 000 Zeitpunkte reduziert. Zur Schätzung der Standardabweichung wird wie beschrieben die durchschnittliche Standardabweichung $\hat{\sigma}_{\bar{s}}$ verwendet.

Es werden hier Grenzen gezeigt, welche nur für den Ausschnitt der Abbildung berechnet wurden: Durch mögliche Überanpassungen dürfte dies die Methodik im Zweifel eher bevorteilen als die Ergebnisse verschlechtern. Mit den Bezeichnungen des Teils 2.4 und für eine Anwendung der 3σ -Regel berechnen sich die Grenzen bei einer Genauigkeit von hier fünf Dezimalstellen zu

$$\begin{aligned} \text{LCL} &= \bar{\bar{y}} - \frac{3}{a_g \sqrt{g}} \bar{s} \\ &= 50.00096 - \frac{3}{0.940 \sqrt{5}} 0.00024 \\ &= 50.00061 \end{aligned}$$

sowie zu

$$\text{UCL} = \bar{\bar{y}} + \frac{3}{a_g \sqrt{g}} \bar{s} = 50.00131$$

Für die zu erwartende Rate an falschen Alarmen wurde damit der gebräuchliche, aus der Verteilungsfunktion Φ der Normalverteilung abgeleitete Wert $2\Phi(-3) = 0.27\%$ gewählt. Real wäre dieser Wert bei den vorliegenden Daten durch die hohe Aufzeichnungsrate offenbar unpraktikabel groß, da 0.27% bedeuten, dass ungefähr bei jeder 370. Messung ein falscher Alarm zu erwarten ist. Bei einer Aufzeichnungsrate von 50 Hertz bedeutet das alle 7 bis 8 Sekunden eine falsche Alarmierung, nach der Bildung von Gruppen der Größe $g = 5$ im Durchschnitt alle 37 Sekunden. Dennoch wird hier und im Weiteren dieser Wert als Referenz verwendet, um Vergleiche mit der Literatur einfach zu halten. Außerdem sollte ein schlüssiges Verfahren ohnehin für verschiedene „Einstellungen“ dieser Rate funktionieren, unabhängig von einer inhaltlichen Nutzbarkeit.

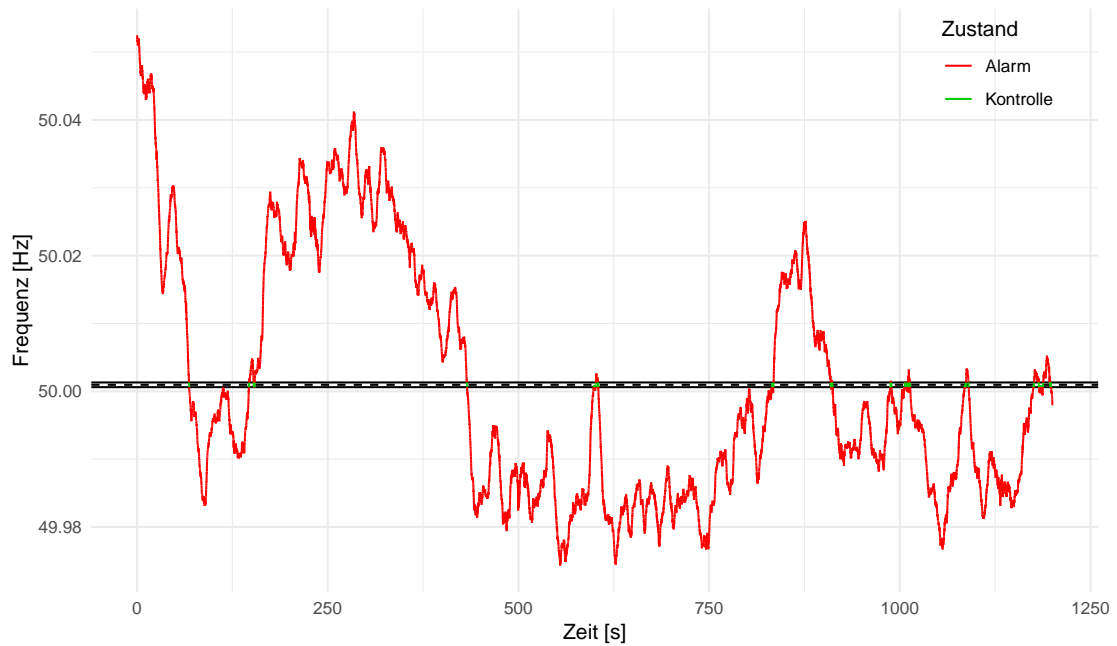


Abbildung 3.1: Wird eine \bar{x} -Karte mit der Gruppengröße 5 ohne weiteren Blick auf die Datenlage erstellt, wird der Systemzustand in der Regel als kritisch eingestuft; der Bereich, welcher den Prozess als unter Kontrolle einstuft, wird extrem klein. Die zwei durchgezogenen Linien zur Markierung der Kontrollgrenzen und die gestrichelte Linie für die mittlere Lage sind hier augenscheinlich kaum voneinander zu trennen. Die Verwendung dieser Karte würde zu einem dauerhaften Alarmzustand führen, nur unterbrochen von einigen Ausnahmen beim Durchlauf des Mittelwerts.

3 Alarmsysteme mit Standard-Methodik

Offensichtlich wird hier die große Mehrheit der Messungen als kritisch beurteilt: Mit diesem Vorgehen wird für über 98.7 Prozent der Messzeitpunkte Alarm gegeben, genauer werden 11 839 von 12 000 Zeitpunkten als kritisch klassifiziert. Der Vergleich mit der eingestellten Rate falscher Alarmer $2\Phi(-3) = 0.27\%$ wird nicht benötigt, um die Unbrauchbarkeit der Karte zu erkennen: Die Grenzen sind augenscheinlich viel zu eng gefasst. Damit ist davon auszugehen, dass die Streuung hier falsch geschätzt wird.

Ein genauerer Blick auf die Daten zeigt, dass die von der Theorie gestellten Annahmen auf mehrere Weisen zugleich verletzt sind: Zum einen sind die Daten nur annähernd normalverteilt, was sich etwa an Schiefe und Wölbung zeigt: Die Schiefe der über die Orte gemittelten Frequenz beträgt für den obigen Zeitausschnitt 0.72, die Wölbungen um 2.40 (vgl. dazu auch die Tabelle 2.1 in Teil 2.2). Zum anderen entstammen die Daten keinem rein stochastischen Prozess: Wie in Teil 2.1.2 zusammenfassend dargestellt, sind die Werte durch ein System produziert, welches durchgehend einer Überwachung und automatischen wie manuellen Eingriffen unterliegt.

Zudem sind die Datenpunkte hochgradig mit ihren Vorgängern korreliert, was zu stark verzerrten Streuungsschätzern führt. Die Autokorrelation der Mittelwertreihe des ersten Zeitabschnitts beispielsweise beträgt 0.99978, ist also nicht nur *groß*, sondern kann ohne Weiteres als *extrem* bezeichnet werden. In der einschlägigen Literatur werden bei Beispielen oder Simulationen zu Kontrollkarten für autokorrelierte Daten durchweg bereits deutlich kleinere Zahlen als „groß“ bezeichnet, an zwei Stellen beispielsweise 0.8 (vgl. den folgenden Teil). Die Probleme, welche sich durch Autokorrelationen beispielsweise bei der Bildung von Urwertkarten ergeben, untersuchen Maragah und Woodall (1992) systematisch. Zu den Auswirkungen der multiplen Verletzungen von Annahmen seien die Ergebnisse von Alwan und Roberts (1995) gegenübergestellt, welche die Folgen von „misplaced control limits“ erörtern und zudem feststellen, dass genau diese eher die Regel als die Ausnahme sind: „In a sample of 235 'expert' control chart applications [...], we found that over 85% of control charts displayed misplaced control limits“ (S. 270).

In der Abbildung 3.2 wird ein vergrößerter Ausschnitt der Abbildung 3.1 gezeigt. In diesem ist die Stärke der Autokorrelation besser zu erkennen: Über einen Zeitraum vom fünf Sekunden ändern sich die Werte, gemessen an der Spannweite des Ausschnitts von Abbildung 3.1, nur in sehr geringem Ausmaß.

Der oben verwendete Wert $g = 5$ wurde für diesen ersten Versuch als *üblicher* Wert gewählt. Offenbar können aber auch andere Gruppengrößen verwendet werden, zudem steht in Form der Urwertkarte ein verändertes Vorgehen zur Verfügung. Die Resultate verschiedener Karten sind in der Tabelle 3.1 zusammengefasst.

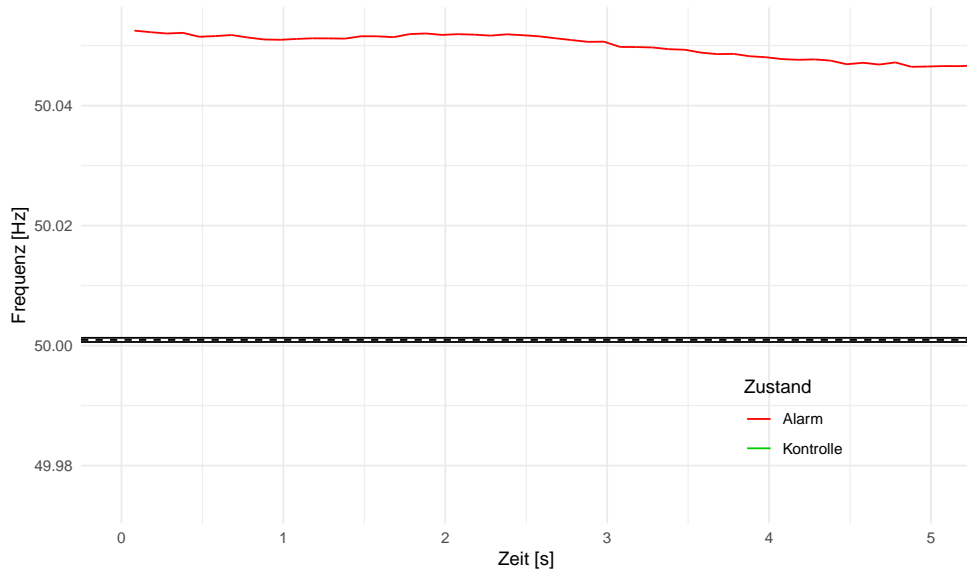


Abbildung 3.2: Werden kleinere zeitliche Ausschnitte betrachtet, so ändert sich die Lage der Karten-Statistik nur in geringem Ausmaß - es besteht eine sehr große Autokorrelation. Dargestellt sind hier die ersten fünf Sekunden der vorigen Grafik.

An den resultierenden Alarmraten (Spalte $\hat{\alpha}$) lässt sich leicht ablesen, dass das Vorgehen hier nicht zielführend ist: Auch bei recht großen Gruppen entstehen keine brauchbaren Mittelwertkarten. Dabei werden mit $g = 500$ Gruppen über 10 Sekunden gebildet, während in den Ausgangsdaten alle 0.02 Sekunden ein Messwert zur Verfügung steht.

Auffällig ist, dass die Rate falscher Alarmer mit der Gruppengröße zunächst noch steigt, während - ausgehend von über 98 Prozent - das Gegenteil zu erwarten wäre. Das kann vermutlich auf die großen Autokorrelationen zurückgeführt werden: Würde eine Mittelwertbildung für unabhängige Werte durchgeführt, würde die Streuung des Prozesses entsprechend reduziert, was in den Streuungsschätzern durch den Faktor $1/g$ berücksichtigt wird. Das ist hier jedoch nicht der Fall: Praktisch wird nur eine Glättung vorgenommen, ohne dass die Spannweite des Prozesses entscheidend verändert wird. Zudem bleibt die Autokorrelation auch für große Gruppen groß.

Auch Urwertkarten führen hier nicht zu einem besseren Verfahren. So wird selbst bei einer Fensterbreite von $q = m (= 60\,000)$ (was bereits einem globalen Varianzschätzer mit den oben kurz angeführten Nachteilen entspricht) noch über jeder zehnte Messzeitpunkt fälschlicherweise einem kritischen Systemstatus zugeordnet. Auch hier ist vor allem die extreme Autokorrelation anzuführen, welche bei Urwertkarten zudem nicht durch die

3 Alarmsysteme mit Standard-Methodik

g	q	Schätzer	$\hat{\sigma}_y$	$\hat{\alpha}$	$\hat{\rho}$
2	-	\bar{s}	0.00018	0.98520	0.99965
3	-	\bar{s}	0.00022	0.98510	0.99959
5	-	\bar{s}	0.00026	0.98658	0.99950
10	-	\bar{s}	0.00029	0.99000	0.99912
20	-	\bar{s}	0.00033	0.99133	0.99814
50	-	\bar{s}	0.00046	0.99250	0.99448
100	-	\bar{s}	0.00069	0.99333	0.98705
500	-	\bar{s}	0.00215	0.99167	0.91130
1	2	\overline{mR}	0.00019	0.97733	0.99979
1	3	\overline{mR}	0.00022	0.97412	0.99979
1	5	\overline{mR}	0.00025	0.97083	0.99979
1	10	\overline{mR}	0.00028	0.96785	0.99979
1	20	\overline{mR}	0.00032	0.96367	0.99979
1	50	\overline{mR}	0.00043	0.95228	0.99979
1	100	\overline{mR}	0.00058	0.93598	0.99979
1	500	\overline{mR}	0.00137	0.86832	0.99979
1	10 000	\overline{mR}	0.00540	0.42812	0.99979
1	60 000	\overline{mR}	0.00924	0.10233	0.99979

Tabelle 3.1: Mit verschiedenen Parameterkombinationen wurden Mittelwert- und Urwertkarten angepasst, und der Streuungsschätzer $\hat{\sigma}_y$, die Fehlalarmrate $\hat{\alpha}$ und der Autokorrelationsschätzer $\hat{\rho}$ festgehalten.

Parameterwahl beeinflusst werden kann: Mit der Breite q ändern sich nur die Kontrollgrenzen, nicht aber die Statistik einer Karte.

Im ersten Schritt wird nun eine Lösung der durch eben diese Autokorrelationen entstehenden Problematiken in den Fokus gesetzt. Mit welchem Verfahren hier am besten Abhilfe geschaffen werden kann ist nicht ohne Weiteres klar. Insgesamt wird dieses Problem in der Literatur mehrfach aufgenommen, aber eben bereits für deutlich kleinere Autokorrelationen. Zahlreiche Veröffentlichungen zu diesem Thema gibt es aus den frühen 1990ern. Es werden verschiedenste Ansätze präsentiert – ein *allgemein bestes* Werkzeug für autokorrelierte Daten scheint aber bisher nicht gefunden. Dazu kommt die hier extreme Form der Autokorrelation. Sowohl analytisch als auch simulativ gewonnene Resultate zeigen, dass die Wahl des Verfahrens zur Autokorrelationskorrektur von weiteren Dateneigenschaften abhängen kann.

3.2.2 Aufspreizen der Grenzen

Im Lehrbuch „Advanced Topics in Statistical Process Control“ widmet der Autor Donald J. Wheeler (1995) der Behandlung von Autokorrelationen bei der Erstellung von Kontrollkarten ein kurzes Kapitel und stellt fest:

„Time series with a large positive, lag-one autocorrelation (say in the neighborhood of 0.8 or larger) will have very *coherent* running records. [...] The large autocorrelation assures that any noise present in the values is relatively minor compared to the signals present.“ (S. 281, Hervorhebung im Original)

Hieraus lassen sich zwei Erkenntnisse ableiten. Zum einen werden bereits Autokorrelationen um 0.8 von Fachleuten als „groß“ bezeichnet, was die Einordnung der hier berechneten Werte bezüglich ihrer Auswirkungen als *extrem* rechtfertigt. Der zweite Satz deutet daraufhin, dass System-Störungen relativ einfach zu erkennen sein sollten. Diese werden hier aber erst betrachtet, wenn die Rate falscher Alarme dem eingestellten Wert zumindest im Wesentlichen entspricht, also sozusagen ebenfalls *unter Kontrolle* ist.

Als Lösung der Verzerrung durch Autokorrelationen wird vorgeschlagen, bei einer Urwertkarte die Abstände der Grenzen zum Zentrum um einen Faktor zu erweitern. Mittels einer Argumentation über Approximationen gelangt der Autor zu dem Vorschlag, dabei als Faktor

$$F_W = \frac{1}{\sqrt{1-r^2}}$$

zu wählen. Mit r wird die bekannte oder geschätzte Autokorrelation der Mittelwertreihe, das heißt der von der Karte betrachteten Größe, bezeichnet. Es ist dann

$$\text{UCL}_W/\text{LCL}_W = \hat{\mu} \pm c F_W \hat{\sigma}_{\overline{mR}} = \bar{x} \pm c \frac{\overline{mR}}{d_q \sqrt{1-r^2}}$$

Auch Gilbert et al. (1997) zeigen Möglichkeiten zum Umgang mit Autokorrelationen auf, auch hier werden bereits Werte „of 0.8 or larger“ als groß bezeichnet (S. 375) – und einer der dort aufgeführten Ansätze für größere Autokorrelationen auch direkt verworfen. Stattdessen vorgeschlagen wird von den Autoren ein auf der Basis von AR(1)-Prozessen abgeleiteter Faktor¹ zur Erweiterung der Grenzen von Urwertkarten mit $q = 2$ (S. 377 f.):

¹Bei Young und Winistrofer (2001) wird der Faktor leicht abweichend als $d_2(2\sqrt{1-r/\pi})^{-1}$ angegeben. Dies wird hier als Übertragungsfehler angesehen, insbesondere da diese Variante bei konkreter Anwendung unplausibel erscheint.

3 Alarmsysteme mit Standard-Methodik

$$F_G = \frac{d_2}{2\sqrt{\frac{1-r}{\pi}}}$$

Die Kontrollgrenzen sind dann folglich

$$UCL_G/LCL_G = \hat{\mu} \pm c F_G \hat{\sigma}_{mR} = \bar{x} \pm c \frac{\overline{mR}}{2} \sqrt{\frac{\pi}{1-r}}$$

Besteht keine Autokorrelation, so ist $r = 0$ und damit $F_W = F_G = 1$. Liegt keine Selbstabhängigkeit der Mittelwertreihe vor, führen beide Varianten also zu Karten wie oben vorgestellt.

Durch die Größe der vorliegenden Stichproben könnte auch in Betracht gezogen werden, als Kontrollgrenzen die empirischen Quantile der Mittelwertreihe zu verwenden, statt zunächst die Streuung der Daten zu schätzen. Dabei werden hier, orientiert an $\Phi(-3)$ und $\Phi(3)$, die 0.1350- und 99.8650-Prozent-Quantile verwendet. Die Anwendung dieser drei Varianten (Faktor nach Wheeler, Faktor nach Gilbert, Quantile) ist in der Abbildung 3.3 beispielhaft dargestellt.

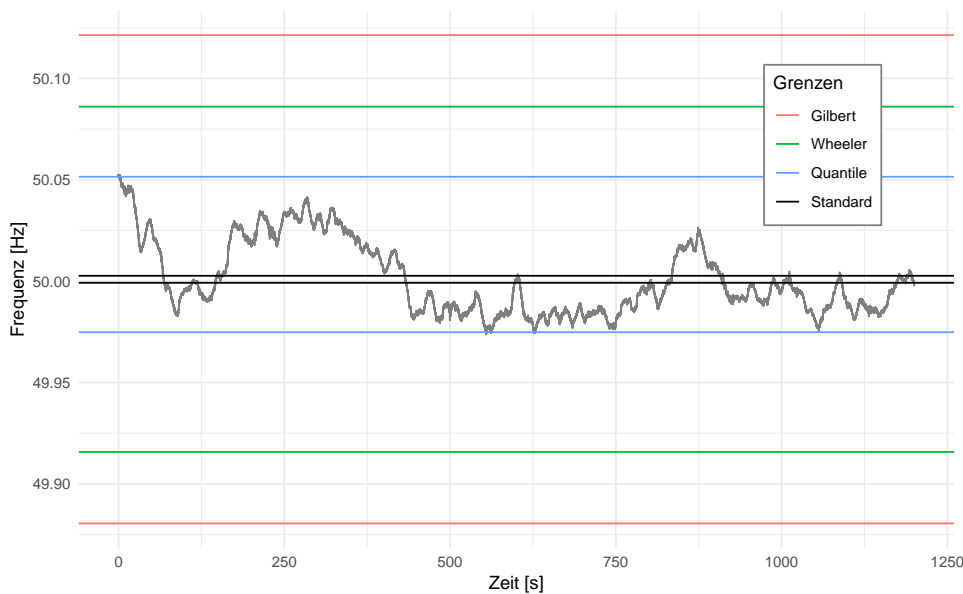


Abbildung 3.3: Mit den beiden angeführten Korrekturfaktoren oder bei Zuhilfenahme der Quantile ergeben sich die hier dargestellten Kontrollgrenzen.

Mit den mittels F_W und F_G aufgespreizten Grenzen ergibt sich ein zur einfachen \bar{x} -Karte gegenteiliges Bild: Statt einem Daueralarm entsteht hier eine durchgehende Einordnung des Systems als unter Kontrolle. Diese Klassifikation entspricht zwar der Realität,

fällt aber dennoch auf: Durch die Wahl der 3σ -Regel sind im betrachteten Ausschnitt 162 falsche Alarmer zu erwarten. Dass diese nicht auftreten ist bei alleiniger Berücksichtigung des Status unter Kontrolle offenkundig positiv zu bewerten. Ein Blick auf die Grafik zeigt jedoch, wie weit die Grenzen für beide Varianten und insbesondere nach Gilbert von den Extremwerten der Reihe entfernt liegen.

Die Raten fallen also deutlich kleiner aus als die oben genannten 0.27 %, was nicht positiv zu bewerten ist: Eine bei weitem nicht ausgenutzte Fehler-Toleranz führt im Allgemeinen dazu, dass auch die Erkennungsrate und die Erkennungsgeschwindigkeit von Störfällen wesentlich schlechter ausfallen als möglich. Fasst man die \bar{x} -Karte als ständig wiederholten Gauß-Test auf, so kann dieser Sachverhalt durch den bekannten trade off zwischen den Fehlern erster und zweiter Art verdeutlicht werden. Die Quantil-Variante gibt gemäß Definition und Parameterwahl 0.26998 Prozent falsche Alarmer aus.

Problematisch bei der Anwendung der Faktoren ist hier vermutlich in erster Linie, dass die vorliegende Autokorrelation nicht nur groß, sondern extrem ist: Beide Faktoren enthalten einen Term $\sqrt{1-r^2}$ beziehungsweise $\sqrt{1-r}$ im Nenner, womit F_W und F_G für ein r nahe der 1 sehr große Werte bei stark steigender Steigung annehmen. Die Faktoren reagieren daher in Bereichen großer Autokorrelationen recht sensibel auf leichte Abweichungen bei der Schätzung von r . Offenbar wirken sich diese Änderungen in starkem Maße auf die Qualität der Grenzen aus, welche in einem Fall ohnehin nur approximativ, und in beiden Fällen mit Blick auf bestimmte Prozesse, nämlich solche vom Typ AR(1), angegeben sind. Dass sich die vorliegenden Daten jedoch nicht durch die ARIMA-Klasse repräsentieren lassen, wird im Weiteren noch thematisiert. Schließlich werden die Faktoren auf beiden Seiten des Zentrums angewandt, so dass sich jeder Fehler doppelt auf die Größe des Akzeptanz-Bereichs auswirkt.

Auch mit den durch die Quantile ermittelten Grenzen wird die Problematik nicht gänzlich aufgelöst. Hier ist der Anteil falscher Alarmer naturgemäß *im Soll*, Abbildung 3.3 lässt aber wiederum erahnen, dass im Fehlerfall viel Zeit bis zur Alarmierung vergehen kann: Auch die Quantile berücksichtigen die zeitliche Abhängigkeit der Daten sowie deren aktuelle Lage keineswegs. Ein starker Anstieg, welcher auf niedrigem Frequenz-Niveau beginnt, würde nur sehr spät erkannt. Eine Verbesserung gegenüber Grenzen auf der Basis von Streuungsschätzern könnte aber gegeben sein und wäre zu untersuchen. Besser erscheint jedoch eine Verwendung von flexiblen Grenzen. Einige der im Folgenden dargestellten Methoden berücksichtigen dies (und eine genauere Analyse der Gestalt von Störfällen wird im Kapitel 6 folgen).

Die Verwendung von Quantilen wird in der im Rahmen dieser Arbeit gesichteten Literatur bisher nicht genannt. Das dürfte zum einen an der Einfachheit dieses Vorgehens

liegen. Zum anderen wird für eine Schätzung mit ausreichender Präzision eine vergleichsweise große Stichprobe benötigt, welche in klassischen Anwendungen wie der industriellen Produktion erst langwierig und teuer zu erzeugen wäre. Alternativ könnten aufwändigere statistische Verfahren zur Quantilsschätzung herangezogen werden, dann aber würde die Kontrollkartenmethodik mit ihrer Einfachheit einen wesentlichen Vorteil verlieren.

3.2.3 EWMA

Auch wenn die jeweiligen Validierungen und Beispiele keine *extremen* Autokorrelationen beinhalten, finden sich in der Literatur weitere Ansätze zur Behandlung der Problematik. In der bereits kurz angesprochenen Veröffentlichung von Young und Winistrofer (2001) wird beispielsweise die Möglichkeit angeführt, statt dem Mittelwert als Karten-Statistik dessen exponentiell gewichtetes gleitendes Mittel zu verwenden. Für autokorrelierte Daten untersuchen auch Lu und Reynolds (1999) dieses Vorgehen. Mit der englischsprachigen Abkürzung dieser *exponentially weighted moving averages* wird dabei von EWMA-Karten gesprochen. Statt der eigentlich beobachteten Größe X_t wie oben wird dabei die geglättete Zeitreihe

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1}$$

betrachtet, wobei $\lambda \in [0, 1]$ ist. Als optimal für den Glättungsparameter hat sich dem erstgenannten Aufsatz folgend $\lambda = 1 - \frac{1-r}{2}$ herausgestellt, wobei r wiederum der Autokorrelationskoeffizient ist. Im vorliegenden Fall ergibt sich also ein Parameter von $\lambda = 0.99989$, was einem sehr großen Gewicht für die aktuelle Beobachtung entspricht.

Wird eine EWMA-Karte mit der genannten Empfehlung für λ erstellt, so lässt sich der Verlauf der Statistik augenscheinlich nicht von der Abbildung 3.1 mit der Standard- \bar{x} -Karten unterscheiden: Durch den extrem hohen Einfluss der aktuellen Beobachtung auf die geglättete Reihe kommt der Vergangenheit kein nennenswerter Einfluss mehr zu. Die neue Zeitreihe besteht praktisch aus den gleichen Werten wie das Original, ergänzt um einen kleinen, wenn auch nicht zufälligen, Summanden. Entsprechend ist auch die Autokorrelation der geglätteten Zeitreihe extrem groß (0.99979), damit besteht das gleiche Problem wie für den Ausgangsdatensatz.

Eine EWMA-Variante wird auch bei Montgomery (2009) vorgeschlagen (vgl. Teil 10.4), wobei die dortigen Betrachtungen und Beispiele im Wesentlichen reine autoregressive Prozesse ersten Grades umfassen. Für die Wahl des Parameters λ wird der Wert vorgeschlagen, welcher die Fehlerquadratsumme der Ein-Schritt-Prognosen optimiert (S. 454; siehe dazu auch Montgomery und Mastrangelo, 1991, S. 183).

Für die vorliegenden Daten wurde dieser Wert mit einer verschachtelten Gittersuche ermittelt, wobei iterativ jeweils um den besten gefundenen Wert ein neues (eindimensionales) Gitter aufgespannt wurde. Der Optimalwert kann bei einer Genauigkeit von 10^{-10} nicht von der 1 unterschieden werden. Das bedeutet offenbar, dass der Historie gar kein Gewicht zukommen sollte – es ergibt sich genau die Zeitreihe, welche geglättet werden soll. Eine weitere Berücksichtigung dieser Variante zur Parameterwahl erübrigt sich.

Eigenschaften von Entscheidungskriterien der Art $|Z_t - z| > c\sqrt{V_0(Z_t)}$ untersuchen auch Schmid und Schöne (1997), wobei z das „Zentrum“ von Z_t sowie $c > 0$ ist. Mit V_0 wird die Varianz aus dem „in-control-process“, also der Phase 1, bezeichnet. Die Autoren zeigen, dass unter recht allgemeinen Voraussetzungen EWMA-Karten bei autokorrelierten Daten immer eine kleinere oder gleiche Rate falscher Alarme haben, als für unabhängige Daten ansonsten gleicher Eigenschaften (s. a. a. O., Theorem 1 und Folgerungen). Sie halten auf der Seite 1 279 weiter fest: „[...] if one falsely assumes that the underlying process is independent, the true ARL is underestimated.“ Mit der ARL ist dabei die durchschnittliche Lauflänge (average run length) gemeint, welche bei der Prüfung der Rate korrekt angezeigter Alarme noch thematisiert wird (vgl. Kapitel 6). Eine zu kurze ARL jedenfalls entspricht bei Systemen unter Kontrolle genau einer zu großen Anzahl falscher Alarme.

Verschiedene Methoden zur Behandlung von Autokorrelationen für die statistische Prozesskontrolle fassen Noffsinger und Anderson (2002) zusammen. Neben dem *Vorgehen nach Gilbert* und der Anpassung von Zeitreihenmodellen (s. u.) wird auch hier die Verwendung von EWMA-Karten in Betracht gezogen. Im Unterschied zum Vorgehen der oben genannten Quellen wird aber ein Glättungsparameter „in the neighborhood of, say 0.2 to 0.4“ (S. 63) vorgeschlagen – und damit in einem ganz anderen Bereich als $\lambda > 0.999$. Für Tests mit den Frequenzdaten wird hier der Wert 0.3 verwendet. Die Stärke der Glättung wird in der Abbildung 3.4 beispielhaft für die Daten aus Almelo und Bukarest, und für einen Ausschnitt von einer Minute verdeutlicht.

Würde die Reihe mit dieser Einstellung des Glättungsparameters betrachtet, so wird das überwiegende Gewicht auf die Historie gelegt, die Autokorrelation steigt in der Folge von 0.99978 auf 0.99998 an. Damit wird die Problematik also lediglich verschärft statt abgemildert. Die Autoren der letztgenannten Veröffentlichung erstellen ihre Kontrollkarte aber nicht auf der Basis der geglätteten Reihe, sondern führen einen weiteren Schritt ein: Statt der geglätteten Werte werden die *one-step-ahead prediction errors* (OSAPes) betrachtet. Mit der Vorhersage \hat{X}_t von X_t zum Zeitpunkt $t - 1$, notiert als $\hat{X}_t(t - 1)$, ist dabei

$$\text{OSAPE}_t = X_t - \hat{X}_t(t - 1) = X_t - Z_{t-1}$$

3 Alarmsysteme mit Standard-Methodik

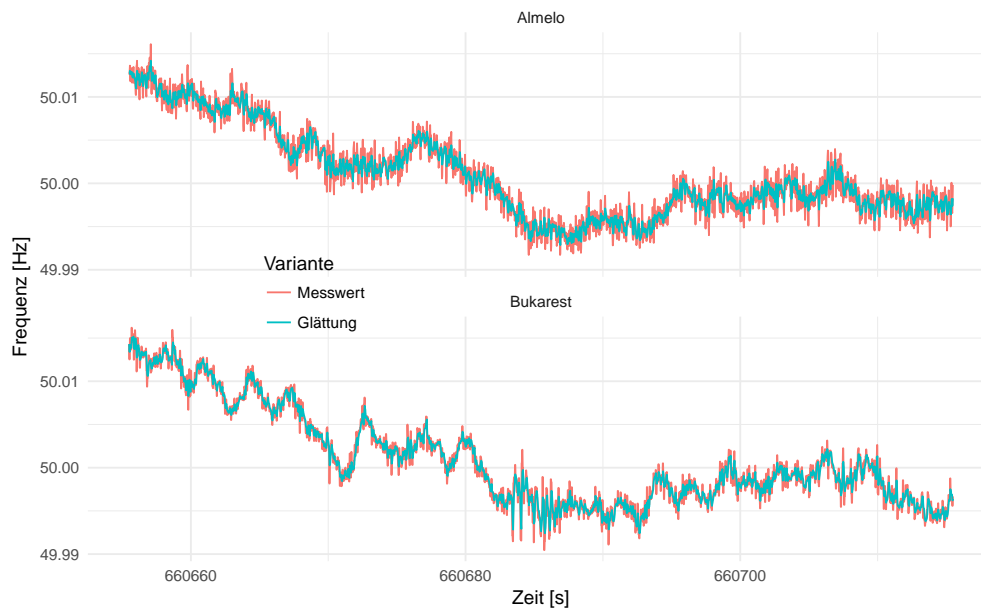


Abbildung 3.4: Glättung durch exponentiell gewichtete gleitende Mittel mit $\lambda = 0.3$. Für diese Verläufe kann nun eine Kontrollkarte erstellt werden, alternativ geschieht dies für die Einschnitt-Prognose-Fehler.

Für die Betrachtung der OSAPes wird also die Mittelwertreihe zunächst geglättet. Aus den Prognosefehlern wird dann eine „normale“ Urwertkarte erstellt. Es ist anzumerken, dass sich eine Karte für Fehler schon in der Konstruktion grundsätzlich von den bisher aufgeführten unterscheidet: Dabei wurde stets eine Statistik aus den Daten berechnet und auf dieser Basis eine konstante Kontrollgrenze für den gesamten Beobachtungszeitraum ermittelt, allenfalls beeinflusst durch das Fehlen von Werten. In diesem Fall wird nun eine konstante Grenze für den Fehler erstellt – im Bezug auf die Ausgangsgröße ist diese Grenze dann aber variabel: Statt einer Forderung der Art $X_t \leq UCL$ wird nun für $X_t \leq UCL(t)$ der Prozess als unter Kontrolle erkannt (und analog für die untere Grenze LCL). Allein diese Flexibilität lässt eine deutliche Verbesserung erwarten.

Das Ergebnis ist in der Abbildung 3.5 dargestellt. Die Vorgehensweise auf Grundlage der OSAPes erscheint als im Prinzip geeignet, dem Problem der extremen Autokorrelationen zu begegnen: Es ergibt sich eine Rate falscher Alarme von 4.09 Prozent. Im Vergleich zu der einfachen Karte sowie zu den Aufspreizungen der Grenzen ist dies mit Abstand der beste Wert. Auch die grafisch dargestellten Verläufe der Grenzen erscheinen nun praktikabel.

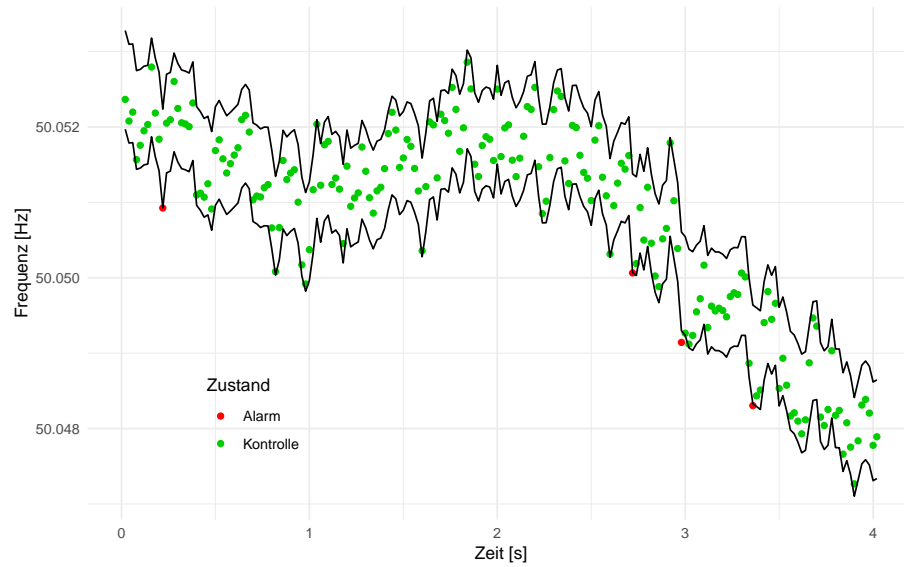


Abbildung 3.5: Im Gegensatz zu den bisherigen Karten sind die Grenzen im EWMA-Verfahren auch abhängig von der Lage der vorangegangenen Messwerte. Dargestellt ist eine Urwertkarte für die OSAPes, wobei die Ungleichung für die Betrachtung der Frequenzwerte umgestellt wurde.

Als *gut* kann das Verfahren aber dennoch nicht gelten, so ist die Alarmrate hier lediglich *relativ gut* im Vergleich mit den anderen bisher angewandten Varianten. Verglichen mit der für ein *absolut gut* funktionierendes Werkzeug zu erwartenden Rate von 0.27 Prozent ist der hier erzielte Wert immer noch um einen Faktor von über 15.1 zu groß. Für den realen Einsatz ist eine solche Abweichung nicht hinnehmbar.

Auch wenn die Daten vorher transformiert wurden, handelt es sich bei diesem Ergebnis um eine Standard- \bar{x} -Karte. Damit können nun weitere Methoden angewandt werden, welche der Autokorrelations-Problematik begegnen: Die geglättete Reihe der Fehler weist einen Autokorrelationskoeffizienten von 0.49 auf. Am einfachsten ist dann die Erweiterung der Grenzen durch die oben bereits ausprobierten Faktoren nach Wheeler und Gilbert. Die OSAPes sind hier zwar keinem AR(1)-Prozess entnommen, deren Definition $X_t - \lambda X_{t-1} - (1 - \lambda)Z_{t-2}$ ist aber zumindest im weiteren Sinne *verwandt* mit diesen. Die Faktoren F_W und F_G berechnen sich hier zu 1.14 und 1.39, und weiten die Grenzen damit nicht so stark auf wie oben. Der Anteil falscher Alarme kann durch die Wheeler-Variante auf 2.3 Prozent gedrückt werden, durch die Alternative nach Gilbert auf 0.96 Prozent. Hier ist nur noch ein Faktor von 3.6 gegenüber dem „Soll“ zu verzeichnen, die EWMA-Variante mit OSAPes und der Autokorrelations-Korrektur nach Gilbert folglich

3 Alarmsysteme mit Standard-Methodik

die bisher beste Variante. Es kann angenommen werden, dass die Verzerrungen durch die Autokorrelationen durch dieses mehrschrittige, kombinierte Vorgehen weitestgehend eliminiert sind. Der noch bestehende Abstand zwischen realisierter und eingestellter Rate falscher Alarme kann dann in den anderen angeführten Annahmeverletzungen vermutet werden, also in unpassenden Verteilungsannahmen, den Regelungen und der nur *groben Eignung* des Gilbert-Faktors.

Im Rahmen der vorgestellten EWMA-Ansätze führt also die Kombination aus OSAPes mit einem eher kleinen Parameter zur besten (und kleinsten) Rate an Zeitpunkten mit falscher Klassifikation – abgesehen von einer Quantilsvariante, welche die Rate per definitionem stets einhalten würde. Wiederum als Ausschnitt vom Beginn der ersten Zeitreihe werden die Grenzen für die Prognose-Fehler in Abbildung 3.6 gezeigt. Hier sind die Verläufe der Grenzen gut voneinander zu trennen.

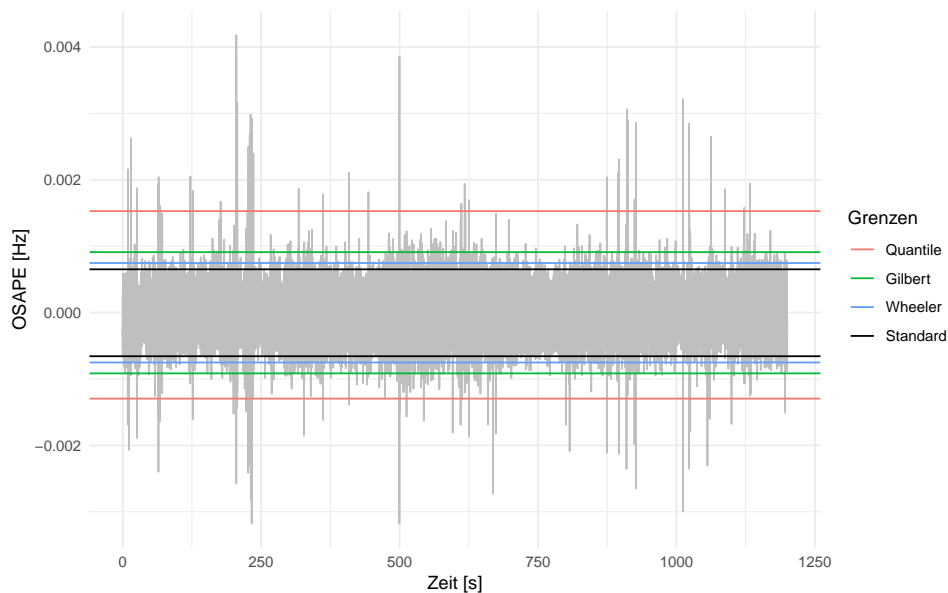


Abbildung 3.6: Zu den besten der bisher erreichten Resultate führt eine Kombination von Vorgehensweisen, nämlich der beiden von Noffsinger und Anderson (2002) und Gilbert et al. (1997). Die Abbildung zeigt den gleichen zeitlichen Rahmen wie die obige Abbildung 3.3 auf der Seite 42.

3.2.4 Subsample Charts

Eine weitere Variante zur Behandlung der beobachteten Problematik besteht in der Verwendung von sogenannten Subsample Charts. Diesen liegt die Idee zugrunde, die vorliegenden Daten auf systematische Weise *auszudünnen*, und somit einen Informationsverlust gegen eine bessere Handhabung der Daten einzutauschen. Es lassen sich dazu zwei Werkzeuge festhalten, die einzeln oder in Kombination angewandt werden können. Dazu stehe eine Zeitreihe x_i der Länge m zur Verfügung, und sei $r \in \mathbb{N}_{>0}$ sowie $s \in \mathbb{N}_{\geq 0}$. Dann können vor der Erstellung einer Urwertkarte angewandt werden:

1. Überspringen von s Beobachtungen zur Senkung der Autokorrelation
2. Zusammenfassen von r aufeinanderfolgenden Beobachtungen durch Mittelwertbildung und Anpassung der Varianzschätzung
3. Die Kombination von 1. und 2.

Zusammenfassen lässt sich dieses Vorgehen durch

$$y_i = \frac{1}{r} \sum_{j=(i-1)(r+s)+1}^{(i-1)(r+s)+r} x_j$$

Diese Vorgehensweise findet sich bei Alwan und Radson (1992) unter dem Stichwort *Subsample Means*, sowie bei Runger und Willemain (1996) unter dem Titel *Batch-means control charts for autocorrelated data*. Die beiden erstgenannten Autoren geben dabei unter Berücksichtigung der Varianz σ_x^2 und der Autokorrelation ρ_x von X_i die Varianz von Y_i an, sofern x_i einem AR(1)-Prozess folgt. Der Unterschied der oben als Variante 2 notierten Idee gegenüber einer gewöhnlichen Mittelwertkarte liegt damit im Streuungsschätzer. Für diesen kann dann verwendet werden:

$$V(Y_i) = \frac{\sigma_x^2}{r} + \frac{2\sigma_x^2 \rho_x^{r+1} - r\rho_x^2 + (r-1)\rho_x}{r^2 (\rho_x - 1)^2}$$

Dieser Term ist insbesondere unabhängig von s , zudem gilt für $\rho_x = 0$ wie bei der gewöhnlichen Mittelwertkarte $V(Y_i) = \sigma_x^2/r$. Ist $r = 1$, ist ebenfalls $V(Y_i) = \sigma_x^2/r (= \sigma_x^2)$.

Den durch autokorrelierte Daten induzierten Problemen wird hier mit einer Methode begegnet, welche offenbar einen Verlust an Information mit sich bringt. Runger und Willemain (1996) betonen dazu: „[T]he rational subgrouping principle [...] should be interpreted in the context of a data-rich environment, wherein clock time, rather than the number of observations, is most important“ (Seite 484).

3 Alarmsysteme mit Standard-Methodik

Die *clock time* ist auch bei der hier vorliegenden Problemstellung ein zentraler Punkt: Wichtig sind in der finalen Begutachtung eines Verfahrens allein die Zeiten zwischen zwei falschen Alarmen, sowie zwischen einem Fehlerfall und der darauf folgenden korrekten Alarmierung. Diese Zeiten können unabhängig von der Messrate beurteilt werden. Anders ist dies beispielsweise bei der industriellen Produktion teurer Produkte wie Automobilen: Dort sind fehlerhaft produzierte Produkte die Kosten verursachende Größe, weniger und nicht allein die verstrichene (Arbeits-)Zeit.

Bei der Anwendung der Strategien auf die Frequenzdaten wird nun wie folgt vorgegangen: Für eine Parameterkombinationen von r und s wird die Mittelwertzeitreihe entsprechend transformiert. Anschließend wird eine Urwertkarte mit $q = 2$ angepasst. Schätzungen für σ_x und ρ_x werden aus den untransformierten Daten gewonnen, wobei für die Streuung die Variante $\hat{\sigma}_{mR}$ verwendet wird. Parallel zu diesem Vorgehen wird eine Karte *ohne Korrektur* erstellt. Dabei werden die transformierten Werte aufgefasst wie Messwerte. Das ist insbesondere für den Einfluss von s , also der Anzahl an übersprungenen Messzeitpunkten, interessant.

Die Abbildung 3.7 zeigt die Resultate dieses Vorgehens. Dazu wurde hier ein Abschnitt mit einer Länge von fünf Stunden verwendet, damit ausreichend Daten für die Zusammenfassungen mit großen Parametern zur Verfügung stehen. Zu erkennen ist, dass die Annahmeverletzungen (etwa gegenüber der Voraussetzung eines AR(1)-Prozesses) hier in einer Größenordnung sind, welche durch das Verfahren nicht mehr toleriert werden können. Die Varianzkorrektur führt nur für sehr große r zu allenfalls akzeptablen Alarmraten, wobei etwa ab $r \geq 600$ eine Überkorrektur stattfindet und die Alarmrate auf Null absinkt: Die Grenzen werden zu weit auseinander gebildet, um eine sinnvolle Alarmierungsstrategie bereitzustellen.

Auch mit Sprüngen von zwei Sekunden ($s = 100$) zwischen zwei Beobachtungen ist die Autokorrelation des Prozesses noch sehr groß, so dass auch das alleinige Ausdünnen der Beobachtungen nicht zu besseren Werten führt. Für größere Wert ist hier zwar eine Verbesserung zu erwarten, es ergibt sich dann aber sicher eine zu große Ausdünnung der Daten – sowohl beurteilt über die Anzahl Beobachtungen, als auch über die *clock time*. Auch das Betrachten der r - s -Kombinationen für Kontrollkarten ohne Varianzkorrektur ist schließlich nicht erfolgreich, wie die Grafik oben links in der Abbildung zeigt.

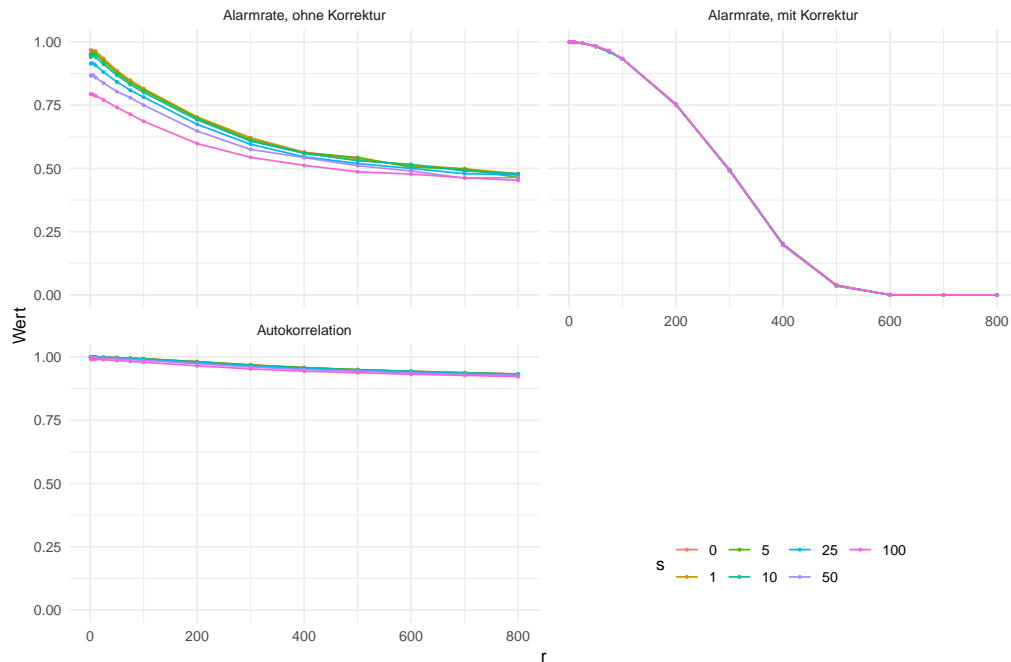


Abbildung 3.7: Angewandt auf die Frequenzdaten führen die Subsample Charts zwar zu Verbesserungen, brauchbare Resultate können aber nicht erzielt werden.

3.2.5 T²-Karten

Statt den Fokus nur auf die Reihe der Mittelwerte der Orts-Frequenzen zu legen, kann auch die Information der gesamten Datenbasis verwendet werden. Dann sind auch die *Quer*-Abhängigkeiten der Datenmatrix zu berücksichtigen, und damit die Abhängigkeiten zwischen jeweils zwei der einzelnen (Orts-)Zeitreihen.

Der Blick auf die entsprechenden Korrelationsschätzer zeigt auch in Zahlen gefasst, dass die Reihen unabhängig von der zeitlichen Komponente stark korreliert sind. Das ist zu erwarten, da alle Messpunkte an das gleiche Stromnetz angebunden sind und somit im Idealfall stets die gleiche zugrunde liegende Größe beurteilt wird (s. o., Teil 2.1). Außerdem wirken sich sämtliche Ereignisse innerhalb des Netzes auch direkt auf alle angeschlossenen Elemente aus, unabhängig davon, ob der Ort des Ereignisses hier aufgenommen wurde oder nicht. Die Entfernung der Netzelemente und die Netzarchitektur bestimmen dabei die Stärke und den zeitlichen Versatz dieser Auswirkungen. In der Tabelle 3.2 sind die Schätzer für die Korrelationen zwischen den fünf Messreihen aufgeführt.

Methoden, die diese Abhängigkeiten per Konstruktion berücksichtigen bestehen in multivariaten Kontrollkarten. Deren bekannteste Vertreterin stellt wohl die T^2 -Karte dar, wie

	Almelo	Bukarest	Dortmund	Ljubljana
Bukarest	0.9933			
Dortmund	0.9969	0.9933		
Ljubljana	0.9949	0.9948	0.9949	
Magdeburg	0.9980	0.9952	0.9980	0.9965

Tabelle 3.2: Die Korrelationskoeffizienten nach Pearson und für paarweise komplette Beobachtungen zeigen an, dass zwischen den einzelnen Zeitreihen sehr große Abhängigkeiten bestehen. Das ist einerseits technisch bedingt, führt aber andererseits zu weiteren Problemen bei der Erstellung von Kontrollkarten.

sie etwa bei Wheeler (1995) vorgestellt wird (S. 334 ff.). Mit dieser Karte wird bei der Konstruktion der verwendeten Statistik auch die Kovarianzmatrix der Daten berücksichtigt. Folglich wird auch die Abhängigkeit der Reihen untereinander aufgenommen. Mit Schätzern für den Mittelwertvektor $\hat{\mu}$ und die Kovarianzmatrix $\hat{\Sigma}$, sowie einem Vektor der Messwerte eines Zeitpunkts t , $\vec{x}'_t = (x_{t1}, x_{t2}, \dots, x_{tm})'$, ist die Statistik der Karte

$$T^2 = (\vec{x}_t - \hat{\mu})' \hat{\Sigma}^{-1} (\vec{x}_t - \hat{\mu})$$

Die weitere Konstruktion der Karte beruht dann auf der F -Verteilung von T^2 , welche für passende Schätzverfahren und multivariat normalverteilte \vec{x}_t gegeben ist. Bei Wheeler (1995) findet sich jedoch der Hinweis, dass diese Methode sehr sensibel gegenüber schiefen Datenstrukturen ist (S. 339). Weiter merkt der Autor auf der gleichen Seite an: „[The T^2 Chart] only works for a narrow set of conditons, and it will be virtually impossible to ever verify if these conditions are completely satisfied.“

Für unabhängige Größen ergibt sich eine Kovarianzmatrix, welche außer auf der Diagonalen nur Nullen enthält. Die Statistik ist dann die Summe aus quadrierten standardisierten Größen. Bei Normalverteilung der Ausgangsgrößen folgt die Konvergenz der Summe gegen eine Chi-Quadrat-Verteilung. Auch mit diesen Erkenntnissen kann eine Kontrollkarte erstellt werden. Die Folgerung der Chi-Quadrat-Verteilung gilt auch, wenn μ und Σ bekannt sind (vgl. dazu auch Weihs und Jessenberger (1999), Teil 5.7.1).

Um die Vergleichbarkeit mit den anderen hier vorgestellten Methoden herzustellen, wird die Wahrscheinlichkeitsmasse des *in-control*-Bereichs hier ebenfalls auf $1 - 2\Phi(-3) = 0.99730$ eingestellt. Im Resultat erkennt die T^2 -Karte 98.49 Prozent der Messzeitpunkte korrekt als „unter Kontrolle“, und sortiert damit 1.51 Prozent fälschlicherweise dem Zustand „kritisch“ zu. Im Vergleich mit andern Verfahren ist dieser Wert einer der besseren,

dennoch wird der Zielwert um den Faktor 5.6 übertroffen. Die Abbildung 3.8 zeigt die T^2 -Karte in Gegenüberstellung der Messwerte selbst.

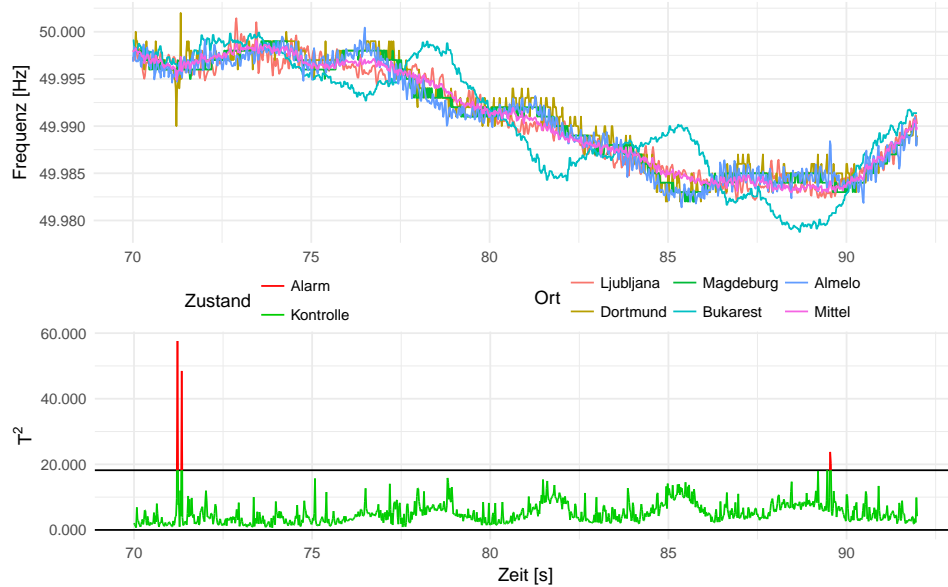


Abbildung 3.8: Die T^2 -Karte verwendet Quadrate und damit ausschließlich positive Werte. Folglich wird nur eine obere Kontrollgrenze betrachtet. Der in diesem Abschnitt von den anderen Orten leicht abweichende Verlauf der Bukarester Werte führt hier nicht zu einem Alarm.

Wie exemplarisch zu erkennen ist, werden zwei Klassen von Alarmen ausgelöst. Als eine Gruppe, im Beispiel bei knapp 90 Sekunden, werden größere, aber augenscheinlich noch normale Schwankungen als Fehlerfall erkannt. Diese Art von Fehlalarmen ist beim Ausnutzen von stochastischen Eigenschaften im Vorhinein berücksichtigt und wird per Festlegung einer zulässigen Rate falscher Alarme in Kauf genommen. Schwieriger ist jedoch die zweite Variante: Durch ein lokales Ereignis in Form eines Fahrplanwechsel-Musters in Dortmund wird ein Alarm ausgelöst. Zumindest im Fall des gezeigten Ausschnitts fällt das Muster aber nicht sonderlich stark aus, dieses sollte als *normal* erkannt werden können.

Zu welchen Anteilen die beiden Fehler-Gruppen insgesamt auftreten kann nicht geklärt werden, ohne dass eine passende automatisierte Erkennung von Fahrplanwechseln vorliegt. Insgesamt erscheint die zweite Gruppe falscher Alarme aber als die einzige Schwäche der T^2 -Karte. Sollte die Möglichkeit bestehen, Fahrplanwechsel anhand der Muster zu erkennen oder dieses Wissen anderweitig einfließen zu lassen, wäre eine erneute Anpassung

mit diesen Karten in jedem Fall ein lohnenswerter Versuch. Dies ist hier jedoch nicht der Fall.

3.3 Ansätze aus der Zeitreihenanalyse

3.3.1 Zur Idee

Zahlreiche in der Literatur vorgeschlagene Verfahren zur Erstellung von Kontrollkarten bei autokorrelierten Daten setzen bestimmte weitere Eigenschaften der Daten voraus. Diese Eigenschaften bestehen oft darin, dass den Daten ein bekanntes, vergleichsweise einfach zu fassendes Zeitreihenmodell zugrunde liegt oder die Daten durch ein solches angemessen repräsentiert werden können. Diese Annahme wird dann als Basis für die Entwicklung von Lösungsvorschlägen verwendet. Auch werden Verfahren zur Behandlung von Selbstabhängigkeiten beurteilt, indem sie für AR- oder ARIMA-Prozesse überprüft werden (bspw. bei Noffsinger und Anderson, 2002).

Wichtige Beiträge zur Behandlung von Autokorrelationen stammen von Layth C. Alwan, heute tätig an der University of Wisconsin–Milwaukee. In einer Veröffentlichung aus dem Jahr 1991 bekräftigt er seine „konservative“ Annahme aus dem Vorjahr, nach der „mindestens 70 %“ aller Daten, die einer statistischen Prozesskontrolle unterzogen werden, mit Autokorrelationen behaftet sind (Alwan, 1991, S. 176). Als Antwort auf diese Problematik wird unter anderem die Variante flexibler Kontrollgrenzen in den Mittelpunkt gestellt.

Flexible Kontrollgrenzen wurden bereits bei der Erstellung von Karten für Ein-Schritt-Prognosen verwendet (vgl. Teil 3.2.3). Wie beschrieben werden sie angefertigt durch die Erstellung von Prognosen für die jeweils folgenden Messwerte und die anschließende Betrachtung der Vorhersagefehler (ebd., S. 185). Wendet man auf diese eine Kontrollkarte an, dann ergeben sich feste Kontrollgrenzen CL_ε für die Fehler $\varepsilon_t = \hat{x}_t - x_t$. Die Grenzen für den Verlauf der Messwerte bilden sich dann durch $CL_\varepsilon + \hat{x}_t$ und sind damit vom Verlauf der Zeitreihe abhängig.

Eine ausführliche Darstellung und Besprechung der Idee, statt der vorliegenden Daten die Residuen von Zeitreihenmodellen für eine Kontrollkarte zu verwenden, stammt von Layth Alwan und Harry Roberts (Alwan und Roberts, 1988). Auch wenn der genannte Artikel Verweise auf zwei kurze, vorherige Beschreibungen dieses Vorgehens enthält, kann er als absolute Grundlagenarbeit angesehen werden. Insbesondere ordnen die Autoren das Vorgehen in den sachlichen Kontext und die „Denkweise“ der statistischen Prozesskontrolle ein (vgl. S. 88 f.).

Einer der beiden genannten Verweise zielt auf Montgomery (1985), die erste Auflage der bereits oben in der sechsten Auflage zitierten *Introduction to Statistical Quality Control*. Dort wird nur kurz die Idee skizziert, in Kontrollkarten Modell-Residuen statt der eigentlichen Daten zu verwenden. Dabei werden (auto-)korrelierte Daten und ein AR(1)-Prozess als Problemstellung genannt (S. 265). In der sechs Jahre jüngeren und insgesamt wesentlich erweiterten zweiten Auflage aus dem Jahr 1991 nehmen solche Karten dann einen deutlich größeren Stellenwert ein (Montgomery, 1991, S. 341 ff.; in den neueren Ausgaben Montgomery, 2009, und Montgomery, 2013, siehe die Seiten 450 ff. beziehungsweise 465 ff.). Die historische Verbreitung der Vorgehensweise lässt sich damit in etwa einordnen.

Die von Alwan und Roberts dargelegte Idee beruht im Wesentlichen auf dem Ansatz, einen Prozess in seiner Zerlegung „Actual = Fitted + Residual“, beziehungsweise bei Prognosen „Actual = Predicted + Error“, zu analysieren (S. 90). Es sollen beide Komponenten betrachtet werden: Die durch das Modell angepassten Werte („Fitted“) können „als Serie von Punktschätzern des bedingten Mittelwerts eines Prozesses“ (ebenda) grafisch analysiert werden. Dabei soll kein Alarmsystem entstehen, sondern ein genaueres Verständnis des Prozesses ermöglicht werden. Die Residuen hingegen sind als Eingangsgrößen einer gewöhnlichen Kontrollkarte zu verwenden.

Der wesentliche Punkt für die sinnvolle Anwendung dieser Methode besteht damit in der Erstellung der Prognosen. Welche Modellvarianten aus der Zeitreihenanalyse zu verwenden sind, ergibt sich offenbar aus den jeweils vorliegenden Datenstrukturen. Von Alwan und Roberts wird die *Box-Jenkins-Methode*, und damit implizit die Verwendung von ARIMA-Modellen vorgeschlagen. Diese stellen sie als ein nützliches Werkzeug vor („a useful set of tools“, ebenda).

Dass sich ARIMA-Modelle an vielen Stellen als nützlich erwiesen haben, muss an dieser Stelle nicht weiter ausgeführt werden, bei autokorrelierten Daten sollten sie stets in Betracht gezogen werden. Weniger weit verbreitet und auch weniger gut untersucht dürften bei der Veröffentlichung des genannten Artikels im Jahr 1988 die erst sechs² beziehungsweise zwei³ Jahre zuvor veröffentlichten ARCH- und GARCH-Modelle gewesen sein. Beide gehören heute zum Standardwerkzeug von Zeitreihenanalytikern, und berücksichtigen heteroskedastische Elemente von Zeitreihen. Solche erscheinen hier nicht unplausibel, so sind bei den Netzfrequenzen wechselnde Phasen mit geringer und mit hoher Volatilität denkbar, gerade da die vorliegenden Werte Eingriffe durch Regelungstechnik enthalten

²Vgl. Engle, 1982

³Vgl. Bollerslev, 1986

Die Frequenzdaten werden nun einer Zeitreihenanalyse unterzogen, die Residuen anschließend für die Erstellung eines Alarmierungssystems verwendet. Wie auch bei der Auswahl der bereits angewandten Kontrollkarten ist dabei zu berücksichtigen, dass ein *vollständiges* Ausschöpfen aller bereitstehenden Verfahren weder realistisch ist, noch angestrebt wird. Zeitreihenmodellierungen können zudem sehr weitreichende und umfassende Analysen bedeuten. Diese werden hier nur verkürzt wiedergegeben, um den Schwerpunkt der Auswertungen nicht allzu weit von der Erstellung des Alarmsystems als eigentlichem Ziel zu entfernen.

Für die Anpassung von Modellen sind einige Eigenschaften der Daten von Interesse, so etwa die Autokorrelationsfunktion. Eine erste entsprechende Betrachtung wurde bereits im obigen Teil 2.2 ab Seite 21 vorgenommen.

3.3.2 ARIMA

Es wird zunächst überprüft, wie gut sich die Frequenzverläufe durch ARIMA-Prozesse modellieren lassen. Die Charakteristiken von ARIMA-Modellen wurden bereits im Teil 2.3 beschrieben und erscheinen hier inhaltlich zunächst nicht unplausibel. So sind AR- und MA-Effekte sowohl bei den Ausgangsreihen, als auch bei Reihen der Differenzen gut vorstellbar: Unterstellt wird, dass sich der aktuelle Wert (oder die aktuelle Änderung) aus den vorangegangenen Frequenzen und aus dem vorangegangenen Rauschen ergibt. Gerade bei der großen Auflösung von 50 Messungen pro Sekunde ist dies zu erwarten.

Allein mit der sogenannten Box-Jenkins-Methode⁴ zur Bestimmung der Ordnungen lassen sich hier keine klaren Modelle ableiten (vgl. Abbildung 2.4 auf der Seite 25). Auch Periodizitäten können für Saisonlängen von einer Minute, einer Stunde und einem Tag nicht ausgemacht werden. Klar wird aber, dass sich durch eine (und genau eine) Differenzierung der Datenreihen das Problem der Autokorrelation deutlich verringert: Die Autokorrelation der Mittelwertreihe sinkt deutlich, die Korrelation zwischen jeweils zwei Datenreihen verschwindet praktisch ganz (der maximale absolute Korrelationskoeffizient beträgt 0.0062, betreffend das Paar Dortmund und Almelo). Auffällig ist, dass neun der zehn Korrelationsschätzer positiv sind. Dies könnte ein Hinweis darauf sein, dass die Zusammenhänge numerisch klein, aber durchaus vorhanden sind.

Die Abbildung 3.9 zeigt die Autokorrelation der Differenzen für die Mittelwertreihe, wobei die Differenz *nach* dem Zusammenfassen der fünf Orte gebildet wurde. Die Rei-

⁴Mit der Box-Jenkins-Methode, benannt nach George E. P. Box and Gwilym M. Jenkins, können in halbwegs übersichtlichen Fällen die Ordnungen p , d und q zur Erstellung eines ARIMA-Modells ermittelt werden. Dabei orientiert sich der Anwender oder die Anwenderin an für bestimmte Ordnungen *typische* grafische Verläufe von Autokorrelations- und partieller Autokorrelationsfunktion. Genaueres dazu beschreiben beispielsweise Judge et al. (1985) ab der Seite 240.

henfolge der Schritte wirkt sich im Allgemeinen nicht aus, bei fehlenden Werten ergeben sich aber geringe Unterschiede.

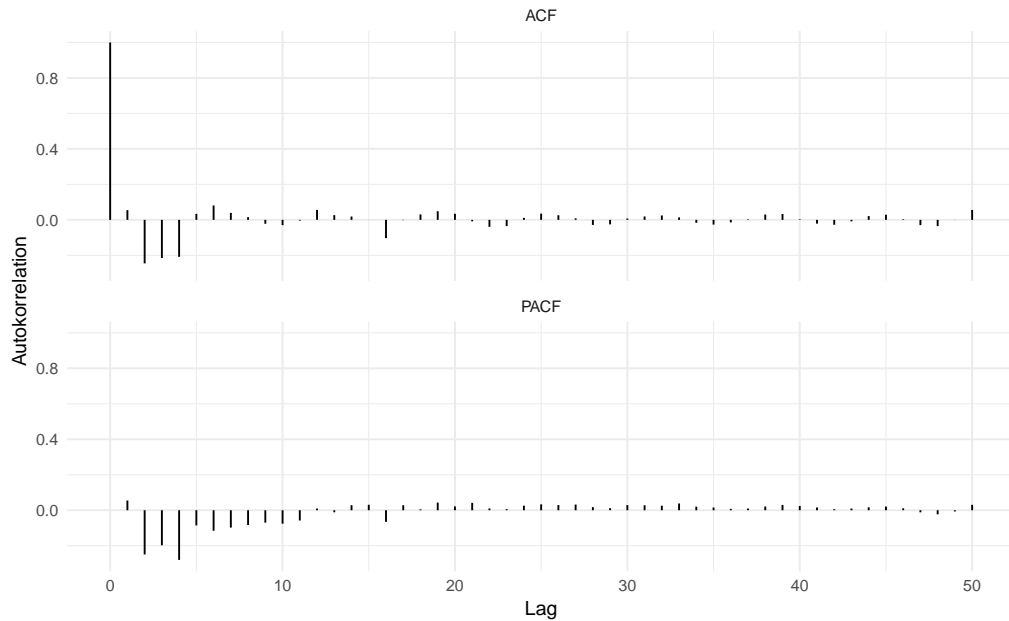


Abbildung 3.9: Die Autokorrelationen (ACF) der Differenzen sowie die partielle Variante (PACF) als Funktionen vom Lag. Die Werte fallen deutlich kleiner aus als für die undifferenzierten Werte (Abbildung 2.4).

Auch die Autokorrelationen der Differenzen unterscheiden sich bezüglich der Größenordnungen der Schätzer deutlich von den vorherigen Ergebnissen für die Messwerte selbst. Außerdem weist der Verlauf der Autokorrelationen der Differenzen ein alternierendes Muster auf. „Gedämpfte Sinus-Schwingungen“ könnten etwa auf einen $ARMA(2,0)$ -Prozess hindeuten, wie Hartung et al. (2005) erklären (S. 687).

Es zeigt sich eine absolut schnell abklingende Autokorrelation, was auch für die partielle Variante gilt. Das deutet daraufhin, dass die Reihe der Differenzen nicht wiederum integriert ist und somit keine weitere Differenzenbildung vonnöten ist. Die nach der zweiten Differenzenbildung resultierenden ACFs zeigen passend dazu auch keine weiteren qualitativen Änderungen. Die zugehörigen partiellen Autokorrelationen hingegen sind für kleine Lags stark negativ und nähern sich dann schnell der Null an. Dieses PACF-Muster kann insbesondere erzeugt werden, indem die Differenzen von weißem Rauschen differenziert werden, also *zu viele* Differenzen gebildet werden. Eine Darstellung der beiden angesprochenen Verläufe für die zweiten Differenzen ist als Abbildung 8.1 auf der Seite 171 im Anhang zu finden.

Die grafischen Betrachtungen führen also zu zwei Erkenntnissen: Zum einen erscheint die einmalige Bildung von Differenzen sinnvoll. Zum anderen zeigen sich aber auch dann keine offensichtlichen Muster von reinen AR- oder MA-Prozessen. Auszuschließen sind ARMA-Modelle hier aber dennoch nicht, da insbesondere Mischprozesse grafisch oft nicht durch einfache Muster aufzuklären sind.

Eine automatisierte Modellwahl kann unter anderem auf Grundlage des BIC durchgeführt werden (vgl. Hartung et al., 2005, S. 687). Für eine mehrfache wiederholte Anpassung von Modellen der Art ARIMA($p,1,q$) werden 50 zufällig ausgewählte Datenabschnitte über jeweils 30 Minuten verwendet. Dies entspricht jeweils 90 000 Beobachtungen. Stammt der gesamte Prozess aus einem bestimmten ARIMA-Prozess, so sollte das auch für alle Teilstücke gelten, die Schätzer entsprechend um den „wahren“ Wert streuen. Die Parameter p und q werden für jeden Abschnitt durch ein schrittweises Verfahren optimiert, wobei in jeder Iteration i die direkten Nachbarn der aktuellen Position (p_i, q_i) in Betracht gezogen werden. Näheres beschreiben Hyndman und Khandakar (2008). Wegen des oben genannten Hinweises auf $p = 2$ wird hier der Startwert gesetzt. Die AR- und MA-Parameter p und q werden jeweils nach oben durch 10 begrenzt, der I-Parameter wird fix auf $d = 1$ gesetzt.

Die folgende Abbildung 3.10 zeigt die Ergebnisse dieser Optimierungen. Es bildet sich kein klares Zentrum heraus, um welches sich die Ordnungsschätzer verteilen. Das wäre jedoch zu erwarten, wenn allen Abschnitten der gleiche ARIMA-Prozess zugrunde läge: Die Anpassungen würden dann einer bestimmten p - q -Kombination entstammen, und das Auftreten anderer Paare würde mit zunehmendem Abstand weniger häufig.

Für die abgeleiteten Karten auf Basis der Modellfehler ergeben sich Fehlerraten zwischen 0.8 und 3.6, im Mittel 1.6 Prozent (als Beispiel Urwertkarten). Dabei ist zu berücksichtigen, dass die Modellparameter hier sogar jeweils eigens optimiert wurden – ein Ansatz *one size fits all* mit festen p und q für die gesamte Reihe wird die Resultate noch verschlechtern. Der Abstand zum eingestellten Wert von 0.27 % fasst die Abweichung von der ARIMA-Annahme zusammen. Ansätze zur Verbesserung durch weitere Schritte der Kontrollkartenmethodik fallen hier nicht auf, so sind beispielsweise die Autokorrelationschätzer der Residuen unauffällig. Als alternativer Streuungsschätzer neben den Standardverfahren der Kontrollkarten bietet sich jedoch das Ergebnis aus der Maximum-Likelihood-Schätzung des Modells an. Wird dieses für die Standardabweichung verwendet, so kann die durchschnittliche Alarmrate auf 1.1 Prozent gesenkt werden.

Gegenüber der Verwendung der differenzierten Reihe, also eines ARIMA-Modells vom Typ (0,1,0), kann das Ergebnis durch die ARIMA-Modellierung nur leicht verbessert werden: Die Differenzierung und direkte Erstellung einer Kontrollkarte führt zu einer durch-

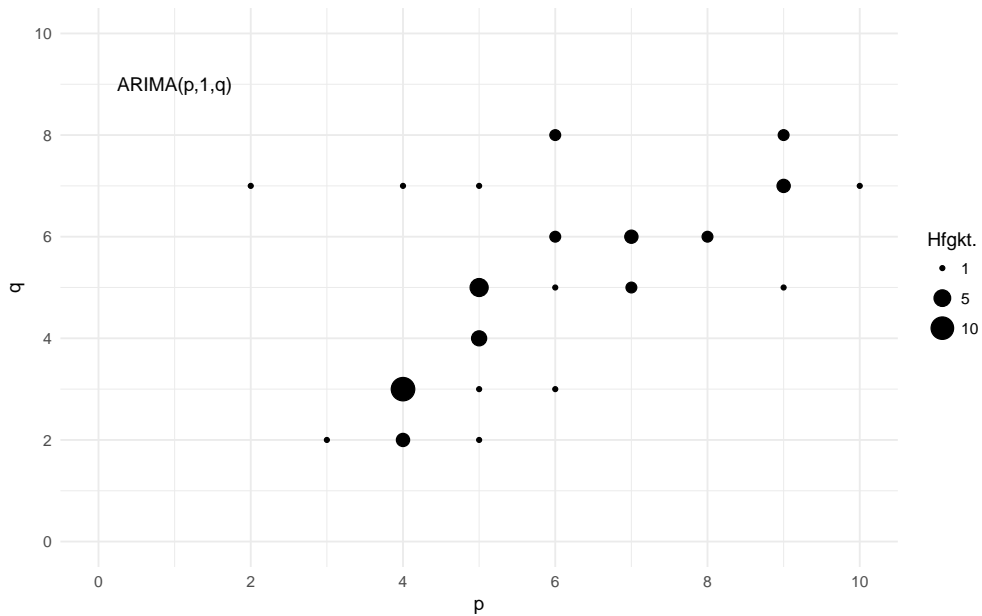


Abbildung 3.10: Die Resultate der BIC-Optimierung in Form der ausgewählten Ordnungen. In zwei Fällen wirkt sich die Begrenzung der Ordnungen durch jeweils 10 aus.

schnittlichen Alarmrate von 2.2 Prozent, beziehungsweise von 1.3 Prozent bei Verwendung des Modellschätzers der Streuung. Gemessen an der Komplexitätszunahme durch die Modelle erscheint der Gewinn zumindest zweifelhaft. Für die vorliegenden Daten müssen klassische Modelle, insbesondere solche mit eher kleinen Parametern p und q und somit übersichtlichen Strukturen, verworfen werden. Eine Abbildung der Daten durch ARIMA-Modelle von ausreichender Qualität für die Verwendung von Kontrollkarten der Modellfehler gelingt damit nicht, der Ansatz wird nicht weiter verfolgt.

Neben ARIMA-Modellen wurden auch weitere Modellvarianten als Grundlage für Kontrollkarten in Betracht gezogen, so etwa (G)ARCH-Modelle (kurz für Generalized Autoregressive Conditional Heteroskedastic) nach Engle (1982) und Bollerslev (1986). Dabei wird ein Prozess nicht über seine Lage-, sondern über seine Varianzstruktur definiert. Auch hiermit konnten für die Bildung von Kontrollkarten keine Erfolge durch Modellierungen erzielt werden, auf weitere Darstellungen wird verzichtet.

Wie bei den ARIMA-Ansätzen kann vermutet werden, dass die in die Frequenzverläufe eingehenden zahlreichen Steuerungen und Regelungen eine Zerlegung der Form $Actual = Fitted + Residual$ mit den genannten Modellklassen nicht ermöglichen, und diese auch nicht ausreichend genau angenähert werden kann. Da die Daten keinem GARCH-

Prozess entstammen, schlagen auch die speziell zugeschnittenen Vorgehensweisen zur Erstellung von „Control Charts for GARCH Processes“ von Schipper und Schmid (2001) hier fehl. Ausnahmen sind die Varianten, die bei der Berechnung der Kontrollgrenzen Parameter enthalten, welche gerade mit Blick auf die eingestellte Fehlerrate optimiert werden. Dieses Vorgehen erscheint jedoch nur sinnvoll, sofern die GARCH-Annahme erfüllt ist. Andernfalls, und gerade bei sehr großen Stichproben, ist ein direktes Schätzen der Quantile der Ausgangsreihe vermutlich die passendere Methode.

3.3.3 VAR-Karten

Neben \bar{x} - oder Urwertkarten wird auch die T^2 -Karte als Karte für Residuen angewandt: Jarrett und Pan (2007) erklären das vector-autoregressive-(VAR-)Modell, welches die multivariate Struktur der Daten berücksichtigt. Im Wesentlichen kann dies durch die folgende Gleichung für einen Vektor $\vec{y}_t \in \mathbb{R}^m$ und den Backshift-Operator B repräsentiert werden (ebd., S. 3 863):

$$\vec{y}_t = \vec{c} + (A_1B + A_2B^2 + \dots + A_pB^p)\vec{y}_t + \vec{\varepsilon}_t$$

Dabei ist \vec{c} ein Vektor bestehend aus Konstanten, die A_i , $1 \leq i \leq p \in \mathbb{N}$, sind Matrizen vom Typ $m \times m$. Durch den Parameter p wird die Ordnung der Autokorrelation ausgedrückt, $\vec{\varepsilon}_t$ stellt den Vektor der Modellfehler dar. Kurz formuliert wird also jeder Wert einer zeitlich geordneten Datenmatrix durch alle Werte der p vorangegangenen Zeilen linear regressiert.

Der Modellfehler stellt dann die Grundlage für die Beurteilung des Prozesses dar. Durch dieses Vorgehen soll die Autokorrelation eliminiert werden. Anschließend werden durch die Verwendung einer T^2 -Karte die verbleibenden Kovarianzstrukturen berücksichtigt. Eine einfache und erfolgreiche Anwendung der VAR-Karte stellen Requeijo und Souza (2011) vor. Auch hier wird auf die Darstellung verschiedener Schritte, Schätzer und dergleichen verzichtet. Es sei aber angemerkt, dass die obige Gleichung für gegebenes p in die Darstellungsweise mehrerer „gewöhnlicher“ linearer Regressionsmodelle überführt werden kann.

Die Tabelle 3.3 enthält die Eigenschaften der resultierenden Karten, welche für einen Auszug der Daten angepasst wurden. Die Autokorrelation der Teststatistik beträgt außer zu Anfang stets um 0.18, weiter werden durchweg etwa 3 Prozent der Zeitpunkte als kritisch eingestuft. Überhaupt sind Veränderungen nur für $p \leq 3$ bei der Autokorrelation zu erkennen. Diese Resultate lassen darauf schließen, dass das VAR-Modell die vorliegenden Datenstrukturen nicht gut erfassen kann.

	p=1	2	3	4	5	6	7	8	9	10
Autokorr.	0.09	0.19	0.18	0.19	0.18	0.18	0.18	0.18	0.17	0.17
Alarmrate [%]	3.09	2.94	3.04	3.02	3.00	2.99	2.98	3.00	3.01	3.02

Tabelle 3.3: Die Selbstabhängigkeit der Statistik und der Anteil falscher Alarme für die VAR-Karte in Abhängigkeit von der Modellordnung p .

Insbesondere ein Vergleich mit dem Teil 3.2.5 zeigt, dass diese Art von Karten nicht geeignet ist: Die „einfachen“ T^2 -Karten erreichen Fehlerraten von unter 2 Prozent, was hier in keinem Fall gelingt. Das Modell verzerrt hier also scheinbar sogar noch Eigenschaften, welche der Mechanik von T^2 -Karten eigentlich zupasskommen.

3.4 Neue Größen

Die Netzfrequenzen lassen sich also auch mit verschiedensten Ansätzen aus dem Bereich der statistischen Prozesskontrolle bestenfalls nur passabel überwachen. Dies wurde deutlich, obwohl bisher nur auf Seiten der Rate falscher Alarme geprüft wurde. Weitere Verfahren, Varianten und Kombinationen können Anwendung finden, wobei die Erfolgswahrscheinlichkeit nach den bisherigen Erkenntnissen eher gering erscheint. Als bestes Verfahren hat sich die EWMA-Karte für OSAPes mit Gilbert-Korrektur erwiesen.

Statt, wie bisher, „vorgefertigte“ Verfahren zu verwenden, sollen nun aus den gewonnenen Erkenntnissen Schlüsse gezogen werden, welche zum Aufbau einer passenden Kontrollkarten-Statistik führen. Als eine der stärksten Annahmeverletzungen stellte sich die extreme Selbstabhängigkeit der Messreihen heraus. Es wurden Versuche unternommen die Autokorrelation, statt wie bisher als Problem, als nützliche Information zu betrachten: Dann könnte etwa der Autokorrelationskoeffizient als Statistik einer Kontrollkarte verwendet werden. Versuche mit gleitenden Fenstern verschiedener Breite zeigten jedoch, dass auch der Autokorrelationskoeffizient selbst wieder hochgradig autokorreliert ist. Dies verhindert die Erstellung von *guten* Karten.

Eine starke Senkung der Autokorrelation konnte aber durch die Verwendung von Differenzen aufeinanderfolgender Messwerte erreicht werden. Varianten mit relativen Differenzen oder der in der Finanzmathematik oft verwendeten „stetigen Rendite“ $\log(X_t) - \log(X_{t-1})$ führten zu qualitativ gleichwertigen Resultaten.

Die Verwendung von Differenzen allein ist aber nicht ausreichend, sondern führt nur zu einer Rate falscher Alarme um 2.0 beziehungsweise 1.3 Prozent (vgl. Teil 3.3.2). Als Ursache für die Abweichungen vom Zielwert können die Verteilungen angesehen wer-

3 Alarmsysteme mit Standard-Methodik

den: Wie die Abbildung 3.11 zeigt, besteht hier keine Normalverteilung. Zwar weicht die Verteilung insgesamt nicht sehr stark ab, sondern ist beispielsweise recht symmetrisch. Zumindest an den Rändern (und gerade diese verwendet eine Kontrollkarte ja) sind die Abweichungen aber bedeutsam.

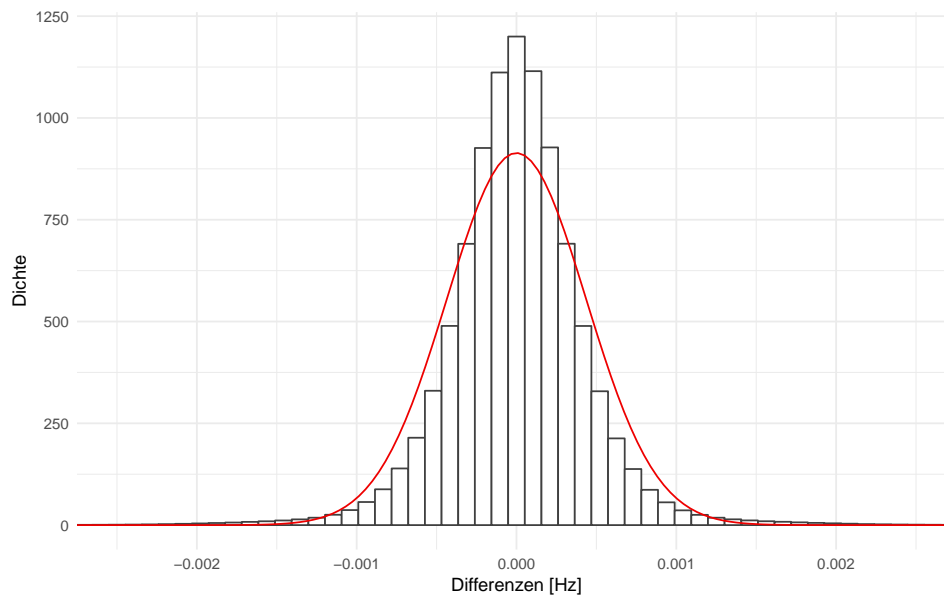


Abbildung 3.11: Die Differenzen stammen aus einer leptokurtischen Verteilung und genügen der Annahme einer Normalverteilung – hier in Form der roten Linie ergänzt – nicht. Die Darstellung zeigt die komplette Zeitreihe 1, getrimmt um einige krasse Ausreißer.

Abhilfe könnte nun ein Vorgehen schaffen, welches aus einer gegebenen Zeitreihe identisch verteilter Elemente x_i mit sonst unbekanntem Eigenschaften eine Zeitreihe y_i erzeugt, welche einer bekannten Verteilung unterliegt. Die Verwendung von Grenzwertsätzen erscheint dabei als praktikabler Ansatz, da diese mitunter sehr wenige Voraussetzungen an die x_i stellen – insbesondere keine bestimmte Verteilung. Mit dem Zentralen Grenzwertsatz (ZGWS, s. u. S. 123 f.) kann dabei eine Normalverteilung als Limes verwendet werden. Dann folgen Beobachtungen aus einer ursprünglich unbekanntem Verteilung nach einer Transformation nicht nur einer definierten, sondern auch einer gut untersuchten Verteilung.

Mit der Wahl dieses Grenzwertsatzes wird 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. Dieses Prinzip wurde bereits in Teil 3.2.4 (Subsample Charts) mit Verweis auf Alwan und

Radson (1992) verwendet. Diese Vorgehensweise wird nun wieder aufgegriffen, wobei hier auch die Überlappung der Abschnitte zweier aufeinander folgender Fenster zugelassen wird:

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

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

wobei mit dem Mittelwert dann folglich $f(x_i, x_{i-1}, \dots, x_{i-w+1}) = 1/w \sum_{j=0}^{w-1} x_{i-j}$ ist.

Für eine gegebene, endliche Zeitreihe der Länge 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 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 $\{x_i, x_{i-1}, \dots, x_{i-w+1}\}$ und $\{x_{i+1}, x_i, \dots, x_{i-w+2}\}$, also die Mengen der Elemente zweier aufeinander folgender Fenster, 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 3.12.

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

$$m = \left\lfloor \frac{n - w}{s} \right\rfloor.$$

Bei der Betrachtung von endlich vielen Beobachtungen, also insbesondere bei der Analyse von Stichproben, kann die Schrittweite über den Umweg der Autokorrelation auch

3 Alarmsysteme mit Standard-Methodik

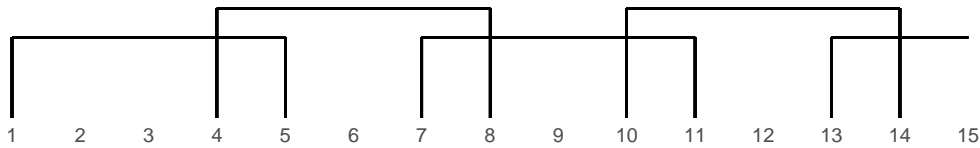


Abbildung 3.12: Mit einer Fensterbreite $w = 5$ und einer Schrittweite $s = 3$ ergeben sich die dargestellten Indexten für die Anwendung der gleitenden Funktion. Statt 15 Werten liegen hier nur noch vier Gruppen vor.

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 empirische Streuung viel geringer ausfallen, als es in der gesamten Datenreihe der Fall wäre. 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 einer stark asymmetrischen, welche viele Ausreißer produziert. Als vorbereitender Schritt wird daher überprüft, inwiefern die Differenzen „symmetrisiert“ werden können. Diese folgen für identisch verteilte Zufallsvariablen grundsätzlich einer symmetrischen Verteilung, da Minuend und Subtrahend den gleichen Verteilungen entstammen.

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, welche dann aber keiner symmetrischen Verteilung mehr folgen. Daher 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 angewandt, 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 könnte also eine Datenvorverarbeitung in Form von

$$\tilde{x}_i = |x_{i+1} - x_i|^d$$

die Konvergenzgeschwindigkeit der Mittelwert-Verteilung erhöhen.⁵

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 an die Normalverteilung „gedrückt“ werden. Neben der Funktion $f(\cdot) = \bar{x}$ als gleitendem Mittelwert wird auch die Variante $f(\cdot) = \bar{x}^k$ berücksichtigt.

Um aus einer gegebenen Zeitreihe Werte zur Verwendung in einer Kontrollkarte zu erzeugen, werden also die nun noch einmal zusammengefassten Schritte unternommen. Dabei ist die Notation so gewählt, dass aus der Realisierung x_1, x_2, \dots, x_n „neue Beobachtungen“ in Form der Reihe y_1, y_2, \dots, y_m gebildet werden, die $n - m$ leeren Elemente werden also gelöscht. 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 Differenzen für eine Potenz $d = 0$ nicht von Interesse ist, wird der Fall „keine Vorverarbeitung der Ausgangsdaten vornehmen“ durch $d = 0$ notiert. Das soll aber keine Fortsetzung (und insbesondere keine stetige Fortsetzung) andeuten.

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: $y_j = \tilde{y}_j^k$ für $k \geq 1$

⁵Eine „komplette“ Box-Cox-Transformation bestünde nicht nur aus der Potenz x^d , sondern aus dem Term $d^{-1}(x^d - 1)$, siehe ebenda.

3 Alarmsysteme mit Standard-Methodik

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 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 3.13 auf der Seite 67 zeigt eine beispielhafte Durchführung des Verfahrens. Dabei wurden Werte aus der χ_5^2 -Verteilung gezogen und für ein bestimmtes Parameter-Setting *normalisiert*. Es ist zu erkennen, dass mit den Transformationen eine Glättung und eine Ausdünnung einhergehen. Gleichzeitig nähert sich die Verteilung der Transformationen der Normalverteilung an.

Zu untersuchen ist nun, für welche Parameter w , s , d und k dieses Vorgehen zum Ziel führt, und wie gleichzeitig die Ausdünnung möglichst gering gehalten werden kann. Dazu bieten sich zwei Vorgehensweisen an: Zum einen können die Parameter einfach mit Blick auf die Alarmrate einer entsprechenden Kontrollkarte optimiert werden. Dieses Vorgehen stellt aber nicht unbedingt sicher, dass *allgemein* gute Parameter gefunden werden. Dazu sollte abgesichert werden, dass die Transformationen auch auf die vorgestellte Art und Weise funktionieren, und nicht nur zufällig einzelne brauchbare Parameterkombinationen gefunden werden. Als Alternative zur speziellen Parameteroptimierung ist also eine allgemeine Empfehlung vorzuziehen.

Soll nun eine Parametereinstellung überprüft werden, ergibt sich somit die Frage, ob die Transformationen normalverteilt sind. Sind sie es nicht, stellt sich die Frage, *wie nah* sie es sind, und ob die Transformationen mit einer bestimmten Parameter-Kombination *näher* an der Normalverteilung sind, als die zu einer anderen. Diese Fragen lassen sich nicht einfach beantworten, so dass zunächst Betrachtungen hierzu erfolgen müssen. Nach diesem Einschub in Form des folgenden Kapitels, wird die Optimierung der Parameter dann im Kapitel 5 vorgestellt.

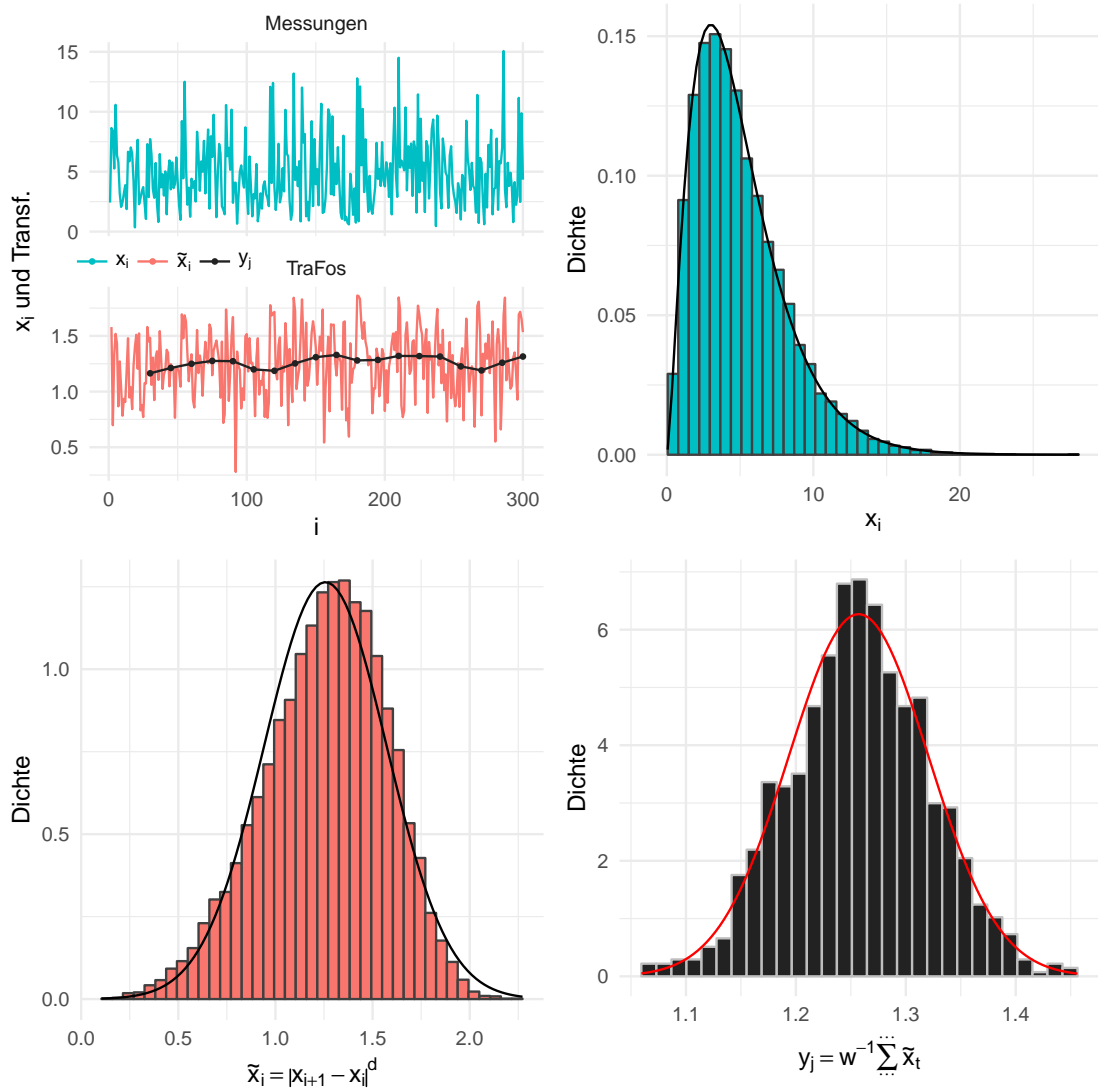


Abbildung 3.13: Das beschriebene Vorgehen beispielhaft für $w = 30$, $s = w/2 = 15$, $d = 0.25$, $k = 1$ und $n = 15\,016$, was $m = 1\,000$ impliziert. Dargestellt ist die Glättung, welche mit jedem Schritt stärker ausfällt. Dabei ist für den Verlauf oben links nur ein kleiner Ausschnitt der Daten dargestellt. Die Histogramme sind ergänzt mit der Dichte der χ^2_5 -Verteilung (oben rechts, x_i), beziehungsweise der Normalverteilung mit den jeweiligen Parameterschätzern.

4 Zur Quantifizierung des Normalverteilungsgrades

Die Normalverteilung ist bekanntlich eines der wichtigsten Werkzeuge eines Statistikers oder einer Statistikerin. So kommt kein umfassendes Lehrbuch ohne die Definition der „Gauß’schen Glockenkurve“ aus, und beispielsweise auch Hartung et al. (2005) oder Genschel und Becker (2005) heben die Normalverteilung als „eine der wichtigsten statistischen Verteilungen“ hervor (Seite 143 f. bzw. Seite 42). Bei zahlreichen weit verbreiteten Methoden wie Regressionsmodellen oder Maximum-Likelihood-Schätzern findet sich die Normalverteilung wieder, etwa als unterstellte Verteilung von Modellfehlern oder als Limes der Verteilung einer Schätzfunktion. Auch und insbesondere der Zentrale Grenzwertsatz verdeutlicht die Bedeutung der Normalverteilung. Wie Hartung et al. (2005) in der Einführung erklären, stammt ein großer Teil der Fundamente heutiger statistischer Verfahren erst aus dem frühen 20. Jahrhundert – die Normalverteilung hingegen wird bereits 1718 von Abraham de Moivre in einer theoretischen Arbeit, und im 19. Jahrhundert von Adolphe Quetelet in praktischer Anwendung (nämlich der Messung des Brustumfangs von Soldaten) verwendet.

Aus der Bedeutung der Normalverteilung erwächst für die praktische Anwendung zahlreicher statistischer Verfahren wie auch im vorliegenden Fall die wichtige Frage, ob die jeweils vorliegenden Daten aus einer Normalverteilung stammen – oder deren Verteilung zumindest nicht allzu stark von dieser abweicht. Letzteres trägt dem Umstand Rechnung, dass in vielen Fällen nur eine *Konvergenz* gegen die Normalverteilung vorliegt. So etwa beim Satz vom zentralen Grenzwert (ZGWS) für eine steigende Anzahl von Summanden, oder bei Dichtfunktionen, welche mit passender Parameterwahl gegen die Dichte der Normalverteilung konvergieren. Hier seien als Beispiele die Konvergenz der t_n -Verteilung gegen die Standardnormalverteilung (für $n \rightarrow \infty$), oder die durch das Galtonbrett gut veranschaulichte Konvergenz der Binomialverteilung genannt.

In den meisten Fällen ist nun jedoch unklar, was eine *nicht zu starke Abweichung* bedeutet. Zur Beurteilung der Normalität eines Datenvektors stehen zwar zahlreiche Methoden zur Verfügung. Diese führen jedoch zumeist entweder nur zur binären Entscheidung

4 Zur Quantifizierung des Normalverteilungsgrades

normal- oder nicht normalverteilt, was insbesondere die Vielzahl an statistischen Tests auf Vorliegen einer Normalverteilung betrifft, oder sie beruhen auf einer grafischen Beurteilung, wie der Form eines Histogramms oder eines Boxplots. Während beide Varianten im Einzelfall praktikabel sind und sich sogar ergänzen können, haben sie bei der Auswertung einer großen Anzahl an Datensätzen klare Mängel: Die Beurteilung von Grafiken ist nicht uneingeschränkt objektiv möglich, erfordert zumeist sogar eine gewisse Vorbildung des Analysten und ist nicht ohne Weiteres automatisierbar. Denkbare Ansätze, beispielsweise ein Histogramm einer statistischen Verteilung automatisiert zuzuordnen, sind aber im Bereich künstlicher Intelligenz gut vorstellbar – müssten dann aber auch mit Verfahren konkurrieren, denen statt dem Histogramm die Daten selbst zur Verfügung stehen. Die Idee einer objektiven Beurteilung von Q-Q-Plots findet sich bei der Gruppe von Teststatistiken wieder, welche auf Regression und Korrelation beruhen (siehe unten, Abschnitt 4.2.1).

Bei der Durchführung von Tests ist es hingegen nicht möglich, graduelle Unterschiede festzustellen. Wird etwa die Normalität einer Summe in Abhängigkeit von der Anzahl der Summanden untersucht, so können Tests trotz stetiger Konvergenz (vgl. ZGWS) nur die „Türschwelle“ zwischen Nicht-Normalität und Normalität erkennen. Über Abstände oder Näherungsgeschwindigkeiten können jedoch keine Aussagen getroffen werden.

Es werden nun Möglichkeiten untersucht, den Abstand einer Stichprobe reellwertiger Zahlen zur Normalverteilungsfamilie zu messen. Diese Abstandsbestimmungen sollen leicht automatisierbar sein, sowie graduelle Unterschiede bezüglich der Nähe zur Normalverteilung angeben können. Eine genauere Beschreibung der Anforderungen folgt als Teil 4.1. Daran anschließend werden die ausgewählten „Methodik-Kandidaten“ zur Messung des Abstands zur Normalverteilung vorgestellt. Es bieten sich insbesondere Metriken für Dichte- oder Verteilungsfunktionen an, auch Teststatistiken erscheinen als sinnvolle Möglichkeit. Neben inhaltlichen Erwägungen wird bei der Auswahl auch die Verfügbarkeit oder einfache Umsetzbarkeit in der statistischen Programmiersprache R berücksichtigt. Eine umfassendere Sicht auf die Verfügbarkeit von Teststatistiken in verschiedenen Programmen geben Yap und Sim (2011) in ihrer Tabelle 1.

Nach einer Gegenüberstellung der verschiedenen Konzepte und Vergleichen theoretischer Eigenschaften der Verfahren folgt die Beschreibung einer durchgeführten Simulationsstudie. Dessen Ergebnisse werden für sich genommen betrachtet, aber auch den theoretisch zu erwartenden Resultaten gegenübergestellt. Theoretische und praktische Eigenschaften werden im Teil 4.4 zusammengefasst und auf dieser Basis dann – soweit möglich – eine Empfehlung für das oder die beste/n Verfahren gegeben. Vor einer Zusammenfassung erfolgen im Teil 4.5 ergänzende Betrachtungen der als am besten eingestuften

Methoden. So sind Abhängigkeiten von der Stichprobengröße sowie von der Bestimmung der jeweiligen Normalverteilungsparameter zu berücksichtigen.

4.1 Ausgewählte Methoden

4.1.1 Anforderungen

Für die Beurteilung des Grades der Annäherung an eine Normalverteilung sind verschiedene Herangehensweisen denkbar. Da eine Normalverteilung für einen Erwartungswert $\mu \in \mathbb{R}$ und eine Varianz $\sigma^2 \in \mathbb{R}_{>0}$ vollständig durch ihre Dichtefunktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

definiert ist, können alle Charakteristiken dieser Verteilung auch aus dieser Funktion abgeleitet werden. Dabei kann die Dichtefunktion offenbar als Ganzes verwendet werden, es können aber auch einzelne Charakteristiken wie beispielsweise die Symmetrie um den Erwartungswert ($f(\mu+x) = f(\mu-x)$) oder Transformationen wie die Verteilungsfunktion als Integral der Dichte ($F(x) = \int_{-\infty}^x f(t) dt$) verwendet werden.

Liegt nun eine Stichprobe $y = (y_1, y_2, \dots, y_n)' \in \mathbb{R}^n$ der Größe $n \in \mathbb{N}$ vor, deren Normalverteilungsgrad beurteilt werden soll, wird also eine Distanzfunktion $d : \mathbb{R}^n \rightarrow \mathbb{R}$ benötigt. Wie im Folgenden dargestellt, ist ein Teil der möglichen Varianten nur zur Beurteilung der Abstands der Daten zu einer *bestimmten* Normalverteilung geeignet, unter Berücksichtigung gegebener Parameter μ und σ^2 handelt es sich dabei also um eine Abbildung $\mathbb{R}^{n+2} \rightarrow \mathbb{R}$.

Wird im Folgenden das Wort „Maß“ verwendet, ist dies nicht nur im mathematischen Sinne einer σ -additiven Abbildung zu verstehen, vielmehr wird es auch als allgemeiner Begriff etwa wie ein Längenmaß verwendet. Als erstrebenswert für ein solches *gutes* Abstandmaß können die folgenden Punkte festgehalten werden.

1. Es können sowohl Abstände zwischen einem Datenvektor und einer Verteilung, als auch zwischen zwei Datenvektoren oder zwei Verteilungen gemessen werden.
2. Es handelt sich bei d um eine stetige Funktion, so dass beispielsweise eine kleine Änderung der Daten nur zu einer kleinen Änderung des Abstands führt.
3. Das Maß ist nach oben und unten durch zwei Größen $a, b \in \mathbb{R}$ beschränkt, so dass ein Funktionswert für sich alleinstehend interpretiert werden kann.

4 Zur Quantifizierung des Normalverteilungsgrades

4. Symmetrie: Sind x, y zwei Datenvektoren oder zwei Verteilungen, so sollte $d(x, y) = d(y, x)$ gelten.
5. Die Zuordnung normal/nicht-normal soll mit möglichst großem Anteil korrekt getroffen werden, sofern dies eindeutig möglich ist.

Eine Erfüllung der ersten Forderung würde nicht nur einen praktikablen Umgang mit dem Maß ermöglichen, es könnten auch weitere Konsistenzüberprüfungen stattfinden: Werden zwei Verteilungen als nah zueinander beurteilt, sollte das auch für zwei Vektoren aus diesen Verteilungen gezogener Zufallszahlen zu erwarten sein.

Eine zunächst naheliegende Variante zur Beurteilung des Grades der Normal- oder auch einer anderen Verteilung stellen Signifikanztests dar. Dabei wird jedoch nur eine binäre Entscheidung ja oder nein erzeugt. Feinere Abstufungen sind nicht möglich, damit ist beispielsweise keine gute Untersuchung des Konvergenzverhaltens von Datenvarianten möglich. Außerdem führen die Stellschrauben des Signifikanzniveaus und der Teststärke (oder des Niveaus und der Stichprobengröße) zu nicht eindeutigen Ergebnissen. Alternativ oder parallel zur Entscheidung eines Tests kann auch der zugehörige p-Wert herangezogen werden, welcher die Forderung nach Abstufung und Begrenzung erfüllt. Auch die mitunter angewandte rein deskriptive Betrachtung des p-Werts ist aber nicht praktikabel: Trifft die Nullhypothese einer bestimmten Verteilung oder Verteilungseigenschaft zu, so ist der p-Wert eine im Intervall $(0,1)$ gleichverteilte Zufallsvariable. Damit kann ein kleiner p-Wert nicht sicher als große Abweichung von einer Normalverteilung aufgefasst werden.

Beeinflusst wird der p-Wert direkt von der jeweiligen Teststatistik. Diese hat deutliche Vorteile gegenüber dem p-Wert: Zwar folgt auch sie einer zufälligen Verteilung, welche abhängig von der Richtigkeit der Nullhypothese ist. Dabei ist es aber nicht ausgeschlossen, dass die Verteilungen der Teststatistik unter Null- und Alternativhypothese zu disjunkten Bereichen führen, oder zumindest weniger Überdeckung aufweisen als bei p-Werten. Des Weiteren liegt einer Teststatistik zumeist eine sachlogische Idee zugrunde, so dass ein großer oder kleiner Wert oft direkt interpretiert werden kann. So wird beispielsweise beim bekannten Kolmogorov-Smirnov-Test der maximale Abstand von empirischer und unterstellter Verteilungsfunktion angegeben – ein Wert, welcher inhaltlich klar und eindeutig verständlich ist. Zudem ist eine Teststatistik, im Gegensatz zum p-Wert oder zur Testentscheidung, frei von Einflüssen der Konvergenzgeschwindigkeit bei approximativ verteilten Teststatistiken.

Neben Signifikanztests bestehen weitere Konzepte, um den Abstand von empirischen oder theoretischen Verteilungen zu beurteilen. Diese entstammen im Allgemeinen der

mathematischen Statistik und werden im folgenden Abschnitt vorgestellt. Die Auswahl berücksichtigt dabei, ob eine Umsetzung in der statistischen Programmiersprache R vorliegt, welche das jeweilige Maß zwischen Daten und Verteilung berechnet.

4.1.2 Distanzen

Zur Beurteilung des Abstands zwischen zwei Verteilungen bestehen verschiedene Möglichkeiten. Oft liegen per definitionem zwei theoretische (nicht empirische) Verteilungen vor, dann müssen die Berechnungsweisen für Datenvektoren entsprechend modifiziert werden (s. u.). Berücksichtigt werden die folgenden Distanzen, welche zum Teil Dichte- und zum Teil Verteilungsfunktionen verwenden. Dabei wäre es auch möglich, alle Abstände nur für Dichten oder nur für Verteilungsfunktionen zu definieren, da sich beide eindeutig ineinander übertragen lassen.

Totalvariationsabstand

Im Teil 3.1.5 „Abstandsmaße und Konvergenzarten für Verteilungen“ führt Rüger (2002) auf den Seiten 41 ff. verschiedene Varianten für Distanzmaße an. Dabei werden P und Q als zwei Verteilungen über dem Borel’schen Messraum $(\mathbb{R}, \mathfrak{B})$ vorausgesetzt. Eine Metrik wird also über der Menge \mathfrak{M} aller Verteilungen dieses Raums gebildet. Der sogenannte Totalvariationsabstand ist dann definiert als

$$TV(P, Q) = \sup_{B \in \mathfrak{B}} |P(B) - Q(B)|$$

In alternativer Darstellung mittels eines Maßes¹ ν und Dichten f und g von P und Q ,

$$TV(P, Q) = \frac{1}{2} \int |f - g| d\nu,$$

wird klar, dass sich der Totalvariationsabstand aus der Fläche von zwei Dichten berechnet: Angegeben wird der Anteil der Fläche unter einer (beliebigen) der Dichtefunktionen, welcher schnittmengenfrei mit der Fläche der anderen Dichte ist, sinngemäß und kurz ist $TV(P, Q) = 1 - \text{Schnittfläche der Dichten}$. In der Abbildung 4.2 auf der Seite 78 ist dies gemeinsam mit den Ideen anderer Verfahren veranschaulicht. Ist im Folgenden jeweils eindeutig, welche Verteilungen beziehungsweise Dichten gemeint sind, wird zumeist statt $TV(P, Q)$ kurz nur TV geschrieben.

¹genauer: mittels eines σ -finiten Maßes ν , „das P und Q dominiert (ein solches existiert stets, z. B. $\nu = P + Q$)“ (Seite 44).

Hellinger-Distanz

Ein ähnliches Maß wie der Totalvariationsabstand stellt der Hellinger-Abstand dar. Dieser beruht auf dem geometrischen Mittel zweier Dichten und gewichtet damit ebenso Bereiche stark, in denen die Dichten beide „groß“ sind. Er berechnet sich durch

$$H(P, Q) = \sqrt{\frac{1}{2} \int (\sqrt{f} - \sqrt{g})^2 d\nu} = \sqrt{1 - \int \sqrt{fg} d\nu}$$

Wie Witting (1985) auf Seite 136 darlegt, können die beiden genannten Distanzen auch allgemeiner als

$$d_r(P, Q) = \left(\int |f^{1/r} - g^{1/r}|^r d\nu \right)^{1/r}$$

formuliert werden, wobei $TV(P, Q) = \frac{1}{2}d_1(P, Q)$ und $H(P, Q) = \frac{1}{\sqrt{2}}d_2(P, Q)$ ist. Beide Distanzmaße sind durch die Null und die Eins begrenzt, stellen Metriken über \mathfrak{M} dar und es gilt

$$TV(P, Q) = 1 \Leftrightarrow H(P, Q) = 1 \Leftrightarrow P \perp Q \Leftrightarrow \exists A \in \mathfrak{B} : P(A) = 1 \wedge Q(A) = 0$$

Lévy-Metrik

Weiter mit Verweis auf Rüger (2002) ist die Metrik nach Paul Lévy anzuführen, welche dort als „das älteste Abstandsmaß für Verteilungen“ (Seite 52) vorgestellt wird. Dabei wird um eine Verteilungsfunktion F eine Lévy-Umgebung so gebildet, dass eine zweite Verteilungsfunktion G komplett in dieser enthalten ist. Die kleinstmögliche Größe dieser Umgebung bildet dann den Abstand zwischen diesen beiden Verteilungen:

$$L(F, G) = \inf\{\varepsilon > 0 : G \in \mathcal{U}_\varepsilon(F)\} (= \inf\{\varepsilon > 0 : F \in \mathcal{U}_\varepsilon(G)\})$$

wobei die Lévy-Umgebung mit der Menge aller eindimensionalen Verteilungsfunktionen \mathcal{F} durch

$$\mathcal{U}_\varepsilon(F) = \{G \in \mathcal{F} : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \forall x \in \mathbb{R}\}$$

definiert wird.

Die Grundidee des Abstands zwischen zwei Verteilungsfunktionen findet sich auch in der durch den Kolmogorov-Smirnov-Test bekannten und unten dargestellten Kolmogorov-

Metrik wieder. Dabei ist anzumerken, dass die Kolmogorov-Variante jeweils nur die Abstände an einer Position x betrachtet. Bei den von Lévy verwendeten Umgebungen wird durch die Betrachtung an den Stellen $x \pm \varepsilon$ eine zusätzliche Einschränkung gemacht, welche eine Stetigkeitsbedingung darstellt. Die Idee der Lévy-Distanz ist in der folgenden Abbildung 4.1 genauer dargestellt.

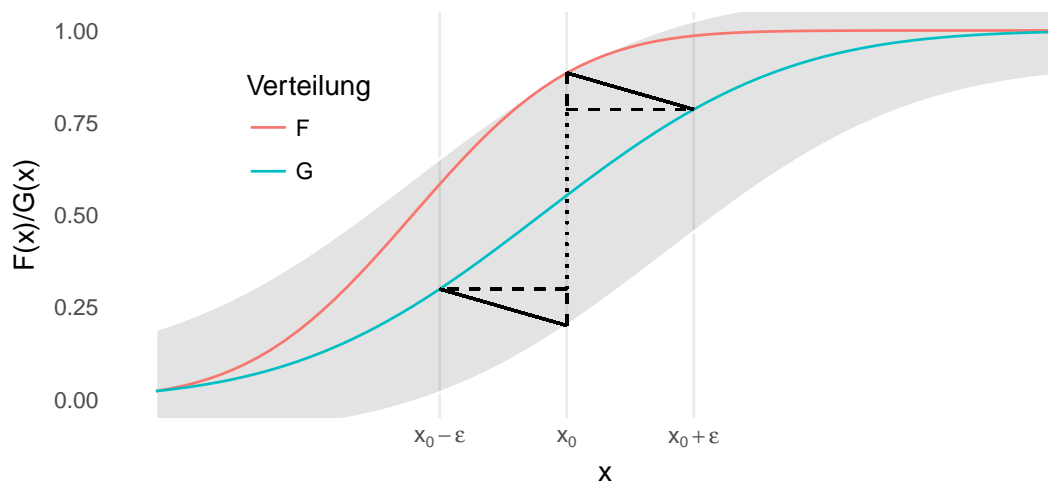


Abbildung 4.1: Die Lévy-Distanz wird gemäß Definition durch das Infimum der Menge $\{\varepsilon > 0 : F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon \forall x \in \mathbb{R}\}$ berechnet. Diese Menge ist hier für das entsprechende Infimum grau hinterlegt. Gemessen wird der Abstand zwischen den beiden Verteilungsfunktionen in 45-Grad-Richtung (bei gleicher Skalierung der Koordinatenachsen): Die beiden Katheten (gestrichelte Linien) des Dreiecks der Punkte mit den Koordinaten $(x_0, G(x_0 + \varepsilon))$, $(x_0, G(x_0 + \varepsilon) + \varepsilon)$ und $(x_0 + \varepsilon, G(x_0 + \varepsilon))$ haben jeweils die Länge ε , die Hypotenuse (durchgezogene Linie) als diagonaler Abstand der beiden Verteilungsfunktionen damit die Länge $\sqrt{2}\varepsilon$. Entsprechendes gilt für das mittels Punktspiegelung erzeugte zweite Dreieck.

Die Lévy-Prohorov-Metrik beruht auf dem gleichen Gedanken wie die Lévy-Metrik und ist auch für mehrdimensionale Verteilungen definiert. An dieser Stelle führt sie aber nur zu einer weniger anschaulichen Definition, so dass die einfachere, ältere Variante vorgezogen wird.

Kolmogorov-Metrik

Der Kolmogorov-Smirnov-Test dürfte zu den bekannteren statistischen Tests zählen und beruht auf dem maximalen Abstand zwischen zwei Verteilungsfunktionen. Dieser Abstand erfüllt alle Eigenschaften einer Metrik und ist zudem nach oben durch die 1 be-

4 Zur Quantifizierung des Normalverteilungsgrades

grenzt. Damit kann ein Maß für die Unterschiedlichkeit zweier Verteilungen gebildet werden durch

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

Die Kolmogorov-Metrik findet sich etwa bei Rüger (2002) auf der Seite 50. Der guten Interpretierbarkeit und hohen Vielseitigkeit dieser Kennzahl steht die Reduzierung der in F und G vorliegenden Informationen auf einen Abstand an einem einzelnen Punkt entgegen. Wie gut diese Verdichtung geeignet ist, wird im Laufe der folgenden Auswertungen untersucht.

Cramér-von-Mises-Metrik

Die Idee der Kolmogorov-Metrik lässt sich leicht auf zwei Weisen weiterentwickeln: Zum einen kann statt der größten Abweichung der beiden Verteilungen auch die kumulative Abweichung der beiden Verteilungsfunktionen verwendet werden. Zum anderen haben sich in der Statistik an vielen Stellen quadratische Abstände als praktikabel erwiesen (etwa bei der Varianz oder dem Kleinste-Quadrate-Schätzer für Regressionsmodelle), so dass sie hier auch als naheliegend erscheinen.

Beide Ideen berücksichtigt die Cramér-von-Mises(CvM)-Distanz, welche nach Harald Cramér und Richard von Mises benannt und durch

$$CM(P, Q) = \int (F(x) - G(x))^2 dF(x)$$

definiert ist. Bei Rieder (1994) findet sich zudem der Hinweis auf die Möglichkeit der Gewichtung durch eine Funktion w , das heißt durch Integration von $(F(x) - G(x))^2 w(x)$. Diese Idee findet etwa beim Anderson-Darling-Test Anwendung, der unten vorgestellt wird.

Vergleich der Ideen

Alle fünf der vorgestellten Metriken lassen sich in Abhängigkeit von Dichte- oder Verteilungsfunktion darstellen. Beispielhaft zeigt die folgende Abbildung 4.2 die jeweiligen „Kernelemente“ der Metriken. Dabei wird auf die zusammenfassenden Elemente wie Integral oder Supremum verzichtet, um die zugrundeliegenden Ideen zu verdeutlichen.

Die Grafik zeigt die Dichte- und Verteilungsfunktionen zweier Normalverteilungen und die entsprechenden Werte der Distanzmaße. Dabei ist festzustellen, dass die Verwendung von Dichten oder Verteilungsfunktionen zu gegenläufigen punktuellen Resultaten führt:

Bei etwa 0.9 weisen die beiden Dichten einen Schnittpunkt auf. Dass die beiden Funktionswerte hier gleich sind bewerten die Hellinger- und TV-Elemente mit dem Minimalwert von 0, während durch den dortigen maximalen Abstand der Verteilungsfunktionen die Methoden von Kolmogorov, Lévy und Cramér/von Mises um 0.9 ihren jeweils größten Wert annehmen.

4 Zur Quantifizierung des Normalverteilungsgrades

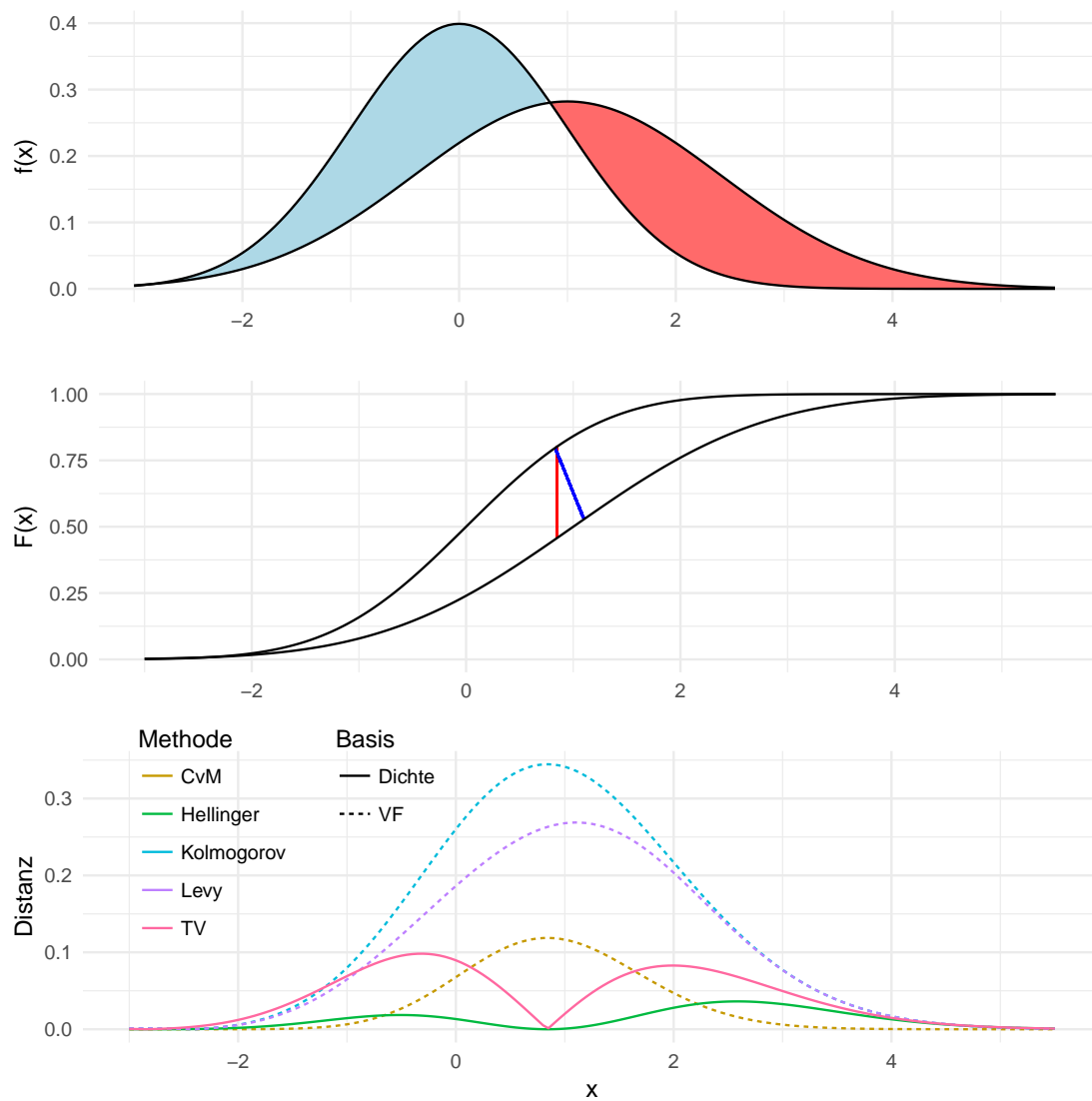


Abbildung 4.2: Dichte, Verteilungsfunktion und punktueller Abstand von zwei Normalverteilungen mit den Parametern $(0,1)$ und $(1,\sigma^2 = 2)$. Die blaue (oder gleichwertig die rote) Fläche ergibt den Totalvariationsabstand, also das Integral der Hälfte der entsprechenden Funktion aus der unteren Grafik. Die vertikale rote Linie zwischen den beiden Verteilungsfunktionen kennzeichnet den Abstand nach Kolmogorov, die diagonale blaue Linie den Abstand nach Lévy. Die Werte der unteren Grafiken werden für die Bildung der Maße in allen Fällen noch zu einer einzelnen Größe verdichtet, etwa durch Integralbildung oder Extremwertbestimmung.

Weitere Distanzmaße

Es existiert ein Vielzahl weiterer Distanzmaße, welche hier nicht berücksichtigt werden. Dabei fußt die vorgenommene Auswahl nicht zuletzt auf der Auswahl der angegebenen Literatur. Des Weiteren sollen keine weiteren Maße hinzugezogen werden, sofern deren Konzept bereits in den einbezogenen Maßen enthalten ist. So beruhen mehrere Maße wie die Bhattacharyya-Distanz auf dem Bhattacharyya-Koeffizienten $BC = \int \sqrt{p(x)q(x)}dx$. Ob dieser Wert groß oder klein ausfällt kann aber stellvertretend an der Hellinger-Distanz durch die Beziehung $H = \sqrt{1 - BC}$ erkannt werden.

Denkbar ist auch, Kombinationen der Ideen von Kolmogorov- und CvM-Metrik zu verwenden, das hieße zusätzlich auch die maximale quadratische Abweichung und die kumulative absolute Distanz zu betrachten (letztere bezeichnet Rieder (1994) als L_1 -Distanz, siehe Seite 125). Da hier jedoch keine wesentlichen weiteren Erkenntnisse zu erwarten sind, wird auf diese Varianten aus Gründen der Übersichtlichkeit verzichtet.

Auf andere Maße wurde verzichtet weil sie per definitionem nachteilige Eigenschaften haben, so ist etwa die Kullback-Leibler-Divergenz nicht symmetrisch in P und Q . Die Energy-Distanz andererseits ist allgemein für Verteilungsfunktionen mehrdimensionaler Beobachtungen konstruiert, die Definition ergibt im eindimensionalen Fall gerade die CvM-Metrik.

Eine bekannte Distanz stellt auch die Mahalanobis-Distanz dar, welche unter anderem in der Diskriminanzanalyse Anwendung findet. Wird ein Datenvektor $(x_1, x_2, \dots, x_n)'$ als Realisierung unabhängig identisch verteilter Größen aufgefasst, ergibt sich die Mahalanobis-Distanz als Summe von zentrierten, standardisierten und insbesondere wiederum unabhängig identisch verteilten Größen. Diese Summe konvergiert gegen eine normalverteilte Zufallsvariable, und das unabhängig von der Ausgangsverteilung. Unterschiede ergeben sich allenfalls bei kleinen Stichprobengrößen, auf welche an dieser Stelle aber nicht eingeschränkt werden soll.

4.1.3 Teststatistiken

Neben den angeführten Metriken werden insgesamt vier Teststatistiken in Betracht gezogen, um den Grad der Normalverteilung einer Stichprobe zu quantifizieren. Diese Statistiken verwenden verschiedene Eigenschaften einer Verteilung, beispielsweise die Form der Dichte oder die erwartete Häufigkeit bestimmter Wertebereiche. Ein Teil der vorgestellten Verfahren kann zur Untersuchung von verschiedenen Verteilungsannahmen verwendet werden, an dieser Stelle wird die Anwendung jedoch jeweils auf die Normalverteilung begrenzt.

4 Zur Quantifizierung des Normalverteilungsgrades

Die folgende Auswahl der Tests soll verschiedene Konzepte der Abweichung von einer Normalverteilung abdecken, zudem sollen die gängigen Methoden enthalten sein. Insbesondere die Einordnung als *gänzig* ist dabei ohne Weiteres sicher als subjektiv zu bewerten. Berücksichtigt werden die Statistiken der folgenden Tests. Diese haben alle eine Nullhypothese der Art „Die Daten stammen aus einer Normalverteilung“. Welche Schlussfolgerungen jeweils daraus gezogen und überprüft werden, wird im Folgenden dargestellt.

χ^2 -Anpassungstest

Die allgemeine Chi-Quadrat-Statistik vergleicht für gegebene Klassen jeweils die beobachtete und die unter einer Nullhypothese erwartete Häufigkeit an Beobachtungen. Für einen Test auf eine stetige Verteilung muss, wie Hartung et al. (2005) auf Seite 182 f. beschreiben, das Intervall $(-\infty, \infty)$ in k disjunkte Abschnitte unterteilt werden. Sind für eine bestimmte Normalverteilung $E_i \in \mathbb{N}$ Beobachtungen in der Klasse $1 \leq i \leq k$ zu erwarten, und werden $O_i \in \mathbb{N}$ tatsächlich beobachtet, so wird die Teststatistik berechnet durch

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Es wird also die relative quadratische Abweichung der absoluten Häufigkeiten von den erwarteten summiert. Bei der Durchführung des Tests ist zu berücksichtigen, wie die Parameter der Normalverteilung hergeleitet wurden, mit denen die E_i berechnet wurden (vgl. ebd.). Das kann hier vernachlässigt werden. Zu berücksichtigen ist jedoch, dass die Statistik eben diese Parameter benötigt, also für μ und σ^2 Schätzer einzusetzen sind. Dazu werden hier das arithmetische Mittel und die Stichprobenvarianz verwendet.

Für die Anzahl der Klassen k wird der Standardwert der verwendeten Software übernommen, das bedeutet $k = \lceil 2n^{2/5} \rceil$. Bei einer *perfekt* normalverteilten Stichprobe ist zwar von $O_i = E_i \forall i$ auszugehen, durch die Quadrierung des Abstandes sind Abweichungen der Summanden von der Null jedoch nur in positiver Richtung möglich und der Erwartungswert von χ^2 ist somit ebenso größer als Null. Die Vergleichsgröße (der kritische Wert) stammt bei einem Test aus der χ_{k-3}^2 -Verteilung, der Erwartungswert beträgt damit $k - 3$. Werden die beiden Parameter einer Normalverteilung nicht geschätzt sondern vorgegeben, sind für die Orientierungsgröße nur $k - 1$ Freiheitsgrade anzusetzen. Nach oben ist diese Statistik offenbar nicht begrenzt.

Anderson-Darling-Test

Der Test von Theodore W. Anderson und Donald A. Darling nutzt aus, dass der Wert der Verteilungsfunktion einer Zufallsvariablen einer Gleichverteilung folgt, was bei der Inversionsmethode zur Erzeugung von Zufallszahlen einer bestimmten Verteilung genau andersherum ausgenutzt wird.

Für die Berechnung der Teststatistik wird die Stichprobe $x_g = (x_{(1)}, x_{(2)}, \dots, x_{(n)})'$ als aufsteigend geordnet vorausgesetzt. Diese wird für einen Test auf Normalverteilung durch die Transformation

$$z_{(i)} = (x_{(i)} - \hat{\mu}) / \hat{\sigma}$$

mit Schätzern für Erwartungswert und Standardabweichung auf eine (0,1)-Verteilung skaliert. Diese empirische Verteilung würde sich bei Vorliegen einer Normalverteilung für die x_i nicht stark von einer $N(0,1)$ -Verteilung unterscheiden, in der Theorie ergibt sich exakt die Standardnormalverteilung.

Die Statistik A^2 des Anderson-Darling-Tests bildet sich daher durch die Verteilungsfunktion Φ der Standardnormalverteilung, ausgewertet an den Stellen $z_{(i)}$. Ist die Anzahl der Datenpunkte bekannt, so können auch die zu erwartenden Werte bestimmt werden. Hinzu kommt nun eine Gewichtung. Die Statistik A^2 wird hier in zwei Varianten angegeben, wobei die zweite einfacher zu interpretieren ist:

$$A^2 = -n - \sum_{i=1}^n \frac{2i-1}{n} (\log(\Phi(z_{(i)})) + \log(1 - \Phi(z_{(n+1-i)})))$$

Diese Form findet sich unter anderem bei Lewis (1961, Seite 119 oben mit abweichender Notation der Indizes), wird dort aber auch nur als „äquivalente Form“ angeführt. In der Grundform wird für empirische und theoretische Verteilungsfunktionen F_n und Φ die Variante

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - \Phi(x))^2}{\Phi(x)(1 - \Phi(x))} d\Phi(x)$$

angegeben. Es handelt sich also wie bei der Distanz nach Cramér und von Mises im Wesentlichen um die Beurteilung des quadratischen Abstands von empirischer und unterstellter Verteilungsfunktion. Dabei wird hier jedoch eine Gewichtung $(\Phi(x)(1 - \Phi(x)))^{-1}$ verwendet, welche bei Cramér und von Mises konstant gleich Eins ist. Dieses Gewicht

4 Zur Quantifizierung des Normalverteilungsgrades

wird beim Anderson-Darling-Test an den Rändern groß, welchen damit mehr Bedeutung zukommt.

Wie der gleichen Quelle in Form der dortigen Tabelle 2 sowie der Gleichung (6) entnommen werden kann, ist die Verteilung von A^2 unter der Nullhypothese abhängig von der Stichprobengröße. Auch ist der kleinste mögliche Wert der Statistik nur asymptotisch gleich Null. Erst für 118 Beobachtungen wird beispielsweise die Schranke 0.01 für den kleinsten *möglichen* Wert unterschritten.

Shapiro-Wilk-Test

Im Jahr 1965 veröffentlichten Samuel S. Shapiro und Martin Wilk den später nach ihnen benannten „analysis of variance test for normality“ (1965). Dabei wird das Verhältnis von zwei Streuungsschätzern verglichen, wobei einer durch die gewöhnliche Stichprobenvarianz gebildet wird, also durch $S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$. Für den zweiten Schätzer wird zunächst angenommen, dass die Beobachtungen x_i einer Normalverteilung mit Erwartungswert μ und Varianz σ^2 entstammen. Dann können die Werte auch durch

$$x_i = \mu + \sigma y_i$$

dargestellt werden, wobei die y_i aus einer Standardnormalverteilung stammen. Auf Basis der Methode der verallgemeinerten kleinsten Quadrate kann nun ein weiterer Schätzer für σ ermittelt werden. Dazu werden die geordneten y_i betrachtet, notiert als $y_{(i)}$. Dann ist für einen Vektor $m = (m_1, m_2, \dots, m_n)$ der Erwartungswerte $m_i = E(y_{(i)})$ und eine Matrix $V = (v_{ij})_{1 \leq i, j \leq n}$ der Kovarianzen $v_{ij} = \text{cov}(y_{(i)}, y_{(j)})$ auch das Quadrat von

$$b = \frac{mV^{-1}}{mV^{-1}m'}x$$

ein Schätzer für die Streuung. Die Teststatistik wird dann gebildet als

$$W = \frac{b^2}{(n-1)S^2}.$$

Für die Berechnung von m und V wird auf die Literatur verwiesen. Wichtig ist an dieser Stelle lediglich, dass bei einer Normalverteilung der x_i ein W von 1 zu erwarten ist. Des Weiteren müssen keine Schätzer für μ und σ herangezogen werden, so dass hier ohne Ergänzungen eine Abbildung der Art $\mathbb{R}^n \rightarrow \mathbb{R}$ vorliegt.

Als einzige Variante der im Rahmen der vorliegenden Ausarbeitung verwendeten Distanzen ist für diese Teststatistik bei Vorliegen einer Normalverteilung der größte Wert zu erwarten. Zu interpretieren ist W damit wie ein Ähnlichkeitsmaß. Um die Interpretati-

onsrichtung „Je größer, desto entfernter von der Normalverteilung“ konsistent anzubieten, wird im Weiteren auch der Wert

$$W^* = 1 - W \in [0, 1]$$

verwendet. Eine kleines W^* deutet demnach auf normalverteilte Daten hin.

Jarque-Bera-Test

Der Jarque-Bera-Test beruht darauf, dass das dritte und das vierte Moment der Normalverteilung unabhängig von den Parametern konstant sind: Alle Vertreter der Normalverteilungsfamilie sind symmetrisch um den Erwartungswert, weisen also eine Schiefe von 0 auf. Die Wölbung, also das vierte zentrale Moment, berechnet sich zur Konstanten 3. Die Nullhypothese der Normalverteilung reduziert sich hier also auf die Aussage, dass Schiefe und Wölbung *nicht zu stark* von denen der Normalverteilung abweichen.

Mit der empirischen Schiefe M_3 und der empirischen Kurtosis M_4 , das heißt mit

$$M_i := \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^i}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{\frac{i}{2}}}$$

berechnet sich die Teststatistik dann durch

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

Die Normalverteilung wird also auf zwei Eigenschaften ihrer Dichtefunktion reduziert, und es werden die entsprechenden quadratischen Abstände betrachtet (wobei offenbar $M_3^2 = (M_3 - 0)^2$ ist). Diese Abstände werden durch die Stichprobengröße n normiert: Je größer die Fallzahl ist, desto geringer dürfen die Abweichungen ausfallen, um JB nicht größer werden zu lassen.

Mit den Faktoren $1/6$ beziehungsweise $1/24$ werden die Abstände so skaliert, dass der Vergleich mit den Quantilen der χ_2^2 -Verteilung zu einem Signifikanztest führt. Diese Verteilung ergibt sich bei einer Normalverteilung der Daten approximativ für die von Jarque und Bera (1987) vorgeschlagene Größe JB (vgl. insbesondere Seite 165).

Wie auch in einer Diplomarbeit von Hain (2010) an der Universität Würzburg ausführlich beschrieben (vgl. Teile 3.2.1 und 3.2.2), bilden die beiden Summanden bereits für sich genommen und ohne Quadrierung die Möglichkeit, Tests durchzuführen: Es gilt für normalverteilte x_i , dass

4 Zur Quantifizierung des Normalverteilungsgrades

$$\sqrt{\frac{n}{6}}M_3 \xrightarrow[n \rightarrow \infty]{D} N(0, 1) \text{ und } \sqrt{\frac{n}{24}}(M_4 - 3) \xrightarrow[n \rightarrow \infty]{D} N(0, 1)$$

Für standardnormalverteilte Zufallsvariablen gilt dann, dass eine Summe von Quadraten gegen die χ^2 -Verteilung konvergiert. Die Anzahl der Freiheitsgrade entspricht dabei gerade der Anzahl an Summanden.

Weitere Tests

Offenbar bilden die vorgestellten Verfahren nur eine Teilmenge aller veröffentlichten und insbesondere aller möglichen Testverfahren ab. Für einzelne Tests sind aber die grundlegenden Ideen bereits durch andere dargestellte Methoden abgebildet. Beispielsweise für die Tests nach Kolmogorov/Smirnov und Cramér/von Mises ist dies durch die gleichnamigen Distanzen der Fall. Der Test nach Lilliefors andererseits stellt nur eine Spezialisierung des Kolmogorov-Smirnov-Tests dar, welche sich bezüglich der Teststatistik nicht von diesem unterscheidet.

Wie der Jarque-Bera-Test basieren auch die Tests von Anscombe/Glynn und von D'Agostino auf der Schiefe beziehungsweise auf der Schiefe und der Wölbung der zu beurteilenden Daten. Während die Grundidee dieses Vorgehens bereits durch die oben dargestellte Größe JB abgedeckt wird, ist die Alternative wie bei D'Agostino et al. (1990) dargestellt zudem deutlich komplexer und damit schwerer zu interpretieren. Damit wird die Variante von Jarque und Bera hier bevorzugt.

An zahlreichen Stellen finden sich weitere Hinweise auf Konzepte, welche zu weiteren Normalverteilungsmaßen führen. So widmen Patel und Read (1996) den „Characterizations“ der Normalverteilung ein ganzes Kapitel. Beispielsweise unter der Nummer 4.2.5 (a) findet sich dort der Sachverhalt, dass für X_i aus einer symmetrischen Verteilung die χ^2_{n-1} -Verteilung von $\sum_{i=1}^n (X_i - \bar{X})/\sigma^2$ äquivalent zur Normalverteilung der X_i ist. Um dies verwendbar zu machen, muss aber zunächst die Verknüpfung mit einer Beurteilung der Symmetrie vorgenommen werden (wobei ein entsprechender Test auch oben bereits genannt ist).

4.1.4 Umsetzung

Für den Großteil der vorgestellten Verfahren werden feste Werte für die Normalverteilungsparameter μ und σ^2 benötigt. Für Schätzungen stehen bekanntlich zahlreiche Methoden bereit, von denen das arithmetische Mittel und die Stichprobenvarianz als Standardmethodik bezeichnet werden können. An dieser Stelle stehen jedoch die Abstände

eines Datenvektors zur Normalverteilung an sich, das heißt zur Familie der Normalverteilungen, im Vordergrund. Damit sollen die Distanzen und Statistiken ausdrücklich auch für nicht-normale Datensituationen berechnet werden. Inwiefern sich Mittelwert und empirische Varianz für diese Verwendung eignen, ist nicht klar. Statt der genannten Schätzer kann auch jeweils der nahestehende Vertreter der Normalverteilung für die Abstandsberechnung herangezogen werden, das heißt eine Distanz d wird für eine Dichte f der Normalverteilung und eine aus den Daten geschätzte Dichtefunktion \hat{f}_n angegeben als

$$d(x) = \min_{\mu, \sigma} d\left(\hat{f}_n(x), f(x|\mu, \sigma)\right)$$

Dieses Vorgehen entspricht der *minimum distance estimation*, wie sie unter anderem bei Boos (1982) für Anderson-Darling-Abstände, bei Rüger (2002, Seite 211) für χ^2 -Abstände oder bei Beran (1977) für Hellinger-Distanzen beschrieben wird. Für Distanzen nach Kolmogorov und Cramér/von Mises wird das Vorgehen bei Rieder besprochen (1994, Seite 232 ff.). Mittelwert und empirische Varianz können dabei als Startwerte für Optimierungen verwendet werden. Eine Untersuchung zu den Vorteilen oder sogar der Notwendigkeit der Optimierung erfolgt später.

Offen ist, auf welche Weise der Schätzer \hat{f}_n gebildet wird. Prinzipiell kommen dabei insbesondere alle Varianten der Kerndichteschätzung in Betracht. An dieser Stelle wird dazu das Vorgehen des R-Pakets *distrEx* verwendet (Ruckdeschel et al., 2006, siehe dazu auch Kohl, 2005). Für die Gegenüberstellung einer Dichtefunktion und einer diskreten und endlichen Menge an Datenpunkten wird dort, wie am Programmcode ersichtlich ist, die stetige Dichte „diskretisiert“ (bei Einstellung des Parameters `asis.smooth.discretize` auf „discretize“). Damit wird das Integral der Distanzfunktionen zu einer Summe. Die Kernidee beruht dabei auf einem Gitter des Bereichs zwischen dem 0.001- und dem 99.999-Prozent-Quantil der zu f gehörenden Verteilung, welches dann den Träger der diskretisierten Verteilung bildet. Für zwei benachbarte Gitterpunkte $y_k < y_{k+1}$ sowie für $y_{k'} = (y_k + y_{k+1})/2$ ist dann $\hat{f}_n(y_{k'}) = F(y_{k+1}) - F(y_k)$.

Das `distrEx`-Paket stellt Funktionen für vier der fünf angeführten Metriken bereit, jedoch keine für die nach Lévy. Dabei ist dieses Paket nicht das einzige mit den jeweiligen Distanzfunktionen, hebt sich aber in einem Punkt ab: Beispielsweise die Funktion `HellingerDist()` akzeptiert als Argumente sowohl zwei Verteilungen, als auch eine Verteilung und einen numerischen Vektor. Für die Anwendung zur Messung der Normalität des Datenvektors wird genau die zweite Variante benötigt – während die erste für theoretische Überlegungen hilfreich ist.

4 Zur Quantifizierung des Normalverteilungsgrades

Für den Lévy-Abstand scheint aktuell keine Implementierung in der Programmiersprache R zur Verfügung zu stehen, somit war hier eine entsprechende Funktion zu erstellen. Es wird dabei auf Funktionen des `distrEx`-Pakets zurückgegriffen. Zwischen diskretisierten theoretischen und empirischen Verteilungsfunktionen wird kein Unterschied gemacht. Die Forderung der Definition „für alle reellwertigen x “ wird durch die Auswertung auf einem Gitter nur näherungsweise realisiert. Dazu werden zwischen den Punkten, an denen jeweils mindestens eine der beiden Verteilungsfunktionen größer als 10^{-5} und kleiner als $1 - 10^{-5}$ ist, äquidistante Gitter bestehend aus 5 000 Punkten verwendet. Für jeden dieser Punkte muss dann die Mengendefinition $F(x - \varepsilon) - \varepsilon \leq G(x) \leq F(x + \varepsilon) + \varepsilon$ erfüllt sein. Die Minimierung des Parameters ε wird dann durch einen Bisektionsalgorithmus realisiert. Dabei wird für die folgenden Anwendungen eine Intervallbreite von maximal 10^{-4} als Stopp-Kriterium verwendet. Bei den Startwerten kann ausgenutzt werden, dass stets $0 \leq L \leq K \leq 1$ gilt.²

Für die Teststatistiken wird auf die R-Pakete `nortest` (für χ^2 - und AD-Test) und `moments` (JB-Test) zurückgegriffen, der Shapiro-Wilk-Test wird durch das Basispaket `stats` bereitgestellt. Genaueres ist bei R Core Team (2018), Groß und Ligges (2015) sowie Komsta und Novomestky (2015) angegeben. Optimierungen der Parameter werden mit dem Algorithmus von Nelder und Mead durchgeführt, wobei für die Varianz eine Restriktion auf den positiven Zahlenraum zu berücksichtigen ist.

Abschließend zu diesem Teil sei daraufhin gewiesen, dass alle bisher genannten Verfahren im Folgenden als „Distanzen“, „Abstände“ oder „Maße“ bezeichnet werden. Dem liegt, neben dem einfacheren Sprachgebrauch, die Idee zugrunde, dass hier alle Methoden als genau solche verwendet werden: Als Werkzeug, *Abstände* beziehungsweise *Distanzen* zu *messen*. Soll zwischen den oben ersichtlichen zwei Gruppen unterschieden werden und gehen diese nicht aus dem Zusammenhang hervor, wird sprachlich zwischen *Metriken* und *Teststatistiken* unterschieden, wobei jeweils der herkömmliche mathematisch-statistische Sinn gemeint ist.

4.1.5 Ausgewählte Verteilungen

Für einen Eindruck der angeführten Kandidaten zur Messung der Normalverteilung werden Simulationen mit verschiedenen Verteilungen durchgeführt. Die Notation $N(\mu, \sigma)$ führt dabei als zweite Größe stets die Standardabweichung, nicht die Varianz auf.

²Dass $0 \leq L$ und $0 \leq K \leq 1$ ist, folgt direkt aus den Definitionen und aus der Begrenzung von Verteilungsfunktionen auf das Intervall $[0,1]$. Wegen $L \leq K$ folgt auch $L \leq 1$. Die Ungleichung von Lévy- und Kolmogorov-Distanz findet sich mit Beweis bei Huber und Ronchetti (2009) auf der Seite 36 als Ungleichung (2.25).

Als Beispiel wird zunächst eine Standardnormalverteilung mit Erwartungswert 0 und Standardabweichung 1 betrachtet. Von dieser ausgehend werden die Abstände zu den Normalverteilungen $N(0,3)$, $N(3,1)$ und $N(3,3)$, sowie zu den Chi-Quadrat-Verteilungen χ_2^2 und χ_5^2 betrachtet. Weiter werden mit den t -Verteilungen t_5 und t_{20} sowie den Gleichverteilungen $U(0,6)$ und $U(-1.5, 1.5)$ zwei symmetrische Verteilungsfamilien einbezogen. Die Poisson-Verteilung $P(2)$ stellt einen Vertreter diskreter Verteilungen dar.

Eine Reduzierung auf einzelne Verteilungen kann nicht allumfassend sein. Diese elf Vertreter sollten aber die Bandbreite der intuitiven Nähe (beziehungsweise des intuitiven Abstands) zur Normalverteilung abbilden. Die folgende Abbildung 4.3 zeigt die Dichten dieser elf Verteilungen, wobei sie aus Gründen der Übersichtlichkeit in Hälften geteilt wurden. Die Verteilungen werden hier und auch im Weiteren kurz durch einen einzelnen Buchstaben oder einen Buchstaben und die Parameter abgekürzt, so wird etwa kurz „die N -Verteilung“ für die Normalverteilung oder „die t_5 “ für die t -Verteilung mit 5 Freiheitsgraden geschrieben.

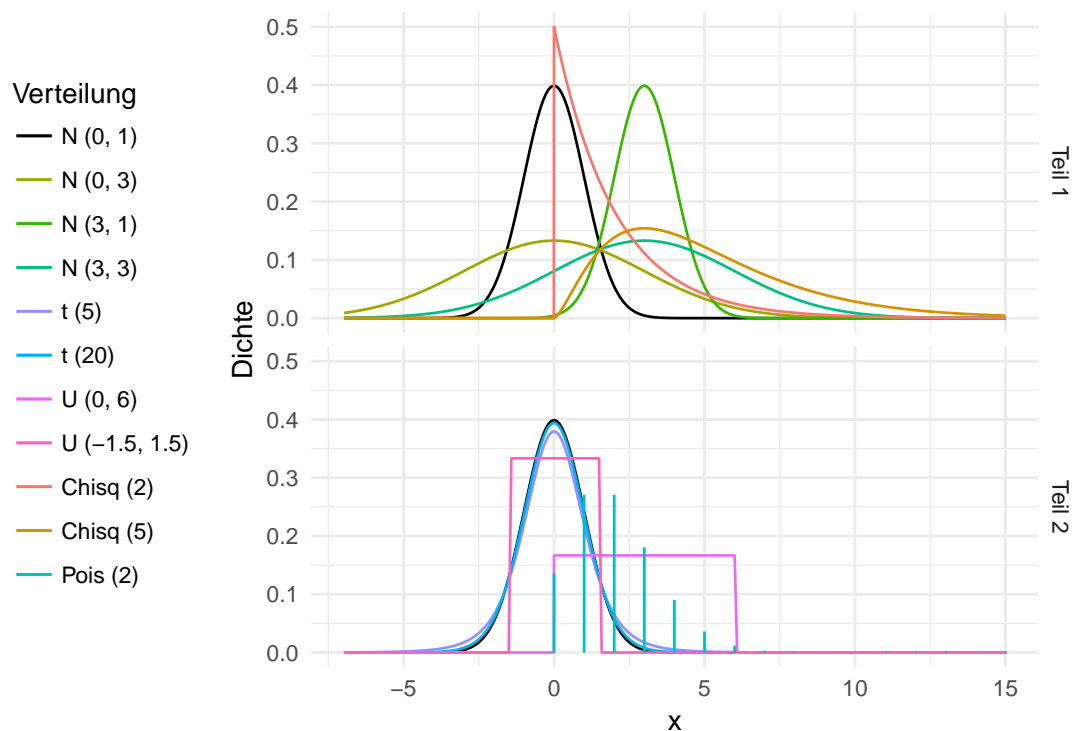


Abbildung 4.3: Die dargestellten elf Verteilungen werden im Folgenden als Basis für Vergleiche verwendet. Betrachtet werden zunächst beispielhaft die jeweiligen Abstände zur schwarz dargestellten $N(0,1)$. Diese ist in beiden Teilen eingezeichnet.

Offenbar ist nun von einem guten Maß eine klare Trennung zwischen der Normal- und der Nicht-Normal-Verteilung eines Datensatzes zu erwarten. Dabei sind Abstufungen erwünscht, so sollte die diskrete Verteilung intuitiv den größten Abstand zur Normalverteilung haben. Die Gleichverteilung teilt mit der Symmetrie immerhin eine zentrale Eigenschaft der Normalverteilung, so auch die t -Verteilung. Durch die Verteilungskonvergenz der t_n -Verteilung gegen die $N(0,1)$ -Verteilung (für $n \rightarrow \infty$) ist *eine irgendwie geartete Nähe* der t - zur N -Verteilung zu erwarten, wobei die Distanz für t_{20} zumindest nicht größer als für t_5 sein sollte. Schließlich kann gefordert werden, dass die vier Normalverteilungen alle zu Werten gleicher Größenordnung führen, wenn lediglich die Distanz zu *irgendeiner* Normalverteilung, also zur Normalverteilungsfamilie insgesamt, untersucht wird.

4.2 Allgemeine Vergleiche

4.2.1 Vergleich der Konzepte

Vor Bewertungen sollen hier die verschiedenen Konzepte der neun oben vorgestellten „Kandidaten“ gegenübergestellt werden. In der Literatur wird beim Vergleich der Methoden im Allgemeinen die Teststärke (Power) in den Mittelpunkt gestellt, so etwa bei Hain (2010), Sürücü (2008), Thadewald und Büning (2007) oder Yap und Sim (2011). Allein diese ist aber, wie oben angeführt, hier nicht von Interesse. Resultate, welche bestimmte Gruppen von Alternativhypothesen betrachten, können aber Hinweise liefern, welche Statistiken gute Indikatoren für Abweichungen von der Normalverteilung sind.

Solche Gruppierungen verwenden Noughabi und Arghami (2011) bei der Konzeption einer Simulationsstudie. Dabei ergeben sich unterschiedliche beste Verfahren, je nach Verteilung der Daten. Symmetrische, nicht-normale Verteilungen mit dem ganzen Raum der reellen Zahlen als Träger werden am besten vom Jarque-Bera-Test erkannt. Das deutet darauf hin, dass die Teststatistik bereits recht sensibel auf von der 3 abweichende Wölbungen reagiert. Bei asymmetrischen Verteilungen mit gleichem Träger fallen die Ergebnisse des Shapiro-Wilk-Tests am besten aus, allerdings werden hier auch ausschließlich Gumbel-Verteilungen betrachtet. Dieser Verteilungstyp weist eine konstante Schiefe von etwa 1.14 und eine konstante Wölbung von 5.4 auf. Dass der JB-Test hier nur (mit wenig Abstand) den zweiten Platz belegt könnte darauf zurückzuführen sein, dass der Shapiro-Wilk-Test ebenfalls gut auf die Asymmetrie reagiert: Schiefe Verteilungen produzieren Ausreißer, welche wiederum die Varianzschätzer stark beeinflussen. Ist dieser Einfluss für die beiden in die Teststatistik W eingehenden Schätzer unterschiedlich, zeigt diese

entsprechend den Unterschied an. Für Verteilungen mit dem Träger $(0, \infty)$ weisen der Shapiro-Wilk-Test und der (hier nicht weiter besprochene und ebenfalls die Streuung betrachtende) Test von Vasicek die besten Werte auf, für solche mit Träger $(0, 1)$ nur letzterer. Es ist jedoch ohne Weiteres nicht klar, wie sich diese Träger auf die Qualität der Tests auswirken.

Auch Monografien und Handbücher liefern neben Kommentaren zur Power zumeist, wie beispielsweise Hartung et al. (2005), lediglich eine Beschreibung einzelner Tests, ohne allgemeine Vor- und Nachteile eingehender und vergleichend zu diskutieren. Auch im Standardwerk *The Theory and Practice of Econometrics* von Judge et al. (1985) findet sich zwar der Verweis auf die „enorme“ Literatur zum Testen auf Normalität (Seite 826), die Autoren geben dann aber nur die Vorschrift zur Berechnung des Shapiro-Wilk-Tests an: Dieser habe demnach „performed reasonably well in a wide variety of circumstances“ (ebenda).

Eine Ausnahme stellt Groß (2004) dar: Hier wird eine Einteilung von Tests auf Normalverteilung in drei Gruppen vorgenommen: Die Gruppe der χ^2 -Tests, die Gruppe der auf der empirischen Verteilungsfunktion beruhenden Tests und die Gruppe der auf Regression und Korrelation beruhenden Tests (vgl. Seite 106 f.). Die dritte Gruppe enthält Methoden, die eine Gegenüberstellung der geordneten Datenreihe mit der bei Normalität zu erwartenden Reihe bewerten. Mit dieser Klassierung werden dann der Chi-Quadrat-Test (erstgenannte Gruppe), der Kolmogorov-(Smirnov-)Test, der Cramér-von-Mises-Test, der Anderson-Darling-Test (zweite Gruppe) und der Shapiro-Wilk-Test (dritte Gruppe) zugeordnet. Die Lévy-Metrik lässt sich zudem klar der zweiten Gruppe zuordnen.

Nicht berücksichtigt werden im Vergleich mit den hier vorgestellten Verfahren die Hellinger-Distanz, der Totalvariationsabstand und der Jarque-Bera-Test. Wie die obige Abbildung 4.2 bereits verdeutlichte, lässt sich für die Distanzmaße eine Aufteilung in Dichte- und Verteilungsfunktion-basierte Methoden erstellen. Somit kann eine ergänzende, vierte Gruppe vorgeschlagen werden: Die der Dichte-basierten Methoden. Dieser kann auch der Test von Jarque und Bera zugeordnet werden, welcher mit Schiefe und Wölbung gerade Eigenschaften der Dichtefunktion untersucht.

Eine Verteilungsfunktion ist bekanntlich stets monoton nichtfallend, rechtsseitig stetig und konvergiert für $x \rightarrow \mp\infty$ gegen 0 beziehungsweise 1 (vgl. bspw. Hartung et al., 2005, Seite 106). Damit ist zumindest die optische Form recht stark eingegrenzt, für die Dichtefunktion bestehen mehr geometrische Möglichkeiten. So fällt beispielsweise der Unterschied von Dichtefunktionen stetiger und diskreter Verteilungen, oder uni- und bimodaler Verteilungen direkt ins Auge. Ob diese Mehrzahl an Charakteristika aber auch von den Methoden gewinnbringend berücksichtigt wird, ist zu untersuchen.

4 Zur Quantifizierung des Normalverteilungsgrades

Wieder bei Groß (2004) im Kapitel 4 finden sich Test-Empfehlungen, welche Eigenschaften der Verfahren berücksichtigen. So wird vom χ^2 -Test abgeraten, nicht zuletzt da das Resultat von der Wahl der Klassen abhängt. Weitere Argumente behandeln dann im Wesentlichen die Teststärke. Bemerkenswert ist das auf Seite 113 angeführte Zitat von Ralph B. D'Agostino, nach dem der Kolmogorov-Smirnov-Test nur eine „historische Kuriosität“ sei, welche „niemals verwendet werden sollte“. Als gut geeignet werden hingegen die Tests nach Anderson/Darling, Cramér/von Mises, Shapiro/Francia und Shapiro/Wilk empfohlen. Inwieweit sich diese Ergebnisse übertragen lassen, wenn die Teststatistiken selbst betrachtet werden, weitere Verfahren hinzugenommen werden und die angeführten Verteilungen als Beispiele verwendet werden, wird die folgende Analyse zeigen.

Es lassen sich weitere Merkmale der Verfahren ausmachen, so wurde die Unterscheidung zwischen Metriken und Teststatistiken bereits deutlich. Auch fällt die Spezialisierung auf die Normalverteilung ins Auge. So sind mit Ausnahme des Shapiro-Wilk- und des Jarque-Bera-Tests alle vorgestellten Verfahren nicht an die Normalverteilung gebunden, sondern lassen sich auch auf andere Verteilungen übertragen. Es ist zu erwarten, dass die Spezialisten bei ausschließlicher Betrachtung der Anwendung für Normalverteilungen bessere Resultate liefern.

Rüger (2002) führt in Teil 3.1.5 (ab Seite 41) Zusammenhänge von Verteilungskonvergenzen an, wenn diese für verschiedene Metriken definiert werden. Eine Folge von Verteilungen P_n ist demnach gegen eine Verteilung P konvergent bezüglich einer Metrik d , wenn $d(P_n, P) \rightarrow 0$ ($n \rightarrow \infty$) gilt. Dies wird als $P_n \xrightarrow{d} P$ notiert. Es bestehen die Zusammenhänge (vgl. a. a. O. (3.41) sowie Satz 3.6)

$$P_n \xrightarrow{TV} P \Leftrightarrow P_n \xrightarrow{H} P \Rightarrow P_n \xrightarrow{K} P \Rightarrow P_n \xrightarrow{L} P$$

Wenn sich also zwei Verteilungen im Sinne von Hellinger oder des Totalvariationsabstands *nah* beieinander befinden, dann auch im jeweils anderen Sinne und zudem nach den Definitionen von Kolmogorov und Lévy. Zeigt die Lévy-Metrik die Nähe zweier Verteilungen an, muss dies aber offenbar nicht für die anderen gelten. Es wird also nicht überraschen, wenn das Lévy-Maß tendenziell häufiger eine Nähe zur Normalverteilung anzeigt als die anderen Maße. Ob das förderlich oder hinderlich ist, und ob eventuell die Kolmogorov-Metrik als „mittleres“ Maß der dargestellten Implikationen die ausgewogensten Zuordnungen trifft, ist zu untersuchen.

4.2.2 Vergleiche für theoretische Verteilungen

Die Frage nach der Distanz zur Normalverteilungsfamilie kann auf zwei Weisen gestellt werden: Zum einen für eine Verteilung als Ganzes, zum anderen für eine Realisierung in Form eines Datenvektors. Die erste Variante wird in diesem Teil betrachtet, die Übertragung auf die Datensituation folgt unten im Teil 4.3. Abstände zwischen jeweils zwei Verteilungen können dabei durch die Metriken per definitionem berechnet werden, also zwischen jeweils zwei theoretischen Verteilungen in ihrer Darstellung durch eine Dichte- oder Verteilungsfunktion. Für die im Teil 4.1.5 angeführten Verteilungen ergeben sich dann Werte wie weiter unten angegeben.

Da die Teststatistiken in ihrer Grundform von einem Datenvektor ausgehen, müssen Anpassungen vorgenommen werden, um theoretische Verteilungen miteinander vergleichen zu können. Die Chi-Quadrat-Statistik etwa vergleicht die erwarteten und beobachteten Häufigkeiten in jeweils einer Klasse miteinander. Für gegebene Klassengrenzen können auch die erwarteten Häufigkeiten zwei theoretischer Verteilungen verglichen werden. Der Vergleich von relativen Häufigkeiten ermöglicht die Unabhängigkeit von der Stichprobengröße n , wobei dann statt der Größe χ^2 der Wert x^2/n betrachtet wird.

Die Festlegung der Klassengrenzen wird hier anhand einer der beiden zu vergleichenden Verteilungen so vorgenommen, dass für jedes Intervall i ein Anteil $E_i \equiv 1/k$ zu erwarten ist. Über Quantils- und Verteilungsfunktionen können dann die erwarteten relativen Häufigkeiten für die zweite Verteilung bestimmt werden. Wird nun außerdem jeweils statt $1/n \cdot \chi^2$ der Wert $1/(k-1) \cdot 1/n \cdot \chi^2$ verwendet, das heißt die durchschnittliche relative Abweichung pro Klasse, ist auch eine Unabhängigkeit von der Anzahl der Klassen gegeben. Dabei sollte k offenbar nicht zu klein gewählt werden, um Streuungseinflüsse zu begrenzen.

Der Anderson-Darling-Test stellt in der Definitions-Variante über das Integral eine empirische und eine theoretische Verteilungsfunktion gegenüber. Dabei kann zwar prinzipiell einfach die empirische Funktion gegen eine theoretische Funktion ausgetauscht und das Integral berechnet werden – je nach Verteilung ändert sich hier aber das (nach Lebesgue–Stieltjes definierte) Integral. Dessen recht aufwendige Berechnung würde im Aufwand den Nutzen vermutlich deutlich überschreiten, so dass der AD-Test hier nicht berücksichtigt wird.

Das Gleiche gilt für den Test nach Shapiro und Wilk. Dabei werden, wie oben beschrieben, zwei Varianzschätzer verglichen, wobei einer der Schätzer auf Erwartungswerten und Kovarianzen von sortierten und transformierten Beobachtungen beruht. Deren Herleitung für alle der hier verwendeten nicht-normalen Verteilungen gestaltet sich recht aufwendig.

4 Zur Quantifizierung des Normalverteilungsgrades

Die Variante nach Jarque und Bera ist nur abhängig von Schiefe und Wölbung, beide Kennzeichen sind für die theoretischen Verteilungen bekannt. Die einzige Variable in der obigen Definition von JB ist dann die Stichprobengröße n , welche für den Vergleich zweier theoretischer Verteilungen jedoch ohne Interpretationsverlust ausgelassen werden kann. Betrachtet wird dann die Größe $1/n \cdot JB$.

Zu beachten ist sowohl für die χ^2 - als auch für die JB -Variante: Bei einer perfekten Übereinstimmung zweier Verteilungen, und simplem Einsetzen der Parameter (Schiefe ist gleich, usw.), ergeben sich Werte von $\chi^2 = JB = 0$. Da aber in beiden Fällen keine symmetrische Verteilung der Teststatistik vorliegt (folgend aus dem Betrachten jeweils quadratischer Abstände), beträgt der Erwartungswert für die Untersuchung eines Datenvektors jeweils nicht Null.

Die Werte für die vier Metriken und die beiden berücksichtigten Teststatistiken finden sich in der folgenden Tabelle 4.1. Dabei wird jeweils gemessen, wie nah die genannten Verteilungen an der Standardnormalverteilung sind.

	Kolm.	TV	Hell.	Lévy	CvM	$\frac{1}{n(k-1)}\chi^2$	$\frac{1}{n}JB$
N(0, 1)	0.000	0.000	0.000	0.000	0.000	0.000	0.000
N(0, 3)	0.242	0.484	0.475	0.166	0.218	0.092	0.000
N(3, 1)	0.866	0.866	0.822	0.552	0.741	0.567	0.000
N(3, 3)	0.625	0.653	0.618	0.391	0.557	0.345	0.000
t(20)	0.008	0.016	0.031	0.005	0.007	0.000	0.006
t(5)	0.030	0.061	0.109	0.018	0.028	0.001	1.500
U(-1.5,1.5)	0.067	0.185	0.284	0.025	0.059	0.003	0.060
U(0,6)	0.687	0.687	0.737	0.483	0.588	0.374	0.060
χ_2^2	0.500	0.500	0.625	0.379	0.359	0.100	1.767
χ_5^2	0.847	0.847	0.844	0.549	0.748	0.643	0.907
P(2)	0.706	1.000	1.000	0.312	0.537	0.262	0.094

Tabelle 4.1: Die fünf Metriken und die zwei angepassten Teststatistiken als Beurteilungsgrundlage des Abstands verschiedener Verteilungen von der N(0,1).

Alle Maße erkennen die Identität, wie die erste Zeile zeigt. Die veränderte Streuungskomponente (N(0,1) verglichen mit der N(0,3), zweite Zeile) bewerten die TV- und H-Metriken nahezu gleich, während die K- und L-Messungen hier weniger stark ausschlagen. Die vergleichsweise geringe Schnittmenge der N(0,1)- und der N(3,1)-Dichten erkennen die fünf Metriken als solche mit Werten zwischen 0.55 und 0.87. Bei der Variante in der vierten Zeile, also beim Vergleich von N(0,1) und N(3,3), erkennen alle vier Metriken einen mittelgroßen Abstand. Numerisch fällt die Lévy-Variante dabei etwas ab.

Die durchschnittliche χ^2 -Statistik schwankt numerisch relativ stark innerhalb der Vergleiche von Normalverteilungen, bewertet aber ebenfalls die vorgenommene Streuungsänderung deutlich schwächer als die Verschiebung. Der Jarque-Bera-Test hingegen kann hier konstruktionsgemäß keine Unterschiede feststellen, da alle Normalverteilungen die gleiche Schiefe und die gleiche Wölbung aufweisen

Für die t-Verteilungen werden wie zu erwarten von fast allen Varianten eher kleine Abstände ausgegeben. Dass etwa die Dichte der t_{20} näher an der $N(0,1)$ ist als die der $N(3,3)$, ist nach Betrachtung der Funktionsverläufe auch eine sinnvolle Interpretation (die Näherung an *irgendeine* Normalverteilung wird erst später betrachtet). Einzig der JB-Abstand erkennt für die t_5 -Verteilung eine recht klare Abweichung von der $N(0,1)$, welche auf die unterschiedliche Wölbung zurückzuführen ist (9 bei der t_5 , 3 bei der N -Verteilung).

Für die Gleichverteilungen hingegen hat der JB-Abstand kaum Indizien auf Abweichungen von der $N(0,1)$: Schiefe und Kurtosis sind Parameter-unabhängig konstant, nämlich 0 und 1.8, so dass nur der Unterschied in der Wölbung zum Tragen kommt. Die Totalvariations- und Hellinger-Metriken erkennen die Unterschiede von $N(0,1)$ - und Gleichverteilungs-Dichten beziehungsweise -Verteilungsfunktionen relativ gut, insbesondere wenn dazu die Lageverschiebung der $U(0,6)$ kommt. Die Methoden nach Kolmogorov, Lévy und CvM, sowie der χ^2 -Wert, scheinen genau diese Lageverschiebung für eine große Sensibilität zu benötigen.

Die Dichten der χ^2 -Verteilungen sind ausschließlich echt positiv wenn auch das Argument größer als Null ist. Mit Schiefen von 2 beziehungsweise 1.3 weichen sie recht gut sichtbar von der $N(0,1)$ -Dichte ab. Gerade für die Trennung der Standardnormalverteilung von der Chi-Quadrat-Verteilung mit zwei Freiheitsgraden scheint die Jarque-Bera-Methode am besten geeignet. Auffällig ist, dass sie als einzige Methode den Abstand von $N(0,1)$ und χ^2_2 größer bewertet, als den von $N(0,1)$ und χ^2_5 (vgl. neunte und zehnte Zeile). Hier zeigen sich die unterschiedlichen Konzepte von Form- und Flächendifferenzen. Innerhalb der Metriken fällt auf, dass hier die Kolmogorov-Variante numerisch näher an TV und Hellinger ist als an der Lévy-Methode, der sie konstruktiv näher ist. Ein solches Paar-Verhalten findet sich auch bei den χ^2 -Verteilungen wieder: Gleiche (gerundete) Werte bei K- und T-Metrik, abweichende Beurteilung durch die H-Metrik. Der Unterschied fällt bei der χ^2_5 -Verteilung jedoch geringer aus.

Die intuitive Forderung nach einem maximal großen Abstand der stetigen $N(0,1)$ zur diskreten Poisson-Verteilung erfüllen der T- und der H-Abstand ohne Einschränkung. Das ist plausibel, da sich anhand von Dichtefunktionen der Unterschied zwischen stetigen und diskreten Trägern auch grafisch sehr gut verdeutlichen lässt. Die K-, L und CvM-Metriken

4 Zur Quantifizierung des Normalverteilungsgrades

erkennen den Abstand wesentlich schlechter, zumindest aber deutlich. Hier kommt zum Tragen, dass die diskrete Verteilungsfunktion eine Treppenfunktion bildet. Nur moderate Abweichungen von der $N(0,1)$ erkennt die χ^2 -Variante. Der Abstand nach Jarque und Bera kann kaum Unterschiede aufzeigen: Offenbar weichen Schiefe und Wölbung der Poisson (2)-Verteilung mit $1/\sqrt{2}$ und $1/2$ nicht stark genug von denen der $N(0,1)$ ab.

Mit diesen Eindrücken erscheinen alle Varianten als grundsätzlich geeignet: Insgesamt ergeben sie jeweils ein schlüssiges Bild, und Schwächen gegenüber anderen Methoden bei einzelnen Vergleichen werden durch Stärken bei anderen Vergleichen kompensiert. Auch inakzeptables Verhalten, wie beispielsweise das Nicht-Erkennen der Gleichheit zweier Verteilungen, fällt nicht auf.

In der Qualität abfallend gegenüber den Alternativen ist aber der χ^2 -Abstand, welcher bei Vorliegen eines anderen der Maße keine neuen Information liefern würde. Dieses Ergebnis ist einerseits bemerkenswert, da dieser Methode ein sehr großer Bekanntheitsgrad attestiert werden kann – welcher sich aber vermutlich vor allem in der vergleichsweise einfachen Idee und der Flexibilität der Methode begründet. Zudem ist zu berücksichtigen, dass hier für Vergleiche von stetigen Verteilungen Klassen gebildet wurden. Dass sich dieser Informationsverlust letztlich in der Qualität niederschlägt, ist nicht verwunderlich.

4.2.3 Folgerungen

Die abgeleiteten Eigenschaften der einzelnen Verfahren werden nun zusammengefasst. Eine Übersicht findet sich anschließend am Ende dieses Teils auf der Seite 96 in Form der Tabelle 4.2. Diese ordnet die im Folgenden dargelegten Charakteristiken jeweils als „erfüllt“, „teilweise erfüllt“ oder „nicht erfüllt“ ein.

Die Gruppierung der Methoden in Metriken und Teststatistiken bei deren Vorstellung im Teil 4.1 war durch die grundsätzliche Herangehensweise motiviert: Ein Teil untersucht die Abstände zweier Funktionen, der andere Teil wurde zur Beurteilung eines Datenvektors konstruiert. Wünschenswert ist es hingegen, die gleiche Methode für beide Fälle anwenden zu können. Es ist, wie oben ausgeführt, prinzipiell auch in allen Fällen möglich, diesen Übergang zu gewährleisten. Durchgeführt wurde dies jedoch nur teilweise durch allgemeine Verfahren wie Diskretisierungen stetiger Verteilungen. Eine effiziente und gut untersuchte, sowie direkte Übertragung der jeweiligen Idee, ist zum jetzigen Zeitpunkt nur für die Methoden nach Anderson-Darling, Cramér-von Mises und Kolmogorov bekannt (und zwar in Form der jeweils gleichnamigen Distanzen und Teststatistiken).

Für die anderen Methoden ist diese Forderung nur als teilweise oder gar nicht erfüllt anzusehen: Etwa der Gedanke des Jarque-Bera-Tests, der Vergleich von Schiefe und Wöl-

bung zweier Verteilungen also, ist auf alle Verteilungen und empirischen Werte übertragbar. Wirklich untersucht, inklusive der Verwendung einer zweckdienlichen Gewichtung der Abstände, sind solche Adaptionen bisher aber nicht. Der Shapiro-Wilk-Test beruht auf dem Vergleich zweier Varianzschätzer. Dieser Vergleich ergibt für theoretische Verteilungen keinen Sinn, da dann die Varianzen eindeutig bekannt sind und nicht geschätzt werden müssen. Werden ersatzweise zwei Varianzen verglichen, um zwei Verteilungen zu vergleichen, hat sich die Vorgehensweise schon weit vom Shapiro-Wilk-Grundgedanken gelöst, und würde auch eher dem F-Test auf Varianzgleichheit als empirischem Analogon entsprechen.

Die oben in Teil 4.1.1 gestellte Forderung nach Symmetrie kann nur für den Vergleich von zwei Verteilungen oder zwei Datenvektoren sinnvoll formuliert werden. Da hier jeweils der Abstand zu einer (Normal-)Verteilung von Interesse ist, kann die Forderung weiter auf den Vergleich von zwei Verteilungen abgeschwächt werden. Die Gruppe der *Metriken* erfüllt eine Definition, welche die Symmetrie-Eigenschaft enthält. Die Teststatistiken hingegen mussten für Betrachtungen zwischen Verteilungen angepasst werden. Dabei läuft mit den Gewichtungen durch die erwartete Klassenhäufigkeit für eine der Verteilungen beim χ^2 -Test direkt eine zentrale Eigenschaft der Symmetrie-Forderung entgegen. Auch der Anderson-Darling-Test verwendet eine bei Integral-Schreibweise leicht erkennbar asymmetrische Vorgehensweise. Beim Jarque-Bera-Test werden die quadratischen Abstände von Schiefe und Wölbung zweier Verteilungen verglichen. Ein Vertauschen der Argumente führt offenbar zu gleichen Resultaten, so dass hier Symmetrie vorliegt. Bei Shapiro-Wilk-Test besteht keine Variante für zwei Verteilungen, und damit insbesondere auch keine symmetrische.

Wichtig für die Interpretierbarkeit der numerischen Resultate sind absolute Bezugsgrößen. So ist etwa für die fünf Metriken klar, dass Werte nahe der Null für zwei ähnliche Verteilungen stehen, Werte nahe Eins entsprechend für sehr unterschiedliche. Mit welchen Zahlenbereichen der Begriff *nahe* belegt wird ist ohne weitere Betrachtungen nicht klar. Ohne beidseitige Beschränkungen der Werte ist aber zumindest einer der Begriffe *klein* und *groß* gar nicht fassbar. Bei der Verwendung im Signifikanztest werden die Größen χ^2 und JB mit $(H_0)\chi^2$ -Verteilungen verglichen. Das verdeutlicht, dass keine oberen Schranken angegeben werden können. Die Normierung durch ein großes Quantil könnte zwar ein praktikables Vorgehen sein, wäre aber weiter zu untersuchen. Für die Chi-Quadrat-Statistik wäre dabei auch der Einfluss der Klassenzahl zu betrachten, diese beeinflusst beim Chi-Quadrat-Test die Verteilung der H_0 -Teststatistik über die Anzahl der Freiheitsgrade.

4 Zur Quantifizierung des Normalverteilungsgrades

Auch für den AD-Test kann keine obere Schranke angegeben werden. Das wird an der hier verwendeten Summen-Schreibweise klar: Neben den Gewichten beinhaltet jeder der Summanden einen Term $-\log(\Phi(z_{(i)})) - \log(1 - \Phi(z_{(n+1-i)}))$. Da keine Anforderungen an die z - oder die zugrunde liegenden x - Werte gestellt werden, kann dieser Term jede beliebige Grenze überschreiten. Einzig der Shapiro-Wilk-Test führt zu einer beschränkten Größe, wobei W wie bereits erwähnt in das Intervall von 0 bis 1 fällt (und folglich ebenso die Variante $W^* = 1 - W$).

Nicht weiter verglichen werden an dieser Stelle die Resultate der theoretischen Abstände verschiedener Verteilungen zur Standardnormalverteilung, wie sie oben in der zugehörigen Tabelle 4.1 (Seite 92) dargestellt wurden: Diese dienen vor allem einem ersten Eindruck der Maße. Auch sollten alle Schlussfolgerungen, die mit Rücksicht auf die Ergebnisse zu ziehen sind, sich auch in der Simulationsstudie wiederfinden – andernfalls wäre gerade die Tatsache dieses nicht konsistenten Bildes zu bewerten.

Eine Zusammenfassung der Eigenschaften folgt nun als Tabelle 4.2. Für weitere Schlussfolgerungen sind die Resultate der im Folgenden beschriebenen Simulationsstudie hinzuzuziehen. Eine gemeinsame Betrachtung erfolgt dann im Abschnitt 4.4.

Eigenschaft	Kolm.	TV	Hell.	Lévy	CvM	χ^2	JB	AD	W
Theorie und Daten	●	◐	◐	◐	●	◐	◐	●	○
Symmetrie	●	●	●	●	●	○	●	○	○
Beschränktheit	●	●	●	●	●	○	○	○	●

Tabelle 4.2: Die im Text genannten Eigenschaften der Methoden als Zusammenfassung. Die geforderten Eigenschaften sind erfüllt oder nicht, was durch die Symbole ● und ○ gekennzeichnet wird. Die teilweise Erfüllung, markiert mit ◐, wird ebenfalls jeweils im Text erläutert.

4.3 Simulationsstudie

Die insgesamt neun Varianten zur Messung der Normalität werden nun nicht nur bezüglich einiger theoretischer Eigenschaften, sondern auch durch eine Simulationsstudie verglichen. Die dazu ausgewählte Vorgehensweise und die Ergebnisse werden an dieser Stelle vorgestellt.

4.3.1 Simulationsdesign

Statt des Abstands zu einer bestimmten Verteilung soll der Abstand von empirischen Daten zur Normalverteilungsfamilie insgesamt, das heißt praktisch zu *irgendeiner* oder *der nahesten* Normalverteilung, beurteilt werden. Das bedeutet, dass statt zwei Verteilungen nun Realisationen der einen und die Familie der anderen Verteilung vorliegen. Wie beschrieben werden dazu mittels Minimierung der Maße Parameterschätzer ermittelt und die Normalverteilungsfamilie somit auf einen Vertreter reduziert. Weiter wird dessen Dichte dann an abzählbar und endlich vielen Stellen in den Vergleich mit den Realisationen gestellt.

Für die Simulation werden Zufallsstichproben aus den oben angeführten elf Verteilungen gezogen und dafür jeweils die Abstandsmessung zur Normalverteilung durchgeführt. Es werden Stichproben der Größen 10, 100 und 1 000 gewählt. Dabei könnten – gerade bei den verwendeten Teststatistiken, aber auch für die Metriken – für den kleinsten Wert 10 gegebenenfalls Probleme durch unterstellte Konvergenzen der Maße aufgezeigt werden. Bei den Varianten 100 und 1 000 sollten diese Probleme allenfalls nur noch in zu vernachlässigender Größenordnung bestehen. Dann kann durch den Faktor 10 gut geprüft werden, ob weitere Abhängigkeiten von der Stichprobengröße bestehen. Diese können offenbar nicht nur in den Maßen selbst bestehen, auch die Diskretisierung der Verteilungen zum Vergleich mit einer Stichprobe können hiervon betroffen sein.

Es werden für jede Stichprobengröße und jede Verteilung 100 mal Zufallszahlen gezogen, und damit insbesondere die Maße für die jeweils gleichen Daten verglichen. Dass diese Anzahl an Wiederholungen ausreicht zeigt die Streuung der im Folgenden präsentierten Ergebnisse. Für die Ziehung der Zufallszahlen wurden die in den „Standard“-Funktionen des R-Pakets *stats* hinterlegten Methoden verwendet, also die Funktionen `rnorm()`, `rchisq()`, usw.

Die Resultate ergeben sich dann für ein Gitter bestehend aus neun Distanzmaßen, elf Verteilungen, drei Stichprobengrößen und einhundert Wiederholungen, das Ergebnis der Simulation besteht also aus $9 \cdot 11 \cdot 3 \cdot 100 = 29\,700$ reellwertigen Zahlen. Diese werden nun bezüglich der verschiedenen Einstellungen verglichen.

4.3.2 Ergebnisse

Die Abbildung 4.4 auf der Seite 99 zeigt einen beispielhaften Einzel-Lauf der Simulation, wobei nur ein Teil der elf Verteilungen einbezogen ist. Dargestellt sind jeweils Histogramme von Zufallsstichproben und eine Kerndichteschätzung. Der Fokus wird auf die Parameterschätzung gelegt, daher sind die für die fünf Metriken optimalen Normalvertei-

4 Zur Quantifizierung des Normalverteilungsgrades

lungen ergänzt. Außerdem ist die Normalverteilung für Stichprobenmittel und -varianz eingezeichnet, was dem Startwert der Optimierungen entspricht. Diese Schätzer werden, soweit nötig, auch für die Berechnung der Teststatistiken verwendet.

Zu erkennen ist, dass der Unterschied zwischen den Schätzern erwartungsgemäß abhängig von der Verteilung der Daten ist: Für die Normal- und die eher „ähnliche“ t-Verteilung ergeben sich praktisch keine Unterschiede zwischen den angepassten Normalverteilungen. Insbesondere für die recht schiefe χ_2^2 -Verteilung hingegen ergeben sich Unterschiede in den angepassten Dichten. Dass hier keine dieser Anpassungen der Kerndichte entspricht, ist zu erwarten: Die χ_2^2 -Verteilung ist eben insbesondere keine Normalverteilung. Geringer, aber nicht weniger offensichtlich, fallen die Abweichungen bei der Gleichverteilung aus. Bei der diskreten Poisson-Verteilung ergeben sich zwei Gruppen, wobei hier auch der gewählte Gauß-Kernel nicht gut geeignet ist, die Dichte der Daten durch eine Kerndichte zu repräsentieren.

Insgesamt ergibt sich der Eindruck, dass sich die Optimierung vor allem für Daten aus besonders Normalverteilungs-fernen Strukturen in nennenswerter Größenordnung auswirkt. Gerade dann aber ist fraglich, ob diese Unterschiede noch von praktischem Interesse sind. Für die Simulation bedeuten die Erkenntnisse aus der Abbildung, dass die verschiedenen Varianten durch die Optimierung Distanzen zu verschiedenen Normalverteilungen berechnen können. Dabei ist unklar, ob diese Unterschiede auch immer relevant sind. Weitere Betrachtungen zur Optimierung werden unten in Teil 4.5.2 folgen.

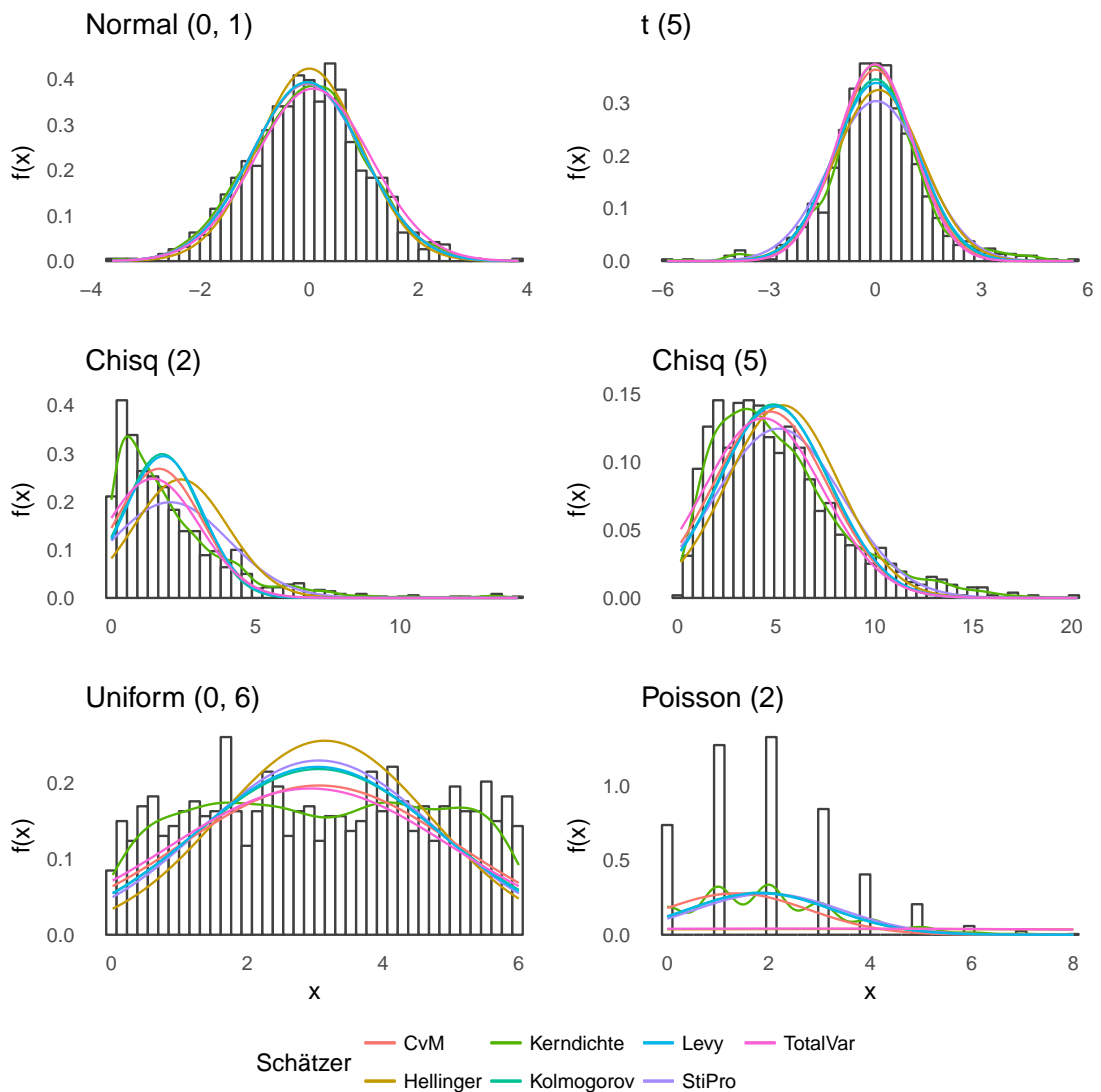


Abbildung 4.4: Sind die Distanzen zur Normalverteilung allgemein zu bestimmen, muss die Verteilungsfamilie für die meisten Varianten auf einen einzelnen Vertreter reduziert werden. Dabei können die Stichprobenschätzer (d. h. hier Mittelwert und Varianz) verwendet werden, oder das jeweilige Maß wird optimiert. Alle Varianten sind hier für die fünf Metriken dargestellt, ergänzt um eine Kerndichteschätzung. Gezeigt werden nur sechs der elf Verteilungen der Simulation, hier für eine Stichprobengröße von 1 000 Beobachtungen

4 Zur Quantifizierung des Normalverteilungsgrades

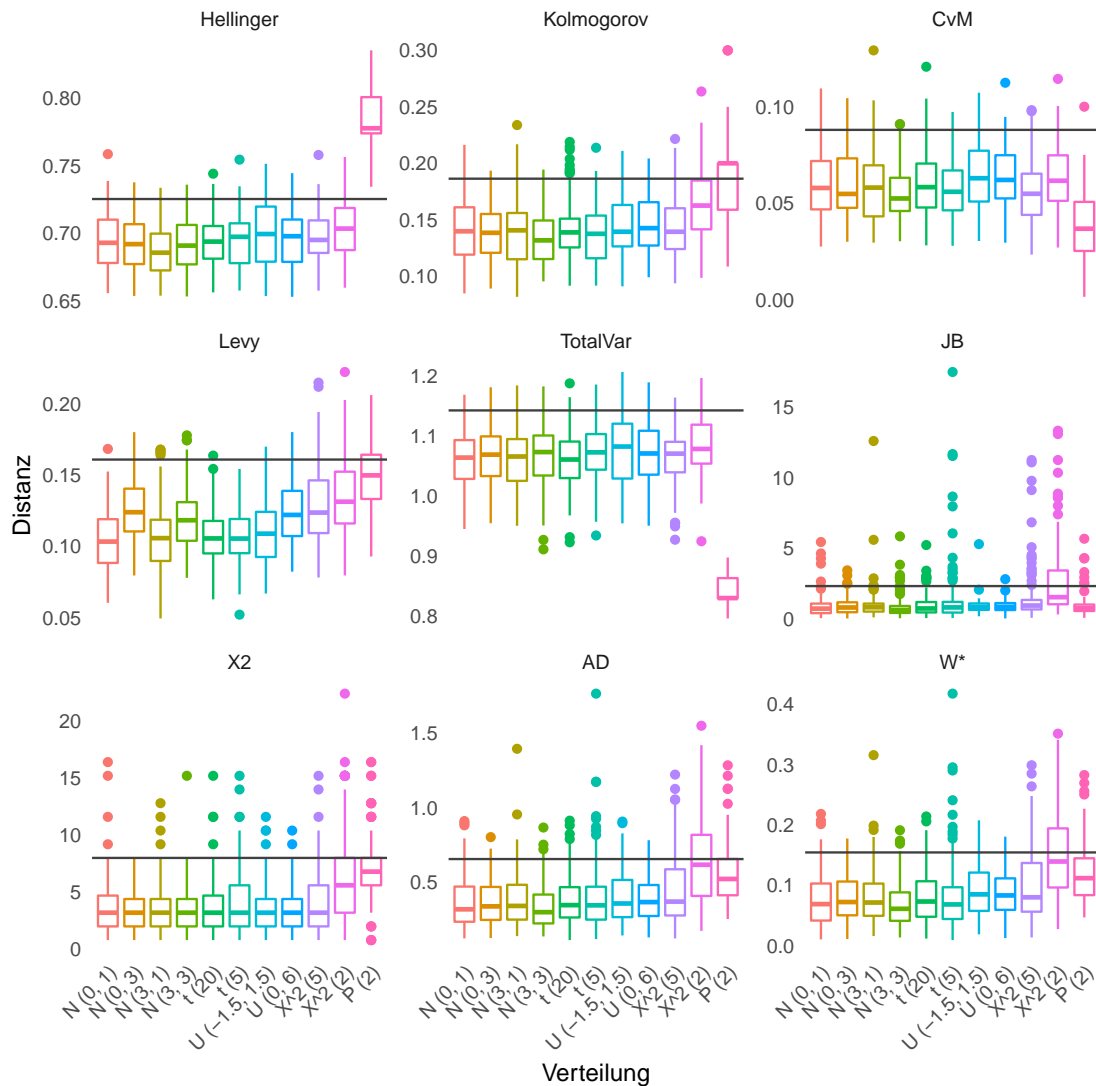


Abbildung 4.5: Die Simulationsergebnisse für die Stichprobengröße 10: Es sind mit allen Methoden kaum Unterschiede auszumachen.

Die Abbildungen 4.5 und 4.6 (sowie die Abbildung 8.2 im Anhang, siehe Seite 172) stellen die Ergebnisse aus der Simulation dar, wobei zwischen den Stichprobengrößen unterschieden wird. Es wird jeweils eine Teilgrafik für jedes Distanzmaß verwendet, die Skalierungen sind entsprechend unterschiedlich und orientieren sich insbesondere an den realisierten Werten, nicht an den gesamten möglichen Intervallen. Eingezeichnet ist jeweils eine horizontale Linie. Diese bildet das 95 %-Quantil ab, welches sich für jeweils alle $4 \cdot 100$ Werte der Normalverteilungen ergibt. Damit soll eine Orientierungshilfe zur

„Trennschärfe“ der Methoden gegeben werden: Idealerweise wären offenbar alle Werte nicht-normaler Daten oberhalb einer solchen Linie.

In der Abbildung 4.5 mit den Resultaten für die Stichprobengröße 10 ist zu erkennen, dass die Unterschiede zwischen einem Datenvektor und der Normalverteilungsfamilie zumindest für kleine Fallzahlen kaum festzustellen sind. Bei derartig kleinen Stichproben sind auch die zwei Freiheitsgrade in Form der zu schätzenden Verteilungsparameter noch relevant. Zunächst fällt auf, dass alle Maße für alle vier Normalverteilungen erwartungs- und forderungsgemäß gleiche Verteilungen bilden, die jeweils vier linken Boxplots lassen keine systematischen Unterschiede zueinander erkennen.

Während die horizontale Anordnung der Verteilungen tendenziell steigende Boxplots erwarten lässt, sind hier kaum Unterschiede zu notieren. Allein die Poisson- und zumeist auch die χ_2^2 -Verteilung fallen augenscheinlich „aus der Reihe“ der anderen grafisch zusammengefassten Ergebnisse. Dabei ist einzig die Hellinger-Distanz in der Lage, den Poisson-Beispielen einen Abstand zuzuordnen, welcher durchweg den für Normalverteilungen üblichen Bereich übersteigt – im Gegenzug erkennt diese Distanz aber kaum anders geartete Abweichungen.

Die kleinsten Werte weisen den Poisson(2)-verteilten Daten die Cramér-von-Mises und die Totalvariations-Distanz zu. Bei der zweitgenannten Methode offenbaren sich zudem weitere Probleme: Während mit Überlegungen zu mathematischen Eigenschaften (s. o.) nur Werte zwischen der Null und der Eins möglich sein sollten, werden hier zu großen Teilen Werte oberhalb dieses Intervalls ausgegeben – die Gründe sind unklar, jedoch in der Diskretisierung bei kleinen Stichproben zu vermuten. Weitere Unterschiede der Maße (oder der zugehörigen Implementierungen) ergeben sich in der Lage der Werte: So sind diese auch für normalverteilte Daten nur teilweise nahe der Null, etwa für Hellinger-Distanzen ist dies gar nicht der Fall (auch wenn hier das Intervall $[0, 1]$ eingehalten wird). Beide Auffälligkeiten könnten jedoch ignoriert werden, wenn sie konsequent auftreten – und sich so dennoch eine Unterscheidungsmöglichkeit zwischen den Verteilungen ergäbe.

Insgesamt ist für kleine Stichproben die offenbar grundsätzlich schwierige Trennung der Verteilungen zu notieren. Dabei fällt keines der Verfahren durchweg positiv auf. Einzig die Hellinger-Distanz trennt die Poisson-Verteilung gut von der Normalverteilung. Die Abweichungen der χ_2^2 -Verteilung erkennen die Teststatistiken tendenziell besser, insbesondere die Variante nach Anderson und Darling. Die Ergebnisse des Totalvariationsabstandes sind zumindest für kleine Stichproben nicht vertrauenswürdig.

Aus der statistischen Testtheorie ist bekannt, dass sich die Power (oder auch Güte, Trennschärfe) eines Verfahrens im Allgemeinen mit steigender Stichprobengröße verbessert. Dieses Prinzip findet sich auch im Vergleich der Abbildung 4.6 für die Ergebnisse der

4 Zur Quantifizierung des Normalverteilungsgrades

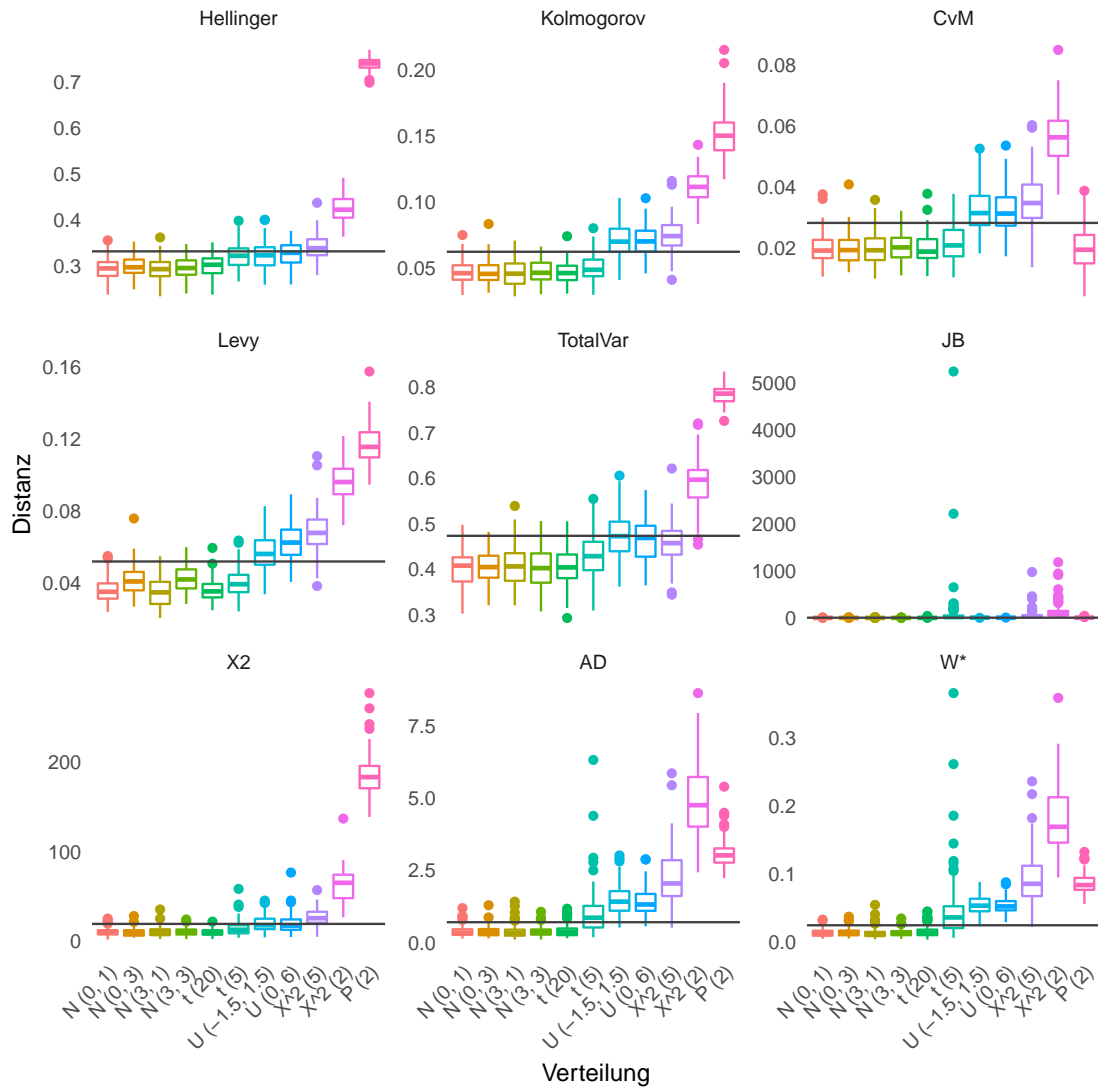


Abbildung 4.6: Die Simulationsergebnisse für die Stichprobengröße 100: Die Maße trennen die Verteilungen unterschiedlich gut.

Simulationen zur Stichprobengröße 100 mit der vorigen Grafik. Je nach Methode können nicht nur die Daten aus Poisson- und χ^2 -Verteilungen relativ sicher von der Normalverteilung getrennt werden: Auch für die anderen Verteilungen können klare Abweichungen notiert werden, so trennen Kolmogorov, CvM und Lévy auch die zweite χ^2 -Verteilung sowie die Gleichverteilungen recht gut von den normalen Verteilungen. Den Teststatistiken nach Anderson/Darling und Shapiro/Wilk gelingt das sogar sehr gut.

Insgesamt hat sich auch die Skala der jeweiligen Größen verändert: Dabei sind für die Metriken insbesondere die aus den Normalverteilungen gewonnenen „Baselines“ nach unten verschoben, zum Teil deutlich. Für die Teststatistiken zeigt sich ein anderer Effekt: Hier führen einzelne bis wenige Datenbeispiele nicht-normaler Verteilungen zu – auch im Vergleich mit dem Bild oben – extremen Abweichungen von der Null.

Für die Simulation mit der Stichprobengröße 1 000 ergeben sich keine qualitativen Unterschiede zu bisherigen Erkenntnissen. Die zugehörige Grafik ist der Vollständigkeit halber im Anhang auf Seite 172 als Abbildung 8.2 zu finden. Die Werte für die Metriken nähern sich hier bei tatsächlichem Vorliegen einer Normalverteilung weiter der Null an; und für feste Verteilungen und feste Maße sinkt die Streuung der Ergebnisse. Der wesentliche Unterschied zu den Resultaten der Stichprobengröße 100 besteht darin, dass für 1 000 Beobachtungen auch die Ergebnisse der t_5 -Verteilung von der Normalverteilungsfamilie getrennt werden können. Im Rahmen der Metriken ist hier die Hellinger-Distanz positiv hervorzuheben.

Insbesondere für die Jarque-Bera-Methode (100 Beobachtungen), aber auch mit Blick auf alle Teststatistiken (1 000 Beobachtungen) ergeben sich mitunter recht schief verteilte Distanzmaße. Das trifft vor allem innerhalb einer Verteilung, zum Teil aber auch für alle Ergebnisse zu. Aus diesem Grund werden in der Abbildung 4.7 die Grafiken der Teststatistiken wiederholt, wobei jeweils eine Vergrößerung des Bereichs geringerer Distanzen enthalten ist. Das Äquivalent für die Metriken ist im Anhang auf Seite 173 als Abbildung 8.3 zu finden, bietet aber nur im Detail neue Erkenntnisse.

Wie mit einer Vergrößerung der Bereiche sofort zu erkennen ist, steigt die Trennschärfe für die Entscheidung Normalverteilung ja/nein mit der Stichprobengröße klar an. Für die Stichproben der Größe 1 000 (rechte Seite der Grafik 4.7) werden die beiden t -Verteilungen zum Prüfstein: Die Werte der vier Normalverteilungen weichen nicht systematisch voneinander ab, alle anderen Verteilungen werden sicher von der Normalverteilung getrennt. Dabei kann eine klare Rangfolge der Verfahren abgeleitet werden.

Die χ^2 -Statistik kann die Nicht-Normalität der t_{20} -Daten nicht erkennen, bei den t_5 -Daten ist die Trefferquote ebenfalls recht gering. Wird weiter nur die t_{20} -Verteilung betrachtet, so folgen auf den Plätzen drei bis eins der AD-, der W^* -Abstand und der JB-

4 Zur Quantifizierung des Normalverteilungsgrades

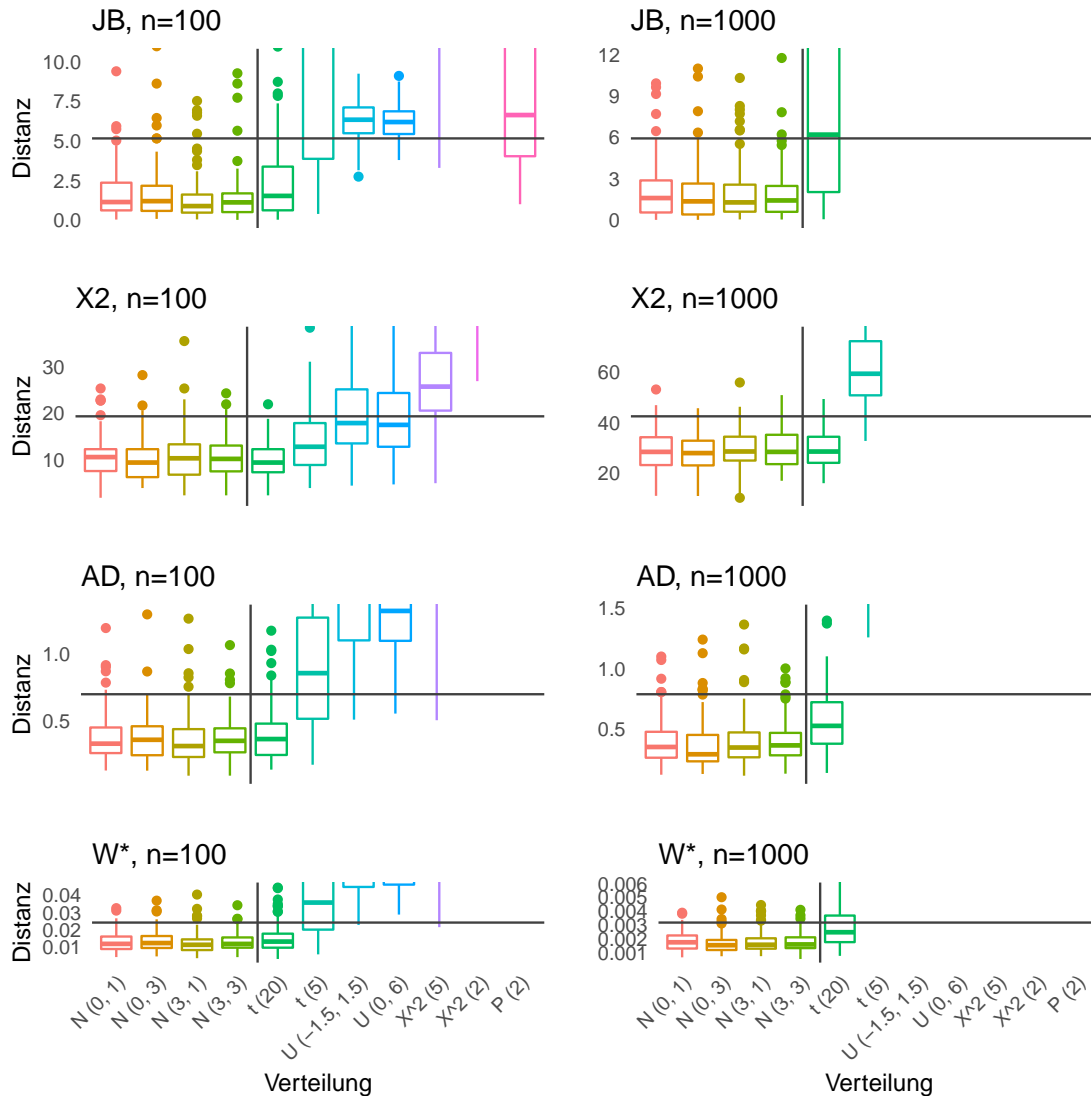


Abbildung 4.7: Auf den vorigen Abbildungen waren die Bereiche entlang der Entscheidungsgrenze „normal oder nicht normal“ mitunter recht klein skaliert. Hier wird eine Vergrößerung gegeben. Das hat zur Folge, dass einige Verteilung-Distanz-Kombinationen hier keine Datenpunkte mehr im Bildausschnitt aufweisen. Dann sind diese aber immer im numerisch größeren, „nördlichen“ Teil der Skala zu verorten.

Abstand. Die Jarque-Bera-Variante ist dabei für gut die Hälfte der 100 Wiederholungen größer als 95 Prozent der Abstände normalverteilter Daten. Dabei ist zu berücksichtigen, dass der Vergleich anhand der dieser Hilfsmarke lediglich punktuell ist, was für den Vergleich der gesamten Verteilung (also hier der Boxplots) nicht gilt.

Für die kleinere Stichprobengröße 100 fällt die JB-Methode jedoch durch eine negative Eigenschaft auf, so kann offenbar die diskrete Verteilung nicht sicher von der Normalverteilung getrennt werden. Für die am schwierigsten zu trennende Verteilung t_{20} liefern alle Methoden qualitativ ähnliche, schlechte Resultate. Insgesamt erscheint die χ^2 -Statistik auch hier im Vergleich als wenig geeignet. Insbesondere bei den Gleichverteilungen, aber auch bei den t_5 -Daten lassen sich leichte Vorteile der W^* -Statistik ausmachen, wobei die Unterschiede besonders zu Anderson-Darling gering sind.

Für größere Stichproben sind damit auf Seiten der Teststatistiken JB , AD und insbesondere W^* als gut zu bewerten. Dabei ist weiter anzumerken, dass nur die Shapiro-Wilk-Variante begrenzt ist, die Werte also alleinstehend am einfachsten interpretierbar sind. Auch ist hier die Skalenverschiebung durch Änderungen der Stichprobengrößen am geringsten.

Diese Liste der „besten“ Methoden ist auf Seiten der Metriken um Kolmogorov/Lévy und Hellinger zu erweitern. Dabei ist die Wahl der Methode abhängig vom inhaltlichen Abstand der Verteilungen: Die Hellinger-Distanz reagiert offenbar am sensibelsten bei den t -Verteilungen, für die Gleichverteilungen fallen die Ergebnisse der Verteilungsfunktionsbasierten Methoden wiederum besser aus.

4.3.3 Vergleich mit theoretischen Größen

Als weitere Eigenschaft der Distanzmaße wird der Abstand der realisierten Werte vom jeweils theoretisch zu erwartenden betrachtet: Dies ist hier möglich, da die den Zufallszahlen zugrunde liegenden Verteilungen bekannt sind. Mit diesen kann nun jeweils die Normalverteilung mit gleichem Erwartungswert und gleicher Varianz der Verteilung gegenübergestellt und so die Distanz zu einer „passenden“ Normalverteilung bestimmt werden. Das bedeutet beispielsweise: Die t -Verteilung mit n Freiheitsgraden hat einen Erwartungswert von 0 und eine Varianz von $n/(n-2)$. Für die t_5 -Verteilung werden die Distanzen zur $N(0, 5/3)$ -Verteilung ermittelt.

Während die Metriken hier per definitionem berechnet werden können, sind bei den Teststatistiken weitere Überlegungen nötig. Insbesondere bei der Anderson-Darling- und der Shapiro-Wilk-Methode sind die jeweiligen Werte nicht trivial zu berechnen (mit Ausnahme der erfüllten Nullhypothese, in diesem Fall ist die Verteilung bekannt oder es

4 Zur Quantifizierung des Normalverteilungsgrades

liegen zumindest tabelliert kritische Werte vor), der Aufwand würde den Nutzen hier übersteigen. Bei der Jarque-Bera-Variante fällt die Berechnung einfacher, da jeweils nur die bekannten Schiefen, Wölbungen und Stichprobengrößen zu verwenden sind. Bei fester Stichprobengröße und fester Regel zur Anzahl der Klassen (s. o.) sind auch die zu berücksichtigenden Einflüsse für die χ^2 -Statistik bekannt.

Die Abbildung 4.8 stellt die zu erwartenden Werte den Realisierungen gegenüber. Dabei findet zum einen die Analyse der Distanzen selbst und implizit auch beispielsweise ihrer Robustheit gegenüber Ausreißern statt. Zum anderen werden hier aber auch und insbesondere gleichzeitig die zur Verfügung stehenden Diskretisierungen und Implementierungen beurteilt. In der Grafik sind jeweils die bereits bekannten Boxplots für einzelne Distanz-Verteilungs-Kombinationen aufgeführt. Die theoretisch zu erwartenden Werte sind jeweils durch ein rotes Kreuz markiert. Die Simulationsdaten werden hier auf die Stichprobengröße 1 000 reduziert. Dass die Distanzschätzer für diesen Wert „gute“ Eigenschaften aufweisen, kann als Minimalforderung angesehen werden: Die Bezeichnung einer Stichprobe als „groß“ ist sicher abhängig vom Bereich der Anwendung. In vielen Gebieten aber ist der Wert 1 000 schon ein aus Kosten- und Zeitgründen nicht zu erreichender Wert, etwaige Approximationen sollten auch für kleinere Werte schon ausreichen.

Bei Interpretationen der Grafik ist berücksichtigen, dass die Simulationsergebnisse Optimierungen der Parameter beinhalten. Damit sind leichte Abweichungen unterhalb der Erwartungen kein Widerspruch zu passenden Werten. Zunächst fällt auf, dass in allen Fällen für die Normalverteilungen eine Null (beziehungsweise für die Chi-Quadrat- und JB-Statistiken ein Wert relativ nahe bei Null) zu erwarten ist. Dieser wird von den Methoden unterschiedlich gut erreicht, was zudem – wie oben gezeigt – abhängig von der Stichprobengröße ist. Augenscheinlich erreichen die beiden Teststatistiken die Null hier am besten bezüglich des relativen Abstandes zum Minimalwert, was aber nicht zuletzt auch auf die relativ großen Skalierungen zurückzuführen ist. Insgesamt ergibt sich ebenso der Eindruck, dass die Realisierungen dieser Varianten am besten den theoretischen Größen gegenübergestellt werden können.

In den mittleren Bereichen (was hier für die Abbildungen jeweils gleichermaßen im horizontalen wie vertikalen Sinne zu verstehen ist) scheinen Hellinger, Lévy und, mit Abstrichen, Kolmogorov gute Resultate zu liefern. Bei der CvM-Methode könnte durch die Optimierungen der Unterschied zwischen einer bestimmten und allen Normalverteilungen hier in der Grafik enthalten sein. Die niedrigen Werte bei der Poisson-Verteilung weichen dennoch auch von inhaltlichen Erwartungen ab. Der Grad der Normalität der χ^2 -Verteilung wird bezüglich der Metriken einzig von der Hellinger-Distanz passend beurteilt.

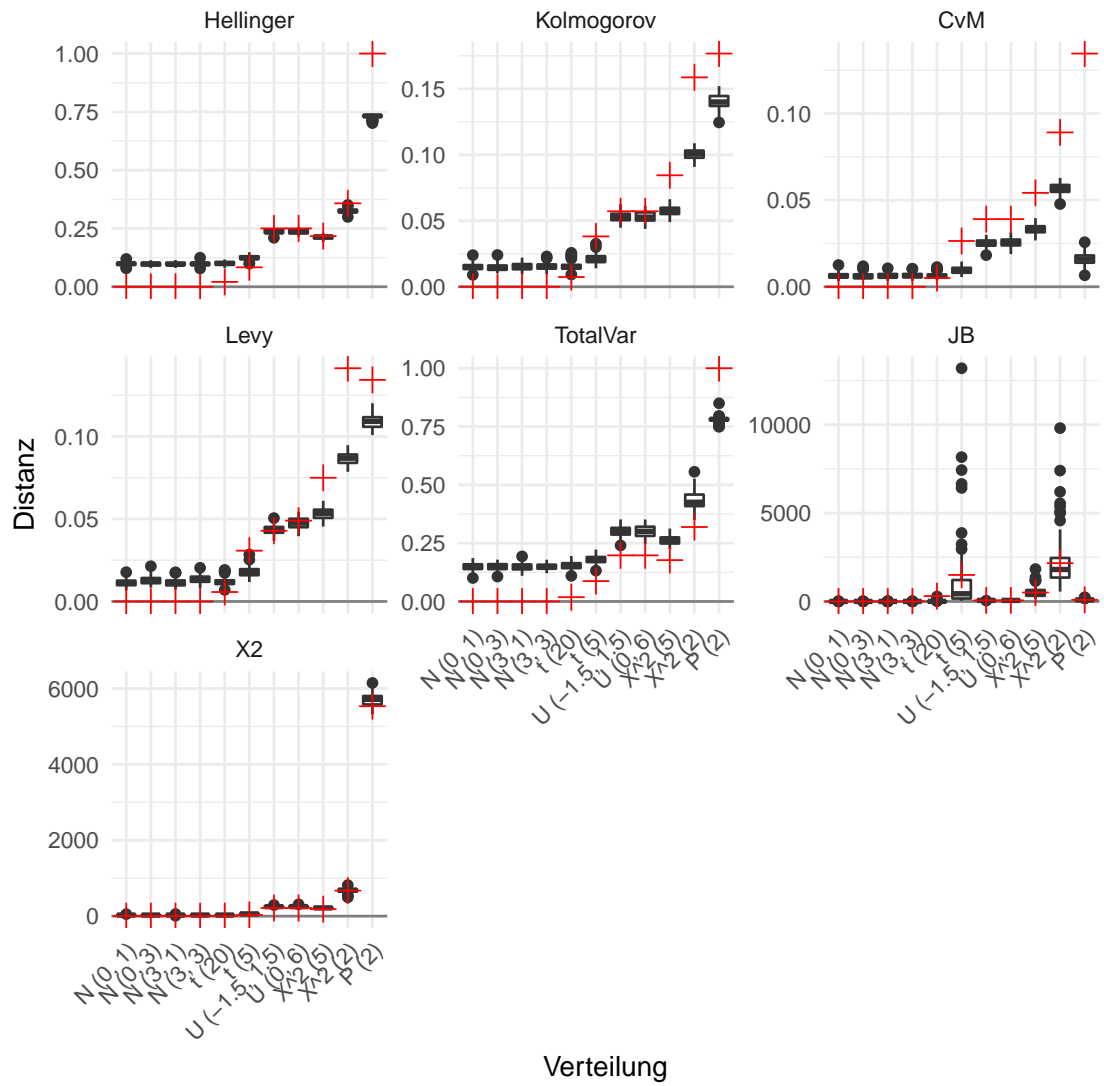


Abbildung 4.8: Die Ergebnisse der Simulation für 1 000 Beobachtungen, ergänzt um die theoretisch zu erwartenden Werte (rot), ohne AD und W^* .

4 Zur Quantifizierung des Normalverteilungsgrades

Auffällig ist hier wiederum der Totalvariationsabstand, dessen Abweichungen nach oben nicht durch die Optimierungen erklärt werden können: Hier ist der gesamte zu erwartende Verlauf nach oben verschoben, mit Ausnahme der Ergebnisse zur Poisson-Verteilung. Dabei ist theoretisch ein Abstand von 1 zu erwarten, die Dichten überdecken sich schließlich nur punktweise. Das wird aber auch vom Hellinger-Abstand nur im Groben erfasst, was bei der Verwendung von Diskretisierungen nicht ganz zu umgehen ist.

Bezogen auf die zu erwartenden Werte weisen alle fünf Metriken eine plausible Rangordnung auf, auch wenn diese nicht immer exakt gleich ist. Für die JB-Methode gilt das nicht unbedingt, so werden die t -Verteilungen hier weiter von der Normalverteilung entfernt als die Gleichverteilungen eingeordnet. Das ist zurückzuführen auf die Symmetrie beider Verteilungstypen und die kleinere Wölbung der Gleichverteilung, ein Effekt der oben bereits kommentiert wurde. Ebenso findet sich bei JB die weniger starke Wölbung der χ_5^2 - gegenüber der t_5 -Verteilung wieder (bei Schiefen von 1.23 und 0).

Die Eigenschaft der Methoden, ihre Werte den theoretisch zu erwartenden anzunähern, kann auch aus einem anderen Blickwinkel betrachtet werden: Unabhängig davon, wie gut der theoretische Wert für eine bestimmte Stichprobengröße realer oder simulierter Daten angenähert wird, kann gefordert werden, dass die Näherung mit steigender Stichprobengröße besser wird. Dieses Prinzip ist als Konsistenz bekannt und in der Schätztheorie elementar. Zur Entscheidung, ob die Verfahren solch ein Verhalten zeigen, wird wiederum die Abbildung 4.8 herangezogen. Für die beiden kleineren Stichprobengrößen sind die Resultate als entsprechende Abbildungen im Anhang unter den Nummern 8.4 und 8.5 auf den Seiten 174 und 175 zu finden. Für alle der sieben diesbezüglich untersuchten Verfahren ergibt sich der Eindruck konsistenten Verhaltens. So bringt die Vergrößerung der Stichproben jeweils eine Verschiebung der Boxplots mit sich: Die Ergebnisse werden differenzierter zwischen den Verteilungen, die Skalierung ändert sich indes. Zugleich ergeben sich dabei „Bewegungen“, wie sie für Konsistenzen zu erwarten sind. Die Distanz nach Cramér und von Mises entfernt sich für die Poisson-Verteilung zwar zunehmend vom Soll, das aber konsequent. Das wird hier als in sich stimmige Eigenschaft aufgefasst, während die große Abweichung von Ist- und Soll-Wert selbst an anderer Stelle berücksichtigt wird.

4.3.4 Folgerungen

Die dargestellten Erkenntnisse sollen nun zusammenfassend bewertet werden. Wie bereits in Teil 4.2.3 erfolgt zunächst eine Diskussion der Eigenschaften der neun Verfahren, anschließend wird eine tabellarische Zusammenfassung gegeben (siehe die Tabelle 4.3).

Die Ergebnisse der Simulationen zeigten für alle Varianten eine Abhängigkeit von der Stichprobengröße, was sowohl die numerische Lage der Ergebnisse als auch die Trennschärfe der Methoden betrifft. Für sehr kleine Stichproben ist keines der Verfahren geeignet. Für kleine Stichproben sind Abweichungen von einer Verteilung offenkundig auch nur sehr schwer zu detektieren. Somit stellt diese Tatsache noch kein grundsätzliches Hindernis zur Verwendung der Methoden dar, zu notieren ist es jedoch. Gegenteilig fällt das Urteil für den Vergleich auf Basis der 1 000-er Stichprobe auf: Hier können alle Verfahren die nicht-normalen Verteilungen angeben, mit Abweichungen für die t_{20} -Daten. Dabei sind die Unterschiede jedoch nur gradueller Natur.

Am besten zur Unterscheidung der Maße eignen sich die Daten mit der Stichprobengröße 100. Die Forderung nach einer guten Trennbarkeit für diese Stichprobengröße kann ohne Weiteres formuliert werden, um ein Maß möglichst allgemein als „gut“ bezeichnen zu können. Das Kriterium der Trennbarkeit von der Normalverteilungsfamilie ist dabei nur vor dem Hintergrund der jeweiligen Verteilungen zu beantworten, wie sich an den Grafiken in Teil 4.3.2 zeigte. Die vier Gruppen Poisson, χ_k^2 , $U(a, b)$ und t_k werden daher jeweils für sich betrachtet. Die Einordnung in eine gute, schlechte oder mittelmäßige erzielte Trennbarkeit erfolgt auf Basis der Abbildungen und kann der folgenden Tabelle entnommen werden.

Eigenschaft	Kolm.	TV	Hell.	Lévy	CvM	χ^2	JB	AD	W
Eignung StiPro 10	○	○	○	○	○	○	○	○	○
Eignung StiPro 100									
Poisson-Vtlg.	●	●	●	●	○	●	◐	●	●
χ_k^2 -Vtlg.	●	◐	◐	●	●	●	●	●	●
$U(a, b)$ -Vtlg.	●	◐	◐	●	●	◐	●	●	●
t_k -Vtlg.	○	○	○	○	○	○	◐	◐	◐
Eignung StiPro 1 000	●	●	●	●	●	●	●	●	●
Konvergenz	◐	○	●	●	○	●	●	x	x
Konsistenz	●	●	●	●	●	●	●	x	x
Implementierung	●	◐	●	●	●	●	●	●	●

Tabelle 4.3: Die Folgerungen aus den Simulationsergebnissen in Zusammenfassung. Die geforderten Eigenschaften sind erfüllt oder nicht, was durch die Symbole ● und ○ gekennzeichnet wird. Die teilweise Erfüllung, markiert mit ◐, wird jeweils im Text erläutert; Eigenschaften mit der Kennzeichnung x sind nicht untersucht worden.

Im vorigen Teil wurden Gleichheit und Unterschiedlichkeit der realisierten und der erwarteten Werte untersucht. Dabei stellte sich heraus, dass auch bei 1 000 Beobachtungen

4 Zur Quantifizierung des Normalverteilungsgrades

nicht alle Maße zum theoretischen Wert passen. Weiter wurde neben der Konvergenz auch die Konsistenz kommentiert, auch diese beiden Eigenschaften sind zusammenfassend in der Tabelle notiert.

Die Eigenschaften der verwendeten Implementierungen enthalten wie bereits erwähnt jeweils zwei bis drei Quellen: Zum einen die Eigenschaften der Methoden selbst, zum anderen die Eigenschaften der gewählten Form der Implementierung. Für einen Teil der Methoden kommen die Eigenschaften der notwendigen Diskretisierung hinzu. Auffälligkeiten ergaben sich beim Totalvariationsabstand, für welchen bei kleinen Stichproben entgegen der Definition Werte oberhalb der Eins ausgegeben werden. Auf welcher Ebene der angeführten Quellen der Eigenschaften sich dies begründet, wird an dieser Stelle nicht weiter untersucht. Für die anderen Methoden waren im Rahmen der hier vorgestellten Arbeiten keine Auffälligkeiten zu notieren, es wird daher – jedoch ohne weitere Prüfung und nach dem Prinzip *in dubio pro reo* – von geeigneten Implementierungen ausgegangen.

4.4 Fazit

In den beiden vorigen Teilen wurden die neun ausgewählten Verfahren für einige beispielhafte Verteilungen mittels verschiedener Kriterien verglichen. Dabei wurde deutlich, dass keines der Abstandsmaße durchweg die beste Wahl zur Behandlung einer spezifischen Fragestellung ist – eine Art *gleichmäßig* bestes Verfahren kann nicht herausgestellt werden. Sowohl die Diskussion der theoretischen Eigenschaften also auch die Ergebnisse aus der Simulationsstudie zeigen, dass die Wahl eines optimalen Verfahrens auch von der Anwendung abhängen kann, etwa wenn verschiedene Formen der Nicht-Normalität verschiedene Auswirkungen haben. Dennoch kann zusammenfassend diskutiert werden, welche Methode zumindest die *durchschnittlich* beste darstellt, welche eventuell gänzlich ungeeignet ist oder welche Auswahl von Maßen die Menge von neun Methoden verlustfrei auf eine praktikablere Größe reduzieren kann.

Eine Zusammenstellung der Ergebnisse bildet die Tabelle 4.4. Diese stellt als Kombination der beiden vorigen Tabellen 4.2 und 4.3 die bisherigen Erkenntnisse zusammen. Es wird dabei auf die Charakteristiken verzichtet, welche keine Unterschiede zwischen den Verfahren aufzeigen. Die letzten beiden Zeilen enthalten eine Gesamtbewertung der neun Methoden, welche durch simples Auszählen der vorher vergebenen „Punkte“ gebildet wurde. Die ideale Gewichtung der einzelnen Aspekte ist dabei jedoch nicht eindeutig und je nach Anwendungsfall zu berücksichtigen, hier wurde eine gleiche Gewichtung aller Merkmale angesetzt.

Eigenschaft	Kolm.	TV	Hell.	Lévy	CvM	χ^2	JB	AD	W
Theorie und Daten	●	◐	◐	◐	●	◐	◐	●	○
Symmetrie	●	●	●	●	●	○	●	○	○
Beschränktheit	●	●	●	●	●	○	○	○	●
Eignung StiPro 100									
Poisson-Vtlg.	●	●	●	●	○	●	◐	●	●
χ_k^2 -Vtlg.	●	◐	◐	●	●	●	●	●	●
$U(a, b)$ -Vtlg.	●	◐	◐	●	●	◐	●	●	●
t_k -Vtlg.	○	○	○	○	○	○	◐	◐	◐
Konvergenz	◐	○	●	●	○	●	●	x	x
Implementierung	●	◐	●	●	●	●	●	●	●
Insgesamt ●	7	3	5	7	6	4	5	5	5
Insgesamt ○	1	2	1	1	3	3	1	2	2

Tabelle 4.4: Die Kombination der nicht-konstanten Zeilen aus den beiden Tabellen 4.2 und 4.3 führt zu einer stark verdichteten, aber übersichtlichen Zusammenfassung der erkannten Unterschiede. Die unteren Zeilen enthalten die Häufigkeiten der Eigenschaften, was eine gleiche Gewichtung aller Kriterien impliziert.

Offenbar ist kein Verfahren durchweg besser als alle anderen: Zwar weisen die beiden Metriken nach Kolmogorov und Lévy, also zwei der auf Verteilungsfunktionen beruhenden Methoden, die meisten erfüllten Eigenschaften auf. Die Bereiche, in denen diese beiden Verfahren jedoch Schwächen aufweisen, werden von anderen Methoden zumindest teilweise erfüllt.

Dominanz kann hier zweckmäßig folgendermaßen definiert werden: Beim Vergleich zweier Verfahren wurde das dominierende Verfahren in allen Bereichen mindestens genau so gut bewertet wie das dominierte Verfahren, sowie in mindestens einem Bereich besser. Dann dominiert dem vorigen Absatz gemäß keine Methode alle anderen. Jedoch werden einige Methoden dominiert und sind damit obsolet. Das betrifft den Totalvariationsabstand (dominiert durch Kolmogorov, Hellinger und Lévy), den Hellinger-Abstand (durch Lévy), den Cramér-von-Mises-Abstand (durch Kolmogorov) und die χ^2 -Statistik (durch Lévy und Anderson-Darling).

Für die übrigen fünf Varianten stellt sich die Frage nach der Gewichtung der spezifischen Vor- und Nachteile: Ist etwa die Sensibilität gegenüber der t-Verteilung von besonderem Interesse, sollte eine der Teststatistiken gewählt werden. Können diskrete Verteilungen von vornherein ausgeschlossen werden, verlieren die CvM- und JB-Kennzahlen einen Nachteil, und so weiter. Insgesamt erscheinen aber Kolmogorov und Lévy als die

4 Zur Quantifizierung des Normalverteilungsgrades

besten Varianten, gefolgt von JB. Mit dem Kolmogorov- und dem Jarque-Bera-Verfahren weisen zwei dieser Methoden zudem die positiven Eigenschaften einer anschaulichen Grundidee und einer vergleichsweise einfachen Berechnungsweise auf.

Statt nur einer Größe kann auch die Verwendung von zwei Methoden in Betracht gezogen werden, was ebenfalls als praktikabel erscheint. Werden die beiden letztgenannten Methoden Kolmogorov und Jarque-Bera gemeinsam betrachtet, so ist jede der tabellierten Forderungen maximal erfüllt („maximal“, da kein Verfahren für die t -Verteilung mehr als die Einordnung „teilweise erfüllt“ erreicht). Von allen 36 Möglichkeiten, zwei der Verfahren zu kombinieren, wird nur für drei Paare dieses maximal mögliche Ergebnis erzielt: Hellinger und Anderson-Darling, Lévy und Anderson-Darling sowie, bereits genannt, Kolmogorov und Jarque-Bera (und entsprechend dominieren dieses drei Paarungen alle anderen). Da die Kombination aus Kolmogorov-Distanz und JB-Test auch für die beiden Verfahren jeweils einzeln betrachtet die besten Werte der drei genannten Paare aufweist, kann sie als die eindeutig beste bezeichnet werden.

4.5 Ergänzungen

Für die Vereinfachung der praktischen Anwendung dieser beiden ausgewählten Maße sollen an dieser Stelle noch zwei Punkte untersucht werden: Zum einen wurden oben Abhängigkeiten von der Stichprobengröße notiert, welche eine stand-alone-Interpretation der Werte verhindern. Zum anderen enthalten die vorgestellten Ergebnisse Parameteroptimierungen. Diese setzen eine vielmalige Auswertung des Distanzmaßes voraus und erhöhen den Rechenaufwand damit im Vergleich zur einmaligen Berechnung der Abstände für die $N(\hat{\mu}, \hat{\sigma}^2)$ -Verteilung enorm. Es ist zu klären, ob dieser Mehraufwand lohnenswert ist.

4.5.1 Zur Stichprobengröße

Die Interpretationen der Werte sind wie gezeigt nur bei Kenntnis des Stichprobenumfangs ohne Weiteres möglich, da auch die Werte für normalverteilte Daten von diesem abhängen. Für die Kolmogorov-Distanz und die Jarque-Bera-Statistik wird die Abhängigkeit an dieser Stelle weiter untersucht. Die Abbildung 4.9 wiederholt die Ergebnisse der Simulationen bei Gegenüberstellung der Stichprobengrößen. Da keines der Maße – den Forderungen und Erwartungen entsprechend – Unterschiede zwischen den vier gewählten Normalverteilungen angezeigt hat, werden diese im Weiteren der Übersichtlichkeit halber nur durch die Standardnormalverteilung repräsentiert.

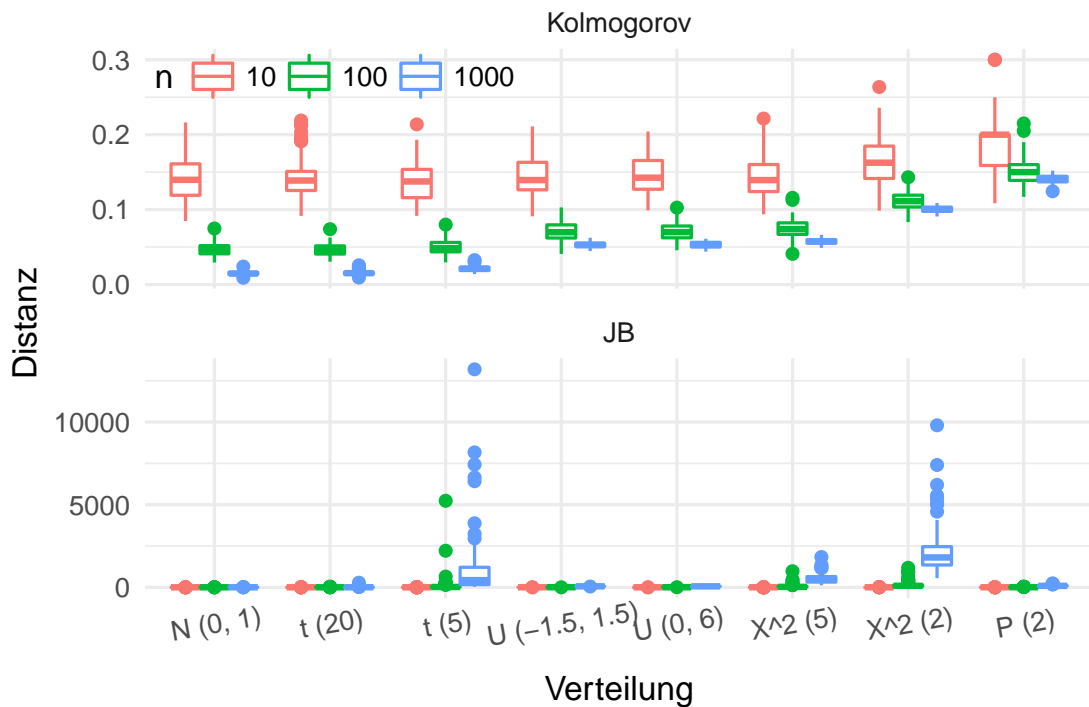


Abbildung 4.9: Die Kolmogorov-Distanzen und Jarque-Bera-Statistiken aus den Simulationen, für verschiedene Verteilungen und verschiedene Stichprobengrößen n .

Auf der Basis dieser Abbildung und auch der vorherigen Erkenntnisse wird die Teststatistik von Jarque und Bera logarithmiert betrachtet: Die mitunter extrem schiefen Verteilungen der Größe führen zu Skalierungen, welche neben einzelnen Ausreißern kaum Systematiken erkennen lassen (vgl. dazu bspw. auch Abbildung 4.6 auf Seite 102). Da der Logarithmus eine streng monotone Funktion ist, führt dies bezüglich der Ordnung mehrerer Distanzen nicht zu Verzerrungen, die Skalierungen werden jedoch praktikabler. Da sich für perfekt normalverteilte Daten ein Wert von Null ergibt, der Logarithmus aber dort nicht definiert ist, wird die JB-Statistik zunächst verschoben:

$$LJB = \log(JB + 1)$$

Diese Verschiebung führt auch dazu, dass die Null wieder den minimalen Abstand bildet, da $\log(0 + 1) = 0$ ist. Alternativ wurde auch die Verwendung der radizierten JB-Statistik in Betracht gezogen, welche aber nicht zielführend ist.

4 Zur Quantifizierung des Normalverteilungsgrades

Die Abbildung 4.9 zeigt (auch bei hier nicht dargestellter Vergrößerung der unteren JB-Skala) ein gegenläufiges Verhalten der beiden Größen an: Während die optimierte Kolmogorov-Distanz gerade bei einer Normal- und ähnlichen Verteilungen große Unterschiede zwischen den Stichprobengrößen aufweist, sind die JB-Werte hier am nächsten beieinander. Das folgt aus der Eigenschaft als Teststatistik, welche unter der Nullhypothese approximativ einer Chi-Quadrat-Verteilung mit zwei Freiheitsgraden folgt. Unterschiede zwischen den drei Boxplots zur Normalverteilung können dieser Approximation sowie Zufallsstreuungen zugeschlagen werden.

Bei der Kolmogorov-Distanz führt die Multiplikation der Ergebnisse mit \sqrt{n} zu gleichen Werten für die normalverteilten Stichproben. Dieser Faktor wird auch beim Kolmogorov-Smirnov-Test genutzt, wie Hartung et al. (2005) auf der Seite 184 entnommen werden kann. Der Nachteil der Betrachtung von

$$K^* = \sqrt{n} K = \sqrt{n} \sup_x |F_n(x) - G_n(x)|$$

besteht in der Verlagerung der Stichprobenabhängigkeit auf die Schranken: Während $K \in [0, 1]$ ist, ist $K^* \in [0, \sqrt{n}]$. Das wird aber als leichter zu interpretierende Variante aufgefasst. Zur Orientierung: Es ist $\sqrt{10} \approx 3.16$, $\sqrt{100} = 10$ und $\sqrt{1000} = 10\sqrt{10} \approx 31.6$. Die Abstände sind also durchweg recht weit entfernt von dieser Schranke, was entsprechend auch bei den Auswertungen der Simulationsstudie im Teil 4.3 bereits zu beobachten war. Die beiden transformierten Größen sind in der Abbildung 4.10 beispielhaft dargestellt.

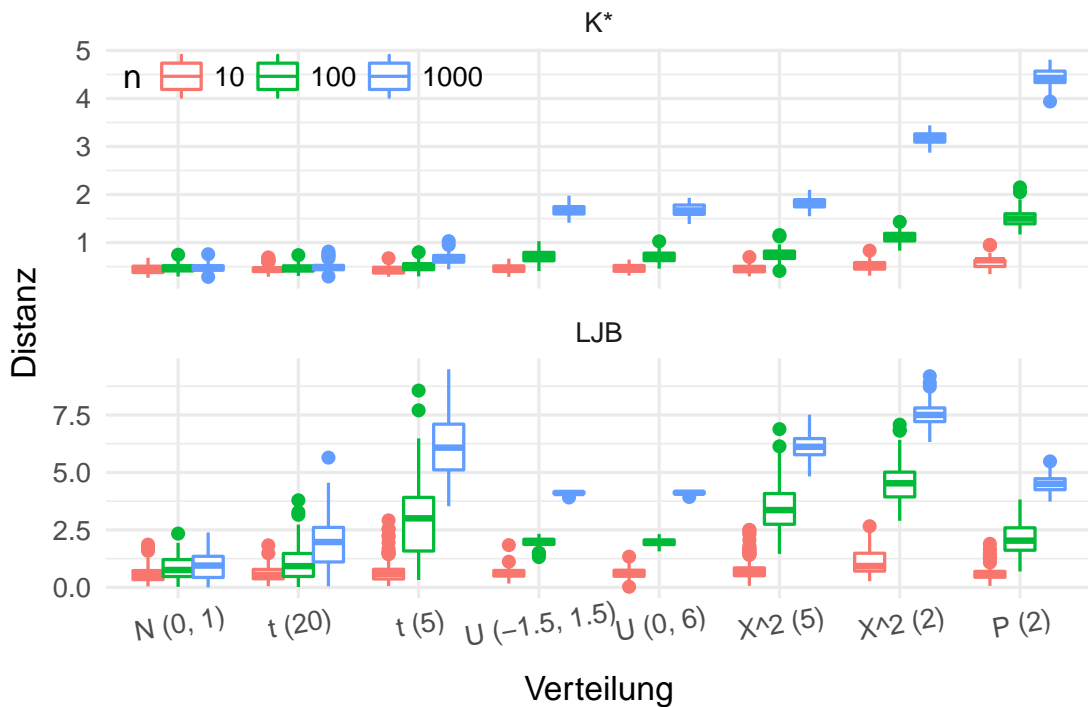


Abbildung 4.10: Mit den beiden angegebenen Transformationen sind zum einen die Skalierungen praktikabler angelegt, zum anderen bestehen keine systematischen Unterschiede für die Werte unterschiedlich großer, normaler Stichproben.

4.5.2 Zur Optimierung

Bei den bisher betrachteten simulierten Kolmogorov-Distanzen wurden die Parameter der Normalverteilung als Variablen aufgefasst. Damit wurde die Metrik dann optimiert. Von Interesse ist nun jedoch die Notwendigkeit dieser Optimierung, beziehungsweise gegebenenfalls die Größenordnung des Vorteils gegenüber der Verwendung geschlossen darstellbarer Parameterschätzer. Für die Jarque-Bera-Variante stellt sich diese Frage nicht, da hier gerade ein von den ersten beiden Momenten unabhängiges Vorgehen gewählt wird.

In einer kleinen Simulationsstudie wurden aus verschiedenen Verteilungen Zufallsstichproben vom Umfang 100 gezogen und die K^* -Distanz jeweils sowohl für die Startwerte der Optimierung (das heißt hier Mittelwert und Stichprobenvarianz), als auch für die optimierte Variante festgehalten. Dieses Vorgehen wird 50 mal wiederholt. Die folgenden beiden Abbildungen 4.11 und 4.12 verdeutlichen die Unterschiede, welche sich durch die Optimierung im Vergleich zu den Startwerten („direkte Bestimmung“) ergeben. In der

4 Zur Quantifizierung des Normalverteilungsgrades

ersten Grafik ist dabei der Unterschied für jede der Stichproben dargestellt, wobei die Distanzen für die Start- und die optimierten Werte jeweils durch eine Gerade verbunden sind (jede Teilabbildung enthält also $3 \cdot 50$ Geraden). Verwendet wurden hier die oben bereits genutzten Verteilungen, wobei wiederum nur Ergebnisse für eine statt für vier Normalverteilungen gezeigt werden.

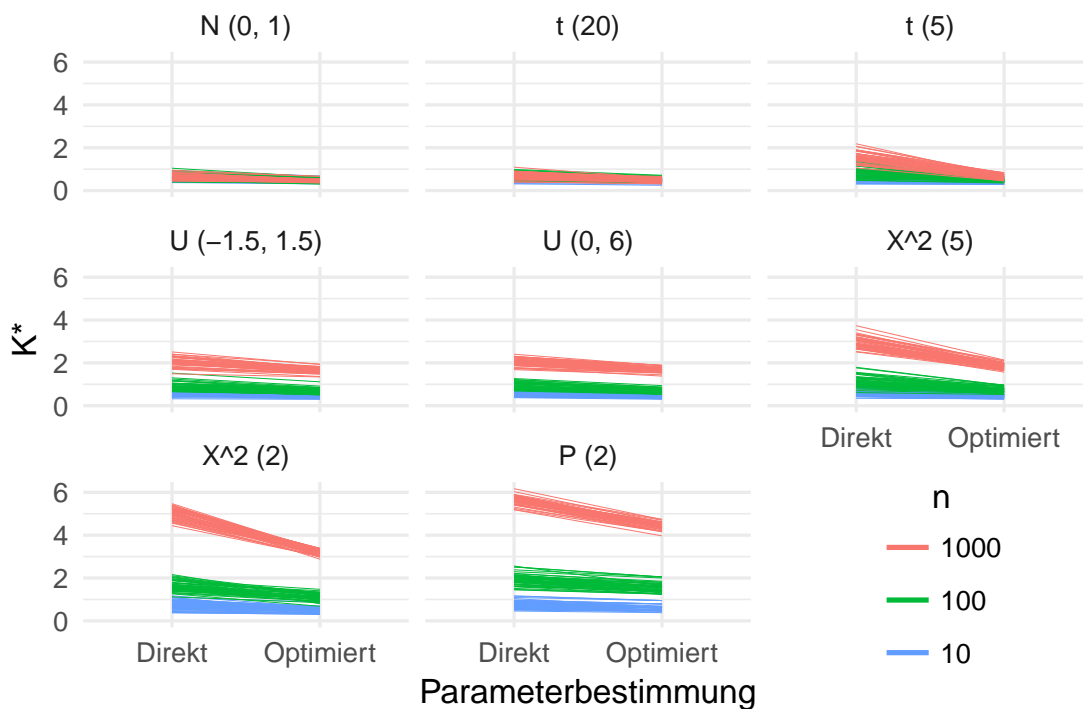


Abbildung 4.11: Die K^* -Distanzen zur Normalverteilung für jeweils 50 Wiederholungen verschiedener Stichprobengrößen. Dargestellt sind die Werte, welche sich für die $N(\hat{\mu}, \hat{\sigma}^2)$ -Verteilung ergeben (geschlossene oder „direkte“ Parameterbestimmung), sowie die Distanzen für die nach Optimierung nächste Normalverteilung. Jede Gerade verbindet diese beiden Werte für die jeweils gleiche Zufallsstichprobe.

Auch unter Berücksichtigung der unterschiedlichen Intervallbreiten ($[0, \sqrt{n}]$, s. o.) fallen Verbesserungen der Abstände ins Auge, wobei diese recht unterschiedlich stark ausfallen: Das Optimierungspotential steigt mit der Distanz, welche sich für die Startparameter berechnet. Damit können vor allem bei den χ^2 - und Poisson-Verteilungen recht starke Verbesserungen erreicht werden, teilweise wird der Distanzwert durch die Optimierungen um ein Drittel und mehr gesenkt. Ebenso steigt der K^* -Wert mit der Stichprobengröße, sofern keine Normalverteilung vorliegt – und mit der Entfernung vom Minimalwert 0 dann

ebenfalls das Optimierungspotential. Dieser Effekt verwundert nicht: Zum einen sind die kleineren Distanzen bereits näher am minimal möglichen Wert von 0. Zum anderen weisen Mittelwert und Stichprobenvarianz gute Eigenschaften wie etwa Erwartungstreue als Schätzer für Erwartungswert und Varianz auf, sofern tatsächlich normalverteilte Daten vorliegen – gerade deswegen sind sie gebräuchliche Schätzer. Steigt der Abstand von gegebenen Daten zur Normalverteilung, fällt auch die Argumentation für deren Verwendung entsprechend schwerer: Es sollen dann optimale Parameter für die Anpassung einer Normalverteilung gefunden werden, obwohl keine Normalverteilung vorliegt. Begriffe wie Erwartungstreue oder Konsistenz sind dann nicht sinnvoll definiert.

Klar ist also, dass die Optimierungen in vielen Fällen zu numerisch anderen Resultaten führen; der beste Vertreter der Normalverteilungsfamilie also nicht immer durch Parameter nahe bei den hier verwendeten geschlossenen Schätzern gekennzeichnet ist. Offen ist jedoch, ob die Unterschiede auch von Relevanz sind: Sollte die Optimierung der Distanzen eine streng monotone Abbildung darstellen, wären die Ergebnisse mit und ohne Optimierung quantitativ zwar unterschiedlich, qualitativ könnten aber keine Unterschiede ausgemacht werden. Die Abbildung 4.12 zeigt aus den Simulationen resultierende Boxplots, wobei wie oben die Linien für das jeweilige 95 %-Prozent-Quantil der Normalverteilungs-Distanzen eingezeichnet sind (auch wenn diese Orientierungsgröße bei nur 50 Beobachtungen jedoch einer recht großen Varianz unterliegt).

Wie bereits vorher klar wurde, zeigt auch diese Darstellung ein zum Teil deutliches Absinken der Distanzmaße durch die Optimierungen. Das bedeutet, dass die Optimierung einen sichtbaren Effekt erzielt hat, und nicht nur die Startwerte oder Werte sehr nahe bei diesen wieder ausgegeben werden. Andererseits wird hier aber auch klar, dass die Ordnungen aus der direkten Parameterschätzung erhalten bleiben: Zwar auf anderem Niveau, ist auch nach den Optimierungen das gleiche „Muster“ wie vorher zu erkennen. Die Boxplots beispielsweise zu den t_5 -Daten heben sich vor den Optimierungen sogar besser von denen der Normalverteilung ab (für $n \geq 100$). Es handelt sich bei letzterer jedoch um eine Beobachtung, welche an dieser Stelle jedoch nicht von zufälligen Einflüssen getrennt werden kann und somit nicht interpretiert werden sollte.

4 Zur Quantifizierung des Normalverteilungsgrades

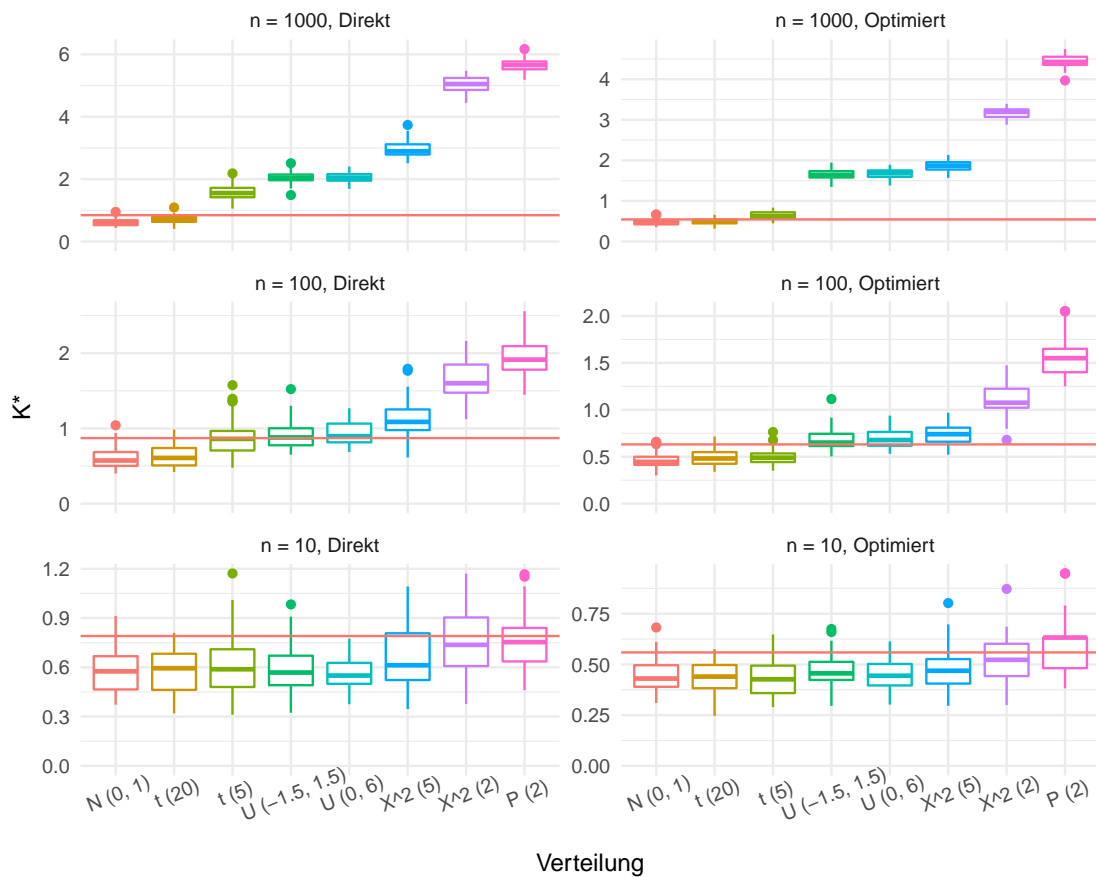


Abbildung 4.12: Die bereits aus der vorangegangenen Abbildung bekannten K^* -Distanzen in anderer Darstellung: Auf der linken Seite die Distanzen zur Normalverteilungsfamilie in Repräsentation durch die Parameterschätzer Mittelwert und Stichprobenvarianz, auf der rechten Seite in Repräsentation durch optimierte Vertreter. Die horizontale Linie markiert jeweils das 95 %-Quantil der Distanzen der $N(0,1)$. Zu beachten sind die unterschiedlichen Skalierungen der Ordinatenachsen bei sonst (zeilenweise) sehr ähnlichen Mustern.

Insgesamt kann mit diesen Ergebnissen festgehalten werden, dass die Optimierung für theoretische Betrachtungen durchaus von Interesse sein kann: Die beste Annäherung einer Stichprobe durch eine Normalverteilung wird nicht ohne Weiteres durch die genannten Schätzer erreicht. In der Praxis sind diese Unterschiede jedoch nicht relevant, so dass auf die Optimierungen verzichtet werden kann. Damit kann Rechenzeit sowie eine mögliche Fehlerquelle eingespart werden.

In diesem Zusammenhang interessant ist eine Erkenntnis von A. Kagan und anderen, welche Patel und Read (1996) auf der Seite 99 folgendermaßen zusammenfassen:

„It turns out that, among all continuous distributions depending on location parameter, and having certain regularity conditions, the location parameter is estimated worst of all in the case of normal samples.“

Ob dieser Sachverhalt einen Einfluss auf den Wert der Maße hat, und wie groß dieser gegebenenfalls ausfällt, kann in weiteren Untersuchungen betrachtet werden. Alternative, geschlossene Parameterschätzer können also in Betracht gezogen werden, um die Berechnung der Maße zu verbessern, ohne den Rechenaufwand deutlich erhöhen zu müssen.

4.6 Zusammenfassung

Die in den vorangegangenen Teilen beschriebenen Untersuchungen hatten das Ziel, eine möglichst gute Methode zur Messung der Normalverteilungsnähe einer Stichprobe zu identifizieren.

Die neun hier ausgewählten Kandidaten bestehen aus fünf Metriken und vier Teststatistiken, wobei zwei der Metriken auf der Dichte- und drei auf der Verteilungsfunktion beruhen. Ein nicht zu vernachlässigender Punkt bei der Auswahl und Anwendung von Metriken stellt die Übertragung auf die Verarbeitung von Stichprobenwerten dar. Hierzu wurden die im verwendeten R-Paket `distrEx` implementierten Diskretisierungen der stetigen Verteilungsfunktionen verwendet. Für nahezu alle Varianten (das heißt für Metriken und Teststatistiken) muss die Familie der Normalverteilungen auf einen einzelnen Vertreter reduziert werden. Dazu wurden die Distanzen zunächst als Funktionen der Normalverteilungsparameter aufgefasst und optimiert.

Theoretische Eigenschaften wie die Gemeinsamkeiten und Unterschiede der Konzepte sowie die zu erwartenden Werte für beispielhafte statistische Verteilungen wurden im Teil 4.2 diskutiert. Neben Minimal-Forderungen wie der Ausgabe des kleinstmöglichen Wertes bei tatsächlichem Vorliegen einer Normalverteilung sind Beschränktheit und Symmetrie der Maße anzustrebende Eigenschaften.

4 Zur Quantifizierung des Normalverteilungsgrades

Als zentrale Entscheidungshilfe bei der qualitativen Unterscheidung der Methoden dient eine Simulationsstudie (s. Teil 4.3). Für Zufallsstichproben verschiedener Verteilungen und verschiedener Größen wurden die Maße unter verschiedenen Gesichtspunkten verglichen. Zusammengefasst mit vorherigen Erkenntnissen ergibt sich schließlich kein eindeutiges Bild: Keine Methode ist allen anderen in sämtlichen Kriterien gleichwertig oder überlegen. Damit stellt sich für die Benennung eines Verfahrens als *das beste* die Frage nach der Gewichtung der einzelnen Kriterien, welche sich je nach Anwendungsfall unterscheiden kann: So kann beispielsweise die gute Trennung der Normalverteilung von diskreten Verteilungen in einigen Anwendungen elementar sein, während in anderen Fällen diskrete Stichproben a priori ausgeschlossen werden können.

In jedem Fall als nicht relevant können der Totalvariations-, der Hellinger- und der Cramér-von-Mises-Abstand sowie die Teststatistik des Chi-Quadrat-Tests angesehen werden: Für diese vier Verfahren (in der hier verwendeten Form inklusive Implementierung) findet sich in jedem Fall ein besseres anderes Vorgehen. Während die Anderson-Darling- und Shapiro-Wilk-Teststatistiken auch als geeignet eingeordnet werden können, scheinen die Distanzen von Kolmogorov und Lévy sowie die Teststatistik nach Jarque und Bera am besten geeignet zu sein. Als Empfehlung wird schließlich die gemeinsame Verwendung von Kolmogorov-Distanz und JB-Teststatistik herausgestellt.

Diese beiden Methoden wurden im Teil 4.5 eingehender betrachtet. Dabei konnten die Vorteile von Transformationen der Art $LJB = \log(JB + 1)$ und $K^* = \sqrt{n} K$ herausgestellt werden. Weiter konnte für K^* gezeigt werden, dass die Optimierungen der Normalverteilungsparameter zwar zu quantitativen Unterschieden der Abstandsmessungen führen, diese qualitativ aber nicht relevant sind. Für die Berechnung von LJB stellt sich diese Frage konstruktionsbedingt nicht.

Die Bindung der Vorgehensweise an eine schnelle Verfügbarkeit in der Programmiersprache R war für die Durchführung der dargestellten Untersuchungen Voraussetzung. Dabei ist eben diese Voraussetzung, trotz der großen Verbreitung von R gerade im wissenschaftlichen Bereich, offenbar nicht objektiven, sondern rein praktischen Gründen anzulasten. Dies gilt auch für die Diskretisierungen der Metriken. Weitergehende Untersuchungen können demnach zusätzliche Methoden, oder die hier vorgestellten Kenngrößen in variiertes Umsetzung betrachten.

Ebenfalls für weitergehende Untersuchungen bieten sich zwei naheliegende Problemstellungen an: Die hier untersuchte Frage nach dem Abstand zur Normalverteilung lässt sich leicht auf die verallgemeinerte Situation eines Abstands zu einer beliebigen Verteilung übertragen. Je nach Methode ist die Übertragung der Grundidee dabei ohne Weiteres möglich (bspw. χ^2) oder erscheint zunächst eher aufwendig (Shapiro-Wilk). Eine Erweite-

rung auf *beliebige* Verteilungen könnte insbesondere auch mehrdimensionale Verteilungen umfassen, für welche sich die hier zugrunde liegende Frage nach dem Grad der Nähe zu einer bestimmten Verteilung oder einer Verteilungsfamilie ebenfalls formulieren lässt.

Wichtig ist im Rahmen der vorliegenden Arbeit jedoch, dass der Grad der Normalverteilung einer gegebenen Stichprobe durch die Verwendung der Kennzahlen K^* und LJB beurteilt werden kann. Damit kann nun überprüft werden, ob eine Transformation eine Stichprobe näher an eine Normalverteilungsannahme heranrückt, oder sie weiter davon entfernt. Es lassen sich also verschiedene Einstellungen der Parameter der auf der Seite 65 beschriebenen Transformationen miteinander vergleichen. Solche Gegenüberstellungen werden im folgenden Kapitel vorgenommen.

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

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* deutlich 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 diesem Kapitel näher beleuchtet werden: Die Anwendung des ZGWS für gleitende Mittelwerte einer Zeitreihe. Unter anderem im Rahmen des auf der Seite 65 vorgeschlagenen Prozedere zur Bildung einer Kontrollkarten-Statistik können Glättungen durchgeführt werden, deren Verteilung dann bekannt ist. Auch wenn kein Wissen über die Verteilung der Zeitreihenelemente selbst vorliegt, stehen dann für die Mittelwerte verschiedenste statistische Werkzeuge zur Verfügung, welche Verteilungsannahmen voraussetzen.

Der Zentrale Grenzwertsatz liegt in mehreren Formulierungen vor. Eine kurze Übersicht findet sich bei Patel und Read (1996) im Kapitel sechs, 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“ (a. a. O., S. 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 der 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 $\sigma^2 := V(X_1) > 0$

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

und dem arithmetischen Mittel $\bar{X} = 1/n \sum_{i=1}^n X_i$ gilt demnach für $n \in \mathbb{N}$ unabhängige und identisch verteilte Zufallsvariablen X_i :

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

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

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 im vorigen Kapitel vorgestellte Kolmogorov-Metrik verwendet werden (s. S. 75 f.). Der ZGWS bedeutet damit auch, dass

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

Es gibt zahlreiche Erweiterungen und Übertragungen dieses prominenten Grenzwertsatzes auf komplexere Anwendungen als das hier zunächst angenommene Szenario unabhängig und identisch verteilter Beobachtungen. Ein Teil dieser veränderten Annahmen an die Datensituation wird unten im Teil 5.2 weiter beleuchtet.

5.1 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 w , die Schrittweite s und die beiden Potenzen d und k variabel, weiter werden mögliche Einflüsse der Stichprobengröße $n \in \mathbb{N}$ berücksichtigt.

5.1.1 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 127 oder der Abbildung 5.1 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 berechnet, die Ergebnissen des vorigen Kapitels nutzend als logarithmierte Jarque-Bera-Statistik sowie nach Kolmogorov mit Berücksichtigung der Stichprobengröße. Als Ankerpunkt für Vergleiche soll die Abbildung 5.2 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 Abbildung 3.13 zeigte auf der Seite 67 ein Beispiel des vorgeschlagenen Verfahrens für Daten aus einer χ_5^2 -Verteilung. Die beiden Maße betragen dabei $K^* = 10.5$ 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 $LJB = 0.39$.

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 für eine Fensterbreite von 100 ergeben sich stark überdeckte Bereiche, welche entsprechend Autokorrelationen aufweisen. 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 $s \leq w$.

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

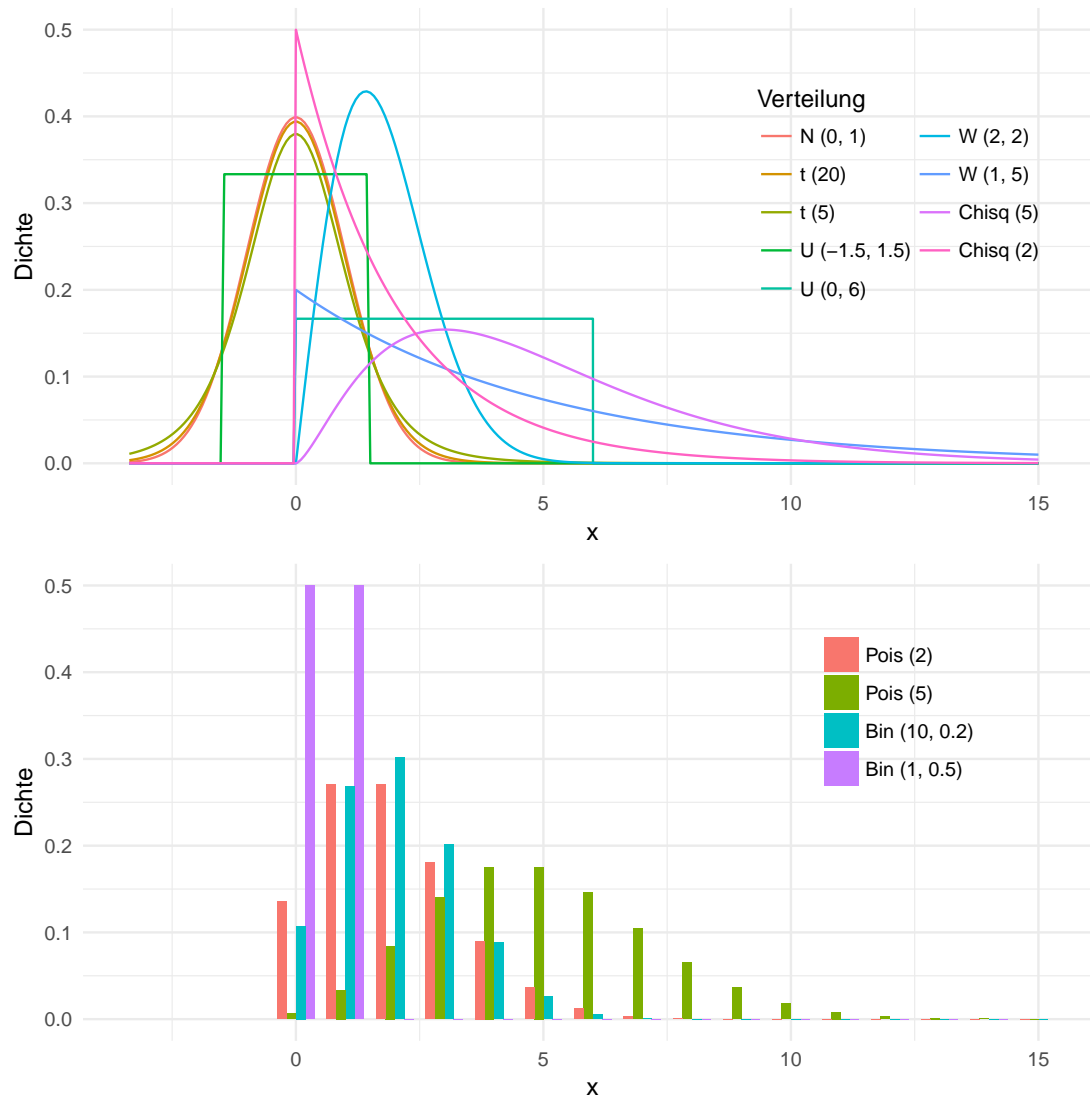


Abbildung 5.1: Die dreizehn im Folgenden betrachteten Verteilungen: Acht stetige und vier diskrete Verteilungen sollen durch das gleitende Mittel *normalisiert* werden. Außerdem wird das Verhalten für Werte betrachtet, welche schon zu Beginn aus einer (Standard-)Normalverteilung stammen.

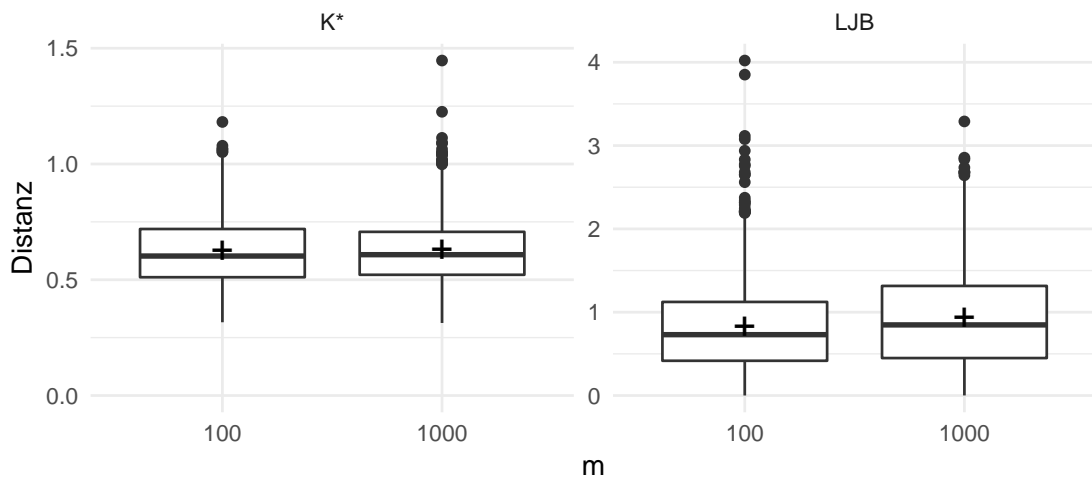


Abbildung 5.2: Wie nah eine Stichprobe an der Normalverteilungsannahme ist, wird durch die oben erläuterten Methoden K^* und LJB gemessen. Stammen die Daten aus einer Normalverteilung der Größe m , so ergeben sich die hier dargestellten Werte (enthalten sind jeweils Ergebnisse zu 1 000 Zufallsstichproben).

Die Ergebnisse werden *nach* den Transformationen verglichen. Sollten durch die Stichprobengrößen Unterschiede bestehen, ist es somit plausibel (und sonst irrelevant), statt der Ausgangsgröße n die Anzahl der resultierenden y -Werte zu betrachten. Diese Anzahl m 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 3.4 beschriebenen Vorgehens berücksichtigt.

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

- Verteilungen:
 - Normalverteilung: $N(0,1)$
 - Chi-Quadrat-Verteilung: χ_2^2, χ_5^2
 - t -Verteilungen: t_5, t_{20}
 - Gleichverteilung: $U(-1.5,1.5), U(0,6)$

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

- Weibull-Verteilung¹: 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 die 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.

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 $w = 1$

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

¹Es wird hier stets von der „englischen Schreibweise“ der Weibull-Verteilung ausgegangen, das heißt von der Dichtefunktion $f(x; \lambda, k) = k/\lambda (x/\lambda)^{k-1} e^{-(x/\lambda)^k}$. 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.

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. Die indirekt festgesetzten Parameter s und n berechnen sich – nach der Bereinigung und ohne Berücksichtigung der Wiederholungen – zu 26 beziehungsweise 218 verschiedenen 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 auch durch die unabhängig von d gleiche Länge der \tilde{x}_i sichergestellt ist (vgl. S. 65).

5.1.2 Ergebnisse

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 5.1 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.

	$d=0$	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 5.1: 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.

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

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, 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 aufgeführte) Gegenüberstellung 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 5.3 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.

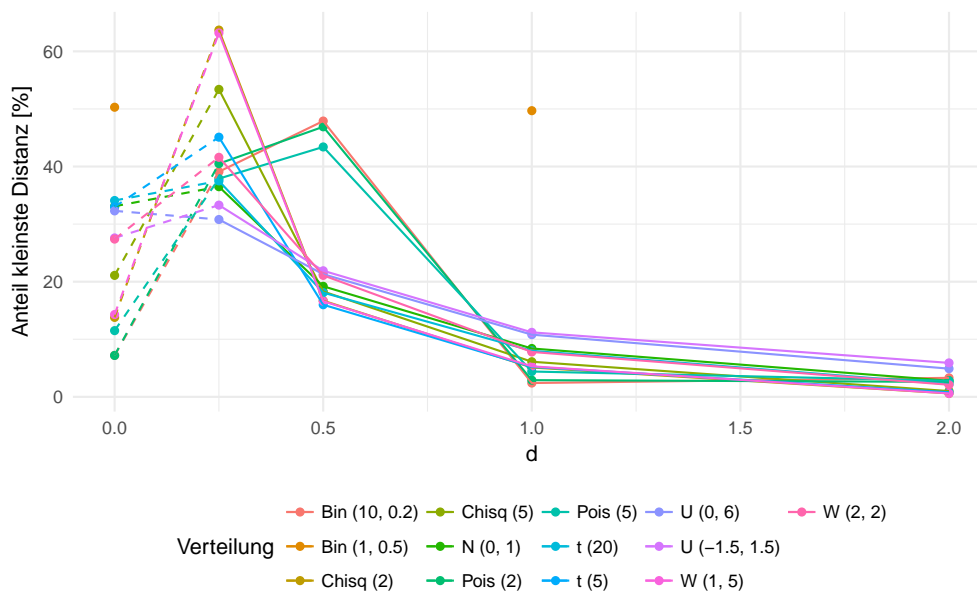


Abbildung 5.3: Je nach Verteilung führt ein bestimmtes d zum kleinsten Distanzmaß. Die Anteile für jedes d sind hier dargestellt. Zu beachten ist, dass der Wert $d = 0$ keine stetige Fortführung des Kontinuums ist. Diese Einstellung symbolisiert die Verwendung der Werte selbst, statt der Differenzen.

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 Schiefe, jedoch kommt hier die diskrete Struktur zum Tragen: Differenzen ungleicher Zahlen betragen dann absolut mindestens Eins, während sie bei stetigen Verteilungen 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 der Verläufe von einer horizontalen Geraden zu erkennen: Diese wäre zu erwarten, wenn alle betrachteten 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 5.1 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 5.4 zeigt die „Zielgröße“ LJB in Abhängigkeit von k und der Ausgangsverteilung.

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

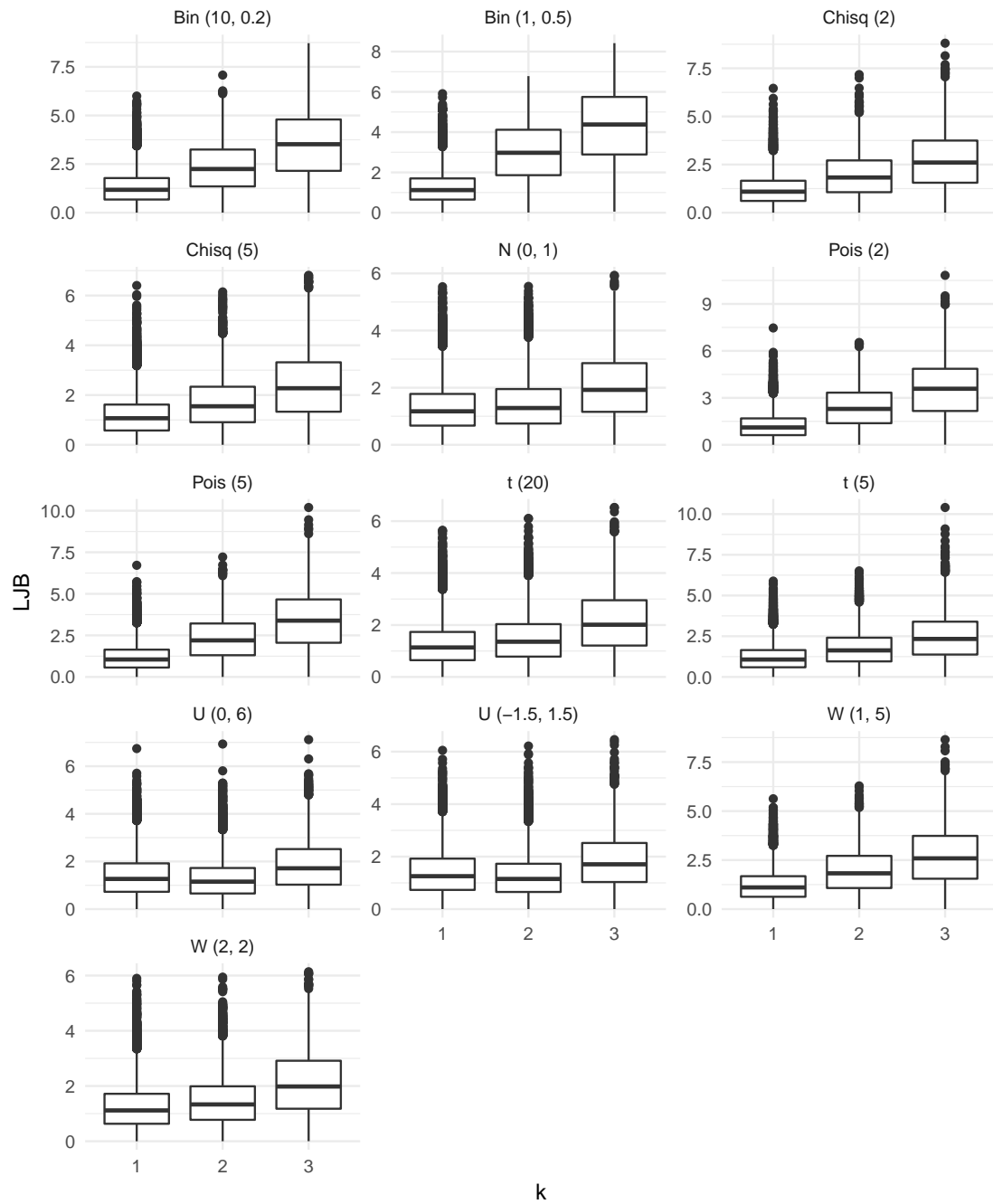


Abbildung 5.4: Für verschiedene w , s und k , sowie für wie oben angegeben ausgewählte d , ergeben sich die dargestellten LJB -Maße. Die Ergebnisse für K^* sind qualitativ gleich, numerisch fallen die Unterschiede etwas geringer aus.

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 65 beschriebenen Vorgehens entfällt damit.

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 den durch die gleitenden Mittel implizite Informationsverlust 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 5.5 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 8.6, 8.7 und 8.8 auf den Seiten 176 bis 178.

Zunächst fallen die Boxplots für kleine s_{rel} auf, insbesondere bei großem w . Das weist darauf hin, dass die oben angestellten Überlegungen zur Unabhängigkeit begründet waren: Werden die Überdeckungsraten aufeinander folgender Fenster beim gleitenden Mittelwert zu groß, kann das Ergebnis nicht mehr als aus einer unabhängigen Verteilung stammend angesehen werden. Dass sich dieser Effekt für $s_{\text{rel}} = 1/w$, das heißt für $s = 1$,

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

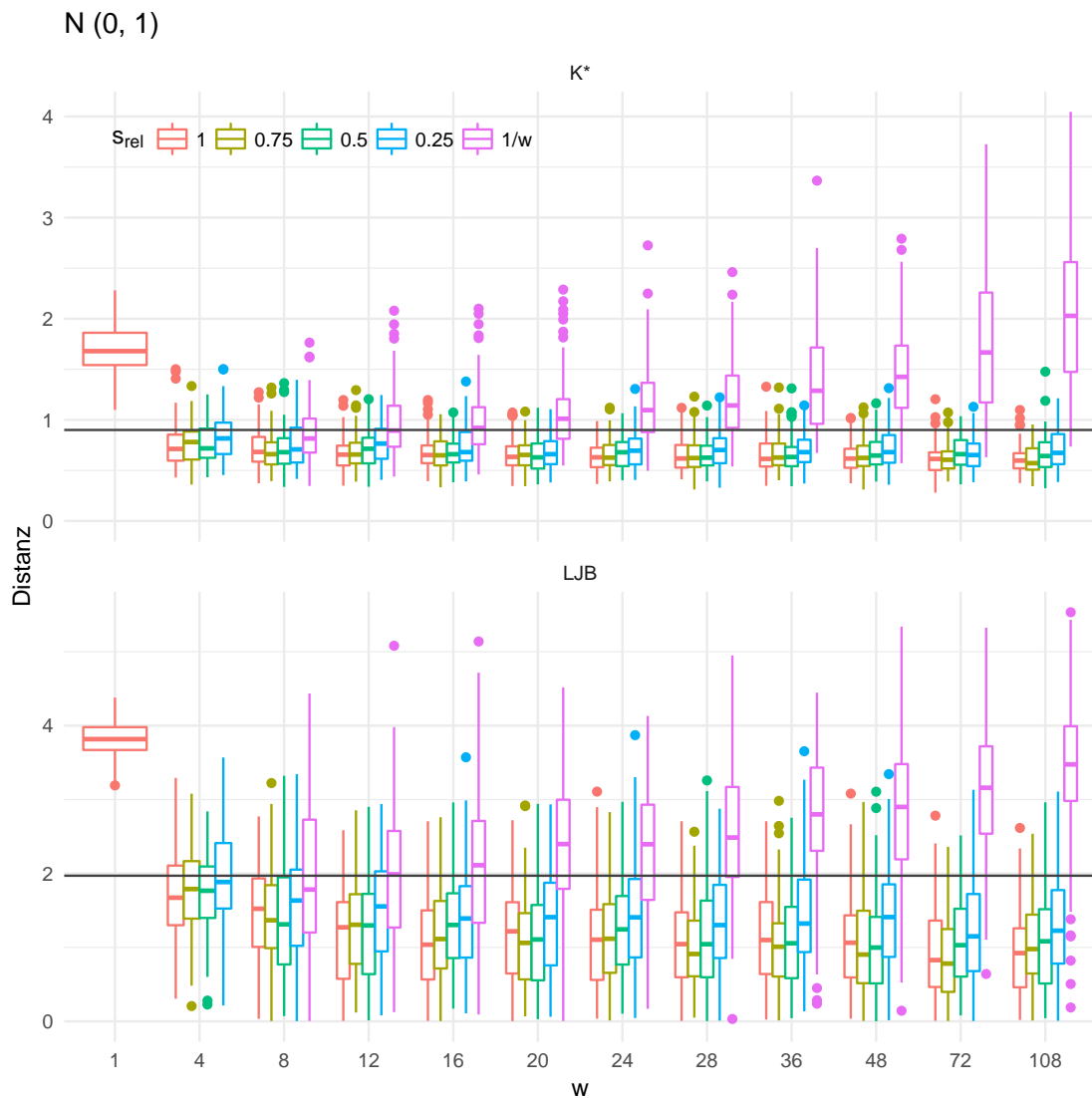


Abbildung 5.5: Die beiden Distanzmaße für a priori normalverteilte Daten, in Abhängigkeit von der Fensterbreite w und der Schrittweite s . Die Ergebnisse wurden auf solche für a posteriori-Stichproben der Größe $m = 1\,000$ reduziert. Die horizontalen Geraden bilden die 95 %-Quantile der Maße für normalverteilte Daten ab und sollen lediglich die Einordnung der numerischen Resultate vereinfachen. Siehe dazu auch den Ankerpunkt in Form der Abbildung 5.2.

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

Die Abbildung 5.6, sowie die Abbildung 8.9 im Anhang auf Seite 179, 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 5.5, 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 $s = w/4$ 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 Sollwerte. 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 $s_{\text{rel}} = 0.5$ als gut verwendbar erscheint. Da kleine Schrittweiten angestrebt werden, wird diese Wahl getroffen.

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

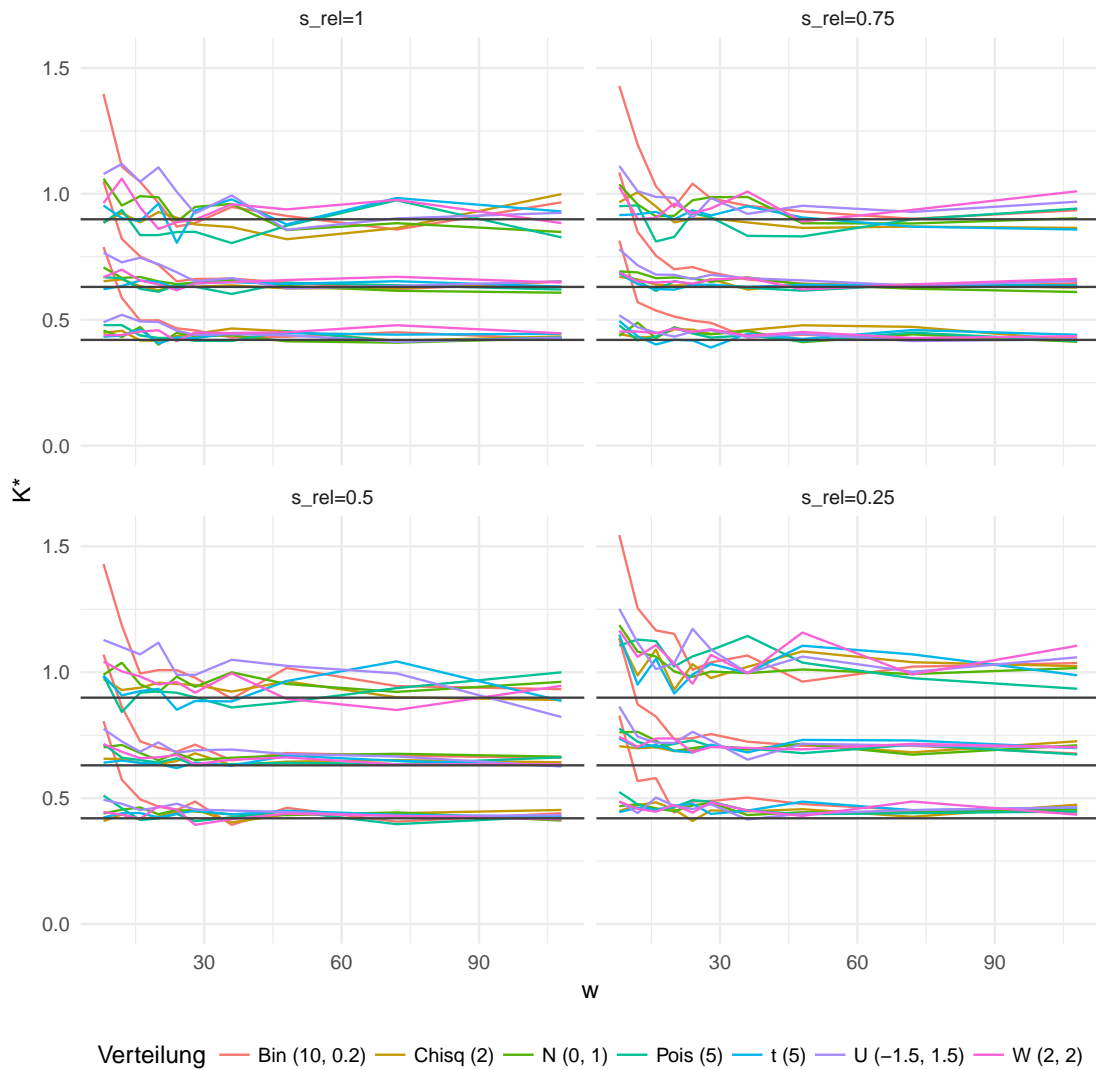


Abbildung 5.6: Je nach Einstellung der Schrittweite s_{rel} ergeben sich andere Zusammenhänge von der Fensterbreite w und dem Abstand zur Normalverteilung. Dargestellt sind hier die Quantile zum 5- und 95-Prozent-Niveau sowie das arithmetische Mittel aus jeweils 100 Wiederholungen. Es werden Ergebnisse für $m = 1\,000$ gezeigt. Die horizontalen Linien geben diese Kennzahlen als für eine Normalverteilung simulierte Werte an.

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.

5.1.3 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 $d = 0.25$ 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 Erstellung der Abbildung 3.13 verwendet, welche auf der Seite 67 einen beispielhaften Durchlauf des Verfahrens zeigt.

5.2 Allgemeinere Szenarien

Für das im Vorangegangenen untersuchte Verfahren wurden durch die verschiedenen simulierten Zeitreihen indirekt auch Beispiele gegeben. Dabei erfüllen diese durchweg die Forderung nach unabhängigen und identisch verteilten Beobachtungen. Es werden nun weitere Situationen betrachtet, welche durch ihr häufiges Vorkommen in der Anwendung Relevanz erlangen. Zunächst wird auch die Konvergenzgeschwindigkeit betrachtet, welche durch das Berry-Esseen-Theorem abgeschätzt werden kann.

5.2.1 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 $y \in \mathbb{R}$ sowie 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 wurde 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 (ebenda).

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 5.7 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.

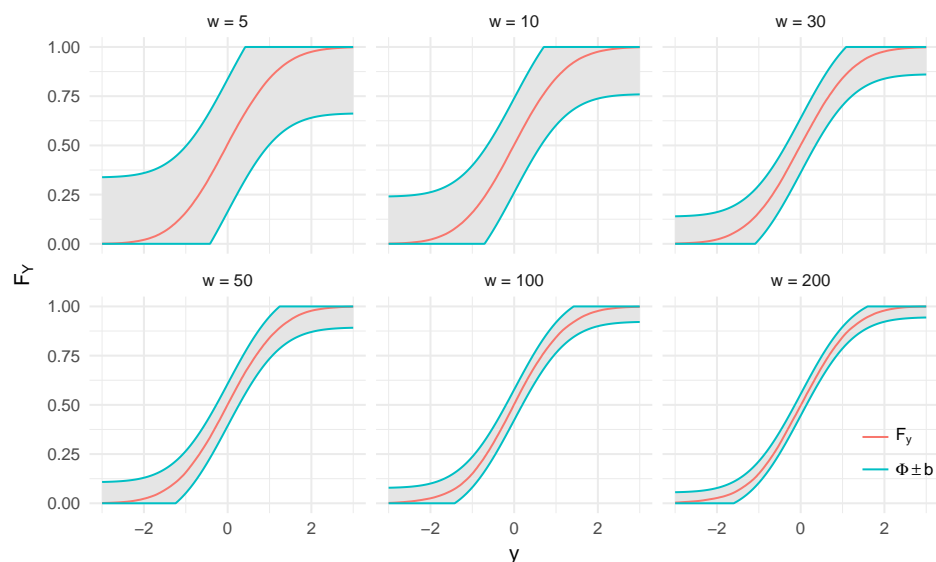


Abbildung 5.7: Für simulierte Daten mit $X_1 \sim \chi_2^2$, $d = 0.25$, $s = 1$ und verschiedene Fensterbreiten sind die empirischen Verteilungsfunktionen, sowie die geschätzten Berry-Esseen-Grenzen dargestellt. Die Grenzen sind hier auf Bereiche zwischen 0 und 1 begrenzt worden. Trotz großer Abhängigkeiten der y_i werden die Grenzen von den empirischen Verteilungsfunktionen eingehalten.

Mit 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 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 $1.54/\sqrt{n}$).

5.2.2 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 der 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 aus der Gleichung

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

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

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

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 sei 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 der S_n 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.

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

Simulativ wird 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 Verteilungen von X_i , Y_i und Z_i gezogen. Um den ZGWS hier für bekannte Ausgangs-Verteilungen zu demonstrieren wird dabei entgegen den obigen Resultaten $d = 0$ gesetzt, also keine Differenzenbildung und Symmetrisierung vorgenommen. Anhand der logarithmierten Jarque-Bera-Statistik wird dann die Nähe zur Normalverteilung beurteilt. Die Abbildung 5.8 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 .

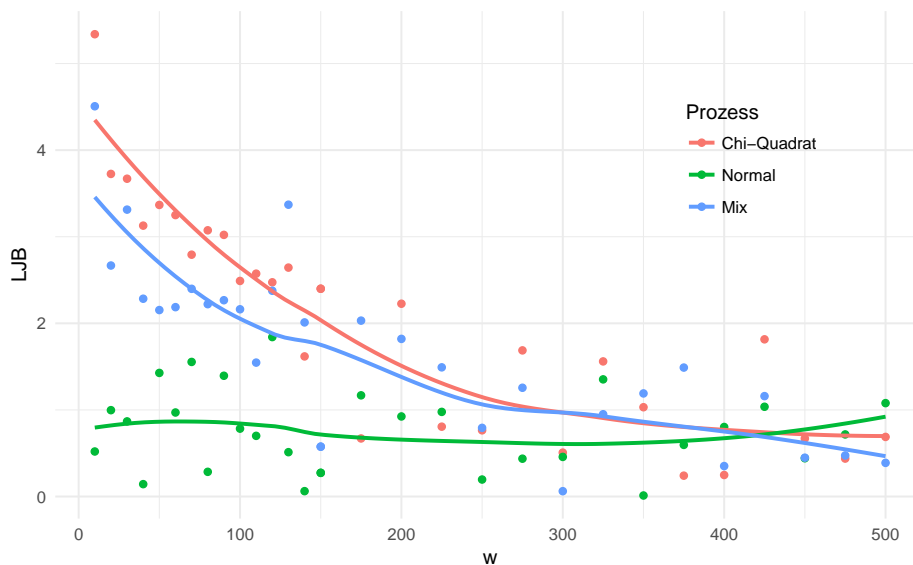


Abbildung 5.8: Wird aus Prozessen der zwei genannten Verteilungen der Wert S_n gebildet, und anschließend der Abstand zur Normalverteilung durch *LJB* gemessen, zeigt sich die durch den ZGWS beschriebene Konvergenz. Dass diese auch für nicht identisch verteilte Größen bestehen kann, zeigt diese beispielhafte Mix-Verteilung. Die Linien stellen polynomielle Regressionen der Ergebnisse dar.

5.2.3 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: 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 ε - δ -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 ermöglicht werden.

Bei Angus (1993) finden sich Herangehensweisen für die Anwendung des ZGWS auf bestimmte ARIMA-Prozesse. Dabei wird jedoch auch deutlich, dass sich die Abweichung von der Unabhängigkeits-Annahme zwar auf die Gestalt von Erwartungswert und Varianz der Limes-Verteilung, nicht jedoch auf die Konvergenz gegen eine Normalverteilung auswirkt (vgl. bspw. 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 $m = 0$ ist 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

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

5 Zur Verteilungs-Konvergenz von gleitenden Mittelwerten

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 den Seiten 123 f. nicht verändert ist. Dort wurde der Term $\sqrt{n}(\bar{X} - \mu)$ 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.

6 Alarmsysteme mit neuer Methodik

Die Konstruktion des vorgestellten Verfahrens sowie die Optimierung der Parameter sollen möglichst allgemein geeignet sein, um auch möglichst allgemein eingesetzt werden zu können. Der eigentliche Anlass für die Arbeiten war aber die Überwachung der Frequenz-Verläufe, die wie in Teil 2.2 beschrieben vorliegen. Nun wird dieses Problem wieder aufgenommen: Nachdem die im zweiten Kapitel beschriebene Problemstellung mit der im dritten Kapitel vorgestellten Standard-Methodik nicht zufriedenstellend gelöst werden konnte, wurde eine neue Variante vorgeschlagen und im vorigen Kapitel 5 untersucht. An dieser Stelle folgt die Anwendung des vorgeschlagenen Verfahrens auf die geschilderte Problemstellung. Neben der Evaluierung der Rate falscher Alarmer im Teil 6.1 erfolgt auch eine Betrachtung der Qualität im Falle eines möglichst schnell zu erkennenden *echten* Alarms. Dabei ist zunächst zu klären, wie sich ein solcher definiert (Teil 6.2).

6.1 Zur Fehlalarmrate

Wird die Reihe der Mittelwerte als Basis für ein Alarmsystem verwendet, so resultiert aus dem beschriebenen Vorgehen eine Rate falscher Alarmer von 15.4 Prozent. Das ist vermutlich auf die Autokorrelation der Daten zurückzuführen: Diese beträgt für die transformierte Reihe knapp 0.77. Quellen dieser Selbstabhängigkeit sind zum einen jene in der Ausgangsreihe, welche sich – wenn auch stark gedämpft – in den absoluten Differenzen wiederfindet. Zum anderen führen eben auch die Überlappungen der Mittelwert-Fenster zu Abhängigkeiten aufeinander folgender Fenster.

Hier ist das Problem der Autokorrelation nun jedoch weitestgehend isoliert, so dass gängige Lösungsstrategien greifen. Die einfachste der im Kapitel 3 vorgestellten Varianten stellt das Aufspreizen der Grenzen dar. Diese Methode führt hier in der Variante nach Gilbert zu einer Rate falscher Alarmer von 0.43 Prozent. Die Zielgröße von $2\Phi(-3) = 0.27\%$ wird damit auch noch verfehlt, der bisherige Bestwert (EWMA mit OSAPE und Gilbert-Faktor) von 0.96 Prozent ist aber verbessert. Zudem bietet sich hier durchaus die Möglichkeit, dem Zielwert noch näher zu kommen: Die Wahl der Parameter w , s und

d kann auch für den jeweils vorliegenden Datensatz optimiert werden. Hier kann jeweils zwischen Aufwand, Dauer in Echtzeit und Nutzen abgewogen werden.

Eine Veränderung der Parametereinstellungen in Form einer Verdopplung von w und s auf 60 beziehungsweise 30 (und damit Beibehaltung von $s = w/2$) führt ohne und mit Aufspreizen der Grenzen zu Fehlalarmraten von 23.4 und 0.30 Prozent. Wird zusätzlich noch die Wahl von d vom „Standard“ 0.25 auf 0.125 halbiert, so sinkt die Fehlerrate noch einmal, und beträgt dann mit aufgespreizten Grenzen 0.2743 Prozent. Dabei ist zu berücksichtigen, dass diese Verbesserung nicht etwa zufälliger Natur ist, sondern durch die Verteilungskonvergenz begründet ist: Mit den Standard-Parametern betragen die Normalverteilungsmaße für die Karten-Statistik (welche sich durch die Gilbert-Korrektur nicht verändert) $K^* = 115.7$ und $LJB = 7.7$. Mit dem veränderten genannten Parameter-Set sinken diese Werte auf 10.0 und 6.2, gerade im Kolmogorov-Maß wird damit eine deutliche Näherung an die Normalverteilung angezeigt.

Die Abbildung 6.1 zeigt einen Ausschnitt aus dem letzten Tag der Zeitreihe 1 mit Kontrollkarten für die beiden genannten Kombinationen von w , s und d . Es ist zu erkennen, dass – in diesem Fall – für beide Parameter-Varianten der gleiche Punkt als Fehlalarm ausgegeben wird. Dabei handelt es sich um eine kurze Phase größerer lokaler Variabilität, welche viele kleine Bewegungen mit sich bringt. Ein Fahrplanwechsel-Muster bei etwa 1 280 Sekunden findet sich für beide Varianten als größerer Statistik-Wert wieder, wird aber durch die fensterweise Mittelwertbildung „aufgefangen“ und führt nicht zum Alarm. Gut zu erkennen ist auch der veränderte Zeitbereich, so sind die Abstände zwischen zwei Statistik-Werten in der unteren Kontrollkarte länger als in der oberen.

6.2 Zur Power

Die Anwendung und der Nutzen einer Kontrollkarte bestehen nun offenbar nicht ausschließlich aus der Vermeidung unnötiger falscher Alarme, vielmehr soll auch und insbesondere eine zuverlässige und schnelle Erkennung von Fehlerfällen gewährleistet sein. Bei den hier untersuchten Frequenzdaten bedeutet das, den Netzbetreiber möglichst schnell auf ein untypisches Verhalten der Verläufe aufmerksam zu machen: Nur wenn ein bestehendes Problem auch zügig erkannt wird, kann schnell, angemessen und möglichst ohne Folgeschäden auf Fehlerfälle reagiert werden. Jede Zeitspanne, die bei der Alarmierung eingespart wird, schafft mehr Möglichkeiten bei der Reaktion auf Fehlerfälle.

Ohne Weiteres ist jedoch nicht klar, wie genau ein Fehlerfall hier dargestellt werden kann, noch weniger sind die Strukturen einer *typischen* oder *durchschnittlichen* Störung eindeutig. Dies hat mehrere Gründe: Zum einen entstehen im europäischen Stromnetz

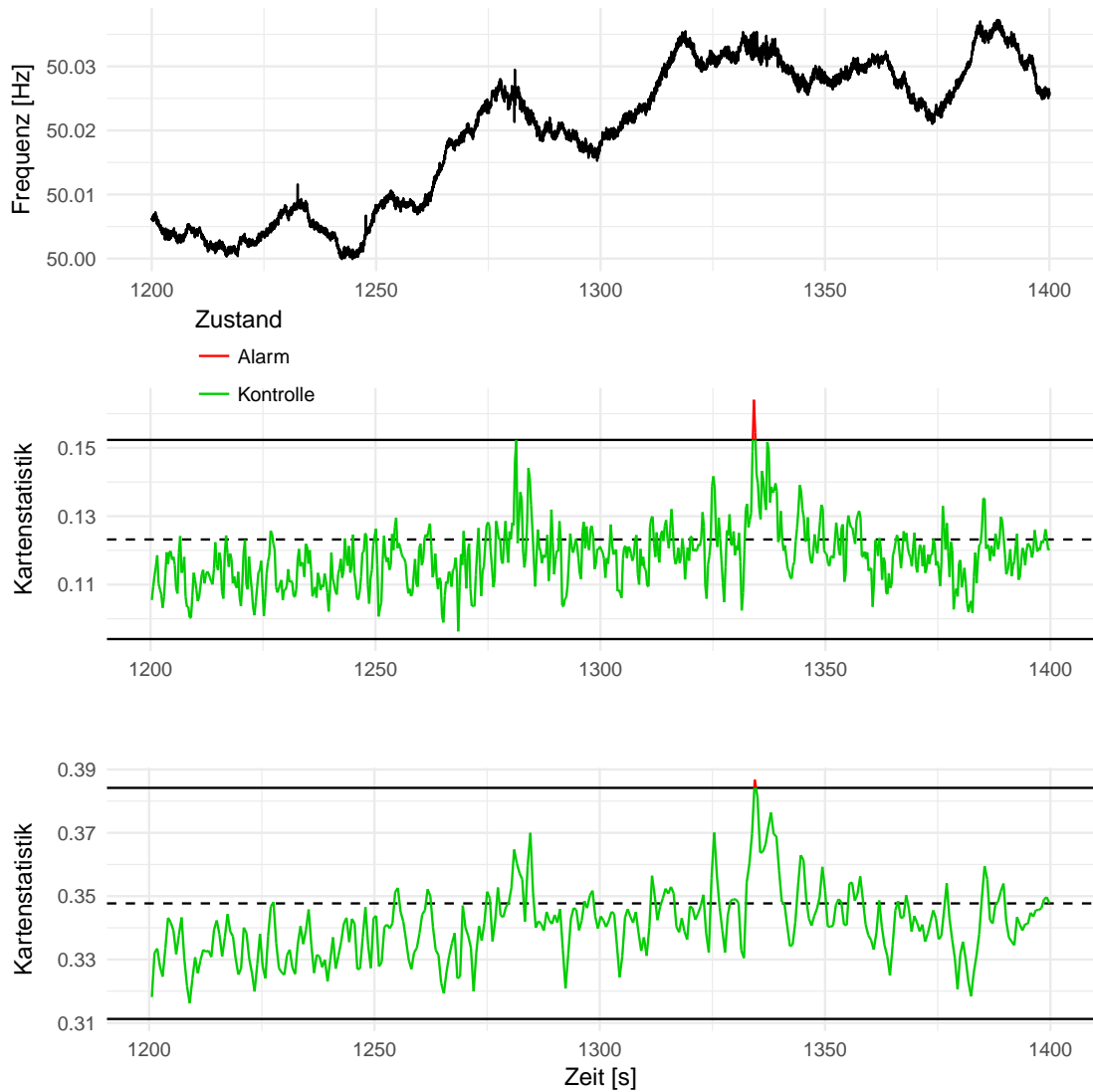


Abbildung 6.1: Für einen Auszug aus der Mittelwertreihe (oben) wurde die Kontrollkarte mit allgemeinen Parametern ($w = 30$, $s = 15$ und $d = 0.25$) wie beschrieben erstellt. Die untere Reihe wurde nach dem gleichen Schema, jedoch für die Variante 60/30/0.125 erstellt.

recht wenige gravierende Störungen, so dass keine Charakteristiken der zugehörigen Frequenzverläufe aus bestehenden Datenaufzeichnungen abgeleitet werden können. Zum anderen begrenzen sich einzelne Großereignisse – wie beispielsweise das „Münsterländer Schneechaos“ von 2005, der Ausfall in mehreren Ländern 2006 oder auch in der Türkei 2015 – weder regional auf die fünf Orte, an denen die hier untersuchten Verläufe erhoben wurden, noch auf den Stand der Netzarchitektur zum Erhebungszeitpunkt. Auch aus inhaltlichen Erwägungen heraus ergibt sich, dass *der typische Ausfall* oder auch *die typischen Ausfälle* nicht einfach beschrieben werden können: Eine Störung führt zumeist zu Kaskaden- und Dominoeffekten, welche wiederum auch von der Art der Störung sowie vom Netzzustand und -design abhängen. Die Modellierung einer Störung erreicht damit Anforderungen, welche sowohl den Rahmen dieser Arbeit, als auch die fachliche Expertise des Verfassers übersteigt.

Eine andere Variante zur Abbildung von Störungen bestünde in der Simulation der technischen Abläufe mit entsprechender Software: Diese ermöglicht es, ein Energienetz mit Erzeugern und Verbrauchern zu erstellen und den Betrieb unter Berücksichtigung einer Vielzahl von physikalischen Effekten zu simulieren. Die dem Verfasser zur Verfügung stehenden Programmpakete berücksichtigen jedoch keine stochastischen Effekte, was keine sinnvolle Anwendung von statistischen Methoden erlaubt.

Stattdessen werden nun zwei Beispiele von realen Großstörungen des europäischen Verbundsystems herangezogen und im Zusammenhang mit dem beschriebenen Verfahren betrachtet. Für eine systematische Analyse folgt dann eine Simulationsstudie auf Basis dreier Varianten von Frequenzverläufen gestörter Systeme. Die Gestalt dieser Störungen wird dazu kurz diskutiert.

6.2.1 Reale Beispiele

Das erste der beiden realen Beispiele stammt aus dem Jahr 2006 und wurde durch die (nicht mehr bestehende) Union for the Coordination of Transmission of Electricity sowie durch die deutsche Bundesnetzagentur ausführlich dokumentiert (vgl. UCTE, 2007 und BNA, 2007). Zur Überführung des Kreuzfahrtschiffes „Norwegian Pearl“ aus der Papenburger Werft über die Ems zur Nordsee wurde am Abend des vierten Novembers eine Höchstspannungsleitung abgeschaltet, was zur Überlastung einer anderen Leitung führte. Vom Emsland ausgehend setzte sich die Störung fort, 15 Millionen Europäer waren bis zu 1.5 Stunden ohne Strom. Die folgende Abbildung 6.2 ist dem Bericht der UCTE entnommen und zeigt den Verlauf der Frequenzen zu Beginn der Störung. Wie deutlich zu erkennen ist, äußert sich die Störung in diesem Fall in einer aufschwingenden Frequenz.

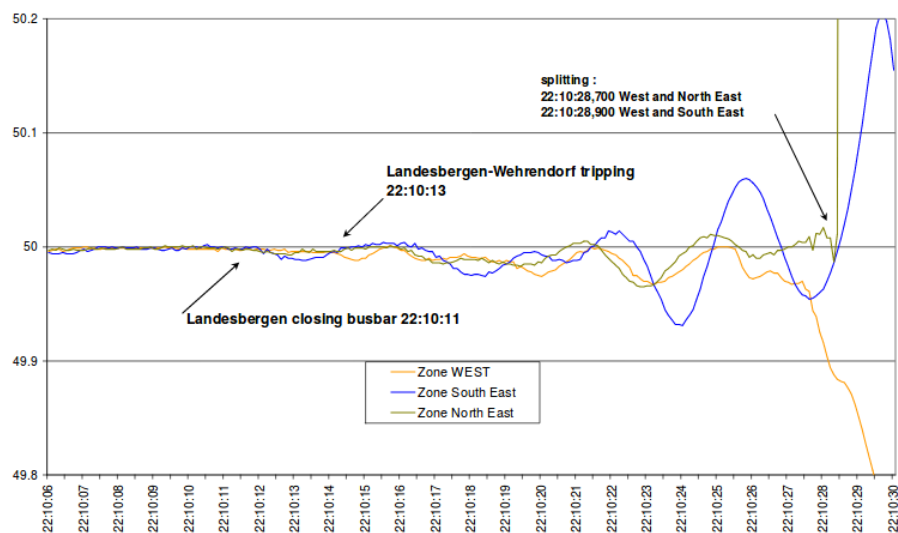


Abbildung 6.2: Dem Ausfall am 4. November 2006 ging die Abschaltung einer Stromsammelschiene (engl. busbar) in Landesbergen voraus. Im Anschluss schaltete sich die Verbindung Landesbergen-Wehrendorf ab und das europäische Verbundnetz zerfiel in drei Teilnetze („Zonen“). Diese Abbildung ist dem Bericht der UCTE entnommen und dort als Figure 5 auf der Seite 22 zu finden.

Ein zweites Beispiel für Frequenzverläufe während einer Störung findet sich ebenfalls in einem Bericht der UCTE (2004), eine weitere Darstellung stellt das schweizerische Bundesamt für Energie zur Verfügung (BFE, 2003, wobei die daraus entstandenen politischen Unstimmigkeiten zwischen Frankreich, Italien und der Schweiz hier nicht von Interesse sind). Demnach kam es in der Nacht zum 28. September 2003 bei einer Leitung durch

den schweizerischen Lukmanierpass zu einem Lichtbogen zwischen einem Leiterseil und einem Baum. In der Folge wurde eine andere Leitung überlastet, welche sich daraufhin abschaltete. „[E]in kaskadenartiger Ausfall weiterer Grenzleitungen nach Italien“ (BFE, 2003, Seite 5) folgte, schließlich ein Stromausfall in ganz Italien. Die UCTE dokumentiert, dass der letzte Kunde erst nach über 18 Stunden wieder mit elektrischer Energie versorgt werden konnte (a. a. O., Seite 22). Die Abbildung 6.3 zeigt den zugehörigen Verlauf der Frequenz.

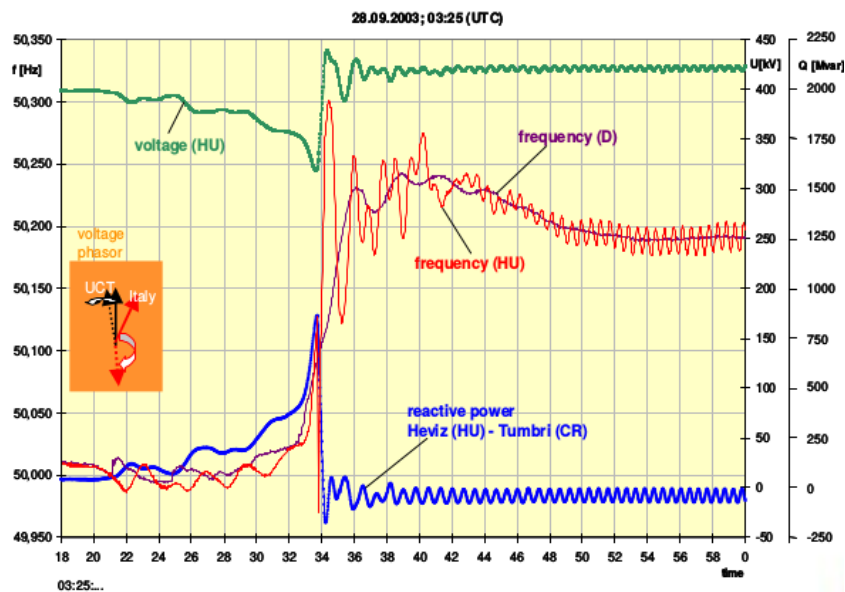


Abbildung 6.3: Der Verlauf der Frequenz (und anderer Messwerte) während der Störung in Italien, gemessen in Uchtelfangen (Saarland/D, lila) und Héviz (Westliches Ungarn/HU, rot). Diese Abbildung ist dem UCTE-Bericht entnommen und dort als Figure 2.2.b. auf der Seite 31 zu finden.

Für beide dargestellten Messorte steigt die Frequenz stark an, zunächst über etwa vier Sekunden (3:25:28 bis 3:25:32 Uhr) langsam, dann über weitere vier Sekunden sehr stark. Anschließend ist das Netz in Deutschland und Ungarn wieder in einem stabilen, wenn auch nicht idealen Zustand. Auffällig sind die unterschiedlichen Strukturen: Während die Werte in Ungarn eine starke (zeitliche-)lokale Variabilität in Form teils sehr starker Schwingungen zeigen, ist dieses Verhalten bei den Messwerten aus Deutschland nur sehr schwach ausgebildet. Die Gründe hierfür sind sicherlich nicht nur in der unterschiedlichen räumlichen Entfernung der Orte zum Ereignis zu suchen, vielmehr wird die gesamte Netzarchitektur Einfluss nehmen. Durch beide Verläufe sollte aber auf einen kritischen Zustand geschlossen werden können.

Um nun den vorgeschlagenen Weg zu Überwachung der Frequenz zu illustrieren, wird ein Abschnitt aus den Frequenzdaten als Lerndatensatz herangezogen. Ein direkt auf diesen Teil folgender Abschnitt wird als Testdatensatz verwendet. Dabei wird dessen Verlauf zur Imitation des 2006er-Ausfalls um einen Summanden der Art $\cos(at \log(t)) \cdot (b \exp(ct))$ mit reellen a , b und c , sowie t als Zeitwert verändert. Beim Ausfall von 2003 wird eine stückweise lineare Funktion verwendet (Messort Uchtelfangen, vgl. Abb. 6.3), beziehungsweise sowohl die lineare Steigung als auch die Schwingung berücksichtigt (Messort Héviz). Die Abbildung 6.4 zeigt das Ergebnis. Wegen der großen strukturellen Ähnlichkeit der Verläufe in den Zonen West und Nord/Ost wird für 2006 nur eines dieser Beispiele betrachtet.

Für die beispielhafte Anwendung der Kontrollkarten wird eine solche nun auf dem gezeigten Abschnitt erstellt. Das Ereignis wird zum Zeitpunkt 0 angenommen. Anschließend werden die fünf Zeitreihen (vier Fehlerfälle sowie die Messwerte selbst) in einem 15 Sekunden langen Abschnitt mit dieser Karte ausgewertet. Über diesen Zeitraum werden alle vier Störungen aufgedeckt, während sich beim Messwert selbst als Vergleichsgröße auf diesem Abschnitt kein falscher Alarm ergibt. Die Abbildung 6.5 zeigt die resultierenden Statistiken samt Alarmgrenzen.

Der Verlauf der Variante *Ungarn 2003*, dargestellt in rot, weist die augenscheinlich größte Abweichung vom eigentlichen Verlauf der Messwerte auf: Hier ist sowohl eine Änderung der mittleren Lage, als auch der lokalen Varianz zu beobachten. Dementsprechend führt dieser Verlauf auch zum größten Wert der Statistik, sowie auch zum größten Anteil an Alarmen: In 35 von 48 Fällen wird die Reihe als auffällig eingestuft, das sind knapp 73 Prozent. Genau so früh wie an dieser Reihe wird der „gleiche“ Ausfall am anderen Messort erkannt: Wie in Ungarn wird der Fehler auch in Deutschland am 14. Wert der Statistik erkannt, und damit nach 4.5 Sekunden. Solange die Frequenz für *Deutschland 2003* weiter steigt, bleibt die Statistik im Bereich der Alarmierung. Wird nur noch ein konstanter Wert zur gemessenen Frequenz addiert, so verschwindet dieser bei der Bildung von Differenzen. Am Verlauf der Statistik sind Original und Verschiebung dann nicht mehr voneinander zu trennen. Eine Interpretation der Alarmrate von 27.1 Prozent kann nur vor diesem Hintergrund stattfinden.

Die Störung aus dem Jahr 2006 wird für beide Messorte später erkannt als die 2003er-Szenarien. Dabei wird die stärkere Schwingung (Süd/Ost, blau) erwartungsgemäß vor der schwächeren (West, orange) erkannt. Zudem wird der dauerhafte Alarm hier erst später ausgegeben: Nach ersten Alarmen zu den Zeitpunkten 5.4 beziehungsweise 7.5 Sekunden wird nach etwa 10 Sekunden die obere Kontrollgrenze dauerhaft überschritten.

6 Alarmsysteme mit neuer Methodik

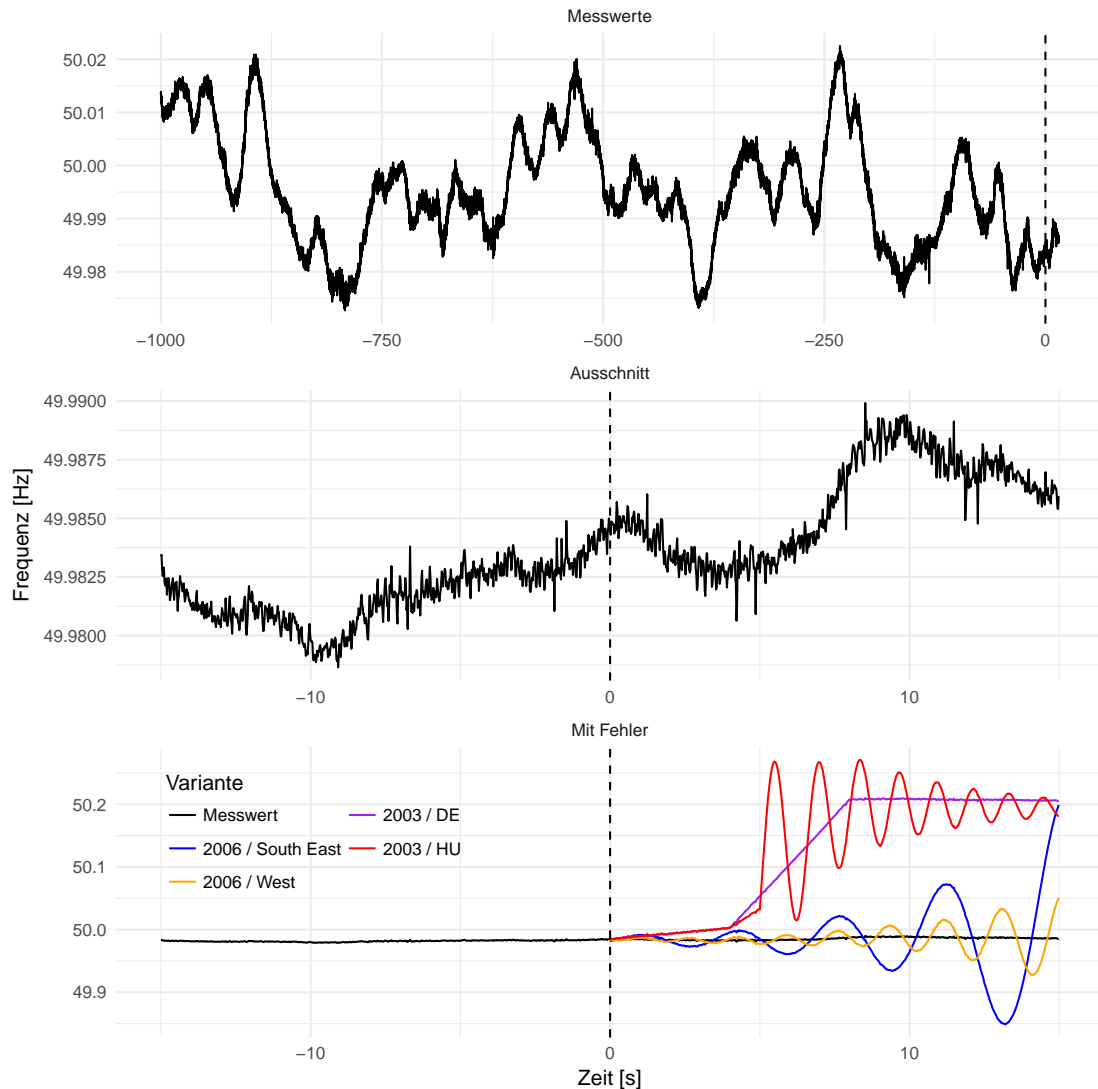


Abbildung 6.4: Den beschriebenen realen Störungen nachempfunden, wird ein Auszug der Daten um einen Fehlerterm additiv erweitert. Vergrößerung und Störungen führen zu recht unterschiedlichen Achsenskalierungen der einzelnen Grafiken, welche hier zu beachten sind. Die Kennungen South East, West, DE und HU sind aus den vorangegangenen Abbildungen 6.2 und 6.3 übernommen.

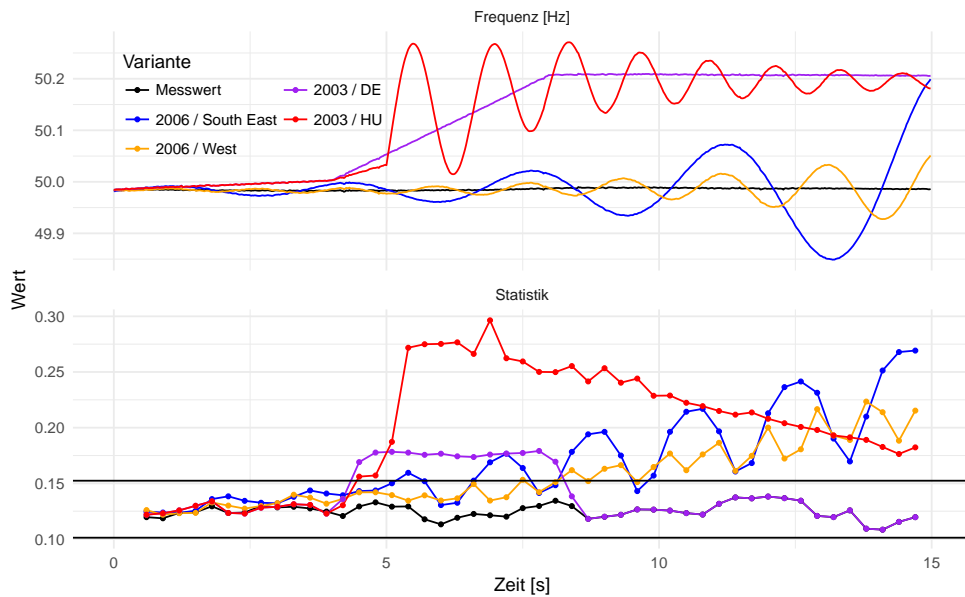


Abbildung 6.5: Nach dem Vorbild der genannten Berichte wurden Frequenz-Verläufe im Fehlerfall nachgebildet. Die Störungen schlagen sich in den gezeigten Verläufen der Kontrollkarten-Statistik nieder, die horizontalen Linien markieren die Kontrollgrenzen.

Insgesamt ist zu bemerken, dass in allen vier Fällen ein Alarm ausgelöst wird, bevor jeweils die Grenze von 50.2 Hertz überschritten wird. Mit Ausnahme von Ungarn 2003 ist der Zeitvorteil deutlich. Gut zu erkennen ist, dass die Statistik – der Verwendung von Differenzen entsprechend – insbesondere auf starke Steigungen reagiert. Je nachdem, wie repräsentativ die betrachteten Fehlerfälle und wie relevant andere Formen in der Praxis sind, erscheint zudem auch die Verwendung einer einseitigen Karte gut möglich.

Neben diesen Betrachtungen von einzelnen Fällen sollen auch systematische Aussagen zur Qualität des Verfahrens getroffen werden. Dabei ist zunächst zu klären, für welche Fehlerstrukturen die Untersuchungen vorgenommen werden.

6.2.2 Struktur der Störungen

Es ist nun Ziel, die Sensitivität der nach dem beschriebenen Verfahren erstellten Kontrollkarten zu analysieren. Neben dem Bezug zur Netzfrequenz soll dies möglichst allgemein durchgeführt werden. Dazu sind idealerweise Testfälle heranzuziehen, welche nach der Sichtung von Beispielen der Literatur zur statistischen Prozesskontrolle als gängig erscheinen. Zumeist handelt es sich dabei um sogenannte *shifts*, das heißt einfache Ver-

schiebungen der Zeitreihe auf einen Erwartungswert, welcher ungleich dem des Prozesses unter Kontrolle ist. Festgehalten wird dann die zu erwartende Lauflänge (average run length, ARL) in Abhängigkeit dieses Erwartungswertes; alternativ die Wahrscheinlichkeit, für eine Beobachtung aus der jeweils neuen Verteilung keinen Alarm zu erhalten. Wird diese Wahrscheinlichkeit als Funktion des Erwartungswertes μ der Karten-Statistik aufgefasst, also als Funktion $L : \mathbb{R} \rightarrow \mathbb{R}$ betrachtet mit

$$L(\theta) := P(\text{Schätzer liegt im Annahmebereich} | \mu = \theta),$$

so wird diese als Operations-Charakteristik- oder kurz als OC-Funktion bezeichnet (vgl. Seite 306 bei Weihs und Jessenberger, 1999). Die Werte von ARL und L können für Karten ohne Gedächtnis ineinander überführt werden durch den Zusammenhang (vgl. ebd., S. 309)

$$\text{ARL}(\theta) = \frac{1}{1 - L(\theta)}$$

Die Größe der Verschiebung besteht bei Wheeler (1983) aus additiven Änderungen des Erwartungswertes zwischen 0.42 und 3 Standardabweichungen, darauf aufbauend ebenso bei Khoo (2013). Eine andere Variante finden sich unter anderem bei Olds (1961): Hier werden Prozesse betrachtet, welche im Fehlerfall aus der Summe von Zufallsvariablen aus einer $N(0, \sigma^2)$ - und aus einer $N(0, \theta^2 \sigma^2)$ -Verteilung stammen (vgl. S. 7 ff), wobei der Prozess für $\theta = 0$ (was als konstanter Summand aufgefasst wird) als unter Kontrolle gilt. Der Prozess außer Kontrolle stammt also aus einer $N(0, (1 + \theta^2)\sigma^2)$ -Verteilung. Dabei werden Parameter $\theta \in [0.25, 3]$ berücksichtigt .

Andere Veröffentlichungen berücksichtigen auch oder insbesondere die Ergebnisse von Karten mit sogenannten run rules (s. u., Teil 6.2.4), so dass sie eher kleinere Änderungen aufdecken können als Varianten ohne solche zusätzlichen Regeln. In diesem Fall gehen Champ und Woodall (1987) von einem Shift formuliert als $d = \sqrt{k}|\mu_1 - \mu_2|\sigma^{-1}$ aus, wobei sie $d = 0.2, 0.4, \dots, 3.0$ betrachten. Dabei ist k die Größe der Gruppen, also der Beobachtungen an einem Zeitpunkt. Bei Page (1955) wird für die Erzeugung des Bereichs außer statistischer Kontrolle der Erwartungswert um 0.2 bis 1 Standardabweichungen verschoben, alternativ die Standardabweichung selbst mit einem Faktor zwischen 1.25 und 2.5 versehen. Bei Lu und Reynolds (1999) beträgt der Shift zwischen 0.5 und 5 Standardabweichungen, wobei gleichzeitig weitere Prozessparameter betrachtet werden. In der Darstellung σ/\sqrt{n} gibt Roberts (1957) die Verschiebung der Lage an, versehen mit Werten zwischen 0 und 3.5. Weitere Einflüsse, wie die Änderung der Stichprobengröße

oder der Autokorrelation (so untersucht etwa von Alwan, 1992, von Wardell et al., 1994 und von Köksal et al., 2008) werden hier nicht weiter betrachtet.

Ein entscheidender Unterschied gegenüber vielen der genannten Beispiele besteht für das vorgeschlagene Verfahren durch die Konstruktion der Kartenstatistik. Wird nämlich nur ein einmaliger Shift in den Ausgangsdaten betrachtet, so wird dieser entweder sofort oder gar nicht detektiert. Welche Zeitspanne den Begriff „sofort“ definiert hängt von der gewählten Fensterbreite und der gewählten Schrittweite ab, der Unterschied macht sich aber in jedem Fall nur zu einzelnen Zeitpunkten bemerkbar. Nach der einmaligen Modifikation des Erwartungswertes und ohne sonstige Änderungen folgen die Differenzen der gleichen Verteilung wie unter Kontrolle, und die Alarmrate gleicht wieder dem Anteil falscher Alarme.

Statt eines einmaligen Sprunges in Ausgangsdaten soll dieser hier in der Statistik erzeugt werden. Dazu können ansteigende Verschiebungen in der beobachteten Zeitreihe verwendet werden. Für das Monitoring des Energienetzes erscheint dies ohnehin realistischer als die punktuelle und endgültige, einmalige Änderung eines stetigen Prozesses.

Ein Fehlerfall kann nun einerseits durch eine Modifikation der Zeitreihe der beobachteten Messwerte x_t selbst mittels einer zeitabhängigen Funktion $f(t)$ simuliert werden, im einfachsten Fall durch eine lineare Verschiebung der Art $x_t + f^l(t) = x_t + a + bt$ mit reellwertigen a und b . Andererseits könnte eine Modifikation auch an den (potenzierten absoluten) Differenzen, oder an der resultierenden Karten-Statistik vorgenommen werden. Die erste Variante ist dabei nicht nur anschaulicher, sie ist auch die einzig relevante Größe: Die Karten-Statistik y_t ist wie die Distanzen \tilde{x}_t ausschließlich ein Hilfsmittel zur Beurteilung der x_t .

Wird, wie bereits als Beispiel genannt, eine lineare Funktion zu den Messwerten addiert, so genügt für die Betrachtung des beschriebenen Verfahrens die Reduktion auf $a = 0$ oder gleichwertig auf $f^l(t) = bt$. Das ergibt sich durch die Betrachtung von Differenzen: Ist

$$x_t^l = x_t + a + bt,$$

so ist

$$\tilde{x}_t^l = |x_{t+1}^l - x_t^l|^d = |(x_{t+1} + a + b(t+1)) - (x_t + a + bt)|^d = |x_{t+1} - x_t + b|^d$$

6 Alarmsysteme mit neuer Methodik

Alternativ könnte auch ein Summand $c \in \mathbb{R}$ mit alternierendem Vorzeichen betrachtet werden um die Statistik zu verschieben, was $f^a(t) = c(-1)^{t \bmod 2}$ entspricht. Für solche $x_t^a = x_t + f^a(t)$ ist dann (beispielsweise für gerade t):

$$\tilde{x}_t^a = |(x_{t+1} - c) - (x_t + c)|^d = |x_{t+1} - x_t - 2c|^d$$

Da so eine Verschiebung auch für ein passendes b aus dem vorigen Beispiel erreicht wird, erübrigt sich die weitere Betrachtung. Soll statt der Lage die Streuung des Prozesses verändert werden, so bietet sich ein einfaches Produkt mit $c \in \mathbb{R}$ der Art

$$x_t^p = cx_t$$

an, für die Differenzen bedeutet das ebenfalls eine Änderung durch einen Faktor:

$$\tilde{x}_t^p = |x_{t+1}^p - x_t^p|^d = |cx_{t+1} - cx_t|^d = |c|^d |x_{t+1} - x_t|^d = |c|^d \tilde{x}_t$$

Die Varianz der Distanzen wird damit zu $V(\tilde{x}_t^p) = |c|^{2d} V(\tilde{x}_t)$ verändert, zudem genügt wegen des Absolutbetrags und der Irrelevanz von $c = 0$ die Berücksichtigung von $c > 0$.

Neben diesen einfachen Transformationen sind, wie oben ersichtlich wurde, auch die Resultate von Schwingungen interessant. Dazu wird hier $f^s(t) = a \cos(bt)$ betrachtet. Die Amplitude $a > 0$ und die Periodenlänge $b > 0$ sind also veränderlich, innerhalb eines Prozesses aber konstant. Ein weiterer Schritt könnte ansteigende Amplituden und kürzer werdende Periodenlänge betrachten. Hier ist aber

$$x_t^s = x_t + a \cos(bt)$$

und damit

$$\tilde{x}_t^s = |x_{t+1} - x_t + a(\cos(b(t+1)) - \cos(bt))|^d$$

Zur Veranschaulichung ist der Summand $a(\cos(b(t+1)) - \cos(bt))$ in der Abbildung 6.6 dargestellt. Offenbar bestehen viele weitere Möglichkeiten, Fehlerfälle zu simulieren. So ist beispielsweise die Autokorrelation im Rahmen der vorliegenden Arbeit immer wieder von Interesse gewesen, so dass es für weitergehende Schritte nahe liegt, eine Veränderung derselben zwischen Trainings- und Teststichprobe zu untersuchen.

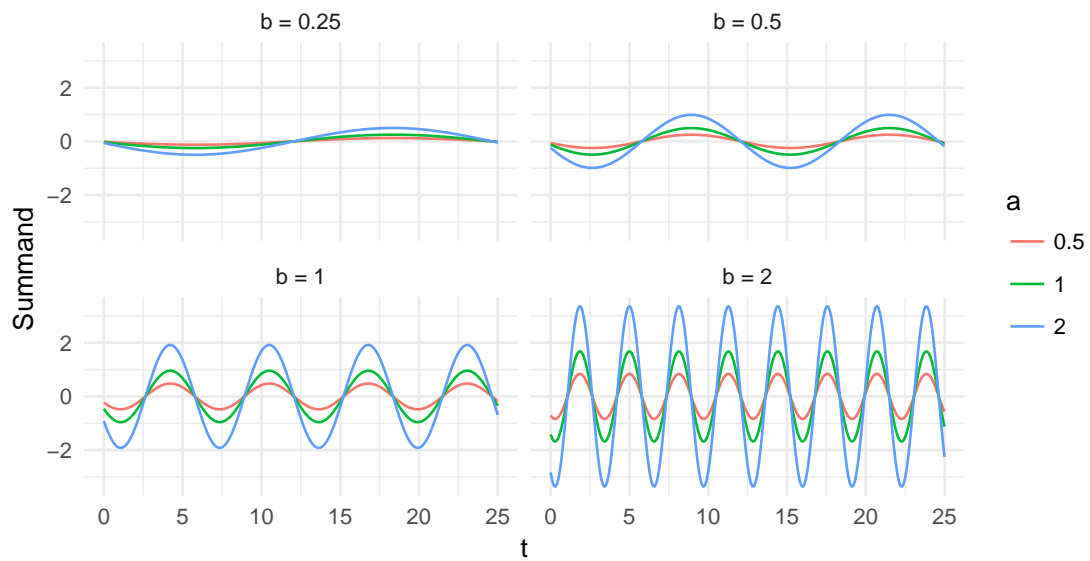


Abbildung 6.6: Der Summand $a(\cos(b(t+1)) - \cos(bt))$ für verschiedene Parameterkombinationen. Es gilt die Gleichheit von $a(\cos(b(t+1)) - \cos(bt))$ und $-2a \sin(bt - b/2) \sin(b/2)$.

6.2.3 Simulation

Für eine aussagekräftige Simulation dieser Alarme sind ausreichend viele Wiederholungen unerlässlich. Diese können auf Basis der Frequenzdaten nicht durchgeführt werden, da hier nicht ausreichend viele *verschiedene* Abschnitte vorliegen, oder diese sehr kurz gewählt werden müssten. Stattdessen werden hier Zufallszahlen aus einfachen AR(1)-Prozessen gezogen. Einen Nachteil zur Beurteilung des Konzepts insgesamt stellt dies nicht dar, im Gegenteil ist dessen allgemeine Verwendbarkeit über eine spezifische Situation hinaus wünschenswert. Bezüglich echter Messwerte bleibt der obige Teil 6.2.1 damit hier der einzige Blickwinkel.

Ein autoregressiver Prozess erster Ordnung, kurz AR(1)-Prozess, wurde bereits oben in Teil 2.3 als ein stochastischer Prozess X_t mit der Eigenschaft

$$X_t = \rho X_{t-1} + \varepsilon_t$$

definiert, wobei $\rho \in \mathbb{R}$ ist und ε_t im Allgemeinen als weißes Rauschen mit einem Erwartungswert von 0 und einer von t unabhängigen Varianz angenommen wird. Beispielsweise Hartung et al. (2005) setzen auf der Seite 678 f. zudem $\rho \neq 0$ voraus. Diese Einschränkung ist hier jedoch nicht zweckmäßig, stattdessen wird $\rho \in [0, 1)$ unterstellt. Offenbar

gilt für die Varianz wegen $V(X_t) = V(X_{t-1})$ und bei Unabhängigkeit der ε_t von X_{t-1} der Zusammenhang

$$V(\varepsilon_t) = (1 - \rho^2)V(X_t)$$

Somit können bei den folgenden Simulationen stets nur zwei der drei genannten Größen eingestellt werden. Während die Autokorrelation ρ dort verändert wird, wird die Varianz der Reihe X_t konstant auf 1 gesetzt. Folglich ist jeweils $V(\varepsilon_t) = 1 - \rho^2$.

Es werden nun Simulationen mit den drei beschriebenen Störungen durchgeführt. Dabei wird eine AR(1)-Stichprobe vom Umfang 10^5 gezogen und damit eine Kontrollkarte nach dem beschriebenen Verfahren erstellt. Eine weitere Stichprobe wird gezogen, nach dem jeweiligen Verfahren transformiert und anschließend die Kontrollkarte darauf angewandt. Variiert werden dabei die jeweiligen Parameter a , b , c sowie die Autokorrelation, welche für AR(1)-Prozesse dem Parameter ρ entspricht. Für jede Kombination wurden 1 000 Wiederholungen durchgeführt, die Parametereinstellungen sind aus den folgenden Grafiken und Tabellen ersichtlich. Die Ergebnisse werden jeweils auf der Basis der neuen Zeitachse betrachtet, also für m statt n Beobachtungen. Auf diese ist der Anteil falscher Alarme eingestellt, eine Umrechnung unter Berücksichtigung der Parameter $w = 30$ und $s = 15$ zudem direkt möglich. An dieser Stelle werden zunächst nur die Anteile der Zeitpunkte mit Alarm-Klassifikation dargestellt, die Lauflängen werden im folgenden Teil betrachtet. Eine direkte Umrechnung ist hier nicht möglich, da bereits durch die überlappenden Fenster die oben zitierte Forderung nach Gedächtnislosigkeit verletzt ist.

Wie beschrieben führt eine Transformation der Art $x_t^l = x_t + a + bt$ zur Betrachtung der Werte $|x_{t+1} - x_t + b|^d$. Für eine sich von $x_{t+1} - x_t$ entfernende Steigung b ist somit eine größere Statistik und folglich eine höhere Alarmierungsrate zu erwarten. Welche Steigungen wiederum als *groß* einzuordnen sind, ist von der Autokorrelation abhängig: Die Größe einer bestimmten Steigung b wirkt sich um so stärker aus, je näher sich x_t und x_{t+1} sind. Der Abstand dieser beiden Werte sinkt wiederum tendenziell für steigende Autokorrelationen, eine bestimmte Steigung ist also bei größerem ρ von größerer Auswirkung als bei kleinen Selbstabhängigkeiten. Das entspricht intuitiven Erwartungen: Je stärker die Autokorrelation ist, desto auffälliger ist ein Sprung in fester Größenordnung.

Sowohl die Abhängigkeit der Alarmrate von b als auch die von ρ ist in der Abbildung 6.7 gut zu erkennen. Dort sind die Alarmierungsraten in Abhängigkeit der Verschiebung dargestellt. Es handelt sich dabei offenbar nicht um eine OC-Funktion wie oben beschrieben, der zugrunde liegende Gedanke ist jedoch der gleiche. Wie zu erkennen ist, sinkt die Stärke der Karte offenbar mit der Autokorrelation. Das ist nicht problematisch, da sie

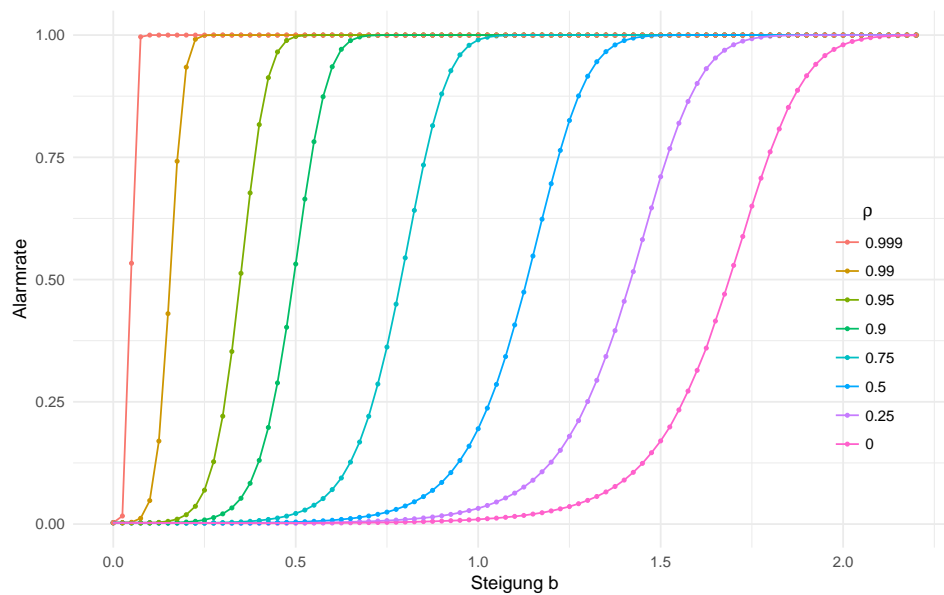


Abbildung 6.7: Für die Addition eines linearen Terms steigt der Anteil richtig erkannter Alarmzustände mit der Steigung b und dem Abhängigkeits-Parameter ρ . Die Punkte zeigen die simulierten Werte, die Verbindungen sind Interpolationen.

gerade mit Blick auf eine große Autokorrelation gebildet wurde und man bei kleineren Selbstabhängigkeiten ohnehin auf Verfahren ohne Ausdünnen der Daten zurückgreifen wird.

Vergleichsweise gering fällt die Abhängigkeit der Alarmrate von der Autokorrelation im Beispiel der Produkt-Transformation aus, wie die Abbildung 6.8 erkennen lässt. Auch hier führt eine steigende Abhängigkeit zu kleineren Abständen aufeinander folgender Messwerte, ein gleicher Faktor transformiert die Reihe etwas öfter in den als *unter Kontrolle* erkannten Bereich. Der deutlich wichtigere Einfluss besteht aber im Faktor c . Bei Halbierung oder Verdopplung der Reihe gegenüber der Trainingsstichprobe gelangt die Alarmrate in Bereiche über 75 Prozent.

Zumindest mit Blick auf den gezeigten Stromausfall von 2006 hat die Schwingungs-Transformation die größte praktische Relevanz. Wie die Abbildung 6.9 zeigt, können auch diese Struktur-Änderungen gerade bei großen Selbstabhängigkeiten der Messwerte gut aufgedeckt werden. Die Systematik ist hier ähnlich zur Variante mit linearen Summanden. Die Alarmierungsrate steigt für eine steigende Amplitude und/oder eine sinkende Periodenlänge. Über beide Varianten wird eine größere Steigung erzeugt, welche dann letztlich detektiert wird. Welche Werte als *groß* zu gelten haben, bestimmt sich wieder-

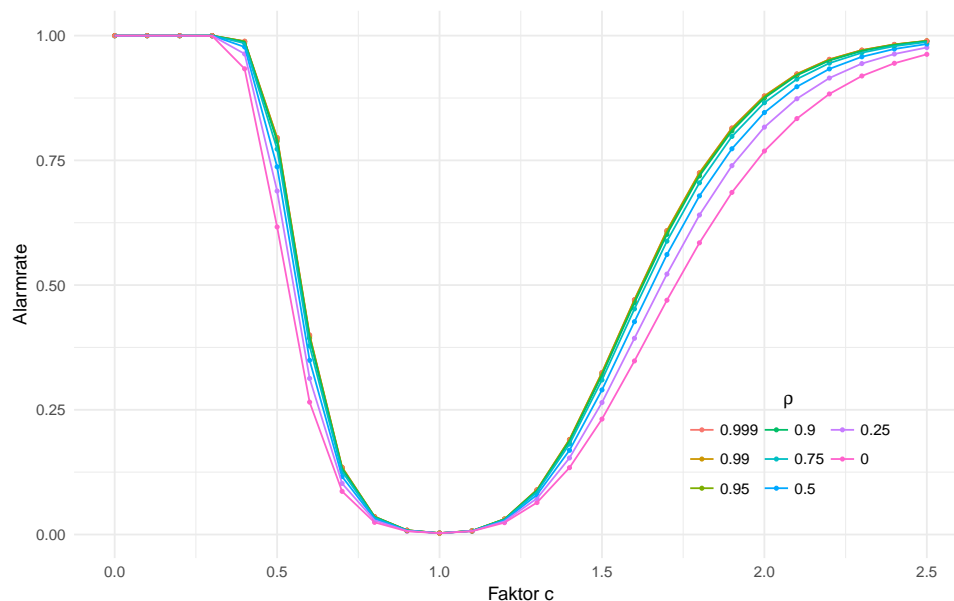


Abbildung 6.8: Die Ergebnisse für Fehlerfälle, welche sich durch die Transformation x_t^p modellieren lassen. Hier weist der Parameter ρ deutlich weniger Einfluss aus als im vorigen Beispiel.

um über die Autokorrelation und damit über den unter statistischer Kontrolle *normalen Abstand* zweier Werte x_t und x_{t+1} . Statt einer Betrachtung von auf- oder abklingenden Schwingungen können die zu erwartenden Alarmierungsrate auch aus Kombinationen der dargestellten Ergebnisse zumindest grob angenähert werden.

Neben den Alarmierungsrate sind offenbar auch die jeweiligen Zeiten bis zum ersten Alarm von Interesse. Diese werden im Folgenden notiert, wobei zusätzlich die Ergebnisse einer versuchsweisen Einbeziehung von run rules dargestellt werden.

6.2.4 Erweiterungen

Bei der allgemeinen Beschreibung der Kontrollkarten-Technik in Teil 2.4 wurde angemerkt, dass viele weitere Varianten und Erweiterungen bestehen. Eine wichtige Erweiterung des Vorgehens besteht in den hier bisher nicht berücksichtigten *run rules*. Diese werden in vielen Varianten verwendet, wie etwa bei Champ und Woodall (1987) oder auch bei Faraz et al. (2014) nachzulesen ist. Auch im vorliegenden Fall könnte eine Verbesserung der Vorgehensweise durch die Hinzunahme solcher Regeln erreicht werden. Dabei ist zum einen die klassische Variante denkbar, bei der die Läufe der Statistik selbst beobachtet werden: Sind $r \in \mathbb{N}$ von $s \in \mathbb{N}$ aufeinander folgenden Werte der Statistik alle ober-

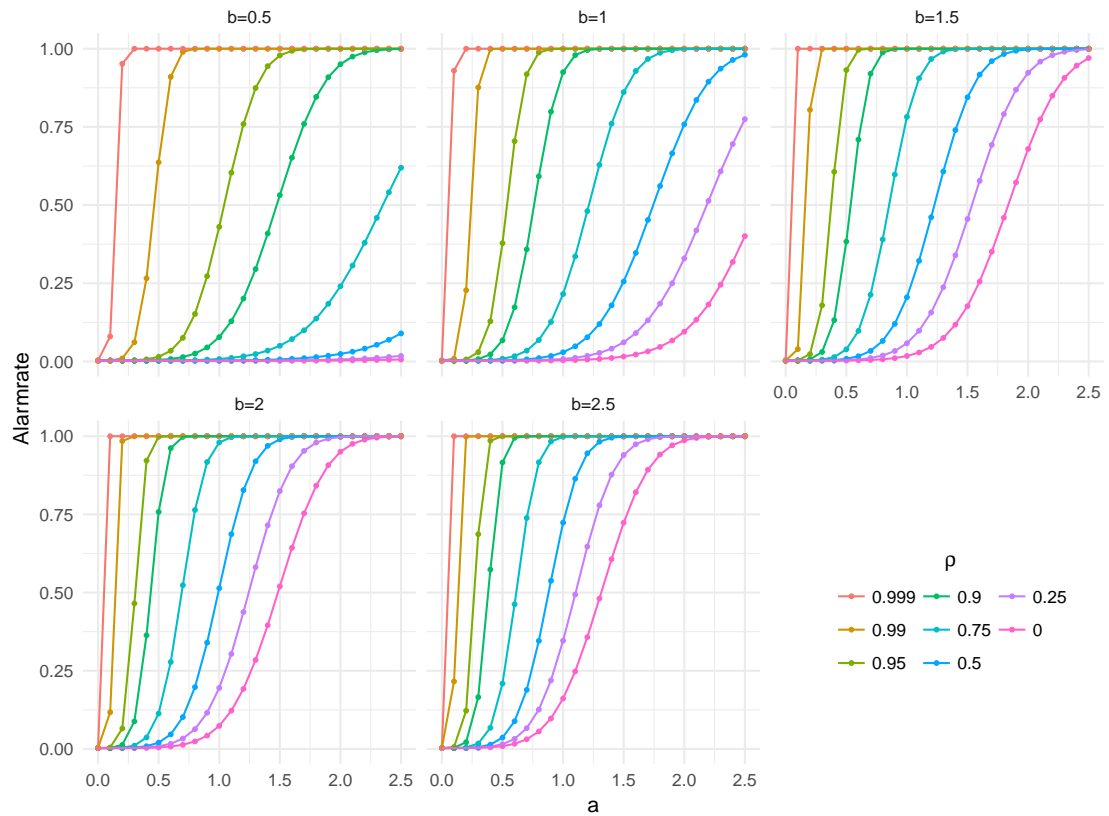


Abbildung 6.9: Die Variante mit Schwingungen, also die Betrachtung eines Fehlerterms $x_t^s = x_t + a \cos(bt)$, führt zu den hier dargestellten Alarmierungsraten. Sowohl für $a = 0$ als auch für $b = 0$ folgt $x_t^s = x_t$.

oder alle unterhalb einer bestimmten Grenze (beispielsweise des geschätzten Erwartungswertes), so wird von einer Lage-Verschiebung des Prozesses ausgegangen. Es wird dann Alarm gegeben, auch wenn die Kontrollgrenzen nicht überschritten werden.

Alternativ können bei der Statistik-Bildung über gleitende Fenster aber nicht nur die Werte dieser Fenster berücksichtigt werden, sondern auch die jeweils darin eingehenden Werte: Haben bei einem Fenster der Breite w insgesamt $u \in \mathbb{N}$ Datenpunkte mit $u \leq w$ eine bestimmte, idealerweise unter Kontrolle unwahrscheinliche Eigenschaft, wird ebenfalls Alarm gegeben. Im Rahmen des vorgeschlagenen Verfahrens würde sich etwa die Betrachtung der Ein-Schritt-Differenzen $x_{t+1} - x_t$ anbieten, wobei der oben verwendete Absolutbetrag an dieser Stelle gezielt ausgelassen wird. Ist der Prozess unter Kontrolle, sind um Null symmetrische Differenzen zu erwarten (und sind Gründe für das Gegenteil bekannt, sollte dieses Wissen ohnehin in die Bildung einer Kontrollkarte einfließen). Es wird also neben der zuerst genannten Regel, welche hier als r -Regel bezeichnet werden soll, eine u -Regel eingeführt: Sind mindestens u Differenzen innerhalb eines Fensters positiv oder negativ, so wird Alarm gegeben.

Bei den im vorigen Teil beschriebenen Simulationen wurden diese Regeln als Beispiel zusätzlich mit berücksichtigt. Dabei wurden $r = 20$, $s = 20$ und $u = 7$ angesetzt. Der Wert für r erscheint im Vergleich mit der Literatur recht hoch, der große Wert hier ist den Autokorrelationen der Teststatistik geschuldet. Dass von $w = 30$ Differenzen für eine Bewertung als „unter Kontrolle“ mindestens 7 und maximal 23 positiv sein sollen, erscheint als möglicher Kompromiss von Spezifität und Sensitivität. Eine Binomialverteilung mit den Parametern 30 und 0.5 lässt hier eine Rate falscher Alarme von 0.14 Prozent erwarten.

Die Ergebnisse werden in Form der durchschnittlichen Lauflängen ARL betrachtet. Die Tabelle 6.1 auf der Seite 163 gibt diese für die Schwingungs-Transformation an. Die Anzahl der Parametereinstellungen ist dabei gegenüber den vorigen Grafiken der Übersichtlichkeit halber etwas gekürzt. Notiert sind die Lauflängen für die ausschließliche Verwendung von Kontrollgrenzen (KG), sowie für die *Hinzunahme* jeweils einer der genannten beziehungsweise beider Regeln. Mit α und β werden die Zustände unter und außerhalb statistischer Kontrolle symbolisiert. Bei den Simulationen wurde die Zeitreihe für den Fehlerfall gegebenenfalls jeweils so lange verlängert, bis mindestens ein Alarm für die Kontrollgrenzen zu notieren war, und damit auch für alle anderen Spalten. Die mittleren Lauflängen basieren somit auf nicht-zensierten Werten.

Beide Regeln weisen eine starke Abhängigkeit von der Autokorrelation auf, so dass optimale Einstellungen nur durch weitergehende Untersuchungen möglich sind. Diese könnten auch den Vergleich mit weiteren Regeln vornehmen. Für einen ersten Eindruck genügen

Status	ρ	a	b	KG	KG & $u = 7$	KG & $r = 20$	KG & $u = 7$ & $r = 20$
α	0	0.00	0.00	411	411	400	400
		0.00	1.00	424	424	415	414
		1.00	0.00	402	402	391	391
	0.5	0.00	0.00	389	389	380	380
		1.00	0.00	388	388	378	378
	0.9	0.00	0.00	391	389	380	379
		0.00	1.00	386	382	375	370
	0.99	0.00	0.00	401	300	385	292
		1.00	0.00	397	295	386	289
	0.999	0.00	0.00	390	262	381	257
		0.00	1.00	400	263	388	255
	β	0	0.10	0.50	387	387	377
0.10			1.00	400	400	393	393
0.10			2.00	423	423	409	409
0.50			0.50	417	417	403	403
0.50			1.00	502	502	470	470
0.50			2.00	227	227	195	195
1.00			0.50	551	551	518	518
1.00			1.00	326	326	189	189
1.00			2.00	17	17	15	15
2.00			0.50	346	345	177	177
2.00		1.00	13	13	11	11	
2.00		2.00	1	1	1	1	
0.5		0.10	0.50	384	384	372	372
		0.10	1.00	388	388	374	374
		0.10	2.00	394	394	382	382
		0.50	0.50	439	439	419	419
		0.50	1.00	319	319	277	277
		0.50	2.00	55	55	43	43
		1.00	0.50	391	389	324	323
		1.00	1.00	40	40	28	28
	1.00	2.00	2	2	2	2	
	2.00	0.50	49	49	28	28	
2.00	1.00	1	1	1	1		
0.99	0.10	0.50	352	280	337	270	
	0.10	1.00	138	128	104	99	
	0.10	2.00	11	11	10	10	
	0.50	0.50	2	2	2	2	
	1.00	0.50	1	1	1	1	
	2.00	0.50	1	1	1	1	

Tabelle 6.1: Durchschnittliche Lauffängen je nach Verwendung von ausschließlich den Kontrollgrenzen (KG) und bei Hinzunahme der beiden genannten run rules, hier für das Schwingungs-Beispiel (x_t^s).

die angewandten Beispiel jedoch, und hier sind die beiden Regeln samt Parameterwahl recht unterschiedlich zu bewerten. So führt die u -Regel für große Autokorrelationen zu einem starken Absinken der durchschnittlichen Zeit bis zum ersten falschen Alarm: Für $\rho = 0.99$ und $a = 0$ und/oder $b = 0$, also den Fall einer Systems unter statistischer Kontrolle, sinkt die durchschnittliche Zeit zwischen zwei falschen Alarmen um etwa ein Viertel. Ein Nachteil, der kaum hinzunehmen ist. Andererseits führt diese Regel auch erst bei großer Autokorrelation zu Effekten im Falle richtig angezeigter Alarme. Zumindest für die gewählten w und u ist das Vorgehen damit nicht vielversprechend.

Besser fallen die Werte der r -Regel aus. Hier wird die Lauflänge unter Kontrolle deutlich weniger verschlechtert, durchschnittlich und je nach Autokorrelation um 9 bis 11 Beobachtungen. Dem gegenüber steht eine mitunter deutliche Reduzierung der Zeit bis zu einem falschen Alarm, wobei diese Reduzierung für alle Werte von ρ zu beobachten ist. Diese Regel ist folglich empfehlenswert, auch wenn die optimalen Einstellungen von der spezifischen Situation und der Autokorrelation abhängen werden.

Die Ergebnisse zu den beiden anderen Störungen, das heißt die Beispiele mit einem Faktor und einem linearen Summanden, sind ebenfalls als Auszug tabelliert im Anhang zu finden, genauer als Tabelle 8.1 auf der Seite 180 und als Tabelle 8.2 auf der Seite 181. Beide Tabellen bieten bezüglich der Kontrollgrenzen keine qualitativ anderen Ergebnisse als die vorgestellten Grafiken zur Alarmrate, und ebenso keine neuen Erkenntnisse gegenüber den Interpretationen der obigen Tabelle 6.1.

7 Zusammenfassung und Ausblick

7.1 Zusammenfassung

Die vorliegende Arbeit wurde mit dem Ziel erstellt, ein automatisierbares Überwachungssystem für Stromnetze zu entwickeln. Die Verwendbarkeit der Netzfrequenz als geeignete Größe aus technischer Sichtweise, sowie die prinzipielle Eignung von Kontrollkarten als statistische Methodik, ergeben sich aus grundsätzlichen Überlegungen. Diese wurden im zweiten Kapitel nach einer Einführung in die Thematik vorgestellt. Ebenso wurden Eigenschaften der Messwerte dargelegt, welche für fünf Orte des europäischen Netzes und zwei Zeiträume von jeweils etwa einer Woche vorliegen.

Die simple Anwendung des Kontrollkarten-Verfahrens in Kombination mit eben diesen Frequenzdaten schlägt dann aber fehl: Extreme Autokorrelationen und Abhängigkeiten zwischen den Messwerten verschiedener Orte führen zum Versagen oftmals erfolgreich eingesetzter Verfahren. Auch dass die statistische Verteilung der Werte keiner bekannten Funktion folgt und die Frequenz ständigen Regelungen unterliegt, führt zu Annahmeverletzungen von klassischen Verfahren: Die Erstellung einer Mittelwertkarte resultiert in einer Fehlalarmrate von mehr als 98 Prozent (Teil 3.2.1). Mit verschiedenen Ansätzen aus der Literatur sollte hier Abhilfe geschaffen werden. Verschiedenste Ansätze wie ein Aufspreizen der Grenzen oder eine Zeitreihenmodellierung führen mitunter zu Verbesserungen, in keinem Fall wird aber die durch die Parameterwahl zu erwartende Rate falscher Alarme ausreichend genau erreicht (vgl. insbes. die Teile 3.2 und 3.3). Als bester Ansatz stellt sich eine Variante mittels exponentiell gewichteter gleitender Mittel, Ein-Schritt-Prognosen und anschließender Aufspreizung der Grenzen heraus, hier werden 0.96 statt der zu erwartenden 0.27 Prozent der Werte als Fehlalarm eingeordnet. Viele andere Verfahren verfehlen den Sollwert um ein Vielfaches dieses Abstands.

Als Alternative wird in Teil 3.4 eine neue Vorgehensweise vorgeschlagen. Dabei werden die Differenzen der Daten verwendet und absolut sowie potenziert betrachtet. Anschließend erfolgt eine Transformation hin zur Normalverteilung, welche durch die Verwendung eines gleitenden arithmetischen Mittelwertes in Konvergenz erreicht wird. Eine ausführ-

liche Darstellung findet sich auf der Seite 65, eine kurze Zusammenfassung folgt im Fazit auf der Seite 167.

Für Anwendung dieses Verfahrens sind – neben der zulässigen Rate falscher Alarme – die Werte von vier Parametern festzulegen: Die Potenz d , die Breite w der Mittelwert-Fenster, die Schrittweite s zwischen zwei Fenstern und eine Potenz k der Mittelwerte. Im Kapitel 5 wurde die möglichst allgemeingültig optimale Einstellung dieser Werte durch eine Simulationsstudie untersucht. Es wird die Empfehlung der Werte $(0.25, 30, 15, 1)$ für das Quadrupel (d, w, s, k) gegeben, der Parameter k ist damit hinfällig. Die Wahl dieser Parameter begründet sich durch die Folgerungen aus einer im Teil 5.1 beschriebenen Simulationsstudie, weitere Szenarien werden im Teil 5.2 untersucht.

Für die Beurteilung von Parametereinstellungen wurde auch der Abstand der transformierten Werte zur Familie der Normalverteilungen herangezogen. Die Frage, wie dieser Abstand gemessen werden kann, stellte sich als eigenständige Problemstellung heraus. Im vierten Kapitel, und damit als Vorarbeit zu der beschriebenen Simulationsstudie, wurde die Eignung von fünf Metriken und vier Teststatistiken überprüft. Dazu wurden die Kriterien einer *guten* Messung diskutiert (Teil 4.2) und die Kandidaten auch hier mit einer Simulationsstudie bezüglich ihrer Eigenschaften verglichen (Teil 4.3). Als Empfehlung wird schließlich die gemeinsame Verwendung der Kolmogorov-Distanz K und der Teststatistik JB von Jarque und Bera gegeben. Im Teil 4.5 konnten weiter die Vorteile von Transformationen der Art $LJB = \log(JB + 1)$ und $K^* = nK$ aufgezeigt werden, wobei n die Stichprobengröße bezeichnet.

Zu Beginn des sechsten Kapitels wurde dann die Anwendung des vorgeschlagenen Verfahrens für die Frequenzdaten dargestellt. Es zeigte sich, dass Eigenschaften der Daten und der Vorgehensweise mit überlappenden Fenstern zu autokorrelierten gleitenden Mittelwerten führen. Hier ist die Autokorrelation aber die einzige Annahmeverletzung, so dass entsprechende Verfahren Anwendung finden können: Mit dem Faktor nach Gilbert wird eine Rate falscher Alarme von 0.43 % erreicht. Soll dieser Werte näher an den Wert 0.27 % herangerückt werden, so kann dies für nur annähernd normalverteilte Daten durch eine Erhöhung der Parameter w und/oder s erreicht werden: Dies führt zu größeren Fenstern und/oder weniger Überlappungen zwischen diesen, was wiederum zu einer größeren Nähe zur Normalverteilung führt. Auch mit dem Parameter d ist eine entsprechende Beeinflussung möglich.

Abschließend wurde die Eignung des Verfahrens zur Detektion von Störfällen untersucht (Teil 6.2). Dazu wurden beispielhaft Messwerte von zwei realen Störungen nachgebildet. In beiden Fällen konnte das Verfahren diese erkennen, und das insbesondere bevor die absoluten Warngrenzen erreicht waren. Das entspricht einem Zeitgewinn bei

der Reaktion auf einen Fehlerfall, sowie auch bei der europaweiten Kommunikation von Maßnahmen. Bei der Untersuchung der Reaktion des Alarmsystems auf weitere Fehlerbilder (linearer Anstieg der Lage, Erhöhung der Varianz, sinusförmige Schwingungen) konnte simulativ dargestellt werden, welche Fehlerklassen das Verfahren aufdecken kann. Als nur teilweise hilfreich stellte sich schließlich die Hinzunahme zwei einfacher run rules heraus.

7.2 Fazit

Ohne Berücksichtigung der Wege und Begründungen für die Auswahl genau dieser Vorgehensweisen ergibt sich die folgende *take-home message*:

Für die Erstellung von Kontrollkarten bei mehrfachen und/oder starken Verletzungen der Annahmen von Standardmethodik ist bei einer Stichprobengröße $n \in \mathbb{N}_{>0}$ für (eventuell bereits pro Messzeitpunkt zusammengefasste) Beobachtungen x_i ($1 \leq i \leq n$) sowie für Parameter $d > 0$ und $s, w \in \mathbb{N}_{>0}$ das folgende Verfahren hilfreich:

1. Bildung von symmetrisch verteilten Differenzen: $\tilde{x}_i = |x_{i+1} - x_i|^d$, $d > 0$
2. Bildung gleitender Mittelwerte: $\tilde{y}_j = \frac{1}{w} \sum_{i=(j-1)s+1}^{(j-1)s+w} \tilde{x}_i$, $j = 1, 2, \dots, \lceil \frac{n-w}{s} \rceil$
3. Erstellen einer klassischen Urwertkarte auf Basis der \tilde{y}_j
4. Gegebenenfalls Behandlung von Autokorrelationen durch überlappende Fenster mit dem Korrekturfaktor nach Gilbert

Als allgemeine Empfehlung können $d = 0.25$, $w = 30$ und $s = w/2$ genannt werden, eine Optimierung je nach Datenlage ist ebenfalls möglich.

Es zeigte sich außerdem, dass zur Quantifizierung des Normalverteilungsgrades einer Stichprobe am besten Transformationen der Kolmogorov-Distanz K und der Teststatistik von Jarque und Bera JB geeignet sind, genauer für eine Stichprobengröße $n \in \mathbb{N}_{>0}$

$$LJB = \log(JB + 1)$$

und

$$K^* = \sqrt{n} K$$

7.3 Ausblick

Ansätze für weitergehende Analysen und Entwicklungen bieten sich an einigen Stellen der im Vorigen beschriebenen Arbeiten. So kann eine Darstellung der „Standard“-Verfahren, wie sie für das dritte Kapitel vorgenommen wurde, nur schwerlich *erschöpfend* oder *vollständig* sein, sondern muss sich auf klassische sowie auf durch ihre Konstruktion vielversprechende Methoden beschränken. Wie am Beispiel von EWMA, OSAPes und der Korrektur nach Gilbert gezeigt, sind zudem oft auch Kombinationen der Verfahren denkbar. Interessante weitere Ansätze könnten verteilungsfreie Kontrollkarten sein, wie sie zusammengefasst etwa von Trautmann (2004) vorgestellt werden.

Die praktische Unmöglichkeit einer *vollständigen* Berücksichtigung von Methoden gilt ebenso an weiteren Stellen, so stehen auch für Zeitreihenanalysen ungezählte Werkzeuge zur Verfügung. Weiter wurden auch die run rules nur im Ansatz getestet, hier könnten sowohl aus der Datenlage abgeleitete, als auch mit durch Expertenwissen aus der Anwendung motivierte Verfahren ergänzt werden.

Mit Blick auf die untersuchten und weitere Verfahren wäre auch ein eventueller Einfluss der Stichprobengröße von Interesse: Beispielsweise wurde bei der fehlgeschlagenen Verwendung der Faktoren F_W und F_G zur Behandlung der Autokorrelationen vermutet, dass das Vorgehen hier unter anderem nicht zum Erfolg führt, weil ein Bruch mit einem Term der Art $1 - r$ im Nenner für r nahe der Eins sehr sensibel auf Änderungen im Autokorrelationsschätzer r reagiert. Ist der verwendete Schätzer konsistent, so wird er für eine steigende Stichprobengröße durchschnittlich näher am wahren Wert liegen – was das Verwenden der Faktoren für *ausreichend große* Fallzahlen wieder zu einem besseren Werkzeug machen könnte.

Auch weitere Ansätze zur Konstruktion eines neuen Verfahrens lassen sich leicht entwickeln. So könnten, auch in Zusammenarbeit mit Experten des Anwendungsfeldes, andere Statistiken verwendet werden: Denkbar ist hier beispielsweise ein gewichteter Abstand vom Sollwert 50 Hertz. Für die Entwicklung neuer Methoden sollte der geringe Aufwand der Rechenoperationen beibehalten werden, um eine online-Beurteilung zu ermöglichen. Kontrollkarten auf der Basis von Differenzen sind, wie das hier vorgeschlagene Verfahren, für die Anwendung auch stets durch feste Grenzen für die Ausgangsreihe zu ergänzen: Durch die Betrachtung von Differenzen allein ist nicht auszuschließen, dass sich auch unauffällige Werte zu extremen Lagen der Messwerte summieren, also sehr langsam sehr große oder sehr kleine Werte entstehen. Mit Blick auf die Netzfrequenzen können hier zunächst die bisherigen, technisch motivierten Grenzen verwendet werden. Neben einer Anpassung der Prozessstatistik kann auch der Streuungsschätzer weiter in den Fokus

gerückt werden (beispielsweise eine Anwendung des Faktors nach Gilbert hat gerade eine Korrektur desselben als Ziel).

Die oben vorgestellte Methode zur Erstellung von Kontrollkarten unter Ausnutzen des Zentralen Grenzwertsatzes geht von einer recht großen Aufzeichnungsfrequenz der Daten aus: Nur dann lässt sich das dauerhafte Überspringen von 14 Messzeitpunkten rechtfertigen (bei $s = 15$). In klassischen Anwendungen wie der stündlichen Entnahme von einzelnen Erzeugnissen einer industriellen Produktionslinie würde sich – gemessen in Echtzeit – ein geradezu absurdes Alarmierungssystem ergeben. Die aus der großen Anzahl an Messungen pro Sekunde resultierende Größe der Datenmenge kann aber ebenfalls weiter ausgenutzt werden: Durch die Verwendung von Grenzen der Form *Zentrum + Faktor · Standardabweichung* soll im Wesentlichen ein bestimmtes Quantil der Verteilung der Daten geschätzt werden. Mit der vorliegenden Stichprobengröße sind hier auch andere Verfahren denkbar, etwa solche aus dem Bereich der Extremwertstatistik. In Kombination mit einer Anwendung des Zentralen Grenzwertsatzes wie oben ist dabei die Näherung von Shore (1986) ein möglicher Kandidat. Vom gleichen Autor wurde zudem eine Abschätzung des Fehlers der Approximation vorgelegt (Shore, 1988).

Wünschenswert wäre schließlich eine eindeutige Beschreibung des zu erkennenden Fehlerfalls, beziehungsweise von dessen Frequenzmustern. Diese liegt aus den im letzten Kapitel angeführten Gründen nicht vor, so dass die ermittelten Lauflängen nur erste Eindrücke bleiben können. Vor einer Anwendung des Verfahrens im hier vorgestellten Fall der Netzfrequenz oder auch in ganz anderen Bereichen ist somit bei der Frage der Qualität auch zu klären, welche Fehlerbilder berücksichtigt werden sollen.

8 Anhang

8.1 Ergänzende Abbildungen

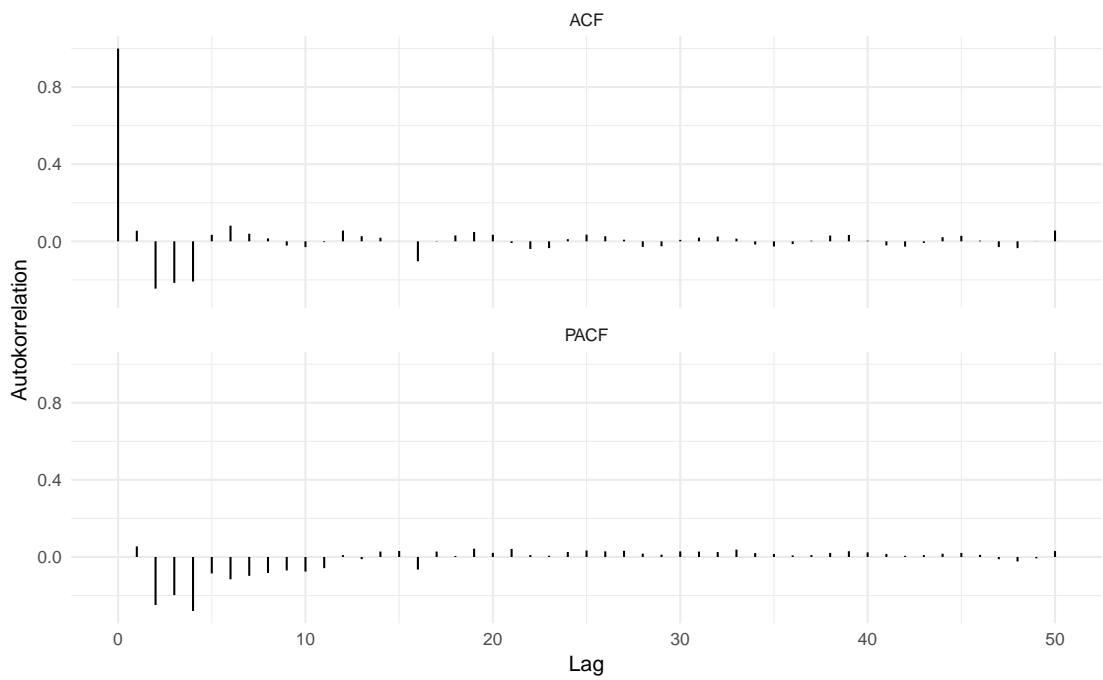


Abbildung 8.1: Ergänzend zur Abbildung 3.9: Die Verläufe der ACF und der PACF für die zweifache Differenzierung der Mittelwertreihe.

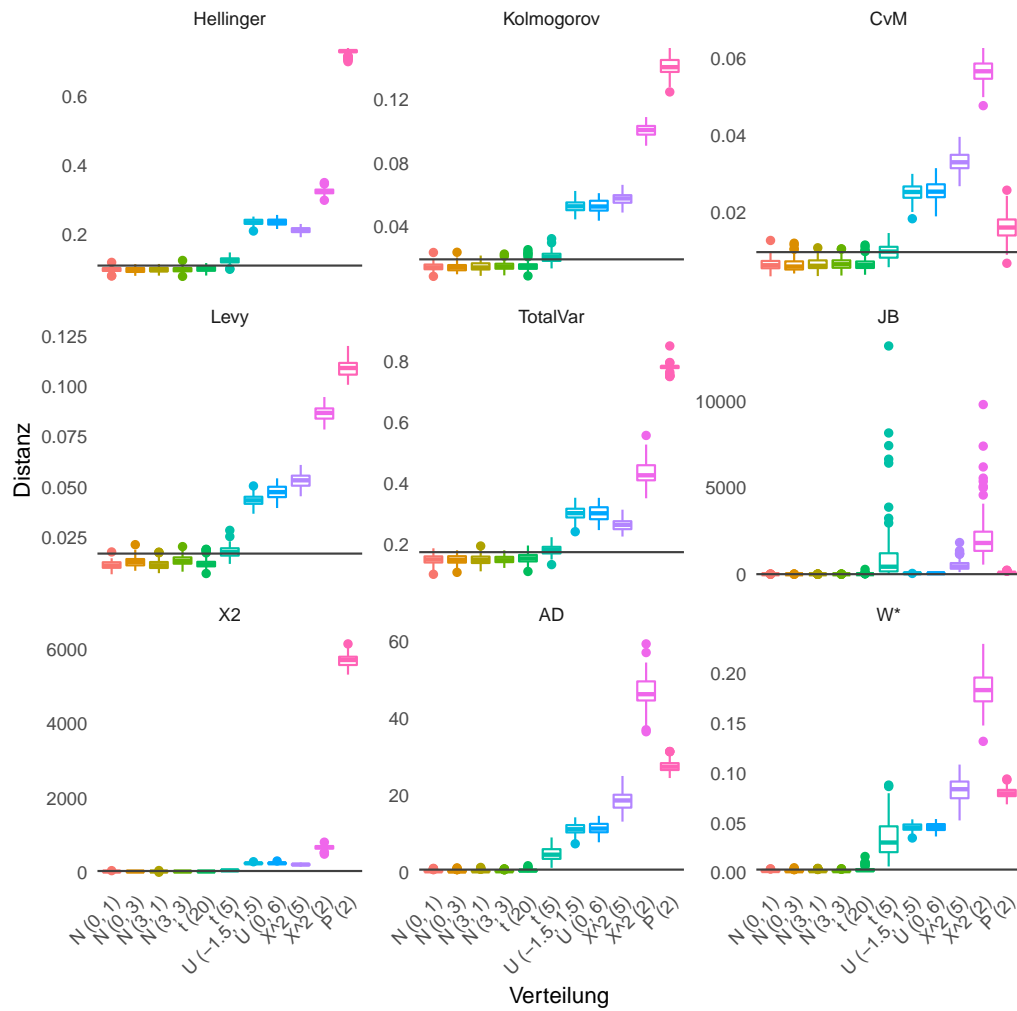


Abbildung 8.2: Ergänzend zu Teil 4.3.2 und den Abbildungen 4.5 und 4.6, hier die Simulationsergebnisse für die Stichprobengröße 1 000: Trotz der recht großen Stichprobe sind grafisch nicht alle Abweichungen der Daten von Normalverteilungen durch die Distanzen zu erkennen.

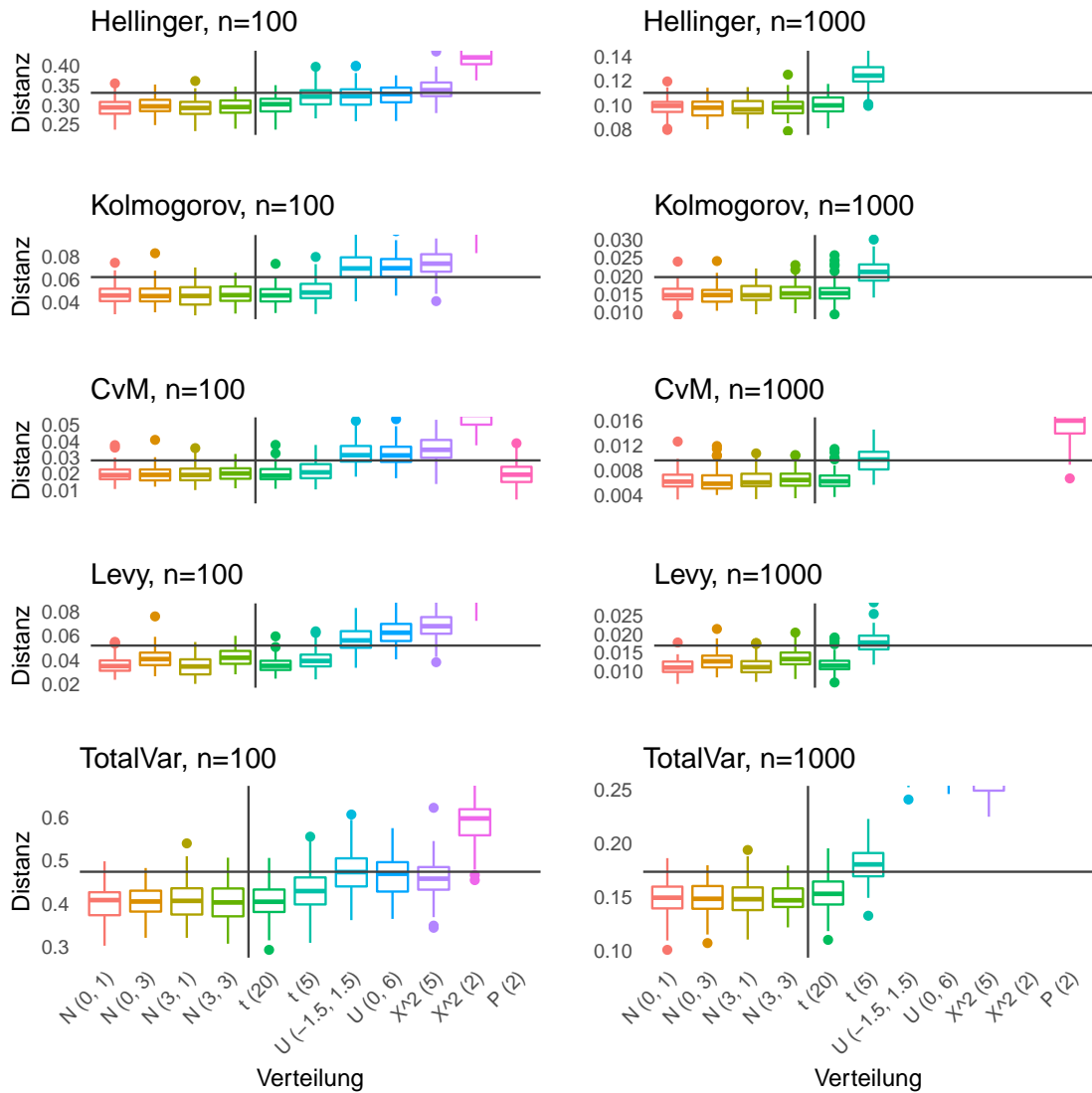


Abbildung 8.3: Ergänzend zu Teil 4.3.2 und Abbildung 4.7, die Simulationsergebnisse im „Zoom“ für kleine Distanzen und die Metriken.

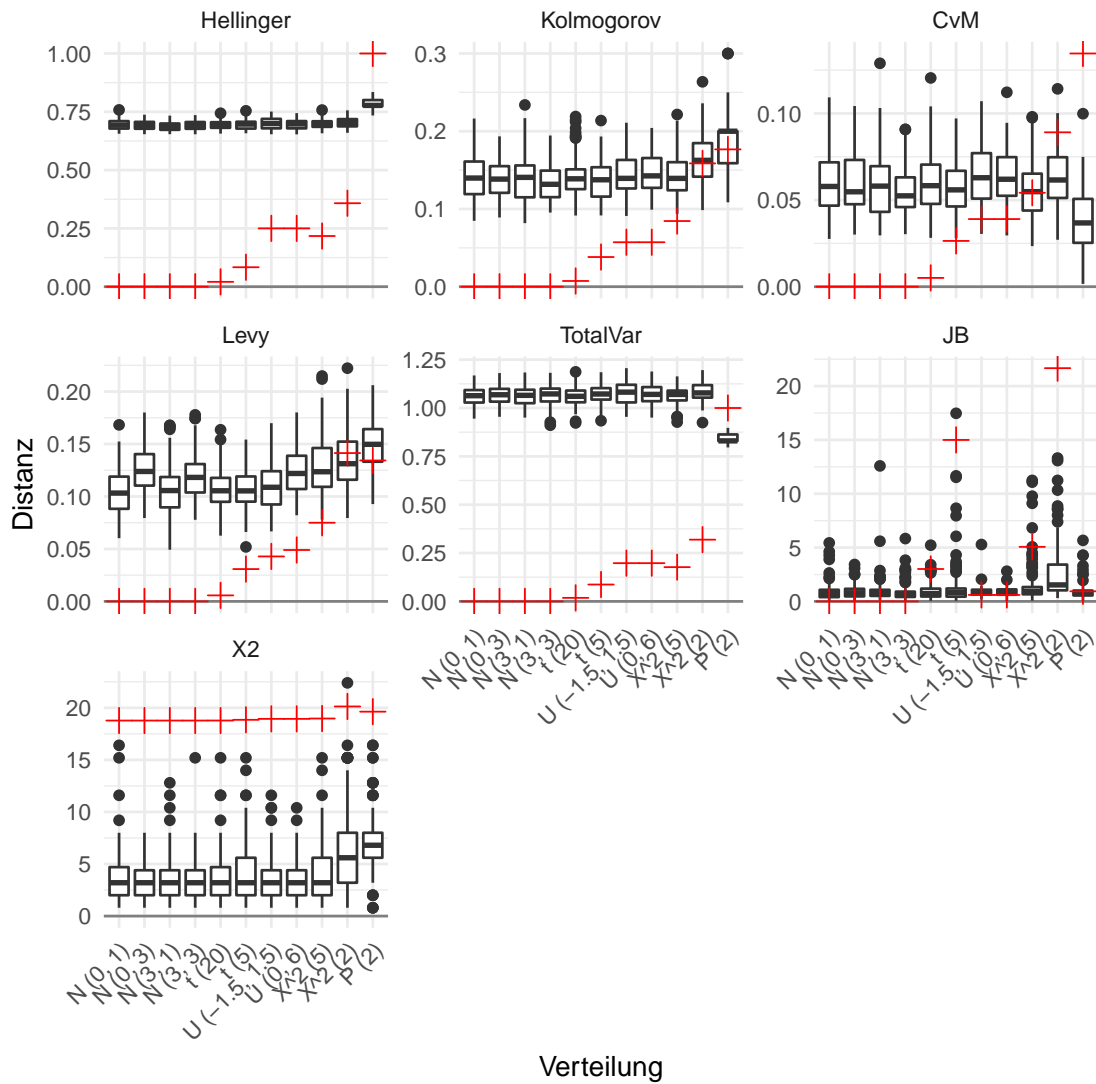


Abbildung 8.4: Ergänzend zu Teil 4.3.3 und Abbildung 4.8: Die Ergebnisse der Simulation mit den theoretisch zu erwartenden Werten (rot). Hier für die Stichprobengröße 10 und ohne AD und W^* .

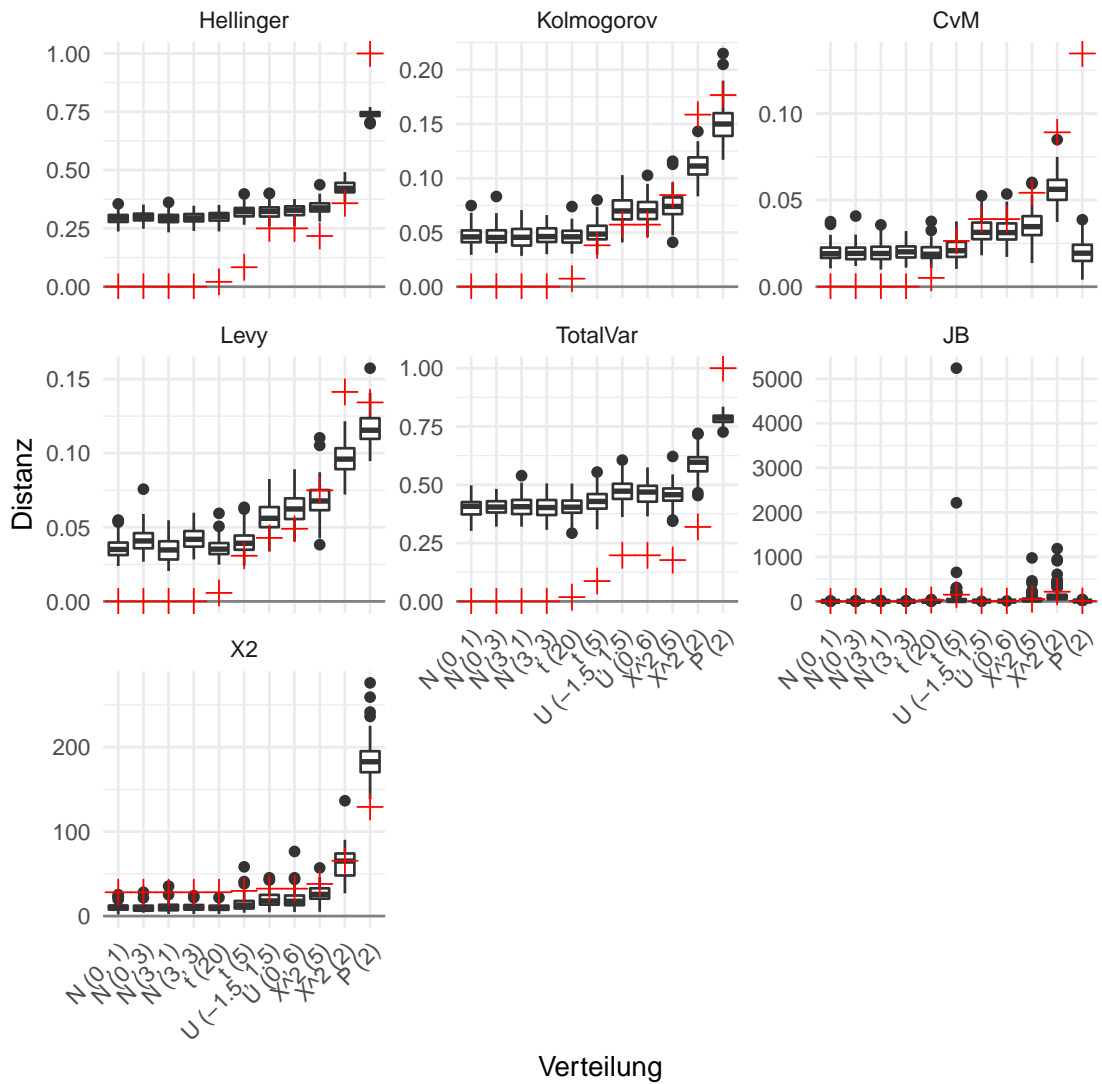


Abbildung 8.5: Ergänzend zu Teil 4.3.3 und Abbildung 4.8: Die Ergebnisse der Simulation mit den theoretisch zu erwartenden Werten (rot). Hier für die Stichprobengröße 100 und ohne AD und W^* .

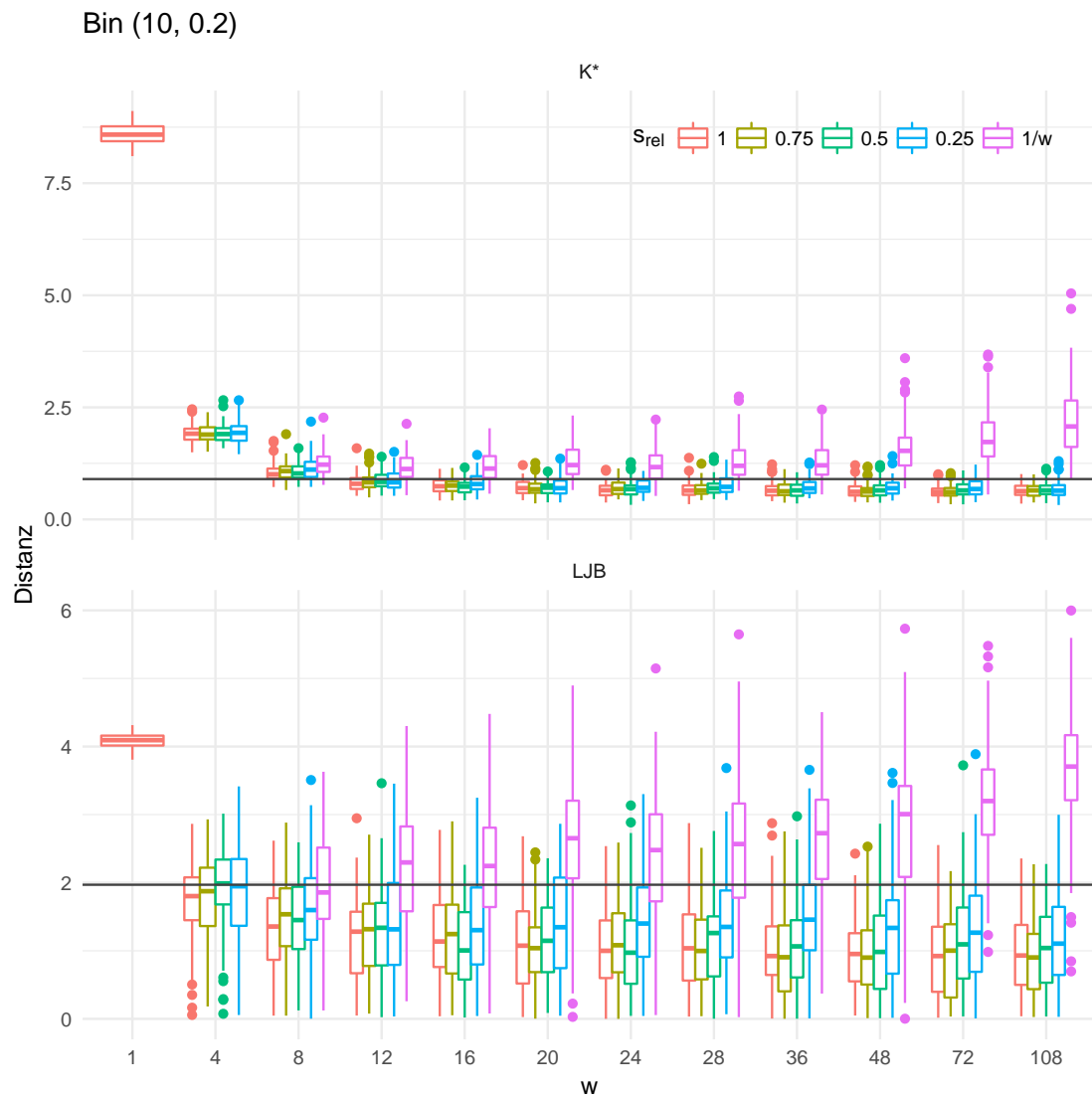


Abbildung 8.6: Ergänzend zu Abbildung 5.5: Die beiden Distanzmaße zur Normalverteilung für a priori Bin(10, 0.2)-verteilte Daten und $m = 1\,000$.

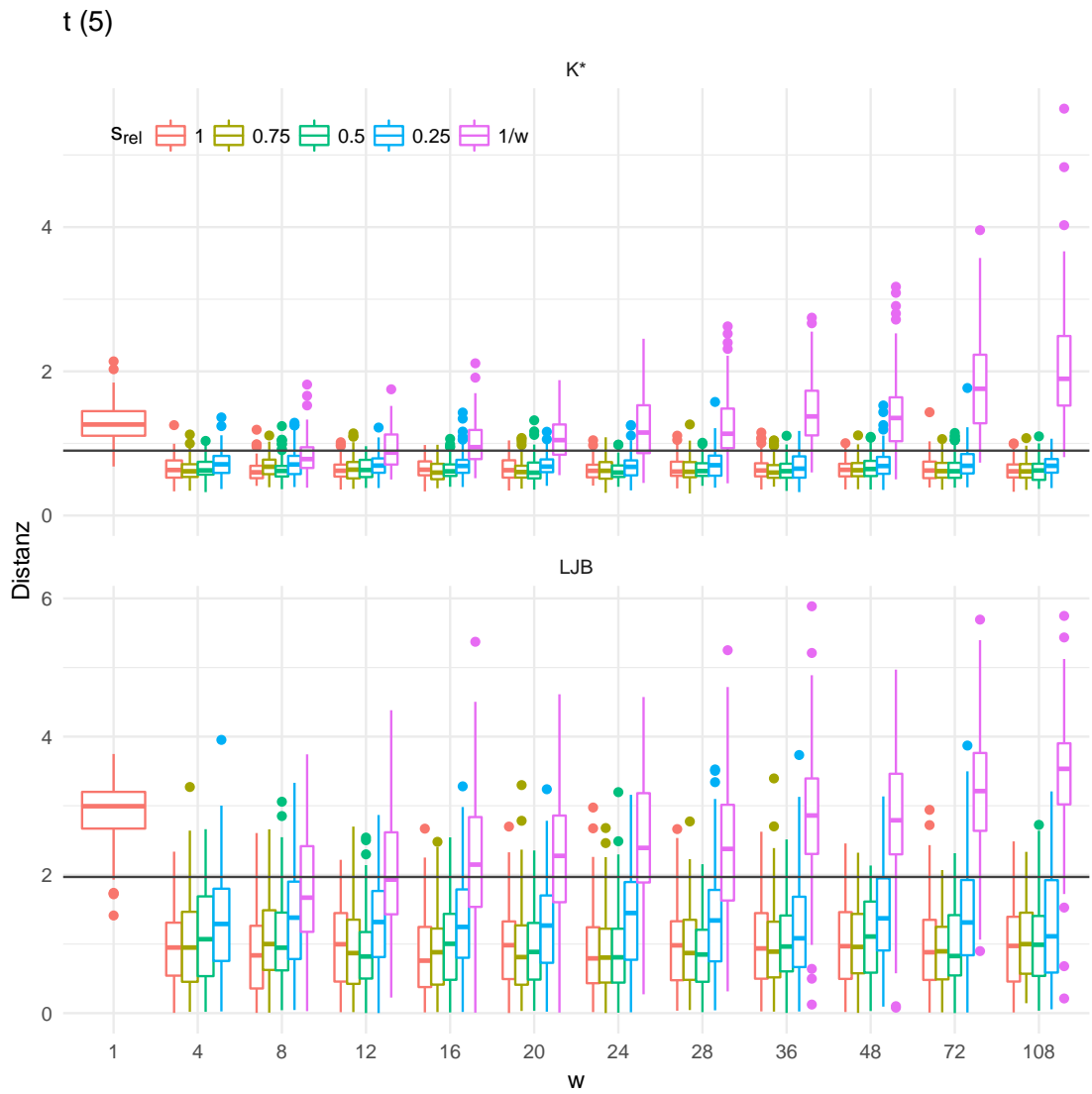


Abbildung 8.7: Ergänzend zu Abbildung 5.5: Die beiden Distanzmaße zur Normalverteilung für a priori t_5 -verteilte Daten und $m = 1\,000$.

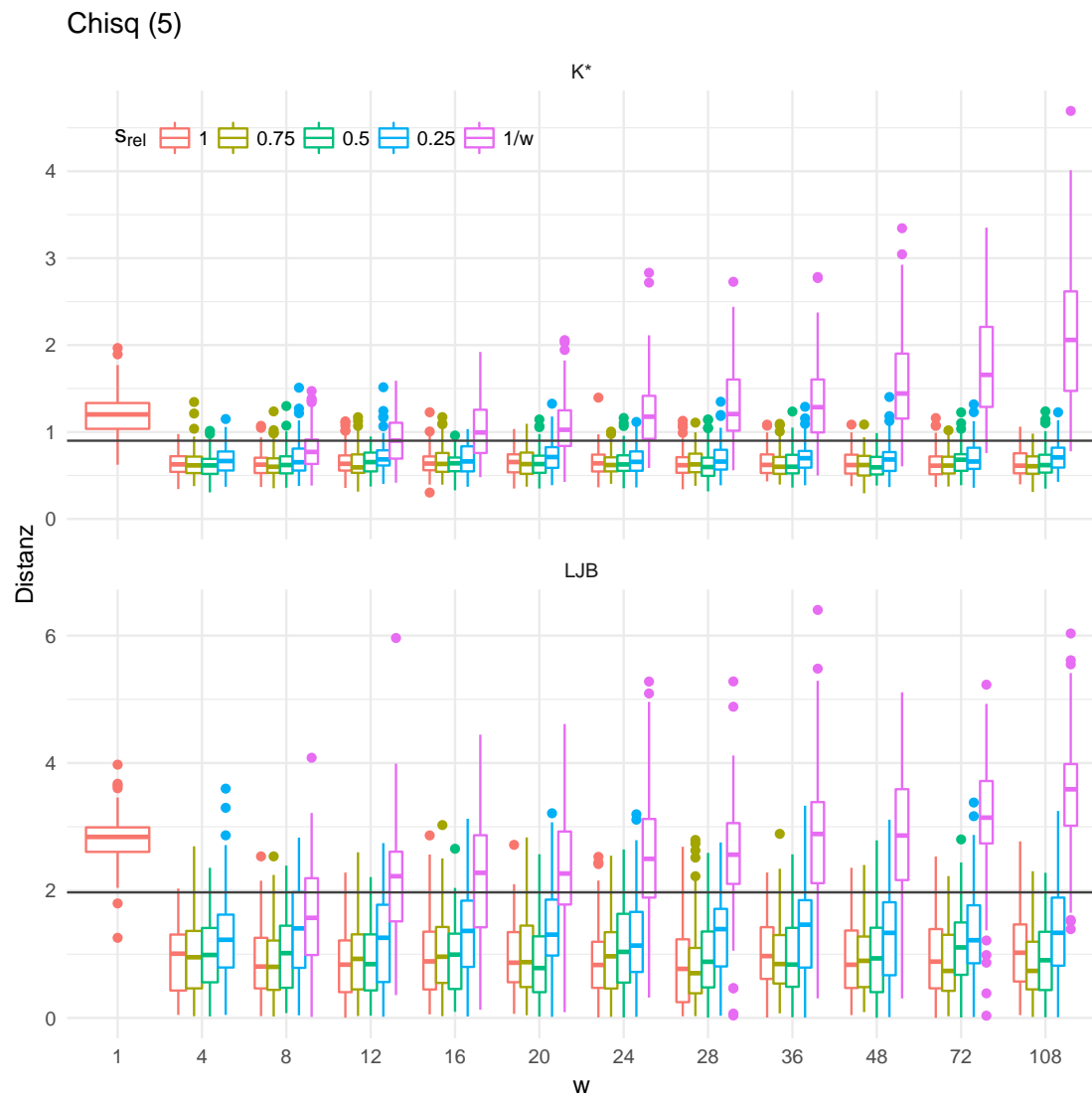


Abbildung 8.8: Ergänzend zu Abbildung 5.5: Die beiden Distanzmaße zur Normalverteilung für a priori χ^2_5 -verteilte Daten und $m = 1\,000$.

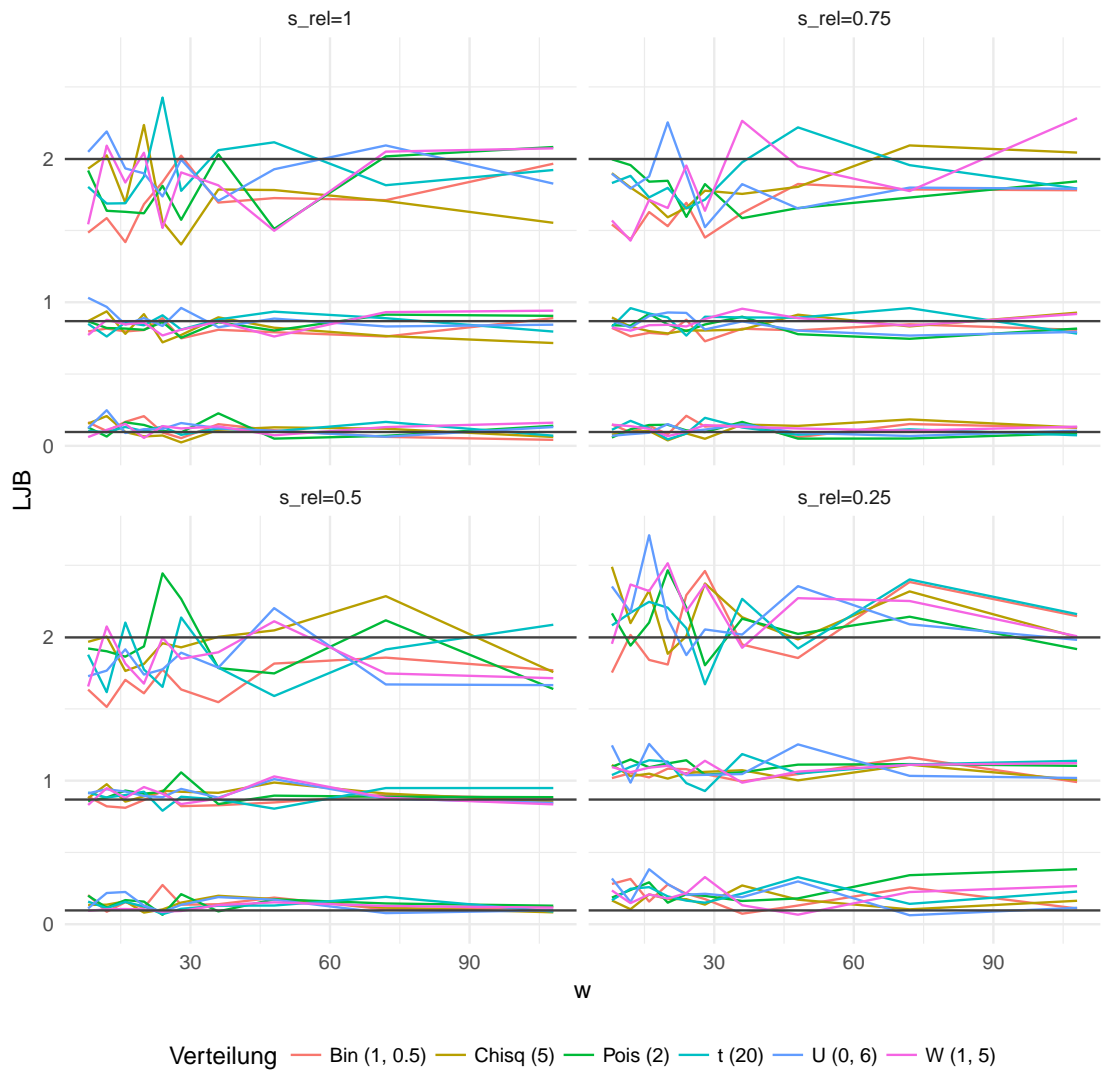


Abbildung 8.9: Ergänzend zu Abbildung 5.6 zeigt diese Grafik Verläufe von Quantilen und Mittelwert. Dabei werden hier die übrigen Verteilungen gezeigt, eine Stichprobengröße von $m = 100$ angesetzt und die LJB -Statistik als Distanzmaß verwendet.

8.2 Ergänzende Tabellen

Status	ρ	Faktor	KG	KG & $u = 7$	KG & $r = 20$	KG & $u = 7$ & $r = 20$
α	0	1.00	381	381	372	372
	0.5	1.00	408	408	393	393
	0.999	1.00	388	264	379	260
β	0	0.20	1	1	1	1
		0.50	2	2	2	2
		0.80	51	51	34	34
		1.50	5	5	5	5
		2.00	1	1	1	1
	0.5	0.20	1	1	1	1
		0.50	2	2	2	2
		0.80	38	38	26	26
		1.50	4	4	4	4
		2.00	1	1	1	1
	0.999	0.50	1	1	1	1
		0.80	36	35	24	23
		1.50	4	4	4	4
		2.00	1	1	1	1

Tabelle 8.1: Ergänzend zu Teil 6.2.4: Durchschnittliche Lauflängen je nach ausschließlicher Verwendung der Kontrollgrenzen (KG) und bei Hinzunahme der run rules, hier für das Beispiel mit einem Faktor (x_t^p).

Status	ρ	Steigung	KG	KG & $u = 7$	KG & $r = 20$	KG & $u = 7$ & $r = 20$
α	0	0.00	418	418	409	409
	0.5	0.00	402	402	391	391
	0.9	0.00	387	384	377	374
	0.99	0.00	383	289	374	282
	0.999	0.00	399	248	393	243
β	0	0.05	407	407	392	391
		0.10	407	407	396	396
		0.25	518	511	501	495
		0.50	550	166	441	154
		1.00	123	3	28	3
	0.5	0.05	390	390	384	384
		0.10	418	418	401	401
		0.25	536	438	489	409
		0.50	251	21	132	20
		1.00	7	1	6	1
	0.9	0.05	418	394	399	375
		0.10	494	267	464	258
		0.25	132	8	72	8
		0.50	2	1	2	1
		1.00	1	1	1	1
	0.05	291	26	253	26	
	0.10	23	3	19	3	
	0.25	1	1	1	1	
	0.05	2	1	2	1	
	0.10	1	1	1	1	

Tabelle 8.2: Ergänzend zu Teil 6.2.4: Durchschnittliche Lauflängen je nach ausschließlicher Verwendung der Kontrollgrenzen (KG) und bei Hinzunahme der run rules, hier für das Beispiel mit linearen Summanden (x_t^l).

Literaturverzeichnis

- [Aichele und Doleski, 2014] Christian Aichele und Oliver D. Doleski (Hrsg.), 2014: *Smart Market. Vom Smart Grid zum intelligenten Energiemarkt*. Springer Vieweg, Wiesbaden.
- [Alwan und Roberts, 1988] Layth C. Alwan und Harry V. Roberts, 1988: *Time-Series Modeling for Statistical Process Control*. Journal of Business & Economic Statistics, Jg. 6, Nr. 1, S. 87-95.
- [Alwan, 1991] Layth C. Alwan, 1991: *Autocorrelation: Fixed versus variable control limits*. Quality Engineering, Jg. 4, Nr. 2, S. 167-188.
- [Alwan und Radson, 1992] Layth C. Alwan und Darrell Radson, 1992: *Time-series investigation of subsample mean charts*. IEEE Transactions, Jg. 24, Nr. 5, S. 66-80.
- [Alwan, 1992] Layth C. Alwan, 1992: *Effects of autocorrelation on control chart performance*. Communications in Statistics – Theory and Methods, Jg. 21, Nr. 4, S. 1 025-1 049.
- [Alwan und Roberts, 1995] Layth C. Alwan und Harry V. Roberts, 1995: *The Problem of Misplaced Control Limits*. Journal of the Royal Statistical Society, Series C (Applied Statistics), Jg. 44, Nr. 3, S. 269-278.
- [Amprion, 2018] Amprion GmbH, 2018: *Netzführung und Systemeinsatz*. Online: <https://www.amprion.net/%C3%9Cbertragungsnetz/Systemf%C3%BChrung/Netzf%C3%BChrung/>. Letzer Abruf am 22. Mai 2018.
- [Anderson und Diaz, 1996] Elizabeth A. Anderson und Joaquin Diaz, 1996: *Using Process Control Chart Techniques to Analyse Crime Rates in Houston, Texas*. Journal of the Operational Research Society, Jg. 47, Nr. 7, S. 871-881.
- [Angus, 1993] John E. Angus, 1993: *A Central limit theorem for autoregressive integrated moving average processes*. Mathematical and Computer Modelling, Jg. 17, Nr. 10, S. 3-9.

- [Bauer, 2001] Heinz Bauer, 2001: *Wahrscheinlichkeitstheorie*. Walter de Gruyter Verlag, Berlin. Fünfte Auflage.
- [Beran, 1977] Rudolf Beran, 1977: *Minimum Hellinger Distance Estimates for Parametric Models*. The Annals of Statistics, Jg. 5, Nr. 3, S. 445-463.
- [BFE, 2003] Bundesamt für Energie, 2003: *Bericht über den Stromausfall in Italien am 28. September 2003*. Online: http://www.bfe.admin.ch/php/modules/publikationen/stream.php?extlang=de&name=de_117511357.pdf (letzter Abruf am 16. Mai 2018).
- [BNA, 2007] Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, 2007: *Bericht über die Systemstörung im deutschen und europäischen Verbundsystem am 4. November 2006*. Online: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Versorgungssicherheit/Berichte_Fallanalysen/Bericht_9.pdf?__blob=publicationFile&v=1 (letzter Abruf am 16. Mai 2018).
- [Boos, 1982] Dennis D. Boos, 1982: *Minimum anderson-darling estimation*. Communications in Statistics – Theory and Methods, Jg. 11, Nr. 24, S. 2 747-2 774.
- [Bollerslev, 1986] Tim Bollerslev, 1986: *Generalized Autoregressive Conditional Heteroskedasticity*. Journal of Econometrics, Jg. 31, Nummer 3, S. 307-327.
- [Burr, 1967] Irving W. Burr, 1967: *The Effect of Non-Normality on Constants for X and R Charts*. Industrial quality control, Jg. 23, Nr. 12, S. 563-569.
- [Champ und Woodall, 1987] Charles W. Champ and William H. Woodall, 1987: *Exact Results for Shewhart Control Charts with Supplementary Runs Rules*. Technometrics, Jg. 29, Nr. 4, S. 393-399.
- [Chen, 1997] Gemai Chen, 1997: *The mean and standard deviation of the run length distribution of \bar{x} charts when control limits are estimated*. Statistica Sinica, Jg. 7, Nr. 3, S. 789-798.
- [D'Agostino et al., 1990] Ralph B. D'Agostino, Albert Belanger und Ralph B. D'Agostino, Jr., 1990: *A suggestion for using powerful and informative tests of normality*. The American Statistician, Jg. 44, Nr. 4, S. 316-321.

- [dena, 2016] Deutsche Energie-Agentur GmbH (dena), 2016: *Analyse: Momentanreserve 2030: Bedarf und Erbringung von Momentanreserve 2030. Endbericht*. Online: https://shop.dena.de/fileadmin/denashop/media/Downloads_Dateien/esd/9142_Studie_Momentanreserve_2030.pdf, letzter Abruf am 22. Mai 2018.
- [DEW, 2017] Dortmunder Energie- und Wasserversorgung GmbH (DEW21), 2017: *120 Jahre Stromversorgung*. Online: <https://www.dew21.de/ueber-dew21/aktuelle-themen/120-jahre-stromversorgung-in-dortmund/>, letzter Abruf am 27. Mai 2018.
- [Dietrich und Schulze, 2014] Edgar Dietrich und Alfred Schulze, 2014: *Statistische Verfahren zur Maschinen- und Prozessqualifikation*. Carl Hanser Verlag, München. Siebte Auflage.
- [Elsberg, 2013] Marc Elsberg, 2013: *Blackout. Morgen ist es zu spät*. Blanvalet, München, 26. Auflage der Taschenbuchausgabe.
- [Engle, 1982] Robert F. Engle, 1982: *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. *Econometrica*, Jg. 50, Nr. 4, S. 987-1 007.
- [ENTSO-E, 2009] European Network of Transmission System Operators for Electricity, 2009: *Continental Europe Operation Handbook, Policy 1: Load-Frequency Control and Performance*. Online: https://www.entsoe.eu/fileadmin/user_upload/_library/publications/ce/oh/Policy1_final.pdf, letzter Abruf am 22. Mai 2018.
- [ENTSO-E, 2010] European Network of Transmission System Operators for Electricity, 2010: *Continental Europe Operation Handbook, Policy 5: Emergency Operations*. Online: https://www.entsoe.eu/fileadmin/user_upload/_library/publications/entsoe/Operation_Handbook/Policy_5_final.pdf, letzter Abruf am 22. Mai 2018.
- [ENTSO-E, 2018a] European Network of Transmission System Operators for Electricity, 2018: *ENTSO-E Member Companies*. Online: <https://www.entsoe.eu/about/inside-entsoe/members>, letzter Abruf am 22. Mai 2018.

- [ENTSO-E, 2018b] European Network of Transmission System Operators for Electricity, 2018: *ENTSO-E Transmission System Map*. Online: <https://www.entsoe.eu/data/map>, letzter Abruf am 22. Mai 2018.
- [Faraz et al., 2014] Alireza Faraz, Giovanni Celano, Erwin Saniga, C. Heuchenne und S. Fichera, 2014: *The variable parameters T^2 chart with run rules*. Statistical Papers, Jg. 55, Nr. 4, S. 933-950.
- [Genschel und Becker, 2005] Ulrike Genschel und Claudia Becker, 2005: *Schließende Statistik. Grundlegende Methoden*. Springer-Verlag, Heidelberg.
- [Gilbert et al., 1997] Kenneth C. Gilbert, Kenneth Kirby und Cheryl R. Hild, 1997: *Charting Autocorrelated Data: Guidelines For Practitioners*. Quality Engineering, Jg. 9, Nr. 3, S. 367-382.
- [Groß, 2004] Jürgen Groß, 2004: *A Normal Distribution Course*. Peter Lang Verlag, Frankfurt am Main.
- [Groß und Ligges, 2015] Jürgen Groß and Uwe Ligges, 2015: *nortest: Tests for Normality*. R-Paket, Version 1.0-4.
- [Hain, 2010] Johannes Hain, 2010: *Comparison of Common Tests for Normality*. Diplomarbeit am Lehrstuhl für Mathematik VIII (Statistik) der Julius-Maximilians Universität Würzburg, Institut für Mathematik und Informatik.
- [Hartung et al., 2005] Joachim Hartung, Bärbel Elpelt und Karl-Heinz Köster, 2005: *Lehr- und Handbuch der angewandten Statistik*. Oldenbourg Wissenschaftsverlag, München, 14. Auflage.
- [Hoeffding und Robbins, 1948] Wassily Hoeffding und Herbert Robbins, 1948: *The central limit theorem for dependent random variables*. Duke Mathematical Journal, Jg. 15, Nr. 3, S. 773-780. Erneut veröffentlicht von N. Fisher und P. Sen (beide Hrsg.), 1994: *The Collected Works of Wassily Hoeffding*. Springer-Verlag, New York.
- [Huber und Ronchetti, 2009] Peter J. Huber und Elvezio M. Ronchetti, 2009: *Robust Statistics*. John Wiley & Sons, Hoboken, New Jersey. Zweite Auflage.
- [Hyndman und Khandakar, 2008] Rob J. Hyndman und Yeasmin Khandakar, 2008: *Automatic Time Series Forecasting: The forecast Package for R*. Journal of Statistical Software, Jg. 28, Nr. 3, S. 1-22.

- [Jarret und Pan, 2007] Jeffrey E. Jarrett und Xia Pan, 2007: *The quality control chart for monitoring multivariate autocorrelated processes*. Computational Statistics & Data Analysis, Jg. 51, Nr. 8, S. 3 862-3 870.
- [Jarque und Bera, 1987] Carlos M. Jarque und Anil K. Bera, 1987: *A Test for Normality of Observations and Regression Residuals*. International Statistical Review, Jg. 55, Nr. 2, S. 163-172.
- [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.
- [Khoo, 2013] Michael B. C. Khoo, 2013: *Power functions of the Shewhart control chart*. Journal of Physics: Conference Series, Ausgabe 423 zur SciTech 2013.
- [Klenke, 2006] Achim Klenke, 2006: *Wahrscheinlichkeitstheorie*. Springer-Verlag, Heidelberg.
- [Köksal et al., 2008] Gülser Köksal, Burcu Kantar, Taylan Ali Ula und Murat Caner Testik, 2008: *The effect of Phase I sample size on the run length performance of control charts for autocorrelated data*. Journal of Applied Statistics, Jg. 35, Nr. 1, S. 67-87.
- [Kohl, 2005] Matthias Kohl, 2005: *Numerical Contributions to the Asymptotic Theory of Robustness*. Dissertation, Universität Bayreuth.
- [Komsta und Novomestky, 2015] Lukasz Komsta und Frederick Novomestky, 2015: *moments: Moments, cumulants, skewness, kurtosis and related tests*. R-Paket, Version 0.14.
- [Kubis et al., 2015] Andreas Kubis, Lena Robitzky, Markus Küch, Sven C. Müller, Boguslaw Jablkowski, Hanno Georg, Nils Dorsch, Sebastian Krey, Christian Langesberg, Dirk Surmann, Daniel Mayorga, Christian Rehtanz, Ulf Häger, Olaf Spinczyk, Christian Wietfeld, Claus Weihs, Uwe Ligges, Johanna Myrzik und Jürgen Götz, 2015: *Validation of ICT-Based Protection and Control Applications in Electric Power Systems*. IEEE PowerTech, Eindhoven, Juni 2015.
- [Langesberg et al., 2018a] 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, <http://hdl.handle.net/2003/36856>, letzter Abruf am 27. Mai 2018.

- [Langesberg et al., 2018b] Christian Langesberg, Uwe Ligges und Claus Weihs, 2018: *Zur Verteilungs-Konvergenz von gleitenden Mittelwerten*. Eldorado, Sammlung der DFG-Forschergruppe: Schutz- und Leittechnik zur zuverlässigen und sicheren Energieübertragung, <http://hdl.handle.net/2003/36861>, letzter Abruf am 27. Mai 2018.
- [Lewis, 1961] Peter A. W. Lewis, 1961: *Distribution of the Anderson-Darling Statistic*. The Annals of Mathematical Statistics, Jg. 32, Nr. 4, S. 1 118-1 124.
- [Lu und Reynolds, 1999] Chao-Wen Lu und Marion R. Reynolds, Jr., 1999: *EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes*. Journal of Quality Technology, Jg. 31, Nr. 2, S. 166-188.
- [Maragah und Woodall, 1992] Hazem D. Maragah und William H. Woodall, 1992: *The effect of autocorrelation on the retrospective X-chart*. Journal of Statistical Computation and Simulation, Jg. 40, Nr. 1-2, S. 29-42.
- [Mayer-Schönberger und Cukier, 2013] Viktor Mayer-Schönberger und Kenneth Cukier, Übersetzung von Dagmer Mallett, 2013: *Big Data*. Redline Verlag, München.
- [Montgomery und Mastrangelo, 1991] Douglas C. Montgomery und Christina M. Mastrangelo, 1991: *Some Statistical Process Control Methods for Autocorrelated Data*. Journal of Quality Technology, Jg. 23, Nr. 3, S. 179-193.
- [Montgomery, 1985] Douglas C. Montgomery, 1985: *Introduction to Statistical Quality Control*. Wiley, New York.
- [Montgomery, 1991] Douglas C. Montgomery, 1991: *Introduction to Statistical Quality Control*. Wiley, New York. Zweite Auflage.
- [Montgomery, 2009] Douglas C. Montgomery, 2009: *Introduction to Statistical Quality Control*. John Wiley & Sons, Hoboken, New Jersey. Sechste Auflage.
- [Montgomery, 2013] Douglas C. Montgomery, 2013: *Statistical Quality Control: A Modern Introduction*. John Wiley & Sons, Singapur. Siebte Auflage (International Student Version).
- [Neidhöfer, 2008] Gerhard Neidhöfer, 2008: *Der Weg zur Normfrequenz 50 Hz*. Bulletin SEV/AES, Nr. 17/2008, Electrosuisse und der Verband Schweizerischer Elektrizitätsunternehmen. S. 29-34.

- [Noffsinger und Anderson, 2002] John R. Noffsinger und R. Bruce Anderson, 2002: *Effect of autocorrelated data on composite panel production monitoring and control: A comparison of SPC techniques*. Forest Products Journal, Jg. 52, Nr. 3, S. 60-67.
- [Noughabi und Arghami, 2011] Hadi Alizadeh Noughbai und Naser Reza Arghami, 2011: *Monte Carlo comparison of seven normality tests*. Journal of statistical computation and simulation. Jg. 11, Nr. 81, S. 965-972.
- [Olds, 1961] Edwin G. Olds, 1961: *Power Characteristics of Control Charts*. Industrial quality control, Jg. 18, Nr. 1, S. 4-10.
- [Page, 1955] E. S. Page, 1955: *Control Charts with Warning Lines*. Biometrika, Jg. 42, Nr. 1/2, S. 243-257.
- [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, 2018] R Core Team, 2018: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Wien.
- [Requeijo und Souza, 2001] José Gomes Requeijo und Adriano Mendonça Souza, 2011: *T² Control Chart to Processes with cross-autocorrelation*. Proceedings of the 41st International Conference on Computers & Industrial Engineering, Los Angeles.
- [Rieder, 1994] Helmut Rieder, 1994: *Robust Asymptotic Statistics*. Springer-Verlag, New York.
- [Roberts, 1957] S. W. Roberts, 1957: *Properties of Control Chart Zone Tests*. Bell System Technical Journal, Jg. 37, Nr. 1, S. 83-114.
- [Ruckdeschel et al., 2006] P. Ruckdeschel, M. Kohl, T. Stabla und F. Camphausen, 2006: *S4 Classes for Distributions*. R News, Jg. 6, Nr. 2 (R-Paket distrEx, Version 2.6.1).
- [Rüger, 2002] Bernhard Rüger, 2002: *Test- und Schätztheorie, Band II: Statistische Tests*. Oldenbourg Wissenschaftsverlag, München.
- [Runger und Willemain, 1996] George C. Runger und Thomas R. Willemain, 1996: *Batch-means control charts for autocorrelated data*. IIE Transactions, Jg. 28, Nr. 6, S. 483-487.

Literaturverzeichnis

- [Schipper und Schmid, 2001] S. Schipper und W. Schmid, 2001: *Control charts for GARCH processes*. Nonlinear Analysis: Theory, Methods & Applications, Jg. 47, Nr. 3, S. 2 049-2 060.
- [Schmid und Schöne, 1997] Wolfgang Schmid und Alexander Schöne, 1997: *Some properties of the EWMA control chart in the presence of autocorrelation*. The Annals of Statistics, Jg. 25, Nr. 3, S. 1 277-1 283.
- [Scrucca, 2004] Luca Scrucca, 2004: *qcc: an R package for quality control charting and statistical process control*. R News, Jg. 4, Nr. 1, S. 11-17.
- [Serfling, 1968] R. J. Serfling, 1968: *Contributions to central limit theory for dependent variables*. The Annals of Mathematical Statistics, Jg. 39, Nr. 4, S. 1 158-1 175.
- [Shapiro und Wilk, 1965] S. S. Shapiro und M. Wilk, 1965: *An analysis of variance test for normality (complete samples)*. Biometrika, Jg. 52, Nr. 3 und 4, S. 591-611.
- [Shevtsova, 2011] Irina Shevtsova, 2011: *On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands*. arXiv:1111.6554 [math.PR].
- [Shewhart, 1931] Walter A. Shewhart, 1931: *Economic Control of Quality of Manufactured Control*. D. van Nostrand Company, New York. Siebte Auflage.
- [Shore, 1986] Haim Shore, 1986: *Simple general approximations for a random variable and its inverse distribution function based on linear transformations of a non-skewed variate*. SIAM Journal on Scientific and Statistical Computing, Jg. 7, Nr. 1, S. 1-23.
- [Shore, 1988] Haim Shore, 1988: *An Approximation for the Error of the Normal Approximation to a Linear Combination of Independently Distributed Random Variables*. IIE Transactions, Jg. 20, Nr. 3, S. 242-246.
- [Sürücü, 2008] Barış Sürücü, 2008: *A power comparison and simulation study of goodness-of-fit tests*. Computers and Mathematics with Applications, Jg. 56, S. 1 617-1 625.
- [SZ, 2011] Susi Wimmer und Jakob Wetzels, sueddeutsche.de, 16.2.2011: *Dem Verbrechen auf der Spur*. Online: <http://www.sueddeutsche.de/muenchen/landeskriminalamt-dem-verbrehen-auf-der-spur-1.1061222>, letzter Abruf am 22. Mai 2018.

- [Tagesschau, 2011] Werner Eckert, Südwestrundfunk: *Ein Netz für „Dinosaurier“ statt für „Kleinvieh“*. Online: <https://www.tagesschau.de/wirtschaft/stromnetze104.html>, letzter Abruf am 27. Mai 2018.
- [Tagesschau, 2017] Malte Pieper, MDR, 18.10.2017: *Wenn der Wecker zu früh klingelt*. Online: <https://www.tagesschau.de/schlusslicht/uhrenchaos-belgien-101.html>, letzter Abruf am 22. Mai 2018.
- [Tagesschau, 2018] Michael Heussen, WDR, 7.3.2018: *Wenn Uhren plötzlich anders ticken*. Online: <https://www.tagesschau.de/ausland/stromnetz-schwankungen-101.html>, letzter Abruf am 22. Mai 2018.
- [Thadewald und Büning, 2007] Thorsten Thadewald und Herbert Büning, 2007: *Jarque-Bera Test and its Competitors for Testing Normality – A Power Comparison*. Journal of Applied Statistics, Jg. 34, Nr. 1, S. 87-105.
- [Tippett, 1925] L. H. C. Tippett, 1925: *On the extreme individuals and the range of samples taken from a normal population*. Biometrika, Jg. 17, Nr. 3-4, S. 364-387.
- [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, Jg. 25, Nr. 4, S. 790-809.
- [Trautmann, 2004] Heike Trautmann, 2004: *Qualitätskontrolle in der Industrie anhand von Kontrollkarten für Wünschbarkeitsindizes – Anwendungsfeld Lagerhaltung*. Dissertation an der Universität Dortmund.
- [UCTE, 2004] Union for the Coordination of Transmission of Electricity, 2004: *Final Report of the Investigation Committee on the 28 September 2003 Blackout in Italy*. Online: https://www.entsoe.eu/fileadmin/user_upload/_library/publications/ce/otherreports/20040427_UCTE_IC_Final_report.pdf (letzter Abruf am 16. Mai 2018).
- [UCTE, 2007] Union for the Coordination of Transmission of Electricity, 2007: *Final Report: System Disturbance on 4 November 2006*. Online: https://www.entsoe.eu/fileadmin/user_upload/_library/publications/ce/otherreports/Final-Report-20070130.pdf (letzter Abruf am 16. Mai 2018).
- [Venables und Ripley, 2002] William N. Venables und Brian D. Ripley, 2002: *Modern Applied Statistics with S*. Springer-Verlag, New York. Vierte Auflage.

Literaturverzeichnis

- [Wardell et al., 1994] Don G. Wardell, Herbert Moskowitz und Robert D. Plante, 1994: *Run-Length Distributions of Special-Cause Control Charts for Correlated Processes*. Technometrics, Jg. 36, Nr. 1, S. 3-17. Taylor & Francis, Abingdon, England.
- [Weihs und Jessenberger, 1999] Claus Weihs und Jutta Jessenberger, 1999: *Statistische Methoden zur Qualitätssicherung und -optimierung: in der Industrie*. Wiley-VCH, Weinheim.
- [Wheeler, 1983] Donald J. Wheeler, 1983: *Detecting a Shift in Process Average: Tables of the Power Function for \bar{X} Charts*. Journal of quality technology, Jg. 15, Nr. 4, S. 150-170.
- [Wheeler, 1995] Donald J. Wheeler, 1995: *Advanced Topics in Statistical Process Control – The Power of Shewhart’s Charts*. SPC Press, Knoxville, Tennessee.
- [Witting, 1985] Hermann Witting, 1985: *Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumfang*. B. G. Teubner, Stuttgart.
- [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.
- [Yap und Sim, 2011] B. W. Yap und C. H. Sim, 2011: *Comparisons of various types of normality tests*. Journal of Statistical Computation and Simulation, Jg. 81, Nr. 12, S. 2 141-2 155.
- [Young und Winistrofer, 2001] Timothy M. Young und Paul M. Winistrofer, 2001: *The effects of autocorrelation on real-time statistical process control with solutions for forest products manufacturers*. Forest Products Journal, Jg. 51, Nr. 11/12, S. 70-77.

Die Bemerkungen auf der Seite 3 sind zitiert nach Larry Webber und Michael Wallace, *Quality Control for Dummies*, 2007, Wiley Publishing, Hoboken, New Jersey, Seite 155, sowie aus *Die Simpsons*, Fernsehserie von Matt Groening, Staffel 2, Episode 22 „Der Lebensretter“.