

KONFIDENZINTERVALLE BEI GERUNDETEN DATEN

DISSERTATION

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
der Technischen Universität Dortmund

Der Fakultät Statistik der Technischen Universität Dortmund
vorgelegt von

Dipl. Stat. Dorothea Schipp

Gutachter:

Prof. Dr. Walter Krämer
Prof. Dr. Götz Trenkler

Tag der mündlichen Prüfung: 31. Januar 2008

Inhaltsverzeichnis

1	Einleitung	1
2	Gerundete Daten aus stetigen Verteilungen	4
2.1	Einführung	4
2.2	Einflüsse des Rundens auf die Momente	7
2.3	Effekte bei Normalverteilung der Grundgesamtheit	10
2.4	Effekte bei Exponentialverteilung der Grundgesamtheit	12
2.5	ML-Schätzung bei gerundeten Daten	15
3	Konfidenzintervalle bei gerundeten Daten	21
3.1	Das Pivot-Verfahren	21
3.2	Das Likelihood-Quotienten-Verfahren	22
3.3	Die χ^2 -Approximation	24
3.4	Das Verfahren nach Wald	27
3.5	Die Methoden im Vergleich	29
3.6	Ein Korrekturverfahren von Lee und Vardeman	32
4	Ein Kombinationsverfahren für Lageparameter	37
4.1	Motivation	37
4.2	Das Grundprinzip	41
4.3	Das Simulationsdesign	44
4.4	Programmiertechnische Details	48
5	Kombinierte Intervalle bei stetigen Verteilungsalternativen	55
5.1	Konfidenzintervalle für die Maximumverteilung	55
5.2	Eine Reanalyse für die Normalverteilung	72
5.3	Konfidenzintervalle für die Logistische Verteilung	88
5.4	Konfidenzintervalle für die Exponentialverteilung	104
6	Transformationen oder Schätzung des Skalenparameters	111

7 Ein Anwendungsbeispiel aus der Hydrologie	114
7.1 Das „Jahrhundert“-Hochwasser der Elbe	114
7.2 Grundbegriffe aus der Hydrologie	115
7.3 Die Genauigkeit von Durchflussdaten	117
7.4 Eine Auswertung der Messreihe Rehefeld1/Wilde Weißeritz	119
8 Zusammenfassung und Ausblick	127
Abbildungsverzeichnis	130
Tabellenverzeichnis	132
Literatur	133

1 Einleitung

Die Erhebung exakter Daten verkörpert ein theoretisches Ideal, dem in der wissenschaftlichen Praxis natürliche Grenzen gesetzt sind. Digitale Messgeräte mit begrenzter Ziffernanzeige registrieren die nächstliegende letzte Ziffer ähnlich wie Ablesungen an einer Skala die nächstliegende Markierung. Die nachfolgende Datenverarbeitung verzichtet meist auf das zeitintensive Einlesen langer Nachkommareihen und reduziert die ohnehin „diskretisierten“ Daten auf ein noch gröberes Messraster zu Lasten der Datengenauigkeit. Allgegenwärtige Rundungsmechanismen verkürzen die theoretisch unendlich langen Ziffernfolgen, vereinen benachbarte Werte zu Intervallen und weisen allen Werten eines Intervalls den gleichen Messwert zu. So sieht sich der Statistiker bei der Auswertung des gesammelten Datenmaterials in der Regel mit folgendem Dilemma konfrontiert: Einerseits weisen die Daten spezifische Kennzeichen auf, die den Schluss auf einen stetigen Verteilungstyp zulassen, andererseits ist die in der statistischen Methodik verlangte Stetigkeit der Zufallsvariablen streng genommen nicht gegeben, denn unendlich viele Nachkommastellen einer zu messenden Variable sind weder ermittelbar noch operabel.

Eine vielfach unterstellte und dennoch in der Realität nie erfüllte Modellannahme verlangt aber gerade, dass die Messung der Zielvariablen mit beliebiger Genauigkeit durchgeführt werden kann.

Yet all of us have been brought up on statistical theory and methods intended to deal with „continuous“ data - data that none of us will ever see! (Burrige, 1989)

Zwar gäbe es auch nichtparametrische Ansätze (vgl. z.B. Büning und Trenkler, 1994), die keineswegs eine Stetigkeit der Daten fordern, aber die Information über die Art der Verteilung ginge verloren. Aus dieser Not heraus werden die ermittelten Daten behandelt wie stetige, obwohl es sich de facto um gerundete, also zu gewissen Intervallen gruppierte, handelt.

Daraus resultierende Schwierigkeiten bei der induktiven Statistik und der Interpretation der Ergebnisse werden typischerweise verkannt oder billigend in Kauf genommen.

Zusätzlich zu einer Parameterschätzung der unterstellten Verteilung wird auch eine Aussage über die Verlässlichkeit dieser Schätzung erwartet. Diese Aussage manifestiert sich üblicherweise in der Berechnung von Konfidenzintervallen, die einen Bereich abstecken, innerhalb dessen sich der tatsächliche Parameter mit einer gewissen Sicherheit befindet. Wie viel Sicherheit aber kann ein Konfidenzintervall bieten, welches die oben skizzierten Gegebenheiten außer Acht lässt?

Die vorliegende Arbeit evaluiert zunächst die Eignung bekannter Methoden zur Konstruktion von Konfidenzintervallen im Umgang mit gerundeten Daten. Dabei interessiert sowohl die Art der Auswirkungen einer Missachtung der Rundung als auch eine sorgfältige Erfassung aller Umstände, die zu einem besonderen Schweregrad der statistischen Fehleinschätzung beitragen können. Als Konsequenz wird eine optimierte Auswertungsstrategie für gerundete Daten entwickelt, die den tatsächlich vorliegenden Modellvoraussetzungen Rechnung trägt. Die vorgeschlagene Methode ist mittels einer differenzierten Situationsanalyse in der Lage, rechtzeitig geeignete Maßnahmen ergreifen und so allzu gravierenden Fehlschlüssen vorzubeugen. Der Handlungsspielraum ist meist durch die steuerbaren Parameter des Studiendesigns vorgegeben. Unter der Prämisse einer vorab regulierbaren Messpräzision lässt sich prospektiv eine Präzisionsstufe fixieren, die gute Qualitäten der Methodik zusichert. Nach der Datenerhebung mit vorgegebener Genauigkeit kann retrospektiv eine auf die konkreten Datengegebenheiten abgestimmte Auswertungsstrategie helfen, Fehlschlüsse zu minimieren.

Ein Schwerpunkt der Arbeit liegt auf der Auswertung kleiner Datensätze. Denn zu den angesprochenen Problemen kommt meist erschwerend hinzu, dass die gewünschten Informationen aus oft nur wenigen Beobachtungen zu gewinnen sind. Man denke etwa an kostspielige technische Apparaturen, die nicht in beliebiger Stückzahl auf ihre Lebensdauer getestet werden können, oder an die moderne Klimaforschung. Hier sind extreme Klimaphänomene zwar in letzter Zeit vermehrt, aber dennoch nur selten zu beobachten. Meist reicht die Wetterbeobachtung nur wenige Jahrzehnte zurück, aber selbst bei einer mehrere Jahrhunderte langen Beobachtungsreihe müssen wenige Messwerte genügen, um die Extreme zu erforschen.

Nach einer allgemeinen Einführung in die Rundungsthematik folgt in Kapitel 2 ein Überblick über historische Beiträge. Einige bedeutende Arbeiten werden vorgestellt, die sich nach damals zeitgemäßer Denkweise mit der Momentenschätzung und der später aufkommenden Maximum-Likelihood-Schätzung befassen. Obwohl die Thematik des Rundens alles andere als neu ist – erste Ansätze datieren zurück in das Ende des 19. Jahrhunderts (vgl. Sheppard, 1898) – hat sie durch die Übertragung der Lösungsansätze auf andere Wissenschaftszweige und weitere Verteilungsmodelle kaum an Brisanz verloren. Im Gegenteil: Die Einbettung in den Kontext der Lebenszeitanalyse hat die Problematik aufs Neue in den Mittelpunkt des Interesses gerückt, da hier das Runden einen Spezialfall im Zusammenhang mit intervallzentrierten Daten darstellt.

Kapitel 3 wendet sich der ML-Intervallschätzung zu, zeigt exakte sowie approximative methodische Ansätze auf und stellt verschiedene Vergleiche neueren Datums vor. Diese gaben Ansporn zu der sich anschließenden Detailanalyse und lieferten

wichtige Denkanstöße für Korrekturmaßnahmen.

Die Kapitel 4 und 5 bilden das eigentliche Kernstück der Arbeit. Sie setzen die Qualitätsprüfung fort und zeigen empirisch auf, inwieweit sich gerundete Daten negativ auf das praxisübliche Pivot-Verfahren auswirken. Dabei werden die bekannten Informationen über wenige Spezialtypen auf ein breiteres Spektrum von Verteilungen erweitert. Am Beispiel einiger gebräuchlicher Verteilungen erfolgt eine Charakterisierung, unter welchen Parameterkonstellationen und Stichprobenumfängen sich jedes einzelne Verfahren bewährt und schließlich die Entwicklung eines Kombinationsverfahrens, das über eine Schwellenwertbedingung für den Rundungsgrad die einzelnen Verfahren miteinander verknüpft. Nach einer vorgeschalteten Einschätzung des Rundungsgrades selektiert der Schwellenwert die geeignete Verfahrensalternative. Anhand umfangreicher Simulationsstudien kann nachgewiesen werden, dass das entwickelte Kombinationsverfahren nicht nur die jeweiligen Ressourcen ausschöpft, sondern sie zu einer besseren Gesamtleistung verbindet.

Kapitel 6 ergänzt das Kombinationsverfahren für Lageparameter durch einige Betrachtungen zu Skalenparametern und Datentransformationen. Diese erweisen sich im abschließenden Datenbeispiel aus der Hydrologie in Kapitel 7 als äußerst nützlich. Starkniederschläge im August 2002, die zu extremen Durchflussmengen im gesamten Flusssystem der Elbe führten, zwangen den Freistaat Sachsen zu einer Überarbeitung seiner Bemessungsgrundlagen für Talsperren. Die Erhebung von Durchflussmengen ist unvermeidlich mit einer teils erheblichen Messungenauigkeit verbunden und eignet sich gut zur Erprobung des entwickelten Verfahrens. Eine Revision des Nebenflusspegels in Rehefeld an der Wilden Weißeritz illustriert die Systematik und verdeutlicht deren Aktualität und Relevanz.

2 Gerundete Daten aus stetigen Verteilungen

2.1 Einführung

Grundlage eines jeden statistischen Verfahrens ist ein Modell, das natürliche Gesetzmäßigkeiten und zufällige Einflüsse, die eine bestimmte Zielgröße dirigieren, auf möglichst einfache Weise und dennoch realitätsgetreu in funktionaler Form abbildet. Aussagen über die Güteeigenschaften der Verfahren sind dabei stets an die zumindest näherungsweise Gültigkeit der im Modell verankerten Annahmen gebunden. So ist bei jeder Anwendung einer statistischen Methode aufs Neue zu verifizieren, ob die Charakteristika der Daten die Anpassung des zugrunde gelegten speziellen Modells erlauben.

Eine typische Annahme fordert, dass die Realisationen einer Zufallsvariablen jeden beliebigen Wert eines vorgegebenen Intervalls annehmen können. Außerdem sollte der Zufallsfehler weder systematisch noch von der Größenordnung des Messfehlers abhängig sein. Dem theoretischen Anspruch stehen Probleme der technischen Realisierbarkeit oder Finanzrestriktionen gegenüber, so dass man sich in der Praxis mit mehr oder weniger unpräzisen, fehlerbehafteten Werten zufrieden geben muss. Dabei werden meist die Grenzen der Messgenauigkeit ohne ausreichende Kenntnis der Konsequenzen ignoriert.

Die Präzision eines Messwertes setzt sich aus einer messtechnischen sowie einer ablesebedingten Komponente zusammen. Beide Komponenten können systematische Fehler wie zufällige Irrtümer beinhalten. Von der Präzision des Messwertes ist die Präzision des Messgerätes zu unterscheiden. Ein gutes Messgerät sollte nach der Kalibration Messwerte am unveränderten Untersuchungsobjekt möglichst oft reproduzieren können. Abweichungen sind dann technisch bedingt zufällig. Die Beschränkung des Sichtfensters an Digitalgeräten oder die Markierungsabstände einer Skaleneinteilung bringen eine systematische Fehlerkomponente ein. Feinere Unterteilungen werden nicht mehr angezeigt, sondern auf den nächstliegenden erfassbaren Wert gerundet. Zufällige Ablesefehler können beispielsweise bei dicht liegenden, schwierig zu unterscheidenden Markierungen auftreten. Häufiger sind planmäßige Rundungen auf eine vorgegebene Stellenzahl oder Markierungseinheit. Obwohl die Geräte eine höhere Präzision anbieten, werden nur wenige Nachkommastellen erfasst, die als „für praktische Zwecke ausreichend“ angesehen werden.

Neben den angesprochenen messtechnischen Abläufen und Ablesevorgängen können gerundete Daten weiterhin im Zuge von arithmetischen Berechnungen, wie etwa des Umrechnens von Brüchen in Dezimalzahlen oder beim Umgang mit irrationalen Zahlen, so etwa π , entstehen.

Um die Präzision der vorliegenden Daten realistisch einschätzen zu können, sollten einige Grundregeln bei der Handhabung gerundeter Daten beachtet werden (Eisenhardt, 1947):

- a) Die ursprüngliche Skalierung der Daten sollte beibehalten werden. Ein Beispiel: Angenommen, Längenmessungen seien in der Maßeinheit *inch* notiert und dabei ganzzahlig gerundet. Anschließend werden die Messungen in *cm* ausgedrückt und besitzen nun, da $1 \text{ inch} = 2.54 \text{ cm}$, folglich 2 Nachkommastellen. Ohne einen expliziten Hinweis auf die Veränderung der Maßeinheiten wird nun ein Leser glauben, die Zahlen seien bis auf $\pm 0.005 \text{ cm}$ exakt, de facto sind es aber nur $\pm \frac{1}{2} \text{ inch} = 1.27 \text{ cm}$.
- b) Stammen die Messungen aus verschiedenen Quellen mit unterschiedlicher Präzision, so muss dieser Umstand bei der Vereinheitlichung berücksichtigt werden. Ein Auffüllen der auf ein gröberes Raster gerundeten Daten mit adäquater Anzahl von Nullen verleitet zu falschen Schlüssen auf deren Präzision. Will man einen Informationsverlust durch „Abschneiden“ der Zahlen mit höherer Messgenauigkeit umgehen, könnte man etwa als Alternative einen der unterstellten Verteilung angepassten Zufallsfehler erzeugen und diesen den gröbereren Daten hinzufügen.
- c) Sukzessives Runden in mehreren Etappen sollte vermieden werden, da die mehrmalige Anwendung der allgemein üblichen Rundungsregeln auf dieselbe Zahl nicht notwendigerweise zu dem selben Ergebnis führt wie sofortiges einmaliges Runden. Ein Beispiel:

0.445 auf 2 Dezimalstellen gerundet ergibt 0.45.

0.45 auf 1 Dezimalstelle gerundet ergibt 0.5.

0.5 ganzzahlig gerundet ergibt 1.

Aber: 0.445 einmalig auf ganze Zahl gerundet ergibt jedoch 0.

Es existieren Gegenvorschläge für abgeänderte Rundungsregeln (Caley, 1874), die in der Lage sind, diese Fehler durch sukzessives Runden abzufangen. Da sie sich aber in der Praxis nicht durchgesetzt haben, soll dieses Thema nicht weiter vertieft werden.

Im Folgenden wird davon ausgegangen, dass beim Runden mit Bedacht vorgegangen wurde und somit die oben aufgeführten Fehlerquellen weitestgehend ausgeschlossen werden können. Dennoch bleibt das Runden ein systematischer Prozess, der in diskreten Daten mündet und dadurch grundlegende Modellannahmen verletzt.

In der einschlägigen Literatur findet man gerundete Daten oft als Spezialfälle in statistischen Verfahren. So sind sie beispielsweise als unvollständige Daten interpretierbar – ein typisches Phänomen in der Lebenszeitanalyse. Unvollständige Daten

entstehen, wenn aus irgendeinem Grund relevante Informationen nicht zugänglich sind, übersehen werden oder verloren gehen. Aus formaler Sicht verschleiert infolgedessen eine Funktion $g(x)$ die wahre Beobachtung x . Unter den Oberbegriff „unvollständige“ Daten fallen zensierte, gruppierte und auch gerundete Werte. Bei solchen Daten bleibt der tatsächliche Wert durch die bloße Erfassung einer Mengenzugehörigkeit verborgen.

Genau besehen ist das Runden ein spezieller Prozess des Gruppierens von Daten, bei dem die mögliche Werteskala in meist gleich lange Intervalle mit äquidistanten Intervallmitten eingeteilt wird. Statt der nicht beobachtbaren, exakten Ursprungswerte der Zufallsvariablen X werden nur die Intervallmitten X_R registriert. Die Zufallsvariable X_R besitzt einen abzählbar unendlichen Wertebereich von äquidistanten Rasterpunkten. Können die Beobachtungswerte als Realisationen der Zufallsvariablen X mit der stetigen Dichtefunktion f beschrieben werden, so tritt ein gerundeter Wert x_R als beobachtbare Intervallmitte mit der Wahrscheinlichkeit

$$\begin{aligned} P(X_R = x_R) &= \int_{x_R - \frac{w}{2}}^{x_R + \frac{w}{2}} f(x) dx \\ &= F\left(x_R + \frac{w}{2}\right) - F\left(x_R - \frac{w}{2}\right) \end{aligned} \quad (2.1)$$

auf. Dabei bezeichnet der Parameter w die Intervallbreite. Er gibt an, auf welche Nachkommastelle gerundet worden ist. Der Wert $w = 1$ entspricht dem Runden auf ganze Zahlen mit den üblichen Regeln, nach denen Abweichungen bis einschließlich 0.5 nach unten aufgerundet, Abweichungen bis ausschließlich 0.5 nach oben abgerundet werden. Dementsprechend bedeutet $w = 0.1$ das Runden auf die erste Nachkommastelle. Die Sequenz der Intervallmitten bezeichnet man auch als Rundungsgitter.

Intuitiv scheint einleuchtend, dass die Intervallbreite w des Rundungsgitters erst zu einer aussagefähigen Größe über die „Schwere“ der Rundung wird, wenn sie im Verhältnis gesehen wird zur Streuung der Daten. Je größer die Skalierung, in der die Daten erfasst sind, also je größer die Intervallbreite, desto stärker fällt eine geringe Streuung der Daten ins Gewicht. Im Extremfall umfasst ein einziges Intervall, beispielsweise $[0.5 ; 1.5[$ das gesamte Datenspektrum, so dass sämtliche Daten auf denselben Wert gerundet werden. Daher benutzt man nach einem Vorschlag von Kendall (1938) das Verhältnis $r := w/\sigma$ zur Angabe für den Rundungsgrad.

Bis heute sind Statistiker darum bemüht, die Tragweite der Verwendung von gerundeten Daten einzuschätzen (vgl. Krämer und Runde, 1996; Durrans, 2001; O'Reilly et. al., 2003). Neben dieser diagnostischen Auseinandersetzung mit gerundeten oder

generell gruppierten Daten ist ein weiteres wichtiges Anliegen die Suche nach Möglichkeiten, über die informationsreduzierten Daten die wichtigsten Kenngrößen der Ursprungsdaten vernünftig zu beschreiben. Dabei sollten auch Richtlinien entwickelt werden, wie der Informationsverlust über eine Regulierung von Rundungsgrad und Stichprobenumfang in Grenzen gehalten werden kann, sofern diese Größen adjustiert werden können und nicht durch technische Gegebenheiten fixiert sind.

2.2 Einflüsse des Rundens auf die Momente

Die frühe Literatur befasst sich in erster Linie mit den Momenten der Verteilungen aus Ursprungsdaten und gruppierten oder gerundeten Daten. Sheppard (1898) kann mit Hilfe der Euler-Maclaurin-Entwicklung die Abweichungen der Momente quantifizieren. Diese Abweichungen können als Korrekturfaktoren verwendet werden, um auf die Momente der Ursprungsverteilung zu schließen.

Bezeichnet M_k das k -te Moment der Ursprungsverteilung

$$\begin{aligned} M_k &= E(X^k) \\ &= \int_{-\infty}^{\infty} x^k f(x) dx, \quad k \in \mathbb{N} \end{aligned} \quad (2.2)$$

und M_{Rk} das k -te Moment der Verteilung der gerundeten Daten

$$\begin{aligned} M_{Rk} &= E(X_R^k) \\ &= \sum_{i=1}^n (x_{Ri})^k \left[F\left(x_{Ri} + \frac{w}{2}\right) - F\left(x_{Ri} - \frac{w}{2}\right) \right], \end{aligned} \quad (2.3)$$

so gelten folgende Approximationen:

$$M_{R1} \approx M_1 = E(X) \quad (2.4)$$

$$M_{R2} \approx M_2 + \frac{w^2}{12} \quad (2.5)$$

$$M_{R3} \approx M_3 + M_1 \quad (2.6)$$

$$M_{R4} \approx M_4 + \left(\frac{w^2}{2}\right) M_2 + \frac{7}{240} w^4. \quad (2.7)$$

Das Stichprobenmittel ist also auch bei gerundeten Daten ein nahezu erwartungstreuer Schätzer für den Erwartungswert der Grundgesamtheit. Allerdings:

$$\begin{aligned} \text{Var}(X_R) &= E(X_R^2) - [E(X_R)]^2 \\ &= M_{R2} - M_{R1}^2 \\ &\approx M_2 - M_1^2 + \frac{w^2}{12} \\ &= \text{Var}(X) + \frac{w^2}{12}. \end{aligned} \quad (2.8)$$

Die Stichprobenvarianz als erwartungstreuer Schätzer von $Var(X_R)$ überschätzt zwar die Varianz der ungerundeten Daten, die Schätzung lässt sich aber durch Abzug von $w^2/12$ verbessern. Der Summand $w^2/12$ entspricht gerade der Varianz einer über dem Intervall $[-w/2, w/2]$ rechteckverteilten Zufallsvariablen. X_R lässt sich demnach als Summe zweier unabhängiger Zufallsvariablen X und U interpretieren, mit $U \sim R(-w/2, w/2)$.

Die Beziehungen (2.4)-(2.7) sind nicht an spezielle Verteilungsannahmen gebunden. Die einzigen Bedingungen sind: Die Dichtefunktion der Ursprungswerte muss stetig sein und sich zu beiden Enden des Definitionsbereiches streng monoton der Null nähern. Außerdem muss der Restterm der für die Herleitung benutzten Euler-Maclaurin-Entwicklung vernachlässigbar klein sein. Aufgrund dieser Forderungen wird die Anwendbarkeit obiger Formeln stark eingeschränkt, Dichtekurven mit abrupten Enden, wie die Rechteckverteilung, sowie U- oder J-förmige Kurven werden ausgeschlossen.

In der Folgezeit werden die Näherungsformeln (2.4)-(2.7) mit korrektiven Zusatztermen ergänzt für Fälle, in denen das Restglied nicht vernachlässigbar ist, aber aus den Daten geschätzt werden kann. Pairman und Pearson (1919), sowie Pearse (1928) finden Korrekturen für Verteilungen mit verdickten und Sandon (1924) für Verteilungen mit abrupten Enden. Fisher (1922) zeigt, dass die Stichprobenmomente periodische Funktionen der Position der Verteilung im Rundungsgitter sind und leitet Sheppards Korrekturen als aperiodischen Term in einer trigonometrischen Reihe ab, und Elderton (1933) schlägt Korrekturen für J-förmige Dichten analog den Korrekturen für die Exponentialverteilung vor.

Abernethy (1933) schließlich kann auf eine Reihenentwicklung und damit verbundene Forderungen gänzlich verzichten und Sheppards Korrekturen über eine Durchschnittsbildung verallgemeinern, indem er das Zentrum der Intervalle nicht mehr als konstant ansieht, sondern zufällig innerhalb des Wertebereiches variieren lässt (zufälliges Rundungsgitter). Bei einer unendlichen Versuchsreihe mit zufällig gewählten Intervallbereichen gleicher Länge bleiben Sheppards Aussagen im Mittel gültig, unabhängig von der Natur der Verteilungsfunktion.

Eine weit verbreitete Situation, in der die Annahmen für eine asymptotische Reihenentwicklung verletzt sind, stellen schiefe Dichten dar, deren Wahrscheinlichkeitsmasse im negativen Wertebereich Null ist. Leider werden auch Daten aus solchen Verteilungen typischerweise auf das nächste Vielfache einer ganzen Zahl gerundet, so dass die Annahme eines zufälligen Rundungsgitters nicht haltbar ist. Außerdem gilt die Argumentation mit einem Zufallsgitter nur für um Null zentrierte Momente. Momente, die um das Stichprobenmittel zentriert sind, wie die Stichprobenvarianz, müssen zusätzliche Auflagen erfüllen, bevor Sheppards Korrekturen anwendbar sind.

Kendall (1938) unterbreitete wenig später einen eleganteren Beweis für Abernethys Aussagen und fügte eine Diskussion der Approximationsgüte im Hinblick auf Stichprobenumfang n und Rundungsgrad r an. Für eine große Bandbreite von n sind die Approximationen durchaus zufrieden stellend. Wird n allerdings zu klein ($n < 5$) oder aber auch zu groß ($n > 100$), so weichen die approximierten Momente der Ursprungsverteilung wieder stärker von den tatsächlichen ab. Gerade letztere Tatsache erstaunt, wo doch sonst eine Erhöhung des Stichprobenumfangs nahezu ein Allheilmittel zu sein scheint. Sie wird aber anhand eines Beispiels (vgl. Lee und Vardeman, 2005) leicht verständlich: Sei X eine normalverteilte Zufallsvariable $X \sim N(4.25; 0.25)$ und X_R die Zufallsvariable der ganzzahlig gerundeten Werte, dann haben die zugehörigen Verteilungsfunktionen die in Abbildung 1 gezeigte Gestalt.

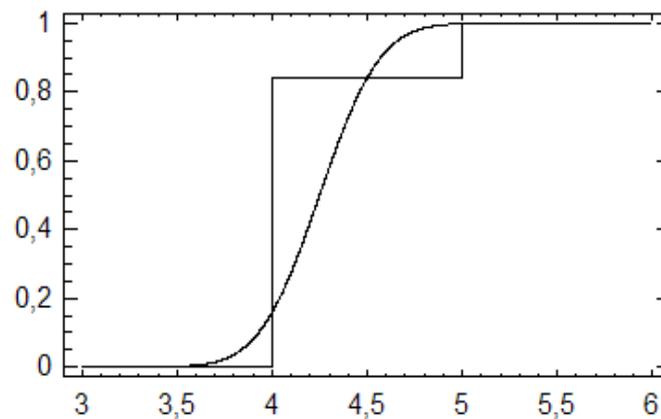


Abbildung 1: Verteilungsfunktionen: $N(4.25; 0.25)$ und gerundete Daten

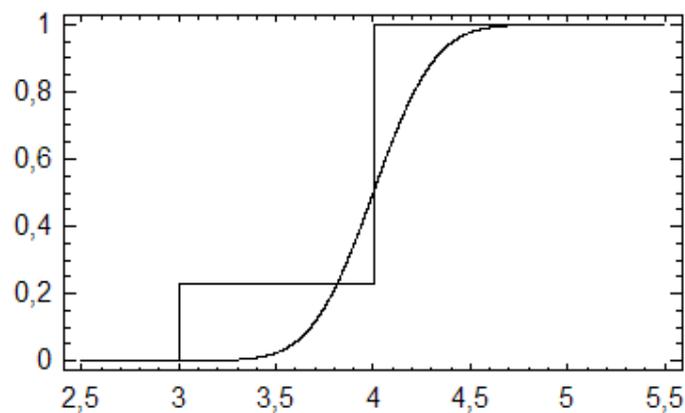


Abbildung 2: Verteilungsfunktionen: $N(4; 0.25)$ und gerundete Daten

Man beachte, dass es sich bei der Treppenfunktion nicht um eine empirische Verteilungsfunktion sondern um die exakte, in Wahrheit ebenso wie die Ursprungs-

verteilung unbekannte Verteilung der ZV X_R handelt. Der wahre Erwartungswert dieser diskreten Verteilung ist in diesem Beispiel 4.1573, und nichts anderes als dieser Wert kann mit Hilfe einer Stichprobe geschätzt werden. Während bei moderaten Stichprobenumfängen noch mit höherer Wahrscheinlichkeit Stichprobenrealisationen entstehen, die gewissermaßen „zufällig“ den wahren Erwartungswert der Grundgesamtheit liefern, so ist dies bei sehr großen Stichprobenumfängen nahezu unmöglich. Der durch die Rundung induzierte Fehler (Verzerrung) dominiert den Zufallsfehler (Varianz), vgl. Gieddebaeck (1957) und Ausführungen auf S. 14. Er wird ferner mitbestimmt von der relativen Position von $E(X)$ und Rundungsgitter (siehe auch Holland, 1975 und Preece, 1981). Bei $X \sim N(4; 0.25)$ und beibehaltenem Rundungsgitter wie oben hätte X_R den gleichen Erwartungswert wie X .

Der systematischen Verzerrung kann nur Abhilfe geschaffen werden, indem die Intervalle möglichst klein gehalten werden. Kendalls Empfehlung lautet:

$$r := w/\sigma < 1.6.$$

An diese allgemein gehaltenen Aussagen schließen sich in den folgenden Abschnitten Untersuchungen unter speziellen Verteilungsannahmen an.

2.3 Effekte bei Normalverteilung der Grundgesamtheit

In den nächsten beiden Unterkapiteln soll die Frage geklärt werden, wie anfällig die populärsten Schätzer und Teststatistiken auf Rundungsfehler reagieren, wenn statt der exakten Daten gerundete verwendet werden. Weit verbreitete Statistiken sind im Falle der Normalverteilung $N(\mu, \sigma)$ die Schätzer

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{als Schätzer für} \quad (2.9)$$

den Erwartungswert μ ,

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{als Schätzer für die Standard-} \quad (2.10)$$

abweichung σ ,

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \quad \text{als Teststatistik für Hypothesen} \quad (2.11)$$

bzgl. des Erwartungswertes und zur
Konstruktion von Konfidenzinter-
vallen für μ .

Was geschieht, wenn man diese durch

$$\bar{X}_R = \frac{1}{n} \sum_{i=1}^n X_{Ri}, \quad (2.12)$$

$$S_R = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_{Ri} - \bar{X}_R)^2} \quad (2.13)$$

$$T_R = \frac{\sqrt{n}(\bar{X}_R - \mu)}{S_R} \quad (2.14)$$

ersetzt?

Zu einer ersten Einschätzung erweisen sich Sheppards Korrekturen als nützlich: Bei n unabhängig, identisch normalverteilten Zufallsvariablen $X_i \sim N(\mu; \sigma^2)$ ist der Mittelwert \bar{X} ein erwartungstreuer Schätzer für μ mit Varianz σ^2/n , und laut Sheppard (1898) gilt:

$$E(X_R) = M_{R1} \approx M_1 = \mu. \quad (2.15)$$

Somit ist \bar{X}_R als nahezu erwartungstreuer Schätzer für $E(X_R)$ ein brauchbarer Schätzer für μ . Brauchbar, falls die Voraussetzungen für eine gute Approximation, wie im vorhergehenden Abschnitt besprochen, tatsächlich vorliegen. Die Varianz von \bar{X}_R ist gegenüber der Varianz von \bar{X} leicht erhöht auf

$$Var(\bar{X}_R) = \frac{\sigma^2}{n} \left(1 + \frac{w^2}{12}\right) \quad (2.16)$$

(vgl. Eisenhart, 1947). Der Schätzer S_R^2 ist zwar erwartungstreu für $Var(X_R)$, tendiert aber dazu, die Varianz der ungerundeten Daten zu überschätzen und muss um den Term $w^2/12$ korrigiert werden,

$$E(S_R^2) - \frac{w^2}{12} \approx \sigma^2. \quad (2.17)$$

Trotz der Erwartungstreue von \bar{X}_R und der Korrekturmöglichkeit für S_R^2 kann es im konkreten Fall zu substantiellen Differenzen zwischen geschätzten und tatsächlichen Parametern kommen. Das Runden kann bei hohem Rundungsgrad und unvorteilhaften Konstellationen von Verteilungsparametern und Rundungsgitter diese Abweichungen verstärken oder gehäuft Stichproben erzeugen, die zu einer Fehleinschätzung führen. Umso dringlicher stellt sich die Frage nach einem geeigneten Konfidenzintervall. Ein solches Konfidenzintervall für den Erwartungswert der Normalverteilung basiert für exakte Daten auf der Statistik T aus (2.10). Die Herleitung ihrer Verteilung unterstellt eine Unabhängigkeit von \bar{X} und S , die aber bei gerundeten Daten nicht gegeben ist (vgl. das Beispiel in Student, 1908). Trotzdem stimmen die Momente von T_R und T für größere Stichprobenumfänge überein (Geary, 1936) und T

kann durch T_R approximiert werden, falls der Rundungsgrad nicht zu groß ist ($n \geq 5$ bei $r < 0.5$). Bei kleineren Stichprobenumfängen und hohen Rundungsgraden häufen sich jedoch die Stichproben, bei denen sämtliche Werte in das gleiche Intervall fallen. Das ergäbe eine Stichprobenvarianz von Null und damit einen unendlich hohen t_R -Wert. Eine Korrektur nach Sheppard verbietet sich, denn sie führt auf die absurde Situation einer negativen Varianz. Mit $w = \sigma$, $\mu = 0$, und $n = 3$ liegt die Wahrscheinlichkeit, eine solche Stichprobe zu erhalten bei 8.5%. Glücklicherweise sinkt diese Wahrscheinlichkeit rasch mit wachsendem n . Auf der anderen Seite werden sich mit wachsendem n die Konfidenzintervalle stark verkleinern, bis schließlich bei Abweichungen von $E(X)$ und $E(X_R)$ der ursprüngliche Erwartungswert nur noch bei seltenen Stichprobenrealisierungen überdeckt wird, was gleichbedeutend ist mit einer Nichteinhaltung des vorgegebenen Signifikanzniveaus. Leider sind weder Größenordnung noch Richtung der Niveauabweichung theoretisch zahlenmäßig erfassbar und vage Empfehlungen zu Rundungsgrad und Stichprobenumfang lediglich ein Versuch, die Niveaueinhaltung abzusichern. Weiterhelfen können da nur Simulationsstudien, die zu verschiedenen Szenarien die tatsächliche Überdeckungswahrscheinlichkeit berechnen.

2.4 Effekte bei Exponentialverteilung der Grundgesamtheit

Die zentrierten Momente der Exponentialverteilung mit Dichtefunktion

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \quad , \quad x \geq 0, \quad \lambda > 0 \quad (2.18)$$

sind ähnlich wie die Momente der Normalverteilung über die Parameter der Verteilung charakterisiert, denn $M_1 = E(X) = \lambda$ und $M_2 = Var(X) = \lambda^2$. Tricker (1984) hat sich in umfangreichen Analysen der Fragestellung gewidmet, ob auch bei Exponentialverteilung Sheppards Formeln, die für J-förmige Dichten nur im Mittel über alle zufälligen Rundungsgitter gelten, ein probates Hilfsmittel zur Annäherung an den Parameter der Grundgesamtheit darstellen. Dazu berechnet er die durch das Runden hervorgerufene Verzerrung der Momente in Abhängigkeit von der Position des Rundungsgitters und des Rundungsgrades: Bei einem zufällig angeordneten Rundungsgitter seien nach Rundung die Realisierungen

$$c, c + w, c + 2w, c + 3w, \dots, c + mw, c + (m + 1)w, \dots, \quad (2.19)$$

möglich. Damit jedes der Gitter aus gleich langen Intervallen besteht, sind als linke Intervallgrenze des ersten Intervalls, das die Null enthält, auch negative Werte zugelassen. Bezeichne c die Mitte dieses Intervalls und sei aus einer Gleichverteilung über $[-\frac{w}{2}, +\frac{w}{2}]$ gewählt, dann ist die Dichte der gerundeten exponentialverteilten

Werte

$$P(X_R = mw + c) = \begin{cases} \int_0^{c+w/2} f(x)dx, & m = 0 \\ \int_{mw+c-w/2}^{mw+c+w/2} f(x)dx, & m \in \mathbb{N} \end{cases} \quad (2.20)$$

$$\Leftrightarrow P(X_R = mw + c) = \begin{cases} 1 - e^{-\frac{1}{\lambda}(c+\frac{w}{2})}, & m = 0 \\ (e^{w/2\lambda} - e^{-w/2\lambda}) e^{-\frac{1}{\lambda}(mw+c)}, & m \in \mathbb{N} \end{cases} \quad (2.21)$$

und in Abhängigkeit vom Rundungsgrad $r = w/\lambda$ gesehen

$$\Leftrightarrow P(X_R = mw + c) = \begin{cases} 1 - e^{-\frac{c}{\lambda} - \frac{r}{2}}, & m = 0 \\ (e^{r/2} - e^{-r/2}) e^{-rm - \frac{c}{\lambda}}, & m \in \mathbb{N}. \end{cases} \quad (2.22)$$

Erwartungswert und Varianz der Dichtefunktion (2.22) bei gerundeten Daten sind gegeben durch

$$E(X_R) = c + \lambda \frac{r e^{c/\lambda}}{e^{r/2} - e^{-r/2}} \quad (2.23)$$

$$Var(X_R) = \lambda^2 \left\{ r^2 e^{-r(\frac{3}{2} + \frac{c}{w})} \frac{\left\{ 1 + e^r - e^{r(\frac{1}{2} - \frac{c}{w})} \right\}}{(1 - e^{-r})^2} \right\}. \quad (2.24)$$

Mit diesen konkreten Angaben kann eine prozentuale Verzerrung im Verhältnis zum wahren Moment tabelliert werden. Für vorgegebene Rundungsgrade lässt sich diese von c abhängige Funktion als Serie von Kurven skizzieren. Die grafische Darstellung (in Tricker, 1984) bietet zumindest überblicksartig eine Einschätzung der Konsequenzen und einige Sachverhalte lassen sich ablesen:

1. Die Verzerrungen, den Erwartungswert betreffend, scheinen sich über alle c -Werte des Definitionsbereiches auszugleichen und somit die Sheppard-Approximation zu untermauern. Ist $r \leq 1/4$, kann angesichts des sehr flachen Kurvenverlaufs der Einfluss durch Wahl des Intervallmittelpunktes als vernachlässigbar gering eingestuft werden. Bei höherem r gewinnt c zunehmend an Bedeutung und das Runden wird mit einer teils erheblichen Erhöhung oder Verminderung des Erwartungswertes quittiert. Da im Regelfall beim Runden eine freie Wahl von c entfällt und $c = 0$, ist ein besonderes Augenmerk auf diesen Fall zu richten. Gerade an dieser Stelle weisen alle Kurven die Spitzenwerte der Reduktion aus.

2. Die Varianzkurven dokumentieren eine Tendenz, die Varianz weitaus häufiger massiv anzuheben als zu senken – auch hierin wird Sheppard bestätigt. Parallelen zeigen zwischen den Verzerrungskurven von Erwartungswert und Varianz in einer Stagnation auf vernachlässigbarem Niveau für kleinere Rundungsgrade, $r \leq 1/4$. Bei größeren Rundungsgraden müsste man einer Varianzerhöhung durch eine Wahl von $c \approx \pm w$ gegensteuern. Im Regelfall $c = 0$ muss man Varianzsteigerungen von bis zu 16% bei $r = 1$ einkalkulieren.

\bar{X}_R ist zwar unverzerrter Schätzer für den Erwartungswert der gerundeten Daten, aber nur in Ausnahmesituationen unverzerrt für den wahren Parameter. Da $E(X_R)$ zudem vom Stichprobenumfang unabhängig ist (vgl. (2.23)), lässt sich die Verzerrung $E(\bar{X}_R) - \lambda$ nicht durch eine Erhöhung des Stichprobenumfangs kompensieren. Eine Erhöhung des Stichprobenumfangs hat im Gegenteil den unerwünschten Nebeneffekt, die Varianz $Var(\bar{X}_R) = Var(X_{Ri})/n^2$ so stark herabzusetzen, dass man kaum eine Chance hat, wenigstens durch eine ungewöhnliche Stichprobe sozusagen „versehentlich“ $E(X)$ zu treffen. Um diese Problematik zu umgehen, sollte im Vorfeld ein nicht zu hoher Rundungsgrad eingeplant werden. Die Abschätzung $r \leq 0.1$ gewährleistet eine Verzerrung, die nicht mehr als 20% der Intervalllänge ausmacht, und bereits ab $r = 0.25$ sind $Var(\bar{X}_R)$ und $Var(\bar{X})$ nahezu gleich.

Zur Orientierung, in welchen Situationen \bar{X}_R in ausreichender Qualität den Parameter schätzt, untersucht Tricker (1984) beispielhaft für einige kleine Stichprobenumfänge den mittleren quadratischen Fehler MQF,

$$\begin{aligned} MQF &= E(\bar{X}_R - \lambda)^2 \\ &= Var(\bar{X}_R) + [E(\bar{X}_R - \lambda)]^2. \end{aligned} \quad (2.25)$$

Auch der mittlere quadratische Fehler wird maßgeblich über r und c gesteuert. Da sich der Beitrag der Varianz mit zunehmendem Stichprobenumfang rasch reduziert, bestimmt für $n \geq 15$ im wesentlichen die Verzerrung den Betrag des MQF: Eine Erhöhung von r schlägt mit einer Erhöhung des MQF negativ zu Buche, die nur über eine Adjustierung von c – soweit in der Realität praktikabel – wieder auszugleichen ist, dabei wäre $c = 0.41w$ ideal. Bei einer Reduktion von r unter 0.1 ist der schwer handhabbare Einfluss von c annähernd beseitigt und die Schätzer \bar{X}_R und \bar{X} sind von vergleichbarer Güte. Tricker (1984) beendet seine Ausführungen mit einer Erörterung der Verteilung der Stichprobensumme $\sum_{i=1}^n X_{Ri}$, die die Grundlage für Tests bezüglich des Verteilungsparameters oder Konfidenzintervalle bildet. Sei

$$S_n := \sum_{i=1}^n X_i$$

die Summe von n unabhängigen und identisch exponentialverteilten Zufallsvariablen, dann folgt $2S_n/\lambda$ unter Verwendung der ungerundeten Daten einer χ^2 -Verteilung mit $2n$ Freiheitsgraden. Ein Nachweis erfolgt später über die charakteristische Funktion der Exponentialverteilung, siehe Gleichung (4.3). Daher ist

$$\left[\frac{2 S_n}{\chi_{(2n, 1-\alpha/2)}^2}; \frac{2 S_n}{\chi_{(2n, \alpha/2)}^2} \right] \quad (2.26)$$

ein geeignetes zweiseitiges Konfidenzintervall für λ . Die Dichte der Stichprobensumme der gerundeten Daten $S_{Rn} = \sum_{i=1}^n X_{Ri}$ kann geschrieben werden als

$$P(S_{Rn} = mw + cn) = \begin{cases} (k')^n, & m = 0 \\ e^{-rm} \sum_{j=1}^n A_{jn} (k')^{n-j} k^j e^{-jr \frac{c}{w}} \prod_{l=1}^{j-1} (m-l), & m \in \mathbb{N}, \end{cases} \quad (2.27)$$

wobei $k' = 1 - e^{-r(\frac{1}{2} + \frac{c}{w})}$, $k = e^{r/2} - e^{-r/2}$, $A_{jn} = \binom{n}{j} / (j-1)!$.

Tricker (1984) diskutiert die Güte der Anpassung der Verteilung von $\sum_{i=1}^n X_{Ri}$ an die χ^2 -Verteilung und zeigt schließlich auf, wie sich das Signifikanzniveau verändert, wenn man in naiver Herangehensweise die χ^2 -Quantile stellvertretend für die Quantile der komplexeren Verteilung von $\sum_{i=1}^n X_{Ri}$ benutzt. Eine Gegenüberstellung der Wahrscheinlichkeiten

$$\begin{aligned} P(S_n \leq \chi_\alpha^2) &=: \alpha_1 \\ \text{und } P(S_{Rn} \leq \chi_\alpha^2) &=: \alpha_2 \end{aligned}$$

lässt erkennen: Selbst bei einem Stichprobenumfang von $n = 25$ fallen die Konfidenzintervalle häufig zu klein aus, was eine deutliche Überschreitung des Signifikanzniveaus nach sich zieht (bei $n = 5$, $r = 2$ liegt α_2 mit Werten bis zu 0.32 weit über dem Sollwert $\alpha_1 = 0.05$). Je massiver der Rundungsgrad, desto höher liegt α_2 . Erst für $r \leq 0.1$ kann eine Einhaltung des Niveaus auch bei kleineren Stichprobenumfängen (ab $n = 5$) garantiert werden. Will man zur Wahrung einer gewissen Qualitätsstufe den Rundungsgrad regulieren und auf den Richtwert einstellen, erfordert das eine vernünftige Einschätzung der Standardabweichung. Bis dato fußt die Schätzung der Varianz ebenfalls auf Sheppards Korrekturen.

2.5 ML-Schätzung bei gerundeten Daten

Die Beschreibung statistischer Eigenschaften der Momentenschätzer aus gerundeten Daten stößt schnell an Grenzen. Exakte funktionale Darstellungen wie in (2.23) und

(2.24) sind nur in Ausnahmefällen bei simpleren Dichteformen, z.B. der Exponentialverteilung, herzuleiten. Und selbst dann vermitteln arbeitsintensive Studien lediglich eine vage Vorstellung ihrer durchschnittlichen Güte, die im konkreten Einzelfall nur wenig Anhaltspunkte für Verbesserungsmaßnahmen bietet. So scheint es geboten, sich nach anderen Schätzmethoden umzusehen, deren statistische Charakteristika besser greifbar sind.

Die Parameterschätzung rückt von einer bloßen Betrachtung der Momente ab und versucht unter der Voraussetzung eines grundsätzlich bekannten Verteilungstyps die Gestalt der Verteilung durch bestmögliche Festlegung der Parameter zu konkretisieren. Ein populäres Verfahren zur Parameterschätzung ist die *Maximum-Likelihood-Methode*.

Betrachtet wird ein parametrisches Modell unabhängiger und identisch verteilter Zufallsvariablen X_i , $i = 1, \dots, n$ mit Dichtefunktion $f(\cdot; \theta)$. O.B.d.A. sei $\theta \in \Theta \subset \mathbb{R}$, wobei entsprechende Aussagen für $\theta \in \mathbb{R}^k$ z.B. Gourieroux und Monfort (1995, S. 175 ff.) zu entnehmen sind. Sobald eine Stichprobe von Beobachtungen $x = x_1, \dots, x_n$ vorliegt, betrachtet man als einen Schätzer $\hat{\theta}$ für θ den Wert, der die gemeinsame Dichte der Zufallsvariablen (genannt Likelihoodfunktion)

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta), \quad \theta \in \Theta \subset \mathbb{R}, \quad (2.28)$$

maximiert, d.h.

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta; x). \quad (2.29)$$

Falls Θ kompakt und die Likelihoodfunktion stetig in θ ist, existiert der Maximum-Likelihood Schätzer. Die Eigenschaften von ML-Schätzern basieren auf der Gültigkeit folgender Regularitätsbedingungen:

(R1) Der Parameterraum Θ ist eine offene Teilmenge von \mathbb{R} ,

(R2) der Träger der Verteilung $\mathcal{X} := \{x : f(x; \theta) > 0\}$ ist unabhängig von θ ,

(R3) $(d \ln f(x; \theta)/d\theta)$ existiert und ist endlich für $\theta \in \Theta$ und $x \in \mathcal{X}$,

(R4) für eine Funktion $h(x)$ sind Differentiation und Integration vertauschbar, d.h.

$$\frac{d}{d\theta} \left(\int \cdots \int h(x) f(x; \theta) dx_1 \cdots dx_n \right) = \int \cdots \int h(x) \left[\frac{d}{d\theta} f(x; \theta) \right] dx_1 \cdots dx_n < \infty.$$

Bedingung (R1) schließt Randpunkte aus und sichert somit die Existenz erster Ableitungen (von beiden Seiten eines Punktes). Bei Gültigkeit der Bedingungen (R1)–(R4) existiert die Fisher-Information \mathcal{I}_n der Stichprobe und es lässt sich zeigen

$$\mathcal{I}_n := E \left[\left(\frac{d \ln f(x; \theta)}{d\theta} \right)^2 \right] = -E \left(\frac{d^2 \ln f(x; \theta)}{d\theta^2} \right).$$

Um pathologische Fälle auszuschließen wird üblicherweise (vgl. Spanos, 1999, S. 662ff.) gefordert:

Stetigkeit:

(R5) $L(\cdot; x) : \Theta \rightarrow [0, \infty)$ ist stetig für alle $\theta \in \Theta$.

Identifizierbarkeit:

(R6) Für alle $\theta_1 \neq \theta_2 \in \Theta$ gilt $f(x; \theta_1) \neq f(x; \theta_2)$.

ML-Schätzer sind im Allgemeinen verzerrt, besitzen aber zwei oftmals nützliche Eigenschaften:

Invarianz:

Sei $\tau(\cdot)$ eine (messbare) Funktion und $\hat{\theta}$ MLS für θ . Dann ist $\tau(\hat{\theta})$ MLS für $\tau(\theta)$.

Pivot-Größen-Eigenschaft:

Sei X Zufallsvariable mit Dichtefunktion $f(\cdot)$. Für Familien von Dichten, die einen Lageparameter α und einen Skalenparameter β besitzen, d.h. deren Dichte mit Hilfe einer Funktion $g(\cdot)$ in der Form

$$f(x) = \frac{1}{\beta} g\left(\frac{x - \alpha}{\beta}, \gamma\right), \quad (\alpha, \beta, \gamma) \in \Theta, \beta > 0 \quad (2.30)$$

darstellbar ist, gilt: Sind $\hat{\alpha}, \hat{\beta}$ und $\hat{\gamma}$ die MLS für α, β und γ , dann sind die Verteilungen von $\hat{\gamma}, \frac{\hat{\beta}}{\beta}, \frac{\hat{\alpha} - \alpha}{\beta}$ und $\frac{\hat{\alpha} - \alpha}{\hat{\beta}}$ unabhängig von α und β und damit Pivot-Größen.

Zur Herleitung asymptotischer Eigenschaften von ML-Schätzern sind in Ergänzung von (R1)–(R6) weitere Regularitätsbedingungen erforderlich.

(R7) $E(\ln f(X; \theta))$ existiert.

(R8) $\frac{1}{n} \ln L(\theta; x) \xrightarrow{f.s.} E(\ln f(X; \theta))$ für alle $\theta \in \Theta$.

(R9) $\ln L(\theta; x)$ ist zweifach differenzierbar in einem offenen Intervall um θ .

Auf Basis der Bedingungen (R1)–(R9) lassen sich nun die folgenden Eigenschaften nachweisen:

Schwache Konsistenz:

Für ein $\epsilon > 0$ gilt

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_{ML} - \theta| < \epsilon) = 1.$$

Starke Konsistenz:

$$P(\lim_{n \rightarrow \infty} \hat{\theta}_{ML} = \theta) = 1.$$

Asymptotische Normalität:

$$\sqrt{n}(\hat{\theta}_{ML} - \theta) \xrightarrow{n \rightarrow \infty} N(0, \mathcal{I}^{-1}(\theta)). \quad (2.31)$$

Die asymptotische Normalität ist der Grundstein, auf dem viele approximative Konfidenzintervalle aufbauen.

Besonders interessant ist, dass die Regularitätsbedingungen lediglich die Stetigkeit der Verteilungsfunktion in θ verlangen, nicht aber die Stetigkeit in x . Dies berechtigt zu der Hoffnung, dass sich die genannten Eigenschaften auf die MLS aus Verteilungen gruppierter Daten übertragen. Entsprechend der Vorgehensweise bei exakten Daten leitet sich der MLS aus der Nullstellenberechnung der logarithmierten und bezüglich θ differenzierten Likelihood $\ln L(\theta; x)$ ab. Die Herleitung einiger leicht zu berechnender Formeln für den MLS sowie Existenzbedingungen sind ein Verdienst von Kulldorff (1961). Aus den vielen Modellvarianten von Exponential- und Normalverteilung, für die Kulldorff solche Formeln bereitgestellt hat, sollen nur einige wenige exemplarisch herausgegriffen werden, auf die in späteren Teilen der Arbeit noch Bezug genommen wird.

Kulldorff (1961) unterstellt ein Rundungsgitter mit k Intervallen, die so eingerichtet sind, dass sie alle Daten der Stichprobe erfassen und jeweils erster und letzter Abschnitt besetzt sind. Mit $n_i, i = 1, \dots, k$ sei die Anzahl der registrierten Werte im i -ten Intervall bezeichnet. Die folgenden Aussagen beziehen sich, obwohl Kulldorff auch allgemeinere Gruppierungen betrachtet, ausschließlich auf den Fall gleich breiter Intervalle.

- *Exponentialverteilung ohne Verschiebungsparameter*

$$F(x) = 1 - e^{-\frac{x}{\lambda}}; \quad x \geq 0. \quad (2.32)$$

Das Abrunden sehr kleiner Werte kann unliebsame Datensituationen auslösen, indem künstlich Nullen entstehen. Wenige Nullen gefährden die Existenz der

ML-Schätzung nicht, aber bei einer Datenreihe von ausschließlich Null-Werten existiert der MLS nicht. Dies gilt nebenbei bemerkt ebenso für weitere Verteilungen wie etwa Lognormal-, Weibull- und Paretoverteilung. Für eine Reihe gleicher, aber von Null verschiedener Werte haben die Likelihoodgleichungen zwar eine Lösung, diese Lösung ist aber nicht eindeutig. Zusammenfassend lässt sich festhalten:

Der MLS für λ existiert und ist eindeutig genau dann, wenn $n_1 < n$ und $n_k < n$. Er ist gegeben durch

$$\hat{\lambda} = \left[\frac{1}{w} \ln \left(1 + \frac{n - n_k}{\sum_{i=1}^k (i-1)n_i} \right) \right]^{-1}. \quad (2.33)$$

Je größer der Stichprobenumfang, desto seltener treten die unerwünschten Störfälle auf, bei denen entweder alle Werte in das erste oder alle in das letzte Intervall fallen, und die Wahrscheinlichkeit für die Existenz des MLS konvergiert gegen 1.

- *Exponentialverteilung mit Verschiebungsparameter*

$$F(x) = 1 - e^{-(x-\mu)/\lambda}; \quad x \geq 0. \quad (2.34)$$

Die MLS für λ und μ existieren genau dann, wenn $n_1 + n_2 < n$ und $n_1 + n_k < n$. Sie sind gegeben durch

$$\hat{\lambda} = \left[\frac{1}{w} \ln \left(1 + \frac{n - n_1 - n_k}{\sum_{i=3}^k (i-2)n_i} \right) \right]^{-1} \quad (2.35)$$

und

$$\hat{\mu} = w - \hat{\lambda} \ln \frac{n}{n - n_1}. \quad (2.36)$$

- *Normalverteilung*

An dieser Verteilung zeichnen sich die ersten Nachteile der ML-Methode ab. Falls $n_1 < n$ und $n_k < n$, existiert zwar der MLS, aber das Ableiten der Loglikelihood

$$\ln L(\mu, \sigma) = \sum_{i=1}^k n_i \ln \left[\Phi \left(\frac{X_{Ri} + \frac{w}{2} - \mu}{\sigma} \right) - \Phi \left(\frac{X_{Ri} - \frac{w}{2} - \mu}{\sigma} \right) \right] \quad (2.37)$$

und die Bestimmung der Nullstellen

$$\frac{d}{d\mu} \ln L(\mu, \sigma) = \sum_{i=1}^k n_i \frac{\phi \left(\frac{X_{Ri} + \frac{w}{2} - \mu}{\sigma} \right) - \phi \left(\frac{X_{Ri} - \frac{w}{2} - \mu}{\sigma} \right)}{\Phi \left(\frac{X_{Ri} + \frac{w}{2} - \mu}{\sigma} \right) - \Phi \left(\frac{X_{Ri} - \frac{w}{2} - \mu}{\sigma} \right)} = 0 \quad (2.38)$$

mit

$\phi(x)$: Dichte der Standardnormalverteilung und

$\Phi(x)$: Verteilungsfunktion der Standardnormalverteilung

gestaltet sich kompliziert und eine Darstellung des MLS ist in geschlossener Form nicht möglich. Daher wird häufig eine numerische Maximierung der Likelihoodfunktion präferiert, die mit Unterstützung von Softwareprogrammen, aufbauend auf Iterationsverfahren (wie dem Newton-Raphson-Verfahren), bewerkstelligt werden kann. Einer der zahlreichen entwickelten Algorithmen ist zum Beispiel in Swan (1969) beschrieben. Einen guten Vergleich diverser Algorithmen hinsichtlich ihrer Konvergenz und Rechenzeiten findet man in Schader und Schmid (1984).

Mit der Anwendung numerischer Iterationsverfahren zur Schätzung erübrigt sich eine theoretische Analyse der Verteilungseigenschaften, da die Güte des Rechenverfahrens erheblich die Qualität der Schätzung beeinflusst: Die Wahl eines guten Startwertes bestimmt das Konvergenzverhalten, ein gewissenhaft ausgesuchter Parameterbereich ist wichtig, um das globale Maximum erfassen zu können und das Abbruchkriterium beeinträchtigt die eventuell verbleibende Restdistanz von der wahren ML-Schätzung. Die Gültigkeit der MLS-Eigenschaften bleibt eine Hoffnung und nur Monte-Carlo-Simulationen können einigermaßen Aufschluss geben.

3 Konfidenzintervalle bei gerundeten Daten

3.1 Das Pivot-Verfahren

Zur Konstruktion von Konfidenzintervallen gibt es verschiedene Ansätze. Eine ad-hoc-Methode, die vorab keinerlei Qualitätsansprüche berücksichtigt, ist die Pivot-Methode:

Sei X_1, \dots, X_n eine Zufallsstichprobe aus einer Verteilung mit Dichte $f(\cdot; \theta)$ und $Q = q(X_1, \dots, X_n; \theta)$ eine Funktion der Zufallsvariablen und Parameter. Wenn die Verteilung von Q nicht von θ abhängt, wird Q Pivotgröße genannt.

Eine Pivotgröße kann demnach von unbekanntem Parametern abhängen. Entscheidend ist, dass diese Parameter die Verteilungseigenschaften von Q nicht beeinflussen. Ist also die Verteilung der Pivotgröße bekannt, kann man auf deren Quantilen einen Konfidenzbereich aufbauen.

Sei Q eine Pivotgröße und q_1 so gewählt, dass $P(q_1 < Q) = 1 - \alpha$. Gibt es zu jedem Beobachtungsvektor eine eindeutige Umformung

$$q_1 < q(X; \theta) \Leftrightarrow \underline{\theta}(X, q_1) < \tau(\theta), \quad (3.1)$$

dann ist $\underline{\theta}(X)$ eine untere Konfidenzschranke für $\tau(\theta)$ zum Niveau α .

Ein klassisches Beispiel für diese Methode sind die gemeinhin benutzten Konfidenzintervalle für den Parameter μ der Normalverteilung $N(\mu, \sigma^2)$, deren Fundament die Quantile der t -Verteilung bilden. Die Größe

$$T = \sqrt{n} \frac{(\bar{X} - \mu)}{S}$$

ist bekanntermaßen t -verteilt und damit Pivotgröße. Aufgrund der Symmetrie der t -Verteilung liefert das Quantil t_α , für das gilt

$$P(t_\alpha \leq T) = 1 - \alpha,$$

über

$$P\left(\bar{X} - \frac{S \cdot t_\alpha}{\sqrt{n}} < \mu\right) = 1 - \alpha$$

die untere Konfidenzschranke

$$\bar{X} - \frac{S \cdot t_\alpha}{\sqrt{n}}.$$

Bislang bleibt ungeklärt, ob zu einem gegebenen Problem tatsächlich eine Pivotgröße existiert oder nicht.

Eine Antwort geben etwa Mood, Graybill und Boes (1974, S. 387f.) für ein $\theta \in \mathbb{R}$. Sie merken an, dass für jede Stichprobe $X = X_1, \dots, X_n$ mit einer in x stetigen Verteilungsfunktion $F(x; \theta)$

$$\prod_{i=1}^n F(X_i; \theta) \quad \text{bzw.} \quad - \sum_{i=1}^n \ln F(X_i; \theta) \quad (3.2)$$

Pivotgrößen darstellen. Welche Verteilungen auch immer zunächst unterstellt werden, $-\sum_{i=1}^n \ln F(X_i; \theta)$ ist Gamma-verteilt, gemäß $\Gamma(n, 1)$. Damit ist die Pivoteigenschaft der obigen Funktionale für beliebige stetige Verteilungen sichergestellt. Für einen Konfidenzbereich sind die Größen allerdings nur nutzbar, wenn zudem die Verteilungsfunktion für jedes x monoton in θ ist und damit auch $\prod_{i=1}^n F(x_i; \theta)$ monoton in θ für alle x_1, \dots, x_n .

3.2 Das Likelihood-Quotienten-Verfahren

Eine zweite Quelle zur Konstruktion von Konfidenzintervallen erschließt sich über den Annahmebereich eines Testverfahrens. Über diese Methodik öffnet sich ein zusätzliches Arsenal von etablierten Statistiken, die den Grundstock zur Konstruktion einer Konfidenzregion bilden. Aufgrund der Dualität zwischen Test- und Konfidenzverfahren ist es erstmals möglich, Konfidenzintervalle nach gewissen Güteeigenschaften zu strukturieren, da sich die Güteermerekmale von Testverfahren auf ihr Pendant als Konfidenzintervall übertragen. Basieren Konfidenzintervalle auf gleichmäßig mächtigsten Tests zum Niveau α , dann sind sie gleichmäßig genaueste $(1 - \alpha)$ -Konfidenzintervalle. (vgl. Lehmann und Romano, 2005, S. 73).

Eine geeignete Statistik für das Testproblem der Hypothesen

$$H_0 : \theta \in \Theta_0 \subset \Theta \quad \text{vs.} \quad H_1 : \theta \in \Theta - \Theta_0 \quad (3.3)$$

leitet sich aus dem Likelihood-Quotienten

$$LQ(x) = \frac{\sup_{\theta \in \Theta_0} L(x; \theta)}{\sup_{\theta \in \Theta} L(x; \theta)} \quad (3.4)$$

ab, der die restringierte mit der unrestringierten Likelihood vergleicht.

Das Likelihood-Quotienten-Verfahren für Konfidenzintervalle nutzt ebenfalls den Quotienten aus (3.4). Dabei beschreibt der Quotient keine konkrete Zahl mehr, sondern eine von den unbekanntem Verteilungsparametern abhängige Funktion. Aufgefasst als Funktion von Zufallsvariablen $X = X_1, \dots, X_n$, stellt die relative Likelihood

$$R(X, \theta) = \frac{L(X; \theta)}{\sup_{\theta \in \Theta} L(X; \theta)} \quad (3.5)$$

eine geeignete Statistik für eine Bereichsschätzung dar: Bedauerlicherweise kann man nur in wenigen Fällen die Verteilung der relativen Likelihood exakt bestimmen. Dennoch nimmt die Methode aus anderen guten Gründen eine zentrale Stellung in der statistischen Theorie ein. Sie verkörpert ein leistungsfähiges Universalwerkzeug, das weder bei Verteilungsannahmen noch bei Hypothesenspezifikationen allzu selektiv vorgeht. Die asymptotische Verteilung ist bekannt und die Eigenschaften gut erforscht. Das Verfahren ist zumindest näherungsweise optimal, und häufig helfen simple Korrekturen, die Approximation an die Grenzverteilung zu beschleunigen. Ihr Einsatzfeld erstreckt sich auch auf Modelle mit Störparametern. Störparameter sind solche, die zwar im Modell integriert, aber nicht von vorrangigem Interesse sind und für die daher auch keine Bereichsschätzung vorgenommen werden soll. Werden die Modellparameter partitioniert in $\theta = (\theta_1, \theta_2)$, wobei θ_1 die relevanten Parameter und θ_2 die Störparameter bezeichnet, dann geht die Likelihood in die Profil-Likelihood

$$\sup_{\theta_2 \in \Theta} L(x; \theta_1, \theta_2) =: L^*(x; \theta_1) \quad (3.6)$$

über und die relative Likelihood aus (3.5) in

$$R(X, \theta_1) = \frac{\sup_{\theta_2 \in \Theta} L(X; \theta_1, \theta_2)}{\sup_{\theta \in \Theta} L(X; \theta)}. \quad (3.7)$$

Die Profil-Likelihood schätzt die Störparameter konsequent nach dem Likelihoodprinzip. Derjenige Parameter, unter dem die konkrete Stichprobe am wahrscheinlichsten ist, wird auch dann unterstellt, wenn eine genaue Spezifikation des Störparameters aus unterschiedlichsten Ursachen nicht vonnöten scheint. Diesen Ansatz kann man sich zunutze machen, wenn trotz Unkenntnis über die Größenordnung der Zusatzparameter ein Konfidenzbereich für eine Teilmenge von Parametern angestrebt wird.

Falls der MLS existiert, dann ist für die Stichprobenrealisation $x = (x_1, \dots, x_n)$

$$R(x, \theta_1) = \frac{L(x; \theta_1, \hat{\theta}_2)}{L(x; \hat{\theta}_1, \hat{\theta}_2)} \quad (3.8)$$

eine Funktion in θ_1 . Dabei bezeichnen $\hat{\theta}_1$ und $\hat{\theta}_2$ die globalen MLS. $\hat{\theta}_2$ dagegen ist ein „partieller“ MLS und kann von θ_1 abhängen. Der Name Profil-Likelihood rührt von der Tatsache her, dass es sich um eine in der Dimension reduzierte Likelihood handelt, bei der die Werte der Likelihood senkrecht zu den Koordinatenachsen von θ_1 projiziert werden. Ist θ_1 eindimensional, sind Profil-Likelihood und damit auch $R(x, \theta_1)$ Kurvenlinien. Letztere entstehen als Schnitt der Oberfläche $(R(x), \theta_1, \theta_2)$ und einer zur Koordinatenebene $(R(x), \theta_1)$ parallel durch $\hat{\theta}_2$ verlaufenden Ebene.

In den nächsten Abschnitten soll auf einige bisher in der Literatur erwähnte Konfidenzintervallmethoden eingegangen werden, die aufgrund ihrer Allgemeingültigkeit auch für gruppierte oder gerundete Daten tauglich sind, oder auch explizit für

diesen Kontext vorgeschlagen wurden. Ausnahmslos alle sind mit dem Likelihood-Quotienten-Verfahren verflochten und haben ihre Wurzeln in der Testtheorie. Nur das letzte der drei vorgestellten Verfahren bildet eine Art Bindeglied aus beiden Verfahrenskategorien, indem der Versuch unternommen wird, die relative Likelihood auf eine bekannte Pivotgröße zurückzuführen.

3.3 Die χ^2 -Approximation

Der Nachweis ihrer asymptotischen Verteilung macht die relative Likelihood für ein breit gefächertes Spektrum statistischer Modelle zu einem leistungsfähigen Instrument. Wilks (1938) widmete sich der Bestimmung der Grenzverteilung:

Unter den Regularitätsbedingungen (R1)-(R4) aus Abschnitt 2.5 konvergiert die Verteilung von $-2 \ln R$ für $n \rightarrow \infty$ gegen eine χ^2 -Verteilung mit k Freiheitsgraden, wobei k die Zahl der zu schätzenden Parameter angibt.

Es lohnt, den Beweis dieser Kernaussage (aus Serfling, 1980, S. 154) in Grundzügen zu rekapitulieren. Ansatzpunkt und Wirkungsweise von Korrekturmaßnahmen, die die Approximation der relativen Likelihood bei kleinen Stichprobenumfängen verbessern sollen – wie etwa Bartletts Korrekturen –, lassen sich so leichter nachvollziehen.

Eine Annäherung an die Verteilung von $-2 \ln R$ erfolgt über eine Taylorreihenentwicklung 2. Ordnung, entwickelt an der Stelle $\hat{\theta}$. Zunächst zerlegt man die relative Likelihood in ihre Komponenten

$$-2 \ln R = -2 \left[\ln L(\hat{\theta}) - \ln L(\theta) \right]. \quad (3.9)$$

Dann folgt eine Reihenentwicklung des zweiten Summanden mit Young's Form von Taylors Theorem (Serfling, 1980, S. 45):

$$\begin{aligned} \ln L(\theta) &= \ln L(\hat{\theta}) + \sum_{i=1}^k \frac{\partial \ln L(\theta)}{\partial \theta_i} \Big|_{\theta=\hat{\theta}} (\theta_i - \hat{\theta}_i) \\ &\quad + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \frac{\partial^2 \ln L(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta=\hat{\theta}} (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) \\ &\quad + o(\|\theta - \hat{\theta}\|^2). \end{aligned} \quad (3.10)$$

Die ersten Ableitungen ergeben für den MLS Null und fallen somit in der Summe weg. Als Größenordnung des Restgliedes verbleibt $o(n^{-1})$, weil

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow 0 \Leftrightarrow (\hat{\theta} - \theta) = o(n^{-\frac{1}{2}}). \quad (3.11)$$

Stellt man die Doppelsumme in (3.10) in matrizieller Schreibweise dar und formuliert die Likelihood als Produkt der Einzeldichten, ergibt sich für

$$2 \ln R = (\hat{\theta} - \theta)' \left[\sum_{m=1}^n \frac{\partial^2 \ln f((x_m; \theta))}{\partial \theta_i \partial \theta_j} \right] \Bigg|_{\theta=\hat{\theta}} (\hat{\theta} - \theta) + o(n^{-1}) \quad (3.12)$$

$$\begin{aligned} &= n(\hat{\theta} - \theta)' \mathcal{I}(\hat{\theta} - \theta) + n(\hat{\theta} - \theta)' [\hat{\mathcal{I}} - \mathcal{I}](\hat{\theta} - \theta) + o(n^{-1}) \\ &\xrightarrow{n \rightarrow \infty} n(\hat{\theta} - \theta)' \mathcal{I}(\hat{\theta} - \theta), \end{aligned} \quad (3.13)$$

denn $(\hat{\mathcal{I}} - \mathcal{I}) \rightarrow 0$ für $n \rightarrow \infty$. Ist ein Zufallsvektor $T = (T_1, \dots, T_n)' \sim N(\mu, \Sigma)$ und C eine symmetrische $(k \times k)$ -Matrix und $\Sigma C \Sigma C \Sigma = \Sigma C \Sigma$, dann ist die quadratische Form $T' C T$ χ^2 -verteilt und die Anzahl der Freiheitsgrade gleich der Spur von $C \Sigma$. Da die Statistik $\sqrt{n}(\hat{\theta} - \theta)$ für $n \rightarrow \infty$ asymptotisch normalverteilt ist gemäß $N(0, \mathcal{I}^{-1})$ (vgl. (2.31)) folgt mit der Wahl von $C = I$, dass $n(\hat{\theta} - \theta)' \mathcal{I}(\hat{\theta} - \theta)$ asymptotisch χ^2 -verteilt ist. Die Anzahl der Freiheitsgrade ist gleich der Spur $(\mathcal{I} \mathcal{I}^{-1}) = \text{Spur}(I) = k$. Mit den gleichen Hilfsmitteln lässt sich zeigen, dass auch in Modellen mit Störparametern die relative Likelihood χ^2 -verteilt ist. Die Anzahl der Freiheitsgrade richtet sich nach der Anzahl der zu schätzenden Parameter, also Gesamtanzahl der Parameter abzüglich nicht relevanter Störparameter (Serfling, 1980, S. 157). Aus diesen Fakten lässt sich ein asymptotisches Konfidenzintervall gewinnen:

$$K I_{LR} := \{ \theta \mid -2 \ln R \leq \chi_{(k, 1-\alpha)}^2 \}. \quad (3.14)$$

Auf den ersten Blick erinnert diese Schreibweise unwillkürlich an ein einseitiges Konfidenzintervall, ist aber nicht damit zu verwechseln. Da R ausgehend vom MLS nach beiden Richtungen fällt, genügt ein einziges Quantil, um zugleich Ober- wie Untergrenze des Konfidenzbereiches festzulegen. Für einen einseitigen Konfidenzbereich schichtet man die Wahrscheinlichkeitsmasse des kritischen Bereiches auf eine Seite. Die Ersetzung des Quantils in (3.14) durch $\chi_{(k, 1-2\alpha)}^2$ erzeugt dann je nach Bedarf eine einseitige untere oder einseitige obere Konfidenzschranke.

Im Gegensatz zu exakten Konfidenzbereichen können asymptotische Methoden eine Einhaltung des Niveaus nicht garantieren. In finiten Situationen mit kleinen Stichprobenumfängen können sowohl Unter- wie Überschreitungen auftreten. Konservative Testverfahren lehnen die Nullhypothese tendenziell seltener ab und unterschreiten damit das Signifikanzniveau. Sie vergrößern den Annahmebereich und weiten die Konfidenzintervalle. Die Lage des wahren Parameters kann nicht mehr präzise genug geortet werden. Antikonservative Verfahren dagegen verursachen eine Einengung des Bereiches und schließen zu häufig den wahren Parameter aus. Durch Simulationsstudien, die die Überdeckungswahrscheinlichkeiten prüfen, kann man nachverfolgen, in welche Richtung ein Testverfahren neigt. Wie auch immer sich das Likelihood-Quotienten-Verfahren im finiten Fall verhält, asymptotisch stellt es eine beachtliche

Qualität unter Beweis: Der zugehörige Likelihood-Quotienten-Test ((3.3) und (3.4)) ist asymptotisch gleichmäßig mächtigster (Wilks, 1962, S. 417) und produziert damit asymptotisch gleichmäßig genaueste Konfidenzbereiche.

Bartlett (1937) unterbreitet zur Verbesserung der finiten Eigenschaften den Vorschlag, die Approximation an die χ^2 -Verteilung durch eine Folge von multiplikativen Konstanten c_n zu beschleunigen (vgl. auch Spanos, 1999, S.716 f.). Mit seinen Korrekturen beabsichtigt er, die Teststatistik näher an den Erwartungswert der asymptotischen Verteilung zu rücken. Dazu muss die Sequenz der Faktoren so konzipiert sein, dass sich die asymptotische Verteilung der Teststatistik nicht ändert, also

$$c_n \xrightarrow{n \rightarrow \infty} 1. \quad (3.15)$$

Außerdem muss sich der Erwartungswert der korrigierten Teststatistik mit dem der asymptotischen Verteilung decken. Da bei der χ^2 -Verteilung der Erwartungswert mit der Anzahl der Freiheitsgrade k übereinstimmt, heißt das formal:

$$E(c_n R) = k. \quad (3.16)$$

Geeignete Korrekturfaktoren lassen sich finden, wenn der Erwartungswert von R über eine Taylorreihenentwicklung in folgender Form darstellbar ist:

$$E(R) = k \left[1 + \frac{b(X_1, \dots, X_n)}{n} + O(n^{-2}) \right]. \quad (3.17)$$

Falls ein konsistenter Schätzer \hat{b} für $b(X_1, \dots, X_n)$ verfügbar ist, setzt man

$$c_n = \left[1 + \frac{\hat{b}}{n} \right]^{-1}. \quad (3.18)$$

Die Verteilung der korrigierten Teststatistik $c_n R$ liegt deutlich dichter an der χ^2 -Verteilung als die Verteilung der unkorrigierten Version. Barndorff-Nielsen und Hall (1988) haben belegt, dass sich die Distanz der Verteilungen durch die Korrektur bis auf eine Größenordnung von $O(n^{-2})$ verringert, d.h.

$$P(c_n R \leq x) = P(X_k^2 \leq x) + O(n^{-2}), \quad (3.19)$$

wobei X_k^2 eine χ_k^2 -verteilte Zufallsvariable darstellt. Exakte analytische Formeln für die Bartlett-Korrekturfaktoren erfordern eine Integration über dem Ereignisraum und sind daher oftmals schwierig herzuleiten. Approximative Formeln (Lawley, 1956) sind hier ein Ausweg. DiCiccio (1987) stellt für die Momente der leicht abgewandelten Teststatistik

$$T(\theta) = \text{sign}(\theta - \hat{\theta}) \sqrt{-2 \ln R} \quad (3.20)$$

Schätzer bereit. Nachfolgend erwies sich (DiCiccio, 1988), dass man im Fall von lediglich einem relevanten Parameter diese Schätzwerte zu einem approximativen Korrekturfaktor vereinen kann, nämlich

$$\hat{b} = m^2 + s^2 \quad (3.21)$$

mit m als Schätzwert für den Erwartungswert von $T(\theta)$ und s^2 als Schätzwert für die Varianz. Ein korrigiertes Konfidenzintervall baut auf der adjustierten Teststatistik und lautet:

$$BLQ = \left\{ \theta : -2 \frac{\ln R}{\hat{b}} \leq \chi_{(k, 1-\alpha)}^2 \right\}. \quad (3.22)$$

BLQ steht für Likelihood-Quotient mit Bartlett-Korrektur. Doganaksoy und Schmee (1993) haben durch Vergleichsstudien für die Normalverteilung und Minimumverteilung (Grenzverteilung der Ordnungsstatistik $X_{(1)}$, Spezialfall der Extremwertverteilung) belegt, dass das LQ-Verfahren von dieser Strategie im Sinne einer verbesserten Approximation bei kleinen Stichprobenumfängen profitiert. Bartlett-korrigierte LQ-Intervalle weisen allerdings in den untersuchten Fällen eine Überdeckungswahrscheinlichkeit auf, die über dem geforderten Niveau liegt, und sind damit tendenziell zu groß und damit von verringerter Aussagekraft.

3.4 Das Verfahren nach Wald

Die asymptotische Normalverteilung des MLS bereitet unmittelbar den Boden für eine zweite Testvariante und damit eine weitere Konfidenzintervallmöglichkeit. Während Likelihood-Quotienten-Intervalle nur rechnergestützt mit iterativen Verfahren zu berechnen sind, bietet diese Form die Möglichkeit einer schnelleren Berechnung über explizite Schätzformeln. Leider haben sich die Autoren einschlägiger Fachliteratur bisher nicht auf eine einheitliche Benennung dieser Methode geeinigt. In der „klassischen“ Statistik wird diese Methode nach ihrem Urheber Wald-Approximation (Wald, 1943) genannt. Im moderneren Zweig der Lebenszeitanalyse spricht man auch wegen ihres Ursprungs aus der asymptotischen Normalverteilung von der Normalapproximation (vgl. Meeker und Escobar, 1994; Lawless, 1982). In einparametrischen Modellsituationen trägt diese Methode ihren Namen zu Recht. Aber im mehrparametrischen Fall kommen für die so genannte Normalapproximation χ^2 -Quantile zum Einsatz. Um Missverständnissen vorzubeugen, wird in der vorliegenden Arbeit der Begriff Wald-Approximation verwendet.

Wald greift anstelle der relativen Likelihood direkt auf die entwickelte Taylorfunktion (3.12) ohne Restglied zurück:

$$W = n(\hat{\theta} - \theta)' \hat{\mathcal{I}}_{\hat{\theta}} (\hat{\theta} - \theta) \quad (3.23)$$

mit

$$\hat{\mathcal{I}}_{\hat{\theta}} = \left[\sum_{m=1}^n \frac{\partial^2 \ln f(x_m; \theta)}{\partial \theta_i \partial \theta_j} \right] \Big|_{\theta=\hat{\theta}}. \quad (3.24)$$

Aufgrund der quadratischen Form von (3.23) scheint es intuitiv nahe liegend, die Verteilung durch eine χ^2 -Verteilung zu beschreiben. In der Tat stimmen Wald-Statistik W und die relative Likelihood für große n asymptotisch überein (vgl. (3.13)) und besitzen damit dieselbe asymptotische Verteilung. Der Vorteil der Waldstatistik liegt in einer oftmals leichteren Berechenbarkeit und in ihrem anschaulichen Charakter: Der Konfidenzbereich markiert ein Ellipsoid um $\hat{\theta}$ und ergibt sich aus allen Punkten, deren Abstand, gewichtet mit der Inversen der geschätzten Kovarianzmatrix, einen kritischen Wert nicht überschreitet. Sind in das Modell Störparameter eingebunden und somit θ partitioniert in (θ_1, θ_2) , werden diese ignoriert und die Wald-Statistik reduziert sich auf

$$W = n(\hat{\theta}_1 - \theta_1)' \hat{\mathcal{I}}_{\hat{\theta}_1} (\hat{\theta}_1 - \theta_1). \quad (3.25)$$

Auch diese „partielle“ Statistik mit der dimensionsreduzierten Informationsmatrix ist χ^2 -verteilt. Die Anzahl der Freiheitsgrade richtet sich nach der verbleibenden Dimension von θ_1 . Walds Statistik definiert also die Konfidenzregion

$$KI_W = \{\theta_1 : n(\hat{\theta}_1 - \theta_1)' \hat{\mathcal{I}}_{\hat{\theta}_1} (\hat{\theta}_1 - \theta_1) \leq \chi_{(k_1, 1-\alpha)}^2\} \quad (3.26)$$

mit k_1 als Dimension von θ_1 .

Für Sonderfälle, in denen nur ein Konfidenzintervall für einen skalarwertigen Parameter gefragt ist, sind die Bereiche

$$n(\hat{\theta} - \theta)' \hat{\mathcal{I}}_{\hat{\theta}} (\hat{\theta} - \theta) \leq \chi_{(1, 1-\alpha)}^2$$

und

$$\sqrt{n} \frac{\hat{\theta} - \theta}{\sqrt{\widehat{Var}(\hat{\theta})}} \leq z_{(1-\alpha/2)}$$

identisch, und das Quantil der Standardnormalverteilung $z_{(1-\alpha/2)}$ kann in gewohnter Weise ein zweiseitiges Konfidenzintervall festlegen:

$$KI_W = \{\theta : \hat{\theta} - z_{(1-\alpha/2)} \sqrt{\widehat{Var}(\hat{\theta})/n} \leq \theta \leq \hat{\theta} + z_{(1-\alpha/2)} \sqrt{\widehat{Var}(\hat{\theta})/n}\}. \quad (3.27)$$

Ausschlaggebend für eine gute Approximation durch die χ^2 -Verteilung ist eine wohl durchdachte Parametrisierung des Modells. Anders als LQ-Konfidenzintervalle sind die Intervalle basierend auf der Wald-Statistik in der Regel transformationsvariant (s. Bsp. in Meeker und Escobar (1994)). Eine geschickte Parameterwahl oder

Transformation kann entscheidend zu einer besseren Annäherung beitragen. Es gibt keine Faustregel, aber grundsätzlich sollte die Transformation eine „Normalisierung“ der Verteilung von $\hat{\theta}$ bewirken. Manche Transformationen peilen eine Verlagerung des Parameterbereiches an, denn der Wertebereich der Normalverteilung erstreckt sich auf den gesamten reellen Raum. Eine zu schätzende Standardabweichung könnte man beispielsweise logarithmisch, oder zu schätzende Wahrscheinlichkeiten $0 < p < 1$ via $\ln[p/(1-p)]$ transformieren.

Sprott (1973, 1975) beurteilt die Approximationsgüte der Wald-Statistik anhand der Symmetrie der relativen Likelihood. Er durchleuchtet die Auswirkungen einer Umparametrisierung und gelangt zu dem Schluss: Bei der Exponentialverteilung bewirkt eine Transformation von $\theta^* = \theta^{-1/3}$ eine gute Normalisierung und schützt vor allzu großen Abweichungen der Wald-Statistik. Gleichzeitig führt er aus, dass man keineswegs immer eine passende Transformation gewinnen kann. Die Suche nach einer Transformation entpuppt sich oft als so kompliziert, dass sich ein erhöhter Rechenaufwand für ein alternatives Konfidenzintervall über die relative Likelihood lohnt.

3.5 Die Methoden im Vergleich

Die bisher vorgestellten Konfidenzintervalle haben sich nach ihrer Entwicklung rasch etabliert und sind besonders in der Lebenszeitanalyse weit verbreitet (vgl. Jeng und Meeker, 1999). Dieses Fachgebiet benutzt überwiegend Verteilungen, deren Parameter nur mit numerischen Verfahren geschätzt werden können. Dadurch entfällt die Möglichkeit, mit theoretischen Mitteln ihre finiten Verteilungseigenschaften zu ergründen. Mangels finiter Aussagen sind hier die asymptotischen Hilfestellungen zur Gewinnung von Konfidenzintervallen unverzichtbar. Gerade der verstärkte Einsatz macht es unerlässlich, die Verfahren für verschiedenartigste Ausgangslagen auf den Prüfstand zu stellen und ihre Güteeigenschaften zu erörtern:

Vom theoretischen Standpunkt aus ist die Sachlage klar. Schon Serflings Beweis (Abschnitt 3.3) offenbart die enge Verbindung von Likelihood-Quotienten und Walds Statistik: Bei großen Stichprobenumfängen sind die Teststatistiken äquivalent. Das Verhalten bei kleineren Stichproben wurde seither in einer Vielzahl von Monte-Carlo-Studien nachgestellt und ausgewertet. Die Studien sollten vor allem die Umstände identifizieren, unter denen die Konfidenzintervalle von unbefriedigender Qualität sind. Doganaksoy (1995) hat eine Übersicht aller bis dato erhältlichen Beiträge zum Thema „Likelihood-Quotienten-Konfidenzintervalle in der Lebenszeitanalyse“ zusammengestellt. Methodisch erstreckt sich seine Dokumentation über sämtliche vorgestellten Verfahren, einschließlich korrigierter Versionen. Modelliert werden un-

terschiedlichste Lebensdauerverteilungen und Zensurformen. Piegorsch (1987) unterstellt beispielsweise eine zweiparametrische Exponentialverteilung und eine Zensur vom Typ I (die gesamte Observationsdauer unterliegt einer Zeitbeschränkung), DiCiccio (1993) die verallgemeinerte Gamma-Verteilung, Griffith (1980) die dreiparametrische Lognormalverteilung und Lawless (1982) sowie Doganaksoy und Schmee (1993) die Weibullverteilung unter Typ I Zensur. Insgesamt vermittelt der Artikel von Doganaksoy einen umfangreichen Überblick und informiert darüber hinaus, wo Einzelheiten zu einer speziellen Thematik nachzulesen sind.

Im Kontext gerundeter Daten hebt sich besonders die Arbeit von Ostruchov und Meeker (1988) hervor. Sie behandeln approximative Konfidenzbereiche für intervallzensierte Weibull- und lognormalverteilte Daten. Wegen des simplen Zusammenhangs über Datentransformationen (das Logarithmieren einer Zufallsvariablen, die der Minimumverteilung genügt, erzeugt eine Weibull-verteilte Zufallsvariable und eine logarithmierte lognormalverteilte Zufallsvariable ist normalverteilt) behalten die Ergebnisse mit einem kleinen Kunstgriff ihre Gültigkeit. In der Untersuchung spielt für Ostruchov und Meeker die Genauigkeit der Konfidenzbereiche unter dem Aspekt der Überdeckungswahrscheinlichkeit die Hauptrolle. Das Modell intervallzensierter Daten legt zugrunde, dass Ausfälle nur zu bestimmten Inspektionszeiten entdeckt werden können und somit nur ein Zeitabschnitt bekannt ist, in dem sich der Ausfall ereignete. In diesem Punkt gleicht das Modell dem der gerundeten Daten. Zwei Besonderheiten aber unterscheiden die Ansätze:

Der Beobachtungsprozess im intervallzensierten Modell bricht nach einem vorher festgelegten Zeitraum ab. Der genaue Ausfall- oder Todeszeitpunkt der bis dahin überlebenden Individuen entzieht sich folglich einer beidseitigen Eingrenzung. Die Lebenszeit dieser zensierten Daten fällt in ein nach oben offenes Intervall, beginnend mit dem letzten Inspektionszeitpunkt als Mindestlebenszeit. Der erwartete Anteil zu beobachtender Ausfälle prägt das Modell maßgeblich und geht als eigenständiger Parameter in die Simulation ein. Ostruchov und Meeker haben in ihre Studien Anteile von 10% und 50% aufgenommen, wobei 100% dem Modell gerundeter Daten entsprochen hätte. Weiterhin können die Inspektionszeiträume variieren. Diese Option haben Ostruchov und Meeker genutzt und die Inspektionszeiten so eingerichtet, dass in jedem Zeitabschnitt gleich viele Ausfälle erwartet werden. Dieser Kunstgriff bewirkt die Invarianz der Ausfallfrequenzen bei Transformation von Daten und Inspektionszeiten. Inspektionsabschnitte mit gleichen Ausfallwahrscheinlichkeiten der Weibullverteilung lassen sich durch Logarithmieren in Inspektionsabschnitte wiederum gleicher Ausfallwahrscheinlichkeit der Minimumverteilung umwandeln. Außerdem sind in Situationen mit wenigen Inspektionszeiten und geringer Anzahl von Testobjekten nicht oder nicht eindeutig lösbare Fälle der ML-Schätzung seltener,

weil sich die Ausfälle gleichmäßig über alle Abschnitte verteilen. Die Praxis zeigt, dass sich solche Inspektionszeiten mit Näherungswerten für die Parameter umsetzen lassen und sich besonders zur Quantilschätzung der Verteilung gut eignen (Meeker, 1986).

Für die Simulationsstudie werden schließlich neben dem zu erwartenden Anteil von Ausfällen p unter allen Testobjekten die erwartete Anzahl von Ausfällen $E(r)$ sowie die Anzahl der Inspektionszeiten modelliert. Der Anteil und die Zahl von Ausfällen lassen auf den Stichprobenumfang rückschließen, der eine Spanne zwischen 20 und 5000 umfasst. Die Einstellung unterschiedlicher Inspektionsfrequenzen ist -auf gerundete Daten übertragen- im weitesten Sinne als Modellierung der Rundungsgitterbreite zu verstehen. Pro Kombination der Modellparameter werden LQ-Konfidenzintervalle mit Wald- und Likelihood-Quotienten-Approximation berechnet und aus allen Simulationsläufen die Überdeckungshäufigkeit des wahren Parameters notiert. Die Konzeption späterer Simulationsstudien in Kapitel 4 orientiert sich in groben Zügen an diesem Aufbau, wird sich aber von der Handhabung nicht schätzbarer Stichproben deutlich distanzieren: Wenn alle Ausfälle sich im selben Intervall, in zwei angrenzenden oder ausschließlich im ersten und letzten Intervallabschnitt einfinden, kommt es immer wieder zu nicht schätzbaren Stichprobensituationen (nach Silvapulle und Burrige, 1986). Ostruchov und Meeker entfernen diese und ersetzen sie durch einen neuen Simulationslauf, solange bis die geforderte Anzahl schätzbarer Stichproben entstanden sind. Unter den in Kapitel 4 behandelten Parameterkonstellationen kommt es so häufig zu nicht schätzbaren Stichprobensituationen, dass man diese Fälle nicht ignorieren kann, sondern mit einer geeigneten Strategie reagieren muss.

Die Ergebnisse der Simulationsstudien für Weibull- und Lognormal-Verteilung weisen eine klare Richtung und sagen übereinstimmend aus: Der erhöhte Rechenaufwand für die LQ-Intervalle zahlt sich in Form von Überdeckungswahrscheinlichkeiten aus, die durchweg dichter am nominalen Niveau liegen als die der Konfidenzintervalle (KI) bei Wald-Approximation. Selbst bei kleinen Stichprobenumfängen bleiben die Fehler nahezu konstant, wohingegen sie mit Wald-Approximation zunehmen. Sonstige Abhängigkeiten von einem der Simulationsparameter können nicht festgestellt werden, auch nicht von der Anzahl der Inspektionen. Die Schätzung mit Wald-Approximation ist für Lage- und Skalenparameter bei kleinem Stichprobenumfang verzerrt. Unterschätzungen treten weitaus häufiger auf als Überschätzungen. Dagegen liegen Unter- und Überschätzungen der LQ-KI etwa in gleicher Größenordnung. Die Wald-Approximation empfehlen die Ostruchov und Meeker daher erst für eine erwartete Ausfallanzahl ab 50 bzw. 100 und 500 für ein zweiseitiges Konfidenzintervall zu Niveau 90% bzw. 95% und 99%. Den erforderlichen Stichprobenumfang

für ein einseitiges KI zum Niveau 95% und mehr beziffern sie aufgrund der hohen Verzerrung mit $E(r) \gg 500$.

Die gewonnenen Erkenntnisse haben für den Fortgang der Arbeit folgende Konsequenzen:

- Die Güte des Verfahrens ist nicht vom erwarteten Ausfallanteil abhängig, daher darf man auf ähnliche Resultate bei fortgesetzter Beobachtung bis zum Ausfall aller Individuen, d.h. Beobachtung ohne Zensur, schließen. Die Eigenschaften der Likelihood-Quotienten-Methode braucht daher nur für Stichprobenumfänge von weniger als 20 ergänzt zu werden.
- Die Wald-Approximation ist umso schlechter, je geringer der Stichprobenumfang ist und wird nicht in weitere Vergleiche einbezogen. Alle entwickelten und noch zu entwickelnden Verfahren müssen sich an den Qualitäten der LQ-Methode messen.

3.6 Ein Korrekturverfahren von Lee und Vardeman

Die bisher vorgestellten Konfidenzintervalle sind in der Lebenszeitanalyse weit verbreitet. Dort hat man sie bei unterschiedlichen Ausgangsmodellen auf den Prüfstand gestellt und miteinander verglichen. Das Runden aber präsentiert sich in der Lebenszeitanalyse nur als untergeordneter Spezialfall (der Fall äquidistanter Inspektionszeiträume und unzensierter Daten) und die Frage nach praktischen Leitfäden in der Handhabung gerundeter Daten bleibt weiterhin offen. Lee und Vardeman (2001, 2003) haben für gerundete Daten aus einer Normalverteilung diesen Klärungsbedarf ansatzweise nachgeholt. Sie haben die Likelihood-Quotienten-Intervalle einer eingehenden Analyse unterzogen und dabei erneut geprüft, welche Folgen die Behandlung von gerundeten Daten als exakte nach sich zieht. Vorkenntnisse über einen der beiden Normalverteilungsparameter werden in ihrem zu Grunde gelegten Modell nicht verlangt. Sie starten ihre Untersuchungen mit Konfidenzintervallen für den Lageparameter. Der Skalenparameter ist von nachrangigem Belang und fungiert als Störparameter.

Ein erster Testlauf bestätigt die Schwächen des traditionellen Pivot-Konfidenzintervalles über die t-Quantile. Eigentlich für exakte Daten geschaffen, werden Situationen mit tatsächlich kleiner Varianz völlig falsch eingeschätzt. Bei Stichproben mit lauter gleichen Werten schrumpft die Intervalllänge auf Null. Das Intervall trifft den Lageparameter häufig nicht mehr, und enorme Einbußen hinsichtlich der Überdeckungswahrscheinlichkeit sind zu verzeichnen. Bei hohen Varianzen sind diese Fälle auch bei sehr kleinen Stichprobenumfängen nahezu

ausgeschlossen und die Einhaltung des Niveaus gelingt. Auch das LQ-Verfahren zeigt sich nicht allen Parameterveränderungen gewachsen. Typisch für approximative Verfahren erreicht die Überdeckungswahrscheinlichkeit das Nominalniveau $(1 - \alpha)$ nur bei größeren Stichprobenumfängen. Für kleine Stichprobenumfänge und kleine Varianzen zeichnet sich ein eher konservatives Verhalten ab, wodurch die Intervallschätzung erheblich an Genauigkeit einbüßt. Bei großen Varianzen dagegen verfehlen die Intervalle oft den tatsächlichen Parameter. Ähnliche Beobachtungen hatten schon u.a. Bartlett (1937) zu Korrekturmaßnahmen veranlasst. Während diese die Teststatistik modifizieren, nehmen Lee und Vardeman Veränderungen der kritischen Werte vor und suchen nach solchen $c(n, \alpha)$, die zwar nach wie vor den Niveaueanforderungen genügen

$$P(-2 \ln R \leq c(n, \alpha)) \geq 1 - \alpha, \quad (3.28)$$

aber den Annahmebereich etwas verbreitern

$$c(n, \alpha) > \chi_{(1; 1-\alpha)}^2 \quad (3.29)$$

und somit die Trefferquote steigern. Die neuen kritischen Werte müssen aber asymptotisch identisch mit denen der χ^2 -Verteilung sein. Lee und Vardeman argumentieren, dass sich für große Standardabweichungen die beiden Teststatistiken, basierend einmal auf gerundeten und ein andermal auf exakten Daten, einander angleichen (vgl. Abschnitt 2.3.). Man könnte also versuchen, aus der Likelihood für exakte Daten solche kritischen Werte zu gewinnen. Tatsächlich lässt sich aus der Teststatistik $R_{\text{exakt}}(X, \mu)$ für den Erwartungswert der Normalverteilung eine Pivotgröße mit bekannter Verteilung isolieren:

$$\begin{aligned} L(X, \hat{\mu}, \hat{\sigma}) &= \left(2\pi \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^{-n/2} \\ &\quad \times \exp - \left[\sum_{i=1}^n (X_i - \bar{X})^2 / \left(\frac{2}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right) \right] \\ &= (2\pi)^{-n/2} \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{-n/2} e^{-n/2} \end{aligned} \quad (3.30)$$

und

$$L(X, \mu, \hat{\sigma}) = (2\pi)^{-n/2} \left[\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \right]^{-n/2} e^{-n/2}, \quad (3.31)$$

mit $\hat{\sigma}$ als partiellem MLS (vgl. (3.8)). Damit schreibt sich die relative Likelihood

der exakten Daten $R_{\text{exakt}}(X, \mu)$ als

$$\begin{aligned}
\Rightarrow -2 \ln R_{\text{exakt}}(X, \mu) &= -2 \ln \left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right]^{-n/2} \\
&= n \ln \left[\frac{\sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \\
&= n \ln \left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \\
&= n \ln \left[1 + \frac{1}{n-1} \left(\sqrt{n} \frac{\bar{X} - \mu}{S} \right)^2 \right] \\
&= n \ln \left[1 + \frac{1}{n-1} T^2 \right]. \tag{3.32}
\end{aligned}$$

Die Quantile der t -Verteilung könnten also als neue kritische Werte

$$c(n, \alpha) = n \ln \left[1 + \frac{1}{n-1} t_{(n-1, 1-\frac{\alpha}{2})}^2 \right] \tag{3.33}$$

fungieren und den Konfidenzbereich definieren:

$$KI_{\text{korrr}} = \left\{ \mu \mid -2 \ln R_{\text{round}} \leq n \ln \left[1 + \frac{1}{n-1} t_{(n-1, 1-\frac{\alpha}{2})}^2 \right] \right\}. \tag{3.34}$$

Diese Modifizierung löst das Problem der unter-nominalen Überdeckungswahrscheinlichkeiten bei großen Varianzen, wie die von Lee und Vardeman durchgeführten Simulationsläufe belegen. Bei kleinen bis moderaten Varianzen zeigt das Verfahren immer noch einen konservativen Hang, ist aber der traditionellen Methode, die die Rundung ignoriert, überlegen. Für den kleinsten untersuchten Stichprobenumfang $n = 2$ und kleine Varianzen ($\sigma \leq 0.25$) sind die angegebenen Intervalle oft inakzeptabel groß. Ohnehin entfallen laut 3-sigma-Regel 99.9% der Normalverteilungswerte auf den Bereich $[\mu - 3\sigma, \mu + 3\sigma]$, womit ein Konfidenzintervall der Länge 6 bei einer Standardabweichung von $\sigma = 0.01$ wertlos wird. Zudem häuft sich die Zahl der Stichproben, in denen sich eine ML-Schätzung verbietet (vgl. Abschnitt 2.3).

Die Analyse der Konfidenzintervalle für die Standardabweichung verläuft nach dem gleichen Schema. Zunächst werden in einem ersten Schritt Likelihood-Quotienten-Intervalle für gerundete Daten und das Standardintervall aus der Pivotgröße

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{(n-1)}^2 \tag{3.35}$$

gegenübergestellt. Entgegen allen Erwartungen schneidet das Likelihood-Quotienten-Verfahren – obwohl auf die Verteilung der gerundeten Daten abgestimmt – im Vergleich schlecht ab. Selbst bei kleinen Stichprobenumfängen liefert die traditionelle Methode höhere Überdeckungswahrscheinlichkeiten, es sei denn, die Standardabweichung ist ausgesprochen klein ($\sigma \leq 0.5$). Wieder setzen Lee und Vardeman über die

relative Likelihood der exakten Daten zu einer Anpassung der kritischen Werte an, um die Überdeckungswahrscheinlichkeiten auf das vorgegebene Niveau anzuheben. Unter Verwendung des partiellen MLS $\hat{\mu}$ (vgl. (3.8)) ist

$$\begin{aligned} L(X, \hat{\mu}, \sigma) &= \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}} \exp \left[- \left(X_i - \hat{\mu} \right)^2 / 2\sigma^2 \right] \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[- \sum_{i=1}^n (X_i - \bar{X})^2 / 2\sigma^2 \right] \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[\frac{1}{2} Y \right] \quad \text{mit} \end{aligned} \quad (3.36)$$

$$Y := \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$$

$$\begin{aligned} L(X, \hat{\mu}, \hat{\sigma}) &= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} \exp \left[- \sum_{i=1}^n (X_i - \bar{X})^2 / 2\hat{\sigma}^2 \right] \\ &= (2\pi\hat{\sigma}^2)^{-\frac{n}{2}} e^{-\frac{n}{2}} \end{aligned} \quad (3.37)$$

$$\begin{aligned} \Rightarrow -2 \ln R_{\text{exakt}}(X, \sigma) &= -2 \ln \left[\left(\frac{\sigma^2}{\hat{\sigma}^2} \right)^{-\frac{n}{2}} \exp \left(-\frac{Y}{2} + \frac{n}{2} \right) \right] \\ &= Y - n - n \ln Y + n \ln n. \end{aligned} \quad (3.38)$$

Die relative Likelihood setzt sich also aus Pivot-Bausteinen zusammen, und ihre Verteilung ist auch im finiten Fall unabhängig von σ . Weil aber die Funktion (3.38) nicht invertierbar ist, lässt sich Y nicht herauslösen, und die χ^2 -Quantile sind für einen Konfidenzbereich unbrauchbar. Die Quantile von $Y - n \ln Y$ für verschiedene Kombinationen von n und α sind nur mit numerischen Methoden zu ermitteln. Eine Monte-Carlo-Simulation der Verteilung liefert die erforderlichen kritischen Werte. Das χ^2 -Quantil in der Formel für das approximative Likelihood-Quotienten-Konfidenzintervall tauschen Lee und Vardeman nun gegen das $(1 - \alpha)$ -Quantil dieser Verteilung aus. Wiederholte Simulationsläufe belegen, dass diese Ersetzung ihren Zweck erfüllt. Für großes σ steigen die Überdeckungswahrscheinlichkeiten deutlich an und ziehen mit dem traditionellen Pivot-Verfahren gleich. Lediglich für moderate σ sinken die Werte noch unter das nominale Niveau. Dieses Absinken ist auf eine übermäßige Produktion von Stichproben mit geringer Stichprobenvarianz zurückzuführen. Ein Beispiel beim ganzzahligen Runden kann diesen Sachverhalt illustrieren:

Unter der Parameterkonstellation $\mu = 0, \sigma = 1$ und $n = 3$ werden als Stichproben Kombinationen der Werte $-1, 0$ und 1 generiert. Und zwar $(-1, -1, -1)$ ebenso wie $(0, 0, 0)$ und $(1, 1, 1)$. Außerdem entstehen häufig Stichproben mit einer Spannweite von 1 , wie $(-1, -1, 0)$, $(1, 1, 0)$ etc. Katalogisiert man eine Vielzahl erzeugter Stichproben, prägen Dreier-Kombinationen mit geringer Stichprobenvarianz das Erscheinungsbild. Sie täuschen über die wahre Varianz hinweg. Diese besitzt aber eine

Größenordnung, die das LQ-Intervall aus dem Möglichkeitsbereich ausgrenzt. Will man dem unerwünschten Absinken der Überdeckungswahrscheinlichkeit entgegenwirken, scheint eine Weitung des Intervalls unumgänglich. Eine zweckmäßige Korrekturmaßnahme bedient sich der bezüglich μ maximierten Auftretswahrscheinlichkeiten. Die maximale Wahrscheinlichkeit für die Realisation einer Stichprobe wird dazu in ihrem Verlauf über σ aufgezeichnet. Um nominelle Überdeckungswahrscheinlichkeiten sicherzustellen, muss die Gesamthäufigkeit aller Stichproben mit einer Spannweite von weniger als zwei (Stichproben, von denen man weiß, dass das LQ-Intervall nicht trifft) unter dem geforderten Niveau bleiben. Als obere Konfidenzgrenze wird dann dasjenige σ gewählt, bei dem erstmalig die Wahrscheinlichkeit all dieser Stichproben das Niveau überschreitet. Die Bestimmung dieser Grenzen erfordert einigen Rechenaufwand. Lee und Vardeman haben eine entsprechende Tabelle beigefügt.

4 Ein Kombinationsverfahren für Lageparameter

4.1 Motivation

Die vorliegende Arbeit entstand in der Intention, das von Lee und Vardeman vorgeschlagene Verfahren auf gerundete Daten aus nicht normalverteilten Grundgesamtheiten zu übertragen. Die Vermutung liegt nahe, dass die Behandlung von gerundeten Daten als exakte ganz analog bei geringer Variabilität der Daten in gravierender Weise die Trefferquote der Konfidenzintervalle senkt. Weil die spezielle Likelihoodfunktion der gerundeten Daten in das LQ-Intervall einfließt, scheint dieses Verfahren prinzipiell adäquat, muss sich aber in realistischen Situationen mit geringem Stichprobenumfang erst einer Bewährungsprobe unterziehen. Bei approximativen Verfahren, angewendet auf kleine Stichproben, zeichnen sich häufig ähnliche Symptome ab: Keine noch so hohe Variabilität kann eine späte Näherung der Verteilung des Likelihoodquotienten an die χ^2 -Verteilung ausgleichen und so eine niedrige Trefferquote verhindern. Bevor sich allerdings Korrekturmaßnahmen in Simulationen als tatsächlich notwendig erweisen, stellt sich die Frage nach der grundsätzlichen Eignung möglicher Verteilungen für das beschriebene Korrekturverfahren. Das heißt, der Likelihoodquotient der exakten Daten muss nicht nur asymptotisch, sondern auch im endlichen Fall Pivotfunktion sein. Er muss sich in eine Funktion einer Zufallsvariablen umformen lassen, deren Verteilung vom zu schätzenden Parameter unabhängig ist. Bestenfalls ist deren Verteilung bereits bekannt, andernfalls kann sie in Simulationen nachgestellt werden. Die entsprechenden Quantile dieser Verteilung und der MLS für gerundete Daten gehen dann in die Berechnung der korrigierten Intervallgrenzen ein. Nur dann besitzt man ein geeignetes Utensil, um eventuelle Unzulänglichkeiten der LQ-Intervalle zu beheben.

Durch ihre häufige Anwendung im industriellen Bereich ist eine Betrachtung folgender Verteilungen lohnend: Lognormal-, Exponential-, Logistische, Weibull-, Pareto-Verteilung und Extremwertverteilung (Verteilung des Minimums und Verteilung des Maximums). Aus der Menge dieser Kandidaten kommen theoretisch drei für das Verfahren in Frage:

- Zweiparametrische Lognormalverteilung

Laut Definition dieser Verteilung ist für eine lognormalverteilte Zufallsvariable X die Zufallsvariable $\ln X$ normalverteilt. Daher gelten analoge Gesetzmäßigkeiten wie für die Normalverteilung, und die relative Likelihood kann analog zur Normalverteilung geschrieben werden als

$$-2 \ln R = -n \ln \left(1 + \frac{1}{n} Z^2 \right)$$

mit

$$Z = \sqrt{n}(\mu - \hat{\mu})/\hat{\sigma}, \quad \hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2,$$

wobei man zur Schätzung die Daten logarithmiert und dann die MLS der Normalverteilung, Mittelwert und Stichprobenvarianz auf die transformierten Daten anwendet. Für alle Verteilungsfamilien, die einen Lage- und einen Skalenparameter besitzen, ist diese Größe für den MLS Pivot. Analog zur Normalverteilung ist diese Größe $t_{(n-1)}$ -verteilt.

- Exponentialverteilung

Ähnlich wie Normal- und Lognormalverteilung ist auch die Exponentialverteilung über einen Lage- und einen Skalenparameter charakterisiert. Sie ist außerdem aufgrund ihrer einfachen Form leicht zu handhaben. ein Konfidenzintervall zum Skalenparameter λ sind Dichteformen mit und ohne Verschiebungsparameter separat zu behandeln.

Die Exponentialverteilung ohne Verschiebungsparameter (vgl. (2.18)) enthält keine nebensächlichen Störparameter, daher entfällt der Übergang auf eine Profil-Likelihood und die relative Likelihood mit dem MLS $\hat{\lambda} = \bar{X}$ kann geschrieben werden als

$$\begin{aligned} -2 \ln R(X, \lambda) &= -2 \left[\ln L(X, \lambda) - \ln L(X, \hat{\lambda}) \right] \\ &= -2 \left[\ln \left(\lambda^{-n} e^{-(1/\lambda) \sum_{i=1}^n X_i} \right) - \ln \left(\bar{X}^{-n} e^{-n} \right) \right] \\ &= -2 \left[n \ln(1/\lambda) - (1/\lambda) \sum_{i=1}^n X_i + n \ln \bar{X} + n \right] \\ &= (2/\lambda) \sum_{i=1}^n X_i - 2n \ln(\bar{X}/\lambda) - 2n \\ &= Z - 2n \ln Z - 2n(1 - \ln 2n). \end{aligned} \tag{4.1}$$

Die Zufallsvariable $Z = (2/\lambda)\sum X_i$ ist χ^2 -verteilt mit $2n$ Freiheitsgraden. Das lässt sich mit Hilfe der momenterzeugenden Funktion m nachweisen:

$$\begin{aligned} m_Z(t) &= E \left(e^{(2t/\lambda) \sum_{i=1}^n X_i} \right) \\ &= E \left(\prod_{i=1}^n e^{(2t/\lambda) X_i} \right) \\ &= \prod_{i=1}^n E \left(e^{(t2/\lambda) X_i} \right) \end{aligned} \tag{4.2}$$

$$\begin{aligned}
&= \prod_{i=1}^n m_{X_i}(2t/\lambda) \\
&= \left(\frac{1/\lambda}{1/\lambda - 2t/\lambda} \right)^n \\
&= \left(\frac{1}{1 - 2t} \right)^n.
\end{aligned} \tag{4.3}$$

Letztere ist die momenterzeugende Funktion der χ^2 -Verteilung mit $2n$ Freiheitsgraden. Folglich ist die Likelihood der exakten Daten Pivotgröße, und ihre Quantile könnten bei Bedarf als Korrektur dienen.

Sind die Daten nach rechts verschoben, muss das Modell zusätzlich einen Verschiebungsparameter aufnehmen, somit ist

$$f(x) = \frac{1}{\lambda} e^{-(x-\mu)/\lambda} I_{[\mu, \infty)}(x), \quad \lambda > 0, \mu \geq 0.$$

Der Lageparameter wird nach der ML-Methode durch das Minimum der Stichprobe geschätzt. Nach Abzug dieses Lageschätzers von den Daten ergibt sich $\hat{\lambda}$ wie gewohnt als Mittelwert

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n (X_i - X_{(1)}).$$

Am Ende resultiert die äußerlich gleiche Formel für die Likelihood der exakten Daten wie bei der Exponentialverteilung ohne Verschiebungsparameter, d.h.

$$\begin{aligned}
-2 \ln R(X, \lambda) &= -2 \left[\ln L(X, \hat{\mu}, \lambda_0) - \ln L(X, \hat{\mu}, \hat{\lambda}) \right] \\
&= Z - 2n \ln Z - 2n(1 - \ln 2n)
\end{aligned} \tag{4.4}$$

und

$$Z = 2n \frac{\hat{\lambda}}{\lambda}.$$

Der subtile Unterschied verbirgt sich in der Verteilung von Z . Die Anzahl der Freiheitsgrade reduziert sich auf $2(n-1)$, da $\hat{\lambda}$ nur noch $(n-1)$ Summanden beinhaltet (vgl. Bain und Engelhardt, 1991, S. 156).

- Paretoverteilung

Die Paretoverteilung sieht zwar auf den ersten Blick nicht die Verwandtschaft mit der Exponentialfamilie an, aber sie lässt sich durch eine simple Transformation in die Exponentialverteilung überführen. Mit der folgenden Parametrisierung

$$F(x) = 1 - \left(\frac{x}{\alpha} \right)^{-\sigma} I_{[\alpha, \infty)}(x), \quad \alpha > 0, \sigma \geq 0, \tag{4.5}$$

ist diese Verknüpfung gut zu erkennen:

$$F(x) = 1 - e^{-\sigma(\ln x - \ln \alpha)}. \quad (4.6)$$

Die logarithmierten Paretodaten $\ln x$ sind also exponentialverteilt mit Skalenparameter $1/\sigma$ und Lageparameter $\ln \alpha$. Damit sind – zumindest bei exakten Daten – die MLS und Umformungen des Likelihood-Quotienten aufgrund der Invarianzeigenschaft des MLS von der Exponentialverteilung auf die Paretoverteilung übertragbar. So ergibt sich der MLS für den Lageparameter α als Minimum der Originaldaten

$$\hat{\alpha} = X_{(1)}$$

und für den Skalenparameter als reziprokes Mittel der zuerst transformierten und im Anschluss verschobenen Daten

$$\hat{\sigma} = \left[\frac{1}{n} \sum_{i=1}^n (\ln X_i - \ln X_{(1)}) \right]^{-1}.$$

Auch der Likelihoodquotient bleibt unverändert

$$-2 \ln R = Z - 2n \ln Z - 2n(1 - \ln 2n) \quad (4.7)$$

mit

$$Z = 2n \frac{\sigma}{\hat{\sigma}} \sim \chi_{2(n-1)}^2.$$

Die Vorgehensweise bleibt für alle Verteilungen gleich. Man versucht, die relative Likelihood so umzustellen, dass sie nur noch von bekannten Pivotgrößen abhängt und macht sich dabei die Pivoteigenschaft der MLS für Verteilungsfamilien mit Lage- und Skalenparameter zunutze.

Auch die übrigen in Erwägung gezogenen Verteilungen besitzen einen Lage- und einen Skalenparameter. Meist existieren sogar schon Tabellen der Quantile. Obwohl damit eine ebensolche Lösung für Weibull-, Extremwert- und Logistische Verteilung zum Greifen nahe scheint, glückt eine Umformung nicht. Bereits der erste Schritt scheitert, weil ihre MLS nicht in geschlossener Form darstellbar sind, sondern iterative Methoden aushelfen müssen. Die relative Likelihood lässt sich nicht formal aufschreiben, geschweige denn in Pivot-Bestandteile zerlegen. Leider ist damit die Menge der Verteilungen, die prinzipiell für das Verfahren von Lee und Vardeman in Frage kommen, schnell erschöpft. Solche Verbesserungsmaßnahmen scheinen einzig und allein Normal- und Exponentialverteilungen (die anderen lassen sich letztendlich auf diese zurückführen) vorbehalten. Dennoch bliebe für alle Verteilungen zu

klären, ob überhaupt Korrekturen nötig sind. In einer direkten Gegenüberstellung des LQ-Verfahrens mit dem Pivotverfahren (seinem größten Kontrahenten für kleine Stichproben) ließe sich herausfinden, welchen Folgen die Vernachlässigung des Rundens nach sich zieht. Falls es denn wirklich nicht möglich sein sollte, das LQ-Verfahren zu überarbeiten, könnte man vielleicht die Leistung des Pivot-Verfahrens steigern? Schließlich besteht doch das einzige Manko des Pivot-Verfahrens darin, Treffsicherheit zu versprechen, wo Vorsicht angebracht wäre, nämlich bei geringer Stichprobenvariabilität. Könnte man nicht verhindern, dass dann die Intervalle auf einen einzigen Punkt schrumpfen? Der Beantwortung dieser Fragen sind die nächsten Abschnitte gewidmet.

4.2 Das Grundprinzip

Dieses Kapitel soll zunächst die grundsätzliche Vorgehensweise beschreiben. Eine detaillierte Schilderung der Endergebnisse, sowie der genaue Studienaufbau folgen im entsprechenden Abschnitt über die spezielle Verteilung bzw. das Simulationsdesign. An dieser Stelle werden nur diejenigen Resultate vorweggenommen, die maßgeblich den Werdegang des Kombinationsverfahrens dokumentieren.

Die Versuchsreihe beginnt mit der Maximumverteilung (Grenzverteilung der Ordnungstatistik $X_{(n)}$; Spezialfall der Extremwertverteilung) mit Verteilungsfunktion

$$F(x) = \exp \left[-e^{-\left(\frac{x-u}{b}\right)} \right], \quad -\infty < u < \infty, \quad b > 0. \quad (4.8)$$

und konzentriert sich zunächst auf den Lageparameter u , der in der Praxis in erster Linie interessiert. Die Beurteilung der Verfahren setzt als Hauptzielkriterium die Überdeckungswahrscheinlichkeit an. Erst in zweiter Instanz werden die Intervalllängen hinzugezogen. Bei gleichen Überdeckungswahrscheinlichkeiten ist das Verfahren mit geringerer Intervalllänge vorteilhafter. Nach etlichen Simulationsläufen mit wechselndem Stichprobenumfang n , Lageparameter u und Skalenparameter b lässt sich eine erste Zwischenbilanz ziehen: In der Gegenüberstellung zwischen LQ (für gerundete Daten) versus Pivot-Verfahren (unter Vernachlässigung des Rundens) kann keines der beiden Verfahren das andere eindeutig ausstechen. Kein Verfahren erzielt global über alle Kombinationen eine höhere Treffsicherheit. Bei kleinem Skalenparameter überdeckt der LQ-Konfidenzbereich den Parameter sogar mit absoluter Sicherheit, während – wie bereits vermutet – die Intervalllänge des Pivots auf Null sinkt und damit meist den Parameter nicht erfasst. Eine größere Variabilität der Daten unterstützt die Treffgenauigkeit des Pivot-Verfahrens, wohingegen sie dem LQ-Verfahren die Schätzung erschwert. Selbst bei sehr großem Skalenparameter stabilisieren sich die Überdeckungswahrscheinlichkeiten nicht auf dem geforderten

Niveau. Erst ab $n = 15$ kann das LQ-Verfahren in diesem Spektrum aufholen. Dazwischen existiert ein Bereich moderat großer Skalenparameter, in dem LQ wie Pivot gleichermaßen Schwächen zeigen, die in einer deutlichen Niveauunterschreitung zutage treten. Dies ist ein klares Signal dafür, dass in irgendeiner Art und Weise Korrekturmaßnahmen in Anspruch genommen werden müssen, um ein Verfahren (oder beide) zu optimieren. Bedauerlicherweise fällt die Maximumverteilung in die Kategorie derjenigen Verteilungen, für die sich die kritischen Werte nach der Methode von Lee und Vardeman nicht abwandeln lassen. Dennoch kann man partiell einen ihrer Gedankengänge verwerten:

Bei einem genauen Blick auf die geschätzten Intervalle stößt man auf die Ursache des Absinkens der Überdeckungswahrscheinlichkeiten in gewissen Parameterbereichen. Allzu häufig werden Stichproben von lauter identischen Werten generiert, wobei dieser Wert vom wahren Lageparameter abweicht und unter beiden Verfahren zu Fehleinschätzungen führt. Typischerweise werden solche Stichproben mit einer Häufigkeit registriert, die das nominale Niveau übersteigt und dadurch die Niveaueinhaltung verhindert. Alle anderen Stichproben mit höherer Spannweite führen zumindest bei einem der Verfahren zu einer treffgenauen Bereichsschätzung. Mit einer ähnlichen Problematik sahen sich Lee und Vardeman bei Konfidenzintervallen für die Varianz der Normalverteilung konfrontiert. Sie entschieden sich dafür, in diesen „Problemzonen“ das LQ-Intervall zu ersetzen. Das neue Intervall sollte alle Parameterwerte ausgrenzen, die die Häufigkeit der problematischen Stichproben über das nominale Niveau des Konfidenzbereiches ansteigen ließen. Dazu wählten sie als neue obere und untere Intervallgrenze diejenigen Parameterwerte, für den die Likelihood gleich dem nominalen Niveau des Konfidenzbereiches ist. Tritt eine Stichprobe gleicher Werte auf, schaltet das Verfahren auf dieses neue Intervall um.

In wiederholten Simulationsläufen wird dieser „Schalter“ zunächst in das Pivotverfahren eingebaut. Die gewünschte Wirkung stellt sich zwar ein, bläht aber bei kleinem Skalenparameter die Intervalllängen unnötig auf. Dies scheint besonders inakzeptabel vor dem Hintergrund, dass das Likelihood-Verfahren doch bei einem Drittel der Länge schon eine hundertprozentige Trefferquote liefert. Dieser Makel – der im übrigen auch dem Verfahren von Lee und Vardeman anhaftet – sollte möglichst vermieden werden. Eine Koppelung der Korrektur mit dem LQ-Verfahren dagegen schützt nicht vor einer schlechten Quote bei hoher Variabilität. Noch immer ist keines der Verfahren dem anderen durchgängig überlegen.

Die Befreiung aus dieser Zwickmühle könnte in einer geschickten Kombination beider Verfahren liegen. Die Fixierung eines Schwellenwertes für den Rundungsgrad könnte dazu dienen, dasjenige Verfahren zu wählen, das im konkreten Fall die meisten Vorteile verspricht. So wäre es möglich, von den Stärken beider Verfahren zu

profitieren. Diese Überlegungen geben den Impuls für das

Kombinationsprinzip:

1. Gesucht sei ein Konfidenzbereich für einen skalaren Parameter. Bestimme für Stichproben $x = x_1, \dots, x_n$ mit Spannweite 0 die korrigierten Intervallgrenzen

$$KI(x_1, \dots, x_n = 0) = \{\mu | L^*(\mu) = \sup_{\sigma > 0} L(\mu | x = 0) \leq \alpha\} = [\Delta_u, \Delta_o], \quad (4.9)$$

bzw. bei Verteilungen mit Lage- und Skalierungsparameter:

$$KI(x_1, \dots, x_n = k) = [k + \Delta_u, k + \Delta_o]. \quad (4.10)$$

Verknüpfe diese Korrekturen mit dem LQ-Verfahren, so dass die Korrekturen bei Entscheidung für LQ zur Anwendung kommen.

2. Lege einen Schwellenwert (Erfahrungswert aus Simulationen) für den Rundungsgrad $r = w/\sigma$ fest.
3. Behandle zunächst die gerundeten Werte als exakte Daten. Berechne die MLS für Lage- und Skalenparameter.
4. Schätze den Rundungsgrad mit Hilfe des MLS für den Skalenparameter. Liegt der geschätzte Rundungsgrad oberhalb des Schwellenwertes, dann verwende das LQ-Verfahren für gerundete Daten. Ansonsten ermittle über die beiden MLS für Lage- und Skalenparameter das Pivot-Konfidenzintervall.

Korrekturintervalle für verschiedene Stichprobenumfänge und Schwellenwert werden einmalig für jede Verteilung berechnet und stehen danach zur Verfügung. Die ersten beiden Schritte der Handlungsvorschrift entfallen ab sofort und der Anwender kann mit den Entscheidungshilfen in Schritt 3 und 4 das jeweils passende Verfahren für seine individuelle Problemstellung aussuchen.

Einige Vorteile gegenüber dem Verfahren von Lee und Vardeman liegen auf der Hand:

- Es lässt sich auf ein weitaus größeres Spektrum von Verteilungen ausdehnen. Viele Verteilungen, die sich den Korrekturmaßnahmen von Lee und Vardeman entziehen, sind auf diese Weise zu behandeln.
- Die Intervalle sind in den meisten Fällen leichter zu berechnen. Gerade wenn der Rundungsgrad den Schwellenwert unterschreitet und die Pivotmethode präferiert wird, kann der Benutzer bereits existierende Algorithmen und Programmpakete zur Schätzung weiterhin problemlos verwenden. Hinzu kommt, dass meistens die zur Verteilung der Pivotgrößen gehörenden Quantile tabelliert vorliegen und nicht neu erzeugt werden müssen.
- Der Schwellenwert beantwortet endlich die bislang ungeklärte Frage, in welchen Situationen das Runden als erheblich und für die Schätzung relevant einzustufen ist: Bei Rundungsgraden unter dem Schwellenwert darf man das Runden zugunsten einfacherer Berechenbarkeit ignorieren. Der Anwender braucht nicht pauschal alle Resultate bisheriger Analysen zu revidieren. Ein Blick auf Rundungsgrad und Schwellenwert genügt, um zu beurteilen, ob sich der Aufwand einer Überarbeitung auszahlt.
- Das künstliche Aufblähen der LQ-Intervalle wird weitestgehend vermieden, so dass man den Parameter wesentlich schärfer eingrenzen kann.
- Lee und Vardeman geben für Stichproben mit geringer Spannweite umfangreiche Tabellen mit Intervallgrenzen vor. In der programmiertechnischen Umsetzung (vgl. „Programmiertechnische Details“) lässt sich ein lästiges Nachblättern völlig umgehen. Die Implementierung kommt mit dem χ^2 -Quantil und Schwellenwert aus. Abgesehen von dem Korrekturintervall werden alle Intervalle vom Programm berechnet, interne Fallabfragen können damit vermieden werden.

In den folgenden verteilungsspezifischen Abschnitten von Kapitel 5 wird dieses Verfahren seine Güte unter Beweis stellen. Dabei wird sich zeigen, dass es als Kombinationssystem mehr ist als die Summe seiner Einzelkomponenten. Es ist nicht nur dem Verfahren mit jeweils höherer Überdeckungswahrscheinlichkeit gleichgestellt, sondern überbietet beide.

4.3 Das Simulationsdesign

Grundsätzlich ist eine Analyse für ganzzahliges Runden hinreichend. Unter der Prämisse, dass die Daten aus einer Verteilungsfamilie mit Lage- und Skalenparameter

stammen (wie hier durchgängig der Fall), darf man die Daten beliebig verschieben und neu skalieren. Auf die Originaldaten lässt sich wegen der Invarianzeigenschaft des MLS über die zugehörige inverse Transformation rückschließen. Die Rücktransformation beziffert sogar die Intervalllänge. Die Pivotgröße, deren Verteilung und die Quantile q_u bzw. q_o bleiben wegen

$$E(Y) = w\mu_X + a, \quad \text{Var}(Y) = w^2\sigma_X^2 \quad (4.11)$$

von der Transformation unberührt

$$q_u \leq \frac{\hat{\mu}_Y - \mu_Y}{\hat{\sigma}_Y} = \frac{w\hat{\mu}_X + a - w\mu_X - a}{w\hat{\sigma}_X} = \frac{\hat{\mu}_X - \mu_X}{\hat{\sigma}_X} \leq q_o, \quad (4.12)$$

und damit sind die gleichen Quantile zu verwenden. Das Intervall der transformierten Daten verlängert sich schließlich um den Faktor w : Das Konfidenzintervall der Originaldaten

$$KI(\mu_X) = [\hat{\mu}_X - q_o\hat{\sigma}_X, \hat{\mu}_X - q_u\hat{\sigma}_X]$$

hat eine Länge von

$$q_u\hat{\sigma}_X - q_o\hat{\sigma}_X \quad (4.13)$$

und das Konfidenzintervall der transformierten Daten

$$KI(\mu_Y) = [w(\hat{\mu}_X + a - q_o\hat{\sigma}_X), w(\hat{\mu}_X + a - q_u\hat{\sigma}_X)]$$

eine Länge von

$$w(q_u\hat{\sigma}_X - q_o\hat{\sigma}_X). \quad (4.14)$$

Aus diesem Grunde ist es gerechtfertigt, die Daten zur weiteren Bearbeitung mit einer Division durch die Abstandsweite w des Rundungsgitters in den natürlichen Zahlenbereich zu übertragen. Das erleichtert eine erste grobe Schätzung des Rundungsgrades, der sich nun zu $1/\hat{\sigma}$ vereinfacht. Der geschätzte Rundungsgrad bleibt bei der Transformation fest:

$$\hat{r} = \frac{w}{\hat{\sigma}_X} = \frac{1}{\frac{1}{w}\hat{\sigma}_X} = \frac{1}{\hat{\sigma}_Y}. \quad (4.15)$$

Ebenfalls wegen der Invarianz gegenüber Verschiebungen genügt eine Modellierung des Lageparameters im Intervall $[0, 1]$. Die Resultate lassen sich für alle anderen Gitterabschnitte übernehmen. In der vorliegenden Studie ist dieser Zwischenraum in Zehnteldistanzen untergliedert, und der Lageparameter nimmt die Werte $\mu = (0, 0.1, 0.2, 0.3, \dots, 0.9)$ an. Genau genommen ist bei symmetrischen Verteilungen eine Aufteilung von 0 bis 0.5 ausreichend. Zu Kontrollzwecken über den korrekten Programmfluss durchläuft der Lageparameter aber weiter den gesamten Abschnitt.

Die Analyse beschränkt sich auf das Niveau $\alpha = 0.05$. Dieses Niveau wird in der Praxis wohl am häufigsten benutzt. Sicherlich ließe sich mit einem anderen nominalen Niveau ähnlich verfahren. Doch der Rechenaufwand ist immens und der Zugewinn an neuer Erkenntnis eher marginal, wenn man bedenkt, dass sich das Niveau nicht auf jeden beliebigen Wert einstellen lässt. Als Beispiel soll die Normalverteilung in Abb. 3 dienen.

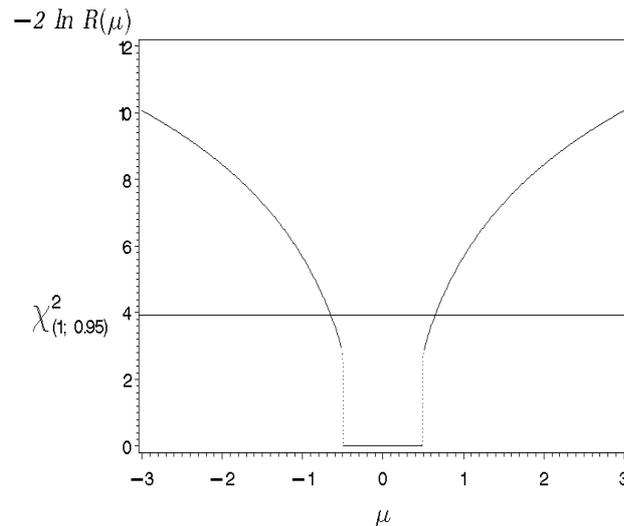


Abbildung 3: $-2 \ln R$ bei Normalverteilung für die Stichprobe $(0,0)$

Auch wenn der MLS selbst nicht eindeutig ist, so existiert dennoch das zugehörige Supremum der Profil-Likelihood. Es ist gleich 1, und damit reduziert sich der Quotient auf

$$\begin{aligned} -2 \ln R &= -2(\ln L^*(\mu) - \ln \sup_{\mu \in \mathbb{R}} L^*(\mu)) \\ &= -2 \ln L^*(\mu). \end{aligned}$$

Für $\mu \in [0, 0.5)$ ist

$$\begin{aligned} -2 \ln L^*(\mu) &= -2 \ln \sup_{\sigma > 0} \prod_{i=1}^n \left\{ \Phi \left(\frac{0.5 - \mu}{\sigma} \right) - \Phi \left(\frac{-0.5 - \mu}{\sigma} \right) \right\} \\ &= -2n \ln \lim_{\sigma \rightarrow 0} \left\{ \Phi \left(\frac{0.5 - \mu}{\sigma} \right) - \Phi \left(\frac{-0.5 - \mu}{\sigma} \right) \right\} \\ &= -2n \ln \left\{ \lim_{x \rightarrow \infty} \Phi(x) - \lim_{x \rightarrow -\infty} \Phi(x) \right\} \\ &= -2n \ln 1 = 0 \end{aligned}$$

und für alle $\mu \in (-0.5, 0]$ wegen $\Phi(x) = 1 - \Phi(-x)$

$$\begin{aligned} -2 \ln L^*(\mu) &= -2n \ln \lim_{\sigma \rightarrow 0} \left\{ 1 - \Phi\left(\frac{0.5 + \mu}{\sigma}\right) + \Phi\left(\frac{0.5 - \mu}{\sigma}\right) \right\} \\ &= -2n \ln \lim_{x \rightarrow \infty} \{1 - \Phi(x) + \Phi(x)\} = 0. \end{aligned}$$

An den Unstetigkeitsstellen $\mu = 0.5$ und $\mu = -0.5$ steigt der Likelihoodquotient abrupt auf

$$\begin{aligned} -2 \ln L^*(0.5) &= -2n \ln \sup_{\sigma > 0} \left\{ \Phi(0) - \Phi\left(-\frac{1}{\sigma}\right) \right\} \\ &= -2n \ln \frac{1}{2} = 2n \ln 2. \end{aligned}$$

Die Schnittstellen von $-2 \ln R$ mit der Geraden auf der Höhe des zugehörigen χ^2 -Quantils begrenzen das Konfidenzintervall. Innerhalb des „Unstetigkeitsbereiches“ kann die Schnittlinie beliebig ohne Auswirkung auf die Konfidenzintervalllänge verschoben werden. Das bedeutet, dass hier eine Regulierung des Niveaus nicht möglich ist. Für einen Stichprobenumfang von 2 ist $2n \ln 2 = 3.307$. Erst das $\chi^2_{(1;0.931)}$ -Quantil übersteigt diesen Wert und kann so eine echte Weitung der Konfidenzintervalle bewirken. Alle anderen $\alpha \leq 0.069$ laufen auf das gleiche Intervall $[-0.5, 0.5]$ hinaus. Liegt der wahre Parameter innerhalb dieses Bereiches, und σ ist sehr klein, werden kaum andere Stichproben als solche mit lauter Nullen erzeugt, und die Überdeckungswahrscheinlichkeit liegt bei 1. Daran ändert auch ein Absenken des nominalen Niveaus nichts. Erst mit der Erhöhung der Varianz, die für ein Auftreten anderer Stichprobenrealisationen sorgt, wird eine Niveausenkung überhaupt erst wirksam.

Die vorliegende Arbeit hat vorwiegend zum Ziel, die Überdeckungsqualitäten bei kleinen Stichprobenumfängen zu evaluieren. Bei größeren Stichprobenumfängen kann man eine gute Approximation des χ^2 -Verfahrens erwarten. In Anlehnung an die Studienergebnisse von Lee und Vardeman erscheint hierzu eine Betrachtung von Stichprobenumfängen bis einschließlich 15 als ausreichend.

Der Bereich für den Skalenparameter wird auf die jeweilige Verteilung abgestimmt. Die einzelnen Verteilungen zeigen unterschiedliches Streuverhalten bei Veränderungen des Skalenparameters, wodurch eine Synchronisation nicht sinnvoll ist.

Am Ende ergibt sich im Studienaufbau eine Variation der Parameter $\mu = (0, 0.1, 0.2, 0.3, \dots, 0.9)$, $n = (2, 3, 4, 5, 10, 15)$, σ und des Schwellenwertes r_0 für den Rundungsgrad. Anhand einer Versuchsreihe mit fixen Stichproben kristallisiert sich im Verlauf ein passender Schwellenwert heraus. Hier gibt es keine allgemeingültige Optimierungsklausel, sondern die Festlegung beruht eher auf Versuch und Irrtum. Der Bereich des mutmaßlichen Schwellenwertes lässt sich aber bei jeder Verteilung anhand spezifischer Kurvenverläufe abstecken. Dazu betrachtet man die Überdeckungswahrscheinlichkeiten von Pivot- und LQ-Verfahren in ihrem Verlauf über

σ . In einem klar umrissenen Bereich des Skalenparameters kreuzen sich die Kurvenverläufe. Für kleinere σ liegt die LQ-Kurve höher, danach holt die Pivot-Methode auf und erreicht schließlich die höheren Überdeckungswahrscheinlichkeiten. Dies ist der Bereich, in dem das Kombinationsverfahren auf die Pivotmethode umschwenken sollte. Testläufe mit verschiedenen Werten in der lokalen Umgebung sollen den Schwellenwert für den Skalenschätzer festlegen. Die genaue Spezifikation ist dabei ein Balanceakt zwischen Anhebung der Überdeckungswahrscheinlichkeiten einerseits und Beibehaltung kurzer Intervalllängen andererseits. Diese beiden Auflagen sind schwer in Einklang zu bringen, denn die Anhebung der Überdeckungswahrscheinlichkeiten zieht meist unweigerlich eine Verlängerung der Konfidenzintervalle nach sich, was jedoch Informationsverlust bedeutet. Wegen der kürzeren Intervalllänge der LQ-Methode sollte der Schwellenwert so gewählt werden, dass das Kombinationsverfahren so spät wie möglich (also für möglichst große σ) auf die Pivotmethode umschaltet. Trotzdem muss eine Niveaueinhaltung garantiert bleiben. Der Schwellenwert für den Schätzer des Skalenparameters lässt sich in eine Vorschrift für den Rundungsgrad überführen.

Sei σ_0 der Schwellenwert für den Skalenparameter, dann kann man die Bedingung

$$\text{wähle LQ, falls } \hat{\sigma} < \sigma_0$$

ersetzen durch

$$\text{wähle LQ, falls } r_0 = \frac{w}{\sigma_0} < \hat{r} = \frac{w}{\hat{\sigma}}.$$

Ein Simulationsprogramm erzeugt für jede Kombination von (n, μ, σ) 1000 Pseudozufallsstichproben (Details zur Erzeugung von Pseudozufallszahlen vgl. Abschnitt „Programmiertechnische Details“). Die generierten Zahlen werden auf ganze gerundet. Danach erfolgt zu jeder dieser Stichproben die Berechnung der Konfidenzintervalle mit allen drei Methoden (LQ-, Pivot-, und Kombinationsverfahren). Das Programm vermerkt die Anzahl der Treffer und die individuellen Intervalllängen und liefert schließlich über die Trefferquote eine Einschätzung der Überdeckungswahrscheinlichkeiten sowie der durchschnittlichen Intervalllängen. Individuelle Vergleiche können dann aufdecken, welches Verfahren sich in speziellen Parametersituationen besonders eignet.

4.4 Programmiertechnische Details

Die verwendeten Algorithmen sind über die Programmiersprache APL implementiert. Sie werden hier verbal oder in Ablaufdiagrammen skizziert. So kann auch ein mit APL nicht vertrauter Leser die Programmschritte nachvollziehen und eventuell

in einer anderen Programmiersprache umsetzen. Der Vorteil von APL liegt darin, dass diese Programmiersprache mit einer kurzen symbolischen Schreibweise ein flexibles Arbeiten mit Vektoren und Matrizen unterstützt. Einige grundlegende Operationen, etwa ein Generator von Pseudozufallszahlen, werden von APL bereitgestellt. Ihre Funktionsweise kann man in einem APL-Handbuch (APL*PLUS, 1992) nachlesen. Teilweise haben andere Benutzer nützliche Prozeduren für oft nachgefragte Problemstellungen zur Verfügung gestellt. Sie sind in einem Ergänzungsband (APL Idiom Library) öffentlich zugänglich. Diesem sind Programmteile wie die Erzeugung einer rechteckverteilten oder normalverteilten Zufallsvariablen und die Berechnung der Verteilungsfunktion der Standardnormalverteilung entnommen. Alle anderen Teilprozeduren sind speziell auf die Anforderungen dieser Simulationsstudie zugeschnitten und neu implementiert. Auf diese Weise kann man den korrekten Ablauf überwachen und Veränderungen bequem vornehmen.

Erzeugung von Zufallswerten

Zufallswerte aus einer beliebigen Verteilung kann man über die inverse Verteilungsfunktion generieren, denn es gilt: Ist X eine Zufallsvariable mit stetiger Verteilungsfunktion F , dann ist die Zufallsvariable $U = F(X)$ rechteckverteilt über dem Intervall $[0,1]$. Umgekehrt: Ist U eine rechteckverteilte Zufallsvariable über dem Intervall $[0,1]$ und F^{-1} die Umkehrfunktion von F , dann besitzt die Zufallsvariable $X = F^{-1}(U)$ die Verteilungsfunktion F .

Mit einem geeigneten Instrumentarium zur Erzeugung einer $[0,1]$ -rechteckverteilten ZV U entstehen die benötigten Zufallswerte einiger anderer Verteilungen nach der Vorschrift:

Exponentialverteilung ohne Verschiebung	:	$F_{EXP}^{-1}(\lambda, u) = -\lambda \ln(1 - u)$
Exponentialverteilung mit Verschiebung	:	$F_{EXP}^{-1}(\mu, \lambda, u) = \mu - \lambda \ln(1 - u)$
Logistische Verteilung	:	$F_{LOG}^{-1}(a, b, u) = a + b \ln\left(\frac{1-u}{u}\right)$
Maximumverteilung	:	$F_{MAX}^{-1}(a, b, u) = a - b \ln(-\ln u)$
Weibull-Verteilung	:	$F_{WEI}^{-1}(a, b, u) = a [-\ln(1 - u)]^{\frac{1}{b}}$

Transformationen bieten eine weitere Möglichkeit, Zufallswerte zu erzeugen:

Ist X normalverteilt, dann ist e^X lognormalverteilt.

Ist X EXP(1)-verteilt, dann ist $e^{X/a}$ Pareto-verteilt mit Parameter a .

Ursprünglich war geplant, ein engmaschiges Gitter über den Parameterbereichen anzulegen. Der Algorithmus sollte sämtliche Lage- und Skalenparameter paarwei-

se kombinieren und für jede Kombination einen Likelihoodwert berechnen. Sowohl globales Maximum, wie auch die lokalen Skalenschätzer zu singulärem fixem Lageparameter wären schnell mit primitiven Vergleichsoperationen zu finden gewesen. Ein Vielfaches der Stichprobenvarianz sollte rechten und linken Endpunkt der Skalenparameterachse programmintern vorgeben. Der Anwender sollte dann über die Größe der Gittermaschen extern wählen können, auf wie viele Stellen genau die Konfidenzgrenze berechnet werden soll. Dieses Vielfache der Stichprobenvarianz zahlenmäßig festzulegen, erwies sich aber als aussichtsloses Unterfangen. Auch nach einer flexibleren Gestaltung über eine weitere Staffelung multiplikativer Faktoren nach der Stichprobenvarianz entstand kein funktionstüchtiges Gittersystem. War das Gitter zu breit bemessen, brach das Programm mit Fehlermeldung ab, weil die Likelihoodwerte an den Gitterrändern zu klein waren. So klein, dass sie sich intern nicht von Null unterschieden und deshalb nicht logarithmiert werden konnten. War das Gitter zu klein ausgelegt, erfasste man die echten Konfidenzgrenzen nicht, und das Programm gab fälschlicherweise rechten und linken Endpunkt aus.

Eine verfeinerte Variante ohne a priori Fixierung der Endpunkte entwickelt das Gitter im Programmverlauf und baut das Konfidenzintervall aus dem Inneren heraus zu den Grenzen hin auf. Dieses Programm muss mehrere Module ineinander verschachteln. Der aus der Informatik bekannte „Von-oben-nach-unten“-Entwurf separiert die einzelnen Etappen und verdeutlicht gleichzeitig ihr Zusammenwirken. Zunächst beschreibt ein Grobentwurf die wichtigsten Arbeitsgänge. Sie werden dabei als „Black Boxes“ behandelt, von der zwar das Endprodukt, nicht aber ihre Arbeitsweise bekannt ist. Weitere Verfeinerungen erhellen das „Innenleben“ einer jeden Black Box und präzisieren die einzelnen Arbeitsschritte.

Algorithmus zur Berechnung des LQ-Konfidenzintervalles für die Lage

Ein Grobentwurf zur Berechnung des LQ-Konfidenzintervalles besteht aus den drei Schritten:

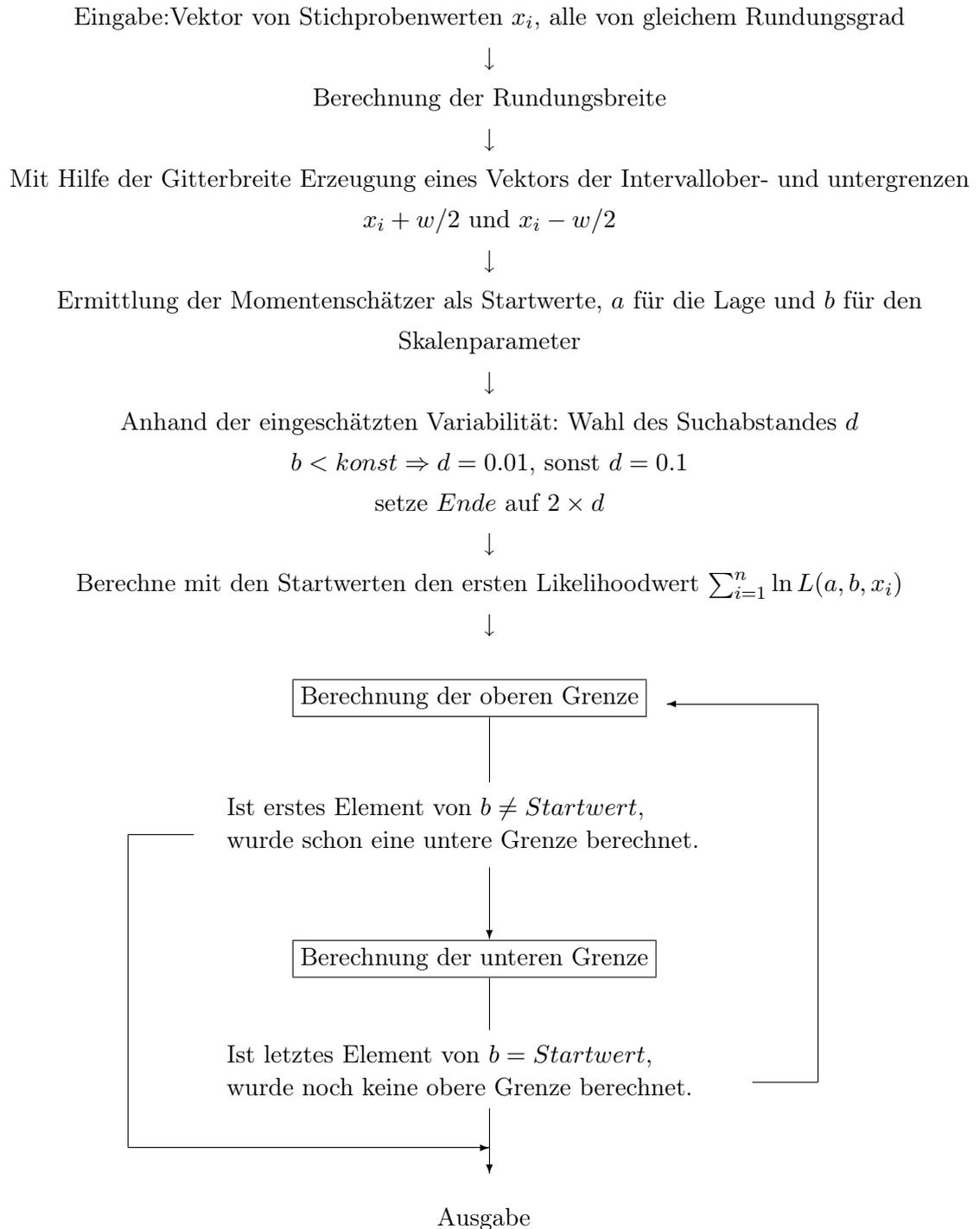
Eingabe: Vektor von Stichprobenwerten mit gleichem Rundungsgrad

Ausgabe: MLS für Lage- und Skalenparameter, sowie KI für die Lage

1. Bestimmung von Startwerten und Gitterfeinheit
2. Berechnung der oberen Konfidenzgrenze
3. Berechnung der unteren Konfidenzgrenze.

Damit das Programm die relative Likelihood nach beiden Seiten korrekt berechnet, muss zuerst die Richtung durchlaufen werden, in der sich das globale Maximum der

Abbildung 4: Algorithmus zur Berechnung des LQ-Konfidenzintervalles für die Lage

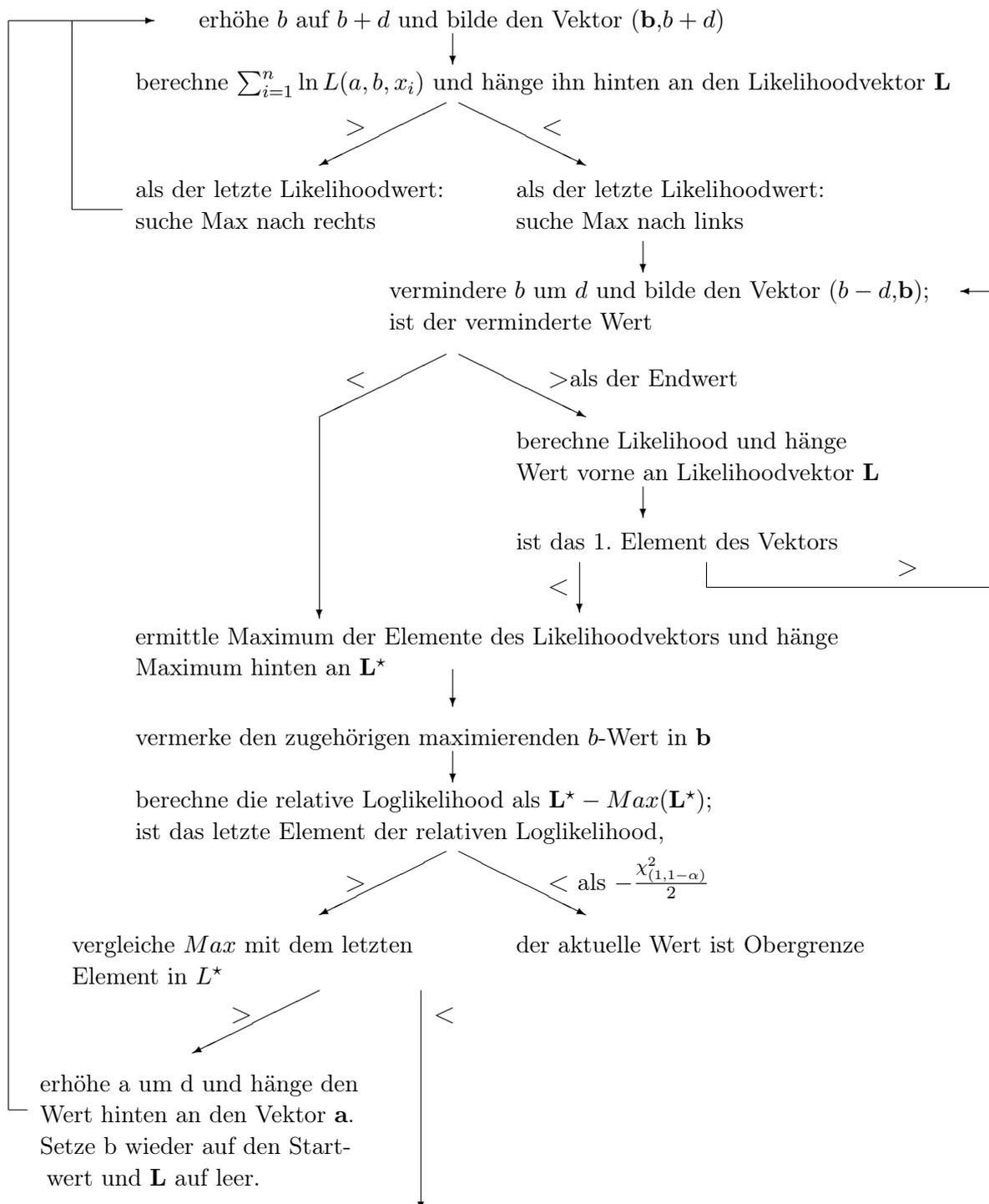


Likelihood befindet. Im Detail betrachtet, kommt man zu folgender Verfeinerung der Arbeitsgänge:

1. **a.** Bestimmung der Startwerte:
Verwende die Momentenschätzer für den Lageparameter als Startwert für a und den Momentenschätzer für den Skalenparameter als Startwert für b .
- b.** Bestimmung der Gitterfeinheit:
Zur Vermeidung überlanger Rechenzeiten kann man sich mit der Gitterfeinheit nach dem Startwert für b richten. Je höher die Varianz, desto länger die Intervalle und somit die Rechenzeiten. Ein Verzicht auf Nachkommastellen höherer Ordnung fällt prozentual zur Intervalllänge nicht ins Gewicht und beschleunigt den Rechenvorgang. Wie rasch die Intervalllänge wächst, hängt vom Verteilungstyp ab und muss individuell jedem Typ angepasst werden. Generell gilt:
 $b < \text{Konstante}$, wähle Gitterbreite 0.01, sonst genügt 0.1.
2. Berechnung der oberen Konfidenzgrenze:
Beginne mit dem Startwert a und
 - a.** berechne den Wert der Profil-Likelihood L^* an dieser Stelle. Dazu beginne mit Startwert b und berechne sukzessive die Likelihoodwerte. Suche in der Richtung, in der die Likelihoodwerte ansteigen. Die Suche kann abgebrochen werden, wenn die Likelihoodwerte wieder sinken. Vermerke diesen Wert im Vektor L^* und den maximierenden Wert \tilde{b} im Vektor \mathbf{b} .
 - b.** Berechne die aktuelle logarithmierte relative Likelihood. Solange der neu hinzugefügte Wert größer ist als $-\chi_{(1,1-\alpha)}^2/2$ ist die Grenze nicht erreicht. Dann muss a um die Gitterfeinheit erhöht und die Schritte **a.** und **b.** müssen wiederholt werden.
3. Berechnung der oberen Konfidenzgrenze:
Analog wie **2.**, verwendet dieselben Module **a** und **b**. Allerdings wird b sukzessive erniedrigt solange die logarithmierte relative Likelihood größer ist als $-\chi_{(1,1-\alpha)}^2/2$.

Der endgültige Algorithmus ist in einem Flussdiagramm (Abb. 4) dargestellt, zunächst noch unter Verwendung einer Black Box für die Berechnung der Grenze. Ihre Arbeitsweise wird im nachfolgenden Diagramm (Abb. 5) entschlüsselt. Die Likelihood bleibt in dieser allgemeinen Darstellung unspezifiziert. Je nach Modell kann der Algorithmus ein approximatives KI für exakte wie für gerundete Daten eines

Abbildung 5: Berechnung der oberen Grenze



Max liegt unterhalb des Startwertes, daher muss zuerst in Richtung der unteren Intervallgrenze gesucht werden.

beliebigen Verteilungstyps ausgeben. Bei exakten Daten entfällt die Berechnung der Rundungsbreite sowie der Rundungsintervallgrenzen $x_i - w/2$, $x_i + w/2$. Für ein Konfidenzintervall zum Skalenparameter braucht man nur a und b zu vertauschen. Darüber hinaus muss man über eine Abbruchbedingung sicherstellen, dass b keine negativen Werte annimmt. Die Voreinstellung für die Likelihood $\sum_{i=1}^n \ln L(a, b, x_i)$ ist 0, der Likelihoodvektor \mathbf{L} sowie \mathbf{L}^* leere Vektoren.

Dieser Algorithmus baut auf dem Fundament guter Startwerte. Sie dürfen keinesfalls außerhalb des zu berechnenden Konfidenzintervalls liegen, sonst stoppt dieser Algorithmus nicht! Allerdings: Für die im nächsten Kapitel untersuchten Verteilungen sind die Momentenschätzer so gut, dass in allen Simulationsläufen nicht ein einziges Mal wegen dieses Fehlers das Programm abbrach.

5 Kombinierte Intervalle bei stetigen Verteilungsalternativen

5.1 Konfidenzintervalle für die Maximumverteilung

Die Extremwerttheorie stellt ein nützliches Instrumentarium für eine variantenreiche Mischung verschiedenster Anwendungsbereiche dar. In der Qualitätskontrolle (z.B. bei Tests zur Materialstärke und der Zuverlässigkeitsanalyse) ist sie ebenso anzutreffen wie in der Klimaforschung zur Vorhersage extremer Klimaereignisse wie Starkregen, Orkanstärken, Fluten oder Dürreperioden. Der Begriff „Extremwertverteilung“ ist darauf zurückzuführen, dass diese Verteilung u.a. als asymptotische Verteilung der größten Ordnungsstatistik für $n \rightarrow \infty$ resultiert, wobei n unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsfunktion F unterstellt werden. Die Grenzwertbetrachtungen zeigten, dass bei wachsendem Stichprobenumfang die asymptotische Verteilung der größten Ordnungsstatistik in genau eine von drei Verteilungskategorien fällt. Gnedenko (1943) fand schließlich notwendige und hinreichende Bedingungen für die schwache Konvergenz, die an den Typ der zugrundeliegenden Verteilung F geknüpft sind. In diesem Abschnitt wird die Maximumverteilung vom Typ I, auch Gumbel-Verteilung genannt, untersucht. Die Verteilungsfunktion hat die Gestalt

$$F(x) = \exp \left[-e^{-\left(\frac{x-u}{b}\right)} \right], \quad -\infty < u < \infty, \quad b > 0. \quad (5.1)$$

Erwartungswert und Varianz sind

$$E(X) \approx u + 0.577 \times b \quad (5.2)$$

$$Var(X) = \frac{1}{6} \pi^2 b^2. \quad (5.3)$$

Beide Momente sind über einfache funktionale Formen mit Lage- und Skalenparameter verknüpft. Die Momentenschätzer lassen sich so leicht berechnen. Sie werden im iterativen Algorithmus zur Bestimmung des MLS als Startwerte dienen.

Der MLS für exakte Daten maximiert die zugehörige Likelihood der Stichprobe $x = (x_1, \dots, x_n)$:

$$L(u, b) = \prod_{i=1}^n \frac{1}{b} e^{-(x_i-u)/b} \exp \left[-e^{-\left(\frac{x_i-u}{b}\right)} \right]. \quad (5.4)$$

Bei gerundeten Daten ändert sich die Likelihood zu

$$L(u, b) = \prod_{i=1}^n \{F(x_i + 0.5) - F(x_i - 0.5)\}, \quad x_i \in \mathbb{N}, \quad (5.5)$$

$$= \prod_{i=1}^n \exp \left[-e^{-\left(\frac{x_i+0.5-u}{b}\right)} \right] - \exp \left[-e^{-\left(\frac{x_i-0.5-u}{b}\right)} \right]. \quad (5.6)$$

Bestimmung des Korrekturintervalls bei Spannweite 0:

Eine Stichprobe von lauter Null-Werten kann auf zwei mögliche Arten entstehen: Entweder der Lageparameter befindet sich in nächster Nähe von Null und die Varianz ist sehr klein, oder aber der Lageparameter ist in Wirklichkeit weiter entfernt und eine leicht erhöhte Variabilität ist für diese Stichprobe verantwortlich. Im zweiten Fall wird es erfahrungsgemäß für beide Verfahren schwierig, das Konfidenzintervall treffend anzugeben. Bei kleinen Stichprobenumfängen sind solche Stichproben leider häufig vorzufinden, häufiger als es für die Niveaueinhaltung zumutbar ist. Hier muss man mit Korrekturmaßnahmen einschreiten. Dazu wird geprüft, unter welchen Varianzvorgaben (Skalenparameter) der Prozentsatz solcher Stichproben unter dem Niveau verharret. Die gewünschte Auskunft gibt die Profil-Likelihood für gerundete Daten:

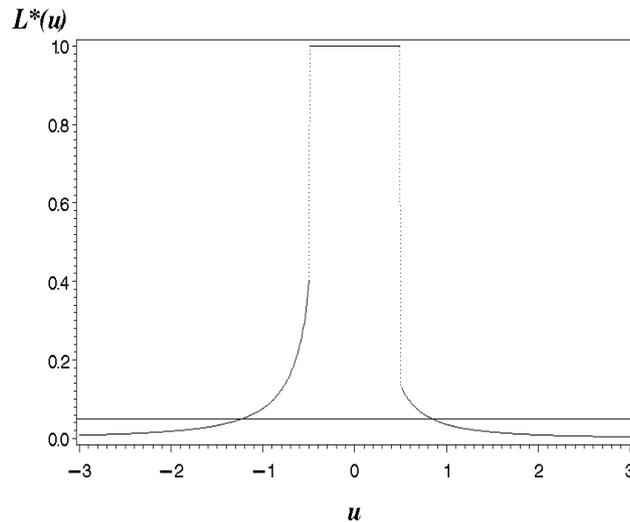


Abbildung 6: Profil-Likelihood der Maximumverteilung für die Stichprobe (0,0)

$$L^*(u | x_1 = x_2 = 0) = \sup_{b>0} \left\{ \exp \left[-e^{-\left(\frac{0.5-u}{b}\right)} \right] - \exp \left[-e^{-\left(\frac{-0.5-u}{b}\right)} \right] \right\}^2. \quad (5.7)$$

Als neue Intervallgrenzen sollen nun diejenigen Werte für u dienen, bei denen die Profil-Likelihood erstmalig die Schranke von 0.05 nach unten durchbricht. Wegen

$$\lim_{u \rightarrow \infty} L^*(u) = [e^0 - e^0] = 0 \quad (5.8)$$

steht die Existenz der Schnittstellen fest. Mit einem Maximierungsalgorithmus lässt sich die Profil-Likelihood berechnen und die Liste von Korrekturintervallen (Tab. 1) zu kleinen Stichprobenumfängen und Niveau $\alpha = 0.05$ erstellen:

n	Δ_1	Δ_2
2	-1.23	0.84
3	-0.77	0.5
4	-0.62	0.5
5	-0.55	0.5

Tabelle 1: Korrekturintervalle für die Maximumverteilung

Mit wachsendem n nähert sich die Profil-Likelihood ausgehend von den Unstetigkeitsstellen bis zu den Rändern des Definitionsbereiches immer mehr der Abszisse. Ab $n = 10$ entfallen somit die Korrekturen. Damit kann man zum zweiten Schritt des Kombinationsprinzips übergehen.

Bestimmung des Schwellenwertes

Eine optische Gegenüberstellung der Überdeckungswahrscheinlichkeiten (ÜW) von LQ- und Pivotverfahren (Grafiken 7-10) bietet einen ersten groben Richtwert. Sofort fällt ins Auge, dass sich die Kurven über alle u hinweg im selben Streckenabschnitt von b kreuzen. Dieser Kreuzungspunkt liegt etwa bei 1.4. Es scheint also angemessen, für $\hat{b} \leq 1.4$ die LQ-Methode zu verwenden. Selbstverständlich sind auch hier zur Einschätzung von b verschiedene Alternativen denkbar. Um Doppelberechnungen zu umgehen, kann man in der ersten Stufe die Daten als exakt behandeln und durch die Programmmodule für exakte Daten lenken. Gibt der Skalenschätzer grünes Licht für das Pivot-Verfahren, können die Schätzer gleich in das Pivot-Konfidenzintervall einfließen, und die Rechenzeit war nicht nutzlos investiert. Nur für kleine Skalenschätzer muss die Prozedur für gerundete Daten aktiviert werden. Unter der Annahme, dass Stichproben mit kleinen Varianzen in der Praxis eher die Ausnahme darstellen, spart diese Vorgehensweise überflüssige Rechenschritte.

Im praktischen Test erweist sich der Wert 1.4 als zu hoch angesetzt. Er wird in Folgeläufen schrittweise gesenkt. Ein Schwellenwert von $b_0 = 0.3$ bewährt sich am besten. Der erprobte Schwellenwert für den Skalenschätzer lässt sich unmittelbar in eine Bedingung für den Rundungsgrad übersetzen:

$$\begin{aligned}
 & \hat{b} \leq b_0 = 0.3 \\
 \Rightarrow & \hat{r} \leq r_0 = \frac{w}{\sigma_0} \\
 \Leftrightarrow & \sqrt{\frac{6}{\pi^2 b_0^2}} \leq \hat{r} \\
 \Leftrightarrow & \hat{r} \geq 2.6 \quad .
 \end{aligned} \tag{5.9}$$

Die Quantile der Pivotgröße

Viele Autoren haben sich damit beschäftigt, Quantile von ML-Pivotgrößen zu tabellieren. Wegen der engen Verbindung zur Weibullverteilung lassen sich aus den Weibull-Quantiltabellen sowohl die Quantile der Minimumverteilung wie auch der Maximumverteilung ableiten, denn es gelten folgende Beziehungen:

Ist X minimumverteilt $MIN(u, b)$, so ist die ZV $W = \exp(X)$ Weibull-verteilt $W = \exp(Y) \sim WEI(e^u, 1/b)$. Ist Y maximumverteilt $MAX(u, b)$, dann ist die Transformation $W = \exp(Y) \sim WEI(e^{-u}, 1/b)$.

Quantile der Weibullverteilung sind für unterschiedlichste Modellversionen hinterlegt. Thoman et al. (1970) beispielsweise simulieren Quantile für unzensierte Daten und große Stichprobenumfänge und Billman et al. (1972) für zensierte Daten. Quantile für sehr kleine Stichprobenumfänge, die hier gefragt sind, waren darunter nicht aufzufinden. Zudem sind alle Textquellen älteren Datums, zu deren Zeit die Rechenleistung der Computer bei weitem nicht dem heutigen Standard entsprach. In der vorliegenden Arbeit wurden deshalb die notwendigen Quantile der Pivotgröße neu erzeugt. Aus je 10000 Pseudozufallsstichproben wurden Schätzer und Pivotgröße berechnet und diese Werte der Größe nach geordnet, woraus sich die in Tabelle 2 enthaltenen Quantile ergaben.

n	$q_{0.975}$	$q_{0.025}$
2	25.239	-10.123
3	4.235	-2.923
4	2.413	-1.967
5	1.782	-1.505
10	0.814	-0.816
15	0.626	-0.619

Tabelle 2: Quantile der Maximumverteilung für $(\hat{u} - u)/\hat{b}$

Ergebnisse der Simulationsstudie

Die Überdeckungswahrscheinlichkeiten der drei Methoden sind in den Grafiken 11-16 abgebildet. Die gepunktete Linie stellt die ÜW des Pivot-Verfahrens dar. Die gestrichelte Linie markiert die ÜW des LQ-Verfahrens für gerundete Daten. Die durchgezogene Kurve schließlich bezieht sich auf das Kombinationsverfahren.

In den Verlaufskurven zur Bestimmung des Schwellenwertes sind schon die ersten Ergebnisse vorweggenommen: Nur wenn der Lageparameter auf eine ganze Zahl fällt, kann das Pivotverfahren in den unteren Skalenbereichen den Parameter eingrenzen. Liegt der Parameter irgendwo in den Zwischenräumen des Rundungsgitters, ist er

mit dem Pivot-Verfahren kaum auffindbar. Eine Erhöhung des Skalenparameters wirkt sich positiv auf die Trefferquote aus und das Niveau wird für große b eingehalten. Je höher der Stichprobenumfang, desto schneller wird das Niveau erreicht. Das LQ-Verfahren dagegen schätzt den Lageparameter bei kleinem b fast immer im richtigen Segment ein. Die Qualität lässt aber zunehmend nach, wenn der Skalenparameter steigt. Die Überdeckungswahrscheinlichkeit pegelt sich für $n = 2$ auf 85% ein und liegt damit ca. 10% unter dem nominalen Niveau. Selbst bei $n = 10$ sind noch minimale Verluste zu verzeichnen. Erst ab $n = 15$ kann man von einer durchgängigen Niveaueinhaltung sprechen, die unabhängig vom zu Grunde gelegten Skalenparameter ist. Das Kombinationsverfahren greift regulierend ein und gleicht die Schwachpunkte beider Verfahren aus. Es zeigt einen geglätteten Verlauf und sorgt für eine gleichmäßig präzise Niveaueinhaltung. Besonders interessant ist, dass die ÜW gerade im Schnittpunkt von Pivot und LQ-Kurve die Quote der anderen deutlich übertrifft. Obwohl die einzelnen ÜW nur bei knapp über 80% liegen ($n = 2$), schafft das Kombinationsverfahren die Anhebung auf Nominalniveau. Und dies gelingt einzig und allein durch eine geschickte Auswahl des jeweils passenden Verfahrens.

Ergänzend geben die Kurven im Verlauf über u Auskunft darüber, wie groß die Aussichten sind, einen Lageparameter in den Gitterzwischenräumen zu erfassen. Gerade das Pivot-Verfahren lässt hier eine hohe Abhängigkeit von der exakten Position des Lageparameters erkennen. Neben einer guten Überdeckungswahrscheinlichkeit an den Gittermarkierungen steigt die Trefferquote bei einer mittigen Lage von u vorübergehend noch einmal an. In dieser Situation bringt allein die Position des Lageparameters eine erhöhte Variabilität ins Spiel. Im Grunde winzige Abweichungen werden durch das Runden übertrieben, so dass die Werte mal auf die ganze Zahl oberhalb, mal unterhalb springen. Die vergrößerte Variabilität weitet die KI und erzeugt dadurch unsystematische „Glückstreffer“. Mit zunehmendem Stichprobenumfang erhebt sich die Spitze in der Mitte immer höher, obwohl doch die Intervalllängen immer kürzer werden. Das rührt von der Tatsache her, dass reine „Spannweite-0“-Stichproben mit Intervalllänge 0 und daher auch keinem Treffer äußerst rar werden und somit die ÜW nicht weiter beeinträchtigen.

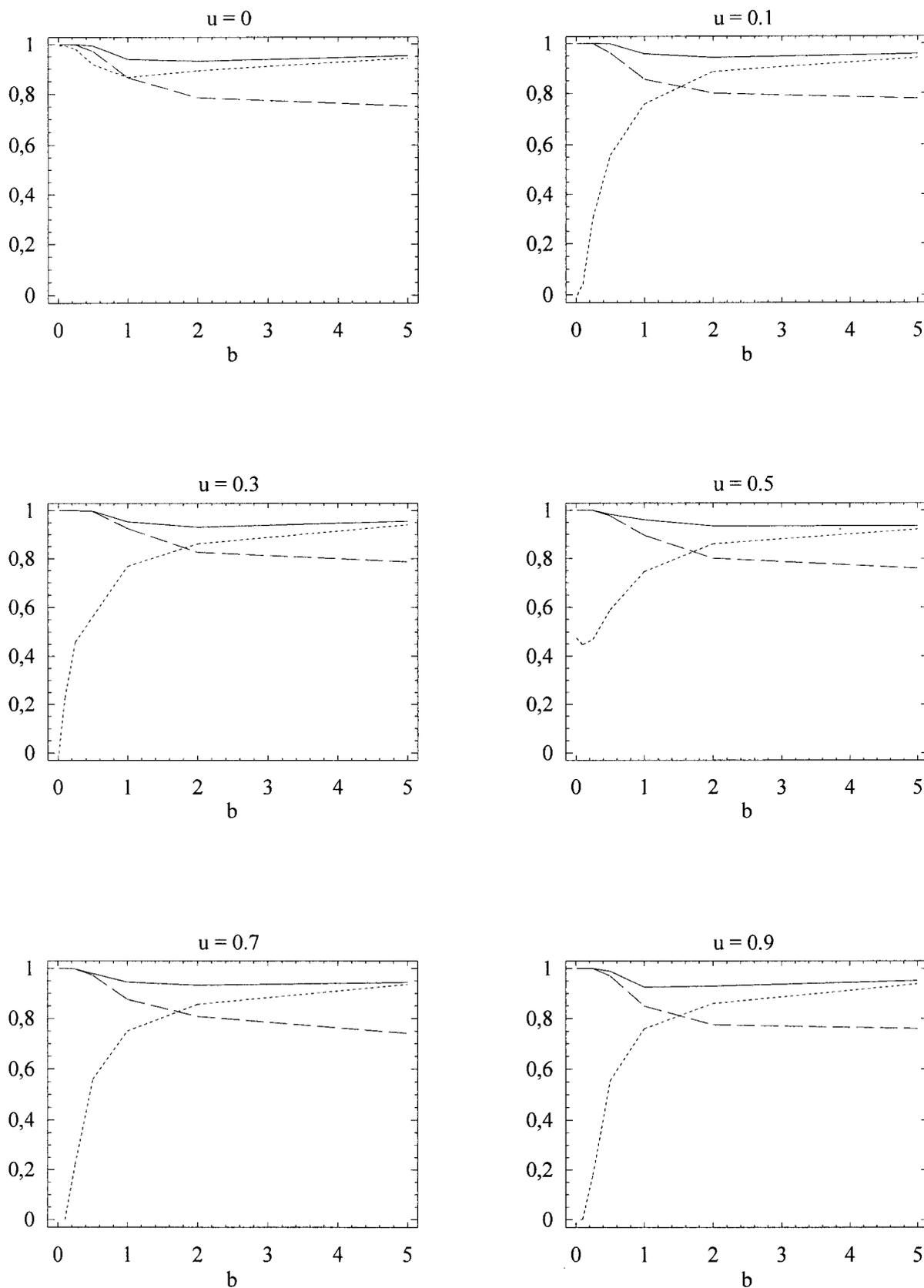
Die ÜW sind nicht symmetrisch um die Gittermitte, besonders in moderaten Skalenbereichen. Dies dürfte auf die Position des Erwartungswertes der Maximumverteilung zurückzuführen sein, in den Lage- und auch Skalenparameter einfließen. Die Werte werden tendenziell eher nach oben verschoben. Zufallstreffer kommen schon unterhalb der Mitte zustande, wohingegen dicht unterhalb von 1 wieder gehäuft „Spannweite-0“-Stichproben entstehen. Das schlägt sich in einer deutlich verminderten ÜW nieder. Ansonsten unterstreichen die Grafiken im Verlauf über u alle

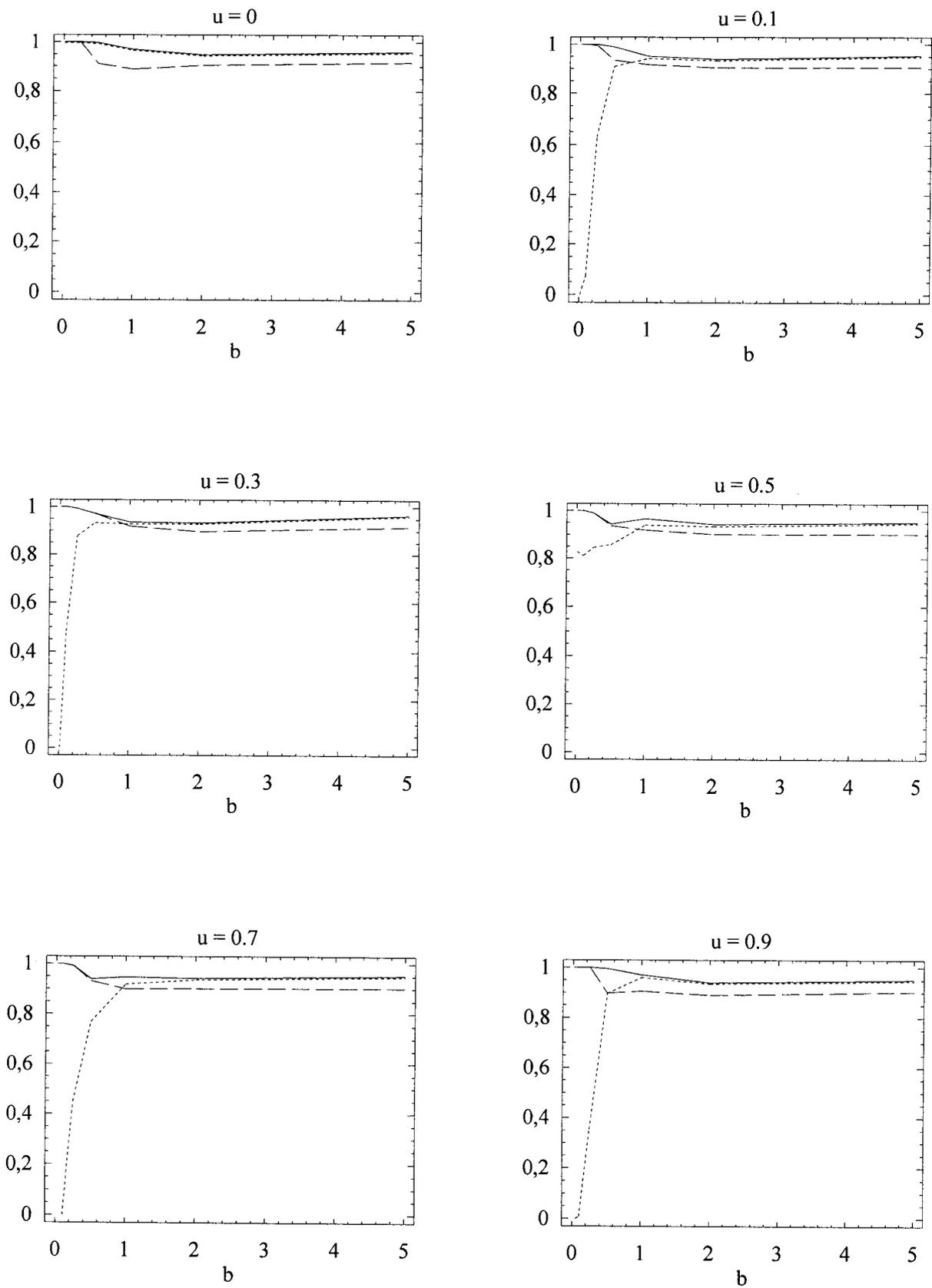
bisherigen Aussagen.

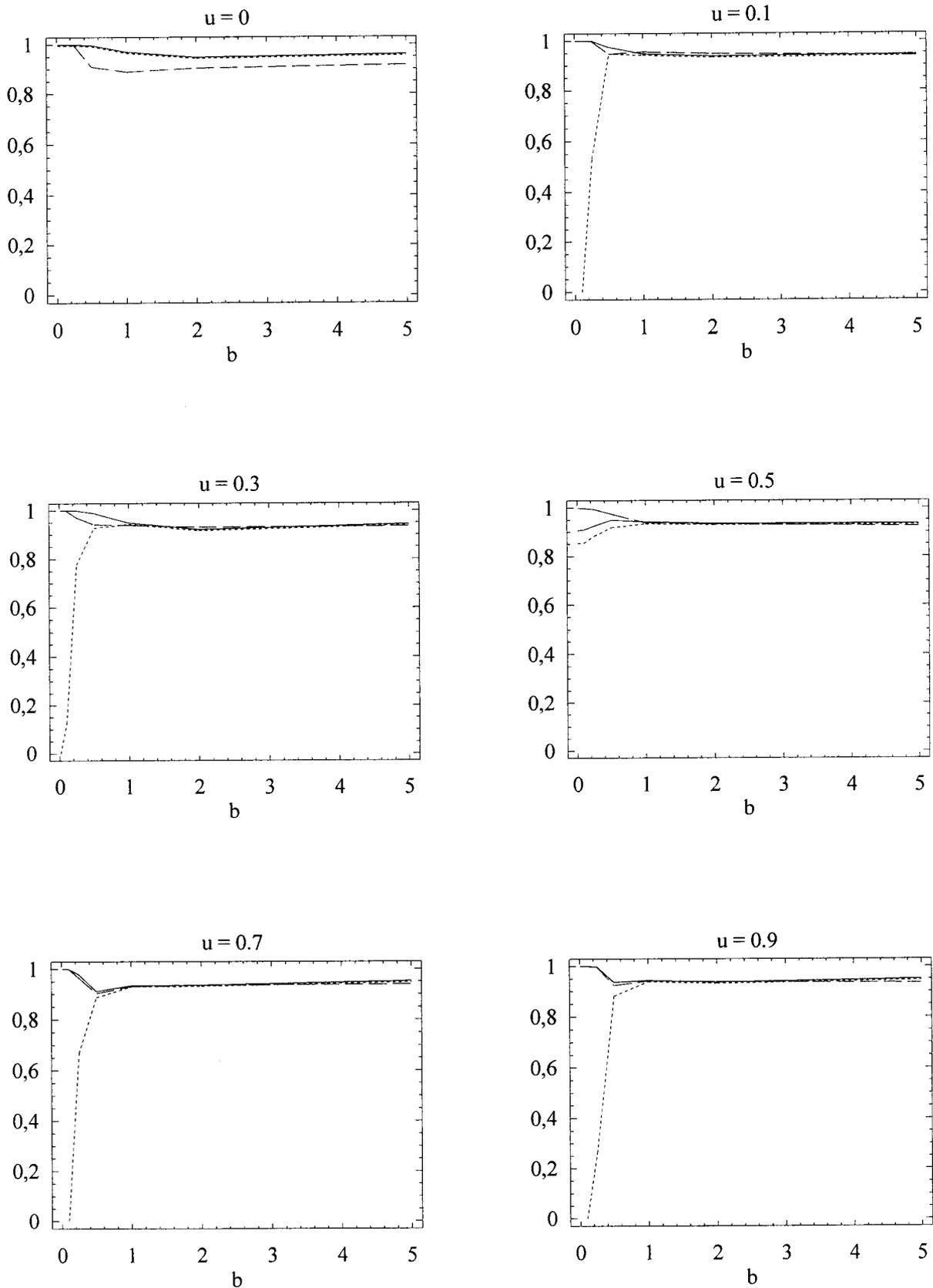
In Tabelle 3 sind die korrespondierenden Intervalllängen aufgeführt: Bei kleinem Skalenparameter behält das Kombinationsverfahren die vorteilhaft kurze Intervalllänge des LQ bei, steigt dann aber rasch an. Das ist einleuchtend, da das Kombinationsverfahren zunehmend auf Pivot umschaltet. Damit werden in moderaten b -Bereichen sogar höhere Intervalllängen als durch das Pivotverfahren erzielt. Wenn man sich vergegenwärtigt, dass es sich um Durchschnittslängen handelt, ist das Ergebnis aber durchaus zufrieden stellend. Besitzt doch das Pivotverfahren nur unwesentlich kürzere durchschnittliche Längen, verschätzt sich aber meistens in der Lage der Parameter.

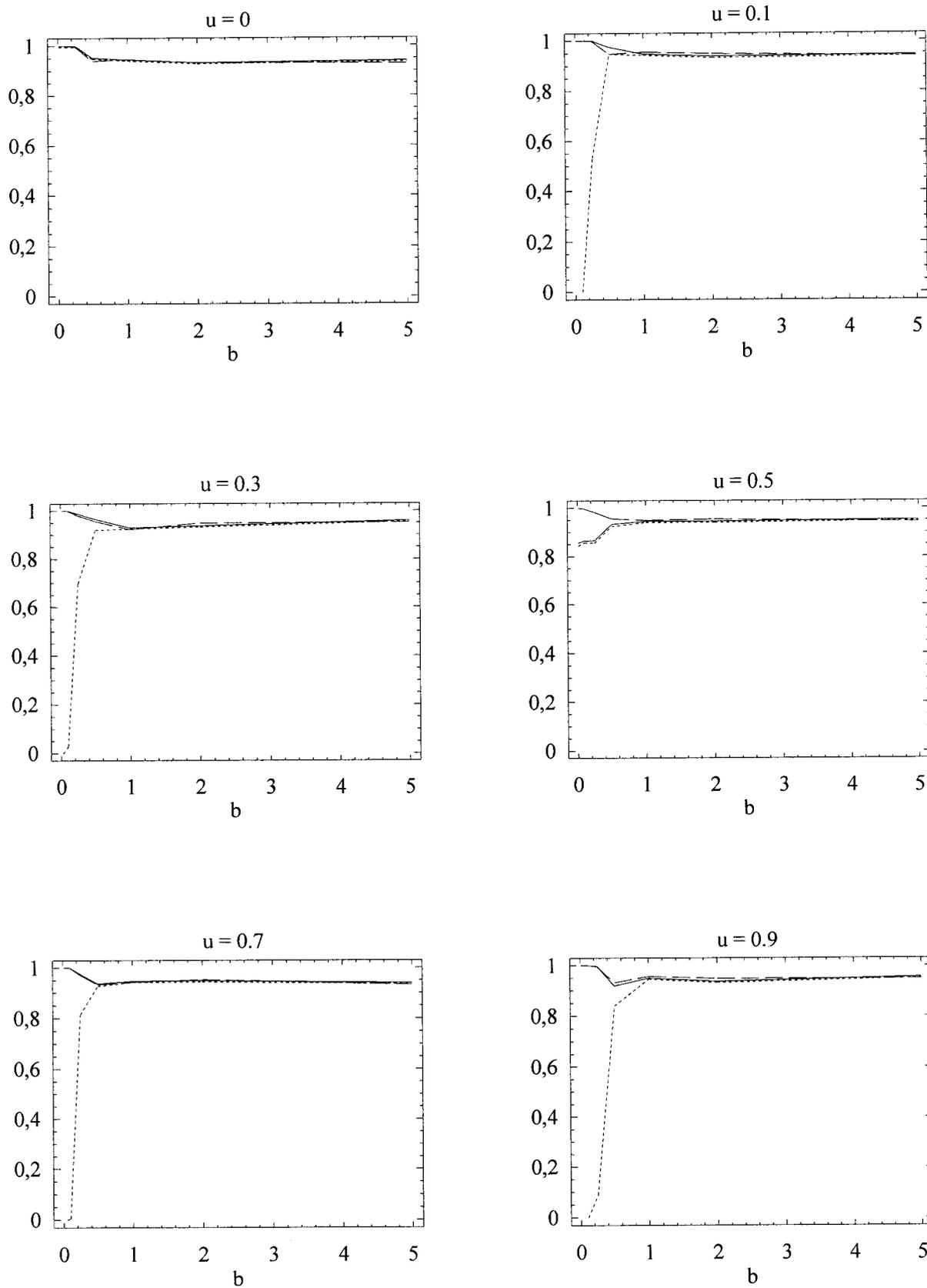
Empfohlene Kombination

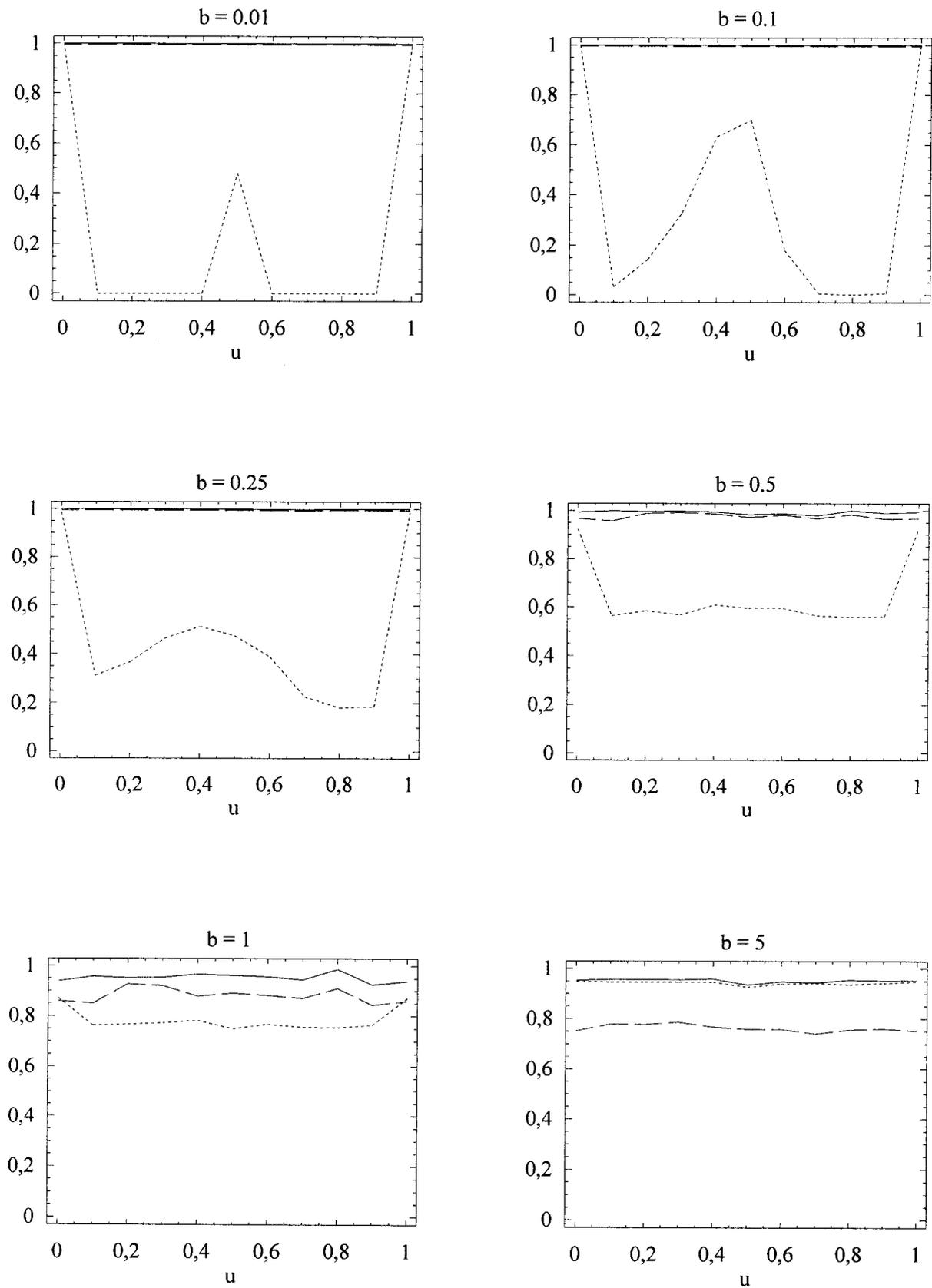
Für $\hat{r} \geq 2.6$ verwende das Pivotverfahren, ansonsten ist das Runden nicht vernachlässigbar und das LQ-Verfahren für gerundete Daten sollte angewendet werden. Ab $n = 15$ ist das LQ-Verfahren uneingeschränkt zu empfehlen. Es besitzt unter allen Parametereinstellungen die besten Überdeckungswahrscheinlichkeiten bei kurzen Intervalllängen. Zudem erübrigt sich das Nachschlagen der Quantile für die Pivotgröße.

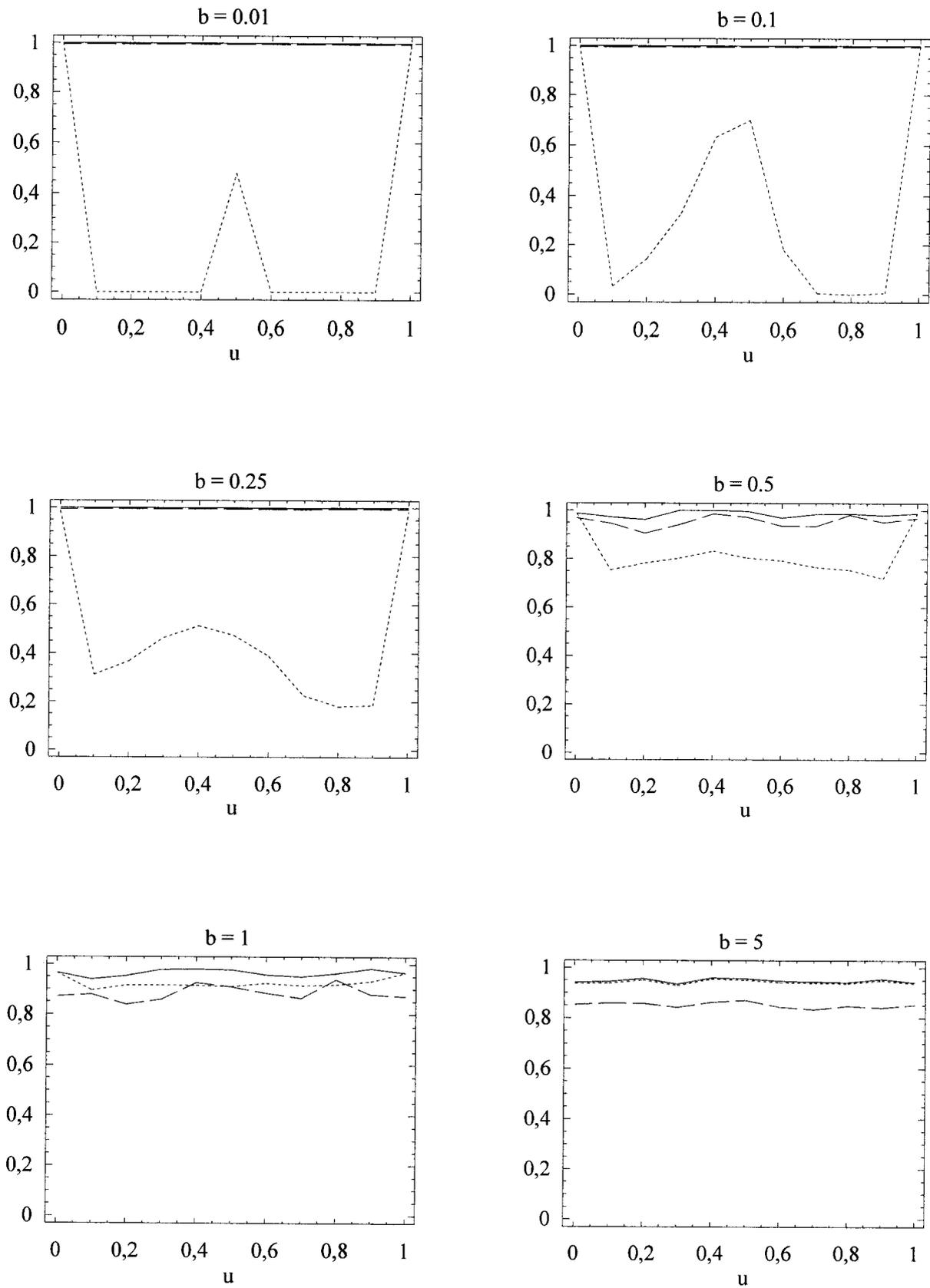
Überdeckungswahrscheinlichkeiten im Verlauf über b Abbildung 7: ÜW Maximumverteilung, $n=2$, Verlauf über b

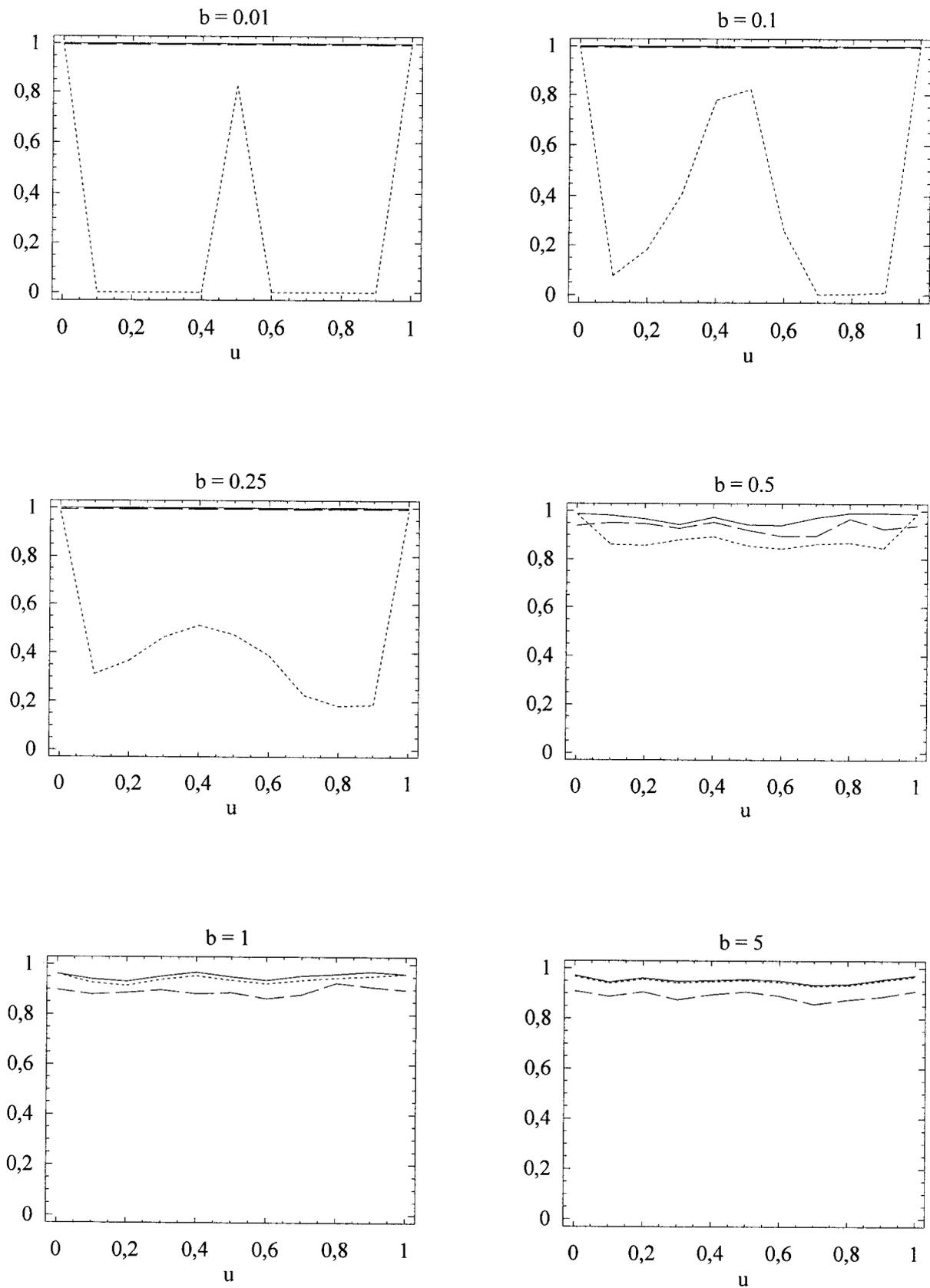
Abbildung 8: ÜW Maximumverteilung, $n=5$, Verlauf über b

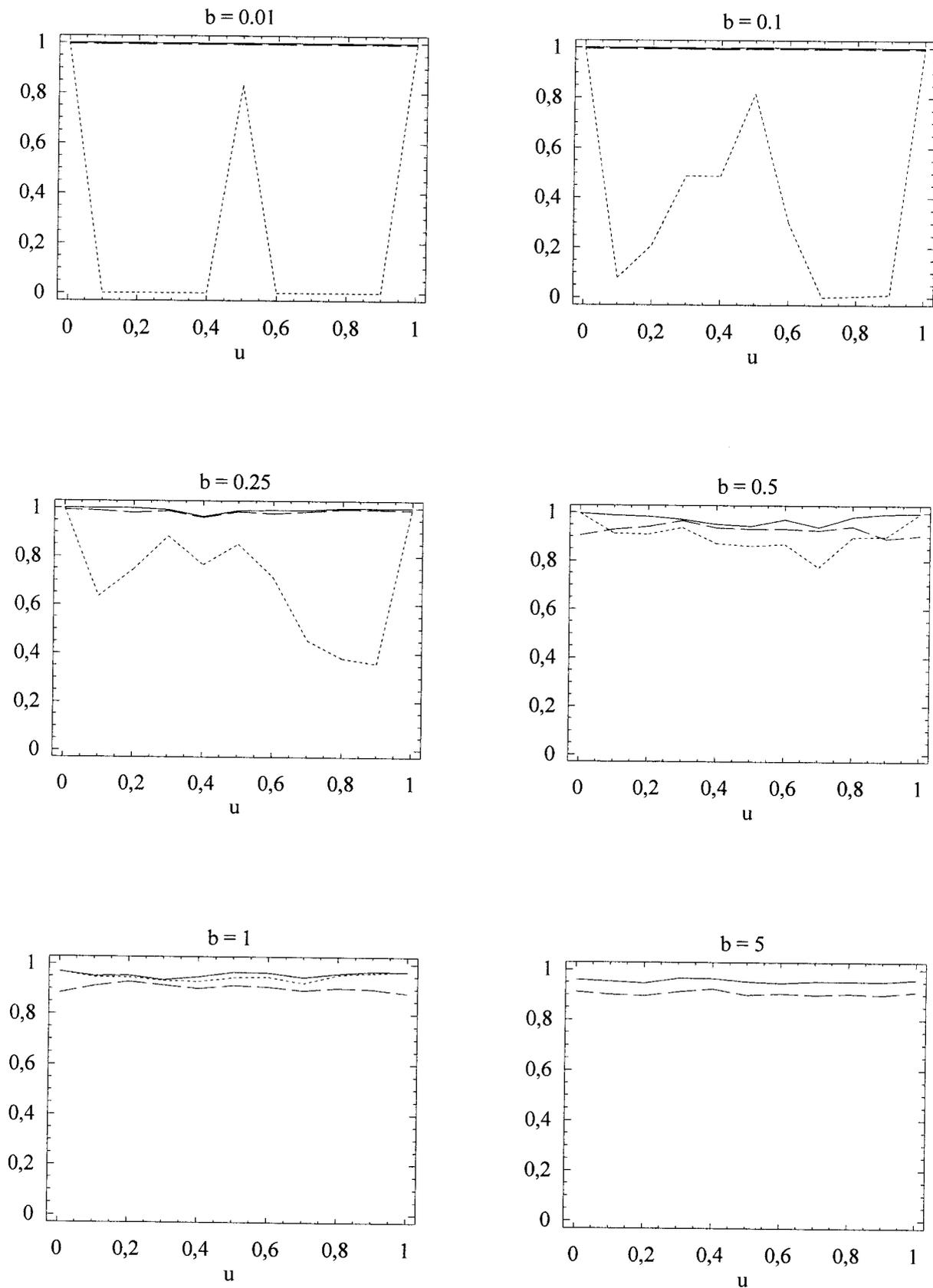
Abbildung 9: ÜW Maximumverteilung, $n=10$, Verlauf über b

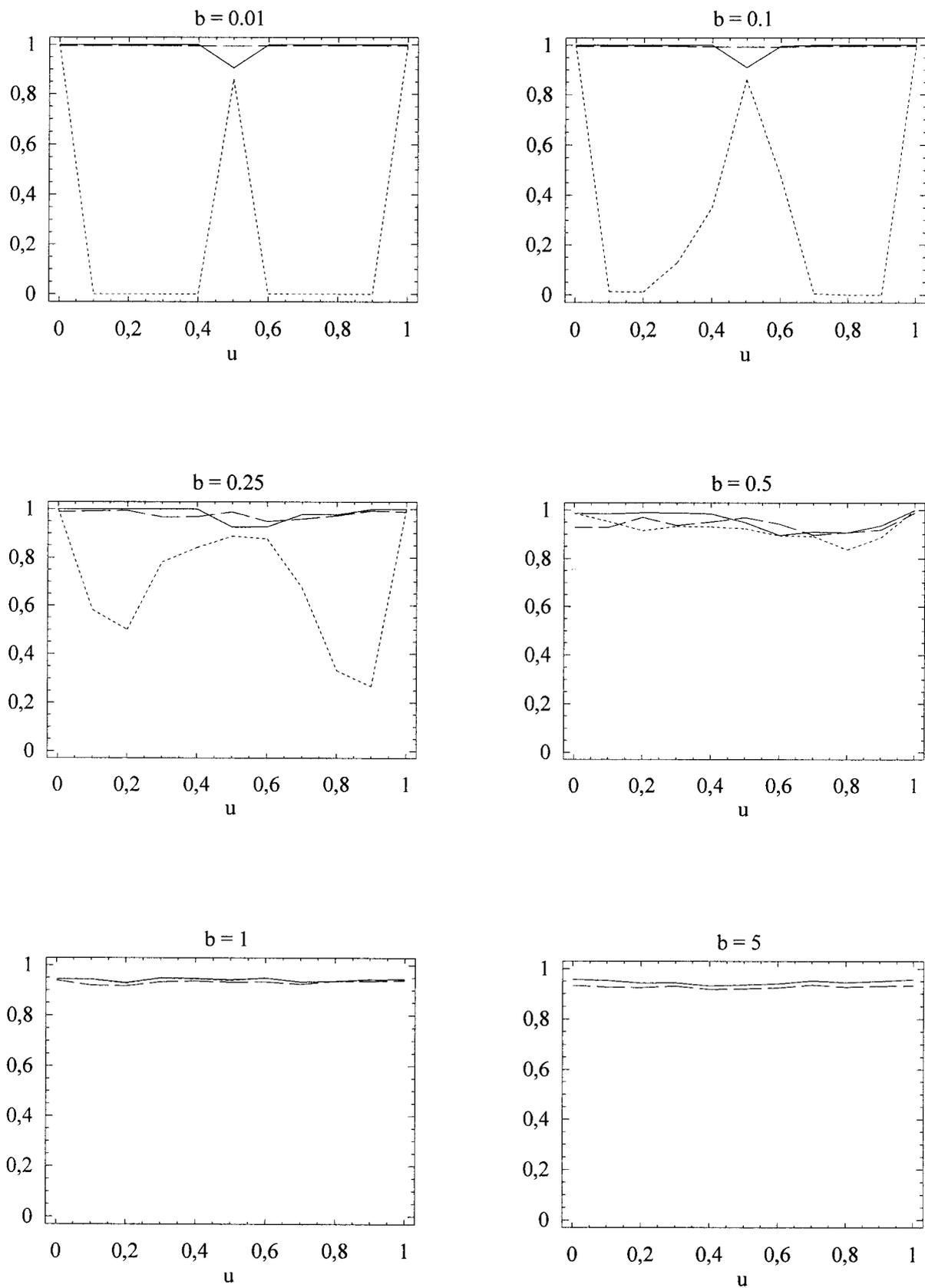
Abbildung 10: ÜW Maximumverteilung, $n=15$, Verlauf über b

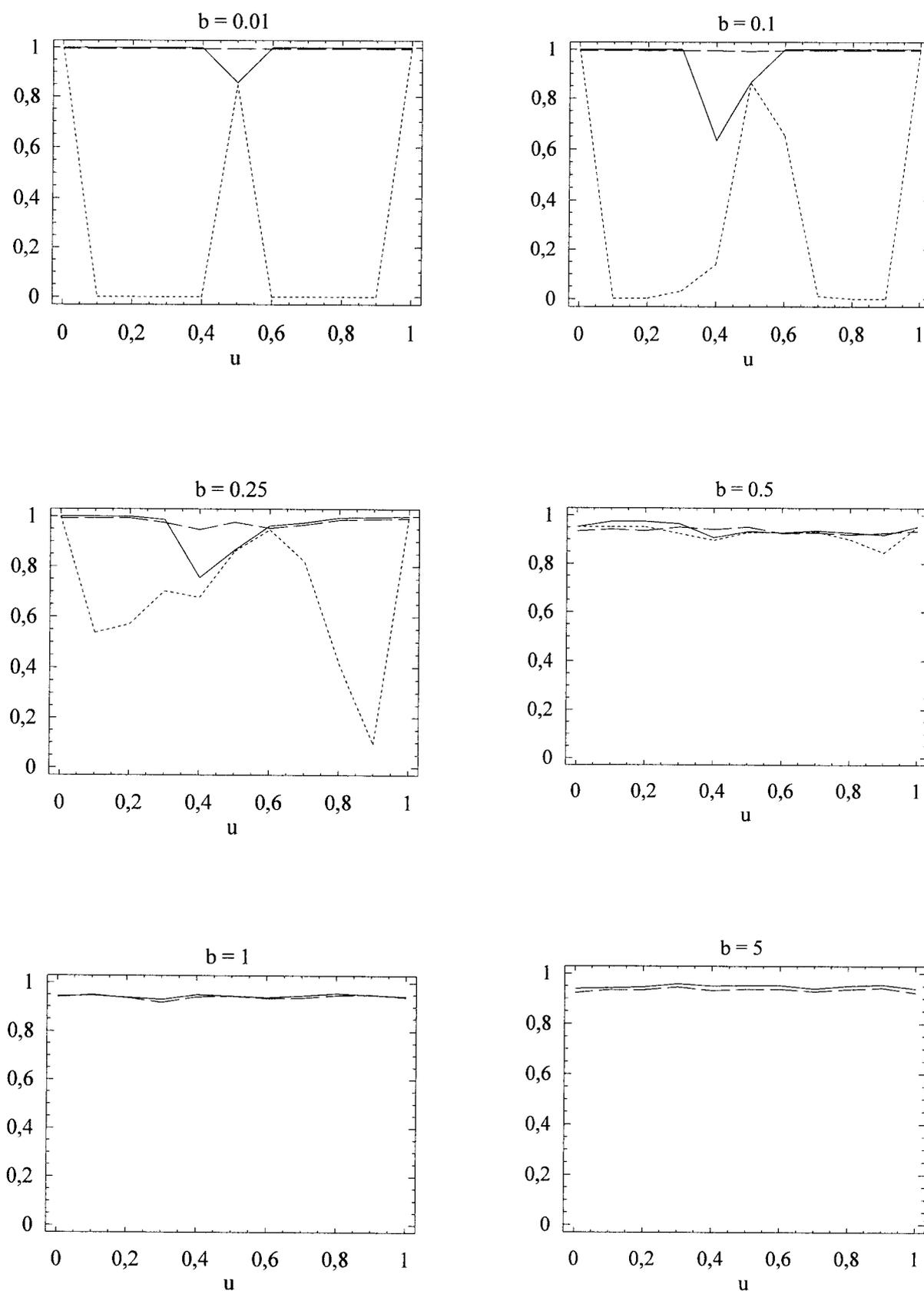
Überdeckungswahrscheinlichkeiten im Verlauf über u Abbildung 11: ÜW Maximumverteilung, $n=2$, Verlauf über u

Abbildung 12: ÜW Maximumverteilung, $n=3$, Verlauf über u

Abbildung 13: ÜW Maximumverteilung, $n=4$, Verlauf über u

Abbildung 14: ÜW Maximumverteilung, $n=5$, Verlauf über u

Abbildung 15: ÜW Maximumverteilung, $n=10$, Verlauf über u

Abbildung 16: ÜW Maximumverteilung, $n=15$, Verlauf über u

Durchschnittliche Intervalllängen von korrigierter LQ-, Pivotmethode und Kombinationsverfahren
Maximumverteilung

$u = 0$															
b	0.01			0.25			0.5			1			2		
n	LQ_{korr}	Pivot	Kombi												
2	2.07	0	2.07	2.11	3.06	4.70	2.44	9.62	10.52	3.78	19.63	20.12	6.67	38.89	39.17
3	1.27	0	1.27	1.32	0.77	1.61	1.65	2.47	2.81	2.83	5.10	5.19	5.43	10.10	10.11
4	1.12	0	1.12	1.12	0.53	1.19	1.38	1.67	1.88	2.38	3.41	3.44	4.63	6.76	6.77
5	1.05	0	1.05	1.00	0.40	1.03	1.20	1.36	1.52	2.11	2.70	2.72	4.04	5.25	5.25
10	0.99	0	0.99	0.84	0.20	0.84	0.83	0.80	0.89	1.47	1.49	1.49	2.83	2.89	2.89
15	0.99	0	0.99	0.78	0.15	0.77	0.67	0.64	0.69	1.20	1.18	1.18	2.29	2.25	2.25

$u = 0.3$															
b	0.01			0.25			0.5			1			2		
n	LQ_{korr}	Pivot	Kombi												
2	2.07	0	2.07	2.14	6.80	7.91	2.42	9.51	10.41	3.63	18.83	19.30	6.48	37.72	38.00
3	1.27	0	1.27	1.38	1.80	2.20	1.66	2.67	2.92	2.79	5.04	5.13	5.54	10.30	10.32
4	1.12	0	1.12	1.11	1.23	1.47	1.36	1.82	1.94	2.45	3.51	3.54	4.65	6.80	6.80
5	1.05	0	1.05	0.97	0.97	1.18	1.18	1.46	1.54	2.14	2.73	2.74	4.03	5.22	5.22
10	0.99	0	0.99	0.69	0.52	0.70	0.80	0.78	0.82	1.45	1.47	1.47	2.79	2.85	2.85
15	0.99	0	0.99	0.58	0.40	0.53	0.65	0.61	0.63	1.18	1.17	1.17	2.30	2.25	2.25

$u = 0.7$															
b	0.01			0.25			0.5			1			2		
n	LQ_{korr}	Pivot	Kombi												
2	2.07	0	2.07	2.11	3.35	4.95	2.45	9.60	10.50	3.64	18.72	19.23	6.72	39.17	39.46
3	1.27	0	1.27	1.34	1.22	1.98	1.68	2.69	2.99	2.73	4.94	5.03	5.32	9.92	9.95
4	1.12	0	1.12	1.09	0.85	1.45	1.41	1.97	2.11	2.43	3.45	3.48	4.67	6.82	6.83
5	1.05	0	1.05	0.96	0.78	1.29	1.23	1.58	1.69	2.12	2.69	2.71	4.09	5.31	5.31
10	0.99	0	0.99	0.65	0.52	0.81	0.82	0.92	0.94	1.46	1.49	1.49	2.82	2.87	2.87
15	0.99	0	0.99	0.50	0.47	0.62	0.66	0.72	0.73	1.19	1.17	1.17	2.29	2.25	2.25

Tabelle 3: Durchschnittliche Intervalllängen; Maximumverteilung

5.2 Eine Reanalyse für die Normalverteilung

Die positive Wirkung des Kombinationsverfahrens auf die Überdeckungswahrscheinlichkeiten verbunden mit der Entwicklung eines Rundungsschwellenwertes motiviert eine Reanalyse von gerundeten Daten aus Normalverteilungen. Eine optimierte Verfahrensauswahl soll nun die von Lee und Vardeman vorgeschlagene Adjustierung der kritischen Werte ersetzen. Der Schwellenwert für den Rundungsgrad könnte Aufschluss über den tatsächlichen Bedarf an korrektiven Eingriffen geben. Mit seiner Hilfe können Zweifel darüber ausgeräumt werden, ob Resultate bereits abgeschlossener Studien, die Rundungseffekte außer Acht gelassen haben, weiterhin als verbindlich anzusehen sind. Auch wäre zu prüfen, inwieweit eine Neugestaltung unliebsame Intervallverlängerungen umgehen kann.

Die Beantwortung dieser Fragen erfordert zunächst eine Reproduktion der Studie von Lee und Vardeman. Zum einen sind in dem Bericht die Überdeckungswahrscheinlichkeiten der unkorrigierten LQ-Version nicht detailgenau überliefert. Zum anderen müsste zu jeder simulierten Stichprobe die geschätzte Stichprobenvarianz protokolliert sein, um zu wissen, für welche der Methoden das Kombinationsverfahren entschieden hätte. n. Zum einen sind in dem Bericht die Überdeckungswahrscheinlichkeiten der unkorrigierten LQ-Version nicht detailgenau überliefert. Zum anderen müsste zu jeder simulierten Stichprobe die geschätzte Stichprobenvarianz protokolliert sein, um zu wissen, für welche der Methoden das Kombinationsverfahren entschieden hätte.

Zum besseren Vergleich der Methoden sollte die Studie in ihrem Aufbau an denjenigen von Lee und Vardeman angeglichen sein: Die Parametereinstellung wird im wesentlichen übernommen, lediglich eine Änderung der Nivauereinstellung ist aus bereits in Abschnitt 4.3 diskutierten Gründen nicht vorgesehen. Das Simulationsprogramm führt jeweils 1000 Läufe mit wechselnden Konstellationen der Parameter $n = 2, 3, 4, 5, 10, 15$, $\mu = (0, 0.1, 0.2, 0.3, \dots, 0.9)$ und $\sigma = (0.01, 0.1, 0.25, 0.5, 1, 2, 5)$ durch, die als Grundlage zur Schätzung der Überdeckungswahrscheinlichkeiten dienen.

Bestimmung des Korrekturfaktors für Spannweite 0

Stichproben mit Spannweite 0 sind bei kleinen Stichprobenumfängen und sehr kleiner bis moderater Standardabweichung keine Seltenheit. In der Tat sind sie unter bestimmten Parameterbedingungen mit einer Häufigkeit von mehr als α vorzufinden, ohne dass eines der Konfidenzintervalle (Pivot oder LQ) die Lage richtig zuordnet. Die Gründe dafür sind bei allen Verteilungen ähnlich gelagert und im Kapitel über

die Maximumverteilung ausführlich besprochen. Es ist daher dringend erforderlich, das Konfidenzintervall auszudehnen. Es wird genau so lange erweitert, bis die Eintrittswahrscheinlichkeit von Spannweite-Null-Stichproben unter das Niveau gedrückt ist, wie auch immer σ in Wahrheit aussehen mag. Man wählt also wieder denjenigen Wert Δ als Grenze, bei dem die Profil-Likelihood, die diese Wahrscheinlichkeit formal ausdrückt, unter 0.05 fällt.

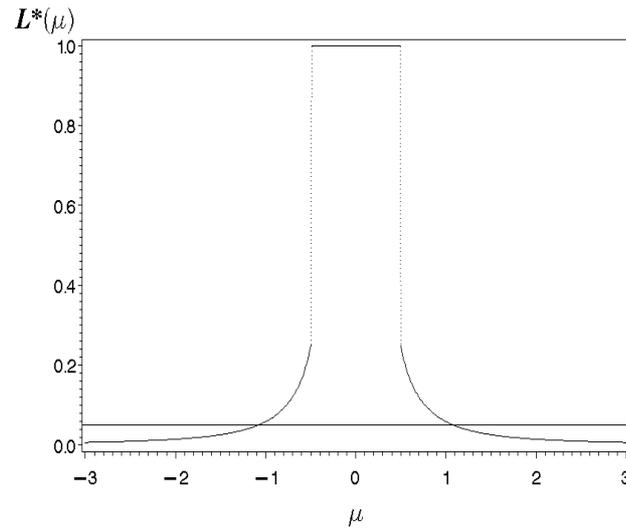


Abbildung 17: Profil-Likelihood der Normalverteilung für die Stichprobe $(0,0)$

Die zugehörige Profil-Likelihood ist gegeben durch

$$L^*(\mu | x_1 = x_2 = 0) = \sup_{\sigma > 0} \left[\Phi \left(\frac{0.5 - \mu}{\sigma} \right) - \Phi \left(\frac{-0.5 - \mu}{\sigma} \right) \right]^2.$$

Wegen der Symmetrieeigenschaft der Normalverteilung genügt es, den Wert Δ ausgehend von k in einer Richtung zu suchen. Für Verteilungen mit Lage- und Skalierungsparameter verschiebt sich außerdem das Intervall bei $x_i = k$, $i = 1, \dots, n$, ohne seine Länge zu verändern. Daher definiert Δ zugleich alle Intervallgrenzen

$$KI(\mu | x_i = k, k \in \mathbb{N}, i = 1, \dots, n) = [k - \Delta_n, k + \Delta_n]. \quad (5.10)$$

Mit

$$\lim_{\mu \rightarrow \infty} \sup_{\sigma > 0} L(\mu, \sigma) = 0$$

ist die Existenz von Δ gesichert. Bereits ab $n = 5$ entfallen die Korrekturen. Die Profil-Likelihood fällt an den Unstetigkeitsstellen abrupt unter Niveau und erzeugt damit keine erweiterten Korrekturintervalle. Tabelle 4 fasst die Korrekturen für kleinere n zusammen.

n	2	3	4
Δ_n	1.1	0.7	0.6

Tabelle 4: Korrekturen für die Normalverteilung

Die Vorgehensweise bei der Maximumverteilung ist mustergültig und wird für sämtliche anderen Verteilungen beibehalten. Die ÜW-Kurven der einzelnen Verfahren (Grafiken 18-21) liefern einen groben Anhaltspunkt für den Schwellenwert. In ihrem Verlauf über σ gibt es einen Schnittpunkt, in dem sich die Verlaufskurven kreuzen. Für Standardabweichungen von unter 1 ($n = 2$) liegt die LQ-Kurve höher als Pivot, danach umgekehrt. Dieser Kreuzungspunkt driftet mit wachsendem Stichprobenumfang weiter nach links, bis sich schließlich die Kurven bei $n = 15$ gar nicht mehr kreuzen, sondern von oben und unten aufeinander zulaufen, auf $1 - \alpha$ einschwenken und sich fortan überdecken. Ab diesem Punkt sind die Verfahren kompatibel.

Der Schwellenwert ist nun so einzurichten, dass das Niveau mit möglichst kleinen Intervalllängen eingehalten wird. Aus Gründen der Einfachheit sollte er für alle n gleichermaßen Bestand haben. Nach den beschriebenen Beobachtungen aus den Verlaufskurven sollte das Kombinationsverfahren spätestens für

$$\sigma_0 = 1 \leq \hat{\sigma} \quad (5.11)$$

zur Pivot-Methode wechseln. Zum Überprüfen dieser Bedingung sind theoretisch viele Varianzschätzer denkbar. Ein gängiger Schätzer wie die Stichprobenvarianz erfüllt diesen Zweck, ohne den Ablauf unnötig zu komplizieren. Die Stichprobenvarianz ist mit elementaren Mitteln zu berechnen und fließt ohnehin als Teilergebnis in das Pivot-Intervall ein. Die Rechenzeit zum Abfragen der Schwellenwertbedingung kann man so in den meisten Fällen wieder gutschreiben. Der erste Versuch mit $\sigma_0 = 1$ genügt jedoch den Niveau-Anforderungen nicht. Zu langsam steigt das Kombinationsverfahren auf das Pivot-Verfahren um, wie eine verminderte Überdeckungsrate bei größeren σ erkennen lässt. Der Schwellenwert wird sukzessive gesenkt. Die besten Quoten erzielt man mit $\sigma = 0.4$. Über den Rundungsgrad formuliert lautet dann die Schwellenwertbedingung:

$$\text{Nehme Pivot, falls } \hat{r} = \frac{1}{\hat{\sigma}} \geq \frac{1}{\sigma_0} = 2.5. \quad (5.12)$$

Ergebnisse der Simulationsstudie

Die Resultate untermauern in ihren Kern die von Lee und Vardeman beschriebenen Sachverhalte (vgl. Grafiken 22-27): Die Pivot-Methode ist bei sehr kleinen Stichprobenumfängen außerstande, den Parameterbereich richtig einzukreisen. Abgesehen von einigen Zufallstreffern, wenn der wahre Parameter in der Mitte zweier Gittermarkierungen platziert ist oder exakt darauf, ist die Trefferquote mangelhaft. Eine

Erhöhung der Varianz wirkt sich vorteilhaft aus, aber bei einem Stichprobenumfang von $n = 2$ bedarf es schon einer außerordentlich hohen Varianz von $\sigma = 5$, um die ÜW annähernd auf Niveau anzuheben. Für $n = 15$ stellt sich eine Niveaueinhaltung erst ab $\sigma = 0.5$ ein. Die LQ-Methode (ohne Korrekturen nach Lee und Vardeman) kann bis einschließlich $\sigma = 0.5$ mit recht akzeptablen Überdeckungswahrscheinlichkeiten aufwarten, sinkt aber dann dramatisch, bei $n = 2$ auf 75%, bei $n = 5$ auf 90%. Selbst bei $n = 15$ verliert die LQ-Methode bei $\sigma = 5$ an Treffgenauigkeit und kann nicht völlig mit der Pivot-Methode gleichziehen.

Das Kombinationsverfahren dagegen hält das Niveau ein und zeigt sich allen Parameter-Szenarien gewachsen. Besonders dann, wenn die ÜW von LQ und Pivot gleichermaßen nachlassen, macht sich eine sorgsame Auswahl des Verfahrens bezahlt. Damit kann sich das Kombinationsverfahren durchaus mit der korrigierten LQ-Version von Lee und Vardeman messen. Letztere ist in den Grafiken der Übersichtlichkeit halber nicht eigens aufgeführt. Die ÜW stimmen mit denen der Kombinationsmethode überein und die Kurvenverläufe wären in den Grafiken nicht voneinander zu trennen. Feine Diskrepanzen lassen sich erst bei einem Blick auf die Intervalllängen ausmachen.

Im großen und ganzen kann man die Verfahren gegeneinander austauschen. Die einzige Ausnahme bilden Stichproben von sehr geringem Umfang ($n = 2$ oder 3), die aus einer wenig streuenden Grundgesamtheit stammen. Hier lässt sich die Intervalllänge durch das Kombinationsverfahren in einer relevanten Größenordnung verkürzen.

Die Praxis

Der große Vorteil des Kombinationsverfahrens besteht gerade in der Einbindung gängiger und in ihren Eigenschaften ausführlich diskutierter Verfahren. Altbewährte Standardmethodik wird nicht generell verdrängt, vielmehr ihr Einsatzgebiet genau klassifiziert. Bei Licht besehen schmälert die Schwellenwertbedingung das Pivot-Verfahren kaum in seiner Geltung. Der Rundungsschwellenwert von 2.5 setzt die Messlatte in einer Höhe an, die kaum jemals von einer Stichprobe übersprungen wird. Oft genug lässt man sich von den geringen Abweichungen unter den Stichprobenwerten blenden und schließt intuitiv auf eine geringe Varianz und hohen Rundungsgrad. Dazu ein Beispiel aus einem Handbuch für kleine Datensätze (Hand et al., 1994):

Ein Gerät soll die spezifische Temperatur zu verschiedenen Wassertemperaturen messen und man wünscht eine Auskunft über die Präzision des Messgerätes. Sechs unterschiedliche Personen eichen das Gerät zunächst auf 15 Grad und messen dann

die spezifische Temperatur mit demselben Gerät bei 10°C . Die Messergebnisse:

10.010 10.021 10.016 10.019 10.017 10.016.

Intuitiv bestehen Bedenken vier Nachkommastellen seien nicht ausreichend, um mit der Pivot-Methode die Präzision des Gerätes verlässlich zu diagnostizieren. Erst wenn man die Daten verschiebt – was bei Lage- und Skalierungsverteilungen völlig legitim ist –, zeichnet sich das echte Ausmaß des Rundungsgrades ab. Die Daten sind dann etwa

10 21 16 19 17 16.

Das ergibt eine Standardabweichung von ≈ 3.73 und einen geschätzten Rundungsgrad von 0.27. Damit liegt der Rundungsgrad weit unter dem Schwellenwert. Das realisierte Pivotkonfidenzintervall zum 5%- Niveau der untransformierten Werte ist

[10.0126; 10.0204].

Es mag für den Hersteller des Gerätes unbefriedigend sein, dass sein Produkt nicht mit der erhofften Genauigkeit von einem Tausendstel, sondern höchstwahrscheinlich nur mit 8 Tausendstel arbeitet, aber an der Glaubwürdigkeit des Ergebnisses aufgrund eines unpassenden Verfahrens besteht kein Zweifel. Ein anderes Extrembeispiel aus Stein (2000) zitieren Lee und Vardeman. Dabei handelt es sich um eine Serie von digitalen Anzeigen eines Messgerätes:

1.3 1.2 1.3 1.3 1.3 1.2 1.2 1.3 1.2 1.3.

Wieder werden die Daten transformiert auf

3 2 3 3 3 2 2 3 2 3.

Deren geschätzte Standardabweichung beläuft sich auf ≈ 0.516 und der geschätzte Rundungsgrad auf ≈ 1.94 . Die Autoren geben das korrigierte LQ-Intervall mit [1.226; 1.294] an und kommentieren: „Steins Annahme, der wahre Parameter liege in der Nähe von 1.26, war überoptimistisch.“ Das aber ist ein generelles Problem der Punktschätzung. Erst die Ergänzung durch ein KI macht deutlich, wie verlässlich diese Angabe ist. Nicht einmal in diesem außergewöhnlichen Beispiel liegt der geschätzte Rundungsgrad über dem Schwellenwert, und auch das Pivot-KI verweist auf den Bereich [1.223; 1.297].

Überdeckungswahrscheinlichkeiten im Verlauf über σ

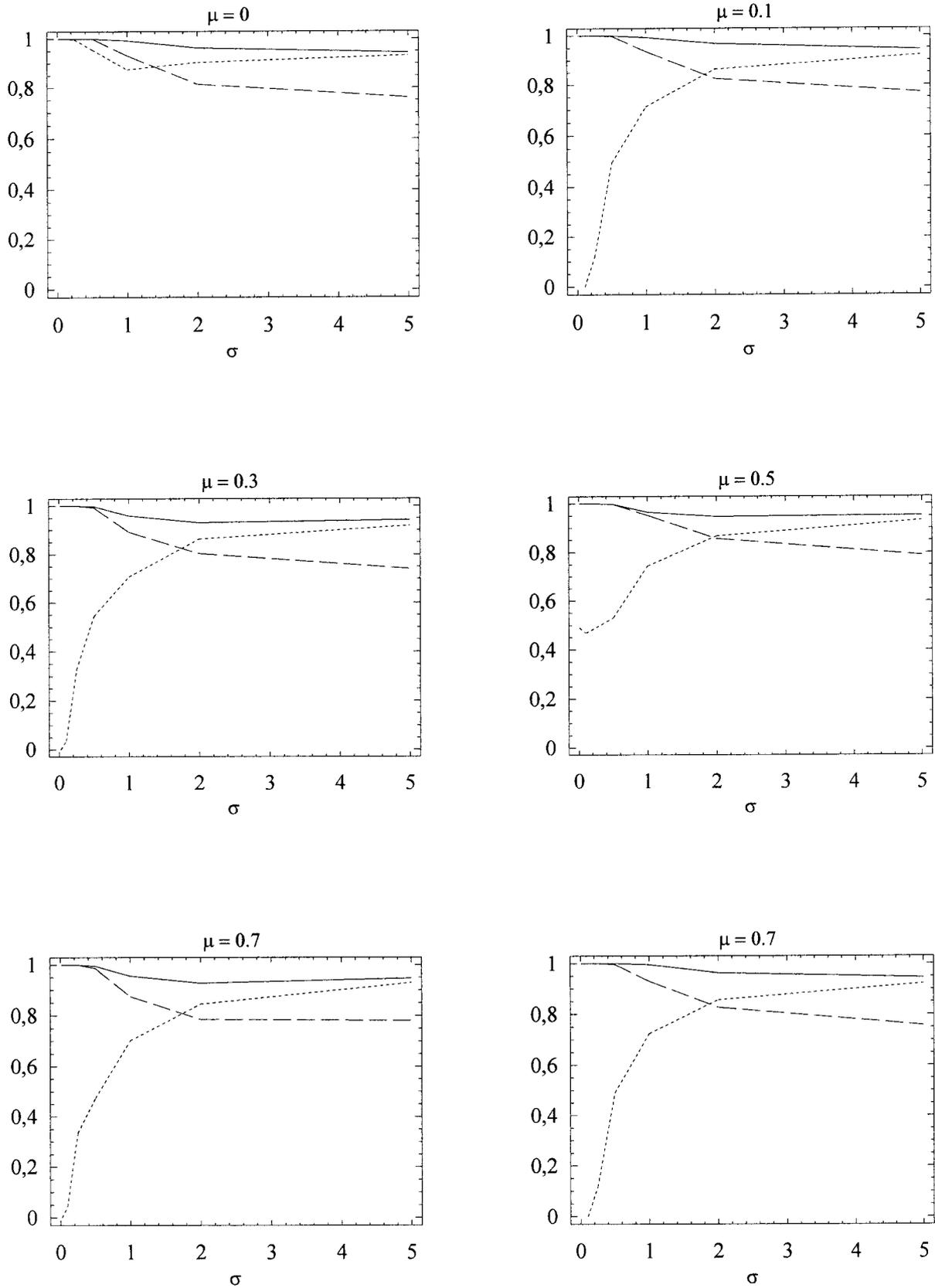
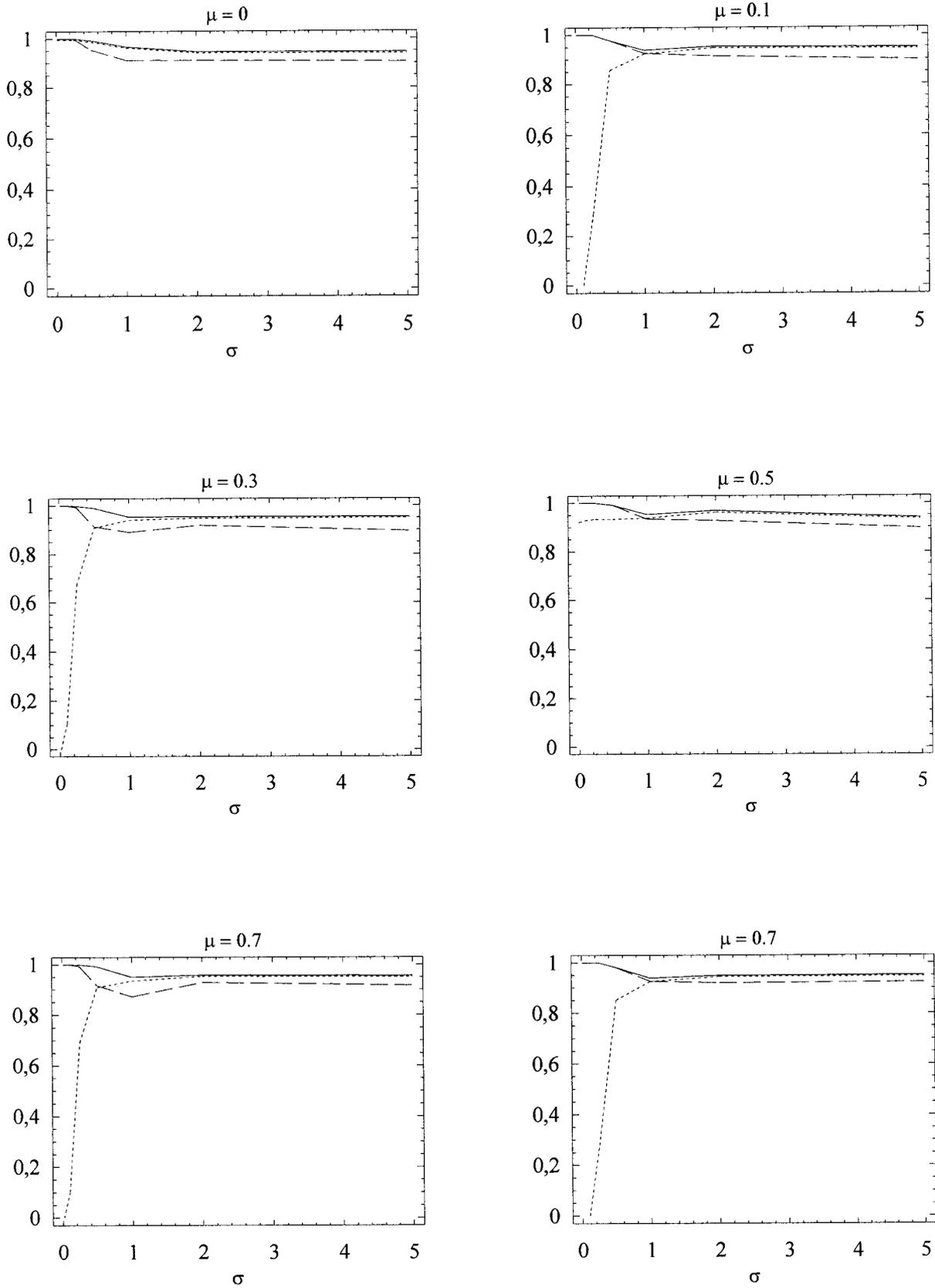
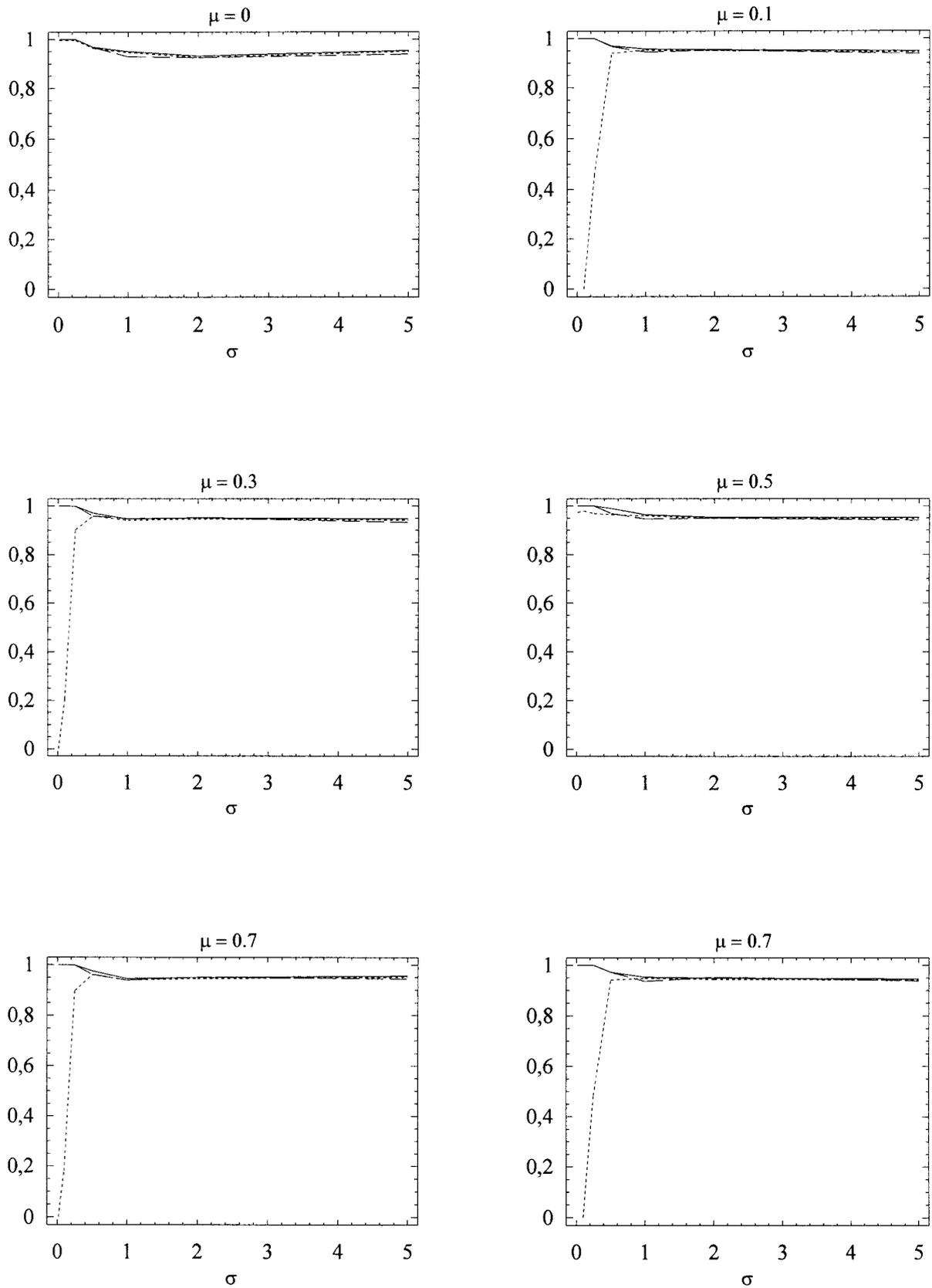
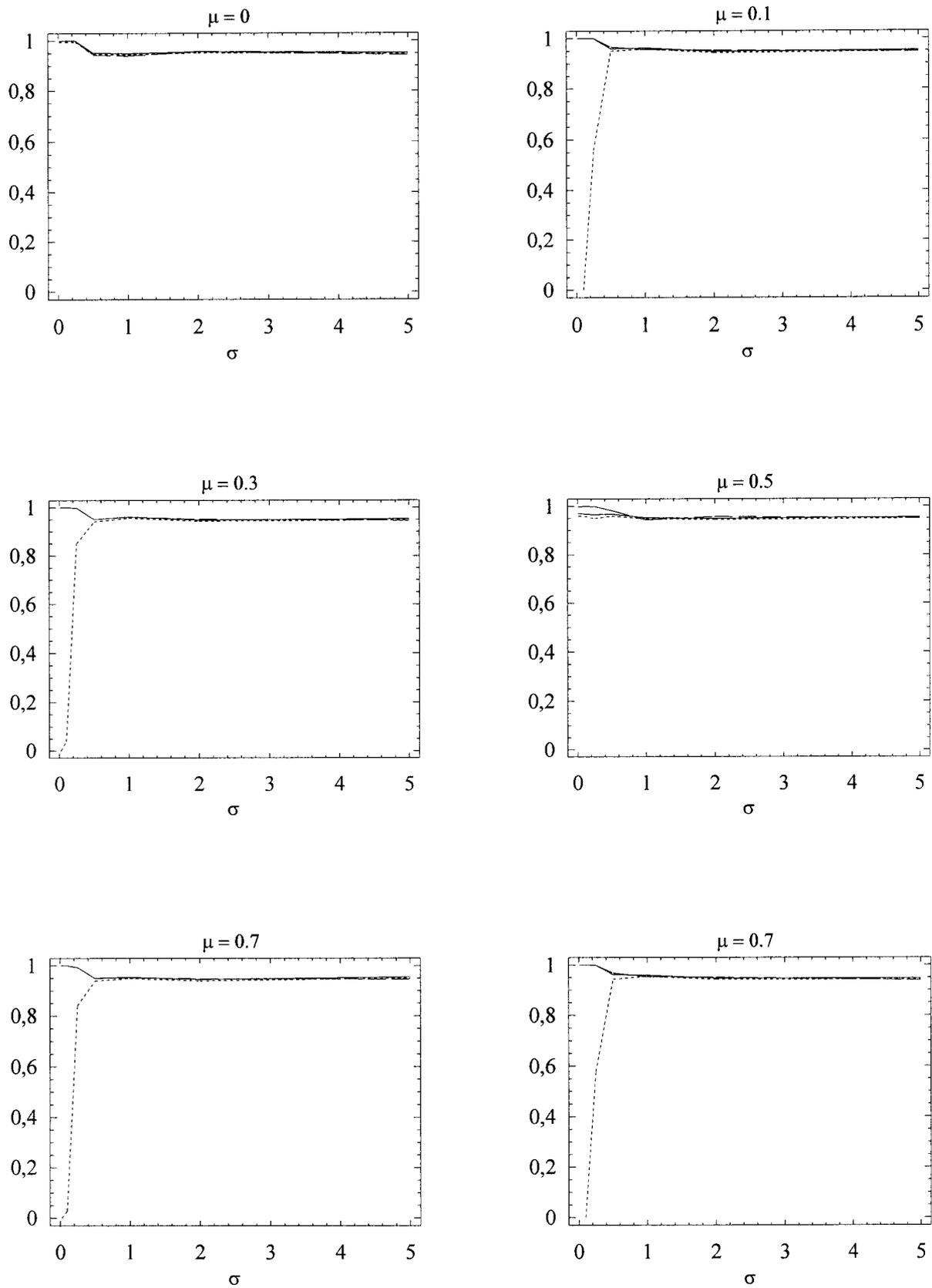
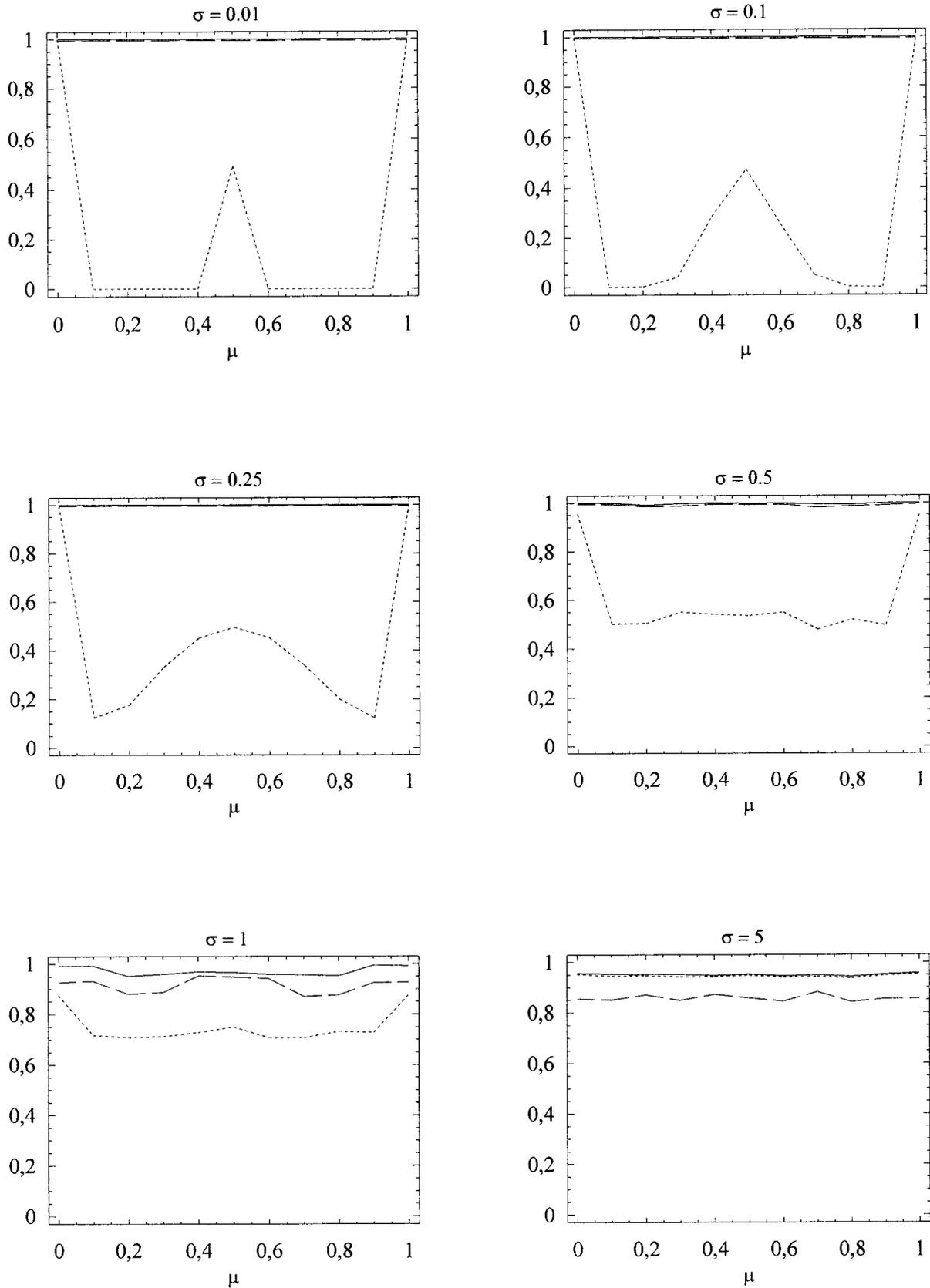


Abbildung 18: ÜW Normalverteilung, $n=2$, Verlauf über σ

Abbildung 19: ÜW Normalverteilung, $n=5$, Verlauf über σ

Abbildung 20: ÜW Normalverteilung, $n=10$, Verlauf über σ

Abbildung 21: ÜW Normalverteilung, $n=15$, Verlauf über σ

Überdeckungswahrscheinlichkeiten im Verlauf über μ Abbildung 22: ÜW Normalverteilung, $n=2$, Verlauf über μ

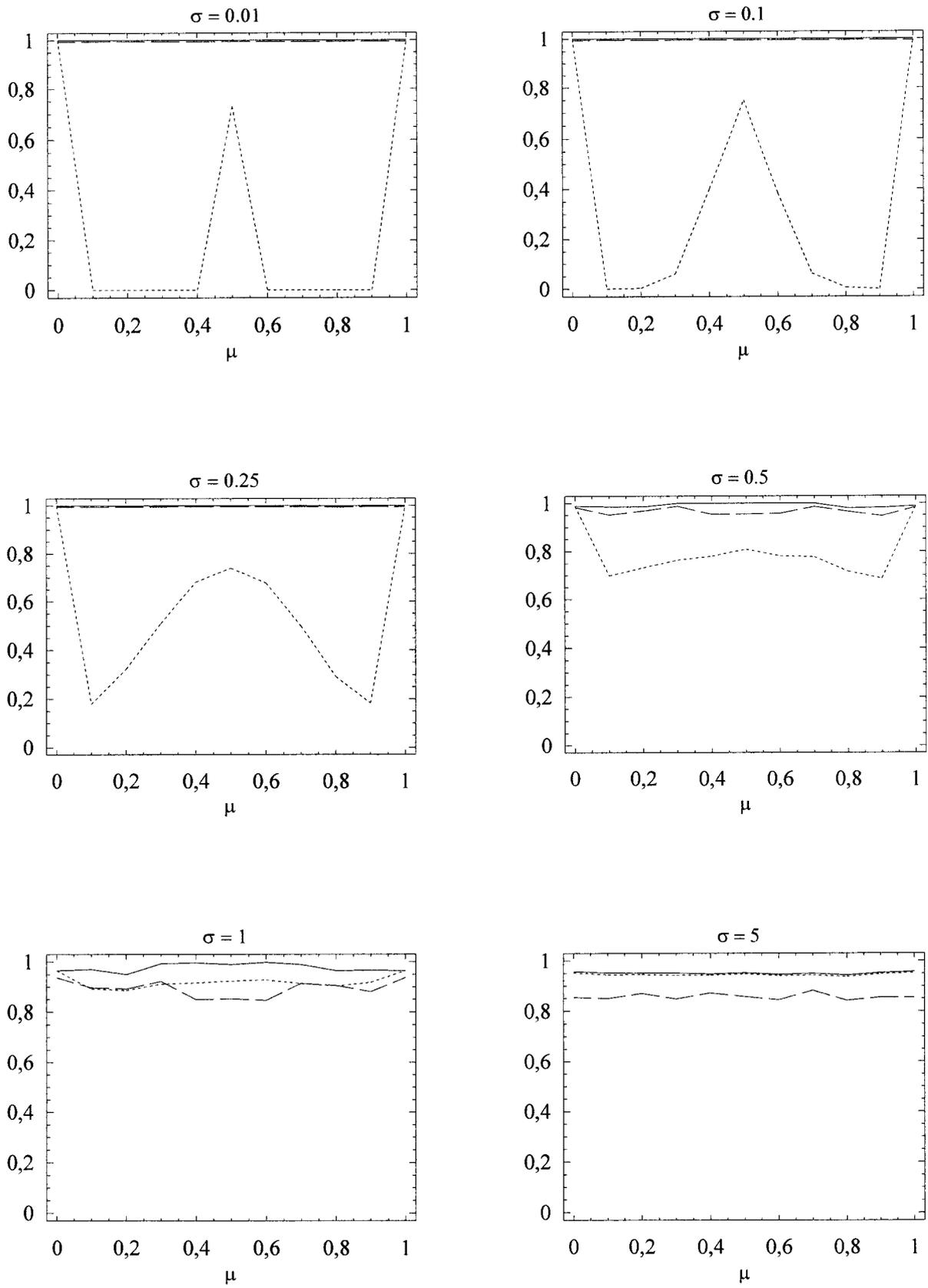


Abbildung 23: ÜW Normalverteilung, $n=3$, Verlauf über μ

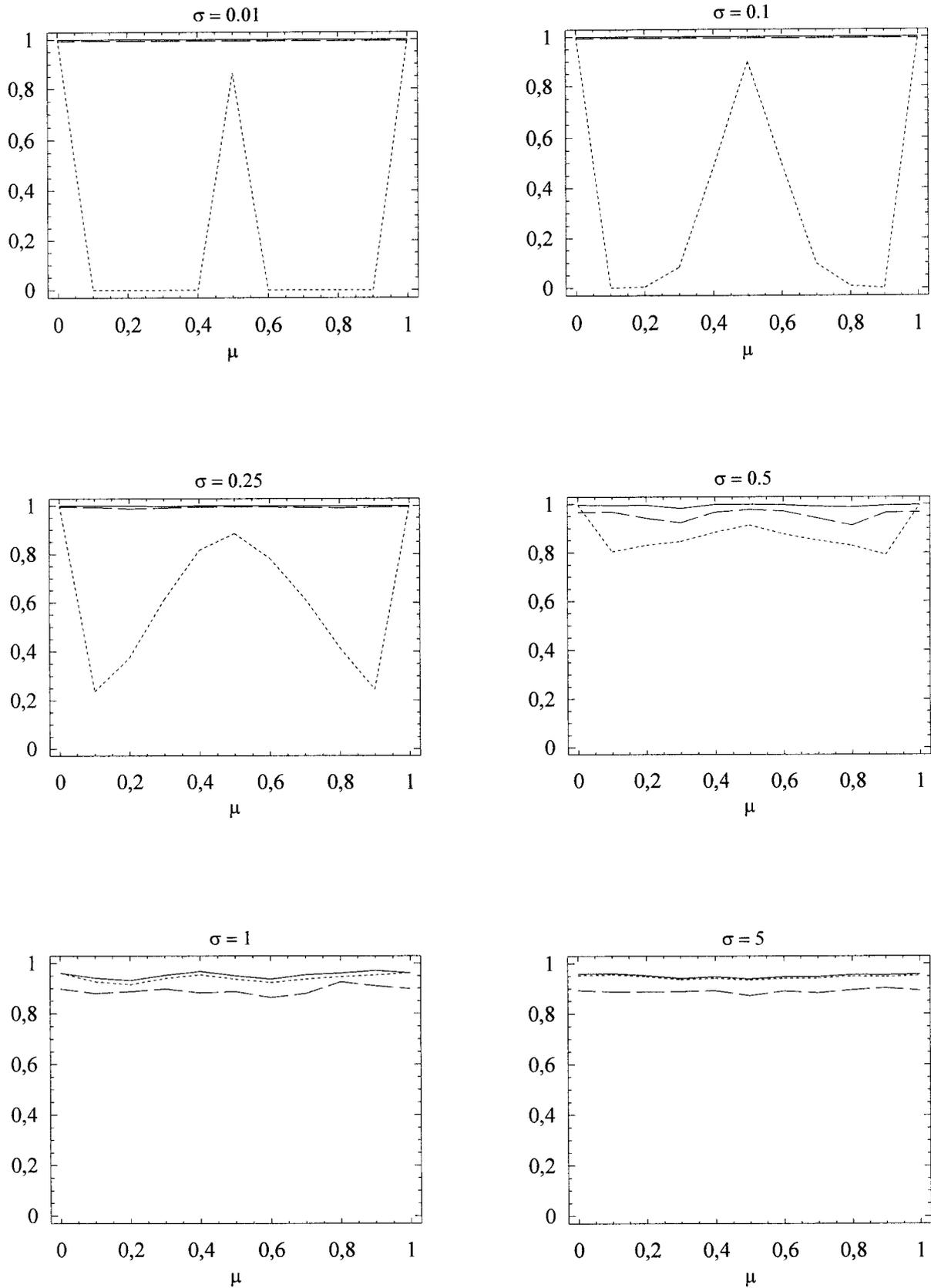
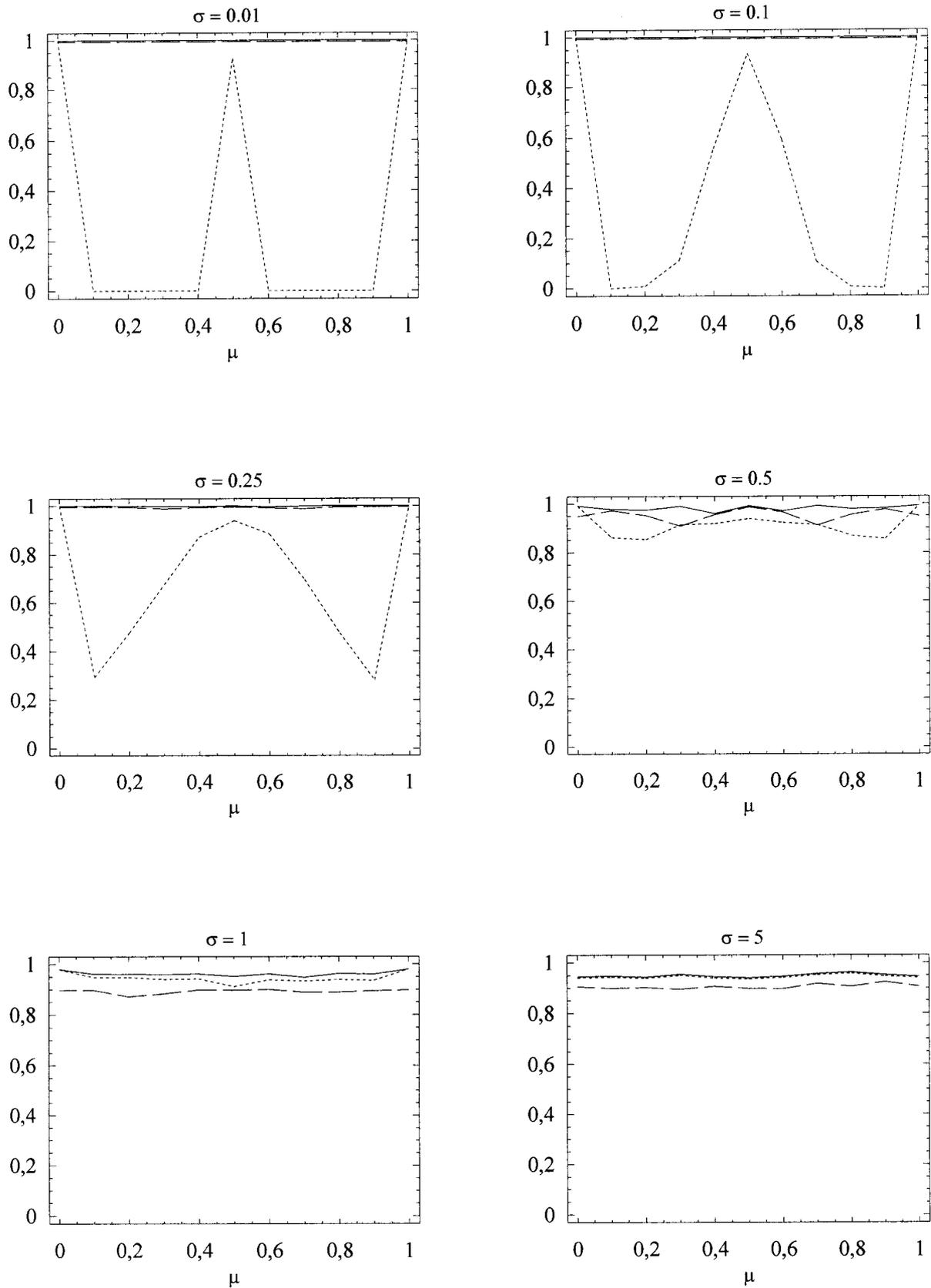
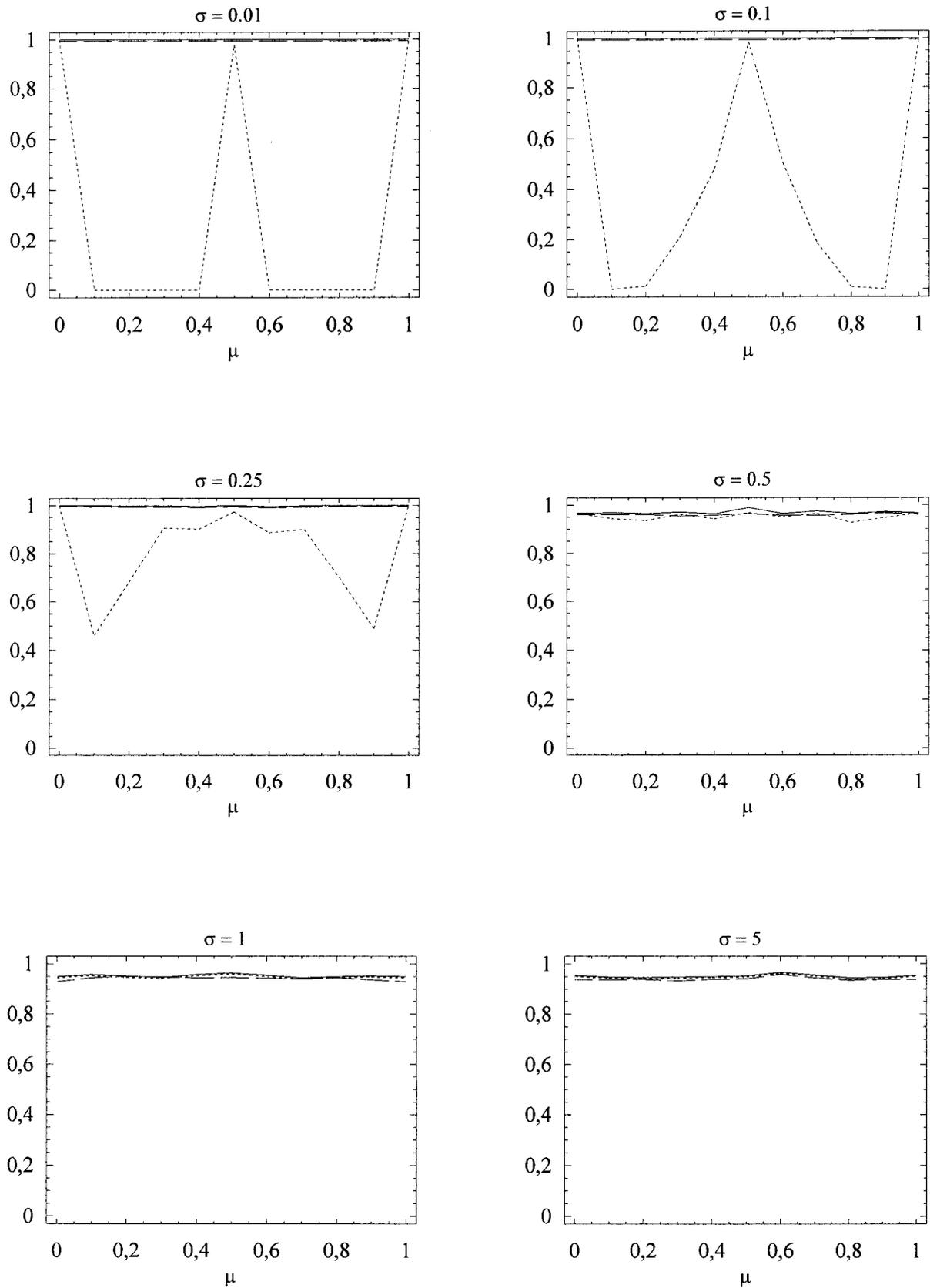
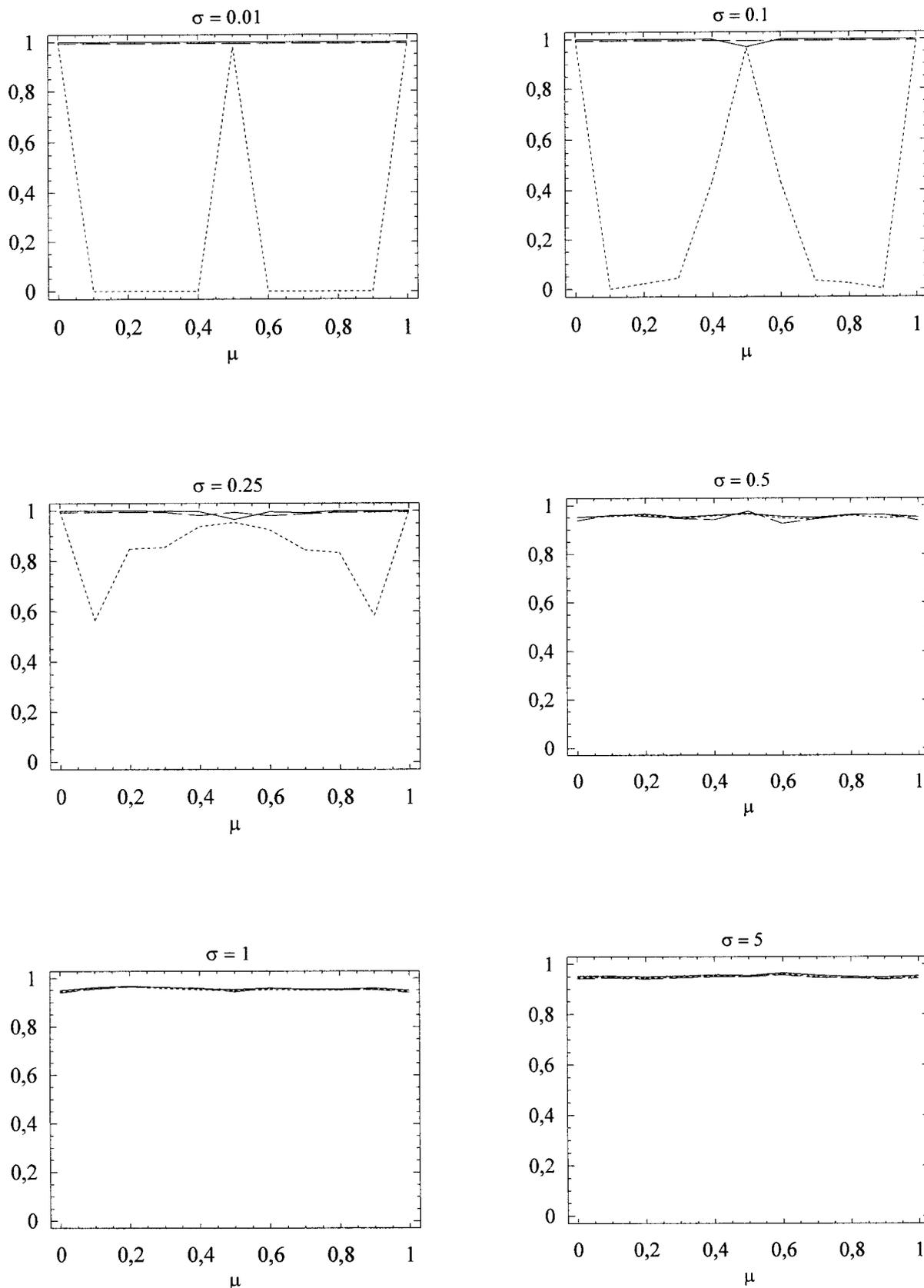


Abbildung 24: ÜW Normalverteilung, $n=4$, Verlauf über μ

Abbildung 25: ÜW Normalverteilung, $n=5$, Verlauf über μ

Abbildung 26: ÜW Normalverteilung, $n=10$, Verlauf über μ

Abbildung 27: ÜW Normalverteilung, $n=15$, Verlauf über μ

**Durchschnittliche Intervalllängen von korrigierter LQ-Methode und
Kombinationsverfahren**

$\mu = 0$										
σ	0.01		0.25		0.5		1		5	
n	LQ _{korrr}	Kombi								
2	6.167	2.20	6.596	3.19	9.726	8.03	16.050	14.84	70.098	70.82
3	1.553	1.40	1.727	1.58	2.710	2.62	4.601	4.79	21.651	21.97
4	1.035	1.20	1.122	1.29	1.753	1.81	3.013	3.11	14.628	14.69
5	1.000	1.00	1.033	1.03	1.402	1.41	2.439	2.44	11.897	11.77

$\mu = 0.5$										
σ	0.01		0.25		0.5		1		5	
n	LQ _{korrr}	Kombi								
2	9.249	7.39	9.077	7.40	10.077	8.49	15.658	15.01	68.904	72.80
3	2.474	2.47	2.473	2.49	2.765	2.82	4.666	4.70	22.446	21.97
4	1.575	1.63	1.557	1.64	1.779	1.83	2.983	3.05	14.713	14.55
5	1.221	1.26	1.219	1.26	1.398	1.41	2.453	2.45	11.680	11.67

Tabelle 5: Durchschnittliche Intervalllängen, Normalverteilung

5.3 Konfidenzintervalle für die Logistische Verteilung

Die Logistische Verteilung ist besonders für die Beschreibung von Wachstumskurven geeignet und wird zu demografischen Zwecken ebenso eingesetzt, wie zur Charakterisierung biologischer Populationen. Im Laufe der Zeit fand sie zunehmend in anderen Gebieten Anwendung, so etwa in der Überlebenszeitanalyse. Ihre Verteilungsfunktion

$$F(x) = \left[1 + \exp \left\{ -\frac{x-a}{b} \right\} \right]^{-1}, \quad b > 0 \quad (5.13)$$

lässt die Verwandtschaft mit anderen Lage- und Skalierungsverteilungen erkennen. Die ersten beiden Momente der Logistischen Verteilung sind

$$\begin{aligned} E(X) &= a \\ \text{Var}(X) &= \frac{b^2 \pi^2}{3}. \end{aligned} \quad (5.14)$$

Die Parameter sind eng mit den Momenten verknüpft, und damit geben die Momentenschätzer brauchbare Startwerte für die Implementierung des LQ-Verfahrens ab. Um die konstanten Faktoren aus der Varianzformel zu eliminieren, wird manchmal eine standardisierte Form der Verteilungsfunktion gewählt mit $\sigma = b\pi/\sqrt{3}$. Wenn nicht sicher ist, ob die vorliegenden Daten aus einer Logistischen oder der Normalverteilung stammen, gestattet die standardisierte Verteilung einen einfacheren Vergleich anhand der Verteilungsform. In den Berechnungen hat das aber keinerlei Relevanz.

Die ML-Schätzer sind nicht in geschlossener Form darstellbar. Eine Lösung ist entweder durch Maximierung der Likelihood

$$L(a, b) = \prod_{i=1}^n b^{-1} [\exp \{-(x_i - a)/b\}]^{-1} [1 + \exp \{-(x_i - a)/b\}]^{-2} \quad (5.15)$$

möglich oder durch numerische Nullstellenbestimmung der logarithmierten und einmalig abgeleiteten Likelihood. Die Umformungen enden in zwei Gleichungen, die sich nicht weiter vereinfachen lassen, sondern ebenfalls nur numerisch zu lösen sind:

$$\begin{aligned} \frac{1}{2} &= \frac{1}{n} \sum_{i=1}^n \left[1 + e^{-(x_i - \hat{a})/\hat{b}} \right]^{-1} \\ \text{und} \quad 1 &= \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \hat{a}}{\hat{b}} \right) \frac{1 - e^{-(x_i - \hat{a})/\hat{b}}}{1 + e^{-(x_i - \hat{a})/\hat{b}}}. \end{aligned} \quad (5.16)$$

Die Likelihood der gerundeten Daten basiert auf den abschnittsweise aufsummierten

Häufigkeiten und schreibt sich als

$$L(a, b) = \prod_{i=1}^n \left[1 + \exp \left\{ -\frac{x_i + 0.5 - a}{b} \right\} \right]^{-1} - \left[1 + \exp \left\{ -\frac{x_i - 0.5 - a}{b} \right\} \right]^{-1}. \quad (5.17)$$

Das Kombinationsverfahren läuft nun nach der bereits bekannten Routine ab:

Bestimmung des Korrekturfaktors

Vor der Berechnung der Korrekturfaktoren betrachtet man zunächst noch einmal die Profil-Likelihood kritisch. Nur wenn sie an den Rändern stark genug abflacht und unter die 5%-Marke fällt, lässt sich die geplante Korrektur auch umsetzen.

$$\begin{aligned} & \lim_{a \rightarrow \infty} L^*(a, b \mid x_i = 0; i = 1, \dots, n) \\ &= \lim_{a \rightarrow \infty} \sup_{b > 0} \prod_{i=1}^n \left[1 + \exp \left\{ -\frac{0.5 - a}{b} \right\} \right]^{-1} - \left[1 + \exp \left\{ -\frac{-0.5 - a}{b} \right\} \right]^{-1} \\ &= 0, \end{aligned} \quad (5.18)$$

denn für beliebiges b und große a sind die Subtrahenden annähernd gleich. Damit sind die Vorkehrungen für eine Korrektur getroffen. Die errechnete Profil-Likelihood für $n = 2$ und $x_1 = x_2 = 0$ hat das in Abbildung 28 dargestellte Aussehen.

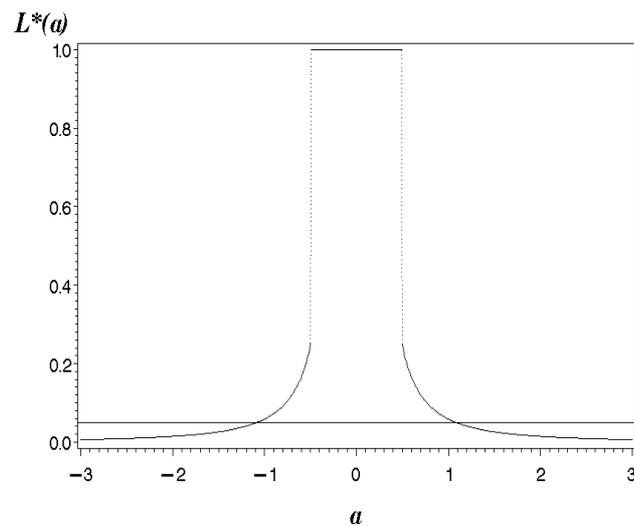


Abbildung 28: Profil-Likelihood der Logistischen Verteilung für die Stichprobe (0,0)

Die 5%-Linie wird bei ± 1.016 überschritten. Ab $n = 5$ liegen die Grenzen der Korrekturintervalle innerhalb der LQ-Intervalle und bewirken somit keine Weitung mehr. Für die Korrekturintervalle

$$KI(x_i = k) = [k + \Delta; k - \Delta]$$

ergibt sich die Liste in Tabelle 6.

n	2	3	4
Δ_n	1.016	0.635	0.518

Tabelle 6: Korrekturen für die Logistische Verteilung

Bestimmung des Schwellenwertes

Die Kurven der ÜW im Verlauf über b (Grafiken 29-32) deuten auf einen Schwellenwert von $\hat{b} \approx 1$ hin, denn an dieser Stelle kreuzen sich die Verläufe. Nach einigen Probeläufen und Vergleichen unter einer schrittweisen Minderung der Schwelle stellt sich schließlich der am besten geeignete Wert bei 0.25 ein. Das Niveau wird eingehalten ohne die durchschnittliche Intervalllänge wesentlich anzuheben. Der Schwellenwert für den Skalenparameter lässt sich aus dem Schwellenwert für den Rundungsgrad errechnen:

$$\hat{r} \geq \frac{1}{\sigma_0} = \frac{\sqrt{3}}{b_0 \pi} \approx 2.21. \quad (5.19)$$

Quantile der Pivotgröße

Die Verzeichnisse für die Quantile der Pivotgröße $(\hat{a} - a)/\hat{b}$ sind umfangreich (siehe etwa Antle et. al., 1970), aber wenig ergiebig, was sehr kleine Stichprobenumfänge angeht. Bain und Engelhardt (1991) z.B. geben Quantile an, wobei bestimmte Anteile der Stichprobe in zensierter Form vorliegen. In den Quantil tafeln sind außerdem nur Stichprobenumfänge in der Folge $n = 10, 20$, usw. aufgelistet. Simulationen müssen die hier benötigten Quantile zuvor erzeugen. Da die Werte der Logistischen Standardverteilung verglichen mit der Standardnormalverteilung breit streut, erfordert der Simulationsvorgang eine hohe Anzahl von Wiederholungen, bis sich die Quantile stabilisieren. Die folgende Quantiltabelle 7 ist nach jeweils 70000 Durchläufen entstanden.

Wegen der Symmetrie der Logistischen Verteilung genügt die Berechnung eines Quantils für beide Grenzen.

n	2	3	4	5	10	15
$q_{0.975}$	21.589	4.783	3.114	2.272	1.289	0.958

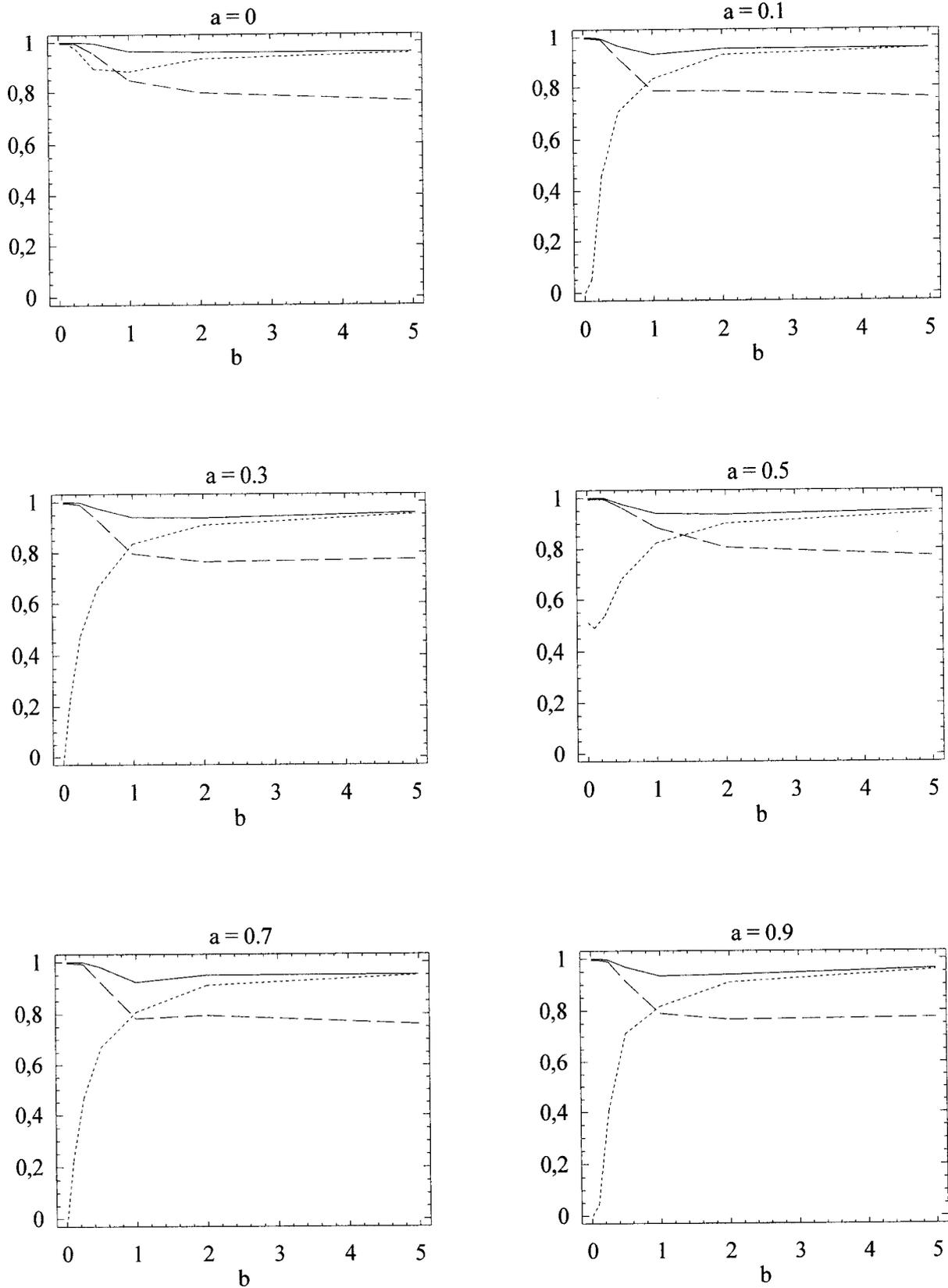
Tabelle 7: Quantile der Logistischen Verteilung für $(\hat{a} - a)/\hat{b}$ *Ergebnisse der Simulationsstudie*

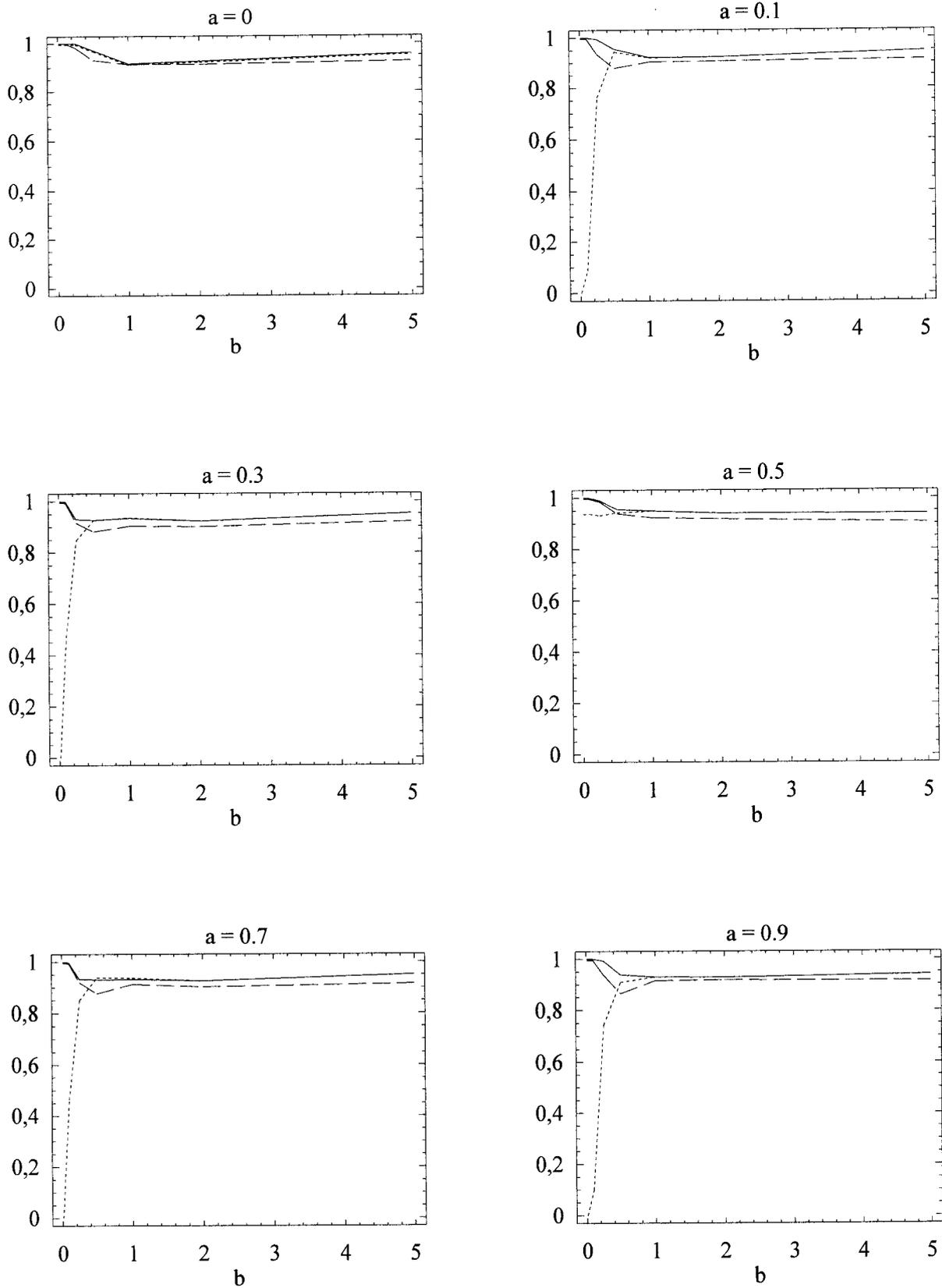
Ähnlich wie in den vorangehenden Abschnitten beobachtet, weist die ÜW der Pivot-Intervalle Defizite auf. Abgesehen von einer kleinen Spitze an der Stelle $a = 0.5$, die von Zufallstreffern hervorgerufen wird, ist das Pivot-Verfahren bei sehr kleinem Skalenparameter nicht imstande, den Lageparameter zu orten (vgl. Grafiken 33-38). Die ÜW erholen sich mit zunehmender Streuungsbreite. Das LQ-Verfahren schätzt bei kleinem b deutlich genauer, büßt aber mit wachsendem Skalenparameter an Präzision ein. Das Kombinationsverfahren eignet sich hingegen auch für Daten aus der Logistischen Verteilung. Auch hier gelingt durch eine rechtzeitige Umstellung von einem auf das andere Verfahren eine gleichmäßige Einhaltung des Niveaus. Über alle Stichprobenumfänge und Parametereinstellungen hinweg verläuft das Abbild der ÜW fast geradlinig auf der 95%-Marke. Dabei erhöht sich die Intervalllänge gegenüber der Vorlage durch das Pivot-Verfahren nur geringfügig. Im großen und ganzen bietet sich dem Betrachter das gleiche Erscheinungsbild wie bei der Normalverteilung. Allerdings:

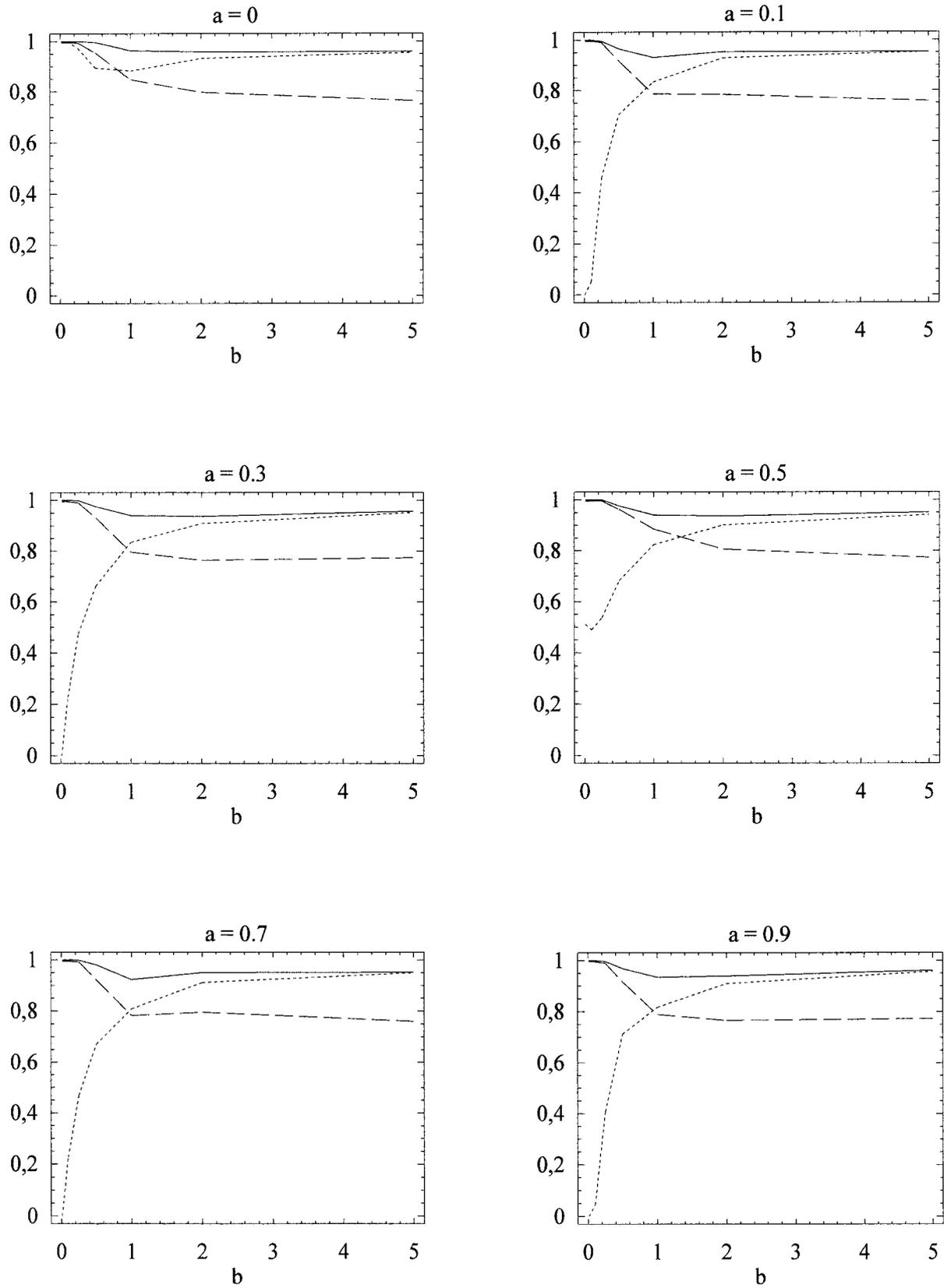
Der Erholungseffekt der Pivot-ÜW stellt sich etwas früher ein, also bei kleineren Werten von b und geringerem Stichprobenumfang. Eine Gegenüberstellung für $n = 2$ zeigt: Die bei $a = 0.1$ tief hängenden Zipfel heben sich rascher als bei der Normalverteilung. Bei $a = 2$ ist die Pivotlinie schon enger an die Niveaumarkierung herangerückt. Dem LQ-Verfahren ergeht es genau umgekehrt. Während sich die Linie der ÜW bei der Normalverteilung mit $\sigma = 1$ noch knapp unter der Marke hält und diese noch an einigen Stellen berührt, kann sie sich bei $a = 1$ der Logistischen Verteilung nur noch schwer an die 90% herantasten. Die meisten Werte liegen um die 80%. Das erklärt sich folgendermaßen: Verglichen mit der Standardnormalverteilung steigt die Dichte der Logistischen Standardverteilung zu ihrem Gipfel hin steiler an und läuft dafür zu den Enden hin länger aus. Sie besitzt bei gleichem Skalenparameter eine viel höhere Streuung. Das wirkt sich auf die Quantilgröße und folglich die Intervalllänge aus. Das 0.975-Quantil der Logistischen Pivotgröße ist fast doppelt so groß wie das t-Quantil und die Intervalle damit doppelt so lang. Eine kleine Erhöhung des Skalenparameters löst eine viel stärkere Streuung aus als eine Anhebung der Standardabweichung der Normalverteilung um den gleichen Betrag. Nach allen bisherigen Beobachtungen trägt eine stärkere Streuung immer zu guten Überdeckungswahrscheinlichkeiten des Pivot-Verfahrens bei. Gleichzeitig hindert sie das LQ-Verfahren an einer guten Intervallschätzung.

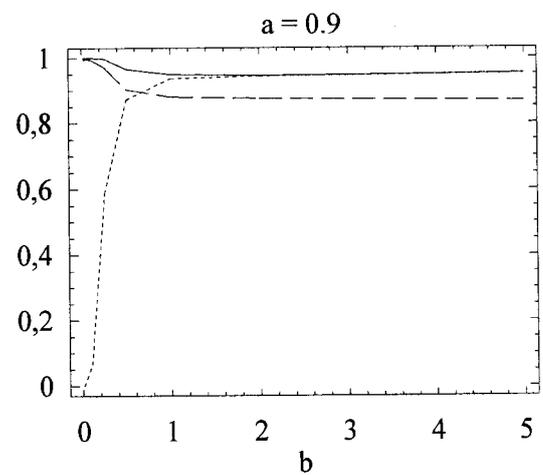
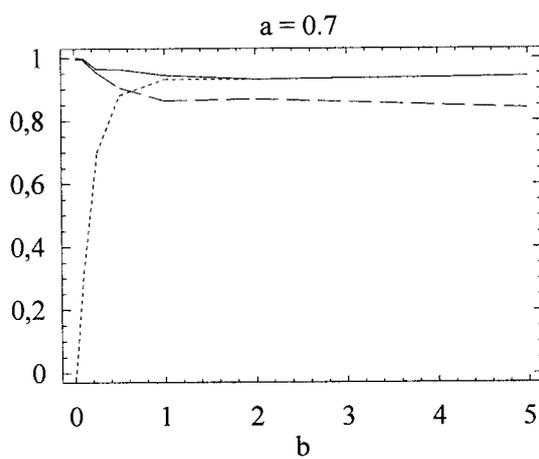
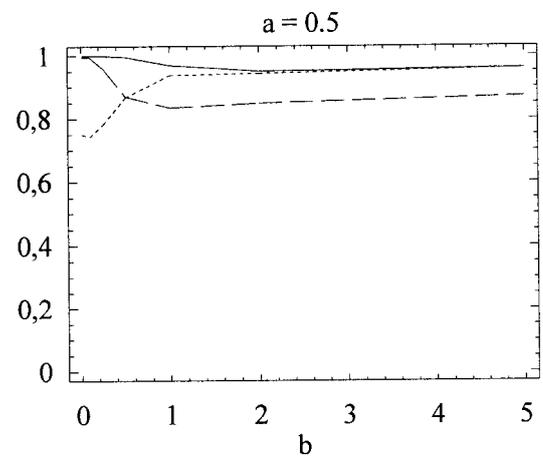
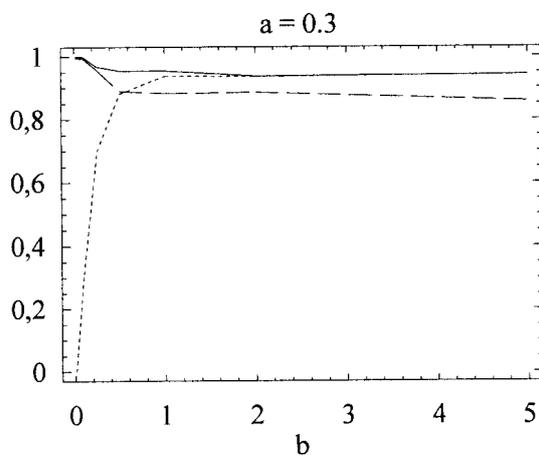
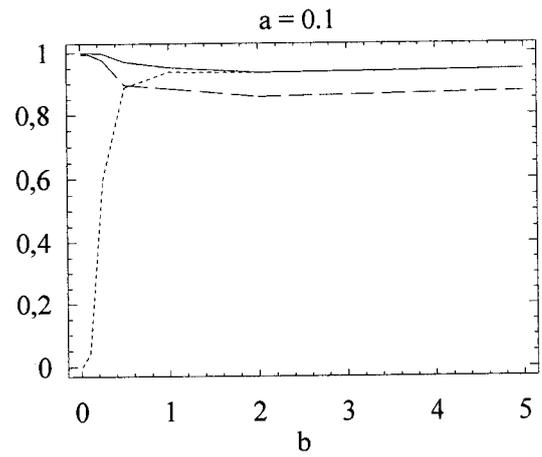
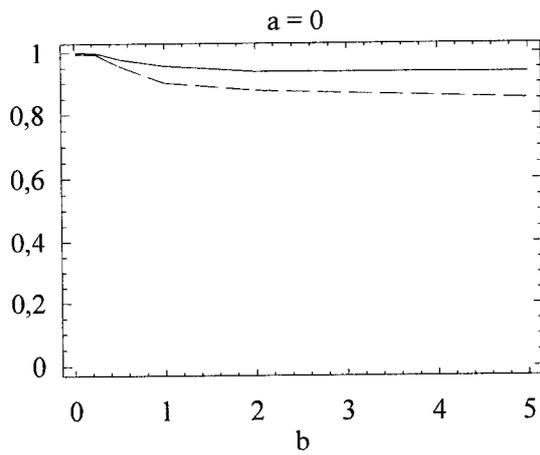
Empfohlene Kombination

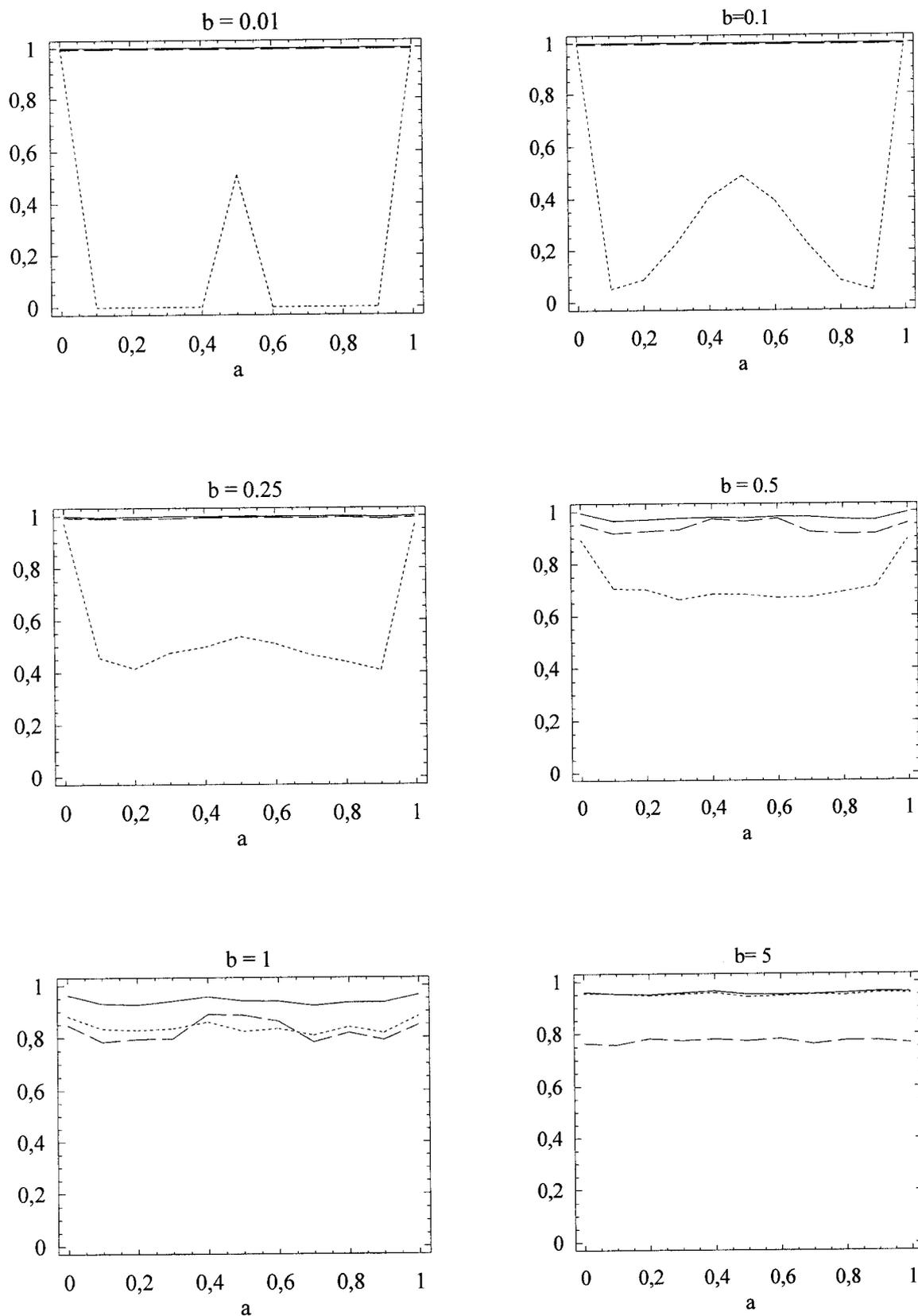
Ist der geschätzte Rundungsgrad größer als 2.21, sollte man die LQ-Methode benutzen. Ansonsten eignet sich das Pivot-Verfahren bei sehr kleinen Stichprobenumfängen besser. Ab $n = 10$ ist die LQ-Methode uneingeschränkt zu empfehlen. Pivot-Quantil-Tabellen sind nur schwer erhältlich und rechnergestützte Verfahren ohnehin unverzichtbar. Daher verliert man durch den Einsatz von LQ weder Rechenzeit noch Treffgenauigkeit und kommt obendrein ohne Quantiltabellen aus.

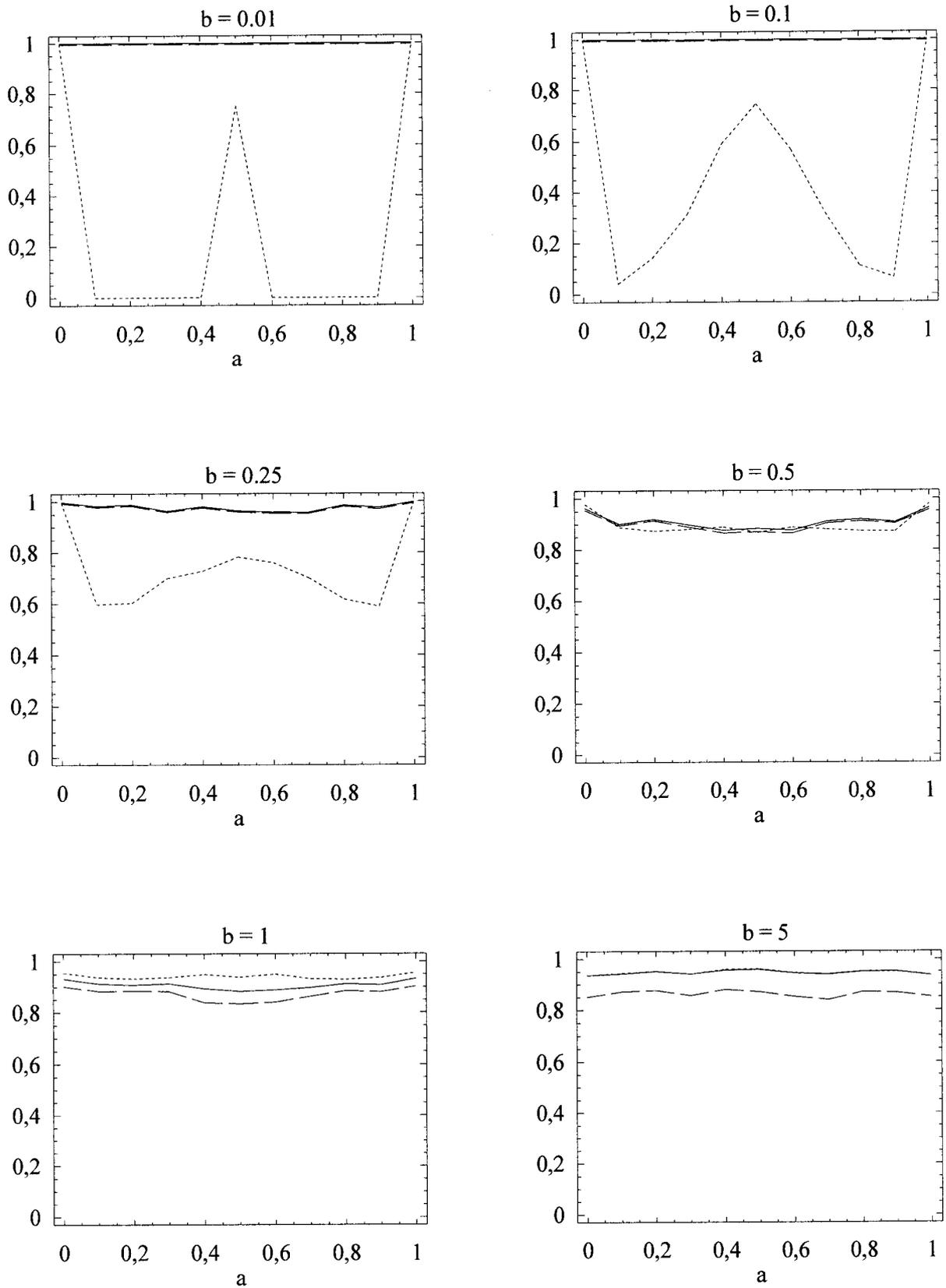
Überdeckungswahrscheinlichkeiten im Verlauf über b Abbildung 29: ÜW Logistische Verteilung, $n=2$, Verlauf über b

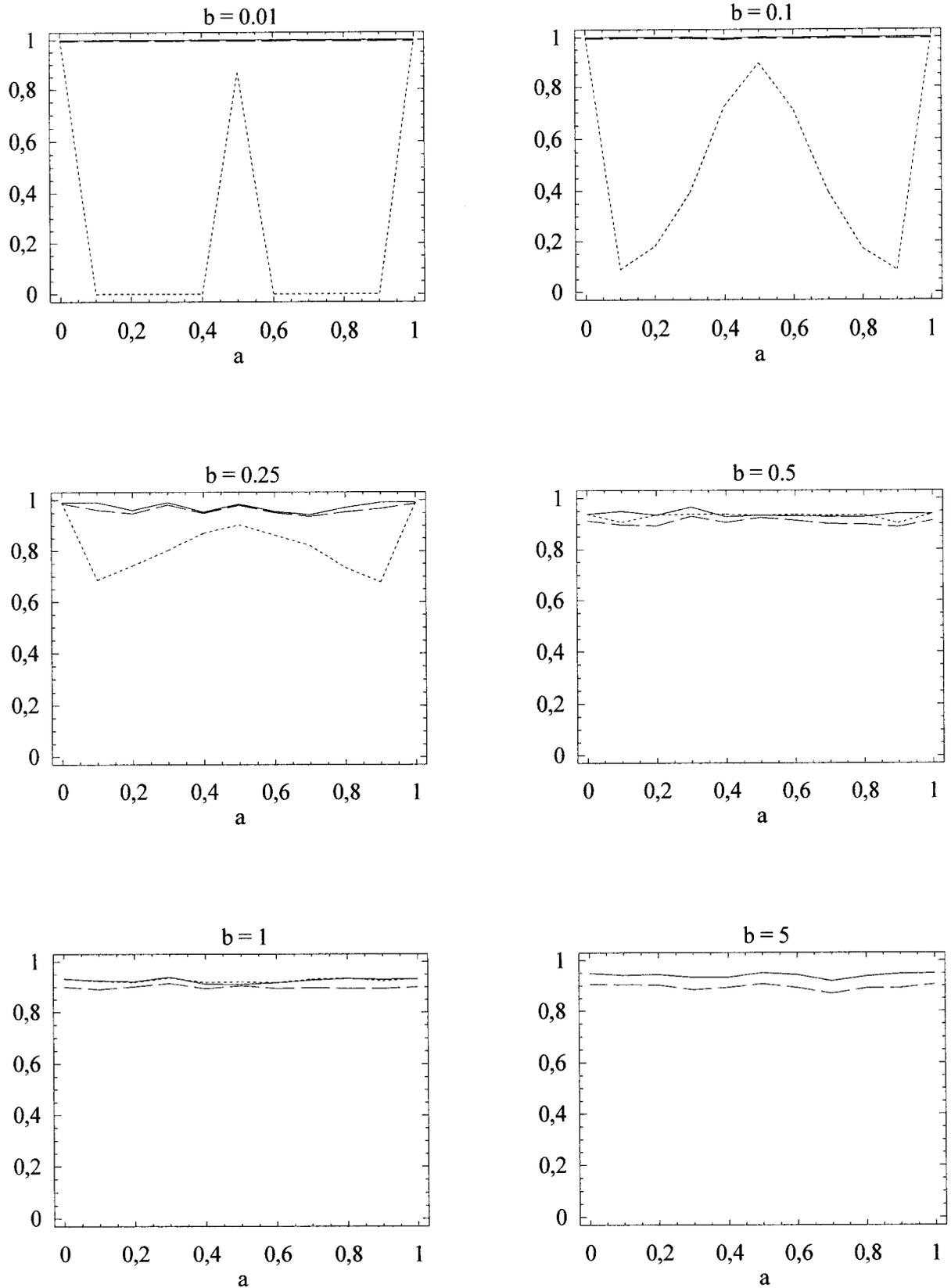
Abbildung 30: ÜW Logistische Verteilung, $n=5$, Verlauf über b

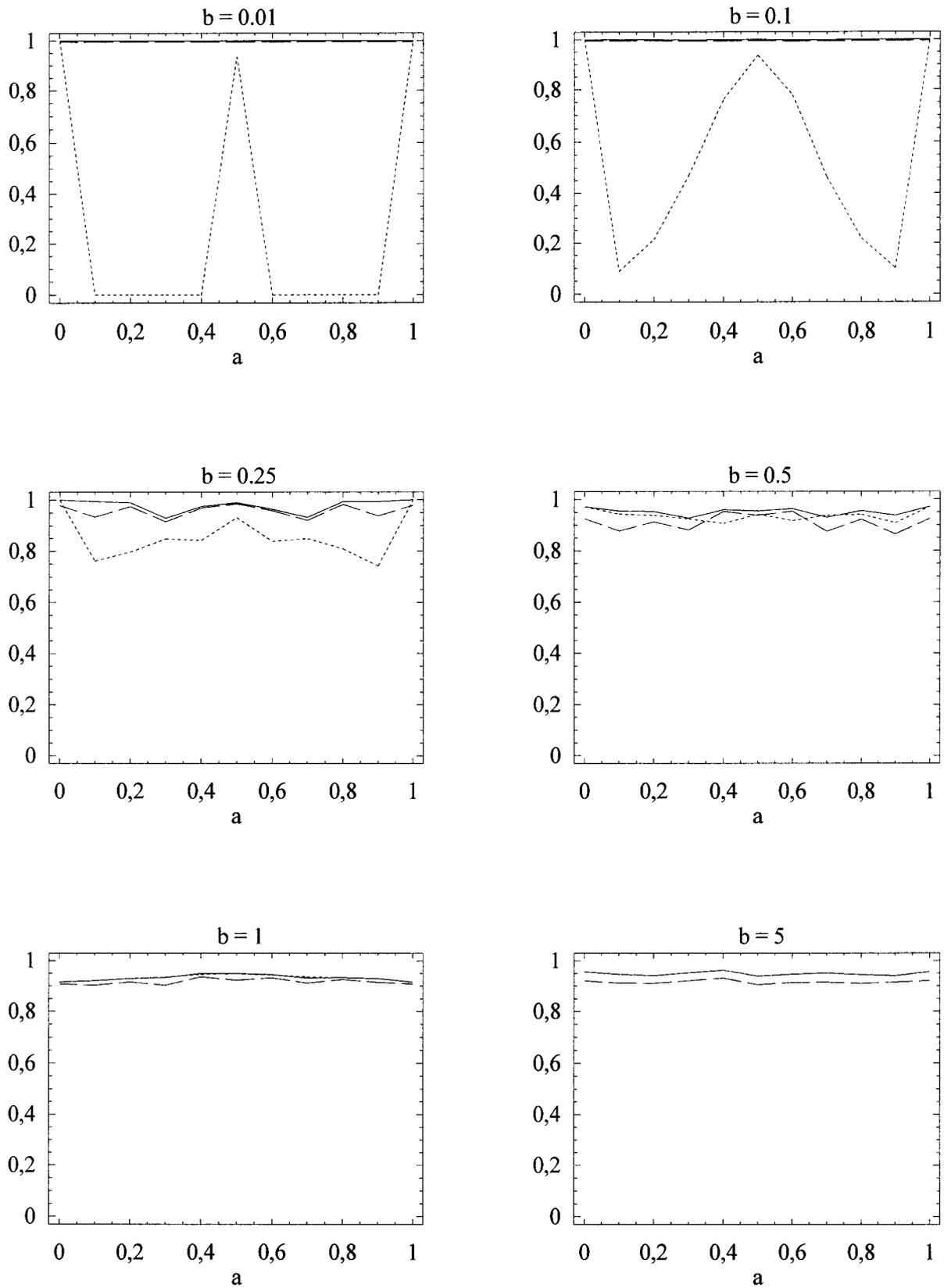
Abbildung 31: ÜW Logistische Verteilung, $n=10$, Verlauf über b

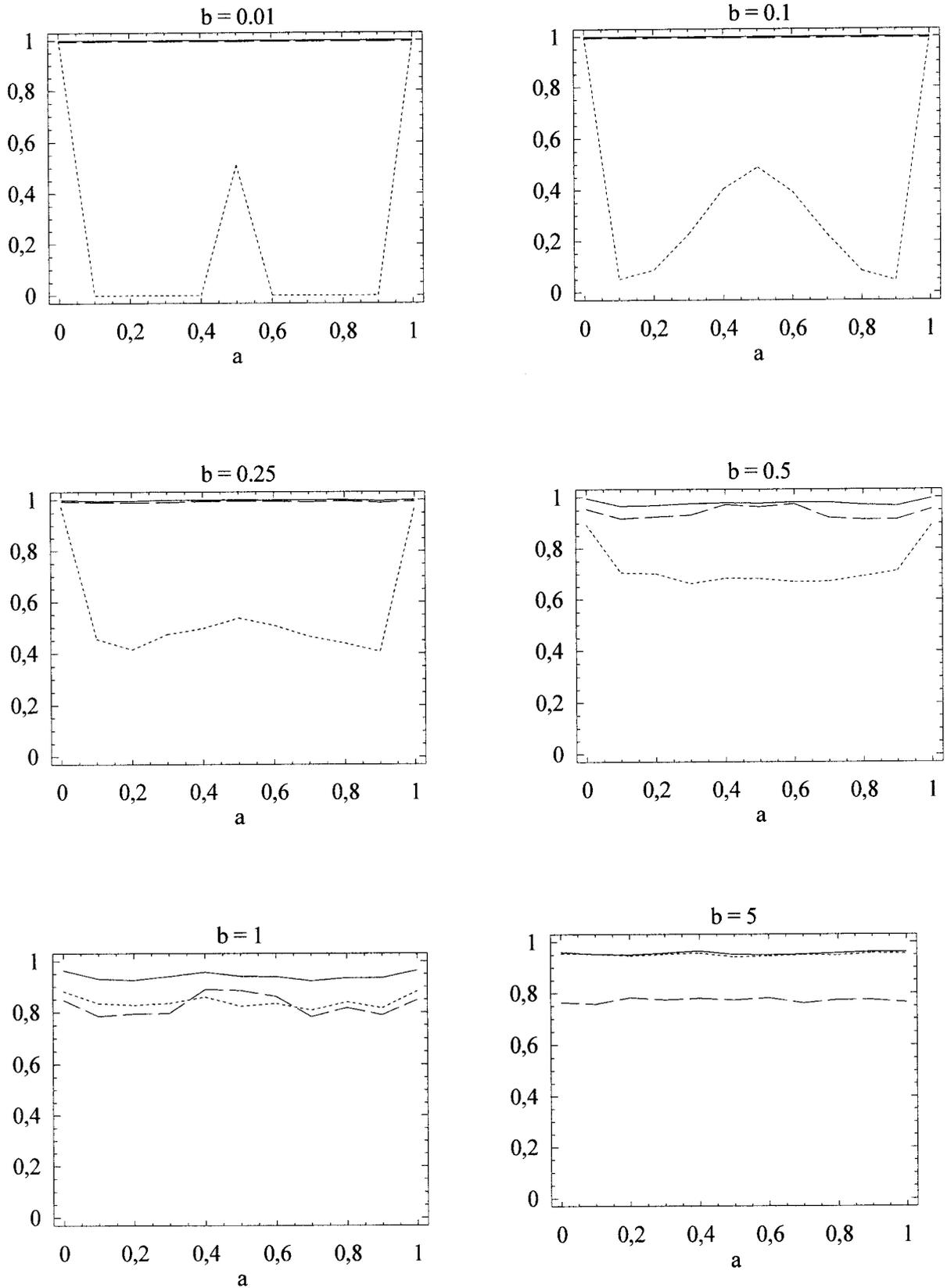
Abbildung 32: ÜW Logistische Verteilung, $n=15$, Verlauf über b

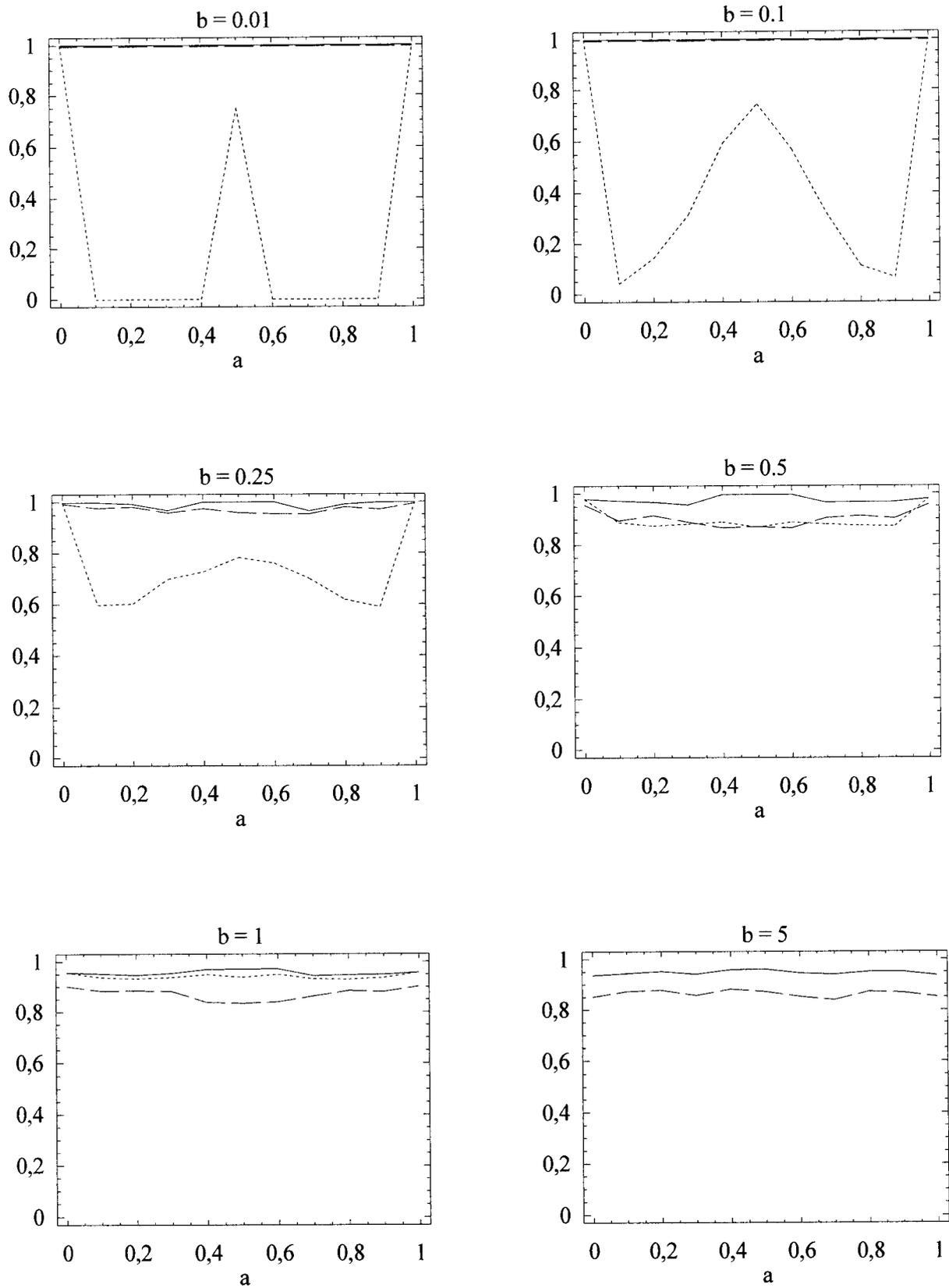
Überdeckungswahrscheinlichkeiten im Verlauf über a Abbildung 33: ÜW Logistische Verteilung, $n=2$, Verlauf über a

Abbildung 34: ÜW Logistische Verteilung, $n=3$, Verlauf über a

Abbildung 35: ÜW Logistische Verteilung, $n=4$, Verlauf über a

Abbildung 36: ÜW Logistische Verteilung, $n=5$, Verlauf über a

Abbildung 37: ÜW Logistische Verteilung, $n=10$, Verlauf über a

Abbildung 38: ÜW Logistische Verteilung, $n=15$, Verlauf über a

Durchschnittliche Intervalllängen von korrigierter LQ-, Pivotmethode und Kombinationsverfahren
Logistische Verteilung

$a = 0$															
b	0.01			0.25			0.5			1			2		
n	LQ_{korr}	Pivot	Kombi												
2	2.03	0	2.03	2.19	5.33	6.58	2.99	12.66	13.26	5.37	25.62	25.95	10.21	51.92	52.10
3	1.27	0	1.27	1.47	1.72	2.25	2.26	3.47	3.61	4.33	6.91	6.95	8.56	14.42	14.43
4	1.04	0	1.04	1.21	1.15	1.36	1.96	2.27	2.26	3.75	4.45	4.45	7.39	9.26	9.26
5	0.98	0	0.98	1.05	0.90	1.15	1.74	1.91	1.94	3.32	3.62	3.62	6.48	7.45	7.45
10	0.94	0	0.94	0.74	0.54	0.74	1.24	1.14	1.15	2.38	2.23	2.23	4.63	4.67	4.67
15	0.94	0	0.94	0.61	0.42	0.59	1.00	0.86	0.86	1.94	1.69	1.69	3.68	3.47	3.47

$a = 0.2$															
b	0.01			0.25			0.5			1			2		
n	LQ_{korr}	Pivot	Kombi												
2	2.03	0	2.03	2.21	5.85	7.04	3.12	13.28	13.88	5.09	24.16	24.51	9.88	50.02	50.21
3	1.27	0	1.27	1.47	1.72	2.28	2.29	3.50	3.65	4.55	7.31	7.34	8.72	14.72	14.74
4	1.04	0	1.04	1.20	1.21	1.36	1.92	2.24	2.24	3.66	4.32	4.32	7.33	9.20	9.20
5	0.98	0	0.98	1.02	0.98	1.17	1.72	1.91	1.93	3.38	3.70	3.70	6.61	7.59	7.59
10	0.94	0	0.94	0.71	0.62	0.74	1.21	1.14	1.14	2.34	2.19	2.19	4.48	4.52	4.52
15	0.94	0	0.94	0.59	0.48	0.58	1.01	0.87	0.87	1.92	1.67	1.67	3.76	3.55	3.55

$a = 0.5$															
b	0.01			0.25			0.5			1			2		
n	LQ_{korr}	Pivot	Kombi												
2	2.12	6.64	7.64	2.23	7.43	8.37	2.96	12.31	12.96	5.18	24.67	25.03	9.85	49.75	49.95
3	1.38	2.02	2.33	1.48	2.23	2.50	2.27	3.43	3.60	4.34	6.992	6.97	8.50	14.33	14.34
4	1.09	1.34	1.35	1.22	1.50	1.49	1.91	2.23	2.22	3.86	4.57	4.57	7.35	9.24	9.24
5	0.93	1.12	1.19	1.04	1.24	1.30	1.74	1.94	1.96	3.40	3.72	3.72	6.50	7.47	7.47
10	0.58	0.72	0.74	0.67	0.79	0.80	1.22	1.15	1.15	2.36	2.21	2.21	4.53	4.57	4.57
15	0.44	0.56	0.57	0.54	0.61	0.62	0.99	0.86	0.86	1.94	1.69	1.69	3.68	3.47	3.47

Tabelle 8: Durchschnittliche Intervalllängen; Logistische Verteilung

5.4 Konfidenzintervalle für die Exponentialverteilung

Die vorhergehenden Kapitel haben Stärken und Schwächen der gebräuchlicher KI-Verfahren demonstriert. Fehlende Adaptionsmöglichkeiten eines von Lee und Varde-man vorgeschlagenen Alternativverfahrens an eine große Schar von Dichten mit Lage- und Skalierungsparameter machen Neuentwicklungen notwendig. Intensive Fallstudien am Beispiel von Intervallschätzungen des Lageparameters inspirieren dabei zu einer Kombinationsstrategie. Die prinzipielle Arbeitsweise der neuen Strategie und ihre Auswirkungen auf die Überdeckungswahrscheinlichkeiten wurden bereits beispielhaft an einigen bedeutenden Verteilungen mit Lage- und Skalierungsparameter vorgeführt. Andere Verteilungen (vgl. 4.1) sind zum Teil über Transformationen eng mit den behandelten Verteilungen verquickt und die Anwendung gleicher Verbesserungsstrategien über das Kombinationsverfahren scheint viel versprechend. Doch eine Übertragung bisheriger Schlüsse ist voreilig. Die Transformation der Daten birgt zahlreichen Tücken in sich, die in einem gesonderten Kapitel zur Sprache kommen. Eine Erörterung der Verteilungen, die mit Transformationen in direktem Zusammenhang stehen (wie etwa Lognormal-, Weibull-, und Paretoverteilung), wird vorerst aus dem Abschnitt der Lageparameter ausgeklammert und erst in Kapitel 6 diskutiert.

Nach den Vorüberlegungen aus Abschnitt 4.1 steht noch eine Untersuchung der Exponentialverteilung aus. Mit der Exponentialverteilung rücken erstmalig neue Aspekte in den Vordergrund. Stand vorher der Lageparameter im Mittelpunkt des Interesses, gilt nun dem Skalenparameter vermehrte Aufmerksamkeit, denn er charakterisiert simultan Erwartungswert und Varianz der Verteilung. In der Lebenszeitanalyse und der Warteschlangentheorie, den zentralen Anwendungsgebieten der Exponentialverteilung, spielt er eine entscheidende Rolle bei der Beschreibung von erwarteter Lebensdauer oder durchschnittlicher Wartezeit (Nelson, 1977).

Vielfach ist ein einparametriges Modell ohne Verschiebungsparameter ausreichend:

$$f(x, \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} I_{[0, \infty)}, \quad \lambda > 0. \quad (5.20)$$

Erwartungswert und Varianz sind

$$E(X) = \lambda \quad \text{und} \quad Var(X) = \lambda^2. \quad (5.21)$$

Mit dieser Parametrisierung ergibt sich als MLS für exakte Daten das Stichprobenmittel $\hat{\lambda} = \bar{X}$ nach Maximierung der Likelihoodfunktion. Aus Abschnitt 4.1 ist bekannt, dass die Pivotgröße $2\lambda \sum_{i=1}^n X_i$ χ^2 -verteilt ist mit $2n$ Freiheitsgraden.

Das Pivot-Konfidenzintervall für exakte Daten ist dann

$$KI_\lambda = \left[\frac{2 \sum_{i=1}^n X_i}{\chi_{(2n; 1-\alpha/2)}^2} ; \frac{2 \sum_{i=1}^n X_i}{\chi_{(2n; \alpha/2)}^2} \right]. \quad (5.22)$$

Eine abgeänderte Likelihood, die die Intervallwahrscheinlichkeiten auf die Beobachtungswerte zentriert, kann Rundungsmechanismen entsprechend berücksichtigen. Ganzzahliges Runden führt zu der Likelihood:

$$L(\lambda) = \prod_{i=1}^n \left[e^{-\left(\frac{x_i-0.5}{\lambda}\right)} - e^{-\left(\frac{x_i+0.5}{\lambda}\right)} \right]. \quad (5.23)$$

Nach Kulldorffs (1961) Hinweis zur ML-Schätzung gruppierter Daten existiert der MLS nur dann, wenn nicht alle Beobachtungswerte in das erste Intervall fallen. Bisher erlaubten alle Verteilungen Werte aus dem gesamten reellen Zahlenbereich, und der MLS war beim Runden mit konstanter Gitterbreite stets berechenbar. Nicht einmal ein einziger Beobachtungswert wäre jemals in das erste Intervall mit $-\infty$ als linke Intervallgrenze gefallen. Die Exponentialverteilung aber schränkt den zulässigen Wertebereich auf die positive Achse ein. Zum ersten Mal sind bei kleinen λ -Werten Stichproben bestehend aus lauter Nullen zu beobachten, die eine nicht schätzbare Situation auslösen. Die Wahrscheinlichkeit dafür, eine einzige Null zu beobachten, liegt beim ganzzahligen Runden bei $1 - \exp^{-0.5/\lambda}$. Obwohl sich eine Punktschätzung verbietet, ist eine Intervallschätzung dadurch aber nicht beeinträchtigt. Mit dem Wissen um die untere Schranke, nämlich links offen zur Null, genügt mit dem LQ-Verfahren die rechtsseitige Suche nach einer Obergrenze.

Eine Vorstudie soll zunächst einen Eindruck von den Auswirkungen auf die Intervallschätzung vermitteln, die eine Vernachlässigung des Rundens nach sich zieht. Dazu schätzen Pivot- und LQ-Methode jeweils ein Intervall aus den simulierten Stichproben von Rundungswerten. Im Vergleich der Überdeckungswahrscheinlichkeiten würde man erwarten, dass eine Vernachlässigung um so stärker ins Gewicht fällt, je kleiner der wahre Parameter und der Stichprobenumfang. Aufgrund der kleinen Varianz λ^2 liegen dann die Werte nahe beieinander und Ausreißer treten selten auf. Eine Durchschnittsbildung der Werte, wie sie der MLS verwendet, verliert ihre ausgleichende Wirkung, denn gerade diese Ausreißer könnten den Schätzwert in Richtung des wahren Parameters rücken. Die Palette der Parameter sollte sich daher in der Simulationsstudie auf den kleinen bis moderaten Bereich erstrecken. Der Stichprobenumfang wird mit $n = (2, 3, 4, 5, 10, 15)$ beibehalten, und λ variiert engmaschig zwischen (0.01 und 10). Ein Zwischenergebnis ist in die Grafik 40 aufgenommen. Die gestrichelte Linie markiert die Überdeckungswahrscheinlichkeiten des LQ-Verfahrens, die punktierte Linie steht für die ÜW des Pivot-Verfahrens. Die

durchgezogenen Linie nimmt bereits die ÜW des Kombinations-Verfahrens vorweg, die sich in diesem Fall mit dem Endergebnis der korrigierten LQ-Version deckt.

Wie vorausgesagt, sind die ÜW des Pivot-Verfahrens umso schlechter, je näher der Parameter der Null. Dann streuen die Stichprobenwerte kaum, und das Pivot-Verfahren schätzt den Skalenbereich viel zu klein. Erst mit höherem Parameter kann das Pivot-Verfahren das Niveau einhalten. Eine Erhöhung des Stichprobenumfanges wirkt sich ebenfalls vorteilhaft aus, weil dann die ÜW steiler ansteigen und das Niveau bei kleinerem Parameter erreicht wird. Das LQ-Verfahren kann bei sehr kleinem Skalenparameter mit außerordentlich guten ÜW aufwarten, allerdings sind in moderaten Bereichen abrupte Einbrüche zu verzeichnen. An der Stelle $\lambda = 0.7$ ist für $n = 3$ der Tiefpunkt mit 85% ausgewiesen. Dieser Tiefpunkt wandert mit zunehmendem Stichprobenumfang in Richtung der Null, gewinnt dabei aber mehr und mehr an Höhe bis schließlich mit $n = 15$ keine Schwankungen mehr zu registrieren sind. Dennoch scheint auch ein kurzfristiges Abfallen um mehr als 10% nicht annehmbar.

Leider kann man zu diesem Zeitpunkt nicht auf Synergieeffekte durch eine Nachahmung des Kombinationsverfahrens bauen. Die Exponentialverteilung in obiger Form besitzt nur einen einzigen Parameter und so ist schwerlich eine Schwellenwertbedingung an einen anderen Parameter zu knüpfen. Einen Schwellenwertbedingung für den zu schätzenden Parameter selbst aufzustellen, würde ein außergewöhnlich präzises Punktschätzverfahren bedingen, denn der Kreuzungsbereich der Kurvenverläufe ist eng gesteckt. Ein detaillierter Vergleich von einzelnen Stichproben und Schätzwerten zeigt, dass in der Hauptsache die „Null-Stichproben“ für die schlechten ÜW verantwortlich sind. Beispielsweise wird $\lambda = 0.8$ vom LQ-Verfahren aus dem Konfidenzbereich einer Null-Stichprobe ausgeschlossen. Die Anzahl solcher Null-Stichproben ist aber so hoch, dass kein Verfahren sich Fehleinschätzungen erlauben kann, ohne damit unter das Niveau zu sinken.

Die gleiche Korrektur, die sich schon bei den anderen Verteilungen als nützlich erwiesen hat, könnte auch jetzt einen Ausweg eröffnen: das Konfidenzintervall sollte in diesem Spezialfall alle λ -Werte einschließen, die Stichproben mit einer Häufigkeit von mehr als α erzeugen.

Bestimmung der Korrekturfaktoren

Die Likelihood einer Nullstichprobe ist

$$L(\lambda \mid x_i = 0, i = 1, \dots, n) = \left[1 - e^{-\left(\frac{0.5}{\lambda}\right)} \right]^n. \quad (5.24)$$

Sie steigt über das Niveau von $\alpha = 0.05$ an den Stellen Δ_n .

Das LQ-Konfidenzintervall in diesen Fällen wird ersetzt durch

$$KI_{\lambda}(x_i = 0, i = 1, \dots, n) = (0, \Delta_n]. \quad (5.25)$$

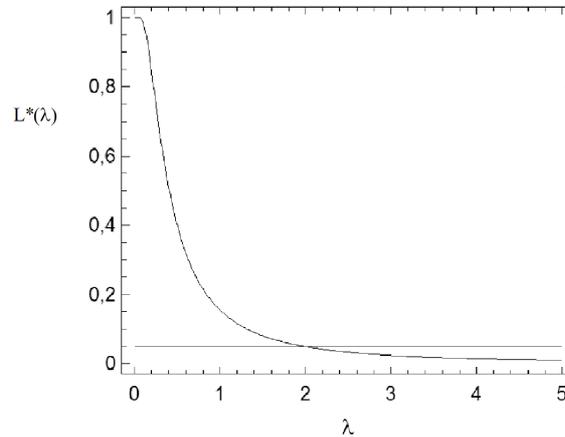


Abbildung 39: Likelihood der Exponentialverteilung für die Stichprobe (0,0)

n	2	3	4	5	10	15
Δ_n	1.97	1.08	0.78	0.62	0.36	0.29

Tabelle 9: Korrekturen für die Exponentialverteilung

Das Ergebnis der korrigierten LQ- Version ist bereits in Abb. 40 eingefügt. Die Korrekturversion hat wie erhofft die ÜW in den heiklen Bereichen annähernd auf das nominale Niveau angehoben. In weiten Teilen (ab $\lambda = 0.3$) vergrößern die Korrekturmaßnahmen die Intervalllängen nur unwesentlich. Die Intervalllänge der Korrekturversion liegt ab $\lambda = 0.6$ sogar unter der Länge der Pivot-Intervalle, deren ÜW sich an dieser Stelle noch weit vom Niveau entfernt.

Bestimmung des Schwellenwertes

Erst mit den ergänzten Korrekturen lässt sich eine Schwellenwertbedingung für den Rundungsgrad finden. Diese Bedingung strebt keine Leistungssteigerung der Verfahren an – das LQ ist in jedem Falle gut genug –, garantiert aber dem Anwender die Qualität seiner Analysen. Angesichts der immensen Zahl von Studien, die Rundungseffekte nicht bedacht haben, und der Tragweite bei großen Rundungsgraden, erscheint ein Richtmaß hilfreich. In Anlehnung an die Vorgehensweise bei Lageparametern bietet sich eine Prüfung der Parameterwerte im Kreuzungsbereich an. Ein Schwellenwert von $\hat{\lambda} = 1$ führt zum gewünschten Erfolg. Die ÜW des korrigierten LQ können aufrecht erhalten werden, und das Verfahren leidet bei kleinerem

Rundungsgrad das in der Regel weniger rechenaufwändige Pivot-Verfahren ein. In Rundungsgraden ausgedrückt:

$$\hat{r} = \frac{1}{\hat{\sigma}} = \frac{1}{\hat{\lambda}} < 1. \quad (5.26)$$

Empfehlung

Das Pivot-Verfahren ist unbedenklich, falls der geschätzte Rundungsgrad unter 1 liegt. Für möglichst kleine Konfidenzintervalle ist in jedem Fall das korrigierte LQ-Verfahren vorzuziehen.

Überdeckungswahrscheinlichkeiten

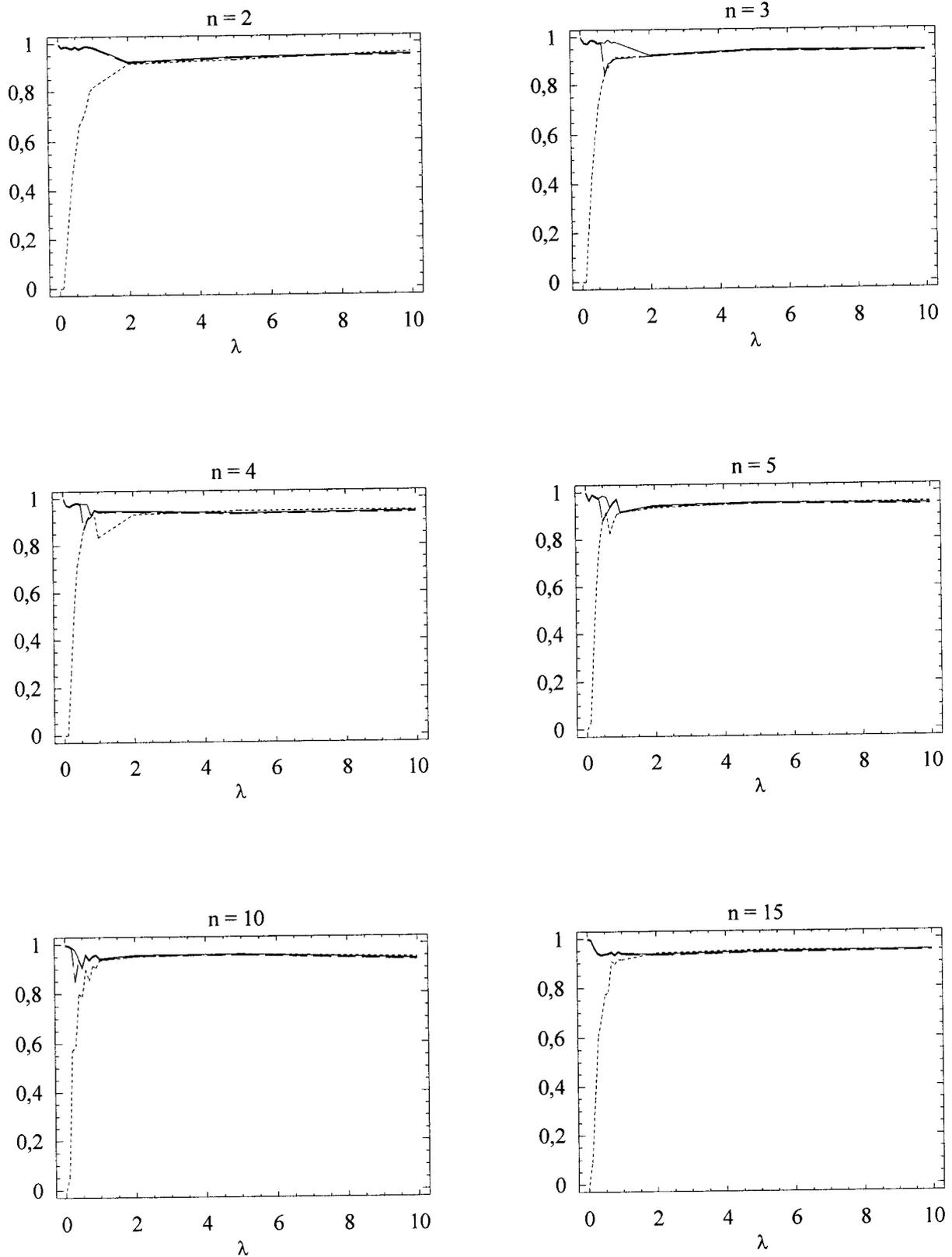


Abbildung 40: Überdeckungswahrscheinlichkeiten Exponentialverteilung

**Durchschnittliche Intervalllängen von korrigierter LQ-, Pivotmethode und Kombinationsverfahren
Exponentialverteilung**

λ	0.01			0.1			0.2			0.3			0.4		
n	Pivot	LQ _{korrr}	Kombi												
2	0	1.97	1.97	0.06	1.99	1.99	0.67	2.20	2.20	1.67	2.58	2.58	2.58	3.02	3.02
3	0	1.08	1.08	0.03	1.09	1.09	0.42	1.25	1.25	0.83	1.44	1.44	1.40	1.75	1.77
4	0	0.78	0.78	0.02	0.79	0.79	0.26	0.89	0.89	0.65	1.08	1.08	0.98	1.27	1.27
5	0	0.62	0.62	0.02	0.63	0.63	0.21	0.71	0.71	0.48	0.85	0.85	0.81	1.05	1.06
10	0	0.36	0.36	0.01	0.36	0.36	0.12	0.40	0.40	0.29	0.50	0.50	0.47	0.63	0.63
15	0	0.29	0.29	0.01	0.29	0.29	0.09	0.32	0.32	0.23	0.39	0.39	0.36	0.49	0.49

λ	0.6			0.8			1			2			5		
n	Pivot	LQ _{korrr}	Kombi												
2	4.40	3.95	4.43	5.77	4.78	5.50	7.44	5.82	7.05	15.72	11.47	15.33	39.67	28.55	39.54
3	2.31	2.31	2.39	3.35	3.04	3.28	4.24	3.70	4.12	8.48	7.01	8.39	22.04	18.01	22.02
4	1.71	1.78	1.81	2.34	2.23	2.35	3.10	2.84	3.08	6.43	5.59	6.40	16.07	13.89	16.06
5	1.37	1.46	1.48	1.95	1.91	1.98	2.49	2.35	2.50	5.18	4.64	5.17	12.92	11.52	12.92
10	0.80	0.89	0.89	1.14	1.17	1.19	1.44	1.42	1.47	2.96	2.76	2.96	7.58	7.14	7.58
15	0.61	0.70	0.70	0.87	0.91	0.92	1.10	1.10	1.13	2.29	2.16	2.29	5.77	5.51	5.77

Tabelle 10: Durchschnittliche Intervalllängen, Exponentialverteilung

6 Transformationen oder Schätzung des Skalenparameters

Die Intervallschätzung des Skalenparameters ist mit ganz eigenen Problemen behaftet, auf die man nicht mit dem Kombinationsverfahren antworten kann. In Modellen ohne Lage-, also ohne Störparameter, lässt sich keine Schwellenwertbedingung an dessen Schätzung binden. Eine Schwellenwertbedingung, die an den zu schätzenden Parameter geheftet ist, verliert ihren Zweck. Sie selbst bedingt eine verlässliche und präzise Schätzung, die angesichts der zu untersuchenden Stichprobenumfänge nicht zu erwarten ist.

Modelle mit Lageparameter entziehen sich aus anderen Gründen einer Schwellenwertbedingung. Aus den Analysen von Lee und Vardeman über die Varianz der Normalverteilung sowie aus einer Prüfung der Exponentialverteilung mit Verschiebung geht hervor, dass die tatsächliche Lage kaum einen Einfluss auf die Überdeckungswahrscheinlichkeiten ausübt. Eine Vorabschätzung liefert daher keinerlei Information zur Wahl der Methodik. Qualitätsmängel lassen sich nicht mit einer Schwellenwertbedingung beheben, sondern nur durch Korrekturintervalle für kleine Skalenparameter. Diese gewinnt man aus der Profil-Likelihood für gerundete Daten. Im Gegensatz zur Intervallschätzung des Lageparameters reicht eine Korrektur bei Stichproben mit Spannweite Null nicht aus für eine Anhebung der ÜW. Weitere Stichproben mit höherer Spannweite leiten zu unrealistisch kleinen Skalenintervallen und das mit inakzeptabler Häufigkeit. Zur Schätzung des Skalenparameters benötigen alle diejenigen Stichproben ein Korrekturintervall, die mit einer Wahrscheinlichkeit von mehr als Niveau auftreten und Konfidenzintervalle erzeugen, die den Parameter verfehlen. Alle Korrekturintervalle müssen so aufeinander abgestimmt sein, dass die Fehleinschätzungen in ihrer Gesamtheit über alle Korrekturintervalle das Niveau nicht übertreten.

Diese Aufgabe bringt mehrere Schwierigkeiten mit sich:

- Ein Richtwert für den Skalenschätzer muss definiert werden, unterhalb dessen Korrekturintervalle zu verwenden sind.
- Eine Ausrichtung einzig und allein nach der Spannweite scheint bei größeren Stichproben willkürlich und zu wenig differenziert.
- In Bezug auf die Aufteilung der Wahrscheinlichkeitsmasse α sind unterschiedliche Gewichtungen denkbar. Damit ist die Festsetzung der Korrekturintervalle nicht eindeutig, auch wenn man typischerweise von jedem einzelnen Korrekturintervall gleiche Genauigkeit einfordern wird.
- Eine Erhöhung des Stichprobenumfanges induziert zunehmend feinere Abstufungen an Skalenschätzern. Der Bereich „kleiner“ Varianzen wird immer weiter

zerstückelt, und jede mögliche Realisation von Stichprobenvarianz braucht ihre eigene Korrektur.

- Schließlich münden die Korrekturen in ein umfangreiches und unübersichtliches Tafelwerk, das für den täglichen Praxiseinsatz eher ungeeignet ist.

Einen interessanten Sonderfall stellen logarithmisch transformierbare Verteilungsmodelle dar, wie Lognormal-, Weibull- oder Paretoverteilung. Die Transformation verwandelt den ursprünglichen Skalenparameter in den Lageparameter. Der im Exponenten befindliche Parameter – wegen seiner formgebenden Eigenschaft meist Formparameter genannt – wird zum neuen Skalenparameter. Auf diese Weise gehen Skalenparameter von Lognormal- und Weibull- in Lageparameter von Normal- und Minimumverteilung über. Ähnliches gilt für den Formparameter der Paretoverteilung, den die Transformation in den Skalenparameter der Exponentialverteilung überführt. Konfidenzintervalle für einige dieser Parameter wurden zwar in den vorangegangenen Kapiteln besprochen, aber die Transformation hat einige Haken:

Durch das Runden können künstlich Werte außerhalb des originalen Definitonsbereiches entstehen (Nullen-Problematik), die sich mit einer anschließenden Transformation nicht mehr vertragen. Dann müsste der Transformation unter Umständen eine Verschiebung vorausgehen. Zum anderen verformt eine nichtlineare Transformation das ursprünglich äquidistante Rundungsgitter. Der Rundungsgrad, nun keine fixe Größe mehr, bietet im transformierten Gitter keine Entscheidungshilfe für das beste Verfahren. Es könnte einfacher sein, die Daten untransformiert zu belassen und dann eine Schwellenwertbedingung aufzustellen, die auf den Rundungsgrad der Originaldaten Bezug nimmt. Alternativ könnte ein Rundungsgitter helfen, das von vorn herein nicht äquidistant sondern in exponentiellen Abständen konzipiert ist. Dazu ein kurzes Gedankenspiel anhand eines von Ismaïl (2004) protokollierten Datensatzes einer Paretoverteilung: Ismaïl untersucht die Einkommensverteilung für 1000 US-Haushalte mit einem Einkommen über 10000 \$. Die Höhe des Einkommens liegt nicht exakt, sondern in Gruppierungen vor, die in Tabelle 11 angegeben sind.

10 000-	zum Vergleich:	$e^9 \approx 8\ 103$
15 000-		$e^{9.5} \approx 13\ 359$
25 000-		$e^{10} \approx 22\ 026$
50 000-		$e^{11} \approx 59\ 874$
100 000-		$e^{11.5} \approx 98\ 715$
150 000-		$e^{12} \approx 162\ 754$
$\geq 500\ 000$		$e^{13} \approx 442\ 413$

Tabelle 11: Einteilung in Einkommensklassen

Die vorgegebene Einteilung ist nahezu exponentiell. So ungewohnt die Gruppengrenzen der rechten Spalte auf den ersten Blick anmuten, beide Staffellungen sind gleichermaßen künstlich und nach ähnlichem Prinzip vorgenommen. Mit der exponentiellen Einteilung aber lassen sich die Daten bequem in eine Exponentialverteilung überleiten und bearbeiten. Die transformierten Daten sind nach einem äquidistanten Rundungsschema einsortiert, und das ermöglicht eine Einschätzung des Rundungsgrades.

Schwierigkeiten ganz anderer Art bereitet die Schätzung des Lageparameters von Weibull- und Lognormalverteilung. Ist ein Verschiebungsparameter in das Weibullmodell involviert, existieren die MLS und sind eindeutig nur dann, wenn wenigstens einer der beiden Parameter, Lage a oder Form c , bekannt ist. Sobald alle drei Parameter unbekannt sind, besitzt die Likelihoodfunktion keine Grenze für $a \rightarrow X_{(1)}$. Setzt man wenigstens $c \geq 1$ voraus, bleibt die Likelihood begrenzt. Ähnliches gilt für die dreiparametrische Lognormalverteilung, $\ln(X - a) \sim N(\mu, \sigma^2)$. Ebenfalls für $a \rightarrow X_{(1)}$ strebt $\hat{\mu}$ gegen $-\infty$ und $\hat{\sigma}^2$ gegen ∞ . Griffiths (1980) resümiert einige Techniken zur Intervallschätzung für die dreiparametrische Lognormalverteilung und zitiert einige bemerkenswerte Vorschläge, wie etwa Giesbrecht und Kempthorne (1976):

Sie gaben zu bedenken, dass viele Probleme, wie die der unbegrenzten Likelihood gelöst werden könnten, wenn man sich eingesteht, dass die Verteilung einer tatsächlich beobachteten Zufallsvariablen niemals wirklich stetig, sondern immer diskret ist. Unter dem diskreten Modell wird angenommen, dass die aufgezeichnete Beobachtung y nur den wahren Datenwert repräsentiert, der irgendwo im Intervall der Weite d um y herum liegt. Die Likelihood in diesem multinomial diskretisierten Modell ist für alle Parameterwerte begrenzt.

Das Modell für gerundete Daten macht demnach eine Parameterschätzung überhaupt erst praktikabel. Als Konsequenz fehlt dem Kombinationsverfahren die Pivot-Alternative, weil ein Verfahren auf exakter Datenbasis nicht ohne Einschränkung existiert. Umso wichtiger wäre gleichzeitig für das LQ-Verfahren, auch bei kleinen Stichprobenumfängen gute Schätzqualitäten aufrecht zu erhalten.

7 Ein Anwendungsbeispiel aus der Hydrologie

7.1 Das „Jahrhundert“-Hochwasser der Elbe

Eine meteorologisch außergewöhnliche Wetterlage im August 2002 löste im Flusssystem der oberen Elbe katastrophale Überschwemmungen aus. Ein „Vb-Tief“ initiierte andauernde, ergiebige Niederschläge, wie sie zuvor in Deutschland noch nicht registriert worden waren. Im Erzgebirge fielen gebietsweise mehr als 300 mm Regen an einem Tag. Am 12. August wurde an der Station des deutschen Wetterdienstes in Zinnwald-Georgenfeld die Rekordniederschlagshöhe von 352,7 mm in der Zeitspanne von 5:00 Uhr mitteleuropäischer Sommerzeit bis zur selben Uhrzeit des Folgetages gemessen. Die extrem hohen Tagesniederschläge trafen auf einen bereits tief gesättigten Boden und verursachten in den Ortschaften im und nahe am Erzgebirge verheerende Sturzfluten. Ihnen folgte ein sehr rasches Ansteigen der Pegelstände der Nebenflüsse der Elbe, die das Erzgebirge nach Norden entwässern. Der ohnehin relativ hohe Wasserstand vieler Flüsse nach den kräftigen Regenfällen in den Wochen zuvor spitzte die Situation zu. Viele Rückhaltebecken konnten die Wassermengen nicht mehr fassen. Das Müglitztal z.B. wurde überflutet und von den Wassermassen verwüstet. Der Wasserstand der Elbe erreichte eine seit Jahrhunderten nicht mehr da gewesene Höchstmarke. Allein in Deutschland entstand durch das August-Hochwasser ein wirtschaftlicher Schaden von 9,2 Milliarden Euro (Rudolf und Rapp, 2003).

Gemäß der weltweiten Schadensstatistik ist keine Bruchursache bei Stauanlagen häufiger als Hochwasser. Stauanlagen sind Bauwerke mit einem in der Regel großen Schadenspotential im Bruchfall. Dies gilt für Hochwasserrückhalteanlagen genauso wie für regelmäßig bewirtschaftete Speicher. Sie müssen daher hohe Sicherheitsanforderungen erfüllen. Die Gewährleistung der erforderlichen Sicherheit einer Stauanlage verlangt vorab, dass das Absperrbauwerk für alle auftretenden Last- und Betriebsfälle korrekt, also mit den angemessenen Sicherheitszuschlägen, bemessen ist. Außerdem müssen die Hochwasserentlastungen einer Stauanlage eine ausreichende Abflussleistung besitzen, sonst kann der Wasserspiegel im Falle eines extremen Hochwassers so hoch ansteigen, dass die Krone überströmt wird. Die Beherrschung der außerordentlichen Belastungen infolge extremer Naturereignisse wie Hochwasser oder Erdbeben bereitet einige Probleme, weil dazu die am Standort größtmöglichen Ereignisse berücksichtigt und abgeschätzt werden müssen. Nach der neuen Fassung der DIN-Norm 19700-11 muss die Standsicherheit von Stauanlagen so angelegt sein, dass ein 10000-jährliches Hochwasser abgehalten wird. Eine Bemessung auf ein kleineres Hochwasser ist nur bei kleinen Rückhaltebecken und auch nur dann zulässig, wenn sich im Bruchfall keine erheblich größeren Überschwemmungen ergeben und

somit keine Menschenleben gefährdet werden (Muth, 1992).

Die statistische Methode zur Berechnung des Bemessungshochwassers stützt sich auf die wahrscheinlichkeitstheoretische Analyse der in der Vergangenheit beobachteten Hochwasser, die als Zufallsereignisse betrachtet werden. Dem Hochwasserkollektiv wird eine Verteilung angepasst und extrapoliert, um die Größe und Wahrscheinlichkeit von Hochwasserpegeln und -durchflüssen in der Zukunft einzuschätzen (Hochwasserprognose). Das schweizerische Sicherheitskonzept (Biedermann, 1988) geht heute davon aus, dass das größtmögliche Hochwasser sowohl bezüglich des Spitzenabflusses als auch der zeitlichen Dauer nicht um mehr als 50% größer sei, als das 1000-jährliche Hochwasser. Damit konkretisiert sich die Vorgabe eines Bemessungshochwassers statt des 10000-jährlichen in der Kalkulation des 1000-jährlichen Hochwassers, das im Mittel nur alle 1000 Jahre übertroffen wird.

Die Landestalsperrenverwaltung des Freistaates Sachsen hat bei der Sicherheitsprüfung bestehender und der Planung neuer Anlagen eine Reihe von Vorgaben und Normen (DIN 19 700 T11, T12) zu beachten und entsprechende Berechnungen durchzuführen. Das Hochwasser vom August 2002, das in zahlreichen Einzugsgebieten des Freistaates Sachsen historische Ausmaße erreichte, veränderte in erheblichem Maße die bisherigen hochwasserstatistischen Aussagen und damit die Grundlagen für die Hochwasserschutzkonzepte in den vom Hochwasser besonders betroffenen Regionen. Die Landestalsperrenverwaltung gab daher eine umfassende Studie bei der WASY GmbH (Gesellschaft für wasserwirtschaftliche Planung und Systemforschung mbH) in Auftrag, mit dem Ziel, nach einer landesweit einheitlichen Methodik das Hochwasserereignis zu bewerten und die Hochwasserverteilungen für künftige Bemessungsaufgaben flächendeckend zu aktualisieren. Den Ergebnisbericht (Hentschel et. al., 2003) sowie die Durchflussdaten der Pegelmessstation Rehefeld1/Wilde Weißeritz hat die Landestalsperrenverwaltung für diese Arbeit freundlicherweise zur Verfügung gestellt.

7.2 Grundbegriffe aus der Hydrologie

Der auf der Erdoberfläche auftreffende Niederschlag fließt auf unterschiedliche Weisen ab. Ein Teil gelangt auf oberirdischem Wege unter dem Einfluss der Schwerkraft in die Flüsse, wobei ein Rest im Gebiet (etwa durch Mulden) zurückgehalten wird. Ein anderer Teil wird durch die Bodeninfiltration in das Grundwasser abgeleitet und erreicht das Flussbett über den Grundwasserabfluss mit einer gewissen zeitlichen Verzögerung. Daher genügt die Angabe der Niederschlagsmenge allein nicht zur Kennzeichnung der Wassermenge, die in einer Speicheranlage eintrifft. Dafür ist die Summe des Wasservolumens aller einfließenden Gewässer ausschlaggebend. Der

Volumenfluss, der pro Zeiteinheit durch einen Gewässerquerschnitt fließt, wird als Durchfluss Q bezeichnet und meist in m^3/s angegeben.

Hochwasser können unter verschiedenen Gesichtspunkten definiert werden. Allgemein versteht man darunter eine zeitlich begrenzte Anschwellung des Durchflusses über den Basisdurchfluss, die eine für jeden Durchflussquerschnitt aus der Statistik oder den örtlichen Gegebenheiten zu bestimmende Grenze überschreitet. Für einzelne Flüsse wurde das Hochwasser als Vielfaches des mittleren Durchflusses (Schwellenwert) festgelegt. Bei statistischen Untersuchungen werden meist die Jahres-Höchstdurchfluss-Werte HQ zugrunde gelegt. Von besonderem Interesse ist der Scheiteldurchfluss. Er kann direkt durch eine Durchflussmessung, indirekt über hydraulische und hydrologische Methoden oder mit Hilfe einer Wasserstands-Durchfluss-Beziehung (Durchflusskurve) ermittelt werden. Dazu ist der Hochwasserstand einer Hochwasserwelle sorgfältig zu messen. Für die Standsicherheit von Deichen und Staumauern ist die Hochwasserdauer von Bedeutung. Eine Hochwasserwelle mit hohem Scheitel und kurzer Dauer hat gewöhnlich eine kleine Durchflusssumme und kann daher durch ein Speicherbecken vollständig beherrscht werden, während dies auf ein Hochwasser mit niedrigem Scheitel, aber langer Dauer und entsprechend großer Durchflusssumme von einem Speicher gleicher Größe oft nicht zutrifft.

Der Bemessungswert ist ein Hochwasser-Scheiteldurchfluss $HQ(T)$ mit einem bestimmten Wiederkehrintervall T (Jährlichkeit). Zu seiner Ermittlung verwenden die Hydrologen Verfahren der Extremwertstatistik unter Beobachtung der Jahreshöchst-durchflüsse. Der einzuschlagende Weg zur Bestimmung des Bemessungsdurchflusses hängt insbesondere von der Reihenlänge der vorhandenen Durchflussdaten ab. Sind langjährige Durchflussbeobachtungen in der beobachteten Flussstrecke vorhanden, kann man die Extremstatistiken der gesuchten Wiederkehrintervalle zu Abschätzung der Jährlichkeiten heranziehen. Das funktioniert ohne Verteilungsannahmen, aber nur solange lediglich Aussagen bis zum n -jährlichen Ereignis der Reihe zu treffen sind. Pegelstandsmessungen an großen Flüssen liegen manchmal über mehrere Jahrhunderte vor. Die Aufzeichnung der Durchflüsse dagegen ist ein relativ junger Bereich und nicht immer gingen Pegelstandsmessung und Durchflussberechnung Hand in Hand. Um historische Hochwasser zu erschließen, müssen Hochwassermarken an alten Gebäuden, Brücken usw. aufgenommen und Chroniken ausgewertet werden. Das Flussbett hat sich aber im Laufe der Zeit durch Erosion, Bewuchs oder Gewässerausbau stark verformt. Dabei ändert sich nicht nur die Geometrie des Profils, sondern auch die Rauigkeiten und die maßgebenden Gefälle, so dass man heutige Berechnungsformeln nicht ohne weiteres auf damalige Pegelstände anwenden kann. Hydraulische Berechnungen anhand des Gewässerprofils an den Hochwassermarken sollen den Durchfluss rekonstruieren. Derartige hydraulische Berechnungen erfor-

dern viel Erfahrung. An kleineren Flüssen hat man teilweise erst vor kurzem mit Pegelstandsmessungen begonnen. Andere Flüsse haben ihren Charakter, etwa nach dem Bau wasserwirtschaftlicher Anlagen so stark verändert, dass nur eine verkürzte Datenreihe zur Auswertung herangezogen werden kann. So ist oft eine Extrapolation der Reihe unumgänglich, um Aussagen über einen Zeitraum von n Jahren hinaus zu erhalten. Dazu ist eine Spezifikation der Verteilungsfunktion nötig.

Trotz umfangreicher Bemühungen ist es bisher wegen der Komplexität der Hochwasser-Ereignisse nicht gelungen, den „wahren“ Verteilungstyp der Hochwasserscheitel aus dem physikalischen Hintergrund abzuleiten. Daher passen Hydrologen verschiedene hypothetische Verteilungsfunktionen an. Im Anschluss an eine Parameterschätzung lässt sich aus Daten und spezifizierter Verteilungsfunktion in zweierlei Hinsicht ein zeitlicher Bezug herstellen: Die Verteilungsfunktion ordnet jedem Beobachtungswert eine Unterschreitungs- bzw. $1 - F(x)$ eine Überschreitungswahrscheinlichkeit zu. Da es sich bei den Daten um jährliche Messwerte handelt, weist die Relation $1/(1 - F(x))$ jedem Hochwasser-Ereignis ein Wiederkehrintervall zu. Das Wiederkehrintervall gibt an, nach wie vielen Jahren in etwa wieder mit einem Hochwasser-Ereignis zu rechnen ist, das einen mindestens ebenso hohen Durchfluss erbringt. Umgekehrt liefern die Perzentile unter gewissen zeitlichen Vorgaben die gesuchten Bemessungsgrößen. Bei Jahresdaten geben demnach

$$F(HQ(100)) = 0.99 \quad \text{und} \quad F(HQ(1000)) = 0.999$$

das 100-jährliche bzw. das 1000-jährliche Hochwasser an.

7.3 Die Genauigkeit von Durchflussdaten

Tägliche Durchflusswerte resultieren als indirekte Beobachtungen aus Wasserstandsablesungen und Berechnungen anhand der Durchflusskurve. An ausgewählten Gewässerquerschnitten wird mit Pegeln der Wasserstand W in cm beobachtet und in hydrologisch begründeten Abständen der zugehörige Durchfluss Q gemessen. Die Beziehung $Q = f(W)$ heißt Durchflusskurve und bildet zusammen mit der Wasserstandsmessung die Grundlage jeder Durchflussbeobachtung.

Die Genauigkeit der Durchflusswerte beginnt mit der Genauigkeit der Wasserstandsbeobachtungen. Die Wasserstände werden aus den kontinuierlichen Aufzeichnungen mechanischer Schreibpegel oder über die Datenfernübertragung gewonnen, die alle 15 Minuten einen Wert liefert. Fehler bei der Messung, etwa durch Gerätedefekte, Fehllage eines Lattenpegels u.ä., sind nie ganz auszuschließen, können aber bei gravierender Abweichung im Rahmen einer intensiven Plausibilitätsprüfung entdeckt werden.

Bei extremen Hochwassern nimmt die Genauigkeit der Wasserstände gravierend ab. Eine Überschwemmung des Pegelhauses und das Anstoßen des Schwimmers können die Messtechnik außer Betrieb setzen. Geodätische Einmessungen der Geschwimmellinien oder Hochwasserlinien an Gebäuden lassen dann oftmals nur noch im Nachhinein auf den Maximalwasserstand schließen.

Weiterhin ist die Genauigkeit der Durchflussmessung zu beachten. Der hydrometrische Flügel misst den Durchfluss im Gewässerquerschnitt. Eichung des Flügels, Pulsation der Geschwindigkeit, die Geschwindigkeitsverteilung im Profil sowie die verwendete Flügelschaukel beeinflussen eine Messung. Darüber hinaus bestimmen die Anzahl der Messpunkte im Querschnitt und die Summe der reinen Messzeit in allen Messpunkten die Genauigkeit einer Durchflussmessung. Ein statistisches Modell fasst später die nach gewissenhafter Durchführung verbleibende Unsicherheit als Zufallsfehler auf.

Als Fehlerquelle ist außerdem die Genauigkeit der Durchflusskurve zu berücksichtigen. Bei der Aufstellung der Durchflusskurve müssen alle Pegelbereiche ausreichend mit Messungen belegt sein. Die Genauigkeit der Randbereiche der Durchflusskurve hat dabei fundamentalen Charakter. Laut Vorschrift müssen für jeden Bereich mindestens fünf Messungen vorliegen. Die Randbereiche der Durchflusskurve sind wegen des seltenen Eintritts extremer Pegelstände (wie etwa beim Augusthochwasser 2002) nur unzureichend oder gar nicht mit Messungen belegt. Technische Schwierigkeiten während extremer Ereignisse verhindern oft eine Messung. Extrapolationsmethoden schreiben deswegen die Durchflusskurve in den Hochwasser-Bereich fort. Stützstellen, bei denen die zu den Wasserständen gehörenden Durchflüsse aus Spendenvergleichen gewonnen werden, liefern größtenteils die Basis einer Extrapolation. Im Gegensatz zu den beschriebenen Fehlerquellen bei Wasserstands- und Durchflussmessung ist der Extrapolationsfehler systematischer Natur, denn er besitzt bei gleichen Messvorgängen immer wieder die gleiche Struktur.

Die Landestalsperrenverwaltung ist sich dieser Fehlerquellen bewusst. Nach eigenen Einschätzungen – speziell zu den Daten der Pegelmessstation Rehefeld1/Wilde Weißeritz – ist der Hochwasserdurchfluss von 65 des Jahres 2002 mit einer Schwankung von 20 bis 30% zu sehen. Schwankung im Sinne eines Konfidenzintervalls für den Zufallsfehler, die aber später einem mittels der Durchflusskurve errechneten Wert nicht anzusehen ist. Die Genauigkeit nimmt für kleinere Werte zu, aber selbst einem mittleren Hochwasser, wie dem des Jahres 1965 mit einem Durchfluss von 6.88, muss eine Schwankung von 10% zugestanden werden. Bisher hat man keine geeignete Maßnahme gefunden, um diesem Faktum Rechnung zu tragen oder seine Tragweite zu beurteilen.

7.4 Eine Auswertung der Messreihe Rehefeld1/Wilde Weißeritz

Die Auswertung soll in erster Linie die in dieser Arbeit entwickelte Methodik illustrieren. Eine vollständige und unter hydrologischen Aspekten korrekte Aufbereitung der Daten (Konsistenz- und Homogenitätsprüfung, Ausreißertests, separate Betrachtungen von Sommer- und Winterhochwasser usw.) bleibt den Hydrologen vorbehalten. Die WASY GmbH hat diese Aufgabe im Auftrag der Landestalsperrenverwaltung übernommen und im Herbst 2003 ihren Ergebnisbericht vorgelegt. Darin wurde eine extremwertstatistische Auswertung vorgenommen unter Einbeziehung verschiedener zeitlicher Extrapolationsverfahren zur Prognose der Jährlichkeiten wie des Kleeberg-Schumann-Verfahrens oder des Schweizer Sicherheitskonzeptes (Kleeberg und Schumann, 2001). Nach einer Faustregel sollte der zeitliche Extrapolationsbereich bei Nutzung einer angepassten analytischen Verteilung das Dreifache des Beobachtungszeitraumes nicht überschreiten. Selbst bei der seltenen Vorlage einer hundertjährigen Beobachtungsreihe wären damit nur Prognosen bis zum 300-jährigen Wiederkehrintervall gestattet. Nach hydrologischen Überlegungen wurden daher oben genannte alternative Näherungsverfahren vorgeschlagen, die eine erweiterte Extrapolation erlauben. In der folgenden Analyse bleibt die Frage nach der Zulässigkeit der Extrapolation außen vor. Als Vergleichswerte werden die geschätzten Jährlichkeiten ohne Korrekturmaßnahmen verwendet. Nachstehende Tabelle zeigt die erhobenen Durchflüsse zum Pegel Rehefeld1 der Wilden Weißeritz aus den Jahren 1961-2002.

1961-1970	2.43	1.97	1.97	1.79	6.88	5.12	2.52	4.3	4	3.76
1971-1980	3.07	2.12	2.12	3.9	11.8	7.92	5.54	5.77	3.26	23.6
1981-1990	10.6	4.64	15.3	4.27	3.92	5.88	10.6	5.3	4.12	4.4
1991-2000	5.62	4.99	3.85	3.06	5.14	6.51	2.66	4.59	5.96	11
2001-2002	3.05	65								

Tabelle 12: Durchflüsse am Pegel Rehefeld1/Wilde Weißeritz

Die Untersuchung soll diese „traditionelle“ Methode ohne Beachtung der Datenfehler einer Variante gegenüberstellen, die die stetige Extrapolation durch eine Intervallskalierung ersetzt. Die Skala soll sich nach den angegebenen Konfidenzintervallen für den zufälligen Messfehler richten und so die Möglichkeit schaffen, zu jeder Ablesung gleichzeitig eine Information über die Verlässlichkeit des Wertes zu übermitteln. Dazu einige Vorüberlegungen: Der Nutzen einer Messskala wird erheblich geschmälert, wenn die Ablesungen in manchen Skalenbereichen genauer vorgenommen werden können als in anderen. Idealerweise sollte sich eine Skala an der tatsächlich tech-

nisch möglichen Präzision orientieren. In dem Beispiel aus Kapitel 5.2. zur Präzision eines Messgerätes wäre eine Skala mit zwei Nachkommastellen vielleicht hilfreicher, als eine zusätzliche dritte Stelle, auf deren Anzeige aber in einer Einzelmessung wenig Verlass ist. Im aktuellen Fall wäre denkbar, die Durchflussskala von vorne herein so auszulegen, dass die Markierungen in den unsicheren Bereichen in weiteren Abständen angebracht sind. Oder anders gesagt: Man gruppiert die stärker schwankenden Bereiche zu größeren Intervallen. Ein Konfidenzintervall für den Zufallsfehler, aus Messwiederholungen errechnet, könnte dazu einen groben Anhaltspunkt bieten. Statt des „unsicheren“ Wertes wird nur die Intervallzugehörigkeit erfasst. Diese ist auf den ersten Eindruck weniger informativ, aber nicht weniger verlässlich. Die Daten können dann als Repräsentanten eines speziellen Wertebereiches interpretiert werden. Lassen sich äquidistante Wertebereiche begründen, stellen die Daten gerundete Werte dar. Unter einer bestimmten Verteilungsannahme kann man mit Hilfe des verteilungsspezifischen Schwellenwertes den Rundungsgrad beurteilen. Auf Basis des Kombinationsverfahrens aus Abschnitt 4.2 wird dann der geschätzte Rundungsgrad auf das geeignete Schätzverfahren verweisen, dem es gelingt, das vorgegebene Niveau für einen Konfidenzbereich einzuhalten.

In einer Vorstufe werden die Rehefeld-Daten zuerst in ein geeignetes Rundungsgitter eingebettet, das der unterschiedlichen Datenqualität gerecht wird. Die Ausführungen über die Ungenauigkeit der Durchflussdaten verdeutlichen die enge Verknüpfung zwischen der Höhe des Datenflusswertes selbst und der Ungenauigkeit. Demnach besteht eine annähernd exponentielle Relation derart, dass sich mit wachsender Durchflussmenge die Ungenauigkeit potenziert. Daher bietet sich die Einordnung der Werte in ein exponentielles Rundungsgitter geradezu an. Nun kann man die Gedanken zu gerundeten transformierbaren Daten aus Kapitel 6 aufgreifen. War in der Ursprungsskala das Gitter exponentiell angelegt, resultiert aus einer logarithmischen Transformation ein äquidistantes Rundungsgitter. So lässt sich die immanente Struktur der Daten nutzen und die Rehefeld-Reihe auf eine Datenreihe mit gleichmäßiger Rundungsbreite zurückführen. Die entwickelten Methoden des Kapitels 5 können die transformierte Reihe dann geeignet bearbeiten.

Die im vorigen Abschnitt genannten Schwankungsbreiten lassen sich in adäquater Weise in Tabelle 13 gezeigter Skala abbilden.

x	$\exp(x)$	x	$\exp(x)$	x	$\exp(x)$
0.25	1.284	1.75	5.755	3.25	25.790
0.5	1.649	2	7.389	3.5	33.115
0.75	2.117	2.25	9.488	3.75	42.521
1	2.718	2.5	12.182	4	54.598
1.25	3.490	2.75	15.643	4.25	70.105
1.5	4.482	3	20.089		

Tabelle 13: Exponentielle Skala der Durchflüsse

Dieses Gitter ist so ausgerichtet, dass die Breite des letzten Intervalles eine Schwankungsbreite von 20 bis 30% des Maximalwertes von 65 widerspiegelt. Die Kontrolle des unteren Wertebereiches zeigt zu große Intervallbreiten an. Der Fehler des Wertes 6.88, eigentlich mit 10% ausgewiesen, würde stark übertrieben. Ein doppelt exponentielles Gitter (Tabelle 14) stellt die Schwankungsangaben in fast idealer Weise nach.

x	$\exp(e^x)$	x	$\exp(e^x)$	x	$\exp(e^x)$	x	$\exp(e^x)$
-0.7	1.64	-0.1	2.47	0.5	5.20	1.1	20.17
-0.6	1.73	0	2.72	0.6	6.18	1.2	27.66
-0.5	1.83	0.1	3.02	0.7	7.49	1.3	39.22
-0.4	1.95	0.2	3.39	0.8	9.26	1.4	57.67
-0.3	2.10	0.3	3.86	0.9	11.7	1.5	88.38
-0.2	2.27	0.4	4.45	1	15.15		

Tabelle 14: Doppelt exponentielle Skala der Durchflüsse

Welchen stellvertretenden Wert des Intervalls man den transformierten Durchflüssen zuweist, ist für die Likelihood der gerundeten Daten irrelevant, denn sie kommt allein mit dem Wissen um die Frequentierung des Intervalls aus. Anders die Likelihood der exakten Daten. Zu ihrer Berechnung stützte sich das Kombinationsverfahren bisher auf die Eingabe der Intervallmittelpunkte als einzig verfügbare Werte. In diesem konkreten Fall gibt es zwei Möglichkeiten: Entweder ersetzt die Mitte des Intervalls (IM) nach der Transformation in das äquidistante Gitter jeden Datenwert, oder die transformierten Daten gehen in die Berechnung ein. Beide Möglichkeiten werden im Folgenden in Betracht gezogen. Tatsächlich führen beide Varianten auf nahezu identische Ergebnisse.

Nach Festlegung der Intervallabschnitte für die Extrapolation durch ein exponentielles Gitter ist nun ein passender Verteilungstyp gesucht. Nach allen bisherigen Erfahrungen gibt es keinen universellen Verteilungstyp, der an alle Beobachtungsreihen von Hochwasserdurchflüssen gleichermaßen gut angepasst ist. Ein Verteilungstyp, der mit einem gewählten Parameterschätzverfahren bei einer gegebenen Stichprobe eine sehr gute Anpassung liefert, kann bei einer Stichprobe aus einem anderen, ohne tiefer gehende Untersuchungen hydrologisch durchaus vergleichbaren Gebiet bei der Anpassung versagen. Die WASY GmbH hat sich für eine Anpassung der so genannten „allgemeinen“ Extremwertverteilung (AE) entschieden. Dabei handelt es sich um eine Parametrisierungsform, die simultan alle drei möglichen Grenzverteilungstypen des Maximums erfasst:

$$F(x) = \exp \left[- \left(1 - a \frac{x - c}{d} \right)^{\frac{1}{a}} \right] \quad (7.1)$$

Die drei Verteilungstypen stehen zu der allgemeinen Form in folgender Beziehung:

$$F(x) = \begin{cases} \exp \left\{ - \left[\frac{(c + \frac{d}{a}) - x}{(c + \frac{d}{a}) - c} \right]^{\frac{1}{a}} \right\} & \text{für } a > 0 \quad (\text{Typ III}), \quad -\infty \leq x \leq c + \frac{d}{a} \\ \exp \left\{ - \left[\frac{c - (c + \frac{d}{a})}{x - (c + \frac{d}{a})} \right]^{-\frac{1}{a}} \right\} & \text{für } a < 0 \quad (\text{Typ II}), \quad c + \frac{d}{a} \leq x \leq \infty \\ \exp \left\{ - \exp \left[\frac{x - c}{d} \right] \right\} & \text{für } a < 0 \quad (\text{Typ I}), \quad c + \frac{d}{a} \leq x \leq \infty \end{cases}$$

Die Parameter der Verteilung werden nach der Methode der wahrscheinlichkeitsgewichteten Momente geschätzt (Kluge, 1996), sind aber im Einzelnen nicht im Ergebnisbericht überliefert.

Gängige Anpassungstests (Chi-Quadrat-Test, Kolmogorow-Smirnow-, Cramer-von-Mises-, Watson-, Anderson-Darling-Test) lassen unter verschiedenen Verteilungsmodellen für die Messreihe Rehefeld in doppelt logarithmierter Skala mehrere Typen offen, die zum 5%-Niveau nicht zur Ablehnung der Verteilungshypothese führen. Die besten Anpassungswerte aber liefert die Logistische Verteilung, dicht gefolgt von der Normalverteilung. Besonders verblüfft, wie sich die Anpassungsgüte durch das zweifache Logarithmieren der Daten gegenüber den Originalwerten verbessert. Das können Quantil-Quantil-Grafiken (Abbildungen 41, 42) bildlich demonstrieren. Die Transformation, die sonst zum Zweck einer besseren Anpassungsgüte mehr oder weniger auf experimenteller Grundlage beruht, ist in diesem Fall durch eine intensive vorausgegangene Fehleranalyse motiviert. Die Fehleranalyse weist den Weg zur besten Transformation.

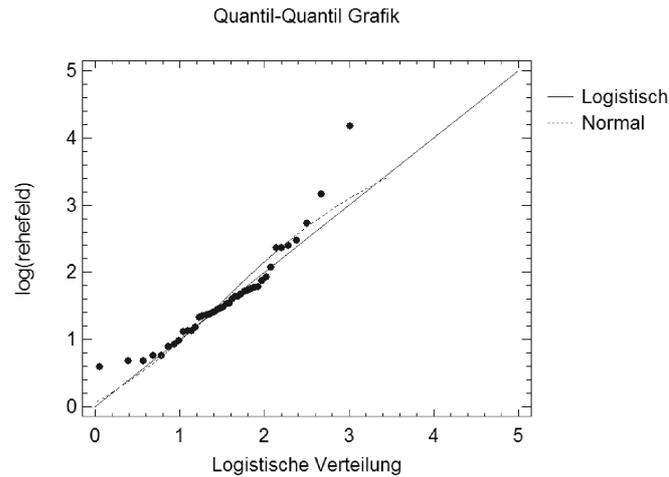


Abbildung 41: QQ-Grafik der einfach logarithmierten Jahres-Höchstwerte

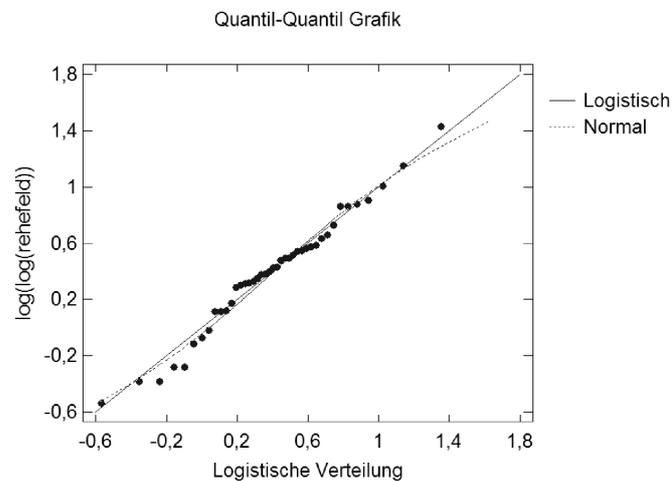


Abbildung 42: QQ-Grafik der doppelt logarithmierten Jahres-Höchstwerte

Nach dem Ablaufprinzip des Kombinationsverfahrens werden die Daten zunächst wie exakte behandelt und der Skalenschätzer zur Abschätzung des Rundungsgrades berechnet. Das ergibt

$$\hat{b}\{\ln(\ln(x))\} = 0.224 \quad \text{und} \quad \hat{b}\{IM\} = 0.222. \quad (7.2)$$

Das Kürzel IM signalisiert die Verwendung der Intervallmitten. Die geschätzten Rundungsgrade sind

$$\hat{r}\{\ln(\ln(x))\} = 0.77 \quad \text{und} \quad \hat{r}\{IM\} = 0.78. \quad (7.3)$$

Sie liegen weit unter dem Schwellenwert von 2.21 und stufen das Ignorieren des Rundens als unbedenklich ein. Demnach wäre die Verwendung des Pivot-Verfahrens gestattet, allerdings ist der Stichprobenumfang mit 42 für eine gute χ^2 -Approximation des Likelihood-Quotienten ausreichend. Wegen der ohnehin nötigen iterativen Verfahren erfolgt daher die Konfidenzintervallberechnung mit einer χ^2 -Approximation

der Likelihood für exakte Daten.

Der Lageparameter wird auf

$$\begin{aligned}\hat{a} \{\ln(\ln(x))\} &= 0.394 \quad \text{und} \\ \hat{a} \{IM\} &= 0.398\end{aligned}\tag{7.4}$$

geschätzt mit den realisierten 95%-Konfidenzintervallen

$$\begin{aligned}KI \{\hat{a}; \ln(\ln(x))\} &= [0.264; 0.524] \quad \text{und} \\ KI \{\hat{a}; IM\} &= [0.268; 0.518].\end{aligned}\tag{7.5}$$

Dies entspricht in der Originalskala einem Konfidenzintervall von

$$\begin{aligned}KI \{\hat{a}; x\} &= [3.676; 5.411] \quad \text{und} \\ KI \{\hat{a}; \exp(\exp(IM))\} &= [3.697; 5.359].\end{aligned}\tag{7.6}$$

Schließlich ergeben sich mit Hilfe der Parameterschätzer die Jährlichkeiten und das zu erwartende Wiederkehrintervall mit einem Durchfluss des Augusthochwassers. Der errechneten Werte, zum einen auf Basis des Schätzers, zum anderen auf Basis der Parameterunter- sowie der -obergrenze, sind der folgenden Tabelle zu entnehmen. Zum Vergleich sind in der Spalte „AE“ (Allgemeine Extremwertverteilung) die geschätzten Jährlichkeiten des WASY-Berichtes aufgenommen.

Erwartete Durchflüsse auf Basis von							
Jahre	$a_u(\ln \ln(X))$	$\hat{a}(\ln \ln(X))$	$a_o(\ln \ln(X))$	$a_u(IM)$	$\hat{a}(IM)$	$a_o(IM)$	AE
2	3.68	4.41	5.41	3.70	4.43	5.36	4.25
10	8.42	11.31	15.84	8.41	11.30	15.40	11.69
50	22.50	34.67	56.72	22.24	34.22	53.68	30.24
100	38.28	63.49	112.98	37.57	62.16	105.24	45.56
200	70.95	128.19	251.51	69.01	124.23	229.74	68.72
m^3/s	Wiederkehrintervall des Durchflusses vom August 2002 in Jahren						
65	57.83	102.54	182.42	61.53	104.93	187.65	182

Tabelle 15: Erwartete Durchflussmengen und Wiederkehrintervalle

Streng genommen handelt es sich bei den oberen und unteren Durchflussangaben nicht um ein Konfidenzintervall für die Jährlichkeiten. Dazu wäre ein Konfidenzintervall für Quantile zu erstellen. Sie vermitteln aber einen ersten Eindruck des Spielraumes, in dem sich die erwarteten Durchflüsse anhand der Parametervorgaben bewegen.

Der nun folgende Abschnitt stellt eine zweite Variante zur Schätzung der Jährlichkeiten vor: Genau betrachtet unterliegt die Definition eines Hochwassers über Jahreswerte einer gewissen Willkür. Ob die Jahreshöchstwerte das Regime solcher Extrema, wie sie einmal im Jahrtausend eine Region heimsuchen, adäquat beschreiben, ist nicht sicher. Die Extrema der Jahrzehnte könnten ebenso gut solche außergewöhnlichen Vorkommnisse charakterisieren. Diese Auffassung ist in Hydrologenkreisen unüblich, weil sie mit einer immensen Datenreduktion verbunden ist. Diese Datenreduktion muss aber keinen Informationsverlust bedeuten, denn die restlichen Datenwerte fließen über einen Prozess der Informationsverarbeitung als bewusste Auswahl des Maximums eines Zehn-Jahres-Zeitraumes in die neue Reihe ein. Diese Vorgehensweise hätte den entscheidenden Vorteil, dass nicht alljährlich sämtliche Kalkulationen zu wiederholen wären, sondern eine 10-jährliche Überarbeitung genügte. Zum Vergleich sollen die Berechnungen auch für die derart reduzierte Datenmenge ausgeführt werden.

Bei Start der Zehn-Jahres-Aufteilung mit Beginn der Pegelaufzeichnung im Jahr 1961, bleiben als 10-Jahres-Maxima:

6.88 23.6 15 11 65.

Diesmal erweist sich die Anpassung einer Normalverteilung an die zweifach logarithmierten Maxima gegenüber der Logistischen Verteilung leicht überlegen (Abbildung 43). Mit Verwendung der Intervallmittelpunkte wird das Mittel auf 1.03 und die

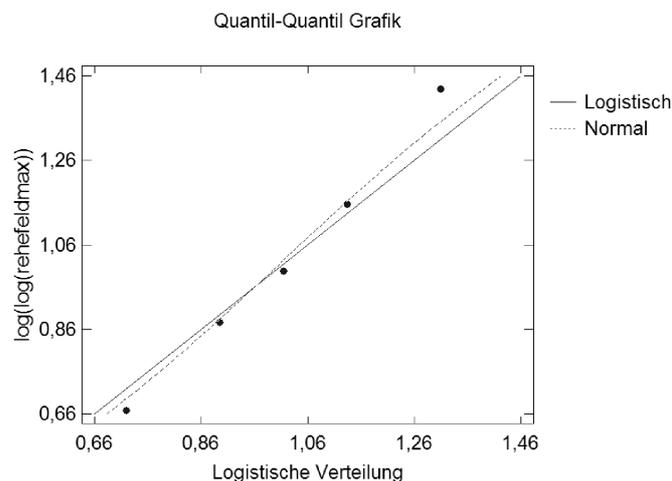


Abbildung 43: QQ-Grafik der doppelt logarithmierten Zehn-Jahres-Höchstwerte

Standardabweichung auf 0.303 geschätzt. Auch hier liegt der geschätzte Rundungsgrad mit 0.33 weit unter dem Schwellenwert. Das Pivot-Konfidenzintervall grenzt den Lageparameterbereich auf $[0.756; 1.304]$ ein, oder in der Originalskala $[8.42; 39.8]$. Verständlicherweise hat sich die Ungenauigkeit verstärkt, interessant aber ist,

dass man mit Hilfe der Parameterschätzer \hat{a} und \hat{b} zu nahezu identischen Schlüssen bezüglich der Jährlichkeiten gelangt wie unter Benutzung aller 42 Daten, wie die folgende Tabelle zeigt.

Jahre	20	100	200	WI(65)
m^3/s	16.46	62.17	289.53	106.44 Jahre

Tabelle 16: Geschätzte Jährlichkeiten auf Basis der 10-Jahres-Maxima

Insgesamt lässt sich folgendes Fazit ziehen: Prognosen über einen Zeitraum von 100 Jahren hinaus muss man auf Grundlage dieser Daten als äußerst spekulativ bezeichnen. Sie sollten nur mit hydrologisch fundierter Zusatzinformation und neu angepasster Modellierung abgegeben werden. Innerhalb des zulässigen Extrapolationszeitraumes von 120 Jahren kommen beide Methoden (bei Verwendung der doppelt logarithmierten Daten oder Intervallmitten) zu vergleichbaren Durchflussschätzungen. Die Schätzwerte durch die allgemeine Extremwertverteilung fallen grundsätzlich kleiner aus. Hohe Durchflussmengen kehren demnach erst nach längeren Zeitabständen wieder. Verglichen mit den Prognosen durch die Logistische Verteilung wirft man damit einen sehr optimistischen Blick in die Zukunft. Während dort ein dem August 2002 vergleichbares Hochwasser erst in 186 Jahren zu erwarten ist, muss der geschätzten Logistischen Verteilung zufolge bereits nach 103 Jahren wieder mit einem solchen Extremereignis gerechnet werden.

8 Zusammenfassung und Ausblick

Gerundete Daten gehören zum Alltag der statistischen Datenanalyse. Allzu oft wird ihre diskrete Natur zugunsten einer bequemerer Handhabung ignoriert und mögliche Folgen werden verkannt. Bereits vor mehr als 100 Jahren hat Sheppard auf ein Ansteigen der Varianz als eine dieser Auswirkungen hingewiesen. Rundungsvorgänge können auch die Qualität der Parameterschätzung mindern. Wie gut sich die Verteilungsparameter approximieren lassen, hängt ganz entscheidend vom Stichprobenumfang und vom Rundungsgrad der Daten ab. Je kleiner der Stichprobenumfang und je größer der Rundungsgrad, desto stärker die Verzerrung und desto schwieriger eine Punktschätzung. Normalerweise sollte das eine parallele Intervallschätzung zum Ausdruck bringen. Jedoch können dies herkömmliche Konfidenzintervalle, die meist auf dem Pivot-Verfahren basieren, nicht leisten. Bei hohen Rundungsgraden und kleinen Stichprobenumfängen häufen sich Stichproben mit gleichen Werten, der ausgewiesene Bereich schrumpft auf eine Singularität und das Niveau wird in nicht mehr akzeptablen Maßen unterschritten. Andererseits zeigen für gerundete Daten konzipierte Verfahren die typischen Schwächen der meisten approximativen Verfahren. Erst mit größeren Stichprobenumfängen ab $n = 15$ ist eine Niveaueinhaltung annähernd gesichert. Dieser Sachverhalt ist bei allen untersuchten Verteilungen zu beobachten und das Ausmaß für Normal-, Maximum-, Exponential- und Logistische Verteilung in der vorliegenden Arbeit dokumentiert.

Das neu entwickelte Kombinationsverfahren koppelt Pivot- und LQ-Verfahren. Mit Hilfe einer Korrekturangabe für gleiche Stichprobenwerte kann das Kombinationsverfahren über eine geeignete Methodenwahl die Mängel jedes einzelnen Verfahrens beheben. Die Entscheidung für das eine oder andere Verfahren fällt ein Schwellenwert für den Rundungsgrad. Ein solcher Schwellenwert wird für jede Verteilung individuell bestimmt, behält aber dann über alle untersuchten Stichprobenumfänge seine Gültigkeit. Liegt der geschätzte Rundungsgrad oberhalb des Schwellenwertes, sollte unbedingt das LQ-Verfahren für gerundete Daten verwendet werden. Unterhalb ist die Benutzung des Pivot-Verfahrens unbedenklich, ja sogar unbedingt anzuraten für $n \leq 15$. Richtet man sich nach diesem Entscheidungskriterium, bleibt eine Niveaueinhaltung gewährleistet. Das Kombinationsverfahren bleibt anders als andere vorgeschlagene Korrekturverfahren auch dann einsatzfähig, wenn die Parameter der Verteilung nicht in geschlossener Form darstellbar sind.

Die Schwellenwertbedingung beantwortet die bis dato ungeklärte Frage, wann der Rundungsgrad so schwerwiegend ist, dass von einer Benutzung der Standardverfahren Abstand genommen werden sollte. Die Schwellenwerte von 2.5 (Normalverteilung), 2.21 (Logistische Verteilung), 2.6 (Maximumverteilung) sind für die Lage-

parameter in vergleichbarer und unerwartet hoher Größenordnung angesiedelt. Das Pivot-Verfahren erweist sich geradezu erstaunlich robust.

In den geschätzten Rundungsgrad fließen die üblicherweise bekannte Rundungsgitterbreite und der geschätzte Skalenparameter ein. Das Kombinationsverfahren bedient sich der dargebotenen Struktur der Daten in Form der Streuung, um retrospektiv eine systematische Verfahrenswahl auszuführen.

Ein Beispiel aus der Praxis zeigt, wie eine angemessene Skaleneinteilung die Präzision der Messtechnik verarbeiten kann. Die Gruppierung in Intervalle nach unterschiedlicher Präzisionsstruktur der Daten ähnelt den bekannten Rundungsmechanismen. Eine geeignete Einteilung erlaubt eine Angabe der Rundungsgitterbreite und eine Intervallschätzung der Parameter mit dem entwickelten Kombinationsverfahren. Die Einrichtung eines exponentiellen Rundungsgitters in Verbindung mit einer anschließenden logarithmischen Datentransformation kann die Ungenauigkeit der Durchflussreihe Rehefeld geeignet modellieren. Eine gute Verteilungsanpassung im hundertjährigen Zeitraum resultiert aus dieser Maßnahme. Die entwickelte Methodik führt schließlich zu verlässlichen Intervallschätzungen für die Verteilungsparameter.

Das Praxisbeispiel vermittelt nebenbei einen Eindruck von den konkreten Herausforderungen an die Statistik und eröffnet damit Perspektiven für zukünftige Forschungsaktivitäten. Konfidenzintervalle für den Lageparameter lassen die Unsicherheit, die mit einer weit vorausgreifenden Extrapolation einhergeht, erahnen. Noch wichtiger wäre ein passendes Konfidenzintervall-Verfahren für Quantile. Der Lageparameter entspricht bei der Normal- und Logistischen Verteilung dem 50%-Quantil und bei der standardisierten Gumbel-Verteilung in etwa dem 58%-Quantil und repräsentiert damit nur eines unter vielen gefragten Quantilen. Grundsätzlich sollte man die Extrapolation über eine Verteilungsanpassung und die Ermittlung ihrer Quantile überdenken. Die Existenz einer sachlogischen oberen Begrenzung der Durchflussmenge wird im zulässigen Wertebereich einer Verteilung nicht berücksichtigt. Die moderne Zeitreihenanalyse modelliert saisonale Schwankungen und Trends und stellt auf diese Zeitreihenmodelle abgestimmte Verfahren bereit, mit denen sich die Prognosen in einer realistischen Größenordnung bewegen. Ein auf gerundete Daten zugeschnittenes Schätzverfahren bei Zeitreihenmodellierung wäre daher ebenfalls von Interesse.

Ein weiteres wichtiges Anwendungsgebiet ist die Qualitätskontrolle in der pharmazeutischen Industrie. Hier gibt es typischerweise Bedarf an Konfidenzintervallverfahren für die Varianz. Die Qualitätssicherung eines Medikamentes erfordert eine geringe Varianz der Wirkstoffkonzentration sowohl innerhalb einer Charge wie auch der Chargen untereinander. Die Varianz innerhalb einer Charge kann meist durch eine hohe Stückzahlanalyse ausreichend gut belegt werden. Viel schwieriger ist eine verlässliche Angabe der Streuung unter den Chargen, denn das Unternehmen kann

aus Kostengründen nicht beliebig viele Chargen zur Untersuchung freigeben. Werden die wenigen Messungen zusätzlich stark gerundet, ist eine Konfidenzangabe überaus heikel. Daher sollte man trotz der angesprochenen Schwierigkeiten die Entwicklung zuverlässiger Intervallschätzungen vorantreiben.

Abbildungsverzeichnis

1	Verteilungsfunktionen: $N(4.25; 0.25)$ und gerundete Daten	9
2	Verteilungsfunktionen: $N(4; 0.25)$ und gerundete Daten	9
3	$-2 \ln R$ bei Normalverteilung für die Stichprobe $(0,0)$	46
4	Algorithmus zur Berechnung des LQ-Konfidenzintervalles für die Lage	51
5	Berechnung der oberen Grenze	53
6	Profil-Likelihood der Maximumverteilung für die Stichprobe $(0,0)$. .	56
7	ÜW Maximumverteilung, $n=2$, Verlauf über b	61
8	ÜW Maximumverteilung, $n=5$, Verlauf über b	62
9	ÜW Maximumverteilung, $n=10$, Verlauf über b	63
10	ÜW Maximumverteilung, $n=15$, Verlauf über b	64
11	ÜW Maximumverteilung, $n=2$, Verlauf über u	65
12	ÜW Maximumverteilung, $n=3$, Verlauf über u	66
13	ÜW Maximumverteilung, $n=4$, Verlauf über u	67
14	ÜW Maximumverteilung, $n=5$, Verlauf über u	68
15	ÜW Maximumverteilung, $n=10$, Verlauf über u	69
16	ÜW Maximumverteilung, $n=15$, Verlauf über u	70
17	Profil-Likelihood der Normalverteilung für die Stichprobe $(0,0)$. . .	73
18	ÜW Normalverteilung, $n=2$, Verlauf über σ	77
19	ÜW Normalverteilung, $n=5$, Verlauf über σ	78
20	ÜW Normalverteilung, $n=10$, Verlauf über σ	79
21	ÜW Normalverteilung, $n=15$, Verlauf über σ	80
22	ÜW Normalverteilung, $n=2$, Verlauf über μ	81
23	ÜW Normalverteilung, $n=3$, Verlauf über μ	82
24	ÜW Normalverteilung, $n=4$, Verlauf über μ	83
25	ÜW Normalverteilung, $n=5$, Verlauf über μ	84
26	ÜW Normalverteilung, $n=10$, Verlauf über μ	85
27	ÜW Normalverteilung, $n=15$, Verlauf über μ	86

28	Profil-Likelihood der Logistischen Verteilung für die Stichprobe (0,0)	89
29	ÜW Logistische Verteilung, $n=2$, Verlauf über b	93
30	ÜW Logistische Verteilung, $n=5$, Verlauf über b	94
31	ÜW Logistische Verteilung, $n=10$, Verlauf über b	95
32	ÜW Logistische Verteilung, $n=15$, Verlauf über b	96
33	ÜW Logistische Verteilung, $n=2$, Verlauf über a	97
34	ÜW Logistische Verteilung, $n=3$, Verlauf über a	98
35	ÜW Logistische Verteilung, $n=4$, Verlauf über a	99
36	ÜW Logistische Verteilung, $n=5$, Verlauf über a	100
37	ÜW Logistische Verteilung, $n=10$, Verlauf über a	101
38	ÜW Logistische Verteilung, $n=15$, Verlauf über a	102
39	Likelihood der Exponentialverteilung für die Stichprobe (0,0)	107
40	Überdeckungswahrscheinlichkeiten Exponentialverteilung	109
41	QQ-Grafik der einfach logarithmierten Jahres-Höchstwerte	123
42	QQ-Grafik der doppelt logarithmierten Jahres-Höchstwerte	123
43	QQ-Grafik der doppelt logarithmierten Zehn-Jahres-Höchstwerte	125

Tabellenverzeichnis

1	Korrekturintervalle für die Maximumverteilung	57
2	Quantile der Maximumverteilung für $(\hat{u} - u)/\hat{b}$	58
3	Durchschnittliche Intervalllängen; Maximumverteilung	71
4	Korrekturen für die Normalverteilung	74
5	Durchschnittliche Intervalllängen, Normalverteilung	87
6	Korrekturen für die Logistische Verteilung	90
7	Quantile der Logistischen Verteilung für $(\hat{a} - a)/\hat{b}$	91
8	Durchschnittliche Intervalllängen; Logistische Verteilung	103
9	Korrekturen für die Exponentialverteilung	107
10	Durchschnittliche Intervalllängen, Exponentialverteilung	110
11	Einteilung in Einkommensklassen	112
12	Durchflüsse am Pegel Rehefeld1/Wilde Weißeritz	119
13	Exponentielle Skala der Durchflüsse	121
14	Doppelt exponentielle Skala der Durchflüsse	121
15	Erwartete Durchflussmengen und Wiederkehrintervalle	124
16	Geschätzte Jährlichkeiten auf Basis der 10-Jahres-Maxima	126

Literatur

- Abernethy, J. R. (1933). On the elimination of systematic error due to grouping. *Annals of Mathematical Statistics*, 4, 263–277.
- Antle, C. E., Klimko, L., & Harkness, W. (1970). Confidence intervals for the parameters of the logistic distribution. *Biometrika*, 57, 397–402.
- Bain, L. J. & Engelhardt, M. (1991). *Statistical analysis of reliability and life testing models, Theory and Methods*. New York: Marcel Dekker, 2nd edition.
- Barndorff-Nielsen, O. E. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika*, 75, 374–378.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. R. Soc. A*, 160, 268–282.
- Biedermann, R. (1988). Sicherheit von Hochwasserrückhaltebecken – schweizerisches Sicherheitskonzept. *Wasserwirtschaft*, 78, 32–35.
- Billman, B., Antle, C., & Bain, L. J. (1972). Statistical inferences from censored Weibull samples. *Technometrics*, 14, 831–840.
- Büning, H. & Trenkler, G. (1994). *Nichtparametrische statistische Methoden*. Berlin: De Gruyter, 2nd edition.
- Burridge, J. (1989). Comment to D. F. Heitjan: Inference from grouped continuous data: A review. *Statistical Science*, 4, 179–181.
- Cayley, A. (1874). Letter to J. W. L. Glaisher. reproduced (in part) in: *Mathematical Tables and other Aids to Computation*, 1, 98.
- DiCiccio, T. J. (1987). Approximate inference for the generalized gamma distribution. *Technometrics*, 29, 33–40.
- DiCiccio, T. J. (1988). Likelihood inference for linear regression models. *Biometrika*, 75, 29–34.
- Doganaksoy, N. (1995). Likelihood ratio confidence intervals in life-data analysis. In N. Balakrishnan (Ed.), *Recent Advances in Life Testing and Reliability* (pp. 359–376). Boca Raton, Florida: CRC Press.
- Doganaksoy, N. & Schmee, J. (1993). Comparison of approximate confidence intervals for distributions used in life-data analysis. *Technometrics*, 35, 175–184.
- Durrans, R. (2001). Maximum likelihood estimators for quantized data. www.societyofwomenengineers.org/docs/sample.paper.pdf.

- Dyck, S. (1980). *Angewandte Hydrologie, Teil 1: Berechnung und Regelung des Durchflusses der Flüsse*. Berlin: VEB Verlag für Bauwesen.
- Eisenhardt, C. (1947). Effects of rounding or grouping data. In C. Eisenhardt, M. W. Hastay, & W. A. Wallis (Eds.), *Selected Techniques of Statistical Analysis* (pp. 185–223). New York: Mc Graw Hill.
- Elderton, W. P. (1933). Adjustment for the moments of j-shaped curves. *Biometrika*, 25, 179.
- Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Statistical Society London A*, 222, 309–368.
- Geary, R. C. (1936). The distribution of student's ratio for non-normal samples. *Supplement to the Journal of the Royal Statistical Society*, 3, 178–184.
- Gieddebaeck, N. F. (1957). Contributions to the study of grouped observations. III The distribution of the mean. *Skandinavisk Aktuarietidskrift*, 40, 20–25.
- Giesbrecht, F. & Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter lognormal distribution. *Journal of the Royal Statistical Society*, 29, 41–49.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44, 423–453. Translated and reprinted in: *Breakthroughs in Statistics I, 1992*, eds. S. Kotz & N. L. Johnson, Berlin: Springer, 195–225.
- Gourieroux, C. & Monfort, A. (1995). *Statistics and Econometric Models*. Cambridge: Cambridge University Press.
- Griffith, D. A. (1980). Interval estimation for the three-parameter lognormal distribution via the likelihood function. *Applied Statistics*, 29, 58–68.
- Haitovsky, Y. (1983). Grouped data. In N. L. Johnson & S. Kotz (Eds.), *Encyclopedia of Statistical Sciences* (pp. 527–536). New York: John Wiley.
- Hand, D. J., Daly, F., Lunn, A. D., McConway, K. J., & Ostrowski, E. (1994). *Handbook of Small Data Sets*. London: Chapman & Hall.
- Heitjan, D. F. (1989). Inference from grouped continuous data: A review. *Statistical Science*, 4, 164–183.
- Hentschel, M., Pruggmayer, H., Schramm, M., & Walther, J. (2003). *Hochwasserstatistische Einordnung des Hochwasserereignisses vom August 2002 and Ermittlung*

- von $HQ(T)$ für die Pegel der oberen Elbe, der Schwarzen Elster, der Spree and der Lausitzer Neiße als Grundlage für die Planung von Maßnahmen des Präventiven Hochwasserschutzes. Technical report, WASY Gesellschaft für wasserwirtschaftliche Planung and Systemforschung mbH, Dresden.
- Holland, B. S. (1975). Some results on the discretization of continuous probability distributions. *Technometrics*, 17, 333–339.
- Ismaïl, S. (2004). A simple estimator for the shape parameter of the pareto distribution with economics and medical applications. *Applied Statistics*, 31, 3–31.
- Jeng, S. L. & Meeker, W. Q. (1999). *Comparisons of Weibull distribution confidence intervals for Type I censored data*. Technical report, Department of Statistics, Iowa State University, Ames, IA.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions*. New York: John Wiley.
- Kleeberg, H.-B. & Schumann, A. H. (2001). Zur Ableitung von Hochwasserabflüssen geringer Überschreitungswahrscheinlichkeiten. *Wasserwirtschaft*, 91, 608.
- Kluge, C. (1996). *Wahrscheinlichkeitsanalyse von Hochwasserdurchflüssen*. TU Dresden: Bericht 7 des Instituts für Siedlungs- und Industrierwasserwirtschaft.
- Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions, Theory and Applications*. London: Imperial College Press.
- Krämer, W. & Runde, R. (1996). Chaos and the compass rose. *Economics Letters*, 54, 113–118.
- Kulldorff, G. (1961). *Contributions to the Theory of Estimation from Grouped and partially Grouped Samples*. Stockholm: Almqvist & Wiksell.
- Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: John Wiley.
- Lawley, D. N. (1956). A general method for approximating the distribution of the likelihood ratio criteria. *Biometrika*, 43, 295–303.
- Lee, C.-S. & Vardeman, S. B. (2001). Interval estimation of a normal process mean from rounded data. *Journal of Quality Technology*, 33, 335–348.
- Lee, C.-S. & Vardeman, S. B. (2002). Interval estimation of a normal process standard deviation from rounded data. *Communications in Statistics - Simulation and Computation*, 31, 13–34.

- Lee, C.-S. & Vardeman, S. B. (2003). Confidence intervals based on rounded data from the balanced one-way normal random effects model. *Communications in Statistics - Simulation and Computation*, 32, 835–856.
- Lehmann, E. L. & Romano, J. P. (2005). *Testing Statistical Hypothesis*. New York: Springer, 3rd edition.
- Meeker, W. (1986). Planning life tests in which units are periodically inspected for failure. *IEEE Transactions on Reliability R*, 35, 571–578.
- Meeker, W. & Escobar, L. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician*, 9, 48–53.
- Meeker, W. Q. & Escobar, L. A. (1994). Maximum likelihood methods for fitting parametric statistical models. In J. L. Stanford & S. B. Vardeman (Eds.), *Statistical methods for physical science* (pp. 211–244). San Diego: Academic Press.
- Meeker, W. Q. & Escobar, L. A. (1998). *Statistical methods for reliability data*. New York: John Wiley.
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. Singapore: Mc Graw Hill, 3rd edition.
- Muth, W. (1992). *Hochwasserrückhaltebecken: Planung, Bau and Betrieb*. Rellingen: Expert Verlag.
- Nelson, W. (1977). Optimum demonstration tests with grouped inspection data from an exponential distribution. *IEEE Transactions on Reliability*, 26, 29–34.
- O'Reilly, F. J., Rueda, R., & Garza-Jinich, M. (2003). How important is the effect of rounding in goodness-of-fit. *Communications in statistics - Simulation and Computation*, 32, 953–976.
- Ostrouchov, G. & Meeker, W. Q. (1988). Accuracy of approximate confidence bounds computed from interval censored weibull and lognormal data. *Journal of Statistical Computation and Simulation*, 29, 43–76.
- Pairman, E. & Pearson, K. (1919). On correction for the moment-coefficients of limited range frequency-distributions when there are finite or infinite ordinates and any slopes at the terminals of the range. *Biometrika*, 12, 231–258.
- Pearse, G. E. (1928). On correction for moment coefficients of frequency distributions when there are infinite ordinates at one or both terminals of the range. *Biometrika*, 20, 314.

- Pearson, K. (1933). Note on Mr. P. Elderton's corrections to moments of j-curves. *Biometrika*, 25, 180.
- Piegorsch, W. W. (1987). Performance of likelihood-based interval estimates for two-parameter exponential samples subject to type I censoring. *Technometrics*, 29, 41–49.
- Preece, D. A. (1981). Distribution of final digits in data. *The Statistician*, 30, 31–60.
- Rudolf, B. & Rapp, J. (2003). *DWD Klimastatusbericht 2002: Das Jahrhunderthochwasser der Elbe: Synoptische Wetterentwicklung und klimatologische Aspekte*. Selbstverlag des Deutschen Wetterdienstes: Offenbach am Main.
- Sandon, F. (1924). Note on the simplification of the calculation of abruptness coefficients to correct crude moments. *Biometrika*, 16, 193–195.
- Schader, M. & Schmid, F. (1984). Computation of maximum likelihood estimates for μ and σ from a grouped sample of a normal population. A comparison of algorithms. *Statistische Hefte*, 25, 245–258.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: John Wiley.
- Sheppard, W. F. (1898). On the calculation of the most probable values of frequency constants, for data arranged according to equidistant divisions of a scale. *Proceedings of the London Mathematical Society*, 29, 353–380.
- Silvapulle, M. J. & Burridge, J. (1986). Existence of maximum likelihood estimates in regression models for grouped and ungrouped data. *Journal of the Royal Statistical Society B*, 48, 100–106.
- Smith, R. M. (1977). Some results on interval estimation for the two-parameter weibull or extreme value distribution. *Communications in Statistics*, A6, 1311–1322.
- Spanos, A. (1999). *Probability Theory and Statistical Inference*. Cambridge: Cambridge University Press.
- Sprott, D. (1973). Normal likelihoods and their relation to large sample theory of estimation. *Biometrika*, 60, 457–465.
- Sprott, D. (1975). Application of maximum likelihood methods to finite samples. *Sankhya*, 37, 259–270.
- Stein, P. (2000). Careful interpolation yields useful information. *Quality Progress*, 33, 67–69.

- Strohmeier, B. (1992). *APL Idiom Library*. USA: STSC Inc.
- STSC (1992). *APL*PLUS II/386 Reference Manual*. USA: STSC Inc.
- Student (1908). The probable error of a mean. *Biometrika*, 6, 1–25.
- Swan, A. V. (1969). Maximum likelihood estimation from grouped and censored normal data. *Applied Statistics*, 18, 110–114.
- Thoman, D. R., Bain, L. J., & Antle, C. E. (1970). Reliability and tolerance limits in the Weibull distribution. *Technometrics*, 12, 363–371.
- Tricker, T. (1984). Effects of rounding data sampled from the exponential distribution. *Journal of Applied Statistics*, 11, 54–87.
- Vardeman, S. B. (2005). Sheppard´s correction for variances and the „quantization noise model“. *IEEE Transactions on Instrumentation and Measurement*, 54, 2117–2119.
- Vardeman, S. B. & Lee, C.-S. (2005). Likelihood-based statistical estimation from quantized data. *IEEE Transactions on Instrumentation and Measurement*, 54, 409–414.
- Wald, A. (1943). Test of statistical hypothesis concerning several parameters when the number of observations is larger. *Philosophical Transactions of the American Mathematical Society London A*, 54, 462–482.
- Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypothesis. *Annals of Mathematical Statistics*, 9, 60–62.
- Wilks, S. S. (1962). *Mathematical Statistics*. New York: John Wiley.
- Wold, H. (1934). Sheppard´s correction formulae in several variables. *Skandinavisk Aktuaritiedskrift*, 17, 248–255.