

Block-Maxima von Zeitreihen mit variablem rechten Randpunkt

Masterarbeit

Vorgelegt von

Torben Staud

aus Neuss, Deutschland

Angefertigt am

Mathematischen Institut

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

30. September 2021

Betreuer: Prof. Dr. Axel Bücher

Danksagung

Mein Dank gilt zuallererst Universitätsprofessor Axel Bücher für den Vorschlag dieses spannenden Themas. Weiterhin bin ich für die Betreuung vor und während des Verfassens dieser Arbeit dankbar.

Außerdem danke ich M.Sc. Tobias Jennessen unter anderem für die fachliche Unterstützung im Rahmen der Extremwertstatistik und M.Sc. Leandra Zanger für die Klärung von Fragen, die sich auf die Programmierung mit *R* bezogen.

Anlässlich des Abschlusses meines Masterstudiums möchte ich den Mitgliedern des Mathematischen Instituts der Universität Düsseldorf, insbesondere denen des Lehrstuhls der Mathematischen Statistik und Wahrscheinlichkeitstheorie an der Universität Düsseldorf für das breite Angebot spannender stochastischer Vorlesungen danken. Diese haben mein Interesse für Extremwertstatistik geweckt und detaillierte Einsichten in stochastische Bereiche ermöglicht.

Zuletzt bedanke ich mich bei meiner Partnerin und meiner Familie, die mir während des Studiums und des Erstellungsprozesses stets unterstützend zur Seite standen.

Inhaltsverzeichnis

Einleitung	2
1 Allgemeine Grundlagen	3
1.1 Statistische Grundlagen	3
1.2 Extremwerttheoretische Grundlagen	19
2 Asymptotische Resultate über die Verteilung Maxima von nicht-stationären Dreiecksschemata	25
2.1 Rahmenbedingungen	25
2.2 Erste Asymptotik zeilenweiser Unabhängigkeit	26
2.3 Zweite Asymptotik zeilenweiser Unabhängigkeit	40
3 Statistische Modellierung	57
3.1 Das explizite Stein'sche Modell	57
3.2 Die klassische Block-Maxima Methode	62
3.3 Die Block-Maxima Methode im expliziten Stein'schen Modell	65
4 Simulationsstudien	69
4.1 Eine Simulationsstudie im expliziten Stein'schen Modell	69
4.2 Eine Simulationsstudie im Weibull-Modell	78
Fazit und Aussicht	84
Appendix	85
Literaturverzeichnis	95

Einleitung

Die Extremwerttheorie ist ein Zweig der Mathematik, welcher sich mit dem Auftreten extremer Ereignisse wie etwa Höchsttemperaturen oder ein Jahrhundertwasser beschäftigt. Ziel ist hierbei, Gesetze im Zufall zu finden und diese zu quantifizieren. Der grundlegende Satz von Fisher-Tippet-Gnedenko [Fisher und Tippet] besagt, dass unter bestimmten Bedingungen nur drei Verteilungsklassen als Grenzwert für eine Maximumsfolge an Zufallsvariablen in Frage kommen, welche sich wiederum zu einer bekannten Verteilungsfamilie zusammenfassen lassen.

In Anwendungen wie der Hydrologie, Klimatologie oder dem Aktienmarkt lässt sich der angesprochene Kontext nicht mehr legitimieren, sodass eine Aufweichung an Voraussetzungen impliziert ist. Der Artikel [Stein] trägt dem Rechnung und enthält unter anderem zwei Grenzergebnisse für nicht-stationäre Zeitreihen, wobei das zweite in einer anderen Grenzverteilungsklasse als der klassische Satz von Fisher-Tippet resultiert.

Ein Ziel dieser Masterthesis ist die Ausarbeitung und Weiterentwicklung zentraler Resultate des Artikels. So zeigen wir, dass die Klasse der Grenzverteilungen für nicht-triviale Wahlen von bestimmten Funktionen disjunkt zu der der generalisierten Extremwertverteilung ist. Außerdem untersuchen wir bestimmte Subklassen der potentiellen Grenzverteilungen, welche bestimmte Tailrelationen im Bezug auf die Weibull-Klasse erfüllen.

Dadurch motiviert, wird eine dieser Klassen im Rahmen von asymptotischer Statistik beleuchtet, wobei vorher die Block-Maxima Methode ausführlich vorgestellt und auf den Kontext des Artikels von Stein erweitert wird. Zuletzt werden Simulationsstudien durchgeführt, in denen unter anderem der gut untersuchte GEV Maximum-Likelihood Schätzer mit einem in dieser Arbeit vorgeschlagenen verglichen wird.

Grundkenntnisse, die diese Arbeit betreffen, wurden [Bücher(a)], [Bücher(b)] und [Bücher(c)] entnommen. Für das Verständnis ist eine vorige Auseinandersetzung mit diesen Themen, trotz der Bemühung um Grundlagenerklärung, unabdingbar. Im Appendix sind Resultate und Verteilungen zu finden, welche innerhalb der Arbeit genutzt werden und so das Nachschlagen dieser ersparen.

1 Allgemeine Grundlagen

Dieses Kapitel hat als Ziel, Grundlagen zu wiederholen oder zu erarbeiten, welche für diese Arbeit zentral sind und den Lesefluss erleichtern. Dabei gliedern wir diese in statistische und extremwerttheoretische Grundlagen.

In dieser Arbeit bilden alle auftretenden Zufallsvariablen X vom selben, ohne Einschränkung reichhaltigen¹, Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ ab.

1.1 Statistische Grundlagen

Definition 1.1.1 (Statistisches Modell).

Ein Tripel $\mathcal{E} := (\mathcal{X}, \mathcal{B}, \mathcal{P})$ heißt **statistisches Modell**, falls

- (i) \mathcal{X} eine beliebige nicht-leere Menge ist, der **Stichprobenraum**,
- (ii) \mathcal{B} eine σ -Algebra auf \mathcal{X} ist, die **Ereignisalgebra**,
- (iii) $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ eine Familie von Wahrscheinlichkeitsmaßen P_θ auf $(\mathcal{X}, \mathcal{B})$ ist, welche durch eine beliebige nicht-leere Menge Θ , der **Parameterraum**, indiziert ist.

Bemerkung 1.1.2.

1. Man beachte, dass so in der Situation $P_{\theta_1} = P_{\theta_2}$ mit $\theta_1 \neq \theta_2$ eines der Maße verworfen wird.² Dies ist unter dem Gesichtspunkt der *Identifizierbarkeit* relevant, die wir später untersuchen werden.
2. Der Anschauung zuträglich lassen sich die P_θ als Bildmaß einer Zufallsvariablen $X : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$ realisieren. Formal nimmt man dafür folgenden Standpunkt ein: $\mathcal{Q} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ und $\mathbb{P}_\theta^X = P_\theta$. Wir werden in dieser Arbeit beide Formulierungen statistischer Modelle nutzen.
3. Oftmals werden wir für konkret gegebene, von dem Parameter θ abhängige, Verteilungsfunktionen F_θ die Notation $\mathcal{P} = \{F_\theta : \theta \in \Theta\} := \{P_\theta : \theta \in \Theta\}$ nutzen, wobei P_θ das eindeutig zu F_θ assoziierte Maß bezeichne.

Beispiel 1.1.3.

1. Ein denkbar simples Modell ist $\mathcal{E} = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P})$ mit $\mathcal{P} := \{\delta_\theta : \theta \in \mathbb{R}\}$. Alternativ geht dieses Modell hervor aus: $X \sim \delta_\theta$ mit $\theta \in \mathbb{R}$ unbekannt.

¹ Damit ist gemeint, dass sich unabhängige Familien von Zufallsvariablen auf dem Wahrscheinlichkeitsraum definieren lassen.

² Es ist durchaus möglich einer anderen Konvention zu folgen, in der die Maße in einer Familie zusammengefasst werden, welche doppeltes Auftreten von Maßen ermöglicht. Dies ist jedoch für die weitere mathematische Behandlung eine rein kosmetische Frage.

2. Das (n -fache) **Produktmodell** ist das aus einem zugrundeliegenden Modell \mathcal{E} hervorgehende Konstrukt $\mathcal{E}^{\otimes n} := (\mathcal{X}^n, \mathcal{B}^{\otimes n}, \mathcal{P}^{\otimes n})$. Mittels der Zufallsvariablencharakterisierung lässt es sich als Realisierung von n vielen unabhängig und identisch nach P_θ verteilten Zufallsvariablen X_1, \dots, X_n auffassen, wobei $\theta \in \Theta$ unbekannt ist.

Ein zentrales Ziel der Statistik ist es, aus der Beobachtung einer Realisierung x unter dem Maß P_θ Rückschlüsse auf den unbekannten Parameter θ zu ziehen und die Qualität dieser Rückschlüsse zu bemessen.

Definition 1.1.4.

Es sei \mathcal{E} ein statistisches Modell und $\gamma : (\Theta, \mathcal{B}_\Theta) \rightarrow (\Gamma, \mathcal{B}_\Gamma)$ messbar. Dann heißt $\hat{\gamma}$ **Schätzer für $\gamma(\theta)$** , falls $\hat{\gamma} : (\mathcal{X}, \mathcal{B}) \rightarrow (\Gamma, \mathcal{B}_\Gamma)$ messbar ist.

Bemerkung 1.1.5.

1. Die Notation $\hat{\gamma}$ ist suggestiv. Da an den Schätzer nur Messbarkeits- und Bildraumsforderungen gestellt werden, kann er völlig relationslos im Hinblick auf den Parameter agieren. Ein simples Beispiel hierauf bezogen ist ein konstanter Schätzer. Insbesondere lässt die Tatsache, dass eine Abbildung ein Schätzer ist, keine unmittelbaren Rückschlüsse auf etwaige *Qualität* dessen zu. Vergleiche auch mit dem auf die Bemerkung folgenden Beispiel.
2. Häufig, so auch in dieser Arbeit, ist der Parameter θ und nicht eine Funktionsanwendung darauf von Interesse. Formal: $\gamma = id_\Theta$ und $\Gamma = \Theta$.

Beispiel 1.1.6.

Im ersten Modell aus Beispiel 1.1.3 ist die Abbildung $\hat{\theta}_0(x) \equiv 0$ zwar ein Schätzer, jedoch nutzt sie keine der vorhandenen Informationen aus x .

Die Abbildung $\hat{\theta}(x) := x$ ist auch ein Schätzer und erfüllt sogar für alle $\theta \in \mathbb{R}$

$$P_\theta\{\hat{\theta} = \theta\} = P_\theta(\{x : \hat{\theta}(x) = \theta\}) = \delta_\theta(\{\theta\}) = 1.$$

Das heißt, für jede unbekannte Wahrheit θ schätzt $\hat{\theta}$ fast sicher den wahren Parameter. Es sei angemerkt, dass dies ein Ausnahmefall ist und auch nur deshalb möglich ist, da das Modell eine nahezu deterministische Struktur hat.

Bei der konkreten Betrachtung oder Erstellung eines Modells sollte im Hinblick auf geeignete Schätzbarkeit des Parameters θ überprüft werden, ob verschiedene Wahrheiten in Realisierungen differenzierbar voneinander sind. Dies führt zu folgendem Begriff.

Definition 1.1.7 (Identifizierbarkeit).

Ein Modell \mathcal{E} heißt **identifizierbar**, falls die Abbildung $\theta \mapsto P_\theta$ injektiv ist.

Beispiel 1.1.8.

Wir betrachten das Modell $\mathcal{E} = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \{\mathcal{N}(\mu^2, 1) : \mu \in \mathbb{R}\})$, was aus $X \sim \mathcal{N}(\mu^2, 1)$

mit unbekanntem $\mu \in \mathbb{R}$ hervorgeht. Instantan sieht man, dass $P_\theta = P_{-\theta}$, sodass das Modell nicht identifizierbar ist und eine sinnvolle Schätzung von μ nur bis auf das Vorzeichen möglich ist. Es ist möglich, den Parameterraum derart einzuschränken, dass das neue Modell identifizierbar wird. Dies muss dann unter Anwendergesichtspunkten kritisch durchdacht werden.

Eine sehr häufige Situation betrifft den Übergang zu unabhängigen Versuchswiederholungen, also dem Produktmodell aus Beispiel 1.1.3. In diesem Falle wäre es wünschenswert, dass die Identifizierbarkeit des Grundmodells sich auf die des Produktmodells fortpflanzt. In der Tat gilt.

Lemma 1.1.9.

Es sei \mathcal{E} ein statistisches Modell mit zugehörigem Produktmodell $\mathcal{E}^{\otimes n}$; dann gilt. \mathcal{E} ist genau dann identifizierbar, wenn $\mathcal{E}^{\otimes n}$ es ist.

Beweis. Für die erste Implikation betrachten wir $\theta_1, \theta_2 \in \Theta$ mit $P_{\theta_1}^{\otimes n} = P_{\theta_2}^{\otimes n}$. Dann gilt

$$P_{\theta_1}(A) = P_{\theta_1}^{\otimes n}(A \times \mathcal{X}^{n-1}) = P_{\theta_2}^{\otimes n}(A \times \mathcal{X}^{n-1}) = P_{\theta_2}(A) \quad (A \in \mathcal{B}),$$

sodass $P_{\theta_1} = P_{\theta_2}$ und die Identifizierbarkeit von \mathcal{E} die Gleichheit $\theta_1 = \theta_2$ schließen lässt.

Für die Rückrichtung sei beachtet, dass $P_{\theta_1} = P_{\theta_2}$ bereits $P_{\theta_1}^{\otimes n} = P_{\theta_2}^{\otimes n}$ und damit nach Voraussetzung auch $\theta_1 = \theta_2$ impliziert. \square

Um die Identifizierbarkeit eines Produktmodells zu überprüfen, reicht es also, das Basismodell diesbezüglich zu überprüfen.

Von nun an bezeichne \mathcal{E} ein statistisches Modell mit $\mathcal{X} = \mathbb{R}^d$ und $\mathcal{B} := \mathcal{B}(\mathbb{R}^d)$, das aus einer d -dimensionalen Zufallsvariable X hervorgeht.

Es sei ein Grenzwertsatzes gegeben. Das heißt es existieren normalisierende Folgen $a_n > 0, b_n \in \mathbb{R}$ und eine Statistik T , sodass für Zufallsvariablen X_1, \dots, X_n unter gewissen Bedingungen die Verteilungskonvergenz

$$\frac{T(X_1, \dots, X_n) - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} G$$

gegen eine Grenzverteilung G aus einer Klasse \mathcal{G} gilt. Man denke bei T an das Maximum oder die Summe der Argumente. Obige Konvergenz lässt sich mittels des Lemmas von Slutsky auf affin-lineare Transformationen ausweiten, sodass man häufig als Grenzwertklasse Verteilungen zulässt, welche vom *selben Typ* wie G sind, siehe Definition 1.1.30, das heißt als $\sigma \cdot G + \mu$ in Verteilung darstellbar sind. Statistisch ist dann häufig die Verteilungsfamilie $\{(\sigma \cdot G + \mu)^{\otimes N} : G \in \mathcal{G}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$ von Interesse. Dies führt zu den folgenden Begriffen; man vergleiche auch für den folgenden Begriff mit [Rüschendorf, Beispiel 2.1.9]

Definition 1.1.10 (Lokationsmodell).

Ein Modell \mathcal{E} heißt **Lokationsmodell**, falls ein Wahrscheinlichkeitsmaß P auf \mathcal{B} mit $\mathcal{P} = \{ \delta_\theta * P : \theta \in \Theta \}$ existiert, wobei $\Theta \subset \mathbb{R}$. Mit der Minkowski-Summe und $e := e_1 + \dots + e_d$, wenn e_i die kanonischen Einheitsvektoren des \mathbb{R}^d bezeichnen, lässt sich die Faltung $(\delta_\theta * P)(\cdot)$ auch als $P(\cdot - e\theta)$ schreiben. Dabei bezeichnen wir hier mit $A - b$ die Menge $A + (-b) := A + \{-b\} = \{a - b : a \in A\}$.

Beispiel 1.1.11.

Das Modell, was aus $X_1, \dots, X_n \sim \mathcal{U}[\theta, \theta + 1]$ hervorgeht, ist ein Lokationsmodell, da für $A_i \in \mathcal{B}(\mathbb{R})$

$$P_\theta \left(\prod_{i=1}^n A_i \right) = \mathcal{U}^{\otimes n}[\theta, \theta + 1] \left(\prod_{i=1}^n A_i \right) = \prod_{i=1}^n \mathcal{U}[0, 1](A_i - e\theta) = \mathcal{U}^{\otimes n}[0, 1] \left(\prod_{i=1}^n A_i - e\theta \right)$$

gilt und somit aus dem Maßeindeutigkeitssatz $P_\theta(\cdot) = \mathcal{U}^{\otimes n}[0, 1](\cdot - e\theta)$ folgt.

Der folgende Begriff formalisiert die Erweiterung eines Modells um Lokationen.

Definition 1.1.12.

Für ein Modell \mathcal{E} heißt $\mathcal{E}_L := (\mathcal{X}, \mathcal{B}, \mathcal{P}_L)$ **Lokationserweiterung von \mathcal{E}** oder auch **das von \mathcal{E} erzeugte Lokationsmodell**, wobei

$$\mathcal{P}_L := \{ P_{\theta, \mu} : \theta \in \Theta, \mu \in \mathbb{R} \} := \{ P_\theta(\cdot - e\mu) : \theta \in \Theta, \mu \in \mathbb{R} \}.$$

Erstrebenswert ist es, eine Beziehung zwischen der Identifizierbarkeit eines Modells und seiner Lokationserweiterung zu etablieren. Evident ist, wenn das von erzeugte Skalenmodell identifizierbar ist, so auch das zugrundeliegende Modell \mathcal{E} , da sich wegen $0 \in \mathbb{R}$ die Verteilungsklasse \mathcal{P} in \mathcal{P}_L einbetten lässt. Die Umkehrung gilt jedoch nicht:

Beispiel 1.1.13.

Betrachte $\mathcal{P} = \{ \mathcal{U}[\theta, \theta + 1] : \theta \in \mathbb{R} \}$ und das dazugehörige erzeugte Skalenmodell mit $\mathcal{P}_L = \{ \mathcal{U}[\theta, \theta + 1](\cdot - \mu) : \theta \in \mathbb{R}, \mu \in \mathbb{R} \}$. Dann ist \mathcal{E} identifizierbar, da für verschiedene θ_1, θ_2 der Träger der Maße $P_{\theta_1}, P_{\theta_2}$ nicht identisch ist.

Jedoch ist \mathcal{P}_L nicht identifizierbar, da $\mathcal{U}[\theta, \theta + 1](\cdot - \mu) = \mathcal{U}[\mu + \theta, \mu + \theta + 1](\cdot)$ und somit beispielsweise $P_{1, -1} = \mathcal{U}[0, 1] = P_{-1, 1}$ gilt.

Das vorangegangene Beispiel motiviert folgendes Attribut.

Definition 1.1.14.

Ein Modell \mathcal{E} heißt **lokationsfrei**, falls für $\theta_1, \theta_2 \in \Theta$ und $\mu \in \mathbb{R}^\times$ bereits $P_{\theta_1}(\cdot) \neq P_{\theta_2}(\cdot - \mu e)$ gilt oder äquivalent: Aus $P_{\theta_1}(\cdot) = P_{\theta_2}(\cdot - \mu e)$ folgt bereits $\mu = 0$.

Beispiel 1.1.15.

1. Das aus $X \sim \mathcal{N}(0, \sigma^2)$ mit $\sigma^2 > 0$ unbekannt hervorgehende Modell ist lokationsfrei, was wir wie folgt einsehen:

Angenommen für $\theta_1, \theta_2 > 0, \mu \in \mathbb{R}$ gilt $\mathcal{N}(0, \sigma_1^2)(\cdot) = \mathcal{N}(0, \sigma_2^2)(\cdot - \mu)$. Wegen $\mathcal{N}(0, \sigma_2^2)(\cdot - \mu) = \mathcal{N}(\mu, \sigma_2^2)(\cdot)$ und Eigenschaften der Normalverteilung muss dann bereits $\mu = 0$ folgen, was die Lokationsfreiheit nach sich zieht.

2. Das Modell aus Beispiel 1.1.13 ist nicht lokationsfrei, da $P_1(\cdot) = P_{-1}(\cdot - 2)$ gilt.

Das nächste Lemma präsentiert ein hinreichendes Kriterium.

Lemma 1.1.16.

Gilt für ein Modell \mathcal{E} , dass aus $P_{\theta_1}(\cdot) = P_{\theta_2}(\cdot - \mu e)$ bereits $P_{\theta_1} = P_{\theta_2}$ folgt, so ist \mathcal{E} lokationsfrei.

Beweis. Sei für $\theta_1, \theta_2 \in \Theta, \mu \in \mathbb{R}$ die Gleichheit der Maße $P_{\theta_1}(\cdot) = P_{\theta_2}(\cdot - \mu e)$ gegeben. Dann erhalten wir nach Voraussetzung $P_{\theta_1} = P_{\theta_2}$, also $P_{\theta_1}(\cdot) = P_{\theta_1}(\cdot - \mu e)$. Angenommen es gilt $\mu \neq 0$. Wir wählen eine d -dimensionale Zufallsvariable X mit $\mathbb{P}^X = P_{\theta_1}$ und schließen mittels Fourier-Transformierten, dass

$$\varphi_X(t) = \varphi_X(t) \cdot \exp(i\langle t, \mu e \rangle) \quad (t \in \mathbb{R}^d).$$

Wegen der Stetigkeit von Fourier-Transformierten und der Tatsache $\varphi_X(0) = 1$ existiert eine (offene) Umgebung U von 0 auf welcher $\varphi_X \neq 0$ gilt. Damit folgt aber $1 \equiv \exp(i\langle t, \mu e \rangle)$ auf U , was wegen

$$|\partial_{t_1} \exp(i\langle t, \mu e \rangle)| = |i\mu \exp(i\langle t, \mu e \rangle)| = |\mu| \neq 0 = |\partial_{t_1} 1|$$

widersprüchlich wäre, sodass $\mu = 0$ und die Lokationsfreiheit folgt. \square

Folgende Verbindung zwischen Lokationserweiterungen und Identifizierbarkeit lässt sich nun formulieren.

Lemma 1.1.17.

Das Modell \mathcal{E} ist genau dann lokationsfrei und identifizierbar, wenn die Lokationserweiterung \mathcal{E}_L identifizierbar ist.

Beweis. Für die Hinrichtung seien $\theta_1, \theta_2 \in \Theta, \mu_1, \mu_2 \in \mathbb{R}$ und $P_{\theta_1}(\cdot - \mu_1 e) = P_{\theta_2}(\cdot - \mu_2 e)$. Dann gilt

$$P_{\theta_1}(A) = P_{\theta_1}(A - (\mu_1 - \mu_1)e) = P_{\theta_2}(A - (\mu_2 - \mu_1)e) \quad (A \in \mathcal{B}),$$

sodass aus der Tatsache, dass \mathcal{E} lokationsfrei ist, $\mu_2 - \mu_1 = 0$ folgt.

Damit erhalten wir $P_{\theta_1} = P_{\theta_2}$ und aus der Identifizierbarkeit von \mathcal{E} folgt damit $\theta_1 = \theta_2$. Das heißt aber $(\theta_1, \mu_1) = (\theta_2, \mu_2)$, sodass \mathcal{E}_L identifizierbar ist.

Im Rahmen der Rückrichtung sei $P_{\theta_1}(\cdot) = P_{\theta_2}(\cdot - \mu e)$ vorgegeben. Dies lässt sich als $P_{\theta_1, 0} = P_{\theta_2, \mu}$ auffassen, sodass die Identifizierbarkeit von \mathcal{E}_L den Schluss validiert. Dass \mathcal{E} identifizierbar ist, folgt völlig analog. \square

Die Lokationsfreiheit im Bezug auf Produktmodelle verhält sich angenehm:

Lemma 1.1.18.

Das Modell \mathcal{E} ist genau dann lokationsfrei, wenn $\mathcal{E}^{\otimes n}$ lokationsfrei ist.

Beweis. Für die Hinrichtung sei $P_{\theta_1}^{\otimes n}(\cdot) = P_{\theta_2}^{\otimes n}(\cdot - \mu e_{n \cdot d})$. Wir schreiben e_m für den Vektor $(1, \dots, 1)^T \in \mathbb{R}^m$. Damit gilt

$$\begin{aligned} P_{\theta_1}(A) &= P_{\theta_1}^{\otimes n}(A \times \mathcal{X}^{n-1}) = P_{\theta_2}^{\otimes n}(A \times \mathcal{X}^{n-1} - \mu e_{n \cdot d}) \\ &= P_{\theta_2}^{\otimes n}((A - \mu e_d) \times (\mathcal{X}^{n-1} - \mu e_{n \cdot d - d})) = P_{\theta_2}^{\otimes n}((A - \mu e_d) \times \mathcal{X}^{n-1}) \\ &= P_{\theta_2}(A - \mu e_d) \quad (A \in \mathcal{B}), \end{aligned}$$

sodass nach Voraussetzung $\mu = 0$ gilt. Wir merken an, dass es an dieser Stelle wichtig ist, dass \mathcal{X} invariant unter Subtraktion ist.

Die Argumentation der Rückrichtung erfolgt analog zu dem Beweis der Rückrichtung von Lemma 1.1.9 und wird an dieser Stelle übersprungen. \square

Eng verwandt mit dem Lokationsmodell ist folgendes Modell, man vergleiche auch mit [Rüschendorf, Bemerkung 8.1.13].

Definition 1.1.19 (Skalenmodell).

Ein Modell \mathcal{E} heißt **Skalenmodell**, falls ein Wahrscheinlichkeitsmaß P auf \mathbb{R}^d derart existiert, dass $\mathcal{P} = \{P(\cdot/\sigma) : \sigma > 0\}$, wobei wir in diesem Fall die Notation $A/\sigma := A \cdot (1/\sigma) := \{a \cdot 1/\sigma : a \in A\}$ für $A \subset \mathbb{R}^d$ nutzen.

Beispiel 1.1.20.

Das durch unabhängige $X_1, \dots, X_n \sim \mathcal{N}(0, \sigma^2)$, $\sigma > 0$ induzierte Modell ist ein Skalenmodell, da aus Eigenschaften der Normalverteilung $\mathcal{N}_d(0, \sigma^2 I_d)(A) = \mathcal{N}_d(0, I_d)(A/\sigma)$ folgt.

Wie im Lokationskontext lässt sich auch hier eine kanonische Erweiterung definieren.

Definition 1.1.21.

Zu einem Modell \mathcal{E} heißt $\mathcal{E}_S := (\mathcal{X}, \mathcal{B}, \mathcal{P}_S)$ **Skalenerweiterung von \mathcal{E}** oder auch **das von \mathcal{E} erzeugte Skalenmodell**, wobei

$$\mathcal{P}_S := \{P_{\theta, \sigma} : \theta \in \Theta, \sigma \in \mathbb{R}^+\} := \{P_{\theta}(\cdot/\sigma) : \theta \in \Theta, \sigma \in \mathbb{R}^+\}.$$

Wie in Beispiel 1.1.13 lässt sich ein identifizierbares Modell angeben, dessen Skalenerweiterung nicht mehr identifizierbar ist. Dies motiviert den Begriff:

Definition 1.1.22.

Das Modell \mathcal{E} heißt **skalenfrei**, falls für alle $\theta_1, \theta_2 \in \Theta$, $\sigma \in \mathbb{R}^+ \setminus \{1\}$ bereits $P_{\theta_1}(\cdot) \neq P_{\theta_2}(\cdot/\sigma)$ folgt. Das ist gleichbedeutend zu: $P_{\theta_1}(\cdot) = P_{\theta_2}(\cdot/\sigma)$ impliziert $\sigma = 1$.

Beispiel 1.1.23.

1. Das Exponentialmodell $\{ \text{Exp}(\lambda) : \lambda > 0 \}$ ist nicht skalenfrei, da $\text{Exp}(1/\lambda)(\cdot) = \text{Exp}(1)(\cdot/\lambda)$ gilt. Man nennt es **Ratenmodell**.
2. Das Modell $\{ \text{Poi}(\lambda^2) : \lambda \in \mathbb{R}^\times \}$ ist skalenfrei, aber offensichtlich nicht identifizierbar.

Beweis. Es gelte $\text{Poi}(\lambda_1^2)(\cdot) = \text{Poi}(\lambda_2^2)(\cdot/\sigma)$. Dann folgt aus der Tatsache, dass der Träger jeder Poissonverteilung \mathbb{N} ist, zunächst $k/\sigma \in \mathbb{N}$ für alle $k \in \mathbb{N}$ und damit auch $\sigma \in \mathbb{N}$. Insbesondere gilt damit für alle Primzahlen, dass $\sigma \in \mathbb{N}$ sie teilt. Mithin folgt $\sigma = 1$, also die Behauptung. \square

Auch hier gilt im Bezug auf den Übergang zu Produktmodellen.

Lemma 1.1.24.

Für ein statistisches Modell \mathcal{E} gilt, dass \mathcal{E} genau dann skalenfrei ist, wenn $\mathcal{E}^{\otimes n}$ skalenfrei ist.

Beweis. Die Argumentation kann analog zu der im Beweis von Lemma 1.1.18 geführt werden. \diamond

Ein hinreichendes Kriterium für Skalenfreiheit lässt sich wie folgt formulieren.

Lemma 1.1.25.

Das Maß δ_0 ist das einzige Wahrscheinlichkeitsmaß $P \in M^1(\mathbb{R}^d)$ für das ein $\sigma \in \mathbb{R}^+ \setminus \{1\}$ mit $P(\cdot) = P(\cdot/\sigma)$ existiert.

Insbesondere gilt daher für statistische Modelle \mathcal{E} mit $\delta_0 \notin \mathcal{P}$, dass \mathcal{E} skalenfrei ist, falls sich aus $P_{\theta_1}(\cdot) = P_{\theta_2}(\cdot/\sigma)$ bereits $P_{\theta_1} = P_{\theta_2}$ schließen lässt.

Beweis. Sei F die zu dem Wahrscheinlichkeitsmaß P assoziierte Verteilungsfunktion $F(t) := P([-\infty, t_1] \times \dots \times [-\infty, t_d])$. Dann gilt $F(t) = F(t/\sigma^n)$ für $n \in \mathbb{N}$. Ohne Einschränkung sei $\sigma > 1$. Es folgt für $t \in (\mathbb{R}^+)^d$ die Gleichheit $F(t) = F(t \cdot \sigma^n) \xrightarrow{n} 1$, sodass die Rechtsstetigkeit $F(t) = 1$ für alle $t \in (\mathbb{R}_0^+)^d$ generiert. Für t außerhalb folgt $F(t) = 0$, womit die Verteilungsfunktion von P mit der von δ_0 übereinstimmt und $P = \delta_0$ impliziert wird. Der Fall $\sigma < 1$ funktioniert sehr ähnlich, sodass insgesamt $P = \delta_0$ folgt. \square

Eine allgemeinere Charakterisierung ist:

Lemma 1.1.26.

Die Skalenfreiheit und Identifizierbarkeit von \mathcal{E} sind gleichbedeutend dazu, dass die Skalenerweiterung identifizierbar ist.

Beweis. Nur der Beweis von links nach rechts:

Wir setzen $P_{\theta_1}(\cdot/\sigma_1) = P_{\theta_2}(\cdot/\sigma_2)$ voraus, sodass für $A \in \mathcal{X}$ folgt

$$P_{\theta_1}(A) = P_{\theta_1}(A \cdot \sigma_1/\sigma_1) = P_{\theta_2}\left(A/\frac{\sigma_2}{\sigma_1}\right). \quad (1.1)$$

Die Skalenfreiheit von \mathcal{E} impliziert $\sigma_2/\sigma_1 = 1$. Die Gleichung (1.1) liefert die Beziehung $P_{\theta_1} = P_{\theta_2}$, sodass aus der Identifizierbarkeit von \mathcal{E} schon $\theta_1 = \theta_2$ und daher $(\theta_1, \sigma_1) = (\theta_2, \sigma_2)$ gilt, was die Behauptung validiert. \square

Als natürliche und in der Einleitung schon angesprochene Erweiterung definieren wir.

Definition 1.1.27.

Ein Modell \mathcal{E} heißt **Lokations- und Skalenmodell**, falls ein Wahrscheinlichkeitsmaß P auf \mathbb{R}^d so existiert, dass $\mathcal{P} = \{P_{\mu, \sigma} : \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+\}$, wobei $P_{\mu, \sigma}(\cdot) = P((\cdot - \mu e)/\sigma)$.

Definition 1.1.28.

Zu einem Modell \mathcal{E} definieren wir die **Lokations- und Skalenerweiterung von \mathcal{E}** oder auch die **von \mathcal{E} erzeugte Lokations- und Skalenklasse** als $\mathcal{E}_{LS} := (\mathcal{X}, \mathcal{B}, \mathcal{P}_{LS})$, wobei

$$\mathcal{P}_{LS} := \{P_{\theta, \mu, \sigma} : \theta \in \Theta, \mu \in \mathbb{R}, \sigma > 0\} := \{P_{\theta}((\cdot - \mu e)/\sigma) : \theta \in \Theta, \mu \in \mathbb{R}, \sigma > 0\}.$$

Nach den Lemmata 1.1.17, 1.1.26 wäre es wünschenswert wenn die Lokations- und Skalenfreiheit bereits die Identifizierbarkeit der Lokations- und Skalenerweiterung, unter Annahme der Identifizierbarkeit des Grundmodells, implizieren würden. Dies gilt jedoch nicht:

Beispiel 1.1.29.

Sei X eine nicht-degenerierte, das heißt \mathbb{P}_X ist kein Dirac-Maß, reellwertige Zufallsvariable und \mathcal{E} das aus $\theta X + \theta - 1$, $\theta > 0$ hervorgehende Modell. Dann ist \mathcal{E} skalenfrei, lokationsfrei und identifizierbar. Jedoch ist \mathcal{E}_{LS} nicht identifizierbar.

Beweis. Lokationsfreiheit: Nehmen wir $\theta_1 X + \theta_1 - 1 \stackrel{d}{=} \theta_2 X + \theta_2 - 1 + \mu$ an, so folgt

$$X \stackrel{d}{=} \frac{\theta_2}{\theta_1} X + \frac{\theta_2 - \theta_1 + \mu}{\theta_1}, \text{ sodass } F(x) = F(ax + b) = F(a^n x + a^{n-1}b + \dots + ab + b),$$

wobei $a := \theta_1/\theta_2$, $b := (\theta_1 - \theta_2 - \mu)/\theta_2$ und F die Verteilungsfunktion von X bezeichne. Mit dem Convergence of Types Theorem, beachte X ist nicht-degeneriert, angewandt auf $a_n := a^n$ und $b_n := a^{n-1}b + \dots + ab + b$ erhalten wir zunächst $a = 1$ und damit $\theta_2 = \theta_1$. Anschließend folgt aus der Konvergenz von $b_n = n \cdot b$, dass $b = 0$, also $\mu = 0$ gelten muss.

Skalenfreiheit: Nehmen wir $\theta_1 X + \theta_1 - 1 \stackrel{d}{=} \sigma \cdot (\theta_2 X + \theta_2 - 1 + \mu)$ an, so folgt wie oben $\sigma = 1$.

Identifizierbarkeit: Da X nicht-degeneriert ist, folgt aus $\theta_1 X + \theta_1 - 1 \stackrel{d}{=} \theta_2 X + \theta_2 - 1$ analog zu oben $\theta_1 = \theta_2$.

\mathcal{E}_{LS} ist nicht identifizierbar: Mit den Wahlen $(\theta_1, \mu_1, \sigma_1) = (1, 2, 2)$ und $(\theta_2, \mu_2, \sigma_2) = (2, 1, 1)$ erhalten wir $P_{\theta_1, \mu_1, \sigma_1} = P_{\theta_2, \mu_2, \sigma_2}$, sodass die Behauptung folgt. \square

Wir erinnern an den Begriff des *selben Typs* und führen die für die Identifizierbarkeit von Lokations- und Skalenerweiterung entscheidende Eigenschaft ein.

Definition 1.1.30 (Typ und Typenfreiheit).

- (a) Zwei Maße P, Q auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ heißen **vom selben Typ**, falls ein $b \in \mathbb{R}$ und ein $a > 0$ mit $P(\cdot) = Q((\cdot - e \cdot b)/a)$ existieren. Mittels assoziierten Zufallsvariablen X, Y mit $\mathbb{P}_X = P, \mathbb{P}_Y = Q$ ist dies gleichbedeutend zu $X \stackrel{d}{=} aY + e \cdot b$.
- (b) Ein Modell \mathcal{E} heißt **typenfrei**, falls aus $P_{\theta_1}(\cdot) = P_{\theta_2}((\cdot - e \cdot b)/a)$ bereits $(b, a) = (0, 1)$ folgt.

Ähnlich wie bei den vorangegangenen Begriffen reicht es auch hier aus, anstelle des Produktmodells, das Basismodell auf Typenfreiheit zu überprüfen.

Lemma 1.1.31.

Das Modell \mathcal{E} ist genau dann typenfrei, wenn das Produktmodell $\mathcal{E}^{\otimes n}$ typenfrei ist.

Beweis. Dies lässt sich ähnlich wie die vorangegangenen Produktmodellresultate begründen. \diamond

Für eine leichtere Nachweismöglichkeit der Typenfreiheit formulieren wir eine Variante eines d -dimensionalen *Convergence of Types Theorems*, wobei eine univariate Formulierung etwa in [Billingsley, Theorem 14.2] zu finden ist.

Lemma 1.1.32 (d -dimensionales Convergence of Types Theorem).

Seien Y, X, X_1, X_2, \dots d - dimensionale reelle Zufallsvariablen und $a_n, \tilde{a}_n > 0, b_n, \tilde{b}_n \in \mathbb{R}^d$ eine Folge von Konstanten mit $(X_n - b_n)/a_n \xrightarrow{d} X$ und $(X_n - \tilde{b}_n)/\tilde{a}_n \xrightarrow{d} Y$; wobei X und Y als derart regulär angenommen werden, dass keine der Marginalverteilungen von X und Y degeneriert ist.

Dann existieren $a > 0, b \in \mathbb{R}^d$ mit $a_n/\tilde{a}_n \xrightarrow{n} a$ und $(b_n - \tilde{b}_n)/\tilde{a}_n \xrightarrow{n} b$. Außerdem gilt $Y \stackrel{d}{=} aX + b$, sodass X und Y vom selben Typ sind.

Beweis. Wir schreiben $X := (X^{(1)}, \dots, X^{(d)})^T$ und analog für die anderen zufälligen Vektoren. Das Continous Mapping Theorem mit den Koordinatenprojektionen angewandt generiert

$$\frac{X_n^{(i)} - b_n^{(i)}}{a_n} \xrightarrow[n \rightarrow \infty]{d} X^{(i)} \quad \text{und} \quad \frac{X_n^{(i)} - \tilde{b}_n^{(i)}}{\tilde{a}_n} \xrightarrow[n \rightarrow \infty]{d} Y^{(i)} \quad (i = 1, \dots, d).$$

Das eindimensionale Convergence of Types Theorem lässt nun auf die Existenz von $a > 0, b^{(i)} \in \mathbb{R}$ schließen, die $a_n/\tilde{a}_n \xrightarrow{n} a, (b_n^{(i)} - \tilde{b}_n^{(i)})/\tilde{a}_n \xrightarrow{n} b^{(i)}$ erfüllen, dabei läuft i von 1 bis d . Setzen wir nun $b := (b^{(1)}, \dots, b^{(d)})^T$, so folgt der erste Teil der Behauptung. Für den zweiten Teil setzen wir $Y_n := (X_n - \tilde{b}_n)/\tilde{a}_n$ und folgern einerseits mittels der Voraussetzung und andererseits mittels des Lemmas von Slutsky

$$X \xleftarrow[n \rightarrow \infty]{d} \frac{X_n - b_n}{a_n} = \frac{Y_n - (b_n - \tilde{b}_n)/\tilde{a}_n}{a_n/\tilde{a}_n} \xrightarrow[n \rightarrow \infty]{d} \frac{Y - b}{a},$$

sodass wegen der Eindeutigkeit schwacher Grenzwerte die Verteilungsgleichheit folgt. \square

Mit dieser Erweiterung können wir die angesprochene hinreichende Bedingung für Typenfreiheit beweisen.

Korollar 1.1.33.

Sei \mathcal{E} ein Modell, dessen Maße kein degeneriertes Randmaß besitzen. Lässt sich aus der Beziehung $P_{\theta_1}(\cdot) = P_{\theta_2}((\cdot - eb)/a)$ bereits $P_{\theta_1} = P_{\theta_2}$ folgern, so ist \mathcal{E} typenfrei.

Beweis. Geben wir $P_{\theta_1}(\cdot) = P_{\theta_2}((\cdot - eb)/a)$ vor und assoziieren eine Zufallsvariable X zu dem Maß P_{θ_1} , so folgt aus der Voraussetzung $X \stackrel{d}{=} aX + eb$. Da keine Marginalverteilung von X degeneriert ist, liefert eine Anwendung von Lemma 1.1.32 auf $X_n := X$ die Tatsachen $b = 0$ und $a = 1$ (vergleiche mit dem Beweis von Beispiel 1.1.29). Mithin ist \mathcal{E} typenfrei. \square

Beispiel 1.1.34.

Das Paretomodell mit $\mathcal{P} = \{ \text{Par}(\alpha^2)^{\otimes n} : \alpha \in \mathbb{R}^\times \}$ ist typenfrei, jedoch nicht identifizierbar.³

Beweis. Seien $X \sim \text{Par}(\alpha_1^2), Y \sim \text{Par}(\alpha_2^2)$ mit $X \stackrel{d}{=} aY + b$. Nach Definition der Pareto-Verteilung, siehe Appendix A.1.6, gilt

$$F_X(t) = \begin{cases} 0 & , t \leq 1 \\ 1 - t^{-\alpha_1^2} & , t \geq 1 \end{cases} \stackrel{!}{=} F_{aY+b}(t) = \begin{cases} 0 & , t \leq a+b \\ 1 - \left(\frac{t-b}{a}\right)^{-\alpha_2^2} & , t \geq a+b \end{cases}$$

sodass zunächst $a + b = 1$ und anschließend für große t die Beziehung $t^{-\alpha_1^2} = ((t - b)/a)^{-\alpha_2^2}$ oder äquivalent $1 \equiv t^{\alpha_1^2 - \alpha_2^2} \cdot ((1 - b/t)/a)^{-\alpha_2^2}$ folgt. Eine Grenzwertbildung über die letztere Gleichung liefert $\alpha_1^2 = \alpha_2^2$, das heißt $P_{\alpha_1} = P_{\alpha_2}$. Aus Korollar 1.1.33 in Verbindung mit Lemma 1.1.31 folgt die Typenfreiheit. \square

Zuletzt geben wir eine Charakterisierung der Identifizierbarkeit von Lokations- und Skalenerweiterungen an.

Lemma 1.1.35.

Die Lokations- und Skalenerweiterung \mathcal{E}_{LS} des Modells \mathcal{E} ist genau dann identifizierbar, wenn \mathcal{E} typenfrei und identifizierbar ist.

Beweis. Die Hinrichtung ist klar.

Für die Rückrichtung nehmen wir $P_{\theta_1, \mu_1, \sigma_1} = P_{\theta_2, \mu_2, \sigma_2}$ an, sodass für $A \in \mathcal{B}$

$$P_{\theta_1}(A) = P_{\theta_1, \mu_1, \sigma_1}(\sigma_1 A + \mu_1 e) = P_{\theta_2, \mu_2, \sigma_2}(\sigma_1 A + \mu_1 e) = P_{\theta_2} \left(\frac{A - e \cdot (\mu_2 - \mu_1)/\sigma_1}{\sigma_2/\sigma_1} \right)$$

³ Die Pareto-Verteilung ist zusätzlich **skaleninvariant**, das heißt es existiert eine Funktion $s(\cdot)$ mit der Eigenschaft, dass $p(\lambda x) = s(\lambda) \cdot p(x)$ für $\lambda \in \mathbb{R}$ gilt. Dabei bezeichnet p die Dichte der Verteilung. Daraus folgt insbesondere $F(\lambda b) - F(\lambda a) = \lambda s(\lambda) \cdot (F(b) - F(a))$ für $a \leq b$. In dem Fall der Pareto-Verteilung gilt $s(\lambda) = \lambda^{-(\alpha+1)}$.

folgt. Da \mathcal{E} typenfrei ist, erhalten wir $\sigma_2/\sigma_1 = 1$, $(\mu_2 - \mu_1)/\sigma_1 = 0$, woraus seinerseits $\sigma_2 = \sigma_1$, $\mu_2 = \mu_1$ und letztlich $P_{\theta_1} = P_{\theta_2}$ folgt. Die Voraussetzung der Identifizierbarkeit von \mathcal{E} validiert daher den Schluss $(\theta_1, \mu_1, \sigma_1) = (\theta_2, \mu_2, \sigma_2)$. \square

Beispiel 1.1.36.

Aus dem Beispiel 1.1.34 und dem Lemma 1.1.35 entnehmen wir: Das von X_1, \dots, X_n unabhängig und identisch nach $\text{Par}(\alpha)$, $\alpha > 0$, verteilten Zufallsvariablen erzeugte Lokations- und Skalenmodell ist identifizierbar.

Bereits in Beispiel 1.1.6 haben wir uns mit der Frage der *Qualität* eines Schätzers beschäftigt, was in der nächsten Definition formalisiert werden wird. Dafür nehmen wir, wenn nicht anders explizit erwähnt, an, dass das statistische Modell \mathcal{E} durch ein σ -endliches Maß μ dominiert ist, sodass nach dem Satz von Radon-Nykodým Dichten $p_\theta := dP_\theta/d\mu$ für $\theta \in \Theta$ existieren.

Definition 1.1.37 (Grundlegende Gütekriterien für Schätzer).

Für $\Gamma \subset \mathbb{R}^k$ sei ein Schätzer $\hat{\gamma} \in L^1(P_\theta)$ von $\gamma(\theta)$ gegeben, wobei $\theta \in \Theta$.

1. Der Vektor $\text{Bias}_\theta(\hat{\gamma}) := E_\theta[\hat{\gamma}] - \gamma(\theta)$ heißt **Bias von $\hat{\gamma}$** und der Schätzer $\hat{\gamma}$ heißt **unbiased** (bisweilen auch **erwartungstreu**, **unverzerrt** oder **unverfälscht**), falls $\text{Bias}_\theta(\hat{\gamma}) = 0$.
2. Die Zahl $\text{MSE}_\theta(\hat{\gamma}) := E_\theta[\|\hat{\gamma} - \gamma(\theta)\|_2^2]$ heißt **Mean Squared Error von $\hat{\gamma}$** .

Bemerkung 1.1.38.

1. Bezeichnen wir mit $\text{Var}_\theta(\hat{\gamma}) := (\text{Var}_\theta(\hat{\gamma}^{(1)}), \dots, \text{Var}_\theta(\hat{\gamma}^{(k)}))^T$ die komponentenweise Varianz einer k -dimensionalen Zufallsvariable, so gilt $\text{MSE}_\theta(\hat{\gamma}) = \|\text{Bias}_\theta(\hat{\gamma})\|_2^2 + \|\text{Var}_\theta(\hat{\gamma})\|_1$. Für einen Beweis siehe Appendix A.2.2
2. Der Mean Squared Error ist ein populäres Beispiel für eine **Risikofunktion**. Es gibt weitere **Verlustfunktionen**, wie zum Beispiel die absolute Abweichung, die Abweichungen anders gewichten und daher zu anderen Risikofunktionen führen; siehe etwa [Casella und Berger, S. 348 ff.].
3. Es ist natürlich Erwartungstreue, also den konkreten Wert der Erwartung unter θ als den wahren Parameter zu fordern, anzusetzen, um Güte des Schätzers $\hat{\gamma}$ zu beschreiben. Die konkrete Forderung lässt sich jedoch nicht auf den MSE übertragen, da dies zur Folge hätte, dass der Schätzer unabhängig von dem Sample x immer die Wahrheit θ_0 schätzen würde, was in den allermeisten Modellen nicht möglich ist. Daher existiert keine simple kanonische Zahl für den MSE. Es existiert jedoch unter *Regularitätsannahmen* eine untere Schranke, **die Cramér-Rao-Schranke**.

Wir betrachten ein Weibull-Lokationsmodell, wobei in diesem Fall die Weibull-Verteilung aus der Extremwerttheorie gemeint ist. Dafür erinnern wir an die Definition

der Gammafunktion

$$\Gamma(z) := \int_0^\infty t^{z-1} \cdot e^{-t} dt, \quad \operatorname{Re}(z) > 0.$$

Mittels partieller Integration und Induktion über n zeigt man $\Gamma(n+1) = n!$ für natürliche Zahlen n .

Beispiel 1.1.39 (Weibull-Lokationsmodell).

Für festes $N \in \mathbb{N}, \sigma > 0$ betrachte $\mathcal{P} = \{ \operatorname{Wei}(\mu, \sigma, 1)^{\otimes N} : \mu \in \mathbb{R} \}$, siehe Appendix 1.2.2.

Der Parameter μ nimmt in diesem Falle die Rolle des rechten Randpunktes ein, so dass der Schätzer $\hat{\mu} := x_{N:N}$ ein wohlmotivierter Schätzer ist; tatsächlich ist er sogar der eindeutige Maximum-Likelihood Schätzer, wie wir später in Beispiel 1.1.42 sehen werden. Um den Bias und Mean Squared Error zu berechnen, bestimmen wir die Verteilung des Schätzers unter P_θ . Seien dafür X_1, \dots, X_N unabhängig und identisch nach $\operatorname{Wei}(\mu, \sigma, 1)$ verteilt.

$$\begin{aligned} P_\mu \{ \hat{\mu} \leq x \} &= \mathbb{P}_\mu \{ X_{N:N} \leq x \} \\ &= \begin{cases} \left(\exp\left(\frac{x-\mu}{\sigma}\right) \right)^N, & x \leq \mu \\ 1, & x > \mu \end{cases} = \begin{cases} \exp\left(\frac{x-\mu}{\sigma/N}\right), & x \leq \mu \\ 1, & x > \mu \end{cases} = \operatorname{Wei}(\mu, \sigma/N, 1)(x). \end{aligned}$$

Wir nutzen $\operatorname{Wei}(\mu, \sigma/N, 1) = \operatorname{GEV}(\mu - \sigma/N, \sigma/N, -1)$ und die Kenntnis, dass das erste Moment beziehungsweise die Varianz einer solchen Verteilung durch $(\mu - \sigma/N) + \sigma/N \cdot (\Gamma(1 - (-1)) - 1)/(-1) = (\mu - \sigma/N) - \sigma/N \cdot (1 - 1) = \mu - \sigma/N$ beziehungsweise durch $(\sigma/N)^2 \cdot (\Gamma(1 + 2) - \Gamma(1 + 1)^2)/(-1)^2 = (\sigma/N)^2$ gegeben ist; siehe Gleichung (1.3).

Wir haben gezeigt: $\operatorname{Bias}_\lambda(\hat{\mu}) = -\sigma/N$ und $\operatorname{MSE}_\lambda(\hat{\mu}) = 2 \cdot (\sigma/N)^2$. \square

Der Bias und der MSE aus dem vorangegangenen Beispiel hängen von der Samplegröße N ab und konvergieren für $N \rightarrow \infty$ gegen Null. Dies motiviert asymptotische Gütekriterien zu betrachten, wenn etwa subasymptotisch Eigenschaften wie Erwartungstreue noch nicht gegeben sind. Bevor wir diese Begriffe einführen, stellen wir die populäre Schätzmethodik der *Maximum-Likelihood Schätzung* vor.

Definition 1.1.40 (Maximum-Likelihood Schätzer).

Für ein Sample $x \in \mathcal{X}$ innerhalb eines dominierten k -parametrischen Modells \mathcal{E} heißt die Abbildung

$$\begin{aligned} L_x: \Theta &\rightarrow \mathbb{R}_0^+, & \theta &\mapsto p_\theta(x) && \text{Likelihood-Funktion,} \\ \ell_x: \Theta &\rightarrow \mathbb{R} \cup \{-\infty\}, & \theta &\mapsto \log L_x(\theta) && \text{Log-Likelihood-Funktion.} \end{aligned}$$

Im Falle der partiellen Differenzierbarkeit der Log-Likelihood-Funktion heißt die Abbildung

$$\dot{\ell}_x := \partial_\theta \ell_x: \Theta \rightarrow \mathbb{R}^k, \quad \theta \mapsto \partial_\theta \ell_x(\theta) \quad \text{Score-Funktion.}$$

Ein Schätzer $\hat{\theta}$ heißt **Maximum-Likelihood Schätzer für θ** , falls er fast-überall Sampleweise die Likelihood-Funktion oder äquivalent die Log-Likelihood-Funktion maximiert, das heißt

$$L_x(\hat{\theta}(x)) = \sup_{\theta \in \Theta} L_x(\theta) \quad [\mu] \quad \text{beziehungsweise}$$

$$\ell_x(\hat{\theta}(x)) = \sup_{\theta \in \Theta} \ell_x(\theta) \quad [\mu].$$

Für eine Funktionalauswertung $\gamma(\theta)$ heißt dann $\hat{\gamma} := \gamma(\hat{\theta})$ **Maximum-Likelihood Schätzer für $\gamma(\theta)$** .

Bevor wir auf das Beispiel 1.1.39 zurückkommen, wollen wir noch ergänzen:

Bemerkung 1.1.41.

1. Existiert die Score-Funktion, so ist eine notwendige Bedingung für das Vorliegen eines Maximums die Beziehung $\dot{\ell}_x(\hat{\theta}(x)) = 0$. In simplen Fällen kann also über regulär-analytische Maximierung ein Maximum-Likelihood Schätzer gefunden werden.
2. Weder Eindeutigkeit noch Existenz sind gesichert. Ist Letzteres fragwürdig, zieht man sich im Falle einer stetigen Log-Likelihood-Funktion oft auf kompakte Teilmengen $\Theta_K \subset \subset \Theta$ zurück, sofern Θ mit einer Topologie ausgestattet werden kann; siehe [Van der Vaart, Abschnitt 5.2.1].
3. Maximum-Likelihood Schätzer sind Spezialfälle sogenannter *M Schätzer*, welche dadurch charakterisiert sind, dass solche Schätzer eine Zielfunktion M_N maximieren.

Nun kommen wir auf das Weibull-Lokationsmodell aus Beispiel 1.1.39 zurück.

Beispiel 1.1.42 (Fortsetzung Weibull-Lokationsmodell).

Im Modell aus Beispiel 1.1.39 ist $\hat{\mu} := x_{N:N}$ tatsächlich der eindeutige Maximum-Likelihood Schätzer für μ .

Beweis. Aus dem Modell bestimmt sich die Log-Likelihood Funktion für ein $x = (x_1, \dots, x_N)$ als

$$\ell_x(\mu) = \begin{cases} -N \cdot \log \sigma - \frac{N\mu}{\sigma} + \sum_{i=1}^N \frac{x_i}{\sigma} & , x_{N:N} \leq \mu \\ -\infty & , x > \mu \end{cases}.$$

Damit ist im Bereich $x_{N:N} \leq \mu$ die Log-Likelihood Funktion antiton in μ , aber ab $\mu > x_{N:N}$ minimal, sodass der eindeutige Maximum-Likelihood Schätzer durch $\hat{\mu} := x_{N:N}$ gegeben ist. \square

In der Stochastik sind Grenzwertsätze von hohem Nutzen, da sie es erlauben über eine Klasse von Verteilungen gleichmäßig Grenzverteilungen zu bestimmen, ohne dass

die eventuelle komplizierte Struktur der zugrundeliegenden Verteilungen in hohem Maße berücksichtigt werden müssen. Bisher haben wir Statistik im Rahmen einer festen Samplegröße beziehungsweise Dimension des Ereignisraumes \mathcal{X} betrieben, sodass Grenzwertsätze nicht ausgenutzt werden konnten. Die *asymptotische Statistik* ermöglicht die Betrachtung von Modellfolgen und daher die Nutzung von Asymptotiken. Der Rahmen dafür ist das Vorliegen von einer Folge von Modellen $\mathcal{E}_N = (\mathcal{X}^n, \mathcal{B}^{\otimes n}, \mathcal{P}_n)$, wobei $(\mathcal{X}, \mathcal{B})$ ein zugrundeliegender Maßraum und $\mathcal{P}_n = \{P_\theta^{(n)} : \theta \in \Theta\}$ eine Folge von Wahrscheinlichkeitsmaßfamilien ist, die von Parametern aus einem Parameterraum Θ abhängen. Der häufigste Fall ist die unabhängige Wiederholung von Zufallsexperimenten, was dem Produktmaß $P_\theta^{(n)} = P_\theta^{\otimes n}$ für eine feste Familie von Verteilungen $\{P_\theta : \theta \in \Theta\}$ auf $(\mathcal{X}, \mathcal{B})$ entspricht. Die Tatsache, dass in dieser Situation Modellfolgen betrachtet werden bewirkt, dass nun Schätzerfolgen und nicht mehr ein einzelner Schätzer im Mittelpunkt stehen.

Es lassen sich damit asymptotische Gütekriterien formulieren.

Definition 1.1.43 (Asymptotische Gütekriterien für Schätzer).

Sei eine Schätzerfolge $(\hat{\gamma}_N)_N$ gegeben, so heißt $\hat{\gamma}_N$

1. **asymptotisch erwartungstreu**, falls für alle $\theta \in \Theta$: $E_\theta[\hat{\gamma}_N - \gamma(\theta)] \rightarrow 0, N \rightarrow \infty$,
2. **(stark) konsistent**, falls $\hat{\gamma}_N \xrightarrow{(f.s)P_\theta} \gamma(\theta)$ für alle $\theta \in \Theta$,
3. **asymptotisch normal**, falls $\Gamma \subset \mathbb{R}^k$ und für alle $\theta \in \Theta$ normalisierende Folgen $a_N = a_N(\theta) > 0, b_N = b_N(\theta) \in \mathbb{R}^k$ derart existieren, dass

$$\frac{\hat{\gamma}_N - b_N}{a_N} \xrightarrow{d_\theta} \mathcal{N}_k(0, \Sigma)$$

für eine positiv semidefinite Kovarianzmatrix $\Sigma := \Sigma(\theta) \in \mathbb{R}^{d \times d} \setminus \{0\}$ gilt.

Asymptotische Normalität wird meist mittels Grenzwertsätze und der Delta-Methode, siehe [A.2.1](#), gezeigt.

Beispiel 1.1.44 (Geometrisches Modell).

Wir betrachten das Produktmodell einer $\text{Geom}(\theta)$ -Verteilung für $\theta \in]0, 1[$, das heißt das aus unabhängig und identisch nach $\text{Geom}(\theta)$ verteilten Zufallsvariablen X_1, \dots, X_N hervorgehende Modell.

Dann ist der eindeutige Maximum-Likelihood Schätzer durch $\hat{\theta}_N := (\bar{x}_N + 1)^{-1}$ gegeben. Außerdem ist er stark konsistent, asymptotisch erwartungstreu und normal. Tatsächlich gilt sogar für alle $k \geq 1$ die Momentenkonvergenz $E_\theta[\hat{\theta}_N^k] \xrightarrow{n} \theta^k$.

Beweis. ML Schätzer: Die Log-Likelihood und Score-Funktion sind durch $\ell_x(\theta) = N \cdot \log \theta + \log(1 - \theta) \sum_{i=1}^N x_i$ beziehungsweise $\dot{\ell}_x(\theta) = N/\theta - \sum_{i=1}^N x_i/(1 - \theta)$ gegeben. Das Nullsetzen der Score-Funktion liefert die eindeutige Nullstelle

$\hat{\theta}_N = (\bar{x}_N + 1)^{-1}$. Wegen $\lim_{\theta \downarrow 0} \ell_x(\theta) = -\infty = \lim_{\theta \uparrow 1} \ell_x(\theta)$ folgt, dass dies das eindeutige globale Maximum ist.

Starke Konsistenz: Aus dem starken Gesetz der großen Zahlen und dem Satz von der stetigen Abbildung folgt die Konvergenz

$$(\bar{X}_N + 1)^{-1} \xrightarrow{f.s.} (\mathbb{E}_\theta[X_1] + 1)^{-1} = \left(\frac{1-\theta}{\theta} + 1 \right)^{-1} = \theta,$$

sodass die starke Konsistenz gezeigt ist.

Asymptotische Erwartungstreue: Zeigen wir die gleichgradige Integrierbarkeit der Familie $(\hat{\theta}_N)_{N \in \mathbb{N}}$ für festes $\theta \in \Theta$, so folgt aus der Konsistenz des Schätzers die L^1 -Konvergenz des Schätzers gegen θ .⁴ Es gilt $\hat{\theta}_N \leq 1$ P_θ -fast-sicher, sodass für $M \geq 2$ der Indikator $\mathbb{1}\{|\hat{\theta}_N| \geq M\}$ fast-sicher zur konstanten Null-Abbildung degeneriert. Insbesondere gilt daher

$$\lim_{M \rightarrow \infty} \sup_{N \in \mathbb{N}} \mathbb{E}_\theta \left[|\hat{\theta}_N| \cdot \mathbb{1}\{|\hat{\theta}_N| \geq M\} \right] = 0$$

und damit die zu zeigende gleichgradige Integrierbarkeit.

Asymptotische Normalität: Für die asymptotische Normalität betrachten wir die Abbildung

$$g:]0, 1[\rightarrow]1/2, 1[, \quad x \mapsto \frac{1}{1+x},$$

mit Ableitung $g'(x) = -(1+x)^{-2}$. Aus dem zentralen Grenzwertsatz folgt

$$\sqrt{N} \left(\bar{X}_N - \frac{1-\theta}{\theta} \right) = \sqrt{N} (\bar{X}_N - \mathbb{E}_\theta[X_1]) \xrightarrow{d_\theta} \mathcal{N} \left(0, \frac{1-\theta}{\theta^2} \right).$$

Wegen $g(\bar{X}_N) = \hat{\theta}_N$, $g((1-\theta)/\theta) = \theta$ und $g'((1-\theta)/\theta) = -\theta^2$ folgt aus der Delta-Methode A.2.1 somit

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d_\theta} -\theta^2 \cdot \mathcal{N} \left(0, \frac{1-\theta}{\theta^2} \right) = \mathcal{N}(0, \theta \cdot (1-\theta)),$$

also die asymptotische Normalität. □

Für genügend *reguläre* Modelle besitzt der Maximum-Likelihood Schätzer wie im vorangegangenen Beispiel wünschenswerte asymptotische Güte, siehe zum Beispiel [Cramér, Abschnitt 33.3]. Das Weibull-Lokationsmodell aus Beispiel 1.1.39 und 1.1.42 besitzt diese Regularität nicht, weil dort der Träger der Verteilungen von dem unbekannten Parameter μ abhängt. Da jedoch die Regularitätsbedingungen nicht not-

⁴ Alternativ nutzt man, dass wegen $\hat{\theta}_N \xrightarrow{f.s.} \theta$ insbesondere Konvergenz des Schätzers in Verteilung vorliegt. Nach [Van der Vaart, 2.21 Example] genügt es daher die Beschränktheit $\limsup_N \mathbb{E}_\theta[|\hat{\theta}_N^p|] < \infty$ für ein $p > 1$ zu zeigen, da dann die Momentenkonvergenz für alle $k < p$ folgt. Wegen der in N und p gleichmäßigen Ungleichung $\hat{\theta}_N^p \leq 1$ folgt daher also sogar die Konvergenz aller Momente.

wendig sondern lediglich hinreichend für asymptotische Normalität sind, lässt sich daraus nicht direkt eine andere Asymptotik schließen. Wir werden jedoch sehen, dass hier keine asymptotische Normalität vorliegt.

Beispiel 1.1.45 (Fortsetzung Weibull-Lokationsmodell).

Im Modell aus Beispiel 1.1.39 gilt für den Maximum-Likelihood Schätzer $\hat{\mu}_N = x_{N:N}$, dass er asymptotisch erwartungstreu, stark konsistent ist und

$$\frac{\hat{\mu}_N - \mu}{1/N} \xrightarrow{d_\mu} \text{Wei}(0, \sigma, 1) \quad (\mu \in \mathbb{R}).$$

Insbesondere ist $\hat{\mu}_N$ nicht asymptotisch normal.

Beweis. Asymptotische Erwartungstreue: Aus Beispiel 1.1.39 entnehmen wir den exakten Bias als $\text{Bias}_\mu(\hat{\mu}) = -\sigma/N$, sodass die Eigenschaft instantan folgt.

Starke Konsistenz: Der Schätzer $\hat{\mu}$ ist ein Maximum über N -viele unabhängig und identisch verteilte Zufallsvariablen, sodass er aufgrund von Lemma A.2.3 gegen den rechten Randpunkt der Zufallsvariablen, der gerade μ entspricht, P_μ fast-sicher konvergiert.

Asymptotische Verteilung: Der Beweis von Beispiel 1.1.39 lässt uns erkennen, dass $\hat{\mu}_N$ einer $\text{Wei}(\mu, \sigma/N, 1)$ -Verteilung folgt. Wir rechnen

$$\begin{aligned} P_\mu \left\{ \frac{\hat{\mu}_N - \mu}{1/N} \leq x \right\} &= \begin{cases} \exp\left(\frac{x/N + \mu - \mu}{\sigma/N}\right) & , x/N + \mu < \mu \\ 1 & , x/N + \mu \geq \mu \end{cases} = \begin{cases} \exp\left(\frac{x}{\sigma}\right) & , x < 0 \\ 1 & , x \geq 0 \end{cases} \\ &= \text{Wei}(0, \sigma, 1), \end{aligned}$$

weswegen die Verteilungsgleichheit sogar subasymptotisch gilt.

Asymptotische Normalität ist nicht möglich, da andernfalls aus dem Convergence of Types Theorem (Lemma 1.1.32) folgen würde, dass Weibull- und Normalverteilung vom selben Typ wären; dem ist jedoch nicht so. Alternativ kann man die nicht-Normalität direkt aus dem Fisher-Tippett 1.2.1 und dem Convergence of Types Theorem folgern, da $\hat{\mu}_N$ als ein Maximum über eine unabhängig und identisch verteilte Folge von Zufallsvariablen gegeben ist. \square

Dieses Resultat lässt sich bis auf einschließlich der Grenze $\alpha = 1$ fortsetzen. Entscheidend ist dabei, dass in diesen Fällen der Faktor $\alpha - 1 \leq 0$ ist, was das Vorzeichen der Singularität des Logarithmus' in der Null umkehrt. Diese Erkenntnisse sind nicht offensichtlich, da für Weibull-Modelle mit Formparameter $\alpha \in]2, \infty[$ asymptotische Normalität des Maximum-Likelihood Schätzer gegeben ist, was in Subkapitel 3.2 genauer besprochen wird. Ein zentraler Begriff, um Maximum-Likelihood Schätzer nicht-regulärer Modelle zu behandeln, ist die *Differenzierbarkeit im quadratischen Mittel*. So gilt unter Annahme eines verallgemeinerten Schrankensatzes, Konsistenz und Invertierbarkeit der Fisher-Information, dass lokale asymptotische Normalität vorliegt, siehe [Van der Vaart, Theorem 5.39]. Dieses Resultat lässt sich jedoch nicht auf das

Weibull-Lokationsmodell anwenden, da für jede Umgebung eines $\mu_0 \in \mathbb{R}$ beliebig viele Parameter μ innerhalb der Umgebung existieren, sodass $x \in \mathbb{R}^N$ mit $p_{\mu_0}(x) > 0$ und $p_\mu(x) = 0$ existieren. Damit existiert jedoch die Log-Likelihood Funktion nicht, was die Theorie unanwendbar macht. Es gibt jedoch Lösungen dieser Problematiken: Für eine Behandlung des GEV-Modells siehe etwa [Bücher und Segers(a), Gleichung (2.2)] für eine derartige Transformation der Log-Likelihood Funktion, dass die Singularität $\log(0)$ nicht mehr auftritt. Dabei wird jedoch unter anderem fortgeschrittene *Empirische Prozess-Theorie* genutzt, sodass wir dies an dieser Stelle nicht weiter untersuchen.

1.2 Extremwerttheoretische Grundlagen

Die Extremwerttheorie ist der Teil der Stochastik, in welchem Gesetzmäßigkeiten von Maxima oder Minima von Familien von Zufallsvariablen erfasst werden. Laut [David und Edwards] wurde durch zwei Artikel von Bortkiewicz, [von Bortkiewicz(a)] und [von Bortkiewicz(b)], die fokussiertere Entwicklung der Extremwerttheorie durch Arbeit seitens von Mises [von Mises] initiiert, welche erste Grundsteine für das berühmte *Fisher-Tippet Theorem*, [Fisher und Tippet] legte.

Die klassische Extremwerttheorie betrachtet Dreiecksschemata Δ von zeilenweisen unabhängig, im gesamten Schema identisch verteilten Zufallsvariablen $X_{i,n}$ und bezeichnet mit M_n die Zufallsvariable $\max(X_{n,1}, \dots, X_{n,n})$.

Aufgrund des Fakts $\min(r_i) = -\max(-r_i)$ für reelle Zahlen r_i betrachtet man ohne Einschränkung nur Maxima. Tatsächlich lässt sich für solche Schemata ein fast-sicheres Konvergenzverhalten gegen den rechten Randpunkt der Verteilungen zeigen, siehe A.2.3, was die Frage nach geeigneten Skalierungsfolgen stellt, sodass eine nicht-triviale Verteilungskonvergenz stattfindet. Das Fisher-Tippet-Theorem präzisiert nun die erstaunliche Tatsache, dass im wesentlichen nur drei verschiedenen Grenzverteilungen für Maxima von Dreiecksschemata in Frage kommen.

Theorem 1.2.1 (Fisher-Tippet Theorem).

Sei Δ ein Dreiecksschema wie oben. Existieren normalisierende Folgen $a_n > 0, b_n \in \mathbb{R}$ derart, dass die Verteilungskonvergenz

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G \quad (1.2)$$

für eine nicht-degenerierte Zufallsvariable G gilt, so muss G schon einer generalisierten Extremwertverteilung $\text{GEV}(\mu, \sigma, \xi)$ entsprechen, wobei μ ein Lokations-, $\sigma > 0$ ein Skalen- und $\xi > 0$ ein Formparameter ist. Letzterer bestimmt, welcher speziellen Verteilungsfamilie G angehört, siehe Abbildung 1.

Die Verteilungsklasse aus dem vorigen Resultat ist wie folgt erklärt.

Definition 1.2.2 (Generalisierte Extremwertverteilung).

Für Parameter $\mu, \xi \in \mathbb{R}, \sigma > 0$ ist die **generalisierte Extremwertverteilung** durch die

Verteilungsfunktion

$$\text{GEV}(\mu, \sigma, \xi)(x) := \begin{cases} 0 & , \xi > 0 \text{ und } x \leq -\sigma/\xi + \mu \\ \exp\left(-\left(1 + \xi \cdot \frac{x-\mu}{\sigma}\right)^{-1/\xi}\right) & , 1 + \xi \cdot (x-\mu)/\sigma > 0 \text{ und } \xi \neq 0 \\ \exp\left(-\frac{x-\mu}{\sigma}\right) & , \xi = 0 \\ 1 & , \xi < 0 \text{ und } x \geq -\sigma/\xi + \mu \end{cases}$$

gegeben. Wir werden sie im Laufe der Arbeit durch $\text{GEV}(\mu, \sigma, \xi)$ abkürzen.

Die Verteilungsfamilie ist eine Parametrisierung von drei Verteilungsfamilien, namentlich der **Gumbel-Verteilung** ($\xi = 0$), der **Fréchet-Verteilung** ($\xi > 0$) und der (**negativen**) **Weibull-Verteilung** ($\xi < 0$).

Die wichtigen Momente sind für eine $\text{GEV}(\mu, \sigma, \xi)$ -verteilte Zufallsvariable G durch

$$\begin{aligned} \mathbb{E}[G] &= \begin{cases} \mu + \sigma \cdot \frac{\Gamma(1-\xi)-1}{\xi} & , \xi \neq 0, \xi < 1 \\ \mu + \sigma\xi & , \xi = 0 \\ \infty & , \xi \geq 1 \end{cases} \\ \text{Var}(G) &= \begin{cases} \sigma^2 \cdot \frac{\Gamma(1-2\xi)-\Gamma^2(1-\xi)}{\xi^2} & , \xi \neq 0, \xi < 1 \\ \frac{\sigma^2\pi^2}{6} & , \xi = 0 \\ \infty & , \xi \geq 1/2 \end{cases} \end{aligned} \quad (1.3)$$

gegeben.

In dieser Arbeit werden wir uns lediglich auf die Weibull-Verteilung beziehen, sodass wir an dieser Stelle lediglich die klassische Parametrisierung dieser angeben.

Diese ist für Parameter $b \in \mathbb{R}, a, \alpha > 0$ durch die Verteilungsfunktion

$$\text{Wei}(b, a, \alpha) := \begin{cases} \exp\left(-\left(-\frac{x-b}{a}\right)^\alpha\right) & , x < b \\ 1 & , x \geq b \end{cases}$$

definiert und wir werden sie dementsprechend mit $\text{Wei}(b, a, \alpha)$ abkürzen.

In Abbildung 1 sind Dichten der $\text{GEV}(\mu, \sigma, \xi)$ -Verteilung für verschiedene Werte von ξ , und im Falle der Gumbel-Verteilung von σ , angegeben. Man erkennt, dass die Gumbel-Verteilung die ganze reelle Achse als Träger hat, wohingegen er im Fréchet-Fall $]\mu, \infty[$ und im Weibull-Fall $]-\infty, \mu[$ ist. Insbesondere ist interessant, dass ab $\xi \leq -1$ die Weibull-Dichte nicht mehr stetig in den rechten Randpunkt μ hinein verläuft.

Es ist nützlich eine konkrete Umrechnungsformel für die beiden Parametrisierungen der Weibull-Verteilung anzugeben, da sie in dieser Arbeit noch von Nutzen sein wird.

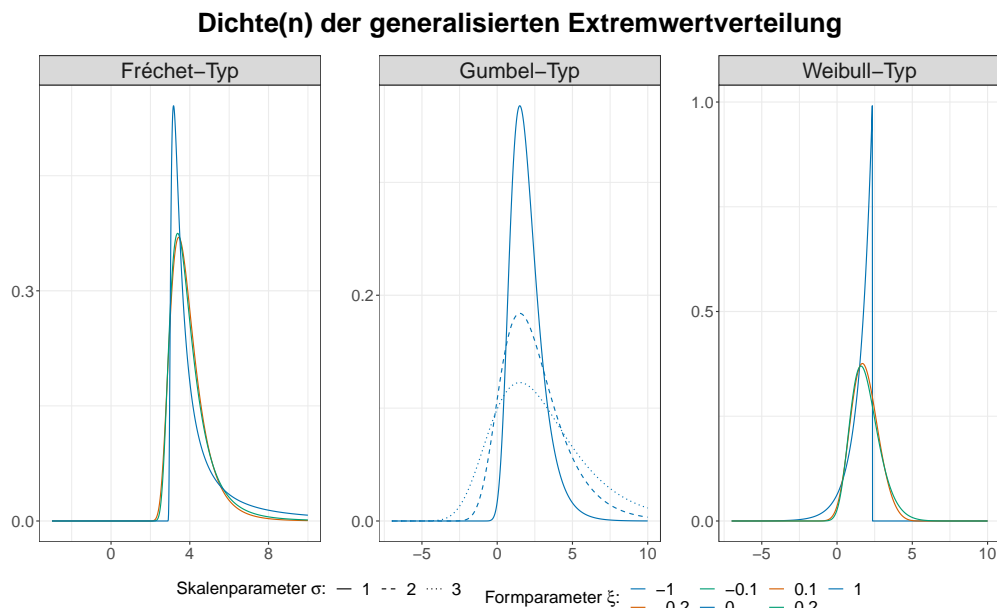


Abbildung 1: Dichte der GEV (μ, σ, ξ) – Verteilung für verschiedene Formparameter und im Fall der Gumbel-Verteilung verschiedene Skalenparameter. Der Lokationsparameter ist als Null fixiert.

Bemerkung 1.2.3 (Umrechnung Weibull-Parametrisierung).

Seien $a, \sigma, \alpha > 0$, $b, \mu \in \mathbb{R}$ und $\xi < 0$. Dann gelten

$$\text{GEV}(\mu, \sigma, \xi) = \text{Wei}(\mu - \sigma/\xi, -\sigma/\xi, -1/\xi)$$

$$\text{Wei}(b, a, \alpha) = \text{GEV}(b - a, a/\alpha, -1/\alpha).$$

Beweis. Wir zeigen nur die erste Umrechnungsformel, da die zweite ähnlich zu beweisen ist. Zunächst renotieren wir

$$\exp\left(-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-1/\xi}\right) = \exp\left(-\left(-\frac{x - (\mu - \sigma/\xi)}{\sigma/\xi}\right)^{-1/\xi}\right),$$

sodass wegen der Äquivalenz von $1 + \xi(x - \mu)/\sigma > 0$ und $x < \mu - \sigma/\xi$ die erste Gleichung folgt. \square

Das Fisher-Tippet Theorem nimmt also für das Maximum eine ähnliche Rolle ein wie der zentrale Grenzwertsatz von Lindeberg-Feller für die Summe von Zufallsvariablen. Angebracht erscheint jedoch der Vergleich mit dem Begriff im weiteren Sinne $(\alpha-)$ stabiler Verteilungen, welche unter verschobener und reskalierten Faltung abgeschlossen sind. So existiert ein Resultat wonach die Qualifikation einer Verteilung als Grenzwert einer normalisierten Summe von unabhängig und identisch verteilten Zufallsvariablen gleichbedeutend zu der Stabilität der Verteilung ist, [Feller, S. 170 ff.].

Das für die Maximumoperation analoge Resultat lautet nun, dass eine Verteilung genau dann *max-stabil* ist, wenn eine reskalierte Verteilungskonvergenz der Maximumfolge stattfindet, [Leadbetter, Lindgren und Rootzén, Theorem 1.3.1]. Über letztere Charakterisierung lässt sich auch das Fisher-Tippet Theorem beweisen, indem genau diese max-stabilen Verteilungen charakterisiert werden. Im Gegensatz zu dem zentralen Grenzwertsatz enthält das Fisher-Tippet Theorem keine hinreichende Bedingungen dafür, dass eine Verteilung F im **Max-Anziehungsbereich** von $\text{GEV}(\mu, \sigma, \xi)$ liegt. Das heißt es existieren normalisierende Folgen $a_n > 0, b_n \in \mathbb{R}$, sodass die Konvergenz in (1.2) gegen eine generalisierte Extremwertverteilung gilt, liegt. Dies ist jedoch für alle drei Typen untersucht und charakterisiert, wobei dabei die Begriffe der *regulären* für die Fréchet, Weibull Anziehungsbereiche und *Pi-* beziehungsweise *Gamma-Variation* für den Gumbel Anziehungsbereich entscheidend sind, siehe [Resnick, S. 54 ff., 59 ff., 38 ff.].

Wie in den statistischen Grundlagen 1.1 angesprochen, ist die Identifizierbarkeit eines statistischen Modells grundlegende Voraussetzung, um Inferenz zu betreiben. Wir weisen sie für das GEV-Modell nur im Weibull-Fall nach. Die anderen Fälle lassen sich ähnlich behandeln.

Lemma 1.2.4 (Identifizierbarkeit des GEV-Modells).

Sei \mathcal{P} das aus X_1, \dots, X_N unabhängig und identisch nach $\text{GEV}(\mu, \sigma, \xi)$ –verteilten Zufallsvariablen hervorgehende Produktmodell, wobei $\mu \in \mathbb{R}, \sigma, \xi > 0$. Dann ist \mathcal{P} identifizierbar.

Beweis. **Nur für den Fall $\xi < 0$:** Wir nutzen Lemma 1.1.31, Korollar 1.1.33 und Lemma 1.1.35, da $\text{GEV}(b, a, \xi)$ die Lokations- und Skalenerweiterung von $\text{GEV}(0, 1, \xi)$ ist.

Seien $a > 0, \xi_1, \xi_2 < 0, b \in \mathbb{R}$ mit $\text{GEV}(0, 1, \xi_1) = \text{GEV}(b, a, \xi_2)$ gegeben. Es gilt daher auf dem Träger beider Verteilungen

$$\exp\left(-(1 + \xi_1 x)^{-1/\xi_1}\right) = \exp\left(-\left(1 + \xi_2 \frac{x - b}{a}\right)^{-1/\xi_2}\right),$$

also insbesondere $1 + \xi_1 x = (1 + \xi_2(x - b)/a)^{\xi_1/\xi_2}$. Differenzieren wir beide Seiten nach x , so folgt $\xi_1 = \xi_1/a \cdot (1 + \xi_2(x - b)/a)^{\xi_1/\xi_2 - 1}$. Damit die rechte Seite konstant in x sein kann, muss also der Exponent Null sein, was $\xi_1 = \xi_2$ impliziert. Es folgt also $\text{GEV}(0, 1, \xi_1) = \text{GEV}(0, 1, \xi_2)$, sodass nach Korollar 1.1.33 die Typenfreiheit folgt.

Die Identifizierbarkeit von $\text{GEV}(0, 1, \xi)$ folgt aus dem gerade geführten Beweis. \square

In dieser Arbeit werden wir jedoch den Rahmen der klassischen Extremwerttheorie verlassen und Zeitreihen betrachten, die nicht (*strikt*) *stationär* sind.

Definition 1.2.5 ((reellwertige) Zeitreihen).

Für eine nicht-leere Indexmenge T , welche man sich als Zeit vorstelle, heißt ein stochastischer Prozess

$$(X_t)_{t \in T}: (\Omega, \mathcal{A}) \rightarrow \left(\prod_{t \in T} \mathbb{R}^d, \bigotimes_{t \in T} \mathcal{B}(\mathbb{R}^d) \right), \quad \omega \mapsto (X_t(\omega))_{t \in T}$$

Zeitreihe.

Eine Zeitreihe heißt **(strikt) stationär**, falls für alle $t_1, \dots, t_n \in T, n \in \mathbb{N}$ und $\tau \in \mathbb{R}$ mit $t_i + \tau \in T$ die Maßgleichheit

$$\mathbb{P}(X_{t_1}, \dots, X_{t_n}) = \mathbb{P}(X_{t_1+\tau}, \dots, X_{t_n+\tau})$$

gilt. Dies impliziert insbesondere, dass die Gleichheit $X_t = X_s$ in Verteilung für alle $s, t \in T$ besteht.

So lässt sich die klassische Extremwerttheorie in den Rahmen unabhängiger und stationärer Zeitreihen einbetten. Verlassen wir jedoch diese Umgebung der Stationarität, gilt im Allgemeinen nicht einmal die Konvergenz des Maximumsprozesses gegen den rechten Randpunkt, siehe A.2.3, da es keinen eindeutigen rechten Randpunkt geben muss.

Beispiel 1.2.6.

Sei $(X_n)_n$ eine unabhängige Zeitreihe. Es seien für ein festes $N \in \mathbb{N}$ die ersten N Zufallsvariablen X_1, \dots, X_N nicht-degeneriert. Zusätzlich besitzen sie denselben rechten Randpunkt x_1^* und die darauffolgenden Zufallsvariablen haben die rechten Randpunkte x_i^* , wobei $\sup_{i>N} x_i^* < x_1^*$.

Dann konvergiert die Zufallsvariable $M_n := \max(X_1, \dots, X_n)$ nicht stochastisch und damit insbesondere nicht fast-sicher gegen ein $c \in \mathbb{R} \cup \{+\infty\}$.

Beweis. Wir zeigen dies durch eine Fallunterscheidung, wofür $c \in \mathbb{R}$ der Kandidat und $\varepsilon > 0$ fest sei.

Fall $c < x_1^*$: In diesem Fall gilt

$$\begin{aligned} \mathbb{P}\{|M_n - c| \leq \varepsilon\} &= 1 - \mathbb{P}\{|M_n - c| > \varepsilon\} \\ &\leq 1 - \mathbb{P}\{X_1 - c > \varepsilon\} = 1 - \mathbb{P}\{X_1 > \varepsilon + c\} < 1, \end{aligned}$$

nach der Definition des rechten Randpunktes für ε nahe Null. Da diese Ungleichung gleichmäßig in n ist, folgt die Behauptung für diesen Fall.

Fall $c > x_1^*, x_1^* < \infty$:

$$\begin{aligned} \mathbb{P}\{|M_n - c| \leq \varepsilon\} &= \mathbb{P}\{c - M_n \leq \varepsilon\} \\ &= 1 - \mathbb{P}\{M_n < c - \varepsilon\} \leq 1 - \mathbb{P}\{M_n \leq x_1^*\} = 0, \end{aligned}$$

für ε nahe Null, wobei die letzte Gleichheit aus der Voraussetzung $\sup_{i \in \mathbb{N}} x_i^* = x_1^*$ folgt.

Fall $c = x_1^*, x_1^* < \infty$:

$$\begin{aligned} \mathbb{P}\{|M_n - x_1^*| \leq \varepsilon\} &= \mathbb{P}\{x_1^* - M_n \leq \varepsilon\} = \mathbb{P}\{M_n \geq x_1^* - \varepsilon\} \\ &= 1 - \mathbb{P}\{M_n < x_1^* - \varepsilon\} = 1 - \prod_{i=1}^N \mathbb{P}\{X_i < x_1^* - \varepsilon\} < 1 \end{aligned}$$

gleichmäßig in n für kleine ε , wobei die vorletzte Gleichheit gilt, aufgrund der Voraussetzung $\sup_{i>N} x_i^* < x_1^*$ und die obere Schranke aus der Annahme der nicht-Degeneriertheit der X_1, \dots, X_N folgt.

Fall $c = x_1^*, x_1^* = \infty$: Wir zeigen, dass für große $K \in \mathbb{N}$ der Grenzwert $\lim_n \mathbb{P}\{M_n \geq K\} < 1$ gilt, sodass M_n nicht in Wahrscheinlichkeit gegen unendlich konvergiert. Sei dafür ohne Einschränkung $K > \sup_{i>N} x_i^*$:

$$\mathbb{P}\{M_n \geq K\} = \mathbb{P}\{M_N \geq K\} = 1 - \mathbb{P}\{M_N < K\} < 1$$

für K derartig groß, dass $\prod_{i \leq N} \mathbb{P}\{X_i > K\} > 0$, was aufgrund der Fallannahme $x_1^* = \infty$ garantiert werden kann. \square

Das Beispiel zeigt, wie der *Grad* von verschiedenen Randpunkten innerhalb der Zeitreihe großen Einfluss auf das Konvergenzverhalten der maximalen Zeitreihe nimmt. Selbst wenn jedoch stochastische Konvergenz vorliegt, folgt nicht ohne Weiteres die fast-sichere Konvergenz, was wir im nächsten Kapitel bemerken werden.

2 Asymptotische Resultate über die Verteilung Maxima von nicht-stationären Dreiecksschemata

In diesem Kapitel werden wir zwei Grenzwertsätze über die asymptotische Verteilung von reskalierten Maxima nicht-stationärer unabhängiger Zeitreihen vorstellen und beweisen. Diese entstammen dem Artikel [Stein]. Motiviert sind diese Resultate von jährlichen Maximaltemperaturuntersuchungen. Dabei werden drei Annahmen an die Temperaturen gestellt, welche die stochastische Behandlung ebenjener ermöglichen.

2.1 Rahmenbedingungen

Zunächst werden Eigenschaften an die Temperaturen gestellt, welche eine asymptotische Untersuchung ermöglichen.

Annahme 2.1.1 (Temperatureigenschaften).

1. Jeder Tag des Jahres besitzt eine endliche Temperaturschranke.
2. Die kleinste obere Schranke verändert sich glatt während eines Jahres.
3. Im Laufe des Jahres nimmt die kleinste obere Temperaturschranke ein eindeutiges globales Maximum an.

Die ersten beiden Bedingungen sind leicht zu motivieren, wohingegen die dritte Bedingung innerhalb mancher Regionen nicht erfüllt sein mag, aber dennoch für den Großteil wohlmotiviert ist.

Damit die Glattheitsbedingung aus 2.1.1 erfüllt werden kann, müssen wir zumindest asymptotisch eine Dichtheit der Beobachtungspunkte in \mathbb{R} garantieren, da sonst der Begriff der Differenzierbarkeit nicht wohldefiniert wäre. Sinngebend ist daher eine Modellierung mittels eines Dreiecksschemas von Zufallsvariablen

$$X_{i,n} : \Omega \rightarrow \mathbb{R}, \omega \mapsto X_{i,n}(\omega) \quad 1 \leq i \leq n, \quad n \in \mathbb{N},$$

das heißt

$$\Delta = \{X_{i,n} : 1 \leq i \leq n, n \in \mathbb{N}\}.$$

Man beachte, hier werden noch keine Forderungen an die Zeilen des Schemas gestellt, sodass die Zufallsvariablen im Allgemeinen nicht identisch verteilt und abhängig sein können. Erweiterungen auf schwache Formen der Abhängigkeit werden in Bemerkung 2.2.15 angesprochen.

Ohne Einschränkung nehmen wir wegen 2.1.1 an, dass $X_{i,n} \in [0, 1]$ gilt, da wir sonst mittels minimaler und maximaler Temperatur reskalieren würden. Die Asymptotik des Dreiecksschemas Δ lässt sich für große n als approximative Verteilung von Jahrestemperaturmaxima sehen, wobei die Anzahl an Messungen als n interpretiert wird. Dies lässt sich in den Zeitreihenrahmen wie folgt einbetten: Man reparametrisiere

$\{(i, n) : 1 \leq i \leq n, n \in \mathbb{N}\}$ mittels $(i, n) \mapsto i/n$ zu der total geordneten Zeitpunktmenge $T_n = \{i/n : 1 \leq i \leq n, n \in \mathbb{N}\}$. Für die Grenzwertbildung in n erhält man somit $T := \bigcup_{n \in \mathbb{N}} \{i/n : 1 \leq i \leq n, n \in \mathbb{N}\} =]0, 1]$ und den linken Endpunkt werden wir mittels Zeitstetigkeitsannahmen einschließen können.

Definition 2.1.2 (Randpunktsfunktion).

Die Abbildung

$$u :]0, 1[\rightarrow [0, 1], \quad \frac{i}{n} \mapsto \sup \{x \in [0, 1] : \mathbb{P}\{X_{i,n} \leq x\} < 1\}$$

heißt **Randpunktsfunktion**. Dabei sei $1 \leq i \leq n$, $n \in \mathbb{N}$ und wir fordern die Stetigkeit von u , wobei wir für irrationale Punkte u mittels Approximation stetig fortsetzen. Man beachte, dass wir hier stillschweigend die Irreduzibilität von i/n fordern, da sonst anschaulich selbe Zeitpunkte unterschiedliche Randpunkte hätten beziehungsweise u nicht wohldefiniert wäre.

Die Randpunktsfunktion u setzt Annahme 2.1.1 partiell um. In den zwei Asymptotiken setzen weitere Forderungen an u die Annahmen vollständig um.

2.2 Erste Asymptotik zeilenweiser Unabhängigkeit

Folgende Forderung setzt auf eine Art die Annahme 2.1.1 um.

Bedingung 2.2.1 (Bedingung an u in der ersten Asymptotik).

Die Randpunktsfunktion u habe eine stetige zweite Ableitung u'' und nehme ein eindeutiges globales Maximum $u_0 \in [0, 1]$ im Punkt $y_0 \in]0, 1[$ an, wobei zusätzlich $u''(y_0) < 0$ gelte.

Das Erzielen von asymptotischen Resultaten erfordert noch Annahmen an die Verteilungen der $X_{i,n}$; im Engeren: wie sich die Zufallsvariablen nahe ihrer Randpunkte verhalten.

Bedingung 2.2.2.

Es existieren positive Funktionen $S :]0, \infty[\times [0, 1] \rightarrow \mathbb{R}^+$, $\alpha : [0, 1] \rightarrow \mathbb{R}^+$ derart, dass

$$\mathbb{P}\{X_{i,n} > u(i/n) - t\} = S(t, i/n) \cdot t^{\alpha(i/n)}, \quad t > 0$$

für kleine t erfüllt ist.

Bedingung 2.2.3.

Es existieren positive Konstanten c_1, c_2 dergestalt, dass für y nahe y_0

$$|\alpha(y) - \alpha(y_0)| \leq c_1 \cdot |y - y_0|^{c_2} \tag{2.1}$$

gilt.

Diese Bedingung lässt sich als eine Verallgemeinerung der lokalen Hölderstetigkeit in y_0 sehen, da letztere in den meisten Fällen lediglich Konstanten $c_2 \in]0, 1[$ zulässt, was in folgender Tatsache begründet liegt.

Bemerkung 2.2.4.

Funktionen $\alpha: [0, 1] \rightarrow \mathbb{R}$ welche (2.1) für beliebige $y_0 \in [0, 1]$ mit $c_2 > 1$ erfüllen, sind konstant.

Beweis. Sei $x \in]0, 1[$. Es gilt für kleine h mittels (2.1) und $c_2 - 1 > 0$:

$$\left| \frac{\alpha(x+h) - \alpha(x)}{h} \right| \leq c_1 \cdot |h|^{c_2-1} \xrightarrow{h \rightarrow 0} 0,$$

womit α auf $]0, 1[$ differenzierbar mit Ableitung $\alpha' \equiv 0$ ist. Der eindimensionale Mittelwertsatz der Differentialrechnung liefert, dass α auf $]0, 1[$ konstant ist und die Stetigkeit setzt dies auf den Rand fort. \square

Die Bedingung 2.2.2 umfasst also in y_0 lokal Hölder-stetige Funktionen und jene, die ein Extremum in y_0 besitzen.

Definition 2.2.5.

Eine Abbildung $f: [0, \infty[$ heißt **regulär variierend in 0 mit Index ρ** , falls

$$\lim_{t \downarrow 0} \frac{f(t \cdot \lambda)}{f(t)} = \lambda^\rho, \quad \lambda > 0.$$

Im Falle $\rho = 0$ spricht man auch von **langsamer Variation in 0**.

Die reguläre Variation in 0 steht in einem engen Zusammenhang mit der klassischen regulären Variation (im Unendlichen), siehe Appendix A.2.7 und A.2.8 für die Definition und eine unvollständige Liste an Eigenschaften.

Bemerkung 2.2.6.

Eine Abbildung f wie oben ist genau dann regulär variierend in Null mit Index ρ , wenn die Abbildung

$$x \mapsto \frac{1}{f(\frac{1}{x})}$$

regulär variierend mit Index ρ ist.

Beweis. Dies erkennt man unmittelbar aufgrund folgender Gleichheit für $\lambda > 0$:

$$\lim_{x \rightarrow \infty} \frac{f(\frac{1}{x})}{f(\frac{1}{\lambda x})} = \lim_{t \downarrow 0} \left(\frac{f(t\lambda^{-1})}{f(t)} \right)^{-1}.$$

\square

Bedingung 2.2.7.

Es existieren Funktionen $S_0: [0, \infty[\rightarrow \mathbb{R}^+$ und $h: [0, 1] \rightarrow \mathbb{R}^+$, wobei S_0 langsam variierend in 0 und h stetig sei. Zusätzlich gelte folgende Konvergenz gleichmäßig in $y \in [0, 1]$:

$$\lim_{t \downarrow 0} \frac{S(t, y)}{S_0(t)} = h(y). \quad (2.2)$$

Die zentralen Theoreme in diesem und dem nächsten Abschnitt beschäftigen sich mit dem Fall, dass Δ ein Dreiecksschema zeilenweise unabhängiger Zufallsvariablen ist. Da wir in unserem Modell explizit erwirken, dass die Zufallsvariablen nicht identisch verteilt sind, ist die klassische Extremwerttheorie nicht anwendbar. Das Hilfsmittel in dieser Situation ist die Poisson-Approximation, siehe Appendix A.2.5.

Zunächst geben wir einige Vorbemerkungen und Notationen.

Bemerkung 2.2.8.

Wir setzen $\alpha_0 := \alpha(y_0)$ und $H(t) := S_0(t) \cdot t^{\alpha_0+1/2}$, sodass mittels Appendix A.2.8 H regulär variierend in 0 mit Index $\alpha_0 + 1/2$ ist.

Es sei zusätzlich $H(\cdot)$ strikt monoton. Nach Appendix A.2.8 ist die dann existierende Inverse H^{-1} regulär variierend in 0 mit Index $(\alpha_0 + 1/2)^{-1} =: \rho$.

Wir definieren zuletzt notationsvereinfachend $H_n := H^{-1}(1/n)$.

Die Strategie ist es nun, die Voraussetzungen einer Poisson-Approximation, siehe Appendix A.2.5 zu verifizieren, wobei wir an der stabilisierten Grenzverteilung von M_n interessiert sind. Damit wir den Kontext einer Poisson-Approximation erreichen, definieren wir

$$B_{i,n} := \mathbb{1}\{X_{i,n} > u_0 - xH_n\} \quad \text{und} \quad S_n := \sum_{i=1}^n B_{i,n}.$$

Dann sind die $B_{i,n}$ Bernoulli-verteilte Zufallsvariablen, welche zeilenweise unabhängig sind. Um die zwei Voraussetzungen der Poisson-Approximation A.2.5 zu zeigen, werden wir einige Hilfslemmata formulieren. Bevor wir dies tun, erinnern wir an die *Betafunktion*. Der Parameter der Poisson-Grenzverteilung ist von komplizierterer Gestalt, sodass wir hier erneut einige abkürzende Schreibweisen einführen. Sei $\delta := (-2/u(y_0)'')^{1/2}$, was wegen 2.2.1 wohldefiniert ist und

$$C := 2^{\alpha_0+1/2} \delta h(y_0) \text{Beta}(\alpha_0 + 1, \alpha_0 + 1)$$

mit der klassischen **Betafunktion**

$$\text{Beta}(x, y) := \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad \text{Re}(x) \wedge \text{Re}(y) > 0.$$

Wir werden in den Beweisen sehen, dass δ aus einer Entwicklung von u um y_0 natürlich auftritt und C als Wert einer limitierenden Reihe auftaucht. Nun formulieren wir das Theorem, vergleiche [Stein, Theorem 1].

Theorem 2.2.9 (Erster Grenzwertsatz für Maxima nicht-stationärer Zufallsvariablen).
Unter den Bedingungen 2.2.1, 2.2.2, 2.2.3 und 2.2.7 gilt in der Situation aus Bemerkung 2.2.8

$$\frac{M_n - u_0}{H_n} \xrightarrow[n \rightarrow \infty]{d} GEV \left(-C^{-1/(\alpha_0+1/2)}, \frac{1}{\alpha_0 + 1/2} C^{-1/(\alpha_0+1/2)}, -\frac{1}{\alpha_0 + 1/2} \right).$$

Die erste Voraussetzung der Poisson-Approximation aus A.2.5 ist einfacher nachzuweisen:

Lemma 2.2.10.

In derselben Situation wie in Theorem 2.2.9 gilt

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \mathbb{P}(X_{i,n} > u_0 - x \cdot H_n) = 0, \quad x > 0. \quad (2.3)$$

Beweis. Zunächst zeigen wir, dass $H_n \xrightarrow{n} 0$:

Mit den aus dem Darstellungssatz von Karamata folgenden Schranken für reguläre Variation, siehe Appendix A.2.8, erhalten wir für $0 < \delta < \rho$ und große n

$$\begin{aligned} n^{\rho-\delta} &\leq \frac{1}{H^{-1}(1/n)} \leq n^{\rho+\delta}, \quad \text{und damit} \\ \left(\frac{1}{n}\right)^{\rho-\delta} &\geq H^{-1}(1/n) \geq \left(\frac{1}{n}\right)^{\rho+\delta}, \end{aligned}$$

sodass eine Grenzwertbildung in n auf beiden Seiten in 0 resultiert und

$$\lim_{n \rightarrow \infty} H_n = \lim_{n \rightarrow \infty} H^{-1}(1/n) = 0$$

impliziert.

Die Bedingung 2.2.1 und $xH_n \rightarrow 0$ erwirken, dass sich im $\lim \max$ -Ausdruck der Quotient i/n dem maximalen Argument y_0 annähert.

Da u_0 eindeutiges Maximum der Randpunktsfunktion ist, lässt sich abschätzen:

$$\begin{aligned} &\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \mathbb{P}(X_{i,n} > u_0 - x \cdot H_n) \\ &\leq \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \mathbb{P}(X_{i,n} > u(i/n) - xH_n) \\ &= \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} S(xH_n, i/n) (xH_n)^{\alpha(i/n)} \\ &= \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{S(xH_n, i/n)}{S_0(xH_n)} S_0(xH_n) (xH_n)^{\alpha(i/n) - \alpha(y_0)} (xH_n)^{\alpha(y_0)} \\ &\leq \lim_{n \rightarrow \infty} S_0(xH_n) (xH_n)^{\alpha(y_0)} \cdot \lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{S(xH_n, i/n)}{S_0(xH_n)} \\ &= 0 \cdot h(y_0) = 0, \end{aligned}$$

wobei die erste Gleichheit aus 2.2.2 und die zweite Gleichheit aus zwei Eins-Multiplikationen

(man beachte S_0 ist positiv) folgt. Die letzte Ungleichung nutzt aus, dass (xH_n) wegen des ersten Schrittes und $\alpha(i/n) - \alpha(y_0)$ wegen (2.1) für große n klein werden, sodass der Potenzausdruck großzügig gegen 1 abgeschätzt werden kann. Final wird bei der vorletzten Gleichheit verwendet, dass $S_0(\cdot)(\cdot)^{\alpha(y_0)}$ regulär variierend in Null mit positivem Index ist und für den rechten Faktor wurde die Bedingung 2.2.7 angewandt. \square

Die folgenden Lemmata dienen dem Nachweis der zweiten Voraussetzung der Poisson-Approximation.

Lemma 2.2.11.

Sei $f: [a, b] \rightarrow \mathbb{R}$ eine $C^2([a, b])$ Abbildung, die ein eindeutiges Maximum M im Punkt x_0 annimmt und $f''(x_0) < 0$ erfülle.

Dann gibt es für Nullfolgen $a_n > 0$ einen Index $N \in \mathbb{N}$ ab dem

$$f^{-1}(]M - a_n, \infty[) =]x_n^1, x_n^2[,$$

wobei $x_0 \in]x_n^1, x_n^2[$, für alle $n \geq N$ gilt.

Dies bedeutet anschaulich, dass nur innerhalb einer Umgebung vom Maximum sich diesem auch beliebig angenähert werden kann. Es kann also nirgendwo eine Konvergenz gegen M stattfinden. Man beachte, dass dabei die Beschränktheit wichtig ist. Andernfalls denke man exemplarisch an ein $M - 1/x$ Verhalten für $x \rightarrow \infty$.

Beweis. Zunächst wählen wir (bezüglich Mengeninklusion) maximale $x_1' < x_0$ und $x_2' > x_0$ mit

$$f'|_{[x_1', x_0[} > 0 \quad \text{beziehungsweise} \quad f'|_{]x_0, x_2']} < 0,$$

was möglich ist, da sonst das Maximum M nicht eindeutig wäre.

Anschließend wählen wir maximale $x_1'' < x_0$ und $x_2'' > x_0$ derart, dass

$$f''|_{[x_1'', x_2''] \setminus \{x_0\}} < 0$$

erfüllt ist, wobei die Existenz solcher Punkte gesichert ist, da $f''(x_0) < 0$ gilt und die zweite Ableitung stetig ist. Wir setzen $U :=]x_1'', x_2''[\cap]x_1', x_2'[,$ und bemerken, dass U sowohl offen ist als auch x_0 enthält. Da U als Schnitt zweier offener Intervalle selbst ein offenes Intervall ist, schreiben wir es als $]x_1, x_2[$. Wegen der Stetigkeit von f und Kompaktheit von $[a, b] \setminus U$ folgt für ein $\varepsilon > 0$

$$\sup_{x \in [a, b] \setminus U} f(x_0) - f(x) = \max_{x \in [a, b] \setminus U} f(x_0) - f(x) \geq \varepsilon > 0,$$

wobei die Positivität aus der Eindeutigkeit des Maximums folgt.

Wähle nun ein N mit $a_n < \varepsilon$ für alle $n \geq N$, sodass für $x \notin U$

$$f(x) = f(x_0) + f(x) - f(x_0) \leq f(x_0) - \varepsilon < f(x_0) - a_n$$

und damit $U^c \cap f^{-1}(]M - a_n, \infty[) = \emptyset$ gilt. Wir haben also eine Monotonieumgebung U von x_0 gefunden außerhalb der Elemente nicht in der Zielmenge liegen können.

Die Aussage folgt nun daraus, dass wegen der Stetigkeit von f in einer Umgebung von x_0 die Funktion f von unten durch $M - a_n$ beschränkt werden kann. Man beachte, dass diese Umgebung natürlich von n abhängt. \square

Lemma 2.2.12.

In derselben Situation wie in Theorem 2.2.9 gilt

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - x \cdot H_n) = Cx^{\alpha_0+1/2}, \quad x > 0. \quad (2.4)$$

Beweis. Wir fixieren $x > 0, \varepsilon \in]0, 1/2[$ und definieren $v_n(y) := u(y) - u_0 + xH_n$, so dass $v_n(y_0) = xH_n > 0$ und $\{y : v_n(y) > 0\} = v_n^{-1}(]0, \infty[)$ offen ist. Tatsächlich ist die Menge sogar für große n ein Intervall, da nach dem Beweis von Lemma 2.2.10 xH_n eine Nullfolge ist und wir daher Lemma 2.2.11 anwenden können.

Definiere y_{1n}, y_{2n} durch $v_n^{-1}(]0, \infty[) =]v_{1n}, v_{2n}[$. Auf dieser Menge gilt mittels Bedingung 2.2.2 und einer Nulladdition

$$\mathbb{P}(X_{i,n} > u_0 - xH_n) = S(v_n(i/n), i/n) \cdot v_n(i/n)^{\alpha(i/n)}. \quad (2.5)$$

Für große n sei $]y_{1n}^*, y_{2n}^*[$ das Intervall $v_n^{-1}(] \varepsilon xH_n, \infty[)$.

Schritt 1: Wir zeigen, dass eine in ε, n gleichmäßige Konstante K existiert, weswegen für große n

$$\sum_{i \geq ny_{2n}^*} \mathbb{P}(X_{i,n} > u_0 - xH_n) \wedge \sum_{i \leq ny_{1n}^*} \mathbb{P}(X_{i,n} > u_0 - xH_n) < K\varepsilon \quad (2.6)$$

gilt. Dies erlaubt es also, uns im Asymptotischen auf die Indexmenge

$I_n := \{i \in \mathbb{N} : ny_{1n}^* < i < ny_{2n}^*\}$ zurückzuziehen, da wegen (2.6) die komplementären i keinen Beitrag liefern, denn ε darf beliebig klein werden und K ist gleichmäßig in n .

Wir zeigen nur die erste Abschätzung aus (2.6). Die zweite folgt völlig analog, da nur symmetrische Abschätzungen genutzt werden.

Für große n und $i > ny_{2n}$ gilt nach Definition $0 \geq v_n(i/n) = u(i/n) - u_0 + xH_n$, sodass $u(i/n) \leq u_0 - xH_n$, was nach Definition des rechten Randpunktes gerade $\mathbb{P}(X_{i,n} > u_0 - xH_n) = 0$ bedeutet. Dies und die Bedingungen 2.2.1, 2.2.2 liefern daher

$$\begin{aligned} \sum_{i \geq ny_{2n}^*} \mathbb{P}(X_{i,n} > u_0 - xH_n) &= \sum_{ny_{2n}^* \leq i \leq ny_{2n}} \mathbb{P}(X_{i,n} > u_0 - xH_n) \\ &\leq \sum_{ny_{2n}^* \leq i \leq ny_{2n}} \mathbb{P}(X_{i,n} > u(i/n) - xH_n) = \sum_{ny_{2n}^* \leq i \leq ny_{2n}} S(xH_n, i/n)(xH_n)^{\alpha(i/n)}. \end{aligned} \quad (2.7)$$

Da u_0 das eindeutige globale Maximum ist, folgt $y_{1n} \rightarrow_n y_0 \leftarrow_n y_{2n}$, sodass auf letzter Indexmenge für große n der Quotient i/n nahe y_0 ist.

Aus der gleichmäßigen Konvergenz in (2.2) und wegen der Stetigkeit von α in y_0 folgen die Konvergenzen

$$\frac{S(xH_n, i/n)}{S_0(xH_n)} \xrightarrow{n \rightarrow \infty} h(y_0), \quad \alpha(i/n) \xrightarrow{n \rightarrow \infty} \alpha_0.$$

Obiges in (2.7) eingesetzt und die langsame Variation von S_0 in der Null liefern

$$\sum_{i \geq ny_{2n}^*} \mathbb{P}(X_{i,n} > u_0 - xH_n) \leq 2h(y_0)S_0(H_n) \max(1, x)^{2\alpha_0} \sum_{ny_{2n}^* \leq i \leq ny_{2n}} H_n^{\alpha(i/n)}. \quad (2.8)$$

Um die Summe abzuschätzen, Taylor-entwickeln wir die Randpunkte der Intervalle um y_0 .

Es gilt wegen der Stetigkeit von u und 2.2.1

$$\varepsilon xH_n = v_n(y_{2n}^*) = xH_n + \frac{u''(\xi_n)}{2}(y_{2n}^* - y_0)^2 \quad (2.9)$$

für ein $\xi_n \in]y_0, y_{2n}^*[$, wobei eine Taylorentwicklung von v_n um y_0 mit dem Lagrange'schen Restglied genutzt wurde. Die zweite Ableitung u'' ist stetig, sodass wegen $y_{2n}^* \rightarrow y_0$ – man beachte, dass oben $y_{2n} \rightarrow y_0$ gezeigt wurde und $y_{2n}^* \leq y_{2n}$ gilt – bereits $u''(\xi_n) = u''(y_0) + o(1)$ folgt. Wegen $\delta^2 = -2/u''(y_0)$ und dem Vorangegangenen erhalten wir mittels (2.9)

$$y_{2n}^* - y_0 = \sqrt{(1 - \varepsilon)xH_n(\delta^2 + o(1))} = \delta\sqrt{(1 - \varepsilon)xH_n} + o(H_n^{1/2}).$$

Das heißt $y_{2n}^* = y_0 + \delta\sqrt{(1 - \varepsilon)xH_n} + o(H_n^{1/2})$ und analog erhalten wir: $y_{1n}^* = y_0 - \delta\sqrt{(1 - \varepsilon)xH_n} + o(H_n^{1/2})$, $y_{1n} = y_0 - \delta\sqrt{xH_n} + o(H_n^{1/2})$ und $y_{2n} = y_0 + \delta\sqrt{xH_n} + o(H_n^{1/2})$. Damit folgt für große n

$$0 \leq y_{2n} - y_{2n}^* \leq \delta\sqrt{xH_n}(1 - \sqrt{1 - \varepsilon}) + o(H_n^{1/2}) \leq \varepsilon\delta\sqrt{xH_n},$$

da $1 - \varepsilon < \sqrt{1 - \varepsilon}$.

Demnach lässt sich die Summe in (2.8) wie folgt abschätzen:

$$\sum_{i \in J_n} H_n^{\alpha(i/n)} \leq 2\varepsilon\delta\sqrt{xH_n}nH_n^{\min\{\alpha(i/n) - \alpha_0: i \in J_n\}} \cdot H_n^{\alpha_0}, \quad (2.10)$$

wobei $J_n := \{i: ny_{2n}^* \leq i \leq ny_{2n}\}$.

Um die Konvergenz von $H_n^{\min\{\alpha(i/n) - \alpha_0: i \in J_n\}}$ gegen 1 zu zeigen, nutzen wir die lokale Hölder-Stetigkeit von α in y_0 , die reguläre Variation in Null von H^{-1} zum Index ρ und eine Regel von de L'Hospital.

Zunächst gilt für kleine $\varepsilon' > 0$ wegen der Schranken für regulär variierende Funktio-

nen aus A.2.8, dass

$$\left| \log H_n^{\min\{\alpha(i/n) - \alpha_0 : i \in J_n\}} \right| \leq \left| \min\{\alpha(i/n) - \alpha_0 : i \in J_n\} \cdot (\rho + \varepsilon') \log(1/n) \right|.$$

Der erste Faktor kann mittels lokaler Hölder-Stetigkeit, der Entwicklung von y_{2n} und der Variation gegen

$$\begin{aligned} \left| \min\{\alpha(i/n) - \alpha_0 : i \in J_n\} \right| &\leq \min_{i \in J_n} c_1 |i/n - y_0|^{c_2} \\ &\leq c_1 |y_{2n} - y_0|^{c_2} \leq \tilde{c}_1 |(\delta\varepsilon')\sqrt{x}|^{c_2} \cdot H_n^{c_2/2} \\ &\leq \tilde{c} \cdot (1/n)^{(\rho - \varepsilon') \cdot c_2/2} \end{aligned}$$

abgeschätzt werden. Beide Abschätzungen gemeinsam resultieren in Verbindung mit einer Anwendung von L'Hospital in der Konvergenz des \log – Ausdrucks gegen Null, was seinerseits wegen der Stetigkeit der Exponentialfunktion die Zielkonvergenz zeigt.

Dies in (2.10) eingesetzt liefert

$$\sum_{i \in J_n} H_n^{\alpha(i/n)} \leq 4\varepsilon\delta x^{1/2} n H_n^{\alpha_0 + 1/2},$$

was wegen $S_0(H_n) \cdot H_n^{\alpha_0 + 1/2} = H(H^{-1}(1/n)) = 1/n$ in Verbindung mit (2.8)

$$\sum_{i \geq ny_{2n}^*} \mathbb{P}(X_{i,n} > u_0 - xH_n) \leq 8h(y_0)\delta n \frac{1}{n} \varepsilon x^{1/2} \max\{1, x\}^{2\alpha_0} =: K\varepsilon$$

und damit das Gewünschte ergibt.

Für große n mit $i \leq ny_{1n}$ verläuft wegen $\mathbb{P}(X_{i,n} > u_0 - xH_n) = 0$ der Beweis analog.

Schritt 2: In diesem Schritt zeigen wir

$$\lim_{n \rightarrow \infty} \sum_{i \in I_n} \mathbb{P}(X_{i,n} > u_0 - xH_n) = Cx^{\alpha_0 + 1/2}.$$

Erneut zeigen wir nur eine Ungleichung, hier für den Limes superior, da die untere Abschätzung für den Limes inferior analog mittels symmetrischer Argumente folgt.

Initial gilt

$$\frac{v_n(i/n)^{\alpha(i/n)}}{v_n(i/n)^{\alpha_0}} = \left(H_n \frac{v_n(i/n)}{H_n} \right)^{\alpha(i/n) - \alpha_0}.$$

Auf I_n wird gerade die Beziehung $v_n(i/n) > \varepsilon x H_n$ erfüllt, sodass der rechte Quotient von der Null weg beschränkt ist. Wie im ersten Schritt folgt nun mittels Logarithmieren und Ausnutzen der vorher etablierten Eigenschaften die Konvergenz des Ausdrucks gegen 1.

Außerdem können wir $v_n(i/n)$ auf I_n durch $\varepsilon x H_n$ und $x H_n$ einschnüren, weswegen der Satz von der lokal gleichmäßigen Konvergenz für reguläre Variation, siehe A.2.8, und die Bedingung 2.2.7

$$\frac{S(v_n(i/n), i/n)}{S_0(H_n)h(y_0)} = \frac{S(v_n(i/n), i/n)}{S_0(v_n(i/n))} \frac{S_0(v_n(i/n))}{S_0(H_n)} \frac{1}{h(y_0)} \xrightarrow{n \rightarrow \infty} \frac{h(y_0)}{h(y_0)} = 1$$

generieren.

Die Gleichung (2.5) und beide Konvergenzen gegen 1 ergeben

$$\begin{aligned} \sum_{i \in I_n} \mathbb{P}(X_{i,n} > u_0 - x H_n) &= \sum_{i \in I_n} S(v_n(i/n), i/n) \cdot v_n(i/n)^{\alpha(i/n)} \\ &= \sum_{i \in I_n} \frac{S(v_n(i/n), i/n)}{S_0(H_n)h(y_0)} \cdot \frac{v_n(i/n)^{\alpha(i/n)}}{v_n(i/n)^{\alpha_0}} \cdot S_0(H_n)h(y_0)v_n(i/n)^{\alpha_0} \\ &< (1 + \varepsilon)^2 S_0(H_n)h(y_0) \sum_{i \in I_n} v_n(i/n)^{\alpha_0}. \end{aligned} \quad (2.11)$$

Die Summe werden wir nun mit einem Integral in Verbindung bringen, woraus auch letztlich der Betafunktionsanteil entsteht.

Wegen $i \in I_n$ gilt $i/n \in [y_{1n}^*, y_{2n}^*]$, sodass sich mittels vorangegangener Entwicklungen $(i/n - y_0)^2 = \delta^2 x H_n (1 - \varepsilon) + o(H_n)$ ergibt. Damit bestimmen wir die Taylorentwicklung von $v_n(i/n)$ um y_0 als

$$\begin{aligned} v_n(i/n) &= x H_n + \frac{u''(\xi_n)}{2} \left(\frac{i}{n} - y_0 \right)^2 \\ &= x H_n + \frac{u''(y_0) + o(1)}{2} (\delta^2 x H_n (1 - \varepsilon) + o(H_n)) \\ &= H_n (\varepsilon x + o(1) + o(1)^2) < H_n (1 + \varepsilon)^{1/\alpha_0} (\varepsilon x + o(1)) \\ &= (1 + \varepsilon)^{1/\alpha_0} (x H_n + (\varepsilon - 1) x H_n + o(H_n)) \\ &= (1 + \varepsilon)^{1/\alpha_0} \left(x H_n + \frac{u''(y_0)}{2} \cdot (\delta^2 x H_n (1 - \varepsilon) + o(H_n)) \right) \\ &= (1 + \varepsilon)^{1/\alpha_0} \left(x H_n + \frac{u''(y_0)}{2} \cdot \left(\frac{i}{n} - y_0 \right)^2 \right). \end{aligned} \quad (2.12)$$

Die Beziehung (2.12) resultiert in

$$\sum_{i \in I_n} v_n(i/n)^{\alpha_0} < (1 + \varepsilon) \sum_{i \in I_n} \left(x H_n + \frac{u''(y_0)}{2} \cdot \left(\frac{i}{n} - y_0 \right)^2 \right)^{\alpha_0} \quad (2.13)$$

Wir setzen abkürzend $\tilde{f}(i) := f(i/n) := (x H_n + u''(y_0)/2 \cdot (i/n - y_0)^2)^{\alpha_0} \geq 0$ auf der Menge $I_n = I_1 \cup I_2 =: \{i : ny_{1n}^* < i \leq ny_0\} \cup \{i : ny_0 < i < ny_{2n}^*\}$ und nehmen ohne Einschränkung an, dass die Randpunkte und y_0 mit n multipliziert keine natürlichen

Zahlen sind. Dabei wird die Abhängigkeit der I_j von n in der Notation unterdrückt.

Unter dieser Notation ist \tilde{f} auf I_1 strikt isoton und auf I_2 strikt antiton und wir zerlegen die Summe in ihre Monotoniebereiche. Es gilt für den ersten Bereich:

$$\begin{aligned} \sum_{i \in I_1} \tilde{f}(i) &= \sum_{i \in I_1, i < \lfloor ny_0 \rfloor} \tilde{f}(i) + \tilde{f}(\lfloor ny_0 \rfloor) \\ &\leq \sum_{i \in I_1, i < \lfloor ny_0 \rfloor} \int_i^{i+1} \tilde{f}(s) ds + \tilde{f}(\lfloor ny_0 \rfloor) = \int_{\lceil ny_{1n}^* \rceil}^{\lfloor ny_0 \rfloor} f\left(\frac{s}{n}\right) ds + \tilde{f}(\lfloor ny_0 \rfloor) \\ &= n \int_{\lceil ny_{1n}^* \rceil / n}^{\lfloor ny_0 \rfloor / n} f(s) ds + f\left(\frac{\lfloor ny_0 \rfloor}{n}\right). \end{aligned} \quad (2.14)$$

Ziel ist es

$$d_n := \left| \int_{y_{1n}^*}^{y_0} f(s) ds - \int_{\lceil ny_{1n}^* \rceil / n}^{\lfloor ny_0 \rfloor / n} f(s) ds \right| < \varepsilon \int_{y_{1n}^*}^{y_0} f(s) ds \quad (2.15)$$

zu zeigen, da dann mittels einer Nulladdition

$$\int_{\lceil ny_{1n}^* \rceil / n}^{\lfloor ny_0 \rfloor / n} f(s) ds < (1 + \varepsilon) \cdot \int_{y_{1n}^*}^{y_0} f(s) ds$$

gelten würde. Der Störsummand in (2.14) lässt sich zuletzt als asymptotisch vernachlässigbar identifizieren.

Dazu vergleichen wir Konvergenzraten. Wir beachten zunächst aufgrund der Isotonie

$$d_n = \int_{\lfloor ny_0 \rfloor / n}^{y_0} f(s) ds + \int_{y_{1n}^*}^{\lceil ny_{1n}^* \rceil / n} f(s) ds \leq \frac{1}{n} \left(f(y_0) + f\left(\frac{\lceil ny_{1n}^* \rceil}{n}\right) \right).$$

Einerseits gilt für $\varepsilon' > 0$ klein genug

$$\begin{aligned} f\left(\frac{\lceil ny_{1n}^* \rceil}{n}\right)^{1/\alpha_0} &= f\left(y_{1n}^* + o\left(\frac{1}{n^{1-\varepsilon'}}\right)\right)^{1/\alpha_0} \\ &= xH_n + \frac{u''(y_0)}{2} \left((xH_n)^{1/2} \left(o(1) - \delta\sqrt{1-\varepsilon} + o\left(\frac{1}{n^{1-\varepsilon'}}\right) \right) \right) \\ &= xH_n (1 - (1 - \varepsilon) + o(1)) = \varepsilon xH_n + o(1), \end{aligned}$$

wobei $(xH_n)^{-1/2} o(n^{-(1-\varepsilon')}) = o(1)$, da wegen der Tatsache $H^{-1} \in RV_0(1/(\alpha_0 + 1/2))$ für genügend kleine ε'

$$\frac{H_n^{-1/2}}{n^{1-\varepsilon'}} \leq \left(\frac{1}{n}\right)^{-(1/(2\alpha_0+1)+\varepsilon'/2)+1-\varepsilon'} = \left(\frac{1}{n}\right)^{2\alpha_0/(2\alpha_0+1)-3\varepsilon'/2} \xrightarrow{n \rightarrow \infty} 0$$

folgt. Damit erhalten wir $d_n \leq (xH_n)^{\alpha_0} / n(1 + \varepsilon^{\alpha_0} + o(1))$.

Andererseits ergibt sich für die untere Schranke

$$\begin{aligned} \int_{y_{1n}^*}^{y_0} f(s) ds &\geq (y_0 - y_{1n}^*) \cdot f(y_{1n}^*) \\ &= (xH_n)^{1/2} \left(\delta \sqrt{1 - \varepsilon} + o(1) \right) (xH_n)^{\alpha_0} (\varepsilon^{\alpha_0} + o(1)) \\ &= (xH_n)^{\alpha_0 + 1/2} \left(\delta \varepsilon^{\alpha_0} \sqrt{1 - \varepsilon} + o(1) \right). \end{aligned}$$

Aus $H^{-1} \in RV_0(1/(1/2 + \alpha_0))$ folgt jedoch wie oben $n \cdot H_n^{1/2} \rightarrow \infty$, sodass aus dieser Tatsache und den beiden erarbeiteten Schranken (2.15) folgt.

Zuletzt gilt für den Störterm in (2.14), dass $|f(\lceil ny_0 \rceil / n)| \leq f(y_0) = (xH_n)^{\alpha_0}$. Wieder folgt wegen $n \cdot H_n^{1/2} \rightarrow \infty$ die asymptotische Vernachlässigbarkeit und final erhalten wir

$$\sum_{i \in I_1} \tilde{f}(i) < (1 + \varepsilon) n \int_{y_{1n}^*}^{y_0} f(s) ds.$$

Verfährt man für die Summe auf I_2 analog, so erhält man insgesamt die Abschätzung

$$\sum_{i \in I_n} \tilde{f}(i) < (1 + \varepsilon) n \int_{y_{1n}^*}^{y_{2n}^*} f(s) ds.$$

Diese Abschätzung in (2.13) eingesetzt liefert mittels der Entwicklungen von y_{1n}^*, y_{2n}^* und einer Substitution

$$\begin{aligned} &\sum_{i \in I_n} v_n(i/n)^{\alpha_0} \\ &< (1 + \varepsilon)^2 \cdot n \int_{y_{1n}^*}^{y_{2n}^*} \left(xH_n + \frac{u''(y_0)}{2} (s - y_0)^2 \right)^{\alpha_0} ds \\ &\leq (1 + \varepsilon)^2 \cdot n \int_{y_0 - \delta \sqrt{xH_n}}^{y_0 + \delta \sqrt{xH_n}} \left(xH_n + \frac{u''(y_0)}{2} (s - y_0)^2 \right)^{\alpha_0} ds \\ &= (1 + \varepsilon)^2 n \delta (xH_n)^{\alpha_0 + 1/2} \int_{-1}^1 (1 - t^2)^{\alpha_0} dt \\ &= (1 + \varepsilon)^2 n \delta (xH_n)^{\alpha_0 + 1/2} 2^{2\alpha_0 + 1} \text{Beta}(\alpha_0 + 1, \alpha_0 + 1), \end{aligned}$$

wobei $t = (s - y_0) / (\delta \sqrt{xH_n})$, im letzten Schritt die Beziehung

$\int_0^1 (1 - t^2)^{\mu-1} dt = 2^{2(\mu-1)} \text{Beta}(\mu, \mu)$, für ein $\text{Re}(\mu) > 0$, und dann die Achsensymmetrie um Null genutzt wurden, siehe [Gradshteyn und Rhyzik, Gleichung 3.249.5].

Setzen wir diese Beziehung in (2.11) ein und nutzen, dass per Definition

$S_0(H_n)H_n^{\alpha_0+1/2} = H(H_n) = 1/n$ gilt, so ergibt sich insgesamt

$$\begin{aligned} & \sum_{i \in I_n} \mathbb{P}(X_{i,n} > u_0 - xH_n) \\ & < (1 + \varepsilon)^4 n \delta(xH_n)^{\alpha_0+1/2} S_0(H_n) h(y_0) 2^{2\alpha_0+1} \cdot \text{Beta}(\alpha_0 + 1, \alpha_0 + 1) \\ & = (1 + \varepsilon)^4 x^{\alpha_0+1/2} C \end{aligned}$$

gleichmäßig in n , sodass in Verbindung mit Schritt 1

$$\limsup_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} < u_0 - xH_n) \leq Cx^{\alpha_0+1/2} \quad (2.16)$$

wegen der Beliebigkeit des $\varepsilon > 0$ folgt.

Die Tatsache, dass wir im gesamten zweiten Schritt symmetrische Argumente genutzt haben, generiert dieselbe Schranke für den Limes inferior wie in (2.16). Die Konklusion auf (2.4) ist mithin valide. \square

Nach dieser Vorarbeit können wir nun die in Theorem 2.2.9 behauptete Verteilungskonvergenz beweisen. Wir bemerken, es an dieser Stelle nicht offensichtlich, dass die existierende Grenzvariable extremwertverteilt ist. Dies muss in diesem allgemeinen Setting nicht identisch verteilter Zufallsvariablen auch nicht immer gelten.

Theorem 2.2.9 (Erster Grenzwertsatz für Maxima nicht-stationärer Zufallsvariablen).
Unter den Bedingungen 2.2.1, 2.2.2, 2.2.3 und 2.2.7 gilt in der Situation aus Bemerkung 2.2.8

$$\frac{M_n - u_0}{H_n} \xrightarrow[n \rightarrow \infty]{d} \text{GEV} \left(-C^{-1/(\alpha_0+1/2)}, \frac{1}{\alpha_0 + 1/2} C^{-1/(\alpha_0+1/2)}, -\frac{1}{\alpha_0 + 1/2} \right).$$

Beweis. Nach der Poisson-Approximation, A.2.5, und den Lemmata 2.2.10, 2.2.12 erhalten wir, die früher eingeführte Notation nutzend,

$$S_n \xrightarrow{d} \text{Poi} \left(\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - xH_n) \right) = \text{Poi} \left(Cx^{1/2+\alpha_0} \right), \quad x > 0.$$

Damit folgt also für $x > 0$ einerseits

$$\begin{aligned} \mathbb{P}(S_n = 0) &= \mathbb{P}\{M_n \leq u_0 - xH_n\} = \mathbb{P}\left\{ \frac{M_n - u_0}{H_n} \leq -x \right\} \\ &\xrightarrow[n \rightarrow \infty]{} \text{Poi} \left(Cx^{\alpha_0+1/2} \right) (0) \\ &= \exp \left(-Cx^{\alpha_0+1/2} \right) \end{aligned} \quad (2.17)$$

und andererseits wegen der Definition von u_0 als globales Maximum der Randpunkts-

funktion aus Definition 2.1.2

$$\begin{aligned} \mathbb{P}(M_n \leq u_0 + xH_n) &\geq \mathbb{P}(M_n \leq u_0) \\ &= \mathbb{P}(\cap_{i=1}^n \{X_{i,n} \leq u_0\}) \geq \mathbb{P}(\cap_{i=1}^n \{X_{i,n} \leq u(i/n)\}) = 1. \end{aligned}$$

Damit konvergiert M_n in Verteilung gegen eine nicht-degenerierte Zufallsvariable, welche wir mit G bezeichnen werden.

Es bleibt zu zeigen, dass G die behauptete Verteilung besitzt.

Falls G extremwertverteilt ist, so gehört sie zwangsläufig zu der Klasse vom Weibull-Typ, da sie den Randpunkt 0 besitzt und somit die Tails dünn beziehungsweise der Formparameter ζ kleiner als Null ist.

Aus der Gleichung (2.17) entnehmen wir die Konvergenz von $(M_n - u_0)/H_n$ gegen eine $\text{Wei}(0, C^{1/(\alpha_0+1/2)}, \alpha_0 + 1/2)$ -verteilte Zufallsvariable. Die Umrechnungsformel aus Bemerkung 1.2.3 liefert somit

$$\begin{aligned} G &= \text{Wei}(0, C^{1/(\alpha_0+1/2)}, \alpha_0 + 1/2) \\ &= \text{GEV}(-C^{-1/(\alpha_0+1/2)}, C^{-1/(\alpha_0+1/2)}/(\alpha_0 + 1/2), -1/(\alpha_0 + 1/2)). \end{aligned}$$

Dies beendet den Beweis. □

Mit diesem Resultat lässt sich leicht ein schwaches Gesetz für Maxima von unabhängigen und nicht-stationären Zufallsvariablen formulieren.

Korollar 2.2.13 (Schwachtes Gesetz für die erste Asymptotik).

In der Situation aus Theorem 2.2.9 gilt

$$M_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} u_0.$$

Beweis. Es gilt $H_n \xrightarrow{n} 0$, sodass $M_n - u_0 = (M_n - u_0)/H_n \cdot H_n \xrightarrow{d} G \cdot 0 = 0$ nach Theorem 2.2.9 und dem Lemma von Slutsky folgt. Die Verteilungskonvergenz gegen eine Konstante impliziert die Konvergenz in Wahrscheinlichkeit. □

Da hier ein Dreiecksschema von Zufallsvariablen betrachtet wird, lässt sich der übliche Schluss auf die fast-sichere Konvergenz wegen der fehlenden Isotonie der M_n nicht tätigen. Jedoch können wir aufgrund des Tailverhaltens aus Bedingung 2.2.2 Folgendes zeigen:

Satz 2.2.14 (Starkes Gesetz für die erste Asymptotik).

In der Situation aus Theorem 2.2.9 gilt $M_n \xrightarrow{f.s.} u_0$.

Beweis. Es bezeichne x^* das Maximum u_0 . Zunächst gilt ähnlich wie im Beweis von

Appendix A.2.3

$$\begin{aligned}
 \mathbb{P}\{M_n \not\xrightarrow{u} x^*\} &= \mathbb{P}\{\liminf M_n < x^*\} = \mathbb{P}\left(\bigcup_{M \in \mathbb{N}} \{\liminf M_n < x^* - 1/M\}\right) \\
 &= \lim_{M \rightarrow \infty} \mathbb{P}\{\liminf M_n < x^* - 1/M\} \leq \lim_M \liminf_n \mathbb{P}\{M_n < x^* - 1/M\} \\
 &= \lim_M \liminf_n \prod_{i=1}^n F_{i,n}(x^* - 1/M), \tag{2.18}
 \end{aligned}$$

wobei die erste Gleichheit aus der Tatsache $M_n \leq x^*$, die dritte aus der Stetigkeit von Maßen, die erste Ungleichheit aus dem Lemma von Fatou und die letzte Gleichheit aus der Unabhängigkeit folgt. Dabei bezeichnen wir mit $F_{i,n}$ die Verteilungsfunktion von $X_{i,n}$. Beschränken wir uns im Produkt auf die Indexmenge

$$I_M := \{i \in \{1, \dots, n\} : u(i/n) > x^* - 1/(2M)\},$$

so kann wegen $F_{i,n} \leq 1$ das Produkt sicherlich nur größer werden. Im Folgenden werden wir $F_{i,n}(x^* - 1/M) < 1$ gleichmäßig in i und n auf I_M zeigen, sodass das Produkt im Grenzwert (in n) gegen Null konvergiert; beachte, dass die Kardinalität von I_M für $n \rightarrow \infty$ auch bestimmt gegen unendlich divergiert, da u stetig und \mathbb{Q} dicht in \mathbb{R} liegt.

Wir schließen mittels Bedingung 2.2.2

$$\begin{aligned}
 &F_{i,n}(x^* - 1/M) \\
 &= 1 - \mathbb{P}\{X_{i,n} > u(i/n) - (1/M + u(i/n) - x^*)\} \tag{2.19} \\
 &= 1 - \frac{S(1/M + u(i/n) - x^*, i/n)}{S_0(1/M + u(i/n) - x^*)} \cdot S_0(1/M + u(i/n) - x^*) \cdot (1/M + u(i/n) - x^*)^{\alpha(i/n)} \\
 &= 1 - \frac{S(1/M + u(i/n) - x^*, i/n)}{S_0(1/M + u(i/n) - x^*)} \cdot S_0(1/M + u(i/n) - x^*) \cdot \\
 &\quad (1/M + u(i/n) - x^*)^{\alpha(y_0)} \cdot (1/M + u(i/n) - x^*)^{\alpha(i/n) - \alpha(y_0)}.
 \end{aligned}$$

Beachte nun, dass $1/M + u(i/n) - x^* \in [1/(2M), 1/M]$ nach Definition von I_M gleichmäßig in i, n gilt. Da $x^* = u_0$ das eindeutige Maximum von u ist und für große M demnach $u(i/n)$ beliebig nahe an x^* liegt, folgt dass für ein beliebig kleines $\varepsilon_M := \varepsilon(M) > 0$ die Beziehung $i/n \in B(y_0, \varepsilon_M)$ gilt. Wegen der lokalen Hölder-Stetigkeit, siehe Bedingung 2.2.3, von $\alpha(\cdot)$ in y_0 folgt daraus $\alpha(i/n) \in B(y_0, c_1 \cdot |i/n - y_0|^{c_2})$, sodass wir

$$\left(\frac{1}{M} + u\left(\frac{i}{n}\right) - x^*\right)^{\alpha(i/n) - \alpha(y_0)} > \left(\frac{1}{2M}\right)^{-\varepsilon_{i,n}},$$

für ein von i, n abhängiges $\varepsilon_{i,n} > 0$, schließen können. Insgesamt können wir wegen der Schranken für in der Null regulär variierender Funktionen, Appendix A.2.8, der

Bedingung 2.2.7 und der Stetigkeit von h_0 aus der Gleichung (2.19)

$$\begin{aligned}
& F_{i,n}(x^* - 1/M) \\
& \leq 1 - \min_{x \in [0,1]} \frac{h_0(x)}{2} \cdot S_0(1/M + u(i/n) - x^*) \left(\frac{1}{2M} \right)^{\alpha(y_0) - \varepsilon_{i,n}} \\
& \leq 1 - \min_{x \in [0,1]} \frac{h_0(x)}{2} \cdot \left(\frac{1}{2M} \right)^{\delta + \alpha(y_0) - \varepsilon_{i,n}} \\
& \leq 1 - \min_{x \in [0,1]} \frac{h_0(x)}{2} \cdot \left(\frac{1}{2M} \right)^{\alpha(y_0)} =: c_M < 1
\end{aligned}$$

gleichmäßig in i, n schließen.

Insgesamt folgt damit aus dieser Schranke für die $F_{i,n}(x^* - 1/M)$ und Gleichung (2.18)

$$\mathbb{P}\left\{M_n \not\rightarrow x^*\right\} \leq \lim_M \liminf_n \prod_{i \in I_M} c_M = \lim_M 0 = 0$$

und die Aussage ist gezeigt. \square

Vergleichen wir diese Version des starken Gesetzes mit der im Kontext von u.i.v Zufallsvariablen bemerken wir, dass in der letzteren Situation alle $X_{i,n}$ dieselben Randpunkte hatten, was hier gerade nicht der Fall ist. Vereinfacht ausgedrückt lässt sich sagen, dass das Maximum asymptotisch von denjenigen Beobachtungen bestimmt werden, welche um den Zeitpunkt y_0 stattgefunden haben. Wir können nun die Aussage des Korollars dahingehend interpretieren, dass diese Zufallsvariablen asymptotisch den rechten Randpunkt u_0 besitzen. Man beachte, dass sich diese Heuristik wegen der Bedingungen am Anfang des Kapitels auch mathematisch konkretisieren lässt und dies auch im vorangegangenen Gesetz passiert ist.

Bemerkung 2.2.15 (Erweiterung auf schwache Abhängigkeit).

Obige Resultate lassen sich erweitern auf Zeitreihen, die eine gewisse zeitliche Abhängigkeit besitzen, siehe [Stein, 2.3 Dependent Case]. Genauer setzt man die Bedingungen $D(u_n)$ und $D'(u_n)$ voraus, wobei erstere Bedingung eine asymptotische Unabhängigkeit der Beobachtungen fordert, falls sie weit genug auseinander liegen und letztere die Wahrscheinlichkeit von mehr als einem Exzess $X_{i,n} > u_n$ beschränkt, siehe [Leadbetter, Lindgren und Rootzén, Abschnitte 3.2 und 3.4]. Dies führen wir in dieser Arbeit jedoch nicht weiter aus.

2.3 Zweite Asymptotik zeilenweiser Unabhängigkeit

In diesem Subkapitel wollen wir ein anderes asymptotisches Verhalten der Randpunktfunktion unterstellen, um so eine nicht-degenerierte Grenzverteilung zu erhalten, welche nicht zur Klasse der Extremwertverteilung gehört. Dies ist insofern interessant, da es eine konkrete nicht-parametrische Verteilungsklasse anbietet, die als

Grenzwert von Maxima nicht-stationärer und unabhängiger Dreiecksschemata auftreten kann. Man beachte, dass Theorem 2.2.9 jedoch auch zeigt, dass die GEV-Verteilung als Grenzwert auftreten kann. Damit sehen wir, dass der Fall von nicht-Stationarität deutlich mehr Variabilität in der Grenzverteilung erwirkt.

Wir setzen $\Delta := \{X_{i,n} : 1 \leq i \leq n, n \in \mathbb{N}\}$ erneut als Dreiecksschema unabhängiger nicht-stationärer Zufallsvariablen an.

Bedingung 2.3.1 (Annahmen an die Funktionen in der zweiten Asymptotik).

Es seien $\bar{\alpha}, h, v: [0, 1] \rightarrow \mathbb{R}$ stetige Abbildungen, wobei v nichtnegativ und h positiv sei. Zusätzlich sei $S: \mathbb{R}^+ \rightarrow \mathbb{R}^+$ eine in der Null langsam variierende Funktion.

Der Beweis wird unter anderem mittels eines Riemannsummenkonvergenz-Arguments seitens Stein geführt. Dies setzt aber insbesondere voraus, dass der Integrationsbereich eine Menge ist, die Riemann-integrierbar ist, was aus dem Beweis von Stein nicht hervorgeht und nach längerer Überlegung auch nicht korrekt scheint; ein Gegenbeispiel geben wir nach der Bedingung an. Daher fordern wir zusätzlich ein nicht übermäßig pathologisches Verhalten der Funktion v :

Bedingung 2.3.2 (Nicht-Pathologie des v).

Für die Abbildung v gilt, dass Urbilder $v^{-1}([x, y])$ durch eine höchstens endliche Vereinigung von Intervallen darstellbar seien, wobei $0 \leq x < y < \infty$.

Gegenbeispiele lassen sich durch stark oszillierende Funktionen konstruieren.

Beispiel 2.3.3 (Pathologisches v).

Man setze die Abbildung v als

$$v: [0, 1] \rightarrow \mathbb{R}_0^+, \quad x \mapsto \frac{1}{2} + \left(x - \frac{1}{2}\right) \cdot \sin\left(\frac{1}{x - 1/2}\right),$$

siehe Abbildung 2. Dann ist das Urbild $v^{-1}([0, 1/2])$ keine endliche Vereinigung von Intervallen, obwohl v stetig und nichtnegativ ist.

Beweis. Die Abbildung ist offensichtlich außerhalb von $1/2$ stetig. Für den Punkt $1/2$ beachten wir, dass der reelle Sinus beschränkt ist. Die nicht-Negativität sieht man mittels einer Fallunterscheidung oder Betrachtung des Graphen, schnell ein.

Es gilt für jedes kleine $\varepsilon > 0$ die Beziehung $\sin(B(1/2, \varepsilon) \setminus \{1/2\}) = [-1, 1]$. Daraus folgt aber insbesondere, dass für all jene ε Punkte $y_1 \in B(1/2, \varepsilon) \setminus \{1/2\}$ mit $v(y_1) < 1/2$ und $y_2 \in B(1/2, \varepsilon) \setminus \{1/2\}$ mit $v(y_2) > 1/2$ existieren. Also kann $v^{-1}([0, 1/2])$ keine endliche Vereinigung von Intervallen sein. \square

Wie schon in der Einleitung angesprochen, wollen wir ein anderes asymptotisches Verhalten der Randpunktfunktion u unterstellen.

Bedingung 2.3.4 (Bedingung an u in der zweiten Asymptotik).

Die Randpunktfunktion sei gegeben durch

$$u: [0, 1] \rightarrow [0, 1], \quad x \mapsto u_0 - G_n \cdot v(x),$$

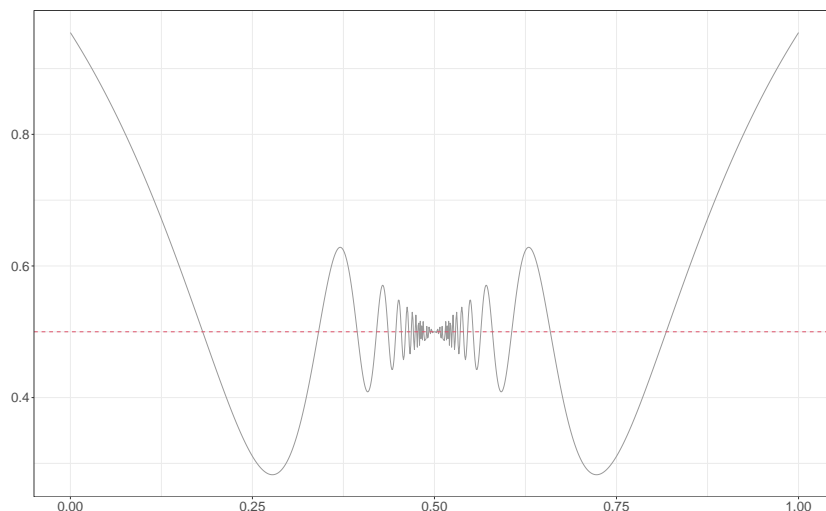


Abbildung 2: Graph der Abbildung v aus Beispiel 2.3.3 mit eingezeichneter Horizontale bei $y = 0.5$.

wobei $G_n := G^{-1}(1/n)$ mit $\alpha_0, u_0 > 0$, $G(x) := S(x) \cdot x^{\alpha_0}$ als strikt isoton vorausgesetzt.

Damit ist G insbesondere regulär variierend in der Null vom Index α_0 und dementsprechend G^{-1} vom Index $1/\alpha_0$, siehe Appendix A.2.8.

Auch werden wir eine gegen Null konvergierende Störung in α_0 zulassen.

Bedingung 2.3.5 (Störung von α_0).

Für $n \in \mathbb{N}$ seien die Abbildungen $\alpha_n(\cdot)$ durch $\alpha_n(x) = \alpha_0 + \bar{\alpha}(x)/\log n$ gegeben.

Zuletzt geben wir eine Form für den Tail der Beobachtungen $X_{i,n}$ vor.

Bedingung 2.3.6 (Tailverhalten).

Für kleine $t > 0$ gelte

$$\mathbb{P}\{X_{i,n} > u(i/n) - t\} = h(i/n) S(t) \cdot t^{\alpha_n(i/n)}. \quad (2.20)$$

Wie in der ersten Asymptotik formulieren wir zuerst das Resultat bevor wir Hilfslemmata für den Beweis erarbeiten, vergleiche [Stein, Theorem 2]

Theorem 2.3.7 (Zweiter Grenzwertsatz für Maxima nicht-stationärer Zufallsvariablen).
 Unter den Bedingungen 2.3.1, 2.3.2, 2.3.4, 2.3.5 und 2.3.6 gilt mit der Bezeichnung
 $M_n := \max \{ X_{i,n} : 1 \leq i \leq n \}$

$$\frac{M_n - u_0}{G_n} \xrightarrow{d} G, \quad (2.21)$$

$$\text{wobei } G(x) = \begin{cases} \exp(-J(-x)) & , x < 0 \\ 1 & , x \geq 0 \end{cases} \quad \text{und}$$

$$\begin{aligned} J(x) &:= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - xG_n) \\ &= \int_0^1 h(y) \cdot \exp\left(-\frac{\bar{\alpha}(y)}{\alpha_0}\right) \cdot (x - v(y))_+^{\alpha_0} dy \end{aligned} \quad (2.22)$$

Wir bemerken, dass die Grenzverteilung eine im Allgemeinen sehr komplizierte Gestalt hat. Man könnte sich fragen, ob sich die GEV-Verteilung in diese Klasse von Verteilungen einbetten lässt. Dies werden wir später negativ beantworten können. Es sei angemerkt, dass Stein auch den Fall behandelt hat, dass die Funktionen von $[a, b[$ mit $-\infty \leq a < b \leq \infty$, wobei für unendliche Grenzen ein offenes Intervall gewählt wird. Dies würde jedoch in der Interpretation einem unendlich langen Jahr oder einem Jahr in dem an gewissen Tagen kein Maximum eintreten kann entsprechen. Für den Fall nicht-endlicher Grenzen konnte er zeigen, dass auch GEV-Verteilungen als Grenzwert entstehen können. Da diese Annahmen künstlich und für die Anwendung schwierig haltbar erscheinen, verweisen wir an dieser Stelle nur auf den Artikel von Stein und gehen nicht weiter auf diesen Spezialfall ein. Die allgemeine Strategie des Beweises von Theorem 2.3.7 besteht erneut darin die Voraussetzungen der Poisson-Approximation aus Appendix A.2.5 nachzuweisen. Folgende Lemmata sind dem dediziert.

Lemma 2.3.8.

In der Situation aus Theorem 2.3.7 gilt

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \mathbb{P}(X_{i,n} > u_0 - G_n x) = 0, \quad x > 0. \quad (2.23)$$

Beweis. Wir fixieren ein $x > 0$. Aus der regulären Variation von G^{-1} in der Null, siehe Bedingung 2.3.4, erhalten wir die Konvergenz $G_n \rightarrow 0$ und da wir $\|v\|_\infty < \infty$ aus der Stetigkeitsvoraussetzung deduzieren können, gilt $G_n \cdot (x - v(i/n)) \rightarrow 0$ gleichmäßig in i . Somit können wir Bedingung 2.3.6 mit $t = G_n \cdot (x - v(i/n))$ nutzen und sehen

$$\begin{aligned} &\mathbb{P}(X_{i,n} > u_0 - G_n x) \\ &= h(i/n) S(G_n \cdot (x - v(i/n))) \cdot (G_n \cdot (x - v(i/n)))^{\alpha_n(i/n)} \\ &\leq \|h\|_\infty S(G_n \cdot (x - v(i/n))) \cdot G_n^{\alpha_n(i/n)} \cdot (x - v(i/n))^{\alpha_n(i/n)}. \end{aligned} \quad (2.24)$$

Die Abbildung $y \mapsto S(y) \cdot y^{\alpha_0}$ ist regulär variierend in der Null mit Index $\alpha_0 > 0$, sodass $S(G_n(x - v(i/n))) \cdot (G_n(x - v(i/n)))^{\alpha_0} \rightarrow 0$. Zusätzlich folgt wie bei v , dass

$\|\bar{\alpha}\|_\infty < \infty$, sodass $\limsup (G_n(x - v(i/n)))^{\bar{\alpha}(i/n)/\log n} \leq 1$ gleichmäßig in i . Diese gleichmäßigen Konvergenzen mit (2.24) verknüpft generieren

$$\begin{aligned} & \limsup_n \max_{1 \leq i \leq n} \mathbb{P}(X_{i,n} > u_0 - G_n x) \\ & \leq \limsup_n \|h\|_\infty o(1) \cdot (G_n(x - v(i/n)))^{\bar{\alpha}(i/n)/\log n} \\ & \leq \limsup_n \|h\|_\infty o(1) \cdot 1 = 0 \end{aligned}$$

und damit (2.24). \square

Das nächste Lemma dient dem Nachweis der zweiten hinreichenden Bedingung für die Poisson-Approximation und ist erneut die aufwendigere Komponente zu zeigen.

Lemma 2.3.9.

In der Situation aus Theorem 2.3.7 gilt

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - x G_n) = \int_0^1 h(y) \cdot \exp\left(-\frac{\bar{\alpha}(y)}{\alpha_0}\right) \cdot (x - v(y))_+^{\alpha_0} dy. \quad (2.25)$$

Beweis. Wir fixieren $x > 0, \varepsilon \in]0, x/2 \wedge 1[$ und definieren vorbereitend die Teilindexbereiche $I_n(c, d) := \{i \in \mathbb{N} : c \leq v(i/n) < d, 1 \leq i \leq n\}$.

Für $i \notin I_n(0, x - \varepsilon) \cup I_n(x - \varepsilon, x)$ gilt gemäß Definition der I_n , dass $v(i/n) \geq x$, sodass $u(i/n) = u_0 - G_n v(i/n) \leq u_0 - G_n x$ folgt. Da $u(i/n)$ gerade den Randpunkt von $X_{i,n}$ darstellt, gilt für solche i also $\mathbb{P}\{X_{i,n} > u_0 - x G_n\} = 0$.

Dies erlaubt uns die Summe über die Tails in zwei Teilsommen zu zerlegen:

$$\begin{aligned} & \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - x G_n) \\ & = \sum_{i \in I_n(0, x - \varepsilon)} \mathbb{P}(X_{i,n} > u_0 - x G_n) + \sum_{i \in I_n(x - \varepsilon, x)} \mathbb{P}(X_{i,n} > u_0 - x G_n) =: S_1 + S_2 \end{aligned} \quad (2.26)$$

und wir schreiben $\bar{\alpha}_m := \min \bar{\alpha}(y), \mu_\varepsilon(x) := \mathbb{A}(v^{-1}(]x - \varepsilon, x[)) \leq \mathbb{A}([0, 1]) = 1$, wobei \mathbb{A} das Lebesguemaß bezeichne.

Schritt 1: Im ersten Schritt zeigen wir, dass der S_1 -Anteil der Summe aus (2.26) asymptotisch vernachlässigbar ist.

Mittels Bedingung 2.3.6 und der in i gleichmäßigen Konvergenz von $G_n(x - v(i/n))$

gegen 0, siehe Beweis von Lemma 2.3.8, erhalten wir

$$\begin{aligned}
 & \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) \\
 &= \sum_{i \in I_n(0, x-\varepsilon)} h(i/n) \cdot S(G_n(x - v(i/n))) \left(G_n(x - v(i/n)) \right)^{\alpha_n(i/n)} \\
 &\leq \|h\|_\infty G_n^{\alpha_0} \sum_{i \in I_n(0, x-\varepsilon)} S(G_n(x - v(i/n))) \cdot G_n^{\bar{\alpha}(i/n)/\log n} \cdot (x - v(i/n))^{\alpha_n(i/n)}.
 \end{aligned} \tag{2.27}$$

Auf $I_n(x - \varepsilon)$ gilt $v(i/n) \in [x - \varepsilon, x]$, sodass $x - v(i/n) \in]0, \varepsilon] \subset [0, 1]$. Die letztere Menge ist eine kompakte Menge, sodass sich aus dem Satz von der lokal gleichmäßigen Konvergenz

$$\begin{aligned}
 \limsup_n \frac{S(G_n(x - v(i/n)))}{S(G_n)} &\leq \limsup_n \sup_{\lambda \in [0,1]} \frac{S(\lambda G_n)}{S(G_n)} \\
 &\leq \limsup_n \sup_{\lambda \in [0,1]} \left| \frac{S(\lambda G_n)}{S(G_n)} - 1 \right| + 1 \\
 &= 0 + 1 = 1
 \end{aligned}$$

ergibt, weswegen $S(G_n(x - v(i/n))) \leq (1 + \varepsilon)S(G_n)$ für große n gleichmäßig in i gilt. Außerdem erhalten wir aus $x - v(i/n) \in [0, \varepsilon]$ und $\bar{\alpha}(i/n)/\log n \geq 0$, dass $(x - v(i/n))^{\alpha_0 + \bar{\alpha}(i/n)/\log n} \leq \varepsilon^{\alpha_0}$. Mittels der eingeführten Notation können wir aufgrund von $G_n \rightarrow 0$, die Beziehung $G_n^{\bar{\alpha}(i/n)/\log n} \leq G_n^{\bar{\alpha}_m/\log n}$ folgern. Diese Tatsachen mit (2.27) und $S(G_n)G_n^{\alpha_0} = G(G_n) = 1/n$ verknüpft generieren

$$\sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) < \|h\|_\infty (1 + \varepsilon) G_n^{\bar{\alpha}_m/\log n} \varepsilon^{\alpha_0} \frac{1}{n} \sum_{i \in I_n(0, x-\varepsilon)} 1. \tag{2.28}$$

Die Summe $\sum_{i \in I_n(0, x-\varepsilon)} 1/n$ lässt sich als Riemann-Summe mit Grenzwert $\int \mathbb{1}(0 \leq y \leq 1, y \in v^{-1}([x - \varepsilon, x])) dy = \mathbb{A}(v^{-1}([x - \varepsilon, x])) = \mu_\varepsilon(x) \leq 1$ identifizieren, da nach Voraussetzung $v^{-1}([x - \varepsilon, x])$ eine maximal endliche Vereinigung von Intervallen ist.

Dies in Verbindung mit (2.28) und $\mu_\varepsilon(x) \leq 1$ ergibt

$$\sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) < \|h\|_\infty (1 + \varepsilon)^2 G_n^{\bar{\alpha}_m/\log n} \varepsilon^{\alpha_0}. \tag{2.29}$$

Somit bleibt lediglich der Ausdruck $G_n^{\bar{\alpha}_m/\log n}$ abzuschätzen. Dazu beachten wir, dass $G^{-1} \in RV_0(1/\alpha_0)$, weswegen $G_n > (1/n)^{2\alpha_0}$ für große n erfüllt ist und damit

$$G_n^{\bar{\alpha}_m/\log n} = \exp\left(\frac{\bar{\alpha}_m}{\log n} \log G_n\right) \leq \exp\left(\frac{2|\bar{\alpha}_m|}{\alpha_0} \cdot \frac{-\log(1/n)}{\log n}\right) \leq \exp\left(\frac{2|\bar{\alpha}_m|}{\alpha_0}\right)$$

unabhängig von n und ε gilt, was in Kombination mit (2.29) und der Beliebigkeit von ε die asymptotische Vernachlässigbarkeit der ersten Summe impliziert, da $\alpha_0 > 0$ vorausgesetzt wurde.

Schritt 2: In diesem Schritt zeigen wir

$$\lim_{n \rightarrow \infty} \sum_{i \in I_n(x-\varepsilon, x)} \mathbb{P}(X_{i,n} > u_0 - xG_n) = J(x), \quad (2.30)$$

woraus sich im dritten Schritt die Aussage einfach folgern lässt.

Genauer zeigen wir die obere Abschätzung für den Limes superior durch J und erhalten mittels eines Symmetrie-Arguments die untere Abschätzung für den Limes inferior.

Für große n und $i \in I_n(0, x - \varepsilon)$ gilt per Definition der I_n , dass $x - v(i/n) \in [\varepsilon, x]$, sodass wegen des Satzes von der lokal gleichmäßigen Konvergenz auf die langsam variierende Funktion S angewandt $S(G_n(x - v(i/n))) < (1 + \varepsilon) \cdot S(G_n)$ gleichmäßig in i hält. Die Voraussetzungen an $\bar{\alpha}$ implizieren die in i gleichmäßige Konvergenz $\bar{\alpha}(i/n)/\log n \rightarrow 0$, sodass sich $(x - v(i/n))^{\bar{\alpha}(i/n)/\log n} < 1 + \varepsilon$ gleichmäßig in i für große n deduzieren lässt. Daher erhalten wir mittels Bedingung 2.3.6

$$\begin{aligned} & \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) \\ &= \sum_{i \in I_n(0, x-\varepsilon)} h(i/n) S\left(G_n(x - v(i/n))\right) \left(G_n(x - v(i/n))\right)^{\alpha_0 + \bar{\alpha}/\log n} \\ &< (1 + \varepsilon) \cdot S(G_n) G_n^{\alpha_0} (1 + \varepsilon) \cdot \sum_{i \in I_n(0, x-\varepsilon)} h(i/n) G_n^{\bar{\alpha}(i/n)\log n} (x - v(i/n))^{\alpha_0} \\ &= (1 + \varepsilon)^2 1/n \cdot \sum_{i \in I_n(0, x-\varepsilon)} h(i/n) G_n^{\bar{\alpha}(i/n)\log n} (x - v(i/n))^{\alpha_0}. \end{aligned} \quad (2.31)$$

Die Funktion G^{-1} ist regulär variierend mit Index $1/\alpha_0$ in der Null, sodass wir sie als $G^{-1}(x) = x^{1/\alpha_0} \cdot L(x)$ mit einer geeigneten langsam variierenden Funktion L schreiben können, siehe Appendix A.2.8. Dies führt zu der Konvergenz

$$\begin{aligned} \left| \frac{\log G_n}{\log n} + \frac{1}{\alpha_0} \right| &= \left| \frac{\log(n^{-1/\alpha_0} \cdot L(1/n))}{\log n} + \frac{1}{\alpha_0} \right| \\ &= \left| \frac{\log L(1/n)}{\log n} \right| = \left| \frac{\log L(1/n)}{\log(1/n)} \right| \xrightarrow{n} 0, \end{aligned}$$

wobei Letzteres aus der langsamen Variation des L folgt. Dies nutzend ergibt sich nach einem Umschreiben

$$G_n^{\bar{\alpha}(i/n)/\log n} = \exp(\bar{\alpha}(i/n)/\log n \cdot \log G_n) < \exp(\|\bar{\alpha}\|_\infty \varepsilon - \bar{\alpha}(i/n)/\alpha_0),$$

was wiederum in (2.31) eingesetzt

$$\begin{aligned} & \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) \\ & < (1 + \varepsilon)^2 \exp(\varepsilon \|\bar{\alpha}\|_\infty) \cdot \sum_{i \in I_n(0, x-\varepsilon)} h(i/n) \exp\left(-\frac{\bar{\alpha}(i/n)}{\alpha_0}\right) (x - v(i/n))^{\alpha_0} \cdot \frac{1}{n} \end{aligned} \quad (2.32)$$

generiert. Der Integrationsbereich $[0, 1] \cap v^{-1}([0, x - \varepsilon])$ ist erneut nach Voraussetzung eine endliche Vereinigung von Intervallen und der Integrand $y \mapsto h(y) \cdot \exp(-\bar{\alpha}(y)/\alpha_0) \cdot (x - v(y))^{\alpha_0}$ als Komposition stetiger Abbildungen selbst stetig. Daher folgt mittels der Konvergenz von Riemann-Summen auf beschränkten Intervallen aus (2.32)

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) \\ & \leq (1 + o_\varepsilon(1)) \cdot \int_{[0,1] \cap v^{-1}([0, x-\varepsilon])} H(y) \cdot \exp\left(-\frac{\bar{\alpha}(y)}{\alpha_0}\right) (x - v(y))^{\alpha_0} dy \\ & \xrightarrow{\varepsilon \downarrow 0} \int_{[0,1] \cap v^{-1}([0, x])} H(y) \cdot \exp\left(-\frac{\bar{\alpha}(y)}{\alpha_0}\right) (x - v(y))^{\alpha_0} dy = J(x), \end{aligned} \quad (2.33)$$

wobei das Vertauschen von Grenzwert und Integral im vorletzten Schritt aus dem Satz von der dominierten Konvergenz resultiert.

Für die untere Schranke bemerken wir, dass im zweiten Schritt jede benutzte Abschätzung symmetrisch nach unten korrigierbar ist, sodass sich

$$\liminf_{n \rightarrow \infty} \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) \geq (1 + o_\varepsilon(1)) \cdot J(x) \xrightarrow{\varepsilon \downarrow 0} J(x) \quad (2.34)$$

ableitet.

Schritt 3: Zusammenfassend erhalten wir als obere Schranke für den Limes superior aus (2.26,) (2.33) und aus Schritt 1

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - xG_n) \\ & = \limsup_{n \rightarrow \infty} \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) + \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) \\ & \leq (1 + o_\varepsilon(1)) \cdot J(x) + o_\varepsilon(1) \xrightarrow{\varepsilon \downarrow 0} J(x); \end{aligned}$$

und mittels (2.3)

$$\begin{aligned}
 & \liminf_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - xG_n) \\
 &= \liminf_{n \rightarrow \infty} \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) + \sum_{i \in I_n(0, x-\varepsilon)} \mathbb{P}(X_{i,n} > u_0 - xG_n) \\
 &\geq (1 + o_\varepsilon(1)) \cdot J + 0 \xrightarrow{\varepsilon \downarrow 0} J(x).
 \end{aligned}$$

Beide Schranken vereint validieren die Gleichung in (2.25). \square

Die vorangegangenen Resultate lassen uns nun die zentrale Verteilungskonvergenz deduzieren.

Theorem 2.3.7 (Zweiter Grenzwertsatz für Maxima nicht-stationärer Zufallsvariablen).
 Unter den Bedingungen 2.3.1, 2.3.2, 2.3.4, 2.3.5 und 2.3.6 gilt mit der Bezeichnung
 $M_n := \max \{ X_{i,n} : 1 \leq i \leq n \}$

$$\frac{M_n - u_0}{G_n} \xrightarrow{d} G, \quad (2.35)$$

wobei $G(x) = \begin{cases} \exp(-J(-x)) & , x < 0 \\ 1 & , x \geq 0 \end{cases}$ und

$$\begin{aligned}
 J(x) &:= \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(X_{i,n} > u_0 - xG_n) \\
 &= \int_0^1 h(y) \cdot \exp\left(-\frac{\bar{\alpha}(y)}{\alpha_0}\right) \cdot (x - v(y))_+^{\alpha_0} dy
 \end{aligned} \quad (2.36)$$

Beweis. Die Beweisstrategie besteht daraus, die Voraussetzung der Poisson-Approximation aus Appendix A.2.5 nachzuweisen. Zu diesem Zwecke setzen wir

$$B_{i,n} := \mathbb{1}(X_{i,n} > u_0 - xG_n), \quad p_{i,n} := \mathbb{P}(B_{i,n} = 1) \text{ und } S_n := \sum_{i=1}^n B_{i,n}.$$

Aus den Lemmata 2.3.8, 2.3.9 und der Poisson-Approximation folgt dann für $x > 0$

$$\begin{aligned}
 \mathbb{P}(M_n \leq u_0 - xG_n) &= \mathbb{P}(S_n = 0) \\
 &\xrightarrow{n \rightarrow \infty} \text{Poi}\left(\lim_{n \rightarrow \infty} \sum_{i=1}^n p_{i,n}\right) \{0\} = \text{Poi}(J(x)) \{0\} \\
 &= \exp(-J(x)).
 \end{aligned}$$

Außerdem folgt aus der Tatsache $u_{i,n} \leq u_0$, dass $\mathbb{P}(M_n \leq u_0 + xG_n) = 1$. \square

Somit findet eine Verteilungskonvergenz von einer normalisierten Maximumsfolge gegen eine nicht-degenerierte Zufallsvariable statt und im Lichte des Satzes von Fisher-Tippet stellt sich natürlicherweise die Frage, ob sich G als Spezialfall der GEV-Verteilung, wie in Theorem 2.2.9, identifizieren lässt.

Bemerkung 2.3.10.

1. Der Form von G entnehmen wir, dass der rechte Randpunkt endlich ist, sodass lediglich die Identifizierung als Weibull-Typ möglich ist.
2. Der rechte Randpunkt von G ist nicht zwangsläufig Null, sondern $-v_m$ mit $v_m := \min_{y \in [0,1]} v(y) \in \mathbb{R}_0^+$. Damit gilt

$$G(x) = \begin{cases} \exp(-J(-x)) & , x < -v_m \\ 1 & , x \geq -v_m. \end{cases} \quad (2.37)$$

Beweis. 1. Die Endlichkeit folgt aus dem zweiten Teil der Bemerkung und der potentielle Typ aus der Tatsache, dass sowohl die Gumbel- als auch die Fréchet-Verteilung einen nicht-endlichen rechten Randpunkt besitzen.

2. Der Form von J , siehe (2.25), entnehmen wir, dass J isoton in x ist. Angesichts der Stetigkeit von v und der Positivität der Faktoren vor dem Indikatorexpression folgt für $x > v_m$, dass $J(x) > 0$. Dies bedeutet jedoch, dass für $x < -v_m$ der Ausdruck $J(-x)$ positiv und damit $\exp(-J(-x)) < 1$ ist. Außerdem gilt $\{v_m - v(y)\}_+^\alpha \equiv 0$, wodurch $G(-v_m) = 1$ folgt. \square

Die der Bemerkung vorangegangene Frage lässt sich negativ beantworten. Bevor wir dies in voller Allgemeinheit zeigen, untersuchen wir einen Fall mit derartigen Wahlen, dass das Integral J explizit wird.

Wir setzen $h(\cdot) \equiv h > 0$, $\bar{\alpha}(\cdot) \equiv 0$, $\mu := u_0 = 0$, $v(y) := |y - 1/2|$ und schreiben fortan $\alpha := \alpha_0$. Diese Wahlen erfüllen die am Anfang des Abschnittes formulierten Forderungen. Verschiebung um $1/2$ und Symmetrie generieren

$$\begin{aligned} J(x) &= \int_0^1 h \cdot \{x - |y - 1/2|\}_+^\alpha dy = h \cdot \int_{-1/2}^{1/2} \{x - |y|\}_+^\alpha dy \\ &= 2h \cdot \begin{cases} \int_0^{1/2} (x - y)^\alpha & , x \geq 1/2 \\ \int_0^x (x - y)^\alpha & , x < 1/2 \end{cases} = \frac{2h}{\alpha + 1} \cdot \begin{cases} x^{\alpha+1} - (x - 1/2)^{\alpha+1} & , x \geq 1/2 \\ x^{\alpha+1} & , x < 1/2 \end{cases}. \end{aligned}$$

Damit erhalten wir

$$G(x) = \begin{cases} \exp\left\{-\frac{2h}{\alpha+1} ((-x)^{\alpha+1} - (-x - 1/2)^{\alpha+1})\right\} & , x \leq -1/2 \\ \exp\left\{-\frac{2h}{\alpha+1} (-x)^{\alpha+1}\right\} & , -1/2 < x < 0, \\ 1 & , 0 \leq x \end{cases} \quad (2.38)$$

wobei wir die Abhängigkeit der Verteilungsfunktion G von den Wahlen $h, \alpha > 0$ zunächst in der Notation unterdrücken. Man beachte, dass G im rechten Tail exakt einem Wei $(\alpha + 1)$ -Typ entspricht. Im linken Tail konvergiert das Argument der Exponen-

tionalfunktion für $x \rightarrow -\infty$ wie der Ausdruck $-x^\alpha$; das sieht man durch eine Anwendung des Mittelwertsatzes ein. Dies zeigt die Nähe zu einer Weibull-Verteilung.

Lemma 2.3.11.

Verteilungsfunktionen G von der Gestalt aus (2.38) sind keine Extremwertverteilungen.⁵

Beweis. Der Gestalt von G lesen wir instantan ab, dass der rechte Randpunkt 0 ist, sodass nach dem Fisher-Tippet Theorem lediglich eine Identifikation als Weibull-Typ mit der Null als rechte Randpunkt möglich wäre. Nehmen wir also nun an, dass $\sigma, \beta > 0$ mit $G(x) = \text{Wei}(\beta)(\sigma x)$ existieren. Der rechte Tail von G ist exakt von Weibull-Gestalt, womit wir $\beta = \alpha + 1$ schließen und für $x \leq -1/2$ die Beziehung

$$\exp\left\{-\frac{2h}{\alpha+1} \left((-x)^{\alpha+1} - (-x-1/2)^{\alpha+1}\right)\right\} = \exp\left(-(-\sigma x)^\beta\right)$$

etablieren können.

Dies ist gleichbedeutend zu $2h/(\alpha+1) \cdot ((x^{\alpha+1} - (x-1/2)^{\alpha+1}) = \sigma^\beta x^{\alpha+1}$, was seinerseits $1 - \sigma^\beta/(2h) \cdot (\alpha+1) \equiv (1 - 1/(2x))^{\alpha+1}$ für $x \geq 1/2$ bedeutet. Die linke Seite ist konstant in x , wohingegen die rechte es nicht ist, ein Widerspruch. Damit kann G keiner Extremwertverteilung entsprechen. \square

Um eine allgemeine Variante von Lemma 2.3.11 beweisen zu können, benötigen wir ein Hilfsresultat, in welchem folgende Notation nützlich ist.

Definition 2.3.12 (\mathcal{G}).

Wir bezeichnen mit \mathcal{G} die Klasse von Verteilungsfunktionen, welche als Grenzwert in Theorem 2.3.7 möglich sind. Das heißt genauer, dass $G \in \mathcal{G}$ die Form (2.37) hat. Weiterhin hängt G damit von $\alpha > 0$ und stetigen Abbildungen $\bar{\alpha}, h, v: [0, 1] \rightarrow \mathbb{R}$ mit $h(\cdot) > 0$ und $v(\cdot) \geq 0$ ab, wobei $v^{-1}([x, y])$ für alle $0 \leq x < y$ als eine endliche Vereinigung von Intervallen darstellbar ist.

Lemma 2.3.13.

Sei $G \in \mathcal{G}$ mit zugehöriger Funktion v . Dann ist die Verteilungsfunktion $\tilde{G}(\cdot) := G(\cdot - \mu)$ für Konstanten $\mu \leq v_m$ auch in \mathcal{G} enthalten, wobei $v_m := \min_{y \in [0, 1]} v(y)$.

Beweis. Sei $G \in \mathcal{G}$ mit zugehörigem $v(\cdot)$. Nach Bemerkung 2.3.10 ist der rechte Randpunkt von G der Wert $-v_m$. Daher lässt sich der rechte Randpunkt von \tilde{G} als $-v_m + \mu$

⁵ Eine alternative Beweismöglichkeit ist es, max-Stabilität von G zu widerlegen. In Lemma 3.1.4 untersuchen wir die Regularität im Punkt des Strukturwechsels. Es stellt sich heraus, dass G darin nicht glatt ist, sodass wir drei verschiedene Beweismöglichkeiten von Lemma 2.3.11 kennengelernt und zwei davon vorgeführt haben.

identifizieren. Wir schreiben die Verteilungsfunktion \tilde{G} wie folgt um

$$\begin{aligned}\tilde{G}(x) &= \begin{cases} \exp(-J(-x + \mu)) & , x - \mu < -v_m \\ 1 & , x - \mu \geq -v_m. \end{cases} \\ &= \begin{cases} \exp\left(-\int_0^1 \gamma(y) \cdot \{-x - (v(y) - \mu)\}_+^\alpha dy\right) & , x - \mu < -v_m \\ 1 & , x - \mu \geq -v_m. \end{cases} \\ &= \begin{cases} \exp\left(-\int_0^1 \gamma(y) \cdot \{-x - \tilde{v}\}_+^\alpha dy\right) & , x < -\tilde{v}_m \\ 1 & , x \geq -\tilde{v}_m, \end{cases}\end{aligned}$$

wobei $\tilde{v}(y) := v(y) - \mu$, $\gamma(y) := h(y) \cdot \exp(-\bar{\alpha}(y)/\alpha)$. Nach Voraussetzung ist $\tilde{v}(\cdot) \geq 0$ und da eine Translation die Bedingungen der Stetigkeit und Oszillationsbegrenzung aus 2.3.2 nicht beeinflusst, ist \tilde{v} eine valide Wahl, sodass $\tilde{G} \in \mathcal{G}$ folgt. \square

Mithilfe dieses Resultats beweisen wir den folgenden Satz.

Satz 2.3.14.

Für alle Grenzverteilungsfunktionen $G \in \mathcal{G}$ gilt: ist die zugehörige Funktion v nicht die konstante Null-Abbildung, so ist G keine Extremwertverteilung.

Beweis. Nach Bemerkung 2.3.10 bleibt zu zeigen, dass für festes $G \in \mathcal{G}$ mit $v(\cdot) \neq 0$ keine $\sigma, \beta > 0$ mit $G \stackrel{d}{=} 1/\sigma \cdot \text{Wei}(\beta) - v_m$ existieren.

Diese Aussage zeigen wir zunächst für den Fall $0 = v_m = \min v(\cdot)$ und anschließend mittels des vorangegangenen Lemmas für den Fall $v_m > 0$.

Schritt 1: Angenommen es gibt $\beta, \sigma > 0$ mit $G \stackrel{d}{=} 1/\sigma \cdot \text{Wei}(\beta)$. Das heißt $\exp(J(-x)) = \exp(-(\sigma x)^\beta)$ für $x < 0$ und damit $J(x) = (\sigma x)^\beta$ für alle $x > 0$. Die Funktion v ist stetig mit kompaktem Definitionsbereich, sodass $\|v\|_\infty$ existiert. Betrachte von nun an $x > \|v\|_\infty$, sodass $\int_0^1 \gamma(y) \{x - v(y)\}^\alpha dy = (\sigma x)^\beta$ hält. Damit muss aber insbesondere die Gleichung

$$\sigma^\beta \equiv x^{\alpha-\beta} \int_0^1 \gamma(y) \cdot \{1 - v(y)/x\}^\alpha dy \quad (x > \|v\|_\infty) \quad (2.39)$$

gelten. Mit $f_x(y) := \gamma(y) \cdot \{1 - v(y)/x\}^\alpha$ gilt $f_x(\cdot) \geq 0$ und die Funktionenfolge $(f_x)_{x > \|v\|_\infty}$ ist isoton in x . Wegen der Beschränktheit von v gilt die punktweise Konvergenz $f_x \xrightarrow{x \rightarrow \infty} \gamma$, sodass mit dem Satz von Beppo Levi $\int_0^1 f_x(y) dy \xrightarrow{x \rightarrow \infty} \|\gamma\|_1 > 0$ folgt, wobei die untere Schranke aufgrund der Positivität von γ gilt. Dies in Verbindung mit (2.39) generiert

$$\sigma^\beta = \|\gamma\|_1 \cdot \lim_{x \rightarrow \infty} x^{\alpha-\beta} = \begin{cases} \infty & , \alpha > \beta \\ 1 & , \alpha = \beta \\ 0 & , \alpha < \beta \end{cases}$$

was in den Fällen $\alpha > \beta$ und $\alpha < \beta$ zu einem Widerspruch führt.

Damit bleibt noch der Fall $\alpha = \beta$ zu untersuchen. Mittels (2.39) erhalten wir $\sigma^\alpha \equiv \int_0^1 f_x(y) dy$ für große x und wieder mittels des Satzes von Beppo Levi die Konvergenz gegen $\|\gamma\|_1$, das heißt für große x

$$\int_0^1 \gamma(y) dy = \int_0^1 \gamma(y) \cdot \{1 - v(y)/(2 \cdot \|v\|_\infty)\}^\alpha dy.$$

Die Abbildung $\gamma(\cdot)$ ist positiv und $v(\cdot)/(2\|v\|_\infty)$ gleichmäßig kleiner als 1, sodass $1 = \{1 - v(\cdot)/(2\|v\|_\infty)\}^\alpha$ fast-überall und damit $v = 0$ fast-überall folgt. Da v stetig ist, muss v schon der konstanten Null-Abbildung entsprechen, was der Annahme des Satzes widerspricht.

Schritt 2: In diesem Schritt betrachten wir diejenigen v mit $v_m > 0$. Erneut nehmen wir an, dass $\sigma, \beta > 0$ mit $G \stackrel{d}{=} 1/\sigma \text{Wei}(\beta) - v_m$ existieren. Wir setzen $\tilde{G}(x) := G(x - v_m)$ und bemerken, dass nach Lemma 2.3.12 $\tilde{G} \in \mathcal{G}$ gilt, da trivialerweise $v_m \leq v_m$ erfüllt ist. Jedoch entspricht der rechte Randpunkt von \tilde{G} der Null und damit folgt insbesondere $\tilde{v}_m = 0$. Damit haben wir $\sigma, \beta > 0$ gefunden, sodass für $\tilde{G} \in \mathcal{G}$ die Verteilungsgleichheit $\tilde{G} \stackrel{d}{=} 1/\sigma \text{Wei}(\beta)$ erfüllt ist, ein Widerspruch zu Schritt 1. \square

Die speziellen Verteilungsfunktionen aus (2.38) bilden eine Subklasse von \mathcal{G} , welche einen rechten Tail besitzt, der exakt und damit insbesondere asymptotisch von Weibull-Gestalt ist. Eine sich daraus ergebende Fragestellung ist, ob sich dieses Resultat, wie etwa die Disjunktheit zu der GEV-Klasse, auf die gesamte Klasse \mathcal{G} erweitern lässt. Diese Frage werden wir im Folgenden negativ beantworten. Dabei werden wir ein konkretes Gegenbeispiel angeben und anschließend ausgehend von diesem Beispiel eine Klasse identifizieren, welche gerade nicht die oben genannte Tail-Asymptotik erfüllt.

Beispiel 2.3.15.

Setze $\bar{\alpha}(\cdot) \equiv 0$, $h(\cdot) \equiv 1$, $\alpha = 1$, und $v(y) := \mathbb{1}(y > 0) \cdot \exp(-1/y) \in \mathcal{C}([0, 1])$. Dann besitzt G für diese konkreten Wahlen den rechten Randpunkt 0 und für alle $\sigma, \beta > 0$ gilt

$$\lim_{x \uparrow 0} \frac{1 - G(x)}{1 - \text{Wei}(\beta)(\sigma x)} = \begin{cases} \infty & , \beta > 1 \\ 0 & , \beta \leq 1 \end{cases}. \quad (2.40)$$

Insbesondere ist der Tail von G nicht asymptotisch äquivalent zu dem Tail einer Weibull-Verteilung und damit auch nicht zu einer GEV-Verteilung.

Beweis. Der Bemerkung 2.3.10 entnehmen wir, dass v_m und damit 0 der rechte Randpunkt von G ist.

Wir rechnen

$$I(x) = \int_0^1 \{x - \exp(-1/y)\}_+ dy = \int_0^{1 \wedge -1/\log x} x - \exp(-1/y) dy.$$

Für $x \downarrow 0$ gilt $1 \wedge -1/\log x = -1/\log x$, sodass aus Obigem

$$J(x) = -\frac{x}{\log x} - \int_0^{-1/\log x} \exp(-1/y) dy$$

folgt, wobei $J(0) = 0$. Damit ist J differenzierbar mit Ableitung $J'(x) = (1 - \log x)/\log^2(x) - 1/\log^2(x) = -1/\log x = 1/|\log x|$, wobei für den zweiten Summanden der Hauptsatz der Differential- und Integralrechnung angewandt wurde.

Eine Anwendung einer Regel von L'Hospital generiert

$$\begin{aligned} \lim_{x \uparrow 0} \frac{1 - G(x)}{1 - \text{Wei}(\beta)(\sigma x)} &= \lim_{x \downarrow 0} \frac{1 - \exp(-J(x))}{1 - \exp(-(\sigma x)^\beta)} \\ &= \lim_{x \downarrow 0} \frac{\exp(-J(x)) \cdot J'(x)}{\exp(-(\sigma x)^\beta) \cdot \sigma^\beta \beta x^{\beta-1}} = \lim_{x \downarrow 0} \frac{J'(x)}{\sigma^\beta \beta x^{\beta-1}} \\ &= \lim_{x \downarrow 0} \frac{1}{|\log x| \cdot x^{\beta-1}} \cdot \frac{1}{\beta \cdot \sigma^\beta} = \begin{cases} \infty & , \beta \geq 1 \\ 0 & , \beta \leq 1 \end{cases} \end{aligned}$$

womit die Behauptung gezeigt ist. □

Wie vor dem letzten Resultat angesprochen, lässt sich die Klasse erheblich vergrößern.

Satz 2.3.16.

Für Verteilungen G der Klasse \mathcal{G} , welche die zugehörige Funktion $v(y) := \mathbb{1}(y > 0) \cdot \exp(-1/y)$ besitzen, gilt Gleichung (2.40), das heißt

$$\lim_{x \uparrow 0} \frac{1 - G(x)}{1 - \text{Wei}(\beta)(\sigma x)} = \begin{cases} \infty & , \beta > \alpha \\ 0 & , \beta \leq \alpha \end{cases} \quad (2.41)$$

für $\beta > 0$ und $\alpha > 0$ als Parameter der Verteilung G .

Beweis. Die Beweisstrategie folgt der aus dem vorigen Beispiel. Das heißt, wir etablieren eine Leibniz-Regel, um dann mittels einer Regel von L'Hospital die Grenzwertaussage zeigen zu können. Der Beweis wird sich hier als aufwendiger herausstellen, da für $\alpha < 1$ Singularitäten innerhalb eines Integrals auftreten. Ohne Einschränkung sei der rechte Randpunkt beider Verteilungen als Null vorausgesetzt.

Schritt 1: Es gilt für die Tails

$$\begin{aligned} \frac{1 - G(x)}{1 - \text{Wei}(0, c, \beta)(x)} &= \frac{1 - \exp(-J(-x))}{1 - \exp(-(-cx)^\beta)} = \frac{1 - \exp(-J(x))}{1 - \exp(-(cx)^\beta)} \\ &\approx \frac{\exp(-J(x)) \cdot J'(x)}{\exp(-(cx)^\beta) \cdot \beta c^\beta x^{\beta-1}} \approx \frac{1}{\beta c^\beta} \cdot \frac{J'(x)}{x^{\beta-1}} \end{aligned} \quad (2.42)$$

für x nahe Null, falls $J(\cdot)$ in einer Rechts-Umgebung von 0 differenzierbar ist.

Schritt 2: In diesem Schritt zeigen wir die Gleichung

$$J'(x) = \int_0^{-1/\log x} \alpha \cdot \gamma(y) (x - \exp(-1/y))^{\alpha-1} dy \quad (2.43)$$

für x nahe Null mit $J(\cdot)$ aus (2.25) und $\gamma(y) = h(y) \cdot \exp(-\bar{\alpha}(y)/\alpha)$. Das heißt die Ableitung nach x darf in das Integral hineingezogen werden, eine Leibniz-Regel.

Sei $h > 0$, es gilt für x nahe Null $x \geq \exp(-1/y)$ genau dann, wenn $y \leq -1/\log x$, sodass

$$\begin{aligned} & \frac{J(x+h) - J(x)}{h} \\ &= \int_0^{-1/\log x} \gamma(y) \cdot \frac{(x+h - \exp(-1/y))^\alpha - (x - \exp(-1/y))^\alpha}{h} dy + \\ & \quad \int_{-1/\log x}^{-1/\log(x+h)} \gamma(y) \cdot \frac{(x+h - \exp(-1/y))^\alpha}{h} dy \\ &=: T_h(x) + R_h. \end{aligned}$$

Zunächst zeigen wir die Konvergenz von R_h gegen Null für $h \downarrow 0$:

$$\begin{aligned} |R_h| &= R_h \leq \|\gamma(\cdot)\|_\infty \cdot h^{\alpha-1} \cdot \left(\frac{1}{\log x} - \frac{1}{\log(x+h)} \right) \\ &= \frac{\|\gamma(\cdot)\|_\infty}{\log x \cdot \log(x+h)} \cdot \frac{\log(x+h) - \log x}{h^{1-\alpha}} \\ &\approx \frac{\|\gamma(\cdot)\|_\infty}{\log^2 x} \cdot \frac{1/(x+h)}{(1-\alpha) \cdot h^{-\alpha}} \\ &\xrightarrow{h \downarrow 0} 0. \end{aligned}$$

Mit dem Mittelwertsatz der Differentialrechnung erhalten wir für ein $\theta := \theta_h \in [0, 1]$

$$\begin{aligned} T_h(x) &= \int_0^{-1/\log x} \alpha \gamma(y) \cdot (x + \theta h - \exp(-1/y))^{\alpha-1} dy \\ &=: \int_0^{-1/\log x} \alpha \gamma(y) \cdot f_h(y) dy. \end{aligned}$$

Offensichtlich gilt die punktweise Konvergenz von $f_h(x) \rightarrow (x - \exp(-1/y))^{\alpha-1}$. Es ist also noch eine (in h) gleichmäßige integrierbare Schranke anzugeben.

Falls $\alpha \geq 1$, so gilt $f_h(y) \leq (x + 1 - \exp(-1/y))^{\alpha-1} \in L^1([0, -1/\log x])$.

Falls $\alpha < 1$, so gilt $f_h(y) \leq (x - \exp(-1/y))^{\alpha-1}$, wobei noch zu zeigen ist, dass letzteres Integral integrierbar ist.

Mittels der Substitution $y = -1/\log u$ erhalten wir

$$\int_0^{-1/\log x} (x - \exp(-1/y))^{\alpha-1} dy = \int_0^x (x-y)^{\alpha-1} \cdot \frac{1}{y \cdot \log^2 y} dy =: I.$$

Da ohne Einschränkung $x < 1$ angenommen werden kann, müssen wir lediglich die Singularitäten in 0 und x untersuchen. Dafür sei $0 < \varepsilon < x$ und wir dreiteilen das Integral I :

$$I = \int_0^\varepsilon (x-y)^{\alpha-1} \cdot \frac{1}{y \cdot \log^2 y} dy + c_{x,\varepsilon} + \int_{x-\varepsilon}^x (x-y)^{\alpha-1} \cdot \frac{1}{y \cdot \log^2 y} dy =: I_1 + c_{x,\varepsilon} + I_2.$$

Für die Behandlung des ersten Integrals nutzen wir die Transformation $u = \log y$, sodass

$$I_1 \leq (x-\varepsilon)^{\alpha-1} \cdot \int_0^\varepsilon \frac{1}{y \log^2 y} dy = d_{x,\varepsilon} \cdot \int_{-\infty}^{\log \varepsilon} \frac{1}{y^2} dy = \frac{-d_{x,\varepsilon}}{\log \varepsilon} < \infty.$$

Bezüglich des zweiten Integrals gilt

$$I_2 \leq \frac{1}{(x-\varepsilon) \cdot \log^2(x-\varepsilon)} \cdot \int_{x-\varepsilon}^x (x-y)^{\alpha-1} dy = e_{x,\varepsilon} \left[\frac{(x-y)^\alpha}{-\alpha} \right]_{x-\varepsilon}^x = e_{x,\varepsilon} \cdot \frac{\varepsilon^\alpha}{\alpha} < \infty.$$

Damit ist I integrierbar, da der innere Teil $c := \int_{[\varepsilon, x-\varepsilon]} (x-y)^{\alpha-1} \cdot 1/(y \cdot \log^2 y) dy$ als Integral einer stetigen Funktion über ein Kompaktum endlich ist.

Damit ist auch für den Fall $\alpha < 1$ die Existenz einer integrierbaren Majorante bewiesen, sodass aus dem Satz von der dominierten Konvergenz

$$T_h(x) \xrightarrow{h \downarrow 0} \int_0^{-1/\log x} \alpha \gamma(y) \cdot (x - \exp(-1/y))^{\alpha-1} dy$$

folgt, wobei $\gamma(\cdot)$ bei der Schrankenfindung vernachlässigt werden durfte, da es als stetige Abbildung auf einem Kompaktum unabhängig von h durch ihr Maximum majorisiert werden kann.

Analog kann man für den Fall $h < 0$ verfahren, sodass wir insgesamt die Zielgleichung (2.43) bewiesen haben.

Schritt 3: Setzen wir die Gleichung (2.43) in die anfängliche Gleichung (2.42) ein, so folgt für $\beta \leq \alpha$ und $\alpha \geq 1$

$$\begin{aligned} \frac{J'(x)}{x^{\beta-1}} &\leq \alpha \|\gamma(\cdot)\|_\infty x^{\alpha-\beta} \cdot \int_0^{-1/\log x} \left(1 - \frac{\exp(-1/y)}{x}\right)^{\alpha-1} dy \\ &\lesssim \frac{x^{\alpha-\beta}}{-\log x} \leq \frac{-1}{\log x} \xrightarrow{x \downarrow 0} 0. \end{aligned}$$

Im Fall $\beta \leq \alpha$ und $\alpha < 1$ teilen wir das Integral auf:

$$\begin{aligned}
 \frac{J'(x)}{x^{\beta-1}} &\lesssim x^{\alpha-\beta} \cdot \int_0^{-1/\log x} \left(1 - \frac{\exp(-1/y)}{x}\right)^{\alpha-1} dy \\
 &= x^{\alpha-\beta} \cdot \int_0^x \left(1 - \frac{y}{x}\right)^{\alpha-1} \cdot \frac{1}{y \log^2 y} dy \\
 &= x^{\alpha-\beta} \cdot \left\{ \int_0^{x/2} \left(1 - \frac{y}{x}\right)^{\alpha-1} \cdot \frac{1}{y \log^2 y} dy + \int_{x/2}^x \left(1 - \frac{y}{x}\right)^{\alpha-1} \cdot \frac{1}{y \log^2 y} dy \right\} \\
 &=: x^{\alpha-\beta} \cdot (J_1 + J_2).
 \end{aligned} \tag{2.44}$$

Das erste Integral lässt sich mit der bekannten Substitution $u = \log y$ behandeln:

$$J_1 \leq \frac{1}{2^{\alpha-1}} \cdot \int_{-\infty}^{\log(x/2)} \frac{1}{y^2} dy = \frac{-1}{2^{\alpha-1} \cdot \log(x/2)} \xrightarrow{x \downarrow 0} 0.$$

Für das zweite Integral gilt

$$J_2 \leq \frac{2}{x \log^2(x/2)} \cdot \int_{x/2}^x \left(1 - \frac{y}{x}\right)^{\alpha-1} dy = \frac{2}{x \log^2(x/2)} \cdot \frac{-x}{\alpha 2^\alpha} \xrightarrow{x \downarrow 0} 0.$$

Beide Konvergenzen in die Gleichung (2.44) eingesetzt ergeben den ersten der Teil der Behauptung in (2.41).

Im Fall $\beta > \alpha$ und $\alpha < 1$ lässt sich wie folgt argumentieren:

$$\begin{aligned}
 \frac{J'(x)}{x^{\beta-1}} &\geq x^{\alpha-\beta} \cdot \int_{-1/\log(x/2)}^{-1/\log(x)} \left(1 - \frac{\exp(-1/y)}{x}\right)^{\alpha-1} dy \\
 &\geq x^{\alpha-\beta} \cdot \frac{1}{2^{\alpha-1}} \cdot \frac{\log x - \log(x/2)}{\log x \log(x/2)} \\
 &\sim \frac{\log(2)}{2^{\alpha-1}} \cdot \frac{x^{\alpha-\beta}}{\log x \log(x/2)} \xrightarrow{x \downarrow 0} \infty.
 \end{aligned}$$

Der Fall $\beta > \alpha$ und $\alpha \geq 1$ funktioniert ähnlich, sodass aus den Konvergenzen gegen unendlich in (2.42) eingesetzt die zweite Behauptung in (2.41) folgt und der Beweis abgeschlossen ist. \square

Zum Abschluss des Kapitels geben wir wie in der ersten Asymptotik ein starkes Gesetz an.

Satz 2.3.17 (Starkes Gesetz für die zweite Asymptotik).

In der Situation aus Theorem 2.3.7 gilt $M_n \xrightarrow{f.s.} u_0$.

Beweis. Der Beweis kann mittels ähnlicher Ansätze und Ideen wie der von Satz 2.2.14 geführt werden, sodass wir ihn an dieser Stelle nicht ausführen. \diamond

3 Statistische Modellierung

In diesem Kapitel untersuchen wir eine konkrete Subklasse der als Grenzwert in 2.3.7 auftretenden Verteilungen, deren Verteilungsfunktionen eine explizite Form annehmen und welche bereits aus 2.38 bekannt ist. Weiterhin umreißen wir die Block Maxima Methode im Fall von unabhängig und identisch verteilten Zufallsvariablen. Wir diskutieren sie für den Fall von nicht identisch aber unabhängig verteilter Beobachtungen und im Hinblick auf den Grenzwertsatz in Theorem 2.3.7.

3.1 Das explizite Stein'sche Modell

Wie in (2.38) gezeigt, geht aus den Wahlen von $v(y) := |y - 1/2|$, $u = 0$ und Reduktion der Funktionen auf Konstanten die Verteilungsfunktion

$$G_{\alpha,h}(x) := \begin{cases} \exp\left\{-\frac{2h}{\alpha+1}((-x)^{\alpha+1} - (-x - 1/2)^{\alpha+1})\right\} & , x \leq -1/2 \\ \exp\left\{-\frac{2h}{\alpha+1}(-x)^{\alpha+1}\right\} & , -1/2 < x < 0, \\ 1 & , 0 \leq x \end{cases}$$

und die Wahrscheinlichkeitsdichte

$$g_{\alpha,h}(x) := \begin{cases} 2h((-x)^{\alpha} - (-x - 1/2)^{\alpha}) \exp\left\{-\frac{2h}{\alpha+1}((-x)^{\alpha+1} - (-x - 1/2)^{\alpha+1})\right\} & , x \leq -1/2 \\ 2h(-x)^{\alpha} \cdot \exp\left\{-\frac{2h}{\alpha+1}(-x)^{\alpha+1}\right\} & , -1/2 < x < 0, \\ 0 & , 0 \leq x \end{cases}$$

hervor. Ein Einführen von Lokations- und Skalenparameter μ beziehungsweise σ resultiert in der Verteilungsfunktion

$$G_{\theta} := G_{\alpha,h,\mu,\sigma}(x) := \begin{cases} \exp\left\{-\frac{2h}{\alpha+1}\left(\left(\frac{\mu-x}{\sigma}\right)^{\alpha+1} - \left(\frac{\mu-x}{\sigma} - 1/2\right)^{\alpha+1}\right)\right\} & , x \leq -\frac{\sigma}{2} + \mu \\ \exp\left\{-\frac{2h}{\alpha+1}\left(\frac{\mu-x}{\sigma}\right)^{\alpha+1}\right\} & , -\frac{\sigma}{2} + \mu < x < \mu, \\ 1 & , \mu \leq x \end{cases} \quad (3.1)$$

und der Dichte $g_{\theta} := g_{\alpha,h,\mu,\sigma}(x) = g_{\alpha,h}((x - \mu)/\sigma)/\sigma$, wobei $\theta := (\alpha, h, \mu, \sigma) \in (\mathbb{R}^+)^2 \times \mathbb{R} \times \mathbb{R}^+ =: \Theta_E$

In Abbildung 3 sieht man für verschiedene Parameter α und h die Dichten $g_{\alpha,h}$ abgetragen. Es ist auffällig, dass die Dichten für $\alpha \leq 1$ im Punkt der Strukturänderung $x = -1/2$ nicht glatt verlaufen.

Definition 3.1.1.

Mit \mathcal{E}_E bezeichnen wir das aus X_1, \dots, X_N unabhängig und identisch nach G_{θ} aus (3.1) verteilt hervorgehende statistische Modell, wobei $(\alpha, h, \mu, \sigma) \in \Theta_E$ unbekannte

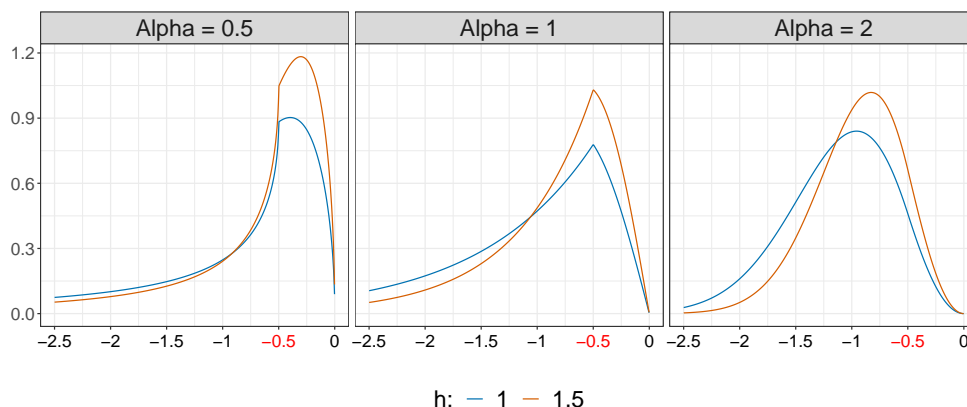


Abbildung 3: Dichten der $G_{\alpha,h}$ -Verteilung für verschiedene Formparameter α und h .

Parameter sind. Das heißt $\mathcal{X} = \mathbb{R}^N$, $\mathcal{B} = \mathcal{B}(\mathbb{R}^N)$ und $\mathcal{P} = \{G_{\theta}^{\otimes N} : \theta \in \Theta_E\}$. Dabei gilt $N \in \mathbb{N}_{\geq 1}$ und wir nennen das Modell **explizites Stein'sche Modell**.

Betrachtet man den rechten Strukturbereich der Verteilungsfunktion aus (3.1), so fällt auf, dass der Formparameter h zu einem Skalen- beziehungsweise transformierten Ratenparameter degeneriert. Wir werden im nächsten Lemma sehen, dass das Modell \mathcal{E}_E identifizierbar ist, woraus wir folgern können, dass h im linken Strukturbereich keine alleinige Skalierungsfunktion innehat. Das Verhalten im rechten Tail führt jedoch dazu, dass eine gewisse Unschärfe in der Schätzung von h zu erwarten ist. Dieses Modell beziehungsweise diese Wahl von $v(\cdot)$ bringen den sofort ersichtlichen Vorteil mit sich, dass die Verteilungsfunktion G_{θ} eine explizite Darstellung besitzt, das heißt es ist gelungen das Integral $J(\cdot)$ in elementarer Form darzustellen. Dies ist bereits für sehr einfache Wahlen von $v(\cdot)$ wie $v(y) = (y - 1/2)^2$ nicht mehr möglich.⁶

Lemma 3.1.2.

Das Modell \mathcal{E}_E aus Definition 3.1.1 ist identifizierbar.

Beweis. Für den Beweis nutzen wir Lemma 1.1.35, weswegen es reicht Identifizierbarkeit des Basismodells $\{G_{\alpha,h} : \alpha, h > 0\}$ und Typenfreiheit von \mathcal{E}_E zu zeigen.

Seien also zunächst $\alpha_1, \alpha_2, h_1, h_2 > 0$ mit $G_{\alpha_1, h_1} = G_{\alpha_2, h_2}$ gegeben. Daraus folgt für $x \in]-1/2, 0[$ die Beziehung $\exp\{-2h_1/(\alpha_1 + 1) \cdot (-x)^{\alpha_1+1}\} = \exp\{-2h_2/(\alpha_2 + 1) \cdot (-x)^{\alpha_2+1}\}$, woraus $h_1/h_2 \cdot (\alpha_1 + 1)/(\alpha_2 + 1) = (-x)^{\alpha_2 - \alpha_1}$ auf dem genannten Bereich folgt. Da die linke Seite konstant in x ist, folgt $\alpha_1 = \alpha_2$ und daraus dann nach Einsetzen $h_1 = h_2$. Damit ist das Basismodell identifizierbar. \square

Auch wenn eine explizite Darstellung der Momente bisher nicht gefunden wurde, lassen sich zumindest Schranken angeben, womit alle Momente existieren.

⁶ Dies ist auch der namensgebende Grund für das explizite Stein'sche Modell.

Satz 3.1.3 (Momentenschranken für $G_{\alpha,h}$).

Sei X für $\alpha, h > 0$ nach $G_{\alpha,h}$ verteilt und $k \geq 1$. Dann gelten die Schranken

$$\begin{aligned} \mathbb{E}[|X^k|] &\leq c_{\alpha,h} \cdot \Gamma\left(\frac{k+1}{\alpha}, 2^{-\alpha}h\right) + 2^{-\alpha}h + \tilde{c}_{\alpha,h} \cdot \gamma\left(1 + \frac{k}{\alpha+1}, 2^{-\alpha}\frac{h}{\alpha+1}\right) \quad (\alpha < 1) \\ \mathbb{E}[|X^k|] &\leq \hat{c}_{\alpha,h} \cdot \Gamma\left(1 + \frac{k}{\alpha}, \frac{h}{2^\alpha}\right) + \frac{\alpha h}{2^{\alpha-1}} \tilde{c}_{\alpha,h} \cdot \gamma\left(1 + \frac{k}{\alpha+1}, 2^{-\alpha}\frac{h}{\alpha+1}\right) \quad (\alpha \geq 1). \end{aligned}$$

Dabei sind $c_{\alpha,h}, \tilde{c}_{\alpha,h}, \hat{c}_{\alpha,h}$ positive von α, h abhängige Konstanten, welche aus dem Beweis hervorgehen und

$$\Gamma(s, t) := \int_t^\infty x^{s-1} \cdot \exp(-x) dx \quad \gamma(s, t) := \int_0^t x^{s-1} \cdot \exp(-x) dx$$

für $\operatorname{Re}(s) \wedge \operatorname{Re}(t) > 0$ die **unvollständige obere** beziehungsweise **untere Gammafunktion**.

Beweis. Zunächst schreiben wir

$$\mathbb{E}[|X^k|] = \mathbb{E}[|X^k| \cdot \mathbb{1}\{X \leq -1/2\}] + \mathbb{E}[|X^k| \cdot \mathbb{1}\{X > -1/2\}] =: I_1 + I_2.$$

Wir behandeln zunächst die zweite Erwartung I_2 , da sie unabhängig von α exakt bestimmt werden kann. Mittels der Substitution $u = 2h/(\alpha+1) \cdot x^{\alpha+1}$ gilt

$$\begin{aligned} I_2 &= 2h \cdot \int_0^{1/2} x^{k+\alpha} \cdot \exp\left(-\frac{2h}{\alpha+1}x^{\alpha+1}\right) dx \\ &= \int_0^{2^{-\alpha}h/(\alpha+1)} \left(\frac{\alpha+1}{2h}\right)^{k/(\alpha+1)} \cdot x^{k/(\alpha+1)} \exp(-x) dx \\ &= \left(\frac{\alpha+1}{2h}\right)^{k/(\alpha+1)} \cdot \gamma\left(1 + \frac{k}{\alpha+1}, 2^{-\alpha}\frac{h}{\alpha+1}\right). \end{aligned}$$

In der Behandlung der ersten Erwartung beginnen wir mit dem Fall $\alpha < 1$:

Wegen der Tatsache $x^\alpha - y^\alpha \leq (x-y)^\alpha$ für $0 \leq y \leq x$ können wir abschätzen

$$\begin{aligned} I_1 &= 2h \cdot \int_{-\infty}^{-1/2} |x|^k \cdot ((-x)^\alpha - (-x-1/2)^\alpha) \exp\left\{-\frac{2h}{\alpha+1}((-x)^{\alpha+1} - (-x-1/2)^{\alpha+1})\right\} dx \\ &\leq 2h \cdot \int_{1/2}^\infty 2^{-\alpha}x^k \cdot \exp\left\{-\frac{2h}{\alpha+1}(x^{\alpha+1} - (x-1/2)^{\alpha+1})\right\} dx \\ &\leq 2h \cdot \int_1^\infty 2^{-\alpha}x^k \cdot \exp\{-2^{-\alpha}hx^\alpha\} dx + 2^{-\alpha}h \\ &= \frac{2}{\alpha} \int_{2^{-\alpha}h}^\infty \left(\frac{2}{h^{1/\alpha}}\right)^{k+1-\alpha} x^{(k+1-\alpha)/\alpha} \cdot \exp(-x) dx + 2^{-\alpha}h \\ &= \frac{2^{k+2-\alpha}}{\alpha} h^{1-(k+1)/\alpha} \cdot \Gamma\left(\frac{k+1}{\alpha}, 2^{-\alpha}h\right) + 2^{-\alpha}h. \end{aligned}$$

Dabei wurde das Integral am Punkt 1 aufgeteilt und das Integral von $1/2$ bis 1 mittels der Grenzen abgeschätzt. Auf Argument der Exponentialfunktion des verbleibenden Integrals wurde der Mittelwertsatz angewandt und die Abschätzung $x - \theta_x/2 \geq x/2$ für $0 \leq \theta_x \leq 1 \leq x$ genutzt. Im zweiten Gleichheitszeichen wurde eine Substitution genutzt.

Im Fall $\alpha \geq 1$ rechnen wir:

$$\begin{aligned}
I_1 &= 2h \cdot \int_{1/2}^{\infty} x^{\frac{\alpha}{2}} (x - \theta_x/2)^{\alpha-1} \cdot \exp \left\{ -\frac{2h}{\alpha+1} \left((-x)^{\alpha+1} - (-x-1/2)^{\alpha+1} \right) \right\} dx \\
&\leq \alpha h \cdot \int_1^{\infty} x^{\alpha+k-1} \cdot \exp(-2^{-\alpha} h x^{\alpha}) dx + \frac{\alpha h}{2^{\alpha-1}} \\
&= \alpha h \int_{h/2^{\alpha}}^{\infty} \frac{2^{\alpha}}{\alpha h} 2^k \left(\frac{x}{h} \right)^{k/\alpha} \cdot \exp(-x) dx + \frac{\alpha h}{2^{\alpha-1}} \\
&= \frac{2^{\alpha+k}}{h^{k/\alpha}} \cdot \Gamma \left(1 + \frac{k}{\alpha}, \frac{h}{2^{\alpha}} \right) + \frac{\alpha h}{2^{\alpha-1}}.
\end{aligned}$$

In der ersten Gleichheit wurde der Mittelwertsatz angewandt, in der ersten Ungleichung das Integral am Punkt 1 aufgeteilt und ähnlich wie im Fall $\alpha < 1$ abgeschätzt. Die vorletzte Gleichheit nutzt wieder eine Substitution. \square

Die Verteilungsfunktion aus (3.1) besitzt drei verschiedene Verhaltensbereiche. Daher ist es naheliegend die Regularität in den ausgewiesenen Wechselstellen zu untersuchen. Wir betrachten hier die Regularität in den rechten Randpunkt hinein nicht genauer, formulieren sie jedoch ohne Beweis nach diesem Lemma.

Lemma 3.1.4 (Regularität von g_{θ}).

Sei g_{θ} wie oben gegeben. Dann gilt $g_{\theta} \in \mathcal{C}^{[\alpha]}(-\infty, 0[)$ und g_{θ} ist nicht $[\alpha] + 1$ mal differenzierbar. Dabei bezeichne $[\cdot]$ hier die echte Gaußklammer, das heißt

$$[x] := \max \{ z \in \mathbb{Z} : z < x \}.$$

Beweis. Da g_{θ} aus $g_{\alpha,h}$ lediglich durch Reskalierung hervorgeht, reicht es die Behauptung für $g_{\alpha,h}$ zu zeigen.

Offensichtlich gilt $g_{\alpha,h} \in \mathcal{C}^{\infty}(-\infty, -1/2[\cup]-1/2, 0[)$, sodass sich die Untersuchung auf die Stelle $x = -1/2$ reduziert.

Schritt 1: Wir zeigen $g_{\alpha,h} \in \mathcal{C}^k(-\infty, \mu[)$ für $k < [\alpha]$. Um die Notation für den Beweis

zu vereinfachen, setzen wir

$$\begin{aligned}
f_l(x) &:= 2h((-x)^\alpha - (-x - 1/2)^\alpha) \exp\left\{-\frac{2h}{\alpha+1}((-x)^{\alpha+1} - (-x - 1/2)^{\alpha+1})\right\} \\
f_r(x) &:= 2h(-x)^\alpha \cdot \exp\left\{-\frac{2h}{\alpha+1}(-x)^{\alpha+1}\right\} \\
f_1(x) &:= 2h \cdot (-x)^\alpha \\
f_2(x) &:= 2h \cdot (-x - 1/2)^\alpha \\
g_1(x) &:= -\frac{2h}{\alpha+1} \cdot (-x)^{\alpha+1} \\
g_2(x) &:= -\frac{2h}{\alpha+1} \cdot (-x - 1/2)^{\alpha+1},
\end{aligned}$$

sodass $f_l(x) = (f_1(x) - f_2(x)) \cdot \exp(g_1(x) - g_2(x))$ und $f_r(x) = f_1(x) \cdot \exp(g_1(x))$.
Damit erhalten wir also $f_l(x) = f_r(x) \cdot \exp(-g_2(x)) - f_2(x) \cdot \exp(g_1(x) + g_2(x))$.

Zunächst untersuchen wir die linksseitige Ableitung mittels Induktion, wobei wegen der Stetigkeit von $g_{\alpha,h}$ in $-1/2$ für $\alpha < 1$ nichts zu zeigen ist. Sei also $\alpha \geq 1$, so erhalten wir durch die letzte Gleichung folgende Darstellung

$$\begin{aligned}
f_l^{(k)}(x) &= \prod_{i=0}^k f_r^{(i)}(x) \cdot (\exp(-g_2(x)))^{(k-i)} \\
&\quad - \prod_{i=0}^k f_2^{(i)}(x) \cdot (\exp(g_1(x) - g_2(x)))^{(k-i)}.
\end{aligned} \tag{3.2}$$

Für $i < k$ gilt $\lim_{x \uparrow -1/2} (\exp(-g_2(x)))^{(k-i)} = 0$ und für $i \leq k$: $\lim_{x \uparrow -1/2} f_2^{(i)}(x) = 0$. Da die anderen Faktoren in einer Umgebung von $-1/2$ beschränkt bleiben, folgt daraus und aus den Tatsachen $f_r \in \mathcal{C}^\infty(]-\infty, 0[)$, $g_2 \in \mathcal{C}^{[\alpha]}(]-\infty, -1/2])$

$$\lim_{x \uparrow -1/2} f_l^{(k)}(x) = f_r^{(k)}(-1/2) \cdot 1 - 0 = f_r^{(k)}(-1/2).$$

Damit, dem Mittelwertsatz und vollständiger Induktion folgt

$$\lim_{\delta \downarrow 0} \frac{f_l^{(k-1)}(-1/2 - \delta) - f_l^{(k-1)}(-1/2)}{-\delta} = \lim_{\delta \downarrow 0} f_l^{(k)}(-1/2 - \delta \cdot \zeta_\delta) = f_r^{(k)}(-1/2),$$

wobei $\zeta_\delta \in [0, 1]$ und die Induktionsvoraussetzung für $k-1$ genutzt wurde. Wegen $f_r \in \mathcal{C}^\infty(]-\infty, 0[)$ folgt trivialerweise die Gleichheit der rechtsseitigen Ableitung und die Aussage des Schrittes 1 ist bewiesen.

Schritt 2: Für $\alpha \notin \mathbb{N}$ zeigen wir, dass $g_{\alpha,h}$ nicht $[\alpha] + 1$ -mal differenzierbar ist.

Innerhalb von (3.2) wird der Term $f_2^{([\alpha]+1)}(x)$ für $x \uparrow -1/2$ unbeschränkt. Da alle anderen Terme beschränkt sind, folgt die nicht-Existenz der linksseitigen Ableitung im Punkt $-1/2$.

Schritt 3: Die Aussage aus Schritt 2 bleibt für $\alpha \in \mathbb{N}$ zu zeigen: Also zeigen wir, dass $g_{\alpha,h}$ in $-1/2$ nicht α -mal differenzierbar ist. Dafür vergleichen wir die linksseitige und rechtsseitige Ableitung. Betrachten wir $f_l^{(\alpha)}(x)$, so entnehmen wir der Darstellung aus (3.2), dass nur die Terme mit $i = \alpha$ einen Beitrag liefern. Es gilt daher

$$\begin{aligned} \lim_{x \uparrow -1/2} f_l^{(\alpha)}(x) &= f_r^{(\alpha)}(-1/2) \cdot 1 - f_2^{(\alpha)}(-1/2) \cdot \exp\left(-\frac{2h}{\alpha+1} 2^{-\alpha-1}\right) \\ &= f_r^{(\alpha)}(-1/2) - 2h(-1)^\alpha \cdot \alpha! \exp\left(-\frac{2h}{\alpha+1} 2^{-\alpha-1}\right). \end{aligned}$$

Daraus schließen wir

$$\begin{aligned} \lim_{\delta \downarrow 0} \frac{f_l^{(\alpha-1)}(-1/2 - \delta) - f_l^{(\alpha-1)}(-1/2)}{-\delta} &= \lim_{\delta \downarrow 0} f_l^{(\alpha)}(-1/2 - \delta \cdot \zeta_\delta) \\ &= f_r^{(\alpha)}(-1/2) - 2h(-1)^\alpha \cdot \alpha! \cdot \exp\left(-\frac{2h}{\alpha+1} 2^{-\alpha-1}\right) \neq f_r^{(\alpha)}(-1/2) \\ &= \lim_{\delta \downarrow 0} \frac{f_r^{(\alpha-1)}(-1/2 + \delta) - f_r^{(\alpha-1)}(-1/2)}{\delta}, \end{aligned}$$

sodass $g_{\alpha,h}$ in $-1/2$ nicht α -mal differenzierbar ist. □

Für die Regularität in den rechten Randpunkt hinein gilt:

Bemerkung 3.1.5.

Sei g_θ wie oben gegeben. Dann gilt $g_\theta \in \mathcal{C}^{[\alpha]}(]-1/2, \mu[)$ und g_θ ist nicht $[\alpha] + 1$ mal differenzierbar. Dabei bezeichne $[\cdot]$ hier wie oben die echte Gaußklammer.

Beweis. Dies kann mittels ähnlicher Überlegungen wie im Beweis von Lemma 3.1.4 gezeigt werden und wird hier nicht ausgeführt. ◇

3.2 Die klassische Block-Maxima Methode

Die **Block-Maxima Methode** im klassischen Kontext bezeichnet eine auf Gumbel, siehe [Gumbel], zurückgehende Anwendung des Fisher-Tippett Theorems auf ein subasymptotisches Modell. Folgende Darstellung folgt [Beirlant et al. S. 132 ff.] und [Domby und Ferreira].

Bemerkung 3.2.1 (Die Block-Maxima Methode im u.i.v. Fall).

Seien $n, m \in \mathbb{N}$ und $X_1, \dots, X_{m \cdot n}$ unabhängig und identisch verteilte Zufallsvariablen, die im Max-Anziehungsbereich einer Extremwertverteilung liegen. Wir unterteilen die endliche Folge der Zufallsvariablen in m viele Blöcke der Größe n , das heißt $B_j := (X_{1+n \cdot (j-1)}, \dots, X_{n \cdot j})$ für $1 \leq j \leq m$. Über diese Blöcke bilden wir wiederum das Maximum, um so das **Block-Maximum** $M_{j,n} := \max B_j$ für $1 \leq j \leq m$ zu erhalten. Nach dem Fisher-Tippett Theorem existieren von n abhängige Konstanten

$b_n \in \mathbb{R}$, $a_n > 0$ und ein $\xi \in \mathbb{R}$ für die

$$\frac{M_{j,n} - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} \text{GEV}(0, 1, \xi) \quad (1 \leq j \leq m)$$

gilt, wenn wir die endliche Folge an Zufallsvariablen mit unabhängig und identisch verteilten Zufallsvariablen auffüllen. Zu bemerken ist hier, dass die normalisierenden Folgen gleichmäßig unterhalb der Blöcke sind und der Grenzwert in n also bedeutet, dass die m Blöcke in ihrer Größe gegen unendlich streben. Obige Gleichung lässt sich für hinreichend große n zu $M_{j,n} \approx a_n \text{GEV}(\xi) + b_n$ in Verteilung umformulieren, was die Annahme der Verteilungsgleichheit $M_{j,n} = a_n \text{GEV}(\xi) + b_n$ nahelegt. Die Block-Maxima Methode besteht nun darin, dass die Verteilungsgleichheit angenommen wird, also bereits die Asymptotik im subasymptotischen greift, symbolisch

$$M_{j,n} \stackrel{d}{=} \text{GEV}(b_n, a_n, \xi) \quad (1 \leq j \leq m). \quad (3.3)$$

Bemerkung 3.2.2.

1. Hierbei ist zu beachten, dass das Fisher-Tippet Theorem natürlich im allgemeinen nicht im endlichen gilt, sodass der Block-Maxima Methode ein systemischer Fehler folgt, man sagt auch, dass das Modell *missspezifiziert* sei, siehe [Dombry und Ferreira]. Mithin ist es in der Praxis indiziert vor einer Anwendung der Block-Maxima Methode die Anpassungsgüte zu betrachten, was beispielsweise mittels QQ-Plots, Anderson-Darling oder Kolomogorov-Smirnov Anpassungstests bewerkstelligt werden kann.
2. In der Praxis wird diese Methode oft in der Klimatologie angewandt, etwa im Rahmen jährlicher Maximaltemperaturen oder maximaler Niederschlagsmengen pro Quadratmeter [Engeland et al.], aber auch bei wirtschafts- oder versicherungswissenschaftlichen Fragen [Yaqiong].
3. Für praktische Anwendungen liegt der Nutzen der Methode darin, aus einer bekannten Verteilung simulieren zu können, falls die Parameter geschätzt wurden.
4. Eine populäre Alternative zu der Block-Maxima Methode ist die *Peak over Thresholds Methode*, welche nur Beobachtungen beachtet, die einen gewissen Schwellenwert übertreffen, sogenannte **Ekzesse**. Auf diese gehen wir in dieser Arbeit nicht tiefer ein.
5. Die Blöcke in der von uns erklärten Block-Maxima Methode wurden disjunkt gewählt, sodass die Unabhängigkeit gewahrt wird und somit das Fisher-Tippet Theorem anwendbar bleibt. Es ist auch möglich sich überlappende Blöcke zu wählen, **Sliding Block-Maxima Method**, sodass die resultierenden Blöcke nicht mehr unabhängig jedoch noch identisch verteilt sind. Dieser Ansatz ist informationserhaltender als die disjunkte Methode, da aus derselben Zeitreihe mehr Maxima extrahiert werden können. Tatsächlich kann ein Schätzer, der auf dieser

Methode basiert asymptotisch effizienter sein als ein auf disjunkten Blöcken basierender Schätzer, siehe dazu zum Beispiel [Bücher und Segers(b)]. Wir weisen darauf hin, dass in dieser Arbeit immer die *disjunkte* Methode gemeint ist, wenn von einer Block-Maxima Methode gesprochen wird.

6. Die Wahl der Anzahl der Blöcke m ist nicht unerheblich für die Güte eines Schätzers. So ist es naheliegend, dass extreme Wahlen wie $m = 1$ oder $m = m \cdot n$ im Allgemeinen ineffizient sind, da im ersten Fall lediglich ein großer Block und im anderen Falle nur Blöcke der Größe 1 gebildet werden, sodass die Asymptotik aus dem Fisher-Tippett Theorem nicht greifen kann. Tiefer wird darauf in [Bücher und Zhou, S. 4] eingegangen und wir verweisen an dieser Stelle lediglich darauf.

Das zugrundeliegende Ziel die Verteilung eines Block-Maximums M_j zu schätzen wird im Rahmen der Block-Maxima Methode wie folgt erreicht.

Bemerkung 3.2.3 (Schätzen von Parametern).

Nach (3.3) wird angenommen, dass die Blockmaxima exakt einer GEV-Verteilung folgen, sodass eine Schätzung der Verteilung der Blockmaxima über eine Schätzung des Formparameters (Extremwertindex) ξ , des Lokationsparameters b_n und des Skalenparameters a_n erreicht wird.

Populäre Methoden dafür sind der Maximum-Likelihood Schätzer und der *gewichtete Momentenschätzer*, siehe [Hoskin et al.], wobei letztere Methode eine Verallgemeinerung der klassischen Momentenmethode darstellt, auf die nicht weiter eingegangen werden soll.

Aus der Definition der GEV-Verteilung 1.2.2 errechnet sich die Log-Likelihood Funktion eines Samples x der Größe N als

$$\begin{aligned} \ell_x(\theta) &= -N \log \sigma - \left(\frac{1}{\xi} + 1 \right) \cdot \sum_{i=1}^N \log \left(1 + \xi \cdot \frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^N \left(1 + \xi \cdot \frac{x_i - \mu}{\sigma} \right)^{-1/\xi}, \end{aligned}$$

falls $x_{1:N} > \mu - \sigma/\xi$ und $\xi \neq 0$, beziehungsweise

$$\begin{aligned} \ell_x(\theta) &= -N \log \sigma - \left(\frac{1}{\xi} + 1 \right) \cdot \sum_{i=1}^N \log \left(1 + \xi \cdot \frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^N \left(1 + \xi \cdot \frac{x_i - \mu}{\sigma} \right)^{-1/\xi}, \end{aligned}$$

falls $\xi = 0$.

Wie bereits unter Beispiel 1.1.45 umrissen, hängt der Träger der Verteilung von den Parametern ab, sodass die klassischen Regularitätsannahmen aus [Cramér, Abschnitt 33.3] nicht gegeben sind, was die Herleitung von Existenz und Gütekriterien aufwendig gestaltet. Wobei zumindest für kompakte Parameterbereiche aufgrund der Stetigkeit der Likelihood-Funktion die Existenz gesichert ist. Wir bezeichnen diesen Maximum-Likelihood Schätzer als $\hat{\theta}_{\text{GEV}}$.

Bemerkung 3.2.4 (Schätzen von Quantilen).

Nach einer Schätzung der Parameter der GEV-Verteilung lassen sich q -Quantile für $q \in]0, 1[$ durch Invertieren der Verteilungsfunktion und einsetzen der Schätzwerte beziehungsweise von q schätzen. Dabei ist die Quantilfunktion der GEV-Verteilung wie folgt gegeben

$$\text{GEV}(\mu, \sigma, \xi)^{-1}(q) = \begin{cases} \mu - \sigma \log(-\log(1-q)) & , \xi = 0 \\ \mu + \frac{\sigma}{\xi} \cdot \{(-\log(1-q))^{-\xi} - 1\} & , \xi \neq 0 \end{cases}. \quad (3.4)$$

Damit ist $\text{GEV}(\hat{\theta}_G)^{-1}(q)$ ein Schätzer für das q -Quantil der Block-Maxima.

Bemerkung 3.2.5 (Güte des Maximum-Likelihood Schätzers in der Blockmaxima Methode).

Der Maximum-Likelihood Schätzer erfüllt unter gewissen Kriterien asymptotische Güteeigenschaften, jedoch ist selbst die Existenz nicht offensichtlich. So gilt in der Tat, dass die Log-Likelihood Funktion im Allgemeinen kein globales Maximum annimmt, siehe [Dombry]. Für lokale Maxima ist dann die Bezeichnung **Pseudo-Maximum-Likelihood Schätzer** geläufig.

Für den Maximum-Likelihood Schätzer ohne Bedingung zweiter Ordnung im misspezifizierten Modell wurde fast-sichere Existenz und starke Konsistenz für $\xi > -1$ in [Dombry, Theorem 2] gezeigt und die Existenz für $\xi < -1$ widerlegt.

Asymptotische Normalität unter der Voraussetzung, dass das Sample unabhängig und identisch GEV-verteilt ist wurde für $\xi > -1/2$ in [Bücher und Segers(a), Proposition 3.3] gezeigt, wobei Kompaktheit des Parameterraumes angenommen wurde. Dass für $\xi \leq -1/2$ keine asymptotische Normalität gilt, wird unter anderem von Beispiel 1.1.45 nahegelegt.

Die asymptotische Normalität im misspezifizierten Modell wird mittels *Bedingungen zweiter Ordnung* behandelt, wobei die *Bedingung erster Ordnung* die Zugehörigkeit zu einem Max-Anziehungsbereich bezeichnet und davon ausgehend äquivalent umformuliert wird, siehe [Dombry und Ferreira]. In demselben Artikel wurde auch gezeigt, dass für $\xi > -1/2$ auch in der misspezifizierten Situation die Normalität vorliegt, siehe [Dombry und Ferreira, Theorem 2.1].

3.3 Die Block-Maxima Methode im expliziten Stein'schen Modell

Die Block-Maxima Methode lässt sich auch in Situationen von nicht identisch verteilten zugrundeliegenden Zufallsvariablen anwenden.

Bemerkung 3.3.1.

Seien für $n, m \in \mathbb{N}$ unabhängige Zufallsvariablen $X_1, \dots, X_{m \cdot n}$ gegeben, welche die Grenzwertrelation aus Theorem 2.3.7 erfüllen. Damit sind die Zufallsvariablen insbesondere nicht-stationär, sodass wir sie mit $X_{i,j}$ für $1 \leq i \leq m$ und $1 \leq j \leq n$ identifizieren. Das heißt wir haben m Zeitreihen der Länge (Blockgröße) n gegeben. Erneut setzen wir $M_{j,n} := \max(X_{j,1}, \dots, X_{j,n})$ als das **Block-Maximum** über den j -ten

Block, wobei $1 \leq j \leq m$. Nach dem zweiten Grenzwertsatz Theorem 2.3.7 existiert eine normalisierende Folge $G_n > 0$ mit

$$\frac{M_{j,n} - u}{G_n} \xrightarrow[n \rightarrow \infty]{d} G_{\alpha,h} \quad (1 \leq j \leq m).$$

Dabei kann dieselbe Folge G_n für alle j gewählt werden. Wie in der klassischen Situation bedeutet diese Konvergenz für große n , dass $M_{j,n}$ nahezu wie $G_n \cdot G_{\alpha,h} + u$ verteilt ist. Man nimmt also im Rahmen der Block-Maxima Methode im Stein'schen Modell an, dass bereits subasymptotisch die Verteilungsgleichheit gilt, das heißt

$$M_{j,n} \stackrel{d}{=} G_n \cdot G_{\alpha,h} + u \quad (1 \leq j \leq m).$$

Bemerkung 3.3.2 (Schätzen von Parametern).

Die Annahme, die Blockmaxima folgen genau der G_θ -Verteilung, führt wie in der klassischen Block-Maxima Methode dazu, dass wir anstelle der tatsächlichen Verteilung der Blockmaxima den Parameter θ schätzen, sodass wir ein parametrisches Schätzproblem identifiziert haben. Das Modell \mathcal{E}_E erfüllt zwar nicht die klassischen Regularitätsbedingungen, aber ähnliche wie das GEV-Modell und ähnelt diesem auch stark. Dies deutet darauf hin, dass der Maximum-Likelihood Schätzer asymptotische Gütekriterien erfüllt. Wir stellen im Folgenden daher den (Quasi-) Maximum-Likelihood Schätzer für dieses Modell vor.

Man beachte, eine (verallgemeinerte) Momentenmethode erscheint initial zielführend, da sie im Rahmen des GEV-Modells erfolgreich ist. In diesem Fall findet jedoch eine Veränderung der Form der Verteilungsfunktion im Punkt $-\sigma/2 + \mu$ statt, welcher also von Parametern des Modells abhängt, wodurch insbesondere im linken Strukturbereich keine explizite Bestimmung der Quantilfunktion möglich ist. Dies führt zu einer Komplikation bei der Momentenbestimmung, sodass dieser Ansatz hier nicht weiter verfolgt wird.

Ein analoges Resultat zum Pickands-Balkema-de Haan Theorem wurde noch nicht etabliert. Auch diesen Ansatz werden wir nicht weiter verfolgen.

Wir merken an, dass die Existenz eines Pseudo-Maximum-Likelihood Schätzers aus der Stetigkeit der Likelihood-Funktion, siehe (3.5), gefolgert werden kann. Die globale Eindeutigkeit ist nicht untersucht. Weiterhin wurden noch keine theoretischen Resultate zu der Güte des Schätzers etabliert. Praktische Untersuchungen asymptotischer Eigenschaften im Rahmen von Simulationsstudien werden im vierten Kapitel durchgeführt.

Bemerkung 3.3.3 (Maximum-Likelihood Schätzung).

Wir nehmen an, dass die Verteilungen der Blockmaxima dem Modell \mathcal{E}_E entstammen; wohl wissend, dass wir hier eine konkrete Missspezifikation billigen. Im Folgenden nutzen wir die Notation $x_\theta := (\mu - x)/\sigma$. Aus der Gestalt der Verteilungsfunktion in (3.1) bestimmen wir die Likelihood-Funktion abhängig von einem Sample x der Größe

1 als

$$L_x(\theta) \quad (3.5)$$

$$= \begin{cases} \frac{2h}{\sigma} \cdot (x_\theta^\alpha - (x_\theta^\alpha - 1/2)^\alpha) \cdot \exp\left\{-\frac{2h}{\alpha+1} \cdot (x_\theta^{\alpha+1} - (x_\theta - 1/2)^{\alpha+1})\right\} & , x < -\frac{\sigma}{2} + \mu \\ \frac{2h}{\sigma} \cdot x_\theta^\alpha \cdot \exp\left\{-\frac{2h}{\alpha+1} \cdot x_\theta^{\alpha+1}\right\} & , -\frac{\sigma}{2} + \mu \leq x < \mu \\ 0 & , \mu \leq x \end{cases}$$

Für ein Sample $x = (x_1, \dots, x_N)$ der Größe N hat die Likelihood-Funktion wegen der Gestalt des Modells die Form $L_x(\theta) = \prod_{i=1}^N L_{x_i}(\theta)$.

Daraus ergibt sich die Log-Likelihood Funktion, wieder zunächst für Samples der Größe 1, als

$$\ell_x(\theta) \quad (3.6)$$

$$= \begin{cases} \log\left(\frac{2h}{\sigma}\right) + \log(x_\theta^\alpha - (x_\theta^\alpha - 1/2)^\alpha) - \frac{2h}{\alpha+1} \cdot (x_\theta^{\alpha+1} - (x_\theta - 1/2)^{\alpha+1}) & , x < -\frac{\sigma}{2} + \mu \\ \log\left(\frac{2h}{\sigma}\right) + \alpha \cdot \log(x_\theta) - \frac{2h}{\alpha+1} \cdot x_\theta^{\alpha+1} & , -\frac{\sigma}{2} + \mu \leq x < \mu \\ -\infty & , \mu \leq x \end{cases}$$

Aus Rechenregeln des Logarithmus' folgt für Samples der Größe N die Gestalt

$$\ell_x(\theta) = \sum_{i=1}^N \ell_{x_i}(\theta).$$

Ein Ableiten der Log-Likelihood Funktion führt zu der Score-Funktion. Wir nutzen die abkürzende Schreibweise $\partial_j l_x(\theta)^{(1)} := \partial_j l_x(\theta) \cdot \mathbb{1}\{x < -\sigma/2 + \mu\}$ und

$\partial_j l_x(\theta)^{(2)} := \partial_j l_x(\theta) \cdot \mathbb{1}\{-\sigma/2 + \mu < x < \mu\}$, sodass $\partial_j l_x(\theta) = \partial_j l_x(\theta)^{(1)} + \partial_j l_x(\theta)^{(2)}$, wobei $j \in \{\alpha, h, \mu, \sigma\}$. Wir erhalten die Ausdrücke

$$\partial_\alpha \ell_x(\theta)^{(1)} \quad (3.7)$$

$$= \frac{\log(x_\theta) \cdot x_\theta^\alpha - \log(x_\theta - 1/2) \cdot (x_\theta - 1/2)^\alpha}{x_\theta^\alpha - (x_\theta - 1/2)^\alpha}$$

$$- 2h \frac{(\alpha + 1) \left\{ \log(x_\theta) x_\theta^{\alpha+1} - \log(x_\theta - 1/2) (x_\theta - 1/2)^{\alpha+1} \right\} - x_\theta^{\alpha+1} + (x_\theta - 1/2)^{\alpha+1}}{(\alpha + 1)^2}$$

$$\partial_\alpha \ell_x(\theta)^{(2)} = \log(x_\theta) - 2h \cdot \frac{(\alpha + 1) \cdot \log(x_\theta) \cdot x_\theta^{\alpha+1} - x_\theta^{\alpha+1}}{(\alpha + 1)^2}$$

$$\partial_h \ell_x(\theta)^{(1)} = \frac{1}{h} - 2 \cdot \frac{x_\theta^{\alpha+1} - (x_\theta - 1/2)^{\alpha+1}}{\alpha + 1}$$

$$\partial_h \ell_x(\theta)^{(2)} = \frac{1}{h} - 2 \cdot \frac{x_\theta^{\alpha+1}}{\alpha + 1}$$

$$\partial_\mu \ell_x(\theta)^{(1)} = \frac{\alpha}{\sigma} \cdot \frac{x_\theta^{\alpha-1} - (x_\theta - 1/2)^{\alpha-1}}{x_\theta^\alpha - (x_\theta - 1/2)^\alpha} - \frac{2h}{\sigma} \cdot (x_\theta^\alpha - (x_\theta - 1/2)^\alpha)$$

$$\partial_\mu \ell_x(\theta)^{(2)} = \frac{\alpha}{\mu - x} - \frac{2h \cdot x_\theta^\alpha}{\sigma}$$

$$\partial_\sigma \ell_x(\theta)^{(1)} = -\frac{1}{\sigma} - \frac{\alpha}{\sigma} \cdot \frac{x_\theta^\alpha - (x_\theta - 1/2)^{\alpha-1} \cdot x_\theta}{x_\theta^\alpha - (x_\theta - 1/2)^\alpha} + \frac{2h}{\sigma} \cdot (x_\theta^{\alpha+1} - x_\theta \cdot (x_\theta - 1/2)^\alpha)$$

$$\partial_\sigma \ell_x(\theta)^{(2)} = -\frac{\alpha + 1 + 2h \cdot x_\theta^{\alpha+1}}{\sigma}.$$

Man beachte, dass in der partiellen Ableitung nach α der Ausdruck $x_\theta - 1/2$ in einer numerischen Berechnung aufgrund von Maschinengenauigkeit als nicht-positiv bestimmt werden kann, sodass diese Pathologie in der Programmierung mittels Korrektursummen behoben werden muss.

Wegen der Stetigkeit der Log-Likelihood Funktion existiert zumindest für kompakte $\Theta_0 \subset \Theta_E$ der Maximum-Likelihood Schätzer, den wir als $\hat{\theta}_S := (\hat{\alpha}_S, \hat{h}_S, \hat{\mu}_S, \hat{\sigma}_S)$ notieren.

Bemerkung 3.3.4 (Schätzen von Quantilen).

Wie in Bemerkung 3.2.4 lassen sich nach Schätzung der Parameter der G_θ -Verteilung durch Invertierung und Einsetzen q -Quantile schätzen. Hier lässt sich jedoch im Gegensatz zu Gleichung (3.4) die Verteilungsfunktion nur im rechten Strukturbereich invertieren. Für $\alpha \rightarrow \infty$ oder $h \rightarrow 0$ gilt $G_\theta(x) \rightarrow 1$, weswegen selbst extreme Quantile für bestimmte Parameterbereiche in den linken Strukturbereich fallen können. Dies führt dazu, dass eine Schätzung der Quantile auf Basis des Maximum-Likelihood Schätzers implizit erfolgt.

4 Simulationsstudien

Im folgenden Kapitel führen wir mittels der Statistiksoftware *R* Simulationsstudien durch, die die vorangegangenen Subkapitel illustrieren und erste Tendenzen konkreter Schätzer aufzeigen. Insbesondere untersuchen wir wie gut sich die Grenzverteilung $G_{\alpha,h}$ an simulierte Zeitreihensamples mittels Maximum-Likelihood Schätzung anpassen lässt. Dies wird mit einer Anpassung durch den klassischen GEV Maximum-Likelihood Schätzers verglichen. Außerdem werden aus beiden Schätzungen Quantile geschätzt und Kennzahlen der Schätzer gegenübergestellt.

4.1 Eine Simulationsstudie im expliziten Stein'schen Modell

Zu Beginn wollen wir Daten simulieren, deren Verteilung exakt einem expliziten Stein'schen Modell folgt. Ausgehend von solchen Samples implementieren wir mehrere Maximum-Likelihood Schätzer, die wir miteinander vergleichen und zuletzt denjenigen weiterverwenden, welcher in den Tests die besten Resultate erzielt. Dieser Ansatz ist dadurch motiviert, dass er nicht vom datengenerierenden Prozess abhängt. Eine Optimierung auf dem später angewandten Prozess würde nicht bedeuten, dass er auf anderen ähnliche Güte bewahrt.

Annahme 4.1.1 (Parameterbereiche im expliziten Stein'schen Modell).

Wir formulieren Bereiche für die Parameter α , h , μ und σ , welche wir im Anschluss motivieren.

$$\begin{aligned} 0.3 \leq \alpha \leq 30, \quad 0.01 \leq h \leq 10, \\ -100 \leq \mu \leq 100, \quad 10^{-8} \leq \sigma \leq 10. \end{aligned}$$

Wie bereits angesprochen entspricht die Form der Verteilungsfunktion von G_θ aus (3.1) exakt einem Weibull-Typ mit Formparameter $\alpha + 1$. Eine *Umrechnung* in den Parameter der GEV-Verteilung liefert daher $\xi \approx -1/(\alpha + 1)$. Der Maximum-Likelihood Schätzer für das GEV-Modell ist jedoch nur für $\xi > -1/2$ asymptotisch normal und für $\xi > -1$ (stark) konsistent; vergleiche Bemerkung 3.2.5. Es ist daher sinnvoll $\xi \geq -1/2$ zu fordern, was nach der Umrechnungsformel 1.2.3 $\alpha \geq 1$ entspricht. Man beachte, dass diese *Umrechnung* eine Heuristik ist, da G_θ im linken Strukturbereich keinem Weibull-Typen entspricht. Die rechte Grenze 30 entspricht $\xi \approx -1/31 \approx 0.03$, was gewählt wird, damit der GEV-Schätzer keine Übergänge zu einem Gumbel oder Fréchet Typ schätzt, was den Bias- und die Varianzbestimmung verzerren würde. Die Restriktionen an h , μ und σ sind Kompaktheitsannahmen, welche die Existenz eines Maximum-Likelihood Schätzers garantieren sollen und in späteren Anwendungskontexten bestätigt werden sollten. In der weiteren Simulation werden wir nur den Formparameter α des expliziten Stein'schen Modells variieren. Diesen Weg verfolgen zum einen deswegen, da der Parameter h in der GEV-Verteilung keine Entsprechung besitzt. Zum anderen degeneriert der Formparameter h im rechten Tail zu einem Skalenparameter, sodass er hier für uns nicht von weiterem Interesse sein soll.

Bemerkung 4.1.2 (Datengrundlage im spezifizierten Modell).

Mittels der *Inversionsmethode* simulieren wir für verschiedene Parameter $N = 100$, die einer von den Parameter abhängenden G_θ -Verteilung folgen. Man beachte, dass hierbei die Quantilsfunktion im linken Strukturbereich der G_θ -Verteilung numerisch bestimmt werden muss, was den Rechenaufwand erhöht. Auf Basis dieser Samples wollen wir Maximum-Likelihood Schätzer unter der Modellannahme $\mathcal{E} = \mathcal{E}_E$ implementieren; das heißt, hier liegt keine Missspezifikation vor.

Für die Implementierung der Maximum-Likelihood Schätzer haben wir die Pakete *optimx* und *nloptr* verglichen, welche umfangreichere Funktionalität als die gewöhnliche *optim*-Funktion aus dem *base* Paket bieten, siehe [Nash und Varadhan] und [Johnson]. Dabei hat sich in unseren Studien das *nloptr* Paket als praktischer erwiesen, da in diesem untere und obere Grenzen für Parameter direkt in die Optimierungsfunktion übergeben werden können. Dies vermindert die Anzahl von ad-hoc Veränderungen an Ziel- oder Score-Funktion, die weitere Ungenauigkeiten produzieren könnten. Wir simulieren je α je 250 Samples der Größe $N = 100$ wobei $\alpha \in \{1, \dots, 10\}$, woraus sich $10 \cdot 250 \cdot 100 = 25 \cdot 10^4$ Beobachtungen ergeben.

Bemerkung 4.1.3 (Implementierung und Wahl des Optimierungsalgorithmus').

Damit ein möglichst leistungsstarker Maximum-Likelihood Schätzer für das misspezifizierte Modell gewählt werden kann, wollen wir einige Schätzer im korrekt spezifizierten Modell betrachten und anhand eines Kriteriums den bestleistenden von ihnen auswählen. Ziel ist ein datenadaptiver Schätzer, sodass im Rahmen von Annahme 4.1.1 keine weitere Supplementierung von Informationen über θ nötig ist. Die Implementierung eines auf der G_θ -Verteilung basierenden Maximum-Likelihood Schätzers wird wegen der Abhängigkeit des Trägers der Verteilung vom Lokationsparameter diffizil, siehe Gleichung (3.1). Daher existiert die Log-Likelihood Funktion nicht überall. Auch kann der Ausdruck $\log((\mu - x)/\sigma - 1/2)$ in ℓ_α aufgrund von Rundungsfehlern in einer Simulation singulär werden, sodass ein Korrektursummand 10^{-20} in der Programmierung der Score-Funktion aus (3.7) eingeführt wurde. Da die Score-Funktion selbst auf dem Träger der Verteilung nicht explizit nach Null aufgelöst werden kann, muss die Log-Likelihood Funktion numerisch maximiert werden, weswegen für gängige (lokale) Optimierungsmethoden ein Startwert für die Parameter α, h, μ, σ angegeben werden muss, von dem aus die Optimierung beginnt. Dabei ist das Finden von geeigneten Anfangswerten für die Optimierung essenziell. Die angesprochene Heuristik $G_\theta \approx \text{Wei}(\alpha + 1)$ ermöglicht es, einen Startwert von α mittels des GEV-Maximum-Likelihood Schätzers zu ermitteln, wobei wir hierfür das Paket *evd*, siehe [Stephenson], nutzen. Es ist möglich, dass der GEV-Schätzer eine Weibull-Beobachtung irrtümlich als Fréchet-Beobachtung interpretiert, was insofern negativ wäre, als dass die G_θ -Verteilung keine Interpretation für $\alpha < 0$ zulässt. Aufgrund der Annahme an α können wir hier jedoch eine Schranke an den Anfangswert für α von $\alpha \leq 15$ setzen, wobei sich letztere Zahl in Simulationen bewährt hat.

Der Lokationsparameter μ lässt sich als rechter Randpunkt interpretieren, siehe Gleichung (3.1), sodass $\mu_{Aw} := x_{1:N} + 0.2$ ein wohlmotivierter Anfangswert ist, wobei der Summand 0.2 ein Störterm ist, damit genügend Abstand zwischen dem Maximum der

Beobachtung und dem Anfangswert für μ gegeben ist. Andernfalls würde die Optimierung außerhalb des Trägers der Log-Likelihood Funktion, siehe Gleichung (3.6), beginnen, was zu Fehlern führen könnte.

Für sinnvolle Werte von h_{Aw} und σ_{Aw} als Anfangswerte ist ein Ansatz, den Strukturwechselpunkt $x = -\sigma/2 + \mu$ zu nutzen. Dort gilt nämlich nach (3.1) die Beziehung $G_\theta(-\sigma/2 + \mu) = \exp(-2h/(\alpha + 1) \cdot 2^{-\alpha-1})$. Stünde uns nun ein Verfahren zur Bestimmung eines Anfangswertes für σ_{Aw} zu Verfügung, so könnten wir mittels der empirischen Verteilungsfunktion $\hat{G}_{\theta,N}(-\sigma_{Aw}/2 + \mu_{Aw})$ als Schätzer für $\exp(-2h/(\alpha + 1) \cdot 2^{-\alpha-1})$ ansetzen. Da wir bereits α_{Aw} gesetzt haben, könnten wir die letzte Gleichung somit nach h auflösen und somit h_{Aw} finden. Dieses Verfahren hängt jedoch maßgeblich von der Leistung der Wahl von α_{Aw} ab. Die oben beschriebene Wahl von α_{Aw} ist jedoch zu unsensibel als dass sie in diesem Verfahren verlässliche Resultate erzielt. Außerdem wurde noch kein Verfahren zur Bestimmung von σ_{Aw} etabliert, so dass dieses Verfahren aktuell nicht datenadaptiv nutzbar ist. Unter Annahmen an die Parameterbereiche, wie in 4.1.1, ist es daher sinnvoller mittels eines Grids von Anfangswerten beziehungsweise Störungen in diesen zu optimieren. Diesen Weg gehen wir in der Wahl der Anfangswerte, wobei wir nur in α_{Aw} und h_{Aw} stören, was sich in Simulationen als effizient herausgestellt hat. Dabei wählen wir als Startwert $\sigma_{Aw} = 2$, da die Optimierungen sich als unsensibel bezüglich der Wahl von σ_{Aw} herausgestellt haben.

Wir haben für die folgenden Optimierungsalgorithmen einen Schätzer implementiert: Den **Broyden-Fletcher-Goldfarb-Shanno ((L)BFGS)**-, den **DIRECT-L**, den **Nelder-Mead** und den **BOBYQA**-Algorithmus. Der LBFGS-Algorithmus ist ein Quasi-Newton Verfahren, also ableitungsbasiert, siehe [Nocedal]. Das heißt, man übergibt eine analytische Ableitung, die in diesem Fall in Gleichung (3.7) gegeben ist und mittels des Paketes *numDeriv*, siehe [Gilbert and Varadhan], auf Korrektheit geprüft wurde. DIRECT-L ist ein globaler Optimierer, welcher die möglichen Parameter in Rechtecke aufteilt und anschließend in diesen Rechtecken weitersucht. Das L steht dafür, dass hier eine Variante genutzt wird, die einen gewissen Zusatzaufwand aufbringt, um nach der globalen Suche ein lokales Minimum auszugeben, siehe [Gablonsky und Kelly]. Die letzten beiden Algorithmen sind lokale und auch ableitungsfreie Optimierer. Letzterer basiert auf einer quadratischen Approximation der Funktion, daher der Suffix QA und ersterer auf einer Optimierung durch Simplexe, siehe [Powell] und [Nelder und Mead]. Damit ist eine gute Funktionalität des BOBYQA Algorithmus' anzunehmen, falls die Zielfunktion zumindest zwei mal differenzierbar ist, wobei der Nelder-Mead Algorithmus auch im Falle nicht-glatte Funktionen funktional sein kann, siehe [Singer und Nelder]. Alle Algorithmen sind Optimierungsmethoden, um nicht-lineare Minimierungsprobleme zu behandeln.

Als Testreferenz nutzen wir die Differenz von der Log-Likelihood Funktion in $\hat{\theta}_{ML}$ und in dem wahren Parameter θ_0 für die jeweiligen Verfahren, wobei ein Algorithmus vorzuziehen ist, falls er einen größeren Wert als die anderen Algorithmen bezüglich dieser Referenz besitzt. Anschließend bilden wir über die verschiedenen α je Verfahren den Mittelwert über obiges $\ell(\hat{\theta}_S) - \ell(\theta_0)$ und die Varianz über letzteren Ausdruck. Ei-

ne niedrige Varianz ist hier wünschenswert, da jene auf ein ähnliches Schätzverhalten für Beobachtungen aus derselben Verteilung hinweist. Zuletzt bilden wir die Differenz aus dem Mittelwert und der Varianz, um beiden Maßen ein Gewicht zuzuordnen.

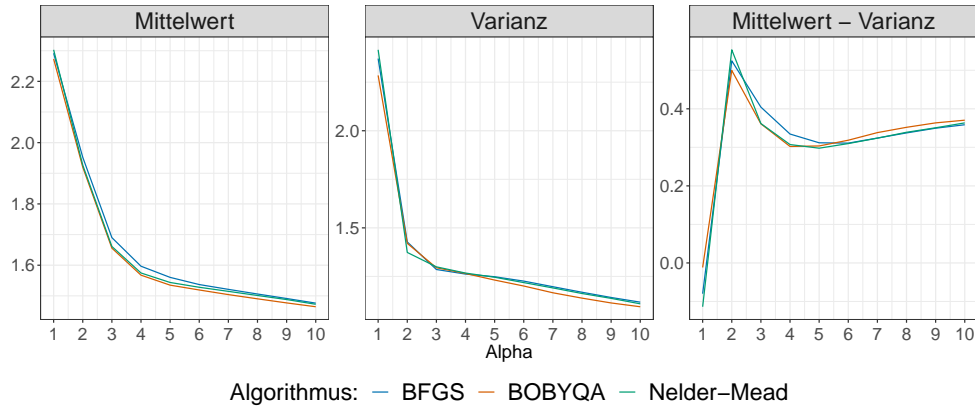


Abbildung 4: Kennzahlen für die Log-Likelihood Funktion in $\hat{\theta}_{ML}$ für die Algorithmen je α , wobei $\alpha \in \{1, \dots, 10\}$ und feste $(h, \mu, \sigma) = (1, 0, 1)$. Datengrundlage ist jene aus Bemerkung 4.1.2.

In Abbildung 4 sind mit Ausnahme des globalen Optimierers die oben erläuterten Kennzahlen für die verschiedenen Optimierungsverfahren je α abgebildet. Die Algorithmen zeigen ähnlich gute Resultate, wobei die Varianz bei den beiden Algorithmen, die Glattheit ausnutzen, für kleine α geringer ist. Der DIRECT-L Algorithmus hat deutlich schlechter als die anderen Algorithmen optimiert, dies ist im Appendix der Abbildung 17 zu entnehmen. Es ist bemerkenswert, dass Maximalstellen der Log-Likelihood Funktion bei einigen Samples auffällig weit von den wahren Parametern entfernt lagen, was an der Gestalt der Funktion liegen kann. Wir wählen für die weitere Optimierung also den Maximum-Likelihood Schätzer, welcher den BFGS Algorithmus nutzt, da er durchschnittlich die besten Werte für die Differenz aus Mittelwert und Varianz aufweist, siehe Abbildung 4. Den Ansatz zuerst mittels einem globalen Algorithmus und anschließend eines lokalen zu optimieren, haben wir hier nicht verfolgt, da die Anfangswertwahl gelungen ist.

Bemerkung 4.1.4 (Datengenerierender Prozess im misspezifiziertem Modell).

Wir arbeiten in der folgenden Situation innerhalb des Stein'schen expliziten Modells. Dies heißt genauer für die Situation aus Theorem 2.3.7, wir setzen $h(\cdot) \equiv h > 0$, $\bar{\alpha}(\cdot) \equiv 0$, $u = 0$ und $v(y) = |y - 1/2|$. Um Daten zu generieren, wählen wir weiterhin die langsam variierende Funktion S als die konstante 1-Abbildung, sodass wir kanonisch

aus der Forderung in (2.20) die Verteilungsfunktionen für $X_{i,n}$ als

$$F_{i,n}(t) = \begin{cases} 0 & , t \leq x_l^* \\ 1 - h \cdot \left(-\frac{|i/n - 1/2|}{n^{1/\alpha}} - t \right)^\alpha & , x_l^* \leq t \leq x_r^* \\ 1 & , x_r^* \leq t \end{cases}$$

wobei $x_r^* = -|i/n - 1/2|/n^{1/\alpha}$ und $x_l^* = x_r^* - h^{-1/\alpha}$, herleiten können.

Um nun Block-Maxima von Zeitreihen der Länge $n \cdot m$ zu simulieren, nutzen wir erneut die Inversionsmethode. Diese Herangehensweise ist hier naheliegend, da die Verteilungsfunktionen $F_{i,n}$ sich exakt invertieren lassen, mithin der Rechenaufwand dadurch verringert werden kann.

Für die weitere Studie simulieren wir für verschiedene α und festem $h = 1$ eine gewisse Anzahl, hier 1.000, von Zeitreihen der Länge $365 \cdot 100$, sodass wir je Parameterkombination $365 \cdot 100 \cdot 1.000 = 36.5 \cdot 10^6$ Beobachtungen generieren. Der Block-Maxima Methode mit Blockgröße $n = 365$ und $m = 100$ Blöcken folgend, erhalten wir je α $100 \cdot 1.000 = 10^5$ Block-Maxima, genauer 500 Samples einer Block-Maxima Beobachtung der (Sample-)Größe 100.

Wir illustrieren die Verteilungskonvergenz von $M_{j,n} \cdot n^{1/\alpha}$ gegen $G_{\alpha,h}$ für verschiedene Werte von α und Blockgrößen n in der Abbildung 5.

Als nächstes wollen wir eine Parameterschätzung innerhalb des Block-Maxima Mo-

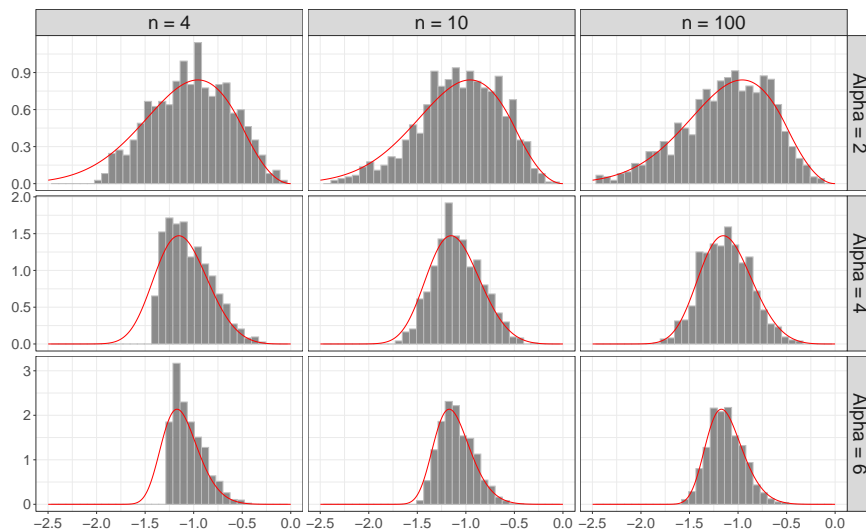


Abbildung 5: Histogramme von jeweils 1000 simulierten standardisierten Block-Maxima für verschiedene Werte von α und Blockgrößen n . Dabei wurden die Parameter als $h = 1$ und $u = 0$ fixiert. Die zugehörige Grenzdichte ist in rot eingezeichnet.

dells aus Subkapitel 3.3 durchführen, wobei wir subasymptotisch arbeiten, das heißt eine Missspezifikation liegt vor.

Bemerkung 4.1.5 (G_θ MLE-Anpassung im misspezifizierten expliziten Stein'schen Modell).

In Bemerkung 4.1.3 haben wir gesehen, dass die Implementierung mittels des LBFGS Algorithmus' die beste Optimierung leistet. Mit diesem Maximum-Likelihood Schätzer passen wir die explizite Stein'sche Verteilung an die gemäß Bemerkung 4.1.4 simulierten Zeitreihen an. Im Appendix werden wir einen kurzen Vergleich von der Variante des Stein'schen Schätzers, die die korrekten Parameter, das heißt $(\alpha, 1, 0, (1/365)^{1/\alpha})$, übergeben bekommt und der datenadaptiven Variante durchführen, siehe Abbildungen 18 und 19. Bevor wir die Güte der simulierten Formparameterschätzer $(\hat{\alpha}_S, \hat{h}_S)$ untersuchen, illustrieren wir in Abbildung 6 exemplarisch mittels QQ-Plots sowohl die Güte der Anpassung als auch die Asymptotik.

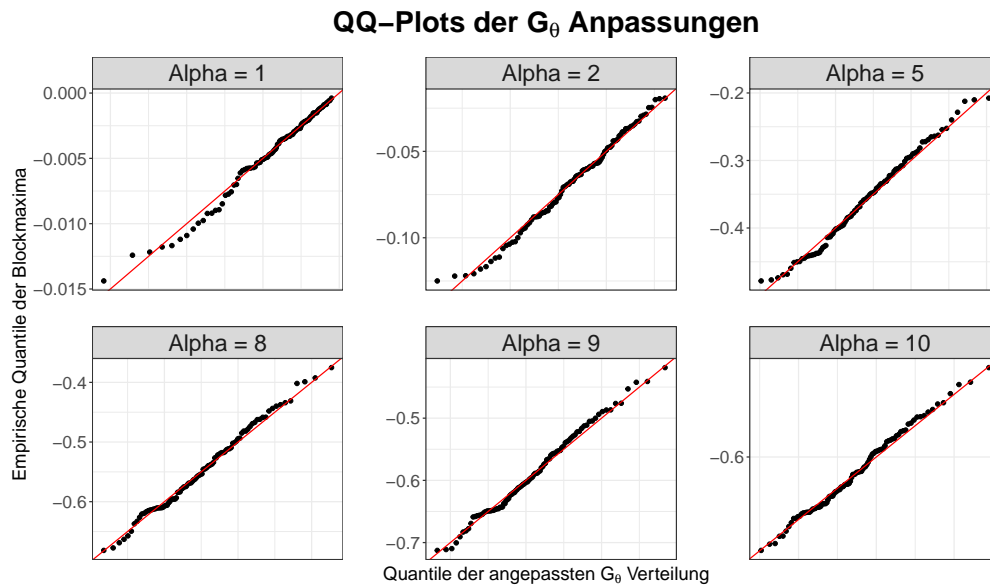


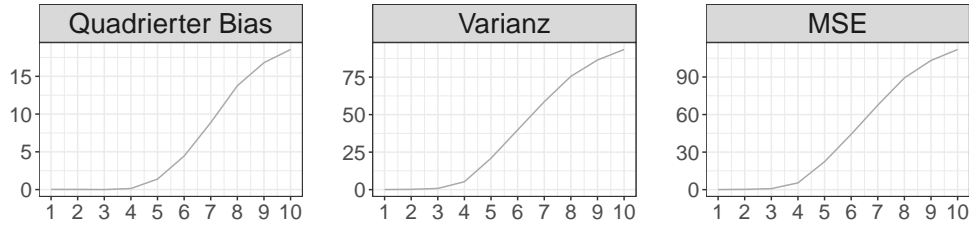
Abbildung 6: QQ-Plots von Quantilen Block-Maximas mit jeweiliger Blockanzahl $m = 100$, Blockgröße $n = 365$ für verschiedene Werte von α und festem Parameter $h = 1$ gegen die Quantile der durch die oben genannte Maximum-Likelihood Methode generierten angepassten G_θ Verteilungen. Eine Winkelhalbierende ist in rot eingezeichnet.

Die Plots aus Abildung 6 zeigen, dass für $\alpha > 1$ zunächst erfolgreich angepasst wurde. So liegt für $\alpha = 1$ ein Teil der Punkte zwar nicht auf der Winkelhalbierenden, jedoch liegen die oberen Quantile auf jener. Insbesondere ist hier zu beachten, dass die Missspezifikation sich nicht visibel schädlich ausgewirkt hat.

In der Abbildung 7 bestimmen wir Kennzahlen des implementierten Schätzers, wobei dieser auf die simulierten Zeitreihen aus Bemerkung 4.1.4 angewandt wurde. Zu bemerken ist hier, dass die Wahrheit $\theta = (\alpha, h, \mu, \sigma)$ streng genommen gar nicht existiert, da wir uns im subasymptotischen Kontext befinden beziehungsweise im misspezifizierten Modell, also die Block-Maxima nur approximativ einer G_θ Verteilung folgen.

Wir werden trotzdem Begriffe wie Bias und MSE in der kanonischen Weise nutzen. Anhand Abbildung 7 erkennen wir, dass Bias und Varianz von $\hat{\alpha}_S$ ab $\alpha = 5$ stark an-

Kennzahlen des Stein'schen ML-Schätzers für α



Kennzahlen des Stein'schen ML-Schätzers für h

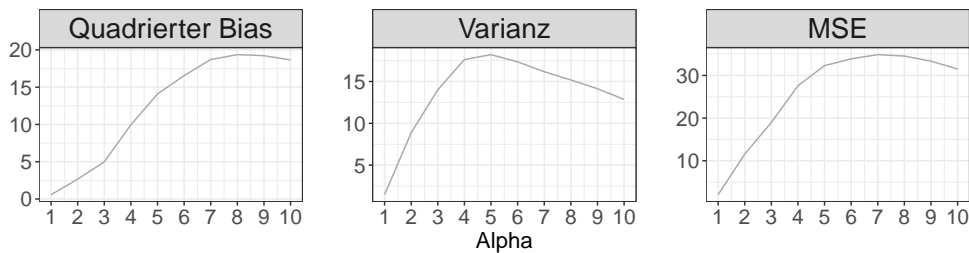


Abbildung 7: Kennzahlen für die Schätzer $(\hat{\alpha}_S, \hat{h}_S)$ für verschiedene $\alpha \in \{1, \dots, 10\}$ und festes $h = 1$. Die Daten wurden gemäß Bemerkung 4.1.4 generiert.

steigen. Der zweite Formparameterschätzer \hat{h}_S steigt generell schneller an, fällt jedoch auch mit wachsendem α und fluktuiert generell nicht so stark wie $\hat{\alpha}_S$, was jedoch auch an der Skalierung von h und an der fehlenden Variation von h liegt.

Die angesprochene Ähnlichkeit der Verteilung $G_{\alpha,h}$ zu einem Wei $(\alpha + 1)$ Typ, vergleiche mit den Ausführungen unter Gleichung (2.38), motiviert zu vergleichen wie gut eine Weibull Anpassung im exakten Stein'schen Modell Kenngrößen der unbekannten Verteilung der Block-Maxima schätzt. Man beachte, dass damit keine Parameter der G_θ -Verteilung geschätzt werden können sondern nur jene einer Weibull-beziehungsweise GEV-Verteilung. Jedoch ist es von natürlichem Interesse, ob die Heuristik, dass asymptotisch $G_{\alpha,h}$ sich in Verteilung wie ein Wei $(\alpha + 1)$ Typ verhält, bereits subasymptotisch in der Parameterschätzung zu erkennen ist.

Bemerkung 4.1.6 (GEV Anpassung im exakten Stein'schen Modell).

Ausgehend vom selben Rahmen wie in Bemerkung 4.1.5 führen wir eine Maximum-Likelihood Schätzung durch, wobei wir hier eine GEV-Verteilung an die Daten anpassen.

Für die Implementierung nutzen wir das R-Paket *evd*, siehe [Stephenson], in welchem die Funktion *fgev* eine Anpassung der GEV-Verteilung an gegebene Daten mittels eines Maximum-Likelihood Schätzers ermöglicht. Dabei wurde diese Funktion mittels des BFGS Algorithmus' innerhalb der *optim*-Funktion programmiert, wobei der Schätzer datenadaptiv ist, da Anfangswerte durch Momentenschätzer bestimmt werden. Wie

in Annahme 4.1.1 erwähnt, werden wir vom Schätzer als größer $-1/31$ geschätzte Formparameter $\hat{\xi}$ auf $\hat{\xi} = -1/31$ begrenzen, damit die Umrechnung in α nicht unter Umständen singuläre Werte produziert.

Anhand der QQ-Plots aus Abbildung 8 erkennen wir, dass sich die Anpassung

QQ-Plots der GEV-Anpassungen

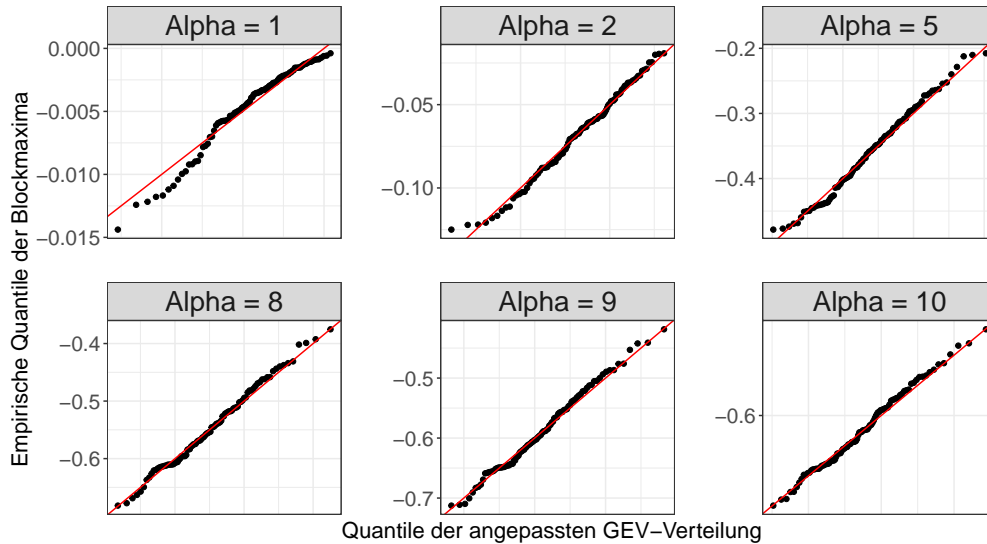


Abbildung 8: QQ-Plots von Quantilen Block-Maximas mit jeweiliger Blockanzahl $m = 100$, Blockgröße $n = 365$ für verschiedene Werte von α und festem Parameter $h = 1$ gegen die Quantile der durch die oben genannte Maximum-Likelihood Methode generierten angepassten G_θ Verteilungen.

durch eine GEV Maximum-Likelihood Schätzung ähnlich wie eine auf Basis der G_θ -Verteilung verhält. Insbesondere fällt auf, dass die Qualität der Anpassung nicht visibel fehlerhaft ist. Dies ist insofern bemerkenswert, da die Grenzverteilung keine Weibullverteilung ist. Die Nähe dieser Grenzverteilung zu einem Weibulltypen wird dadurch erneut verdeutlicht.

In Abbildung 9 sind die Kennzahlen von dem GEV-Maximum-Likelihood Schätzer abgebildet, wobei die Umrechnung $\alpha = -1/\xi$ genutzt wurde. Man beachte, dass hier keine kanonische Angabe eines Bias' möglich ist, da die Beobachtungen weder subasymptotisch noch approximativ einer GEV-Verteilung folgen.

Der Abbildung entnehmen wir, dass die Variation von dem Parameter α innerhalb der Datengenerierung im expliziten Stein'schen Modell eine gleichgerichtete Variation des geschätzten Weibullparameters bewirkt. Außerdem kann man erkennen, dass die vor dieser Bemerkung genannte Heuristik für größere α sogar durch die Maximum-Likelihood Schätzung zu erahnen ist.⁷

⁷ Für einen Vergleich mit dem Benchmark Stein Schätzer siehe Abbildung 18 im Appendix.

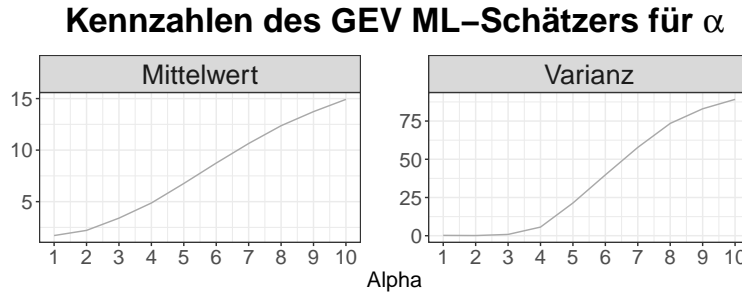


Abbildung 9: Kennzahlen des Schätzers $\hat{\alpha}_{\text{GEV}}$ für verschiedene $\alpha \in \{1, \dots, 10\}$ und festem $h = 1$. Die Daten wurden gemäß Bemerkung 4.1.4 generiert.

Nachdem wir innerhalb des Rahmens von Bemerkung 4.1.4 sowohl eine Anpassung mittels der G_θ -Verteilung als auch der GEV-Verteilung durchgeführt haben, wollen wir die dadurch gemäß Bemerkung 3.2.4 beziehungsweise 3.3.4 geschätzten Quantile untereinander und insbesondere mit dem *wahren* Quantil der Block-Maxima vergleichen. Dabei ist von Interesse, ob die GEV-Anpassung auch innerhalb des expliziten Stein'schen Modells extreme Quantile ähnlich gut wie die Anpassung mittels des naheliegenden Maximum-Likelihood Schätzers der Asymptotik zu schätzen vermag.

Bemerkung 4.1.7 (Vergleich der Quantile).

Abbildung 10 zeigt Kennzahlen der 25%- und 99%-Quantilschätzer, wobei jene auf den in Bemerkung 4.1.6 und 4.1.5 vorgestellten Implementierungen basieren. Diese Werte vergleichen wir durch Bestimmung des quadrierten Bias mit den *wahren* Quantilen, welche aus einer Simulation mit einer sehr großen Blockanzahl $m = 10^6$ hervorgehen, vergleiche mit Bemerkung A.2.9 im Appendix.

Beide Schätzer unterscheiden sich in ihrer Qualität nur marginal. In der Schätzung des 25%-Quantils ist zu erkennen, dass der Bias des Stein Maximum-Likelihood Schätzers für wachsende α geringer als derjenige des GEV Schätzers ist. Überraschend ist die Tatsache, dass in beiden Quantilen die Varianzen fast identisch sind, obwohl in der Situation des Stein'schen Schätzers vier und nicht drei Parameter geschätzt werden. Die Tatsache, dass das 25%-Quantil mit geringerem Bias durch den Stein'schen Schätzer geschätzt wird, lässt sich durch die andere Struktur im Bereich $x < -\sigma/2 + \mu$ erklären, siehe Gleichung (3.1). In diesem Bereich entspricht die Verteilung keiner Weibull-Verteilung mehr, sodass ein größerer Bias in niedrigeren Quantilen auftritt. Sind geringere Quantile zu schätzen, empfiehlt es sich also den Stein'schen Schätzer zu wählen. Die Ergebnisse der 99%-Quantils Schätzung deuten auf nahezu identisches Verhalten hin, da die Kennzahlen dort sich maximal um Größenordnung 10^{-4} unterscheiden. Ein möglicher Grund dafür ist die unter Gleichung (2.38) angesprochene Ähnlichkeit von dem linken Tail zu einem $\text{Wei}(\alpha) -$ und dem rechten zu einem $\text{Wei}(\alpha + 1)$ Typ.⁸

⁸ Für einen Vergleich mit dem Benchmark Schätzer siehe Abbildung 19 Appendix.

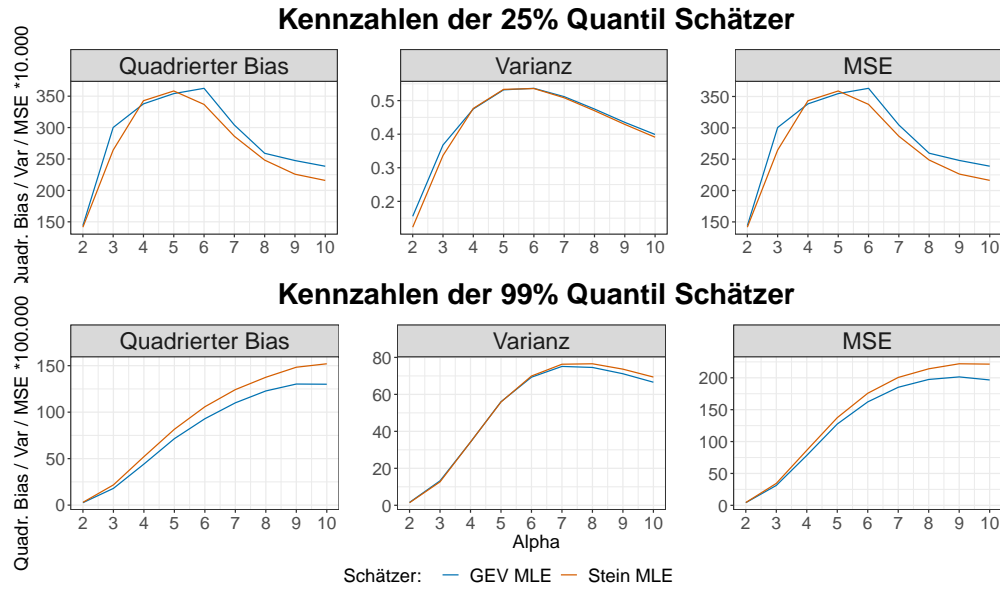


Abbildung 10: Skalierte Kennzahlen vom GEV- und vom Stein'schen Maximum-Likelihood Schätzer für das 99% Quantil als Funktion von α abgetragen, wobei $h = 1$ fest ist. Die Datengrundlage ist diejenige aus Bemerkung 4.1.4, das heißt je α 100 Blöcke der Größe 365. Wir haben mit 1000 Samples simuliert.

4.2 Eine Simulationsstudie im Weibull-Modell

In Analogie zu dem Subkapitel 4.1 wollen wir nun dieselben Untersuchungen in einem Weibull-Modell durchführen. Dabei nutzen wir also erneut die Block-Maxima Methode. Motivierend hierbei ist es zu überprüfen, ob sich die Nähe der expliziten Stein'schen Verteilung zu einem Weibull-Typen innerhalb der Schätzungen symmetrisch verhält, genauer: Die durchgeführte Studie in 4.1.7 hat durch Simulationen indiziert, dass eine GEV Maximum-Likelihood Anpassung im Rahmen des Stein'schen Modells für die Schätzung von extremen Quantilen nicht fehlschlägt ähnliche Qualität wie die Stein'sche Anpassung hat. Im Folgenden wollen wir die Rollen vertauschen und daher aus einer im Anziehungsbereich der Weibull-Verteilung liegenden Verteilung simulieren.

Annahme 4.2.1 (Parameterbereiche im Weibull-Modell).

Wie in dem expliziten Stein'schen Modell 4.1.1 stellen wir Annahmen an die Parameter innerhalb des Modells. Von nun an bezeichnen wir mit β den Weibull-Formparameter, um ihn vom Parameter α der G_θ -Verteilung abzugrenzen. Der Heuristik $\alpha + 1 \approx \beta$ folgend nehmen wir

$$1.3 \leq \beta \leq 31, \quad -100 \leq \mu \leq 100, \quad 10^{-8} \leq \sigma \leq 10$$

an. Dies übersetzt sich zu $-0.77 \leq \xi \leq -0.03$ als mögliches Schätzergebnis für den Formparameter $\xi = -1/\beta$.

Im Gegensatz zu der Simulationsstudie aus Abschnitt 4.1 werden wir hier nicht das korrekt spezifizierte Modell untersuchen, da wir bereits eine effizient leistende Implementierung des Maximum-Likelihood Schätzers für das explizite Stein'sche Modell ausgewählt haben, siehe Abbildung 4, und in Anwendungen Beobachtungen sich lediglich im subasymptotischen Rahmen befinden. Sprechen wir im Folgenden vom *Weibull-Modell*, so ist damit immer das misspezifizierte Modell gemeint. Um eine Vergleichbarkeit zu der Studie aus Abschnitt 4.1 wahren zu können, nutzen wir die Implementierungen aus dem vorherigen Kapitel.

Bemerkung 4.2.2 (Datengenerierung im Weibull-Modell).

Die Basis der Studie werden unabhängig und identisch verteilte Zeitreihen bilden, so dass insbesondere Stationarität der Zeitreihe gegeben ist, was einen ersten Unterschied zu der Studie aus 4.1 darstellt.

Die Verteilungsfunktionen für die $X_{i,n}$ seien für $\beta > 0$ durch

$$F_{i,n}(x) = F(x) = \begin{cases} 1 - (-x)^\beta & , -1 < x < 0 \\ 0 & , \text{sonst} \end{cases}$$

gegeben. Die Wahl der $F_{i,n}$ wurde derart getroffen, dass sie im Max-Anziehungsbereich der Weibull-Verteilung liegt, für einen Beweis siehe A.2.4.

Die Quantilfunktion lässt sich erneut explizit bestimmen, sodass die Inversionsmethode hier anwendbar ist. Für die weitere Studie simulieren wir für verschiedene $\beta > 0$ 1000 Zeitreihen der Länge $365 \cdot 100$, wobei wieder $n = 365$ die Blockgröße und $m = 100$ die Blockanzahl darstellen. Die Lokations- beziehungsweise Skalenparameter setzen wir fest als $b = 0$ und $a = 1$ an. Somit erhalten wir 1000 Samples der Größe 100 und je β werden $365 \cdot 100 \cdot 1000 = 36.5 \cdot 10^6$ Beobachtungen simuliert.

In Abbildung 11 illustrieren wir die Verteilungskonvergenz von $M_{j,n} \cdot n^{1/\alpha}$ gegen die $\text{Wei}(\alpha)$ – Verteilung, welche im Appendix A.2.4 bewiesen wird. Wir erkennen, dass bereits für kleine n die Asymptotik eintritt und wenig Fluktuation auftritt. Dies ist vielversprechend für die folgenden Anpassungen.

Bemerkung 4.2.3 (GEV Anpassung im Weibull Modell).

Wir passen die GEV/Wei-Verteilung mittels Maximum-Likelihood Schätzung an die Daten aus Bemerkung 4.2.2 an. Zunächst betrachten wir exemplarisch QQ-Plots in Abbildung 12, um die Güte der Anpassung qualitativ einschätzen zu können. Es ist anhand der Plots zu erkennen, dass die Anpassung nicht misslungen ist, sodass eine weitere Studie der Schätzleistung indiziert ist.

Anschließend bestimmen wir in Abbildung 13 Kennzahlen vom Formparameterschätzer $\hat{\beta}_{\text{GEV}}$. Dabei verhalten sich die Kennzahlen ähnlich wie in dem Stein-Modell mit isotonem MSE, wobei die Varianz dominiert.

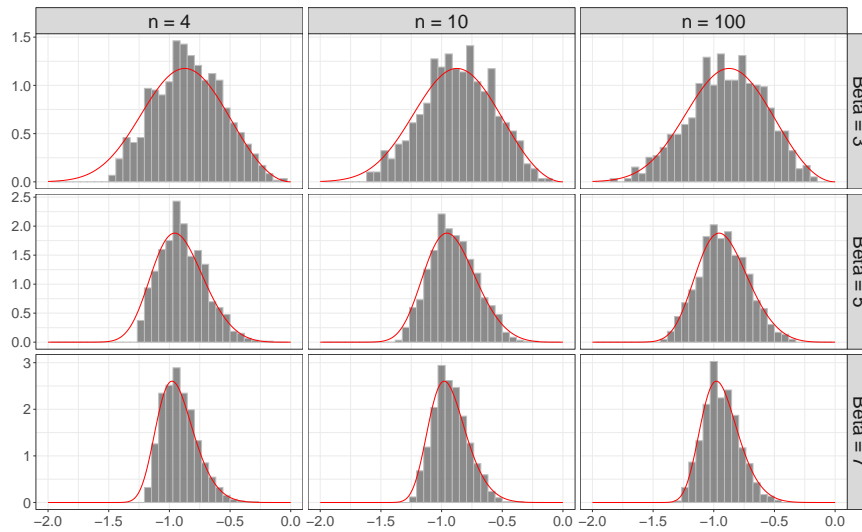


Abbildung 11: Histogramme von jeweils 1000 simulierten standardisierten Block-Maxima für verschiedene Werte von β und Blockgrößen n . Die zugehörige Grenzdichte der Wei (β) – Verteilung ist in rot eingezeichnet.

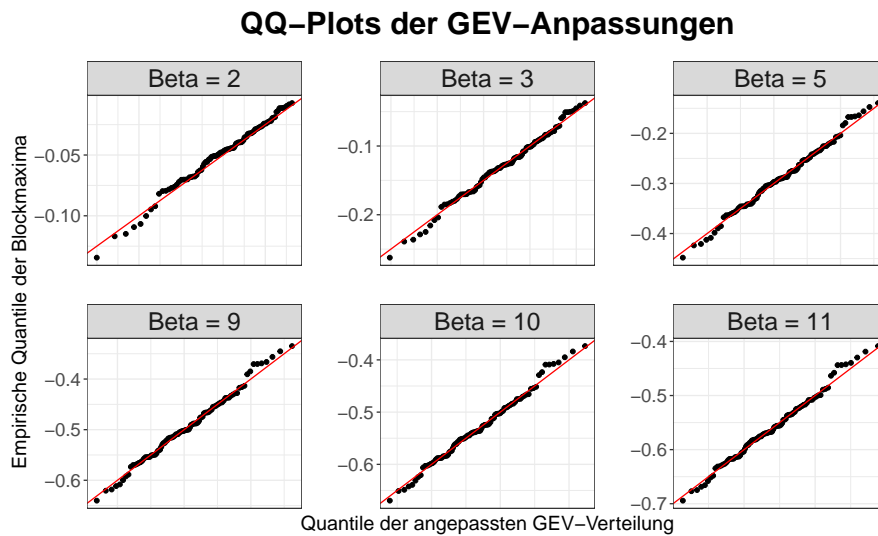


Abbildung 12: QQ-Plots von Quantilen Block-Maximas mit jeweiliger Blockanzahl $m = 100$, Blockgröße $n = 365$ für verschiedene Werte von β gegen die Quantile der durch die oben genannte Maximum-Likelihood Methode generierten angepassten G_θ Verteilungen. Die Parameter b, a sind fest als 0 beziehungsweise 1 gewählt.

Kennzahlen des GEV ML-Schätzers für β

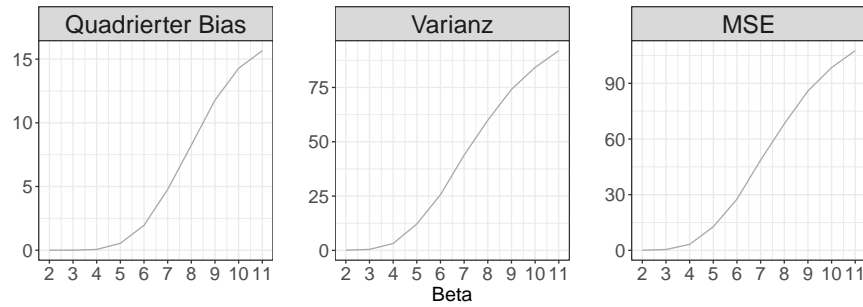


Abbildung 13: Auf Basis der Daten aus Bemerkung 4.2.2 bestimmte Kennzahlen des Formparameterschätzers $\hat{\beta}_{\text{GEV}}$. Dabei variiert β in $\{2, \dots, 11\}$ und die Parameter b, a sind fest als 0 beziehungsweise 1 gewählt.

Bemerkung 4.2.4 (G_θ MLE-Anpassung im Weibull Modell).

Wie auch für die Anpassung mittels GEV-Verteilung betrachten wir auch für die G_θ -Verteilung in Abbildung 14 QQ-Plots, um exemplarische Qualität der Schätzungen feststellen zu können. Dabei ist aufgrund der Verteilungsähnlichkeiten eine gelungene Anpassung naheliegend. Vergleicht man diese QQ-Plots mit jenen aus Abbildung 12,

QQ-Plots der G_θ Anpassungen

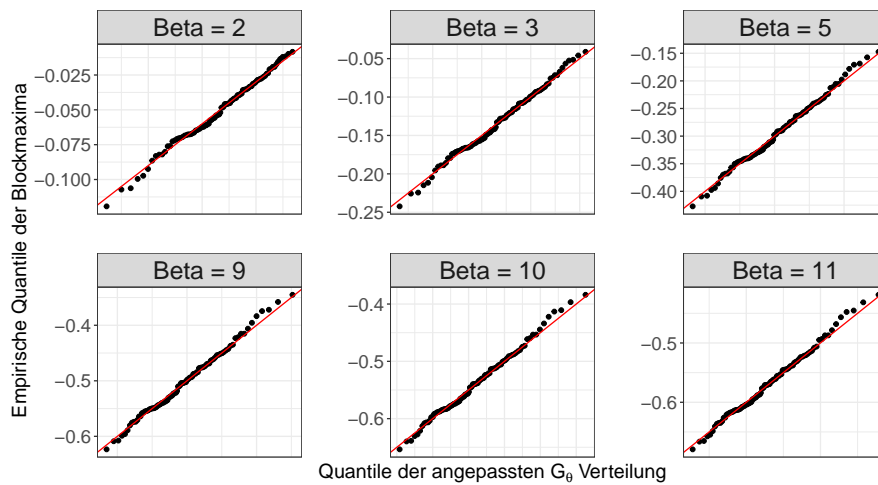


Abbildung 14: QQ-Plots von Quantilen Block-Maximas mit jeweiliger Blockanzahl $m = 100$, Blockgröße $n = 365$ für verschiedene Werte von β gegen die Quantile der durch die oben genannte Maximum-Likelihood Methode generierten angepassten G_θ Verteilungen. Die Parameter b, a sind fest als 0 beziehungsweise 1 gewählt.

so kann visuell keine als besser klassifiziert werden. Die Qualitäten der Anpassungen nach Maßgabe der QQ-Plots sind jedoch nicht zu unterscheiden, sodass die genauere

Untersuchung des Schätzverhaltens indiziert wird.

Daher sind in Abbildung 15 Mittelwert und Varianz der Schätzer $\hat{\alpha}_S$ und \hat{h}_S abgebildet. Ein Bias und damit auch ein MSE sind hier nicht kanonisch definierbar, da wir uns in dem Weibull-Modell befinden und die Parameter α und h demnach gar nicht im Modell exakt wiederzufinden sind. Wir erinnern an die Heuristik $\beta \approx \alpha + 1$. In den Plots ist diese für größere β zu erkennen, wobei für kleinere α die Tendenz $\hat{\alpha} \approx \beta$ gilt. Wie in den Plots 7, 9 und 13 ist auch hier die Varianz isoton in der Variation von β . Auch die Größenordnung der Kennzahlen ist ähnlich. Obwohl der Parameter h im Weibull-Modell keine Entsprechung besitzt, verhält sich der Schätzer \hat{h}_S hier wie Im expliziten Stein'schen Modell, siehe Abbildung 9.

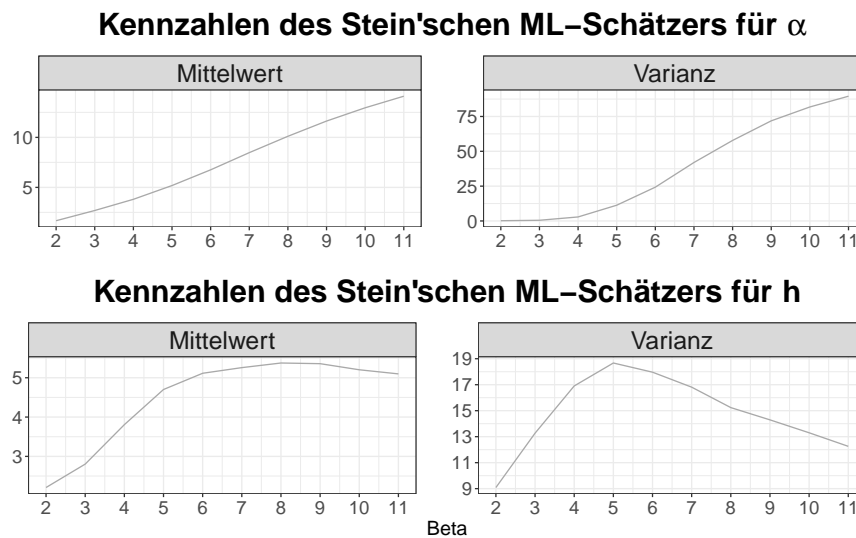


Abbildung 15: Kennzahlen für die Schätzer $(\hat{\alpha}_S, \hat{h}_S)$ für verschiedene $\beta \in \{2, \dots, 11\}$. Die Daten wurden gemäß Bemerkung 4.2.2 generiert und die Parameter b, a fest als 0 beziehungsweise 1 gewählt.

Bemerkung 4.2.5 (Vergleich der Quantilschätzer im Weibull-Modell).

Zuletzt wollen wir analog zu Abbildung 10 die Quantilschätzer vergleichen. Zu diesem Zweck sind in Abbildung 16 diese für die beiden Schätzer je β abgebildet, wobei die *exakten* Quantile gemäß Bemerkung A.2.9 mit $m = 10^6$ bestimmt wurden.

Es fällt auf, dass der Stein Schätzer verzerrter als der GEV Schätzer ist, wobei sich die Differenz in einer Größenordnung von 10^{-6} bis 10^{-4} bewegt. Für beide Quantile dominiert die Varianz in Bereichen von 10^{-3} bis 10^{-4} , was im Fall des Stein-Modells umgekehrt war, siehe Abbildung 10. Überraschend ist dabei die geringere Varianz des Stein Schätzers, was jedoch auch an der Implementierung über ein Gitter von Anfangswerten liegen mag. Insgesamt lässt sich auch hier kein evident besserer Schätzer ermitteln, da die Differenzen des MSE in einer Größenordnung 10^{-6} bis 10^{-4} liegen. Ein möglicher Grund dafür ist die Ähnlichkeit beider Verteilungsklassen, siehe Bemerkung 3.3.4.

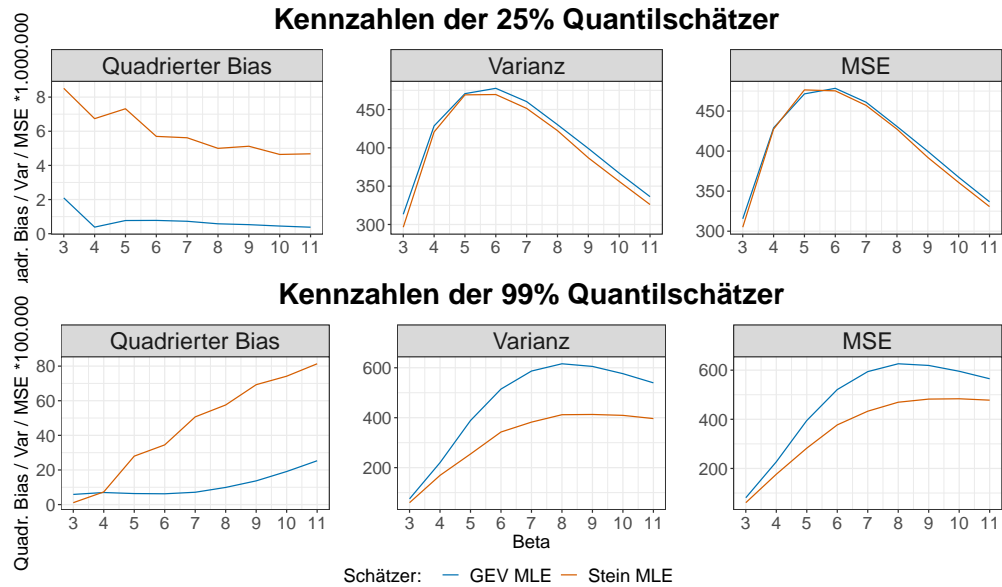


Abbildung 16: Skalierte Kennzahlen vom GEV- und vom Stein'schen Maximum-Likelihood Schätzer für die 25- und 99% Quantile als Funktion von β abgetragen. Die Datengrundlage ist diejenige aus Bemerkung 4.2.2, das heißt je β 100 Blöcke der Größe 365. Wir haben mit 1000 Samples simuliert.

Fazit und Aussicht

Innerhalb des Kapitels 2 wurden zwei Grenzwertsätze aus [Stein] vorgestellt und ausgearbeitet, die von Temperaturuntersuchungen motiviert waren. Dabei haben wir im Rahmen von Satz 2.3.14 für die zweite Asymptotik gezeigt, dass eine andere Grenzverteilungsklasse als die der GEV-Verteilungen resultiert. Diese Klasse \mathcal{G} ist für bestimmte Wahlen von Funktionen bezüglich des Tailverhaltens weit von einer GEV-Verteilung entfernt, kann jedoch auch starke Ähnlichkeiten aufweisen, was wir in Satz 2.3.16 und Beispiel 2.3.8 sahen. Daneben haben wir für beide Asymptotiken in Satz 2.2.14 und 2.3.17 starke Gesetze etabliert, wobei wir nur den Beweis für die erste Asymptotik vollständig ausführten.

Die angesprochenen Ähnlichkeiten motivierten eine statistische Behandlung der Block-Maxima Methode im Bezug auf den zweiten Grenzwertsatz aus Theorem 2.3.7. Hierbei wurden zunächst Ähnlichkeiten und Unterschiede zu der klassischen Block-Maxima Methode aufgeführt. Wir untersuchten das explizite Stein'sche Modell auf Regularität im Strukturwechsellpunkt und stellten so einen weiteren Unterschied fest. Zuletzt führten wir detaillierte Simulationsstudien innerhalb zwei Block-Maxima Settings durch, welche indizierten, dass der von uns vorgestellte (explizite) Stein'sche Maximum-Likelihood Schätzer in beiden misspezifizierten Modellen ähnlich gute Resultate liefert wie der gut untersuchte GEV Maximum-Likelihood Schätzer.

Es gibt einige interessante Untersuchungspunkte, welche in dieser Arbeit nicht behandelt wurden wie zum Beispiel theoretische Resultate über die asymptotische Güte des (expliziten) Stein'schen Maximum-Likelihood Schätzers. So ist es naheliegend, dass er ähnliche wie der GEV-Schätzer erfüllt, was aber noch nicht bewiesen wurde. Außerdem haben wir eine zweite Subklasse von \mathcal{G} in Beispiel 2.3.15 kennengelernt, welche ein grundlegend anderes Tailverhalten besaß. Dies motiviert die Frage, ob sich für ein daraus entstehendes Modell Schätzer etablieren lassen, welche ähnliche Güte besitzen wie die des expliziten Stein'schen Modells.

Appendix

A.1 Verteilungen

Definition A.1.1 (Dirac-Verteilung).

Eine Y -wertige Zufallsvariable X heißt **Dirac-verteilt**, falls ein $y \in Y$ existiert mit $X = y$ fast-sicher, das heißt $\mathbb{P}^X = \delta_y$.

Definition A.1.2 (Geometrische Verteilung).

Eine \mathbb{N}_0 -wertige Zufallsvariable X heißt **geometrisch zum Parameter p_0 verteilt**, falls sie die Zähldichte

$$p(k) = (1 - p_0)^k p_0 \cdot \mathbb{1}_{\mathbb{N}_0}(k)$$

besitzt, wobei $p_0 \in]0, 1[$. Wir schreiben dafür **Geom(p_0)**.

Definition A.1.3 (Stetige Gleichverteilung).

Eine \mathbb{R}^d -wertige Zufallsvariable heißt **stetig gleichverteilt zum Parameter A** , falls sie die Lebesgue-Dichte

$$p(x) = \frac{1}{\mathbb{A}(A)} \cdot \mathbb{1}_A(x)$$

besitzt, wobei $A \subset \mathcal{B}(\mathbb{R}^d)$. Wir schreiben dafür **$\mathcal{U}(A)$** und im eindimensionalen Fall schreiben wir **$\mathcal{U}([a, b])$** anstelle von $\mathcal{U}([a, b])$.

Definition A.1.4 (Normalverteilung).

Eine \mathbb{R}^d -wertige Zufallsvariable heißt **standard normalverteilt**, falls sie die Lebesgue-Dichte

$$p(x) = \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{\sum_{i=1}^d x_i^2}{2}\right)$$

besitzt.

Eine \mathbb{R}^d -wertige Zufallsvariable X heißt **normalverteilt mit Erwartungswert μ und Kovarianzmatrix Σ** , falls für eine standard-normalverteilte Zufallsvariable N die Beziehung

$$\mathbb{P}^X = \mathbb{P}^{AX+\mu}$$

gilt, wobei $A \in \mathbb{R}^{d \times d}$ die positiv semidefinite Wurzel von Σ ist und $b \in \mathbb{R}^d, \Sigma \in \mathbb{R}^{d \times d}$ positiv semidefinit. Wir schreiben dafür **$\mathcal{N}_d(\mu, \Sigma)$** und für den eindimensionalen Fall **$\mathcal{N}(\mu, \sigma^2)$** , wobei Σ mit $\sigma^2 \geq 0$ identifiziert wurde.

Definition A.1.5 (Exponentialverteilung).

Eine \mathbb{R} -wertige Zufallsvariable heißt **exponential verteilt zum Parameter λ** , falls sie die Lebesgue-Dichte

$$p(x) = \lambda \exp(-\lambda \cdot x) \cdot \mathbb{1}_{\mathbb{R}^+}(x)$$

besitzt, wobei $\lambda > 0$. Wir schreiben dafür **Exp(λ)**.

Definition A.1.6 (Pareto-Verteilung).

Eine \mathbb{R} -wertige Zufallsvariable heißt **Pareto-verteilt zum Parameter α** , falls sie die

Lebesgue-Dichte

$$p(x) = \alpha x^{-(\alpha+1)} \cdot \mathbb{1}_{[1,\infty[}(x)$$

besitzt, wobei $\alpha > 0$. Wir schreiben dafür $\mathbf{Par}(\alpha)$.

A.2 Hilfsresultate

Theorem A.2.1 (Delta-Methode).

Sei $\Phi: D \subset \mathbb{R}^k \rightarrow \mathbb{R}^m$ differenzierbar in $\theta \in D$ für $k, m \in \mathbb{N}$ und T_n, T Zufallsvariablen, die Werte in D annehmen. Gilt für eine gegen unendlich bestimmt divergierende Folge r_n , die Verteilungskonvergenz $r_n \cdot (T_n - \theta) \xrightarrow{d} T$, so folgt

$$r_n \cdot (\Phi(T_n) - \Phi(\theta)) \xrightarrow[n \rightarrow \infty]{d} \Phi'(\theta) \cdot T,$$

wobei $\Phi'(x)$ die Jacobi-Matrix von Φ in x bezeichne.

Beweis. Siehe [Van der Vaart, Theorem 3.1] ◇

Bemerkung A.2.2 (Formel für den Mean Squared Error).

Bezeichnen wir mit $\text{Var}_\theta(\hat{\gamma}) := (\text{Var}_\theta(\hat{\gamma}^{(1)}), \dots, \text{Var}_\theta(\hat{\gamma}^{(k)}))^T$ die komponentenweise Varianz einer k -dimensionalen Zufallsvariable, so gilt

$$\text{MSE}_\theta(\hat{\gamma}) = \|\text{Bias}_\theta(\hat{\gamma})\|_2^2 + \|\text{Var}_\theta(\hat{\gamma})\|_1.$$

Beweis. Wir rechnen

$$\begin{aligned} \text{MSE}_\theta(\hat{\gamma}) &= \mathbb{E}_\theta \left[\|\hat{\gamma} - \gamma(\theta)\|_2^2 \right] = \mathbb{E}_\theta \left[\sum_{i=1}^k (\hat{\gamma}^{(i)} - \gamma^{(i)}(\theta))^2 \right] = \sum_{i=1}^k \text{MSE}_\theta(\hat{\gamma}^{(i)}) \\ &= \sum_{i=1}^k \text{Bias}_\theta(\hat{\gamma}^{(i)})^2 + \text{Var}_\theta(\hat{\gamma}^{(i)}) = \|\text{Bias}_\theta(\hat{\gamma})\|_2^2 + \|\text{Var}_\theta(\hat{\gamma})\|_1. \end{aligned}$$

□

Lemma A.2.3 (Fast-sichere Randpunktskonvergenz).

Sei $\{X_{i,n}: 1 \leq i \leq n, n \in \mathbb{N}\}$ ein zeilenweise unabhängiges und im gesamten Schema identisch verteiltes Dreiecksschema von reellwertigen Zufallsvariablen. Dann gilt mit der Notation $M_n(x) := \max\{X_{n,1}, \dots, X_{n,n}\}$

$$M_n \xrightarrow{f.s.} x^*,$$

wobei hier x^* den rechten Randpunkt der Verteilungen bezeichnet.

Beweis. Wir zeigen, dass die Konvergenz mit Wahrscheinlichkeit Null nicht auftritt.

$$\begin{aligned}
\mathbb{P}\{M_n \not\rightarrow x^*\} &= \mathbb{P}\left\{\liminf_n M_n < x^*\right\} = \mathbb{P}\left(\bigcup_{M \in \mathbb{N}} \left\{\liminf_n M_n < x^* - 1/M\right\}\right) \\
&= \lim_M \mathbb{P}\left\{\liminf_n M_n < x^* - 1/M\right\} \leq \lim_M \mathbb{P}\left(\liminf_n \{M_n \leq x^* - 1/M\}\right) \\
&\leq \lim_M \liminf_n \mathbb{P}\{M_n < x^* - 1/M\} = \lim_M \liminf_n \left(F(x^* - 1/M)\right)^n \\
&= 0,
\end{aligned}$$

wobei die erste Gleichheit aus der Tatsache $M_n \leq x^*$ fast-sicher, die dritte aus der Stetigkeit von unten und die letzte aus der Tatsache $F(x^* - 1/M) < 1$ für festes $M \in \mathbb{N}$ folgt. \square

Beispiel A.2.4.

Sei $\{X_{i,n} : 1 \leq i \leq n, n \in \mathbb{N}\}$ ein zeilenweise unabhängiges und identisch verteiltes reellwertiges Dreiecksschema von Zufallsvariablen mit

$$F_{i,n}(x) = F(x) \begin{cases} 1 - (-x)^\alpha & , -1 < x < 0 \\ 0 & , \text{sonst} \end{cases}$$

für ein $\alpha > 0$. Dann gilt mit der Notation $M_{j,n}(x) := \max\{X_{j,1}, \dots, X_{j,n}\}$ die Verteilungskonvergenz

$$\frac{M_{j,n}}{n^{-1/\alpha}} \xrightarrow[n \rightarrow \infty]{d} \text{Wei}(\alpha)$$

für festes $j \in \mathbb{N}$.

Beweis. Für den Fall $x \geq 0$ ist die Konvergenz beider Verteilungsfunktionen gegen 0 klar. Wir rechnen für $x < 0$

$$\mathbb{P}\{M_{j,n} \cdot n^{1/\alpha} \leq x\} = F^n\left(x \cdot n^{-1/\alpha}\right) = \left(1 - \left(-\frac{x}{n^{1/\alpha}}\right)^\alpha\right)^n \xrightarrow{n} \exp\{-(-x)^\alpha\},$$

wobei wir im ersten Schritt die zeilenweise Unabhängigkeit ausgenutzt haben. Damit stimmen Grenzwert der Verteilungsfunktion und die der $\text{Wei}(\alpha)$ -Verteilung überein. Die Behauptung folgt aus dem Portmanteau-Theorem. \square

Theorem A.2.5 (Poisson-Approximation).

Seien $(p_{j,n})_{j,n \in \mathbb{N}} \subset [0, 1]$ Zahlen, für $j \in \mathbb{N}$ seien unabhängige Folgen von Zufallsvariablen $(X_{j,n})_{n \in \mathbb{N}}$ mit $X_{j,n} \sim \text{Ber}(p_{j,n})$ gegeben und $\Delta := \{X_{j,n} : j, n \in \mathbb{N}\}$ das induzierte Dreiecksschema mit unendlichen Zeilen.

Gilt zusätzlich

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} p_{j,n}^2 = 0$$

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} p_{j,n} \in]0, \infty[,$$

so folgt die Verteilungskonvergenz der Summe

$$\sum_{j=1}^{\infty} X_{j,n} \xrightarrow[n \rightarrow \infty]{d} \text{Poi} \left(\lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} p_{j,n} \right).$$

Beweis. Siehe [Klenke, Satz 3.7]

◇

Für die *reguläre Variation* ist folgende Konvention nützlich.

Definition A.2.6.

Eine Abbildung reellwertige Abbildung f heißt **letztendlich auf \mathbb{R} definiert**, falls ein $A > 0$ existiert, sodass f auf $]A, \infty[$ definiert ist.

Definition A.2.7 (Reguläre Variation).

Eine letztendlich auf \mathbb{R} definierte Funktion heißt **regulär varierend mit Index ρ** , falls

$$\lim_{t \uparrow \infty} \frac{f(\lambda \cdot t)}{f(t)} = \lambda^\rho, \quad \lambda > 0.$$

Wir schreiben $f \in RV(\rho)$ und im Falle $\rho = 0$ spricht man auch von **langsamer Variation**.

Wir geben ohne Beweis eine unvollständige Liste von Eigenschaften regulär variierender Funktionen an. Für eine ausschöpfende Behandlung dieser verweisen wir auf [Bingham et al.].

Satz A.2.8 (Eigenschaften regulär variierender Funktionen).

Sei $f \in RV(\alpha)$, $g \in RV(\beta)$.

1. Die Funktion f lässt sich letztendlich als Produkt $f(x) = x^\alpha \cdot L(x)$ mit einer langsam variierenden Funktion $L(\cdot)$ schreiben.
2. Es gilt $f \cdot g \in RV(\alpha + \beta)$.
3. Für beliebige Kompakta $K \subset \mathbb{R}^+$ gilt

$$\lim_{x \rightarrow \infty} \sup_{\lambda \in K} \left| \frac{f(\lambda x)}{f(x)} - \lambda^\alpha \right| = 0.$$

4. Für beliebige $\delta > 0$ gilt

$$\lim_{x \rightarrow \infty} \frac{x^{\alpha-\delta}}{f(x)} = 0, \quad \lim_{x \rightarrow \infty} \frac{f(x)}{x^{\alpha+\delta}} = 0$$

und insbesondere letztendlich

$$x^{\alpha-\delta} < f(x) < x^{\alpha+\delta}.$$

5. Falls $\alpha > 0$, gilt für $r_n \rightarrow \infty$

$$\lim_n f(r_n) = \infty.$$

6. Falls $\alpha > 0$ und f letztendlich injektiv ist, existiert die auf dem Bildbereich definierte Inverse und ist regulär variierend zum Index $1/\alpha$.

7. Die Resultate lassen sich mittels Bemerkung 2.2.6 für die reguläre Variation in der Null umformulieren. So wird etwa die vierte Eigenschaft zu der Aussage:

Sei $f \in RV_0(\alpha)$, so gilt für alle $\delta > 0$

$$\lim_{t \downarrow 0} \frac{t^{\alpha+\delta}}{f(t)} = 0, \quad \lim_{t \downarrow 0} \frac{f(t)}{t^{\alpha-\delta}} = \infty$$

und insbesondere letztendlich

$$t^{\alpha+\delta} < f(t) < t^{\alpha-\delta}.$$

Beweis. Wir verweisen auf das Buch [Bingham et al.]. ◇

Bemerkung A.2.9 (Numerische Bestimmung von *exakten* Quantilen).

Es sei eine Zufallsvariable X simulierbar, deren Verteilungsfunktion nicht explizit oder nur kompliziert darstellbar ist. Zum Zwecke der nahezu exakten Bestimmung der Quantile dieser Zufallsvariable ist es möglich, sehr oft aus dieser Verteilung zu simulieren und anschließend einen Quantilschätzer, wie etwa mittels der *quantile* Funktion aus dem *base R*-Paket, darauf anzuwenden. Die theoretische Legitimation für diese Praxis ist der Satz von Glivenko-Cantelli, siehe zum Beispiel [Klenke, Satz 5.23]. Nach diesem gilt für obigen Fall

$$\sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - F(x) \right| \xrightarrow{f.s.} 0,$$

wobei \hat{F}_n die empirische und F die Verteilungsfunktion von X bezeichne.

A.3 Abbildungen

In Abbildung 17 ist der Algorithmenvergleich gemäß Bemerkung 4.1.3 inklusive des auf dem globalen Optimierer DIRECT-L basierenden Schätzers zu sehen. Offensicht-

lich ist die Leistung des globalen Optimierers deutlich schlechter als die der anderen.

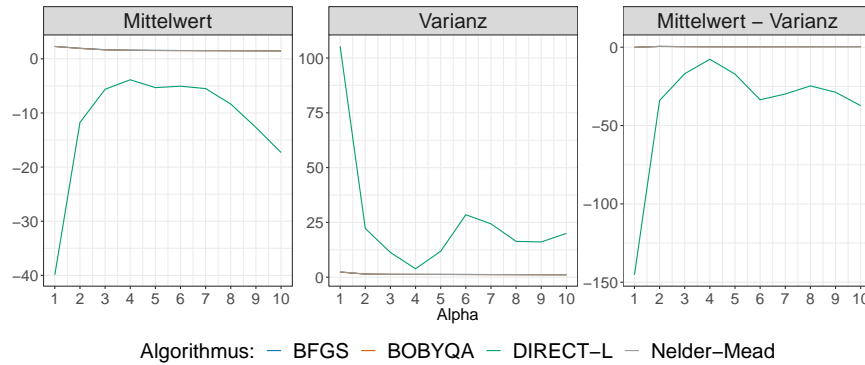


Abbildung 17: Kennzahlen für die Differenz $\ell(\hat{\theta}_S) - \ell(\theta_0)$ für die verschiedenen Algorithmen je α , wobei $\alpha \in \{1, \dots, 10\}$ und feste $(h, \mu, \sigma) = (1, 0, 1)$, vergleiche Bemerkung 4.1.3. Datengrundlage ist jene aus Bemerkung 4.1.2.

Wie in Bemerkung 4.1.5 angekündigt, vergleichen wir das Verhalten des Benchmark Stein Schätzers und der datenadaptiven Implementierung.

In Abbildung 18 sind die Kennzahlen der Schätzung des Formparameters α abgebildet.

Wir erkennen, dass der Benchmark Schätzer keine bessere Schätzung auf α bezogen

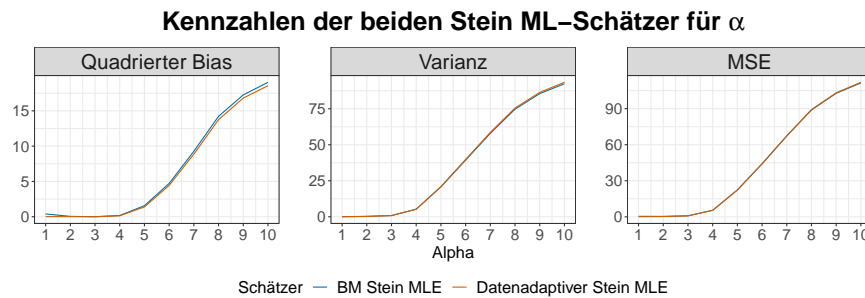


Abbildung 18: Kennzahlen des ML Schätzers $\hat{\alpha}_S$ für die datenadaptive und die Benchmark Implementierung je α . Datengrundlage ist jene aus Bemerkung 4.1.4.

generiert, sodass unter der Datengrundlage aus Bemerkung 4.1.4 der datenadaptive Schätzer gute Ergebnisse erzielt.

Zuletzt wollen wir ähnliche Kennzahlen für die Schätzung der Quantile vergleichen. In der Abbildung 19 sehen wir Kennzahlen der Quantilschätzer für die datenadaptive und Benchmark Implementierung.

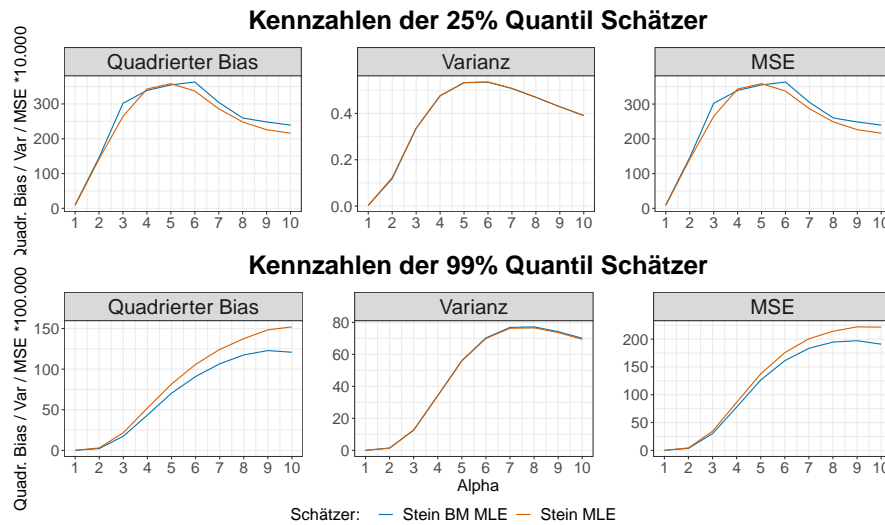


Abbildung 19: Kennzahlen des Stein 25- und 99%-Quantil Schätzers für die datenadaptive und die Benchmark Implementierung je α . Datengrundlage ist jene aus Bemerkung 4.1.4.

Auch hier fällt auf, dass keine signifikante bessere Schätzung durch den Benchmark Schätzer erreicht wird.

Literatur

- [Beirlant et al.] Jan Beirlant u. a. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.
- [Billingsley] Patrick Billingsley. *Probability and Measure*. Bd. Anniversary Edition. John Wiley & Sons, 2012.
- [Bingham et al.] Nicholas H Bingham, Charles M Goldie und Jef L Teugels. *Regular variation*. Bd. 1. Cambridge university press, 1989.
- [von Bortkiewicz(a)] Ladislaus von Bortkiewicz. *Die Variationsbreite beim Gausschen Fehlergesetz*. Nordisk Statistisk Tidsskrift, 1922.
- [von Bortkiewicz(a)] Ladislaus von Bortkiewicz. *Variationsbreite und mittlerer Fehler*. Berliner Mathematische Gesellschaft, 1921.
- [Bücher(a)] Axel Bücher. *Wahrscheinlichkeitstheorie. Skriptum zur Vorlesung an der Heinrich-Heine-Universität Düsseldorf*. 2019.
- [Bücher(b)] Axel Bücher. *Extremwerttheorie. Manuskriptum zur Vorlesung an der Heinrich-Heine-Universität Düsseldorf*. 2019.
- [Bücher(c)] Axel Bücher. *Mathematische Statistik I. Skriptum zur Vorlesung an der Heinrich-Heine-Universität Düsseldorf*. 2020.
- [Bücher und Segers(a)] Axel Bücher und Johan Segers. "On the maximum likelihood estimator for the generalized extreme-value distribution". In: *Extremes* 20.4 (2017), S. 839–872.
- [Bücher und Segers(b)] Axel Bücher und Johan Segers. "Inference for heavy tailed stationary time series based on sliding blocks". In: *Electronic Journal of Statistics* 12.1 (2018), S. 1098–1125.
- [Bücher und Zhou] Axel Bücher und Chen Zhou. "A horse racing between the block maxima method and the peak-over-threshold approach". In: *arXiv preprint arXiv:1807.00282* (2018).
- [Casella und Berger] George Casella und Roger L Berger. *Statistical inference*. Duxbury, 2021.
- [Cramér] Harald Cramér. *Mathematical Methods of Statistics*. Princeton university press, 1946.

- [David und Edwards] Herbert Aron David und Anthony William Fairbank Edwards. *Annotated readings in the history of statistics*. Springer, 2001.
- [Dombry] Clément Dombry. "Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework". In: *Bernoulli* 21.1 (2015), S. 420–436.
- [Dombry und Ferreira] Clément Dombry und Ana Ferreira. "Maximum likelihood estimators based on the block maxima method". In: *Bernoulli* 25.3 (2019), S. 1690–1723.
- [Engeland et al.] Kolbjørn Engeland, Hege Hisdal und Arnoldo Frigessi. "Practical extreme value modelling of hydrological floods and droughts: a case study". In: *Extremes* 7.1 (2004), S. 5–30.
- [Feller] William Feller. "An introduction to probability theory and its applications". In: *John Wiley & Sons Inc., New York* 2 (1971).
- [Fisher und Tippett] Ronald Aylmer Fisher und Leonard Henry Cabel Tippett. "Limiting forms of the frequency distribution of the largest or smallest member of a sample". In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Bd. 24. 2. Cambridge University Press. 1928, S. 180–190.
- [Gablonsky und Kelly] Joerg M Gablonsky und Carl T Kelley. "A locally biased form of the DIRECT algorithm". In: *Journal of Global Optimization* 21.1 (2001), S. 27–37.
- [Gilbert and Varadhan] Ravi Gilbert Paul und Varadhan. *numDeriv: Accurate Numerical Derivatives*. R package version 2016.8-1.1. 2019. URL: <https://CRAN.R-project.org/package=numDeriv>.
- [Gradshteyn und Ryzhik] Izrail Solomonovich Gradshteyn und Iosif Moiseevich Ryzhik. *Table of integrals, series, and products*. Academic press, 2014.
- [Gumbel] Emil Julius Gumbel. *Statistics of extremes*. Columbia university press, 1958.
- [Henningesen und Toomet] Arne Henningsen und Ott Toomet. "maxLik: A package for maximum likelihood estimation in R". In: *Computational Statistics* 26.3 (2011), S. 443–458. URL: <http://dx.doi.org/10.1007/s00180-010-0217-1>.

- [Hoskin et al.] Jonathan Richard Morley Hosking, James R Wallis und Eric F Wood. "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments". In: *Technometrics* 27.3 (1985), S. 251–261.
- [Johnson] Steven G. Johnson. "The NLOpt nonlinear-optimization package". In: (). URL: <https://CRAN.R-project.org/package=nloptr>.
- [Klenke] Achim Klenke. *Wahrscheinlichkeitstheorie*. Bd. 3. Springer, 2012.
- [Leadbetter, Lindgren und Rootzén] Malcolm R Leadbetter, Georg Lindgren und Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Science & Business Media, 2012.
- [von Mises] R von Mises. "Über die variationsbreite einer Beobachtungsreihe". In: *Sitzungsberichte der Berliner Mathematischen Gesellschaft* 22.3 (1923).
- [Nash und Varadhan] John C. Nash und Ravi Varadhan. "Unifying Optimization Algorithms to Aid Software System Users: optimx for R". In: *Journal of Statistical Software* 43.9 (2011), S. 1–14. URL: <https://www.jstatsoft.org/v43/i09/>.
- [Nelder und Mead] John A Nelder und Roger Mead. "A simplex method for function minimization". In: *The computer journal* 7.4 (1965), S. 308–313.
- [Nocedal] Jorge Nocedal. "Updating quasi-Newton matrices with limited storage". In: *Mathematics of computation* 35.151 (1980), S. 773–782.
- [Powell] Michael JD Powell. "The BOBYQA algorithm for bound constrained optimization without derivatives". In: *Cambridge NA Report NA2009/06, University of Cambridge, Cambridge* (2009), S. 26–46.
- [Resnick] Sidney I Resnick. *Extreme values, regular variation and point processes*. Springer, 2013.
- [Rüschendorf] Ludger Rüschendorf. *Mathematische statistik*. Bd. 62. Springer, 2014.
- [Singer und Nelder] Saša Singer und John Nelder. "Nelder-mead algorithm". In: *Scholarpedia* 4.7 (2009), S. 2928.

- [Stein] ML Stein. "Should annual maximum temperatures follow a generalized extreme value distribution?" In: *Biometrika* 104.1 (2017), S. 1–16.
- [Stephenson] A. G. Stephenson. "evd: Extreme Value Distributions". In: *R News* 2.2 (2002), S. 0. URL: <https://CRAN.R-project.org/package=evd>.
- [Van der Vaart] Aad W Van der Vaart. *Asymptotic statistics*. Bd. 3. Cambridge university press, 2000.
- [Yaqiong] Tian Yaqiong. "Extreme Risk Analysis of Personal Insurance Claim Based on Block Maxima Method". In: *International Journal of Economics, Finance and Management Sciences* 6.4 (2018), S. 192.

Erklärung

Hiermit versichere ich, dass ich die Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Düsseldorf, den 30. September 2021

(Torben Staud)