

# Extremwertstatistik

Arne Huckemann

September 2023

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>3</b>
1.1	Grundbegriffe . . . . .	3
<b>2</b>	<b>Univariate Extremwerttheorie</b>	<b>8</b>
2.1	Eindeutigkeit von $\alpha$ -Stabilen Verteilungen . . . . .	8
2.2	Die drei Extremwertverteilungen . . . . .	12
2.3	Darstellung der drei Extremwertverteilungen . . . . .	16
2.4	Max-Anziehungsbereiche . . . . .	22
2.5	Anziehungsbereiche . . . . .	24
2.6	Überschreitungen . . . . .	33
<b>3</b>	<b>Univariate Schätzverfahren</b>	<b>37</b>
3.0.1	Graphische Werkzeuge zur Modellverifikation: . . . . .	38
3.0.2	Block-Maxima - Anpassung einer GEV . . . . .	40
3.0.3	Peaks-Over-Threshold (POT) - Anpassung einer GPD . . . . .	45
3.1	Weitere Schätzverfahren für den Extremwertindex . . . . .	50
3.1.1	Pickands-Schätzer: $\gamma \in \mathbb{R}$ . . . . .	50
3.1.2	Hill-Schätzer: $\gamma > 0$ . . . . .	52
3.1.3	Dekkers-Einmahl-de Haan-Schätzer (Momenten-Schätzer): . . . . .	53
3.2	Extrapolation in Risikobereiche . . . . .	54
<b>4</b>	<b>Multivariate Extremwerttheorie (und -statistik)</b>	<b>57</b>
4.0.1	Max-Stabilität . . . . .	63
4.1	Poisson-Punktprozesse . . . . .	63
4.1.1	Hintergrund: Punktprozesse . . . . .	64

4.1.2	Poisson-Punktprozesse . . . . .	66
4.1.3	Eigenschaften eines Poisson-Punktprozesses . . . . .	66
4.2	Räumliche Statistik . . . . .	68
4.2.1	Brownsche Bewegung . . . . .	68
4.3	Simulation von PPP . . . . .	73
4.3.1	Methode 1: (Acceptance Rejection) . . . . .	73
4.3.2	Methode 2:() . . . . .	73
4.4	Spektraldarstellung... . . . .	74

# Kapitel 1

## Einführung

In 1.1 führen wir neue Begriffe ein, nämlich die ausgeartete Verteilungsfunktionen, langsam variierende Funktionen und  $\Sigma$ -Anziehungsbereiche.

In 2.2. Behandeln wir Aussagen zur Eindeutigkeit von  $\Sigma$ -Stabilen Verteilungen. Außerdem führen wir die schwache Konvergenz bzgl. Verteilungsfunktionen (Es konvergiert dann auch bzgl. er Pseudoinversen). Den Satz von Skorohod sagt dann aus, dass Aussagen bzgl. fast Sicherer Konvergenz äquivalent zu zeigen sind auf einem anderen W-Raum. Diese Aussage benötigen wir dann, wenn wir zeigen wollen, dass nicht ausgeartete Verteilungsfunktionen gegen Verteilungen konvergieren, die vom gleichen Typen sind.

In 2.3. werden wir die drei Extremwertverteilungen einführen und ein paar Eigenschaften von Ihnen diskutieren.

In 2.4. behandeln wir Max-Anziehungsbereiche, definieren Extremwertverteilungen. Außerdem definieren wir zusätzlich Zufallsvariablen, die vom gleichen Typ sind. Schließlich zeigen wir, dass die 3 Extremwertverteilungen gleich der sogenannten Jenkinson-von Mises-Darstellung sind.

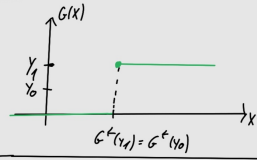
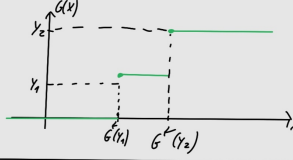
In 2.5

### 1.1 Grundbegriffe

**Definition 1.1.1.** Eine Verteilungsfunktion heißt ausgeartet bzw. degeneriert, falls diese nur Werte in  $\{0, 1\}$  annimmt.

Also  $\exists c$  mit  $\mathbb{P}[X = c] = 1$ , d.h.

$$G(t) = \begin{cases} 0, & t < c \\ 1, & t \geq c \end{cases}$$

Sei $G$ degeneriert:	Sei $G$ nicht degeneriert:
	
$\nexists y_1, y_0 : 0 < y_0 < y_1$	$\exists y_1, y_2 : 0 < y_1 < y_2$

**Definition 1.1.2.** ( $\alpha$ -stabile Verteilungen) Seien  $X_1, X_2, \dots$  u.i.v. so liegen die Verteilungen dieser Zufallsvariablen im  $\Sigma$ -Anziehungsbereich von einer sogenannten  $\alpha$ -stabilen Verteilung  $G$ , falls  $\exists a_n > 0$  und  $b_n \in \mathbb{R}$  mit:

$$\frac{1}{a_n} \left( \sum_{i=1}^n X_i - b_n \right) \rightarrow_d G$$

für eine nicht ausgeartete Verteilungsfunktion  $G$ , d.h.  $G$  nennen wir  $\alpha$ -stabil, falls diese eine charakteristische Funktion der Form

$$\Phi(x) := \int_{\mathbb{R}} e^{itx} G(dx) = \exp(i\gamma t - c|t|^\alpha (1 - \text{sign}(t)i\beta z(t, \alpha)))$$

besitzt mit

$$z(t, \alpha) = \begin{cases} \tan(\pi\alpha/2), & \alpha \neq 1 \\ -2/\pi \log(t), & \alpha = 1 \end{cases}$$

Hierbei ist  $\gamma \in \mathbb{R}, c > 0, \alpha \in (0, 2], \beta \in [-1, 1]$ .

- (Dies ist die Fourietransformierte, Irgendwas mit Sin und Cos und Hilbertraum...)
- Dabei beschreibt  $c$  die Varianz bei einer normalverteilung mit irgend einem Faktor
- $\alpha$  beschreibt on das  $\alpha$ -te Moment hat
- $\beta$  ist relevant bzgl. Symmetrie, d.h. ob es schief oder symmetrisch ist.
- Für  $\alpha = 2$  gilt:  $z(t, \alpha) = 0 \Rightarrow \Phi(t) = e^{i\gamma t - ct^2}$ . Das ist die charakteristische Funktion der Normalverteilung mit evtl:  $c = \frac{\sigma^2}{2}$ .
- $\alpha = 1$  dann ist  $\Phi(t)$  cauchy
- $\beta = 0$  dann ist  $\Phi(t)$  symmetrisch

**Beispiel 1.1.3.** Bei der Definition der  $\alpha$ -stabilen Verteilungen gelten die folgenden Fälle

- Normalverteilung ( $\alpha = 2$ ),
- Cauchyverteilung ( $\alpha = 1, \beta = 0, \gamma = 0, c = 1$ ).
- In dem Term von  $z(t, \alpha)$  ist ein Logarithmus auf, da irgendwo  $\int \frac{1}{x^2} dx$  auftaucht.

**Definition 1.1.4.** (Multivariate Normalverteilung)

Eine multivariate Verteilung  $G$  heißt, multivariate Normalverteilung, falls (d-varianz) gilt:

$$t \mapsto \int_{\mathbb{R}} e^{it^T x} G(dx) = e^{i\gamma^T t - t^T C t}$$

Dabei ist  $\gamma \in \mathbb{R}^d$  und  $C \in \mathbb{R}^{d \times d}$  und  $C$  positiv semi-definit. (Dabei ist  $C$  die Kovarianzmatrix)

**Bemerkung 1.1.5.** Wenn wir diese Definition zu der Definition von der multivariaten Normalverteilung in Stochastik 2 vergleichen, stellen wir fest, dass:

- Die Definition in Stochastik 2 war schlecht, da verschiedene  $A$  die gleiche Verteilung ergeben haben
- Der Vorteil dieser Definition ist, dass für semi-definite  $C$  es keine Dichten gibt, aber wir dieses  $C$  einfach stehen lassen können??? und wir müssen es nicht invertieren??

**Definition 1.1.6.** (Sehr wichtige Definition)

Eine messbare Funktion  $L : [x_0, \infty) \rightarrow \mathbb{R}$  heißt langsam variierend, falls  $L$  schließlich positiv ist (das heißt  $L$  ist positiv ab einem  $x_1 \geq x_0$ ) und für alle  $t > 0$ :

$$\lim_{x \rightarrow \infty} \frac{L(xt)}{L(x)} \rightarrow 1$$

( $x_0$  wird so sehr nach rechts verschoben, bis irgendwann  $L$  positiv ist.

**Beispiel 1.1.7.**

- $L(x) = c$  mit  $c > 0$
- $L(x) = c + o(1), c > 0, L$  messbar

$$\lim_{x \rightarrow \infty} \frac{L(xt)}{L(x)} = \frac{c + o(1)}{c + o(1)} = 1$$

- $L(x) = \log(x)$  oder  $L(x) = 1/\log(x)$

$$\lim_{x \rightarrow \infty} \frac{L(xt)}{L(x)} = \lim_{x \rightarrow \infty} \frac{\log(t \cdot x)}{\log(x)} = \lim_{x \rightarrow \infty} \frac{\log(t) + \log(x)}{\log(x)} = 1 + \lim_{x \rightarrow \infty} \frac{\log(t)}{\log(x)} = 1 \quad \forall t > 0$$

- $L(x) = \log(\log(x)), x > 1$ . Hier das gleiche Prinzip. Man kann sogar die Logs iterieren, das heißt abzählbar oft aufeinander anwenden und dann gilt die Eigenschaft immernoch.

**Bemerkung 1.1.8. Achtung!** Es gibt auch messbare Funktionen  $L$  die folgende beide Eigenschaften haben:

- $\limsup_{x \rightarrow \infty} L(x) = \infty$
- $\liminf_{x \rightarrow \infty} L(x) = 0$

Das heißt, Funktionen die solche Schweinereien erfüllen, sind auch zugelassen, d.h. die können auch die Eigenschaften aus der vorherigen Definition erfüllen.

Man kann sich das vorstellen als schwingende Funktion wo die periode kürzer wird aber die amplitude größer???

BILD

Falls  $L$  so eine schweinerei ist, dann können wir mittels einem Vorfaktor die Funktion monoton, machen, da diese viel schneller wächst/fällt als  $L$ . Die Monotonie ist für uns relevant, da Verteilungsfunktionen monoton sind.

$$X^{\pm\alpha}L(x), \text{ dabei ist } \alpha \text{ das Gleiche wie in der ersten Definition}$$

Frage: Wie sehen die  $\sum$ -Anziehungsbereiche aus?

**Satz 1.1.9.**

- i) Die Verteilungsfunktion  $F$  gehört zum  $\sum$ -Anziehungsbereich der Normalverteilung genau dann, wenn die Funktion

$$\xi : x \mapsto c \int_{|y| \leq x} y^2 dF(y)$$

langsam variiert.

- ii) Die Verteilungsfunktion  $F$  gehört zum  $\sum$ -Anziehungsbereich einer  $\alpha$ -stabilen Verteilung mit  $\alpha < 2$  genau dann, wenn

$$F(-x) = \frac{c_1 + o(1)}{x^\alpha} L(x) \quad \text{und} \quad 1 - F(x) = \frac{c_2 + o(1)}{x^\alpha} L(x)$$

für  $x \rightarrow \infty$ , wobei  $L$  langsam variierend ist,  $c_1, c_2 \geq 0$  mit  $c_1 + c_2 > 0$ . (Es gilt in diesem Fall  $\beta = (c_2 - c_1) / (c_1 + c_2)$ .) langsam variiert.

**Bemerkung 1.1.10.** • Zunächst betrachten wir uns den ersten Teil des vorherigen Satzes:

- Das  $\sum$  bezieht bzw. steht für die Summe von u.i.v. ZV, d.h.  $\sum_{i=1}^n X_i$
- Berry Essen besagt:.....???
- Wir wollen die Dichte konstruieren, die im  $\sum$ -Anziehungsbereich von der Normalverteilung liegt (OHNE VARIANZ??)

Dabei ist unser Ziel das bei  $\xi$  ein log rauskommt. D.h. sei  $f$  die Dichte von  $F$ , dann sieht sie folgendermaßen aus:

$$f(x) = \frac{const.}{1 + |y|^3}$$

”JEDE endliche Summe hat unendliche Varianz aber falls die Summe gegen unendlich geht, wird die Varianz gegen 0 gehen”??

*BILD*

Für  $\alpha \rightarrow 0$  existieren immer weniger Momente. (Das ist die Gleiche Ecke wie bei der Stoch. Konvergenz??) und wir finden dieses Verhalten böseartig.

Wenn  $\alpha = 1$  existiert der Erwartungswert nicht, da:

$$const. \int x^2 |x|^{-3} dx = \infty, \forall const.$$

- Als nächstes betrachten wir genauer die Aussage des zweiten Teil des Satzes:

Wir ”tun jetzt so, als könnte man die Verteilungsfunktion  $F$  differenzieren an den Grenzen”, dann gilt:

*BILD*

Dabei ist dann  $L \neq \tilde{L}$ , da man bei  $o(1)$  jede ”Schweinerei” reinmachen kann.



# Kapitel 2

## Univariate Extremwerttheorie

### 2.1 Eindeutigkeit von $\alpha$ -Stabilen Verteilungen

**Bemerkung 2.1.1.** (Erinnerung an Stochastik 1)

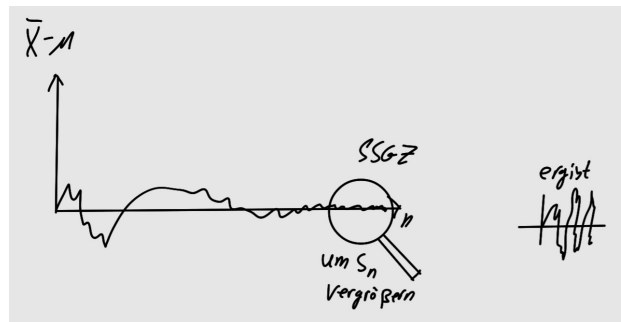
- SGGZ mit  $\mathbb{E}[X_i] = \mu < \infty$  und  $X_i$  u.i.v., dann gilt

$$\frac{1}{n} \sum_{i=1}^n X_i \longrightarrow \mu$$

- ZGWS: Dabei ist  $s_n$  egal wie es aussieht.

$$\frac{(\frac{1}{n} \sum_{i=1}^n X_i) - \mu}{s_n} \longrightarrow Y \sim \mathcal{N}(0, 1)$$

- Hier betrachten wir nochmal kurz den Einfluss von  $s_n$  auf die Konvergenz von  $\bar{X} - \mu$ : D.h. wenn wir



um " $s_n$ " vergrößern, dann konvergiert dieser Ausdruck nicht gegen Null. Das ist ein Analogon zu der Taylor entwicklung.

**Satz 2.1.2.** (Satz von Skorohod) Seien  $Y_n, n \geq 0$  reellwertige Zufallsvariablen, so dass  $Y_n \rightarrow^d Y_0$ . Dann gibt es Zufallsvariablen  $Y'_n, n \geq 0$  auf dem Wahrscheinlichkeitsraum  $([0, 1], \mathcal{B}([0, 1]), \lambda)$ , wobei  $\lambda$  das Lebesguemaß ist, so dass

i)  $\forall n \in \mathbb{N}_0 : Y'_n = {}^d Y_n$

ii)  $Y'_n \rightarrow Y'_0$  fast sicher.

**Definition 2.1.3.** • Eine Folge von monotonen Funktionen  $F_n$  auf einer offenen Teilmenge von  $\mathbb{R}^d$  (z.B. eine Folge von Verteilungsfunktionen) konvergiert schwach gegen  $F$ , falls  $F_n(x) \rightarrow F(x)$  für alle Stetigkeitspunkte  $x$  von  $F$ . Wir schreiben  $F_n \rightarrow^d F$ .

- Eine Folge von Zufallsvariablen  $X_n$  konvergiert schwach, d.h. konvergiert in Verteilung gegen eine Zufallsvariable  $X$ , falls die zu  $X_n$  gehörenden Verteilungsfunktionen schwach gegen die Verteilungsfunktion von  $X$  konvergiert. Wir schreiben  $X_n \rightarrow^d X$ .

**Definition 2.1.4.** Pseudoinverse

**Satz 2.1.5.** (Billingsley, 1995, Ferreira, de Haan 2006) Falls die Verteilungsfunktionen  $F_n$  schwach gegen  $F$  konvergieren, so konvergieren auch die Pseudoinversen  $F_n^{\leftarrow}$  schwach gegen  $F^{\leftarrow}$ .

**Bemerkung 2.1.6.** Wir werden folgende Extremwertverteilungen konstruieren: Die Gumbel-Verteilung  $\Lambda$ , die Fréchet-Verteilung  $\Phi_\alpha$  (wobei  $\alpha > 0$ ) und die Weibull-Verteilung  $\Psi_\alpha$  (wobei  $\alpha > 0$ ). Weitere Beispiele von Extremwertverteilungen können konstruiert werden, indem wir auf die oben genannten Verteilungen lineare Transformationen anwenden. Deshalb definieren wir Verteilungen vom selben Typ.

**Definition 2.1.7.** Zwei Zufallsvariablen  $Z_1$  und  $Z_2$  sind vom gleichen Typ, wenn es  $c > 0$  und  $d \in \mathbb{R}$  gibt mit

$$Z_1 \stackrel{d}{=} cZ_2 + d$$

Notation:  $Z_1 \bowtie Z_2$ .

Bezeichnen wir mit  $F_{Z_1}$  und  $F_{Z_2}$  die Verteilungsfunktionen von  $Z_1$  und  $Z_2$ , so kann man die obige Bedingung wie folgt formulieren:

$$F_{Z_1}(t) = \mathbb{P}[Z_1 \leq t] \stackrel{VSS}{=} \mathbb{P}[cZ_2 + d \leq t] = \mathbb{P}\left[Z_2 \leq \frac{t-d}{c}\right] = F_{Z_2}\left(\frac{t-d}{c}\right).$$

**Definition 2.1.8.** Zwei Verteilungsfunktionen  $F_1$  und  $F_2$  sind vom gleichen Typ, wenn es  $c > 0$  und  $d \in \mathbb{R}$  gibt, so dass für alle  $t \in \mathbb{R}$

$$F_1(t) = F_2\left(\frac{t-d}{c}\right).$$

Notation:  $F_1 \bowtie F_2$ .

**Beispiel 2.1.9.** Die Gleichverteilung auf dem Intervall  $[0, 1]$  ist vom gleichen Typ wie die Gleichverteilung auf einem beliebigen Intervall  $[a, b]$ . Die Normalverteilung mit beliebigen Parametern ist vom gleichen Typ wie die Standardnormalverteilung.

**Bemerkung 2.1.10.** Unsere Frage ist nun, ob der Grenzwerte der folgenden Darstellung eindeutig sind:

$$\frac{1}{a_n} \left( \sum_{i=1}^n X_i - b_n \right) \rightarrow_d G \quad \text{und} \quad \frac{1}{\tilde{a}_n} \left( \sum_{i=1}^n X_i - \tilde{b}_n \right) \rightarrow_d \tilde{G}$$

für Konstanten  $a_n, \tilde{a}_n > 0$  und  $b_n, \tilde{b}_n \in \mathbb{R}$  mit zwei verschiedenen nicht ausgearteten Grenzverteilungen  $G, \tilde{G}$ . Haben  $G$  und  $\tilde{G}$  dann etwas miteinander zu tun? Ja das kommt im nächsten Satz!

**Satz 2.1.11.** (Khinchine, Konvergenz gegen Typen) Seien  $Z_1, Z_2, \dots$  Zufallsvariablen mit Verteilungsfunktionen  $F_1, F_2, \dots$ . Seien  $G$  und  $\tilde{G}$  nicht ausgeartete Verteilungsfunktionen.

a) Seien  $c_n, \tilde{c}_n > 0, d_n, \tilde{d}_n \in \mathbb{R}$ , so dass

$$\frac{1}{c_n} (Z_n - d_n) \rightarrow_d G \quad \text{und} \quad \frac{1}{\tilde{c}_n} (Z_n - \tilde{d}_n) \rightarrow_d \tilde{G}, \quad (1)$$

dann folgt die Existenz der Limites

$$c = \lim_{n \rightarrow \infty} \frac{\tilde{c}_n}{c_n} > 0 \quad \text{und} \quad d = \lim_{n \rightarrow \infty} \frac{\tilde{d}_n - d_n}{c_n} \quad (2)$$

und es gilt  $\tilde{G}(t) = G(ct + d)$ . D.h.  $\tilde{G}$  und  $G$  sind vom selben Typ.

b) Seien  $c_n, \tilde{c}_n > 0, d_n, \tilde{d}_n \in \mathbb{R}$ , so dass die Limites in (2) existieren, dann folgt:

$$\frac{1}{c_n} (Z_n - d_n) \rightarrow_d G \quad \Longleftrightarrow \quad \frac{1}{\tilde{c}_n} (Z_n - \tilde{d}_n) \rightarrow_d \tilde{G}$$

mit  $\tilde{G} = G(ct + d)$ .

*Beweis.* b)

”  $\Rightarrow$  ”

Wir nehmen an, dass (2) gilt. Sei  $Y_n := (Z_n - d_n)/c_n$  und es gilt  $(Z_n - d_n)/c_n \rightarrow_d G$ . D.h. in schön geschrieben, dass  $Y_n \rightarrow_d W \sim G$ .

z.z.  $(Z_n - \tilde{d}_n)/\tilde{c}_n \rightarrow_d \tilde{G}$ , wobei  $\forall t \in \mathbb{R}: \tilde{G}(t) = G(ct + d)$ .

Denn dann gelten

- Sei  $\frac{W-d}{c} \sim \tilde{G}$ , dann gilt:  $\tilde{G}_{\frac{W-d}{c}}(t) = \mathbb{P}[\frac{W-d}{c} < t] = \mathbb{P}[W < ct + d] = G(ct + d)$
- Damit ist es äquivalent z.z., dass  $(Z_n - \tilde{d}_n) / \tilde{c}_n \rightarrow^d \frac{W-d}{c}$

Nach dem Satz von Skorohod existieren Zufallsvariable  $Y'_n$  und  $W'$  auf  $[0, 1]$  mit

- $Y'_n \stackrel{d}{=} Y_n$
- $Y'_n \xrightarrow{f.s.} W' \stackrel{d}{=} W$

Wir definieren nun  $Z'_n := c_n Y'_n + d_n$ . Dann gilt  $Z'_n \stackrel{d}{=} Z_n = Y_n c_n + d_n$ . Damit folgt

$$\begin{aligned} \frac{1}{\tilde{c}_n} (Z_n - \tilde{d}_n) &= \stackrel{d}{=} \frac{1}{\tilde{c}_n} (Z'_n - \tilde{d}_n) = \frac{1}{\tilde{c}_n} (c_n Y'_n + d_n - \tilde{d}_n) = \frac{c_n}{\tilde{c}_n} \left( Y'_n - \frac{\tilde{d}_n - d_n}{c_n} \right), \quad (2) \text{ anwenden und Skorohod} \\ &\xrightarrow{f.s.} \frac{1}{c} (W' - d) = \stackrel{d}{=} (W - d)/c \sim F_{\frac{W-d}{c}}(t) = F_W(ct + d) \stackrel{VSS}{=} G(ct + d) \stackrel{???}{=} \tilde{G}(t). \end{aligned}$$

Damit haben wir die erste Implikation gezeigt. Die Rückrichtung zeigen wir nicht.

a)

Gelte also (1), d.h.  $\frac{Z_n - d_n}{c_n} \rightarrow^d G$  und  $\frac{Z_n - \tilde{d}_n}{\tilde{c}_n} \rightarrow^d \tilde{G}$ , dass ist äquivalent zu (Erinnerung  $F_n$  war  $\forall n \in \mathbb{N}$  die Verteilungsfunktion von  $Z_n$ ):

$$F_n(c_n x + d_n) \rightarrow^d G(x) \quad \text{und} \quad F_n(\tilde{c}_n x + \tilde{d}_n) \rightarrow^d \tilde{G}(x).$$

Wir haben also GEZEIGT? das  $F_n$  schwach konvergiert gegen  $G$  bzw.  $\tilde{G}$ .??? Damit ist auch die Pseudoinverse schwach konvergent wegen Satz ...:

$$(F_n^{\leftarrow}(y) - d_n) / c_n \rightarrow^d G^{\leftarrow}(y) \quad \text{und} \quad (F_n^{\leftarrow}(y) - \tilde{d}_n) / \tilde{c}_n \rightarrow^d \tilde{G}^{\leftarrow}(y)$$

Wegen Definition 1.1.1 gilt: Da  $G$  und  $\tilde{G}$  nicht degeneriert sind, gibt es insbesondere zwei Stetigkeitsstellen  $y_1$  und  $y_2$  von  $G$  und  $\tilde{G}$  mit

$$G^{\leftarrow}(y_1) < G^{\leftarrow}(y_2) \quad \text{und} \quad \tilde{G}^{\leftarrow}(y_1) < \tilde{G}^{\leftarrow}(y_2),$$

Damit erhalten wir das folgende System von Konvergenzen:

$$(F_n^{\leftarrow}(y_1) - d_n) / c_n \rightarrow G^{\leftarrow}(y_1) \quad (2.1)$$

$$(F_n^{\leftarrow}(y_2) - d_n) / c_n \rightarrow G^{\leftarrow}(y_2) \quad (2.2)$$

$$\left(F_n^{\leftarrow}(y_1) - \tilde{d}_n\right) / \tilde{c}_n \rightarrow \tilde{G}^{\leftarrow}(y_1) \quad (2.3)$$

$$\left(F_n^{\leftarrow}(y_2) - \tilde{d}_n\right) / \tilde{c}_n \rightarrow \tilde{G}^{\leftarrow}(y_2) \quad (2.4)$$

Wenn wir (2.2)-(2.1) und (2.4)-(2.3) erhalten wir:

$$(F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)) / c_n \rightarrow G^{\leftarrow}(y_2) - G^{\leftarrow}(y_1) > 0, \quad (2.5)$$

$$(F_n^{\leftarrow}(y_2) - F_n^{\leftarrow}(y_1)) / \tilde{c}_n \rightarrow \tilde{G}^{\leftarrow}(y_2) - \tilde{G}^{\leftarrow}(y_1) > 0. \quad (2.6)$$

Wenn wir nun (2.5) durch 21.6) teilen ergibt sich:

$$\frac{\tilde{c}_n}{c_n} \xrightarrow{(2)} \frac{G^{\leftarrow}(y_2) - G^{\leftarrow}(y_1)}{\tilde{G}^{\leftarrow}(y_2) - \tilde{G}^{\leftarrow}(y_1)} =: c > 0.$$

(Wir durften durch (2.6) teilen, da  $\tilde{G}^{\leftarrow}(y_1) < \tilde{G}^{\leftarrow}(y_2)$ .)

Insbesondere folgt mit (2.3)

$$\frac{\left(F_n^{\leftarrow}(y_1) - \tilde{d}_n\right)}{c_n} = \frac{\left(F_n^{\leftarrow}(y_1) - \tilde{d}_n\right)}{\tilde{c}_n} \cdot \frac{\tilde{c}_n}{c_n} \rightarrow \tilde{G}^{\leftarrow}(y_1) c$$

und damit gilt schließlich:

$$\left(\tilde{d}_n - d_n\right) / c_n = \frac{\left(F_n^{\leftarrow}(y_1) - d_n\right)}{c_n} - \frac{\left(F_n^{\leftarrow}(y_1) - \tilde{d}_n\right)}{c_n} \rightarrow G^{\leftarrow}(y_1) - \tilde{G}^{\leftarrow}(y_1) c =: d.$$

Folglich existieren die Limites in (2).

Damit haben wir die Äquivalenz von a) und b) gezeigt. Da wir in b) gezeigt haben, dass  $\tilde{G}(t) = G(ct + d)$  gilt dies auch für die Richtung die wir in a) haben.  $\square$

## 2.2 Die drei Extremwertverteilungen

In diesem Abschnitt werden wir drei Extremwertverteilungen Definieren und charakterisieren. Bevor wir damit anfangen folgt eine kleine Motivation, warum wir es überhaupt machen und die Definition einer Ex-

tremwertverteilung Als nächstes betrachten wir das SGGZ(schwache) für Extremwerte

**Satz 2.2.1.** Seien  $X_1, X_2, \dots$  u.i.v. Zufallsvariablen mit Verteilungsfunktion  $F$ . Sei weiterhin

$$M_n := \max_{i=1}^n X_i$$

das Maximum. Dann ist die Verteilungsfunktion von  $M_n$  gegeben durch  $F^n$ , d.h.

$$\mathbb{P}[M_n \leq t] = F^n(t)$$

*Beweis.*

$$\begin{aligned}\mathbb{P}(M_n \leq t) &= \mathbb{P}(\max(X_1, \dots, X_n) \leq t) \\ &= \mathbb{P}(X_1 \leq t, \dots, X_n \leq t) \\ &= \mathbb{P}(X_1 \leq t) \dots \mathbb{P}(X_n \leq t) \\ &= F(t)^n = F^n(t).\end{aligned}$$

□

**Satz 2.2.2.** Es gilt  $M_n \rightarrow x^*$  fast sicher, wobei  $x^* := \sup\{x \in \mathbb{R} \mid F(x) < 1\}$  der obere Endpunkt von  $F$  ist.

*Beweis.* siehe Übung (Hinweis: Borell-Cantelli).

□

**Bemerkung 2.2.3.** Der obere Endpunkt  $x^*$  kann endlich oder  $\infty$  sein. Skizzieren!

Obiger Satz charakterisiert das Verhalten von  $M_n$  für  $n \rightarrow \infty$  nur sehr grob. Wir sind an Stabilitätsaussagen der Form

$$\frac{M_n - b_n}{a_n} \rightarrow^d G$$

interessiert (ähnlich wie beim ZGWS), wobei  $G$  eine nicht ausgeartete Verteilungsfunktion ist.

**Definition 2.2.4.** (Extremwertverteilung)

Falls Normierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$  existieren, so dass

$$\frac{M_n - b_n}{a_n} \rightarrow^d G$$

gilt, wobei  $G$  eine nicht ausgeartete Verteilungsfunktion ist, sagen wir, dass  $G$  eine Extremwertverteilung ist und  $X$  (bzw. die Verteilung  $F$  von  $X$ ) in ihrem Max-Anziehungsbereich liegt. D.h. es gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}\left[\frac{M_n - b_n}{a_n} \leq t\right] = G(t)$$

Frage: Gibt es überhaupt Extremwertverteilungen? Wir betrachten drei Beispiele:

**Beispiel 2.2.5.** Seien  $X_1, X_2, \dots$  exponentialverteilt mit Parameter 1, d.h.

$$F(t) = 1 - e^{-t}, \quad t \geq 0.$$

Dann gilt (für  $n$  groß genug, so dass  $t + \log n > 0$ )

$$\mathbb{P}(M_n - \log n \leq t) = F^n(t + \log n) = \left(1 - e^{-(t + \log n)}\right)^n = \left(1 - \frac{e^{-t}}{n}\right)^n \rightarrow e^{-e^{-t}}$$

für  $n \rightarrow \infty$ .

Interpretation: Für großes  $n$  nimmt das Maximum  $M_n$  Werte an, die sich von dem Wert  $\log n$  um eine approximativ Gumbel-verteilte "Fluktuation" unterscheiden

**Definition 2.2.6.** Eine Zufallsvariable hat die Gumbel-Verteilung, falls ihre Verteilungsfunktion die Gestalt  $\Lambda(t) := e^{-e^{-t}}$  hat.

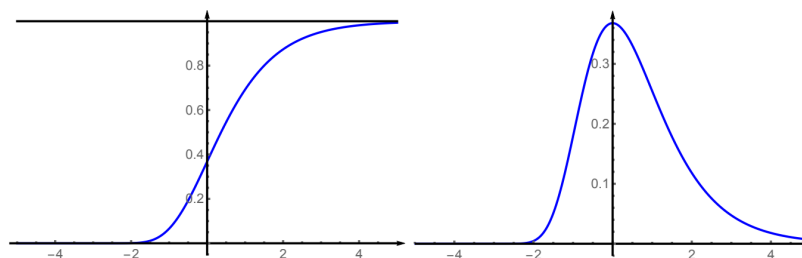


ABBILDUNG 3. Links: Verteilungsfunktion der Gumbel-Verteilung. Rechts: Dichte der Gumbel-Verteilung.

**Beispiel 2.2.7.** Seien  $X_1, X_2, \dots$  Pareto-verteilt mit Parameter  $\alpha > 0$ , d.h.

$$F(t) = \begin{cases} 1 - t^{-\alpha} & t \geq 1 \\ 0 & t \leq 1 \end{cases}$$

Dann gilt (für  $n$  groß genug:  $tn^{1/\alpha} > 1$ )

$$\mathbb{P}\left(\frac{M_n}{n^{1/\alpha}} \leq t\right) = F^n\left(tn^{1/\alpha}\right) = \left(1 - \frac{1}{t^\alpha n}\right)^n \rightarrow e^{-1/t^\alpha}$$

für  $n \rightarrow \infty$ .

Interpretation: Für großes  $n$  nimmt das Maximum  $M_n$  sehr große Werte auf der Skala  $n^{\frac{1}{\alpha}}$  an. Reskaliert man  $m_n$  mit dem Faktor  $n^{-\frac{1}{\alpha}}$ , so erhält man approximativ Fréchet-verteilte Werte

**Definition 2.2.8.** Eine Zufallsvariable hat die Fréchet-Verteilung, falls ihre Verteilungsfunktion die Gestalt

$$\Phi_\alpha(t) := \begin{cases} e^{-1/t^\alpha} & t > 0 \\ 0 & \leq 0 \end{cases}$$

, heißt Fréchet-Verteilung mit Parameter  $\alpha > 0$ .

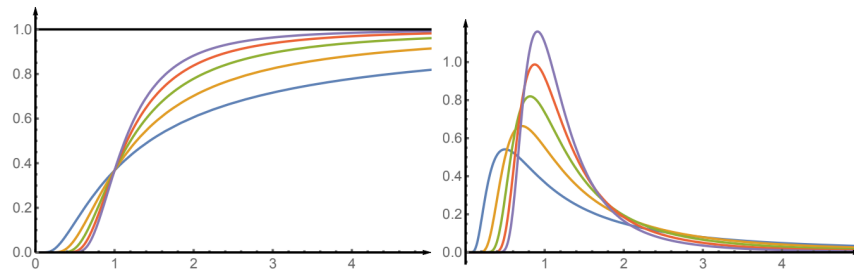


ABBILDUNG 4. Verteilungsfunktionen (links) und Dichten (rechts) der Fréchet-Verteilungen mit  $\alpha = 1.0, 1.5, 2.0, 2.5, 3.0$ .

**Beispiel 2.2.9.** (siehe Übungen) Sei  $\alpha > 0$  und

$$\Psi_\alpha(t) = \begin{cases} 0, & t \leq 0 \\ 1 - (1 - t)^\alpha, & t \in [0, 1] \\ 1, & t \geq 1 \end{cases}$$

Dann konvergieren die Zufallsvariablen  $n^{1/\alpha} M_n$  für  $n \rightarrow \infty$  in Verteilung gegen  $\Psi_\alpha$ , d.h. es gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}\left[n^{1/\alpha} M_n \leq t\right] = \begin{cases} e^{-(-t)^\alpha}, & t \leq 0, \\ 1, & t \geq 0. \end{cases}$$



Dies gilt, da einmal für  $n$  hinreichend groß ist  $tn^{-1/\alpha} \in [-1, 0]$  das folgende gilt:

$$\mathbb{P} \left[ n^{1/\alpha} M_n \leq t \right] = \mathbb{P} \left[ M_n \leq tn^{-1/\alpha} \right] = F^n \left( tn^{-1/\alpha} \right).$$

Aus der Formel für die Verteilungsfunktion  $F$  folgt, dass

$$\mathbb{P} \left[ n^{1/\alpha} M_n \leq t \right] = \left( 1 - \left( -tn^{-1/\alpha} \right)^\alpha \right)^n = \left( 1 - \frac{(-t)^\alpha}{n} \right)^n \xrightarrow{n \rightarrow \infty} e^{-(-t)^\alpha}.$$

Für  $t \geq 0$  gilt  $\mathbb{P} \left[ n^{1/\alpha} M_n \leq t \right] = 1$ , denn  $M_n \leq 0$  f.s. Daraus folgt die Behauptung.

Interpretation: Für großes  $n$  nähert sich das Maximum  $M_n$  dem Wert 0 von unten an. Dabei nimmt  $M_n$  sehr kleine negative Werte auf der Skala  $n^{-1/\alpha}$  an. Reskaliert man  $M_n$  mit dem Faktor  $n^{1/\alpha}$ , so erhält man approximativ Weibull-verteilte Fluktuationen.

**Definition 2.2.10.** Die Verteilungsfunktion  $\Psi_\alpha(t) = e^{-(-t)^\alpha}$ ,  $t \leq 0$  heißt Weibull-Verteilung mit Parameter  $\alpha > 0$ .

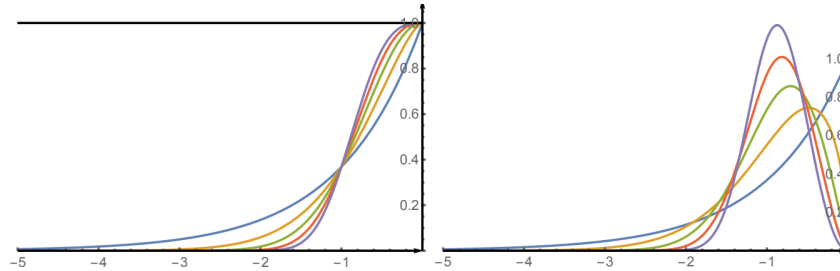


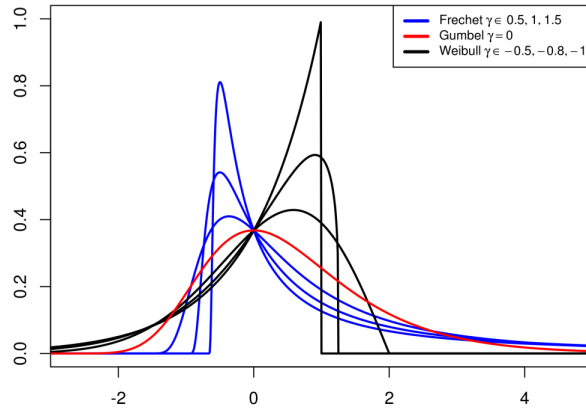
ABBILDUNG 5. Verteilungsfunktionen (links) und Dichten (rechts) der Weibull-Verteilungen mit  $\alpha = 1.0, 1.5, 2.0, 2.5, 3.0$ .

**Bemerkung 2.2.11.** Eine Fréchet-verteilte Zufallsvariable nimmt nur positive Werte an. Eine Weibull-verteilte Zufallsvariable nimmt nur negative Werte an. Eine Gumbel-verteilte Zufallsvariable kann sowohl positive als auch negative Werte annehmen.

## 2.3 Darstellung der drei Extremwertverteilungen

**Bemerkung 2.3.1.** Wir haben 3 Extremwertverteilungen konstruiert. Weitere Beispiele von Extremwertverteilungen können konstruiert werden, indem wir auf die oben genannten Verteilungen lineare Transformationen anwenden.

Es gibt nämlich eine Darstellung (Jenkinson-von Mises-Darstellung), die alle drei Familien (Gumbel, Fréchet



und Weibull) als Spezialfälle beinhaltet. Betrachte nämlich die folgende Familie von Verteilungsfunktionen (parametrisiert durch  $\gamma \in \mathbb{R}$ )

**Satz 2.3.2.** Hat eine Zufallsvariable  $Z$  (mit Verteilungsfunktion  $G(t)$ ) eine Extremwertverteilung, dann gilt

$\forall c > 0$  und  $d \in \mathbb{R} : cZ + d$  hat mit der Verteilungsfunktion  $G\left(\frac{t-d}{c}\right)$  eine Extremwertverteilung.

*Beweis.* Die Voraussetzung, dass die Zufallsvariable  $Z$  einer Extremwertverteilung gehorcht, bedeutet, dass es unabhängige identisch verteilte Zufallsvariablen  $X_1, X_2, \dots$  und Folgen  $a_n \in \mathbb{R}, b_n > 0$  gibt, so dass

$$\frac{\max\{X_1, \dots, X_n\} - a_n}{b_n} \xrightarrow[n \rightarrow \infty]{d} Z.$$

Daraus folgt, dass

$$\frac{\max\{cX_1, \dots, cX_n\} - (ca_n - db_n)}{b_n} = c \cdot \frac{\max\{X_1, \dots, X_n\} - a_n}{b_n} + \frac{db_n}{b_n} \xrightarrow[n \rightarrow \infty]{d} cZ + d.$$

Somit hat die Zufallsvariable  $cZ + d$  ebenfalls eine Extremwertverteilung.  $\square$

**Bemerkung 2.3.3.** Beim Satz von der Konvergenz gegen Typen haben wir Zufallsvariablen betrachtet, die in einer bestimmten Form gegen Verteilungen konvergieren, die bei anders gewählten Konstanten vom selben Typen sind. Diese Aussage gilt natürlich auch, falls wir das Maximum von Zufallsvariablen (was ja wieder eine Zufallsvariable ist) betrachten:

Falls also gilt

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G \quad \frac{M_n - \tilde{b}_n}{\tilde{a}_n} \xrightarrow{d} \tilde{G}$$

dann folgt, dass  $\tilde{G}$  und  $G$  vom selben Typ sind, genauer

$$\tilde{G}(t) = G(at + b)$$

mit  $a = \lim_{n \rightarrow \infty} \tilde{a}_n/a_n$  sowie  $b = \lim_{n \rightarrow \infty} (\tilde{b}_n - b_n)/a_n$ .

**Satz 2.3.4.** (Fisher-Tippett '29, Gnedenko '43, de Haan '70) Eine Verteilung  $G$  ist eine Extremwertverteilung  $\iff \exists a > 0, b \in \mathbb{R}$  und  $\gamma \in \mathbb{R}$  gibt mit

$$G(t) = G_\gamma(at + b)$$

wobei

$$G_\gamma(t) = \exp\left(-(1 + \gamma t)_+^{-1/\gamma}\right)$$

Hierbei bedeutet  $(1 + \gamma t)_+ = \max(1 + \gamma t, 0)$ .

**Bemerkung 2.3.5.**

- Bevor wir diesen Satz beweisen, werden wir im nächsten Lemma zeigen, dass der sogenannte Extremwertindex eindeutig ist, D.h. der Beweis folgt.
- Falls  $\gamma > 0$ , so ist  $G_\gamma$  vom gleichen Typ wie  $\Phi_\alpha$  mit  $\alpha = 1/\gamma$  (Fréchet-Verteilung).
- Falls  $\gamma = 0$ , interpretieren wir

$$(1 + \gamma t)^{-1/\gamma} \rightarrow e^{-t}, \quad \gamma \rightarrow 0,$$

so dass  $G_0(t) = \Lambda(t)$  (Gumbel-Verteilung).

- Falls  $\gamma < 0$ , so ist  $G_\gamma$  vom gleichen Typ wie  $\Psi_\alpha$  mit  $\alpha = -1/\gamma$  (Weibull-Verteilung). Es können also nur diese drei Typen von Verteilungen als Extremwertverteilungen auftreten. Deshalb heißt obiger Satz auch Drei-Typen-Satz.

**Definition 2.3.6.** Der Parameter  $\gamma$  in Satz 1.2.10 heißt Extremwertindex. Die Verteilungsfunktion

$$G_{\gamma;\mu,\sigma}(x) = G_\gamma((x - \mu)/\sigma) = \begin{cases} \exp\left(-(1 + \gamma(x - \mu)/\sigma)^{-1/\gamma}\right), & \gamma \neq 0 \\ \exp\left(-e^{-(x - \mu)/\sigma}\right), & \gamma = 0 \end{cases},$$

wobei  $1 + \gamma(x - \mu)/\sigma > 0$ , heißt verallgemeinerte Extremwertverteilung (Generalized Extreme Value distribution = *GEV*).

**Bemerkung 2.3.7.** Wir wollen zeigen, dass der Extremwertindex  $\gamma$  eindeutig ist. Dies folgt, dann aus dem folgenden Satz

Mit dem Satz der Konvergenz gegen Typen folgt: Der Extremwertindex  $\gamma$  in Satz von Fisher-Tippet ist eindeutig durch  $F$  festgelegt, die affine Transformation  $t \mapsto at + b$  nicht.

**Lemma 2.3.8.** Sei  $F$  eine nicht ausgeartete Verteilungsfunktion. Falls

$$F(ax + b) = F(cx + d)$$

für alle  $x \in \mathbb{R}$  und feste  $a, c > 0, b, d \in \mathbb{R}$

Dann gilt:  $a = c$  und  $b = d$ .

*Beweis.* (von Satz 1.2.10)

” $\Rightarrow$ ”

Nach Definition ist  $G$  ist eine Extremwertverteilung  $\iff$  Es gibt eine Verteilungsfunktion  $F$  und Normierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$  gibt mit  $F^n(a_n x + b_n) \rightarrow G(x)$  für alle Stetigkeitsstellen  $x$  von  $G$ , und wenn  $G$  nicht ausgeartet ist.

Sei  $t > 0$  beliebig aber fest (unser Ziel ist es eine Funktion in Abhängigkeit von  $t$  zu konstruieren, da  $\lceil nt \rceil \rightarrow \infty, n \rightarrow \infty$ ), dann gilt

$$F^{\lceil nt \rceil}(a_{\lceil nt \rceil} x + b_{\lceil nt \rceil}) \rightarrow G(x)$$

und es folgt

$$F^{\lceil nt \rceil}(a_n x + b_n) = (F^n(a_n x + b_n))^{\lceil nt \rceil/n} \rightarrow G^t(x).$$

Denn  $\forall \epsilon > 0$  :

$$\frac{\lceil nt \rceil}{n} = \frac{nt + \epsilon}{n} = \frac{nt}{n} + \frac{\epsilon}{n} \rightarrow t, n \rightarrow \infty$$

Der Satz der Konvergenz gegen Typen liefert:

$$\exists A(t), B(t) \in \mathbb{R} : \quad G^t(x) = G(A(t)x + B(t))$$

mit

$$A(t) = \lim_{n \rightarrow \infty} \frac{a_n}{a_{\lceil nt \rceil}} \quad \text{und} \quad B(t) = \lim_{n \rightarrow \infty} \frac{b_n - b_{\lceil nt \rceil}}{a_{\lceil nt \rceil}}$$

Insbesondere sind  $A(t)$  und  $B(t)$  messbar??.

Da  $G$  nicht ausgeartet ist, existiert ein  $x_0$  mit  $G(x_0) \in (0, 1)$ , d. h. wir können  $G(x_0) = e^{-q}$  schreiben mit

einer positiven Konstanten  $q$ , denn  $e^{-q} < 1$ . Dies führt zu

$$G(A(t)x_0 + B(t)) = G^t(x_0) = \exp(-qt). \quad (1)$$

Wäre die Abbildung  $t \mapsto A(t)x_0 + B(t)$  bekannt und invertierbar, wäre der Beweis beendet.

Für  $s, t > 0$  erhalten wir

$$\begin{aligned} G(A(st)x + B(st)) &= G^{st}(x) = (G^s(x))^t = G^t(A(s)x + B(s)) \\ &= G(A(t) \cdot (A(s)x + B(s)) + B(t)) \\ &= G(A(s)A(t)x + A(t)B(s) + B(t)). \end{aligned}$$

Lemma 2.3.8 liefert

$$A(st) = A(s)A(t) \quad (2)$$

$$B(st) = A(t)B(s) + B(t) \quad (3)$$

für  $s, t > 0$ .

Da  $A(t)$  messbar ist, folgt  $A(t) = t^{-\theta}$  für ein  $\theta \in \mathbb{R}$  (siehe Übungsaufgabe Blatt 2).

Wir unterscheiden drei Fälle.

Fall  $\theta = 0$ :

Damit ist  $A(t) \equiv 1$  und aus (3) erhalten wir:

$$B(st) = B(s) + B(t).$$

Wir wollen jetzt zeigen, dass  $B'(s' + t') = B'(s') + B'(t')$ .

Dazu definieren wir  $B' := B \circ \exp$  und  $s' := \log(s), t' := \log(t)$ , so ergibt sich

$$\begin{aligned} B'(s' + t') &= B'(\log(s) + \log(t)) \\ &= B'(\log(s \cdot t)) \\ &= (B' \circ \log)(s \cdot t) \\ &= B(s \cdot t) \\ &= B(s) + B(t) = (B \circ e \circ \log)(s) + (B \circ e \circ \log)(t) \\ &= (B \circ e)(s') + (B \circ e)(t') \\ &= B'(s') + B'(t') \end{aligned}$$

Damit folgt dann wieder mittels Blatt 2 Aufgabe 3, dass:

$$\exists c \in \mathbb{R} \forall s > 0: \quad B(s) = (B' \circ \log)(s) = B'(s') \stackrel{\text{Blatt 2}}{=} -c \cdot s' = -c \cdot \log(s)$$

Gleichung (1) liefert (d.h. wir setzen  $A(t) = 1$  wegen VSS und eben gezeigt, dass  $B(t) = -c \cdot \log(s)$ ):

$$G(x_0 - c \log t) = \exp(-qt),$$

Wir sehen, dass  $c \neq 0$  gelten muss, da sonst die rechte Seite von  $t$  abhängt, aber die Linke nicht.

Mit  $u := x_0 - c \log(t)$  erhalten wir

$$G(u) = \exp(-qt) = \exp\left(-qe^{-(u-x_0)/c}\right) = \exp\left(-e^{-(u-x_0)/c+\log(q)}\right) = \exp\left(-e^{-((u-x_0)/c-\log(q))}\right).$$

Dies ist eine Verteilungsfunktion  $\iff c > 0$ .

Damit vom Typ Gumbel.

Fall  $\theta > 0$ :

Aus (2) folgt (tausche Rollen von  $s$  und  $t$ ), dass

$$\begin{aligned} B(st) &= B(ts) \\ \iff \\ A(t)B(s) + B(t) &= A(s)B(t) + B(s) \end{aligned}$$

bzw.

$$\frac{B(s)}{1-A(s)} = \frac{B(t)}{1-A(t)}, \quad s, t \neq 1,$$

damit gilt

$$B(t) = (1-A(t)) \cdot \underbrace{\frac{B(s)}{1-A(s)}}_{=:d}$$

Dann ist der Term  $\frac{B(s)}{1-A(s)}$  eine Konstante bzgl. der Funktion  $B(t)$ , d. h.

$$B(t) = d(1-A(t)) = d(1-t^{-\theta})$$

und dabei ist nach dem oben gezeigten  $A(t) = t^{-\theta}$ .

Mittels (1) gilt dann wieder

$$G(t^{-\theta}x_0 + d(1 - t^{-\theta})) = G(t^{-\theta}(x_0 - d) + d) = \exp(-qt)$$

wieder muss gelten  $x_0 \neq d$ , da sonst eine Seite von  $t$  abhängt, aber die Andere nicht.

Mit  $\alpha = 1/\theta$  und  $u = t^{-\theta}(x_0 - d) + d$  haben wir  $t = [(u - d)/(x_0 - d)]^{-\alpha}$  und somit

$$\begin{aligned} G(u) &= \exp\left(-q \left[\frac{u-d}{x_0-d}\right]^{-\alpha}\right) = \exp\left(-\left[\frac{u-d}{q^{1/\alpha}(x_0-d)}\right]^{-\alpha}\right) \\ &= \Phi_\alpha\left(\frac{u-d}{q^{1/\alpha}(x_0-d)}\right), \quad u \geq d, \end{aligned}$$

und  $x_0 > d$ , damit dies eine Verteilungsfunktion ergibt. (Fréchet-Verteilung)

Fall  $\theta < 0$ : Ähnlich wie im vorherigen Fall, erhält man mit  $\alpha = -1/\theta$

$$G(u) = \exp\left(-\left[\frac{u-d}{q^{1/\alpha}(x_0-d)}\right]^\alpha\right).$$

Da  $G$  nichtfallend ist, ist  $x_0 - d$  negativ und  $u \leq d$ . (Weibull-Verteilung) □

## 2.4 Max-Anziehungsbereiche

Zur Erinnerung:  $X \sim F$  liegt im Max-Anziehungsbereich von  $G$ , falls für u.i.v. Kopien  $X_1, X_2, \dots$  und  $M_n := \max(X_1, \dots, X_n)$  mit geeigneten Normierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$  gilt

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G \text{ oder, äquivalent, } F^n(a_n x + b_n) \xrightarrow{d} G(x).$$

Bis auf eine monotone affine Transformation ist der Limes stets eindeutig (selber Typ). Alle möglichen nicht ausgearteten Funktionen  $G$ , die als Limes infrage kommen, haben wir bereits durch den Extremwertindex  $\gamma \in \mathbb{R}$  klassifiziert.

**Lemma 2.4.1.** (Poisson-Approximation) Für  $\tau \in [0, \infty]$  und eine Folge  $(u_n) \subset \mathbb{R}$  gilt

$$n(1 - F(u_n)) = n\mathbb{P}(X > u_n) \rightarrow \tau \quad (n \rightarrow \infty)$$

$\Longleftrightarrow$

$$F^n(u_n) = \mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau} \quad (n \rightarrow \infty)$$

*Beweis.*  $\tau < \infty$  :

Zunächst können wir in beiden Fällen schon spezifizieren, dass  $\tau \geq 0$  ist, denn

- $\mathbb{P}[M_n \leq u_n] \rightarrow e^{-\tau}$
- $n\mathbb{P}[X > u_n] \rightarrow \tau$

da Wahrscheinlichkeiten nicht negativ sind und nach oben durch 1 beschränkt sind.

Eine weitere Beobachtung ist, dass die Bedingung, dass

$$n(1 - F(u_n)) \rightarrow \tau, n \rightarrow \infty$$

impliziert, dass

$$(1 - F(u_n)) \rightarrow 0, n \rightarrow \infty$$

Nun zum restlichen Beweis:

” $\Rightarrow$ ”

Dann ist

$$F^n(u_n) = [1 - (1 - F(u_n))]^n = \left[1 - n \frac{1 - F(u_n)}{n}\right]^n \sim [1 - \tau/n + o(1/n)]^n \rightarrow e^{-\tau}$$

da  $n(1 - F(u_n)) \rightarrow \tau$  nach VSS

” $\Leftarrow$ ” : Es gilt für  $n \rightarrow \infty$ , dass

$$\begin{aligned} \log F^n(u_n) &= n \log F(u_n) = n \log(1 - (1 - F(u_n))) \\ &\sim -n((1 - F(u_n)) + (1 - F(u_n)) \cdot o(1 - F(u_n))), \quad \text{mit } o(1 - F(u_n)) \rightarrow 0 \\ &\sim -n(1 - F(u_n)) \rightarrow -\tau \end{aligned}$$

Damit gilt dann:

$$F^n(u_n) = e^{\log(F^n(u_n))} \sim e^{-n(1 - F(u_n))} \rightarrow e^{-\tau}$$

und die Taylorentwicklung der rechten Seite liefert die Behauptung.

$\tau = \infty$  : Übung. □

**Bemerkung 2.4.2.** Das Lemma von eben stellt eine Verbindung zwischen Überschreitungswahrscheinlichkeiten und der Verteilung der Maxima von u.i.v. Kopien her.



**Bemerkung 2.4.3.** (Warum "Poisson-Approximation"?) Setze für u.i.v.  $X_i \sim F$

$$B_n := \sum_{i=1}^n \mathbf{1}_{X_i > u_n} \sim \text{Bin}(n, 1 - F(u_n)).$$

Die Zufallsvariable  $B_n$  zählt die Anzahl der Punkte über der Schwelle  $u_n$  und ist offenbar binomialverteilt (vgl. Skizze).

BILD

Es ist bekannt, dass eine ZV  $X \sim \text{Bin}(n, p)$  gegen eine ZV  $Y \sim \text{Poi}(\lambda)$  konvergiert, falls  $np \rightarrow \lambda$

D.h. in unserem Fall gilt  $B_n \rightarrow \text{Poi}(\tau)$  genau dann, wenn  $\mathbb{E}(B_n) = n(1 - F(u_n)) \rightarrow \tau$ . In diesem Fall folgt auch  $\mathbb{P}(M_n \leq u_n) = \mathbb{P}(B_n = 0) \rightarrow e^{-\tau}$ . (Zur Erinnerung: Eine Poisson-verteilte Zufallsvariable  $Y \sim \text{Poi}(\lambda)$  ist definiert durch  $\mathbb{P}(Y = k) = e^{-\lambda} \lambda^k / k!, k = 0, 1, 2, \dots$ )

**Korollar 2.4.4.** Eine Verteilungsfunktion  $F$  gehört zum Max-Anziehungsbereich einer Extremwertverteilung  $G$  genau dann, wenn Konstanten  $a_n > 0, b_n \in \mathbb{R}$  existieren, so dass

$$n(1 - F(a_n x + b_n)) \rightarrow^d -\log G(x).$$

*Beweis.* Die Aussage folgt direkt aus Lemma 1.3.1. □

In diesem Korollar ist also unser  $u_n = a_n x + b_n$  eine reelle Folge. D.h. die Poisson approximation ist eine verallgemeinerung vom Max-Anziehungsbereich.

## 2.5 Anziehungsbereiche

### Charakterisierungen der einzelnen Anziehungsbereiche

In den Beispielen 1.2.4, 1.2.6 und 1.2.8 haben wir explizit Normierungskonstanten  $a_n$  und  $b_n$  gefunden, mit der die Konvergenz (9) gilt. Zum Beispiel haben wir gezeigt, dass die Exponentialverteilung im Gumbel-Max-Anziehungsbereich liegt und dass die Pareto-Verteilung im Fréchet-Max-Anziehungsbereich liegt.

Wir wollen im Folgenden Kriterien herleiten, die für eine gegebene Verteilungsfunktion  $F$  sicherstellen, dass sie im Max-Anziehungsbereich einer Extremwertverteilung  $G$  liegt, ohne dass wir explizit Normierungskonstanten bestimmen müssen. Mit dem bereits gezeigten, genügt es, die drei Typen Fréchet, Weibull und Gumbel oder die Jenkinson-Von-Mises-Familie  $G_\gamma$  genauer zu untersuchen. Sicherlich wird es bei solchen Kriterien nur auf das Tail-Verhalten der Verteilungen  $F$  ankommen.

Dazu definieren wir folgende Äquivalenzrelation:  $f(t) \sim g(t) : \iff \lim_{t \rightarrow \infty} f(t)/g(t) = 1$ .

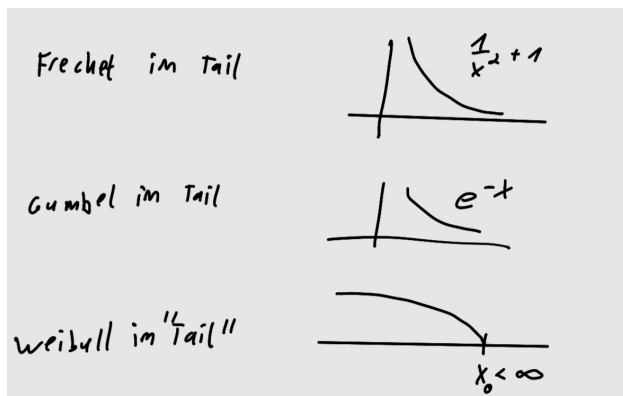
Wir beobachten, dass

$$1 - \Phi_\alpha(t) \sim t^{-\alpha}, \quad \text{für } t \uparrow \infty,$$

$$1 - \Lambda(t) \sim e^{-t}, \quad \text{für } t \uparrow \infty,$$

$$1 - \Psi_\alpha(t) \sim (-t)^\alpha, \quad \text{für } t \uparrow 0,$$

Wir wollen nun untersuchen, inwiefern Funktionen im Max-Anziehungsbereich von diesem Verhalten abweichen dürfen. Ein wichtiger Begriff hierfür ist die reguläre Variation. Wir benötigen ein paar Vorbereitungen.



**Definition 2.5.1.** (Regulär variierende Funktionen)

- (i) Eine messbare Funktion  $f : (x_0, \infty) \rightarrow \mathbb{R}$ , welche schließlich positiv ist, heißt regulär variierend (in  $\infty$ ) mit Index  $\alpha \in \mathbb{R}$ , falls

$$\lim_{\lambda \rightarrow \infty} \frac{f(\lambda t)}{f(\lambda)} = t^\alpha \quad \text{für alle } t > 0.$$

Wir schreiben  $f \in RV_\alpha$ . Falls  $\alpha = 0$ , so heißt  $f$  langsam variierend.

- (ii) Eine messbare Funktion  $f : (x_0, x_1) \rightarrow \mathbb{R}$ , welche in einer Umgebung von  $x_0$  (bzw. in einer Umgebung von  $x_1$ ) nur positive Werte annimmt, heißt regulär variierend in  $x_0$  (bzw. in  $x_1$ ) mit Index  $\alpha \in \mathbb{R}$ , falls

$$\begin{aligned} \lim_{\lambda \downarrow 0} \frac{f(x_0 + \lambda t)}{f(x_0 + \lambda)} &= t^\alpha \quad \text{für alle } t > 0 \\ \left( \text{bzw. } \lim_{\lambda \downarrow 0} \frac{f(x_1 - \lambda t)}{f(x_1 - \lambda)} &= t^\alpha \quad \text{für alle } t > 0 \right). \end{aligned}$$

Wir schreiben  $f \in RV_\alpha^{x_0}$  (bzw.  $f \in RV_\alpha^{x_1}$ ).

**Bemerkung 2.5.2.** Die zweite Definition von regulär variierend brauchen wir, um den Max-Anziehungsbereich von der Gumbel-Verteilung zu ermitteln.

**Anziehungsbereich der Fréchet-Verteilung  $\Phi_\alpha$**

**Satz 2.5.3.** Eine Verteilungsfunktion  $F$  mit oberem Endpunkt  $x^* = F^{\leftarrow}(1)$  liegt im Max-Anziehungsbereich der Fréchet-Verteilung  $\Phi_\alpha(t) = \exp(-t^{-\alpha})$  mit  $\alpha > 0$ ,  $\iff$

$$x^* = \infty \quad \text{und} \quad 1 - F \in RV_{-\alpha}.$$

Ist dies erfüllt, so können die Normierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$ , für die  $(M_n - b_n)/a_n \xrightarrow{d} \Phi_\alpha$  gilt, mit

$$a_n = F^{\leftarrow}\left(1 - \frac{1}{n}\right) \quad \text{und} \quad b_n = 0$$

gewählt werden.

*Beweis.* Übersprungen □

### Anziehungsbereich der Weibull-Verteilung $\Psi_\alpha$

**Satz 2.5.4.** Eine Verteilungsfunktion  $F$  mit oberem Endpunkt  $x^* = F^{\leftarrow}(1)$  liegt im Max-Anziehungsbereich der Weibull-Verteilung  $\Psi_\alpha(t) = \exp(-(-t)^\alpha)$  mit  $\alpha > 0$ ,  $\iff$

$$x^* < \infty \quad \text{und} \quad 1 - F \in RV_\alpha^{x^*}$$

Ist dies erfüllt, so können die Normierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$ , für die  $(M_n - b_n)/a_n \xrightarrow{d} \Psi_\alpha$  gilt, mit

$$a_n = x^* - F^{\leftarrow}\left(1 - \frac{1}{n}\right) \quad \text{und} \quad b_n = x^*$$

gewählt werden.

**Definition 2.5.5.** Sei  $F$  die Verteilungsfunktion einer Zufallsvariablen  $X$ . Dann schreiben wir im Folgenden

$$\bar{F}(t) := 1 - F(t) = \mathbb{P}(X > t)$$

für die Survivalfunktion (oder Tailfunktion).

**Beispiel 2.5.6.** (siehe Beispiel 1.2.8 KS)

Auf dem dritten Übungsblatt haben wir gezeigt, dass die folgende Funktion im Anziehungsbereich von der Weibull-Verteilung ist. Mit unserem neuen Kriterium können wir das jetzt auch wie folgt zeigen:

Sei  $\alpha > 0$  und

$$F(t) = \begin{cases} 0, & t \leq 0, \\ 1 - (1 - t)^\alpha, & t \in [0, 1] \\ 1, & t \geq 1, \end{cases}$$

Dann gilt  $x^* = 1 < \infty$  und sogar exakt

$$1 - F(x^* - t) = 1 - F(1 - t) = t^\alpha,$$

D.h.  $1 - F \in RV_\alpha^{x^*}$  und damit können wir die Normierungskonstanten wie folgt wählen:  $a_n = x^* - F^{\leftarrow}(1 - \frac{1}{n}) = 1 - (1/(1 - F))^{\leftarrow}(n) = n^{-1/\alpha}$  und  $b_n = 1$  geeignete Normierungskonstanten liefern.

BILD

### Anziehungsbereich der Gumbel-Verteilung $\Lambda$

Es sind nur recht komplizierte explizite hinreichende und notwendige Bedingungen bekannt, da die Survivalfunktion  $1 - F$  noch wesentlich stärker variieren kann als in den anderen beiden Fällen. Erste Charakterisierung:

**Satz 2.5.7.** Eine Verteilungsfunktion  $F$  liegt im Max-Anziehungsbereich der Gumbel-Verteilung  $\Lambda(t) = \exp(-e^{-t})$ ,  $\iff \exists$  positive und messbare Funktion  $a(t)$  mit

$$\lim_{t \rightarrow x^*} \frac{1 - F(t + xa(t))}{1 - F(t)} = e^{-x} \quad \text{für alle } x \in \mathbb{R}$$

Nun folgt eine hinreichende Bedingung, wo wir überprüfen können, ob eine Zufallsvariable im Anziehungsbereich von  $G_\gamma$ .

**Satz 2.5.8.** (Von-Mises-Bedingung)

Falls die Verteilungsfunktion  $F$  mit differenzierbarer Lebesgue Dichte eine der folgenden Äquivalenten Aussagen erfüllt

$$\lim_{x \rightarrow x_0} \left( \frac{1 - F}{f} \right)'(x) = \gamma \quad \text{oder äquivalent} \quad \lim_{x \rightarrow x_0} \frac{-f'(1 - F)}{f^2}(x) = 1 - \gamma$$

Dann liegt  $F$  im Max-Anziehungsbereich von  $G_\gamma$  und die Normierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$  die  $\frac{M_n - b_n}{a_n} \xrightarrow{d} G_\gamma$  erfüllen können wie folgt gewählt werden:

$$a_n := \frac{1}{n \cdot f(b_n)} \quad \text{und} \quad b_n := F^{\leftarrow}\left(1 - \frac{1}{n}\right)$$

**Bemerkung 2.5.9.** Bei der Normalverteilung ist  $\gamma = 0$ . Um dies zu zeigen brauchen wir erst einen Supersatz, der gleich folgt.

### Tail-Äquivalenz und Bestimmen von Normalisierungskonstanten

Das Bestimmen von Normalisierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$ , so dass  $(M_n - b_n)/a_n \rightarrow^d G$  für eine Extremwertverteilung  $G$  ist mitunter sehr schwierig. Wir werden im folgenden sehen, dass es dabei sehr nützlich sein kann, eine "Tail-äquivalente" Verteilung zu betrachten.

#### Definition 2.5.10. (Tail-Äquivalenz)

Zwei Verteilungsfunktionen  $F_1$  und  $F_2$  heißen Tail-äquivalent, falls sie den selben oberen Endpunkt  $x^*$  haben und der Limes

$$\lim_{x \uparrow x^*} \frac{1 - F_1(x)}{1 - F_2(x)} \in (0, \infty)$$

#### Beispiel 2.5.11. Die Verteilungsfunktionen

- $F(x) = 1 - x^{-\alpha} \cdot L(x)$
- $F_*(x) = 1 - x^{-\alpha^*} \cdot L_*(x)$

sind Tail-Äquivalent, (d.h.  $\frac{x^{-\alpha} L(x)}{x^{-\alpha^*} L_*(x)} \rightarrow const.$ )  $\iff \alpha = \alpha^*$  und  $\frac{L(x)}{L_*(x)} \rightarrow const.$

#### Satz 2.5.12. (Tail-Äquivalenz)

Sei die Verteilungsfunktion  $F$  im Max-Anziehungsbereich der Verteilungsfunktion  $G$ , das heißt es existieren Normalisierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$ , so dass

$$F^n(a_n x + b_n) \rightarrow^d G(x).$$

Dann gilt für eine Verteilungsfunktion  $F_*$  und eine nicht-ausgeartete Verteilung  $G_*$ , dass

$$F_*^n(a_n x + b_n) \rightarrow^d G_*(x)$$

genau dann, wenn die folgenden Bedingungen erfüllt sind:

- $G$  und  $G_*$  sind vom selben Typ mit  $G_*(t) = G(at + b)$  für  $t \in \mathbb{R}$  für ein  $a > 0$  und ein  $b \in \mathbb{R}$ .
- $F$  und  $F_*$  sind Tail-äquivalent (mit gemeinsamem Endpunkt  $x^*$ ).
- Falls

(i)  $G = \Phi_\alpha$  (Fréchet-  $\alpha$ ), dann  $b = 0$  und  $\lim_{t \uparrow \infty} (1 - F(t)) / (1 - F_*(t)) = a^\alpha$ ,

- (ii)  $G = \Psi_\alpha$  (Weibull-  $\alpha$ ), dann  $b = 0$  und  $\lim_{t \uparrow x^*} (1 - F(t)) / (1 - F_*(t)) = a^{-\alpha}$ ,
- (iii)  $G = \Lambda$  (Gumbel), dann  $a = 1$  und  $\lim_{t \uparrow x^*} (1 - F(t)) / (1 - F_*(t)) = e^b$ .

**Bemerkung 2.5.13.** Es ist wichtig zu merken, die Verbindung zum Satz von der Konvergenz gegen Typen.

- In diesem Satz konvergieren unterschiedliche Funktionen mit den gleichen Konstanten gegen Extremwertverteilungen, die vom gleichen Typ sind. Im anderen Satz konvergieren die gleichen Verteilungsfunktionen mit unterschiedlichen konstanten gegen Extremwertverteilungen die vom gleichen Typen sind.
- Der Unterschied ist, dass wir im Satz der Konvergenz gegen Typen der  $\Sigma$ -Anziehungsbereich betrachten und hier den Max-Anziehungsbereich. DAS STIMMT SO NICHT

**Beispiel 2.5.14.** Sei  $X \sim \mathcal{N}(0, 1)$  standardnormalverteilt, dann erfüllt die Dichte  $\varphi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ , dass  $\varphi'(x) = -x\varphi(x)$ . Somit folgt mit der Regel von L'Hospital für  $x \rightarrow \infty$ , dass

$$\frac{-\varphi'(x)(1 - \Phi(x))}{\varphi^2(x)} = \frac{1 - \Phi(x)}{x^{-1}\varphi(x)} \sim \frac{-\varphi(x)}{-x^{-2}\varphi(x) + x^{-1}\varphi'(x)} = \frac{1}{x^{-2} + 1} \rightarrow 1$$

Die Von-Mises-Bedingung (14) ist also für  $\gamma = 0$  erfüllt, denn

$$\lim_{x \rightarrow x_0} \frac{-f'(1 - F)}{f^2}(x) \stackrel{Def.}{=} \lim_{x \rightarrow x_0} \frac{-\varphi'(x)(1 - \Phi(x))}{\varphi^2(x)} = 1 \stackrel{!}{=} 1 - \gamma$$

Die Normalverteilung liegt im Gumbel-Anziehungsbereich.

**Beispiel 2.5.15.** (Normalisierungskonstanten für die Normalverteilung)

Sei  $X \sim \mathcal{N}(0, 1)$  standardnormalverteilt mit Dichte  $\varphi$  und Verteilungsfunktion  $\Phi$ . Aus Beispiel 1.3.40 KS wissen wir bereits, dass  $X$  im Gumbel-Anziehungsbereich liegt. Wir wollen nun Normalisierungskonstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$  für die Konvergenz

$$\Phi^n(a_n x + b_n) \rightarrow^d \Lambda(x)$$

bestimmen. Da  $\varphi$  differenzierbar ist und die Von-Mises-Bedingung (14) gilt (siehe Bsp. 1.3.40), ist  $\Phi$  nach Lemma 1.3.39 eine Von-Mises-Funktion und die Hilfsfunktion  $a$  in den Sätzen 1.3 .30 und 1.3 .32 ist mit  $a = (1 - \Phi)/\varphi$  wählbar. Geeignete Normalisierungskonstanten sind also

$$a_n = \frac{1}{n\varphi(b_n)} \text{ und } b_n = \Phi^{\leftarrow}\left(1 - \frac{1}{n}\right),$$

und da  $b_n = \Phi^{\leftarrow} \left(1 - \frac{1}{n}\right) \iff \Phi(b_n) = 1 - \frac{1}{n}$  gilt

$$a_n = \frac{1 - \Phi(b_n)}{\varphi(b_n)}$$

welche jedoch die Berechnung der Quantilsfunktion der Normalverteilung erfordern. Wir wollen noch einen weiteren Ansatz betrachten, der zu expliziten Konstanten führt. Da die Normalverteilung bereits im Anziehungsbereich der Gumbel Verteilung gilt, ist aus Bsp. 1.3.40 bereits bekannt, dass für  $x \rightarrow \infty$  wegen

$$\frac{\varphi'(x) \cdot (1 - \Phi(x))}{\varphi^2(x)} = \frac{x \cdot \varphi(x) \cdot (1 - \Phi(x))}{\varphi^2(x)} = \frac{1 - \Phi(x)}{x^{-1}\varphi(x)} \rightarrow 1 \quad (1)$$

Damit gilt dann

$$1 - \Phi(x) \sim x^{-1}\varphi(x) = (\sqrt{2\pi}x)^{-1} \exp(-x^2/2).$$

Da dieser Ausdruck in einer Umgebung von  $\infty$  (zum Beispiel  $[1, \infty)$ ) monoton fällt und gegen 0 konvergiert, dürfen wir ihn dort als Survivalfunktion  $1 - F$  für eine Verteilungsfunktion  $F$  auffassen. Per Definition sind  $\Phi$  und  $F$  dann Tail-äquivalent mit

$$\lim_{t \rightarrow \infty} (1 - F(t))/(1 - \Phi(t)) = 1 = e^0.$$

Nach Satz 1.3.43 gilt somit, dass  $\Phi^n(a_n x + b_n) \rightarrow^d \Lambda(x + 0) = \Lambda(x)$  genau dann, wenn  $F^n(a_n x + b_n) \rightarrow^d \Lambda(x)$  mit denselben (!) Konstanten  $a_n > 0$  und  $b_n \in \mathbb{R}$ . Weil Die Normalverteilung im Anziehungsbereich liegt, gilt also

$$a_n = \frac{1 - F(b_n)}{f(b_n)} \text{ und } b_n = F^{\leftarrow} \left(1 - \frac{1}{n}\right)$$

eine weitere zulässige Wahl (da  $F$  auch als Von-Mises-Funktion gewählt werden kann). Nach dem Satz von der "Konvergenz gegen Typen" ist  $a_n$  jedoch nur bis auf asymptotische Äquivalenz festgelegt, so dass mit  $\varphi(x) \sim f(x)$  (nachrechnen!) und

$$\frac{1 - F(x)}{f(x)} \sim \frac{1 - \Phi(x)}{\varphi(x)} \stackrel{(1)}{\sim} \frac{1}{x}$$

Damit

$$a_n = \frac{1}{b_n} \quad \text{und} \quad b_n = F^{\leftarrow} \left(1 - \frac{1}{n}\right).$$

eine zulässige Wahl für die Normierungskonstanten darstellt. Wir haben also eine berechnbare Darstellung

für  $a_n$  gefunden. Nun kann man sich  $b_n$  im Detail anschauen. Hier die Skizze: Schließlich ist

$$\begin{aligned} b_n = F^{\leftarrow}(1 - 1/n) &\Leftrightarrow -\log(1 - F(b_n)) = \log n \\ &\Leftrightarrow \frac{1}{2}b_n^2 + \log b_n + \frac{1}{2}\log(2\pi) = \log n. \end{aligned}$$

Dies führt auf (Details, siehe Resnick S. 72)

$$b_n = \sqrt{2 \log n} - \frac{\log \log n + \log(4\pi)}{2\sqrt{2 \log n}} + o\left(\frac{1}{\sqrt{\log n}}\right).$$

Da  $a_n$  nur bis auf asymptotische Äquivalenz festgelegt ist und  $b_n$  nur bis auf ein additives Element aus  $o(a_n)$ , folgt schließlich, dass

$$a_n = \frac{1}{\sqrt{2 \log n}} \quad \text{und} \quad b_n = \sqrt{2 \log n} - \frac{\log \log n + \log(4\pi)}{2\sqrt{2 \log n}}.$$

eine zulässige Wahl darstellt.

**Lemma 2.5.16.** Seien  $X > 0$  und  $Y > 0$  zwei unabhängige Zufallsvariablen im  $\Phi_\alpha$ -Anziehungsbereich,  $\alpha > 0$ . Dann liegt auch  $X + Y$  im  $\Phi_\alpha$ -Anziehungsbereich und es gilt

$$\frac{\mathbb{P}(X + Y > u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} \rightarrow 1 \quad \text{für } u \rightarrow \infty.$$

*Beweis.* Nach Voraussetzung gibt es zwei langsam variierende Funktionen  $L_1$  und  $L_2$ , so dass  $\mathbb{P}(X > u) = u^{-\alpha}L_1(u)$  und  $\mathbb{P}(Y > u) = u^{-\alpha}L_2(u)$ , da  $X$  bzw.  $Y$  im  $\Phi_\alpha$ -Anziehungsbereich sind, gilt

$$1 - \Phi_\alpha(u) \sim u^{-\alpha} \text{ d.h. } \lim_{u \rightarrow \infty} \frac{1 - \Phi_\alpha(u)}{u^{-\alpha}} = 1$$

Weil  $L_1$  bzw.  $L_2$  langsam variierend sind gilt:

$$\lim_{x \rightarrow \infty} \frac{L_1(\lambda x)}{L_1(x)} = 1 \text{ d.h. } L_1(\lambda x) \sim L_1(x) \quad \text{also} \quad \lim_{x \rightarrow \infty} \frac{L_1(\lambda x)}{L_1(x)} = 1$$

Dies können wir dann oben einsetzen:

$$\lim_{u \rightarrow \infty} \frac{1 - \Phi_\alpha(u)}{u^{-\alpha}} \cdot \frac{L_1(\lambda u)}{L_1(u)} = 1 \quad \text{d.h. } 1 - \Phi_\alpha(u) \sim u^{-\alpha} \cdot \frac{L_1(u)}{L_1(\lambda u)}$$

.....



Wir können das Ereignis  $\{X + Y > u\}$  nun wie folgt einschachteln: Sei dazu  $\delta \in (0, 1/2)$ . Dann ist

$$\begin{aligned} \{X > u\} \cup \{Y > u\} &\subset \{X + Y > u\} \\ &\subset \{X > (1 - \delta)u\} \cup \{Y > (1 - \delta)u\} \cup \{X > \delta u, Y > \delta u\}. \end{aligned}$$

Dies gilt, da:

- Wenn  $X + Y > u \Rightarrow \forall \delta > 0 \ X > (1 - \delta)u$
- Wenn  $X + Y > u \Rightarrow \forall \delta > 0 \ Y > (1 - \delta)u$
- Wenn  $X + Y > u \Rightarrow \forall \delta > 0 \ X > \delta u$  und  $Y > \delta u$

Wir erhalten dann also für  $u \rightarrow \infty$

$$\begin{aligned} \frac{\mathbb{P}(X + Y > u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} &> \frac{\mathbb{P}(X > u) + \mathbb{P}(Y > u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} \\ &\geq \frac{\mathbb{P}(X > u) + \mathbb{P}(Y > u) - \mathbb{P}(X > u)\mathbb{P}(Y > u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} = 1 - o(1) \end{aligned}$$

aus der ersten Inklusion sowie

$$\begin{aligned} &\frac{\mathbb{P}(X + Y > u)}{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)} \\ &\leq \frac{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u) + \mathbb{P}(X > \delta u)\mathbb{P}(Y > \delta u)}{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)} \\ &= \frac{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)}{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)} + \frac{\mathbb{P}(X > \delta u)\mathbb{P}(Y > \delta u)}{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)} \\ &= 1 + o(1) \end{aligned}$$

aus der zweiten Inklusion.

Somit folgt

$$\begin{aligned}
1 - o(1) &\leq \frac{\mathbb{P}(X + Y > u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} \\
&= \frac{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} \frac{\mathbb{P}(X + Y > u)}{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)} \\
&\leq \frac{\mathbb{P}(X > (1 - \delta)u) + \mathbb{P}(Y > (1 - \delta)u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} (1 + o(1)) \\
&\leq \frac{((1 - \delta)u)^{-\alpha} \cdot L_1((1 - \delta)u) + ((1 - \delta)u)^{-\alpha} \cdot L_2((1 - \delta)u)}{u^{-\alpha} \cdot L_1(u) + u^{-\alpha} \cdot L_1(u)} (1 + o(1)) \\
&= \frac{((1 - \delta)u)^{-\alpha}}{u^{-\alpha}} \frac{L_1((1 - \delta)u) + L_2((1 - \delta)u)}{L_1(u) + L_1(u)} (1 + o(1)) \\
&= (1 - \delta)^{-\alpha} \cdot \frac{L_1((1 - \delta)u) + L_2((1 - \delta)u)}{L_1(u) + L_1(u)} (1 + o(1)) \\
&\leq (1 - \delta)^{-\alpha} (1 + o(1)) (1 + o(1))
\end{aligned}$$

Die letzte Gleichheit gilt, da die punktweise Summe langsam variierender Funktionen wieder langsam variierend ist (Blatt 2).

Schließlich gilt für jedes  $\delta \in (0, 1/2)$ .....

$$1 = \liminf_{u \rightarrow \infty} \frac{\mathbb{P}(X + Y > u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} \leq \limsup_{u \rightarrow \infty} \frac{\mathbb{P}(X + Y > u)}{\mathbb{P}(X > u) + \mathbb{P}(Y > u)} = (1 - \delta)^{-\alpha}$$

Mit  $\delta \rightarrow 0$  folgt das Resultat (Warum? Kleine Übung!). □

## 2.6 Überschreitungen

BILD

Frage: Welcher Verteilung folgen die Punkte über dem Threshold?

In der Theorie verhält sich  $u_n$  ungefähr wie  $c \cdot \log(u)$  oder  $c \cdot n^\alpha$ . In der Praxis bringt die Asymptotik nichts, da nicht genügend Daten vorliegen.

Eine erste Verbindung zwischen Überschreitungswahrscheinlichkeiten und der Verteilung der Maxima haben wir bereits in Lemma der Poisson-Approximation gesehen. Zur Erinnerung (in unserer Standardnotation  $M_n := \max(X_1, \dots, X_n)$  wobei  $X_1, \dots, X_n \sim X \sim F$ ) :

Für  $\tau \in [0, \infty]$  gilt

$$n\bar{F}(u_n) = n\mathbb{P}(X > u_n) \rightarrow \tau \quad \Leftrightarrow \quad F^n(u_n) = \mathbb{P}(M_n \leq u_n) \rightarrow e^{-\tau}.$$

Desweiteren haben wir in Satz KS1.3.30 und Übung KS1.3.31 gesehen, dass eine Verteilungsfunktion  $F$  genau dann zum Anziehungsbereich der Extremwertverteilung  $G_\gamma$  gehört, wenn es eine positive und messbare Funktion  $a(u)$  gibt mit

$$\lim_{u \rightarrow x^*} \frac{1 - F(u + xa(u))}{1 - F(u)} = -\log G_\gamma(x) = \begin{cases} (1 + \gamma x)_+^{-1/\gamma} & \gamma \neq 0 \\ e^{-x} & \gamma = 0 \end{cases}$$

für alle  $x \in \mathbb{R}$  mit  $1 + \gamma x > 0$ . (Der erste Fall ist aus dem Beispiel und der zweite vom Satz zum Anziehungsbereich von der Gumbel Verteilung).

Falls zusätzlich  $x > 0$  ist, folgt  $u + xa(u) > u$ , d.h.  $a(u) > 0$  und wir können die linke Seite mit einer Zufallsvariablen  $X \sim F$  wie folgt als bedingte Überschreitungswahrscheinlichkeit interpretieren:

(Vorher ist wichtig zu erwähnen, dass  $X > u + xa(u) \Rightarrow X > u$ , damit hat dieses Ereigniss Masse 1, falls das vorherige gilt)

$$\begin{aligned} \frac{1 - F(u + xa(u))}{1 - F(u)} &= \frac{\mathbb{P}(X > u + xa(u))}{\mathbb{P}(X > u)} = \frac{\mathbb{P}(X > u + xa(u), X > u)}{\mathbb{P}(X > u)} \\ &= \mathbb{P}(X > u + xa(u) \mid X > u) = \mathbb{P}\left(\frac{X - u}{a(u)} > x \mid X > u\right) \end{aligned}$$

Nun können wir die vorherige Aussage umschreiben zu mittels unserer neuen Schreibweise, d.h.:

Wenn  $X$  mit oberem Endpunkt  $x^*$  im Max-Anziehungsbereich von  $G_\gamma$  liegt, impliziert dies also für  $x > 0$  und  $1 + \gamma x > 0$ , dass

$$\lim_{u \rightarrow x^*} \mathbb{P}\left(\frac{X - u}{a(u)} > x \mid X > u\right) = -\log G_\gamma(x) = \begin{cases} (1 + \gamma x)^{-1/\gamma} & \gamma \neq 0 \\ e^{-x} & \gamma = 0 \end{cases}$$

Nun betrachten wir noch weitere Fälle:

Im Fall  $x < 0$  ist der Grenzwert offenbar 1 (denn wir bedingen darauf, dass  $X > u$  und da  $a$  eine positive Funktion ist, ist der Quotient immer  $> 0$ ) und im Fall  $1 + \gamma x < 0$  (der nur für  $\gamma < 0$  eintreten kann), ist der Grenzwert auf der rechten Seite 0. Somit gilt die folgende Konvergenz von bedingten Verteilungen:

$$\lim_{u \rightarrow x^*} \mathbb{P}\left(\frac{X - u}{a(u)} \leq x \mid X > u\right) = (1 + \log G_\gamma(x))_+ = \begin{cases} \left(1 - (1 + \gamma x)_+^{-1/\gamma}\right)_+ & \gamma \neq 0 \\ (1 - e^{-x})_+ & \gamma = 0 \end{cases}$$

**Definition 2.6.1.** Die Verteilungsfunktion

$$H_{\gamma;\mu,\sigma}(x) = H_{\gamma}\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R},$$

wobei  $\mu \in \mathbb{R}, \sigma > 0$  und  $\gamma \in \mathbb{R}$  mit

$$H_{\gamma}(x) = (1 + \log G_{\gamma}(x))_+ = \begin{cases} \left(1 - (1 + \gamma x)_+^{-1/\gamma}\right)_+ & \gamma \neq 0 \\ (1 - e^{-x})_+ & \gamma = 0 \end{cases}$$

heißt verallgemeinerte Pareto-Verteilung (= Generalized Pareto Distribution = GPD). Falls  $\mu = 0$ , schreiben wir auch  $H_{\gamma;\sigma} := H_{\gamma,0,\sigma}$ .

EVTL VORHERIGE SEITE EINBEZIEHEN AUS KS...

Die Aussage des Satzes 1.4.3 kann sogar noch verschärft werden. Die Konvergenz in (ii) ist in einer Umgebung des oberen Endpunktes  $x^*$  von  $X$  in folgendem Sinne gleichmäßig.

**Satz 2.6.2.** (Pickands, Balkema, de Haan). Eine Zufallsvariable  $X$ , die stetig an ihrem oberen Endpunkt  $x^*$  ist, liegt genau dann im Max-Anziehungsbereich der Extremwertverteilung  $G_{\gamma}$ , wenn es eine positive messbare Funktion  $\sigma(u)$  gibt mit

$$\lim_{u \uparrow x^*} \sup_{t \in [0, x^* - u]} |\mathbb{P}(X - u \leq t \mid X > u) - H_{\gamma;\sigma(u)}(t)| = 0.$$

**Bemerkung 2.6.3.** Ist  $X$  im Max-Anziehungsbereich einer Extremwertverteilung, so hat  $X$  notwendigerweise keinen Sprung am oberen Endpunkt, vgl. Korollar 1.3.4.

**Definition 2.6.4.** Sei  $X$  eine Zufallsvariable mit Verteilungsfunktion  $F$  und oberem Endpunkt  $x^*$  sowie  $u < x^*$ . Dann heißt

$$F_u(x) := \mathbb{P}(X - u \leq x \mid X > u), \quad x \geq 0$$

Exzess-(Verteilungs-)Funktion. Falls außerdem  $\mathbb{E}|X| < \infty$  (bzw.  $\mathbb{E}X_+ < \infty$ ), so heißt

$$e(u) := \mathbb{E}[X - u \mid X > u], \quad u < x^* \quad (\text{bzw. } 0 \leq u < x^*)$$

mittlere Exzess-Funktion (mean excess function).

Die mittlere Exzess-Verteilung spielt vor allem in der Finanz- und Versicherungsmathematik eine wichtige

Rolle. Dort heißt die Quantilsfunktion auch Value at Risk  $\text{VaR}(q) := F^{\leftarrow}(q)$  und die folgende Funktion Expected Shortfall

$$\text{ES}(q) := \mathbb{E}(X \mid X > \text{VaR}(q)) = \text{VaR}(q) + e(\text{VaR}(q)).$$

BILD

Beides sind häufig betrachtete Riskikomaße. (Vorsicht: In der Literatur finden sich unterschiedliche Konventionen bezüglich der Verwendung von  $q$  oder  $1 - q$ . .) Andere Namen für die mittlere Exzess-Funktion schließen mean excess loss function (Versicherungskontext) oder mean residual life function (Medizin) ein.

## Kapitel 3

# Univariate Schätzverfahren

Viele Schätzverfahren basieren darauf, dass man eine stochastische Größe durch ihr empirisches Pendant ersetzt oder mit einem empirischen Pendant vergleicht. Wir fassen ein paar Vorüberlegungen/Wiederholungen zur Ordnungsstatistik und zur empirischen Verteilungsfunktion zusammen.

**Definition 3.0.1.** (Ordnungsstatistik)

Für Zufallszahlen  $X_1, \dots, X_n$  heißt ein monoton aufsteigendes Tupel  $X_{1:n}, \dots, X_{n:n}$ , welches (realisationsweise) durch Umordnen aus  $X_1, \dots, X_n$  entsteht, eine Ordnungsstatistik der  $X_1, \dots, X_n$ . Die Zufallsvariable  $X_{i:n}$  heißt auch  $i$ -te Ordnungsstatistik.

**Definition 3.0.2.** Sei  $X_1, \dots, X_n$  eine Stichprobe. Dann heißt die (zufällige) Funktion

$$F_n(t) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}} = \frac{|\{X_i \leq t \mid i \in \{1, \dots, n\}\}|}{n}$$

empirische Verteilungsfunktion. Es gelten folgende Aussagen (falls  $X_{1:n} < \dots < X_{n:n}$ ):

$$\begin{aligned} F_n(t) &= \frac{1}{n} \sum_{i=1}^n i \mathbf{1}_{[X_{i:n}, X_{i+1:n})}(t) \quad (\text{mit der Konvention } X_{n+1:n} := \infty) \\ F_n(X_{k:n}) &= \frac{k}{n}, k = 1, \dots, n \\ X_{k:n} &= F_n^{\leftarrow} \left( \frac{k}{n} \right) = F_n^{\leftarrow} \left( \frac{k}{n+1} \right) = F_n^{\leftarrow} \left( \frac{k-0.5}{n} \right), k = 1, \dots, n \\ X_{n-k:n} &= F_n^{\leftarrow} \left( \frac{n-k}{n} \right) = F_n^{\leftarrow} \left( 1 - \frac{k}{n} \right) = \left( \frac{1}{1-F_n} \right)^{\leftarrow} \left( \frac{n}{k} \right), k = 0, \dots, n-1 \end{aligned}$$

Das zu  $F_n$  gehörende Maß ist  $\frac{1}{n} \sum_{i=1}^n \delta_{X_i} = \frac{1}{n} \sum_{i=1}^n \delta_{X_{i:n}}$ . Der folgende Satz ist fundamental für die nicht-parametrische Statistik.

**Beispiel 3.0.3.** Sei  $F$  eine Verteilungsfunktion und man habe  $X_1, \dots, X_n$  Realisierungen dieser Verteilung. Wenn wir einen Schätzer für  $U(y) := F^{\leftarrow}(1 - \frac{1}{y})$  mit  $y \geq 1$  suchen, ist ein offensichtlicher Ansatz:

BILD

Im Fall, dass  $y := \frac{n}{k}$ ,  $k \leq n$  gilt

$$U(\frac{n}{k}) = F^{\leftarrow}(\frac{n-k}{n})$$

Wir ordnen die Realisierungen:  $X_{1:n} \leq \dots \leq X_{n:n}$  und wählen wegen oben den Schätzer für  $U(y)$  als

$$\widehat{U(\frac{n}{k})} = X_{n-k:n}$$

Dies gilt denn

$$\mathbb{P}(X \geq F^{\leftarrow}(1 - \frac{n}{k})) = \mathbb{P}(F(X) \geq 1 - \frac{n}{k}) = 1 - \mathbb{P}(F(X) \leq 1 - \frac{n}{k}) = 1 - (1 - \frac{n}{k}) = \frac{n}{k}$$

**Satz 3.0.4.** (Glivenko-Cantelli) Für eine u.i.v. Folge  $X_1, X_2, \dots \sim F$  gilt  $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0$  fast sicher.

Vor jeder tiefergehenden Datenanalyse sollte man sich stets einen Überblick über die vorliegenden Daten verschaffen, zum Beispiel mit Hilfe von Boxplots, Histogrammen oder empirischen Verteilungsfunktionen (oder Scatterplots bei multivariaten Daten). Bevor wir uns mit spezifischen Schätzverfahren der Extremwertstatistik auseinandersetzen, gehen wir kurz auf zwei nützliche allgemeine Diagnosewerkzeuge zur Modellüberprüfung ein.

### 3.0.1 Graphische Werkzeuge zur Modellverifikation:

#### PIT-Histogramm und QQ-Plot

Bei den folgenden graphischen Werkzeugen ist die Philosophie eher: SSpricht nichts gegen das Modell? (Darf die Hypothese, dass das vorliegende Modell wahr ist, als nicht signifikant ablehnbar eingestuft werden?) Ein "guter" Plot muss nicht bedeuten, dass das Modell "richtig" ist. Vielmehr zeigt ein "schlechter" Plot an, dass etwas nicht stimmt.

#### **PIT-Histogramm:**

Für eine Stichprobe  $X_1, \dots, X_n$  und eine stetige Verteilungsfunktion  $F$  heißt das Histogramm mit  $J$  Bins

mit jeweiliger Größe  $\left(\frac{j-1}{J}, \frac{j}{J}\right], j = 1, \dots, J$  mit den Höhen

$$f_j := \frac{\text{Anzahl der } i \in \{1, \dots, n\}, \text{ so dass } F(X_i) \in \left(\frac{j-1}{J}, \frac{j}{J}\right]}{n/J}$$

PIT-Histogramm (PIT=Probability Integral Transform). Ist  $F$  zusätzlich streng monoton, so gilt für die Höhen  $f_j$  im  $j$ -ten Bin, dass

$$J \cdot f_j = F_n \circ F^{\leftarrow} \left( \frac{j}{J} \right) - F_n \circ F^{\leftarrow} \left( \frac{j-1}{J} \right) .\text{????} \text{Nicht } 1/J?$$

Falls die Stichprobe  $X_1, \dots, X_n$  u.i.v. nach der stetigen (!) Referenzverteilung  $F$  generiert ist, sollte das PIT-Histogramm wie das Histogramm einer Gleichverteilung auf  $(0, 1]$  aussehen (vgl. Beweis von Satz 1.1.15 und Übung 1.1.13):

$$\mathbb{P}(F(X) \leq t) = \mathbb{P}(F^{\leftarrow}(F(X)) \leq F^{\leftarrow}(t)) = \mathbb{P}(X \leq F^{\leftarrow}(t)) = F(F^{\leftarrow}(t)) = t, \quad t \in (0, 1],$$

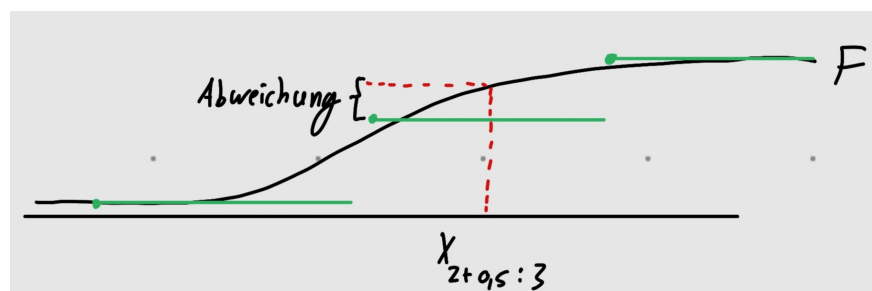
das heißt das Histogramm sollte aus ungefähr gleichhohen Balken der Höhe 1"bestehen. Bemerkung 2.0.14. Das PIT-Histogramm spielt bei der Kalibrierung von Vorhersageverteilungen in der Meteorologie eine wichtige Rolle. Meist werden dort (je nach Größe des Datensatzes) 10 oder 20 Bins betrachtet, siehe zum Beispiel Gneiting et al. (2007).

### QQ-Plot:

Für eine Stichprobe  $X_1, \dots, X_n$  und eine Verteilungsfunktion  $F$  heißt die folgende Menge QQ-Plot:

$$\{(F^{\leftarrow}(p_{k,n}), X_{k:n}) : k = 1, \dots, n\}.$$

Hierbei sind typische Wahlen  $p_{k,n} = k/(n+1)$  oder  $p_{k,n} = (k-0.5)/n$ , welche beide  $X_{k:n} = F_n^{\leftarrow}(p_{k,n})$  garantieren.





Der QQ-Plot liefert durch den Vergleich von theoretischen mit empirischen Quantilen eine graphische Methode, die Güte eines Fits zu beurteilen. Eigenschaften (siehe Embrechts, Klüppelberg, Mikosch):

(i) Vergleich von Verteilungen: Falls die Stichprobe  $X_1, \dots, X_n$  u.i.v. nach der Referenzverteilung  $F$  generiert ist, sollte der Plot ungefähr linear auf der Winkelhalbierenden liegen. (Die Hypothese, dass  $F$  die Daten gut beschreibt, kann nicht abgelehnt werden. Theoretische Rechtfertigung mittels Konvergenz empirischer Prozesse.)

(Siehe auch Resnick, 2006, für eine praktische Empfehlung ungefähr linear zu beurteilen. (Im Zweifel zusätzlich 100 Datensätze aus der Referenzverteilung  $F$  generieren und (in anderer Farbe) hinzufügen. Falls der ursprüngliche QQ-Plot innerhalb dieses Bandes liegt, ist die Variabilität akzeptabel.))

(ii) Ausreißer können leicht identifiziert werden.

(iii) Lokations- und Skalenparameter: Falls  $F$  nur bis auf Lokations- und Skalenparameter mit der Verteilung der u.i.v. Stichprobe  $X_1, \dots, X_n$  übereinstimmt, ist der Plot immer noch linear (nur nicht mehr auf der Winkelhalbierenden) und Lokations- und Skalenparameter können mittels Anstieg und Achsenabschnitt (durch lineare Regression) geschätzt werden (oftmals Startwert für weitere Schätzverfahren).

(iv) Formparameter: Einige Formunterschiede können abgeleitet werden, zum Beispiel, wenn die Daten einen schwereren oder leichteren Tail aufweisen als  $F$ .

**Bemerkung 3.0.5.** Ein Vorteil gegenüber dem PP-Plot  $\{(F(X_{k:n}), p_{k,n}) \mid 1 \leq k \leq n\}$  ist, dass man beim QQ-Plot besser erkennen kann, ob auch in den extremen Bereichen eine gute Anpassung vorliegt.

BILD

#### Annahme MAB( $\gamma$ )

Im Folgenden wollen wir stets annehmen (!), dass eine Stichprobe  $X_1, X_2, \dots$  von u.i.v. Beobachtungen aus einer Verteilung  $F$  im Max-Anziehungsbereich einer Extremwertverteilung  $G_\gamma$  vorliegt.

### 3.0.2 Block-Maxima - Anpassung einer GEV

**Beispiel 3.0.6.** (Wasserstände an einem Deich) Es liegen Daten von täglich gemessenen Wasserhöchstständen an einem Deich über einen Zeitraum von 100 Jahren vor. Frage: Wie hoch muss der Deich gebaut (die Höhe ist  $x_p$ ) werden, damit die Wahrscheinlichkeit für eine Überflutung innerhalb eines Jahres kleiner als eine vorgegebene Wahrscheinlichkeit  $p = 1/10000$  ist? Notieren wir mit  $X_i^{(j)}, i = 1, \dots, 365, j = 1, \dots, 100$  den

Wasserpegel am  $i$ -ten Tag im  $j$ -ten Jahr und unterstellen wir, dass die Jahresmaxima

$$M^{(j)} := \max \left( X_1^{(j)}, \dots, X_{365}^{(j)} \right), \quad j = 1, \dots, 100$$

u.i.v. sind wie eine Zufallsvariable  $M_{365}$ ??, so suchen wir also das Niveau  $x_p$ , so dass

$$\mathbb{P}(M_{365} > x_p) = p \quad \text{bzw.} \quad \mathbb{P}(M_{365} \leq x_p) = 1 - p. \quad (*)$$

Da  $p = 1/10000$  sehr, sehr klein ist, können wir sicherlich davon ausgehen, dass alle bisher gemessenen Wasserstände viel kleiner als die zu schätzende Deichhöhe  $x_p$  sind. Unter den getroffenen Modellannahmen ist (\*) äquivalent zu

$$\mathbb{E}(\text{Anzahl der Überflutungsjahre bei Deichhöhe } x_p \text{ in } 1/p = 10000 \text{ Jahren}) = 1,$$

denn

$$\mathbb{E} \left( \sum_{j=1}^{10000} \mathbf{1}_{\{M^{(j)} > x_p\}} \right) = 10000 \cdot \mathbb{P}(M_{365} > x_p) = 1/p \cdot p = 1.$$

Somit können wir obige Frage auch so formulieren: Frage: Wie hoch muss der Deich gebaut werden, damit ein Jahr mit Flut im Mittel nur alle  $1/p = 10000$  Jahre eintritt?

Übung 2.1.2. Man zeige außerdem  $\mathbb{E} \min \{j \geq 1 : M^{(j)} > x_p\} = 1/p$ .

In dieser Formulierung wird offenbar, warum man  $x_p$  auch als Wiederkehrniveau (Return Level)  $x_p$  zur Wiederkehrzeit  $1/p = 10000$  bezeichnet.

Wir wählen folgendes Modell:  $x_1, \dots, x_n$  sind Realisierungen von  $X_1, \dots, X_n$ , die u.i.v. Zufallsvariablen mit einer GEV-Verteilung  $G_{\gamma, \mu, \sigma}$  mit Parametervektor  $\theta = (\gamma, \mu, \sigma) \in \mathbb{R}^2 \times \mathbb{R}_+$  sind.

Achtung: oft schreiben wir  $G_\gamma$  und meinen eigentlich  $G_{\gamma; 0, 0}$ .

**Bemerkung 3.0.7.** Den jährlichen Maxima eine GEV-Verteilung zu unterstellen, ist eine natürliche Wahl, da jedes  $X_i$  ein Maximum von vielen u.i.v. Zufallsvariablen ist. Wir haben in früheren Kapiteln gezeigt, dass solche Maxima unter sehr allgemeinen Bedingungen gegen Extremwertverteilungen konvergieren. Denn der Satz (Fisher-Tippett '29, Gnedenko '43, de Haan '70) legt nahe, dass Maxima gleichmäßig gruppierter u.i.v. Daten annähernd GEV-verteilt sind, falls die Daten im Max-Anziehungsbereich einer Extremwertverteilung liegen ( $M^{(j)} \approx G_{\gamma; b_{365}, a_{365}}$ ).

Natürlich braucht man für die Konvergenz Normierungskonstanten, in unserem Fall kann man aber annehmen, dass die Normierungskonstanten bereits in den Parametern  $\mu$  und  $\sigma$  enthalten sind.

Damit man aber überhaupt Annehmen kann, dass die Maxima gegen eine GEV-Verteilung konvergieren, ist vorausgesetzt, dass diese u.i.v. verteilt sind. Dies ist zu Überprüfen.

**Bemerkung 3.0.8.** Da wir im obigen Modell voraussetzen, dass die  $X_i$  identisch verteilt sind, kann das Modell nur auf stationäre Daten angewendet werden, d.h. Daten, die keinen Trend aufweisen. Werden die jährlichen Maxima mit der Zeit immer größer (kleiner), muss ein anderes Modell verwendet werden, siehe unten.

**Beispiel 3.0.9.** (Fortsetzung Beispiel 1.2.4) Uns ist bereits bekannt, dass das Maximum von u.i.v. Standard-exponentialverteilten Zufallsvariablen  $X_1, \dots, X_n$  wie folgt durch eine Gumbelverteilung approximiert werden kann

$$M_n \approx \Lambda(x - \log n) = G_{0; \log n, 1}(x), \quad n \gg 0.$$

Liegen nun  $n \cdot m$  Beobachtungen  $X_i^{(j)}, \dots, X_i^{(j)}, i = 1, \dots, n, j = 1, \dots, m$  vor, so sollten die (unnormierten) Maxima

$$M^{(j)} := \max_{i=1, \dots, n} X_i^{(j)}, \quad j = 1, \dots, m,$$

annähernd eine u.i.v. Stichprobe aus der GEV-Verteilung  $G_{0; \log n, 1}(x)$  repräsentieren. Satz 1.2.10 (Fisher-Tippet '29, Gnedenko '43, de Haan '70) motiviert folgendes Vorgehen:

BILD

### Block-Maxima – Anpassung einer GEV

- (1) Gruppieren der Daten in aufeinanderfolgende Blöcke gleicher Länge,
- (2) Anpassung einer GEV an die Block-Maxima,
- (3) Verifikation des geschätzten Modells,
- (4) Schätzen weiterer Größen von Interesse (Extrapolation).

Zu (1): Wahl der Blockgröße: Ein praktisches Problem ist die Wahl der Blockgröße, die einen Kompromiss zwischen Bias und Varianz darstellt:

- Sind die Blöcke zu klein, ist die Approximation durch eine GEV noch nicht gerechtfertigt. Dies führt zu einem hohen Bias bei der Schätzung und Extrapolation.
- Sind die Blöcke sehr groß, so liegen nur sehr wenige Block-Maxima vor. Dies führt zu einer hohen Varianz der Schätzung.

- Die Wahl der Blockgröße unterliegt oftmals praktischen Überlegungen. Bei Zeitreihen mit täglichen Daten sind Jahresmaxima typisch. Dadurch werden auch saisonale Effekte umgangen.

Zu (2): Anpassung einer GEV: Wir wollen an dieser Stelle auf die Schätzung der GEV-Parameter  $\theta = (\gamma, \mu, \sigma) \in \mathbb{R}^2 \times \mathbb{R}_+$  mittels Maximum-Likelihood näher eingehen.

Einschub: Maximum-Likelihood (ML) für u.i.v. Beobachtungen

Es wird angenommen, dass es eine Verteilung  $F_{\theta_0}$  gibt, welche die u.i.v. Beobachtungen  $Y_1, \dots, Y_m$  generiert und dass  $F_{\theta_0}$  zu einer parametrischen Familie  $\{F_{\theta} : \theta \in \Theta\}$  von Verteilungen gehört mit  $\Theta \subset \mathbb{R}^d$ . Ein ML-Schätzer für den wahren Parameter  $\theta_0 \in \Theta$  maximiert die (log-)Likelihood, d.h.

$$\hat{\theta}_0 := \operatorname{argmax}_{\theta \in \Theta} L(\theta | Y) = \operatorname{argmax}_{\theta \in \Theta} \ell(\theta | Y)$$

wobei  $L(\theta | y)$  die Dichte-Funktion von u.i.v. Variablen  $Y_1, \dots, Y_m$  mit Verteilung  $F_{\theta}$  ist und  $\ell = \log L$ , d.h.

$$L(\theta | y) = \prod_{i=1}^m f_{\theta}(y_i) \quad \text{bzw.} \quad \ell(\theta | y) = \sum_{i=1}^m \log f_{\theta}(y_i).$$

Dazu benötigt man die Dichte  $f_{\gamma, \mu, \sigma}(z)$  der GEV-Verteilung. Durch Ableiten der Verteilungsfunktion  $G_{\gamma, \mu, \sigma}$  erhält man, dass für  $\gamma \neq 0$

$$f_{\theta}(z) = f_{\gamma, \mu, \sigma}(z) = \begin{cases} \frac{1}{\sigma} \left(1 + \gamma \frac{z - \mu}{\sigma}\right)^{-\frac{1}{\gamma} - 1} \exp \left\{ - \left(1 + \gamma \frac{z - \mu}{\sigma}\right)^{-\frac{1}{\gamma}} \right\}, & 1 + \gamma \frac{z - \mu}{\sigma} > 0, \\ 0, & \text{sonst} \end{cases}$$

während für  $\gamma = 0$

$$f_{\theta}(z) = f_{0, \mu, \sigma} = \frac{1}{\sigma} e^{-\frac{z - \mu}{\sigma}} \exp \left\{ -e^{-\frac{z - \mu}{\sigma}} \right\}, \quad \text{für } z \in \mathbb{R}.$$

Mit Hilfe der Dichten kann man die Log-Likelihoodfunktion aufstellen und unsere Dichte einsetzen:

Die log-Likelihood ist hier

$$\begin{aligned} \ell(\gamma, \mu, \sigma | x) &= \begin{cases} -m \log \sigma - (1 + 1/\gamma) \sum_{i=1}^m \log \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]_+ \\ \quad - \sum_{i=1}^m \left[ 1 + \gamma \left( \frac{x_i - \mu}{\sigma} \right) \right]_+^{-1/\gamma} & \gamma \neq 0, \\ -m \log \sigma - \sum_{i=1}^m \left( \frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left[ - \left( \frac{x_i - \mu}{\sigma} \right) \right] & \gamma = 0. \end{cases} \end{aligned}$$

Es stellt sich nun die Frage, ob die Regularitätsbedingungen für eine ML-Schätzung erfüllt sind. Diese wurde von Smith (1985) detailliert untersucht.

- Falls  $\gamma > -0.5$  sind ML-Schätzer regulär mit den gewöhnlichen asymptotischen Eigenschaften.
- Falls  $\gamma \in (-1, -0.5)$  sind ML-Schätzer anwendbar, besitzen aber nicht mehr die Standard-Eigenschaften.
- Falls  $\gamma < -1$ , sind ML-Schätzer nicht verfügbar.

Asymptotische Eigenschaften:

Unter geeigneten Regularitätsbedingungen (siehe zum Beispiel Casella, Berger (2001) oder Newey, MacFadden (1994)) ist der MLSchätzer

- konsistent  $(\hat{\theta}_0 \xrightarrow{P} \theta_0 \text{ für } m \rightarrow \infty)$  oder sogar
- stark konsistent  $(\hat{\theta}_0 \xrightarrow{f.s.} \theta_0 \text{ für } m \rightarrow \infty)$  und
- asymptotisch normal mit

$$\hat{\theta}_0 \rightarrow^d \mathcal{N}(\theta_0, I_E(\theta_0)^{-1}), \text{ für } m \rightarrow \infty, \text{ wobei } I_E(\theta_0) = \mathbb{E} \left( - \frac{\partial^2 \ell(\theta | Y)}{\partial \theta_k \partial \theta_\ell} \Big|_{\theta=\theta_0} \right)_{k, \ell=1, \dots, d}.$$

Die Fisher-Informationsmatrix  $I_E(\theta_0)$  kann mittels  $I_O(\hat{\theta}_0) = - \frac{\partial^2 \ell(\theta | Y)}{\partial \theta_k \partial \theta_\ell} \Big|_{\theta=\hat{\theta}_0}$  geschätzt werden, um Konfidenzbereiche zu konstruieren. Genauere (aber rechenintensivere) Konfidenzbereiche ergeben sich in der Regel aus der asymptotischen  $\chi^2$ -Verteilung der Abweichung (Deviance)  $D(\theta_0) = 2(\ell(\hat{\theta}_0) - \ell(\theta_0)) \approx \chi_d^2$  (oder unter Einbezug der Profile-log-Likelihood).

Vorteile der ML-Methode:

- Die asymptotischen Eigenschaften lassen sich in der Regel unmittelbar auf  $g(\theta)$  übertragen, falls  $g(\theta)$  eine von  $\theta$  abhängige Größe ist (zum Beispiel ein hohes Quantil). Der ML-Schätzer für  $g(\theta)$  ist dann  $g(\hat{\theta}_0)$ . Konfidenzbereiche lassen sich mit der Delta-Methode ermitteln.
- Das ursprüngliche Modell mit u.i.v. Beobachtungen kann leicht zu einem komplexeren Modell erweitert werden. Die  $k$ -te Komponente  $\theta_k$  des Parameters  $\theta$  könnte beispielsweise wie bei einer linearen Regression mit einem linearen Trend  $\theta_k(i) = \alpha_k + \beta_k w_k(i)$  modelliert werden, wobei  $w_k(i)$  ein bekanntes Gewicht für die  $i$ -te Beobachtung ist. Der ML-Schätzer für  $(\alpha, \beta)$  besitzt dann ähnliche asymptotische Eigenschaften. ("Verträglichkeit mit Projektion (auf andere Größen) und Einbettung (komplexerer Modelle)")

Zu (3): Modellverifikation: Bevor weitere Schlüsse gezogen werden, sollte nach dem Schätzen der GEV-Parameter die Plausibilität des geschätzten Modells überprüft werden (z.B. PIT-Histogram, QQPlot,

Return-Level-Plot).

Zu (4): Schätzen weiterer Größen (Extrapolation):

Nachdem der Parameter  $\theta$  geschätzt wurde, können wir die Deichhöhe  $x_p$  schätzen. Wir erinnern, dass  $x_p$  die Deichhöhe ist, bei der eine Überflutung mit Wahrscheinlichkeit  $p$  in einem Jahr stattfindet. Das Problem besteht also darin, dass  $(1 - p)$ -Quantil des jährlichen Maximums zu schätzen. Wir schätzen  $x_p$  indem wir die Gleichung

$$G_{\hat{\gamma}, \hat{\mu}, \hat{\sigma}}(\hat{x}_p) = 1 - p$$

lösen (falls es mehrere Lösungen gibt, betrachten wir die kleinste):

Da die Pseudoinverse für beide Fälle der GEV bijektiv ist, können wir einfach die Umkehrfunktion betrachten, d.h. für den Fall  $\gamma = 0$  (der Fall  $\gamma \neq 0$  völlig Analog)

$$\begin{aligned} y &:= \exp\left(-e^{-\frac{x-\mu}{\sigma}}\right) &\iff & \log(y) = -e^{-\frac{x-\mu}{\sigma}} \\ &\iff \log(-\log(y)) = -\frac{x-\mu}{\sigma} &\iff & -\sigma \log(-\log(y)) = x - \mu \\ &\iff \mu - \sigma \log(-\log(y)) = x \end{aligned}$$

Damit können wir die Schätzer für die Parameter der GEV einsetzen

$$\hat{x}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}} \{1 - (-\log(1 - p))^{-\hat{\gamma}}\}, & \hat{\gamma} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log(-\log(1 - p)), & \hat{\gamma} = 0. \end{cases}$$

Für  $\hat{\gamma} < 0$  (im Fall der Weibull-Verteilung) besitzt die Verteilung  $G_{\hat{\gamma}, \hat{\mu}, \hat{\sigma}}$  einen endlichen rechten Endpunkt, der übrigens per Definition  $x_0$  ist. In diesem Fall gehen wir davon aus, dass es einen absolut höchsten Wasserstand gibt, der niemals überschritten wird. Der Schätzer für  $x_0$  ist dann gegeben durch:

$$\hat{x}_0 = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\gamma}}$$

### 3.0.3 Peaks-Over-Threshold (POT) - Anpassung einer GPD

Bei der Block-Maxima-Methode fällt auf, dass ein Großteil der Daten gar nicht in die Schätzung mit einbezogen wird. Der Satz 1.4.4 (Pickands, Balkema, de Haan) motiviert eine alternative Methode (die schon lange in der Hydrologie verwendet wird). Zur Erinnerung: Falls  $X$  im Max-Anziehungsbereich von  $G_\gamma$  ist,

gilt

$$X - u \mid X > u \approx H_{\gamma; \sigma(u)} \quad \text{für } u \gg 0.$$

BILD

### Peaks-Over-Threshold (POT) - Anpassung einer GPD

- (1) Bestimmen einer Schwelle  $u_0$ , so dass die Exzesse  $X_i - u_0 \mid X_i > u_0$  annähernd GPD-verteilt sind,
- (2) Anpassung einer GPD an die Exzesse  $X_i - u_0 \mid X_i > u_0$ ,
- (3) Verifikation des geschätzten Modells,
- (4) Schätzen weiterer Größen von Interesse (Extrapolation).

#### Zu (1): Wahl der Schwelle $u_0$

Falls  $Y_0 := X - u_0 \mid X > u_0 \sim H_{\gamma; \sigma_0}$ , so gilt auch für jedes  $u_1 > u_0$ , dass

$$\begin{aligned} Y_1 := X - u_1 \mid X > u_1 &= Y_0 + u_0 - u_1 \mid Y_0 + u_0 > u_1 \\ &= Y_0 - (u_1 - u_0) \mid Y_0 > (u_1 - u_0) \stackrel{\text{Blatt 6}}{\sim} H_{\gamma; \sigma_0 + \gamma(u_1 - u_0)} \quad (*) \end{aligned}$$

Man versucht stets eine möglichst kleine Schwelle  $u_0$  zu verwenden, ab der eine Approximation der Exzesse durch eine GPD gerade noch gerechtfertigt erscheint.

Die Wahl der Schwelle  $u_0$  stellt (ähnlich wie bei der Blockgrößenwahl bei der BlockmaximaMethode) wiederum einen Kompromiss zwischen Bias und Varianz dar:

- Ist die Schwelle zu klein, ist die Approximation durch eine GPD noch nicht gerechtfertigt. Dies führt zu einem hohen Bias bei der Schätzung und Extrapolation.
- Ist die Schwelle sehr groß, so bleiben nur sehr wenige Überschreitungen übrig. Dies führt zu einer hohen Varianz bei der Schätzung.
- Theoretisch ist eine asymptotisch optimale Wahl von  $u$  möglich. Bei realen Daten gibt es verschiedene Ansätze.

Eine Möglichkeit besteht darin (unter der zusätzlichen Annahme  $\gamma < 1$ ) eine Schwelle  $u_0$  anhand eines Mean-Excess-Plots zu ermitteln.

Mean-Excess-Plot/Mean-Residual-Life-Plot: Für eine Stichprobe  $X_1, \dots, X_n$  ist wegen

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}} \quad \text{d.h. es ist} \quad \bar{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i > t\}}$$

wie folgt

$$\begin{aligned} e_n(u) &\stackrel{Blatt6}{=} \frac{1}{\bar{F}_n(u)} \int_u^\infty \bar{F}_n(y) dy \\ &= \frac{1}{\sum_{i=1}^n \mathbf{1}_{\{X_i > u\}}} \sum_{i=1}^n \int_u^\infty \mathbf{1}_{\{X_i > y\}} dy \\ &= \frac{1}{\sum_{i=1}^n \mathbf{1}_{\{X_i > u\}}} \sum_{i=1}^n \int_u^{X_i} dy \\ &= \frac{1}{\sum_{i=1}^n \mathbf{1}_{\{X_i > u\}}} \sum_{i=1}^n (X_i - u)_+, \quad u \geq 0 \end{aligned}$$

die empirische mittlere Exzess-Funktion. Dann heißt die folgende Menge Mean-ExcessPlot (oder Mean-Residual-Life-Plot):

$$\{(X_{k:n}, e_n(X_{k:n})) : k = 1, \dots, n\}.$$

(Bemerkung: Konfidenzintervalle können aus asymptotischer Normalität von Stichprobenmittelwerten gewonnen werden.)

BILD

Weil wir angenommen/verifiziert haben, dass für ein groß genug  $u$   $X - u$  ungefähr der Generalisierten Paretoverteilung folgt und da für  $X \sim \text{Pareto}(\sigma, \gamma)$ :

$$\mathbb{E}[u] = \frac{\sigma}{1 - \gamma}$$

Falls  $Y_0 := X - u_0 \mid X > u_0 \sim H_{\gamma; \sigma_0}$  mit  $\gamma < 1$ , so folgt für  $u_1 > u_0$

$$e(u_1) = \mathbb{E}(X - u_1 \mid X > u_1) = \mathbb{E}(Y_0 - (u_1 - u_0) \mid Y_0 > (u_1 - u_0)) \stackrel{Blatt6}{=} \frac{\sigma_0 + \gamma(u_1 - u_0)}{1 - \gamma}.$$

Das heißt der Mean-Excess-Plot sollte ab  $u_0$  "ungefähr linear" sein, denn  $e(u_1)$  ist für  $u_1 > u_0$  eine Lineare Funktion. In der Praxis ist die Beurteilung von "ungefähr linear" nicht immer ganz einfach. Für zu große  $u$ , ist die Information durch den Plot nicht mehr verlässlich, da kaum noch Daten zur Ermittlung von  $e(u)$  vorliegen.



**Bemerkung 3.0.10.** Darüber hinaus liefert der Mean-Exzess-Plot eine graphische Methode, die "Schwere" eines Tails zu beurteilen. Allerdings ist große Vorsicht geboten, da der Plot gegenüber kleinen Veränderungen in den Daten nicht sehr robust ist (Eine robustere Alternative wäre ein Median-Excess-Plot).

Eine andere Methode besteht darin, erst für mehrere Schwellenwerte  $u$  eine GPD an die Exzesse  $X_i - u \mid X_i > u$  anzupassen (siehe Schritt (2)) und die Stabilität der Parameterschätzung auszuwerten.

Wir haben ja gezeigt, dass falls die Exzesse mit der Threshold  $u_0$  Generalisiert Pareto verteilt sind, dann sind auch die Exzesse für  $u_1 > u_0$  Generalisiert Pareto verteilt. Wenn wir aber die Parameter der Generalisierten Pareto Verteilung für die Exzesse mit Threshold  $u_1$  betrachten, dann gilt gemäß (\*) dass die Funktion des Zweiten Parameters in Abhängigkeit von  $u$  wie folgt aussieht.

$$\sigma(u) = \sigma(u) + \gamma(u)$$

D.h. die Schätzer für  $\sigma^*$  sollten für Große  $u$  konstant sein, denn wir haben ja Anfang gesagt, dass nur für große  $u$  der Excess mit Threshold  $u$  GPD ist.

BILD

Abbildung 5: Verschiedene Schätzwerte für  $\sigma^*(u)$  und  $\gamma(u)$  für verschieden Schwellen  $u$  bei Regendaten (Beispiel aus Coles, 2001)

Zu(2) : Anpassung einer GPD an die Exzesse  $X_i - u \mid X_i > u$  :

Die Exzesse  $y_1, \dots, y_k$  sind Realisierungen von unabhängigen und identisch verteilten Zufallsvariablen  $Y_1, \dots, Y_k$ , die gemäß einer verallgemeinerten Pareto-Verteilung  $H_{\gamma, \sigma}$  verteilt sind. Dabei sind  $\gamma \in \mathbb{R}$  und  $\sigma > 0$  unbekannte Parameter. Die Dichte von  $H_{\gamma, \sigma}$  ist gegeben durch

$$f_{\gamma, \sigma}(t) = \frac{1}{\sigma} \left( 1 + \frac{\gamma t}{\sigma} \right)^{-\frac{1}{\gamma} - 1} \quad \text{mit} \quad \begin{cases} t > 0, & \text{falls } \gamma > 0, \\ t \in \left[ 0, -\frac{\sigma}{\gamma} \right], & \text{falls } \gamma < 0, \end{cases}$$

und  $f_{0, \sigma}(t) = (1/\sigma)e^{-t/\sigma}, t > 0$ , für  $\gamma = 0$ . Damit ergibt sich für die Log-Likelihoodfunktion

$$l(\gamma, \sigma) := \sum_{i=1}^k \log f_{\gamma, \sigma}(y_i) = -k \log \sigma - \left( 1 + \frac{1}{\gamma} \right) \sum_{i=1}^k \log \left( 1 + \frac{\gamma y_i}{\sigma} \right),$$

zumindest für  $\gamma \neq 0$ . Der ML-Schätzer

$$(\hat{\gamma}, \hat{\sigma}) = \underset{\gamma, \sigma}{\operatorname{argmax}} l(\gamma, \sigma)$$

muss numerisch ermittelt werden.

Zu (3): Modellverifikation: Bevor weitere Schlüsse gezogen werden, sollte nach dem Schätzen der GPD-Parameter die Plausibilität des geschätzten Modells überprüft werden (z.B. PIT-Histogram, QQPlot, Return-Level-Plot).

Zu (4): Schätzen weiterer Größen von Interesse (Extrapolation):

Nun werden wir für ein gegebenes kleines  $p$  die Deichhöhe  $x_p$  schätzen, die an einem Tag mit Wahrscheinlichkeit  $p$  überflutet wird.

Es sei  $X$  die Zufallsvariable, die den Wasserstand an einem Tag beschreibt. Mit Berücksichtigung des Satzes von Pickands-Balkema-de Haan gehen wir davon aus, dass für großes  $u$ , dass  $X - u \mid X > u \sim H_{\gamma;\sigma}$ , dann gilt für die Excess Funktion :

$$F_u(x) = \mathbb{P}(X - u \leq x \mid X > u) = (1 + \log(G_{\gamma;\sigma}(t)))_+ = \begin{cases} (1 - (1 + \gamma \frac{x}{\sigma})_+^{-\frac{1}{\gamma}})_+ & , \gamma \neq 0 \\ (1 - e^{-\frac{x}{\sigma}})_+ & , \gamma = 0 \end{cases}$$

damit gilt für die Tailfunktion

$$\bar{F}_u(x) = \mathbb{P}[X - u > t \mid X > u] \approx \left(1 + \frac{\gamma t}{\sigma}\right)^{-\frac{1}{\gamma}}.$$

Mit  $t = x_p - u$  folgt also:

$$\mathbb{P}[X > x_p] \approx \mathbb{P}[X > u] \left(1 + \gamma \frac{x_p - u}{\sigma}\right)^{-\frac{1}{\gamma}}.$$

Nun setzen wir die rechte Seite gleich  $p$ . Wenn die Gleichung nach  $x_p$  umgestellt wird, erhält man schließlich

$$x_p \approx u + \frac{\sigma}{\gamma} \left[ \left( \frac{\mathbb{P}[X > u]}{p} \right)^{\gamma} - 1 \right].$$

Dabei haben wir Schätzer für  $\sigma$  und  $\gamma$  bereits hergeleitet. Die Wahrscheinlichkeit  $\mathbb{P}[X > u]$  kann durch  $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i > u\}}$  geschätzt werden. Dieser ist  $\text{Bin}(n, \mathbb{P}(X > u_0))$  Verteilt. Es ergibt sich der Schätzer

$$\hat{x}_p = \begin{cases} u + \frac{\hat{\sigma}}{\hat{\gamma}} \left[ \left( \frac{\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i > u\}}}{p} \right)^{\hat{\gamma}} - 1 \right] & \gamma \neq 0 \\ u_0 + \hat{\sigma} \log \left( \frac{\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{x_i > u\}}}{p} \right) & \gamma = 0 \end{cases}$$

bzw. auch für den oberen Endpunkt der Verteilung

$$\hat{x}_0 = u_0 - \frac{\hat{\sigma}}{\hat{\gamma}}, \quad \text{falls } \hat{\gamma} < 0$$

Im Vergleich mit der Block-Maxima-Methode würde man wegen

$$\begin{aligned} 1 &= \mathbb{E} \left( \text{Anzahl an Überschreitungen von } x_{p/365} \text{ innerhalb von } 365/p \text{ Tagen} \right) \\ &= \mathbb{E} \left( \text{Anzahl an Überschreitungen von } x_{p/365} \text{ innerhalb von } 1/p \text{ Jahren} \right) \end{aligned}$$

das geschätzte Wiederkehrniveau  $\hat{x}_{p/365}$  betrachten (obwohl innerhalb eines Jahres natürlich mehrere Überschreitungen stattfinden können).

BILD

## 3.1 Weitere Schätzverfahren für den Extremwertindex

Frage: Wie ist der Extremwertindex definiert? Warum interessiert man sich dafür?

- Welches Verhalten ist "typisch"? (Finanzdaten:  $\gamma \in (1/4, 1/3)$ , Versicherungsdaten  $\gamma \in (1/2, 1)$ , Winddaten:  $\gamma \in (-0.2, 0.1)$ )
- Existiert Korrelation? Greift ZGWS?
- Kann man noch Konfidenzintervalle konstruieren? Auch für die Korrelationsschätzer? (Existieren noch 4. Momente?)
- Grundlage für Schätzverfahren für hohe Quantile, stochastische Größen in Risikobereichen.
- Diversifikation von Portfolios sinnvoll? (Siehe Übungen.)

Die folgenden Schätzer sind jeweils von der Form  $\hat{\gamma}_{k,n}$ . Was bedeuten hierbei  $n$  und  $k$ ?

### 3.1.1 Pickands-Schätzer: $\gamma \in \mathbb{R}$

Idee: Aus den Übungen wissen wir bereits, dass eine positive Funktion  $a$  existiert, so dass für  $U(y) := F^{\leftarrow}(1 - y^{-1})$  gilt

$$\lim_{s \rightarrow \infty} \frac{U(sy) - U(s)}{a(U(s))} = \begin{cases} \frac{y^\gamma - 1}{\gamma}, & \gamma \neq 0, \\ \log(y), & \gamma = 0. \end{cases}$$

Es folgt für  $x, y > 0, y \neq 1$ , dass

$$\lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{a(U(s))} \frac{a(U(s))}{U(sy) - U(s)} = \lim_{s \rightarrow \infty} \frac{U(sx) - U(s)}{U(sy) - U(s)} = \begin{cases} \frac{x^\gamma - 1}{y^\gamma - 1}, & \gamma \neq 0, \\ \frac{\log(x)}{\log(y)}, & \gamma = 0. \end{cases}$$

.....Aufschrieb..

Insbesondere mit  $x = 2$  und  $y = 1/2$  (und Multiplizieren mit -1 )

$$\lim_{s \rightarrow \infty} \frac{U(2s) - U(s)}{U(s) - U(s/2)} = \begin{cases} -\frac{2^\gamma - 1}{(1/2)^\gamma - 1}, & \gamma \neq 0, \\ -\frac{\log(2)}{\log(1/2)}, & \gamma = 0. \end{cases} = 2^\gamma.$$

denn

$$-\frac{2^\gamma - 1}{(1/2)^\gamma - 1} = 2^\gamma \cdot (-1) \cdot \frac{1 - 2^{-\gamma}}{(1/2)^\gamma - 1} = 2^\gamma \frac{2^{-\gamma} - 1}{(1/2)^\gamma - 1} = 2^\gamma$$

Ersetzen wir nun  $2s$  durch  $n/k$  sowie  $U = (1/(1-F))^\leftarrow$  durch die empirische Variante  $U_n = (1/(1 - F_n))^\leftarrow$ , erhalten wir

$$\lim_{s \rightarrow \infty} \frac{U(\frac{n}{k}) - U(\frac{n}{2k})}{U(\frac{n}{2k}) - U(\frac{n}{4k})} = 2^\gamma \iff \lim_{s \rightarrow \infty} \frac{1}{\log(2)} \log \left( \frac{U(\frac{n}{k}) - U(\frac{n}{2k})}{U(\frac{n}{2k}) - U(\frac{n}{4k})} \right) = \gamma$$

den folgenden (Pickands-) Schätzer

$$\hat{\gamma}_{k,n}^{(P)} := \frac{1}{\log(2)} \log \left( \frac{X_{n-k:n} - X_{n-2k:n}}{X_{n-2k:n} - X_{n-4k:n}} \right)$$

(Siehe Unterkapitel mit Ordnungstatistik)

Pickands-Plot Für eine Stichprobe  $X_1, \dots, X_n$  heißt die folgende Menge PickandsPlot

$$\left\{ \left( k, \hat{\gamma}_{k,n}^{(P)} \right) : k = 1, \dots, [n/4] \right\}$$

Bei steigendem Stichprobenumfang  $n$  sollte sicherlich auch  $k = k_n$  größer gewählt werden, da wir uns für eine Eigenschaft des Tails interessieren. Andererseits müssen wir sicherstellen, dass  $2s = n/k = n/k_n \rightarrow \infty$ , da der Schätzer durch die Konvergenz (20) motiviert ist.

**Definition 3.1.1.** Eine Folge  $(k_n)_{n \in \mathbb{N}}$  mit  $k_n \rightarrow \infty$  und  $k_n/n \rightarrow 0$  für  $n \rightarrow \infty$  heißt intermediäre Folge.

### 3.1.2 Hill-Schätzer: $\gamma > 0$

Idee: Da zusätzlich  $\gamma > 0$  angenommen wird, dass  $F$  im Anziehungsbereich von  $G_\gamma$  liegt, wissen wir nach Satz 1.3.18, dass  $1 - F \in RV_{-1/\gamma}$ , denn  $\frac{1}{\gamma} > 0$ . Außerdem schätzen wir beim Hill-Schätzer nur den positivteil von  $\gamma$ , d.h. wenn der Hill-Schätzer für  $\gamma$  gleich Null ist, dann wissen wir nur das der Postivteil Null ist und haben keine Information den Negativteil. Entsprechend gilt

$$h(x) = \frac{1 - F(x)}{x} \in RV_{-(\frac{1}{\gamma}+1)},$$

wobei  $-(\frac{1}{\gamma} + 1) < -1$ . Nach Satz 1.3.13 (Karamata) folgt

$$\lim_{t \rightarrow \infty} \frac{th(t)}{\int_t^\infty h(s)ds} = - \left( - \left( \frac{1}{\gamma} + 1 \right) + 1 \right) = \frac{1}{\gamma}$$

bzw.

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty 1 - F(s) \frac{ds}{s}}{1 - F(t)} = \gamma.$$

Partielle Integration liefert

$$\lim_{t \rightarrow \infty} \frac{\int_t^\infty \log(s) - \log(t) dF(s)}{1 - F(t)} = \gamma.$$

Ersetzen wir nun  $F$  durch die empirische Verteilungsfunktion  $F_n$  und  $t$  durch die Ordnungsstatistik  $X_{n-k:n}$  für eine intermediäre Folge  $k = k_n$ , erhalten wir den (Hill-)Schätzer

$$\hat{\gamma}_{k,n}^{(H)} := \frac{1}{k} \sum_{i=0}^{k-1} \log(X_{n-i:n}) - \log(X_{n-k:n}).$$

denn

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\int_{X_{n-k:n}}^\infty \log(s) - \log(X_{n-k:n}) dF_n(s)}{1 - F_n(X_{n-k:n})} &= \lim_{k \rightarrow \infty} \frac{\frac{1}{n} \sum_{i=0}^{k-1} \log(X_{n-i:n}) - \log(X_{n-k:n})}{1 - \frac{n-k}{n}} \\ &= \lim_{k \rightarrow \infty} \frac{\frac{1}{n} \sum_{i=0}^{k-1} \log(X_{n-i:n}) - \log(X_{n-k:n})}{\frac{k}{n}} \\ &= \lim_{k \rightarrow \infty} \frac{\sum_{i=0}^{k-1} \log(X_{n-i:n}) - \log(X_{n-k:n})}{k} = \gamma \end{aligned}$$

Hill-Plot Für eine Stichprobe  $X_1, \dots, X_n$  heißt die folgende Menge Hill-Plot

$$\left\{ \left( k, \hat{\gamma}_{k,n}^{(H)} \right) : k = 1, \dots, n-1 \right\}.$$

Eigenschaften:

**Satz 3.1.2.** (EKM, Theorem 6.4.6)

Unter Annahme  $MAB(\gamma)$  gilt:

a) (Schwache Konsistenz) Falls  $k \rightarrow \infty, k/n \rightarrow 0$  für  $n \rightarrow \infty$ , dann

$$\hat{\gamma}_{k,n}^{(H)} \xrightarrow{P} \gamma \quad (\text{stochastisch}).$$

b) (Starke Konsistenz) Falls  $k/n \rightarrow 0, k/\log \log n \rightarrow \infty$  für  $n \rightarrow \infty$ , dann

$$\hat{\gamma}_{k,n}^{(H)} \xrightarrow{\text{f.s.}} \gamma \quad (\text{fast sicher}).$$

c) (Asymptotische Normalität) Unter weiteren Annahmen (zweiter Ordnung, siehe später) gilt

$$\sqrt{k} \left( \hat{\gamma}_{k,n}^{(H)} - \gamma \right) \rightarrow^d \mathcal{N}(0, \gamma^2)$$

### 3.1.3 Dekkers-Einmahl-de Haan-Schätzer (Momenten-Schätzer):

$\gamma \in \mathbb{R}$

Idee: Der Hill-Schätzer gibt keine Auskunft über den Fall  $\gamma \leq 0$ . Tatsächlich gilt für eine intermediäre Folge  $k = k_n$ , dass  $\hat{\gamma}_{k,n}^{(H)} \rightarrow^P 0$  für Daten aus dem Gumbel- oder Weibull-Anziehungsbereich, das heißt

$$\gamma_{k,n}^{(H)} \rightarrow^P \gamma_+.$$

Um den Hill-Schätzer  $\hat{\gamma}_{k,n}^{(H)}$  um eine Komponente zu ergänzen, die gegen  $\gamma_- = \min(\gamma, 0)$  konvergiert, führen Dekkers, Einmahl, de Haan (1989) die folgenden Ausdrücke ein

$$M_{k,n}^{(r)} := \frac{1}{k} \sum_{i=0}^{k-1} (\log(X_{n-i:n}) - \log(X_{n-k:n}))^r, \quad r = 1, 2.$$

Für  $r = 1$  ist dies einfach der Hill Schätzer.

Dann gilt

$$\gamma_{k,n}^{(DedH)} := M_{k,n}^{(1)} + \left(1 - \frac{1}{2} \left(1 - \frac{M_{k,n}^{(1)}}{M_{k,n}^{(2)}}\right)^{-1}\right) \rightarrow \delta_+ + \delta_- = \delta$$

Ein Dekkers-de Haan bzw. Hill Plot sieht dann wie folgt aus

BILD

Man nehme das  $k$ , wo der Plot einigermaßen stabil und konstant ist. Für dieses  $\tilde{k}$  ist dann der  $\gamma_{\tilde{k},n}^{(DedH)}$  der Schätzer für  $\gamma$ .

## 3.2 Extrapolation in Risikobereiche

Ausgehend von den bisher betrachteten Schätzern wollen wir noch Schätzer zur Extrapolation in Risikobereiche konstruieren. Neben der Block-Maxima- und der POT-Methode kann man hierfür auch die im Abschnitt 2.3 KS betrachteten Schätzer als Grundlage nehmen. Wir gehen auf den Fall  $\gamma = 1/\alpha > 0$  näher ein, also Frechet. Wir machen also genau das Selbe, was wir schon davor gemacht haben, jedoch nun mit einem anderen Ansatz, anderen Schätzern und auch nur auf den für Freshet Verteilte Zufallsvariablen.

- Bei den Blockmaxima haben wir das Maximum über eine gleiche Menge an ZV betrachtet. Unser Schätzer für  $x_p$  war dort einfach die Schätzer der Parameter der GEV eingesetzt in das Pseudoinverse der GEV. Denn das die Verteilung der Maxima läuft gegen die GEV für  $n$  gegen unendlich.
- Bei der POT-Methode wird der einmal der Threshold verwendet, sodass der Abstand der Punkte die drüber liegen zum Threshold generalisiert pareto Verteilt sind. Andererseits wird dann die Pseudoinverse der Tailfunktion betrachtet vom Punkt  $x_p$ .
- Jetzt verwenden wir eine ähnliche Methode zu POT-Methode, die jedoch speziell für den Freshet Fall konstruiert wurde.

**Schätzer für die Wahrscheinlichkeit  $p_x = \bar{F}(x) = \mathbb{P}(X > x)$  eine hohe Schwelle  $x$  zu überschreiten**

Wir wissen bereits, dass für  $F$  im  $G_\gamma$ -Anziehungsbereich gilt

$$n\bar{F}(a_nt) \rightarrow t^{-\alpha}, \quad \text{mit } a_n = F^{\leftarrow}(1 - 1/n).$$

Ist  $k_n := \frac{n}{k}$  eine intermediäre Folge, so gilt die eben gezeigte Aussage auch für  $k_n$  :

$$\underbrace{\frac{n}{k} \bar{F}(F^{\leftarrow}\left(1 - \frac{k}{n}\right)t)}_{=:x} \rightarrow t^{-\alpha}.$$

$\Longleftrightarrow$

$$\underbrace{\bar{F}(F^{\leftarrow}\left(1 - \frac{k}{n}\right)t)}_{=:x} \rightarrow \frac{k}{n} \left( t \cdot \frac{F^{\leftarrow}\left(1 - \frac{k}{n}\right)}{F^{\leftarrow}\left(1 - \frac{k}{n}\right)} \right)^{-\alpha}.$$

$\Longleftrightarrow$

$$p_x = \bar{F}(x) \approx \frac{k}{n} \left( \frac{x}{F^{\leftarrow}\left(1 - \frac{k}{n}\right)} \right)^{-\alpha} \quad (*)$$

Einsetzen von konsistenten Schätzern liefert

$$\hat{p}_{x,k,n} := \frac{k}{n} \left( \frac{x}{X_{n-k:n}} \right)^{-1/\hat{\gamma}_{k,n}} \quad (**)$$

als Schätzer für die Überschreitungswahrscheinlichkeit einer hohen Schwelle  $x$ .

**Schätzer für ein hohes Quantil**  $x_p := F^{\leftarrow}(1 - p)$  (so dass  $\bar{F}(x) = \mathbb{P}(X > x) = p$  )

Ausgehend von (\*) erhalten wir (für hinreichend kleines  $p$  ) die Approximation

$$p = \bar{F}(x) \approx \frac{k}{n} \left( \frac{x_p}{F^{\leftarrow}\left(1 - \frac{k}{n}\right)} \right)^{-\alpha}.$$

Umstellen und Einsetzen von konsistenten Schätzern liefert

$$\hat{x}_{p,k,n} := \left( \frac{np}{k} \right)^{-\hat{\gamma}_{k,n}} X_{n-k:n}$$

als Schätzer für ein hohes Quantil (d.h.  $p$  klein).

**Schätzer für stochastische Größen, die nur von  $F$  eingeschränkt auf  $[u, \infty)$  für hohes  $u$  abhängen**

Bsp. Nettorisikoprämie. Sei  $X \sim F$  im  $G_\gamma$ -Anziehungsbereich mit  $\gamma < 1$ . Dann gilt

$$\mathbb{E}(X - u)_+ = \int_u^\infty 1 - F(x) dx.$$

Falls  $u$  hinreichend groß ist (wenigstens  $u > X_{n-k:n}$ ), so können wir  $\mathbb{E}(X - u)_+$  durch

$$\mathbb{E}(\widehat{X - u})_+ := \int_u^\infty \hat{p}_{x,k,n} dx \quad (***)$$



schätzen. Mit dem Schätzer  $(**)$  ergibt sich in der Situation  $\gamma > 0$

$$\begin{aligned}\mathbb{E}(\widehat{X - u})_+ &= \frac{k}{n} \int_u^\infty \left( \frac{x}{X_{n-k:n}} \right)^{-1/\hat{\gamma}_{k,n}} dx = \frac{k}{n} (X_{n-k:n})^{1/\hat{\gamma}_{k,n}} \left( \frac{u^{1-1/\hat{\gamma}_{k,n}}}{1-1/\hat{\gamma}_{k,n}} - \frac{0}{1-1/\hat{\gamma}_{k,n}} \right) \\ &= \frac{k}{n} (X_{n-k:n})^{1/\hat{\gamma}_{k,n}} \frac{u^{1-1/\hat{\gamma}_{k,n}}}{1-1/\hat{\gamma}_{k,n}}\end{aligned}$$

**Bemerkung 3.2.1.** Ist speziell  $\hat{\gamma}_{k,n}$  der Hill-Schätzer, dann ist der Schätzer in  $(***)$  unter geeigneten Annahmen (2. Ordnung) ebenso asymptotisch normal.

**Bemerkung 3.2.2.** Für die Anwendung dieser Verfahren gelten nach wie vor die Warnhinweise dieses Abschnitts!

## Kapitel 4

# Multivariate Extremwerttheorie (und -statistik)

Anstelle einer u.i.v. Folge  $X_1, \dots, X_n$  von Zufallsvariablen betrachten wir nun eine u.i.v. Folge  $X_1, \dots, X_n$  Zufallsvektoren mit Werten in  $\mathbb{R}^d$ .

Wir wollen das Extremwertverhalten dieser Vektoren analysieren und daraus Schätzverfahren zur Extrapolation in Risikobereiche gewinnen. Zunächst ist gar nicht klar, was wir unter einer „extremen“ Beobachtung verstehen wollen. Folgende Varianten wären sicherlich denkbar.

- Wenigstens eine Komponente ist groß.
- Eine Norm/lin. Funktional des Vektors ist groß. (Das heißt z.B. der Betrag wenigstens einer Komponente ist groß).
- Alle Komponenten sind groß.

Wir orientieren uns am folgenden Beispiel.

**Beispiel 4.0.1.** Es seien  $X_n = (X_n^{(1)}, \dots, X_n^{(d)})$ ,  $n = 1, 2, \dots$  die Pegelstände zum Zeitpunkt  $n$  an  $d$  verschiedenen Küstenorten, die wir mit  $1, \dots, d$  durchnummeriert haben. Dann erscheint es sinnvoll, das Grenzverhalten des Vektors der komponentenweise Maxima  $(M^{(1)}, \dots, M^{(d)})$  für  $n \rightarrow \infty$  zu untersuchen. Ein schwerwiegendes Ereignis in dem Zeitraum  $[0, n]$  ist bereits gegeben, wenn wenigstens eines dieser Maxima  $M_n^{(i)}$  die Höhe  $u^{(i)}$  des Deiches am Ort  $i$  überschreitet. („Ein Deichbruch an wenigstens einem Ort ist schon schwerwiegend genug.“)

Die Wahrscheinlichkeit für ein solches Ereignis ist dann

$$\begin{aligned}
\mathbb{P}\left(M_n^{(1)} > u^{(1)} \text{ oder } \dots \text{ oder } M_n^{(d)} > u^{(d)}\right) &= \mathbb{P}\left(\{M_n^{(1)} \leq u^{(1)}\}^C \cup \dots \cup \{M_n^{(d)} > u^{(d)}\}\right) \\
&= \mathbb{P}\left(\left(\{M_n^{(1)} \leq u^{(1)}\} \cap \{M_n^{(2)} > u^{(2)}\}^C \cap \dots \cap \{M_n^{(d)} > u^{(d)}\}^C\right)^C\right) \\
&= 1 - \mathbb{P}\left(\{M_n^{(1)} \leq u^{(1)}\} \cap \dots \cap \{M_n^{(d)} \leq u^{(d)}\}\right) \\
&= 1 - F^n\left(u^{(1)}, \dots, u^{(d)}\right),
\end{aligned}$$

falls  $F$  die Verteilungsfunktion von  $X_1 = (X_1^{(1)}, \dots, X_1^{(d)})$  ist. (Überprüfen!)

Weitere Situationen, die dem Muster folgen: "Der Ausfall eines Systems an einer einzigen Stelle bedeutet schon den Zusammenbruch des gesamten Systems."

- Ist zum Beispiel die Schadstoffkonzentration in der Luft an einer einzigen Stelle zu hoch, ist das erlaubte Level in der gesamten Stadt schon überschritten.
- Ist die Windstärke, die ein Wolkenkratzer aushalten muss, an einer einzigen Stelle zu hoch, kann dies schon katastrophale Folgen haben.

Es stellt sich heraus, dass die Theorie, die sich aus der Analyse der komponentenweise Maxima ergibt, sogar reichhaltig genug ist, dass sie auch auf andere Arten von extremen Beobachtungen übertragen werden kann.

**Definition 4.0.2.** Für die Randverteilung  $F_i(t) = \mathbb{P}(X_i \leq t) = \mathbb{P}(X_1 \in \mathbb{R}, \dots, X_i \leq t, \dots, X_d \in \mathbb{R})$  folgt aus der Stetigkeit von Maßen

$$F_i(t_i) = \lim_{t_k \rightarrow \infty, \forall k \neq i} F(t_1, \dots, t_d) =: F_X(+\infty, \dots, t_i, \dots, +\infty).$$

Mit dieser Formel ist klar, wie aus der gemeinsamen Verteilungsfunktion aller  $X_i$  die Verteilungsfunktion eines einzelnen  $X_i$  berechnet werden kann: Man schickt einfach in der gemeinsamen Verteilungsfunktion alle anderen Variablen  $t_k$  nach  $+\infty$ .

## Notation

Sei im Folgenden  $X_n = (X_n^{(1)}, \dots, X_n^{(d)})$ ,  $n \geq 1$  eine Folge von Zufallsvektoren mit Werten in  $\mathbb{R}^d$  mit gemeinsamer (multivariater) Verteilungsfunktion  $F$ . Mit  $\mathbf{x}$  bezeichnen wir stets einen  $d$ -dimensionalen Vektor und mit  $x^{(1)}, x^{(2)}, \dots, x^{(d)}$  seine Komponenten, so dass

$$F(\mathbf{x}) = F\left(x^{(1)}, \dots, x^{(d)}\right) = \mathbb{P}\left(X_1^{(1)} \leq x^{(1)}, \dots, X_1^{(d)} \leq x^{(d)}\right).$$

Die Randverteilungen seien  $F_i(x^{(i)}) = F(\infty, \infty, \dots, \infty, x^{(i)}, \infty, \dots, \infty) = \mathbb{P}(X_i \leq x^{(i)})$ ,  $i = 1, \dots, d$ .

Operationen mit Vektoren (zum Beispiel Maxima, Produkte, Summen) sind stets komponentenweise gemeint, wenn nicht anders spezifiziert. Für das Maximum schreiben wir " $\vee$ " und " $\bigvee$ " und für das Minimum " $\wedge$ " und " $\bigwedge$ ". Relationszeichen wie " $\leq$ " sind dabei ebenso immer komponentenweise zu verstehen, d. h.  $\mathbf{x} \leq \mathbf{y}$  bedeutet  $x^{(i)} \leq y^{(i)}$  für alle  $i \in \{1, \dots, d\}$ .

Man beachte:  $\mathbf{x} \not\leq \mathbf{y}$  bedeutet  $x^{(i)} > y^{(i)}$  für ein  $i \in \{1, \dots, d\}$  und somit im Allgemeinen nicht  $\mathbf{x} > \mathbf{y}$ . Schließlich schreiben wir  $d$ -dimensionale Rechtecke stets in der Form  $[\mathbf{a}, \mathbf{b}] := \{\mathbf{x} : \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$ . Entsprechend  $(\mathbf{a}, \mathbf{b}] := \{\mathbf{x} : \mathbf{a} < \mathbf{x} \leq \mathbf{b}\}$  etc.

**Definition 4.0.3.** Falls in der obigen Situation (komponentenweise) Normierungskonstanten  $a_n^{(i)} > 0$  und  $b_n^{(i)} \in \mathbb{R}$ ,  $i = 1, \dots, d$  existieren, so dass die Konvergenz

$$\mathbb{P}\left(\frac{M_n^{(i)} - b_n^{(i)}}{a_n^{(i)}} \leq x^{(i)}, i = 1, \dots, d\right) = F^n\left(a_n^{(1)}x^{(d)} + b_n^{(1)}, \dots, a_n^{(d)}x^{(d)} + b_n^{(d)}\right) \rightarrow G(\mathbf{x})$$

gilt, wobei  $G$  eine Verteilungsfunktion mit nicht-ausgearteten Randverteilungen  $G_i$ ,  $i = 1, \dots, d$  ist, sagen wir, dass  $G$  eine (multivariate) Extremwertverteilung ist und  $X_1$  (bzw. die Verteilung  $F$  von  $X_1$ ) in ihrem Max-Anziehungsbereich liegt. Denn wir betrachten ein Folge von u.i.v. Zufallsvektoren  $X_1, \dots, X_n$ .

### Randverteilungen

Man beachte, dass wir nicht nur annehmen, dass  $G$  nicht ausgeartet ist, sondern, dass jede Randverteilung von  $G$  nicht ausgeartet ist. Aus der Konvergenz aus der vorherigen Definition können wir unmittelbar auf die Konvergenz der Randverteilungen  $F_i$  von  $F$  schließen

$$\mathbb{P}\left(\frac{M_n^{(i)} - b_n^{(i)}}{a_n^{(i)}} \leq x^{(i)}\right) = F_i^n\left(a_n^{(i)}x^{(i)} + b_n^{(i)}\right) \rightarrow G_i\left(x^{(i)}\right).$$

Da  $G_i$  nicht ausgeartet ist, ist  $G_i$  notwendigerweise eine (univariate) Extremwertverteilung, also vom Typ  $\Phi_\alpha$ ,  $\Lambda$  oder  $\Psi_\alpha$ . Diese Randverteilungen kann man durch monotone und stetige Transformationen ineinander überführen.

**Bemerkung 4.0.4.** Man zeige, dass für eine Zufallsvariable  $X$  gilt

$$X \text{ hat Vtlgsfkt } \Phi_\alpha \Leftrightarrow \ln X^\alpha \text{ hat Vtlgsfkt } \Lambda \Leftrightarrow -X^{-1} \text{ hat Vtlgsfkt } \Psi_\alpha.$$

Desweiteren sind die entsprechenden Anziehungsbereiche in folgendem Sinne verträglich.

Die folgende Proposition sagt, dass man in der Praxis immer verlustfrei auf  $\phi_1$  (oder  $\Lambda$  oder  $\Psi$ ) die Rand-

verteilung normieren kann. Im folgenden wird nämlich das Pseudoinverse von der Freshet 1 benutzt, aber es kann genauso gut die Pseudoinverse von den anderen GEV verwendet werden. Dann müsste das  $U_i$  angepasst werden. Denn  $U_i$  ist asymptotisch ähnlich zu  $\Psi_i$

BILD

Die Idee des Satzes ist wie folgt: Wir haben zwei Verteilungen  $F_1, F_2$  und suchen eine monotone Transformation  $T$ , sodass für  $X \sim F_1$  gilt  $T(X) \sim F_2$ . Aufgrund von

$$\mathbb{P}(F_2^{\leftarrow}(F_1(X)) \leq t) = \mathbb{P}(F_1(X) \leq F_2(t)) = F_2(t)$$

dies gilt, da

$$\text{Sei } X \sim F \Rightarrow F(X) \sim \mathcal{U}([0, 1]) \Rightarrow G^{\leftarrow}(F(X)) \sim G, \text{ für eine beliebige Verteilung } G.$$

Wir können also ohne die Ordnung der Daten zu ändern, dessen Verteilung ändern, die dann gegen eine GEV mit Freshet 1 Rändern läuft.

Unsere monotone Transformation sollte also  $T(x) := F_2^{\leftarrow}(F_1(X))$  sein.

**Lemma 4.0.5.** Proposition

Seien  $F$  und  $G$  d-variate Verteilungsfunktionen und  $G$  habe stetige Ränder. Sei  $U_i = (1/(1 - F_i))^{\leftarrow}, i = 1, \dots, d$  und

$$F_*(\mathbf{x}) := F\left(U_1\left(x^{(1)}\right), \dots, U_d\left(x^{(d)}\right)\right).$$

Sei  $\Psi_i = (1/(-\log G_i))^{\leftarrow}, i = 1, \dots, d$  und

$$G_*(\mathbf{x}) := G\left(\Psi_1\left(x^{(1)}\right), \dots, \Psi_d\left(x^{(d)}\right)\right).$$

Dann gilt

- (a)  $G_*$  hat  $\Phi_1$ -Ränder.
- (b)  $G$  ist eine multivariate Extremwertverteilung genau dann, wenn es  $G_*$  ist.
- (c) Falls  $F$  im  $G$ -Anziehungsbereich liegt, so liegt  $F_*$  im  $G_*$ -Anziehungsbereich mit  $F_*^n(n\mathbf{x}) \rightarrow G_*(\mathbf{x})$ .
- (d) Falls  $F_*^n(n\mathbf{x}) \rightarrow G_*(\mathbf{x})$  und zusätzlich  $F_i$  im  $G_i$ -Anziehungsbereich liegt, dann ist  $F$  im  $G$ -Anziehungsbereich.

*Beweis.* (a) Das haben wir in der Vorüberlegung gezeigt. Betrachte die Randverteilung von  $G_*$ .

(b)

folgt aus c) und d)

### c) Vorüberlegung

Vorüberlegung: Falls  $F_i$  im  $G_i$ -Anziehungsbereich liegt, folgt für geeignete Normierungskonstanten, dass

$$n \left( 1 - F_i \left( a_n^{(i)} x + b_n^{(i)} \right) \right) \rightarrow -\log G_i(x)$$

(wegen dem Korollar nach dem Lemma der Poisson-Approximation).

Wenn wir durch den Ausdruck auf der linken Seite teilen, konvergiert dies auch gegen 1 durch die Rechte Seite:

$$\frac{\frac{1}{(1-F_i)} \left( a_n^{(i)} x + b_n^{(i)} \right)}{n} \rightarrow \frac{1}{-\log G_i}(x).$$

Nun wollen wir die Umkehrfunktion/Pseudoinverse von der linken Seite bestimmen:

$$\begin{aligned} & \frac{\frac{1}{(1-F_i)} \left( a_n^{(i)} x + b_n^{(i)} \right)}{n} =: y \\ \iff & \frac{1}{(1-F_i)} \left( a_n^{(i)} x + b_n^{(i)} \right) = ny \\ \iff & a_n^{(i)} x + b_n^{(i)} = \left( \frac{1}{1-F_i} \right)^{\leftarrow} (ny) \\ \iff & a_n^{(i)} x + b_n^{(i)} = U_i(ny) \\ \iff & x = \frac{U_i(ny) - b_n^{(i)}}{a_n^{(i)}} \end{aligned}$$

Da wenn  $F_n$  schwach konvergiert gegen  $F$ , dann konvergiert auch die Pseudoinverse  $F_n^{\leftarrow}$  schwach  $F^{\leftarrow}$ . Wir betrachten nun die rechte Seite:

$$\frac{1}{-\log G_i}(x) =: y \iff x = \left( \frac{1}{-\log G_i} \right)^{\leftarrow} (y) = \Psi(y)$$

$$\frac{U_i(ny) - b_n^{(i)}}{a_n^{(i)}} \rightarrow \Psi_i(y), \quad y > 0$$

c)

Falls nun zusätzlich  $F$  im  $G$ -Anziehungsbereich liegt (mit notwendigerweise diesen Normierungskonstanten

als mögliche Wahl), d.h. es gilt

$$\exists a_n^{(i)} > 0 \text{ und } b_n^{(i)} \in \mathbb{R} : \quad \frac{M_n^{(i)} - b_n^{(i)}}{a_n^{(i)}} \xrightarrow{(d)} Y \sim G.$$

Dann gilt:

$$\begin{aligned} F_*^n(n\mathbf{x}) &= \mathbb{P}(M_n^{(i)} \leq U_i n x^{(i)}, i = 1, \dots, d) = \mathbb{P}\left(\underbrace{\frac{M_n^{(i)} - b_n^{(i)}}{a_n^{(i)}}}_{\rightarrow Y^{(i)}} \leq \underbrace{\frac{U_i(n x^{(i)}) - b_n^{(i)}}{a_n^{(i)}}}_{\rightarrow \Psi_i(x^{(i)})}, i = 1, \dots, d\right) \\ &\rightarrow \mathbb{P}(Y^{(i)} \leq \Psi_i(x^{(i)}), i = 1, \dots, d) = G_*(\mathbf{x}). \end{aligned}$$

d)

Gelte nun  $F_*^n(n\mathbf{x}) \rightarrow G_*(\mathbf{x})$ , und sei  $F_i$  im Max-Anziehungsbereich von  $G_i$  so folgt

$$\begin{aligned} &F^n(a_n^{(1)}x^{(1)} + b_n^{(1)}, \dots, a_n^{(d)}x^{(d)} + b_n^{(d)}) \\ &= F^n\left(\left(\frac{1}{(1-F_i)}\right)^\leftarrow \circ \left(\frac{1}{(1-F_i)}\right)\left(a_n^{(1)}x^{(1)} + b_n^{(1)}\right), \dots, \left(\frac{1}{(1-F_i)}\right)^\leftarrow \circ \left(\frac{1}{(1-F_i)}\right)\left(a_n^{(d)}x^{(d)} + b_n^{(d)}\right)\right) \\ &= F_*^n\left(\frac{1}{(1-F_i)}\left(a_n^{(1)}x^{(1)} + b_n^{(1)}\right), \dots, \frac{1}{(1-F_i)}\left(a_n^{(d)}x^{(d)} + b_n^{(d)}\right)\right) \\ &= F_*^n\left(n \underbrace{\frac{\frac{1}{(1-F_i)}\left(a_n^{(1)}x^{(1)} + b_n^{(1)}\right)}{n}}_{\rightarrow -\frac{1}{\log G_1}(x^{(1)})}, \dots, n \underbrace{\frac{\frac{1}{(1-F_i)}\left(a_n^{(d)}x^{(d)} + b_n^{(d)}\right)}{n}}_{\rightarrow -\frac{1}{\log G_d}(x^{(d)})}\right) \\ &\rightarrow G_*\left(\Psi_1^\leftarrow(x^{(1)}), \dots, \Psi_d^\leftarrow(x^{(d)})\right) \\ &= G\left((\Psi_1 \circ \Psi_1^\leftarrow)(x^{(1)}), \dots, (\Psi_d \circ \Psi_d^\leftarrow)(x^{(d)})\right) = G(\mathbf{x}) \end{aligned}$$

□

Typischerweise standardisiert man die Randverteilungen von  $G$ , wenn man nur noch an der Charakterisierung der Abhängigkeitsstruktur interessiert ist. Wir werden im Folgenden meist wie in obiger Proposition Extremwertverteilungen mit  $\Phi_1$ -Rändern (StandardFréchet-Ränder) betrachten. Frage: Wie kann der Raum aller Extremwertverteilungen  $G_*$  mit  $\Phi_1$ -Rändern (vollständig) charakterisiert werden?

### 4.0.1 Max-Stabilität

**Definition 4.0.6.** Eine  $d$ -variate Verteilungsfunktion  $G$  heißt max-stabil, falls für jedes  $n \in \mathbb{N}$  Normierungskonstanten  $a_n^{(i)} > 0$  und  $b_n^{(i)} \in \mathbb{R}, i = 1, \dots, d$  existieren, so dass

$$G^n \left( a_n^{(1)} x^{(1)} + b_n^{(1)}, \dots, a_n^{(d)} x^{(d)} + b_n^{(d)} \right) = G(x).$$

Ähnlich wie im univariaten Fall gilt:

**Lemma 4.0.7.** Die Klasse der multivariaten Extremwertverteilungen stimmt überein mit der Klasse der max-stabilen Verteilungen mit nicht ausgearteten Rändern.

*Beweis.* Nicht behandelt

Da die Randverteilungen die Normierungskonstanten bereits eindeutig bestimmen, gilt für die Klasse der Extremwertverteilungen  $G_*$  mit  $\Phi_1$ -Rändern notwendigerweise

$$G_*^n(n\mathbf{x}) = G_*(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, n \in \mathbb{N}$$

denn  $\forall i = 1, \dots, d: G_{*i} \sim \Phi_1$ , damit gilt

$$G_{*i}^n(nx_i) = (e^{-(nx_i)^{-1}})^n = e^{-\frac{n}{nx_i}} = e^{-x_i^{-1}} = G_{*i}(x_i) \sim \Phi_1$$

□

Desweiteren ist jede max-stabile Verteilung insbesondere max- $\infty$ -oft teilbar. (Warum?)

**Definition 4.0.8.** Eine (multivariate) Verteilungsfunktion  $G$  heißt max -  $\infty$ -oft teilbar, falls es für jedes  $n \in \mathbb{N}$  eine Verteilungsfunktion  $G_n$  gibt, so dass  $G = G_n^n$ .

Bevor wir die Klasse der multivariaten Extremwertverteilungen weiter untersuchen können, benötigen wir noch zwei mathematische Schlüsselkonzepte. Wir beschäftigen uns zuerst mit Poisson-Punktprozessen und anschließend mit einem Strukturresultat für max- $\infty$ -oft teilbare Verteilungen.

## 4.1 Poisson-Punktprozesse

Wir formalisieren im Folgenden den Begriff eines "zufälligen Punktmusters im Raum". Aus mathematischer Sicht hat es sich als vorteilhaft herausgestellt einen solchen "Punktprozess" als zufälliges diskretes Maß zu



verstehen. Wir tragen zunächst die nötigen Definitionen und einige wichtige Konzepte zusammen.

**Motivation:**

### 4.1.1 Hintergrund: Punktprozesse

Sei im Folgenden  $E$  stets ein lokal kompakter, separabler metrischer Raum (oder etwas allgemeiner ein lokal kompakter Hausdorffraum mit abzählbarer Basis für die Topologie).  $E$  kann zum Beispiel eine offene oder abgeschlossene Teilmenge von  $\mathbb{R}^d$  oder  $(0, \infty]^d$  oder  $[-\infty, \infty]^d$  sein.

Polnische Räume sind eine Verallgemeinerung vom  $\mathbb{R}^d$  und garantiert die Def. eines bedingten Erwartungswerts.

**Definition 4.1.1.** (Poisson'scher Grenzwertsatz)

Sei  $(p_n)_{n \in \mathbb{N}} \in (0, 1)^{\mathbb{N}}$  eine Folge von Parametern der Binomialverteilung  $Bin(n, p_n)$ . Falls  $np_n \rightarrow \lambda, n \rightarrow \infty$ , dann gilt

$$Bin(n, p_n) \xrightarrow{d} Poi(\lambda)$$

**Beispiel 4.1.2.** Ein Sanhaufen ist ein polnischer Raum, da jedes Sandkorn eine kompakte Menge ist und endliche Vereinigung kompakter Mengen wieder kompakt ist. Damit ist es eine zufällige kompakte Menge, wenn ein Kind drin spielt.

**Definition 4.1.3.** Sei  $E$  unser Zustandsraum. Dieser kann z.B.  $E = \mathbb{R}^d$  sein. Man muss ein paar VSS an  $E$  fordern, die wir überspringen.  $\mathbb{B}(E)$  sei die Borel- $\sigma$  Algebra von  $E$ . Sei  $(E, \mathbb{B}(E))$  ein Maßraum. Ein Maß  $\mu$  heißt lokal endlich bzw. Radon-Maß, falls

$$\forall \text{ kompakten } B \subset E : \mu(B) < \infty$$

**Beispiel 4.1.4.** Das Maß  $\mu := \sum_{x \in \mathbb{Q}} \delta_x$  ist kein Radon-Maß auf  $\mathbb{R}$ , da  $\mathbb{Q}$  dicht in  $\mathbb{R}$  liegt. Damit gilt dann

$$\mu([0, 1]) = \infty$$

**Definition 4.1.5.** Ein Radon-Maß  $\mu$  heißt Zählmaß, falls

$$\forall \text{ kompakten } A \subset E : \mu(A) \in \mathbb{N}_0$$

Man kann zeigen, dass

$$M_p(E) := \{\mu : \mu \text{ lokal endliches Zählmaß auf } E\}$$

ein (bezüglich der vagen Topologie) abgeschlossener Unterraum von  $M_+(E)$  ist (also wieder ein polnischer Raum). Dabei beschreibt  $p$  ein Punkt im Raum???

**Satz 4.1.6.** Sei  $E$  ein lokal kompakter separabler metrischer Raum. Dann lässt sich jedes Zählmaß  $\mu$  auf  $E$  schreiben als

$$\mu = \sum_{i=1}^n \delta_{x_i}$$

darstellen. Dabei ist  $n \in \mathbb{N} \cup \{\infty\}$  und  $\forall i \leq n: x_i \in E$ .

Diesen Satz benötigen wir weil wir später viele Zählmaße in dieser Form schreiben werden.

**Definition 4.1.7.** Ein Punktprozess  $\Phi$  (mit Zustandsraum  $E$ ) ist eine Zufallsvariable mit Werten in  $M_p(E)$ , das heißt eine messbare Abbildung

$$\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (M_p(E), \mathcal{B}(M_p(E))),$$

wobei  $(\Omega, \mathcal{A}, \mathbb{P})$  ein W'raum ist. Mit anderen Worten, ein Punktprozess ist ein zufälliges lokal endliches Zählmaß. D.h. ein Zählmaß ist eine Realisierung vom Punktprozess, d.h.

$$\forall \omega \in \Omega : \Phi(\omega) =: \mu_\omega \in M_p(E)$$

D.h. es gilt

$$\forall \text{ kompakten } B \subset E : \mu_\omega(B) < \infty$$

**Bemerkung 4.1.8.** Wenn wir eine Teilmenge  $B \subset E$  betrachten bezüglich eines Punktprozesses, schreiben wir für die Anzahl der Punkte in  $B$

$$\Phi(\cdot)(B) =: \Phi(B)$$

D.h. die Anzahl der Punkte in  $B$  ist eine Zufallsvariable die der Punktprozess ist.

**Beispiel 4.1.9.** Mikroskopische Bilder von Mittelpunkten von Körnern:

??? BILD

### 4.1.2 Poisson-Punktprozesse

**Definition 4.1.10.** Sei  $\Lambda$  ein Radonmaß auf  $E$ . Ein Punktprozess  $\Phi$  auf dem Zustandsraum  $E$  heißt Poisson-Punktprozess mit Intensitätsmaß  $\Lambda$ , falls Folgendes gilt:

- (i) Für jedes  $B \in \mathcal{B}(E)$  ist  $\Phi(B) \sim \text{Poi}(\Lambda(B))$ .
- (ii) Für jedes  $k \geq 1$  und  $B_1, \dots, B_k \in \mathcal{B}(E)$  paarweise disjunkt sind  $\Phi(B_1), \dots, \Phi(B_k)$  stochastisch unabhängig.

Wir schreiben auch kurz  $\Phi \sim \text{PPP}(\Lambda)$ .

**Beispiel 4.1.11.** (Kästche mit Lebesguemaß und Punkten??)

BILD

**Beispiel 4.1.12.** • Wenn  $\Lambda(B) = 0$  dann gilt:  $\phi(B) \equiv 0$

- Wenn  $\Lambda(B) = \infty$  dann gilt:  $\phi(B) \equiv \infty$  (z.B. wenn  $B = E = \mathbb{R}^d$  und  $\Lambda$  das Lebesguemaß

**Beispiel 4.1.13.** Gegenbeispiele, d.h. Punktprozesse die keine Poisson-Punktprozesse sind:

- $\phi(\{0\}) \stackrel{???}{=} \phi(\mathbb{R}^d) = 1$  ist kein PPP, da ein fester Punkt nicht Poisson verteilt ist.

Es gilt  $\phi \equiv \mathbf{1}_{\{0\}}$  ist aber unabhängig.

BILD

- Nun ein Beispiel zu einem Punktprozess, der Poisson verteilt ist aber nicht unabhängig ist.

**Bemerkung 4.1.14.** Simulation eines Poisson-Punktprozesses:

- Ziehe die Anzahl der Punkte in der Menge als  $\sim \text{Poi}(C|A|)$
- Streue gleichverteilt die Punkte unabhängig und in alle koordinaten.

### 4.1.3 Eigenschaften eines Poisson-Punktprozesses

1. Teilgebiete eines Poisson-Punktprozesses (PPP) sind wieder PPP.

BILD

Wieso? Die Punkte innerhalb der Teilgebiete sind wieder unabhängig voneinander und Poisson verteilt.

2. Superpositionssatz:

Seien  $\forall k = 1, 2, \dots$  die  $\Phi_k \sim \text{PPP}(\Lambda_k)$  unabhängige PPP und wir def.  $\Lambda := \sum_{k=1}^{\infty} \Lambda_k$  (dies ist ein radon-Maß).

Dann ist  $\Phi := \sum_{k=1}^{\infty} \Phi_k \sim PPP(\Lambda)$

BILD

### 3. Abbildungssatz:

Sei  $T : E_1 \rightarrow E_2$  eine messbare Abbildung und  $\Phi \sim PPP(\Lambda)$  auf  $E_1$ , dann ist  $T(\Phi)$  ein PPP auf  $E_2$ , falls  $T(\Phi)$  ein Radonmaß ist. (Dies wäre z.B. der Fall wenn für die Abbildung  $T$  die Urbilder kompakter Mengen wieder kompakt sind). D.h. wenn die Bedingung gilt, ist  $T(\Phi) \sim PPP(T(\Lambda))$  für das zufällige Bildmaß  $T(\Phi)$  auf  $E_2$ .

BILD

### 4. Verdünnung:

Sei  $\Phi$  ein PPP auf einem polnischen Raum  $E$  und sei  $\forall x \in E: Z_x$  stochastisch unabhängig Bernoulli verteilt. (D.h. wir machen für jeden Punkt im Raum eine Bernoulli Ziehung).

Sei  $\Phi(\omega) = \sum_{i \in I} \delta_{x_i}(\omega)$ ,  $x_i \in E$  und  $i \in I$  (Indexmenge über  $E$ ). Da das Diracmaß messbar ist, ist die abzählbare Summe darüber auch messbar. Damit können wir den Abbildungssatz anwenden:

Es sind  $\sum_{i \in I} \delta_{x_i}(Z_{x_i})$  und  $\sum_{i \in I} \delta_{x_i}(1 - Z_{x_i})$  PPP und stochastisch unabhängig.

MITTLERE ANZAHL AN PUNKTEN???

### 5. Unabhängige Markierung:

VERSTEHE ICH NICHT

### 6. PPP-Darstellung einer univariaten reellwertigen Variablen:

Sei  $Y_1, Y_2, \dots \sim \Lambda^* := \frac{\Lambda}{\Lambda(E)}$ , dann ist mit Blatt 9

$$\Phi = \sum_{i=1}^{\infty} \delta_{Y_i} \sim PPP(\Lambda).$$

Wir definieren,  $X := \max \text{supp}(\Phi)$

BILD

Dann gilt:

$$\mathbb{P}(X \leq x) = \mathbb{P}(\Phi((x, \infty)) = 0) \stackrel{Poi}{=} e^{-\Lambda((x, \infty))}.$$

D.h. die Kunst liegt darin, das Maß  $\Lambda$  so zu wählen, dass am Ende auf der rechten Seite eine Verteilung herauskommt.

Wir wählen  $\Lambda((x, \infty)) := \frac{1}{x}$ , dann gilt

$$\mathbb{P}(X \leq x) = \mathbb{P}(\Phi((x, \infty)) = 0) \stackrel{Poi}{=} e^{-\Lambda((x, \infty))} = e^{-x^{-1}}$$

Damit haben wir also eine Freshet Verteilungsfunktion.

## 4.2 Räumliche Statistik

### 4.2.1 Brownsche Bewegung

**Definition 4.2.1.** EINDEUTIGKEITSSATZ KOLMOGOROV

**Definition 4.2.2.** (Brownsche Bewegung klassisch)

Ein reellwertiger stochastischer Prozess  $(B_t)_{t \geq 0}$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  heißt (standard) Brownsche Bewegung, falls

- (i)  $B(0) = 0$   $\mathbb{P}$ -fast sicher.
- (ii)  $B$  hat stationäre und unabhängige Inkremente, d.h.
  - die Verteilung von  $B(t+h) - B(t)$  hängt von  $h$  aber nicht von  $t$ , ab
  - die Inkremente  $B(t_1) - B(t_0), \dots, B(t_n) - B(t_{n-1})$  sind unabhängige Zufallsvariablen für alle  $0 \leq t_0 \leq \dots \leq t_n$ .
- (iii) Der Pfad  $t \mapsto B(t)$  sind  $\mathbb{P}$ -fast sicher stetig.
- (iv)  $B(t) - B(s) \sim \mathcal{N}(0, |t - s|)$  für alle  $t, s \in I$  mit  $s < t$  beliebig.

**Definition 4.2.3.** Ein (reellwertiger) Gauß-Prozess ist ein stochastischer Prozess  $(X_t)_{t \in T}$  auf einer beliebigen Indexmenge  $T$ , wenn seine endlichdimensionalen Verteilungen mehrdimensionale Normalverteilungen (auch Gauß-Verteilungen genannt) sind.

Es soll also für alle  $n \in \mathbb{N}$  und alle Indizes  $t_1, t_2, \dots, t_n \in T$  die multivariate Verteilung von  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  durch eine  $n$ -dimensionale Normalverteilung gegeben sein.

Eine ein- oder mehrdimensionalen Normalverteilung ist durch die Angabe des Erwartungswertvektors und der Kovarianzmatrix vollständig bestimmt. Die Wahrscheinlichkeitsverteilung eines Gauß-Prozesses liegt durch die Angabe aller endlichdimensionalen Verteilungen fest. Daher ist ein Gaußprozess durch eine Erwartungswertfunktion

$$\mu(t) := \mathbb{E}[X_t], \quad t \in T$$

und eine Kovarianzfunktion

$$\gamma(s, t) := \text{Cov}[X_s, X_t] := \mathbb{E}[(X_s - \mu(s))(X_t - \mu(t))], \quad s, t \in T$$

eindeutig bestimmt. Durch das Funktionenpaar  $(\mu : T \rightarrow \mathbb{R}, \gamma : T \times T \rightarrow \mathbb{R})$  liegen die Parameter aller mehrdimensionalen Normalverteilungen fest und damit liegt die Wahrscheinlichkeitsverteilung des Gauß-Prozesses fest.

**Bemerkung 4.2.4.** In der Geostatistik wird die Brownsche Bewegung anders eingeführt mit einer zu der vorherigen Definition äquivalent ist.

**Definition 4.2.5.** (Brownsche Bewegung Geostatistik)

Ein reelwertiger stochastischer Prozess  $(B_t)_{t \geq 0}$  auf einem Wahrscheinlichkeitsraum  $(\Omega, \mathcal{A}, \mathbb{P})$  heißt (standard) Brownsche Bewegung, falls

- es ist ein Gausscher Prozess mit Erwartungswert 0
- $\mathbb{V}[B(t) - B(s)] = |t - s|^\alpha$  mit  $\alpha = 1$

**Bemerkung 4.2.6.**

- Falls  $\alpha \in (0, 2]$  dann nennen wir den obigen Prozess: fraktionale Brownsche Bewegung.
- Die beiden Punkte in der zweiten Definition der Brownsche Bewegung sind äquivalent zu ii) und iv) in der ersten Definition.  
Dies gilt wegen den (existenz) und Eindeutigkeitssatz von Kolmogorov.
- In der Geostatistik lässt man i) und iii) weg.

**Definition 4.2.7.** In der Geostatistik definiert man

- die sogenannte Kovarianzfunktion als

$$c(t, s) := \text{Cov}(Z(t), Z(s))$$

- das zentrierte Semi-Variogramm als

$$\gamma(t, s) := \frac{1}{2} \mathbb{V}[Z(t) - Z(s)]$$

Dabei ist dann  $Z$  ein stochastischer Prozess mit Kovarianzfunktion  $c$

**Bemerkung 4.2.8.** Man kann die Indizes  $t, s$  als Zeitpunkte oder koordinaten in einem Zustandsraum interpretieren.

**Definition 4.2.9.** Ein stochastische Prozess heißt stationär, falls seine Verteilung über die Zeit gleich bleibt.

Er heißt isotrop, falls die Kovarianzfunktion symmetrisch ist, d.h.

$$c(s, t) = c(t, s)$$

**Bemerkung 4.2.10.** Es gilt nun mit der obigen Definition

$$\begin{aligned}\gamma(t, s) &:= \frac{1}{2} \mathbb{V}[Z(t) - Z(s)] \\ &\stackrel{\text{Bienyame}}{=} \frac{1}{2} (\mathbb{V}[Z(t)] + \mathbb{V}[Z(s)] - 2 \cdot \text{Cov}(Z(t), Z(s))) \\ &= \frac{1}{2} (c(t, t) + c(s, s) - 2 \cdot c(t, s))\end{aligned}$$

Wenn der stochastische Prozess  $Z$  stationär ist, und die Kovarianzfunktion  $c(t, s) = c_0(\|t - s\|)$ , dann ist er auch isotrop, denn

$$c(t, s) = c_0(\|t - s\|) \stackrel{\text{Norm symm.}}{=} c_0(\|s - t\|) = c(s, t)$$

Mit dieser neuen Schreibweise erhalten wir für die obige Gleichung:

$$\gamma(t, s) = c_0(0) - c_0(\|t - s\|)$$

Wenn wir nun definieren  $h := \|t - s\|$ , dann können wir schreiben

$$\gamma_0(h) = c_0(0) - c_0(h)$$

BILD

**Bemerkung 4.2.11.** Hier formulieren wir die Idee von Variogrammen. In der Geostatistik ist man beispielsweise Interessiert an der Korrelation zweier Punkte innerhalb eines Raumes. Dort stellt man dann fest, dass Punkte, die in der Nähe voneinander liegen eine höhere Korrelation aufweisen. (Z.b. wenn wir Höhe eines Messpunktes zum Meerespiegel betrachten). Dann geben Punkte, die in der Nähe voneinander liegen, weniger Informationen über den Raum, als zwei Punkte, die weiter weg voneinander liegen. Deswegen ist der standard Mittelwertschätzer nicht geeignet. Man muss Punkte weniger gewichten, die Nah beieinander sind.

BILD

Wenn wir nun den Abstand eines beliebigen Vektors  $y$  zu einem festen Punkt  $x$  betrachten, dann beschreibt

$h = \|x - y\|$  dessen Abstand. Wir wollen nun eine Kovarianzfunktion in Abhängigkeit von dessen Abstands machen. Dies muss nicht unbedingt ein Linearer Zusammenhang sein, sondern kann z.B.  $e^{-h^2}$  oder  $e^{-x}$ .

BILD

Wo die Funktion den Wert 1 annimmt ist offensichtlich der Abstand 0, da  $x=y$  und dessen Korrelation ist 1, da sie am gleichen Punkt sind.

Dann ist das Variogramm von diesen beiden Kovarianzfunktionen wie folgt

BILD

D.h. weil bei der von uns definierten Kovarianzfunktionen gilt  $c_0(0) = e^0 = 1$  erhalten wir folgendes Variogramm

$$\gamma_1(h) = 1 - e^{-h^2} \quad \text{und} \quad \gamma_2(h) = 1 - e^{-h}$$

**Bemerkung 4.2.12.** In dieser Bemerkung diskutieren wir, wieso wir Variogramme verwenden.

Was wir ja bisher gemacht haben für die Kovarianzfunktion, ist sie geklappt haben und einige Dinge definiert haben. Also bisher noch keinen Vorteil erhalten.

- Für den Schätzer der Kovarianz braucht man einen Schätzer für den Erwartungswert. Dieser ist aber in der räumlichen Statistik nicht so einfach wie in Stochastik 2, denn Z.B.

BILD

Um den Erwartungswert gut zu Schätzen wollen wir den BLUE ausrechnen, d.h. wir wollen das der lineare Schätzer das Risiko zum wahren EW minimiert. Dazu brauchen wir nach dem Steinerschen Satz die Kovarianzstruktur, denn für das Risiko gilt

$$R(\theta, \hat{\theta}) = \mathbb{V}_{\theta}[\hat{\theta}(X)] + \left( \text{Bias}(\hat{\theta}(\theta)) \right)^2$$

D.h. unser Problem ist: Wir wollen die Cov schätzen, brauchen dafür einen Schätzer für den EW, dieser braucht aber die Cov Struktur. (Katze in schwanz gebissen).

Dieser Problem umgeht der Ansatz mit Variogrammen.

- Irgendwas mit  $C_0$  und  $\gamma_0$  schätzen???
- Das Variogramm ist allgemeiner als ...

Sei z.B.  $Z(\cdot)$  stationärer isotropischer Gaußprozess und  $Y \sim \text{Cauchy}$ , dann hat  $X(t) := Z(t) + Y$  ein  $\gamma_0$  aber kein  $c_0$  (da ...???)



- Wir hatten ja  $\gamma_0(h) = c_0(0) - c_0(h)$  wenn man dies umstellt erhält man

$$c_0(h) = c_0(0) - \gamma_0(h)$$

Hier könnte man eine Taylorentwicklung mit  $x=0$  sehen, d.h.

$$c_0(h) = c_0(0) - \dots + o(\dots)$$

Man erhält aus dem himmel fallend (wahrscheinlich polynom vom Grad  $\alpha$ ) mit  $c \in \mathbb{R}$

$$c_0(h) = c_0(0) - c|h|^\alpha + o(|h|^\alpha)$$

Wir betrachten nun die Kovarianzfunktion für kleine  $h$ , d.h. bei unserer Taylorentwicklung fällt der Restterm weg.

Also erhalten wir  $c_0(h) = c_0(0) - c|h|^\alpha$

mit  $h \ll 1$ .  $\gamma(0, h) = c|h|^\alpha$ , was genau das Variogramm von der Fractionellen Brownschen Bewegung ist. Damit haben beide die gleiche Kovarianzfunktion. BILD

Wenn man also nur kleine Abstände betrachtet, dann ist der stochastische Prozess dahinter eine Fraktionelle BB.

**Bemerkung 4.2.13.** Wir wollen hier den letzten Teil vorherigen Bemerkung hier genauer für Extremwerte betrachten. Angenommen wir haben  $n$  viele Gaußprozesse.

BILD

Wenn wir nun den Maximumsprozess nehmen, wissen wir, dass dieser Punktweise gegen Gumbel geht, falls wir  $n$  gegen unendlich schicken. Außerdem sehen wir, dass der entstehende Maximumsprozess sehr hügelig wird und gegen random noise geht (die Werte sind unabhängig verteilt).

BILD

Wenn wir diesen Prozess reskalieren, d.h. zoomen

BILD

sehen wir lokal Fraktionelle Brownsche Bewegung. D.h. beim zoomen betrachten wir nur kleine Abstände und wegen oben wissen wir, dass dort fraktionelle BB ist auf einem Pfad mit deterministischen Trend.

Dieser Limesprozess ist der Brown-Resnick-Prozess.

BILD

## 4.3 Simulation von PPP

### 4.3.1 Methode 1: (Acceptance Rejection)

Angenommen wir wollen  $\Phi \sim PPP(\Lambda)$  auf  $[0, \infty)$  simulieren, wobei die Dichte des Intensitätsmaß gegeben ist durch  $\Lambda'(x) = \frac{1}{x^2}$ . Dann liegen die Punkte auf  $[0, \infty)$  wie folgt

BILD

Bis einer hohen oberen Schranke  $x_0$  kann man ohne allzu hohe Rechenkosten, die Dichte des Intensitätsmaß via Acceptance-Rejection simulieren. Deswegen wählen wir die obere Schranke  $x_0$  so groß, sodass die Masse oberhalb dieser sehr klein ist, d.h.  $\lambda((-\infty, x_0))$  fast 1. D.h. wir ignorieren den Tail komplett.

### 4.3.2 Methode 2:()

Hier ist die Idee, dass wir einen PPP  $\Phi$  auf einer kleinen Menge simulieren und dann mittels dem Abbildungssatz auf einer größeren Menge abbilden, d.h. dort ist dann  $T(\Phi)$  ein PPP, falls  $T$  messbar ist.

Ein kleines Beispiel:

Sei  $\Phi \sim PPP(c\lambda)$  mit  $\lambda$  das Lebesgue-Maß und  $c \in \mathbb{R}$  ein PPP auf  $[0, x_0)$  mit  $x_0 \in \mathbb{R}_+$ .

BILD

Auf dem kleinen Intervall können wir einfach gleichverteilt simulieren. Unsere messbare Abbildung die den PPP auf  $[0, \infty)$  Abbildet ist  $f(x) := \frac{1}{x}$ , denn stetig. Wir überprüfen nun, ob das Bild  $f(\Phi) = [0, \infty)$  ist.

Wir wollen nun zeigen, dass die  $\forall u \in (0, \infty)$  die Wahrscheinlichkeit, dass das  $\max_{i=1, \dots, n} \frac{1}{x_i}$  positiv ist bzw. dass der obere Endpunkt  $x^* = \infty$ . Wir werden zeigen, dass es Freshet verteilt ist und damit ist der obere Endpunkt unendlich.

$$\begin{aligned} \mathbb{P} \left( \max_{i=1, \dots, n} \frac{1}{x_i} \leq u \right) &= \mathbb{P} \left( \min_{i=1, \dots, n} \frac{1}{x_i} \geq \frac{1}{u} \right) \\ &= \mathbb{P} \left( \Phi \left( \left[0, \frac{1}{u}\right] \right) = 0 \right) \\ &= e^{-c\lambda([0, \frac{1}{u}])} \\ &= e^{-\frac{c}{u}} \sim \text{Freshet} \end{aligned}$$

BILD

Damit ist  $\frac{1}{x}$  die richtige Abbildung. ...

## 4.4 Spektraldarstellung...

In diesem Abschnitt definieren wir einen Vektor mithilfe eines PPP und zeigen, dass dieser Max stabil ist, berechnen seine multivariate Verteilung und zeigen dass die Randverteilungen Freshet 1 Verteilt sind.

Sei  $\Phi \sim PPP(c \cdot du)$  auf  $[0, \infty]$ . Dabei ist  $du$  die Ableitung einer absolutstetigen Verteilung (Lebesgue Maß??).

**Definition 4.4.1.** Sei  $\Phi \sim PPP(\Lambda)$ , dann nennen wir  $\Phi_c \sim PPP(c\Lambda)$  den geschlürften PPP, falls  $c \in (0, 1)$

**Bemerkung 4.4.2.** Wir definieren

$$x := \min_{y \in \text{supp}(\Phi)} |y| \stackrel{\text{PPP auf } [0, \infty]}{=} \min_{y \in \text{supp}(\Phi)} y$$

$x$  ist also dann der kleinste Punkt bei einer Realisierung bei einem PPP.

Dann ist für  $X \sim \Phi$

$$\mathbb{P}(X \geq x) = \mathbb{P}(\Phi([0, x] = 0)) = e^{-cx}$$

BILD

**Bemerkung 4.4.3.** Sei  $\Phi_c = \sum_{k=1}^{\infty} \delta_{(u_k, e_k)} \sim PPP(cdu \times \Lambda(de))$  auf  $[0, \infty) \times E$  und die Funktion  $f : E \rightarrow [0, \infty)^d$  komponentenweise integrierbar d.h.

$$\forall k \in \mathbb{N} : \int_E f_k(e) \Lambda(de) < \infty$$

$$\text{D.h. } f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_d \end{pmatrix} \text{ und } \forall k \in \mathbb{N} : f_k = \begin{pmatrix} f_{k,1} \\ \vdots \\ f_{k,d} \end{pmatrix}.$$

Jetzt definieren wir

$$\forall i = 1, \dots, d : X_i := \bigvee_{k=1}^{\infty} u_k^{-1} f_{k,i}.$$

Nun zeigen wir dass dieser Vektor  $X = (X_1, \dots, X : d)$  max-stabil ist :

**Max-Stabil:**

Erinnerung: Sei  $f$  eine Funktion auf  $X$ , dann ist  $\text{supp}(f) := \{x \in X \mid f(x) \neq 0\}$

Dann ist  $\Phi_c$  durch  $n$  mal Verdünnung wo jeweils  $n - \frac{1}{n}$  Punkte gestrichen werden. D.h. die Verdünnten  $\Phi_c^{(i)}$ ,  $i = 1, \dots, d$  mit jeweils  $\frac{1}{n}$  Punkten sind dann PPP und müssen mit  $\frac{1}{n}$  geschlürft werden, damit mittels der

Superpositionseigenschaft

$$\Phi_c \stackrel{d}{=} \sum_{i=1}^d \Phi_{\frac{c}{n}}^{(i)}, \quad \Phi_{\frac{c}{n}}^{(i)} \stackrel{u.i.v.}{\sim} \Phi_{\frac{c}{n}}$$

Wir definieren

$$X_c := \max \text{Supp}(\Phi_c) \quad X_{\frac{c}{n}}^{(i)} = \max \text{supp}(\Phi_{\frac{c}{n}}^{(i)})$$

Dann gilt

$$\begin{aligned} \max \text{supp}(\sum_{i=1}^d \Phi_{\frac{c}{n}}^{(i)}) &= \max(\bigcup_{i=1}^d \text{supp}(\Phi_{\frac{c}{n}}^{(i)})) \\ &= \max_{i=1, \dots, d} \max(\text{supp}(\Phi_{\frac{c}{n}}^{(i)})) \\ &= \max_{i=1, \dots, d} X_{\frac{c}{n}}^{(i)} \end{aligned}$$

und aufgrund von

$$X_c := \max \text{supp}(\Phi_c) \stackrel{d}{=} \max \text{supp}(\sum_{i=1}^d \Phi_{\frac{c}{n}}^{(i)})$$

gilt

$$X_c \stackrel{d}{=} \max_{i=1, \dots, d} X_{\frac{c}{n}}^{(i)}$$

Damit haben wir schon einmal Max- $\infty$ -oft teilbarkeit gezeigt. D.h. wir haben eine Verteilung  $G_n$  gefunden, sodass  $G_n^n = G$  gilt, falls

$$X_c \sim G, \quad \max_{i=1, \dots, d} X_{\frac{c}{n}}^{(i)} \sim G_n, \quad X_{\frac{c}{n}}^{(1)} \sim G_n^n$$

Nun müssen wir zeigen, dass  $G_n$  und  $G$  vom gleichen Typen sind und dann haben wir max-stabilität gezeigt.

Da

$$\begin{aligned} F_{X_{\frac{c}{n}}^{(i)}}(t) &\stackrel{oben}{=} F_{X_{\frac{c}{n}}^{(1)}}^n(t) \\ &= \mathbb{P}(X_{\frac{c}{n}}^{(1)} \leq t)^n \\ &= \mathbb{P}\left(\frac{1}{n} \cdot (c\lambda \otimes \Lambda(\lambda))(t, \infty) = 0\right)^n \\ &= \mathbb{P}\left(\frac{1}{n} X_c^{(1)} \leq t\right)^n \\ &= F_{X_{\frac{c}{n}}^{(i)}}(nt) \end{aligned}$$

Also vom gleichen Typen und damit Max Stabil, d.h.  $G_n(t) = G(nt)$

**Multivariate Verteilung:**

Es gilt

$$\begin{aligned}
\forall i = 1, \dots, d : \quad X_i \leq x^{(i)} &\stackrel{Def}{\iff} \forall i = 1, \dots, d : \quad \bigvee_{k=1}^{\infty} u_k^{-1} f_{k,i} \leq x^{(i)} \\
&\iff \forall i = 1, \dots, d \text{ und } \forall k \in \mathbb{N} : u_k^{-1} f_{k,i} \leq x^{(i)} \\
&\iff \forall i = 1, \dots, d \text{ und } \forall k \in \mathbb{N} : \frac{f_{k,i}}{x^{(i)}} \leq u_k \\
&\iff \forall k \in \mathbb{N} : \bigvee_{i=1}^d \frac{f_{k,i}}{x^{(i)}} \leq u_k \\
&\iff \nexists k \in \mathbb{N} : \bigvee_{i=1}^d \frac{f_{k,i}}{x^{(i)}} \geq u_k \\
&\iff \text{Es gilt nicht} : \bigvee_{i=1}^d \frac{f_i}{x^{(i)}} \geq \mathbf{u} \\
&\iff \Phi \left( \left\{ (u, f) \mid \bigvee_{i=1}^d \frac{f_i}{x^{(i)}} \geq \mathbf{u} \right\} \right) = 0
\end{aligned}$$

Dies können wir nun einsetzen und erhalten, da mit  $de$  und  $du$  das Lebesgue Maß gemeint ist:

$$\begin{aligned}
\mathbb{P}(X_1 \leq x^{(1)}, \dots, X_d \leq x^{(d)}) &= \mathbb{P} \left( \Phi \left( \left\{ (u, f) \mid \bigvee_{i=1}^d \frac{f_i}{x^{(i)}} \geq \mathbf{u} \right\} \right) = 0 \right) \\
&= e^{-c \cdot \lambda \otimes \Lambda(\lambda) \left( \left\{ (u, f) \mid \bigvee_{i=1}^d \frac{f_i}{x^{(i)}} \geq \mathbf{u} \right\} \right)}
\end{aligned}$$

Mit Stochastik 1 gilt für zwei Maße:

$$\begin{aligned}
&\mu_1 \otimes \mu_2(A) \\
&\stackrel{\text{Produktmaß}}{=} \int_{\Omega_1} \mu_2(A_{\omega_1}) \, d\mu_1(\omega_1)
\end{aligned}$$

Dies nun angewendet für unser Produktmaß oben

$$\begin{aligned}
\mathbb{P}(X_1 \leq x^{(1)}, \dots, X_d \leq x^{(d)}) &= e^{-c \cdot \lambda \otimes \Lambda(\lambda) \left( \left\{ (u, f) \mid \bigvee_{i=1}^d \frac{f_i}{x^{(i)}} \geq \mathbf{u} \right\} \right)} \\
&= e^{\int_E -c \cdot \lambda \left( \left\{ (u, f) \mid \bigvee_{i=1}^d \frac{f_i(e)}{x^{(i)}} \geq \mathbf{u} \right\} \right) \Lambda(de)} \\
&= e^{\int_E -c \left( \bigvee_{i=1}^d \frac{f_i(e)}{x^{(i)}} - 0 \right) \Lambda(de)} \\
&= e^{\int_E -c \bigvee_{i=1}^d \frac{f_i(e)}{x^{(i)}} \Lambda(de)}
\end{aligned}$$

Schließlich betrachten wir noch die Randverteilungen:

**Randverteilungen:**

Es gilt  $\forall i = 1, \dots, d$ :

$$\begin{aligned} \mathbb{P}(X_i \leq x^{(i)}) &\stackrel{\text{Randverteilung}}{=} \mathbb{P}(X_i \leq x^{(i)}, X_j \leq \infty \forall i \neq j) \\ &\stackrel{\text{oben}}{=} e^{\int_E -c \frac{f_i(e)}{x^{(i)}} \Lambda(de)} \\ &= e^{-c \frac{\int_E f_i(e) \Lambda(de)}{x^{(i)}}} \end{aligned}$$

Dabei ist  $c \frac{\int_E f_i(e) \Lambda(de)}{x^{(i)}} \in (0, \infty)$  und damit ist  $X_i \sim \Phi_1$ , falls

$$c \int_E f_i(e) \Lambda(de) = 1$$

denn dann ist

$$\mathbb{P}(X_i \leq x^{(i)}) = e^{-(x^{(i)})^{-1}}$$

**Definition 4.4.4.** Das Tupel  $(E, \mathcal{B}(E), \Lambda, f)$  heißt in diesem Fall auch Spektraldarstellung von  $X$ . Allgemeiner gilt, dass jede Funktion von der Form

$$G_*(\mathbf{x}) = \int_{\Omega} \bigvee_{i=1}^d \frac{f_i(\omega)}{x^{(i)}} \Lambda(d\omega), \quad \mathbf{x} > \mathbf{0}.$$

mit einem Maßraum  $(\Omega, \mathcal{A}, \Lambda)$  und (komponentenweise) integrierbarem  $f : \Omega \rightarrow [0, \infty)^d$  eine max-stabile Verteilungsfunktion mit  $\Phi_{1, \Lambda(f_i)}$ -Rändern ist.

**Bemerkung 4.4.5.** Wir werden im Folgenden zeigen, dass jeder max-stabile Zufallsvektor (mit  $\Phi_1$ -Rändern) eine Spektraldarstellung besitzt und dass der Maßraum hierfür nicht zu groß gewählt werden muss. (Es genügt sogar (unabhängig von der Dimension des Zufallsvektors) das Intervall  $(\Omega, \mathcal{A}, \Lambda) = ([0, 1], \mathcal{B}([0, 1]), dt)$  zu wählen.)

Man beachte weiterhin, dass die Spektraldarstellung eines max-stabilen Zufallsvektors nicht eindeutig ist. Wir werden ebenfalls zeigen, dass durch die Wahl zweier Normen auf  $\mathbb{R}^d$  ein eindeutiges "Winkelmaß" als Spektralmaß erzwungen werden kann. Zunächst benötigen wir hierfür ein Strukturresultat für max- $\infty$ -oft teilbare Verteilungen.

**Bemerkung 4.4.6.** Wir können das Spektralmaß auch auf der Einheitssphäre definieren:

$$\begin{aligned}
G_*(\mathbf{x}) &= \int_{\Omega} \bigvee_{i=1}^d \frac{f_i(\omega)}{x^{(i)}} \Lambda(d\omega), \quad \mathbf{x} > \mathbf{0} \\
&= \int_{\Omega} \bigvee_{i=1}^d \frac{f_i(\omega)}{x^{(i)}} \|\omega\|_p^{-1} \cdot \|\omega\|_p \Lambda(d\omega) \\
&= \int_{\Omega} \bigvee_{i=1}^d \frac{f_i(\omega)}{x^{(i)}} \|\omega\|_p^{-1} \cdot (\|\omega\|_p \Lambda(d\omega)) \\
&\stackrel{???}{=} \int_{\{\omega \mid \|\omega\|_p=1\}} \bigvee_{i=1}^d \frac{f_i(\omega)}{x^{(i)}} \|\omega\|_p^{-1} \cdot \tilde{\Lambda}(d\omega)
\end{aligned}$$

Mit  $\tilde{\Lambda}(d\omega) = \|\omega\|_p \Lambda(d\omega)$  und  $\mathbf{x} \in (0, \infty]$

**Satz 4.4.7.** (Spektral-/Winkelmaß-Charakterisierung max-stabiler Verteilungen)

Sei  $\mathbb{S}_2 := \{\omega \mid \|\omega\|_2 = 1\}$ .

Die Menge der d-variaten max-stabilen Verteilungen  $G_*$  mit  $\Phi_1$ -Rändern steht durch die Wahl zweier Normen  $\|\cdot\|_1, \|\cdot\|_2$  auf  $\mathbb{R}^d$  durch die Beziehung

$$G_*(\mathbf{x}) = \exp \left( - \int_{\mathbb{S}_2} \bigvee_{i=1}^d \frac{\omega^{(i)}}{x^{(i)}} \|\omega\|_1^{-1} \sigma_*(d\omega) \right), \quad \mathbf{x} \in [\mathbf{0}, \infty]$$

in 1:1-Korrespondenz mit der Menge der aller Radonmaße auf  $\mathbb{S}_2$ , welche die Bedingung

$$\int_{\mathbb{S}_2} \omega^{(i)} \|\omega\|_1^{-1} \sigma_*(d\omega) = 1, \quad i = 1, \dots, d$$

erfüllen.

**Beispiel 4.4.8.** X d-variater max stabil...

**Lemma 4.4.9.**  $G$  ist max- $\infty$ -oft-Teilbar  $\iff \forall t > 0 : G^t$  ist eine Verteilungsfunktion

*Beweis.*  $G$  ist max- $\infty$ -oft-Teilbar

$$\iff \forall n \in \mathbb{N} : G_n^n = G \text{ und } G_n \text{ ist eine Verteilungsfunktion}$$

$$\iff \forall n \in \mathbb{N} : G_n = G^{\frac{1}{n}} \text{ und } G^{\frac{1}{n}} \text{ ist eine Verteilungsfunktion}$$

$$\iff \forall n \in \mathbb{N} : G^{\frac{k}{n}} \text{ ist eine Verteilungsfunktion}$$

$$\Rightarrow \forall t > 0 : G^t \text{ ist eine Verteilungsfunktion}$$

□

**Bemerkung 4.4.10.** Sei  $\mathbf{l} \in [-\infty, \infty]^d$  der "lower Bound" einer Verteilung und  $\mu$  ein Radonmaß auf  $E := [1, \infty) \times \{\mathbf{l}\}$  und

$$\Phi := \sum_{k=1}^{\infty} \delta_{(t_k, j_k)} \sim PPP(dt \times \mu)$$

Sei außerdem für  $t > 0$ :

$$Y(t) := \mathbf{l} \vee \bigvee_{t_k \leq t} j_k$$

Wir sind nun an der Verteilung interessiert. Dabei ist  $t \in (0, \infty)$  und  $y \in E$ .

BILD

Hier ist  $\mathbf{l} = 0$ . D.h. wir sind an der Verteilung der Maxima der Tupel, wobei die erste Komponente kleiner sein soll als  $t$ . Dann ist

$$\begin{aligned} \mathbb{P}(Y(t) \leq \mathbf{y}) &= \mathbb{P}(\Phi([0, t] \times [-\infty, \mathbf{y}]^C) = 0) \\ &\stackrel{Poi}{\sim} e^{-(\lambda \otimes \mu)([0, t] \times [-\infty, \mathbf{y}]^C)} \\ &= \begin{cases} e^{-t\mu[-\infty, \mathbf{y}]} & , \mathbf{y} \geq \mathbf{l} \\ 0 & , \text{sonst} \end{cases} \end{aligned}$$

Ein exponent measure hat bestimmte Eigenschaften...

**Satz 4.4.11.** Eine multivariate Verteilungsfunktion  $G$  ist genau dann max- $\infty$ -oft teilbar, wenn es ein  $\mathbf{l} \in [-\infty, \infty)^d$  und ein Exponent measure  $\mu$  auf  $E := [\mathbf{l}, \infty] \setminus \{\mathbf{l}\}$  gibt, so dass  $G$  in der Form

$$G(\mathbf{y}) = \begin{cases} \exp(-\mu([- \infty, \mathbf{y}]^c)), & \mathbf{y} \geq \mathbf{l}, \\ 0, & \text{sonst} \end{cases}$$

dargestellt werden kann.

*Beweis.* " $\Leftarrow$ ":

1. Zunächst zur max- $\infty$ -oft-Teilbarkeit: Wir wollen  $\forall n \in \mathbb{N}$  eine Verteilungsfunktion  $G_n$  finden, die  $G_n^n = G$  erfüllt.



Dazu definieren wir  $\tilde{\mu}_n := \frac{1}{n}\mu$  und die Funktion

$$G_n(y) := \begin{cases} e^{-\tilde{\mu}_n([-\infty, y]^C)} & , y \geq l \\ 0 & , y < l \end{cases}$$

Dann gilt offensichtlich

$$G_n^n(y) = G(y) \quad \forall y \in \mathbb{R}$$

- $G$  ist eine Verteilungsfunktion, da

i)  $G(\infty) = 1$  und  $G(-\infty) = 0$

ii) Bild  $[0,1]$ , klar mit i) und iv)

iii) rechtsstetig

iv) Rechtecksmotonomie:

Sei  $d = 2$  und  $y \leq x : \iff \forall i = 1, 2 : y_i \leq x_i$ . Damit gilt dann

$$G((x_1, x_2)) - G((y_1, x_1)) + G((y_1, y_2)) - G((x_1, y_1)) \geq 0$$

BILD

Da Maße Monoton sind, d.h.  $\forall A \subseteq B : \mu(A) \leq \mu(B)$ .

Alternative Beweisidee mit PPP???

" $\Rightarrow$ ":

Wir gehen in zwei Schritten vor:

1. Zunächst zeigen wir, dass  $A := \{x \in \mathbb{R}^d : G(x) > 0\}$  ein Rechteck ist, d. h.

$$x \in A \text{ und } x \leq y \Rightarrow G(y) > 0 \Rightarrow y \in A$$

und

$$z.z. \mathbf{x}, \mathbf{y} \in A \Rightarrow x \wedge y \in A$$

Für letztere Bedingung genügt es, die Ungleichung  $G(\mathbf{x} \wedge \mathbf{y}) \geq G(\mathbf{x})G(\mathbf{y})$  zu zeigen, bzw.

$$\log G(\mathbf{x} \wedge \mathbf{y}) \geq \log G(\mathbf{x}) + \log G(\mathbf{y}).$$

Wir wollen zeigen  $G^{1/n}(x) + G^{1/n}(x) \leq G^{1/n}(x \wedge y)$ .

Sei nun  $Y \sim G^{1/n}$ . Wir schätzen  $G^{1/n}(x)$  nach oben ab, sodass wir das erhalten, was wir zeigen wollen:

$$\begin{aligned} G^{1/n}(\mathbf{x}) &= \mathbb{P}(Y \leq \mathbf{x}) \\ &= \mathbb{P}(Y \leq \mathbf{x}, Y \leq \mathbf{y}) + \mathbb{P}(Y \leq \mathbf{x}, \text{ nicht } Y \leq \mathbf{y}) \\ &\leq \mathbb{P}(Y \leq \mathbf{x} \wedge \mathbf{y}) + (1 - \mathbb{P}(Y \leq \mathbf{y})) \\ &= G^{1/n}(\mathbf{x} \wedge \mathbf{y}) + 1 - G^{1/n}(\mathbf{y}). \end{aligned}$$

Wegen

$$n \left( G^{1/n}(x) - 1 \right) \sim n \log G^{1/n}(x) = \log(G(x))$$

gilt für  $n \rightarrow \infty$  :

$$\begin{aligned} \log G(\mathbf{x}) + \log G(\mathbf{y}) &\leftarrow n \left( G^{1/n}(x) - 1 \right) + n \left( G^{1/n}(y) - 1 \right) \\ &= n \left( G^{1/n}(x) + G^{1/n}(y) - 1 \right) \\ &\stackrel{\text{oben}}{\leq} n \left( G^{1/n}(x \wedge y) - 1 \right) \rightarrow \log G(x \wedge y) \end{aligned}$$

2. Setze  $\mathbf{l} := \inf A$  und  $E := [\mathbf{l}, \infty] \setminus \{\mathbf{l}\}$ . Auf  $E$  ist durch  $\mu_n([\mathbf{l}, \mathbf{x}]) := nG^{1/n}(\mathbf{x})$  ein Maß  $\mu_n$  definiert. Dann ist (wegen dem Korr. nach der Poi-Approximation):

$$\mu_n([-\infty, x]^c) = n \left( 1 - G^{1/n}(x) \right) \rightarrow -\log G(x).$$

Man kann zeigen, dass  $\mu_n \rightarrow^v \mu$  mit  $\mu([-\infty, x]^c) = -\log G(x)$ , d. h.

$$G(\mathbf{x}) = \exp(-\mu([-\infty, \mathbf{x}]^c))$$

z.z.  $G$  ist W-Maß (weggelassen)

□

**Bemerkung 4.4.12.** Im vorherigen Kapitel haben wir die GEV  $G_*$  mit  $\Phi_1$  Rändern charakterisiert und gezeigt, dass

$$G_*^n(nx) = G_*(x) \quad \forall n \in \mathbb{N}$$

gilt.

Dann gilt mit Lemma..., dass auch

$$\forall t > 0 : \quad G_* t(tx) = G_*(x)$$

gilt.

Bei der Freshet Verteilung ist glücklicherweise der lower bound  $\mathbf{1} = 0$ , da die Verteilung nur auf  $(0, \infty)$  definiert ist. Da  $G_*$  max- $\infty$ -oft teilbar ist existiert ein expent measure  $\mu_*$  auf  $[0, \infty] \setminus \{0\}$ , sodass ...

**Bemerkung 4.4.13.** Sei  $(t_1, \dots, t_n) \in [0, \infty)^n$  und  $\Phi = \sum_{i=1}^{\infty} \delta_{X_i} \sim PPP(\Lambda)$  auf  $[0, \infty)^{(t_1, \dots, t_n)}$ . Sei im folgenden  $n=2$ .

D.h.  $X_i = \begin{pmatrix} x_{t_1}^{(i)} \\ x_{t_2}^{(i)} \end{pmatrix}$  wird interpretiert als Gewitter stärke einmal in  $t_1$  und in  $t_2$ . Wenn wir nicht darauf eingehen, wie die Gewitterverläufe mit den Punkten in verbindung liegen, kann eine Realisierung dieses PPP sein

BILD

Man kann auch die Vektoren  $X_i$  betrachten und auf die Einheitssphäre mit der 1 Norm Projizieren, mittels

$$\tilde{X}_i := \frac{X_i}{\|X_i\|} :$$

BILD

Wir erinnern uns an die Definition der  $X_i$ :

$$X_i = \begin{pmatrix} x_{t_1}^{(i)} \\ x_{t_2}^{(i)} \end{pmatrix} \sim \Lambda =: \Lambda_1 \times \Lambda_2$$

Fall 1: Seien  $\forall i \in \mathbb{N}$  die  $x_{t_1}^{(i)}$  und  $x_{t_2}^{(i)}$  vollständig abhängig. Damit liegt der Vektor  $X_i$  auf der Winkelhalbierenden und der Vektor  $\tilde{X}_i$  wie folgt auf der Einheitsphäre:

BILD

D.h. das Intensitätsmaß ist  $\Lambda = \delta_{(\frac{1}{2}, \frac{1}{2})}$ . Das ganze können wir nochmal sauber nachrechnen. Dazu brauchen wir aber das  $\Phi_1$ -Ränder um schöne Ergebnisse zu bekommen.

Sei wieder

$$X_i := \begin{pmatrix} X_{t_1}^{(i)} \\ X_{t_2}^{(i)} \end{pmatrix} \quad \text{mit} \quad X_1^{(i)} = X_2^{(i)} \sim \Phi_1$$

D.h. die komponenten sind vollständig abhängig. Dann ist die Verteilungsfunktion gegeben durch:

$$\begin{aligned} G(\mathbf{c}) &= \mathbb{P}(X_i \leq \mathbf{c}) = \mathbb{P}(X_1^{(i)} \leq c_1, X_2^{(i)} \leq c_2) = \mathbb{P}(X_1^{(i)} \leq \min\{c_1, c_2\}) \\ &= e^{-\frac{1}{\min\{c_1, c_2\}}} = e^{-\max\{c_1, c_2\}} \end{aligned}$$

Jetzt wollen die Aussage verwenden, dass für jede max-stabile Verteilungsfunktion  $G$  mit  $\Phi_1$  Rändern

in der folgenden Form geschrieben werden kann

$$G(\mathbf{c}) = G_*(\mathbf{c}) = \exp \left( \int_{\Omega} \bigvee_{i=1}^2 \frac{\omega_i}{c_i} \Lambda(d\omega) \right) = \exp \left( \int_{\mathbb{S}_1} \bigvee_{i=1}^2 \frac{\omega_i}{c_i} \|\omega\|^1 \sigma_*(d\omega) \right)$$

Um das Spektralmaß  $\sigma_*$  zu finden überprüfen wir die Randbedingung. Nach Satz ... existiert ja eine Bijektion aller Radonmaße auf  $\mathbb{S}_2$ , die die Randbedingung erfüllen, zu der Radonmaße, die die obige Bedingung erfüllen. Die Randbedingung ist

$$_1 = 1, \dots, n : \quad \int_{\mathbb{S}_2} \omega_i \|\omega\|^{-1} \sigma_*(d\omega) \stackrel{???}{=} \int_{\mathbb{S}_2} \omega_i 2 \cdot \delta_{(\frac{1}{2}, \frac{1}{2})}(d\omega) = \frac{1}{2} \cdot 2 = 1$$