

Theorem 13 ($MDA(\wedge)$, alternative Charakterisierung)

Eine Verteilungsfunktion F gehört zu $MDA(\wedge)$ dann und nur dann wenn es eine positive Funktion \tilde{a} existiert, sodass

$$\lim_{x \uparrow x_F} \frac{\bar{F}(x + u\tilde{a}(x))}{\bar{F}(x)} = e^{-u}, \forall u \in \mathbb{R}$$

Eine mögliche Wahl für \tilde{a} ist $\tilde{a}(x) = a(x)$

$$a(x) = \int_x^{x_F} \frac{\bar{F}(t)}{\bar{F}(x)} dt$$

Die Funktion $a(x)$ heißt durchschnittliche Überschufunktion (mean excess function):

$$a(x) = E(X - x | X > x), \forall x \leq x_F$$

Einige Verteilungen, die dem $MDA(\wedge)$ gehören:

- Normal: $F(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}$, $x \in \mathbb{R}$.
- Exponential: $f(x) = \lambda^{-1} \exp\{-\lambda x\}$, $x > 0$, $\lambda > 0$.
- Lognormal: $f(x) = (2\pi x^2)^{-1/2} \exp\{-(\ln x)^2/2\}$, $x > 0$.
- Gamma: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}$, $x > 0$, $\alpha, \beta > 0$.

Graphische Methoden zur Untersuchung des Verteilungsrandes

- Histogramm
- Quantil-Quantil Plots

X_1, X_2, \dots, X_n sind i.i.d. ZV mit einer unbekanntem Verteilung \tilde{F} . Es wird vermutet, dass \tilde{F} am Rand von einer bekannten Verteilung F approximiert wird. Wie kann man diese Vermutung testen?

Sei $X_{n,n} \leq X_{n-1,n} \leq \dots \leq X_{1,n}$ eine sortierte Stichprobe aus X_1, X_2, \dots, X_n .

qq-plot: $\{(X_{k,n}, F^{\leftarrow}(\frac{n-k+1}{n+1})) : k = 1, 2, \dots, n\}$.

Bei einer plausiblen Vermutung stellt der qq-plot eine einigermaßen lineare Abhängigkeit dar. Diese Eigenschaft bleibt auch dann erhalten wenn die echte Verteilung und die Referenz-Verteilung nicht übereinstimmen, sondern vom selben Typus sind.

Faustregel: Je größer das Quantil um so mehr “heavy tailed” ist die Verteilung!

Der Hill Schätzer

Seien X_1, X_2, \dots, X_n i.i.d. mit Verteilungsfunktion F , sodass $\bar{F} \in RV_{-\alpha}$, $\alpha > 0$, d.h. $\bar{F}(x) = x^{-\alpha}L(x)$ mit $L \in RV_0$.

Ziel: Schätzung von α !

Theorem 14 (Satz von Karamata)

Sei L eine langsam variierende und lokal beschränkte Funktion auf $[x_0, +\infty)$ für ein $x_0 \in \mathbb{R}$. Dann gilt:

(a) Für $\kappa > -1$:

$$\int_{x_0}^x t^\kappa L(t) dt \sim \frac{1}{\kappa + 1} x^{\kappa+1} L(x) \text{ für } x \rightarrow \infty$$

(b) Für $\kappa < -1$:

$$\int_x^{+\infty} t^\kappa L(t) dt \sim -\frac{1}{\kappa + 1} x^{\kappa+1} L(x) \text{ für } x \rightarrow \infty$$

Beweis in Bingham et al. 1987.

Annahme: L ist lokal beschränkt in $[u, +\infty)$.

Aus dem Satz von Karamata folgt:

$$E(\ln(X) - \ln(u) | \ln(X) > \ln(u)) = \lim_{u \rightarrow \infty} \frac{1}{\bar{F}(u)} \int_u^\infty (\ln x - \ln u) dF(x) = \alpha^{-1}. \quad (8)$$

Für die empirische Verteilung $F_n(x) = \frac{1}{n} \sum_{k=1}^n I_{[x_k, \infty)}(x)$ und eine hohe stichprobenabhängige Schwelle $x_{k,n}$, erhält man:

$$E(\ln(X) - \ln(x_{k,n}) | \ln(X) > \ln(x_{k,n})) \approx \frac{1}{\bar{F}_n(x_{k,n})} \int_{x_{k,n}}^\infty (\ln x - \ln x_{k,n}) dF_n(x) = \frac{1}{k-1} \sum_{j=1}^{k-1} (\ln x_{j,n} - \ln x_{k,n}).$$

Wenn $k = k(n) \rightarrow \infty$ und $k/n \rightarrow 0$, dann $x_{k,n} \rightarrow \infty$ für $n \rightarrow \infty$, und dann folgt aus (8):

$$\lim_{n \rightarrow \infty} \frac{1}{k-1} \sum_{j=1}^{k-1} (\ln x_{j,n} - \ln x_{k,n}) \stackrel{d}{=} \alpha^{-1}$$

Der untenstehende Hill-Schätzer ist also konsistent:

$$\hat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k (\ln x_{j,n} - \ln x_{k,n}) \right)^{-1}$$

Wie wird ein passendes k für eine gegebene Stichprobengröße n gewählt?

k zu klein: hohe Varianz des Schätzers!

k zu gross: Schätzer basiert auf zentrale Werte der Verteilung \implies
Der Schätzer ist verzerrt!

Grafische Inspektion des Hill Plots: $\left\{ \left(k, \hat{\alpha}_{k,n}^{(H)} \right) : k = 2, \dots, n \right\}$

Für einen gegebenen Schätzer $\hat{\alpha}_{k,n}^{(H)}$ von α erhält man folgenden Schätzer für die Randverteilung \hat{F} :

$$\hat{F}(x) = \frac{k}{n} \left(\frac{x}{x_{k,n}} \right)^{-\hat{\alpha}_{k,n}^{(H)}}.$$

und folgenden Quantil-Schätzer:

$$\hat{q}_p = \hat{F}^{\leftarrow}(p) = \left(\frac{n}{k}(1-p) \right)^{-1/\hat{\alpha}_{k,n}^{(H)}} x_{k,n}.$$

Die POT Methode (Peaks over Threshold)

Definition 18 (Die verallgemeinerte Pareto Verteilung (GPD))

Die standard GPD G_γ :

$$G_\gamma(x) = \begin{cases} 1 - (1 + \gamma x)^{-1/\gamma} & \text{für } \gamma \neq 0 \\ 1 - \exp\{-x\} & \text{für } \gamma = 0 \end{cases}$$

wobei $x \in D(\gamma)$

$$D(\gamma) = \begin{cases} 0 \leq x < \infty & \text{für } \gamma \geq 0 \\ 0 \leq x \leq -1/\gamma & \text{für } \gamma < 0 \end{cases}$$

oder $G_\gamma(x) = 1 - (1 + \gamma x)^{-1/\gamma}$, $x \in D(\gamma)$ für $\gamma \neq 0$, und $G_0 = \lim_{\gamma \rightarrow 0} G_\gamma$.

Sei $\nu \in \mathbb{R}$ und $\beta > 0$. Eine GPD ist durch die untenstehende Verteilungsfunktion gegeben

$$G_{\gamma, \nu, \beta} = 1 - \left(1 + \gamma \frac{x - \nu}{\beta}\right)^{-1/\gamma}$$

wobei $x \in D(\gamma, \nu, \beta)$ und

$$D(\gamma, \nu, \beta) = \begin{cases} \nu \leq x < \infty & \text{für } \gamma \geq 0 \\ \nu \leq x \leq \nu - \beta/\gamma & \text{für } \gamma < 0 \end{cases}$$

Theorem 15 (Charakterisierung von $MDA(H_\gamma)$)

Sei $\gamma \in \mathbb{R}$. Die untenstehenden Aussagen sind äquivalent:

(i) $F \in MDA(H_\gamma)$

(ii) Es existiert eine positive messbare Funktion $a(\cdot)$, sodass für $x \in D(\gamma)$

$$\lim_{u \uparrow x_F} \frac{\bar{F}(u + xa(u))}{\bar{F}(u)} = \bar{G}_\gamma(x).$$

Definition 19 (Exzess-Verteilung)

Sei X eine ZV mit Verteilungsfunktion F und rechtem Endpunkt x_F .

Für $u < x_F$

$$F_u(x) = P(X - u \leq x | X > u), x \geq 0$$

heißt Exzess-Verteilungsfunktion über die Schwelle u .

Theorem 16 (Eine weitere Charakterisierung von $MDA(H_\gamma)$)

Sei $\gamma \in \mathbb{R}$. Folgende Aussagen sind äquivalent:

(i) $F \in MDA(H_\gamma)$

(ii) Es existiert eine positive messbare Funktion $\beta(\cdot)$, sodass

$$\lim_{u \uparrow x_F} \sup_{x \in (0, x_F - u)} |F_u(x) - G_{\gamma, 0, \beta(u)}(x)| = 0$$

POT: Schätzer für den Tail und das Quantil der Exzess-Verteilung

Seien X_1, \dots, X_n i.i.d. ZV mit Verteilungsfunktion $F \in MDA(H_\gamma)$ für $\gamma \in \mathbb{R}$.

- Wähle eine hohe Schwelle u (unter Verwendung von geeigneten stat. Verfahren) und berechne

$$N_u = \# \{i \in \{1, 2, \dots, n\}: X_i > u\}$$

- Sei Y_1, Y_2, \dots, Y_{N_u} eine Stichprobe von Exzess-Beobachtungen. Bestimme $\hat{\beta}$ und $\hat{\gamma}$, sodass folgendes gilt:

$$\bar{F}_u(y) \approx \bar{G}_{\hat{\gamma}, 0, \hat{\beta}(u)}(y),$$

wobei $\bar{F}_u(y) = P(X - u > y | X > u)$

- Kombiniere die obigen zwei Schritte um folgende Schätzer zu berechnen:

$$\widehat{\bar{F}}(u + y) = \frac{N_u}{n} \left(1 + \hat{\gamma} \frac{y}{\hat{\beta}} \right)^{-1/\hat{\gamma}} \quad (9)$$

$$\hat{q}_p = u + \frac{\hat{\beta}}{\hat{\gamma}} \left(\left(\frac{n}{N_u} (1 - p) \right)^{-\hat{\gamma}} - 1 \right) \quad (10)$$