

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. Es wird vermutet, dass eine bekannte Verteilung F die unbekanntem Verteilung approximiert. 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 sind 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 !

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

Aus dem Satz von Karamata (siehe Bingham et al. 1987) 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:

$$\frac{1}{\bar{F}_n(X_{k,n})} \int_{X_{k,n}}^\infty (\ln x - \ln X_{k,n}) dF_n(x) =$$

$$E(\ln(X) - \ln(X_{k,n}) | \ln(X) > \ln(X_{k,n})) = \frac{1}{k-1} \sum_{j=1}^{k-1} (\ln X_{j,n} - \ln X_{k,n}).$$

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

$$\lim_{n \rightarrow \infty} \frac{1}{k-1} \sum_{j=1}^{k-1} (\ln X_{j,n} - \ln X_{k,n}) = \alpha^{-1} \implies \hat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^k (\ln X_{j,n} - \ln X_{k,n}) \right)^{-1}$$

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

k zu klein: hohe Varianz des Schätzers!

k zu gross: Schätzer basiert auf zentrale Werte der Verteilung \implies Bias!

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 Tail-Schätzer:

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

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 16 (Die verallgemeinerte Pareto Verteilung (GPD))

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

Sei $\nu \in \mathbb{R}$ und $\beta > 0$. Eine GPD ist die Funktion

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

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 14 (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 $x \in D(\gamma)$

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

Definition 17 (Exzess-Verteilung)

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

Für $x < x_F$

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

heißt Exzess-Verteilungsfunktion über die Schwelle u .

Theorem 15 (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

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