

Robuste Statistik

Blatt 3

Aufgabe 3.1: (3 Punkte)

Betrachten Sie den Datensatz

2, 3, 5, 6, 9.

- a) Bestimmen Sie die Lokations-Tiefe der Parameter $\mu \in \{-1, 2.5, 5, 7, 10\}$ mittels einer dafür selbst erstellten R-Funktion.
- b) Plotten Sie die Lokations-Tiefe für $\mu \in [-1, 10]$ auf Grundlage des gegebenen Datensatzes. Welches μ liefert die maximale Lokations-Tiefe?
- c) Welche der in a) untersuchten Werte für μ sind Lokations-Nonfits? Begründen Sie Ihre Aussage!

Aufgabe 3.2: (3 Punkte)

Wir betrachten Lokations-M-Schätzer $\hat{\mu}_\rho(y)$ basierend auf Score-Funktionen ρ aus Definition 3.3.1. und Realisationen $y = (y_1, \dots, y_N)^\top$ von unabhängig, identisch verteilten Zufallsvariablen. Zeigen Sie, dass der Lokations-M-Schätzer mit Score-Funktion $\rho(z) = |z|$ einen Maximum-Likelihood-Schätzer der Lokation $\mu \in \mathbb{R}$ definiert.

Aufgabe 3.3: (4 Punkte)

Betrachten Sie den Datensatz

0, 0, 0, 0, 10.

Die 10 ist dabei als unerwünschter Ausreißer zu betrachten. Betrachten Sie die beiden Schätzfunktionen

$$\hat{\theta}_1^p(\mathbf{y}) = p \cdot \arg \min_{\theta \in \mathbb{R}} \sum_{n=1}^N (y_n - \theta)^2 + (1 - p) \cdot \arg \min_{\theta \in \mathbb{R}} \sum_{n=1}^N |y_n - \theta|, \quad p \in [0, 1],$$

$$\hat{\theta}_2^p(\mathbf{y}) = \arg \min_{\theta \in \mathbb{R}} \sum_{n=1}^N (p \cdot (y_n - \theta)^2 + (1 - p) \cdot |y_n - \theta|), \quad p \in [0, 1].$$

- a) Berechnen Sie eine Schätzung für den gegebenen Datensatz mithilfe von R basierend auf $\hat{\theta}_1^p(\mathbf{y})$ und $\hat{\theta}_2^p(\mathbf{y})$. Sie können dazu die Funktion `optimize()` verwenden. Plotten Sie beide Schätzungen in einer gemeinsamen Grafik in Abhängigkeit von p und erklären Sie die Bedeutung des Parameters p .
- b) Interpretieren Sie Ihre Ergebnisse aus a): Bis zu welcher Größe von p bleiben die Schätzer vom Ausreißer unberührt? Welchen der beiden Schätzer würden Sie bevorzugen? Begründen Sie Ihre Aussage!

Abgabe bis spätestens 23.04.2019, 13 Uhr