Übungsblatt 3

Abgabe per mail an `hoepfner@mathematik.uni-mainz.de` bis **MO 17.06.19**

<u>Aufgabe 3.1 (Lokations- und Skalenmodelle):</u> Sei P ein Wahrscheinlichkeitsmass auf $(I\!R, \mathcal{B}(I\!R))$ mit zugehöriger Verteilungsfunktion F und Dichte f, beliebig aber fest. Dann erzeugt P ein Lokations- und Skalenmodell

$$\left(I\!R^n \, , \, \mathcal{B}(I\!R^n) \, , \, \left\{ P^n_{m,c} := \bigotimes_{i=1}^{n} P_{m,c} \, : \, m \in I\!R, c > 0 \right\} \right)$$

durch die Festsetzung

$$P_{m,c}(dx) \;:=\; \frac{1}{c}\, f\left(\frac{x-m}{c}\right)\, dx \quad , \quad x \in I\!R \,.$$

Man betrachte Schätzer

$$\widehat{m}_n \;:=\; median(X_1, \ldots, X_n)$$

für den Lokationsparameter und

$$\begin{aligned} \widehat{c}_n \;&:=\; \frac{1}{n} \sum_{i=1}^{n} |X_i - \widehat{m}_n| \\ \check{c}_n \;&:=\; median(Y_1, \ldots, Y_n) \quad , \quad Y_i := |\, X_i - median(X_1, \ldots, X_n)\,| \end{aligned}$$

für den Skalenparameter.

a) Zeige: in dem von der Doppelexponentialverteilung erzeugten Lokations- und Skalenmodell

$$(\diamond) \qquad P_{m,c}(dx) \;=\; \frac{1}{2c}\, e^{-|\frac{x-m}{c}|}\, dx$$

ist $T_n := (\widehat{m}_n, \widehat{c}_n)$ Maximum-Likelihood-Schätzer für den unbekannten Parameter (m, c).

b) Man schreibe eine Funktion im Argument n, mit der man sich im Modell $(\diamond)$ die Verteilungen

$$r_n \;:=\; \mathcal{L}\left(\frac{\sqrt{n}\, \widehat{m}_n}{\check{c}_n} \mid P^n_{0,1} \right) \;,\quad v_n \;:=\; \mathcal{L}\left(\frac{\sqrt{n}\, \widehat{m}_n}{\widehat{c}_n} \mid P^n_{0,1} \right) \;,\quad w_n \;:=\; \mathcal{L}\left(\frac{\sqrt{n}\, \overline{X}_n}{\sqrt{\widetilde{S}^2_n}} \mid P^n_{0,1} \right)$$

und ihre Quantile approximativ verschafft, in der folgenden Weise: aufgrund von $N = 5000$ Simulationsläufen mit je n doppelexponentialverteilten iid ZV bilde man die empirische Verteilungsfunktion für jede der drei genannten Grössen, ermittle empirische Quantile, und zeichne diese in ein gutes Histogramm ein. Idealerweise setze man in gleichem Massstab drei Graphiken in Querformat auf eine

Seite, oben ein gutes Histogramm der empirischen Verteilung von r_n mit empirischen Quantilen, in der Mitte dasselbe für v_n, unten für w_n. Man setze dabei den Masstab so an, dass obere und untere 1%, 2.5%, 5% Quantile eingezeichnet sind und optisch verglichen werden können.

<u>Hinweis:</u> von einer Verteilung t_{n-1} der Statistik w_n unter $P_{0,1}^n$ kann hier natürlich keine Rede mehr sein: man ist nicht mehr im Normalverteilungsmodell! Es gibt keine Sätze über die Verteilung des Paares (empirischer Mittelwert, empirische Varianz) in Lokations- und Skalenmodellen, die von einem allgemeinen F mit endlichen zweiten Momenten erzeugt werden: also ist es sinnvoll, die genannten Verteilungen empirisch zu ermitteln.

c) Mit Hilfe der in b) empirisch ermittelten Quantile konstruiere man drei Typen von Konfidenzintervallen für den unbekannten Lokationsparameter, die im Modell ($\diamond$) zu vorgegebenem $\beta \in (0,1)$ den Konfidenzkoeffizient β liefern. Diese drei basiere man auf

$$T_n^{(1)} := (\widehat{m}_n, \breve{c}_n) \quad , \quad T_n^{(2)} := (\widehat{m}_n, \widehat{c}_n) \quad , \quad T_n^{(3)} := \left(\overline{X}_n, \sqrt{\widetilde{S}_n^2}\right)$$

als Schätzer für das Paar $(m,c) \in I\!R \times (0,\infty)$.

d) Anhand des Datensatzes

6.0490067,	20.0324319,	8.0917892,	5.3001702,	8.4231110,	8.4178589,
7.9412088,	6.0004114,	7.6231672,	3.2640157,	7.9500893,	7.1471330,
−3.7385962,	13.1849629,	15.3080767,	6.6531491,	9.1205368,	0.9610968,
4.9916255,	11.5371074,	3.1072696,	5.7450847,	5.0906057,	9.5253568,
4.7945269,	4.0560317,	10.1374325			

vergleiche man die drei in c) gebildeten Konfidenzintervalle für den unbekannten Lokationsparameter $m \in I\!R$. Man simuliere sich selbst weitere Datensätze (variieren von n, m_0, c_0 ...) und führe auf diesen denselben Vergleich aus. Welches Verfahren ist im Modell ($\diamond$) vorzuziehen?

<u>Hinweis:</u> Doppelexponential P_{m_0,c_0}-verteilte Zufallszahlen verschafft man sich so (man überlege sich die Begründung!): man ziehe unabhängige ZV U, V_1, V_2 mit $U \sim \mathcal{B}(1, \frac{1}{2})$ und $V_i \sim \mathcal{E}xp(1)$ und definiere damit $X := m_0 + c_0 \,[\, UV_1 - (1-U)V_2 \,]$.

<u>Aufgabe 3.2 :</u> Man erzeuge in $M = 500$ Läufen je einen Datensatz von $n_0 = 37$ mit Parametern $m_0 = 1.34$ und $c_0 = 0.77$ doppelexponentialverteilten ZV, berechne für diesen Datensatz die Werte der Schätzer $T_n^{(i)}$ nach c) für $i = 1,2,3$, und gebe dann für jeden der drei Schätzer je ein Histogramm der ermittelten Schätzwerte für den Lokationsparameter und ein Histogramm der ermittelten Schätzwerte

für den Skalenparameter. Man achte darauf, zu vergleichende Histogramme jeweils über einem gemeinsamen Intervall darzustellen, welches erlaubt, die Schätzfehler der drei Schätzer im Modell ($\diamond$) direkt zu vergleichen.
Was können Sie über die Schätzfehler der Schätzer $T_n^{(2)}$ oder $T_n^{(3)}$ im Modell ($\diamond$) im Vergleich mit den Schätzfehlern des Maximum-Likelihood-Schätzers $T_n^{(1)}$ aussagen?

Abgabe: Graphiken und Programm

Aufgabe 3.3 : Zum Kontrast betrachte man das von der Cauchyverteilung erzeugte Lokations- und Skalenmodell

$$(\star) \qquad P_{m,c}(dx) \ = \ \frac{1}{\pi} \, \frac{c}{c^2 + (x-m)^2} \, dx \ .$$

Man erzeuge man in $M = 500$ Läufen je einen Datensatz von $n = 37$ mit Parametern $m_0 = 1.34$ und $c_0 = 0.77$ cauchyverteilten Zufallszahlen, berechne aufgrund dieser Daten die Schätzer $T_n^{(i)}$ nach c) für $i = 1, 2, 3$, und gebe für jeden der drei Schätzer je ein Histogramm der ermittelten Schätzwerte für den Lokationsparameter und ein Histogramm der ermittelten Schätzwerte für den Skalenparameter. Wieder achte man darauf, zu vergleichende Histogramme jeweils über einem gemeinsamen Intervall darzustellen, welches erlaubt, die Schätzfehler der drei Schätzer im Modell ($\diamond$) direkt zu vergleichen.
Was fällt auf ? Welchen der drei Schätzer empfehlen Sie für das Modell ($\star$) – und warum ?

Abgabe: Graphiken und Programm