R. Höpfner: Schätzer und Tests, Sommersemester 2012

Übungsblatt 3

Abgabe per mail an `hoepfner@uni-mainz.de` bis **MI 30.05.12**

Besprechung voraussichtlich FR 01.06.12

<u>Aufgabe 3.1 ('Schöne' Likelihoodflächen I):</u> Man betrachte das Normalverteilungsmodell und generiere einen kleinen (etwa mit $n = 4$ oder $n = 5$) sowie einen grösseren Datensatz (ausprobieren: etwa mit $n = 59$ oder $n = 121$ ...) von n nach $\mathcal{N}(\mu_0, \sigma_0^2)$ verteilten Zufallszahlen $X_1, \ldots, X_n$ und visualisiere die Log-Likelihood-Fläche basierend auf dieser Beobachtung
i) als Contour Plot mit geeigneten Beschriftungen
ii) als Fläche im $I\!R^3$
über geeigneten Rechtecken $(\mu_0 - c, \mu_0 + d) \times (\sigma_0 - c', \sigma_0 + d')$. Dabei wähle man jeweils ein kleines Rechteck, das die Maximalstelle enthält und die lokale Struktur der Log-Likelihoodfläche in Umgebungen ihrer Maximalstelle gut sichtbar macht, und ein grosses Rechteck, welches das Verhalten der Log-Likelihoodfläche für Parameterwerte (μ, σ^2) in grösserer Entfernung von (μ_0, σ_0^2) verdeutlicht. (Abgabe: Graphiken, Programm)

<u>Aufgabe 3.2 ('Schöne' Likelihoodflächen II):</u> Das von einer vorgegebenen stetigen Verteilungsfunktion F auf $(I\!R, \mathcal{B}(I\!R))$ erzeugte Lokations- und Skalenmodell ist die Verteilungsfamilie

$$P_{m,c} := \mathcal{L}(cZ + m) \;:\; Z \sim F \,,\; m \in I\!R \,,\; c > 0$$

bzw. das durch unabhängige Versuchswiederholung entstehende statistische Experiment

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

Man betrachte sehr genau für geeignet kleine n (insbesondere n gerade, etwa $n = 4$) die entstehende Gestalt der Log-Likelihood-Flächen und diskutiere die sich daraus ergebenden Konsequenzen für Maximum-Likelihood Schätzer in den folgenden Fällen:
a) Doppelexponentialverteilung F:

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

(die zugehörigen ZV simuliert man als $U(m + cZ_1) + (1 - U)(m - cZ_2)$ für U, Z_1, Z_2 unabhängig, wobei $U \sim \mathcal{B}(1, \frac{1}{2})$ binomialverteilt und Z_1, Z_2 exponentialverteilt mit Parameter 1);

b) F eine Mischung von Doppelexponentialverteilungen:

$$P_{m,c}(dx) \ := \ \frac{1}{4c}\left(e^{-|\frac{x-[m+0.2]}{c}|} + e^{-|\frac{x-[m-0.2]}{c}|}\right) dx \ ;$$

c) Cauchyverteilung F:

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

(Abgabe: Graphiken, Programm)

Aufgabe 3.3: Man überlege sich, dass im Lokations- und Skalenmodell aus Aufgabe 3.2 a) (F Doppelexponentialverteilung)

$$(\widehat{m}_n, \widehat{c}_n) \quad \text{mit} \quad \widehat{m}_n := median(X_1, \ldots, X_n) \ , \ \widehat{c}_n := \frac{1}{n}\sum_{i=1}^{n} |X_i - \widehat{m}_n|$$

ein Maximum Likelihood Schätzer für den unbekannten Parameter ist. (Keine Abgabe)

Aufgabe 3.4: In jedem aus einer stetigen Verteilungsfunktion F erzeugten Lokations- und Skalenmodell kann man aufgrund von n iid Beobachtungen $X_1, \ldots, X_n$ den Lokationsparameter durch

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

und den Skalenparameter durch

$$\widetilde{c}_n := median(Y_1, \ldots, Y_n) \quad \text{wobei} \quad Y_i := |X_i - \widetilde{m}_n|$$

(Medianabstand) schätzen. Nach einer geeigneten Zahl von Simulationsläufen überzeuge man sich anhand eines Histogramms der erzielten Schätzfehler, dass diese Methode im Falle des durch die Cauchyverteilung erzeugten Lokations- und Skalenmodells aus Aufgabe 3.1 c) zu brauchbaren Ergebnissen führt, während der Schätzer $\widehat{c}_n$ aus Aufgabe 3.3 hier völlig ungeeignet ist. Warum ist dies so (und warum könnte man anstelle der Cauchyverteilung genausogut eine symmetrisch stabilen Verteilung mit Parameter $0 < \alpha < 1$ nehmen ...)? (Abgabe: Graphiken, Programm)