Angewandte Stochastik Blatt 4

1. (Ausreißer und Maximum Likelihood Schätzung) Um gelegentliche Ausreißer bei Meßdaten zu berücksichtigen, nimmt man an, daß die Meßwerte $x_1, \ldots, x_n$ mit Wahrscheinlichkeit $9/10$ gemäß $N(\mu, \sigma^2)$ verteilt sind, aber mit Wahrscheinlichkeit $1/10$ Ausreißer mit Verteilung $N(\mu, 1)$ sind.

a) Berechne die Likelihood $L(\mu, \sigma^2; x_1, \ldots, x_n)$ ($\mu \in \mathbb{R},\ \sigma^2 > 0$) im Produktmodell.

b) Zeige, daß es keinen MLE für $\theta := (\mu, \sigma^2)$ gibt, und daß

$$\sup_{\sigma} L(\hat{\mu}, \sigma^2; x_1, \ldots, x_n) = \sup_{\mu,\sigma} L(\mu, \sigma^2; x_1, \ldots, x_n)$$

genau dann gilt, wenn $\hat{\mu}$ einen der Werte $x_1, \ldots, x_n$ annimmt.

2. (Suffiziente Statistiken) Gegeben sei ein diskretes statistisches Modell $(\mathfrak{X}, \mathcal{B}, P_\theta : \theta \in \Theta)$. Zeige, daß eine Statistik $T : \mathfrak{X} \to \mathbb{R}^k$ genau dann suffizient für den Parameter θ ist, wenn für die Dichten gilt:

$$f(x \,|\, \theta) = g(\theta, T(x))h(x),$$

mit geeigneten Funktionen $g : \Theta \times \mathbb{R}^k \to [0, \infty)$, $h : \mathfrak{X} \to [0, \infty)$.

3. (Wright-Fisher-Modell) Ein Gen besitze die beiden Allele A und a. In einer Population von N Lebewesen mit diploidem Chromosomensatz kommt das Gen $2N$ mal vor. Jede Generation bestehe aus N Individuen und entstehe aus der vorhergehenden durch Zufallspaarung: Jedes Gen der Nachkommengeneration "sucht sich" unabhängig von allen anderen ein Eltern-Gen und nimmt dessen Ausprägung an. Sei X_n die Anzahl der $A-$Gene in der $n-$ten Generation. X_n ist offenbar eine Markov-Kette auf $E = \{0, \ldots, 2N\}$. Bestimme die Übergangsmatrix und berechne für beliebiges $x \in E$

$$h_{2N} := P\left[X_n = 2N \,|\, X_0 = x\right],$$

für alle hinreichend großen n.

4. (Maximum-Likelihood-Schätzung von Ausfallszeiten). Wir betrachten die Überprüfung von n Speicherchips aus einer neuen Serienproduktion. Es werden sämtliche n Speicherchips simultan in Betrieb genommen, und dann die sukzessive eintretenden Ausfallszeiten registriert. Es wird dabei angenommen, daß die Lebensdauern der Speicherchips unabhängig und exponentialverteilt mit Parameter $\theta \in (0, \infty)$ sind, d.h. die Lebensdauern sind durch die Dichte

$$f(x) = \theta e^{-\theta x} \cdot 1_{(0,\infty)}(x), \quad x \in \mathbb{R},$$

gegeben. Sei X_i die Zeit des i-ten Ausfalls, es gilt also

$$X_1 \leq \ldots \leq X_n.$$

Zur Schätzung der erwarteten Lebensdauer werden nur die Lebensdauern der $k < n$ zuerst ausgefallenen Chips registriert. Der Stichprobenraum ist daher

$$\mathfrak{X} = \{x = (x_1, \ldots, x_k) \,|\, x_1 \leq x_2 \leq \ldots \leq x_k\} \subset (0, \infty)^k.$$

Zeige, dass die Dichte von $(X_1, \ldots, X_k)$ gegeben ist durch

$$f_\theta(x_1, \ldots, x_k) = \frac{n!}{(n-k)!} \theta^k e^{-\theta(\sum_{i=1}^{k-1} x_i + (n-k+1)x_k)} \ ,$$

und bestimme den Maximum-Likelihood-Schätzer für θ.

5. (Das Ruinproblem) Ein Mann hat sich k Euro für einen Jaguar gespart, der N Euro kostet $(0 < k < N)$. Um den fehlenden Betrag zu gewinnen, läßt er sich auf folgendes Spiel ein: Er wirft wiederholt eine faire Münze. Erscheint Kopf, dann erhält er einen Euro , kommt hingegen Zahl, dann zahlt er einen Euro an die Bank. Er spielt solange bis eine der folgenden zwei Möglichkeiten eintritt: Entweder er verliert das ganze Geld, oder er gewinnt genug, um sich den Jaguar leisten zu können. Sei A das Ereignis, daß der Mann alles verliert, und $p_k = P_k[A]$ die Wahrscheinlichkeit dafür zum Startwert k. Zeige:

$$p_k = \frac{1}{2}(p_{k+1} + p_{k-1}), \quad 0 < k < N,$$

und löse diese Differenzengleichung zu den Randbedingungen $p_0 = 1, p_N = 0$.