Fachbereich Mathematik
Prof. Dr. Michael Kohler
Dipl.-Math. Andreas Fromkorth
Dipl.-Inf. Jens Mehnert

TECHNISCHE UNIVERSITÄT DARMSTADT

SS 09
6.7.2009

10. Übungsblatt zur „Einführung in die Stochastik“

Aufgabe 38 (3 Punkte)

Die Zufallsvariablen $X_1, \dots, X_n$ seien unabhängig identisch auf $[\theta, 2\theta]$ gleichverteilt, d.h. sie sind unabhängig und besitzen (jeweils) eine Dichte $f_\theta : \mathbb{R} \to \mathbb{R}_+$ mit

$$f_\theta(x) = \begin{cases} \frac{1}{\theta} & \text{für} \quad \theta \leq x \leq 2\theta, \\ 0 & \text{für} \quad x \notin [\theta, 2\theta]. \end{cases}$$

Hierbei ist $\theta \in \mathbb{R}_+$ ein Parameter der Dichte f_θ.

(a) Zeigen Sie, dass der Schätzer

$$T_n(X_1, \dots, X_n) = \frac{2}{3 \cdot n} \sum_{i=1}^{n} X_i$$

ein erwartungstreuer Schätzer für θ ist.

(b) Ist der Schätzer in a) auch stark konsistent ? Begründen Sie ihre Antwort.

Lösung: a)

$$\begin{aligned} EX_1 &= \int_{\mathbb{R}} x \cdot f(x) dx \\ &= \int_{\theta}^{2\theta} x \cdot \frac{1}{\theta} dx \\ &= \frac{3}{2}\theta. \end{aligned}$$

$T_n(X_1, \dots, X_n)$ ist erwartungstreuer Schätzer für θ,da

$$\begin{aligned} E_\theta\left[T_n(X_1, \dots, X_n)\right] &= E_\theta\left[\frac{2}{3 \cdot n} \sum_{i=1}^{n} X_i\right] \\ &= \frac{2}{3 \cdot n} E_\theta\left[X_1\right] \\ &= \frac{2}{3} \cdot \frac{3}{2}\theta \\ &= \theta \end{aligned}$$

für alle $\theta > 0$.

b) Der Schätzer ist auch stark konsistent, da nach dem Gesetz der großen Zahlen gilt:

$$T_n(X_1, \dots, X_n) \overset{f.s.}{\to} \frac{2}{3} \cdot E_\theta(X_1) = \theta$$

für alle $\theta > 0$.

Aufgabe 39 (3 Punkte)

Die zufällige Lebensdauer einer Leuchtstoffröhre hängt nicht von der gesamten Brenndauer, sondern nur von der Anzahl der Ein– und Ausschaltvorgänge ab. Die Wahrscheinlichkeit, dass eine Röhre beim k–ten Einschaltvorgang ausfällt, sei $p^{k-1} \cdot (1-p)$ $(k \in \mathbb{N})$, wobei der Parameter $p \in (0,1)$ als Maß für die Güte der Röhre angesehen werden kann.

In einer Glühlampenfabrik wird die Qualität der produzierten Röhren dadurch kontrolliert, dass n Röhren unabhängig voneinander durch Relais ständig ein– und ausgeschaltet werden. Dabei wird registriert, wann die einzelnen Röhren ausfallen. Das Ergebnis dieser Versuche sei $k_1, \ldots, k_n \in \mathbb{N}$, d.h. die i–te Röhre ist beim k_i–ten Einschaltvorgang ausgefallen.

Bestimmen Sie durch Anwendung des Maximum–Likelihood–Prinzips eine Schätzung des Parameters p ausgehend von $k_1, \ldots, k_n$.

Lösung: Wahrscheinlichkeit, dass die Röhre beim k-ten Versuch ausfällt, beträgt $p^{k-1} \cdot (1-p)$

Maximum-Likelihood Schätzer:

$$\widehat{p}(k_1, \ldots, k_n) = \text{argmax}_{p \in (0,1)} P[X_1 = k_1, \ldots, X_n = k_n]$$

Es gilt

$$\begin{aligned} P[X_1 = k_1, \ldots, X_n = k_n] &\overset{\text{Unabhängigkeit}}{=} \prod_{i=1}^{n} P[X_i = k_i] \\ &= \prod_{i=1}^{n} p^{k_i - 1} \cdot (1-p) \\ &= (1-p)^n \cdot p^{(\sum\limits_{i=1}^{n} k_i) - n}, \end{aligned}$$

wobei die Maximierung dieses Ausdrucks äquivalent zur Maximierung von

$$L(p) = n \cdot \log(1-p) + \left(\sum_{i=1}^{n} k_i - n \right) \cdot \log p.$$

Es gilt dann

$$\begin{aligned} L'(p) &= -\frac{n}{1-p} + \frac{\left(\sum\limits_{i=1}^{n} k_i - n \right)}{p} \\ &= \frac{-n \cdot p + \sum\limits_{i=1}^{n} k_i - n - p \cdot \sum\limits_{i=1}^{n} k_i + np}{p(1-p)}, \end{aligned}$$

wobei dieser Ausdruck gerade Null ist, falls

$$\begin{aligned} & \sum_{i=1}^{n} k_i - n - p \sum_{i=1}^{n} k_i \overset{!}{=} 0 \\ \Leftrightarrow \quad & (1-p) \sum_{i=1}^{n} k_i = n \\ \Leftrightarrow \quad & 1 - p = \frac{n}{\sum\limits_{i=1}^{n} k_i} \\ \Leftrightarrow \quad & p = 1 - \frac{n}{\sum\limits_{i=1}^{n} k_i} \end{aligned}$$

Wir erhalten somit als Maximum-Likelihood Schätzer von p

$$\widehat{p}(k_1, \ldots, k_n) = 1 - \frac{n}{\sum\limits_{i=1}^{n} k_i}.$$

Aufgabe 40 (3 Punkte)

Ein Flugunternehmen möchte die zufällige Anzahl X der Personen, die nach Erwerb eines Flugtickets nicht (rechtzeitig) zum Abflug erscheinen, stochastisch modellieren. Nimmt man an, dass bei $n = 240$ verkauften Flugtickets jede einzelne Person, die ein Flugticket erworben hat, unbeeinflusst von den anderen Käufern der Flugtickets mit Wahrscheinlichkeit $p \in [0,1]$ nicht zum Abflug erscheint, so ist die zufällige Zahl X der nicht zum Abflug erscheinenden Personen binomialverteilt mit Parametern $n = 240$ und p, d.h.

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad (k \in \{0, 1, \dots, n\}).$$

Bei den letzten zehn Abflügen sind

$$x_1 = 10, x_2 = 6, x_3 = 15, x_4 = 1, x_5 = 2, x_6 = 5, x_7 = 6, x_8 = 16, x_9 = 11, x_{10} = 3$$

der jeweils $n = 240$ Personen, die ein Flugticket gekauft hatten, nicht zum Abflug erschienen. Konstruieren Sie mit Hilfe des Maximum-Likelihood-Prinzips ausgehend von diesen Daten eine Schätzung von p.

Lösung: X sei $b(240, p)$-vrteilt.

Für den Maximum-Likelihood Schätzer $\widehat{p}$ von p muss gelten

$$\widehat{p} = \mathrm{argmax}_{p \in (0,1)} \prod_{i=1}^{10} P[X_i = x_i].$$

Es gilt

$$\begin{aligned}
\prod_{i=1}^{10} P[X_i = x_i] &= \prod_{i=1}^{10} \tbinom{246}{x_i} p^{x_i} (1-p)^{240-x_i} \\
&= \prod_{i=1}^{10} \tbinom{240}{x_i} \cdot p^{\sum_{i=1}^{10} x_i)} \cdot (1-p)^{(2400 - \sum_{i=1}^{10} x_i} \\
&= \prod_{i=1}^{10} \tbinom{240}{x_i} \cdot p^{75} \cdot (1-p)^{2325}.
\end{aligned}$$

Die Maximierung dieses Ausdrucks ist äquivalent zur Maximierung der Funfktion

$$f(p) = p^{75} \cdot (1-p)^{2325}$$

bzw. wegen der Monotonie der Logarithmusfunktion der Maximierung von:

$$l(p) = 75 \log(p) + 2325 \log(1-p).$$

Es gilt nun

$$\begin{aligned}
&l'(p) = \tfrac{75}{p} + \tfrac{2325}{(1-p)} \cdot (-1) = \tfrac{75 - 75p - 2325p}{p(1-p)} \stackrel{!}{=} 0 \\
&\Leftrightarrow p = \tfrac{75}{2400} = 0,03125
\end{aligned}$$

und

$$l''(3125) \leq 0$$

und somit hat l eine Maximalstelle in $0,03125$. Als ML-Schätzer erhalten wir daher $\widehat{p} = 0,03125$.

Aufgabe 41 (3 Punkte)

Wirtschaftswissenschaftler W. möchte die Dauer von Arbeitslosigkeit stochastisch modellieren. Dazu beschreibt er sie durch eine $\exp(\lambda)$–Verteilung, d.h. durch eine Verteilung, die eine Dichte $f : \mathbb{R} \to \mathbb{R}_+$ besitzt mit

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x} & \text{für} \quad x \geq 0, \\ 0 & \text{für} \quad x < 0. \end{cases}$$

Um den unbekannten Parameter $\lambda > 0$ zu schätzen, lässt er sich vom Arbeitsamt für vier zufällig herausgegriffene Arbeitslose ermittelten, dass diese genau $x_1 = 12$ bzw. $x_2 = 2$ bzw. $x_3 = 18$ bzw. $x_4 = 8$ Monate nach Verlust ihres bisherigen Arbeitsplatzes eine neue Arbeitsstelle gefunden haben.
(a) Konstruieren Sie den Maximum–Likelihood–Schätzer für λ und geben Sie an, was man im Falle der obigen Stichprobe als Schätzung für λ erhält.
(b) Zeigen Sie, dass der Schätzer

$$T_n(X_1, \ldots, X_n) = \frac{1}{\frac{1}{n}\sum_{i=1}^n X_i}$$

ein stark konsistenter Schätzer für λ ist.

Lösung: $X_1, \ldots, X_n$ $\exp(\lambda)$-verteilt, d. h. sie haben die Dichte

$$f_\lambda(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & , x \geq 0 \\ 0 & , x < 0 \end{cases}$$

a) Für den Maximum-Likelihood Schätzer von λ muss gelten, dass er das Produckt der Dichten bei gegebenen Schätzwerten maximiert. In unserem Fall heißt es, dass er definiert ist als

$$\begin{aligned} &\text{argmax}_{\lambda>0} \prod_{i=1}^n f_\lambda(x_i) \\ &= \text{argmax}_{\lambda>0} \prod_{i=n}^n \lambda \cdot e^{-\lambda x_i} \cdot 1_{[0,\infty)}(x_i). \\ &= \text{argmax}_{\lambda>0} \lambda^n \cdot e^{-\lambda \cdot \sum\limits_{i=1}^n x_i} \cdot 1_{[0,\infty)^n}(x_1, \ldots, x_n). \end{aligned}$$

Dieser Ausdruck wird genau dann maximal, wenn $x_1, \ldots, x_n > 0$ und

$$\log\left(\lambda^n \cdot e^{-\lambda \cdot \sum\limits_{i=1}^n x_i}\right) := L(\lambda)$$

maximal wird. Es gilt nun

$$L(\lambda) = n \cdot \log(\lambda) - \lambda \cdot \sum_{i=1}^n x_i$$

und somit

$$L'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^n x_i.$$

Weiterhin gilt

$$\frac{n}{\lambda} - \sum_{i=1}^n x_i \stackrel{!}{=} 0 \Leftrightarrow \lambda = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^n x_i}$$

und da $L''(\lambda) < 0$ gilt für alle $\lambda > 0$ erhalten wir als ML-Schätzer

$$\widehat{\lambda} = \frac{1}{\frac{1}{n}\sum\limits_{i=1}^n x_i},$$

also in unserem Fall

$$\widehat{\lambda} = \frac{1}{\frac{1}{4} \cdot 40} = \frac{1}{10}.$$

b) Nach dem starken Gesetz der großen Zahlen gilt:

$$P\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = E(X_i)\right] = 1.$$

In unserem Fall sind die X_i unabhängig identisch $\exp(\lambda)$-verteilt mit Erwartungswert $\frac{1}{\lambda}$ und somit gilt

$$P\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i = \frac{1}{\lambda}\right] = 1$$

und daher auch

$$P\left[\lim_{n\to\infty}\frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n}X_i} = \lambda\right] = 1.$$

Also ist $T_n(X_1,\ldots,X_n)$ stark konsistenter Schätzer für λ.