Einführung in die Stochastik
11. Übungsblatt

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
M. Kohler
A. Fromkorth
D. Furer

SS 2011
08.07.2011

Gruppen und Hausübung

Aufgabe 41
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:

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

Es gilt

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

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

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

wobei dieser Ausdruck gerade Null ist, falls

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

Wir erhalten somit als Maximum-Likelihood Schätzer von p

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

Aufgabe 42
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 $n = 555$ 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 $n = 555$ und p, d.h.

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

Bei den letzten zehn Abflügen sind

$$x_1 = 20, x_2 = 11, x_3 = 28, x_4 = 3, x_5 = 3, x_6 = 7, x_7 = 13, x_8 = 17, x_9 = 22, x_{10} = 5$$

der jeweils $n = 555$ 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 $b(555,p)$-vrteilt.

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

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

Es gilt

$$\begin{aligned}
\prod_{i=1}^{10} P[X_i = x_i] &= \prod_{i=1}^{10} \tbinom{555}{x_i} p^{x_i}(1-p)^{555-x_i} \\
&= \prod_{i=1}^{10} \tbinom{555}{x_i} \cdot p^{\sum\limits_{i=1}^{10} x_i} \cdot (1-p)^{(5550-\sum\limits_{i=1}^{10} x_i)} \\
&= \prod_{i=1}^{10} \tbinom{555}{x_i} \cdot p^{129} \cdot (1-p)^{5421}.
\end{aligned}$$

Die Maximierung dieses Ausdrucks ist äquivalent zur Maximierung der Funfktion

$$f(p) = p^{129} \cdot (1-p)^{5421}$$

bzw. wegen der Monotonie der Logarithmusfunktion der Maximierung von:

$$l(p) = 129\log(p) + 5421\log(1-p).$$

Es gilt nun

$$\begin{aligned}
&l'(p) = \tfrac{129}{p} + \tfrac{5421}{(1-p)} \cdot (-1) = \tfrac{129-129p-5421p}{p(1-p)} \stackrel{!}{=} 0 \\
&\Leftrightarrow p = \tfrac{129}{5550} = 0,02324
\end{aligned}$$

und

$$l''(0,02324) \leq 0$$

und somit hat l eine Maximalstelle in $0,02324$. Als ML-Schätzer erhalten wir daher $\widehat{p} = 0,02324$.

Aufgabe 43
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 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

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

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(λ)-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 $T_n(X_1,\ldots,X_n)$ konsistenter Schätzer für λ.

Aufgabe 44

Student S. vermutet, dass die zufällige Zeit (in Minuten), die Dozent K. bei seiner Statistik Vorlesung immer zu früh kommt, durch eine stetig verteilte Zufallsvariable X mit Dichte

$$f_\alpha(x) = \begin{cases} \alpha & \text{für} \quad 0 \le x \le 10, \\ \frac{1}{10} - \alpha & \text{für} \quad 10 < x \le 20, \\ 0 & \text{für} \quad x < 0 \text{ oder } x > 20 \end{cases}$$

beschrieben werden kann. Um den Parameter $\alpha \in \left[0, \frac{1}{10}\right]$ der Dichte von X zu schätzen, notiert sich Student S., dass Dozent K. bei den letzten $n = 5$ Vorlesungen

$$x_1 = 10 \text{ bzw. } x_2 = 3 \text{ bzw. } x_3 = 4 \text{ bzw. } x_4 = 1 \text{ bzw. } x_5 = 15$$

Minuten zu früh kam.

a) Bestimmen Sie die Likelihood-Funktion

$$L(\alpha) = \prod_{i=1}^{5} f_\alpha(x_i), \quad (\alpha \in [0, \frac{1}{10}]).$$

b) Bestimmen Sie – ausgehend von den angegebenen Werten von $x_1, \ldots, x_5$ – die zugehörige Maximum-Likelihood-Schätzung von α.

c) Seien $X_1, \ldots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte f_α. Zeigen Sie: Der Schätzer

$$T_n(X_1,\ldots,X_n) = \frac{1}{10\cdot n}\sum_{i=1}^{n} 1_{[0,10]}(X_i)$$

mit

$$1_{[0,10]}(x) = \begin{cases} 1 & \text{für} \quad 0 \le x \le 10, \\ 0 & \text{für} \quad x < 0 \text{ oder } x > 10, \end{cases}$$

ist ein erwartungstreuer Schätzer für α.

d) Ist der Schätzer in c) auch konsistent ? Begründen Sie ihre Antwort.

Lösung: a) Es gilt

$$
\begin{aligned}
L(\alpha; x_1, \ldots, x_5) &= \prod_{i=1}^{5} f_\alpha(x_i) \\
&= \alpha^{\#\{x_i : 0 \le x_i \le 10, i \in \{1,\ldots,10\}\}} \cdot \left(\frac{1}{10} - \alpha\right)^{\#\{x_i : 10 < x_i \le 20, i \in \{1,\ldots,10\}\}}
\end{aligned}
$$

b) Es gilt

$$
\begin{aligned}
L(\alpha; 10, 3, 4, 1, 15) &= f_\alpha(10) \cdot f_\alpha(3) \cdot f_\alpha(4) \cdot f_\alpha(1) \cdot f_\alpha(15) \\
&= \alpha^4 \cdot \left(\frac{1}{10} - \alpha\right) \\
&= -\alpha^5 + \frac{1}{10}\alpha^4
\end{aligned}
$$

Es gilt nun

$$
L'(\alpha; 10, 3, 4, 1, 15) = -5\alpha^4 + \frac{2}{5} \cdot \alpha^3
$$

und

$$
L''(\alpha; 10, 3, 4, 1, 15) = -20\alpha^3 + \frac{6}{5}\alpha^2.
$$

Nulltellen der ersten Ableitung sind, 0 (dreifach) und $\frac{2}{25}$. Das gesuchte Maximum kann also nur $\alpha = \frac{2}{25}$ (sonst wäre $L(\alpha; 10, 3, 4, 1, 15)$=0). Da die Randwerte ebenfalls eine 0 liefern und $L'' < 0$ für $\alpha = \frac{2}{25}$ ist gilt dies damit auch.

c) Es gilt

$$
\begin{aligned}
E\left[T_n\left(X_1, \ldots, X_n\right)\right] &= \frac{1}{10 \cdot n} \sum_{i=1}^{n} E\left[1_{[0,10]}\left(X_i\right)\right] \\
&= \frac{1}{10 \cdot n} \sum_{i=1}^{n} 1 \cdot P\left[1_{[0,10]}\left(X_i\right) = 1\right] \\
&= \frac{1}{10 \cdot n} \sum_{i=1}^{n} 10 \cdot \alpha \\
&= \alpha
\end{aligned}
$$

d) Es gilt nach dem starken Gesetz der großen Zahlen

$$
T_n\left(X_1, \ldots, X_n\right) = \frac{1}{10 \cdot n} \sum_{i=1}^{n} 1_{[0,10]}\left(X_n\right) \rightarrow E\left[1_{[0,10]}\left(X_1\right)\right] = \alpha
$$

fast sicher, womit der Schätzer auch konsistent ist.

Dieses Übungsblatt wird im Rahmen der Übungen am 11. bzw. 12.07.2011 besprochen.