Fachbereich Mathematik & Informatik
Freie Universität Berlin
Carsten Hartmann, Stefanie Winkelmann

Musterlösung für den 10. Übung zur Vorlesung

Stochastik I

Wintersemester 2011/2012

1. Aufgabe (Parameterschätzung I, 4 Punkte)
In einer Urne sind $N \geq 2$ Lose mit den Nummern $1, ..., N$, wobei N unbekannt ist. Es wird $n \leq N$ mal ohne Zurücklegen gezogen mit den Ergebnissen $x_1, ..., x_n$.

a) Bestimmen Sie den Maximum-Likelihood-Schätzer T für N.

b) Berechnen Sie $\mathbb{E}_N(T)$ und konstruieren Sie daraus einen erwartungstreuen Schätzer T^* für N.
Tipp: Sie können Aufgabe 3 vom 5. Übungszettel und die Formel $\sum_{k=n}^{N} \binom{k}{n} = \binom{N+1}{n+1}$ *verwenden.*

Lösung:
a) Ein Schätzer T ist definitionsgemäß Maximum-Likelihood-Schätzer für N, falls

$$\mathbb{P}_{T(x)}(X = x) = \max_N \mathbb{P}_N(X = x),$$

wobei $\mathbb{P}_N$ das Wahrscheinlichkeitsmaß in Abhängigkeit des unbekannten Parameters N ist und $x = (x_1, ..., x_n)$ das Ergebnis der Stichprobe. Für festes N gibt es $\binom{N}{n}$ mögliche gleichwahrscheinliche Ergebnisse, so dass

$$\mathbb{P}_N(X = x) = \frac{1}{\binom{N}{n}} = \frac{n!}{N \cdot (N-1) \cdot ... \cdot (N-n+1)}.$$

Dieser Bruch wird umso größer, je kleiner N ist. N ist aber mindestens so groß wie $\max(x_1, ..., x_n)$. Daher ist $T(x) = \max(x_1, ..., x_n)$ der Maximum-Likelihood-Schätzer.
b) Betrachte den Schätzer $T(x) = \max(x_1, ..., x_n)$ aus a). Nach Aufgabe 3 vom 5. Übungszettel gilt $\mathbb{E}_N(T) = \sum_{i=1}^{\infty} \mathbb{P}_N(T \geq i)$. Da $\mathbb{P}_N(T \geq i) = 0$ für $i > N$ folgt

$$\mathbb{E}_N(T) = \sum_{i=1}^{N} \mathbb{P}_N(T \geq i).$$

Dies lässt sich weiter umschreiben zu

$$\mathbb{E}_N(T) = \sum_{i=1}^{N} (1 - \mathbb{P}_N(T < i)) = N - \sum_{i=1}^{N} \mathbb{P}_N(T < i).$$

Es gilt

$$\mathbb{P}_N(T < i) = \begin{cases} 0 & : \quad i \leq n \\ \frac{\binom{i-1}{n}}{\binom{N}{n}} & : \quad n < i \leq N \end{cases},$$

und somit

$$\mathbb{E}_N(T) = N - \sum_{i=n+1}^{N} \frac{\binom{i-1}{n}}{\binom{N}{n}} = N - \frac{1}{\binom{N}{n}} \sum_{i=n}^{N-1} \binom{i}{n} = N - \frac{1}{\binom{N}{n}} \binom{N}{n+1},$$

wobei im letzten Schritt der Tipp verwendet wurde. Lösen wir die Binomialkoeffizienten auf, erhalten wir

$$\mathbb{E}_N(T) = N - \frac{N-n}{n+1} = \frac{n}{n+1} \cdot (N+1)$$

und somit $\mathbb{E}_N(T) \neq N$ (für $N \neq n$), d.h. T ist nicht erwartungstreu. Um einen erwartungstreuen Schätzer zu konstruieren, nutzen wir die Linearität des Erwartungswertes und definieren $T^* = \frac{n+1}{n} \cdot T - 1$, denn dann gilt

$$\mathbb{E}_N(T^*) = \frac{n+1}{n} \cdot \mathbb{E}_N(T) - 1 = N.$$

2. Aufgabe (Parameterschätzung II, 4 Punkte)
Anhand von n unabhängigen Stichproben $x_1, ..., x_n$ soll die Varianz einer Zufallsvariable mit unbekannter Verteilung und unbekanntem Erwartungswert geschätzt werden. Vorausgesetzt wird nur, dass Erwartungswert und Varianz existieren. Konstruieren Sie einen erwartungstreuen Schätzer für die Varianz.

Lösung:
Wir schätzen den Erwartungswert mit dem erwartungstreuen Schätzer $\bar{X} = \frac{1}{n}(X_1 + ... + X_n)$. Die Varianz ist die mittlere quadratische Abweichung vom Erwartungswert. Da liegt es nahe, die Varianz durch

$$T(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

zu schätzen - das ist die mittlere quadratische Abweichung vom geschätzten Erwartungswert. Jedoch ist

$$\begin{aligned}
\mathbb{E}(T(X)) &= \mathbb{E}\left(\frac{1}{n}\sum_i (X_i - \bar{X})^2\right) \\
&= \frac{1}{n}\sum_i (\mathbb{E}(X_i^2) - 2\mathbb{E}(X_i\bar{X}) + \mathbb{E}(\bar{X}^2)) \\
&= \frac{1}{n}\sum_i \mathbb{E}(X_i^2) - 2\frac{1}{n^2}\sum_i\sum_j \mathbb{E}(X_iX_j) + \frac{1}{n^2}\sum_i\sum_j \mathbb{E}(X_iX_j) \\
&= \frac{1}{n}\sum_i \mathbb{E}(X_i^2) - \frac{1}{n^2}\sum_i\sum_j \mathbb{E}(X_iX_j) \\
&= \frac{1}{n}\sum_i \mathbb{E}(X_i^2) - \frac{1}{n^2}\sum_i \mathbb{E}(X_i^2) - \frac{1}{n^2}\sum_i\sum_{j\neq i} \mathbb{E}(X_iX_j) \\
&= \frac{1}{n}n\mathbb{E}(X^2) - \frac{1}{n^2}n\mathbb{E}(X^2) - \frac{1}{n^2}n(n-1)\mathbb{E}(X)^2 \\
&= \frac{n-1}{n}\mathbb{E}(X^2) - \frac{n-1}{n}\mathbb{E}(X)^2 \\
&= \frac{n-1}{n}\mathrm{Var}(X),
\end{aligned}$$

d.h., der Schätzer T ist nicht erwartungstreu. Um einen erwartungstreuen Schätzer T^* für die Varianz zu erhalten, müssen wir den Vorfaktor $\frac{n-1}{n}$ loswerden. Dies machen wir durch die Wahl von

$$T^*(X) = \frac{n}{n-1}T(X) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Dann ist

$$\mathbb{E}(T^*(X)) = \mathbb{E}\left(\frac{n}{n-1}T(X)\right) = \frac{n}{n-1}\mathbb{E}(T(X)) = \mathrm{Var}(X),$$

und T^* ist erwartungstreu.

3. Aufgabe (Konfidenzintervall, 4 Punkte)
Bei einer Prüfung mit Notenvergabe $1, 2, 3, 4, 5$ gibt es eine gewisse (unbekannte) Verteilung $\mathbb{P}(X = i) = p_i$, $i = 1, ..., 5$. Es soll $\mathbb{E}(X)$ geschätzt werden, wobei bekannt ist, dass $\mathrm{Var}(X) \leq 0.5$ gilt. Bei einer Prüfung von 25 unabhängigen Kandidaten entsteht folgender Notenspiegel:

1	2	3	4	5
2	7	4	6	6

Konstruieren Sie mithilfe der Tschebyschev-Ungleichung ein Konfidenzintervall für $\mathbb{E}(X)$ um den Mittelwert $\bar{x}$ zum Niveau $\alpha = 0.25$. Ist der Wert $\mathbb{E}(X) = 3$ noch plausibel?

Lösung:
Es ist $\bar{x} = \frac{1}{n}(x_1 + ... + x_n) = 3.28$. Das gesuchte Konfidenzintervall hat die Form $C(\bar{x}) = (\bar{x} - \varepsilon, \bar{x} + \varepsilon)$, wobei ε bestimmt werden soll. Der unbekannte Parameter ist $\nu = \mathbb{E}(X)$, und es soll gelten

$$\mathbb{P}_\nu(\nu \in C(\bar{x})) = \mathbb{P}_\nu(|\bar{x} - \nu| < \varepsilon) = 1 - \mathbb{P}_\nu(|\bar{x} - \nu| \geq \varepsilon) \geq 1 - \alpha = 0.75,$$

d.h. $\mathbb{P}_\nu(|\bar{x} - \nu| \geq \varepsilon) \leq 0.25$. Mit Tschebyschev gilt

$$\mathbb{P}_\nu(|\bar{x} - \nu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\bar{x})}{\varepsilon^2} = \frac{\mathrm{Var}(X)}{n\varepsilon^2} \leq \frac{0.5}{n\varepsilon^2}.$$

Ein $\varepsilon > 0$ funktioniert also, sobald

$$\frac{0.5}{n\varepsilon^2} \leq 0.25 \Leftrightarrow \varepsilon^2 \geq \frac{0.5}{0.25 \cdot n} = \frac{0.5}{0.25 \cdot 25} = 0.08$$

d.h. $\varepsilon \geq 0.2828$. Der Wert $\mathbb{E}(X) = 3$ ist damit gerade noch plausibel, d.h. er liegt in $C(\bar{x})$. (Die Bedingung $\varepsilon \geq 0.2828$ ist hinreichend, aber nicht unbedingt notwendig: Es kann sein, dass auch noch kleinere ε möglich sind. Dies liegt daran, dass die Tschebyschev-Ungleichung nur eine Abschätzung nach oben liefert.)

4. Aufgabe (Hypothesentest, 4 Punkte)
Es werden n unabhängige Stichproben $x_1, ..., x_n$ aus der Gleichverteilung auf dem Intervall $[0, \nu]$ ($\nu > 0$) gezogen. Dabei soll der Test ϕ mit H_0: $\nu = 1$ gegen H_1: $\nu \neq 1$ durchgeführt werden, wobei der Annahmebereich als $A = \{(x_1, ..., x_n) : \frac{1}{2} < \max(x_1, ..., x_n) \leq 1\}$ festgelegt wird. Stellen Sie die Gütefunktion $G_\phi(\nu)$ auf. Welches Niveau hat der Test?

Lösung:
Der Ablehnungsbereich ist $R = \mathbb{R}^+ \setminus A = [0, \frac{1}{2}] \cup (1, \infty)$. Die zugehörige Testregel ist

$$\varphi(x) = \begin{cases} 0 & : \quad x \in (\frac{1}{2}, 1] \\ 1 & : \quad \text{sonst} \end{cases},$$

wobei $x = \max(x_1, ..., x_n)$. Da es sich um einen deterministischen Test handelt, gilt für die Gütefunktion $G_\varphi(\nu) = \mathbb{E}_\nu(\varphi) = \mathbb{P}_\nu(x \in R)$, d.h.

$$\begin{aligned} G_\varphi(\nu) &= \mathbb{P}_\nu(x \in R) \\ &= \mathbb{P}_\nu(x \leq 0.5) + \mathbb{P}(x > 1) \\ &= \begin{cases} 1 & : \quad \nu \leq 0.5 \\ \left(\frac{0.5}{\nu}\right)^n & : \quad 0.5 < \nu \leq 1 \\ \left(\frac{0.5}{\nu}\right)^n + \left(\frac{\nu-1}{\nu}\right)^n & : \quad \nu > 1 \end{cases}. \end{aligned}$$

Demnach ist $G_\varphi(1) = \left(\frac{1}{2}\right)^n$, und dies ist die Wahrscheinlichkeit für einen Fehler erster Art und damit auch das Niveau des Testes.