④ Übungen zu Grundlagen der induktiven Statistik

Zoltán Zomotor

Versionsstand: 13. Mai 2015, 10:11

Aufgabe 1: Reviewfragen

(a) Was ist der Unterschied zwische deskriptiver und induktiver Statistik?

(b) Was versteht man unter einer *einfachen Stichprobe*?

(c) Welchen Wert nimmt die Varianz des Stichprobenmittels $\mathrm{Var}(\bar{X})$ für eine einfache Stichprobe $X_1, \ldots, X_n$ mit beliebiger Verteilung und $\mathrm{Var}(X_i) = \sigma^2$ an, wenn $n \to \infty$?

(d) Was ist die Likelihoodfunktion $f(x_1, \ldots, x_n|\vartheta)$?

(e) Wie wird die Likelihoodfunktion $f(x_1, \ldots, x_n|\vartheta)$ bei einer einfachen Stichprobe $X_1, \ldots, X_n$ bestimmt? Was bedeutet hier ϑ? Geben Sie ein Beispiel für ϑ.

(f) Unter welcher Voraussetzung für die einfache Stichprobe $X_1, \ldots, X_n$ gilt für ihre Stichprobenvarianz

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)?$$

Welchen Wert haben unter dieser Voraussetzung Erwartungswert und Varianz $\mathrm{E}(T)$ und $\mathrm{Var}(T)$ mit

$$T = \frac{\bar{X} - \mu}{S} \sqrt{n}?$$

(g) Wie hängen Erwartungswert der Stichproben-Standardabweichung $\mathrm{E}(S)$ und die Standardabweichung σ zusammen?

Aufgabe 2: Likelihoodfunktion

Sie machen eine Blitzumfrage in Ihrem Umfeld, ob jemand raucht oder nicht. Sie bekommen folgende Antworten: *Raucher, Nichtraucher, Raucher, Nichtraucher, Nichtraucher.* Fassen Sie die Blitzumfrage als einfache Zufallsstichprobe „mit Zurücklegen“ auf. Bestimmen Sie die Likelihoodfunktion $f(R, \bar{R}, R, \bar{R}, \bar{R}|p)$, wobei p der Anteil der Raucher ist.

Aufgabe 3: Likelihoodfunktion

Gegeben sei eine einfache Stichprobe $X_1, \ldots, X_n$ mit

(a) $X_i \sim Exp(\lambda)$ (b) $X_i \sim P(\mu)$ (c) $X_i \sim B(1, p)$

Geben Sie jeweils die Likelihoodfunktion $f(x_1, \ldots, x_n|\vartheta)$ in kompakter Form an. Verwenden Sie falls möglich das Produktsymbol $\prod$ und/oder Summensymbol $\sum$.

Aufgabe 4: Erwartungswert und Varianz von Stichprobenfunktionen

Leiten Sie für die einfache Stichprobe $X_1, \ldots, X_n$ mit $\mathrm{E}(X_i) = \mu$ und $\mathrm{Var}(X_i) = \sigma^2$ jeweils den Erwartungswert und die Varianz folgender Stichprobenfunktionen her:

(a) $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ (b) $Y = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$

Aufgabe 5: Fraktile

X_i (mit $i = 1, \ldots, n$) seien unabhängige, jeweils $N(\mu, \sigma)$-verteilte Zufallsvariable. Geben Sie für jede der folgenden Größen die Zahl $x_{0.95}$ an, die mit 5 % Wahrscheinlichkeit überschritten wird (zum Beispiel für (a): $P(x_{0.95} < A) = 0.05$).

(a) $A = \frac{1}{\sigma^2} \sum_{i=1}^{7} (X_i - \bar{X})^2$

(b) $B = \frac{1}{\sigma^2} \sum_{i=1}^{137} (X_i - \bar{X})^2$

(c) $C = \frac{1}{\sigma^2} \sum_{i=1}^{12} (X_i - \mu)^2$

(d) $D = \frac{\sum_{i=1}^{16} X_i - 16\mu}{4S}$ (mit S^2: Stichprobenvarianz von X)

(e) $E = \frac{\sum_{i=1}^{1600} X_i - 1600\mu}{40S}$

(f) $F = \sum_{i=1}^{9} \frac{X_i - \mu}{3\sigma}$

Lösung 1:

(a)
- deskriptive (auch beschreibende oder empirische) Statistik: Vorliegende Daten in geeigneter Weise beschreiben, aufbereiten und zusammenfassen, verdichten quantitativer Daten zu Tabellen, graphischen Darstellungen und Kennzahlen.
- induktive (auch mathematische oder schließende) Statistik: Aus den Daten einer Stichprobe Eigenschaften der Grundgesamtheit ableiten.

(b) Alle Stichprobenvariablen $X_1, \ldots, X_n$ sind i.i.d. (=unabhängig und identisch verteilt)

(c) $\mathrm{Var}(\bar{X}) = 0$ für $n \to 0$

(d) Gemeinsame Dichte / Wahrscheinlichkeitsfunktion von $X_1, \ldots, X_n$

(e) $f(x_1, \ldots, x_n|\vartheta) = f(x_1) \cdot f(x_2) \cdots f(x_n)$. ϑ ist der, oder die, Parameter einer Dichte- oder Wahrscheinlichkeitsfunktion, zum Beispiel μ und σ der $N(\mu, \sigma)$-Verteilung.

(f) $X_i \sim N(\mu, \sigma)$, $\mathrm{E}(T) = 0$, $\mathrm{Var}(T) = \dfrac{n-1}{n-3}$

(g) $\mathrm{E}(S) \leq \sigma$

Lösung 2:

$f(R, \bar{R}, R, \bar{R}, \bar{R}|p) = f(R) \cdot f(\bar{R}) \cdot f(R) \cdot f(\bar{R}) \cdot f(\bar{R}) = p^2(1-p)^3$

Lösung 3:

(a) $f(x_1, \ldots, x_n|\lambda) = \lambda^n \mathrm{e}^{-\lambda \sum_{i=1}^n x_i} = \lambda^n \mathrm{e}^{-\lambda n \bar{x}}$

(b) $f(x_1, \ldots, x_n|\mu) = \dfrac{\mu^{\sum_{i=1}^n x_i} \mathrm{e}^{-n\mu}}{\prod_{i=1}^n x_i!} = \dfrac{\mu^{n\bar{x}} \mathrm{e}^{-n\mu}}{\prod_{i=1}^n x_i!}$

(c) $f(x_1, \ldots, x_n|p) = p^{\sum_{i=1}^n x_i}(1-p)^{n - \sum_{i=1}^n x_i} = p^{n\bar{x}}(1-p)^{n\,(1-\bar{x})}$

Lösung 4:

(a)

$$\mathrm{E}(\bar{X}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}\mu\sum_{i=1}^n 1 = \mu$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \sum_{i=1}^n \frac{1}{n^2}\,\mathrm{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{n}{n^2}\sigma^2 = \frac{1}{n}\sigma^2$$

(b)

$$\mathrm{E}(Y) = \mathrm{E}(\frac{\bar{X}-\mu}{\sigma}\sqrt{n}) = \frac{\mathrm{E}(\bar{X})-\mu}{\sigma}\sqrt{n} = \frac{\mu-\mu}{\sigma}\sqrt{n} = 0$$

$$\mathrm{Var}(Y) = \mathrm{Var}(\frac{\bar{X}-\mu}{\sigma}\sqrt{n}) = \frac{n}{\sigma^2}\,\mathrm{Var}(\bar{X}) = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1$$

Lösung 5:

(a) $A \sim \chi^2(n-1) = \chi^2(6)$

$$P(A < x_{0.95}) = F(x_{0.95}) = 0.95$$
Tabelle 1 im Skript: $x_{0.95} = 12.59$

(b) $B \sim \chi^2(n-1)$, $n > 30 \Rightarrow$ Näherung mit $\tilde{x}_{0.95}$ aus $N(0,1)$–Verteilung:

$$\begin{aligned}
\tilde{x}_{0.95} &= 1.64 + 0.01 \frac{0.95 - 0.9495}{0.9505 - 0.9495} \\
&= 1.645 \\
x_{0.95} &= \frac{1}{2}(\tilde{x}_{0.95} + \sqrt{2 \cdot 136 - 1})^2 \\
&= 164
\end{aligned}$$

(c) $C \sim \chi^2(n) = \chi^2(12)$, $x_{0.95} = 21.03$

(d) $D \sim t(n-1) = t(15)$, $x_{0.95} = 1.753$

(e) $E \sim t(1599) \approx N(0,1)$, $x_{0.95} = 1.645$,

(f) $F \sim N(0,1)$, $x_{0.95} = 1.645$