Einführung in die Stochastik
Übungsblatt 13

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Dr. Robert Schlicht
Dr. Mehdi Slassi
Dipl. Math. Andreas Fromkorth

Sommersemester 2010
9. Juli 2010

Approximative Konfidenzintervalle

In der Statistik wird häufig durch die Normalverteilung approximiert. Ein Standardverfahren zur Anwendung auf die Binomialverteilung sieht wie folgt aus: $X_1, \dots, X_n$ seien unabhängig $\{0,1\}$-wertig mit $P_\theta(X_i = 1) = \theta$ für alle $\theta \in \Theta = (0,1)$, d. h. die Summe $X_1 + \dots + X_n$ ist unter P_θ $B_{n,\theta}$-verteilt. Setze $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Für $c \in \mathbb{R}$ und jedes $\theta \in \Theta$ gilt dann

$$P_\theta\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \theta}{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}} \le c\Big) \xrightarrow[n\to\infty]{} \Phi(c)$$

(Φ Verteilungsfunktion von $N(0,1)$), denn für alle $\epsilon > 0$ gilt

$$P_\theta\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \theta}{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}} \le c\Big) = P_\theta\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \theta}{\sqrt{\theta(1-\theta)}} \le c\frac{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}}{\sqrt{\theta(1-\theta)}}\Big)$$
$$\begin{cases} \le P_\theta\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i-\theta}{\sqrt{\theta(1-\theta)}} \le c+\epsilon\Big) + P_\theta\Big(c\frac{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}}{\sqrt{\theta(1-\theta)}} > c+\epsilon\Big) \to \Phi(c+\epsilon) + 0 \\ \ge P_\theta\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i-\theta}{\sqrt{\theta(1-\theta)}} \le c-\epsilon\Big) - P_\theta\Big(c\frac{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}}{\sqrt{\theta(1-\theta)}} < c-\epsilon\Big) \to \Phi(c-\epsilon) - 0 \end{cases}$$

wegen zentralem Grenzwertsatz und $\hat{\theta}_n \to \theta$ P_θ-f. s.

Ist n hinreichend groß und $X = X_1 + \dots + X_n$, so ist also

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i - \theta}{\sqrt{\hat{\theta}_n(1-\hat{\theta}_n)}} = \sqrt{n}\frac{\frac{1}{n}X - \theta}{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}}$$

approximativ $N(0,1)$-verteilt.

Nun sei $1-\alpha$ ein vorgegebenes Konfidenzniveau und c so, dass $N(0,1)([-c,c]) = 1-\alpha$. Dann gilt also

$$P_\theta\Big(\sqrt{n}\frac{\frac{1}{n}X - \theta}{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}} \in [-c,c]\Big) \approx N(0,1)([-c,c]) \ge 1-\alpha\,.$$

Auflösen der Bedingung in P_θ nach θ führt zu einem **approximativen Konfidenzintervall** zum Niveau $1-\alpha$,

$$\Big[\frac{1}{n}X - c\cdot\frac{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}}{\sqrt{n}},\ \frac{1}{n}X + c\cdot\frac{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}}{\sqrt{n}}\Big]$$

mit $c = \Phi^{-1}(1-\frac{\alpha}{2})$.

Die Verwendung solcher Konfidenzintervalle ist, obwohl eine Standardvorgehensweise, aus theoretischer Sicht problematisch, weil die Normalapproximation für θ nah bei 0 oder 1 bekanntermaßen zu Fehlern führt. Für $n = 1000$ und $\theta = 0.2\%$ kann man z. B. nachrechnen

$$N(0,1)([-1.96, 1.96]) \ge 0.95\,,$$

aber

$$P_\theta\Big(\sqrt{n}\frac{\frac{1}{n}X - \theta}{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}} \in [-1.96, 1.96]\Big) \le P_\theta(0.5 < X < 7.3) = B_{n,\theta}(\{1,\dots,7\}) < 0.87\,.$$

Das Konfidenzniveau wird also nicht einmal approximativ eingehalten. Ein approximatives Konfidenzintervall kann also allenfalls als eine deskriptive Größe dienen, wenn man nicht über die Aussage $\theta \in (0,1)$ hinausgehendes Vorwissen über θ hat.

Aufgabe 13.1 — 4 Punkte

Bei der Shell Jugendstudie 2006 wurden 1231 Mädchen befragt. Dabei gaben 55 Prozent der befragten Mädchen an, das Abitur anzustreben. Leiten Sie aus dem zentralen Grenzwertsatz ein approximatives Konfidenzintervall $[u, 1]$ zum Konfidenzniveau 0.95 für den Anteil der Mädchen her, die das Abitur anstreben. (Hinweis: Die Verteilungsfunktion von $N(0,1)$ nimmt den Wert 0.95 an der Stelle 1.645 an.)

Aufgabe 13.2 — 4 Punkte

Die Zufallsvariablen $X_1, ..., X_n$ seien unabhängig und identisch verteilt mit der Verteilungsfunktion

$$F_\theta(x) = \begin{cases} 1 - \theta^2/x^2, & \text{falls } x \geq \theta, \\ 0, & \text{sonst.} \end{cases}$$

wobei der Parameter $\theta > 0$ unbekannt ist.
Bestimmen Sie den Maximum-Likelihood-Schätzer T_n für $\tau(\theta) = \theta$.

Aufgabe 13.3 — 4 Punkte

Bei der Größenmessung in einer Gruppe von 9 Personen ergaben sich folgende Körpergrößen [in cm]:

184.2, 182.6, 185.3, 184.5, 186.2, 183.9, 185.0, 187.1, 184.4.

Diese Messwerte werden als Realisationen der Zufallsvariablen $X_1, \ldots, X_9$ angenommen, die unabhängig und identisch $N(\mu, \sigma^2)$-verteilt seien.

(a) Geben Sie ein Konfidenzintervall zum Konfidenzniveau 0.99 für den Erwartungswert μ an, falls die Standardabweichung bekannt ist und $\sigma = 2.4$ [cm] beträgt. (Hinweis: Die Verteilungsfunktion von $N(0,1)$ nimmt den Wert 0.995 an der Stelle 2.576 an.)

(b) Welches Konfidenzintervall ergibt sich in (a) für dasselbe Konfidenzniveau, falls die Standardabweichung als unbekannt angenommen wird? (Hinweis: Die Verteilungsfunktion der t-Verteilung mit 8 Freiheitsgraden nimmt den Wert 0.995 an der Stelle 3.355 an.)

Aufgabe 13.4 — 4 Punkte

Ein Autohersteller behauptet, dass der Benzinverbrauch für einen neuentwickelten Typ im Mittel 6 Liter pro 100km beträgt. Dabei kann er davon ausgehen, dass der Verbrauch normalverteilt ist mit der Standardabweichung $\sigma = 0.3$.

(a) Eine Verbraucherzentrale vermutet, dass der Hersteller einen zu niedrigen Mittelwert μ angegeben hat. Sie überprüft deshalb 20 Autos des neuen Typs auf ihren Verbrauch und berechnet aus diesen Werten das arithmetische Mittel $\bar{x} = 6.1$. Kann man hiermit die Behauptung des Herstellers widerlegen? Führen Sie einen Test zum Niveau $\alpha = 0.05$ durch.

(b) Eine Autozeitschrift bekommt von 152 Käufern des neuen Typs Beschwerden über den zu hohen Verbrauch zugesandt. Sie errechnet aus diesen Werten das arithmetische Mittel $\bar{x} = 7.3$. Kann man hiermit die Behauptung des Herstellers widerlegen?

Keine Abgabe