Lösungsvorschläge für die Übungsaufgaben der Vorlesung “Statistik“

Thomas Rupp
thomas@7t7.de

24. Januar 2002

Aufgabe 33 Wir folgen der Vorlesung vom 13.12.01:
$\mathcal{X} = (x_1, x_2, x_3)$ ist multinomialverteilt mit den Wahrscheinlichkeitsgewichten

$$g_\vartheta(\mathcal{X}) = \binom{x_1 + x_2 + x_3}{x_1 x_2 x_3} \left((1-\vartheta)^2\right)^{x_1} (2\vartheta(1-\vartheta))^{x_2} \left(\vartheta^2\right)^{x_3}.$$

Dann ist

$$L(\vartheta, \mathcal{X}) = \log g_\vartheta(\mathcal{X}) = \log \frac{m}{x_1 x_2 x_3} + x_2 \log 2 + (2x_1 + x_2)\log(1-\vartheta) + (x_2 + 2x_3)\log\vartheta.$$

a)
Der ML-Schätzer ist dasjenige ϑ, für das dieses maximal ist:

$$\begin{aligned} \frac{dL}{d\vartheta} &= \frac{x_2 + 2x_3}{\vartheta} - \frac{2x_1 + x_2}{1-\vartheta} = \frac{68 + 224}{\vartheta} - \frac{20 + 68}{1-\vartheta} = 0 \\ \Leftrightarrow \quad \vartheta &= \frac{292}{380} = \frac{73}{95} \Rightarrow \hat{\vartheta} = \frac{73}{95} \end{aligned}$$

b)
Die Asymptotische Varianz ist $\frac{1}{n}\frac{1}{I(\vartheta)}$. Dabei ist n die Anzahl der durchgeführten Experimente; hier Eins.[1] Wegen $I(\vartheta) = -\mathbb{E}_\vartheta[L''(\vartheta, \mathcal{X})]$ können wir diese berechnen:

$$\begin{aligned} \mathbb{E}(x_1) &= 190(1-\vartheta)^2 \\ \mathbb{E}(x_2) &= 190 \cdot 2\vartheta(1-\vartheta) \\ \mathbb{E}(x_3) &= 190\vartheta^2 \\ -\mathbb{E}_\vartheta[L''(\vartheta, \mathcal{X})] &= -\mathbb{E}\left[-\frac{x_2 + 2x_3}{\vartheta^2} - \frac{2x_1 + x_2}{(1-\vartheta)^2}\right] \\ &= \frac{1}{\vartheta^2}\left(\mathbb{E}(x_2) + 2\mathbb{E}(x_3)\right) + \frac{1}{(1-\vartheta)^2}\left(2\mathbb{E}(x_1) + \mathbb{E}(x_2)\right) \\ &= 380\frac{1-\vartheta}{\vartheta} + 380 + 380 + 380\frac{\vartheta}{1-\vartheta} \\ &\Rightarrow \quad I(\hat{\vartheta}_n) = \frac{1714750}{803} \approx 2135.43 \end{aligned}$$

Also ist die Asymptotische Varianz des ML-Schätzers $\frac{803}{1714750} \approx 0.000468$.

c)
Das 99er Konfidenzintervall ist dann

$$[\hat{\vartheta}_n - \frac{1}{\sqrt{n}}\frac{1}{\sqrt{I(\hat{\vartheta}_n)}}\texttt{qnorm}(0.995), \hat{\vartheta}_n + \frac{1}{\sqrt{n}}\frac{1}{\sqrt{I(\hat{\vartheta}_n)}}\texttt{qnorm}(0.995)] = [0.71268, 0.824]$$

für $n = 1$. Wenn wir den Weg über 190 Einzelexperimente gewählt hätten, dann wäre n nun 190. In jedem Fall taucht das $n = 190$ entweder bei den Erwartungswerten oder bei der asymptotischen Varianz auf; nicht aber bei beiden.

Aufgabe 34
Diese Aufgabe führen wir mit R aus.
Wir speichern die beiden Tabellen mittels `p<-matrix(c(235,38,35,7),nc=2)` und `q<-matrix(c(122,103,93,69),nc=2)` in Variablen gespeichert. Wir können nun den Chi-Quadrat-Test ohne oder mit "Yates' continuity correction" durchführen.

[1] Ein Experiment mit 190 Komponenten. Genauso könnten wir auch 190 Einzelexperimente durchfhren und $n = 190$ setzen; dann müssen die Erwartungswerte aber für nu jeweils ein Einzelexperiment berechnet werden.

Mittels `chisq.test(p,correct=F)`, `chisq.test(q,correct=F)` und (für die Summe der Einzeltabellen) `chisq.test(p+q,correct=F)` erhalten wir die folgenden (rechtsseitigen), nicht geglätteten p-Werte: $0.6357, 0.5339$ und 0.01993.
Mit Glättung ergibt sich durch `chisq.test(p)`, `chisq.test(q)`, `chisq.test(p+q)` die p-Werte $0.8128, 0.6042$ und 0.02524.

In beiden Fällen lässt sich bestätigen, das es nur bei der Betrachtung der gemeinsamen Tabelle einen Grund gibt, die Gliechbehandlung der Geschlechter abzulehnen. Das liegt daran, dass es bei den einzelnen Tabellen noch um 'zu kleine' Datensätze handelt, so dass sich die Daten noch nicht nahe um ihren wahren Erwartungswert gruppieren.

Fisher's exakter Test (wie im ersten Übungsblatt fördert ganz analoge Ergebnisse zutage.

Aufgabe 35
Wie sieht der gesuchte ML-Schätzer $\hat{\vartheta}_n$ aus? In jedem Fall $\geq \max(x_1, \ldots, x_n)$.
Es ist aber auch $g_\vartheta(x_i) = \frac{1}{\vartheta} \cdot 1_{[0,\vartheta]}(x_i)$. Somit ist $\sum_{i=1}^n L(\vartheta, x_i) = -n \log \vartheta$. Da diese Funktion monoton fallend ist, ist sie umsogrößer, umsokleiner ϑ ist. Letzteres ist jedoch durch $\max(x_1, \ldots, x_n)$ beschränkt. Somit ist $\hat{\vartheta}_n := \max(x_1, \ldots, x_n)$.
Jetzt berechnen wir nurnoch die gesuchte Verteilungsfunktion:

$$
\begin{aligned}
P(n(\vartheta - \hat{\vartheta}_n) \leq x) &= P(\vartheta - \max(x_1, \ldots, x_n) \leq \frac{x}{n}) \\
&= P(\min(x_1, \ldots, x_n) \leq \frac{x}{n}) = 1 - P(x_1 > \frac{x}{n}) \cdots P(x_n > \frac{x}{n}) \\
&= 1 - \left(\frac{\vartheta - \frac{x}{n}}{\vartheta}\right)^n = 1 - \left(1 - \frac{\frac{x}{\vartheta}}{n}\right)^n
\end{aligned}
$$

Daraus folgt für den Limes und danach für die Dichte

$$
\lim_{n\to\infty} 1 - \left(1 - \frac{\frac{x}{\vartheta}}{n}\right)^n = 1 - e^{-\frac{x}{\vartheta}} \text{ und } \frac{d(1 - e^{-\frac{x}{\vartheta}})}{dx} = \frac{1}{\vartheta} e^{-\frac{x}{\vartheta}}.
$$

Somit ist die Folge der ML-Schätzer asymptotisch nicht normal, sondern exponentiell verteilt.

Aufgabe 36 Es ist $g_{(\mu,\sigma^2)}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)$ und somit $\log g_{(\mu,\sigma^2)}(x) = -\log\sqrt{2\pi} - \log\sigma - \frac{x^2 - 2x\mu + \mu^2}{2\sigma^2}$. Die Fisher-Informationsmatrix ist dann

$$
\mathbb{E}_{(\mu,\sigma^2)} \left[\begin{array}{cc} \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\mu} \cdot \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\mu} & \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\mu} \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\sigma^2} \\ \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\sigma^2} \cdot \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\mu} & \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\sigma^2} \cdot \frac{\partial \log g_{(\mu,\sigma^2)}(x)}{\partial\sigma^2} \end{array} \right].
$$

MAPLE rechnet dann aus, dass dies

$$
\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}^{-1} \text{ ist.}
$$

Im Sinne der Vorlesung vom 17.12. ist $m(\vartheta) = m((\mu, \sigma^2) = \sigma^2$. Die asymptotische Varianz $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ ist dann

$$
\mathrm{grad} m(\vartheta) \cdot I(\vartheta)^{-1} \cdot \mathrm{grad} m(\vartheta)^\top = (0,1) \cdot I(\vartheta)^{-1} \cdot (0,1)^\top = 2\sigma^4.
$$

Nach Aufgabe ist die tatsächliche Varianz $\cdot\sqrt{n}$ des ML-Schätzers (also skaliert) $\sqrt{n}\mathrm{Var}\hat{s}_n^2 = n\frac{1}{n}(m_4 - \frac{n-3}{n-1}\sigma^4) \to 2\sigma^4$, da $m_4 = \mu^4 + 3\sigma^4 + 6\mu^2\sigma^2$ und $n \to \infty$. Die asymptotische Varianz ist also im Limes gleich der tatsächlichen.