Lösungsvorschläge für die Übungsaufgaben der Vorlesung “Statistik“

Thomas Rupp
thomas@7t7.de

13. Dezember 2001

Aufgabe 25 Auf den ersten Blick bietet sich natürlich sofort an, den Logarithmus der X_i zu betrachten.
Denn

$$\log(X_i) = \log(\beta) + Z_i', \qquad Z_i' \sim \mathfrak{N}(0, \sigma^2).$$

Wegen der Normalverteilung der Z_i' bietet sich jetzt natürlich der (gegen Streckung robuste) T-Test an.
Ein `t.test(log(daten))` bringt dann folgendes:
ein 95er Konfidenzintervall für $\log\beta$: $[-0.876, 1.679]$ und einen geschätzten Mittelwert von 0.4017.
Daraus ergibt sich das Konfidenzintervall für β: $[0.4166, 5.36]$.
Der Median ist `median(log(daten))`$= 0.0938$.
Die p-Werte des t.tests sind so gross, das sie nicht zur Ablehnung taugen und auf ein gutes Ergebnis hoffen lassen.
Daraus ergibt sich dann ein aus dem Mittelwert geschätztes $\beta = e^{0.4017} \approx 1.5$, oder aus dem Median geschätztes $\exp(0.0938) = 1.09$ welche dem tatsächlichen $\beta = 2$ einigermaßen nahe kommen (σ war 3).

Aufgabe 26
Folgen wir einfach den Buchstaben des Satzes:
Sei f_0 die Dichte von $X = (X_1, X_2, \ldots, X_n), X_i \sim \mathfrak{N}(0,1)$ und analog f_1 die Dichte einer $\mathfrak{N}(1,1)$ verteilten, $n-$dimensionalen ZV.
Die Erwartungswerte < 0 und > 1 müssen wir nicht betrachten, da, wie wir dann sehen, die Fehler nur kleiner werden.

$$\begin{aligned}
A(c) &= \{x \in \mathbb{R}^d | f_0(x) \geq c f_1(x)\} \\
&= \{x \in \mathbb{R}^d | \frac{1}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^{n} e^{-\frac{x_i^2}{2}} \geq c \frac{1}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^{n} e^{-\frac{x_i^2-1}{2}}\} \\
&= \{x \in \mathbb{R}^d | \prod_{i=1}^{n} e^{-\frac{x_i^2}{2} + \frac{x_i^2-2x_i+1}{2}} \geq c\} \\
&= \{x \in \mathbb{R}^d | \sum_{i=1}^{n} x_i \leq \frac{n}{2} - \log c\}
\end{aligned}$$

Mit $\alpha = 0.05$ und $n = 10$ soll nun ja gelten

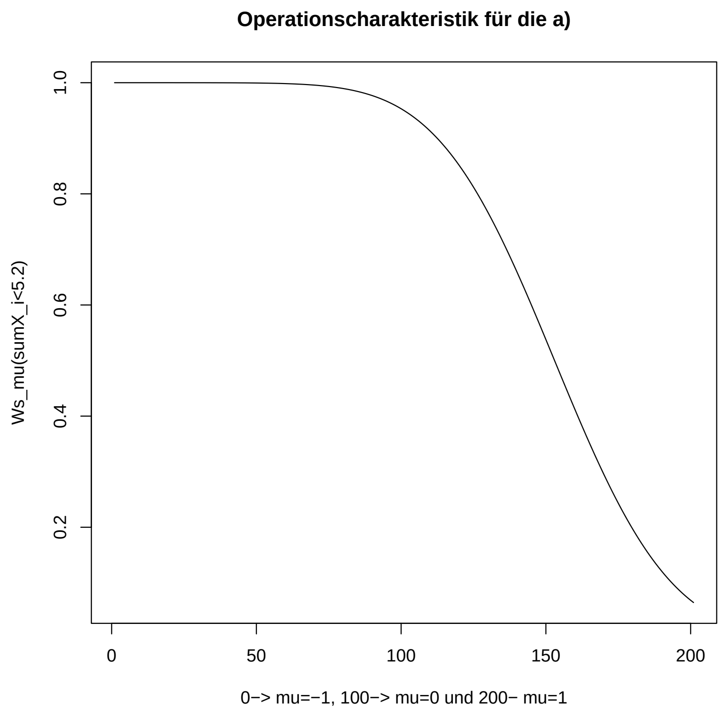
$$\begin{aligned}
0.05 &= P_0(\{X \notin A(c)\}) \\
\Leftrightarrow 0.95 &= P_0(\{X \in A(c)\}) \\
&= P_0(\{\sum_{i=1}^{10} X_i \leq 5 - \log c\}) \\
&\Leftrightarrow\ 5 - \log c = \texttt{qnorm(0.95,0,}\sqrt{10}\texttt{)} = 5.2 \\
(&\Leftrightarrow\ c = 0.8175)
\end{aligned}$$

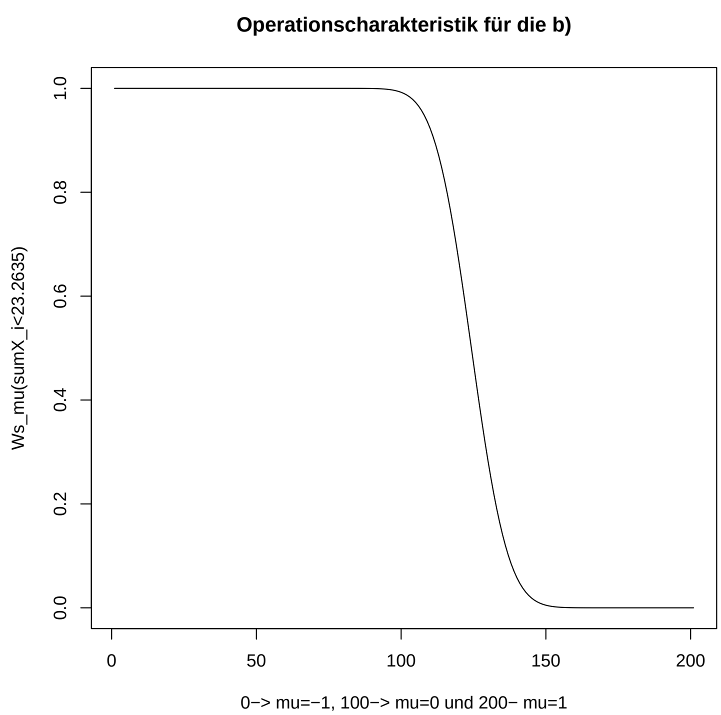
Denn die Summe von n identischen, gemäß $\mathfrak{N}(\mu, \sigma^2)$ verteilten ZVs ist $\mathfrak{N}(n\mu, n\sigma^2)$ verteilt.
Wir lehnen die Hypothese $\mu \leq 0$ also nur ab, wenn $\sum_{i=1}^{10} X_i > 5.2$. Der Fehler zweiter Art ist dann

$$P_1(\sum_{i=1}^{10} X_i \leq 5.2) = \texttt{pnorm(5.2,10,}\sqrt{10}\texttt{)} = 0.06458.$$

Für $n = 100$ und $\alpha = 0.01$ gilt dann:

$50 - \log c =$`qnorm(0.99,0,`$\sqrt{100}$`)`$= 23.2635 \quad (\Rightarrow c = 4.088 \cdot 10^{-11})$.
Und der Fehler zweiter Art ist: `pnorm(23.2635,100,`$\sqrt{100}$`)`$= 8.358 \cdot 10^{-15}$.
Die beiden dazugehörigen Operationscharakteristiken sind





Ein äquivalenter, aber vielleicht einsichtiger Test ist folgender:
Da ja $\mathfrak{N}(n\mu, n\sigma^2) \sim \sum X_i = n\bar{X}$ folgt $\bar{X} \sim \mathfrak{N}(\mu, \frac{\sigma^2}{n})$.

Nun befinden wir uns nurnoch im $\mathbb{R}^d$ und $A(c)$ muss die Form $(-\infty, a]$ haben. Mit $P(\bar{X} \leq a) = 0.95$ folgt $a =$`qnorm(0.95,0,`$\sqrt{0.1}$`)`$= 0.52$. Der Fehler zweiter Art ist dann `pnorm(0.52,1,`$\sqrt{0.1}$`)`$= 0.0645$ natürlich gleich wie oben.

Analog ist bei $\alpha = 0.01$ und $n = 100$ der Fehler zweiter Art ist:
`pnorm(qnorm(0.99,0,`$\sqrt{0.01}$`),1,`$\sqrt{0.01}$`)`$= 8.358 \cdot 10^{-15}$.

Aufgabe 27

Der P-Wert ist unter der Hypothese uniform auf $[0,1]$ verteilt: der P-Wert ist c, wenn der beobachtete Datensatz nur mit Wahrscheinlichkeit $\leq c$ auftritt.

Oder: wie wahrscheinlich ist es, dass P kleiner als c ist? Genauso wahrscheinlich wie das ein Datensatz nur mit Wahrscheinlichkeit $\leq c$ eintritt; und dies ist nunmal c.

Dann haben wir

$$P_0(-2\log P \leq x) = P_0(P \geq \exp(-\frac{x}{2})) = 1 - P_0(P \leq \exp(-\frac{x}{2})) = 1 - e^{-\frac{x}{2}}.$$

Und die die Dichte davon ist

$$\frac{d(1-\exp(-\frac{x}{2}))}{dx} = \frac{e^{-\frac{x}{2}}}{2}.$$

Rufen wir uns jetzt in Erinnerung, dass die Dichte der χ^2-Verteilung

$$\frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \text{ für } x > 0 \text{ ist.}$$

Also ist $-2\log P \sim \chi^2(2)$. Jetzt benutzen wir nurnoch, dass

$$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n) \qquad (\text{mit } X \sim \mathfrak{N}(0,1))\,.$$

Also ist $\sum_{i=1}^{k}(-2\log P_i) \sim \chi^2(2k)$.

Aufgabe 28

a)

$$\begin{aligned}
\mathbb{E}(X-c)^2 &= (\mu - c)^2 + \mathrm{Var}X = \mu^2 - 2\mu c + c^2 + \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \\
\Leftrightarrow \mathbb{E}(X^2 - 2cX + c^2) &= c^2 - 2\mu c + \mathbb{E}(X^2) \\
\Leftrightarrow \mathbb{E}(X^2) - 2c\mu + c^2 &= c^2 - 2\mu c + \mathbb{E}(X^2) \\
\Leftrightarrow 0 &= 0
\end{aligned}$$

b)

Setze $c := m(\vartheta)$ und $X := S$.

c)

Erstmal eine kleine "Zwischenrechnung". Weil die X_i unabhängig und identisch verteilt sind, gilt mit $Z_i := X_i - \mu$:

$$\begin{aligned}
\mathbb{E}(\hat\sigma^2) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar X)^2\right) \\
&= \mathbb{E}\left(\frac{1}{n}\sum(X_i^2 - 2X_i\bar X + \bar X^2)\right) \\
&= \frac{1}{n}\left(\sum \mathbb{E}X_i^2 - \frac{2}{n}\mathbb{E}(\sum X_i \sum X_j) + \frac{1}{n^2}\mathbb{E}(\sum X_i)^2\right) \\
&= \mathbb{E}X^2 - \frac{2}{n}\mathbb{E}X^2 - \frac{2}{n^2}\mathbb{E}(\sum_{i<j} X_i X_j) + \frac{1}{n}\mathbb{E}X^2 + \mathbb{E}(\sum_{i<j} X_i X_j) \\
&= \frac{n-1}{n}\mathbb{E}X^2 - \frac{1}{n^2}\sum_{i=1}^{n-1}(n-i)\mu^2 \\
&= \frac{n-1}{n}\mathbb{E}X^2 - \frac{1}{n^2}n(n-1)\mu^2 = \mathrm{Var}(X)\frac{n-1}{n} \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

Und mit dem Ergebnis aus Aufgabe 24 und der b) ergibt sich

$$\begin{aligned}
\mathbb{E}_\pi[(\hat\sigma^2 - \sigma^2)^2] &= \big(\mathbb{E}(\hat\sigma^2) - \sigma^2\big) + \mathrm{Var}\hat\sigma^2 \\
&= \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 + \mathrm{Var}\left[\frac{1}{n}\sum(X_i - \bar X)^2\right] \\
&= \frac{\sigma^4}{n^2} + \left(\frac{n-1}{n}\right)^2 \mathrm{Var}\left[\frac{1}{n-1}\sum(X_i - \bar X)^2\right] \\
&= \frac{\sigma^4}{n^2} + \left(\frac{n-1}{n}\right)^2 \frac{1}{n}\left(m_4 - \frac{n-3}{n-1}\sigma^4\right)
\end{aligned}$$