Lösungsvorschläge für die Übungsaufgaben der Vorlesung "Statistik"

Thomas Rupp
thomas@7t7.de

7. März 2002

Aufgabe 45
Das Modell sieht ja im Grunde so aus:

$$\mathfrak{Y} = \vec{\mu} + \sigma^2 \mathfrak{Z}, \quad \mathfrak{Z} \sim \mathcal{N}(\vec{0}, I).$$

Da das Rauschen in Erwartung Null ist würden wir analog zu dem Fall $\vec{\mu} \in D$ folgende Schätzer erwarten: $(\hat{\vec{\mu}}, \hat{\sigma}^2) = \left(\mathfrak{Y}, \frac{\sum (Y_i - \hat{\mu}_i)^2}{n}\right)$. Dies werden auch die ML-Schätzer sein, denn $\mathfrak{Y}$ ist normalverteilt im $\mathbb{R}^n$ und deswegen

$$\begin{aligned}
F_{(\vec{\mu},\sigma^2)}(\mathfrak{Y}) &= \log\left[\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu_i)^2}{2\sigma^2}\right)\right] \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{\sum_{i=1}^n (y_i - \mu_i)^2}{2\sigma^2} \\
\Rightarrow \frac{\partial F}{\partial \mu_i} &= \frac{2y_i}{2\sigma^2} - \frac{2\mu_i}{2\sigma^2} = 0 \Leftrightarrow \mu_i = y_i \\
\Rightarrow \frac{\partial F}{\partial \sigma^2} &= -2\pi\frac{n}{2}\frac{1}{2\pi\sigma^2} + \frac{\sum_{i=1}^n (y_i - \mu_i)^2}{2(\sigma^2)^2} = 0 \Leftrightarrow \sigma^2 = \frac{\sum_{i=1}^n (y_i - \mu_i)^2}{n}
\end{aligned}$$

Aufgabe 46 Das Modell kann folgendermaßen aussehen:

$$\mathfrak{Y} = C \cdot (w_1, w_2) + \sigma^2 \mathfrak{Z} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot (w_1, w_2) + \sigma^2 \mathfrak{Z}.$$

Dabei ist zu beachten, dass die Summe mi 7 und die Differenz mit 1 gemessen wurde.

b)

Der KQ-Schätzer für $\begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix}$ ist (siehe Vorlesung)

$$\begin{aligned}
\begin{pmatrix} \hat{w}_1 \\ \hat{w}_2 \end{pmatrix} &= \left(C^\top C\right)^{-1} C^\top \mathfrak{Y} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ 3 \\ 7 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 11 \\ 9 \end{pmatrix} = \begin{pmatrix} \frac{11}{3} \\ 3 \end{pmatrix}
\end{aligned}$$

Mit $M = \{(m_1, m_2, m_1 + m_2, m_1 - m_2) | m_1, m_2 \in \mathbb{R}\}, \dim M = k = 2$ ist die Streuung gegeben durch

$$s^2 = \frac{\|\mathcal{P}_{M^\perp}\mathfrak{Y}\|^2}{n-k} = \frac{\|\mathfrak{Y} - \overbrace{(\hat{w}_1, \hat{w}_2, \hat{w}_1 + \hat{w}_2, \hat{w}_1 - \hat{w}_2)^\top}^{=:\hat{\mathfrak{Y}}}\|^2}{n-k}.$$

Somit ist $s^2 = (3 - \frac{11}{3})^2 + (3-3)^2 + (7 - \frac{20}{3})^2 + (1 - \frac{2}{3})^2)/2 = \frac{1}{3}$.

c)

Der Quotient der Streuung unter der Hypothese H_0 und der Streuung ohne Hypothese ist ein Quotient zweier, durch ihre Freiheitsgrade normierte χ^2-verteilten Statistiken. Diese ist Fisher-verteilt und somit ist

$$F = \frac{\frac{\|\mathcal{P}_{L^\perp}\mathfrak{Y}\|^2}{k-r}}{\frac{\|\mathcal{P}_{M^\perp}\mathfrak{Y}\|^2}{n-k}} = \frac{\frac{\|\mathfrak{Y}-\hat{\mathfrak{Y}}_{H_0}\|^2}{k-r}}{\frac{\|\mathfrak{Y}-\hat{\mathfrak{Y}}\|^2}{n-k}} \sim F(1,2).$$

Dabei ist $L = \{(l,l,2l,0)|l \in \mathbb{R}\}, \dim L = r = 1$. MaW.: M ist der Raum, in denen die wahren und geschätzten Gewichte liegen, L ist der Unterraum, in dem die beiden Gewichte gleich sind.
Wir brauchen also noch $\hat{\mathfrak{Y}}_{H_0}$; dies ist der Punkt in L, der dem Punkt $\mathfrak{Y}$ am nächsten liegt. Wir suchen also das Minimum von $(3-l)^2 + (3-l)^2 + (7-2l) + (1-0)^2$. Dies ist für $l = \frac{10}{3}$ der Fall. Somit ist $\hat{\mathfrak{Y}}_{H_0} = (10/3, 10/3, 20/3, 0)^\top$.
Letzlich ist somit

$$F = \frac{(3-10/3)^2 + (3-10/3)^2 + (7-20/3)^2 + (1-0)^2}{1/3} = \frac{2/3}{1/3} = 2.$$

Der P-Wert ist somit $1-\texttt{pf}(2,1,2) \approx 0.29$. Also besteht kein Grund die Hyptohese des gleichen Gewichts abzulehnen.

Aufgabe 47
a)
Wegen $\beta := \beta_1 = -\beta_2, \gamma := \gamma_1 = -\gamma_2$ und $\delta := \delta_{11} = -\delta_{12} = -\delta_{21} = \delta_{22}$ können wir das Modell so schreiben:

$$\begin{pmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{pmatrix} = \begin{pmatrix} \alpha \\ \alpha \\ \alpha \\ \alpha \\ \alpha \\ \alpha \\ \alpha \\ \alpha \end{pmatrix} + \begin{pmatrix} \beta \\ \beta \\ \beta \\ \beta \\ -\beta \\ -\beta \\ -\beta \\ -\beta \end{pmatrix} + \begin{pmatrix} \gamma \\ \gamma \\ -\gamma \\ -\gamma \\ \gamma \\ \gamma \\ -\gamma \\ -\gamma \end{pmatrix} + \begin{pmatrix} \delta \\ \delta \\ -\delta \\ -\delta \\ -\delta \\ -\delta \\ \delta \\ \delta \end{pmatrix} + \sigma\mathfrak{Z}.$$

M, der Raum in der die wahren Parameter liegen, ist dann

$$M = \{(\alpha+\beta+\gamma+\delta, \alpha+\beta+\gamma+\delta, \alpha+\beta-\gamma-\delta, \dots, \alpha-\beta-\gamma+\delta)|\alpha,\beta,\gamma,\delta \in \mathbb{R}\}$$

mit $\dim M = k = 4$.
Der von $\mathfrak{Y}$ auf M projezierte Punkt für den KQ-Schätzer ist der Punkt in M, der minimalen Abstand zu $\mathfrak{Y}$ hat. Also das Minimum der Abbildung $M \ni x \mapsto (Y_{111} - (\alpha+\beta+\gamma+\delta))^2 + (Y_{112} - (\alpha+\beta+\gamma+\delta))^2 + \dots + (Y_{222} - (\alpha-\beta-\gamma+\delta))^2$. In δ-Richtung ist das Minimum bei $\frac{1}{8}(Y_{111} + Y_{112} + Y_{211} + Y_{222} - Y_{121} - Y_{122} - Y_{211} - Y_{212}) =: \hat{\delta}$.

b)

Mit $\sigma^2 = \frac{\|\mathfrak{Y}-\hat{\mathfrak{Y}}\|^2}{8-4} = 1$ (die Schätzer für die anderen Parameter kann man analog ausrechnen) und einsetzen kommt folgende Bedingung zu Tage:

$$(Y_{111} - Y_{112})^2 + (Y_{121} - Y_{122})^2 + (Y_{211} - Y_{212})^2 + (Y_{221} - Y_{222})^2 = 8.$$

Gesucht sind nun Datenvektoren $\mathfrak{Y}$, so dass

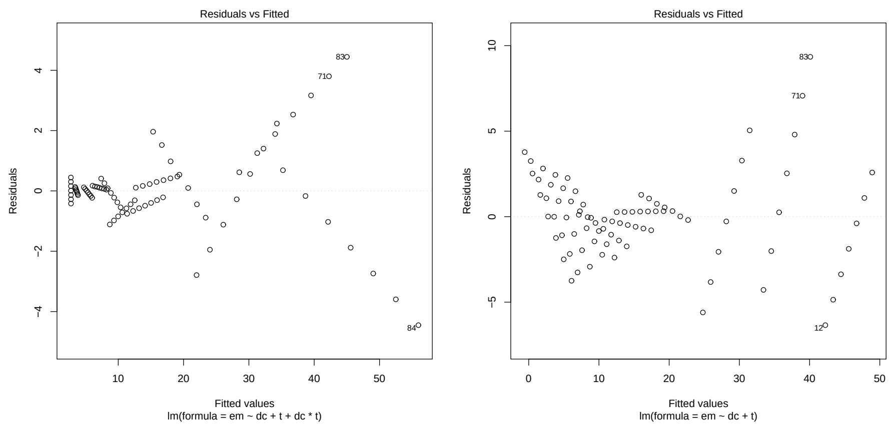
$$\frac{\frac{\|\mathfrak{Y}-\hat{\mathfrak{Y}}_{\delta=0}\|^2}{4-3}}{\sigma^2} <> \texttt{qf}(0.95, 1, 4) = 7.7.$$

Dies kann man wieder auf zu Afugabe 46 analoge Weise machen: der gesuchte Punkt $\hat{\mathfrak{Y}}_{\delta=0}$ ist derjenige aus der Ebene $(\alpha+\beta+\gamma, \alpha+\beta+\gamma, \alpha+\beta-\gamma, \dots, \alpha-\beta-\gamma)$ welcher

dem Vektor $\mathfrak{Y}$ am nächsten liegt. Also dies wieder als Funktion aufschreiben, nach α, β, γ ableiten und gleich Null setzen. Mit diesen neu gewonnen $\hat{\alpha}', \hat{\beta}'$ und $\hat{\gamma}'$ kann man die Aufgae durch geschicktes Raten schnell lösen. Der Term $\|\mathfrak{Y} - \hat{\mathfrak{Y}}_{\delta=0}\|^2$ ist
i) 8.5 für $\mathfrak{Y} = (2,4,0,0,2,2,3,1)^\top$
ii) 4 für $\mathfrak{Y} = (2,2,2,0,2,2,2,0)^\top$.

In beiden Fällen ist $\sigma^2 = 1$ da $(2-4)^2+0+0+(3-1)^2 = 8$ bzw. $0+2^2+0+2^2 = 8$ ist.

Aufgabe 48 Mit dem beiliegenden Programm kann man sich alles schön ansehen. Mit `source`("a48.r") wird das Programm ausgeführt und die Plots werden ausgegeben.
Insbesondere der Plot der Residuen gegen die gefilterten (also die geschätzten systematischen Komponenten) sollte im Idealfall eine zufällige Punktwolke ergeben:



Dies ist jedoch offensichtlich nicht der Fall; der Plot enthält soviel Systematik, so dass das zugrunde liegende Modell eher abwegig erscheint. Eine einfache Modifikation durch weglassen der multiplikativen Komponente ist nicht ausreichend, wie man hier sehen kann.
Was jedoch abzulesbar scheint, ist eine zunehmende Variabilität in Abhängigkeit der Temperatur bei hohen Dicumylperoxidgehalten.

Desweiteren zweigen auch die Plots der Residuen gegen die Normalverteilung, dass die Residuen nicht wirklich normalverteilt sind (also nicht auf einer Geraden liegen).