Universität Konstanz WS 2007/08
Fachbereich Mathematik und Statistik
Prof. Dr. J. Beran
07. Januar 2008

4. Computerübung zur Mathematischen Statistik

Aufgabe 1 (Jackknife Methode)
Sei $(X_1, ..., X_n)$ eine Stichprobe und $T_n(X_1, ..., X_n)$ ein Schätzer für den Parameter θ. Nun werden n Schätzer $T_n^{(i)}, i = 1...n$ berechnet, wobei $T_n^{(i)} = T_{n-1}(X_1...X_{i-1}, X_{i+1}...X_n), i = 2, ..., n-1$ bzw. $T_n^{(1)} = T_{n-1}(X_2...X_n)$ und $T_n^{(n)} = T_{n-1}(X_1...X_{n-1})$.
Der Jackknife Schätzer ist dann gegeben durch $\hat{T}_n = nT_n - \frac{n-1}{n}\sum_{i=1}^{n} T_n^{(i)}$

(a) Simulieren Sie über den Befehl rbinom(...) 20 Binomialverteilte Daten mit $p = 0.3$ und speichern sie diese in x.

(b) Es soll p^2 durch $\overline{X}^2$ geschätzt werden.
Berechnen Sie theoretisch die Verzerrung von $\overline{X}^2$.

(c) Bestimmen Sie basierend auf dem Schätzer aus b) den Jackknife-Schätzer $\hat{\theta}_{JK}$ für p^2.

(d) Demonstrieren Sie durch Simulation, dass sich die Verzerrung von $\hat{\theta}_{JK}$ reduziert.

Aufgabe 2 (Logistische Regression)
Am 28. Januar 1986 explodierte die US Raumfähre Challenger 73 Sekunden nach dem Start. Es hatte sich der bis dahin schwerste Unfall in der Geschichte des Space Shuttles ereignet.Ein Dichtring in der Feststoffrakete wurde für die Katastrophe verantwortlich gemacht. Durch niedrige Nachttemperaturen wurde er porös und ließ heiße Gase entweichen. Im Vergleich zu den 23 vorhergegangenen Starts war die Temperatur mit 31°F sehr niedrig (siehe Tabelle challenger.xls) Nun könnte man vermuten,dass ein Zusammenhang zwischen der niedrigen Temperatur und dem Unfall besteht. Dies soll im Folgenden untersucht werden.

(a) Lesen Sie den Datensatz challenger.xls in R ein.Speichern Sie die Temperaturen in einer Variable Temp und die binäre Größe in einer Variable Failure.

(b) Berechnen Sie die relative Häufigkeit eines Schadens für den gesamten Datensatz und für Temperaturen $< 65°$! Zeichnen Sie außerdem einen Scatterplot der beiden Variablen und interpretieren Sie diesen.

(c) Nun soll eine Logistische Regression durchgeführt werden. Der Befehl dazu :

```
summary(out.int <- glm(Failure  ~ Temp, family=binomial))
```

Dabei berechnet die Funktion glm ein allgemeines lineares Modell und die Option family=binomial spezifiziert das Logistische Modell.
Hinweis:
Die Idee dahinter ist folgende: Die binäre Variable Y_i kann die Werte $0, 1$ annehmen,wobei 1 einen Schaden bedeutet.Nun möchte man die Wahrscheinlichkeit $P(Y_i = 1) =: \pi_i$ in Verbindung zur erklärenden Variable X_i (in unserem Beispiel die Temperatur) setzen.Ein einfacher linearer Zusammenhang ist hier nicht gegeben, daher bietet sich der Ansatz der logistischen Regression an :

$$\pi_i = \frac{exp(a + bX_i)}{1 + exp(a + bX_i)}$$

was durch Umformen folgenden Zusammenhang unterstellt:

$$\lambda_i := ln(\frac{\pi_i}{1 - \pi_i}) = a + bX_i$$

(d) Schreiben Sie das gerade geschätzte Modell auf und interpretieren Sie die Koeffizienten.

(e) Transformieren Sie die geschätzen Werte für λ_i in die Wahrscheinlichkeiten π_i.

(f) Verwenden Sie nun das Modell um die Wahrscheinlichkeit eines Ausfalls bei 31°F vorherzusagen. Interpretieren Sie!