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

5. Computerübung zur Mathematischen Statistik

Aufgabe 1 (Gammaverteilung)
Im Folgenden sind die Bedienungszeiten (in Minuten) von 15 Bankkunden gegeben:
8.77, 9.34, 20.37, 6.14, 14.84, 41.23, 7.91, 8.73,
4.68, 9.16, 2.46, 26.79, 3.33, 7.23, 6.91
Für die Bedienungszeiten wird angenommen, dass sie unabhängig und identisch gammaverteilt sind mit unbekannten Parametern $\alpha > 0$ und $\lambda > 0$.Die Dichte der Gammaverteilung lautet

$$f(x) = \frac{1}{\Gamma(\alpha)}\lambda^{\alpha}x^{\alpha-1}e^{-\lambda x}$$

ihr Erwartungswert $E(X) = \frac{\alpha}{\lambda}$ und ihre Varianz $Var(X) = \frac{\alpha}{\lambda^2}$.

(a) Geben Sie die Daten in R ein. Setzen Sie den Parameter $\alpha = 2$ und berechnen Sie Likelihood und Log-Likelihood Funktion.Zeichnen Sie beide in Abhängigkeit von λ.Zeichnen Sie zusätzlich die sog. Score-Funktion(die erste Ableitung der Log-Likelihoodfunktion nach dem Parameter).Was gilt für den optimalen Schätzer für λ?

(b) Berechnen Sie den Maximum-Likelihood-Schätzer für λ mittels des Newton-Raphson- Algorithmus. Warum sind Startwerte $\lambda_0 \geq 0.34$ ungeeignet?

(c) Nun sind beide Parameter unbekannt.Berechnen Sie mit Hilfe des Newton-Raphson Verfahrens Maximum-Likelihood-Schätzer für die Parameter α und λ.Verwenden Sie dafür geeignete Startwerte,die Sie aus den ersten beiden Momenten berechnen können.(Hinweis: Die erste bzw. zweite Ableitung der Funktion $ln(\Gamma(\alpha))$ erhält man durch die R-Befehle $digamma(\alpha)$ bzw. $trigamma(\alpha)$.)

Anleitung zum Newton-Raphson-Verfahren zur Bestimmung eines Parameters θ:

(1) Wähle einen Startwert θ_0, eine vorgegeben Genauigkeit ϵ (hier sinnvoll z.b.$\epsilon = 0.000001$) und setze $i = 0$.

(2) Berechne $\theta_{i+1} = \theta_i - \frac{l'(\theta_i)}{l''(\theta_i)}$.
l'/l" ist hierbei die 1./2. Ableitung der Log-Likelihood-Funktion nach dem Parameter θ.

(3) Gehe zu Schritt 4, wenn $|\theta_{i+1} - \theta_i| < \epsilon$,ansonsten erhöhe i um 1 und fahre mit 2. fort.

(4) θ_i ist der Wert des Maximums.

Anmerkung: Bei der Schätzung beider Parameter gleichzeitig ist θ der Vektor aus α und λ, die erste Ableitung l' ist der Vektor der Ableitungen nach den beiden Parametern, und l" ist die 2x2 Matrix der 2.Ableitungen(symmetrisch!).Durch eine Matrix teilen bedeutet hier mit der inversen zu multiplizieren (Befehl in R: solve(A))

Aufgabe 2 (Cauchy-Verteilung)
Die Dichte der Cauchy Verteilung mit Parametern μ und σ ist:

$$f(x) = \frac{1}{\sigma\pi} * \frac{1}{1 + (\frac{x-\mu}{\sigma})^2}$$

Wir nennen μ den location Parameter und σ den scale Parameter. Da die Verteilung symmetrisch um μ is, ist der sample-Median median(x) ein guter Schätzer für μ. Ein robuster Schätzer für σ ist der halbe Interquartil-Abstand $\frac{IQR(x)}{2}$.

(a) Berechnen Sie die Log-Likelihood Funktion für eine $Cauchy(\mu, \sigma)$-Verteilung theoretisch. In R kann man dies ganz schnell mithilfe des Befehls dcauchy machen.

(b) Simulieren Sie dann 100 (1000 , 10000) Cauchy(2,4)-verteilte Zufallsvariablen x (Befehl:rcauchy(..)) und minimieren Sie die negative Log-Likelihood Funktion über den Befehl nlm(Schauen Sie sich die help-Datei zu diesem Befehl an!) Dazu brauchen Sie einen geeigneten Startwert (über Median und IQR ausrechnen).

Hinweis: Definieren Sie als θ den Vektor aus μ und σ.

Aufgabe 3 (Einfacher Gauß-Test)
Eine Tischlerei muss für Gartenmöbel, die in Massen produziert werden, Kanthölzer zuschneiden, deren Länge 40 cm betragen soll. Die Länge der Kanthölzer kann als normalverteilte Zufallsvariable angenommen werden, die Standardabweichung der CNC-Maschine ist bekannt, sie beträgt $\sigma = 0,2cm$ (Maschineneinstellung). Bei der Qualitätssicherung wird in regelmäßigen Zeitabständen durch Stichproben geprüft, ob der Sollwert 40 cm eingehalten wird. Eine Stichprobe ergibt folgende Werte:

40,2 ; 39,9 ; 40,4 ; 40,0 ; 40,3

(a) Befindet sich der Produktionsprozess unter statistischer Kontrolle für das Signifikanzniveau $\alpha = 0.05$? Testen Sie dazu das Testproblem
Nullhypothese $H_0 : \mu = 40$ gegen die Alternative $H_1 : \mu \neq 40$
Die Teststatistik ergibt sich als

$z = \frac{(\bar{x}-\mu)\sqrt{n}}{\sigma}$ die $N(0,1)$ verteilt ist.

(b) Berechnen Sie den p-value (d.h. das Signifikanzniveau zu dem der Test gerade nicht mehr abgelehnt werden kann)