Albert-Ludwigs-Universität Freiburg Wintersemester 2011/2012

Statistische Methoden der Datenanalyse

Markus Schumacher

Übung VI

Markus Warsinsky

5.12.2011

Anwesenheitsaufgaben

Aufgabe 38 *Maximum-Likelihood-Anpassung an exponentiell verteilte Daten*

In Aufgabe 26 wurde eine ROOT-Datei mit simulierten Zerfallszeiten erzeugt.

In dieser Übung sollen Sie versuchen, herauszufinden, welchen Wert der Zerfallskonstante τ ein anderer Übungsteilnehmer benutzt hat, um diese Datei zu erzeugen.

Gehen Sie dazu wie folgt vor:

(i) Lassen Sie sich von einem anderen Übungsteilnehmer die erzeugte Datei geben und öffnen diese in ROOT. Ein Beispiel dafür findet sich im ROOT-Makro `/home/warsinsk/sd_wise1112/ueb6/aufgabe_38_anfang.C`. Kopieren Sie sich dieses Makro und öffnen Sie es. Versuchen Sie die Funktionsweise zu verstehen. Tragen Sie in Zeile 3 die Datei eines anderen Übungsteilnehmers ein.

(ii) Berechnen Sie für einen sinnvollen Wertebereich der Zerfallskonstante τ den Logarithmus der Likelihood für die Exponentialverteilung und die in Aufgabe 26 erzerzeugte Meßreihe:

$$\ln \mathcal{L}(\tau_j) = \sum_{i=1}^{N} \left[\ln\left(\frac{1}{\tau_j}\right) - \frac{x_i}{\tau_j} \right] \quad (1)$$

und stellen Sie sie mittels eines `TGraph` graphisch dar, in dem $\ln \mathcal{L}(\tau_j)$ gegen τ_j aufgetragen wird. Gehen Sie dazu wie folgt vor:

a) Erstellen Sie eine weitere `for`-Schleife, in der Sie verschiedene Werte für τ durchfahren.

b) Summieren Sie für jeden Durchlauf dieser `for`-Schleife die Werte der Likelihood auf. Füllen Sie die so erhaltenen Punkte in einen `TGraph` mittels der `SetPoint`-Methode.

(iii) Ermitteln Sie aus der erzeugten Abbildung eine Abschätzung für die Zerfallskonstante, $\hat{\tau}$, und deren Standardabweichung $\sigma_{\hat{\tau}}$ darauf. Vegleichen Sie den Wert mit dem Ergebnis, das aus der "Minimum Variance Bound" Methode ($\hat{V}(\hat{\tau}) = \frac{\hat{\tau}^2}{n}$) folgt.

(iv) In `/home/warsinsk/sd_wise1112/ueb6` liegen einige Dateien mit verschiedenem Stichprobenumfang. Lassen Sie ihr Skript über diese laufen und beobachten, was mit der Likelihood passiert.

(v) Nun wollen wir die in ROOT integrierte Anpassungsroutine benutzen, um eine Maximum-Likelihood-Anpassung an die Daten durchzuführen. Erstellen Sie dazu ein neues Skript, in dem Sie wieder den `TTree` einlesen. Die Schleife über alle Einträge ist jetzt nicht mehr notwendig.

Stellen Sie eine eindimensionale Exponentialverteilung bereit:

```
TF1 expFunk = TF1("expFunk", "(1.0/[0])*TMath::Exp(-x/[0])", 0.0, 100.0);
```

Setzen Sie den Parameter auf einen sinnvollen Startwert:

```
expFunk.SetParameter(0, 2.0);
```

Und führen Sie dann die Maximum-Likelihood-Anpassung an die Daten mittels des Kommandos

```
tree.UnbinnedFit("expFunk", "expzufall", "", "V");
```

durch. Die Option `"expzufall"` steht für den Namen der Zufallsvariablen im `TTree`. Die Option `"V"` steht für detaillierte Ausgabe. Vergleichen Sie mit dem Ergebnis, das aus dem Minos-Fit (Option `"VE"`) folgt.

(vi) Sollten Sie ein wenig mehr Zeit zur Verfügung haben, benutzen Sie die „Monte-Carlo"-Methode, um den Mittelwert des Schätzwertes der Zerfallskonstante $\overline{\hat{\tau}}$ und dessen Standardabweichung $s_{\hat{\tau}}$ zu bestimmen. Gehen Sie dazu wie folgt vor:

- Nehmen Sie den Schätzwert $\hat{\tau}$ aus Unteraufgabe (v). Sie können den Wert des Parameters der Funktion mittels `expFunk.GetParameter(0)` abfragen.
- Erstellen Sie einen neuen `TTree` (mit anderem Namen als den ursprünglichen) und bereiten Sie eine Fliesskommazahl vor, um sie in den `TTree` einzufüllen (wie dies geht, wurde in Aufgabe 26 besprochen).
- Generieren Sie eine weitere Stichprobe von 100 exponentiell verteilten Zufallszahlen mit $\tau' = \hat{\tau}$. Verwenden Sie dazu einen Zufallszahlengenerator vom Typ `TRandom3` und die Methode `Exp(tau)`. Füllen Sie diese Zufallszahlen in den `TTree`.
- Ermitteln Sie unter Verwendung der ROOT-Anpassungsroutine den Maximum-Likelihood-Schätzwert $\hat{\tau}'$ für diese neue Stichprobe.
- Wiederholen Sie diese Schritte 1000 mal und füllen Sie die Schätzwerte für die Zerfallskonstante aus jedem Schritt in ein Histogramm. Dabei sollten Sie jedesmal den `TTree` erst leeren, bevor wieder 100 Zufallszahlen eingefüllt werden. Dies kann mittels der `Reset()`-Methode erreicht werden.
- Lesen Sie aus diesem Histogramm die Werte für $\overline{\hat{\tau}}$ und $s_{\hat{\tau}}$ ab.

Hausaufgaben

Aufgabe 39 *„Maximum Likelihood" für die Normalverteilung* **4 Punkte**

Betrachten Sie n verschiedene Messungen x_i, wobei jede einen bekannten Fehler σ_i besitzt. Jede Variable ist also gemäß der Gaussverteilung $G(x_i;\mu,\sigma_i)$ verteilt. In dieser Übung werden wir den besten Schätzwert für den Mittelwert μ dieser Messungen ermitteln.

(i) Stellen Sie die Likelihoodfunktion für die n gaussischen Messungen auf.

(ii) Berechnen Sie den „Maximum Likelihood"-Schätzer $\hat{\mu}$ für den Mittelwert.

(iii) Berechnen Sie die Varianz des „Maximum Likelihood"-Schätzers $\hat{\mu}$, indem Sie die MVB-Formel

$$V(\hat{\mu}) = \frac{1}{\left(-\frac{\partial^2 \ln L}{\partial \mu^2}\right)_{\mu=\hat{\mu}}} \tag{2}$$

benutzen.

Aufgabe 40 *„Maximum Likelihood" für die Poissonverteilung* **3 Punkte**

Betrachten Sie eine einmalige Messung einer Meßgröße n, die gemäß der Poisson-Wahrscheinlichkeitsdichtefunktion $f(n;\nu) = \frac{\nu^n}{n!}e^{-\nu}$ verteilt sein soll.

(i) Berechnen Sie den „Maximum Likelihood"-Schätzer $\hat{\nu}$ für den Mittelwert.

(ii) Ist der Schätzer erwartungstreu?

(iii) Benutzen Sie die MVB-Methode, um die Varianz des Schätzers $V(\hat{\nu})$ auszurechnen.

Aufgabe 41 *Beweis der Rao-Cremér-Frechet-Ungleichung* **7 Punkte**

Eine Zufallsvariable folge der Wahrscheinlichkeitsdichtefunktion $f(x;\theta)$, wobei θ ein unbekannter Parameter sei. Betrachten Sie eine Stichprobe $\vec{x} = (x_1,\ldots,x_n)$, die benutzt wird, um einen Schätzer $\hat{\theta}(\vec{x})$ für θ zu ermitteln. Gehen Sie von diesem Startpunkt aus, um die Rao-Cremér-Frechet-Ungleichung

$$V[\hat{\theta}] \geq \frac{(1+\frac{\partial b}{\partial \theta})^2}{-E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]} \tag{3}$$

zu beweisen, wobei $b = E[\hat{\theta}] - \theta$ die Verzerrung (bzw. der Bias) des Schätzers ist. Gehen Sie dazu wie folgt vor:

(i) Benutzen Sie die Cauchy-Schwarz-Ungleichung für zwei Funktionen u und v

$$\left(\int u^2 \mathrm{d}x_1 \ldots \mathrm{d}x_n\right)\left(\int v^2 \mathrm{d}x_1 \ldots \mathrm{d}x_n\right) \geq \left(\int uv \mathrm{d}x_1 \ldots \mathrm{d}x_n\right)^2, \tag{4}$$

mit $u = (\hat{\theta} - E[\hat{\theta}])\sqrt{L}$ und $v = \frac{\partial \ln L}{\partial \theta}\sqrt{L}$, wobei $L = f_{\text{gem}}(\vec{x};\theta)$ die Likelihoodfunktion ist, welche auch gleichzeitig die gemeinsame Wahrscheinlichkeitsdichteverteilung für $\vec{x}$ ist.

Schreiben Sie die Ungleichung in Form von Erwartungswerten und Varianzen und ermitteln Sie daraus eine untere Grenze auf $V[\hat{\theta}]$.

(ii) Nehmen Sie an, dass $\vec{x} = (x_1,\ldots,x_n)$ nicht von θ abhängt, so dass θ und die Ableitung nach θ vor das Integral gezogen werden können. Zeigen Sie damit, dass

$$E\left[\frac{\partial \ln L}{\partial \theta}\right] = \int \frac{\partial \ln f_{\text{gem}}(\vec{x};\theta)}{\partial \theta} f_{\text{gem}}(\vec{x};\theta)\mathrm{d}x_1 \ldots \mathrm{d}x_n = 0.$$

(iii) Zum Beweis der Rao-Cremér-Frechet-Ungleichung zeigen Sie nun, dass

$$E\left[\hat{\theta}\frac{\partial \ln L}{\partial \theta}\right] = 1 + \frac{\partial b}{\partial \theta},$$

und in ähnlicher Weise:

$$E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right].$$

Nehmen Sie dazu wieder an, dass die Ableitung nach θ und die Integration über $\vec{x}$ vertauscht werden können.

Aufgabe 42 *Messung der Polarisationsasymmetrie* **6 Punkte**

Eine wichtige Messung der Paritätsverletzung in der elektroschwachen Wechselwirkung des Standardmodells wurde durchgeführt, indem die inelastische Streuung von polarisierten Elektronen an einem Deuteriumtarget vermessen wurde. Die Paritätsverletzung führt dabei zu einer Differenz im Streuwirkungsquerschnitt zwischen links- und rechts-händig polarisierten Elektronen. Die Meßgröße war dabei die Polarisationsasymmetrie α, die gegeben ist durch

$$\alpha = \frac{\sigma_R - \sigma_L}{\sigma_R + \sigma_L}, \tag{5}$$

wobei σ_L (σ_R) die Wirkungsquerschnitte für links- (rechts)-händige Elektronen bei der inelastischen Streuung an einem Deuteriumtarget sind.

Für eine gegebene Luminosität L ist die Anzahl der beobachteten Streuereignisse bei links- bzw. rechthändigen Elektronen eine poissonverteilte Zufallsvariable, n_L bzw. n_R, mit Mittelwerten ν_L bzw. ν_R. Die Mittelwerte stehen mit den Wirkungsquerschnitten über $\nu_L = \sigma_L L$ bzw. $\nu_R = \sigma_R L$ in Beziehung, wobei die Luminosität L für die beiden Datensätze identisch ist.

(i) Benutzen Sie die Tatsache, dass die beobachtete Anzahl von Streuereignissen poissonverteilt ist, um einen „Maximum Likelihood"-Schätzer $\hat{\alpha}$ für die Asymmetrie α aufzustellen.

(ii) Benutzen Sie Fehlerfortpflanzung (und nehmen Sie dabei an, dass die Messungen der Streuereignisse für links- bzw. rechtshändige Elektronen unabhängig voneinander sind), um die Standardabweichung $\sigma_{\hat{\alpha}}$ auf diesen Schätzer als Funktion von α und $n_{TOT} = n_R + n_L$ zu ermitteln.

(iii) Bevor das Experiment aufgebaut wurde, wurde erwartet, dass die Asymmetrie ungefähr 10^{-4} betragen würde. Welche Gesamtzahl von Streuereignissen musste man dafür beobachten, um die Asymmetrie mit einer relativen Genauigkeit von 10% zu messen?