

Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

Übung VI

03.12.2013, 04.12.2013

Anwesenheitsaufgaben

Aufgabe 33 *Maximum-Likelihood-Anpassung an exponentiell verteilte Daten*

In Aufgabe 21 wurde eine ROOT-Datei mit simulierten Zerfallszeiten erzeugt.

In dieser Übung sollen Sie versuchen, herauszufinden, welchen Wert der mittleren Lebensdauer 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/slai/StatisticsCourse/PS6/aufgabe_33_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 Lebensdauer τ den Logarithmus der Likelihood für die Exponentialverteilung und die in Aufgabe 21 erzeugte 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 Lebensdauer, $\hat{\tau}$, und deren Standardabweichung $\sigma_{\hat{\tau}}$ darauf. Vergleichen 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/slai/StatisticsCourse/PS6/` 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 Lebensdauer $\bar{\hat{\tau}}$ 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 21 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 Lebensdauer 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 $\bar{\hat{\tau}}$ und $s_{\hat{\tau}}$ ab.

(vii) Wiederholen Sie das Vorgehen von Unteraufgabe (vi), aber anstatt der Lebensdauer τ , nutzen Sie die Zerfallskonstante $\lambda = \frac{1}{\tau}$.

- Ihr TF1 soll so aussehen:
`TF1 expFunk = TF1("expFunk", "[0]*TMath::Exp(-x*[0])", 0.0, 100.0);`
- Was passiert bei geringem Stichprobenumfang? (5 oder weniger) Ist der Maximum-Likelihood-Schätzer erwartungstreu?

Hausaufgaben

Aufgabe 34 *Integration der Planck-Verteilung*

12 Punkte

Wir betrachten die Erzeugung von Photonen, die dem Gesetz von der Strahlung des Schwarzen Körpers folgen. Die Verteilung der skalierten Frequenz $x = h\nu/kT$ ist dann gegeben durch die Plancksche Strahlungsformel:

$$f(x) = c \frac{x^3}{e^x - 1}$$

wobei c eine Normierungskonstante ist. Im folgenden soll besprochen werden, wie man diese Funktion möglichst effizient zwischen $x = 0$ und einem bestimmten x_{max} integrieren könnte.

- (i) Warum lässt sich die Transformationsmethode nicht anwenden?
- (ii) Skizzieren Sie den Verlauf der Funktion.
- (iii) Ermitteln Sie einen Ausdruck für die Lage des Maximums und dessen Höhe. Die erhaltenen Gleichungen lassen sich nur numerisch lösen. Als Lage des Maximums ergibt sich $x \approx 2.82$ mit $f_{max} \approx 1.42c$.
- (iv) Beschreiben Sie, wie man nach der Akzeptanz-Zurückweisungsmethode vorgehen würde, um die Funktion zu integrieren bzw. Zufallszahlen zu erzeugen. Wie ist die Effizienz gegeben? Warum wird dieses Verfahren für hohe x sehr ineffizient?
- (v) Nun soll ein Ansatz mit einer stückweise definierten Majorante gemacht werden. Im unteren Bereich bis zu einem Punkt x_1 machen wir den Ansatz $g_I(x) = f_{max}$, oberhalb von x_1 hingegen den folgenden Ansatz:
 - a) Betrachten Sie zunächst das asymptotische Verhalten von $f(x)$ für große x . Welchen Einfluss hat der x^3 -Term?
 - b) Warum könnte der Ansatz

$$g_{II}(x) = K \cdot c \cdot x^{-\epsilon} \exp(-x^{1-\epsilon})$$

ein geeigneter Ansatz für eine Majorante sein?

- c) Ermitteln Sie die Transformationsvorschrift, die zu dieser Majorante gehört.
- d) Beschreiben Sie nun noch das allgemeine Verfahren für die Erzeugung von Zufallsvariablen bzw. zur Monte-Carlo-Integration. Wie wird entschieden, ob eine Zufallszahl $x < x_1$ oder $x > x_1$ genommen wird?
- e) Wie groß ist der Effizienzgewinn durch diese Majorante im Gebiet $x > x_1$ für $c = 1$, $K = 200$, $\epsilon = 0.1$, $x_{max} = 20$, $x_1 = 6$?

Aufgabe 35 *Maximum Likelihood für die Poissonverteilung*

8 Punkte

Beweisen Sie folgende wesentlichen Eigenschaften der Poissonverteilung $f(n; \nu) = \frac{\nu^n}{n!} e^{-\nu}$ für $n \geq 0$.

- (i) Betrachten Sie eine einmalige Messung einer Messgröße n , die gemäß der Poisson-Wahrscheinlichkeitsdichtefunktion verteilt sein soll. Berechnen Sie den Maximum Likelihood-Schätzer $\hat{\nu}$ für ν .
- (ii) Ist der Schätzer erwartungstreu?
- (iii) Betrachten Sie m unabhängige Messungen von dieser Messgröße $\vec{n} = \{n_1, n_2, \dots, n_m\}$. Berechnen Sie den Maximum Likelihood-Schätzer $\hat{\nu}$ für ν in diesem Fall.
- (iv) Ist dieser Schätzer in diesem Fall erwartungstreu?
- (v) Was ist die Effizienz dieses Schätzers?