Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

Übung XI

22.1.2013, 25.1.2013

Anwesenheitsaufgaben

Aufgabe 61 *Anpassungsgüte bei einer Maximum Likelihood Anpassung*

In dieser Übung sollen Sie die Ergebnisse der "Binned"Maximum Likelihood Anpassung an die $e^+e^- \to \mu^+\mu^-$-Ereignisse aus Übung VII benutzen. Betrachtet werden sollen die Verhältnisse der Likelihoodfunktionen

$$\lambda = \frac{\mathcal{L}(\vec{n}|\vec{\nu})}{\mathcal{L}(\vec{n}|\vec{n})}.$$

Damit soll gezeigt werden, dass

$$\chi^2_M = -2\ln\lambda_M$$

gemäß einer χ^2-Verteilung mit $N - m - 1$ Freiheitsgraden verteilt ist. Dabei ist N die Anzahl der Datenbins und m die Anzahl der angepassten Parameter.

Das Makro `/home/slai/StatisticsCourse/PS11/aufgabe61_anfang.C` gibt ein Beispiel, wie man die generierten $e^+e^- \to \mu^+\mu^-$-Ereignisse (Anzahl N_{tot}) einliest, ihre $\cos\theta$-Verteilung berechnet und in ein Histogramm namens `hist` einfüllt.

(i) Definieren Sie eine `TF1`-Funktion gemäß

$$f(x;\alpha,\beta) = \frac{1+\alpha x+\beta x^2}{2+\frac{2\beta}{3}},$$

und führen Sie eine Anpassung dieser Funktion an das $\cos\theta$-Histogramm mit dem Befehl

```
hist.Fit("FunkName","IL");
```

durch. Die Option `"IL"` steht für eine Maximum Likelihood Anpassung unter Benutzung der Integrale über die Bins des Histogramms `hist`.

(ii) Definieren Sie als nächstes eine weitere `TF1`-Funktion gemäß $f(x;\alpha,\beta)$ unter Benutzung der angepassten Werte $\hat{\alpha}$ und $\hat{\beta}$. Die Werte von $\hat{\alpha}$ und $\hat{\beta}$ können Sie mit

```
TF1::GetParameter(int i)
```

ermitteln, wobei der Wert des `i`-ten Parameters zurückgegeben wird. Zum Fixieren der Parameter in der neuen Funktion benutzen Sie

```
TF1::FixParameter(int i, float wert)
```

um den Wert des `i`-ten Parameters auf `wert` zu setzen. Benutzen Sie dann

```
TF1::GetRandom(),
```

um eine gemäß der so definierten Funktion verteilte Zufallszahl zu erhalten.

(iii) Erzeugen Sie N_{tot} neue Werte für $\cos\theta$ und füllen Sie diese in ein neues Histogramm. Führen Sie eine Anpassung der ursprünglichen Funktion $f(x;\alpha,\beta)$ an das neue $cos(\theta)$-Histogramm durch.

(iv) Ermitteln Sie als nächstes den χ^2-Wert für das erzeugte Histogramm, indem Sie die Formel

$$\chi^2_M = 2\sum_{i=1}^{N}\left(n_i \ln \frac{n_i}{\hat{\nu}_i}\right)$$

benutzen. Dabei sind N die Anzahl der Histogrammbins und $\hat{\nu}_i$ die Erwartungswerte der angepassten Funktion $f(x;\hat{\alpha},\hat{\beta})$, die gegeben sind durch

$$\hat{\nu}_i = N_{\text{tot}} \int_{x_i^{\min}}^{x_i^{\max}} f(x;\hat{\alpha},\hat{\beta})\mathrm{d}x.$$

Zum Integrieren der Funktion $f(x;\hat{\alpha},\hat{\beta})$ sollten Sie die Methode

```
TF1::Integral(int binMin,int binMax)
```

verwenden. Die untere Bingrenze des `i`-ten Bins erhalten Sie mittels

```
TH1::GetBinLowEdge(int i).
```

Schreiben Sie eine Schleife über alle Histogrammbins, in der Sie die Einzelbeiträge zu χ^2_M aufsummieren.

(v) Wiederholen Sie dieses Vorgehen 1000 mal und füllen Sie jedesmal den Wert von χ^2_M in ein Histogramm ein. Denken Sie daran, das neue Histogramm für die zufällig ermittelten Werte von $\cos\theta$ am Anfang jedes Zufallsexperiments wieder zurückzusetzen. Dies geschieht am einfachsten mit der Methode

```
TH1::Reset();
```

(vi) Stellen Sie das Histogramm am Bildschirm dar und überzeugen Sie sich davon, dass es einer χ^2-Verteilung folgt. Führen Sie, falls Sie noch Zeit haben, eine Anpassung einer χ^2-Funktion an das Histogramm durch. Stimmt die Anzahl der Freiheitsgrade mit der Erwartung überein?

Hausaufgaben

Aufgabe 62 *F-Test für zwei verschiedene Messungen* **7 Punkte**

Betrachten Sie zwei unabhängige Chi-Quadrat verteilte Variablen, u_1 und u_2, mit ν_1 und ν_2 Freiheitsgraden, d.h. u_1 ist gemaß $\chi^2(\nu_1)$ verteilt und u_2 nach gemaß $\chi^2(\nu_2)$. Dann ist die Variable F, definiert durch

$$F \equiv \frac{u_1/\nu_1}{u_2/\nu_2} \qquad 0 \leq F \leq \infty;\ \nu_1,\nu_2 > 0 \tag{1}$$

verteilt nach der WDF

$$f(F;\nu_1,\nu_2) = \frac{\Gamma\left(\frac{1}{2}(\nu_1+\nu_2)\right)}{\Gamma\left(\frac{1}{2}\nu_1\right)\Gamma\left(\frac{1}{2}\nu_2\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{1}{2}\nu_1} \frac{F^{\frac{1}{2}\nu_1-1}}{\left(1+\frac{\nu_1 F}{\nu_2}\right)^{\frac{1}{2}(\nu_1+\nu_2)}} \tag{2}$$

welche "F-Verteilung für (ν_1,ν_2) Freiheitsgrade"genannt wird (siehe Abb. 1).

Für zwei Datensätze $x_1,x_2,\ldots\ldots,x_n$, gaussverteilt nach $N(\mu_1,\sigma_1^2)$, und $y_1,y_2,\ldots\ldots,y_n$, gaussverteilt nach $N(\mu_2,\sigma_2^2)$, wobei die Mittelwerte μ_1 und μ_2 beider Verteilungen bekannt sind, ist die Größe

$$F = \frac{s_1}{s_2} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu_1)^2}{\frac{1}{m-1}\sum_{i=1}^{m}(y_i-\mu_2)^2} \tag{3}$$

durch die F-Verteilung mit (n,m) Freiheitsgraden verteilt, wenn $\sigma_1^2 = \sigma_2^2$. Daher kann das Verhältnis s_1/s_2 benutzt werden, um die Hypothese, dass beide Verteilungen die selbe Varianz ($H_0 : \sigma_1^2 = \sigma_2^2$) aufweisen, gegen die Hypothese, dass beide Varianzen verschieden sind ($H_1 : \sigma_1^2 > \sigma_2^2$), zu testen.

Kehren wir noch einmal zur Messung des Teilchenimpulses aus Aufgabe 60 zurück. Beide Messungen haben den selben Mittelwert $\mu_1 = \mu_2 = \mu_0$. Betrachten Sie hier nun folgende beiden Hypothesen für die Varianzen des inversen Impulses:

$$H_0 : \frac{1}{\sigma_1^2} = \frac{1}{\sigma_2^2}$$
$$H_1 : \frac{1}{\sigma_1^2} < \frac{1}{\sigma_2^2}$$

(i) Berechnen Sie den Wert von F für beide Messungen.

(ii) Wie viele Freiheitsgrade haben die Messungen?

(iii) Was ist der kritische Wert von F bei einer Signifikanz von 5%? Sollte der Test auf einer oder auf zwei Seiten durchgeführt werden?

(iv) Sind daher die Präzisionen der beiden Messungen miteinander konsistent?

Aufgabe 63 *Zählexperiment für eine Signal- und Untergrundmessung* **13 Punkte**

Betrachtet wird ein Experiment, dessen Ziel die Entdeckung eines neuen Teilchens oberhalb eines von momentanen Theorien vorhergesagten Untergrundes ist. Dabei könnte es sich beispielsweise um das Higgs-Boson oder auch um supersymmetrische Teilchen handeln, wobei die Untergrundvorhersage durch das Standardmodell der Teilchenphysik erfolgt.

Die zu betrachtenden Hypothesen sind also die Nullhypothese H_0, dass nur Ereignisse aus Untergrundprozessen gemessen wurden, sowie die Alternativhypothese H_1, dass sowohl Signal- als auch Untergrundereignisse beobachtet wurden.

Im Experiment wurde eine Gesamtanzahl x von Ereignissen aufgezeichnet. Weiterhin wurde eine andere, signalfreie kinematische Region definiert, aus der man die Normierung des Untergrundes bestimmen kann. In dieser Region wurden y Ereignisse gefunden. Das Verhältnis der Untergrundereignisse in der signalfreien Region zu denjenigen in der Signalregion sei τ. Die mittlere Anzahl der Untergrundereignisse in der Signalregion sei b und die mittlere Anzahl der Signalereignisse im Falle von Hypothese H_1 sei s.

(i) Stellen Sie die Likelihoodfunktionen für die Hypothesen H_0 und H_1 auf. Nehmen Sie dabei an, dass die Gesamtanzahlen von Ereignissen in Signal- und Kontrollregion jeweils Poissonverteilt sind.

(ii) Betrachten Sie jetzt die Schätzer für s und b unter der Hypothese H_1 ($\hat{s}$ bzw. $\hat{b}$) sowie den Schätzer für b unter der Nullhypothese, $\hat{\hat{b}}$.

a) Stellen Sie die Profile Likelihood λ auf.

b) Bestimmen Sie Ausdrücke für $\hat{s}$, $\hat{b}$ und $\hat{\hat{b}}$. Betrachten Sie dazu

$$\left.\frac{\partial L(H_0)}{\partial b}\right|_{\hat{\hat{b}}}, \tag{4}$$

sowie die die beiden gleichzeitigen Einschränkungen

$$\left.\frac{\partial L(H_1)}{\partial s}\right|_{\hat{s}} \quad \text{und} \quad \left.\frac{\partial L(H_1)}{\partial b}\right|_{\hat{b}}. \tag{5}$$

c) Berechnen Sie $q = -2\ln\lambda$ und zeigen Sie dadurch, dass die in der Vorlesung angegebene Gleichung

$$q = 2\left[x\ln x + y\ln y - (x+y)\ln\left(\frac{x+y}{1+\tau}\right) - y\ln\tau\right]$$

korrekt ist.

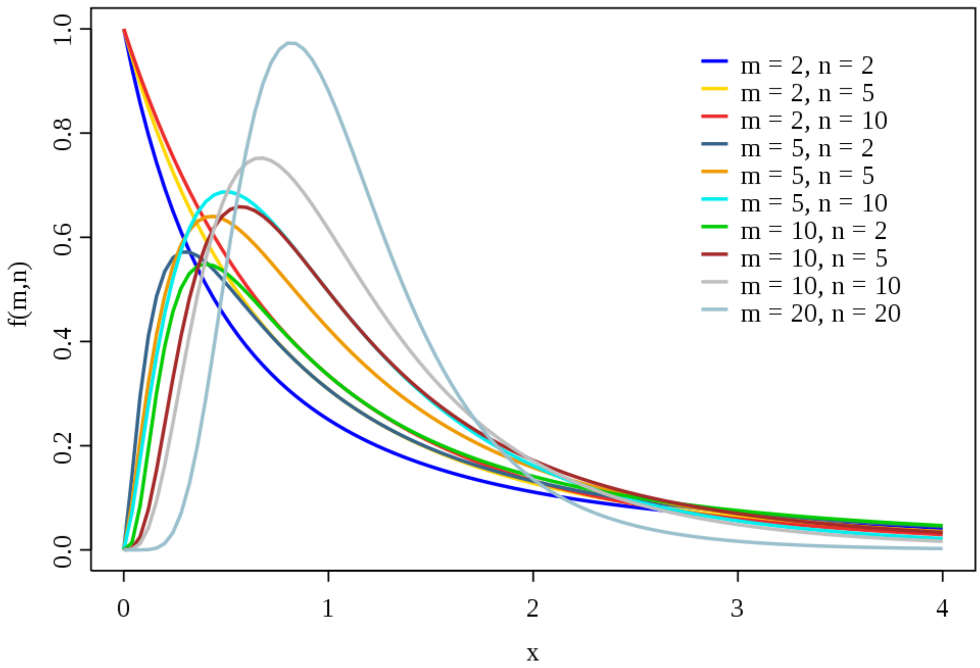

Abbildung 1: Die F-Verteilung