Albert-Ludwigs-Universität Freiburg Wintersemester 2011/2012

Statistische Methoden der Datenanalyse

Markus Schumacher

Übung XI

Markus Warsinsky

23.1.2012

Anwesenheitsaufgaben

Aufgabe 63 *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/warsinsk/sd_wise1112/ueb11/aufgabe63_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

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

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 1000mal 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 64 *Zählexperiment für eine Signal- und Untergrundmessung* **12 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

$$\left.\frac{\partial L(H_0)}{\partial b}\right|_{\hat{\hat{b}}}, \tag{1}$$

sowie die die beiden gleichzeitigen Einschränkungen

$$\left.\frac{\partial L(H_1)}{\partial s}\right|_{\hat{s}} \quad \text{und} \quad \left.\frac{\partial L(H_1)}{\partial b}\right|_{\hat{b}}. \tag{2}$$

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.

Aufgabe 65 *Zählexperiment für eine Signal- und Untergrundmessung 2* **8 Punkte**

Monte-Carlo-Studien von Proton-Proton-Kollisionen im ATLAS-Detektor haben gezeigt, dass der Wirkungsquerschnitt für $pp \to H+X \to \gamma\gamma+X$-Ereignisse, die die Ereignisselektion passieren, gegeben ist durch $\sigma_S = 25{,}4\,\mathrm{fb}$. Der Wirkungsquerschnitt für Untergrundereignisse, die dieselbe Ereignisselektion passieren, beträgt $\sigma_B = 947\,\mathrm{fb}$. In einer weiteren Analyse kann ein reiner Untergrunddatensatz mit einem Wirkungsquerschnitt von $\sigma_T = 10300\,\mathrm{fb}$ ausgewählt werden.

(i) Benutzen Sie die Relation

$$N = \mathcal{L}\sigma, \tag{3}$$

um die Anzahlen von Signal- (x), Untergrundereignissen (y) sowie die Anzahl von Ereignissen in der Seitenbandregion (τb) für eine integrierte Luminosität von $\mathcal{L} = 10\,\mathrm{fb}^{-1}$ auszurechnen.

(ii) Berechnen Sie die Schätzer für die Anzahl der Signalereignisse $\hat{s}$, der Untergrundereignisse $\hat{b}$ unter der Hypothese von Signal plus Untergrund , sowie den Schätzer $\hat{\hat{b}}$ auf die Anzahl der Untergrundereignisse in der Nur-Untergrund Hypothese.

(iii) Berechnen Sie die Größe

$$q = -2\ln\lambda. \tag{4}$$

(iv) Berechnen Sie daraus die Signifikanz des vorhergesagten Signals.

(v) Nehmen Sie nun an, dass die Untergrundrate in der Signalregion mit einer relativen Genauigkeit von $\Delta b/b = 5\%$ abgeschätzt werden kann. Wenn man annimmt, dass es sich dabei um einen Poissonfehler handelt, kann man eine effektive Seitenbandregion konstruieren mit Ereignisanzahl $y' = \tau' b$. Dazu setzt man die relative statistische Unsicherheit in der hypothetischen Seitenbandregion (Poissonfehler) gleich der relativen Genauigkeit der Untergrundvorhersage:

$$\frac{\sqrt{\tau' b}}{\tau' b} = \frac{\Delta b}{b} \Leftrightarrow \tau' = \frac{b}{(\Delta b)^2} \tag{5}$$

Berechnen Sie die Werte für τ' und y' für die effektive Seitenbandregion.

(vi) Benutzen Sie die Werte für τ' und y' sowie die ursprüngliche Anzahl von Ereignissen in der Signalregion x, um den neuen Wert für q und daher der Signifikanz für eine Messung mit einem Fehler von 5% auf die Untergrundvorhersage zu bekommen.