Albert-Ludwigs-Universität Freiburg Wintersemester 2011/2012

Statistische Methoden der Datenanalyse

Markus Schumacher

Übung IX

Markus Warsinsky

9.1.2012

Anwesenheitsaufgaben

Aufgabe 54 *χ^2-Anpassung an simulierte Z^0-Produktionsdaten*

In diesem Beispiel soll eine χ^2-Anpassung an einen simulierten Datensatz von Produktionswirkungsquerschnitten für Z^0-Bosonen in e^+e^--Kollisionen durchgeführt werden. Mit einem (Ihnen vorerst nicht zur Verfügung stehenden) ROOT-Makro `ZPeakGen.C` wurden Z^0-Boson-Produktionswirkungsquerschnitte als Funktion der Schwerpunktsenergie E_{cm} generiert und jeder Datenpunkt gemäß einer Gaussverteilung verschmiert. Die Gaussverteilung hatte dabei nicht für jedes E_{cm} dieselbe Breite. Der Fehler auf E_{cm} wird als vernachlässigbar angenommen. Die resultierenden Wirkungsquerschnitte befinden sich in einer Textdatei `/home/warsinsk/sd_ws1112/ueb9/data_aufgabe54.txt`. In dieser Übungsaufgabe sollen diese Pseudomessdaten in ROOT eingelesen werden, graphisch dargestellt, und eine χ^2-Anpassung einer Breit-Wigner-Verteilung durchgeführt werden. Gehen Sie dazu wie folgt vor:

(i) Das Einlesen der Messdaten kann relativ einfach in ein Objekt vom Typ `TGraphErrors` erfolgen. Sie können dazu einfach die folgende Syntax beim Erstellen des Objektes benutzen:

```
TGraphErrors graph=TGraphErrors("daten.txt","%lg %lg %lg");
```

Damit werden die Daten aus der Textdatei `daten.txt` eingelesen. Es wird dabei angenommen, dass die Messpunkte zeilenweise in der Textdatei stehen, und jeweils in jeder Zeile erst der x-Wert, dann der y-Wert und dann der Fehler auf y stehen.

(ii) Stellen Sie diesen TGraphErrors graphisch auf dem Bildschirm dar:

```
graph.Draw("AP");
```

(iii) Definieren Sie eine ROOT `TF1`-Funktion gemäß der Breit-Wigner-Verteilung

$$f(x) = A\frac{\frac{\Gamma}{2}}{(x-x_0)^2+\frac{\Gamma^2}{4}},$$

wobei Γ die Breite der Verteilung, x_0 die Position des Maximums (also die Masse des betreffenden Teilchens) und A eine Normierungskonstante ist.

(iv) Benutzen Sie diese Funktion für eine Anpassung an den `TGraphErrors`, der die simulierten Datenpunkte enthält, indem Sie den Befehl

```
graph.Fit("fitfunkname");
```

benutzen. Sie sollten vor der Ausführung der Anpassung die Parameter auf Startwerte setzen, die schon relativ nah an denen, die das χ^2 minimieren sind. Ansonsten kann es sein, dass die Anpassung aufgrund numerischer Probleme nicht konvergiert. Dies können Sie mit der Methode `TF1::SetParameter(int parnumber, float parvalue)` machen.

(v) Benutzen Sie danach die Methoden von `TF1`

```
TF1::GetChisquare()
TF1::GetNDF()
TF1::GetProb(),
```

um Informationen über die durchgeführte Anpassung zu erhalten (unter Umständen werden diese Werte auch am Bildschirm automatisch angezeigt).

(vi) Im folgenden soll die Monte-Carlo-Methode benutzt werden, um den sogenannten P-Wert zu berechnen, und mit demjenigen aus der ROOT-Anpassung zu vergleichen. Dazu sollen Zufallsexperimente basierend auf der soeben angepassten Funktion sowie den relativen Messfehlern des Datensatzes durchgeführt werden. Gehen Sie dazu wie folgt vor (am Besten fügen Sie alles nach der Anpassung ein):

a) Stellen Sie einen Zufallszahlengenerator vom Typ `TRandom3` bereit.

b) Stellen Sie ein eindimensionales Histogramm vom Typ `TH1F` bereit.

c) Stellen Sie eine weitere `TF1`-Funktion bereit, die für die Breit-Wigner-Anpassung in jedem Zufallsexperiment benutzt werden soll.

d) Stellen Sie einen weiteren `TGraphErrors` bereit. Um zu wissen, wieviele Punkte dieser haben muss, können Sie den ursprünglichen Graphen fragen:

```
int npoints=graph.GetN();
```

e) Um ein Zufallsexperiment zu erstellen schreiben Sie eine `for`-Schleife über alle Punkte des ursprünglichen `TGraphErrors`. Lesen Sie die x und y-Werte zum Beispiel für den `i`-ten Punkt ein wie folgt:

```
double xval, yval;
graph.GetPoint(i , xval, yval);
```

Den Fehler auf den y-Wert ermitteln Sie mit

```
double yerr=graph.GetErrorY(i);
```

Ermitteln Sie nun die Vorhersage der ursprünglich angepassten Funktion für diesen x-Wert:

```
double yfunk=funkname.Eval(xval);
```

Verschmieren Sie diesen neuen y-Wert mittels des Zufallszahlengenerators und der Methode `TRandom3::Gaus(float mean, float sigma)`. Wählen Sie für die Breite der gaussschen Verschmierung die relative Unsicherheit auf den Messwert (also `yerr/yval`), setzen den Mittelwert auf 1 und multiplizieren den Funktionswert mit der erhaltenen Zufallszahl. Ermitteln Sie ausserdem die Unsicherheit auf diesen Wert (`yfunk*yerr(yval)`). Setzen Sie dann den `i`-ten Punkt des neuen `TGraphErrors` auf die ermittelten Werte wie folgt:

```
newgraph.SetPoint(i, xval, newyval);
newgraph.SetPointError(i, 0., newyerr);
```

f) Passen Sie nach dieser `for`-Schleife die bereitgestellte Funktion an die Zufallsdaten an (unter Umständen müssen die Startwerte der Parameter angepasst werden). Wichtig: Passen Sie nicht die ursprüngliche Funktion an, ansonsten werden die bereits ermittelten Parameter überschrieben. Ermitteln Sie das minimale χ^2 und füllen Sie es in das bereitgestellte Histogramm.

g) Erstellen Sie auf diese Weise 1000 Zufallsexperimente, indem Sie das obige Vorgehen in eine zweite `for`-Schleife einschachteln.

h) Stellen Sie das Histogramm der χ^2-Werte auf dem Bildschirm dar.

i) Bestimmem Sie aus dem Histogramm der χ^2-Werte den sogenannten P-Wert dafür, ein gleich großes oder noch grösseres χ^2 wie in der ursprünglichen Anpassung zu erhalten. Benutzen Sie dazu die Methode von `TH1`

```
TH1::Integral(int binMin, int binMax),
```

mittels der man das Integral zwischen den Bins `binMin` und `binMax` erhält. Sie können die Methode von `TH1`

```
TH1::FindBin(float x)
```

benutzen, um die Binnummer für die Position `x` auf der x-Achse zu erhalten.

Hausaufgaben

Aufgabe 55 *Elektron-Positron-Kollisionen* **3 Punkte**

Die Anzahl von Ereignissen in Elektron-Positron-Kollisionen mit bestimmten kinematischen Eigenschaften kann als poissonverteilte Zufallsvariable angesehen werden. Nehmen Sie an, dass bei einer bestimmten integrierten Luminosität (dies wäre ein Datensatz für eine gegebene Strahlintensität) 3,9 Ereignisse aus bekannten Prozessen erwartet werden, aber 16 Ereignisse beobachtet wurden.

Berechnen Sie den P-Wert für die Hypothese, dass kein neuer Prozess zu der Zahl der Ereignisse beiträgt. Um Poissonwahrscheinlichkeiten zu summieren, kann die Relation

$$\sum_{n=0}^{m} P(n;\nu) = 1 - F_{\chi^2}(2\nu; n_{\text{dof}})$$

benutzt werden, wobei $P(n;\nu)$ die Poissonwahrscheinlichkeit für n Ereignisse bei einem Mittelwert von ν bezeichnet, sowie F_{χ^2} die kumulative χ^2-Verteilung mit $n_{\text{dof}} = 2(m+1)$ Freiheitsgraden ist (diese Werte können Sie tabelliert in Standardwerken vorfinden).

Aufgabe 56 *Teilchenidentifikation* **4 Punkte**

Durchquert ein geladenes Teilchen ein Gasvolumen, so erzeugt dies in dem Medium Ionisation. Die mittlere Menge hängt dabei von der Masse und dem Impuls der Teilchen ab. Daher können durch Ausarbeitung einer Testhypothese, basierend auf der gemessenen Ionisation im Gasvolumen bei bekanntem Impuls, verschiedene Teilchen identifiziert werden.

Betrachten Sie einen Strahl von Teilchen, welcher entweder Pionen oder Elektronen enthält. Man kann nun, als Funktionen der Ionisation t, die WDF $g(t|e)$ der Hypothese, dass es sich um ein Elektron handelt, sowie die WDF $g(t|\pi)$ der Hypothese, dass es sich um ein Pion handelt, aufstellen. Man wählt nun Elektronen aus, indem gefordert wird, dass $t \leq t_{cut}$:

Nehmen Sie an, dass die beiden WDFs gegeben sind durch eine um $t = 0$ zentrierte Gaussverteilung für die Elektronen und durch eine um $t = 2$ zentrierte Gaussverteilung für die Pionen. Beide Gaussverteilungen haben eine Standardabweichung von Eins. Ein Test zur Elektronselektion wird konstriert, imdem eine Ionisation $t \leq 1$ gefordert wird.

(i) Was ist die Signifikanz des Tests (d.h. die Wahrscheinlichkeit, Elektronen zu akzeptieren)?

(ii) Wie groß ist die Mächtigkeit des Tests gegen die Hypothese, dass das Teilchen ein Pion ist? Wie groß ist die Wahrscheinlichkeit, dass ein Pion als Elektron akzeptiert wird?

(iii) Betrachten Sie eine Strahlzusammensetzung von 99% Pionen und 1% Elektronen. Wie groß ist die Reinheit der durch $t \leq 1$ selektierten Auswahl?

Aufgabe 57 *Beweis des Neyman-Pearson-Lemmas* **6 Punkte**

Das Neyman-Pearson-Lemma besagt, dass der Test einer einfachen Hypothese H_0 bezüglich der einfachen Hypothese H_1 ein bester Test ist (also für gegebene Signifikanz α eine maximale Mächtigkeit $1 - \beta$ besitzt), wenn die kritische Region S_C im Stichprobenraum so gewählt wird, dass:

$$r = \frac{f(\vec{x}|H_0)}{f(\vec{x}|H_1)} \begin{cases} \leq c & \text{für } \vec{x} \in S_C \\ \geq c & \text{für } \vec{x} \notin S_C \end{cases}$$

Dabei ist c eine Konstante, die passend zu einem gegebenem α gewählt wird. r wird das Likelihoodverhältnis genannt.

Beweisen Sie diese Aussage. Gehen Sie dazu wie folgt vor:

(i) Nehmen Sie an, dass es eine weitere kritische Region S mit gleicher Signifikanz α gibt. Zerlegen Sie den Stichprobenraum in eine Region A, die nur in der kritischen Region S_C liegt, eine Region B, die nur in S liegt und eine weitere Region C, die die Schnittmenge zwischen S_C und S darstellt.

(ii) Zeigen Sie, dass das Integral über $f(\vec{x}|H_0)$ über die Regionen A und B identisch ist.

(iii) Stellen Sie Ungleichungen zwischen den Integralen über $f(\vec{x}|H_0)$ und $f(\vec{x}|H_1)$ separat in den Regionen A und B auf.

(iv) Benutzen Sie diese Ergebnisse, um zu zeigen, dass die Mächtigkeit für die kritische Region S_C, definiert über das Likelihoodverhältnis, immer größer gleich der Mächtigkeit für Region S ist.

Aufgabe 58 *Zerfallszeit eines Teilchens unter Verwendung des Likelihoodverhältnis* **7 Punkte**

In einem Experiment wird ein Satz von Zerfallszeiten $\vec{t} = (t_1,t_2,\ldots,t_n)$ eines Teilchens aufgenommen. Betrachten Sie einen Test der Nullhypothese $H_0 : \tau = 1$, wobei τ die wahre Lebensdauer der Teilchens ist, gegenüber der Alternativhypothese $H_1 : \tau > 1$.

(i) Wie lauten die Likelihoodfunktionen sowie der Schätzer der Lebensdauer $\hat{\tau}$ für n Messungen eines exponentiellen Zerfalls eines Teilchens mit Lebensdauer τ?

(ii) Wie lauten somit die zwei Likelihoodfunktionen $\alpha(\vec{x},\vec{\lambda}(\omega))$ für H_0 und für den vollen Parameterraum $\alpha(\vec{x},\vec{\lambda}(\Omega))$, gegeben durch $H_0 + H_1$?

(iii) Zeigen Sie, dass das Likelihoodverhältnis gegeben ist durch

$$T = \frac{\alpha(\vec{x},\hat{\vec{\lambda}}(\omega))}{\alpha(\vec{x},\hat{\vec{\lambda}}(\Omega))} = \bar{t}^n \exp\left(-n(\bar{t}-1)\right), \tag{1}$$

wobei $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$ der arithmetische Mittelwert der gemessenen Zerfallszeiten ist.

(iv) Für große n ist $\bar{t}$ gemäß der Gaußschen WDF $N(1,\frac{1}{n})$ verteilt. Berechnen Sie damit den kritischen Wert von $\bar{t}$ für eine Signifikanz $\alpha = 0{,}05$.

(v) Berechnen Sie den kritischen Wert T_{cr} des Likelihoodverhältnisses und zeigen Sie, dass $T_{\mathrm{cr}} \to 0$ für $n \to \infty$.