

Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

Übung X

14.01.2014, 15.01.2014

Anwesenheitsaufgaben

Aufgabe 49 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/slai/Statistics/PS10/data_aufgabe49.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 gaussischen 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) Bestimmen 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 50 Zerfallszeit eines Teilchens unter Verwendung des Likelihoodverhältnis

9 Punkte

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

- Wie lauten die Likelihoodfunktionen sowie der Schätzer der Lebensdauer $\hat{\tau}$ für n Messungen eines exponentiellen Zerfalls eines Teilchens mit Lebensdauer τ ?
- Wie lauten somit die Maxima der zwei Likelihoodfunktionen $L_{max}(\tau = 1, \vec{t})$ für H_0 und für den vollen Parameterraum $L_{max}(\tau \geq 1, \vec{t})$, gegeben durch $H_0 + H_1$?
- Zeigen Sie, dass das Likelihoodverhältnis gegeben ist durch

$$T = \frac{L_{max}(\tau = 1, \vec{t})}{L_{max}(\tau \geq 1, \vec{t})} = \bar{t}^n \exp(-n(\bar{t} - 1)), \quad (1)$$

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

- 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 $\alpha = 0.05$.
- Berechnen Sie den kritischen Wert T_{cr} des Likelihoodverhältnisses und zeigen Sie, dass $T_{cr} \rightarrow 0$ für $n \rightarrow \infty$.

Aufgabe 51 Studentische t -Verteilung

7 Punkte

Betrachten Sie zwei Zufallsvariablen: Die erste, x , sei nach einer Standard-Normalverteilung $N(0,1)$ und die zweite, u , nach der Chi-Quadrat-Verteilung mit ν Freiheitsgraden $\chi^2(\nu)$ verteilt. x und u seien unabhängig. Betrachten Sie nun die neue Zufallsvariable t , definiert als:

$$t \equiv \frac{x}{\sqrt{u/\nu}} \quad -\infty \leq t \leq \infty, \nu > 0$$

- Zeigen Sie, dass t gemäß der Wahrscheinlichkeitsdichtefunktion

$$f(t; \nu) = \frac{\Gamma(\frac{1}{2}(\nu + 1))}{\sqrt{\pi\nu}\Gamma(\frac{1}{2}\nu)} \cdot \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}}$$

verteilt ist. Diese wird auch Studentische t -Verteilung mit ν Freiheitsgraden genannt (siehe Abb. 1). Betrachten Sie dazu die kombinierte Wahrscheinlichkeitsdichtefunktion $f(x, u; \nu)$ für x und u , transformieren Sie auf die WDF $f(t, v; \nu)$ mit $t = \frac{x}{\sqrt{u/\nu}}$ und $v = u$ und marginalisieren Sie schließlich über v , um $f(t; \nu)$ als Randverteilung zu erhalten. Tipp: $\int_0^\infty x^n \exp(-a \cdot x) = \Gamma(n+1)/a^{n+1}$.

- Zeigen Sie, dass sich $f(t; \nu)$ für $\nu = 1$ als Cauchyverteilung ergibt. (Tipps: $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$.)

Aufgabe 52 Medizinische Wirksamkeit

4 Punkte

Eine neue medizinische Behandlung soll an 100 Patienten auf ihre Wirksamkeit überprüft werden. Wie beurteilt man, ob diese Behandlung überhaupt eine Wirksamkeit herbeiführt, wenn ohnehin 60% aller Patienten ohne Behandlung geheilt werden? Nutzen Sie ein Signifikanzniveau von 5%.

(Zur Berechnung der Mindestanzahl an erfolgreich behandelten Patienten kann die Normalapproximation verwendet werden.)

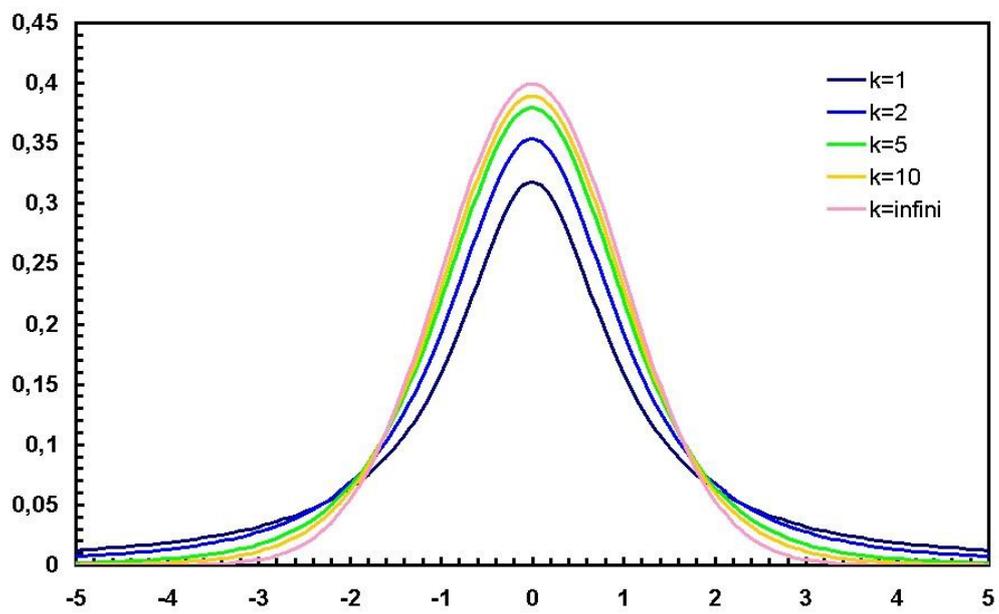


Abbildung 1: Die Studentsche t -Verteilung.