

# Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

## Übung VII

10.12.2013, 11.12.2013

### Anwesenheitsaufgaben

#### Aufgabe 36 'Binned' Maximum Likelihood Fit zu $e^+e^- \rightarrow \mu^+\mu^-$ mit Monte Carlo Statistik

Im Streuungsprozess  $e^+e^- \rightarrow \mu^+\mu^-$  folgt der Winkel  $\theta$  des  $\mu^+$  Teilchens der theoretischen Verteilung:

$$N(\cos \theta) \propto 1 + \alpha \cos \theta + \beta \cos^2 \theta. \quad (1)$$

Der Winkel  $\theta$  wird relativ zur Strahlachse gemessen. In der Übung sollen Sie einen 'Binned' Maximum Likelihood Fit durchführen, um die Parameter zu finden, welche zur Erstellung der Monte Carlo Ereignisse benutzt wurden. Ein solcher simulierter Datensatz befindet sich in `/home/slai/StatisticsCourse/PS7/MuMuEvGen.root`.

Um eine 'Binned' Maximum Likelihood Statistik zu erstellen betrachte man ein Histogramm mit  $N$  Bins, wobei jedes  $n_i$  Einträge  $\vec{n} = (n_1, \dots, n_N)$  enthält. Für eine Wahrscheinlichkeitsdichtefunktion (WDF)  $f(x; \vec{\theta})$  ist der Erwartungswert gegeben durch  $\vec{\nu} = (\nu_1, \dots, \nu_N)$ , wobei  $\nu_i(\vec{\theta})$  gegeben ist durch das Integral der WDF über die Breite des Bins

$$\nu_i(\vec{\theta}) = n_{tot} \int_{x_i^{min}}^{x_i^{max}} f(x; \vec{\theta}) dx \quad (2)$$

wobei  $x_i^{min}$  und  $x_i^{max}$  für die Grenzen jedes Bins  $i$  stehen.

Daher ist die gemeinsame WDF,  $f_g(n; \vec{\nu})$ , gegeben durch die Multinomialverteilung

$$f_g(n; \vec{\nu}) = \frac{n_{tot}!}{n_1! \dots n_N!} \left( \frac{\nu_1}{n_{tot}} \right)^{n_1} \dots \left( \frac{\nu_N}{n_{tot}} \right)^{n_N} \quad (3)$$

und demzufolge ist die log-Likelihoodfunktion für die gemeinsame WDF gegeben durch

$$\ln \mathcal{L}(\vec{\theta}) = \sum_{i=1}^N n_i \ln \nu_i(\vec{\theta}) \quad (4)$$

Dies ist eine Funktion, welche maximiert werden muss, um die zur Erstellung des Monte Carlo Datensatzes verwendeten Parameter  $\vec{\theta}$  zu finden. Beachten Sie diesbezüglich, dass  $n_{tot}$  nicht in der Berechnung der log-Likelihoodfunktion benötigt wird, da es sich hierbei um eine additive Konstante handelt.

Um die Parameter Ihrer Monte Carlo Verteilung zu berechnen gehen Sie folgende Schritte durch:

- (i) Zum Berechnen der Variablen  $\cos \theta$  findet sich in `/home/slai/StatisticsCourse/PS7/aufgabe36_i.C` ein ROOT-Skript. Kopieren Sie sich dieses Skript in ein Verzeichnis Ihrer Wahl. Öffnen Sie es in einem Texteditor, lassen es laufen, und versuchen Sie die Funktionsweise zu verstehen.

- (ii) Modifizieren Sie dieses Skript nun so, dass Sie die Werte von  $\cos\theta$  in ein Histogramm mit 200 Bins zwischen -1 und +1 einfüllen. Stellen Sie das Histogramm am Ende des Skripts graphisch auf dem Bildschirm dar.
- (iii) Definieren Sie eine TF1 Funktion der Form

$$f(x; \alpha, \beta) = \frac{1 + \alpha x + \beta x^2}{2 + \frac{2\beta}{3}}, \quad (5)$$

welche in dem Maximum Likelihood Fit benutzt werden soll. Dies können Sie mit folgendem Befehl bewerkstelligen:

```
TF1 fitFunk = TF1("fitFunk",
    "( 1.0 + [0]*TMath::Power(x,2) + [1]*x)/(2.0+2.0*[0]/3.0)" , -1., 1.);
```

Um die Parameter auf bestimmte Werte zu setzen, können Sie die Funktion `SetParameter` wie folgt benutzen:

```
fitFunk.SetParameter(0, 1.0);
fitFunk.SetParameter(1, 0.1);
```

- (iv) Definieren Sie - analog zu dem Maximum Likelihood Fit aus Aufgabe 33 - eine Schleife, um über mögliche Werte von  $\alpha$  zu iterieren. Definieren Sie innerhalb dieser Schleife eine weitere Schleife über möglichen Werte von  $\beta$ . Sinnvolle Werte für die Variation von  $\alpha$  sind zwischen 0 und 0.25 und für  $\beta$  zwischen 0.8 und 1.3. Nehmen Sie sowohl für  $\alpha$  als auch für  $\beta$  50 äquidistante Werte an (somit wird die Likelihood für 2500 Kombinationen ausgerechnet).
- (v) Erstellen Sie für jeden Wert von  $\alpha_i$  und  $\beta_i$  eine Schleife über die Bins in Ihrem  $\cos\theta$  Histogramm und berechnen Sie den log-Likelihood Wert. Gehen Sie wie folgt vor:

- Setzen Sie zunächst die Parameter in der bereitgestellten TF1-Funktion auf die gerade aktuellen Werte von  $\alpha_i$  und  $\beta_i$ .
- Erstellen Sie eine Schleife von  $i=1$  (das niedrigste Bin des Histogramms - Bin 0 - ist das Underflow Bin) bis  $i=\text{hist.GetNbinsX}()$ . Dies ist bereits die dritte und gleichzeitig innerste `for`-Schleife.
- Die Anzahl an Einträgen im  $i$ ten Bin erhalten Sie mit der TH1 Funktion `GetBinContent(int i)`
- Um Ihre Funktion über die Binbreite zu integrieren benutzen Sie die TF1 Funktion `Integral(float a, float b)`. `a` steht für die untere, `b` für die obere Integrationsgrenze. Um diese Werte aus dem  $\cos\theta$  Histogramm zu erhalten, benutzen Sie die TH1 Funktion `GetBinLowEdge(int i)`, welche die untere Grenze des  $i$ -ten Bins ausgibt.
- Summieren Sie die log-Likelihoodfunktion für alle Bins des Histogramms auf.

- (vi) Füllen Sie für jedes  $\alpha_i$  und  $\beta_i$  den log-Likelihood Wert in einen zweidimensionalen Graph. Dies funktioniert analog zu einem eindimensionalen Graphen:

```
TGraph2D likeGraph = TGraph2D(likepoints);
```

wobei `likepoints` die Gesamtanzahl an Kombinationen von  $\alpha$  und  $\beta$  ist, für die die Likelihood ausgewertet werden soll.

Jeder Punkt des Graphen kann gesetzt werden durch die `TGraph2D` Funktion:

```
SetPoint(int pointNumber, float alpha, float beta, float loglikelihood)
```

Denken Sie daran, sich `pointNumber` auszurechnen (dies ist nicht die Zählvariable in den Schleifen über die Werte von  $\alpha$  und  $\beta$ !).

- (vii) Zeichnen Sie den Graphen (log-Likelihood gegen  $\alpha$  und  $\beta$ ) unter Benutzung von

```
likeGraph.Draw("colz");
```

wobei die Option `"colz"` ROOT die Anweisung gibt, einen Surface-Plot zu erstellen. Weitere Optionen des `Draw`-Befehls können im 'ROOT User's Guide' in der `TGraph2D` Sektion gefunden werden.

- (viii) Ermitteln Sie aus den erzeugten Graph eine Abschätzung für  $\hat{\alpha}$  und  $\hat{\beta}$  und deren Standardabweichungen  $\sigma_{\hat{\alpha}}$  und  $\sigma_{\hat{\beta}}$ . Dazu könnten folgende Befehle nützlich sein:

```
float maximum=likeGraph.GetHistogram().GetMaximum();
```

gibt das Maximum des Graphen zurück.

```
likeGraph.GetHistogram().SetMinimum(maximum-0.5);
```

setzt für die Darstellung die minimale Zeichenebene auf das Maximum des Graphen minus 0.5.

- (ix) Vergleichen Sie die Werte Ihres Fits mit denen des RooT Fitting Paketes, indem Sie die Histogramm-Anpassungsfunktion

```
hist.Fit("functionName", "LI");
```

benutzen, wobei "functionName" der Name der WDF Funktion ist und die Option "LI" einen Likelihood Fit ausführt, indem die gegebene Funktion über jedes Bin des Histogramms integriert wird.

# Hausaufgaben

## Aufgabe 37 *Messung der Polarisationsasymmetrie*

8 Punkte

Eine wichtige Messung der Paritätsverletzung wurde durchgeführt, indem die inelastische Streuung von polarisierten Elektronen an einem Deuteriumtarget vermessen wurde. Die Paritätsverletzung führt dabei zu einer Differenz im Streuwirkungsquerschnitt zwischen links- und rechts-händig polarisierten Elektronen. Die Messgröße war dabei die Polarisationsasymmetrie  $\alpha$ , die gegeben ist durch

$$\alpha = \frac{n_R - n_L}{n_R + n_L},$$

wobei  $n_L$  ( $n_R$ ) die Wirkungsquerschnitte für links- (rechts)-händige Elektronen bei der inelastischen Streuung an einem Deuteriumtarget sind.

Die Anzahl der beobachteten Streuereignisse bei links- bzw. recht-händigen Elektronen ist eine poissonverteilte Zufallsvariable,  $n_L$  bzw.  $n_R$ , mit Mittelwerten  $\nu_L$  bzw.  $\nu_R$ .

- (i) Benutzen Sie die Tatsache, dass die beobachtete Anzahl von Streuereignissen poissonverteilt ist, um einen "Maximum Likelihood"-Schätzer  $\hat{\alpha}$  für die Asymmetrie  $\alpha$  aufzustellen.  
*Hinweis:* Wie kann man die Wahrscheinlichkeit, dass man  $n_R$  und  $n_L$  Ereignisse beobachtet, abhängig von  $\alpha$  ausdrücken?
- (ii) Benutzen Sie Fehlerfortpflanzung (und nehmen Sie dabei an, dass die Messungen der Streuereignisse für links- bzw. rechtshändige Elektronen unabhängig voneinander sind), um zu zeigen dass die Standardabweichung  $\sigma_{\hat{\alpha}}$  auf diesen Schätzer  $\sigma_{\hat{\alpha}} = \sqrt{\frac{1-\hat{\alpha}^2}{n_{tot}}}$  geschrieben werden kann (wobei  $n_{tot} = n_R + n_L$ ).
- (iii) Bevor das Experiment aufgebaut wurde, wurde erwartet, dass die Asymmetrie ungefähr  $10^{-4}$  betragen würde. Welche Gesamtzahl von Streuereignissen musste man dafür beobachten, um die Asymmetrie mit einer relativen Genauigkeit von 10% zu messen?

## Aufgabe 38 *Maximum Likelihood für eine Gleichverteilung*

8 Punkte

Gegeben seien  $N$  Zufallsvariablen  $\vec{x} = \{x_i\}$  einer Gleichverteilung mit Wahrscheinlichkeitsdichtefunktion:

$$f(x) = \begin{cases} 1/\beta & \text{für } 0 \leq x \leq \beta \\ 0 & \text{sonst} \end{cases}.$$

- (i) Wie lautet der Schätzer  $\hat{\beta}_{ML}$  nach der Maximum-Likelihood-Methode?
- (ii) Zeigen Sie, dass dieser Schätzer  $\hat{\beta}_{ML}$  für die Fälle  $N = 1$  und  $N = 2$  nicht erwartungstreu ist.  
*Hinweis:* Für den  $N = 2$  Fall soll man die bedingte Wahrscheinlichkeit  $f(x_2|x_2 > x_1)$  berechnen.
- (iii) Ist der Schätzer  $\hat{\beta}_{ML}$  erwartungstreu für den Grenzfall  $N \rightarrow \infty$ ?

## Aufgabe 39 *Kombinierte Likelihoodfunktion*

4 Punkte

Gegeben seien zwei unabhängige Messungen  $s \pm \sigma$  und  $c \pm \sigma$  von  $\sin \theta$  beziehungsweise  $\cos \theta$ . Nehmen Sie an, dass die gauss'schen Fehler der Messungen gleich sind. Was ist der Maximum-Likelihood Schätzer  $\hat{\theta}$ ?