

# Statistische Methoden der Datenanalyse

Markus Schumacher

## Übung IX

Matthew Beckingham und Henrik Nilsen

7.01.2010

### Computerübung

#### Aufgabe 37 *Kleinste Quadrat und Maximum Likelihood Anpassung an $e^+e^- \rightarrow \mu^+\mu^-$ Monte Carlo Daten*

Betrachtet wird ein Prozess aus der Teilchenphysik, in dem sich ein Elektron und ein Positron gegenseitig vernichten und ein positiv und ein negativ geladenes Myon entstehen ( $e^+e^- \rightarrow \mu^+\mu^-$ ). Die Winkelverteilung in diesem Prozess ist gegeben durch folgende Wahrscheinlichkeitsdichtefunktion:

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

wobei  $\theta$  der Winkel zwischen dem positiv geladenen Myon und der  $z$ -Achse ist. In dieser Übung werden Sie die Methoden der kleinsten Quadrate und der Maximum Likelihood anwenden, um diese Verteilung an Monte Carlo Daten anzupassen.

Das Makro `EvGen_i.C` stellt Code bereit, mittels dessen Sie die obige WDF benutzen und die Vierervektoren der beiden Endzustandsmyonen  $P_{1,2} = (E_{1,2}, \vec{p}_{1,2})$  generieren können. Die Werte der Energie  $E$  und die kartesischen Komponenten der Impulse  $p_x, p_y, p_z$  werden dann in einem Root-Tree (`TTree`) innerhalb eines Root-Files (`TFile`) abgespeichert.

Betrachten Sie das Makro und versuchen Sie zu verstehen, wie es funktioniert. Lassen Sie es dann laufen, generieren Sie einige Ereignisse und versuchen Sie die Verteilung von  $\cos \theta$  anzupassen.

- (i) Lassen Sie das Makro `EvGen_i.C` laufen, um einen Monte Carlo Datensatz für die weitere Analyse zu erstellen.
- (ii) Im Makro `2DFit_i.C` findet sich ein Beispiel, wie man das erstellte `TFile` öffnet, auf die Daten im `TTree` zugreift und eine Schleife über alle Ereignisse innerhalb des `TTree` laufen lässt. Benutzen Sie diesen Code, um über den `TTree` zu laufen und ein Histogramm mit den Werten von  $\cos \theta$  zu füllen.
- (iii) Berechnen Sie die Maximum Likelihood Schätzer  $\hat{\alpha}$  und  $\hat{\beta}$ . Gehen Sie dazu wie folgt vor:
  - a) Schreiben Sie eine Schleife über alle möglichen Werte für  $\alpha$ . Schreiben Sie innerhalb dieser Schleife eine weitere Schleife über alle Werte für  $\beta$  für jeden möglichen Wert für  $\alpha$ .
  - b) Laufen Sie für jedes Wertepaar  $(\alpha_i, \beta_i)$  über die  $N$  Ereignisse in dem `TTree` und berechnen Sie die Likelihoodfunktion

$$\log L(\cos \theta; \alpha_i, \beta_i) = \sum_{j=0}^N \log \left( \frac{1 + \alpha_i \cos \theta_j + \beta_i \cos^2 \theta_j}{2 + \frac{2\beta_i}{3}} \right). \quad (2)$$

- c) Füllen Sie die erhaltenen Werte der Likelihood in einen zweidimensionalen Graph (`TGraph2D`), der  $L$  in Abhängigkeit von  $\alpha$  und  $\beta$  darstellt. Dieser kann durch
 

```
TGraph2D* likeGraph = new TGraph2D();
```

 definiert werden und wird mittels der Methode
 

```
TGraph2D::SetPoint(Int_t point, Double_t x, Double_t y, Double_t z)
```

gefüllt, wobei `point` der Index des  $n$ -ten Punktes in dem Graph ist, sowie  $x$ ,  $y$  und  $z$  die  $x$ ,  $y$  und  $z$  Werte des  $n$ -ten Punktes sind.

- d) Benutzen Sie diesen Graph, um die Maximum Likelihood Schätzer  $\hat{\alpha}$  und  $\hat{\beta}$  sowie die jeweiligen Fehler zu erhalten.
- (iv) Als nächstes sollen Schätzer mittels der gebinnten Maximum Likelihood Methode und der gebinnten Kleinsten Quadrate erhalten werden, indem das erstellte Histogramm von  $\cos \theta$  benutzt wird. Gehen Sie dazu wie folgt vor:

- a) Definieren Sie eine eindimensionale Anpassungsfunktion (TF1) von der Form  $f(\cos \theta; \alpha, \beta)$ , mit der die gebinnten Maximum Likelihood und  $\chi^2$ -Anpassungen durchgeführt werden sollen.
- b) Schreiben Sie wie im vorherigen Aufgabenteil eine Schleife über alle möglichen Werte für  $\alpha$ . Schreiben Sie innerhalb dieser Schleife eine weitere Schleife über alle Werte für  $\beta$  für jeden möglichen Wert für  $\alpha$ .
- c) Laufen Sie für jedes Wertepaar  $(\alpha_i, \beta_i)$  über die Bins des  $\cos \theta$  Histogramms und berechnen Sie die Likelihood

$$\log L(\cos \theta; \alpha_i, \beta_i) = \sum_{j=1}^N n_j \log \nu_j(\alpha_i, \beta_i), \quad (3)$$

wobei  $\nu_j(\alpha_i, \beta_i)$  die erwartete Anzahl von Einträgen im Bin  $j$  ist. Diese ist gegeben durch

$$\nu_j(\alpha_i, \beta_i) = n_{tot} \int_{\cos \theta_{min}}^{\cos \theta_{max}} f(\cos \theta; \alpha_i, \beta_i) d(\cos \theta) \quad (4)$$

Einige nützliche RooT-Funktionen für Histogramme sind:

`TH1::GetEntries()`

Liefert die Gesamtanzahl der Einträge in einem Histogramm,

`TH1::GetBinLowEdge(Int_t i)`

Gibt den  $x$ -Wert der unteren Grenze von Bin  $i$  aus.

`TH1::GetBinContent(Int_t i)`

ergibt die Anzahl der Einträge im Bin  $i$ .

Eine weitere nützliche Methode für RooT-Funktionen ist:

`TF1::Integral(Double_t a, Double_t b)`

Diese integriert die Funktion zwischen den Werten  $a$  and  $b$ .

- d) Berechnen Sie gleichzeitig für jedes Wertepaar  $(\alpha_i, \beta_i)$  die  $\chi^2$ -Werte, indem sie die folgende Gleichung benutzen:

$$\chi^2(\alpha_i, \beta_i) = \sum_{j=1}^N \frac{(n_j - \nu_j(\alpha_i, \beta_i))^2}{\sigma_j^2} \quad (5)$$

wobei  $\sigma_j$  der Fehler auf die Anzahl der Einträge im Bin  $j$  ist. Eine weitere nützliche RooT-Funktion ist:

`TH1::GetBinError(Int_t i)`

Diese gibt den Fehler auf die Anzahl der Einträge im Bin  $i$  zurück.

- e) Füllen Sie die Werte der Likelihood und von  $\chi^2$  in zweidimensionale Graphen (`TGraph2D`) in Abhängigkeit von  $\alpha$  und  $\beta$ .
- f) Finden Sie daraus die Schätzwerte für  $\alpha$  und  $\beta$  sowie ihre Fehler mittels der Maximum Likelihood und der Methode der kleinsten Quadrate.
- g) Vergleichen Sie die so gefundenen Werte mit denen, die man erhält, wenn man RooT direkt benutzt, um  $f(\cos \theta; \alpha, \beta)$  an das  $\cos \theta$  Histogramm anzupassen. Benutzen Sie dazu die Methode

`TH1::Fit(const char* formula, Option_t* option)`

wobei `formula` der Name der TF1 Anpassungsfunktion ist. Setzen die die Variable `option` auf "I", um eine Kleinste Quadrat Anpassung durchzuführen, bei der die Anpassungsfunktion über die Breite jedes Bins integriert wird. Benutzen Sie die Option "IL", um eine gebinnte Maximum Likelihood Anpassung durchzuführen.