

Blatt 11

Abgabe von Marian Bruns und Kai Brügge

Aufgabe 31

a)

Für den Ereignisvektor gilt

$$\vec{g} = \begin{pmatrix} (1 - \epsilon) \cdot f_1 + \epsilon \cdot f_2 \\ \epsilon \cdot f_1 + (1 - \epsilon) \cdot f_2 \end{pmatrix}$$

Der Zusammenhang zwischen \mathbf{A} , \vec{g} und \vec{f} lautet

$$\vec{f} = \mathbf{A} \cdot \vec{g}$$

Damit ergibt sich für die Antwortmatrix \mathbf{A}

$$\mathbf{A} = \begin{pmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{pmatrix}$$

Da 20% der Messwerte verloren gehen, muss die Matrix noch mit 0.8 multipliziert werden:

$$\mathbf{A} = 0.8 \cdot \begin{pmatrix} 1 - \epsilon & \epsilon \\ \epsilon & 1 - \epsilon \end{pmatrix}$$

b)

In Matrix Schreibweise gilt für \vec{f}

$$\vec{f} = \mathbf{A}^{-1} \cdot \vec{g}$$

Nach Regel von Sarrus für 2x2 Matrizen folgt aus \mathbf{A} die Inverse

$$\mathbf{A}^{-1} = \frac{1}{0.8 \cdot (1 - 2\epsilon)} \cdot \begin{pmatrix} 1 - \epsilon & -\epsilon \\ -\epsilon & 1 - \epsilon \end{pmatrix}$$

Mit der Determinante $\det(\mathbf{A}) = 0.8 \cdot (1 - 2\epsilon)$.

c)

BVB-Formel:

$$\mathbf{V}[f] = \mathbf{A}^{-1} \cdot \mathbf{V}[g] \cdot (\mathbf{A}^{-1})^T$$

Da die Einträge von \vec{g} unkorreliert sind ist $\mathbf{V}[g]$ diagonal. Wir nehmen an, dass die gemessenen Werte gleich den Erwartungswerten der jeweiligen Verteilung sind, wodurch sich für Poissonverteilte g_i (Erwartungswert=Varianz) ergibt:

$$\mathbf{V}[g] = \begin{pmatrix} g_1 & 0 \\ 0 & g_2 \end{pmatrix}$$

Dann ergibt sich:

$$\mathbf{V}[f] = \frac{1}{0.64 \cdot (1 - 2\epsilon)^2} \begin{pmatrix} (1 - \epsilon)^2 g_1 + \epsilon^2 g_2 & (\epsilon^2 - \epsilon)(g_1 + g_2) \\ (\epsilon^2 - \epsilon)(g_1 + g_2) & (1 - \epsilon)^2 g_2 + \epsilon^2 g_1 \end{pmatrix}$$

d)

$$\vec{f} = \begin{pmatrix} 254.84 \\ 207.97 \end{pmatrix}$$

$$\mathbf{V}[f] = \begin{pmatrix} 399.63 & -81.08 \\ -81.08 & 339.09 \end{pmatrix}$$

Die Fehler auf f_1 und f_2 sind die Quadratwurzeln der Diagonaleinträge dieser Matrix:

$$\sigma_{f_1} = 19.99$$

$$\sigma_{f_2} = 18.41$$

Der Korrelationskoeffizient beträgt

$$\rho = -0.22$$

e)

$$\vec{f} = \begin{pmatrix} 327.5 \\ 133.75 \end{pmatrix}$$

$$\mathbf{V}[f] = \begin{pmatrix} 3868.75 & -3459.34 \\ -3459.34 & 3626.56 \end{pmatrix}$$

Die Fehler auf f_1 und f_2 sind die Quadratwurzeln der Diagonaleinträge dieser Matrix:

$$\sigma_{f_1} = 62.20$$

$$\sigma_{f_2} = 60.22$$

Korrelationskoeffizient:

$$\rho = -0.92$$

Offensichtlich sind f_1 und f_2 wesentlich stärker korreliert als zuvor. Gleichzeitig werden die Fehler auf f_1 und f_2 deutlich größer. Das ist eine logische Konsequenz der Tatsache, dass die Zuordnung der Ereignisse bei wachsendem ϵ immer schlechter wird.

f)

Für $\epsilon = 0.5$ wird **A** singulär und ist somit nicht invertierbar ($\det = 0$). Anschaulich bedeutet $\epsilon = 0.5$, dass ein Signal genauso wahrscheinlich richtig wie falsch zugeordnet wird, d.h. die Messdaten enthalten kaum Information.

Aufgabe 32

a)

Der Tree mit dem Namen **Signal_MC_Akzeptanz** wird eingelesen und mit der Variable **AnzahlHits** entfaltet. Die Entfaltung soll im test modus ausgeführt werden mit einem Daten/MC Verhältniss von 0.9. Die wichtigen Zeilen aus der **parameter.config** lauten entsprechend

```
mode: test

pseudo_data_fraction: 0.9

source_file_moca: ./Blatt7_TRUEEE.root
roottree_moca: Signal_MC_Akzeptanz
```

b)

Die Anzahl der Bins wird über das Schlüsselwort **number_bins** festgelegt. Der erlaubte Bereich der Zielvariable soll zwischen 1 und 300 TeV liegen. Logarithmisch also von $\log(1) = 0$ bis $\log(300) \approx 2.5$. Die Verteilung soll in 9 Bins verteilt werden. Die Zeilen aus der **parameter.config** dazu sehen wie folgt aus.

```
branch_x: Energie log
limits_x: 0 2.5
```

```
number_bins: 9
```

Die Energie Verteilung folgt näherungsweise einer Exponentialverteilung. Durch das logarithmieren wird die Verteilung annähernd linear. Dadurch werden die Ereignisse gleichmäßiger auf die Bins verteilt.

c)

Es werden die Observablen **AnzahlHits**, **x**, **y** eingelesen. Die Ortsinformationen zu logarithmieren macht keinen Sinn deshalb wird nur **AnzahlHits** logarithmisch eingelesen. Damit die die Grammatik der Konfigurationssyntax auch ja nicht Kontextfrei bleibt muss noch die Anzahl der eingelesenen Variablen und die Anzahl der Bins für jede Variable mit angegeben werden.

```
number_all_variables: 3
```

```
branch_y: AnzahlHits log
number_y_bins: 9
```

```
branch_y: x
number_y_bins: 9
```

```
branch_y: y
number_y_bins: 9
```

d)

Die Korrelationsplots nach der Entfaltung.

Offensichtlich erkennt man nur bei der Variable **AnzahlHits** eine Korrelation. Im folgenden wird nur noch diese Variable benutzt.

e)

Der Übersichtsplot zeigt die Werte verschiedener Teststatistiken bei verschiedenen Parametern. In unserem Beispiel wurden über die Parameter wie folgt geloopt:

```
number_deg_free: 4
max_number_deg_free: 8
```

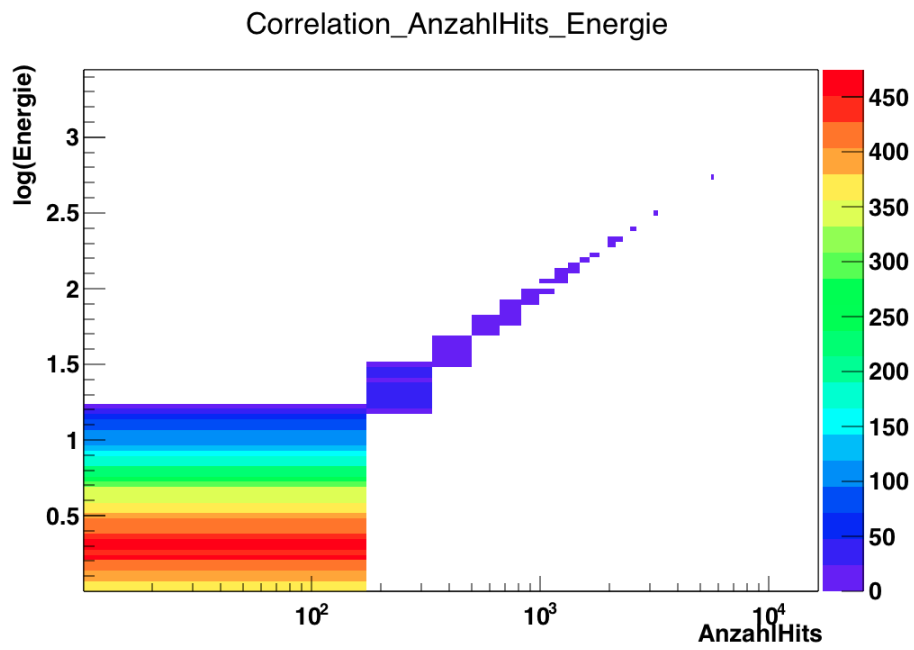


Figure 1: Energie gegen AnzahlHits logarithmisch

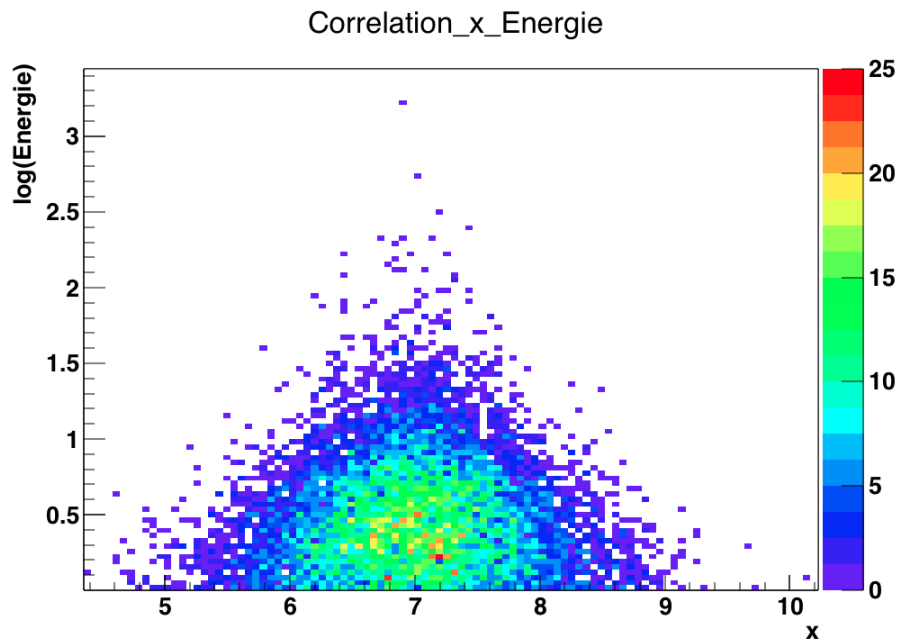


Figure 2: Energie gegen x

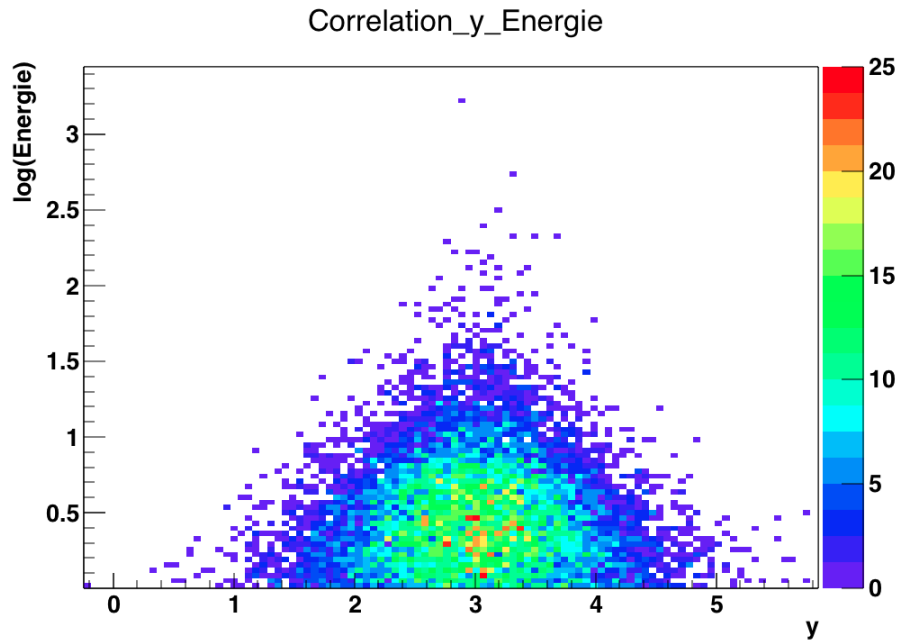


Figure 3: Energie gegen y

```
number_knots: 4
max_number_knots: 8
```

Wir nehmen das Parameterpaar welches möglichst kleine Werte für den Zweistichproben KS-Test und den χ^2 -Test hat. Ob der übersichtsplot die p-Werte der Tests oder deren Teststatistik anzeigt wird aus der Dokumentation nicht ersichtilich

In der Dokumentation steht, dass ein Wert nahe 1 für den KS-Test ein gute übereinstimmung zeigt. Das scheint aber nicht dem output der aktuellen Version zu entsprechen. Tatsächlich scheint die Berechnung des KS-Werts nicht zu funktionieren da nur 0 oder 1 als Werte im Plot vorhanden sind.

Per optischem Vergleich der Testplots scheint das Wertepaar (5,5) gut zu sein. Das entsprechende Testresultat sieht wie folgt aus

f)

Im Bereich von 1 bis 300 TeV sind 24926 Signaleinträge im Baum. Aus der Integration über die Enrgieverteilung mit einem Index von -2.7 folgt

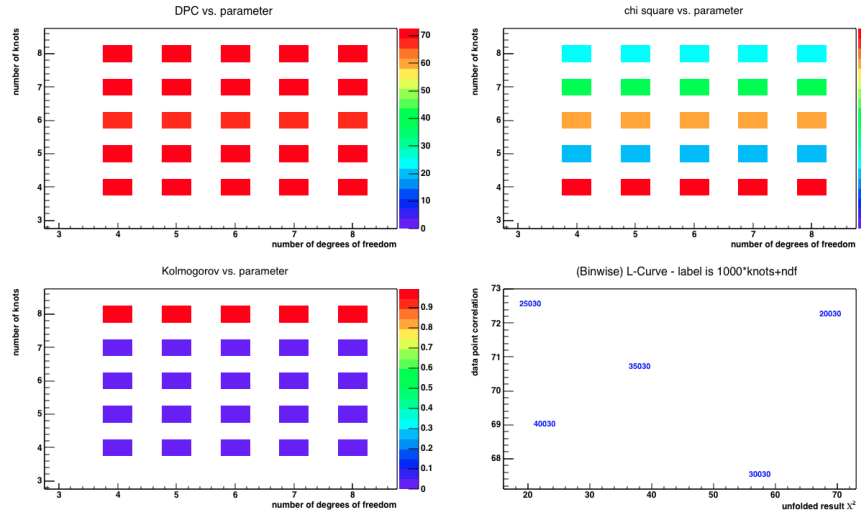


Figure 4: Übersichts plot von TRUEE

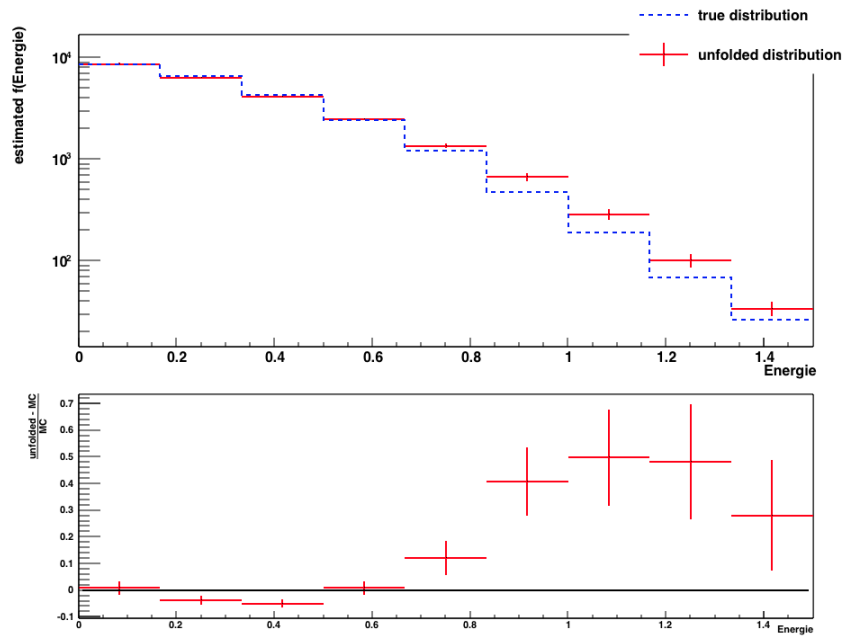


Figure 5: Vergleich der Verteilungen

$$N \int_1^{300} E^{-2.7} dE \stackrel{!}{=} 24926 \implies N \approx 42390$$

Da unser input logarithmiert war müssen wir die entsprechende Umkehrfunktion zum 10er Logarithmus berücksichtigen. Die Zeile aus der Konfigurationsdatei lautet also

```
mc_func: 42390*pow(10,x)^(-2.7)
```

g)

Der Pull-Mode ist auf meinem System nicht ausführbar. Es erscheint nur folgende Fehlermeldung

```
Illegal instruction: 4
```