

# Statistische Methoden der Datenanalyse WS 2017/18

Prof. Dr. Ulrich Landgraf

## Aufgabenblatt 7 vom 13.12.2017

### Aufgabe 1 (Schätzer für eine Landauverteilung) (5 Punkte)

Wenn ein geladenes Teilchen durch Materie fliegt, verliert es seine Energie durch viele Stöße mit den Atomen. Die dabei pro Atom abgegebene Energie wird für hochenergetische Teilchen durch die sogenannte Landau-Verteilung beschrieben. Sie kann nur über ein Integral definiert werden:

$$f(u) = \frac{1}{\pi} \int_0^{\infty} e^{-t \ln t - ut} \sin(\pi t) dt,$$

kann aber in ROOT als `TMath::Landau` berechnet werden. In dieser Definition liegt der wahrscheinlichste Wert (“most probable value”=mpv) beinahe bei Null (genau bei -0.22278) und die Verteilung hat eine Breite nahe bei Eins. Mit der Transformation

$$u = \frac{x - \text{mpv}}{\text{width}}$$

kann man einen anderen Mittelwert und eine andere Breite erhalten.

a) Veranschaulichen Sie sich zunächst diese Landauverteilung, indem Sie sie für verschiedene Werte von `mpv` und `width` mit der Klasse `TF1` zeichnen.

b) Generieren Sie (mit `Random::Landau`) 100 unabhängige Sätze von je 1000 Zufallsereignissen mit `mpv = 2` und `width = 0.5` und testen Sie, ob man mit dem Schätzer

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

einen stabilen Schätzwert für den wahrscheinlichsten Wert der Verteilung erhält. Tragen Sie dazu die Werte von  $\bar{x}$  in ein Histogramm ein. Warum ist  $\bar{x}$  kein guter Schätzer?

c) Wir modifizieren nun  $\bar{x}$ , indem wir jeweils die Zufallsereignisse mit den höchsten Werten von  $\bar{x}$  weglassen, d.h. die Werte, die im “Schwanz” der Verteilung liegen (sogenanntes “truncated mean”). Man kann z.B. die 10% der Zufallswerte weglassen, die am höchsten liegen, oder auch 20%. Wir wollen den entsprechenden Schätzer  $\hat{x}_{10}$  bzw.  $\hat{x}_{20}$  nennen. Berechnen und histogrammieren Sie  $\hat{x}_{10}$  und  $\hat{x}_{20}$  ebenso wie  $\bar{x}$  in Teilaufgabe b). ( Benutzen Sie `TMath::Sort!`)

Bemerkung: Diese Methode des “truncated mean” wird in der Teilchenphysik verwendet, um aus wenigen (typischerweise weniger als 20) Messwerten des Energieverlusts eine Schätzung des wahrscheinlichsten Energieverlustes zu erhalten, der (zusammen mit der Energie des Teilchens) eine Identifikation der Teilchenart erlaubt. Das “truncated mean” kann aber auch in anderen Situationen sinnvoll sein, in denen eine Verteilung einen langen “Schwanz” besitzt; so etwa, wenn man den Durchschnittsverdienst in verschiedenen Ländern vergleichen will, der durch ein paar wenige Superreiche verzerrt werden kann.