Statistische Tests in ROOT

VL ROOT Datenanalyse - Jörg Marks

Maurice Morgenthaler

17.05.2019

Universität Heidelberg

Table of contents

Grundlagen

Was sind statistische Tests?

Grundlegende Problemstellung der Physik:
Untersuchung von Hypothesen anhand aufgenommener Daten

Es existieren zu einem Datenset zumindest zwei Hypothesen:

- Nullhypothese H_0: Grundannahme. Wird erst bei ausreichend geringer Wahrscheinlichkeit falsch zu liegen verworfen
- Alternativhypothese H_1: Die Hypothese, die man testen möchte. Bei entsprechend überzeugender Sachlage wird sie angenommen.

Was heißt das? Braucht Quantifizierung, anhand man entscheidet ob H_0 widerlegt $\Rightarrow$ Teststatistik t mit Verteilung g(t|H)
$\Rightarrow$ Wähle kritischer Punkt t_c mit Hinblick auf mögliche Fehler

Was sind statistische Tests?

Typ I Fehler: Ablehnung von H_0

obwohl wahr. Wahrscheinlichkeit:

$$\int_{[t_c}^{\infty} g(t|H_0)dt = \alpha$$

- α wird Test *size* genannt
  $\Rightarrow$ sollte klein sein

Typ II Fehler: Annahme von H_0
obwohl falsch. Wahrscheinlichkeit:

$$\int_{-\infty}^{t_c[} g(t|H_1)dt = \beta$$

- $1 - \beta$ wird Test *power* genannt
  $\Rightarrow$ sollte groß sein

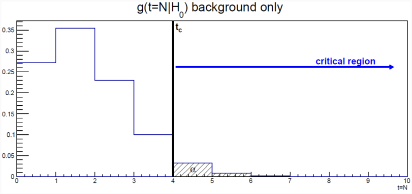

Figure 1: Darstellung von α

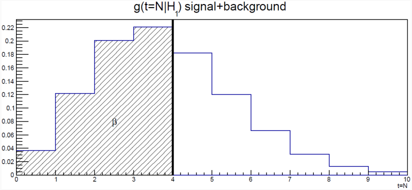

Figure 2: Darstellung von β

Was sind statistische Tests?

Der Testverlauf:

- Definiere H_0 und H_1
- Wähle Teststatistik t und bestimme dessen Verteilung $g(t|H_0)$
- Wähle size α mit Hinblick auf Typ I und Typ II Fehler
  $\Rightarrow$ oft $\alpha = 0.05$
- Bestimme t aus den observierten Daten
  $\Rightarrow$ verwerfe H_0 falls $t_{obs} > t_c$

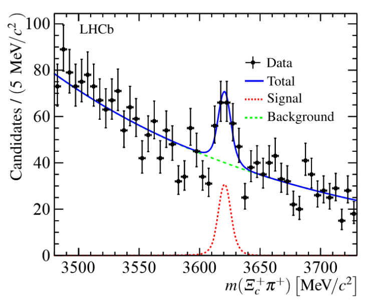

Figure 3: Beispiel Anwendung für statistische Tests: Signal über Background

Was sind statistische Tests?

Noch ein verwandtes Konzept: *p-Wert*

$$p = \int_{t_{obs}}^{\infty} g(t|H)dt$$

- Quantifiziert, wie wahrscheinlich, das H durch statistische Fluktuation verworfen
- Sind Wahrscheinlichkeiten für Schwanz der Teststatistik
  $\Rightarrow$ Vergleich mit Normalverteilungs-Schwanz ergibt *Signifikanz Z*

$$Z = \Phi^{-1}(1 - p), \quad \Phi : \text{CDF Normalverteilung}$$

In High Energy Physics:

- Hinweis/Evidence: Z = 3 σ (p < 0.00135)
- Observation: Z = 5 σ (p $< 2.867 \times 10^{-7}$)

Statistische Tests in ROOT

χ^2-Test

In ROOT implementiert als *TH1::Chi2Test*

Charakteristik:

- Oft Prüfe ob zwei Histogramme der selben Verteilung folgen
- Für diskrete und kontinuierliche Verteilungen verwendbar
- Ergebnis von Binweite abhängig
- Gilt für hohe ANzahl an Elementen
- Gibt einen p-Wert aus, an dem man Signifikanz Z festlegen kann
- p-Wert aus allgemeiner χ^2 Funktion berechnet:

$$\text{Test Statistik } \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

- Mit O_i: gemessener Wert, E_i: Erwartungswert, in Bin i

Im folgenden Prozedur für zwei nicht gewichtete Histogramme erklärt

χ^2-Test

- Gesamtanzahl an Elementen in jeweiligen Histogrammen gegeben durch

$$N = \sum_{i=1}^{r} n_i \qquad M = \sum_{i=1}^{r} m_i$$

- Annahme: Die Histogramme folgen der gleichen Verteilung $\Rightarrow$ Wahrscheinlichkeiten für Event in Bin i zu sein p_i sind für beide Histogramme gleich und es gilt:

$$\sum_{i=1}^{r} p_i = 1$$

- Die Verteilungen auf die Bins folgen Poisson:

$$\frac{e^{-Np_i}(Np_i)^{n_i}}{n_i\,!} \qquad \frac{e^{-Mp_i}(Mp_i)^{m_i}}{m_i\,!}$$

χ^2-Test

- Die durch die Daten gegebenen erwarteten Binwahrscheinlichkeiten $\hat{p}_i$ sind:

$$\hat{p}_i = \frac{n_i + m_i}{N + M}$$

- Ist der sog. *maximum likelihood estimator* (mehr dazu später in der Vorlesung)
- Einsetzen der Terme der Histogramme in allgemeine χ^2 Form:

$$\chi^2 = \sum_{i=1}^{r} \frac{(n_i - N\hat{p}_i)^2}{N\hat{p}_i} + \sum_{i=1}^{r} \frac{(m_i - M\hat{p}_i)^2}{M\hat{p}_i} = \frac{1}{MN} \sum_{i=1}^{r} \frac{(Mn_i - Nm_i)^2}{n_i + m_i}$$

- Folgt einer χ^2_{r-1} Verteilung (für viel Statistik)

Kolmogorov-Smirnov-Test

In ROOT implementiert als *TH1::Kolmogorov*

Charakteristik:

- Kann wie der χ^2-Test dazu genutzt werden, um zu überprüfen, ob zwei Verteilungen equivalent sind
  $\Rightarrow$ Zuverlässiger als χ^2-Test für wenig Statistik
- Nur für kontinuierliche Verteilungen
- Für binned und unbinned Daten implementiert
- Binned Data: Positiver Bias
  $\Rightarrow$ Sinkt mit hoher Binzahl

Kolmogorov-Smirnov-Test

- Gegeben: N Datenpunkte, von klein nach groß sortiert
- Basiert auf *Empirischen Verteilungsfunktionen* (ECDF):

$$F_{X,N} = \frac{n(i)}{N}$$

- n(i): Anzahl an Punkten kleiner als Punkt Y_i

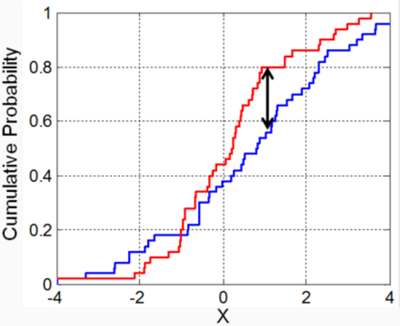

Figure 4: Beispiel der ECDF

- Teststatistik: $D = sup_i |F_{X,N}(Y_i) - F_{Y,M}(Y_i)|$
- H_0: gleiche Verteilung abgelehnt, falls ...

$$D > \sqrt{\frac{(N+M)\ln(\frac{2}{a})}{2NM}} = D_c$$

Anderson-Darling-Test

Implementiert als: *TH1::AndersonDarlingTest*

Charakteristik:

- Modifizierung des Kolmogorov-Smirnov-Test
- Gewichtet die Funktionsausläufer stärker
- Kritischer Wert abhängig von Verteilung (für 1 Sample)
  $\Rightarrow$ Vorteil: Test sensitiver
  $\Rightarrow$ Nachteil: Wert muss für alle Verteilungen berechnet werden
- Nur für kontinuierliche Verteilungen

Anderson-Darling-Test

Test Statistik:

- Für N Datenpunkte Y_N, M Datenpunkte X_M und dem kombinierten Sample Z_{N+M} sortiert von niedrig zu hoch

$$A = \frac{1}{NM} \sum_{i=1}^{N+M} (N_i Z_{(N+M-i)})^2 \frac{1}{iZ_{(N+M-i)}}$$

- N_i: Anzahl Punkte in Y_N, die kleiner als der i-te Wert in $Z_{(N+M)}$
- Kritischer Wert abhängig von N, M und test size α

N	M	α = 0.01	α = 0.05	α = 0.10
∞	∞	3.857	2.492	1.933

Table 1: Kritischer Wert um für Normalverteilung zu testen

Übungsaufgabe

Übungaufgabe

1. Lade die 1000 Events von Daten.txt herunter (https://github.com/MMorgenthaler/ROOT_Marks)
2. Welche der folgenden Verteilungen folgen die Daten?
   - Normalverteilung [TRandom::Gaus(mean = 0., sigma = 1.)]
   - Landau-Verteilung [TRandom::Landau(mu = 0., sigma = 1.)]
   - Exponential-Verteilung [TRandom::Exp(tau = 1.)]

Anmerkungen:
Benutze folgende Optionen:

- Chi2Test: ”UUCHI2”
- KolmogorovTest: ”M”
- AndersonDarlingTest: ”T”

Benutze den kritischen Wert für Normalverteilungen

References

[1] Webpage: Anderson-darling test. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm. Accessed: 2019-05-16.

[2] Webpage: Kolmogorov-smirnov goodness-of-fit test. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm. Accessed: 2019-05-16.

[3] Webpage: Th1 class reference. https://root.cern.ch/root/html608/classTH1.html. Accessed: 2019-05-16.

[4] Webpage: Chi-square goodness of fit test. `https://stattrek.com/chi-square-test/goodness-of-fit.aspx`. Accessed: 2019-05-16.

[5] Sonja Engmann and Denis Cousineau. Comparing distributions: the two-sample anderson–darling test as an alternative to the kolmogorov–smirnov test. *Journal of Applied Quantitative Methods*, 6:1–17, 09 2011.

[6] Sebastian Neubert. Statistical methods in particle physics - lecture notes. `https://uebungen.physik.uni-heidelberg.de/vorlesung/20182/smipp`. Accessed: 2019-03-13.

[7] Anthony Pettitt. A two-sample anderson–darling rank statistic. *Biometrika*, 63:161–168, 04 1976. doi: 10.1093/biomet/63.1.161.

[8] Frank C. Porter. Testing Consistency of Two Histograms. 2008.

Quellen iii

[9] **Bild 1,2,5:** https://uebungen.physik.uni-heidelberg.de/vorlesung/20182/smipp

[10] **Bild 3:** http://lhcb-public.web.cern.ch/lhcb-public/Images2018/XicPiPeak_s.png

[11] **Bild 4:** https://upload.wikimedia.org/wikipedia/commons/3/3f/KS2_Example.png

Back-Up: p-Wert

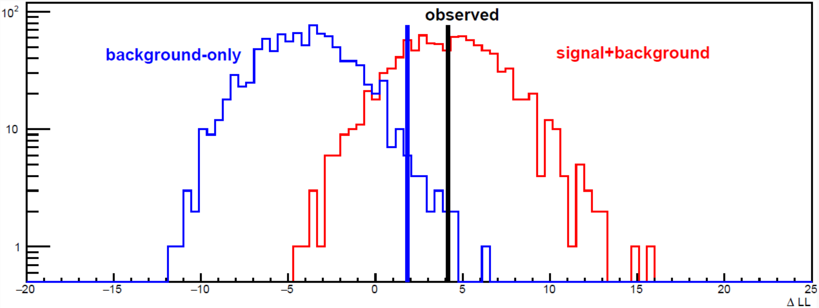

Figure 5: Darstellung des p-Werts

Back-Up: Estimator eines Poisson

Likelihood Funktion des Poisson:

$$L(\lambda; x_1, ..., x_N) = \prod_{i=1}^{N} \exp(-\lambda) \frac{1}{x_i!} \lambda^{x_i}$$

Log-Likelihood Funktion des Poisson:

$$l(\lambda; x_1, ..., x_N) = -N\lambda - \sum_{i=1}^{N} \ln(x_i!) + \ln(\lambda) \sum_{i=1}^{N} x_i$$

Maximum Likelihood Estimator:

$$\hat{\lambda}_N = \frac{1}{N} \sum_{i=1}^{N} x_i$$