

# Statistische Methoden der Datenanalyse

Markus Schumacher, Stan Lai, Florian Kiss

## Übung X

15.01.2013, 18.01.2013

### Anwesenheitsaufgaben

#### Aufgabe 57 Vergleich von Messungen einer gaussverteilten Variablen

Betrachten Sie den Fall, Sie hätten einen Satz von  $N$  Messungen einer gaussverteilten Variablen  $\vec{x} = (x_1, x_2, \dots, x_N)$  aufgenommen, wobei  $\vec{x}$  gemäß  $f_G(x; \mu_0, \sigma_0)$  verteilt sei. In dem vorliegenden Beispiel sollen Sie zwei verschiedene Hypothesentests betrachten, um sowohl den Mittelwert als auch Varianz Ihrer Messungen mit der erwarteten Verteilung  $f_G(x; \mu_0, \sigma_0)$  zu vergleichen.

Das Makro `/home/slai/StatisticsCourse/PS10/aufgabe_57_anfang.C` beinhaltet Code, welcher einen Satz von  $M$  Experimenten generiert, jeweils mit  $N$  Messungen einer gaussverteilten Variablen. Jede Messung wird in ein Histogramm gefüllt, welches am Ende angezeigt wird.

- (i) Vergleichen Sie zuerst den Mittelwert der generierten Messdaten mit der Gaussverteilung, welche Sie dazu verwendeten, die Messungen zu erstellen.
- a) Nehmen Sie an, Sie kennen den Mittelwert  $\mu$  wie auch  $\sigma$  der Gaussfunktion, die zum Generieren der Daten benutzt wurde. Um zu prüfen, ob Ihre Daten den Mittelwert  $\mu = \mu_0$  besitzen, berechnen Sie für jedes Experiment die Teststatistik

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N}}$$

und füllen Sie diese in ein bereitgestelltes Histogramm. Diese Variable sollte nach der gaussischen WDF  $f_G(t; 0, 1)$  verteilt sein. Überzeugen Sie sich davon, indem Sie die Methode

```
hist.Fit("gaus");
```

verwenden, um eine Gaussverteilung an Ihr Histogramm von  $t$  anzupassen. Stimmen die angepassten Parameter mit der Erwartung überein?

- b) Nehmen Sie nun an, Sie würden lediglich den Mittelwert  $\mu$ , jedoch nicht die Breite der den Messungen zugrunde liegenden Gaussverteilung kennen. Folglich prüfen Sie, ob Ihre Daten den Mittelwert  $\mu = \mu_0$  besitzen, indem Sie für jedes Experiment die Teststatistik

$$t' = \frac{\bar{x} - \mu_0}{s/\sqrt{N}}$$

berechnen, wobei die Standardabweichung  $s$  gegeben ist durch

$$s^2 = \frac{1}{N-1} \sum (x_i - \bar{x})^2 = \frac{N}{N-1} (\overline{x^2} - \bar{x}^2)$$

Füllen Sie Ihre Werte von  $t'$  in ein Histogramm. Die Variable  $t'$  sollte entsprechend einer Studentischen  $t$ -Verteilung  $f_t(t; N-1)$  mit  $N-1$  Freiheitsgraden verteilt sein. Überzeugen Sie sich davon, dass dies der Fall ist, indem Sie eine Studentische  $t$ -Verteilung an Ihr Histogramm anpassen. Verwenden Sie

```
TF1 tFit=TF1("tFit", "[1]*TMath::Student(x, [0])", -50., 50.);
```

um eine Anpassungsfunktion in Form einer Studentischen  $t$ -Verteilung bereitzustellen. Der `[0]`-te Parameter steht für die Anzahl der Freiheitsgrade. Passen Sie diese Funktion an Ihr Histogramm von  $t'$  an (`thist.Fit("tFit");`).

- (ii) Als nächstes vergleichen Sie die Breite der generierten Daten mit der der Gaussverteilung, die zum Generieren der Messungen benutzt wurde.

Gehen Sie davon aus, dass Ihnen der Mittelwert  $\mu$ , jedoch nicht die Breite der Gaussverteilung bekannt ist. Um nun zu prüfen, ob die Daten die Breite  $\sigma = \sigma_0$  aufweisen, berechnen Sie für jedes Experiment die Teststatistik

$$t'' = \frac{(N-1)s^2}{\sigma_0^2}$$

und füllen Sie diese in ein weiteres Histogramm. Die Variable  $t''$  sollte nach einer  $\chi^2$  WDF  $f_{\chi^2}(t''; N-1)$  mit  $N-1$  Freiheitsgraden verteilt sein. Überzeugen Sie sich davon, dass dies der Fall ist, indem Sie eine  $\chi^2$ -Verteilung an Ihr Histogramm anpassen. Verwenden Sie

```
TF1 tChi2Fit = TF1("tChi2Fit", "[0]*(1.0/(TMath::Power(2,[1]/2.0)
*TMath::Gamma([1]/2.0)))
*TMath::Power(x,([1]/2.0)-1.0)
*TMath::Exp(-x/2.0)", 0., 50.);
```

um eine  $\chi^2$ -Verteilung zur Anpassung bereitzustellen. Der [0]-te Parameter steht für den Normierungsfaktor und der [1]-te für die Anzahl an Freiheitsgraden. Passen Sie diese Funktion mittels `tthist.Fit("tChi2Fit");` an Ihr Histogramm von  $t''$  an.

- (iii) Wie verändern sich die Verteilungen von  $t$ ,  $t'$  und  $t''$ , wenn Sie einen systematischen Fehler hinzufügen, der alle Messungen um einen Wert von 1 erhöht, in der Art

$$\vec{x} \rightarrow \vec{x}' = (x_1 + 1, x_2 + 1, \dots, x_N + 1)?$$

- (iv) Wie verändern sich die Verteilungen von  $t$ ,  $t'$  und  $t''$ , wenn Sie einen gaussischen Fehler mit Standardabweichung  $\sigma = 0.1$  als zusätzliche Verschmierung zu allen Messungen hinzufügen?

# Hausaufgaben

**Aufgabe 58** *Zerfallszeit eines Teilchens unter Verwendung des Likelihoodverhältnis* **8 Punkte**

In einem Experiment wird ein Satz von Zerfallszeiten  $\vec{t} = (t_1, t_2, \dots, t_n)$  eines Teilchens aufgenommen. Betrachten Sie einen Test der Nullhypothese  $H_0 : \tau = 1$ , wobei  $\tau$  die wahre Lebensdauer des Teilchens ist, gegenüber der Alternativhypothese  $H_1 : \tau > 1$ .

- (i) Wie lauten die Likelihoodfunktionen sowie der Schätzer der Lebensdauer  $\hat{\tau}$  für  $n$  Messungen eines exponentiellen Zerfalls eines Teilchens mit Lebensdauer  $\tau$ ?
- (ii) Wie lauten somit die Maxima der zwei Likelihoodfunktionen  $L_{max}(\tau = 1, \vec{t})$  für  $H_0$  und für den vollen Parameterraum  $L_{max}(\tau \geq 1, \vec{t})$ , gegeben durch  $H_0 + H_1$ ?
- (iii) Zeigen Sie, dass das Likelihoodverhältnis gegeben ist durch

$$T = \frac{L_{max}(\tau = 1, \vec{t})}{L_{max}(\tau \geq 1, \vec{t})} = \bar{t}^n \exp(-n(\bar{t} - 1)), \quad (1)$$

wobei  $\bar{t} = \frac{1}{n} \sum_{i=1}^n t_i$  der arithmetische Mittelwert der gemessenen Zerfallszeiten ist.

- (iv) Für große  $n$  ist  $\bar{t}$  gemäß der Gaußschen WDF  $N(1, \frac{1}{n})$  verteilt. Berechnen Sie damit den kritischen Wert von  $\bar{t}$  für eine Signifikanz  $\alpha = 0.05$ .
- (v) Berechnen Sie den kritischen Wert  $T_{cr}$  des Likelihoodverhältnisses und zeigen Sie, dass  $T_{cr} \rightarrow 0$  für  $n \rightarrow \infty$ .

**Aufgabe 59** *Studentsche  $t$ -Verteilung - Teil I* **7 Punkte**

Betrachten Sie zwei Zufallsvariablen: Die erste,  $x$ , sei nach einer Standard-Normalverteilung  $N(0,1)$  und die zweite,  $u$ , nach der Chi-Quadrat-Verteilung mit  $\nu$  Freiheitsgraden  $\chi^2(\nu)$  verteilt.  $x$  und  $u$  seien unabhängig. Betrachten Sie nun die neue Zufallsvariable  $t$ , definiert als:

$$t \equiv \frac{x}{\sqrt{u/\nu}} \quad -\infty \leq t \leq \infty, \nu > 0$$

- (i) Zeigen Sie, dass  $t$  gemäß der Wahrscheinlichkeitsdichtefunktion

$$f(t; \nu) = \frac{\Gamma(\frac{1}{2}(\nu + 1))}{\sqrt{\pi\nu}\Gamma(\frac{1}{2}\nu)} \cdot \frac{1}{\left(1 + \frac{t^2}{\nu}\right)^{\frac{1}{2}(\nu+1)}}$$

verteilt ist. Diese wird auch Studentsche  $t$ -Verteilung mit  $\nu$  Freiheitsgraden genannt (siehe Abb. 1). Betrachten Sie dazu die kombinierte Wahrscheinlichkeitsdichtefunktion  $f(x, u; \nu)$  für  $x$  und  $u$ , transformieren Sie auf die WDF  $f(t, v; \nu)$  mit  $t = \frac{x}{\sqrt{u/\nu}}$  und  $v = u$  und marginalisieren Sie schließlich über  $v$ , um  $f(t; \nu)$  als Randverteilung zu erhalten. Tipp:  $\int_0^\infty x^n \exp(-a \cdot x) = \Gamma(n + 1)/a^{n+1}$ .

- (ii) Zeigen Sie, dass sich  $f(t; \nu)$  für  $\nu = 1$  als Cauchyverteilung ergibt. (Tipps:  $\Gamma(1/2) = \sqrt{\pi}$ ,  $\Gamma(1) = 1$ .)

**Aufgabe 60** *Studentsche  $t$ -Verteilung - Teil II* **5 Punkte**

Die Studentsche  $t$ -Verteilung kann dazu benutzt werden, um auf einem Datensatz eine Nullhypothese  $H_0$  zu testen.

Gegeben sei eine Stichprobe vom Umfang  $n$  aus einer Gaussverteilung  $N(\mu, \sigma^2)$ . Falls  $\sigma$  bekannt ist, ist die Verteilung für

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (2)$$

eine Gaussverteilung  $N(0,1)$ . Wenn  $\sigma^2$  jedoch nicht bekannt ist, dann ist  $t$  gegeben durch:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad (3)$$

mit der Stichprobenvarianz  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ . In diesem Fall ist  $t$  nach der Studentischen  $t$ -Verteilung mit  $n - 1$  Freiheitsgraden verteilt.

Betrachten Sie als Beispiel die Messung eines monoenergetischen Strahls von Teilchen mit Impuls  $P_0 = 24.90 \text{ GeV}/c$ . Dieser trifft auf eine Blasenkammer und durch Messung der Krümmung entlang der Teilchen spur wird der inverse Impuls  $1/P_i$  bestimmt. Nehmen Sie an, dass  $1/P$  für 20 Teilchen durch zwei verschiedene Detektoren  $A$  und  $B$  mit den Ergebnissen  $1/P_A = (40.12 \pm 0.46) \times 10^{-3} (\text{GeV}/c)^{-1}$  und  $1/P_B = (40.25 \pm 0.25) \times 10^{-3} (\text{GeV}/c)^{-1}$  gemessen wurde.

Um zu testen, ob beide Messungen mit der Bestimmung des inversen Impulses der einfallenden Teilchen,  $1/P_0$ , konsistent sind, sollten Sie diese beiden Hypothesen betrachten:

$$H_0 : \frac{1}{P_i} = \frac{1}{P_0}$$
$$H_1 : \frac{1}{P_i} \neq \frac{1}{P_0}$$

- (i) Was sind, unter Hinzunahme von Gleichung 3, die Werte von  $t$  für beide Messungen?
- (ii) Wie viele Freiheitsgrade hat jede Messung?
- (iii) Nutzen Sie die zur Verfügung gestellte Tabelle, um die Grenze der kritischen Region mit einer Signifikanz von  $\alpha = 0.05$  zu finden. Bedenken Sie hierbei, dass Sie einen beidseitigen Test durchführen. Wieso muss dieser Test auf zwei Seiten durchgeführt werden?
- (iv) In Bezug auf den inversen Impuls der einfallenden Teilchen: Sind beide Messungen damit konsistent?

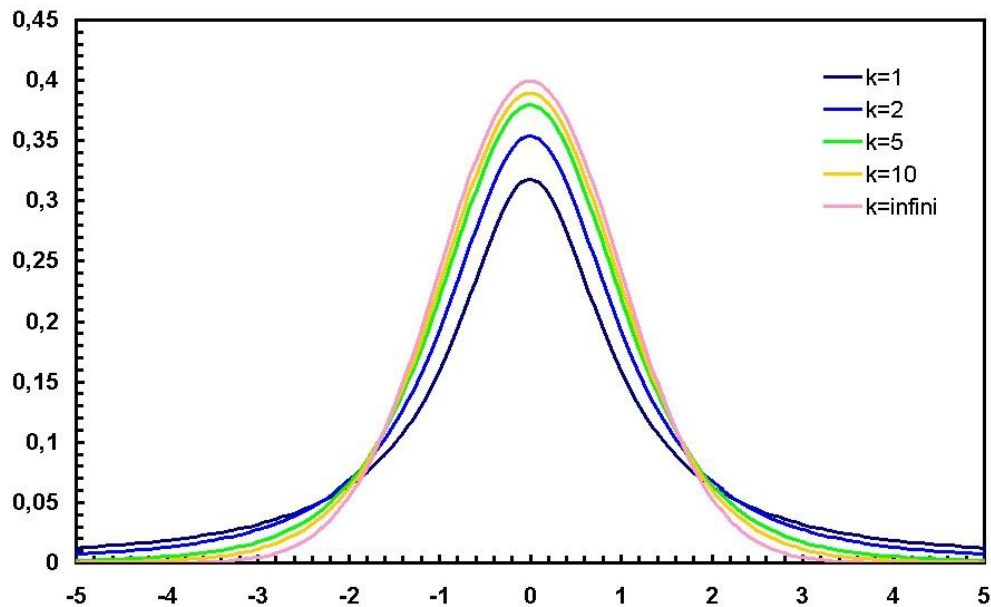


Abbildung 1: Die Studentsche  $t$ -Verteilung.

TABLE 7.2  
CRITICAL VALUES OF  $t$   
For various values of confidence levels and  $n$

Confidence (2 tailed)	60%	80%	90%	95%	98%	99%
(1 tailed)	80%	90%	95%	97.5%	99%	99.5%
$n = 1$	1.376	3.078	6.314	12.706	31.820	63.651
2	1.061	1.886	2.920	4.303	6.965	9.925
3	0.978	1.638	2.353	3.182	4.541	5.841
4	0.941	1.533	2.132	2.776	3.747	4.604
5	0.920	1.476	2.015	2.571	3.365	4.032
6	0.906	1.440	1.943	2.447	3.143	3.707
7	0.896	1.415	1.895	2.365	2.998	3.499
8	0.889	1.397	1.860	2.306	2.896	3.355
9	0.883	1.383	1.833	2.262	2.821	3.250
10	0.879	1.372	1.812	2.228	2.764	3.169
11	0.876	1.363	1.796	2.201	2.718	3.106
12	0.873	1.356	1.782	2.179	2.681	3.055
13	0.870	1.350	1.771	2.160	2.650	3.012
14	0.868	1.345	1.761	2.145	2.624	2.977
15	0.866	1.341	1.753	2.131	2.602	2.947
16	0.865	1.337	1.746	2.120	2.583	2.921
17	0.863	1.333	1.740	2.110	2.567	2.898
18	0.862	1.330	1.734	2.101	2.552	2.878
19	0.861	1.328	1.729	2.093	2.539	2.861
20	0.860	1.325	1.725	2.086	2.528	2.845
21	0.859	1.323	1.721	2.080	2.518	2.831
22	0.858	1.321	1.717	2.074	2.508	2.819
23	0.858	1.319	1.714	2.069	2.500	2.807
24	0.857	1.318	1.711	2.064	2.492	2.797
$\infty$	0.842	1.282	1.645	1.960	2.326	2.576

Abbildung 2: Kritische Werte für  $t$ .