11. Übungsblatt zur Vorlesung
Statistische Methoden der Datenanalyse
Abgabe: 24.01.2019 23:59

WS 2018/2019
Prof. W. Rhode

Zeit	Raum	Abgabe im Moodle; Mails mit Betreff: [SMD1819]
Di. 10-12	CP-03-150	tobias.hoinka@udo.edu, felix.geyer@udo.edu und jan.soedingrekso@udo.edu
Di. 16-18	CP-03-150	simone.mender@udo.edu und alicia.fattorini@udo.edu
Mi. 10-12	CP-03-150	mirco.huennefeld@udo.edu und kevin3.schmidt@udo.edu

Aufgabe 32: χ^2-*Test* **5 P.**

In einem Experiment werden 7 verschiedene Energiedifferenzen mit den Werten [1]

$$31{,}6\,\text{meV},\quad 32{,}2\,\text{meV},\quad 31{,}2\,\text{meV},\quad 31{,}9\,\text{meV},$$
$$31{,}3\,\text{meV},\quad 30{,}8\,\text{meV},\quad 31{,}3\,\text{meV}$$

mit jeweils einem Fehler von 0,5 meV gemessen.

a) Hypothese A sagt einen Wert von 31,3 meV für diese Messgröße voraus. Machen Sie einen χ^2-Test und entscheiden Sie, ob die These bei 5 % gewählter Signifikanz verworfen werden muss, oder nicht.

b) Wie a), aber mit der Hypothese B, die den Wert 30,7 meV vorhersagt.

Aufgabe 33: *Kolmogorow–Smirnow-Test* **5 P.**

In dieser Aufgabe sollen Sie mithilfe des Kolmogorow–Smirnow-Tests die Ähnlichkeit der Poisson- und Gauß-Verteilung untersuchen.

a) Welche Werte müssen Sie für μ und σ einer Gauß-Verteilung wählen, damit sie einer Poisson-Verteilung mit Erwartungswert λ möglichst ähnlich ist?

b) Implementieren Sie den Kolmogorow–Smirnow-Test für gebinnte Daten.

c) Der Kolmogorow–Smirnow-Test überprüft die Nullhypothese H_0, ob die Daten aus der gleichen Wahrscheinlichkeitsverteilung stammen. Wir wollen nun untersuchen, ab welchem Erwartungswert λ sich Poisson- und Gauß-Verteilung so ähnlich sind, dass der Kolmogorow–Smirnow-Test beide nicht mehr unterscheiden kann. Ziehen Sie dazu jeweils 10 000 Zufallszahlen aus einer Poisson-Verteilung und aus der entsprechenden Gauß-Verteilung für ein zu testendes λ. Beachten Sie:

- Runden Sie die aus der Gauß-Verteilung gezogenen Werte auf ganze Zahlen.
- Nutzen Sie jeweils 100 Bins im Intervall $[\mu - 5\sigma, \mu + 5\sigma]$

[1] Das Beispiel kommt aus der Festkörperphysik

- Ermitteln Sie durch Iteration den Wert für λ ab dem Sie H_0 auf Grund des Kolmogorow–Smirnow-Tests bei einem Konfidenzniveau von $\alpha = 5\,\%$ nicht mehr verwerfen können.

d) Ermitteln Sie analog zu oben den Wert für λ für die Konfidenzniveaus $\alpha = 2{,}5\,\%$ und $\alpha = 0{,}1\,\%$.

Aufgabe 34: *Ballon-Experiment* **5 P.**

Bei einem Ballon-Experiment zur Messung des Flusses der kosmischen Strahlung in der oberen Atmosphäre wird über einen Zeitraum von einer Stunde Protonen mit einer Energie zwischen 1 GeV und 100 GeV gezählt. Über einen Zeitraum von einer Woche wird jeden Tag ein Messdurchgang von einer Stunde Dauer vorgenommen. Ihre Messdaten lauten:

Tag	1	2	3	4	5	6	7
Counts	4135	4202	4203	4218	4227	4231	4310

a) Nehmen sie an, dass der Fluss der kosmischen Strahlung konstant im Messzeitraum ist. Berechnen Sie mit Hilfe der Maximum-Likelihood-Methode die wahrscheinlichste Zählrate. Geben Sie zudem den Likelihood-Wert an.

b) Ihr Kollege sieht sich die Messwerte an und stellt die Hypothese auf, dass der Fluss der kosmischen Strahlung einen dramatischen Zuwachs erlebt. Nehmen sie einen linear ansteigenden Fluss an und berechnen sie numerisch mit Hilfe der Maxmimum-Likelihood-Methode die wahrscheinlichsten Flussparameter. Geben sie zudem den Likelihoodwert für diese Parameter an.

c) Berechnen sie mit Hilfe eines Likelihood-Quotienten-Tests die Signifikanz seiner Beobachtung. Beurteilen sie die erreichte Signifikanz.
Hinweis: Nehmen sie an, dass Wilks' Theorem hier gültig ist. Warum darf man dies annehmen?

d) Ihr Kollege vollzieht eine Woche später eine weitere Messung um seine These zu untermauern. Seine Messung ergibt sich zu

Tag	14
Counts	4402

Berechnen sie erneut (a) bis (c) für diesen neuen Datensatz.

Aufgabe 35: *Likelihood-Quotienten-Test* **5 P.**

In einer Honigfabrik wird je eine Portion Honig in ein Glas zu μ_0 Millilitern abgefüllt. Es wird angenommen, dass die Füllmengen produktionsbedingt einer Normalverteilung mit Mittelwert $\mu = \mu_0$ und einer unbekannten Varianz σ^2 folgen.

a) Stellen Sie die Testbedingung für einen Likelihood-Quotienten-Test auf, in dem Sie die Nullhypothese von oben gegen die Gegenhypothese, dass die Füllmenge einer Normalverteilung folgt, die *nicht* den Mittelwert μ_0 hat, testen.

b) Für jeweils welche Wahl der Parameter μ und σ^2 werden die Likelihood-Funktionen der einzelnen Hypothesen auf dem jeweiligen Parameterbereich maximal?

c) Setzen Sie die in b) erhaltenen Parameter ein und reduzieren Sie die Testbedingung auf einen Ausdruck der einer t-Statistik folgt.
Hinweis: Die Größe $T = \sqrt{N}(\bar{x}-\mu_0)/s$ folgt unter der Nullhypothese der t-Statistik, wobei s die Stichprobenvarianz ist.

d) Es wird aufgrund der eingesetzten Maschinen eine Füllmenge von $\mu_0 = 200\,\text{ml}$ erwartet. Aus einer Stichprobe mit 25 Messungen zur Qualitätskontrolle wird ein Mittelwert von $\bar{x} = 205\,\text{ml}$ bei einer geschätzten Standardabweichung von $s = 10\,\text{ml}$ gemessen. Wird die oben aufgestellte Nullhypothese bei einer Signifikanz von 5 % abgewiesen oder beibehalten?

Aufgabe 36: *Zwei Histogramme* **5 P.**

Gegeben sind zwei Histogramme mit dem gleichen Binning (r Bins). Die Hypothese ist, dass beide Histogramme Zufallszahlen mit der gleichen Verteilung repräsentieren. Das heißt es existieren r Wahrscheinlichkeiten $p_1, \dots, p_r$ dafür, dass eine Beobachtung im i-ten Bin liegt ($\sum_{i=1}^{r} p_i = 1$). Die Einträge im ersten i-ten des ersten Histogramm werden als n_i und im zweiten m_i bezeichnet. Die Anzahl der Beobachtungen im ersten Histogramm ist $N = \sum_{i=1}^{r} n_i$ und im zweiten $M = \sum_{i=1}^{r} m_i$.

a) Welcher Verteilung folgen die Zählraten in den einzelnen Bins? Stellen sie die PDF für ein einzelnes Bin für beide Histogramme (n_i und m_i) auf unter Annahme der Nullhypothese auf.

b) Stellen sie die Likelihood Funktion für die Nullhypothese auf. Finden sie den Schätzer $\hat{p}_i$, der die Likelihood maximiert.

c) Stellen sie die χ^2-Test-Statistik unter Annahme der Nullhypothese auf. (Keine Vereinfachung des Terms nötig)

d) Wie viele Freiheitsgrade hat die X^2 Verteilung? Folgt die Test-Statistik für kleine Bininhalte ($n_i, m_i < 10$) immer noch einer χ^2-Verteilung? Falls nein, warum nicht?

e) Gegeben sind die Histogramme:

n_1	n_2	n_3
111	188	333

m_1	m_2	m_3
15	36	30

Es kann gezeigt werden, dass die Test-Statistik zu $X^2 = \frac{1}{NM}\sum_{i=1}^{r}\frac{(Nm_i - Mn_i)^2}{n_i + m_i}$ vereinfachen lässt. Prüfen Sie, ob für die gegeben Histogramme die Nullhypothese für $\alpha = 0.1, 0.05, 0.01$ verworfen wird.