Universität Tübingen
Mathematisches Institut
Dr. Stefan Keppeler

Sommersemester 2010

Mathematik II für Biologen

Übungsblatt 4 (Abgabe am 7.5.2010)

Aufgabe 10 (Fortsetzung von Aufgabe 8) (10 Punkte)
Die folgende Tabelle zeigt die Anzahl der google-Treffer für den Ausdruck *"k Katze(n) und h Hund(e)"*.

h\k	1	2	3	4	5	6	7	8
1	135	1550	187	115	4830	930	460	6
2	9980	1450	5250	8	3	2	2	3
3	4	191	3730	9	5	1	0	0
4	4	119	4590	162	9	4	1	1
5	3	9250	888	154	92	3	1	0
6	0	3880	349	37	4	0	0	1
7	4	1520	70	37	4	2	0	1
8	0	453	87	7	1	2	3	2

Daraus wurde der Korrelationskoeffizient $r \approx -0{,}3$ bestimmt. (Wie und für was? Vgl. auch Webforum.)

a) Wie könnte man das google-Experiment verbessern, um evt. ein aussagekräftigeres Ergebnis zu erhalten?
b) Führen Sie ein eigenes google-Experiment mit einem Satz oder Ausdruck, der zwei Zahlen enthält, durch, und bestimmen und deuten Sie den zugehörigen Korrelationskoeffizienten.

Aufgabe 11 (10 Punkte)
Im folgenden werden Fragestellungen beschrieben, die jeweils durch die (statistische) Auswertung der Messergebnisse eines "Experimentes" beantwortet werden könnten. Identifizieren Sie jeweils die Nullhypothese H_0 und die Alternativhypothese H_A und beschreiben Sie, wie solch ein Experiment aussehen könnte.

Zur Erinnerung: Nur H_A kann "statistisch bewiesen" werden, nicht jedoch H_0. H_0 ist sozusagen der Angeklagte, der nur bei hinreichender Beweislast verurteilt (=abgelehnt) werden kann. Bei mangelhaften Beweisen muss H_0 freigesprochen werden. Man will sich also relativ sicher sein, dass H_0 tatsächlich falsch ist, wenn man H_0 ablehnt.

Beispiel: Ein Hersteller von Angelschnüren will zeigen, dass die Schnüre eines Konkurrenten minderwertig sind, nämlich schon bei einer Belastung von weniger als 15 Kilogramm reißen. Dann ist H_0, dass die Schnüre des Konkurrenten erst bei einer (durchschnittlichen) Belastung von (mindestens) 15 kg reißen, und H_A, dass sie bei einer (durchschnittlichen) Belastung von echt weniger als 15 kg reißen. Ein Experiment, das die Daten dazu liefert, besteht darin, (viele) Angelschnüre des Konkurenten zu kaufen, jeweils mit bis zu 15 kg zu belasten, und zu notieren, ob sie bei einer Belastung von weniger als 15 kg reißen.

a) Ein Herpetologe will untersuchen, ob der Anteil p der Froscheier, aus denen Jungtiere schlüpfen, auf über 45% steigt, wenn man die Eier mit UV-Licht bestrahlt.

b) Beim Spielen von "Siedler von Catan" keimt der Verdacht, dass die Summe der beiden Würfel mit einer höheren Wahrscheinlichkeit 7 ist, als dies bei fairen Würfeln der Fall sein sollte.

c) Ein Forscherteam möchte beweisen, dass das Herzinfarktrisikio sinkt, wenn man täglich 7,5 g Schokolade isst (vgl. `www.spiegel.de/wissenschaft/medizin/0,1518,686558,00.html` und Buijsse et al., Eur. Heart J. (2010) `dx.doi.org/10.1093/eurheartj/ehq068`).

d) Korrigiert Ihr Übungsgruppenleiter tatsächlich viel strenger als der der Parallelgruppe?

e) Sie nehmen an, dass die Luft in Barcelona mehr als 500 Pikogramm Kokain pro Kubikmeter enthält (vgl. `www.spiegel.de/wissenschaft/mensch/0,1518,624811,00.html`).

Aufgabe 12 MATLAB (10 Punkte)

In dieser Aufgabe werden wir mittels (Monte-Carlo-)Simulation zu unterscheiden versuchen zwischen Ereignissen, die "durchaus" eintreten können, und solchen, die "praktisch unmöglich" sind.

Angenommen, Sie vermuten, dass Ihr Würfel gezinkt, d.h. nicht ideal, nicht fair ist. Und zwar vermuten Sie, dass die 1 häufiger oder seltener erscheint, als sie es sollte, wenn der Würfel fair wäre. Um Ihre Vermutung zu testen, werfen Sie den Würfel 100 Mal. Sei X die Anzahl der dabei geworfenen Einsen. Sie beobachten $X = 12$. Nun stellt sich die Frage, ob diese Beobachtung mit der (Null-)Hypothese vereinbar ist, dass der Würfel fair ist, oder ob sie die (Alternativ-)Hypothese unterstützt, dass der Würfel nicht fair ist.

Um herauszufinden, welche Werte von X durchaus beobachtet werden können, falls der Würfel fair ist, simulieren wir das Experiment, einen fairen Würfel 100 Mal zu werfen, und wiederholen diese Simulation 10000 Mal. MATLAB Programm dazu:

```
>> n=10000;
>> mat=unidrnd(6,100,n); % Ergibt eine 100 x n Matrix mat mit Eintraegen,
      % die zufaellig und unabhaengig voneinander 1,2,3,4,5 oder 6 sind,
      % wobei keine Zahl bevorzugt wird. Auf diese Weise wird ein idealer
      % Wuerfel simuliert. (Versuchen Sie es mit kleineren Zahlen n und
      % lassen Sie das Semikolon weg, wenn Sie sehen wollen, wie mat
      % aussieht.)
>> ist1= mat==1; % neue Matrix mit Eintrag 1, falls entsprechender Eintrag
                 % in mat eine 1 ist und 0 sonst.
>> wieoft1=sum(ist1);  % summiert ueber Spalten
>> r=hist(wieoft1,0:35)/n;
      % r = Vektor der relativen Haeufigkeit von Experimenten, in denen
      % keinmal 1 gewuerfelt wurde, genau einmal, genau zweimal, ...,
      % genau 34 Mal, mindestens 35 Mal.
>> bar(0:35,r) % zugehoeriges Stabdiagramm (ein Histogramm)
```

a) Welche Zeile oder Spalte der Matrix `mat` repräsentiert Ihr 7. (simuliertes) Experiment? ("Experiment" = 100maliges Werfen eines fairen Würfels)

b) In wieviel Prozent der Fälle ist bei Ihnen in 100 Würfen genau 12 Mal die 1 aufgetreten? Wie liest man dies aus der Computer-Ausgabe ab, und mit welchem MATLAB-Befehl erhalten Sie diesen Wert?

c) In wieviel Prozent der Fälle ist bei Ihnen in 100 Würfen mindestens 12 Mal die 1 aufgetreten? Wie liest man dies aus der Computer-Ausgabe ab, und mit welchem MATLAB-Befehl erhalten Sie diesen Wert?

d) Benutzen Sie den Vektor r, den Sie erhalten haben, um zu entscheiden, welche Werte "praktisch unmöglich" sind. Nehmen Sie hierbei $\alpha = 5\%$ als Signifikanz-Niveau, d.h. erklären Sie (etwa) $\alpha/2 = 2.5\%$ der kleinsten tatsächlich beobachteten Werte von X und $\alpha/2 = 2.5\%$ der größten beobachteten Werte von X für "praktisch nicht beobachtbar", wenn das Experiment nur ein einziges Mal durchgeführt wird. Gehört $X = 12$ demgemäß zu den (für $\alpha = 5\%$) "praktisch unmöglichen" Werten? HINWEIS:

```
>> cumsum(r)     % kumulative Summe von r: An der n-ten Stelle steht
                 % die Summe aller Eintraege von r links der n-ten
                 % Stelle einschliesslich der n-ten Stelle selbst.
```

e) Angenommen, jemand schlägt vor, genau dann die (Null-)Hypothese H_0 : *Der Würfel ist fair* zu verwerfen und stattdessen an die (Alternativ-)Hypothese H_A : *Der Würfel ist unfair* zu glauben, wenn bei der einmaligen Durchführung des Experimentes entweder $X \leq 10$ oder $X \geq 22$ beobachtet wurde.
In wieviel Prozent Ihrer oben simulierten 10000 Fälle würde man dann H_0 verwerfen müssen, obwohl wir wissen, dass H_0 wahr ist?