Universität Tübingen
Mathematisches Institut
Dr. Stefan Keppeler

Sommersemester 08

Mathematik II für Biologen

Übungsblatt 6 (Abgabe am 28.05.2008)

Aufgabe 17 (χ^2**-Test**) (10 Punkte)

Eine Maschine soll je 500 Nüsse einer Mischung bestehend aus Erd-, Hasel-, Cashew- und Pekannüssen im Verhältnis 5:2:2:1 in Tüten abfüllen.

a) Wieviele Nüsse von jeder Sorte müsste jede Tüte enthalten (bei einer Gesamtzahl von 500), wenn das Verhältnis der genannten Sorten *exakt* 5:2:2:1 betragen sollte? (Dies sind die "erwarteten" Anzahlen.)

Nun wird jedoch nicht verlangt, dass genau 50% aller Nüsse in den Tüten Erdnüsse sind etc. (so wie nicht erwartet werden kann, dass bei 900 Würfen eines fairen Würfels jede Augenzahl genau 150 mal geworfen wird), sondern es wird verlangt, dass sich die Maschine genauso verhält, als würde sie aus einem sehr sehr großen Container, der die genannten Nüsse gut gemischt genau im Verhältnis 5:2:2:1 enthält, ohne Zurücklegen nacheinander Nüsse ziehen und in die Tüte legen. Falls die Maschine dies erfüllt, kann der konkrete Inhalt einer Tüte zwar zufällig schwanken, aber "im Durchschnitt" sind die Nüsse in den verlangten Anteilen vorhanden.

Um zu testen, ob die Maschine richtig eingestellt ist und sich so verhält wie oben beschrieben (dies ist H_0) oder ob sie nicht richtig eingestellt ist (dies ist H_A), wird der Produktion eine Tüte entnommen und ihr Inhalt untersucht. Er besteht aus 269 Erd-, 112 Hasel-, 74 Cashew- und 45 Pekannüssen. (Dies sind die "beobachteten" Anzahlen.) Wir wählen als Teststatistik

$$\chi^2 := \sum_{i=1,2,3,4} \frac{(\text{ beobachtete Anzahl Nüsse der Sorte } i \text{ - erwartete Anzahl Nüsse der Sorte } i)^2}{\text{erwartete Anzahl Nüsse der Sorte } i},$$

vgl. T aus Aufgabe 16. Je größer der Wert von χ^2 desto größer der Verdacht, dass H_A gilt, d.h. dass die Maschine nicht richtig eingestellt ist. Man könnte nun wie in Aufgabe 16 den p-Wert mittels Monte-Carlo-Simulation schätzen und dann je nachdem, ob der (geschätzte) p-Wert kleiner als α ist oder nicht, H_0 verwerfen oder nicht. Für den Fall $\alpha = 5\%$ kann man jedoch stattdessen folgende Faustregel[2] für den Verwerfungsbereich anwenden: *Verwirf H_0 auf dem Signifikanz-Niveau $\alpha = 5\%$, falls $\chi^2 > \nu + 2\sqrt{2\nu}$, wobei ν die sog. Anzahl der Freiheitsgrade ist.* Die Anzahl ν der Freiheitsgrade ist gleich der Anzahl Klassen (hier =4) minus 1, d.h. hier $\nu = 4 - 1 = 3$.

b) Berechnen Sie diese "kritische Grenze" $\nu + 2\sqrt{2\nu}$.
c) Berechnen Sie den Wert der Teststatistik χ^2.
d) Wie lautet demnach die Testentscheidung?

---

[2]Diese Faustregel kann angewandt werden, wenn für etwa 80% der betrachteten Klassen gilt, dass die erwartete Anzahl jeweils ≥ 4 ist, und für die restlichen Klassen die erwartete Anzahl jeweils ≥ 1 ist. Dies ist hier erfüllt.

Aufgabe 18 (10 Punkte)

Man kann die χ^2-Teststatistik aus Aufgabe 17 auch benutzen, um zu testen, ob die Daten zu gut zum Modell passen, um echt zu sein.

a) (Matlab) Angenommen jemand behauptet, er hätte 900 Mal einen fairen Würfel geworfen und dabei jede Augenzahl genau 150 Mal beobachtet. Würden Sie ihm glauben? Schliesslich entsprechen die beobachteten Anzahlen n_i von Würfen der Zahl i genau den für einen fairen Würfel erwarteten Zahlen, d.h. die Teststatistik χ^2 ist 0. Oder ist die Übereinstimmung hier zu gut, um wahr zu sein? Führen Sie das Experiment, einen fairen Würfel 900 mal zu werfen, n-mal durch (n groß) und überprüfen Sie, in wievielen Fällen $\chi^2 = 0$ beobachtet wurde. Interpretieren Sie das Ergebnis. (Hier ist sozusagen H_0*: Daten sind echt* und H_A*: Daten sind gefälscht.*)

```
>> n=10000; % n kleiner waehlen, falls Rechner zu langsam
>> chi_quadrat=zeros(1,n);
>> anzahl=zeros(1,6);
>> for k=1:n
       wuerfe=unidrnd(6,1,900);
       for i=1:6
           anzahl(i)=sum(wuerfe==i);
       end
       chi_quadrat(k)=sum((anzahl-[150 150 150 150 150 150]).^2/150);
       % Ob man hier noch durch 150 teilt oder nicht, spielt keine Rolle.
   end
>> sum(chi_quadrat==0)/n % sum(T==0) zaehlt, wie oft T==0 beobachtet wurde.
```

b) (Matlab) *Hat Mendel geschummelt?* Ermitteln Sie mittels Monte-Carlo-Simulation, wie groß in etwa die Wahrscheinlichkeit dafür ist, bei Gültigkeit der Mendelschen Gesetze für 556 Erbsen einen Wert von χ^2 zu erhalten, der *kleiner* ist als der in Aufgabe 16 b) berechnete. Spricht dies dafür, dass Mendel seine Zahlen geschönt hat?

c) Mendel hat viele Versuche durchgeführt, nicht nur den oben erwähnten. Angenommen, in Aufgabe b) konnte gezeigt werden, dass die Zahlen in obigem Beispiel so gut zu den theoretisch zu erwartenden passen, dass es statistisch signifikant auf dem Signifikanz-Niveau von $\alpha = 5\%$ ist, dass sie nicht mehr rein zufällig zustande gekommen sein können. Welches Argument könnte man dennoch zur Verteidigung von Mendel anführen?

Aufgabe 19 (10 Punkte)

Beim Spiel Würfelzwerge von Selecta gibt es Kärtchen, auf denen Zwerge mit drei Kleidungsstücken (Mütze, Jacke und Hose) dargestellt sind. Die Kleidungsstücke kommen in 6 unterschiedlichen Farben (rot, gelb, grün, blau, lila und pink) vor. Man würfelt mit drei Farbwürfeln und muss schnell einen Zwerg finden, der Kleidung in der entsprechenden Farbkombination trägt. Zu jeder möglichen Farbkombination gibt es genau einen Zwerg, d.h., wenn es einen Zwerg mit gelber Mütze, gelber Jacke und blauer Hose gibt, dann kann es keinen mit gelber Mütze, blauer Jacke und gelber Hose geben, aber sehr wohl noch einen mit gelber Mütze, blauer Jacke und blauer Hose.

a) Wieviele Kärtchen hat das Spiel?
b) Wieviele ein-, zwei- und dreifarbige Zwerge gibt es jeweils?
c) Wie groß ist jeweils die Wahrscheinlichkeit, mit fairen Farbwürfeln die Kombination für einen ein-, zwei- oder dreifarbigen Zwerg zu würfeln?

Aufgabe 20 (10 Punkte)

Ungefähr 2% der Bevölkerung leiden an rheumatoider Arthritis. Etwa 70% davon haben das humane Leukozyten Antigen HLA-DR4, das sonst nur 25% der allgemeinen Bevölkerung haben. Angenommen, eine zufällig ausgewählte Person hat HLA-DR4. Mit welcher Wahrscheinlichkeit leidet diese Person unter rheumatoider Arthritis?

Bitte definieren Sie, ähnlich wie im Vorlesungsbeispiel *Diagnostischer Test* zum Satz von Bayes, klar die Ereignisse und Wahrscheinlichkeiten, die es zu betrachten gilt, und zeigen Sie alle Rechenschritte.