Universität Tübingen, Mathematisches Institut
Dr. Stefan Keppeler Sommersemester 2013

Mathematik II für Biologen

Übungsblatt 6 (Abgabe am 29.5.13 vor der Vorlesung oder bis 31.5.13, 11:10, in die grüne Mappe vor C6P43)

Aufgabe 18 (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 17. 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 17 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[4] 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?
e) Testen Sie mit einem χ^2-Test ohne Simulation, ob die Daten aus Aufgabe 17 auf dem Signifikanz-Niveau $\alpha = 5\%$ mit Mendels Hypothese vereinbar sind.

Aufgabe 19 (10 Punkte)

Zwei Frauen und ein Mann bewerben sich für zwei gleiche offene Stellen. Mindestens eine dieser Personen bekommt eine dieser Stellen. Die Wahrscheinlichkeit, dass der Mann als einziger eingestellt wird, beträgt 0,1. Mit Wahrscheinlichkeit 0,3 werden die beiden Stellen von den beiden Frauen besetzt. Mit Wahrscheinlichkeit 0,4 bleibt eine der beiden Stellen offen.
Bestimmen Sie eine geeignete Grundmenge Ω, die diese Situation beschreibt. Was ist über das dazugehörige Wahrscheinlichkeitsmaß P bekannt? Berechnen Sie mit Hilfe der Rechenregeln aus der Vorlesung die Wahrscheinlichkeit dafür, dass der Mann eingestellt wird.

Aufgabe 20 (20 Zusatzpunkte)

Erreichen Sie bis spätestens 16.06.13 auf `www.khanacademy.org` *Proficiency* in den *Skills Counting 2*, *Probability space*, *Basic set notation*, *Probability 1* und *Independent probability.*
HINWEISE siehe Aufgabe 3 (Blatt 1).

---

[4] 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 21 (10 Punkte)

Man kann die χ^2-Teststatistik aus Aufgabe 18 (und 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 600 Mal einen fairen Würfel geworfen und dabei jede Augenzahl genau 100 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 600 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=10^5; % n kleiner waehlen, falls Rechner zu langsam
>> chi_quadrat=zeros(1,n);
>> anzahl=zeros(1,6);
>> for k=1:n
       wuerfe=unidrnd(6,1,600);
       for i=1:6
           anzahl(i)=sum(wuerfe==i);
       end
       chi_quadrat(k)=sum((anzahl-[100 100 100 100 100 100]).^2/100);
   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 17 a berechnete. Spricht dies dafür, dass Mendel seine Zahlen geschönt hat?

c) Mendel hat viele Versuche durchgeführt, nicht nur den oben erwähnten. Angenommen, in Teil 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 $\alpha = 10\%$ ist, dass sie nicht mehr rein zufällig zustande gekommen sein können. Welches Argument könnte man dennoch zur Verteidigung von Mendel anführen?

Aufgabe 22 (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) Wie groß ist jeweils die Wahrscheinlichkeit, mit fairen Farbwürfeln die Kombination für einen ein-, zwei- oder dreifarbigen Zwerg zu würfeln?
b) Wieviele ein-, zwei- und dreifarbige Zwerge gibt es jeweils?
c) Wieviele Kärtchen hat das Spiel?

Aufgabe 23 (10 Punkte)

Die Kiesel in einem Flussbett stammen entweder aus Gebirge A oder Gebirge B, und zwar stammen 40% der Kiesel aus Gebirge A und 60% aus Gebirge B. 80% der Kiesel aus Gebirge A und 30% der Kiesel aus Gebirge B sind Quarzit-Kiesel. Betrachten Sie die folgenden Ereignisse:

A := Kiesel stammt aus Gebirge A
B := Kiesel stammt aus Gebirge B
Q := Kiesel ist ein Quarzit-Kiesel

Übersetzen Sie obige Aussagen in mathematische Formeln über (bedingte) Wahrscheinlichkeiten. Berechnen Sie dann die Wahrscheinlichkeit dafür, dass (a) ein zufällig aus dem Flussbett ausgewählter Kiesel ein Quarzit-Kiesel ist, und dass (b) ein zufällig ausgewählter Quarzit-Kiesel aus Gebirge B stammt.