

Mathematik II für Biologen

Übungsblatt 9 (Abgabe am 18.06.2008)

Aufgabe 30 (Vergleich der Macht zweier Tests)

(10 Punkte)

Oft gibt es mehrere Tests, die man auf dieselben Daten zum selben Signifikanz-Niveau α anwenden kann. Während α die Wahrscheinlichkeit für einen Fehler erster Art bestimmt und damit für alle Tests gleich ist, können die Wahrscheinlichkeiten β für einen Fehler zweiter Art für die verschiedenen Tests sehr verschieden sein. Es ist für Anwendungen wichtig, den Test zu nehmen, der (für die in Frage kommenden Alternativen) das kleinste β und damit die größte Macht $1 - \beta$ hat.

Beispiel: Gegeben sei ein Würfel, bei dem alle Zahlen gleich wahrscheinlich sind, nur die Zahl 6 kommt im Durchschnitt nur 80% so häufig vor, wie sie dies für einen fairen Würfel sollte. (Somit erscheint mit Wahrscheinlichkeit $P[W=6] = 0,8/6$ eine 6 und mit Wahrscheinlichkeit jeweils $P[W=i] = (1 - 0,8/6)/5$ eine der anderen fünf Zahlen $i = 1, 2, 3, 4, 5$.) Im folgenden werden ein χ^2 -Test und ein Vorzeichentest hinsichtlich ihrer Fähigkeit verglichen, zwischen einem ungezinkten und einem in dieser speziellen Weise gezinkten Würfel zu unterscheiden.

- a) Ungefähr mit welcher Wahrscheinlichkeit (=Macht= $1 - \beta$) findet ein χ^2 -Test zu $\alpha = 5\%$ heraus, dass dieser Würfel nicht fair ist, wenn der Würfel 900 Mal geworfen wird? MATLAB-Code dazu

```
n=10000; verworfen=zeros(1,n); % n kleiner waehlen, falls Rechner zu langsam
krit=5+2*sqrt(2*5); % kritische Grenze fuer alpha=5% und 6-1 Freiheitsgrade,
% vgl.Aufg.17

anzahl=zeros(1,6);
for k=1:n
    simwuerfe=ceil(5.8*rand(900,1)); % simuliert 900 Würfe des unfairen Würfels
    for i=1:6
        anzahl(i)=sum(simwuerfe==i);
    end
    T=sum((anzahl-[150 150 150 150 150 150]).^2)/150;
    verworfen(k)=T>krit; % Notiert 1, falls T>krit und 0 sonst.
end
Macht=??? % Anteil der Tests, die verwerfen konnten
```

- b) Ungefähr mit welcher Wahrscheinlichkeit findet ein Vorzeichentest heraus, dass dieser Würfel nicht fair ist, wenn der Würfel 900 Mal geworfen wird? Der Vorzeichentest verwende als Teststatistik T die Anzahl Würfe, die eine Zahl $< 3,5$ ergeben. Wenn der Würfel fair ist, sollte im Durchschnitt die Hälfte der Würfe, d.h. $900/2 = 450$ Würfe, eine Zahl $< 3,5$ ergeben. Gemäß Faustregel aus der Vorlesung verwirft dieser Test zu $\alpha = 5\%$, falls die tatsächlich beobachtete Zahl T um mindestens $\sqrt{900} = 30$ nach oben oder unten von der erwarteten Zahl 450 abweicht. Ersetzen Sie dazu im obigen MATLAB-Code die entsprechenden Zeilen durch $T=\text{sum}(\text{simwuerfe}<3.5)$; und $\text{verworfen}(k)=\text{abs}(T-450)>=30$;
- c) Interpretieren Sie die Ergebnisse aus a) und b). Welcher Test scheint mächtiger zu sein?
- d) Wie müsste man obige Programme abändern, um zu überprüfen, ob die jeweiligen Verwerfungskriterien tatsächlich zu $\alpha = 5\%$ passen?

Aufgabe 31

(10 Punkte)

Die Zufallsvariable X habe folgende Dichte:

$$f(t) = \begin{cases} c(1-t) & \text{falls } 0 \leq t \leq 1 \\ 0 & \text{sonst} \end{cases}$$

- a) Skizzieren Sie (von Hand) $f(t)$ und bestimmen Sie die Konstante c so, dass f eine Dichte ist. (Hinweis: Die Fläche unter der Kurve muss 1 sein.)
- b) Berechnen und skizzieren Sie (von Hand) die zugehörige (kumulative) Verteilungsfunktion $F(t)$.
- c) Bestimmen Sie die Wahrscheinlichkeit, dass X einen Wert zwischen 0,2 und 0,4 annimmt, also $P[0,2 \leq X \leq 0,4]$. Zeichnen Sie diese Wahrscheinlichkeit in die Skizzen von a) und b) ein.
- d) Berechnen Sie den Erwartungswert, die Varianz und den Median von X .

Aufgabe 32

(10 Punkte)

Eine räumlich homogene Ackerfläche wurde in 100 gleich große Quadrate eingeteilt. In der Mitte eines jeden Quadrates wurde eine Falle aufgestellt. Wir bezeichnen mit x_i die Anzahl Exemplare des Laufkäfers *CARABUS GRANULATUS*, die nach einer gewissen Zeit in der i -ten Falle gefunden wurden ($i = 1, 2, \dots, 100$). Die folgende Tabelle gibt die Anzahl $n(x)$ der Fallen mit genau x solchen Käfern an.⁵

x	0	1	2	3	4	5	6	7	8	9	10	≥ 11	Total
$n(x)$	22	17	24	10	14	2	1	3	3	3	1	0	100

Wir nehmen als (Null-)Hypothese H_0 an, dass die Käfer „gleichmäßig“ über das Feld verteilt sind und sich unabhängig voneinander bewegen.

- a) Warum ist die Anzahl von Käfern in jeder Falle in etwa poissonverteilt, wenn H_0 stimmt?

Laut Vorlesung sind der Erwartungswert $E[X]$ und die (theoretische) Varianz $\text{Var}(X)$ einer poissonverteilten Zufallsvariable mit Parameter λ beide gleich λ . Wenn obige Daten x_1, \dots, x_{100} also von unabhängigen poissonverteilten Zufallsvariablen stammen, so sollten auch ihre empirischen Entsprechungen, nämlich der arithmetische Mittelwert \bar{x} und die empirische Varianz s_x^2 , in etwa übereinstimmen, d.h. der Quotient s_x^2/\bar{x} sollte in etwa gleich 1 sein.

- b) Berechnen Sie (ähnlich wie in Aufgabe 3) den empirischen Mittelwert \bar{x} und die empirische Varianz s_x^2 der Stichprobe x_1, \dots, x_{100} sowie den Quotienten s_x^2/\bar{x} , entweder per Hand und mit dem Taschenrechner (Rechenweg angeben) oder mit MATLAB (Code angeben und erläutern).

Nun kann man natürlich auch bei perfekter Übereinstimmung der Daten mit dem Modell nicht erwarten, dass der empirische Quotient s_x^2/\bar{x} genau gleich 1 ist. Er wird nach wie vor zufällig um 1 herum schwanken. Nur wie stark?

- c) Um dies zu untersuchen, erzeugen Sie bitte 100 poissonverteilte Zufallszahlen mit Parameter λ , wobei Sie als λ den Wert \bar{x} nehmen, den Sie in Aufgabe b) berechnet haben, und berechnen Sie zu dieser simulierten Stichprobe den empirischen Quotienten s_x^2/\bar{x} (=Teststatistik). Wiederholen Sie das Ganze 1000 Mal und tragen Sie die so erhaltenen Quotienten in ein Histogramm ein. MATLAB-Code:

```
pois=poissrnd(???, 100, 1000) % ??? = arithm. Mittelwert der echten Daten
quotienten=var(pois)./mean(pois);
hist(quotienten)
```

Wo befindet sich der Quotient s_x^2/\bar{x} der echten Daten in diesem Histogramm? Wie lautet demnach in etwa der p-Wert für diesen Test? Was schließen Sie daraus für H_0 ?

Aufgabe 33

(10 Punkte)

Die Inkubationszeit X (in Monaten) einer bestimmten ansteckenden Krankheit wird als *log-normal* verteilt mit $\mu = 0$ und $\sigma = 0,4$ modelliert, d.h. man nimmt an, dass $\log X \sim \mathcal{N}(0; 0,4^2)$, wobei \log der natürliche Logarithmus ist.

- a) Geben Sie die Wahrscheinlichkeit an, dass die Inkubationszeit mehr als 2 Monate beträgt.
HINWEIS: Der MATLAB-Befehl `normcdf(x)` berechnet $\Phi(x)$.
- b) Berechnen Sie den Median `med` von X .
- c) Für eine $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariable Y gilt bekanntlich

$$P[\mu - \sigma \leq Y \leq \mu + \sigma] \approx 68\% \quad \text{und} \\ P[\mu - 1.96\sigma \leq Y \leq \mu + 1.96\sigma] \approx 95\%.$$

Finden Sie Konstanten c_1 und c_2 , so dass

$$P\left[\frac{\text{med}}{c_1} \leq X \leq \text{med} \cdot c_1\right] \approx 68\% \quad \text{und} \\ P\left[\frac{\text{med}}{c_2} \leq X \leq \text{med} \cdot c_2\right] \approx 95\%.$$

⁵Quelle: G. Welzl, F. P. Wahl. Statistische Modelle zur Erfassung der räumlichen Verteilung von Laufkäfern in einer Agrarlandschaft. Tagungsberichte der Arbeitsgruppe Ökologie, Dt. Region der Internat. Biometr. Ges., Heft 6, Mai 1995.