Universität Tübingen, Mathematisches Institut
Dr. Stefan Keppeler Sommersemester 2013

Mathematik II für Biologen

Übungsblatt 13 (Abgabe am 19.7.2013)

Aufgabe 48 (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 Durschnitt 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=sum(simwuerfe<3.5);` und `verworfen(k)=abs(T-450)>=30;`.

c) Interpretieren Sie die Ergebnisse aus den Teilen 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 49 (10 Zusatzpunkte)

Im Einführungsbeispiel aus den Vorlesungen 3 & 4 haben wir eine Rinderherde von 1000 Tieren betrachtet. Durch Untersuchen einer Stichprobe von 10 Tieren (3 davon waren krank) wollten wir rückschließen, wieviele Tiere der Herde krank sind. Dazu haben wir Herden mit unterschiedlich vielen kranken Tieren simuliert und ein einseitiges 95%-Vertrauensintervall für die Anzahl kranker Tiere bestimmt: $\{88, 89, \ldots, 1000\}$. Wir können dies auch als Vertrauensintervall $[8{,}8\%, 100\%]$ für den Anteil kranker Tiere interpretieren.

a) Bestimmen Sie zum Vergleich mittels Bootstrap ein einseitiges 95%-Vertrauensintervall für den Anteil kranker Tiere.

   ANLEITUNG:

   (i) Schätzen Sie den Anteil kranker Tier der Herde durch den Anteil kranker Tiere der Stichprobe, d.h. 30%. Nur mit welcher Genauigkeit? Hier kommt der Bootstrap ins Spiel.

   (ii) Definieren Sie sich einen Datenvektor, der die beobachtete Stichprobe repräsentiert, z.B. einen Vektor mit drei Einsen und 7 Nullen. Wie erhält man daraus die Anzahl und den Anteil kranker Tiere an der Stichprobe?

   (iii) Erzeugen Sie nun mittels Bootstrap viele (z.B. 100 000) ähnliche Stichproben, und ermitteln Sie für jede die Anzahl (oder den Anteil) kranker Tiere. Visualisieren Sie das Ergebnis, z.B. durch ein Histogramm, und bestimmen das gesuchte Vertrauensintervall.

b) Bestimmen Sie nun auch – ebenfalls mittels Bootstrap – ein zweiseitiges 90%-Vertrauensintervall für den Anteil kranker Tiere.

c) Hätten wir die Ergebnisse der Aufgabenteile a und b auch ermitteln können, ohne die Bootstrap-Stichproben tatsächlich zu erzeugen? Wie?