Tests für den Parameter p einer Bernoulliverteilung $\boxed{\text{S 2006 Version 1.0}}$

Es wäre naheliegend, dieses Kapitel schon in der Angewandten Statistik I zu behandeln. Aber dafür hatte die Zeit nicht gereicht.

Einige Beispiele sollen Situationen zeigen, in denen man einen Test für den Parameter p einer Bernoulliverteilung durchführen könnte.

Beispiel 1: Der englische Mathematiker John Kerrich warf während seiner Kriegsgefangenschaft im zweiten Weltkrieg eine Münze zehntausendmal. Fünftausendundsiebenundsechzigmal lag die Seite mit dem Kopf oben, die relative Häufigkeit des Ergebnisses *Kopf* war also 0.5067. Spricht diese Abweichung von $\frac{1}{2}$ dafür, dass bei der verwendeten Münze die Wahrscheinlichkeit für *Kopf* größer als $\frac{1}{2}$ ist?

Beispiel 2: Ein Meinungsforschungsinstitut befragt eine Stichprobe von 1000 bundesdeutschen Haushalten danach, ob bei ihnen innerhalb des letzten Jahres das Waschmittel Superweiß verwendet wurde. Die Herstellerfirma will nämlich wissen, ob ihr Waschmittel bei weniger als 15 Prozent der bundesdeutschen Haushalte (Hypothese H) oder bei mehr als 15 Prozent (Gegenhypothese K) im letzten Jahr verwendet wurde. Bei 180 Haushalten in der Stichprobe wird mit *Ja* geantwortet.

Beispiel 3: Student S behauptet, ziemlich sicher das Mineralwasser A und das Mineralwasser B am Geschmack erkennen zu können. Deshalb werden ihm zehn Gläser mit Mineralwasser vorgesetzt, bei denen er nicht anhand irgendwelcher Indizien wissen kann, ob sie jeweils A oder B enthalten, und er soll das am Geschmack erkennen. Bei 7 Gläsern tippt er auf den richtigen Inhalt. Spricht das Ergebnis dafür, dass S wirklich die beiden Sorten am Geschmack erkennen kann, oder rät er wohl einfach nur?

Wenn man es allgemein formuliert, so liegen im ersten Beispiel exakt und bei geeigneter Durchführung der Umfrage bzw. des Mineralwassertests in den beiden anderen Beispielen zumindest näherungsweise folgende Voraussetzungen vor:

Gegeben sind Realisierungen $x_1, x_2, \ldots, x_n$ von n stochastisch unabhängigen Zufallsgrößen $X_1, X_2, \ldots, X_n$, die alle gemäß derselben Bernoulliverteilung mit unbekanntem Parameter p verteilt sind.

Zu testen ist $H : p \leq p_0$ gegen $K : p > p_0$

Im ersten Fall ist dies wohl klar, es ist $n = 10000$, es ist p_0 gleich $\frac{1}{2}$. Die Zufallsgröße X_i nimmt den Wert $x_i = 1$ genau dann an, wenn beim i-ten Wurf *Kopf* geworfen wird, sonst den Wert $x_i = 0$. Die Formulierung "Spricht diese Abweichung dafür, dass bei der verwendeten Münze die Wahrscheinlichkeit für *Kopf* größer als $\frac{1}{2}$ ist? " deutet zumindest darauf hin, dass man $K : p > \frac{1}{2}$ wählt, denn man will ja gegebenenfalls diese Aussage statistisch absichern.

Im zweiten Fall muss man streng genommen voraussetzen, dass eine reine Zufallsstichprobe *mit* Zurücklegen aus der Menge aller Haushalte Deutschlands genommen wird - damit exakte stochastische Unabhängigkeit vorliegt. Da allerdings die Zahl 1000 verglichen mit der Gesamtzahl aller deutschen Haushalte klein ist, kann der Unterschied zwischen Entnahme mit und Entnahme ohne Zurücklegen praktisch vernachlässigt werden. Die Zufallsgröße X_i nimmt den Wert $x_i = 1$ genau dann an, wenn der i-te Haushalt im letzten Jahr das Waschmittel Superweiß verwendet hat (bzw. das behauptet), sonst den Wert $x_i = 0$. Es ist laut obiger Formulierung zu testen $H : p \leq 0.15$ gegen $K : p > 0.15$

Im dritten Fall dürfen sich die verschiedenen Proben nicht gegenseitig irgendwie beeinflussen. (Das könnte vielleicht dadurch erreicht werden, dass der Student zwischendurch etwas wartet, oder vielleicht auch dadurch, dass er zwischendurch etwas trockenes Brot isst, oder dergleichen.) Die Zufallsgröße X_i nimmt den Wert $x_i = 1$ bzw. $x_i = 0$ an, wenn der Student beim i-ten Glas die Mineralwassersorte korrekt feststellt bzw. daneben tippt. Auch hier deutet die Formulierung "Spricht das Ergebnis dafür, dass ..." darauf hin, dass man $H : p \leq \frac{1}{2}$ gegen $K : p > \frac{1}{2}$ testen soll.

Test für $H : p \leq p_0$ gegen $K : p > p_0$

Das Rezept für den Test $H : p \leq p_0$ gegen $K : p > p_0$ ist naheliegend. Man entscheidet sich für die Gegenhypothese K genau dann, wenn $\sum_{i=1}^{n} x_i$ groß genug ist. ($\sum_{i=1}^{n} x_i$ ist ja die Zahl der "Treffer", man kann auch so argumentieren, dass die Summe $\sum_{i=1}^{n} x_i$ bis auf einen Faktor gleich der Schätzung $\hat{p} = \overline{x}$ für p ist.)

Bekanntlich ist $\sum_{i=1}^{n} X_i$ binomialverteilt mit den Parametern n und p, für $p = p_0$ also binomialverteilt mit den Parametern n und p_0. Die Regel lautet dann:

Entscheidung für K, wenn $\sum_{i=1}^{n} x_i > c$, andernfalls Entscheidung für H.

Zur Bestimmung von c legt man p_0 zugrunde, denn für ein kleineres p ist die Wahrscheinlichkeit $w_p(\sum_{i=1}^{n} X_i > c)$ kleiner als für p_0 - was anschaulich klar sein sollte. Aber im Gegensatz zu den bekannten Normalverteilungstests beispielsweise ist hier die Prüfgröße natürlich nicht stetig verteilt (weil sie ja binomialverteilt ist), und deshalb kann man nicht exakt $w_{p_0}(\sum_{i=1}^{n} X_i > c) = \alpha$ erreichen. Aber man muss natürlich dafür sorgen, dass diese Wahrscheinlichkeit gegebenenfalls möglichst knapp unter α liegt, damit nichts verschenkt wird (d.h. die Fehlerwahrscheinlichkeit zweiter Art nicht unnötig groß wird). Das bedeutet für die Wahl von c:

c ist die kleinste ganze Zahl mit der Eigenschaft

$\sum_{k=c+1}^{n} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \alpha$

Analog überlegt man sich für den einseitigen Test in der Gegenrichtung, also für $H : p \geq p_0$ gegen $K : p < p_0$:

Entscheidung für K, wenn $\sum_{i=1}^{n} x_i < c$, andernfalls Entscheidung für H. Dabei ist c die *größte* ganze Zahl mit der Eigenschaft

$\sum_{k=0}^{c-1} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \alpha$

Beim zweiseitigen Test $H : p = p_0$ gegen $K : p \neq p_0$ verteilt man die Wahrscheinlichkeit α je zur Hälfte auf beide Seiten, also gilt:

Entscheidung für K, wenn $\sum_{i=1}^{n} x_i < c_1$ oder $\sum_{i=1}^{n} x_i > c_2$, andernfalls Entscheidung für H. Dabei ist c_1 die *größte* ganze Zahl mit der Eigenschaft

$\sum_{k=0}^{c_1-1} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \frac{\alpha}{2}$

und c_2 die *kleinste* ganze Zahl mit der Eigenschaft

$\sum_{k=c_2+1}^{n} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \frac{\alpha}{2}$

Es sei bemerkt, dass man bei großem n natürlich die obigen Summen schwer berechnen kann und deshalb gerne mit der Näherung nach dem Satz von de Moivre-Laplace (Spezialfall des zentralen Grenzwertsatzes) arbeitet. Das ist schon im Skript zur Angewandten Statistik I erläutert: Man ersetzt in der Prüfgröße für den Mittelwert einer Normalverteilung bei bekannter Varianz die Größe σ durch $\sqrt{p_0 \cdot (1-p_0)}$ und testet ansonsten so wie bei dem genannten Testproblem. Es ist klar, dass dieser Test dann nur näherungsweise die Fehlerwahrscheinlichkeit erster Art kleinergleich α hält.

Der Test wird manchmal auch als Test für den Parameter p einer *Binomialverteilung* bezeichnet. Das ist insofern nicht ganz unsinnig, als die Summe $\sum_{i=1}^{n} X_i$ binomialverteilt ist und der Test sich auf das p dieser Verteilung bezieht. Aber es ist insofern nicht sehr treffend, als die Verteilung von X_1 (gleich Verteilung von X_i) eine Bernoulli-Verteilung (d.h. die spezielle Binomialverteilung mit $n = 1$ ist).

Zum Merken:
Voraussetzung: $x_1, x_2, \ldots, x_n$ Realisierungen von stochastisch unabhängigen Zufallsgrößen, die alle gemäß derselben Bernoulliverteilung mit unbekanntem Parameter p verteilt sind.
1. $H : p \leq p_0$ gegen $K : p > p_0$
Entscheidung für K, wenn
$\sum_{i=1}^{n} x_i > c$
Andernfalls Entscheidung für H. Dabei ist c die kleinste ganze Zahl mit der Eigenschaft
$\sum_{k=c+1}^{n} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \alpha$
2. $H : p \geq p_0$ gegen $K : p < p_0$
Entscheidung für K, wenn
$\sum_{i=1}^{n} x_i < c$
Andernfalls Entscheidung für H. Dabei ist c die größte ganze Zahl mit der Eigenschaft
$\sum_{k=0}^{c-1} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \alpha$
3. $H : p = p_0$ gegen $K : p \neq p_0$
Entscheidung für K, wenn
$\sum_{i=1}^{n} x_i < c_1$ oder $\sum_{i=1}^{n} x_i > c_2$
Andernfalls Entscheidung für H. Also Entscheidung für H, wenn $c_1 \leq \sum_{i=1}^{n} x_i \leq c_2$. Dabei ist
c_1 die größte ganze Zahl mit der Eigenschaft
$\sum_{k=0}^{c_1-1} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \frac{\alpha}{2}$
und
c_2 die kleinste ganze Zahl mit der Eigenschaft
$\sum_{k=c_2+1}^{n} \binom{n}{k} p_0^k \cdot (1-p_0)^{n-k} \leq \frac{\alpha}{2}$

Aufgabe: Führen Sie für das dritte Beispiel den Test exakt durch!

Aufgabe: Führen Sie für die ersten zwei Beispiele den Test näherungsweise durch!

Aufgabe: Berechnen Sie für die ersten zwei Beispiele ein näherungsweises Konfidenzintervall für p!