4 Beurteilende Statistik

Dieses ist das zweite Kapitel in diesem Skript, das die Überschrift "Statistik" trägt, doch im Gegensatz zu Kapitel 2, in dem es nur darum ging, die Daten ansprechend und charakteristisch aufzubereiten, wollen wir hier Schlüsse aus unseren Daten ziehen. Genauer versuchen wir gewissermaßen, die inverse Aufgabe der Wahrscheinlichkeitstheorie zu lösen: Wir versuchen nun von den Daten auf die zugrunde liegende Verteilung zu schließen. Dieses Problem existiert in verschiedenen Schwierigkeitsgraden. Das schwierigste und realistischste davon, dass wir nämlich nichts oder so gut wie nichts a priori über die Verteilung der Daten wissen, können wir hier noch nicht einmal ansatzweise behandeln. Dies wäre die sogenannte nicht-parametrische Situation.

In der parametrischen Situation wissen wir schon von vornherein, dass die Verteilung der beobachteten Daten aus einer Klasse stammt, die wir mit einem endlich-dimensionalen Vektor beschreiben können. Diesen Fall wollen wir genauer anschauen. Wir werden hierbei archetypisch auch nur eine Situation genauer studieren, den n-fachen Münzwurf: Es seien also n unabhängige Beobachtungen $x_1, \ldots, x_n$ gegeben, die alle vom Wurf einer Münze stammen. Diese habe Erfolgswahrscheinlichkeit p, also

$$\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0), \tag{4.1}$$

und p sei dabei unbekannt. Die Familie der zugrunde liegenden Wahrscheinlichkeiten ist also $(\mathbb{P}_p)_{p\in[0,1]}$, p ist also der (eindimensionale) Parameter, um unsere Wahrscheinlichkeiten zu parametrisieren. Ziel ist es nun, mehr Informationen über p zu sammeln. Genauer wollen wir die folgenden Problemkreise untersuchen:
1. Schätze p.
2. Teste Hypothesen über p.

Prinzipiell kennt die beurteilende Statistik auch noch eine dritte Fragestellung, nämlich Bereiche anzugeben, in denen sich p mit großer Wahrscheinlichkeit befindet, sogenannte Konfidenzintervalle. Diese sollen hier aber nicht untersucht werden.

Für den Rest des Kapitels sei also $\Omega = \{0, 1\}$, $\mathcal{A} = \mathcal{P}\Omega$ und $\mathbb{P}_p$ die in (4.1) beschriebene Klasse von Verteilungen auf Ω. Weiter seien unabhängige Zufallsvariablen $X_1, \ldots, X_n$, die nach $\mathbb{P}_p$ verteilt sind, gegeben. Insbesondere gilt also

$$\begin{aligned} \mathbb{P}_p(X_1 = x_1, \ldots, X_n = x_n) &= \mathbb{P}_p(X_1 = x_1) \cdot \ldots \cdot \mathbb{P}_p(X_n = x_n) = p^k \cdot (1-p)^{n-k} \\ &\quad \text{für} \quad \sum_{i=1}^n x_i = k. \end{aligned}$$

4.1 Das Schätzproblem

In diesem Kapitel sollen Mechanismen beschrieben werden, die uns eine möglichst gute Schätzung von p erlauben. Eine solche Schätzung sollte natürlich tunlichst von den gemachten Beobachtungen $X_1 = x_1, \ldots, X_n = x_n$ abhängen. Wir definieren die Schätzfunktion deshalb in folgender Weise:

Definition 4.1 *Eine Schätzfunktion bzw. ein Schätzer für p ist eine Funktion*

$$\begin{aligned} \hat{p} : \mathbb{R}^n &\to \mathbb{R} \\ (x_1, \ldots, x_n) &\mapsto \hat{p}(x_1, \ldots, x_n). \end{aligned}$$

Bemerkung 4.2

1. *Die Idee bei der obigen Definition ist die, dass die $x_1, \ldots, x_n$ die Ausgänge der Zufallsvariablen $X_1, \ldots, X_n$ sind. $\hat{p}$ ist also ein zufälliger Wert, wobei der Zufall daher kommt, dass $x_1, \ldots, x_n$ zufällige Werte sind.*

2. *Die einfachste Schätzfunktion ist stets zu vermuten, dass die Münze fair ist, d. h. man wählt*
$$g(x_1, \ldots, x_n) = \frac{1}{2}$$
für alle $x_1, \ldots, x_n \in \{0, 1\}^n$.

3. *Die bekannteste Schätzfunktion für p ist das arithmetische Mittel (das sicher beinahe jeder nennen wird, den man nach einer Schätzung für p fragt), also*
$$\hat{p}(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{4.2}$$
Der Rest dieses Abschnitts beschäftigt sich damit, den naiven Schätzer in (4.2) zu rechtfertigen.

Dies geschieht so, dass wir uns zunächst fragen, was ein sinnvolles Prinzip wäre, um einen Schätzer zu konstruieren. Dieses Prinzip ist das sogenannte Maximum-Likelihood-Prinzip. Ausgangspunkt dieses Prinzips ist die Häufigkeitsinterpretation der Wahrscheinlichkeit. Danach werden wir eher ein wahrscheinliches als ein unwahrscheinliches Ereignis beobachten. Das Maximum-Likelihood-Prinzip ist es nun, einen Schätzer für p so zu konstruieren, dass unter dem geschätzten Wert für p die Beobachtung, die wir gemacht haben, maximale Wahrscheinlichkeit hat. Dieser Schätzer heißt Maximum-Likelihood-Schätzer.

Definition 4.3 *Ein Maximum-Likelihood-Schätzer für p ist jedes $\hat{p}$ mit*

$$\mathbb{P}_{\hat{p}}(x_1, \ldots, x_n) = \max_{p \in [0,1]} \mathbb{P}_p(x_1, \ldots, x_n).$$

Um dieses Konzept besser zu verstehen, berechnen wir einfach den Maximum-Likelihood-Schätzer in unserer Situation.

Satz 4.4 *Sei $0 \leq p \leq 1$. In der zu analysierenden Situation ist Maximum-Likelihood-Schätzer der naive Schätzer*

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Beweis: Eine Beobachtung $x = (x_1, \ldots, x_n)$ mit $k := \sum_{i=1}^{n} x_i$ 1en und $(n-k)$ 0en hat die Wahrscheinlichkeit

$$P_p(\{x\}) = \mathbb{P}_p(\{(x_1, \ldots, x_n)\}) = p^k (1-p)^{n-k}.$$

Diese wollen wir in p maximieren, d. h. wir suchen das Maximum der Funktion

$$p \mapsto L_x(p) := p^k (1-p)^{n-k}, \quad k := \sum_{i=1}^{n} x_i.$$

Diese Funktion ist nicht so einfach zu maximieren, jedenfalls ist

$$p \mapsto \log L_x(p)$$

leichter zu maximieren. Das ist auch völlig ausreichend, denn der natürliche Logarithmus ist wie jeder andere Logarithmus eine monotone Funktion, d. h. ein Maximum von $\log L_x(\cdot)$ ist auch ein Maximum von L_x. Nun ist

$$\mathcal{L}_x(p) := \log L_x(p) = k \log p + (n-k) \log(1-p).$$

Weiter ist

$$\begin{aligned} \frac{d}{dp} \mathcal{L}_x(p) &= \frac{k}{p} - \frac{n-k}{1-p} \qquad \text{und} \\ \frac{d^2}{dp^2} \mathcal{L}_x(p) &= -\frac{k}{p^2} - \frac{n-k}{(1-p)^2} < 0. \end{aligned}$$

Eine Nullstelle der ersten Ableitung wird also in der Tat ein lokales Maximum von $\mathcal{L}_x(\cdot)$ sein. Nun ist

$$\begin{aligned} & \frac{k}{p} - \frac{n-k}{1-p} = 0 \\ \Leftrightarrow \quad & k - kp = np - kp \\ \Leftrightarrow \quad & p = \frac{k}{n}. \end{aligned}$$

Man vergewissert sich zudem, dass für $0 < k < n$ $\mathbb{P}_1(\{x\})$ und $\mathbb{P}_0(\{x\})$ echt kleiner sind als $\mathbb{P}_{\hat{p}}(\{x\})$. Also ist (man erinnere sich, dass $k = \sum_{i=1}^{n} x_i$ war)

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

in der Tat ein Maximum-Likelihood-Schätzer. $\square$

Die Maximum-Likelihood-Methode ist *die* Methode schlechthin, um gute Schätzer zu konstruieren und dies gilt für eine sehr große Klasse von Problemen. Für den Fall der Binomialverteilung haben wir gesehen, dass der naive Schätzer (das arithmetische Mittel) der Maximum-Likelihood-Schätzer ist. Dies bedeutet, dass er aus einem uns vernünftig erscheinenden Prinzip konstruiert werden kann, aber es bedeutet natürlich nicht, dass es auch per se ein guter Schätzer sein muss. Wir wollen hier noch zwei Qualitätsmerkmale für den Maximum-Likelihood-Schätzer überprüfen:

Das erste geht davon aus, dass wir vielleicht in jeder einzelnen Schätzung, d. h. bei jeder einzelnen Beobachtung $(x_1, \ldots, x_n)$ einen Schätzfehler haben, dass wir aber wenigstens im Mittel den richtigen Wert für p schätzen wollen (im Durchschnitt über die $(x_1, \ldots, x_n)$ nicht über p).

Definition 4.5 *Ein Schätzer $\bar{p} = \bar{p}(X_1, \ldots, X_n)$ für p heißt* erwartungstreu, *wenn für alle $p \in [0, 1]$*

$$\mathbb{E}_p \bar{p} = p$$

gilt. Dabei besagt der untere Index p, dass wir den Erwartungswert bei zugrunde liegendem p bilden.

Es stellt sich heraus, dass unser naiver Schätzer $\hat{p} = \frac{1}{n} \sum_{i=1}^n x_i$ erwartungstreu ist:

Satz 4.6 *Der naive Schätzer*

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

ist erwartungstreu (wir schreiben große X_i um zu zeigen, dass $\hat{p}$ eine Zufallsvariable ist).

Beweis: Aus den Eigenschaften des Erwartungswerts folgt

$$\begin{aligned} \mathbb{E}_p \hat{p} &= \mathbb{E}_p \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_p X_i \\ &= \frac{1}{n} \sum_{i=1}^{n} p = \frac{1}{n} \cdot np = p. \end{aligned}$$

$\hat{p}$ ist also in der Tat erwartungstreu. □

Das zweite Qualitätsmerkmal geht von der Überlegung aus, dass für endliches n jeder Schätzer fehlerbehaftet sein mag, dass wir uns aber wünschen, dass für immer größere Stichproben die Qualität unseres Schätzers immer besser wird, und der Schätzer schließlich gegen den zu schätzenden Wert konvergiert.

Definition 4.7 *Ein Schätzer $\bar{p} = \bar{p}(X_1, \ldots, X_n)$ heißt konsistent, wenn für alle $\varepsilon > 0$ gilt*

$$\lim_{n\to\infty} \mathbb{P}_p(|\bar{p}(X_1, \ldots, X_n) - p| \geq \varepsilon) = 0$$

für alle $p \in [0, 1]$.

Auch diese Eigenschaft hat unser naiver Schätzer:

Satz 4.8 *Der naive Schätzer $\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i$ ist für die Münzwurf-Situation konsistent.*

Beweis: Es gilt für alle $\varepsilon > 0$

$$\mathbb{P}_p(|\hat{p} - p| \geq \varepsilon) = \mathbb{P}_p(|\frac{1}{n}\sum_{i=1}^n X_i - p| \geq \varepsilon)$$

und daher folgt die Behauptung aus dem schwachen Gesetz der großen Zahlen für den Münzwurf. □

Der naive Schätzer wird für uns noch von großem Nutzen sein, wenn wir im nächsten Abschnitt Hypothesen über das zugrunde liegende p testen wollen.

4.2 Testtheorie im Münzwurf

Im vorhergehenden Abschnitt hatten wir gesehen, dass ein sehr guter Schätzer für das unbekannte p der empirische Mittelwert, d. h. der naive Schätzer, ist. In der Klasse der erwartungstreuen Schätzer ist er sogar in gewissem Sinne der beste (dies werden wir hier aber weder spezifizieren noch beweisen). Ist nun aber unser p (der Erfolgsparameter) eine irrationale Zahl, so werden wir p mit keiner Schätzung unseres naiven Schätzers richtig schätzen: egal, wie groß n ist und welche Stichprobe $x_1, \ldots, x_n$ wir beobachten, stets wird $\frac{1}{n}\sum_{i=1}^n x_i$ ein Bruch sein und damit verschieden von p. Wir wollen in diesem Abschnitt bescheidener sein und sehen, ob wir zumindest Aussagen über p mit großer Wahrscheinlichkeit als richtig oder falsch nachweisen können.

Wir wollen dies zunächst an einem Beispiel kennenlernen:

Beispiel 4.9 Wir wollen uns der Frage zuwenden, ob ein neugeborenes Küken Körner erkennen kann, oder ob es dies durch Erfahrung lernen muss. Hierzu führen wir das folgende Experiment durch: Sobald ein Küken geschlüpft ist, werden ihm falsche Körner aus Papier vorgesetzt. Die Hälfte sind kleinere Kreise, die andere Hälfte kleine Dreiecke von gleicher Fläche. Nun beobachten wir das Küken 16 mal beim Picken. Davon picke es X Kreise. Wir haben nun zwei Hypothesen:

H_0: Das Verhalten des Kükens ist völlig indifferent gegenüber der Form der Körner, d. h. es pickt runde Körner und eckige Körner mit gleicher Wahrscheinlichkeit.

H_1: Das Küken bevorzugt runde Körner.

Dass das Küken dreieckige Körner lieber mag, scheint uns unwahrscheinlich. Man kann dieses Experiment dadurch mathematisieren, dass man sich die gepickten Körner als eine Folge von 16 0en und 1en vorstellt. Dabei schreiben wir 1, falls das Küken ein rundes Korn pickt und 0 für ein dreieckiges Korn. Wir nehmen an, dass diese 0en und 1en aus stochastisch unabhängigen Experimenten stammen (wir sind also nun in der Situation des Münzwurfs), und wir versuchen aufgrund der Stichprobe zwischen den beiden Hypothesen

$$H_0 : p = \frac{1}{2} \qquad \text{und} \qquad H_1 : p > \frac{1}{2}$$

zu entscheiden. Egal, wie wir uns entscheiden und welche Entscheidungsregel wir verwenden, können wir zwei Fehler begehen: Den Fehler erster Art: H_0 ist wahr und wird verworfen. Den Fehler zweiter Art: H_0 ist falsch und wird angenommen. Offensichtlich ist es unmöglich, beide Fehler gleichzeitig in den Griff zu bekommen. Versucht man beispielsweise den Fehler erster Art zu eliminieren, so wird man H_0 stets annehmen müssen. Dies aber maximiert den Fehler zweiter Art.

In der statistischen Theorie und Praxis hat sich das folgende Verfahren eingebürgert: Man sucht nach einem Test, bei dem der Fehler 1. Art kleiner ist als ein vorgegebenes Signifikanzniveau α. Typische Werte für α sind 5 %, 2,5 %, 1 %, 0,5 % oder 0,1 %. Wir wollen unsere Hypothesen auf das Niveau α=5 % testen (weil wir es nicht so tragisch finden, wenn wir uns irren – je gravierender ein Irrtum ist, desto kleiner müssen wir α wählen). Das folgende Testverfahren liegt nahe: Wir schätzen p aus den Daten mit Hilfe des naiven Schätzers $\hat{p}$. Wenn $\hat{p}$ so aussieht, wie wir es unter H_0 vermuten würden (also z. B. nicht zu sehr von 1/2 entfernt liegt), so nehmen wir H_0 an, sonst H_1. Dabei müssen wir, da wir den Fehler 1. Art klein halten wollen, H_0 auch dann annehmen, wenn $\hat{p}$ sogar größer ist als 1/2, aber nicht zu groß. Genauer sieht das Testverfahren so aus:

1. Schätze p durch $\hat{p}$
2. Akzeptiere H_0, falls $\hat{p} \leq \Gamma = \Gamma(\alpha)$
3. Verwerfe H_0, falls $\hat{p} > \Gamma = \Gamma(\alpha)$.

Das Intervall $]\Gamma(\alpha), 1]$ heißt dabei Ablehnungsbereich für diesen Test. Die Schranke für $\Gamma = \Gamma(\alpha)$ berechnet sich dabei so, dass unter H_0

$$\mathbb{P}(\hat{p} > \Gamma) \leq \alpha, \qquad \text{also} \qquad \mathbb{P}_{1/2}(\hat{p} > \Gamma) \leq \alpha$$

gilt. In unserem Fall wollen wir also

$$\mathbb{P}_{1/2}(\hat{p} > \Gamma) \leq 0,05$$

sicherstellen. Wegen $\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i$ ist dies gleichbedeutend mit

$$\begin{aligned} \mathbb{P}_{1/2}(\sum_{i=1}^n X_i > n\Gamma) &\leq 0,05, \quad \text{d. h.} \\ \sum_{k=n\Gamma}^n b(k,16,\frac{1}{2}) &\leq 0,05. \end{aligned}$$

Berechnet man die Werte der Binomialverteilung $b(k, 16, \frac{1}{2})$, sieht man, dass

$$\sum_{k=12}^{16} b(k,16,\frac{1}{2}) \cong 0,0384 < 0,05,$$

während $\sum_{k=11}^{16} b(k,16,\frac{1}{2}) > 0,05$ gilt. Wir wählen also

$$n\Gamma = 12, \quad \text{d. h.} \quad \Gamma = \frac{12}{16} = \frac{3}{4}.$$

Ist $\hat{p} \leq 0,75$, werden wir H_0 annehmen, ansonsten verwerfen.

Dieses Beispiel erlaubt es uns auch schon, die Regeln für sogenannte einseitige Testprobleme aufzustellen: Zu testen sei

$$\begin{array}{llll} H_0 : p \leq p_0 & \text{gegen} & H_1 : p > p_0 & \text{bzw.} \\ H_0 : p < p_0 & \text{gegen} & H_1 : p \geq p_0 & \text{oder} \\ H_0 : p = p_0 & \text{gegen} & H_1 : p > p_0 & \end{array}$$

auf dem Niveau α (d. h. die Wahrscheinlichkeit für einen Fehler 1. Art darf höchstens α sein). Dann befolge man folgendes Verfahren:

1. Schätze p durch $\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i$ aus den Daten.
2. Bestimme $\Gamma = \Gamma(\alpha)$ möglichst klein, so dass

   $$\mathbb{P}_{p_0}(\hat{p} > \Gamma) \leq \alpha$$

   und dabei Γ möglichst klein.
3. Falls $\tilde{p} \leq \Gamma$, nehme H_0 an, ansonsten verwerfe H_0.

Bemerkung 4.10 *1. Dieses Testverfahren ist insofern optimal, dass es beispielsweise für $H_0 : p = p_0$ gegen $H_1 : p = p_1$, $p_1 > p_0$ bei festgehaltenem Fehler 1. Art den kleinsten Fehler 2. Art liefert. Dies ist der Inhalt des sogenannten Neyman-Pearson-Lemmas, das wir hier nicht beweisen können.*

2. *Dass wir auch für $H : p < p_0$ bzw. $H_0 : p \leq p_0$ nur*

$$\mathbb{P}_{p_0}(\hat{p} > \Gamma) \leq \alpha$$

kontrollieren müssen, liegt daran, dass

$$p \mapsto \mathbb{P}_p(\hat{p} > \Gamma)$$

eine monotone Funktion ist.

3. *Analog kann man $H_0 : p = p_0$ gegen $H_1 : p < p_0$ und ähnliche Hypothesen testen. Das Verfahren ist dann:*
   1. *Schätze $\hat{p}$ aus den Daten.*
   2. *Finde $\Gamma = \Gamma(\alpha)$ möglichst groß, so dass $\mathbb{P}_{p_0}(\hat{p} < \Gamma) \leq \alpha$.*
   3. *Akzeptiere H_0, falls $\hat{p} \geq \Gamma$, ansonsten verwerfe H_0.*

Beispiel 4.11 *Zwei Spieler A und B würfeln. Dabei behauptet B, dass der Würfel gezinkt sei und weniger 6en würfelt als normalerweise. A hingegen behauptet, der Würfel sei fair. Es werden 10 Probewürfe gemacht, dabei fällt eine 6. Kann man die Hypothese*

$$\begin{aligned} & H_0: \text{ Der Würfel ist fair } \quad (p = \frac{1}{6}) \\ \text{gegen} \qquad & H_1 : p < \frac{1}{6} \end{aligned}$$

auf dem Niveau $\alpha = 0,05$ verwerfen? Hierbei ist p die unbekannte Wahrscheinlichkeit, eine 6 zu würfeln. Wir tabellieren die Wahrscheinlichkeiten $\mathbb{P}_{1/6}(X = k)$, die für uns relevant sind (X sei die Anzahl von 6en in 10 Würfen)

k	$\mathbb{P}_{1/6}(X = k)$	$\mathbb{P}_{1/6}(X \leq k)$
0	0,161	0,161

Damit sehen wir bereits, dass wir mit 10 Würfen die Hypothese H_0 nie verwerfen können. Wir akzeptieren also H_0.
(Aufgabe: Wie groß muss die Anzahl der Würfe mindestens sein, damit man einen nicht leeren Ablehnungsbereich hat?)

Neben den bisher behandelten einseitigen Tests gibt es auch zweiseitige Tests. Diese sind von der Form

$$H_0 : p = p_0 \qquad \text{gegen} \qquad H_1 : p \neq p_0.$$

Es liegt auf der Hand, den Test folgendermaßen zu gestalten:

1. Schätze p durch $\hat{p} = \frac{1}{n}\sum_{i=1}^n x_i$.

2. Bestimme $\Gamma = \Gamma(\alpha)$ möglichst klein mit
$$\mathbb{P}_{p_0}(\hat{p} \notin [p_0 - \Gamma, p_0 + \Gamma]) \leq \alpha.$$
3. Akzeptiere H_0, falls $\hat{p} \in [p_0 - \Gamma, p_0 + \Gamma]$, ansonsten verwerfe H_0.

Wir wollen dies an einem Beispiel kennenlernen.

Beispiel 4.12 *Sind Ratten farbenblind? Wir wollen klären, ob Ratten eine der Farben rot oder grün vorziehen. Hierfür planen wir den folgenden Versuch: Ratten werden durch einen Gang geschickt, der sich in zwei Gänge verzweigt, die grün bzw. rot gestrichen sind. Je nach dem Ausgang des Experiments entscheiden wir uns für*

H_0 : *Die Ratten entscheiden sich für die beiden Gänge mit gleicher Wahrscheinlichkeit* $(p = \frac{1}{2})$ *oder*

H_1 : *Die Ratten bevorzugen einen der Gänge, d. h. eine der Farben* $(p \neq \frac{1}{2})$.

Hierbei steht p für die Wahrscheinlichkeit, z. B. den roten Gang zu betreten. Wir wenden nun den oben beschriebenen Weg an: Es stehen uns 10 Ratten zur Verfügung, d. h. wir können eine Stichprobe vom Umfang $n = 10$ erheben. Wir suchen nun unser Γ und damit das kritische Gebiet. Dazu müssen wir Γ so wählen, dass
$$\mathbb{P}_{1/2}(\frac{1}{n}S_n \notin [\frac{1}{2} - \Gamma, \frac{1}{2} + \Gamma)) \leq \alpha, \quad S_n := \sum_{i=1}^{n} X_i,$$
und Γ dabei möglichst klein. Wir fertigen die folgende Tabelle an:

k	$\mathbb{P}(S_n = k)$
0	$\frac{1}{1024}$
10	$\frac{1}{1024}$
1	$\frac{10}{1024}$
9	$\frac{10}{24}$
2	$\frac{45}{1024}$
8	$\frac{45}{1024}$

Wir sehen also
$$\mathbb{P}_{1/2}(|S_n - 5| \geq 4) = \frac{22}{1024} < 0,05$$
und
$$\mathbb{P}_{1/2}(|S_n - 5| \geq 3) > 0,05.$$
Wir gehen also folgendermaßen vor: Gehen die Ratten 9 oder 10 mal in denselben Gang, so werden wir die Hypothese H_0, dass Ratten farbenblind sind, verwerfen, anderenfalls werden wir sie auf dem Niveau $\alpha = 0,05$ akzeptieren.

Bemerkung 4.13 *1. Bei einem einseitigen Test, beispielsweise der Form*

$$H_0 : p < p_0 \qquad \text{gegen} \qquad H_1 : p > p_0$$

ist prinzipiell auch ein Vertauschen der Hypothesen H_0 und H_1 möglich. Dabei tauscht man auch die Fehler 1. Art und 2. Art, man sichert somit einen anderen Fehler ab. Ein solches Vertauschen ist bei einem zweiseitigen Test

$$H_0 : p = p_0 \qquad \text{gegen} \qquad H_1 : p \neq p_0$$

nicht möglich.

2. Man sollte beachten, dass die Verwerfungsbereiche von ein- und zweiseitigen Tests verschieden sind. Eine einseitige Hypothese kann verworfen werden, während die zweiseitige angenommen wird. Weiß man z. B. in Beispiel 4.13, dass Ratten – wenn überhaupt – rot bevorzugen, d. h. testet man einseitig

$$H_0 : p = \frac{1}{2} \qquad \text{gegen} \qquad H_1 : p > \frac{1}{2},$$

so sieht man, dass der Ablehnungsbereich für α=10 % so konstruiert wird, dass man H_0 bei 8, 9 oder 10 Entscheidungen für den grünen Gang H_0 ablehnt, während man dies im Originalproblem nur bei 9 oder 10 Entscheidungen für "grün" täte. Das ist gewissermaßen paradox: Obwohl man eine stärkere Evidenz für H_0 hat (denn $p < \frac{1}{2}$ ist ja a priori ausgeschlossen im modifizierten Probem), lehnt man H_0 im modifizierten Problem eher ab (nämlich schon bei 8 grünen Versuchen).

Für große Versuchsumfänge n können wir wieder den Zentralen Grenzwertsatz verwenden, den wir schon in Kapitel 3 kennengelernt haben.

Beispiel 4.14 *Eine Münze wird 1 000 mal geworfen. Es soll wieder*

$$H_0 : p = \frac{1}{2} \qquad \text{gegen} \qquad H_1 : p \neq \frac{1}{2}$$

getestet werden, also fragen wir, ob die Münze fair ist. Das Testniveau betrage α=5 % und wir beobachten 550 mal Kopf. Können wir H_0 bestätigen oder verwerfen? Wir konstruieren den Ablehnungsbereich. Gesucht ist Γ, so dass

$$\mathbb{P}_{1/2}(|S_{1000} - 500| \geq \Gamma) \leq \alpha$$

und dabei Γ möglichst klein. Es ist nach dem Zentralen Grenzwertsatz:

$$\mathbb{P}_{1/2}(|S_{1000} - 500| \geq \Gamma) = \mathbb{P}_{1/2}(|\frac{S_{1000} - 500}{\sqrt{1000 \cdot \frac{1}{4}}}| \geq \frac{2\Gamma}{\sqrt{1000}}) \approx 2 - 2\Phi(\frac{2\Gamma}{\sqrt{1000}}).$$

Dabei ist $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ gesetzt. Es soll also

$$2 - 2\Phi(\frac{2\Gamma}{\sqrt{1000}}) = 0,05$$

sein und deshalb

$$\Phi(\frac{2\Gamma}{\sqrt{1000}}) = 0,975.$$

Aus einer $\mathcal{N}(0,1)$*-Tafel entnehmen wir, dass dazu*

$$\frac{2\Gamma}{\sqrt{1000}} \simeq 1,96$$

sein muss, also

$$\Gamma \cong 31.$$

Wir werden H_0 also annehmen, wenn wir zwischen 469 und 531 mal Kopf sehen und ansonsten verwerfen. In diesem Fall (550 mal Kopf) würden wir H_0 also verwerfen. Ist stattdessen $\alpha = 0,001$ vorgegeben, so führt dieselbe Rechnung zu

$$\begin{array}{rcll} 2 - 2\Phi(\frac{2\Gamma}{\sqrt{1000}}) & = & 0,001, & \textit{d. h.} \\ \Phi(\frac{2\Gamma}{\sqrt{1000}}) & = & 0,9995, & \textit{also} \\ \frac{2\Gamma}{\sqrt{1000}} & \approx & 3,3, & \textit{d. h.} \\ \Gamma & \cong & 52. & \end{array}$$

In diesem Falle würde man H_0 annehmen.