Statistik

von Peter Pfaffelhuber

Version: 28. September 2015

Inhaltsverzeichnis

0 Einführung

Es ist nicht übertrieben zu behaupten, dass in der heutigen Welt immer mehr *Daten* jeglicher Art erhoben werden. Diese zu ordnen und aus Daten Schlussfolgerungen zu ziehen ist Aufgabe der Statistik.

Man teilt dabei diese Aufgaben in zwei Gebiete auf. Die *deskriptive Statistik* dient rein der Beschreibung der Daten, etwa durch geeignete Wahl von Statistiken, die die Daten zusammenfassen. Anders ist dies bei der hier behandelten *schließenden* oder *induktiven Statistik*. Die Aufgabe ist hier, mit Hilfe von stochastischen Modellen Aussagen darüber zu treffen, welchen Annahmen den Daten zugrunde liegen könnten. Wir beginnen mit einem Beispiel.

Bemerkung 0.1 (Ein einführendes Beispiel: Erfolgswahrscheinlichkeit beim Münzwurf).
Bevor wir statistische Konzepte einführen, betrachten wir folgendes Beispiel: eine Münze wird 53 mal geworfen. Dabei ist die Wahrscheinlichkeit für *Kopf* (was wir im Folgenden als Erfolg werten wollen) noch unbekannt. Von den 53 Würfen sind 23 ein Erfolg.

Unsere statistischen Überlegungen gehen nun von der Vorstellung aus, dass die 53 Münzwürfe die Realisierung einer Zufallsvariable $X = (X_1, \dots, X_{53})$ sind, wobei $X_1, \dots, X_{53}$ unabhängig und identisch verteilt sind mit

$$X_i = \begin{cases} 1, & \text{falls der } i\text{-te Wurf } \textit{Kopf} \text{ zeigt}, \\ 0, & \text{sonst.} \end{cases}$$

und es gilt

$$\mathbf{P}(X_i = 1) = p.$$

Jetzt ist $X_1 + \cdots + X_n$ die Gesamtzahl der Erfolge. Als Summe von n unabhängigen Bernoulli-verteilten Zufallsvariablen ist diese Summe $B(n = 53, p)$-verteilt. Wichtig ist, dass zwar $n = 53$ bereits fest steht (schließlich wissen wir ja, dass wir 53 mal die Münze geworfen haben), nicht jedoch p. In dieser Situation gibt es zwei *statistische Probleme*.

- *Schätzproblem:* Wir können versuchen, den Erfolgsparameter p zu schätzen. Entweder werden wir dazu einen aus den Daten (23 Erfolge aus 53 Versuchen) abgeleiteten Wert $\widehat{p}$ angeben (*Punktschätzer*), oder ein aus den Daten abgeleitetes Intervall $[a, b]$, in dem der wahre Parameter p mit hoher Wahrscheinlichkeit liegt (*Intervallschätzer*).

- *Testproblem:* Stellen wir uns vor, der Werfer der Münze behauptet, dass die Münze fair ist, also $p = \frac{1}{2}$ gilt. Dieser Meinung können wir skeptisch gegenüber stehen, da ja nur 23 aus 53 Würfen (etwa 43 %) ein Erfolg waren. Wir können versuchen, die Hypothese $p = \frac{1}{2}$ zu testen. Das bedeutet, dass wir untersuchen, wie gut die Hypothese mit den Daten in Einklang steht.

Die Erfolgswahrscheinlichkeit schätzen: Wir versuchen nun, die Erfolgswahrscheinlichkeit des Münzwurfes zu schätzen. Dies muss auf Grundlage der Daten erfolgen, also basierend auf dem Wissen, dass 23 aus 53 Erfolge zu verzeichnen waren. Ein einfacher Ansatz ist es, zu vermuten, dass die 23 Erfolge aus der 53 Würfen in etwa der Erfolgsquote entspricht. Also ist

$$\widehat{p} = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{23}{53} \approx 0.43$$

ein Schätzer für den unbekannten Parameter p. (Im Folgenden werden wir einen Schätzer für einen Parameter $\bullet$ meistens mit $\widehat{\bullet}$ bezeichnen.) Da $\widehat{p}$ von den Daten abhängt, die wir uns als Realisierung von einer Zufallsvariable gedacht haben, ist $\widehat{p}$ also auch eine Zufallsvariable. Warum ist der Schätzer $\widehat{p}$ gut? Nehmen wir an, wir wüssten den wahren Parameter p. Dann leistet $\widehat{p}$ zumindest im Mittel das gewünschte: (Wir schreiben hier und im Folgenden $\mathbf{P}_p(\cdot)$ und $\mathbf{E}_p[\cdot]$, wenn wir Wahrscheinlichkeiten und Erwartungswerte unter der Hypothese ausrechnen wollen, dass p der wahre Parameter ist.)

$$\mathbf{E}_p[\widehat{p}] = \frac{1}{n}\mathbf{E}_p[X_1 + \cdots + X_n] = p.$$

Wir sagen auch, der Schätzer $\widehat{p}$ ist erwartungstreu (oder unverzerrt oder unbiased).

Eine weitere wünschenwerte Eigenschaft eines Schätzers ist, dass er immer besser wird, je größer die zu Grunde liegende Datenmenge ist. Eine große Datengrundlage bedeutet in unserem Fall, dass die Münze oft geworfen wurde, also n groß ist. Aus dem schwachen Gesetz großer Zahlen wissen wir, dass

$$\mathbf{P}_p(|\widehat{p} - p| \geq \varepsilon) = \mathbf{P}_p\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - \mathbf{E}_p[X_1]\Big| \geq \varepsilon\Big) \xrightarrow{n\to\infty} 0$$

für alle $\varepsilon > 0$. Die Eigenschaft, dass $\widehat{p}$ mit hoher Wahrscheinlichkeit immer näher am wahren Wert p liegt, wenn mehr Daten zur Verfügung stehen, nennen wir *Konsistenz*.

Stellen wir uns vor, in zwei aufeinander folgenden Experimenten bekommen wir zunächst 23 von 53 Erfolge, und dann 23000 von 53000 Erfolge. In beiden Fällen ist $\widehat{p} = \frac{23}{53}$. Es ist jedoch klar, dass wir dem Wert von $\widehat{p}$ im zweiten Experiment eine viel höhere Bedeutung zumessen und wir uns sicherer sind, dass der wahre Wert in der Nähe von $\widehat{p}$ liegt. Diese Sicherheit können wir mit einem *Intervallschätzer*, also einem Intervall, in dem der wahre Wert mit hoher Wahrscheinlichkeit liegt, zu Ausdruck bringen. Dazu wählen wir ein (kleines) $\alpha \in (0,1)$, etwa $\alpha = 5\,\%$. Aus dem zentralen Grenzwertsatz folgt, dass es eine nach $N(0,1)$ verteilte Zufallsvariable Z gibt, so dass[1]

$$\mathbf{P}_p\Big(-1.96 \leq \frac{n\widehat{p} - np}{\sqrt{np(1-p)}} \leq 1.96\Big) \approx \mathbf{P}\big(-1.96 \leq Z \leq 1.96\big) \approx 0.95.$$

Weiter ist

$$\begin{aligned} 0.95 &\approx \mathbf{P}_p\Big(-1.96 \leq \frac{\widehat{p} - p}{\sqrt{\widehat{p}(1-\widehat{p})/n}} \leq 1.96\Big) \\ &= \mathbf{P}_p\Big(\widehat{p} - 1.96\sqrt{\widehat{p}(1-\widehat{p})/n} \leq p \leq \widehat{p} + 1.96\sqrt{\widehat{p}(1-\widehat{p})/n}\Big). \end{aligned}$$

Wir haben gerade gezeigt, dass

$$\big[\widehat{p} - 1.96\sqrt{\widehat{p}(1-\widehat{p})/n}; \widehat{p} + 1.96\sqrt{\widehat{p}(1-\widehat{p})/n}\big]$$

ein *Konfidenzintervall* für $\widehat{p}$ zum *Niveau* $95\,\%$ ist. Das bedeutet, dass der wahre Wert mit Wahrscheinlichkeit etwa $95\,\%$ in diesem (zufälligen) Intervall liegt. Haben wir 23 Erfolge in

[1] Wir schreiben in dieser einführenden Bemerkung öfter approximative Formeln mittels '$\approx$'. Es sei bemerkt, dass dieses Symbol keine mathematisch exakten, beweisbaren Aussagen trifft. Für Anwendungen sinnvolle Resultate liefert es allerdings allemal.

$n = 53$ Versuchen, ist unser Konfidenzintervall also $[0.30, 0.57]$. Haben wir hingegen 23000 Erfolge bei 53000 Versuchen, ist das Konfidenzintervall etwa $[0.430, 0.438]$, also wesentlich kleiner.

Die Erfolgswahrscheinlichkeit testen: Nehmen wir an, der Werfer der Münze behauptet, sie sei fair, also $p = \frac{1}{2}$. Können wir diese Hypothese aufgrund der Daten verwerfen? Zunächst stellen wir fest, dass wir prinzipiell zwei Arten von Fehlern mit unserer Entscheidung machen können. Wenn wir die Hypothese verwerfen, könnte sie doch wahr sein, und wenn wir die Hypothese nicht verwerfen, könnte sie doch richtig sein.

Da wir nicht grundlos dem Werfer der Münze widersprechen wollen, wollen wir die Wahrscheinlichkeit, dass wir die Hypothese ablehnen (wir dem Werfer der Münze widersprechen), obwohl sie wahr ist (die Hypothese des Werfers richtig ist), kontrollieren. Das bedeutet, dass

$$\mathbf{P}_{p=1/2}(\text{Hypothese verwerfen}) \leq \alpha$$

für ein anfangs gewähltes $\alpha \in (0, 1)$ sein soll. Klar ist, dass damit die Hypothese umso seltener abgelehnt werden kann, je kleiner α ist. Nun kommen wir zu der Regel, mit der wir die Hypothese ablehnen wollen. In unserem Beispiel haben wir für die Hypothese $p = \frac{1}{2}$ zu wenig (23 von 53) Erfolge. Wir würden die Hypothese ablehnen wollen, wenn

$$\mathbf{P}_{p=1/2}(\text{Daten extremer als tatsächliche Daten}) \leq \alpha. \tag{0.1}$$

Wir wählen (wie oben beim Konfidenzintervall) $\alpha = 5\,\%$. Für Y_n nach $B(n = 53, p = \frac{1}{2})$ verteilt, erwarten wir 26.5 Erfolge. Um die Wahrscheinlichkeit einer Abweichung, die größer ist als die der Daten zu berechnen, betrachten wir eine nach $N(0, 1)$ verteilte Zufallsvariable Z und berechnen

$$\begin{aligned}
&\mathbf{P}_{p=1/2}\big(|X_1 + \cdots + X_n - 26.5| \geq 3.5\big) \\
&\qquad = 1 - \mathbf{P}_{p=1/2}\Big(- \frac{3.5}{\sqrt{np(1-p)}} < \frac{Y_n - np}{\sqrt{np(1-p)}} < \frac{3.5}{\sqrt{np(1-p)}}\Big) \\
&\qquad \approx 1 - \mathbf{P}_{p=1/2}(-0.962 \leq Z \leq 0.962\big) \approx 33.6\,\%
\end{aligned}$$

Da dieser Wert größer als $\alpha = 5\,\%$ ist, kann die Hypothese nicht verworfen werden, siehe (0.1).

Wir beginnen, das soeben gesagte in einen formalen Rahmen zu fügen. Siehe auch Abbildung 0.1.

Definition 0.2 (Statistisches Modell). *Seien $S, \Theta, Theta'$ Mengen. Ein* statistisches Modell *ist ein Paar $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$, wobei X eine Zufallsvariable mit Zielbereich S ist, bei deren Verteilung noch ein Parameter $\vartheta \in \Theta$ frei, also unbestimmt, ist. Das bedeutet, dass es eine Funktion $\vartheta \mapsto \rho_\vartheta$ gibt mit*[2]

$$\mathbf{P}_\vartheta(X \in da) = \rho_\vartheta(a)da.$$

[2] Wir wollen im Folgenden die dauernde Unterscheidung zwischen diskreten Zufallsvariablen und Zufallsvariablen mit Dichten durch die Notation $\mathbf{P}(X \in da)$ vermeiden. Ist der Wertebereich S von X diskret und $a \in S$, ist damit

$$\mathbf{P}(X \in da) := \mathbf{P}(X = a)$$

gemeint. Hat X die Dichte $f(a)da$, ist

$$\mathbf{P}(X \in da) := f(a)da.$$

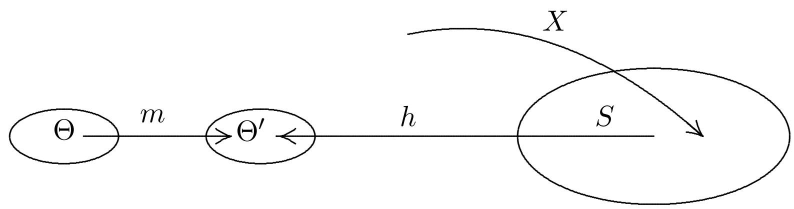

Abbildung 0.1: Veranschaulichung von statistischen Modellen

Die Menge Θ *heißt* Parameterraum*, die Menge* S Beobachtungsraum. *Jedes Zufallsvariable* $h(X)$ *mit* $h : S \to \Theta'$ *heißt* Statistik.

Bemerkung 0.3 (Schätz- und Testprobleme in der Statistik). Die in Beispiel 0.1 angesprochenen Test- und Schätzprobleme ordnen sich nun folgendermaßen ein:

- *Schätzproblem:* Gegeben sei eine Funktion $m : \Theta \to \Theta'$. Bestimme eine Funktion $h : S \to \Theta'$, so dass $h(X)$ den Wert von $m(\vartheta)$ (möglichst gut) schätzt. Was dabei *gute Schätzer* sind, werden wir im nächsten Abschnitt besprechen.
- *Testproblem:* Sei $\Theta_0 \subset \Theta$. Kann die Hypothese $\vartheta \in \Theta_0$ auf Grundlage der Daten verworfen werden?

Beispiel 0.4 (Erfolgswahrscheinlichkeit beim Münzwurf). In der Situation aus Beispiel 0.1 heißt der freie Parameter p anstatt ϑ. Es ist $S = \{0, \dots, 53\}$ und X die Anzahl der Erfolge in $n = 53$ Versuche. Damit ist $\mathbf{P}_p = B(n = 53, p)$ mit dem freien Parameter $p \in \Theta = [0, 1]$. Es ist mit $\Theta' = \Theta$

$$h : \begin{cases} S & \to \Theta, \\ x & \mapsto \frac{1}{53}x. \end{cases}$$

und $\widehat{p} = h(X)$ ist ein Schätzer für p. Für das in Beispiel 0.1 angesprochene Testproblem ist $\Theta_0 = \{\frac{1}{2}\}$. Die Hypothese $p \in \Theta_0$ kann nicht verworfen werden, wie oben gezeigt.

1 Punktschätzer

Der in einem statistischen Modell freie Parameter (den wir meist mit ϑ bezeichnen), kann aus Daten, d.h. der Realisierung der Zufallsvariable X geschätzt werden. Führt so eine Schätzung auf einen einzigen Wert, sprechen wir von Punktschätzern.

1.1 Unverzerrte und konsistente Schätzer

Definition 1.1 (Punktschätzer, unverzerrte und konsistente Schätzer).

1. *Sei* $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ *ein statistisches Modell und* $m : \Theta \to \Theta'$. *Jede Statistik* $\widehat{m} := \widehat{m}(X)$ *mit* $\widehat{m} : S \to \Theta'$ *heißt (Punkt-)Schätzer für* $m(\vartheta)$.

   Der Schätzer $\widehat{m}$ *heißt unverzerrt (erwartungstreu, unbiased), falls*

   $$\mathbf{E}_\vartheta[\widehat{m}] = m(\vartheta)$$

   für alle $\vartheta \in \Theta$.

2. *Sei* $(X^n, (\mathbf{P}^n_\vartheta)_{\vartheta\in\Theta})_{n=1,2,\dots}$ *eine Folge statistischer Modelle mit derselben Parametermenge* Θ *und* $\widehat{m}_1(X^1), \widehat{m}_2(X^2), \dots$ *eine Folge von Schätzern für* $m(\vartheta)$. *Die Folge* $\widehat{m}_1(X^1), \widehat{m}_2(X^2), \dots$ *heißt konsistent, falls*

$$\mathbf{P}^n_\vartheta[|\widehat{m}_n(X^n) - m(\vartheta)| \geq \varepsilon] \xrightarrow{n\to\infty} 0$$

für alle $\varepsilon > 0, \vartheta \in \Theta$.

Beispiel 1.2 (Erfolgswahrscheinlichkeit beim Münzwurf). Betrachten wir noch einmal das einführende Beispiel der Schätzung des Erfolgsparameters in einem Münzwurf in n Versuchen. Wir wählen dazu die Notation aus den Beispielen 0.1 und 0.4. Der Schätzer $\widehat{p} = h(X)$ ist unverzerrt und konsistent, wie wir in Theorem 1.4 zeigen werden.

Ein weiterer Schätzer für p wäre $\widehat{p}' = X_1$. Das bedeutet, dass $\widehat{p}'$ nur die beiden Werte, 0 und 1 annehmen kann, und genau dann 1 ist, wenn der erste Wurf einen Erfolg zeigt. Der Schätzer $\widehat{p}'$ ist ebenfalls erwartungstreu, denn

$$\mathbf{E}_p[\widehat{p}'] = \mathbf{E}_p[X_1] = p.$$

Allerdings ist $\widehat{p}'$ nicht konsistent, da

$$\mathbf{P}_p(|\widehat{p}' - p| > \varepsilon) = 1,$$

falls $\varepsilon < \min(p, 1-p)$, unabhängig von n.

Definition 1.3 (Mittelwert und empirische Varianz). *Sei* $X = (X_1, \dots, X_n)$ *ein Vektor von Zufallsvariablen. Dann heißt*

$$\overline{X} := \tfrac{1}{n}(X_1 + \dots + X_n)$$

Mittelwert von X *und*

$$s^2(X) := \tfrac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$$

empirische Varianz von X.

Theorem 1.4 (Der Mittelwert als Schätzer des Erwartungswertes).
Sei $(X = (X_1, \cdots, X_n), (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ *ein statistisches Modell, so dass* $X_1, \dots, X_n$ *unter allen* $\mathbf{P}_\vartheta$ *identisch verteilt sind und* $\mu_\vartheta := \mathbf{E}_\vartheta[X_1]$ *existiert. Dann ist* $\overline{X}$ *ein unverzerrter Schätzer für* μ_ϑ. *Sind außerdem* $X_1, \dots, X_n$ *unter allen* $\mathbf{P}_\vartheta$ *unabhängig mit endlichem* $\mathbf{V}_\vartheta[X_1]$, *so ist* $\overline{X}$ *auch konsistent.*

Beweis. Zunächst ist

$$\mathbf{E}_\vartheta[\overline{X}] = \tfrac{1}{n}(\mathbf{E}_\vartheta[X_1] + \dots + \mathbf{E}_\vartheta[X_n]) = \mathbf{E}_\vartheta[X_1] = \mu_\vartheta,$$

was bereits die Unverzerrtheit von $\overline{X}$ als Schätzer von μ_ϑ zeigt. Für die Konsistenz berechnen wir für $\varepsilon > 0$

$$\mathbf{P}_\vartheta[|\overline{X} - \mu_\vartheta| \geq \varepsilon] = \mathbf{P}_\vartheta\Big(\Big|\frac{X_1 + \dots + X_n}{n} - \mathbf{E}_\vartheta[X_1]\Big| \geq \varepsilon\Big) \xrightarrow{n\to\infty} 0$$

nach dem schwachen Gesetz der großen Zahlen. □

Theorem 1.5 (Die empirische Varianz als Schätzer der Varianz). *Sei $n \geq 2$ und $X = (X_1, \cdots, X_n)$ so, dass $X_1, \ldots, X_n$ unter allen $\mathbf{P}_\vartheta$ paarweise unkorreliert und identisch verteilt sind mit $\sigma_\vartheta^2 := \mathbf{V}_\vartheta[X_1] < \infty$. Dann ist die empirische Varianz $s^2(X)$ ein unverzerrter Schätzer für σ_ϑ^2.*

Beweis. Wir schreiben zunächst

$$\mathbf{E}_\vartheta[s^2(X)] = \tfrac{1}{n-1} \sum_{i=1}^n \mathbf{E}_\vartheta[X_i^2 - 2X_i\overline{X} + \overline{X}^2] = \tfrac{n}{n-1}\mathbf{E}_\vartheta[X_1^2 - 2X_1\overline{X} + \overline{X}^2].$$

Nun ist

$$\begin{aligned} \mathbf{E}_\vartheta[X_1^2] &= \mu_\vartheta^2 + \sigma_\vartheta^2, \\ \mathbf{E}_\vartheta[X_1\overline{X}] &= \mu_\vartheta^2 + \tfrac{1}{n}\sigma_\vartheta^2, \\ \mathbf{E}_\vartheta[\overline{X}^2] &= \mathbf{E}_\vartheta[X_1\overline{X}], \end{aligned}$$

also

$$\mathbf{E}_\vartheta[s^2(X)] = \tfrac{n}{n-1}\mathbf{E}_\vartheta[X_1^2 - X_1\overline{X}] = \sigma_\vartheta^2,$$

was die Unverzerrtheit bereits zeigt. □

1.2 Risikofunktion

Definition 1.6 (Mittlerer quadratischer Fehler). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und $m : \vartheta \mapsto m(\vartheta)$.*

1. *Für einen Schätzer $h(X)$ von $m(\vartheta)$ ist der* mittlere quadratische Fehler *(oder die* Risikofunktion*) definiert als*

$$R_{h(X)} : \begin{cases} \Theta & \to [0, \infty], \\ \vartheta & \mapsto \mathbf{E}_\vartheta[(h(X) - m(\vartheta))^2]. \end{cases}$$

2. *Sei $\mathcal{S} = \{h_i(X) : i \in \mathcal{I}\}$ eine Menge von Schätzern von $m(\vartheta)$. Falls für ein $h(X) \in \mathcal{S}$ gilt, dass*

$$R_{h(X)}(\vartheta) = \inf_{i\in\mathcal{I}} R_{h_i(X)}(\vartheta),$$

so heißt $h(X)$ bester Schätzer in $\mathcal{S}$.

Beispiel 1.7 (Erfolgswahrscheinlichkeit beim Münzwurf). Wir haben in Beispiel 1.2 gesehen, dass sowohl $\widehat{p} = \overline{X}$ als auch $\widehat{p}' = X_1$ erwartungstreue Schätzer von p sind. Weiter ist

$$\begin{aligned} R_{\widehat{p}}(p) &= \mathbf{V}_p[\widehat{p}] = \mathbf{V}_p\Big[\frac{X_1 + \cdots + X_n}{n}\Big] = \tfrac{1}{n^2}\sum_{i=1}^n \mathbf{V}_p[X_i] = \tfrac{1}{n}\mathbf{V}[X_1] = \tfrac{1}{n}p(1-p), \\ R_{\widehat{p}'}(p) &= \mathbf{V}_p[X_1] = p(1-p), \end{aligned}$$

und damit hat $\widehat{p}$ eine kleinere Risikofunktion als $\widehat{p}'$ (was nicht erstaunen sollte).

Bemerkung 1.8. Es gilt

$$R_{h(X)}(\vartheta) = \mathbf{V}_\vartheta[h(X)] + \big(\mathbf{E}_\vartheta[h(X)] - m(\vartheta)\big)^2.$$

Ist insbesondere $h(X)$ erwartungstreu, so gilt

$$R_{h(X)}(\vartheta) = \mathbf{V}_\vartheta[h(X)].$$

Sei zum Beispiel $X = (X_1, \dots, X_n)$ unter jedem $\mathbf{P}_\vartheta$ unabhängig identisch verteilt und existiert $m(\vartheta) := \mu_\vartheta := \mathbf{E}_\vartheta[X_1]$. Sei weiter

$$\mathcal{S} = \{\overline{X} + c : c \in \mathbb{R}\}$$

eine Menge von Schätzern von $m(\vartheta)$. Dann ist $\overline{X} \in \mathcal{S}$ bester Schätzer in $\mathcal{S}$. Wir werden in Beispiel 1.18 zumindest im Fall, in dem die X_i Poisson-verteilt sind, zeigen, dass $\overline{X}$ bester erwartungstreuer Schätzer des Erwartungswertes, also bester Schätzer in

$$\{h(X) : \mathbf{E}_\vartheta[h(X)] = \mu_\vartheta\}$$

ist.

1.3 Maximum-Likelihood-Schätzer

Das Konzept von Maximum-Likelihood-Schätzern geht davon aus, dass bereits Daten $X = a$ erhoben wurden. Der Maximum-Likelihood-Schätzer ist dann dasjenige ϑ, für das die Wahrscheinlichkeit, die Daten zu erhalten, maximiert wird.

Definition 1.9 (Maximum Likelihood). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell. Die Abbildung*

$$L : \begin{cases} S \times \Theta & \to [0,1] \\ (a, \vartheta) & \mapsto \mathbf{P}_\vartheta(X \in da) \end{cases}$$

heißt Likelihood-Funktion*. Für eine Abbildung $h : S \mapsto \Theta$ mit*

$$L(a, h(a)) = \max_{\vartheta\in\Theta} L(a, \vartheta)$$

heißt $\widehat{\vartheta}_{ML} = h(X)$ Maximum-Likelihood-Schätzer *von ϑ.*

Bemerkung 1.10 (Interpretation von Maximum-Likelihood-Schätzern). Sei X diskret und $X = a$. Da die Vorstellung die ist, dass die erhobenen Daten Ergebnis eines Zufallsexperiments (d.h. die Realisierung einer Zufallsvariable X) sind, sagt man auch, dass die Daten $X = a$ sind. Ein Maximum-Likelihood-Schätzer ist also ein Parameter ϑ, unter dem die Wahrscheinlichkeit, die Daten $X = a$ zu beobachten – das ist $\mathbf{P}_\vartheta(X = a)$ – maximal ist.

Beispiel 1.11 (Erfolgswahrscheinlichkeit beim Münzwurf). Betrachten wir also wieder das Beispiel der Schätzung der Erfolgswahrscheinlichkeit beim Münzwurf. Wir hatten $X = 23$ Erfolge bei $n = 53$ Erfolgen verzeichnet. Unter $\mathbf{P}_p$ ist X nach $B(n = 53, p)$ verteilt, also ist

$$L(X, p) = \binom{53}{X} p^X (1-p)^{n-X}.$$

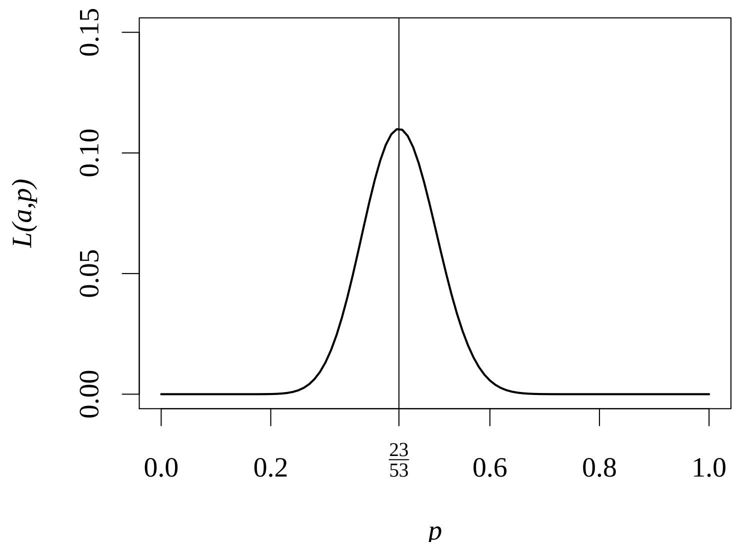

Abbildung 1.1: Die Likelihood-Funktion beim Münzwurf für $X = 23$ aus Beispiel 1.11.

die Likelihood-Funktion; siehe auch Abbildung 1.1. Um diese für gegebenes X zu maximieren, berechnen wir den Wert p, für den $\log L(X,p)$ maximal ist. Die Bestimmung des Maximums der log-Likelihood-Funktion $\log L(a,p)$ genügt, da log eine streng monotone Funktion ist. Wir berechnen

$$\frac{\partial \log L(X,p)}{\partial p} = \frac{\partial}{\partial p}\left(\log\binom{53}{X} + X\log p + (n-X)\log(1-p)\right) = \frac{X}{p} - \frac{n-X}{1-p}.$$

Am Maximum $\widehat{p}_{ML}$ muss $\frac{\partial \log L(X,p)}{\partial p} = 0$ sein, also ist

$$(1-\widehat{p}_{ML})X = \widehat{p}_{ML}(n-X), \qquad \widehat{p}_{ML} = \frac{X}{n}$$

ein Maximum-Likelihood-Schätzer für p. Da es nur ein einziges Maximum der Likelihood gibt, ist dies auch der einzige Maximum-Likelihood-Schätzer.

Beispiel 1.12 (Maximum-Likelihood-Schätzer für $\boldsymbol{\mu}$ und $\boldsymbol{\sigma^2}$ von Normalverteilungen). Wir betrachten den Fall einer unabhängigen, normalverteilten Stichprobe. Sei also $(X, (\mathbf{P}_{(\mu,\sigma^2)})_{\mu\in\mathbb{R},\sigma^2\in\mathbb{R}_+})$ so, dass $X = (X_1,\dots,X_n)$ und $X_1,\dots,X_n$ unter $\mathbf{P}_{(\mu,\sigma^2)}$ unabhängig und identisch nach $N(\mu,\sigma^2)$ verteilt sind.

Wir berechnen nun die Maximum-Likelihood-Schätzer für μ und σ^2. Genau wir im letzten Beispiel berechnen wir zunächst die log-Likelihood-Funktion

$$\begin{aligned}\log L((X_1,\dots,X_n),(\mu,\sigma^2)) &= \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Big(-\sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2}\Big)\right)\\ &= -n\log\sigma - \sum_{i=1}^n \frac{(X_i-\mu)^2}{2\sigma^2} + C,\end{aligned}$$

wobei C weder von μ noch von σ abhängt. Ableiten nach μ und σ ergibt

$$\frac{\partial \log L((X_1,\dots,X_n),(\mu,\sigma^2))}{\partial\mu} = \sum_{i=1}^n \frac{X_i-\mu}{\sigma^2},$$
$$\frac{\partial \log L((X_1,\dots,X_n),(\mu,\sigma^2))}{\partial\sigma} = -\frac{n}{\sigma} + \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^3}.$$

Für die Maximum-Likelihood-Schätzer $\widehat{\mu}_{ML}$ und $\widehat{\sigma}^2_{ML}$ gilt notwendigerweise

$$\sum_{i=1}^n (X_i - \widehat{\mu}_{ML}) = 0,$$
$$\frac{n}{\widehat{\sigma}_{ML}} - \sum_{i=1}^n \frac{(X_i-\widehat{\mu}_{ML})^2}{\widehat{\sigma}^3_{ML}} = 0.$$

Die Maximum-Likelihood-Schätzer sind also gegeben durch

$$\widehat{\mu}_{ML} = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X},$$
$$\widehat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2 = \frac{n-1}{n} s^2(X).$$

Insbesondere sehen wir, dass $\overline{X}$ nicht nur erwartungstreu und konsistent (siehe Theorem 1.4) ist, sondern auch ein Maximum-Likelihood-Schätzer für μ. Allerdings ist der Maximum-Likelihood-Schätzer für σ^2 nicht erwartungstreu, wie man aus Theorem 1.5 abliest. Immerhin ist $\widehat{\sigma}^2_{ML}$ für große n annähernd erwartungstreu, da $\widehat{\sigma}^2_{ML} - s^2(X) \xrightarrow{n\to\infty} 0$.

Beispiel 1.13 (Maximum-Likelihood-Schätzer des Parameters einer Poisson-Verteilung).
Sei $(X, (\mathbf{P}_\lambda)_{\lambda\geq 0})$ so, dass $X = (X_1,\dots,X_n)$ und $X_1,\dots,X_n$ unabhängig und identisch nach $\text{Poi}(\lambda)$ verteilt ist. Wir wissen bereits, dass $\mathbf{E}_\lambda[X_1] = \mathbf{V}_\lambda[X_1] = \lambda$. Damit folgt, dass sowohl $\overline{X}$ als auch $s^2(X)$ unverzerrte Schätzer für λ sind, siehe Theoreme 1.4 und 1.5. Wir berechnen nun den Maximum-Likelihood-Schätzer für λ (welcher sich als $\overline{X}$ herausstellen wird). Später, in Beispiel 1.18, werden wir sehen, dass dieser dem Schätzer $s^2(X)$ vorzuziehen ist, da er eine kleinere Risikofunktion besitzt.

Für den Maximum-Likelihood-Schätzer für λ berechnen wir zunächst wieder die log-Likelihood-Funktion

$$\log L((X_1,\dots,X_n),\lambda) = \log \prod_{i=1}^n e^{-\lambda}\frac{\lambda^{X_i}}{X_i!} = -n\lambda + \log(\lambda)\sum_{i=1}^n X_i + C,$$

wobei C nicht von λ abhängt. Also ist

$$\frac{\partial \log L((X_1,\dots,X_n),\lambda)}{\partial\lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^n X_i. \tag{1.1}$$

Die log-Likelihood-Funktion ist also maximal für

$$-n + \frac{1}{\widehat{\lambda}_{ML}} \sum_{i=1}^n X_i = 0, \qquad \widehat{\lambda}_{ML} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Damit ist $\widehat{\lambda}_{ML} = \overline{X}$ der Maximum-Likelihood-Schätzer für λ. $\square$

Maximum-Likelihood-Schätzer sind in vielen Fällen konsistent, was sicher eine wünschenswerte Eigenschaft ist. Der nächste Satz diese Konsistent von Maximum-Likelihood-Schätzern in einem einfachen Fall.

Theorem 1.14 (Konsistenz von Maximum-Likelihood-Schätzern). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell mit $\Theta \subseteq \mathbb{R}$ endlich, so dass $X = (X_1, \dots, X_n)$ unabhängig und identisch verteilt sind und X_1 eine Zufallsvariable mit Zielbereich $\mathbb{R}$ und einer beschränkten Dichte f_ϑ ist. Dann ist die Folge von Maximum-Likelihood-Schätzern von ϑ für $n = 1, 2, \dots$ konsistent.*

Bemerkung 1.15. Der Satz gilt auch unter schwächeren Voraussetzungen. (Etwa sollte Θ kompakt sein und $\vartheta \mapsto L(a, \vartheta)$ stetig.)

Beweis von Theorem 1.14. Zunächst ist für alle ϑ, ϑ' wegen der Jensen'schen Ungleichung, und, da $x \mapsto \log(x)$ konkav ist,

$$\mathbf{E}_{\vartheta'}\Big[\log \frac{f_\vartheta(X)}{f_{\vartheta'}(X)}\Big] \le \log \mathbf{E}_{\vartheta'}\Big[\frac{f_\vartheta(X)}{f_{\vartheta'}(X)}\Big] = \log \int f_{\vartheta'}(x) \frac{f_\vartheta(x)}{f_{\vartheta'}(x)} dx = \log \int f_\vartheta(x) dx = \log 1 = 0.$$

Der Maximum-Likelihood-Schätzer maximiert die Funktion, die ϑ auf

$$\frac{1}{n} \log L((X_1, \dots, X_n), \vartheta) - \frac{1}{n} \log L((X_1, \dots, X_n), \vartheta_0) = \frac{1}{n} \sum_{i=1}^n \log \frac{f_\vartheta(X_i)}{f_{\vartheta_0}(X_i)}$$

abbildet. Wegen dem starken Gesetz der großen Zahlen ist

$$\frac{1}{n} \log L((X_1, \dots, X_n), \vartheta) - \frac{1}{n} \log L((X_1, \dots, X_n), \vartheta_0) \xrightarrow{n \to \infty} \mathbf{E}_\vartheta\Big[\log \frac{f_\vartheta(X)}{f_{\vartheta_0}(X)}\Big] \le 0$$

mit $= 0$ genau dann, wenn $f_\vartheta = f_{\vartheta_0}$. Da Θ diskret ist, konvergiert die Folge der Maximum-Likelihood-Schätzer für ϑ gegen ϑ_0. $\square$

Die Likelihood-Funktion liefert sogar noch mehr Informationen über ein statistisches Modell. Wir werden in Theorem 1.17 sehen, dass man aus ihr eine untere Schranke für die Risikofunktion eines Schätzers ableiten kann.

Definition 1.16 (Fisher-Information und effiziente Schätzer). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell und $(a, \vartheta) \mapsto L(a, \vartheta)$ die Likelihood-Funktion.*

1. *Die Abbildung*

$$\vartheta \mapsto \mathcal{I}(\vartheta) := \mathbf{E}_\vartheta\Big[\Big(\frac{\partial}{\partial \vartheta} \log L(X, \vartheta)\Big)^2\Big]$$

   heißt Fisher-Information.

2. *Sei $h(X)$ ein Schätzer für ϑ. Dann heißt $h(X)$* effizient, *falls für alle $\vartheta \in \Theta$*

$$R_{h(X)}(\vartheta) = \frac{1}{\mathcal{I}(\vartheta)}.$$

Theorem 1.17 (Cramér-Rao-Schranke, ohne Beweis). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und $\vartheta \mapsto \mathcal{I}(\vartheta)$ die Fisher-Information. Ist $h(X)$ ein unverzerrter Schätzer für ϑ, dann gilt*

$$R_{h(X)}(\vartheta) \geq \frac{1}{\mathcal{I}(\vartheta)}.$$

Beispiel 1.18 (Effizienz des Schätzers für den Parameter einer Poisson-Verteilung).
Sei $(X, (\mathbf{P}_\lambda)_{\lambda\geq 0})$ so, dass $X = (X_1, \dots, X_n)$ und $X_1, \dots, X_n$ unabhängig nach Poi(λ) verteilt ist. Wir haben bereits in Beispiel 1.13 berechnet, dass

$$\widehat{\lambda}_{ML} = \tfrac{1}{n}\sum_{i=1}^n X_i$$

der Maximum-Likelihood-Schätzer für λ ist. Wir wollen nun sehen, ob $\widehat{\lambda}_{ML}$ auch effizient ist. Dafür berechnen wir die Fisher-Information (siehe (1.1) für die Ableitung der log-Likelihood-Funktion)

$$\mathcal{I}(\lambda) = \mathbf{E}_\lambda\Big[\Big(\frac{\partial}{\partial\lambda}\log L(X,\lambda)\Big)^2\Big] = \mathbf{E}_\lambda\Big[\Big(-n + \frac{1}{\lambda}\sum_{i=1}^n X_i\Big)^2\Big] = \mathbf{V}_\lambda\Big[\frac{1}{\lambda}\sum_{i=1}^n X_i\Big] = \frac{n}{\lambda}.$$

Der Schätzer $\widehat{\lambda}_{ML}$ ist also effizient, da

$$\mathbf{V}_\lambda[\widehat{\lambda}_{ML}] = \frac{\lambda}{n} = \frac{1}{\mathcal{I}(\lambda)}.$$

1.4 Suffiziente Statistiken

Wir betrachten statistische Modelle $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ (wobei X Wertebereich S hat). Zum Schätzen von ϑ haben wir bisher Funktionen $h : S \to \Theta$ verwendet. In manchen Situationen genügt es, Schätzer zu betrachten, die von einer Funktion von X abhängen. In diesem Fall gibt es also ein $\nu : S \to \widetilde{S}$ und der Schätzer ist eine Funktion $h \circ \nu$ mit $h : \widetilde{S} \to \Theta$. Man fasst also die erhobenen Daten X durch eine Funktion $\nu(X)$ zusammen, und der Schätzer für ϑ hängt nur von dieser Statistik $\nu(X)$ ab. Der Begriff der Suffizienz einer Statistik ν gibt an, für welche Funktionen ν man dabei sogar vorhandene Schätzer besser (im Sinne einer kleineren Risikofunktion) machen kann.

Definition 1.19 (Suffiziente Statistik). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und $\mathbf{P}(X \in da) = \rho_\vartheta(da)$. Eine Zufallsvariable $\nu(X)$ mit Zielbereich $\widetilde{S}$ heißt suffiziente Statistik für ϑ, falls für alle $b \in \widetilde{S}$*

$$\mathbf{P}_\vartheta(X \in \cdot \mid \nu(X) \in db) \textit{ nicht von } \vartheta \textit{ abhängt.}$$

Beispiel 1.20 (Münzwurf). Sei $X = (X_1, \dots, X_n) \in \{0,1\}^n$ ein p-Münzwurf mit noch unbestimmtem p. Der statistische Raum ist also $(X, (\mathbf{P}_p)_{p\in[0,1]})$, wobei

$$\mathbf{P}_p(X = (a_1, \dots, a_n)) = p^k(1-p)^{n-k}, \text{ falls } k = \sum_{i=1}^n a_i.$$

Die Statistik

$$\nu(X_1, \dots, X_n) = X_1 + \cdots + X_n$$

ist suffizient für p. Denn es gilt für $k = a_1 + \cdots + a_n$

$$\begin{aligned}\mathbf{P}_p(X = (a_1, \dots, a_n) \mid X_1 + \cdots + X_n = k) &= \frac{\mathbf{P}_p(X = (a_1, \dots, a_n))}{\mathbf{P}_p(X_1 + \cdots + X_n = k)} = \frac{p^k(1-p)^{n-k}}{\binom{n}{k}p^k(1-p)^{n-k}} \\ &= \frac{1}{\binom{n}{k}},\end{aligned}$$

unabhängig von p. Für $k \neq \sum_{i=1}^n a_i$ gilt

$$\mathbf{P}_p(X = (a_1, \dots, a_n) \mid X_1 + \cdots + X_n = k) = 0,$$

ebenfalls unabhängig von p.

Theorem 1.21 (Charakterisierung von Suffizienz). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell und der Wertebereich S von X diskret. Eine Statistik $\nu(X)$ ist genau dann suffizient für ϑ, wenn es eine (von ϑ unabhängige) Funktion $P : \widetilde{S} \times S \to [0,1]$ gibt mit*

$$\mathbf{P}_\vartheta(X = a) = \mathbf{P}_\vartheta(\nu(X) = \nu(a)) \cdot P(\nu(a), a) \quad \text{für alle } a \in S.$$

Beweis. '$\Rightarrow$': Definiere $P(b, a) = \mathbf{P}_\vartheta(X = a|\nu(X) = b)$. Da $\nu(X)$ suffizient für ϑ ist, hängt dies nicht von ϑ ab.

'$\Leftarrow$': Klar, da

$$\mathbf{P}_\vartheta(X = a|\nu(X) = \nu(a)) = \frac{\mathbf{P}_\vartheta(X = a)}{\mathbf{P}_\vartheta(\nu(X) = \nu(a))} = P(\nu(a), a)$$

nicht von ϑ abhängt. □

Beispiel 1.22 (Münzwurf). Im Beispiel des Münzwurfs aus Beispiel 1.20 ist $\nu(X) = X_1 + \cdots + X_n$. Die Funktion $P(\cdot, \cdot)$ aus dem letzten Theorem ist in diesem Fall

$$P(k, (a_1, \dots, a_n)) = \begin{cases} \frac{1}{\binom{n}{k}}, & a_1 + \cdots + a_n = k, \\ 0, & \text{sonst.} \end{cases}$$

Wir kommen nun dazu, Schätzer für ϑ aufzustellen, die nur von einer suffizienten Statistik $\nu(X)$ abhängen; siehe die Eingangsbemerkung zu diesem Abschnitt.

Theorem 1.23 (Satz von Rao-Blackwell). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, $m : \Theta \to \Theta'$, $h : S \to \Theta'$ mit $h(X)$ einem Schätzer für $m(\vartheta)$. Ist $\nu(X)$ suffizient für ϑ, so ist*

$$\widetilde{h}(X) = \mathbf{E}_\vartheta[h(X)|\nu(X)] \tag{1.2}$$

unabhängig von ϑ und

$$\mathbf{E}_\vartheta[(\widetilde{h}(X) - m(\vartheta))^2] \le \mathbf{E}_\vartheta[(h(X) - m(\vartheta))^2].$$

Bemerkung 1.24 (Interpretation). Das Theorem definiert eine Funktion $\widetilde{h} : S \to \Theta$. Wichtig ist, dass $\widetilde{h}(X)$ nur von $\nu(X)$ abhängt. Es gilt nämlich

$$\widetilde{h}(a) = \mathbf{E}_\vartheta[h(X)|\nu(X) = \nu(a)]$$

nach der Definition der bedingten Erwartung, und die rechte Seite hängt wegen der Definition der bedingten Erwartung nur von $\nu(a)$ ab. Die Aussage des Satzes ist, dass der Schätzer $\widetilde{h}(X)$ eine kleinere Risikofunktion hat als $h(X)$.

Beispiel 1.25 (Die Schätzer $\widehat{p}$ und $\widehat{p}'$ beim Münzwurf). Sei wieder $X = (X_1, \dots, X_n)$ ein p-Münzwurf mit noch unbestimmten p. In Beispiel 1.2 und 1.7 haben wir zwei erwartungstreue Schätzer für p kennen gelernt, nämlich

$$\widehat{p} = \widehat{p}(X) = \tfrac{1}{n}(X_1 + \cdots + X_n), \qquad \widehat{p}' = \widehat{p}'(X) = X_1$$

und hatten auch festgestellt, dass $\widehat{p}$ eine kleinere Risikofunktion besitzt als $\widehat{p}'$. Weiter wissen wir aus Beispiel 1.20, dass $\nu(X) = X_1 + \cdots + X_n$ suffizient für p ist. Wir können also nun den Schätzer $\widehat{p}'$ mit Hilfe des Satzes von Rao-Blackwell verbessern, indem wir $h(X) = \widehat{p}'(X)$ setzen und $\widetilde{h}(X)$ aus (1.2) berechnen. Dies ergibt aus Symmetriegründen

$$\begin{aligned}\widetilde{h}(X) = \mathbf{E}_p[\widehat{p}' \mid \nu(X)] = \mathbf{E}_p[X_1 \mid X_1 + \cdots + X_n] &= \tfrac{1}{n}\sum_{i=1}^n \mathbf{E}_p[X_i \mid X_1 + \cdots + X_n] \\ &= \tfrac{1}{n}\sum_{i=1}^n \mathbf{E}_p[X_1 + \cdots + X_n \mid X_1 + \cdots + X_n] = \tfrac{1}{n}\sum_{i=1}^n X_1 + \cdots + X_n \\ &= \widehat{p}(X).\end{aligned}$$

Insbesondere sehen wir hier ein konkretes Beispiel dafür, dass $\widetilde{h}(X)$ eine echt kleinere Risikofunktion besitzt als $h(X)$, siehe Beispiel 1.7.

Beweis von Theorem 1.23. Wir beschränken den Beweis auf den Fall diskreter Zufallsvariablen X. Der Fall von Zufallsvariablen mit Dichten geht analog. Zunächst ist

$$\mathbf{E}_\vartheta[h(X) \mid \nu(X)] = \sum_{a \in S} h(a)\mathbf{P}_\vartheta(X = a \mid \nu(X)),$$

was wegen der Suffizienz von $\nu(X)$ nicht von ϑ abhängt. Wegen (siehe Kapitel 4)

$$\mathbf{E}_\vartheta[\widetilde{h}(X)] = \mathbf{E}_\vartheta[\mathbf{E}_\vartheta[h(X) \mid \nu(X)]] = \mathbf{E}_\vartheta[h(X)]$$

gilt mit der Varianzformel (ebenfalls aus Kapitel 4)

$$\begin{aligned}\mathbf{E}_\vartheta[(\widetilde{h}(X) - m(\vartheta))^2] &= \big(\mathbf{E}_\vartheta[\widetilde{h}(X)] - m(\vartheta)\big)^2 + \mathbf{Var}_\vartheta[\mathbf{E}_\vartheta[h(X) \mid \nu(X)]] \\ &\le \big(\mathbf{E}_\vartheta[h(X)] - m(\vartheta)\big)^2 + \mathbf{Var}_\vartheta[\mathbf{E}_\vartheta[h(X)|\nu(X)]] + \mathbf{E}_\vartheta[\mathbf{Var}_\vartheta[h(X)|\nu(X)]] \\ &= \mathbf{E}_\vartheta[(h(X) - m(\vartheta))^2]\end{aligned}$$

und die Behauptung ist gezeigt. □

2 Von der Normalverteilung abgeleitete Verteilungen

Die Normalverteilung hat eine zentrale Stellung in der Statistik. Grund hierfür ist der Zentrale Grenzwertsatz: kommt eine zufällige Größe durch viele unabhängige Einflüsse zu Stande, so ist sie annähernd normalverteilt. Wir lernen nun einige aus Normalverteilungen abgeleitete Verteilungen kennen.

Definition 2.1 (Die χ^2- und die t-Verteilung). *Seien $X, X_1, \ldots, X_n$ unabhängige, nach $N(0,1)$ verteilte Zufallsvariablen.*

1. *Die Verteilung der Zufallsvariable*

$$X_1^2 + \cdots + X_n^2$$

heißt χ^2-Verteilung mit n Freiheitsgraden *und wird mit $\chi^2(n)$ bezeichnet.*

2. *Die Verteilung von*

$$\frac{X}{\sqrt{(X_1^2 + \cdots + X_n^2)/n}}$$

heißt t-Verteilung mit n Freiheitsgraden *und wird mit $t(n)$ bezeichnet.*

2.1 Die Gamma-Funktion

Die Dichten der $\chi^2(n)$- und $t(n)$-Verteilung können explizit angegeben werden. Hierfür benötigen wir jedoch ein paar Hilfsresultate über die Gamma-Funktion.

Definition 2.2 (Gamma- und Beta-Funktion). *Die Gamma-Funktion hat Definitionsbereich $(0, \infty)$ und ist definiert durch*

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Die Beta-Funktion hat Definitionsbereich $(0, \infty)^2$ und ist definiert durch

$$B(x, y) := \int_0^1 t^{x-1}(1-t)^{y-1} dt.$$

In Propositionen 2.6 und 2.7 werden wir die Dichten der $\chi^2(n)$- und $t(n)$-Verteilung herleiten. Um einige der Eigenschaften dieser Dichten zu zeigen, benötigen wir zunächst ein paar Aussagen über die Gamma-Funktion.

Lemma 2.3. 1. *Für $x > 0$ ist*

$$\Gamma(x+1) = x\Gamma(x).$$

2. *Für $x, y > 0$ gilt*

$$\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = B(x, y).$$

3. Es gilt

$$\Gamma(1) = 1, \qquad \Gamma(1/2) = \sqrt{\pi}.$$

Beweis. Mittels partieller Integration folgt 1. aus

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t} dt = -t^x e^{-t} \mid_0^\infty + \int_0^\infty x t^{x-1} e^{-t} dt = x\Gamma(x).$$

Für 2. berechnen wir direkt, mit Hilfe des Transformationssatzes,

$$\begin{aligned}
\Gamma(x)\Gamma(y) &= \int_0^\infty \int_0^\infty s^{x-1} t^{y-1} e^{-(s+t)} ds dt \\
&\overset{r=t+s}{=} \int_0^\infty \int_0^r s^{x-1}(r-s)^{y-1} e^{-r} ds dr \\
&\overset{u=s/r}{=} \int_0^\infty \int_0^1 r u^{x-1} r^{x-1} r^{y-1} (1-u)^{y-1} e^{-r} du dr \\
&= \Gamma(x+y) \cdot \int_0^1 u^{x-1}(1-u)^{y-1} du.
\end{aligned}$$

und die Aussage folgt. Für 3. gilt erst einmal durch einfache Integration

$$\Gamma(1) = \int_0^\infty e^{-t} dt = 1.$$

Für die zweite Aussage verwenden wir 2. und schreiben

$$\Gamma(1/2)^2 = B(1/2, 1/2) = \int_0^1 \frac{dt}{\sqrt{t(1-t)}} \overset{s=2t-1}{=} \int_{-1}^1 \frac{ds}{\sqrt{1-s^2}} = \arcsin(s)|_{-1}^1 = \pi$$

und das Lemma ist bewiesen. □

2.2 Die Dichte der Summe und des Quadrats von Zufallsvariablen

Eine $\chi^2(n)$-verteilte Zufallsvariable ist die Summe von Quadraten von normalverteilten Zufallsvariablen. Um deren Dichte zu berechnen, zeigen wir nun zwei Aussagen, die allgemein die Dichte von Summen und von Quadraten von bekannten Verteilungen liefern.

Lemma 2.4. *Sei X eine Zufallsvariable mit Zielbereich $S \subseteq \mathbb{R}$. Hat die Verteilung von X eine Dichte $f(a)da$, dann hat X^2 die Dichte*

$$\frac{1}{2\sqrt{a}}(f(\sqrt{a}) + f(-\sqrt{a}))1_{a \geq 0} da.$$

Beweis. Es gilt

$$\begin{aligned}
\mathbf{P}(X^2 \leq x) &= \mathbf{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = \int_0^{\sqrt{x}} f(a) + f(-a) da \\
&\overset{b=a^2}{=} \int_0^x \frac{1}{2\sqrt{b}}(f(\sqrt{b}) + f(-\sqrt{b})) db
\end{aligned}$$

nach dem Transformationssatz. Dies zeigt bereits, dass X^2 die angegebene Dichte besitzt. □

Lemma 2.5. *Seien X, Y unabhängige Zufallsvariablen mit Zielbereich $S \subseteq \mathbb{R}$. Haben X und Y die Dichten $f(a)da$ und $g(a)da$, so hat $X+Y$ die Dichte*

$$\Big(\int_{-\infty}^{\infty} f(b)g(a-b)db\Big)da.$$

Beweis. Wir schreiben

$$\begin{aligned}\mathbf{P}(X+Y \le z) &= \int_{-\infty}^{\infty} \mathbf{P}(X \in db, Y \le z-b) = \int_{-\infty}^{\infty} \mathbf{P}(X \in db) \cdot \mathbf{P}(Y \le z-b) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-b} f(b)g(a)dadb = \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(b)g(a-b)dbda.\end{aligned}$$ □

2.3 Die χ^2- und die t-Verteilung

Nachdem wir bereits in Definition 2.1 sowohl die $\chi^2(n)$- als auch die $t(n)$-Verteilung definiert haben, leiten wir nun deren Dichten her. Anschließend folgt mit dem Satz von Fisher in Theorem 2.8 eine Aussage, die wir in den nächsten Abschnitten oft brauchen werden.

Proposition 2.6 (Dichte der χ^2-Verteilung). *Sei X_n eine $\chi^2(n)$-verteilte Zufallsvariable. Diese hat die Dichte*

$$\frac{1}{2^{n/2}\Gamma(n/2)} a^{n/2-1} e^{-a/2} da.$$

Weiter ist

$$\mathbf{E}[X_n] = n, \qquad \mathbf{Var}[X_n] = 2n.$$

Beweis. Zunächst zeigen wir, dass die angegebene Formel wirklich eine Dichte ist, d.h. zu 1 integriert. Dies folgt sofort aus

$$\int_0^{\infty} a^{n/2-1}e^{-a/2}da \overset{b=a/2}{=} 2^{n/2}\int_0^{\infty} b^{n/2-1}e^{-b}db = 2^{n/2}\cdot\Gamma(n/2).$$

Seien nun $Y_1, Y_2, \dots$ unabhängige, nach $N(0,1)$ verteilte Zufallsvariablen sowie $X_n = Y_1^1 + \cdots + Y_n^2$. Um die Form der Dichte einer $\chi^2(n)$-verteilten Zufallsvariablen zu zeigen, gehen wir mit Induktion vor. Für $n=1$ ist X_1 genauso verteilt wie Y_1^2. Nach Lemma 2.4 gilt also

$$\mathbf{P}(X_1 \le x) = \int_0^x \frac{1}{\sqrt{a}}\frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{a}{2}\Big)da.$$

Da diese zu 1 integriert, folgt die Aussage für $n=1$. Für den Induktionsschritt sei die Aussage also schon für n gezeigt. Es ist X_{n+1} genauso verteilt wie $X_n + X_{n+1}^2$. Wir verwenden Lemma 2.5 und schreiben, für Konstanten C_n die nur von n abhängen, und von Zeile zu Zeile verschieden sein können,

$$\begin{aligned}\mathbf{P}(X_{n+1} \in da) &= \mathbf{P}(X_n + Y_{n+1}^2 \in da) \\ &= C_n \cdot \int_0^a b^{n/2-1}e^{-b/2}(a-b)^{-1/2}e^{-(a-b)/2}db \\ &= C_n \cdot e^{-a/2}\int_0^a b^{n/2-1}(a-b)^{-1/2}db \\ &\overset{c=b/a}{=} C_n \cdot e^{-a/2}aa^{n/2-1}a^{-1/2}\int_0^1 c^{n/2-1}(1-c)^{-1/2}dc \\ &= C_n \cdot e^{-a/2}a^{(n+1)/2-1}.\end{aligned}$$

Damit folgt die Aussage, weil die Funktion in der letzten Zeile zu 1 integrieren muss.

Um Erwartungswert und Varianz von X abzuleiten, berechnen wir

$$\mathbf{E}[X_n] = \sum_{i=1}^n \mathbf{E}[Y_i^2] = n.$$

Zur Berechnung der Varianz von X_n ist zunächst

$$\int_{-\infty}^{\infty} x^3 \cdot x \exp\Big(-\frac{x^2}{2}\Big) dx = -x^3 \exp\Big(-\frac{x^2}{2}\Big)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} 3x^2 \exp\Big(-\frac{x^2}{2}\Big) dx = 3\sqrt{2\pi},$$

also $\mathbf{E}[Y_1^4] = 3$. Damit gilt

$$\mathbf{Var}[X_n] = \sum_{i=1}^n \mathbf{Var}[Y_i^2] = n(\mathbf{E}[Y_1^4] - 1) = 2n. \qquad \square$$

Proposition 2.7 (Dichte der t-Verteilung). *Sei X_n eine nach $t(n)$-verteilte Zufallsvariable. Diese hat Dichte*

$$\frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \cdot \frac{1}{(1+a^2/n)^{(n+1)/2}} da.$$

Weiter ist

$$\begin{aligned} \mathbf{E}[X_n] &= 0 \textit{ für } n \geq 2, \\ \mathbf{Var}[X_n] &= \frac{n}{n-2} \textit{ für } n \geq 3. \end{aligned}$$

Beweis. Wir zeigen zunächst, dass die angegebene Funktion eine Dichte ist. Wegen Lemma 2.3 und $a^2/n = t/(1-t) \iff t = a^2/(n+a^2)$ ist

$$\begin{aligned} \int_{-\infty}^{\infty} \frac{1}{(1+a^2/n)^{(n+1)/2}} da &= \int_0^1 \frac{1}{(1+\frac{t}{1-t})^{(n+1)/2}} \sqrt{\frac{1-t}{t}} \frac{\sqrt{n}}{(1-t)^2} dt \\ &= \sqrt{n} \int_0^1 t^{-1/2}(1-t)^{n/2-1} dt = \sqrt{n} B(1/2, n/2) \\ &= \sqrt{n} \frac{\Gamma(1/2)\Gamma(n/2)}{\Gamma((n+1)/2)} = \sqrt{n\pi} \frac{\Gamma(n/2)}{\Gamma((n+1)/2)}. \end{aligned}$$

Für $x \in \mathbb{R}$ sei nun

$$A := \Big\{(a,b) : \frac{a}{\sqrt{b/n}} \leq x\Big\}.$$

Sei Y eine $N(0,1)$-verteilte Zufallsvariable und Z eine unabhängige, $\chi^2(n)$-verteilte Zufallsvariable Z. Für C_n's, die nur von n abhängen, und alle verschieden sind, ist

$$\begin{aligned} \mathbf{P}(X_n \leq x) &= \mathbf{P}((Y,Z) \in A) = C_n \int_A \int e^{-a^2/2} b^{n/2-1} e^{-b/2} da db \\ &\overset{c=\sqrt{n}a/\sqrt{b}}{=} C_n \int_{-\infty}^x \int_0^{\infty} e^{-(c^2/n+1)b/2} b^{n/2-1} \sqrt{b}\, db dc \\ &\overset{d=(1+c^2/n)b/2}{=} C_n \int_{-\infty}^x \int_0^{\infty} e^{-d} \frac{(2d)^{(n-1)/2}}{\big(1+c^2/n\big)^{(n+1)/2}} dd\, dc \\ &= C_n \int_{-\infty}^x \frac{1}{\big(1+c^2/n\big)^{(n+1)/2}} dc \end{aligned}$$

Da die Dichte zu 1 integrieren muss, folgt die Aussage. Für die Berechnung des Erwartungswertes und der Varianz, siehe Übung!

□

Wir kommen nun zu einem Satz, den wir später noch oft brauchen werden. Wir erinnern an die Begriffe des Mittelwertes und der empirischen Varianz aus Definition 1.3.

Theorem 2.8 (Satz von Fisher). *Sei $n > 1$, $X = (X_1, .., X_n)$ unabhängig nach $N(\mu, \sigma^2)$ verteilt, $\overline{X}$ der Mittelwert und $s^2(X)$ die empirische Varianz. Dann ist*

$$\overline{X} \text{ nach } N(\mu, \sigma^2/n)$$

und

$$(n-1)s^2(X)/\sigma^2 \text{ nach } \chi^2(n-1)$$

verteilt. Außerdem sind $\overline{X}$ und $s^2(X)$ stochastisch unabhängig und

$$T := \frac{\overline{X} - \mu}{\sqrt{s^2(X)/n}} \tag{2.1}$$

ist nach $t(n-1)$ verteilt.

Bemerkung 2.9 (Interpretation). Stellen wir uns vor, es liegt uns eine Stichprobe $X = (X_1, \ldots, X_n)$ vor, wobei wir davon ausgehen dürfen, dass $X_1, \ldots, X_n$ unabhängig sind und X_i normalverteilt sind. Wollen wir etwa die Verteilung des Mittelwertes berechnen, wissen wir, dass

$$\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}}$$

nach $N(0,1)$ verteilt ist. Eine gängige Situation ist nun die, dass wir eine Vermutung haben, wie groß μ ist, jedoch nicht wissen, wie groß σ^2 ist. Es liegt nun nahe, σ^2 in der letzten Formel durch $s^2(X)$ zu ersetzen, wie es in T aus (2.1) geschehen ist. Durch das Ersetzen der festen Größe σ^2 durch die Zufallsvariable $s^2(X)$ verändert sich natürlich die Verteilung. Der Satz von Fisher besagt nun, dass die Verteilung der Standardisierung von $\overline{X}$, wenn man σ^2 durch $s^2(X)$ ersetzt, nach $t(n-1)$ verteilt ist. Bemerkenswert ist dabei, dass die Verteilung von T nicht mehr von σ^2 abhängt.

Beweis von Theorem 2.8. Wir setzen $Z = (Z_1, \ldots, Z_n)$ mit $Z_i := (X_i - \mu)/\sigma$. Wir wissen bereits, dass die Z_i unabhängig und nach $N(0,1)$ verteilt sind. Weiter ist

$$\overline{Z} = \tfrac{1}{n} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} = (\overline{X} - \mu)/\sigma$$

und

$$(n-1)s^2(Z) = \sum_{i=1}^{n} \Big(\frac{X_i - \mu - \overline{X} + \mu}{\sigma} \Big)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 = (n-1)s^2(X)/\sigma^2.$$

Damit ist

$$T = \frac{\overline{X} - \mu}{\sqrt{s^2(X)/n}} = \frac{\sigma \cdot \overline{Z}}{\sqrt{\sigma^2 s^2(Z)/n}} = \frac{\overline{Z}}{\sqrt{s^2(Z)/n}}$$

und damit genügt es, die Behauptungen für den Vektor Z zu zeigen.

Wir wählen nun eine orthogonale Matrix $O = (o_{ij})_{1\leq i,j\leq n}$ mit

$$a_{11} = \cdots = a_{1n} = \frac{1}{\sqrt{n}}$$

und setzen $W = OZ$. Wir verwenden im Beweis die aus Kapitel 3 bekannte Tatsache, dass W ein Vektor unabhängiger, nach $N(0,1)$ verteilter Zufallsvariablen ist. Weiter ist

$$\begin{aligned} W_1 &= a_{11}Z_1 + \cdots + a_{1n}Z_n = \sqrt{n}\cdot\overline{Z}, \\ (n-1)s^2(Z) &= \sum_{i=1}^n (Z_i - \overline{Z})^2 = \sum_{i=1}^n Z_i^2 - n\overline{Z}^2 \\ &= \sum_{i=1}^n Z_i^2 - W_1^2 = \sum_{i=2}^n W_i^2 \end{aligned}$$

da O eine orthogonale Matrix ist und damit

$$\sum_{i=1}^n Z_i^2 = \sum_{i=1}^n (OZ)_i^2 = \sum_{i=1}^n W_i^2.$$

Insbesondere haben wir eben gezeigt, dass $(n-1)s^2(X)$ die Summe von $n-1$ unabhängigen, nach $N(0,1)$ verteilten Zufallsvariablen ist und damit $\chi^2(n-1)$ verteilt. Da weiter W_1 von $(W_2,\dots,W_n)$ unabhängig ist, folgt auch, dass $\overline{X}$ von $s^2(X)$ unabhängig ist.

Es bleibt nun zu zeigen, dass T nach $t(n-1)$ verteilt ist. Wir schreiben

$$T = \frac{\overline{Z}}{\sqrt{s^2(Z)/n}} = \frac{\sqrt{n}\overline{Z}}{\sqrt{s^2(Z)}} = \frac{W_1}{\sqrt{(W_2^2 + \cdots + W_n^2)/(n-1)}}.$$

Da $W_1,\dots,W_n$ unabhängig und nach $N(0,1)$ verteilt sind, folgt, dass T nach $t(n-1)$ verteilt ist. □

3 Intervallschätzer

In Abschnitt 1 haben wir bereits Punktschätzer kennen gelernt. Diese geben keinen Aufschluss darüber, wie sicher man sich beim Schätzen ist. In diesem Abschnitt lernen wir nun Intervallschätzer kennen. Diese geben nicht einen punktgenauen Wert für den zu schätzenden Parameter, sondern geben ein Intervall an, in dem der Parameter mit hoher Wahrscheinlichkeit liegt.

3.1 Konfidenzintervalle und Quantile

Wir beginnen mit der Definition von Konfidenzintervallen und des wichtigen Begriffes des Quantils.

Definition 3.1 (Konfidenzintervall). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, S der Zielbereich von X und $m : \Theta \to \Theta' \subseteq \mathbb{R}$ ein Intervall.*

1. *Wir bezeichnen mit $\mathcal{I}(\Theta')$ die Menge der Intervalle, die Teilmenge von Θ' sind. Für eine Abbildung $K : S \to \mathcal{I}(\Theta')$ heißt $K(X)$ Konfidenzintervall für $m(\vartheta)$.*

2. *Ist* $\alpha \in (0,1)$ *und gilt* $\mathbf{P}_\vartheta(m(\vartheta) \in K(X)) \geq 1-\alpha$ *für alle* $\vartheta \in \Theta$*, so heißt* $K(x)$ *Konfidenzintervall zum (Konfidenz-)Niveau* $1-\alpha$*.*

Bemerkung 3.2 (Triviales Konfidenzintervall und Interpretation).

1. Das einfachste Konfidenzintervall ist sicherlich $K(X) = \Theta'$. Immerhin gilt ja auch $\mathbf{P}_\vartheta(m(\vartheta) \in \Theta') = 1$, und damit hat dieses Konfidenzintervall Niveau 1. Allerdings ist dieses Intervall natürlich völlig wertlos, weil es keine Aussage über den wahren Wert $m(\vartheta)$ trifft.

2. Man kann sagen, dass ein Konfidenzintervall zum Niveau $1-\alpha$ ein Intervall ist, in dem der wahre Wert mit Wahrscheinlichkeit $1-\alpha$ liegt. Wichtig ist es jedoch zu bemerken, dass natürlich nicht der wahre Wert *zufällig* ist, sondern dass $K(X)$ von der Zufallsvariable X abhängt und $K(X)$ damit ein zufälliges Intervall darstellt. Die Wahrscheinlichkeit von $1-\alpha$ kommt also dadurch zustande, dass mit einer Wahrscheinlichkeit von höchstens α die Daten so sind, dass das von den Daten abgeleitete, zufällige, Konfidenzintervall den wahren Wert $m(\vartheta)$ nicht beinhaltet.

Für das Folgende ist der Begriff des Quantils zentral.

Definition 3.3 (Quantil). *Sei* X *eine Zufallsvariable mit Zielbereich* $S \subseteq \mathbb{R}$*. Eine Zahl* $q_\alpha \in \mathbb{R}$ *heißt* α-Quantil *von (der Verteilung von)* X*, falls*

$$\mathbf{P}(X < q_\alpha) \leq \alpha \leq \mathbf{P}(X \leq q_\alpha).$$

Jedes $q_{0.5}$*-Quantil heißt* Median *von (der Verteilung von)* X*.*

Bemerkung 3.4 (Die Quantilfunktion als Umkehrung der Verteilungsfunktion). Sei F die Verteilungsfunktion von X, also $F(x) = \mathbf{P}(X \leq x)$. Für eine Funktion $\alpha \mapsto q_\alpha$, die jedem α ein α-Quantil zuordnet, gilt

$$F(q_\alpha-) = \mathbf{P}(X < q_\alpha) \leq \alpha \leq P(X \leq q_\alpha) = F(q_\alpha).$$

Ist insbesondere F stetig, was genau dann gilt, wenn X eine Dichte besitzt, so ist also

$$F(q_\alpha) = \alpha.$$

In diesem Fall ist also $\alpha \mapsto q_\alpha$ eine Umkehrfunktion von $x \mapsto F(x)$. Dabei ist die Umkehrfunktion von F genau dann eindeutig, wenn F bijektiv ist, was genau dann der Fall ist, wenn $F' > 0$, die Verteilung von X also eine Dichte besitzt, die nirgends 0 ist.

3.2 Zwei Beispiele: Binomialverteilung und Median

Wir lernen nun zwei Beispiele von Konfidenzintervallen kennen. Zunächst beschäftigen wir uns nochmal mit der Erfolgswahrscheinlichkeit p eines Münzwurfes. Anschließend schätzen wir den Median einer Verteilung.

Proposition 3.5 (Konfidenzintervall für den Erfolgsparameter der Binomialverteilung). *Sei* $n \in \mathbb{N}, \Theta = [0,1]$ *und* $(X, (\mathbf{P}_p)_{p\in\Theta}$*, so dass* X *unter* $\mathbf{P}_p$ *binomialverteilt ist. Für* $\alpha \in (0,1)$ *definieren wir*

$$p_0(k) := \sup\{p : \mathbf{P}_p(X \geq k) \leq \alpha/2\},$$
$$p_1(k) := \inf\{p : \mathbf{P}_p(X \leq k) \leq \alpha/2\}.$$

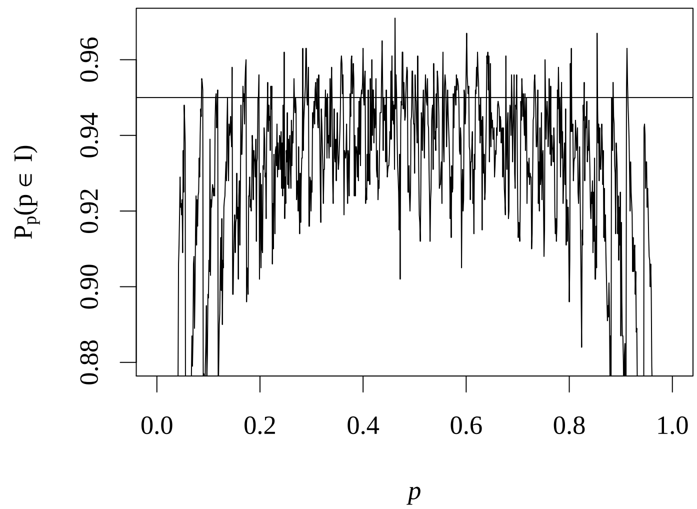

Abbildung 3.1: Simulationsstudie zum approximativen Konfidenzintervall aus Beispiel 3.6.

mit $\sup \emptyset = 0, \inf \emptyset = 1$. *Dann ist*

$$K(X) := [p_0(X), p_1(X)]$$

ein Konfidenzintervall für p *zum Niveau* $1 - \alpha$.

Beispiel 3.6 (p-Münzwurf). Wir haben in Abschnitt 1 Punktschätzer für die Erfolgswahrscheinlichkeit beim p-Münzwurf kennen gelernt. Der erfolgversprechendste Schätzer aus Beispiel 1.2 war $\widehat{p}$. Obige Proposition gibt nun ein exaktes Konfidenzintervall für p. Im Beispiel war $X = 23$. Für $\alpha = 5\,\%$ finden wir

$$\begin{aligned}\mathbf{P}_{0.298}(X \geq 23) &\approx 0.0246,\\ \mathbf{P}_{0.578}(X \leq 23) &\approx 0.0243,\end{aligned}$$

also ist $(0.298, 0.578)$ das Konfidenzintervall aus der Proposition zum Niveau $95\,\%$ für $X = 23$.

Weiter haben wir bereits in Beispiel 0.1 gesehen, dass approximativ

$$I := \left[\widehat{p} - 1.96\sqrt{\widehat{p}(1-\widehat{p})/n}; \widehat{p} + 1.96\sqrt{\widehat{p}(1-\widehat{p})/n}\right]$$

ein Konfidenzintervall für p ist. (Man beachte, dass hier $\widehat{p}$ eine Zufallsvariable ist und dieses Intervall damit auch zufällig ist.) Im Beispiel war $I = [0.30, 0.57]$, was nahe am exakten Intervall liegt. Um zu überprüfen, wie gut diese Approximation ist, simulieren wir für jeden Wert von $p = 0.001, \ldots, 0.999$ jeweils 1000 p-Münzwürfe der Länge $n = 53$. In Abbildung 3.1 tragen wir p gegen den Anteil der Münzwürfe, in denen $p \in I$ gilt, auf. Wie man sieht, liegt das tatsächliche Konfidenzniveau von I gerade an an den Rändern unterhalb von $1 - \alpha$.

Beweis von Proposition 3.5. Zunächst ist sicher $p_0(k) \leq p_1(k)$. Es gilt

$$\begin{aligned}p \leq p_0(k) &\iff \mathbf{P}_p(X \geq k) \leq \alpha/2\\ p_1(k) \leq p &\iff \mathbf{P}_p(X \leq k) \leq \alpha/2\end{aligned}$$

und damit

$$\mathbf{P}_p(p \leq p_0(X)) \leq \alpha/2,$$
$$\mathbf{P}_p(p_1(X) \leq p) \leq \alpha/2.$$

Daraus folgt

$$\mathbf{P}_p(p_0(X) \leq p \leq p_1(X)) = 1 - \mathbf{P}_p(p < p_0(X)) - \mathbf{P}_p(p_1(X) \leq p) \geq 1 - \alpha.$$

Damit ist also $(p_0(X), p_1(X))$ ein Konfidenzintervall für p zum Konfidenzniveau $1 - \alpha$. □

Wir kommen nun in Proposition 3.8 zu einem Konfidenzintervall für den Median einer Verteilung. Dieses erhält man mit Hilfe von Ordnungsstatistiken, die wir zunächst einführen.

Definition 3.7 (Ordnungsstatistiken). *Sei $x_1, \dots, x_n \in \mathbb{R}$. Dann bezeichnen wir die kleinste der Zahlen mit $x_{(1)}$, die zweitkleinste mit $x_{(2)}, \dots$, die größte mit $x_{(n)}$.*

Proposition 3.8 (Konfidenzintervall für den Median).
Sei $\alpha \in (0,1)$, $(X = (X_1, \dots, X_n), (\mathbf{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell mit reellwertigen $X_1, \dots, X_n$, die unter allen $\mathbf{P}_\vartheta$ unabhängig sind. Falls

$$m : \vartheta \mapsto \text{ Median von } X \text{ unter } \mathbf{P}_\vartheta,$$

und $q_{\alpha/2} \in \{0, \dots, n\}$ ein $\alpha/2$-Quantil von $B(n, 1/2)$, dann ist

$$[X_{(q_{\alpha/2})}, X_{(n - q_{\alpha/2} + 1)}]$$

ein Konfidenzintervall von $m(\vartheta)$ zum Niveau $1 - \alpha$.

Bemerkung 3.9 (Verteilungsfreie Verfahren). Man beachte, dass in der Proposition keinerlei Bedingungen an die Verteilung von X_i gestellt wurde. Solche Verfahren bezeichnet man auch als verteilungsfrei.

Beweis von Proposition 3.8. Sei Y eine nach $B(n, 1/2)$ verteilte Zufallsvariable. Dann gilt

$$2\mathbf{P}(Y < q_{\alpha/2}) \leq \alpha. \tag{3.1}$$

Es ist

$$\mathbf{P}(X_1 < m(\vartheta)) \leq 1/2$$
$$\mathbf{P}(X_1 > m(\vartheta)) = 1 - \mathbf{P}(X_1 \leq m(\vartheta)) \leq 1/2.$$

nach Definition des Medians. Weiter gilt

$$\begin{aligned}
\mathbf{P}_\vartheta(m(\vartheta) < X_{(q_{\alpha/2})}) &= \mathbf{P}_\vartheta(\text{weniger als } q_{\alpha/2} \text{ der } X_i\text{'s kleiner als } m(\vartheta)) \\
&\leq \mathbf{P}(Y < q_{\alpha/2}), \\
\mathbf{P}_\vartheta(m(\vartheta) > X_{(n - q_{\alpha/2} + 1)}) &= \mathbf{P}_\vartheta(\text{weniger als } q_{\alpha/2} \text{ der } X_i\text{'s größer als } m(\vartheta)) \\
&\leq \mathbf{P}(Y < q_{\alpha/2})
\end{aligned}$$

und mit (3.1) folgt die Behauptung. □

3.3 Intervallschätzer bei normalverteilten Zufallsgrößen

Wir kommen nun zu dem wichtigen Fall der Intervallschätzung des Erwartungswertes und der Varianz bei normalverteilten Zufallsvariablen.

Theorem 3.10 (Schätzung des Erwartungswertes und der Varianz bei normalverteilten Zufallsvariablen).
Sei $n \geq 2$ und $(X = (X_1, \dots, X_n), (\mathbf{P}_{\mu,\sigma^2})_{\mu\in\mathbb{R},\sigma^2\in\mathbb{R}_+})$ ein statistisches Modell, so dass $X_1, \dots, X_n$ unter $\mathbf{P}_{\mu,\sigma^2}$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt sind. Weiter seien $t_{n,a}$ und $\chi^2_{n,a}$ das a-Quantil der t- und der χ^2-Verteilung mit n Freiheitsgraden. Weiter sei $\alpha \in (0, 1)$.

1. *Es ist*
$$\Big[\overline{X} - \sqrt{s^2(X)/n} \cdot t_{n-1,1-\alpha/2}, \overline{X} + \sqrt{s^2(X)/n} \cdot t_{n-1,1-\alpha/2}, \Big]$$
ein Konfidenzintervall für μ zum Niveau $1 - \alpha$.

2. *Es ist*
$$\Big[\frac{n-1}{\chi^2_{n-1,1-\alpha/2}} s^2(X), \frac{n-1}{\chi^2_{n-1,\alpha/2}} s^2(X)\Big]$$
ein Konfidenzintervall für σ^2 zum Niveau $1 - \alpha$.

Bemerkung 3.11 (Vorwissen über die Verteilung von X). Man beachte, dass das Konfidenzintervall für μ keinerlei Angabe darüber benötigt, wie groß die Varianz der zu Grunde liegenden Normalverteilung ist. Allerdings muss auch gesagt werden, dass das Theorem voraussetzt, dass alle X_i's dieselbe Varianz besitzen.

Beispiel 3.12 (Normalverteilte Schlafdauern). Wir untersuchen ein Beispiel aus der Medizin. Ein Medikament wir daraufhin untersucht, ob es den Schlaf von Probanden verlängert. Dazu wird jeweils die Schlafdauerdifferenz bei zehn Patienten notiert. Man erhält

$$1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4.$$

Diese Beobachtungen betrachten wir als Realisierungen von Zufallsvariablen $X_1, \dots, X_{10}$, die nach $N(\mu, \sigma^2)$ verteilt sind, wobei weder μ noch σ^2 bekannt sind. Wir berechnen

$$\overline{X} = 2.33, \qquad s^2(X) = 4.01.$$

Damit ergibt sich ein Konfidenzintervall

$$[0.897, 3.762] \text{ für } \mu$$

und

$$[1.897, 13.361] \text{ für } \sigma^2,$$

jeweils zum Niveau 95 %.

Beweis von Theorem 3.10. In der Situation des Theorems besagt der Satz von Fisher, Theorem 2.8, dass (für jedes der $\mathbf{P}_{\mu,\sigma^2}$)

$$T := \frac{\overline{X} - \mu}{\sqrt{s^2(X)/n}}$$

nach $t(n-1)$ verteilt ist. Für 1. folgt daraus sofort, da die t-Verteilung eine positive Dichte hat und damit die Quantile eindeutig sind,

$$\begin{aligned}\mathbf{P}_{\mu,\sigma^2}\big(\overline{X} - \sqrt{s^2(X)/n}\cdot t_{n-1,1-\alpha/2} \le \mu \le \overline{X} + \sqrt{s^2(X)/n}\cdot t_{n-1,1-\alpha/2}\big)\\ &= \mathbf{P}_{\mu,\sigma^2}\big(T - t_{n-1,1-\alpha/2} \le 0 \le T + t_{n-1,1-\alpha/2}\big)\\ &= \mathbf{P}_{\mu,\sigma^2}\big(T \le t_{n-1,1-\alpha/2}\big) - \mathbf{P}_{\mu,\sigma^2}\big(T \le -t_{n-1,1-\alpha/2}\big)\\ &= 1-\alpha/2-\alpha/2 = 1-\alpha,\end{aligned}$$

da die t-Verteilung symmetrisch um 0 ist. Für 2. ist (wieder wegen dem Satz von Fisher)

$$\chi^2 := (n-1)s^2(X)/\sigma^2$$

nach $\chi^2(n-1)$ verteilt. Damit gilt

$$\begin{aligned}\mathbf{P}_{\mu,\sigma^2}\Big(\frac{n-1}{\chi^2_{n-1,1-\alpha/2}}s^2(X) \le \sigma^2 \le \frac{n-1}{\chi^2_{n-1,\alpha/2}}s^2(X)\Big)\\ &= \mathbf{P}_{\mu,\sigma^2}\Big(\frac{\chi^2}{\chi^2_{n-1,1-\alpha/2}} \le 1 \le \frac{\chi^2}{\chi^2_{n-1,\alpha/2}}\Big)\\ &= \mathbf{P}_{\mu,\sigma^2}\big(\chi^2 \le \chi^2_{n-1,1-\alpha/2}\big) - \mathbf{P}\big(\chi^2 \le \chi^2_{n-1,\alpha/2}\big)\\ &= 1-\alpha,\end{aligned}$$

und das Theorem ist gezeigt. □

4 Tests

Neben Schätzproblemen sind Testprobleme das wichtigste Thema der induktiven Statistik. In der empirischen Forschung ist es oftmals so, dass aufgrund eines ersten Verständnisses eine Hypothese über die erworbenen Daten aufgestellt werden kann. Das Überprüfen solcher Hypothesen erfolgt dann mittels statistischer Tests. Wie üblich geht man davon aus, dass die Daten die Realisierung einer Zufallsvariable sind.

4.1 Definition und einführende Beispiele

Wir beginnen mit der Einführung wichtiger Begriffe wie Teststatistik, Nullhypothese, Alternative, Ablehnungsbereich, Signifikanzniveau und p-Wert.

Definition 4.1 (Statistischer Test). *Sei $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell, S der Zielbereich von X, und $\Theta_0, \Theta_A \subseteq \Theta$ disjunkt mit $\Theta_0 \cup \Theta_A = \Theta$.*

1. *Ein Paar (Y, C) mit $Y = t(X)$ für $t : S \to S'$ und $C \subseteq S'$ heißt* statistischer Test *von*

$$H_0 : \vartheta \in \Theta_0 \text{ gegen } H_A : \vartheta \in \Theta_A.$$

Hier heißt Y Teststatistik, *C* kritischer *oder* Ablehnungsbereich *des Tests, H_0 heißt* Nullhypothese *und H_A heißt* Alternativhypothese. *Man sagt, der Test (T, C) hat* (Signifikanz-)Niveau *$\alpha \in [0, 1]$, falls*

$$\sup_{\vartheta\in\Theta_0} \mathbf{P}_\vartheta(Y \in C) \le \alpha. \tag{4.1}$$

Falls $Y \in C$, sagt man, dass H_0 abgelehnt (und damit H_A angenommen) ist. Falls $Y \notin C$, sagt man, dass H_0 nicht abgelehnt ist (und H_A abgelehnt ist).

Sei nun (Y, C) ein Test der Nullhypothese H_0 gegen H_A.

2. *Die Hypothese $H : \vartheta \in \Theta_H$ (die entweder H_0 oder H_A sein kann) heißt* einfach *wenn $\Theta_H = \{\vartheta^*\}$ für ein $\vartheta^* \in \Theta$. Andernfalls heißt H* zusammengesetzt.

3. *Ist $\Theta = (\underline{\vartheta}, \overline{\vartheta})$ ein Intervall (wobei $\underline{\vartheta} = -\infty$ und $\overline{\vartheta} = \infty$ zugelassen sind und die Intervalle auch abgeschlossen sein können), so heißt der Test (Y, C)* einseitig, *falls $\Theta_H = (\underline{\vartheta}, \vartheta^*)$ oder $\Theta_H = (\vartheta^*, \overline{\vartheta})$. Falls $\Theta_H = (\vartheta^+, \vartheta^*)$ mit $\underline{\vartheta} < \vartheta^+ \leq \vartheta^* < \overline{\vartheta}$, so heißt der Test (Y, C)* zweiseitig.

4. *Gilt*

$$\sup_{\vartheta \in \Theta_0} \mathbf{P}_\vartheta(Y \in C) < \alpha,$$

so heißt der Test (Y, C) konservativ *zum Niveau α.*

5. *Der Test (Y, C) heißt* unverfälscht, *falls*

$$\mathbf{P}_{\vartheta_0}(Y \in C) \leq \mathbf{P}_{\vartheta_A}(Y \in C)$$

für alle $\vartheta_0 \in \Theta_0, \vartheta_A \in \Theta_A$ gilt.

Bemerkung 4.2 (Interpretation und Fehler eines Tests).

1. Einen statistischen Test hat man sich am besten so vorzustellen (siehe auch das nächste Beispiel): die Daten sind gegeben durch die Zufallsvariable X. Diese Daten fasst man durch die meist reellwertige Funktion t zusammen zur Teststatistik $Y = t(X)$. Die Daten können entweder nach $\mathbf{P}_\vartheta$ mit $\vartheta \in \Theta_0$ (d.h. die Nullhypothese ist richtig) oder mit $\vartheta \in \Theta_A$ (d.h. die Alternativhypothese ist richtig) verteilt sein. Ziel ist es, die Nullhypothese genau dann (anhand der Daten X) abzulehnen, wenn H_A richtig ist. Der Ablehnungsbereich C ist so gewählt, dass H_0 genau dann abgelehnt wird, wenn $Y \in C$. Dabei können zwei verschiedene Arten von Fehler auftreten; siehe auch Tabelle 4.1.

	H_0 abgelehnt	H_0 nicht abgelehnt
H_0 richtig	Fehler erster Art	richtige Entscheidung
H_0 falsch	richtige Entscheidung	Fehler zweiter Art

Tabelle 4.1: Die möglichen Fehler eines statistischen Tests.

Gehen wir zunächst davon aus, dass $\vartheta \in \Theta_0$. Hat der Test ein Niveau α, so wissen wir, dass $\mathbf{P}_\vartheta(Y \in C) \leq \alpha$. Da H_0 genau dann abgelehnt wird, wenn $Y \in C$, wissen wir also, dass die Nullhypothese höchstens mit Wahrscheinlichkeit α abgelehnt wird, wenn sie zutrifft. Damit hat man also die Wahrscheinlichkeit, die Nullhypothese abzulehnen, falls sie zutrifft, durch α beschränkt. Falls $Y \in C$, aber $\vartheta \in \Theta_0$, die Nullyhpothese also irrtümlicherweise verworfen wird, sprechen wir über einen *Fehler erster Art* (dessen Wahrscheinlichkeit durch α kontrolliert wird).

Geht man davon aus, dass $\vartheta \in \Theta_A$, liegt eine Fehlentscheidung genau dann vor, wenn $Y \notin C$, die Nullhypothese also nicht abgelehnt wird. In diesem Fall sprechen wir von einem *Fehler zweiter Art.* Das Niveau des Tests liefert keinen Anhaltspunkt dafür, mit welcher Wahrscheinlichkeit ein solcher Fehler auftritt.

2. Entscheidend bei der Feststellung des Niveaus des Tests ist, dass die Verteilung von Y für jedes $(\mathbf{P}_\vartheta)_{\vartheta\in\Theta_0}$ bekannt sein muss. (Andernfalls kann man ja die linke Seite aus (4.1) nicht berechnen.) Aus diesem Grund haben wir uns in Abschnitt 2 mit der Verteilung von Größen beschäftigt, die von Normalverteilungen abgeleitet waren.

3. Auf den ersten Blick besteht eine scheinbare Symmetrie zwischen H_0 und H_A. Schließlich lehnen wir H_0 genau dann ab (und nehmen H_A an), wenn $Y \in C$ und wir lehnen H_0 nicht ab (und lehnen damit H_A ab) wenn $Y \notin C$. Allerdings wird diese Symmetrie durch das Niveau des Tests gebrochen. Weiß man, dass (Y, C) ein Test zum Niveau α ist, bedeutet das, dass die Wahrscheinlichkeit, die Nullhypothese H_0 abzulehnen, obwohl sie wahr ist, höchstens α ist. Mit anderen Worten ist die Wahrscheinlichkeit für einen Fehler erster Art höchstens α. Allerdings hat man keine Kontrolle über den Fehler zweiter Art.

   Wegen dieser Asymmetrie ist in der Praxis die Nullhypothese genau so zu wählen, dass eine Ablehnung der Nullhypothese möglichst sicher auf die Richtigkeit der Alternativhypothese zurückzuführen ist. Wir betrachten das Beispiel der Schlafdauern aus Beispiel 3.12. Bevor wir die Daten des Experimentes erhoben haben, haben wir die Vorstellung, dass das Medikament die Schlafdauer verändert. Außerdem legen wir ein Signifikanzniveau α fest (was in der Praxis oft $\alpha = 5\,\%$ ist). Um unsere Vorstellung über die Schlafdauer zu überprüfen, testen wir

$$\begin{gathered} H_0 : \text{die Verteilung der Schlafdauer ändert sich nicht} \\ \text{gegen} \\ H_A : \text{die Verteilung der Schlafdauer ändert sich.} \end{gathered}$$

   Kommt es nämlich jetzt zu einer Ablehnung von H_0, so wissen wir, dass dies mit Wahrscheinlichkeit höchstens α dann passiert, wenn H_0 wahr ist, es also gar keine Veränderung der Schlafdauer gibt. Damit können wir uns relativ sicher sein, dass die Ablehnung der Nullhypothese darauf zurückzuführen ist, dass H_A zutrifft. Damit ist unsere Vorstellung, dass das Medikament die Schlafdauer verändert, höchstwahrscheinlich bestätigt.

4. Ein möglichst großer Ablehnungsbereich bei möglichst kleinem Niveau α ist für jeden Test wünschenswert. Schließlich soll die Nullhypothese in möglichst vielen Fällen abgelehnt werden, ohne dass die Wahrscheinlichkeit, sie irrtümlicherweise abzulehnen, größer als α wird. Das bedeutet, dass man meist versucht, keine konservativen Tests zu konstruieren. Allerdings ist dies nicht immer möglich. Etwa kann eine Vergrößerung des Ablehnungsbereichs um einen einzigen Wert dazu führen, dass der Test das Niveau α nicht mehr einhält (wie es zum Beispiel beim Binomialtest, Proposition 4.4, passiert). Dann kann der Ablehnungsbereich nicht vergrößert werden und der Test ist unter Umständen konservativ.

5. Die Forderung von unverfälschten Tests ist klar zu verstehen: Da wir H_0 dann ablehnen, wenn $Y \in C$, soll zumindest die Wahrscheinlichkeit, dass H_0 abgelehnt wird, unter $\mathbf{P}_{\vartheta_A}$, $\vartheta_A \in \Theta_A$ größer sein als für $\mathbf{P}_{\vartheta_0}$, $\vartheta_0 \in \Theta_0$.

Bemerkung 4.3 (p-Werte und alternative Definition eines Tests).

1. Sei $Y = y$, d.h. dass die Teststatistik Y, angewendet auf die echten Daten, ergibt y. Dann heißt der Wert
$$p_y := \sup_{\vartheta\in\Theta_0} \mathbf{P}_\vartheta(Y \text{ extremer als } y)$$

p-Wert des Tests für $Y = y$. Dabei hängt die Bedeutung davon, was 'extremer' heißt davon ab, was genau die Alternative ist. (Dies ist oftmals in konkreten Beispielen einfach zu verstehen, siehe etwa den Binomialtest, Abbildung 4.1 und den Gauss-Test, Abbildung 4.6.) Immer gilt jedoch $p_y \leq p_{y'}$, falls y extremer als y' ist. Es ist wichtig zu beachten, dass es dadurch einen engen Zusammenhang zwischen dem Niveau α des Tests und dem p-Wert gibt. Ist nämlich (Y, C) ein Test zum Niveau α und

$$C = \{y : y \text{ extremer als } y_0\}$$

für ein y_0, so wird H_0 genau dann abgelehnt, wenn

$$\alpha \geq \sup_{\vartheta \in \Theta_0} \mathbf{P}_\vartheta(Y \text{ extremer als } y_0) = p_{y_0}.$$

Ist $Y = y$ und gilt $p_y \leq p_{y_0}$, so wird H_0 also abgelehnt. Es genügt also, für einen Test zum Niveau α und $Y = y$ den Wert p_y zu bestimmen. Ist $p_y \leq \alpha$, so wird H_0 abgelehnt. Dieses Vorgehen wird bei vielen Statistik-Programmen angewendet, bei denen ausschließlich p-Werte ausgegeben werden. Dabei muss man meist angeben, was genau die Alternative ist (einseitig oder zweiseitig), damit das Programm weiß, in welche Richtungen Abweichungen als extrem zu betrachten sind.

2. Oft definiert man die Teststatistik als Abbildung $t : S \to \{0, 1\}$, also $S' = \{0, 1\}$ und den Ablehnungsbereich $C = \{1\}$. Dies ist keine Einschränkung gegenüber obiger Definition. Sei (Y, C) ein Test zum Niveau α nach Definition 4.2 und $\varphi = 1_C : S' \to \{0, 1\}$. Dann ist $(\varphi(Y), \{1\})$ ebenfalls ein Test zum Niveau α, denn es gilt

$$\sup_{\vartheta \in \Theta_0} \mathbf{P}_\vartheta(\varphi(Y) = 1) = \sup_{\vartheta \in \Theta_0} \mathbf{P}_\vartheta(Y \in C) \leq \alpha.$$

Wir kommen nun zunächst zu einem ersten konkreten Beispiel für einen Test. Diesen haben wir auch schon in unserem Eingangsbeispiel 0.1 kennen gelernt.

Proposition 4.4 (Binomialtest). *Sei $\alpha \in [0, 1], n \in \mathbb{N}$ und $(X, (\mathbf{P}_p)_{p \in [0,1]})$ ein statistisches Modell, so dass X unter $\mathbf{P}_p$ nach $B(n, p)$ verteilt ist.*

(a) Ist $\Theta_0 = p^*$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(X, \{0, \dots, k\} \cup \{l, \dots, n\})$ ein unverfälschter Test zum Niveau α, falls

$$\mathbf{P}_{p^*}(X \leq k) \leq \alpha/2, \qquad \mathbf{P}_{p^*}(X \geq l) \leq \alpha/2.$$

(b) Ist $\Theta_0 = [0, p^*]$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(X, \{k, \dots, n\})$ ein unverfälschter Test zum Niveau α, falls

$$\mathbf{P}_{p^*}(X \geq k) \leq \alpha.$$

(c) Ist $\Theta_0 = [p^*, 1]$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(X, \{0, \dots, k\})$ ein unverfälschter Test zum Niveau α, falls

$$\mathbf{P}_{p^*}(X \leq k) \leq \alpha.$$

Der Binomialtest

überprüft, ob bestimmte Erfolgswahrscheinlichkeiten einer Binomialverteilung angenommen werden.

Statistisches Modell	X unter $\mathbf{P}_p$ nach $B(n,p)$ verteilt
Hypothesen	(a) $H_0 : p \in \{p^*\}$ gegen $H_A : p \notin \{p^*\}$ (b) $H_0 : p \in [0, p^*]$ gegen $H_A : p \in (p^*, 1]$ (c) $H_0 : p \in [p^*, 1]$ gegen $H_A : p \in [0, p^*)$
Teststatistik	X unter $\mathbf{P}_p$ verteilt nach $B(n,p)$
Ablehnungsbereich	(a) $\{0,\dots,k,l,\dots,n\}$ mit $\mathbf{P}_{p^*}(X \le k), \mathbf{P}_{p^*}(X \ge l) \le \alpha/2$, (b) $\{l,\dots,n\}$ mit $\mathbf{P}_{p^*}(X \ge l) \le \alpha$ (c) $\{0,\dots,k\}$ mit $\mathbf{P}_{p^*}(X \le k) \le \alpha$,
p-Wert, falls $X = x$	(a) $\mathbf{P}_{p^*}(X \le x \wedge (2p^*n - x)) + \mathbf{P}_{p^*}(X \ge x \vee (2p^*n - x))$ (b) $\mathbf{P}_{p^*}(X \ge x)$ (c) $\mathbf{P}_{p^*}(X \le x)$

Abbildung 4.1: Schematische Darstellung des Binomialtests aus Proposition 4.4

Beweis. Wir beweisen nur (c), die anderen beiden Aussagen folgen analog. Klar ist, dass der Test unverfälscht ist. Es ist außerdem

$$\sup_{p\in\Theta_0} \mathbf{P}_p(X \in \{0,\dots,k\}) = \mathbf{P}_{p^*}(X \le k) \le \alpha$$

nach Voraussetzung. Also folgt bereits die Aussage. □

Beispiel 4.5 (Binomialtest). Sei $\alpha = 5\,\%$, $n = 53$ und $(X, (\mathbf{P}_p)_{p\in[0,1]})$ wie in der Proposition. Wir wollen nun

$$H_0 : p = 1/2 \text{ gegen } H_A : p \neq 1/2$$

testen, wenn wir in 53 Versuchen 23 Erfolge erzielt haben. Nach Proposition 4.4 ist der kritische Bereich von der Form $\{0,\dots,k\} \cup \{l,\dots,53\}$. Es ist

$$\mathbf{P}_{p=1/2}(X \le 18) + \mathbf{P}_{p=1/2}(X \ge 35) \approx 2.70\,\%.$$

Da $18 < 23 < 35$, liegt 23 nicht im Ablehnungsbereich von H_0. Damit kann die Nullhypothese aufgrund der Daten ($X = 23$) nicht abgelehnt werden. Auf dasselbe Ergebnis kommt man mit Hilfe des p-Wertes. Es ist

$$\mathbf{P}_{p=1/2}(X \text{ extremer als } 23) = \mathbf{P}_{p=1/2}(X \le 23) + \mathbf{P}_{p=1/2}(X \ge 30) \approx 41.01\,\%.$$

Da dieser Wert größer als $\alpha = 5\,\%$ ist, kann man die Nullhypothese nicht ablehnen.

Proposition 4.6 (Gauss-Test). *Sei* $\alpha \in [0,1], \sigma^2 \in \mathbb{R}_+, \mu^* \in \mathbb{R}$ *und* $(X = (X_1, \dots, X_n), (\mathbf{P}_\mu)_{\mu\in\mathbb{R}})$ *ein statistisches Modell, so dass* $X_1, \dots, X_n$ *unter* $\mathbf{P}_\mu$ *unabhängig und nach* $N(\mu, \sigma^2)$ *verteilt sind. Weiter sei*

$$Z := \frac{\overline{X} - \mu^*}{\sqrt{\sigma^2/n}}.$$

und z_p *für* $p \in [0,1]$ *das* p*-Quantil von* $N(0,1)$*.*

(a) *Ist* $\Theta_0 = \{\mu^*\}$, $\Theta_A = \Theta \setminus \Theta_0$*, so ist* $(Z, (-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty))$ *ein unverfälschter Test zum Niveau* α*.*

(b) *Ist* $\Theta_0 = (-\infty, \mu^*]$, $\Theta_A = \Theta \setminus \Theta_0$*, so ist* $(Z, [z_{1-\alpha}, \infty))$ *ein unverfälschter Test zum Niveau* α*.*

(c) *Ist* $\Theta_0 = [\mu^*, \infty)$, $\Theta_A = \Theta \setminus \Theta_0$*, so ist* $(Z, (-\infty, z_\alpha])$ *ein unverfälschter Test zum Niveau* α*.*

Beweis. Wieder beweisen wir nur (c). Es ist klar, dass der Test unverfälscht ist. Wir wissen, dass unter $\mathbf{P}_{\mu^*}$ die Zufallsvariable Z nach $N(0,1)$ verteilt ist. Damit gilt

$$\sup_{\mu\geq\mu^*} \mathbf{P}_\mu(Z \leq z_\alpha) = \mathbf{P}_{\mu^*}(Z \leq z_\alpha) = \alpha,$$

woraus die Behauptung sofort folgt. □

Definition 4.7 (Gütefunktion, Macht eines Tests). *Sei* $(X, (\mathbf{P}_\vartheta)_{\vartheta\in\Theta})$ *ein statistisches Modell und* (Y, C) *ein statistischer Test von* $H_0 : \vartheta \in \Theta_0$ *gegen* $H_A : \vartheta \in \Theta_A$*.*

1. *Die Funktion*

$$g_Y : \begin{cases} \Theta & \to [0,1] \\ \vartheta & \mapsto \mathbf{P}_\vartheta(Y \in C) \end{cases}$$

heißt Gütefunktion *von* (Y, C)*. Für* $\vartheta \in \Theta_A$ *heißt* $g_Y(\vartheta)$ *auch die* Macht *des Tests* (Y, C) *in* ϑ*.*

2. *Ist* $\alpha \in [0,1]$ *und* (Y', C') *ein weiterer Test von* H_0 *gegen* H_A*. Sind sowohl* (Y, C) *als auch* (Y', C') *Tests zum Niveau* α*, so heißt* (Y, C) gleichmäßig besser *als* (Y', C')*, falls*

$$g_Y(\vartheta) \geq g_{Y'}(\vartheta)$$

für alle $\vartheta \in \Theta_A$ *gilt.*

Bemerkung 4.8 (Idealer Test). Ein idealer Test (Y, T) (den es in realen Situationen nie gibt) hätte die Eigenschaft, dass $Y \in C$ nur für $\vartheta \in \Theta_A$ möglich ist und weiter, dass $\mathbf{P}_\vartheta(Y \in C) = 1$ für $\vartheta \in \Theta_A$. Das bedeutet, dass $g_Y(\vartheta) = 1_{\vartheta\in\Theta_A}$ die Gütefunktion eines idealen Tests ist. Für einen solchen Test wäre die Macht für alle $\vartheta \in \Theta_A$ gleich 1.

Der Gauss-Test

überprüft, ob der Erwartungswert einer Normalverteilung (oder irgendeiner Verteilung bei approximativ unendlich großer Samplegröße) gleich einer vorgegebenen Größe μ^* ist, wenn die Varianz bekannt ist.

Statistisches Modell	$X = (X_1, \ldots, X_n)$ und $X_1, \ldots, X_n$ unter $\mathbf{P}_\mu$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt
Hypothesen	(a) $H_0 : \mu \in \{\mu^*\}$ gegen $H_A : \mu \notin \{\mu^*\}$ (b) $H_0 : \mu \in (-\infty, \mu^*]$ gegen $H_A : \mu \in (\mu^*, \infty)$ (c) $H_0 : \mu \in [\mu^*, \infty)$ gegen $H_A : \mu \in (-\infty, \mu^*)$
Teststatistik	$Z = \dfrac{\overline{X} - \mu^*}{\sqrt{\sigma^2/n}}$ unter $\mathbf{P}_{\mu^*}$ verteilt nach $N(0,1)$
Ablehnungsbereich	(a) $(-\infty, z_{\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$, (b) $(z_{1-\alpha}, \infty)$ (c) $(-\infty, z_\alpha)$
p-Wert, falls $Z = z$	(a) $2(1 - \mathbf{P}(Z \leq \lvert z \rvert))$, wenn Z nach $N(0,1)$ verteilt ist (b) $\mathbf{P}(Z \geq z))$ (c) $\mathbf{P}(Z \leq z))$

Abbildung 4.2: Schematische Darstellung des Gauss-Tests aus Proposition 4.6

Beispiel 4.9 (Gauss-Test). Betrachten wir die Situation des Gauss-Tests aus Proposition 4.6 (bzw. Abbildung 4.2). Sei $\widetilde{Z}$ eine (unter allen $\mathbf{P}_\mu$) nach $N(0,1)$ verteilte Zufallsvariable. Hier gilt im Fall des einseitigen Tests (c)

$$\begin{aligned} g_Z(\mu) &= \mathbf{P}_\mu(Z \leq z_\alpha) = \mathbf{P}_\mu\Big(\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} + \frac{\mu - \mu^*}{\sqrt{\sigma^2/n}} \leq z_\alpha\Big) = \mathbf{P}\Big(\widetilde{Z} \leq z_\alpha + \frac{\mu^* - \mu}{\sqrt{\sigma^2/n}}\Big) \\ &= \Phi\Big(z_\alpha + \frac{\mu^* - \mu}{\sqrt{\sigma^2/n}}\Big), \end{aligned}$$

wobei $\Phi(.)$ die Verteilungsfunktion von $N(0,1)$ ist. Analog ist für den zweiseitigen Test (a)

$$\begin{aligned} g_Z(\mu) &= \mathbf{P}_\mu(Z \leq z_{\alpha/2}) + \mathbf{P}_\mu(Z \geq z_{1-\alpha/2}) \\ &= \mathbf{P}\Big(\widetilde{Z} \leq z_{\alpha/2} + \frac{\mu^* - \mu}{\sqrt{\sigma^2/n}}\Big) + \mathbf{P}\Big(Z \geq z_{1-\alpha/2} + \frac{\mu^* - \mu}{\sqrt{\sigma^2/n}}\Big) \\ &= 1 - \Phi\Big(z_{1-\alpha/2} + \frac{\mu^* - \mu}{\sqrt{\sigma^2/n}}\Big) + \Phi\Big(z_{\alpha/2} + \frac{\mu^* - \mu}{\sqrt{\sigma^2/n}}\Big). \end{aligned}$$

Diese zwei Gütefunktionen sind in Abbildung 4.3 dargestellt.

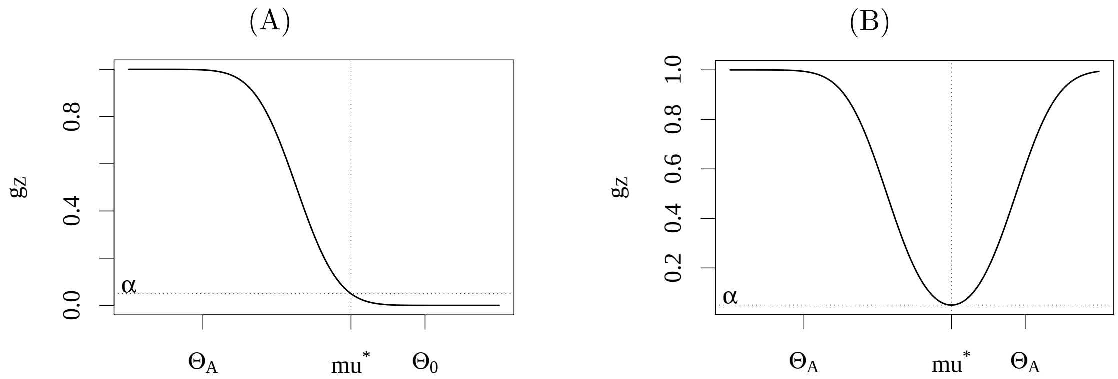

Abbildung 4.3: Die Gütefunktionen für den einseitigen (A) und den zweiseitigen (B) Gauss-Test aus Beispiel 4.9.

4.2 Parametertests

Der eben besprochene Gauss-Test fällt bereits in die Klasse der Parametertests. Unter diesem Stichwort versteht man statistische Tests, die eine Stichprobe daraufhin testen, ob Parameter der zugrunde liegenden Verteilung gewisse Werte annehmen. Auch in diesem Abschnitt beschäftigt uns der Fall von unabhängigen, normalverteilten Stichproben.

Wir werden Tests für folgende Situationen kennenlernen: Test, ob der Erwartungswert einer normalverteilten Stichprobe einen bestimmten Wert annimmt (einfacher t-Test, Proposition 4.10 und Abbildung 4.4); Test, ob die Erwartungswerte von zwei verbundenen Stichproben identisch sind (gepaarter t-Test, Korollar 4.13 und Abbildung 4.5); Test, ob die Erwartungswerte von zwei unabhängigen Stichproben identisch sind (doppelter t-Test, Proposition 4.15 und Abbildung 4.6).

Proposition 4.10 (Einfacher *t*-Test).
Sei $\alpha \in [0,1]$, $\mu^* \in \mathbb{R}$ und $(X = (X_1, \dots, X_n), (\mathbf{P}_{\mu,\sigma^2})_{\mu\in\mathbb{R},\sigma^2\in\mathbb{R}_+})$ ein statistisches Modell, so dass $X_1, \dots, X_n$ unter $\mathbf{P}_{\mu,\sigma^2}$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt sind. Weiter sei

$$T := \frac{\overline{X} - \mu^*}{\sqrt{s^2(X)/n}}.$$

und $t_{n,p}$ für $p \in [0,1]$ das p-Quantil von $t(n)$.

(a) Ist $\Theta_0 = \{\mu^*\} \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(T, (-\infty, t_{n-1,\alpha/2}) \cup (t_{n-1,1-\alpha/2}, \infty))$ ein unverfälschter Test zum Niveau α.

(b) Ist $\Theta_0 = (-\infty, \mu^*] \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(T, [t_{n-1,1-\alpha}, \infty))$ ein unverfälschter Test zum Niveau α.

(c) Ist $\Theta_0 = [\mu^*, \infty) \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, so ist $(T, (-\infty, t_{n-1,\alpha}])$ ein unverfälschter Test zum Niveau α.

Beweis. Wieder beweisen wir nur (c), da die anderen beiden Aussagen analog folgen. Genau wie im Gauss-Test ist klar, dass der Test unverfälscht ist. Aus dem Satz von Fisher, Theorem 2.8, folgt dass T unter $\mathbf{P}_{\mu^*,\sigma^2}$ nach $t(n-1)$ verteilt ist. Damit gilt

$$\sup_{\mu\geq\mu^*} \mathbf{P}_{\mu,\sigma^2}(T \leq t_{n-1,\alpha}) = \mathbf{P}_{\mu^*,\sigma^2}(T \leq t_{n-1,\alpha}) = \alpha,$$

Der einfache t-Test

überprüft, ob der Erwartungswert einer Normalverteilung (oder irgendeiner Verteilung bei approximativ unendlich großer Samplegröße) gleich einer vorgegebenen Größe μ^* ist, wenn die Varianz unbekannt ist.

Statistisches Modell	$X = (X_1, \dots, X_n)$ und $X_1, \dots, X_n$ unter $\mathbf{P}_{\mu,\sigma^2}$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt
Hypothesen	(a) $H_0 : \mu \in \{\mu^*\}$ gegen $H_A : \mu \notin \{\mu^*\}$ (b) $H_0 : \mu \in (-\infty, \mu^*]$ gegen $H_A : \mu \in (\mu^*, \infty)$ (c) $H_0 : \mu \in [\mu^*, \infty)$ gegen $H_A : \mu \in (-\infty, \mu^*)$
Teststatistik	$T = \dfrac{\overline{X} - \mu^*}{\sqrt{s^2(X)/n}}$ unter $\mathbf{P}_{\mu^*,\sigma^2}$ verteilt nach $t(n-1)$
Ablehnungsbereich	(a) $(-\infty, t_{n-1,\alpha/2}) \cup (t_{n-1,1-\alpha/2}, \infty)$, (b) $(t_{n-1,1-\alpha}, \infty)$ (c) $(-\infty, t_{n-1,\alpha})$
p-Wert, falls $T = t$	(a) $2(1 - \mathbf{P}(T \leq \lvert t \rvert))$, wenn T nach $t(n-1)$ verteilt ist (b) $\mathbf{P}(T \geq t))$ (c) $\mathbf{P}(T \leq t))$

Abbildung 4.4: Schematische Darstellung des einfachen t-Tests aus Proposition 4.10

woraus die Behauptung sofort folgt. $\square$

Bemerkung 4.11 (Vergleich von Gauss-Test und einfachem t-Test). Sowohl der Gauss-Test aus Proposition 4.6 (und Abbildung 4.2), als auch der einfache t-Test basieren auf unabhängigen, normalverteilten Stichproben. Der entscheidende Unterschied der beiden Tests besteht darin, dass beim Gauss-Test die Varianz der zugrunde liegenden Normalverteilung bekannt sein muss, und beim t-Test nicht. Dies sieht man etwa daran, dass beim Gauss-Test die Nullhypothese nur aus einem Bereich für μ, nicht jedoch für σ^2 besteht. Außerdem kann man ja nur bei Kenntnis von σ^2 die Teststatistik Z aus Proposition 4.6 berechnen. Beim t-Test ersetzt man σ^2 durch die empirische Varianz $s^2(X)$, und erhält die Teststatistik T.

Bemerkung 4.12 (Gütefunktion des einfachen t-Tests und Stichprobengröße). Wir betrachten die Situation aus Proposition 4.10(b) mit $\mu^* = 0$ und $\alpha = 5\,\%$. Durch das Signifikanzniveau α wird die Wahrscheinlichkeit für einen Fehler erster Art kontrolliert. Die Wahrscheinlichkeit für einen Fehler zweiter Art wird durch die Gütefunktion angegeben. Ist nämlich $\mu > 0$, so ist

$$\mathbf{P}_{\mu,\sigma^2}(\text{Fehler zweiter Art}) = \mathbf{P}_{\mu,\sigma^2}(T \notin C) = 1 - g_T(\mu).$$

Wir fragen nun, wie gut wir diesen Fehler zweiter Art kontrollieren können, wenn $\mu > 0$ gegeben ist. Die Gütefunktion ist gegeben als

$$\begin{aligned} g_T(\mu) &= \mathbf{P}_{\mu,\sigma^2}(T \in C) = \mathbf{P}_{\mu,\sigma^2}\big(T \geq t_{n-1,1-\alpha}\big) \\ &= \mathbf{P}_{\mu,\sigma^2}\Big(\widetilde{T} + \frac{\mu}{\sqrt{s^2(X)/n}} \geq t_{n-1,1-\alpha}\Big) \end{aligned}$$

für die unter $\mathbf{P}_{\mu,\sigma^2}$ nach $t(n-1)$ verteilte Zufallsgröße

$$\widetilde{T} = \frac{\overline{X} - \mu}{\sqrt{s^2(X)/n}}.$$

Ist n groß, so ist $\widetilde{T}$ etwa nach $N(0,1)$ verteilt und $s^2(X) \approx \sigma^2$. Damit ist für eine nach $N(0,1)$ verteilte Zufallsvariable Z

$$\mathbf{P}_{\mu,\sigma^2}(\text{Fehler zweiter Art}) \approx \mathbf{P}\Big(Z + \frac{\mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha}\Big).$$

Wollen wir etwa erreichen, dass die Wahrscheinlichkeit für einen Fehler zweiter Art nicht größer als 5 % ist, bedeutet das, dass

$$\frac{\mu}{\sqrt{\sigma^2/n}} \geq 3.29 \tag{4.2}$$

gelten muss, denn

$$\mathbf{P}\Big(Z + 3.29 \leq 1.645\Big) = \mathbf{P}\Big(Z \leq -1.645\Big) = 5\,\%.$$

Man beachte, dass (4.2) eine Bedingung für die Stichprobengröße liefert. Will man etwa $\mu = 0.1$ (bei $\sigma^2 = 1$) noch mit einem Fehler zweiter Art von 5 % ablehnen können, so muss

$$\sqrt{n} \geq 3.29/0.1 = 32.9, \qquad n \geq 1083$$

gewählt werden.

Korollar 4.13 (Gepaartert-Test).
Sei $\alpha \in [0,1]$ *und* $((X,Y) = (X_1,\dots,X_n,Y_1,\dots,Y_n)),(\mathbf{P}_{\mu,\sigma^2})_{\mu\in\mathbb{R},\sigma^2\in\mathbb{R}_+})$ *ein statistisches Modell, so dass* $Y - X := (Y_1 - X_1,\dots,Y_n - X_n)$ *unter* $\mathbf{P}_{\mu,\sigma^2}$ *unabhängig und nach* $N(\mu,\sigma^2)$ *verteilt sind. Weiter sei*

$$T := \frac{\overline{Y} - \overline{X}}{\sqrt{s^2(Y-X)/n}}.$$

und $t_{n,p}$ *für* $p \in [0,1]$ *das* p*-Quantil von* $t(n)$*.*

(a) Ist $\Theta_0 = \{0\} \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, *so ist* $(T,(-\infty,t_{n-1,\alpha/2}) \cup (t_{n-1,1-\alpha/2},\infty))$ *ein unverfälschter Test zum Niveau* α*.*

(b) Ist $\Theta_0 = (-\infty,0] \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, *so ist* $(T,[t_{n-1,1-\alpha},\infty))$ *ein unverfälschter Test zum Niveau* α*.*

(c) Ist $\Theta_0 = [\mu^*,\infty) \times \mathbb{R}_+$, $\Theta_A = \Theta \setminus \Theta_0$, *so ist* $(T,(-\infty,t_{n-1,\alpha}])$ *ein unverfälschter Test zum Niveau* α*.*

Der gepaarte t-Test

überprüft, ob die Erwartungswerte einer Normalverteilung (oder irgendeiner Verteilung bei approximativ unendlich großer Samplegröße) zweier verbundener Stichproben gleich ist, wenn deren Varianz unbekannt ist.

Statistisches Modell	$Y - X = (Y_1 - X_1, \dots, Y_n - X_n)$ und $Y_1 - X_1, \dots, Y_n - X_n$ unter $\mathbf{P}_{\mu,\sigma^2}$ unabhängig und nach $N(\mu, \sigma^2)$ verteilt
Hypothesen	(a) $H_0 : \mu = 0 \quad$ gegen $\quad H_A : \mu \neq 0$ (b) $H_0 : \mu \in (-\infty, 0] \quad$ gegen $\quad H_A : \mu \in (0, \infty)$ (c) $H_0 : \mu \in [0, \infty) \quad$ gegen $\quad H_A : \mu \in (-\infty, 0)$
Teststatistik	$T = \dfrac{\overline{Y} - \overline{X}}{\sqrt{s^2(Y - X)/n}}$ unter $\mathbf{P}_{0,\sigma^2}$ verteilt nach $t(n-1)$
Ablehnungsbereich	(a) $(-\infty, t_{n-1,\alpha/2}) \cup (t_{n-1,1-\alpha/2}, \infty)$, (b) $(t_{n-1,1-\alpha}, \infty)$ (c) $(-\infty, t_{n-1,\alpha})$
p-Wert, falls $T = t$	(a) $2(1 - \mathbf{P}(T \leq \lvert t\rvert))$, wenn T nach $t(n-1)$ verteilt ist (b) $\mathbf{P}(T \geq t))$ (c) $\mathbf{P}(T \leq t))$

Abbildung 4.5: Schematische Darstellung des gepaarten t-Tests aus Korollar 4.13

Beweis. Man wendet einfach Proposition 4.10 auf den Vektor $Y - X$ an. □

Beispiel 4.14 (Schlafdauern). Betrachten wir noch einmal die Schlafdauern aus Beispiel 3.12. Hier wurden die Schlafdauern von Personen vor und nach Einnahme des Medikamentes notiert. Seien also X und Y die Schlafdauern vor und nach Einnahme, so ist

$$Y - X = (1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4).$$

Aus Beispiel 3.12 wissen wir bereits, dass $\overline{Y} - \overline{X} = 2.33$ und $s^2(Y - X) = 4.01$. Damit ist $T = 2.33/\sqrt{4.01/10} \approx 3.68$.

Testet man also

$$H_0 : \mathbf{E}_{\mu,\sigma^2}[\overline{Y} - \overline{X}] = 0 \text{ gegen } \mathbf{E}_{\mu,\sigma^2}[\overline{Y} - \overline{X}] \neq 0,$$

auf dem Niveau 5 %, so ist dessen Ablehnungsbereich $C = (-\infty, -2.262) \cup (2.262, \infty)$. Da $3.68 \in C$, kann man H_0 auf dem Niveau 5 % verwerfen. Das bedeutet, dass vermutlich eine Veränderung der Schlafdauer durch Einnahme des Medikamentes stattfand.

Proposition 4.15 (Doppelter t-Test).
Sei $\alpha \in [0,1]$ und $((X,Y) = (X_1,\dots,X_m,Y_1,\dots,Y_n)),(\mathbf{P}_{\mu_X,\mu_Y,,\sigma^2})_{\mu_X,\mu_Y\in\mathbb{R},\sigma^2\in\mathbb{R}_+})$ ein statistisches Modell, so dass $X_1,\dots,X_m,Y_1,\dots,Y_n$ unter $\mathbf{P}_{\mu_X,\mu_Y,\sigma^2}$ unabhängig sind, sowie $X_1,\dots,X_m$ nach $N(\mu_X,\sigma^2)$ und $Y_1,\dots,Y_n$ nach $N(\mu_Y,\sigma^2)$ verteilt sind. Weiter sei

$$T := \frac{\overline{Y}-\overline{X}}{\sqrt{s^2(X,Y)(m+n)/(mn)}}$$

mit

$$s^2(X,Y) := \frac{1}{m+n-2}\Big(\sum_{i=1}^m (X_i-\overline{X})^2 + \sum_{i=1}^n (Y_i-\overline{Y})^2\Big), \tag{4.3}$$

und $t_{n,p}$ für $p\in[0,1]$ das p-Quantil von $t(n)$.

(a) Ist $\Theta_0 = \{(\mu_X,\mu_Y) : \mu_X = \mu_Y\}\times\mathbb{R}_+$, $\Theta_A = \Theta\setminus\Theta_0$, so ist $(T,(-\infty,t_{n-1,\alpha/2})\cup(t_{m+n-2,1-\alpha/2},\infty))$ ein unverfälschter Test zum Niveau α.

(b) Ist $\Theta_0 = \{(\mu_X,\mu_Y) : \mu_Y \le \mu_X\}\times\mathbb{R}_+$, $\Theta_A = \Theta\setminus\Theta_0$, so ist $(T,[t_{m+n-2,1-\alpha},\infty))$ ein unverfälschter Test zum Niveau α.

(c) Ist $\Theta_0 = \{(\mu_X,\mu_Y) : \mu_Y \ge \mu_X\}\times\mathbb{R}_+$, $\Theta_A = \Theta\setminus\Theta_0$, so ist $(T,(-\infty,t_{m+n-2,\alpha}])$ ein unverfälschter Test zum Niveau α.

Beweis. Wir müssen zunächst zeigen, dass T nach $t(m+n-2)$ verteilt ist. Ohne Einschränkung der Allgemeinheit sei $\sigma^2 = 1$. Da $X_1,\dots,X_m,Y_1,\dots,Y_n$ unabhängig sind, sind $s^2(X)$ und $s^2(Y)$ unabhängige Zufallsvariablen. Aus dem Satz von Fisher folgt, dass $\overline{X},\overline{Y},s^2(X),s^2(Y)$ unabhängig sind, $\overline{Y}-\overline{X}$ unter $\mathbf{P}_{0,\sigma^2}$ nach $N(0,\frac{1}{m}+\frac{1}{n})$ und $(m-1)s^2(X)$ nach $\chi^2(m-1)$ und $(n-1)s^2(Y)$ nach $\chi^2(n-1)$ verteilt ist. Damit ist

$$T = \frac{\frac{1}{\sqrt{1/m+1/n}}(\overline{Y}-\overline{X})}{\sqrt{((m-1)s^2(X)+(n-1)s^2(Y))/(m+n-2)}}$$

nach $t(m+n-2)$ verteilt. Der Rest des Beweises folgt wie beim einfachen t-Test, Proposition 4.10. □

Bemerkung 4.16 (Gleichheit der Varianzen). Im doppelten t-Test ist die Annahme der Gleichheit der Varianzen wichtig. Sind die Varianzen nicht gleich, ist nämlich die Teststatistik T unter H_0 nicht $t(m+n-2)$-verteilt. Um die Gleichheit der Varianzen zu überprüfen, gibt es wiederum statistische Tests, etwa den F-Test, auf den wir hier nicht weiter eingehen. Außerdem sei bemerkt, dass es auch eine Variante des doppelten t-Tests für den Fall ungleicher Varianzen gibt. Hierzu wird ein k (in Abhängigkeit von $s^2(X)$ und $s^2(Y)$) ermittelt, so dass approximativ T unter H_0 nach $t(k)$ verteilt ist.

Beispiel 4.17 (Geburtsgewichte). In einer Kölner Klinik wurden im Jahr 1985 $m = 269$ Mädchen und $n = 288$ Jungen geboren. (Die einzelnen Geburtsgewichte sind also $X_1,\dots,X_{269}$ und $Y_1,\dots,Y_{288}$.) Das Durchschnittsgewicht und die empirische Varianz der Mädchen in Gramm war $\bar{X} = 3050$ und $s^2(X) = 211600$, das der Jungen $\bar{Y} = 3300$ und $s^2(Y) = 220900$.

Der doppelte t-Test

überprüft, ob die Erwartungswerte einer Normalverteilung (oder irgendeiner Verteilung bei approximativ unendlich großer Samplegröße) zweier unverbundener Stichproben gleich ist, wenn deren Varianz gleich, aber unbekannt ist.

Statistisches Modell	$X_1, \dots, X_m$ unter unter $\mathbf{P}_{\mu_X,\mu_Y,\sigma^2}$ unabhängig und nach $N(\mu_X, \sigma^2)$ verteilt, $Y_1, \dots, Y_n$ unter unter $\mathbf{P}_{\mu_X,\mu_Y,\sigma^2}$ unabhängig und nach $N(\mu_Y, \sigma^2)$ verteilt
Hypothesen	(a) $H_0 : \mu_X = \mu_Y \quad \text{gegen} \quad H_A : \mu_X \neq \mu_Y$ (b) $H_0 : \mu_Y \leq \mu_X \quad \text{gegen} \quad H_A : \mu_Y > \mu_X$ (c) $H_0 : \mu_Y \geq \mu_X \quad \text{gegen} \quad H_A : \mu_Y < \mu_X$
Teststatistik	$T = \dfrac{\overline{Y} - \overline{X}}{\sqrt{s^2(X,Y)(m+n)/(mn)}}$, $s^2(X,Y)$ wie in (4.3), unter $\mathbf{P}_{\mu,\mu,\sigma^2}$ verteilt nach $t(m+n-2)$
Ablehnungsbereich	(a) $(-\infty, t_{m+n-2,\alpha/2}) \cup (t_{m+n-2,1-\alpha/2}, \infty)$, (b) $(t_{m+n-2,1-\alpha}, \infty)$ (c) $(-\infty, t_{m+n-2,\alpha})$
p-Wert, falls $T = t$	(a) $2(1 - \mathbf{P}(T \leq \lvert t \rvert))$, wenn T nach $t(m+n-2)$ verteilt ist (b) $\mathbf{P}(T \geq t))$ (c) $\mathbf{P}(T \leq t))$

Abbildung 4.6: Schematische Darstellung des doppelten t-Tests aus Proposition 4.15

Es soll zum Signifikanzniveau $\alpha = 0.01$ getestet werden, ob Jungen und Mädchen das gleiche erwartete Geburtsgewicht haben (Fall (a) in Proposition 4.15). Wir berechnen

$$s^2(X,Y) = \frac{1}{269 + 288 - 2}(268 \cdot s^2(X) + 287 \cdot s^2(Y)) = 216409$$

und

$$T = \frac{3300 - 3050}{\sqrt{216409 \cdot (269 + 288)/(269 \cdot 288)}} = 6.338 > 2.585 = t_{555,0.995}.$$

Also können wir die Nullhypothese $\mu_X = \mu_Y$ auf dem Signifikanzniveau $\alpha = 0.01$ ablehnen. Unter der (sehr plausiblen) Annahme der Normalverteilung und der Annahme der Varianzhomogenität kann man zum Niveau $\alpha = 0.01$ schließen, dass Jungen im Mittel schwerer sind als Mädchen.

4.3 Anpassungs- und Unabhängigkeitstests

Die bisher behandelten Tests basierten alle auf normalverteilten Stichproben. Etwa können wir mit den zuletzt behandelten t-Tests überprüfen, ob der Mittelwert einer Stichprobe vermutlich einen bestimmten Wert annimmt. Auffällig ist, dass diese t-Tests immer nur auf den Erwartungswert der zu Grunde liegenden Verteilung testen. Dies ist bei den χ^2-Tests, die wir in diesem Abschnitt kennen lernen, anders. Beim χ^2-Anpassungstest (Theorem 4.19) wird überprüft, ob die gesamte Verteilung einer Stichprobe (und nicht nur der Erwartungswert) eine bestimmte Form hat. Der χ^2-Unabhängigkeitstest (Theorem 4.25) überprüft, ob sich zwei Merkmale unabhängig verhalten.

Beispiel 4.18 (Mendel's Experimente). Der Naturforscher Gregor Mendel kreuzte rot- und weißblühende Erbsenpflanzen. In der ersten Nachkommensgeneration erhielt er ausschließlich rosa-blühende Pflanzen. Kreuzte er diese weiter, erhielt er in der nächsten Generation rot- rosa- und weißblühende Pflanzen. Seine Theorie sagte voraus, dass diese Farben in der zweiten Nachkommensgeneration im Verhältnis 1:2:1 auftreten sollten.

Ziel des χ^2-Anpassungstests in diesem Beispiel ist es, zu überprüfen, ob das Farbverhältnis von 1:2:1 von einer Stichprobe vermutlich eingehalten wird. Betrachten wir dazu $n = 400$ Pflanzen der zweiten Generation, von denen wir annehmen, dass deren Farbe Realisierung von unabhängigen Experimenten ist. Sei $\mathcal{I} = \{\text{rot, rosa, weiß}\}$ und $X = (X_1, \ldots, X_n)$ ein Vektor von unabhängigen, $\mathcal{I}$-wertigen Zufallsvariablen. Falls die Theorie von Mendel stimmt, gilt für alle i

$$\mathbf{P}(X_i = \text{rot}) = \tfrac{1}{4}, \qquad \mathbf{P}(X_i = \text{rosa}) = \tfrac{1}{2}, \qquad \mathbf{P}(X_i = \text{weiß}) = \tfrac{1}{4}.$$

Gleichbedeutend ist es, dass die Verteilungsgewichte der Verteilung von X_i durch den Vektor $\underline{\pi} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$ gegeben sind. Die Stichprobe liefert $S_{\text{weiß}} = 115$ weiße Blüten, $S_{\text{rosa}} = 171$ rosa Blüten und $S_{\text{rot}} = 114$ rote Blüten. Kann aufgrund dieser Beobachtung die Nullhypothese über das 1:2:1-Farbverhältnis abgelehnt werden?

Theorem 4.19 (χ^2-Anpassungstest). *Sei $\alpha \in [0,1]$, $\mathcal{I}$ eine endliche Menge mit $|\mathcal{I}| = r$ und*

$$\Theta = \{\underline{p} = (p_i)_{i\in\mathcal{I}} \text{ Verteilungsgewichte einer Verteilung auf } \mathcal{I}\}.$$

Weiter sei $(X = (X_1, \ldots, X_n), (\mathbf{P}_{\underline{p}})_{\underline{p}\in\Theta})$ ein statistisches Modell, so dass $X_1, \ldots, X_n$ unter $\mathbf{P}_{\underline{p}}$ unabhängig und nach $\underline{p}$ verteilt sind (d.h. $\mathbf{P}(X_k = i) = p_i$). Setze für $i \in \mathcal{I}$

$$S_i := |\{k : X_k = i\}|.$$

Weiter sei $\chi^2_{m,p}$ für $p \in [0,1]$ das p-Quantil von $\chi^2(m)$ und $\Theta_0 = \{\underline{\pi}\}$, $\Theta_A = \Theta \setminus \Theta_0$ für ein $\underline{\pi} \in \Theta$. Für

$$\chi^2_n := \sum_{i\in\mathcal{I}} \frac{(S_i - n\pi_i)^2}{n\pi_i}.$$

ist $(\chi^2_n, (\chi^2_{r-1,1-\alpha}, \infty))$ im Grenzwert $n \to \infty$ ein Test zum Niveau α.

Bemerkung 4.20 (Approximativer Test). Der χ^2-Anpassungstest ist ein approximativer Test für große Stichproben. Unter H_0 gilt nämlich für jedes x (also insbesondere für $x = \chi^2_{r-1,\alpha}$)

$$\lim_{n\to\infty} \mathbf{P}_{\underline{\pi}}(\chi^2_n > x) = \mathbf{P}(X > x)$$

für eine nach $\chi^2(r-1)$ verteilte Zufallsgröße X. In der Praxis ist die Stichprobengröße natürlich nie beliebig groß. Deswegen hat man Faustregeln für die Anwendbarkeit des χ^2-Anpassungstests aufgestellt. Man verwendet den χ^2-Test dann, wenn entweder

$$\begin{aligned} &n\pi_i \geq 3 \text{ für alle } i \in \mathcal{I} \qquad \text{oder} \\ &r \geq 10 \quad \text{und} \quad n\pi_i \geq 1 \text{ für alle } i \in \mathcal{I} \end{aligned} \tag{4.4}$$

gilt.

Beweisskizze von Theorem 4.19. Wir werden nur einige heuristische Bemerkungen anstelle eines kompletten Beweises des Theorems machen. Zunächst bemerken wir, dass die Teststatistik genau dann klein ist, wenn alle S_i nahe an $n\pi_i$ sind. Unter H_0 ist schließlich auch $\mathbf{E}[S_i] = n\pi_i$. Also misst die Teststatistik quadratische Abweichungen von $(S_i)_{i\in\mathcal{I}}$ von den erwarteten Anzahlen $(n\pi_i)_{i\in\mathcal{I}}$. Sind diese quadratischen Abweichungen zu groß, wird H_0 abgelehnt, da ja $(\chi^2_{r-1,1-\alpha}, \infty)$ der Ablehnungsbereich ist.

Wir zeigen nun noch, dass im speziellen Fall $r = 2$ die Teststatistik χ^2_n approximativ $\chi^2(1)$ verteilt ist. Sei also $\mathcal{I} = \{1, 2\}$. Dann ist S_1 nach $B(n, \pi_1)$ verteilt, $S_2 - n\pi_2 = (n - S_1 - n(1 - \pi_1) = -(S_1 - n\pi_1)$, $\frac{1}{\pi_1} + \frac{1}{\pi_2} = \frac{1}{\pi_1\pi_2} = \frac{1}{\pi_1(1-\pi_1)}$ und, wegen dem zentralen Grenzwertsatz, für jedes $x > 0$, für eine nach $N(0, 1)$ verteilte Zufallsvariable Z

$$\begin{aligned} \lim_{n\to\infty} \mathbf{P}_{\underline{\pi}}\Big(\frac{(S_1 - n\pi_1)^2}{n\pi_1} + \frac{(S_2 - n\pi_2)^2}{n\pi_2} \leq x\Big) &= \lim_{n\to\infty} \mathbf{P}_{\underline{\pi}}\Big(\frac{(S_1 - n\pi_1)^2}{n\pi_1(1-\pi_1)} \leq x\Big) \\ &= \lim_{n\to\infty} \mathbf{P}_{\underline{\pi}}\Big(-\sqrt{x} \leq \frac{S_1 - n\pi_1}{\sqrt{n\pi_1(1-\pi_1)}} \leq \sqrt{x}\Big) \\ &= \mathbf{P}(-\sqrt{x} \leq Z \leq \sqrt{x}) = \mathbf{P}(Z^2 \leq x). \end{aligned}$$

Da Z^2 nach $\chi^2(1)$ verteilt ist, haben wir gezeigt, dass approximativ χ^2_n für große n nach $\chi^2(1)$ verteilt ist. □

Beispiel 4.21 (Mendel's Experiment). Wir führen nun den χ^2-Anpassungstest für das Mendel'sche Experiment aus Beispiel 4.18 mit $\alpha = 0.05$ durch. Hier ist $n = 400$,

$$H_0 : \underline{p} = \underline{\pi} := \big(\tfrac{1}{4}, \tfrac{1}{2}, \tfrac{1}{4}\big).$$

Außerdem liefert die Stichprobe $S_{\text{weiß}} = 115$, $S_{\text{rosa}} = 171$ und $S_{\text{rot}} = 114$. Die erste Bedingung aus (4.4) ist sicher erfüllt, so dass wir den χ^2-Test anwenden können. Wir berechnen

$$\chi^2_{400} = \frac{(115-100)^2}{100} + \frac{(171-200)^2}{200} + \frac{(114-100)^2}{100} \approx 8.46 > 5.99 = \chi^2_{2,0.95},$$

H_0 wird also abgelehnt. (Nun darf man nach den biologischen Ursachen für die Abweichung von den Mendel'schen Verhältnissen forschen: z.B. könnte Selektion gegen Heterozygote im Spiel sein, oder Wechselwirkung des Gens für die Blütenfarbe mit anderen Teilen des Genoms.)

Bemerkung 4.22 (Erweiterung des χ^2-Anpassungstests für unbekannte Parameter). Nicht immer ist die Verteilung, nach der die Daten verteilt sein sollen, so klar wie im Beispiel 4.18. Oftmals will man wissen, ob die Daten einer Verteilungsklasse (etwa die der Poisson-Verteilungen mit Parameter $\lambda > 0$) angehören, die noch einen oder mehr Parameter besitzt. In diesem Fall muss man zunächst die Parameter aus den Daten schätzen und kann erst anschließend den χ^2-Anpassungstest durchführen. In einer solchen Situation ist klar, dass schon durch das Schätzen der Parameter eine Anpassung der Verteilung an die Daten vollzogen wird. Den χ^2-Anpassungstest kann man jedoch nachwievor durchführen, indem man die Anzahl der Freiheitsgrade der χ^2-Verteilung reduziert, falls die verwendeten Schätzer Maximum-Likelihood-Schätzer sind. Man geht also folgendermaßen vor (siehe auch Abbildung 4.7):

1. Schätzung der l fehlenden Parameter der Verteilung mittels Maximum-Likelihood, basierend auf $X_1, \dots, X_n$.
2. Die Teststatistik χ^2 aus Theorem 4.19 ist dann approximativ (unter den Annahmen (4.4)) nach
   $\chi^2(r - l - 1)$ verteilt.

Also ist dann $(\chi^2_n, (\chi^2_{r-l-1,1-\alpha}, \infty))$ für große n ein Test zum Niveau α.

Der χ^2 Anpassungstest

überprüft, ob die empirische Verteilung von gruppierten Daten mit einer vorgegebenen Verteilungsklasse übereinstimmt.

Statistisches Modell	$\mathcal{I}$ endlich, $X_1, \dots, X_n$ unabhängig mit Verteilungsgewichten $\underline{p}$ $S_i := \lvert\{k : X_k = i\}\rvert$
Hypothesen	$H_0 : \underline{p} \in \{\underline{\pi}(\lambda) : \lambda \in \Lambda\}$ für eine Verteilungsfamilie $\{\underline{\pi}(\lambda) : \lambda \in \Lambda\}$ $H_A : \underline{p} \notin \{\underline{\pi}(\lambda) : \lambda \in \Lambda\}$
Teststatistik	$\chi^2_n = \sum_{i \in \mathcal{I}} \frac{(S_i - n\pi_i(\widehat{\lambda}))^2}{n\pi_i(\widehat{\lambda})}$ unter H_0 approximativ nach $\chi^2(r - 1 - l)$ verteilt
Bemerkungen	$l =$ Länge von $\widehat{\lambda}$ $= \#$ der mittels Maximum Likelihood zur Bestimmung von π geschätzten Parameter $n\pi_i \geq 3$ für alle $i \in \mathcal{I}$ oder $r \geq 10$ und $n\pi_i \geq 1$ für alle $i \in \mathcal{I}$
p-Wert	$\mathbf{P}(X > \chi^2_n)$, wobei X nach $\chi^2(r - l - 1)$ verteilt ist

Abbildung 4.7: Schematische Darstellung des χ^2-Anpassungstest, wenn l Parameter aus den Daten geschätzt werden müssen.

Beispiel 4.23 (Hufschlagtote). Wir betrachten das klassische Beispiel von Hufschlagtoten der preussischen Armee. Hierbei wurden 14 Regimenter über 20 Jahre beobachtet, und für jedes Regiment und jedes Jahr die Zahl der Hufschlagtoten aufgezeichnet. Folgende Häufigkeiten wurden dabei beobachtet:

Anzahl der Todesfälle	0	1	2	3	4
Häufigkeit	144	91	32	11	2

Es liegt nahe zu vermuten, dass die Anzahl der Hufschlagtoten in jedem Jahr eine Poisson-verteilte Zufallsvariable ist. Um dies zu überprüfen, testen wir mit $\alpha = 0.01$ mittels eines χ^2-Anpassungstests.

Zunächst berechnen wir den Maximum-Likelihood-Schätzer für λ. Dieser ist nach Beispiel 1.13 gegeben durch

$$\widehat{\lambda} = \frac{1}{280}\big(0 \cdot 144 + 1 \cdot 91 + 2 \cdot 32 + 3 \cdot 11 + 4 \cdot 2\big) = 0.7.$$

Damit ergeben sich folgende erwartete Größen:

Anzahl der Todesfälle	0	1	2	3	4
Häufigkeit	139.04	97.33	34.07	7.95	1.39

Da wir fordern, dass alle erwarteten Anzahlen mindestens 5 sind, müssen wir die Fälle mit vielen Hufschlagtoten gruppieren:

Anzahl der Todesfälle	0	1	2	3 oder mehr
Häufigkeit	139.04	97.33	34.07	9.56

Wir setzen also $\mathcal{I} = \{0, 1, 2, 3 \text{ oder mehr}\}$,

$$H_0 : \underline{p} \in \Big\{e^{-\lambda}\Big(\frac{\lambda^0}{0!}, \frac{\lambda^1}{1!}, \frac{\lambda^2}{2!}, \dots\Big) \text{ für ein } \lambda > 0\Big\}.$$

Damit ergibt sich die Teststatistik

$$\begin{aligned}\chi^2_{280} &= \frac{(144 - 139.04)^2}{139.04} + \frac{(91 - 97.33)^2}{97.33} + \frac{(32 - 34.07)^2}{34.07} + \frac{(11 - 7.95)^2}{7.95} + \frac{(2 - 1.61)^2}{1.61} \\ &= 1.95 \leq 9.21 = \chi^2_{2,0.99}.\end{aligned}$$

Damit können wir die Nullhypothese Poisson-verteilter Daten auf dem Signifikanzniveau $\alpha = 0.01$ nicht ablehnen.

Beispiel 4.24 (Nochmal Mendel's Experimente). Eine spezielle Form des Anpassungstest ist der χ^2-Unabhängigkeitstest. Hier beobachtet man zwei Merkmale der Stichprobe. Beispielsweise könnte ein Züchter von Pflanzen zählen, wieviele seiner Pflanzen rot, rosa und weiß blühen, und wieviele glatte und gesprenkelte Blätter haben. Wir gehen davon aus, dass (X_i, Y_i) die Farbe und die Blättermusterung der i-ten Pflanze ist. Ziel ist es herauszufinden, ob X_i und Y_i stochastisch unabhängig sind, d.h. ob etwa

$$\mathbf{P}(X_i = \text{rot}, Y_i = \text{glatt}) = \mathbf{P}(X_i = \text{rot}) \cdot \mathbf{P}(Y_i = \text{glatt})$$

(und ähnliche Gleichungen für die anderen Farben/Blattmusterungen) gilt. Bei einem Versuch, in dem man 400 Blumen als Stichprobe untersucht hat, erhält man folgende Daten, die wir in einer Kontingenztabelle zusammen fassen:

	glatt	gesprenkelt	$\sum$
weiß	41	74	115
rosa	49	122	171
rot	47	67	114
$\sum$	137	263	400

Wir bezeichnen diese Anzahlen mit $S_{\text{weiß, glatt}}, \dots, S_{\text{rot, gesprenkelt}}$. Die Marginalien bezeichnen wir mit $S_{\text{weiß}\bullet}, \dots, S_{\text{rot}\bullet}$ und $S_{\bullet\text{glatt}}$, $S_{\bullet\text{gesprenkelt}}$. Kann aufgrund dieser Daten die Hypothese der Unabhängigkeit verworfen werden?

Theorem 4.25 (χ^2-Unabhängigkeitstest). *Sei $\alpha \in [0,1]$, $\mathcal{I}, \mathcal{J}$ endliche Mengen mit $|\mathcal{I}| = r$ $|\mathcal{J}| = s$ und*

$$\Theta = \{\underline{p} = (p_{ij})_{i\in\mathcal{I}, j\in\mathcal{J}} \text{ Verteilungsgewichte einer Verteilung auf } \mathcal{I}\times\mathcal{J}\}.$$

Weiter sei $((X,Y) = ((X_1,Y_1),\dots,(X_n,Y_n)), (\mathbf{P}_{\underline{p}})_{\underline{p}\in\Theta})$ ein statistisches Modell, so dass $(X_1,Y_1),\dots,(X_n,Y_n)$ unter $\mathbf{P}_{\underline{p}}$ unabhängig und nach $\underline{p}$ verteilt sind (d.h. $\mathbf{P}(X_k = i, Y_k = j) = p_{ij}$). Setze für $i \in \mathcal{I}, j \in \mathcal{J}$

$$\begin{aligned} S_{ij} &:= |\{k : X_k = i, Y_k = j\}|, \\ S_{i\bullet} &:= |\{k : X_k = i\}| = \sum_{j\in\mathcal{J}} S_{ij}, \\ S_{\bullet j} &:= |\{k : Y_k = j\}| = \sum_{i\in\mathcal{I}} S_{ij} \end{aligned}$$

und $p_{i\bullet} := \sum_{j\in\mathcal{J}} p_{ij}, p_{\bullet j} := \sum_{i\in\mathcal{I}} p_{ij}$ für $\underline{p} \in \Theta$. Weiter sei $\chi^2_{m,p}$ für $p \in [0,1]$ das p-Quantil von $\chi^2(m)$ und

$$\Theta_0 = \{\underline{p} \in \Theta : p_{ij} = p_{i\bullet}\cdot p_{\bullet j} \text{ für } i \in \mathcal{I}, j \in \mathcal{J}\},$$

$\Theta_A = \Theta \setminus \Theta_0$. Mit

$$\chi^2_n := \sum_{i=1}^r \sum_{j=1}^s \frac{(S_i - \frac{S_{i\bullet}\cdot S_{\bullet j}}{n})^2}{\frac{S_{i\bullet}\cdot S_{\bullet j}}{n}}$$

ist dann $(\chi^2_n, (\chi^2_{(r-1)(s-1),1-\alpha}, \infty))$ im Grenzwert $n\to\infty$ ein Test zum Niveau α.

Bemerkung 4.26 (Approximativer Test). Genau wie der χ^2-Anpassungstest ist der χ^2-Unabhängigkeitstest nur approximativ für große Stichproben anwendbar. In der Praxis haben sich folgende Regeln für die Gültigkeit des χ^2-Unabhängigkeitstests herausgestellt:

$$\begin{aligned} &S_{i\bullet}\cdot S_{\bullet j}/n \geq 5 \text{ für } r = s = 2, \\ &S_{i\bullet}\cdot S_{\bullet j}/n \geq 2 \text{ für } r = 2, s = 3, \\ &S_{i\bullet}\cdot S_{\bullet j}/n \geq 5 \text{ für alle bis auf ein Paar } i,j \text{ und } \min S_{i\bullet}\cdot S_{\bullet j}/n \geq 1 \text{ für größere } r,s. \end{aligned} \tag{4.5}$$

Beweisskizze zu Theorem 4.25. Wir stellen fest, dass wir uns in derselben Situation wie beim χ^2-Anpassungstest befinden, indem wir $\mathcal{I}\times\mathcal{J}$ als den Zielbereich der beobachteten Zufallsvariablen (X_i,Y_i) ansehen. Wie in Bemerkung 4.22 (und Abbildung 4.7) müssen wir unter H_0 für die Verteilung auf $\mathcal{I}\times\mathcal{J}$ noch die Marginalien $p_{i\bullet}, p_{\bullet j}$ für $i \in \mathcal{I}, j \in \mathcal{J}$ als Parameter schätzen. Die Anzahl der Marginalien, die zu schätzen ist, ist $r-1+s-1$, da $\sum_{i\in\mathcal{I}} p_{i\bullet} = \sum_{j\in\mathcal{J}} p_{\bullet j} = 1$ gelten muss. Wir schätzen hierbei (unter H_0)

$$\widehat{p}_{i\bullet} = \frac{S_{i\bullet}}{n}, \qquad \widehat{p}_{\bullet j} = \frac{S_{\bullet j}}{n}.$$

Damit ist die Teststatistik χ_n^2 approximativ nach $\chi^2(rs-(r-1)-(s-1)-1) = \chi^2((r-1)(s-1))$ verteilt und die Behauptung folgt. □

Der χ^2 Unabhängigkeitstest

überprüft, ob zwei Merkmale einer Stichprobe stochastisch unabhängig sind.

Statistisches Modell: $(X_1, Y_1), \dots, (X_n, Y_n)$ unabhängig, identisch verteilt,

$$\begin{aligned} S_{ij} &:= |\{k : X_k = i, Y_k = j\}|, \\ S_{i\bullet} &:= |\{k : X_k = i\}| = \sum_{j\in\mathcal{J}} S_{ij}, \\ S_{\bullet j} &:= |\{k : Y_k = j\}| = \sum_{i\in\mathcal{I}} S_{ij}. \end{aligned}$$

Hypothesen: $H_0 : \mathbf{P}(X_k = i, Y_k = j) = \mathbf{P}(X_k = i)\cdot\mathbf{P}(Y_k = j)$ für alle i, j

$H_A : \mathbf{P}(X_k = i, Y_k = j) \neq \mathbf{P}(X_k = i)\mathbf{P}(Y_k = j)$ für mindestens ein Paar i, j

Teststatistik: $\chi_n^2 = \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} \frac{\left(S_{ij} - \frac{S_{i\bullet}\cdot S_{\bullet j}}{n}\right)^2}{\frac{S_{i\bullet}\cdot S_{\bullet j}}{n}}$ unter H_0 approximativ nach $\chi^2((r-1)(s-1))$ verteilt

Bemerkungen: Für die approximative Gültigkeit muss gelten, dass

$$\begin{aligned} &S_{i\bullet}\cdot S_{\bullet j}/n \geq 5 \text{ für } r = s = 2 \\ &S_{i\bullet}\cdot S_{\bullet j}/n \geq 2 \text{ für } r = 2, s = 3 \\ &S_{i\bullet}\cdot S_{\bullet j}/n \geq 5 \text{ für alle bis auf ein Paar } i, j \text{ und} \\ &\min S_{i\bullet}\cdot S_{\bullet j}/n \geq 1 \text{ für größere } r, s \end{aligned}$$

Ablehnungsbereich: $\chi_n^2 > \chi^2_{(r-1)(s-1),1-\alpha}$

p-Wert: $\mathbf{P}(X \geq \chi_n^2)$, wobei X nach $\chi^2((r-1)(s-1))$ verteilt ist

Abbildung 4.8: Schematische Darstellung des χ^2-Unabhängigkeitstests

Beispiel 4.27 (Nochmal Mendel's Experimente). Wir betrachten nochmal die beiden Merkmale der Blütenfarbe und Blattmuster des Züchters aus Beispiel 4.24. Aus den Werten für $S_{\text{weiß}\bullet}, \ldots, S_{\bullet\text{gesprenkelt}}$ ergibt sich eine Tabelle der Werte $S_{\text{weiß}\bullet} \cdot S_{\bullet\text{glatt}}/400, \ldots S_{\text{rot}\bullet} \cdot S_{\bullet\text{gesprenkelt}}/400$:

	glatt	gesprenkelt	$\sum$
weiß	39.4	75.6	115
rosa	58.6	112.4	171
rot	39.0	75.0	114
$\sum$	137	263	400

Wir wollen die Annahme der Unabhängigkeit der Blütenfarbe und Blattmuster auf dem Niveau $\alpha = 0.05$ testen. Wir berechnen

$$\chi^2_{400} = \frac{(41 - 39.4)^2}{39.4} + \cdots + \frac{(67 - 75.6)^2}{75.6} \approx 4.94 < 5.99 = \chi^2_{2,0.95}.$$

Damit kann die Nullhypothese einer unabhängigen Ausprägung der Blütenfarbe und Blattmusterung nicht abgelehnt werden.