KAPITEL 10

Tests statistischer Hypothesen

In der Statistik muss man oft Hypothesen testen, z.B. muss man anhand einer Stichprobe entscheiden, ob ein unbekannter Parameter einen vorgegebenen Wert annimmt. Zuerst betrachten wir ein Beispiel.

10.1. Ist eine Münze fair?

Es sei eine Münze gegeben. Wir wollen testen, ob diese Münze fair ist, d.h. ob die Wahrscheinlichkeit von "Kopf", die wir mit θ bezeichnen, gleich $1/2$ ist. Dazu werfen wir die Münze z.B. $n = 200$ Mal. Sei S die Anzahl der Würfe, bei denen die Münze Kopf zeigt. Nun betrachten wir zwei Hypothesen:

- Nullhypothese H_0: Die Münze ist fair, d.h., $\theta = 1/2$.
- Alternativhypothese H_1: Die Münze ist nicht fair, d.h., $\theta \neq 1/2$.

Wir müssen entscheiden, ob wir die Nullhypothese H_0 verwerfen oder beibehalten. Die Entscheidung muss anhand des Wertes von S getroffen werden. Unter der Nullhypothese gilt, dass $\mathbb{E}_{H_0} S = 200 \cdot \frac{1}{2} = 100$. Die Idee besteht nun darin, die Nullhypothese zu verwerfen, wenn S stark von 100 abweicht. Die Größe $|S - 100|$ bezeichnet man in diesem Fall als *Teststatistik*. Dabei sind große Werte von $|S - 100|$ ein Hinweis darauf, dass die Nullhypothese verworfen werden muss, d.h. große Werte sind *signifikant*.
Wir wählen also eine Konstante $c \in \{0, 1, \ldots\}$ und verwerfen H_0, falls $|S - 100| > c$. Andernfalls behalten wir die Hypothese H_0 bei. Bei diesem Vorgehen können wir zwei Arten von Fehlern machen:

- *Fehler* 1. *Art*: H_0 wird verworfen, obwohl H_0 richtig ist.
- *Fehler* 2. *Art*: H_0 wird nicht verworfen, obwohl H_0 falsch ist.

Wie sollte nun die Konstante c (der sogenannte *kritische Wert*) gewählt werden? Man möchte natürlich die Wahrscheinlichkeiten der beiden Arten von Fehlern klein halten. In diesem Beispiel ist es allerdings nicht möglich, die Wahrscheinlichkeit eines Fehlers 2. Art zu bestimmen. Der Grund dafür ist, dass man für die Berechnung dieser Wahrscheinlichkeit den Wert von θ kennen muss, bei einem Fehler 2. Art ist allerdings nur bekannt, dass $\theta \neq 1/2$ ist. Die Wahrscheinlichkeit eines Fehlers 1. Art kann aber sehr wohl bestimmt werden und ist

$$\mathbb{P}_{H_0}[|S - 100| > c] = 2\mathbb{P}_{H_0}[S > 100 + c] = 2 \sum_{k=100+c+1}^{200} \binom{200}{k} \frac{1}{2^{200}},$$

da $S \sim \text{Bin}(200, 1/2)$ unter H_0. Wir wollen nun c so wählen, dass die Wahrscheinlichkeit eines Fehlers 1. Art nicht größer als ein kleines vorgegebenes Niveau $\alpha \in (0, 1)$ ist. Normalerweise wählt man $\alpha = 0.01$ oder 0.05. Hier wählen wir das Niveau $\alpha = 0.05$. Nun rechnet man nach,

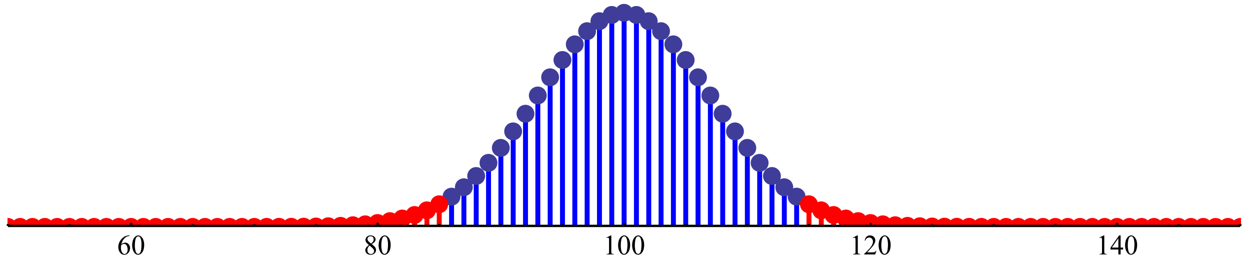

Abbildung 1. Zähldichte der Binomialverteilung mit Parametern $n = 200$ und $\theta = 1/2$. Rot: Ablehnungsbereich. Blau: Annahmebereich.

dass

$$\mathbb{P}_{H_0}[|S-100| > c] = \begin{cases} 0.05596, & \text{für } c = 13, \\ 0.04003, & \text{für } c = 14. \end{cases}$$

Damit die Wahrscheinlichkeit eines Fehlers 1. Art kleiner als $\alpha = 0.05$ ist, müssen wir also $c \geq 14$ wählen. Dabei ist es sinnvoll, c möglichst klein zu wählen, denn sonst vergrößert man die Wahrscheinlichkeit eines Fehlers 2. Art. Also wählen wir $c = 14$. Unsere Entscheidungsregel lautet nun wie folgt:

- Wir verwerfen H_0, falls $|S - 100| > 14$.
- Sonst behalten wir die Hypothese H_0 bei.

Das Beibehalten von H_0 bedeutet nicht, dass H_0 "bewiesen" wurde. Es kann ja immer noch sein, dass die Münze unfair mit einem $\theta = 1/2 + 10^{-10}$ ist und ein dermaßen kleiner Unterschied kann bei 200 Würfen sowieso nicht erkannt werden. Das Beibehalten von H_0 bedeutet lediglich, dass in den vorhandenen Daten keine ausreichenden Hinweise gegen H_0 gefunden wurden.

Eine wichtige Größe zur Auswertung von statistischen Tests ist der p–Wert.

Definition 10.1.1. Als *p–Wert* bezeichnet man die Wahrscheinlichkeit (unter der Nullhypothese), dass die Teststatistik einen mindestens so extremen Wert annimmt, wie der in der Stichprobe beobachtete Wert.

Hat man z.B. bei 200 Würfen 150 Mal "Kopf" beobachtet, so ist der p–Wert gegeben durch

$$\mathbb{P}_{H_0}[|S-100| \geq |150-100|] = 2\mathbb{P}_{H_0}[S \geq 150] = 2\sum_{k=150}^{200} \binom{200}{k} \frac{1}{2^{200}} \approx 8.393 \cdot 10^{-13}.$$

Der Wert 150 weicht vom unter der Nullhypothese erwarteten Wert 100 um 50 ab. Bei einer richtigen Nullhypothese H_0 hat eine Abweichung von mindestens 50 eine Wahrscheinlichkeit von lediglich $8.393 \cdot 10^{-13}$. Deshalb muss in diesem Fall die Nullhypothese ohne große Zweifel verworfen werden.

Der p–Wert liegt immer zwischen 0 und 1. Ein kleiner p–Wert ist ein Hinweis darauf, dass die Nullhypothese verworfen werden muss.

Asymptotischer Test. Im obigen Beispiel kann man für die Berechnung der Wahrscheinlichkeiten die Approximation durch die Normalverteilung benutzen. Es soll ein c mit

$$\mathbb{P}_{H_0}[S - 100 < -c] \leq \frac{\alpha}{2}$$

bestimmt werden. Um die Güte der Approximation zu verbessern, benutzen wir den $\frac{1}{2}$–Trick. Da c ganz ist, ist die obige Ungleichung äquivalent zu

$$\mathbb{P}_{H_0}[S - 100 \leq -c - 0.5] \leq \frac{\alpha}{2}.$$

Unter H_0 gilt $S \sim \text{Bin}(200, 1/2)$ und somit $\mathbb{E}_{H_0} S = 100$, $\operatorname{Var}_{H_0} S = 200 \cdot \frac{1}{2} \cdot \frac{1}{2} = 50$. Die obige Ungleichung ist äquivalent zu

$$\mathbb{P}_{H_0}\left[\frac{S-100}{\sqrt{50}} \leq -\frac{c+0.5}{\sqrt{50}}\right] \leq \frac{\alpha}{2}.$$

Nun können wir die Normalverteilungsapproximation benutzen und die obige Ungleichung durch die folgende ersetzen:

$$\Phi\left(-\frac{c+0.5}{\sqrt{50}}\right) \leq \frac{\alpha}{2}$$

Somit muss für c die folgende Ungleichung gelten:

$$\frac{c+0.5}{\sqrt{50}} \geq -z_{\frac{\alpha}{2}},$$

wobei $z_{\frac{\alpha}{2}}$ das $\frac{\alpha}{2}$–Quantil der Standardnormalverteilung ist. Wegen der Symmetrie der Standardnormalverteilung gilt $-z_{\frac{\alpha}{2}} = z_{1-\frac{\alpha}{2}}$. Für $\alpha = 0.05$ ist $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ und somit ist die obige Ungleichung äquivalent zu $c \geq 13.36$. Somit müssen wir $c = 14$ wählen. Die Entscheidungsregel bleibt genauso wie oben.

10.2. Tests für die Parameter der Normalverteilung

Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ unabhängige und mit Parametern (μ, σ^2) normalverteilte Zufallsvariablen. Wir wollen Hypothesen über die Parameter μ und σ^2 testen. Wir werden die folgenden vier Fälle betrachten:

(1) Tests für μ bei bekanntem σ^2.
(2) Tests für μ bei unbekanntem σ^2.
(3) Tests für σ^2 bei bekanntem μ.
(4) Tests für σ^2 bei unbekanntem μ.

Fall 1: Tests für μ bei bekanntem σ^2 (Gauß–z–Test).

Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma_0^2)$ unabhängig, wobei die Varianz σ_0^2 bekannt sei. Wir wollen verschiedene Hypothesen über μ testen, z.B. $\mu = \mu_0$, $\mu \geq \mu_0$ oder $\mu \leq \mu_0$, wobei μ_0 ein vorgegebener Wert ist. Wir betrachten die Teststatistik

$$T := \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{\sigma_0}.$$

Unter $\mu = \mu_0$ gilt $T \sim \mathrm{N}(0,1)$. Wir betrachten drei Fälle in Abhängigkeit davon, wie die zu testende Hypothese formuliert wird.

Fall 1A. $H_0 : \mu = \mu_0$; $H_1 : \mu \neq \mu_0$. Die Nullhypothese H_0 sollte verworfen werden, wenn $|T|$ groß ist. Dabei sollte die Wahrscheinlichkeit eines Fehlers 1. Art höchstens α sein. Dies führt zu der Entscheidungsregel, dass die Nullhypothese H_0 verworfen wird, falls $|T| > z_{1-\frac{\alpha}{2}}$, s. Abbildung 2 (links).

Fall 1B. $H_0 : \mu \geq \mu_0$; $H_1 : \mu < \mu_0$. Die Nullhypothese H_0 sollte verworfen werden, wenn T klein ist. Also verwerfen wir H_0, falls $T < z_\alpha$, s. Abbildung 2 (Mitte). Unter $\mu = \mu_0$ ist die Wahrscheinlichkeit eines Fehlers erster Art gleich α. Man kann zeigen, dass für $\mu > \mu_0$ (was auch zu H_0 gehört), die Wahrscheinlichkeit, H_0 irrtümlich zu verwerfen, kleiner als α ist.

Fall 1C. $H_0 : \mu \leq \mu_0$; $H_1 : \mu > \mu_0$. Hier sollte H_0 verworfen werden, wenn T groß ist. In diesem Fall wird H_0 verworfen, wenn $T > z_{1-\alpha}$, , s. Abbildung 2 (rechts).

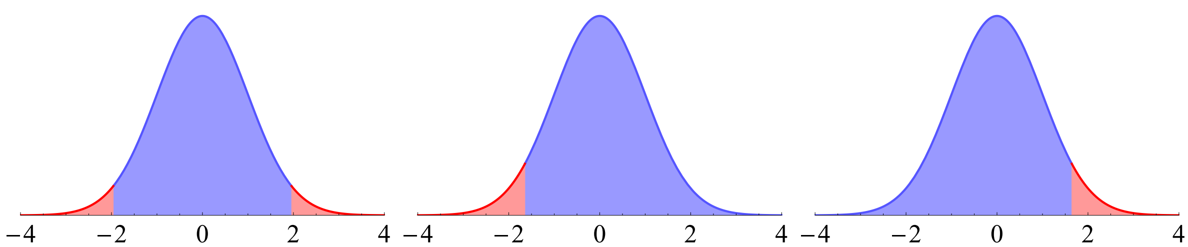

Abbildung 2. Vorgehensweise beim Gauß–z–Test. Rot: Ablehnungsbereich. Blau: Annahmebereich. Links: Zweiseitiger Test (Fall 1A). Mitte: Einseitiger Test (Fall 1B). Rechts: Einseitiger Test (Fall 1C).

Fall 2: Tests für μ bei unbekanntem σ^2 (Student–t–Test).

Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt seien. Wir möchten Hypothesen über μ testen, z. B. $\mu = \mu_0$, $\mu \geq \mu_0$ oder $\mu \leq \mu_0$, wobei μ_0 vorgegeben ist. Die Teststatistik aus Fall 1 können wir dafür nicht verwenden, denn sie enthält den unbekannten Parameter σ^2. Deshalb schätzen wir zuerst σ^2 durch

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Wir betrachten die Teststatistik

$$T := \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{S_n}.$$

Dann gilt unter $\mu = \mu_0$, dass $T \sim t_{n-1}$.

Fall 2A. $H_0 : \mu = \mu_0$; $H_1 : \mu \neq \mu_0$. Die Nullhypothese H_0 sollte verworfen werden, wenn $|T|$ groß ist. Dabei sollte die Wahrscheinlichkeit eines Fehlers 1. Art höchstens α sein. Wegen der

Symmetrie der t–Verteilung erhalten wir die folgende Entscheidungsregel: H_0 wird verworfen, falls $|T| > t_{n-1,1-\frac{\alpha}{2}}$.

Fall 2B. $H_0 : \mu \geq \mu_0$; $H_1 : \mu < \mu_0$. Die Nullhypothese H_0 wird verworfen, wenn $T < t_{n-1,\alpha}$.

Fall 2C. $H_0 : \mu \leq \mu_0$; $H_1 : \mu > \mu_0$. Die Nullhypothese H_0 wird verworfen, wenn $T > t_{n-1,1-\alpha}$.

Fall 3: Tests für σ^2 bei bekanntem μ (χ^2–Streuungstest).

Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu_0, \sigma^2)$ unabhängig, wobei der Erwartungswert μ_0 bekannt sei. Wir wollen verschiedene Hypothesen über die quadratische Streuung σ^2 der Stichprobe testen, wie z. B. $\sigma^2 = \sigma_0^2$, $\sigma^2 \geq \sigma_0^2$ oder $\sigma^2 \leq \sigma_0^2$, wobei σ_0^2 vorgegeben ist. Ein natürlicher Schätzer für σ^2 ist

$$\widetilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)^2.$$

Unter $\sigma^2 = \sigma_0^2$ gilt

$$T := \frac{n\widetilde{S}_n^2}{\sigma_0^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu_0}{\sigma_0} \right)^2 \sim \chi_n^2.$$

Fall 3A. $H_0 : \sigma^2 = \sigma_0^2$; $H_1 : \sigma^2 \neq \sigma_0^2$. Die Nullhypothese H_0 sollte abgelehnt werden, wenn T zu groß oder zu klein ist. Die χ^2–Verteilung ist nicht symmetrisch. Dies führt zu folgender Entscheidungsregel: H_0 wird verworfen, wenn $T < \chi^2_{n,\frac{\alpha}{2}}$ oder $T > \chi^2_{n,1-\frac{\alpha}{2}}$.

Fall 3B. $H_0 : \sigma^2 \geq \sigma_0^2$; $H_1 : \sigma^2 < \sigma_0^2$. Die Nullhypothese H_0 sollte verworfen werden, wenn T zu klein ist. Dies führt zu folgender Entscheidungsregel: H_0 wird verworfen, wenn $T < \chi^2_{n,\alpha}$ ist.

Fall 3C. $H_0 : \sigma^2 \leq \sigma_0^2$; $H_1 : \sigma^2 > \sigma_0^2$. Die Nullhypothese H_0 sollte verworfen werden, wenn T zu groß ist. Dies führt zu folgender Entscheidungsregel: H_0 wird verworfen, wenn $T > \chi^2_{n,1-\alpha}$ ist.

Fall 4: Tests für σ^2 bei unbekanntem μ (χ^2–Streuungstest).

Seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$, wobei μ und σ^2 unbekannt seien. Wir wollen Hypothesen über σ^2 testen, z. B. $\sigma^2 = \sigma_0^2$, $\sigma^2 \geq \sigma_0^2$ oder $\sigma^2 \leq \sigma_0^2$, wobei σ_0^2 vorgegeben ist. Ein natürlicher Schätzer für σ^2 ist in diesem Fall

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$

Unter $\sigma^2 = \sigma_0^2$ gilt

$$T := \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2_{n-1}.$$

Die Entscheidungsregeln sind also die gleichen wie in Fall 3, lediglich muss man die Anzahl der Freiheitsgrade der χ^2–Verteilung durch $n-1$ ersetzen.

10.3. Zweistichprobentests für die Parameter der Normalverteilung

Nun betrachten wir zwei Stichproben $(X_1, \ldots, X_n)$ und $(Y_1, \ldots, Y_m)$. Wir wollen verschiedene Hypothesen über die Lage und die Streuung dieser Stichproben testen. Z. B. kann man sich für die Hypothese interessieren, dass die Erwartungswerte (bzw. Streuungen) der beiden Stichproben gleich sind. Wir machen folgende Annahmen:

(1) $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ sind unabhängige Zufallsvariablen.
(2) $X_1, \ldots, X_n \sim \mathrm{N}(\mu_1, \sigma_1^2)$.
(3) $Y_1, \ldots, Y_m \sim \mathrm{N}(\mu_2, \sigma_2^2)$.

Wir wollen nun Hypothesen über $\mu_1 - \mu_2$ und σ_1^2/σ_2^2 testen. Dabei werden wir uns auf die Nullhypothesen der Form $\mu_1 = \mu_2$ bzw. $\sigma_1^2 = \sigma_2^2$ beschränken. Nullhypothesen der Form $\mu_1 \geq \mu_2$, $\mu_1 \leq \mu_2$, $\sigma_1^2 \geq \sigma_2^2$, $\sigma_1^2 \leq \sigma_2^2$ können analog betrachtet werden.

Fall 1: Test für $\mu_1 = \mu_2$ bei bekannten σ_1^2 und σ_2^2 (Zweistichproben–z–Test).

Es seien also σ_1^2 und σ_2^2 bekannt. Wir können $\mu_1 - \mu_2$ durch $\bar{X}_n - \bar{Y}_m$ schätzen. Unter der Nullhypothese $H_0 : \mu_1 = \mu_2$ gilt, dass

$$T := \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathrm{N}(0, 1).$$

Die Nullhypothese H_0 wird verworfen, wenn $|T|$ groß ist, also wenn $|T| > z_{1-\frac{\alpha}{2}}$.

Fall 2: Test für $\mu_1 = \mu_2$ bei unbekannten aber gleichen σ_1^2 und σ_2^2 (Zweistichproben–t–Test).

Es seien nun σ_1^2 und σ_2^2 unbekannt. Um das Problem zu vereinfachen, werden wir annehmen, dass die Varianzen gleich sind, d.h. $\sigma^2 := \sigma_1^2 = \sigma_2^2$. Wir schätzen σ^2 durch

$$S = \frac{1}{n+m-2} \left(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 + \sum_{j=1}^{m} (Y_j - \bar{Y}_m)^2 \right).$$

Wir betrachten die folgende Teststatistik:

$$T := \frac{\bar{X}_n - \bar{Y}_m}{S\sqrt{\frac{1}{n} + \frac{1}{m}}}.$$

Wir haben bei der Konstruktion der Konfidenzintervalle gezeigt, dass $T \sim t_{n+m-2}$ unter $\mu_1 = \mu_2$. Somit wird die Nullhypothese H_0 verworfen, wenn $|T| > t_{n+m-2, 1-\frac{\alpha}{2}}$.

Fall 3: Test für $\sigma_1^2 = \sigma_2^2$ bei unbekannten μ_1 und μ_2 (F–Test).

Seien also μ_1 und μ_2 unbekannt. Wir wollen die Nullhypothese $H_0 : \sigma_1^2 = \sigma_2^2$ testen. Natürliche Schätzer für σ_1^2 und σ_2^2 sind gegeben durch

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2, \quad S_Y^2 = \frac{1}{m-1} \sum_{j=1}^{m} (Y_j - \bar{Y}_m)^2.$$

Bei der Konstruktion der Konfidenzintervalle haben wir gezeigt, dass für $\sigma_1^2 = \sigma_2^2$

$$T := \frac{S_X^2}{S_Y^2} \sim F_{n-1, m-1}.$$

Die Hypothese H_0 sollte verworfen werden, wenn T zu klein oder zu groß ist. Dabei ist die F–Verteilung nicht symmetrisch. Die Nullhypothese wird also verworfen, wenn $T < F_{n-1,m-1,\frac{\alpha}{2}}$ oder $T > F_{n-1,m-1,1-\frac{\alpha}{2}}$.

Fall 4: Test für $\sigma_1^2 = \sigma_2^2$ bei bekannten μ_1 und μ_2 (F–Test).

Analog zu Fall 3 (Übung).

10.4. Allgemeine Modellbeschreibung

Wir beschreiben nun allgemein das statistische Testproblem. Sei $(\mathbb{P}_\theta)_{\theta\in\Theta}$ eine Familie von Wahrscheinlichkeitsmaßen auf dem Stichprobenraum $(\mathfrak{X}, \mathcal{A})$. Der Parameterraum Θ sei in zwei disjunkte Teilmengen Θ_0 und Θ_1 aufgeteilt, d.h.

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \varnothing.$$

Sei X eine Stichprobe, die zufällig aus $\mathfrak{X}$ gemäß $\mathbb{P}_\theta$ gezogen wird, wobei $\theta \in \Theta$ unbekannt sei. Wir betrachten nun zwei Hypothesen:

(1) Die Nullhypothese H_0: $\theta \in \Theta_0$.
(2) Die Alternativhypothese H_1: $\theta \in \Theta_1$.

Wir sollen anhand der Stichprobe X entscheiden, ob wir H_0 verwerfen oder beibehalten.

Definition 10.4.1. Ein *Test* ist eine messbare Funktion $\varphi : \mathfrak{X} \to \{0, 1\}$.

Interpretation:

- H_0 wird verworfen, falls $\varphi(X) = 1$.
- H_0 wird beibehalten, falls $\varphi(X) = 0$.

Die Menge $K := \{x \in \mathfrak{X} : \varphi(x) = 1\}$ heißt der *Ablehnungsbereich*, denn H_0 wird verworfen, falls $X \in H_0$.

Wir werden auch einen allgemeineren Begriff benötigen:

Definition 10.4.2. Ein *randomisierter Test* ist eine messbare Funktion $\varphi : \mathfrak{X} \to [0, 1]$.

Interpretation: Um die Entscheidung zu treffen, ob H_0 verworfen werden soll, berechnet man zuerst $p := \varphi(X)$. Danach führt man ein Bernoulli–Experiment mit Erfolgswahrscheinlichkeit p durch. Bei Erfolg verwirft man H_0, bei Misserfolg behält man H_0 bei.

Im folgenden betrachten wir immer randomisierte Tests.

Definition 10.4.3. Die Funktion $G : \mathfrak{X} \to [0, 1]$ mit $G(\theta) = \mathbb{E}_\theta\varphi(X)$ heißt die *Gütefunktion* eines Tests.

Dabei ist $\mathbb{E}_\theta\varphi(X)$ die Wahrscheinlichkeit (unter $\mathbb{P}_\theta$), dass der Test φ die Hypothese H_0 verwirft. Es gilt also:

- Für $\theta \in \Theta_0$ ist $G(\theta)$ die Wahrscheinlichkeit, dass H_0 irrtümlicherweise verworfen wird (Fehler 1. Art).
- Für $\theta \in \Theta_1$ ist $1-G(\theta)$ die Wahrscheinlichkeit, dass H_0 irrtümlicherweise beibehalten wird (Fehler 2. Art).

Definition 10.4.4. Für $\theta \in \Theta_1$ heißt $G(\theta) = \mathbb{E}_\theta\varphi(X)$ die *Macht* des Tests. Die Macht ist also die Wahrscheinlichkeit, dass die falsche Nullhypothese entlarvt wird.

Beim Testen können wir zwei Arten von Fehlern machen:

- Fehler 1. Art: H_0 wird verworfen, obwohl H_0 richtig ist.
- Fehler 2. Art: H_0 wird nicht verworfen, obwohl H_0 falsch ist.

Normalerweise versucht man φ (bzw. den Ablehnungsbereich K) so zu wählen, dass die Wahrscheinlichkeit eines Fehlers 1. Art durch ein vorgegebenes Niveau $\alpha \in (0,1)$ beschränkt ist, typischerweise $\alpha = 0.01$ oder 0.05.

Definition 10.4.5. Ein Test $\varphi : \mathfrak{X} \to [0,1]$ hat *Signifikanzniveau* $\alpha \in (0,1)$, falls

$$\mathbb{E}_\theta\varphi(X) \leq \alpha \ \text{ für alle } \theta \in \Theta_0.$$

Sei Φ_α die Menge aller Tests zum Signifikanzniveau α. Unter allen Tests zum Niveau α möchte man nun denjenigen finden, der eine möglichst große Macht hat.

Definition 10.4.6. Wir sagen, dass $\varphi : \mathfrak{X} \to [0,1]$ *gleichmäßig bester Test* zum Niveau α ist, wenn $\varphi \in \Phi_\alpha$ und

$$\mathbb{E}_\theta\varphi(X) = \sup_{\psi \in \Phi_\alpha} \mathbb{E}_\theta\psi(X) \text{ für alle } \theta \in \Theta_1.$$

Diese Bedingung besagt, dass für alle $\theta \in \Theta_1$ der Test φ eine kleinere Wahrscheinlichkeit eines Fehlers 2. Art unter $\mathbb{P}_\theta$ hat als jeder andere Test $\psi \in \Phi_\alpha$.

Zum Schluss definieren wir noch den p–Wert. Stellen wir uns vor, dass wir unsere Testentscheidung auf dem Wert einer Statistik $T : \mathfrak{X} \to \mathbb{R}$ basieren. Es kann z.B. sein, dass große Werte von T für eine Ablehnung von H_0 sprechen. In diesem Fall hat der Test die Form $\varphi(x) = \mathbb{1}_{T(x)\geq c}$ für einen kritischen Wert c.

Definition 10.4.7. Der p–Wert einer Beobachtung $x \in \mathfrak{X}$ ist gegeben durch

$$p\text{–Wert(x)} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta[T(X) \geq T(x)].$$

Ein p–Wert, der kleiner als α ist, führt zur Ablehnung von H_0.

Beispiel 10.4.8. Wir berechnen die Gütefunktion des Gauß–z–Tests. Seien $X_1, \dots, X_n \sim$ N$(\mu, 1)$ unabhängig. Wir wollen $H_0 : \mu = 0$ gegen $H_1 : \mu \neq 0$ testen, wobei wir hier der Einfachheit halber angenommen haben, dass $\mu_0 = 0$ und $\sigma_0^2 = 1$. Der zweiseitige Gauß–z–Test ist gegeben durch

$$\varphi(x_1, \dots, x_n) = \mathbb{1}_{|\sqrt{n}\bar{x}_n| > z_{1-\frac{\alpha}{2}}}.$$

Unter $\mathbb{P}_\mu$ gilt $\bar{X}_n \sim \mathrm{N}(\mu, 1/n)$, somit $\sqrt{n}\bar{X}_n - \sqrt{n}\mu \sim \mathrm{N}(0,1)$. Die Gütefunktion berechnet sich zu

$$\begin{aligned}
G(\mu) &= \mathbb{E}_\mu \varphi(X_1, \dots, X_n) \\
&= \mathbb{P}_\mu \left[|\sqrt{n}\bar{X}_n| > z_{1-\frac{\alpha}{2}} \right] \\
&= \mathbb{P}_\mu \left[\sqrt{n}\bar{X}_n - \sqrt{n}\mu < -z_{1-\frac{\alpha}{2}} - \sqrt{n}\mu \right] + \mathbb{P}_\mu \left[\sqrt{n}\bar{X}_n - \sqrt{n}\mu > z_{1-\frac{\alpha}{2}} - \sqrt{n}\mu \right] \\
&= 1 - \Phi(\sqrt{n}\mu + z_{1-\frac{\alpha}{2}}) + \Phi(\sqrt{n}\mu - z_{1-\frac{\alpha}{2}}),
\end{aligned}$$

s. Abbildung 3 (links), wobei $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$ die Standardnormalverteilungsfunktion bezeichnet. Die Gütefunktion ist gleich α an der Stelle 0 (Wahrscheinlichkeit eines Fehlers 1. Art) und konvergiert gegen 1 für $\mu \to \pm\infty$ (somit wird die Alternative bei großem $|\mu|$ mit großer Wahrscheinlichkeit richtig entlarvt).

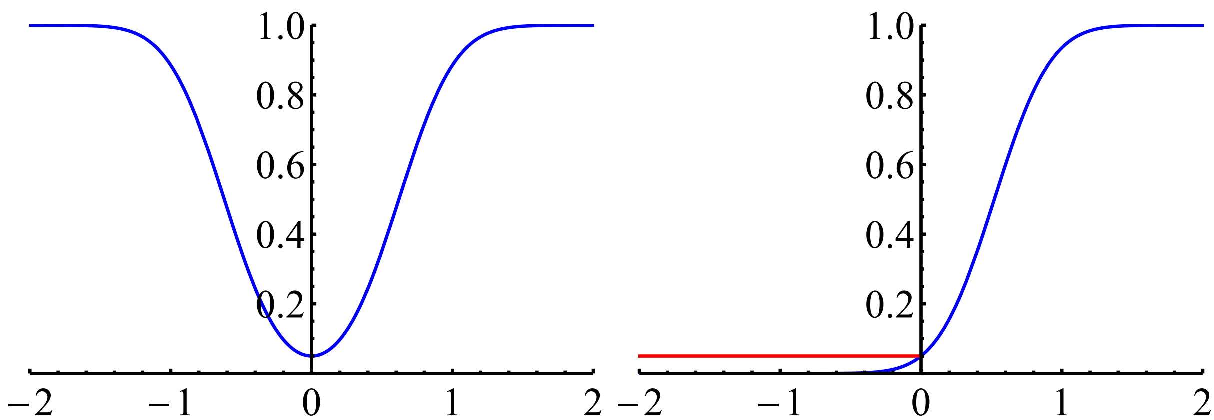

Abbildung 3. Gütefunktion des Gauß–z–Tests für $\mu_0 = 0$, $\sigma_0^2 = 1$ und $n = 10$. Links: Zweiseitiger Test (Fall 1A). Rechts: Einseitiger Test (Fall 1C).

Aufgabe 10.4.9. Seien $X_1, \dots, X_n \sim \mathrm{N}(\mu, 1)$ unabhängig. Betrachten Sie die einseitigen Hypothesen $H_0 : \mu \leq 0$ und $H_1 : \mu > 0$.

(a) Bestimmen Sie die Gütefunktion des einseitigen Gauß–z–Tests

$$\varphi(x_1, \dots, x_n) = \mathbb{1}_{\sqrt{n}\bar{x}_n > z_{1-\alpha}}.$$

(b) Zeigen Sie, dass die Gütefunktion für alle $\mu \leq 0$ unterhalb von α bleibt, s. Abbildung 3 (rechts). D.h. es handelt sich tatsächlich um einen Test zum Niveau α.

10.5. Tests einfacher Hypothesen: Neyman–Pearson–Theorie

In diesem Kapitel betrachten wir den Fall, wenn beide Hypothesen *einfach* sind, d.h.

$$\Theta_0 = \{\theta_0\},\ \Theta_1 = \{\theta_1\},\ \Theta = \{\theta_0, \theta_1\}.$$

Zur Vereinfachung der Notation schreiben wir im Folgenden $\mathbb{P}_0$ bzw. $\mathbb{P}_1$ für $\mathbb{P}_{\theta_0}$ bzw. $\mathbb{P}_{\theta_1}$.

Annahme: Die Wahrscheinlichkeitsmaße $\mathbb{P}_0$ und $\mathbb{P}_1$ besitzen Dichten h_0 und h_1 bzgl. eines σ-endlichen Maßes λ auf $(\mathfrak{X}, \mathcal{A})$.

Wir werden nun zeigen, wie man einen gleichmäßig besten Test zum Niveau α konstruiert. Eine ganz natürliche Vorgehensweise ist diese: man entscheidet sich für H_1 bzw. H_0 wenn der sogenannte *Likelihood–Quotient* $h_1(x)/h_0(x)$ größer bzw. kleiner als ein vorgegebener Wert k ist. Ist der Quotient gleich k, so ist man sich nicht sicher und randomisiert mit Erfolgswahrscheinlichkeit γ.

Definition 10.5.1. Seien $k \in [0, \infty]$ und $\gamma \in [0, 1]$. Ein *Likelihood–Quatienten–Test* (oder *LQ–Test*) ist ein Test der Form

$$\varphi(x) = \begin{cases} 1, & \text{falls } \frac{h_1(x)}{h_0(x)} > k, \\ 0, & \text{falls } \frac{h_1(x)}{h_0(x)} < k, \\ \gamma, & \text{falls } \frac{h_1(x)}{h_0(x)} = k. \end{cases} \tag{10.5.1}$$

Mögliche Unbestimmtheiten der Form $0/0$ werden wir im Folgenden ignorieren, denn die Menge $A := \{x \in \mathfrak{X} \colon h_0(x) = h_1(x) = 0\}$ ist eine Nullmenge bzgl. $\mathbb{P}_0$ und $\mathbb{P}_1$. Somit ist die Wahrscheinlichkeit, dass die Stichprobe X in A landet gleich 0 sowohl unter H_0 als auch unter H_1.

Abbildung 4. Jerzy Neyman und Karl Pearson

Lemma 10.5.2 (Neyman–Pearson, Teil 1)**.** Sei φ ein LQ–Test mit $\mathbb{E}_0\varphi(X) = \alpha$. Dann gilt

$$\mathbb{E}_1\varphi(X) = \sup_{\psi\colon \mathbb{E}_0\psi(X)\le\alpha} \mathbb{E}_1\psi(X),$$

d.h. φ ist gleichmäßig bester Test zum Niveau α.

Beweis. Sei ψ ein Test zum Niveau α, d.h. $\mathbb{E}_0\psi(X) \le \alpha$. Es reicht zu zeigen, dass

$$\mathbb{E}_1\varphi(X) \ge \mathbb{E}_1\psi(X).$$

Wir behaupten, dass

$$(\varphi(x) - \psi(x))(h_1(x) - kh_0(x)) \ge 0 \text{ für alle } x \in \mathfrak{X}. \tag{10.5.2}$$

Um dies zu zeigen, betrachten wir drei Fälle:

Fall 1: $h_1(x) - kh_0(x) > 0$. Dann gilt $\varphi(x) = 1$ und folglich $\varphi(x) - \psi(x) \ge 0$, woraus sich die Behauptung (10.5.2) ergibt.

Fall 2: $h_1(x) - kh_0(x) < 0$. Dann gilt $\varphi(x) = 0$. Es folgt $\varphi(x) - \psi(x) \le 0$, und (10.5.2) ist richtig.

Fall 3: $h_1(x) - kh_0(x) = 0$. In diesem Fall ist das Produkt auf der linken Seite von (10.5.2) gleich 0 und (10.5.2) stimmt.

Indem wir (10.5.2) bzgl. λ integrieren, erhalten wir

$$\int_{\mathfrak{X}} (\varphi - \psi)h_1 d\lambda \ge k \int_{\mathfrak{X}} (\varphi - \psi)h_0 d\lambda.$$

Nachdem wir Integrale als Erwartungswerte darstellen, ergibt sich

$$\mathbb{E}_1[\varphi(X) - \psi(X)] \ge k\,\mathbb{E}_0[\varphi(X) - \psi(X)].$$

Es gilt allerdings $\mathbb{E}_0\varphi(X) = \alpha$, während $\mathbb{E}_0\psi(X) \le \alpha$. Somit ist der Erwartungswert auf der rechten Seite nichtnegativ und es folgt, dass $\mathbb{E}_1[\varphi(X) - \psi(X)] \ge 0$. Das beweist die Behauptung. $\square$

Lemma 10.5.3 (Neyman–Pearson, Teil 2)**.** Zu jedem $\alpha \in (0, 1)$ lassen sich $k \in [0, \infty)$ und $\gamma \in [0, 1]$ finden, so dass für den durch (10.5.1) definierten Test φ

$$\mathbb{E}_0\varphi(X) = \alpha$$

gilt. Laut Teil 1 des Neyman–Pearson–Lemmas ist φ gleichmäßig bester Test zum Niveau α.

Beweis. Sei φ gegeben durch (10.5.1). Durch eine geeignete Wahl von k und γ wollen wir erreichen, dass

$$\mathbb{E}_0\varphi(X) \overset{\text{def}}{=} \mathbb{P}_0[T > k] + \gamma\mathbb{P}_0[T = k] \overset{!}{=} \alpha.$$

Die Statistik T ist $\mathbb{P}_0$–fast sicher endlich, denn es gilt

$$\mathbb{P}_0[T = +\infty] = \mathbb{P}_0[h_0 = 0] = \int_{\{h_0=0\}} h_0 d\lambda = \int_{\{h_0=0\}} 0 d\lambda = 0.$$

Die Funktion $y \mapsto \mathbb{P}_0[T > y]$ ist rechtsstetig, monoton nichtsteigend und es gilt $\lim_{y\to+\infty} \mathbb{P}_0[T > y] = 0$ und $\mathbb{P}_0[T > y] = 1$ falls $y < 0$. Startend mit $y = +\infty$ verkleinern wir den Wert von y solange $\mathbb{P}_0[T > y] \le \alpha$ gilt. D.h. wir definieren

$$k = \inf\{y > 0 \colon \mathbb{P}_0[T > y] \le \alpha\} \in [0, \infty).$$

Nun gibt es zwei Fälle.

Fall 1: Der Wert α wird erreicht, d.h. $\mathbb{P}_0[T > k] = \alpha$. Dann können wir $\gamma = 0$ setzen.

Fall 2: Der Wert α wird übersprungen, d.h. $\mathbb{P}_0[T > k] < \alpha$ und $\mathbb{P}_0[T \ge k] \ge \alpha$. Wir definieren dann

$$\gamma := \frac{\alpha - \mathbb{P}_0[T > k]}{\mathbb{P}_0[T = k]} \in [0, 1],$$

wobei wir bemerken, dass $\alpha - \mathbb{P}_0[T > k] \le \mathbb{P}_0[T \ge k] - \mathbb{P}_0[T > k] = \mathbb{P}_0[T = k]$ und folglich $\gamma \le 1$. □

Beispiel 10.5.4 (Likelihood–Quotienten–Test für den Parameter der Binomialverteilung)**.** Es sei $X \sim \text{Bin}(n, \theta)$ mit einem unbekannten $\theta \in (0, 1)$. Wir betrachten die einfachen Hypothesen

$$H_0 : \theta = \theta_0 \quad \text{und} \quad H_1 : \theta = \theta_1,$$

wobei $\theta_0, \theta_1 \in (0, 1)$ vorgegeben seien und wir der Einfachheit halber $\theta_0 < \theta_1$ voraussetzen. Der gesunde Menschenverstand sagt, dass wir H_0 verwerfen müssen, wenn X "zu groß" ist. Diese Intuition wird durch die Neyman–Pearson–Theorie bestätigt.
Der Stichprobenraum ist $\mathfrak{X} = \{0, \ldots, n\}$. Sei λ das Zählmaß auf $\{0, \ldots, n\}$, d.h. $\lambda(\{x\}) = 1$ für alle $x \in \mathfrak{X}$. Die Zähldichten h_0 und h_1 sind gegeben durch

$$h_i(x) = \binom{n}{x} \theta_i^x (1 - \theta_i)^{n-x}, \quad x \in \{0, \ldots, n\}, \quad i \in \{0, 1\}.$$

Für den Likelihood–Quotienten erhalten wir

$$\Lambda(x) := \frac{h_1(x)}{h_0(x)} = \frac{\binom{n}{x}\theta_1^x(1-\theta_1)^{n-x}}{\binom{n}{x}\theta_0^x(1-\theta_0)^{n-x}} = \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^x \cdot \left(\frac{1-\theta_1}{1-\theta_0}\right)^n.$$

Der Likelihood–Quotienten–Test fällt die Entscheidung in Abhängigkeit davon, ob $\Lambda(x)$ größer, kleiner oder gleich einem bestimmten Wert k ist. Da aber $\frac{\theta_1}{\theta_0} > 1$ und $\frac{1-\theta_0}{1-\theta_1} > 1$, ist $\Lambda(x)$ eine monoton steigende Funktion von x. Somit gilt

$$\Lambda(x) > k \Longleftrightarrow x > k^*, \quad \Lambda(x) < k \Longleftrightarrow x < k^*, \quad \Lambda(x) = k \Longleftrightarrow x = k^*,$$

für einen passenden Wert k^*. Wir können also den Likelihood–Quotienten–Test auch folgendermaßen darstellen:

$$\varphi(x) = \begin{cases} 1, & \text{falls } x > k^*, \\ 0, & \text{falls } x < k^*, \\ \gamma^*, & \text{falls } x = k^* \end{cases}$$

für $x \in \{0, 1 \ldots, n\}$. Wir müssen noch die Werte $k^* \in \{0, \ldots, n\}$ und $\gamma^* \in [0, 1]$ so bestimmen, dass die Wahrscheinlichkeit eines Fehlers 1. Art gleich α wird, d.h.

$$\mathbb{P}_0[X > k^*] + \gamma^* \mathbb{P}_0[X = k^*] = \alpha. \tag{10.5.3}$$

Startend mit $k = n$ (in welchem Fall $\mathbb{P}_0[X > k] = 0$ ist) verkleinern wir den Wert k solange die Wahrscheinlichkeit $\mathbb{P}_0[X > k]$ unterhalb von α bleibt. Sei $k = k^*$ der kleinstmögliche Wert, für den noch $\mathbb{P}_0[X > k] \leq \alpha$ gilt, d.h.

$$k^* = \min \left\{ k \in \{0, \ldots, n\} : \mathbb{P}_0[X > k] \stackrel{\text{def}}{=} \sum_{i=k}^{n} \binom{n}{i} \theta_0^i (1-\theta_0)^{n-i} \leq \alpha \right\}.$$

Würden wir nun H_0 verwerfen, wenn $X > k^*$, und sonst H_0 beibehalten, so wäre das Signifikanzniveau $\leq \alpha$. Bei einer weiteren Verkleinerung von k^* würde das Signifikanzniveau den Wert α übersteigen. Um das Niveau von *exakt* α zu erreichen, müssen wir im Fall $X = k^*$ eine zufällige Entscheidung treffen. Entscheiden wir uns für H_1 mit Wahrscheinlichkeit

$$\gamma^* = \frac{\alpha - \mathbb{P}_0[X > k^*]}{\mathbb{P}_0[X = k^*]} \in [0, 1],$$

so ist (10.5.3) erfüllt und das Niveau ist exakt α.

Beispiel 10.5.5 (Erkennung eines Signals im Rauschen)**.** Es seien $X_1, \ldots, X_n \sim N(\mu, \sigma_0^2)$ unabhängig, wobei σ_0^2 bekannt sei. Fasst man $X_1, \ldots, X_n$ als Messungen eines Signals zu verschiedenen Zeitpunkten auf, so kann man folgende Hypothesen aufstellen:

$$\begin{aligned} &H_0: \ \mu = 0, \ \text{d.h. es wurde nur Rauschen empfangen,} \\ &H_1: \ \mu = \mu_1, \ \text{d.h. es wurde ein verrauschtes Signal der Stärke } \mu_1 \text{ empfangen.} \end{aligned}$$

Dabei sein $\mu_1 > 0$ bekannt. Der Stichprobenraum ist $\mathfrak{X} = \mathbb{R}^n$. Die Dichten von $\mathbb{P}_0$ und $\mathbb{P}_1$ bzgl. des Lebesgue–Maßes λ sind

$$h_0(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma_0} \right)^n e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n x_i^2}, \quad h_1(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma_0} \right)^n e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu_1)^2}.$$

Nach dem Lemma von Neyman–Pearson basiert der gleichmäßig beste Test auf dem Wert der LQ–Statistik

$$\Lambda(x_1, \ldots, x_n) := \frac{h_1(x_1, \ldots, x_n)}{h_0(x_1, \ldots, x_n)} = e^{\frac{\mu_1}{\sigma_0^2} \sum_{i=1}^n x_i - \frac{n}{2\sigma_0^2} \mu_1^2}.$$

Da Λ eine monoton steigende Funktion von $\bar{x}_n$ ist, hat der gleichmäßig beste Test die Form

$$\varphi(x_1, \ldots, x_n) = \begin{cases} 1, & \text{falls } \bar{x}_n > k^*, \\ 0, & \text{falls } \bar{x}_n < k^*, \end{cases}$$

wobei der Fall $\bar{x}_n = k^*$ wegen der Stetigkeit der Normalverteilung ignoriert werden kann und k^* so gewählt werden muss, dass $\mathbb{P}_0[\bar{X}_n > k^*] = \alpha$. Da $\sqrt{n}\bar{X}_n$ unter H_0 standardnormalverteilt ist, müssen wir $k^* = z_{1-\alpha}/\sqrt{n}$ wählen. Der gleichmäßig beste Test sieht also folgendermaßen aus:

$$\varphi(x_1, \ldots, x_n) = \begin{cases} 1, & \text{falls } \sqrt{n}\bar{x}_n > z_{1-\alpha}, \\ 0, & \text{falls } \sqrt{n}\bar{x}_n < z_{1-\alpha}. \end{cases}$$

Es sei bemerkt, dass der resultierende Test nicht von μ_1 abhängt.

Aufgabe 10.5.6. Bestimmen Sie die Wahrscheinlichkeit eines Fehlers 2. Art für den obigen Test. Zeigen Sie, dass diese Wahrscheinlichkeit für $n \to \infty$ gegen 0 konvergiert.

Aufgabe 10.5.7. Sei φ ein gleichmäßig bester Test zum Niveau α für $H_0 = \{\theta_0\}$ gegen $H_1 = \{\theta_1\}$. Zeigen Sie, dass $\mathbb{E}_{\theta_1}\varphi(X) \geq \alpha$.

10.6. Tests für einseitige Hypothesen bei monotonen Dichtequotienten

Sei $(\mathbb{P}_\theta)_{\theta\in\Theta}$ eine Familie von Wahrscheinlichkeitsmaßen auf dem Stichprobenraum $(\mathfrak{X}, \mathcal{A})$. In diesem Abschnitt nehmen wir an, dass $\Theta = (\theta_-, \theta_+) \subset \mathbb{R}$ ein (möglicherweise unendliches) Intervall ist und betrachten die einseitigen Hypothesen

$$H_0 : \theta \leq \theta_0 \quad \text{und} \quad H_1 : \theta > \theta_0,$$

wobei $\theta_0 \in \Theta$ vorgegeben sei.

Die Aufgabe, H_0 gegen H_1 zu testen, ist sicherlich schwieriger, als die Aufgabe, die einfachen Hypothese $\theta = \theta_1$ und $\theta = \theta_2$ gegeneinander zu testen, wobei $\theta_1 \leq \theta_0 < \theta_2$. Die in Beispielen 10.5.4 und 10.5.5 konstruierten gleichmäßig besten Tests für einfache Hypothesen basieren jeweils auf einer Statistik T, die unabhängig von der Wahl von θ_1 und θ_2 ist. Diese Kohärenzeigenschaft lässt hoffen, dass auch der gleichmäßig beste Test für die einseitigen Hypothesen auf derselben Statistik basieren muss. Wir werden nun eine allgemeine Eigenschaft von statistischen Modellen formulieren, die garantiert, dass alle LQ–Tests von einfachen Hypothesen auf derselben Statistik basieren.

Annahme: $(\mathbb{P}_\theta)_{\theta\in\Theta}$ ist eine *dominierte* Familie von Wahrscheinlichkeitsmaßen auf $(\mathfrak{X}, \mathcal{A})$, d.h. es gibt ein σ-endliches Maß λ auf $(\mathfrak{X}, \mathcal{A})$, sodass für jedes $\theta \in \Theta$ das Wahrscheinlichkeitsmaß $\mathbb{P}_\theta$ eine Dichte h_θ bzgl. λ besitzt.

Definition 10.6.1. Die Familie $(\mathbb{P}_\theta)_{\theta\in\Theta}$ besitzt *monotone Dichtequotienten* in einer Statistik $T : \mathfrak{X} \to \mathbb{R}$, wenn für alle $\theta_1 < \theta_2$ eine monoton steigende Funktion $H_{\theta_1,\theta_2} : \mathbb{R} \to (0, \infty)$ existiert mit

$$\frac{h_{\theta_2}(x)}{h_{\theta_1(x)}} = H_{\theta_1,\theta_2}(T(x)) \quad \mathbb{P}_{\theta_1}\text{– und } \mathbb{P}_{\theta_2}\text{–f.s.}$$

Beispiel 10.6.2. Sei $\mathbb{P}_\theta$ die Binomialverteilung $\text{Bin}(n, \theta)$ auf $\mathfrak{X} = \{0, \ldots, n\}$. Der Parameterraum ist dabei $\Theta = (0, 1)$. Als λ nehmen wir das Zählmaß auf $\{0, \ldots, n\}$, d.h. $\lambda(\{x\}) = 1$ für alle $x \in \mathfrak{X}$. Die Zähldichte h_θ ist dann gegeben durch

$$h_\theta(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x \in \{0, \ldots, n\}.$$

Die Dichtequotienten sehen für $\theta_1 < \theta_2$ folgendermaßen aus:

$$\frac{h_{\theta_2}(x)}{h_{\theta_1(x)}} = \frac{\binom{n}{x}\theta_1^x(1-\theta_1)^{n-x}}{\binom{n}{x}\theta_2^x(1-\theta_2)^{n-x}} = \left(\frac{\theta_2(1-\theta_1)}{\theta_1(1-\theta_2)}\right)^x \cdot \left(\frac{1-\theta_1}{1-\theta_0}\right)^n,$$

was eine monoton steigende Funktion von $T(x) := x$ ist, denn $\frac{\theta_2}{\theta_1} > 1$ und $\frac{1-\theta_1}{1-\theta_2} > 1$. Somit liegen hier monotone Dichtequotienten vor.

Beispiel 10.6.3. Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit Dichte/Zähldichte $h_\theta(x)$. Bildet $h_\theta(x) = a(\theta)b(x)e^{c(\theta)d(x)}$ eine Exponentialfamilie, so gilt für den Likelihood–Quotienten

$$\frac{h_{\theta_2}(x_1)\dots h_{\theta_2}(x_n)}{h_{\theta_1}(x_1)\dots h_{\theta_1}(x_n)} = \exp\left\{(c(\theta_2) - c(\theta_1))\sum_{i=1}^n d(x_i)\right\}.$$

Ist die Funktion $c(\theta)$ monoton steigend, so liegen monotone Dichtequotienten in der Statistik $T(x_1, \dots, x_n) := \sum_{i=1}^n d(x_i)$ vor. Diese Beobachtung liefert mehrere Beispiele von statistischen Modellen mit monotonen Dichtequotienten, etwa

(a) $X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma_0^2)$ unabhängig mit bekanntem $\sigma_0^2 > 0$. Dabei ist $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$.
(b) $X_1, \dots, X_n \sim \mathrm{N}(\mu_0, \sigma^2)$ unabhängig mit bekanntem $\mu_0 \in \mathbb{R}$. Dabei ist $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i^2$.
(c) $X_1, \dots, X_n \sim \mathrm{Poi}(\theta)$ unabhängig. Dabei ist $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$.
(d) $X_1, \dots, X_n \sim \mathrm{Exp}(1/\theta)$ unabhängig. Dabei ist $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$.

Der nächste Satz beschreibt den gleichmäßig besten Test zum Niveau α für einseitige Hypothesen bei monotonen Dichtequotienten in einer Statistik T. Dieser Test basiert auf dem Wert der Statistik T.

Satz 10.6.4 (Gleichmäßig bester Test für einseitige Hypothesen)**.** Sei $(\mathbb{P}_\theta)_{\theta\in\Theta}$ mit $\Theta = (\theta_-, \theta_+) \subset \mathbb{R}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\mathfrak{X}, \mathcal{A})$ mit monotonen Dichtequotienten in einer Statistik $T : \mathfrak{X} \to \mathbb{R}$. Sei $\theta_0 \in \Theta$ vorgegeben und betrachte die einseitigen Hypothesen $H_0 : \theta \le \theta_0$ und $H_1 : \theta > \theta_0$.

(a) Zu jedem $\alpha \in (0, 1)$ existieren $k^* \in \mathbb{R}$ und $\gamma^* \in [0, 1]$ mit

$$\mathbb{P}_{\theta_0}[T > k^*] + \gamma^*\mathbb{P}_{\theta_0}[T = k^*] = \alpha.$$

(b) Der durch

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } T(x) > k^*, \\ 0, & \text{falls } T(x) < k^*, \\ \gamma^*, & \text{falls } T(x) = k^*. \end{cases}$$

definierte Test φ^* ist gleichmäßig bester Test für H_0 gegen H_1 zum Niveau α.

Beweis. Wir zeigen nur Teil (b), denn der Beweis von Teil (a) verläuft genauso wie im zweiten Teil des Neyman–Pearson–Lemmas.

Schritt 1: Die Macht von φ^*. Sei $\theta > \theta_0$. Wir zeigen, dass für jeden Test φ mit $\mathbb{E}_{\theta_0}\varphi(X) \le \alpha$ gilt

$$\mathbb{E}_\theta\varphi^*(X) \ge \mathbb{E}_\theta\varphi(X).$$

Das bedeutet, dass φ^* eine kleinere Wahrscheinlichkeit eines Fehlers 2. Art besitzt als jeder Test φ zum Niveau α. Wir behaupten, dass

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} > k, \\ 0, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} < k, \\ \gamma^*, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} = k, \end{cases} \tag{10.6.1}$$

wobei $k = H_{\theta_0,\theta}(k^*)$. In der Tat,

$$\frac{h_\theta(x)}{h_{\theta_0}(x)} > k \iff H_{\theta_0,\theta}(T(x)) > k = H_{\theta_0,\theta}(k^*) \iff T(x) > k^*,$$

wobei wir die Monotonie von $H_{\theta_0,\theta}$ benutzt haben. Analoge Äquivalenzen gelten für "$<$" und "$=$".
Es folgt aus (10.6.1), dass φ^* ein LQ–Test für die einfache Hypothese $\tilde{H}_0 := \{\theta_0\}$ gegen $\tilde{H}_1 := \{\theta_1\}$ ist. Das Niveau dieses Tests ist

$$\mathbb{E}_{\theta_0}\varphi^*(X) = \mathbb{P}_{\theta_0}[T > k^*] + \gamma^*\mathbb{P}_{\theta_0}[T = k^*] = \alpha.$$

Es folgt aus dem ersten Teil des Neyman–Pearson–Lemmas, dass φ^* eine nicht kleinere Macht besitzt, als jeder andere Test zum Niveau $\leq \alpha$, also z.B. $\mathbb{E}_\theta\varphi^*(X) \geq \mathbb{E}_\theta\varphi(X)$. Das beweist die Behauptung von Schritt 1.

Schritt 2: Das Niveau von φ^*. Wir zeigen, dass φ^* ein Test zum Niveau α ist, d.h.

$$\mathbb{E}_\theta\varphi^*(X) \leq \alpha \text{ für alle } \theta \leq \theta_0.$$

Für $\theta = \theta_0$ fogt das aus Teil (a) des Satzes. Sei also $\theta < \theta_0$. Wir behaupten, dass

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} < k', \\ 0, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} > k', \\ \gamma^*, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} = k', \end{cases} \tag{10.6.2}$$

wobei $k' = 1/H_{\theta,\theta_0}(k^*)$. In der Tat,

$$\frac{h_\theta(x)}{h_{\theta_0}(x)} < k' \iff \frac{h_{\theta_0}(x)}{h_\theta(x)} > \frac{1}{k'} \iff H_{\theta,\theta_0}(T(x)) > \frac{1}{k'} = H_{\theta,\theta_0}(k^*) \iff T(x) > k^*,$$

wegen der Monotonie von H_{θ,θ_0}. Ähnliche Äquivalenzen gelten für "$>$" und "$=$", was die Behauptung (10.6.2) beweist. Wir betrachten nun

$$\psi^*(x) := 1 - \varphi^*(x) = \begin{cases} 0, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} < k', \\ 1, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} > k', \\ 1-\gamma^*, & \text{falls } \frac{h_\theta(x)}{h_{\theta_0}(x)} = k', \end{cases}$$

Somit ist ψ^* ein LQ–Test für $\tilde{H}_0 := \{\theta_0\}$ gegen $\tilde{H}_1 := \{\theta\}$ zum Niveau

$$\mathbb{E}_{\theta_0}\psi^*(X) = 1 - \mathbb{E}_{\theta_0}\varphi^*(X) = 1 - \alpha.$$

Aus dem ersten Teil des Neyman–Pearson–Lemmas folgt, dass ψ^* gleichmäßig bester Test zum Niveau $1-\alpha$ ist, d.h.

$$\mathbb{E}_\theta\psi^*(X) = \sup_{\psi:\, \mathbb{E}_{\theta_0}\psi(X) = 1-\alpha} \mathbb{E}_\theta\psi(X).$$

Somit gilt für φ^*, dass

$$\begin{aligned}\mathbb{E}_\theta\varphi^*(X) = 1 - \mathbb{E}_\theta\psi^*(X) = 1 - &\sup_{\psi:\, \mathbb{E}_{\theta_0}\psi(X) = 1-\alpha} \mathbb{E}_\theta\psi(X) \\ &= \inf_{\psi:\, \mathbb{E}_{\theta_0}\psi(X) = 1-\alpha} \mathbb{E}_\theta[1-\psi(X)] = \inf_{\varphi:\, \mathbb{E}_{\theta_0}\varphi(X) = \alpha} \mathbb{E}_\theta\varphi(X),\end{aligned}$$

wobei wir beim letzten Übergang $\varphi := 1 - \psi$ gesetzt haben. Nun ist aber $\varphi(X) := \alpha$ ein gültiger Test mit $\mathbb{E}_{\theta_0}\varphi(X) = \mathbb{E}_\theta\varphi(X) = \alpha$, somit ergibt sich

$$\mathbb{E}_\theta\varphi^*(X) = \inf_{\varphi:\, \mathbb{E}_{\theta_0}\varphi(X)=\alpha} \mathbb{E}_\theta\varphi(X) \le \alpha,$$

was die Behauptung beweist. □

Beispiel 10.6.5 (Einseitiger Binomial–Test ist gleichmäßig bester Test)**.** Ein alterprobtes Medikament führe zu einer Besserung mit bekannter Wahrscheinlichkeit $\theta_0 \in (0,1)$. Ein neues Medikament wurde in n Fällen erprobt und führte zu einer Besserung in $x \in \{0,\ldots,n\}$ Fällen. Wir betrachten die Hypothesen

$$\begin{aligned}
&H_0: \text{neues Medikament ist nicht besser als das alte},\\
&H_1: \text{neues Medikament ist besser}.
\end{aligned}$$

Das folgende statistische Modell erscheint natürlich: Die Beobachtung $X \sim \mathrm{Bin}(n,\theta)$ ist binomialverteilt mit einem unbekanntem $\theta \in (0,1)$, der Stichprobenraum ist $\mathfrak{X} = \{0,\ldots,n\}$. Dann lauten unsere Hypothesen $H_0: \theta \le \theta_0$ und $H_1: \theta > \theta_0$. Die Bedingung der monotonen Dichtequotienten gilt mit $T(x) = x$, wie in Beispiel 10.6.2 gezeigt wurde. Somit hat der gleichmäßig bester Test die Form

$$\varphi(x) = \begin{cases} 1, & \text{falls } x > k^*,\\ 0, & \text{falls } x < k^*,\\ \gamma^*, & \text{falls } x = k^* \end{cases}$$

für $x \in \{0,1\ldots,n\}$. Es bleibt nur noch, die Werte k^* und γ^* so zu wählen, dass die Wahrscheinlichkeit H_0 unter $\theta = \theta_0$ irrtümlich zu verwerfen gleich α wird, d.h.

$$\mathbb{E}_{\theta_0}\varphi(X) = \mathbb{P}_{\theta_0}[T > k^*] + \gamma^*\mathbb{P}_{\theta_0}[T = k^*] = \alpha.$$

Die Parameter k^* und γ^* können nun genauso wie in Beispiel 10.5.4 bestimmt werden, nämlich

$$\begin{aligned}
k^* &= \min\left\{k \in \{0,\ldots,n\} : \mathbb{P}_{\theta_0}[X > k] \overset{\text{def}}{=} \sum_{i=k}^{n}\binom{n}{i}\theta_0^i(1-\theta_0)^{n-i} \le \alpha\right\},\\
\gamma^* &= \frac{\alpha - \mathbb{P}_{\theta_0}[X > k^*]}{\mathbb{P}_{\theta_0}[X = k^*]}.
\end{aligned}$$

Beispiel 10.6.6 (Einseitiger Gauß–z–Test ist gleichmäßig bester Test)**.** Betrachte unabhängige Beobachtungen $X_1,\ldots,X_n \sim \mathrm{N}(\mu,\sigma_0^2)$, wobei $\sigma_0^2 > 0$ bekannt sei. Wir interessieren uns für die Hypothesen

$$H_0: \mu \le \mu_0 \quad \text{und} \quad H_1: \mu > \mu_0,$$

wobei μ_0 vorgegeben sei. Der einseitige Gauß–z–Test basiert auf der Teststatistik

$$T := \sqrt{n}\frac{\bar{X}_n - \mu_0}{\sigma_0},$$

die unter $\mu = \mu_0$ standardnormalverteilt ist, und verwirft H_0 falls $T > z_{1-\alpha}$. Um zu zeigen, dass dieser Test gleichmäßig bester Test zum Niveau α ist, müssen wir die Bedingung der monotonen Dichtequetienten überprüfen. Der Stichprobenraum ist $\mathfrak{X} = \mathbb{R}^n$ und die Dichte von $\mathbb{P}_\mu$ bzgl. des n-dimensionalen Lebesgue–Maßes ist

$$h_\mu(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i - \mu)^2\right\}.$$

Für beliebige $\mu_1 < \mu_2$ ist der Dichtequotient gegeben durch

$$\begin{aligned}
\frac{h_{\mu_2}(x_1, \ldots, x_n)}{h_{\mu_1}(x_1, \ldots, x_n)} &= \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n \left((x_i - \mu_2)^2 - (x_i - \mu_1)^2\right)\right\} \\
&= \exp\left\{\frac{(\mu_2 - \mu_1)\sum_{i=1}^n x_i}{\sigma_0^2} + \frac{n}{2\sigma_0^2}(\mu_1^2 - \mu_2^2)\right\} \\
&= \exp\left\{\frac{\mu_2 - \mu_1}{\sigma_0^2} n\left(\frac{T\sigma_0}{\sqrt{n}} + \mu_0\right) + \frac{n}{2\sigma_0^2}(\mu_1^2 - \mu_2^2)\right\},
\end{aligned}$$

wobei wir beim letzten Übergang die Identität $\sum_{i=1}^n x_i = n(\frac{T\sigma_0}{\sqrt{n}} + \mu_0)$ benutzt haben. Die rechte Seite ist eine monoton steigende Funktion von T, folglich liegen monotone Dichtequotienten vor.

Aufgabe 10.6.7. Seien $X_1, \ldots, X_n \sim \mathrm{N}(0, \sigma^2)$ unabhängig, wobei der Erwartungswert gleich 0 ist und die Varianz $\sigma^2 > 0$ unbekannt sei. Zeigen Sie, dass die Bedingung der monotonen Dichtequotienten mit $T(x_1, \ldots, x_n) = \sum_{i=1}^n x_i^2$ gilt und konstruieren Sie den gleichmäßig besten Test zum Niveau α für $H_0 : \sigma^2 \le \sigma_0^2$ gegen $H_1 : \sigma^2 > \sigma_0^2$.

Aufgabe 10.6.8. Seien $X_1, \ldots, X_n \sim \mathrm{Poi}(\theta)$ unabhängig, wobei $\theta > 0$. Zeigen Sie, dass die Bedingung der monotonen Dichtequetienten mit $T(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$ gilt und beschreiben Sie den gleichmäßig besten Test von $H_0 : \theta \le \theta_0$ gegen $H_1 : \theta > \theta_0$ zum Niveau α.

10.7. Verallgemeinerter Likelihood–Quotienten–Test

Für einfache Hypothesen ist der Likelihood–Quotienten–Test ein gleichmäßig bester Test. Für nicht-einfache Hypothesen lässt sich der LQ–Test wie folgt verallgemeinern.

Annahme: $(\mathbb{P}_\theta)_{\theta\in\Theta}$ ist eine dominierte Familie von Wahrscheinlichkeitsmaßen auf dem Stichprobenraum $(\mathfrak{X}, \mathcal{A})$, d.h. es gibt ein σ-endliches Maß λ auf $(\mathfrak{X}, \mathcal{A})$, sodass für jedes $\theta \in \Theta$ das Wahrscheinlichkeitsmaß $\mathbb{P}_\theta$ eine Dichte h_θ bzgl. λ besitzt.

Wir betrachten die Hypothesen

$$H_0 : \theta \in \Theta_0 \quad \text{und} \quad H_1 : \theta \in \Theta_1,$$

wobei $\Theta = \Theta_0 \cup \Theta_1$ eine disjunkte Zerlegung des Parameterraumes Θ sei.

Definition 10.7.1. Die *verallgemeinerte Likelihood–Quotienten–Statistik* ist definiert durch
$$\Lambda(x) = \frac{\sup_{\theta \in \Theta_0} h_\theta(x)}{\sup_{\theta \in \Theta} h_\theta(x)} \in [0,1], \quad x \in \mathfrak{X}.$$

Kleine Werte von Λ sprechen für eine Ablehnung von H_0.

Definition 10.7.2. Der *verallgemeinerte LQ–Test* ist definiert durch
$$\varphi(x) = \begin{cases} 1, & \text{falls } \Lambda(x) \le c, \\ 0, & \text{falls } \Lambda(x) > c, \end{cases}$$

Dabei soll die Wahl von $c \in [0,1]$ sicherstellen, dass der Test Niveau α besitzt, d.h.
$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta \varphi(X) = \alpha.$$

Beispiel 10.7.3 (Zweiseitiger Student–t–Test als verallgemeinerter LQ–Test)**.** Betrachte unabhängige normalverteilte Beobachtungen $X_1, \ldots, X_n \sim \mathrm{N}(\mu, \sigma^2)$. Der Stichprobenraum ist $\mathfrak{X} = \mathbb{R}^n$ und der Parameterraum ist eine Halbebene:
$$\Theta = \{(\mu, \sigma^2) \colon \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

Wir betrachten die Hypothesen $H_0 : \mu = \mu_0$ und $H_1 : \mu \neq \mu_0$ mit einem vorgegebenen $\mu_0 \in \mathbb{R}$ und einem unbekanntem σ^2, d.h.
$$\Theta_0 = \{(\mu_0, \sigma^2) : \sigma^2 > 0\}, \quad \Theta_1 = \{(\mu, \sigma^2) \colon \mu \neq \mu_0, \sigma^2 > 0\}.$$

Die Dichte von $\mathbb{P}_{\mu,\sigma^2}$ bzgl. des Lebesgue–Maßes auf $\mathbb{R}^n$ ist
$$h_{\mu,\sigma^2}(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}.$$

Um das Supremum der Likelihood–Funktion $h_{\mu,\sigma^2}(x_1, \ldots)$ über Θ bzw. Θ_0 zu bestimmen, erinnern wir uns an die Maximum–Likelihood–Schätzer
$$\operatorname*{arg\,max}_{(\mu,\sigma^2) \in \Theta} h_{\mu,\sigma^2}(x_1, \ldots, x_n) = \left(\bar{x}_n, \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right)$$

und
$$\operatorname*{arg\,max}_{(\mu,\sigma^2) \in \Theta_0} h_{\mu,\sigma^2}(x_1, \ldots, x_n) = \left(\mu_0, \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2\right).$$

Einsetzen der optimalen Werte von (μ, σ^2) in die Dichte h_{μ,σ^2} ergibt
$$\sup_{(\mu,\sigma^2) \in \Theta} h_{\mu,\sigma^2}(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2\right)^{-n/2} e^{-n/2}.$$

bzw.

$$\sup_{(\mu,\sigma^2)\in\Theta_0} h_{\mu,\sigma^2}(x_1,\dots,x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{n}\sum_{i=1}(x_i-\mu_0)^2\right)^{-n/2} e^{-n/2}.$$

Die Likelihood–Quotienten–Statistik ergibt sich somit zu

$$\Lambda(x_1,\dots,x_n) = \left(\frac{\sum_{i=1}(x_i-\bar{x}_n)^2}{\sum_{i=1}(x_i-\mu_0)^2}\right)^{n/2}.$$

Um den Zusammenhang zur Student–t-Statistiki herzustellen, benutzen wir die Steiner–Formel $\sum_{i=1}(x_i-\mu_0)^2 = \sum_{i=1}^n(x_i-\bar{x}_n)^2 + n(\bar{x}_n-\mu_0)^2$ und schreiben Λ wie folgt um:

$$\Lambda(x_1,\dots,x_n) = \left(\frac{1}{1+\frac{n(\bar{x}_n-\mu_0)^2}{\sum_{i=1}(x_i-\bar{x}_n)^2}}\right)^{n/2} = \left(\frac{1}{1+T^2/(n-1)}\right)^{n/2},$$

was eine monoton fallende Funktion von $|T|$ ist, wobei

$$T(x_1,\dots,x_n) := \sqrt{n}\,\frac{\bar{x}_n-\mu_0}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n(x_i-\bar{x}_n)^2}}$$

die Student–t–Statistik bezeichnet. Die Bedingung $\Lambda(x_1,\dots,x_n)\le c$ ist äquivalent zur Bedingung $|T(x_1,\dots,x_n)| > c'$ für ein passendes c'. Wählt man c bzw. c' so, dass beide Tests das gleiche Signifikanzniveau α besitzen, so treffen die beiden Tests die gleiche Entscheidung.

10.8. Asymptotische Tests für die Erfolgswahrscheinlichkeit bei Bernoulli–Experimenten

Manchmal ist es nicht möglich oder schwierig, einen exakten Test zum Niveau α zu konstruieren. In diesem Fall kann man versuchen, einen Test zu konstruieren, der zumindest approximativ (bei großem Stichprobenumfang n) das Niveau α erreicht. Wir werden nun die entsprechende Definition einführen. Seien $X_1, X_2, \dots$ unabhängige und identisch verteilte Zufallsvariablen mit Dichte bzw. Zähldichte h_θ, wobei $\theta\in\Theta$. Es sei außerdem eine Zerlegung des Parameterraumes Θ in zwei disjunkte Teilmengen Θ_0 und Θ_1 gegeben:

$$\Theta = \Theta_0\cup\Theta_1, \quad \Theta_0\cap\Theta_1 = \varnothing.$$

Wir wollen die Nullhypothese $H_0:\theta\in\Theta_0$ gegen die Alternativhypothese $H_1:\theta\in\Theta_1$ testen.

Definition 10.8.1. Eine Folge von Borel–Funktionen $\varphi_1,\varphi_2,\dots$ mit $\varphi_n:\mathbb{R}^n\to\{0,1\}$ heißt *asymptotischer Test* zum Niveau $\alpha\in(0,1)$, falls

$$\limsup_{n\to\infty}\sup_{\theta\in\Theta_0}\mathbb{P}_\theta[\varphi_n(X_1,\dots,X_n)=1]\le\alpha.$$

Dabei ist φ_n die zum Stichprobenumfang n gehörende Entscheidungsregel.

Wir werden nun asymptotische Tests für die Erfolgswahrscheinlichkeit θ bei Bernoulli–Experimenten konstruieren. Seien $X_1,\dots,X_n$ unabhängige und mit Parameter $\theta\in(0,1)$

Bernoulli–verteilte Zufallsvariablen. Wir wollen verschiedene Hypothesen über den Parameter θ testen, z. B. $\theta = \theta_0$, $\theta \geq \theta_0$ oder $\theta \leq \theta_0$. Ein natürlicher Schätzer für θ ist $\bar{X}_n$. Wir betrachten die Teststatistik

$$T_n := \sqrt{n}\,\frac{\bar{X}_n - \theta_0}{\sqrt{\theta_0(1-\theta_0)}}.$$

Unter der Hypothese $\theta = \theta_0$ gilt nach dem zentralen Grenzwertsatz

$$T_n \xrightarrow[n\to\infty]{d} \mathrm{N}(0,1).$$

Wir betrachten nun drei verschiedene Fälle.

Fall A. $H_0 : \theta = \theta_0$; $H_1 : \theta \neq \theta_0$. In diesem Fall sollte H_0 verworfen werden, wenn $|T_n|$ groß ist. Entscheidungsregel: H_0 wird verworfen, wenn $|T_n| \geq z_{1-\frac{\alpha}{2}}$.

Fall B. $H_0 : \theta \geq \theta_0$; $H_1 : \theta < \theta_0$. Die Nullhypothese H_0 sollte verworfen werden, wenn T_n klein ist. Entscheidungsregel: H_0 wird verworfen, wenn $T_n \leq z_\alpha$.

Fall C. $H_0 : \theta \leq \theta_0$; $H_1 : \theta > \theta_0$. Die Nullhypothese H_0 sollte verworfen werden, wenn T_n groß ist. Entscheidungsregel: H_0 wird verworfen, wenn $T_n \geq z_{1-\alpha}$.

Nun betrachten wir ein Zweistichprobenproblem, bei dem zwei Parameter θ_1 und θ_2 von zwei Bernoulli–verteilten Stichproben verglichen werden sollen. Wir machen folgende Annahmen:

(1) $X_1, \ldots, X_n, Y_1, \ldots, Y_m$ sind unabhängige Zufallsvariablen.
(2) $X_1, \ldots, X_n \sim \mathrm{Bern}(\theta_1)$.
(3) $Y_1, \ldots, Y_m \sim \mathrm{Bern}(\theta_2)$.

Es sollen nun Hypothesen über die Erfolgswahrscheinlichkeiten θ_1 und θ_2 getestet werden, z. B. $\theta_1 = \theta_2$, $\theta_1 \geq \theta_2$ oder $\theta_1 \leq \theta_2$. Ein natürlicher Schätzer für $\theta_1 - \theta_2$ ist $\bar{X}_n - \bar{Y}_m$. Wir definieren uns die Größe

$$\tilde{T}_{n,m} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{\theta_1(1-\theta_1)}{n} + \frac{\theta_2(1-\theta_2)}{m}}}.$$

Satz 10.8.2. Unter $\theta := \theta_1 = \theta_2$ gilt

$$\tilde{T}_{n,m} \xrightarrow{d} \mathrm{N}(0,1) \text{ für } n, m \to \infty.$$

Beweis. Wir haben die Darstellung $\tilde{T}_{n,m} = Z_{1;n,m} + \ldots + Z_{n+m;n,m}$, wobei

$$Z_{k;n,m} = \begin{cases} \frac{X_k - \theta}{n\sqrt{\theta(1-\theta)}\sqrt{\frac{1}{n}+\frac{1}{m}}}, & \text{falls } k = 1, \ldots, n, \\ -\frac{Y_{k-n} - \theta}{m\sqrt{\theta(1-\theta)}\sqrt{\frac{1}{n}+\frac{1}{m}}}, & \text{falls } k = n+1, \ldots, n+m. \end{cases}$$

Wir wollen den zentralen Grenzwertsatz von Ljapunow verwenden. Es gilt:

(1) Die Zufallsvariablen $Z_{1;n,m}, \ldots, Z_{n+m;n,m}$ sind unabhängig.
(2) $\mathbb{E}Z_{k;n,m} = 0$.
(3) $\sum_{k=1}^{n+m} \mathbb{E}Z_{k;n,m}^2 = 1$.

Die letzte Eigenschaft kann man folgendermaßen beweisen:

$$\sum_{k=1}^{n+m} \mathbb{E} Z_{k;n,m}^2 = \frac{1}{\theta(1-\theta)\left(\frac{1}{n}+\frac{1}{m}\right)} \left(n \cdot \frac{\theta(1-\theta)}{n^2} + m \cdot \frac{\theta(1-\theta)}{m^2} \right) = 1.$$

Wir müssen also nur noch die Ljapunow–Bedingung überprüfen. Sei $\delta > 0$ beliebig. Es gilt

$$\begin{aligned}
\sum_{k=1}^{n+m} \mathbb{E}|Z_{k;n,m}|^{2+\delta} &= \frac{1}{\sqrt{\theta(1-\theta)}^{2+\delta}\left(\frac{1}{n}+\frac{1}{m}\right)^{\frac{2+\delta}{2}}} \left\{ \frac{n}{n^{2+\delta}} \mathbb{E}|X_1-\theta|^{2+\delta} + \frac{m}{m^{2+\delta}} \mathbb{E}|Y_1-\theta|^{2+\delta} \right\} \\
&\leq \frac{C(\theta)}{\left(\frac{1}{n}+\frac{1}{m}\right)^{\frac{2+\delta}{2}}} \left\{ \frac{1}{n^{1+\delta}} + \frac{1}{m^{1+\delta}} \right\} \\
&= \frac{C(\theta)}{n^{\frac{\delta}{2}}\left(1+\frac{n}{m}\right)^{\frac{2+\delta}{2}}} + \frac{C(\theta)}{m^{\frac{\delta}{2}}\left(\frac{m}{n}+1\right)^{\frac{2+\delta}{2}}},
\end{aligned}$$

was für $n, m \to \infty$ gegen 0 konvergiert. Dabei ist $C(\theta)$ eine von n, m unabhängige Größe. Nach dem zentralen Grenzwertsatz von Ljapunow folgt die Behauptung des Satzes. □

Die Größe $\tilde{T}_{n,m}$ konvergiert zwar gegen die Standardnormalverteilung, wir können diese Größe allerdings nicht direkt zur Konstruktion von asymptotischen Tests verwenden, denn $\tilde{T}_{n,m}$ beinhaltet die unbekannten Parameter θ_1 und θ_2. Deshalb betrachten wir eine Modifizierung von $\tilde{T}_{n,m}$, in der θ_1 und θ_2 durch die entsprechenden Schätzer $\bar{X}_n$ und $\bar{Y}_m$ ersetzt wurden:

$$T_{n,m} = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n} + \frac{\bar{Y}_m(1-\bar{Y}_m)}{m}}}.$$

Nach dem Gesetz der großen Zahlen gilt $\bar{X}_n \to \theta_1$ und $\bar{Y}_m \to \theta_2$ fast sicher für $n, m \to \infty$. Aus dem Satz von Slutsky kann man dann herleiten (Übungsaufgabe), dass

$$T_{n,m} \xrightarrow{d} \mathrm{N}(0,1) \text{ für } n, m \to \infty.$$

Wir betrachten nun drei verschiedene Nullhypothesen.

Fall A. $H_0 : \theta_1 = \theta_2$; $H_1 : \theta_1 \neq \theta_2$. In diesem Fall sollte H_0 verworfen werden, wenn $|T_{n,m}|$ groß ist. Entscheidungsregel: H_0 wird verworfen, wenn $|T_{n,m}| \geq z_{1-\frac{\alpha}{2}}$.

Fall B. $H_0 : \theta_1 \geq \theta_2$; $H_1 : \theta_1 < \theta_2$. Die Nullhypothese H_0 sollte verworfen werden, wenn $T_{n,m}$ klein ist. Entscheidungsregel: H_0 wird verworfen, wenn $T_{n,m} \leq z_\alpha$.

Fall C. $H_0 : \theta_1 \leq \theta_2$; $H_1 : \theta_1 > \theta_2$. Die Nullhypothese H_0 sollte verworfen werden, wenn $T_{n,m}$ groß ist. Entscheidungsregel: H_0 wird verworfen, wenn $T_{n,m} \geq z_{1-\alpha}$.

10.9. Pearson–χ^2–Test

Beispiel 10.9.1. In seinen klassischen Versuchen hat Gregor Mendel Vererbung von Merkmalen bei Erbsen untersucht. In einem Experiment aus dem Jahre 1865 züchtete er 556 Erbsen der folgenden 4 Typen:

- 315 rund und gelb,
- 101 kantig und gelb,
- 108 rund und grün,

- 32 kantig und grün.

Theoretisch sollten diese Typen im Verhältnis $9:3:3:1$ stehen. Die theoretischen Erwartungswerte der 4 Typen sind $312.75, 104.25, 104.25, 34.75$, was von Mendels Werten abweicht. Sind diese Abweichungen nun so groß, dass Mendelsche Theorie verworfen werden kann? Diese Frage kann man mit dem Pearson–χ^2–Test beantworten.

Wir beginnen mit der Definition der Multinomialverteilung. Man betrachte n Bälle, die unabhängg voneinander in Behälter geworfen werden. Die Anzahl der Behälter sei d und die Wahrscheinlichkeit, dass ein Ball im Behälter $i \in \{1, \ldots, d\}$ landet, sei $p_i \geq 0$, wobei $p_1 + \ldots + p_d = 1$. Bezeichnen wir mit X_i die Anzahl der Bälle, die im Behälter i landen, so ist der Zufallsvektor $(X_1, \ldots, X_d)$ multinomialverteilt mit Parametern $(n; p_1, \ldots, p_d)$. Es gilt

$$\mathbb{P}[X_1 = x_1, \ldots, X_d = x_d] = \binom{n}{x_1, \ldots, x_d} p_1^{x_1} \ldots p_d^{x_d}$$

für alle $x_1, \ldots, x_d \in \mathbb{N}_0$ mit $x_1 + \ldots + x_d = n$. Notation: $(X_1, \ldots, X_d) \sim \text{Mult}(n; p_1, \ldots, p_d)$.

Aufgabe 10.9.2. Zeigen Sie, dass für die Marginalverteilungen $X_i \sim \text{Bin}(n, p_i)$ gilt.

Beispiel 10.9.3. Wir werfen einen fairen Würfel n Mal. Es sei X_1 die Anzahl der Einsen, X_2 die Anzahl der Zweien, usw. Dann ist $(X_1, \ldots, X_6)$ multinomialverteilt mit Parametern $(n; \frac{1}{6}, \ldots, \frac{1}{6})$.

Mit dem Pearson–χ^2–test kann man Hypothesen über die Parameter $(p_1, \ldots, p_d)$ testen. Wir betrachten eine multinomialverteilte Stichprobe

$$(X_1, \ldots, X_d) \sim \text{Mult}(n; p_1, \ldots, p_d),$$

wobei n, d bekannt und $p_1, \ldots, p_d$ unbekannt seien. Für einen vorgegebenen Wahrscheinlichkeitsvektor $(p_1^*, \ldots, p_d^*)$ betrachten wir die Hypothesen

$$H_0 : (p_1, \ldots, p_d) = (p_1^*, \ldots, p_d^*) \text{ und } H_1 : (p_1, \ldots, p_d) \neq (p_1^*, \ldots, p_d^*).$$

Wir stellen das dazugehörige statistische Modell auf. Der Stichprobenraum ist die endliche Menge

$$\mathfrak{X} = \{(x_1, \ldots, x_d) \in \mathbb{N}_0^d \colon x_1 + \ldots + x_d = n\}.$$

Der Parameterraum ist ein Simplex

$$\Theta = \{(p_1, \ldots, p_d) : p_1, \ldots, p_d \geq 0,\ p_1 + \ldots + p_d = 1\}.$$

Für ein $(p_1, \ldots, p_d) \in \Theta$ ist $\mathbb{P}_{p_1, \ldots, p_d}$ ein Wahrscheinlichkeitsmaß auf $\mathfrak{X}$ mit

$$\mathbb{P}_{p_1, \ldots, p_d}[A] = \sum_{(x_1, \ldots, x_d) \in A} \binom{n}{x_1, \ldots, x_d} p_1^{x_1} \ldots p_d^{x_d}, \quad A \subset \mathfrak{X}.$$

Um die oben formulierte Hypothese H_0 zu testen, werden wir die quadratischen Abweichungen der beobachteten Werte $x_1, \ldots, x_d$ von den erwarteten Werten $np_1^*, \ldots, np_d^*$ mit speziellen Gewichten summieren.

Definition 10.9.4. Die *Pearson–Statistik* ist definiert durch

$$T_n(x_1, \ldots, x_d) = \sum_{i=1}^{d} \frac{(x_i - np_i^*)^2}{np_i^*}.$$

Der Pearson–χ^2–Test verwirft H_0, wenn T_n größer als ein kritischer Wert ist. Um den kritischen Wert zu bestimmen, müssen wir die Verteilung von T_n unter der Nullhypothese kennen.

Satz 10.9.5 (Pearson). Unter H_0 gilt $T_n \xrightarrow[n\to\infty]{d} \chi^2_{d-1}$.

Also verwerfen wir H_0, wenn $T_n > \chi^2_{d-1,1-\alpha}$. Dies ist ein asymptotischer Test zum Niveau α, denn nach dem Satz von Pearson gilt für die Wahrscheinlichkeit eines Fehlers 1. Art

$$\lim_{n\to\infty} \mathbb{P}_{H_0}[T_n > \chi^2_{d-1,1-\alpha}] = \alpha.$$

Bemerkung 10.9.6. Man beachte, dass die Anzahl der Freiheitsgrade der χ^2–Verteilung gleich $d-1$ und nicht d ist. Ein Freiheitsgrad ist durch die Relation $\sum_{i=1}^{d}(x_i - np_i^*) = n - n = 0$ verlorengegangen.

Beweis von Satz 10.9.5. Seien $\xi_1, \xi_2, \ldots$ unabhängige identisch verteilte Zufallsvariablen mit Werten in der Menge $\{1, \ldots, d\}$ und Wahrscheinlichkeiten

$$\mathbb{P}[\xi_k = 1] = p_1^*, \quad \ldots, \quad \mathbb{P}[\xi_k = d] = p_d^*, \quad k \in \mathbb{N}.$$

Wir können ξ_k als die Nummer des Behälters interpretieren, in dem der k-te Ball landet. Somit ist der Vektor $(X_1, \ldots, X_d)$ mit

$$X_1 = \sum_{k=1}^{n} \mathbb{1}_{\{\xi_k = 1\}}, \quad \ldots, \quad X_d = \sum_{k=1}^{n} \mathbb{1}_{\{\xi_k = d\}}.$$

multinomialverteilt mit Parametern $(n; p_1^*, \ldots, p_d^*)$. Wir wollen zeigen, dass

$$T_n(X_1, \ldots, X_d) \xrightarrow[n\to\infty]{d} \chi^2_{d-1}.$$

Schritt 1: Kovarianzmatrix. Definitionsgemäß gilt

$$(X_1, \ldots, X_d) = \sum_{k=1}^{n} \left(\mathbb{1}_{\{\xi_k=1\}}, \ldots, \mathbb{1}_{\{\xi_k=d\}}\right).$$

Also ist $(X_1, \ldots, X_d)$ eine Summe von n unabhängigen identisch verteilten Zufallsvektoren. Um auf diese Summe den multidimensionalen zentralen Grenzwertsatz anzuwenden, müssen wir den Erwartungswert und die Kovarianzmatrix der Summanden ausrechnen. Für alle $i \in \{1, \ldots, d\}$ gilt

$$\mathbb{E}\mathbb{1}_{\{\xi_1 = i\}} = \mathbb{P}[\xi_1 = i] = p_i^*.$$

Außerdem gilt für alle $i, j \in \{1, \ldots, d\}$, dass

$$\begin{aligned}\operatorname{Cov}\left(\mathbb{1}_{\{\xi_1=i\}}, \mathbb{1}_{\{\xi_1=j\}}\right) &= \mathbb{P}[\xi_1 = i, \xi_1 = j] - \mathbb{P}[\xi_1 = i]\mathbb{P}[\xi_1 = j] \\ &= \begin{cases} -p_i^* p_j^*, & \text{falls } i \neq j, \\ p_i^*(1 - p_i^*), & \text{falls } i = j. \end{cases}\end{aligned}$$

wobei wir beim letzten Übergang benutzt haben, dass Ereignisse $\{\xi_1 = i\}$ und $\{\xi_1 = j\}$ für $i \neq j$ disjunkt sind.

Schritt 2: Zentraler Grenzwertsatz. Mit dem $(d-1)$-dimensionalen zentralen Grenzwertsatz ergibt sich

$$U_n := \left(\frac{X_1 - np_1^*}{\sqrt{n}}, \ldots, \frac{X_{d-1} - np_{d-1}^*}{\sqrt{n}} \right) \xrightarrow[n\to\infty]{d} (Z_1, \ldots, Z_{d-1}),$$

wobei $Z = (Z_1, \ldots, Z_{d-1})$ ein $(d-1)$-dimensionaler Gauß–verteilter Zufallsvektor ist mit $\mathbb{E}Z_1 = \ldots = \mathbb{E}Z_{d-1} = 0$ und der Kovarianzmatrix

$$r_{ij} := \operatorname{Cov}(Z_i, Z_j) = \begin{cases} -p_i^* p_j^*, & \text{falls } i \neq j, \\ p_i^*(1 - p_i^*), & \text{falls } i = j. \end{cases}$$

Es sei bemerkt, dass wir die letzte Koordinate weggelassen haben, denn sonst würden wir wegen der Relation $X_1 + \ldots + X_d = n$ im Grenzwert einen Gauß–Vektor $(Z_1, \ldots, Z_d)$ mit $Z_1 + \ldots + Z_d = 0$ bekommen. Dieser Vektor ist degeneriert, wir müssen aber im nächsten Schritt die Kovarianzmatrix invertieren.

Schritt 3: Normierung auf die Standardnormalverteilung. Das Inverse der Kovarianzmatrix $\Sigma = (r_{ij})_{1\leq i,j\leq d-1}$ berechnet sich zu (Übung)

$$\Sigma^{-1} = (s_{ij})_{1\leq i,j\leq d-1} \quad \text{mit} \quad s_{ij} = \begin{cases} 1/p_d^*, & \text{falls } i \neq j, \\ 1/p_d^* + 1/p_j^*, & \text{falls } i = j. \end{cases}$$

Der Zufallsvektor $\Sigma^{-1/2}Z$ ist standardnormalverteilt auf $\mathbb{R}^{d-1}$. Mit dem Satz von der stetigen Abbildung ergibt sich, dass

$$\Sigma^{-1/2}U_n = \Sigma^{-1/2} \left(\frac{X_1 - np_1^*}{\sqrt{n}}, \ldots, \frac{X_{d-1} - np_{d-1}^*}{\sqrt{n}} \right) \xrightarrow[n\to\infty]{d} (N_1, \ldots, N_{d-1}),$$

wobei $(N_1, \ldots, N_{d-1})$ standardnormalverteilt auf $\mathbb{R}^{d-1}$ ist. Durch nochmalige Anwendung desselben Satzes folgt

$$\left(\Sigma^{-1/2}U_n\right)^\top \left(\Sigma^{-1/2}U_n\right) \xrightarrow[n\to\infty]{d} N_1^2 + \ldots + N_{d-1}^2 \sim \chi_{d-1}^2.$$

Auf der anderen Seite gilt

$$\begin{aligned}
\left(\Sigma^{-1/2}U_n\right)^\top \left(\Sigma^{-1/2}U_n\right) &= U_n^\top \Sigma^{-1} U_n \\
&= \sum_{i=1}^{d-1}\sum_{j=1}^{d-1} s_{ij} \left(\frac{X_i - np_i^*}{\sqrt{n}}\right)\left(\frac{X_j - np_j^*}{\sqrt{n}}\right) \\
&= \sum_{j=1}^{d-1} \frac{1}{p_j^*}\frac{(X_j - np_j^*)^2}{n} + \sum_{i=1}^{d-1}\sum_{j=1}^{d-1} \frac{1}{p_d^*}\frac{(X_i - np_i^*)(X_j - np_j^*)}{n} \\
&= \sum_{j=1}^{d-1} \frac{(X_j - np_j^*)^2}{np_j^*} + \frac{1}{np_d^*}\left(\sum_{i=1}^{d-1}(X_i - np_i^*)\right)^2 \\
&= \sum_{j=1}^{d} \frac{(X_j - np_j^*)^2}{np_j^*} \\
&= T_n(X_1, \ldots, X_d),
\end{aligned}$$

wobei wir beim letzten Übergang die Relation $\sum_{i=1}^d (X_i - np_i^*) = 0$ benutzt haben. Fasst man alles zusammen, so ergibt sich die zu beweisende Aussage

$$T_n(X_1, \ldots, X_d) = \left(\Sigma^{-1/2}U_n\right)^\top \left(\Sigma^{-1/2}U_n\right) \xrightarrow[n\to\infty]{d} \chi^2_{d-1}.$$

□

Aufgabe 10.9.7. Zeigen Sie, dass $T_n(x_1, \ldots, x_d) = \sum_{i=1}^d \frac{x_i^2}{np_i^*} - n$.

Der nächste Satz besagt, dass der Pearson–Test jeden festen Parameterwert aus der Alternativhypothese mit einer für $n \to \infty$ gegen 1 konvergierenden Wahrscheinlichkeit erkennt.

Satz 10.9.8 (Punktweise Konsistenz des Pearson–χ^2–Tests)**.** Für eine Stichprobe

$$(X_1, \ldots, X_d) \sim \text{Mult}(n; p_1, \ldots, p_d)$$

betrachten wir die Hypothesen

$$H_0 : (p_1, \ldots, p_d) = (p_1^*, \ldots, p_d^*), \quad \tilde{H}_1 : (p_1, \ldots, p_d) = (p_1', \ldots, p_d'),$$

wobei $(p_1^*, \ldots, p_d^*) \neq (p_1', \ldots, p_d')$ vorgegeben seien. Dann gilt

$$\mathbb{P}_{\tilde{H}_1}[T_n(X_1, \ldots, X_d) > \chi^2_{d-1,1-\alpha}] = 1.$$

Beweis. Nach Voraussetzung gibt es mindestens ein $i \in \{1, \ldots, d\}$ mit $p_i' \neq p_i^*$. Unter $\tilde{H}_1$ gilt nach dem starken Gesetz der großen Zahlen

$$\frac{X_i}{n} = \frac{1}{n}\sum_{k=1}^n \mathbb{1}_{\{\xi_k = i\}} \xrightarrow[n\to\infty]{f.s.} p_i' \neq p_i^*.$$

Es folgt, dass unter $\tilde{H}_1$

$$T_n = T_n(X_1, \dots, X_d) \geq \frac{(X_i - np_i^*)^2}{np_i^*} = \frac{n}{p_i^*}\left(\frac{X_i}{n} - p_i^*\right)^2 \xrightarrow[n\to\infty]{f.s.} +\infty.$$

Daraus folgt, dass T_n gegen $+\infty$ auch in Wahrscheinlichkeit konvergiert, d.h. $\lim_{n\to\infty} \mathbb{P}_{H_1}[T_n > c] = 1$ für jedes feste $c \in \mathbb{R}$. □

Beispiel 10.9.9. Für den oben beschriebenen Versuch von Mendel mit $n = 556$ Erbsen lautet die Nullhypothese

$$(p_1^*, p_2^*, p_3^*, p_4^*) = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right).$$

Die von Mendel beobachteten Werte sind

$$(x_1, x_2, x_3, x_4) = (315, 101, 108, 32).$$

Der Wert der Pearson–Statistik berechnen sich zu $T_n(x_1, x_2, x_3, x_4) = 0.47$. Nach dem Satz von Pearson sollte T_n unter der Nullhypothese approximativ χ_3^2–verteilt sein. Das 0.95–Quantil der χ_3^2–Verteilung ist laut Tabelle $\chi_{3,0.95}^2 = 7.81$, was viel größer als der beobachtete Wert von T_n ist. Also kann die Nullhypothese nicht verworfen werden. Die Daten zeigen keine signifikante Abweichung von der Mendelschen Theorie.

Der p–Wert ist die Wahrscheinlichkeit, dass bei einer unabhängigen Wiederholung des Experiments ein Wert der Pearson–Statistik T_n beobachtet wird, der ≥ 0.47 ist, und beträgt in unserem Fall 0.92. Von 10 Biologen, die das Experiment von Mendel unabhängig wiederholen, würde im Durchschnitt nur einer eine bessere Übereinstimmung mit der Theorie beobachten, als Mendel. Auch in anderen Experimenten von Mendel war die Übereinstimmung mit den theoretischen Werten "zu gut". Aus diesem Grund warf Fisher in einer Arbeit aus dem Jahre 1936 Mendel vor, seine Ergebnisse beschönigt zu haben, was zur sogenannten Mendel–Fisher–Kontroverse führte. An dieser Stelle verweisen wir auf das Buch von A. Franklin, A. W. F. Edwards, D. J. Fairbanks, D. L. Hartl, and T. Seidenfeld, "*Ending the Mendel–Fisher Controversy*", Univ. Pittsburgh Press, 2008.

Beispiel 10.9.10. Man kann sich fragen, ob in der Dezimaldarstellung der Zahl

$$\pi = 3.141592653589793238\dots$$

alle 10 Ziffern ungefähr gleich oft vorkommen. Unter den ersten $n = 10^7$ Dezimalstellen von π finden sich so viele verschiedene Ziffern:

x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
999.440	999.333	1.000.306	999.965	1.001.093	1.000.466	999.337	1.000.206	999.814	1.000.040

Wir möchten nun die Hypothese, dass jede Ziffer mit einer Häufigkeit von 1/10 vorkommt, testen. Die Pearson–Statistik berechnet sich zu

$$T_n(x_0, \dots, x_9) = 2.7834.$$

Nach dem Satz von Pearson sollte T_n approximativ χ^2_9-verteilt sein. Das 0.95–Quantil der χ^2_9–Verteilung ist laut Tabelle $\chi^2_{9,0.95} = 16.92$, was viel größer als der beobachtete Wert von T_n ist. Somit können wir die Nullhypothese nicht verwerfen.

Wir berechnen noch den p–Wert des Pearson–Tests. Dieser ist definiert als die Wahrscheinlichkeit, dass eine χ^2_9–verteilte Zufallsvariable $\geq$ als der beobachtete Wert 2.7834 ist. Mit der entsprechenden Software erhält man einen p–Wert von 0.9723, was auffallend hoch ist!

Wie kann man nun diesen extrem kleinen Wert der Pearson–Statistik (bzw. den extrem hohen p–Wert) interpretieren? Die Abweichungen der beobachteten Werte $x_0, \ldots, x_9$ (s. die obige Tabelle) von dem erwarteten Wert 10^6 sind viel kleiner, als das, was man bei einer Folge von u.i.v. Zufallsvariablen mit Gleichverteilung auf $\{0, 1, \ldots, 9\}$ erwarten würde. Der kleine Wert der Pearson–Statistik ist ein Hinweis darauf, dass die von uns untersuchte Folge von Ziffern nicht zufällig ist.