Kapitel 4.3
Testen von Verteilungshypothesen

Beispiele

1) Landwirt wird Verfahren zur Änderung des Geschlechtsverhältnisses von Nutztieren angeboten, so dass z.B. mehr Kuhkälber als Stierkälber geboren werden. Um die Wirksamkeit des Verfahrens zu überprüfen, muss getestet werden, ob sich das Verhältnis tatsächlich ändert. Da ein Test aber lange dauert, sollte die Anzahl der Tests gering bleiben.

2) Faire Münze: Wie oft muss eine Münze geworfen, d.h. „getestet“ werden, um mit hinreichender Sicherheit sagen zu können, dass sie fair ist?

3) Qualitätskontrolle: Hypothese: „Ein Produkt ist besser als ein anderes.“

In allen Beispielen gibt es zwei Möglichkeiten: Die Annahme (Verhältnis ändert sich/Münze ist fair/Produkt ist besser) trifft zu oder nicht. Anhand eines Tests soll eine Entscheidung getroffen werden, welche der zwei Möglichkeiten tatsächlich vorliegt.

Mathematische Formulierung

Gegeben: Parametrisches Modell $(P_\vartheta)_{\vartheta\in\Theta}$, mit $\Theta \subseteq \mathbb{R}$ oder $\Theta \subseteq \mathbb{R}^n$. Beobachtungen $x_1, \dots, x_n$, die Realisierungen von u.i.v. nach P_ϑ verteilten ZVn $X_1, \dots, X_n$ seien, wobei ϑ unbekannt ist.

Aufgabe: Entscheide anhand der Beobachtungen, ob der unbekannte Parameter ϑ in $\Theta_0 \subset \Theta$ liegt oder im Komplement:

$$\text{Nullhypohese } H_0 : \vartheta \in \Theta_0$$
$$\text{Alternative } H_1 : \vartheta \in \Theta_1 = \Theta \setminus \Theta_0 = \Theta_0^c$$

Definition 4.15

Ein Test ist eine Entscheidungsregel der Form

$$E = e(X_1, \dots, X_n) = \begin{cases} 1, & \textit{für Annahme von } H_1 \textit{, d.h. Ablehnung von } H_0, \\ 0, & \textit{für Annahme von } H_0 \textit{, d.h. Ablehnung von } H_1, \end{cases}$$

die aufgrund der Beobachtungen $x_i = X_i(\omega)$, $1 \leq i \leq n$, eine Entscheidung $E(\omega) = e(x_1, \dots, x_n)$für oder gegen die Hypothese trifft.

Bemerkung:

a) Allgemeiner wird ein Test definiert als (messbare) Abbildung der Beobachtungen $e(X_1, \ldots, X_n) \to [0, 1]$, wobei $e(x_1, \ldots, x_n)$ die Wahrscheinlichkeit angibt, sich aufgrund der beobachteten Daten für die Alternative zu entscheiden.

   Ein Test ist somit eine Zufallsvariable mit Werten in $[0, 1]$.

   Wir betrachten im Folgenden nur Tests mit eindeutiger Entscheidung, d.h. $e(X_1, \ldots, X_n) \to \{0, 1\}$.

b) Ein Test ist in diesem Fall eindeutig charakterisiert durch seinen *kritischen Bereich*

$$K = \{\omega \in \Omega \mid E(\omega) = 1\}$$

   d.h. durch die Menge der Fälle, bei denen die Nullhypothese H_0 verworfen wird.

c) Sind $\Theta_0 = \{\vartheta_0\}$ oder $\Theta_1 = \{\vartheta_1\}$ einelementig, spricht man von *einfachen Hypothesen bzw. Alternativen*, andernfalls von zusammengesetzten Hypothesen/Alternativen.

Bei der Testentscheidung sind zwei Fehler/Irrtümer möglich:

Fehler 1. Art: Nullhypothese H_0 wird verworfen, obwohl sie zutrifft.
Fehler 2. Art: Nullhypothese H_0 wird nicht verworfen, obwohl sie falsch ist.

Definition 4.16

Ein Test hat das **(Signifikanz-)Niveau** α, *wenn gilt*

$$P_\vartheta(K) \leq \alpha \qquad \forall\, \vartheta \in \Theta_0.$$

Beachte, dass $P_\vartheta(K)$ für $\vartheta \in \Theta_0$ gerade die Wahrscheinlichkeit für einen Fehler 1. Art ist!

Typische Vorgehensweise: Man fixiert ein α (Wahrscheinlichkeit für Fehler 1. Art) und sucht dann unter allen Tests zum Niveau α denjenigen, der die Wahrscheinlichkeit für einen Fehler 2. Art minimiert, d.h. für den $P_\vartheta(K^c)$ minimal wird für $\vartheta \in \Theta_1$.

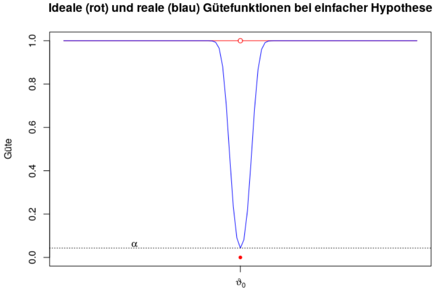
$\longrightarrow$ Asymmetrie bzgl. der beiden Fehler, nur einen kann man kontrollieren.

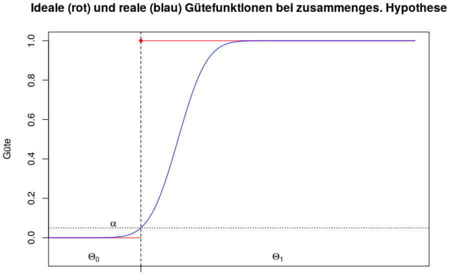
Definition 4.17

Die durch $\beta(\vartheta) = P_\vartheta(K)$ *definierte Funktion* $\beta : \Theta \to [0, 1]$ *heißt* **Gütefunktion** *des Tests.*

Wahrscheinlichkeit für Fehler 1. Art: $\beta(\vartheta),\ \vartheta \in \Theta_0$
Wahrscheinlichkeit für Fehler 2. Art: $1 - \beta(\vartheta),\ \vartheta \in \Theta_1$

Für $\vartheta \in \Theta_1$ heißt $\beta(\vartheta)$ die Macht (engl. power) oder auch Güte/Schärfe des Tests in ϑ.





Kälberbeispiel revisited (Beispiel 4.18)

Verfahren, das Geschlechtsverhältnis bei Kälbern verbessern soll:

$\Omega = \{\omega = (x_1, \ldots, x_n) \mid x_i \in \{0, 1\}\}$,
wobei 0 kein Erfolg (Stier) entspricht und 1 Erfolg (Kuh).

Annahme: $n = 20$, d.h. 20 Kälbergeburten werden beobachtet

P_ϑ sei Binomialverteilung mit $p = \vartheta$ und $n = 20$, $\Theta = \left[\frac{1}{2}, 1\right]$.

(Einfache) Hypothese: $\Theta_0 = \{\frac{1}{2}\}$ (kein Effekt)
Alternative: $\Theta_1 = \left(\frac{1}{2}, 1\right]$ (Verfahren wirkt)

(ebenso möglich wäre $\Theta = [0, 1]$, $\Theta_0 = \left[0, \frac{1}{2}\right]$, $\Theta_1 = \left(\frac{1}{2}, 1\right]$)

Testidee: Lehne Hypothese ab (d.h. Entscheid für Wirksamkeit und Kauf des Verfahrens), falls $\sum_{i=1}^{20} x_i > c$.

Problem: Wie ist c zu wählen, damit Effekt deutlich und Entscheidung klar wird?
Züchter will aus Kostengründen vermeiden, das Verfahren einzuführen, wenn es in Wahrheit wirkungslos ist (Fehler 1. Art), und entscheidet sich für das Niveau $\alpha = 0.05$.

Kälberbeispiel revisited (Beispiel 4.18)

Wahrscheinlichkeit für Fehler 1. Art ist

$$P_{\frac{1}{2}}\left[\sum_{i=1}^{20} X_i > c\right] = \sum_{k=c+1}^{20} \binom{20}{k} \left(\frac{1}{2}\right)^{20} \leq \alpha = 0.05$$

Man findet $P_{\frac{1}{2}}\left[\sum_{i=1}^{20} X_i > 14\right] = 0.0207,\ P_{\frac{1}{2}}\left[\sum_{i=1}^{20} X_i > 13\right] = 0.0577$

$$\Longrightarrow \quad c = 14,\ \text{d.h. } K = \left\{\omega \in \Omega \,\middle|\, \sum_{i=1}^{20} x_i > 14\right\}$$

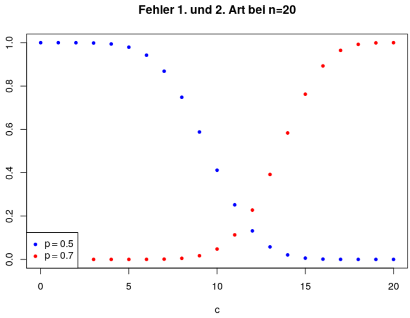
Betrachte nun den Fehler 2. Art: Angenommen, der Fall $\vartheta = 0.7$ sei bereits wirtschaftlich interessant. Die Macht $\beta(\vartheta)$ des Tests bei $\vartheta = 0.7$ ist

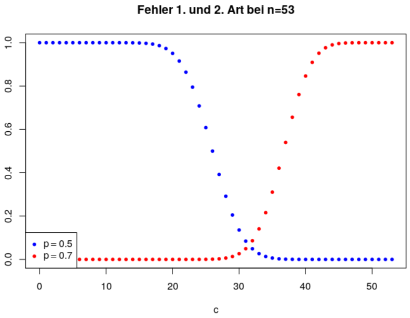
$$P_{0.7}(K) = \sum_{k=15}^{20} \binom{20}{k} \cdot 0.7^k \cdot 0.3^{20-k} = 0.417$$

d.h. mit Wahrscheinlichkeit $1 - \beta(0.7) = 0.583$ wird *nicht* entdeckt, dass das Verfahren gut funktioniert: unbefriedigend!

Kälberbeispiel revisited (Beispiel 4.18)

Konsequenz: Im vorliegenden Beispiel reichen Daten nicht für größere *Trennschärfe* aus! Um z.B. $P_{0.5}(K) \leq 0.05$ und $P_{0.7}(K) \geq 0.9$ zu erreichen, braucht man mindestens $n = 53$ Versuche (mit $c = 32$).





Allgemeine Konstruktion von Tests

Fall der einfachen Alternative: $\Theta_1 = \{\vartheta_1\}$
Wähle K so, dass $P_{\vartheta_1}(K)$ maximal wird (d.h. maximiere die Macht des Tests) unter der Nebenbedingung

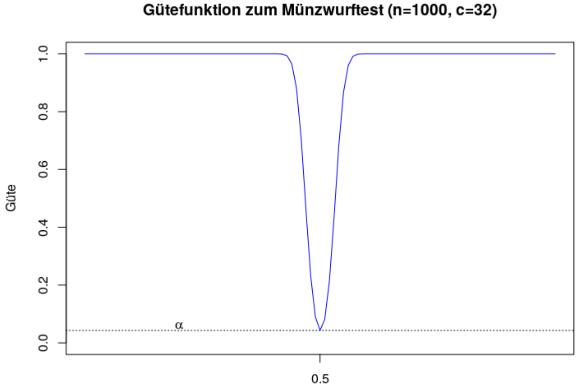
$$P_\vartheta(K) \leq \alpha \quad \forall\, \vartheta \in \Theta_0 \quad \text{(wahre Signifikanzniveau } \alpha\text{)}.$$

Fall der zusammengesetzten Alternative: $|\Theta_1| > 1$
Im Allgemeinen wird das optimale K von $\vartheta_1 \in \Theta_1$ abhängen. Ist das nicht der Fall, so heißt der durch das eindeutige K festgelegte Test *gleichmäßig bester Test (UMP-Test, uniformly most powerful) zum Niveau* α.

Beispiel 4.19 (Münzwurf): $\Theta = [0,1]$, $\Theta_0 = \left\{\frac{1}{2}\right\}$ (Nullhypothese: Münze ist fair). Plausibler Test ist

$$E(\omega) = \begin{cases} 1, & \text{falls } \left|\sum_{i=1}^n x_i - \frac{n}{2}\right| \geq c, \\ 0, & \text{sonst} \end{cases}$$

für einen bestimmten kritischen Wert c.



Es gilt $P_{0.5}(K) = 0.043 \leq 0.05$, d.h. der Test hält das Niveau $\alpha = 0.05$ ein.

Für ϑ mit $|\vartheta - \frac{1}{2}| \geq 0.06$ ist auch die Fehlerwahrscheinlichkeit 2. Art ≤ 0.05, d.h. der Test ist in diesem Sinne trennscharf für gebiete mit Abstand ≥ 0.06.

Kapitel 4.4
Likelihood-Quotienten-Tests

Grundannahmen

Grundraum $(\Omega, \mathcal{A}) = (\mathbb{R}^n, \mathcal{B}_n)$, $\mathfrak{P} = \{P_\vartheta \mid \vartheta \in \Theta\}$

Für diskrete Wahrscheinlichkeitsmaße sei

$$P_\vartheta(\{(x_1, \ldots, x_n)\}) = \mu_\vartheta(x_1) \cdot \ldots \cdot \mu_\vartheta(x_n),$$

im stetigen Fall habe P_ϑ eine Dichte

$$p((x_1, \ldots, x_n), \vartheta) = f(x_1, \vartheta) \cdot \ldots \cdot f(x_n, \vartheta).$$

Einfache Hypothese und Alternative
Sei $\Theta = \{\vartheta_0, \vartheta_1\}$, die Nullhypothese sei $\Theta_0 = \{\vartheta_0\}$.

Die *Likelihoodfunktion* ist gegeben durch

$$L((x_1, \ldots, x_n), \vartheta_i) = \begin{cases} f(x_1, \vartheta_i) \cdot \ldots \cdot f(x_n, \vartheta_i) & \text{im stetigen Fall,} \\ \mu_{\vartheta_i}(x_1) \cdot \ldots \cdot \mu_{\vartheta_i}(x_n) & \text{im diskreten Fall,} \end{cases} \quad i = 0, 1.$$

Der **Likelihood-Quotient** ist

$$\frac{L(x, \vartheta_1)}{L(x, \vartheta_0)}.$$

Neyman-Pearson-Lemma

Idee: Ein hoher Wert des Likelihood-Quatienten spricht für ϑ_1, d.h. Ablehnung der Hypothese, ein niedriger Wert für ϑ_0, d.h. Hypothese kann aufrecht erhalten werden.

Definition 4.20

Ein **Likelihood-Quotienten-Test** *(LQT) von ϑ_0 gegen ϑ_1 ist ein Test $E = e(X_!, \ldots, X_n)$ der Form*

$$e(X_1, \ldots, X_n) = \begin{cases} 1, & \text{falls } \frac{L((X_1,\ldots,X_n),\vartheta_1)}{L((X_1,\ldots,X_n),\vartheta_0)} \geq c, \\ 0, & \text{sonst}, \end{cases}$$

mit kritischem Bereich $K = \left\{ \frac{L((X_1,\ldots,X_n),\vartheta_1)}{L((X_1,\ldots,X_n),\vartheta_0)} \geq c \right\}$ und Signifikanzniveau

$$\alpha = P_{\vartheta_0}\left(\frac{L((X_1, \ldots, X_n), \vartheta_1)}{L((X_1, \ldots, X_n), \vartheta_0)} \geq c \right).$$

Neyman-Pearson-Lemma

Satz 4.21 (Lemma von Neyman-Pearson)

Jeder Likelihood-Quotienten-Test E ist im folgenden Sinne optimal: Ist $\tilde{E}$ ein weiterer Test mit demselben Signifikanzniveau wie E, so hat $\tilde{E}$ eine mindestens ebenso große Fehlerwahrscheinlichkeit 2. Art, d.h.

$$P_{\vartheta_0}(\tilde{K}) \leq P_{\vartheta_0}(K) \quad \Longrightarrow \quad P_{\vartheta_1}(\tilde{K}) \leq P_{\vartheta_1}(K).$$

Beispiel 4.22 (Nachrichtentechnik)
Über Nachrichtenkanal wird in n sukzessiven Perioden festes Signal S gesendet, und zwar $S = 1$, falls zu Beginn der Sendezeit ein bestimmtes Ereignis eingetreten ist („Notsignal"), und $S = 0$ sonst („Funkstille").
Problem: Signal wird durch Rauschen überlagert, d.h. der Empfänger hört tatsächlich die Signal $X_1, \ldots, X_n$ mit

$$X_i = S + Z_i, \quad i = 1, \ldots, n,$$

wobei die Z_i zufällige Störungen sind.

Annahme: „*weißes Rauschen*“, d.h. die Z_i sind u.i.v. nach $N(0, 1)$ verteilt.

Mögliche Signalfolgen	
1. Folge	2. Folge
1.022	-1.420
-0.472	0.489
1.279	-1.711
3.521	-1.186
0.571	0.754
-1.851	-0.732
0.194	-0.066
1.192	1.006
-0.501	-0.798
-0.273	0.162
4.682	-3.502

Aufgrund der empfangenen Signale soll entschieden werden, ob die Hypothese $S = 1$ aufrecht erhalten werden kann oder verworfen werden muss.

Wichtig ist, das Signal, falls es gesendet wird (Notfall), auch wirklich zu entdecken (d.h. die Wahrscheinlichkeit für einen Fehler 1. Art sollte klein sein). Empfangenes Signal ist $N(\vartheta, 1)$-verteilt, wobei $\vartheta = 0$ oder $\vartheta = 1$.

Hypothese: $S = 1$ wird gesendet, d.h. die X_i sind $N(1, 1)$-verteilt.
Alternative: $S = 0$, d.h. die X_i sind $N(0, 1)$-verteilt.

Der Likelihood-Quotient ist

$$\frac{L((x_1, \dots, x_n), 0)}{L(x_1, \dots, x_n), 1)} = \frac{e^{-\frac{1}{2}\sum_{i=1}^n x_i^2}}{e^{-\frac{1}{2}\sum_{i=1}^n (x_i-1)^2}} \geq c$$

Ausmultiplizieren des Nenners und Kürzen ergibt

$$e^{\frac{n}{2}-\sum_{i=1}^n x_i} \geq c \iff \frac{n}{2} - \sum_{i=1}^n x_i \geq \ln(c) \iff \sum_{i=1}^n x_i \leq k \quad \text{mit} \quad k = \frac{n}{2} - \ln(c),$$

d.h. $K = \left\{\sum_{i=1}^n X_i \leq k\right\}$.

Der Test hat das Niveau $\alpha = 0.05$, falls gilt

$$P_{\vartheta_0}\left[\sum_{i=1}^{n} X_i \leq k\right] = P_{\vartheta_0}\left[\frac{\sum_{i=1}^{n} X_i - n}{\sqrt{n}} \leq \frac{k-n}{\sqrt{n}}\right] = \Phi\left(\frac{k-n}{\sqrt{n}}\right) = \alpha.$$

Somit muss gelten

$$\frac{k-n}{\sqrt{n}} = \Phi^{-1}(\alpha) \approx -1.65 \quad \Longrightarrow \quad k = n - \sqrt{n} \cdot 1.65.$$

Konkret für $n = 10$ ergibt dies $k \approx 4.79$, d.h. $K = \{\sum_{i=1}^{10} X_i \leq 4.79\}$. Damit lehnt der Test für beide Datenreihen die Hypothese $S = 1$ ab, d.h. man geht aufgrund der empfangenen Signale von „Funkstille" aus.

Die Macht des Tests in $\vartheta_1 = 0$ ist

$$P_{\vartheta_1}(K) = P_{\vartheta_1}\left[\sum_{i=1}^{10} X_i \leq 4.79\right] = P\left[\frac{\sum_{i=1}^{10} X_i}{\sqrt{10}} \leq \frac{4.79}{\sqrt{10}}\right] = \Phi\left(\frac{4.79}{\sqrt{10}}\right) \approx 0.934,$$

d.h. die Fehlerwahrscheinlichkeit 2. Art ist $1 - P_{\vartheta_1}(K) \approx 0.066$, insofern ist der Test in Bezug auf beide Fehler recht befriedigend.

Plausibel wäre auch der folgende Test (der aber nach dem Neyman-Pearson-Lemma nicht besser sein kann!):

$$\tilde{K} = \{\max(X_1, \dots, X_n) \leq c\}$$

Dann ist

$$P_{\vartheta_0}(\tilde{K}) = \prod_{i=1}^{n} P_{\vartheta_0}[X_i \leq c] = \prod_{i=1}^{n} P_{\vartheta_0}[X_i - 1 \leq c - 1] = \big(\Phi(c-1)\big)^n \leq \alpha = 0.05,$$

falls $c - 1 \leq \Phi^{-1}\big(\alpha^{\frac{1}{n}}\big)$.
Speziell für $n = 10$ und $\alpha = 0.05$ ist $\alpha^{\frac{1}{10}} = 0.74$ und $\Phi^{-1}(0.74) = 0.64$, d.h. $c = 1.64$.
Damit würde die Hypothese $\vartheta = 1$ für die erste Datenreihe nicht verworfen, sondern nur bei der zweiten.
Die Macht dieses Tests an der Stelle $\vartheta_1 = 0$ ist

$$P_{\vartheta_1}(\tilde{K}) = \big(\Phi(c)\big)^{10} = (0.9495)^{10} \approx 0.6,$$

d.h. die Fehlerwahrscheinlichkeit 2. Art ist ungefähr $1 - 0.6 = 0.4$ und damit etwa sechsmal so groß wie beim Likelihood-Quotiententest!

Zusammengesetzte Hypothesen und Alternativen

Sei nun $|\Theta| > 2$ und wiederum Θ_0 die Hypothesenmenge sowie $\Theta_1 = \Theta \setminus \Theta_0$ die Alternativenmenge.
Der Likelihood-Quotient ist nun

$$\lambda(x_1, \dots, x_n) = \frac{\sup_{\vartheta \in \Theta} L((x_1, \dots, x_n), \vartheta)}{\sup_{\vartheta \in \Theta_0} L((x_1, \dots, x_n), \vartheta)}$$

Es gilt $\lambda(x_1, \dots, x_n) \geq 1$ Große Werte von λ lassen wiederum $\vartheta \in \Theta_1$ vermuten. Dies führt zu einem Test der Form $\{K = \{\lambda(X_1, \dots, X_n) \geq c\}$ mit $c > 1$.

Für $\Theta = \{\vartheta_0, \vartheta_1\}$ und $c > 1$ ist

$$\begin{aligned}
\lambda(x_1, \dots, x_n) &= \frac{\max\{L((x_1, \dots, x_n), \vartheta_0), L((x_1, \dots, x_n), \vartheta_1)\}}{L((x_1, \dots, x_n), \vartheta_0)} \geq c \\
\iff \quad & \frac{L((x_1, \dots, x_n), \vartheta_1)}{L((x_1, \dots, x_n), \vartheta_0)} \geq c
\end{aligned}$$

Beispiel 4.23: Nachrichtentechnik revisited
Es kann irgendein „Text“ $\vartheta = (\vartheta_1, \ldots, \vartheta_n)$, $\vartheta_i \in \mathbb{R}$, gesendet werden,
d.h. $\Theta = \mathbb{R}^n$.
Hypothese: $\vartheta_0 = (0, \ldots, 0)$ („Funkstille“), Alternative: $\vartheta \neq \vartheta_0$

$$\lambda(x_1, \ldots, x_n) = \frac{\sup_{\vartheta\in\Theta} L((x_1, \ldots, x_n), \vartheta)}{L((x_1, \ldots, x_n), \vartheta_0)}$$

$$\begin{aligned} L((x_1, \ldots, x_n), \vartheta) &= (2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\vartheta_i)^2} \text{ maximal für } \widehat{\vartheta} = (x_1, \ldots, x_n) \\ &\implies \lambda(x_1, \ldots, x_n) = \frac{(2\pi)^{-\frac{n}{2}}}{(2\pi)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2}} = e^{\frac{1}{2}\sum_{i=1}^n x_i^2} \geq c \\ &\implies \sum_{i=1}^n x_i^2 \geq \tilde{c} \quad \text{mit } \tilde{c} = 2\ln(c), \end{aligned}$$

d.h. der kritische Bereich ist $K = \{\sum_{i=1}^n X_i^2 \geq \tilde{c}\}$.
$\tilde{c}$ ist so zu wählen, dass $P_{\vartheta_0}(K) = P_{\vartheta_0}\left[\sum_{i=1}^n X_i^2 \geq \tilde{c}\right] \leq \alpha$.
$\sum_{i=1}^n X_i^2$ ist unter der Hypothese χ_n^2-verteilt $\implies \tilde{c} = q_{1-\alpha}^{\chi_n^2}$.

Verallgemeinerungen

Unter einigen weiteren technischen Voraussetzungen charakterisiert das allgemeine Neyman-Pearson-Lemma die in diesem Rahmen optimalen Tests (d.h. $\beta(\vartheta) \leq \alpha$, $\vartheta \in \Theta_0$, $\beta(\vartheta)$ maximal für $\vartheta \in \Theta_1$). Diese haben die folgende Struktur:

Einseitiger Test: $\Theta = \mathbb{R}$, $\Theta_0 = (-\infty, \vartheta_0)$, $\Theta_1 = [\vartheta_0, \infty)$

$$e(X_1, \ldots, X_n) = \begin{cases} 1, & \text{falls } T_0(x_1, \ldots, x_n) > c_1, \\ 0, & \text{falls } T_0(x_1, \ldots, x_n) < c_1. \end{cases}$$

(vertausche 0 und 1, falls $\Theta_0 = (\vartheta_0, \infty)$ und $\Theta_1 = (-\infty, \vartheta_0]$)

Zweiseitiger Test: $\Theta = \mathbb{R}$, $\Theta_0 = \{\vartheta_0\}$, $\Theta_1 = \mathbb{R} \setminus \{\vartheta_0\}$

$$e(X_1, \ldots, X_n) = \begin{cases} 1, & \text{falls } T_0(x_1, \ldots, x_n) \notin [c_{21}, c_{22}], \\ 0, & \text{sonst.} \end{cases}$$

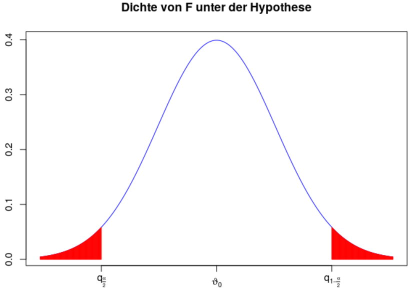
Strategie zur Konstruktion von Tests

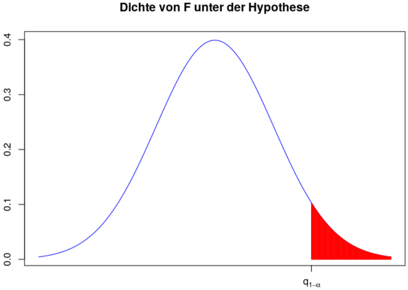
Bilde aus den Beobachtungen eine Teststatistik T (d.h. finde eine Abbildung $T : \mathbb{R}^n \to \mathbb{R}$), *deren Verteilung* F auf der Hypothese $\Theta_0 = \{\vartheta_0\}$ bzw. dem linken oder rechten Randpunkt der Hypothesen $\Theta_0 = (-\infty, \vartheta_0)$ oder $\Theta_0 = (\vartheta_0, \infty)$ *bekannt ist*.

Wählt man dann c_1 als das $1-\alpha$-Quantil $q^F_{1-\alpha}$ von F bzw. c_{21}, c_{22} als die $\frac{\alpha}{2}$- bzw. $1-\frac{\alpha}{2}$-Quantile $q^F_{\frac{\alpha}{2}}, {}^F_{1-\frac{\alpha}{2}}$ von F, erhält man einen Niveau-α-Test.

Ist $T = T_0$ (mit T_0 wie oben, was jeweils zu beweisen wäre), kann man sicher sein, dass der Test optimal ist in dem Sinne, dass er die größtmögliche Güte auf der Alternativenmenge hat.

Man beachte die Asymmetrie bzgl. der Fehler 1. und 2. Art!
Niveau-α-Tests halten den Fehler 1. Art unter Kontrolle, d.h. lehnt ein Test die Hypothese ab, kann man größerer Wahrscheinlichkeit davon ausgehen, dass sie tatsächlich falsch ist.
Verwirft ein Test jedoch die Hypothese nicht, ist damit noch nicht bewiesen, dass sie tatsächlich zutrifft! Je nach Testausgang kann die Wahrscheinlichkeit dafür recht gering sein.





Einfache Hypothese
$\Theta_0 = \{\vartheta_0\}$: Test lehnt Hypothese ab, wenn der Wert der Teststatistik T in einem der roten Bereiche liegt.

Zusammengesetzte Hypothese
$\Theta_0 = (-\infty, \vartheta_0)$: Test lehnt Hypothese ab, wenn der Wert der Teststatistik T im roten Bereiche liegt.

Fehlerhafte Testinterpretationen

Man beachte, dass die Fehlerwahrscheinlichkeiten 1. und 2. Art genau genommen *bedingte Wahrscheinlichkeiten* gegeben die Hypothese bzw. die Alternative sind:

$$P_\vartheta(\text{Fehler 1. Art}) = P_\vartheta(e(X_1, \dots, X_n) = 1 \mid \vartheta \in \Theta_0)$$
$$P_\vartheta(\text{Fehler 2. Art}) = P_\vartheta(e(X_1, \dots, X_n) = 0 \mid \vartheta \in \Theta_1)$$

Die Wahrscheinlichkeit, dass die Hypothese tatsächlich falsch ist, wenn der Test sie verwirft, ist dagegen

$$P_\vartheta(\vartheta \in \Theta_1 \mid e(X_1, \dots, X_n) = 1)$$

und diese kann deutlich kleiner als $1 - \alpha$ sein! Nehmen wir vereinfachend an, dass $P_\vartheta(\text{Fehler 1. Art}) \equiv \alpha$ und

$$P_\vartheta(e(X_1, \dots, X_n) = 1 \mid \vartheta \in \Theta_1) \equiv 1 - \alpha,$$

d.h der Test ist in beiden Fällen gleich trennscharf.

Fehlerhafte Testinterpretationen

Gilt $P(\vartheta \in \Theta_0) = 0.9 = 1 - P(\vartheta \in \Theta_1)$, so ist für $\alpha = 0.05$ nach der Bayesschen Regel

$$
\begin{aligned}
&P_\vartheta(\vartheta \in \Theta_1 \mid e(X) = 1) \\
&= \frac{P_\vartheta(e(X) = 1 \mid \vartheta \in \Theta_1) \cdot P(\vartheta \in \Theta_1)}{P_\vartheta(e(X) = 1 \mid \vartheta \in \Theta_1) \cdot P(\vartheta \in \Theta_1) + P_\vartheta(e(X) = 1 \mid \vartheta \in \Theta_0) \cdot P(\vartheta \in \Theta_0)} \\
&= \frac{0.95 \cdot 0.1}{0.95 \cdot 0.1 + 0.05 \cdot 0.9} = 0.6785714,
\end{aligned}
$$

d.h. nur in etwa zwei Dritteln der Fälle, in denen der Test die Hypothese verwirft, ist sie tatsächlich falsch.
Hat der Test eine geringere Macht auf der Alternative, d.h. ist $P_\vartheta(e(X_1, \dots, X_n) = 1 \mid \vartheta \in \Theta_1) < 1 - \alpha$, fällt obiger Wert noch geringer aus. Insbesondere wird deutlich, dass man ohne Kenntnis von $P(\vartheta \in \Theta_0)$ und $P(\vartheta \in \Theta_1)$ aus einem Testentscheid gegen die Hypothese nicht zwingend schließen kann, dass sie unzutreffend ist.