Taschenbuch der Statistik

von
Horst Rinne

4., überarb. u. erw. Aufl.

Thematische Gliederung:

Mathematische Statistik

Harri Deutsch 2008

Verlag C.H. Beck im Internet:
www.beck.de
ISBN 978 3 8171 1827 4

Inhaltsverzeichnis: Taschenbuch der Statistik – Rinne

Taschenbuch der Statistik

Prof. Dr. Horst Rinne

4., vollständig überarbeitete und erweiterte Auflage

Verlag Harri Deutsch

Dr. Horst Rinne ist emeritierter Professor für Statistik und Ökonometrie am Fachbereich Wirtschaftswissenschaften der Justus-Liebig-Universität Gießen. Er ist Autor zahlreicher Monographien über Ökonometrie, Zeitreihenanalyse sowie statistische Qualitätssicherung.

Wissenschaftlicher Verlag Harri Deutsch GmbH
Gräfstraße 47
60486 Frankfurt am Main
verlag@harri-deutsch.de
www.harri-deutsch.de

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

ISBN 978-3-8171-1827-4

4., vollständig überarbeitete und erweiterte Auflage 2008

Druck: Clausen & Bosse, Leck
Printed in Germany

3 Testtheorie

Eine **statistische Hypothese** *[hypothesis, statistical]* ist eine Behauptung über Eigenschaften einer oder mehrerer Zufallsvariablen, etwa über deren Parameter (→ Parameterhypothesen, vgl. Abs. C3.2 und Abs. C3.3) oder über deren Verteilung (→ Verteilungshypothesen, vgl. Abs. C3.4). Mit H_0 wird die zu prüfende Hypothese (= **Nullhypothese** *[null hypothesis]*[1]) bezeichnet, mit H_1 die **Gegen-** oder **Alternativhypothese** *[hypothesis, alternative]*. Ein Verfahren zur Überprüfung einer statistischen Hypothese auf der Basis einer Zufallsstichprobe aus der unter der Nullhypothese stehenden Verteilung heißt ein **statistischer Test**.

3.1 Grundbegriffe der Testtheorie

Anknüpfend an die zur Prüfung von H_0 herangezogene Stichprobenfunktion $\Psi(\boldsymbol{X})$ unterscheidet man zwischen

- **verteilungsgebundenen** Testverfahren *[parametric test]* (Die Verteilung von $\Psi(\boldsymbol{X})$ hängt von der Verteilung in der Grundgesamtheit ab, auf die sich H_0 bezieht.),
- **verteilungsfreien** (syn.: **parameterfreien**) Testverfahren *[parameter–free test]*. (Die Verteilung von $\Psi(\boldsymbol{X})$ hängt nicht von der Verteilung in der Grundgesamtheit ab, auf die sich H_0 bezieht.)

Anknüpfend an den Inhalt der Hypothesen unterscheidet man zwischen

- **parametrischen Tests** (syn.: **Parametertests**) *[parameter test]*, wenn sich die Hypothese auf Verteilungsparameter bezieht, und
- **nichtparametrischen Tests** *[non–parameter test]* sonst.

3.1.1 Ablauf eines Tests

Eine Nullhypothese H_0 wird als statistisch widerlegt angesehen und abgelehnt (= verworfen), wenn der Stichprobenbefund — ausgedrückt durch eine Stichprobenfunktion — in deutlichem (syn.: **signifikantem**) Gegensatz zu ihr steht, d. h. unter H_0 nur eine sehr geringe Eintrittswahrscheinlichkeit α hat. Entsprechend konzipierte Prüfverfahren heißen daher **Signifikanztests** *[significance test]*.

Ein Signifikanztest umfasst folgende Schritte:

1. Festlegung von Nullhypothese H_0 und Alternativhypothese H_1;
2. Festlegung des Stichprobenumfangs n und ggf. der Art der Stichprobenziehung;
3. Wahl des **Signifikanzniveaus** *[level of significance]* α;
4. Wahl einer Stichprobenfunktion $G := \Psi(\boldsymbol{X})$ als **Prüfgröße** (syn.: **Testfunktion**);
5. Konstruktion eines **Ablehnbereichs** (syn.: **kritischer Bereich**) *[critical region]*K_α für die Prüfgröße G;
6. Stichprobenziehung, Auswertung von $\Psi(.)$ mit dem realisierten Stichprobenvektor $\boldsymbol{x}$ zu $g = \Psi(\boldsymbol{x})$ und Entscheidung:
   - Lehne H_0 auf dem Niveau α ab, wenn $g \in K_\alpha$.
   - Lehne H_0 nicht ab, wenn $g \notin K_\alpha$.

Hinweis: Statt der Schritte 5 und 6 kann man äquivalent auch einen Schritt 5* durchführen: Berechne das **empirische Signifikanzniveau** *[level of significance, empirical]* α^* als Wahrscheinlichkeit dafür, dass G unter H_0 den realisierten Wert g „in Richtung auf H_1 überschreitet“, d. h. man

[1] Der Zusatz „Null“ mag damit erklärt werden, dass beim Signifikanztest (vgl. unten) diese Hypothese als Null und nichtig erwiesen werden soll.

berechnet das Niveau, bei dem der Test gerade noch ablehnen würde.[2] H_0 wird dann auf dem Niveau α verworfen, wenn $\alpha^* \leq \alpha$.

Die obigen Schritte seien am Parametertest erläutert. Gegeben ist eine Verteilungsfunktion $F_X(x \mid \Theta)$, die von einem Parameter Θ abhängt, der Werte aus einem **Parameterraum** Ω annehmen kann: $\Theta \in \Omega$. Der wahre Wert von Θ ist unbekannt.

1. Schritt

Hypothesenbildung[3]

$$H_0\colon\ \Theta \in \Omega_0; \quad H_1\colon\ \Theta \in \Omega_1 = \Omega \setminus \Omega_0, \ \text{ wobei weder } \ \Omega_0 = \emptyset \ \text{ noch } \ \Omega_0 = \Omega.$$

Es heißt H_i $(i = 0, 1)$

- **einfach** *[hypothesis, simple]*, wenn $|\Omega_i| = 1$, d. h. Ω_i einelementig ist,
- **zusammengesetzt** *[hypothesis, composite]*, wenn $|\Omega_i| > 1$. Ω_i ist dann i. A. ein Intervall oder Bereich.

Der Test heißt

- **einseitig** *[test, one–sided]*, wenn Ω_1 einseitig beschränkt ist,
- **zweiseitig** *[test, two–sided]*, wenn Ω_1 nach oben und unten unbeschränkt ist.

Beispiele:
1. $H_0\colon\ \mu = 2; \quad H_1\colon\ \mu \neq 2$
   $\Longrightarrow$ einfache Hypothese gegen eine zusammengesetzte Alternative mit einem zweiseitigen Test,
2. $H_0\colon\ P \leq 0{,}1; \quad H_1 : P > 0{,}1$
   $\Longrightarrow$ zusammengesetzte Hypothese gegen eine zusammengesetzte Alternative mit einem einseitigen Test.

2. Schritt

Der Stichprobenumfang n und die Art der Stichprobenziehung beeinflussen die Varianz der Prüfgröße und damit die Güte des Tests *[power of a test]* (vgl. Abs. C3.1.2): Je kleiner die Varianz von $\Psi(\boldsymbol{X})$, desto trennschärfer ist der Test.

3. Schritt

Das Signifikanzniveau α ist die maximal tolerierte Wahrscheinlichkeit dafür, dass H_0 zu Unrecht abgelehnt wird. Als **Fehler erster Art** *[error of the first kind]* bezeichnet man das **Verwerfen von** H_0, **wenn** H_0 **zutrifft** ($\widehat{=}$ **irrtümliches Verwerfen**). Je kleiner α gewählt wird (gebräuchliche Werte sind 0,1; 0,05; 0,01; 0,001), desto unwahrscheinlicher wird diese Fehlentscheidung. Da man H_0 bei kleinem α–Wert selten ablehnt, steigt aber die Wahrscheinlichkeit, auch H_0 dann nicht abzulehnen, wenn H_0 falsch ist. Als **Fehler zweiter Art** *[error of the second kind]* bezeichnet man das **Nichtverwerfen von** H_0 (kurz: Annahme von H_0), **wenn** H_0 **nicht zutrifft** ($\widehat{=}$ **irrtümliche Annahme**). Während beim Signifikanztest der Fehler erster Art durch Vorgabe von α unter Kontrolle ist, kann der Fehler zweiter Art eine sehr hohe Wahrscheinlichkeit haben, bis zu $1 - \alpha$, vgl. Abb. C3/1. Da Ablehnung von H_0 mit dem Ereignis $G \in K_\alpha$ äquivalent ist, wird ein Signifikanztest zum Niveau α im Falle $H_0\colon\ \Theta \in \Omega_0$ durch[4]

$$\alpha = \sup_{\Theta \in \Omega_0} \Pr(G \in K_\alpha \mid \Theta)$$

[2] Diese Vorgehensweise wird in vielen Statistik–Software–Paketen realisiert, da die numerische Bestimmung von Wahrscheinlichkeiten einer Prüfverteilung i. d. R. einfacher ist als die ihrer Perzentile.

[3] Formal gesehen zerlegt man mit der Hypothesenbildung den Parameterraum in zwei disjunkte, nichtleere Teilmengen. Sachlich–inhaltliche Aspekte der Hypothesenformulierung werden in Abs. C3.1.3 behandelt.

[4] Als **Fehler dritter Art** *[error of the third kind]* wird die Wahl eines falschen Modells oder einer falschen Methode bezeichnet.

definiert. Das Supremum aller Ablehnwahrscheinlichkeiten, wenn Θ in Ω_0 liegt (also H_0 zutrifft), muss gleich dem Signifikanzniveau sein. Im Falle einer einfachen Nullhypothese $H_0: \; \Theta = \Theta_0$ ergibt sich

$$\alpha = \Pr(G \in K_\alpha \,|\, \Theta_0).$$

Tab. C3/1: Entscheidungsalternativen und deren Wahrscheinlichkeiten* beim Signifikanztest

Wert der Entscheidung / Entscheidung	H_0 ist wahr.	H_0 ist falsch.
H_0 nicht verwerfen (Wert der Entscheidung)	$\Pr(G \notin K_\alpha \| H_0) \geq 1-\alpha$ (richtige Entscheidung)	$\Pr(G \notin K_\alpha \| H_1)$ (Fehler 2. Art)
H_0 verwerfen (Wert der Entscheidung)	$\Pr(G \in K_\alpha \| H_0) \leq \alpha$ (Fehler 1. Art)	$\Pr(G \in K_\alpha \| H_1)$ (richtige Entscheidung)

$*$) Die Wahrscheinlichkeiten in dieser Tabelle sind bedingte Wahrscheinlichkeiten.
Innerhalb einer Spalte ergänzen sie sich zu Eins, nicht aber innerhalb einer Zeile.

Bemerkung: Im Falle einer diskreten Prüfgröße G lassen sich die obigen Bestimmungsgleichungen für α nicht immer genau erfüllen. Zwei Auswege bieten sich an:

- Durchführung eines **randomisierten Tests** *[test, randomized]*, vgl. Abs. C3.1.7;
- **Gehen–auf–die–sichere–Seite**, indem man nur

$$\sup_{\Theta \in \Omega_0} \Pr(G \in K_\alpha \,|\, \Theta) < \alpha$$

erfüllt, d. h. α nicht voll ausschöpft.

4. Schritt

Die Wahl der Prüfgröße $G = \Psi(\boldsymbol{X})$ ist im Prinzip beliebig; i. d. R. wählt man G so, dass sie gute stochastische Eigenschaften hat. Zu beachten ist aber:

- Die Verteilung von G muss von den zu testenden Hypothesen H_0 und H_1 abhängen.
- Die Verteilung von G unter H_0 muss bekannt sein, denn nur dann lässt sich feststellen, ob das gegebene Signifikanzniveau α eingehalten wird.

5. Schritt

Der Ablehnbereich K_α bzw. sein Komplement $\overline{K}_\alpha$, der **Annahmebereich** *[acceptance region]*, wird unter Beachtung der Verteilung von G unter H_0 so festgelegt, dass

1. $\sup_{\Theta \in \Omega_0} \Pr(G \in K_\alpha \,|\, \Theta) = \alpha$ (bzw. $\leq \alpha$) bei zusammengesetzter H_0
   $\Pr(G \in K_\alpha \,|\, \Theta_0) = \alpha$ (bzw. $\leq \alpha$) bei einfacher H_0,
2. $\sup_{\Theta \in \Omega_1} \Pr(G \in \overline{K}_\alpha \,|\, \Theta)$ minimal wird.

Forderung 2 erfüllt man, indem K_α so gelegt wird, dass ein Wert g von G, der in K_α fällt, stark gegen H_0 und damit für H_1 spricht.

6. Schritt

Die **Nullhypothese** H_0 wird dann **abgelehnt**, wenn der Wert g von G in den kritischen Bereich K_α fällt. Man sagt, die Beobachtung steht im signifikanten Widerspruch zu H_0 und fasst dies als **Bestätigung der**

Gegenhypothese H_1 auf. Die Schlussfolgerung, dass dann H_1 zutreffend sei, ist eine statistische Entscheidung, die richtig oder falsch sein kann.[5] Die Fehlentscheidung (für H_1, wenn H_0 und nicht H_1 gilt) ist unter Kontrolle, da die Irrtumswahrscheinlichkeit (Wahrscheinlichkeit für den Fehler 1. Art) höchstens α beträgt.

Kann die **Nullhypothese nicht abgelehnt** werden, da g nicht in den kritischen Bereich K_α fällt, so bedeutet dies **nicht die Bestätigung von** H_0, sondern nur, dass die Beobachtung nicht zur Verwerfung ausreicht (Freispruch mangels Beweis). Man nimmt dann zwar H_0 an, da eine Entscheidung getroffen werden muss, d. h. man arbeitet mit H_0 weiter, bis man ggf. vom Gegenteil überzeugt wird. Die Entscheidung „Annahme von H_0" ist aber nicht unter statistischer Kontrolle, da man für den Fehler 2. Art beim Signifikanztest keine vorgegebene Höchstwahrscheinlichkeit β einhalten kann.

Bemerkung: Das Konfidenzintervall zum Niveau $1-\alpha$ (vgl. Abs. C2.2.2) für den zu testenden Parameter Θ ist die Menge jener hypothetischen Werte unter H_0, für die das Stichprobenergebnis zum Niveau α nicht signifikant ist.

3.1.2 Beurteilung eines Tests

Zum Testen einer Nullhypothese bieten sich oft verschiedene Verfahren an. Die Auswahl eines Tests verlangt einen Beurteilungsmaßstab. Dieser ist die **Gütefunktion** *[power function]* (syn.: Machtfunktion) oder die dazu (auf Eins) komplementäre **OC–Funktion** *[operating characteristic]* (syn.: Operationscharakteristik).

Definition: Die Gütefunktion eines Tests für den Parameter Θ ist gegeben durch:

$$GF:\ \Omega\ \rightarrow\ [0;1]\ \text{ mit }\ GF(\Theta) := \Pr(G \in K_\alpha \,|\, \Theta).$$

Die **Gütefunktion** ist also die **Ablehnwahrscheinlichkeit** des Tests in Abhängigkeit von Θ, wobei

- $GF(\Theta)$ eine Fehlentscheidungswahrscheinlichkeit (für den Fehler 1. Art) ist, wenn $\Theta \in \Omega_0$,
- $GF(\Theta)$ die Wahrscheinlichkeit einer richtigen Entscheidung ist, wenn $\Theta \in \Omega_1$.

Es ist $\alpha = \sup_{\Theta\in\Omega_0} GF(\Theta)$.

Definition: Die OC–Funktion eines Tests für den Parameter Θ ist gegeben durch:

$$OC:\ \Omega\ \rightarrow\ [0;1]\ \text{ mit }\ OC(\Theta) := \Pr(G \notin K_\alpha \,|\, \Theta).$$

Die **OC–Funktion** ist also die **Annahmewahrscheinlichkeit** in Abhängigkeit von Θ, wobei

- $OC(\Theta)$ eine Fehlentscheidungswahrscheinlichkeit (für den Fehler 2. Art) ist, wenn $\Theta \in \Omega_1$,
- $OC(\Theta)$ die Wahrscheinlichkeit einer richtigen Entscheidung ist, wenn $\Theta \in \Omega_0$.

Es ist $\beta := \sup_{\Theta\in\Omega_1} OC(\Theta)$.

Definition: Ein Test mit dem Signifikanzniveau α heißt **konservativ** *[test, conservative]*,[6] wenn

$$\sup_{\Theta\in\Omega_0} GF(\Theta) < \alpha\,.$$

Interpretation: Der Test schöpft die vorgegebene Wahrscheinlichkeit für den Fehler 1. Art nicht voll aus.

Definition: Ein Test mit dem Signifikanzniveau α heißt **unverfälscht** *[test, unbiased]* (syn.: **unverzerrt**), wenn

$$GF(\Theta) \geq \alpha\ \ \forall\, \Theta \in \Omega_1.$$

[5] Merke: Ein statistischer Test, der nur auf einer Teilinformation (Stichprobe) über eine Grundgesamtheit beruht, kann H_0 (bzw. H_1) niemals im Sinne der Logik als wahr oder falsch nachweisen.

[6] Man spricht auch von einem „Gehen auf die sichere Seite" .

Interpretation: Bei einem unverfälschten Test wird die Nullhypothese, wenn sie nicht zutrifft (d. h. $\Theta \in \Omega_1$), mit einer mindestens so hohen Wahrscheinlichkeit verworfen wie im Falle ihres Zutreffens.

Definition: Ein Test heißt **konsistent** *[test, consistent]*, wenn

$$\lim_{n\to\infty} \Pr(G_n \in K_\alpha \,|\, \Theta \in \Omega_1) = 1.$$

Interpretation: Mit wachsendem Stichprobenumfang wird die Ablehnung einer unzutreffenden Nullhypothese immer sicherer.

Definition: Ein Test heißt **gleichmäßig bester** *[test, uniformly most powerful]* (syn.: **trennscharfer**) Test, wenn dessen Gütefunktion $GF(\Theta)$ über Ω_1 mindestens so groß ist wie die Gütefunktion $\widetilde{GF}(\Theta)$ eines beliebigen anderen Niveau–α–Tests:

$$GF(\Theta) \geq \widetilde{GF}(\Theta) \ \ \forall\, \Theta \in \Omega_1.$$

Beispiel „Mittleres Abfüllgewicht“[7]

Es soll 1 kg Zucker in Tüten abgefüllt werden. Das Abfüllgewicht X sei unstrittig normalverteilt, $X \sim No(\mu; \sigma^2)$, mit bekannter Standardabweichung $\sigma = 2\,g$. Der Erwartungswert μ des Abfüllgewichts ist unbekannt, da sich die Abfüllvorrichtung in ihrer Einstellung im Zeitablauf verändert. Die Nullhypothese lautet zunächst:

$$H_0 : \mu = \mu_0 \ \text{ mit } \ \mu_0 = 1000\,g.$$

Die Alternativhypothese wird nach der jeweiligen Interessenlage anders ausfallen (vgl. Abb. C3/3).

a) Eine **Verbraucherorganisation** möchte nachweisen, dass im Mittel zu wenig abgefüllt wird, also lautet hier

$$H_1 : \ \mu < \mu_0 \ \text{ mit } \ H_0 : \mu \geq \mu_0.$$[8]

b) Der **Abfüller** möchte nachweisen, dass er im Mittel mehr als erforderlich abfüllt, also lautet hier

$$H_1 : \ \mu > \mu_0 \ \text{ mit } \ H_0 : \mu \leq \mu_0.$$[8]

c) Die **Gewerbeaufsicht** möchte nachweisen, dass im Mittel nicht korrekt abgefüllt wird, also lautet hier

$$H_1 : \ \mu \neq \mu_0 \ \text{ mit } \ H_0 : \mu = \mu_0.$$[8]

Wegen der günstigen Eigenschaften von $\overline{X}_n$ als Schätzer für μ wird $\overline{X}_n$ als Prüfgröße G genommen: $\overline{X}_n \sim No(\mu; \sigma^2/n)$. Dann wird H_0 abgelehnt, wenn

a) im Falle der Verbraucherorganisation $\overline{X}_n$ viel kleiner als μ_0,
b) im Falle des Abfüllers $\overline{X}_n$ viel größer als μ_0 und
c) im Falle der Gewerbeaufsicht $|\overline{X}_n - \mu_0|$ groß ausfällt.

Der kritische Bereich K_α bei gegebenem α folgt unter H_0 aus $No(\mu_0; \sigma^2/n)$ als:

a) $\overline{X}_n < \mu_0 - u_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}$ bzw. $K_\alpha^a = \left(-\infty;\ \mu_0 - u_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}\right)$,

b) $\overline{X}_n > \mu_0 + u_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}$ bzw. $K_\alpha^b = \left(\mu_0 + u_{1-\alpha}\dfrac{\sigma}{\sqrt{n}};\ \infty\right)$,

[7] Beispiel in Anlehnung an BAMBERG/BAUR: Statistik; 12. Aufl., Oldenbourg, München/ Wien (2002), p. 173 ff.

[8] Da durch H_0 und H_1 der gesamte Parameterraum (hier: $\Omega = \mathbb{R}$) aufgeteilt wird, ist diese Form von H_0 erforderlich.

c) $\overline{X}_n < \mu_0 - u_{1-\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}$ oder $\overline{X}_n > \mu_0 + u_{1-\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}$

bzw. $K_\alpha^c = \left(-\infty; \mu_o - u_{1-\alpha/2}\,\dfrac{\sigma}{\sqrt{n}}\right) \cup \left(\mu_0 + u_{1-\alpha/2}\,\dfrac{\sigma}{\sqrt{n}};\ \infty\right)$.

Würde man α nicht hälftig auf den oberen und unteren Teil des kritischen Bereichs aufteilen, würde der Test eine gleich große Abweichung des unbekannten μ von μ_0 nach oben oder nach unten mit unterschiedlichen Wahrscheinlichkeiten entdecken, was sicherlich von der Gewerbeaufsicht nicht erwünscht ist.

Sei $\alpha = 0{,}01$, so dass $u_{0,99} \approx 2{,}33$ und $u_{0,995} \approx 2{,}58$. In einer Stichprobe von $n = 16$ habe sich $\bar{x}_{16} = 998{,}5\,g$ eingestellt.

a) $\bar{x}_{16} = 998{,}5 \in K_\alpha^a = (-\infty; 998{,}835)$

$\Longrightarrow$ Die Verbraucherorganisation kann $H_0\colon\ \mu \geq 1000$ zugunsten von $H_1\colon\ \mu < 1000$ verwerfen.

b) $\bar{x}_{16} = 998{,}5 \notin K_\alpha^b = (1001{,}165; \infty)$

$\Longrightarrow$ Der Abfüller kann $H_0\colon\ \mu \leq 1000$ nicht zugunsten von $H_1\colon\ \mu > 1000$ verwerfen.

c) $\bar{x}_{16} = 998{,}5 \in K_\alpha^c = (-\infty; 998{,}71) \cup (1001,29; \infty)$

$\Longrightarrow$ Die Gewerbeaufsicht kann $H_0\colon\ \mu = 1000$ zugunsten von $H_1\colon\ \mu \neq 1000$ verwerfen.

Eine Testdurchführung mit dem empirischen Signifikanzniveau α^* liefert äquivalent:

a)
$$\begin{aligned}\alpha^* &= \Pr(\overline{X}_{16} < \bar{x}_{16} \mid \mu = 1000; \sigma_{\overline{X}} = 0{,}5)\\ &= \Pr\left(U < \frac{998{,}5 - 1000}{0{,}5}\right) = \Pr(U < -3) \approx 0{,}00135\end{aligned}$$

$\Longrightarrow$ Ablehnung von H_0, da $\alpha^* \approx 0{,}00135 < \alpha = 0{,}01$.

b)
$$\begin{aligned}\alpha^* &= \Pr(\overline{X}_{16} > \bar{x}_{16} | \mu = 1000; \sigma_{\overline{X}} = 0{,}5)\\ &= \Pr\left(U > \frac{998{,}5 - 1000}{0{,}5}\right) = \Pr(U > -3) \approx 0{,}99865\end{aligned}$$

$\Longrightarrow$ Keine Ablehnung von H_0, da $\alpha^* \approx 0{,}99865 > \alpha = 0{,}01$.

c) Sei $\eta := \overline{X}_{16} - \mu_0$. Unter H_0 ist $\eta \sim No(0; 0{,}5^2)$.

$$\begin{aligned}\alpha^* &= \Pr(|\eta| > |\bar{x}_{16} - \mu_0|) = \Pr(|\eta| > 1,5)\\ &= \Pr(\eta < -1{,}5 \vee \eta > 1{,}5)\\ &= \Pr(U < -3 \vee U > 3) = 2\Pr(U < -3) \approx 0{,}0027\end{aligned}$$

$\Longrightarrow$ Ablehnung von H_0, da $\alpha^* \approx 0{,}0027 < 0{,}01$.

Die Gütefunktion des Tests für die Verbraucherorganisation ergibt sich als

$$\begin{aligned}GF^a(\mu) &= \Pr(\overline{X}_{16} \in K_\alpha^a \mid \mu) = \Pr(\overline{X}_{16} < 998{,}835 \mid \mu)\\ &= \Pr\left(U < \frac{998{,}835 - \mu}{0{,}5}\right)\\ &= \Phi\left(\frac{998{,}835 - \mu}{0{,}5}\right),\end{aligned}$$

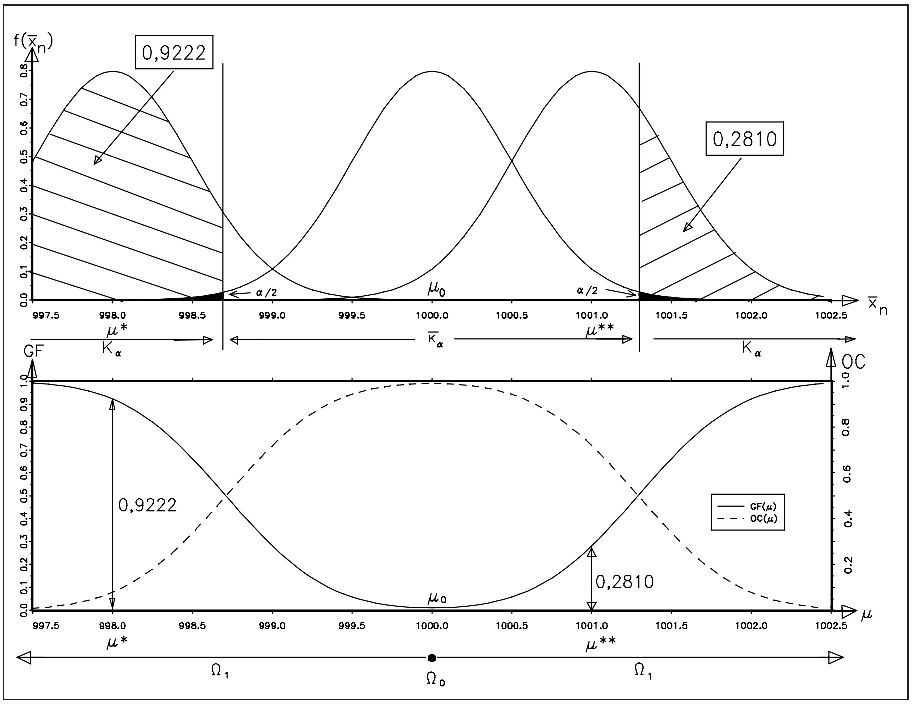
die des Abfüllers als

$$\begin{aligned} GF^b(\mu) &= \Pr(\overline{X}_{16} \in K_\alpha^b \,|\, \mu) = \Pr(\overline{X}_{16} > 1001{,}165 \,|\, \mu) \\ &= \Pr\left(U > \frac{1001{,}165 - \mu}{0{,}5}\right) \\ &= 1 - \Phi\left(\frac{1001{,}165 - \mu}{0{,}5}\right) \end{aligned}$$

und die der Gewerbeaufsicht als

$$\begin{aligned} GF^c(\mu) &= \Pr(\overline{X}_{16} \in K_\alpha^c \,|\, \mu) \\ &= \Pr(\overline{X}_{16} < 998{,}71 \,|\, \mu) + \Pr(\overline{X}_{16} > 1001{,}29 \,|\, \mu) \\ &= \Pr\left(U < \frac{998{,}71 - \mu}{0{,}5}\right) + \Pr\left(U > \frac{1001{,}29 - \mu}{0{,}5}\right) \\ &= 1 - \Phi\left(\frac{1001{,}29 - \mu}{0{,}5}\right) + \Phi\left(\frac{998{,}71 - \mu}{0{,}5}\right). \end{aligned}$$

Abb. C3/1: Güte- und OC–Funktion eines zweiseitigen Tests



Die letzte Gütefunktion ist — zusammen mit der komplementären OC–Funktion — im unteren Teil der Abb. C3/1 dargestellt. Diese Gütefunktion ist symmetrisch um $\mu_0 = 1000$. Sie nimmt mit wachsender Abweichung $|\mu - \mu_0|$ zu. Im oberen Teil der Abb. C3/1 sind drei Verteilungen für $\overline{X}_n = \overline{X}_{16}$ angegeben.

Unter der Verteilung mit $\mu = \mu_0 = 1000$ findet man jeweils $\alpha/2 = 0{,}005$ im linken und rechten Teil von K_α. Für die Verteilung mit $\mu^* = 998$ ergibt sich

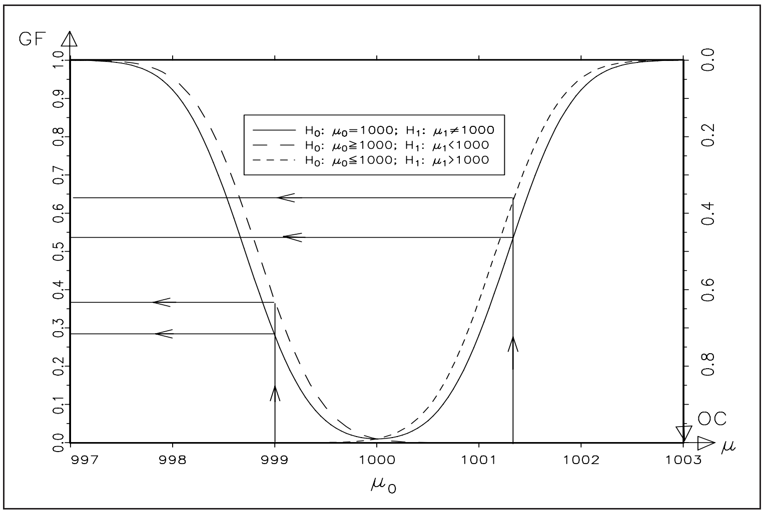
$$GF^c(998) = 1 - \Phi\left(\frac{1001{,}29 - 998}{0{,}5}\right) + \Phi\left(\frac{998{,}71 - 998}{0{,}5}\right)$$
$$= \underbrace{1 - \Phi(6{,}58)}_{\approx 0} + \underbrace{\Phi(1{,}42)}_{\approx 0{,}9222} \approx 0{,}9222,$$

für die Verteilung mit $\mu^{**} = 1001$

$$GF^c(1001) = 1 - \Phi\left(\frac{1001{,}29 - 1001}{0{,}5}\right) + \Phi\left(\frac{998{,}71 - 1001}{0{,}5}\right)$$
$$= \underbrace{1 - \Phi(0{,}58)}_{\approx 0{,}2810} + \underbrace{\Phi(-4{,}58)}_{\approx 0} \approx 0{,}2810.$$

Abb. C3/2 zeigt den Verlauf der drei Gütefunktionen $GF^a(\mu)$, $GF^b(\mu)$ und $GF^c(\mu)$. Man sieht, dass bei den zwei einseitigen Tests das vorgegebene Signifikanzniveau α jeweils am „Rand" (bei μ_0) des unter H_0 stehenden Parameterteilraums Ω_0 angenommen wird und die Wahrscheinlichkeit des Fehlers 1. Art mit wachsender „Entfernung" des Parameters μ vom Rand abnimmt. Man erkennt ferner, dass ein einseitiger Test über Ω_1 trennschärfer ist als unter sonst gleichen Bedingungen der zweiseitige Test, dessen Gütefunktion flacher verläuft. Wo immer möglich, sollte man daher einseitige Hypothesen formulieren und testen.

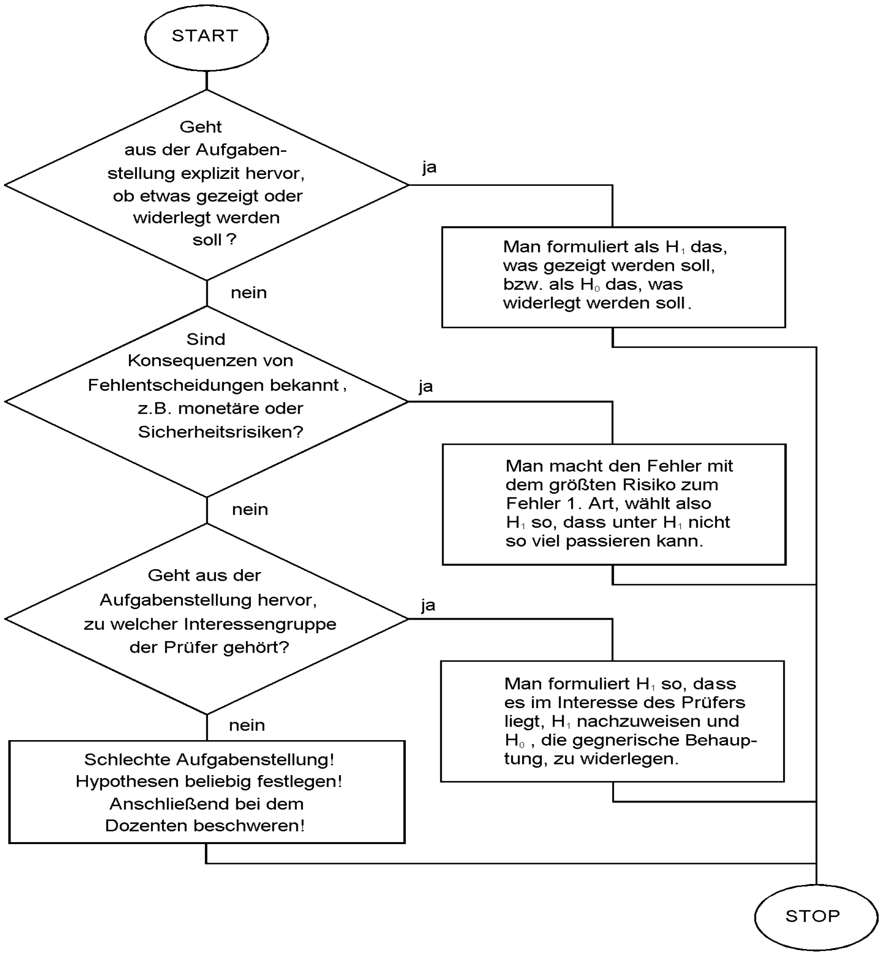
Abb. C3/2: Vergleich der Gütefunktionen ein- und zweiseitiger Tests



3.1.3 Hypothesenformulierung

Beim Signifikanztest werden Null- und Alternativhypothese ungleich behandelt. Das Hauptinteresse gilt H_0, die möglichst selten und nur mit kontrollierter Wahrscheinlichkeit irrtümlich abgelehnt werden soll. Auch stellen Ablehnung und Nichtablehnung von H_0 Aussagen unterschiedlicher Qualität dar, was seinen Grund in der bei der Aufstellung des Ablehnbereichs deutlich gewordenen Ungleichbehandlung der Fehler erster und zweiter Art hat. Als **Faustregel** gilt, dass man diejenige Hypothese unter H_1 stellt, die man bestätigen oder statistisch untermauern möchte. Bei Ablehnung von H_0 gilt H_1 als statistisch bestätigt, während die Nichtablehnung von H_0 nicht als statistische Bestätigung von H_0 aufzufassen ist. Abb. C3/3 enthält einen ausführlichen Leitfaden zur Hypothesenbildung.

Abb. C3/3: Hypothesenformulierung beim Signifikanztest



Beispiel: Eine Maschine M1 hat bisher Werkstücke mit einer Ausschussquote von $P_0 = 0{,}2$ produziert. Diese Maschine soll durch eine neue Maschine M2 ersetzt werden, aber nur dann, wenn sie besser ist, also mit einer niedrigeren Ausschussquote P als M1 arbeitet. In einem Probelauf beim Hersteller von M2 wird eine Stichprobe von $n = 30$ Werkstücken gefertigt und geprüft. Um die Beziehung $P < P_0 = 0{,}2$ (M2 ist besser als M1) statistisch abzusichern, wird sie unter H_1 gestellt und man erhält

$$H_0:\ P \geq 0{,}2\,;\quad H_1:\ P < 0{,}2\,.$$

Es sei $\alpha = 0{,}05$. Die Prüfgröße

$$X_n^T = \sum_{i=1}^{30} X_i \text{ mit } X_i = \begin{cases} 0 - & i\text{–tes Werkstück gut} \\ 1 - & i\text{–tes Werkstück defekt} \end{cases}$$

ist mit P und $n = 30$ binomialverteilt. Je kleiner X_n^T ausfällt, desto deutlicher spricht dies für H_1 und gegen H_0, so dass der Ablehnbereich von der Gestalt

$$K_\alpha = \{0, 1, 2, \ldots, c\}$$

ist, wobei zu fordern ist

$$\sup_{P \geq 0{,}2} \Pr\big(X_n^T \leq c \,|\, P\big) \leq \alpha = 0{,}05$$

und

$$\sup_{P \geq 0{,}2} \Pr\big(X_n^T \leq c + 1 \,|\, P\big) > \alpha = 0{,}05\,.$$

Da $\Pr\big(X_n^T \leq c \,|\, P\big)$ monoton fallend in P ist, ergeben sich die beiden supremalen Wahrscheinlichkeiten aus $Bi(30; 0{,}2)$ als:

$$\sum_{i=0}^{2} \binom{30}{i} 0{,}2^i \, 0{,}8^{30-i} \approx 0{,}0442 < \alpha = 0{,}05$$

$$\sum_{i=0}^{3} \binom{30}{i} 0{,}2^i \, 0{,}8^{30-i} \approx 0{,}1227 > \alpha = 0{,}05,$$

so dass der Ablehnbereich konkret lautet:

$$K_\alpha = \{0, 1, 2\}.$$

Finden sich also bis zu zwei defekte Stücke in $n = 30$, wird H_0 abgelehnt, d. h. man entscheidet sich für die neue Maschine M2 und hat ein Risiko von höchstens $4,42\%$, dann eine Maschine installiert zu haben, die gleich oder schlechter ist als die alte Maschine M1. Wäre die neue Maschine besser als M1, so sieht man am Komplement der Gütefunktion über Ω_1 in Abb. C3/4, dass die Wahrscheinlichkeit der Nichtablehnung von H_0 bis zu 95,58% betragen kann, man also eine sehr hohe Chance hat, bei M1 zu bleiben, wenn M2 besser ist.

Will das Unternehmen das Risiko unter Kontrolle halten, bei der alten und schlechteren Maschine M1 zu bleiben, so muss es

$$H_0^*:\ P < 0{,}2 \text{ gegen } H_1^*:\ P \geq 0{,}2$$

testen und bei der alten Maschine bleiben, wenn der Test zur Ablehnung von H_0^* führt, also gezeigt ist, dass die neue Maschine statistisch gesichert schlechter arbeitet. Seien wieder $\alpha = 0{,}05,\ n = 30$ und X_n^T die Prüfgröße. Je größer jetzt X_n^T, desto deutlicher spricht dies für H_1^* und gegen H_0^*, so dass der Ablehnbereich die Gestalt

$$K_\alpha^* = \{d, d + 1, \ldots, 30\}$$

hat und zu fordern ist

$$\sup_{P < 0{,}2} \Pr\big(X_n^T \geq d \,|\, P\big) \leq \alpha = 0{,}05$$

und

$$\sup_{P < 0{,}2} \Pr\big(X_n^T \geq d - 1 \,|\, P\big) > \alpha = 0{,}05.$$

Da $\Pr(X_n^T \geq d \mid P)$ monoton steigend in P ist, ergeben sich die supremalen Wahrscheinlichkeiten mit $P = 0{,}2$ als

$$\sum_{i=11}^{30} \binom{30}{i} 0{,}2^i \, 0{,}8^{30-i} \approx 0{,}0256 < \alpha = 0{,}05,$$

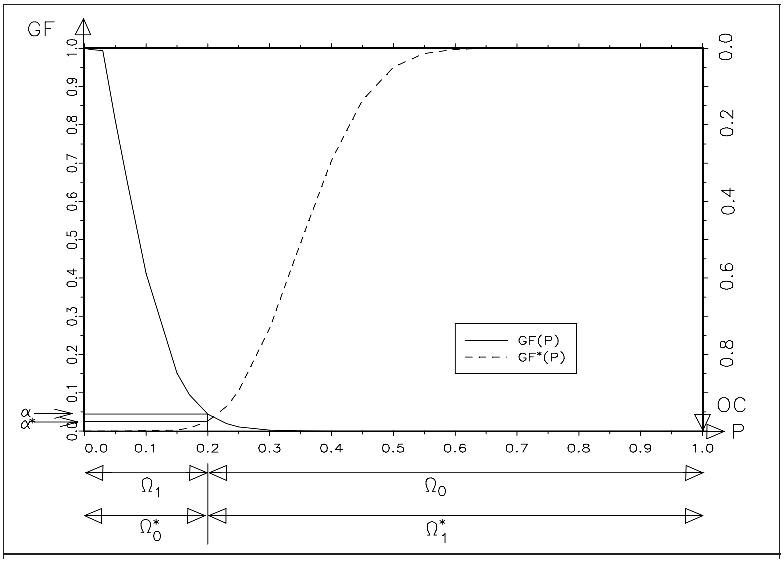
$$\sum_{i=10}^{30} \binom{30}{i} 0{,}2^i \, 0{,}8^{30-i} \approx 0{,}0611 > \alpha = 0{,}05,$$

so dass der Ablehnbereich definitiv lautet

$$K_\alpha^* = \{11, 12, \ldots, 30\}.$$

Findet man elf oder mehr defekte Stücke in $n = 30$, wird H_0^* abgelehnt und weiter mit M1 produziert. Wird H_0^* nicht abgelehnt, wird man wohl M2 einführen. Die Chance, dass man bei M1 bleibt, wenn M2 besser ist $(P < 0{,}2)$, liest man in Abb. C3/4 unter $GF^*(P)$ über Ω_0^* bei $P \approx 0{,}2$ mit 2,56% ab. Das Risiko, M2 einzuführen und damit eine schlechtere Maschine als M1 zu haben, ist als Wahrscheinlichkeit des Fehlers 2. Art das Komplement von $GF^*(P)$ über Ω_1^*. Bei $P + \varepsilon$, ε klein, kann diese Wahrscheinlichkeit bis zu 97,44% betragen.

Abb. C3/4: Gütefunktionen der Tests von $H_0: \; P \geq 0{,}2$ gegen $H_1: \; P < 0{,}2$ und $H_0^*: \; P < 0{,}2$ gegen $H_1^*: \; P \geq 0{,}2$ mit $n = 30$ und $\alpha = 0{,}05$



3.1.4 Likelihood–Quotienten–Tests

Sei $\boldsymbol{X} = (X_1, \ldots, X_n)'$ eine Stichprobe mit der Dichte[9] $f(\boldsymbol{x} \mid \boldsymbol{\Theta})$ für $\boldsymbol{\Theta} \in \Omega$, wobei $\boldsymbol{\Theta}$ bzgl. eines Vektors ζ absolut stetig ist. Zu testen ist $H_0: \; \boldsymbol{\Theta} \in \Omega_0$ gegen $H_1: \; \boldsymbol{\Theta} \in \Omega_1 = \Omega \backslash \Omega_0$. Der **LQ–Test** (Likelihood–

[9] Im diskreten Fall ist von der gemeinsamen Wahrscheinlichkeitsfunktion $\Pr(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{\Theta})$ auszugehen.

Quotienten–Test) *[likelihood ratio test]* basiert auf dem folgenden Likelihood–Quotienten als Prüfgröße:

$$\lambda(\boldsymbol{X}) := \frac{\sup\limits_{\boldsymbol{\Theta}\in\Omega_0} f(\boldsymbol{X}\,|\,\boldsymbol{\Theta})}{\sup\limits_{\boldsymbol{\Theta}\in\Omega} f(\boldsymbol{X}\,|\,\boldsymbol{\Theta})}\,.$$

Der Test hat zum Niveau α für H_0 den Ablehnbereich

$$K_\alpha = \big\{\boldsymbol{x} : \lambda(\boldsymbol{x}) < c\big\}, \tag{†}$$

wobei c so zu wählen ist, dass

$$\int_{K_\alpha} f(\boldsymbol{x}\,|\,\boldsymbol{\Theta})\;\mathrm{d}\zeta(\boldsymbol{\Theta}) \le \alpha\;\forall\;\boldsymbol{\Theta}\in\Omega_0.$$

Dieser Test geht auf NEYMAN (J. NEYMAN, 1894–1981) und PEARSON (E.S. PEARSON, 1895–1980) zurück.[10]

Wenn H_0 und H_1 jeweils einfache Hypothesen sind $\big[H_0\colon\ \boldsymbol{\Theta}=\boldsymbol{\Theta}_0$ versus $H_1\colon \boldsymbol{\Theta}=\boldsymbol{\Theta}_1\big]$, ist der LQ–Test eng verwandt, jedoch nicht identisch mit dem durch das NEYMAN/PEARSON–Lemma[11] gegebenen besten Niveau–α–Test. Die obige Testfunktion $\lambda(\boldsymbol{X})$ und der **alternative Likelihood–Quotient**

$$LQ := f(\boldsymbol{X}\,|\,\boldsymbol{\Theta}_1)\big/f(\boldsymbol{X}\,|\,\boldsymbol{\Theta}_0)$$

ordnen die in den kritischen Bereich K_α aufzunehmenden Punkte $\boldsymbol{x}$ im Stichprobenraum identisch mit der möglichen Ausnahme jener Punkte, für die $\lambda(\boldsymbol{x}) = 1$. In den meisten Anwendungsfällen mit kleinem α stimmen beide Tests überein, da $\lambda(\boldsymbol{x}) = 1$ bedeutet, dass die Likelihood unter Ω_0 mindestens so groß ist wie unter Ω_1 und daher jene $\boldsymbol{x}$ bei beiden Tests in den Annahmebereich fallen.

Die Bestimmung des Ablehnbereichs eines LQ–Tests kann im Einzelfall schwierig sein, da die Verteilung von $\lambda(\boldsymbol{X})$ nicht immer leicht zu ermitteln ist. Bisweilen lässt sich, wie in den beiden folgenden Beispielen gezeigt, die Verteilung einer geeigneten Transformation von $\lambda(\boldsymbol{X})$ finden.

1. Beispiel: Seien $X_i \overset{iid}{\sim} No(\mu;\sigma^2);\ i=1,\dots,n$; mit bekanntem σ^2 und unbekanntem μ. Zu testen ist die einfache Hypothese $H_0\colon\ \mu=\mu_0$ gegen die einfache Alternative $H_1\colon\ \mu=\mu_1$, wobei o. B. d. A. $\mu_1 > \mu$ gelten möge. Dann ist:

$$\begin{aligned}\boldsymbol{\Theta} &= (\mu),\\ \Omega &= \{\mu_0;\mu_1\},\\ \Omega_0 &= \{\mu_0\},\\ f(\boldsymbol{X}\,|\,\boldsymbol{\Theta}) = f(\boldsymbol{X}\,|\,\mu) &= (2\,\pi\,\sigma^2)^{-n/2}\,\exp\big[-\Sigma(X_i-\mu)^2\big/(2\,\sigma^2)\big].\end{aligned}$$

Weil hier die Angabe des Ablehnbereichs (†) durch

$$\frac{f(\boldsymbol{X}\,|\,\mu_0)}{\max\big\{f(\boldsymbol{X}\,|\,\mu_0);f(\boldsymbol{X}\,|\,\mu_1)\big\}} < c$$

äquivalent ist mit

$$LQ := \frac{f(\boldsymbol{X}\,|\,\mu_1)}{f(\boldsymbol{X}\,|\,\mu_0)} > c^*,$$

kann K_α auch über den Quotienten LQ gefunden werden, also

$$LQ = \frac{f(\boldsymbol{X}\,|\,\mu_1)}{f(\boldsymbol{X}\,|\,\mu_0)} = \exp\left[-\frac{1}{2\,\sigma^2}\left(\sum_{i=1}^n (X_i-\mu_0)^2 - \sum_{i=1}^n (X_i-\mu_1)^2\right)\right] > c^*.$$

[10] NEYMAN, J. / PEARSON, E.S. (1928): Biometrika **20A**, 175–240 und 263–294.
[11] NEYMAN, J. / PEARSON, E.S. (1933): Philos. Trans. Roy. Soc. **A231**, 289–337.

Aus dieser Ungleichung erhält man unter Beachtung von $\sum X_i = n\,\overline{X}_n$ für die handlichere Prüfgröße $\overline{X}_n$ äquivalent:

$$\overline{X}_n > \frac{2\,\sigma^2\,\ln c^* + n\,(\mu_1^2 - \mu_0^2)}{2\,n\,(\mu_1 - \mu_0)} =: C\,.$$

Soll der Test das Signifikanzniveau α haben, ist C so zu wählen, dass gilt

$$\Pr(\overline{X}_n > C \,|\, \mu_0) = \alpha,$$

d. h.

$$\begin{aligned} 1 - \Phi\left(\frac{C-\mu_0}{\sigma/\sqrt{n}}\right) &= \alpha \qquad (*)\\ \iff \frac{C-\mu_0}{\sigma/\sqrt{n}} &= u_{1-\alpha}\\ \iff C &= \mu_0 + u_{1-\alpha}\,\frac{\sigma}{\sqrt{n}}\,. \end{aligned}$$

Das ist die Grenze des kritischen Bereichs eines einseitigen GAUSS–Tests, vgl. Abs. C3.2.1. Arbeitet man mit der originären Prüfgröße LQ, so ist der zugehörige kritische Wert bei einem Signifikanzniveau α:

$$c^*(\alpha) = \exp\left\{\frac{n\,(\mu_1-\mu_0)\left[2\,u_{1-\alpha}\,\dfrac{\sigma}{\sqrt{n}} - (\mu_1-\mu_0)\right]}{2\,\sigma^2}\right\}.$$

Die Wahrscheinlichkeit des Fehlers 2. Art ist

$$\beta = \Pr(\overline{X}_n \le C \,|\, \mu_1) = \Phi\left(\frac{C-\mu_1}{\sigma/\sqrt{n}}\right). \qquad (**)$$

Mit $(*)$ und $(**)$ hat man zwei Gleichungen, in denen die vier Größen $\alpha,\ \beta,\ C$ und n vorkommen. Es gibt $\binom{4}{2} = 6$ Möglichkeiten, aus diesen vier Größen zwei vorzugeben und dann die anderen beiden zu berechnen.

1) Vorgabe von α und β

$$n = -\left[-\frac{\sigma^2}{(\mu_0-\mu_1)^2}(u_\beta - u_{1-\alpha})^2\right]_G,$$

$$C = \mu_0 + u_{1-\alpha}\,\sigma/\sqrt{n}\,; \qquad (+)$$

2) Vorgabe von α und C

$$n = -\left[-\frac{\sigma^2\,u_{1-\alpha}^2}{(A-\mu_0)^2}\right]_G,$$

β gemäß $(**)$;

3) Vorgabe von β und C

$$n = -\left[-\frac{\sigma^2\,u_\beta^2}{(A-\mu_1)^2}\right]_G,$$

α gemäß $(*)$;

4) Vorgabe von α und n (Vorgehensweise des Signifikanztests)

C aus (+),

β aus $(**)$;

5) <u>Vorgabe von β und n</u>
   $C = \mu_1 + u_\beta\, \sigma/\sqrt{n}$,
   α aus $(*)$;
6) <u>Vorgabe von C und n</u>
   α aus $(*)$,
   β aus $(**)$.

<u>Bemerkung:</u> In den Fällen 1) bis 3) wird n aufgerundet, d. h. man geht auf die sichere Seite, so dass die zugehörigen effektiven Fehlerwahrscheinlichkeiten nicht größer werden als die Vorgaben α und/oder β.

<u>2. Beispiel:</u> Seien $X_i \overset{iid}{\sim} No(\mu; \sigma^2);\ i = 1, \ldots, n$; mit unbekanntem μ und σ^2. Zu testen ist

$$H_0\colon\ \mu = 0 \ \text{ gegen } \ H_1\colon\ \mu \neq 0.$$

Dann ist

$$\begin{aligned}
\boldsymbol{\Theta} &= (\mu; \sigma^2)', \\
\Omega &= \{(\mu; \sigma^2) \mid \mu \in \mathbb{R};\ \sigma^2 \in \mathbb{R}^+\}, \\
\Omega_0 &= \{\mu; \sigma^2) \mid \mu = 0;\ \sigma^2 \in \mathbb{R}^+\}, \\
f(\boldsymbol{X} \mid \boldsymbol{\Theta}) &= (2\,\pi\,\sigma^2)^{-n/2} \exp\big[-\Sigma(X_i - \mu)^2/(2\,\sigma^2)\big].
\end{aligned}$$

Der ML–Schätzer von $\boldsymbol{\Theta}$ unter Ω_0 ist

$$\widehat{\boldsymbol{\Theta}}_{\Omega_0} = \left(0;\ \frac{1}{n}\sum_{i=1}^{n} X_i^2\right)' = \left(0;\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 + \overline{X}_n^2\right)'$$

und unter Ω

$$\widehat{\boldsymbol{\Theta}}_{\Omega} = \left(\overline{X}_n;\ \frac{1}{n}\sum(X_i - \overline{X}_n)^2\right)'.$$

Also gelten

$$\sup_{\boldsymbol{\Theta}\in\Omega_0} f(\boldsymbol{X} \mid \boldsymbol{\Theta}) = f(\boldsymbol{X} \mid \widehat{\boldsymbol{\Theta}}_{\Omega_0}) = \left\{2\,\pi\left[\frac{1}{n}\,\Sigma(X_i - \overline{X}_n)^2 + \overline{X}_n^2\right]\right\}^{-n/2} e^{-n/2},$$

$$\sup_{\boldsymbol{\Theta}\in\Omega} f(\boldsymbol{X} \mid \boldsymbol{\Theta}) = f(\boldsymbol{X} \mid \widehat{\boldsymbol{\Theta}}_{\Omega}) = \left\{2\,\pi\left[\frac{1}{n}\,\Sigma(X_i - \overline{X}_n)^2\right]\right\}^{-n/2} e^{-n/2}$$

und somit

$$\lambda(\boldsymbol{X}) = \left\{\frac{\frac{1}{n}\big[\Sigma(X_i - \overline{X}_n)^2 + \overline{X}_n^2\big]}{\frac{1}{n}\Sigma(X_i - \overline{X}_n)^2}\right\}^{-n/2}.$$

Da H_0 zu verwerfen ist, wenn $\lambda(\boldsymbol{X}) > c$, folgt hier als Ablehnbedingung

$$\frac{1}{n}\,\frac{\overline{X}_n^2}{\Sigma(X_i - \overline{X}_n)^2} > c^{-2/n} - 1$$

oder nach Umformung

$$|\,t\,| := \frac{\sqrt{n}\,\big|\overline{X}_n\big|}{\sqrt{\dfrac{1}{n-1}\,\Sigma(X_i - \overline{X}_n)^2}} > \sqrt{(n-1)\left(c^{-2/n} - 1\right)}.$$

Damit hat man den t–Test (vgl. Abs. C3.2.1) zum Niveau α, wenn man

$$\sqrt{(n-1)\left(c^{-2/n}-1\right)} = t_{1-\alpha/2;n-1}$$

wählt. Daraus ergibt sich c als

$$c = \sqrt{\left(\frac{n-1}{t^2_{1-\alpha/2;n-1}+n-1}\right)^n}.$$

In Situationen, wo man nicht direkt oder über eine Transformation die Verteilung von $\lambda(\boldsymbol{X})$ bestimmen kann, muss man auf asymptotische Ergebnisse zurückgreifen und den Ablehnbereich approximativ bestimmen. Dabei stützt man sich auf ein Ergebnis von C.R. RAO (1973).

Sei $\boldsymbol{X} = (X_1, \ldots, X_n)'$ eine Stichprobe mit der Dichte[12] $f(\boldsymbol{x} \mid \boldsymbol{\Theta})$, die hinreichend regulär ist, so dass der ML–Schätzer $\widehat{\boldsymbol{\Theta}}$ als asymptotisch (multivariat) normalverteilt gelten kann. Für die m Parameter Θ_i in $\boldsymbol{\Theta}$ sollen — Das ist nun die Nullhypothese! — ℓ unabhängig lineare Restriktionen $(\ell < m)$ in Gleichungsform vorliegen:

$$\boldsymbol{B}\,\boldsymbol{\Theta} = \boldsymbol{b}$$

mit $\boldsymbol{B}$ als $(\ell \times m)$–Matrix und $\boldsymbol{b}$ als ℓ–elementigem Vektor. Der Teilvektorraum Ω_0 unter H_0 besteht aus allen Vektoren, die $\boldsymbol{B}\,\boldsymbol{\Theta} = \boldsymbol{b}$ erfüllen:

$$\Omega_0 = \{\boldsymbol{\Theta} \mid \boldsymbol{\Theta} \in \Omega \wedge \boldsymbol{B}\,\boldsymbol{\Theta} = \boldsymbol{b}\}.$$

Es ist

$$H_0\colon\ \boldsymbol{\Theta} \in \Omega_0 \ \text{ gegen } \ H_1\colon\ \boldsymbol{\Theta} \in \Omega\backslash\Omega_0$$

zu testen. RAO hat gezeigt, dass unter H_0 gilt:

$$LR := -2\,\ln\left[\lambda(\boldsymbol{X})\right] \overset{approx}{\sim} \chi^2(\ell),$$

d. h. H_0 ist auf dem Niveau α zu verwerfen, wenn

$$LR = -2\,\ln\left[\sup_{\Theta\in\Omega} f(\boldsymbol{X} \mid \boldsymbol{\Theta})\right] - 2\,\ln\left[\sup_{\Theta\in\Omega_0} f(\boldsymbol{X} \mid \boldsymbol{\Theta})\right] > \chi^2_{\ell;1-\alpha}.$$

Wichtig für die asymptotische Verteilung von $-2\,\ln\left[\lambda(\boldsymbol{X})\right]$ ist, dass die Nullhypothese die Form von Gleichheits- und nicht etwa von Ungleichheitsrestriktionen hat.

3.1.5 WALD– und LAGRANGE–Multiplikatoren–Tests

Zu der asymptotischen χ^2–Variante des LQ–Tests gibt es zwei asymptotisch äquivalente Verfahren, um Nullhypothesen über $\boldsymbol{\Theta}$ in Form von Gleichheitsrestriktionen zu testen.

Bei dem von A. WALD (1902–1950) vorgeschlagenen Verfahren (vgl. Ann. Math. Statist. **16**, 117–186) zum Testen von ℓ Nebenbedingungen über die $m > \ell$ Parameter Θ_i in $\boldsymbol{\Theta}$ unter H_0 in der Form

$$\boldsymbol{B}\,\boldsymbol{\Theta} = \boldsymbol{b}$$

benötigt man den unrestringierten ML–Schätzer $\widehat{\boldsymbol{\Theta}}$. Die Prüfgröße des **WALD–Tests** lautet

$$W := (\boldsymbol{B}\,\widehat{\boldsymbol{\Theta}} - \boldsymbol{b})'\left[\boldsymbol{B}\,\boldsymbol{I}^{-1}(\widehat{\boldsymbol{\Theta}})\,\boldsymbol{B}'\right]^{-1}(\boldsymbol{B}\,\widehat{\boldsymbol{\Theta}} - \boldsymbol{b}),$$

[12] Bei diskretem $\boldsymbol{X}$ ist von der gemeinsamen Wahrscheinlichkeitsfunktion $\Pr(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{\Theta})$ auszugehen.

mit $\boldsymbol{I}(\widehat{\boldsymbol{\Theta}})$ als FISHER–Informationsmatrix, die mit dem ML–Schätzer $\widehat{\boldsymbol{\Theta}}$ für $\boldsymbol{\Theta}$ auszuwerten ist. Unter H_0 gilt

$$W \overset{\text{approx}}{\sim} \chi^2(\ell).$$

Der **LAGRANGE–Multiplikator–Test** (LM–Test), auch **Score–Test** genannt, baut die unter H_0 stehenden Nebenbedingungen $\boldsymbol{B}\,\boldsymbol{\Theta} = \boldsymbol{b}$ in die zu maximierende Log–Likelihood-Funktion ein, erweitert diese also zur LAGRANGE–Funktion, und schätzt den Parametervektor unter den Nebenbedingungen. Mit $\widehat{\boldsymbol{\Theta}}_R$ sei dieser restringierte Schätzer bezeichnet. Die Prüfgröße

$$LM := \mathcal{D}(\widehat{\boldsymbol{\Theta}}_R)'\,\boldsymbol{I}^{-1}(\widehat{\boldsymbol{\Theta}}_R)\,\mathcal{D}(\widehat{\boldsymbol{\theta}}_R)$$

ist unter H_0 wiederum approximativ χ^2–verteilt:

$$LM \overset{\text{approx}}{\sim} \chi^2(\ell).$$

Es ist

$$\mathcal{D}(\widehat{\boldsymbol{\Theta}}_R) = \frac{\partial \ln[f(\boldsymbol{X}\,|\,\boldsymbol{\Theta})]}{\partial \boldsymbol{\Theta}}\bigg|_{\boldsymbol{\Theta}=\widehat{\boldsymbol{\Theta}}_R},$$

also die Ableitung der Log–Likelihood–Funktion an der Stelle des restringierten Schätzers $\widehat{\boldsymbol{\Theta}}_R$. Entsprechend ist $\boldsymbol{I}(\widehat{\boldsymbol{\Theta}}_R)$ die mit $\widehat{\boldsymbol{\Theta}}_R$ ausgewertete FISHER–Informationsmatrix.

Unter H_0 sind der LQ–, der WALD– und der LM–Test asymptotisch äquivalent. Während der WALD–Test ausschließlich mit dem unrestringierten Schätzer und der LM–Test nur mit dem restringierten Schätzer arbeitet, benötigt der LQ–Test beide Schätzer.

3.1.6 Sequentieller Likelihood–Quotienten–Test

Der klassische, auf NEYMAN und PEARSON zurückgehende LQ–Test arbeitet mit einem **festen Stichprobenumfang** n und **einem kritischen Wert** $c^*(\alpha)$. Der 1947 von A. WALD[13] eingeführte **sequentielle Likelihood–Quotienten–Test** (SPRT) *[sequential probability ratio test]* arbeitet mit **zwei kritischen Werten** $A = A(\alpha, \beta)$ und $B = B(\alpha, \beta)$ und einem **variablen Stichprobenumfang**, der zufällig ist. Der SPRT besitzt folgende Entscheidungsregel:

$$LQ_n(\boldsymbol{X}) = \frac{f(\boldsymbol{X}\,|\,\Theta_1)}{f(\boldsymbol{X}\,|\,\Theta_0)} \begin{cases} \gg 1 \Longrightarrow H_1 \text{ annehmen,} \\ \ll 1 \Longrightarrow H_0 \text{ annehmen,} \\ \approx 1 \Longrightarrow \text{neue Beobachtung } x_{n+1} \text{ ziehen.} \end{cases}$$

Vorgehensweise:

1. Wähle $A,\ B$ mit $0 < A < 1 < B$, setze $n = 1$ und ziehe das erste Stichprobenelement.
2. Nimm H_0 an, falls $LQ_n(\boldsymbol{X}) \le A$. (Ende des Tests)
3. Nimm H_1 an, falls $LQ_n(\boldsymbol{X}) \ge B$. (Ende des Tests)
4. Ziehe ein weiteres Stichprobenelement, wenn $A < LQ_n(\boldsymbol{X}) < B$ (Fortsetzung des Tests), und bilde $LQ_{n+1}(\boldsymbol{X})$. Setze $LQ_{n+1}(\boldsymbol{X}) = LQ_n(\boldsymbol{X})$ und gehe zu 2.

Mit

$$A = \frac{\beta}{1-\alpha}, \quad B = \frac{1-\beta}{\alpha}$$

erhält man einen konservativen Test. I. d. R. wird mit dem logarithmierten Likelihood–Quotienten

$$\ln\big[LQ_n(\boldsymbol{X})\big] = \begin{cases} \sum\limits_{i=1}^{n} \big\{ \ln\big[f(X_i\,|\,\Theta_1)\big] - \ln\big[f(X_i\,|\,\Theta_0)\big]\big\} & - X \text{ stetig,} \\ \sum\limits_{i=1}^{n} \big\{ \ln\big[\Pr(X = X_i\,|\,\Theta_1)\big] - \ln\big[\Pr(X = X_i\,|\,\Theta_0)\big]\big\} & - X \text{ diskret,} \end{cases}$$

[13] Die Originalarbeit entstand bereits um 1942 als Forschungsauftrag (Auftraggeber: die US–Militärverwaltung), in der gezeigt wurde, wie man bei der zerstörenden Qualitätsprüfung den Stichprobenumfang bei gegebener OC–Funktion reduzieren kann. Als Top–Secret konnten die Ergebnisse erst nach Kriegsende publiziert werden.

gearbeitet und dieser verglichen mit

$$b := \ln B, \quad a := \ln A\,.$$

Unter allen Tests mit nicht größeren Wahrscheinlichkeiten α und β für die Fehler erster und zweiter Art ist der erwartete Stichprobenumfang beim SPRT am kleinsten. Diese Aussage macht den SPRT bei der zerstörenden Qualitätsprüfung und teueren Testobjekten interessant.

Beispiel: Die Arbeitsweise eines SPRT sei ausführlicher an der Überprüfung von

$$H_0\colon\ \mu \leq \mu_0 \ \text{ gegen } \ H_1\colon\ \mu \geq \mu_1 \ \text{ mit } \ \mu_1 \geq \mu_0$$

und $X_i \overset{iid}{\sim} No(\mu;\sigma^2)$ bei bekannter Varianz σ^2 gezeigt. Aus dem Fortsetzungsbereich für $LQ_n(\boldsymbol{X})$

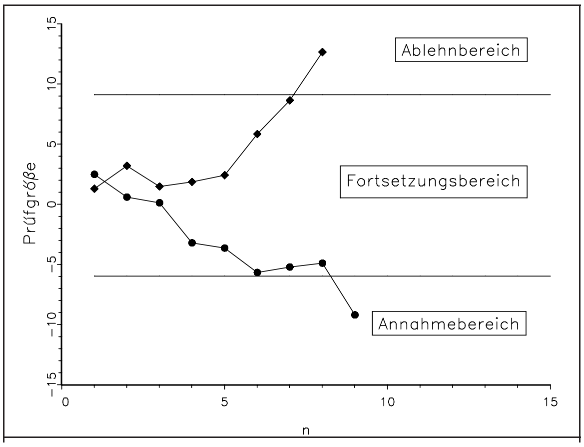
$$A \ < \ \frac{\exp\left[-\dfrac{\Sigma(X_i-\mu_1)^2}{2\,\sigma^2}\right]}{\exp\left[-\dfrac{\Sigma(X_i-\mu_0)^2}{2\,\sigma^2}\right]} \ < \ B$$

erhält man nach Logarithmierung und Umformung

$$\frac{\sigma^2}{\mu_1-\mu_0}\,\ln A < \underbrace{n\left(\overline{X}_n - \frac{\mu_0-\mu_1}{2}\right)}_{\text{Prüfgröße}} < \frac{\sigma^2}{\mu_1-\mu_0}\,\ln B.$$

Zur graphischen Durchführung des Tests trägt man die Prüfgröße (auf der Ordinate) über n (auf der Abszisse) ab. Die Grenzen des Fortsetzungsbereichs sind zwei achsenparallele Geraden mit dem Abstand $\sigma^2\,[\ln B - \ln A]\big/[\mu_1-\mu_0]$, vgl. Abb. C3/5 mit zwei eingezeichneten Prüftrajektorien.

Abb. C3/5: Skizze zur Durchführung eines SPRT mit zwei Prüftrajektorien



Unter Verwendung von

$$h := \frac{\mu_0+\mu_1-2\,\mu}{\mu_1-\mu_0} \quad \text{und} \quad k := \frac{-h\,(\mu_1-\mu_0)^2}{2\,\sigma^2}$$

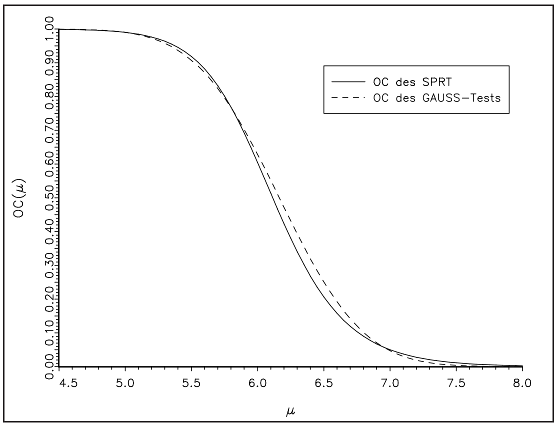
ergibt sich die OC–Funktion als

$$OC(\mu) = \begin{cases} \dfrac{B^h - 1}{B^h - A^h} & \text{für} \quad \mu \neq \dfrac{\mu_0 + \mu_1}{2}, \\[2ex] \dfrac{\ln B}{\ln B - \ln A} & \text{für} \quad \mu = \dfrac{\mu_0 + \mu_1}{2}, \end{cases}$$

und die ASN–Funktion *[average sample number]* für den erwarteten Stichprobenumfang

$$ASN(\mu) = \begin{cases} \dfrac{OC(\mu)\ln A + \big[1 - OC(\mu)\big]\ln B}{k} & \text{für} \quad \mu \neq \dfrac{\mu_0 + \mu_1}{2}, \\[2ex] \dfrac{-(\ln A)\,(\ln B)\,\sigma^2}{(\mu_1 - \mu_0)^2} & \text{für} \quad \mu = \dfrac{\mu_0 + \mu_1}{2}. \end{cases}$$

Abb. C3/6: OC–Funktion eines SPRT und des OC–äquivalenten GAUSS–Tests



Für die Datenkonstellation

$$\mu_0 = 5; \quad \mu_1 = 7; \quad \alpha = 0{,}01; \quad \beta = 0{,}05; \quad \sigma = 2$$

zeigen Abb. C3/6 und C3/7 die Verläufe von $OC(\mu)$ und $ASN(\mu)$. Der OC–äquivalente GAUSS–Test mit

$$OC(\mu_0) = 1 - \alpha \quad \text{und} \quad OC(\mu_1) = \beta$$

und festem Stichprobenumfang hat — berechnet gemäß dem Beispiel in Abs. C3.1.4 — den Stichprobenumfang

$$n = -\left[-\frac{\sigma^2}{(\mu_0 - \mu_1)^2}\,\big(u_\beta - u_{1-\alpha}\big)^2\right]_G = -\left[-\frac{2^2}{(5-7)^2}\,(-1{,}6449 - 2{,}3263)^2\right]_G = 16$$

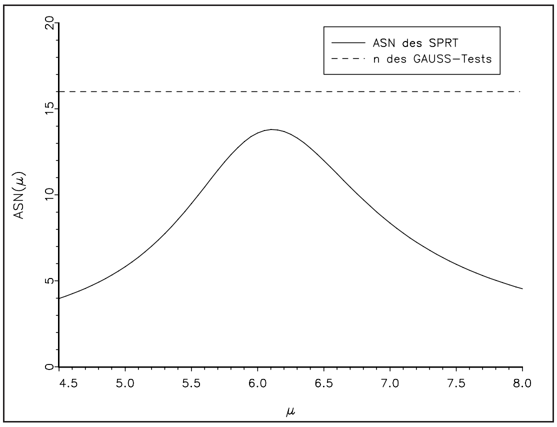
und den Ablehnbereich für $H_0\colon\ \mu \leq \mu_0 = 5$

$$\overline{X}_{16} > \mu_0 + u_{1-\alpha}\,\frac{\sigma}{\sqrt{16}} \approx 6{,}1632.$$

Man sieht, dass $n = 16 > \max\limits_{\mu}\big[ASN(\mu)\big] \approx 14$. GHOSH (1970) zeigt, dass

- für $0{,}005 < \alpha,\ \beta < 0{,}1$ gilt: $1{,}67\, ASN(\mu_0) < n < 3{,}29\, ASN(\mu_0)$,
- für $\alpha = \beta > 0{,}1$ gilt: $n > ASN(\mu)\ \forall\ \mu$.

Abb. C3/7: ASN–Funktion eines SPRT mit OC–äquivalentem festem Stichprobenumfang



Ist für eine BERNOULLI–Verteilung

$$H_0:\ P \leq P_0 \quad \text{gegen} \quad H_1:\ P \geq P_1 \ \text{ mit } \ P_1 > P_0$$

zu testen und wird eine Stichprobe mit Zurücklegen gezogen, so bekommt man für die Prüfgröße

$$X_n^T := \sum_{i=1}^{n} X_i, \ \text{ die kumulierte Anzahl der „Erfolge“},$$

den Fortsetzungsbereich

$$a + c \cdot n < X_n^T < a + b \cdot n$$

mit

$$a = \frac{\ln A}{\ln\left[\dfrac{P_1\,(1-P_0)}{P_0\,(1-P_1)}\right]}\,; \quad b = \frac{\ln B}{\ln\left[\dfrac{P_1\,(1-P_0)}{P_0\,(1-P_1)}\right]}\,; \quad c = \frac{\ln\left[\dfrac{1-P_0}{1-P_1}\right]}{\ln\left[\dfrac{P_1\,(1-P_0)}{P_0\,(1-P_1)}\right]}\,.$$

Zur Berechnung der OC– und ASN–Funktion benötigt man eine Hilfsfunktion für $h \in \mathbb{R}$:

$$P = \psi(h) = \begin{cases} \dfrac{\ln\left(\dfrac{1-P_0}{1-P_1}\right)^h - 1}{\left[\dfrac{P_1\,(1-P_0)}{P_0\,(1-P_1)}\right]^h - 1} & \text{für } \ h \neq 0, \\ \qquad c & \text{für } \ h = 0. \end{cases}$$

Dann ist

$$OC(P) = OC\big[\psi(h)\big] = \begin{cases} \dfrac{\left(\dfrac{\alpha}{1-\beta}\right)^h - 1}{\left[\dfrac{\alpha\,\beta}{(1-\alpha)\,(1-\beta)}\right]^h - 1} & \text{für } \; P \neq c, \;\; \text{d.h. } \; h \neq 0, \\ \dfrac{\ln\left(\dfrac{\alpha}{1-\beta}\right)}{\ln\left[\dfrac{\alpha\,\beta}{(1-\alpha)\,(1-\beta)}\right]} & \text{für } \; P = c, \;\; \text{d.h. } \; h = 0, \end{cases}$$

und

$$ASN(P) = \begin{cases} \dfrac{b - (b-a)\,OC(P)}{P - c} & \text{für } \; P \neq c, \\ \dfrac{b\,a}{c\,(c-1)} & \text{für } \; P = c. \end{cases}$$

Ist für eine POISSON–Verteilung

$$H_0\colon\; \lambda \leq \lambda_0 \quad \text{gegen} \quad H_1\colon\; \lambda \geq \lambda_1 \quad \text{mit} \quad \lambda_1 > \lambda_0$$

zu testen, so ist für die Prüfgröße

$$X_n^T := \sum_{i=1}^{n} X_i, \quad \text{die kumulierte Anzahl der Ereignisse,}$$

der Fortsetzungsbereich durch

$$a^* + c^*\,n < X_n^T < b^* + c^*\,n$$

mit

$$a^* = \frac{\ln A}{\ln(\lambda_1/\lambda_0)}\,; \;\; b^* = \frac{\ln B}{\ln(\lambda_1/\lambda_0)}\,; \;\; c^* = \frac{\lambda_1 - \lambda_0}{\ln(\lambda_1/\lambda_0)}$$

gegeben. Zur Berechnung der OC– und ASN–Funktion benötigt man eine Hilfsfunktion für $h \in \mathbb{R}$:

$$\lambda = \psi^*(h) = \begin{cases} \dfrac{\lambda_1 - \lambda_0}{\ln(\lambda_1 - \lambda_0)}\,\dfrac{\ln h}{h-1} & \text{für } \; h \neq 1, \\ c^* & \text{für } \; h = 1. \end{cases}$$

Dann ist

$$OC(\lambda) = OC\big[\psi^*(h)\big] = \begin{cases} \dfrac{1 - h^{b^*}}{h^{a^*} - h^{b^*}} & \text{für } \lambda \neq c^*, \;\; \text{d.h. } \; h \neq 1, \\ \dfrac{b^*}{b^* - a^*} & \text{für } \lambda = c^*, \;\; \text{d.h. } \; h = 1, \end{cases}$$

und

$$ASN(\lambda) = \begin{cases} \dfrac{b^* - (b^* - a^*)\,OC(\lambda)}{\lambda - c^*} & \text{für } \; \lambda \neq c^*, \\ \dfrac{b^*\,(-a^*)}{c^*} & \text{für } \; \lambda = c^*. \end{cases}$$

3.1.7 Randomisierte Tests

Bei einem Test mit diskreter Prüfgröße G lässt sich ein vorgegebenes Signifikanzniveau α nicht immer und durch ausschließliche Vorgabe des Ablehnbereichs K_α einhalten. Um α voll auszuschöpfen, muss die Gütefunktion des konservativen, α unterschreitenden Tests durch einen Trick, **Randomisierung** genannt, angehoben werden.

Seien g_i $(i = 1, \ldots, I)$ die möglichen äquidistanten Realisationen von G. Für den **zweiseitigen Test** von $H_0 : \ \Theta = \Theta_0$ mit $H_1 : \ \Theta \neq \Theta_0$ sind die konservativen Eingriffsgrenzen mit g_u und g_o berechnet, und es gelten

$$\text{Pr}(G \leq g_u \,|\, \Theta_0) = \alpha_u \leq \frac{\alpha}{2}\,,$$
$$\text{Pr}(G \geq g_o \,|\, \Theta_0) = \alpha_o \leq \frac{\alpha}{2}\,.$$

Dann lautet die Entscheidungsregel des randomisierten zweiseitigen Tests:

1) Ist $G \leq g_u$ oder $G \geq g_o$ eingetreten, so lehne H_0 ab.

2) Ist $g_{u+1} < G < g_{o-1}$ eingetreten, so lehne H_0 nicht ab.

3a) Ist $G = g_{u+1}$ eingetreten, so lehne H_0 mit der Wahrscheinlichkeit

$$\alpha_u^* = \frac{\frac{\alpha}{2} - \alpha_u}{\text{Pr}(G = g_{u+1} \,|\, \Theta_0)}$$

ab.

3b) Ist $G = g_{o-1}$ eingetreten, so lehne H_0 mit der Wahrscheinlichkeit

$$\alpha_o^* = \frac{\frac{\alpha}{2} - \alpha_o}{\text{Pr}(G = g_{o-1} \,|\, \Theta_0)}$$

ab.

4) In den Fällen 3a) oder 3b) muss eine in $[0; 1]$ gleichverteilte Zufallszahl Z erzeugt werden (Das ist die Randomisierung!), und es wird H_0 abgelehnt, wenn

$$Z \leq \alpha_u^* \quad \text{bzw.} \quad Z \leq \alpha_o^*.$$

Beim **einseitigen Test** von $H_0 : \ \Theta \leq \Theta_0$ gegen $H_1 : \ \Theta > \Theta_0$ mit der konservativen Eingriffsgrenze g_o und

$$\text{Pr}(G \geq g_o \,|\, \Theta_0) = \alpha_o \leq \alpha$$

ist die Entscheidungsregel:

- Fällt $G \geq g_o$ aus, lehne H_0 ab.
- Fällt $G < g_{o-1}$ aus, lehne H_0 nicht ab.
- Realisiert sich g_{o-1}, so lehne H_0 mit der Wahrscheinlichkeit

$$\alpha_o^* = \frac{\alpha - \alpha_o}{\text{Pr}(G = g_{o-1} \,|\, \Theta_0)}$$

ab.

Beim einseitigen Test von $H_0 : \ \Theta \geq \Theta_0$ gegen $H_1 : \ \Theta < \Theta_0$ mit der konservativen Eingriffsgrenze g_u und

$$\text{Pr}(G \leq g_u \,|\, \Theta_0) = \alpha_u \leq \alpha$$

heißt es:

- Im Fall $G \leq g_u$ lehne H_0 ab.
- Im Fall $G > g_{u+1}$ lehne H_0 nicht ab.

- Im Fall $G = g_{u+1}$ lehne H_0 mit Wahrscheinlichkeit

$$\alpha_u^* = \frac{\alpha - \alpha_u}{\Pr(G = g_{u+1} \mid \Theta_0)}$$

ab.

3.2 Verteilungsgebundene Parametertests

Die nachfolgend zusammengestellten Testverfahren gehen von einer bestimmten Verteilung in der Grundgesamtheit aus, in den meisten Fällen von einer Normalverteilung, oder verwenden letztere approximativ als Prüfgrößenverteilung.

3.2.1 Tests für einen Mittelwert

Tab. C3/2: GAUSS–Test — Hypothesen über μ einer Normalverteilung bei bekanntem σ^2

H_0	H_1	Prüfgröße	Ablehnbereich	Gütefunktion $GF(\mu)$
$\mu \le \mu_0$	$\mu > \mu_0$	$\frac{(\overline{X}_n - \mu_0)\sqrt{n}}{\sigma}$	$(u_{1-\alpha}; \infty)$	$\Phi\left[-u_{1-\alpha} - \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right]$
$\mu \ge \mu_0$	$\mu < \mu_0$	$\frac{(\overline{X}_n - \mu_0)\sqrt{n}}{\sigma}$	$(-\infty; -u_{1-\alpha})$	$\Phi\left[-u_{1-\alpha} + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right]$
$\mu = \mu_0$	$\mu \ne \mu_0$	$\frac{\lvert\overline{X}_n - \mu_0\rvert\sqrt{n}}{\sigma}$	$(u_{1-\alpha/2}; \infty)$	$\Phi\left[-u_{1-\alpha/2} + \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right] + {}$ $+\,\Phi\left[-u_{1-\alpha/2} - \frac{(\mu_0 - \mu)\sqrt{n}}{\sigma}\right]$

Tab. C3/3: t–Test — Hypothesen über μ einer Normalverteilung bei unbekanntem σ^2

H_0	H_1	Prüfgröße	Ablehnbereich	Gütefunktion
$\mu \le \mu_0$	$\mu > \mu_0$	$\frac{(\overline{X}_n - \mu_0)\sqrt{n}}{S_n}$	$(t_{n-1;1-\alpha}; \infty)$	vgl. unten
$\mu \ge \mu_0$	$\mu < \mu_0$	$\frac{(\overline{X}_n - \mu_0)\sqrt{n}}{S_n}$	$(-\infty; -t_{n-1;1-\alpha})$	vgl. unten
$\mu = \mu_0$	$\mu \ne \mu_0$	$\frac{\lvert\overline{X}_n - \mu_0\rvert\sqrt{n}}{S_n}$	$(t_{n-1;1-\alpha/2}; \infty)$	vgl. unten

Bemerkungen zum t–Test:

1. $S_n := \sqrt{\frac{1}{n-1}\sum(X_i - \overline{X}_n)^2}$
2. Sind die X_i nicht normalverteilt, sondern iid mit $\mu = \mathrm{E}(X_i)$ und bekanntem $\sigma^2 = \mathrm{Var}(X_i)\ \forall\, i$, so lässt sich bei nicht zu kleinem n approximativ nach Tab. C3/2 vorgehen. Ist σ^2 unbekannt, lässt sich approximativ mit Tab. C3/3 arbeiten, da der t–Test gegen Abweichungen von der Normalverteilung **robust** ist.