Neyman-Pearson

- Schulen statistischer Inferenz -

Christian Schnell
betreut von Dr. Marco Cattaneo

Seminar für methodologische und historische Grundlagen
des Wahrscheinlichkeitsbegriffs und der statistischen Inferenz

Institut für Statistik
Ludwig-Maximilians-Universität
München

11. März 2013

Gliederung

1 Einleitung

John Arbuthnot als Vorreiter des statistischen Tests

(190)

Anno.	Christened. Males.	Females.	Anno.	Christened. Males.	Females.
1667	5616	5322	1689	7604	7167
68	6073	5560	90	7909	7302
69	6506	5829	91	7662	7392
70	6278	5719	92	7602	7316
71	6449	6061	93	7676	7483
72	6443	6120	94	6985	6647
73	6073	5822	95	7263	6713
74	6113	5738	96	7632	7229
75	6058	5717	97	8062	7767
76	6552	5847	98	8426	7626
77	6423	6203	99	7911	7452
78	6568	6033	1700	7578	7061
79	6247	6041	1701	8102	7514
80	6548	6299	1702	8031	7656
81	6822	6533	1703	7765	7683
82	6909	6744	1704	6113	5738
83	7577	7158	1705	8366	7779
84	7575	7127	1706	7952	7417
85	7484	7246	1707	8379	7687
86	7575	7119	1708	8239	7623
87	7737	7214	1709	7840	7380
88	7487	7101	1710	7640	7288

Abbildung: Schotte John Arbuthnot (*1667 - †1735) und ein Ausschnitt der 82 Jahre der Londoner Geburtenstatistik

Arbuthnot in der Geschichte des Wahrscheinlichkeitsbegriffs

- **Huygens** (*1629 - †1695):
  Zusammenfassung seiner Erkenntnisse über Wahrscheinlichkeitsrechnung

  → „Van Rekeningh in Spelen van Geluck“ (1657)

- **Arbuthnot** (*1667 - †1735) :
  Übersetzung auf Englisch mit eigenen Anmerkungen

  → anonym unter dem Titel „Of the Laws of Chance“ (1692)

  Veröffentlichung seiner Arbeit „An argument for Divine Providence“ (1710)

  ⇒ Verwendung des Begriffs „Value of Expectation“ (nach Huygens)

- **Bernoulli** (*1655 - †1705) :
  Danach Veröffentlichung von „Ars Conjectandi“ (1713)

Arbuthnots Herangehensweise

- **Ansatz:** Überwiegen in einem Jahr Jungen- oder Mädchengeburten?
- **Beweisführung:**
  Annahme: zunächst Gleichverteilung der Geschlechter bei der Geburt

  $\rightarrow$ Berechnung wie bei einem Münzwurf („zweiseitiger Würfel“)

  $\rightarrow$ „Erwartungswert“: in einem Jahr werden mehr Jungen geboren $\hat{=} \frac{1}{2}$

  $\Rightarrow$ Wiederhole 82 mal: in jedem Jahr werden mehr Jungen geboren $\hat{=} (\frac{1}{2})^{82}$
- **Schlussfolgerung:** Annahme einer Gleichverteilung ist nicht gerechtfertigt!

  $\Rightarrow$ Arbuthnot: „göttliche Fügung“ ist bewiesen.

Aus der heutigen Sichtweise

Bemerkenswert:
Erweiterung einer logischen Schlussweise zu einer **statistischen Schlussweise**

$\Rightarrow$ Gegenteil dessen, was nachgewiesen werden soll als Nullhypothese:

> *„Die Wahrscheinlichkeit, dass ein neugeborenes Kind ein Junge bzw. ein Mädchen ist liegt jeweils bei 50%.“*

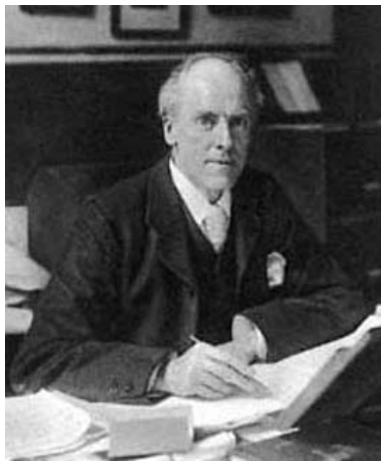
$\Rightarrow$ Durch mathematische Überlegungen folgt der Schluss:

> *„Die Wahrscheinlichkeit, dass ein neugeborenes Kind ein Junge ist liegt bei über 50% und ist somit größer als die Wahrscheinlichkeit, dass ein neugeborenes Kind ein Mädchen ist.“*

2 Neyman-Pearson Schule

2.1 Konstellation zwischen den wichtigsten Wissenschaftlern

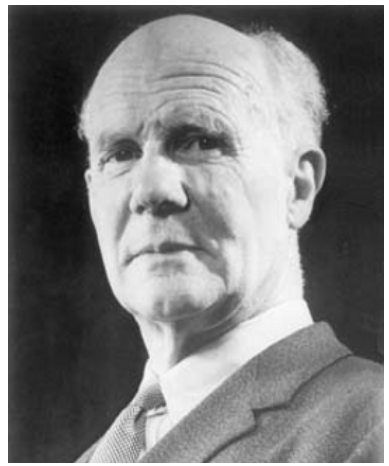
Karl Pearson
(*1857 - †1936)
→ * und † in England



Ronald Aylmer Fisher
(*1890 - †1962)
→ * in England
→ † in Australien

Jerzy Neyman
(*1894 - †1981)
→ * in Moldawien
→ † in USA/Kalifornien

Egon Sharpe Pearson
(*1895 - †1980)
→ * und † in England



Anfänge des Konflikts

Karl Pearson:

- Mitbegründer und Lektor des Magazins „Biometrika“ (ab 1901)
- Professur für Eugenik am University College in London (ab 1913)
- Gründung: „Department of Applied Statistics“
- Statistische Errungenschaften: Korrelations-/Kontingenzkoeffizient, χ^2-Test

Fisher:

- Stellenangebot am Institut von Pearson (1919)
- Forschungsarbeiten: Maximum-Likelihood, t-Test, Fisher-Test, Suffizienz, Varianzanalyse und Versuchsplanung

$\Rightarrow$ Karl Pearson blockierte Veröffentlichungen von Fisher in der „Biometrika“ aufgrund von inhaltlichen Uneinigkeiten bei der Auswertung von Stichproben

Verschärfung des Konflikts

Egon Sharpe Pearson:

- Ab 1924 am Institut und ab 1933 Institutsleitung gemeinsam mit Fisher
- Weiterhin Blockierung von Fisher in der „Biometrika“ als Lektor
- Intensive Zusammenarbeit mit Neyman von 1928 bis 1933[1]
- Anstellung von Neyman am Institut von 1934 bis 1938

[1] In dieser Zeit Veröffentlichung von solch grundlegenden Papers wie z.B.:
„On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference“ (1928)
„On the Problem of the Most Efficient Tests of Statistical Hypotheses“ (1933)

Höhepunkt des Konflikts (zwischen Neyman und Fisher)

Vorwurf an Fisher: Unsaubere Beweisführung und umständliche Formulierungen

→ Neyman über Fishers Nullhypothesentest:

„im mathematischen Sinne schlechter als nutzlos“

Vorwurf an Neyman: Beschränkt sich auf mathematische Denkweise

→ Fisher über Neyman:

„Neyman verstehe nichts von Statistik“ und
„hat keinen Bezug zu Naturwissenschaften“

Nullhypothese in Fishers Signifikanztest

- Aufwendige Berechnung der Verteilungstabellen beim t-Test (1908)
- **Festlegung des Signifikanzniveaus** auf 1% und 5% durch Fisher (1925)

  $\rightarrow$ anwenderfreundliche Verwendung des Signifikanztests von Fisher
- Benennung einer **Nullhypothese**: Unterscheidet sich die tatsächliche Verteilung einer Stichprobe von einer hypothetischen Verteilung?
- Allgemeine Vorgehensweise: Nullhypothese ohne bedeutenden Effekt

  $\rightarrow$ Ziel ist die **Verwerfung der Nullhypothese**

Alternativhypothese durch Neyman-Pearson

- Weiterentwicklung der Ideen von Fisher (Ende 1920er)
- Betrachtung von Hypothesenpaaren (mit **konkreter Alternativhypothese**):

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

- Disjunkte Parameterräume Θ_0 und Θ_1 können sich zu Θ ergänzen
- Einfachstes Szenario: H_0 und H_1 durch einzelne Parameterwerte repräsentiert
  $\rightarrow$ **„einfache“ Hypothesen** vs. „zusammengesetzte“ Hypothesen

Mögliche Testentscheidungen

	H_0 beibehalten	H_0 verwerfen
H_0 trifft zu	richtige Schlussfolgerung **Spezifität** (Wahrscheinlichkeit $1-\alpha$)	falsche Schlussfolgerung **Fehler 1. Art** (auch α-Fehler)
H_1 trifft zu	falsche Schlussfolgerung **Fehler 2. Art** (auch β-Fehler)	richtige Schlussfolgerung **Power** (Wahrscheinlichkeit $1-\beta$)

$$\Rightarrow P(\text{Fehler 1. Art}) = P(H_0 \text{ verwerfen} | H_0 \text{ trifft zu}) = \alpha$$

$$\begin{aligned}\Rightarrow P(\text{Fehler 2. Art}) &= P(H_0 \text{ beibehalten} | H_1 \text{ trifft zu}) \\ &= 1 - P(H_0 \text{ verwerfen} | H_1 \text{ trifft zu}) = \beta\end{aligned}$$

Fishers p-Wert

- Mögliche Annahme: kleine Werte sprechen **gegen Nullhypothese**
- p-Wert als **Maß** der Evidenz **gegen Nullhypothese**:

$$p = F(x) = P(X \leq x)$$

- Kumulierung der Wahrscheinlichkeit aller **möglichen** Beobachtungen kleiner oder gleich der eigentlichen Beobachtung x
- Je **kleiner** der p-Wert, desto **weniger plausibel** ist es, dass die beobachteten Werte für die Nullhypothese sprechen.

Beispiel: Geschlechterverhältnis bei der Geburt

- **Gleichverteilungsannahme unter H_0:**
  Erwarte bei 1000 Geburten ein Verhältnis von 500 Jungen zu 500 Mädchen
- **Beobachtung:** Verhältnis 950 zu 50 (Jungen (J) zu Mädchen (M))
- **p-Wert unter H_0:** $p = P(M = 0) + P(M = 1) + ... + P(M = m)$

  $\rightarrow$ mit der tatsächlichen Beobachtung $m = 50$

α-Niveau nach Neyman-Pearson

- **Annahme:** kleine Werte sprechen **gegen Nullhypothese**
  Zusätzlich: Standardnormalverteilte Prüfgröße Z mit α-Quantil z_α

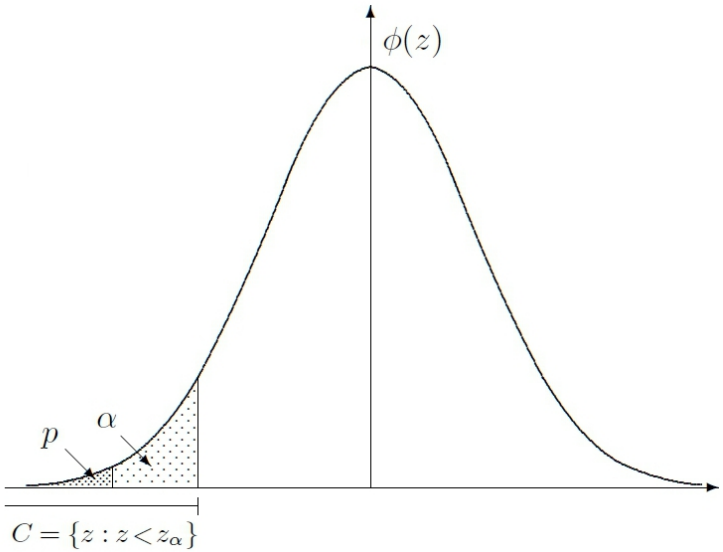
- **Ablehnbereich** der Nullhypothese: $P_{H_0}(Z \leq z_\alpha) = \alpha$

  $\rightarrow$ z_α entspricht dem kritischen Wert und wird durch α bestimmt

  $\rightarrow$ Kontrolle des Ablehnbereichs durch α

- Umso **größer** α gewählt wird, desto **weniger wahrscheinlich** spricht eine beobachtete Prüfgröße für die Nullhypothese.

Vergleich von p-Wert und α-Niveau



$\Rightarrow$ **Fisher:** H_0 wird abgelehnt, falls p-Wert kleiner als α-Niveau (z.B. 5%)

$\Rightarrow$ **Neyman-Pearson:** H_0 wird abgelehnt, falls der Prüfgrößenwert in den kritischen Bereich C fällt

Bestimmung des Signifikanzniveaus

1. Fisher (1935): Lege Signifikanzniveau im Sinne einer Konvention fest

   $\rightarrow$ vor der Durchführung des Tests

2. Fisher (1956): Berechne exaktes Signifikanzniveau aus den Daten (p-Wert)

   $\rightarrow$ nach der Durchführung des Tests

3. Neyman-Pearson: Lege α und β fest nach Kosten-Nutzen Analyse eines möglichen Fehlers 1. oder 2. Art

   $\rightarrow$ vor der Durchführung des Tests

Motivation für Kosten-Nutzen Analyse

Beispiel: H_0: „Angeklagter ist schuldig“ vs. H_1: „Angeklagter ist unschuldig“

$\rightarrow$ Fehler 1. Art (α): Angeklagte wird freigesprochen, obwohl er schuldig ist.

$\rightarrow$ Fehler 2. Art (β): Angeklagte wird verurteilt, obwohl er unschuldig ist.

Wichtig: Konsequenzen einer Fehleinschätzung

1. Fehler 1. Art: Gefahr für die Gesellschaft durch Beschuldigten
2. Fehler 2. Art: Art der Bestrafung (z.B. Todesstrafe oder Gefängnisstrafe)

Interpretation von Signifikanz

- Aussage bezieht sich auf Szenario, in dem die Nullhypothese Gültigkeit hat
- Wahrscheinlichkeit der Beobachtungen unter der Annahme: „H_0 trifft zu“
- Die beobachteten Werte sind so klein/groß, dass es **(extrem) unwahrscheinlich** ist (z.B. 5%), dass sie noch unter H_0 vorkommen.
- **Keine Aussage** über die Wahrscheinlichkeit von Hypothesen möglich

  $\rightarrow$ Nur im Rahmen der Bayes-Inferenz

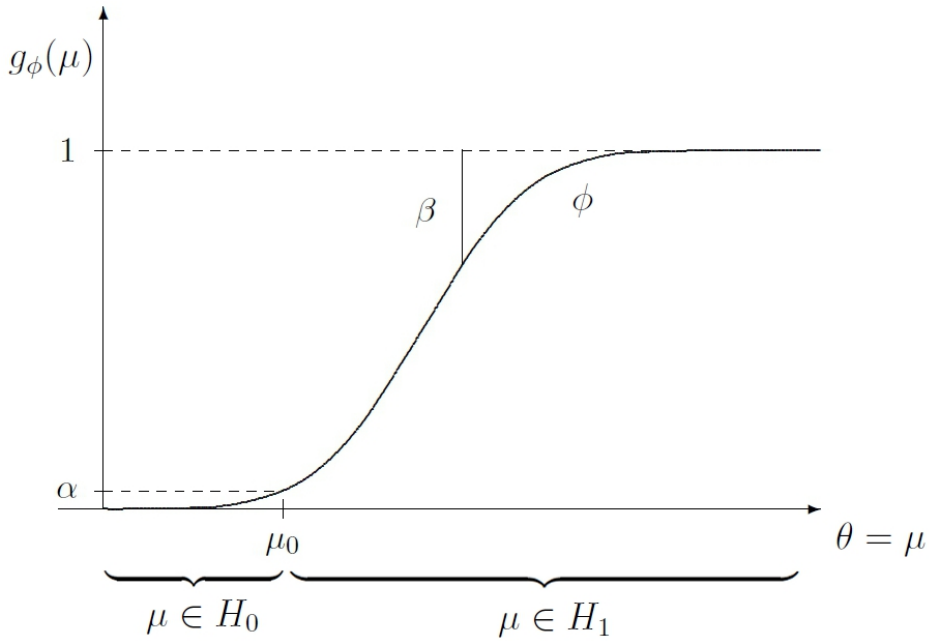
Zusammensetzung einer Gütefunktion

$$g_\phi(\theta) = P(H_0 \text{ verwerfen}|\theta) \leq \alpha, \quad \text{falls } \theta \in \Theta_0$$

$$1 - g_\phi(\theta) = 1 - P(H_0 \text{ verwerfen}|\theta) = \beta, \quad \text{falls } \theta \in \Theta_1$$

Testproblem einseitiger Gaußtest:

$H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$



$\rightarrow$ Gütefunktion $g_\phi(\mu) = 1 - \Phi(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma}\sqrt{n})$ eines einseitigen Gaußtests

Eigenschaften einer Gütefunktion

Beachte:
Ein Test ist **besser** als ein anderer Test, wenn er im Bereich der Alternative eine **größere Power** besitzt $\rightarrow$ tatsächlich existierende Effekte werden öfter entdeckt

1. Für Werte aus H_1 heißt die Gütefunktion Power eines Tests.
2. Für Werte aus H_0 ist die Gütefunktion kleiner gleich α.
3. Für wachsendes Stichprobenumfang n wird die Power eines Tests größer, d.h. die Gütefunktion wird steiler.

   $\vdots$

Gütefunktion bei verschiedenen Stichprobenumfängen

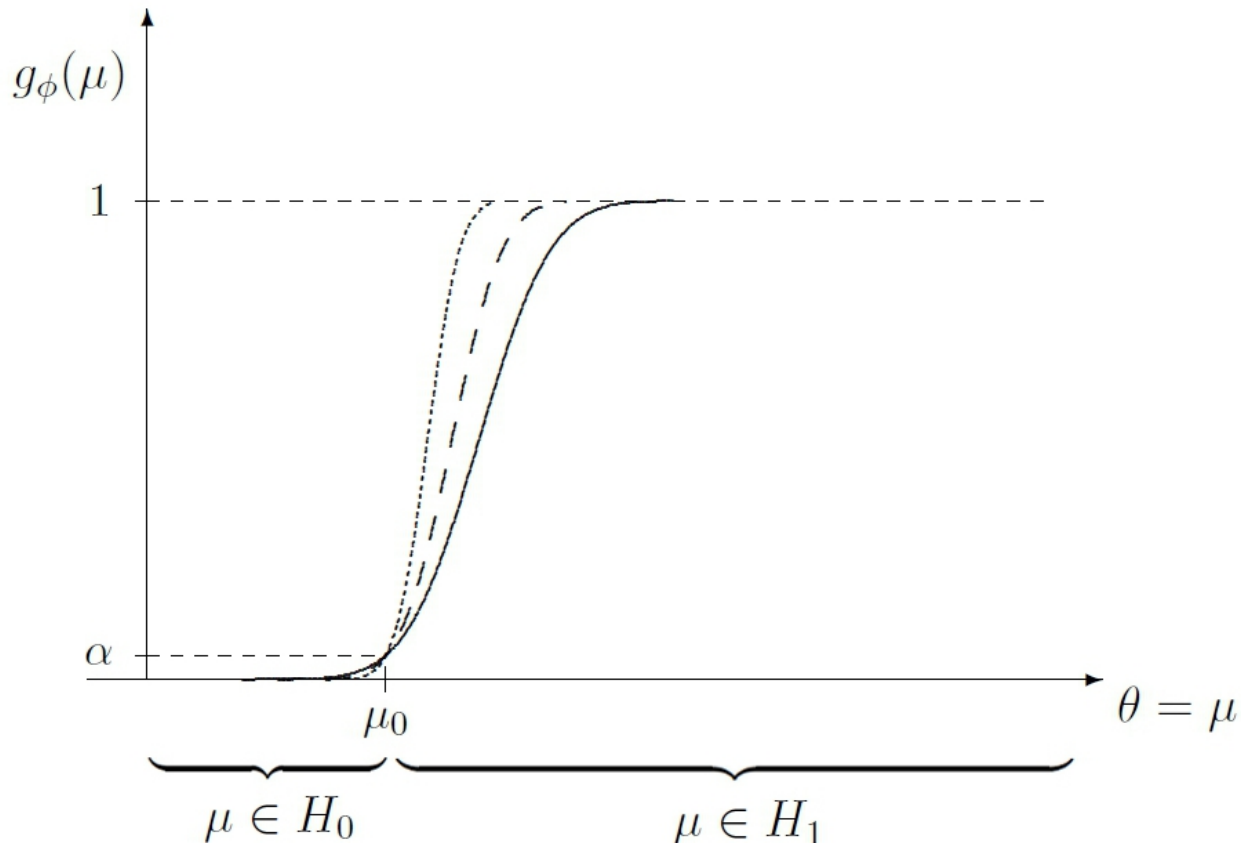


Abbildung: Gütefunktion $g_\phi(\mu) = 1 - \Phi(z_{1-\alpha} - \frac{\mu-\mu_0}{\sigma}\sqrt{n})$ für verschiedene Stichprobenumfänge $n = 10$ (—), $n = 20$ (- - -) und $n = 50$ ($\cdots$) und konstantem σ

Eigenschaften einer Gütefunktion

1. Für Werte aus H_1 heißt die Gütefunktion Power eines Tests.
2. Für Werte aus H_0 ist die Gütefunktion kleiner gleich α.
3. Für wachsenden Stichprobenumfang n wird die Power eines Tests größer, d.h. die Gütefunktion wird steiler.
4. Für wachsendes α wird die Power eines Tests größer.

$\vdots$

Gütefunktion bei verschiedenen Signifikanzniveaus

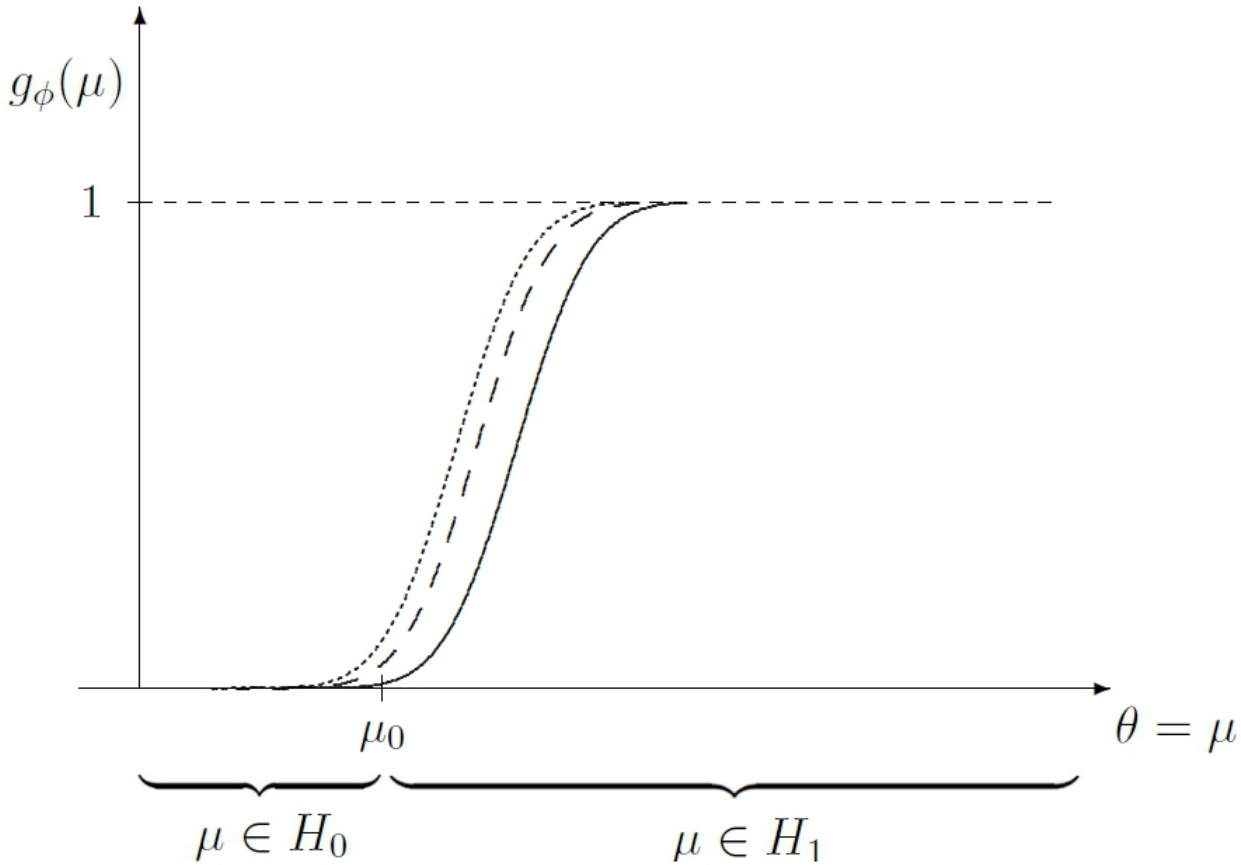


Abbildung: Gütefunktion $g_\phi(\mu) = 1 - \Phi(z_{1-\alpha} - \frac{\mu-\mu_0}{\sigma}\sqrt{n})$ für verschiedene Signifikanzniveaus $\alpha = 0,01$ (—), $\alpha = 0,05$ (- - -) und $\alpha = 0,1$ ($\cdots$) und konstantem σ

Eigenschaften einer Gütefunktion

1. Für Werte aus H_1 heißt die Gütefunktion Power eines Tests.
2. Für Werte aus H_0 ist die Gütefunktion kleiner gleich α.
3. Für wachsenden Stichprobenumfang n wird die Power eines Tests größer, d.h. die Gütefunktion wird steiler.
4. Für wachsendes α wird die Power eines Tests größer.
5. Für eine wachsende Abweichung zwischen Werten aus H_1 und H_0 wird die Power eines Tests größer.

Power bei wachsender Abweichung zwischen Werten aus H_1 und H_0

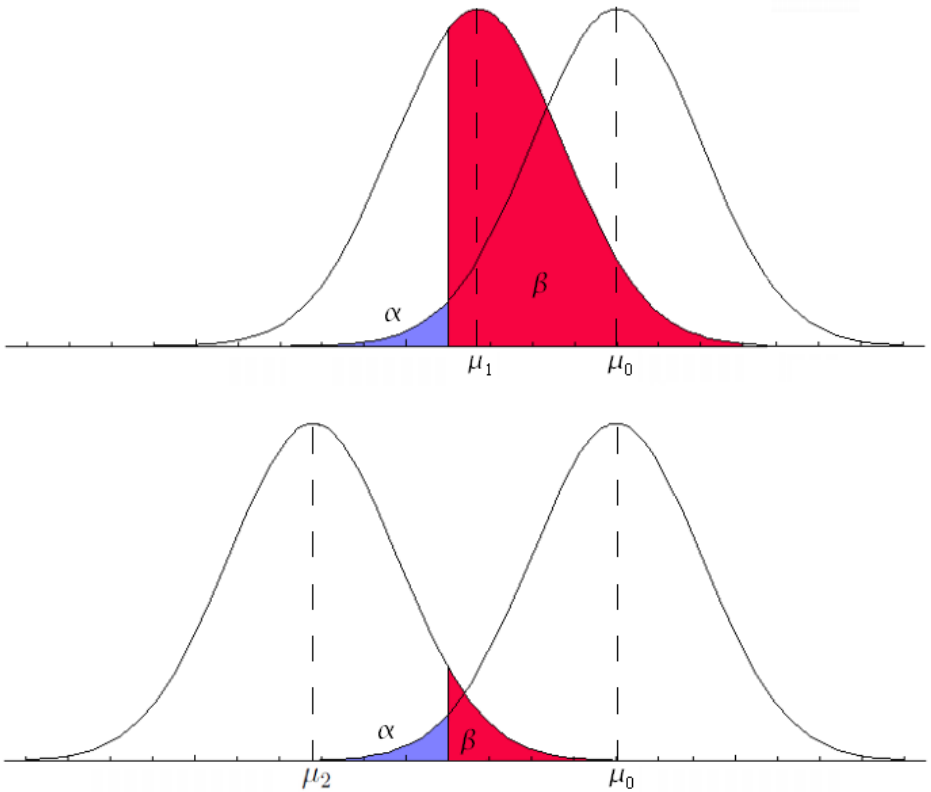


Abbildung: Vergleich α- und β-Fehler

Wann ist ein Test optimal?

1. Vergleichbarkeit nur für ein bestimmtes Testproblem

2. Zusätzlich gleiches und festes α-Niveau

   $\rightarrow$ fest vorgegeben Wahrscheinlichkeit des Fehlers 1. Art:
   d.h. für $\theta \in \Theta_0$ ist $P(H_0 \text{ verwerfen}|\theta) = \alpha$ fest

3. Test besitzt gleichzeitig minimalst mögliches β

   $\rightarrow$ Minimierung des Fehlers 2. Art:
   d.h. für $\theta \in \Theta_1$ wird $1 - P(H_0 \text{ verwerfen}|\theta) = \beta$ minimiert

$\Rightarrow$ Test hat **größtmögliche Power** $(1 - \beta)$ und ist somit **optimal** im Vergleich zu allen anderen Tests unter den beschriebenen Voraussetzungen.

Der Likelihood-Quotienten-Test (1/2)

Betrachtung eines **einfachen** Testproblems ($\theta_0 \neq \theta_1$):

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

Likelihood-Quotient: Dichte ist für H_0 gleich $f_0(x)$ und für H_1 gleich $f_1(x)$

$$\Lambda(x) = \frac{f_1(x)}{f_0(x)}$$

$\rightarrow$ **Relatives Maß** für bzw. gegen die Nullhypothese

$\Rightarrow$ Für besten Test wähle kritischen Wert k_α für Einhaltung des α-Niveaus:

$$H_0 \text{ ablehnen} \quad \Leftrightarrow \quad \Lambda(x) > k_\alpha$$

Der Likelihood-Quotienten-Test (2/2)

Theoretisches Problem:
α-Niveau wird bei diskreter Wahrscheinlichkeitsverteilung nicht immer ausgeschöpft

$\rightarrow$ Notwendigkeit einer **Randomisierung**

Testfunktion im diskreten Fall:

$$\phi^*(x) = \begin{cases} 1 & , \text{falls } f_1(x) > k_\alpha f_0(x) \Leftrightarrow \Lambda(x) > k_\alpha \\ \gamma_\alpha & , \text{falls } f_1(x) = k_\alpha f_0(x) \Leftrightarrow \Lambda(x) = k_\alpha \\ 0 & , \text{falls } f_1(x) < k_\alpha f_0(x) \Leftrightarrow \Lambda(x) < k_\alpha \end{cases}$$

Testfunktion im stetigen Fall:

$$\phi^*(x) = \begin{cases} 1 & , \text{falls } f_1(x) > k_\alpha f_0(x) \Leftrightarrow \Lambda(x) > k_\alpha \\ 0 & , \textit{sonst}. \end{cases}$$

Das Lemma

1. **Optimalität:** Für jedes k_α und γ_α hat der Test ϕ^* maximale Power unter allen Tests, deren Niveau höchstens gleich dem α-Niveau von ϕ^* ist.

2. **Existenz:** Zu vorgegebenem $\alpha \in [0, 1]$ existieren Konstanten k_α^* und γ_α^*, so dass der LQ-Test ϕ^* mit diesem k_α^* und $\gamma_\alpha = \gamma_\alpha^*$ für alle x exakt das Niveau α besitzt.

   $\rightarrow$ Es existieren k_α^* und γ_α^*, so dass α-Niveau für ϕ^* voll ausgeschöpft wird

3. **Eindeutigkeit:** Falls ein Test ϕ mit Niveau α maximale Power (= kleinsten Fehler 2. Art) unter allen anderen Tests mit Niveau α besitzt, dann ist ϕ ein LQ-Test.

   $\rightarrow$ Optimaler Test ist Likelihood-Quotienten-Test

Allgemeine Interpretation eines Tests

- **Zentraler Streitpunkt:** Unterscheidung zwischen induktivem Verhalten (Neyman-Pearson) und induktivem Schließen (Fisher)
- **Fisher:** Erlangen konkreter Erkenntnisse durch signifikante Ergebnisse

  $\rightarrow$ konkrete Aussagen, wie „Alle Schwäne sind weiß!“

  $\rightarrow$ kognitivistische Herangehensweise
- **Neyman-Pearson:** Verhaltensempfehlung durch signifikante Ergebnisse

  $\rightarrow$ Verhalten besteht darin Nullhypothese zu verwerfen oder nicht

  $\rightarrow$ dezisionistische Herangehensweise

Häufige Vorgehensweise in der Praxis

1. Benennung einer Null- und einer Alternativhypothese (nach Neyman-Pearson)
2. Berechnung eines exakten p-Wertes (nach Fisher);
   meist automatisiert durch Programme
3. Nachträgliche Kennzeichnung des p-Wertes mit Sternen
   z.B. * bei $p < 0,05$, ** bei $p < 0,01$ und *** bei $p < 0,001$

$\Rightarrow$ Problematik: keine Kontrolle des Fehlers 1. Art über α

$\Rightarrow$ Gefahr: Anpassung des α-Niveaus nach der Berechnung des p-Werts

Allgemeine Kritik

- Verwendung der Nullhypothese als „Strohmann“
- Ritualisierte Anwendung von Signifikanztests

  $\rightarrow$ signifikante Ergebnisse sind notwendig, um beachtet zu werden
- Stupide Verwendung von Signifikanztests auf Kosten von geeigneteren Methoden[2]

[2] Siehe hierfür z.B.: Loftus, G. R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age, *Behavior Research Methods* 25(2): 250-256.

Die wichtigsten Aspekte im Überblick (1/2)

- Konflikt zwischen Neyman-Pearson und Fisher hat das heutige Verständnis der Statistik maßgeblich beeinflusst
- Unterscheidung zwischen Signifikanztest nach Fisher und Hypothesentest nach Neyman-Pearson
- Benennung einer **konkreten Alternativhypothese** durch Neyman-Pearson ermöglicht:
  1. Bestimmung des β-**Fehlers**
  2. Betrachtung einer **Gütefunktion**
  3. Vergleich von Tests über die **Power** $(1 - \beta)$
  4. Konstruktion eines **optimalen Tests** über das Neyman-Pearson Lemma

$\rightarrow$ Fisher lehnt Benennung einer konkreten Alternativhypothese ab

Die wichtigsten Aspekte im Überblick (2/2)

- Ein **optimaler Test** hat bei gleichem Testproblem **maximale Power** im Vergleich zu jedem anderen α-Niveau Test
- Induktives **Verhalten** (Neyman-Pearson) versus induktives **Schließen** (Fisher)

  $\rightarrow$ **Fisher:** Erlangen konkreter Erkenntnisse durch signifikante Ergebnisse

  $\rightarrow$ **Neyman-Pearson:** Verhaltensempfehlung durch signifikante Ergebnisse
- Heutzutage: inkonsistente **Vermischung** der Ideen von Fisher und Neyman-Pearson bei der Verwendung von Signifikanztests

Empfohlene Literatur

→ Zusätzlich zu den zitierten und genannten Veröffentlichung von Neyman und E. S. Pearson im Vorbereitungsmaterial:

1. **Lehmann**, E. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two?, *Journal of the American Statistical Association* 88(424): 1242-1249.

2. **Saint-Mont**, U. (2011). Statistik im Forschungsprozess: Eine Philosophie der Statistik als Baustein einer integrativen Wissenschaftstheorie, Physica-Verlag HD.

Abschlussdiskussion

Wie ist eure Vorgehensweise bei der Durchführung eines Signifikanztests?

Achtet ihr auf eine „Kosten-Nutzen Abwägung“ nach Neyman-Pearson oder ist für euch grundsätzlich der p-Wert von Bedeutung?

Beweis der Optimalität (1/2)

Sei ϕ^* bester Test und ϕ beliebiger anderer Test mit gleichem α-Niveau.

$\rightarrow \mathbb{E}_{\theta_0}[\phi^*(X)] = \alpha \mathrel{\hat{=}}$ Wahrscheinlichkeit H_0 zu verwerfen, falls $\theta \in \Theta_0$

$\rightarrow \mathbb{E}_{\theta_1}[\phi^*(X)] = 1 - \beta \mathrel{\hat{=}}$ Wahrscheinlichkeit H_0 zu verwerfen, falls $\theta \in \Theta_1$

Für Optimalität von ϕ^* gilt es zu zeigen:

$$\mathbb{E}_{\theta_1}[\phi(X)] \leq \mathbb{E}_{\theta_1}[\phi^*(X)]$$

Betrachte hierfür Hilfsfunktion $U(x)$:

$$\begin{aligned} U(x) &= [\phi^*(x)f_1(x) - \phi^*(x)k_\alpha f_0(x)] - [\phi(x)f_1(x) - \phi(x)k_\alpha f_0(x)] \\ &= (\phi^*(x) - \phi(x))(f_1(x) - k_\alpha f_0(x)) \end{aligned}$$

Beweis der Optimalität (2/2)

Da $U(x) \geq 0$ für alle x, gilt somit:

$$
\begin{aligned}
0 &\leq \int U(x)\,dx \\
&= \int (\phi^*(x) - \phi(x))(f_1(x) - k_\alpha f_0(x))\,dx \\
&= \int \phi^*(x) f_1(x)\,dx - \int \phi(x) f_1(x)\,dx + k_\alpha \left(\int \phi(x) f_0(x)\,dx - \int \phi^*(x) f_0(x)\,dx \right) \\
&= \mathbb{E}_{\theta_1}[\phi^*(X)] - \mathbb{E}_{\theta_1}[\phi(X)] + \underbrace{k_\alpha(\mathbb{E}_{\theta_0}[\phi(X)] - \mathbb{E}_{\theta_0}[\phi^*(X)])}_{\leq 0\text{, da } \mathbb{E}_{\theta_0}[\phi(X)] \leq \mathbb{E}_{\theta_0}[\phi^*(X)]}
\end{aligned}
$$

$\Rightarrow \mathbb{E}_{\theta_1}[\phi^*(X)] \geq \mathbb{E}_{\theta_1}[\phi(X)]$,
d.h. die Power von ϕ^* ist größer oder zumindest genauso groß wie die Power von ϕ.

Hinweis: $\mathbb{E}_{\theta_0}[\phi^*(X)] = \int_{-\infty}^{+\infty} \phi^*(x) f_0(x)\,dx$

frequentistisch vs. nicht-frequentistisch

1. **frequentistisch:**
   Güte einer Schlussfolgerung basiert darauf, wie häufig diese im Durchschnitt zu einer wahren Aussage führt
   (bei Anwendung auf viele verschiedene Beobachtungen).

   $\rightarrow$ klassische Inferenz

2. **nicht-frequentistisch:**
   Güte einer Schlussfolgerung basiert darauf, wie plausibel diese ist
   (in Hinblick auf die vorliegende Beobachtung).

   $\rightarrow$ Bayes-Inferenz

kognitivistisch vs. dezisionistisch

1. **kognitivistisch:**
   Schlussfolgerung als konkrete Erkenntnis

   $\rightarrow$ klassische Inferenz und Bayes-Inferenz

2. **dezisionistisch:**
   Schlussfolgerung dient dem Treffen von Entscheidungen

   $\rightarrow$ entscheidungstheoretische Inferenz

objektivistisch vs. subjektivistisch

1. **objektivistisch:**
   Nur Verwendung von objektiven Informationen aus den Beobachtungen

   $\rightarrow$ klassische Inferenz

2. **subjektivistisch:**
   Zusätzlich subjektive Annahmen wie z.B. das a priori Wissen

   $\rightarrow$ Bayes-Inferenz

Anhang

Herleitung Konfidenzintervall zweiseitiger Gauß-Test

Seien $X_1, ..., X_n$ ZV mit $X_i \sim N(\mu, \sigma^2)$ bei $\mathbb{E}(X_i) = \mu$, $\mathbb{V}(X_i) = \sigma^2$ und $n \geq 30$, wobei σ^2 bekannt ist. Für $\mu = \mu_0$ ist: $\bar{X} \sim N(\mu_0, \frac{\sigma}{\sqrt{n}}) \Rightarrow Z = \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} \sim N(0, 1)$

$$\text{Betrachte:} \qquad H_0 : \mu = \mu_0 \quad \text{gegen} \quad H_1 : \mu \neq \mu_0$$

Die Nullhypothese H_0 ist beizubehalten falls:

$$|z| = \left| \frac{\bar{x} - \mu_0}{\sigma}\sqrt{n} \right| \leq z_{1-\alpha/2} \quad \Leftrightarrow \quad |\bar{x} - \mu_0| \leq z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \quad \bar{x} - \mu_0 \geq -z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{und} \quad \bar{x} - \mu_0 \leq z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow \quad \mu_0 \leq \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{und} \quad \mu_0 \geq \bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Daraus lässt sich folgendes $(1 - \alpha)$-Konfidenzintervall für μ bestimmen:

$$\left[\bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad , \quad \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

Anhang

Herleitung Gütefunktion einseitiger Gaußtest

Seien $X_1, ..., X_n$ ZV mit $X_i \sim N(\mu, \sigma^2)$ bei $\mathbb{E}(X_i) = \mu$, $\mathbb{V}(X_i) = \sigma^2$ und $n \geq 30$, wobei σ^2 bekannt ist. Für $\mu = \mu_0$ ist: $\bar{X} \sim N(\mu_0, \frac{\sigma}{\sqrt{n}}) \Rightarrow Z = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim N(0, 1)$

Betrachte: $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$

$$
\begin{aligned}
g_\phi(\mu) &= P(H_0 \text{ verwerfen} \,|\, \mu) = P\left(\frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} > z_{1-\alpha} \,|\, \mu\right) \\
&= P\left(\frac{\bar{X} - \mu_0 + \mu - \mu}{\sigma}\sqrt{n} > z_{1-\alpha} \,|\, \mu\right) \\
&= P\left(\frac{\bar{X} - \mu}{\sigma}\sqrt{n} + \frac{\mu - \mu_0}{\sigma}\sqrt{n} > z_{1-\alpha} \,|\, \mu\right) \\
&= P\left(\frac{\bar{X} - \mu}{\sigma}\sqrt{n} > z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma}\sqrt{n} \,|\, \mu\right) = 1 - \Phi(z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma}\sqrt{n})
\end{aligned}
$$

Da mit μ als wahren Parameter $\frac{\bar{X} - \mu}{\sigma}\sqrt{n} \sim N(0, 1)$ gilt. Graphische Darstellung der Gütefunktion ist nur in Abhängigkeit von α und n und als Funktion von μ möglich.