Statistische Methoden der Datenanalyse

Wintersemester 2012/2013

Albert-Ludwigs-Universität Freiburg

Prof. Markus Schumacher, Dr. Stan Lai

Physikalisches Institut Westbau 2 OG

E-Mail: Markus.Schumacher@physik.uni-freiburg.de

stan.lai@cern.ch

http://terascale.physik.uni-freiburg.de/lehre/ws_1213/statmethoden_ws1213

Kapitel 8:
Statistische Hypothesentests

Statistische Hypothesentests: Einführung

Ziel: Vergleich der Beobachtung bzw. Auswertung der Messdaten
einer Stichprobe mit Hypothesen
→ Entscheidung welche Hypothese bevorzugt wird,
welche Hypothesen verworfen oder behalten werden

a) Vergleich mit Theorie
- Gauss-WDF + Annahme über Mittelwert, Varianz
- Gute der Anpassung von Parametern
- Hinweis auf neues Phänomen

b) Vergleich von zwei Stichproben
- Mittelwert, Varianz
- Form der Verteilung (selbe Grundgesamtheit)

c) Unterscheidung von Hypothesen
- Exponential- oder Gauss-WDF
- linearer oder quadratischer Zusammenhang zwischen Messpaaren
- Diskriminierung von Ereignisklassen (Signal oder Untergrund), Spam-Mail oder “gute Mail”, Teilchensorten: e, μ, π, γ….)

Statistische Hypothesentests: Einführung

Methode:

- Konstruktion einer Größe zur Quantifzierung der Übereinstimmung/Diskrepanz mit Hypothesen → Teststatistik t

- Quantifizierung der Übereinstimmung mittels Wahrscheinlichkeitsdichtefunktion f(t) für Teststatistik

- Entscheidung über Verwerfung der Hypothese oder Auswahl unter Alternativhypothesen

Grundbegriffe für statistische Tests: Hypothesen

Hypothese(n):
klare Aussage(n), die man falsifizieren bzw. unterscheiden kann

Fall (a): eine allgemein anerkannte Hypothese, die falsifiziert werden soll
$\rightarrow$ Nullhypothese H_0

Beispiele:
- Lebensdauer des Teilchens ist τ
- kein Anzeichen für neue Physik
- Theorie und Stichprobe (2 Stichproben) stimmen überein in Mittelwert oder Form der Verteilung

Bemerkung: oft Aussage (Nullhypothese) = Negation der Aussage, die einen interessiert

Beispiel: Suche nach neuem Teilchen $\rightarrow$ H_0: nur Untergrund

Lebensdauer $\tau \neq \tau_0$ $\rightarrow$ H_0: $\tau = \tau_0$

Grundbegriffe für statistische Tests: Hypothesen

Fall (b): mehrere Hypothesen zwischen den Unterschieden werden soll

"Standardannahme", Nullhypothese: H_0

Alternativhypothesen: $H_1, H_2, H_3, \ldots$

Beispiele: Lebensdauer H_0: $\tau = 2s$ H_1: $\tau = 1s$ H_2: $\tau > 2s$

Higgssuche: H_0: nur Untergrund H_1:Signal+Untergrund

Signatur im Detektor von: H_0 = Pion, H_1= Myon, H_2=Elektron

Arten von Hypothesen:

x sei ZV und $f(x;\lambda)$ Wahrscheinlichkeitsdichtefunktion

- einfach: wenn f(x) durch Hypothese vollständig fixiert
  entweder kein Parameter oder Parameter λ festgelegt
- zusammengesetzt: wenn mindestens einer der Parameter nicht bekannt ist
  $f(x;\lambda)$ mit λ unbekannt oder λ aus Intervall [a,b]

WDF für Hypothesen werden mit $f(x|H_0)$ und $f(x|H_1)$ bezeichnet

Grundbegriffe für statistische Tests: Teststatistik

Teststatistik $t(x_1, \ldots x_n)$: Funktion der Stichprobenwerte $(x_1,..,x_n)$

zur Quantifzierung der Übereinstimmung mit Hypothesen H_i mit Ziel
- Verwerfung von H_0 (H_1)
- Unterscheidung von H_0, H_1, H_2

$t(x_1,...,x_n)$ kann Vektor sein z.B. $t_i = x_i$ (nicht sehr nützlich)

Ziel: Reduzierung der Dimension von t auf skalare Größe t
unter optimaler Ausnutzung der Information in Stichprobe (x_i)
bzgl. der Hypothesen H_i

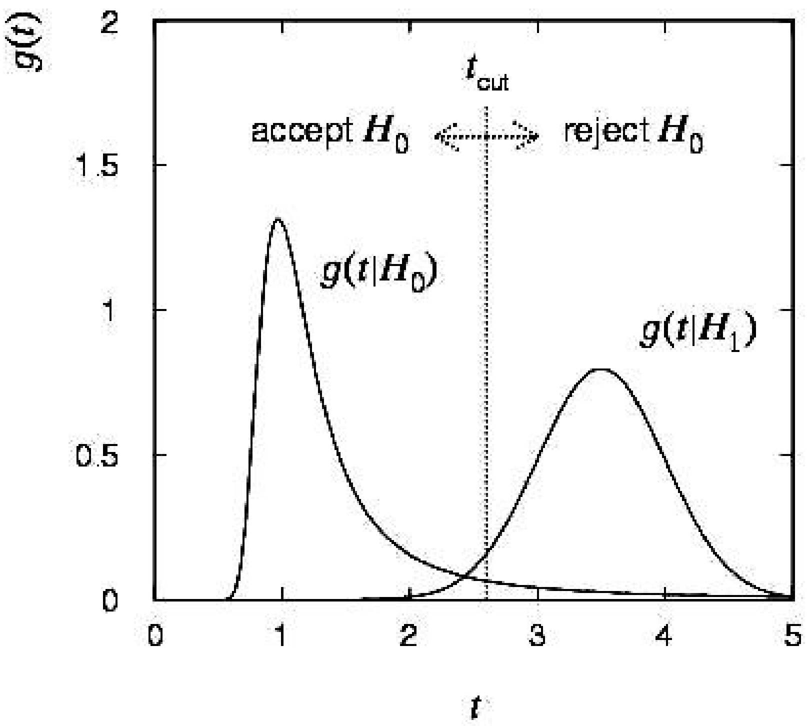
Aufgaben: a) Definition/Auswahl von Teststatistik t
b) Bestimmung der WDFs für t unter den Hypothesen $f(t \mid H_i)$
c) Festlegung eines Kriteriums für Verwerfung/Unterscheidung der Hypothesen

Grundbegriffe für stat. Tests:Entscheidungsgrenze

Entscheidungsgrenze t_{cut} oder t_{krit}

$$t(x_1, \ldots, x_n) = t_{\mathrm{cut}}$$

Annahme: wir können die WDFs für die Testatistik unter beiden Hypothesen ausrechnen $g(t|H_0),\ g(t|H_1),\ \ldots$



Festlegung der Entscheidungsgrenze (kritischer Wert) t_{krit} oder t_{cut}

kritische Region/Verwerfungregion: $t>t_{krit}$

Verwerfung von H_0, Nicht-Verwerfung (Akzeptanz) von H_1

Komplement zu kritischer Region/Akzeptanzregion: $t<t_{krit}$

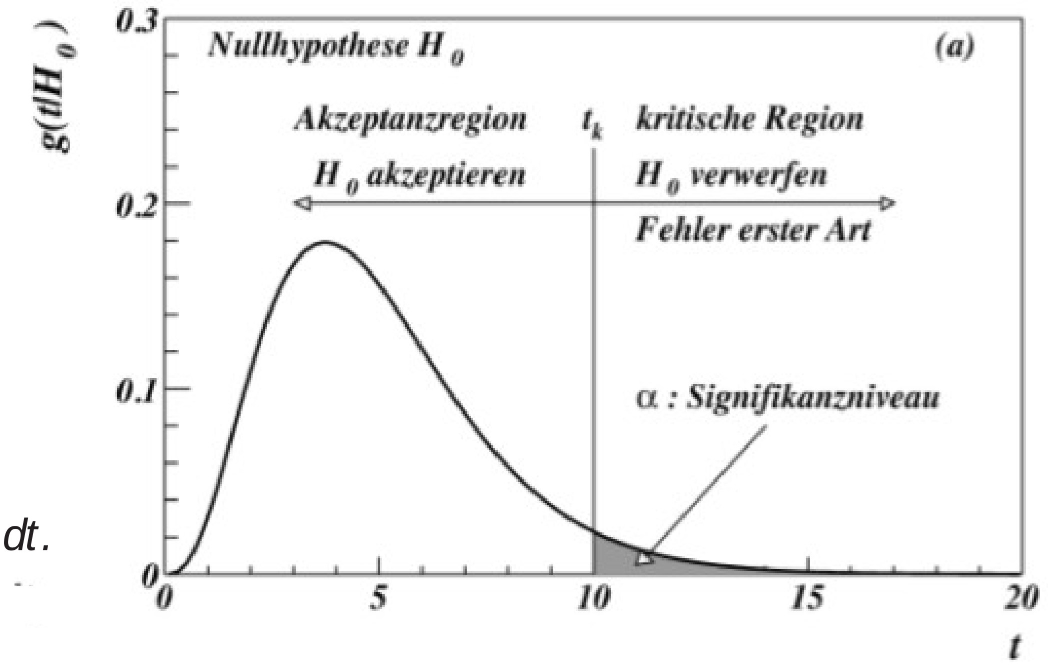
keine Verwerfung von H_0, Verwerfung von H_1

Grundbegriffe: Signifikanzniveau

kritische Region Ω_k

Akzeptanzregion $\Omega - \Omega_k$

$$\alpha = \int_{\Omega_k} g(\vec{t}|H_0)d\vec{t}. \qquad \alpha = \int_{t_k}^{\infty} g(t|H_0)dt.$$



α: Signifikanzniveau, Fehler erster Art
ist Wahrscheinlichkeit H_0 zu verwerfen, obwohl die Hypothese war ist

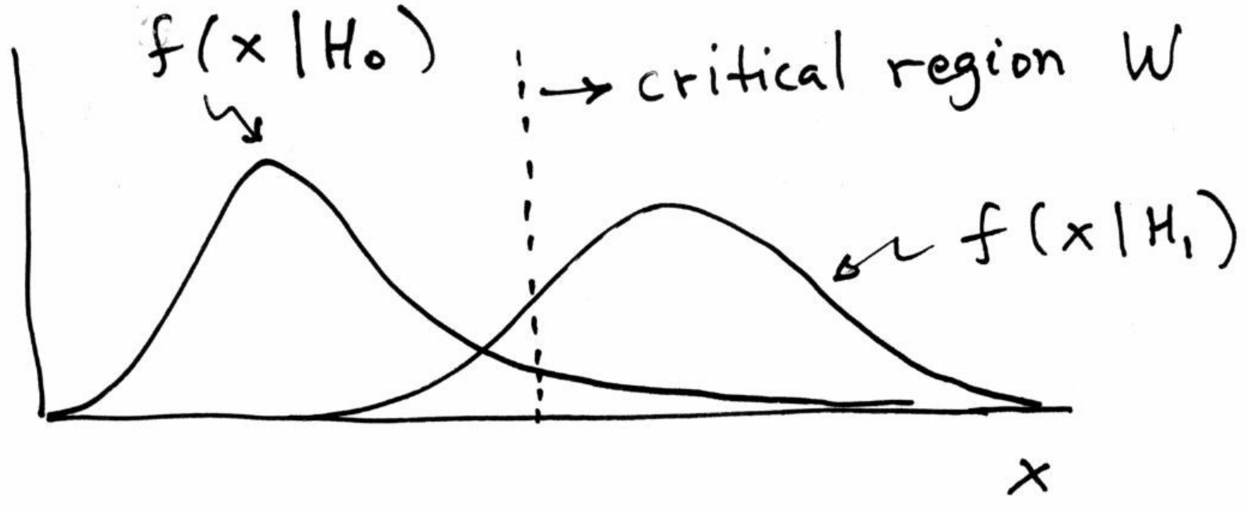
Bemerkung: α (=5%,10%,1%, 2.85x10^{-7}) vor Experiment festlegen,
wenn H_0 verworfen/getestet werden soll
α ist keine Zufallsvariable sondern ein festgelegter Wert

Definition of a test (2)

Aber im allgmeinen gibt es unendliche viele Möglichkeiten kritische Regionen zu wählen, die alle das gleiches Signifinkanzniveau α besitzen.

Also muss die Wahl der kritischen Region für die Nullhypothese H_0 die Alternativhypothese H_1 berücksichtigen.

Ungefähr ausgedrückt heisst das Kriterium: wähle die kritische Region so, dass es eine kleine Wkt gibt dort eine Messung zu finden, wenn H_0 wahr ist, aber eine große, wenn H_1 ist wahr ist.



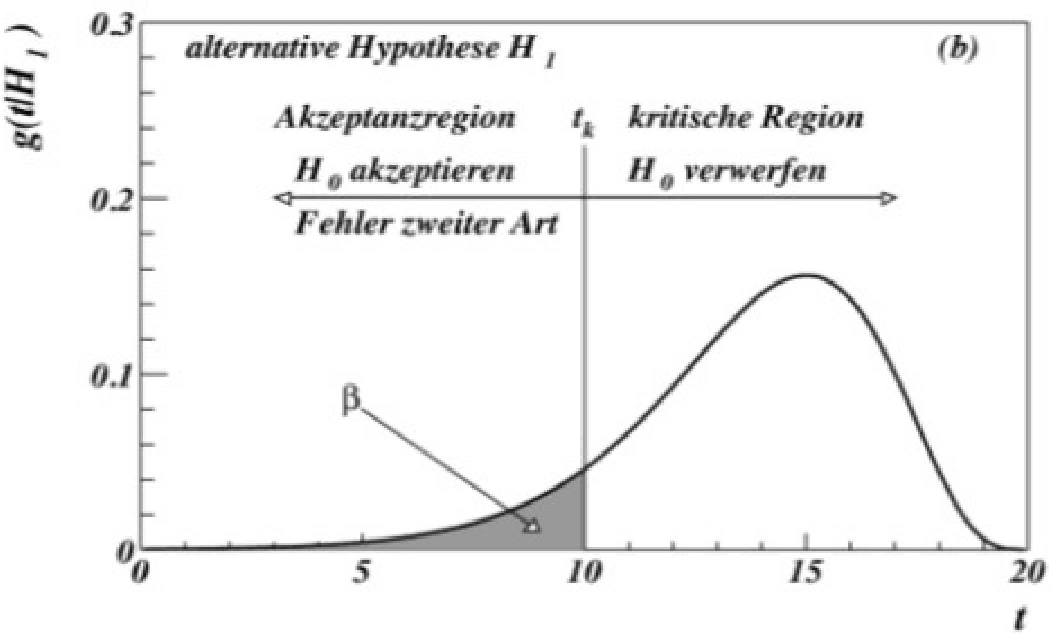
Grundbegriffe: Mächtigkeit

kritische Region Ω_k

Akzeptanzregion $\Omega - \Omega_k$

$$\beta = \int_{\Omega - \Omega_k} g(\vec{t}|H_1)d\vec{t}.$$

$$\beta = \int_{-\infty}^{t_k} g(t|H_1)dt.$$



β: Fehler zweiter Art   1-β: Mächtigkeit des Tests

β ist Wkt H_1 zu verwerfen obwohl die Hypothese wahr ist

1-β ist Wkt. H_1 zu akzeptieren, wenn Hypothese wahr ist

Ziel: α,β minimieren (ideal = 0) und 1-β maximieren (ideal=1)
nicht gleichzeitig möglich → Kompromiss
Wahl der besten kritischen Region/Teststatistik für gegebenes α

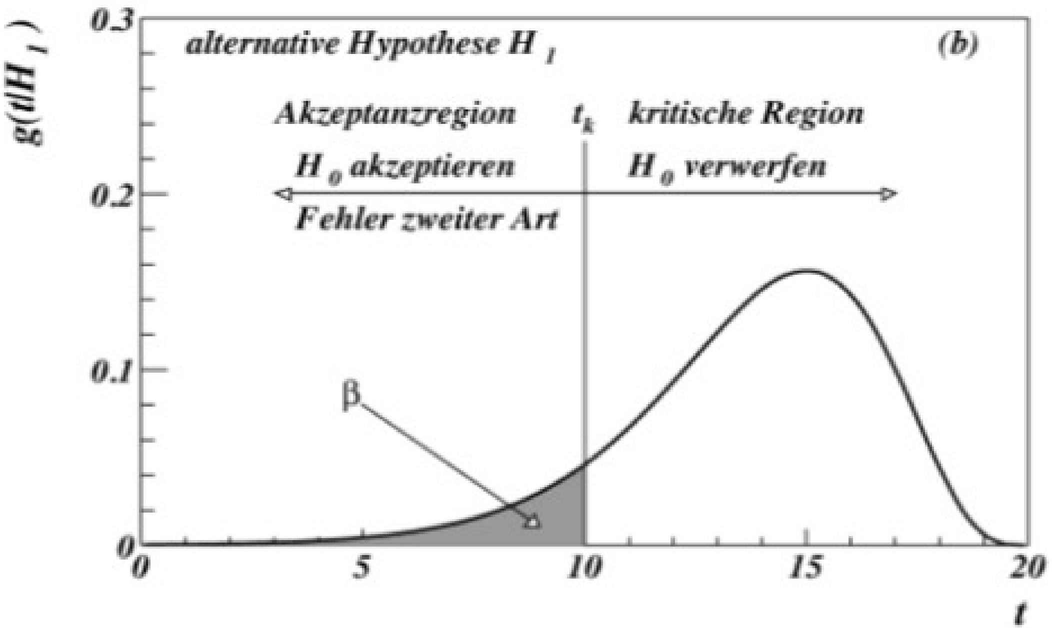
Grundbegriffe: Mächtigkeit

kritische Region Ω_k

Akzeptanzregion $\Omega - \Omega_k$

$$\beta = \int_{\Omega - \Omega_k} g(\vec{t}|H_1)\,d\vec{t}.$$

$$\beta = \int_{-\infty}^{t_k} g(t|H_1)\,dt.$$



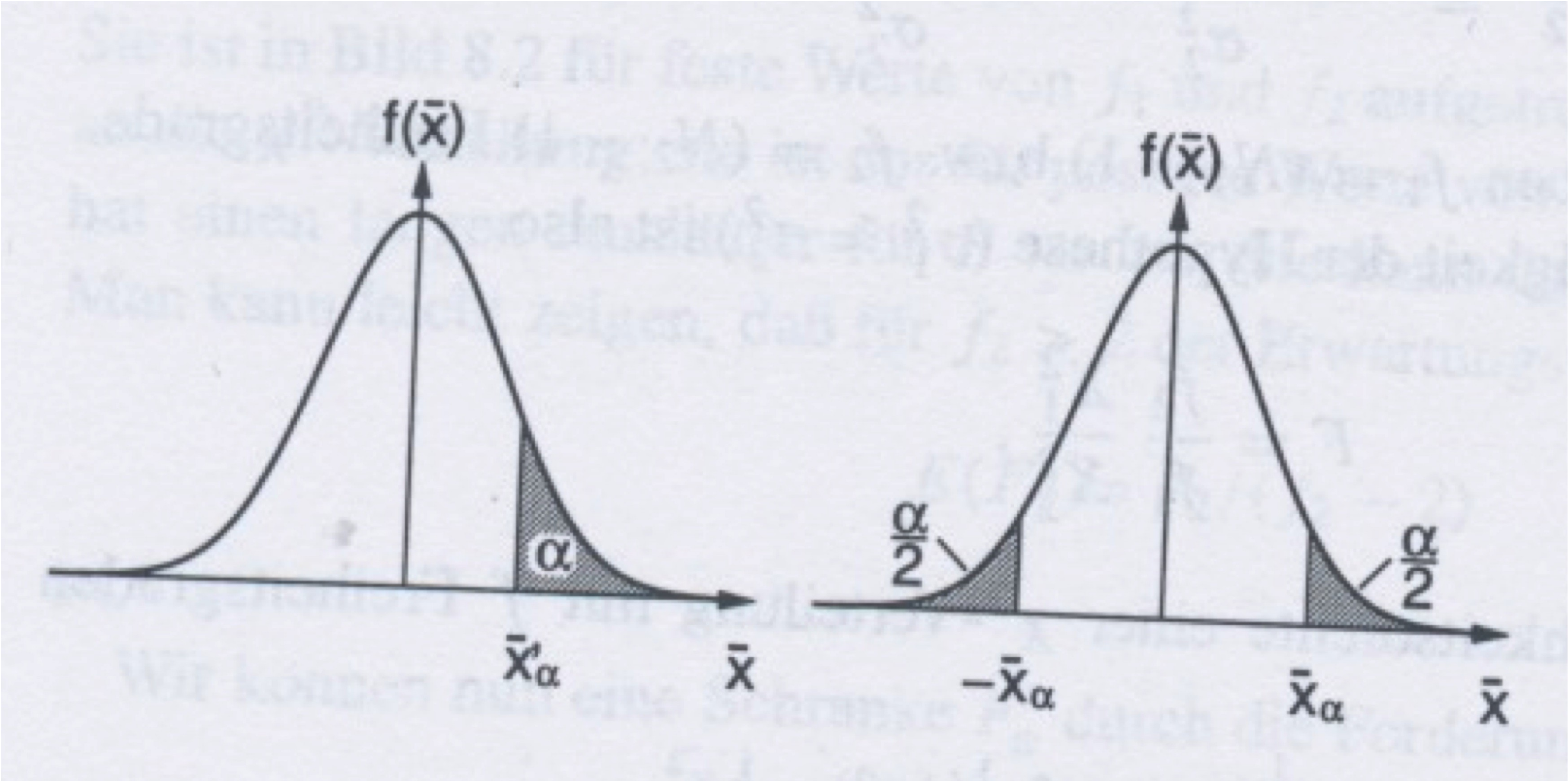
β: Fehler zweiter Art 1-β: Mächtigkeit des Tests

β ist Wkt H_1 zu verwerfen obwohl die Hypothese war ist

1-β ist Wkt. H_1 zu akzeptieren, wenn Hypothese war ist

Ziel: α,β minimieren (ideal = 0) und 1-β maximieren (ideal=1)
nicht gleichzeitig möglich → Kompromiss
Wahl der besten kritischen Region/Teststatistik für gegebenes α

Ein- und zweiseitige Tests



Je nach Problem sind Abweichungen in beide Richtungen interessant
→ dann zweiseitiger Test z.B. Toleranzen in der industriellen Produktion

verteile das Signifikanzniveau i.a. zur Hälfte auf Ausläufer nach oben u. unten