

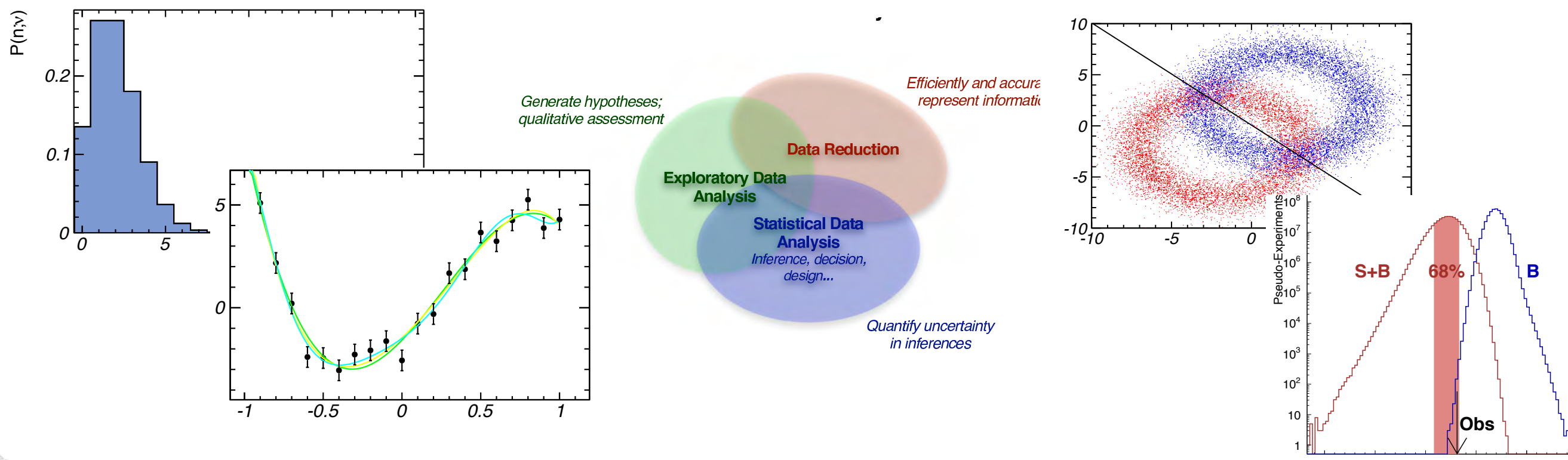
Rechnernutzung in der Physik

Teil 3 – Statistische Methoden der Datenanalyse

Karlsruher Institut für Technologie
Wintersemester 2012/2013

Ulrich Husemann

Institut für Experimentelle Kernphysik, Karlsruher Institut für Technologie



QISPOS-Anmeldung

- Rechnernutzung jetzt bei QISPOS freigeschaltet
→ bitte anmelden

- Name: Rechnernutzung
- Prüfungsnummer: 172
- Anmeldebeginn: 15.10.12
- Anmeldeende: 07.02.13
- Rücktrittsende: 07.02.13
- Prüfungsdatum: 08.02.13

Kurze Wiederholung

- Systematische Methoden zur Parameterschätzung
 - Maximum-Likelihood-Methode: Maximierung der Likelihood-Funktion
 - Kleinste Quadrate (LS): Minimierung der χ^2 -Funktion
- Kleinste Quadrate: $\chi^2 = (\vec{y} - \vec{\lambda})^T V^{-1} (\vec{y} - \vec{\lambda})$
 - Analytisch lösbar, wenn λ linear in Parametern
 - Goodness-of-fit: $\chi^2/n \approx 1$ (n Zahl der Freiheitsgrade)
- Graphische Methode zur Bestimmung der Varianz von Schätzern
 - ML: Funktionswert von $-\ln L$ am Minimum + 1/2
 - LS: Funktionswert von χ^2 am Minimum + 1
- Parameteranpassung mit ML und LS: Voraussetzungen, Unterschiede, Behandlung von Unsicherheiten kennen/beachten!

Werbung in eigener Sache

- Vorlesung im Sommersemester 2013:
„Moderne Methoden der Datenanalyse”
 - Vorlesung Dienstags 15:45-17:15 Uhr
 - Computerpraktikum Donnerstags 15:30–18:00 Uhr
 - Dozenten: Ralf Ulrich, UH
- Voraussichtliche Inhalte
 - Ausführlichere Einleitung in Grundlagen der Statistik
 - Konfidenzintervalle, Limits etc.
 - Statistische Klassifikation mit multivariaten Methoden
 - Faltungs- und Entfaltungsmethoden
 - Numerische Optimierung
 - Econo-Physics

Kapitel 3.5

Parameterschätzung

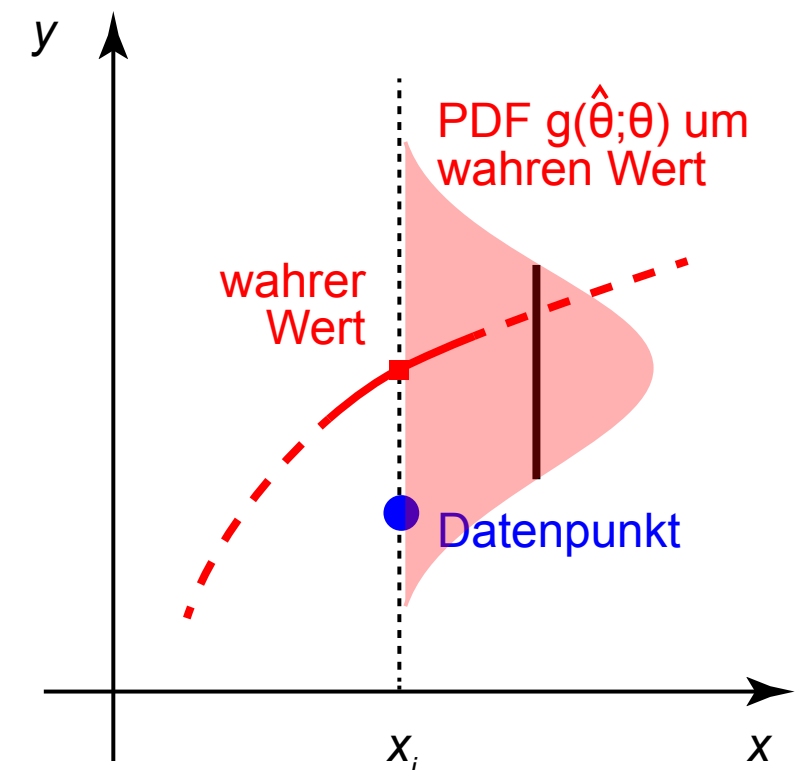
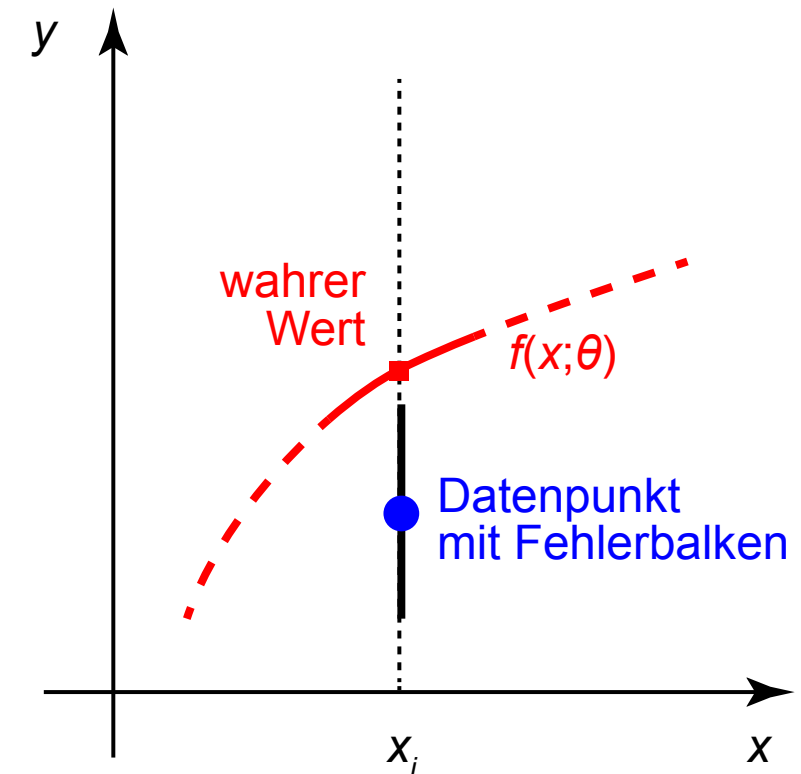
Was ist ein „Fehlerbalken“?

■ Sichtweisen auf „Fehlerbalken“

1. Fehlerbalken = Unsicherheit der Messung
2. Datenpunkt = Teil einer Stichprobe = Zufallsvariable aus PDF $g(\hat{\theta}; \theta)$ um wahren Wert (häufig: $\pm 1\sigma$ einer Gaußverteilung)

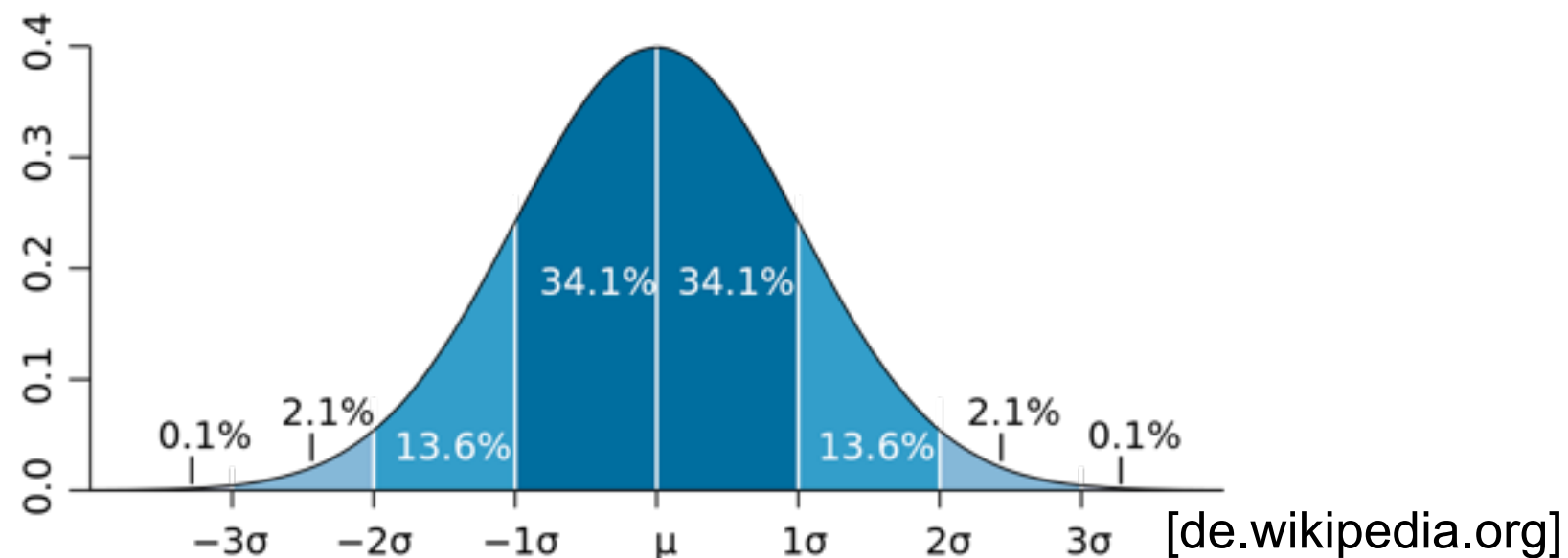
■ Größe des Fehlerbalkens aus Unsicherheit der Parameterschätzung

- Bisher: Standardabweichung (bzw. Kovarianz) von ML- oder LS-Schätzern, ggf. nach Fehlerfortpflanzung
- Problem: Standardabweichung ungünstig für asymmetrische Verteilungen (z. B. Exponentialverteilung: 86% Wahrscheinlichkeit in $\pm 1\sigma$)



Vertrauensintervalle

- Besserer Ansatz: Vertrauensintervalle
 - Verallgemeinerung der Varianz von Gaußverteilungen
 - Idee: Angabe der Unsicherheit als Vertrauensintervalle" mit fester Wahrscheinlichkeit („Vertrauensniveau“, engl.: confidence level, CL)
- Typische Konventionen für Vertrauensniveaus:
 - 68.3% Wahrscheinlichkeit (entspricht $\pm 1\sigma$ bei Gaußverteilung)
 - 90% Wahrscheinlichkeit (entspricht $\pm 1.645\sigma$ bei Gaußverteilung)
 - 95% Wahrscheinlichkeit (entspricht $\pm 1.960\sigma$ bei Gaußverteilung)

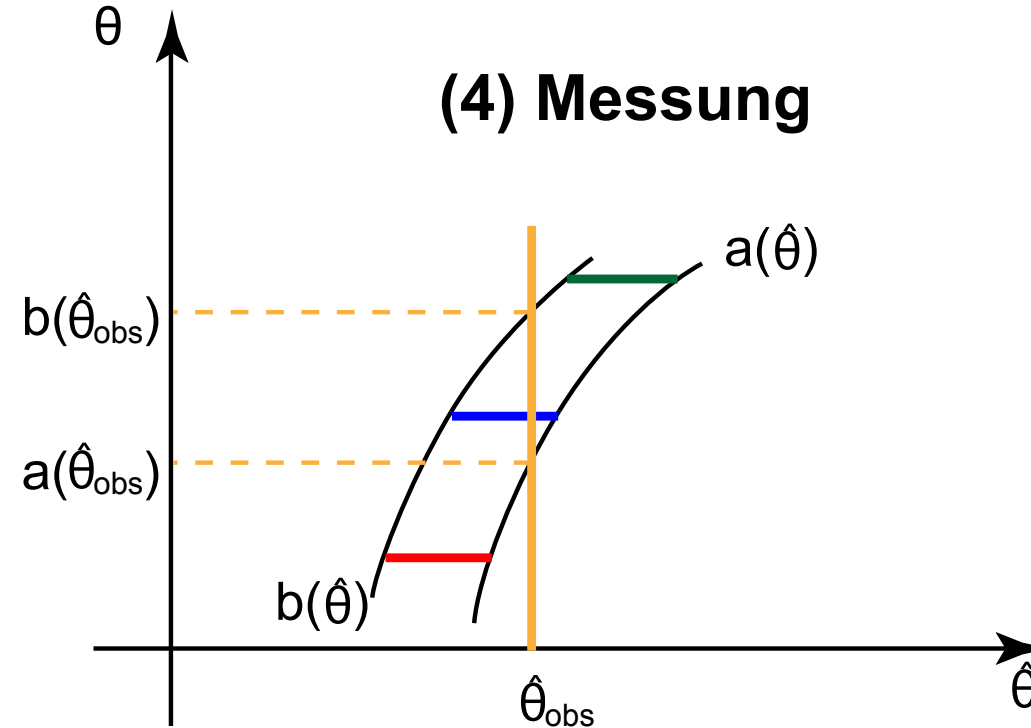
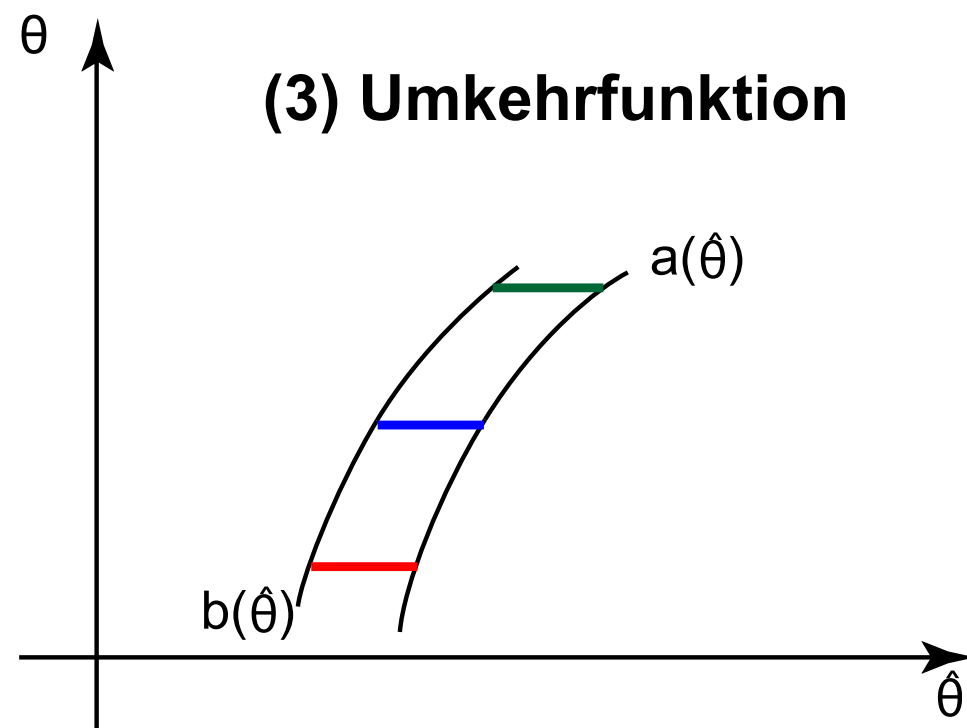
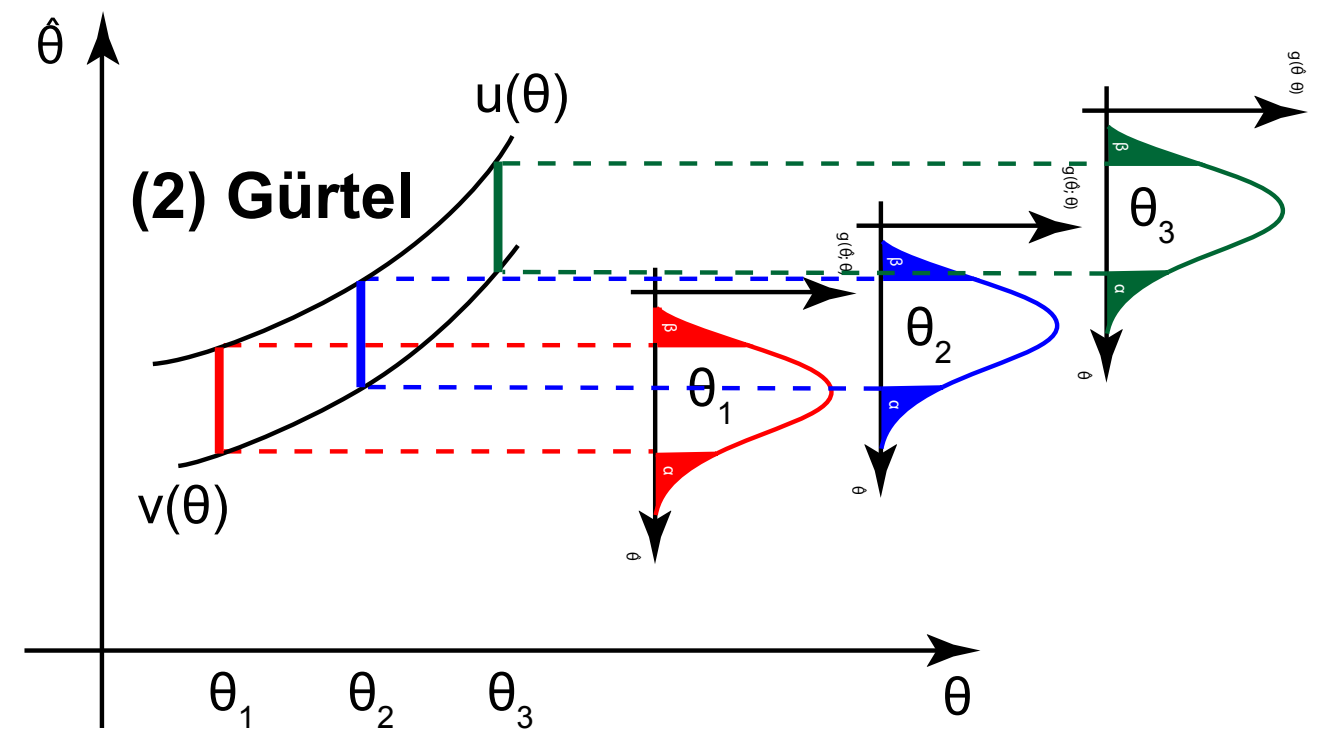
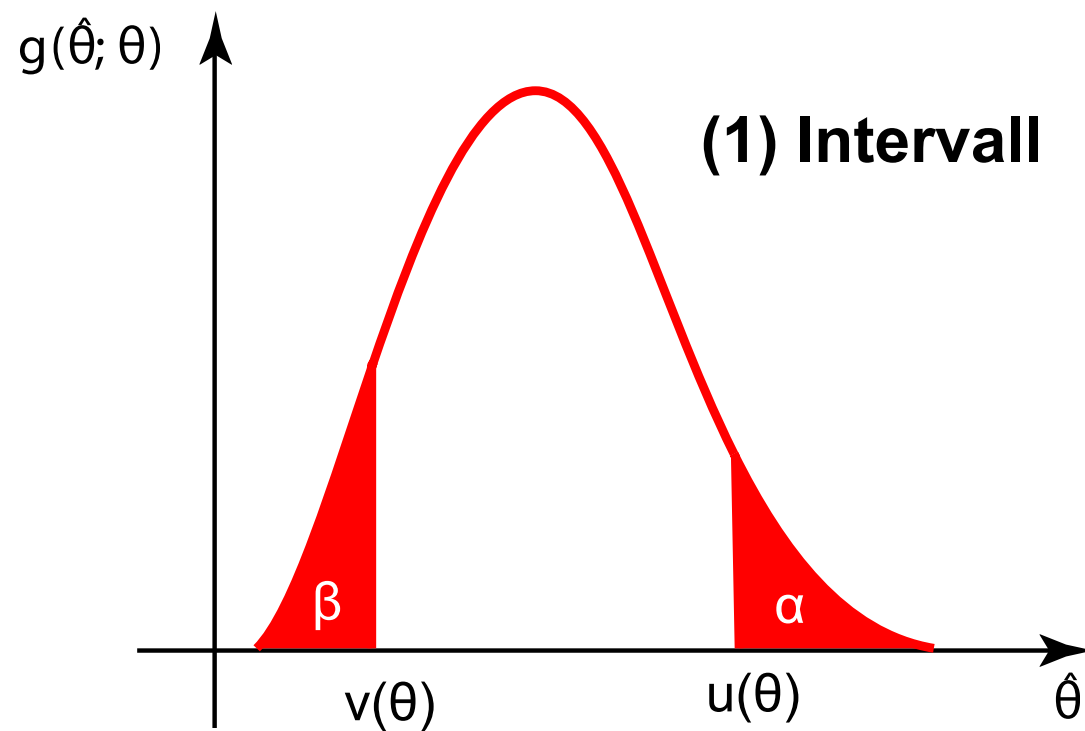


Neyman-Konstruktion

- Systematische Vorschrift aus frequentistischer Statistik: Konstruktion von Vertrauensintervallen mit gegebenem Vertrauensniveau (J. Neyman, 1937)
 - Konstruktion eines Intervalls $[a,b]$ im Parameterraum aus Daten, so dass wahrer Wert von Parameter θ in festen Bruchteil wiederholter Experimente (z. B. 68.3%) in $[a,b]$ liegen würde
 - Aussage über die „Abdeckung“ (engl.: coverage) des wahren Werts durch das Intervall, nicht über Wahrscheinlichkeit des wahren Werts selbst
- Bemerkungen
 - Bei festem $1-\alpha-\beta$: Aufteilung in α, β im Prinzip beliebig
→ Messungen: zentrales Intervall: $\alpha = \beta$
 - Volle Neyman-Konstruktion aufwändig
→ Praxis: Bei bekannter Likelihood-Funktion gilt in guter Näherung

$$-\ln L(\hat{\theta}_{-d}^{+c}) = -\ln(\hat{\theta}) + \frac{1}{2}$$

Neyman-Konstruktion

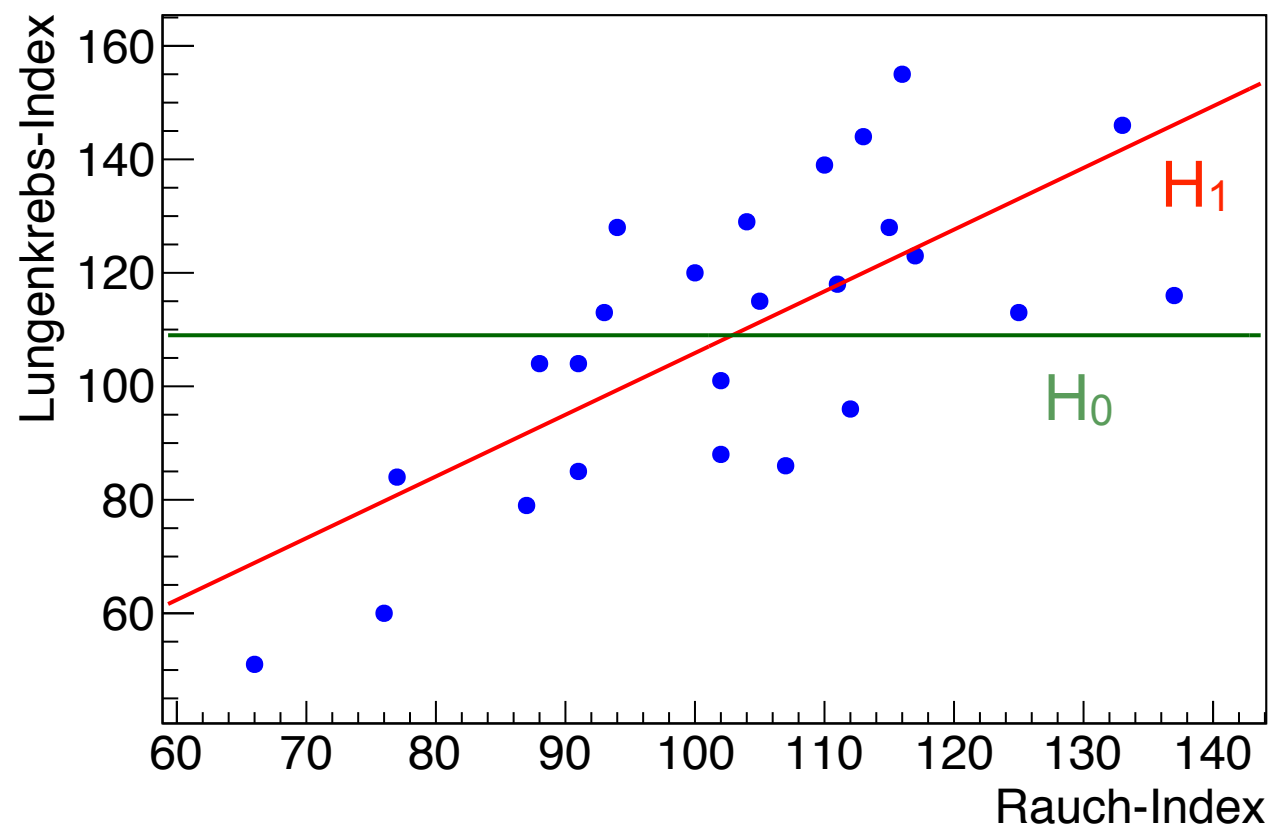


Kapitel 3.6

Hypothesentests

Verursacht Rauchen Lungenkrebs?

Rauchen und Lungenkrebs

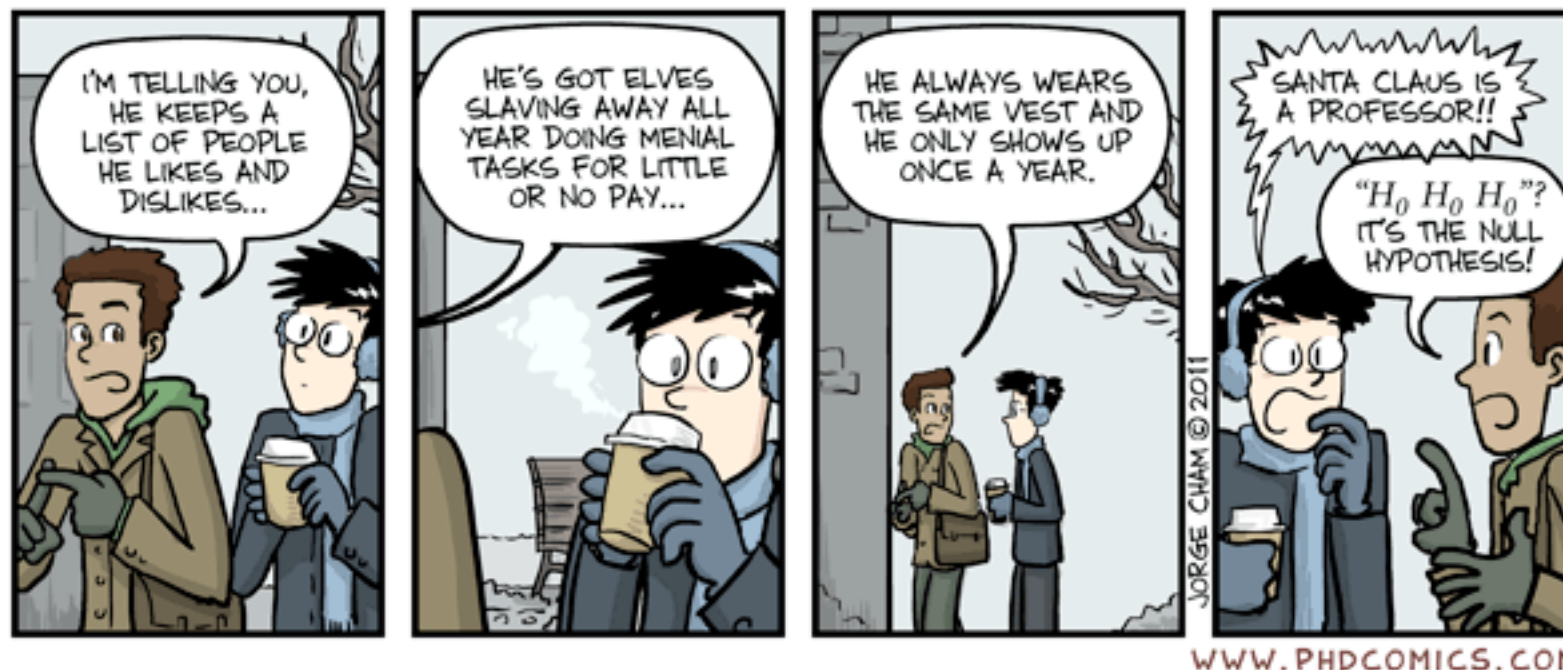


Daten aus: *Occupational Mortality: The Registrar General's Decennial Supplement for England and Wales, 1970-1972*, Her Majesty's Stationery Office, London, 1978.

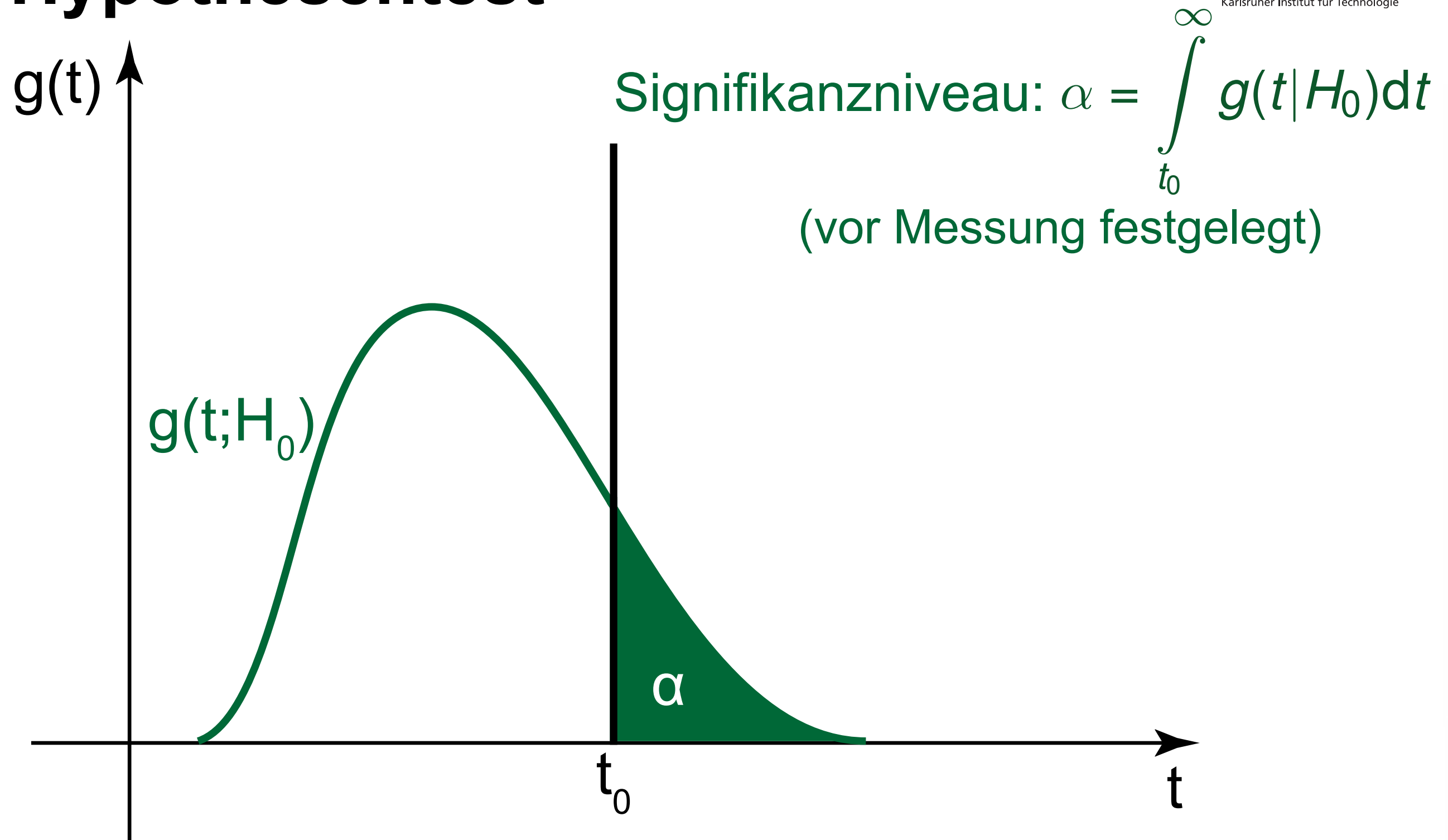
- Datensatz:
 - 25 Personengruppen
 - Rauch-Index: 100, wenn Zahl der Zigaretten pro Tag dem Durchschnitt von allen Männern desselben Alters entspricht
 - Lungenkrebs-Index: 100, wenn Zahl der Lungenkrebstoten dem Durchschnitt entspricht
- Korrelationskoeffizient: 0.716
- Frage: Kann „Nullhypothese“ H_0 (= kein Zusammenhang) verworfen werden?

Statistische Testverfahren

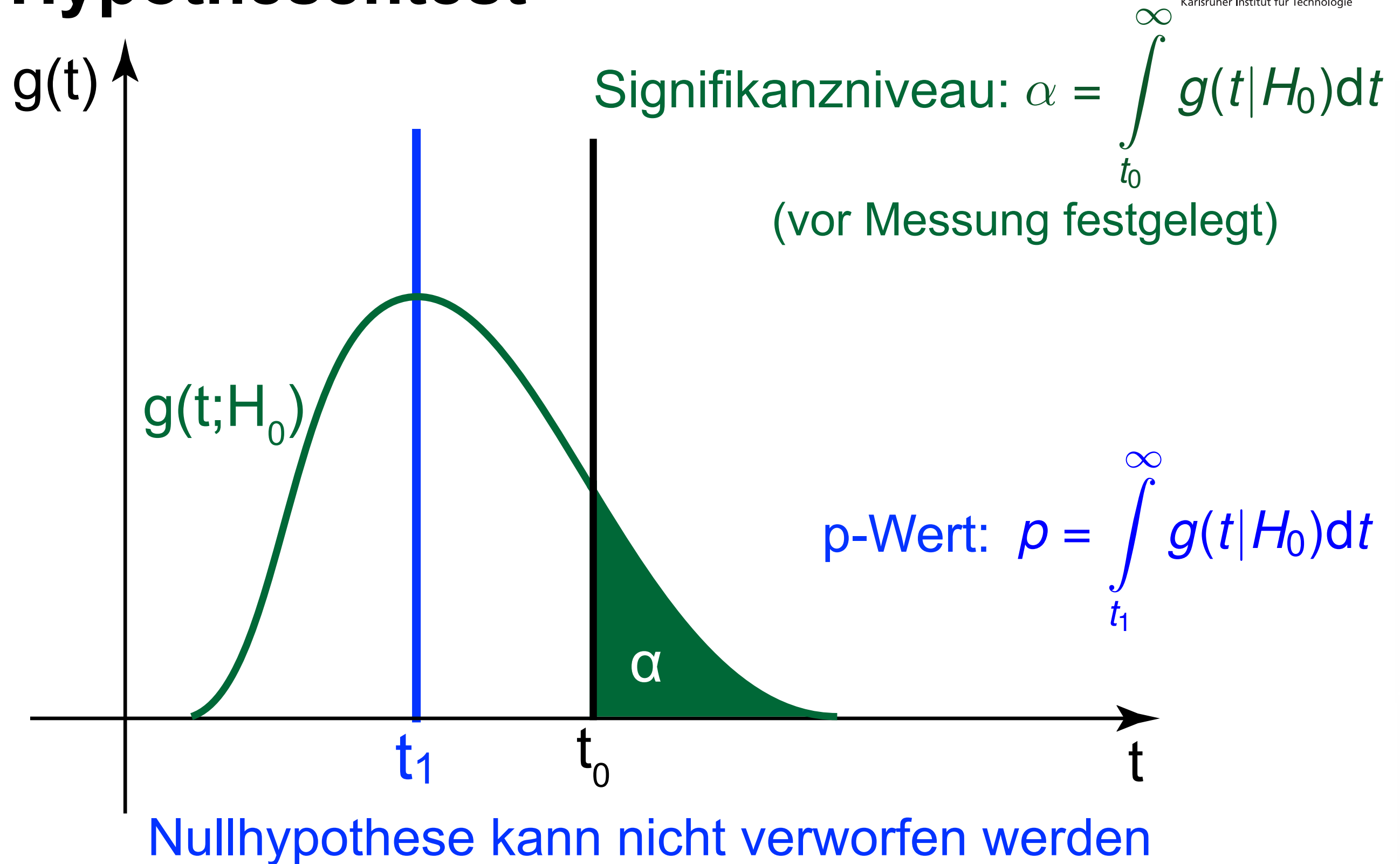
- Aufgabenstellung: vergleiche Stichprobe aus Daten mit einer oder mehreren Hypothesen H_i
 - Kann H_i aufgrund der Daten **verworfen** werden?
 - Wie sicher sind wir uns, dass Verwerfen **korrekt** war?
- Statistische Testverfahren:
 - **Quantitative** Methoden für Hypothesentest
 - Oft: Vergleich mit ausgezeichneten Hypothese $H_0 \rightarrow$ **Nullhypothese**



Hypothesentest



Hypothesentest



Fehler erster Art

(engl. „type-I error“,
auch „false positive“)

Nullhypothese mit
Wahrscheinlichkeit α verworfen,
selbst wenn sie wahr ist

Beispiele:

Krankheit fälschlich bei gesundem
Patienten diagnostiziert, fälschliche
Entdeckung eines neuen Teilchens

Fehler zweiter Art

(engl. „type-II error“,
auch „false negative“)

Nullhypothese nicht verworfen,
selbst wenn sie falsch ist

Beispiele:

Krankheit bei krankem Patienten nicht
diagnostiziert, neues Teilchen nicht
entdeckt, obwohl in Daten vorhanden

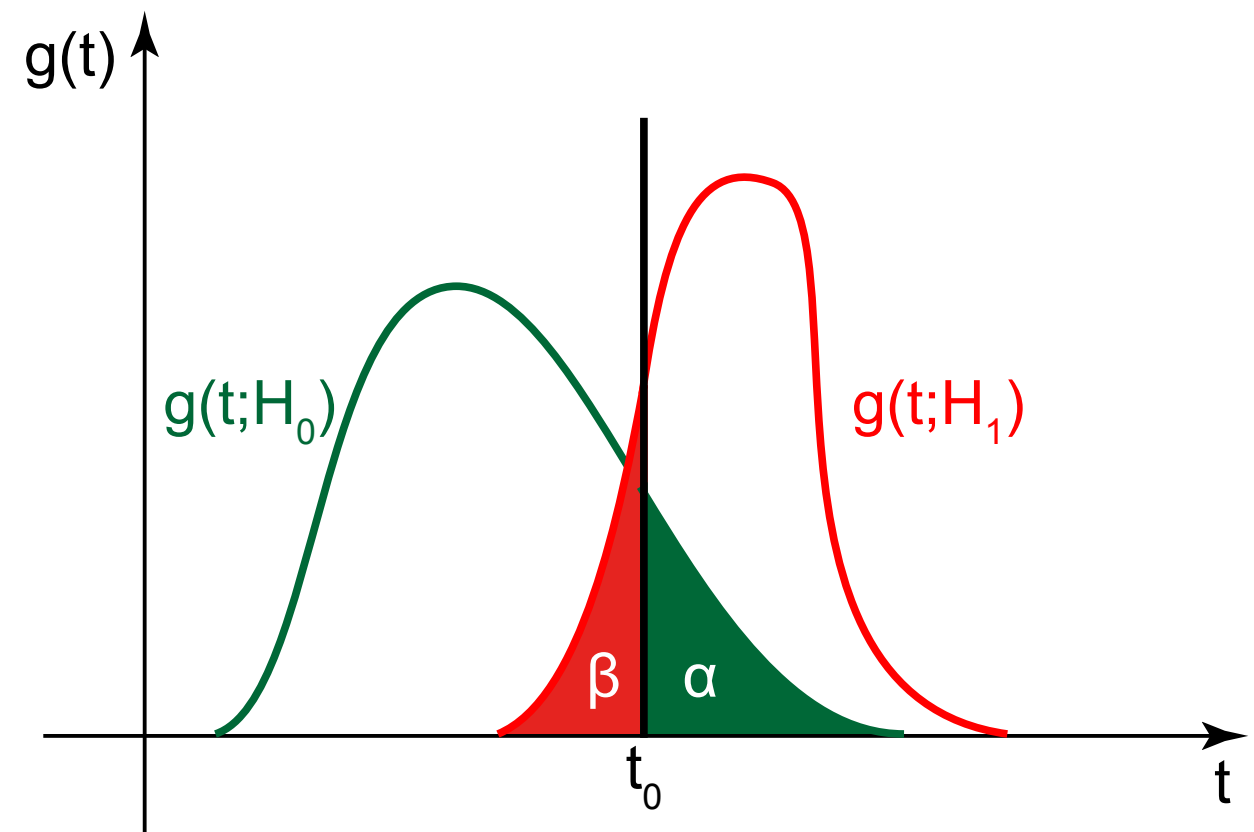
Wichtig: Signifikanzniveau/Irrtumswahrscheinlichkeit α **vor** dem Experiment festlegen

Was ist ein p-Wert?

- Unterscheidung: Signifikanzniveau und p-Wert
 - α = Wahrscheinlichkeit für Fehler erster Art (vor Messung festgelegt)
 - p = Wahrscheinlichkeit, dass Werte für Prüfgröße von $t \geq t_1$ gemessen würden, wenn Nullhypothese wahr ist (nachdem t_1 gemessen wurde)
- Häufige Missverständnisse
 - p-Wert ist nicht Wahrscheinlichkeit, dass Nullhypothese wahr
 - p-Wert ist nicht Wahrscheinlichkeit, Nullhypothese fälschlich zu verwerfen
 - p-Wert ist nicht Wahrscheinlichkeit, dass Messung „nur eine Fluktuation“

Vergleich von Hypothesen

- Alternative Hypothese H_1
- Wahrscheinlichkeit für Fehler zweiter Art:
$$\beta = \int_0^{t_0} g(t|H_1) dt$$
- Teststärke (engl. „power“): $1 - \beta$
- Bei gegebenem Signifikanzniveau: wie gut kann H_1 von H_0 unterschieden werden?



Wahl von Prüfgrößen

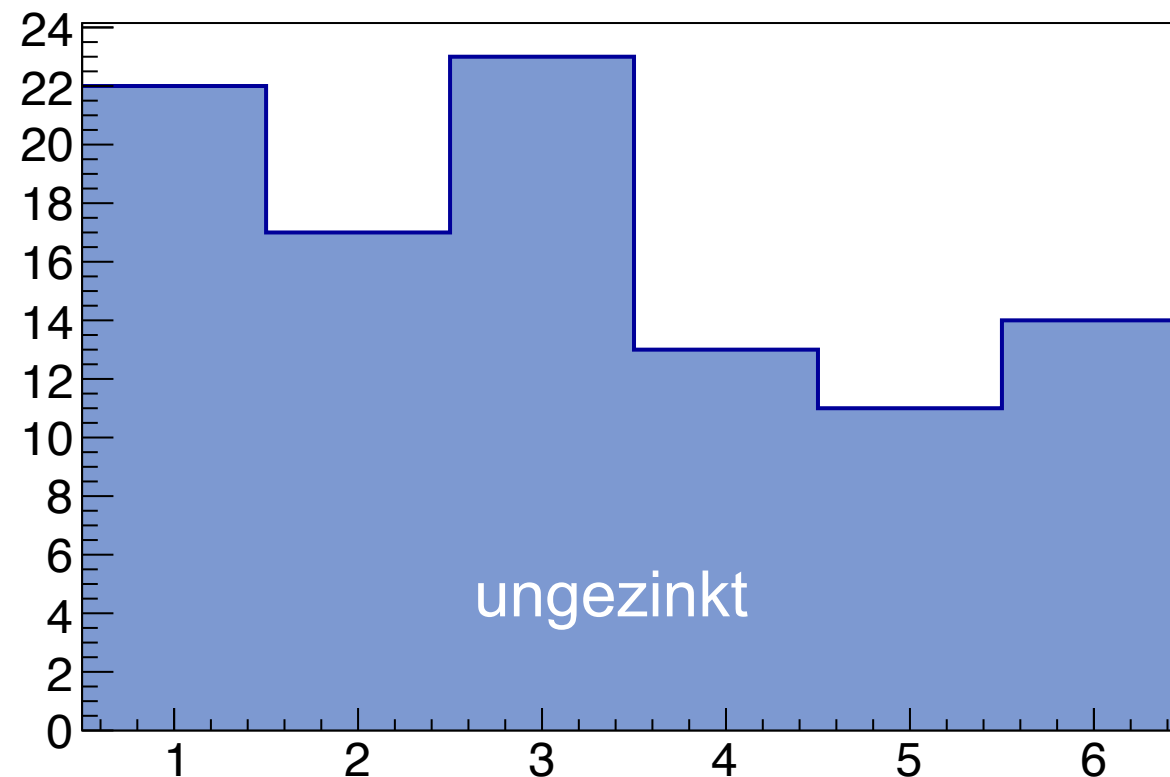
- Falls Likelihood-Funktionen $L(H_0)$ und $L(H_1)$ bekannt:
beste Prüfgröße ist **Likelihood-Quotient** (Neyman-Pearson-Lemma)

$$r(\vec{x}) = \frac{L(H_0)}{L(H_1)} = \frac{\prod_{i=1}^n f(x_i|H_0)}{\prod_{i=1}^n f(x_i|H_1)}$$

- Problem: Likelihood-Funktionen oft **unbekannt**
 - Lösung 1: Ansatz für funktionale Form der Prüfgröße
 - Lösung 2: Aufwändige Monte-Carlo-Simulation
 - Lösung 3: Asymptotische Lösungen (Satz von Wilks)
- Zwei Klassen von Prüfgrößen
 - **Güte der Anpassung**: Übereinstimmung nur mit Nullhypothese
 - Statistische **Klassifikation** (→ Vorlesung Datenanalyse)

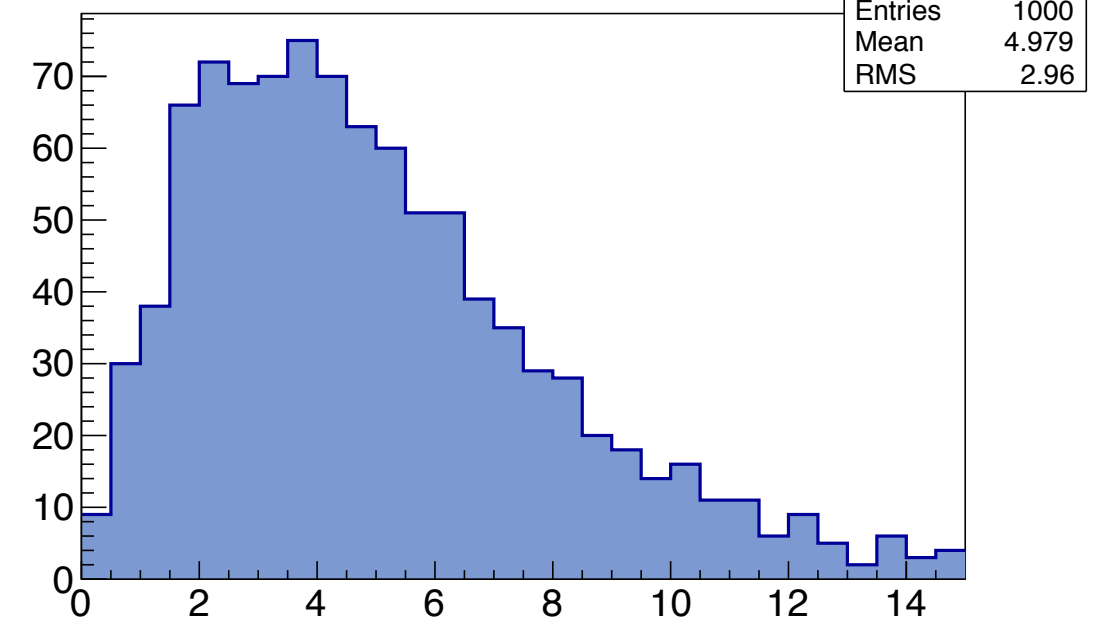
Gezinkter Würfel?

Gezinkter Wuerfel?

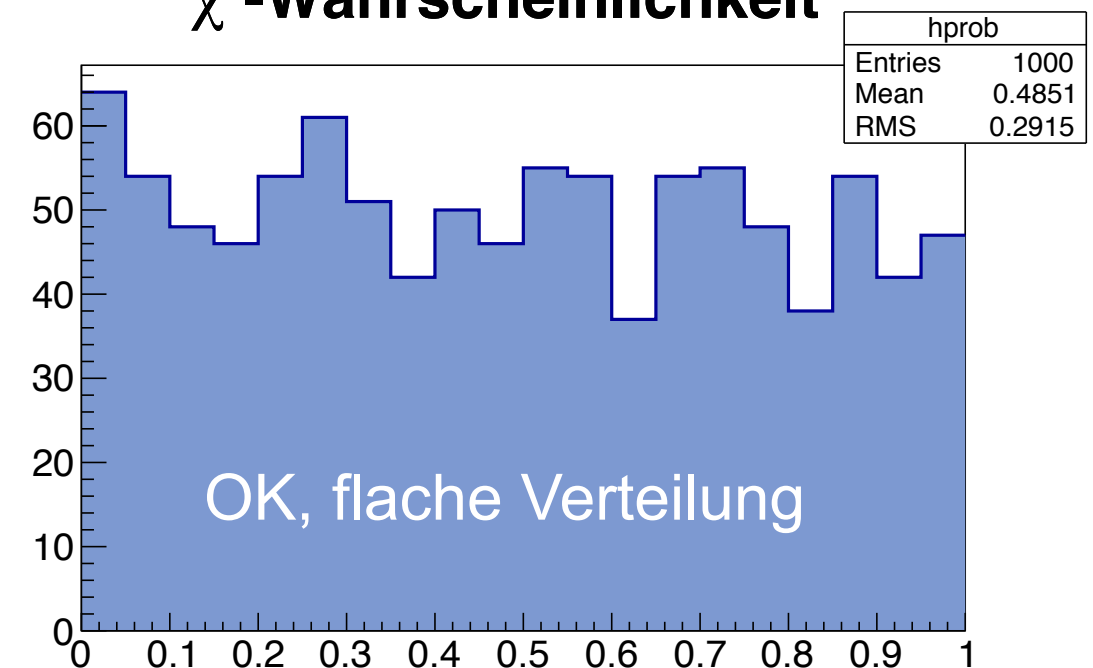


- 100 mal würfeln:
- Wahrscheinlichkeit 1/6 für jeden Bin
- Experiment 1000 wiederholen:
Prüfgröße χ^2 -verteilt, χ^2 -
Wahrscheinlichkeit flach

χ^2 -Pruefgroesse

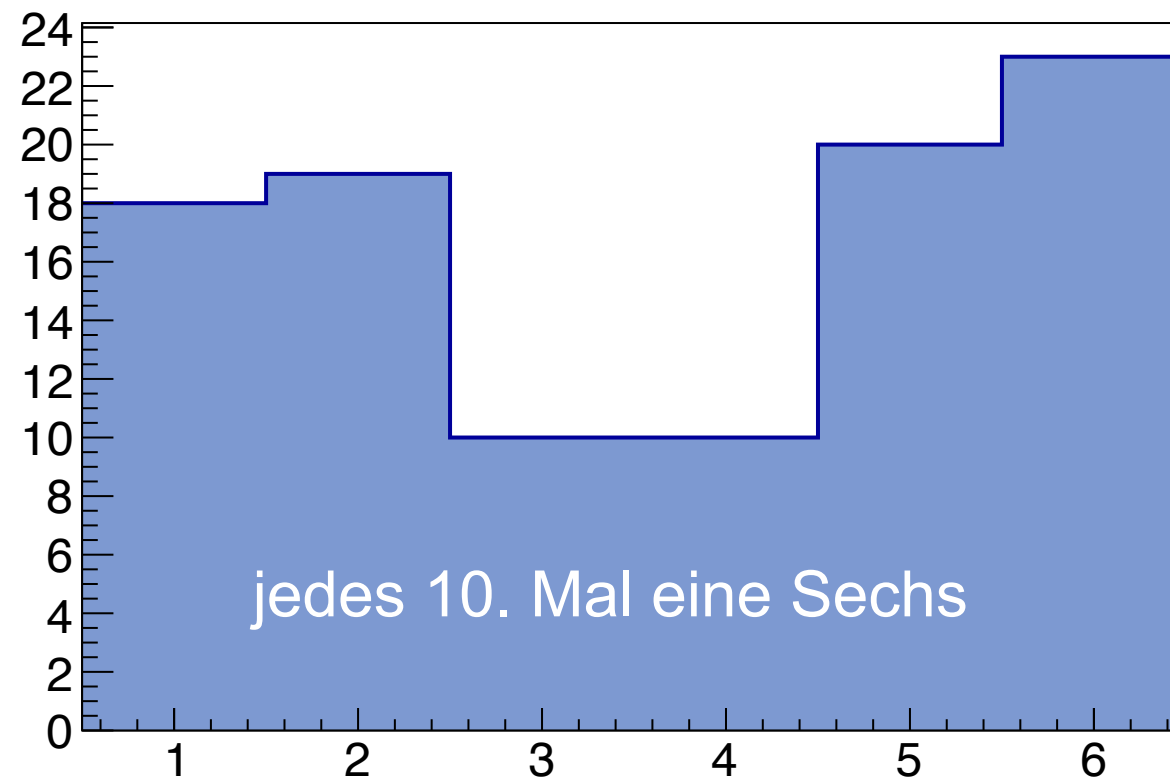


χ^2 -Wahrscheinlichkeit



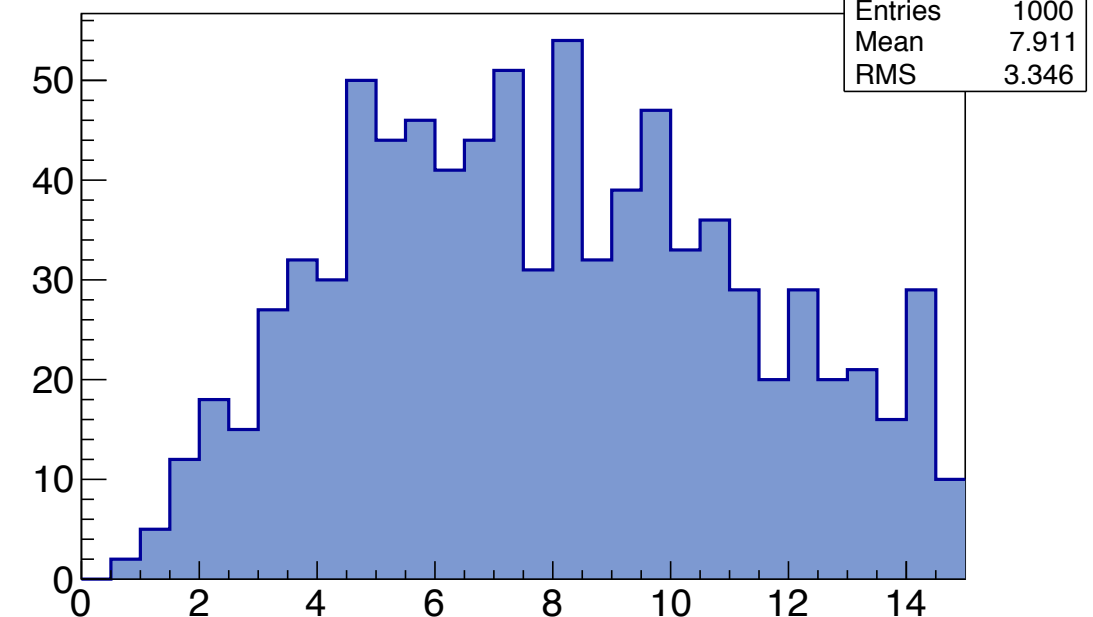
Gezinkter Würfel?

Gezinkter Wuerfel?

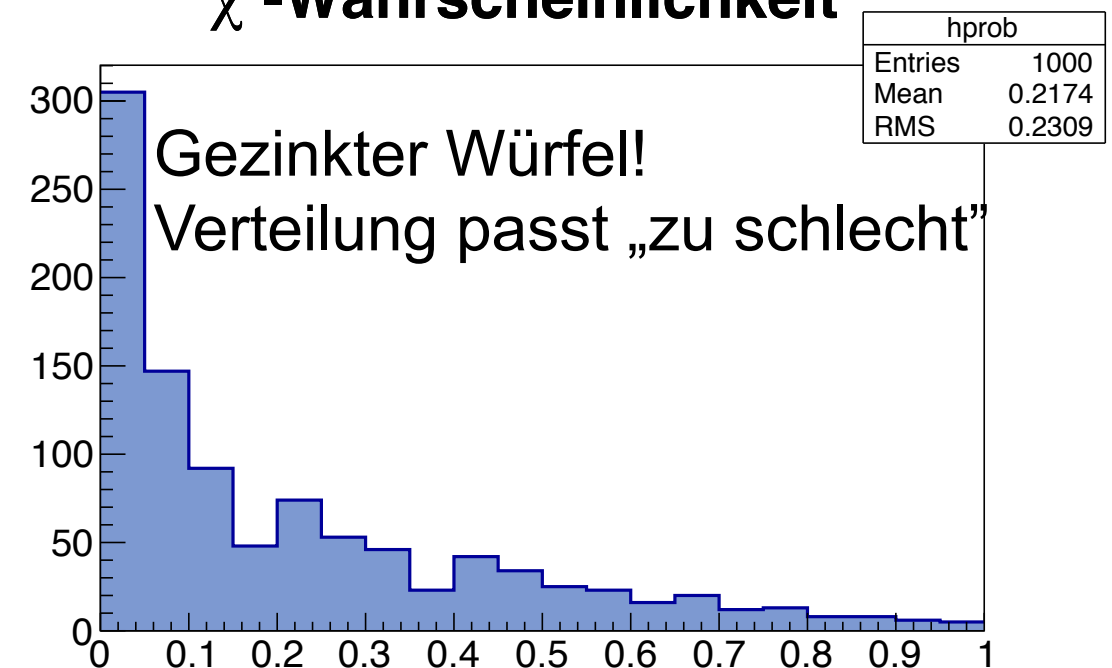


- 100 mal würfeln:
- Wahrscheinlichkeit 1/6 für jeden Bin
- Experiment 1000 wiederholen:
Prüfgröße χ^2 -verteilt, χ^2 -
Wahrscheinlichkeit flach

χ^2 -Pruefgroesse

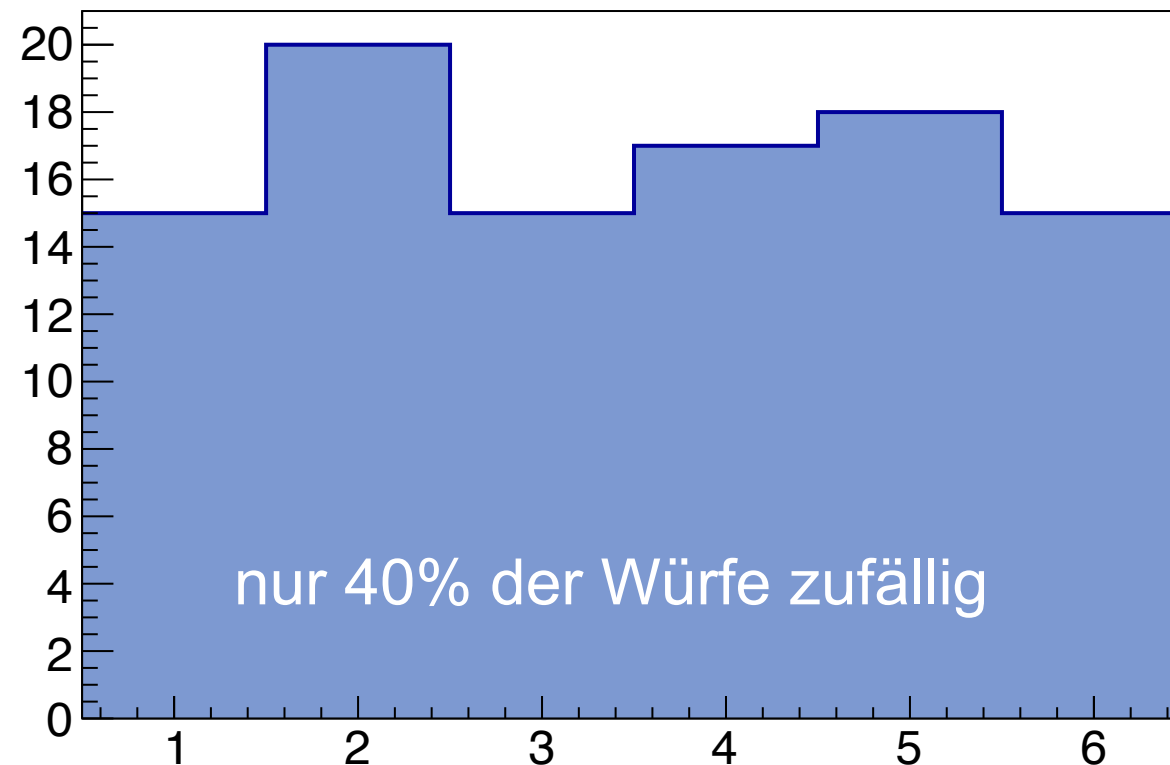


χ^2 -Wahrscheinlichkeit



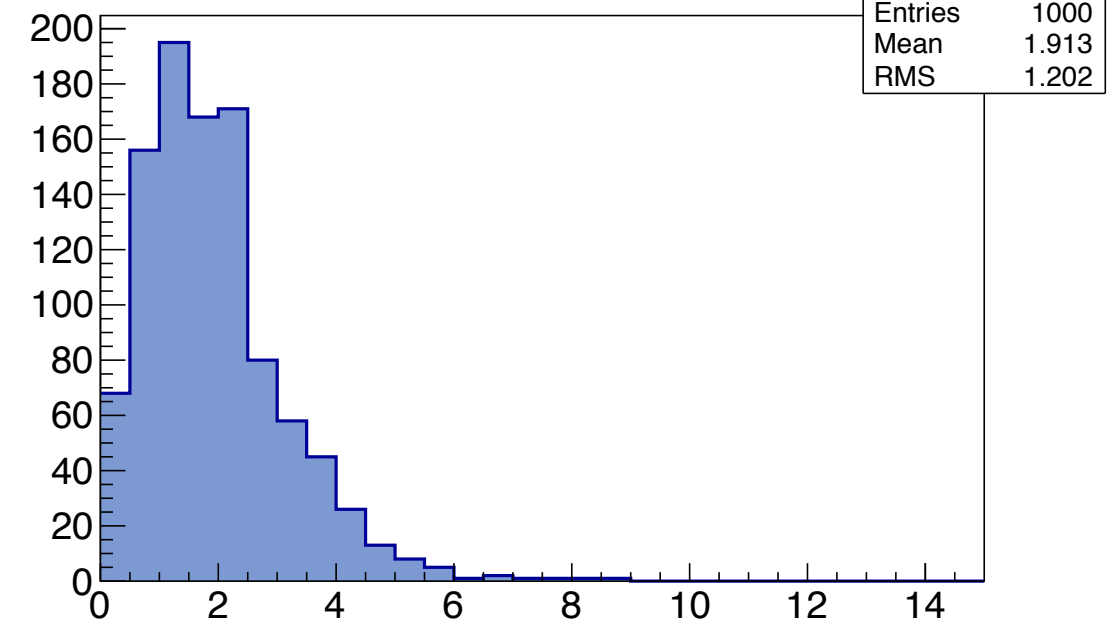
Gezinkter Würfel?

Gezinkter Wuerfel?



- 100 mal würfeln:
- Wahrscheinlichkeit 1/6 für jeden Bin
- Experiment 1000 wiederholen:
Prüfgröße χ^2 -verteilt, χ^2 -
Wahrscheinlichkeit flach

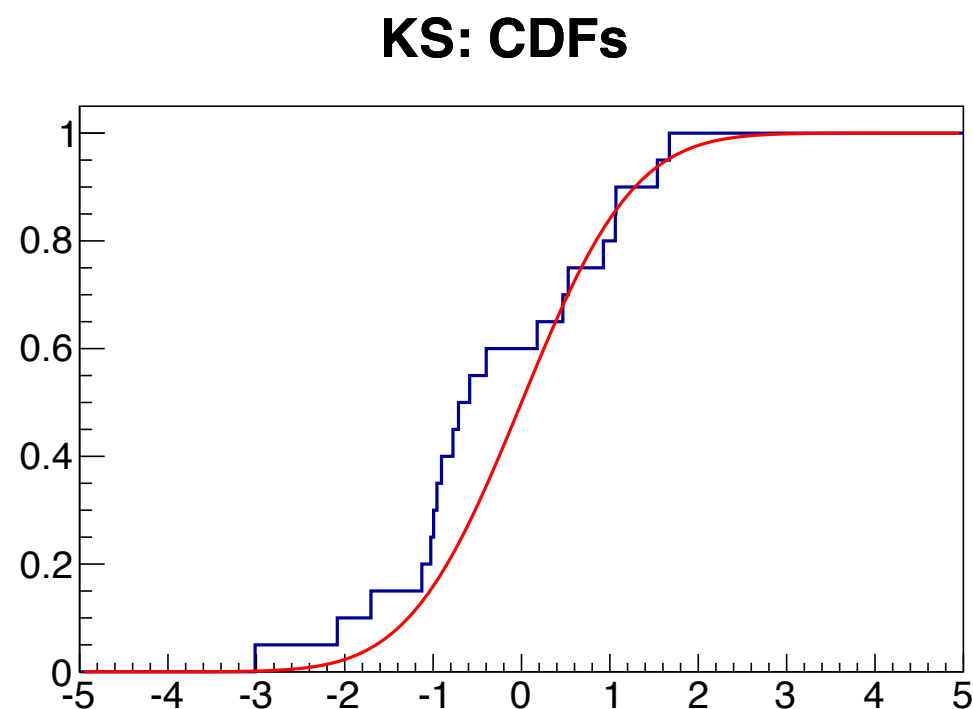
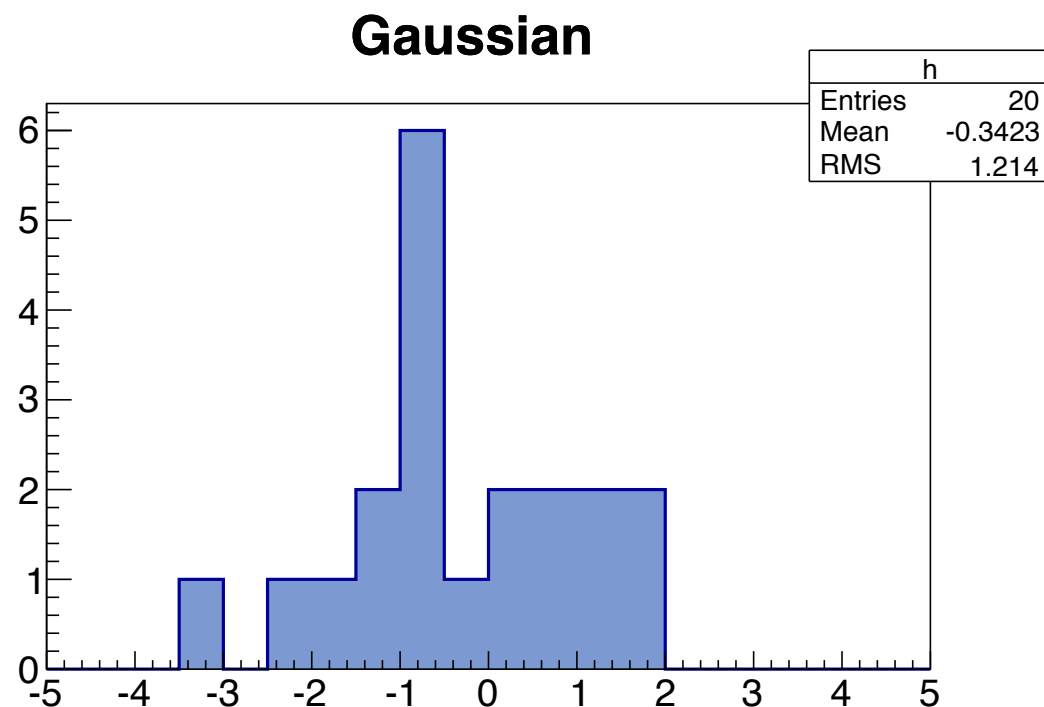
χ^2 -Pruefgroesse



χ^2 -Wahrscheinlichkeit



Kolmogorov-Smirnov-Test



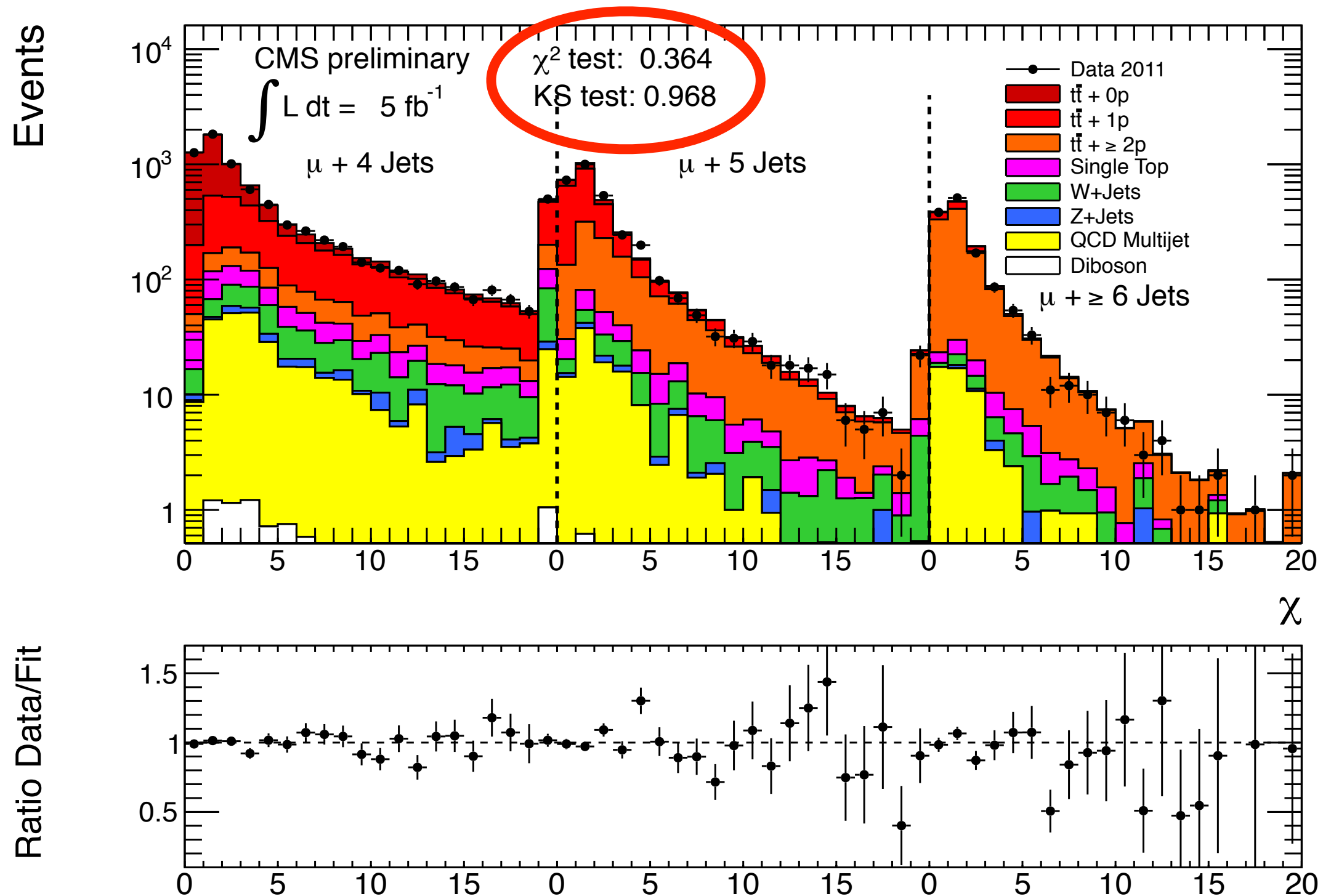
- Beispiel: 20 Zufallszahlen aus Gaußverteilung
- Vergleich der empirischen CDF mit Gauß-CDF
 - KS-Abstand: 0.271
 - p-Wert: 0.106

Vergleich: χ^2 - und KS-Test

	χ^2 -Test	Kolgomorov-Smirnov-Test
Anwendung	Nur Kategorien von Daten (z. B. Bins eines Histogramms)	Rohdaten und Kategorien
Art des Vergleichs	Kategorie für Kategorie	Globaler Vergleich über CDFs
Vorzeichen der Abweichung	Irrelevant	Wichtig (z. B. Daten systematisch zu hoch/niedrig)

Komplementäre „goodness-of-fit“-Tests → Praxis: beide Tests anwenden

χ^2 - und KS-Test: Beispiel



[CMS-PAS-12-018]

- Was sollten Sie aus dem dritten Teil der Vorlesung Rechnernutzung mitnehmen?
 - Statistische Werkzeuge sehr wichtig der Physik:
Was passiert „im Inneren“?
 - „Kochrezepte“ der statistischen Datenanalyse (z. B. Gauß'sche Fehlerfortpflanzung, Ausgleichsgerade):
Welche Voraussetzungen müssen erfüllt sein?
 - Angewandte Statistik: Monte-Carlo-Methoden sehr nützlich
- Haben wir Ihr Interesse geweckt?
 - Bachelorarbeiten am Institut für Experimentelle Kernphysik im Bereich der Datenanalyse auf Anfrage
 - Besuchen Sie die Vorlesung Datenanalyse