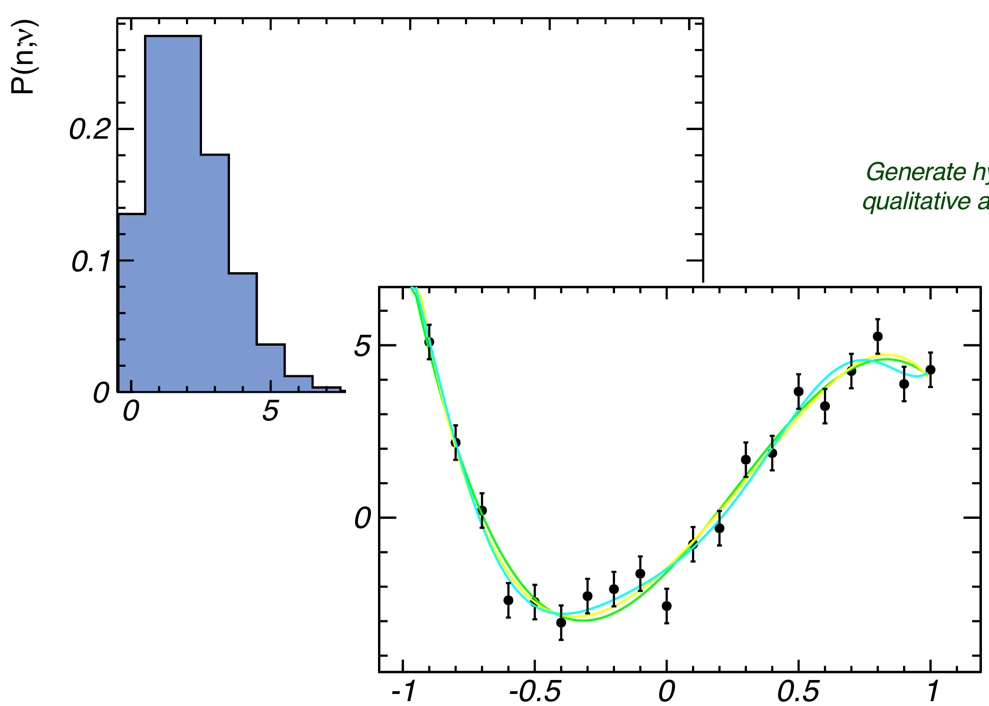
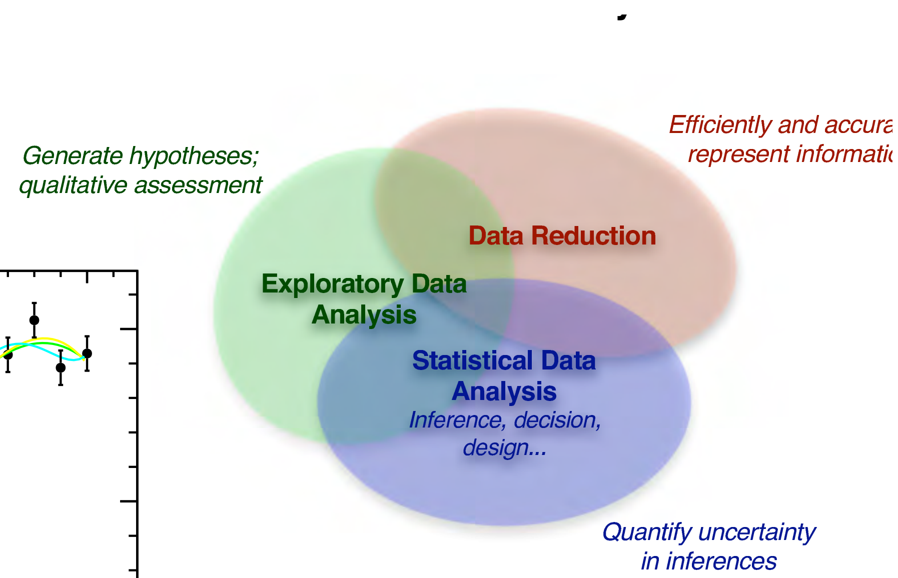
Rechnernutzung in der Physik

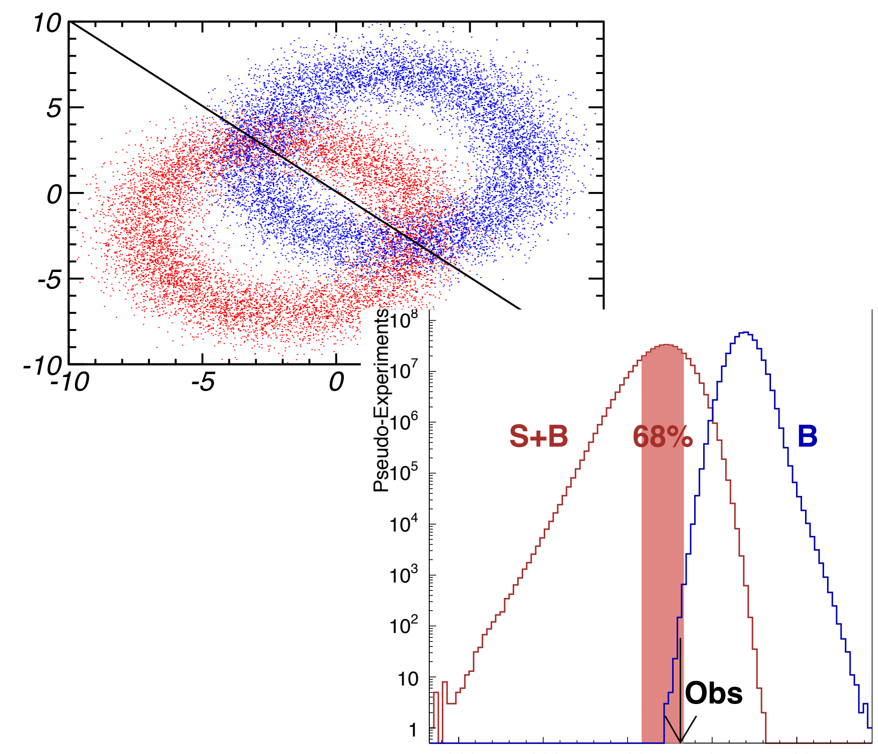
Teil 3 – Statistische Methoden der Datenanalyse

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

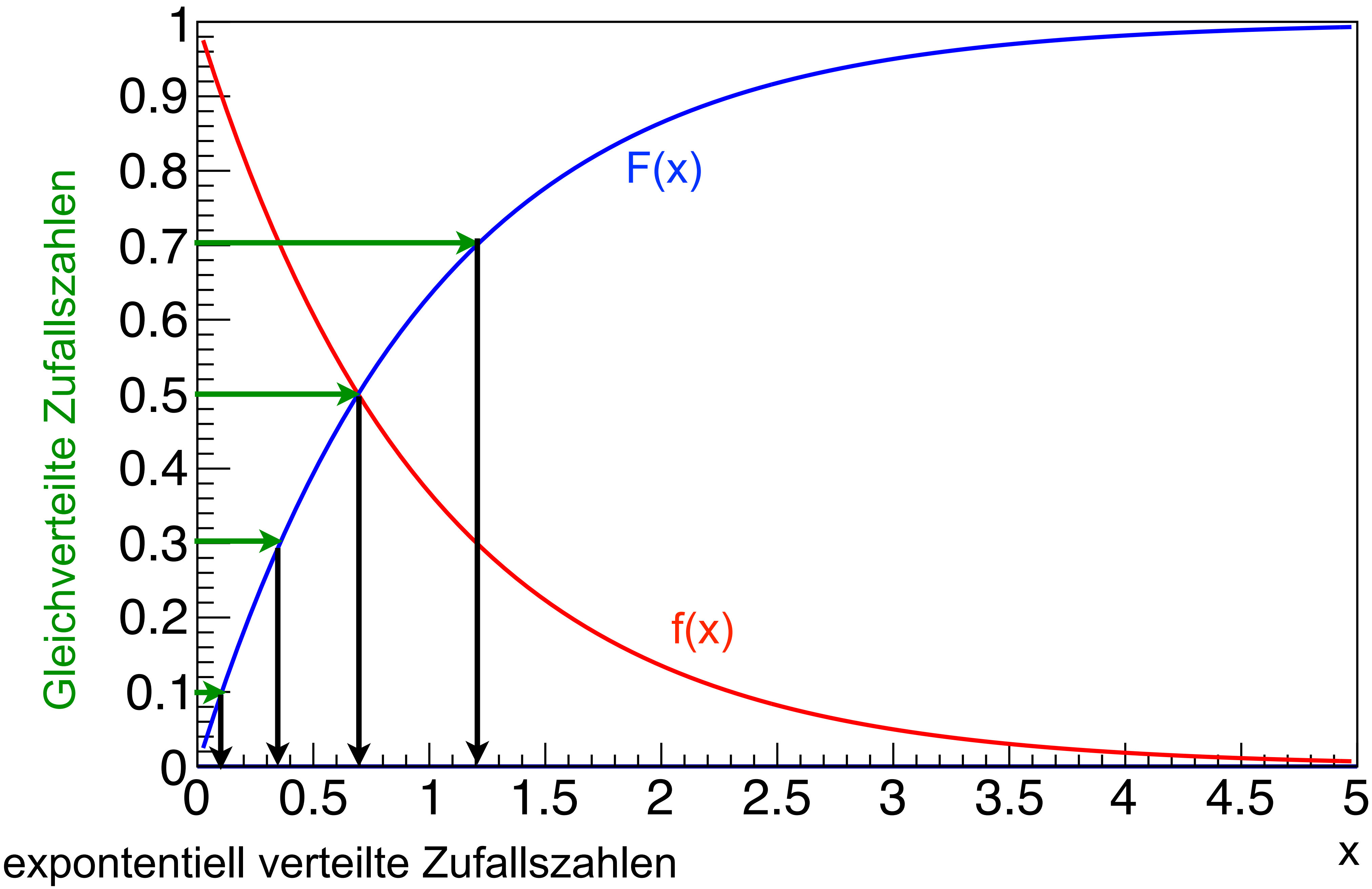
Karlsruher Institut für Technologie

- Beispiele für Wahrscheinlichkeitsverteilungen
  - Binomialverteilung („Erfolg-Miserfolg”) $$P(n; N, p) = \binom{N}{n} p^n (1-p)^{N-n}$$
  - Poissonverteilung („Zählexperiment”) $$P(n; \nu) = \frac{\nu^n}{n!} e^{-\nu}$$
  - Gaußverteilung (zentraler Grenzwertsatz) $$G(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$
  - Log-Normalverteilung $$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$$
  - χ²-Verteilung $$\chi^2(x; n) = \frac{x^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \exp\left[-\frac{x}{2}\right]$$
  - Cauchy-Lorentz-Verteilung (Resonanzen) $$f(x; x_0, \Gamma) = \frac{1}{\pi} \frac{\Gamma}{(x-x_0)^2 + \Gamma^2}$$
- Monte-Carlo-Methoden
  - Numerische Techniken zur Berechnung von W.keiten mit Zufallszahlen
  - Pseudozufallszahlen im Computer: lange Periode, so wenig Korrelation wie möglich, Startwert („seed”) geschickt wählen (z. B. Mersenne-Twister)

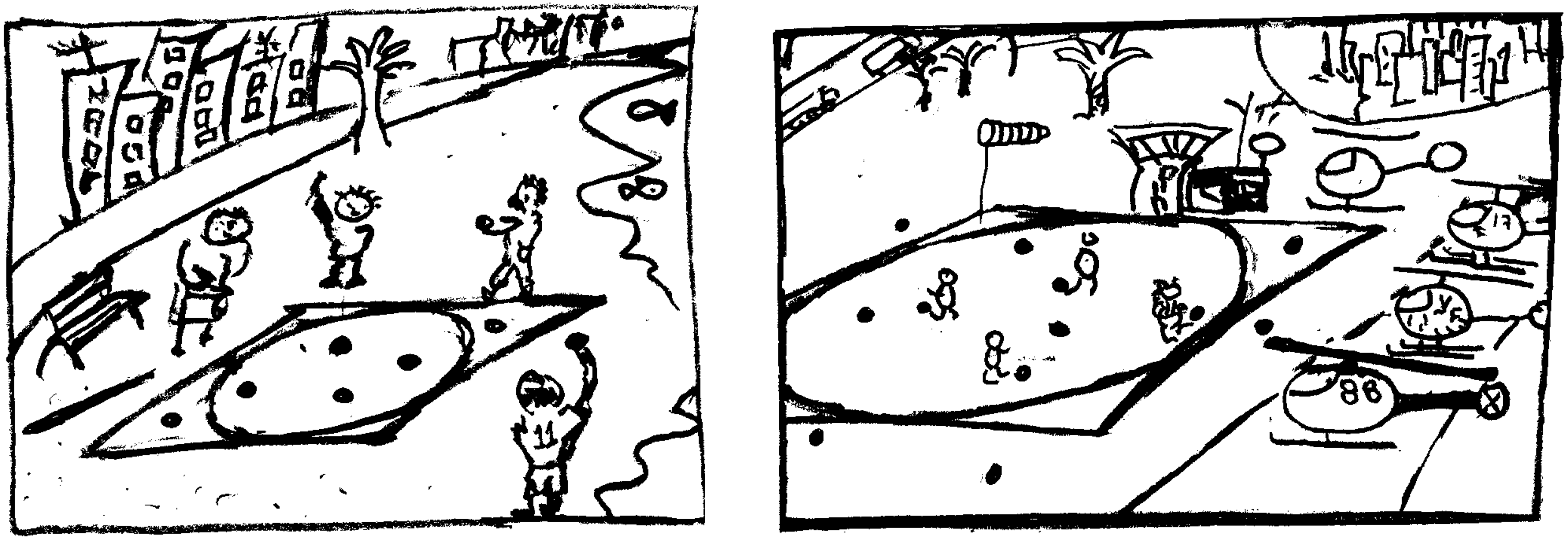
Kapitel 3.4

Monte-Carlo-Methoden

Transformationsmethode



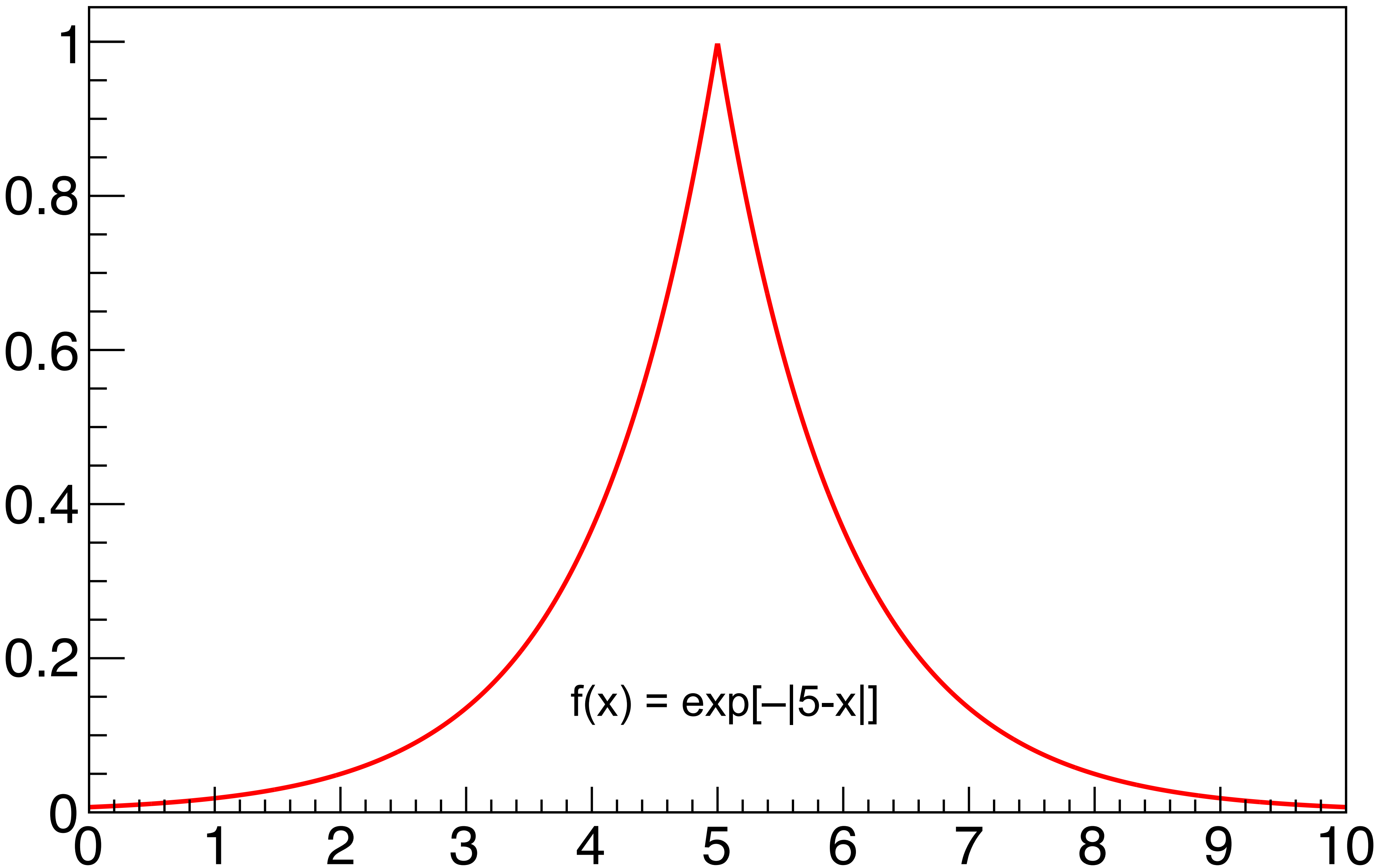
Beispiel 1: Bestimmung von π



[W. Krauth, arXiv:cond-mat/9612186v2]

- Idee: Verhältnis der Kreisfläche zur Quadratfläche π/4
  - Würfle gleichverteilte Zufallszahlen x, y
  - Verwerfungsmethode: liegt (x,y) innerhalb des Kreises?
  - Vgl. Übungen

Beispiel 2: MC-Integral



Analytische Lösung: 2(1–exp[–5]) = 1.9865

Genauigkeit der MC-Methode

- MC-Integration: $\int_{x_1}^{x_2} f(x)\,\mathrm{d}x = \lim_{N\to\infty} \frac{N_{hit}}{N} \cdot (x_2 - x_1) \cdot y_{max}$
  - Effizienz $\varepsilon = N_{hit}/N$ → Genauigkeit aus Binomialverteilung: $V[\varepsilon] = \varepsilon(1-\varepsilon)/N$
  - Verbesserung mit $\sqrt{N}$ unabhängig von Dimension d des Integrals
- Vergleich: Quadraturformeln, z. B. Trapezregel
  - Effizienz skaliert mit Zahl der Stützstellen n: $V[\varepsilon] \sim n^{-4}$
    → viel besser als MC-Methode
  - Höherdimensionale Integrale (d Dimensionen): $V[\varepsilon] \sim n^{-4/d}$
    → MC-Integration ab 4 Dimensionen effizienter

Anwendungen der MC-Methode

- Numerische Mathematik, z. B. Integration, Optimierung
- Angewandte Statistik, z. B. PDFs, Fehlerfortpflanzung, Hypothesentests
- Simulation stochastischer Prozesse
  - Vielteilchensysteme, z. B. Ising-Modell für gekoppelte Spins
  - Teilchenphysik: quantenmechanische Prozesse, z. B. Produktion von Elementarteilchen → „MC-Generatoren” (z. B. Pythia)
  - Wechselwirkung von Teilchen in Materie, z. B. Neutrondiffusion, Streuung

Kapitel 3.5

Parameterschätzung

Grundbegriffe

- Bislang: Wahrscheinlichkeitstheorie → jetzt: Arbeit mit Daten

Wahrscheinlichkeitstheorie	Messung
PDF = ideale Wahrscheinlichkeitsverteiliung	Beobachtete Ereignisse = endliche zufällige Stichprobe aller mögliche Ergebnisse (engl.: random sample)
Bekanntes W.keitsmodel mit bekannten Parametern	Konstruktion einer "Likelihood-Funktion" (später) aus Daten → Schätzung der Parameter

- Ziel: Schätzung einer (hoffentlich kleinen) Zahl von Modellparametern aus (potenziell großer) Datenmenge

Grundbegriffe

- Stichprobe der Größe n:
  - n-dimensionaler Vektor $\mathbf{x} = (x_1, \ldots, x_n)$ unabhängiger Zufallsvariablen x_i,
  - x_i stammen aus einer Grundgesamtheit (auch: Population) der Größe N
  - x_i folgen derselben PDF f(x), ggf. mit weiteren Parametern
  - Beispiel: 10 Längenmessungen
- Prüfgröße (auch: Kenngröße, engl.: statistic)
  - Beliebige Funktion der Daten der Stichprobe ohne freie Parameter (Stichprobe wird dabei als „eine Messungen” interpretiert)
  - Beispiel: arithmetisches Mittel der Längenmessung $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Schätzfunktion (auch: Schätzer, engl.: estimator)
  - Prüfgröße zur Schätzung von Eigenschaften der unbekannten PDF f(x)
  - Beispiel: arithmetisches Mittel als Schätzer für „wahren” Erwartungswert

Parameterschätzung: Methoden

- Ziel: Schätzung eines Parameters θ aus Daten
  - z. B. Steigung einer Ausgleichgeraden
- Gesucht: systematische Methoden, Schätzer mit „guten” Eigenschaften zu konstruieren (z. B. erwartungstreu, möglichst kleine Varianz, …)
- Methode der kleinsten (Fehler-)Quadrate (engl.: least squares method, LS)
  - Gauß, Laplace, Legendre, ca. 1800
  - Bekannt z. B. aus Praktikum („Ausgleichsgerade”)
- Maximum-Likelihood-Methode (ML)
  - Fisher, ca. 1912
  - Sehr universell einsetzbar

LS und ML: Erster Vergleich

Karlsruher Institut für Technologie

- Ausgangsfrage: welches Modell passt am besten zu den Daten?

Maximum Likelihood: maximiere Höhe der PDF

(PDF muss bekannt sein)

Kleinste Quadrate: minimiere Abstand vom Erwartungswert

(nur Erwartungswert bekannt)