

# Experimentelle Methoden der Teilchenphysik

Sommersemester 2011/2012

Albert-Ludwigs-Universität Freiburg



Prof. Markus Schumacher

Physikalisches Institut, Westbau, 2. OG Raum 008

Telefon 07621 203 7612

E-Mail: [Markus.Schumacher@physik.uni-freiburg.de](mailto:Markus.Schumacher@physik.uni-freiburg.de)

Kapitel 9: Grundlagen der stat. Methoden der Datenanalyse

<http://terascale.physik.uni-freiburg.de/lehre/Sommersemester%202012>

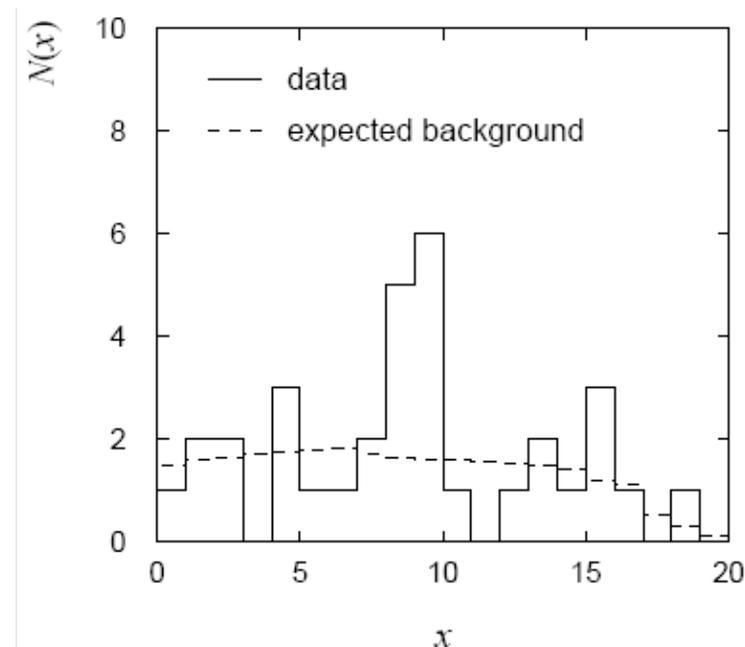
# Motivation

- Erkenntnisgewinn in den “exakten” Wissenschaften basiert auf Wechselspiel von Theorie (Modellbildung) und Experiment (Messungen und Datenanalyse)
- **Verbindung:** quantitative Interpretation mit statistischen Methoden
- **Modellbildung in der Theorie:**
  - Objekte die beschrieben werden (bekannte und hypothetisch postulierte)  
Gibt es weitere unbekannte bzw. unentdeckte Objekte?
  - **Gesetzmäßigkeiten/ Verteilungen abhängig von Parametern**  
Stimmen die Gesetze?  
Brauchen wir neue Gesetze oder Erweiterungen der bekannten?  
Welche Werte haben die Parameter und wie genau kennen wir diese?

# Aufgaben der statistischen Datenanalyse

- Darstellung und Beschreibung der Daten
- Bestimmung des “besten” Wertes für einen unbekannt Parameter
- Bestimmung eines Intervalls oder von Grenzen, innerhalb der unbekannte Parameter mit einer gewissen Konfidenz liegen sollten
- Quantifizierung der Übereinstimmung zwischen Messdaten und Modellen (zwischen verschiedenen Datensätzen)
- Vergleich mehrerer Hypothesen bzgl. der Übereinstimmung mit den Daten
- Treffen von Entscheidungen auf der Grundlage der Messdaten

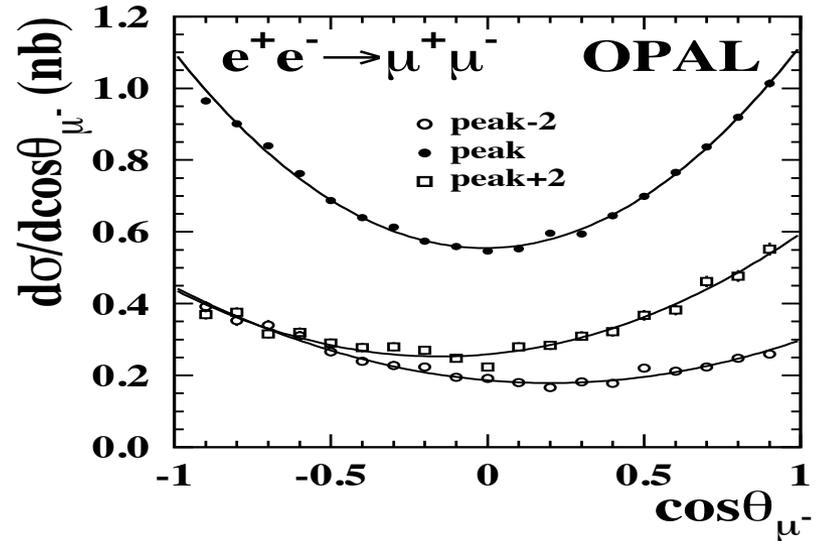
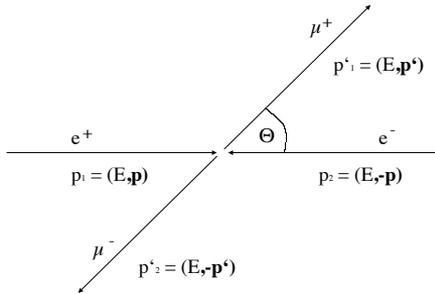
- richtige und wichtige Fragen stellen
- Ergebnisse kritisch hinterfragen
- bei Interpretation Annahmen und Methoden klar darstellen



# Beispiele

Bestimmung von Parametern

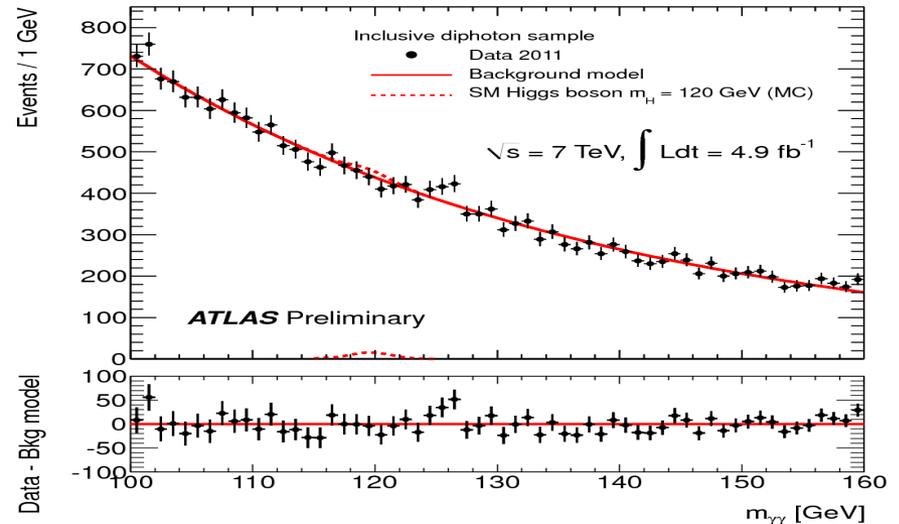
Streuhäufigkeiten:  $x = \cos\theta$   
 $d\sigma/dx = (1 + a x + b x^2)$



Hypothesentest:

Suche nach neuen Teilchen

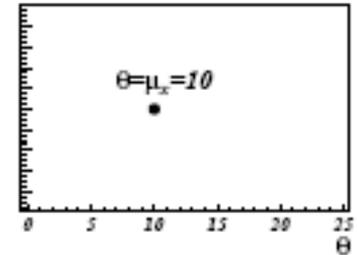
Gibt es ein Higgs-Signal  
 oder nur den Untergrund?



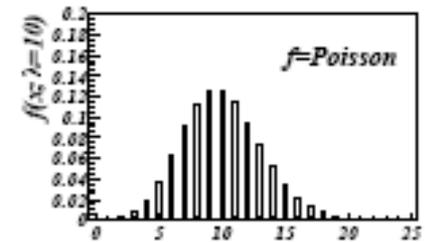
# Bsp: Theorie = Poissonverteilung mit Mittelwert 10

- Modell: Zählrate folgt Poissonverteilung mit unbekanntem Mittelwert
- Theorie sagt Verteilung der Zählrate  $f(n; \text{Mittelwert})$  vorher
- Messung = Stichprobe von  $m$  Zählraten
- Quantifizierung der Übereinstimmung zwischen Modell (Poisson) und Stichprobe und Modellen
- Bestimmung eines Schätzwertes für den Mittelwert und dessen Fehler

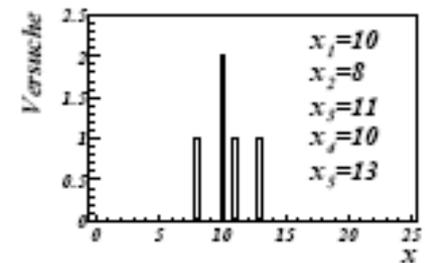
Modell  
mit Parameter  $\theta$



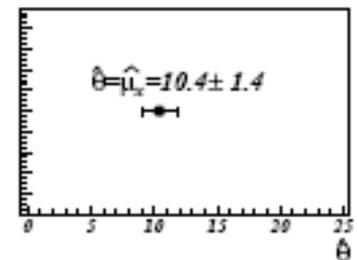
Observable  $x$   
mit Verteilung  $f(x; \theta)$



Stichprobe  
 $x_1, \dots, x_N$



Schätzwert  $\hat{\theta}$   
Fehler von  $\hat{\theta}$



# Der Begriff der Wahrscheinlichkeit

In den Naturwissenschaften gibt es verschiedene Elemente der Unsicherheit:

Unsicherheit: Ausgang eines Experimentes/Messung  
unvorhersagbar bei Wiederholung



Gründe: die Theorie ist nicht deterministisch z. B. Quantenmechanik,

zufällige Messfehler (Messapparatur, Experimentator/Beobachter)  
anwesend auch ohne Quanteneffekte

Randbedingungen, die man im Prinzip kennen könnte,  
aber es nicht tut, z.B. durch Begrenzungen an Zeit, Geld, ...

Wir können diese Unsicherheit durch das Konzept der  
**WAHRSCHEINLICHKEIT** quantifizieren

# Axiomatische Definition der Wahrscheinlichkeit

Betrachte Menge  $S$  (Grundgesamtheit) mit Untermengen  $A, B, \dots$

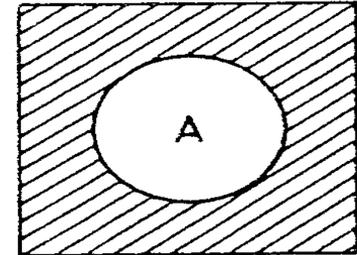
Ordne jeder Teilmenge eine Zahl  $P$  zwischen 0 und 1 zu, so dass gilt:

For all  $A \subset S, P(A) \geq 0$

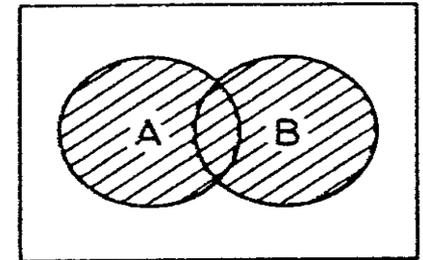
$$P(S) = 1$$

If  $A \cap B = \emptyset, P(A \cup B) = P(A) + P(B)$

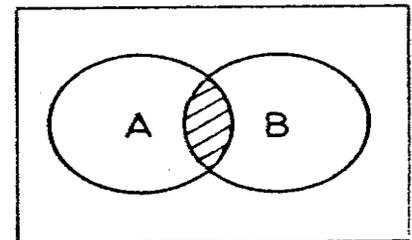
**Kolmogorov  
Axiome (1933)**



$\bar{A}$



$A \cup B$



$A \cap B$

Aus diesen Axiomen können wir  
weitere Eigenschaften ableiten

$$P(\bar{A}) = 1 - P(A)$$

$$P(A \cup \bar{A}) = 1$$

$$P(\emptyset) = 0$$

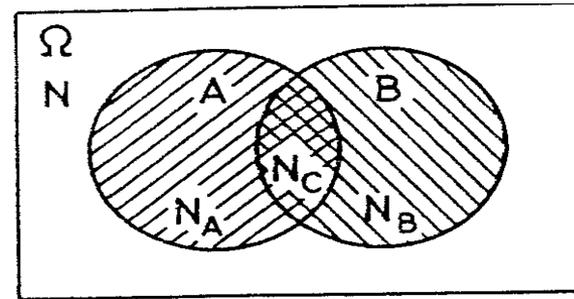
if  $A \subset B$ , then  $P(A) \leq P(B)$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

# Bedingte Wahrscheinlichkeit

Wir definieren weiterhin die bedingte Wahrscheinlichkeit:  
A gegeben B (mit  $P(B) \neq 0$ )

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



z.B. Würfeln:  $P(n < 3 | n \text{ even}) = \frac{P((n < 3) \cap n \text{ even})}{P(\text{even})} = \frac{1/6}{3/6} = \frac{1}{3}$

Wenn Teilmengen A, B **unabhängig**, dann gilt  $P(A \cap B) = P(A)P(B)$

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

Nicht zu verwechseln mit disjunkten Teilmengen: i.e.  $A \cap B = \emptyset$

Die bedingten Wahrscheinlichkeiten erfüllen Kolmogorov-Axiome

# Das Theorem von Bayes

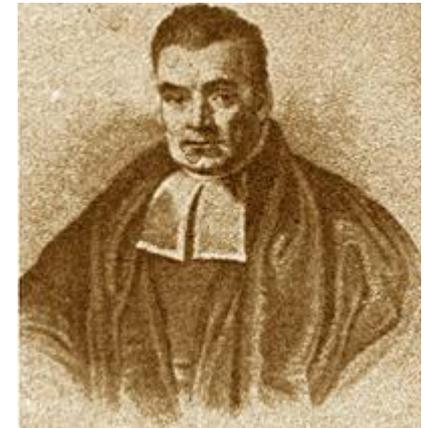
Aus der Definition der bedingten Wahrscheinlichkeit erhalten wir

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{und} \quad P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$P(A \cap B) = P(B \cap A)$$

Bayes-Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



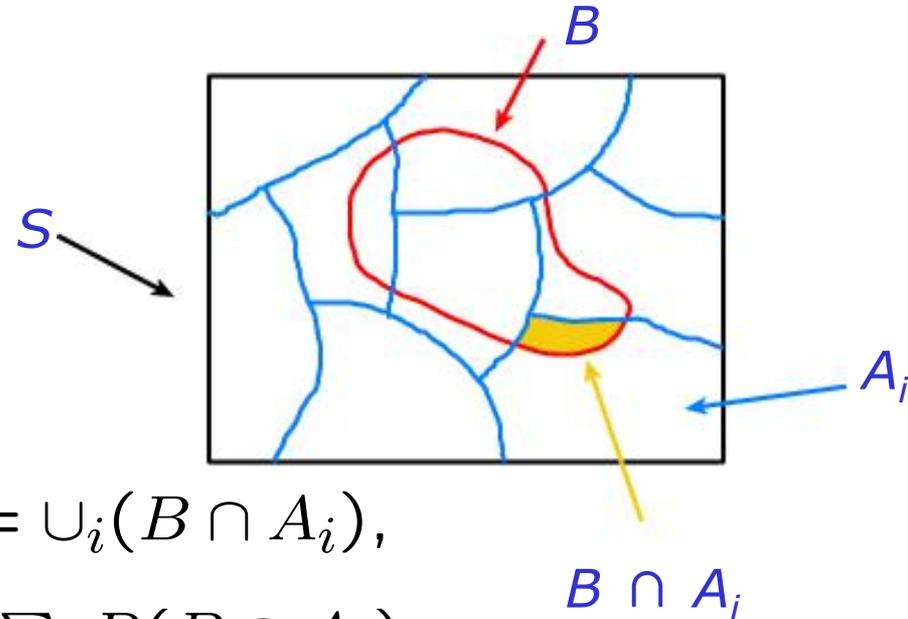
Erstmals publiziert (posthum) durch den Reverend Thomas Bayes (1702–1761)

*An essay towards solving a problem in the doctrine of chances*, Philos. Trans. R. Soc. **53** (1763) 370; Reprint in Biometrika, **45** (1958) 293

# Das Gesetz der Gesamtwahrscheinlichkeit

Betrachte eine Untermenge  $B$   
der Grundgesamtheit  $S$ ,

Die in disjunkte Untermengen  $A_i$   
zerlegt ist, so dass gilt  $\cup_i A_i = S$ ,



$$\rightarrow B = B \cap S = B \cap (\cup_i A_i) = \cup_i (B \cap A_i),$$

$$\rightarrow P(B) = P(\cup_i (B \cap A_i)) = \sum_i P(B \cap A_i)$$

$$\rightarrow P(B) = \sum_i P(B|A_i)P(A_i) \quad \text{Gesetz der Gesamtwkt.}$$

Bayes-Theorem wird zu

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$

# Ein Beispiel zur Anwendung von Bayes-Theorem

Annahme: die Wahrscheinlichkeit (für jeden) AIDS zu haben ist:

$$P(\text{AIDS}) = 0.001$$

$$P(\text{no AIDS}) = 0.999$$

← A-Priori-Wahrscheinlichkeiten, i.e. bevor ein Test durchgeführt wurde

Betrachte einen AIDS-Test: Ergebnis ist + or -

$$P(+|\text{AIDS}) = 0.98$$

$$P(-|\text{AIDS}) = 0.02$$

← Wahrscheinlichkeiten dafür, eine infizierte Person (in-)korrekt zu identifizieren

$$P(+|\text{no AIDS}) = 0.03$$

$$P(-|\text{no AIDS}) = 0.97$$

← Wahrscheinlichkeiten dafür, eine nicht infizierte Person (in-)korrekt zu identifizieren

Nehmen Sie an, dass Ihr Ergebnis “+” ist.  
Wie beunruhigt sollten Sie sein?

# Ein Beispiel zum Bayes-Theorem (Fortsetzung)

Die Wahrscheinlichkeit AIDS zu haben, wenn der Test “+” ist, ergibt sich zu

$$\begin{aligned}P(\text{AIDS}|+) &= \frac{P(+|\text{AIDS})P(\text{AIDS})}{P(+|\text{AIDS})P(\text{AIDS}) + P(+|\text{no AIDS})P(\text{no AIDS})} \\ &= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999} \\ &= 0.032 \quad \leftarrow \text{Posterior-Wahrscheinlichkeit}\end{aligned}$$

i.e. “wahrscheinlich” sind Sie nicht erkrankt !

Ihre Sichtweise: mein Grad an Glaube, dass ich AIDS habe ist 3.2%

Die Sichtweise ihres Arztes: 3.2% von Leuten unter diesen  
Bedingungen haben AIDS

Erstaunliches Ergebnis? → Deshalb oft B-Probe bei solchen Tests

# Interpretation der Wahrscheinlichkeit

Bisher nur axiomatische Definition. Nicht hilfreich in der Anwendung.

Wir brauchen:

- was sind die Elemente der Menge?
- wie ordnen wir diesen die Wahrscheinlichkeitswerte zu?

2 Schulen:

- a) Frequentisten / Klassische Wahrscheinlichkeit
- b) Bayesianer / Subjektive Wahrscheinlichkeit

Bemerkung: Bayes-Theorem gilt in beiden und wird von beiden benutzt.  
Aber die Interpretation ist unterschiedlich.

# Frequentistische Definition/Interpretation

Grundgesamtheit  $S$  = alle mögliche Ergebnisse eines Experimentes,  
das hypothetisch wiederholbar ist

Untermengen/Ereignis  $A, B$ , = Messung liefert Ergebnis, dass im  
Wertebereich von  $A$  bzw.  $B$  liegt...

Elementarereignis: Menge mit einem Element

Wahrscheinlichkeit = Grenzfall der relativen Häufigkeit

$$P(A) = \lim_{n \rightarrow \infty} \frac{\text{Anzahl der Ergebnisse } A \text{ in } n \text{ Messungen}}{n}$$

# Frequentistische Statistik: generelle Philosophie

In frequentistischer Statistik, werden Wahrscheinlichkeiten nur Daten zugeordnet, i.e. Ergebnissen von wiederholbaren Beobachtungen

Wahrscheinlichkeit = Grenzfall der relativen Häufigkeit

Wahrscheinlichkeit für

$P$  (Higgs boson existiert),

$P(9.81 \text{ m/s}^2 < g < 9.82 \text{ m/s}^2)$

$P(\text{morgen 7.7 2012 regnet es in Freiburg})$

sind entweder 0 or 1, aber wir wissen nicht welche.

Die Werkzeuge der frequentistischen Statistik sagen uns, was wir zu erwarten haben, unter der Annahme über gewisse Wahrscheinlichkeiten, über hypothetisch wiederholbare Beobachtungen.

Die bevorzugten Theorien (Modelle, Hypothesen, ...) sind diejenigen, für die unsere Beobachtungen als “normal” betrachtet werden.

# Bayessianische Interpretation/Definition

Grundgesamtheit  $S$  = Menge der Hypothesen

Untermengen  $A, B$  = eine oder mehrere Hypothesen (exklusive, d.h. sich ausschließende Aussagen, die wahr oder falsch sind)

Wahrscheinlichkeit = Grad des Glaubens (Zuversicht), dass Hypothese  $A$  wahr ist

$P(A)$  = degree of belief that  $A$  is true

Zuordnung der Wahrscheinlichkeiten im Prinzip beliebig

Kann Wahrscheinlichkeiten beliebigen Aussagen zuordnen  
z.B.  $P(\text{morgen 7.7 .2011 regnet es in Freiburg}) = 0.95$

Umfasst frequentistische Def. "Ereignis tritt mit  $XY\%$  Häufigkeit auf".

# Bayesianische Statistik: Generelle Philosophie

In Bayesianischer Statistik: weise "subjektive Wahrscheinlichkeit"  
= "Grad des Glaubens" den verschiedenen Hypothesen zu:

Wkt. solche Daten zu beobachten unter  
der Annahme der Hypothese  $H$  (the likelihood)

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H) dH}$$

Posterior-Wkt., i.e. nach  
Auswertung der Daten

A-Priori-Wkt., i.e.  
vor der Datennahme

Normierung beinhaltet Summe  
über alle möglichen Hypothesen

Bayes-Theorem hat einen "wenn-dann"-Charakter:

**Wenn** Ihre A-Priori-Wkt  $\pi(H)$  wären, **dann** sagt es Ihnen wie diese  
Wkt. sich im Licht der Beobachtung/Daten ändern.

Keine generelle Regel für Wahl der A-Priori-Wkt! → Subjektiv!

# Bemerkung zu den zwei Schulen

Subjektivität in der Annahme der A-Priori-Wahrscheinlichkeit in der Bayesianischen Schule erscheint vielen als nicht "wissenschaftlich". Das ist falsch! Aber Abhängigkeit von A-Priori-Wahrscheinlichkeiten enthält gewisse Beliebigkeit.

Häufige Annahme: flach=gleichverteilt

Aber in welcher Größe?

z.B. Erkrankung unabhängig vom Alter oder proportional zum Alter?

In Physik (Naturwissenschaften) frequentistische Interpretation oft nützlicher, aber subjektive Wahrscheinlichkeit kann eine "natürlichere" Behandlung nicht wiederholbarer Messungen/ sich nicht wiederholender Phänomene liefern:

z.B. - systematische Fehler

- Wahrscheinlichkeit, dass ein neues Phänomen existiert, ...

# Zufallsvariablen und Wkt.dichteverteilungen

Eine Zufallsvariable (ZV) ist eine numerische Charakteristik, die einem Element der Grundgesamtheit zugeordnet wird. ZV können diskret oder kontinuierlich sein.

Annahme: das Ergebnis eines Experimentes ist kontinuierliche ZV  $x$   
Wkt. Dafür, genau den Wert  $x_i$  zu messen ist beliebig klein.

Wahrscheinlichkeit  $x$  im Intervall  $[x, x + dx]$  zu erhalten  $= f(x)dx$ .

→  $f(x)$  = Wahrscheinlichkeitsdichtefunktion (WDF)

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad x \text{ muss irgendwo sein}$$

Für diskrete Zufallsvariable mit Ergebnissen  $x_i$  mit z.B.  $i = 1, 2, \dots$

$$P(x_i) = p_i \quad \text{Wahrscheinlichkeits-Massenfunktion}$$

$$\sum_i P(x_i) = 1 \quad x \text{ muss einen der möglichen Werte annehmen}$$

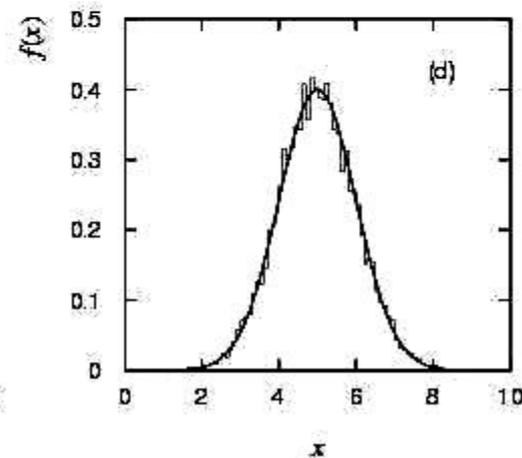
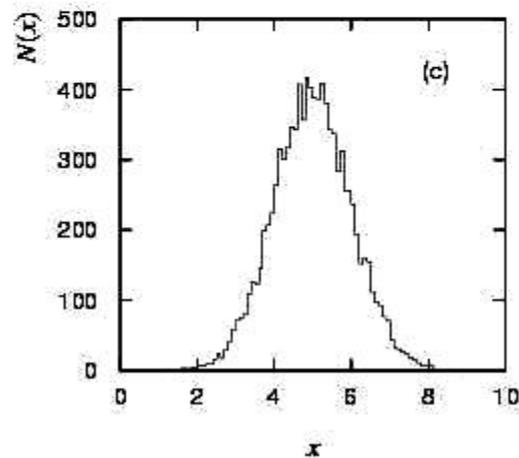
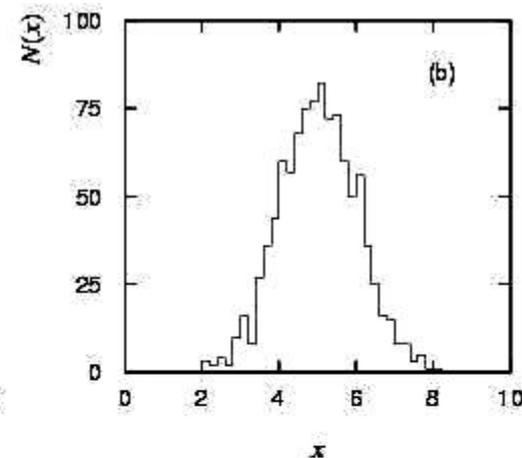
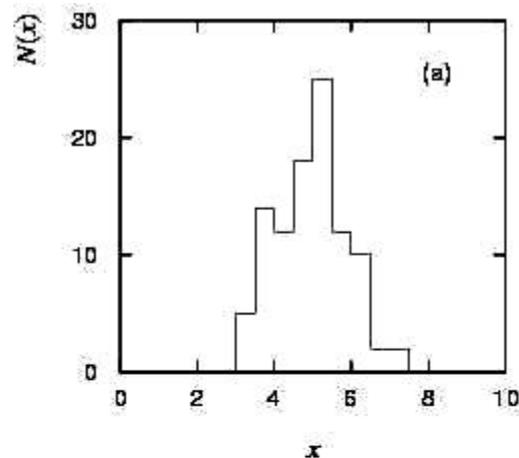
# Approximative WDF aus Histogramm

WDF = Histogramm mit  
unendlich vielen Einträgen  
(unendlich gr. Sample),  
verschwindender (0) Binbreite  
normiert auf "eins"

$$f(x) = \frac{N(x)}{n\Delta x}$$

$n$  = number of entries

$\Delta x$  = bin width

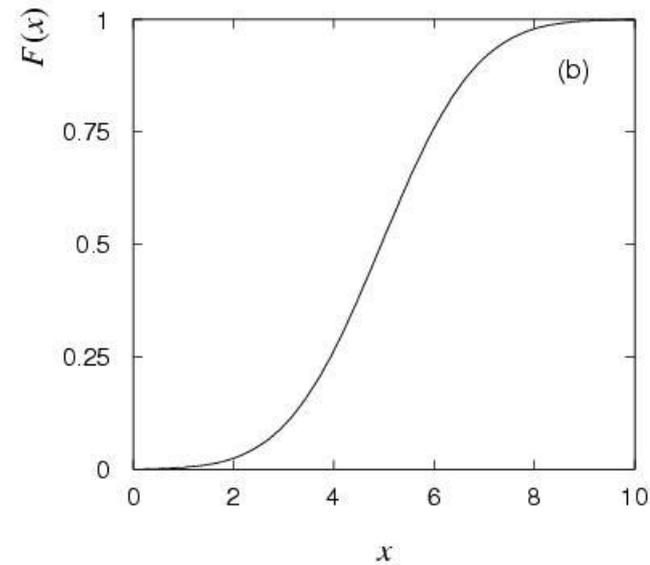
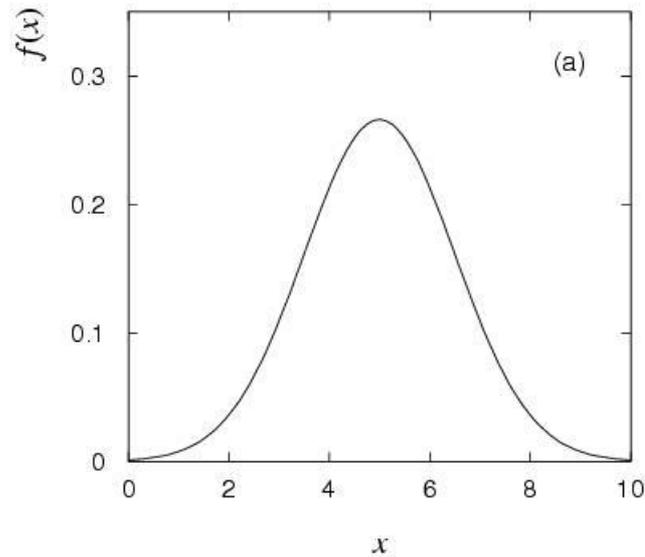


# Kumulativverteilung

Wahrscheinlichkeit ein Ergebnis kleiner oder gleich  $x$  zu erhalten

$$\int_{-\infty}^x f(x') dx' \equiv F(x)$$

Kumulativverteilung



Alternativ kann WDF über  $f(x) = \frac{\partial F(x)}{\partial x}$  definiert werden

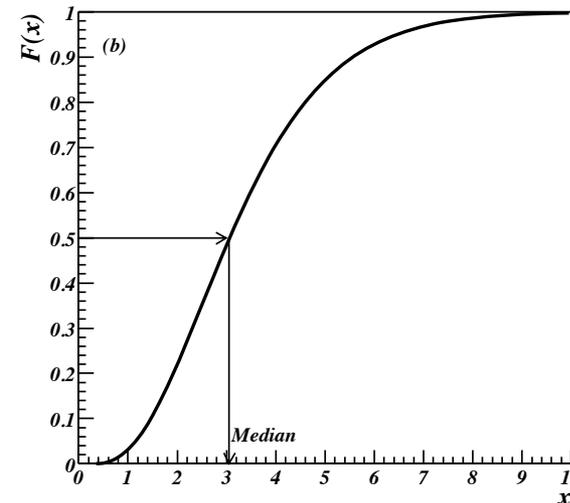
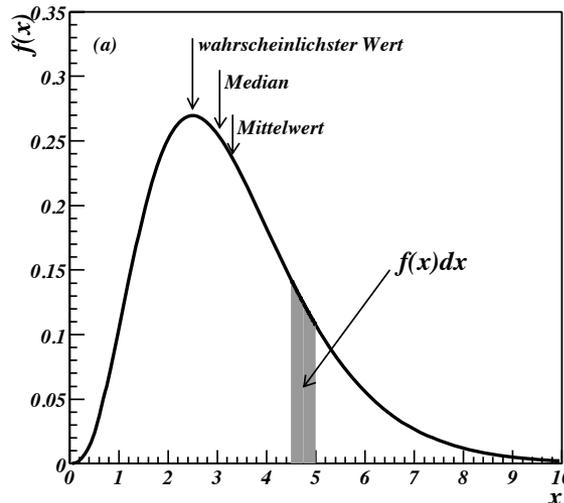
Für diskrete ZV gilt  $F(x) = \sum_{x_i \leq x} P(x_i)$

# Quantile

Das Quantil  $x_\alpha$  ist der Wert der ZV  $x$  für den die Wahrscheinlichkeit, dass die Messung einen Wert kleiner  $x_\alpha$  ergibt gerade  $\alpha$  ist.

$$F(x_\alpha) = \int_{-\infty}^{x_\alpha} f(x) dx = \alpha \Rightarrow x_\alpha = F^{-1}(\alpha)$$

Das Quantil zu  $\alpha=0.5$  ist der Median  $x_{1/2}$



# Multivariate WDF: mehrere Zufallsvariablen

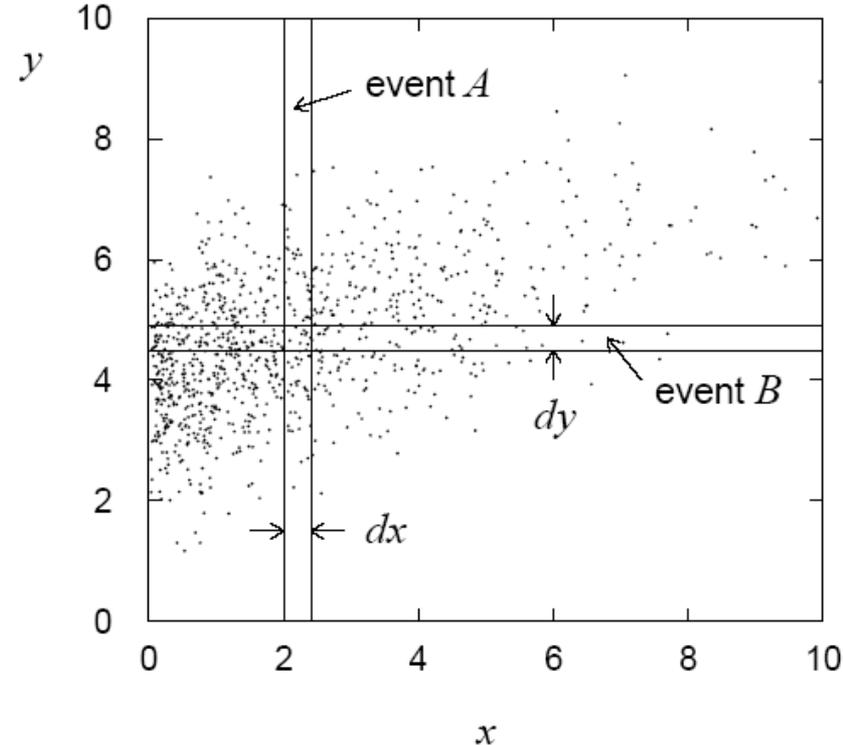
Ergebnis des Experiment wird durch mehrere Observablenwerte charakterisiert, z.B. einen  $N$ -komponentigen Vektor:  $(x_1, \dots, x_n)$

$$P(A \cap B) = \int f(x, y) dx dy$$

Gemeinsame WDF  $f(x, y)$

$f(x, y) dx dy =$  Wahrscheinlichkeit, dass  $x \in [x, x + dx]$  und  $y \in [y, y + dy]$

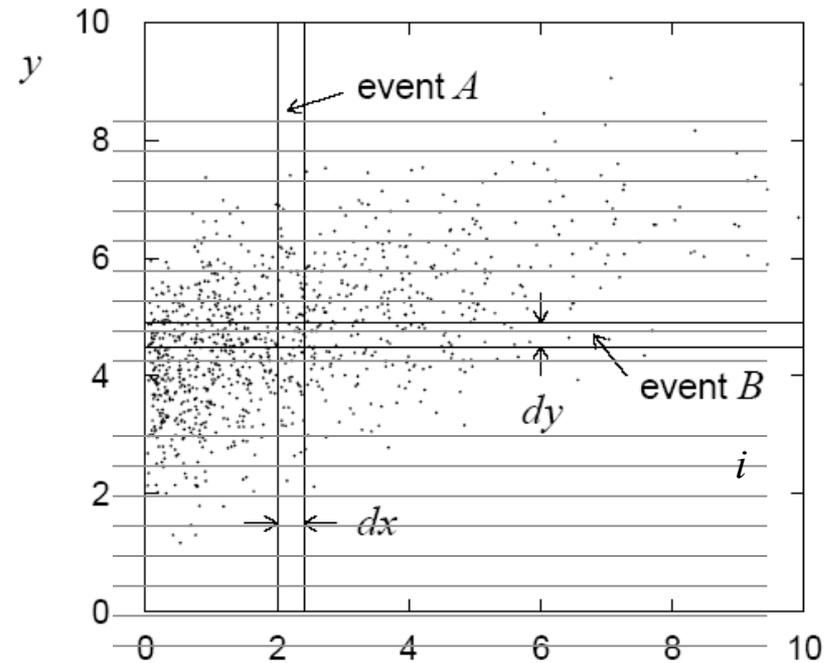
Normierung: 
$$\int \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n = 1$$



# Marginal-WDF oder Randverteilung

Menschmal interessiert uns nur WDF für einige oder eine der Zuvallsvariablen:

$$\begin{aligned} P(A) &= \sum_i P(A \cap B_i) \\ &= \sum_i \int f(x, y_i) dy dx \\ &\rightarrow \int f(x, y) dy dx \end{aligned}$$



Integration über uninteressante ZV ("Marginalisieren")

→ **Marginal-WDF oder Randverteilung**

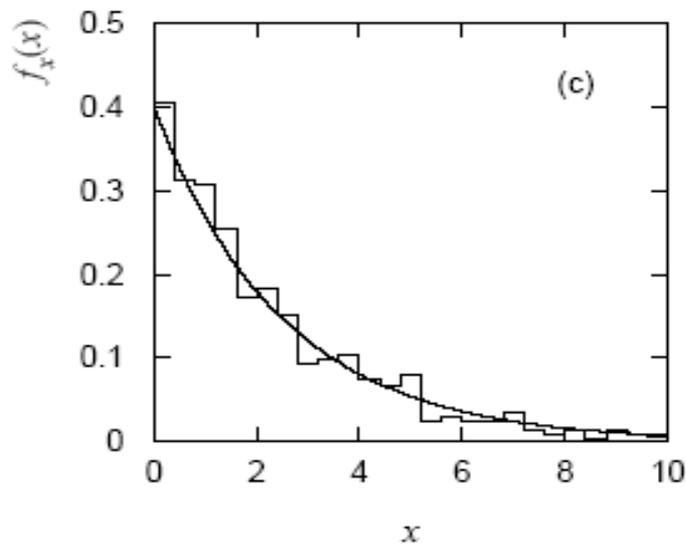
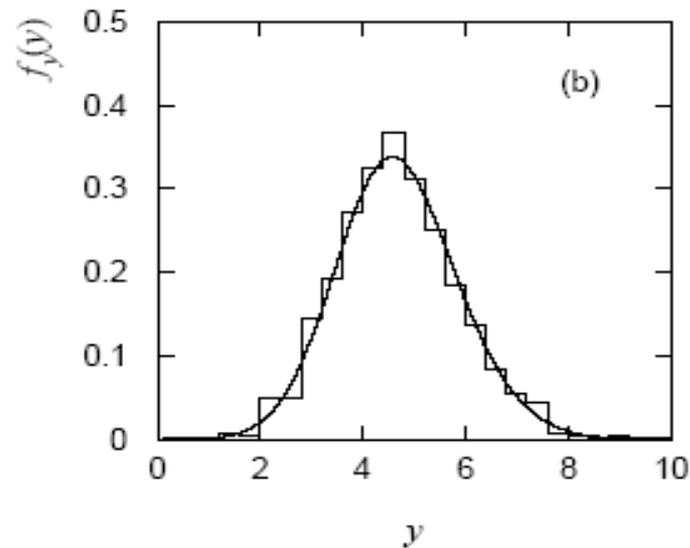
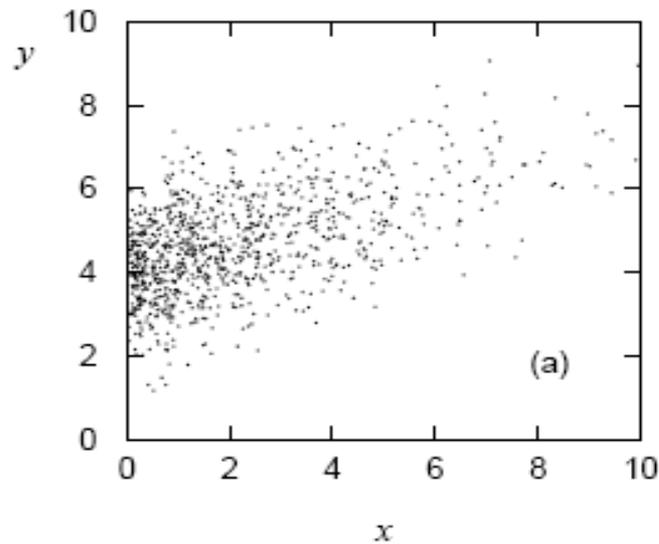
$$f_1(x_1) = \int \cdots \int f(x_1, \dots, x_n) dx_2 \dots dx_n$$

$$f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

$x_1, x_2$  sind unabhängig genau dann, wenn gilt  $f(x_1, x_2) = f_1(x_1)f_2(x_2)$

# Marginal-WDF oder Randverteilung



Marginal-WDF:  
Projektion der gemeinsamen  
WDF auf eine der ZV-Achsen

# Bedingte WDF

Manchmal soll eine der Zufallsvariablen in der gemeinsamen WDF konstant gehalten werden.

Erinnerung an bedingte Wahrscheinlichkeit:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\int f(x, y) dx dy}{\int f_x(x) dx}$$

→ bedingte WDF:  $h(y|x) = \frac{f(x, y)}{f_x(x)}$ ,  $g(x|y) = \frac{f(x, y)}{f_y(y)}$

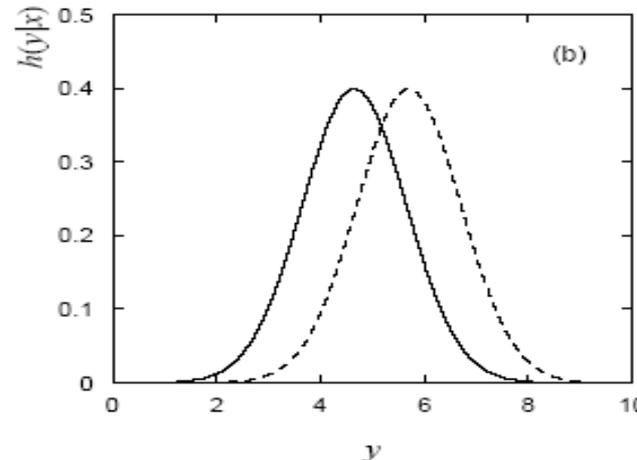
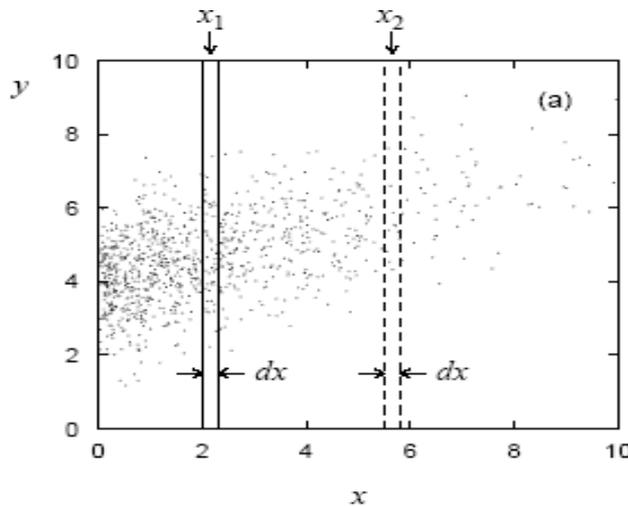
Bayes-Theorem wird zu:  $g(x|y) = \frac{h(y|x)f_x(x)}{f_y(y)}$ .

Erinnerung:  $A, B$  unabhängig, wenn  $P(A \cap B) = P(A)P(B)$ .

→  $x, y$  unabhängig wenn  $f(x, y) = f_x(x)f_y(y)$ .

# Bedingte WDF (Fortsetzung)

Beispiel: : aus gemeinsamer WDF  $f(x,y)$  werden bedingte WDFs  $h(y|x_1)$ ,  $h(y|x_2)$  bestimmt



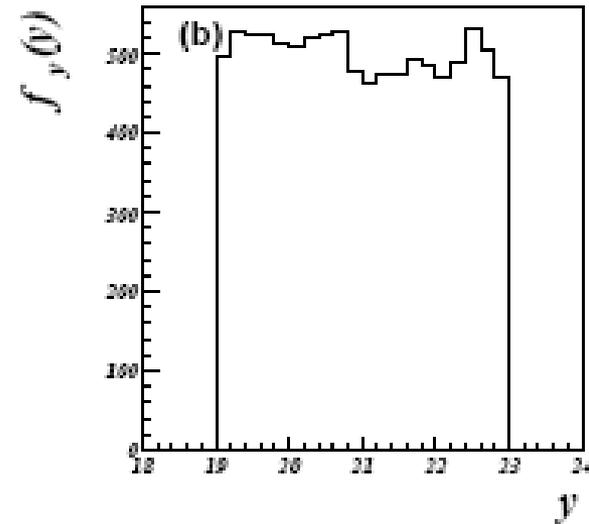
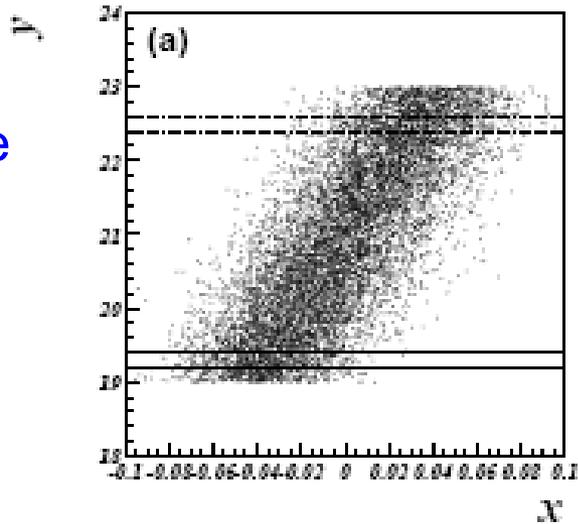
$$\int h(y|x) dy = 1 .$$

$$h(y|x) = \frac{f(x, y)}{f_x(x)} , \quad g(x|y) = \frac{f(x, y)}{f_y(y)}$$

- behandle eine der ZV als konstant in gemeinsamer WDF
- teile die gemeinsame WDF durch die Marginal-WDF der Variablen, die konstant gehalten werden
- was ist übrig bleibt ist bedingte WDF mit korrekter Normierung

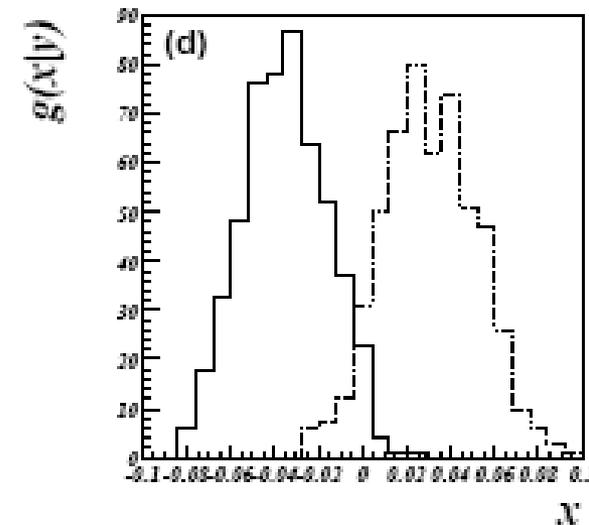
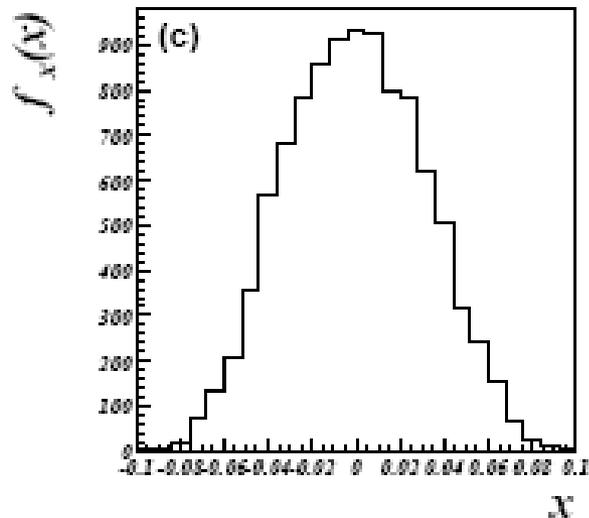
# Marginal- und Bedingte-WDF

2-dim.  
gemeinsame  
WDF



Rand-  
verteilung  
für y

Rand-  
verteilung  
für x



Bedigte  
WDFs  
für 2  
Y-Werte

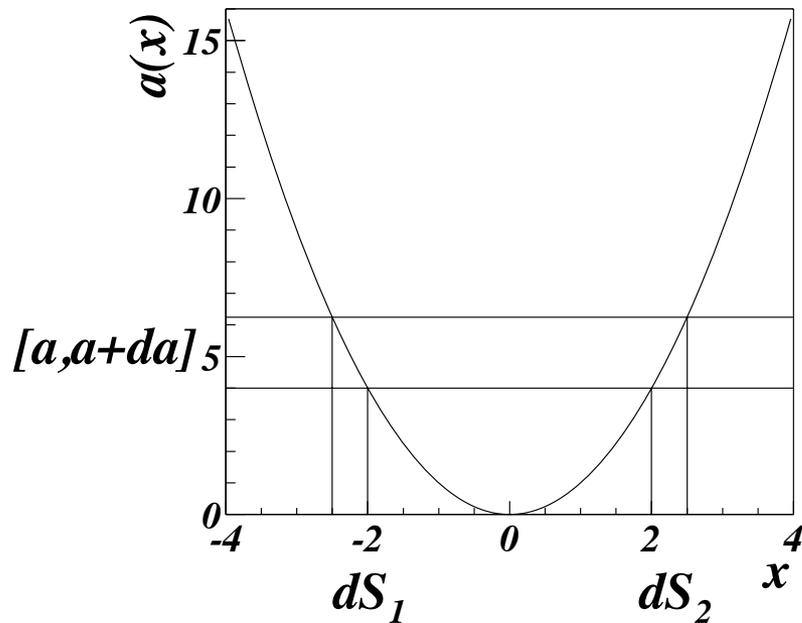
# Funktionen / Transformation von Zufallsvariablen

Ein Funktion  $a(x)$  von einer ZV  $x$  ist selbst ebenfalls eine Zufallsvariable.

Annahme: ZV  $x$  folgt einer WDF  $f(x)$

Betrachte Abhängigkeit/Funktion:  $x \rightarrow a(x)$ .

Was ist die WDF für die ZV  $g(a)$ ?

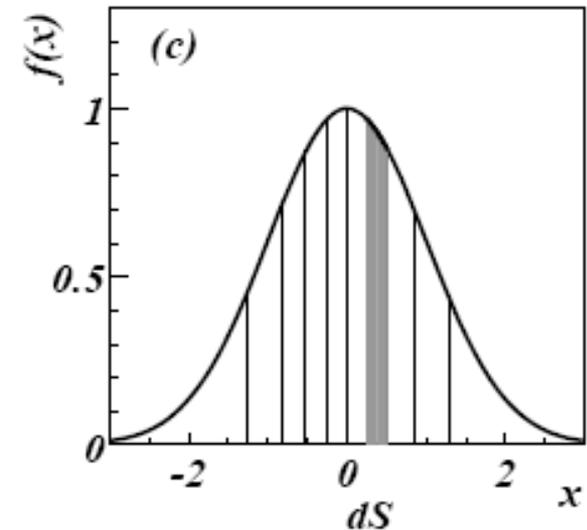
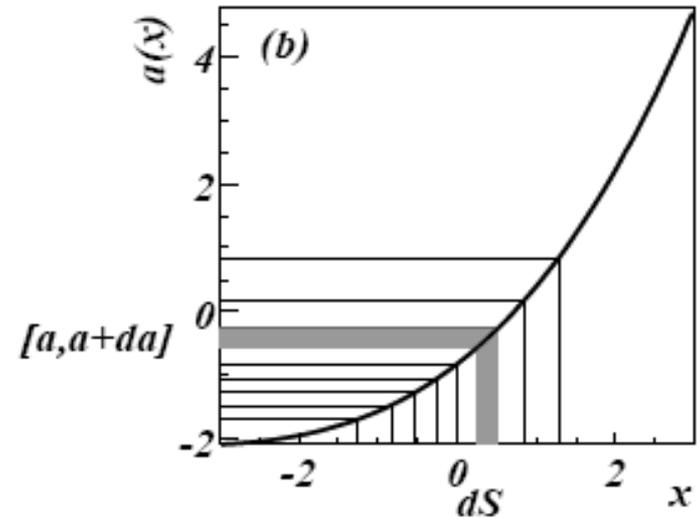
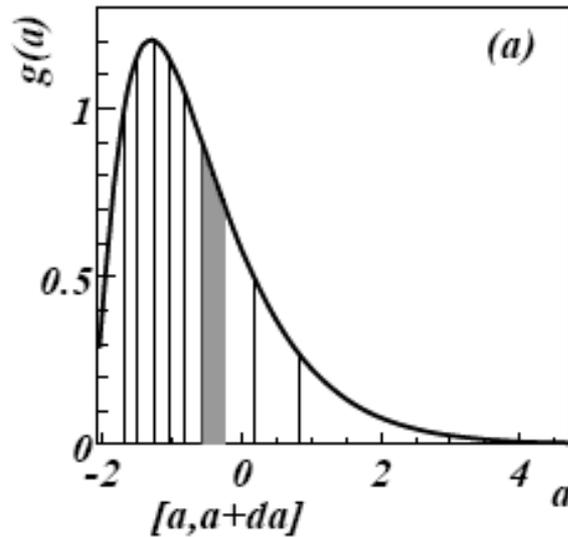


Sinnvolle Bedingung:

$$g(a) da = \int_{dS} f(x) dx$$

$dS$  = Region im Raum der  $x$ -Werte für die  $a$  in  $[a, a+da]$  liegt.

# Transformation von Zufallsvariablen



Wahrscheinlichkeit im a-Raum  $g(a)da$   
Entspricht der im x-Raum  $f(x)dS$

# Funktionen / Transformation von Zufallsvariablen

i)  $a(x)$  eindeutig umkehrbar

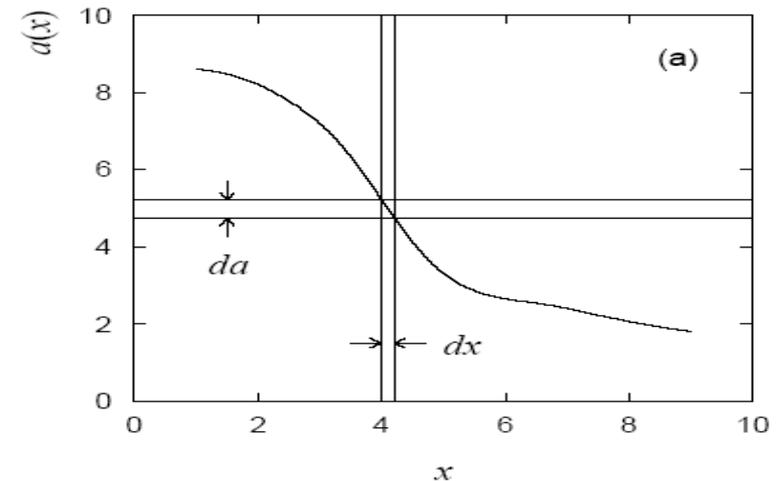
Forderung:

Wahrscheinlichkeit, dass

$a$  in  $[a, a + da]$

= Wahrscheinlichkeit,

dass  $x$  in  $[x(a), x(a + da)]$



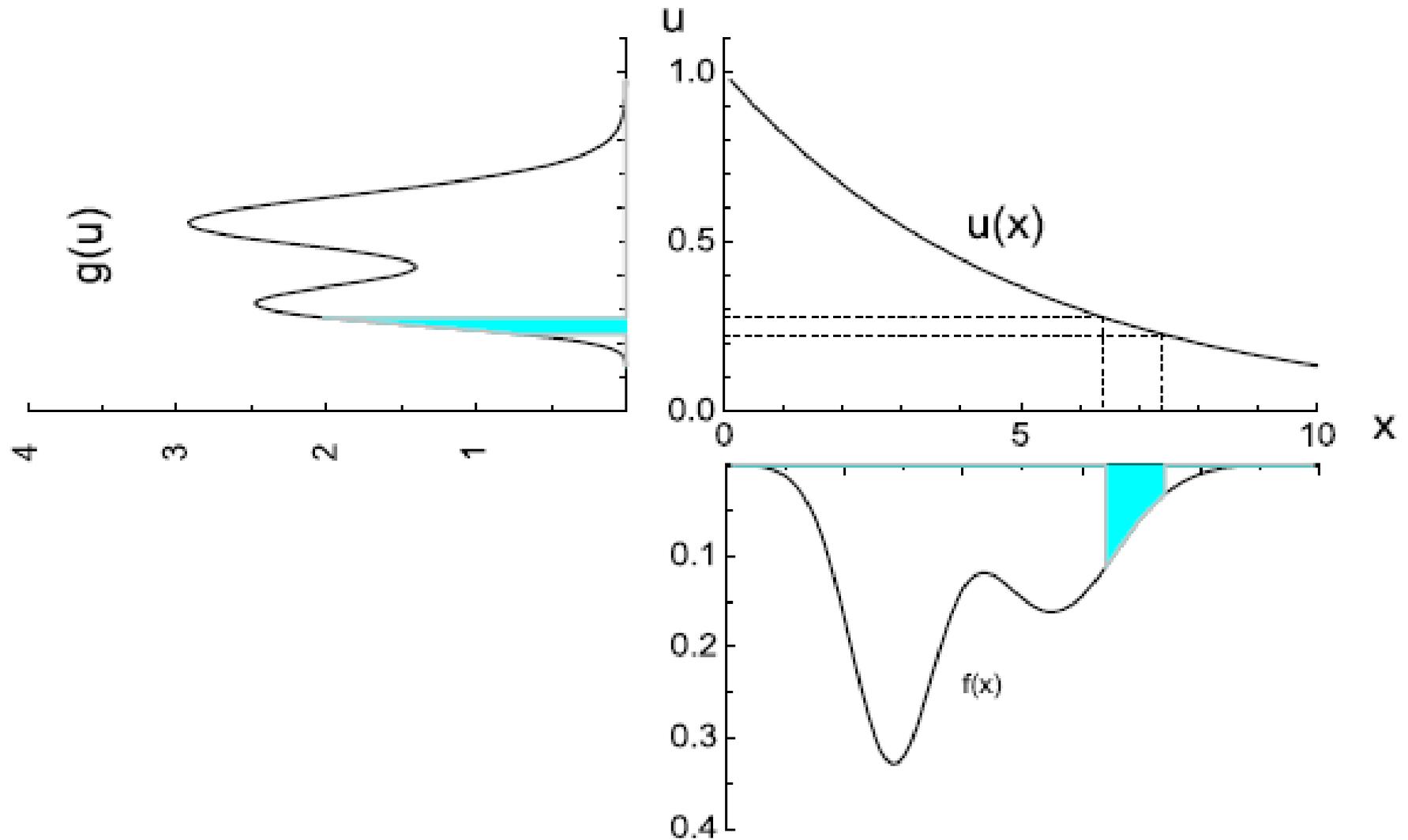
Falls gilt  $x(a) > x(a + da)$  wäre Integral  $< 0 \rightarrow$  Betragstriche

$$g(a)da = \left| \int_{x(a)}^{x(a+da)} f(x') dx' \right| = \int_{x(a)}^{x(a) + \left| \frac{dx}{da} \right| da} f(x') dx' = f(x(a)) \left| \frac{dx}{da} \right| da$$

Also gilt zusammengefasst:

$$g(a) = f(x(a)) \left| \frac{dx}{da} \right|$$

# Transformation von Zufallsvariablen



Transformation / Funktion:  $a(x) = \exp(x)$

Wktdichtefunktion für ZV  $x$ :  $f(x) = \frac{1}{2}x^2, x \text{ aus } [0, 2]$

Eindeutige Umkehrfunktion:  $x(a) = \ln(a)$

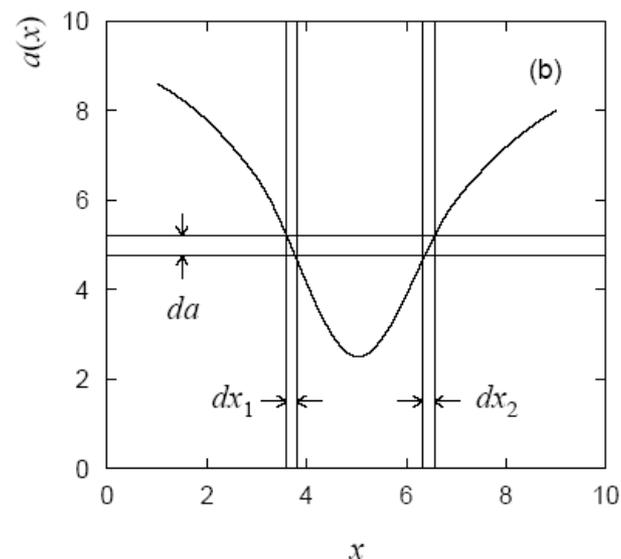
Jakobifaktor / Differential:  $\frac{dx}{da} = \frac{1}{a} > 0$

Und alles zusammengesetzt ....

$$g(a) = \frac{1}{2}(\ln(a))^2 \frac{1}{a} \quad \text{oder kompakt} \quad g(a) = \frac{\ln^2(a)}{2a}$$

# Funktionen ohne eindeutige Umkehrung

Wenn es keine eindeutige Inverse von  $a(x)$  gibt, dann summiere über alle  $dx$ -Intervalle in  $dS$ , die nach  $a$  abgebildet werden

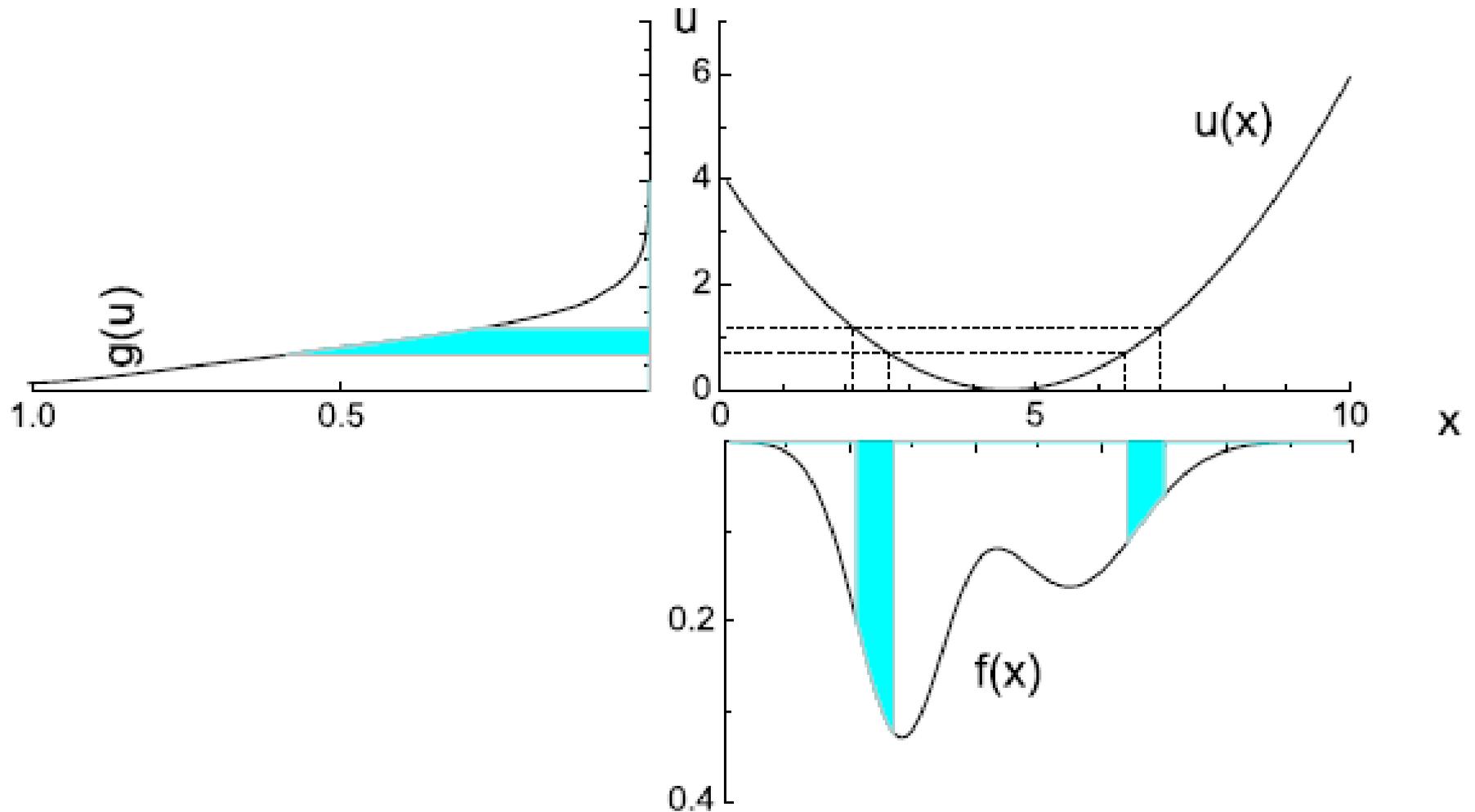


Beispiel:  $a = x^2, x = \pm\sqrt{a}, dx = \pm\frac{da}{2\sqrt{a}}$ .

$$dS = \left[ \sqrt{a}, \sqrt{a} + \frac{da}{2\sqrt{a}} \right] \cup \left[ -\sqrt{a} - \frac{da}{2\sqrt{a}}, -\sqrt{a} \right]$$

$$g(a) = \frac{f(\sqrt{a})}{2\sqrt{a}} + \frac{f(-\sqrt{a})}{2\sqrt{a}}$$

# Transformation von ZV: Mehrdeutige Inverse



# Funktionen von mehreren Zufallsvariablen

Betrachte Zufallsvariablen  $\vec{x} = (x_1, \dots, x_n)$

Und eine Funktion  $a(\vec{x})$ .

Die Bedingung der Gleichheit von Wahrscheinlichkeiten lautet:

$$g(a')da' = \int \dots \int_{dS} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Wobei  $dS$  = Region im  $x$ -Raum zwischen den Hyperflächen, die definiert werden durch

$$a(\vec{x}) = a', \quad a(\vec{x}) = a' + da'$$

# Funktionen von mehreren ZV: Fourier-Faltung

Beispiel: 2 unabhängige ZV  $x, y$  mit gemeinsamer WDF  $f_x(x) f_y(y)$

Betrachte Funktion  $a = x + y$ . Was ist  $g(a)$ ?

Anwendung:  $x$  Meßgröße,  $y$  Auflösungsvermögen des Messinstrumentes

$$\begin{aligned}g(a)da &= \int_{-\infty}^{\infty} \int_{(a-y)}^{(a-y+da)} f_x(x) f_y(y) dx dy \\ &= \int_{-\infty}^{\infty} f_x(a-y) f_y(y) da dy \\ \Rightarrow g(a) &= \int_{-\infty}^{\infty} f_x(a-y) f_y(y) dy\end{aligned}$$

(Fourier-Faltung)

# Mehr zu Variablentransformationen

Betrachte eine Vektor von ZV  $\vec{x} = (x_1, \dots, x_n)$  mit gemeinsamer WDF  $f(\vec{x})$ .

Bilde  $n$  linear unabhängige Funktionen  $\vec{y}(\vec{x}) = (y_1(\vec{x}), \dots, y_n(\vec{x}))$

Für die, die inversen Funktionen existieren  $x_1(\vec{y}), \dots, x_n(\vec{y})$

Dann ist die gemeinsame WDF für den Vektor der  $y$ -Funktionen gegeben durch

$$g(\vec{y}) = |J| f(\vec{x})$$

wobei  $J$  die

Jacobi-Determinante ist

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & & & \vdots \\ \cdots & & & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

Um z.B..  $g_1(y_1)$  zu erhalten, integriere über andere  $y$  in  $g(\vec{y})$ .

# Erwartungswerte

Betrachte kontinuierliche Zufallsvariable  $x$  with WDF  $f(x)$ .

Definiere den Erwartungswert (arithmetischen.Mittelwert) als

$$E[x] = \int x f(x) dx$$

Notation (meist):  $E[x] = \mu$

Achtung:

$E[\ ]$  ist keine Zufallsvariable, keine Funktion der Stichprobe  
ist ein linearer Operator auf der Funktion z.B.  $a(x)=x$

$$E[\alpha x + \beta] = \alpha E[x] + \beta$$

ist eine Charakteristik der Wahrscheinlichkeitsdichtefunktion  
ist abhängig von den Parametern der WDF

# Erwartungswerte

Für ein Funktion  $y(x)$  with WDF  $g(y)$ , gilt

$$E[y] = \int y g(y) dy = \int y(x) f(x) dx \quad (\text{äquivalent})$$

n-tes algebraisches Moment (n=1 ergibt Mittelwert):

$$E[x^n] = \int_{-\infty}^{\infty} x^n f(x) dx = \mu'_n$$

n-tes zentrales Moment (n>1):

$$E[(x - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx = \mu_n$$

# Erwartungswerte (II)

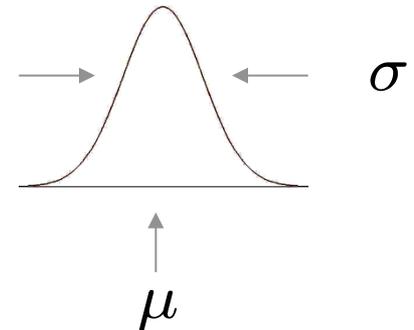
Varianz = 2tes zentrales Moment:

$$V[x] = E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \quad V[x] = \sigma^2$$

$$V[x] = E[(x - \mu)^2] = E[x^2] - 2E[x]\mu + \mu^2 = E[x^2] - \mu^2.$$

Standardabweichung:  $\sigma = \sqrt{\sigma^2}$

$$\sigma_x = \sqrt{V[x]}$$



# Erwartungswerte (III)

Weitere zentrale Momente (selten gebraucht):

$$\text{Schiefe } \gamma = \mu_3 / \sigma^3$$

misst Asymmetrie der Verteilung:  
positiv (negativ) für Ausläufer nach rechts (links)

$$\text{Wölbung oder Kurtosis } \kappa = \mu_4 / \sigma^4 - 3$$

misst Anteil im Zentrum zum Anteil in Ausläufern  
relativ zur Gauss-/Normalverteilung

0: für Gaussverteilung

positiv: für mehr Ausläufer als Gaussverteilung bei gleicher Varianz

negativ: für weniger Ausläufer als Gaussverteilung bei gleicher Varianz

# Erwartungswerte (IV)

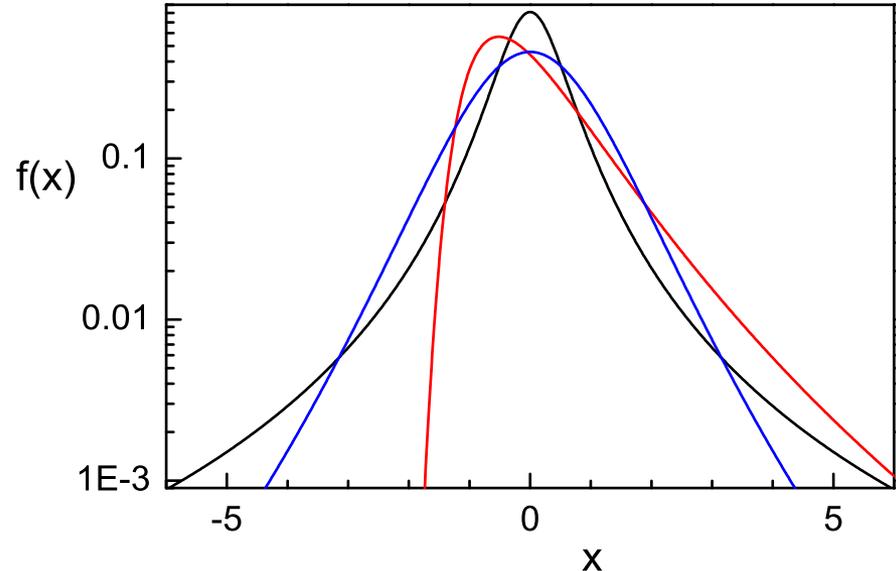
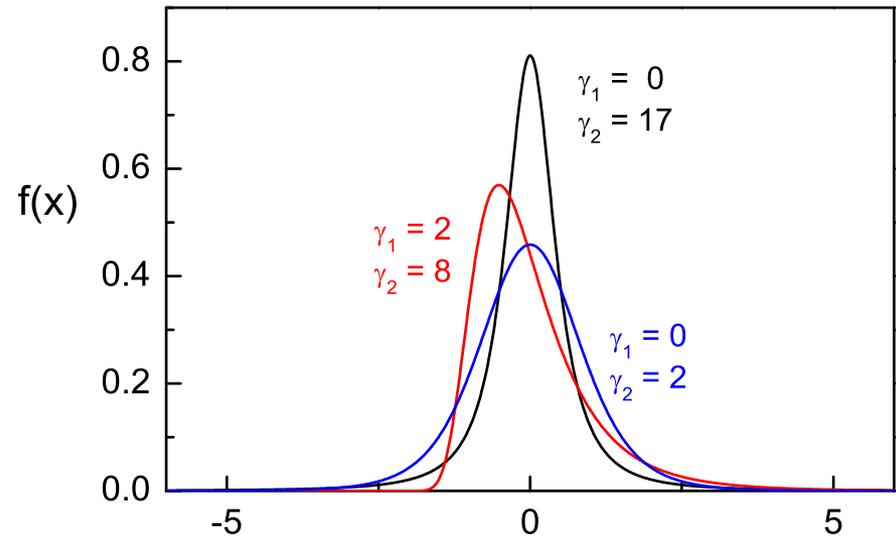
**Schiefe**  $\gamma = \mu_3/\sigma^3$

$$\begin{aligned}\gamma_1 &= E[(x - \mu)^3] / \sigma^3 \\ &= E[x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3] / \sigma^3 \\ &= \{E(x^3) - 3\mu [E(x^2) - \mu E(x)] - \mu^3\} / \sigma^3 \\ &= \frac{E(x^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3}.\end{aligned}$$

**Wölbung oder Kurtosis**  $\kappa =$

$$\mu_4/\sigma^4 - 3$$

Alle Verteilungen haben  
Mittelwert = 0 und Varianz = 1



# Kovarianz und Korrelation

Definiere Kovarianz  $\text{cov}[x,y]$  (oft auch Matrixnotation  $V_{xy}$ ) als

$$\text{COV}[x, y] = E[xy] - \mu_x \mu_y = E[(x - \mu_x)(y - \mu_y)]$$

Korrelationskoeffizient (dimensionlos) definiert als

$$\rho_{xy} = \frac{\text{COV}[x, y]}{\sigma_x \sigma_y}$$

Wenn  $x, y$ , unabhängig,  $f(x, y) = f_x(x) f_y(y)$ , dann gilt

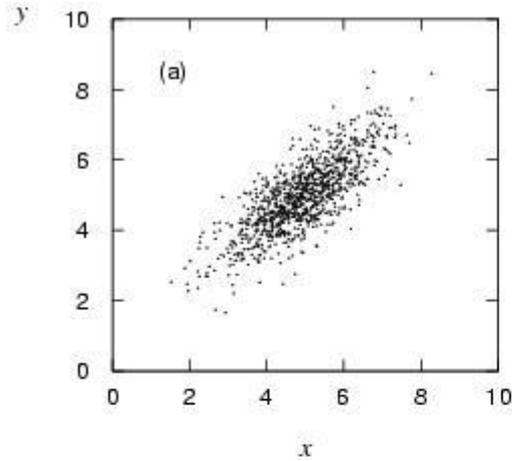
$$E[xy] = \int \int xy f(x, y) dx dy = \mu_x \mu_y$$

→  $\text{COV}[x, y] = 0$       $x$  und  $y$  sind 'unkorreliert'

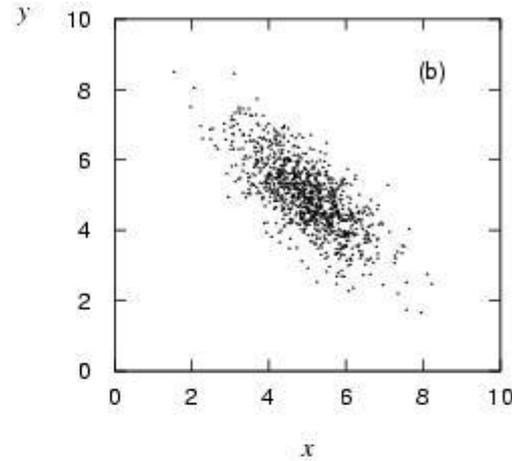
Bemerkung: Umkehrschluss i. a. nicht wahr.

# Korrelation (Fortsetzung)

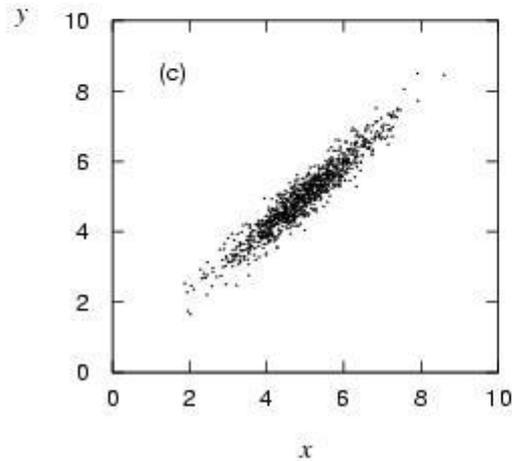
$$\rho = 0.75$$



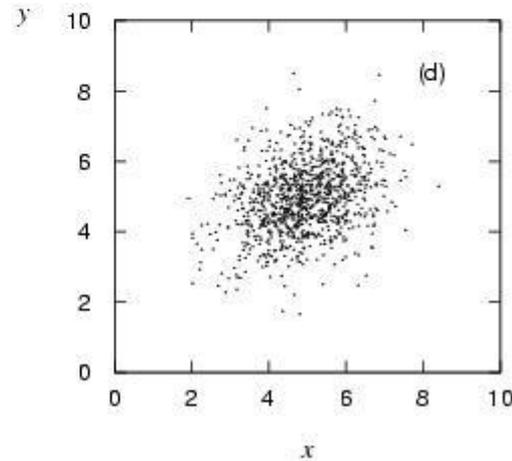
$$\rho = -0.75$$



$$\rho = 0.95$$



$$\rho = 0.25$$



# Fehlerfortpflanzung

Annahme: wir haben eine Stichprobe der ZV  $x$   $\vec{x} = (x_1, \dots, x_n)$

und wir kennen/schätzen die Kovarianzen  $V_{ij} = \text{COV}[x_i, x_j]$

die die Messfehler in the  $x_i$  quantifizieren

Nun betrachten wir ein Funktion  $y(\vec{x})$ .

Frage: was ist die Varianz von  $y(\vec{x})$  ?

- a) WDF  $f(\vec{x})$  unbekannt
- b) WDF  $f(\vec{x})$  bekannt  $\rightarrow$  "orthodoxe" Weg:  
nutze gemeinsame WDF  $f(\vec{x})$  um die WDF  $g(y)$ , zu bestimmen.  
Dann bestimme aus  $g(y)$  die Varianz  $V[y] = E[y^2] - (E[y])^2$ .

Oft nicht praktikabel,  $f(\vec{x})$  nicht vollständig bekannt.

# Fehlerfortpflanzung (II)

Annahme: wir kennen:  $\vec{\mu} = E[\vec{x}]$

In der Praxis nur Schätzwert aus der Stichprobe  $\vec{x}$

Entwickle  $y(\vec{x})$  in einer Taylor-Serie bis zum Glied erster Ordnung um  $\vec{\mu}$

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{i=1}^n \left[ \frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i)$$

Um die Varianz  $V[y]$  zu finden, brauchen wir  $E[y^2]$  and  $E[y]$ .

$$E[y(\vec{x})] \approx y(\vec{\mu}) \quad \text{weil gilt} \quad E[x_i - \mu_i] = 0$$

# Fehlerfortpflanzung (III)

$$\begin{aligned} E[y^2(\vec{x})] &\approx y^2(\vec{\mu}) + 2y(\vec{\mu}) \sum_{i=1}^n \left[ \frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} E[x_i - \mu_i] \\ &+ E \left[ \left( \sum_{i=1}^n \left[ \frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}} (x_i - \mu_i) \right) \left( \sum_{j=1}^n \left[ \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} (x_j - \mu_j) \right) \right] \\ &= y^2(\vec{\mu}) + \sum_{i,j=1}^n \left[ \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij} \end{aligned}$$

Zusammenfügen der Teilergebnisse liefert Varianz für  $y(\vec{x})$

$$\sigma_y^2 \approx \sum_{i,j=1}^n \left[ \frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

$$V_y = \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \frac{\partial y}{\partial x_3} \\ \dots \\ \frac{\partial y}{\partial x_n} \end{pmatrix}^T V_x \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \frac{\partial y}{\partial x_3} \\ \dots \\ \frac{\partial y}{\partial x_n} \end{pmatrix} = (\vec{\nabla} y)^t V_x \vec{\nabla} y$$

# Fehlerfortpflanzung (IV)

Wenn die  $x_j$  unkorreliert sind, i.e.  $V_{ij} = \sigma_i^2 \delta_{ij}$ , dann wird daraus

$$\sigma_y^2 \approx \sum_{i=1}^n \left[ \frac{\partial y}{\partial x_i} \right]_{\vec{x}=\vec{\mu}}^2 \sigma_i^2$$

Ähnlich gilt für einen Satz von Funktionen  $\vec{y}(\vec{x}) = (y_1(\vec{x}), \dots, y_m(\vec{x}))$

$$U_{kl} = \text{COV}[y_k, y_l] \approx \sum_{i,j=1}^n \left[ \frac{\partial y_k}{\partial x_i} \frac{\partial y_l}{\partial x_j} \right]_{\vec{x}=\vec{\mu}} V_{ij}$$

Oder in Matrixnotation  $U = AVA^T$ , wobei

$$A_{ij} = \left[ \frac{\partial y_i}{\partial x_j} \right]_{\vec{x}=\vec{\mu}}$$

# Fehlerfortpflanzung (V): Beispiel 1

Betrachte Stichprobe der ZV  $x_i$  vom Umfang  $n$ .

Die unkorrelierten Messungen haben Messfehler:  $\sigma^2 = \sigma_i^2$

Die Funktion  $y(x_i)$  ist der Stichprobenmittelwert:  $y(\vec{x}) = \bar{x} = \frac{1}{n} \sum_i^n x_i$

Was ist der Fehler auf den Stichprobenmittelwert?

Partielle Ableitungen:  $\frac{\partial y}{\partial x_i} = \frac{1}{n}$

$$V(y) = V(\bar{x}) = \sum_i^n \left( \frac{\partial y}{\partial x_i} \right)^2 \sigma^2 = \frac{n}{n} \sigma^2 = \frac{\sigma^2}{n}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

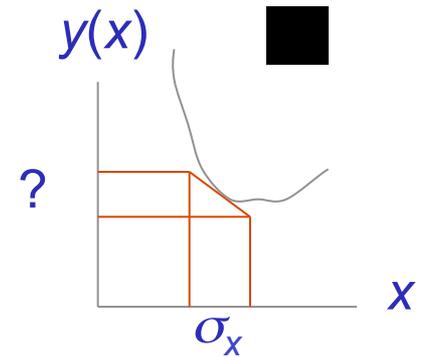
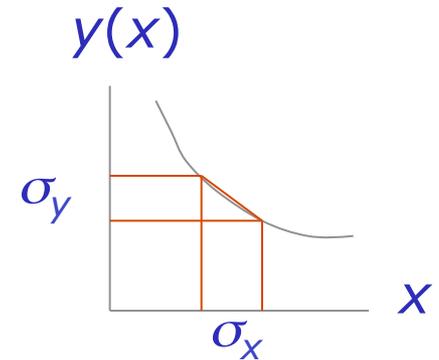
# Fehlerfortpflanzung (VI)

Die 'Fehlerfortpflanzungsformel' sagt uns, wie wir die Kovarianzen eines Satzes von Funktionen  $\vec{y}(\vec{x}) = (y_1(\vec{x}), \dots, y_m(\vec{x}))$  durch die Kovarianzen der ursprünglichen Variablen approximativ auswerten können.

Limitierung: exakt nur wenn  $\vec{y}(\vec{x})$  linear.

Näherung bricht zusammen, wenn die Funktion nicht linear ist in einem Bereich der vergleichbar ist mit der Größe von  $\sigma_x$ .

Bemerkung: keine Annahme über die WDF von  $x_j$ ,  
i.e., sie muss nicht die Gaußverteilung sein



# Fehlerfortpflanzung (VII) – Spezialfall

$$y = x_1 + x_2 \quad \rightarrow \quad \sigma_y^2 = \sigma_1^2 + \sigma_2^2 + 2\text{COV}[x_1, x_2]$$

$$y = x_1 x_2 \quad \rightarrow \quad \frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + 2 \frac{\text{COV}[x_1, x_2]}{x_1 x_2}$$

wenn die  $x_i$  unkorreliert sind, gilt  
addiere Fehler quadratisch für die Summe (oder Differenz),  
addiere relative Fehler quadratisch für Produkt (oder Quotient).



Aber Korrelationen können dies komplett ändern...

# Fehlerfortpflanzung (VIII) – Spezialfall (2)

Betrachte  $y = x_1 - x_2$  mit

$$\mu_1 = \mu_2 = 10, \quad \sigma_1 = \sigma_2 = 1, \quad \rho = \frac{\text{COV}[x_1, x_2]}{\sigma_1 \sigma_2} = 0.$$

$$V[y] = 1^2 + 1^2 = 2, \rightarrow \sigma_y = 1.4$$

Nur nehmen wir an, dass  $\rho = 1$ . Dann gilt

$$V[y] = 1^2 + 1^2 - 2 = 0, \rightarrow \sigma_y = 0$$

i.e. für 100% Korrelation verschwindet der Fehler in der Differenz.

Analog: für die Summe werden die Fehler linear addiert

# Fehlerfortpflanzung (IX): Beispiel 2

Betrachte Messung einer Spur/Trajektorie in Zylinderkoordinaten  $r, \phi, z$

Die Messfehler seien unkorreliert und es gelte  $\sigma_r = 0, \sigma_\phi, \sigma_z$

Die Kovarianzmatrix lautet dann:

$$V_{r,\phi,z} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \sigma_\phi^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

Wie sehen die Fehler in kartesischen Koordinaten aus?

$$x = r \cos \phi$$

$$y = r \sin \phi$$

$$z = z$$

# Fehlerfortpflanzung (X): Beispiel 2

Die Jakobi-Matrix der Ableitungen lautet:

$$A = \begin{pmatrix} \cos \phi & -r \sin \phi & 0 \\ \sin \phi & r \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Die Kovarianzmatrix in x,y, z lautet:

$$V_{x,y,z} = AV_{r,\phi,z}A^T = \begin{pmatrix} \sigma_\phi^2 y^2 & -\sigma_\phi^2 xy & 0 \\ -\sigma_\phi^2 xy & \sigma_\phi^2 x^2 & 0 \\ 0 & 0 & \sigma_z^2 \end{pmatrix}$$

Bem.: - Fehler auf z unverändert

- Fehler in x,y  $\sigma_x, \sigma_y$   
abhängig von x,y  
korreliert

Die Korrelation ergibt sich zu:  $\rho_{xy} = -\frac{\sigma_\phi^2 x y}{\sigma_\phi x \sigma_\phi y} = -1$