

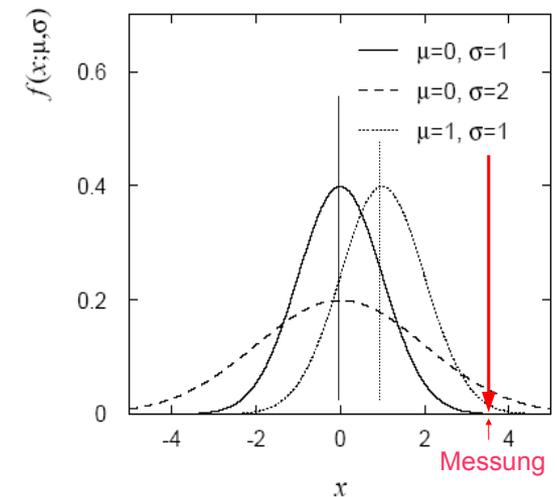
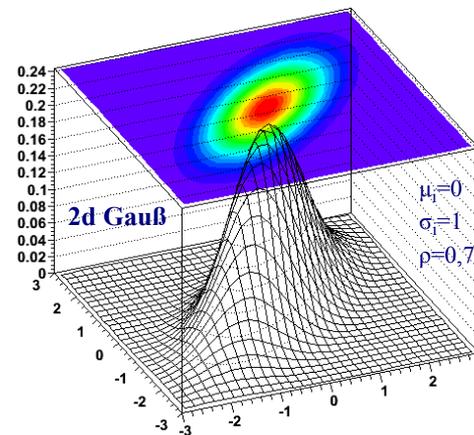
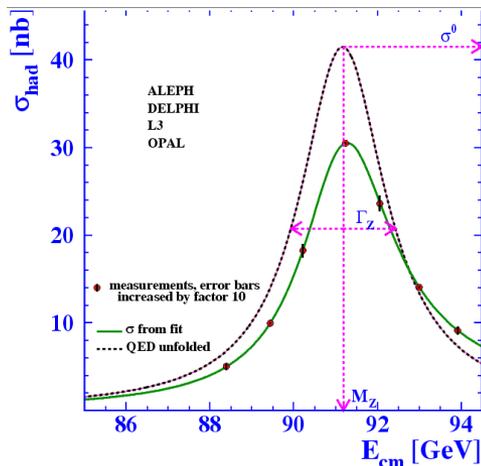
Moderne Methoden der Datenanalyse

Einführung (2)

Andreas B. Meyer  

2. Mai 2017

Fakultät für Physik
Institut für Experimentelle Kernphysik - IEKP



Vorlesungsprogramm

- Einführung: Überblick und grundlegende Konzepte (1)
- Heute: Einführung: Überblick und grundlegende Konzepte (2)
- Zufallszahlen und Monte-Carlo Methoden

A.MEYER

- Hypothesentests (1)
- Parameterschätzung
- Parameterschätzung (Goodness-of-fit)
- Optimierungs- und Parametrisierungsmethoden

R.ULRICH

- Konfidenzintervalle
- Hypothesentests (2)
- Klassifikation
- Klassifikation
- Klassifikation
- Messen und Entfalten
- Systematische Unsicherheiten

A.MEYER

1. Einführung (2) - Übersicht

- Zufallszahlen
- Zentraler Grenzwertsatz
- Experiment und Theorie
- Statistische und systematische Fehler

- Wahrscheinlichkeiten
- Erwartungswert und Varianz
- Grundlegende Verteilungen
- Korrelationen
- Fehlerfortpflanzung

- Kleinste Quadrate
- Maximum-Likelihood

- Hypothesentests
- Konfidenz-Intervalle, Signifikanz und Signalstärke
- Entfaltung und systematische Fehler

1.1
MOTIVATION

1.2
WAHRSCHEIN-
LICHKEITSVER-
TEILUNGEN

1.3
PARAMETER-
SCHÄTZUNG

1.4
AUSBLICK

1.2

WAHRSCHEINLICHKEITS- VERTEILUNGEN

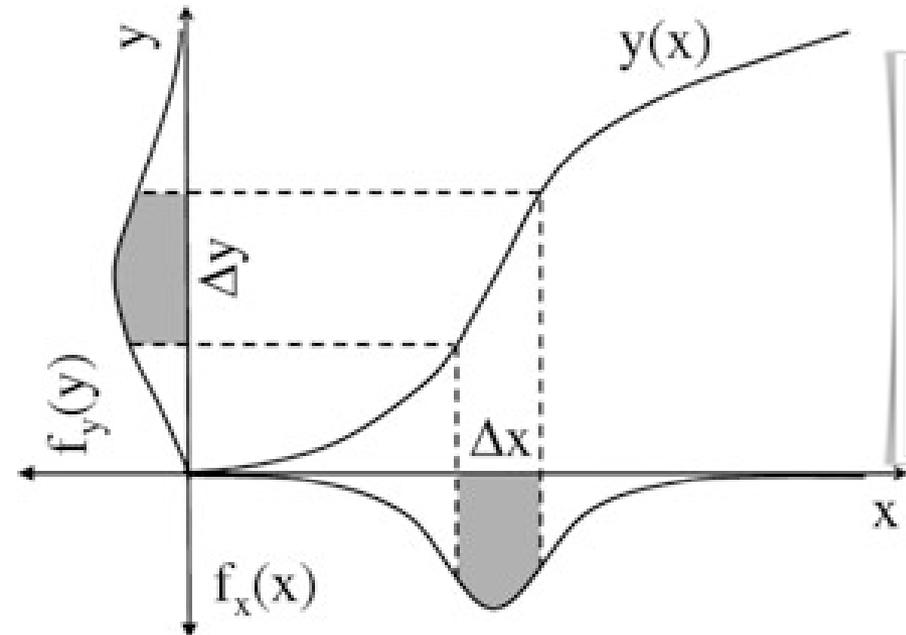
Fortsetzung

■ Betrachte $y = y(x)$, wenn x eine Varianz σ_x^2 hat

■ $f_x(x)$ ist die PDF für Zufallsvariable x ,
 $f_y(y)$ ist die PDF für Zufallsvariable y

■ Wahrscheinlichkeitserhaltung =
Gleichheit der Verteilungsfunktionen,

Flächen unter f_y und f_x in den Intervallen Δx und Δy sind gleich groß, d.h.: $f_y(y)dy = f_x(x)dx$



Erdmann/Hebbeker S.66

$$f_y(y) = f_x(x) \left| \frac{dx}{dy} \right|$$

siehe auch Blobel/Lohrmann S.101 ff

Transformation von Wahrscheinlichkeitsdichten

- Betrachte $y = y(x)$, wenn x eine Varianz σ_x^2 hat
- Mittelwert und Varianz von y durch Taylor-Entwicklung:

$$y(x) = y(\langle x \rangle) + (x - \langle x \rangle) \left. \frac{dy}{dx} \right|_{\langle x \rangle} + \dots$$

$$E[y] = \underbrace{E[y(\langle x \rangle)]}_{y(\langle x \rangle)} + \underbrace{E[x - \langle x \rangle]}_{=0} \left. \frac{dy}{dx} \right|_{\langle x \rangle} + \dots$$

lineare Näherung,
innerhalb der
Variation von x sei
 $y(x)$ linear

$$\langle y \rangle \approx y(\langle x \rangle)$$

$$V[y] = E[(y - \langle y \rangle)^2] \approx E[(x - \langle x \rangle)^2] \left(\left. \frac{dy}{dx} \right|_{\langle x \rangle} \right)^2$$

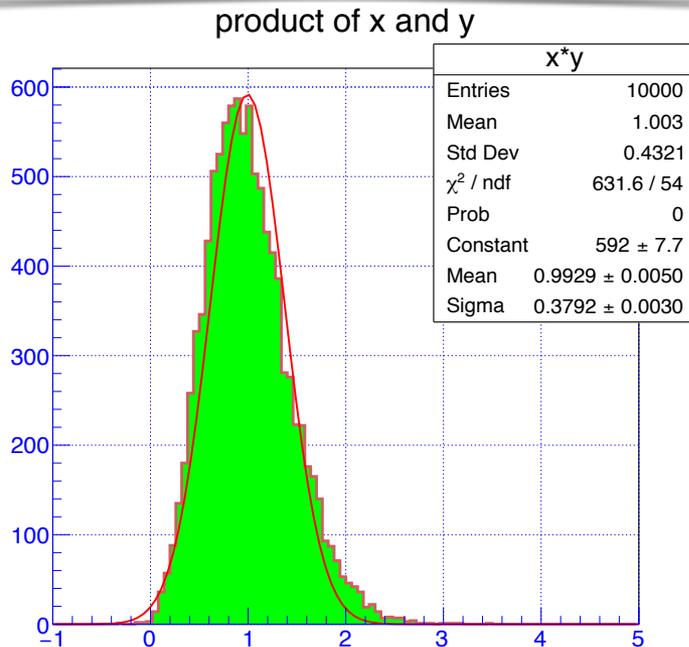
$$= V[x] \cdot \left(\left. \frac{dy}{dx} \right|_{\langle x \rangle} \right)^2$$

$$\sigma_y = \sqrt{\left(\left. \frac{dy}{dx} \right|_{\langle x \rangle} \right)^2} \cdot \sigma_x$$

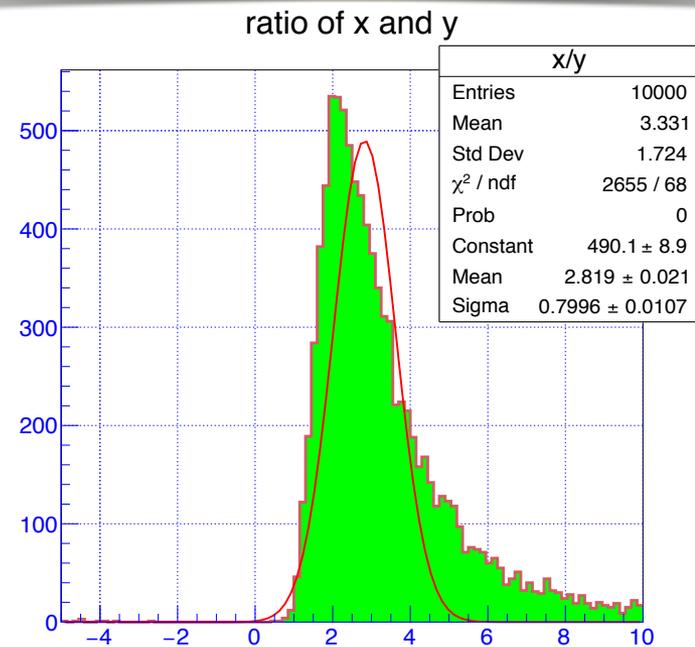
Messfehler und Fehlerfortpflanzung

Beispiel: x und y seien Gauß-verteilt

$$f(x,y)=x*y \text{ mit } \mu_x=\mu_y=1 \text{ und } \sigma_x=\sigma_y=0.3$$



$$f(x,y)=x/y \text{ mit } \mu_x=6, \mu_y=2 \text{ und } \sigma_x=\sigma_y=1.0$$



Fehlerfortpflz. liefert: $E = 1$, $\sigma \sim 0.42$

$E = 3$, $\sigma \sim 2.1$

- Beachte: allgemein ist $E[y(x)] \neq y(E[x])$
- Überprüfung der statistischen Verfahren z.B. durch Simulation: "Pseudo-Experimente" oder "Toy Monte Carlo"

Mehrdimensionale Verteilungen

■ Wahrscheinlichkeitsverteilungen können von mehr als einer Variablen abhängig sein.

■ Mittelwerte (Erwartungswerte):

$$\langle x \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) dx dy$$

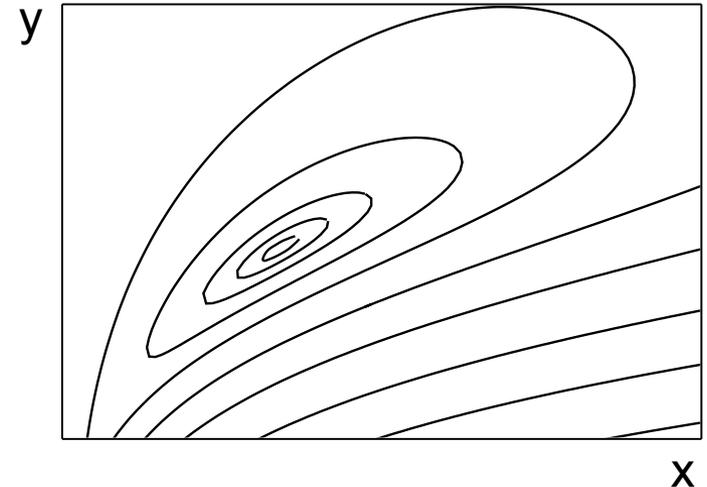
■ Varianz:

$$V[x] = \sigma_x^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle x \rangle)^2 \cdot f(x, y) dx dy \quad \text{und entsprechend für } y$$

■ Kovarianzen σ_{xy} beschreiben die Korrelation zwischen x und y :

$$\sigma_{xy} = \sigma_{yx} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \langle x \rangle) \cdot (y - \langle y \rangle) \cdot f(x, y) dx dy$$

Linien: Punkte gleicher Wahrscheinlichkeit



Korrelation in zwei Dimensionen

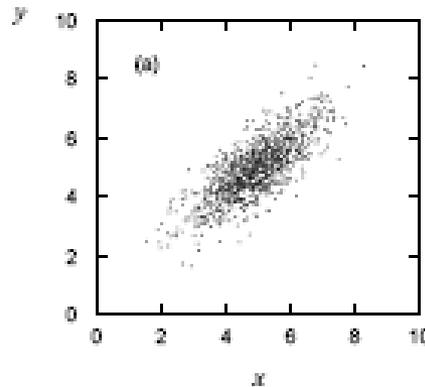
■ Kovarianzmatrix V

$$V = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{pmatrix}$$

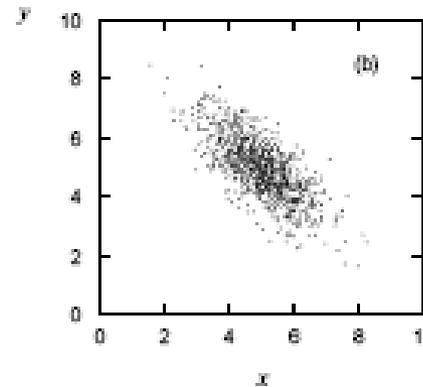
■ Korrelationsmatrix: Transformiere V so, dass Diagonalelemente alle 1. Dann sind ρ_{xy} die Korrelationskoeffizienten.

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad -1 \leq \rho_{xy} \leq 1$$

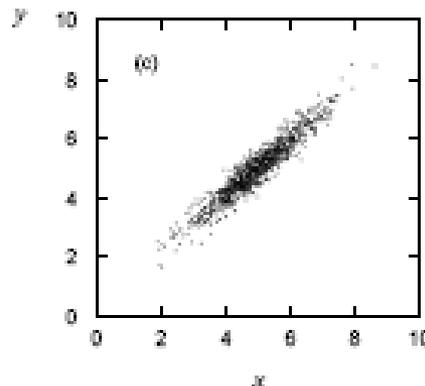
$$\rho = 0.75$$



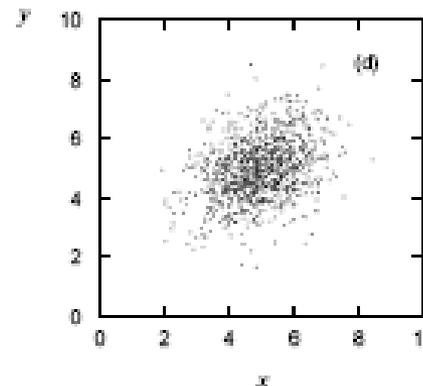
$$\rho = -0.75$$



$$\rho = 0.95$$



$$\rho = 0.25$$



Beispiel: Gauß-Verteilung in n Dimensionen

■ Wahrscheinlichkeitsdichte der Gauß-Verteilung:

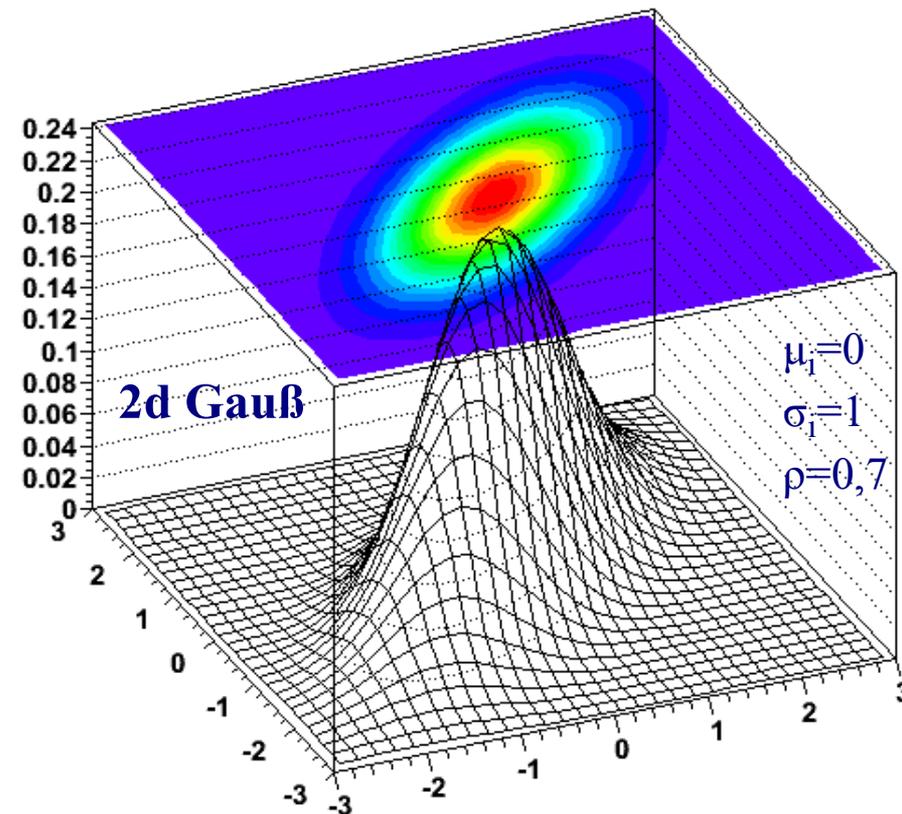
$$P(\vec{x}, \vec{\mu}, V) = \frac{1}{(2\pi)^{n/2} \sqrt{|V|}} \exp \left(-\frac{1}{2} (\vec{x} - \vec{\mu})^T V^{-1} (\vec{x} - \vec{\mu}) \right)$$

■ In zwei Dimensionen:

$$V = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$|V| = \det(V) = \sigma_1^2 \sigma_2^2 - (\rho \sigma_1 \sigma_2)^2$$

$$V^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \times \begin{pmatrix} \sigma_2^2 & -\rho \sigma_1 \sigma_2 \\ -\rho \sigma_1 \sigma_2 & \sigma_1^2 \end{pmatrix}$$



Fehlerfortpflanzung mit mehreren Variablen

- Die lineare Transformation für $\mathbf{x}=(x_1, \dots, x_n)$ und $\mathbf{y}=(y_1, \dots, y_m)$ ist gegeben durch:

$$\vec{y} = B \cdot \vec{x} \quad \text{mit} \quad B = \begin{pmatrix} \partial y_1 / \partial x_1 & \dots & \partial y_1 / \partial x_n \\ \partial y_2 / \partial x_1 & \dots & \partial y_2 / \partial x_n \\ \dots & \dots & \dots \\ \partial y_m / \partial x_1 & \dots & \partial y_m / \partial x_n \end{pmatrix}$$

- Erwartungswert:

$$E[\vec{y}] = B \cdot E[\vec{x}]$$

- Allgemeine Messung: multivariate Verteilungsdichte $f_x(x_1, x_2, \dots, x_n)$ wird transformiert in $f_y(y_1, y_2, \dots, y_m)$:

$$f_x(\vec{x}) dx_1 dx_2 \dots dx_n = f_y(\vec{y}) dy_1 dy_2 \dots dy_m$$

- Für $m=n$ gilt:

$$f_y(\vec{y}) = f_x(\vec{x}) |J| \quad |J| = \det(B)$$

Jacobi-Determinante

Fehlerfortpflanzung mit mehreren Variablen

- Die lineare Transformation für $\mathbf{x}=(x_1, \dots, x_n)$ und $\mathbf{y}=(y_1, \dots, y_m)$ ist gegeben durch:

$$\vec{y} = B \cdot \vec{x} \quad \text{mit} \quad B = \begin{pmatrix} \partial y_1 / \partial x_1 & \dots & \partial y_1 / \partial x_n \\ \partial y_2 / \partial x_1 & \dots & \partial y_2 / \partial x_n \\ \dots & \dots & \dots \\ \partial y_m / \partial x_1 & \dots & \partial y_m / \partial x_n \end{pmatrix}$$

- Erwartungswert:

$$E[\vec{y}] = B \cdot E[\vec{x}]$$

- Für $m \neq n$: Kovarianz-Matrix i.a. durch Näherung:

$$V[\vec{y}] = B \cdot E [(\vec{x} - \langle \vec{x} \rangle)^2] \cdot B^T \\ B \cdot V[\vec{x}] \cdot B^T$$

“Allgemeines Fehlerfortpflanzungsgesetz”

Fehlerfortpflanzung mit mehreren Variablen

- Messung besteht häufig aus Kombination mehrerer Größen in eine einzelne $y=y(x_1, \dots, x_n)$. B ist dann eine $1 \times n$ Matrix
- Beispiel für $m=1$ und $n=2$:

$$\begin{aligned} V[y] &= B \cdot V[\vec{x}] \cdot B^T \\ &= \begin{pmatrix} \frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} \end{pmatrix} \cdot \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \end{pmatrix} \\ &= \left(\frac{\partial y}{\partial x_1} \right)^2 \sigma_1^2 + 2 \frac{\partial y}{\partial x_1} \frac{\partial y}{\partial x_2} \sigma_{12} + \left(\frac{\partial y}{\partial x_2} \right)^2 \sigma_2^2 \end{aligned}$$

- Falls die $x_1 \dots x_n$ unkorreliert, dann sind die $\sigma_{i \neq j} = 0$, d.h. V ist diagonal

$$V[y_j] = \sigma_{y_j}^2 = \sum_{i=1}^n \left(\frac{\partial y_j}{\partial x_i} \right)^2 \sigma_{x_i}^2$$

Fehlerfortpflanzungsgesetz für unkorrelierte Variablen x_i

Konstruktion einer Kovarianzmatrix

- Beispiel Praktikum: 3 Gruppen a 2 Personen, 6 Messungen
- Unsicherheiten g_i der $i=1\dots 6$ Messungen setzen sich zusammen aus:
 - s_i systematische Fehler der Messgeräte: korreliert innerhalb der Gruppe, unkorreliert zwischen Gruppen
 - f_i statistische Fehler der Messung: unkorreliert zwischen allen i
 - t_i Theoriefehler: korreliert für alle Messungen
 - g_i Gesamtfehler: $g_i^2 = s_i^2 + f_i^2 + t_i^2$

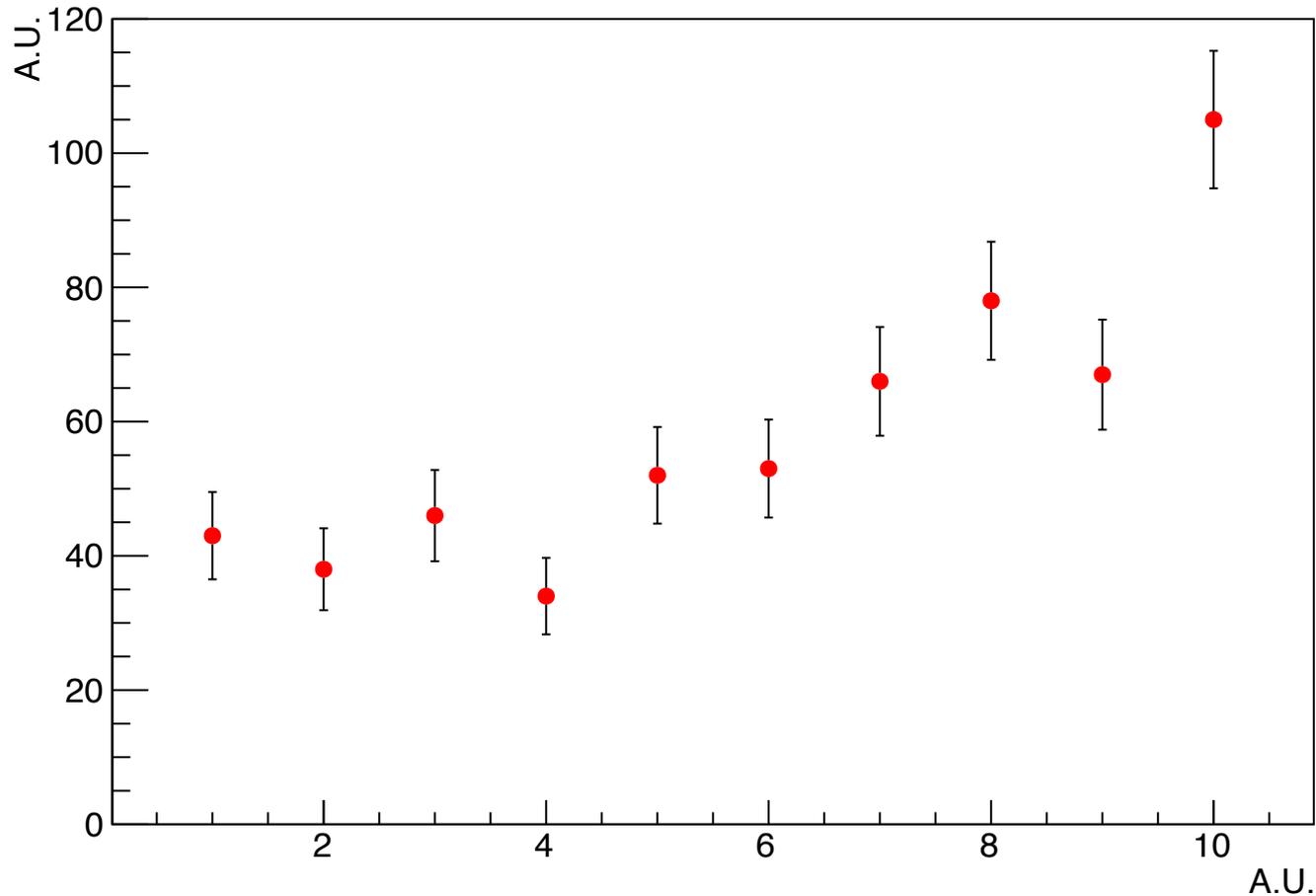
$$V = \begin{pmatrix} g_1^2 & s^2 + t^2 & t^2 & t^2 & t^2 & t^2 \\ s^2 + t^2 & g_2^2 & t^2 & t^2 & t^2 & t^2 \\ t^2 & t^2 & g_3^2 & s^2 + t^2 & t^2 & t^2 \\ t^2 & t^2 & s^2 + t^2 & g_4^2 & t^2 & t^2 \\ t^2 & t^2 & t^2 & t^2 & g_5^2 & s^2 + t^2 \\ t^2 & t^2 & t^2 & t^2 & s^2 + t^2 & g_6^2 \end{pmatrix}$$

1.3

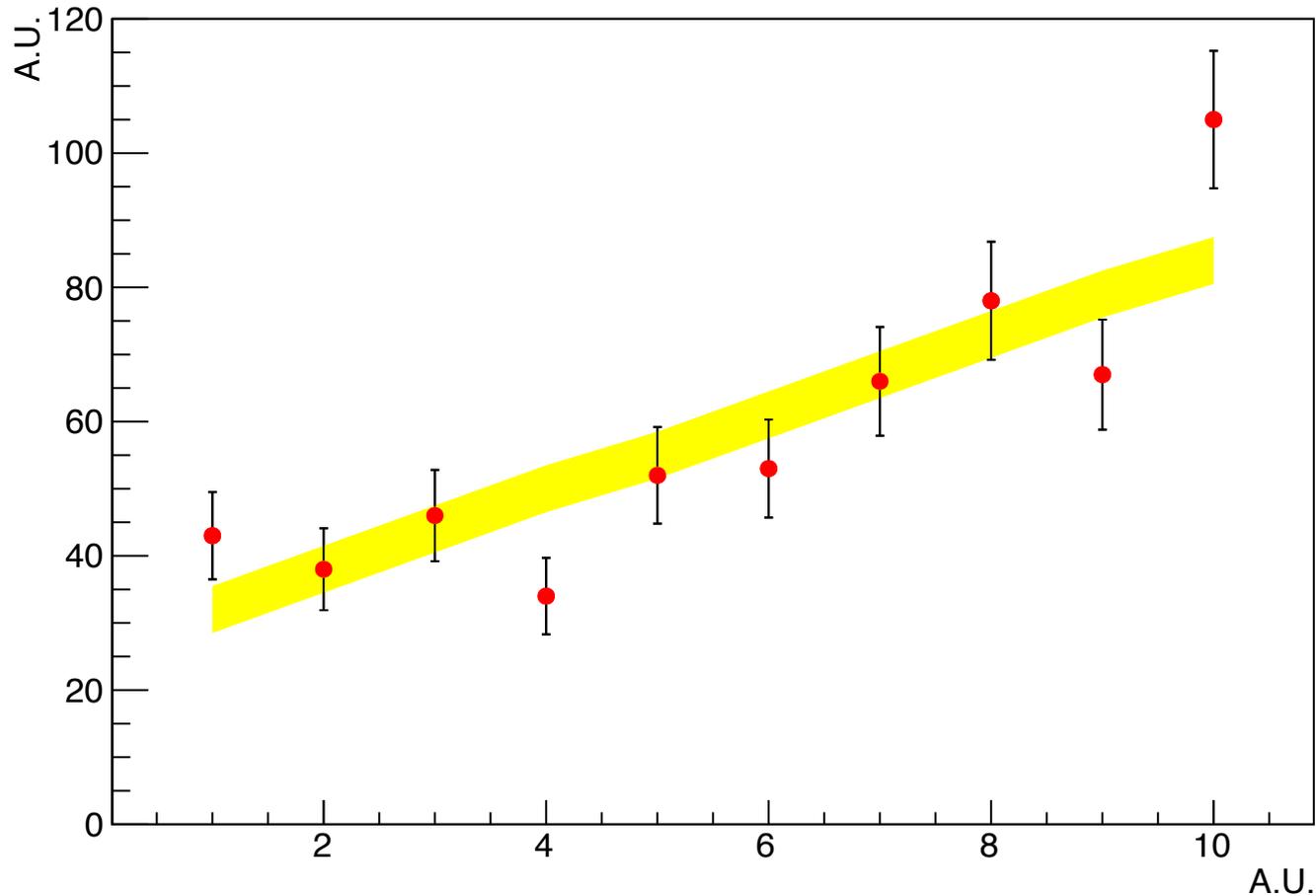
PARAMETERSCHÄTZUNG

Parameterschätzung

- Wahrscheinlichkeitsverteilungen (z.B. Gauß-, Poisson und Binomialverteilung) sind durch “**wahre**” **Parameter a** charakterisiert
- Erinnerung: Eine PDF $f(x)$ ist durch ihre Momente (Erwartungswert, Varianz und höhere vollständig bestimmt).
- **Messungen** von Verteilungen **bestehen aus Stichproben**, d.h. aus n einzelnen Elementen $x_1, x_2 \dots x_n$
- Aus der Stichprobe können **Schätzwerte \hat{a}** für die Parameter der PDF bestimmt werden.
- Die Schätzwerte \hat{a} sind selbst Zufallszahlen
- Forderungen an Methoden zur Bestimmung von Schätzwerten:
 - Konsistenz: $\lim_{n \rightarrow \infty} \hat{a} = a$
 - Erwartungstreue: $E(\hat{a}) = a$ auch für $n < \infty$ (!)
 - Effizienz: $V(\hat{a})$ möglichst klein.
 - Robust: Stabil gegen falsche Annahmen oder Daten



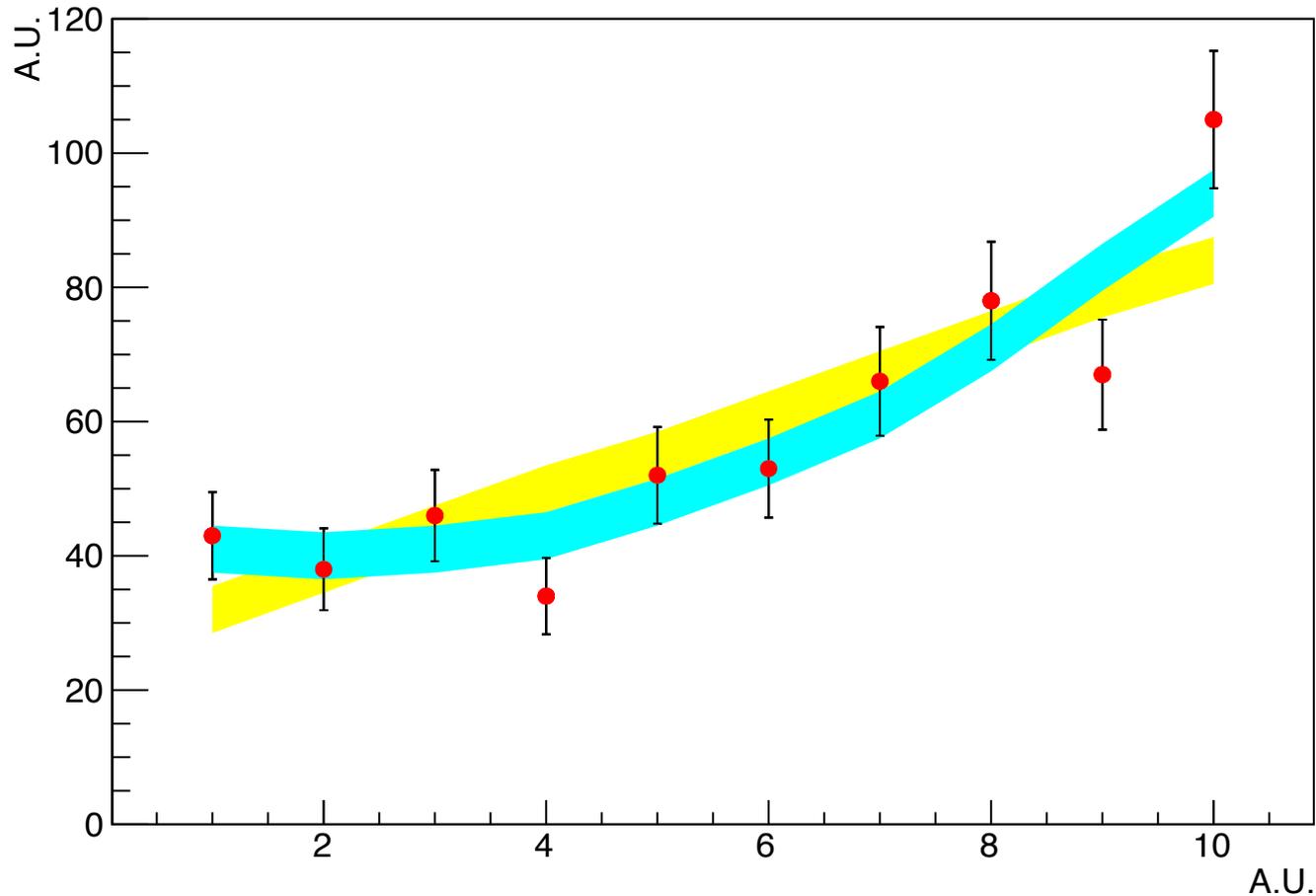
- 10 Datenpunkte x_i
- Annahme hier: Datenpunkte sind Gauß-verteilt und unkorreliert



■ Vergleich Daten und Theorie

■ Theorie dargestellt als Band (theoretische Fehler sind häufig korreliert)

Beispiel: `datavstheory.C`



- Vergleich Daten und Theorie
- Welche Theorie beschreibt die Daten besser?

“Goodness of Fit” (am Beispiel: χ^2 - Verteilung)

- Gegeben eine Vorhersage: Wie gut stimmt sie mit den Daten überein ?
- Beispiel: Gewichtete Summe der Abstands- (auch: Residuen-) quadrate zwischen Daten (mit Werten x_i und Unsicherheiten σ_i) und Funktionswerten $f(x_i)$

$$S = \chi^2 = \sum_{i=1}^N \left(\frac{y_i - f(x_i, \{p\})}{\sigma_i} \right)^2$$

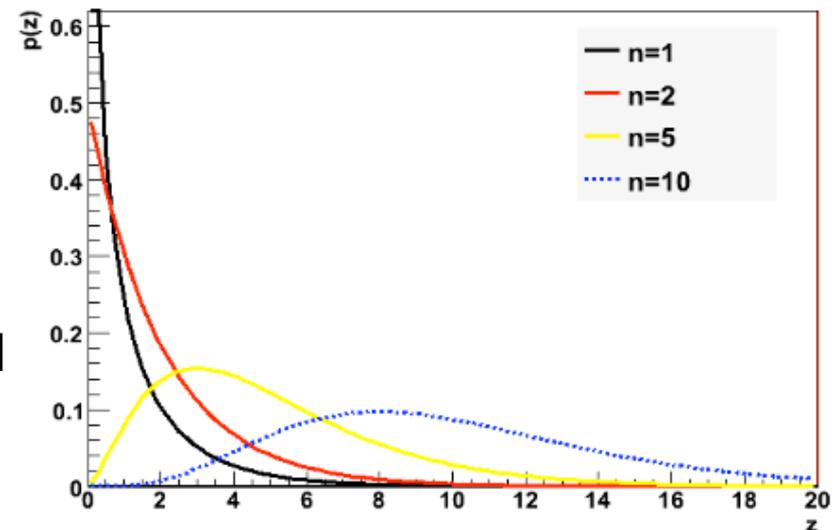
N Messwerte y_i
k Parameter p

- Bei Gauß-förmig verteilten y_i mit Unsicherheiten σ_i folgt S einer χ^2 -Verteilung mit $n = N - k$ Freiheitsgraden

$$f_n(\chi^2) = \frac{\frac{1}{2} \left(\frac{\chi^2}{2} \right)^{\frac{n}{2}-1} e^{-\frac{\chi^2}{2}}}{\Gamma\left(\frac{n}{2}\right)}$$

■ Erw.wert: $\langle \chi^2 \rangle = n \Rightarrow \langle \chi^2/n \rangle = 1$

■ Varianz: $V[\chi^2] = 2n$

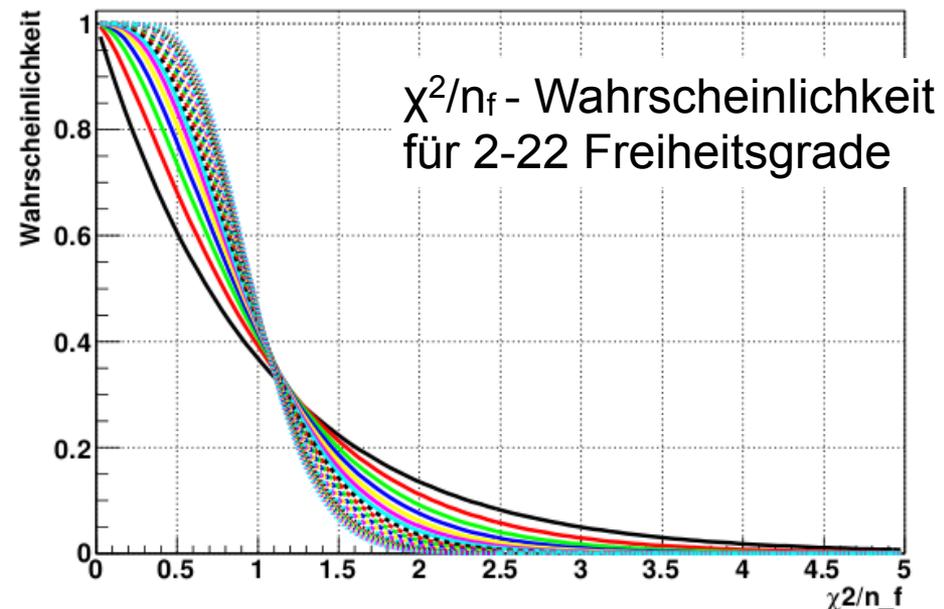
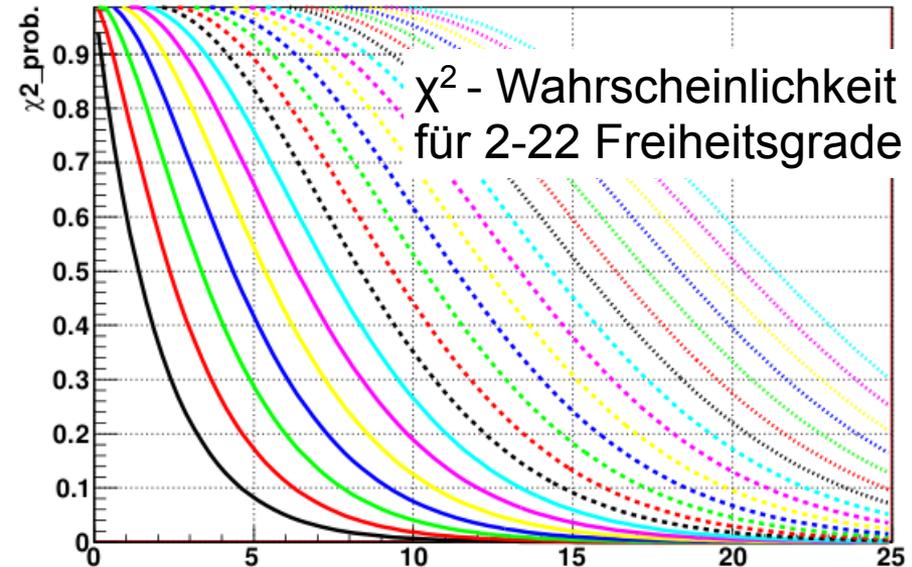


χ^2 - Wahrscheinlichkeit

Beispiele: [chi2PDF.C](#)
oder [chi2PDF-pyroot.py](#)

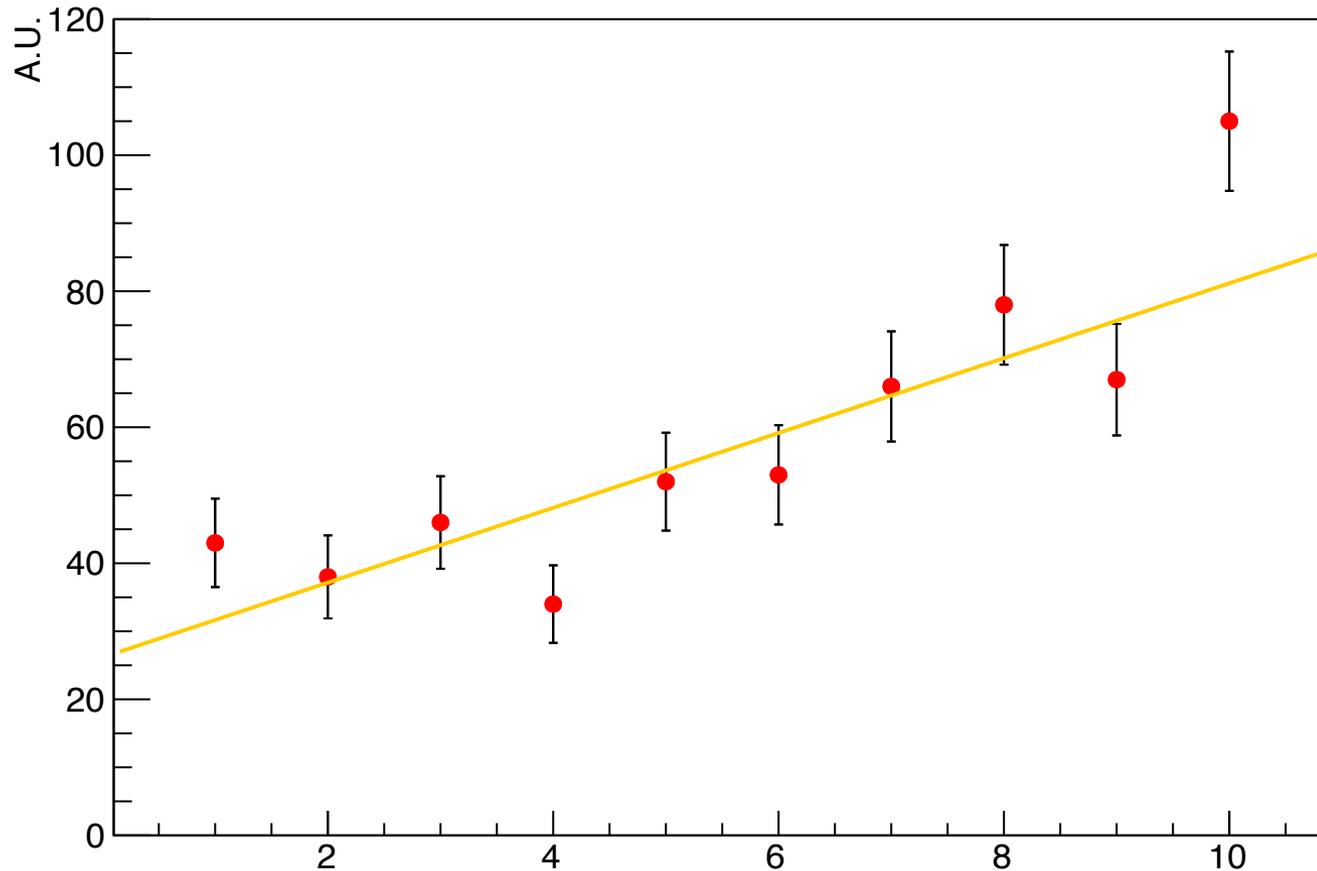
$$\begin{aligned}\chi_{\text{Prob}}^2 &= \int_{\chi^2}^{\infty} f_n(v) dv \\ &= 1 - \int_0^{\chi^2} f_n(v) dv\end{aligned}$$

- dient zur Quantifizierung der Übereinstimmung
- Aussage, mit welcher Wahrscheinlichkeit ein größerer Wert für χ^2 als der tatsächlich gemessene zu beobachten wäre
- Wahrscheinlichkeit für χ^2/n einen Wert > 1 zu beobachten ist $\sim 40\%$ (ziemlich unabhängig von n)



Parameterschätzung (“kleinste Quadrate”)

- Anpassung einer Funktion an die Daten: Lineare Funktion $ax+b$, d.h. 2 Parameter



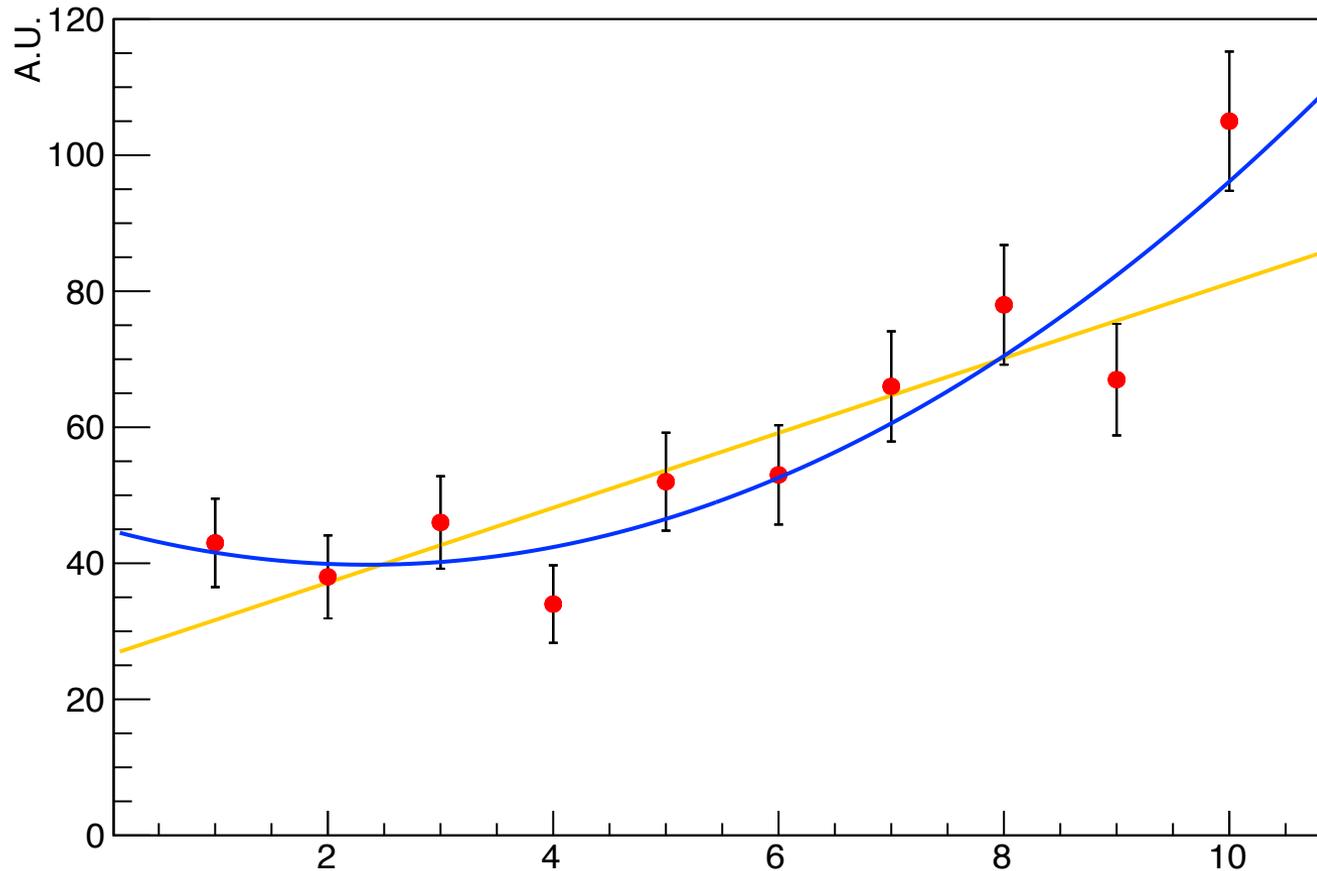
- Minimiere Summe der Abstandsquadrate:
- $\chi^2 / \text{ndof} = 17.6 / 8$ (p-Wert: 2.4%)

$$\chi^2 = \sum_{i=1}^N \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$

siehe auch Cowan Abschnitt 7.3

Parameterschätzung (“kleinste Quadrate”)

- Anpassung einer Funktion an die Daten: Parabel ax^2+bx+c , d.h. 3 Parameter



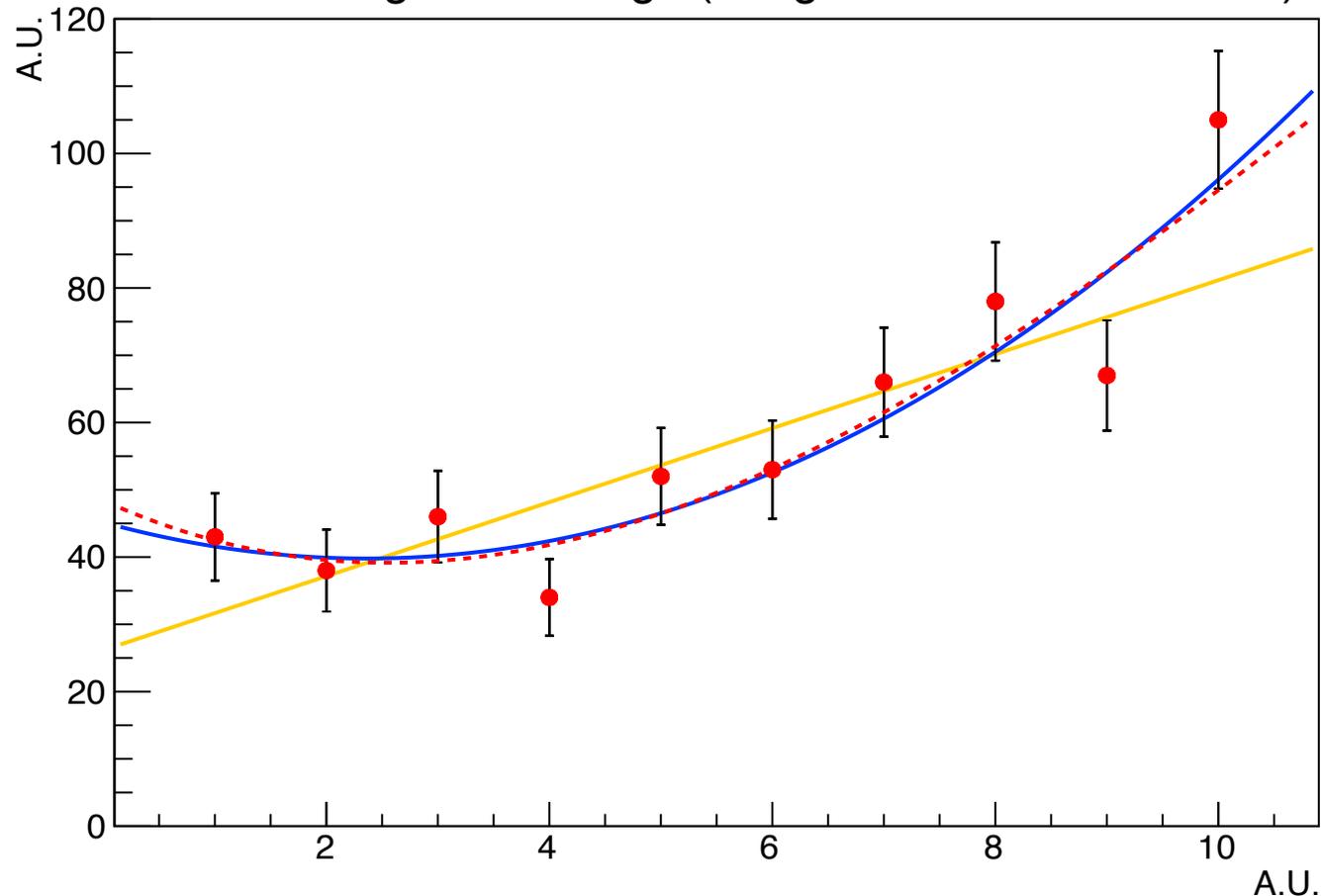
- Minimiere Summe der Abstandsquadrate:
- $\chi^2 / \text{ndof} = 9.1 / 7$ (p-Wert: 24%)

$$\chi^2 = \sum_{i=1}^N \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2$$

Parameterschätzung (“kleinste Quadrate”)

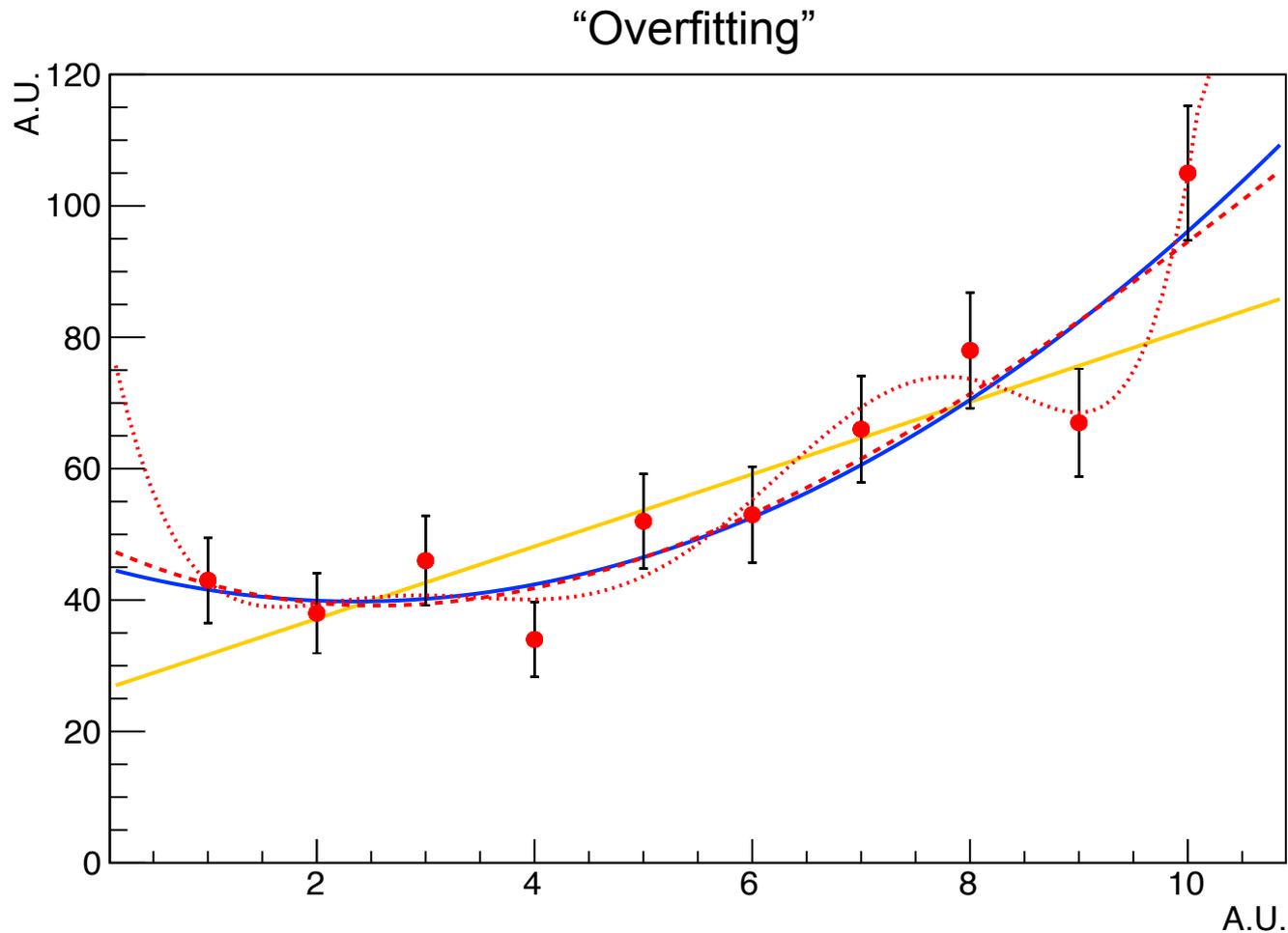
Frage: Wieviel freie Parameter soll man nehmen?

Antwort: möglichst wenige (hängt auch von Theorie ab)



- Anpassung einer Funktion an die Daten: Polynom 3. Ordnung, d.h. 4 Parameter
- $\chi^2 / \text{ndof} = 8.9 / 6$ (p-Wert: 17%)

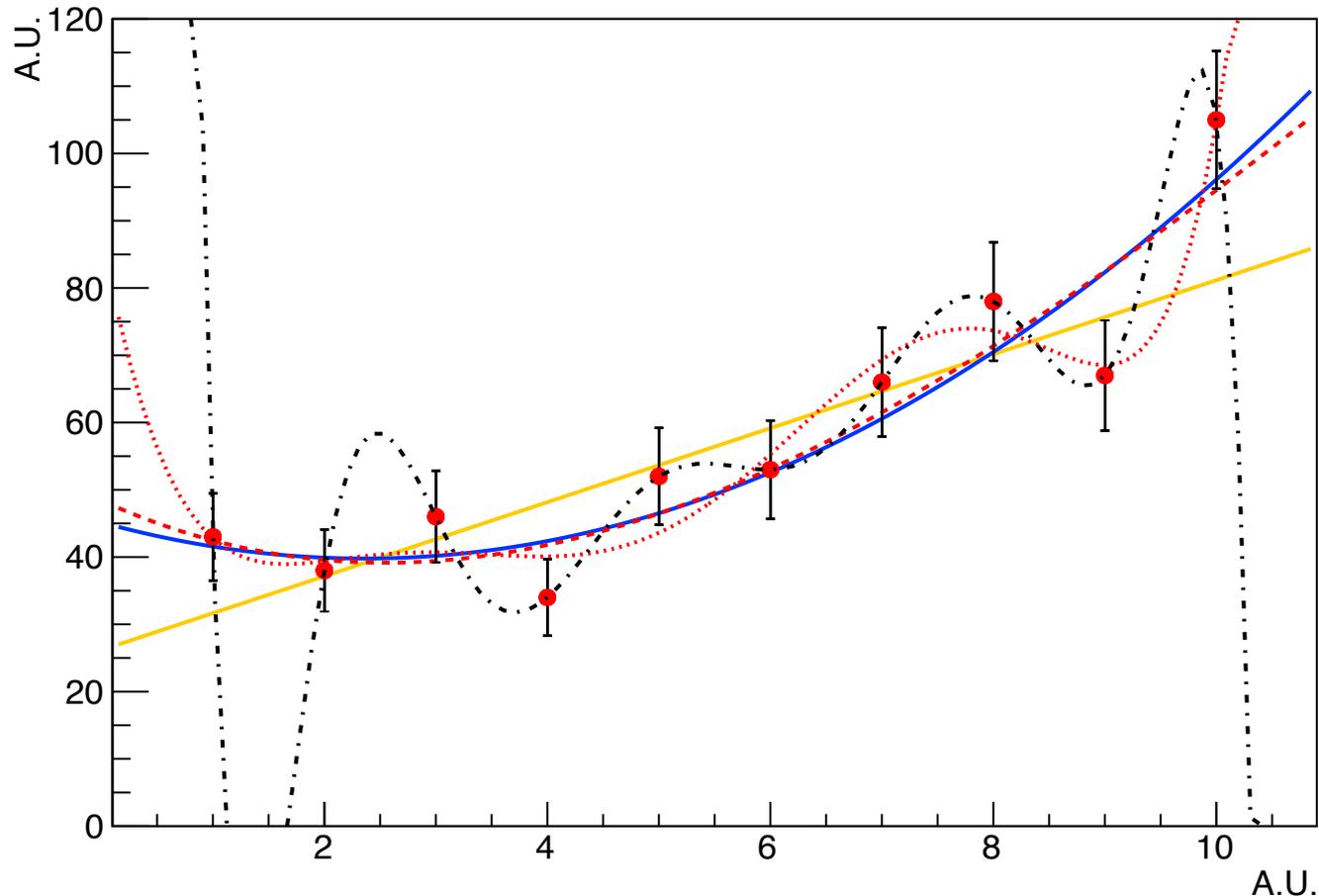
Parameterschätzung (“kleinste Quadrate”)



- Anpassung einer Funktion an die Daten: Polynom 7. Ordnung, d.h. 8 Parameter
- $\chi^2 / \text{ndof} = 3.7 / 2$ (p-Wert: 15%)

Parameterschätzung (“kleinste Quadrate”)

Im Limit vieler Parameter geht der Fit i.a. sicher durch die Daten, aber man hat nichts gelernt (Anzahl der Parameter = Anzahl der Datenpunkte)



- Anpassung einer Funktion an die Daten: Polynom 9. Ordnung, d.h. 10 Parameter
- $\chi^2 / \text{ndof} \sim 0. / 0$

Korrelation der Parameter

■ Betrachte erneut eine Geraden-Anpassung (in 2D)

Beispiel: fitf1.C

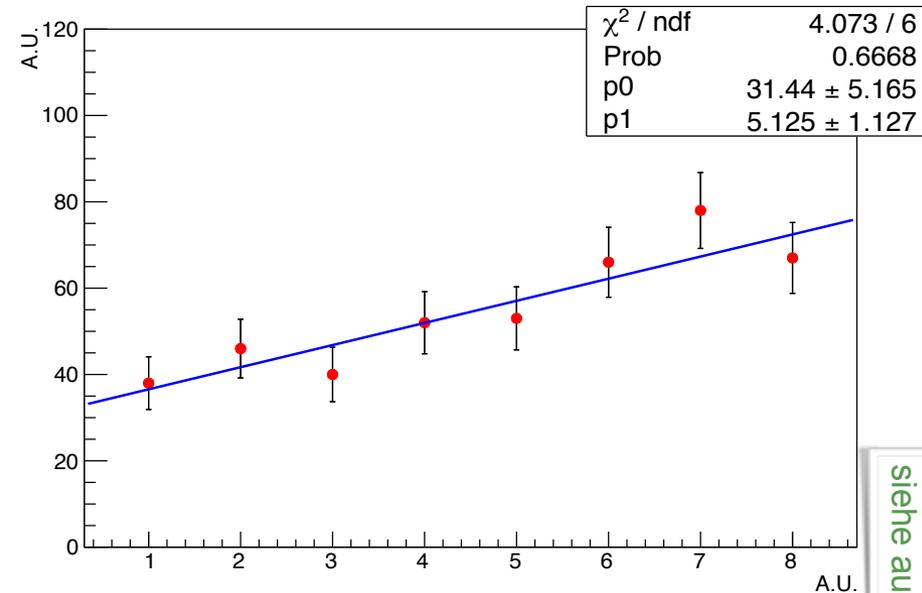
■ Ansatz: $f(x) = p_0 + p_1 x$

■ Ergebnis:

■ Parameter und χ^2 siehe Abb.

■ Kovarianz: $\begin{pmatrix} 26.7 & -5.07 \\ -5.07 & 1.27 \end{pmatrix}$

■ Korrelation: $\begin{pmatrix} 1 & -0.86 \\ -0.86 & 1 \end{pmatrix}$



siehe auch Cowan Abschnitt 7.3

■ Parameter p_0 und p_1 sind stark (anti-) korreliert!

(Anschaulich: Größeres p_0 verlangt kleineres p_1 , so dass $f(x)$ die Daten noch beschreibt)

Korrelation der Parameter

■ Betrachte erneut eine Geraden-Anpassung (2D)

Beispiel: fitf1.C

■ Ansatz: $f(x) = p_0 + p_1 (x - 4)$

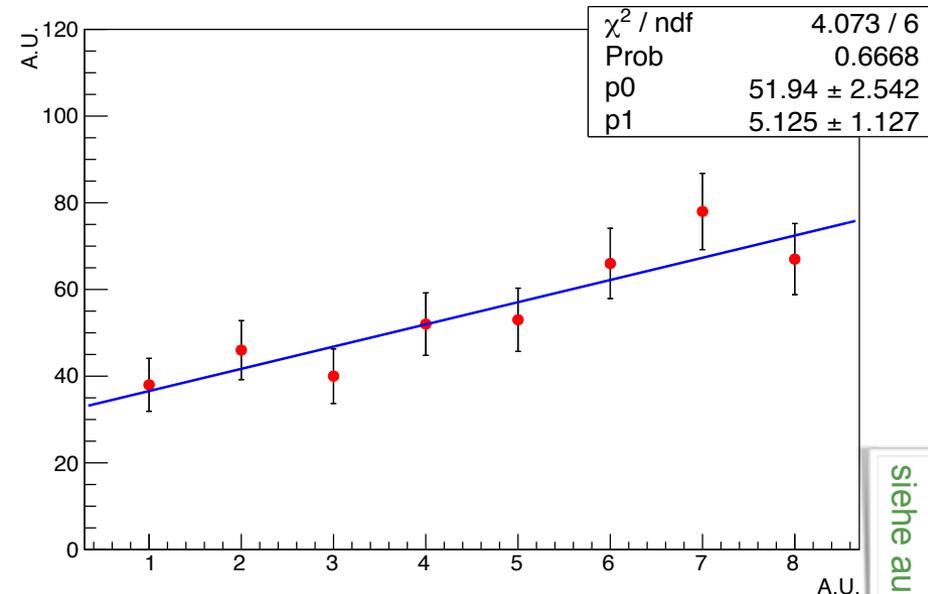
■ Ergebnis

■ Parameter und χ^2 siehe Abb.
 χ^2 , p-Wert, p_1 unverändert (!)

■ Kovarianz: $\begin{pmatrix} 6.4 & 0.013 \\ 0.013 & 1.27 \end{pmatrix}$

■ Korrelation: $\begin{pmatrix} 1 & 0.004 \\ 0.004 & 1 \end{pmatrix}$

■ Parameter p_0 und p_1 hier kaum noch korreliert

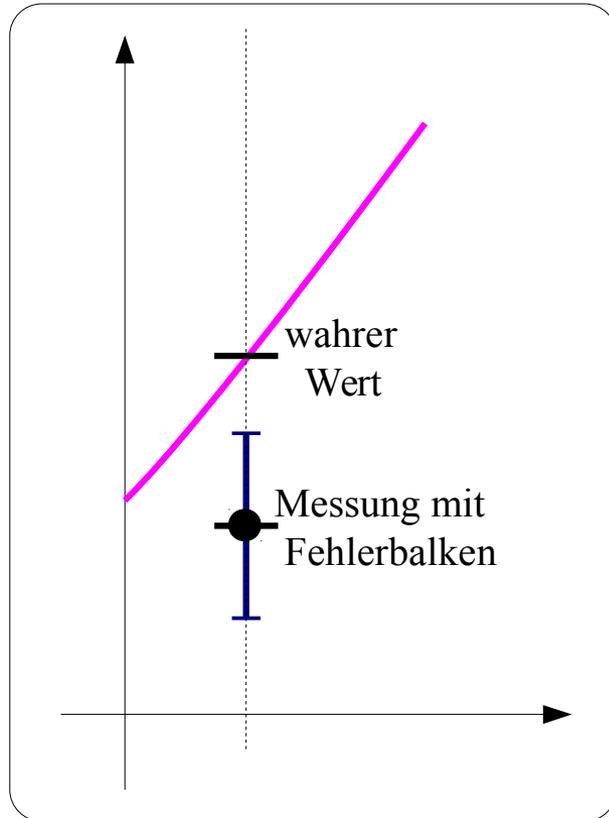


siehe auch Cowan Abschnitt 7.3

Dekorrelation durch geeignete Transformation der Koordinaten

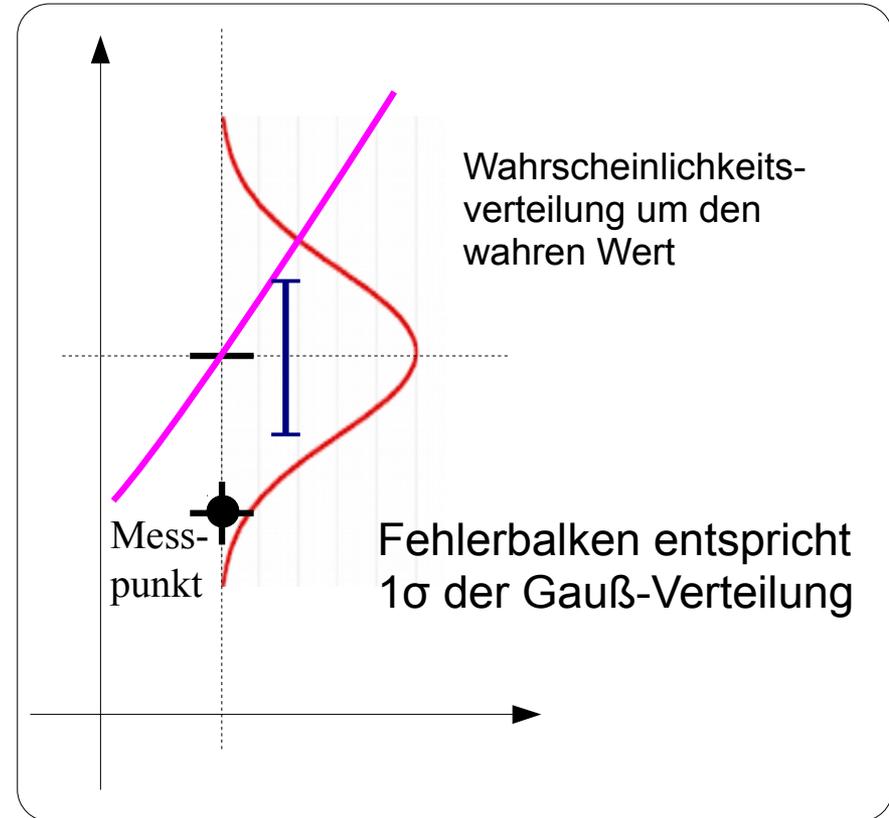
- Messungen sind Stichproben aus einer (wahren) Verteilung
- Abweichungen der Messung vom wahren Wert durch Unsicherheiten
 - Streuung (statistisch)
 - Verzerrung (systematisch)
- Theoretische Vorhersage der wahren Verteilung von experimentell bestimmten Modellparametern abhängig
 - Goodness-of-fit: Quantifizierung der Übereinstimmung von Daten und Vorhersage
 - Parameterschätzung: systematische Prozedur zur Bestimmung einer möglichst kleinen Zahl von Modellparametern und deren Unsicherheit (beschrieben durch Kovarianz-Matrix)

Übliche Darstellung



=

Eigentliche Bedeutung

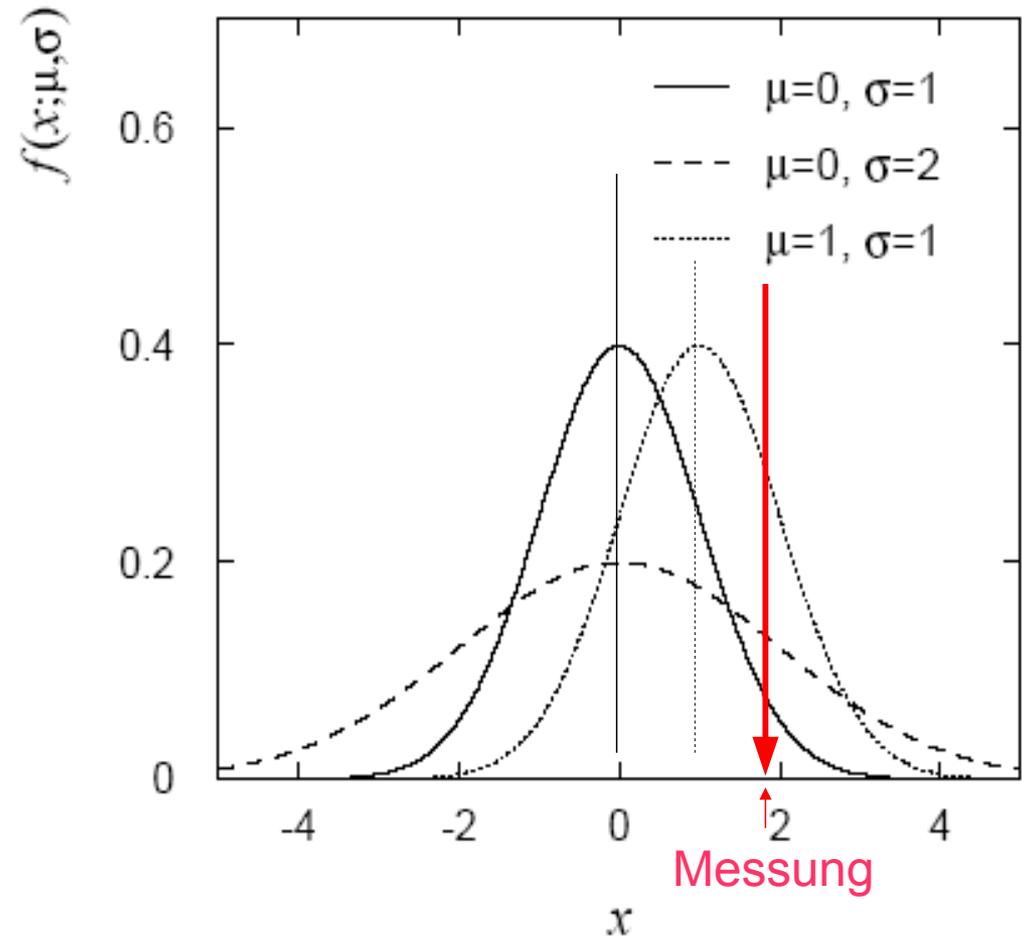


■ Messung mit Fehlerbalken bedeutet:

Beobachtung eines Messergebnisses, das einer Zufallszahl aus einer Verteilungsdichte um den wahren Wert entspricht

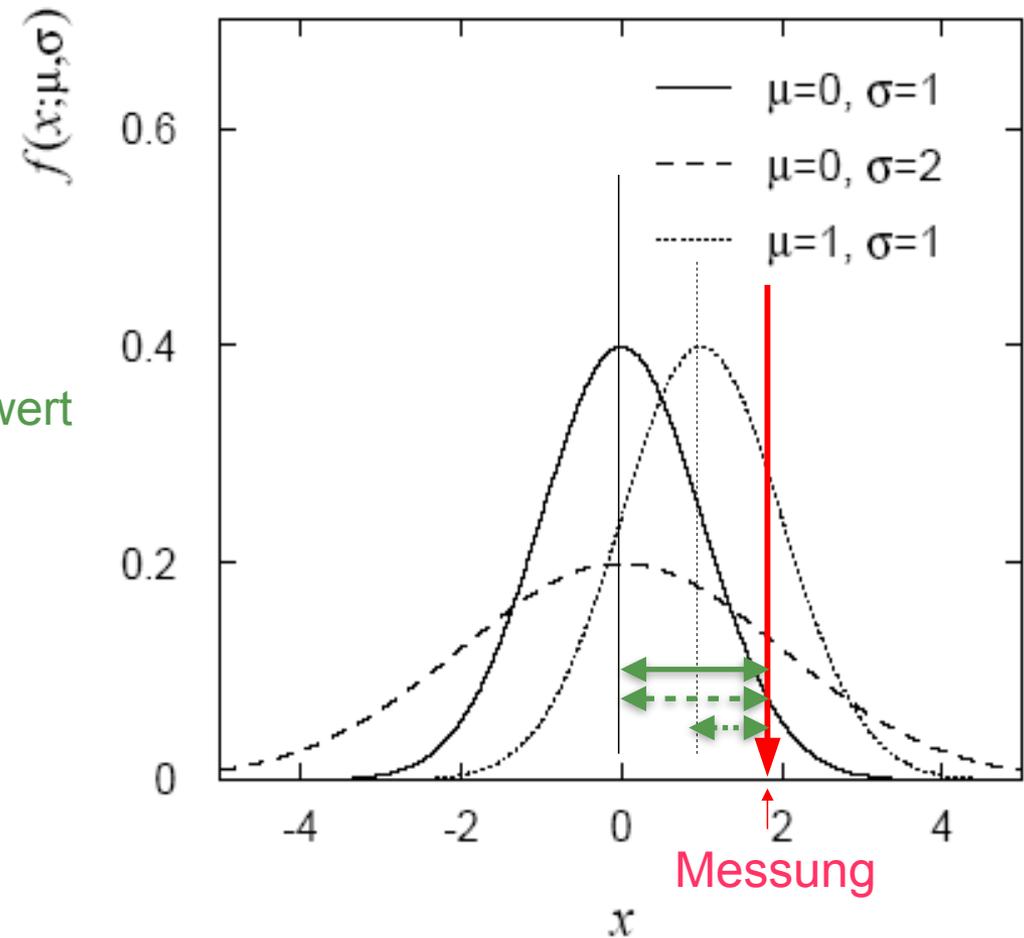
Welche der Verteilungen passt am besten zur Messung ?

Vergleiche Messung
mit drei alternativen
Gauß-Verteilungen



- LS: Least Squares:
Kleinsten Abstand zwischen
Messung und Erwartungswert

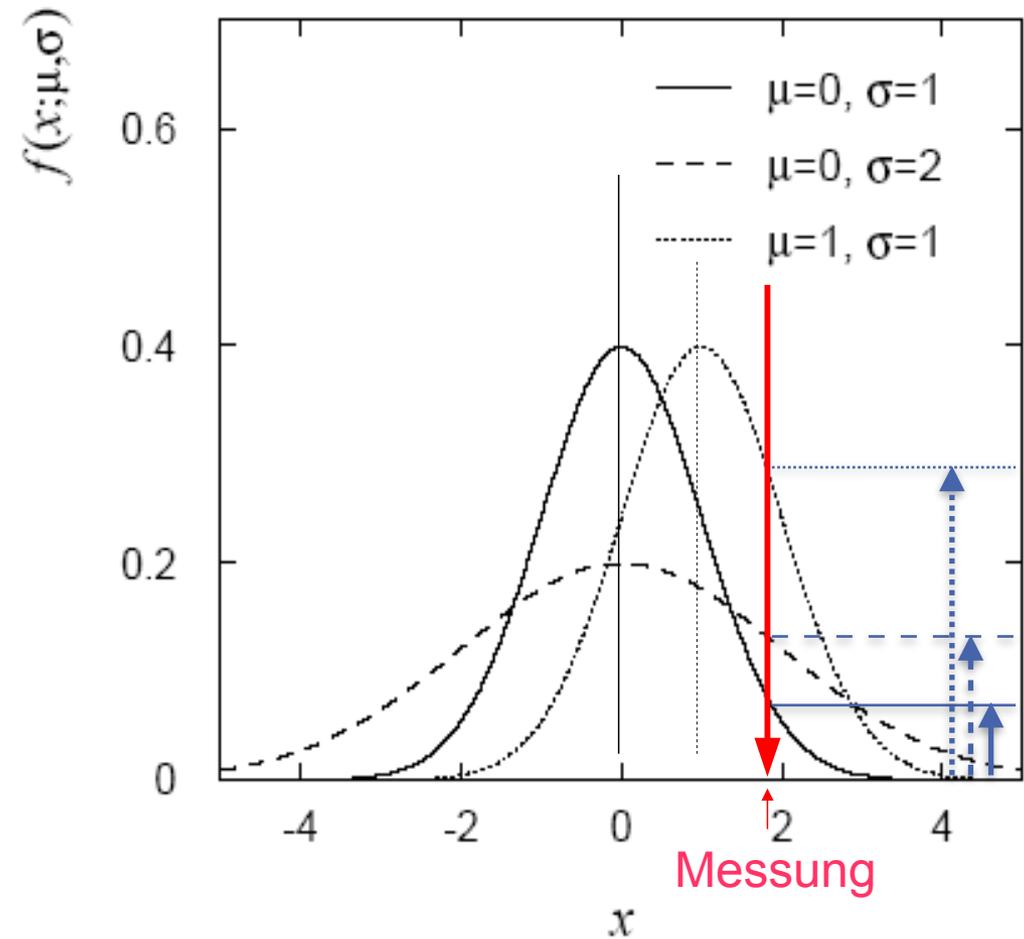
Minimiere Abstand vom Sollwert



ML: Maximum Likelihood

- ML: Maximum Likelihood:
Maximale Wahrscheinlichkeit

Maximiere Höhe der PDF



Vergleich: ML und LS

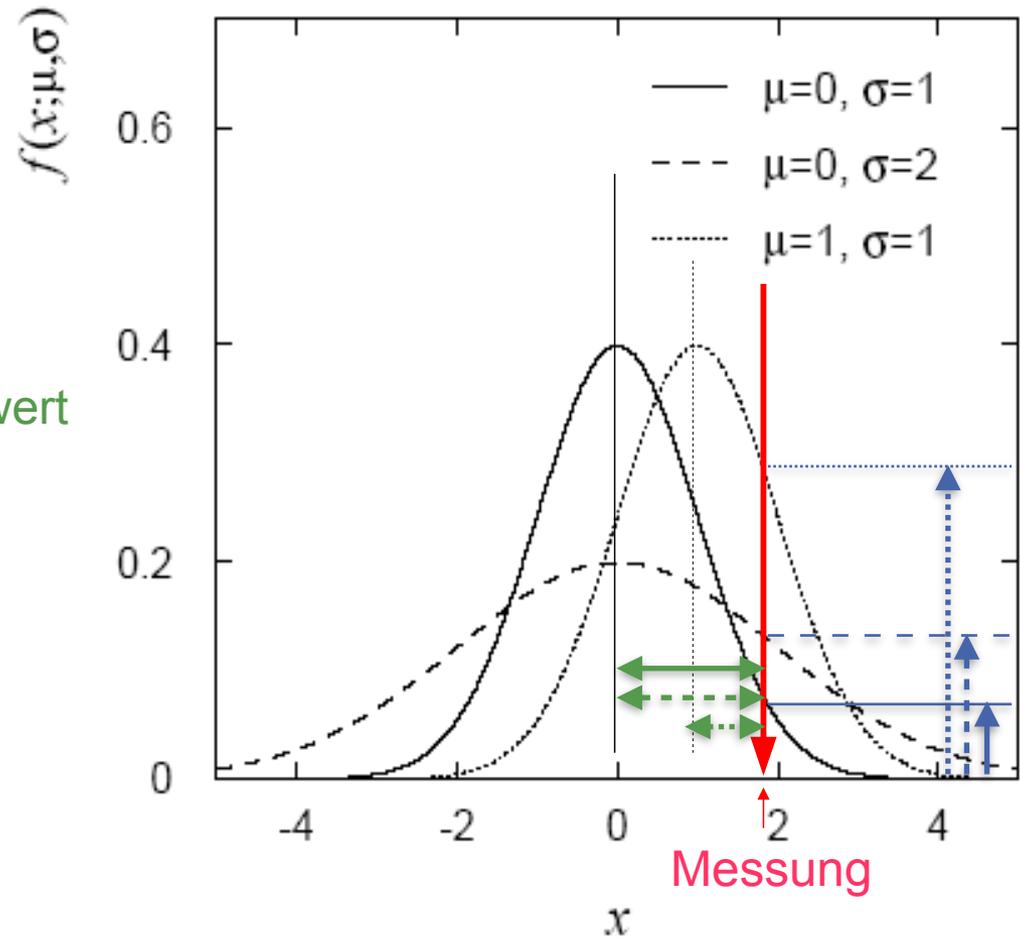
- ML: Maximum Likelihood:
Maximale Wahrscheinlichkeit

Maximiere Höhe der PDF

- LS: Least Squares:
Kleinster Abstand zwischen
Messung und Erwartungswert

Minimiere Abstand vom Sollwert

- Für diese Messung: gleiches
Ergebnis: $\mu=1, \sigma=1$ bevorzugt



Vergleich: ML und LS

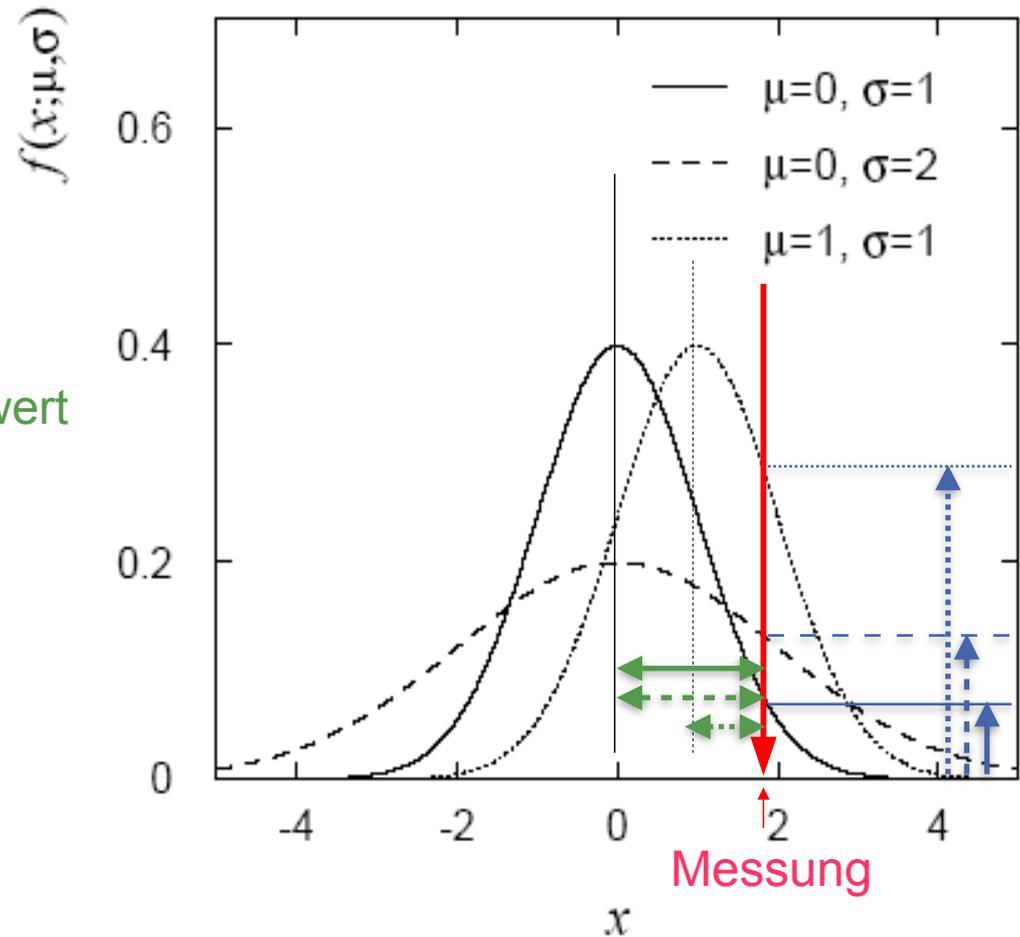
- ML: Maximum Likelihood:
Maximale Wahrscheinlichkeit

Maximiere Höhe der PDF

- LS: Least Squares:
Kleinsten Abstand zwischen
Messung und Erwartungswert

Minimiere Abstand vom Sollwert

- Für diese Messung: gleiches
Ergebnis: $\mu=1, \sigma=1$ bevorzugt



Im allgemeinen geben ML und LS unterschiedliche Ergebnisse!

Vergleich: ML und LS

- ML: Maximum Likelihood:
Maximale Wahrscheinlichkeit

Maximiere Höhe der PDF

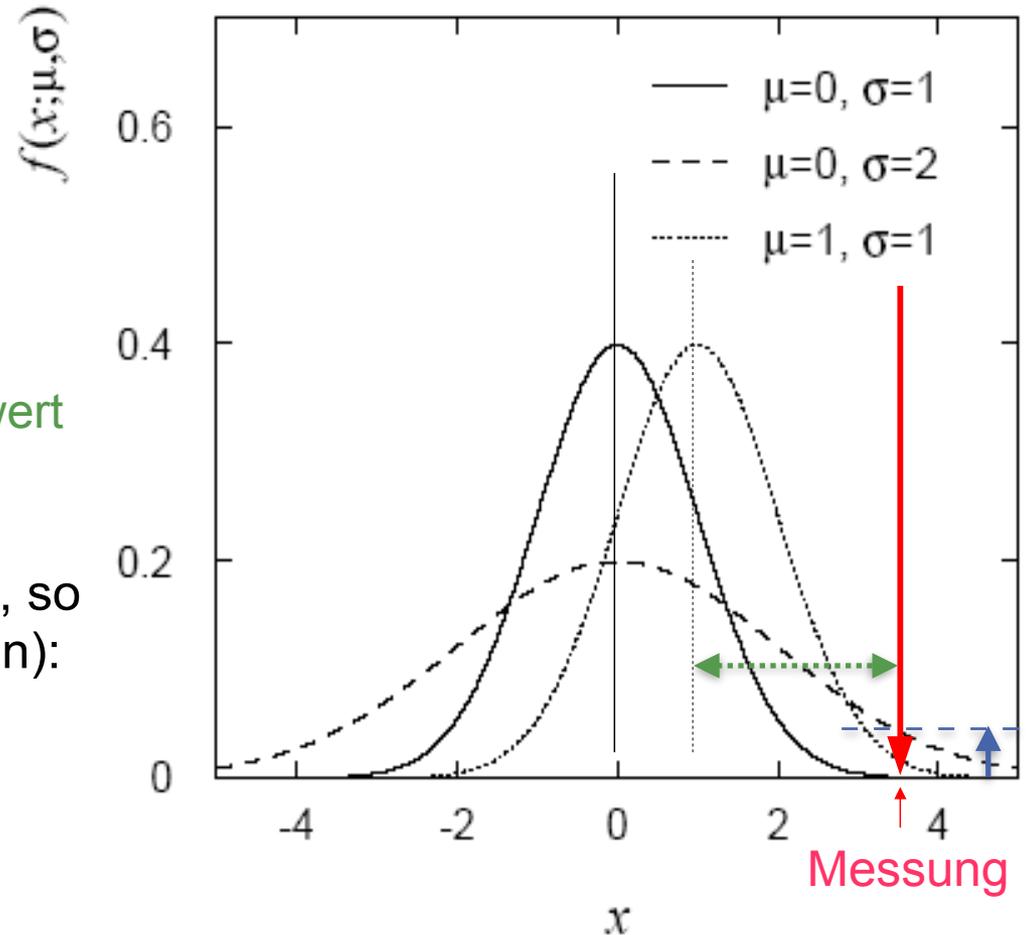
- LS: Least Squares:
Kleinster Abstand zwischen
Messung und Erwartungswert

Minimiere Abstand vom Sollwert

- Läge der Messwert z.B. bei 3.5, so
wäre (für diese drei Verteilungen):

- ML: $\mu=0, \sigma=2$

- LS: $\mu=1, \sigma=1$



Im allgemeinen geben ML und LS unterschiedliche Ergebnisse!

Maximum Likelihood

■ Likelihood-Funktion:

- Die Werte $\mathbf{x} = (x_1, \dots, x_n)$ seien n unabhängige Messpunkte
- Das Produkt der Werte der Wahrscheinlichkeitsdichten $P(\mathbf{x}|\mathbf{a})$

$$\mathcal{L}(\mathbf{a}) = P(x_1|\mathbf{a}) \cdot P(x_2|\mathbf{a}) \cdot \dots \cdot P(x_n|\mathbf{a}) = \prod_{i=1}^n P(x_i|\mathbf{a})$$

hängt nur von den Parametern \mathbf{a} ab.

■ Maximum-Likelihood Prinzip:

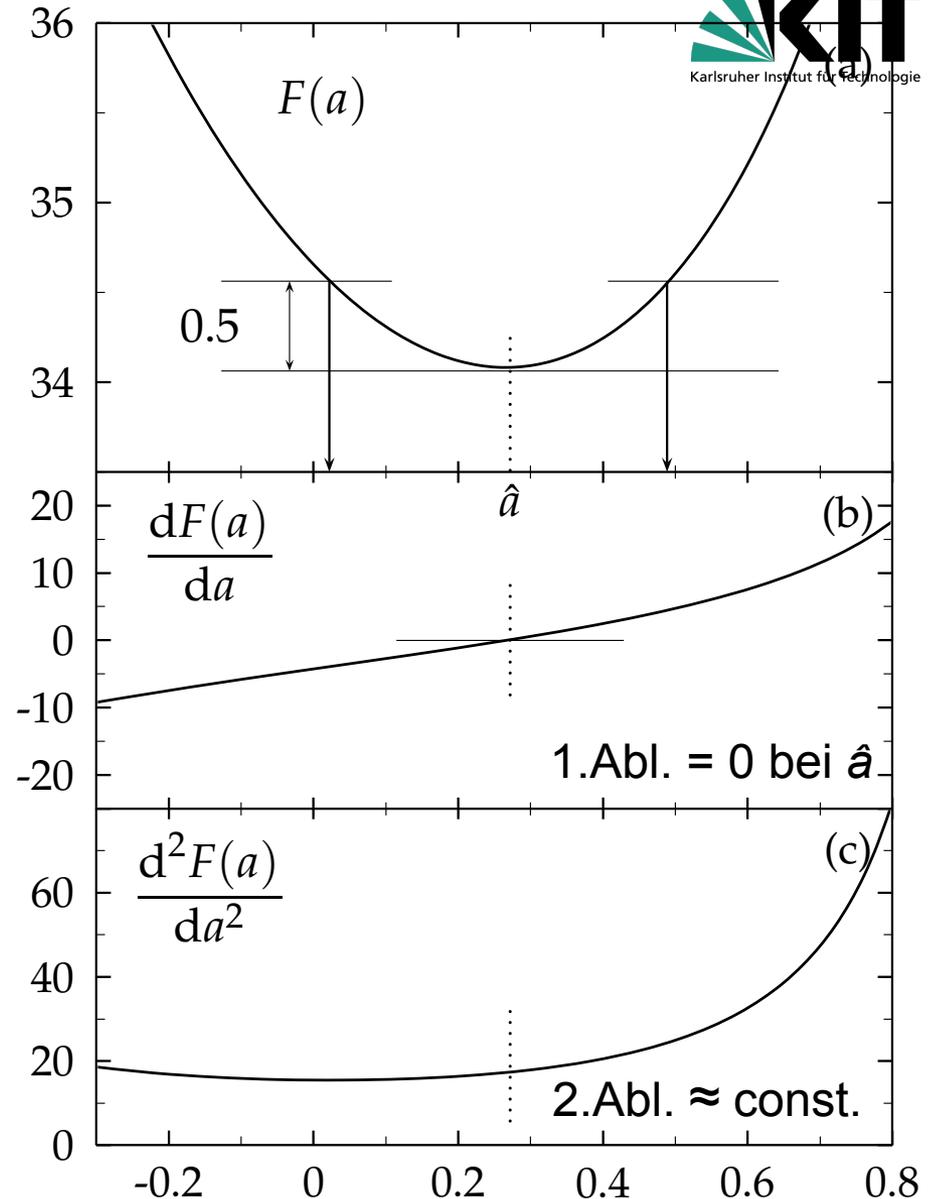
- Der beste Schätzwert $\hat{\mathbf{a}}$ für den Parametervektor \mathbf{a} ist der, der die Likelihood-Funktion $\mathcal{L}(\mathbf{a})$ maximiert
- In der Praxis: minimiere den negativen Logarithmus:

$$F(\mathbf{a}) = -\ln \mathcal{L}(\mathbf{a}) = -\sum_{i=1}^n \ln (P(x_i|\mathbf{a}))$$

Maximum Likelihood

- Der beste Schätzwert \hat{a} für a liegt am Minimum F_{min} von $F(\hat{a})$
- Fehlerabschätzung:
 - Bestimmung des $\pm 1\sigma$ Intervalls aus der Bedingung $F_{min}(\hat{a} + \sigma_{r,l}) = F(\hat{a}) + 0.5$
 - Die Unsicherheit kann direkt aus $F(\hat{a})$ abgelesen werden.
- Allgemein: Varianz = 1/Krümmung bei mehreren Parametern:

$$(\text{cov}^{-1})_{ij} = \partial^2 F / \partial a_i \partial a_j$$
- Beispiel hier: fast Gauß-förmige PDF. Für Gauß ist die 2.Abl. = const. und die Likelihood ist eine Parabel



Blobel/Lohrmann Abbildung 6.3

Vergleich: ML und LS

- Für Gauß-förmig verteilte Messungen sind LS und ML äquivalent:

- Gauß-PDF:
$$f(x_i|a) = \frac{1}{\sqrt{2\pi}\sigma_i} \cdot \exp\left[-\frac{(x_i - a)^2}{2\sigma_i^2}\right]$$

- Der negative Logarithmus dieser Likelihood, $F(a) = -\ln \prod_i f(x_i|a)$ ist:

$$F(a) = -\ln \mathcal{L}(a) = \underbrace{\frac{1}{2} \sum_i \frac{(x_i - a)^2}{\sigma_i^2}}_{\chi^2} + \underbrace{\sum_i \ln(\sqrt{2\pi}\sigma_i)}_{\text{konstant bzgl. } a}$$

- Also:

$$\Delta(-\ln L) = \frac{1}{2} \Delta\chi^2 \quad \text{und} \quad \partial^2(-\ln L)/\partial a_i \partial a_j = \frac{1}{2} \partial\chi^2/\partial a_i \partial a_j$$

- Die χ^2 -Methode ist ein (wichtiger) Spezialfall der ML Methode, nämlich für die Annahme Gaußischer Unsicherheiten der Messwerte

Vergleich: ML und LS

- Fehlerbestimmung: $\Delta(-\ln L)$ und $\Delta\chi^2$:
in Quantilen der Gauß-Verteilung

	$\Delta(-\ln L)$	$\Delta\chi^2$
1σ	0.5	1
2σ	2.0	4
3σ	4.5	9
$n\sigma$	$n^2/2$	n^2

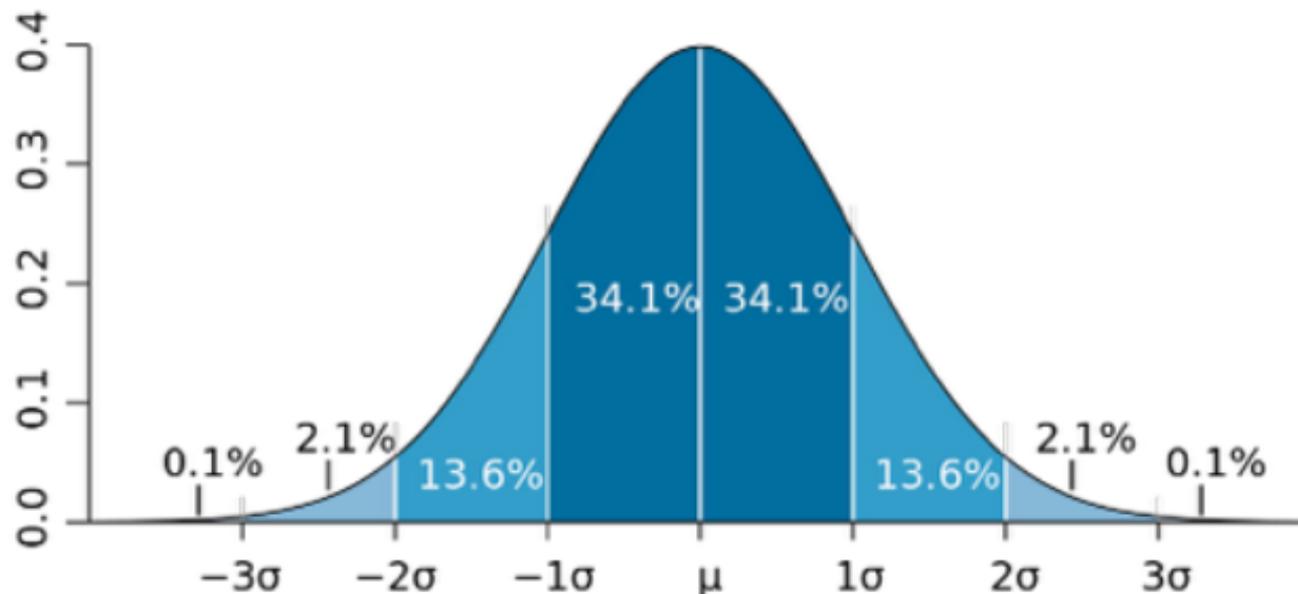
	Maximum Likelihood	Kleinste Quadrate
Methode	Höhe der PDF	Abweichung vom Mittelwert
Voraussetzungen	PDF ist bekannt	Mittelwert und Varianz
Effizienz	maximal	maximal bei linearen Problemen
Komplexität	aufwändig	oft exakt analytisch lösbar
Güte der Anpassung	nein	ja: χ^2 -Wahrscheinlichkeit
Robustheit	nein	nein

1.4 AUSBlick

Konfidenz-Intervalle

- Frequentistische Sichtweise: Es existiert ein fester wahrer Wert a . Das Konfidenz-Intervall zu einem Konfidenz-Niveau $CL=p\%$ enthält den wahren Wert a in $p\%$ aller Fälle.
- Annahme bisher: Schätzwert \hat{a} ist Gauß-verteilt um wahren Wert $a \rightarrow$ Angabe des Intervalls in Einheiten der Standardabweichung σ

“Punktschätzer”



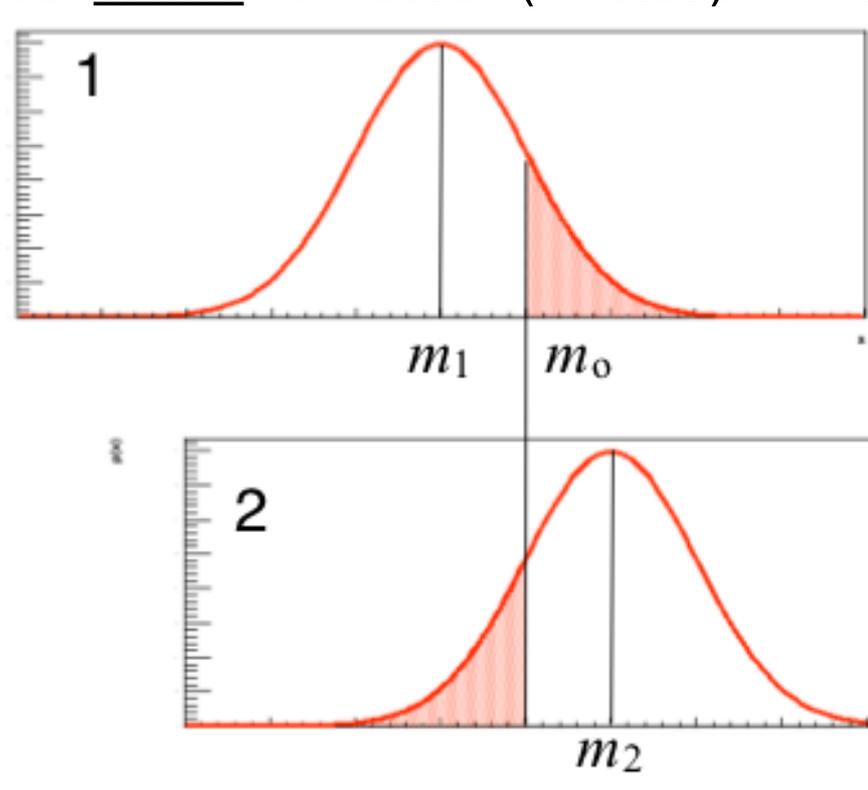
Konfidenz-Intervalle

- Betrachte ein 2-seitiges 68%-Konfidenz-Intervall für Gauß-förmige PDF und Messwert m_0
- Untere Intervall-Grenze m_1 : Für einen wahren Wert m_1 ist die Wahrscheinlichkeit einen Wert m_0 oder größer zu messen $(1 - 0.68)/2 = 16\%$
- Obere Intervall-Grenze m_2 : Für einen wahren Wert m_2 ist die Wahrscheinlichkeit einen Wert m_0 oder kleiner zu messen $(1 - 0.68)/2 = 16\%$

Das Intervall $[m_1, m_2]$ soll den wahren Wert in 68% aller Fälle enthalten

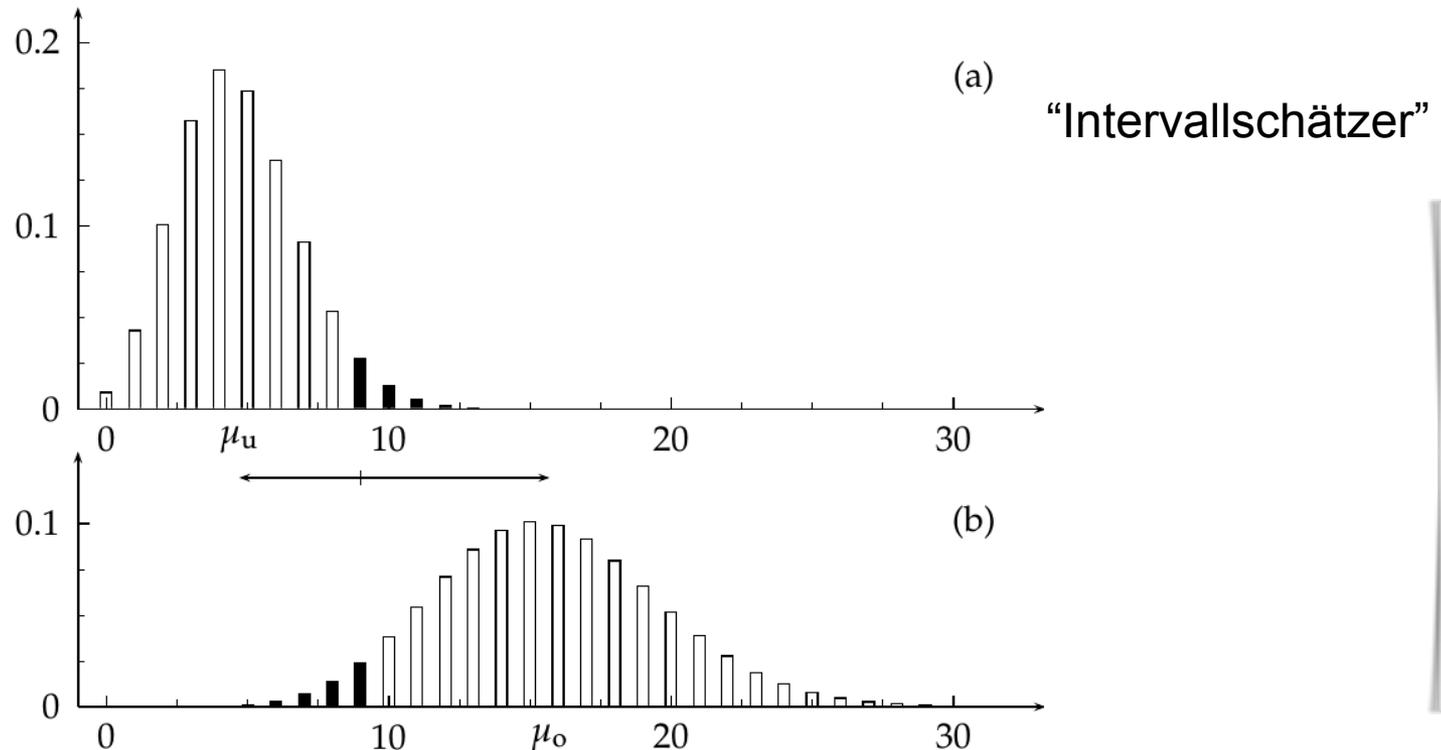
Gauß-Verteilung:

$$m_{21} = m_0 \pm \sigma$$



“Intervallschätzer”

- Nicht immer Gauß-verteilt: Beispiel: Poisson-Verteilung mit $n=9$



- Für eine Messung von $n=9$ bekommt man ein 90% Konfidenz-Intervall (5% oben und 5% unten) von $\mu_u = 4.7$ und $\mu_o = 15.7$, asymmetrisch.
- (Vergleich Annahme Gauß-PDF für $n=9$: $9 \pm 1.645 \cdot \sqrt{9} = 9 \pm 4.9$)

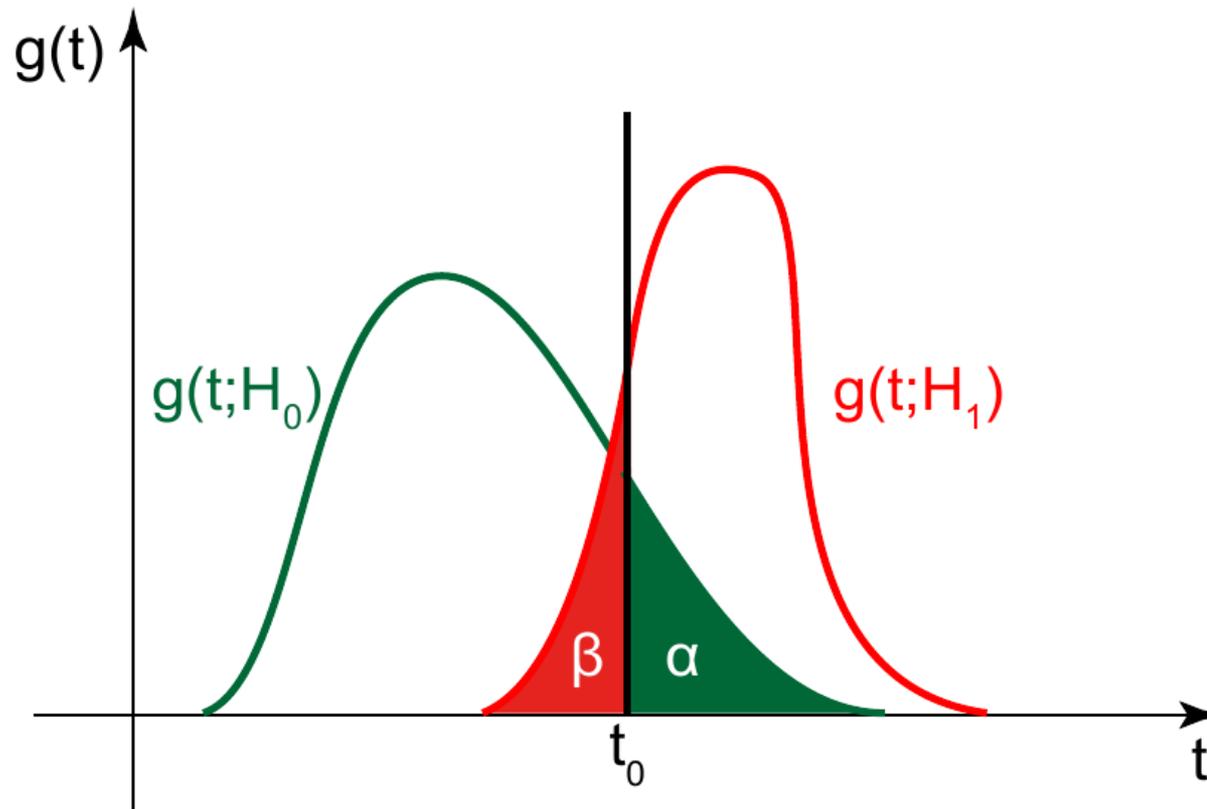
- Verwende statistische Methoden zur Entscheidungsfindung

- Beispiele:
 - Gibt es eine globale Erwärmung ?
 - Wirkt ein Medikament wirklich ?
 - Gibt es einen Unterschied bei Prüfungsergebnissen von verschiedenen Gruppen (Studienfach, Geschlecht, Jahrgang, Herkunft ...)?
 - Werden Vorlesungen durch Evaluation besser ?
 - Soll ich morgen einen Regenschirm mitnehmen ?
 - Welche Werbung zeige ich Personen, die eine Webseite besuchen?

- Allgemein: Vergleich der Daten mit Hypothesen H_i

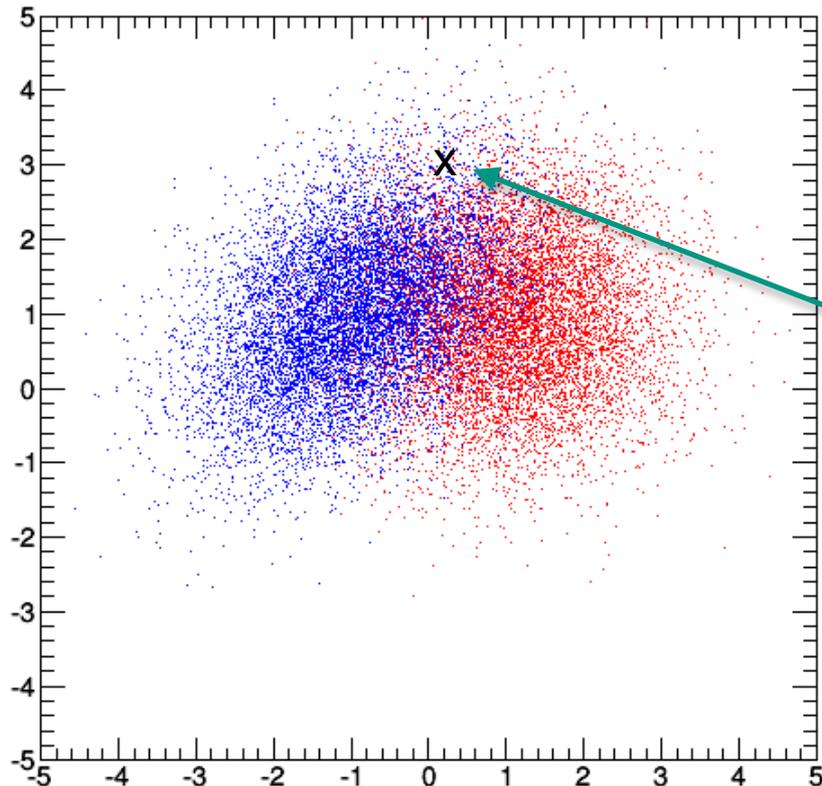
Hypothesentests

- Hypothesen werden durch PDF einer Prüfgröße t beschrieben.
- Quantitative Auswertung liefert Entscheidung (mit definiertem p -Wert)



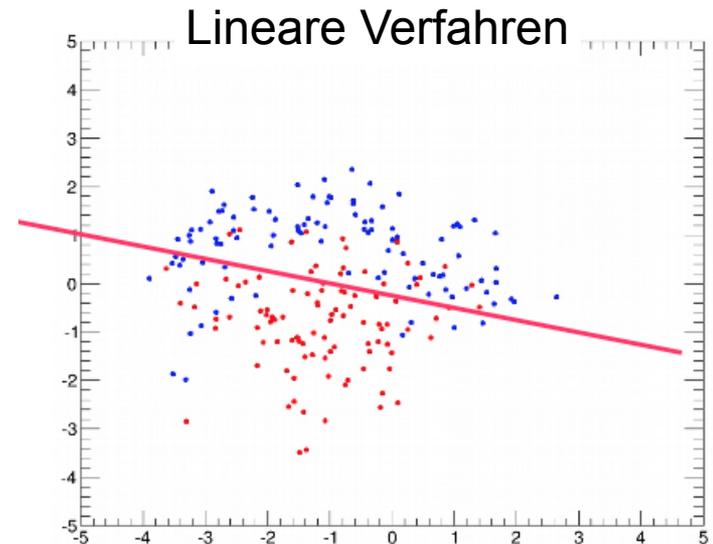
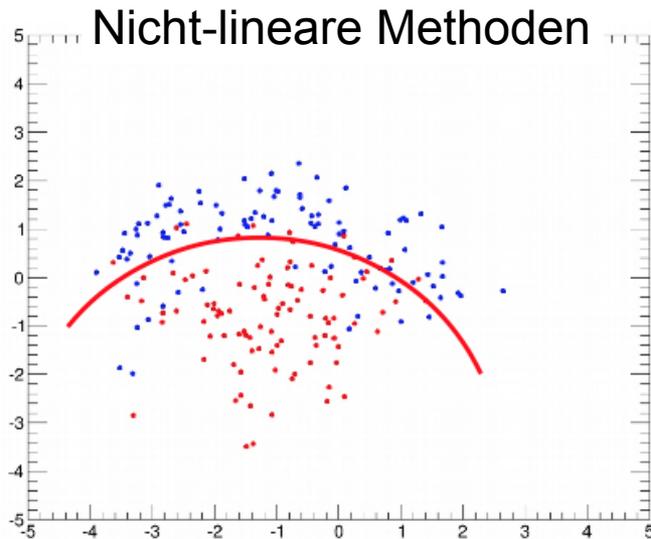
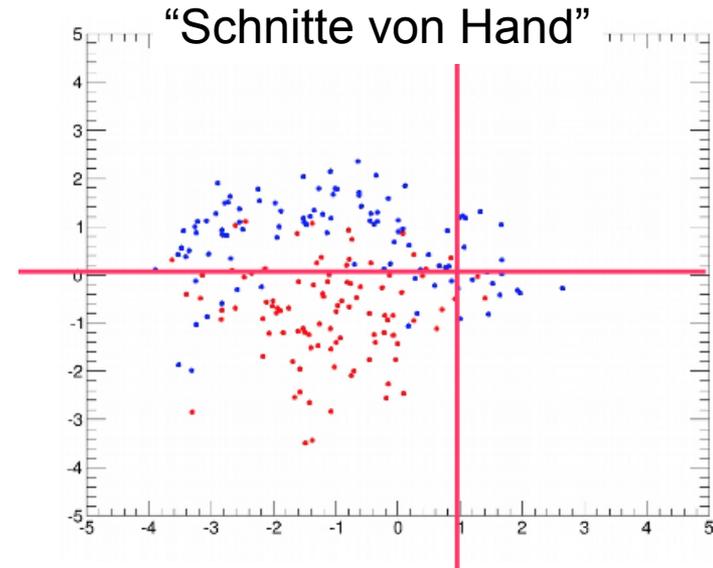
Prüfgrößen für zwei Hypothesen: Wahl von t_0 bestimmt Entscheidung

- Gehört ein Ereignis zu einer von zwei oder mehreren Klassen, z.B. Signal oder Untergrund ?
- Zufallsereignis beschrieben durch n Zufallsvariable x_1, \dots, x_n
- Klasse k beschrieben durch PDF $f_k(x_1, \dots, x_n)$



Gehört Punkt x
zu Klasse 0 oder 1?

- Maschinelles Lernen: Optimaler Klassifikator wird algorithmisch ermittelt.
- Besonders wichtig bei nicht-linearen Methoden
- Diese Vorlesung: Focus auf Supervised Learning (Lernen anhand von vorklassifizierten Daten)



- Fehlerfortpflanzung
- Korrelationen
- Goodness of Fit
- Parameterschätzung: Maximum-Likelihood und kleinste Quadrate

- Ausblicke
 - Hypothesentests
 - Konfidenz-Intervalle
 - Klassifikation
 - Messen, Entfalten und Systematischer Fehler

- Am 9.5.: Zufallszahlen und die Monte-Carlo Methode