Karlsruher Institut für Technologie

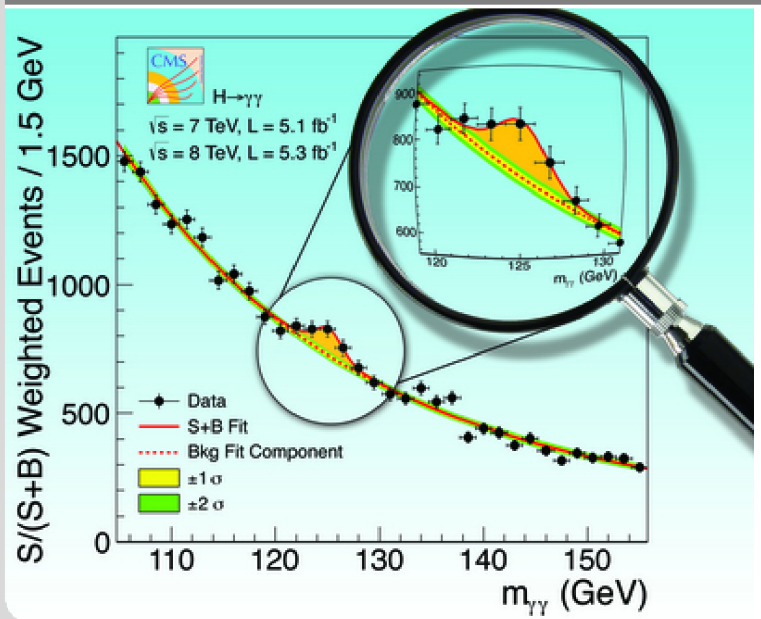
# Moderne Methoden der Datenanalyse

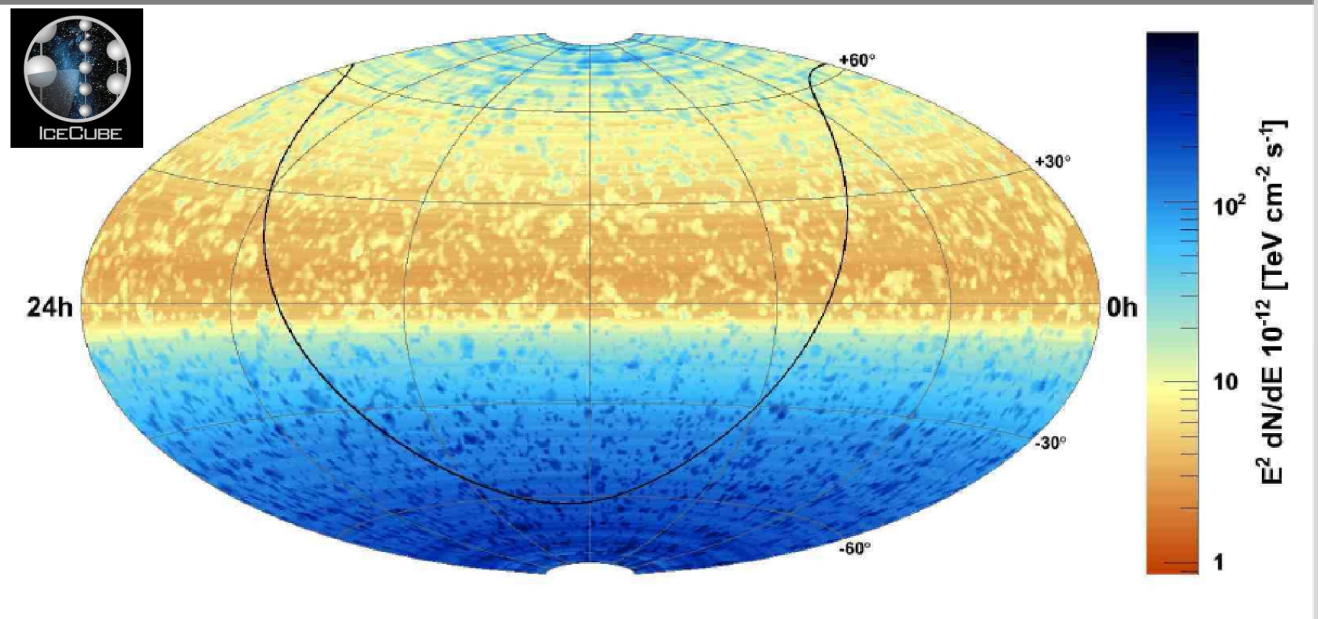
23. Mai

**Ralf Ulrich**, Andreas Meyer

Institut für Kernphysik, Institut für Experimentelle Kernphysik

**Master-Kurs SS 2017**

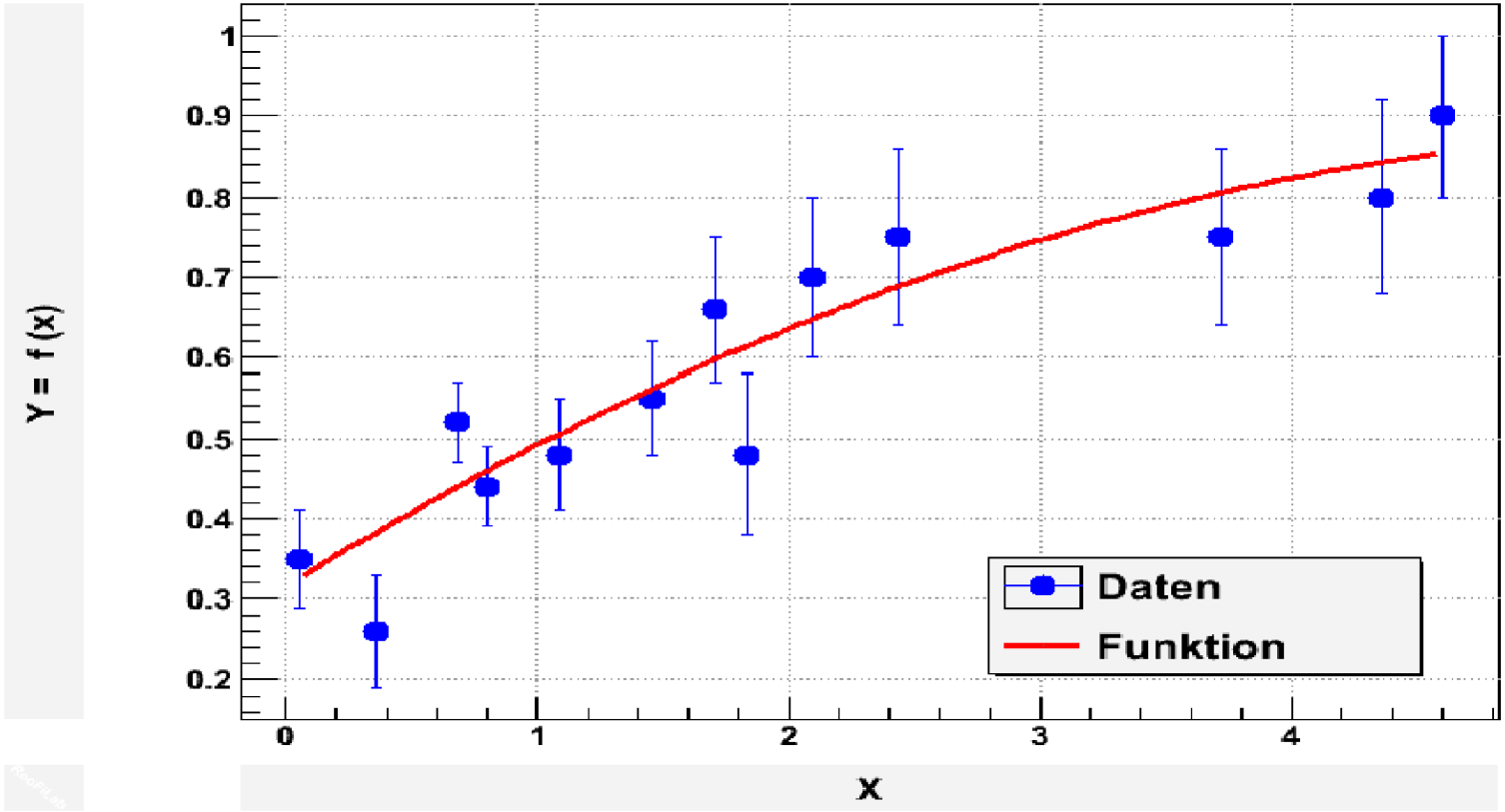


# Themen, Volesungsprogramm

- Einführung: Überblick und grundlegende Konzepte (1)
- Einführung: Überblick und grundlegende Konzepte (2)
- Zufallszahlen und Monte-Carlo Methoden
- **Hypothesentests (1)**
- **Parameterschätzung** ← Heute
- **Parameterschätzung (Goodness-of-fit)**
- **Optimierungs- und Parametrisierungsmethoden**
- Konfidenzintervalle
- Hypothesentests (2)
- Klassifikation
- Klassifikation
- Klassifikation
- Messen und Entfalten
- Systematische Unsicherheiten

# Schätzung von Parametern




**Anpassung** von *Modellen* = *parameterabhängige Funktionen*
an *statistische Daten* = *Messwerte*

# Quantitative Wissenschaft

## Messung von Parametern

Gemessene Werte weichen durch (statistische und systematische) Messfehler vom wahren Wert des Parameters ab.

➔ Beobachtungen (Daten) sind **Stichproben** aus einer Verteilung, die durch einen Parameter festgelegt ist.

**Schätzung**: Prozedur zur Bestimmung eines Parameterwertes (und seines Fehlers) aus zufallsverteilten Daten

Beispiele:

- Messung der mittleren Lebensdauer $\equiv$ eines atomaren Zustands aus N Messungen von Zerfallszeiten $t_i$
- Messung einer Ereignisrate k aus N Messungen von Ereignishäufigkeiten

# Was sind Schätzer?

**Schätzung der mittleren Größe von StudentInnen**

**Daten**: Größen von N (repräsentativ ausgewählten) StudentInnen

1) Alle Größen addieren und durch N teilen
2) Nur die ersten 10 mitteln, den Rest verwerfen
3) Alle Größen addieren und durch N-1 teilen
4) Alle Daten ignorieren und 1.8 m nehmen
5) Alle Größen multiplizieren und N-te Wurzel ziehen
6) Die am häufigsten auftretenden Größen mitteln
7) Kleinste und größte Größe addieren und durch 2 teilen
8) Nur jede zweite Größe nehmen und mitteln

- Alle Methoden sind Schätzer
- Aber welche Methoden sind sinnvoll? Welche ist die Beste?

# Eigenschaften von Schätzern

Bezeichnung des Schätzers für einen Parameter a: $\hat{a}$

$\hat{a}$ ist Funktion von Zufallsvariablen, also selbst eine Zufallsvariable !

Wahrer Wert: $a_0$

- Konsistenz: $\lim_{N \to \infty} \hat{a} = a_0$
- Erwartungstreue: $E[\hat{a}] = a_0$ Bias $b = E[\hat{a}] - a_0$
- Effizienz: Die Varianz $V[\hat{a}]$ soll minimal sein
- Robustheit: Nicht sensitiv auf Ausreißer und Fehler im Modell

➔ Eigenschaften eines Schätzers hängen von der Verteilung ab.

# Aufgabe: Zeigen Sie, dass 3. verzerrt ist

(Alle Größen addieren und durch N-1 teilen)

$$\hat{a} = \frac{1}{N-1} \sum_{i=1}^{N} x_i$$

$$E[\hat{a}] = E\left[\frac{1}{N-1} \sum_{i=1}^{N} x_i\right] = \frac{1}{N-1} E\left[\sum_{i=1}^{N} x_i\right] = \frac{1}{N-1} \sum_{i=1}^{N} E[x_i]$$

$$= \frac{N}{N-1} \underbrace{E[x_i]}_{\langle x \rangle = a_0} = \frac{N}{N-1} a_0 \neq a_0$$

Bias: $E[\hat{a}] - a_0 = a_0 \frac{1}{N-1} = b$

$$\lim_{N \to \infty} b = 0$$

# Beispiel: Schätzung von Mittelwert und Varianz (schon bekannt)

**Mittelwert**:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

- Konsistent und erwartungstreu
- Effizienz und Robustheit hängt von der Verteilung ab

**Varianz, Mittelwert bekannt:**

$$\hat{V} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

**Varianz, Mittelwert unbekannt:**

$$\hat{V} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \hat{\mu})^2$$

- Verzerrt

$$\hat{V} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \hat{\mu})^2 \qquad b = 0$$

- **Bessel-Korrektur**

(siehe S. Brandt S. 133)

# Gewichtetes Mittel

- Mehrere Messungen, mit **unterschiedlicher** Genauigkeit

$N$ Messungen

$x_1 \pm \sigma_1$ , $x_2 \pm \sigma_2$ ....

- **Arithmetisches Mittel**

$$\hat{\mu} = \frac{1}{N} \sum x_i$$

- **Gewichtetes Mittel**

$$\hat{\mu} = \frac{1}{\sum_i \frac{1}{\sigma_i^2}} \sum_i \frac{1}{\sigma_i^2} x_i$$

# Likelihood

- Wahrscheinlichkeitsdichte $\int f(x|\vec{a})\,dx = 1$

- Stichprobe (Messung)

  N Messungen $x_i$ mit $i = 1 \ldots N$

- Likelihood $\mathcal{L}(X|\vec{a}) = f(x_1|\vec{a}) \cdot f(x_2|\vec{a}) \cdots f(x_N|\vec{a}) = \prod_{i=1}^{N} f(x_i|\vec{a})$

- Keine Wahrscheinlichkeitsdichte in a

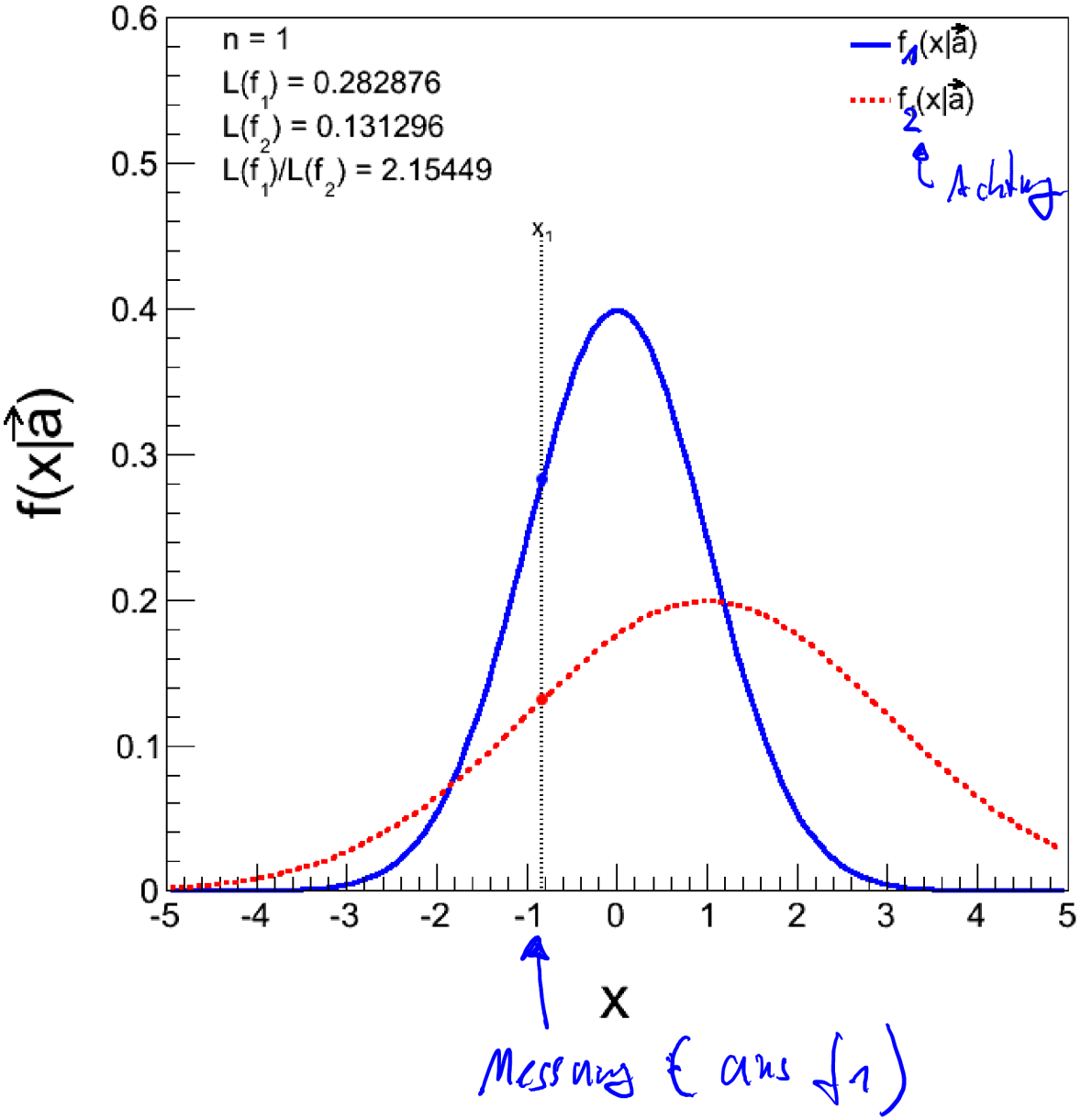
  $\int \mathcal{L}(\vec{a})\,d\vec{a} \neq 1$

- Wahrscheinlichkeitsdichte in x

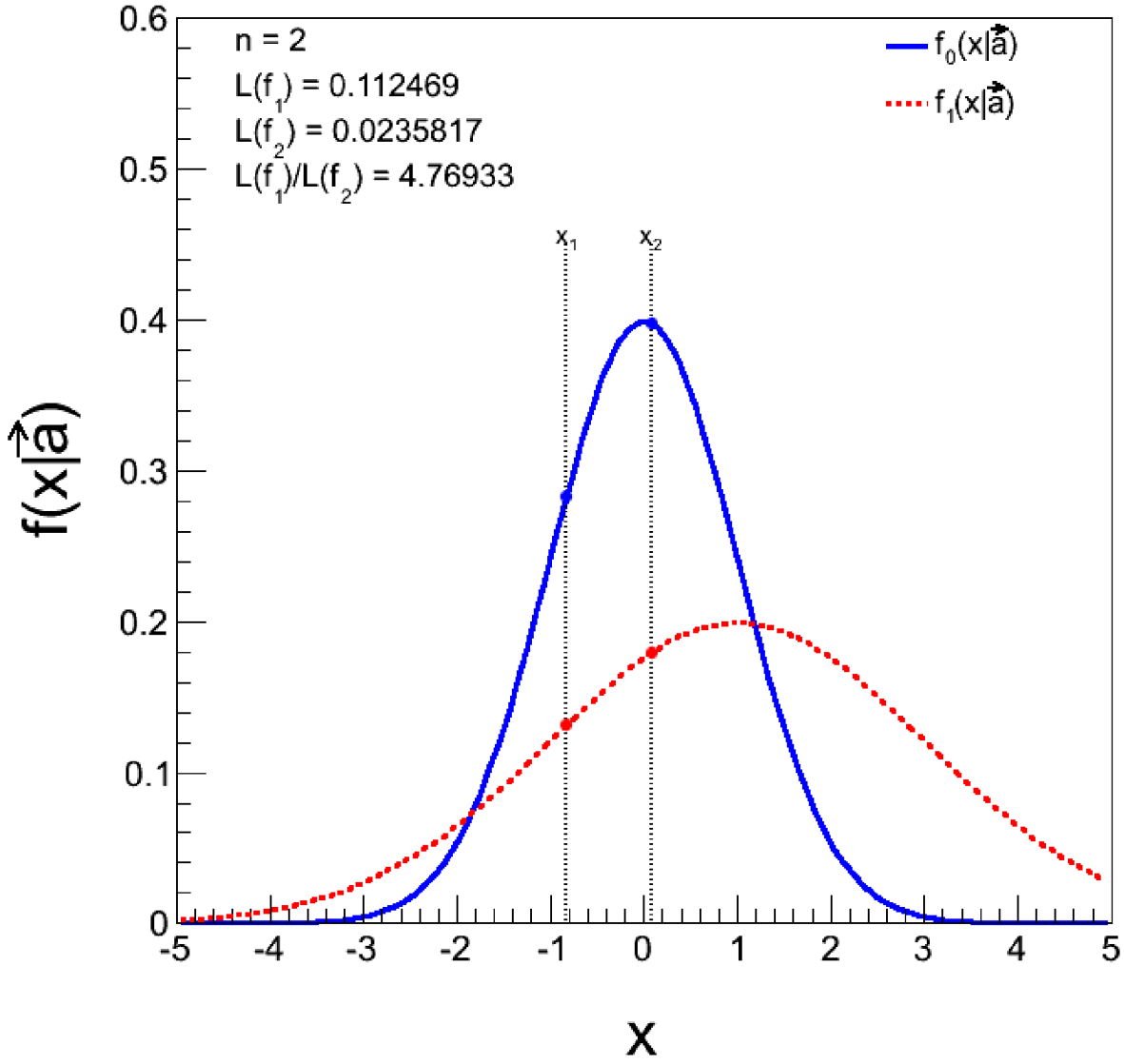
  $\int \mathcal{L}(X|\vec{a})\,dX = \int f(x_1|\vec{a}) \cdot f(x_2|\vec{a}) \cdots f(x_N|\vec{a}) \prod_{i=1}^{N} dx_i = 1$

- Erwartungswert: $E[\alpha] = \int \alpha\, \mathcal{L}(X|\vec{a})\,dX$

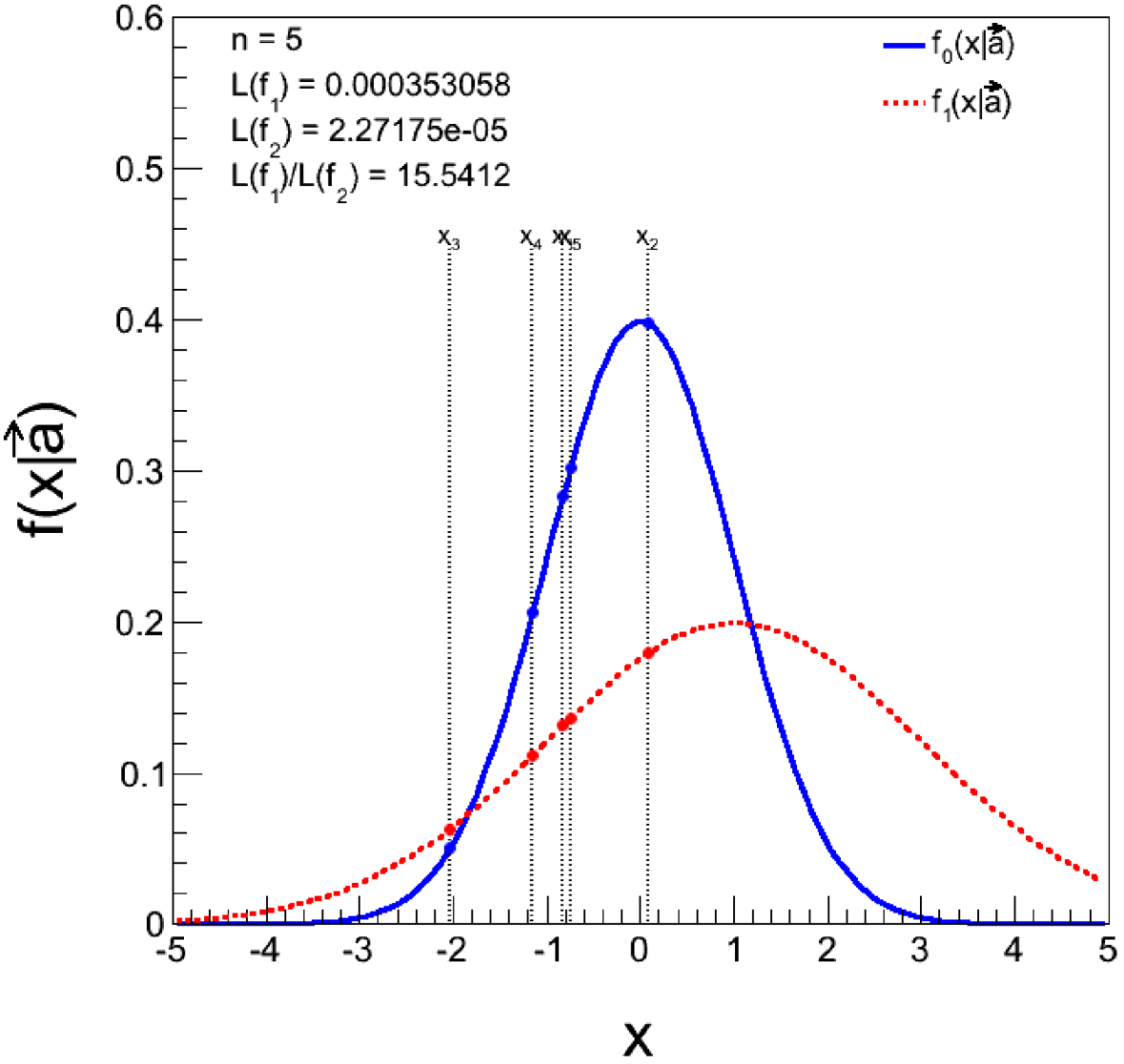
# Beispiel, Likelihood



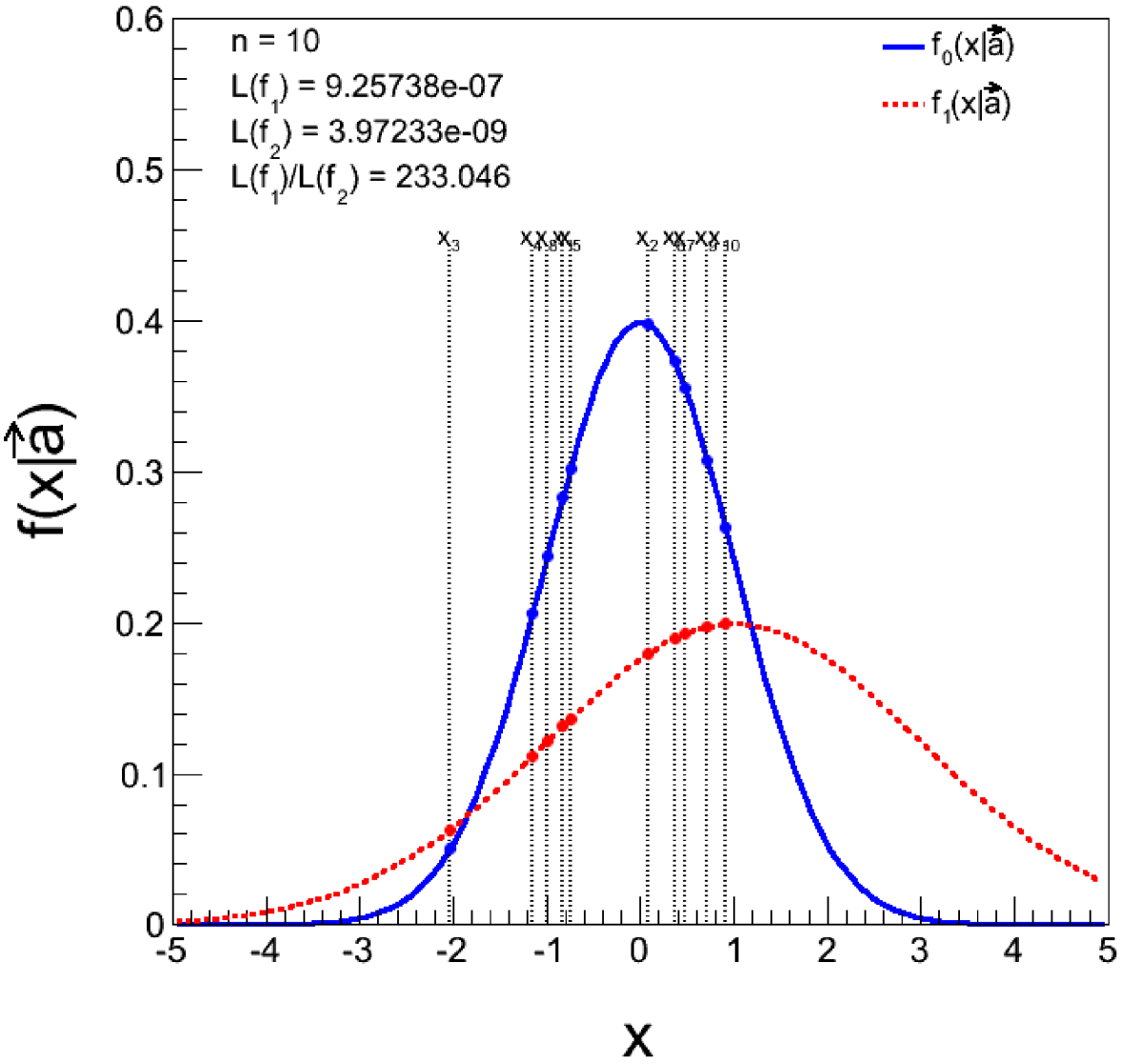
# Beispiel, Likelihood



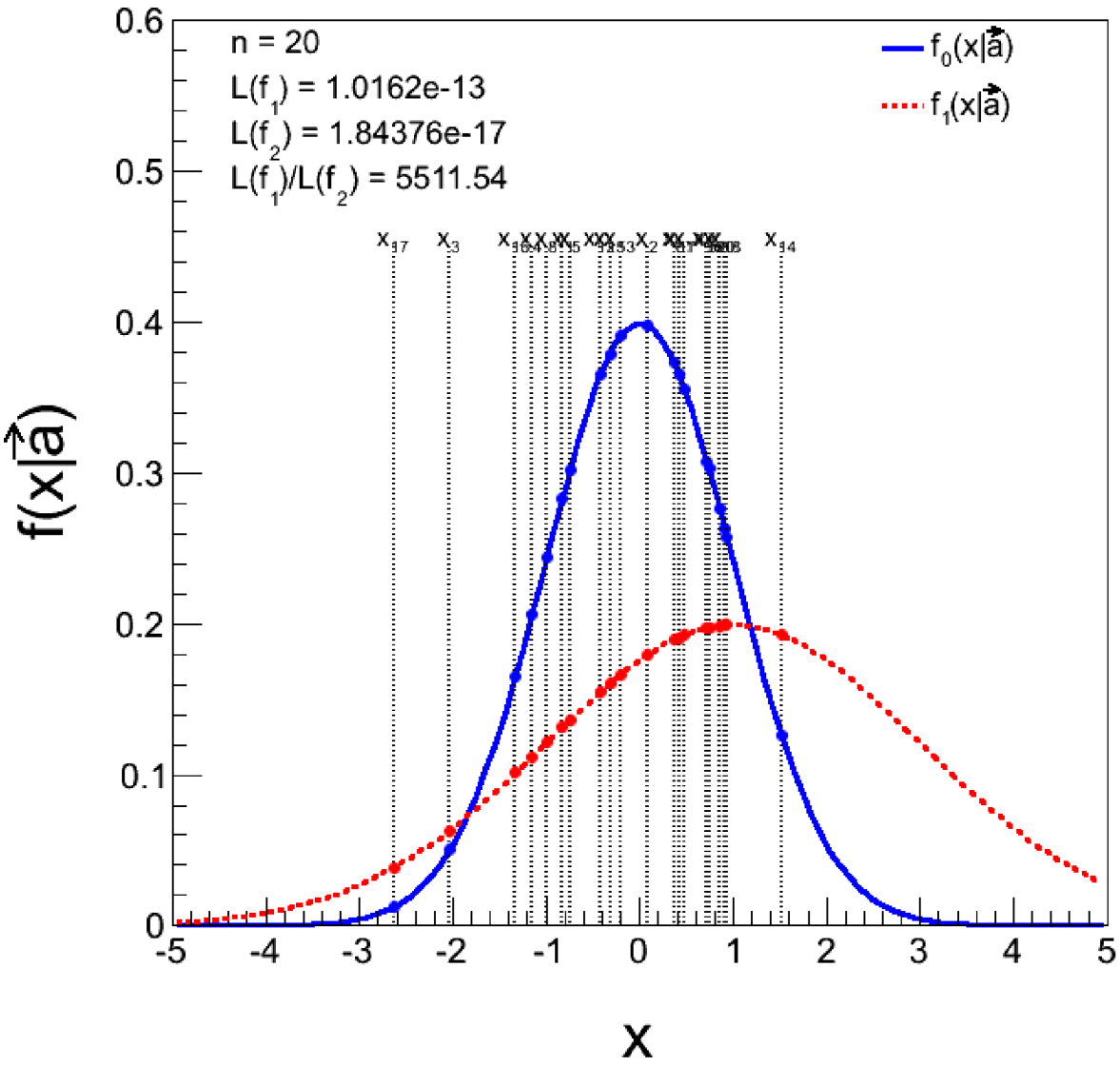
# Beispiel, Likelihood



# Beispiel, Likelihood

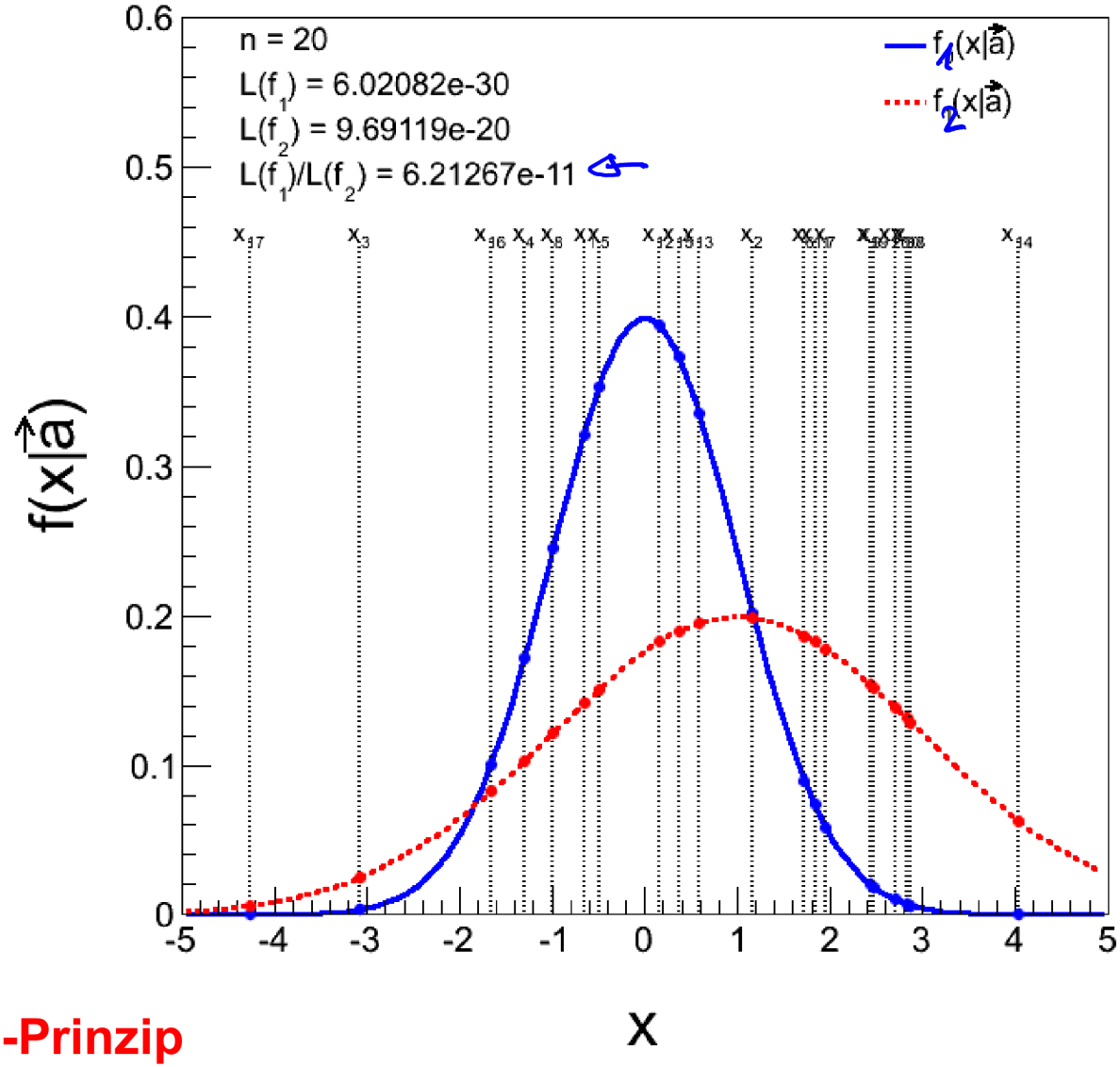


# Beispiel, Likelihood



# Beispiel, Likelihood

Messung aus f2




**Maximum-Likelihood-Prinzip**

Der beste Schätzwert für den Parametervektor **a** maximiert die Likelihood Funktion.

# Likelihood Schätzung

$\hat{a}$: $\mathcal{L}$ soll maximiert werden

$$\Rightarrow \frac{d\mathcal{L}}{da} = 0$$

Aus technischen und nummerischen Gründen

$\hat{a}$: $-\ln \mathcal{L}$ minimiert

$$\Rightarrow \frac{d \ln \mathcal{L}}{da} = 0$$

$$-\ln \mathcal{L} = -\ln \left( \prod_i f(x_i | \vec{a}) \right) = -\sum_i \ln f(x_i | \vec{a})$$

kombiniert man un-korrelierte Messungen / Experimente:

$$\mathcal{L} = \mathcal{L}_1 \cdot \mathcal{L}_2 \quad \Rightarrow \quad -\ln \mathcal{L} = -\ln \mathcal{L}_1 - \ln \mathcal{L}_2$$

(Produkt) (Summe)

### ⋆5.3.5 Notes on Maximum Likelihood

The principle of maximum likelihood is not a rule that requires justification —it does not need to be proved. It is merely a sensible way of producing an estimator. But although the name 'maximum likelihood' has a nice ring to it—it suggest that your estimate is the 'most likely' value—this is an unfair interpretation; it is the estimate that makes your data most likely—another thing altogether.

For large samples, $\hat{a}$ has a probability distribution that is unbiased and normally distributed about the true value $a$, with variance equal to the minimum variance bound. You cannot ask for better than that. So in the large $N$ limit (sometimes referred to as the 'asymptotic limit') ML can indeed claim to be the 'best' estimator, though for smaller samples this is not necessarily true.

It has other obvious advantages. There is no loss of information through binning; all the experimental information is used. It is very suitable for problems where several variables are to be estimated. The method readily gives the errors on its estimate: the $1\sigma$ points are those where the log likelihood falls by 0.5 from its peak value.

Even so, it is not all roses. For small $N$, ML estimators are (usually) biased. You have been warned. And you have to be careful not to apply large $N$ formulae to small $N$ situations.

You have to know the form of the parent distribution function. If your assumptions about $P(x; a)$ are wrong then there is no way of telling this from the results of the fit; it does not give you any quality factor or goodness-of-fit number.

Sometimes the differentiation of $\ln L$ gives you an equation you can solve analytically. (When differentiating the likelihood, do not forget the normalisation factors.) Sometimes you cannot, and it has to be done numerically and/or graphically.

This probably requires you to use a computer and code your likelihood function as a program. If your computer complains about non-existent logarithms of negative numbers, your form for the likelihood function has gone negative and is unphysical, so think again.

*aus: Barlow*

# Messung einer Zerfallskonstanten

- Analytisches Modell: $f(x|\tau) = \frac{1}{\tau} e^{-t/\tau}$
- $N$ Messwerte der Verteilung: $x_i$ mit $i = 1 \ldots n$
- Gesamte Likelihood: $\mathcal{L} = \prod_{i=1}^{N} f(t_i|\tau) = \prod_i \frac{1}{\tau} e^{-t_i/\tau}$

$$-\ln \mathcal{L} = -\ln \prod_i \frac{1}{\tau} e^{-t_i/\tau} = -\sum_i \ln \frac{1}{\tau} e^{-t_i/\tau}$$

$$= -\sum_i \ln \frac{1}{\tau} - \frac{x_i}{\tau}$$

# Bester Parameter:

$\mathcal{L}(\tau)$ ist maximal $\rightarrow$ $-\ln \mathcal{L}(\tau)$ ist minimal

$$\frac{d \ln \mathcal{L}(\tau)}{d\tau} = 0 = \sum_i \frac{1}{\tau} - \frac{x_i}{\tau^2} = N\frac{1}{\tau} - \sum_i \frac{t_i}{\tau^2}$$

$$\frac{n}{\tau} - \frac{1}{\tau^2}\sum_i t_i = 0$$

$$\hat{\tau} = \frac{1}{N}\sum_i t_i$$

- Estimator: $\hat{t} = \langle t \rangle$
- Analytisches Ergebnis
- Aber: Guter Schätzer?

# Bias, Erwartungstreue

- Erwartungswert des Schätzers

$$E[\hat{\tau}] = E\left[\frac{1}{N}\sum_i t_i\right] = \frac{1}{N} E\left[\sum_i t_i\right] = \frac{1}{N}\sum_i \underbrace{E[t_i]}_{\langle t \rangle}$$

$$= \frac{1}{N}\langle t \rangle = \tau$$

- Schätzer hat **keinen Bias**

$$E[\hat{\tau}] - \tau = 0 \simeq b$$

# Unsicherheit

- Varianz

$$V[\hat{\tau}] = V\left[\frac{1}{N}\sum_i t_i\right]$$

$$= \int \cdots \int \left(\frac{1}{N}\sum_i t_i\right)^2 \mathcal{L}(\tau) \prod_i dt_i$$

$$- \left(\int \cdots \int \frac{1}{N}\sum x_i \mathcal{L}(\tau) \prod_i dt_i\right)^2$$

$$= \frac{\tau^2}{N}$$

- Standardabweichung $\sigma = \frac{\tau}{\sqrt{N}}$
- Schätzer ist **effizient** $V \propto \frac{1}{N}$

# Die Cramér-Rao-Grenze für Varianzen

- Für die Varianz eines ML Schätzers gilt

$$V[a] \geq \left(1 + \frac{\partial(E[\hat{a}] - a_0)}{\partial a}\bigg|_{a=\hat{a}}\right) \cdot \underbrace{E\left[-\frac{\partial^2 \ln \mathcal{L}}{\partial^2 a}\right]^{-1}}_{\text{Information}}$$

*(siehe z.B. S. Brandt, Datenanalyse)*

- Nur in Spezialfällen analytisch ausrechenbar, z.B. Exponentialgesetz
- In der Praxis wird meistens $b = E[\hat{a}] - a = 0$ angenommen
- Für effiziente Schätzer gilt die Exaktheit $\text{"}\geq\text{"} \quad \rightarrow \quad \text{"}=\text{"}$
- ML Schätzer für große N sind effizient

# Anwendung Cramér-Rao

(Herleitung Barlow 5.1.3)

- Mehrere Dimensionen: **a**=($a_1$,...,$a_n$)
- Kein Bias: b=0
- Inverse Kovarianz-Matrix: $(V^{-1})_{ij} = -E\left[\left.\frac{\partial^2 \ln \mathcal{L}}{\partial a_i \partial a_j}\right|_{\hat{a}_i \hat{a}_j}\right]$

$$(V^{-1})_{ij} = \int \cdots \int -\frac{\partial^2 \ln \mathcal{L}}{\partial a_i a_j} \mathcal{L} \prod_{i=1}^{N} dx_i$$

$$= -N \cdot \int f(x|\vec{a}) \frac{\partial^2}{\partial a_i \partial a_j} \ln f(x|\vec{a})\, dx$$

$$\propto N \qquad \Rightarrow V_{ij} \propto \frac{1}{N}$$

- Numerisch Approximiert, z.B. in Minuit: MIGRAD, HESSE

exakt für effiziente Schätzer ohne Bias!

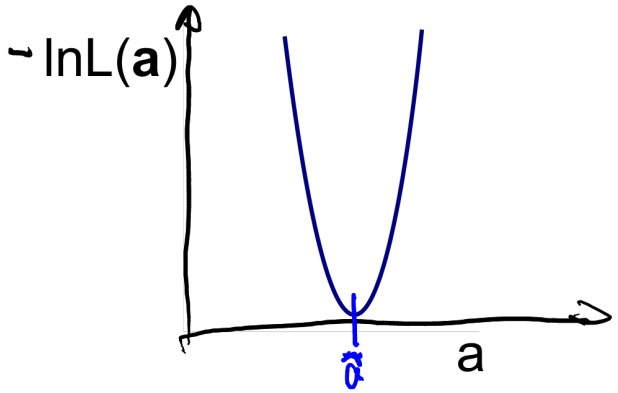
$$\hat{V}^{-1}_{ij} = -\frac{\partial^2 \ln \mathcal{L}}{\partial a_i \partial a_j}$$

→ mit nur einem Parameter:

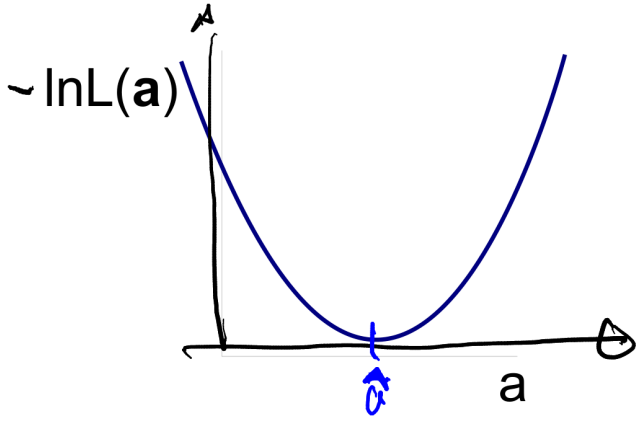
$$\frac{1}{V} = \frac{1}{\sigma^2} = -\frac{\partial^2 \ln \mathcal{L}}{\partial a^2}$$

# Parameterfehler, Intuitiv

Je schärfer das Minimum von lnL(**a**) desto kleiner die Parameterunschärfe




Große Krümmung: kleine Unsicherheit




Kleine Krümmung: große Unsicherheit

Fehler der Parameter a sind umgekehrt proportional zur Krümmung von lnL(a) am Minimum

$$\frac{1}{\sigma_a^2} = V_{\hat{a}}^{-1} = -\left.\frac{\partial^2 \ln \mathcal{L}(a)}{\partial a^2}\right|_{\hat{a}} \qquad \left(V_{\hat{a}}^{-1}\right)_{ij} = -\left.\frac{\partial^2 \ln \mathcal{L}(\vec{a})}{\partial a_i \partial a_j}\right|_{\hat{a_i},\hat{a_j}}$$

# Fehler des ML Schätzers, graphisch

Taylorentwicklung von ML Funktion am Minimum

$$-\ln \mathcal{L}(a) = -\ln \mathcal{L}(\hat{a}) + (a-\hat{a}) \underbrace{\frac{d-\ln \mathcal{L}(a)}{da}\Big|_{a=\hat{a}}}_{\doteq 0} + (a-\hat{a})^2 \cdot \frac{d^2-\ln \mathcal{L}(a)}{2\,da^2} + \ldots$$

$$\frac{d \ln \mathcal{L}}{da}\Big|_{\hat{a}} = 0$$

$$\approx -\ln \mathcal{L}(\hat{a}) + \frac{1}{2}(a-\hat{a})^2 \frac{d-\ln \mathcal{L}(\hat{a})}{\partial a^2}$$

Falls höhere Terme vernachlässigbar sind (lnL(a) parabelförmig)

$$\mathcal{L}(a) = e^{\ln \mathcal{L}(a)} = k \exp\left[-\frac{1}{2}(a-\hat{a})^2 \frac{d^2 \ln \mathcal{L}(a)}{da^2}\right]$$

Gauss-Verteilung mit

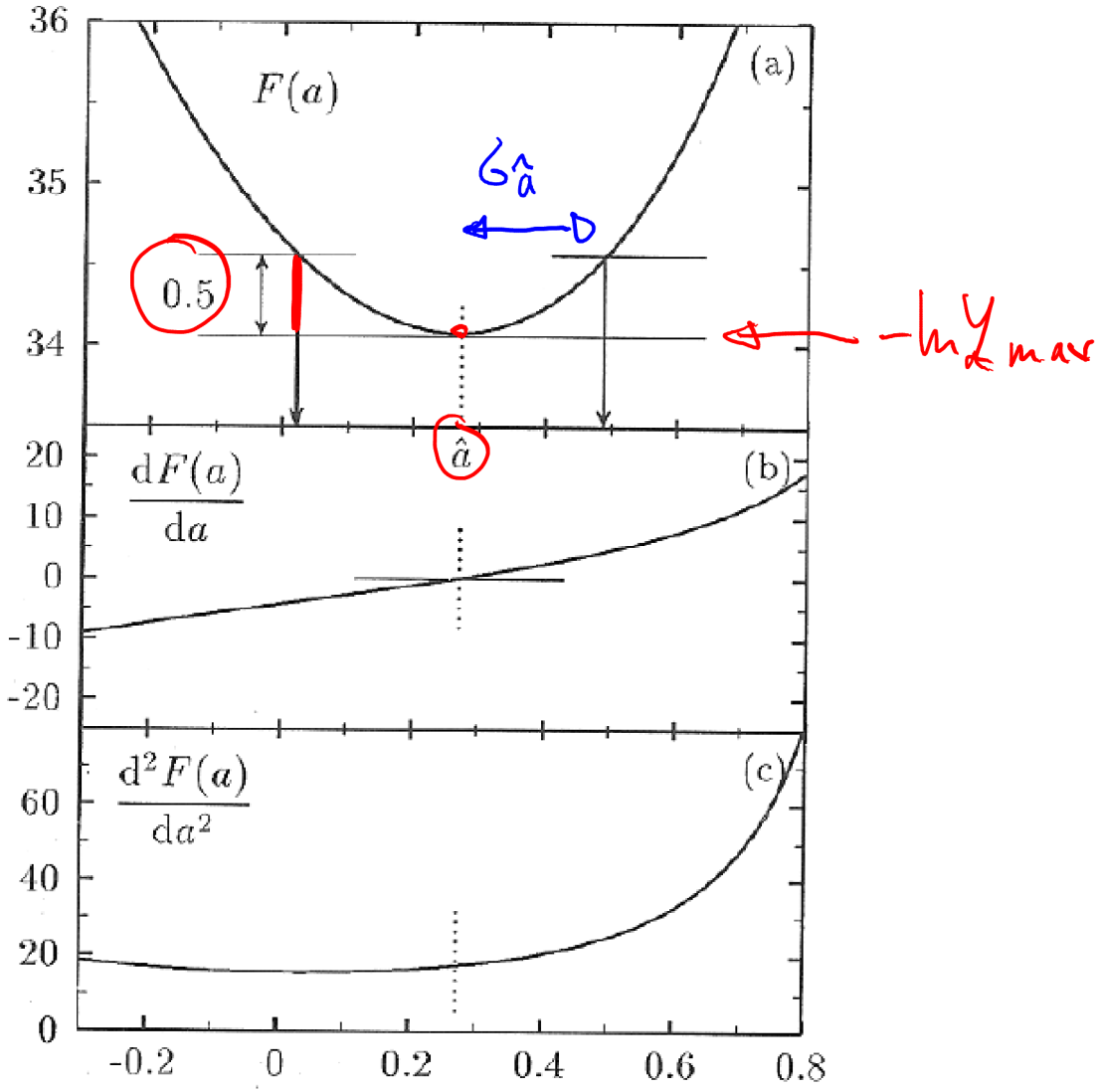
$$\mu = \hat{a}$$

und

$$\sigma = \left.\frac{d^2 \ln \mathcal{L}(a)}{2a^2}\right|^{-1}_{a=\hat{a}}$$

# Fehler des ML Schätzers, graphisch

$$\ln \mathcal{L}(\hat{a} \pm \widehat{\sigma_a}) = \ln \mathcal{L}_{\max} - 0.5$$

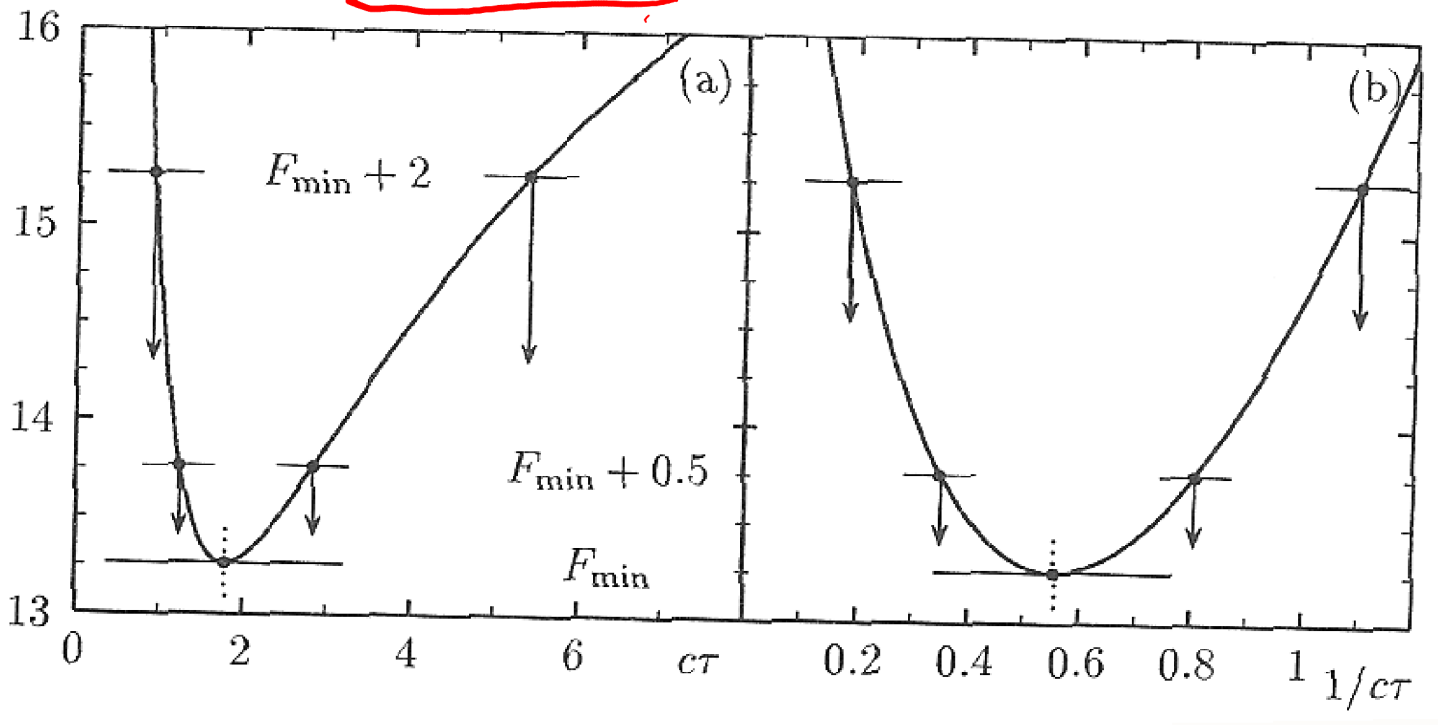


# Fehler bei nicht parabelförmigen F(a)

F(a) parabelförmig: $\ln \mathcal{L}(\hat{a} \pm \widehat{\sigma_a}) = \ln \mathcal{L}_{\max} - 0.5$

Allgemeine verwendet: $\ln \mathcal{L}(\hat{a} \pm \widehat{\sigma_a}) = \ln \mathcal{L}_{\max} - 0.5n^2$

Fehler können asymmetrisch sein!

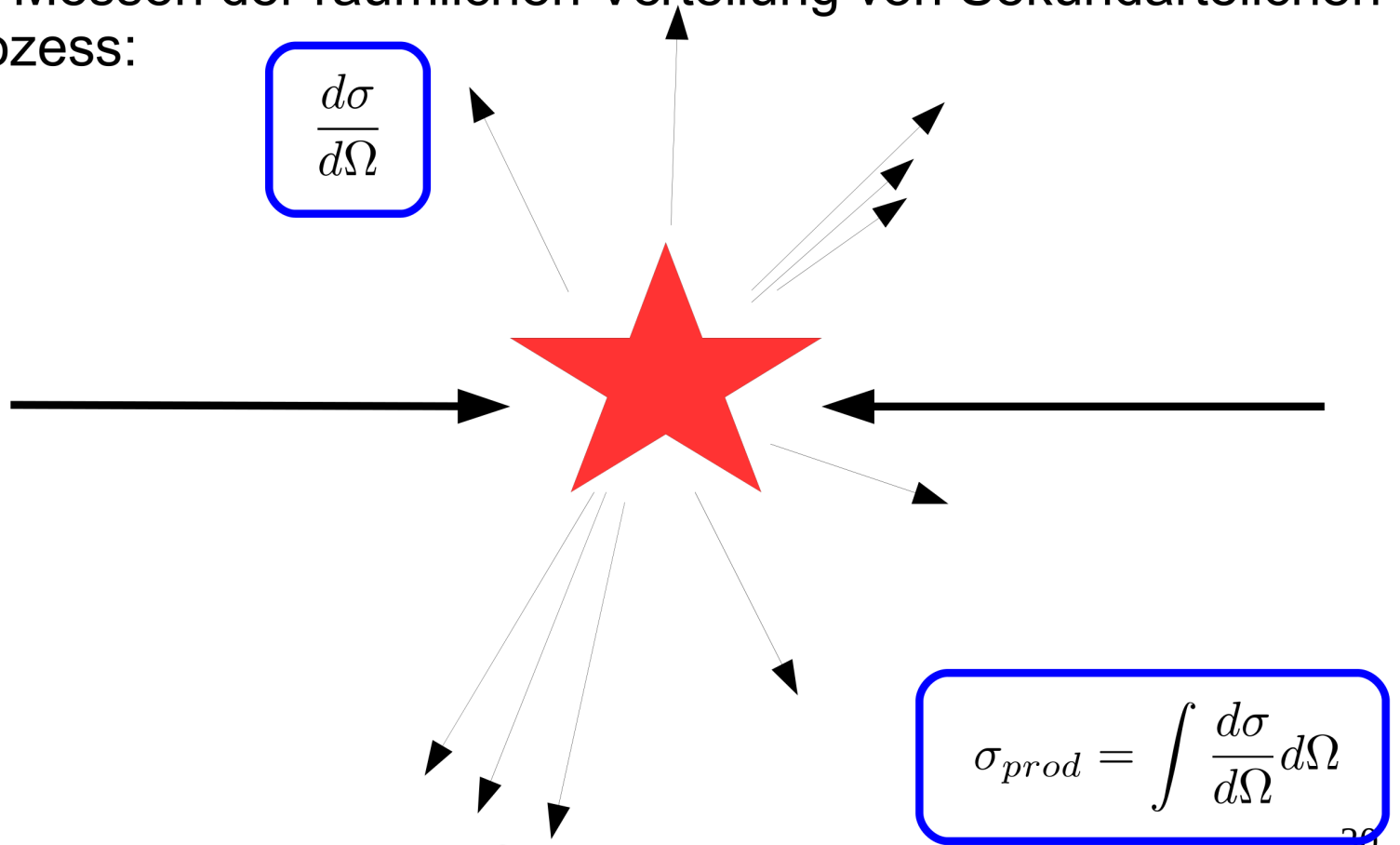





(z.B. in Minuit: MINOS, gilt auch in beliebig vielen Dimensionen)

# Erweiterte Likelihood Methode

- Oft ist auch die **Anzahl der Daten** (→ Normierung) ein relevanter Parameter
- Zum Beispiel: Messen der räumlichen Verteilung von Sekundärteilchen nach Streuprozess:




- Aber auch gesamter Produktions-Querschnitt:

# Erweiterte Likelihood Methode

- Erwarte $\nu$ Ereignisse in der Analyse
- Der Wert von $\nu$ kann bekannt sein ($\rightarrow$ Constraint), oder ein unbekannter Parameter

$$\mathcal{L} = \prod_i \left( f(x_i | \vec{a}) \right) \cdot \frac{\nu^n}{n!} e^{-\nu} = \frac{e^{-\nu}}{n!} \prod_i \nu f(x_i | \vec{a})$$

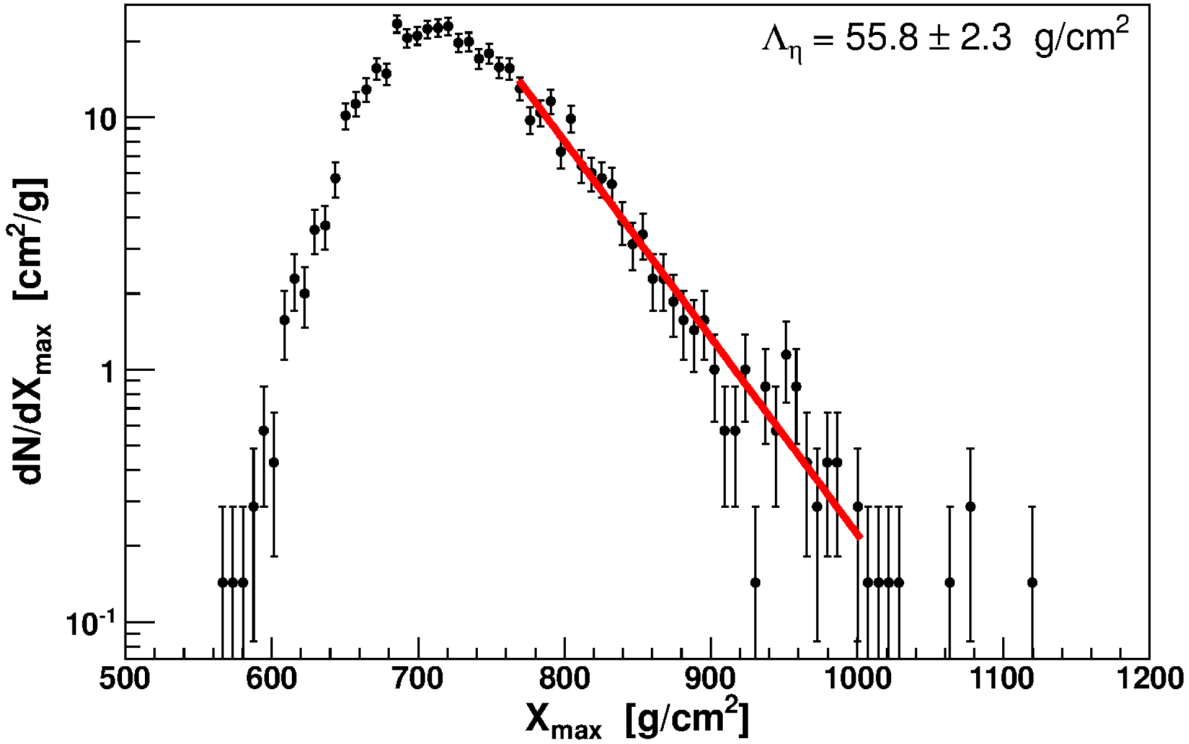
$$\ln \mathcal{L}(\vec{a}) = -\nu + \sum_i \ln \nu f(x_i | \vec{a})$$

vernachlässigt sind alle Terme, die nicht vom Parameter abhängen.

- Falls kein Constraint vorliegt, kann es doch sinnvoll sein, z.B. bei der gemeinsamen Bestimmung der Häufigkeit von Hintergrund und Signal Ereignissen in einer kombinierten Likelihood Verteilung

# Beispiel aus der Praxis: Proton-Luft Wechselwirkungsquerschnitt

- Pierre Auger Observatory, *PRL 109 (2012) 062002*
- Absorption von Luftschauern in der Atmosphäre
- Erweiterte Ungebinnte Log-Likelihood Analyse
- Auf einem Intervall $[X_{Start}, X_{Ende}]$



# Binned Likelihood

- Für große Zahl von Daten kann Likelihood sehr rechenaufwändig sein
- Daten können in N Bins zusammengefasst werden:

  Die erwartete Häufigkeit in Bin i ist dann: $\nu_i(\vec{a}) = n_{\mathrm{tot}} \int_{x_i^{\min}}^{x_i^{\max}} f(x|\vec{a})dx$

- Nennt man $n_i$ die beobachtete Häufigkeit in Bin i, dann ist

  Multinomiale-Verteilung:

  $$f_{\mathrm{joint}}(\vec{n}|\vec{\nu}) = \mathcal{L}(\vec{a}) = \frac{n_{\mathrm{tot}}!}{n_1! \ldots n_N!} \left(\frac{\nu_1}{n_{\mathrm{tot}}}\right)^{n_1} \ldots \left(\frac{\nu_N}{n_{\mathrm{tot}}}\right)^{n_N}$$

  die gesamte Wahrscheinlichkeit aller Bins $\vec{n}$

- Log-Likelihood, unter Vernachlässigung aller konstanten Terme

$$\log \mathcal{L}(\vec{a}) = \sum_{i=1}^{N} n_i \log \nu_i(\vec{a})$$

# Binned Likelihood mit Normierung

- Erweitert man diese Likelihood mit der gesamt Anzahl $n_{tot}$ der Daten sowie der Normierung $\nu_{tot}$

$$n_{\text{tot}} = \sum_i n_i \qquad \nu_{\text{tot}} = \sum_i \nu_i$$

- Erhält man die erweiterte Likelihood

$$f_{\text{joint}}(\vec{n}|\vec{\nu}) = \mathcal{L}(\vec{a}) = \frac{\nu_{\text{tot}}^{n_{\text{tot}}}}{n_{\text{tot}}!} e^{-\nu_{\text{tot}}} \frac{n_{\text{tot}}!}{n_1! \dots n_N!} \left(\frac{\nu_1}{n_{\text{tot}}}\right)^{n_1} \dots \left(\frac{\nu_N}{n_{\text{tot}}}\right)^{n_N}$$

Multinomial-Verteilung ≠ Poisson-Verteilung

Wobei diese dem Produkt der einzelnen Poisson Wahrscheinlichkeiten pro Bin entspricht

$$f_{\text{joint}}(\vec{n}|\vec{\nu}) \overset{!}{=} \mathcal{L}(\vec{a}) = \Pi_i \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$$

Mit der log-Likelihood:

$$\log \mathcal{L}(\vec{a}) = \sum_{i=1}^{N} [-\nu_i + n_i \log \nu_i] = -\nu_{\text{tot}} + \sum_{i=1}^{N} n_i \log \nu_i$$

# Anwendung Likelihood auf einen Blick

- Allgemein:
  - Beurteilung von Messwerten mithilfe beliebiger Modellannahmen (beliebiger Verteilungen)
  - Wenn effiziente Schätzer für Parameter existieren, dann findet die Likelihood Methode diese! Also i.A. gilt

$$\sigma(\hat{a}) \propto \frac{1}{\sqrt{N}}$$

  - Zudem gilt für große N meistens:

$$b \to 0$$

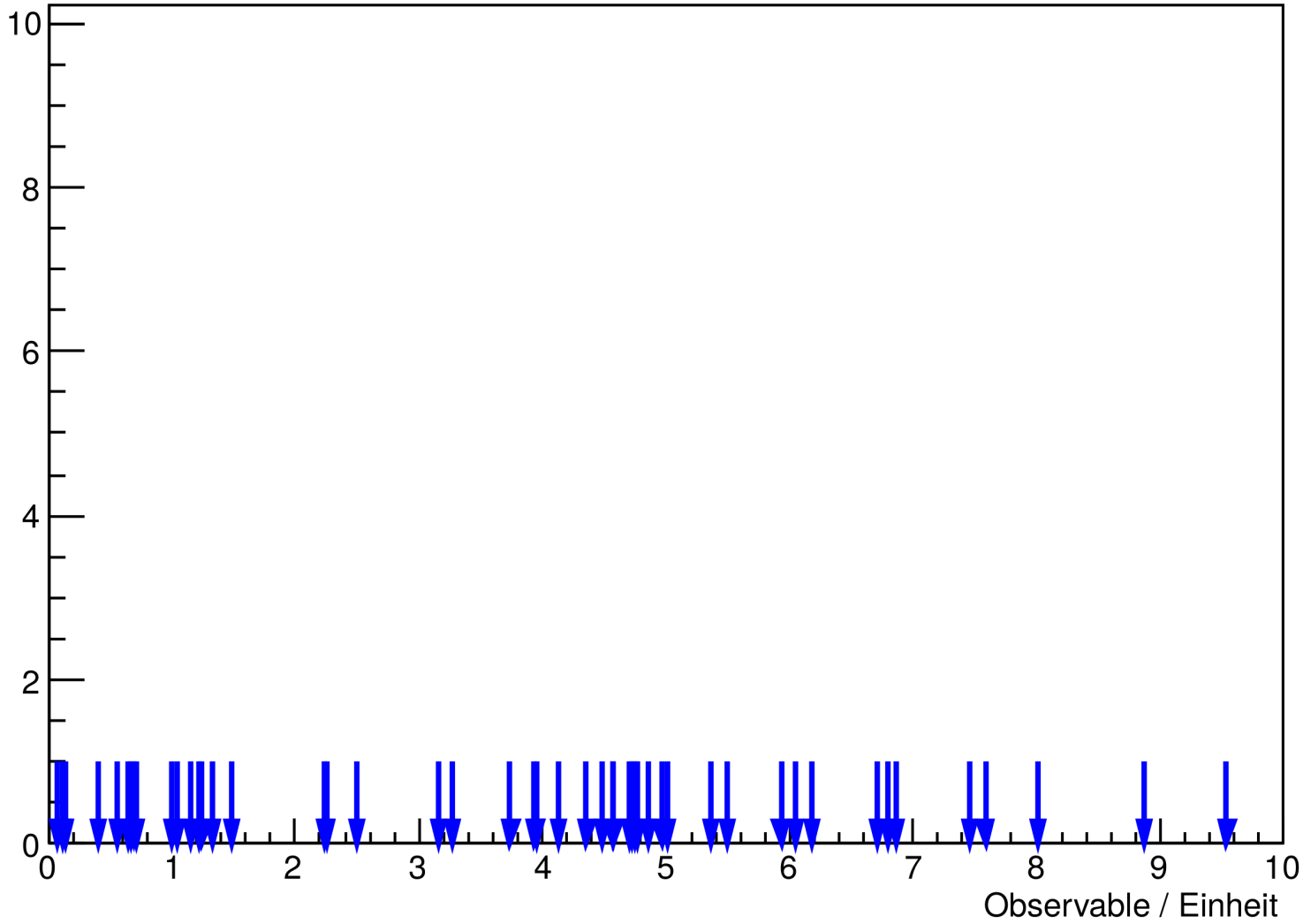
- Histogramme:
  - Anzahl von Daten in Bin: Poisson-Statistik

    → Dies ist die fundamental richtige Verteilung für die Einträge in Bins

  - Histogramm-Likelihood für große N entspricht der „normalen“ Likelihood
  - Histogramm mit Normierung entspricht erweiterter Likelihood Methode

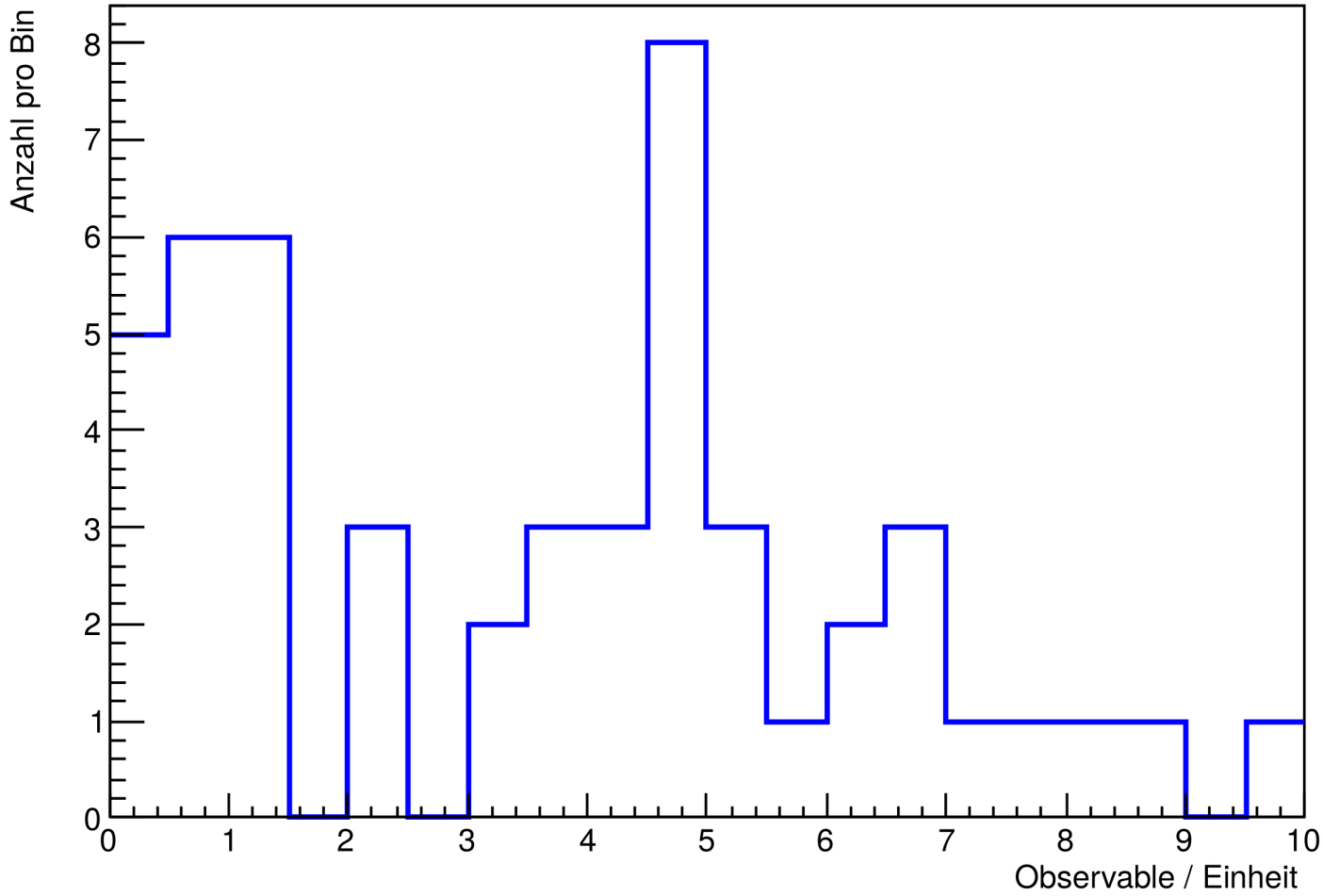
# Zählexperimente, Histogramme

- Typisches Experiment:
  N Messungen einer bestimmten Observablen



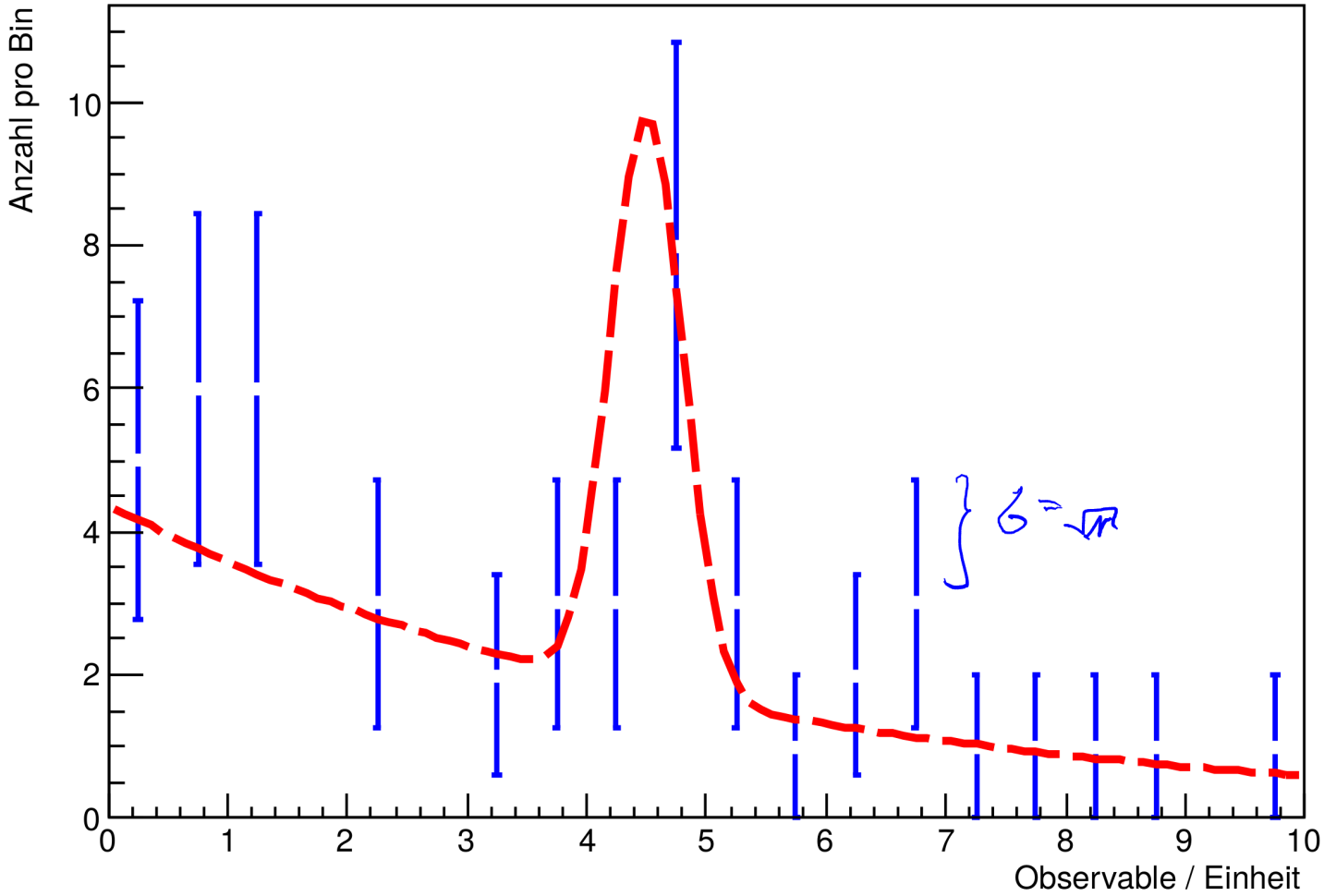
# Zählexperimente, Histogramme

- Histogrammieren der Daten
  Aber: Information geht verloren!



# Zählexperimente, Histogramme

- Histogramm: Mittelwert $\mu$ und Varianz $\mu$ wegen Poisson Statistik




- Fehler NICHT Gauß-förmig !
- Leere Bins !

# Zusammenfassung

- Statistische Schätzer sind die Grundlage der quantitativen Wissenschaft
- Schätzen ist der Prozess des Ableitens von Modellparametern/Parametern aus einer Stichprobe (~Messung)
- Mit der Likelihood Funktion kann die Übereinstimmung einer Messung mit einem Modell beurteilt werden
- Das Verhältnis von Likelihood Funktionen ist ein sehr guter Test von Modellen
- Das Maximum der Likelihood Funktion definiert einen „effektiven“ Schätzer der Parameter der Likelihood Funktion, falls dieser existiert
- Die Krümmung der Likelihood Funktion am Maximum ist ein Schätzer der Varianz der Parameter. Unter der Annahme effektiver Schätzer ohne Bias
- Die Cramer-Rao Grenze gilt

# Beweis Minimum-Variance Bound (Barlow 5.1.3)

$E[\hat{a}] = \int \hat{a}\, \mathcal{L}\, dx = a$ und $\hat{a}$ unbiassed

$\frac{dE[\hat{a}]}{da} = \int \hat{a} \frac{d\mathcal{L}}{da} dx \stackrel{!}{=} 1$ oder I $\int \hat{a} \frac{d\ln\mathcal{L}}{da} \mathcal{L}\, dx = 1$

also $\int \mathcal{L}\, dx = 1$ $\Rightarrow$ II $\int \frac{d\ln\mathcal{L}}{da} \mathcal{L}\, dx = 0 = E\left[\frac{d\ln\mathcal{L}}{da}\right]$

multiply by $a$, and subtract I-II

$\Rightarrow \int (\hat{a}-a) \frac{d\ln\mathcal{L}}{da} \mathcal{L}\, dx = 1$

mit Schwartz-Ungleichung:

$$\underbrace{\left(\int (\hat{a}-a)^2 \mathcal{L}\, dx\right)}_{V[\hat{a}]} \underbrace{\left(\int \left(\frac{d\ln\mathcal{L}}{da}\right)^2 \mathcal{L}\, dx\right)}_{E\left[\left(\frac{d\ln\mathcal{L}}{da}\right)^2\right]} \geq 1 = \left(1 + \left.\frac{db}{da}\right|_{a=\hat{a}}\right)^2$$

mit Bias $b \neq 0$

Definiere „Information": $I = E\left[\left(\frac{d\ln\mathcal{L}}{da}\right)^2\right]$ oder auch $I = -E\left[\frac{d^2\ln\mathcal{L}}{da^2}\right]$