

# Statistische Methoden der Datenanalyse

## Wintersemester 2011/2012

### Albert-Ludwigs-Universität Freiburg



Prof. Markus Schumacher

Physikalisches Institut Westbau 2 OG Raum 008

Telefonnummer 07621 203 7612

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

Vorlesung basiert in weiten Teilen auf der von Glen Cowan.

[http://terascale.physik.uni-freiburg.de/lehre/ws\\_1112/statmethoden\\_ws1112](http://terascale.physik.uni-freiburg.de/lehre/ws_1112/statmethoden_ws1112)

# Eigenschaften von ML-Schätzern

- Konsistenz:** Wenn Erwartungswert und Varianz des Schätzers endlich und Grundgesamtheit unabhängig vom Parameter, dann ist der ML-Schätzer konsistent.
- Erwartungstreue:** Keine allgemeine Aussage. Muss für jeden Schätzer untersucht werden.  
Selten analytisch. Meist mit MC-Methode.
- Effizienz:** Wenn es einen effizienten Schätzer für den Parameter gibt, dann wird er durch die ML-Methode gegeben.
- Wenn Grundgesamtheit beschrieben wird durch
- $$f(x; \theta) = \exp(B(\theta)C(x) + D(\theta) + E(x))$$
- und Grundgesamtheit unabhängig von Parameter, dann gibt es einen effizienten Schätzer.

# Asymptotische Eigenschaften von ML-Schätzern

Asymptotisch = Grenzfall unendlich großen Stichprobenumfangs

**Erwartungstreue:** ML-Schätzer ist asymptotisch erwartungstreu, wenn Varianz endlich ist, da er konsistenz ist.

**WDF für ML-Schätzer:** Geht gegen Gauss-WDF.

**Effizienz:** ML-Schätzer wird 100% effizient, d.h. Varianz = SMV wenn Grundgesamtheit unabhängig von Parameter.

**Form der Likelihoodfunktion:**

Likelihoodfunktion geht gegen Gauss-Funktion.  
log-Likelihoodfunktion geht gegen Parabel.

$$\mathcal{L}(\theta) = \mathcal{L}(\hat{\theta}) \exp\left(-\frac{(\hat{\theta} - \theta)^2}{2V[\hat{\theta}]}\right) \quad \log \mathcal{L}(\theta) = \log \mathcal{L}(\hat{\theta}) - \frac{(\hat{\theta} - \theta)^2}{2V[\hat{\theta}]}$$

# Varianz für Schätzer: Analytische Methode

In weniger Fällen kann die Varianz analytisch berechnet werden.

Benötigt Berechnung von:

$$E[\hat{\theta}] = \int \hat{\theta}(x_1, \dots, x_n) f(x_1, \dots, x_n; \theta_0) dx_1 \dots dx_n \equiv \mu(\theta_0)$$

$$V[\hat{\theta}] = \int (\hat{\theta}(x_1, \dots, x_n) - \mu(\theta_0))^2 f(x_1, \dots, x_n; \theta_0) dx_1 \dots dx_n \equiv \sigma_{\theta_0}^2$$

Erwartungswert und Varianz hängen vom wahren Parameterwert ab.

In der Praxis: schätze wahren Wert durch ML-Schätzer.

Für Exponential-WDF kann dies noch hingeschrieben werden und liefert bekanntes Ergebnis.

Allerdings wird diese Methode in der Praxis quasi nicht verwendet.

# Varianz für Schätzer: Monte-Carlo-Methode

Nachdem wir den Schätzwert für den Parameter bestimmt haben müssen wir nun den statistischen Fehler des Schätzers bestimmen, i.e., wie weit die Verteilung der Schätzwerte wäre, wenn wir die gesamte identische Messung sehr oft wiederholen würden.

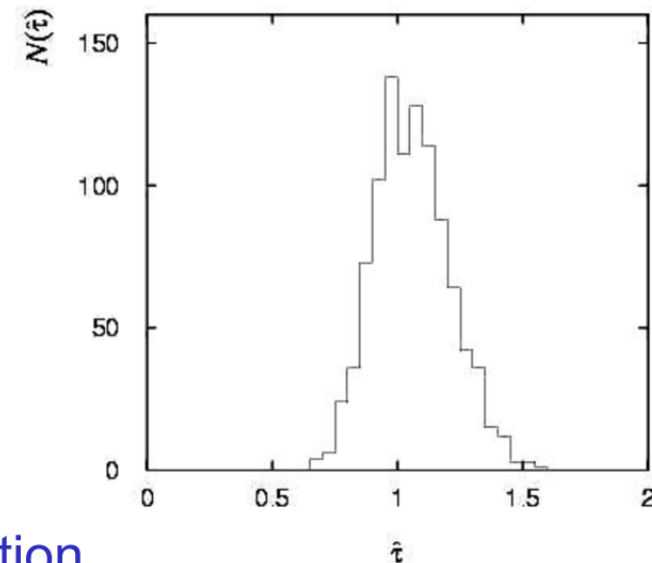
Ein Weg (und der einzig 100% korrekte) dies zu tun, ist die identischen Messungen viele male mit der MC-Methode zu simulieren.

$$\sigma_S^2 = \frac{1}{N-1} \sum_{i=1}^N (\hat{\theta}^{(k)} - \overline{\hat{\theta}^{(k)}})^2$$

Für unser Exponential-WDF-Beispiel erhalten wir aus der Varianz der Schätzwerte:

$$\hat{\sigma}_{\hat{\tau}} = 0.151$$

Beachte: Verteilung ist nahezu gaussförmig. Fast immer wahr für ML-Methode im Grenzfall großer Stichproben.



Wähle Datenschatzwert als Input für MC-Simulation.

Prüfe Abhängigkeit der Varianz von angenommenen Wert in der Simulation.

# Varianz der Schätzer aus SMV

Die Informationsungleichung (RCF) setzt eine untere Schranke auf die Varianz für beliebigen Schätzer (nicht nur ML-Schätzer):

$$V[\hat{\theta}] \geq \left(1 + \frac{\partial b}{\partial \theta}\right)^2 \bigg/ E \left[ -\frac{\partial^2 \ln L}{\partial \theta^2} \right] \quad (b = E[\hat{\theta}] - \theta)$$

Oft ist der Bias  $b$  klein, und die Ungleichung gilt exakt oder ist eine gute Näherung (z.B. im Grenzfall großer Stichproben). Dann:

$$V[\hat{\theta}] \approx -1 \bigg/ E \left[ \frac{\partial^2 \ln L}{\partial \theta^2} \right] \quad (V^{-1})_{ij} = E \left[ -\frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} \right]$$

Wir schätzen diesen Wert durch die 2te Ableitung von  $\ln L$  im Maximum:

$$\hat{V}[\hat{\theta}] = - \left( \frac{\partial^2 \ln L}{\partial \theta^2} \right)^{-1} \bigg|_{\theta = \hat{\theta}} \quad (\widehat{V^{-1}})_{ij} = - \frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} \bigg|_{\vec{\theta} = \hat{\theta}}$$

# Varianz des Schätzers: Graphische Methode

Entwickle  $\ln L(\theta)$  in Taylorreihe um das Maximum:

$$\ln L(\theta) = \ln L(\hat{\theta}) + \left[ \frac{\partial \ln L}{\partial \theta} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2!} \left[ \frac{\partial^2 \ln L}{\partial \theta^2} \right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \dots$$

Erster Term ist  $\ln L_{\max}$ , zweiter Term verschwindet, für den dritten Term verwende die Informationsungleichung: (unter der Annahme der Gleichheit):

$$\ln L(\theta) \approx \ln L_{\max} - \frac{(\theta - \hat{\theta})^2}{2\widehat{\sigma}_{\hat{\theta}}^2}$$

$$\text{i.e.,} \quad \ln L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) \approx \ln L_{\max} - \frac{1}{2}$$

→ um  $\hat{\sigma}_{\hat{\theta}}$  zu erhalten, ändere  $\theta$  von  $\hat{\theta}$  weg, bis  $\ln L$  um  $\frac{1}{2}$  kl. ist als im Max.

$$\text{Grenzen des } k \times \sigma\text{-Intervalls: } \log \mathcal{L} = \log \mathcal{L}_{\max} - \frac{k^2}{2}$$

# Beispiel: WDF-Exponentialfunktion

Graphisch:

$$[1.02 - 0.12, 1.02 + 0.16]$$

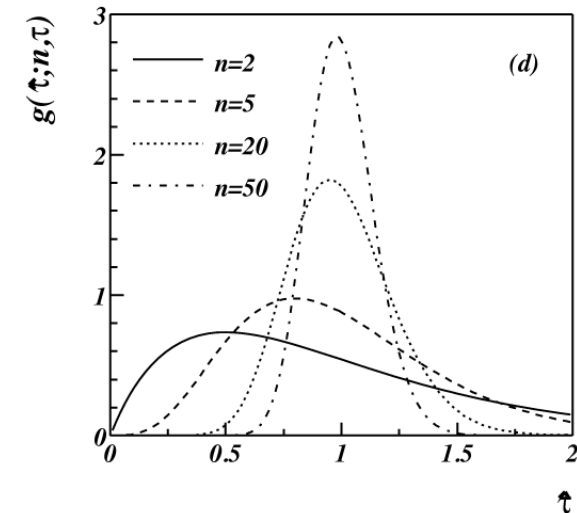
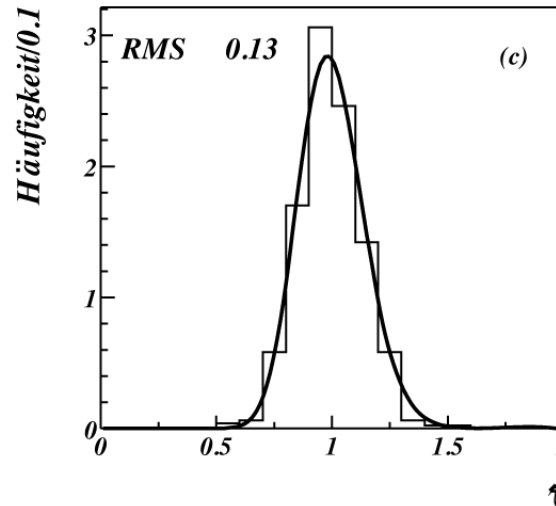
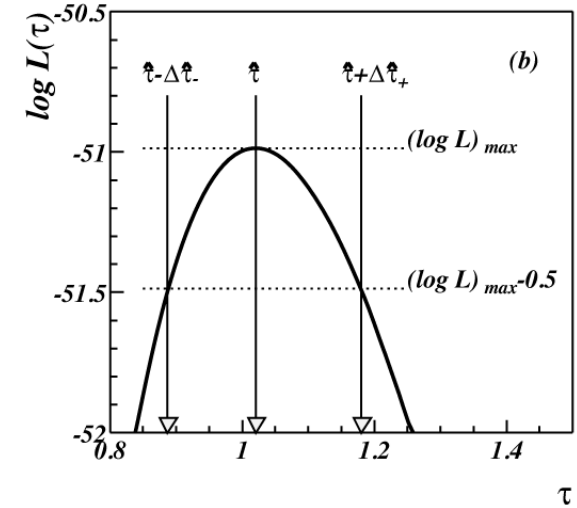
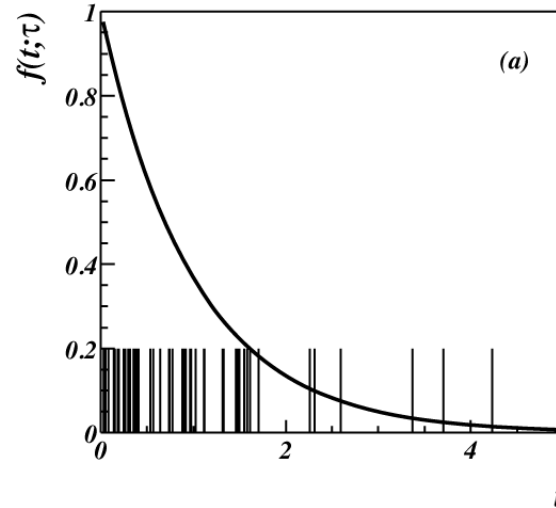
Analytisch:

$$\widehat{V}[\hat{\tau}] = \frac{\hat{\tau}^2}{n} \approx 0.021$$

$$\hat{\sigma} \approx 0.14$$

MC-Methode:

0.13,

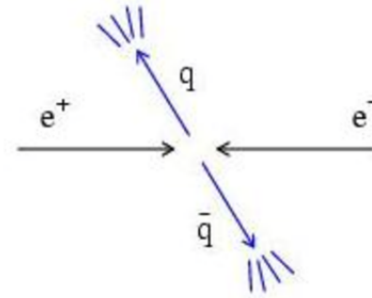




# Beispiel: ML-Methode mit 2 Parametern

Betrachte die Verteilung eines Streuwinkels  $x = \cos \theta$ ,

$$f(x; \alpha, \beta) = \frac{1 + \alpha x + \beta x^2}{2 + 2\beta/3}$$



Oder wenn  $x_{\min} < x < x_{\max}$ , müssen wir korrekt normieren:

$$\int_{x_{\min}}^{x_{\max}} f(x; \alpha, \beta) dx = 1 .$$

Beispiel:  $\alpha = 0.5$ ,  $\beta = 0.5$ ,  $x_{\min} = -0.95$ ,  $x_{\max} = 0.95$ ,  
generiere  $n = 2000$  Messungen/Ereignisse mit der MC-Methode.

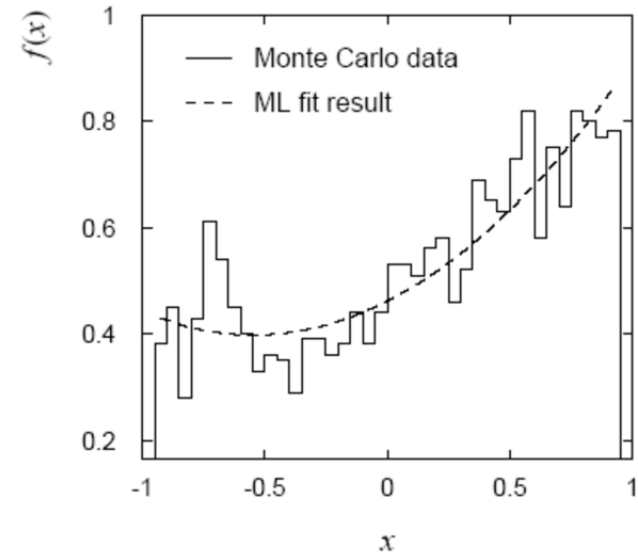
# Beispiel: ML-Methode mit 2 Parametern- Fit-Ergebnis

Numerische Bestimmung des Maximums von  $\ln L(\alpha, \beta)$  (**MINUIT**) liefert

$$\hat{\alpha} = 0.508$$

$$\hat{\beta} = 0.47$$

Bemerkung: Kein Binning der Daten im Fit.  
Histogramm nur für Visualisierung



Kovarianzen aus  $(\widehat{V}^{-1})_{ij} = -\left. \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right|_{\vec{\theta} = \vec{\hat{\theta}}}$  (**MINUIT Routine HESSE**)

$$\hat{\sigma}_{\hat{\alpha}} = 0.052$$

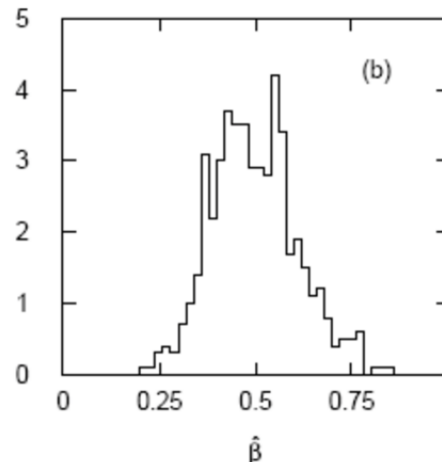
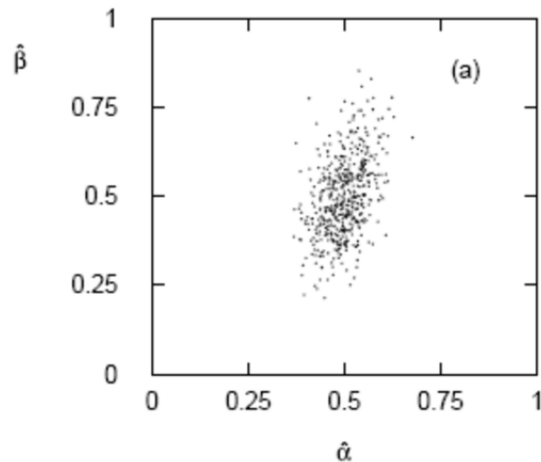
$$\text{cov}[\hat{\alpha}, \hat{\beta}] = 0.0026$$

$$\hat{\sigma}_{\hat{\beta}} = 0.11$$

$$r = 0.46$$

# Zwei-Parameter-Fit: MC-Studie

Wiederhole ML-Fit für 500 Experimente, alle mit jeweils  $n = 2000$  Ereignissen



$$\bar{\hat{\alpha}} = 0.499$$

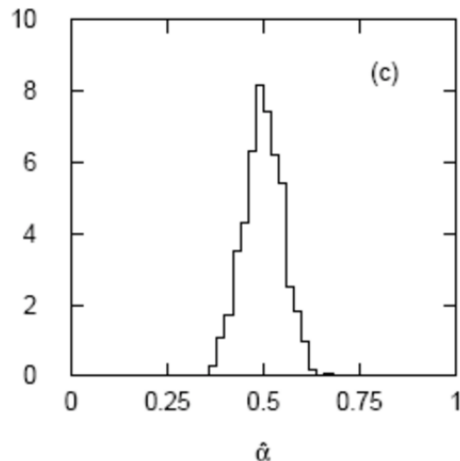
$$s_{\hat{\alpha}} = 0.051$$

$$\bar{\hat{\beta}} = 0.498$$

$$s_{\hat{\beta}} = 0.111$$

$$\widehat{\text{cov}}[\hat{\alpha}, \hat{\beta}] = 0.0024$$

$$r = 0.42$$



Mittelwerte der Schätzer  $\sim$  wahre Werte  
Kovarianzen nahe an früheren Abschätzungen.  
Randverteilungen ungefähr gaussförmig.

# Die “ $\ln L_{\max} - 1/2$ ”-Kontur

Für großen Stichprobenumfang  $n$ , wir  $\ln L$  Paraboloid (quadratische Form) in der Nähe des Maximums:

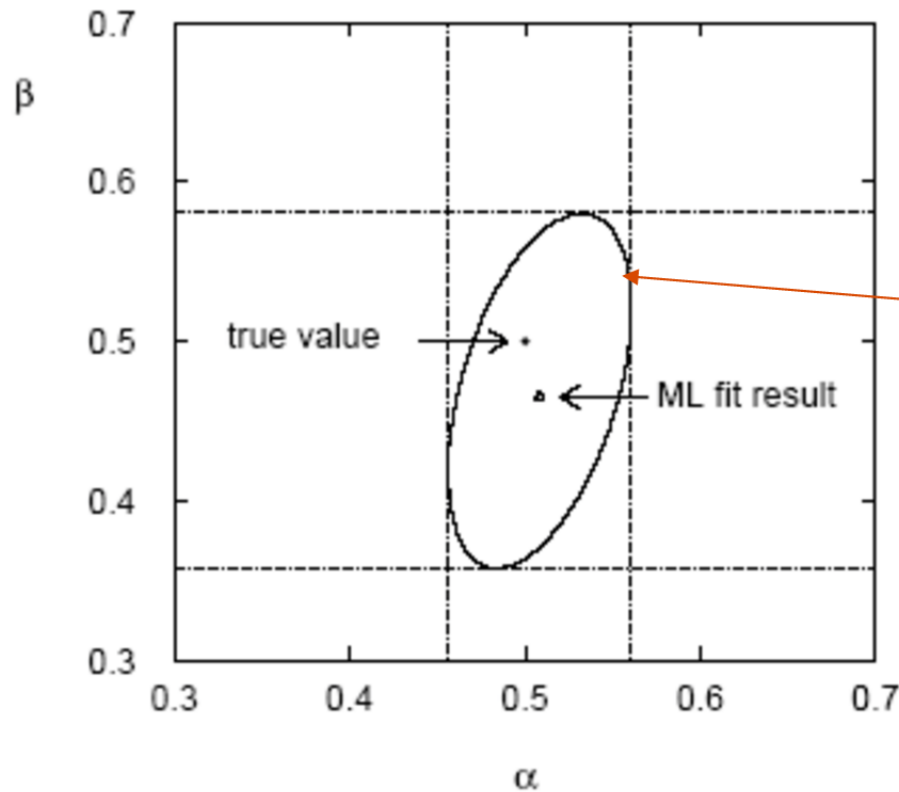
$$\ln L(\alpha, \beta) \approx \ln L_{\max}$$

$$-\frac{1}{2(1-\rho^2)} \left[ \left( \frac{\alpha - \hat{\alpha}}{\sigma_{\hat{\alpha}}} \right)^2 + \left( \frac{\beta - \hat{\beta}}{\sigma_{\hat{\beta}}} \right)^2 - 2\rho \left( \frac{\alpha - \hat{\alpha}}{\sigma_{\hat{\alpha}}} \right) \left( \frac{\beta - \hat{\beta}}{\sigma_{\hat{\beta}}} \right) \right]$$

Die Kontur  $\ln L(\alpha, \beta) = \ln L_{\max} - 1/2$  ist eine Ellipse:

$$\frac{1}{(1-\rho^2)} \left[ \left( \frac{\alpha - \hat{\alpha}}{\sigma_{\hat{\alpha}}} \right)^2 + \left( \frac{\beta - \hat{\beta}}{\sigma_{\hat{\beta}}} \right)^2 - 2\rho \left( \frac{\alpha - \hat{\alpha}}{\sigma_{\hat{\alpha}}} \right) \left( \frac{\beta - \hat{\beta}}{\sigma_{\hat{\beta}}} \right) \right] = 1$$

# Kovarianzen aus "In L"-Kontur



Die  $(\alpha, \beta)$ -Ebene für die erste Messung

$$\ln L(\alpha, \beta) = \ln L_{\max} - 1/2$$

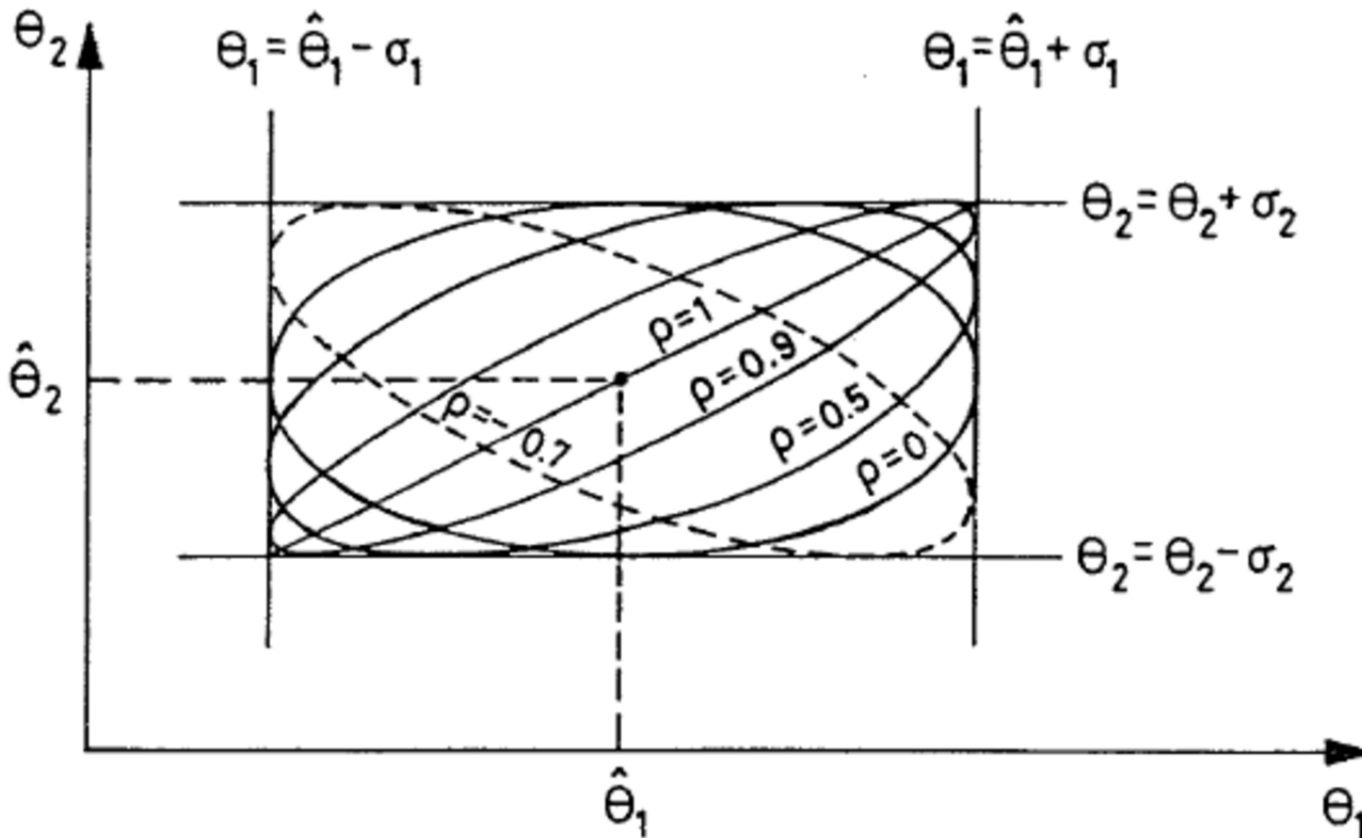
→ Tangenten an Kontur ergeben Standardabweichungen

→ Winkel der Hauptachse der Ellipse zur Horizontalen ergibt Korrelation:

$$\tan 2\phi = \frac{2\rho\sigma_{\hat{\alpha}}\sigma_{\hat{\beta}}}{\sigma_{\hat{\alpha}}^2 - \sigma_{\hat{\beta}}^2}$$

Korrelation zwischen Schätzern resultieren in einer Vergrößerung der Standardabweichungen (stat. Fehler).

# Kovarianzen aus "In L"-Kontur



Teilweise findet man: bilde senkrechte im Punkt des Schätzwertes  
→ Liefert zu kleine Fehler (bei Korrelation=1 verschwindende Fehler).

# Ein weiteres Beispiel

Gegeben sei Messgröße  $x > 0$

Annahme: folgt WDF  $f(x; s, \mu) = s \frac{1}{\sqrt{\pi}} e^{-(x-\mu)^2} + (1-s)e^{-x}$

Ein Anteil aus Gauss-WDF, ein Anteil aus Exponential WDF.

Wahre Wert der Parameter:

$$s_0 = 0.5$$
$$\mu_0 = 2.0$$

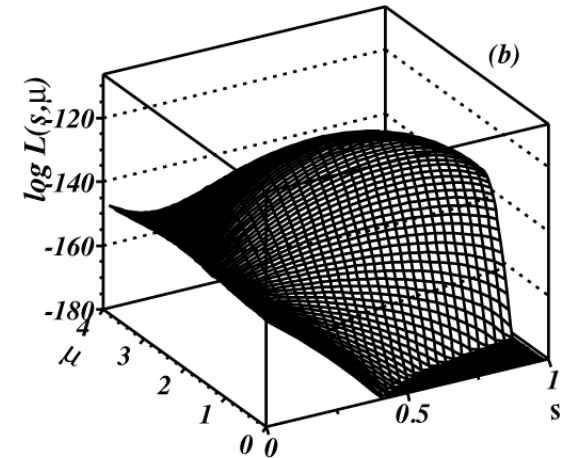
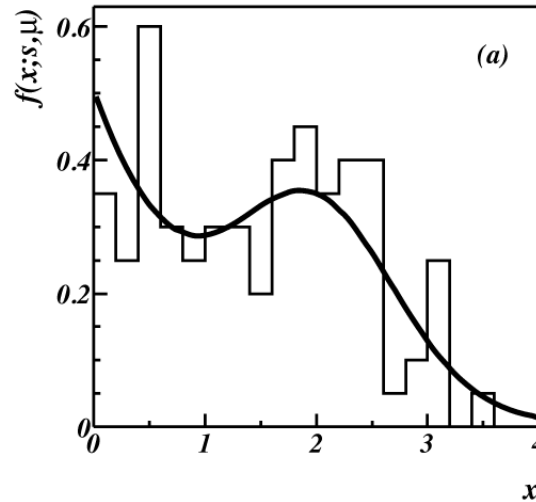
Messung (Simulation) liefert Stichprobe vom Umfang  $n=100$

Berechne Log-Likelihoodfunktion  $\log \mathcal{L}(s, \mu) = \sum_{i=1}^n \log f(x_i; s, \mu),$

# Ein weiteres Beispiel (2)

Links: Verteilung  
der Stichprobe vgl.  
mit wahrer WDF

Recht: Log-Likelihood-  
Funktion



Numerische Minimierung  
von  $-\log L$  mit  
MIGRAD in MINUIT

---

```

MIGRAD MINIMIZATION HAS CONVERGED.
MIGRAD WILL VERIFY CONVERGENCE AND ERROR MATRIX.
COVARIANCE MATRIX CALCULATED SUCCESSFULLY
FCN=118.51 FROM MIGRAD      STATUS=CONVERGED      34 CALLS      49 TOTAL
                        EDM=9.7183e-009    STRATEGY= 1      ERROR MATRIX ACCURATE

EXT PARAMETER
NO.   NAME      VALUE          ERROR          STEP          FIRST
      NAME      VALUE          ERROR          SIZE          DERIVATIVE
  1   sig       5.63270e-001  8.44144e-002  1.17212e-003 -6.04609e-004
  2   mean     1.93964e+000  1.32226e-001  4.53681e-004 -1.09683e-003
                                ERR DEF= 0.5
EXTERNAL ERROR MATRIX.      NDIM= 25   NPAR= 2   ERR DEF=0.5
7.196e-003 -4.669e-003
-4.669e-003 1.751e-002
PARAMETER CORRELATION COEFFICIENTS
NO.   GLOBAL      1      2
  1   0.41595    1.000 -0.416
  2   0.41595   -0.416 1.000
    
```

---

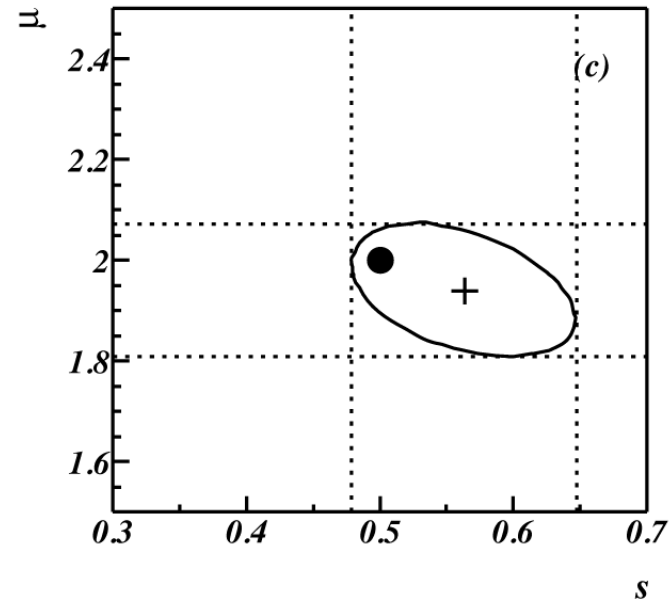


# Ein weiteres Beispiel (3)

## ML-Schätzer

$$\hat{s} = 0.56 \pm 0.08$$

$$\hat{\mu} = 1.94 \pm 0.13$$



Aus Krümmung der  
log-Likelihoodfkt. Im  
Minum ergibt sich Kovarianzmatrix

$$\begin{aligned} (\widehat{V^{-1}})_{ij} &= -\frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} \Big|_{\vec{\theta}=\hat{\theta}} \\ \Rightarrow \widehat{V} &= \begin{pmatrix} 7.196 \times 10^{-3} & -4.669 \times 10^{-3} \\ -4.669 \times 10^{-3} & 1.751 \times 10^{-2} \end{pmatrix} \end{aligned}$$

Negative Koorrelation:  $\rho=-0.42$

# Ein weiteres Beispiel (4)

Alternative Abschätzung der Varianz über  $\Delta \ln L = \frac{1}{2}$ -Regel

Für großen Stichprobenumfang geht  $\ln L$  in Paraboloid über

$$\log \mathcal{L}(s, \mu) = \log \mathcal{L}_{max} - \frac{1}{2(1 - \rho_{\hat{s}\hat{\mu}})^2} \left( \left( \frac{s - \hat{s}}{\sigma_{\hat{s}}} \right)^2 + \left( \frac{\mu - \hat{\mu}}{\sigma_{\hat{\mu}}} \right)^2 - 2\rho_{\hat{s}\hat{\mu}} \left( \frac{s - \hat{s}}{\sigma_{\hat{s}}} \right) \left( \frac{\mu - \hat{\mu}}{\sigma_{\hat{\mu}}} \right) \right)$$

Die Kontur für  $\log \mathcal{L}(s, \mu) = \log \mathcal{L}_{max} - 0.5$

ist gegeben durch die Ellipsengleichung:

$$1 = \frac{1}{(1 - \rho_{\hat{s}\hat{\mu}})^2} \left( \left( \frac{s - \hat{s}}{\sigma_{\hat{s}}} \right)^2 + \left( \frac{\mu - \hat{\mu}}{\sigma_{\hat{\mu}}} \right)^2 - 2\rho_{\hat{s}\hat{\mu}} \left( \frac{s - \hat{s}}{\sigma_{\hat{s}}} \right) \left( \frac{\mu - \hat{\mu}}{\sigma_{\hat{\mu}}} \right) \right) \quad \tan(2\phi) = \frac{2\rho_{\hat{s}\hat{\mu}}\sigma_{\hat{s}}\sigma_{\hat{\mu}}}{\sigma_{\hat{s}}^2 - \sigma_{\hat{\mu}}^2}$$

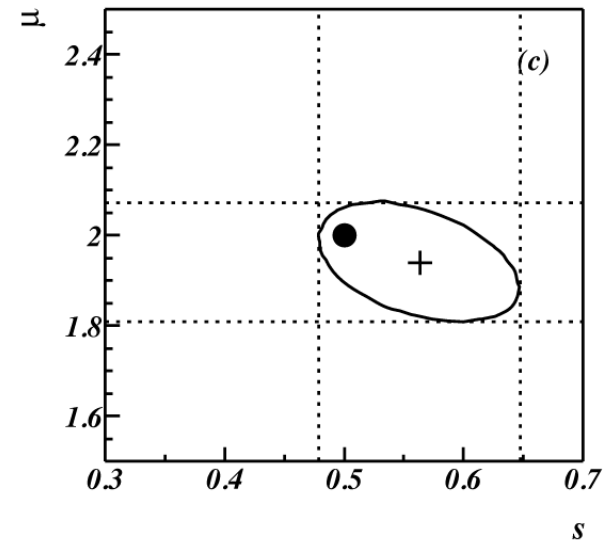
# Ein weiteres Beispiel (5)

## Numerische Bestimmung mit MINOS in MINUIT:

---

```
FCN=118.51 FROM MINOS      STATUS=SUCCESSFUL      46 CALLS      95 TOTAL
                        EDM=9.7183e-009      STRATEGY= 1      ERROR MATRIX ACCURAT
EXT PARAMETER              PARABOLIC              MINOS ERRORS
NO.  NAME      VALUE      ERROR      NEGATIVE      POSITIVE
  1  sig      5.63270e-001  8.44144e-002 -8.50699e-002  8.40686e-002
  2  mean     1.93964e+000  1.32226e-001 -1.30473e-001  1.34331e-001
                        ERR DEF= 0.5
```

---



Ln L keine perfekter Paraboloid → asymmetrische Fehler

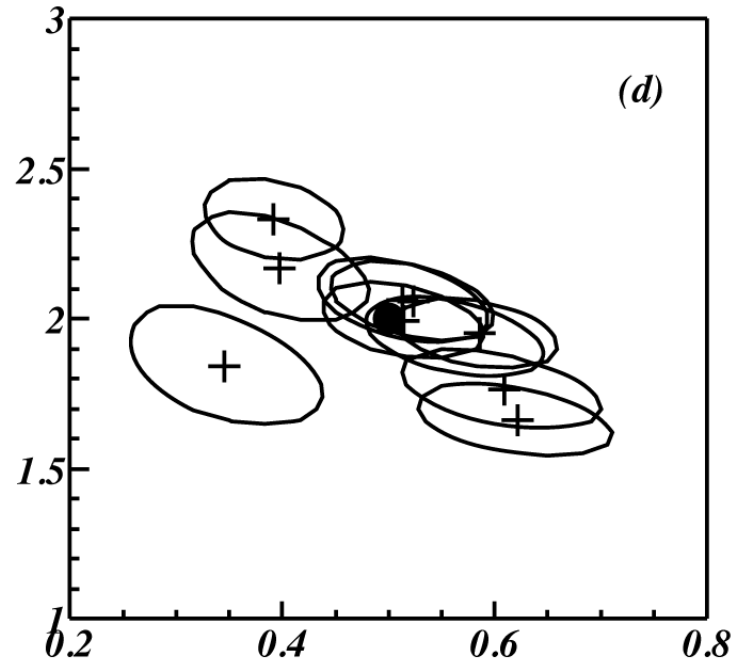
Fehler stimmen innerhalb Unsicherheit mit MIGRAD-Methode überein

# Ein weiteres Beispiel (6)

Simulation von 10 identischen  
Messreihe je a 100 Messwerte.

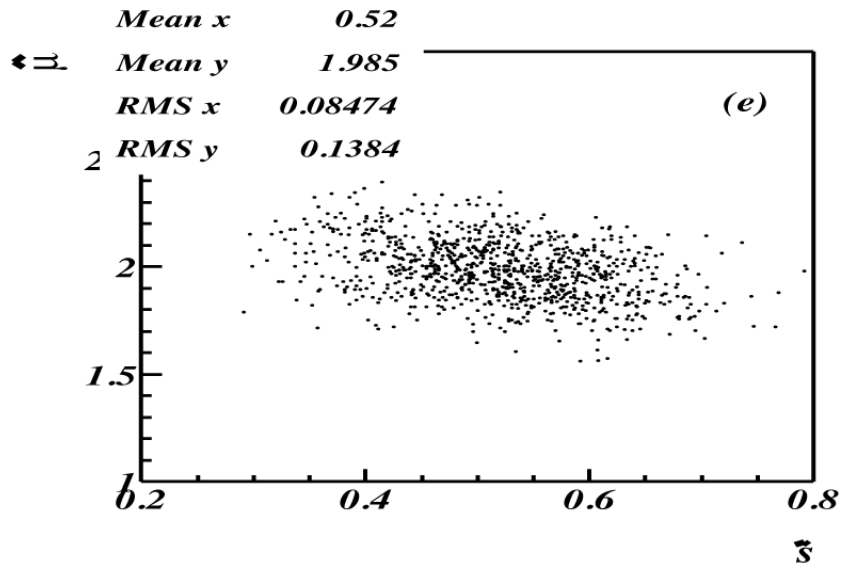
In 4 von 10 Fällen liegt wahrer Wert  
in der Fehlerellipse.

(für 2-dim Gaussverteilung erwartet  
man, dass in 39% der Fällen die  
Kovarianzellipse den wahren Wert abdeckt)



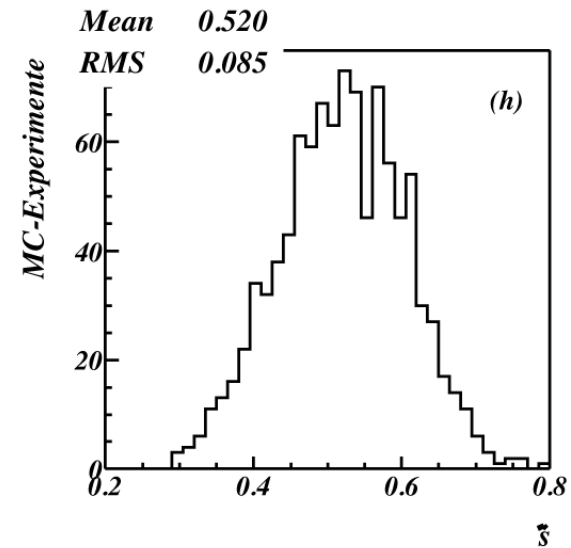
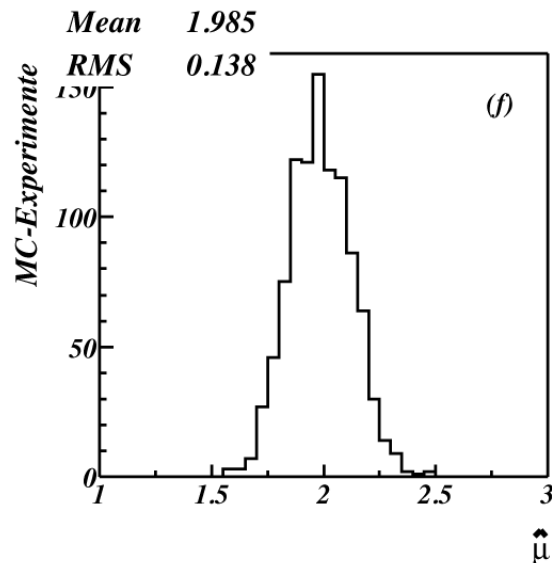
Jetzt Simulation von 1000 Messreihen und Betrachtung  
der Verteilung der ML-Schätzer

# Ein weiteres Beispiel (7)



Korrelation sichtbar.

Breite der Verteilung stimmt  
Mit abgeschätzten Varianzen in  
Einzelmessung überein



# Erweiterte ML-Methode

Bisher haben wir nur die Form der WDF  $f(x, \theta)$  verwendet.

Nicht auch die absolute Normierung.

Betrachte Anzahl der Beobachtungen als Poission-verteilte ZV

Das Ergebnis eines Experimentes ist dann:  $n, x_1, \dots, x_n$ .

Die erweiterte Likelihoodfunktion lautet:

$$L(\nu, \vec{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^n f(x_i; \vec{\theta})$$

Annahme die Theorie gibt einen Zusammenhang  $\nu = \nu(\theta)$ , dann ist die erweiterte log-Likelihoodfunktion (EML) gegeben durch:

$$\ln L(\vec{\theta}) = -\nu(\vec{\theta}) + \sum_{i=1}^n \ln(\nu(\vec{\theta}) f(x_i; \vec{\theta})) + C$$

wobei  $C$  Terme repräsentiert, die nicht von den Parametern  $\theta$  abhängen.

# Erweiterte ML-Methode

Beispiel: erwartete Anzahl von Ereignissen  $\nu(\vec{\theta}) = \sigma(\vec{\theta}) \int L dt$

Wobei der totale Wirkungsquerschnitt  $\sigma(\theta)$  als Funktion der Parameter in der Theorie vorhergesagt wird, ebenso wie die Verteilung einer Variablen  $x$  (normierter differentieller Wirkungsquerschnitt).

Erweiterte ML nutzt mehr Information  $\rightarrow$  kleinere Varianz für  $\hat{\vec{\theta}}$

Wenn  $\nu$  nicht von  $\theta$  abhängt, sondern ein freier Parameter ist, dann liefert die EML-Methode:

$$\hat{\nu} = n$$

$$\hat{\theta} = \text{same as ML}$$

(Dennoch kann es sinnvoll sein die EML-Methode zu verwenden.)

# Erweiterte ML Beispiel

Betrachte zwei Arten von Ereignissen (z.B. Signal und Untergrund), von denen für jede in der Theorie die WDF für eine Observable  $x$   $f_s(x)$  and  $f_b(x)$  vorhergesagt wird.

Wir beobachten im Experiment eine Überlagerung der beiden Arten von Ereignissen.

Der Anteil an Signalereignissen sei  $\theta$ , die erwartete Gesamtzahl  $\nu$ ,  
Die Anzahl beobachteter Ereignisse  $n$ .

Dann gilt  $\mu_s = \theta\nu$ ,  $\mu_b = (1 - \theta)\nu$ , Ziel ist Schätzung von  $\mu_s, \mu_b$ .

$$f(x; \mu_s, \mu_b) = \frac{\mu_s}{\mu_s + \mu_b} f_s(x) + \frac{\mu_b}{\mu_s + \mu_b} f_b(x)$$

$$P(n; \mu_s, \mu_b) = \frac{(\mu_s + \mu_b)^n}{n!} e^{-(\mu_s + \mu_b)}$$

$$\rightarrow \ln L(\mu_s, \mu_b) = -(\mu_s + \mu_b) + \sum_{i=1}^n \ln [(\mu_s + \mu_b) f(x_i; \mu_s, \mu_b)]$$



# Erweiterte ML Beispiel (2)

Monte-Carlo-Beispiel aus einer Kombination von Exponential-WDF (Untergrund) und Gauss-WDF (Signal):

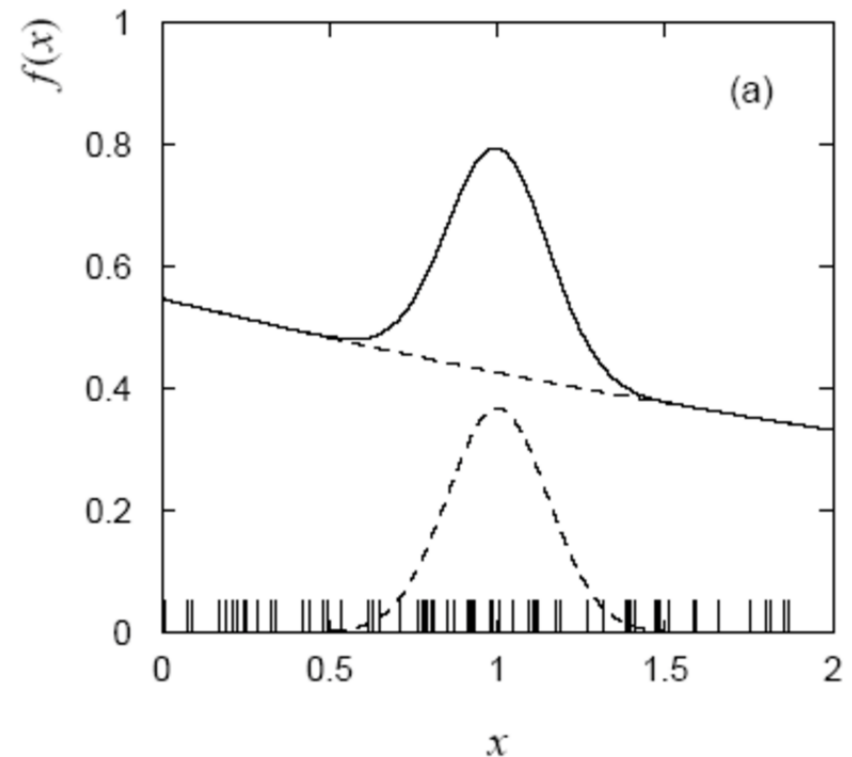
$$\mu_s = 6$$

$$\mu_b = 60$$

Maximiere log-Likelihood bzgl.  $\mu_s$  and  $\mu_b$ :

$$\hat{\mu}_s = 8.7 \pm 5.5$$

$$\hat{\mu}_b = 54.3 \pm 8.8$$



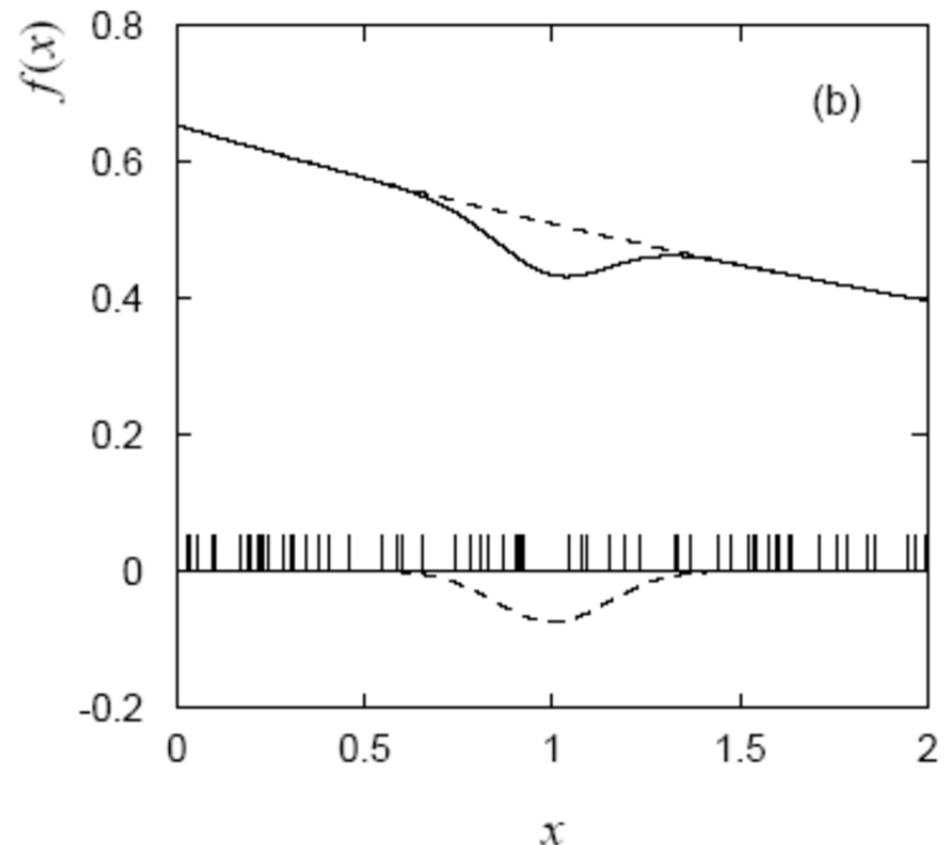
Fehler reflektieren Poisson-Fluktuation  
In der Gesamtanzahl, als auch im Verhältnis  
von Signal/Untergrund

# EML-Beispiel: ein unphysikalischer Schätzwert

Eine Abwärtsfluktuation in den Daten in der Signalregion kann dazu führen, dass wenige Ereignisse geschätzt werden, als unter der Annahme, dass nur Untegruncereignisse existieren.

Der Schätzwert für  $\mu_S$  wird dann zu negativen (unphysikalischen) Werten gezwungen.

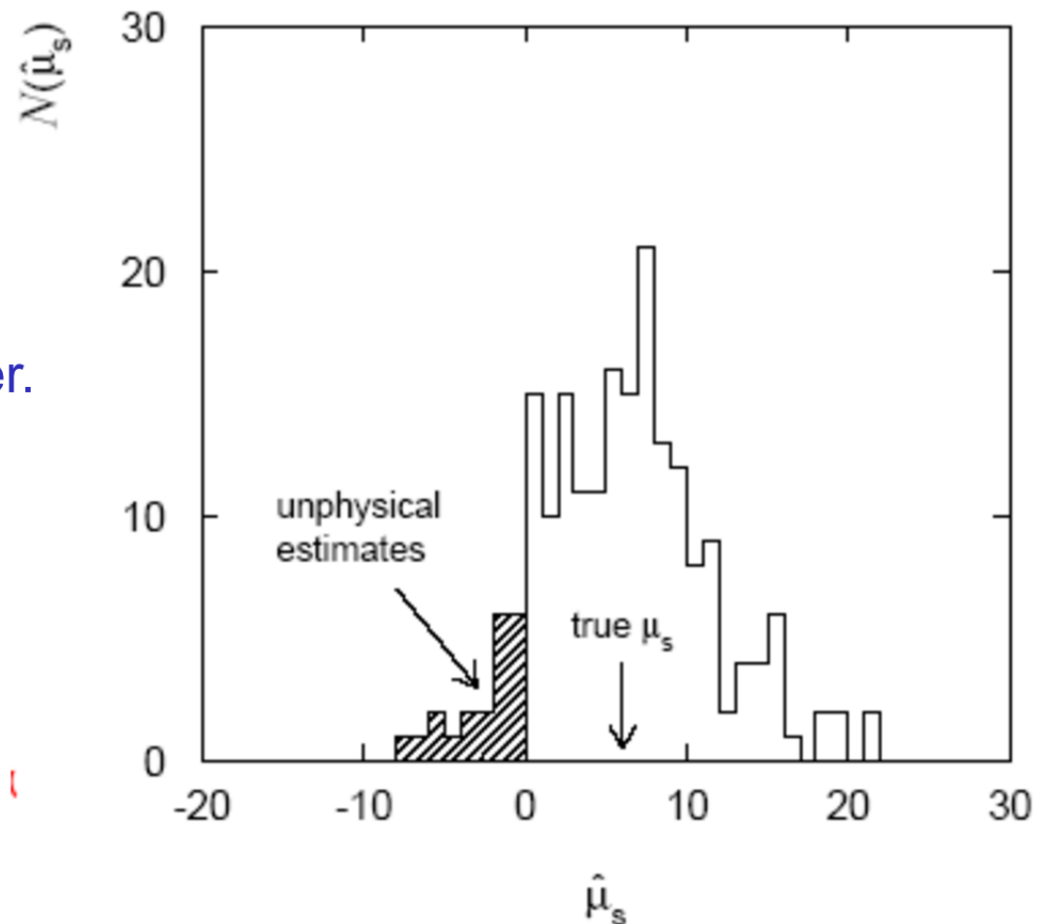
Das kann akzeptiert werden Solange die Gesamt-WDF überall positiv bleibt.



# Unphysikalische Schätzwerte (2)

In diesem Fall ist der unphysikalische Schätzer erwartungstreu und sollte deshalb trotzdem als Ergebnisse berichtet werden, weil der Erwartungswert über eine große Anzahl von Schätzungen gegen den wahren Wert konvergiert.

Rezept: wiederhole Experiment viele male und erlaube unphysikalische Schätzer.



# ML-Methode für Histogramme

Oft werden Daten in Histogramm gefüllt  
(um Rechenzeit bei Auswertung der Likelihood  
Für sehr große Stichproben zu sparen)

Oder Daten liegen nur in Form eines Histogrammes vor

Oder WDF der Theorie kann nicht analytisch berechnet  
Sondern nur durch MC-Simulation (z.B. Berücksichtigung  
Von Akzeptanz. Und Auflösungseffekten (→Faltung))

Stichprobe liefert Histogramm mit Einträgen:

$$\vec{n} = (n_1, \dots, n_N), n_{\text{tot}} = \sum_{i=1}^N n_i$$

Hypothese im Modell mit Parametern  $\theta$  ist (mit  $\nu_{\text{tot}} = n_{\text{tot}}$ ):

$$\vec{\nu} = (\nu_1, \dots, \nu_N), \nu_{\text{tot}} = \sum_{i=1}^N \nu_i$$

# ML-Methode für Histogramme

Bei analytischer bekannter WDF der Theorie ergeben sich  $\nu_i$

$$\nu_i(\vec{\theta}) = \nu_{\text{tot}} \int_{\text{bin } i} f(x; \vec{\theta}) dx$$

oder direkt aus dem Histogramm der Simulation für die WDF.

Da  $n_{\text{tot}}$  konstant, folgt die gemeinsame WDF für den Stichprobenvektor einer Multinomialverteilung.

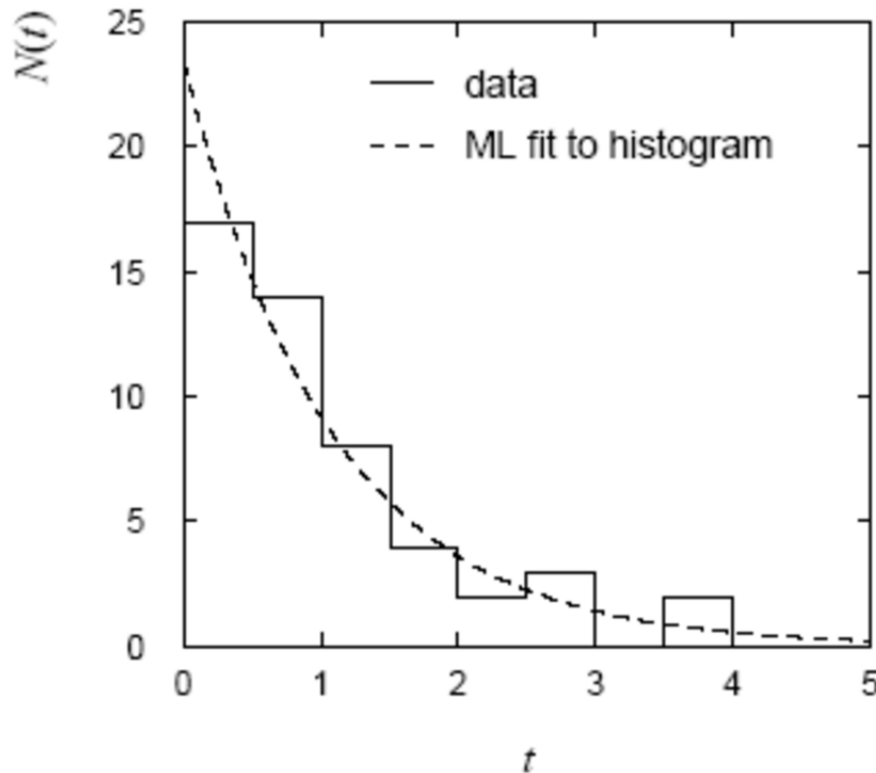
$$f(\vec{n}; \vec{\nu}) = \frac{n_{\text{tot}}!}{n_1! \dots n_N!} \left( \frac{\nu_1}{n_{\text{tot}}} \right)^{n_1} \dots \left( \frac{\nu_N}{n_{\text{tot}}} \right)^{n_N}$$

Die log-Likelihood-Funktion ergibt sich zu:

$$\ln L(\vec{\theta}) = \sum_{i=1}^N n_i \ln \nu_i(\vec{\theta}) + C \quad C = \text{const.} = \text{Terme unabhängig von } \theta_i$$

# ML-Beispiel für Histogramm

Unser beliebtes Beispiel: die Exponential-WDF, aber nun als Histogramm



$$\hat{\tau} = 1.07 \pm 0.17$$

( $1.06 \pm 0.15$  for unbinned  
ML with same sample)

Im Grenzfall verschwindender Binbreite  $\rightarrow$  normale ML-Methode  
Einteilung in Bins führt zu Informationsverlust  $\rightarrow$  Varianz größer

# Erweiterte ML-Beispiel für Histogramm

Bisher nur Form der Histogramme.

Zahl der Einträge in Theorie durch Beobachtung fixiert (mit  $\nu_{\text{tot}} = n_{\text{tot}}$ )

$$\vec{n} = (n_1, \dots, n_N), \quad n_{\text{tot}} = \sum_{i=1}^N n_i \quad \vec{\nu} = (\nu_1, \dots, \nu_N), \quad \nu_{\text{tot}} = \sum_{i=1}^N \nu_i$$

Betrachte nun  $n_{\text{tot}}$  als Poisson-ZV mit Mittelwert  $\nu_{\text{tot}} = n_{\text{tot}}$ .

Die gemeinsame WDF ergibt sich zu:

$$f(\vec{n}; \vec{\nu}) = \frac{n_{\text{tot}}!}{n_1! \dots n_N!} \left( \frac{\nu_1}{n_{\text{tot}}} \right)^{n_1} \dots \left( \frac{\nu_N}{n_{\text{tot}}} \right)^{n_N} \quad \times \text{Pois}(n_{\text{tot}}; \nu_{\text{tot}})$$

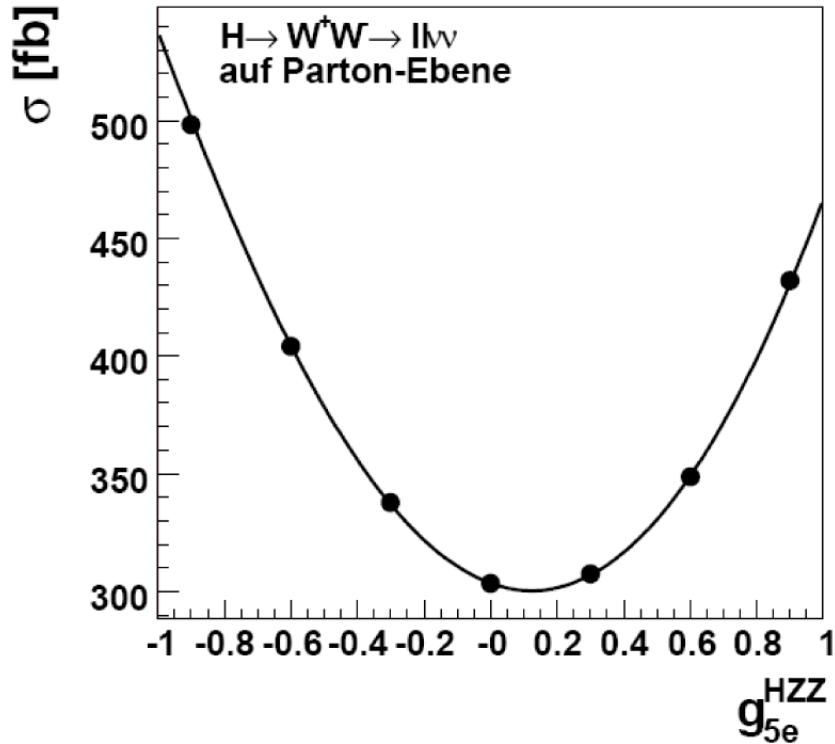
= Produkt von N unabhängigen Poisson-WDF in jedem Bin.

Und die erweiterte Likelihoodfunktion ist:

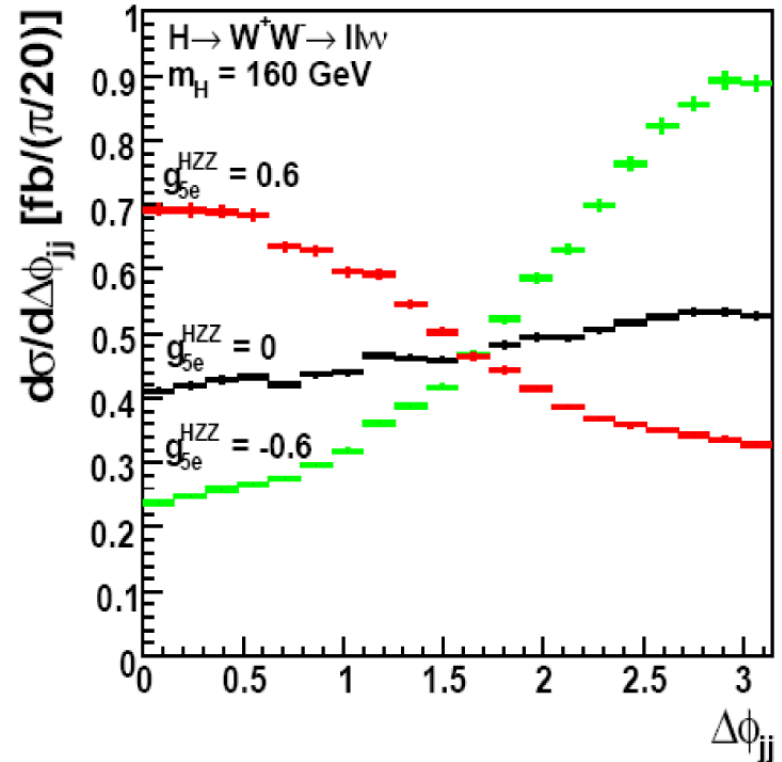
$$\ln L(\nu_{\text{tot}}, \vec{\theta}) = -\nu_{\text{tot}} + \sum_{i=1}^N n_i \ln \nu_i(\nu_{\text{tot}}, \vec{\theta}) + C$$

# Erweiterte ML-Methode: Beispiel

Ziel: Messung von  $g$ .



Anzahl der beobachteten Ereignisse hängt quadratisch von  $g$  ab.



Form hängt nicht trivial von  $g$  ab.



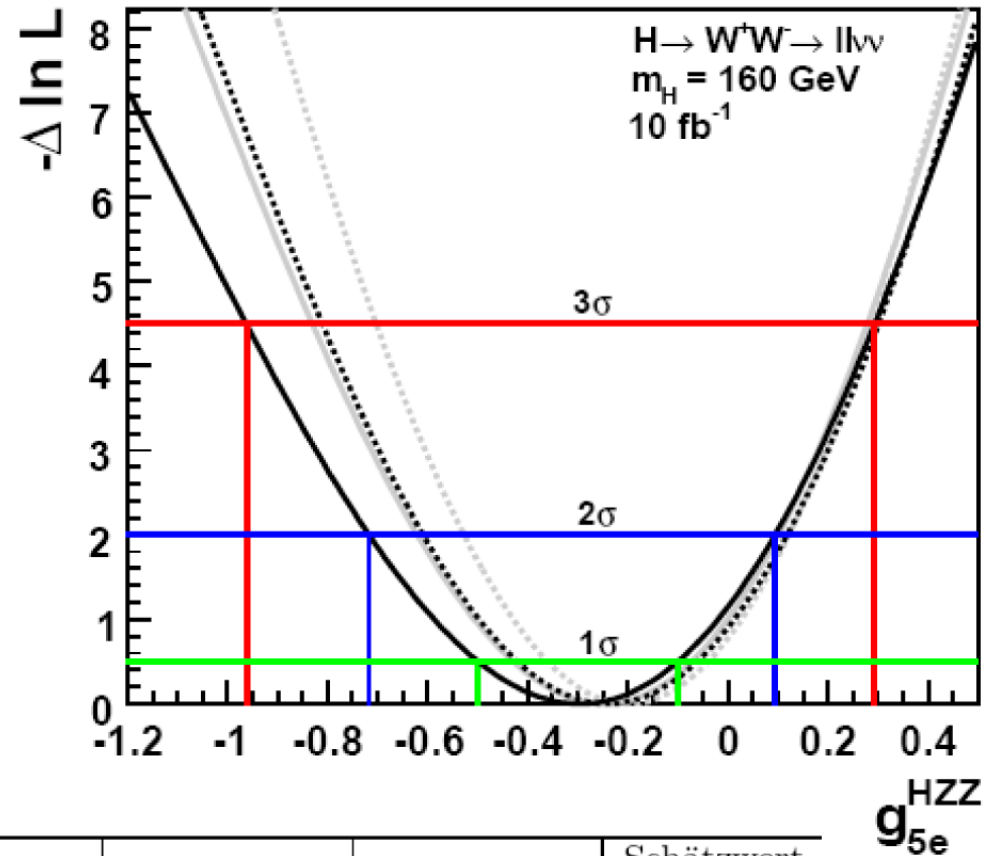
# Erweiterte ML-Methode: Beispiel-Fit Ergebnis

ML: durchgezogene Linien

$$\ln L(g_{5e}^{HZZ}) = \sum_{i=1}^N n_i \ln \nu_i(g_{5e}^{HZZ})$$

EML: unterbochene Linien

$$\ln L(g_{5e}^{HZZ}) = -\nu_{tot}(g_{5e}^{HZZ}) + \sum_{i=1}^N n_i \ln \nu_i(g_{5e}^{HZZ})$$



	Minimum $-\Delta \ln L$	$1\sigma$ -Intervall	$2\sigma$ -Intervall	$3\sigma$ -Intervall	Schätzwert für $\sigma$
$H \rightarrow W^+W^- \rightarrow ll\nu\nu$					
$10 \text{ fb}^{-1}$ einfach	-0.30	[-0.50, -0.10]	[-0.72, 0.10]	[-0.96, 0.30]	0.20
erweitert	-0.24	[-0.42, -0.06]	[-0.61, 0.12]	[-0.81, 0.30]	0.18

EML reduziert den statistischen Fehler

# Beziehung zwischen ML und Bayes-Schätzern

In Bayesianischer Statistik, sind beide,  $\theta$  und  $\mathbf{x}$  Zufallsvariablen.

$$L(\theta) = L(\vec{x}|\theta) = f_{\text{joint}}(\vec{x}|\theta)$$

Erinnerung an Bayesianische Interpretation:

Benutze subjektive Wahrscheinlichkeit für Hypothesen ( $\theta$ );

vor Messung/Experiment: Kenntnis zusammengefasst in A-Priori-WDF  $\pi(\theta)$ ;

verwende Bayes-Theorem, um die A-Priori-Wkt. durch die Daten aus Experiment abzuändern/anzupassen.

$$p(\theta|\vec{x}) = \frac{L(\vec{x}|\theta)\pi(\theta)}{\int L(\vec{x}|\theta')\pi(\theta') d\theta'}$$

Posterior-WDF für Hypothesen über Parameter  $\theta$   
(Bedingte WDF für  $\theta$  gegeben die Stichprobe  $\mathbf{x}$ )

# Beziehung zwischen ML und Bayes-Schätzern

Reiner Bayesianer:  $p(\theta | x)$  enthält gesamte Kenntnis über Parameter  $\theta$ . Brauche nichts anderes.

Pragmatischer Bayesianer:  $p(\theta | x)$  könnte komplizierte Funktion sein,  
→ fasse Kenntnis über Parameter in Schätzwert zusammen  $\hat{\theta}_{\text{Bayes}}$

Bestimme Modalwert (oder Erwartungswert) von  $p(\theta | x)$

Was verwenden wir für "Prior"  $\pi(\theta)$ ? Es gibt keine goldene Regel, Often wird "Unkenntnis des Prior" durch Gleichverteilung  $\pi(\theta) = \text{konstant}$ , präsentiert. Dann gilt für Modalwert=Maximum von  $p(\theta | x)$

$$\hat{\theta}_{\text{Bayes}} = \hat{\theta}_{\text{ML}}$$

Aber... wir könnten einen anderen Parameter verwendet haben z.B.  $\lambda = 1/\theta$ , und wenn "Prior"  $\pi_{\theta}(\theta)$  konstant ist, dann nicht der Prior  $\pi_{\lambda}(\lambda)$  !

"Komplette Ignoranz des Prior" ist nicht wohldefiniert.

# Kombination von Messungen mit ML-Methode:

Annahme: zwei Observable  $x$  und  $y$  mit unterschiedlichen Abhängigkeiten vom Parameter  $\theta$ , der geschätzt werden soll.

$$f_x(x; \theta) \quad f_y(y; \theta)$$

Zwei unabhängige Messreihen/stichproben für  $x$  und  $y$ .

Wenn für beide Observablen die Likelihoodfunktionen bekannt sind, dann ergibt sich die gemeinsame Likelihoodfunktion zu:

$$\mathcal{L}(\theta) = \prod_{i=1}^n f_x(x_i; \theta) \cdot \prod_{j=1}^m f_y(y_j; \theta) = \mathcal{L}_x(\theta) \cdot \mathcal{L}_y(\theta)$$

Bestimmung von Schätzwert und Varianz für  $\theta$  mit normaler ML-Methode wie gehabt.

# Kombination von Messungen mit ML-Methode:

Annahme: nur Schätzwerte  $\hat{\theta}_x$  und  $\hat{\theta}_y$  und Varianzen  $\widehat{V}[\hat{\theta}_x]$  und  $\widehat{V}[\hat{\theta}_y]$  bekannt

Im Grenzfall großer Stichproben werden die WDF für die Schätzer durch Gauss-Verteilungen beschrieben:  $g_x(\hat{\theta}_x; \theta)$  und  $g_y(\hat{\theta}_y; \theta)$

Die Messungen können dann kombiniert werden mittels der Likelihoodfunktion

$$\mathcal{L}(\theta) = g_x(\hat{\theta}_x; \theta) \cdot g_y(\hat{\theta}_y; \theta)$$

Der ML-Schätzer ergibt sich zu: 
$$\hat{\theta} = \frac{\frac{\hat{\theta}_x}{\widehat{V}[\hat{\theta}_x]} + \frac{\hat{\theta}_y}{\widehat{V}[\hat{\theta}_y]}}{\frac{1}{\widehat{V}[\hat{\theta}_x]} + \frac{1}{\widehat{V}[\hat{\theta}_y]}}$$

und dessen Varianz zu:

$$\widehat{V}[\hat{\theta}] = \frac{1}{\frac{1}{\widehat{V}[\hat{\theta}_x]} + \frac{1}{\widehat{V}[\hat{\theta}_y]}}$$

# Kapitel 6

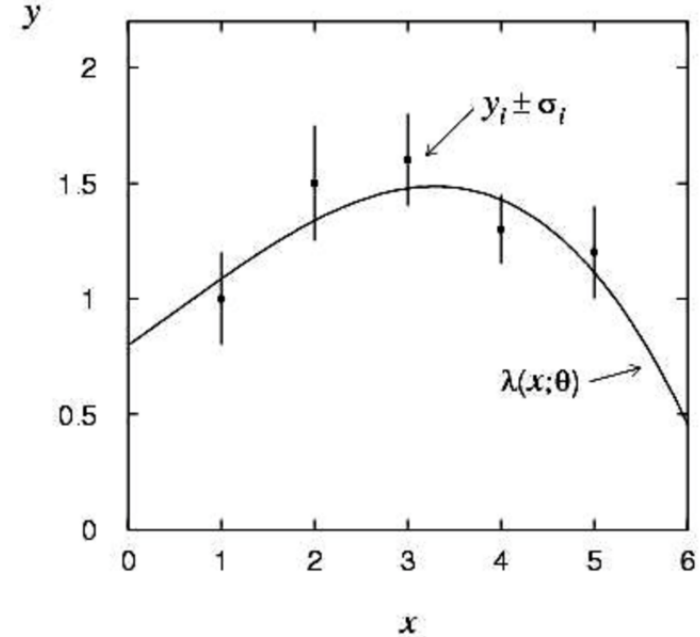
## Die Methode der kleinsten Quadrate

# Die Method der kleinsten Quadrate KQ

Betrachte  $N$  unabhängige Messungen von Wertepaaren  $(x_i, y_i)$

$x_i$  perfekt bekannt

$y_i$  mit Varianz  $V[y_i] = \sigma_i^2$  .  
aus beliebiger WDF



Theorievorhersage für Zusammenhang  $\lambda$  zwischen  $x$  und  $y$

$E[y_i] = \lambda(x_i; \theta)$  . Anzahl der Parameter  $\theta_i <$  Anzahl der Messpaare

z.B:  $y = s_0 + v_0 x + 0.5 g^2 x$  Weg-Zeit-Gesetz  $x$ =Zeit

$y$  Häufigkeit des Auftretens eines numerischen Wertes  $x$

# Das KQ-Prinzip

Bestimme Schätzwerte für die Parameter aus der Minimierung der folgenden Größe:

$$\chi^2 \equiv \sum_{i=1}^N w_i (y_i - f_i)^2 \quad w_i = \text{Gewicht für die } i\text{-te Beobachtung}$$

Theoriezusammenhang hier:  $f_i = f_i(\theta_1, \theta_2, \dots, \theta_L; \mathbf{x}_i)$

Wenn Varianz bekannt dann:  $w_i = 1/\sigma_i^2 \quad \chi^2 = \sum_{i=1}^N \left( \frac{y_i - f_i}{\sigma_i} \right)^2$

Wenn Varianz unbekannt oder gleich dann setze  $w_i=1$ :  $\chi^2 = \sum_{i=1}^N (y_i - f_i)^2$

Wenn  $y_i$  = Zählrate:, dann verwende erwartete Poisson-Varianz  $\sigma_i^2 \approx f_i$

$$\chi^2 = \sum_{i=1}^N \frac{(y_i - f_i)^2}{f_i}$$

Nicht erwartete Varianz  $\sigma_i^2 \approx y_i$  liefert verzerrte Schätzer  $\chi^2 = \sum_{i=1}^N \frac{(y_i - f_i)^2}{y_i}$



# Verbindung zusammen mit ML-Methode

Betrachte  $N$  unabhängige Messungen von Wertepaaren  $(x_i, y_i)$

$x_i$  perfekt bekannt       $y_i$  mit Varianz  $V[y_i] = \sigma_i^2$  . nun aus Gauss-WDF

Theoriezusammenhang:  $E[y_i] = \lambda(x_i; \theta)$  .

Die Likelihoodfunktion ist dann gegeben durch:

$$L(\theta) = \prod_{i=1}^N f(y_i; \theta) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[ -\frac{(y_i - \lambda(x_i; \theta))^2}{2\sigma_i^2} \right]$$

und die log-Likelihoodfunktion ergibt sich zu:

$$\ln L(\theta) = -\frac{1}{2} \sum_{i=1}^N \frac{(y_i - \lambda(x_i; \theta))^2}{\sigma_i^2} + \text{terms not depending on } \theta$$

# Zusammenhang: ML- und KQ-Methoden

$$\ln L(\theta) = -\frac{1}{2} \sum_{i=1}^N \frac{(y_i - \lambda(x_i; \theta))^2}{\sigma_i^2} + \text{terms not depending on } \theta$$

Vergleich mit: 
$$\chi^2(\theta) = \sum_{i=1}^N \frac{(y_i - \lambda(x_i; \theta))^2}{\sigma_i^2}$$

Es gilt also:  $-\frac{1}{2} \ln L = \chi^2$

und  $\Delta \ln L = -1/2$  entspricht  $\Delta \chi^2 = 1$

d.h. für gaussverteilte  $y_i$  gilt: ML-Schätzer = KQ-Schätzer  
dann “erben” KQ-Schätzer die Eigenschaften von ML-Schätzern

wenn  $y_i$  nicht gaussverteilt, dann ist KQ eine “ad hoc” Methode  
Keine generelle Aussage über Eigenschaften von KQ-Schätzer