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

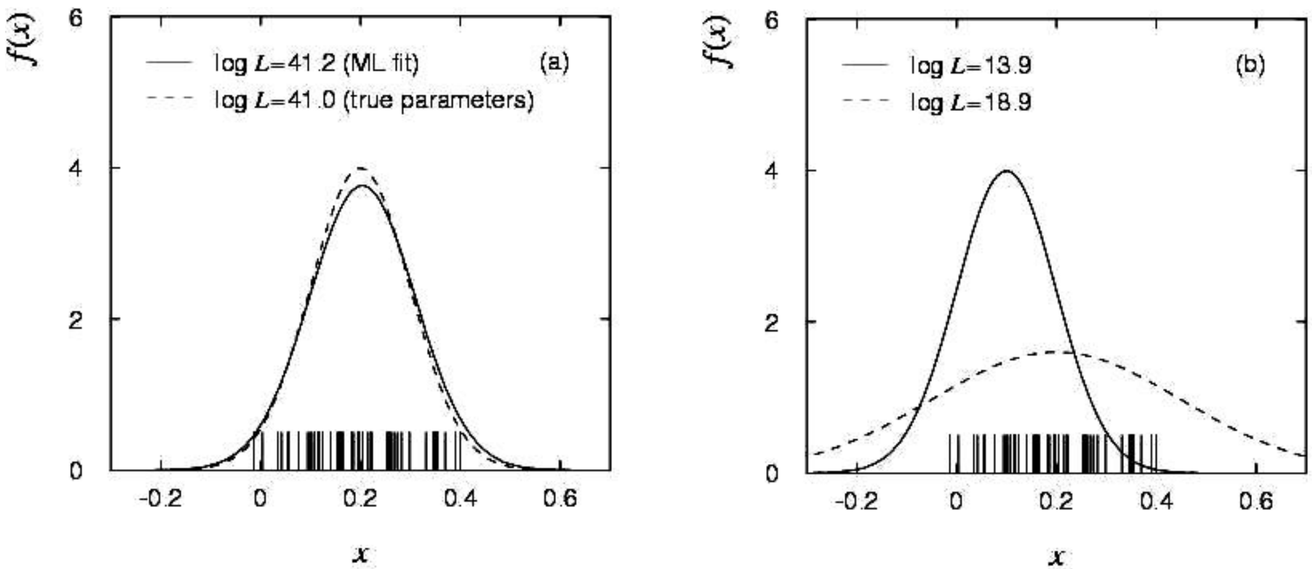
E-Mail: Markus.Schumacher@physik.uni-freiburg.de

Kapitel 13: Die Maximum-Likelihood-Methode

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

“Maximum Likelihood”-Schätzer: Grundidee

Wenn der hypothetisch angenomme Wert θ nahe am wahren Wert ist, dann erwarten wir, dass die Wahrscheinlichkeit groß ist, eine Stichprobe zu finden, wie wir sie aktuell gemessen haben.



Also definieren wir den “Maximum Likelihood (ML)”-Schätzer als den Parameterwert, der die Likelihoodfunktion maximiert.

Für ML-Schätzer gibt es keine Garantie, dass sie “optimale” Eigenschaften für kleine Stichproben besitzen. Aber in der Praxis sind sie sehr gut.

"Maximum Likelihood"-Schätzer: Formal

Das Maximum-Likelihood-Prinzip: finde Wert des Parameters so, dass

$$\mathcal{L}(\hat{\theta}) = \prod_{i=1}^{n} f(x_i; \hat{\theta}) = \mathrm{Maximum}.$$

Oft praktischer die log-Likelihood-Funktion (oder ln-Likelihood-Fkt.) zu verwenden. (ln und log meinen hier immer den natürlichen Logarithmus.)

$$\log \mathcal{L}(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

(Summieren ist einfacher als Multiplizieren.)

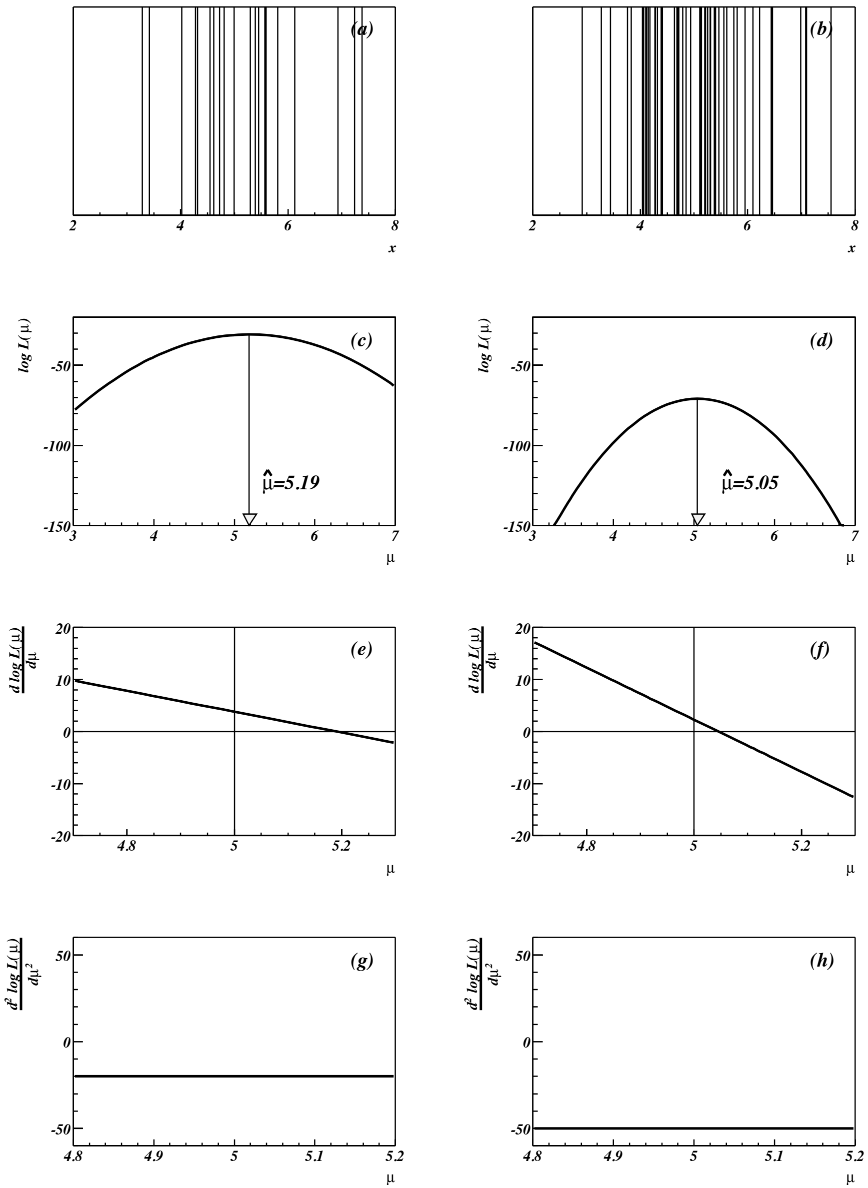
ML-Schätzer gegeben durch die Gleichung: $$\left.\frac{\partial \log \mathcal{L}}{\partial \theta}\right|_{\hat{\theta}} = 0$$

Für mehrere Parameter: $$\left.\frac{\partial \log \mathcal{L}}{\partial \theta_i}\right|_{\hat{\theta}_i} = 0, \qquad i = 1, ..., m$$

Gauss-WDF
$\mu=5$
$\sigma=1$
Stichprobe
n= 20

$$\hat{\mu} \approx 5.19$$

$$I(\mu) = 20$$



Gauss-WDF
$\mu=5$
$\sigma=1$
Stichprobe
n= 50

$$\hat{\mu} \approx 5.05$$

$$I(\mu) = 50$$

ML-Beispiel: Parameter der Exponential-WDF

Betrachte Exponential-WDF: $f(t;\tau) = \frac{1}{\tau} e^{-t/\tau}$

und unabhängige Messungen $t_1, \ldots, t_n$

Die Likelihoodfunktion lautet: $L(\tau) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-t_i/\tau}$

Der Wert von τ der $L(\tau)$ maximiert, liefert auch den Maximalwert seines Logarithmus (der Log-Likelihoodfunktion):

$$\ln L(\tau) = \sum_{i=1}^{n} \ln f(t_i;\tau) = \sum_{i=1}^{n} \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

ML-Beispiel: Parameter der Exponential-WDF (2)

$$\log \mathcal{L}(\tau) = \sum_{i=1}^{n} \log f(t_i;\tau) = \sum_{i=1}^{n} \left(\log \frac{1}{\tau} - \frac{t_i}{\tau} \right) = n \log \frac{1}{\tau} - \frac{1}{\tau} \sum_{i=1}^{n} t_i$$

Bestimmung des Maximums $\quad 0 = \left. \frac{\partial \log \mathcal{L}(\tau)}{\partial \tau} \right|_{\tau=\hat{\tau}} = -n\frac{1}{\hat{\tau}} + \frac{1}{\hat{\tau}^2} \sum_{i=1}^{n} t_i$

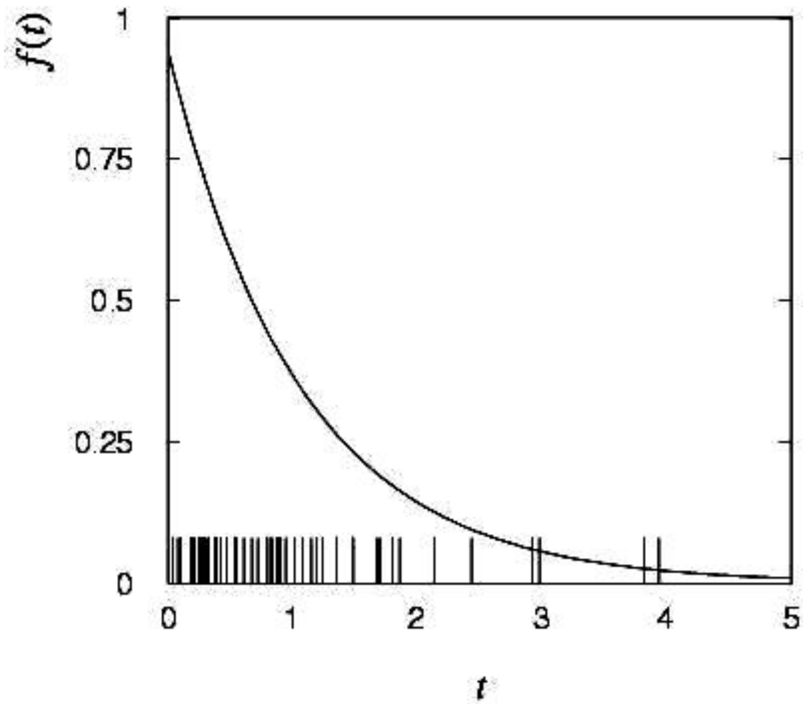
Liefert den ML-Schätzer: $\quad \hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i$

Hier also den arithmetischen Mittelwert der Stichprobe (daher konsistent und erwartungstreu)

Monte Carlo-Test:
generiere 50 Messwerte für $\tau = 1$.

Wir bestimmen den ML-Schätzer zu:

$$\hat{\tau} = 1.062$$



ML-Beispiel: Parameter der Exponential-WDF (3)

Fehler auf arithmetischen Mittelwert kennen wir: $V[\hat{\tau}] = \frac{1}{n} V[t] = \frac{1}{n} \tau^2$

Vergleich mit der Schranke minimaler Varianz:

$$\frac{\partial^2 \log \mathcal{L}}{\partial \tau^2} = \frac{n}{\tau^2} \left(1 - \frac{2}{n\tau} \sum_{i=1}^{n} t_i \right) = \frac{n}{\tau^2} \left(1 - \frac{2\hat{\tau}}{\tau} \right)$$

$$V[\hat{\tau}] \geq \frac{-1}{E\left[\frac{n}{\tau^2}\left(1 - \frac{2\hat{\tau}}{\tau}\right)\right]} = \frac{-1}{\frac{n}{\tau^2}\left(1 - \frac{2E[\hat{\tau}]}{\tau}\right)} = \frac{\tau^2}{n}$$

Also ML-Schätzer ist für dieses Probem auch effizient.

Schätzwert für die Varianz: $\widehat{V[\hat{\tau}]} = \frac{\hat{\tau}^2}{n}$

Invarianz unter Parametertransformationen

Sei $\hat{\theta}$ ML-Schätzer für den Parameter θ

$a(\theta)$ eine beliebige Funktion des Parameters

Frage: was ist der ML-Schätzer für a?

Es gilt: $$\left.\frac{\partial\mathcal{L}}{\partial\theta}\right|_{\hat{\theta}} = \left.\frac{\partial\mathcal{L}}{\partial a}\right|_{a(\hat{\theta})} \left.\frac{\partial a}{\partial\theta}\right|_{\hat{\theta}}$$

So lange $\partial a/\partial\theta \neq 0$ folgt damit:

$$\left.\frac{\partial\mathcal{L}}{\partial\theta}\right|_{\hat{\theta}} = 0 \quad \Rightarrow \quad \left.\frac{\partial\mathcal{L}}{\partial a}\right|_{a(\hat{\theta})} = 0 \quad \Rightarrow \quad \widehat{a(\theta)} = a(\hat{\theta})$$

D.h.: der ML-Schätzer für den transformierten Parameter ist gleich der Transformierten des ML-Schätzers des ursprünglichen Parameters. Dies ist eine sehr angenehme Eigenschaft, aber sie hat einen Preis.

Funktionen von ML-Schätzern

Annahme: wir hätten Exponential-WDF geschrieben als:

$$f(t;\lambda) = \lambda e^{-\lambda t} \ ,$$

i.e., wir verwenden $\lambda = 1/\tau$. Was ist der ML-Schätzer für λ ?

Für eine Funktion $\alpha(\theta)$ eines Parameters θ, ist es egal, ob wir die Likelihoodfunktion L als Funktion von α oder θ schreiben

Der ML-Schätzer einer Funktion $\alpha(\theta)$ ist einfach: $\hat{\alpha} = \alpha(\hat{\theta})$.

Für die Zerfallskonstante erhalten wir also: $\hat{\lambda} = \frac{1}{\hat{\tau}} = \left(\frac{1}{n} \sum_{i=1}^{n} t_i \right)^{-1}$.

Caveat: $\hat{\lambda}$ ist nicht erwartungstreu,
obwohl $\hat{\tau}$ erwartungstreu ist.

Man kann zeigen $E[\hat{\lambda}] = \lambda \frac{n}{n-1}$. (Bias $\rightarrow 0$ für $n \rightarrow \infty$)

Beispiel für ML: Parameter der Gauss-WDF

Betrachte n unabhängige Messungen $x_1, ..., x_n$,
wobei x_i derselben Gauss-WDF $f(x;\mu,\sigma^2)$ folgen

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

Die Log-Likelihood-Funktion ergibt sich zu:

$$\begin{aligned}
\ln L(\mu,\sigma^2) &= \sum_{i=1}^{n} \ln f(x_i;\mu,\sigma^2) \\
&= \sum_{i=1}^{n} \left(\ln \frac{1}{\sqrt{2\pi}} + \frac{1}{2} \ln \frac{1}{\sigma^2} - \frac{(x_i-\mu)^2}{2\sigma^2} \right) .
\end{aligned}$$

Beispiel für ML: Parameter der Gauss-WDF (2)

Setze Ableitungen nach μ, σ^2 zu Null, und löse die Gleichungen:

$$0 = \left.\frac{\partial \log \mathcal{L}(\mu, \sigma^2)}{\partial \mu}\right|_{\mu=\hat{\mu}} = \sum_{i=1}^{n} \frac{(x_i - \hat{\mu})}{\sigma^2} \qquad 0 = \left.\frac{\partial \log \mathcal{L}(\mu, \sigma^2)}{\partial \sigma^2}\right|_{\sigma^2=\widehat{\sigma^2}}$$

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2 \,.$$

Wir wissen bereits, dass der Schätzer für μ erwartungstreu ist.

Aber wir finden $E[\widehat{\sigma^2}] = \frac{n-1}{n}\sigma^2$, also hat ML estimator

für σ^2 einen Bias, aber es gilt: $b \to 0$ für $n \to \infty$.

Erinnerung: $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \widehat{\mu})^2$$

ist ein erwartungstreuer Schätzer für die Varianz σ^2.
Dies ist aber nicht der ML-Schätzer

Eigenschaften von ML-Schätzern

Konsistenz: Wenn Erwartungwert 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

$$f(x;\theta) = \exp\left(B(\theta)C(x) + D(\theta) + E(x)\right)$$

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 konsistent 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.

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

Varianz für Schätzer: Analytische Methode

In wenigen Fällen kann die Varianz analytisch berechnet werden.

Benötigt Berechnung von:

$$
\begin{aligned}
E[\hat{\theta}] &= \int \hat{\theta}(x_1, ..., x_n) f(x_1, ..., x_n; \theta_0)\, dx_1 ... dx_n \equiv \mu(\theta_0) \\
V[\hat{\theta}] &= \int (\hat{\theta}(x_1, ..., x_n) - \mu(\theta_0))^2 f(x_1, ..., x_n; \theta_0)\, dx_1 ... dx_n \equiv \sigma^2_{\theta_0}
\end{aligned}
$$

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

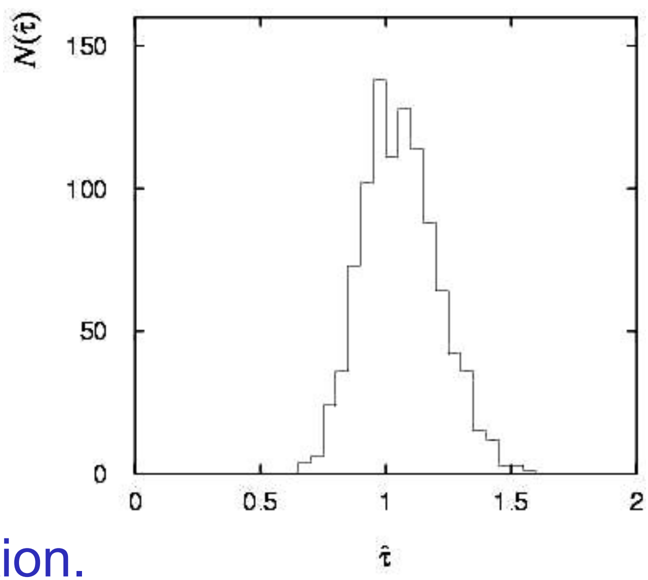
Nachdem wir den Schätzwert für den Parameter bestimmt haben müssen wir nun den statistischen Fehler des Schätzers bestimmen, i.e., wie weit die Verteilung der Schätzwerte wäre, wenn wir die gesamte identische Messung sehr oft wiederholen würden.

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:

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



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

Wähle Datenschätzwert 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):

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

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

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

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

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

Varianz des Schätzers: Graphische Methode

Entwickele ln L (θ) in Taylorreihe um das Maximum:

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

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

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

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

$\rightarrow$ um $\widehat{\sigma}_{\widehat{\theta}}$ zu erhalten, ändere θ von $\widehat{\theta}$ weg, bis ln L um ½ kl. ist als im Max.

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

Beispiel: WDF-Exponentialfunktion

Graphisch:

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

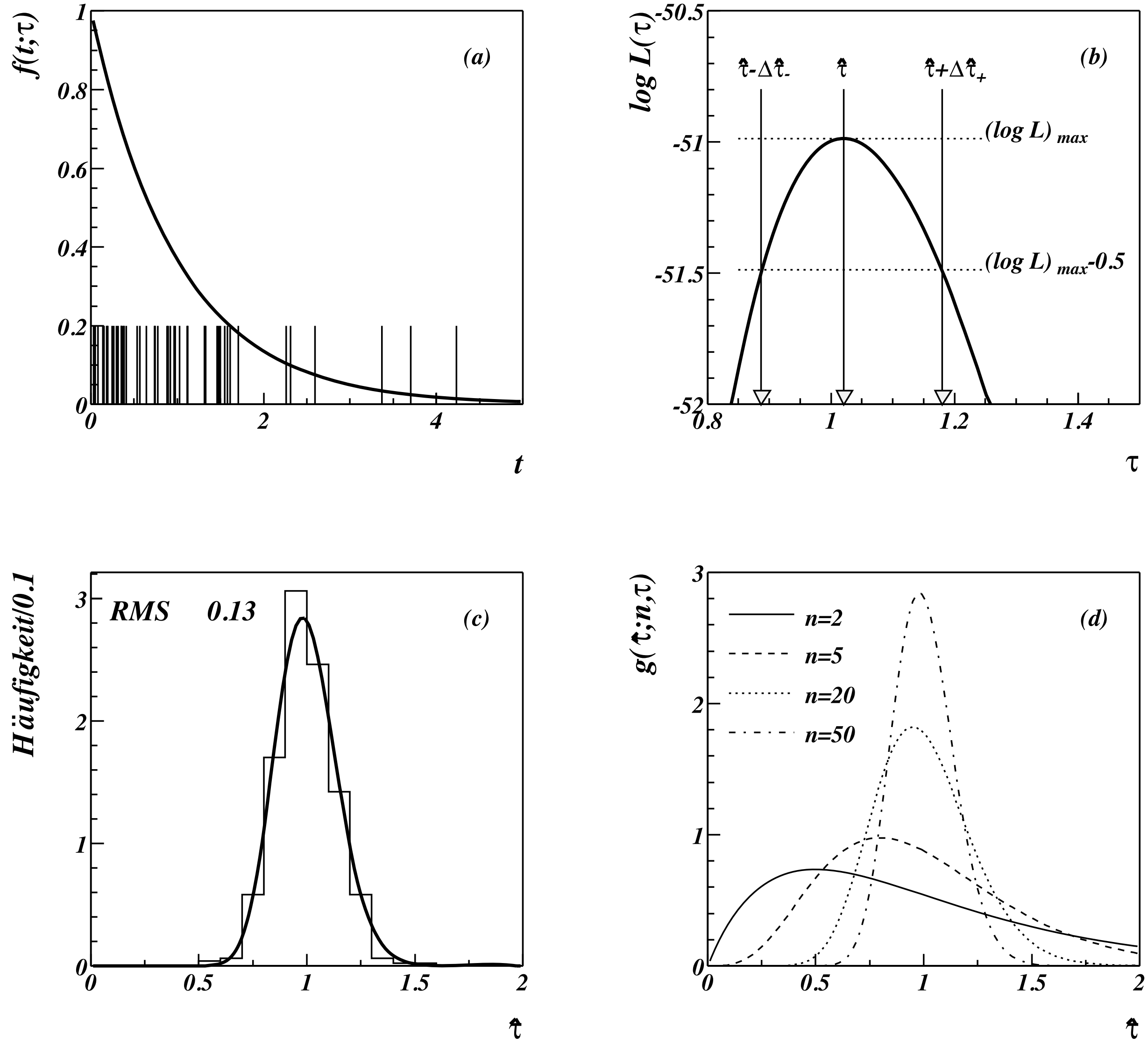
Analytisch:

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

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

MC-Methode:

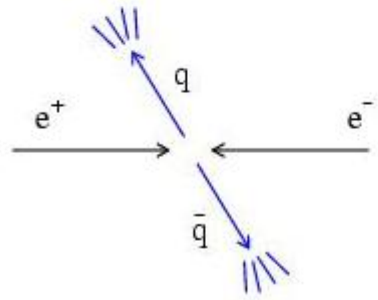
$$0.13,$$



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:

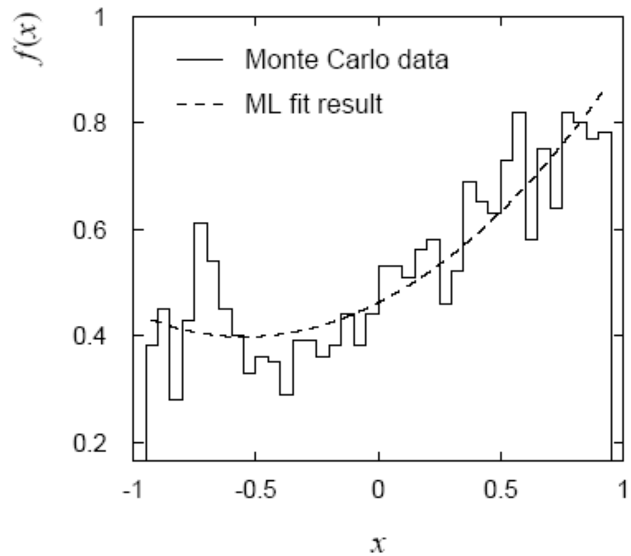
$$\int_{x_{\min}}^{x_{\max}} f(x;\alpha,\beta)\,dx = 1\ .$$

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

$$\widehat{\alpha} = 0.508$$
$$\widehat{\beta} = 0.47$$

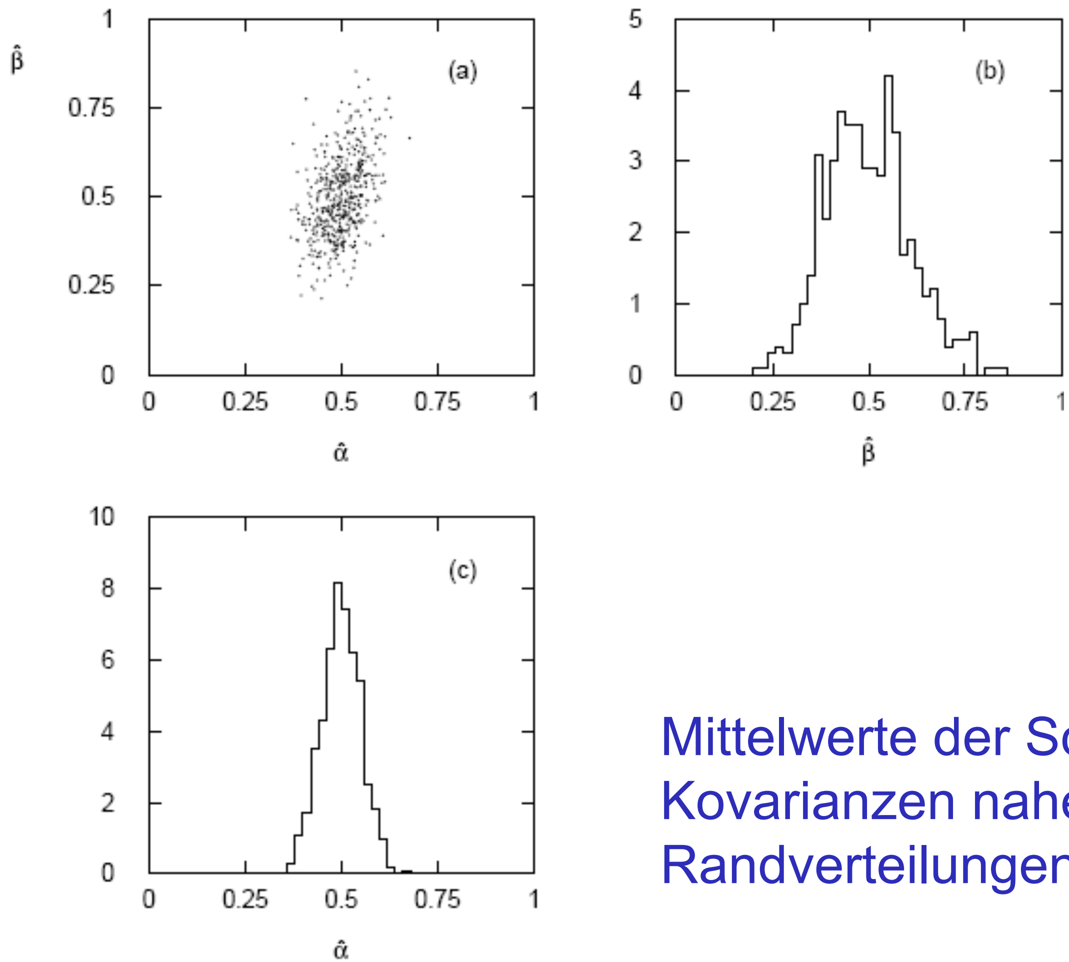


Bemerkung: Kein Binning der Daten im Fit.
Histogramm nur für Visualisierung

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

$$\widehat{\sigma}_{\widehat{\alpha}} = 0.052 \qquad \mathrm{cov}[\widehat{\alpha}, \widehat{\beta}] = 0.0026$$
$$\widehat{\sigma}_{\widehat{\beta}} = 0.11 \qquad r = 0.46$$

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



$$\begin{aligned}\overline{\hat{\alpha}} &= 0.499\\ s_{\hat{\alpha}} &= 0.051\\ \overline{\hat{\beta}} &= 0.498\\ s_{\hat{\beta}} &= 0.111\\ \widehat{\mathrm{cov}}[\hat{\alpha},\hat{\beta}] &= 0.0024\\ r &= 0.42\end{aligned}$$

Mittelwerte der Schätzer ~ wahre Werte
Kovarianzen nahe an früheren Abschätzungen.
Randverteilungen ungefähr gaussförmig.

Die "ln L_{max} - ½"-Kontur

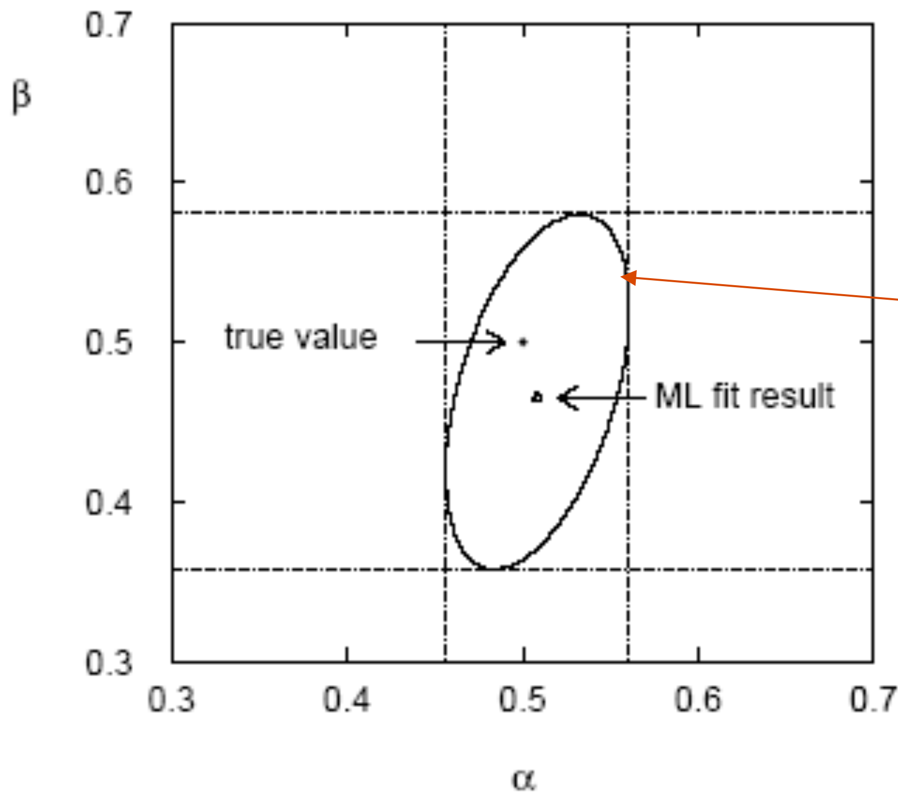
Für großen Stichprobenumfang *n*, wir ln *L* Paraboloid (quadratische Form) in der Nähe des Maximums:

$$\ln L(\alpha,\beta) \approx \ln L_{\text{max}} - \frac{1}{2(1-\rho^2)}\left[\left(\frac{\alpha-\widehat{\alpha}}{\sigma_{\widehat{\alpha}}}\right)^2 + \left(\frac{\beta-\widehat{\beta}}{\sigma_{\widehat{\beta}}}\right)^2 - 2\rho\left(\frac{\alpha-\widehat{\alpha}}{\sigma_{\widehat{\alpha}}}\right)\left(\frac{\beta-\widehat{\beta}}{\sigma_{\widehat{\beta}}}\right)\right]$$

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

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

Kovarianzen aus "ln *L*"- Kontur



Die (α, β)-Ebene für die erste Messung

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

→ Tangenten an Kontur

ergeben Standardabweichungen

→ Winkel der Hauptache 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).

Erweiterte ML-Methode

Bisher haben wir nur die Form der WDF f(x,θ) verwendet.

Nicht auch die die absolute Normierung.

Betrachte Anzahl der Beobachtungen als Poisson-verteilte ZV

Das Ergebnis eines Experimentes ist dann: n, x_1, ..., 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 ln-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 θ 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 → kleinere Varianz für $\hat{\vec{\theta}}$

Wenn ν nicht von θ 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.)

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 analystisch berechnet sondern nur durch MC-Simulation (z.B. Berücksichtigung von Akeptanz- und Auflösungseffekten (→Faltung))

Stichprobe liefert Histogramm mit Einträgen:

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

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

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

ML-Methode für Histogramme

Bei analytischer bekannter WDF der Theorie ergeben sich νi

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

oder direkt aus dem Histogramm der Simulation für die WDF.

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

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

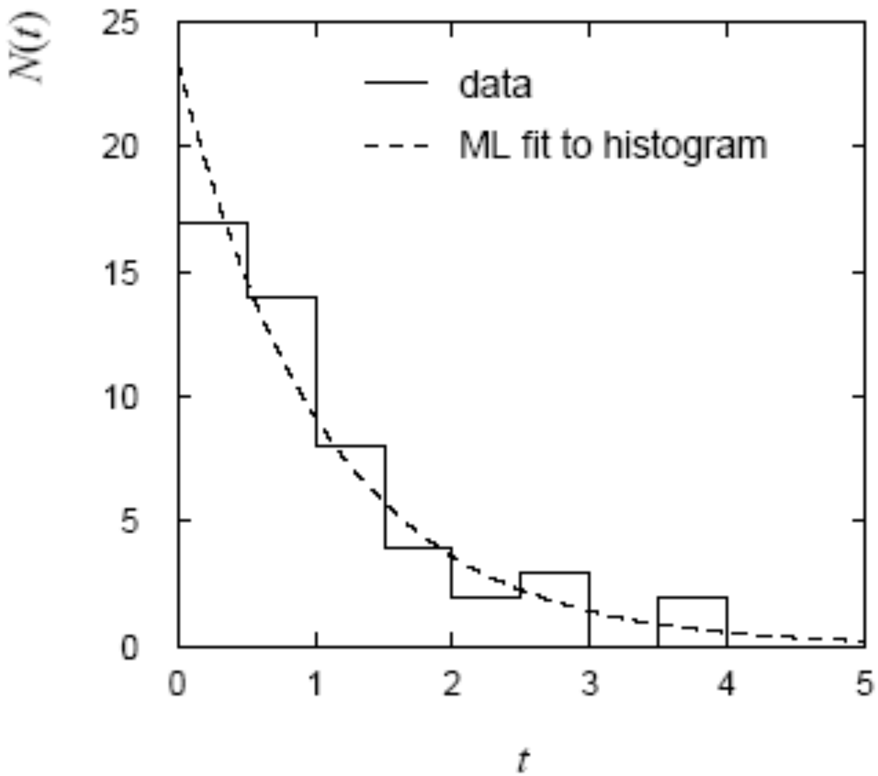
Die log-Likelihood-Funktion ergibt sich zu:

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

C=const.=Terme unabhängig von θ_i

ML-Beispiel für Histogram

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 → normale ML-Methode

Einteilung in Bins führt zu Informationsverlust → Varianz größer

Erweiterte ML-Beispiel für Histogram

Bisher nur Form der Histogramme.
Zahl der Einträge in Theorie durch Beobachtung fixiert (mit $\nu_{tot}=n_{tot}$)

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

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

Die gemeinsame WDF ergibt sich zu:

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

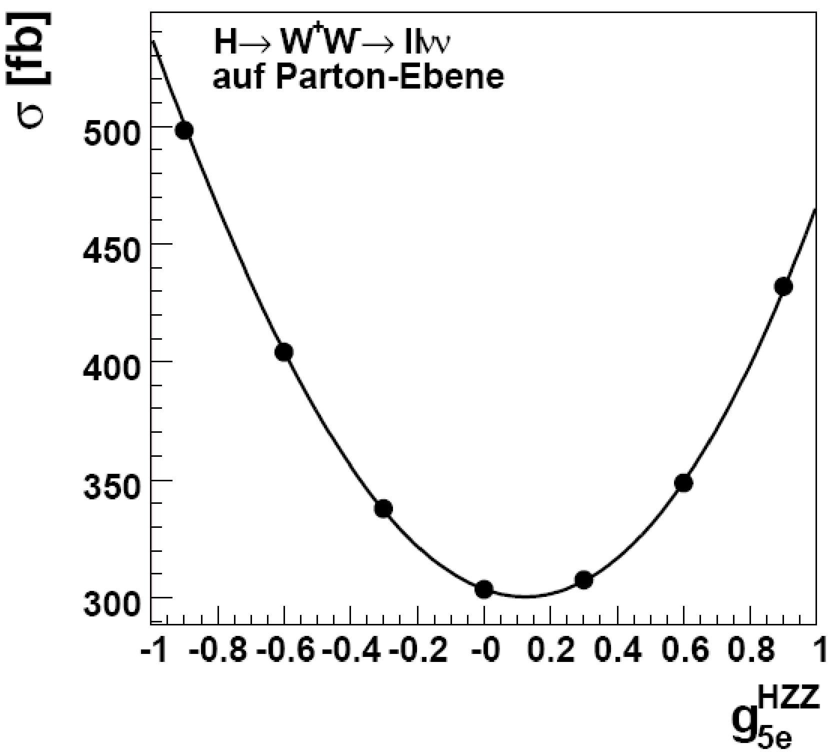
= Produkt von N unabhängigen Poisson-WDF in jedem Bin.

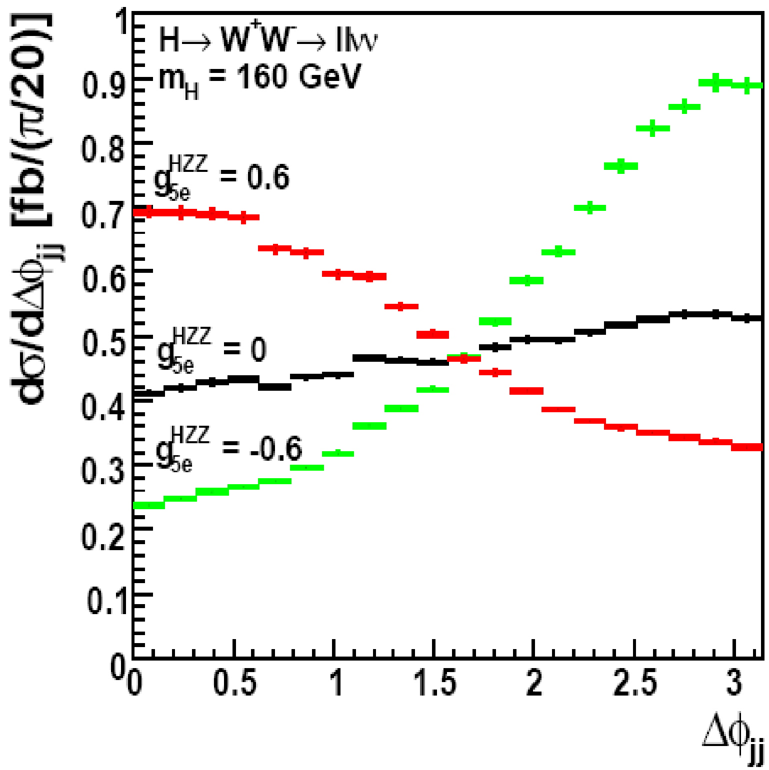
Und die erweiterte Likelihoodfunktion ist:

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

Erweiterte ML-Methode: Beispiel

Ziel: Messung von g.





Anzahl der beobachteten Ereignisse hängt quadratisch von g ab.

Form hängt nicht trivial von g ab.

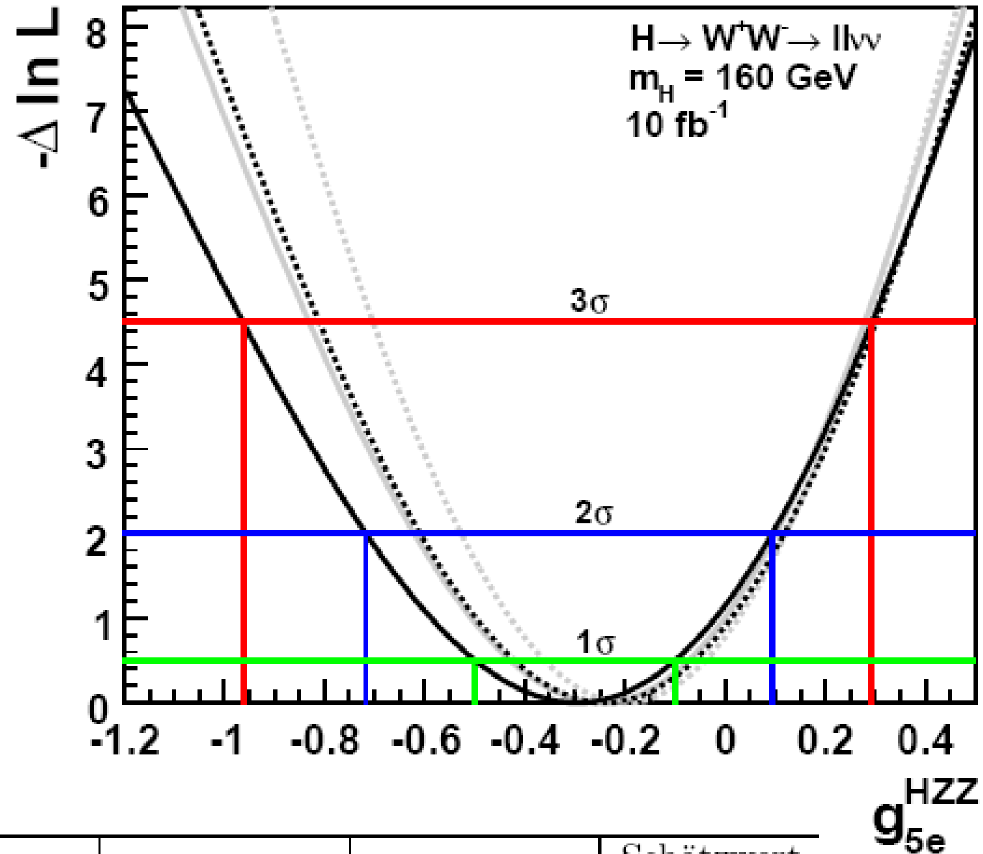
Erweiterte ML-Methode: Beispiel-Fit Ergebnis

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 $-\Delta \ln L$	1σ-Intervall	2σ-Intervall	3σ-Intervall	Schätzwert für σ
$H \to W^+W^- \to ll\nu\nu$						
$10\,fb^{-1}$	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

Kombination von Messungen mit ML-Methode:

Annahme: zwei Observable *x* und *y* mit unterschiedlichen Abhängigkeiten vom Parameter θ, der geschätzt werden soll.

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

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 θ 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 $g_y(\hat{\theta}_x;\theta)$

Die Messungen können dann kombiniert werden mittels der Likelihoodfunktion

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

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:

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