Statistische Methoden der Datenanalyse

Wintersemester 2012/2013

Albert-Ludwigs-Universität Freiburg

Prof. Markus Schumacher, Dr. Stan Lai

Physikalisches Institut Westbau 2 OG

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

stan.lai@cern.ch

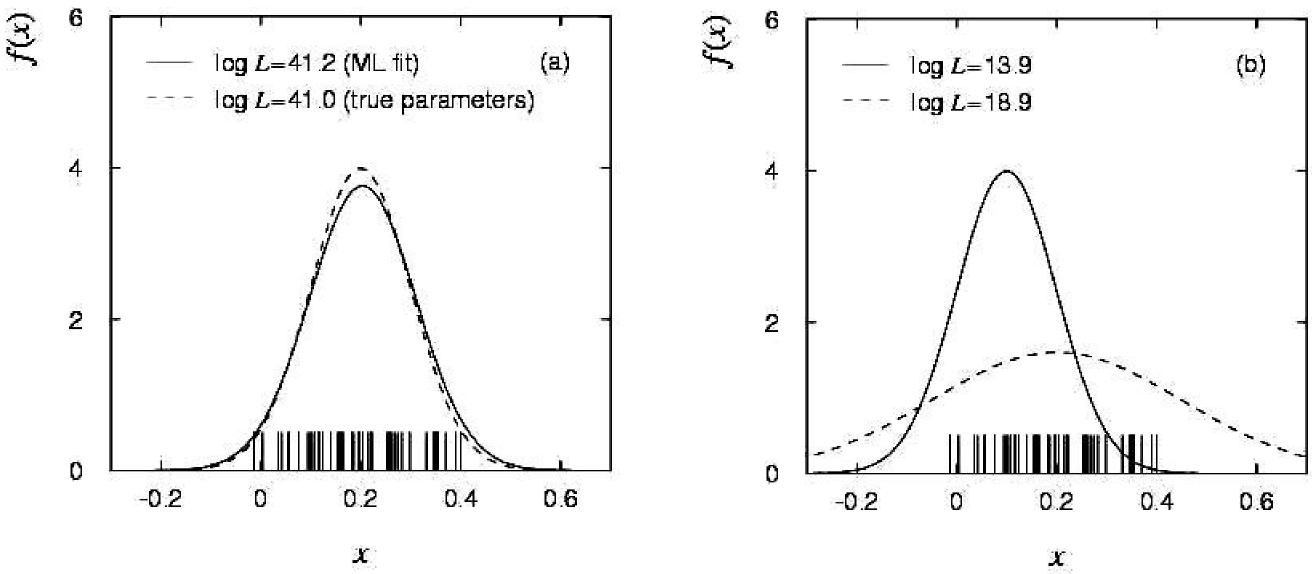
http://terascale.physik.uni-freiburg.de/lehre/ws_1213/statmethoden_ws1213

Kapitel 6

Die “Maximum Likelihood”-Methode

"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 fen Parameterwert, der die Likelihoodfunktion maximiert.

Für ML-Schätzer gibt es keine Garantie, dass sie "optimale" Eigenschaften besitzen. Aber in der Praxis sind sie sehr gut.

"Maximum Likelihood"-Schätzer: Formal

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

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

Oft praktischer die log-Likelihood-Funktion zu verwenden

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

(Summieren ist einfacher als multiplizieren.)

ML-Schätzer gegeben durch die Gleichung:

$$\left.\frac{\partial \log L}{\partial \theta}\right|_{\hat{\theta}} = 0$$

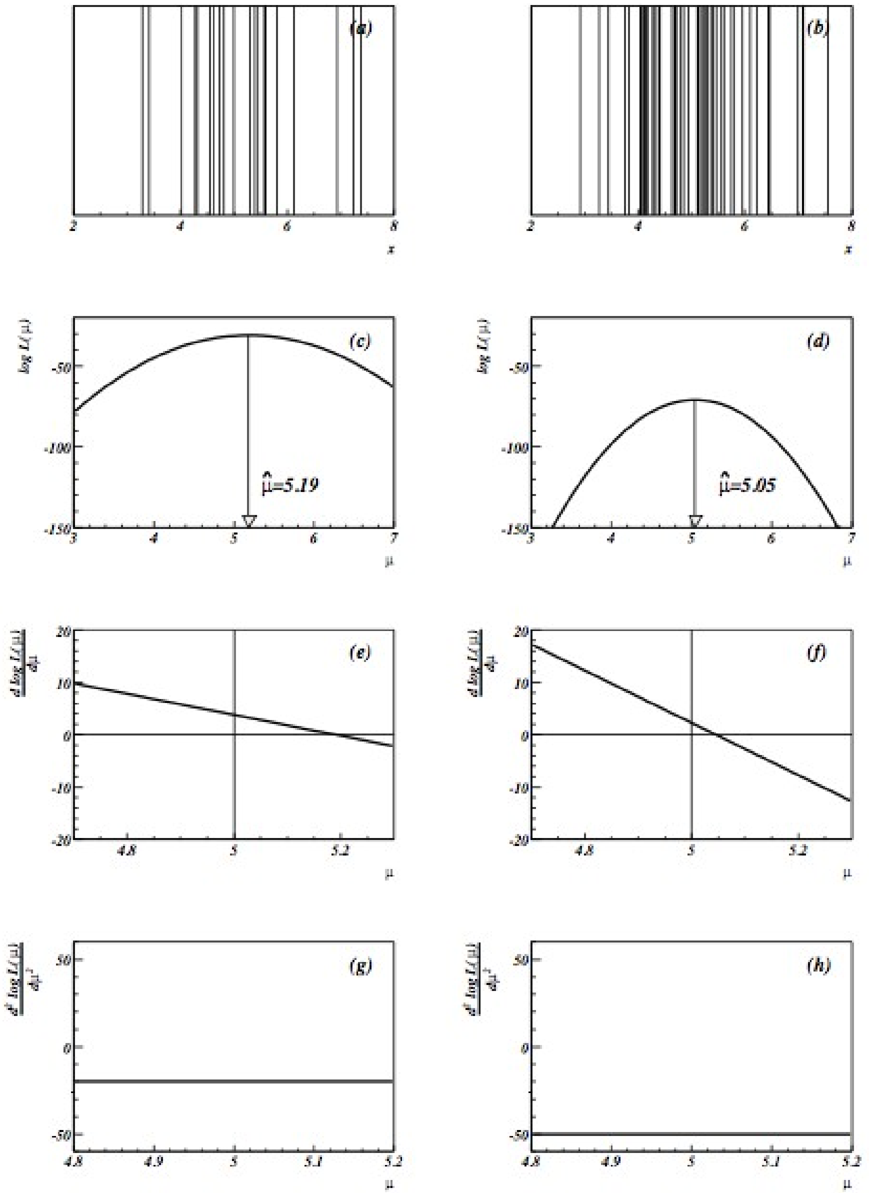
Für mehrere Parameter:

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

“Maximum Likelihood”-Schätzer: Gauss-Bsp.

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

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



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

$$I(\mu) = 20$$

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

$$I(\mu) = 50$$

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 Maximal-wert 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)

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

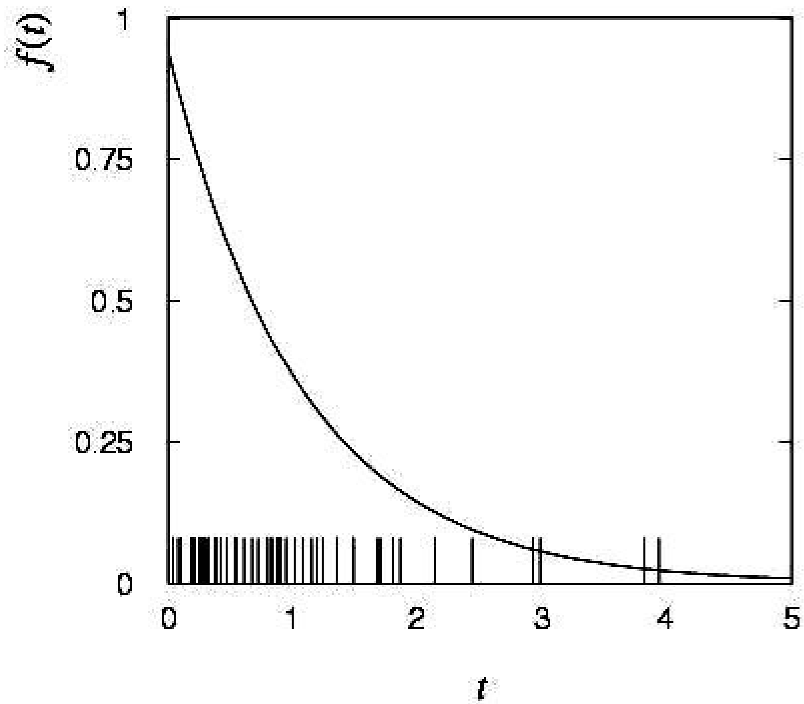
Bestimmung des Maximums

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

Liefert den ML-Schätzer:

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

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



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

Wir bestimmen den ML-Schätzer zu:

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

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

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

Vergleich mit der Schranke minimaler Varianz:

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

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

Also ML-Schätzer ist für dieses Problem 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: $$\left.\frac{\partial L}{\partial \theta}\right|_{\hat{\theta}} = \left.\frac{\partial L}{\partial a}\right|_{a(\hat{\theta})} \left.\frac{\partial a}{\partial \theta}\right|_{\hat{\theta}}$$

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

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

D.h.: der ML-Schätzer für den transformierten Parameter ist gleich der Transformation 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 $\to 0$ für $n \to \infty$)

Beispiel für ML: Parameter der Gauss-WDF

Betrachte n unabhängige Messungen $x_1, \ldots, x_n$,
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:

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

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

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

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.

Erwartunsgtreue: 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

Erwartunsgtreue: ML-Schätzer ist asymptotisch erwartunsgtreu, wenn Varianz endlich ist, da er konsitent 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-Likelihhoodfunktion geht gegen Parabel.

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

Varianz für Schätzer: Analytische Methode

In weniger Fällen kann die Varianz analytisch berechnet werden.

Benötigt Berechnung von:

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

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

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

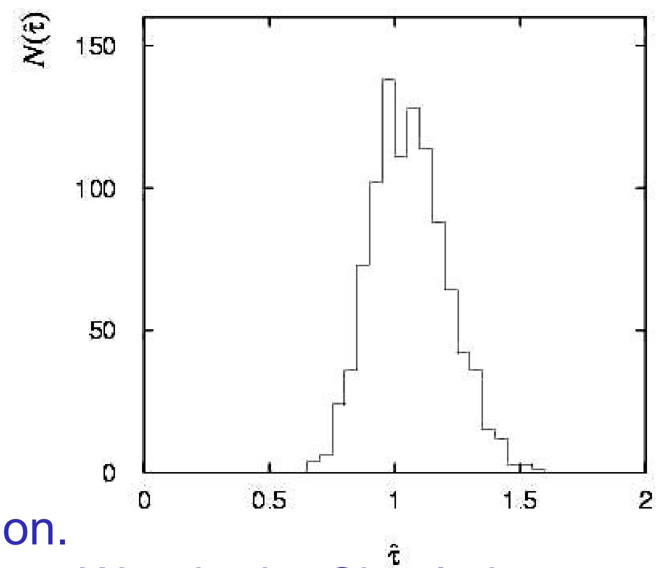
Nachdem wir den Schätzwert für den Parameter bestimmt haben müssen wir nun den statistischen Fehler des Schätzers bestimmen, i.e., wie weit die Verteilung der Schätzwerte wäre, wenn wir die gesamte identische Messung sehr oft wiederholen würden.

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 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):

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

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

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

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

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

Varianz des Schätzers: Graphische Methode

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

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

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

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

i.e.,

$$\ln L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) \approx \ln L_{\max} - \frac{1}{2}$$

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

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

Beispiel: Varianzschätzung durch graph. Methode

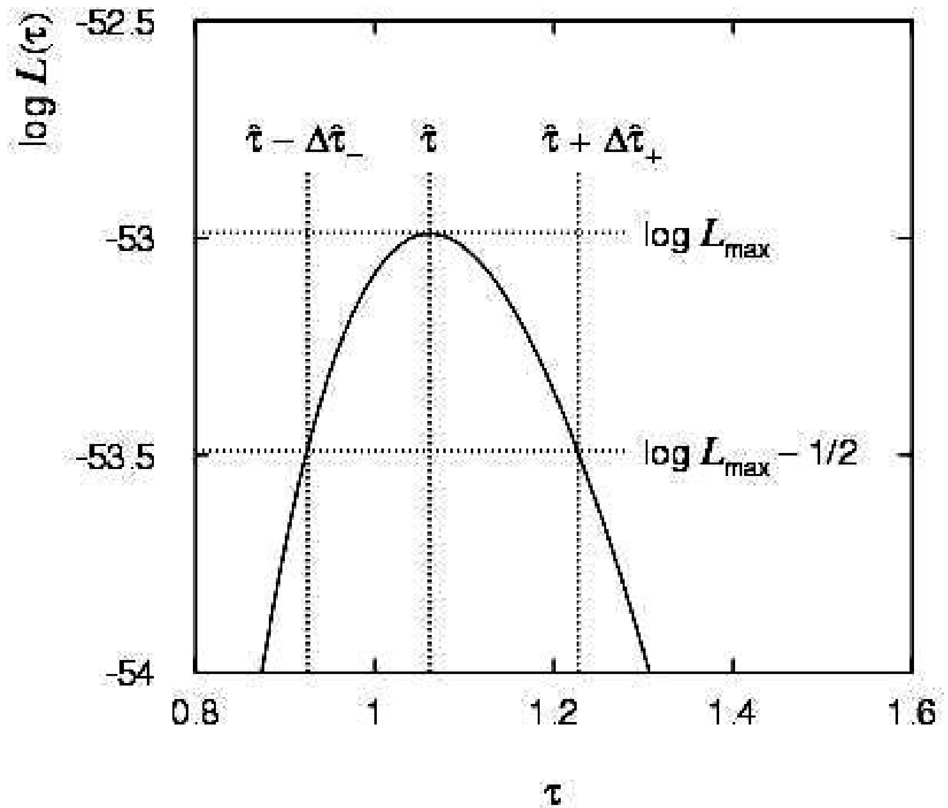
ML-Beispiel mit Exponential-WDF:

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

$$\Delta\hat{\tau}_- = 0.137$$

$$\Delta\hat{\tau}_+ = 0.165$$

$$\hat{\sigma}_{\hat{\tau}} \approx \Delta\hat{\tau}_- \approx \Delta\hat{\tau}_+ \approx 0.15$$



In L nicht ganz parabelförmig, da endliche Stichprobe (n = 50).
→ gib asymmetrische Fehler an.

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