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

Vorlesung basiert in weiten Teilen auf der von Glen Cowan.

http://terascale.physik.uni-freiburg.de/lehre/ws_1112/statmethoden_ws1112

Kapitel 4

Grundlagen der Parameterschätzung

Grundgesamtheit, Stichprobe, Statistik, Schätzer

Betrachte ZV x mit WDF $f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}$ oft abhängig von Parametern

Grundgesamtheit = Menge möglicher Ergebnisse beschrieben durch f(x)
Stichprobe vom Umfang n = Satz von n unabhängigen Messungen der Zv x

$$\vec{x} = (x_1, \ldots, x_n)$$

Ziel: Ableitung von Eigenschaften von f(x) aus der Stichprobe

Statistik: beliebige Funktion der Stichprobenwerte der ZV

Schätzer: Statistik um Eigenschaften der WDF zu bestimmen

a) Form der WDF unbekannt → Schätzwerte für Erwartungswert, Varianz, …
b) Form der WDF bekannt → Schätzwerte für Parameter $\hat{\theta}(\vec{x})$

Schätzer = Funktioneller Zusammenhang, Schätzwert = numerischer Wert
beide werden mit “Hut ^” beschrieben

Geminsame WDF der Stichprobe

Betrachte n Beobachtungen der ZV x: $x_1, \ldots, x_n$,
wobei x der WDF $f(x; \theta)$ folgt.

Unter den Annahmen:

1) Messungen sind unabhängig
2) WDF/Grundgesamtheit ändert sich zwischen Beobachtung nicht

folgt die gemeinsame WDF der Stichprobe:

$$f(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

Achtung! Ist nicht immer erfüllt in der Praxis!

Gegenbeispiele: - Lotterie (Ziehen von Kugeln ohne zurücklegen)
- Längenmessung von Stab bei Temperaturschwankungen

Die Likelihoodfunktion

Das Ergebnis eines Experimente (Satz von Messungen) $\vec{x} = (x_1, \ldots, x_n)$ sei eine Menge von Zahlen ***x***, und die gemeinsame WDF der Stichprobe ist eine Funktion/Statistik, welche von den Parametern der WDF ***θ***: abhängt:

$$f(\vec{x}; \vec{\theta})$$

Nun werte diese Funktion mit den Werten der Stichprobe aus und betrachte sie als Funktion der Parameter:

$$L(\vec{\theta}) = f(\vec{x}; \vec{\theta})$$ Likelihoodfunktion

(***x*** konstant)

Im Falle von unabhängigen, identischen Messungen gilt:

$$L(\vec{\theta}) = \prod_{i=1}^{n} f(x_i; \vec{\theta})$$ (x_i konstant)

Schätzer für den Erwartungswert μ von f(x)

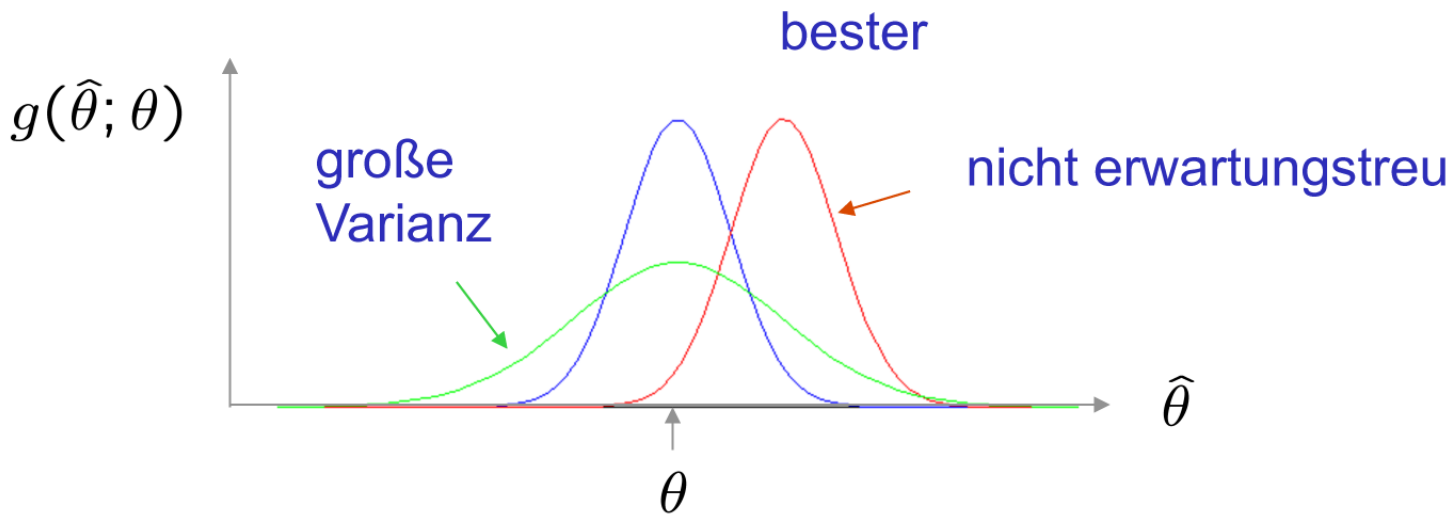
Mögliche Schätzer für den Erwartungswert der WDF der Grundgesamtheit

$\widehat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	den Mittelwert der Stichprobe
$\widehat{\mu} = \frac{1}{10}\sum_{i=1}^{10} x_i$	den Mittelwert der ersten 10 Punkte der Stichprobe
$\widehat{\mu} = \frac{1}{n-1}\sum_{i=1}^{n} x_i$	$n/(n-1)$ mal den Mittelwert der Stichprobe
$\widehat{\mu} = 42$	
$\widehat{\mu} = (\min(x_i) + \max(x_i))/2$	Mittelwert des größten und kleinsten Wertes
$\widehat{\mu} =$ Median der Stichprobe	

Fragen: - welcher Schätzer ist gut, der “Beste”?
- welche Kriterien sollte ein guter Schätzer erfüllen?
- wie findet man den optimalen Schätzer für ein Problem?

Eigenschaften von Schätzern

Wenn mir das Experiment (aus m Messungen bestehend) oft wiederholen.
Würden die Schätzwerte einer WDF folgen:



Wir wollen kleinen (oder Null) Bias (systematischen. Fehler): $b = E[\hat{\theta}] - \theta$

$\rightarrow$ Mittelwert der wiederholten Messungen sollte = wahren Wert sein

Und wir wollen kleine Varianz (statistischer Fehler): $V[\hat{\theta}]$

$\rightarrow$ kleiner Bias und keine Varianz sind i.a. gegenläufige Kriterien

Eigenschaften von Schätzern

Aus der Informationstheorie lässt sich zeigen, dass es eine untere Schranke für die Varianz eines Schätzers für einen Parameter gibt

$$V\left[\hat{\theta}\right] \geq \frac{\left(1+\frac{\partial b}{\partial \theta}\right)^2}{E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]}. \qquad V\left[\hat{\theta}\right] \geq \frac{\left(1+\frac{\partial b}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right]}.$$

Schranke Minimaler Varianz (SMV)

Rao-Cramér-Frechet-Ungleichung

Information nach R.A Fisher: $I(\theta) \equiv E\left[\left(\frac{\partial \log \mathcal{L}}{\partial \theta}\right)^2\right] = E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]$

$\rightarrow$ je größer die Information, desto kleiner der statistische Fehler.

Kriterien für gute Schätzer

Konsistenz $\widehat{\theta}^{(n)} \xrightarrow{n\to\infty} \theta$

für jedes $\varepsilon > 0$ gilt $\lim_{n\to\infty} P(|\widehat{\theta}^{(n)} - \theta| > \varepsilon) = 0$

Verzerrung (Bias) $b^{(n)} = E[\widehat{\theta}^{(n)}] - \theta$

möglichst keinen Bias d.h. Schätzer ist erwartungstreu.
Konsistente Schätzer mit endlicher Varianz sind asymptotisch (n→unendlich) erwartungstreu

Effizienz

$$\text{Effizienz}\left[\widehat{\theta}^{(n)}\right] = \frac{SMV}{V\left[\widehat{\theta}^{(n)}\right]}$$

Effizienz sollte nahe an “1” sein.

Robustheit Varianz des Schätzers ist unabhängig von der WDF der Grundgesamtheit

Ein Schätzer für den Erwartungswert

Parameter: $\mu = E[x]$

Schätzer: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \overline{x}$ ('Stichprobenmittelwert')

Man kann zeigen: ist konsistent

$b = E[\hat{\mu}] - \mu = 0$ ist erwartungstreu

ist effizient für Gauss-WDF (aber nicht für alle WDFs)

Der Fehler auf den Schätzer für den Erwartungswert ist gegeben durch

$$V[\hat{\mu}] = \frac{\sigma^2}{n} \qquad \left(\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}} \right)$$

Arith. Mittelwert für Gauss-WDF

Die log-Likelihood-Funktion lautet:

$$\begin{aligned}\log \mathcal{L} &= \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)\right) \\ &= -n\log\left(\sigma\sqrt{2\pi}\right) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}\end{aligned}$$

$$\frac{\partial^2 \log \mathcal{L}}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

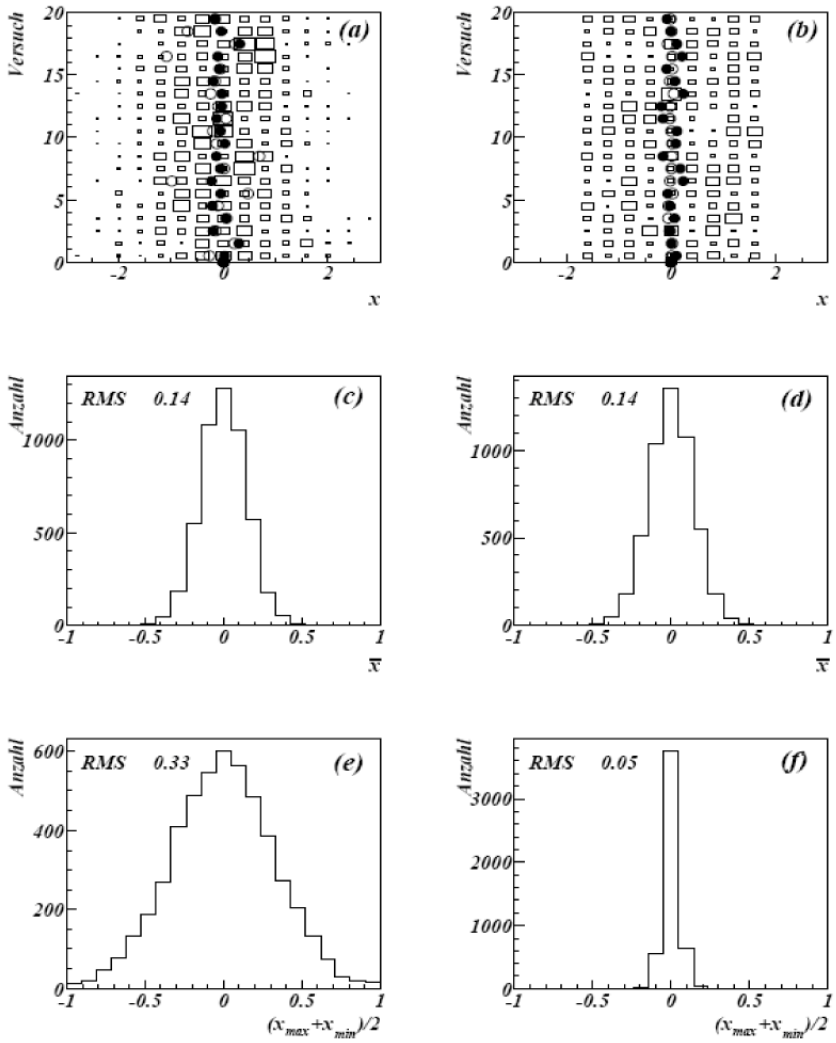
Die Schranke minimaler Varianz ergibt sich zu:

$$\mathrm{SMV} = \frac{1}{E\left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right]} = \frac{\sigma^2}{n} = \frac{V[x]}{n} = V[\overline{x}]$$

D.h. der Schätzer hat eine Effizienz von 100%

Vergleich von 2 Schätzern für den Erwartungswert

Gauss-WDF | Gleichverteilung



$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \overline{x}$$

robust

$$\hat{\mu} = (\min(x_i) + \max(x_i))/2$$

effizienter für Gleich-verteilung

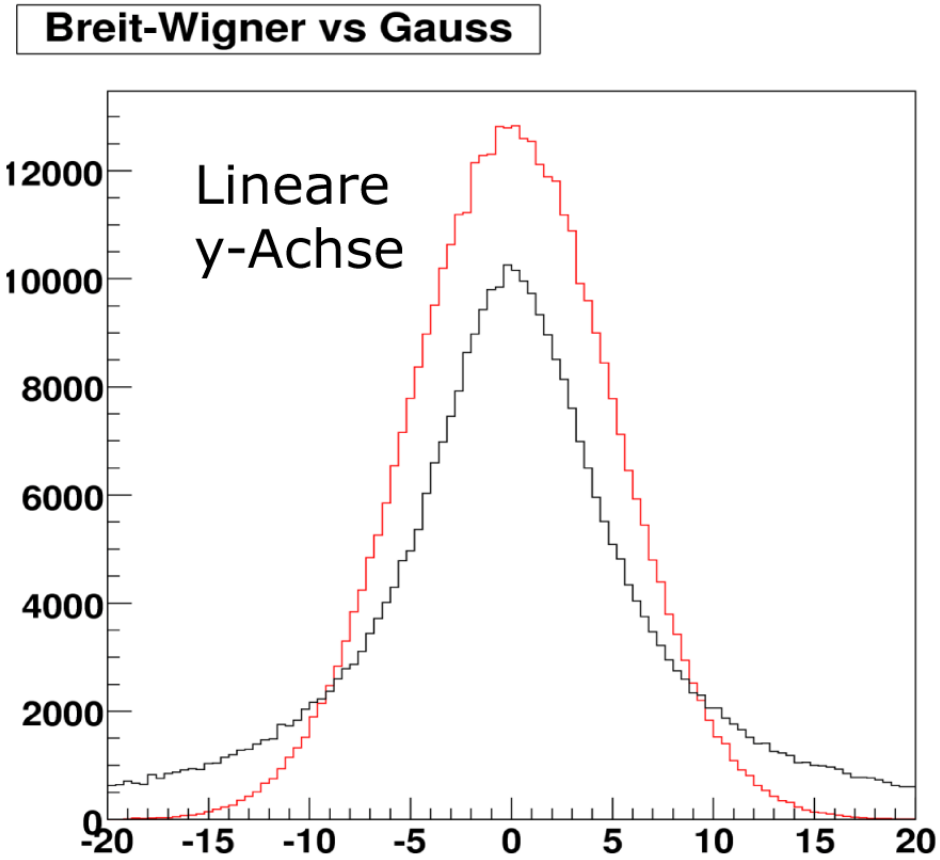
Robustheit: „truncated mean“

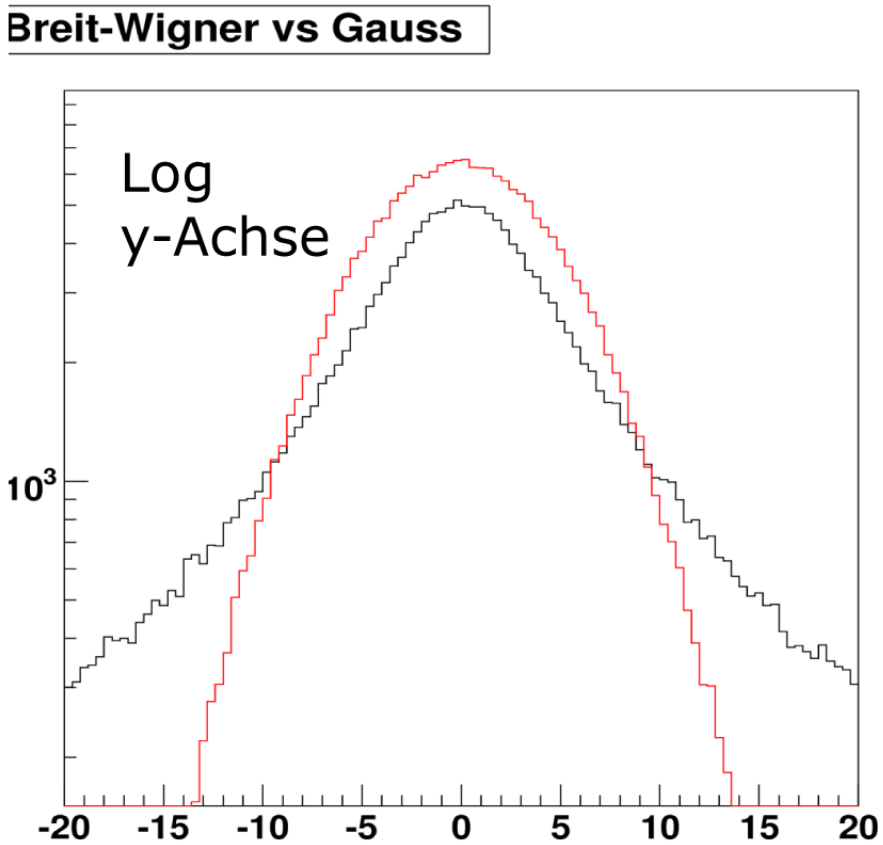
Gaussian pdf Breit-Wigner pdf

Ungefähr gleiche Varianz in Wertebereich

Aufgabe: gute Schätzer finden für den Symmetriepunkt der Verteilung (x=0)

Arithmetischer Mittelwert?





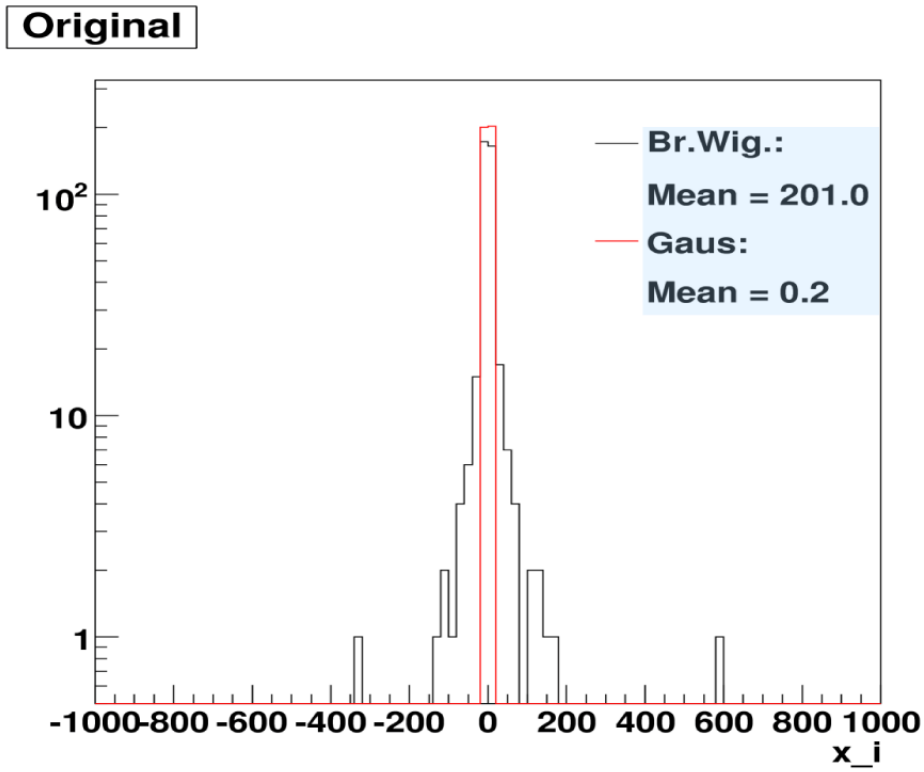
Robustheit: „truncated mean“

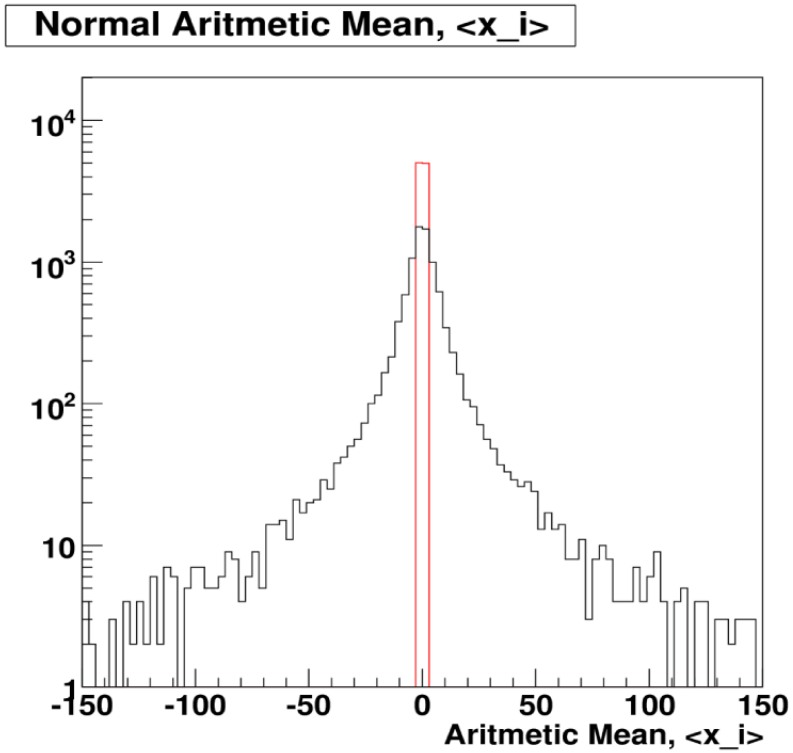
Stichprobe von 101 Zufallszahlen x*i* von jeder PDF.

Breit Wigner: sehr grosse Werte $|x_i|$ führt zu großen Werte für arithmetischen Mittelwert $\langle x_i \rangle$.

100 000 Stichproben
==> WDF für arithmetischen Mittelwert, $\langle x_i \rangle$

Breit Wigner: viel größere Varianz als für Gauß wegen Stichproben mit großen Einzelwerten x_i (siehe unten, links)





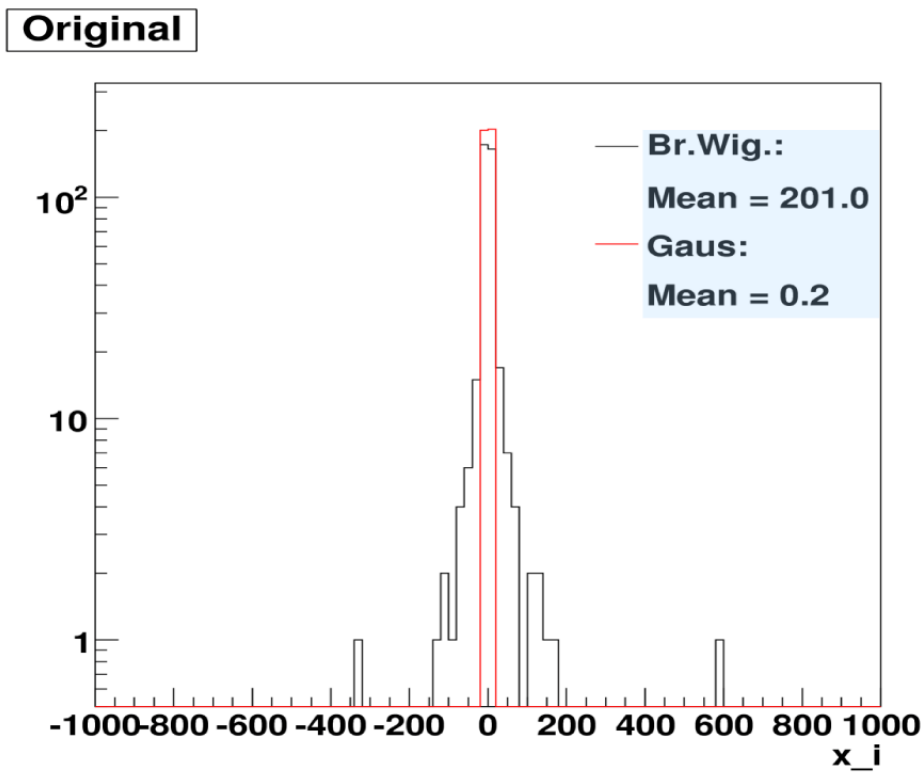
Robustheit: „truncated mean“

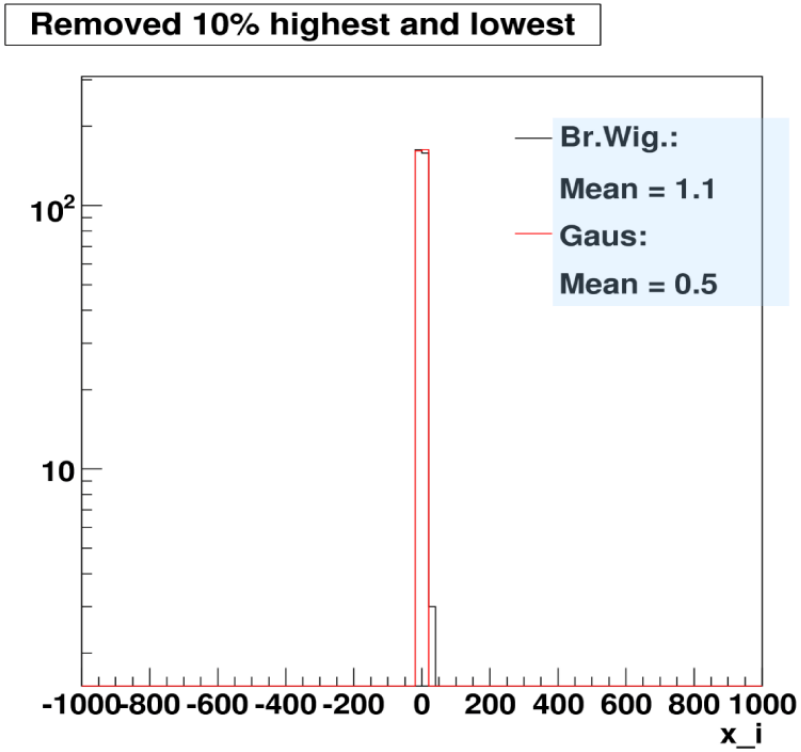
Stichprobe von 101 Zufallszahlen x_i von jeder PDF.

Breit Wigner: sehr grosse Werte $|x_i|$ führt zu grosse Werte für aritmetische Mittlewert $<x_i>$.

Verbesserte Schätzer für Symmetriepunkt für Breit Wigner: 10% grösste und kleinste Werte für x_i ignorieren, dann $<x_i>$ berechnen (Englisch: “truncated mean”)

==> viel kleinere Varianz für $<x_i>$ für Breit Wigner

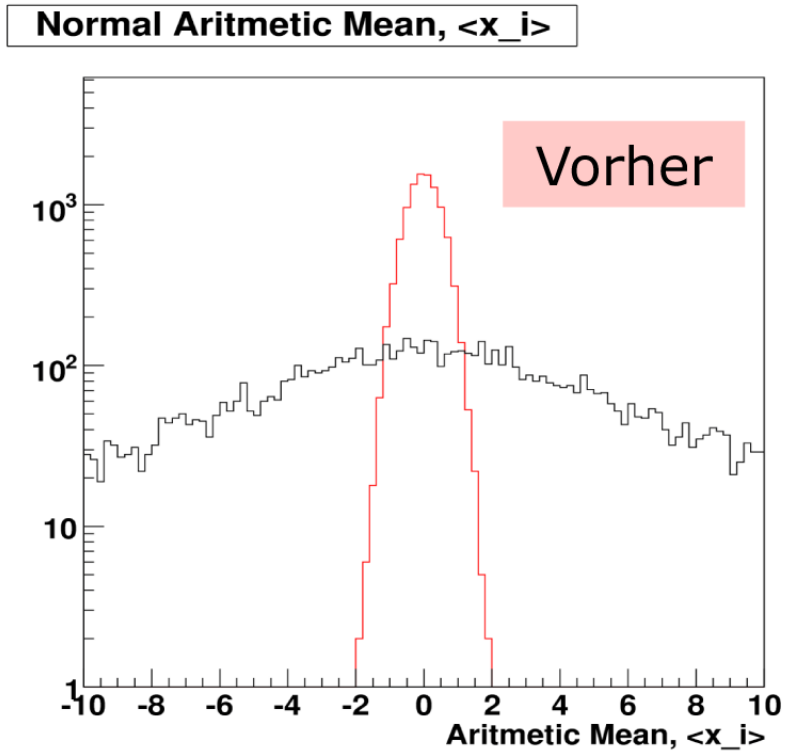




Robustheit: „truncated mean“

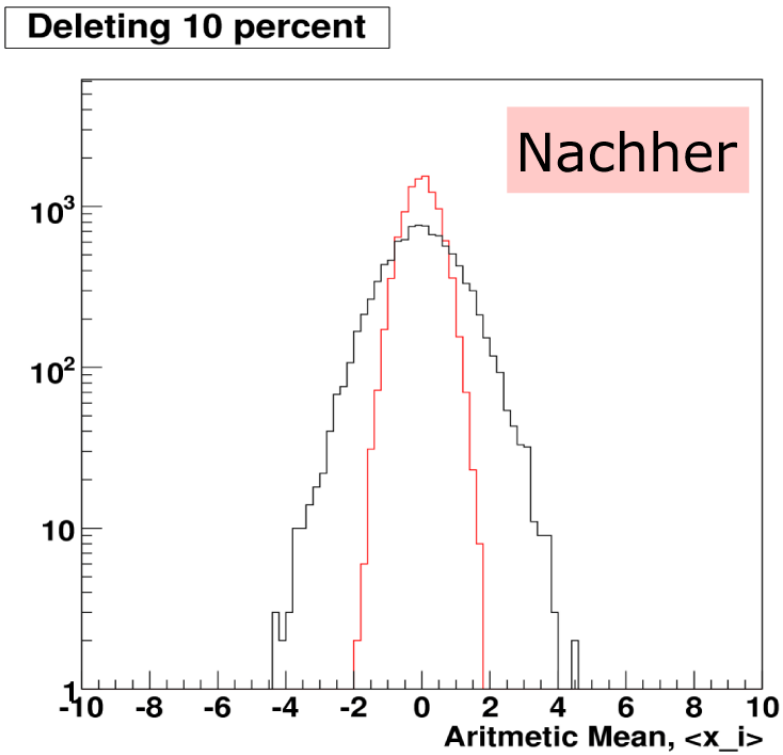
WDF für aritmetische Mittelwert, $<x_i>$

NORMAL

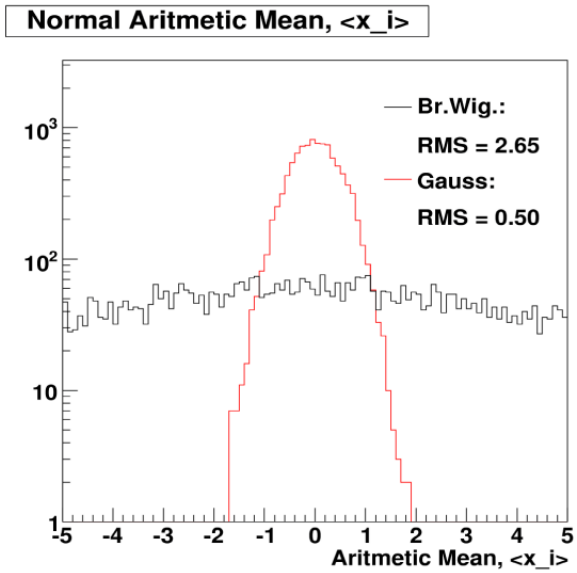


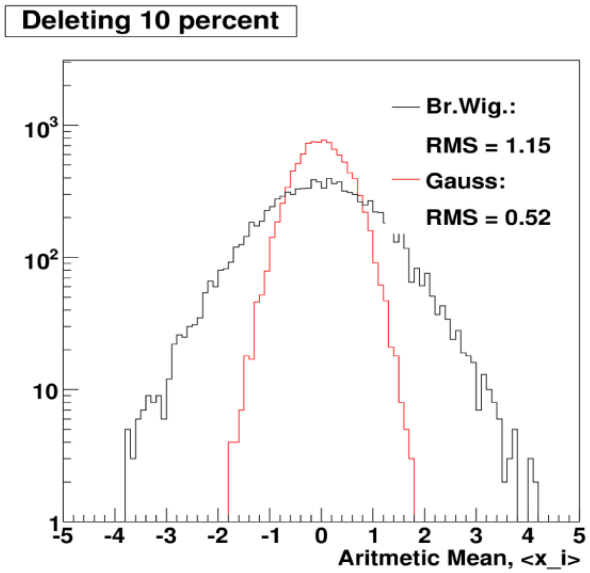
WDF für aritmetische Mittelwert, $<x_i>$

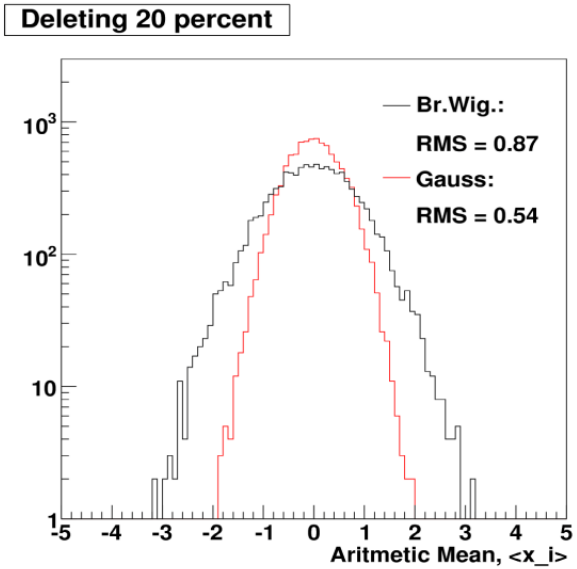
10% größte und kleinste Werte ignoriert

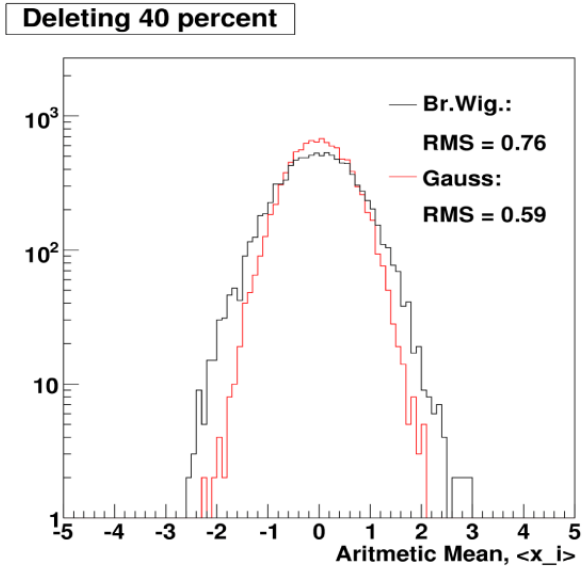


Robustheit: „truncated mean“

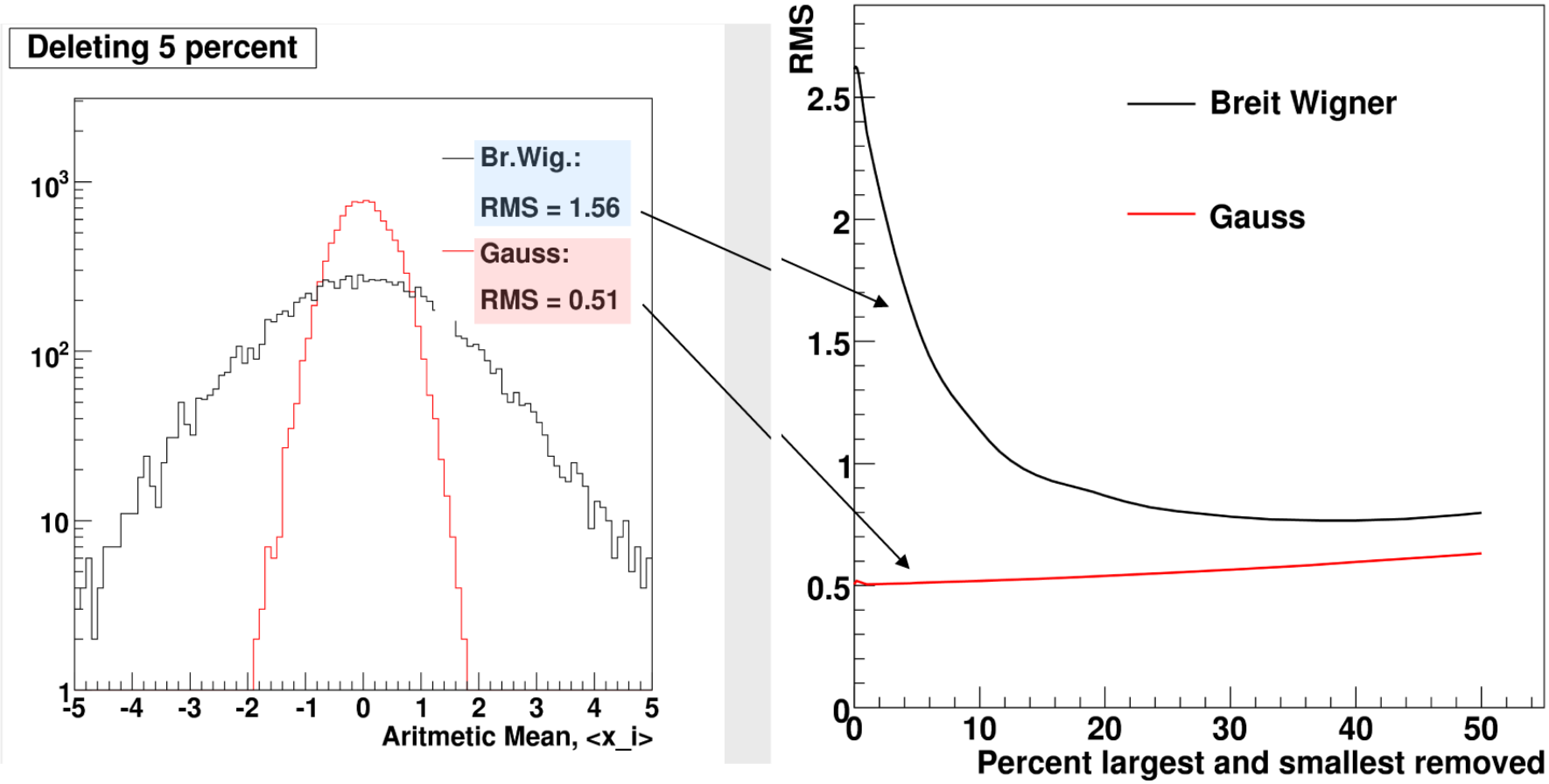








Robustheit: „truncated mean“



“truncated” Mittelwert kann den Fehler auf den Schätzer verbessern.
Hängt von Menge des Abschneidens und WDF ab.

Ein Schätzer für die Varianz

Parameter: $\sigma^2 = V[x]$

Schätzer: $$\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \equiv s^2$$ ('Stichprobenvarianz')

Man kann zeigen: ist konsistent (mit und ohne Besselkorrektur (*n/n*-1))

$$b = E[\widehat{\sigma^2}] - \sigma^2 = 0$$

erwartungstreu (ohne *n/n*-1 Korrektur nur asymptotisch erwartungstreu)

Der Fehler auf den Schätzer für die Varianz ist gegeben durch

$$V[\widehat{\sigma^2}] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-1}\mu_2\right) , \qquad \mu_k = \int (x-\mu)^k f(x)\, dx$$

Ein Schätzer für die Varianz

oder:

$$
\begin{aligned}
V[\sigma_S^2] &= V\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2\right] \\
&= \frac{V[x]^2}{(n-1)^2}V\left[\sum_{i=1}^{n}\frac{(x_i-\overline{x})^2}{V[x]}\right]
\end{aligned}
$$

Wenn Grundgesamtheit einer Gauss-WDF folgt, dann folgt der Ausdruck in der eckigen Klammer einer Chi-Quadrat-WDF mit n-1 FG, Deren Varianz 2(n-1) ist.

$$
V[\sigma_S^2] = \frac{2V[x]^2}{(n-1)} \quad \text{für gaußverteilte Grundgesamtheit}
$$

Ein Schätzer für die Kovarianz

Ein konsistenter und erwartungstreuer Schätzer für die Koarianzen ist:

$$\widehat{V}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{n}{n-1}\left(\overline{xy} - \overline{x}\,\overline{y}\right)$$

Schätzer für die Korrelationskoeffizienten sind :

$$\widehat{\rho} = \frac{\widehat{V}_{xy}}{\sigma_{S,x}\sigma_{S,y}} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}}$$

Für 2-dimensionale Gauss-WDF gilt:

$$\begin{aligned} E[\widehat{\rho}] &= \rho - \frac{\rho(1-\rho^2)}{2n} + \mathcal{O}(n^{-2}) \\ V[\widehat{\rho}] &= \frac{1}{n}(1-\rho^2)^2 + \mathcal{O}(n^{-2}) \end{aligned}$$

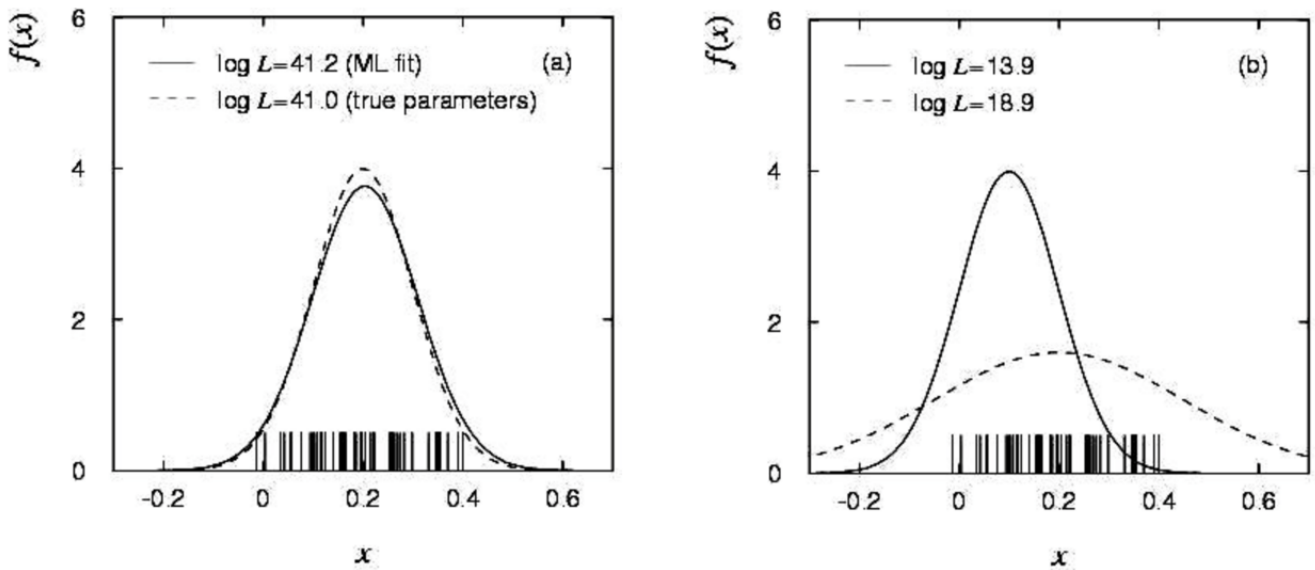
d.h. nur asymptotisch unverzerrt, obwohl $\widehat{V}_{xy}, \sigma_{S,x}, \sigma_{S,y}$ erwartungstreu sind.

Kapitel 5

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

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

Oft praktischer die log-Likelihood-Funktion zu verwenden

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

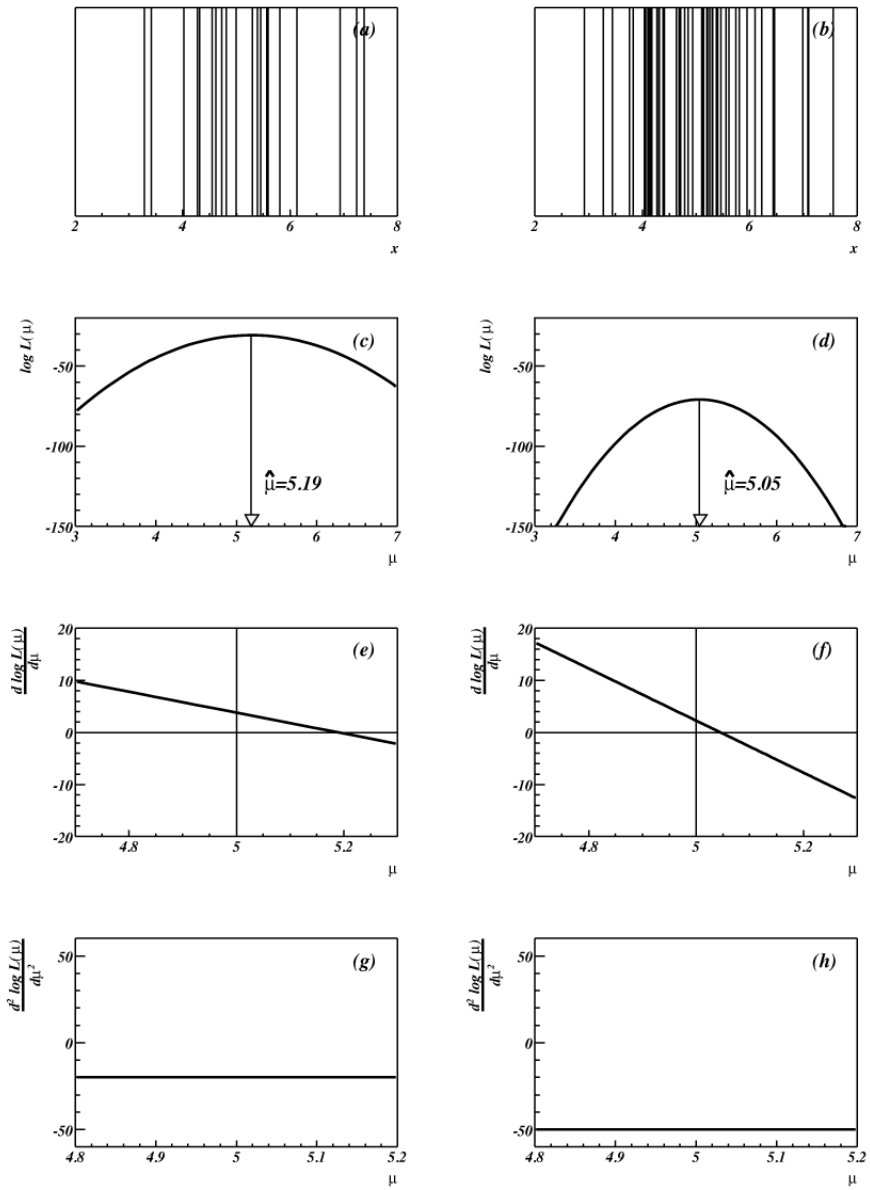
(Summieren ist einfacher als multiplizieren.)

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

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

“Maximum Likelihood”-Schätzer: Gauss-Bsp.

Gauss-WDF
μ=5
σ=1
Stichprobe
n= 20



Gauss-WDF
μ=5
σ=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 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 $0 = \left. \frac{\partial \log \mathcal{L}(\tau)}{\partial \tau} \right|_{\tau=\hat{\tau}} = -n \frac{1}{\hat{\tau}} + \frac{1}{\hat{\tau}^2} \sum_{i=1}^{n} t_i$

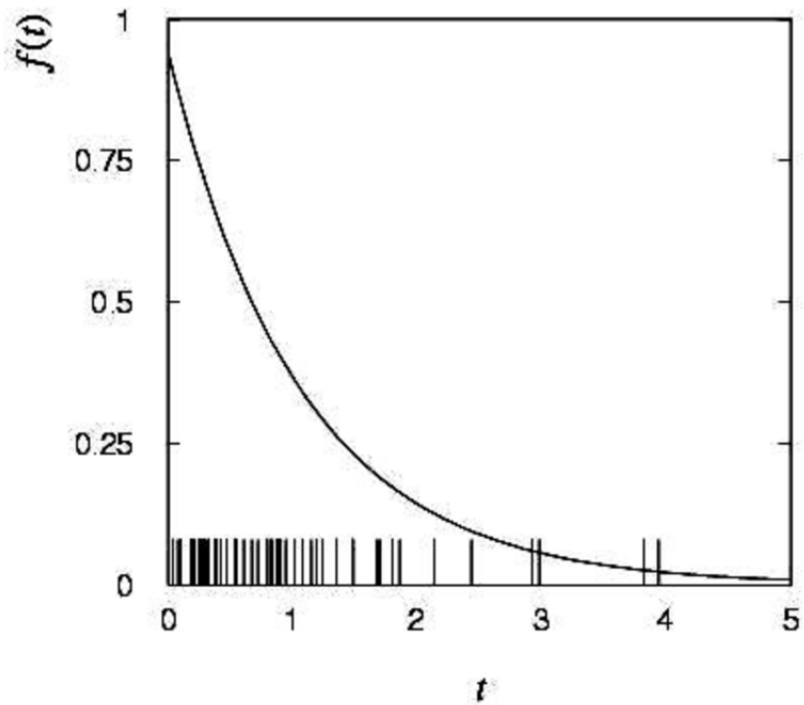
Liefert den ML-Schätzer: $$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

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

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:

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

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

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: $\widehat{\alpha} = \alpha(\widehat{\theta})$.

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

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

Man kann zeigen $E[\widehat{\lambda}] = \lambda \frac{n}{n-1}$. (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:

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

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

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 konsitenz 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.

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

Varianz für Schätzer: Analytische Methode

In 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 Exponetial-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

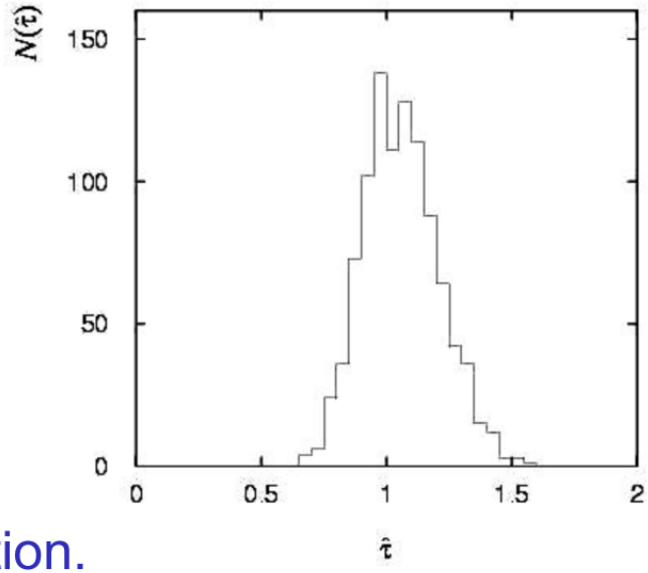
Nachdem wir den Schätzwert für den Parameter bestimmt haben müssen wir nun den statistischen Fehler des Schätzers bestimmen, i.e., wie weit die Verteilung der Schätzwerte wäre, wenn wir die gesamte identische Messung sehr oft wiederholen würden.

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:

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

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

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

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 Gleicheit):

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

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

→ 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: $\log \mathcal{L} = \log \mathcal{L}_{max} - \frac{k^2}{2}$

Beispiel: Varianzschätzung durch graph. Methode

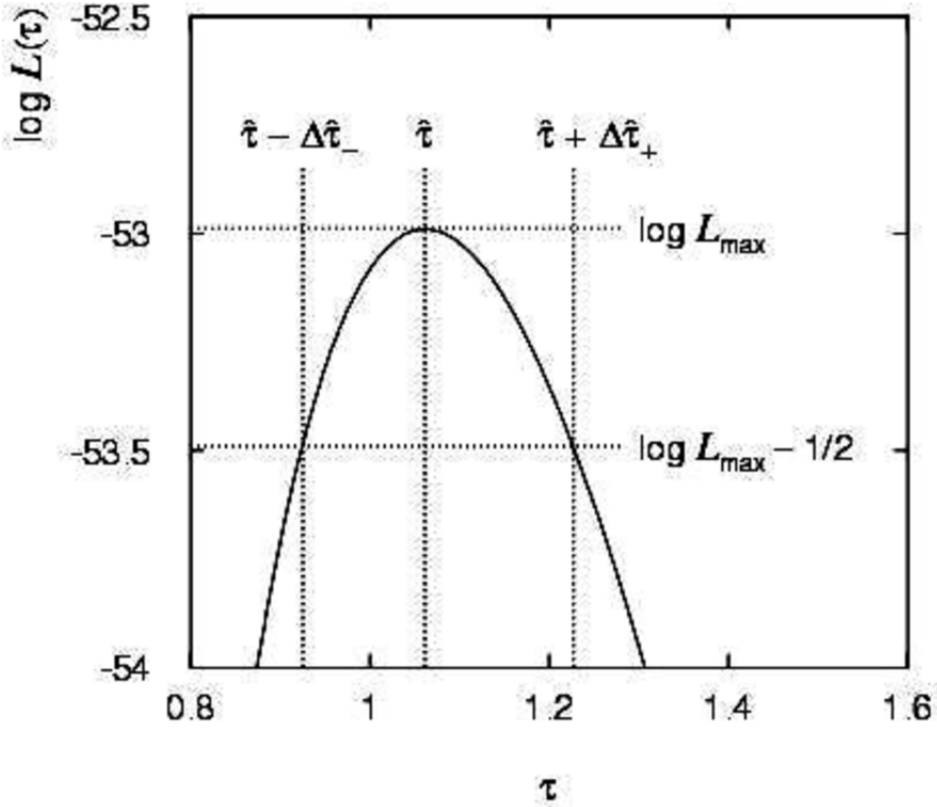
ML-Beispiel mit Exponential-WDF:

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

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

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

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



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

Beispiel: WDF-Exponentialfunktion

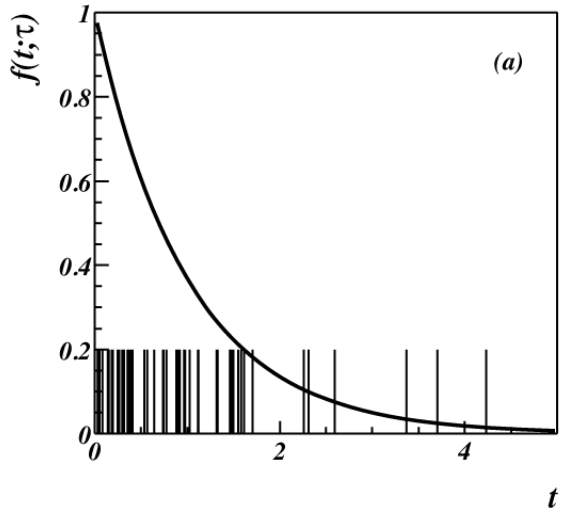
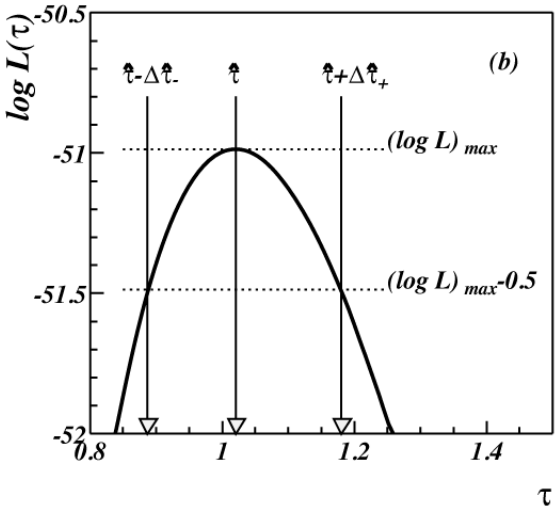
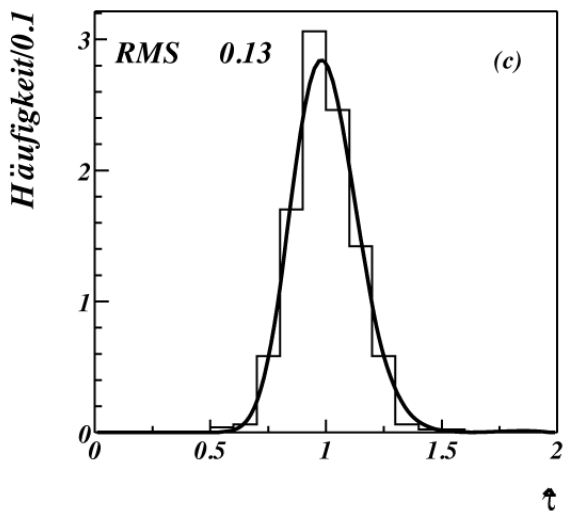
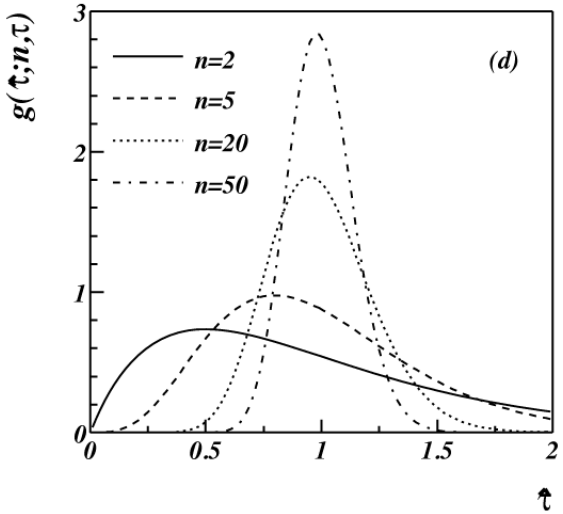
Graphisch:

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

Analytisch:

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

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

MC-Methode:

$$0.13,$$