Statistische Methoden der Datenanalyse

Wintersemester 2011/2012

Albert-Ludwigs-Universität Freiburg

Dr. Stan Lai und Prof. Markus Schumacher

Physikalisches Institut Westbau 2 OG Raum 008

Telefonnummer 07621 203 8408 (SL) / 7612 (MS)

E-Mail: Stan.Lai@physik.uni-freiburg.de

Markus.Schumacher@physik.uni-freiburg.de

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

Kapitel 5
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 $\widehat{\theta}(\vec{x})$

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

Gemeinsame 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)

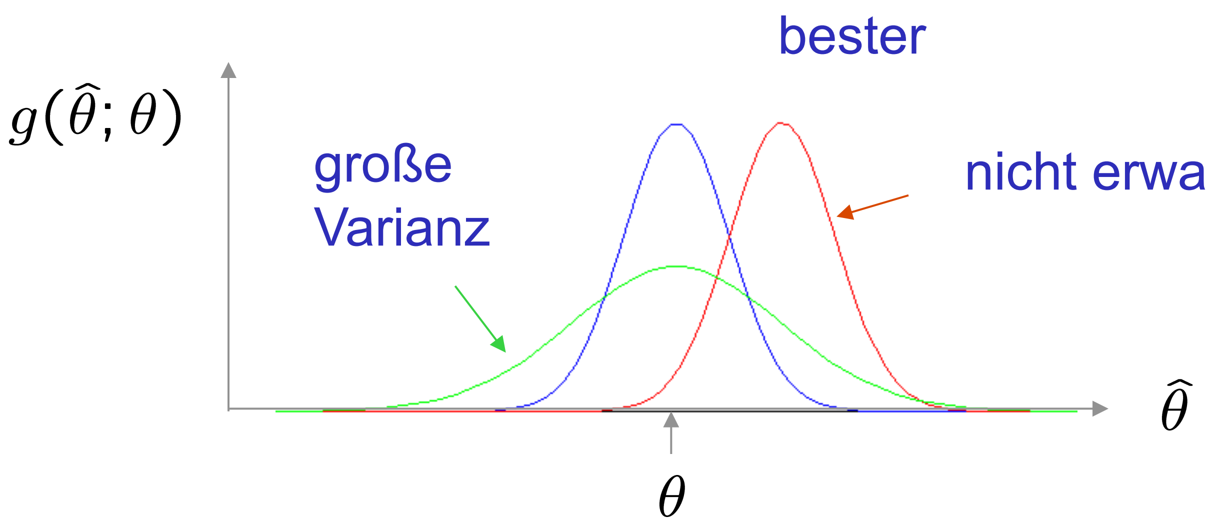
Mögliche Schätzer für den Erwartungswert der WDF der Grundgesamtheit

$\widehat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	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
$\widehat{\mu} = \text{Median der Stichprobe}$	

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

→ je größer die Information, desto kleiner der statistische Fehler.

→ Mittelwert der wiederholten Messungen sollte = wahren Wert sein

Kriterien für gute Schätzer

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

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

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

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 $\quad$ 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:

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

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

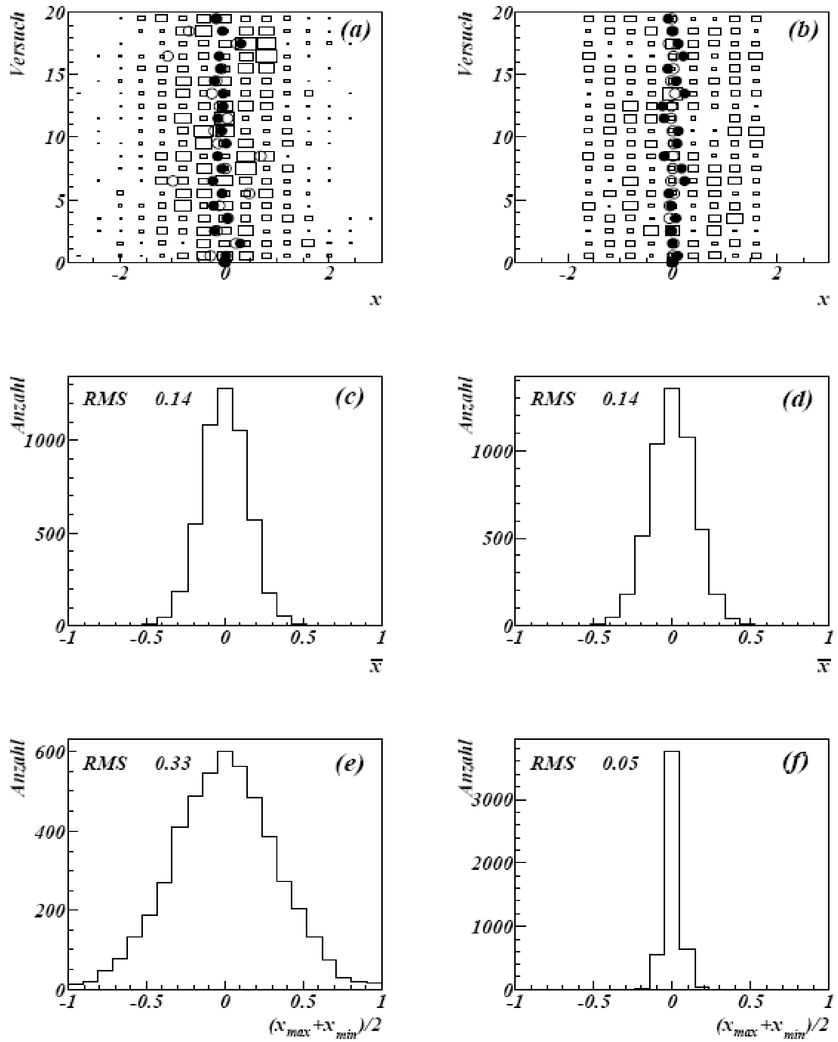
Die Schranke minimaler Varianz ergibt sich zu:

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

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

Vergleich von 2 Schätzern für den Erwartungswert

Gauss-WDF | Gleichverteilung



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

robust

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

effizienter für Gleich-verteilung

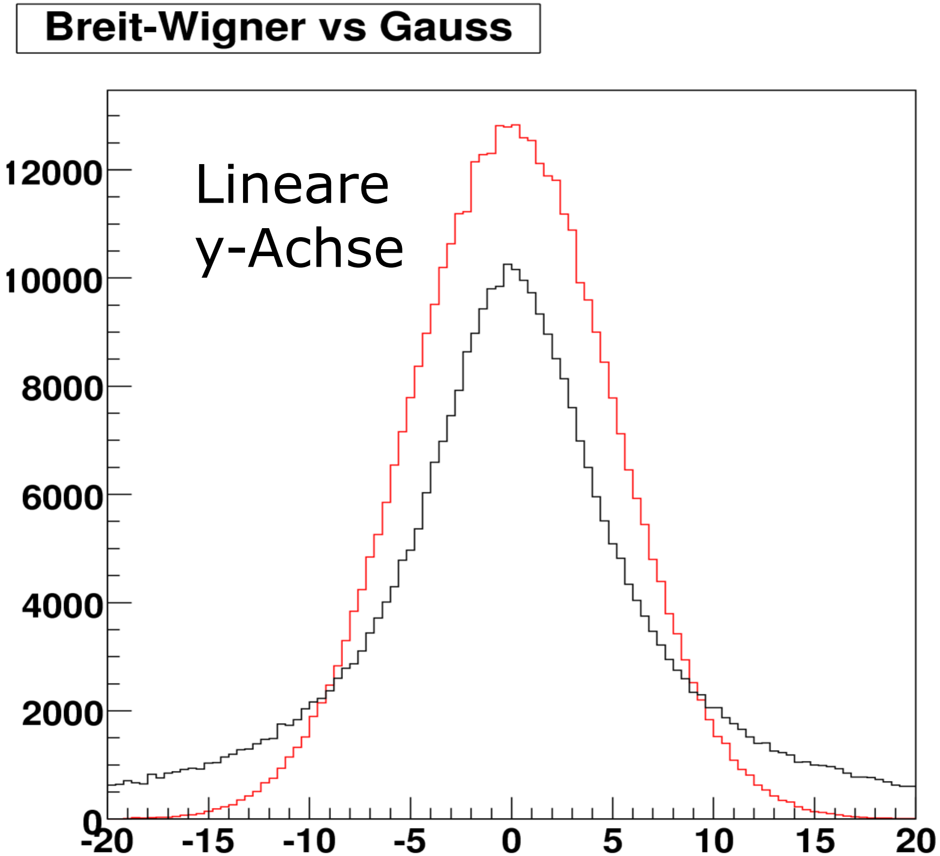
Robustheit: „truncated mean“

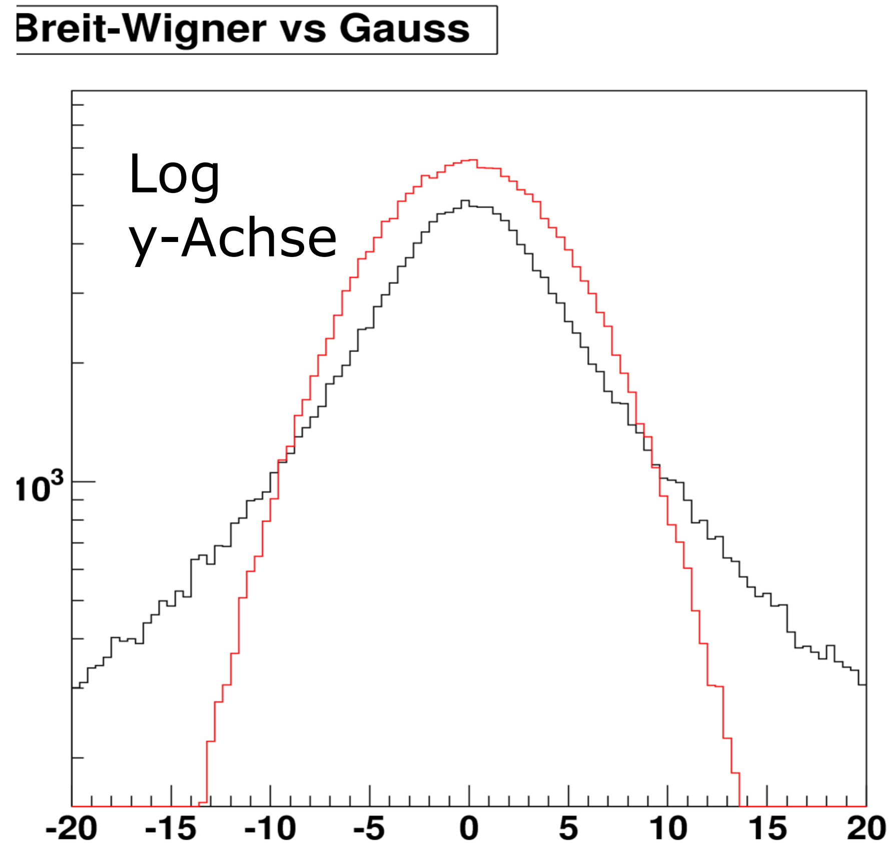
Gaussische WDf Breit-Wigner WDF

Ungefähr gleiche Varianz in Wertebereich

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

Arithmetischer Mittelwert?

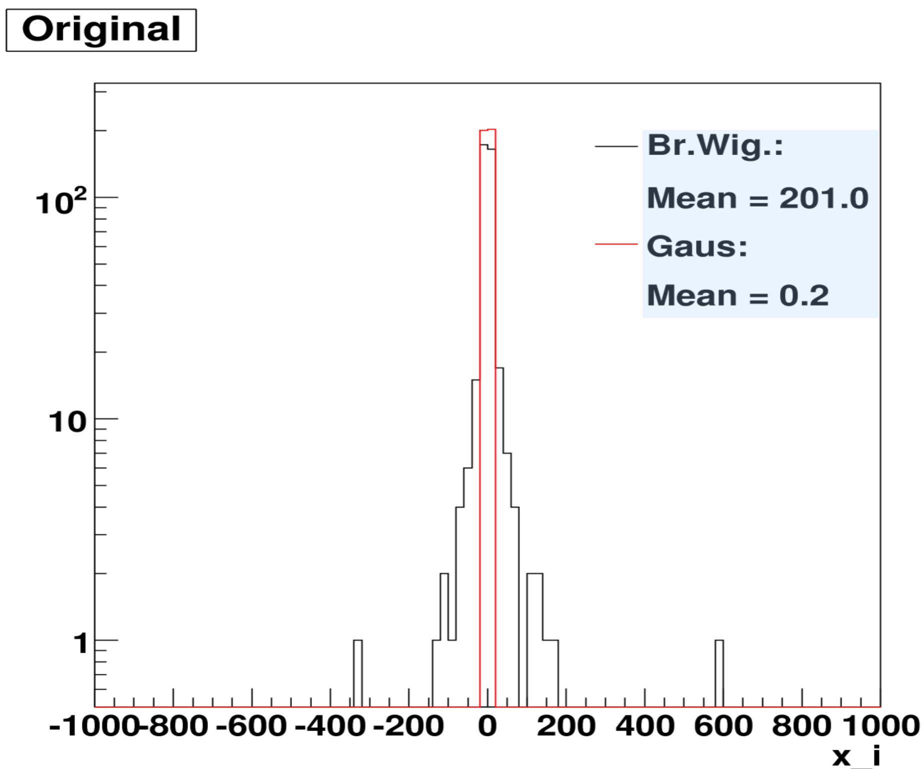




Robustheit: „truncated mean“

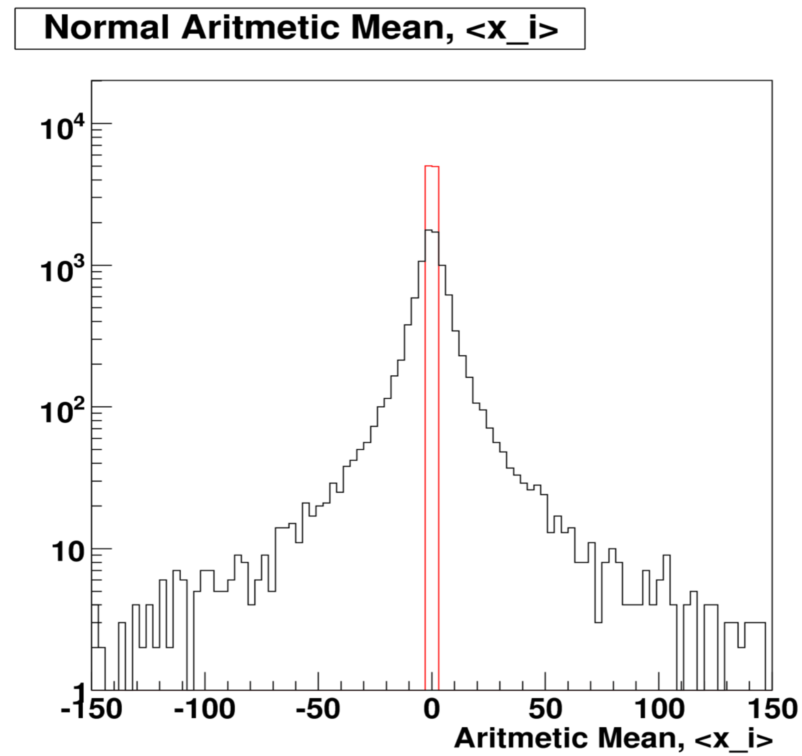
Stichprobe von 101 Zufallszahlen x_i von jeder PDF.

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, $\langle x_i \rangle$

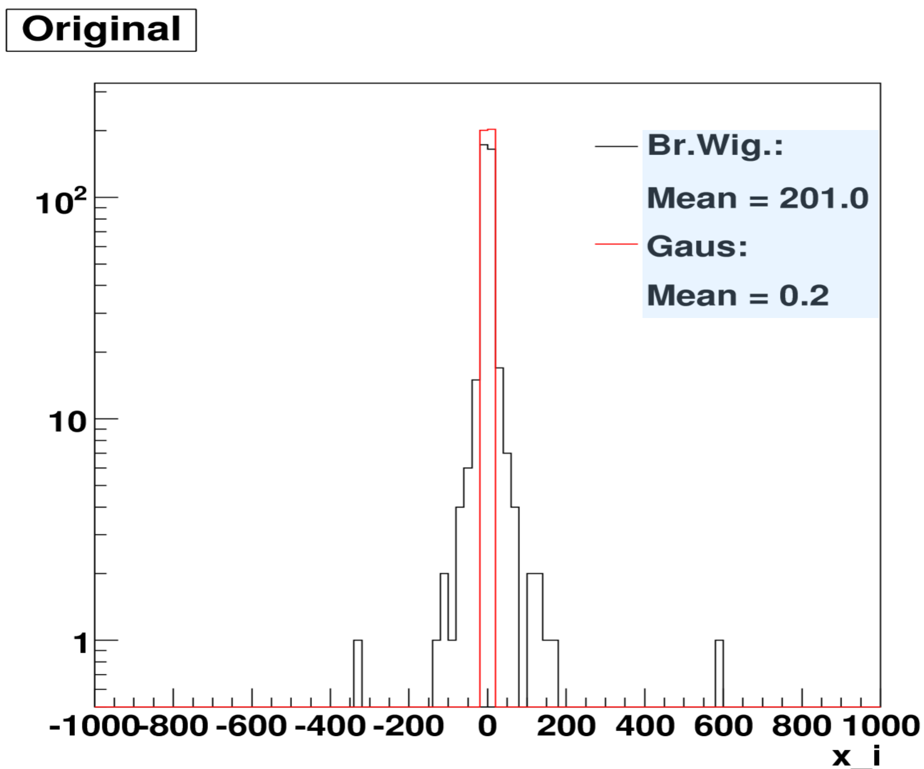
Breit Wigner: viel größere Varianz als für Gauß wegen Stichproben mit großen Einzelwerten x_i (siehe unten, links)



Robustheit: „truncated mean“

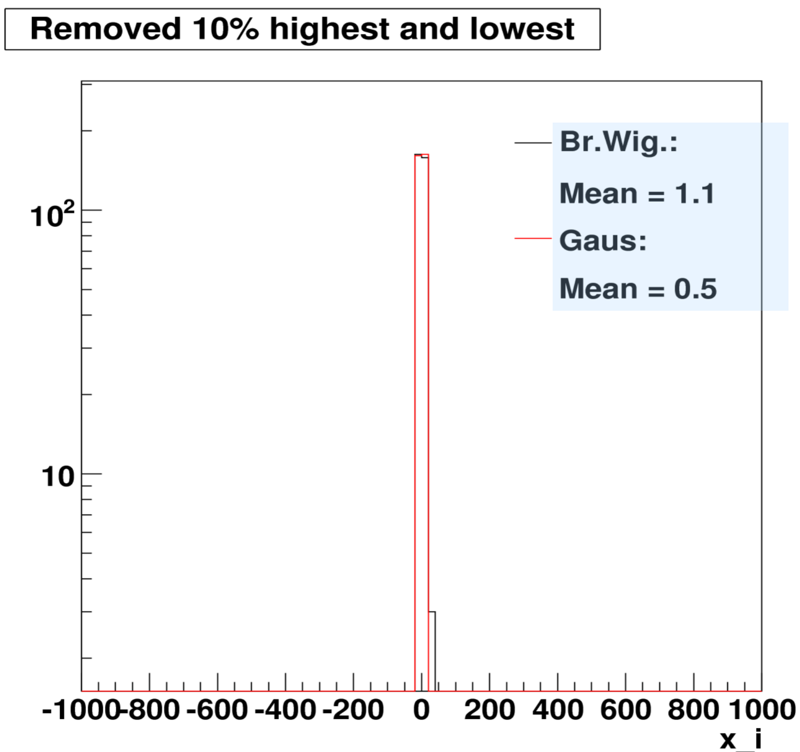
Stichprobe von 101 Zufallszahlen x_i von jeder PDF.

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”)

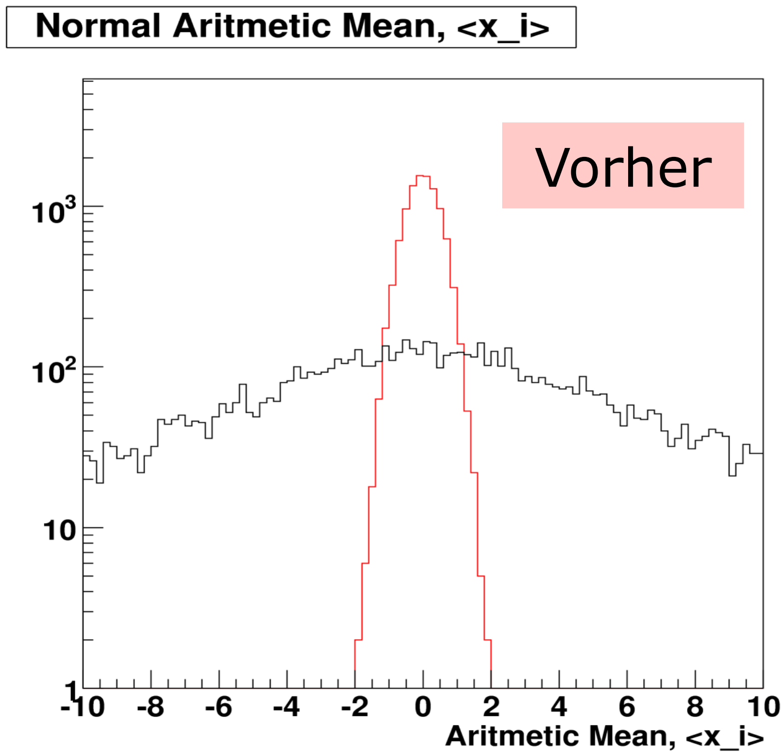
==> viel kleinere Varianz für $<x_i>$ für Breit Wigner



Robustheit: „truncated mean“

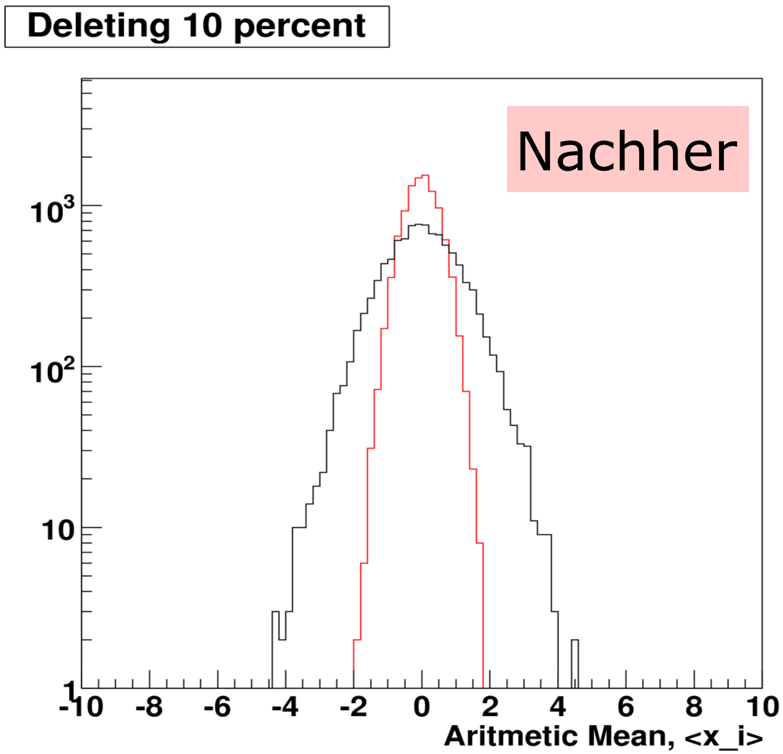
WDF für aritmetische Mittelwert, $<x_i>$

NORMAL

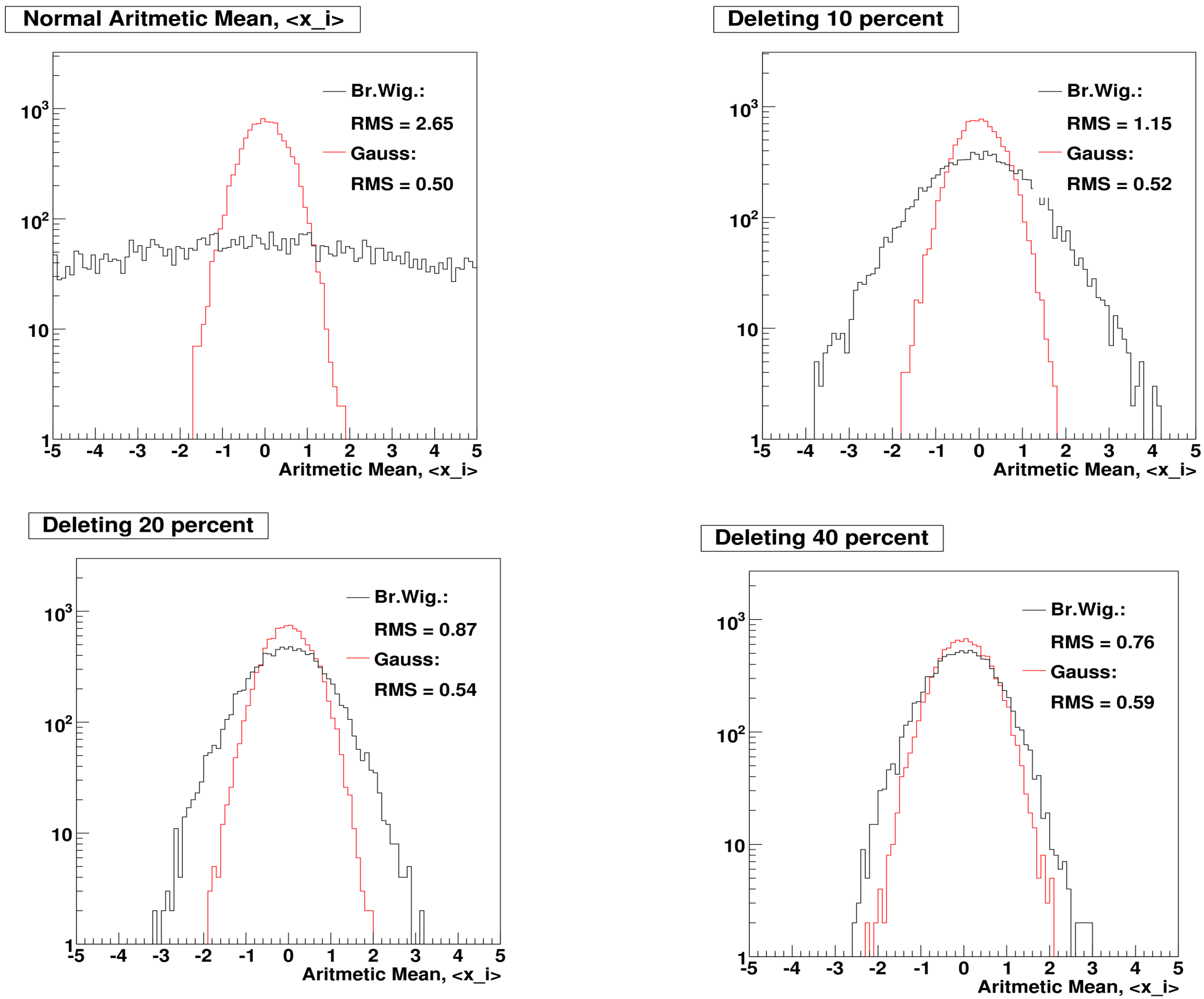


WDF für aritmetische Mittelwert, $<x_i>$

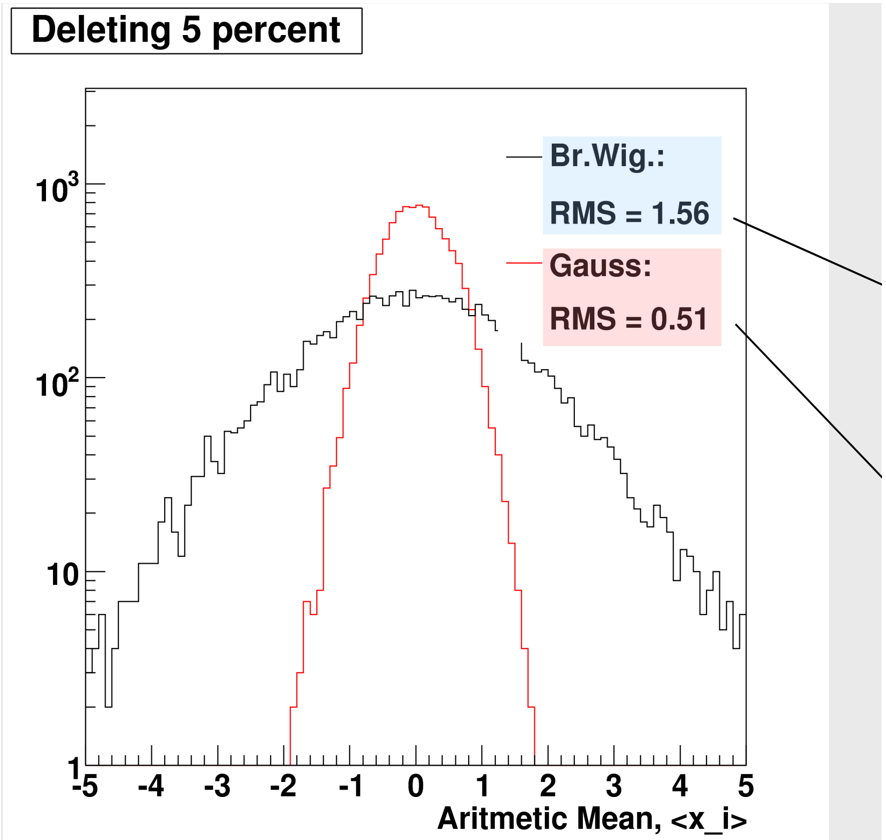
10% größte und kleinste Werte ignoriert

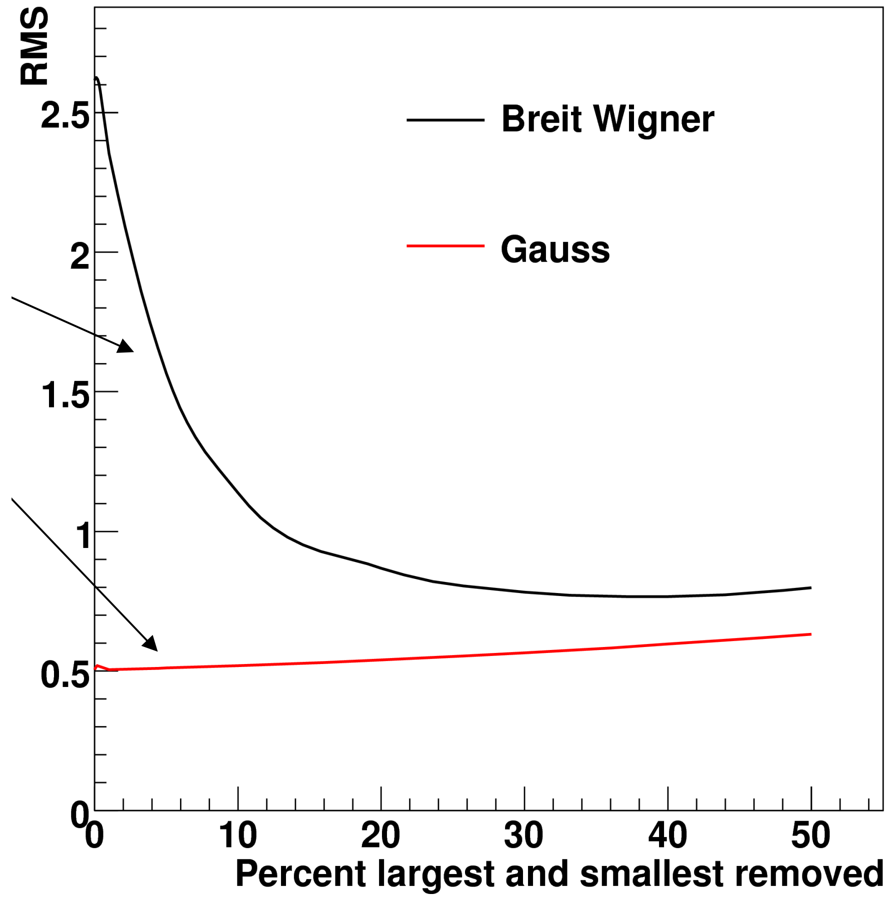


Robustheit: „truncated mean“



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.