

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 12: Grundlagen der Parameterschätzung

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

Grundgesamtheit, Stichprobe, Statistik, Schätzer

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

Grundgesamtheit = Menge möglicher Werte von x beschrieben durch $f(x)$

Stichprobe vom Umfang n = Satz von n unabhängigen Messungen der ZV x

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

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 \rightarrow Schätzwerte für Erwartungswert, Varianz, ...

b) Form der WDF bekannt \rightarrow Schätzwerte für Parameter $\hat{\theta}(\vec{x})$

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

Gemeinsame WDF der Stichprobe

Betrachte n Beobachtungen der ZV x : x_1, \dots, 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:

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

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 Experimentes (Satz von Messungen) $\vec{x} = (x_1, \dots, x_n)$ sei eine Menge von Zahlen \mathbf{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:

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

(\mathbf{x} konstant)

Im Falle von unabhängigen, identischen Messungen gilt:

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

Schätzer für den Erwartungswert μ von $f(x)$

Mögliche Schätzer für den Erwartungswert μ der WDF der Grundgesamtheit

$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ den Mittelwert der Stichprobe

$\hat{\mu} = \frac{1}{10} \sum_{i=1}^{10} x_i$ den Mittelwert der ersten 10 Punkte der Stichprobe

$\hat{\mu} = \frac{1}{n-1} \sum_{i=1}^n x_i$ $n/(n-1)$ mal den Mittelwert der Stichprobe

$\hat{\mu} = 42$

$\hat{\mu} = (\min(x_i) + \max(x_i))/2$ Mittelwert des größten und kleinsten Wertes

$\hat{\mu} = \text{Median der Stichprobe}$

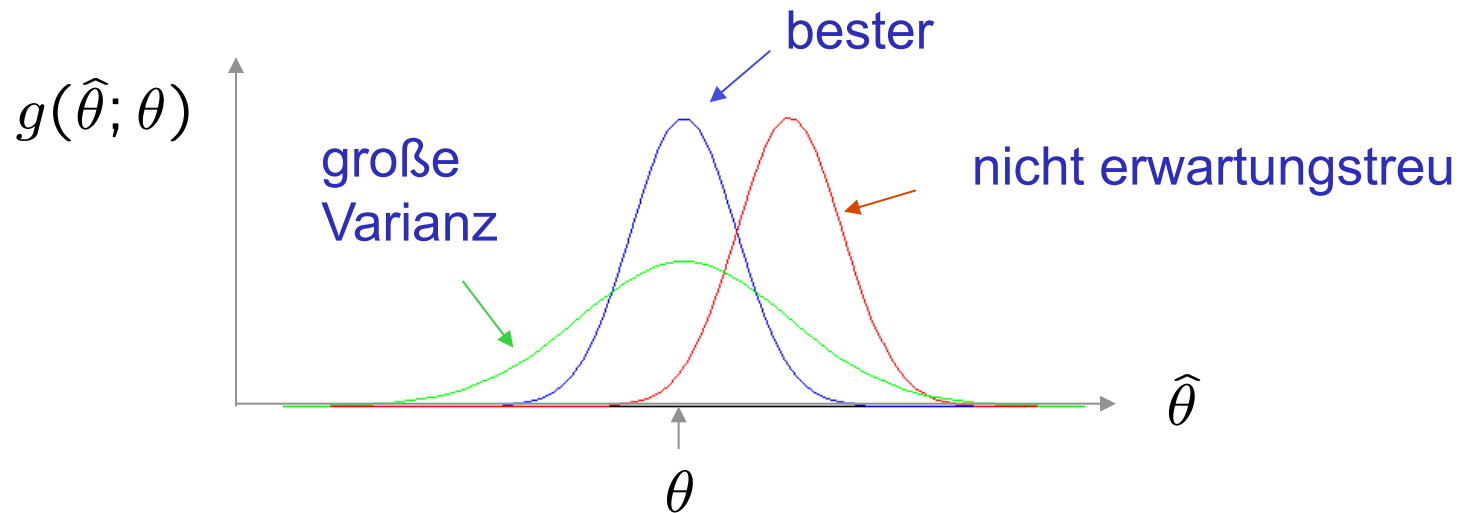
Fragen: - welcher Schätzer ist gut, der “Beste”?

- welche Kriterien sollte ein guter Schätzer erfüllen?

- wie findet man den optimalen Schätzer für ein Problem?

Eigenschaften von Schätzern

Wenn das Experiment (jeweils aus m Messungen) oft wiederholt wird.
Folgen die Schätzwerte einer WDF:



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

→ Mittelwert der vielen Schätzwerte sollte = wahren Wert sein

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

→ kleiner Bias und kleine 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

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

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.

Kriterien für gute Schätzer

Konsistenz $\hat{\theta}^{(n)} \xrightarrow{n \rightarrow \infty} \theta.$

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

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

möglichst keinen *Bias* d.h. Schätzer ist erwartungstreu.
Konsistente Schätzer mit endlicher Varianz sind
asymptotisch ($n \rightarrow$ unendlich) erwartungstreu

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

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: $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \equiv \bar{x}$ ('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

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

Arithmetischer 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:

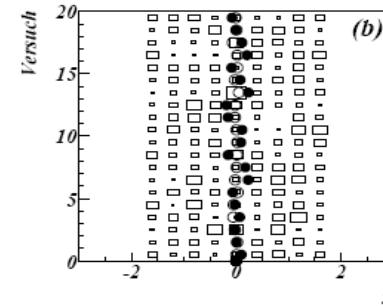
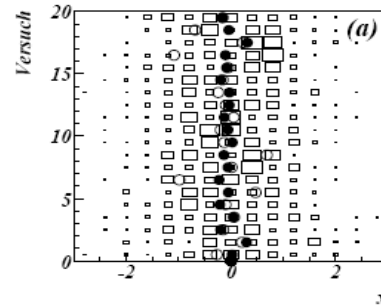
$$\text{SMV} = \frac{1}{E \left[-\frac{\partial^2 \log \mathcal{L}}{\partial \theta^2} \right]} = \frac{\sigma^2}{n} = \frac{V[x]}{n} = V[\bar{x}]$$

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

Vergleich von 2 Schätzern für den Erwartungswert

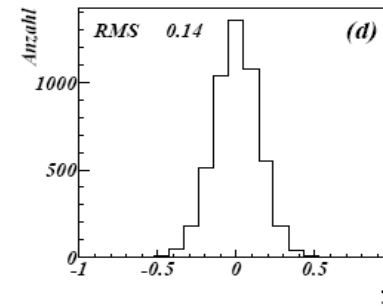
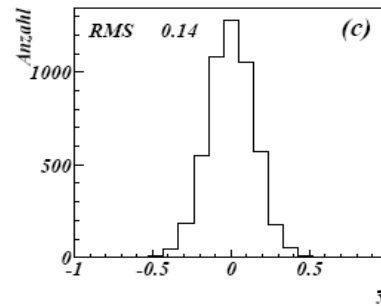
Gauss-WDF

Gleichverteilung



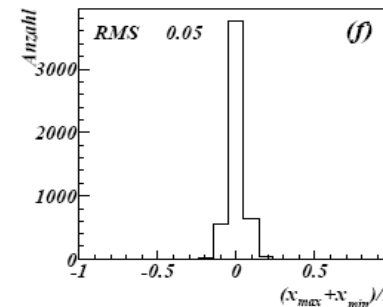
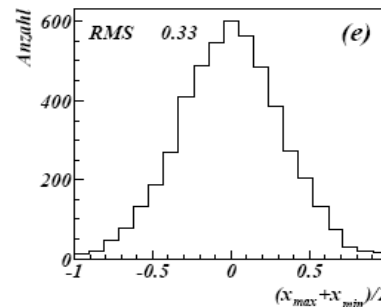
mit gleicher
Varianz = 1

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



robust

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



effizienter
für
Gleich-
verteilung

Ein Schätzer für die Varianz

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

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

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

$$b = E[\hat{\sigma}^2] - \sigma^2 = 0 \quad \text{erwartungstreu (ohne } n/n-1 \text{ Korrektur nur asymptotisch erwartungstreu)}$$

Der Fehler auf den Schätzer für die Varianz ist gegeben durch

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

Ein Schätzer für die Varianz

oder:

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

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 Kovarianzen ist:

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

Schätzer für die Korrelationskoeffizienten sind :

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

Für 2-dimensionale Gauss-WDF gilt:

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

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

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