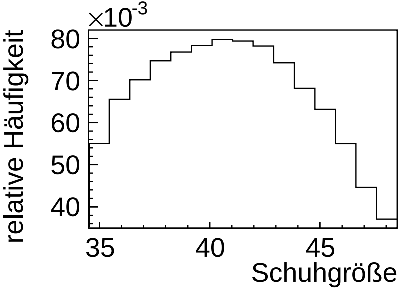
Gliederung

1. Parameterschätzung
   - Schätzwerte
   - Kriterien für Schätzwerte
2. Maximum Likelihood
   - Idee von Maximum Likelihood
   - Fehlerberechnung
3. Kleinste Quadrate
   - Prinzip der kleinsten Quadrate
   - Fehlerfortpflanzung und Hypothesentest

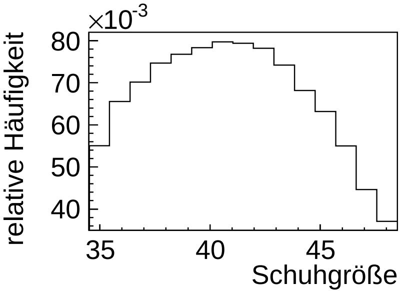
Schätzwerte

- Wahrer Wert nicht messbar, stattdessen folgen Messwerte einer Verteilung
- Schätzung eines Parameters a einer Verteilung anhand einer Stichprobe $\{x_i\}$
- Schätzfunktion $\hat{a}(\{x_i\})$



Schätzwerte

- Wahrer Wert nicht messbar, stattdessen folgen Messwerte einer Verteilung
- Schätzung eines Parameters a einer Verteilung anhand einer Stichprobe $\{x_i\}$
- Schätzfunktion $\hat{a}(\{x_i\})$



Beispiele:

- $\hat{a}_1(\{x_i\}) = \frac{1}{N}\sum_i x_i$
- $\hat{a}_2(\{x_i\}) = \sqrt[N]{\prod_i x_i}$
- $\hat{a}_3(\{x_i\}) = \mathrm{median}(\{x_i\})$
- $\hat{a}_4(\{x_i\}) = \max(\{x_i\})$

Konsistenz

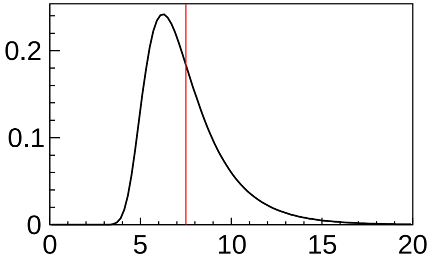
- Der Schätzwert $\hat{a}$ soll sich für große Stichproben dem wahren Wert a_0 annähern
- $\lim_{n\to\infty} \hat{a}(\{x_i\}) = a_0$

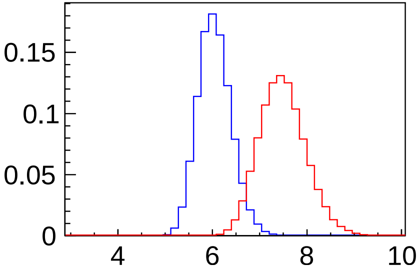
$\hat{a}_1(\{x_i\}) = \frac{1}{N}\sum_i x_i$	konsistent
$\hat{a}_2(\{x_i\}) = \sqrt[N]{\prod_i x_i}$	inkonsistent
$\hat{a}_3(\{x_i\}) = \mathrm{median}(\{x_i\})$	im allgemeinen inkonsistent
$\hat{a}_4(\{x_i\}) = \max(\{x_i\})$	inkonsistent

Erwartungstreue

- Schätzwert $\hat{a}$ ist selbst Zufallsvariable: Andere Stichprobe $\Rightarrow$ anderer Schätzwert
- Erwartungstreu wenn Erwartungswert des Schätzwerts dem wahren Wert entspricht
- $< \hat{a} >= a_0$
- Bias $b =< \hat{a} > -a_0$ kann korrigiert werden wenn bekannt
- Im allgemeinen gilt nicht $< \hat{a} >= a_0 \Rightarrow < f(\hat{a}) >= f(a_0)$

Erwartungstreue: Beispiel

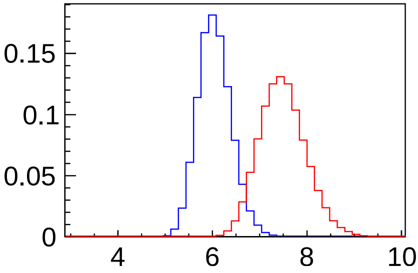




- Moyal-Verteilung mit Mittelwert $\mu = 7{,}5$
- Stichprobe aus 20 Messwerten, $\{x_i\}$ seien aufsteigend sortiert
- $\hat{a}_1\left(\{x_i\}\right) = \frac{1}{20}\sum_{i=1}^{20} x_i$
- $\hat{a}_2\left(\{x_i\}\right) = \frac{1}{11}\sum_{i=1}^{11} x_i$

Effizienz

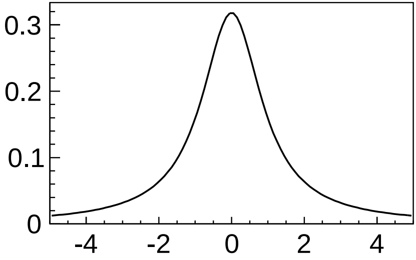
- Kleine Varianz von $\hat{a}$ wünschenswert
- Führt zu großer Wahrscheinlichkeit dass $\hat{a}$ nahe am Erwartungswert liegt
- Effizienz definiert als $\frac{V(\hat{a}_{min})}{V(\hat{a})}$
- $V(\hat{a}_{min})$ aus Maximum-Likelihood-Schätzung oder Rao-Cramér-Frechet-Grenze.



- $V(\hat{a}_2) < V(\hat{a}_1)$
- Verzerrung kann korrigiert werden

Robustheit

- Robustheit gegenüber „Ausreißern"oder falschen Voraussetzungen
- Schlecht quantifizierbar
- Mehr Robustheit geht meist auf Kosten der Effizienz



- Beispiel: Ausreißer bei Cauchy-Verteilung
- Mittelwert nicht robust
- Median robust; Effizienz $\approx 70\%$

Idee von Maximum Likelihood

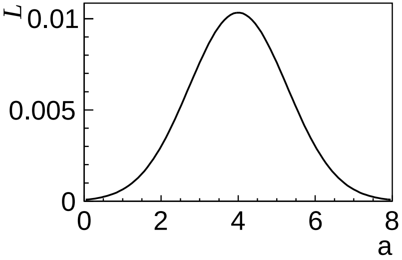
Definition

$$\mathcal{L}\left(\mathbf{x}|\mathbf{a}\right) := f_1\left(x_1|\mathbf{a}\right) \cdot f_2\left(x_2|\mathbf{a}\right) \cdot \ldots = \prod_i f_i\left(x_i|\mathbf{a}\right)$$

- $\mathcal{L}$ heißt Likelihood-Funktion
- $\mathcal{L}$ ist Wahrscheinlichkeitsdichte in $\mathbf{x}$, abhängig von den zu bestimmenden Parametern $\mathbf{a}$
- $\mathcal{L}$ ist nicht Wahrscheinlichkeitsdichte in $\mathbf{a}$: $\int \mathcal{L}\left(\mathbf{x}|\mathbf{a}\right) \mathrm{d}\mathbf{a} \neq 1$
- Suche $\mathbf{a}$ für den die gemessene Messwerte $\{x_i\}$ am wahrscheinlichsten sind
- $\mathcal{L}\left(\{x_i\}|\mathbf{a}\right) = \text{Maximum}$
- In der Praxis oft: Minimiere Log-Likelihood-Funktion $\mathcal{F} = -(2) \log \mathcal{L}$

Beispiel zu Maximum Likelihood

Beispiel



- Messung einer um wahren Wert gaußverteilten Größe
- $f(x|a) = \frac{1}{2\pi\sqrt{\sigma}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$
- Zwei Messungen, $x_1 = 3$, $x_2 = 5$
- $\mathcal{L} \propto \exp\left(-\frac{(x_1-a)^2+(x_2-a)^2}{2\sigma^2}\right)$

- $\mathcal{F} = C + \frac{(x_1-a)^2+(x_2-a)^2}{2\sigma^2}$
- $\frac{\partial\mathcal{F}}{\partial a} = -\frac{2(x_1-a)+2(x_2-a)}{2\sigma^2} = 0$ wenn $a = \frac{x1+x2}{2} = 4.$

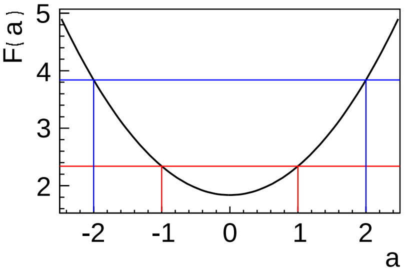
Beispiel: Lineare Regression

Beispiel

- Wahre Werte liegen auf Ursprungsgerade $g(x|a) = a \cdot x$
- Messe $\{x_i, y_i\}$-Paare
- Messwerte gaußverteilt mit Varianz σ^2 um wahren Wert
- Wahrscheinlichkeitsdichte ist $f_i(x_i, y_i|a) = \text{Gauss}(y_i - g(x_i|a), \sigma)$
- $\mathcal{L} = \prod_i f_i(x_i, y_i|a) = \prod_i \frac{1}{2\pi\sqrt{\sigma}} \exp\left(\frac{(y_i - x_i \cdot a)^2}{2\sigma^2}\right)$
- $\mathcal{F} = C + \sum_i \frac{(y_i - x_i \cdot a)^2}{2\sigma^2}, \quad \frac{\partial\mathcal{F}}{\partial a} = -\sum_i \frac{2x_i \cdot (y_i - x_i \cdot a)}{2\sigma^2}$
- $\frac{\partial\mathcal{F}}{\partial a} = 0$ wenn $a = \left(\sum_i (x_i \cdot y_i)\right) / \left(\sum_i (x_i)^2\right)$

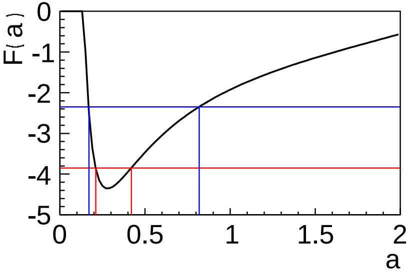
Parabolischer Fall

- Likelihood-Funktion geht gegen Gaußkurve für große Stichproben
- Entwickle $\mathcal{F}(a) = \mathcal{F}(\hat{a}) + \frac{1}{2} \cdot \frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}a^2} \cdot (a - \hat{a})^2 + \dots$
- Vergleich mit Gaußkurve $\exp\left(-\frac{(a-\hat{a})^2}{2\sigma^2}\right)$
- $\sigma = \left(\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}a^2}\right)^{-1/2}$
- $\mathcal{F}(\hat{a} \pm r\sigma) = \mathcal{F}(\hat{a}) + \frac{1}{2}\frac{1}{\sigma^2}(\hat{a} \pm r\sigma - \hat{a})^2 + \dots = \mathcal{F}(\hat{a}) + \frac{r^2}{2}$



Nichtparabolischer Fall

- Sei $z := z(a)$ so, dass $\mathcal{F}_z(z) := \mathcal{F}\left(a^{-1}(z)\right)$ parabolisch ist
- Bestimme σ_z wie vorher
- $\mathcal{F}_z(\hat{z} + r\sigma_z) = \mathcal{F}_z(\hat{z}) + \frac{r^2}{2} = \mathcal{F}(\hat{a}) + \frac{r^2}{2} = \mathcal{F}(\hat{a} + r\sigma_r)$
  $\mathcal{F}_z(\hat{z} - r\sigma_z) = \mathcal{F}_z(\hat{z}) + \frac{r^2}{2} = \mathcal{F}(\hat{a}) + \frac{r^2}{2} = \mathcal{F}(\hat{a} - r\sigma_l)$

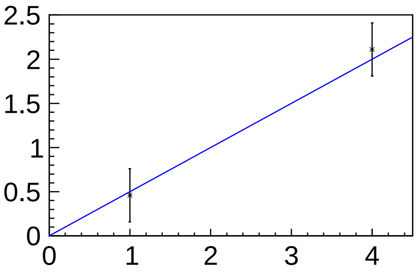


- Lese $\mathcal{F}$ dort ab wo es um $\frac{r^2}{2}$ zugenommen hat
- Im allgemeinen asymmetrische Fehler

Fehlerberechnung bei Ausgleichsgerade

Beispiel

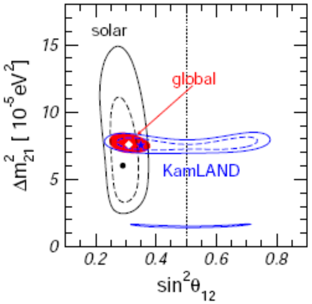
- $$\frac{\partial \mathcal{F}}{\partial a} = -\sum_i \frac{x_i \cdot (y_i - x_i \cdot a)}{\sigma^2}$$
- $$\frac{\partial^2 \mathcal{F}}{\partial a^2} = \sum_i \frac{x_i^2}{\sigma^2}$$
- $$\sigma_a = \left(\frac{\partial^2 \mathcal{F}}{\partial a^2}\right)^{-1/2} = \frac{\sigma}{\sqrt{\sum_i x_i^2}}$$



- Fehler nur abhängig von x-Werten
- Beschreibt nur Fortpflanzung der y-Fehler in *a*
- Kein Goodness-of-fit

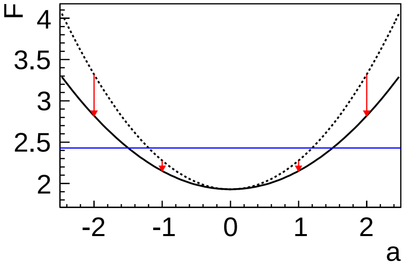
Mehrere Parameter

- Gleiche Ideen funktionieren weiter
- $\frac{\partial \mathcal{F}}{\partial a_i} = 0$ für alle a_i.
- $G_{ij} = \frac{\partial^2 \mathcal{F}}{\partial a_i \partial a_j}$ liefert das Inverse der Kovarianzmatrix $V = G^{-1}$
- $\mathcal{F}(\mathbf{a}) = \mathcal{F}(\hat{\mathbf{a}}) + \frac{r^2}{2}$ definieren Konturen gleicher Likelihood
- Schließen $r\sigma$-Gebiete ein



Profile Likelihood

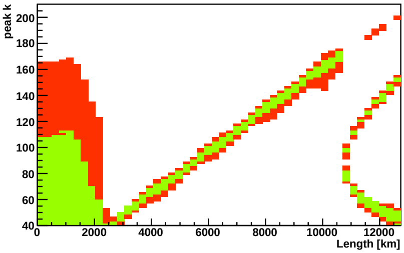
- Oft: Modell enthält Störparameter ν
- Benötigt zur Modellierung, aber am Ende uninteressant
- Wie wirkt sich Fehler bei der Bestimmung von ν auf interessanten Parameter *a* aus?
- Minimiere $\mathcal{F}$ für jedes *a* als Funktion von ν.



- Kurve wird flacher als bei konstantem ν
- Führt zu größerem σ_a

Neyman-Konstruktion

- Bisher vorgestellte Ansätze problematisch bei kleinen Stichproben
- Bestimme Konfidenzregion die a_0 bei N Versuchen $C \cdot N$ mal enthält (Frequentist)
- C heißt Coverage, $\alpha = 1 - C$ ist Irrtumswahrscheinlichkeit
- a ist in der Konfidenzregion enthalten wenn $\int \mathcal{L}(\mathbf{x}|a)\,\mathrm{d}\mathbf{x} = C$ die gemessenen $\{x_i\}$ enthält



- Freiheit: Wahl des Integrationsgebiets, zum Beispiel:
- $\int_{-\infty}^{x_0} \mathcal{L}(x|a)\,\mathrm{d}x = C$
- $\int_{x_0}^{\infty} \mathcal{L}(x|a)\,\mathrm{d}x = C$
- $\int_{-\infty}^{x_u} \mathcal{L}(x|a)\,\mathrm{d}x = \alpha/2 \wedge \int_{x_o}^{+\infty} \mathcal{L}(x|a)\,\mathrm{d}x = \alpha/2$

Bayes'sche Statistik

- Bayes-Theorem gilt auch für Wahrscheinlichkeitsdichten
- $\mathcal{P}(\mathbf{a}|\mathbf{x}) = \frac{\mathcal{L}(\mathbf{x}|\mathbf{a}) \cdot f(\mathbf{a})}{g(\mathbf{x})}$
- Intepretiere $\mathcal{P}(\mathbf{a}|\mathbf{x})$ als Glaube, dass **a** einen bestimmten Wert hat
- $g(\mathbf{x})$ konstant in **a**, legt Normierung fest
- $f(\mathbf{a})$ heißt Prior: Glaube vor der Messung dass a einen bestimmten Wert hat

- **a** liege mit Wahrscheinlichkeit $1 - \alpha$ in der Kredibilitätsregion
- $\int \mathcal{P}(\mathbf{a}|\{x_i\})\,\mathrm{d}\mathbf{a} = 1 - \alpha$
- Erneut Freiheit bei der Wahl des Integrationsbereichs

Prinzip der kleinsten Quadrate

- Gaußverteilte Fehler führen für $\mathcal{F}$ zu Ausdrücken der Form

Definition

$$S := \frac{(y_1 - a_1)^2}{\sigma_1^2} + \frac{(y_2 - a_2)^2}{\sigma_2^2} + \ldots = \sum_i \frac{(y_i - a_i)^2}{\sigma_i^2} = \sum_i \frac{\Delta y_i^2}{\sigma_i^2}$$

- Δy_i heißen Residuen
- Prinzip der kleinsten Quadrate: $S =$ Minimum
- Äquivalent zu ML für gaußverteilte Fehler, sonst „eigenes“ Verfahren
- Allgemein mit Kovarianzmatrix $\mathbf{V}$ und $\Delta\mathbf{y} = \begin{pmatrix} \Delta y_1 \\ \Delta y_2 \\ \vdots \end{pmatrix}$:

$$S = \Delta\mathbf{y}^{\mathrm{T}}\mathbf{V}^{-1}\Delta\mathbf{y} = \text{Minimum}$$

Lineare kleinste Quadrate

- Lineares Modell $f(x|\mathbf{a}) = a_1 f_1(x) + a_2 f_2(x) + \ldots$
- Vereinfachung: $\{y_i\}$ seien unkorreliert und gleiche Varianz σ^2
- $S = \frac{1}{\sigma^2} \sum_i (y_i - f(x_i|\mathbf{a}))^2$
- $\frac{\partial S}{\partial a_j} = -\frac{2}{\sigma^2} \sum_i f_j(x_i) \cdot (y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \ldots) \stackrel{!}{=} 0$
- $\Rightarrow$ Normalengleichungen:

$$a_1 \sum_i f_j(x_i) f_1(x_i) + a_2 \sum_i f_j(x_i) f_2(x_i) + \ldots = \sum_i y_i f_j(x_i)$$

Matrixschreibweise

Normalengleichungen:

$$a_1 \sum_i f_j(x_i) f_1(x_i) + a_2 \sum_i f_j(x_i) f_2(x_i) + \ldots = \sum_i y_i f_j(x_i)$$

Definiere Matrix für n Messungen und p Parameter

$$\mathbf{A} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \ldots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \ldots & f_p(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \ldots & f_p(x_n) \end{pmatrix} \qquad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Schreibe Normalengleichungen in Matrixform:

$$\left(\mathbf{A}^{\mathrm{T}} \mathbf{A}\right) \mathbf{a} = \mathbf{A}^{\mathrm{T}} \mathbf{y}$$

Matrixschreibweise

Normalengleichungen:

$$a_1 \sum_i f_j(x_i) f_1(x_i) + a_2 \sum_i f_j(x_i) f_2(x_i) + \ldots = \sum_i y_i f_j(x_i)$$

Definiere Matrix für n Messungen und p Parameter

$$\mathbf{A} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \ldots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \ldots & f_p(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \ldots & f_p(x_n) \end{pmatrix} \qquad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Schreibe Normalengleichungen in Matrixform:

$$\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)\mathbf{a} = \mathbf{A}^{\mathrm{T}}\mathbf{y} \Rightarrow \mathbf{a} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{y}$$

Allgemeine Lösung und Fehlerfortpflanzung

- Definiere Gewichtsmatrix $\mathbf{W} := \mathbf{V}^{-1}$
- $S = \Delta\mathbf{y}^{\mathrm{T}}\mathbf{V}^{-1}\Delta\mathbf{y} = (\mathbf{y} - \mathbf{Aa})^{\mathrm{T}}\mathbf{W}(\mathbf{y} - \mathbf{Aa})$
- S wird minimal bei $\mathbf{a} = \underbrace{\left(\mathbf{A}^{\mathrm{T}}\mathbf{WA}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}}_{\mathbf{B}}\,\mathbf{y}$

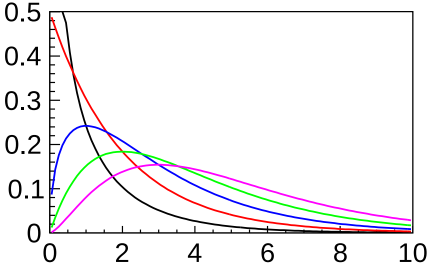
Fehlerfortpflanzung:

- Gegeben sei lineare Transformation $\mathbf{y} = \mathbf{Bx}$
- Dann gilt Fehlerfortpflanzungsgesetz: $\mathbf{V}_{\mathrm{y}} = \mathbf{BV}_{\mathrm{x}}\mathbf{B}^{\mathrm{T}}$
- Hier:

$$\begin{aligned}
\mathbf{V}_{\mathrm{a}} = \mathbf{BVB}^{\mathrm{T}} &= \left(\mathbf{A}^{\mathrm{T}}\mathbf{WA}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{WV}\left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{WA}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\right)^{\mathrm{T}} \\
&= \left(\mathbf{A}^{\mathrm{T}}\mathbf{WA}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{A}\right)^{-1} \\
&= \left(\mathbf{A}^{\mathrm{T}}\mathbf{WA}\right)^{-1}
\end{aligned}$$

χ^2-Test

- χ^2_n-Verteilung gibt Wahrscheinlichkeitsdichte für Summe aus n standard-normalverteilter Zufallsvariablen
- S folgt χ^2-Verteilung mit $n-p$ Freiheitsgraden (für gaußverteilte Fehler)
- Wahrscheinlichkeit diesen oder größeren Wert für S zu erhalten:
$$\int_S^\infty f(\chi^2_{n-p})\mathrm{d}\chi^2$$



- Erwartungswert ist $E[\chi^2_n] = n$
- Für $n-p > 2$ ist auch Wahrscheinlichkeit für kleines S gering

Zusammenfassung

	Maximum Likelihood	**Kleinste Quadrate**
Input	PDFs	Varianzen
Anwendbarkeit	PDFs bekannt	Daten unverzerrt, Varianzen endlich
Konsistent	Ja	Ja
Erwartungstreu	nur für große Stichproben	Ja im lin. Fall
Effizient	asymptotisch maximal	maximal (im lin. Fall)
Robust	Nein	Nein
Goodness-of-fit	Nein	Für gaußsche Fehler
Rechenaufwand	im allgemeinen hoch	Im lin. Fall relativ gering

Literatur

Allgemein:

- V. Blobel, E. Lohrmann, Statistiche und numerische Methoden der Datenanalyse, Teubner (1998)
- C. Amsler et al. (Particle Data Group), Physics Letters B667, 1 (2008)

Plots:

- T. Schwetz, M. Tortola, J. W.F. Valle, New J.Phys.10:113011, 2008
- K. Scholberg, A. Burgmeier, R. Wendell, (2009) 0910.3174