Karlsruher Institut für Technologie

Einführung in die
Computergestützte Datenauswertung

Ulrich Husemann, Karlsruher Institut für Technologie, Institut für Experimentelle Teilchenphysik

Sommersemester 2024 – Kapitel 4

Kapitel 4.5

Ausblick: Moderne Methoden der Datenanalyse

Maximum-Likelihood-Methode

Wiederholung

- Methode der **kleinsten Quadrate** (LS-Methode): systematische Methode zur Schätzung von Parametern und deren Unsicherheit
- Minimierung der **Summe der Residuenquadrate**:

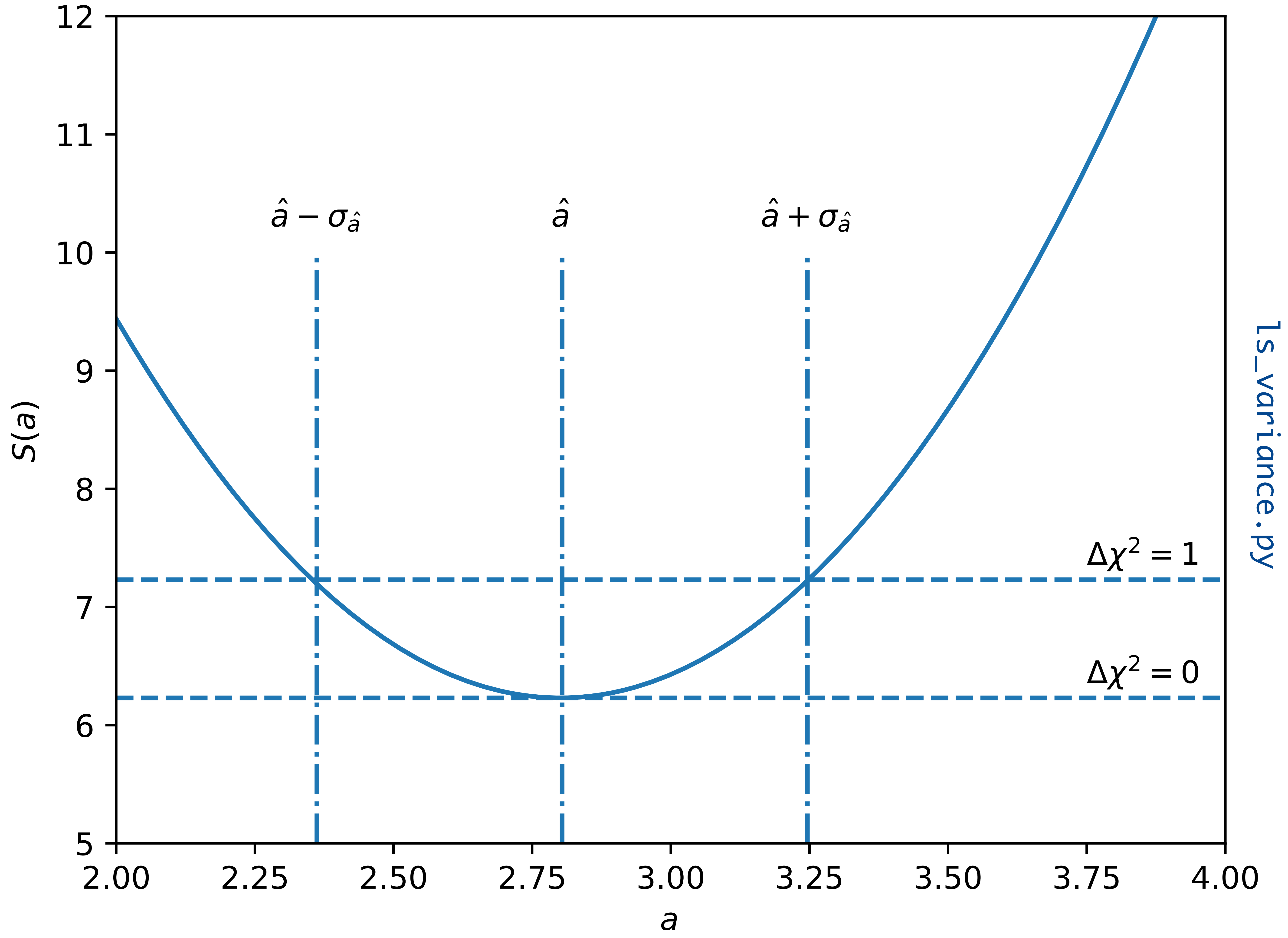
$$S \equiv \chi^2 \equiv \sum_{i=1}^{n} \frac{(y_i - \lambda_i(x_i;\mathbf{a}))^2}{\sigma_i^2} \stackrel{!}{=} \min \qquad S \equiv (\mathbf{y} - \boldsymbol{\lambda}(\mathbf{x};\mathbf{a}))^{\mathrm{T}}\,\mathbf{V}^{-1}\,(\mathbf{y} - \boldsymbol{\lambda}(\mathbf{x};\mathbf{a})) \stackrel{!}{=} \min$$

(meist numerisch, für lineare Funktionen der Parameter auch analytisch)

- **Güte der Anpassung** über χ^2-Wahrscheinlichkeit
$\rightarrow$ Faustformel: $\chi^2/n_{\mathrm{dof}} \approx 1$

Wiederholung

- Methode der kleinsten Quadrate: **Unsicherheit** der Parameterschätzung
  - **Hesse-Matrix**: Varianz des Parameters proportional zu (2. Ableitung der Summe der Residuenquadrate S am Minimum)$^{-1}$
  - **Grafischer Ansatz**: Unsicherheit der Parameterschätzung = Intervall um Minimalwert mit $\Delta\chi^2 = 1$



Eigenschaften von Schätzwerten

Karlsruher Institut für Technologie

- Günstige **Eigenschaften** von Schätzwerten (nur in Worten):
  - **Erwartungstreue**: keine (oder geringe) **Verzerrung** (engl.: bias)
    = Abweichung des erwarteten Schätzwerts von wahrem Wert
    (Beispiel: Besselkorrektur → erwartungstreue Stichprobenvarianz)
  - **Konsistenz**: im Grenzfall großer Stichproben konvergiert Schätzwert gegen wahren Wert (Beispiel: Stichprobenvarianz, auch ohne Besselkorrektur)
  - **Effizienz**: kleinstmögliche **Varianz** des Schätzwerts
    → Parameter kann mit möglichst geringer Unsicherheit geschätzt werden
  - Große **Genauigkeit**: Kombination aus Varianz und Verzerrung (mittlere quadratische Abweichung, engl.: mean square error, MSE) klein
  - **Robustheit**: geringe Sensitivität auf „Ausreißer" in Daten oder fehlerhaftes Modell
  - **Einfachheit** der Definition des Schätzwerts

Maximum-Likelihood-Methode

- Maximum-Likelihood-(ML-)Methode (Fisher, ca. 1912):
  - Parameterschätzung durch **Maximierung einer Funktion** („Likelihood-Funktion”), so dass das statistische Modell am **besten zu den Daten passt**
  - **Universell einsetzbar** für fortgeschrittene statistische Methoden (Beispiel später: Hypothesentests)
- Man kann u. a. zeigen:
  - ML-Schätzwerte sind **effizient**, aber im allgemeinen nicht erwartungstreu
  - Für **normalverteilte** Zufallsvariablen sind Methode der kleinsten Quadrate und Maximum-Likelihood-Methode **äquivalent**

LS und ML: erster Vergleich

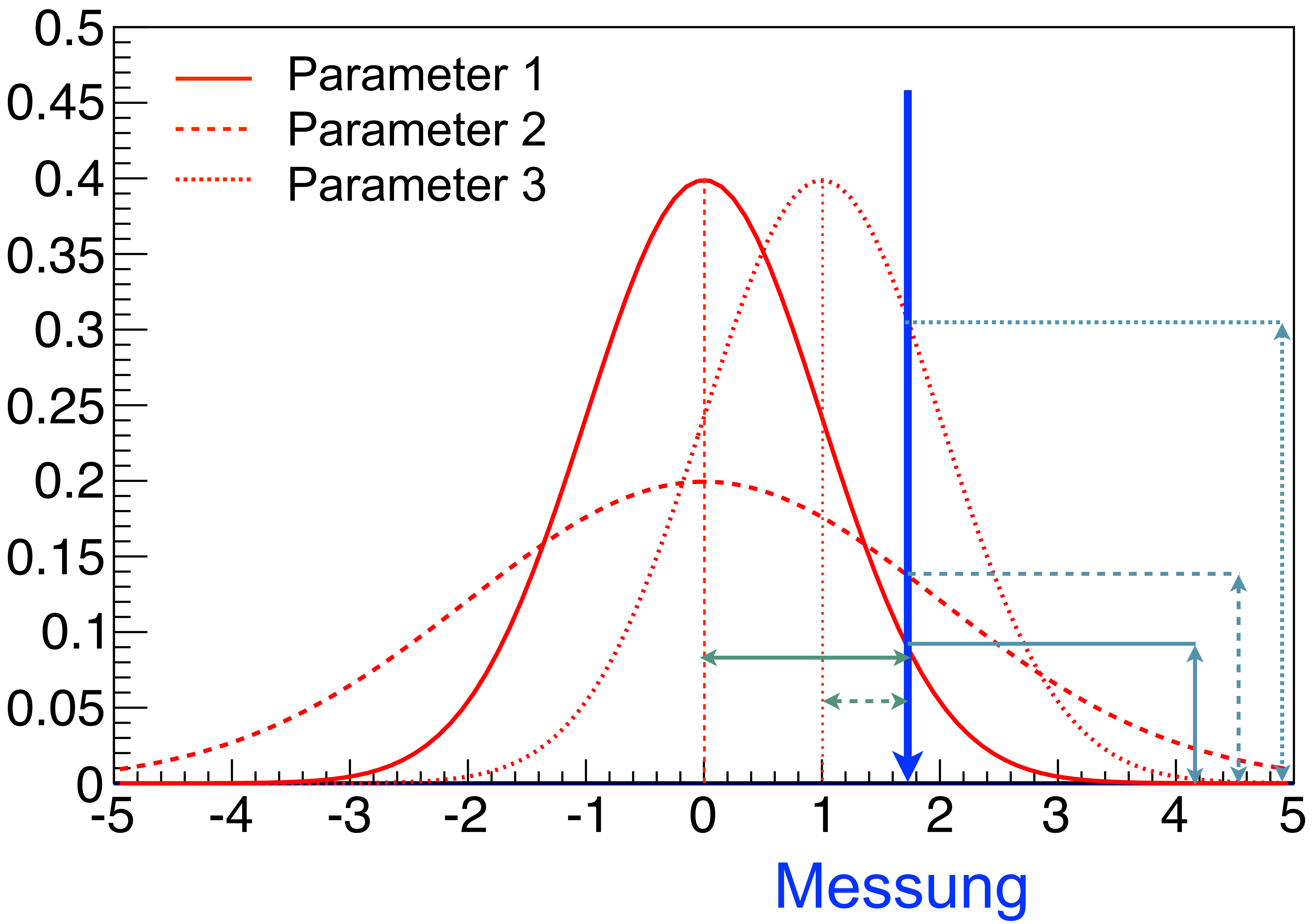
- Frage: welche Modellparameter passen am besten zu den Daten?

Kleinste Quadrate:
minimiere (normierten) Abstand vom Erwartungswert

(nur Erwartungswert/Varianz müssen bekannt sein)

Maximum Likelihood:
maximiere Höhe der PDF

(gesamte PDF muss bekannt sein)



Likelihood-Funktion

- Messwerte x_i: Stichprobe der Größe n aus **bekannter** Wahrscheinlichkeitsdichte $f(x;\mathbf{a})$ mit **Normierung** bezüglich x:

$$\int f(x;\mathbf{a})\,\mathrm{d}x = 1$$

- Konstruktion der Likelihood-Funktion: **Produkt** der Wahrscheinlichkeitsdichtefunktionen (PDFs) für alle Messwerte

$$L(\mathbf{a}) \equiv L(\mathbf{a}|\mathbf{x}) \equiv \prod_{i=1}^{n} f(x_i;\mathbf{a})$$

  - Große Werte, wenn **a** „passt” (und PDF korrekt)
  - Reinterpretation als **Funktion im Parameterraum** $\{ a_1, \ldots, a_n \}$ (d. h. **nach** der Messung, x_i fest), dort ist L aber **keine PDF** (nicht normiert)

Maximum-Likelihood-Prinzip

- Der **Maximum-Likelihood-Schätzwert $\hat{\mathbf{a}}$** für einen Parametervektor **a** ist derjenige Wert von **a**, für den die Likelihood-Funktion den **größten Wert** annimmt: $L(\hat{\mathbf{a}}) \geq L(\mathbf{a})$ für alle **a**

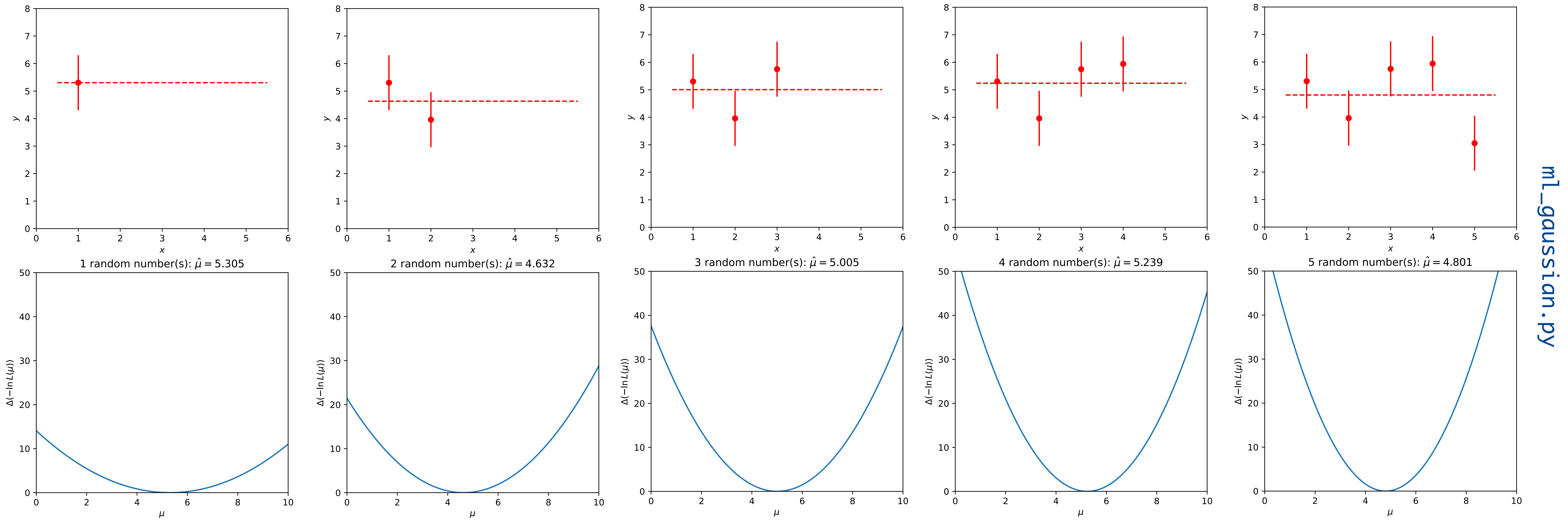
- Praxis: –ln $L(\mathbf{a})$ wird **minimiert** („negative Log-Likelihood”)
  - Logarithmus streng monotone Funktion → **dieselben Extrema**
  - Aus Produkten werden **Summen**: $-\ln L(\mathbf{a}) = -\sum_{i=1}^{n} f(x_i; \mathbf{a})$
    → konstante Summanden irrelevant für Minimum
  - Schätzwert $\hat{\mathbf{a}}$ aus Bedingungen:

$$\frac{\partial(-\ln L(\mathbf{a}))}{\partial a_j} \stackrel{!}{=} 0 \qquad \qquad \left.\frac{\partial^2(-\ln L(\mathbf{a}))}{\partial a_i\, \partial a_j}\right|_{\mathbf{a}=\hat{\mathbf{a}}} \quad \text{negativ definit}$$

  - Minimierung: meist **numerisch**, in wenigen Fällen auch analytisch

Beispiel: Normalverteilung

- n = 1–5 normalverteilte Zufallszahlen mit $\mu = 5$ und $\sigma^2 = 1$:



$$\Delta(-\ln L(\mu)) \equiv -(\ln L(\mu) - \ln L(\hat{\mu}))$$

$\rightarrow$ Parabeln um $\hat{\mu}$, die mit wachsendem n stärker gekrümmt werden

Analytische Lösung: Normalverteilung

Karlsruher Institut für Technologie

- Gegeben: n normalverteilte Messwerte $x_i \rightarrow$ Likelihood-Funktion:

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

$$\rightarrow \quad -\ln L(\mu, \sigma) = \frac{1}{2} \sum_{i=1}^{n} \left[\ln(2\pi) + \ln(\sigma^2) + \frac{(x_i - \mu)^2}{\sigma^2}\right]$$

- Bemerkungen:
  - Gleichung für **Parabel** in μ
  - **Äquivalenz** zur Methode der kleinsten Fehlerquadrate für Normalverteilung:

$$-\ln L = \frac{1}{2} S + \text{const}$$

(konstante Terme irrelevant für Minimum von $-\ln L$ $\rightarrow$ **derselbe** Schätzwert)

Varianz von ML-Schätzwerten: grafisch

- Taylorentwicklung von –ln L um Minimum (analog zu LS-Methode):

$$-\ln L(a) \approx -\ln L(\hat{a}) + \left.\frac{\partial(-\ln L(a))}{\partial a}\right|_{a=\hat{a}} (a-\hat{a}) + \frac{1}{2}\left.\frac{\partial^2(-\ln L(a))}{\partial a^2}\right|_{a=\hat{a}} (a-\hat{a})^2$$

am Minimum: 0

Cramér-Rao-Ungleichung: $\hat{V}[\hat{a}]^{-1}$

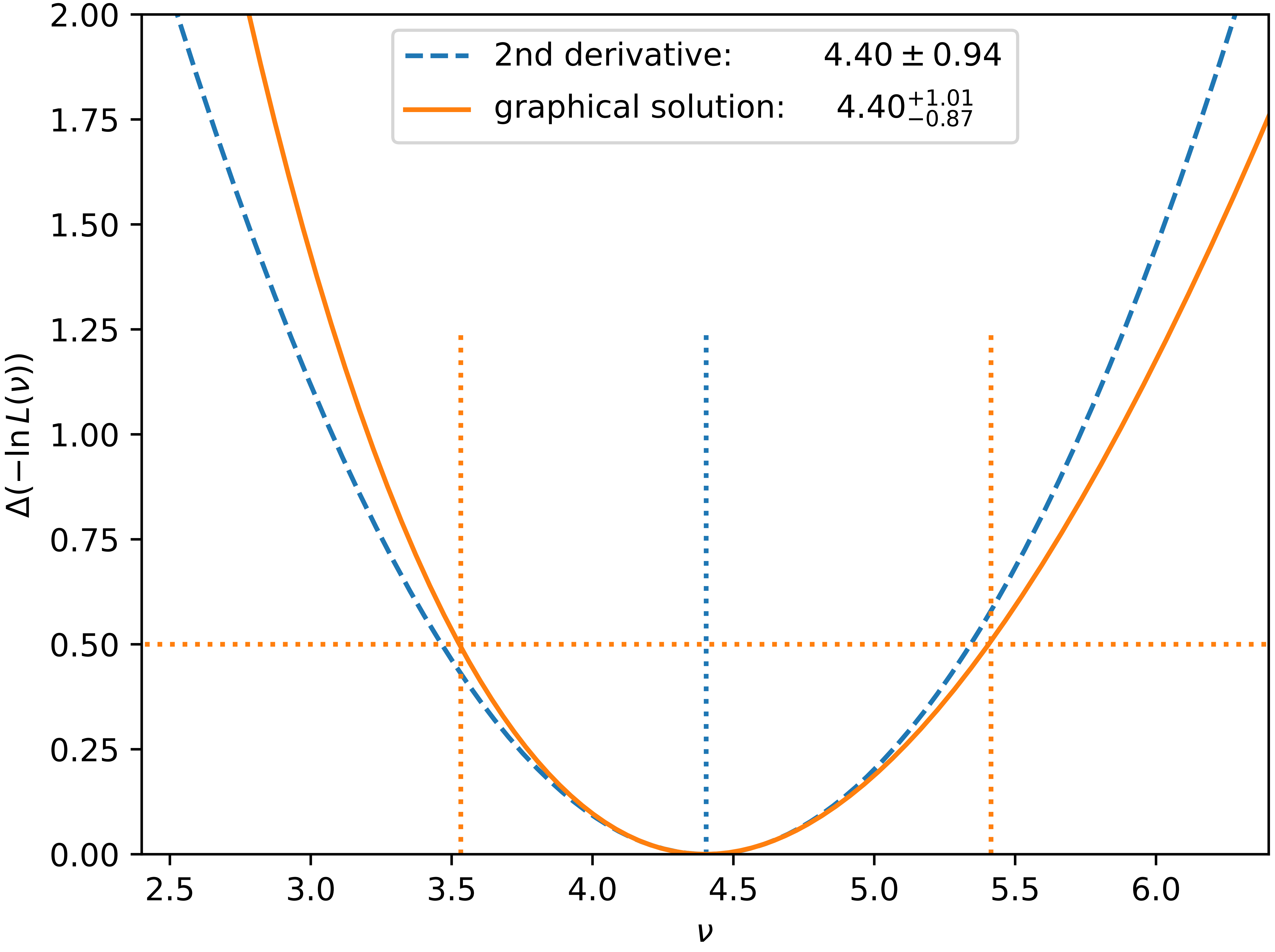
- Damit:

$$\Delta(-\ln L(a)) \equiv -(\ln L(a) - \ln L(\hat{a})) \approx \frac{1}{2}\left.\frac{\partial^2(-\ln L(a))}{\partial a^2}\right|_{\hat{a}} (a-\hat{a})^2 = \frac{1}{2}\frac{(a-\hat{a})^2}{\hat{V}[\hat{a}]}$$

$$\rightarrow \text{für } \Delta(-\ln L(a)) = \frac{1}{2}: \; a = \hat{a} \pm \sqrt{\hat{V}[\hat{a}]}$$

Unsicherheit der Parameterschätzung: Scan der negativen Log-Likelihood für Funktionswert –ln $L(\hat{a})$ + 1/2

Varianz von ML-Schätzwerten: grafisch



Zusammenfassung

- **Maximum-Likelihood-Methode**: universelles und mächtiges Werkzeug der mathematischen Statistik
  - **Likelihood-Funktion**: Produkt der (als bekannt vorausgesetzten) Wahrscheinlichkeitsdichten aller Messungen
  - Maximum-Likelihood-Prinzip: Schätzwerte für Parameter durch **Maximierung der Likelihood-Funktion**
  - Praxis: (numerische) **Minimierung** der negativen Log-Likelihood-Funktion
  - Für normalverteilte Zufallsgrößen: dieselben Schätzwerte wie Methode der kleinsten Quadrate