# Datenanalyse in der Physik

# Vorlesung 6

# Maximum-Likelihood-Methode

Prof. Dr. J. Mnich

DESY und Universität Hamburg



## Parameterschätzung

Problem:

- Aus Messungen, die mit Unsicherheiten (Fehlern) behaftet sind

$\Longrightarrow$

- bestmögliche Bestimmung von physikalischen Parametern, inklusive deren Unsicherheiten

Dieses Problem wird allgemein als Parameterschätzung bezeichnet

Es läuft darauf hinaus, die Parameter eines physikalischen Modells möglichst gut an die Messdaten anzupassen

Anpassungsrechnung oder Fit

Wir werden hier die beiden gebräuchlichsten Verfahren besprechen:

- Maximum-Likelihood-Methode
- Methode der kleinsten Quadrate ($\chi^2$-Methode)

# Parameterschätzung

Allgemeine Bemerkungen:

- Messdaten:
  Es sollen $n$ Messwerte $x_1, \ldots, x_n$ aufgenommen worden sein, die wir als Stichproben von Zufallsvariablen behandeln
  Die $x_i$ können dabei einzelne Zufallsvariablen sein, aber auch Zufallsvektoren

  Im Prinzip muß die Form der pdf bekannt sein
  oder man muß Annahmen darüber machen
  $\Longrightarrow$ erfordert ein physikalisches Modell
  siehe die Diskussion zur Bayes' Statistik

- Fehler:
  Die Fehler der einzelnen Messungen sowie deren Korrelationen (Kovarianzmatrix) müssen ermittelt werden

  Man unterscheidet
  - statistische Fehler
    Wiederholung der Messung reduziert Unsicherheit
  - systematische Fehler
    Wiederholung reduziert Unsicherheit nicht
    typisches Beispiel sind Kalibrationsunsicherheiten



# Schätzung des Mittelwertes

- Zunächst einmal betrachten wir die einfachste Methode um den Mittelwert einer pdf aus einer Stichprobe abzuschätzen

$$\langle x \rangle = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Man kann zeigen, dass das bereits die optimale Methode für eine zu Grunde liegende Gauß-Verteilung ist
(folgt aus Maximum-Likelihood-Methode)

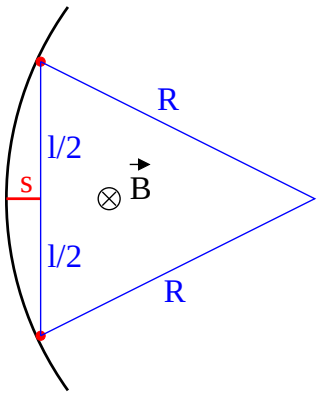
- Falls die pdf nicht so schnell abfällt wie die Gauß-Verteilung, d.h. lange Schwänze hat, oder gar asymmetrisch ist, ist Vorsicht geboten

Bei kleinen Stichproben kann dies zu falschen Ergebnissen führen (Methode konvergiert langsam)

Nach Möglichkeit sollte man versuchen, Parametertransformationen auszuführen, so dass man gaußverteilte Zufallsvariablen erhält

# Beispiel: Impulsmessung durch Messung der Sagitta

Oft misst man den Impuls geladener Teilchen im Magnetfeld durch die Abweichung von einer Geraden auf einer Länge $l \rightarrow$ Sagitta $s$




Es gilt $p = q\,B\,R$ und für $s \ll l$: $s \approx \frac{l^2}{8\,R}$

$$\Longrightarrow p = q\,B\,\frac{l^2}{8\,s} = K\,\frac{1}{s}$$

Meist ist die Genauigkeit der Impulsmessung allein durch die Messung der Sagitta $s$ bestimmt, von der man annehmen kann, dass sie gaußverteilt ist
Es gilt:

$$\langle p \rangle \approx p(\langle s \rangle) = \frac{K}{\langle s \rangle}$$

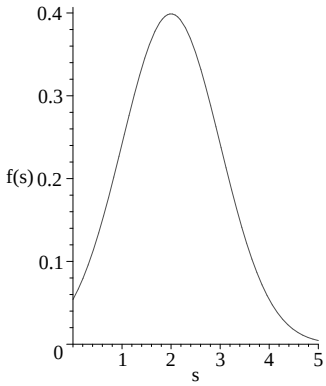
$$\sigma_p^2 \approx \left(\frac{\mathrm{d}p}{\mathrm{d}s}\right)^2 \sigma_s^2 = \frac{p^4}{K^2}\,\sigma_s^2$$

$$\Longrightarrow g(p) = \frac{f(s)}{\left|\frac{\mathrm{d}p}{\mathrm{d}s}\right|} = \frac{1}{\sigma_p^2}\,e^{-\frac{p^4\left(\frac{1}{p}-\frac{1}{\langle p\rangle}\right)^2}{2\,\sigma_p^2}}$$



# Beispiel: Impulsmessung durch Messung der Sagitta

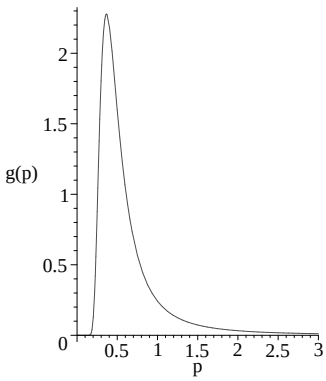
Die gaußförmige pdf der Sagitta-Messung mit $\langle s \rangle = 2$ und $\sigma_s^2 = 1$ sieht so aus:
(willkürliche Einheiten)




Offenbar sind hier auch Messungen in der Nähe von $s = 0$ nicht unwahrscheinlich

Negative $s$-Werte ergeben auch anderes Vorzeichen für den Impuls

Das führt zu folgender, stark asymmetrischer Verteilung für den Impuls:




Messungen von $s$ nahe bei Null ergeben hohe Impulswerte

Es empfiehlt sich die gaußverteilte Größe $1/p$ zu betrachten

# Maximum-Likelihood-Methode

- Es liege eine Stichprobe aus $n$ Messungen einer Zufallsvariablen $x$ (oder eines Zufallsvektors $\vec{x}$) vor
- Die Wahrscheinlichkeitsdichteverteilung (pdf) sei bekannt und hänge von einem Parameter $a$ (oder mehreren Parametern $\vec{a}$) ab:

$$f(x;a)$$

- Das Prinzip der Maximum-Likelihood-Methode ist die Likelihood-Funktion

$$L(a) = f(x_1;a) \cdot f(x_2;a) \ \ldots \ f(x_n;a) = \prod_{i=1}^{n} f(x_i;a)$$

zu maximieren um so den besten Schätzwert für den (oder die) Parameter $a$ zu erhalten

- Man muss darauf achten, dass die pdf für alle möglichen Werte von $a$ korrekt normiert ist

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



# Maximum-Likelihood-Methode

Das Maximum wird durch Differenzieren bestimmt

- 1 Parameter $a$:

$$\frac{\mathrm{d}L(a)}{\mathrm{d}a} = 0$$

- Mehrere Parameter $a_1, \ldots, a_m$ (bzw. $\vec{a}$):

$$\frac{\partial L(\vec{a}\,)}{\partial a_k} = 0 \qquad \text{für alle} \ \ k = 1, \ldots, m$$

- Aus praktischen Gründen minimiert man den negativen Logarithmus der Likelihood-Funktion

$$\ell(a) \equiv -\ln L(a) = -\ln \prod_{i=1}^{n} f(x_i;a) = -\sum_{i=1}^{n} \ln f(x_i;a)$$

  - Der Logarithmus vermeidet, dass man für große $n$ sehr kleine Zahlen für $L(a)$ erhält und in numerische Probleme läuft
  - Das Vorzeichen wird so gewählt um die gleichen numerischen Minimierungsalgorithmen zu verwenden wie bei der Methode der kleinsten Quadrate ($\chi^2$-Methode)

# Maximum-Likelihood-Methode

Bemerkungen zur Maximum-Likelihood-Methode

- Die Methode erfordert die Kenntnis der zu Grunde liegenden Wahrscheinlichkeitsdichteverteilung $f(x;a)$

  Annahme einer falschen pdf führt zu falschen Ergebnissen

- Wichtige Kontrolle:
  Man überprüfe, ob die pdf mit den ermittelten Parametern die Messdaten richtig wieder gibt

- Probleme treten auf, falls das Minimum am Rande des Parameterraumes von $a$ gefunden wird

Universität Hamburg



# Beispiel zur Maximum-Likelihood-Methode

Experiment zur Radioaktivität: Unter jeweils gleichen Bedingungen werden $n$ Messungen der Zahl der Zerfälle $r_i$ gemacht

Frage: Wie groß ist die mittlere Zahl der Zerfälle $\mu$?
(z.B. zur Berechnung der Zerfallskonstanten $\lambda$)

I. Einfacher Mittelwert

$$\mu = \frac{1}{n}\sum_{i=1}^{n} r_i$$

II. Maximum-Likelihood-Methode
Die Zufallsvariable $r$ ist Poisson-verteilt (für große $\mu$ kann natürlich auch Gauß-Verteilung benutzt werden)

$$P(r_i;\mu) = \frac{\mu^{r_i}\, e^{-\mu}}{r_i!}$$

$\Longrightarrow$ Likelihood-Funktion

$$L(\mu) = \prod_{i=1}^{n} P(r_i;\mu) = \prod_{i=1}^{n} \frac{\mu^{r_i}\, e^{-\mu}}{r_i!}$$

Universität Hamburg

# Beispiel zur Maximum-Likelihood-Methode

Negativer Logarithmus der Likelihood-Funktion

$$\ell(\mu) = -\ln L(\mu) = -\sum_{i=1}^{n} \ln \frac{\mu^{r_i}\, e^{-\mu}}{r_i!} = \sum_{i=1}^{n} (-r_i \ln \mu + \mu + \ln r_i!)$$

Ableiten nach dem Parameter $\mu$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\ell(\mu) = \frac{\mathrm{d}}{\mathrm{d}\mu} \sum_{i=1}^{n} (-r_i \ln \mu + \mu + \ln r_i!) = \sum_{i=1}^{n} \left(-r_i \frac{1}{\mu} + 1\right)$$

und zu Null setzen

$$0 \stackrel{!}{=} \sum_{i=1}^{n} \left(-r_i \frac{1}{\mu} + 1\right) = n - \frac{1}{\mu} \sum_{i=1}^{n} r_i$$

$$\Longrightarrow \mu = \frac{1}{n} \sum_{i=1}^{n} r_i$$

In diesem einfachen, analytisch zu lösenden Beispiel liefert die Maximum-Likelihood-Methode das gleiche Resultat wie die einfache Mittelung

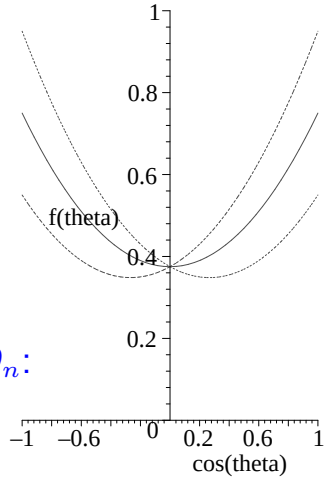


# 2. Beispiel Maximum-Likelihood-Methode

Die Verteilung des Polarwinkels $\theta$ von Elementarteilchen ist oft proportional $(1 + 8/3\, A \cos\theta + \cos^2\theta)$, z.B. in der Reaktion $\mathrm{e^+e^-} \to \mu^+\mu^-$, worin der Parameter $A$ die Vorwärts-Rückwärts-Asymmetrie ist

Die normierte pdf ist

$$f(\theta; A) = \frac{3}{8}\left(1 + \frac{8}{3} A \cos\theta + \cos^2\theta\right) \qquad 0 \le \theta \le \pi$$




Bestimmung des Parameters $A$ aus einer Stichprobe $\theta_1, \ldots \theta_n$:

- Zählen der Vorwärts- und Rückwärts-Ereignisse

$$A = \frac{N(\theta < \frac{\pi}{2}) - N(\theta > \frac{\pi}{2})}{N(\theta < \frac{\pi}{2}) + N(\theta > \frac{\pi}{2})}$$

- Nach der Maximum-Likelihood-Methode

$$\ell(A) = -\ln L(A) = -\sum_{i=1}^{n} \ln \frac{3}{8}\left(1 + \frac{8}{3} A \cos\theta_i + \cos^2\theta_i\right)$$

Die Maximum-Likehood-Methode liefert hier eine bessere Schätzung für den Parameter $A$, d.h. die Varianz der Schätzung bzw. der Fehler auf $A$ ist kleiner

$\Longrightarrow$ Übung

# Fehlerberechnung mit der Maximum-Likelihood-Methode

Der mit der Maximum-Likelihood aus einer statistischen Stichprobe ermittelte beste Schätzwert des Parameters $a$ kann selbst wieder als Zufallsvariable aufgefasst werden

Wie groß ist die Varianz $\sigma_a^2$ bzw. wie groß ist die Standardabweichung $\sigma_a$ des Schätzwertes?

- Im Grenzfall $n \to \infty$ nähert sich die Likelihood-Funktion $L(a)$ einer Gauß-Funktion
- mit dem Mittelwert $\hat{a}$, der dem wahren Wert des Parameters $a$ entspricht
- und der Varianz $\sigma_a^2 \to 0$

(ohne Beweis, aber intuitiv verständlich)

Wir können $\ell(a)$ um den wahren Wert $\hat{a}$ entwickeln $\left.\frac{\mathrm{d}\ell(a)}{\mathrm{d}a}\right|_{a=\hat{a}} = 0$

$$\ell(a) = \ell(\hat{a}) + \frac{1}{2}\left.\frac{\mathrm{d}^2\ell(a)}{\mathrm{d}a^2}\right|_{a=\hat{a}} (a-\hat{a})^2 + \ldots$$

oder für die Likelihood-Funktion:

$$L(a) \approx \text{const} \cdot e^{-\frac{1}{2}\left\{\left.\frac{\mathrm{d}^2\ell(a)}{\mathrm{d}a^2}\right|_{\hat{a}} (a-\hat{a})^2\right\}} \qquad \Longrightarrow \qquad \sigma_a^2 = \left(\left.\frac{\mathrm{d}^2\ell(a)}{\mathrm{d}a^2}\right|_{\hat{a}}\right)^{-1}$$

Universität Hamburg



# Beispiel

Wir kehren zurück zu unserem Beispiel des radioaktiven Zerfalls

Die Likelihood-Funktion der Poisson-verteilten Zahl der Zerfälle $r_i$ ist:

$$L(\mu) = \prod_{i=1}^{n} P(r_i;\mu) = \prod_{i=1}^{n} \frac{\mu^{r_i}\, e^{-\mu}}{r_i!}$$

$$\frac{\mathrm{d}}{\mathrm{d}\mu}\ell(\mu) = \sum_{i=1}^{n}\left(-r_i\frac{1}{\mu} + 1\right) \quad \stackrel{!}{=} \quad 0 \qquad \Longrightarrow \qquad \mu = \frac{1}{n}\sum_{i=1}^{n} r_i$$

Wie groß ist die Varianz des Parameters $\mu$?

Bilden der 2. Ableitung bei $\mu = \hat{\mu}$

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}\mu^2}\ell(\mu)\right|_{\mu=\hat{\mu}} = \frac{1}{\hat{\mu}^2}\sum_{i=1}^{n} r_i = \frac{1}{\hat{\mu}^2}\,\hat{\mu}\,n = \frac{n}{\hat{\mu}} = \frac{1}{\sigma_\mu^2}$$

$$\Longrightarrow \sigma_\mu^2 = \frac{\hat{\mu}}{n}$$

Auch das ist ein einleuchtendes Ergebnis!

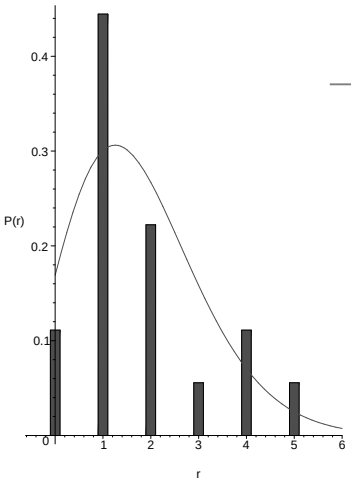
Wenn der wahre Wert $\hat{\mu}$ nicht bekannt ist, wird dieser zur Abschätzung der Varianz (Fehlers) durch den Schätzwert $\mu$ ersetzt

Universität Hamburg

# Numerisches Beispiel

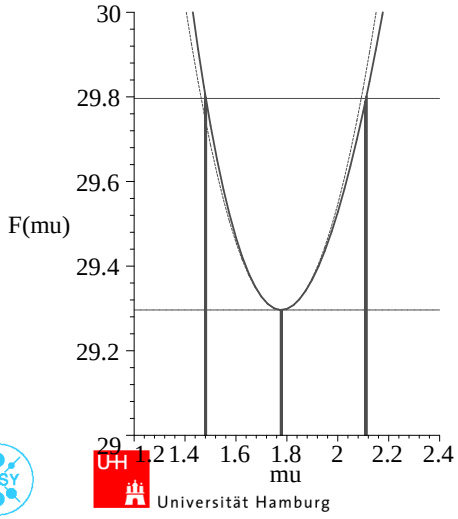
Wir betrachten eine Messreihe mit $n = 18$ Einzelmessungen $r_i$ zur Radioaktivität

$$r_i = [1, 1, 5, 4, 2, 0, 3, 2, 4, 1, 2, 1, 1, 0, 1, 1, 2, 1]$$




Die Daten sind mit der erwarteten Poisson-Verteilung verträglich

Der geschätzte, beste Wert für die mittlere Zahl der Zerfälle ist $\mu = \frac{1}{n} \sum_{i=1}^{n} r_i = 1,78$ und der geschätzte Fehler auf diesen Mittelwert ist $\sigma_\mu = \sqrt{\mu/n} = 0,31$




Der Logarithmus der Likelihood-Funktion $\ell(\mu) = -\ln L(\mu)$ kann in der Nähe des Minimums durch eine Parabel genähert werden

$$\ell(\mu) \approx \ell(\hat{\mu}) + \frac{1}{2} \frac{(\mu - \hat{\mu})^2}{\sigma_\mu^2}$$

Die Standardabweichung $\sigma_\mu$ kann von dieser Parabel an den Schnittpunkten mit der Horizontalen

$$\ell_{\min} + \frac{1}{2}$$

abgelesen werden



# Fehlerberechnung mit der Maximum-Likelihood-Methode

Im Allgemeinen kann die 2. Ableitung der Funktion $\ell(a)$ nicht analytisch berechnet werden. Wie bestimmt man den Fehler auf den Parameter $a$?

Die Antwort steht auf der vorherigen Seite:
Man bestimme die Stellen, bei der $\ell(a)$ auf $\ell_{\min} + \frac{1}{2}$ angewachsen ist:

$$\boxed{\ell(\hat{a} \pm \sigma_a) = \ell_{\min} + \frac{1}{2}}$$

- In der parabolischen Näherung ist $L(a) = e^{-\ell(a)}$ eine Gauß-Verteilung um den wahren Wert $\hat{a}$
  In einem Experiment ist die Wahrscheinlicheit $68\%$, den Parameter $a$ im Intervall $\hat{a} - \sigma_a \leq a \leq \hat{a} + \sigma_a$ zu finden
- Man gibt das Resultat des Experimentes durch den ermittelteten Schätzwert $a$ und die Standardabweichung $\sigma_a$ als Fehler an

$$a \pm \sigma_a$$

# Fehlerberechnung mit der Maximum-Likelihood-Methode

- Falls die parabolische Näherung nicht gut ist, ermittelt man asymmetrische Fehler $\sigma_l$, $\sigma_r$ durch

$$\boxed{\ell(\hat{a} - \sigma_l) = \ell(\hat{a} + \sigma_r) = \ell_{\min} + \frac{1}{2}}$$

  Man kann zeigen, dass dieses Intervall immer $68\%$ Wahrscheinlichkeit enthält
  - durchführen einer nichtlinearen Parametertransformation $b = b(a)$, so dass $\ell(b)$ parabolisch wird
  - die Transformation selber muss man nicht kennen, um Aussagen über $a$ zu machen (Erhaltung der Wahrscheinlichkeit)

- Das Resultat des Experimentes mit asymmetrischen Fehlern wird so angegeben:

$$a\,^{+\sigma_r}_{-\sigma_l}$$

- Für unser Experiment zur Radioaktivität heißt dass:
  Statt $\mu = 1,78 \pm 0,31$ würde man schreiben

$$\mu = 1,78\,^{+0,33}_{-0,30}$$



# Fehlerberechnung mit der Maximum-Likelihood-Methode

- Intervalle, die $k$ Standardabweichungen enthalten sollen, können analog ermittelt werden. Es gilt

$$\boxed{\ell(\hat{a} - k\sigma_l) = \ell(\hat{a} + k\sigma_r) = \ell_{\min} + \frac{k^2}{2}}$$

- So ist z.B. der $2\,\sigma$ Bereich des Parameters $a$, der $95\%$ Wahrscheinlichkeit enthält, definiert durch

$$\ell_{\min} + 2$$

- Dieser Zusammenhang folgt sofort aus den Eigenschaften der integrierten Gauß-Verteilung

# Verallgemeinerung für mehrere Parameter

Allgemeiner Fall der Maximum-Likelihood-Methode für $m$ Parameter $a_1, \ldots a_m$

Wir führen wieder den $m$-komponentigen Spaltenvektor $\vec{a}$ ein
Die Likelihood-Funktion für $n$ Messdaten schreibt sich dann:

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

Entwicklung des Logarithmus $\ell(\vec{a}\,) = -\ln L(\vec{a}\,)$ um die wahren Werte $\vec{\hat{a}}$

$$\begin{aligned} \ell(\vec{a}\,) &= \ell(\vec{\hat{a}}\,) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \ell(\vec{a}\,)}{\partial a_i\, \partial a_j}\,(a_i - \hat{a}_i)\,(a_j - \hat{a}_j) + \ldots \\ &= \ell(\vec{\hat{a}}\,) + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} G_{ij}\,(a_i - \hat{a}_i)\,(a_j - \hat{a}_j) + \ldots \end{aligned}$$

Die Likelihood-Funktion nähert sich asymptotisch einer $m$-dimensionalen Gauß-Funktion an (bis auf Normierungsfaktor)

$$L(\vec{a}\,) = e^{-\ell(\vec{a}\,)}$$

und die Matrix der doppelten partiellen Ableitungen ist die inverse Kovarianzmatrix des Vektors $\vec{a}$

$$G = V^{-1} \qquad \text{mit} \qquad G_{ij} = \frac{\partial^2 \ell(\vec{a}\,)}{\partial a_i\, \partial a_j}$$



# Integrierte Likelihood-Funktionen

In der Gauß'schen Näherung der Likelihood-Funktion können alle Resultate zur integrierten mehrdimensionalen Gauß-Verteilung übernommen werden

Die 1-$\sigma$-Kontur ist definiert durch $\ell(\vec{a}) = \ell(\vec{\hat{a}}) + \frac{1}{2}$
Die 2-$\sigma$-Kontur ist definiert durch $\ell(\vec{a}) = \ell(\vec{\hat{a}}) + 2$ etc.

und die entsprechenden Wahrscheinlichkeitsinhalte können durch Integration über Gauß-Funktionen berechnet werden

- Likelihood für 2 Parameter
  Die Wahrscheinlichkeit, das Parameter-Paar innerhalb der 1-$\sigma$-Kontur zu finden ist $39\%$
  In der parabolischen Näherung sind diese Konturen Ellipsen in der Ebene $a_1, a_2$
  Sonst ergeben sich asymmetrische Fehler und andere Kurven, deren Wahrscheinlichkeitsinhalt aber dem im Gauß'schen Fall entspricht
- Um den Fehler eines Parameters zu ermitteln, muss $\ell(\vec{a})$ bzgl. aller anderen Parameter minimiert werden
  Mit dieser Funktion $\ell'(\vec{a})$ kann dann durch Erhöhen des Minimums um $\frac{1}{2}$ der Fehler eines Parameters bestimmt werden

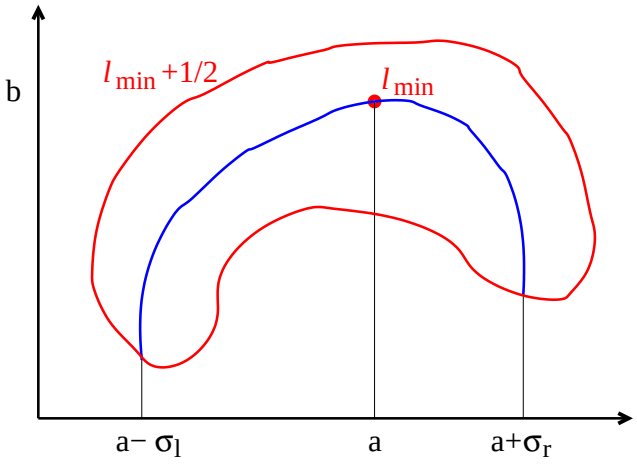
  $\Longrightarrow$ MINUIT: Programm zur Lösung des Minimierungsproblems

# Beispiel

Nebenstehend ist die Kontur des negativen Likelihood für 2 Parameter $a, b$ gezeigt

$$\ell(a,b) = \ell_{\min} + \frac{1}{2}$$

In diesem Beispiel ist die parabolische Näherung nicht gut erfüllt
Trotzdem gibt die rote Kontur den Bereich an, in dem das Paar $a, b$ mit $39\%$ Wahrscheinlichkeit gefunden wird (1-$\sigma$-Kontur)




Um die Standardabweichungen des einzelnen Parameters $a$ zu finden, muss $\ell(a,b)$ für festes $a$ bezüglich $b$ minimiert werden

$\Longrightarrow$ blaue Kurve

An den Stellen $\ell_{\min} + \frac{1}{2}$ dieser Kurve kann die Standardabweichung von $a$ abgelesen werden (bzw. $\sigma_l$ und $\sigma_r$)