

6 Jenseits der Methode der kleinsten Fehlerquadrate: Fitten eines Histogramms mit Maximum Likelihood

Lernziele Lektion 6

- Sie wissen, wie man eine Kurve an Histogrammdataen fittet, wenn die Statistik der einzelnen Bins einer Poissonverteilung genügt.
- Sie wissen, wie man Schätzwerte für die Kurvenparameter bestimmt und deren Genauigkeit graphisch oder numerisch abschätzt.

6.1 Das Experiment

Hintergrundinformation I: In einem Experiment messen Sie die Energieverteilung der emittierten Photonen einer radioaktiven Probe. Im Detektor wird jedes einzelne Photon in einen elektrischen Strompuls umgewandelt. Die Peakhöhe des Pulses ist ein Mass für die Energie des Photons. Mit Hilfe eines Vielkanalanalysators messen Sie die Peakhöhenverteilung und damit die Energieverteilung der Photonen in Form eines Histogramms. Solch ein gemessenes Histogramm ist in Abb. 28 dargestellt. Die Pulshöhen sind in Intervallen, sogenannten Bins, zusammengefasst. Die Höhe der über den Bins dargestellten Balken gibt die Zahl der gemessenen Photonen im Intervall an.

Sie erwarten in dieser Messung eine gaussförmige Resonanz bei der Energie E_0 , die im Histogramm bereits als der Ursprung der Energieskala gewählt wurde. Aus früheren Messungen ist die Breite der Resonanz bereits bekannt. Die Energieskala wurde gemäss $x = (E - E_0)/\sqrt{2}\Delta$ skaliert, wobei Δ die Breite der Resonanz ist. Die Breite der Bins ist eine halbe Einheit. In der Umgebung dieser Resonanz gibt es allerdings Hintergrundstrahlung, die einen konstanten Hintergrund zum Spektrum beiträgt.

Modell der Messung M. Entsprechend erwarten wir eine Energieverteilung der Form

$$f(x) = A \exp[-x^2] + B,$$

wobei $x = (E - E_0)/\Delta$, A die Amplitude der Resonanz und B der Hintergrund ist. Innerhalb eines Bins mit Zentrum bei x_j erwarten wir daher

$$\mu_j = \int_{x_j - \Delta x/2}^{x_j + \Delta x/2} f(x) dx \quad (35)$$

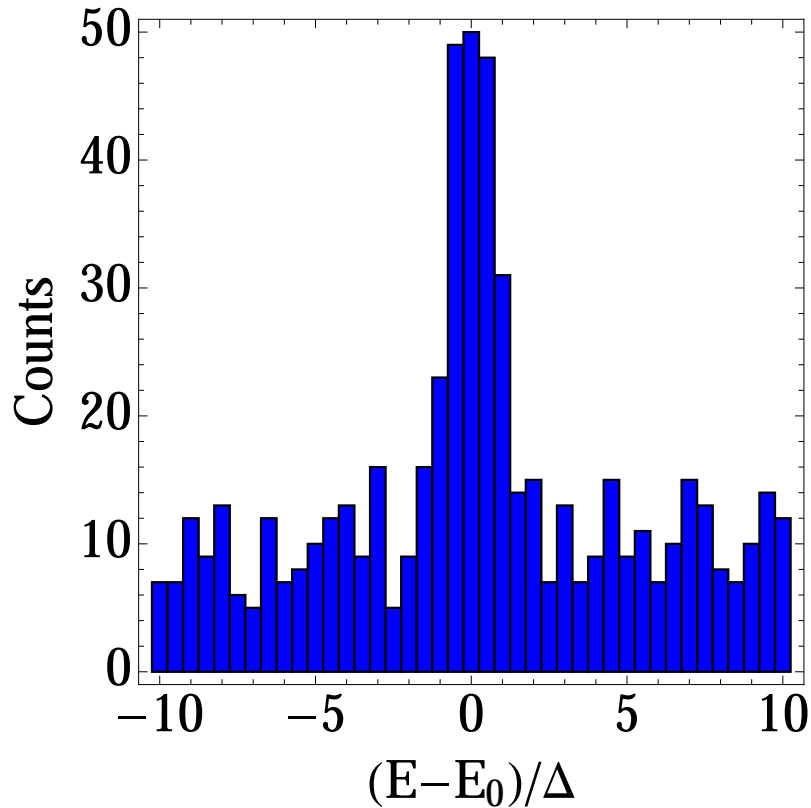


Abbildung 28: Histogramm der Energieverteilung der emittierten Photonen einer radioaktiven Probe. Der Nullpunkt und die Skala der Energie wurden so gewählt, dass dort eine gaussförmige Resonanz der Breite $\sqrt{2}$ liegt.

gezählte Photonen. Ist die Binbreite Δx genügend klein, dann kann dieser Ausdruck durch

$$\mu_j \approx f(x_j)\Delta x$$

genähert werden. In unserem Fall ist $\Delta x = 1/2$.

Die Wahrscheinlichkeit im j -ten Bin k_j Photonen zu messen ist dann durch die Poissonverteilung

$$\mathcal{P}(k_j; \mu_j) = \frac{\mu_j^{k_j}}{k_j!} e^{-\mu_j}$$

gegeben. Die Messung liefert uns einen Satz Wertepaare (x_j, k_j) (siehe Abb. 28).

Frage: wie gross sind die Amplitude A der Resonanz und der Hintergrund B ?

Likelihood: Mit Hilfe der x_j und der Gl. (35) sind wir in der Lage Werte $\mu_j(A, B)$ bei gegebenen Parametern A und B zu berechnen. Bei einer Gesamtzahl von N Bins ist die Wahrscheinlichkeitsverteilung für das Ergebnis $\{k_1, \dots, k_N\}$ gegeben durch

$$\text{pdf}(\{k_j\}|A, B, M, I) = \prod_{j=1}^N \mathcal{P}[k_j; \mu_j(A, B)] = \prod_{j=1}^N \frac{\mu_j^{k_j}}{k_j!} e^{-\mu_j}.$$

6.2 Abschätzung der Modellparameter A und B

Priorwahrscheinlichkeit. Als Priorwahrscheinlichkeit nehmen wir gemäss dem Indifferenzprinzip die Verteilung

$$\text{pdf}(A, B|M, I) = \begin{cases} \text{const.} & \text{für } A, B > 0 \\ 0 & \text{sonst.} \end{cases}$$

Posteriorverteilung. Mit Hilfe des Bayes'schen Theorems erhalten wir dann die Posteriorverteilung für die zwei Parameter A und B

$$\text{pdf}(A, B|\{k_j\}, \{x_j\}, M, I) \propto \prod_{j=1}^N \frac{\mu_j^{k_j}(A, B)}{k_j!} e^{-\mu_j(A, B)}.$$

Wegen der konstanten Priorverteilung ist das gerade die Likelihood-Funktion. Diese Funktion ist in Abb. 29 dargestellt.

Abschätzung der gesuchten Parameter. Um die gesuchten Parameter A und B abzuschätzen, suchen wir das Maximum dieser Verteilung. Dieses Verfahren ist als *Maximum Likelihood Method* bekannt.

Wie bereits früher ist es praktisch, anstelle die Posteriorverteilung zu maximieren, deren negativen Logarithmus zu minimieren. Wir bilden daher

$$\begin{aligned} L(A, B) = -\ln \text{pdf}(A, B|\{k_j\}, \{x_j\}, M, I) &= \text{const.} - \sum_{j=1}^N \ln \left(\frac{\mu_j^{k_j}(A, B)}{k_j!} e^{-\mu_j(A, B)} \right) \\ &= \text{const.} - \sum_{j=1}^N [k_j \ln \mu_j(A, B) - \mu_j(A, B)]. \end{aligned}$$

Diese Funktion ist in Abb. 30 dargestellt. Die Minimierung führen wir numerisch durch und erhalten so die Schätzwerte \hat{A} und \hat{B} (roter Punkt in der Abbildung). Abbildung 31 zeigt die Daten zusammen mit der Kurve, die mit \hat{A} und \hat{B} gezeichnet wurde.

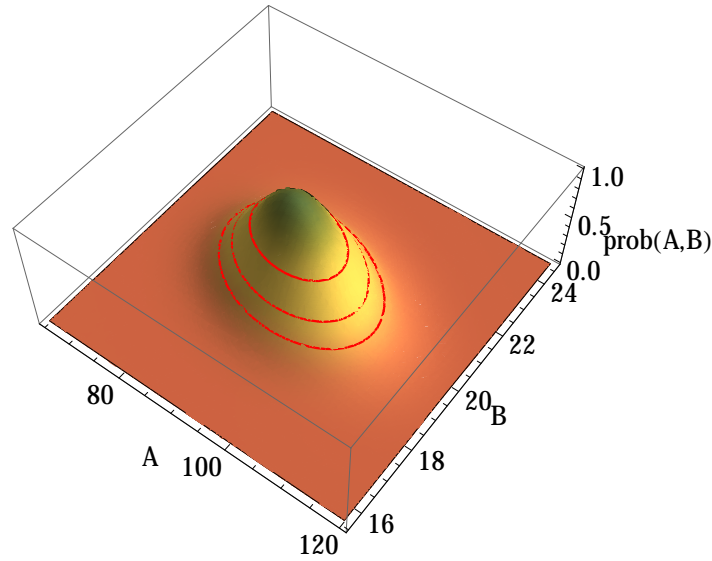


Abbildung 29: Posteriorverteilung der beiden Parameter A und B . Der Wert am Maximum der Verteilung wurde auf Eins skaliert. Die drei roten Linien sind Höhenlinien bei $1/\sqrt{e}$, $1/e$ und $1/e^{3/2}$ vom Maximalwert.

Genauigkeit der abgeschätzten Parameter. Wie beim Fitten nichtlinearer Kurven an Messdaten geben wir die Genauigkeit der abgeschätzten Parameter als den Parameterbereich an, innerhalb dessen die Posteriorverteilung auf ihren $1/\sqrt{e}$ -ten Teil abgefallen ist. Die Höhenlinie $2(L(A, B) - L_{\min}) = 1$ hilft uns bei der graphischen Bestimmung (siehe Abb. 30). Die roten gestrichelten Linien in der Abbildung zeigen uns, welche Werte wir für die Fehlerbalken angeben:

$$A = 91.6 \begin{matrix} +8.9 \\ -8.4 \end{matrix} \quad \text{und} \quad B = 20.0 \pm 1.1.$$

Die Tatsache, dass die Fehlerbalken nach oben und unten verschieden sind rührt daher, dass die Höhenlinie $2(L(A, B) - L_{\min}) = 1$ keine exakte Ellipse ist. Dennoch erkennen wir an der Verkipfung der Höhenlinie bezüglich der Koordinatenachsen, dass es eine gewisse (negative) Korrelation zwischen den beiden Parametern A und B gibt.

Oft kann man die Genauigkeit der abgeschätzten Parameter auch durch eine Entwicklung der Funktion $L(A, B)$ um das Minimum (\hat{A}, \hat{B}) erhalten. Das ergibt

$$L(A, B) \approx L_{\min} + \frac{1}{2} \begin{pmatrix} A - \hat{A} \\ B - \hat{B} \end{pmatrix} \underbrace{\begin{pmatrix} \frac{\partial^2 L}{\partial A^2} & \frac{\partial^2 L}{\partial A \partial B} \\ \frac{\partial^2 L}{\partial A \partial B} & \frac{\partial^2 L}{\partial B^2} \end{pmatrix}}_{:=H} \Big|_{(\hat{A}, \hat{B})} \begin{pmatrix} A - \hat{A} \\ B - \hat{B} \end{pmatrix}.$$

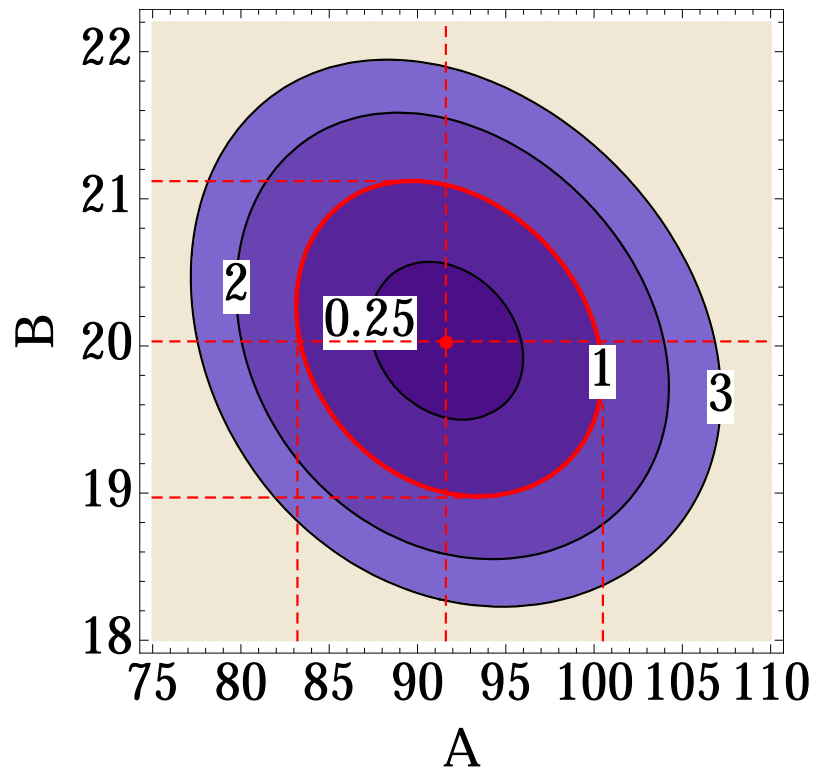
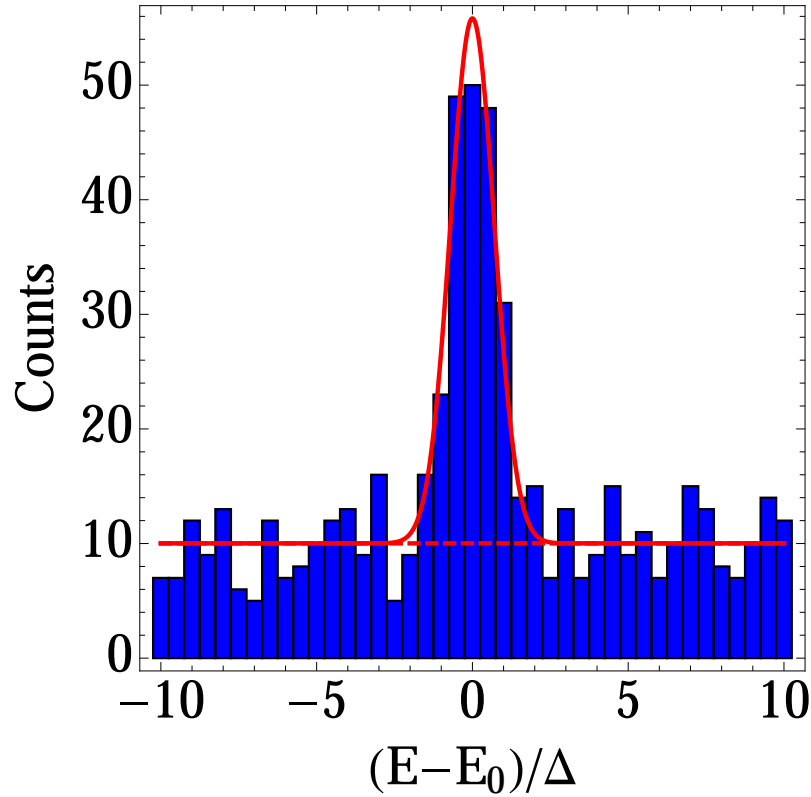


Abbildung 30: Die Funktion $2(L(A, B) - L_{\min})$ in der Nähe ihres Minimums. Der rote Punkt kennzeichnet die Lage des Minimums. Die eingezeichneten Höhenlinien liegen bei 0.25, 1, 2, und 3. Die gestrichelten roten Linien geben die Fehlerbalken für A und B an.



Abbildungung 31: Maximum likelihood fit an die Daten. Die gestrichelte rote Linie gibt die Position des Hintergrundes an.

Die Matrix H der partiellen Ableitungen ist die Hesse Matrix.

Die Elemente der Hesse Matrix können in der Regel nur numerisch berechnet werden. Dazu bietet es sich zum Beispiel an, die zweiten Ableitungen durch finite Differenzen anzunähern. In diesem Sinne hätten wir in unserem Beispiel

$$\begin{aligned} \left. \frac{\partial^2 L}{\partial A^2} \right|_{\hat{A}, \hat{B}} &\approx \frac{L(\hat{A} + h, \hat{B}) - 2L(\hat{A}, \hat{B}) + L(\hat{A} - h, \hat{B})}{h^2} \\ \left. \frac{\partial^2 L}{\partial B^2} \right|_{\hat{A}, \hat{B}} &\approx \frac{L(\hat{A}, \hat{B} + k) - 2L(\hat{A}, \hat{B}) + L(\hat{A}, \hat{B} - k)}{k^2} \\ \left. \frac{\partial^2 L}{\partial A \partial B} \right|_{\hat{A}, \hat{B}} &\approx \frac{L(\hat{A} + h, \hat{B} + k) - L(\hat{A} + h, \hat{B} - k) - L(\hat{A} - h, \hat{B} + k) + L(\hat{A} - h, \hat{B} - k)}{4hk}. \end{aligned}$$

Die Werte von h und k sind dabei hinreichend klein zu wählen, damit die Näherung gut

wird. Mit etwas Erfahrung kann man geeignete Werte aus dem Plot von L (siehe Abb. 30) ablesen.

In unserem konkreten Beispiel findet man

$$H = \begin{pmatrix} 0.01399 & 0.02453 \\ 0.02453 & 0.91125 \end{pmatrix}.$$

Die Kovarianzmatrix der abgeschätzten Parameter ist dann durch

$$\begin{pmatrix} \sigma_A^2 & \sigma_A \sigma_B \rho \\ \sigma_A \sigma_B \rho & \sigma_B^2 \end{pmatrix} = H^{-1}$$

gegeben. In unserem konkreten Fall ergibt sich

$$\begin{pmatrix} \sigma_A^2 & \sigma_A \sigma_B \rho \\ \sigma_A \sigma_B \rho & \sigma_B^2 \end{pmatrix} = \begin{pmatrix} 75.04 & -2.020 \\ -2.020 & 1.152 \end{pmatrix}.$$

Daraus gewinnt man

$$A = 91.6 \pm 8.7 \quad \text{und} \quad B = 20.0 \pm 1.1$$

in Übereinstimmung mit dem graphisch gewonnenen Ergebnis. Zudem findet man den Korrelationskoeffizienten

$$\rho = -0.22.$$

Es bestätigt sich die Vermutung, dass eine negative Korrelation zwischen den beiden Parametern besteht. Dieses numerische Verfahren der Fehlerabschätzung lässt sich auf Datenfits mit mehreren Parametern verallgemeinern.

6.3 Aufgaben und Fragen zum tieferen Verständnis

1. Warum ist der Schätzwert $\hat{B} \approx 20$, wohingegen der Wert des Hintergrundsignals in Abb. 31 bei 10 liegt?
2. Begründen Sie, warum die Höhenlinie $2(L(A, B) - L_{\min}) = 1$ gewählt wurde, um den Fehler graphisch zu ermitteln. Benutzen Sie dabei die Analogien zur Methode der kleinsten Fehlerquadrate.
3. Diskutieren Sie, ob die Auswertung der Daten $\{(x_j, k_j)\}$ auch durchgeführt werden könnte, ohne die Counts auf Bins eines Histogramms aufzuteilen. In diesem Fall bestünde das Messergebnis aus einer Serie von Werten $\{x_j\}$, bei denen ein Photon beobachtet wurde. Wie würde sich die Analyse ändern? Welchen Vorteil, bzw. welchen Nachteil bringt das ‘binning’ der Daten mit sich?