

Histogramme fitten

T. Ihn

8. November 2016

1 Statistische Prozesse mit konstanter Rate

In der Physik treten häufig Prozesse auf, die wegen der zugrunde liegenden Quantenmechanik statistischer Natur sind. Beispiele sind der radioaktive Zerfall, die Photonenemission eines angeregten Atoms, oder das Tunneln eines Elektrons durch eine Potentialbarriere. Diese Prozesse haben eine Gemeinsamkeit: sie werden durch eine zeitlich konstante Rate beschrieben, so dass die Wahrscheinlichkeit, dass das Ereignis im infinitesimal kleinen Zeitintervall dt passiert, durch

$$\text{prob}(\text{Event happens}|\lambda) = \lambda dt$$

gegeben ist, wobei λ die Rate ist, die in Einheiten s^{-1} gemessen wird. Entsprechend ist die Wahrscheinlichkeit, dass das Ereignis in diesem Zeitintervall nicht eintritt, durch

$$\text{prob}(\text{Event does not happen}|\lambda) = 1 - \lambda dt$$

gegeben.

2 Verteilung der Wartezeit

Aus der obigen einfachen Grundannahme der konstanten Rate lässt sich die Verteilung der Wartezeit (engl. ‘waiting time distribution’) ableiten. Dabei fragt man nach der Wahrscheinlichkeit, dass das Ereignis erst nach einer Wartezeit t eintritt. Dazu berechnen wir zunächst die Wahrscheinlichkeit, dass innerhalb der Wartezeit t das Ereignis nicht eintritt. Sie ist durch

$$\text{prob}(\text{nothing happens until } t|\lambda) = \lim_{n \rightarrow \infty} \left(1 - \lambda \frac{t}{n}\right)^n = e^{-\lambda t}.$$

Hier wurde die Idee verwendet, das Zeitintervall t in n kleine Subintervalle t/n aufzuteilen, und anschliessend den Grenzwert für infinitesimal kleine Intervalle durch $n \rightarrow \infty$ zu realisieren. Die Gesamtwahrscheinlichkeit ist damit

$$\text{prob}(\text{nothing happens until } t \text{ and at } t \text{ the event happens}|\lambda) = e^{-\lambda t} \lambda dt.$$

Die Exponentialverteilung

$$\text{pdf}(t|\lambda) = \lambda e^{-\lambda t}$$

ist daher die Verteilung der Wartezeiten für einen statistischen Prozess mit zeitlich konstanter Rate λ .

Eigenschaften der Exponentialverteilung. Die Exponentialverteilung besitzt den Mittelwert

$$\langle t \rangle = \frac{1}{\lambda}$$

und die Varianz

$$\text{Var}(t) = \frac{1}{\lambda^2}.$$

Ihr Maximum (Mode) liegt bei $t = 0$.

3 Zählstatistik

Für die gleiche Klasse von Prozessen mit konstanter Rate lässt sich auch eine andere statistische Verteilung, die sogenannte Zählstatistik berechnen. Dabei stellt man die Frage nach der Zahl k von Ereignissen in einem gegebenen endlichen Zeitintervall Δt . Hier hilft uns die Binomialverteilung weiter, also

$$\text{prob}(k|\lambda, \Delta t) = \lim_{n \rightarrow \infty} \binom{n}{k} (\lambda \Delta t / n)^k (1 - \lambda \Delta t / n)^{n-k}.$$

Dabei haben wir das Zeitintervall Δt in n kleine Subintervalle zerlegt, deren Grösse wir mit dem Limes $n \rightarrow \infty$ infinitesimal klein werden lassen. Wir bezeichnen $\mu = \lambda \Delta t$ und finden damit

$$\begin{aligned} \text{prob}(k|\mu) &= \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} (\mu/n)^k (1 - \mu/n)^{n-k} \\ &= \frac{\mu^k}{k!} \lim_{n \rightarrow \infty} \overbrace{\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}}^{k \text{ factors}} (1 - \mu/n)^{n-k} \\ &= \frac{\mu^k}{k!} \lim_{n \rightarrow \infty} (1 - 1/n)(1 - 2/n)\dots(1 - (k+1)/n)(1 - \mu/n)^{n-k} \\ &= \frac{\mu^k}{k!} e^{-\mu}. \end{aligned}$$

Die Verteilung

$$\text{prob}(k|\mu) = \frac{\mu^k}{k!} e^{-\mu}$$

ist die Poissonverteilung. Sie ist die Zählstatistik für Prozesse mit konstanter Rate, wobei $\mu = \lambda \Delta t$ ist, mit der Rate λ und dem Zeitintervall Δt während dem gezählt wird.

Eigenschaften der Poissonverteilung. Der Erwartungswert für die Zahl k der Events ergibt sich aus der Poissonverteilung gemäss

$$\langle k \rangle = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = \mu \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} e^{-\mu} = \mu \sum_{k=0}^{\infty} \frac{\mu^k}{k!} e^{-\mu} = \mu.$$

Entsprechend ist

$$\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 \frac{\mu^k}{k!} e^{-\mu} = \mu \sum_{k=0}^{\infty} (k+1) \frac{\mu^k}{k!} e^{-\mu} = \mu(\mu+1) = \mu^2 + \mu.$$

Damit ist die Varianz der Poissonverteilung gegeben durch

$$\text{Var}(k) = \langle k^2 \rangle - \langle k \rangle^2 = \mu,$$

und die Standardabweichung durch

$$\sigma_k = \sqrt{\text{Var}(k)} = \sqrt{\mu}.$$

4 Experimente, die Histogramme generieren

4.1 Das Experiment

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 über eine bestimmte Zeit Δt . Solch ein gemessenes Histogramm ist in Abb. 1 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 gemäss der theoretischen Beschreibung des Prozesses 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

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

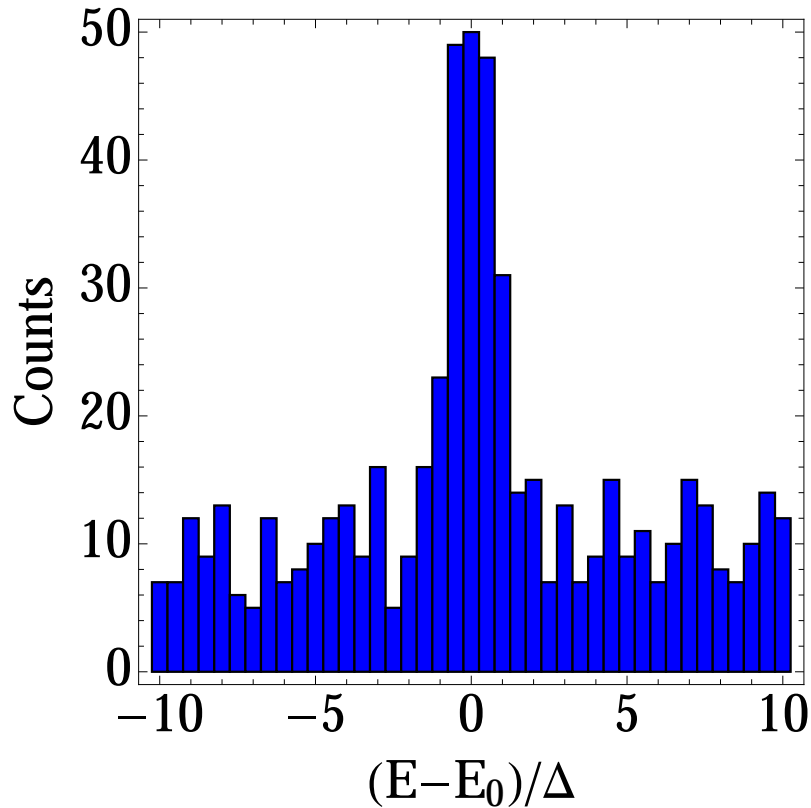


Abbildung 1: Histogramm der Energieverteilung der emittierten Photonen einer radioaktiven Probe. Der Nullpunkt und die Skala der Energie wurden so gewählt, dass dort eine gaussförmige Resonanz der Breite $\sqrt{2}$ liegt.

wobei $x = (E - E_0)/\Delta$, A die Amplitude der Resonanz und B der Hintergrund ist. Diese Funktion ist die Fitfunktion, die wir durch geeignete Wahl der Parameter A und B an das Histogramm anpassen wollen.

Innerhalb eines Bins mit Zentrum bei x_j erwarten wir daher

$$\mu_j(A, B) = \int_{x_j - \Delta x/2}^{x_j + \Delta x/2} f(x; A, B) dx \quad (1)$$

gezählte Photonen im Zeitintervall Δt . Entsprechend ist $\mu_j/\Delta t$ die (zeitlich konstante) Rate für bin j . Ist die Binbreite Δx genügend klein, dann kann dieser Ausdruck durch

$$\mu_j(A, B) \approx f(x_j; A, B) \Delta x$$

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 (die Zählstatistik des Bins j)

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

gegeben.

Die Messung liefert uns einen Satz Wertepaare (x_j, k_j) (siehe Abb. 1). Das ist völlig analog zur Gauss'schen Datenanalyse. Dort konnten wir jedem Messpunkt einen Fehlerbalken $\pm\sigma$ gemäss der Standardabweichung σ der Normalverteilung zuordnen. Der entsprechende Fehlerbalken in unserem Fall des Histogramms wäre die Wurzel aus der Varianz der Poissonverteilung, also

$$\sigma_j = \sqrt{\mu_j}.$$

Frage: Wie gross sind die Amplitude A der Resonanz und der Hintergrund B bei den gegebenen Messdaten (x_j, k_j) ?

Likelihood: Mit Hilfe der x_j und der Gl. (1) 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

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

4.2 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. 2 dargestellt.

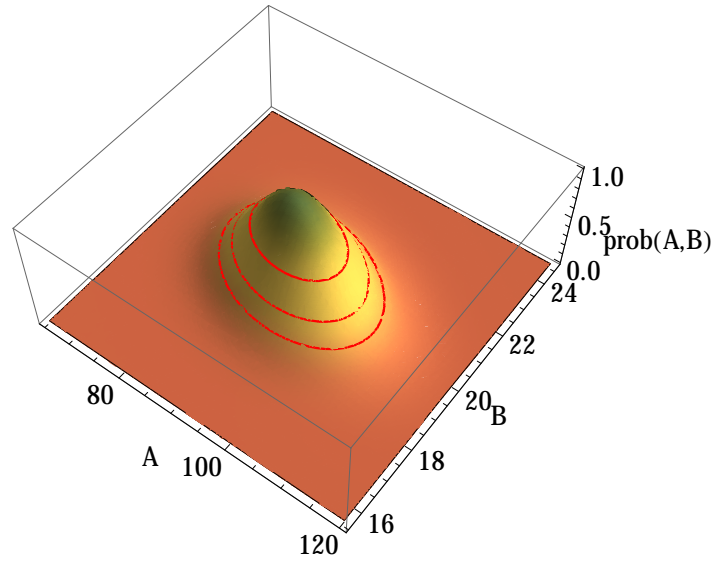


Abbildung 2: Posteriorverteilung der beiden Parameter A und B . Der Wert am Maximum der Verteilung wurde auf Eins skaliert. Die drei roten Linien sind Höhenlinien bei $1/\sqrt{e}$, $1/e$ und $1/e^{3/2}$ vom Maximalwert.

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. 3 dargestellt. Die Minimierung führen wir numerisch durch.

Numerisches Minimieren der log-Likelihood in Python. Wir führen die numerische Minimierung in Python mit Hilfe der Funktion `fmin` aus dem Paket `scipy.optimize` durch. Der folgende Codeausschnitt zeigt die Definition der Fitfunktion, der Erwartungswerte für die einzelnen Bins, sowie die Definition der log-Likelihood. Anschliessend wird die Likelihood minimiert, und die Parameterwerte am Minimum werden extrahiert.

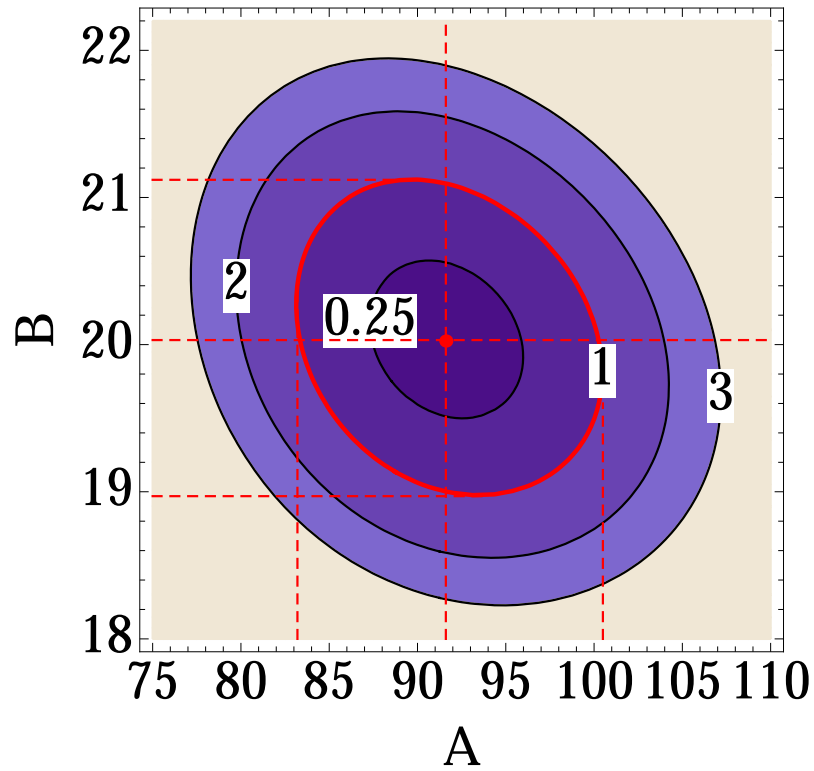


Abbildung 3: Die Funktion $2(L(A, B) - L_{\min})$ in der Nähe ihres Minimums. Der rote Punkt kennzeichnet die Lage des Minimums. Die eingezeichneten Höhenlinien liegen bei 0.25, 1, 2, und 3. Die gestrichelten roten Linien geben die Fehlerbalken für A und B an.

```

import numpy as np
from scipy.optimize import fmin

#
# Function calculating the expected number of counts in a bin given A,B
#
def mu(A,B,x):
    return (A*np.exp(-x**2) + B)/2.

#
# Function calculating the negative log-likelihood for given A,B
#
def L((A,B),xi,ci):
    mui = mu(A,B,xi)
    return -np.sum(np.multiply(ci,np.log(mui)) - mui)

#
# Minimization of the negative log-likelihood
#
parms,Lmin,_,_,_ = fmin(L,[100,20],args=(xi,ci),full_output=True)
A0 = parms[0]
B0 = parms[1]

```

So erhalten wir die Schätzwerte \hat{A} und \hat{B} (roter Punkt in der Abbildung). Abbildung 4 zeigt die Daten zusammen mit der Kurve, die mit \hat{A} und \hat{B} gezeichnet wurde.

Genauigkeit der abgeschätzten Parameter (graphisch). 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. 3). 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.

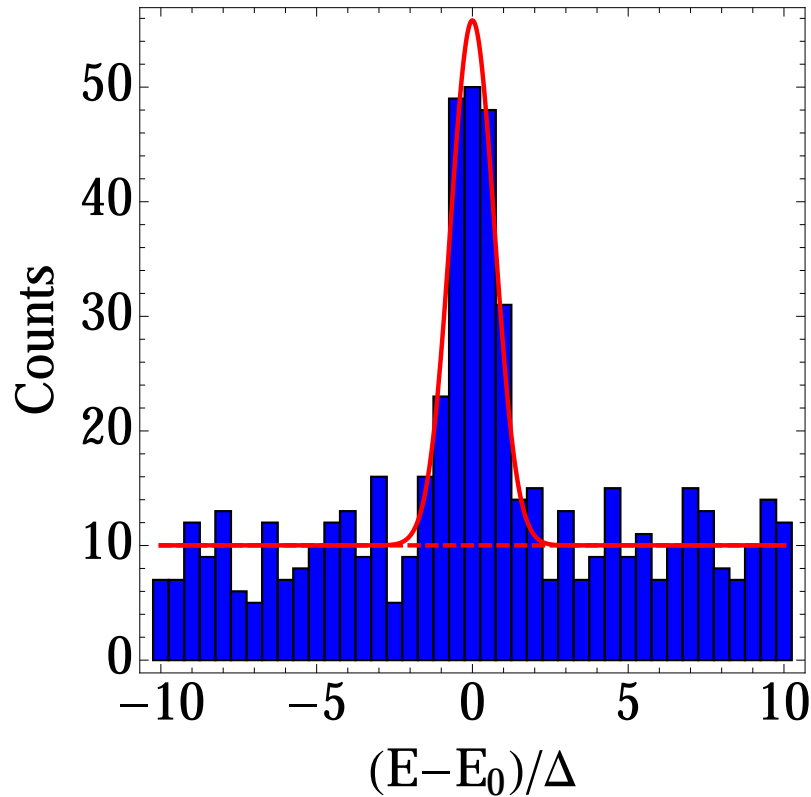


Abbildung 4: Maximum likelihood fit an die Daten. Die gestrichelte rote Linie gibt die Position des Hintergrundes an.

Genauigkeit der abgeschätzten Parameter (numerisch). 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

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

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

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

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. 3) 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.

Berechnung der Kovarianzmatrix mit Python.

```
#
# Determine the elements of the Hesse matrix numerically
#
```

```

h = 1.;
k = 0.1;
d2LdA2 = (L([A0+h,B0],xi,ci) - 2*L([A0,B0],xi,ci) + L([A0-h,B0],xi,ci))/(h*h)
d2LdB2 = (L([A0,B0+k],xi,ci) - 2*L([A0,B0],xi,ci) + L([A0,B0-k],xi,ci))/(k*k)
d2LdAdB = (L([A0+h,B0+k],xi,ci) - L([A0+h,B0-k],xi,ci) - L([A0-h,B0+k],xi,ci) + L([A0-h,B0-
# invert the Hesse matrix
cov = np.linalg.inv([[d2LdA2,d2LdAdB],[d2LdAdB,d2LdB2]])

# extract errors and correlation coefficient
sigmaA = np.sqrt(cov[0,0])
sigmaB = np.sqrt(cov[1,1])
rho = cov[0,1]/sigmaA/sigmaB

```

5 Gauss'sche Näherung der Poissonverteilung

Im Grenzfall für grosse Werte von μ geht die Poissonverteilung näherungsweise in eine Gaussverteilung über. Ist die Verteilung um einen grossen Wert von μ mit einer Breite von $\pm\sqrt{\mu}$ konzentriert, so sind auch nur grosse Werte von $k \approx \mu$ relevant. Wir nähern daher $k!$ mit der Stirling Formel

$$\begin{aligned}
k! &\approx \sqrt{2\pi k} e^{-k} k^k = \sqrt{2\pi(\mu + \delta x)} e^{-(\mu + \delta x)} (\mu + \delta x)^{\mu + \delta x} \\
&= \sqrt{2\pi\mu(1 + \delta x/\mu)} e^{-\mu(1 + \delta x/\mu)} (\mu(1 + \delta x/\mu))^{\mu(1 + \delta x/\mu)}.
\end{aligned}$$

Hier haben wir die Abweichung δx eingeführt, so dass $\delta x = k - \mu$. Die Idee dahinter ist, dass für grosse μ die relative Abweichung $\delta x/\mu \sim 1/\sqrt{\mu}$ ein kleiner Parameter ist. Wir bezeichnen $\epsilon = \delta x/\mu$ und haben für die Poissonverteilung den Ausdruck

$$P(k|\mu) = \frac{\mu^{\mu(1+\epsilon)}}{\sqrt{2\pi\mu(1+\epsilon)} e^{-\mu(1+\epsilon)} (\mu(1+\epsilon))^{\mu(1+\epsilon)}} e^{-\mu}.$$

Logarithmieren wir diesen Ausdruck, und entwickeln ihn in eine Taylorreihe bis zur zweiten Ordnung in ϵ , so erhalten wir

$$\ln P(k|\mu) \approx -\frac{1}{2} \ln(2\pi\mu) - \frac{1}{4(1-2\mu)} + \frac{1-2\mu}{4} \left(\epsilon - \frac{1}{1-2\mu} \right)^2.$$

Wir ersetzen $\epsilon = (k - \mu)/\mu$ und finden

$$\ln P(k|\mu) \approx -\frac{1}{2} \ln(2\pi\mu) - \frac{1}{4(1-2\mu)} + \frac{1-2\mu}{4} \left(\frac{k-\mu}{\mu} - \frac{1}{1-2\mu} \right)^2.$$

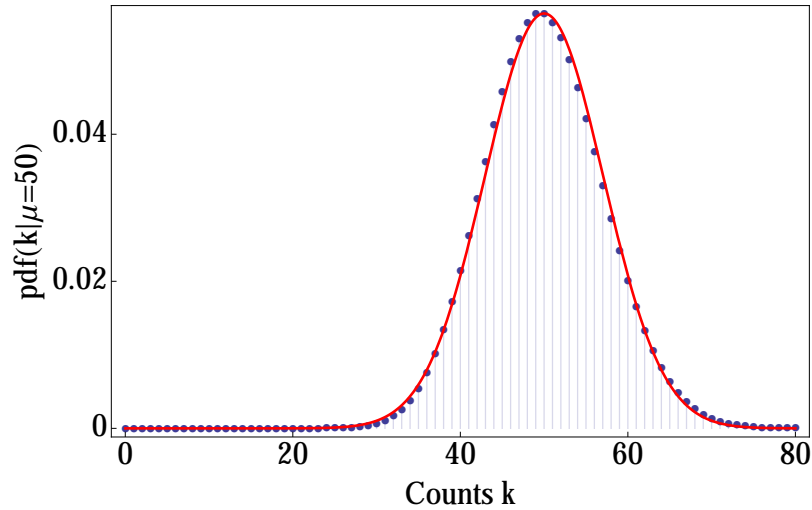


Abbildung 5: Vergleich der Poissonverteilung für $\mu = 50$ (blue Punkte) mit der Gauss'schen Näherung (rote Linie).

Im Grenzfall für grosse μ wird daraus

$$\ln P(k|\mu) \approx -\frac{1}{2} \ln(2\pi\mu) - \frac{1}{2} \frac{(k - \mu)^2}{\mu},$$

so dass

$$P(k|\mu) \approx \frac{1}{\sqrt{2\pi\mu}} \exp\left(-\frac{1}{2} \frac{(k - \mu)^2}{\mu}\right).$$

Das ist eine Normalverteilung mit Mittelwert μ und Standardabweichung $\sqrt{\mu}$. Abbildung 5 zeigt die Poissonverteilung für $\mu = 50$ im Vergleich zur Gauss'schen Näherungskurve. Diese Kurve wird hier als kontinuierliche Linie dargestellt. Tatsächlich nimmt k jedoch nur ganzzahlige Werte an. Die Näherung wird immer besser, je grösser μ ist.