2
Parameter estimation

Olaf Behnke and Lorenzo Moneta

Exercise 2.1: Efficiencies and averages

a) Determine the total decay rate R. For this, correct the decay rates N_1, N_2 of the two counters for their respective efficiencies ϵ_1 and ϵ_2 and then add the corrected numbers:

$$R = \frac{N_1}{\epsilon_1} + \frac{N_2}{\epsilon_2} = \frac{99}{0.99} + \frac{4}{0.04} = 100 + 100 = 200\,.$$

For the uncertainty of this estimate find from error propagation

$$\sigma_R = \sqrt{\left(\frac{\sigma_{N_1}}{\epsilon_1}\right)^2 + \left(\frac{\sigma_{N_2}}{\epsilon_2}\right)^2} = \sqrt{(9/0.99)^2 + 50^2} = 50.8\,.$$

The uncertainty is dominated by the single measurement with the largest uncertainty (in this case the second measurement).

b) Here we have two independent measurements of the total rate R:

Measurement 1:

$$R_1 = \frac{N_1}{\epsilon_1 \cdot \epsilon_{geom1}} = 200 \pm 18\,.$$

Here ϵ_{geom1} denotes the geometric acceptance of the first detector which is assumed to be 0.5.

Measurement 2:

$$R_2 = \frac{N_2}{\epsilon_2 \cdot \epsilon_{geom2}} = 200 \pm 100\,.$$

Denoting the uncertainties of the two rate measurements as σ_1 and σ_2, determine the weighted average $\hat{R}$ and its uncertainty $\sigma_{\hat{R}}$ (see equations (2.18) and (2.19) in the book):

$$\hat{R} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \cdot \left(\frac{R_1}{\sigma_1^2} + \frac{R_2}{\sigma_2^2}\right) = \frac{1}{\frac{1}{18^2} + \frac{1}{100^2}} \cdot \left(\frac{200}{18^2} + \frac{200}{100^2}\right) = 200\,;$$

$$\sigma_{\hat{R}} = \left(\frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \right)^{-0.5} = \left(\frac{1}{\frac{1}{18^2} + \frac{1}{100^2}} \right)^{-0.5} = 17.7\,.$$

In this case, the uncertainty is smaller than for the individual measurements, and it is dominated by the single measurement with the smallest uncertainty.

Exercise 2.2: Weighted average and χ^2

First note that the weighted average $\hat{\theta}$ and its variance $\sigma^2_{\hat{\theta}}$ are given by (see equations (2.18) and (2.19) in the book)

$$\hat{\theta} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} \cdot \left(\frac{y_1}{\sigma_1^2} + \frac{y_2}{\sigma_2^2} \right) = \frac{g_1 y_1 + g_2 y_2}{g_1 + g_2}\,, \quad \text{with } g_i := 1/\sigma_i^2 \text{ for } i = 1, 2\,, \text{ and}$$

$$\sigma^2_{\hat{\theta}} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{1}{g_1 + g_2}\,.$$

Then rewrite χ^2 as

$$\begin{aligned}
\chi^2 &= \frac{(y_1-\theta)^2}{\sigma_1^2} + \frac{(y_2-\theta)^2}{\sigma_2^2} = \frac{(y_1-\hat{\theta}-(\theta-\hat{\theta}))^2}{\sigma_1^2} + \frac{(y_2-\hat{\theta}-(\theta-\hat{\theta}))^2}{\sigma_2^2} \\
&= \frac{(y_1-\hat{\theta})^2}{\sigma_1^2} + \frac{(y_2-\hat{\theta})^2}{\sigma_2^2} - 2(\theta-\hat{\theta}) \cdot \underbrace{\left[\frac{(y_1-\hat{\theta})}{\sigma_1^2} + \frac{(y_2-\hat{\theta})}{\sigma_2^2} \right]}_{=0} + (\theta-\hat{\theta})^2 \cdot \underbrace{\left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right]}_{=1/\sigma^2_{\hat{\theta}}} \\
&= \frac{(y_1-\hat{\theta})^2}{\sigma_1^2} + \frac{(y_2-\hat{\theta})^2}{\sigma_2^2} + \frac{(\theta-\hat{\theta})^2}{\sigma^2_{\hat{\theta}}}\,.
\end{aligned}$$

The last term is already one of the two χ^2 terms to be shown in the problem. The first two terms define the minimum of the χ^2. They can be rewritten as

$$\begin{aligned}
\chi^2_{min} &= \frac{(y_1-\hat{\theta})^2}{\sigma_1^2} + \frac{(y_2-\hat{\theta})^2}{\sigma_2^2} = g_1 \cdot \left(y_1 - \frac{g_1 y_1 + g_2 y_2}{g_1 + g_2} \right)^2 + g_2 \cdot \left(y_2 - \frac{g_1 y_1 + g_2 y_2}{g_1 + g_2} \right)^2 \\
&= g_1 \cdot \left(\frac{g_2 y_1 - g_2 y_2}{g_1 + g_2} \right)^2 + g_2 \cdot \left(\frac{g_1 y_2 - g_1 y_1}{g_1 + g_2} \right)^2 \\
&= \frac{g_1 g_2^2}{(g_1+g_2)^2} \cdot (y_1 - y_2)^2 + \frac{g_2 g_1^2}{(g_1+g_2)^2} \cdot (y_1 - y_2)^2 \\
&= \frac{g_1 g_2 (g_1 + g_2)}{(g_1+g_2)^2} \cdot (y_1 - y_2)^2 = \frac{g_1 \cdot g_2}{g_1 + g_2} \cdot (y_1 - y_2)^2 \\
&= \frac{1}{1/g_1 + 1/g_2} \cdot (y_1 - y_2)^2 = \frac{(y_1 - y_2)^2}{\sigma_1^2 + \sigma_2^2}\,,
\end{aligned}$$

providing the other χ^2 term stated in the problem.

Exercise 2.3: Unbinned fits 1

The properly normalised probability density function f in this problem is $f(x;\lambda) = 0.5(1+\lambda\, x)$ with $x = \cos\theta$. In the top left of figure 2.1, the data points (the ten spikes) and the fitted curve (dashed line) are shown. The top right shows the log-likelihood curve based on the likelihood

$$L = \prod_{i=1}^{10} f(x_i;\lambda) = \prod_{i=1}^{10} 0.5(1+\lambda x_i)\,,$$

with x_i denoting the individual data values.

a) The λ point with the maximum $\ln L$ defines the maximum-likelihood estimate $\hat{\lambda}$. Read off from figure 2.1: $\hat{\lambda} = 0.3$. The two λ points where $\ln L$ drops by 0.5 from the maximum value define a 68% confidence interval for λ; find (see figure 2.1) the interval: $[-0.8, 1.3]$. Express the result in the short-hand notation of equation (2.26) in the book:

$$\lambda = -0.3^{+1.0}_{-1.1}\,.$$

b) With the small statistics at hand it is not easy to judge on the compatibility of theory and data. It looks as if the theory is compatible, although other models, e.g. with a distribution $\sim 1-(\cos\theta)^2$, would also fit the data.

c) The bottom left of figure 2.1 shows the data points (the ten spikes) and in the bottom right the log-likelihood curve. The fact that all data points lie in the region of positive $\cos\theta$ values leads to a problem: The log-likelihood function rises continuously with increasing λ, so that no maximum can be found at finite λ values! The theory does not look consistent with the data. If λ is positive and *large* (however, λ is bounded to not be larger than unity in order to have a positive probability density for all values of $\cos\theta$), we would expect the data points to be clustered towards $\cos\theta = 1$ and not to be uniformly distributed. If λ is positive but *small* compared to unity, we would expect some values to lie on the negative side. So the experimenter or fitter should also question the data and check if everything is correct or if perhaps a bug is present.

Exercise 2.4: Unbinned fits 2

a) The log-likelihood function is given by

$$\ln L(\lambda) = \sum_{i=1}^{N} \ln(\lambda e^{-\lambda t_i}) = N\ln(\lambda) - \lambda T, \quad \text{with} \quad T = \sum_{i=1}^{N} t_i\,.$$

The maximum-likelihood estimate $\hat{\lambda}$ is the point at which $\ln L$ is maximal; find it by:

$$\frac{\partial \ln L}{\partial\lambda} = \frac{N}{\lambda} - T = 0 \qquad \Rightarrow \quad \hat{\lambda} = \frac{N}{T}\,.$$

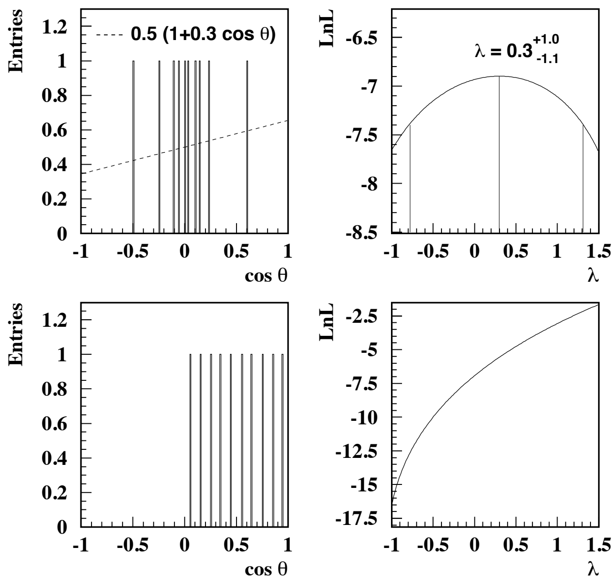

Figure 2.1 The solution of exercise 2.3. For details see the text.

Estimate the error of $\hat{\lambda}$ (see equations (2.22) and (2.23) in the book):

$$\hat{\sigma}_{\hat{\lambda}} = \left[-\frac{\partial^2 \ln L(\lambda)}{\partial \lambda^2} \bigg|_{\lambda=\hat{\lambda}} \right]^{-0.5} = \left[\frac{N}{\hat{\lambda}^2} \right]^{-0.5} = \frac{\hat{\lambda}}{\sqrt{N}} .$$

This estimate is based on approximating the likelihood function around its maximum by a Gaussian. This works only well for reasonably large N, e.g. $N = 1$ is too small, but $N = 20$ is already fine unless a very high precision is required. For small N follow the prescription (see equation (2.24) in the book and the discussion thereafter) and estimate a 68% confidence interval $[\lambda_{low}, \lambda_{up}]$ for λ by finding the two points λ_{low} and λ_{up} where $\ln L$ drops by 0.5 from its maximum value at $\hat{\lambda}$. Find these points e.g. by scanning or plotting $\ln L$ as a function of λ (for an example see the solution of exercise 2.7).

b) The detector efficiency is $e^{-\nu t}$. Write the proper probability density function f of the decays as

$$f(t) = \lambda' e^{-\lambda' t} \quad \text{with} \quad \lambda' = \lambda + \nu .$$

Determine results for λ' exactly in the same way as in task a) for λ and obtain at the end $\lambda = \lambda' - \nu$. Since ν is fixed, the uncertainty of $\hat{\lambda}$ is the same as the one of $\hat{\lambda}'$.

c) Obtain the probability density of the observed decays from a convolution of the theoretical rate with the resolution function:

$$\begin{aligned} f(t) &= \int_0^\infty \mathrm{d}t' \, \lambda e^{-\lambda t'} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-t')^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \lambda e^{\frac{\sigma^2\lambda^2}{2} - \lambda t} \int_0^\infty \mathrm{d}t' \, e^{-\frac{1}{2\sigma^2}\left(t' + (\lambda - \frac{t}{\sigma^2})\sigma^2\right)^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \lambda e^{\frac{\sigma^2\lambda^2}{2} - \lambda t} \int_{-t+\lambda\sigma^2}^\infty \mathrm{d}t' \, e^{-\frac{t'^2}{2\sigma^2}} . \end{aligned}$$

According to equations (2.3) and (2.4) in the book, the minimal standard deviation of the estimated parameter $\hat{\lambda}$ is given by

$$\sigma_{\hat{\lambda}} \geq I(\lambda)^{-0.5} , \quad \text{with the information} \quad I(\lambda) = N \int_{-1}^{1} \frac{1}{f} \left(\frac{\partial f}{\partial \lambda}\right)^2 \mathrm{d}t \, .$$

Calculate the derivative of the probability density function with respect to λ:

$$\frac{\partial f}{\partial \lambda} = \frac{f}{\lambda} + (\lambda\sigma^2 - t) f - \frac{1}{\sqrt{2\pi}} \lambda \sigma e^{-\frac{t^2}{2\sigma^2}} .$$

Determine numerically

$$\sigma_{\hat{\lambda}} = \left[\int_0^\infty \frac{1}{f} \left(\frac{\partial f}{\partial \lambda}\right)^2 \mathrm{d}t \right]^{-0.5} / \sqrt{N} \, .$$

The solution is displayed in figure 2.2, for a fixed value $\lambda = 1$ and setting $N = 1$. For $\sigma \ll 1$ find $\sigma_{\hat{\lambda}} = 1$, corresponding to task a). For increasing σ the uncertainty $\sigma_{\hat{\lambda}}$ is growing, reflecting the loss of information due to the detector resolution (smearing of events). For $\sigma > 2$ the uncertainty $\sigma_{\hat{\lambda}}$ slowly approaches σ. This is clear since the observed decay-rate distribution gradually evolves into a simple Gaussian with mean value λ and width σ.

Exercise 2.5: Correlated fit parameters

We follow here the *Bayesian* reasoning and interprete the likelihood function as a two-dimensional Gaussian probability density of the parameter vector $\mathbf{x}$. First calculate the inverse of the covariance matrix $\mathbf{C}$ and its determinant:

$$\mathbf{C}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} , \quad det(\mathbf{C}^{-1}) = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \, .$$

In the following it is very convenient to work with reduced variables:

$$\frac{x_1 - \lambda_1}{\sigma_1} \to x \quad \text{and} \quad \frac{x_2 - \lambda_2}{\sigma_2} \to y \, .$$

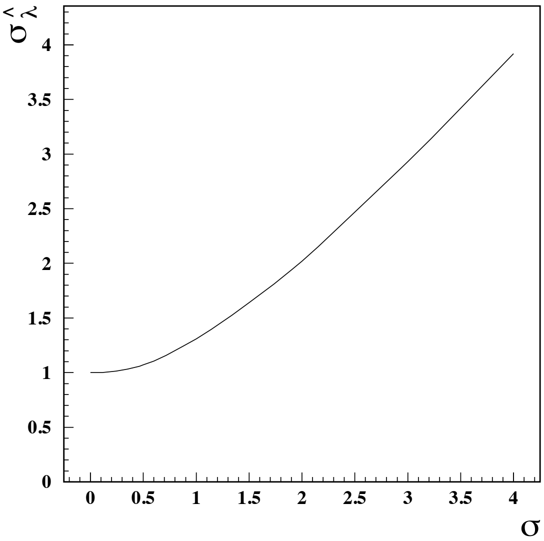

Figure 2.2 The solution of exercise 2.4. For details see the text.

The results obtained from here on can be easily transferred back at the end to the original variables. Determine the probability density $f(x,y) \equiv G(\mathbf{x})$:

$$f(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left[\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right].$$

For the first task calculate the probability density for x while y can have any value. For this integrate (marginalise) the density over y using the following integral formula:

$$\mathcal{I} = \int_{-\infty}^{\infty} \exp\left(\frac{-y^2}{2a^2} + by\right) \mathrm{d}y = \sqrt{2\pi}a \cdot \exp\left(\frac{a^2b^2}{2}\right).$$

The relations to our variables are

$$a^2 = 1-\rho^2, \quad b = \frac{\rho x}{1-\rho^2} \quad \Rightarrow \frac{a^2b^2}{2} = \frac{\rho^2 x^2}{2(1-\rho^2)}.$$

Using $\mathcal{I}$, calculate the marginal probability density of x (denoted as $\tilde{f}(x)$):

$$\begin{aligned}\tilde{f}(x) &= \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-x^2}{2(1-\rho^2)} + \frac{\rho^2x^2}{2(1-\rho^2)}\right] \\ &= \frac{1}{\sqrt{2\pi}} \exp\left[\frac{-x^2}{2}\right].\end{aligned}$$

This is a unit Gaussian distribution in x; thus the original variable x_1 follows a Gaussian distribution with expectation value λ_1 and standard deviation σ_1.

For the second task x_2 is fixed to the value λ_2'. This corresponds to fixing the transformed variable y to $y' = \frac{\lambda_2' - \lambda_2}{\sigma_2}$. Now renormalise the probability density for x. For this integrate f over x, keeping y fixed to y'. This corresponds exactly to the marginalisation step in the first task (but swapping x and y). Hence the renormalisation factor is

$$\sqrt{2\pi} \exp\left[\frac{y'^2}{2}\right] ,$$

and we find for the renormalised probability density of x:

$$f(x|_{y=y'}) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[\frac{-1}{2(1-\rho^2)}(x - \rho y')^2\right] .$$

This is a Gaussian distribution with expectation value $\rho y'$ and standard deviation $\sqrt{1-\rho^2}$. Thus the original variable x_1 follows a Gaussian distribution with expectation value $\lambda_1 + \frac{\sigma_1}{\sigma_2}\rho(\lambda_2' - \lambda_2)$ and standard deviation $\sigma_1\sqrt{1-\rho^2}$. This demonstrates the role of the correlation coefficient ρ: if x_2 is shifted (e.g. by an external constraint from another measurement) by one standard deviation, that is $\lambda_2' - \lambda_2 = \sigma_2$, then the expectation value of x_1 is shifted by ρ times its own standard deviation σ_1. Furthermore the fixing of x_2 reduces the uncertainty of x_1 by a factor $\sqrt{1-\rho^2}$.

Exercise 2.6: Evaluation of the information

According to equations (2.3) and (2.4) in the book, the minimal standard deviation of the estimated parameter $\hat{\lambda}$ is given by

$$\sigma_{\hat{\lambda}} \geq I(\lambda)^{-0.5} , \quad \text{with the information} \quad I(\lambda) = n \int_{-1}^{1} \frac{1}{f}\left(\frac{\partial f}{\partial \lambda}\right)^2 \mathrm{d}t .$$

Here n denotes the number of recorded decays, and the probability density function f is proportional to the rate $N_{B^0 \to J/\psi K_S^0}(t)$ or to $N_{\bar{B}^0 \to J/\psi K_S^0}(t)$.

Let us generalise the problem and introduce time-dependent rates N and $\bar{N}$ with arbitrary oscillation frequency ω:

$$B^0 : \quad N \propto e^{-t}\left[1 + \lambda \sin(\omega t)\right] , \tag{2.1}$$

$$\bar{B}^0 : \quad \bar{N} \propto e^{-t}\left[1 - \lambda \sin(\omega t)\right] . \tag{2.2}$$

The integrals of the rates are given by:

$$\begin{aligned}
\int_0^\infty N(t)\mathrm{d}t &= 1 + \lambda \int_0^\infty \frac{1}{2i}\left[e^{-t(1-i\omega)} - e^{-t(1+i\omega)}\mathrm{d}t\right] \\
&= 1 + \frac{\lambda\omega}{1+\omega^2} \\
&= 1 + \lambda\delta \quad \text{with} \quad \delta := \frac{\omega}{1+\omega^2} , \\
\int_0^\infty \bar{N}(t)\mathrm{d}t &= 1 - \lambda\delta .
\end{aligned}$$

Thus obtain properly normalised probability density functions:

$$
\begin{aligned}
N &= \frac{1}{1+\lambda\delta} e^{-t} \left[1 + \lambda sin(\omega t)\right], \\
\bar{N} &= \frac{1}{1-\lambda\delta} e^{-t} \left[1 - \lambda sin(\omega t)\right] .
\end{aligned}
$$

Determine the derivatives of these functions with respect to λ:

$$
\begin{aligned}
\frac{\partial N}{\partial \lambda} &= \frac{-\delta e^{-t}}{(1+\lambda\delta)^2} \left[1 + \lambda sin(\omega t)\right] + \frac{e^{-t}}{1+\lambda\delta} sin(\omega t) \\
&= \frac{e^{-t}}{(1+\lambda\delta)^2} \left[sin(\omega t) - \delta\right], \\
\frac{\partial \bar{N}}{\partial \lambda} &= \frac{-e^{-t}}{(1-\lambda\delta)^2} \left[sin(\omega t) - \delta\right].
\end{aligned}
$$

Therefore find:

$$
\begin{aligned}
\frac{1}{N}\left(\frac{\partial N}{\partial \lambda}\right)^2 &= \frac{e^{-t}}{(1+\lambda\delta)^2} \frac{[sin(\omega t) - \delta]^2}{(1+\lambda sin(\omega t)}, \\
\frac{1}{\bar{N}}\left(\frac{\partial \bar{N}}{\partial \lambda}\right)^2 &= \frac{e^{-t}}{(1-\lambda\delta)^2} \frac{[sin(\omega t) - \delta]^2}{(1-\lambda sin(\omega t)} .
\end{aligned}
$$

The estimate of the uncertainty on λ is based on

$$
\sigma_{\hat{\lambda}} = \left[\int_0^\infty \frac{1}{N}\left(\frac{\partial N}{\partial \lambda}\right)^2 \mathrm{d}t\right]^{-0.5} / \sqrt{n} .
$$

Performing the integration numerically, obtain for $n = 500$ $\lambda = 0.3$ and $\omega = 0.7$ the solutions for tasks a) and b):

$$
\begin{aligned}
\sigma_{\hat{\lambda}} &= 0.165 \quad \text{for} \quad B^0, \\
\sigma_{\hat{\lambda}} &= 0.099 \quad \text{for} \quad \bar{B}^0.
\end{aligned}
$$

The sensitivity is larger for $\bar{B}^0$ because of the minus sign in the rate (2.2); this causes the rate to be closer to zero at the first maximum of $sin(\omega t)$, compared to the B^0 case, and a small change of λ can lead to a rather large relative rate variation.

If we produce both B^0 and $\bar{B}^0$ and observe their time-dependent decays, we have not only to integrate the rates over the time but also to sum up over both processes to obtain properly normalised PDFs. Let us assume that both mesons have been produced by strong interactions with the same probability $N(t = 0) = \bar{N}(t = 0)$. Using (2.2) and (2.2), this leads to the following normalisation condition:

$$
\int_0^\infty (N(t) + \bar{N}(t))\mathrm{d}t = 2 \quad \Rightarrow \quad \text{total normalisation factor } = 1/2 .
$$

Using the normalised rates gives:

$$\left.\begin{array}{l}\frac{1}{N}\left(\frac{\partial N}{\partial\lambda}\right)^2\\ \frac{1}{\bar{N}}\left(\frac{\partial\bar{N}}{\partial\lambda}\right)^2\end{array}\right\} = \frac{e^{-t}(sin(\omega t))^2}{1\pm\lambda sin(\omega t)}\,.$$

Finally obtain the solution for task c) by numerical integration:

$$\sigma_{\hat{\lambda}} = \left[\int_0^\infty \frac{1}{N}\left(\frac{\partial N}{\partial\lambda}\right)^2 \mathrm{d}t + \int_0^\infty \frac{1}{\bar{N}}\left(\frac{\partial\bar{N}}{\partial\lambda}\right)^2 \mathrm{d}t\right]^{-0.5} /\sqrt{500} = 0.075\,.$$

The improved sensitivity is caused by the anticorrelation of the two rates: When λ changes, one rate increases, but the other decreases. This is similar to the fit of the slope of a straight line (assuming the constant term to be fixed): When we have one measurement point to the left and another one to the right of the coordinate origin (with the B^0 rates corresponding to one point and the $\bar{B}^0$ to the other point), this increases the lever arm of the measurement compared to the case of only one of the two measurements available.

Exercise 2.7: Unbinned maximum-likelihood versus least-squares estimate

a) The log-likelihood function is given by

$$\ln L(\lambda) = \sum_{i=1}^{N} \ln(\lambda e^{-\lambda t_i}) = N\ln(\lambda) - \lambda T \quad \text{with} \quad T = \sum_{i=1}^{N} t_i\,.$$

Using the solution of exercise 2.4, determine the maximum-likelihood estimate

$$\hat{\lambda} = \frac{N}{T} = 40/270.4 = 0.148\,.$$

Please note that for simplicity we have dropped all physical units, but since the decay times are given in minutes, the proper unit for $\hat{\lambda}$ is min^{-1}.
Estimate the error on $\hat{\lambda}$ by approximating the likelihood function around the maximum by a Gaussian (see again solution of exercise 2.4):

$$\hat{\sigma}_{\hat{\lambda}} = \frac{\hat{\lambda}}{N} = 0.023\,.$$

Figure 2.3 shows, in the top, the log-likelihood function as a function of λ in a wider (left) and in a narrow (right) region around the maximum of $\ln L$. In the wider region one can see an asymmetric curve around the maximum which deviates strongly from a parabola — the Gaussian approximation fails. However, in the region around the maximum where $\ln L$ drops by about one unit (right plot), the Gaussian approximation seems to be ok. In fact the positive and negative uncertainties (see equation (2.24) in the book), which are defined as the distances from $\hat{\lambda}$ to the two λ points at which $\ln L$ drops by 0.5, differ only marginally from the (above) value of 0.023 (0.024 and 0.022, respectively). In summary the result based on the maximum-likelihood estimate is

$$\lambda = 0.148^{+0.024}_{-0.022} \,.$$

Figure 2.3 shows, in the bottom, a binned comparison of the observed data and the theoretical expectation using the fitted value $\hat{\lambda} = 0.148$. The left (right) panel shows the distributions in linear (logarithmic) vertical scale. The theoretical ex-

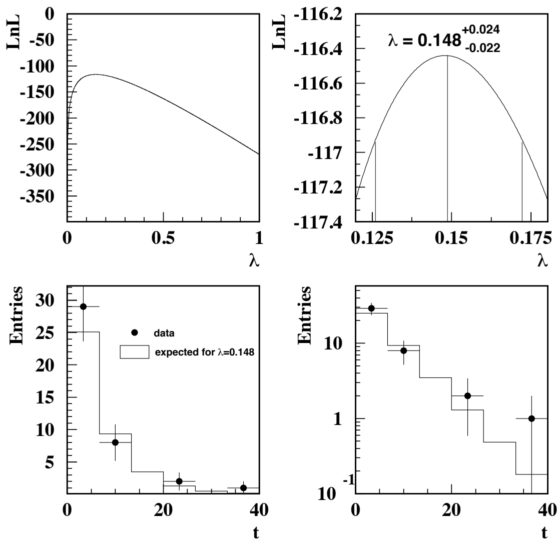

Figure 2.3 The solution of exercise 2.7 using the maximum-likelihood estimate. For details see the text.

pectation value ν in a bin is obtained from the integral

$$\nu = N \int_{t_{low}}^{t_{up}} \lambda e^{-\lambda t} \mathrm{d}t = N(e^{-t_{low}} - e^{-t_{up}}) \,.$$

Here t_{low} and t_{up} denote the lower and upper decay-time bin limits. The binning is chosen (arbitrarily) as $= 0.15^{-1}$ so it close to $\hat{\lambda}^{-1}$. One therefore expects a drop of the number of events from one bin to the next by a factor $\sim e^{-1}$. For a correct hypothesis the data are expected to fluctuate around the estimated theoretical expectation values approximately according to Poisson statistics. This seems to be the case for most of the bins; the most significant deviation is in the third bin where 3.4 entries are expected and none is recorded — the probability for such a Poisson fluctuation is about 3%, which is not significant. In summary the hypothesis of an exponential decay seems to be compatible with the data.

b) For the least-squares method we can use either the χ^2 defined by Pearson (see equation (2.90) in the book) or the one defined by Neyman (see equation (2.91) in the book). Choose the popular Neyman variant,

$$\chi^2 = \sum_{i=1}^{B} \frac{[n_i - \nu_i(\lambda)]^2}{n_i} ,$$

where B denotes the number of bins and n_i and ν_i the observed and expected number of decays in bin i. Figure 2.4 shows in the top the χ^2 function as a function of λ in a wider (left) and in a narrow (right) region around the minimum of χ^2. The

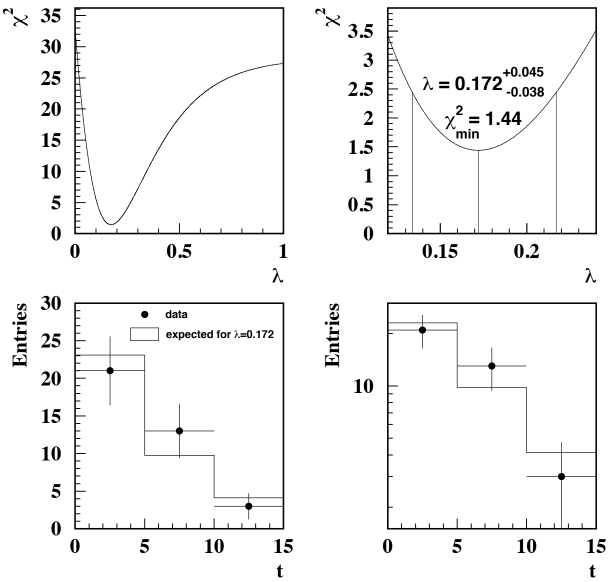

Figure 2.4 The solution of exercise 2.7 using the minimum χ^2 estimate. For details see the text.

minimal χ^2 value of $\chi^2_{min} = 1.44$ is obtained at $\hat{\lambda} = 0.172$. This λ value together with the two points where χ^2 increases by one unit defines the estimator value and its (estimated) uncertainty:

$$\lambda = 0.172^{+0.045}_{-0.038} \, .$$

The result is consistent with the one obtained in task a), however the χ^2 method yields a larger uncertainty on $\hat{\lambda}$. One reason for this is the information loss due to the binning; another is the restriction of the total fit interval to times smaller than 15, discarding three decays at larger decay times. The χ^2_{min}/ndf value of $1.44/2$ is indicating a good agreement of the exponential-decay hypothesis with the data. This is also demonstrated by the good agreement of the data and fit histogram shown in the lower plots of figure 2.4 in linear (left) and logarithmic (right) vertical scale.

Exercise 2.8: Best-fit parameters

The least-squares fits using polynomials of orders zero to three are shown in figure 2.5. The following information is also provided in the plots:

- the χ^2/ndf, denoting the minimum χ^2 value per number of degrees of freedom, obtained for the fitted parameter values;
- the fitted parameter values and their errors. The parameters A0, A1, A2 and A3 denote the coefficients of the terms of the polynomial of degres zero to three.

a) The fit with a constant function $y = a$ describes the data poorly. This is also indicated by the χ^2/ndf value of $90.8/5$.
b) The fit with a straight line $ax + b$ describes the data much better. However, the χ^2/ndf value of $12.8/4$ is still rather large. The probability to find for repeated experiments an equally large or larger χ^2 value is only $\sim 1.3\%$.
c) The fit with a parabola $ax^2 + bx + c$ describes the data somewhat better than the fit with the straight line. This is indicated by the χ^2/ndf of $7.9/3$. The probability to find for repeated experiments a larger χ^2 value is $\sim 5\%$, which is acceptable.
d) The fit with a cubic polynomial $ax^3 + bx^2 + cx + d$ does not really improve the description compared to the parabola. The χ^2 drops only by about one unit, but this is what is expected for an additional parameter which is not really needed (the parameter only helps to better describe the local fluctuations of the data).

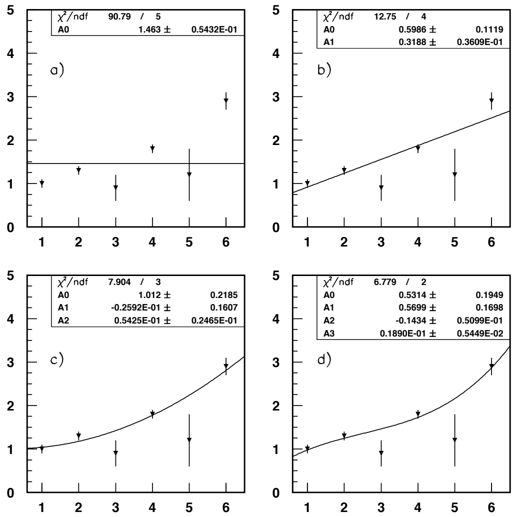

Figure 2.5 The solution of exercise 2.8. Shown are the input data points (triangles with error bars) and the fitted polynomials of degrees zero to three (solid curves). For further details see the text.