PHYS 509 - Problem Set 3

Tobias Faehndrich

October 24, 2025

Contents

1 Problem 1: Least Squares Fit of Exponential Decay (1000 Events)

File `exp1000.txt` contains a simulated list of decay times from an experiment.

(a) Perform a least squares fit to this distribution for the lifetime τ, to find the best estimate and also its uncertainty. Plot your fit results on your histogram of the data.

(b) Plot a scan of the χ^2 about the minimum over a range that shows at least the $\pm 3\sigma$ region with lines indicating the 1, 2 and 3σ levels on the χ^2 plot. Make a table with the symmetric and asymmetric confidence intervals for 1, 2 and 3σ limits.

Tutor's Hint:

- For least squares fitting of histogram data, use $\chi^2 = \sum_i \frac{(n_i - \mu_i)^2}{\sigma_i^2}$ where n_i is the observed count in bin i, and μ_i is the expected count.
- For Poisson statistics (counting), $\sigma_i^2 = n_i$ (variance equals mean).
- The exponential decay PDF is $p(t|\tau) = \frac{1}{\tau} e^{-t/\tau}$ for $t \geq 0$.
- Expected count in bin i with edges $[t_{i-1}, t_i]$: $\mu_i = N \int_{t_{i-1}}^{t_i} p(t|\tau) dt = N(e^{-t_{i-1}/\tau} - e^{-t_i/\tau})$.
- Minimize χ^2 to find best τ. Uncertainty from $\Delta\chi^2 = 1$ (1 parameter).
- For confidence intervals: $\Delta\chi^2 = 1, 4, 9$ for 1, 2, 3σ levels.

1.1 Solution

2 Problem 2: Maximum Likelihood Fit of Exponential Decay (1000 Events)

Use the same file as above, but perform a maximum likelihood fit for τ.

(a) Perform a likelihood fit to this distribution for the lifetime τ, to find the best estimate and also its uncertainty. Plot your fit result on a histogram of the data.

(b) Plot a scan of the likelihood L about the minimum over a range that shows at least the $\pm 3\sigma$ region with lines indicating the 1, 2 and 3σ levels on the plot. Make a table with the symmetric and asymmetric confidence intervals for 1, 2 and 3σ limits.

Tutor's Hint:

- For unbinned maximum likelihood, the likelihood function is $L(\tau) = \prod_{i=1}^{N} p(t_i|\tau)$ where t_i are the individual measurements.
- It's easier to work with log-likelihood: $\ln L = \sum_{i=1}^{N} \ln p(t_i|\tau) = -N \ln \tau - \frac{1}{\tau} \sum_{i=1}^{N} t_i$.
- Maximize $\ln L$ (or minimize $-\ln L$) to find best τ.

- The ML estimator for exponential lifetime: $\hat{\tau} = \bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_i$ (sample mean).
- Uncertainty from $\Delta(-\ln L) = 0.5$ for 1 parameter (equivalent to $\Delta\chi^2 = 1$).
- For confidence intervals: $\Delta(-\ln L) = 0.5, 2, 4.5$ for 1, 2, 3σ levels.

2.1 Solution

3 Problem 3: Fits with Small Sample (30 Events)

Do the same fits as above, but with the file `exp30.txt`, which only has 30 entries.

(a) Do the least squares fit on this file, with the same plots as above. Explain your histogram binning, how you determined it, and any other special treatment you had to make.

(b) Do the ML fit, again with the same plots as above.

Tutor's Hint:

- With only 30 events, histogram binning becomes critical. Too many bins $\rightarrow$ empty bins, too few $\rightarrow$ loss of information.
- Rule of thumb: $N_{\text{bins}} \approx \sqrt{N} \approx 5-6$ bins for $N = 30$.
- For Poisson statistics with small counts, use $\sigma_i^2 = n_i$ but be careful with empty bins ($n_i = 0$).
- Some options for handling empty bins: exclude from χ^2, use $\sigma_i^2 = 1$, or use expected count μ_i.
- ML fit doesn't require binning - works directly with unbinned data. Should be more reliable for small samples.
- Compare results: Which method gives tighter uncertainties? Which is more stable?

3.1 Solution

4 Problem 4: Convolution of Exponential with Gaussian

Derive an expression for the **convolution** of an exponential decay with lifetime τ, with a Gaussian resolution function. Assume the Gaussian is centered at 0, with variance σ^2. You should be able to express this function using an exponential and an erfc function.

Tutor's Hint:

- Convolution of two PDFs $f(t)$ and $g(t)$: $(f * g)(t) = \int_{-\infty}^{\infty} f(t')g(t-t')dt'$.
- Exponential: $f(t) = \frac{1}{\tau}e^{-t/\tau}$ for $t \geq 0$, zero otherwise.
- Gaussian: $g(t) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-t^2/(2\sigma^2)}$.

- The convolution integral: $h(t) = \int_0^\infty \frac{1}{\tau} e^{-t'/\tau} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-t')^2/(2\sigma^2)} dt'$.
- Trick: Complete the square in the exponent to get a Gaussian integral.
- Result involves $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-u^2} du$ (complementary error function).
- Expected form: $h(t) = \frac{A}{\tau} e^{a+bt} \cdot \mathrm{erfc}(c + dt)$ where you need to find A, a, b, c, d.

4.1 Solution

5 Problem 5: ML Fit with Gaussian Smearing

Download the file `exp_smeared.txt`, which contains a set of simulated data from an experiment measuring the lifetime of a particle, but with Gaussian smearing due to detector resolution. Perform a ML fit to the data to extract the lifetime τ and detector resolution σ, along with the full covariance matrix of these estimators. Use the Kolmogorov-Smirnov test to determine the goodness of fit for this model.

Plot your fit result on top of a histogram of your data.

Tutor's Hint:

- The PDF is now the convolution result from Problem 4: $h(t|\tau, \sigma)$.
- Two parameters to fit: τ and σ. Use unbinned ML.
- Log-likelihood: $\ln L(\tau, \sigma) = \sum_{i=1}^{N} \ln h(t_i|\tau, \sigma)$.
- Covariance matrix from inverse Hessian: $V = H^{-1}$ where $H_{jk} = -\frac{\partial^2 \ln L}{\partial\theta_j \partial\theta_k}$.
- Diagonal elements $V_{jj} = \sigma_j^2$ (variances), off-diagonal V_{jk} are covariances.
- Correlation: $\rho_{jk} = V_{jk}/(\sigma_j \sigma_k)$.
- Kolmogorov-Smirnov test: Compare empirical CDF with fitted CDF. $D = \max |F_{\text{data}}(t) - F_{\text{fit}}(t)|$.
- KS p-value tells you goodness of fit. $p > 0.05$ usually means acceptable fit.

5.1 Solution

6 Problem 6: ML Fit with Signal and Background

Download the file `exp_smear_bckgnd.txt`, which contains data from a signal which is a smeared exponential similar to the one above, but also with a background of random events with uniform probability over the experiment's region of operation, which is from $t = -3$ to $t = 10$ seconds.

The detector's time resolution is determined elsewhere to be a Gaussian $\mathcal{N}(0, \sigma^2)$ with $\sigma = 0.2$ s.

Perform a ML fit to the data to extract the lifetime τ, fraction of the sample that is signal, along with the full covariance matrix of these estimators.

Plot your fit result on a histogram of the data, and also on the same plot show the fitted background and (smeared) signal fits.

Tutor's Hint:

- Combined PDF: $p(t|\tau, f) = f \cdot h(t|\tau, \sigma) + (1 - f) \cdot p_{\text{bkg}}(t)$ where f is signal fraction.
- Background PDF: $p_{\text{bkg}}(t) = \frac{1}{13}$ for $t \in [-3, 10]$, zero otherwise (uniform).
- Signal PDF: $h(t|\tau, \sigma)$ from Problem 4, with $\sigma = 0.2$ s (fixed).
- Two parameters: τ and f. Use unbinned ML.
- For plotting: $N_{\text{sig}} = f \cdot N_{\text{total}}$, $N_{\text{bkg}} = (1 - f) \cdot N_{\text{total}}$.
- Signal component: $f \cdot h(t|\tau, \sigma) \cdot N_{\text{total}} \cdot \Delta t$ (for histogram).
- Background component: $(1 - f) \cdot p_{\text{bkg}}(t) \cdot N_{\text{total}} \cdot \Delta t$.
- Total fit: sum of signal and background components.

6.1 Solution

7 Problem 7: Gaussian Signal with Flat Background

File `gauss_bckgnd.txt` contains the counts from an experiment with a Gaussian signal of unknown mean and sigma, along with a flat background. There are 50 bins, with edges listed in the file.

(a) Do a least squares fit on the data for the amplitude, mean, and sigma of the signal (S, μ_0, σ_0), and the background level μ_B. The **significance** of your signal is defined as S/σ_S, where σ_S^2 is the variance of S returned by your fit.

(b) Plot the histogram of the data along with the total fit result, the signal component and the background.

(c) Choose the bins that you consider contain the bulk of the signal based upon the fit. How many background events B do you estimate are in this region, and what would you estimate is the statistical uncertainty σ_B on this number. Explain your reasoning.

(d) Given the number of signal events S you obtained from your fit, what would the uncertainty σ_S on this be if the data were background free? Compare this to what you actually obtained in the fit.

(e) Considering all the bins you chose in your signal region, what is your estimate of the uncertainty on the size of statistical fluctuations in this region, expressed in terms of S and B? What is your expected signal significance, then, and how does it compare with your fitted value?

Tutor's Hint:

- Model for bin counts: $\mu_i = S \cdot g_i(\mu_0, \sigma_0) + \mu_B$ where $g_i = \int_{x_{i-1}}^{x_i} \mathcal{N}(\mu_0, \sigma_0^2)dx$.
- Four parameters: S (total signal events), μ_0 (signal mean), σ_0 (signal width), μ_B (background per bin).
- $\chi^2 = \sum_i \frac{(n_i - \mu_i)^2}{n_i}$ where n_i is observed count in bin i.
- Signal significance from fit: S/σ_S where σ_S comes from covariance matrix.
- For signal region (e.g., $[\mu_0 - 3\sigma_0, \mu_0 + 3\sigma_0]$): count bins in this range.
- Background in region: $B = \mu_B \cdot N_{\text{bins}}$, uncertainty $\sigma_B = \sqrt{\mu_B \cdot N_{\text{bins}}}$ (Poisson).
- If background-free: $\sigma_S = \sqrt{S}$ (Poisson).
- With background: total counts in region $N = S + B$, so $\sigma_N = \sqrt{N} = \sqrt{S + B}$.
- Signal significance: $\frac{S}{\sqrt{S+B}}$ (approximation when background is known).

7.1 Solution

8 Problem 8: Measuring Gravitational Acceleration

In an undergraduate lab, students perform an experiment to measure g, by dropping a long clear plastic plate with a dark line marked every 10 cm over a 1 m length. The lines are sensed by a photosensor, and the time t that each passes the sensor is recorded. The clock of the timer starts when the first line is sensed, so there are 11 measurements, starting with $t = 0$ for the line at 0 cm. The plate is held by hand some point above the photosensor before releasing.

The file `gmeasure.txt` contains the times for the 11 markings, each with an uncertainty of 5 ms.

(a) What is the equation for the position of the plate $z(t)$, where z = vertical position of the 0 cm line marking wrt the photosensor position? (For uniformity, let's take z increasing the downward direction.)

(b) Since your experimentally measured variable is t, invert this so t is the dependent variable and then fit for g. Plot the data, with errorbars, along with your fit result.

(c) Instead, treat z as the dependent variable, and fit for g using the original equation you had. You will need to convert the uncertainty on t to an approximate uncertainty on z at each point that will change as your fit parameters change. Plot the data, with errorbars, along with your fit result. Compare with the previous results.

Tutor's Hint:

- Free fall from rest at initial position z_0 above sensor: $z(t) = z_0 + \frac{1}{2}gt^2$.
- At $t = 0$, the 0 cm line is at the sensor, so $z(0) = 0 \Rightarrow z_0 = 0$.
- Actually, the plate was held above and released. When does the 0 cm line reach the sensor?

- More carefully: If released at height h above sensor at t_{release}, then at sensor crossing: $0 = -h + \frac{1}{2}g(t_0)^2$.
- Redefine clock: $t = 0$ when 0 cm line passes sensor. Then for subsequent lines: $z_i = \frac{1}{2}gt_i^2 + v_0 t_i$ where v_0 is velocity at $t = 0$.
- Simplification: If we take z as position of each line wrt 0 cm line, then $z_i = i \cdot 10$ cm for line $i = 0, 1, 2, \ldots, 10$.
- For part (b): $t = \sqrt{\frac{2z}{g}}$ (if starting from rest). Fit t vs z data.
- For part (c): $z = \frac{1}{2}gt^2$. Error propagation: $\sigma_z = \left|\frac{dz}{dt}\right| \sigma_t = gt \cdot \sigma_t$.
- Note: uncertainty on z depends on g (the parameter you're fitting!). Iterate: fit, recalculate errors, refit.

8.1 Solution