Chapter 7

The Method of Least Squares

Exercise 7.1: Galileo's studies of motion included experiments with a ball and an inclined ramp. The ball's trajectory is made horizontal before it falls over the edge, as shown in Fig. 7.1. The horizontal distance d from the edge to the point of impact is measured for different values of the initial height of the ball h. Five data points obtained by Galileo in 1608 are shown in Table 7.1.[1]

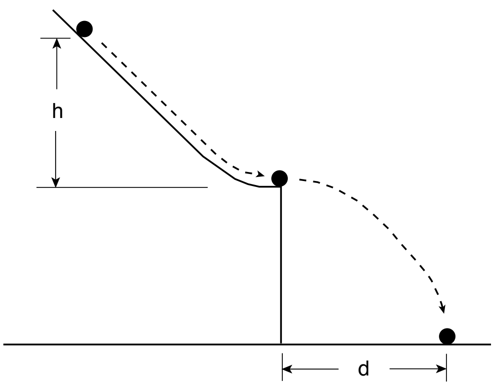

Figure 7.1: The configuration of the ball and ramp experiment performed by Galileo.

Table 7.1: Galileo's data on horizontal distance before impact d for five values of the starting height h. The units are punti (points); one punto is slightly less than 1 mm.

h	d
1000	1500
828	1340
800	1328
600	1172
300	800

Assume the heights h are known with negligible error, and that the horizontal distances d can be regarded as independent Gaussian random variables with standard deviations of $\sigma = 15$ punti.

[1] See Stillman Drake and James Maclachlan, Galileo's discovery of the parabolic trajectory, *Scientific American* **232** (March 1975) 102; Stillman Drake, *Galileo at Work*, University of Chicago Press, Chicago (1978).

(It is not actually known how Galileo estimated the measurement uncertainties, but 1–2% is plausible.) In addition, we know that if $h = 0$, then the horizontal distance d will be zero, i.e. if the ball is started at the very edge of the ramp, it will fall straight down to the floor.

(a) Consider relations between h and d of the form

$$d = \alpha h \tag{7.1}$$

and

$$d = \alpha h + \beta h^2. \tag{7.2}$$

Find the least-squares estimators for the parameters α and β. What are the values of the minimized χ^2 and the P-values for the two hypotheses?

(b) Assume a relation of the form

$$d = \alpha h^{\beta}\,. \tag{7.3}$$

Write a computer program to perform a least squares fit of α and β. Note that this is a nonlinear function of the parameters and must be solved numerically. A solution using the MINUIT minimization routines from the CERN library is given in `fit_galileo.f, fcn_galileo.f`.

(c) Galileo regarded the motion as the superposition of horizontal and vertical components, where the horizontal motion is of constant speed, and the vertical speed is zero at the lower edge of the ramp, but then increases in direct proportion to the time. Show that this leads to a relation of the form

$$d = \alpha\sqrt{h}\,. \tag{7.4}$$

Find the least squares estimate for α and the value of the minimized χ^2. What is the P-value for this hypothesis?

Exercise 7.2: Consider a least-squares fit to a histogram with y_i entries in bins $i = 1, \ldots, N$, and predicted values

$$\lambda_i(\boldsymbol{\theta}) = n \int_{x_i^{\min}}^{x_i^{\max}} f(x;\boldsymbol{\theta})dx\,, \tag{7.5}$$

where $f(x;\boldsymbol{\theta})$ depends on unknown parameters $\boldsymbol{\theta}$. Suppose that one replaces the total number of entries n by a parameter ν, and that this is adjusted simultaneously with the other parameters when minimizing

$$\chi^2(\boldsymbol{\theta},\nu) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\boldsymbol{\theta},\nu))^2}{\sigma_i^2}\,. \tag{7.6}$$

(a) Show that taking $\sigma_i^2 = \lambda_i$ leads to the estimator

$$\hat{\nu}_{\mathrm{LS}} = n + \frac{\chi^2_{\mathrm{min}}}{2} \tag{7.7}$$

for the total number of entries.

(b) Show that using $\sigma_i^2 = y_i$ (modified least squares) gives the estimator

$$\hat{\nu}_{\mathrm{MLS}} = n - \chi^2_{\mathrm{min}} \, . \tag{7.8}$$

Exercise 7.3: Consider an LS fit to a histogram with y_i entries in bins $i = 1, \ldots, N$, with predicted values $\lambda_i(\boldsymbol{\theta})$. Suppose the total number of entries n is treated as a constant, so that the y_i are multinomially distributed.

(a) What is the covariance matrix $V_{ij} = \mathrm{cov}[y_i, y_j]$? Why does the inverse of this matrix not exist?

(b) Consider the fit using only the first $N-1$ bins. Find the inverse covariance matrix, and show that this is equivalent to fitting to all N bins but without consideration of the correlations.

Exercise 7.4: Suppose that a data sample of size n has resulted in measurements of N quantities $y_1, \ldots, y_N$, which are to be used in a least-squares fit of some unknown parameters. If the measurements are correlated, one requires the inverse covariance matrix V^{-1} in order to construct the χ^2. Often this is obtained by first determining the matrix of correlation coefficients, $\rho_{ij} = V_{ij}/(\sigma_i \sigma_j)$, e.g. by means of a Monte Carlo calculation.

(a) Recall that for efficient estimators, the inverse covariance matrix is proportional to the sample size n. Show that if this is the case, then the matrix of correlation coefficients is independent of the sample size.

(b) Show that the inverse covariance matrix is given by

$$(V^{-1})_{ij} = \frac{(\rho^{-1})_{ij}}{\sigma_i \sigma_j}. \tag{7.9}$$

(Start with the identity

$$\begin{aligned} \delta_{ij} &= \sum_k (V^{-1})_{ik} V_{kj} \\ &= \sum_k (V^{-1})_{ik} \rho_{kj} \sigma_k \sigma_j . \end{aligned} \tag{7.10}$$

Multiply both sides of (7.10) by ρ^{-1} and sum over the appropriate indices to obtain (7.9).)

Exercise 7.5: Consider two partially overlapping samples of a random variable x, with n and m observations, c of which are common to both. Suppose the variance of x $V[x] = \sigma^2$ is known. Consider the sample means

$$y_1 = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{7.11}$$

and

$$y_2 = \frac{1}{m} \sum_{i=1}^{m} x_i \,. \tag{7.12}$$

(a) Show that the covariance is

$$\mathrm{cov}[y_1, y_2] = \frac{c\sigma^2}{nm} \,. \tag{7.13}$$

(b) Using the results of Section 7.6, find the weighted average of y_1 and y_2 and its variance (or standard deviation).

Exercise 7.6: The astronomer Claudius Ptolemy performed experiments on the refraction of light using a circular copper disk submerged to its center in water, as illustrated in Fig. 7.2. Angles of refraction θ_r for 8 values of the angle of incidence θ_i obtained by Ptolemy around 140 A.D. are shown in Table 7.2.[2]

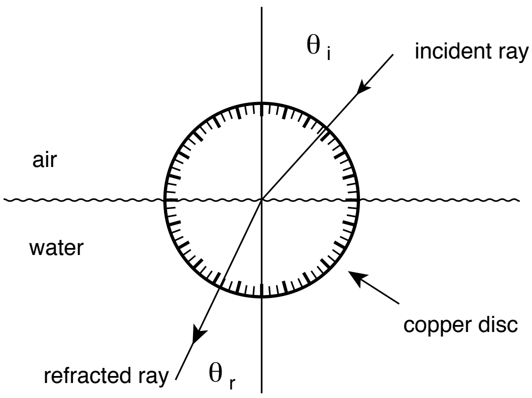

Figure 7.2: The apparatus used by Ptolemy to investigate the refraction of light.

Table 7.2: Angles of incidence and refraction (in degrees).

θ_i	θ_r
10	8
20	$15\frac{1}{2}$
30	$22\frac{1}{2}$
40	29
50	35
60	$40\frac{1}{2}$
70	$45\frac{1}{2}$
80	50

For purposes of this exercise we will take the angles of incidence to be known with negligible error and treat the angles of reflection as independent Gaussian variables with standard

[2]From Olaf Pedersen and Mogens Pihl, *Early Physics and Astronomy: A Historical Introduction*, MacDonald and Janes, London, 1974.

deviations of $\sigma = \frac{1}{2}^\circ$. (This is a reasonable guess given that the angles are reported to the nearest half degree. Note that we can absorb an error in θ_i into an effective error in θ_r.)

(a) The correct law of refraction was not discovered until the 17th century. Until then, a commonly used hypothesis was

$$\theta_\mathrm{r} = \alpha\theta_\mathrm{i}, \tag{7.14}$$

although it is reported that Ptolemy preferred the formula

$$\theta_\mathrm{r} = \alpha\theta_\mathrm{i} - \beta\theta_\mathrm{i}^2. \tag{7.15}$$

Find the LS estimates of the parameters for both hypotheses and determine the minimized χ^2. Comment on the goodness-of-fit for both hypotheses. Is it plausible that all of the numbers are based on actual measurements?[3]

(b) The law of refraction discovered by Snell in 1621 is

$$\theta_\mathrm{r} = \sin^{-1}\left(\frac{\sin\theta_\mathrm{i}}{r}\right), \tag{7.16}$$

where $r = n_\mathrm{r}/n_\mathrm{i}$ is the ratio of indices of refraction of the two media. Determine the LS estimate for r and find value of the minimized χ^2. Comment on the validity of the Gaussian assumption for θ_r with $\sigma = \frac{1}{2}^\circ$.

Exercise 7.7: Consider again the problem of Exercise 6.5: N independent Poisson variables $\mathbf{n} = (n_1, \ldots, n_N)$ have mean values $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N)$, where the means are related to a controlled variable x by a relation of the form

$$\nu(x) = \theta a(x). \tag{7.17}$$

(a) Consider first the LS method, where the denominators in the χ^2 use the variances $\sigma_i^2 = \nu_i$. Show that the LS estimator for θ is given by

$$\hat{\theta} = \left(\frac{\sum_{i=1}^N \frac{n_i^2}{a(x_i)}}{\sum_{i=1}^N a(x_i)}\right)^{1/2}. \tag{7.18}$$

By expanding $\hat{\theta}(\mathbf{n})$ in a Taylor series to second order about $\boldsymbol{\nu}$ and computing the expectation value, show that the bias of (7.18) is given by

$$b = \frac{N-1}{2\sum_{i=1}^N a(x_i)} + O(E[(n_i - \nu_i)^3]). \tag{7.19}$$

(Use $\mathrm{cov}[n_i, n_j] = \delta_{ij}\nu_j$ for the covariance of independent Poisson variables.)

[3] See R. Feynman, R. Leighton and M. Sands, *The Feynman Lectures on Physics*, Vol. I, Addison-Wesley, Menlo Park, 1963, Section 26-2.

(b) Repeat (a) using the modified LS method, where the χ^2 uses variances based on the observed values: $\sigma_i^2 = n_i$. Show that the MLS estimator for θ is given by

$$\hat{\theta} = \frac{\sum_{i=1}^{N} a(x_i)}{\sum_{i=1}^{N} \frac{a(x_i)^2}{n_i}}, \tag{7.20}$$

and that its bias is given by

$$b = -\frac{N-1}{\sum_{i=1}^{N} a(x_i)} + O(E[(n_i - \nu_i)^3]). \tag{7.21}$$

Compare the biases from (a) and (b) to the results of Exercise 7.2.

(c) Estimate the variance $\hat{\theta}$ for both the LS and MLS cases using error propagation.

Note that since it was shown in Exercise 6.5 that the ML estimator for θ is both unbiased and has minimum variance, the LS and MLS estimators are not preferred here. For sufficiently large data samples, however, the three methods are very similar; cf. Exercise 7.8.

Exercise 7.8: Consider again Perrin's data on the number of mastic particles as a function of height (Exercise 6.5). Determine the LS estimates for Boltzmann's constant k (and equivalently Avogadro's number $N_{\mathrm{A}} = R/k$) and the coefficient ν_0 by minimizing

$$\chi^2(k, \nu_0) = \sum_{i=1}^{N} \frac{(n_i - \nu_i(k, \nu_0))^2}{\sigma_i^2}. \tag{7.22}$$

(a) Take the standard deviation σ_i of n_i to be $\sqrt{\nu_i}$ (the usual method of least squares).

(b) Take σ_i to be $\sqrt{n_i}$ (the modified method of least squares).

Compare the results from (a) and (b) to the estimates obtained by maximum likelihood in Exercise 6.5.