1 Introduction

In this project you will be given a theoretical prediction for the outcome of an experiment and a set of actual data. The prediction is not completely specified, however, but rather contains free parameters. The goal is to find the values of the parameters such that the prediction gives the best agreement with the data using the method of least squares. In addition you will evaluate the level of agreement between data and hypothesis.

The experiment we will consider was performed by Galileo using a ball and an inclined ramp as shown in Fig. 1. The ball starts at a height h above the edge of the ramp, and its trajectory is forced to be horizontal before it falls over the edge. The horizontal distance d from the edge to the point of impact is measured for different values of h. Five data points obtained by Galileo in 1608 are shown in Table 1.[1]

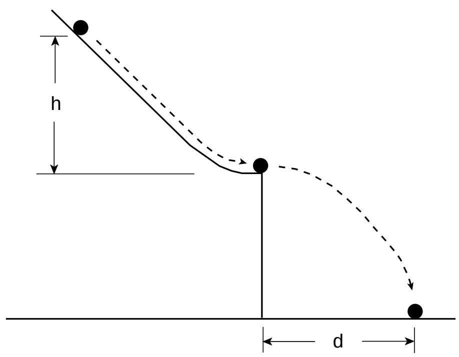

Figure 1: The configuration Galileo's ball and ramp experiment.

Table 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

We will assume the heights h are known with negligible error, and that the horizontal distances d have uncertainties of $\sigma = 15$ punti. (It is not actually known what the measurement uncertainties were, 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.

[1] 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).

We could proceed by applying the known laws of mechanics to the system and deriving the relationship between d and h. For purposes of this exercise, however, we will pretend we don't yet know Newton's laws (as was the case for Galileo), and we will simply try different hypotheses and compare their predictions with the data.

2 The assignment

We will consider three hypotheses for the functional relationship between d and h: a linear relation with a single free parameter α,

$$d = \alpha h \ , \tag{1}$$

a quadratic relation with two parameters, α and β,

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

and something which is a nonlinear function of the parameters,

$$d = \alpha h^{\beta} \ . \tag{3}$$

The goal of the project is to investigate these three hypotheses (or to think up new ones), to determine which gives the best level of agreement with the data, and to determine the best values of the parameters. Fitting the parameters should be done with the method of least squares, described below. This will involve finding the parameter values which minimize a certain function, and you should investigate different ways of doing this.

You should produce plots of the fitted curves along with the data points and their errors, and you should quantify the level of agreement between the data and each hypothesis.

3 The method of least squares

Once the functional form of the prediction for $d(h)$ has been specified, we need to determine the values of the parameters such that we achieve the best agreement with the data. This can be done with the method of least squares. Consider one of the predictions involving two parameters, α and β. The idea is to construct the quantity (called chi-square),

$$\chi^2(\alpha, \beta) = \sum_{i=1}^{N} \frac{(d_{i,\text{meas}} - d(h_i; \alpha, \beta))^2}{\sigma_i^2} \ , \tag{4}$$

where the $d_{i,\text{meas}}$ is the ith measured value for which the height was h_i, $d(h_i; \alpha, \beta)$ is the corresponding predicted value for d using the parameter values α and β, and σ_i^2 is the 'uncertainty' squared of the ith measurement. (More precisely, the measured values are treated as random variables, and σ_i is the standard deviation of the ith measurement.)

In the method least squares, you choose the parameter values such that the χ^2 is a minimum. That is, the sum of squares of the differences between the predicted and measured values should be made a minimum.

The value of χ^2 at its minimum is a measure of the 'goodness-of-fit'. If the hypothesis is correct, then you expect a contribution to the minimized χ^2 of about 1 per data point, since the deviation between the data and the prediction should be on the order of the uncertainty σ. In fact one can show the expected value of the minimized χ^2 is equal to $N - m$ (called the *number of degrees of freedom*), where N is the number of data points and m is the number of fitted parameters.

4 Numerical minimization

Finding the parameter values which minimize $\chi^2(\alpha, \beta)$ can in general be a challenging task. The hypotheses (1) and (2) are linear functions of the parameters, and for these you can simply set the derivatives of χ^2 with respect to the parameters equal to zero and solve.

For equation (3), however, you will not be able to find an analytic solution, and you will have to find the minimum numerically. There are many techniques you can try. A simple possibility is brute force: guess different values for the parameters, e.g. on a grid of points covering a region of the parameter space, compute the χ^2 for each point, and try to find the point which give the smallest possible χ^2.

There are of course more systematic algorithms which you can implement on the computer. For example, start with some guess for the parameters. Hold α constant and determine the value of β which minimizes the χ^2. This can be done by finding the solution to $\partial\chi^2/\partial\beta = 0$ using, say, the Newton–Raphson method. Then hold β constant at this new value, and determine the value of α which minimizes the χ^2. Repeat the procedure as many times as necessary until the change in the parameters from one iteration to the next drops below some threshold, or alternatively, until the change in the minimized χ^2 between iterations falls below a specified value. It will help to make a plot of the successive values of α and β and successive values of the minimized χ^2.

5 Teamwork

You will need to divide up the work so that each member has specific tasks to accomplish and the team achieves its goals to the fullest extent possible. Below are some suggestions on possible ways of separating the project into individual tasks.

In Java you will probably define a class called, say, `RampData`, whose data members contain the measured values of d and h. The class could contain a member function to read the data from a file. A `Hypothesis` class can contain a (static) method to calculate the predicted distance $d(h; \alpha, \beta)$.

Another class called, say, `MyStatUtils` could contain a method called `Chi2` that takes as input the data from `RampData` and computes the χ^2 statistic defined by equation (4). A method of the `MyStatUtils` class could implement some strategy for finding the values of the parameters that minimize the χ^2. Alternatively, you could create a separate `Minimiser` class. In either case it should be constructed in a modular way so that one may start with a fairly simple but slow method, and then simply replace it at a later stage with more refined one.

You will need a main method that creates the various objects and executes the methods in the desired way. The classes should contain methods that allow you to access the data and parameter values so that you can pass them to a plotting routine.

6 References

Numerical minimization and the method of least squares are described in:

S. Brandt, *Statistical and Computational Methods in Data Analysis*, Springer, New York (1997);

L. Lyons, *Statistics for Nuclear and Particle Physicists*, CUP (1986);

R.J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, John Wiley, Chichester (1989).

7 Appendix

Galileo's solution to the problem is shown in Fig. 2 (from the Biblioteca Nazionale Centrale di Firenze, `http://www.bncf.firenze.sbn.it/`).

G. Cowan, 19 November, 2001

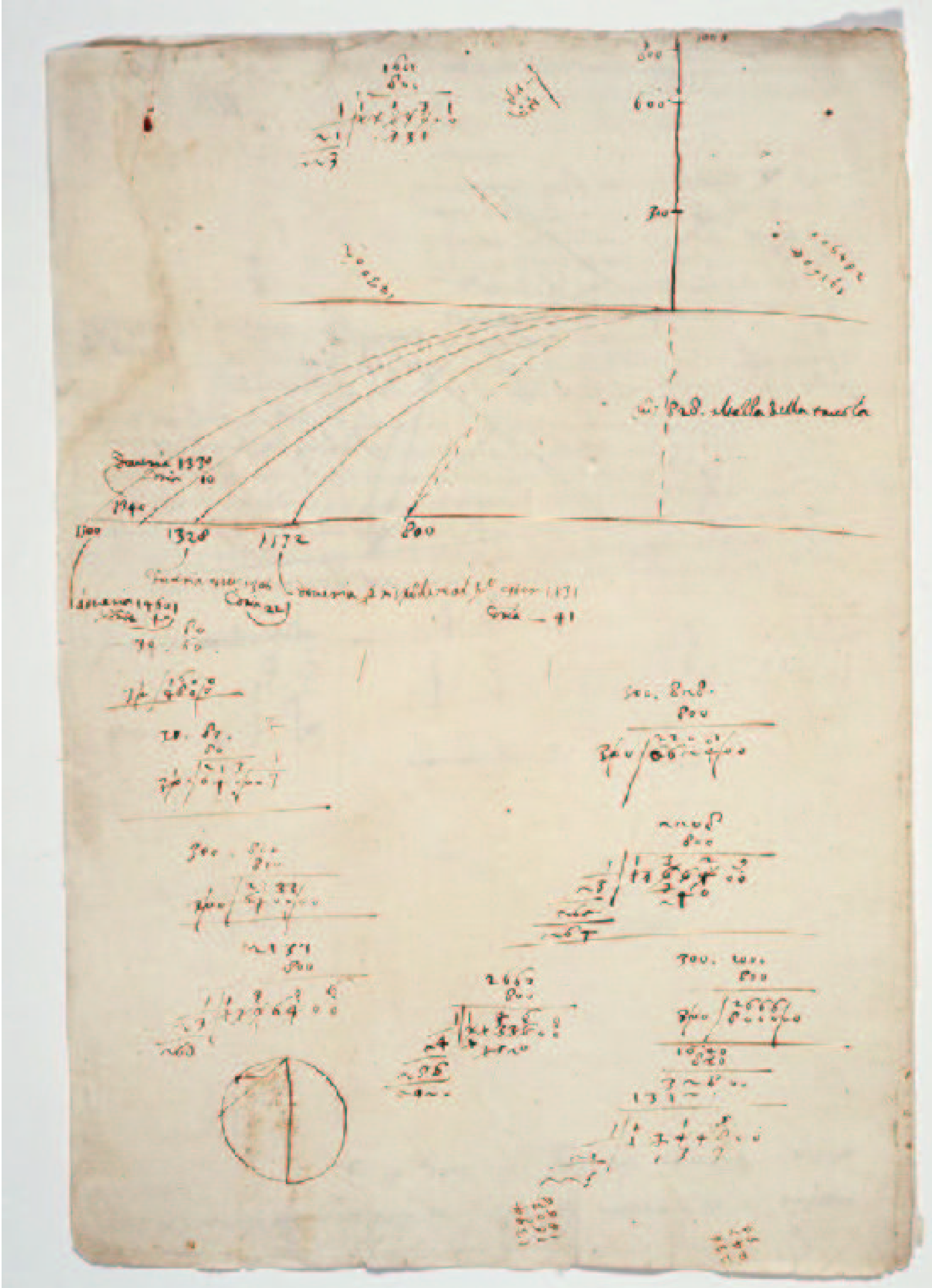

Figure 2: Manuscript *f*.116 from Galileo's notebooks showing the ball and ramp experiment.