Physics through Computational Thinking

Data Analysis

Auditya Sharma and Ambar Jain
Dept. of Physics, IISER Bhopal

Outline

In this lecture you will

1. learn to read in data and store it into suitable arrays.
2. learn about error bars: how they are computed and how they can throw light.

Reading and Plotting a Data file

Create a function and generate some data.

```
f[x_] = Exp[-x^2];

f[2];
```

```
f[2.0];

mytable = Table[f[n], {n, 1, 10}];
mytable = Table[f[n], {n, 1, 10}] // N;

ListPlot[mytable];

ListLogPlot[mytable];

Export["~/Desktop/Gaussiandata.dat", mytable];

Import["~/Desktop/Gaussiandata.dat", "List"];
```

Linear Combination of Random Variables

Consider the sum of N random variables:

$$S = x_1 + x_2 + ... + x_N \tag{1}$$

Let us say that all the x_i are independent identically distributed random variables drawn from the same distribution $f_{\mu,\sigma}(x)$ with mean μ and standard deviation σ. We have the obvious result that the mean of the sum of a bunch of random variables is equal to the sum of the means of the individual random variables:

$$\mu_S = E(S) = E(x_1) + E(x_2) + ... + E(x_N) = N\,\mu \tag{2}$$

Also, a less obvious similar result holds for the variances (when the random variables are independent):

$$\sigma_S{}^2 = \text{var}(S) = \text{var}(x_1) + \text{var}(x_2) + ... + \text{var}(x_N) = N\,\sigma^2. \tag{3}$$

It is also true that if you take a random variable and multiply it by a scalar mean of the resulting random variable is also simply multiplied by a scalar:

$$E(\alpha\, x) = \alpha\, E(x) \tag{4}$$

but the variance has to be multiplied by the square of the scalar:

$$\text{var}(\alpha\, x) = \alpha^2\, \text{var}(x). \tag{5}$$

Therefore combining the above results if we consider the random variable which is called the sample mean:

$$X = \frac{x_1 + x_2 + ... + x_N}{N} \tag{6}$$

we have the two important results:

$$\mu_X = E(X) = \frac{E(x_1) + E(x_2) + ... + E(x_N)}{N} = \mu \tag{7}$$

and crucially:

$$\sigma_X{}^2 = \text{var}(X) = \frac{\text{var}(x_1) + \text{var}(x_2) + ... + \text{var}(x_N)}{N^2} = \frac{\sigma^2}{N}. \tag{8}$$

Often times when we perform experiments we acquire data whose distribution is unknown, so the precise μ and σ are not available. However the sample mean X turns out to be an excellent estimate of the true mean of the distribution. The error in our estimate of the actual mean of the quantity is contained in the standard deviation σ_X, which we have seen is given by $\frac{\sigma^2}{N}$. Unfortunately the σ of the distribution itself is an unknown, and with some (beautiful, not very difficult) arguments, one can show that the sample variance:

$$s^2 = \frac{(x_1 - X)^2 + (x_2 - X)^2 + ... + (x_N - X)^2}{N} \tag{9}$$

can yield an excellent estimate of the variance of the actual distribution with the aid of the simple relation:

$$\sigma^2 = \frac{N}{N-1} s^2. \tag{10}$$

The overall consequence of all this is that given a sample of N data points x_i, the mean along with error bars is given by the compact relation:

$$X \pm \frac{s}{\sqrt{(N-1)}}. \tag{11}$$

Galileo's *Mathematica* Experiment!

Let us generate some data from a thought experiment, which we will then analyze by the basic statistical tools we have described above. This is a recurring set of techniques that would come in handy for analyzing a variety of experimental/simulation data.

```
g = 9.8;

data = Prepend[Table[ Table[Round[1/2 g i^2 (1 + 0.2 RandomReal[]), 0.01], {i, 1, 10}], {10}], Range[1, 10]] // Transpose;

data // TableForm;

Export["~/Desktop/mydata.csv", data];

data1 = Import["~/Desktop/mydata.csv"];

set = data[[1, 2 ;;]];


Mean[set];

StandardDeviation[set] / Sqrt[Length[set] - 1];

meandata = Table[
 set = data[[n, 2 ;;]];
 {n, Mean[set],  StandardDeviation[set] / Sqrt[Length[set] - 1]}
 , {n, 1, 10}
 ];

TableForm[meandata];

Needs["ErrorBarPlots`"]

ErrorListPlot[meandata[[ ;; , 2 ;; 3]], Joined → True];
```