Error analysis for IPhO contestants

Heikki Mäntysaari

University of Jyväskylä, Department of Physics

Abstract

In the experimental part of IPhO (and generally when performing measurements) you have to estimate the accuracy of your results. In this lecture I go trough the most important aspects of error analysis.

1 Introduction

A rule of thumb: a result without an estimate of its accuracy is worthless. This holds also in the experimental part of IPhO (if it is not stated in the problem that the error estimation is **not** required). Example: the *fine structure constant* (important in e.g. quantum mechanics and particle physics) is measured to be

$$\alpha = (7.297\,352\,537 \pm 0.000\,000\,005) \times 10^{-3} \tag{1}$$

Look at that accuracy! Uncertainty is only 0.37 parts per billion!

On the other hand if the densities of the samples are measured to be $\rho_1 = 1.1 \pm 0.5\ \mathrm{kg/m^3}$ and $\rho_2 = 1.5 \pm 0.4\ \mathrm{kg/m^3}$, can one say whether or not these samples consist of same/different material? What if the error bars were not given?

In the following chapters we shall go trough quickly the most important topics in error analysis. There are also a few excercises for you to do.

Updated June 13, 2012
Electronic address: `heikki.mantysaari@jyu.fi`

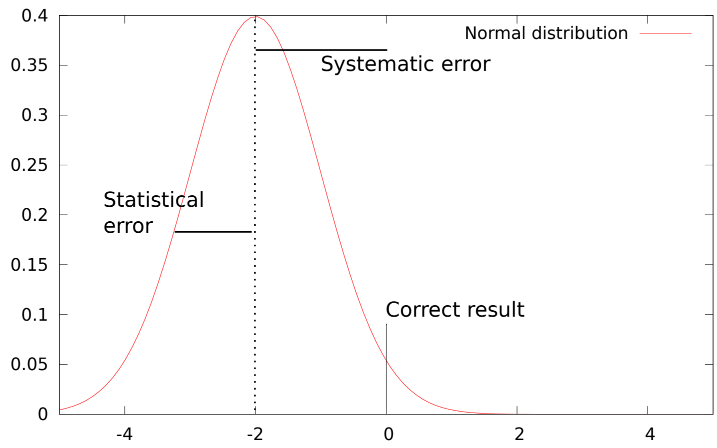

Figure 1: Distribution of measurements if the correct result is $x = 0$ and systematic error is -2 units.

Terminology

- If result is $a \pm b$, b is *the absolute error* and $b/|a|$ is the *relative error.*

- *Systematic error* is an error caused by a *bias* of the measurement device. This can be fixed (at least partly) by calibration.

- *Statistical error* is caused by many independent, unknown effects. Measurements obey usually the normal distribution.

The situation is shown in Fig. 1. If the correct value of some quantity is 0 and due to an incorrect calibration our measurement device gives results which are two units smaller than the real value, our measured values are distributed around -2. Usually one can assume that this distribution is Gaussian ($\sim e^{-k(x-x_0)^2}$ for some k around $x_0 = -2$ in this example).

As a sidenote: one can actually derive exactly the results presented in this lecture assuming that the measured values follow the normal (Gaussian) distribution. However, we will not go trough the rigorous derivations here as these are not needed in IPhO. You will derive these results during your first year in the university.

2 Errors in calculations

2.1 Max–Min

"Max–Min" method is the simplest possible way to estimate the error. As a straightforward way it is a good tool for IPhO: even tough it gives quite large error estimates, you will get full points :)

Let us consider a situation where you want to measure some quantity f which is a function of some other quantities a and b (this method generalizes easily for an arbitrary number of quantities). For example if f was velocity, you would determine it by measuring quantities $a = s$ (length) and $b = t$ (time): $f = f(a, b) = a/b$.

Denote the absolute error of a by δa and similarly δb for b. We can now evaluate the largest and smallest possible values for f by calculating $f(a \pm \delta a, b \pm \delta b)$, and choosing the signs in such a way that we get the largest/smallest possible result. In this way we get f_{max} and f_{min}, and the error estimation for f is

$$\delta f = \max\left\{f_{\text{max}} - f, f - f_{\text{min}}\right\}. \tag{2}$$

Another (equally correct) way to define δf would be an average of these terms.

Exercise 1. *(a trivial one) You have measured* $s = 10 \pm 1$ cm *and* $t = 5 \pm 0.2$ s. *Evaluate velocity and its accuracy.*

Exercise 2. *Evaluate the error bars for the (quite an arbitrary) quantity*

$$q = \frac{a^c}{b} \sin(\pi a) \ln d, \tag{3}$$

where we have measured $a = 0.8 \pm 0.1$, $b = 2 \pm 0.3$, $c = 0.4 \pm 0.01$ *and* $d = 10 \pm 2$.

Exercise 3. *For a quantity* $q = a \cdot b$ *show the result*

$$\frac{\delta q}{|q|} = \frac{\delta a}{|a|} + \frac{\delta b}{|b|}. \tag{4}$$

You may assume that errors are small compared to the actual values of the quantities, namely $\delta a/|a| \ll 1$.

In addition consider a situation $q = a/b$, *and convince yourself that the above result holds also in this case. You may need an approximation*

$$\frac{1}{1 \pm x} \approx 1 \mp x \tag{5}$$

valid for small $x \ll 1$ (this kind of approximations are useful to know; can you derive this?).

Do you see how does these results generalize to the case where

$$q = \frac{a_1 \cdot a_2 \cdots a_N}{b_1 \cdot b_2 \cdots b_M} \tag{6}$$

(A derivation is not required)

2.2 Addition of errors

Usually errors do not move result to the same direction, so min-max-method gives quite large error estimate as it gives the worst-case scenario. Thus if we have two independent errors (say, statistical error and systematical error), more realistic method is to add errors in quadrature.

Consider for example a case in which $q = x + y$. In this case $\delta q = \delta x + \delta y$ clearly overestimates the error: if x and y are independent, a better idea is to compute uncertainty as

$$\delta q = \sqrt{(\delta x)^2 + (\delta y)^2}. \tag{7}$$

Compare with the Pythagora's theorem! This can be interpreted such that the independent error sources move the result to orthogonal directions (there is no correlation between the errors), and here we actually compute the most probable error instead of the maximum error.

For relative errors one can derive

$$\frac{\delta q}{q} = \sqrt{\left(\frac{\delta a_1}{a_1}\right)^2 + \cdots + \left(\frac{\delta a_N}{a_N}\right)^2} \tag{8}$$

for $q = a_1 \cdots a_k / (a_{k+1} \cdots a_N)$.

Addition in quadrature is usually used when you have got two (or more) different errors from different sources. Let us say that the statistical error of q is estimated to be δq_{stat} and that the systematic error is estimated as δq_{sys}. In that case $\delta q = \sqrt{(\delta q_{\text{stat}})^2 + (\delta q_{\text{syst}})^2}$.

For example we might measure the length of a pen by a ruler to be $L = 7.50$ cm, and we estimate the accuracy in which we can read the ruler is 0.10 cm. In addition, we can't be sure that that cheap ruler is really "calibrated" correctly: we estimate (or the manufacturer gives the estimate) that the ruler is "calibrated" in accuracy 0.05 cm. As a result the error estimate is $\delta L = \sqrt{(0.10 \text{ cm})^2 + (0.05 \text{ cm})^2} = 0.12$ cm.

Exercise 4. *The efficiency of an electric motor which lifts a mass m to an altitude h in time t is*

$$e = \frac{mgh}{VIt}, \tag{9}$$

where V and I are the voltage and the current that the motor uses, respectively. Let us suppose that m, h and V are measured in 1% accuracy, I in 3% accuracy and t only in 7% accuracy. Compute the relative error of the efficiency using a) the max-min method b) addition of errors quadratively. How does you result change if you neglect the smallest errors?

3 Fitting parameters and linearizing data

In order to obtain more accurate results, one usually has to vary some quantity and measure how some other quantity changes. For example, one can measure resistance of a resistor by varying the current in the circuit and measuring the voltage over the resistor. As $U = RI$, plotting current as a function of voltage the data points should fall on a same straight line. Fitting a straight line to this data set gives the resistance, as it is now just the slope of the line (there are mathematical algorithms to do that, but in practice [=IPhO] one draws the line by hand such that it agrees with the data points as well as possible).

Of course one could just measure current at one voltage and obtain a result. However, especially in more complicated measurements one data point may hide some errors. For example if in the above example the fitted line would not go trough the origin, there would probably be some systematic error in the measurement. But we could still obtain the resistance by evaluating the slope of the fitted line. Also, if the data point would not fall on a same curve, we might conclude that we see some new phenomena. In this example, at sufficiently large currents the temperature of the resistor increases significantly and the resistance probably changes also.

If the two quantities (above voltage and current) do not depend linearly on each other, fitting procedure is more complicated. As an example, let us consider a radioactive decay: if the half-life of an isotope is $t_{1/2}$, the number of active nuclei at time t is

$$N(t) = N_0 \exp(-\lambda t) = N_0 \exp\left(-\frac{\ln 2}{t_{1/2}} t\right), \tag{10}$$

where N_0 is the number of nuclei at time $t = 0$ and $\lambda = \ln 2/t_{1/2}$ is the decay constant. From this we can calculate the *activity* of the sample, namely

Table 1: Activity of the sample.

t/h	1	2	4	6	8	10	12	14
$A(t) \cdot 10^3$	80 ± 10	61 ± 7	31 ± 6	20 ± 5	9.9 ± 3	4.5 ± 1	3.1 ± 1	1.7 ± 1

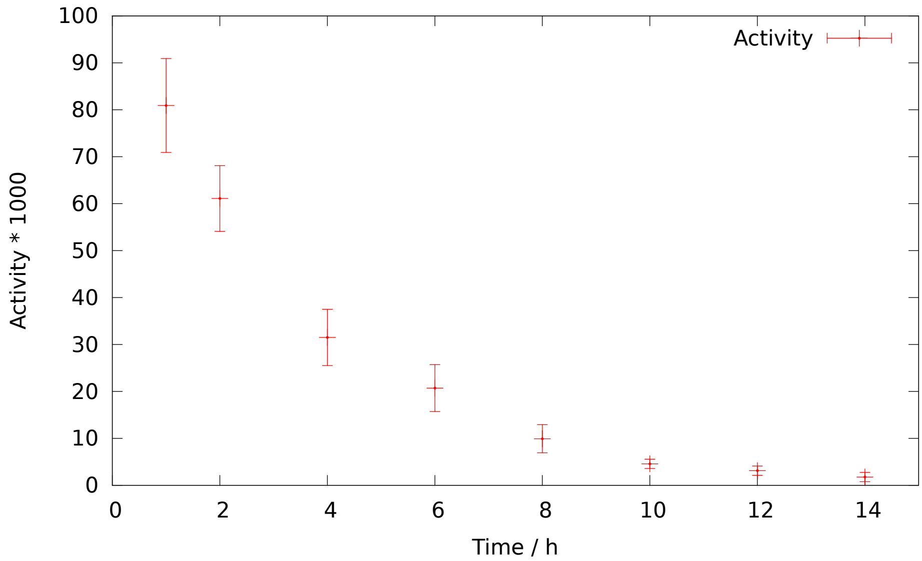

Figure 2: Activity as a function of time

number of decays per unit time which is nothing but a derivative (do you see why?)

$$A(t) = |N'(t)| = N_0 \lambda \exp(-\lambda t). \tag{11}$$

Let us say that we want to measure the half life of the given nucleus, and a Geiger counter is used to measure the activity as a function of time. The results with their corresponding error estimations are given in the table 1.

The results are plotted in figure 2. How can we find out the decay constant λ and thus the half-life? (Of course modern computers can fit a function $f(x) = A \cdot \exp(-Bt)$ and give you the constants, but you don't have one in IPhO :)

The idea is to *linearize* the data, plot it and fit a straight line to data set. Taking a logarithm on both sides of eq. (11) one gets

$$\ln A(t) = \ln \left[N_0 \lambda e^{-\lambda t} \right] = \ln(N_0 \lambda) - \lambda t \tag{12}$$

(Notice that here the world "logarithm" refers to the natural (base-e) logarithm). So we see that if we plot the logarithm of activity as a function of time and fit a straight line to the data set, the slope of the line will give us directly the decay constant λ.

As we learned at the beginning, we also have to estimate the error of our result. So in addition to the "best fit" you just did, you also have to fit (taking into account the error bars) lines which have largest and smallest possible slopes in such a way that they still somehow go trough all the points (within their error bars). In this way you get maximum and minimum values for the slope and thus for the decay constant, so you can estimate the error.

Exercise 5. *Follow the previously explained procedure and figure out the half-life of the nuclei (result is probably something like 2.5 hours) and estimate the error. Can you also figure out the number of nuclei at the beginning ($t = 0$)?*

4 Statistical analysis

In order to get more reliable results, one has to make a lot of measurements. But how to get an error estimation from large amount of data?

In section 3 we learned how to linearize data and fit some curves to it. This is not very useful method if you measure the same quantity many times and want to somehow combine the results.

Note: The methods described in this chapter work only if you have a lot of data points. How much is a lot? It depends, but let us say that less than ten is clearly too few. But on the other hand, keep in mind that in IPhO you shouldn't waste your time by making *too* many measurements.

Let us first define a quantity called *standard deviation* (in Finnish *keski-hajonta*) which measures how much the data points vary around the average. If the average is denoted by $\bar{x}$ and we have measured values $x_1, x_2, \ldots, x_N$, the standard deviation is

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (\bar{x} - x_i)^2}{N}}. \tag{13}$$

An another interpretation is that σ gives the most probable error of a single measurement.

(To be precise there should be $N - 1$ in the denominator, but as N is large, $N - 1 \approx N$. This is actually one possible way to determine whether you are allowed to do statistical analysis: if $N \approx N - 1$, then N is probably large enough).

Once the standard deviation is computed, it is easy to obtain an error estimation:

$$\delta\bar{x} = \frac{\sigma}{\sqrt{N}}. \tag{14}$$

If the datapoints have different error estimates, then it is no longer possible to just calculate the mean value, because it would mean that one considered measurements with huge errorbars as important as more accurate measurements.

Let us say that we have measured a quantity x N times, the results are x_i and the error estimation of each data point is δx_i. Then we should use $1/\delta x_i^2$ as a weight, and the weighted mean is

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i/\delta x_i^2}{\sum_{i=1}^{N} 1/\delta x_i^2} \tag{15}$$

Because the weight is square of the inverse error, measurements with poor accuracy contribute little to the final result.

The error estimation is

$$\delta\bar{x} = \frac{1}{\sqrt{\sum_{i=1}^{N} 1/\delta x_i^2}}. \tag{16}$$

Derivation of this result is problem 8.

Exercise 6. *(A boring one...)*
Estimate an error for the quantity x if measurements are $x_1 = 5$, $x_2 = 8$, $x_3 = 10$, $x_4 = 7$, $x_5 = 6$, $x_7 = 6$, $x_8 = 22$, $x = 9$, $x_{10} = 7$, $x_{11} = 4$ and $x_{12} = 5$.

5 General formula for error propagation

So far we have gone trough a few (more or less) straightforward ways to compute an error of the measured quantity. What you have learned this far is enough for IPhO. However, the general law presented in this section is, in my opinion, so easy and quick to use that it might be worth of knowing also in IPhO. Plus it gives smaller errors than min-max method!

The idea is that we use derivative of the function to approximate how much the error in some quantity x affects our result $q = q(x, y, z, \dots)$. As you may know, the derivative gives the local slope of the function. Thus, if the absolute value of the derivative is large, only a small error in x causes huge difference in the final result q.

The situation is shown in Fig. 3, where all other variables y, z and so on are kept constant. Now as $q'(x)$ gives the local slope of the function $q(x)$, we can approximate that

$$q(x + \delta x) = q(x) + q'(x)\delta x. \tag{17}$$

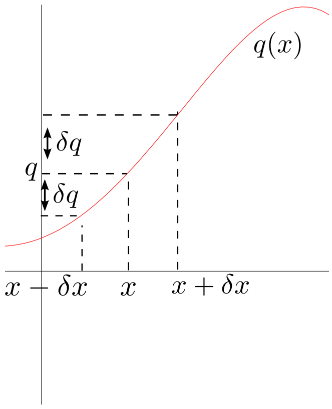

Figure 3: $q(x)$ as a function of x, and effect of δx on δq.

which gives

$$\delta q = q(x + \delta x) - q(x) = q'(x)\delta x. \tag{18}$$

Thus it is enough to calculate the derivative of the function q in order to find out the error estimation! Notice that here we basically approximate function using its first order Taylor polynomial around x to see what happens when x changes a bit. Higher order terms are neglected, which requires that δx is small ($\delta x^2 \ll \delta x$).

We can of course do the same calculation for other variables y, z, ... and add these errors in quadrature:

$$\delta q = \sqrt{\left(\frac{\partial q}{\partial x}\delta x\right)^2 + \left(\frac{\partial q}{\partial y}\delta y\right)^2 + \left(\frac{\partial q}{\partial z}\delta z\right)^2} \tag{19}$$

if $q = q(x, y, z)$ *and all variables are independent.*

A nice thing is that now you don't have to think about which combination of parameters gives the largest/smallest value: it is enough to just calculate the derivatives and add the results in quadrature.

Exercise 7. *In order to measure the lattice constant of a given material one measured that the second intensity maxima was seen in angle* $\theta = 30 \pm 2°$, *when the laser wavelength was* $\lambda = 450 \pm 30$ nm. *Calculate the lattice constant and its error using the general formula for error propagation. (Recall:* $d \sin\theta = k\lambda$, d *is the lattice constant and* k *the number of maxima).*

Exercise 8. *Derive equations* (16) *and* (8) *(in the case where* $q = x_1 \cdots x_N$*). Hint: Notice that the weighted average is a function of original measured values.*

Exercise 9. *First, show that the relative uncentainty for the quantity* $q = x^2$ *is*

$$\frac{\delta q}{q} = 2\frac{\delta x}{|x|}. \tag{20}$$

Now, compare result (20) *with the result you get from the equation* (8) *by writing* $q = x \cdot x$*, so that*

$$\frac{\delta q}{q} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta x}{x}\right)^2} = \sqrt{2}\frac{\delta x}{|x|}. \tag{21}$$

According to the previous problem, both of these results follow from the general rule for error propagation. Why do we get different results? Which one is correct?

Exercise 10. *Consider a quantity* $q = x(\ln x - 1)$*. After a careful measurement of* x*, one obtains* $x = 1.00 \pm 0.07$*. Calculate* q *and estimate its error using a) general formula for error propagation b) max-min method. Which one is correct here?*

6 About measurement devices

First, notice that according to IPhO syllabus, "Candidates must be aware that instruments affect measurements". (If you haven't done it yet, take a look of the syllabus [list of topics you should be familiar with], you find it from page `http://ipho.phy.ntnu.edu.tw/syllabus.html`). For example, multimeter has a finite resistance when it measures voltage/current.

There are two clearly different sources of errors when one measures something with some device: the measurement procedure might be difficult (e.g. the setup geometry may make it difficult to use ruler to measure distances) and there is some internal error attached in each device. First one is more or less easy to estimate, the last one requires some calculation.

Device manufacturer usually tells the accuracy of the device e.g. in the manual. In easiest case the error is given to be e.g. 1%, so the error of the measured value a is just $0.01 \cdot a$. At the end this error and other estimated errors (e.g. from the difficulties in reading the values or using the device) are added in quadrature.

The accuracy of a mutlimeter is usually given in a form "$\pm 0.9\% + 1$ dig". Let us say we use a multimeter in range where you read the voltage to be 10.00 V. The accuracy is then calculated as $0.009 \cdot 10.00$ V+ last digit that counts (the last one that the device shows you), in this case 0.01 V. So the total error estimation is $(0.09 + 0.01)$ V $= 0.1$ V.

Exercise 11. *To measure the resistance of a resistor, one measures the voltage over it to be* $U = 2.0$ V *and the current in the circuit to be* 0.15 A. *In the range used the multimeter accuracy for voltage measurements is* $\pm 1\% + 2$ *digits and for current measurements* $\pm 2\% + 1$ *digit. Calculate the resistance and its accuracy. Use the general formula for error propagation.*

Exercise 12. *List a few possible experimental setups (that might be possible in IPhO) where one has to take into account that instruments affect measurements.*

7 Final words

One more thing to learn: rounding. Errors are always rounded up. In the final result you should have as many significant numbers in the error estimation as in the result. For example $1.2345 \pm 0.322456 \approx 1.2 \pm 0.4$.

In real experiment you have a lot of error sources. In order to avoid unnecessary work one can include in the error analysis only uncertainties which are dominant. For example, if in some experiment you measure the weight of a ball using an accurate weighing scale in 0.0001 g accuracy and then use a stopwatch to measure the time it takes for the ball to fall from a given height, it makes no sense to include this tiny mass uncertainty in the error analysis. This is especially important in IPhO as there you do not have the luxury of having an extra time to do "too accurate" error estimates. Keep it simple!

Error analysis may not be the most interesting topic, but is something you just have to handle in order to get points in IPhO (and to become a physicist).

References

[1] John R. Taylor. An Introduction to Error Analysis. ISBN 978-0935702750.