Problem Solving II: Experimental Methods

For more about the theory of uncertainty analysis, see this handout, or for a more introductory take, this handout and this comic. For practical tips for real experiments, especially at the IPhO, see chapter 7 of *Physics Olympiad: Basic to Advanced Exercises*. For some entertaining general discussion, see chapters I-5 and I-6 of the Feynman lectures. There is a total of **81** points.

1 Basic Statistics

Idea 1

If a quantity X has the probability distribution $p(x)$, that means

$$\text{the probability that } a \leq X \leq b \text{ is } \int_a^b p(x)\,dx.$$

In particular, the total probability has to sum to one, so

$$\int_{-\infty}^{\infty} p(x)\,dx = 1.$$

Using the probability distribution, we can calculate expectation values, i.e. averages. For example, the expectation value of X, also called the mean, is

$$\langle X \rangle = \int_{-\infty}^{\infty} x p(x)\,dx$$

while the expectation value of an arbitrary function of X is

$$\langle f(X) \rangle = \int_{-\infty}^{\infty} f(x) p(x)\,dx.$$

One especially important quantity is the variance of X, defined as

$$\operatorname{var} X = \langle X^2 \rangle - \langle X \rangle^2.$$

The standard deviation is defined by $\sigma_X = \sqrt{\operatorname{var} X}$. It describes how "spread out" the distribution of X is, and it will play an important role in uncertainty analysis.

2 Uncertainty Analysis

Idea 2

When a physical quantity is measured in an experiment and reported as $x \pm \Delta x$, it is uncertain what the true value of the quantity is. If the quantity has a probability distribution $p(x)$, then the reported uncertainty Δx is essentially the standard deviation of $p(x)$.

Remark

In practice, you'll have to use intuition and experience to assign uncertainties for real measurements. For example, if you're using a clock that times only to the nearest second, you might take $\Delta t = 0.5\,\text{s}$. If you're using a good ruler, which has millimeter markings, you might take $\Delta x = 0.5\,\text{mm}$, though you can actually do a bit better if you look carefully. Of course, the ultimate test is the results: if you assigned the uncertainties right, your final uncertainty should encompass the true result most (but not all) of the time.

Remark

Note how this differs from "high school" uncertainty analysis. In school, you might be told to show uncertainty using significant figures, and when adding two things, to keep only the figures that are significant in both of them. That corresponds to

$$\Delta(x+y) = \max(\Delta x, \Delta y)$$

which is an underestimate. Or, you might be told that the uncertainty needs to encapsulate all the possible values, which implies that

$$\Delta(x+y) = \Delta x + \Delta y$$

which is an overestimate, since the uncertainties could cancel.

Idea 3

If x has uncertainty Δx, and $f(x)$ can be approximated by its tangent line, $f(x') \approx f(x) + (x'-x)f'(x)$ within the region $x \pm \Delta x$, then $f(x)$ has approximate uncertainty $f'(x)\,\Delta x$.

Idea 4

For practical computations, it is often useful to use relative uncertainties. The relative uncertainty of x is $\Delta x/x$, and can be expressed as a percentage.

Remark

In this problem set, we have given rules for calculating the mean and standard deviation of derived quantities. But in general, probability distributions can have all kinds of weird features, which aren't captured by those two numbers. The reason we focus on them anyway is because of the central limit theorem, which roughly states that if we have many independent random variables, the distribution of the sum will approach a normal distribution. As you saw in problem 4, normal distributions are characterized entirely by their mean and standard deviation, so we don't lose any information by reporting only those two quantities.

Remark

There are many situations where the rules above can't be used. For example, consider the uncertainty of $x + y^2/x$, where x and y have independent uncertainties. You can calculate the uncertainty of either term with the standard rules, but you can't calculate the uncertainty of their sum, because the terms are not independent (both contain x).

In these cases, you can use the multivariable equivalent of the tangent line approximation,

$$f(x', y') \approx f(x, y) + (x' - x)\frac{\partial f}{\partial x} + (y' - y)\frac{\partial f}{\partial y}.$$

Adding the two contributions to the uncertainty in quadrature gives

$$\Delta f = \sqrt{\left(\frac{\partial f}{\partial x}\,\Delta x\right)^2 + \left(\frac{\partial f}{\partial y}\,\Delta y\right)^2}.$$

This is the general rule that includes the rules you derived above as special cases. However, it shouldn't be necessary in Olympiad problems. For example, if you run into such situations in an experiment, often one of the uncertainties is much smaller, and can be neglected entirely.

Remark

As you saw in problem 9, the tangent line approximation can sometimes fail. The proper way to handle situations like these would be to find the full probability distribution of the desired quantity, rather than just describing it crudely with its standard deviation. However, this can't be done analytically except in the simplest of cases. So when professional physicists run into situations like these, which are quite common, they often just numerically compute a few million or billion values, starting with randomly drawn inputs each time, and use that to infer the probability distribution. This technique is called Monte Carlo. It's very powerful, but certainly not relevant to Olympiad physics, or even to most physicists! On Olympiads, you should just fall back to something reasonable, such as taking the minimum and maximum possible values.

3 Data Analysis

Idea 5

All graphical data analysis for the USAPhO and IPhO can be performed by drawing a line and measuring its slope and intercept. This is a bit artificial, but it's necessary because of the limited calculation equipment you have during these exams. Despite this, drawing lines can be surprisingly powerful.

Remark

When performing data analysis in practice, you should neatly organize your work. Always make a data table that explicitly shows what you're calculating, and make a neat graph with a ruler and graph paper. Set the axis scale so that the graphed data points cover almost the entire page, and let the x-axis include $x = 0$ if you need to find a y-intercept. The computation of the slope should be explicitly shown. For each line you should use at least about five points; you don't have to use them all. If you have a calculator that can find best-fit slopes for you, don't use it, as these features are generally not allowed on real Olympiads.

Idea 6

Historically, uncertainty analysis has only appeared on the $F = ma$, and data analysis has only appeared on the USAPhO, but the two appear together in the IPhO.

To perform uncertainty analysis for best fit lines, plot the uncertainties of the data points as error bars. Then draw the steepest and shallowest lines that still pass through most of the error bars. These will give you the bounds on your slope and intercept. We'll see some examples of this procedure in later problem sets. It isn't the most mathematically rigorous method, but it gives decent results.

4 Estimation

Estimation is a useful skill for checking the answers to real-world problems.