Lecture note - Error analysis

ZHAO, Yongfeng
Center for Soft Condensed Matter Physics and Interdisciplinary Research & School of Physical Science and Technology, Soochow University, 215006 Suzhou, China
(Dated: August 8, 2022)

CONTENTS

I. INTRODUCTION

A. Objectives of experimental courses

Physics, especially theoretical physics, is built on experiments. Unlike biology and chemistry, we do not have many standard protocols, with which one must be skilled in research and applications in physics. Thus, the training in the undergraduate physics experimental courses is mainly about the following.

- **Complement theoretical courses.**

  We try to convince the students that the theories they learn in the other courses are not bullshit. The theories we learned in the mechanics and electromagnetism courses are built to describe and analyze experimental phenomena. Finally, we must return to the real world to see how they work.

- **The basic concept of academic integrity.**

  To build correct knowledge in physics, one must respect academic integrity. The first step is to respect the reality: Experimental data are from experiments. **Experimental data can not be faked or plagiarized.**

  Due to the randomness of nature, the numbers of individual measurements are not important at all. There is no perfect answer for the data, and it will not contribute to the final mark. Faking or plagiarizing data does not bring more grades. The student will lose all the data grades for one experiment if one fakes or plagiarizes data.

- **Error analysis.**

  It will be the topic of this note.

B. Measurement

Measurement is at the center of physics experiments. Physics is built on quantitative and precise measurements of quantities, from which we find relations of quantities and formulate theories to describe nature. Some quantities can

be measured **directly**. E.g., One can measure the thickness of a plate using an outside micrometer. Some quantities have to be inferred **indirectly**, using the existing knowledge of physics law. E.g., One can use Ohm's law to estimate the resistance of a resistor by measuring the voltage and the current across the resistor.

Philosophically, we believe there is a **ground truth** in each physical quantity. We want to know the ground truth through measurements. It is, however, a basic fact that **no measurements can be exact in reality**. It is safe to say that all new physics and technologies come from increasing preciseness. Hence, improving the preciseness of measurements is a crucial question. The answer will require knowledge of errors.

C. Error

Error is the deviation between the measurements and the ground truth. Error analysis is about the origin of errors and the estimation of uncertainties. Through these analyses, we will know:

- How good is a measurement result? To what extent can we trust our measurements.
- How to improve the accuracy of a measurement, considering the limited preciseness of instruments?

We usually classify errors into two types based on their origin and statistical behavior.

- **Systematic error** is a systematic bias in the measurements. It can come from the systematic bias of instruments, inaccurate methods, systematic change in the environment, and a bias from the experimentalist.

  We can use better instruments, improve the methods, and control the environment to reduce systematic errors.

- **Random error** varies from one observation to another. It usually comes from thermal motion, random events, and the intrinsic randomness of nature.

  The primary method to reduce random error is repeating the same observation and averaging the result.

Note that an error is not a mistake in the experiment. **Error is inevitable and can not be eliminated.** We assume the experiment has been correctly carried out when talking about errors.

D. Uncertainty

Uncertainty is the range in which the ground truth of a measurement lies with a certain probability. It denotes the confidence interval of a measurement result. A measurement result x is usually denoted as

$$x \pm u_x \ , \tag{1}$$

where u_x is the uncertainty of the quantity X. The expression indicates that the ground truth of X lies with a certain probability in the interval $(x - u_x, x + u_x)$. **A measurement result can be complete only when accompanied by uncertainty and unit.**

Here we note that only 1-2 significant figures in u_x are enough for the purpose. Since the preciseness of the result is bounded by u_x, the significant figures beyond u_x become meaningless. The figures in the final result should align with the uncertainty.

Example 1: Consider a measurement of length gives $l = 3.9514$ cm and $u_l = 0.1261$ cm. Firstly, we keep only one significant figure in $u_l = 0.1$. Then since the ground truth lies probably in $(3.9514 - 0.1 \text{ cm}, 3.9514 + 0.1 \text{ cm})$, more than 2 significant figures in l becomes meaningless. The final result should be expressed as

$$l = 4.0 \pm 0.1 \text{ cm} \ .$$

Here we apply the rounding rule of "only on and no off", such that the last digit should always be plus one. Note that the last digit 0 in l cannot be neglected.

Example 2: A measurement of length gives $l = 2.6 \pm 0.3$ cm. If the error is normally distributed and the uncertainty is reported as standard deviation, the probability that the ground truth lies in $(2.3 \text{ cm}, 2.9 \text{ cm})$ is 68.3%. Because for normal distribution

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\sigma}^{\sigma} e^{-\frac{x^2}{2\sigma^2}} \approx 0.683 \ . \tag{2}$$

Due to the random error, the measurement result is a random variable from a specific (but unknown) probability distribution. The uncertainty is commonly taken as the standard deviation for this probability distribution. In the next section, we will talk about how to estimate it.

II. THE ESTIMATION OF UNCERTAINTIES

A. The uncertainty from the finite precision of instruments of a single measurement

All the instruments, in reality, have finite precision and accuracy, thus limiting the uncertainty of a single measurement. Let us first discuss the instrumental uncertainty in a single direct observation.

The error of an instrument is usually calibrated and measured by the manufacturer. One needs to read the manual of the instrument to know how to determine its uncertainty. It can vary from one instrument to another.

Usually, the error of an instrument is given in terms of tolerance. The error in most instruments can be considered uniformly distributed within a range bounded by the tolerance. For example, suppose we measure a length of 1.56 cm with a ruler of division 0.1 cm. The ground truth will not leave the interval $(1.51, 1.61)$ cm. The tolerance is $0.1/2 = 0.05$ cm, and the measurement result can be assumed uniformly distributed in $(1.51, 1.61)$ cm.

If a random variable x is distributed uniformly in $(\bar{x}-\Delta_I, \bar{x}+\Delta_I)$, where Δ_I is the tolerance, its standard deviation can be calculated as

$$\sigma_x^2 = \frac{1}{2\Delta_I}\int_{\bar{x}-\Delta_I}^{\bar{x}+\Delta_I}(x-\bar{x})^2\,dx = \frac{\Delta_I^2}{3}\,. \tag{3}$$

Thus the uncertainty of the instrument u_I, taken as standard deviation σ_x, is given from the tolerance Δ_I by

$$u_I = \frac{\Delta_I}{\sqrt{3}}\,. \tag{4}$$

Δ_I is usually read as a half division or calculated from the information from the manufacturer of the instrument.

Remark: The instrumental error is not necessarily random but can be systematic.

Example 3: For the resistor box used in the single arm electric bridge (Fig. 1), the manufacturer provides the error table as Tab. I. According to the manual, the tolerance of the resistor box is given by $\Delta_I = \sum_i R_i \times \alpha_i\% + R_0$, where R_i is the resistance of the i-th dial, and the residual resistance $R_0 = 0.02\ \Omega$.

Figure 1. A resistor box.

If one sets the resistor box at 31935.7 Ω, the tolerance is given by

$$\Delta_I = 30000\ \Omega\times 0.1\% + 1000\ \Omega\times 0.1\% + 900\ \Omega\times 0.1\% + 30\ \Omega\times 0.1\% + 5\ \Omega\times 0.5\% + 0.7\ \Omega\times 2\% + 0.02\ \Omega = 31.969\ \Omega\,, \tag{5}$$

disk multiples	×10000	×1000	×100	×10	×1	×0.1
accuracy class α	0.1	0.1	0.1	0.1	0.5	2

Table I. The error table of the resistor box in Fig. 1.

and the uncertainty of the resistor box is

$$u_I = \frac{\Delta_I}{\sqrt{3}} = 18.457\ \Omega = 20\ \Omega\ .$$

The resistance of the resistor box should be written as $(3.194 \pm 0.002) \times 10^4\ \Omega$.

Remark: Using scientific notation to present measurement results can keep the number of significant figures correct.

B. The uncertainty from statistics of repeated measurements

We do not require the students to be skilled in the theory of uncertainty in this section. Some knowledge will help understand the important formulas coming in the following sections.

Let us consider the measurement of the ground truth X and focus on the random error from other than finite instrument precision. Due to random error, the measurement result x is a random variable. Both X and the probability distribution of x are unknown, so we can only estimate the uncertainty from a finite series of measurement results x_i of n trials, $i = 1, 2, \cdots, n$.

It is the mathematical theorem, **Central Limit Theorem**, that enables the estimation of uncertainties. We now state the theorem roughly as following. For n random variables r_1, r_2, $\cdots$, r_n, the mean value

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i \tag{6}$$

is apparently a random variable as well. No matter what probability distribution each r_i obeys, if they are identically distributed and if they are independent, in the $n \to \infty$ limit, the probability distribution of $\bar{r}$ converges to a normal distribution almost surely. If μ and σ are the mean value (expectation) and the (finite) standard deviation of each r_i, the normal distribution of $\bar{r}$ has mean μ and standard deviation $\sigma/\sqrt{n}$.

How does the theorem apply to the analysis of uncertainties? If the experimental condition of measuring X is well-controlled, we can assume each observation x_i is independent and identically distributed. If the systematic error has been correctly accounted for, the error $\Delta_i = x_i - X$ has a mean zero. Thus when the number of trials n is large enough, the mean value $\bar{x}$ converge to X, and the uncertainty taken as standard deviation is given by $u_{\bar{x}} = \sigma_\Delta/\sqrt{n}$.

$\sigma_\Delta = \lim_{n\to\infty}\sqrt{\sum_i \Delta_i^2/n}$ is still unknown. To estimate σ_Δ, we define $\delta_i = x_i - \bar{x}$. We note that

$$\sum_i \Delta_i = \sum_i x_i - nX = n\bar{x} - nX\ , \tag{7}$$

$$\bar{x} = X + \frac{1}{n}\sum_i \Delta_i\ . \tag{8}$$

Then we have

$$\sum_i \delta_i^2 = \sum_i (x_i - X - \frac{1}{n}\sum_j \Delta_j)^2 = \sum_i (\Delta_i - \frac{1}{n}\sum_j \Delta_j)^2 = \sum_i \Delta_i^2 - \frac{2}{n}\sum_{i,j}\Delta_i\Delta_j + \frac{n}{n^2}\sum_i \Delta_i^2\ . \tag{9}$$

Since Δ_i are independent and has zero mean, the second term is always zero if $i \neq j$. We thus have

$$\sum_i \delta_i^2 = \frac{n-1}{n}\sum_i \Delta_i^2\ . \tag{10}$$

Then an estimation of σ_Δ gives

$$\sigma_\Delta = \sqrt{\frac{\sum_i \Delta_i^2}{n}} = \sqrt{\frac{\sum_i \delta_i^2}{n-1}} = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}}\ . \tag{11}$$

Now we have the following estimation on $\bar{x}$ and its uncertainty $u_{\bar{x}}$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \,, \qquad\qquad u_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}} \,. \tag{12}$$

Remarks:

1. Note that the standard deviation σ_Δ of each measurement x_i is a constant of the number of measurement n, which characterizes the nature of the measurement error distribution. $u_{\bar{x}} = \sigma_\Delta/\sqrt{n}$ is the standard deviation of the mean value $\bar{x}$, which shows how fluctuating $\bar{x}$ is.
2. Now we see why we need multiple repeats and averages in measurements. The mean $\bar{x}$ converges to X if the number of trials is large. And the uncertainty $u_{\bar{x}}$ scales as $o(n^{-1/2})$, which decays slowly with n.
3. Because of the central limit theorem, it is safe to assume the error $\Delta_i = x_i - X$ is always normally distributed. After many trials and averaging, we will not see the difference if it is not a normal distribution.

C. The combined uncertainty of repeated measurements

If we measure a quantity X directly with real instrument n times, the estimation by the mean value can be expressed as

$$\bar{x} = X + \Delta = X + \Delta_I + \Delta_R \,, \tag{13}$$

where Δ_I is the error from the instrument precision, and Δ_R is the random error. The uncertainties of Δ_I and Δ_R are given by u_I and u_R, respectively. The uncertainty of the final result is thus a combination of the two sources of error. If we assume the instrumental and random errors are uncorrelated, we have the following.

The combined uncertainty of n measurements x_i can be estimated as

$$u = \sqrt{u_I^2 + u_R^2} \,, \tag{14}$$

where u_I is the uncertainty from instrument precision given by Eq. (4), and u_R is the uncertainty of random error given by (12).

Remark: If we assume Δ_I and Δ_R are both normally distributed with zero mean, the sum is also a random variable of normal distribution. The standard deviation of the sum is given by $\sqrt{u_I^2 + u_R^2}$. The mean of Δ_I, however, cannot be predicted and may not be zero since it can be a systematic error. It is why u_I given by (4) does not have the $1/\sqrt{n}$ factor: The systematic error cannot be reduced by repeating the measurements. Nevertheless, this unknown systematic error has to be captured by uncertainty. Hence, the combination rule (14) is a convention (by the International Standards Organization) without rigorous proof.

Example 4: 5 measurements of length with a ruler give the following results: 1.78 cm, 1.75 cm, 1.79 cm, 1.80 cm, 1.75 cm, and the division of the ruler is 0.1 cm. The average gives

$$\bar{l} = \frac{1.78 + 1.75 + 1.79 + 1.80 + 1.75}{5} \text{ cm} = 1.774 \text{ cm} \,.$$

The tolerance of the ruler is half a division 0.05 cm, and the instrumental uncertainty is given by

$$u_I = \frac{0.05}{\sqrt{3}} \text{ cm} = 0.03 \text{ cm} \,.$$

The uncertainty from random error is

$$u_R = \sqrt{\frac{(1.78-1.774)^2 + (1.75-1.774)^2 + (1.79-1.774)^2 + (1.80-1.774)^2 + (1.75-1.774)^2}{5\times(5-1)}} \text{ cm} = 0.01 \text{ cm} \,.$$

The combined uncertainty is

$$u = \sqrt{u_I^2 + u_R^2} = 0.03 \text{ cm} .$$

Finally, the measurement result is expressed as

$$l = 1.77 \pm 0.03 \text{ cm} .$$

Note that the precision of the instrument now dominates the error.

D. The uncertainty of indirect measurements

Not many physical quantities can be measured directly. For instance, the resistance R of a resistor can be measured from the current I and voltage U across the resistor by Ohm's law $R = U/I$. The measurements in I and U both have uncertainties, bringing uncertainty in the calculated R. Thus we need to calculate the uncertainty of R based on the uncertainties of I and U.

Consider a general case, where we have m independent direct measurements x_i, $i = 1, 2, \cdots, m$. The uncertainty of x_i is u_i. The measurement of y is calculated through a function $y = f(x_1, x_2, \cdots, x_m)$. What is the uncertainty u_y of y?

Each quantity x_i can have an error of $\delta x_i = x_i - \bar{x}_i$. Assuming the error in each measurement is small, we can estimate the error δy in y by the partial derivatives

$$\delta y = \sum_{i=1}^{m} \left(\frac{\partial f}{\partial x_i} \right) \delta x_i . \tag{15}$$

If all the errors δx_i are normally distributed, the linear combination in Eq. (15) gives another normal distribution. The standard deviation of δy is given by

$$\sigma_y = \sqrt{\sum_{i=1}^{m} \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2} . \tag{16}$$

This can be shown by calculating the sum $\sum_{j=1}^{n} \delta y_j^2$, where n is the number of experiments, and j labels different measurement. We have

$$\sum_{j=1}^{n} \delta y_j^2 = \sum_{j=1}^{n} \sum_{i=1}^{m} \left(\frac{\partial f}{\partial x_i} \right)^2 \delta x_{i,j}^2 + 2 \sum_{j=1}^{n} \sum_{i \neq k} \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_k} \right) \delta x_{i,j} \delta x_{k,j} . \tag{17}$$

If all the quantities are uncorrelated, we have $\sum_{j=1}^{n} \delta x_{i,j} \delta x_{k,j} = 0$ for $i \neq k$, and the second term equals zero. By the definition of standard deviation $\sigma_i = \sqrt{\sum_{j=1}^{n} (\bar{x}_i - x_{i,j})^2/(n-1)}$, we have Eq. (16). Thus we have the following.

The uncertainty u_y of a quantity y measured by a function $y = f(x_1, x_2, \cdots, x_m)$ is given by

$$u_y = \sqrt{\sum_{i=1}^{m} \left(\frac{\partial f}{\partial x_i} \right)^2 u_i^2} , \tag{18}$$

where u_i is the uncertainty of each x_i given by (14).
If it is easier to calculate the derivatives of $\ln f$ than of f, Eq. (18) is equivalent to

$$\frac{u_y}{y} = \sqrt{\sum_{i=1}^{m} \left(\frac{\partial \ln f}{\partial x_i} \right)^2 u_i^2} . \tag{19}$$

Function $z = f(x, y)$	Formulas for u_z
$z = x + y$	$u_z = \sqrt{u_x^2 + u_y^2}$
$z = x - y$	$u_z = \sqrt{u_x^2 + u_y^2}$
$z = xy$	$\frac{u_z}{\bar{z}} = \sqrt{\left(\frac{u_x}{\bar{x}}\right)^2 + \left(\frac{u_y}{\bar{y}}\right)^2}$
$z = \frac{x}{y}$	$\frac{u_z}{\bar{z}} = \sqrt{\left(\frac{u_x}{\bar{x}}\right)^2 + \left(\frac{u_y}{\bar{y}}\right)^2}$
$z = kx$	$u_z = \lvert k\rvert u_x$
$z = x^k$	$\frac{u_z}{\bar{z}} = \frac{\lvert k\rvert u_x}{\bar{x}}$
$z = \sqrt[k]{x}$	$\frac{u_z}{\bar{z}} = \frac{1}{k}\frac{u_x}{\bar{x}}$
$z = \sin x$	$u_z = \lvert\cos x\rvert u_x$
$z = \ln x$	$u_z = \frac{u_x}{\bar{x}}$
$z = e^x$	$\frac{u_z}{\bar{z}} = u_x$

Table II. Uncertainty of commonly used functions.

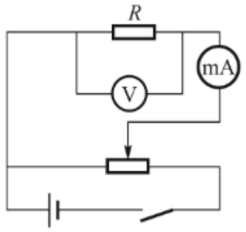

Figure 2. The circuit to measure the resistance of a resistor.

For convenience, we show a list of commonly used functions as in Tab. II.

Example 5: We measure the resistance of a resistor using the circuit shown in Fig. 2.

Because the voltmeter has a finite inner resistance R_V, the resistance R of the resistor is given by

$$R = \frac{U}{I - U/R_V} = \frac{R_V U}{R_V I - U} \,. \tag{20}$$

The three quantities U, I, R_V can all have an uncertainty. To calculate the uncertainty of R, we firstly need to calculate the derivatives of R

$$\frac{dR}{R} = \left(\frac{1}{U} + \frac{1}{IR_V - U}\right) dU + \left(\frac{1}{R_V} - \frac{I}{IR_V - U}\right) dR_V - \frac{R_V}{IR_V - U} dI \,. \tag{21}$$

Then the uncertainty of R is given by

$$\frac{u_R}{R} = \sqrt{\left(\frac{1}{U} + \frac{1}{IR_V - U}\right)^2 u_U^2 + \left(\frac{1}{R_V} - \frac{I}{IR_V - U}\right)^2 u_{R_V}^2 + \left(\frac{R_V}{IR_V - U}\right)^2 u_I^2} \,. \tag{22}$$

If one measurements gives $R_V = 1000\ \Omega$, $U = 7.0 \pm 0.2$ V, $I = 35 \pm 3$ mA, and the uncertainty u_{R_V} can be neglected,

we have

$$R = \frac{1000 \times 7}{1000 \times 0.035 - 7}\ \Omega = 250\ \Omega\ .$$

And

$$u_R = 250 \times \sqrt{\left(\frac{1}{7} + \frac{1}{1000 \times 0.035 - 7}\right)^2 \times 0.2^2 + \left(\frac{1000}{1000 \times 0.035 - 7}\right)^2 \times 0.003^2}\ \Omega = 28\ \Omega\ .$$

The final result is

$$R = (2.5 \pm 0.3) \times 10^2\ \Omega\ .$$

E. The uncertainty of linear least-squares fitting

Some physical quantities can be linearly related, and some important quantities are measured from the coefficients. Consider n data points (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) obeying a linear relation $y = kx + b$. Because of errors, they do not lie on a straight line perfectly. How do we estimate k, b, and their uncertainties?

To estimate k and b, we can find a line which is as close as possible to all the data points. Specifically, we define the mean-square error

$$\epsilon(k, b) = \sum_{i=1}^{n} (y_i - kx_i - b)^2 \tag{23}$$

to measure the "closeness" of the line to the data points. $\epsilon = 0$ if they are perfectly lining. The best estimations of k and b are those minimizing ϵ. Taking $\partial\epsilon/\partial k = \partial\epsilon/\partial b = 0$, we have the equations

$$\frac{\partial \epsilon}{\partial k} = -\sum_{i=1}^{n} 2(y_i - kx_i - b)x_i = 0\ , \tag{24}$$

$$\frac{\partial \epsilon}{\partial b} = -\sum_{i=1}^{n} 2(y_i - kx_i - b) = 0\ . \tag{25}$$

They can be solved to give the following.

For n data points (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) obeying a linear relation $y = kx+b$, the least-square estimations of k and b are

$$k = \frac{\sum_i x_i y_i - \sum_i x_i \sum_j y_j}{n\sum_i x_i^2 - (\sum_i x_i)^2} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}\ , \qquad b = \bar{y} - k\bar{x}\ . \tag{26}$$

If we assume the residue $y_i - kx_i - b$ is a normally distributed variable centered at zero, we can estimate the uncertainties of k and b through the uncertainties of each point and simple error propagation. Lengthy calculations give

$$u_k = \sqrt{\frac{1}{n(\overline{x^2} - \bar{x}^2)} \cdot \frac{\epsilon}{n-2}}\ , \qquad u_b = \sqrt{\frac{\overline{x^2}}{n(\overline{x^2} - \bar{x}^2)} \cdot \frac{\epsilon}{n-2}}\ . \tag{27}$$

We often talk about the correlation coefficient r in linear least-squares fitting, which is defined as

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}\ . \tag{28}$$

$|r| = 1$ if the points are perfectly lining. $r = 0$ means x and y are uncorrelated.

Remark: According to (27), the uncertainty in intercept b can be large.