Instrumentation Skills in Physics-II

Unit II- Error Analysis and Measurement Uncertainty

Pritam Das, Salbari College

1 Errors and error analysis

When one tries to classify the errors in numerical computations, it might be useful to study the sources of the errors and the growth of the individual errors. The sources of the errors are essentially static while the growth takes place dynamically. There are essentially three sources of errors:

- Initial Errors
- Local round-off errors
- Local truncation errors

The initial errors are errors in the initial data, a simple example of which is when data are obtained from a physical or chemical apparatus.

Round-off errors depend on the fact that practically each number in a numerical computation must be rounded (or chopped) to a certain number of digits.

Truncation errors arise when an infinite process (in some sense) is replaced by a finite one. Examples of this include the computation of a definite integral through approximation by a sum or the numerical integration of an ordinary differential equation by some finite difference method.

1.1 Absolute & Relative Errors

Absolute Error: The absolute value of the difference between the number x and its finite representation $f(x)$.

$$A.E. = |x - f(x)| \tag{1}$$

Relative Error: The ratio of the absolute error and the number x:

$$R.E. = \frac{|x - f(x)|}{|x|} \geq 0 \tag{2}$$

When does one use one or the other definition of error? Consider $x = 0.33$ and $f(x) = 0.30$. Clearly, A.E $= 0.03$ and R.E $= 0.03/0.33 = 0.09091 \simeq 9.1\%$. When $x = 0.33 \times 10^{-5}$ and $f(x) = 0.30 \times 10^{-5}$, A.E. $= 0.03 \times 10^{-5} = 3 \times 10^{-7}$ but R.E. $= 0.09091$. Note that the relative error is unchanged, while the absolute error changed by a factor of 10^5. Several conclusions can be drawn:

- The absolute error is strongly dependent on the magnitude of x.
- The absolute error is misleading unless it is stated what it is an error of.
- The relative error is a measure of the number of significant digits of x that are correct.
- A relative error has meaning even when x is not known. It is given as a percentage value.

1.2 Truncation error

In numerical analysis and scientific computing, truncation error is an error caused by approximating a mathematical process.

Example: Truncation error can cause $(A + B) + C \neq A + (B + C)$ within a computer when $A = -10^{25}, B = 10^{25}, C = 1$ because $(A+B)+C = (0)+C = 1$ (like it should), while $A+(B+C) = A + (B) = 0$. Here, $A + (B + C)$ has a truncation error equal to 1. This truncation error occurs because computers do not store the least significant digits of an extremely large integer.

1.3 Round-off error

When a calculator or digital computer is used to perform numerical calculations, an unavoidable error, called round-off error, must be considered. This error arises because the arithmetic performed in a machine involves numbers with only a finite number of digits, with the result that many calculations are performed with approximate representations of the actual numbers. There are two major approaches: chopping and rounding.

- When chopping a number to a specified number of decimal places, say m, the first m digits of the mantissa are retained, simply chopping off the remainder.
- When rounding a number, the computer chooses the closest number that is representable by the computer.

1.4 Error Propagation

Assume the operation

$$x \oplus y = (1 + \delta)(x + y), \tag{3}$$

Here, δ is the relative error of this operation and $\oplus$ be any arithmetic operation. Errors propagate across multiple operations. For example, the addition of three numbers:

$$(x \oplus y) \oplus z = f\{x(1 + \delta) + y(1 + \delta)) + z(1 + \delta)\} = (1 + 3\delta)(x + y + z) \tag{4}$$

We have assumed that all the relative errors are equal. The result is that the addition of three numbers gives a maximum relative error of three times the relative error of one of the input numbers. In reality, different signs for the error lead to error cancellations, so that the results are not so detrimental.

Note: Suppose we want to compute $f(x)$ where x is a real number and f is a real function. In practical computations, the number x must be approximated by a rational number r since no computer can store numbers with an infinite number of decimals. The difference $|r - x|$ constitutes the initial error while the difference $\epsilon_0 = |f(r) - f(x)|$ is the corresponding propagated error. In many cases, f is such a function that it must be replaced by a simpler function f_1 (often a truncated power series expansion of f).

The difference $\epsilon_1 = |f_1(r) - f(r)|$ is then the *truncation error*. The calculations performed by the computer, however, are not exact but pseudo-operations of a type discussed earlier. The result is that instead of getting $f_1(r)$ we get another value $f_2(r)$ which is then a wrongly computed value of a wrong function of a wrong argument.

The difference $\epsilon_2 = |f_2(r) - f_1(r)|$ could be termed the propagated error from the roundings. The total error is $\epsilon = \epsilon_0 + \epsilon_1 + \epsilon_2$.

2 Statistical methods in data analysis

Statistical analysis is the process of collecting and analyzing data in order to discern patterns and trends. It is a method for removing bias from evaluating data by employing numerical analysis. This technique is useful for collecting the interpretations of research, developing statistical models, and planning surveys and studies.

Statistical analysis is a scientific tool in Artificial Intelligence (AI) and Machine Learning (ML) that helps collect and analyze large amounts of data to identify common patterns and trends to convert them into meaningful information. In simple words, statistical analysis is a data analysis tool that helps draw meaningful conclusions from raw and unstructured data. Given below are the 6 types of statistical analysis:

- **Descriptive Analysis:** Descriptive statistical analysis involves collecting, interpreting, analyzing, and summarizing data to present them in the form of charts, graphs, and tables. Rather than drawing conclusions, it simply makes the complex data easy to read and understand.
- **Inferential Analysis:** The inferential statistical analysis focuses on drawing meaningful conclusions on the basis of the data analyzed. It studies the relationship between different variables or makes predictions for the whole population.
- **Predictive Analysis:** Predictive statistical analysis is a type of statistical analysis that analyzes data to derive past trends and predict future events on the basis of them. It uses machine learning algorithms, data mining, data modelling, and artificial intelligence to conduct the statistical analysis of data.
- **Prescriptive Analysis:** The prescriptive analysis conducts the analysis of data and prescribes the best course of action based on the results. It is a type of statistical analysis that helps you make an informed decision.
- **Exploratory Data Analysis:** Exploratory analysis is similar to inferential analysis, but the difference is that it involves exploring the unknown data associations. It analyzes the potential relationships within the data.
- **Causal Analysis:** The causal statistical analysis focuses on determining the cause and effect relationship between different variables within the raw data. In simple words, it determines why something happens and its effect on other variables. This methodology can be used by businesses to determine the reason for failure.

2.1 Importance of Statistical Analysis

Statistical analysis eliminates unnecessary information and catalogs important data in an uncomplicated manner, making the monumental work of organizing inputs appear so serene. Once the data has been collected, statistical analysis may be utilized for a variety of purposes. Some of them are listed below:

- The statistical analysis aids in summarizing enormous amounts of data into clearly digestible chunks.
- The statistical analysis aids in the effective design of laboratory, field, and survey investigations. Statistical analysis may help with solid and efficient planning in any subject of study.
- Statistical analysis aid in establishing broad generalizations and forecasting how much of something will occur under particular conditions.
- Statistical methods, which are effective tools for interpreting numerical data, are applied in practically every field of study. Statistical approaches have been created and are increasingly applied in physical and biological sciences, such as genetics.
- Statistical approaches are used in the job of a businessman, a manufacturer, and a researcher. Statistics departments can be found in banks, insurance businesses, and government agencies.
- A modern administrator, whether in the public or commercial sector, relies on statistical data to make correct decisions.
- Politicians can utilize statistics to support and validate their claims while also explaining the issues they address.

2.2 Statistical Analysis Process

Given below are the 5 steps to conduct a statistical analysis that you should follow:

- Step 1: Identify and describe the nature of the data that you are supposed to analyze.
- Step 2: The next step is to establish a relation between the data analyzed and the sample population to which the data belongs.
- Step 3: The third step is to create a model that clearly presents and summarizes the relationship between the population and the data.
- Step 4: Prove if the model is valid or not.
- Step 5: Use predictive analysis to predict future trends and events likely to happen.

2.3 Statistical Analysis Methods

Although there are various methods used to perform data analysis, given below are the 5 most used and popular methods of statistical analysis:

- **Mean:** Mean or average mean is one of the most popular methods of statistical analysis. Mean determines the overall trend of the data and is very simple to calculate. Mean is calculated by summing the numbers in the data set together and then dividing it by the number of data points. Despite the ease of calculation and its benefits, it is not advisable to resort to mean as the only statistical indicator as it can result in inaccurate decision making.
- **Standard Deviation:** Standard deviation is another very widely used statistical tool or method. It analyzes the deviation of different data points from the mean of the entire data set. It determines how data of the data set is spread around the mean. You can use it to decide whether the research outcomes can be generalized or not.
- **Regression:** Regression is a statistical tool that helps determine the cause and effect relationship between the variables. It determines the relationship between a dependent and an independent variable. It is generally used to predict future trends and events.
- **Hypothesis Testing:** Hypothesis testing can be used to test the validity or trueness of a conclusion or argument against a data set. The hypothesis is an assumption made at the beginning of the research and can hold or be false based on the analysis results.
- **Sample Size Determination:** Sample size determination or data sampling is a technique used to derive a sample from the entire population, which is representative of the population. This method is used when the size of the population is very large. You can choose from among the various data sampling techniques such as snowball sampling, convenience sampling, and random sampling.

3 Estimation and reporting of measurement uncertainty

Regardless of method, repeated measurements on the same sample will generally produce different results if the system is sufficiently sensitive. The dispersion of results obtained from such repeated measurements (imprecision) can be described approximately by a normal probability (Gaussian) distribution, with some 95% of the results falling within ± 2 standard deviations (SD) of the mean value. Measurement results are thus unreliable and should be regarded as best estimates of the true value of the quantities being measured. Some knowledge of the result variability expected from a given measurement procedure is required if results are to be meaningfully compared with other results of the same kind or with decision and legal limits.

In the 1990s it was recognized that measurement comparability between laboratories and methods required an internationally agreed approach to estimating and expressing measurement uncertainty, which is described in the 'Guide to the Expression of Uncertainty in Measurement' (GUM). Although GUM is primarily for measurements in physics, the principles are applicable to biological and chemical measurements. The terminology of the science of measurement (Metrology) is defined in the 'International Vocabulary of Basic and General Terms in Metrology' (VIM).

The basic parameter of MU is the SD, and the symbol for uncertainty is u. Some definitions to be remembered:

- Coefficient of variation (CV); also termed relative standard measurement uncertainty $\implies$ Standard measurement uncertainty (SD) divided by the absolute value of the measured quantity value. $CV = SD/x$ or $SD/mean\ value$.
- Combined standard measurement uncertainty (u_c) $\implies$ Standard measurement uncertainty that is obtained using the individual standard measurement uncertainties associated with the input quantities in a measurement model.
- Coverage factor (k) $\implies$ Number larger than one by which a combined standard measurement uncertainty is multiplied to obtain an expanded measurement uncertainty.
- Coverage interval $\implies$ Interval containing the set of true values of a measurand with a stated probability, based on the information available.
- Expanded measurement uncertainty (U) $\implies$ Product of a combined standard measurement uncertainty and a coverage factor larger than the number one.
- Measurand $\implies$ Quantity intended to be measured.
- Measurement uncertainty $\implies$ Non-negative parameter characterising the dispersion of the quantity values being attributed to a measurand, based on the information used.
- Quantity $\implies$ Property of a phenomenon, body, or substance, where the property has a magnitude that can be expressed as a number and a reference.
- Standard measurement uncertainty (u) $\implies$ Measurement uncertainty expressed as a standard deviation.
- Trueness $\implies$ Closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value.
- True value $\implies$ Quantity value consistent with the definition of a quantity.