

NE 155, S21
Probability and Statistics MC board work

Probability Distributions

We need to figure out all of these things about how particles are moving around in our system, but how do we do it?

We get functional expressions of the probability that various things will happen and try to take enough samples to effectively capture those expressions.

- For a random variable, x , the probability that x will have a value between a and b is $P\{a \leq x \leq b\}$.
- The *probability density function* expresses the likelihood that x' will take on a value between x and $x + \Delta x$:

$$\lim_{\Delta x \rightarrow 0} f(x) \Delta x = P\{x \leq x' \leq x + \Delta x\}$$
$$\int_a^b f(x) dx = P\{a \leq x \leq b\}$$

- We often normalize this PDF to integrate to one, using one of

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad \text{or} \quad \int_{x^-}^{x^+} f(x) dx = 1$$

- To get the probability that our random variable x' is less than or equal to some value x , we use a *cumulative distribution function*:

$$F(x) = P\{x' \leq x\}$$
$$F(x) = \int_{-\infty}^x f(x') dx'$$
$$\lim_{x \rightarrow \infty} F(x) \equiv F(\infty) = 1$$
$$\lim_{x \rightarrow -\infty} F(x) \equiv F(-\infty) = 0$$
$$P\{a \leq x' \leq b\} = F(b) - F(a)$$

Various physical phenomena can be represented by probability distributions, e.g.

- Photon emission energy: Each possible energy has a different probability (intensity)
- Scattering cross-sections: Each possible scattering angle has a different probability as a function of the energy
- Transmission through a medium: Probability of reaching a particular position depends on the cross-section

We in one way or another get these PDFs and/or CDFs and use random numbers to select values for use in simulation.

Statistics

The “true” mean value, μ , of any PDF is the expected value, $E(x)$

$$\mu = E(x) = \int x f(x) dx$$

Because we can’t usually do this, we use random samples and estimate the true mean from the “sample” mean, \bar{x}

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \lim_{N \rightarrow \infty} \bar{x} \rightarrow \mu .$$

The variance of a PDF is the measure of spread in that PDF. If we knew the PDF, we could directly say:

$$\begin{aligned} \sigma^2 &= E[(x - \mu)^2] = \int (x - \mu)^2 f(x) dx \\ &= \int x^2 f(x) dx - 2\mu \int x f(x) dx + \mu^2 \int f(x) dx \\ &= E(x^2) - \mu^2 \end{aligned}$$

However, we don’t know the PDF so we use the samples to get the sample variances (how much the samples vary from the true mean)

$$\begin{aligned} S_x^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \\ &= \frac{1}{N-1} \left[\sum_{i=1}^N x_i^2 - 2\bar{x} \sum_{i=1}^N x_i + \bar{x}^2 \sum_{i=1}^N 1 \right] \\ &\approx \bar{x}^2 - \bar{x}^2 \end{aligned}$$

One step beyond the sample variances gives us the variance of the sample mean (how much the sample mean varies from the true mean).

$$\begin{aligned} S_{\bar{x}}^2 &= E[(\bar{x} - \mu)^2] = E \left[\left(\frac{1}{N} \sum_{i=1}^N x_i - \mu \right)^2 \right] \\ &= \frac{1}{N^2} E \left[\sum_{i=1}^N (x_i - \mu) \sum_{j=1}^N (x_j - \mu) \right] = \frac{1}{N^2} E \left[\sum_{i=1}^N \sum_{j=1}^N (x_i - \mu)(x_j - \mu) \right] \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[(x_i - \mu)(x_j - \mu)] = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N S_x^2 \delta_{ij} = \frac{1}{N^2} \sum_{i=1}^N S_x^2 \\ &= \frac{N S_x^2}{N^2} = \boxed{\frac{S_x^2}{N}} \end{aligned}$$

The error in the results decreases with the square of increasing the number of histories.

We can apply the variance through the standard deviation (the square root of the variance) to express the confidence we have in our answer.

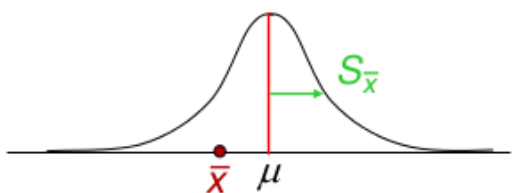
The Central Limit Theorem states that For N *independent* random variables, x_i , sampled from *identical distributions*, their mean follows a Normal (Gaussian) distribution.

(Note: this is the *IID* requirement for MC)

We can use this information to define *confidence intervals*

$$\begin{aligned}\bar{x} - S_{\bar{x}} < E(x) < \bar{x} + S_{\bar{x}} & \text{ about 68\% of the time} \\ \bar{x} - 2S_{\bar{x}} < E(x) < \bar{x} + 2S_{\bar{x}} & \text{ about 95\% of the time}\end{aligned}$$

The **standard deviation** of the mean is a measure of the error in the result.



Relative Error is

$$R = \frac{S_{\bar{x}}}{\bar{x}} = \sqrt{\frac{\sum_{i=1}^N x_i^2}{(\sum_{i=1}^N x_i)^2} - \frac{1}{N}}$$

If x_i are equal and non-zero, $R=0$.

Thus, we can reduce the error by reducing the spread in x_i .

Accuracy vs. Precision

The distinction between accuracy and precision can be seen in Fig. 1.

Accuracy is the degree of closeness of measurements of a quantity to that quantity's true value.

The *precision* of a measurement system, related to reproducibility and repeatability, is the degree to which repeated measurements under unchanged conditions show the same results.

Accuracy can be affected by systematic errors in simulation: physical and mathematical models; errors in geometry or source model; incorrect code use by user.

Usually unknown.

Conversely, precision can usually be improved: run more histories; use variance reduction; adjust your measurement (fewer scoring bins).

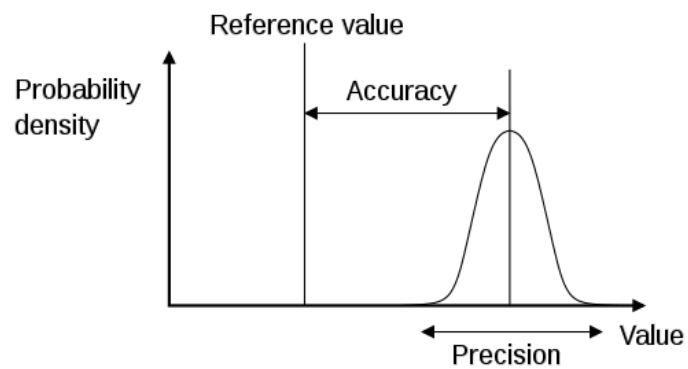


Figure 1: "Accuracy and precision" by Pekaje at English Wikipedia - Transferred from en.wikipedia to Commons.. Licensed under GFDL via Commons - https://commons.wikimedia.org/wiki/File:Accuracy_and_precision.svg#/media/File:Accuracy_and_precision.svg