Lecture 1
Probability and Statistics

Wikipedia:

- Benjamin Disraeli, British statesman and literary figure (1804 – 1881):
  - ★ There are three kinds of lies: lies, damned lies, and statistics.
    - ◆ popularized in US by Mark Twain
    - ◆ the statement shows the persuasive power of numbers
      - ☞ use of statistics to bolster weak arguments
      - ☞ tendency of people to disparage statistics that do not support their positions

- The purpose of P3700:
  - ★ how to understand the statistical uncertainty of observation/measurement
  - ★ how to use statistics to argue against a weak argument (or bolster a weak argument?)
  - ★ how to argue against people disparaging statistics that do not support their positions
  - ★ how to lie with statistics?

Why there is statistical uncertainty?

- You sell 7 cryogenic equipment last month
  - ★ You know how to count and 7 is the exact number of equipment sold
    - ☞ there is no uncertainty on 7!
  - ★ However if you used the statistics of 7 to predict the future sale or compare with past sale
    - ☞ there is an uncertainty on "7"
    - ☞ the number of equipment sold could be 5, 8, or 10!
    - ☞ must include the uncertainty in the calculation
    - ◆ What is the uncertainty on "7"?
      - ☞ Lecture 2: $\sqrt{7} = 2.6$
      - ☞ there is a 68% chance that the expected number of equipment sold per month is 4.4-9.6
    - ◆ However the number of equipment sold per month is a discrete number
      - ☞ there is a ~68% chance that the expected number of equipment sold per month is 4-10
      - ☞ should use Poisson statistics as in Lecture 2 for more precise prediction

Introduction:

- Understanding of many physical phenomena depend on statistical and probabilistic concepts:
  - ★ Statistical Mechanics (physics of systems composed of many parts: gases, liquids, solids.)
    - ◆ 1 mole of anything contains 6×10^{23} particles (Avogadro's number)
    - ◆ impossible to keep track of all 6×10^{23} particles even with the fastest computer imaginable
      - ☞ resort to learning about the group properties of all the particles
      - ☞ partition function: calculate energy, entropy, pressure... of a system
  - ★ Quantum Mechanics (physics at the atomic or smaller scale)
    - ◆ wavefunction = probability amplitude
      - ☞ probability of an electron being located at (x,y,z) at a certain time.

- Understanding/interpretation of experimental data depend on statistical and probabilistic concepts:
  - ★ how do we extract the best value of a quantity from a set of measurements?
  - ★ how do we decide if our experiment is consistent/inconsistent with a given theory?
  - ★ how do we decide if our experiment is internally consistent?
  - ★ how do we decide if our experiment is consistent with other experiments?
    - ☞ In this course we will concentrate on the above experimental issues!

Definition of probability:

- Suppose we have N trials and a specified event occurs r times.
    - example: rolling a dice and the event could be rolling a 6.
  - define probability (P) of an event (E) occurring as:

    $P(E) = r/N$ when $N \to \infty$
    - examples:
      - six sided dice: $P(6) = 1/6$
      - coin toss: $P(\text{heads}) = 0.5$
        - P(heads) should approach 0.5 the more times you toss the coin.
        - for a single coin toss we can never get $P(\text{heads}) = 0.5$!
  - by definition probability is a non-negative real number bounded by $0 \leq P \leq 1$
    - if $P = 0$ then the event never occurs
    - if $P = 1$ then the event always occurs
    - sum (or integral) of all probabilities if they are mutually exclusive must = 1.
      - events are independent if: $P(A \cap B) = P(A)P(B)$

        $\cap \equiv$ intersection, $\cup \equiv$ union
        - coin tosses are independent events, the result of next toss does not depend on previous toss.
      - events are mutually exclusive (disjoint) if: $P(A \cap B) = 0$ or $P(A \cup B) = P(A) + P(B)$
        - in coin tossing, we either get a head or a tail.

- Probability can be a discrete or a continuous variable.
  - Discrete probability: P can have certain values only.
    - examples:
      - tossing a six-sided dice: $P(x_i) = P_i$ here $x_i = 1, 2, 3, 4, 5, 6$ and $P_i = 1/6$ for all x_i.
      - tossing a coin: only 2 choices, heads or tails.
    - for both of the above discrete examples (and in general) when we sum over all mutually exclusive possibilities:

$$\sum_i P(x_i) = 1$$

  - Continuous probability: P can be any number between 0 and 1.
    - define a "probability density function", pdf, $f(x)$

$$f(x)dx = dP(x \le \alpha \le x + dx) \quad \text{with } \alpha \text{ a continuous variable}$$

    - probability for x to be in the range $a \le x \le b$ is:

$$P(a \le x \le b) = \int_a^b f(x)dx$$

    - just like the discrete case the sum of all probabilities must equal 1.

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

      - ☞ $f(x)$ is normalized to one.
    - probability for x to be exactly some number is zero since:

$$\int_{x=a}^{x=a} f(x)dx = 0$$

Notation:
x_i is called a random variable

- Examples of some common $P(x)$' s and $f(x)$' s:

Discrete = $P(x)$	Continuous = $f(x)$
binomial	uniform, i.e. constant
Poisson	Gaussian
	exponential
	chi square

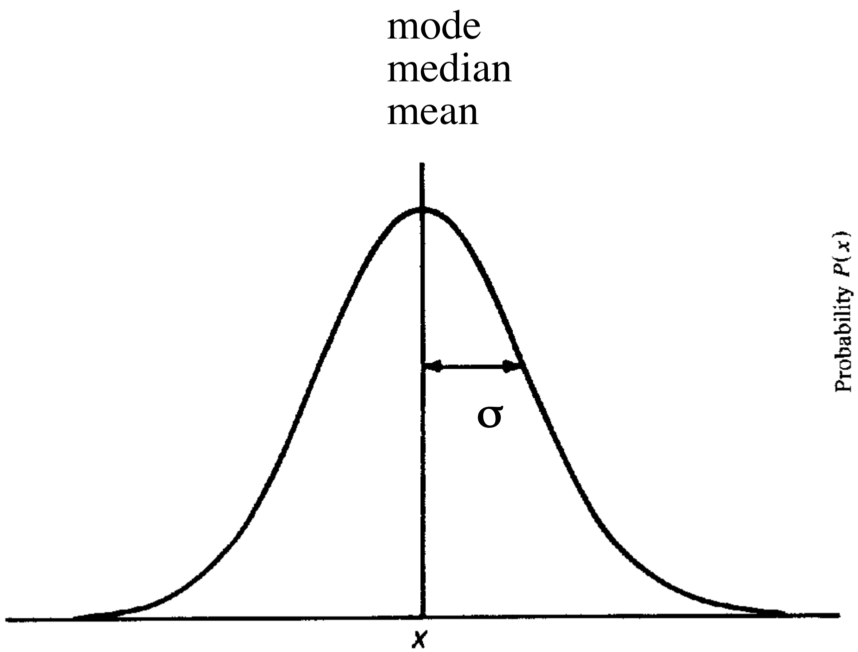
- How do we describe a probability distribution?
  - mean, mode, median, and variance
  - for a continuous distribution, these quantities are defined by:

Mean	Mode	Median	Variance
average	most probable	50% point	width of distribution
$\mu = \int_{-\infty}^{+\infty} x f(x)dx$	$\left.\frac{\partial f(x)}{\partial x}\right\vert_{x=a} = 0$	$0.5 = \int_{-\infty}^{a} f(x)dx$	$\sigma^2 = \int_{-\infty}^{+\infty} f(x)(x-\mu)^2 dx$

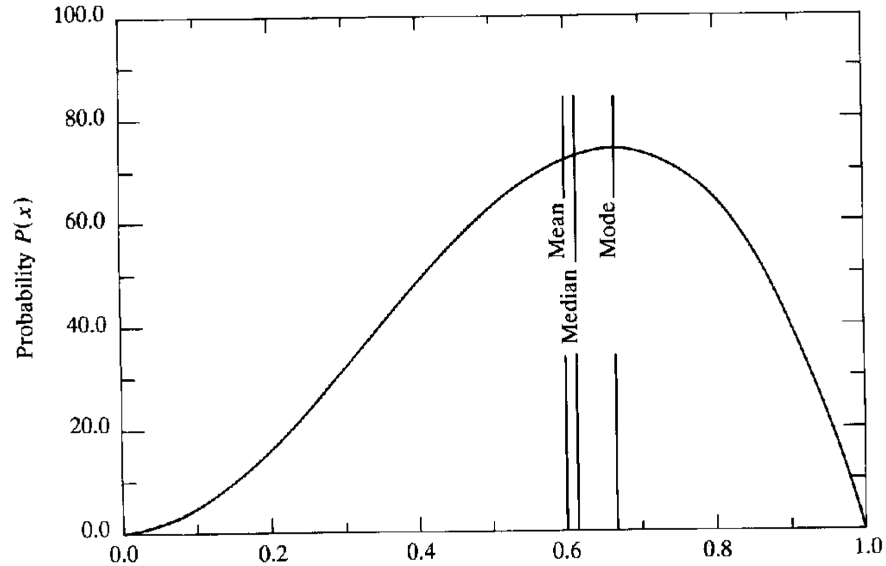
  - for a discrete distribution, the mean and variance are defined by:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

- Some continuous *pdf*:
  - Probability is the area under the curves!



symmetric distribution (gaussian)



Asymmetric distribution showing the mean, median and mode

For a Gaussian pdf, the mean, mode, and median are all at the same x.

For most pdfs, the mean, mode, and median are at different locations.

- Calculation of mean and variance:
  - example: a discrete data set consisting of three numbers: $\{1, 2, 3\}$
    - average (μ) is just:

$$\mu = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{1+2+3}{3} = 2$$

    - complication: suppose some measurement are more precise than others.
      - if each measurement x_i have a weight w_i associated with it:

$$\mu = \sum_{i=1}^{n} x_i w_i / \sum_{i=1}^{n} w_i$$

"weighted average"

    - variance (σ^2) or average squared deviation from the mean is just:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

variance describes the width of the pdf!

      - σ is called the standard deviation
      - rewrite the above expression by expanding the summations:

$$\begin{aligned}\sigma^2 &= \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \mu^2 - 2\mu \sum_{i=1}^{n} x_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 + \mu^2 - 2\mu^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 \\ &= \left\langle x^2 \right\rangle - \left\langle x \right\rangle^2\end{aligned}$$

$< > \equiv$ average

      - n in the denominator would be n -1 if we determined the average (μ) from the data itself.

★ using the definition of μ from above we have for our example of {1,2,3}:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 = 4.67 - 2^2 = 0.67$$

★ the case where the measurements have different weights is more complicated:

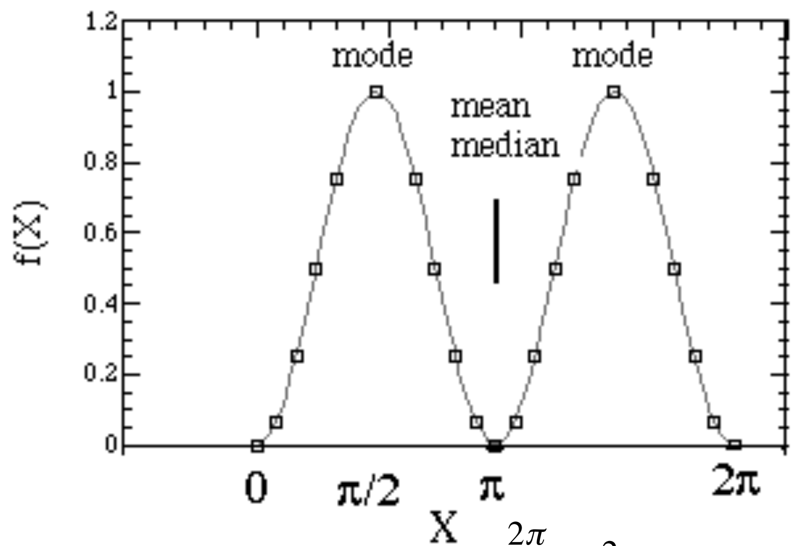
$$\sigma^2 = \sum_{i=1}^{n} w_i (x_i - \mu)^2 / \sum_{i=1}^{n} w_i = \sum_{i=1}^{n} w_i x_i^2 / \sum_{i=1}^{n} w_i - \mu^2$$

- μ is the weighted mean
- if we calculated μ from the data, σ^2 gets multiplied by a factor $n/(n-1)$.

◆ example: a continuous probability distribution, $f(x) = \sin^2 x$ for $0 \le x \le 2\pi$

★ has two modes!

★ has same mean and median, but differ from the mode(s).



★ $f(x)$ is not properly normalized: $\int_0^{2\pi} \sin^2 x dx = \pi \neq 1$

☞ normalized pdf: $f(x) = \sin^2 x / \int_0^{2\pi} \sin^2 x dx = \frac{1}{\pi}\sin^2 x$

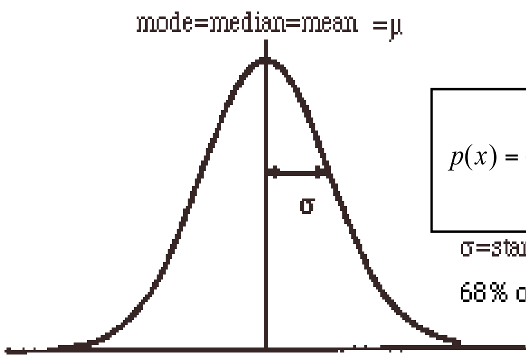
- for continuous probability distributions, the mean, mode, and median are calculated using either integrals or derivatives:

$$\mu = \frac{1}{\pi}\int_0^{2\pi} x \sin^2 x dx = \pi$$

$$\text{mode}: \frac{\partial}{\partial x}\sin^2 x = 0 \Rightarrow \frac{\pi}{2}, \frac{3\pi}{2}$$

$$\text{median}: \frac{1}{\pi}\int_0^{\alpha} \sin^2 x dx = \frac{1}{2} \Rightarrow \alpha = \pi$$

- example: Gaussian distribution function, a continuous probability distribution



$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \textit{gaussian}$$

σ=standard deviation

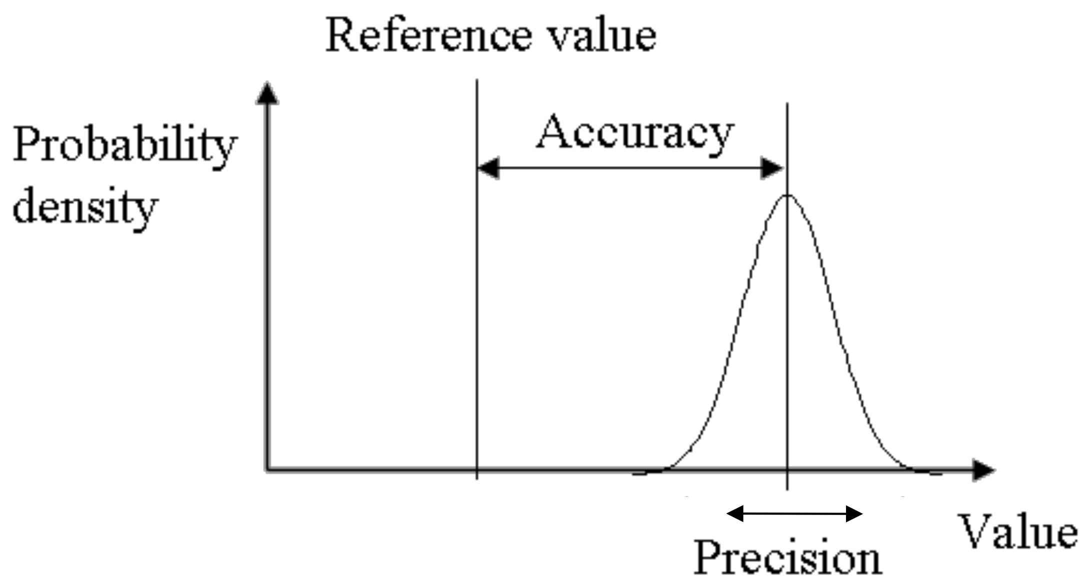
68% of area within ±1σ

Accuracy and Precision:

- Accuracy: The accuracy of an experiment refers to how close the experimental measurement is to the true value of the quantity being measured.
- Precision: This refers to how well the experimental result has been determined, without regard to the true value of the quantity being measured.
  - just because an experiment is precise it does not mean it is accurate!!

accurate but not precise

precise but not accurate



Measurement Errors (Uncertainties)

- Use results from probability and statistics as a way of indicating how "good" a measurement is.
  - most common quality indicator:
    relative precision = [uncertainty of measurement]/measurement
    - example: we measure a table to be 10 inches with uncertainty of 1 inch.
      relative precision = 1/10 = 0.1 or 10% (% relative precision)
  - uncertainty in measurement is usually square root of variance:
    σ = standard deviation
    - usually calculated using the technique of "propagation of errors" (Lecture 4).

Statistics and Systematic Errors

- Results from experiments are often presented as:

  $N \pm XX \pm YY$

N: value of quantity measured (or determined) by experiment.

XX: statistical error, usually assumed to be from a Gaussian distribution.

- with the assumption of Gaussian statistics we can say (calculate) something about how well our experiment agrees with other experiments and/or theories.
  - Expect an 68% chance that the true value is between N - XX and $N + XX$.

YY: systematic error. Hard to estimate, distribution of errors usually not known.

- examples: mass of proton = 0.9382769 $\pm$ 0.0000027 GeV (only statistical error given)
  mass of W boson = 80.8 $\pm$ 1.5 $\pm$ 2.4 GeV

- What's the difference between statistical and systematic errors?

  $N \pm XX \pm YY$

  - statistical errors are "random" in the sense that if we repeat the measurement enough times:

    XX -> 0

  - systematic errors do not -> 0 with repetition.
    - examples of sources of systematic errors:
      - voltmeter not calibrated properly
      - a ruler not the length we think is (meter stick might really be < meter!)
  - because of systematic errors, an experimental result can be precise, but not accurate!
- How do we combine systematic and statistical errors to get one estimate of precision?
  - big problem!
  - two choices:
    - $\sigma_{\text{tot}} = XX + YY$ add them linearly
    - $\sigma_{\text{tot}} = (XX^2 + YY^2)^{1/2}$ add them in quadrature
      - widely accepted practice if XX and YY are not correlated
        - errors not of same origin, e.g. from the same voltmeter
      - ☺ smaller σ_{tot}!
- Some other ways of quoting experimental results
  - lower limit: "the mass of particle X is > 100 GeV"
  - upper limit: "the mass of particle X is < 100 GeV"
  - asymmetric errors: mass of particle $X = 100^{+4}_{-3}$ GeV

How to present your measured values:

- Don't quote any measurement to more than three significant digits
  - three significant digits means you measure a quantity to 1 part in a thousand or 0.1% precision
  - difficult to achieve 0.1% precision
  - acceptable to quote more than three significant digits if you have a large data sample (e.g. large simulations)
- Don't quote any uncertainty to more than two significant digits
- Measurement and uncertainty should have the same number of digits
  - 991 ± 57
  - 0.231 ± 0.013
  - $(5.98 \pm 0.43) \text{x} 10^{-5}$
- follow this rule in the lab report!