Phys 443
Computation Physics I
Probability&Statistics
Experimental Uncertainty

Probability and Statistics

I) The understanding of many physical phenomena relies 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)

Even though the force between particles (Newton's laws) is known it is impossible to keep track of all 6×10^{23} particles even with the fastest computer imaginable.

We must resort to learning about the group properties of all the particles:

use the partition function: calculate average energy, entropy, pressure... of a system

Quantum Mechanics (physics at the atomic or smaller scale, $< 10^{-10}$m)

wavefunction = probability amplitude

talk about the probability of an electron being located at (x,y,z) at a certain time.

II) Our understanding/interpretation of experimental data relies 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?

Statistics

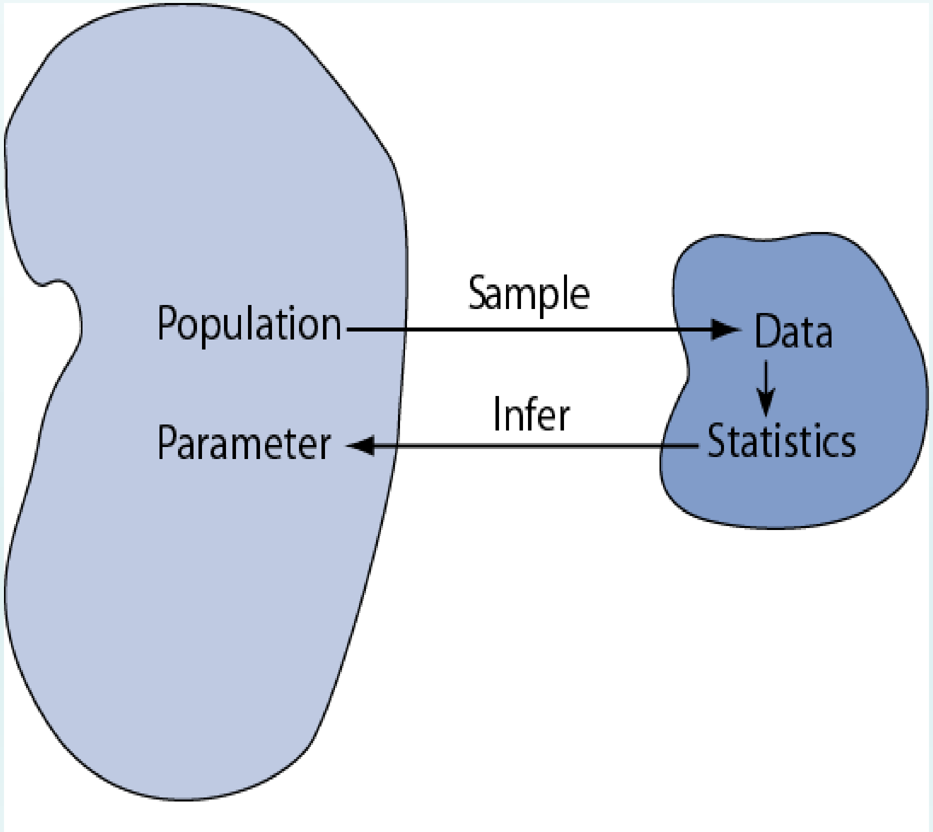
- Type of analysis which includes the planning, summarizing and interpreting of observations of a system followed by predicting or forecasting future events.

- **Descriptive Statistics:** Describe or summarize observed measurements of a system
- **Inferential Statistics:** Infer, predict, or forecast future outcomes, tendencies, behaviors of a system

Inferential Statistics

Statistical inference is the act of generalizing from a **sample** to a **population** with calculated degree of certainty.

We want to learn about population *parameters* …



…but we can only calculate *sample statistics*

Questions Answered by Statistics!

- How much energy will a 1-MeV proton lose in its next collision with an atomic electron?
- Will a 400-keV photon penetrate a 2-mm lead shield without interacting?
- How many disintegrations will occur during the next minute with a given radioactive source?

- **Repeated measurements result in a spread of values. How certain, then, is a measurement?**

- **'Uncertainties' in scatter, photon penetration, and decay are inherent due to quantum physics that interprets such events as probabilistic.**

Probability and Statistics

- The theory of probability is a branch of pure mathematics. From a certain set of axioms and definitions one builds up the theory by deductions..

- Suppose it is know that when tossing a coin it has an a priori probability p(*p=1/2*) of landing "heads" and a probability *1-p* of landing "tails". We ask: What is the probability of observing r heads out of n tossing. <u>This is a question in probability theory, and an answer is provided by the binomial distribution law, which states that the probability to obtain r heads ad *(n-r)* tails is given by the number</u>

$$P(r) = \frac{n!}{r!(n-r)!}\ p^r \cdot q^{n-r}$$

Probability and Statistics

- **A completely different situation exists if one has no a priori knowledge on the probability p and decides to perform an experiment to "determine" this parameter**. A simple experiment would consist in tossing the coin repeatedly and counting how many times the outcome heads would occur. It would then be a question of statistics to ask what the parameter p is like, given that in n tosses, r heads were observed. A reasonable answer to this question is to say that "the most likely value" of the parameter is given by the ratio of the number of heads observed to the total number of tosses,

$$p = \hat{p} = \frac{r}{n}$$

- The experiment has then given a *point estimate* p=$\hat{p}$ for the unknown parameter .
- we could also say that the value of p was inferred from the observations made by the experiment. However, from the very mature of the experiment, we could not be completely sure that the value $\hat{p}$ thus obtained would be identical to the true value of the parameter p. Indeed we would feel that if the experiment were repeated, with new sequences of tosses, then presumably different estimates $\hat{p}$ would be obtained.

Probability and Statistics

- Instead of stating the result of the experiment in terms of a single number $\hat{p}$ we could give an interval estimate for the parameter p. We would then finalize our experiment by giving two numbers p_1,p_2, hoping that

$$p_1 \leq p \leq p_2$$

Represents a true statement about p. The faith we attach to the statment could be expressed by assigning a *confidence level*.

Given the observation of r heads in n tosses it is again a case of statistical inference to determine an interval $[p_1,p_2]$ which is such that it has a certain probability of including the true value of p. In general, the larger we take the interval the more certain we would be that this interval really includes the true value of p. but at the same time a large interval means less precise knowledge on p. On the other hand, a small interval corresponds to a better precision in the determination of p, but the statement that $[p_1,p_2]$ includes p then has a greater chance of not being true.

What is Probability?

There is not universal agreement on what probability means, and this does affect you. Three suggested interpretations:

- Propensity. Ex.: an unbiased coin has a “propensity” 1/2 to land heads or tails. i.e., 1/2 is a property of the coin.
- Degree of belief. Ex.: your belief that a coin is unbiased means you assign probability 1/2 to the proposition “the coin will land heads.”
- Relative frequency. Ex.: the relative frequency with which heads appears in a sequence of infinite tosses of an unbiased coin is 1/2.

Questions:

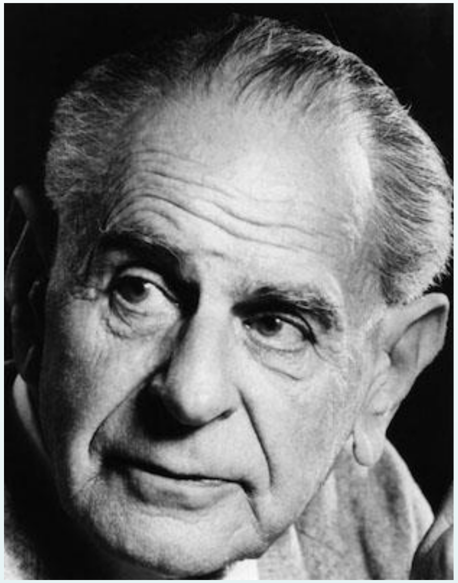
- Which of these seems most reasonable to you?
- Is one more “fundamental” than the others?
- How can scientists compare independent measurements and draw conclusions?

What is Probability?

Propensity

Q: why does a given kind of experiment generate a given outcome at a persistent rate?

- A: probability is physical property, like particle decay in Quantum Mechanics.
- Problem: we don’t observe the intrinsic probability, just the outcome.
- Problem: how to quantify ignorance?

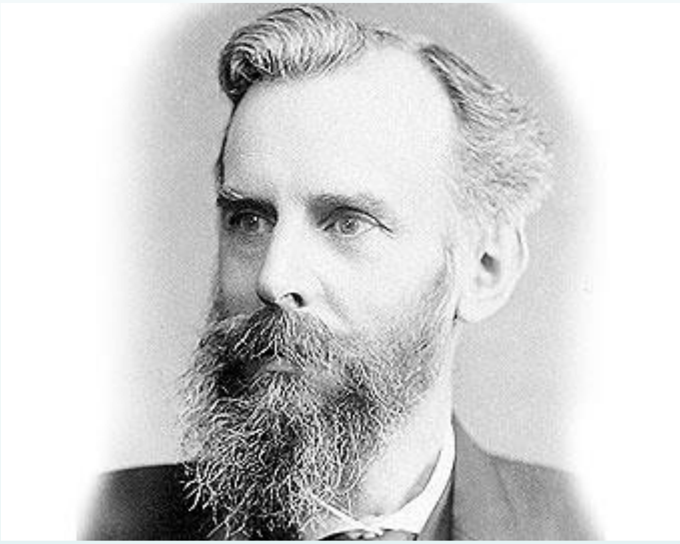


Karl Popper:“In so far as a scientific statement speaks about reality, it must be falsifiable; and in so far as it is not falsifiable, it does not speak about reality .”

What is Probability?

Relative Frequency, or “Frequentist Statistics”

- Probability & chance come from ignorance of physical causes, not intrinsic physical properties.
- We can’t or don’t usually need to care about causes, just the “objective” consistency of long-run outcomes.
- Appeal: we observe outcomes and relative frequencies all the time in controlled experiments!
- Problem: what about non-repeatable phenomena. Invoke “ensembles” of hypothetical identical systems.
- Problem: how to “objectively” identify an ensemble?



John Venn:“On the view here adopted we are concerned only with averages, or with the single event as deduced from an average and conceived to form one of a series”

What is Probability?

Degree of Belief, or “Bayesian Statistics”

Probability quantifies our knowledge or ignorance of a situation given “data” plus our a **priori beliefs**.

- Degree of belief is always subjective, but can/should be revised as we learn (“acquire new data”).
- **Appeal**: this is common sense, and describes scientific inference very well.
- Problem: how do we quantify our a priori beliefs, i.e., our beliefs before taking any data?
- Problem: if probability is subjective, how can we achieve consensus?

Laplace:“Probability theory is only common sense reduced to calculation.”

How do we define Probability?

Statisticisans do not seem to agree about the best way to define probability. A rather simple approach, among physicists, and define probability in terms of the limit of relative frequency of occurence of specified class, or the event A, occur r times, then the probability of A is defined as the limiting value

How do we define Probability?

- **Laplace's Classical Definition:** The Probability of an event A is defined a-priori without actual experimentation as

$$P(A) = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of possible outcomes}}$$

provided all these outcomes are *equally likely*.

- **Relative Frequency Definition**: The probability of an event A is defined as

$$P(A) = \lim_{n \to \infty} \frac{n_A}{n}$$

where n_A is the number of occurrences of A and n is the total number of trials

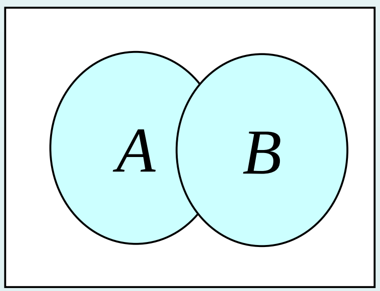
From this definition the probability of the event E is some number satisfying
$0 \leq P(A) \leq 1$

Probability

- The axiomatic approach to probability, due to Kolmogorov developed through a set of axioms
- For any Experiment *E*, has a set *S or* Ω of all possible outcomes called sample space, Ω.

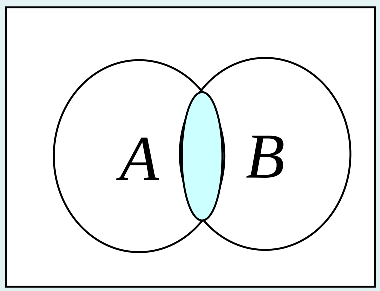
- Ω has subsets *{A, B, C, ...}* called events. If $A \cap B = \phi$, the empty set, then *A* and *B* are said to be **mutually exclusive events**.

Venn diagrams



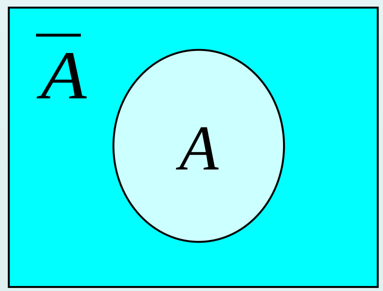
$A \cup B$

union



$A \cap B$

intersection



$\overline{A}$

complement

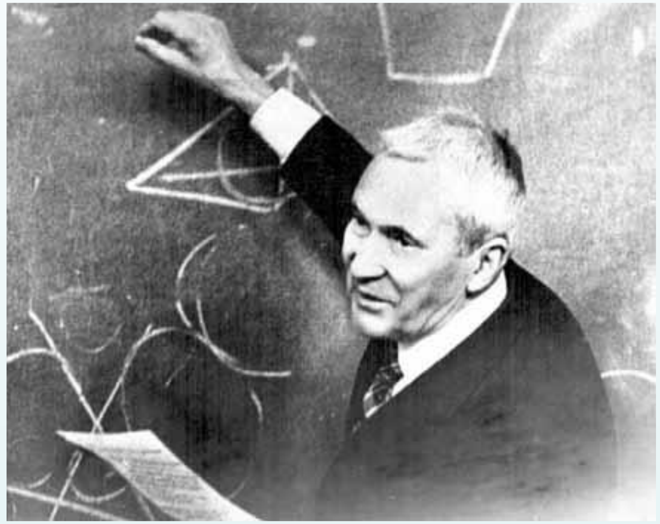
Probability: Axioms of Probability

- For any event A, we assign a number $P(A)$, called the probability of the event A. This number satisfies the following three conditions that act the axioms of probability.

(i) $P(A) \geq 0$ (Probability is a nonnegative number)

(ii) $P(\Omega) = 1$ (Probability of the whole set is unity)

(iii) If $A \cap B = \phi$, then $P(A \cup B) = P(A) + P(B)$.



Kolmogorov axioms (1933)

(Note that (iii) states that if A and B are mutually exclusive (M.E.) events, the probability of their union is the sum of their probabilities.)

How do we define Probability?

$\cap \equiv$ intersection
$\cup \equiv$ union

A= {1,2,3}
B={1,3,5}
$A \cap B$={1,3}
$A \cup B$= {1,2,3,5}

Example: the trial could be rolling a dice and the event could be rolling a 6.

six sided dice: $P(6) = 1/6$

for an honest dice: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

coin toss: $P(\text{heads}) = P(\text{tails}) = 0.5$

P(heads) should approach 0.5 the more times you toss the coin.

For a single coin toss we can never get $P(\text{heads}) = 0.5$!

Let A and B be subsets of S then $P(A) \geq 0$, $P(B) \geq 0$

Events are independent if: $P(A \cap B) = P(A)P(B)$

Coin tosses are independent events, the result of the 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 tossing a coin we either get a head or a tail.

Sum (or integral) of all probabilities if they are mutually exclusive must = 1.

Probability: Example

Example: Assume the probability of having pizza for lunch is 40%, the probability of having pizza for dinner is 70%, and the probability of having pizza for lunch and dinner is 30%. Also, this person always skips breakfast. We can recast this example using:

$P(A)$= probability of having pizza for lunch =40%

$P(B)$= probability of having pizza for dinner = 70%

$P(A \cap B)$=30% (pizza for lunch and dinner)

1) What is the probability that pizza is eaten <u>at least</u> once a day?

The key words are "at least once", this means we want the union of A & B

$P(A \cup B)=P(A)+P(B)-P(A \cap B) = .7+.4-.3 =0.8$

2) What is the probability that pizza is not eaten on a given day?

Not eating pizza (Z') is the complement of eating pizza (Z) so $P(Z)+P(Z')=1$

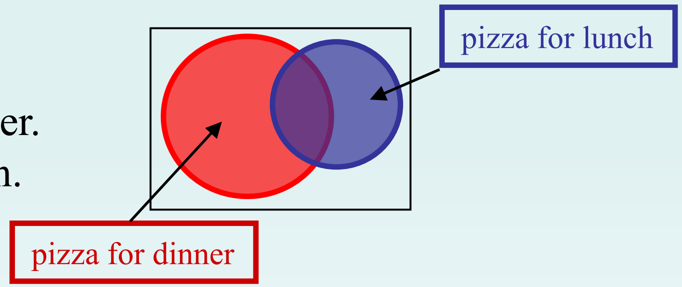
$P(Z')=1-P(Z) =1-0.8 = 0.2$

3) What is the probability that pizza is only eaten <u>once</u> a day?

This can be visualized by looking at the Venn diagram and realizing we need to exclude the overlap (intersection) region.

$P(A \cup B)-P(A \cap B) = 0.8-0.3 =0.5$

The non-overlapping blue area is pizza for lunch, no pizza for dinner.
The non-overlapping red area is pizza for dinner, no pizza for lunch.



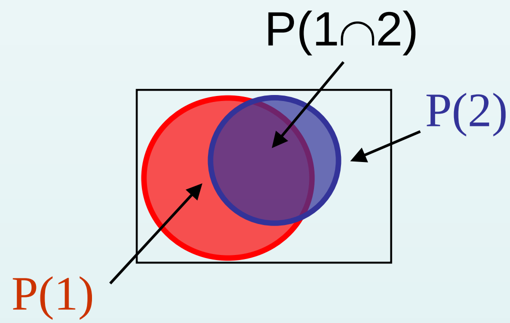
Independence Example

Let's consider a situation where we are trying to determine the number of events (N) we have in our data sample. This is a "classic" problem that comes in many different situations:

data analysis: using computer algorithms to find rare events

Consider the case where "data1" finds N_1 events and "data2" finds N_2 events. The number of events found by both data sample is N_{12}.

What can we say about the total number of events, N?



$$P(1) = \frac{N_1}{N} \quad P(2) = \frac{N_2}{N} \quad P(1 \cap 2) = \frac{N_{12}}{N}$$

If the data are independent then: $P(1\cap 2)=P(1)P(2)$

$$P(1)P(2) = P(1 \cap 2) \Rightarrow \frac{N_1}{N} \times \frac{N_2}{N} = \frac{N_{12}}{N}$$

$$\boxed{N = \frac{N_1 N_2}{N_{12}}}$$

Bayes's Theorem

Bayes's Theorem relates conditional probabilities. It is widely used in many areas of the physical and social sciences.

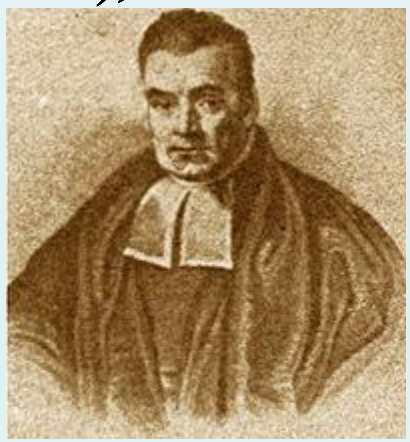
1) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

2) Conditional probability: $P(A \cap B) = P(B)\, P(A|B)$. Read "the probability of B times the probability of A given B".

3) A special case of conditional probability: if A and B are independent of each other (nothing connects them), then $P(A \cap B) = P(A)\, P(B)$

4) Bayes' Theory

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



Conditional Probability

Frequently we must calculate a probability assuming something else has occurred.

This is called conditional probability.

Here's an example of conditional probability:

Suppose a day of the week is chosen at random. The probability the day is Thursday is 1/7.

P(Thursday)=1/7

Suppose we also know the day is a weekday. Now the probability is conditional, =1/5.

P(Thursday|weekday)=1/5

the notation is: probability of it being Thursday given that it is a weekday

Formally, we define the conditional probability of A given B has occurred as:

$$P(A|B)=P(A\cap B)/P(B)$$

We can use this definition to calculate the intersection of A and B:

$$P(A\cap B)=P(A|B)P(B)$$

For the case where the A_i's are both mutually exclusive and exhaustive we have:

For our example let B=the day is a Thursday, A_1= weekday, A_2=weekend, then:

P(Thursday)=P(thursday|weekday)P(weekday)+P(Thursday|weekend)P(weekend)

P(Thursday)=(1/5)(5/7)+(0)(2/7)=1/7

Example

- Eighty percent of all students taking Quantum Physics pass the course while seventy-five percent of those taking Electromagnetic pass the course. However, of those that have passed Quantum Physics, 90 percent will pass Electromagnetic. What fraction of students passes both courses? What fraction will pass at least one course?

Let A = the event, student passes Quatum Physics
B = the event, student passes Electromagnetic
Given: $P(A) = .80$; $P(B) = .75$; $P(B|A) = .90$
Required: $P(A \cap B) = ?$ and $P(A \cup B) = ?$

$$P(A \cap B) = P(B|A)\, P(A) = (.90)\,(.80) = .72$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = .80 + .75 - .72 = .83$$

Bayes' Theorem

$$P(H|D,I)=\frac{P(H|I)P(D|H,I)}{P(D|I)}$$

This just follows from laws of conditional probability---even frequentists agree, but they give it a different interpretation

H = a hypothesis (e.g. “this material is a superconductor”)
I = prior knowledge or data about H
D = the data

P(H|I) = the “prior probability” for H

P(D|H,I) = the probability of measuring D, given H and I. Also called the “likelihood”

P(D|I) = a normalizing constant: the probability that D would have happened anyway, whether or not H is true.

Note: you can only calculate P(D|I) if you have a “hypothesis space” you're comparing to. A hypothesis is only “true” relative to some set of alternatives.

Bayes' Theorem

- Disease and a test for the disease
- 0.1% of the population have the disease(prior)
- If one has the disease, the test is positive with 85% probability
- If one does not have the disease, the test is positive with 1% probability
- What is the posterior probability to have the disease, given a positive test?

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)}$$

A: positive test
B: has disease
<u>Prior</u>: P(B)=0.001
Likelihood: P(A|B)=0.85
P(A|~B) =0.01

Normalisation:
$P(A)=P(A\cap B)+P(A\cap \sim B)=$
$P(A|B)*P(B)+P(A|\sim B)*P(\sim B)=$
$0.85*0.001+0.01*0.999= 0.01084$

<u>Posterior</u>: P(B|A)=
0.85*0.001/0.01084 = 9%
Because the disease is so rare, the probability is only 7.8%.

The test has to be improved, 1% of false-positive tests is too much

Triple Scree Test

The incidence of Down's syndrome is 1 in 1000 births. A triple screen test is a test performed on the mother's blood during pregnancy to diagnose Down's. The manufacturer of the test claims an 85% detection rate and a 1% false positive rate. You (or your partner) test positive. What are the chances that your child actually has Down's?

Consider 100,000 mothers being tested. Of these, 100,000/1000=100 actually carry a Down's child, while 99,900 don't. For these groups:

85 are correctly diagnosed with Down's.
15 are missed by the test
999 are incorrectly diagnosed with Down's
98901 are correctly declared to be free of Down's

Fraction of fetuses testing positive who really have the disorder:
85/(85+999) = 7.8%

Interpretations of probability

There are multiple, sometimes mutually exclusive, ways to interpret probability. WHICH DO YOU BELIEVE?

1) Probability is a statement about frequency. If you repeat a measure 1000 times and get the same outcome 200 times, the probability of that outcome is 0.2. Relative frequency

$$P(A) = \lim_{n \to \infty} \frac{\text{times outcome is } A}{n}$$

Interpretations of probability

2) Probability is an intrinsic property of objects with its own objective existence. QM predicts probabilities---they somehow are intrinsic properties of systems.

Interpretations of probability

3) Probability is a statement about our knowledge, and so is subjective. While I say the probability of rain tomorrow is 1/3, you may have reason to believe otherwise and assign a different probability.

Subjective probability:

$$P(A) = \text{degree of belief that } A \text{ is true}$$

Both interpretations consistent with Kolmogorov axioms.

• In physics frequency interpretation often most useful, but subjective probability can provide more natural treatment of non-repeatable phenomena: systematic uncertainties

Problems with the frequency interpretation

1) We naturally want to talk about the probability of events that are not repeatable even in principle. Tomorrow only happens once--- can we meaningfully talk about it? Maybe we want to talk about the probability of some cosmological parameter, but we only have one universe! A strict interpretation of probability as frequency says that we cannot use the concept of probability in this way.

2) Probability depends on the choice of ensemble you compare to. The probability of someone in a crowd of people being a physicist depends on whether you are talking about a crowd at a soccer game, a crowd at a university club, or a crowd at a APS meeting.

In spite of these conceptual problems, the “frequentist interpretation” is the most usual interpretation used in science.

The "subjective" interpretation

In "Bayesian statistics" probability is a way of quantifying our knowledge of a situation. P(E)=1 means that it is 100% certain that E is the case. Our estimation of P depends on how much information we have available, and is **subject to revision**.

The Bayesian interpretation is the cleanest conceptually, and actually is the oldest interpretation. Although it is gaining in popularity in recent years, it's still considered "heretical". The main objections are:

1) "Subjective" estimates have no place in science, which is supposed to be an objective subject.

2) It is not always obvious how to quantify the prior state of our knowledge upon which we base our probability estimate.

the most common reason for scientists not to be Bayesian is because they think that most other scientists aren't.

Frequentist vs. Bayesian: does it matter?

- The interpretive framework determines which questions we ask, how we try to answer them, and what conclusions we draw.
- In general frequentist methods are most commonly encountered, but will cover Bayesian techniques thoroughly enough that you can use them when appropriate or desired. In many cases the Bayesian approach is simpler to understand.

Frequentist vs. Bayesian Comparison

Bayesian Approach

- "The probability of the particle's mass being between 1020 and 1040 MeV is 98%."
- Considers the data to be known and fixed, and calculates probabilities of hypotheses or parameters.
- Requires prior knowledge.
- Well-defined, automated "recipe" for handling almost all problems.
- Requires a model of the data.

Frequentist Approach

- "If the true value of the particle's mass is 1030 MeV, then if we repeated the experiment 100 times only twice would we get a measurement smaller than 1020 or bigger than 1040."
- Considers the model parameters to be fixed (but unknown), and calculates the probability of the data given those parameters.
- Uses "random variables" to model the outcome of unobserved data.
- Requires a model of the data.

Probability Distributions

- Probability theory is the foundation for ***statistical inference***. A ***probability distribution*** is a device for indicating the values that a random variable may have.
- There are two categories of random variables. These are:
  - ***discrete random variables***, and
  - ***continuous random variables***.

Discrete random variable

The ***probability distribution of a discrete random variable*** specifies all possible values of a discrete random variable along with their respective probabilities

Discrete random variables: weather might be sunny, rainy, cloudy, snow

P(Weather=sunny)
P(Weather=rainy)
P(Weather=cloudy)
P(Weather=snow)

Discrete probability

- 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:

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

NOTATION
x_i is called a random variable

Continuous probability

Continuous probability: P can be any number between 0 and 1.

define a "probability density function", *pdf*, *f(x)*:

$$f(x)dx = dP(x \le \alpha \le x + dx)$$ with α a continuous variable

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

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

We say that *f(x)* is *normalized* to one.

Probability for x to be <u>exactly</u> some number is zero since:

$$f(x)dx = dP(x \le \alpha \le x + dx)$$

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

Probability="area under the curve"

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

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

Note: in the above example the pdf depends on only 1 variable, x. In general, the pdf can depend on many variables, i.e. $f=f(x,y,z,\ldots)$. In these cases the probability is calculated using from multi-dimensional integration.

Probability can be a discrete or a continuous 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 normalized continuous distribution, these quantities are defined by:

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

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

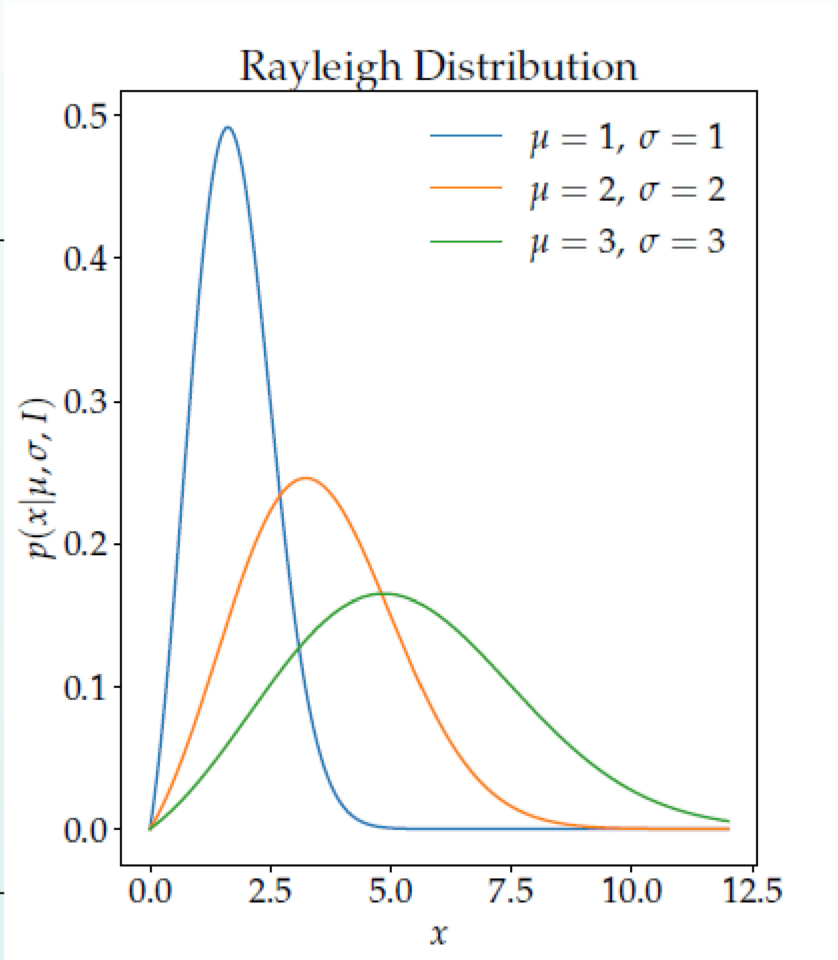
- for a discrete distribution, the mean and variance are defined by:

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

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

Basic parameters

Mean: "location", central value
Variance: "width" or "spread"
Mode: most probable value
Median: central value
Percentiles: rank/scoring
Skew: asymmetry of PDF
Kurtosis: "peakedness"

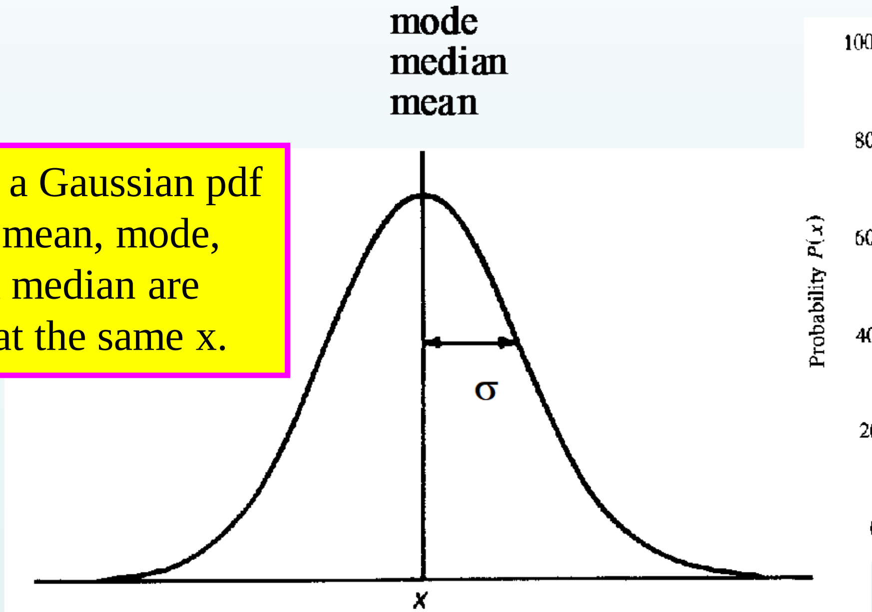


Probability can be a discrete or a continuous variable

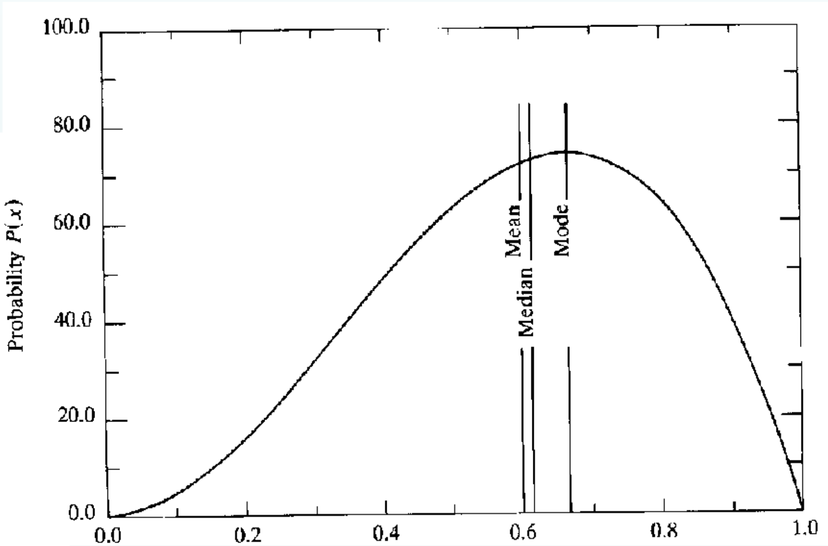
Remember: Probability is the area under these curves!

For many pdfs its integral can not be done in closed form, **use a table to calculate probability.**

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

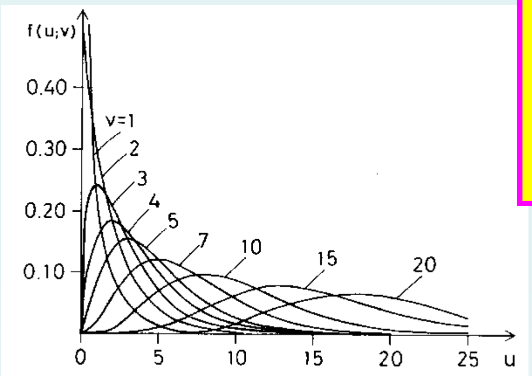


symmetric distribution (gaussian)

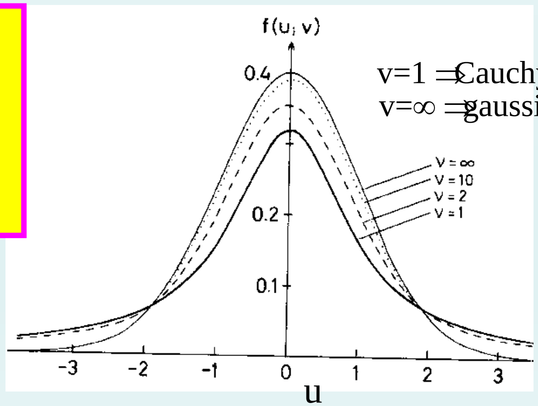


Asymmetric distribution showing the mean, median and mode

For many pdfs the mean, mode, and median are in different places.



Chi-square distribution



v=1 ⇒ Cauchy (Breit-Wigner)

v=∞ ⇒ gaussian

Student t distribution

The Centre of the Data: Mean, Median, & Mode

Mean of a PDF = expectation value of x

$$\mu \equiv \langle x \rangle \equiv \int dx\, P(x)\, x$$

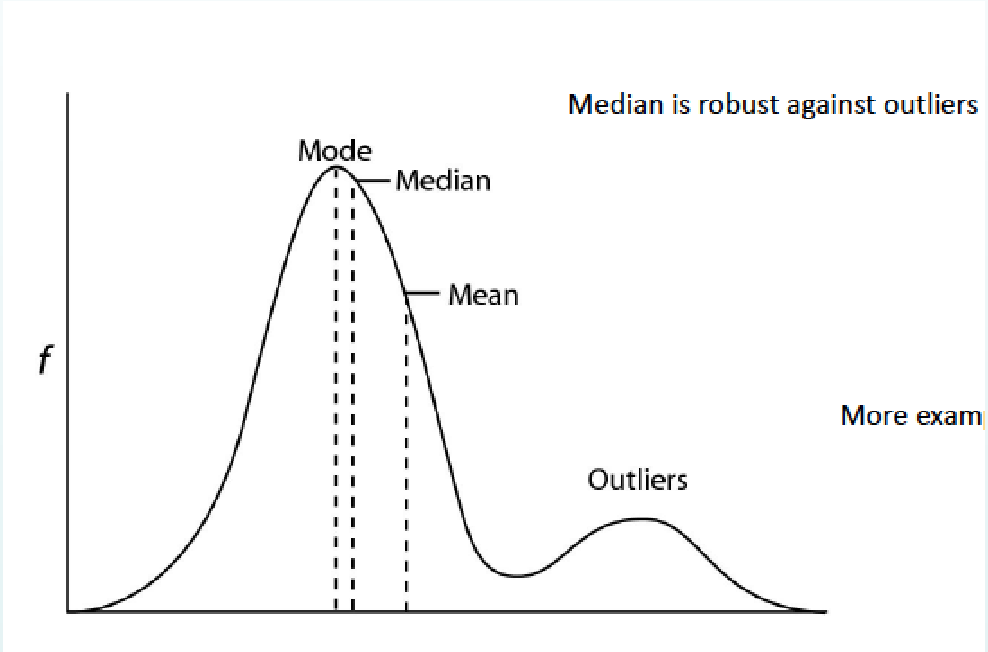
Other notations: E (x) and $\bar{x}$.

Typical usage: μ, <x>, and E (x) refer to the expectation value of a PDF, while $\bar{x}$ refers to the mean of a set of measurements {x}:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Weighted mean: if not all data should contribute equally to the sum,

$$\bar{x} = \frac{\sum_{i-1}^{N} w_i x_i}{\sum_{i-1}^{N} w_i}$$



Variance V & Standard Deviation σ

It measures the spread of x about its mean value we take the variance V(x), or dispersion, of x

Variance of a distribution: $$V(x)=\sigma^2=\int dx\, P(x)(x-\mu)^2$$

$$V(x)=\int dx\, P(x)\, x^2-2\mu\int dx\, P(x)\, x+\mu^2\int dx\, P(x)=\langle x^2\rangle-\mu^2=\langle x^2\rangle-\langle x\rangle^2$$

Variance of a data sample (has same notation as variance of a distribution---be careful!):

$$\hat{V}(x)=\sigma^2=\frac{1}{N-1}\sum_i (x_i-\bar{x})^2$$

Variance V & Standard Deviation σ

Note that the calculation of the variance of a data set will differ if the mean is known vs. calculated from the data.

This is unbiased if you must estimate the **mean from the data**.

$$\hat{V}(x) = \sigma^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

Use this if **you know the true** mean of the underlying distribution.

$$V(x) = \sigma^2 = \frac{1}{N} \sum_i (x_i - \mu)^2$$

If we compute ¯x from the data but use the formula on the rigth , our estimate of the variance of the PDF will be too small (biased).

Underestimating var (x), in this or any other way, can result in serious mistakes. For example, for small N you could underestimate the probability of observing a particular x_i.

Variance V & Standard Deviation σ

Suppose you have a detector that is measuring events xi in real time. How do you calculate var (x) as the data are recorded?

If you use the formula

$$\hat{V}(x) = \sigma^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2$$

then you need to estimate $\bar{x}$ and then recalculate all of the deviations from $\bar{x}$, requiring a second pass through the data. Inefficient!

But, if you realize that

$$\mathrm{var}(x) = \langle (x-\mu)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = \overline{x^2} - \bar{x}^2$$

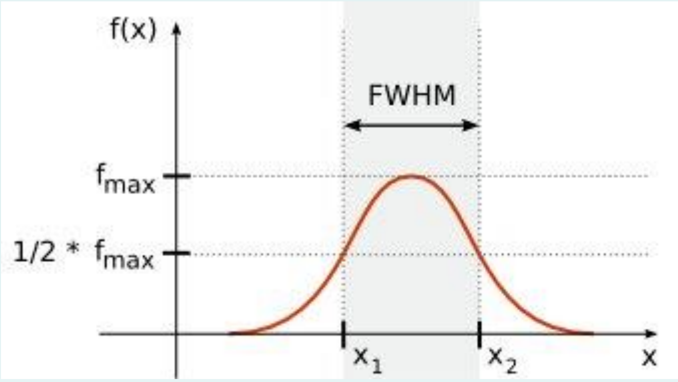
then you can write an algorithm that computes both the mean and variance on the fly.

FWHM & Quartiles/Percentiles

FWHM = Full Width Half Max. It means that the width of a distribution at the point where $P(x)=(1/2)(P_{max})$. For Gaussian distributions, FWHM=2.35σ

Quartiles, percentiles, and even the median are "rank statistics". Sort the data from lowest to highest. The median is the point where 50% of data are above and 50% are below. The quartile points are those at which 25%, 50%, and 75% of the data are below that point. You can also extend this to "percentile rank", just like on a GRE exam.

FWHM or some other width parameter, such as "75% percentile data point – 25% data point", are often robust in cases where the RMS is more sensitive to events on tails.



For describing core distribution
use FWHM, for describing tails use RMS

Median

The median is the halfway point: Half the population has a lower value, and half a higher value.

Aside from scoring data like exams, when is the median ever useful?

It is a measure of centrality that is less sensitive to the tails of of a PDF than other measures like the mean.

Example

Let $\{x_i\} = 1, 2, 1, 1, 1, 2, 3, 1, 1000$. The mean and median are given by

$$\bar{x} \approx 112.4$$
$$\text{median}\,(x) = 1$$

The mean in the example is sensitive to an outlier far from the main cluster of values, while the median is not. It is said to be "robust" against outliers.

Median/Mean

Suppose an insect life span of $p(x) = \frac{1}{72}x, \quad 0 \le x \le 12$ months, and 0 elsewhere. What's the median life span?

We need

$$\frac{1}{2} = \int_0^M \frac{1}{72} x\,dx = \frac{1}{144} x^2 \Big|_0^M = \frac{1}{144} M^2$$

Then solve for M:

$$\frac{1}{2} = \frac{1}{144} M^2$$

$$72 = M^2$$

$$M = \sqrt{72} \approx 8.49$$ months

Mean

What most people think of as the average:

Add up n values and divide by n.

How should we represent this for something represented by a density function p(x)?

Mean

Previous insect population example:

$p(x) = \frac{1}{72}x, \quad 0 \le x \le 12$ and p(x) = 0 elsewhere

mean

$$\int_{-\infty}^{\infty} x\,p(x)\,dx = \int_0^{12} x\left(\tfrac{1}{72}x\right)dx$$
$$= \int_0^{12} \tfrac{1}{72}x^2\,dx$$
$$= \frac{1}{216}x^3\Big|_0^{12}$$
$$= 8$$

Higher Moments

The mean ("central value") is the first moment of a PDF and the variance ("spread") is the second moment.

Of course you can calculate the r^{th} moment of a distribution if you really want to. For example, the third central moment is called the skew, and is sensitive to the asymmetry of the distribution (exact definition may vary---here's a unitless definition):

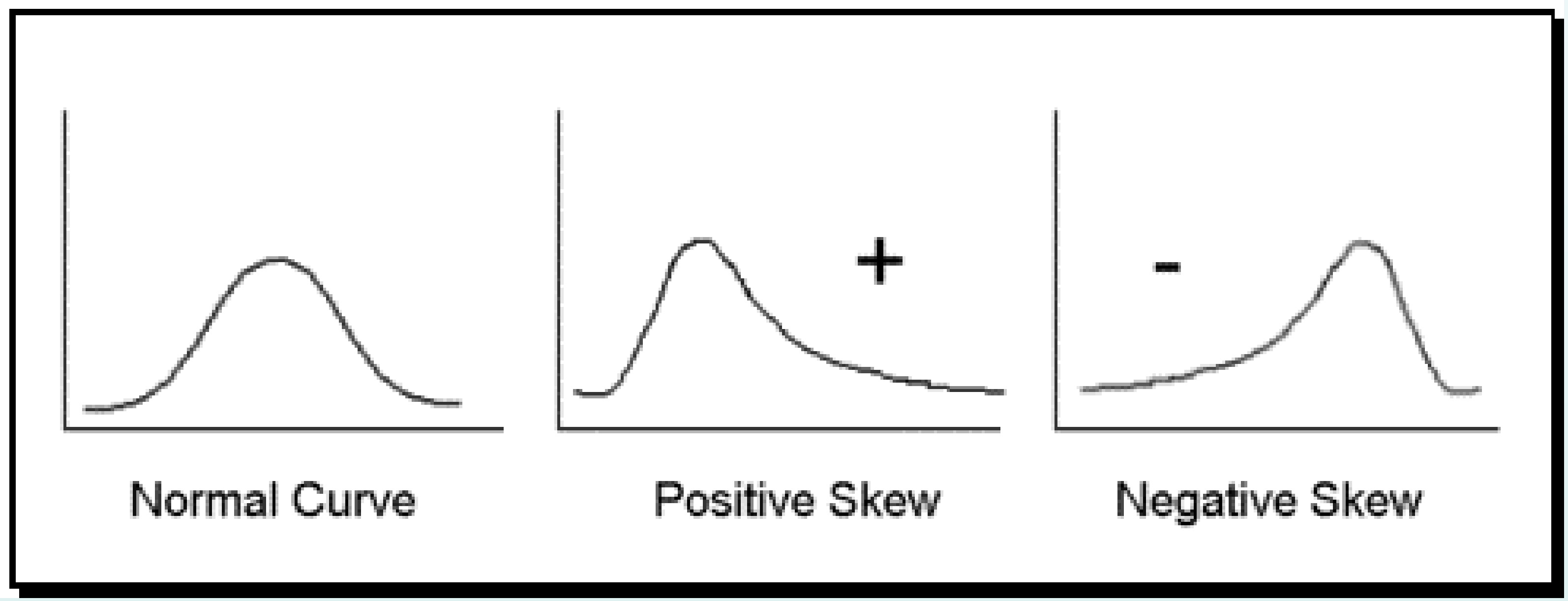
$$\text{skew} = \gamma = \frac{1}{N\sigma^3} \sum_i (x_i - \bar{x})^3$$

Kurtosis (or curtosis) is the fourth central moment, with varying choices of normalizations.

Warning: Not every distribution has well-defined moments. The integral or sum will sometimes not converge!

Skewed Data

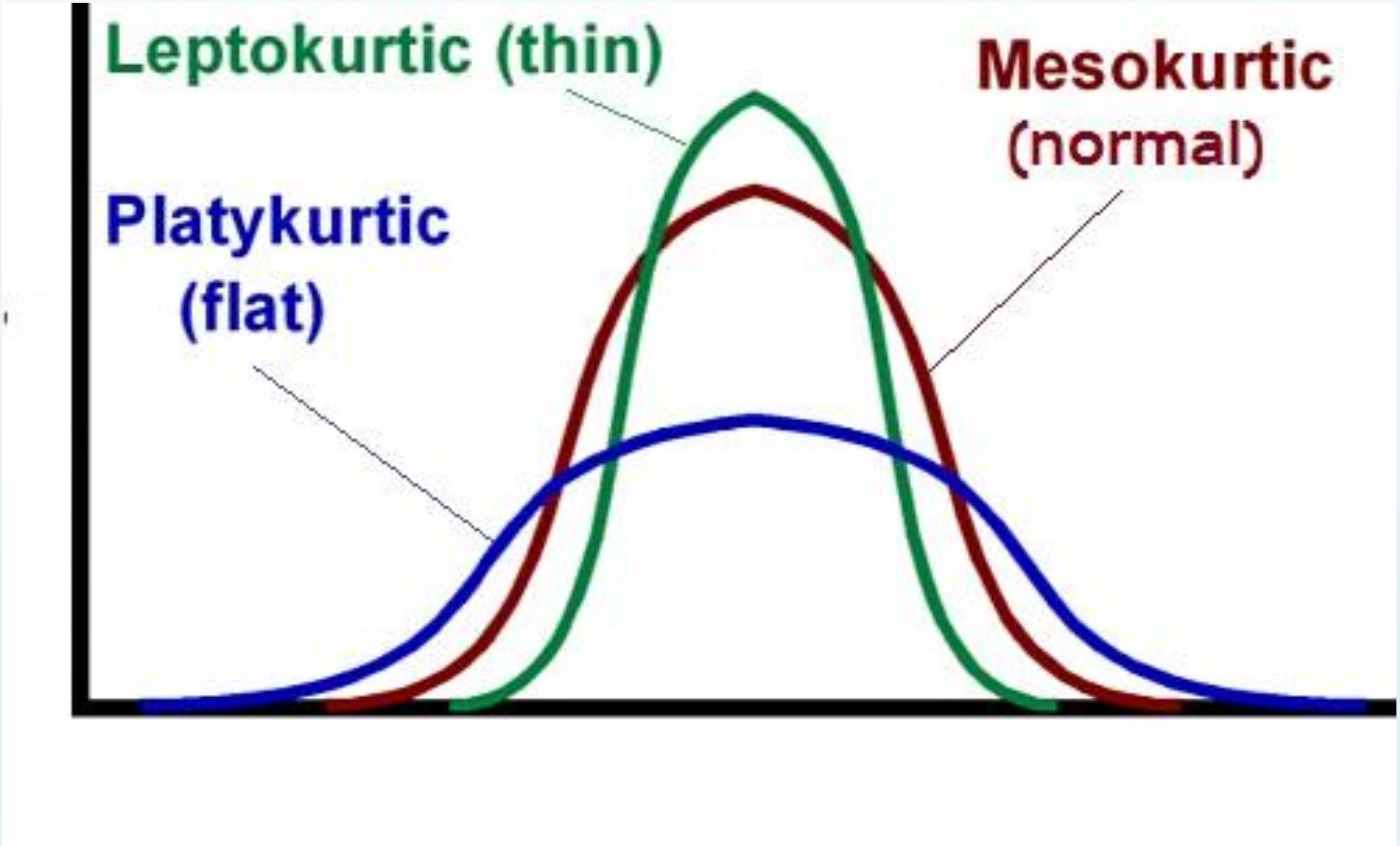
- Data may have a ***positive skew*** (long tail to the right, or a **negative skew** (long tail to the left).



Kurtosis Data

- ***Kurtosis*** indicates data that are bunched together or spread out.
- Data that are bunched together give a tall, think distribution which is not normal. This is called ***leptokurtic***.
- Data that are spread out give a low, flat distribution which is not normal. This is called ***platykurtic***.

Kurtosis Data



52

Experimental Uncertainity

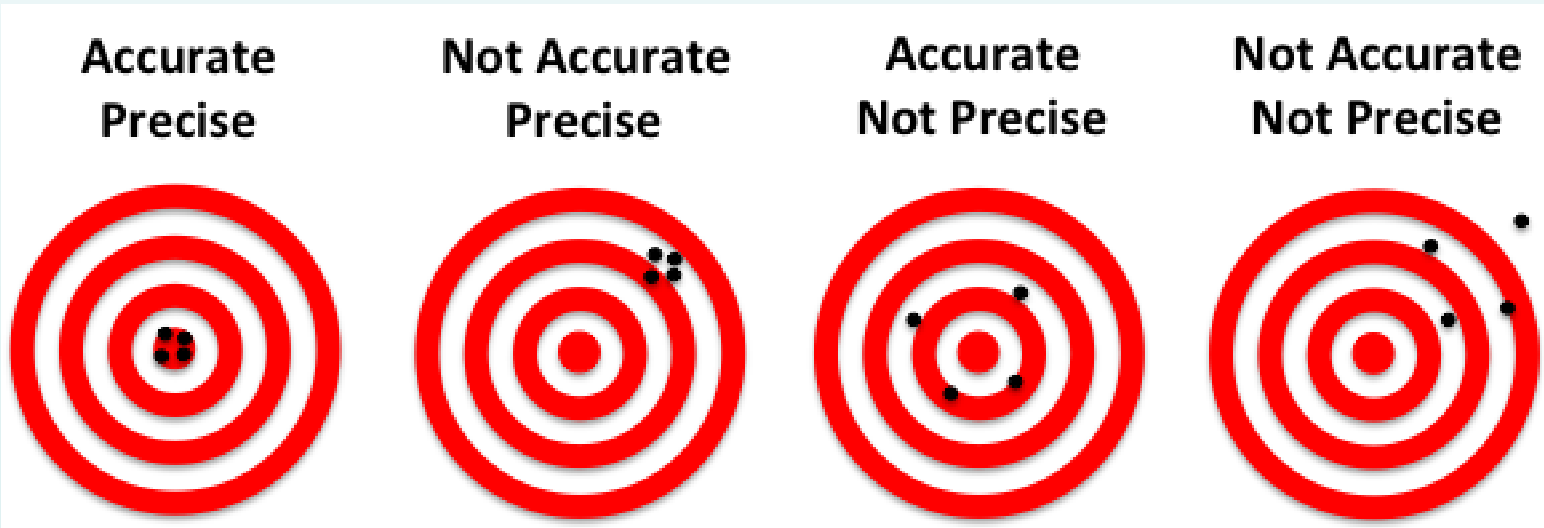
Experiments

- must be repeatable.
- result should be as accurate as possible
- all factors intervening in the result must be known i.e.
  - capabilities of the apparatus used to perform the measurement.

Accuracy & Precision

Accuracy:The accuracy of an experiment refers to how close the experimental measurement is to the true value of the quantity being measured. Accuracy gives information regarding 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. Precision gives information regarding the quality of measurement.



Error

An experimental measurement has no sense if it is not accompanied by its error.

Note that:

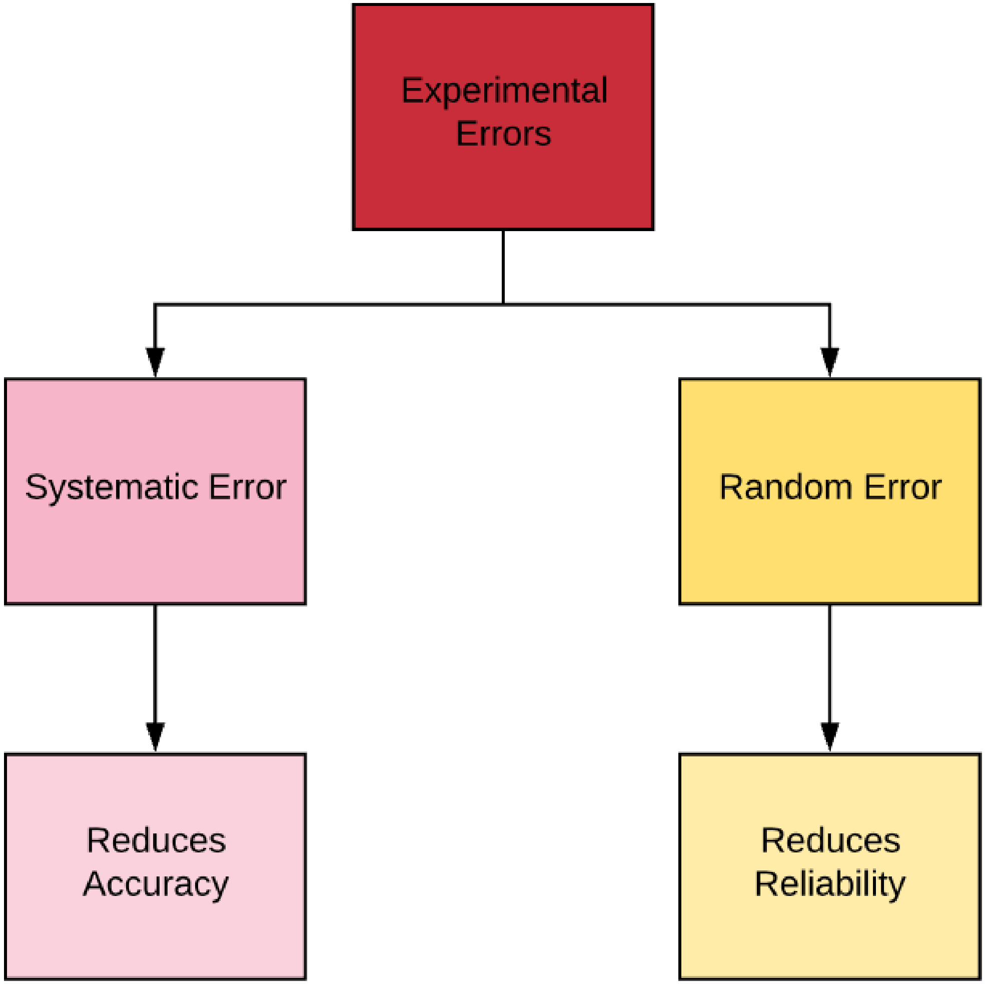
- **An error is not a mistake**, it is the difference between the measured and the true value of a given quantity.

- Errors cannot be avoided, but the experimentalist can make them as small as possible.

- The error can result from **inaccurate equipment, limited statistics, or other factors beyond your control.**

- There are many different types of errors that can occur in an experiment, but they will generally fall into one of two categories: **random errors or systematic errors.**

Random/Systematic errors



Random Errors(Precision Errors)

- arise from inherent instrument limitation (e.g. electronic noise) and/or the inherent nature of the phenomena (e.g. Radiactive decay, counting statistics).
- each measurement fluctuates independently of previous measurements, i.e. no constant offset or bias.
- measurements with a low level of random error have a high precision.

Random Errors(Precision Errors)

- Random errors usually result from human and from accidental errors. **Accidental errors are brought about by changing experimental conditions that are beyond the control of the experimenter;** examples are vibrations in the equipment, changes in the humidity, fluctuating temperatures, etc. Human errors involve such, the incorrect reading of an instrument, or a personal bias in assuming that particular readings are more reliable than others.

- By their nature, random errors cannot be quantified exactly since the magnitude of the random errors and their effect on the experimental values is different for every repetition of the experiment. **So statistical methods are usually used to obtain an estimate of the random errors in the experiment.**

Systematic Errors

- uncertainties in the bias of the data, such as an unknown constant offset, instrument mis-calibration.
- implies that all measurements are shifted the same (but unknown) amount from the truth.
- measurements with a low level of systematic error, or bias, have a high accuracy.

Systematic Errors

- **Systematic error is an error that will occur consistently in only one direction each time the experiment is performed**, i.e., the value of the measurement will always be greater (or lesser) than the real value. Systematic errors most commonly arise from defects in the instrumentation or from using improper measuring techniques.

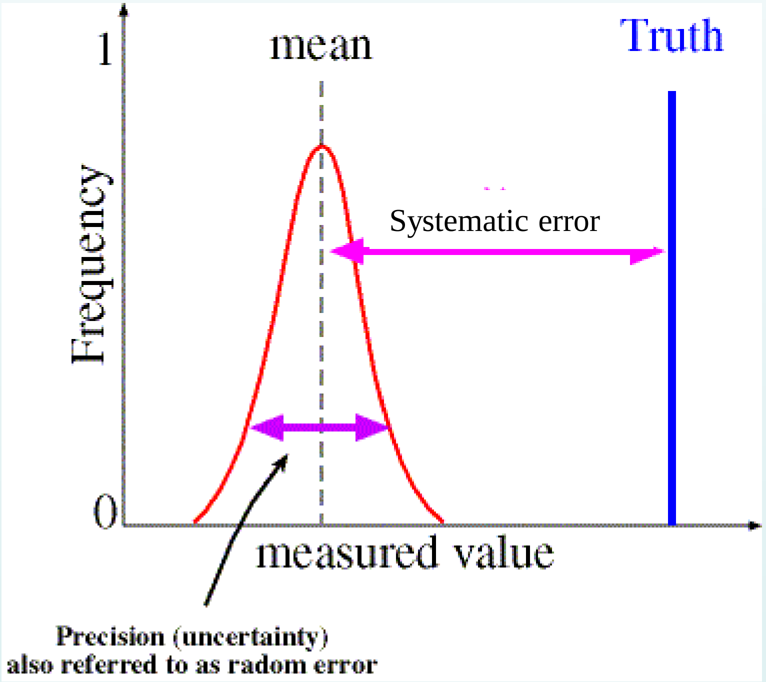
- For example, measuring a distance using an instrument that is not calibrated, or incorrectly neglecting the effects of viscosity, air resistance and friction, are all factors that can result in a systematic shift of the experimental outcome.

- Although the nature and the magnitude of systematic errors are difficult to predict in practice, some attempt should be made to quantify their effect whenever possible.

- **Proper calibration and adjustment of the equipment will help reduce the systematic errors leaving only the accidental errors to cause any spread in the data.**

Random/Systematic Errors

- Systematic/bias errors are consistent and repeatable (constant offset)



Systematic Errors

- Systematic errors typically cannot be characterized with statistical methods but rather must be analyzed case-by-case.
- Measurement standards should be used to avoid systematic errors as much as possible.
  - – double-check equipment against known values established by standards.
- If there is a fatal flaw in a study, it is usually from an overlooked systematic error (i.e. bias).!
- Attention to experimental detail is the only defense.

Random/Systematic Errors

Error	Description	Random/Systematic
Scale Error	If the equipment is not calibreted correctly, all measurements will be offset by the same fraction	Systematic Error
Zero error	If the equipment has an offset (e.g. A mass scale shows a reading that is not zero when there is nothing on it), all measurements will be offset by the same amount	Systematic Error
Parallax error	If you make a measurement by comparing an indicator against a scale (e.g. Reading a voltmeter), the angle at which you view it will affect the reading	Systemtatic error if you always view the dial from the same angle. Random error if you view the dial from a random angle each time
Errors arising from the environment	Ideally, the control variables are kept constant, but some may be beyond your control, e.g. Air pressure, temperature, humidity , vibrations	Changes to the control variables can result in both systematic and random errors. One consistent change give a systematic error. Random changes will give random error.
Reaction Time	If a measurement relies on your reaction time, then you may react too early or too late by different amounts of time	Random error
Measurement errors from insufficient precision	If you are measuring something that falls between two markings on a scale(e.g. You are using a ruler to measure something that is 10.25mm long) you cannot measure its precise value and will need to round it up or down.	Random error.

How to reduce Random/Systematic Errors

- Random errors can shift values both higher and lower, they can be eliminated through **repetition and averaging**

- Systematic errors arises from equipment, to eliminate it is to use **calibrated equipment and eliminate any zero or parallax error**.

Measures of Central Tendency:

- Mode - Most Frequent Measurements (not necessarily unique)
- Median - Central Value dividing data set into 2 equal parts (unique term)
- Mean (Arithmetic Mean)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Statistical Methods

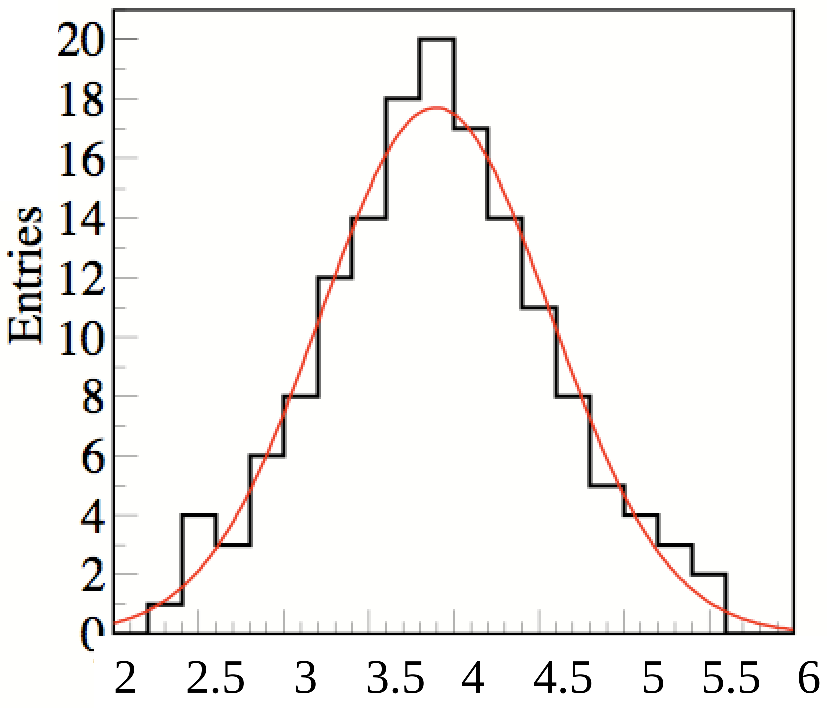
- When several independent measurements of a quantity are made, an expected result to report for that quantity is represented by the average of the measurements. For a set of experimental data containing N elements, or measurements, given by $\{S_1, S_2, S_3, \ldots, S_N\}$, the average S_{avg}, or expectation value $\overline{S}$, is calculated using the formula

$$
\begin{aligned}
\bar{S} &\equiv \frac{1}{N}\sum_{i=1}^{N} S_i\,, \\
&= \frac{S_1 + S_2 + S_3 + \cdots + S_N}{N}\,.
\end{aligned}
$$

- The fact that the average represents the closest approximation that is available to the true value of the quantity being measured**. It is sometimes referred to as the best estimate of the true value.**

Finding Standard Error in the Mean

- Example: Measure circumference of 150 times the same length – *make a histogram*
- The values appears to be centered on the average value (mean, $\bar{c}$)
- The values appears to occur more frequently near the average and less frequently further away (symmetric distribution)

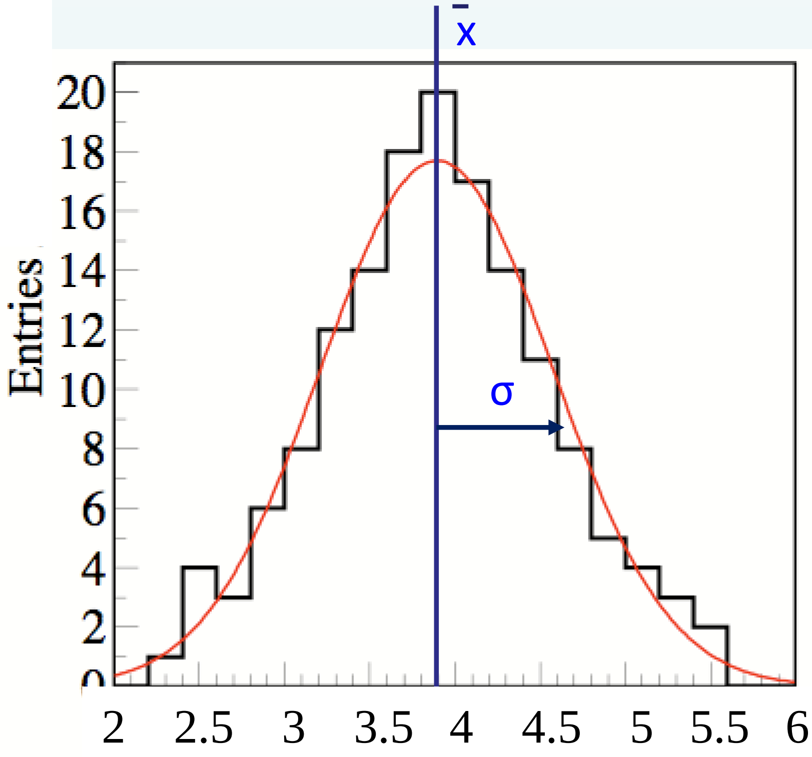


$$\bar{c} = \langle c \rangle = \frac{1}{N}\sum c_i = 4.39\text{m}$$

$$\sigma_c = \sqrt{\frac{\sum_{i=1}^{N}(c_i - \langle c \rangle)^2}{N-1}} = 0.70 \text{ m}$$

Finding Standard Error in the Mean

- Mean($\bar{x}$) and width (standard deviation, σ) describe distribution



What is our uncertainty in the estimate of the Mean we get from the N measurements of the histogram?

- Standard deviation of the mean

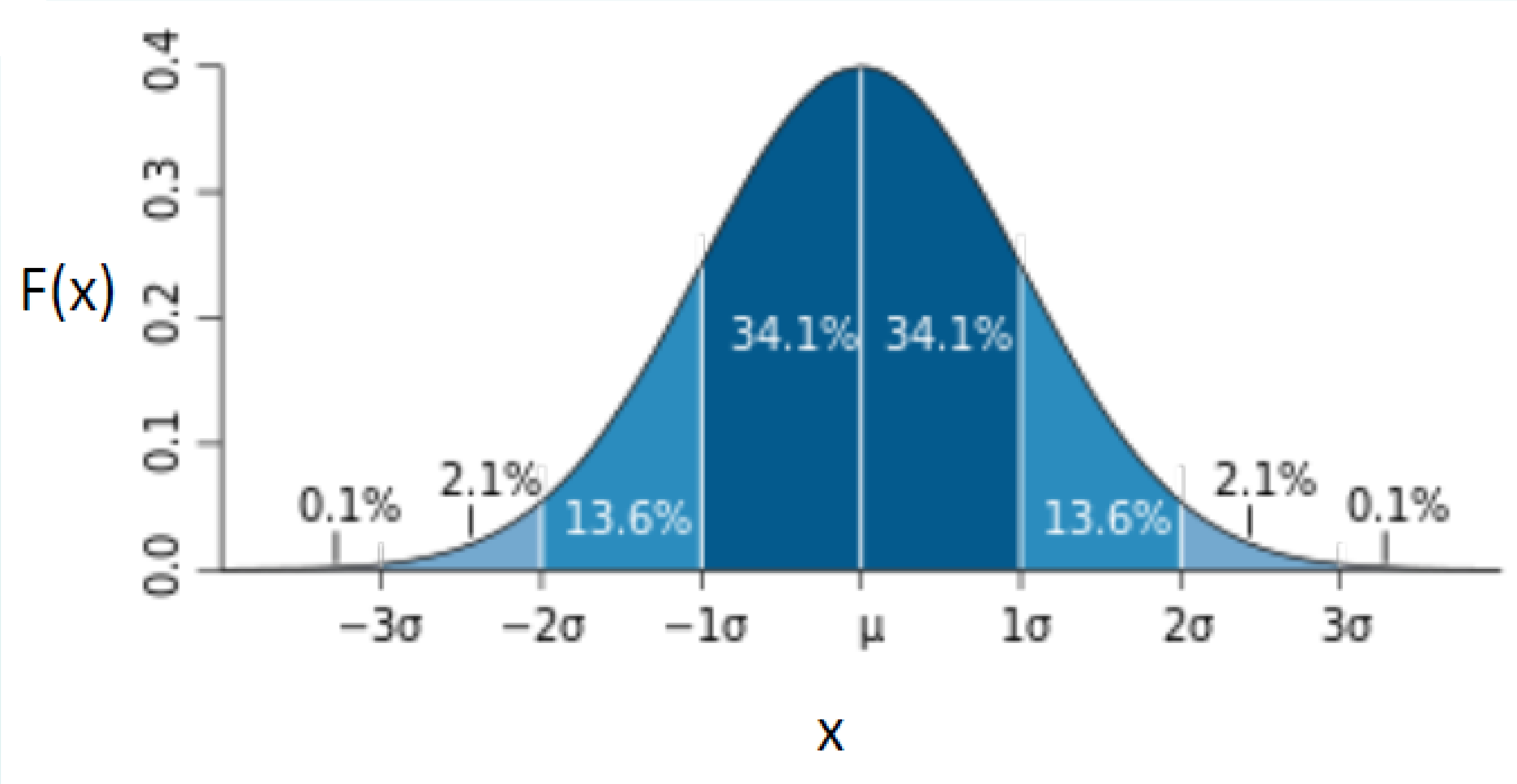
$$\sigma_{\overline{x}} = \sigma_x / \sqrt{N}$$

- We know the mean to much better than the uncertainty on individual measurements for our 150 measurements.

$$\sigma_{\overline{x}} = 0.7 \text{ m} / \sqrt{150} = 0.06 \text{ m}$$

Finding Standard Error in the Mean

- Mean 4.39, STD: 0.70m
- 68% of the measured length is between 4.39-0.70 and 4.39 + 0.70



Example of uncertainty quotation

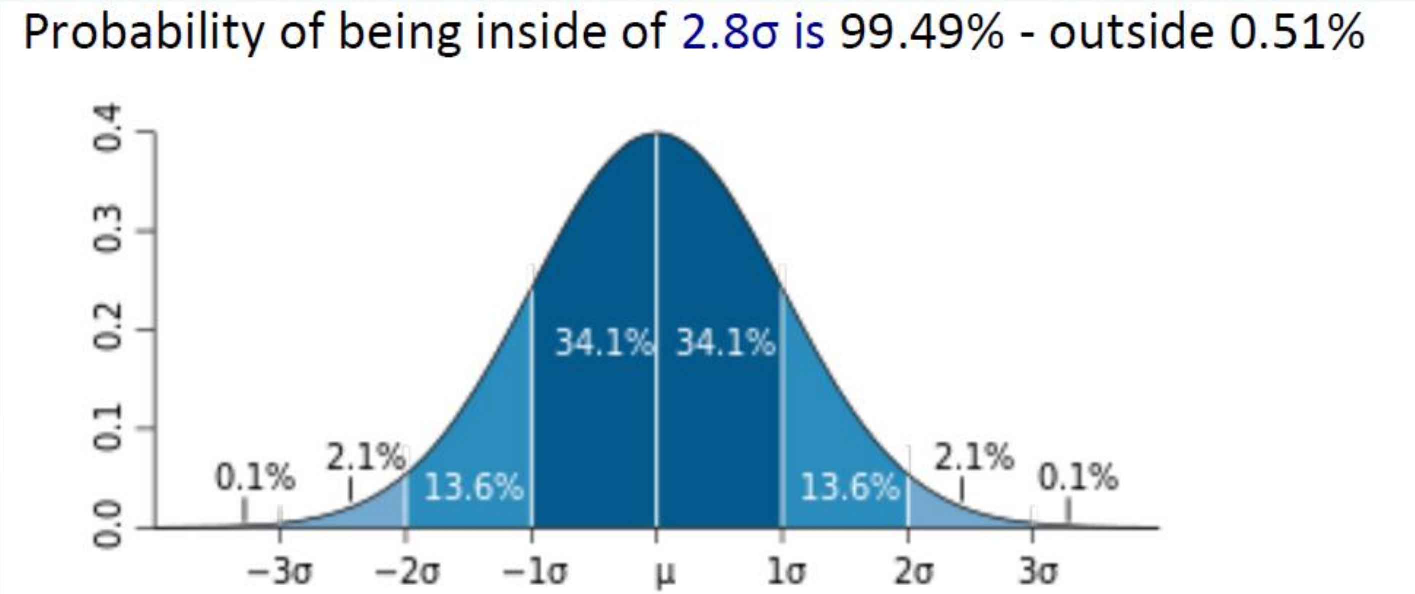
- Consider a measurement of e/m
- Measured value: $1.946 \pm 0.019(\text{stat.}) \pm 0.064(\text{sys.})$ C/kg
- Accepted value : **1.75882×10^{11}** C/kg
- Discrepancy from the expected valu $[1.946-1.75882] \times 10^{11}$ C/ kg
- Significance of Discrepancy: Ratio of discrepancy and σ_{total}

$$\frac{\text{discrepancy}}{\sigma_{\text{total}}} = \frac{0.187}{\sqrt{(0.019)^2 + (0.064)^2}} = 2.8$$

Saying 'off' 2.8σ

Example of uncertainty quaotation

Is it a good measurement ?



- No! You will need to discuss more about the sourcess of error/ go back and fix mistakes
- 2σ is often used as a cutoff (5%)

Propagation of Errors

- In many experiments, the quantities measured are not the quantities of final interest. Since all measurements have uncertainties associated with them, clearly any calculated quantity will have an uncertainty that is related to the uncertainties of the direct measurements. The procedure used to estimate the error for the calculated quantities is called the propagation of errors.
- Consider the general case first. Suppose the variables $A,B,C, \ldots$ represent independent measurable quantities that will be used to obtain a value for some calculated quantity U. Since U is a function of $A,B,C, \ldots$, it can be written as $U = f(A,B,C, \ldots)$. The measurements of the quantities $A,B,C, \ldots$ yield estimates of the expectation values written as $\bar{A}, \bar{B}, \bar{C}, \ldots$ and the associated uncertainties $\Delta A, \Delta B, \Delta C, \ldots$ for each variable. To find the expectation value, or best estimate, for the quantity U, the expectation value of each measured variable is substituted into the equation for U:

$$\bar{U} = f(\bar{A}, \bar{B}, \bar{C}, \ldots) .$$

- If the errors for A,B,C, . . . are independent, random, and sufficiently small, it can be shown that the uncertainty for U is given by

$$\Delta U = \sqrt{\left(\frac{\partial U}{\partial A}\right)^2 (\Delta A)^2 + \left(\frac{\partial U}{\partial B}\right)^2 (\Delta B)^2 + \left(\frac{\partial U}{\partial C}\right)^2 (\Delta C)^2 + \cdots}$$

Addition and Subtraction

- The correct notation to express the final estimate for the calculated quantity U is given by

$$U = \bar{U} \pm \Delta U$$

- Suppose two quantities A and B are added and the uncertainties associated with each variable are ΔA and ΔB. It follows that

$$U = A + B \qquad \longrightarrow \qquad \frac{\partial U}{\partial A} = 1\,, \quad \frac{\partial U}{\partial B} = 1$$

$$\Delta U = \sqrt{(\Delta A)^2 + (\Delta B)^2}\,.$$

- If the two values are subtracted then

$$U = A - B \qquad \longrightarrow \qquad \frac{\partial U}{\partial A} = 1\,, \quad \frac{\partial U}{\partial B} = -1$$

$$\Delta U = \sqrt{(\Delta A)^2 + (\Delta B)^2}\,.$$

- Hence these equations can be generalized for any combination of addition and subtraction of any number of variables, i.e.

$$\begin{aligned} U &= \pm A \pm B \pm C \pm \cdots\,, \\ \Delta U &= \sqrt{(\Delta A)^2 + (\Delta B)^2 + (\Delta C)^2 + \cdots} \end{aligned}$$

Multiplication and Division

- Suppose two quantities A and B are multiplied and the uncertainties associated with each variable are ΔA and ΔB. It follows that

$$U = A\,B \qquad \longrightarrow \qquad \frac{\partial U}{\partial A} = B\,, \quad \frac{\partial U}{\partial B} = A$$

$$\Delta U = U\sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2}$$

- Likewise if the two values are divided then

$$U = \frac{A}{B} \qquad \longrightarrow \qquad \frac{\partial U}{\partial A} = \frac{1}{B}\,, \quad \frac{\partial U}{\partial B} = -\frac{A}{B^2}$$

$$\Delta U = U\sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2}$$

- Hence these equations can be generalized for any combination of multiplication and division of any number of variables

$$U = \frac{A\,B\,\cdots}{X\,Y\,\cdots},$$

$$\Delta U = U\sqrt{\left(\frac{\Delta A}{A}\right)^2 + \left(\frac{\Delta B}{B}\right)^2 + \cdots + \left(\frac{\Delta X}{X}\right)^2 + \left(\frac{\Delta Y}{Y}\right)^2 + \cdots}\,.$$

Exponents and Roots

- Suppose that a calculation involves the use of exponents or roots of a quantity A whose uncertainty is ΔA. It follows in the most general case that,

$$\begin{aligned} U &= k\, A^{\alpha}\,, \\ \Delta U &= k\, \alpha\, A^{\alpha-1}\, \Delta A \end{aligned}$$

Example

- Suppose the acceleration due to gravity can be calculated from experimental data using the following equation

$$g = \frac{2\,x\,\ell}{h\,t^2}$$

where the estimate measurements are given by

$$\begin{aligned}
x &= \bar{x} \pm \Delta x = 74.11 \pm 0.05 \text{ cm} \\
\ell &= \bar{\ell} \pm \Delta \ell = 100.1 \pm 0.2 \text{ cm} , \\
h &= \bar{h} \pm \Delta h = 1.10 \pm 0.1 \text{ cm} , \\
t &= \bar{t} \pm \Delta t = 3.708 \pm 0.003 \text{ s}
\end{aligned}$$

$$\Delta g = \bar{g}\sqrt{\left(\frac{\Delta x}{\bar{x}}\right)^2 + \left(\frac{\Delta \ell}{\bar{\ell}}\right)^2 + \left(\frac{\Delta h}{\bar{h}}\right)^2 + \left(\frac{\Delta (t^2)_{t=\bar{t}}}{\bar{t}^2}\right)^2} ,$$

$$\begin{aligned}
&= \bar{g}\sqrt{\left(\frac{\Delta x}{\bar{x}}\right)^2 + \left(\frac{\Delta \ell}{\bar{\ell}}\right)^2 + \left(\frac{\Delta h}{\bar{h}}\right)^2 + \left(\frac{2\,\Delta t}{\bar{t}}\right)^2} , \\
&= 981\sqrt{\left(\frac{0.05}{74.11}\right)^2 + \left(\frac{0.2}{100.1}\right)^2 + \left(\frac{0.1}{1.10}\right)^2 + \left(\frac{0.006}{3.708}\right)^2} , \\
&\simeq 9 \text{ cm}\,\text{s}^{-2} .
\end{aligned}$$

$$g = 981 \pm 9 \text{ cm}\,\text{s}^{-2}$$

Numerical Errors

1. ***round-off errors***: due to a limited number of significant digits

2. ***truncation errors***: due to the truncated terms e.g. infinite Taylor series

3. ***propagation errors***: due to a sequence of operations. It can be reduced with a good computational order. e.g.

Errors in numerical computation

True Value = Approximation + Error

E_t = True value – Approximation (+/-)

True error

$$\text{True fractional relative error} = \frac{\text{true error}}{\text{true value}}$$

$$\text{True percent relative error, } \varepsilon_t = \frac{\text{true error}}{\text{true value}} \times 100\%$$

Errors in numerical computation

- For numerical methods, the true value will be known only when we deal with functions that can be solved analytically (simple systems). In real world applications, we usually not know the answer a priori. Then

$$\varepsilon_{\text{a}} = \frac{\text{Approximate error}}{\text{Approximation}} \times 100\%$$

- *Iterative approach*, example Newton's method

In many occasions, the error is calculated as the difference between the previous and the actual approximations.

$$\varepsilon_{\text{a}} = \frac{\text{Current approximation - Previous approximation}}{\text{Current approximation}} \times 100\%$$

(+ / -)

Errors in numerical computation

- Use absolute value.
- Computations are repeated until stopping criterion is satisfied.

$$|\varepsilon_a| \langle \varepsilon_s$$

Pre-specified % tolerance based on the knowledge of your solution

- It is convenient to relate the errors with the number of significant figures. If the following relation holds, one can be sure that at least n significant figures are correct.

$$\varepsilon_s = (0.5 \times 10^{(2-n)})\%$$

you can be sure that the result is correct to at least n significant figures.

Errors in numerical computation

- Example: Numerical Errors Analysis

$$x^3 - 3x^2 - 6x + 3 = 0$$

$$x = \sqrt{3x + 6 - \frac{8}{x}}$$

The initial estimate $x_0 = 2$

$$x_1 = \sqrt{3x_0 + 6 - \frac{8}{x_0}} = 2.828427$$

error: $e = x_1 - x_0 = 0.828427$

Errors in numerical computation

Trail			
0	2.000000	-	2.000000
1	2.828427	0.828427	1.171573
2	3.414214	0.582786	0.585786
3	3.728203	0.313989	0.271797
4	3.877989	0.149787	0.122011
5	3.946016	0.068027	0.053984
6	3.976265	0.030249	0.023735
7	3.989594	0.013328	0.010406
8	3.995443	0.005849	0.004557
9	3.998005	00002563	0.001995
10	3.999127	0.001122	0.000873

Round-off Errors

- Numbers such as π, e, or $\sqrt{7}$ cannot be expressed by a fixed number of significant figures.
- Computers use a base-2 representation, they cannot precisely represent certain exact base-10 numbers.
- Fractional quantities are typically represented in computer using "floating point" form, e.g.,

Integer part

$$m.b^e$$

exponent

mantissa

Base of the number system used

Chopping

Example:

π=3.14159265358 to be stored on a base-10 system carrying 7 significant digits.

π=3.141592 chopping error ε_t=0.00000065

If rounded

π=3.141593 ε_t=0.00000035

- Some machines use chopping, because rounding adds to the computational overhead. Since number of significant figures is large enough, resulting chopping error is negligible.

In Class Exercise

1. Write a python (use numpy, scipy and matplotlib)script

   -read data.txt file
   -calculate mean, median, standard deviation of each column data
   -Find the direction skew

Frequentist and Bayesian Statistics

The "propensity" view is not common but you will often encounter both frequentist and Bayesian approaches.

Bayesian

- Data are known and fixed; calculate probability of a hypothesis given data.
- Fundamental; a generalization of deductive logic.
- Well-defined procedure for calculating a probability.
- Must quantify prior knowledge of parameters/hypotheses, even if your "knowledge" is total ignorance. Can be HARD

Frequentist

- Model parameters are unknown but fixed; calculate probability of data given a hypothesis.
- Under limited circumstances, guarantees experiments obey long-run relative frequencies.
- Uses "random variables" to model the outcome of unobserved data.
- Ad hoc, "cookbook" approach to statistics.