Reminder

• I am back!

• HW10 due **Sunday 15 April** electronically by midnight

• Join Piazza!
  – NB: It is easiest to join Piazza by first authenticating through NetBadge, accessing our UVaCollab site, and then connecting to Piazza from the link on the left-hand sidebar

• Office hours: all held in our computer lab, room 022-C of this bldg
  – Me: After lecture 3:30-4:30 every Tuesday
  – TAs:
    • Mondays: 3-5pm and 6-8pm
    • Wednesdays: 5-9pm
Review and Outline

• Last time:
  – Histograms
  – Plot making tools
  – Gaussian statistics

• Today:
  – More on the gravity problem
  – Gaussian statistics
  – Other probability distributions
  – Gaussian limit
  – Interpreting experimental uncertainties
  – Problem solving
  – Debugging
  – Good coding style
Example: Structs and Functions
Example: The Gravity Problem

The Problem:
We want to read in the position vectors, initial velocities, and masses of a bunch of objects. Then, using this data, we want to calculate the gravitational force on each object, due to the others. (For the first part, we'll ignore the initial velocities.)

Here's one of our objects. It has mass “m”, and it's located at position $X$. The calculated force on it is $F$. 
Example: The Gravity Problem

Adding the Forces:

To find the total force on one mass, we just add the force vectors due to each of the other forces.

\[ \vec{F} = \vec{F}_b + \vec{F}_c + \vec{F}_d \]
Example: The Gravity Problem

Finding Distance and Direction:

We'll need to know the distance and direction to each other object.

This is the vector from a to b:

$$ \vec{r} = \vec{X}_b - \vec{X}_a $$

The magnitude of this vector gives us the distance:

$$ r = |\vec{r}| $$

Once we know these, we can make a unit vector pointing from a to b:

$$ \vec{u} = \frac{\vec{r}}{r} $$
Example: The Gravity Problem

**Calculating a Single Force:**

Newton tells us that the magnitude of the gravitational force between two objects is:

\[ F = G \frac{m_a m_b}{r^2} \]

The force will point toward the other object, so the force vector will just be:

\[ \vec{F} = F \hat{u} \]
Example: The Gravity Problem

Data Structure:

To solve this problem programatically, we'll first need a data structure to store information about each body:

```c
typedef struct{
    double s_vec[3];    // space(position) vector
    double v_vec[3];    // velocity vector
    double f_vec[3];    // force vector
    double mass;
} body;

const int MAX_BODIES = 100;
body bodies[MAX_BODIES];    // array of bodies
```
Example: The Gravity Problem

Reading Data from a File:

```c
int read_data(char* file, body *bodies) {
    int num=0; // number of entries read from file
    int status;
    FILE *file_p = fopen(file,"r");

    while(num<MAX_BODIES) {
        status=fscanf(file_p,"%lf %lf %lf %lf %lf %lf %lf",
                      &bodies[num].s_vec[0],
                      &bodies[num].s_vec[1],
                      &bodies[num].s_vec[2],
                      &bodies[num].v_vec[0],
                      &bodies[num].v_vec[1],
                      &bodies[num].v_vec[2],
                      &bodies[num].mass);
        if (status==EOF) break;
        num++;
    }
    return num;
}
```
Example: The Gravity Problem

Some Useful Functions:

```c
// Find distance between two points:
double distance(double *svec1, double *svec2)
{
    double dist2=0;
    int i;
    for (i=0; i<3; i++)
        dist2 += (svec1[i]-svec2[i])* (svec1[i]-svec2[i]);
    return sqrt(dist2);
}

// Find difference of two vectors:
void vsub(double *v1, double *v2, double *v1m2)
{
    int i;
    for (i=0; i<3; i++)
        v1m2[i] = v1[i]-v2[i];
}
```

Calculating the Forces:

```c
void forces(body *bodies, int nbodies){
    double dist, force;
    double dirvec[3];
    const double G = 6.67e-11;

    for(int i=0; i<nbodies; i++){
        bodies[i].f_vec[0]=0;
        bodies[i].f_vec[1]=0;
        bodies[i].f_vec[2]=0;
        for(int j=0; j<nbodies; j++){
            if (i!=j) {
                dist = distance(bodies[i].s_vec, bodies[j].s_vec);
                vsub(bodies[j].s_vec, bodies[i].s_vec, dirvec);
                dirvec[0] /= dist;
                dirvec[1] /= dist;
                dirvec[2] /= dist;
                force = G*bodies[i].mass*bodies[j].mass/(dist*dist);
                for(int k=0; k<3; k++) {
                    bodies[i].f_vec[k] += force*dirvec[k];
                }
            }
        }
    }
}
```
Example: The Gravity Problem

\[ \vec{F} = m \vec{a} \]
\[ \vec{a} = \vec{F} / m \]

\[ \Delta \vec{V} = \vec{a} \Delta t \]

\[ \vec{V}_{\text{new}} = \vec{V}_{\text{old}} + \Delta \vec{V} \]
Calculating Trajectories:

Here's one simple way to approximate the motion of the objects. Here, we assume a constant velocity during each time step:

```c
void evolve(body *bodies, int nbodies, double delta_t) {
    for (int i=0; i<nbodies; i++) {
        for (int j=0; j<3; j++) {
            double acceleration_j = bodies[i].f_vec[j] / bodies[i].mass;
            bodies[i].s_vec[j] += bodies[i].v_vec[j] * delta_t;
            bodies[i].v_vec[j] += acceleration_j * delta_t;
            a_j = F_j/m
            x_j^new = x_j + v_j Δt
            v_j^new = v_j + a_j Δt
        }
    }
}
```
Example: The Gravity Problem

A Better Approximation:

Here's a better approximation. In this version, we only assume a constant acceleration during each time step:

```c
void evolve(bodies, int nbodies, double delta_t){
    for (int i=0; i<nbodies; i++) {
        for (int j=0; j<3; j++){
            double acceleration_j =
                bodies[i].f_vec[j] / bodies[i].mass;
            a_j = F_j/m
            x_j^new = x_j + v_j Δt + \frac{1}{2} a_j Δt^2
            bodies[i].s_vec[j] +=
                (bodies[i].v_vec[j]*delta_t+0.5*acceleration_j*delta_t*delta_t);
            v_j^new = v_j + a_j Δt
            bodies[i].v_vec[j] += acceleration_j * delta_t;
        }
    }
}
```
Statistics: Our First Entrypoint
Making a Measurement: Truth

• Let’s say I want to measure some physical observable.
• Examples:
  – the acceleration due to gravity near the Earth’s surface
  – the speed of light
  – the mass of the Higgs boson
• Each such measureable has a single value – the truth:

Make N repeated measurements of some quantity $x$. A perfect apparatus and data collection scheme would yield the true value every time!
Why do we see a range of measured values?
- imperfections in our instruments
- limitations in our measurements
How do we go from to arrive at?

Use Statistics:

The best we can achieve is the center distribution. Its shape depends on the measurement’s **uncertainty**.
The Story of Measurement

• A measurement is not about the central value one finds
• A measurement is truly about its **uncertainty**
  – The central value in fact is MEANINGLESS and OF NO USE without understanding and reporting the associated uncertainty

• Two types of uncertainty:
  – **Systematic uncertainty**: Features of the measurement device or technique that shift (aka “bias”) the measured result away from the true value
    • Biases can often be corrected if discovered
    • Eg: Imagine you are measuring athletes running the 40-yard dash. You discover however the length they run is in fact 42 yards long. This can in principle be corrected
  – **Statistical or random uncertainty**: Features of the measurement device or technique that shift the measured result by a different amount in each attempt (hence, random)
“Uncertainty” and “error” are used interchangeably, often inappropriately. I will try to use “uncertainty” universally – and you should as well.
The Gaussian Distribution

Measures of random processes, in the limit of many samples often tend to produce a characteristic Gaussian or Normal distribution. (The classic “Bell Curve”.)

\[
P(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

\(\mu = \text{“mean”, average value of the distributions}\)
\(\sigma^2 = \text{“variance”, characterizes the width}\)
\(\sigma = \text{“standard deviation”}\)
The Gaussian Distribution

Measures of random processes, in the limit of many samples often tend to produce a characteristic Gaussian or Normal distribution. (The classic “Bell Curve”.)

Central Limit Theorem: Sufficiently large number of randomly distributed measurements always approaches a Gaussian distribution.

\( \mu = \text{“mean”}, \) average value of the distributions

\( \sigma^2 = \text{“variance”}, \) characterizes the width

\( \sigma = \text{“standard deviation”} \)
The red curve represents the probability of observing a particular value. This is the **Parent Distribution** from which our measurements will be drawn. Our experiment draws measurements, at random, from this. These measurements form our **Sample Distribution**.

In the limit of many samples our Sample Distribution can approach the Parent Distribution.
What Can We Get From the Sample Distribution?

Our best estimates of the mean and variance of the Parent Distribution come from the mean and variance of the Sample Distribution. This is something we can study with appropriate programs.

We can determine the parameters of the Sample Distribution, such as its mean ($\bar{x}$) and standard deviation ($s$).

Note: variance = $s^2$

From these parameters we can make statements about the parent distribution.
To clearly distinguish between parent and sample distributions, we use $\mu$ and $\sigma$ for the Parent Distribution and $\bar{x}$ and $s$ for the Sample Distribution.

Most probable value $= \mu$ (the mean)

$P(x) = \text{Probability of seeing a value between } x \text{ and } x+dx.$

68% of Area

68% probability of seeing a value between $+/- \sigma$ of the mean.

This is a probability distribution, so the total area under the curve is equal to 1. (There's a 100% probability that something happened.)

$$\int_{-\infty}^{\infty} P(x) \, dx = 1$$
The Normal distribution is a **Probability Density Function** (PDF), so the total area under the curve must be equal to 1. One way to see this is to look at the integral of the Normal distribution. The integral of a PDF is called a **Cumulative Distribution Function** (CDF).

The CDF tells us the probability of observing a value **less than or equal to** $x$. 

The CDF approaches 1 as $x$ increases.
Systematic errors (aka “biases”) must be corrected with expert knowledge of your apparatus and procedures. Statistics will not improve a mis-calibrated detector or a flawed design for your experiment!

Systematic errors might shift our results away from the true values we want to measure.

Perhaps our clock is running fast? Maybe our yardstick is worn down?
Recall: Monte Carlo Integration

The Whale Example:

$m$ = Number of points in whale shape
$M$ = Total number of points

$m/M$ as $M \to \infty$ \[ \frac{A_{\text{whale}}}{A_{\text{box}}} \]

$A_{\text{whale}} \approx A_{\text{box}} \times \frac{m}{M}$

If we repeat this experiment with a different set of random points, our resulting estimate of the whale's area will be slightly different. By coming up with several estimates of the area, and looking at how much variation they display, we can get an idea of how accurate our estimate is.
Recall: Multiple Measurements

The Sample Mean:

Experiment 1  Experiment 2

Experiment 3  Experiment 4

Each experiment contains $M$ points. $N = \text{Number of experiments.}$

The mean (average) value will be:

$$\overline{A} = \frac{1}{N} \sum_{i=1}^{N} A_i$$

Say we repeat the experiment $N$ times. From our $N$ experiments, we get a set of independent estimates of the area: $A_1, A_2, A_3, \ldots, A_N$.

If we made a histogram of our estimates, it might look like this:
Multiple Measurements

The Standard Deviation:

Each of our $A_i$ values will deviate from the true value ($A_{\text{whale}}$) by some amount, $d_i$.

Since these deviations may be positive or negative, the average deviation will tend toward zero.

How, then, can we come up with an estimate for a “typical” deviation? We can sum the squares of the $d_i$ values, instead!

We define the “variance” ($\sigma^2$) as the average squared deviation from the true value:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} d_i^2 = \frac{1}{N} \sum_{i=1}^{N} (A_i - A_{\text{whale}})^2$$

Its square root, $\sigma$, is called the “Standard Deviation” or “Standard Error”. It's also sometimes called the Root-Mean-Square (RMS) deviation, for obvious reasons.
Multiple Measurements

Important vocabulary:
- mean
- variance
- standard deviation or RMS

The standard deviation has an intimate relationship to the definition of the Gaussian (aka normal) distribution.

Here the standard deviation of the sample of measurements is usually the best estimate of the standard deviation of the parent distribution.
The standard deviation tells us about the distribution of our $A_i$ values. In particular, if the values follow a “normal” (gaussian) distribution, it tells us that we should expect that about 68% of our $A_i$ values will fall within $+/- \sigma$ of the true value of the area.
• If plotted, our measurements will look approximately like a Gaussian Distribution.
  – Is this magic?
  – No!

• Recall the Central Limit Theorem, paraphrased
  – Any collection of random measurements of some variable will resemble a Gaussian
The Importance of the Gaussian Distribution

• If plotted, our measurements will look approximately like a Gaussian Distribution.
  – Is this magic?
  – No!

• Recall the Central Limit Theorem, paraphrased
  – Any collection of random measurements of some variable will resemble a Gaussian

• Back to the whale:
  – Our random measurements here are the results of MC integration of the whale shape
  – An infinite number of such measurements would be a perfect Gaussian
But…What about the True Value?

The Sample Standard Deviation:

The problem with computing $\sigma$ is that we don't know the true value, $A_{\text{whale}}$. Our best approximation of this value is the average of all of our $A_i$ measurements, $\bar{A}$.

But, it can be demonstrated that just plugging $\bar{A}$ in place of $A_{\text{whale}}$ will systematically underestimate the value of sigma.

We correct for this by dividing by $N-1$ instead of $N$, and arrive at the following approximation:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (A_i - \bar{A})^2 \sim \sigma^2$$

We call $s$ the Sample Standard Deviation.
But...What about the True Value?

The Sample Standard Deviation:

The problem with computing $\sigma$ is that we don't know the true value, $A$. To correct for this, the change to using $N-1$ is called Bessel’s Correction. Read more [here](#).

What happens below as $N$ gets very large?

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (A_i - \overline{A})^2 \approx \sigma^2$$

We call $s$ the Sample Standard Deviation.
For purposes of writing computer programs, it’s useful to rearrange our expression for $s^2$ a little. Here are our expressions for $s^2$ and the average, $\overline{A}$:

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (A_i - \overline{A})^2$$

$$\overline{A} = \frac{1}{N} \sum_{i=1}^{N} A_i$$

Substituting the expression for $\overline{A}$ into the first equation, we can collect terms and arrive at this:

$$s^2 = \frac{1}{N-1} \left[ \sum_{i=1}^{N} A_i^2 - \frac{1}{N} \left( \sum_{i=1}^{N} A_i \right)^2 \right]$$

This is easy to code, because our program just needs to do one loop, and keep two sums: the sum of the values and the sum of their squares.
We said that our best guess of the true value, $A_{\text{whale}}$, is the mean of our sample values, $\bar{A}$. Now let’s get back to the problem of quantifying just how good that guess is. Earlier, we defined the mean as:

$$\bar{A} = \frac{1}{N} \sum_{i=1}^{N} A_i = \frac{1}{N} (A_1 + A_2 + A_3 \ldots)$$

If we know the standard deviations of the $A_i$ values, we can use them to calculate the standard deviation of the mean, through the regular propagation of errors process:

$$\sigma_f(x,y,z\ldots) = \sigma_x^2 \left( \frac{\partial f}{\partial x} \right)^2 + \sigma_y^2 \left( \frac{\partial f}{\partial y} \right)^2 + \sigma_z^2 \left( \frac{\partial f}{\partial z} \right)^2 \ldots$$

Where we treat $\bar{A}$ as a function of the variables $A_1, A_2, A_3 \ldots A_N$. 
Uncertainty on the Mean

Applying this to our expression for $\bar{A}$, we get:

$$\sigma^2_A = \sigma_1^2 \frac{1}{N^2} + \sigma_2^2 \frac{1}{N^2} + \sigma_3^2 \frac{1}{N^2} \ldots$$

What are the $\sigma_i$ values? They're the estimates of how far each $A_i$ value typically deviates from the true value. In our case, the values for $\sigma_i$ will all be equal to $\sigma$, the standard deviation of our collection of $A_i$ values. So, we can simplify the equation above and write:

$$\sigma^2_A = \frac{1}{N} \sigma^2$$

or

$$\sigma_{\bar{A}} = \frac{\sigma}{\sqrt{N}}$$

Our uncertainty in $\bar{A}$ decreases like the $\sqrt{N}$ as $N$ increases. For example, to get a 10-times better value for $\bar{A}$, we need to do 100 times as many measurements. Also, because we don’t know the exact value of $\sigma$, we say that $\sigma_{\bar{A}} \approx s/\sqrt{N}$. 

10
Confidence Levels

When we cite an experimental result, we also state the uncertainty in the result. So, in our whale-shape example, we might give our final estimate of the area as $\bar{A} + \sigma_{\bar{A}}$ (for example “1.12 +/- 0.01”).

What we mean:

“There’s a 68% probability that our average value, $\bar{A}$, lies within a distance +/- $\sigma_{\bar{A}}$ from the true value, $A_{\text{whale}}$.”

Probability of getting a given value of $\bar{A}$. 68% of Area
Back to the Whale!

Number of Points per Experiment:

What about “M”, the number of points in each experiment? So far, everything we’ve said has been independent of M.

Each experiment contains M points. N = Number of experiments.

Fewer points will increase the uncertainty in each Ai value (σ). But, since

$$\sigma_{A_i} = \sigma/\sqrt{N}$$

we can compensate for a factor of ten increase in σ by just increasing N (the number of experiments) by a factor of 100.

M = 10
σ = 0.2

M = 1000
σ = 0.02
The thing that matters is the TOTAL number of points that are thrown!

It does not matter if they are in 100, 10 or even 1 experiment – it is the number of points.

What if we treated each single randomly-thrown point as a single measurement?
Treating Each Point as an Experiment:

What would happen if we set M to 1? Then the number of points within the whale for each experiment \( m_i \) would always either be 0 or 1.

Each experiment’s estimate of \( A_{\text{whale}} \) is given by:

\[
A_i = \frac{m_i}{M} A_{\text{box}}
\]

So, our mean value for the area could be written:

\[
\overline{A} = \frac{1}{N} \sum_{i=1}^{N} A_i = \frac{1}{N} \sum_{i=1}^{N} \frac{m_i}{M} A_{\text{box}} = \frac{A_{\text{box}}}{NM} \sum_{i=1}^{N} m_i
\]

If \( M=1 \), then \( m_i \) must be either 0 or 1, and:

\[
\overline{A} = \frac{A_{\text{box}}}{N} \sum_{i=1}^{N} m_i = \frac{n}{N} A_{\text{box}}
\]

where “n” is just the sum of all of the \( m_i \) values (in other words, the total number of points within the whale).
Standard Deviation when M=1:

So, if we now define the mean value of A like this:

\[
\bar{A} = \frac{A_{box}}{N} \sum_{i=1}^{N} m_i
\]

And (by analogy with what we've done before) the standard deviation of m like this:

\[
s^2_m = \frac{1}{N-1} \left[ \sum_{i=1}^{N} m_i^2 - \frac{1}{N} \left( \sum_{i=1}^{N} m_i \right)^2 \right]
\]

We arrive at the following expression for the standard error of the mean, in the case where we have only one experiment with many points:

\[
\sigma_{\bar{A}} \approx \frac{A_{box}}{\sqrt{N}} s_m
\]
More Probability Distributions
The Bernoulli Distribution: Success or Failure?
The Bernoulli Distribution:

- Only two possible outcomes (true or false, success or failure).
- The probability, $p$, of one possible outcome is known.

$$P(\text{heads}) = p$$
$$P(\text{tails}) = 1 - p$$
The Bernoulli Distribution:

- Only two possible outcomes (true or false, success or failure).
- The probability, $p$, of one possible outcome is known.

$$P(\text{heads}) = p$$
$$P(\text{tails}) = (1-p)$$

Questions:
The Bernoulli Distribution:

- Only two possible outcomes (true or false, success or failure).
- The probability, $p$, of one possible outcome is known.

$$P(\text{heads}) = p$$
$$P(\text{tails}) = (1-p)$$

Questions:

What is the $p$ for tossing heads in one trial?
The Bernoulli Distribution

The Bernoulli Distribution:

- Only two possible outcomes (true or false, success or failure).
- The probability, $p$, of one possible outcome is known.

\[
P(\text{heads}) = p \\
P(\text{tails}) = (1-p)
\]

Questions:

What is the $p$ for tossing heads in one trial?

Say I have thrown ten heads in a row – what is the probability of doing that?
The Bernoulli Distribution

Questions:

What is the p for tossing heads in one trial?

Say I have thrown ten heads in a row – what is the probability of doing that?

After ten heads, what is probability that next one is heads too?
The Bernoulli Distribution

The outcomes need not have the same probability! For example:

Outcome 1: Roll a 6.

\[ P(6) = p = \frac{1}{6} \approx 0.17 \]

Outcome 2: Roll something else.

\[ P(\text{tails}) = (1-p) = \frac{5}{6} \approx 0.83 \]
The Binomial Distribution:
How Many Successes in Many Attempts?
The Binomial Distribution

What if we flip a coin many times, or if we flip many coins? How many instances of a given outcome (say, heads) should we expect?

The **Binomial distribution** gives the probability of observing a certain number of true results, $x$, after doing $n$ tests.

$p$ is the probability of success on each test.

![Graph showing binomial distribution with $n = 20$, $p = 0.5$]

It's used when we are interested in a number of TRUE/FALSE experiments.

Consider $n$ coins tosses:

TRUE = HEADS,
FALSE = TAILS

What is the probability of getting $x$ HEADs-up if you flip the coin $n$ times?
Understanding the Binomial Distribution:

- \( x \) = number of successes (e.g., how many heads?)
- \( n \) = number of trials (e.g., how many tosses?)
- \( p \) = probability of success in a single trial.

\[ P(x; n, p) = \frac{n!}{x!(n - x)!} p^x (1 - p)^{n-x} \]

- \( P(x; n, p) \) = Probability of seeing \( x \) successes after \( n \) trials, given probability \( p \) of success.

Number of possible ways to arrange the \( x \) successes and \( (n-x) \) failures.

Probability of getting \( x \) successes and \( (n-x) \) failures.
The Binomial Distribution

\[ P(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \]

\*P(x;n,p) = Probability of seeing x successes after n trials, given probability p of success.\*

1. Consider a single fair 6-sided die
2. Define “success” as rolling a 5
3. Roll the dice 10 times – what is probability of rolling *three* 5’s?
   - n=10, p=1/6, x=3
   - prob(5) = 1/6 hence prob(three 5’s) = (1/6)^3
   - prob(else) = 1/6 hence prob(seven non-5’s) = (5/6)^7
   - But there are many ways to get three 5’s in 10 rolls:
     - 5-5-5-3-2-6-1-2-4-1
     - 5-5-5-3-3-3-2-2-4-6
     - ...
   - The combination term “n choose x” calculates how many such combinations exist. Here “n choose x” = 10 choose 3 = 10! / (3! (10-3!)) = 120
The Binomial Distribution

Binomial Distribution for One Fair Coin Flip:

Here's the Binomial distribution with:
- \( n = 1 \) (one coin flip)
- \( p = 0.5 \), 50% chance of heads on a flip
- \( 1 - p = 0.5 \), 50% chance of tails on a flip

One toss, 50% heads/tails

\[
P(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}
\]

x = 0: getting zero heads
x = 1: getting one head
The Binomial Distribution

**Binomial Distribution for One Biased Coin Flip:**

Here's the Binomial distribution with

\( n = 1 \) (one coin flip)
\( p = 0.3, \) 30% chance of heads on a flip
\( 1 - p = 0.7, \) 70% chance of tails on a flip

One toss, 30% heads/70% tails

\[ P(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} \]
The Binomial Distribution

Binomial Distribution for 20 Fair Coin Flips:

20 tosses of a fair coin

\[ P(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} \]

\[ P(x; 20, 0.5) \]

Notice that the mean value for \( x \) is 10. Could we have predicted this?
The Binomial Distribution

Mean and Variance for the Binomial Distribution:

Given the binomial distribution:

\[ P(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} \]

We can calculate the mean value (\( \mu \)) of \( x \):

\[ \mu = np \]

and the variance (\( \sigma^2 \)) of \( x \):

\[ \sigma^2 = np(1 - p) \]
The Binomial Distribution

The Effect of Varying $p$:

20 tosses, vary $p$ from 5% to 50%

- $p = 5\%$: Very lopsided coin.
- $p = 30\%$
- $p = 50\%$: Fair coin.
A Special Case of the Binomial Distribution

The Poisson Distribution: How Many Successes in MANY MANY MANY Attempts if \( p \) is REALLY SMALL?
An interesting special case of the binomial distribution is the one in which:

- The number of trials, $n$, approaches infinity,
- The probability of success, $p$, approaches zero,
- The mean number of successes, $\mu = np$, remains fixed.

As these limits are approached, the binomial distribution can be approximated by the following (much simpler) expression:

$$P(x; \mu) = \frac{\mu^x}{x!} e^{-\mu}$$

This is called the Poisson Distribution, and it is valid when $p$ is small, $n$ is large and $\mu$ is some intermediate value.
The Poisson Distribution

How Good is the Poisson Approximation?

binomial distribution, $n=20$, $p=0.05$

Poisson distribution, $\mu=1$

$P(x; 20, p)$

$n = 20$
$p = 0.05$
$\mu = np = 1$

$x = Number of heads$
The Poisson Distribution

• Rule of thumb:
  – The Poisson distribution is **fairly good** approximation of the Binomial distribution if
    • \( n > 20 \)
    • \( p < 0.05 \)
  – The Poisson distribution is an **excellent** approximation of the Binomial distribution if
    • \( n \geq 100 \)
    • \( np \leq 10 \)

  – So with just a moderate number of trials and reasonably small probability of success, the Poisson is good.
Example: The Poisson Distribution

Radioactive Decay:

Consider a sample of a radioactive material like Uranium.

We point a detector at the sample and count the number of radioactive decays that happen during a five-minute period.

- \( n \) is large, on the order of Avogadro's Number.
- \( p \) is small (it's unlikely that a particular nucleus will decay while we're looking at it).
- \( \mu = np \), the average number of decays in five minutes, is still an appreciable number, since \( n \) is so large.

The results of several 5-minute observations would be Poisson distributed.
Example: The Poisson Distribution

http://www.engineerguy.com/videos/video-lines.htm
The Poisson Distribution

Mean and Variance of the Poisson Distribution:

Given the Poisson distribution:

\[ P(x; \mu) = \frac{\mu^x}{x!} e^{-\mu} \]

We can derive the very simple expression for the variance of \( x \):

\[ \sigma^2 = \mu \]

So the standard deviation is:

\[ \sigma = \sqrt{\mu} \]
The Poisson Distribution and Histograms

Consider the following:

We fill a histogram with a large number of entries, \( n \).

The probability, \( p \), that any given entry will land in a particular bin is small.

This implies that we can use Poisson statistics to describe the variations in the number of counts in a given histogram bin.

If the count in a given bin is \( m \), then the best estimate of the uncertainty in the bin count is \( \sigma = \sqrt{m} \).
The Poisson Distribution and Histograms

Histogram with Error Bars:
The Gaussian Limit
The Poisson Distribution For Large $\mu$

**Poisson Distribution for $\mu >> 0$:**

Atoms out of 50 decaying w/in 0.1, 0.5, 1 half life

- $\mu = 3.35$
- $\mu = 14.6$
- $\mu = 25$

As $\mu$ increases, the distribution approaches a Gaussian ("normal") shape.
The Poisson Distribution For Large $\mu$

How Good is this Approximation?

Atoms out of 50 decaying w/in 0.1, 0.5, 1 half life

- $\mu = 3.35$
- $\mu = 14.6$
- $\mu = 25$

Gaussian approximation.
As we noticed earlier, the Binomial distribution itself looks like a Gaussian when the mean is sufficiently far away from zero.

\[ \mu = np >> 0 \]

20 tosses of a coin, vary P(heads) 5%-50%
Calculating the binomial distribution for large $n$ is hard...

DeMoivre figured out that large $n$ behavior is as a Gaussian.

$\mu = np \gg 0$
Relations Between Distributions

\[ P(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1 - p)^{n-x} \]

\[ P(x; \mu) = \frac{\mu^x}{x!} e^{-\mu} \]

**Binomial** → **Poisson**

- \( n \rightarrow \text{Large}, \ p \rightarrow \text{Small} \)
- \( n \rightarrow \text{Large}, \ np = \mu \gg 0 \)
- \( \mu \gg 0 \)

**Gaussian**

\[ P(x; \mu; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]
Relations Between Distributions

**Picking the Right Distribution:**

Need full Binomial description when there is a good chance you see events at both limits: $x=0$ and $x=n$.

Can use Poisson limit when you are likely to observe events at $x=0$, but not likely to see events at $x>>\mu$ approaching upper limit.

- **Binomial**
  - $n \rightarrow $ Large, $p \rightarrow $ Small
  - $n \rightarrow $ Large, $np = \mu >> 0$

- **Poisson**
  - $\mu >> 0$

- **Gaussian**

Can use Normal/Gaussian limit when number of trials is large and observed data are not likely to land near either the upper or the lower limits.
We’ll pick up from here next time.

See you Thursday at labs!