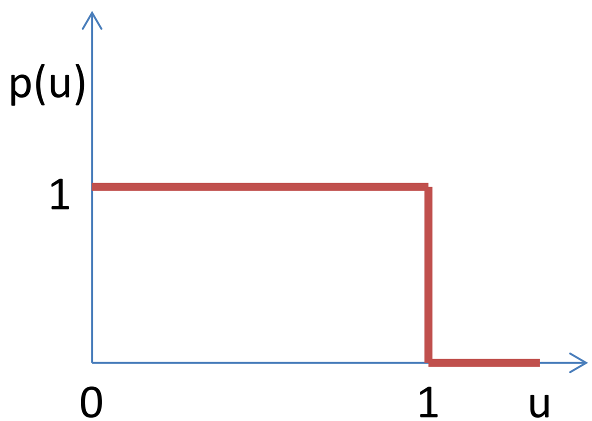
11 – Some Important Probability Distributions

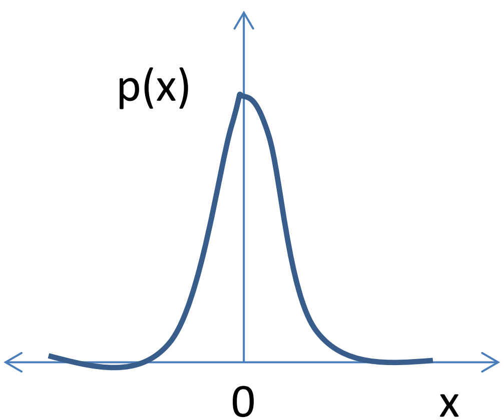
Phys 281 – Class 11

Grant Wilson

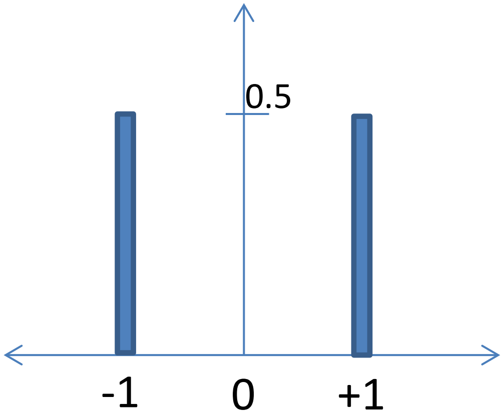
A “random” process does not imply that all probabilities are equal.

A “random” process does not imply that all probabilities are equal.

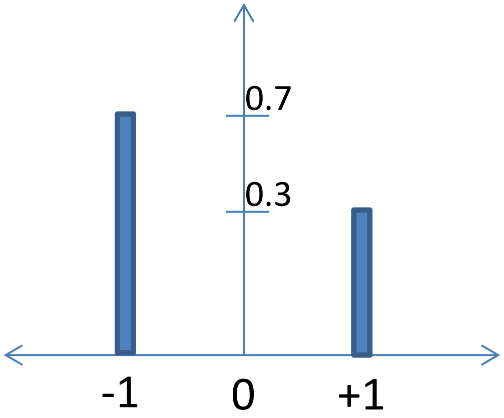




A “random” process does not imply that all probabilities are equal.



A “random” process does not imply that all probabilities are equal.



Gaining Insight Through Simulations

- It is not enough to analytically work your way through problems.
- It is also not enough to simulate your way through problems.
- The real power and insight comes when you can connect the two realms.

Exercise – Coin Flipping

- Given a “biased coin” with a probability of turning up heads = 0.3, what is the probability of obtaining 3 heads in six flips of the coin.

Brute-Force Answer

```
#what is probability of achieving n-heads of a coin with bias, b, after s tosses?
#n - number of heads achieved
#b - probability that heads will come up
#s - number of tosses of the coin
#nTrials - number of times we run the experiement

nTrials = 100000.
b = 0.3
s = 6.
n = np.zeros(nTrials)
for i in range(int(nTrials)):
   u = np.random.uniform(0,1,s)
   x = np.where(u<=0.3)
   n[i] = len(x[0])*1.

#at this point, n is an array of the number of heads found in each set of 6 coin-flips
#all that's left is to count up how many times n=3
w = np.where(n == 3)
nThree = len(n[w])*1./nTrials
```

Exercise – Coin Flipping

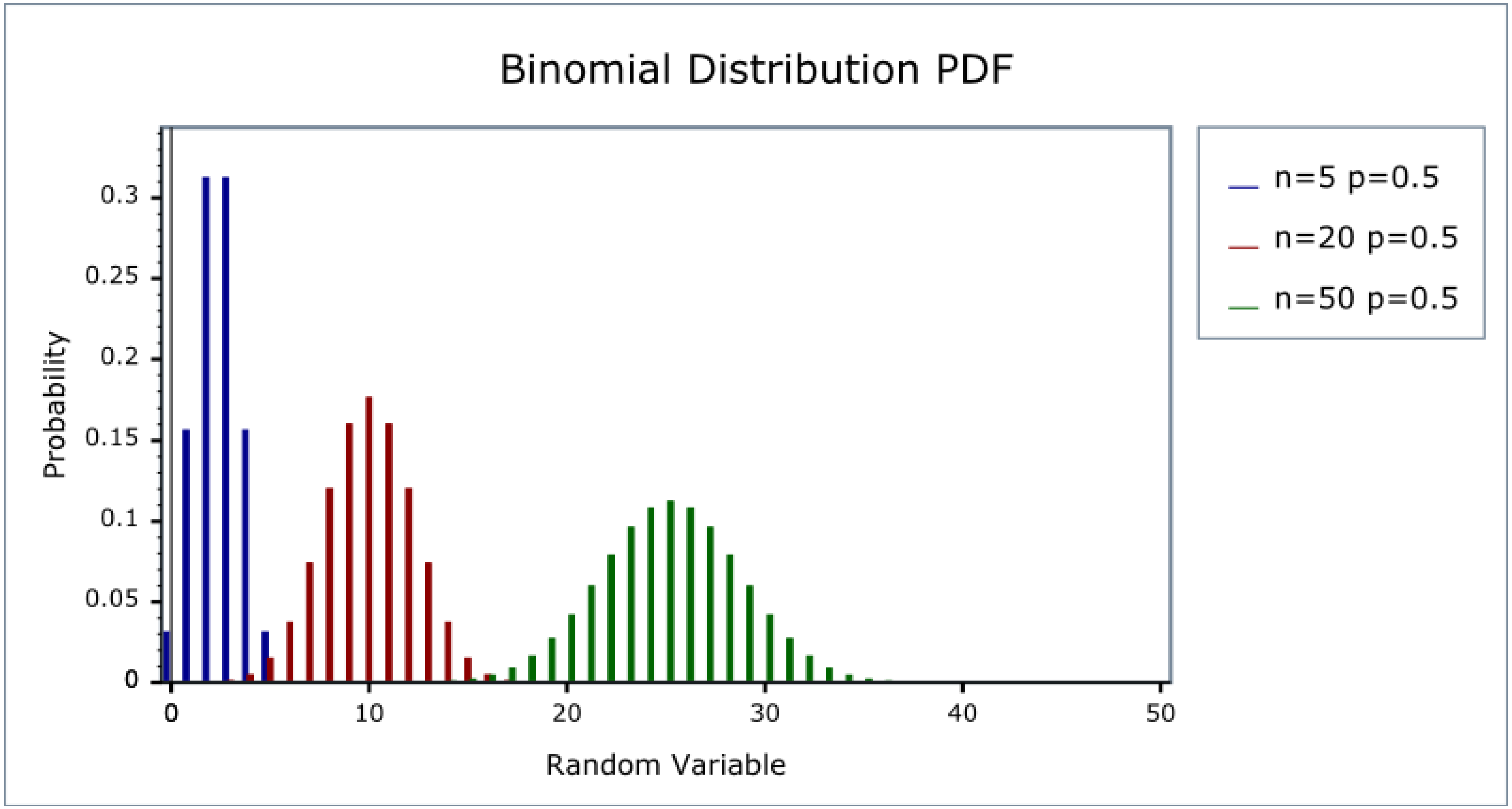
- Given a “biased coin” with a probability of turning up heads = 0.3, what is the probability of obtaining 3 heads in six flips of the coin.

This is an example of a random process governed by the “binomial distribution.”

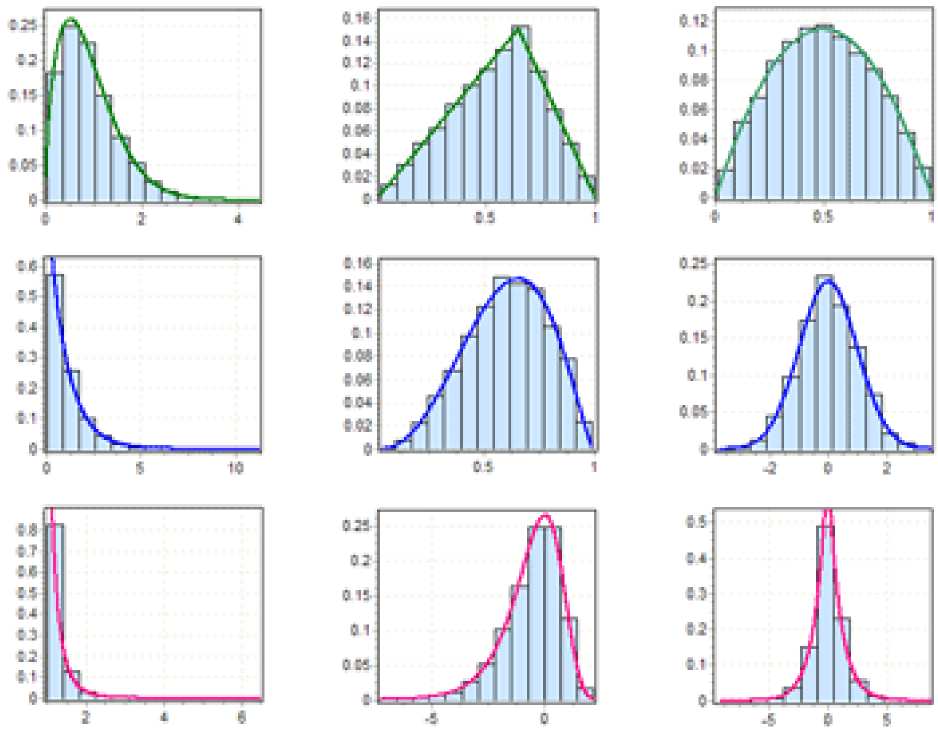
The Binomial Distribution

- Relevant when the result of an experiment can be described as yes/no or success/failure outcome of a trial and the probability of obtaining success is known.
  - e.g.,
    - What is the probability of obtaining 2 heads in 3 flips of a coin?
    - What is the probability of exactly two out of six rolled dice coming up with 1 facing up?
    - If I toss 10 coins 100 times, what is the mean number of heads? What's the standard deviation of the number of heads?

Binomial Distribution

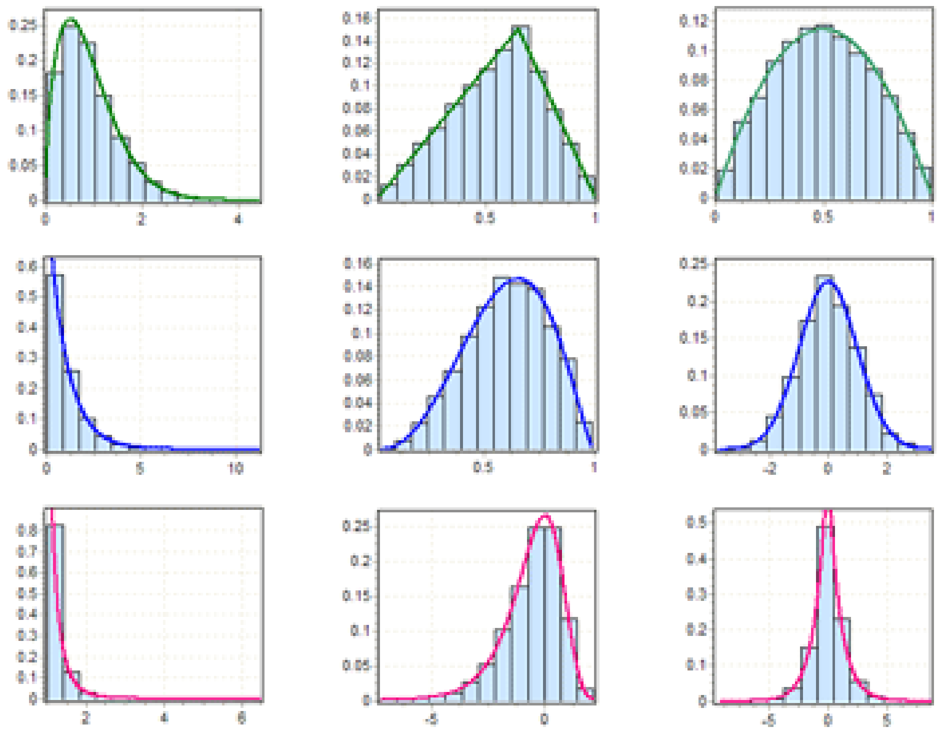


An Aside – Characterizing probability distributions



Given a probability distribution, what metrics can we use to characterize it?

An Aside – Characterizing probability distributions



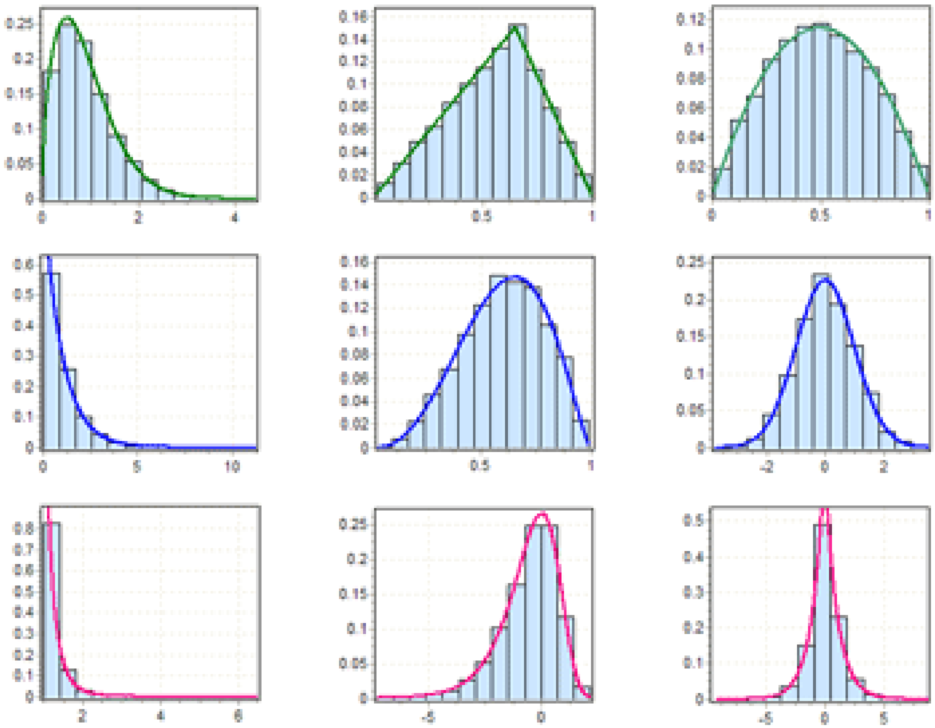
Given a probability distribution, what metrics can we use to characterize it?

Easiest to think about:

1. The mean value

2. The standard deviation.

An Aside – Characterizing probability distributions



Given a probability distribution, what metrics can we use to characterize it?

Easiest to think about:

1. The mean value, μ
2. The standard deviation, σ

For a series of N-deviates, x_i, drawn from a given probability distribution:

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

or in Python notation: $\mu =$ x.mean(), $\sigma = x.std()$

The Binomial Distribution

- The probability, P_B of observing ν successes in N trials, where the probability of success per trial is p, is given by

$$P_B(\nu, N; p) = \frac{N!}{\nu!\,(N-\nu)!} p^{\nu}(1-p)^{N-\nu}$$

- The average number of successes, $\bar{\nu}$, is given by

$$\bar{\nu} = Np$$

- The standard deviation of the number of successes is

$$\sigma_{\nu} = \sqrt{Np(1-p)}$$

The example in the python documentation is illustrative

- A company drills 9 wild-cat oil exploration wells, each with an estimated probability of success of 0.1. All nine wells fail. What is the probability of that happening?
    - p = 0.1
    - nAttempts = 9
    - # success = 0

```
#Draw 20,000 deviates from a binomial distribution with n=9 and p=0.1.
#Count the number of deviates equal to zero and divide by 20,000.
```

Answer

#draw 20,000 deviates from a binomial
#distribution with n=9 and p=0.1

import numpy as np

b = np.random.binomial(9,0.1,20000)

#find the indexes of b where b_i=0

w = np.where(b == 0)

#count the elements of b that are zero

count = len(b[w])

#divide by the number of trials in the simulation to get the percentage

answer = count/20000.

The Poisson Distribution

- We can flip the binomial distribution on its head and ask, given a **mean rate** of events where:

    1. the event is something that can be counted in whole numbers,
    2. the occurrences are independent,
    3. the average frequency of occurrence for the time period in question is known; and
    4. it is possible to count how many events **have** occurred but meaningless to count how many events **have not** occurred,

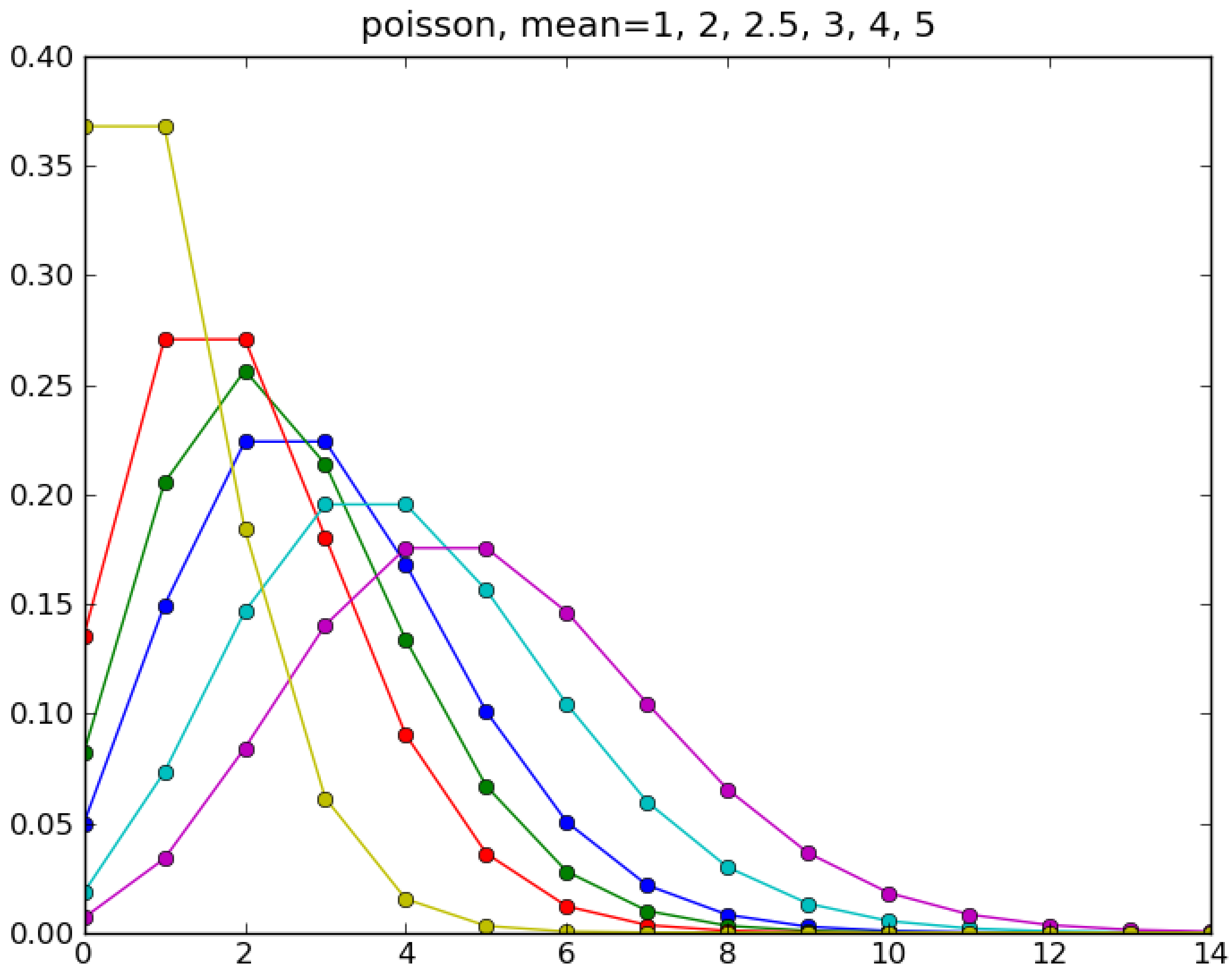
  What is the likelihood (probability) of some particular number of events occurring?

The Poisson Distribution

- The Poisson probability distribution is

$$p(k; r) = \frac{r^k}{k!} e^{-r}$$

  - where
    - r = average rate of occurrence of some event per module (usually of time)
    - k = the actual number of events we wish to know the probability for


poisson, mean=1, 2, 2.5, 3, 4, 5
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
12
14

Poisson Distribution Example

- Graduate student admissions – In Astronomy we typically seek a class of 4 graduate students. On average, 1/3 of the students we admit each year actually accept and so we accept 12 students to the program. What fraction of years do we actually get 0 students, 4 students, and 8 students?

$$p(k;r) = \frac{r^k}{k!}e^{-r}$$

r = average rate of occurrence of some event per module (usually of time)

k = the actual number of events we wish to know the probability for

Answer

- The mean rate of acceptance is 4 per year.

$$r = 4$$

- We want the Poisson probabilities for k=0, 4, and 8

$$f(0,4) = \frac{4^0 e^{-4}}{4!} = 0.02 = 2\%$$

$$f(4,4) = \frac{4^4 e^{-4}}{4!} = 0.18 = 18\%$$

$$f(0,8) = \frac{4^0 e^{-4}}{4!} = 0.02 = 2\%$$

Exercise – Poisson Distribution

- The average number of whales seen during a whale cruise out of Boston is 5. What is the probability that on a given cruise you see more than 8 whales?

(Do this problem both by using the analytic Poisson distribution and by drawing Poisson deviates using numpy.random.poisson)

The mean and standard deviation of the Poisson distribution

- The mean of the Poisson distribution is

$$\mu = r$$

- The standard deviation of the Poisson distribution is

$$\sigma = \sqrt{r}$$

Implications!

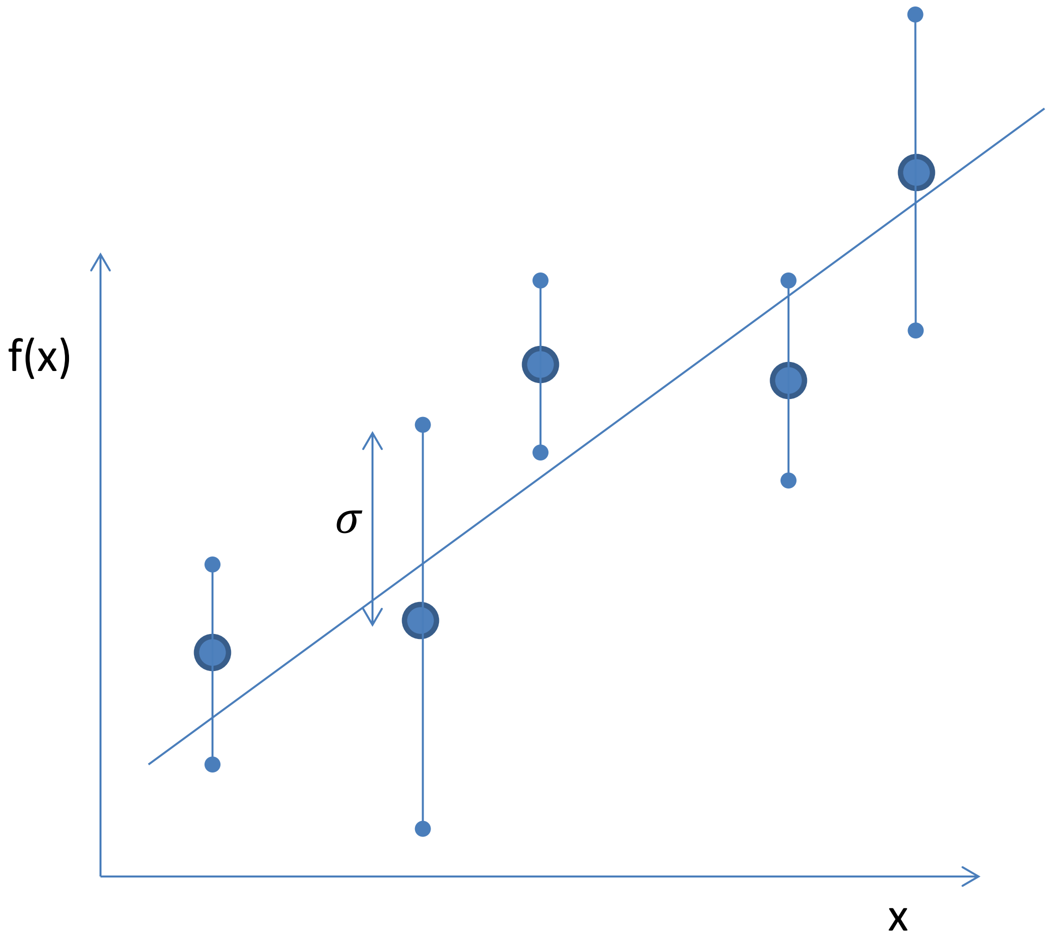
Exercise

- The average rate of muons hitting the Earth's surface is 1/minute/cm^2. Suppose you have a 1cm^2 muon detector. How long would you expect to have to wait to measure the actual muon rate with an accuracy of 1%?

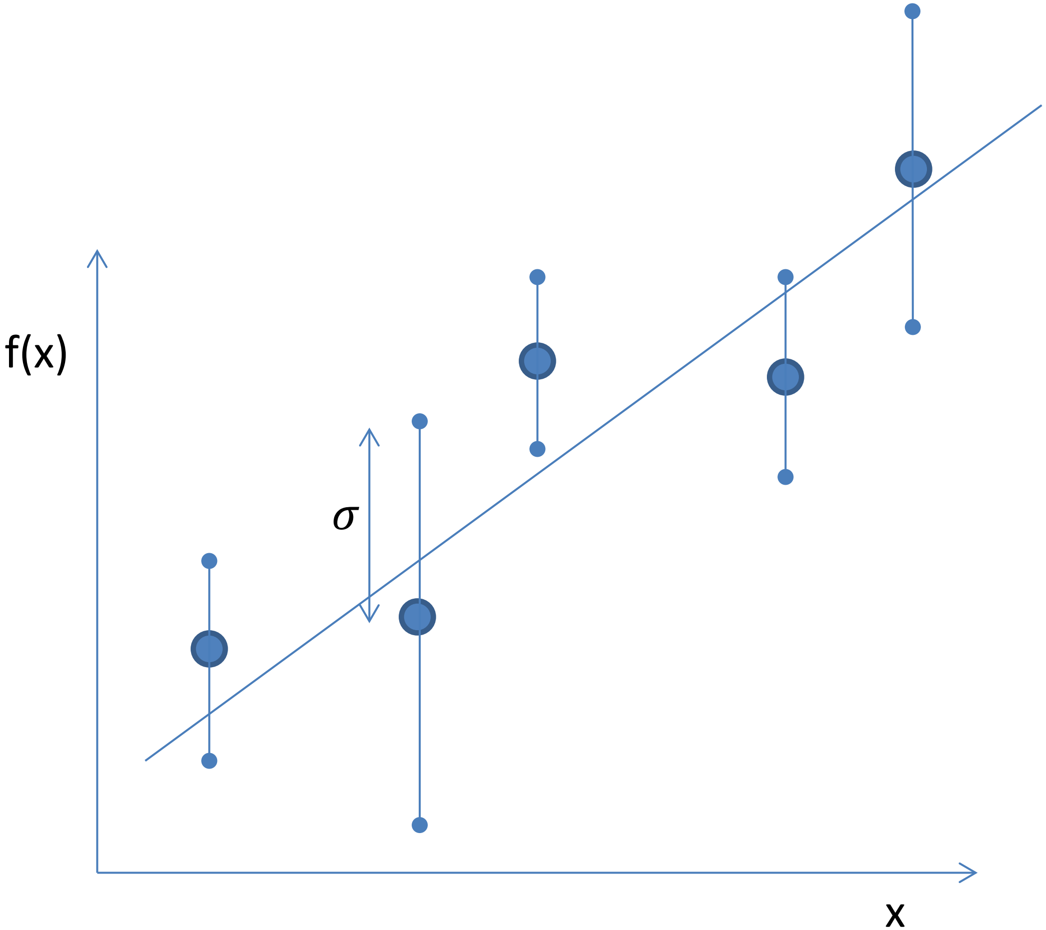
The χ^2- Distribution

- There are many ways to discuss this, I will approach this from the perspective of data analysis.

- Suppose you have N-measurements of some quantity y(x) where
  1. the measurements are independent
  2. the measurement errors are normally distributed with standard deviations σ_i



then $$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i, a))^2}{\sigma_i^2}$$ is drawn from a chi-square distribution with ν degrees of freedom where $\nu = \mathrm{N} -$ the number of free parameters, a.



What is this?

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - f(x_i, a))^2}{\sigma_i^2}$$

Key Features of χ^2- distribution

- ν is called the number of degrees of freedom

- The mean of a χ^2- distribution with ν degrees of freedom is:

$$\mu = \nu$$

- The standard deviation of a χ^2- distribution with ν degrees of freedom is:

$$\sigma = \sqrt{2\nu}$$

Exercise

- Recover a chi-square distribution with 18 degrees of freedom by following these steps:
  1. create a line with 20 x-values – your line should have a user-input slope, a, and y-intercept, b.
  2. take the y-values of your line and add noise drawn from a normal distribution with mean = 0 and standard deviation, σ = 2.
  3. compute the χ^2statistic for your line and values in step 2.
  4. repeat steps 2 and 3, 10,000 times, saving all the values of χ^2
  5. make a histogram of the χ^2 values. Overplot the chisq function (use scipy.stats.chi2.pdf()) for 18 degrees of freedom to show that they match.