Stochastic Programs, Probability and Distribution

Chapter 17

You don’t have to be so stochastic

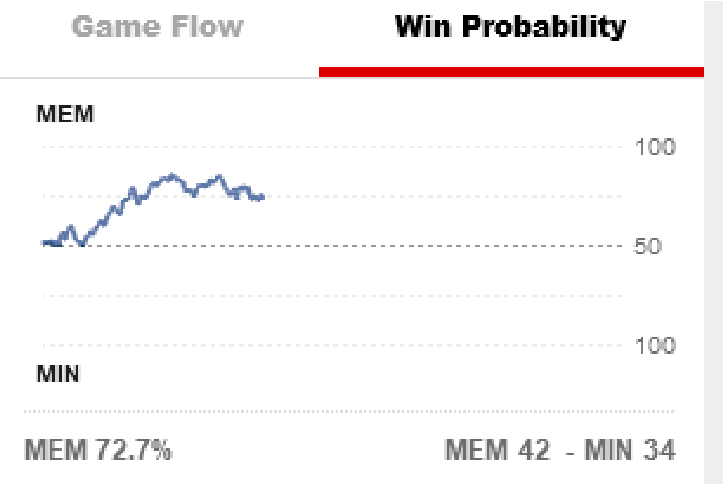
- What do we mean when we say “stochastic”
  - Describes a value or event that occur at random or is randomly distributed

- Newtonian physics:
  - If A then B
- Quantum physics:
  - If A then maybe B (or maybe C or maybe both?)

- No such thing as *homo* economicus

Nondeterminism

- Causal nondeterminism
  - Not every event is caused by a previous event
  - Bohr/Heisenberg
- Predictive nondeterminism
  - It is impossible to make precise predictions about future states
  - Einstein/Schodinger



Stochastic problems

- Games of chance
  - Gambling
  - Is that car really going to turn?
  - Dating
  - Games
- Predictions
  - Weather
  - Stocks
  - Actuary

Randomness

- What is the output of `roll_n(10)`?

```
def roll_die():
    """Returns a random integer between 1 and 6"""
    return random.choice([1,2,3,4,5,6])

def roll_n(n):
    result = ''
    for i in range(n):
        result = result + str(rollDie())
    return result
```

- How likely are we to see s-s-s-snake eyes (1111111111)

Calculating simple probabilities

- What is **probability**
  - The fraction of all possible results that have a given outcome
  - Always represented as a number between 0 and 1
    - 0 = absolute impossibility
    - 1 = absolute certainty
  - Multiply by 100 to get a percentage
- The probability of an event **not** happening
  - 1 – probability it will happen

Independent probability

- No event is **dependent** on the result of another event
- The roll of 1 die does not depend on the previous roll
- Probability of two independent events occurring
  - Probability of event 1 $\times$ Probability of event 2
  - Multiplicative law
- Selection **with** replacement

Craps – what is the probability of 7

- Probability of the two rolls totaling 7
- Each die is independent
  - 1 ● 6
  - 2 ● 5
  - 3 ● 4
  - 4 ● 3
  - 5 ● 2
  - 1 ● 6
- 6(1/6 * 1/6) = 1/6 = 0.166667

Dependent probability

- The probability of an event depends on prior events
- Dealing cards
  - Probability the first card is an ace
    - 4/52 = 1/13 = 0.0769
  - Probability the second card is also an ace
    - 3/51 = 0.0577
  - Probability the first two cards dealt are both aces?
    - Multiplicative law still applies
    - 4/52 * 3/52 = 0.0044

Inferential statistics

- Guiding principle:
  - A random sample tends to have the same properties as the population from which it was drawn
- So, we can **infer** about the whole based on the sample
  - If our sample is good (Chapter 19)
- Chance of two aces in a row? 0.0044 (0.44%)
  - Happens once – wow, that is rare
  - Happens twice – something might be fishy
  - Happens thrice - prestidigitator

Flipping a coin

- Fair coin
  - Probability of heads = Probability of tails = 0.50

```
def flip(num_flips):
    heads = 0.0
    for i in range(num_flips):
        if random.random() < 0.5:
            heads += 1
    return heads/num_flips

def flip_sim(numFlipsPerTrial, numTrials):
    fracHeads = []
    for i in range(numTrials):
        fracHeads.append(flip(numFlipsPerTrial))
    mean = sum(fracHeads)/len(fracHeads)
    return mean
```

What is “average”

- Measures of **central tendency**
  - **Mean** – total / number of samples
    - Expressed as ‘mu’ - μ
  - **Median** – mid point between the largest and smallest samples
    - Same number of samples before and after the mid-point
  - **Mode** – Most common value

Law of large numbers

- Bernoulli's theorem
  - in repeated independent experiments with the same probability (*p*) of the expected value, the chance the number of times differs from *p* converges to 0 as the number of experiments goes to infinity.
- Regression to the mean
  - Following an extreme event
  - The next random event will more *likely* be *closer* to the mean
- Gambler's fallacy
  - Following an extreme even
  - Things will "even out"

Example:
If I flip a coin 6 times and get 6 heads

Gambler's fallacy

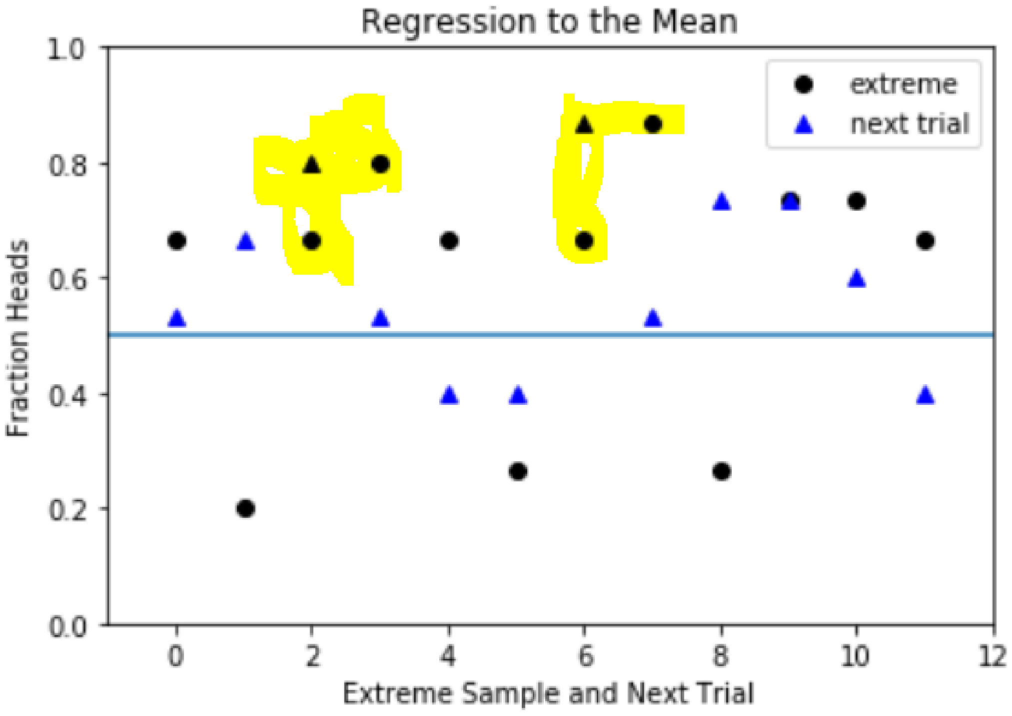
- The next 6 flips will have more tails than heads

Regression to the mean

- The next 6 flips will have some heads and some tails, but still may have more heads than tails

Plotting regression to mean

```
def regress_to_mean(num_flips,
num_trials):
  frac_heads = []
  for t in range(num_trials):
    frac_heads.append(flip(num_flips))
    extremes, next_trials = [], []
  for i in range(len(frac_heads) - 1):
    if frac_heads[i] < 0.33 or
       frac_heads[i] > 0.66:
      extremes.append(frac_heads[i])
      next_trials.append(frac_heads[i+1])
  #Plot results
  plt.plot(range(len(extremes)),
           extremes, 'ko',
           label = 'Extreme')
  plt.plot(range(len(next_trials)),
           next_trials, 'k^',
           label = 'Next Trial')
```



Question?

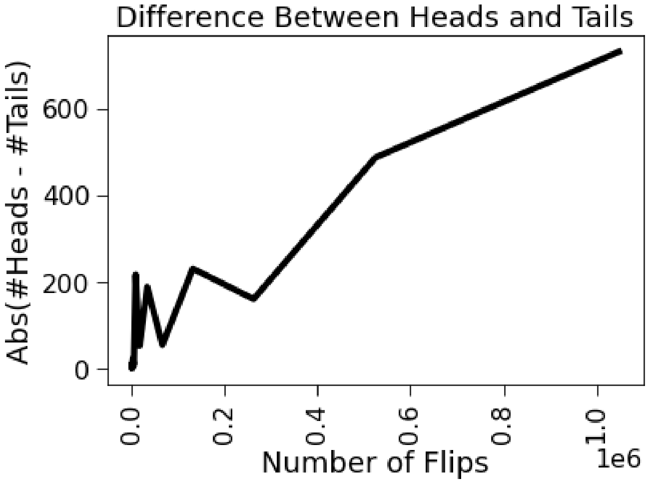
- Sally averages 5 strokes per hole when she plays golf.
- One day, she took 40 strokes to complete the front nine.
  - Averaging 4.444 strokes per hole
- Her partner suggests that should would probably regress to the mean and shoot a 50 on the back nine.

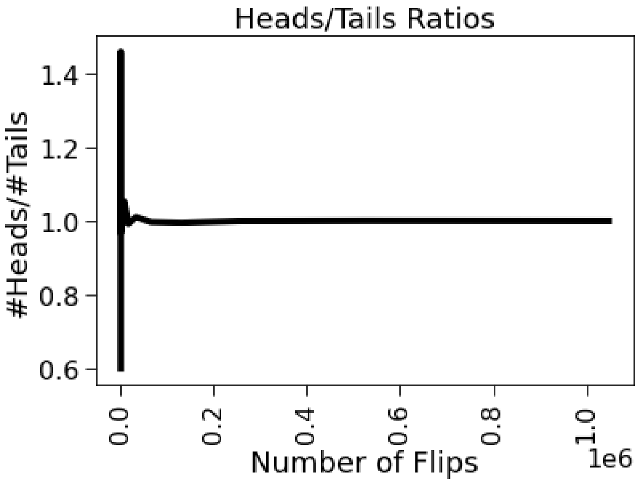
- What to you think?

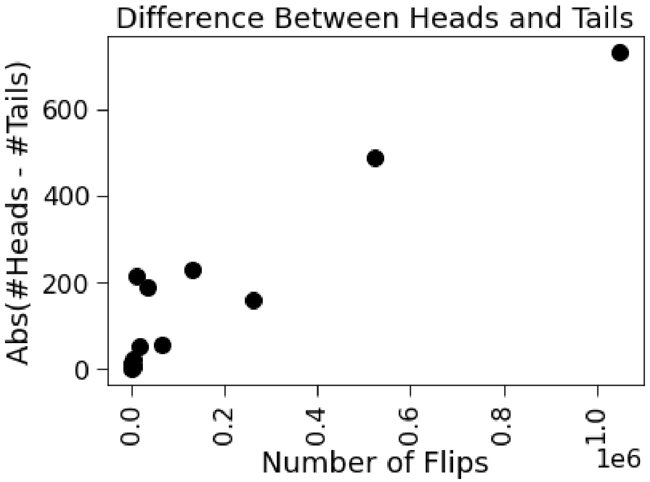
Differences between heads and tails

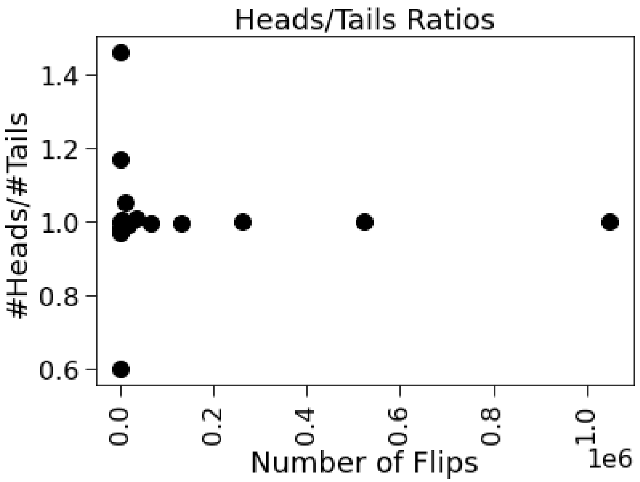
```
def flip_plot(min_exp, max_exp, style='k'):
    """Assumes min_exp and max_exp positive ints; min_exp < max_exp
       Plots results of 2**min_exp to 2**max_exp coin flips"""
```

Plotting the absolute difference and ratios










```
16
32
64
128
256
512
1024
2048
4096
8192
16384
32768
65536
131072
262144
524288
1048576
```

Variance

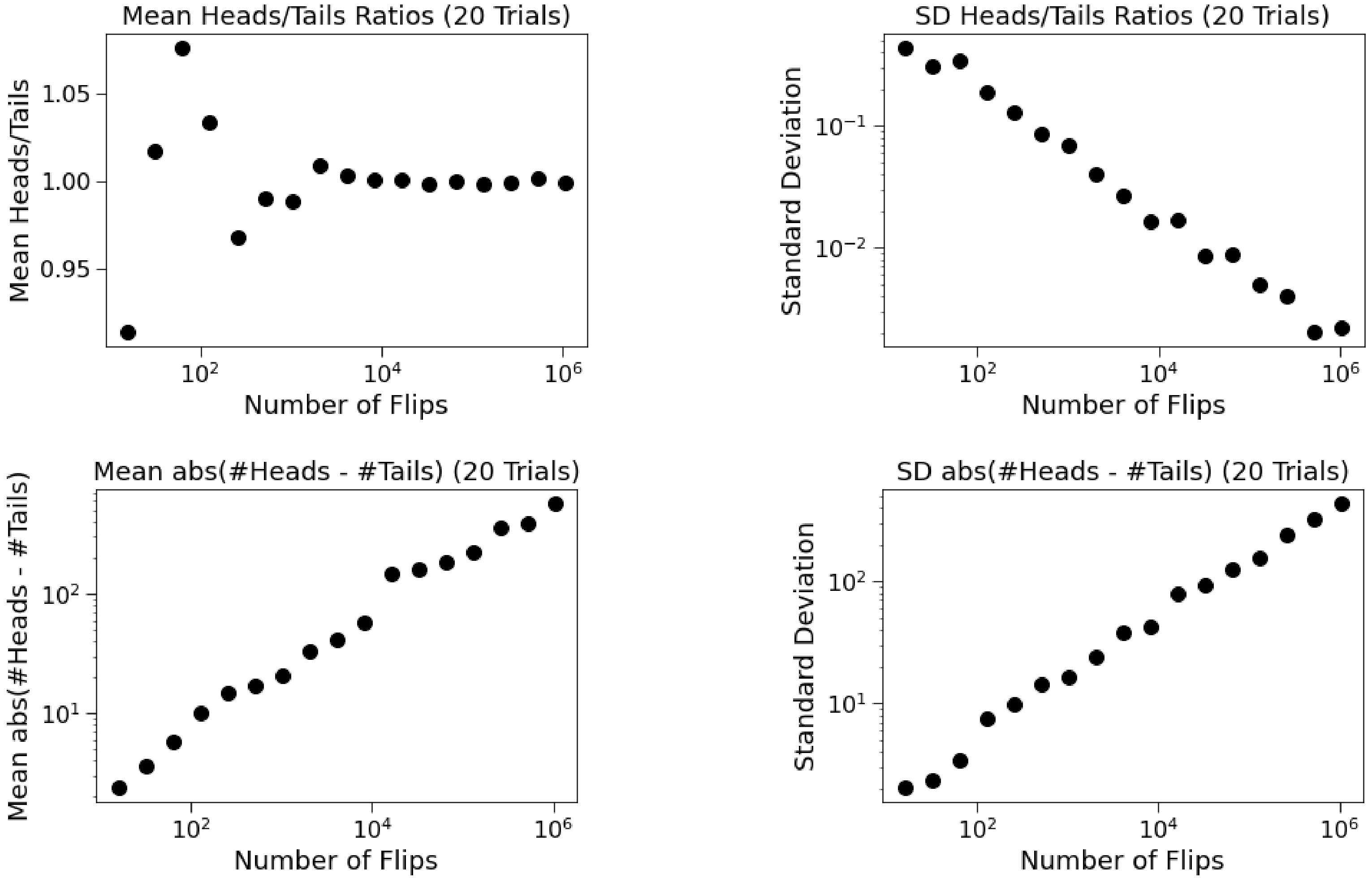
- Measure of how much **spread** there is in the data
  - How much it varies
- $variance(X) = \frac{\sum_{x \epsilon X}(x - \mu)^2}{|X|}$
- Informally, the fraction of values that are close to the mean

Standard Deviation

- Square Root of the Variance
- It is in the same scale as the variable under consideration
- σ , s , sd

- We can use the relationship between the sample size and standard deviation to help determine how much confidence we have in a result

Standard deviation for absolute difference and ratios

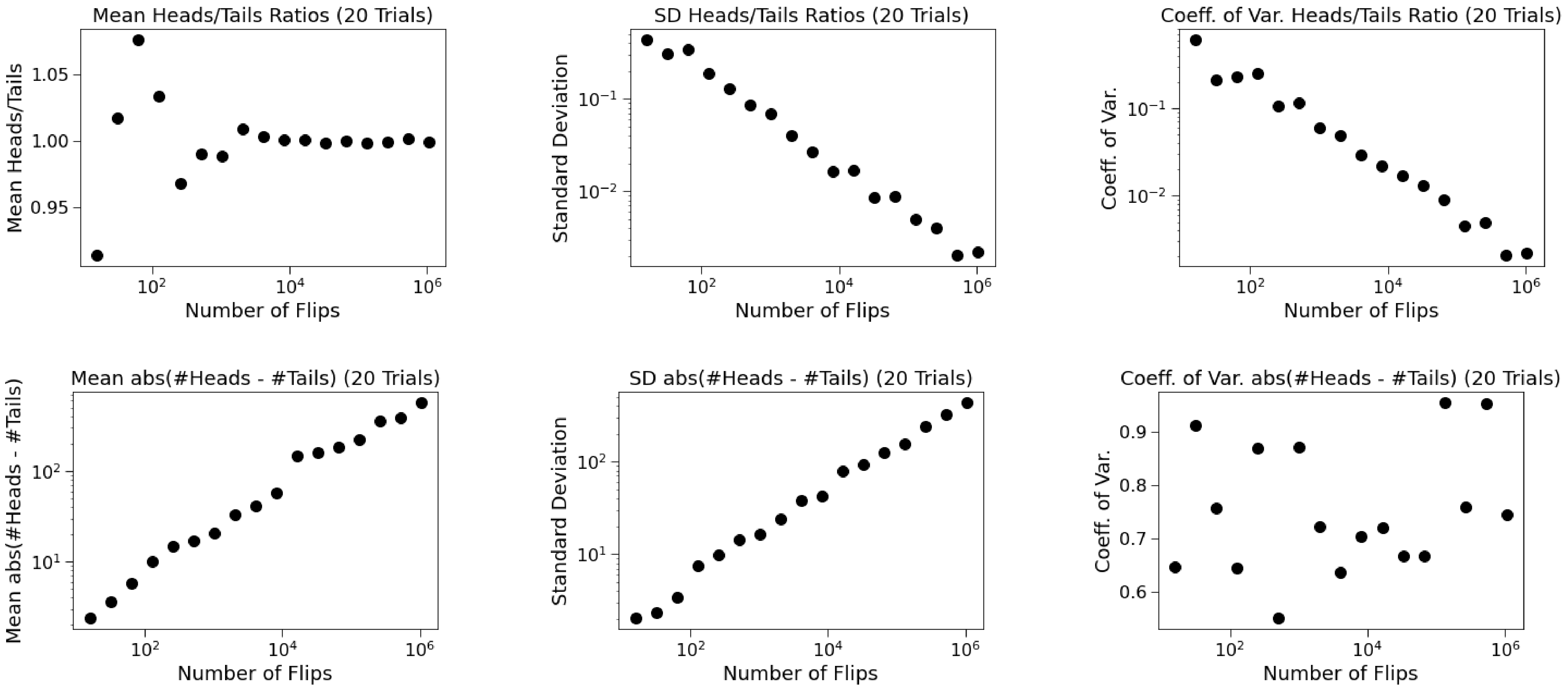


Coefficient of Variance (CV)

- **Coefficient of Variance** shows the relationship between the mean and the standard deviation
  - Why not variance?

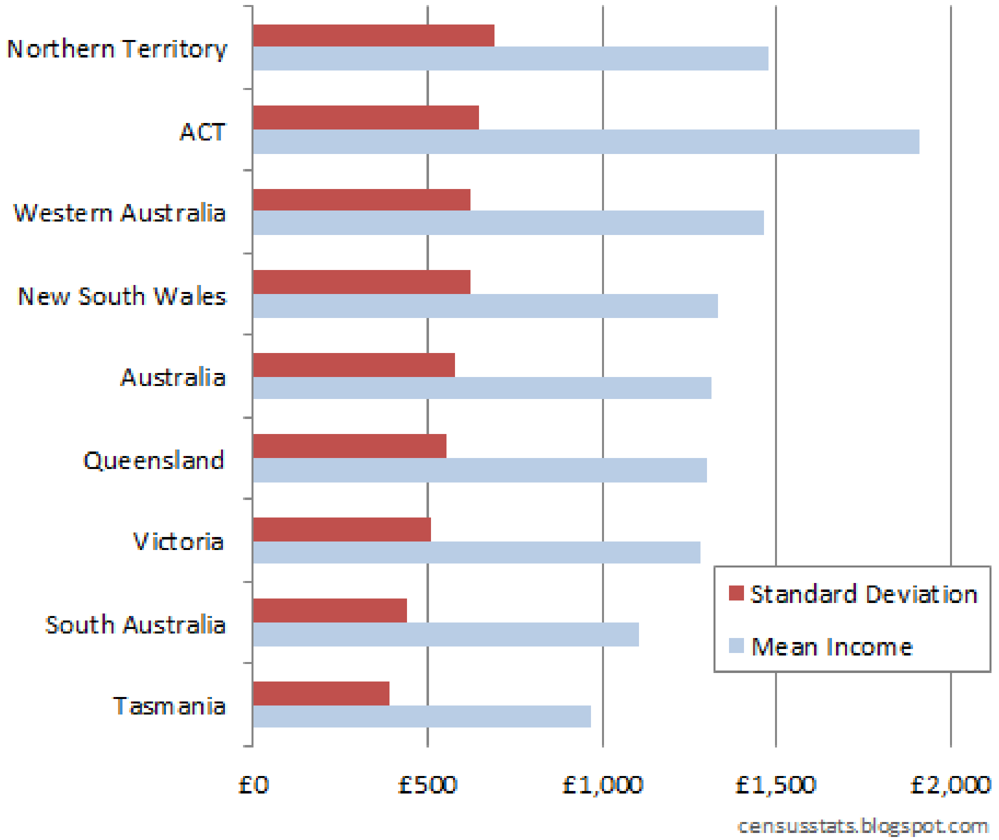
- In general. Distributions with CV < 1.0 are considered low variance

Adding in CV to our plots



CV - Continued

- Allows comparison of sets with different means



CV ACT= 0.32

CV Tas = 0.42

Confidence?

- As mean approaches 0, CV varies considerably
- Mean of 0 does not have a value for CV

- We can build a confidence interval with SD but not CV

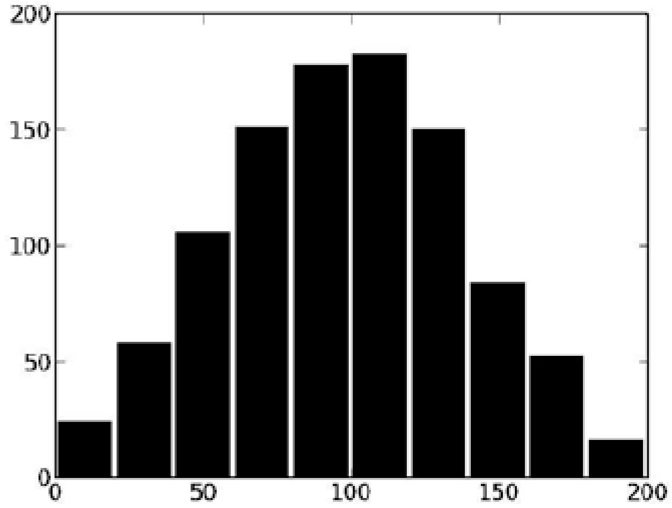
Distributions

- Probability
- Normal
- Continuous and Discrete Uniform Distributions
- Binomial and Multinomial
- Exponential and Geometric
- Benford’s Distribution

Histograms

- A plot designed to show the distribution of values in a set of data
  - Not the same as a bar chart
  - Values are sorted
  - Divided into equal-width bins
- Frequency Distribution

```
vals = []
for i in range(1000):
    num1 = random.choice(range(0, 101))
    num2 = random.choice(range(0, 101))
    vals.append(num1+num2)
plt.close()
plt.hist(vals, bins = 10)
plt.xlabel('Number of Occurences')
plt.show()
```



Probability Distribution

- Relative frequency
- Shows the **probability** of a random value falling within a specific range
  - As opposed to raw counts
- Discrete (Random Variable)
  - Possible values are finite (e.g. roll of a die)
- Continuous (Random Variable)
  - Infinite values within a range (e.g. speed of a car)

Discrete Probability Distribution

Point	Possible Rolls						Count	Probability
1							0	0
2	1, 1						1	0.027777778
3	1, 2	2, 1					2	0.055555556
4	1, 3	2, 2	3, 1				3	0.083333333
5	1, 4	2, 3	3, 2	4, 1			4	0.111111111
6	1, 5	2, 4	3, 3	4, 2	5, 1		5	0.138888889
7	**1, 6**	**2, 5**	**3, 4**	**4, 3**	**5, 2**	**6, 1**	**6**	**0.166666667**
8		2, 6	3, 5	4, 4	5, 3	6, 2	5	0.138888889
9			3, 6	4, 5	5, 4	6, 3	4	0.111111111
10				4, 6	5, 5	6, 4	3	0.083333333
11					5, 6	6, 5	2	0.055555556
12						6, 6	1	0.027777778

Continuous Probability Distribution

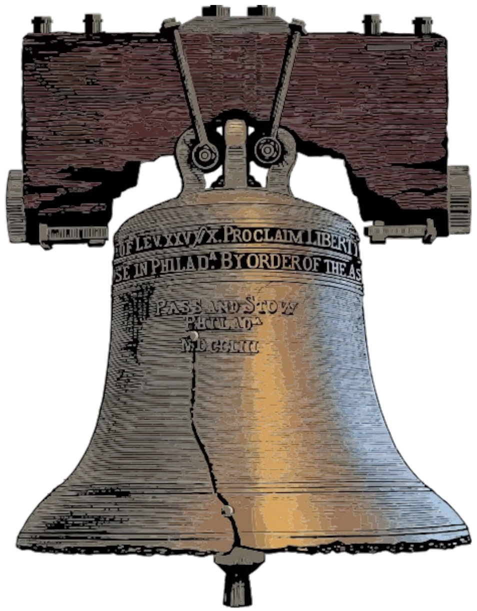
- What is the probability that someone is driving 59.1567853214787421487845245748975121877874 mph?
  - Likely 0
- We can create a **Probability Density Function (PDF)**
  - Describes the probability of a random variable being between two values
    - What is the probability that someone is driving between 59 and 60 mph?
  - Defines a curve
  - The probability of a *rv* occurring between x_1 and x_2 is the area under the curve between those two points

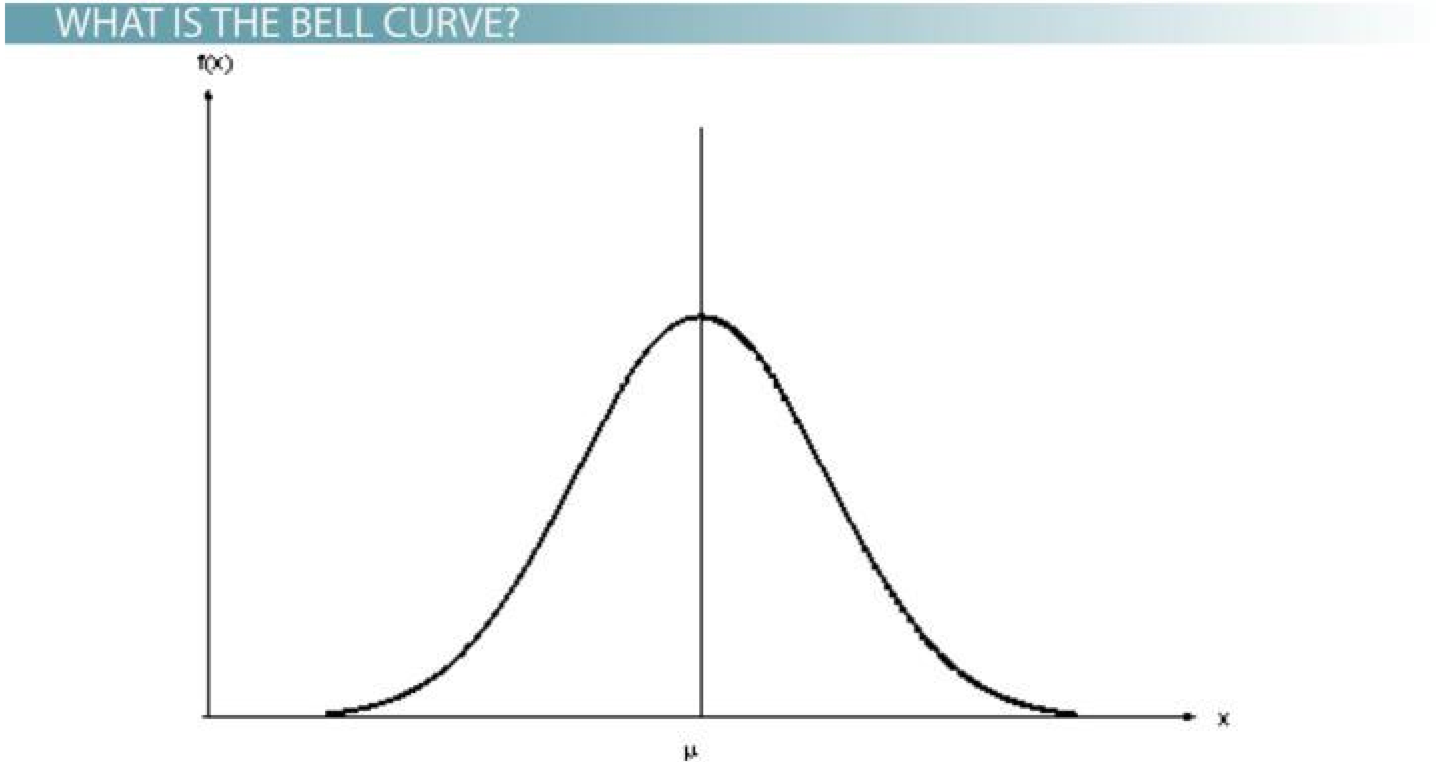
PDF examples (p370)

- `random.random()` produces a number between 0.0 and 1.0
  - The probability of any specific value is evenly distributed
  - The area under the curve is 1.0
  - The area under the curve for two points will be max – min
- `random.random() + random.random()` produces a number between 0.0 and 2.0
  - The probability of any specific value is not evenly distributed (like the dice)
  - The area under the curve is 1.0
  - The area under the curve for two points will be would have to be calculated

Normal (Gaussian) Distribution

- Peaks in the middle
- Decreases symmetrically on both sides
- Asymptotically approaches 0 (at both ends)
- Standard **bell curve**





Specifying a Normal Distribution

- **Mean** defines the peak
- **Standard deviation** defines the shape
  - Width and slope of the side

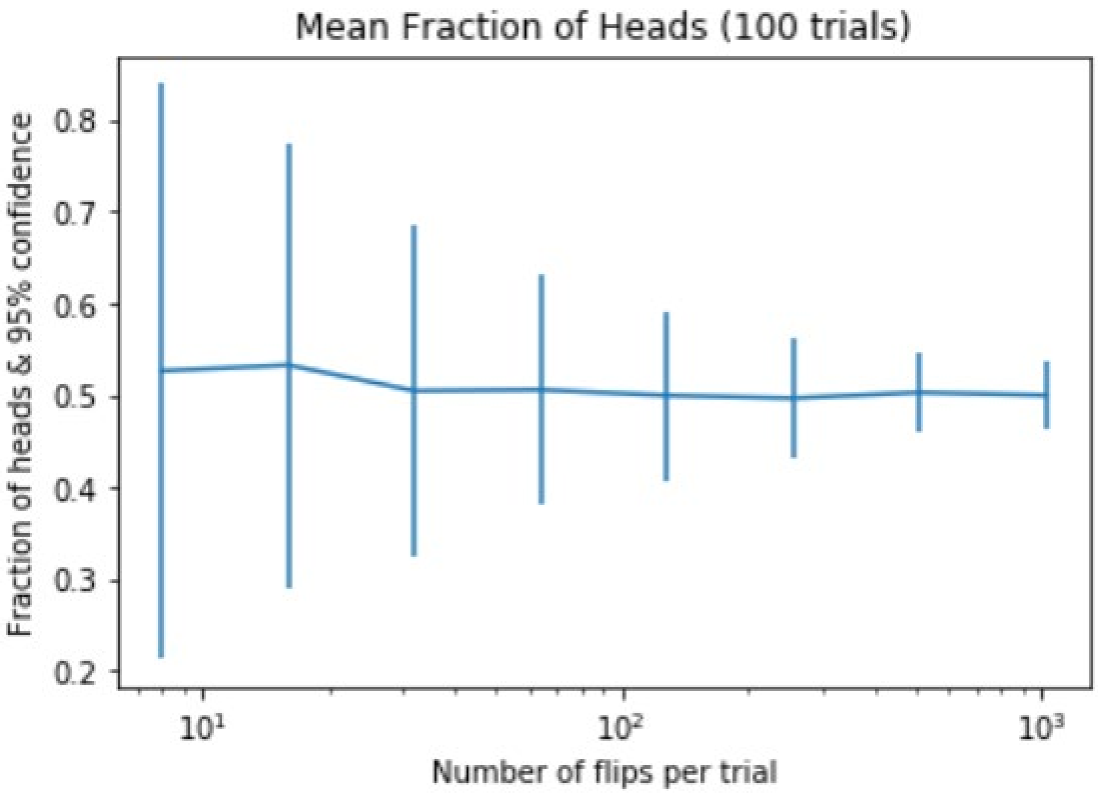
- Often the law of large numbers drives us to a Normal Distribution
- **Empirical rule**
  - Approximately 67.27% of all values are + or – 1σ from the mean
  - Approximately 95.45% are within 2σ
    - ~95% are within 1.96 σ
  - Approximately 99.7% are within 3σ

Normal Distribution and Confidence Intervals

- Provides a **range** that is likely to contain an unknown value
- And a **degree of confidence** that the value lies within that range

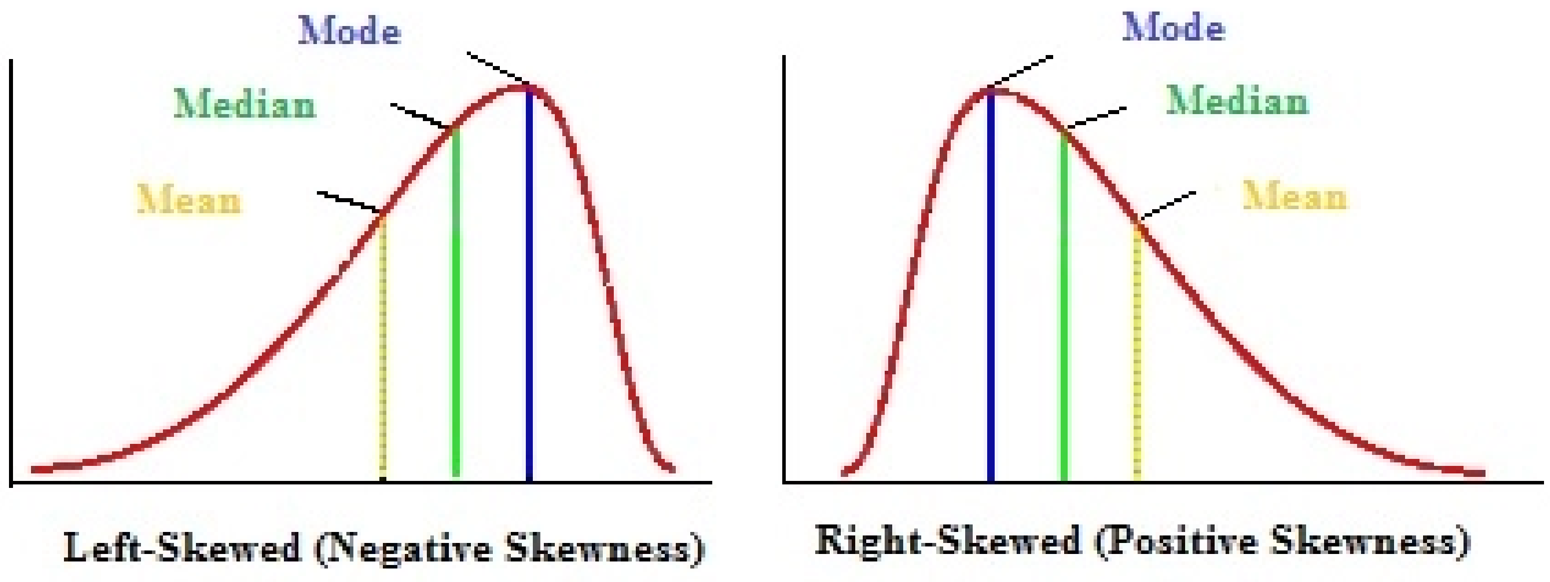
- Range is determined by mean +/- standard deviation
- Confidence level is given by the empirical rule

Introducing Error Bars



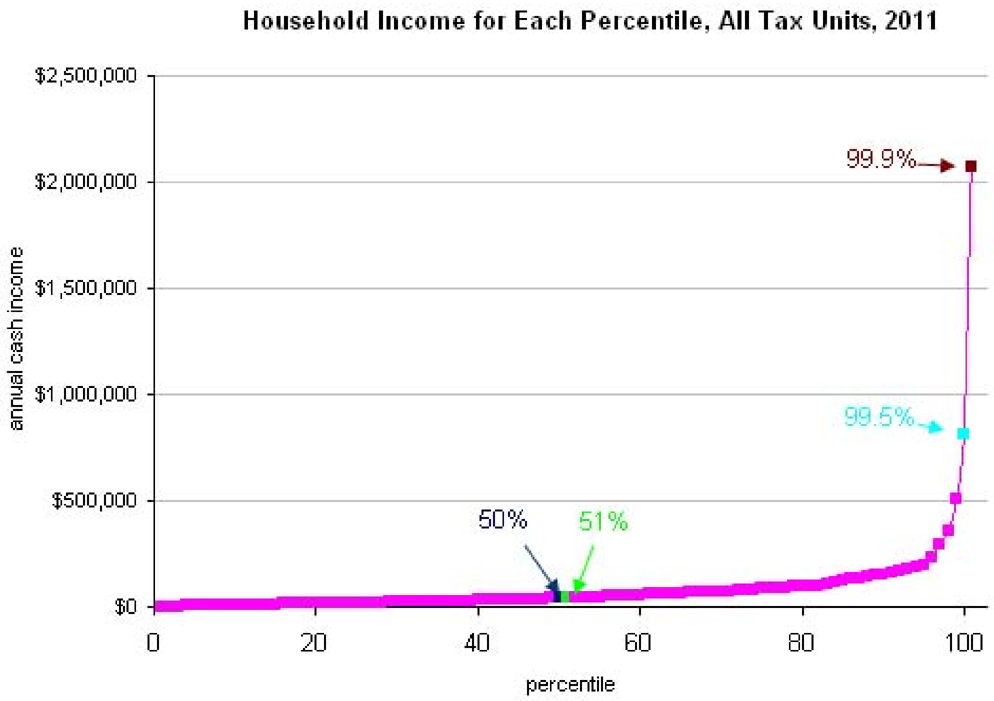
When Normal isn't

- Often data is "skewed" to the left or right
  - The direction of the skew corresponds with the side of the "long tail"
  - Is also referenced numerically

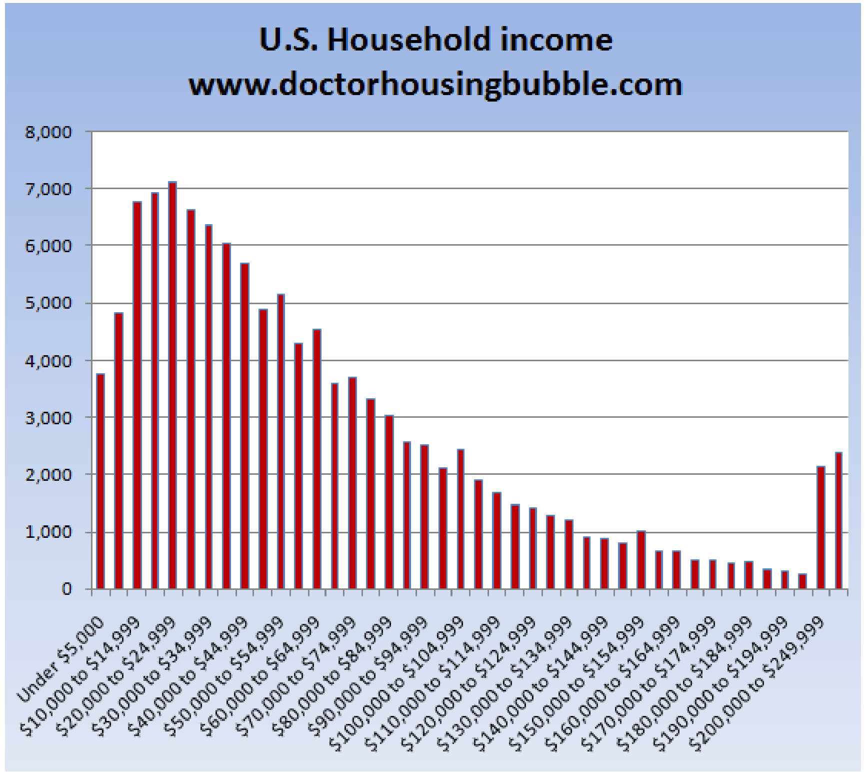


Income Example

Left (Negatively Skewed)



Right (Positive) Skewed



Uniform Distribution

- Continuous Uniform Distribution
  - Rectangular distribution
  - All intervals of the same length have the same probability
  - Probability for the interval (x,y)
    - $P(x,y) = \begin{Bmatrix} \frac{y-x}{max-min} \; if \; x \geq \min \quad and \; y \leq max \\ else \; 0 \end{Bmatrix}$
  - `random.uniform(min, max)`

- Continuous Discrete Distribution

Binomial and Multinomial Distribution

- Can only be discrete
  - **Nominal** or **Categorical**

- **Binomial** only two possible values (True/False, Pass/Fail)

- The probability of exactly *k* successes in *n* trials
- **Binomial coefficient** $\binom{n}{k}$ n choose k
  - Number of subsets (k) that can be constructed from a set of size n

Finger Exercise

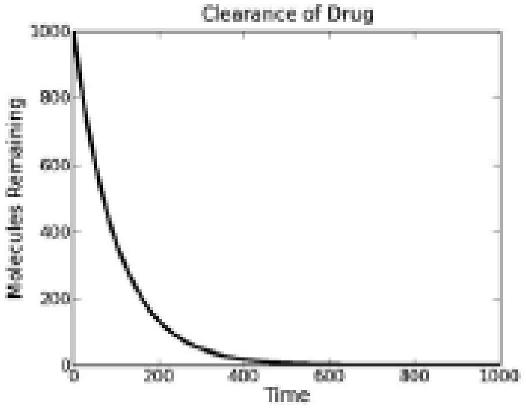
- Implement a function that calculates the probability of rolling exactly two 3’s in *k* rolls of a fair die. Use this function to plot the probability as k varies from 2 to 100

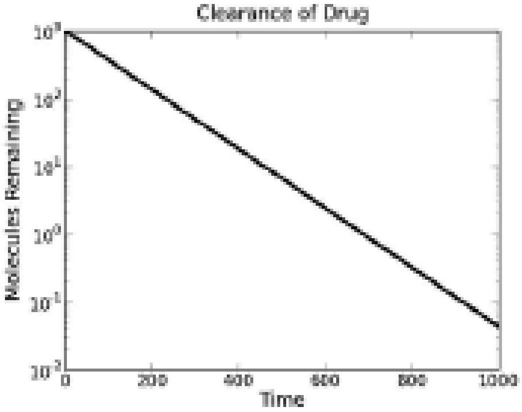
Multinomial Distribution

- A generalized case of binomial distribution
  - More than two possible outcomes
  - n trials each with m possible mutually exclusive outcomes
  - Multinomial distribution gives the probability of any combination of numbers of occurrences of outcomes

Exponential and Geometric Distributions

- Exponential **Growth** or **Decay**
  - Commonly found in nature
  - Half-life is the expected time for 50% decay
  - Use logarithmic plot for straight line

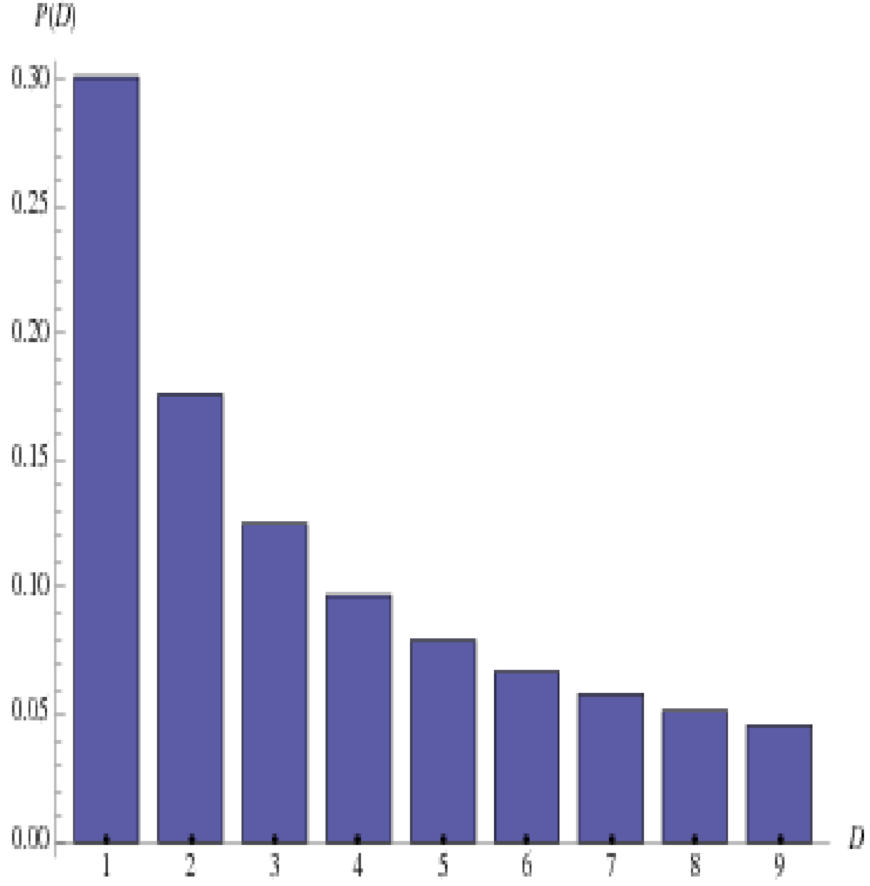




- Geometric – discrete analog to exponential

Benford's Distribution

- Given a large set of decimal integers, how often will any particular digit (1 through 9) be the first digit?
  - You would think this would be uniform, no?
  - **Benford's law** $P(d) = \ log_{10}\left(1 + \frac{1}{d}\right)$
- Occurs more often than you would think
- http://testingbenfordslaw.com/



Examples

- Determining hashing collisions
- World Series
  - That’s why we play the game
  - And that is why I leave it up to you to explore ;)