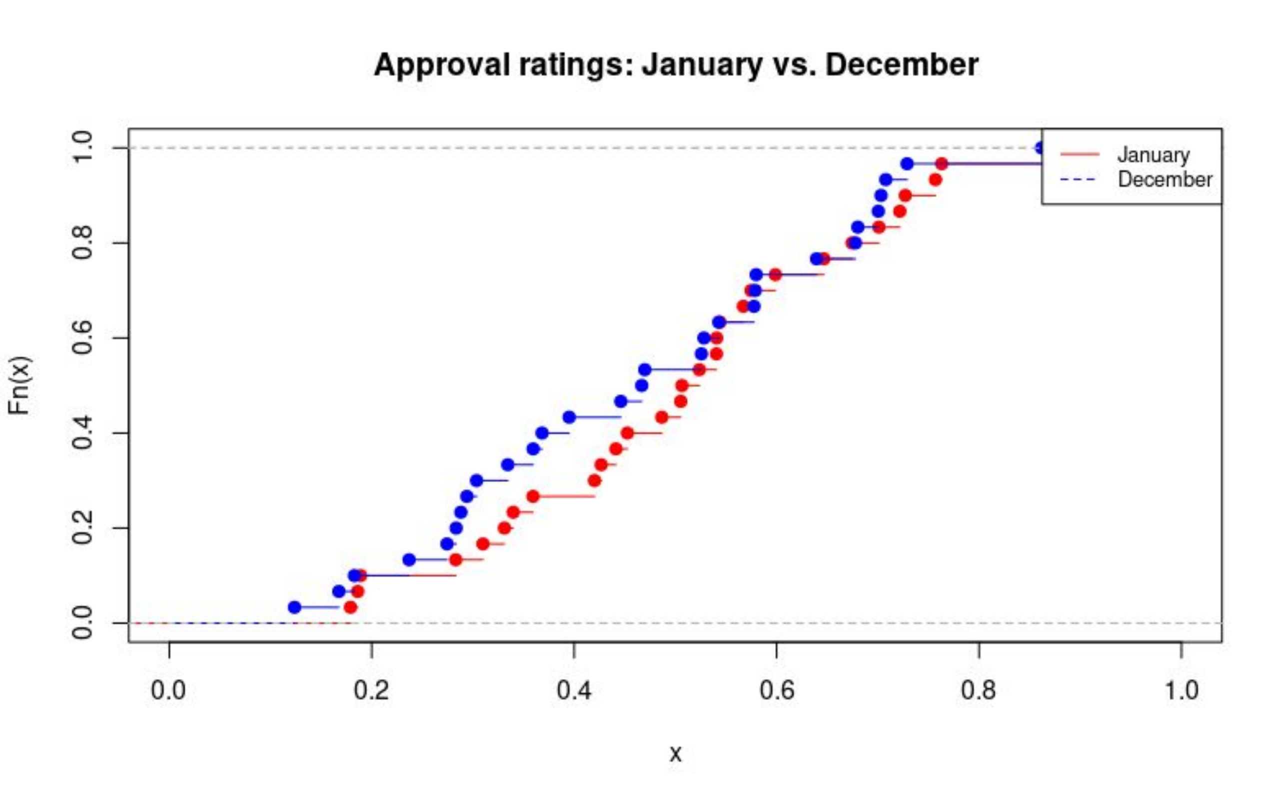
Useful Probability Distributions

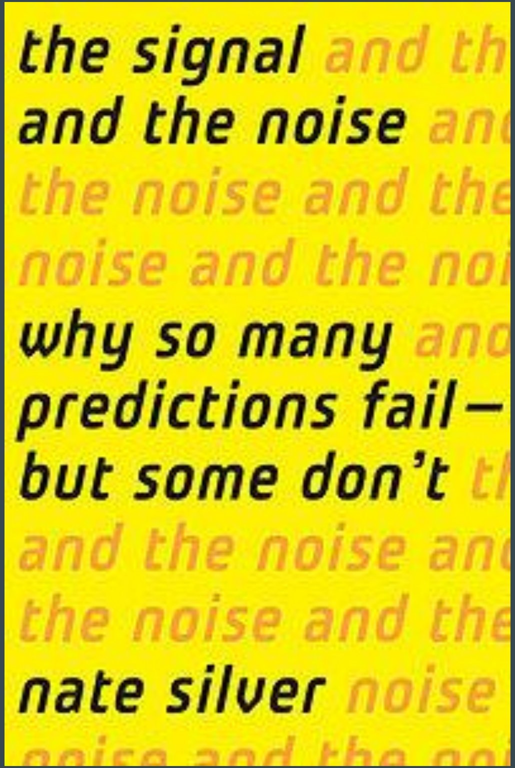
4 February 2019

PLSC 309

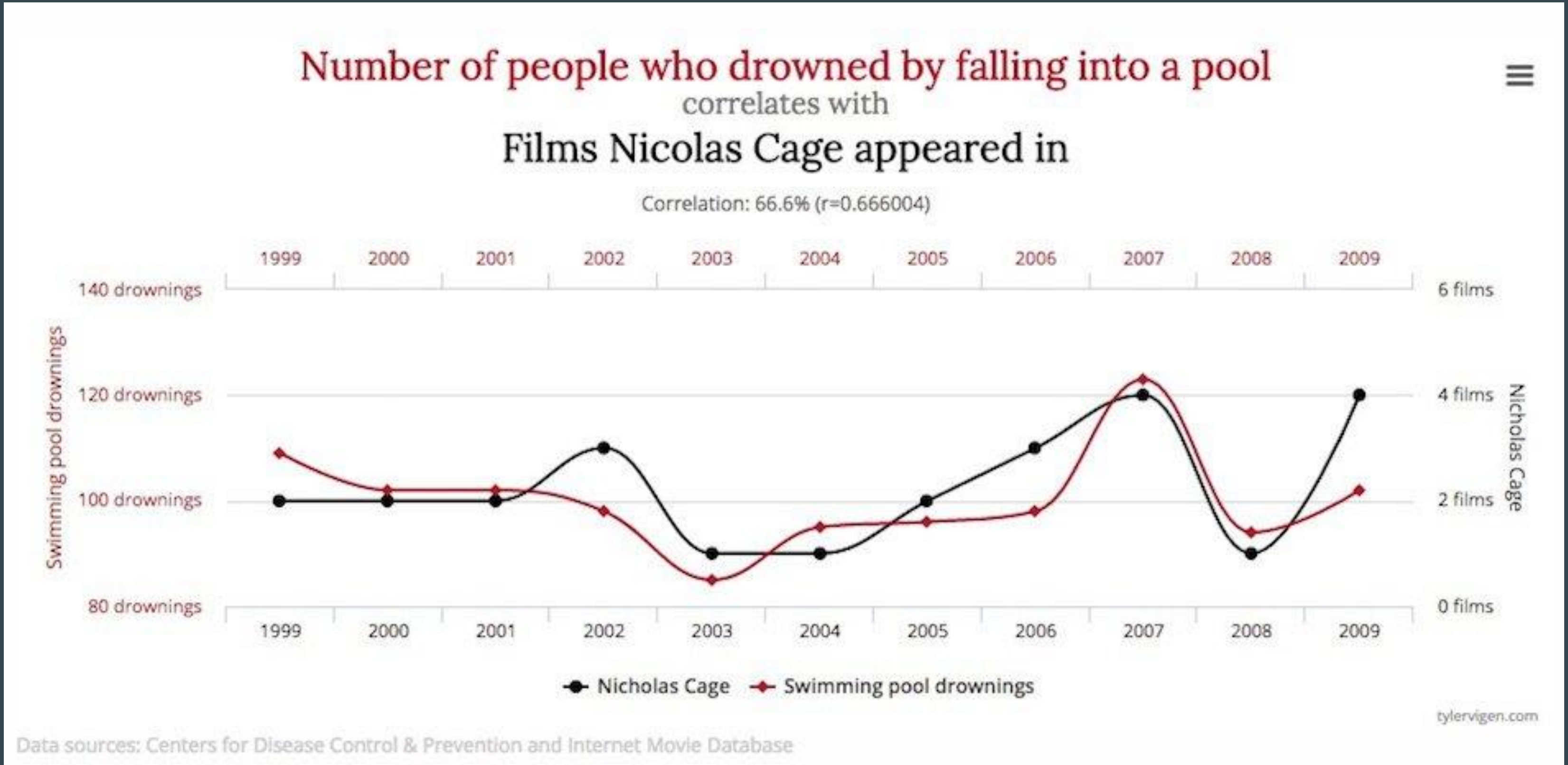
The fundamental problem of statistics



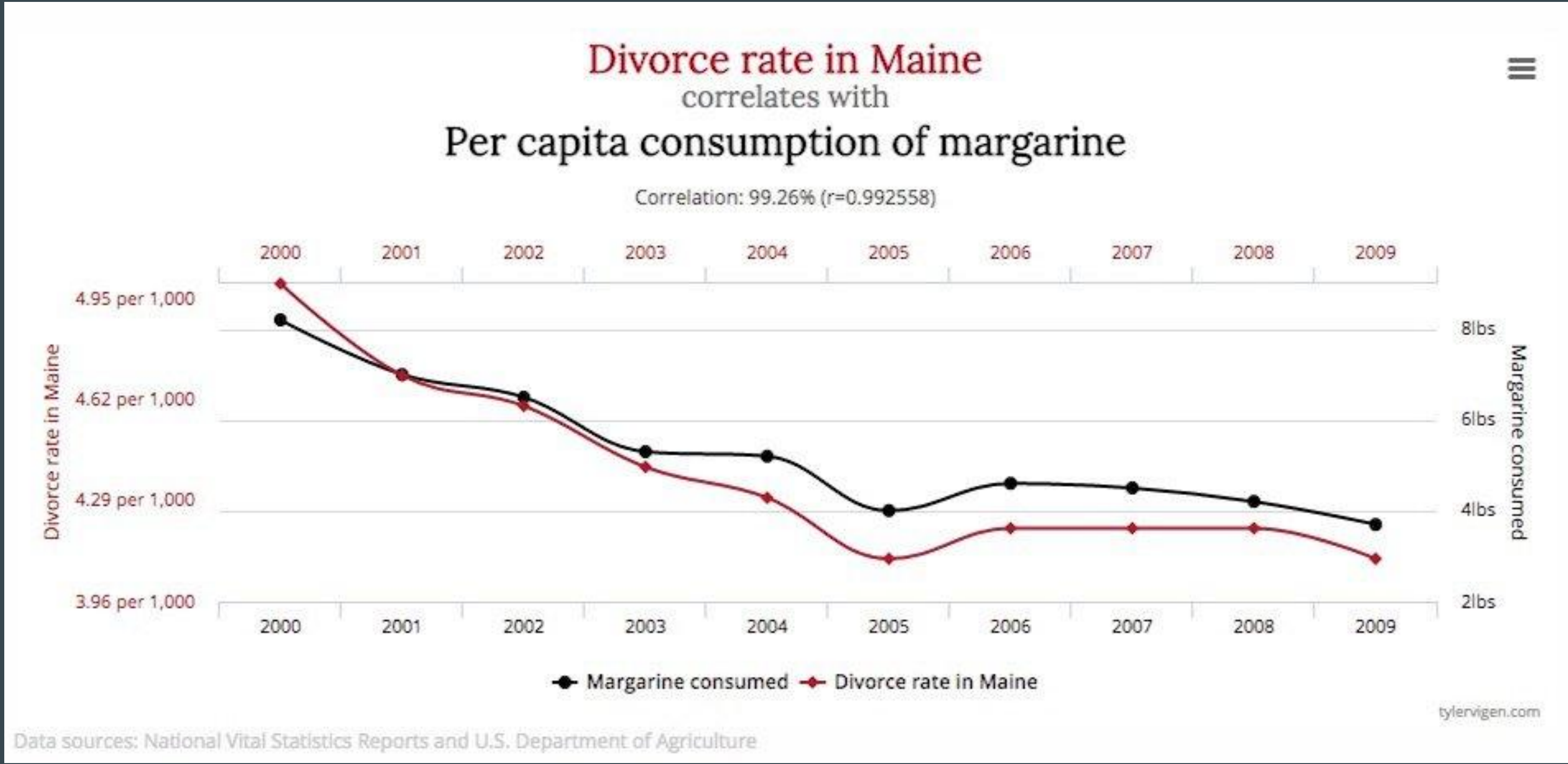
The fundamental problem of statistics



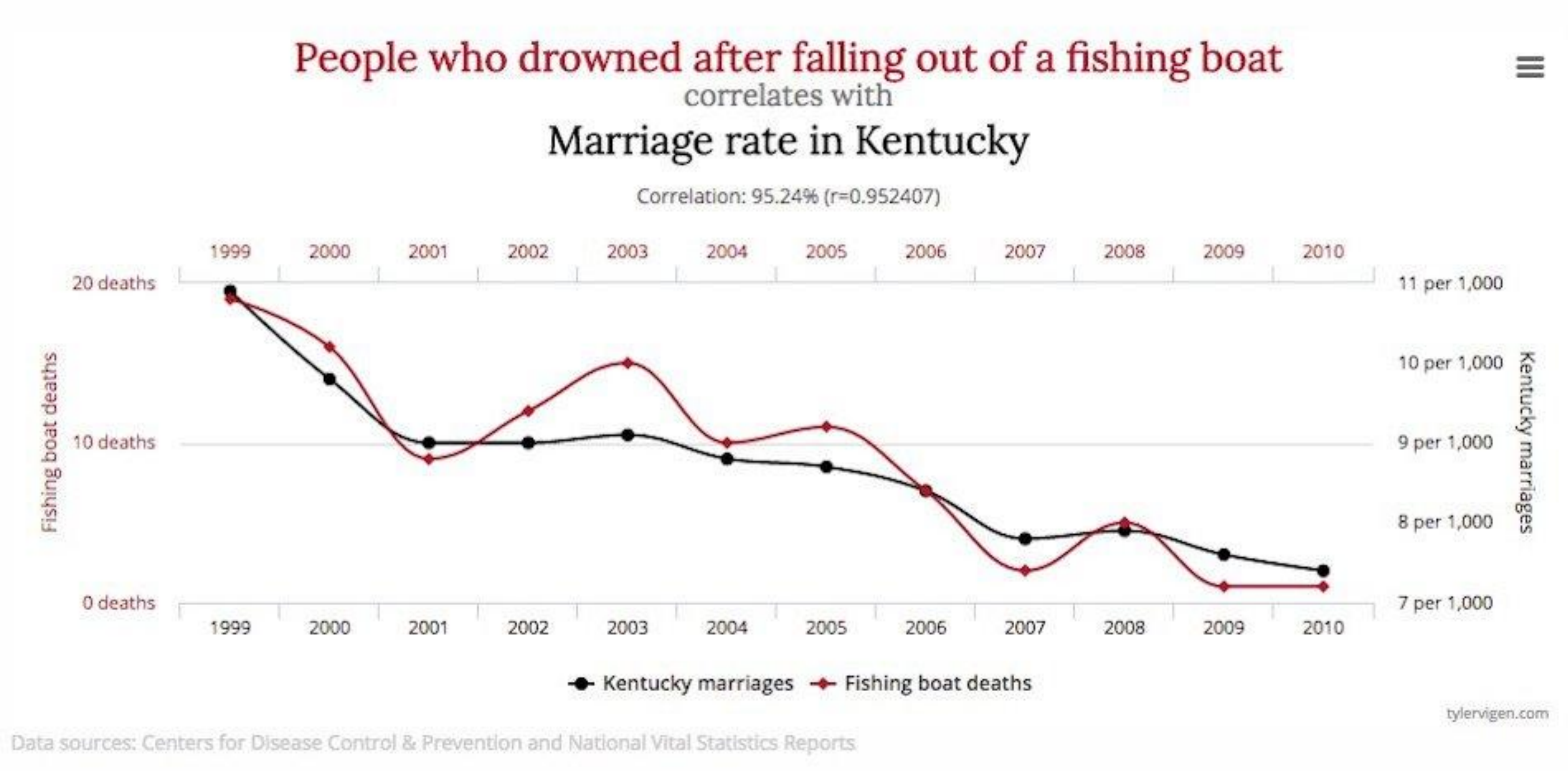
The fundamental problem of statistics



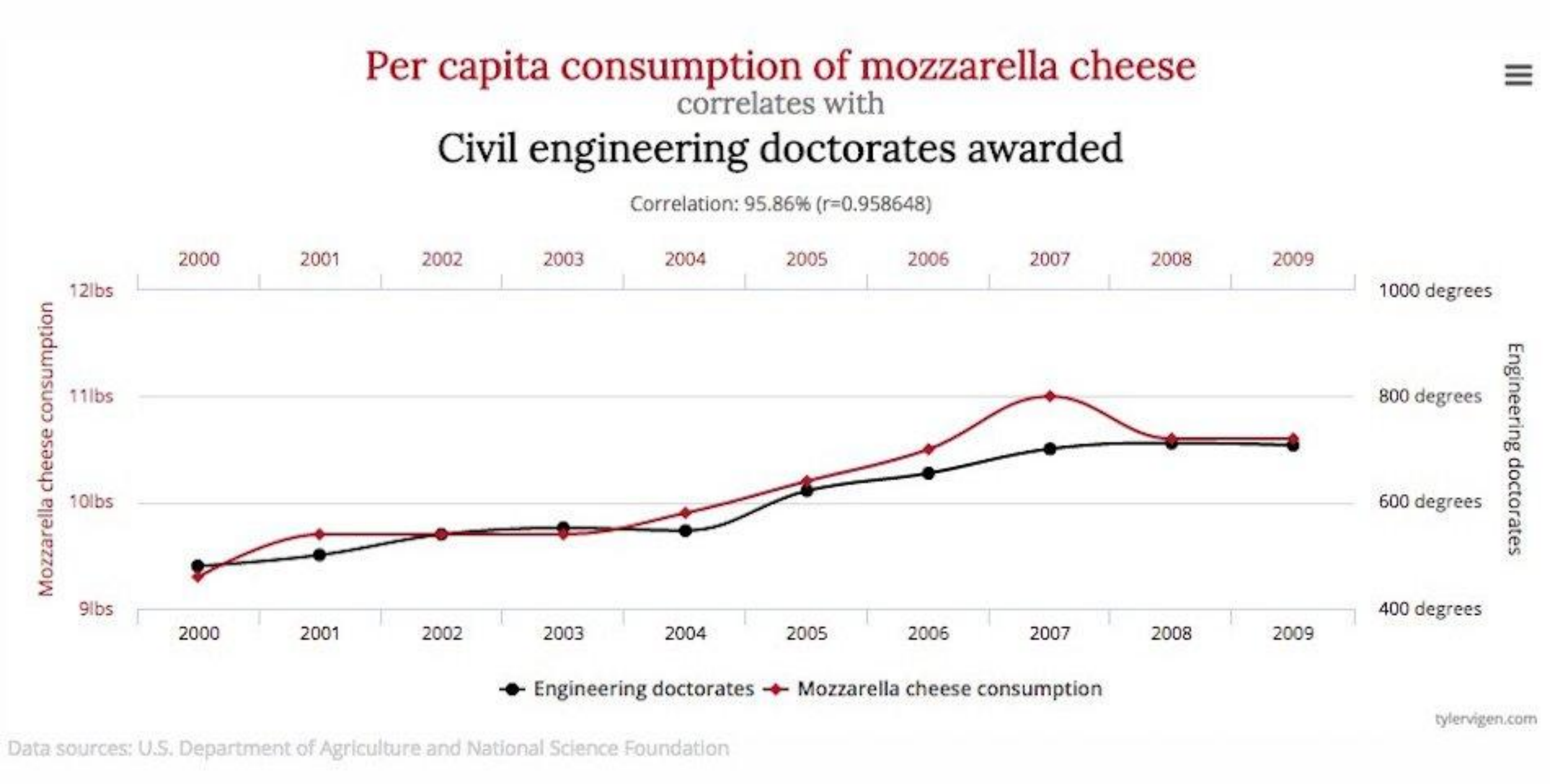
The fundamental problem of statistics



The fundamental problem of statistics



The fundamental problem of statistics



The fundamental problem of statistics

- We need to distinguish random noise from systematic changes
- If we assume a probability distribution for what we're trying to explain...
- ...we have a good guess for what the random noise will be!

Preview

- It is difficult to directly estimate a probability distribution from data
- Instead we look for *naturally occuring distributions*
- We then make an assumption (backed up by evidence) that our data is drawn from one of those distributions

But first, some vocabulary

- We have seen random variables, which we typically label X
- The distributions we see have other mathematical variables called *parameters*
    - These are usually constants
    - They determine the shape of the distribution
    - These can be calculated or directly estimated from data

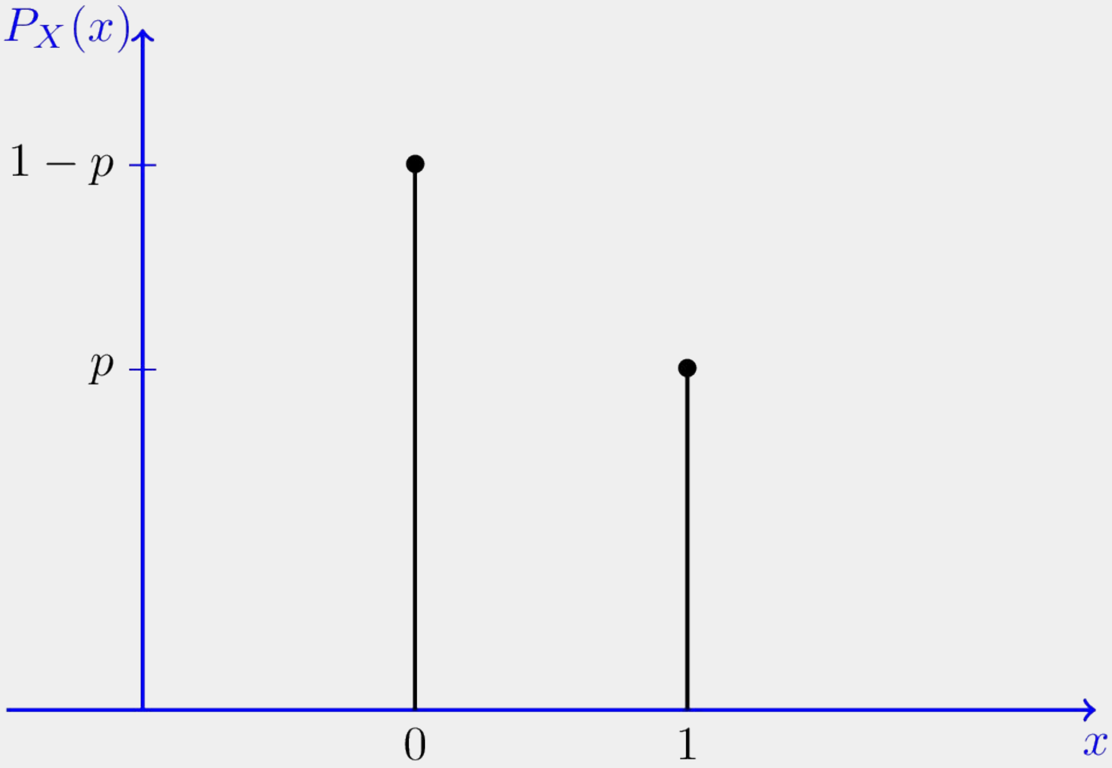
Bernoulli Distribution

Successes and failures

- A Bernoulli distribution describes any event with a binary outcome
  - Outcome = 1 (*success*)
  - Outcome = 0 (*failure*)
- Two parameters
  - *p:* probability of success
  - *q:* probability of failure (equivalent to 1-p)

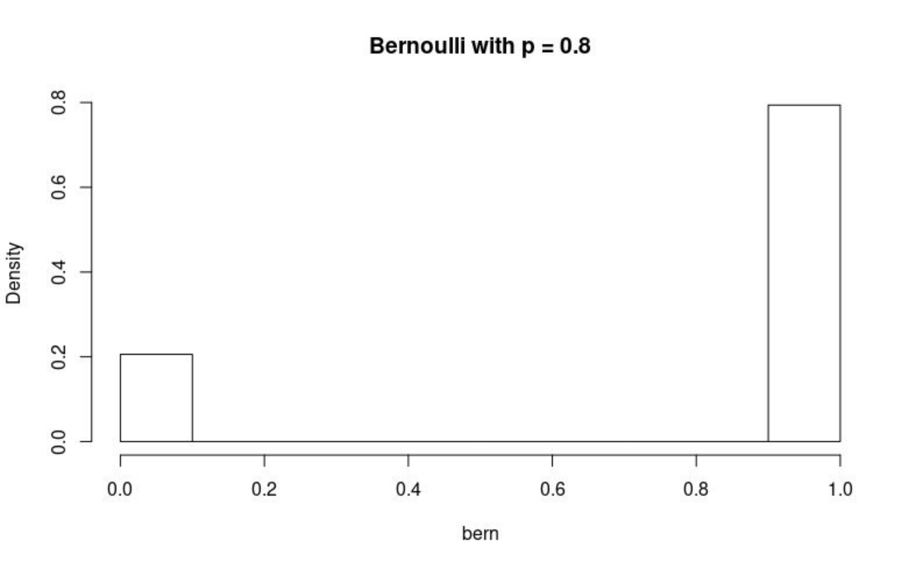
Bernoulli Distribution

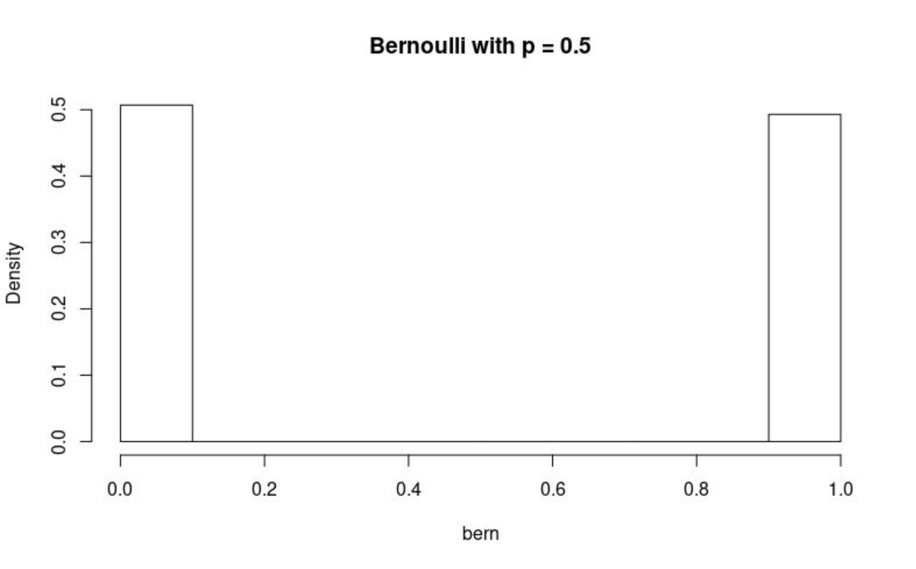
$$X \sim Bernoulli(p)$$



$$\begin{cases} q = 1-p & \text{if } k=0 \\ p & \text{if } k=1 \end{cases}$$

Changing parameter p





Mean and Variance for Bernoulli

- A bernoulli distribution is one trial that ends in either a success or a failure
- Each trial is independent from the other
- $E(X) = p$
- $Var(X) = p(1-p)$ or pq

Examples of Bernoulli processes

- A candidate is running for election. The chance they win is 70%
  - 0 = candidate loses
  - 1 = candidate wins
  - p = .7
  - E(X) = .7
  - Var(X) = .21
- There is a 5% chance of a coup in Thailand
  - 0 = no coup
  - 1 = coup
  - P = .05
  - E(X) = .05
  - Var(X) = .0475

Binomial Distribution

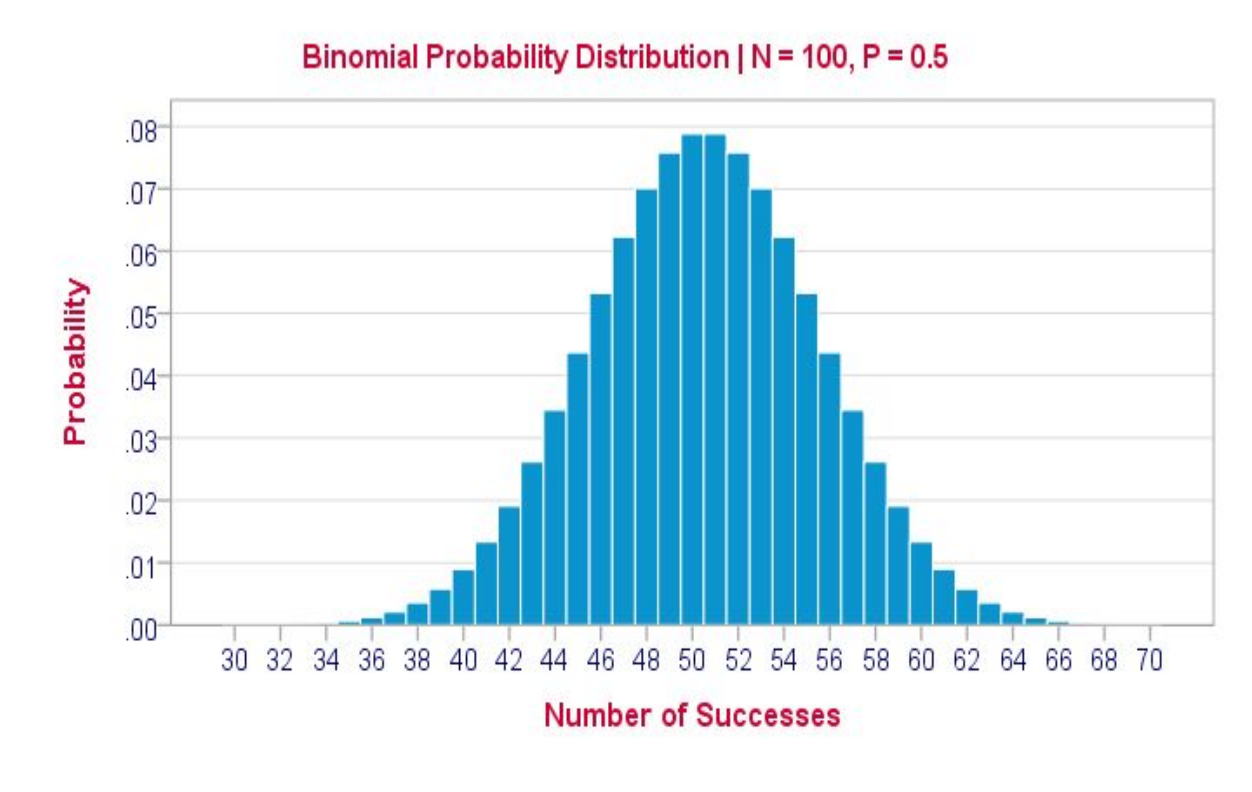
An extension of the Bernoulli

- In the Bernoulli, we had one event that yields an outcome of success or failure
- What if we had a bunch of events that yielded 0-1 outcomes?
- In other words, given a series of n trials, what is the probability that we have k successes?
- Three parameters
    - p: probability of success
    - n: number of trials
    - k: number of successes

Interlude: combinations

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Binomial Distribution



$$\binom{n}{k} p^k (1-p)^{n-k}$$

Binomial PMF

Let's say we're trying to flip a coin ten times. A success is a heads. We only get one head. Probability of success is *p*. Probability of failure is 1-*p*.

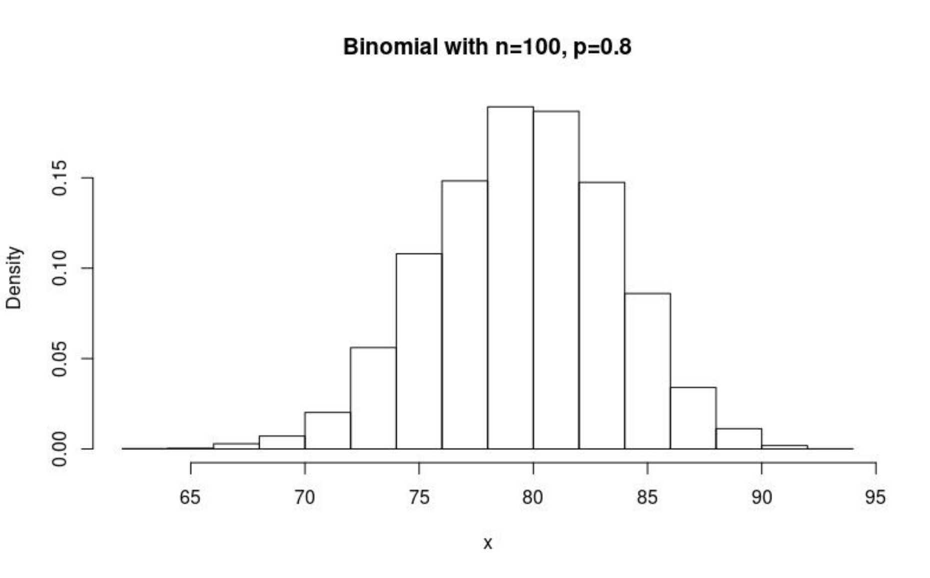
$\{S, F, F, F, F, F, F, F, F, F\} = p * (1\text{-}p)^9$

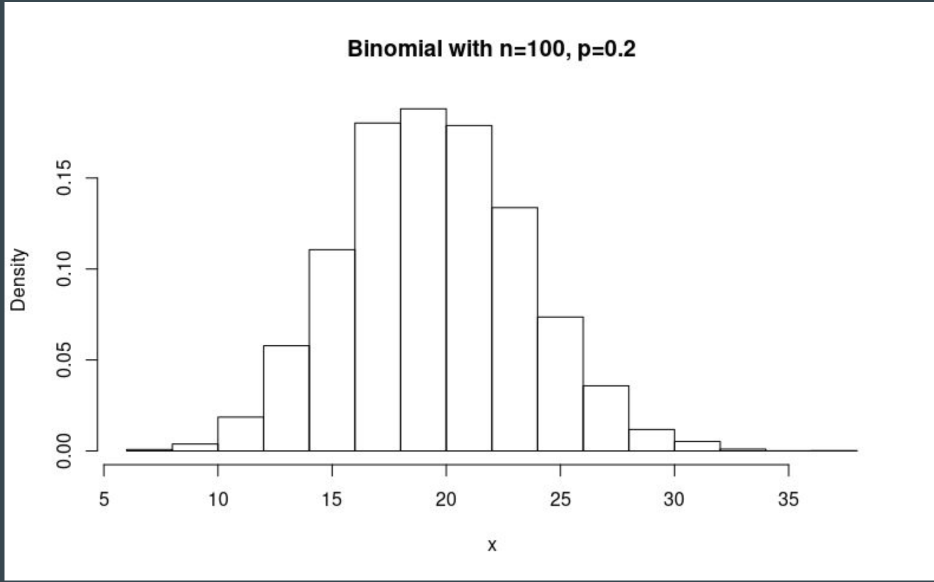
$\{F, S, F, F, F, F, F, F, F, F\} = (1\text{-}p) * p * (1\text{-}p)^8 = p * (1\text{-}p)^9$

$\{F, F, S, F, F, F, F, F, F, F\} = (1\text{-}p)^2 * p * (1\text{-}p)^7 = p * (1\text{-}p)^9$

In other words, each different way of getting one success has a probability of $p * (1\text{-}p)^9$. Summing all these up we get $10 * p * (1\text{-}p)^9$. 10 choose 1 = 10.

Effect of parameters on Binomial Distribution





Mean and Variance for Binomial distribution

- Binomial distribution extends the Bernoulli to n trials
  - E(X) for Bernoulli was p
  - …E(X) for Binomial is *np*
- Var(X) = np(1-p)

Examples of Binomial processes

- 18 candidates are running for office. Incumbents win 90% of the time. How many incumbents will win?
  - $E(X) = np = 16.2$
  - $Var(X) = np(1-p) = 1.62$
- A political party needs has 60 senators in Congress. Senators vote with their own party 92% of the time. What is the probability they get 60 votes?
  - $E(X) = np = 55.2$
  - $Var(X) = 4.416$

Multinomial Distribution

Categories as binary choices

- The binomial distribution dealt with k success from n trials
- A multinomial distribution is about choosing from k categories in n trials
- If there are k categories, each category has a probability, p_i

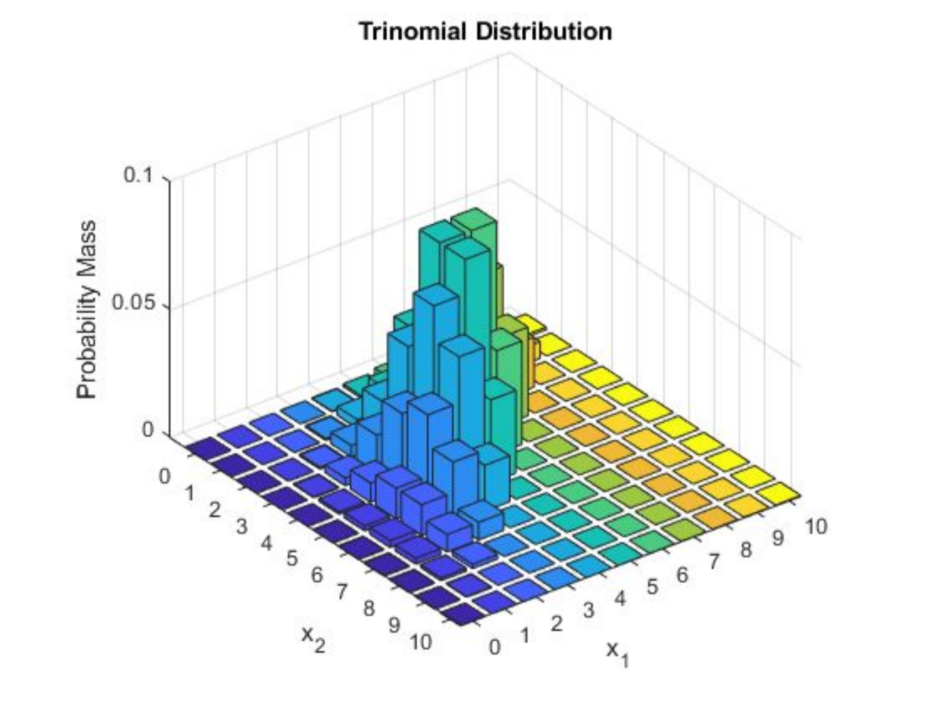
Categorical variables are condensed binary outcomes

Political party: 0 = Democrat; 1 = Republican; 2 = Green

Political Party
0
2
1

Democrat	Republican	Green
1	0	0
0	0	1
0	1	0

Multinomial Distribution



$$\frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

Mean and Variance for Multinomial distribution

- There is no global expected value or variance, just one for each category
- Calculations are exactly the same as the binomial distribution
  - $E(X_i) = np_i$
  - $Var(X_i) = np_i(1-p_i)$

Examples of Multinomial Processes

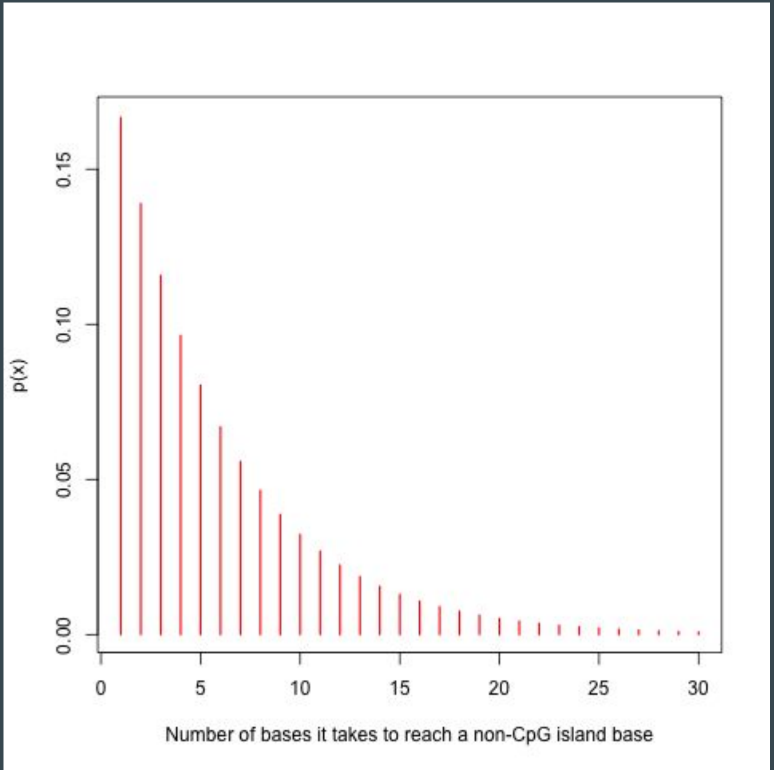
- Any consumer choice
  - Type of car to buy
  - Type of phone to buy
  - Etc.
- The Green, Democrat, and Republican party candidates are running against one another in a district with 1000 voters. The probability that someone votes Green is 2%, Democrat is 49%, and Republican is 49%.
  - $E(\text{Green}) = np_i = 20$
  - $E(\text{Dem}) = np_i = 490$

Geometric Distribution

How long does it take to get a success?

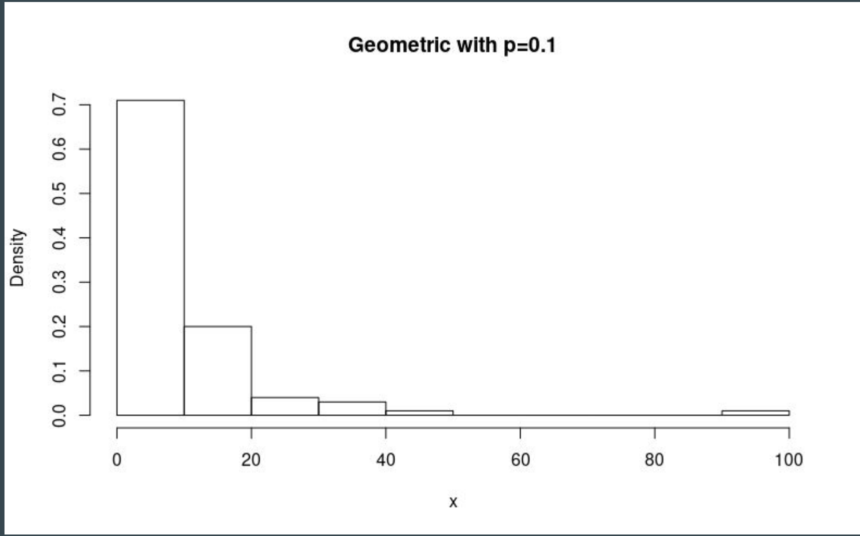
- For *k* trials, how long does it take to get to one success?
- Bernoulli: one trial, one success
- Binomial: *n* trials, *k* successes
- Geometric: *k* trials, one success
  - *p*: probability of success
  - *k:* how many trials it takes to get to one success

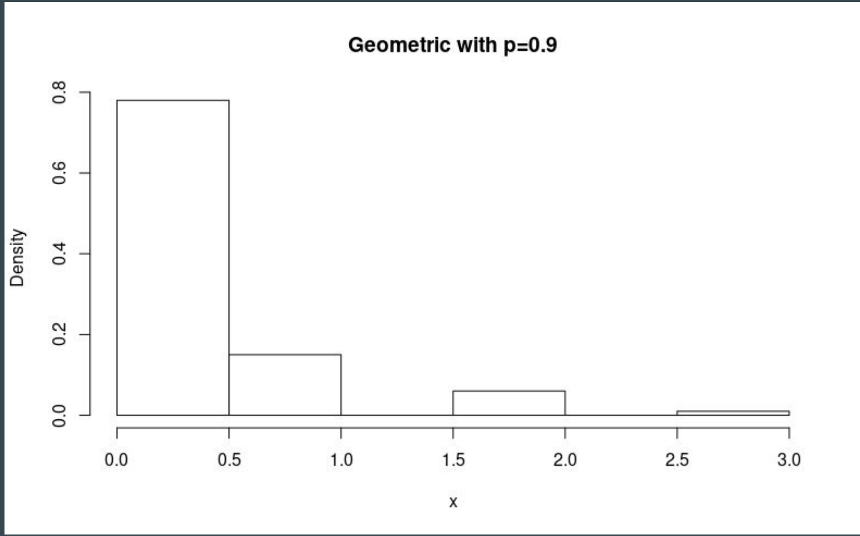
Geometric distribution



$$(1-p)^{k-1}p$$

Changing parameters for geometric





Mean and Variance for the geometric distribution

- Expected value for the geometric is telling us, on average, how many trials we should expect to do before we come to a single success
  - $E(X) = 1/p$
- $Var(X) = (1-p) / p^2$

Examples of geometric processes

- A factory is making widgets. Their failure rate is 0.01. How many trials will it take until there is a failure?
  - $E(X) = 1/p = 100$
  - $Var(X) = (1-p) / p^2 = 9900$
- Somebody won't stop gambling until they win at blackjack. The chance you win at blackjack is 49%. How many hands will they sit through until they win?
  - $E(X) = 1/p = 49$
  - $Var(X) = (1-p) / p^2 = 2.12$

Review

Today we learned about commonly occuring discrete distributions:

- Bernoulli - a single trial that can end in either success or failure
- Binomial - a collection of Bernoulli trials
- Multinomial - a categorical extension of the binomial distribution
- Geometric - number of times it takes for a series of Bernoulli trials to produce a success