41000: Business Statistics

Vadim Sokolov

Chicago Booth

Spring 2021

Getting Started

HW, Slides
Midterms

https://vsokolov.org/courses/41000/

- General Expectations
  1. Read the notes/Practice
  2. Be on schedule
  3. Add R to your friend list!

Course Expectations

Midterm: 35% T/F and three long questions. Cheat Sheet allowed.

Final Project: 35% 50% on Writing/Presentation skills 50% on Modeling.

Homework: 30% Bi-weekly Assignments. Groups of 3-4. Otherwise it's no fun!

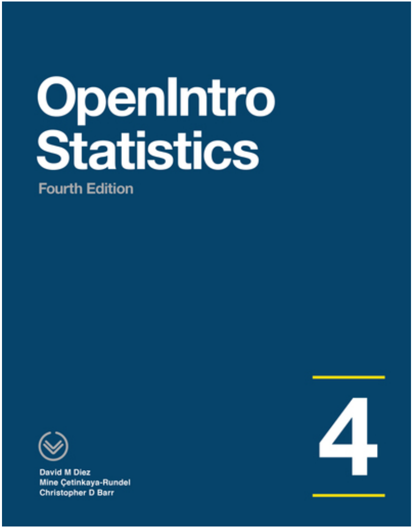
Grading is $\checkmark+, \checkmark, \checkmark-$.

~~Fixed Grading Curve: 3.33.~~

Course Overview

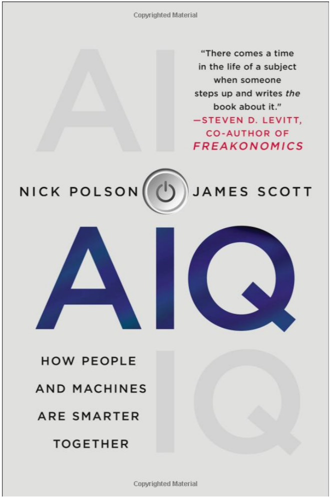
Weeks 1&2	Probability & Bayes OpenIntro Statistics, Chapters 2&3
Weeks 3&4	Data Analytics Chapters 4,5&6
Week 5	Modeling and Linear Regression Chapters 6&7
Week 6	Midterm
Week 7	Logistic Regression
Week 8	Predictive Analytics
Week 9	Artificial Intelligence (AI)
Week 10	Deep Learning (DL)

Course Book: OpenIntro Statistics



AIQ: People & Robots Smarter Together

Additional Reading!!



Seven AI-IQ Stories

- Abraham Wald (October 31, 1902 – December 13, 1950)
- Henrietta Leavitt (July 4, 1868 – December 12, 1921)
- John Craven (August 16, 1940 –)
- Grace Hopper (December 9, 1906 – January 1, 1992)
- Isaac Newton (January 4,1643 – March 31, 1727)
- Florence Nightingale (May 12, 1820 – August 13, 1910)
- Joe DiMaggio (November 25, 1914 – March 8, 1999)

Business Statistics: 41000

Section 1: Introduction

Probability and Bayes

Vadim Sokolov

Suggested Reading

OpenIntro Statistics, Chapters 2&3

This Section

How to deal with uncertainty?

- Random Variables and Probability Distributions
- Joint and Conditional (Happy/Rich), Independence (Sally Clark),
- Expectation and Variance (Bookies vs Betters, Tortoise and Hare)
- Binomial Distribution (Patriot Coin Toss), Normal distribution (Crash of 1987)
- Decision Making under uncertainty (Marriage Problem and Probability and Decision Trees)
- Bayes Rule (Practice Hard $\neq$ Play in NBA)

Review of Basic Probability Concepts

Probability lets us talk efficiently about things that we are uncertain about.

- What will Amazon's sales be next quarter?
- What will the return be on my stocks next year?
- How often will users click on a particular Google ad?

All these involve estimating or predicting unknowns!!

Random Variables

Numeric

Event → Outcome (Random Variable)

Random Variables are numbers that we are not sure about. There's a list of potential outcomes. We assign probabilities to each outcome.

Example: Suppose that we are about to toss two coins. Let X denote the number of heads. We call X the random variable that stands for the potential outcome.

Probability

Probability is a language designed to help us communicate about uncertainty. We assign a number between 0 and 1 measuring how likely that event is to occur It's immensely useful, and there's only a few basic rules.

1. If an event A is certain to occur, it has probability 1, denoted $P(A) = 1$
2. Either an event A occurs or it does not.

   $P(A) + P(\text{not } A) = 1$

   $$P(A) = 1 - P(\text{not } A)$$

3. If two events are mutually exclusive (both cannot occur simultaneously) then

   $$P(A \text{ or } B) = P(A) + P(B)$$

   OR => +

4. Joint probability, when events are independent

   $$P(A \text{ and } B) = P(A)P(B)$$

   AND => *

Probability Distribution

Map: Outcome → Probab.
Value of R.V

We describe the behavior of random variables with a Probability Distribution

Example: Suppose we are about to toss two coins. Let X denote the number of heads.

$$X = \begin{cases} 0 \textit{ with prob. } 1/4 \\ 1 \textit{ with prob. } 1/2 \\ 2 \textit{ with prob. } 1/4 \end{cases}$$

HH
HT
TT
TH

X is called a Discrete Random Variable

Question: What is $P(X = 0)$? How about $P(X \geq 1)$?

$P(X \geq 1) = P(X = 1 \text{ or } X = 2) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}$

Pete Rose Hitting Streak

Proob of hitting safely
is 0.3

Pete Rose of the Cincinnati Reds set a National League record of hitting safely in 44 consecutive games ...

- Rose was a 300 hitter. Assume he comes to bat 4 times each game.
- Each at bat is assumed to be independent, i.e., the current at bat doesn't affect the outcome of the next.

What probability might reasonably be associated with that hitting streak?

Joe DiMaggio's record is 56! His batting average was .325

Pete Rose Hitting Streak

Let A_i denote the event that "Rose hits safely in the ith game"

$P(\text{Rose Hits Safely in 44 consecutive games}) =$

$P(A_1 \text{ and } A_2 \ldots \text{and } A_{44}) = P(A_1)P(A_2)\ldots P(A_{44})$

We now need to find $P(A_i) \ldots$ where $P(A_i) = 1 - P(\text{not } A_i)$

$$
\begin{aligned}
P(A_1) &= 1 - P(\text{not } A_1) \\
&= 1 - P(\text{Rose makes 4 outs}) \\
&= 1 - (0.7)^4 = 0.76
\end{aligned}
$$

So for the winning streak we have $(0.76)^{44} = 0.0000057$!!!

10^{-6}

Pete Rose Hitting Streak

$p \approx 10^{-6}$

$\frac{p}{1-p}$

There are three basic inferences

- This means that the odds for a particular player as good as Pete Rose starting a hitting streak today are $175,470$ to 1
- Doesn't mean that the run of 44 won't be beaten by some player at some time: the Law of Very Large Numbers
- Joe DiMaggio's record is 56!!!! It's going to be hard to beat. We have $(0.792)^{56} = 2.13 \times 10^{-6}$ or 455,962 to 1

New England Patriots and Coin Tossing

Patriots won 19 out of 25 coin tosses in 2014-15 season! What is the probability of that happening?

- Let X be a random variable equal to 1 if the Patriots win and 0 otherwise. It's reasonable to assume $P(X = 1) = \frac{1}{2}$
- There are 25 choose 19 or $177,100$ different sequences of 25 games where the Patriots win 19. Each potential sequence has probability 0.5^{25} why?

$$Pr\left(\textit{Patriots win } 19 \textit{ out } 25 \textit{ tosses}\right) = 177,100 \times 0.5^{25} = 0.005$$

0.5%

$P(X=1) = \frac{1}{2}$

Prob of 19 wins out 25

$$\underbrace{1\ 1\ \dots\ 1}_{19}\ \underbrace{0\ 0\ \dots\ 0}_{6}$$

$$\boxed{(0.5)^{25} \cdot \frac{25!}{19!\,6!}}$$

$$\underset{0.5}{P(X=1)}\,P(X=1)\,\dots\,P(X=1)\,\underset{0.5}{P(X=0)}\,\dots\,P(X=0)$$

$$= (0.5)^{25}$$

$$\frac{25 \cdot 24 \cdot 23 \cdots 1}{19!\quad 6!} \quad 25! \quad = \frac{25!}{19!\,6!}$$

Conditional, Joint and Marginal Distributions

Use probability to describe outcomes involving more than one variable at a time.

Need to be able to measure what we think will happen to one variable relative to another

In general the notation is ...

$,$ = AND

- $P(X = x, Y = y)$ is the joint probability that $X = x$ and $Y = y$
- $P(X = x \mid Y = y)$ is the conditional probability that X equals x given $Y = y$
- $P(X = x)$ is the marginal probability of $X = x$

known fact

calculated from joint.

Conditional, Joint and Marginal Distributions

Relationship between the joint and conditional ...

$$
\begin{aligned}
P(x, y) &= P(x)P(y \mid x) \\
&= P(y)P(x \mid y)
\end{aligned}
$$

Relationship between the joint and marginal ...

$$
P(x) = \sum_{y} P(x, y)
$$

$$
P(y) = \sum_{x} P(x, y)
$$

Example: $P(y=2, X=1) = 0.11$

"happiness index" as a function of salary.

Joint pobability table

Salary (X)	Happiness (Y) 0 (low)	1 (medium)	2 (high)
low 0	0.03	0.12	0.07
medium 1	0.02	0.13	$\boxed{0.11}$
high 2	0.01	0.13	0.14
very high 3	0.01	0.09	0.14

$0.24 = P(X=3)$

0.58   0.46 $0.18 + 0.28 = 0.46$

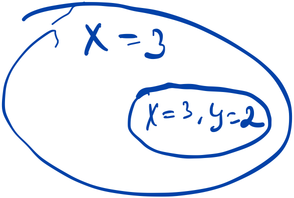
Is $P(Y = 2 \mid X = 3) > P(Y = 2)$?

proportion of happy people among those who have very high salary.

avg. person

$$P(y=2) = P(y=2, X=0) + P(y=2, X=1) + P(y=2, X=2) + P(y=2, X=3)$$

$$P(y=2 \mid X=3) = \frac{P(y=2, X=3)}{P(X=3)} =$$



$$= \frac{0.14}{0.24} = 0.58$$

Independence

$P(2H \mid 1H) = \frac{1}{2} = P(2H)$

Two random variable X and Y are independent if

$$P(Y = y \mid X = x) = P(Y = y)$$

for all possible x and y values. Knowing $X = x$ tells you nothing about Y!

Example: Tossing a coin twice. What's the probability of getting H in the second toss given we saw a T in the first one?

Bayes Rule

$$\boxed{P(x|y) = \frac{P(y,x)}{P(y)}}$$

Bayes Rule

The computation of $P(x \mid y)$ from $P(x)$ and $P(y \mid x)$ is called Bayes theorem ...

$$P(x \mid y) = \frac{P(y,x)}{P(y)} = \frac{P(y,x)}{\sum_x P(y,x)} = \frac{P(y \mid x)P(x)}{\sum_x P(y \mid x)P(x)}$$

This shows now the conditional distribution is related to the joint and marginal distributions.

You'll be given all the quantities on the r.h.s.

$$P(y,x) = P(y)P(x|y)$$

$$P(x,y) = P(x)P(y|x)$$

Bayes Rule

$$P(NBA \mid PH) = \frac{\overset{1}{P(PH \mid NBA)}\;\overset{\approx 0}{P(NBA)}}{P(PH)}$$

Key fact: $P(x \mid y)$ is generally different from $P(y \mid x)$!

Example: Most people would agree

$$Pr\left(Practice\ hard \mid Play\ in\ NBA\right) \approx 1$$

$$Pr\left(Play\ in\ NBA \mid Practice\ hard\right) \approx 0$$

The main reason for the difference is that $P(Play\ in\ NBA) \approx 0$.

Sally Clark Case: Independence Plays a Huge Role

Famous London crime case: Both babies died of SIDS

$$p(S_1, S_2) = p(S_1)p(S_2) = (1/8500)(1/8500) = (1/73,000,000)$$

The odds ratio for double SIDS to double homicide at between 4.5:1 and 9:1

Under Bayes

$\times\, p(S_2)$

$$p(S_1, S_2) = p(S_1)p(S_2 \mid S_1) = (1/8500)(1/100) = (1/850,000)$$

The $1/100$ comes from taking into account genetics.

That's a big difference! Under dependence assumption she'd be acquitted.

Travel Time

from Gleacher to ORD

Prob	t F
0.25	35
0.3	40
.45	30

Expected Travel Time (Avg):

$0.25 \cdot 35 + 0.3 \cdot 40 + 0.45 \cdot 30$

Expected Value of a Random Variable

$$\boxed{E(X) = \sum_x x\, P(x)}$$

Random Variables: Expectation $E(X)$

Example

- The expected value of a random variable is simply a weighted average of the possible values X can assume.
- The weights are the probabilities of occurrence of those values.

$$E(X) = \sum_{x} xP(X = x)$$

- With n equally likely outcomes with values $x_1, \ldots, x_n$, $P(X = x_i) = 1/n$

$$E(X) = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Roulette Expectation

- European Odds: 36 numbers (red/black) + zero
- You bet $\$1$ on 11 Black (pays 35 to 1)
- X is the return on this bet

$$E(X) = \frac{1}{37} \times 36 + \frac{36}{37} \times 0 = 0.97$$

- If you bet $\$1$ on Black (pays 1 to 1)

$$E(X) = \frac{18}{37} \times 2 + \frac{19}{37} \times 0 = 0.97$$

Casino is guaranteed to make money in the long run!

Standard Deviation $sd(X)$ and Variance $Var(X)$

Units $\quad X \quad Var(X)$

$\qquad\quad \$ \quad \$^2$

The variance is calculated as

$$Var(X) = E\left((X - E(X))^2\right) = \sum_x (x - E(x))^2 P(x)$$

A simpler calculation is $Var(X) = E(X^2) - E(X)^2$.

The standard deviation is the square-root of variance.

$X \quad Var(x) \quad sd(x)$

$\$ \quad \$^2 \quad \$$

$$sd(X) = \sqrt{Var(X)}$$

Roulette Variance

- European Odds: 36 numbers (red/black) + zero
- You bet $1 on 11 Black (pays 35 to 1)
- X is the return on this bet

$$Var(X) = \frac{1}{37} \times (36 - 0.97)^2 + \frac{36}{37} \times (0 - 0.97)^2 = 34$$

- If you bet $1 on Black (pays 1 to 1)

$$Var(X) = \frac{18}{37} \times (2 - 0.97)^2 + \frac{19}{37} \times (0 - 0.97)^2 = 1$$

If your goal is to spend as much time as possible in the casino (free drinks): place small bets on black/red

Example: $E(X)$ and $Var(X)$

Var:

T: $(0-1.5)^2 \cdot 0 + (1-1.5)^2 \cdot 0.5 + (2-1.5)^2 \cdot 0.5 = 0.25$

Tortoise and Hare are selling cars.

Probability distributions, means and variances for X, the number of cars sold

	X				Mean	Variance	sd
cars sold	0	1	2	3	$E(X)$	$Var(X)$	$\sqrt{Var(X)}$
100% Tortoise	0	0.5	0.5	0	1.5	0.25	0.5
Hare	0.5	0	0	0.5	1.5	2.25	1.5

T $E(X)$: $1.5 = 0.5 \cdot 1 + 0.5 \cdot 2$

H $1.5 = 0.5 \cdot 3$

Expectation and Variance

Let's do Tortoise expectations and variances

- The Tortoise

$$
\begin{aligned}
E(T) &= (1/2)(1) + (1/2)(2) = 1.5 \\
Var(T) &= E(T^2) - E(T)^2 \\
&= (1/2)(1)^2 + (1/2)(2)^2 - (1.5)^2 = 0.25
\end{aligned}
$$

- Now the Hare's

$$
\begin{aligned}
E(H) &= (1/2)(0) + (1/2)(3) = 1.5 \\
Var(H) &= (1/2)(0)^2 + (1/2)(3)^2 - (1.5)^2 = 2.25
\end{aligned}
$$

Expectation and Variance

What do these tell us above the long run behavior?

- Tortoise and Hare have the same expected number of cars sold.
- Tortoise is more predictable than Hare.

  He has a smaller variance

  The standard deviations $\sqrt{Var(X)}$ are 0.5 and 1.5, respectively
- Given two equal means, you always want to pick the lower variance.

Covariance

$Cov(X, Y)$ is +ve if X & y move in the same direction

Suppose that we have two random variables X and Y

−ve if opposite direction

We need to measure whether they move together or in opposite directions

The Covariance is defined by

≈ 0 if sometime in the same dir & sometime in opposite.

$$Cov(X, Y) = E\left((X - E(X))(Y - E(Y))\right)$$

In terms of probability distributions, we need to calculate

$$Cov(X, Y) = \sum_{x,y} (x - E(X))(y - E(Y))P(x, y)$$

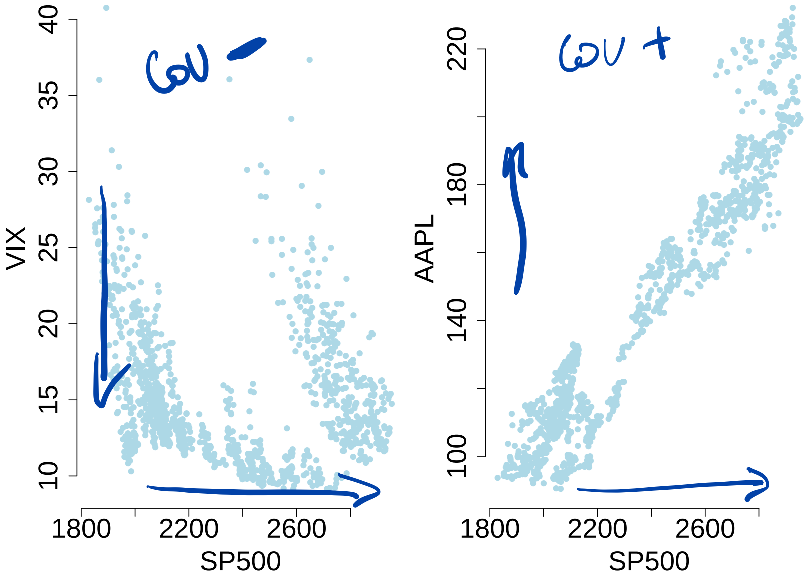
$x\nearrow$ $y\nearrow$: +

$x\swarrow$ $y\swarrow$: +

$x\nearrow$ $y\swarrow$: −

$x\swarrow$ $y\nearrow$: −

Lets look at Covariance on Markets



Correlation

$Cov(X, y) = \sum_{x,y} (X - E(x))(y - E(y))P(x,y)$

The Correlation is defined by

$$Corr(X, Y) = \frac{\overset{\$ \quad kg}{Cov(X, Y)}}{\underset{\$ \quad kg}{sd(X)sd(Y)}}$$

X \$ y kg

$Cov(X, y)$ \$ · kg

- What are the units of $Corr(X, Y)$?

  They don't depend on the units of X or Y!

- $-1 \leq Corr(X, Y) \leq 1$

If $Cov(Apple\ Sales, Economy) = 5$, $sd(Apple\ Sales) = 2$ and $sd(Economy) = 3.5$, then there's a 71.4% correlation

$$Corr(Apple\ Sales, Economy) = \frac{5}{2 \times 3.5} = \frac{5}{7} = 0.714$$

How often x, y move in the same dir

Linear Combinations of Random Variables

Two key properties:

$a = b = \frac{1}{2}$

Let a, b be given constants

$E(0.5X + 0.5Y) =$
$0.5E(X) + 0.5E(Y)$

Example

- Expectations and Variances

$$E(aX + bY) = aE(X) + bE(Y)$$

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) \boxed{+ 2abCov(X, Y)}$$

where $Cov(X, Y)$ is the covariance between random variables.

Tortoise and Hare

What about Tortoise and Hare?

We need to know $Cov(Tortoise, Hare)$.

Let's take $Cov(T, H) = -1$ and see what happens

Suppose $a = \frac{1}{2}, b = \frac{1}{2}$

Expectation and Variance

$$E\left(\frac{1}{2}T + \frac{1}{2}H\right) = \frac{1}{2}E(T) + \frac{1}{2}E(H) = \frac{1}{2} \times 1.5 + \frac{1}{2} \times 1.5 = 1.5$$
$$Var\left(\frac{1}{2}T + \frac{1}{2}H\right) = \frac{1}{4}0.25 + \frac{1}{4}2.25 - 2\frac{1}{2}\frac{1}{2} = 0.625 - 0.5 = \boxed{0.125}$$

Much lower!

Building a Portfolio of ETFs

What's the appropriate investment decision for you?

ETF=Exchange Traded Fund. There's many funds to choose from!

You have to decide which ones? and how much?

We'll see how the expected return (mean) and risk (volatility) math works for you

...

There's no free lunch! You'll have to take some risk ...

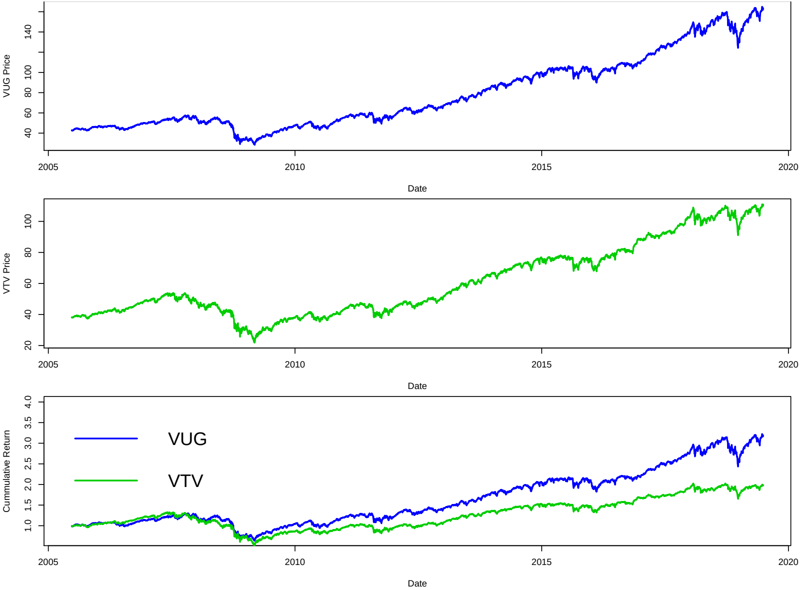
Building a Portfolio of ETFs: Assignment 1

Vanguard has a suite of ETFs. Here's a couple of combinations to choose from

1. Stocks and Bonds? `SPY` and `TLT`
2. Growth and Value? `VUG` and `VTV`
3. European or China? `VGK` and `FXI`

Let $P = aX + bY$ be your portfolio. What a, b do you choose?

Growth vs Value



Dummy Variables

A random variable that assigns a 1 when some event occurs and a 0 otherwise is called a **dummy variable**. It is called this because the number 1 “stands in” for the event.

Event	Value	Probability
Head	1	1/2
Tail	0	1/2

As before, we assign each outcome a probability, which together constitute the distribution of the random variable. It describes how the total probability mass is distributed across the various outcomes.

Recap:

X - R.V number we are not certain about

Distribution: Map from X to its probability

Joint $P(x, y)$

Conditional $P(x \mid y)$

Marginal $P(x)$ calculated from joint $P(x, y)$

Independence $P(x \mid y) = P(x)$

Bayes Rule $P(x \mid y) = \dfrac{P(x, y)}{P(x)} = \dfrac{P(y \mid x) P(y)}{P(x)}$

Binomial Distribution

Binomial R.V: Flip coin n times each time you have prob. p to win the toss. Binom R.V counts # of wins successes.

$n = 20$, $p = 1/2$

X Binom R.V. Take values from 0 to 20

Pattern Example: $n = 25$, $p = \frac{1}{2}$

$X \sim Bin(25, \frac{1}{2})$; $P(X = 19)$

$X \sim Bin(n, p)$

$$P(X = x) = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

$$\underbrace{1\,1\,1 \ldots 1}_{x}\;\underbrace{0\,0 \ldots 0}_{n-x}$$

Patriot

$X \sim Bin(25, \frac{1}{2})$

$$P(X = 19) = \frac{25!}{19!\,6!} \underbrace{\frac{1}{2}^{19} \frac{1}{2}^{6}}_{\frac{1}{2}^{25}}$$

Binomial Distribution

Coin flip (Binary event)

Bernoulli Trials: A sequence of repeated experiments are Bernoulli trials if:

1. The result of each trial is either a success or failure.
2. The probability p of a success is the same for all trials.
3. The trials are *independent*.

If X is the number of successes it is a Binomial Random Variable.

Binomial Distribution

$\binom{n}{x}$ n choose x $= \dfrac{n!}{x!(n-x)!}$

Normal approx $N(np, np(1-p))$

We calculate probabilities using:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$\binom{n}{x}$ counts the number of ways of getting x successes in n trials.

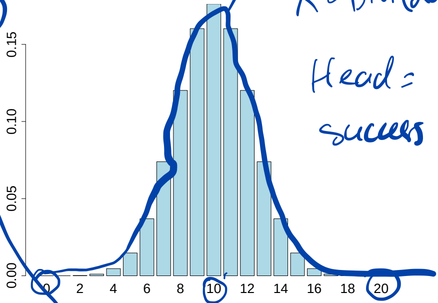
- The formula for $\binom{n}{x}$ is

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

where $n! =$

$n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$.

Binomial Mass Function



$X \sim Bin(20, \frac{1}{2})$

Head = success

In R: `dbinom(x, n, p)` and `rbinom(1000, n, p)`

$E(X) = \boxed{n \cdot p}$

R: B B B B B | B $= \sum_x x P(X=x)$

Binomial Distribution

	Var
$p=0$	0
$p=1$	0
n small	small
n large	large

The Mean and Variance of the Binomial are:

Binomial Distribution	Parameters
Expected value	$\mu = E(X) = np$
Variance	$\sigma^2 = Var(X) = np(1-p)$

$$Var(X) = \sum_x (x - np)^2 P(X=x)$$

Binomial: Example

Hits ~ Bin(4, 0.325)

P(hits > 2) = (

Assuming the Joe DiMaggio's batting average is 0.325 per at-bat and his hits are independent. What is the probability of getting more than 2 hits in 4 at-bats.

$$
\begin{aligned}
P(hits > 2) &= \underline{P(hits = 3)} + \underline{P(hits = 4)} \\
&= \binom{4}{3} p^3 (1-p) + \binom{4}{4} p^4 \\
&= \boxed{10.4\%}
\end{aligned}
$$

r = random # generator

d = density

Bernoulli Process = process of flipping coin n times

Let X represent either *success* ($X = 1$) or *failure* ($X = 0$)

- A Bernoulli process is a binary outcome with

$$P(X = 1) = p \text{ and } P(X = 0) = 1 - p$$

- We assume that trials are independent.

  The probability of two successes in a row is

$$\begin{aligned} P(X_1 = 1, X_2 = 1) &= P(X_1 = 1)P(X_2 = 1) \\ &= p \cdot p \\ &= p^2 \end{aligned}$$

Binomial distribution counts the successes of a *Bernoulli process*.

EPL Odds

$\frac{2}{5}$ prob of winning $\frac{2}{7}$

Best Odds Underlined

Odds Shortening

Odds Drifting

Sort By

	bet365	sky bet	Ladbrokes	William Hill	Marathon Bet	betfair	BetVictor	Paddy Power	Unibet	Coral	Betfred	betway	Black Type	Redzone	BoyleSports	SportPesa	For the bettor	sportingbet	Bethard	888sport	MoPlay	Spreadex	Royal Panda	SportNation	betfair Exchange	BETDAQ	Matchbook	Smarkets
Sign Up Offers	£	£20	£20	£30		£100	£40	£25	£40	£20	£40	£30	£10	£10	£25	£10	£50	£10	£50	£30	£30	£300	£20	£50	£20	£20	£10	£10
Special Offers		✓	✓ 2	✓ 2			✓ 3		✓	✓		✓			✓ 2	✓	✓			✓ 3				✓	✓	✓		
Each-way terms	2 1/4		2 1/4	2 1/3	2 1/3	2 1/4	2 1/3	2 1/4		2 1/4	2 1/3	2 1/4		2 1/3	2 1/4	2 1/4	2 1/4		2 1/4		2 1/4							
QuickBet																												
Man City	2/5	4/9		<u>1/2</u>		2/5	2/5	2/5	21/50	2/5	2/5			4/9	2/5	2/5	2/5	5/11	2/5	21/50	4/11		2/5	4/9	4/9		4/9	4/9
Liverpool ★	<u>3</u>	11/4		11/4		11/4	5/2	11/4	<u>3</u>	<u>3</u>	<u>3</u>			9/4	11/4	<u>3</u>	<u>3</u>	13/5	<u>3</u>	<u>3</u>	<u>3</u>		<u>3</u>	9/4	7/2	11/5	7/2	<u>7/2</u>
Tottenham	<u>20</u>	18		16		<u>20</u>	16	<u>20</u>	<u>20</u>	16	<u>20</u>			<u>20</u>	16	<u>20</u>	<u>20</u>	16	<u>20</u>	<u>20</u>	<u>20</u>		<u>20</u>	<u>20</u>	20	81/5	<u>21</u>	<u>21</u>
Chelsea	40	<u>50</u>		33		<u>50</u>	33	<u>50</u>	<u>50</u>	33	33			<u>50</u>	33	<u>50</u>	<u>50</u>	40	<u>50</u>	<u>50</u>	<u>50</u>		40	<u>50</u>	51	38	49	<u>53</u>
Man Utd	40	<u>50</u>		33		<u>50</u>	28	40	40	33	33			40	33	40	40	33	40	40	40		40	40	47		44	48
Arsenal	<u>50</u>	<u>50</u>		40		45	40	45	40	33	40			<u>50</u>	40	<u>50</u>	<u>50</u>	40	<u>50</u>	40	<u>50</u>		<u>50</u>	<u>50</u>	47	38	49	48
Everton	150	<u>250</u>		200		<u>250</u>	200	<u>250</u>	150	150	200			200	200	200	200	200	200	150	200		150	200	<u>446</u>	244	78	293
Wolves	200	200		200		<u>250</u>	150	<u>250</u>	150	150	150			200	200	200	200	200	200	150	200		200	200	<u>427</u>	293	78	293
Leicester	<u>300</u>	250		200		250	200	250	200	200	250			250	200	250	250	200	250	200	250		<u>300</u>	250	<u>531</u>		255	489
West Ham	500	<u>750</u>		500		500	400	500	500	500	500			500	500	500	500	500	500	500	500		500	500	<u>750</u>			489
Newcastle	1000	<u>1500</u>		750		500	400	<u>1500</u>	1000	750	250			1000	500	1000	1000	750	1000	1000	1000		1000	1000	949			979
Aston Villa	1000	<u>1500</u>		750		500	<u>1500</u>	500	1000	750	1000				1000	750	750	750	750	1000	750		Bet 1000	750	949		882	979

EPL 2017 Data

Data

We have a historical set of data on scores

home team	results		visit team
Chelsea	2	1	West Ham
Chelsea	5	1	Sunderland
Watford	1	2	Chelsea
Chelsea	3	0	Burnley
	...		

Tomorrow Manchester United (MU) is playing Hall U. I want to place a bet.

How can I predict the outcome of this game?

Poisson R.V.

Continious process

- Calls to call center
- Visits to website
- Attacks to IT system
- Goals scored by a team

\# Goals is Poisson R.V.

One parameter.

λ - average rate

EPL

\# goals per game

Poisson Distribution

$\lambda = 0$ $\checkmark$ or 0

$\lambda = 100$ λ

The Poisson distribution counts the occurrence of events

Given the rate λ we calculate probabilities as follows

$$P(X = x) = \boxed{\frac{e^{-\lambda}\lambda^{x}}{x!}} \text{ where } x = 0, 1, 2, 3, \ldots$$

The Poisson Mean and Variance are:

Poisson Distribution	Parameters
Expected value	$\mu = E(X) = \lambda$
Variance	$\sigma^2 = Var(X) = \lambda$

λ is the rate of occurrence of an event.

EPL MU

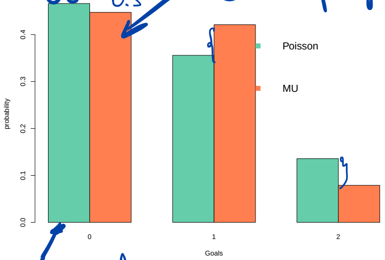
McL: λ against $\lambda_{for} = 1.6$

I can build a model assuming goals follow Poisson distribution.

I calculate λ by taking an average.

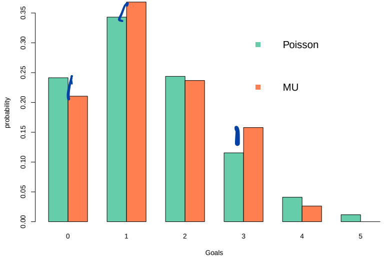
actual proportion from data

0.51 0.5



Predicted by Poisson model

(a) MU against



(b) MU for

Our Poisson model fits the empirical data!!

English Premier League: EPL

Calculate Odds for the possible scores in a match?

$$0-0,\ 1-0,\ 0-1,\ 1-1,\ 2-0, \ldots$$

analytical formula is hard.

Let

$X =$ Goals scored by Hall U

$Y =$ Goals scored by MU

$P(X - Y < 0)$ $P(X - Y = 0)$

What's the odds of a MU winning? $P(X < Y)$ Odds of a draw? $P(X = Y)$

```
x = rpois(100,0.6)
y = rpois(100,1.4)
sum(x<y)/100    # Team 2 wins
sum(x==y)/100    # Draw
```

λ Hall

λ MU

Simulate 100 games

sum(X-Y < 0)/100

EPL: Attack and Defence Strength

Each team gets an “attack” strength and “defence” weakness rating Adjust home and away average goal estimates

Team	Points	Goals for	'Attack strength'	Goals against	'Defence weakness'
Man United	87	67	1.46	24	0.52
Liverpool	83	74	1.61	26	0.57
Chelsea	80	65	1.41	22	0.48
Arsenal	69	64	1.39	36	0.78
Everton	60	53	1.15	37	0.80
Aston Villa	59	53	1.15	48	1.04
Fulham	53	39	0.85	32	0.70
Tottenham	51	44	0.96	42	0.91
West Ham	48	40	0.87	44	0.96
Man City	47	57	1.24	50	1.09
Stoke	45	37	0.80	51	1.11
Wigan	42	33	0.72	45	0.98
Bolton	41	41	0.89	52	1.13
Portsmouth	41	38	0.83	56	1.22
Blackburn	40	40	0.87	60	1.30
Sunderland	36	32	0.70	51	1.11
Hull	35	39	0.85	63	1.37
Newcastle	34	40	0.87	58	1.26
Middlesbrough	32	27	0.59	55	1.20
West Brom	31	36	0.78	67	1.46

EPL: Hull vs ManU

Poisson Distribution

Man U @ Home

adj factor 1.1

ManU Average away goals $= 1.47$. Prediction: $1.47 \times 1.46 \times 1.37 = \boxed{2.95}$

Average $\times$ Attack strength $\times$ $\boxed{\text{Defense weakness}}$

Hull Average home goals $= 1.47$. Prediction: $1.47 \times 0.85 \times 0.52 = \boxed{0.65}$.

	Team	Expected Goals	0	1	2	3	4	5
Simulation	Man U	2.95	7	22	26	12	11	13
	Hull City	0.65	49	41	10	0	0	0

EPL Predictions

A model is only as good as its predictions

- In our simulation Man U wins 88 games out of 100, we should bet when odds ratio is below 88 to 100.
- Most likely outcome is 0-3 (12 games out of 100)
- The actual outcome was 0-1 (they played on August 27, 2016)
- In out simulation 0-1 was the fourth most probable outcome (9 games out of 100)

Bookies vs Betters: The Battle of Probabilistic Models

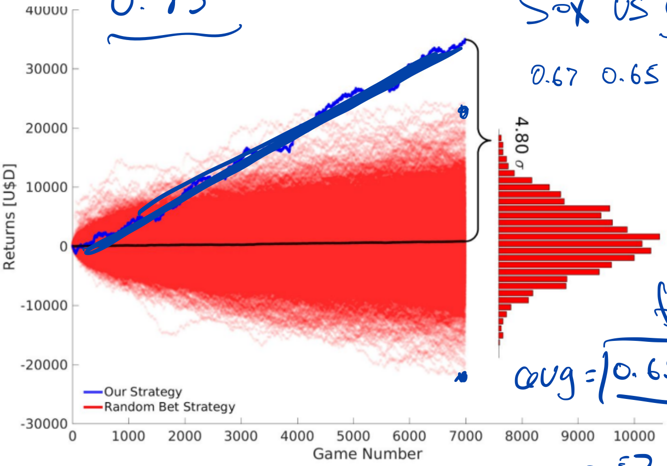
0.85

Sox vs Cubs

0.67 0.65 0.63 $\boxed{0.57}$

adjusted for Cubs

avg = $\boxed{0.65}$



0.57

bet on Sox

Source: The Secret Betting Strategy That Beats Online Bookmakers

Bookies vs Betters: The Battle of Probabilistic Models

- Bookies set odds that reflect their best guess on probabilities of a win, draw, or loss. Plus their own margin
- Bookies have risk aversion bias. When many people bet for an underdog (more popular team)
- Bookies hedge their bets by offering more favorable odds to the opposed team
- Simple algorithm: calculate average odds across many bookies and find outliers with large deviation from the mean

Discrete R.V. : Positive integer #
0, 1, 2, 3. . . .

Goog closed 2050.75

Prob that goog closes 2075.68 next trading day?

closes between 2000 & 2100

95% : Goog close price as continuous R.V. Assign prob to interval

Continuous Random Variables

Suppose we are trying to predict tomorrow's return on the S&P500...

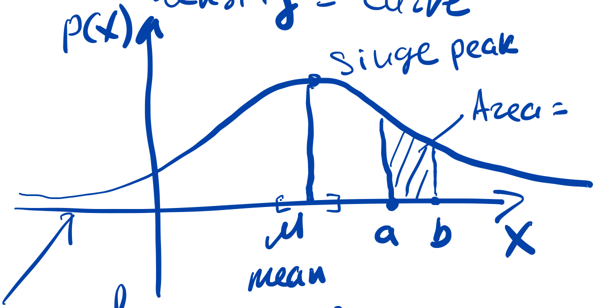
There's a number of questions that come to mind

- What is the random variable of interest?
- How can we describe our uncertainty about tomorrow's outcome?
- Instead of listing all possible values we'll work with intervals instead. The probability of an interval is defined by the area under the probability density function.

They are continuous (as opposed to discrete) random variables

Density for Continious R.V.

density = curve



Normal a.k.a Bell curve

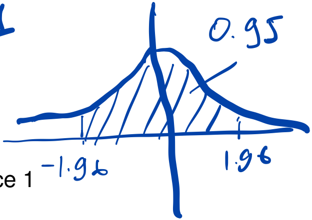
- Symetric
- Exponentially goes to 0 from μ

Corresponding R.V is called Normal

Variance (σ^2) = How fast it goes to 0
"thickness of the bell"

Normal Distribution

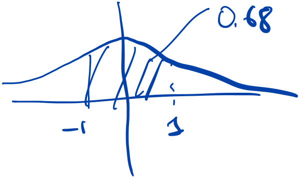
$\mu = 0$ & $\sigma^2 = 1$



Z is a standard normal random variable

- The standard Normal has mean 0 and has a variance 1

$$Z \sim N(0, 1)$$



- We have the probability statements

Calculates area under the curve

$$P(-1 < Z < 1) = 0.68$$

$$\boxed{P(-1.96 < Z < 1.96) = 0.95}$$

`qnorm` and $\boxed{\texttt{pnorm}}$ We can simulate 1000 draws using `rnorm(1000,0,1)`

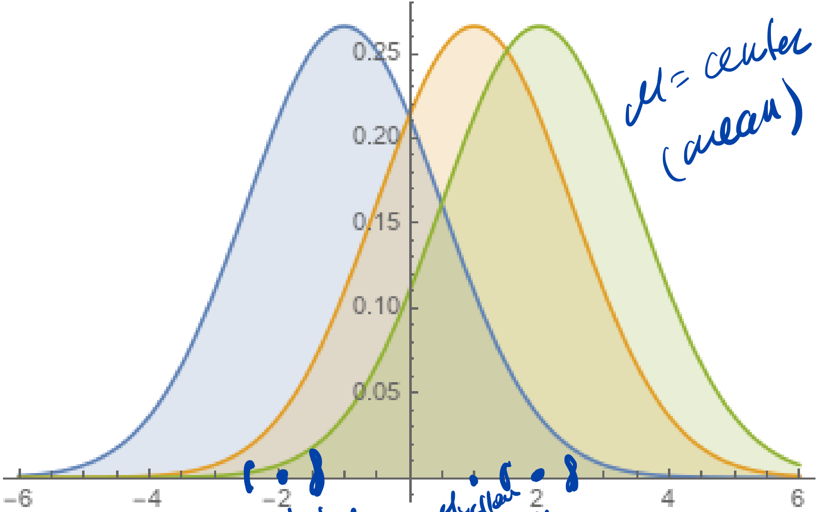
$P(X \leq a)$

How many μ σ

generates random #

Normal Distribution

By changing the mean parameter μ, we change the center of the bell curve



μ blue

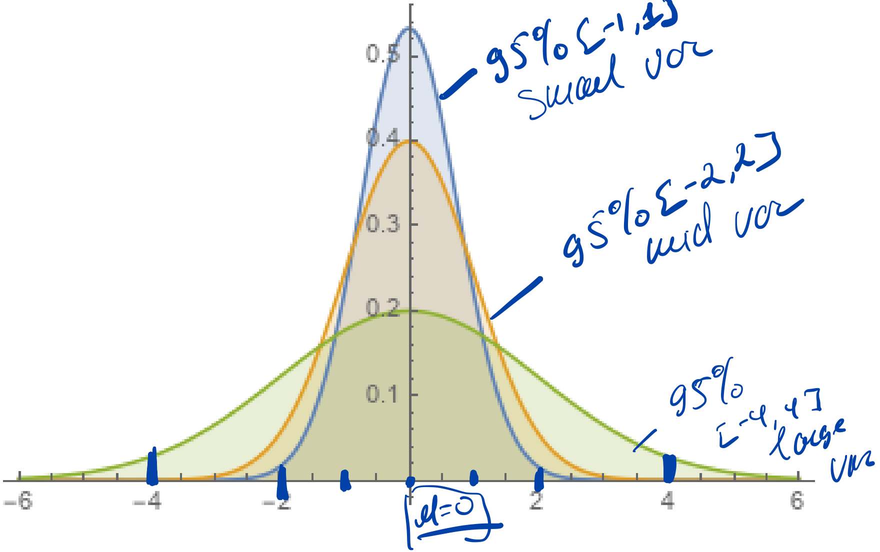
μ yellow

μ green

μ = center (mean)

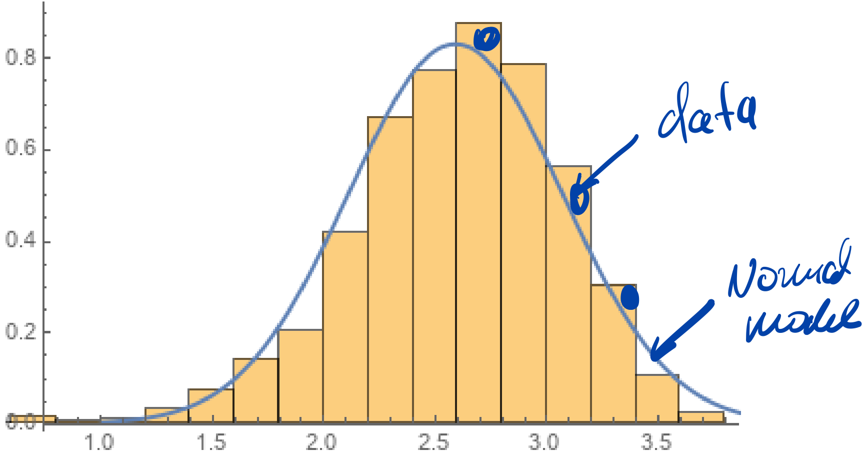
Normal Distribution

By changing the variance parameter σ^2, we change the “fatness” of the bell curve



Normal Distribution

Chicago Wind Speed (2007-2014) data on a log scale seem to be well described by the Normal distribution



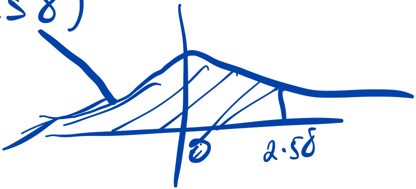
pnorm and qnorm

pnorm(2.58) = pnorm(2.58, 0, 1)

We can find probabilities and quintiles in R. Here are the important values

```
>pnorm(2.58)
[1] 0.9950
>pnorm(1.96)
[1] 0.9750
>pnorm(1.64)
[1] 0.9499
```

$= P(Z \le 2.58)$



qnorm is the inverse of pnorm. Simulation rnorm.

```
N=1000, x=rnorm(N,0,1), p=sum(x<1.96)/N
```

$P(Z \geq 2.58) = 1 - $ pnorm(2.58)

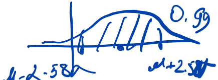
= pnorm(2.58, lower.tail = TRUE)

95% of vals are within 1.96 sigmas away from μ

Normal Distribution with General Mean and Variance

$Z \sim N(0, 1)$ 95% of values inside $[-1.96, 1.96]$

Here are two useful facts: If $X \sim \mathrm{N}(\mu, \sigma^2)$, then



$$P(\mu - 2.58\sigma < X < \mu + 2.58\sigma) = 0.99$$
$$\boxed{P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95.}$$

$X \sim N(\mu, \sigma^2)$ 95% of vals inside $[\mu - 1.96\sigma, \mu + 1.96\sigma]$

- The chance that X will be within 2.58σ of its mean is 99%, and the chance that it will be within 2σ of its mean is about 95%.

$X \sim N(0, \sigma^2)$ 95% of values inside $[-1.96\sigma, 1.96\sigma]$

$X \sim N(\mu, \sigma^2)$

$X \sim N(3, 1)$

$X - 3 \sim N(0, 1)$

$X \sim N(0, 4)$

$\sigma = 2$

$\frac{X}{2} \sim N(0, 1)$

$\sigma = 1$

$X \sim N(\mu, \sigma^2)$

$\frac{X - \mu}{\sigma} \sim N(0, 1)$

Normalize

95% of vals are in [-4, 4]

95% of vals are in [-2, 2]

$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

$X_1 \sim N(\mu_1, \sigma_1^2)$; $X_2 \sim N(\mu_2, \sigma_2^2)$ $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

The Normal Distribution

Our probability model is written $X \sim N(\mu, \sigma^2)$ μ is the mean, σ^2 is the variance

- Standardization if $X \sim N(\mu, \sigma^2)$ then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

- μ : the center of the distribution σ : how spread out the data are

  95% probability X is inside $\mu \pm 1.96\sigma$.

In R and Excel

In R: For a lower tail area, use *pnorm(z, mean, sd)*

Here *z* is where you want to compute the tail area, while “mean" and “sd” are the mean and standard deviation of the normal distribution, respectively.

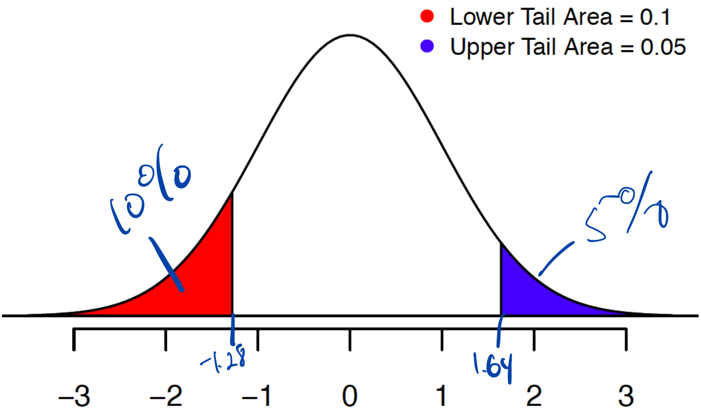
To compute an upper tail area, use *pnorm(z, mean, sd, lower.tail=F)*

In Excel: To compute a lower tail area, first click on a cell. In the formula bar, type NORMDIST(z,mean,sd,TRUE).

To find the *z* value corresponding to a specified lower tail area *p*, use the expression NORMINV(p,mean,sd) in the formula par, where *p* is the lower tail area.

Pictorially

$z \sim N(\mu = 0, \sigma^2 = 1)$



Examples of upper and lower tail areas. The lower tail area of 0.1 is at $z = -1.28$. The upper tail area of 0.05 is at $z = 1.64$

Example: The Crash

1. Future cannot be predicted from past (sometimes)

How extreme was the 1987 crash of -21.76%?

1. Prior to the October, 1987 crash SP500 monthly returns were 1.2% with a risk/volatility of 4.3% = s.d

$$X \sim N\left(0.012, 0.043^2\right);$$

$P(X \le -0.2176)$

pnorm(-0.2176, 0.012, 0.043) = $\boxed{10^{-8}}$

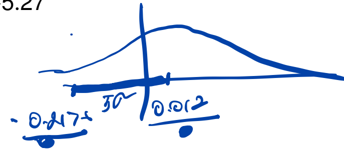
Standardize:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 0.012}{0.043} \sim N(0, 1)$$

2. Calculate the observed Z:

$$Z = \frac{-0.2176 - 0.012}{0.043} = -5.27$$

That's a 5-sigma event!

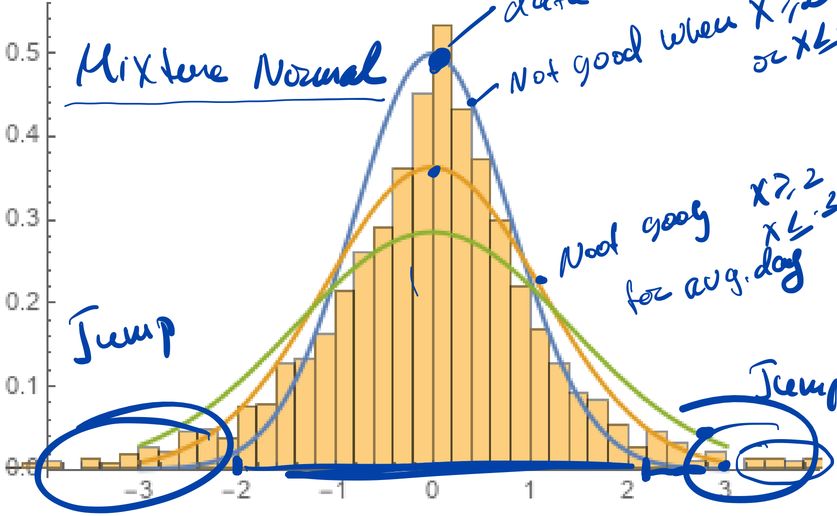


Example: The Crash

Merton: Normal + Jump

Nobel Prize

We assumed returns follow normal distribution. Using an inaccurate model can lead to inaccurate results.



Mixture Normal

data

Not good when $x \geq 2$ or $x \leq -2$

Noot goog for avg. day $x \geq 2$ $x \leq -3$

Jump

Jump

Normal as Approximaiton to Binomial

$X \sim Bin(1000, 0.05)$

$P(X \geq 125) = \boxed{P(X=125) + P(X=126) \dots P(X=1000)}$

A real estate firm in Florida offers a free trip to Florida for potential customers. Experience has shown that of the people who accept the free trip, 5% decide to buy a property. If the firm brings 1000 people, what is the probability that at least 125 will decide to buy a property?

In order to find the probability that *at least* 125 decide to buy, the binomial distribution would require calculating the probabilities for 125-1000. Instead, we use the normal approximation for the binomial.

Normal Approx

$X \approx N(50, 47.5)$

$$\mu = np = 50.$$

$\boxed{1 - \text{pnorm}(125, 50, 6.89)}$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1000 \times .05 \times .95} = \sqrt{47.5} = 6.89$$

$= 10^{-16}$

Calculating the Z-score for 125: $Z = \frac{125-50}{6.89} = 10.9$. and $P(Z \geq 10.9) = 0$.

10-Sigma event

Dating Problem.

250 Dates $\approx$ 70 days of dating

4 dates a day

1 2 3 250

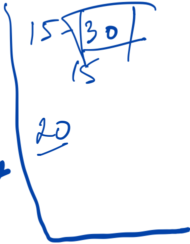
$\uparrow$

Decision making under uncertainty.

$\approx$ 80 dates (1/3 of budget)

Pick the best match = 8/10

Continue dating & pick first person with match $\geq$ 8



Probability and Making Decision

The Secretary Problem: also called the matching or marriage problem

- You will see items (spouses) from a distribution of types $F(x)$.

  You clearly would like to pick the maximum.

  You see these chronologically.

  After you decide no, you can't go back and select it.

- Strategy: wait for the length of time

$$\frac{1}{e} = \frac{1}{2.718281828} = 0.3678$$

  Select after you observe an item greater than the current best.

A Simple Example

Say I have a budget to consider 10 candidate, and they have the following scores (which I do not know)

```
1   7   4   5   3   8   10   3   0   11
```

I decide to screen 3 candidates before making decision

- My best candidate from the screening pool is 7
- Next is 5 –> dump
- Next is 3 –> dump
- Next is 8 –> accept

Probability and Decision

What's your best strategy? $1/3$

- Turns out its insensitive to the choice of distribution.
- Although there is the random sample i.i.d. assumption lurking.
- You'll not doubt get married between 18 and 60.

  Waiting $\frac{1}{e}$ along this sequence gets you to the age 32!

  Then, pick the next best person!

$P(x, y) = P(y, x)$

or

and

Business Statistics: 41000

Bayes Rules

Vadim Sokolov

The University of Chicago Booth School of Business

http://vsokolov.org/courses/41000/

posterior

prior

$$P(x \mid y) = \frac{P(x, y)}{P(y)} = \frac{P(y \mid x)\, P(x)}{P(y)}$$

Joint

marginal

Total

Bayes and AI

Old Style AI
Rule based:
If like phone then suggest case

Many applications in social media and business ...

1. Google Translate: 3 billion words a day
2. Amazon Alexa: Speech Recognition
3. Driverless Car: Waymo 1T/sec

Computing
Modeling
Data available

Shannon's autonomous mouse: Theseus

Bayes solves these problems!!

Modern day AI
$P(\text{Buy case} \mid \text{like phone})$
1. Probabilistic
2. Learned from data

Bayes and AI

What Does “AI" Really Mean? Think of an algorithm.

Two distinguishing features of AI algorithms:

1. Algorithms typically deal with probabilities rather than certainties.
2. There’s the question of how these algorithms “know" what instructions to follow.

Abraham Wald

How Abraham Wald improved aircraft survivability. Raw Reports from the Field

Type of damage suffered	Returned (316 total)	Shot down (60 total)
Engine	29	?
Cockpit	36	?
Fuselage	105	?
None	146	0

This fact would allow Wald to estimate:

$$P(\text{damage on fuselage} \mid \text{returns safely}) = 105/316 \approx 32\%$$

You need the inverse probability :

$$P(\text{returns safely} \mid \text{damage on fuselage})$$

Completely different!

Abraham Wald

Wald invented a method to implement the missing data, which is called by data scientist as “imputation". Wald Invented A Method for Reconstructing the Full Table

Type of damage suffered	Returned (316 total)	Shot down (60 total)
Engine	29	31
Cockpit	36	21
Fuselage	105	8
None	146	0

Then Wald got:

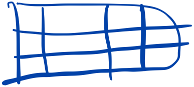
$$P(\text{returns safely} \mid \text{damage on fuselage}) = \frac{105}{105+8} \approx 93\%$$

$$P(\text{returns safely} \mid \text{damage on engine}) = \frac{29}{29+31} \approx 48\%$$

“Personalization" = “Conditional Probability"

Youtube

old



Conditional probability is how AI systems express judgments in a way that reflects their partial knowledge.

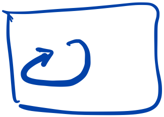
Google: Ad $P(\text{Click} \mid \text{Your History} + \text{Search query})$

Personalization runs on conditional probabilities, all of which must be estimated from massive data sets in which you are the conditioning event.

$P(\text{like a clip} \mid \text{Your History})$

Many Business Applications!! Suggestions vs Search, ...

New



watched
liked
searches

How does Netflix Give Recommendations?

Will a subscriber like Saving Private Ryan, given that he or she liked the HBO series Band of Brothers?

Both are epic dramas about the Normandy invasion and its aftermath.

100 people in your database, and every one of them has seen both films.

Their viewing histories come in the form of a big “ratings matrix".

	Liked Band of Brothers	Didn't like it
Liked Saving Private Ryan	56 subscribers	6 subscribers
Didn't like it	14 subscribers	24 subscribers

$$P(\text{likes Saving Private Ryan} \mid \text{likes Band of Brothers}) = \frac{56}{\underbrace{56+14}_{70}} = \boxed{80\%}$$

User 101

How does Netflix Give Recommendations?

	Users		New
movie titles	1	1	1
	0		1
		0	1
	1		1
		1	

But real problem is much more complicated:

1. Scale. It has 100 million subscribers and ratings data on more than 10,000 shows. The ratings matrix has more than a trillion possible entries.
2. "Missingness". Most subscribers haven't watched most films. Moreover, missingness pattern is informative.
3. Combinatorial explosion. In a database with 10,000 films, no one else's history is exactly the same as yours.

The solution to all three issues is careful modeling.

How does Netflix Give Recommendations?

User Feat

Movie Feat

Netflix Avg

Avg Movie

Avg for user

The fundamental equation is:

$$\underline{\text{Predicted Rating}} = \text{Overall Average} + \text{Film Offset} + \text{User Offset} + \boxed{\text{User-Film Interaction}}$$

These three terms provide a baseline for a given user/film pair:

- The overall average rating across all films is 3.7.
- Every film has its own offset. Popular movies have positive offsets.
- Every user has an offset. Some users are more or less critical than average.

Movie Title → Movie Features

Main Actor
Genre
Foreign
Post/Modern

User → User Features: Gender
Age
Like War Drama

Netflix

The User-Film Interaction is calculated based on a person's ratings of similar films exhibit patterns because those ratings are all associated with a latent feature of that person.

There's not just one latent feature to describe Netflix subscribers, but dozens or even hundreds. There's a “British murder mystery" feature, a “gritty character-driven crime drama" feature, a “cooking show" feature, a “hipster comedy films" feature, ...

The Hidden Features Tell the Story

P(Avg person likes a movie | Features of the movie)

High.

These latent features are the magic elixir of the digital economy–a special brew of data, algorithms, and human insight that represents the most perfect tool ever conceived for targeted marketing.

Your precise combination of latent features–your tiny little corner of a giant multidimensional Euclidean space–makes you a demographic of one.

Netflix spent $130 milion for 10 episodes on The Crown. Other network television: $400 million commissioning 113 pilots, of which 13 shows made it to a second season.

360 mil wasted

≈10%

Bayes's Rule in Medical Diagnostics

Dr. Survey

maj: ~~$P(C|+)$~~ $=0.8$

Alice is a 40-year-old women, what is the chance that she really has breast cancer when she gets positive mammogram result, given the conditions:

1. The prevalence of breast cancer among people like Alice is 1%. $P(C)=0.01$
2. The test has an 80% detection rate. $P(+|C)=0.8$
3. The test has a 10% false-positive rate. $P(+|\bar{C})=0.1$

$\bar{C}$ = not C

The posterior probability

$$P(\text{cancer} \mid \text{positive mammogram})?$$

$P(C|+)$ v.s $P(+|C)$

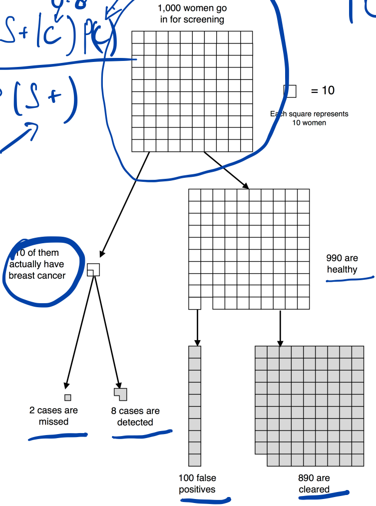
$$P(C \mid S+) \quad 0.074$$

$$= \frac{P(S+ \mid C)\, P(C)}{P(S+)} \quad 0.8$$

100 squares × 10

Very low

$$P(C \mid +) = \frac{P(+ \mid C)\, P(C)}{P(+)}$$



Each of the 1000 cases in light grey is equally likely.

Of 1000 cases:

- 108 positive mammograms. 8 are true positives. The remaining 100 are false positives.
- 892 negative mammograms. 2 are false negatives. The other 890 are true negatives.

positives = 100+8

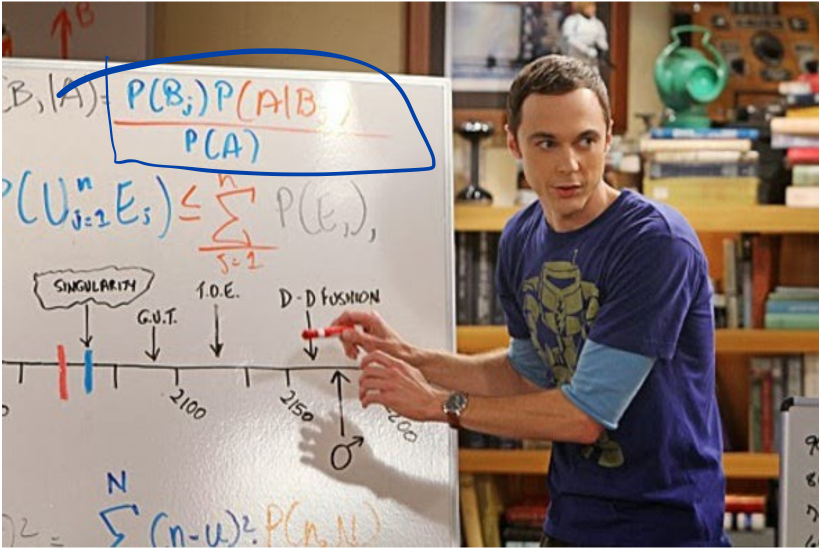
True positives = 8

$$P(+) = P(+, C) + P(+, \bar{C})$$
$$= P(+ \mid C)P(C) + P(+ \mid \bar{C})P(\bar{C})$$
$$P(\bar{C}) = 1 - P(C) = 0.99$$

$$\frac{8}{108} \simeq 0.074$$

Only 7.4%

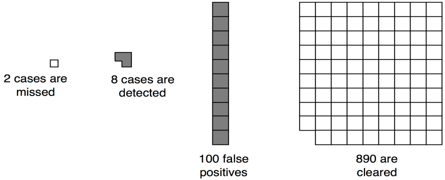
Bayes Rule



Calculation of Posterior Probability

$$P(\text{cancer} \mid \text{positive mammogram}) = \frac{8}{108} \approx 7.4\%$$

The 892 cases in white correspond to a negative test.
They now have zero probability.



The 108 cases in dark grey correspond to a positive test.
They are all equally likely.

Most women who test positive on a mammogram are healthy, because the vast majority of women who receive mammograms in the first place are healthy.

The Reverend and the Submarine (Thomas Bayes & John Craven)

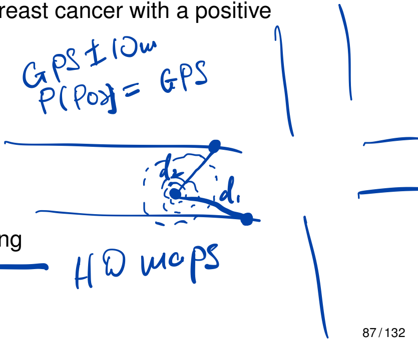
How did John Craven found a lost submarine from 140 square miles of ocean floor?

How do self-driving car find themselves and dodge bicycle, snow, and kangaroo?

How large is the probability of actually having breast cancer with a positive mammogram result?

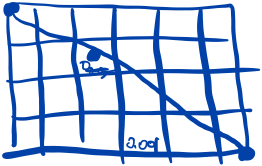
– Bayes's Rule $P(H \mid D) = \frac{P(H) \cdot P(D|H)}{P(D)}$

– SLAM: Simultaneous Localization And Mapping



The Story of Scorpion

Lazy Stone



The Scorpion is famous because one day, in 1968, it went missing, somewhere along a stretch of open ocean spanning thousands of miles. Despite the long odds, navy officials threw everything they had into the search.

To lead the search, the Pentagon turned to Dr. John Craven. To sort through this thicket of unknowns, Craven turned to his preferred strategy: Bayesian search. This methodology had been pioneered during World War II, when the Allies used it to help locate German U-boats.

Bayesian Search

Bayes

- Combine information from multiple sources to identify the location
- Air France Flight 447
- Brazilian navy has searched for 2 years. US Navy found in 2 weeks using Bayesian search
- MH370: Bayesian Methods in the Search for MH370 by Davey et.al.

Four essential steps of Bayesian search

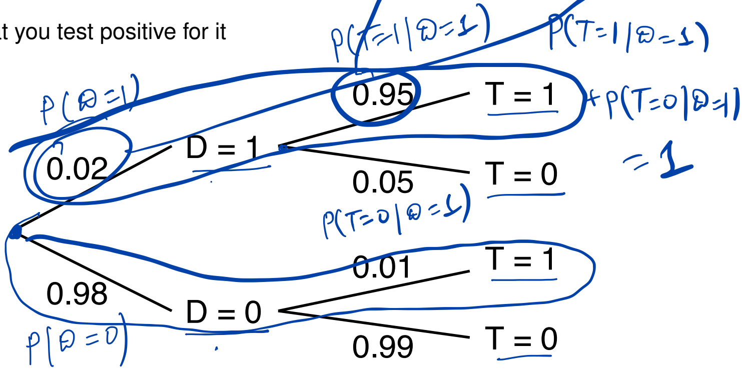
1. Create a map of prior probabilities over your search grid. Two sources of information, the presearch opinions of various experts and the capability of search instruments, are combined into prior.
2. Search the location of highest prior probability, for example, square C5.
3. If find nothing, revise your beliefs. Reduce the probability around square C5 and bump up the probability in the other regions accordingly.
4. Iterate steps 2 and 3.

Bayes's rule: prior belief + facts = revised belief.

Bayes Rule

$$P(D=1 \mid T=1) = \frac{P(T=1 \mid D=1)\, P(D=1)}{P(T=1)}$$

Disease Testing example .... Let $D = 1$ indicate you have a disease Let $T = 1$ indicate that you test positive for it



If you get a positive result, you are really interested in the question:

Given that you tested positive, what is the chance you have the disease?

$$P(T=1) = \underbrace{P(T=1 \mid D=1)}_{0.95}\, \underbrace{P(D=1)}_{0.02} + \underbrace{P(T=1 \mid D=0)}_{0.01}\, \underbrace{P(D=0)}_{0.98}$$

Bayes Rule

We have a joint probability table

		D	
		0	1
T	0	0.9702	0.001
	1	0.0098	0.019

Bayes Probability

$$p\,(D = 1 \mid T = 1) = \frac{0.019}{(0.019 + 0.0098)} = \boxed{0.66}$$

Bayes Rule

Let's think about this intuitively ... imagine you are about to test $100,000$ people.

- We assume that $2,000$ of those have the disease.
- We also expect that 1% of the disease-free people to test positive, i.e., 980 95% of the sick people to test positive. i.e. 1900. In total, we expect $2,880$ positive tests.
- Now choose one of those people at random ... what is the probability that he/she has the disease?

$$p\left(D=1 \mid T=1\right)=\frac{1,900}{2,880}=0.66$$

We get the same answer!!

Sensitivity and Specificity

Binary Predictive Rule

Type II

Type I

Two errors: An infected person may test negative, a well person tests positive.

False Negative

False positive

Sensitivity (or power) = true positive rate (or recall) % sick people who are correctly identified $P(T \mid D)$.

In a perfect world, we'd like $P(\bar{T} \mid D) \approx 0$

Sens = $\frac{\text{\# of positive tests among sick people}}{\text{\# sick people}}$

Specificity = true negative rate % of negatives correctly identified as such $P(\bar{T} \mid \bar{D})$.

- False negative = 1 - sensitivity, β where β is the type I error
- False positive rate = $1 -$ specificity, $\alpha =$ type I error

We want the probability: $P(D \mid T)$

Accuracy = $\frac{\text{\# of correct pred}}{\text{\# people}}$

Spec = $\frac{\text{\# neg test among healthy}}{\text{\# healthy people}}$

Confusion Matrix

False Positive Low Cost

False Negative High Cost

We can use accuracy rate:

$$\text{accuracy} = \frac{\text{\# of Correct answers}}{n}$$

$\boxed{\alpha = 0.2}$

$P(\text{Default} \mid \text{Credit History})$

or its dual, error rate

$$\text{error rate} = 1 - \text{accuracy}$$

$= 0.3 > \alpha \Rightarrow$ no loan

You remember, we haveP two types of errors. We can use confusion matrix to quantify those

	Predicted: YES	Predicted: NO
Actual: YES	TPR	FNR
Actual: NO	FPR	TNR

True positive rate (TPR) is the sensitivity and false positive rate (FPR) is the specificity of our predictive model

Bayes Classifier

$P(\text{Spam} \mid \text{Content})$
$\quad\;\; S \qquad\quad C$

$C = (w_1, w_2, \ldots w_n)$

$$P(S|C) = \frac{P(C|S)P(S)}{P(C)}$$

Training Data

$C_1 \quad S_1$
$C_2 \quad S_2 \qquad S_i$ is 0 or 1
$\ldots$
$C_m \quad S_m$

Output Y and input X:

Y : image of a cat X : identify "`cat`"

Silicon Valley: Season 4: Not Hotdog

What's our best decision?

Pick $\hat{Y}$ as the *most likely* category given that $X = x$, namely

$$\hat{Y} = \operatorname{argmax}_Y p(Y = y \mid X = x)$$

I need the probability table $P(X = x, Y = y)$ and marginal $P(X = x)$.

$$P(S|c) = \frac{P(c|S)\,P(S)}{P(c)}$$

w_i = sale

$P(w_i) = 0.001$

$P(w_i|S) = 0.01$

$$P(S) = \frac{\#\text{ of spam msg}}{m}$$

$$P(c|S) = P(w_1, w_2, \ldots, w_n | S) = P(w_1|S)P(w_2|S)\ldots P(w_n|S)$$

$$P(w_i|S) = \frac{\#\text{ of msg that have word } w_i}{\#\text{ of spam msg}}$$

$$P(c) = P(w_1, w_2 \ldots, w_n) = P(w_1)P(w_2)\ldots P(w_n)$$

$$P(w_i) = \frac{\#\text{ of msg that have word } w_i}{m}$$

Test Marketing a New Product

Basic Problem:

- Your company is developing a new product and will be test marketing to better gauge the sales of the new product.
- Based on positive, neutral or negative reactions, *what are the probability of high and low sales?*

NetFlix Bayes and AI

Test Marketing

$P(H) = 0.08$

$P(L) = 0.92$

$$P(H \mid Pos) = \frac{P(Pos \mid H)P(H)}{P(Pos)}$$

$$P(Pos) = P(Pos \mid H)P(H) + P(Pos \mid L)P(L)$$

Suppose you are given the following information

- New products introduced in the marketplace have high sales 8% of the time and low sales 92% of the time.
- A marketing test has the following accuracies:

  If sales are high, then consumer test reaction is positive 70%, neutral 25% and negative 5%.

  If sales are low, then consumer test reaction is positive 15%, neutral 35% and negative 50%.

$P(Pos \mid H) = 0.7$ $\quad$ $P(Pos \mid L) = 0.15$

$P(Neut \mid H) = 0.25$ $\quad$ $P(Neu \mid L) = 0.35$

$P(Neg \mid H) = 0.05$ $\quad$ $P(Neg \mid L) = 0.5$

Test Marketing

Step 1: Set-up your notation. Let

$$H = \text{high sales} \quad L = \text{low sales}$$

$$\text{Pos} = \text{positive} \quad \text{Neu} = \text{Neutral} \quad \text{Neg} = \text{Negative}$$

Step 2: List the known conditional probabilities For the marketing test we have

$$P(Pos \mid H) = 0.70, P(Neu \mid H) = 0.25, P(Neg \mid H) = 0.05$$

$$P(Pos \mid L) = 0.15, P(Neu \mid L) = 0.35, P(Neg \mid L) = 0.50$$

Finally, the base rates are $P(H) = 0.08$ and $P(L) = 0.92$

Test Marketing

Step 3: Describe the posterior probabilities that are required:

$$P(H \mid Pos)$$

The probability of high sales given a positive marketing test. Compute the probability of a positive test

$$\begin{aligned} \underline{P(Pos)} &= P(Pos \mid H)P(H) + P(Pos \mid L)P(L) \\ &= 0.70 \times 0.08 + 0.15 \times 0.92 = \boxed{0.194} \end{aligned}$$

Test Marketing

Now use Bayes Rule

$$
\begin{aligned}
P(H \mid Pos) &= \frac{P(Pos \mid H)P(H)}{P(Pos)} \\
&= \frac{0.70 \times 0.08}{0.194} = 0.288
\end{aligned}
$$

Hence 28.8% you'll have high sales in the market.

We should interpret this *relative* to our initial probability of only 8%.

Prior Posterior

$0.08 \xrightarrow[\text{Positive}]{} 0.288$

Two Headed Coin

$P(2H) = \frac{1}{1024}$

Evid = 10 Heads

$P(2H \mid 10H) = \frac{1}{2}$

Large jar containing 1024 fair coins and one two-headed coin.

- You pick one at random and flip it 10 times and get all heads.
- What's the probability that the coin is the two-headed coin?

$\frac{1}{1025}$ probability of initially picking the two headed coin. $\frac{1}{1024}$ chance of getting 10 heads in a row from a fair coin Therefore, it's a 50/50 bet.

$P(10H \mid \text{Fair Coin}) = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$

$P(2H \mid 10H) = \frac{\overset{1}{P(10H \mid 2H)}\, \overset{1/1025}{P(2H)}}{P(10H)}$

$P(10H) = P(10H \mid 2H) P(2H) + P(10H \mid FC) P(FC)$

Two Headed Coin

Let E be the event that you get 10 Heads in a row

$$P\left(\text{two headed} \mid E\right) = \frac{P\left(E \mid \text{two headed}\right) P\left(\text{two headed}\right)}{P\left(E \mid \text{fair}\right) P\left(\text{fair}\right) + P\left(E \mid \text{two headed}\right) P\left(\text{two headed}\right)}$$

Therefore, the posterior probability

$$P\left(\text{two headed} \mid E\right) = \frac{1 \times \frac{1}{1025}}{\frac{1}{1024} \times \frac{1024}{1025} + 1 \times \frac{1}{1025}} = \underline{0.50}$$

Apple Watch Series 4 ECG and Bayes' Theorem

$P(T \mid FB) = 98\%$ $\quad P(\bar{T} \mid \bar{FB}) = 99\%$

The Apple Watch Series 4 can perform a single-lead ECG and detect atrial fibrillation. The software can correctly identify 98% of cases of atrial fibrillation (true positives) and 99% of cases of non-atrial fibrillation (true negatives).

However, what is the probability of a person having atrial fibrillation when atrial fibrillation is identified by the Apple Watch Series 4?

$$P(FB \mid T) = \frac{P(T \mid FB)\, P(FB)}{P(T)} \quad ? \; 0.02$$

Bayes' Theorem:

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

$$P(T \mid \bar{FB}) = 1 - P(\bar{T} \mid \bar{FB})$$

$$P(T) = P(T \mid FB)P(FB) + P(T \mid \bar{FB})P(\bar{FB})$$

Apple Watch

Predicted	atrial fibrillation	no atrial fibrillation
atrial fibrillation	1960	980
no atrial fibrillation	40	97020

$$\boxed{0.6667} = \frac{0.98 \cdot 0.02}{0.0294}$$

The conditional probability of having atrial fibrillation when the Apple Watch Series 4 detects atrial fibrillation is about 67%. ?

The Game Show Problem: Assignment 1

Monte Hall *Let's make a Deal*.

You pick a door. Monty then opens one of the other two doors, revealing a goat.

Monty can't open your door or show you a car

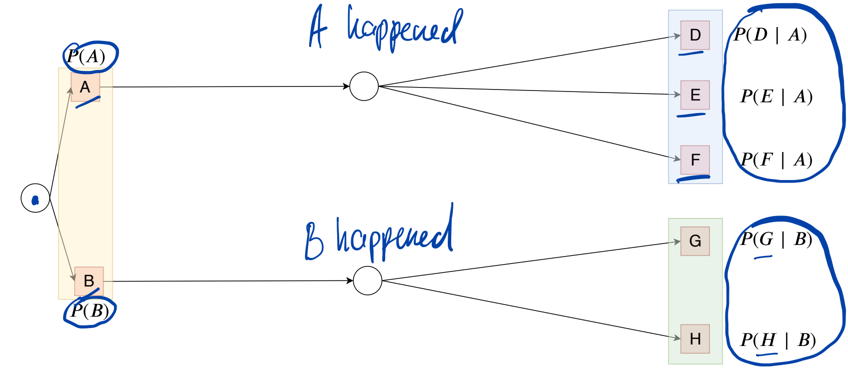
You have the choice of switching doors.

Is it advantageous to switch?

Assume you pick door A at random. Then $P(A) = (1/3)$.

You need to figure out $P(A \mid MB)$ after Monte reveals door B is a goat.

Decision Trees



$P(A) + P(B) = 1$

$P(D \mid A) + P(E \mid A) + P(F \mid A) = 1$

$P(G \mid B) + P(G \mid B) = 1$

H

Catastrophe Modeling

$P(S) = 0.01$

You live in a house that is somewhat prone to mud slides.

- Each rainy season there is a 1% chance of a mud slide occurring.
- You estimate that a mud slide would do $1 million in damage.
- You have the option of building a retaining wall that would help reduce the chance of a devastating mud slide.

  The wall costs $40,000 to build, and if the slide occurs, the wall will hold with a 95% probability.
- You also have the option of a Geologist's opinion.

Should you build the wall? Should you use the Geologist's Test and Bayes Rule?

60% Yes

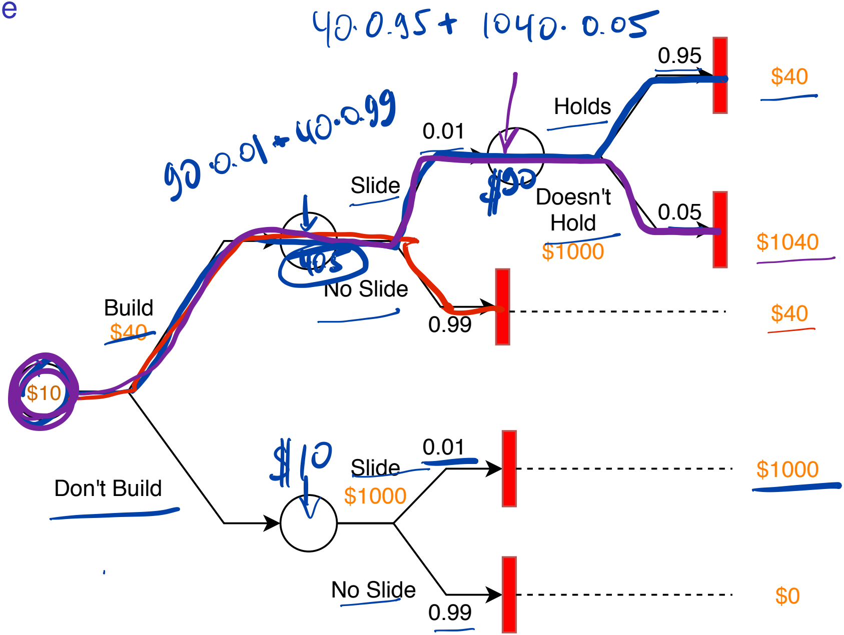
Decision Tree

Let's formally solve this as follows:

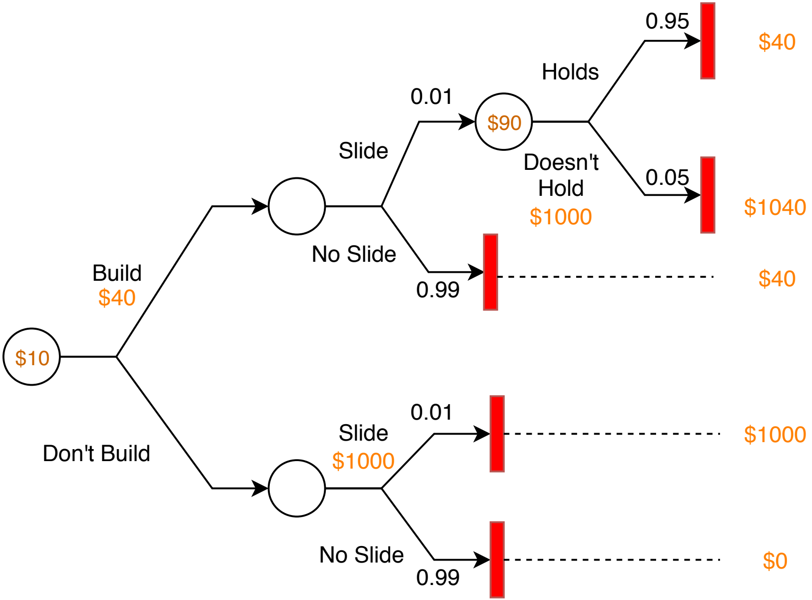
- Build a decision tree.
- The tree will list the probabilities at each node. It will also list any costs there are you going down a particular branch.
- Finally, it will list the expected cost of going down each branch, so we can see which one has the better risk/reward characteristics.

There's also the possibility of a further test to see if the wall will hold.

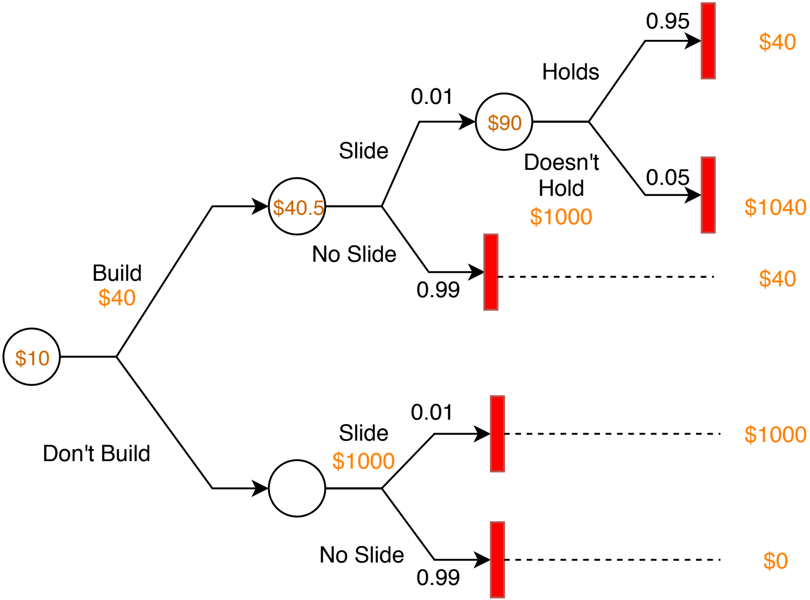
Tree


40·0.95 + 1040·0.05
90·0.01 + 40·0.99
0.95
$40
Holds
0.01
$90
Slide
Doesn't Hold
$1000
0.05
$1040
405
No Slide
0.99
$40
Build
$40
$10
$10
0.01
Slide
$1000
$1000
Don't Build
No Slide
0.99
$0

Tree

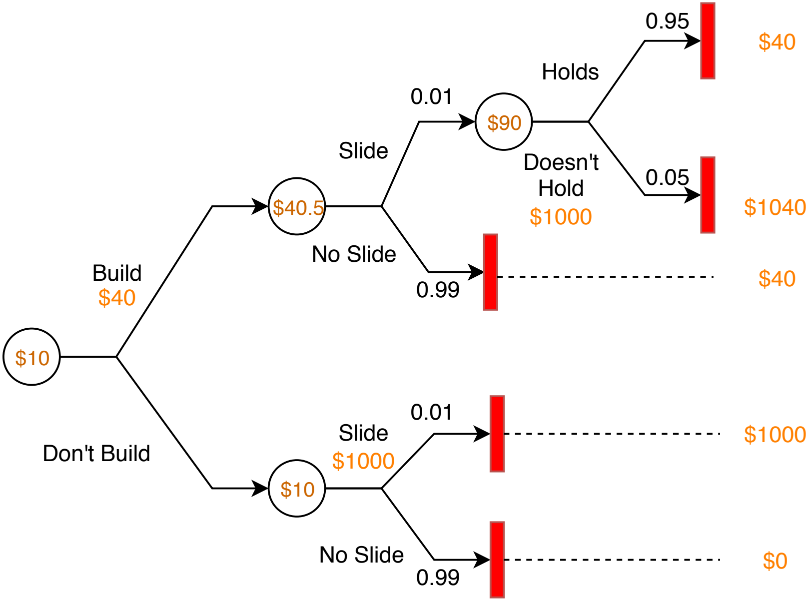

0.95
$40
Holds
0.01
$90
Slide
Doesn't
Hold
0.05
$1040
$1000
No Slide
$40
Build
$40
0.99
$10
0.01
Slide
$1000
Don't Build
$1000
$1000
No Slide
$0
0.99

Tree



0.95
$40
Holds
0.01
$90
Slide
Doesn't
Hold
0.05
$40.5
$1040
$1000
No Slide
$40
Build
0.99
$40
$10
0.01
Slide
$1000
Don't Build
$1000
No Slide
$0
0.99

Tree



0.95
$40
Holds
0.01
$90
Slide
Doesn't
Hold
0.05
$1000
$1040
$40.5
No Slide
0.99
$40
Build
$40
$10
0.01
$1000
Slide
$1000
Don't Build
$10
No Slide
0.99
$0

Testing

Let's include the testing option

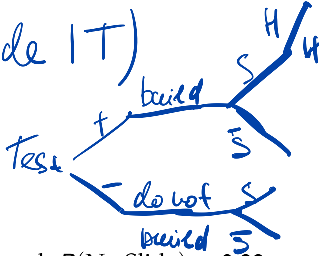
- You also have the option of having a test done to determine whether or not a slide will occur in your location.
- The test costs \$3000 and has the following accuracies.

$$P(T \mid \text{Slide}) = 0.90 \quad \text{and} \quad P(\bar{T} \mid \text{No Slide}) = 0.85$$

If you choose the test, then should you build the wall?

Bayes Rule

$P(\text{Slide} \mid T)$



Bayes Rule is as follows:

- The initial prior probabilities are

$$P(Slide) = 0.01 \quad \text{and} \quad P(\text{No Slide}) = 0.99$$

- Therefore

$$P(S \mid T) = \frac{P(T \mid S)\,P(S)}{P(T)}$$

$$P(T) = P(T \mid \text{Slide})P(\text{Slide}) + P(T \mid \text{No Slide})P(\text{No Slide})$$

$$P(T) = 0.90 \times 0.01 + 0.15 \times 0.99 = 0.1575$$

We'll use this to find our optimal course of action.

Bayes Probabilities

$P(Slide \mid T)$

The Bayes probability given a positive test is

$$
\begin{aligned}
P(Slide \mid T) &= \frac{P(T \mid Slide)P(Slide)}{P(T)} \\
&= \frac{0.90 \times 0.01}{0.1575} = 0.0571
\end{aligned}
$$

$$0.01 \xrightarrow[\uparrow \text{ Test}+]{} 0.057$$

Bayes Probabilities

$P(Slide \mid Not\ T)$

The Bayes probability given a negative test is

$$
\begin{aligned}
P\left(\text{Slide} \mid \bar{T}\right) &= \frac{P(\bar{T} \mid \text{Slide})P(\text{Slide})}{P(\bar{T})} \\
&= \frac{0.1 \times 0.01}{0.8425} \\
&= \underline{0.001187}
\end{aligned}
$$

Compare this to the initial base rate of a 1% chance of having a mud slide.

$0.01 \xrightarrow[\text{Test}]{} 0.001$

Probability Lose Everything

You build the wall without testing, what's the probability that you lose everything?

With the given situation, there is one path (or sequence of events and decisions) that leads to losing everything:

1. Build without testing (given)
2. Slide (0.01)
3. Doesn't hold (0.05)

$$P(\text{losing everything} \mid \text{build w/o testing}) = 0.01 \times 0.05 = 0.0005$$

Probability Lose Everything

You choose the test, what's the probability that you'll lose everything?

There are two paths that lead to losing everything:

1. First Path: There are three things that have to happen to lose everything

   Test +ve ($P = 0.1575$), Build, Slide ($P = 0.0571$), Doesn't Hold ($P = 0.05$)
2. Second Path: Now you lose everything if Test -ve ($P = 0.8425$), Don't Build,

   Slide given negative ($P = 0.001187$)

Conditional Probabilities

Expected loss

For the first path

$(3+40+1000) \cdot 0.00045$

$$P(\text{first path}) = 0.1575 \times 0.0571 \times 0.05 = 0.00045$$

For the second path

$(3+1000)\, 0.001$

$$P(\text{second path}) = 0.8425 \times 0.001187 = 0.00101$$

Hence putting it all together

$$P(\text{losing everything} \mid \text{testing}) = 0.00045 + 0.00101 = \boxed{0.00146}$$

Risk and Reward

Risk-Return Trade-off $\underline{\$760}$ diff

Choice	Expected Cost	Risk	P
Don't Build	$10,000	0.01	1 in 100
Build w/o testing	~~$40,500~~	0.0005	1 in 2000
Test	$10,760	0.00146	1 in 700

Expected Cost: Fee + Build + Loss

Expected cost: $3 + 40 \times 0.1575 + 1000 \times 0.00146$ or $\boxed{\$\ 10,760}$

What do you choose?

Summary

How to deal with uncertainty?

- Random Variables and Probability Distributions
- Joint and Conditional (Happy/Rich), Independence (Sally Clark),
- Expectation and Variance (Bookies vs Betters, Tortoise and Hare)
- Binomial Distribution (Patriot Coin Toss), Normal distribution (Crash of 1987)
- Decision Making under uncertainty (Marriage Problem and Probability and Decision Trees)
- Bayes Rule (Practice Hard $\neq$ Play in NBA)

Probability

- Random events
- Probability Distribution
- Independence
- Conditional Probability
- Decision trees
- Bayes and Decisions
- Bayes Examples
- Distributions
  - Normal: (a) (b) (c)
  - Binomial (a) (b) (c)

Working with Data

- Data Basics (variables, data tables, numerical vs categorical)
- Observational Studies & Experiments
- Sampling and sources of bias
- Experimental Design
- Scatter plots and histograms
- Mean and mode
- Variance
- log transform

Hypothesis Testing

- Another Introduction to Inference
- Hypothesis Testing (for a mean)
- HT (for the mean) examples
- Confidence Interval
  - Confidence Interval (for a mean)
  - Accuracy vs. Precision7
  - Required Sample Size for ME
  - CI (for the mean) examples

Significance and Errors

- Inference for Other Estimators
- Decision Errors
- Significance vs. Confidence Level
- Statistical vs. Practical Significance

Comparing means and t-distribution

- Introduction
- t-distribution
- Inference for a mean
- Inference for comparing two independent means
- Inference for comparing two paired means
- Power

Comparing proportions

Linear Model

- Introduction
- Correlation
- Residuals
- Least Squares Line
- Prediction and Extrapolation
- Conditions for Linear Regression
- R Squared
- Regression with Categorical Explanatory Variables
- Outliers in Regression
- Inference for Linear Regression

Multiple Linear Model

Classification

- Overview
- Logistic Regression
- Logistic Regression Details

Deep Learning

- What is a neural network?
- Supervised Learning with Neural Networks
- Why is Deep Learning taking off?
- Binary Classification
- Logistic Regression