Probability

Carlos Carvalho and Rob McCulloch

1. Discrete Random Variables

Probability and statistics let us talk about things we are unsure about.

- How much will Amazon sell next quarter?
- What will the return of my retirement portfolio be next year?
- How often will users click on a particular Facebook ad?
- If I give a patient a certain drug, how long are they likely to live?
- Will the Leafs win the Stanley Cup any time soon?

All of these involve inferring or predicting unknown quantities!!

Random Variables are numbers that we are *not sure about* but we might have some idea of how to describe the "likelihood" of the potential outcomes.

Example:

Suppose we are about to toss two coins.
Let X denote the number of heads that we are going to get.

We say that X, is the random variable that stands for the number we are not sure about.

- We describe our beliefs about a random variables with a Probability Distribution

- Example: If X is the random variable denoting the number of heads in two coin tosses, we can describe our beliefs through the following probability distribution:

x	$P(X = x)$
0	.25
1	.5
2	.25

x: a possible outcome
$P(X = x)$: the probability X turns out to be x.

In general

A random variable is *a number we're not sure about*.

Its *distribution* describes what we think it might turn out to be.

For a discrete random variable, we specify the distribution by:

- Listing all the possible numbers it can turn out to be.
- Assigning a probability to each possible outcome.
- Each probability is between 0 and 1.
- The probabilities add up to 1.

Note: "discrete" refers to the situation where can make the list (we have a countable set of possible outcomes).

Later we will look at continuous random variable where such a list is not practical.

The Bernoulli Distribution

A very common situation is that we are wondering whether something will happen or not.

Heads or tails, respond or don't respond, ...

It turns out to be very convenient to code up one possibility as a 0, and the other possibility as a 1.

The gives us the *Bernoulli distribution*.

$X \sim$ Bernoulli(p) means:

x	$P(X = x)$
0	1-p
1	p

Example:

I am about to toss a single coin.

X is the random variable which is 1 if the coin turns out to be heads and 0, if it is tails.

$$X \sim \text{Bernoulli}(.5)$$

Example:

Wall Street Journal, June 23, 2022.
Fed Chair Jerome Powell Says Higher Interest Rates Could Cause a Recession, by Nick Timiraos

> *Economists surveyed by the Wall Street Journal last week saw a 44% likelihood of a U.S recesssion in the next 12 months..*

R is 1 if recession in the next 12 months, 0 otherwise.

$$R \sim \text{Bernoulli}(.44)$$

Powell: "You should know that no one is very good at forecasting recessions very far out."

Example:

L is the random variable which is 1 if the Leafs win the Stanley Cup and 0, if not.

$$L \sim \text{Bernoulli}(?????)$$

it's a nightmare ...

Probability gets used in all kinds of ways.

Two key ways it is used are:

- We build models probabilistically.
- We measure our uncertainy given the information in the data and our model.

For example, in our fundamental regression model will have a random "error term" representing the part of y we cannot predict from the information in x.

We will use confidence intervals to probabilistically assess our uncertainty about key model parameters given the information in the data.

2. Conditional, Joint and Marginal Distributions

In general we want to use probability to address problems involving more than one variable at the time.

Let's suppose you are thinking about your sales next quarter.

Let S denote your sales (in thousands of units sold).

S is a number you are not sure about !!

In thinking about S, you find you are thinking about what will happen next quarter for the overall economy.

We need to think about two things we are uncertain about, the economy and sales!

Let E denote the performance of the economy next quarter.

Let $E = 1$ if the economy is expanding and $E = 0$ if the economy is contracting (what kind of random variable is this?).

Let's assume $P(E = 1) = 0.7$.

Let S denote my sales next quarter... and let's suppose we have the following probability statements:

s	$P(S = s \mid E = 1)$	s	$P(S = s \mid E = 0)$
1	0.05	1	0.20
2	0.20	2	0.30
3	0.50	3	0.30
4	0.25	4	0.20

These are called *Conditional Distributions*, they describe our beliefs about S conditional on knowing what happens for E.

We are more likely to have larger sales when $E = 1$ than when $E = 0$!!

How would $P(S = s \mid E = 1)$ relate to $P(S = s \mid E = 0)$ if E and S "had nothing to do with each other" ?

s	$P(S = s \mid E = 1)$	s	$P(S = s \mid E = 0)$
1	0.05	1	0.20
2	0.20	2	0.30
3	0.50	3	0.30
4	0.25	4	0.20

- In blue is the conditional distribution of S given $E = 1$
- In red is the conditional distribution of S given $E = 0$
- We read: *the probability of Sales of 4 ($S = 4$)* **given(or conditional on)** *the economy is growing ($E = 1$) is 0.25*

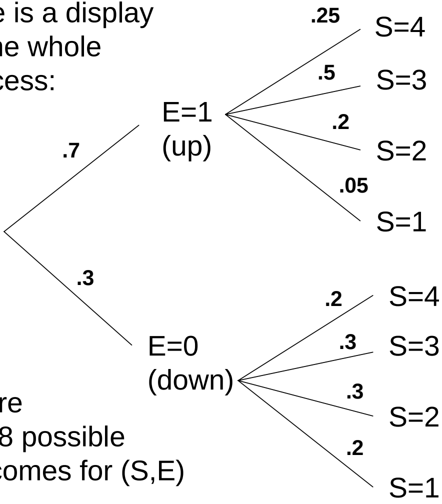
Our probability model for (E, S) captures the relationship between E and S through the difference in the two conditional distributions.

The conditional distributions tell us about about what can happen to S for a given value of E... but what about S and E jointly?

$$\begin{aligned} P(S = 4 \quad \text{and} \quad E = 1) &= P(E = 1) \times P(S = 4 | E = 1) \\ &= 0.70 \times 0.25 = 0.175 \end{aligned}$$

In english, 70% of the times the economy grows and 25% of those times sales equals 4, so 25% of 70% is 17.5%.

here is a display of the whole process:



p(S=4 and E=1) = .7*.25 = .175

p(S=3 and E=1) = .7*.5 = .35

p(S=2 and E=1) = .7*.2 = .14

p(S=1 and E=1) = .7*.05 = .035

p(S=4 and E=0) = .3*.2 = .06

p(S=3 and E=0) = .3*.3 = .09

p(S=2 and E=0) = .3*.3 = .09

p(S=1 and E=0) = .3*.2 = .06

There are 8 possible outcomes for (S,E)

We can specify the distribution of the pair of random variables (S, E) by listing all possible pairs and the corresponding probability.

(s, e)	$p(S = s, E = e)$
(1,1)	.035
(2,1)	.14
(3,1)	.35
(4,1)	.175
(1,0)	.06
(2,0)	.09
(3,0)	.09
(4,0)	.06

Question: What is $P(S = 1)$?

$$P(S=1) = P(S=1, E=1) + P(S=2, E=0) =$$
$$.035 + .06 = .095$$

```
> pv=c( .035, .14, .35, .175, .06, .09, .09, .06)
> cat('p(E=1) = ',sum(pv[1:4]),'\n')
> cat('p(E=0) = ',sum(pv[5:8]),'\n')
> cat('p(S=s): \n')
> print(pv[1:4] + pv[5:8])

p(E=1) =  0.7
p(E=0) =  0.3
p(S=s):
[1] 0.095 0.230 0.440 0.235
```

Summary:

In general the notation is:

- $P(Y = y, X = x)$ is the joint probability of that the random variable Y equals y **AND** the random variable X equals x.
- $P(Y = y|X = x)$ is the conditional probability that the random variable Y takes the value y **GIVEN** that X equals x.
- $P(Y = y)$ and $P(X = x)$ are the marginal probabilities of $Y = y$ and $X = x$

Fundamental Equation:

$$P(Y = y, X = x) = P(X = x)\, P(Y = y \,|\, X = x)$$

Warning:

The notation can get tricky.

Sometimes rather than writing

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

someone might write just,

$$p(x, y)$$

for the same thing!!

Usually, *but not always*, capitals are used for random variables and small case is used for possible values.

Marginals from Joint

$$p(X = x) = \sum_{y'} p(X = x, Y = y')$$

Or,

$$p(x) = \sum_{y'} p(x, y')$$

Example

$$p(E = 1) = \sum_{s'} p(E = 1, S = s') =$$

```
> temp = c(.035,.14,.35,.175)
> sum(temp)
[1] 0.7
```

Conditional, Joint and Marginal Distributions and Two-way Tables

Why we call marginals marginals... the table represents the joint and at the margins, we get the marginals.

		S				
		1	2	3	4	
E	0	.06	.09	.09	.06	.3
	1	.035	.14	.35	.175	.7
		.095	.23	.44	.235	1

Conditionals from Joints

We derived the joint distribution of (E, S) from the marginal for E and the conditional $S \mid E$.

You can also calculate the conditional from the joint by doing it the other way

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

$\Rightarrow$

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

Example... Given $E = 1$ what is the probability of $S = 4$?

		S				
		1	2	3	4	
E	0	.06	.09	.09	.06	.3
	1	.035	.14	.35	.175	.7
		.095	.23	.44	.235	1

$$P(S = 4|E = 1) = \frac{P(S = 4, E = 1)}{P(E = 1)} = \frac{0.175}{0.7} = 0.25$$

Example... Given $S = 4$ what is the probability of $E = 1$?

		S				
		1	2	3	4	
E	0	.06	.09	.09	.06	.3
	1	.035	.14	.35	.175	.7
		.095	.23	.44	.235	1

$$P(E = 1|S = 4) = \frac{P(S = 4, E = 1)}{P(S = 4)} = \frac{0.175}{0.235} = 0.745$$

What is the *conditional distribution* of $S \mid E = 1$?

We just compute $P(S = s|E = 1)$ for each $s = 1, 2, 3, 4$.

s	$P(S = s \mid E = 1)$
1	$.035/.7$
2	$.14/.7$
3	$.35/.7$
4	$.175/.7$

```
> pv = c(.035,.14,.35,.175)
> pv = pv/sum(pv)
> pv
[1] 0.05 0.20 0.50 0.25
```

$$P(S = s \mid E = e) \propto P(S = s, E = e)$$

In general, for X and Y

Given $X = x$,

$$P(Y = y \,|\, X = x) \propto P(X = x, Y = y).$$

where we think of x as fixed and y as varying over all possible values.

Independence

Two random variables X and Y are *independent* if

$$P(Y = y|X = x) = P(Y = y)$$

for all possible x and y.

In other words,

knowing X tells you nothing about Y!

e.g.,tossing a coin 2 times... what is the probability of getting H in the second toss given we saw a T in the first one?

Example:

You are about to toss two coins.

Let X_1 be 1 if the first coin is a head and 0 if tails.
Let X_2 be 1 if the second coin is a head and 0 if tails.

$$X_1 \sim \text{Bernoulli}(.5), \quad X_2 \sim \text{Bernoulli}(.5)$$

What is the probability of getting two heads in a row?

$$\begin{aligned} P(X_1 = 1, X_2 = 1) &= P(X_1 = 1)\, P(X_2 = 1 \mid X_1 = 1) \\ &= P(X_1 = 1)\, P(X_2 = 1) \\ &= (.5) \times (.5) \\ &= .25 \end{aligned}$$

IID

Our two coins X_1 and X_2 both have the same distribution and they are independent.

We say they are IID:

- I: independent
- ID: identically distributed

This terminology gets used a lot in statistics.

Example:

We say the two coins are IID Bernoulli with $p = .5$.

Suppose I am about to toss two dice.
Y_1 is the number on the face of the first die.
Y_2 is the number on the face of the second die.

Are Y_1, Y_2 IID?

Are Y_1, Y_2 IID Bernoulli?

Note:

If X and Y are independent then,

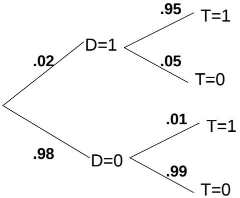
$$\begin{aligned} P(X = x, Y = y) &= P(X = x)P(Y = y|X = x) \\ &= P(Y = y)P(X = x|Y = y) \\ &= P(X = x)\,P(Y = y) \end{aligned}$$

The joint is the product of the marginals.

Disease Testing Example

You are about to be tested for a disease.

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



$\Rightarrow$

		D	
		0	1
T	0	.9702=.98*.99	.001
	1	.0098	.019

If you take the test and the result is positive, you are really interested in the question: Given that you tested positive, what is the chance you have the disease?

		D	
		0	1
T	0	.9702	.001
	1	.0098	.019

$$P(D = 1|T = 1) = \frac{0.019}{(0.019 + 0.0098)} = 0.66$$

Note:

In this example the *sensitivity* is .95.

The probability of a true postitive.

In this example the *specificity* is .99.

The probability of a true negative.

Bayes Theorem:

In the disease testing problem we formulated our understanding of the variables T and D using

$$p(t, d) = p(d)p(t|d).$$

Then we use probability theory to compute the quantity we really want which is

$$p(d \mid t).$$

This process of getting the probability "the other way" from how the modeling describes things is called *Bayes Theorem*.

We can develop a more formal statement of Bayes Theorem by writing things out using our basic properties of probability.

Suppose we have $p(y)$ and $p(x \mid y)$.

$$p(y|x) = \frac{p(y,x)}{p(x)} = \frac{p(y,x)}{\sum_{y'} p(y',x)} = \frac{p(y)p(x|y)}{\sum_{y'} p(y')p(x|y')}$$

For binary y (y is 0 or 1, as in our Disease testing problem), we have:

$$p(Y=1|x) = \frac{p(Y=1)\,p(x|Y=1)}{p(Y=0)\,p(x|Y=0) + p(Y=1)\,p(x|Y=1)}$$

$$p(Y=1|x) = \frac{p(Y=1)\, p(x|Y=1)}{p(Y=0)\, p(x|Y=0) + p(Y=1)\, p(x|Y=1)}$$

In the disease testing example Y is D and X is T:

$$p(D=1|T=1) = \frac{p(T=1|D=1)p(D=1)}{p(T=1|D=1)p(D=1)+p(T=1|D=0)p(D=0)}$$

$$p(D=1|T=1) = \frac{.95*.02}{.95*.02+.01*.98} = \frac{0.019}{(0.019+0.0098)} = 0.66$$

Or,

$P(D = d \mid T = 1) \propto (.0098, .019)$ for $d = 0, 1.$

```
> jv = c(.0098,.019)
> jv/sum(jv)
[1] 0.3402778 0.6597222
```

$P(D = d \mid T = 1) = (.34, .66)$ for $d = 0, 1.$

4. More Than Two Random Variables

Of course, we may want to think about more than two uncertain quantities at a time!!

Our ideas extend nicely to any number of variables.

For example with three random variables X_1, X_2, and X_3 we might want to think about:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$$

The probability that X_1 turns out to be x_1 and X_2 turns out to be x_2 and X_3 turns out to be x_3.

We can immediately extend our basic inutitive ideas:

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) =$$

$$P(X_1 = x_1)\, P(X_2 = x_2 \mid X_1 = x_1)\, P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2).$$

Example:

Suppose we have 10 voters.
4 are republican and 6 are democratic.

We "randomly" choose 3.

Let Y_i be 1 if the i^{th} voter is a democrat and 0 otherwise, $i = 1, 2, 3$.

What is

$$P(Y_1 = 1, Y_2 = 1, Y_3 = 1)$$

What is the probability of getting three democrats in a row ??

$P(Y_1 = 1, Y_2 = 1, Y_3 = 1) =$

$P(Y_1 = 1)\, p(Y_2 = 1 \mid Y_1 = 1)\, P(Y_3 = 1 \mid Y_1 = 1, Y_2 = 1)$

$= (6/10)(5/9)(4/8)$

$= (1/6) = .167.$

When we randomly pick a person from a population of people, and then randomly pick a second from the ones left, and so on, we call it *sampling without replacement*.

If we put the person back each time and randomly choose from the whole group each time, then we call it *sampling with replacement*.

Random Sampling in R

```
> x = 1:10
> x
 [1]  1  2  3  4  5  6  7  8  9 10
>
> set.seed(99)
> sample(x,5)
[1] 1 6 9 5 3
> sample(x,5)
[1] 10  2  8  6  5
>
> set.seed(99)
> sample(x,5)
[1] 1 6 9 5 3
>
> sample(1:10000,20)
 [1] 2922 7102 3200  358 8973 9188 7071 6724 6409 5278 4694 5972 1533 6560  398
[16] 4228   78    7 4818 6343
>
> set.seed(34)
> sample(1:10,5,replace=TRUE)
[1]  1  9 10  2  8
> sample(1:10,5,replace=TRUE)
[1] 6 6 4 3 3
```

Example:

Suppose we are tossing 100 coins.
Let X_i be 1 if the i^{th} coin is a head and 0 otherwise.

What is the probability of 100 heads in a row?

$$P(X_1 = 1, X_2 = 1, \ldots, X_{100} = 1) =$$

$$P(X_1 = 1)\, P(X_2 = 1 \mid X_1 = 1) \ldots P(X_j = 1 \mid X_1 = 1, X_2 = 1, \ldots, X_{j-1} = 1)$$
$$\ldots P(X_{100} = 1 \mid X_1 = 1, \ldots, X_{99} = 1)$$

$$= .5^{100} = 7.888609e - 31.$$

The 100 X_i are IID Bernoulli, with $p = .5$.

Question:

Suppose I get 100 heads in a row.

What is the probability the next one is a head?

Example:

Suppose I toss 100 dice.

Let Y_i be number on the ith die.

Are the Y_i IID?

Are the Y_i IID Bernoulli?

5. Probability and Decisions

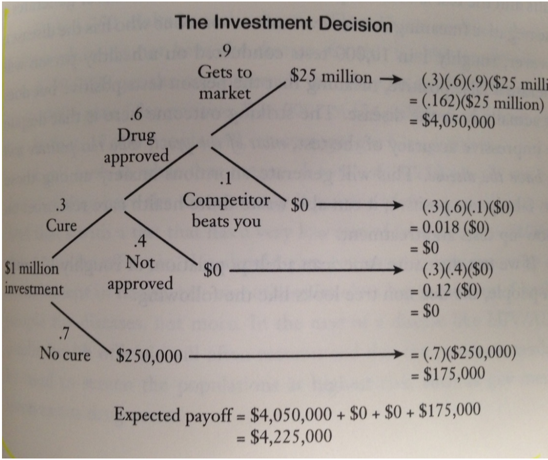
Suppose you are presented with an investment opportunity in the development of a drug... probabilities are a vehicle to help us build scenarios and make decisions.

You make a 1 million investment to develop the drug.

If "no cure" (drug does not work) you get 250,000 back you don't spend.

If you do find a cure you have to worry about whether it is approved and whether a competitor beats you out.

Notice at each outcome endpoint, the value is weighted by the probability of ending up there.



We basically have a new random variable, i.e, our revenue, with the following probabilities...

Note:
$.3^*.6^*.9 = 0.162$

Revenue	$P(Revenue)$
\$250,000	0.7
\$0	0.138
\$25,000,000	0.162

To get the expected revenue we add up the probability weighted revenues.

The expected revenue is then \$4,225,000...
So, should we invest or not?

What if you could choose this investment instead?

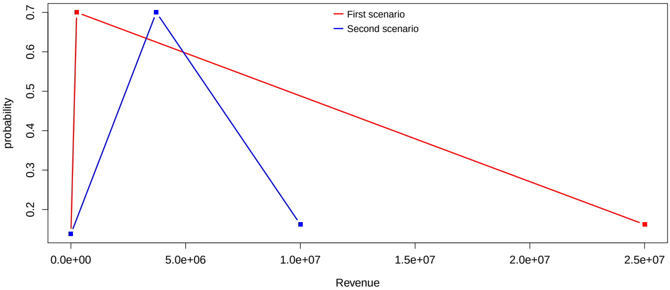
Revenue	*P*(*Revenue*)
\$3,721,428	0.7
\$0	0.138
\$10,000,000	0.162

```
rv2 = c(3721428,0,10000000)
pv = c(.7,.138,.162)
sum(rv2*pv)
[1] 4225000
```

The expected revenue is still \$4,225,000...

What is the difference?

Here is a plot of the two distributions for the two drug discovery scenarios.



Not clear you should use the probability weighted sum (*the expected value*) but people use it a lot.

Target Marketing

It costs you 80 cents to send out a promotion.
If the customer responds, you get a profit of 40 dollars.

Should we send the promotion ???

Well, it depends on how likely it is that the customer will respond!!

If they respond, you get 40-0.8=\$39.20.

If they do not respond, you lose \$0.80.

Let's assume your "predictive analytics" team has studied the conditional probability of customer responses given customer characteristics... (say, previous purchase behavior, demographics, etc)

Suppose that for a particular customer, the probability of a response is 0.05.

Profit	*P(Profit)*
\$-0.8	0.95
\$39.20	0.05

Should you do the promotion?

.95*(-.8) + .05*39.20 = 1.2.

6. Mean and Variance of a Random Variable

The probability weighted average is called the *expected value*.

The Mean or Expected Value is defined as (for a discrete X):

$$E(X) = \sum_{i=1}^{n} P(x_i) \times x_i$$

We weight each possible value by how likely they are... this provides us with a measure of centrality of the distribution... a "good" prediction for X!

We have already used this in our drug development and target marketing examples.

Expected Value of a Bernoulli

Suppose $X \sim \text{Bernoulli}(p)$,

$$X = \begin{cases} 1 & \text{with prob.} \quad p \\ 0 & \text{with prob.} \quad 1 - p \end{cases}$$

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} P(x_i) \times x_i \\ &= 0 \times (1 - p) + 1 \times p \\ E(X) &= p \end{aligned}$$

The Variance is defined as (for a discrete X):

$$Var(X) = \sum_{i=1}^{n} P(x_i) \times [x_i - E(X)]^2$$

Weighted average of squared prediction errors... This is a measure of spread of a distribution. More risky distributions have larger variance.

Variance of a Bernoulli

Suppose

$$X = \begin{cases} 1 & \text{with prob.} \quad p \\ 0 & \text{with prob.} \quad 1-p \end{cases}$$

$$\begin{aligned} Var(X) &= \sum_{i=1}^{n} P(x_i) \times [x_i - E(X)]^2 \\ &= (0-p)^2 \times (1-p) + (1-p)^2 \times p \\ &= p(1-p) \times [(1-p) + p] \\ Var(X) &= p(1-p) \end{aligned}$$

Question: For which value of p is the variance the largest?

The Standard Deviation

- What are the units of $E(X)$? What are the units of $Var(X)$?
- A more intuitive way to understand the spread of a distribution is to look at the standard deviation:

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

- What are the units of $sd(X)$?

Mean and Variance for Drug Development Example

Previously we computed the expected value for the drug development examples.
Let's review those calculations and compute the variances as well.

Same mean,

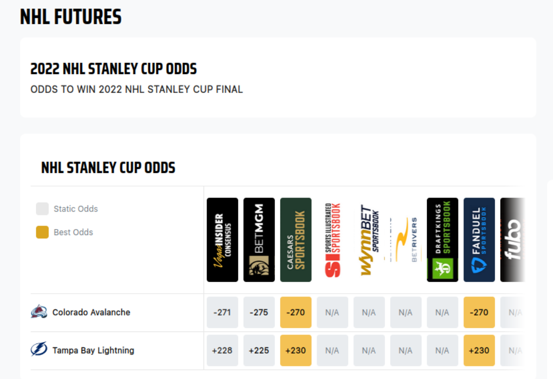
Different standard deviations!!

```
> pv = c(.7,.138,.162)
>
> d1 = c(250000,0,25000000)
> d2 = c(3721428,0, 10000000)
>
> cat("E1:",sum(pv*d1),"\n")
E1: 4225000
> cat("E2:",sum(pv*d2),"\n")
E2: 4225000
>
> M = sum(pv*d1)
>
> v1 = sum(pv*(d1-M)^2)
> v2 = sum(pv*(d2-M)^2)
> s1 = sqrt(v1)
> s2 = sqrt(v2)
>
> cat("s1,s2: ",s1,", ",s2,"\n")
s1,s2:  9134721 ,  2836141
```

The larger standard deviation in the second distribution reflects the fact that the second distribution is more "spread out", more "uncertain".

Example:

https://www.vegasinsider.com/nhl/odds/futures/, June 17th, 2022.



NHL FUTURES

2022 NHL STANLEY CUP ODDS

ODDS TO WIN 2022 NHL STANLEY CUP FINAL

NHL STANLEY CUP ODDS

Static Odds

Best Odds

	VegasInsider Consensus	BetMGM	Caesars Sportsbook	SI Sportsbook	WynnBET Sportsbook	BetRivers	DraftKings Sportsbook	FanDuel Sportsbook	fubo
Colorado Avalanche	-271	-275	-270	N/A	N/A	N/A	N/A	-270	N/A
Tampa Bay Lightning	+228	+225	+230	N/A	N/A	N/A	N/A	+230	N/A

What do these "odds" mean.

June 21, 1:24 pm, 2024.

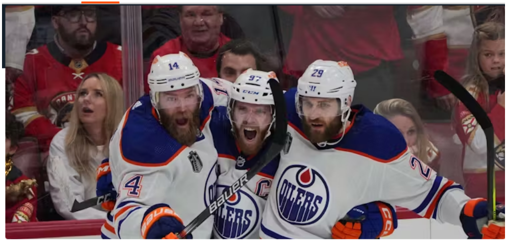


Photo By - USA TODAY Sports

The Stanley Cup odds are tightening up as the Edmonton Oilers defeated the Florida Panthers in Game 5, 5-3, to make it two straight wins with their backs against the wall.

The Cats are still favored over the Oilers as the series heads back to Alberta for Game 6 tonight, but by a much smaller margin then they were ahead of Game 4. Here are the latest NHL odds to hoist Lord Stanley's hardware ahead of this evening's contest.

Stanley Cup betting odds

We have the latest NHL odds, including odds to win the Stanley Cup, Game 6 odds, and series score odds.

Odds at a glance

- Edmonton Oilers to win: +270
- Florida Panthers to win: -345
- Florida Panthers in 6 games: -110
- Edmonton Oilers in 7 games: +260
- Florida Panthers in 7 games: +230

Odds courtesy of FanDuel Sportsbook.

https://www.thelines.com/betting/moneyline/

San Francisco 49ers (-380) at Detroit Lions (+290)

- Lions are the underdog, 49ers are favored to win.
- A $100 wager at +290 would pay $290 in profit if the Lions had won the game.
- In this example, the moneyline on the favored 49ers was -380. A bettor would need to wager $380 to win $100.

```
How do we figure out the implied probability from a moneyline?
Substitute the absolute value of the American odds
for "x" into these equations:

    Negative odds: x/(x+100)
    Positive odds: 100/(x+100)

Then, multiply the result by 100.

In the above example, San Francisco has a 79.16% to beat Detroit
 while the Lions have a 25.64% chance of pulling the upset.
```

```
> 380/(380+100)
[1] 0.7916667
> 100/(290+100)
[1] 0.2564103
```

https://www.thelines.com/betting/moneyline/

```
"Hang on," you may be saying to yourself.
"Those percentages add up to more than 100."

Good eye. That brings us to:
How does the sportsbook earn money booking moneyline bets?

Notice the "gap" between the two numbers in San Francisco vs. Detroit.
For example, San Francisco is -380 while Detroit is +290,
instead of Detroit being +380.
That difference in the numbers represents the vigorish,
commonly called the vig or the "juice"
{ what the bookmaker charges for accepting your action.

An easy way to see this is to imagine betting both sides.
If you put $380 on San Francisco and $100 on Detroit,
you would get back your original $480 no matter which team
won if Detroit was +380 instead of +290.
```

back to hockey ....

Tampa was $+230$.
So they are the underdog.

What is the probability (approximately)??

```
> 100/(230+100)
[1] 0.3030303
```

how would you get those formulas ???:

A bet is often called a *fair bet* if the expected value of winnings is 0.

Suppose we have an underdog.
With probability p the underdog wins so you win x if you bet on them.
With probability $(1 - p)$ you "win" (-100).

Solve for expected value of winnings equals 0.

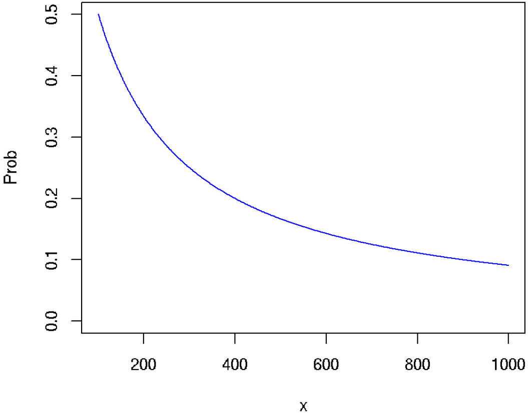
$$px + (1 - p)(-100) = 0$$

$$p(x + 100) = 100$$

$$p = \frac{100}{x+100}$$

Of course, your probability that the underdog wins may be much higher and that is why you bet might bet on them !!

```
x = seq(from=100,to=1000,length.out=500)
Prob = 100/(x+100)
plot(x,Prob,type="l",col='blue',ylim=c(0,.5))
```



Odds Ratio

A simple way to think about this is with the *odds ratio*.

Again assume we have an underdog. Again let p be the probability of winning on an underdog.

Let $q = (1 - p)$.

Then our equation is

$$p\,x - 100\,q = 0 \Rightarrow \frac{p}{q} = \frac{100}{x},\ \frac{q}{p} = \frac{x}{100}.$$

The quantity $\frac{p}{q}$ is often called the *odds ratio* (or just the odds).

So if something is $+200$, the odds is $200/100 = 2$ against the underdog winning.

June 19, 2023:

```
Odds

Colorado Avalanche +800

Toronto Maple Leafs +1100

Edmonton Oilers +1100

New Jersey Devils +1300

Boston Bruins +1300

Vegas Golden Knights +1300

Carolina Hurricanes +1300
```

So Leafs are 11 to 1, the odds is 11
(against the Leafs winning the Stanley Cup).

7. Covariance and Correlation

- A measure of *dependence* between two random variables...
- It tells us how two unknown quantities tend to move together

The Covariance is defined as (for discrete X and Y):

$$Cov(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \times [x_i - E(X)] \times [y_j - E(Y)]$$

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

Example:

$\mu_X = .1, \quad \mu_Y = .1.$

$\sigma_X = .05, \quad \sigma_Y = .05.$

		X	
		.05	.15
Y	.05	.4	.1
	.15	.1	.4

```
   x    y prob x-E(X) y-E(Y)    prod
0.05 0.05  0.4  -0.05  -0.05  0.0025
0.15 0.05  0.1   0.05  -0.05 -0.0025
0.05 0.15  0.1  -0.05   0.05 -0.0025
0.15 0.15  0.4   0.05   0.05  0.0025
```

$\text{Cov}(X, Y) = \sigma_{XY}$
$= .4 * .0025 + .1 * (-.0025) + .1 * (-.0025) + .4 * .0025 = .0015.$

Intuition: There is an 80% chance X and Y move *in the same direction*.

Example:

$\mu_X = .1, \quad \mu_Y = .1.$

$\sigma_X = .05, \quad \sigma_Y = .05.$

		X	
		.05	.15
Y	.05	.1	.4
	.15	.4	.1

x	y	prob	x-E(X)	y-E(Y)	prod
0.05	0.05	0.1	-0.05	-0.05	0.0025
0.15	0.05	0.4	0.05	-0.05	-0.0025
0.05	0.15	0.4	-0.05	0.05	-0.0025
0.15	0.15	0.1	0.05	0.05	0.0025

$$\text{Cov}(X, Y) = \sigma_{XY}$$
$$= .1 * .0025 + .4 * (-.0025) + .4 * (-.0025) + .1 * .0025 = -.0015.$$

Intuition: There is an 80% chance X and Y move *in opposite directions*.

Ford vs. Tesla

- Assume a very simple joint distribution of monthly returns for Ford (F) and Tesla (T):

	t=-7%	t=0%	t=7%	P(F=f)
f=-4%	0.06	0.07	0.02	**0.15**
f=0%	0.03	0.62	0.02	**0.67**
f=4%	0.00	0.11	0.07	**0.18**
P(T=t)	**0.09**	**0.80**	**0.11**	**1**

Let's summarize this table with some numbers...

	t=-7%	t=0%	t=7%	P(F=f)
f=-4%	0.06	0.07	0.02	**0.15**
f=0%	0.03	0.62	0.02	**0.67**
f=4%	0.00	0.11	0.07	**0.18**
P(T=t)	**0.09**	**0.80**	**0.11**	**1**

- $E(F) = 0.12$, $E(T) = 0.14$
- $Var(F) = 5.25$, $sd(F) = 2.29$,
  $Var(T) = 9.76$, $sd(T) = 3.12$

Tesla has the bigger mean *and* the bigger standard deviation.

Which is the better stock?

	t=-7%	t=0%	t=7%	P(F=f)
f=-4%	0.06	0.07	0.02	**0.15**
f=0%	0.03	0.62	0.02	**0.67**
f=4%	0.00	0.11	0.07	**0.18**
P(T=t)	**0.09**	**0.80**	**0.11**	**1**

$$\begin{aligned} Cov(F,T) =& (-7-0.14)(-4-0.12)0.06 + (-7-0.14)(0-0.12)0.03+ \\ & (-7-0.14)(4-0.12)0.00+(0-0.14)(-4-0.12)0.07+ \\ & (0-0.14)(0-0.12)0.62 + (0-0.14)(4-0.12)0.11+ \\ & (7-0.14)(-4-0.12)0.02 + (7-0.14)(0-0.12)0.02+ \\ & (7-0.14)(4-0.12)0.07 = 3.063 \end{aligned}$$

Okay, the covariance in positive... makes sense, but can we get a more intuitive number?

Correlation

$$Corr(X, Y) = \frac{Cov(X, Y)}{sd(X)sd(Y)}$$

- What are the units of $Corr(X, Y)$? It doesn't depend on the units of X or Y!
- $-1 \leq Corr(X, Y) \leq 1$

In our first example:
Corr(X,Y) = .0015/(.05*.05) = 0.6

In our second example:
Corr(X,Y) = -.0015/(.05*.05) = -0.6

In our Ford vs. Tesla example:

$$Corr(F, T) = \frac{3.063}{2.29 \times 3.12} = 0.428 \text{ (not too strong!)}$$

8. Linear Combinations of Random Variables

Is it better to hold Ford or Tesla? How about half and half?

What do we mean by "half and half"?

We mean the *portfolio* where we put half of our money into Ford and half into Tesla.

Return On a Portfolio:

Suppose we form a portfolio in which we put fraction w_1 of our wealth into an asset with return R_1 and fraction w_2 of our wealth into an asset with return R_2. Note that $w_1 + w_2 = 1$.
Let P be the return on the portfolio.

$$P = w_1 \, R_1 + w_2 \, R_2$$

In our Ford/Tesla use use R_1 = Ford, R_2 = Tesla, and $w_1 = w_2 = .5$.

Since the return on Ford and Tesla are random variables, so of course is the return on the portfolio!

Here is the joint distribution of $(R_1, R_2) = (F, T)$ *and* $P = .5\,F + .5\,T$.

	Ford	Tesla	P	prob
1	-4	-7	-5.5	0.06
2	0	-7	-3.5	0.03
3	4	-7	-1.5	0.00
4	-4	0	-2.0	0.07
5	0	0	0.0	0.62
6	4	0	2.0	0.11
7	-4	7	1.5	0.02
8	0	7	3.5	0.02
9	4	7	5.5	0.07

Is it better to hold Ford or Tesla? How about half and half?

We can compare the random returns based on the means and variances:
big mean: good, big variance: bad.

We could could compute the mean and variance of P directly from its distribution, but there are some very handy formulas for the mean and variance of a *linear combination* of random variables.

Let X and Y be two random variables, a, b, and c are **known** constants:

- $E(c + aX + bY) = c + aE(X) + bE(Y)$
- $Var(c + aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab \times Cov(X, Y)$

Applying this to the Ford vs. Tesla example...

- $E(0.5F + 0.5T) = 0.5E(F) + 0.5E(T) =$
  $0.5 \times 0.12 + 0.5 \times 0.14 = 0.13$
- $Var(0.5F + 0.5T) =$
  $(0.5)^2 Var(F) + (0.5)^2 Var(T) + 2(0.5)(0.5) \times Cov(F, T) =$
  $(0.5)^2(5.25) + (0.5)^2(9.76) + 2(0.5)(0.5) \times 3.063 = 5.28$
- $sd(0.5F + 0.5T) = \sqrt{5.28} = 2.297$

so, what is better? Holding Ford, Tesla or the combination?

```
asset: Ford, Tesla, Portfolio
mean: .12, .14, .13
sd:  2.29, 3.12, 2.297
```

Let's check the mean and variance of P from our basic formulas by computing them in R.

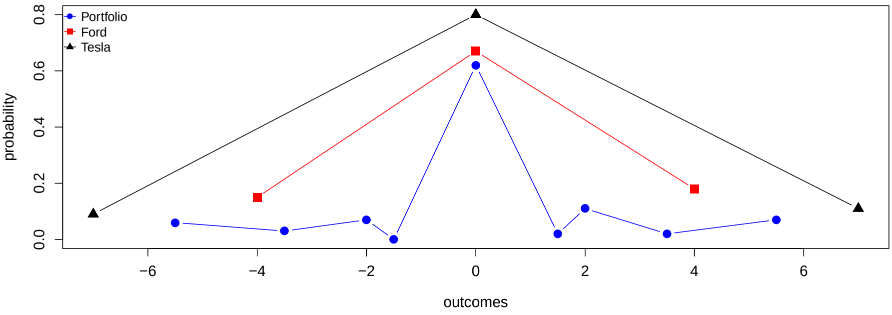
```
> head(ddf)
  Ford Tesla    P prob
1   -4    -7 -5.5 0.06
2    0    -7 -3.5 0.03
3    4    -7 -1.5 0.00
4   -4     0 -2.0 0.07
5    0     0  0.0 0.62
6    4     0  2.0 0.11

> EP = sum(ddf$prob * ddf$P)
> cat("Expected value of port: ",EP,"\n")
Expected value of port:  0.13
> VP = sum(ddf$prob * (ddf$P-EP)^2)
> cat("Variance of port: ",VP,"\n")
Variance of port:  5.2931
```

Same numbers!!

Let's see what is going on graphically!

- plot the three distributions for Ford, Tesla, and the portfolio
- possible values on the x axis, probabilities on the y axis
- easy to see that Tesla has a higher variance than Ford.
- not so easy to see the difference in the means, this is realistic
- you can see you the diversification killed the tails



Note:

If X and Y are independent, then $Cov(X, Y) = 0$.

Covariance measures linear dependence.

If they have nothing to do with each other (independence), then they certainly have nothing to do with each other linearly.

More generally...

- $E(w_0 + w_1X_1 + w_2X_2 + ...w_pX_p) =$
  $w_0 + w_1E(X_1) + w_2E(X_2) + ... + w_pE(X_p) = w_0 + \sum_{i=1}^{p} w_iE(X_i)$

- $Var(w_0 + w_1X_1 + w_2X_2 + ...w_pX_p) =$
  $w_1^2 Var(X_1) + w_2^2 Var(X_2) + ... + w_p^2 Var(X_p)$
  $+2w_1w_2Cov(X_1, X_2) + 2w_1w_3Cov(X_1, X_3) + ... =$
  $\sum_{i=1}^{p} w_i^2 Var(X_i) + 2\sum_{i=1}^{p}\sum_{j>i} w_iw_jCov(X_i, X_j)$

Example:

Ford, Tesla, GM.

$$P = w_1 F + w_2 T + w_3 G$$

$$E(P) = w_1 E(F) + w_2 E(T) + w_3 E(G)$$

$$\begin{aligned} Var(P) = & \, w_1^2 Var(F) + w_2^2 Var(T) + w_3^2 Var(G) \\ & + 2w_1 w_2 Cov(F, T) + 2w_1 w_3 Cov(F, G) + 2w_2 w_3 Cov(T, G). \end{aligned}$$

With lots of assests this gets complicated!! There many possible pairs of assests and corresponding covariance pairs representing the high dimensional dependence of the many input assets.

In practice you have to estimate all the covariances from data, another good reason to index!!

Example, Sum and Mean of IID

Suppose you play a game n times and the winning from the ith play is represented by the random variable X_i, $i = 1, 2, \ldots, n$.

We assume the each play of the game is independent of the others and it is the same game each time.

So, the X_i are IID.

What are the mean and variance of the *total* winnings?

$$T = X_1 + X_2 + X_3 + \ldots + X_n.$$

Let $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

$$T = X_1 + X_2 + X_3 + \ldots + X_n.$$

The Mean:

$$\begin{aligned} E(T) &= E(X_1) + E(X_2) + \ldots + E(X_n) \\ &= \mu + \mu + \ldots + \mu \\ &= n\,\mu \end{aligned}$$

The variance:

For the variance note that *because all the X_i are independent, the covariances are all 0 !!!!!*

$$\begin{aligned} Var(T) &= Var(X_1) + Var(X_2) + \ldots + Var(X_n) \\ &= \sigma^2 + \sigma^2 + \ldots + \sigma^2 \\ &= n\,\sigma^2 \end{aligned}$$

And the average:

$$\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \frac{1}{n}X_3 + \ldots + \frac{1}{n}X_n.$$

The Mean:

$$\begin{aligned} E(\bar{X}) &= \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \ldots + \frac{1}{n}E(X_n) \\ &= \frac{1}{n}\mu + \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu \\ &= n\,(\frac{1}{n})\,\mu = \mu \end{aligned}$$

$$\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \frac{1}{n}X_3 + \ldots + \frac{1}{n}X_n.$$

The variance:

$$\begin{aligned}
Var(\bar{X}) &= \frac{1}{n^2}Var(X_1) + \frac{1}{n^2}Var(X_2) + \ldots + \frac{1}{n^2}Var(X_n) \\
&= \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \ldots + \frac{1}{n^2}\sigma^2 \\
&= n\,(\frac{1}{n^2})\sigma^2 \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

Note:

We have, for X_i, IID, $E(X_i) = \mu$, $Var(X_i) = \sigma^2$,

$$E(\bar{X}) = \mu, \quad Var(\bar{X}) = \frac{\sigma^2}{n}.$$

Intuitively this says the average of a lot of IID draws tends to be closer to the mean μ than an individual draw.

We do a lot of averaging in statistics !!

This will turn out to be important !!

Portfolio vs. Single Project (from Misbehaving)

In a meeting with 23 executives plus the CEO of a major company economist Richard Thaler poses the following question:

Suppose you were offered an investment opportunity for your division (each executive headed a separate/independent division) that will yield one of two payoffs. After the investment is made, there is a 50% chance it will make a profit of $2 million, and a 50% chance it will lose $1 million. Thaler then asked by a show of hands who of the executives would take on this project. Of the twenty-three executives, only three said they would do it.

Then Thaler asked the CEO a question. If these projects were independent, that is, the success of one was unrelated to the success of another, how many of the projects would he want to undertake? His answer: all of them!

What ?????!!!!

How can we understand this?

for an individual executive:

$X_i, i = 1, 2, \ldots, 23.$

x	$P(X = x)$
-1	.5
2	.5

μ: $.5 * (-1) + .5 * 2 = 0.5$

σ^2: $.5 * (-1 - .5)^2 + .5 * (2 - .5)^2 = 2.25$
σ: 1.5
μ/σ: $.5/1.5 = 0.3333333$

for CEO:

$$T = X_1 + X_2 + X_3 + \ldots + X_n.$$

$E(T)$: 23*.5 = 11.5

$Var(T)$: 23*2.25 = 51.75

$sd(T)$: sqrt(51.75) = 7.193747

$E(T)/sd(T)$: 11.5/7.193747 = 1.598611

For the CEO, the mean is much bigger relative to the standard deviation that is it for the individual managers !!!

Companies, CEO's, managers have to be careful in setting incentives that avoid what psychologist and behavior economists call "narrow framing"... otherwise, what can be perceived to be bad for one manager may be very good for the entire company!

Note, the Sharpe Ratio

Wikipedia:

$$S_a = \frac{E(R_a - R_b)}{\sigma_a}$$

where:

- R_a is a risky asset (and hence a random variable)
- R_b is the risk free return, and hence a constant (such as a U.S. Treasury security)
- σ_a standard deviation of $R_a - R_b$, which is the standard deviation of R_a.

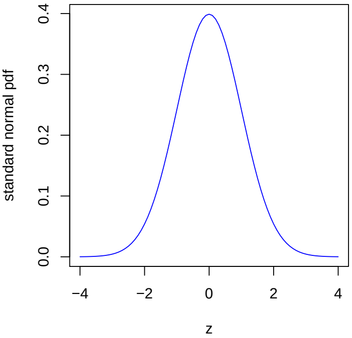
9. Continuous Random Variables

- Suppose we are trying to predict tomorrow's return on the S&P500...
- Question: What is the random variable of interest?
- Question: How can we describe our uncertainty about tomorrow's outcome?
- Listing all possible values seems like a crazy task... we'll work with intervals instead.
- These are called continuous random variables.
- The probability of an interval is defined by the area under the *probability density function, the "pdf"*.

The Normal Distribution

- A random variable is a number we are NOT sure about but we might have some idea of how to describe its potential outcomes.
- The probability the number ends up in an interval is given by the area under the curve (pdf)

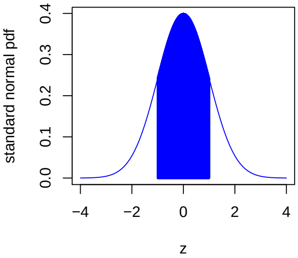
This is the pdf for the *standard normal* distribution.

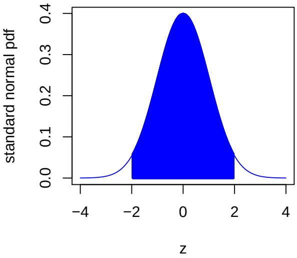


Notation: We often use Z, to denote a standard normal random variable.

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

$$P(-1.96 < Z < 1.96) = 0.95$$





Note:

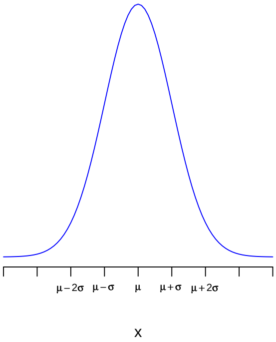
For simplicity we will often use $P(-2 < Z < 2) \approx 0.95$

Questions:

- What is $P(Z < 1)$? How about $P(Z \leq 1)$?
- What is $P(Z < 0)$?

- The standard normal is not that useful by itself. When we say "the normal distribution", we really mean a family of distributions.
- We obtain pdfs in the normal family by shifting the bell curve around and spreading it out (or tightening it up).

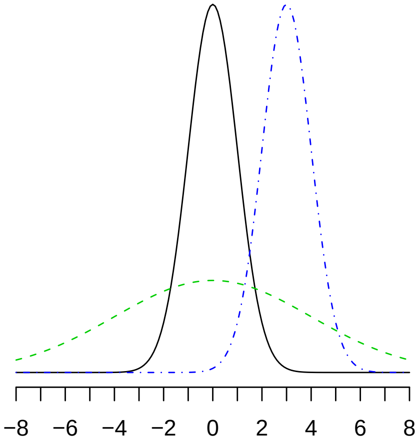
- We write $X \sim N(\mu, \sigma^2)$.
- The parameter μ determines where the curve is. The center of the curve is μ.
- The parameter σ determines how spread out the curve is. The area under the curve in the interval $(\mu - 2\sigma, \mu + 2\sigma)$ is 95%. $P(\mu - 2\,\sigma < X < \mu + 2\,\sigma) \approx 0.95$



- For the normal family of distributions we can see that the parameter μ talks about "where" the distribution is *located* or *centered*.
- We often use μ as our best guess for a *prediction*.
- The parameter σ talks about how *spread out* the distribution is. This gives us and indication about how *uncertain* or how *risky* our prediction is.
- $Z \sim N(0, 1)$.

Example:

- Below are the pdfs of
  $X_1 \sim N(0,1)$, $X_2 \sim N(3,1)$, and $X_3 \sim N(0,16)$.
- Which pdf goes with which X?



Mean and Variance of a Continuous Random Variable

Continuous random variables have expected values and variances analogous to what we have defined for discrete random variables.

But, the definition requires calculus and we don't want to have to remember all that stuff.

Fortunately, our intuition is the same!!!

- The expected value of a random variable is the probability weighted average value.
- The variance of a random variable is the probability weighted average squared distance to the expected value.

Mean and Variance for a Normal

For

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

$$E(X) = \mu, \quad Var(X) = \sigma^2$$

μ is the mean and σ^2 is the variance !!

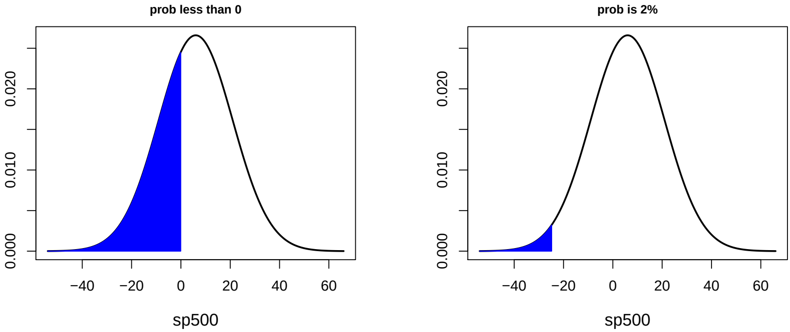
Note:

$$E(Z) = 0, \quad Var(Z) = 1.$$

The Normal Distribution – Example

- Assume the annual returns on the SP500 are normally distributed with mean 6% and standard deviation 15%. SP500 $\sim N(6, 225)$. (Notice: $15^2 = 225$).
- Two questions:
  (i) What is the chance of losing money on a given year?
  (ii) What is the value such that there's only a 2% chance that the return is less than the value?
- Lloyd Blankfein: *"I spend 98% of my time thinking about 2% probability events!"*
- (i) $P(SP500 < 0)$ and (ii) $P(SP500 <?) = 0.02$

The Normal Distribution – Example



$(i)\ P(SP500 < 0) = 0.35$ and $(ii)\ P(SP500 < -25) = 0.02$

In R:

For $X \sim N(\mu, \sigma^2)$, `pnorm(c,...)` gives $P(X < c)$, which is the CDF (cumulative distribution function) evaluated at c.

`qnorm(q,...)` gives the value c such that $P(X < c) = q$.

```
> 1-2*pnorm(-1.96,mean=0,sd=1)
[1] 0.9500042
> 1-2*pnorm(-1.00,mean=0,sd=1)
[1] 0.6826895

> pnorm(0,mean=6,sd=15)
[1] 0.3445783
> pnorm(-25,mean=6,sd=15)
[1] 0.01938279
>
> qnorm(.02,mean=6,sd=15)
[1] -24.80623
```

In Excel see: NORMDIST and NORMINV

The Probability of an Interval:

What is $P(0 < SP500 < 20)$?

That is, what is the probability that the return value ends up being in the interval $(0, 20)$?

```
> pnorm(0,6,15)
[1] 0.3445783
> pnorm(20,6,15)
[1] 0.8246761
> 0.8246761 - 0.3445783
[1] 0.4800978
```

The probability of the interval (a, b) is $CDF(b) - CDF(a)$.

In general

If X is a random variable, then the cumulative distribution function or *CDF* is

$$F(x) = P(X \leq x)$$

If X is a continuous random variable then

$$P(a \leq X \leq b) = F(b) - F(a), \ \ b > a.$$

Standardization

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

μ is the mean and σ^2 is the variance.

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

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

$\frac{X-\mu}{\sigma}$ should look like a Z !!

The number of standard deviations, X is away from the mean.

Example:

Prior to the 1987 crash, monthly S&P500 returns (R) followed (approximately) a normal with mean 0.012 and standard deviation equal to 0.043. How extreme was the crash of -0.2176? The standardization helps us interpret these numbers...

$$R \sim N(0.012, 0.043^2)$$

The month of the crash, R turned out to be $r = -0.2176$. Correspondingly, for the crash,

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

which is pretty wild for a standard normal!!

5 standard deviations away from the mean!!

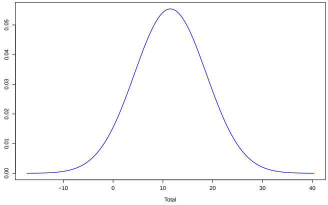
The Normal Distribution – Approximating Combinations of RVs

Recall the Thaler example (Portfolios of projects vs. single project).

A linear combination of independent random variables is approximately normal (the CLT: Central Limit Theorm), so

$$T \sim N(11.5, 7.2^2) \quad \text{approximately}$$

```
> .5 * 23
[1] 11.5
> 2.25*23
[1] 51.75
> sqrt(51.75)
[1] 7.193747
> 1 - pnorm(0,11.5,7.2)
[1] 0.9448919
```



much more compelling than the simple Sharpe ratio we looked at before !!

In summary, in many situations, if you can figure out the mean and variance of the random variable of interest.

If the random variable is a combination of many variables (often the case) it may be approximately normal.

If you know it is (approximately) normal and you know the mean and variance, you know the distribution (approximately).

Portfolios, once again...

- As before, let's assume that the annual returns on the SP500 are normally distributed with mean 6% and standard deviation of 15%, i.e., $SP500 \sim N(6, 15^2)$
- Let's also assume that annual returns on bonds are normally distributed with mean 2% and standard deviation 5%, i.e., $Bonds \sim N(2, 5^2)$
- What is the best investment?
- What else do I need to know if I want to consider a portfolio of SP500 and bonds?

- Additionally, let's assume the correlation between the returns on SP500 and the returns on bonds is -0.2.
- How does this information impact our evaluation of the best available investment?

Recall that for two random variables X and Y:

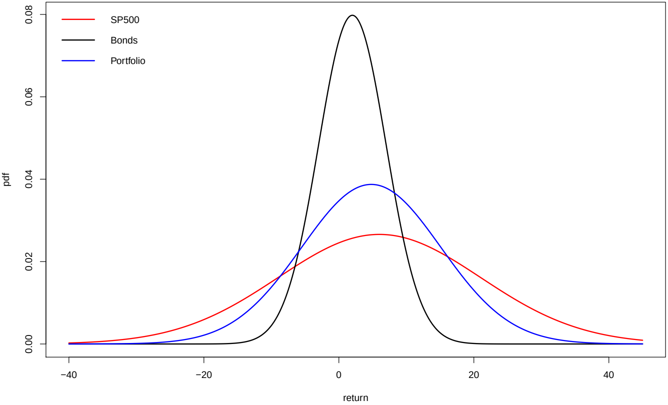
- $E(aX + bY) = aE(X) + bE(Y)$
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab \times Cov(X, Y)$

- What is the behavior of the returns of a portfolio with 70% in the SP500 and 30% in Bonds?
- $E(0.7SP500 + 0.3Bonds) = 0.7E(SP500) + 0.3E(Bonds) = 0.7 \times 6 + 0.3 \times 2 = 4.8$
- $Var(0.7SP500 + 0.3Bonds) = (0.7)^2 Var(SP500) + (0.3)^2 Var(Bonds) + 2(0.7)(0.3) \times Corr(SP500, Bonds) \times sd(SP500) \times sd(Bonds) = (0.7)^2(15^2) + (0.3)^2(5^2) + 2(0.7)(0.3) \times -0.2 \times 15 \times 5 = 106.2$

Let's assume the linear combination is normal, which is often the case when both of the input random variables are normal.

$$Portfolio \sim N(4.8, 10.3^2)$$

Here are the normal pdfs for our three assets, SP500, Bonds, and the portfolio.



The Uniform Distribution

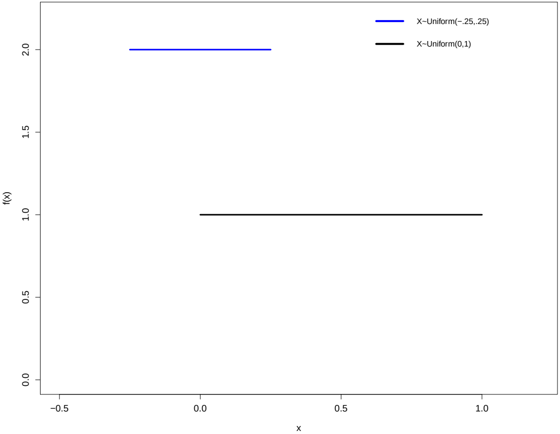
Suppose we think a random variable X can turn out to be any number between -.25 and .25 and the numbers in (-.25,25) are equally likely?

How would we describe this??

Suppose we think a random variable X can turn out to be any number between 0 and 1 and the numbers are equally likely?

How would we describe this?

If X can be any number in (a, b) and the numbers are equally likely, then we say $X \sim Uniform(a, b)$



The density is $\frac{1}{b-a}$ inside (a, b) and 0 elswhere.

10. A Little Calculus

In the discrete case we have explicit formulas for things like the expected value of a random variable.

In the continuous case we need calculus, in particular, the integral.

For the record we record the formulas for the continuous case.

However, it is just fine to get your intuition from the discrete case and not worry about the integrals.

Discrete

$$P(X \in A) = \sum_{x \in A} p(x)$$

p is the probability mass function

Continuous

$$P(X \in A) = \int_A f(x)\,dx$$

f is the probability density function

Discrete

$$E[x] = \sum_x x\,p(x)$$

In general,

$$E(f(x)) = \sum_x f(x)\,p(x)$$

Continuous

$$E[x] = \int x f(x)\,dx$$

$$E\{g(x)\} = \int g(x)\,f(x)\,dx$$

$E\{x\} = \mu_x \qquad E\{y\} = \mu_y$

Variance

$$\mathrm{Var}(X) = \sigma_x^2 = \int (x-\mu_x)^2 f(x)\, dx$$

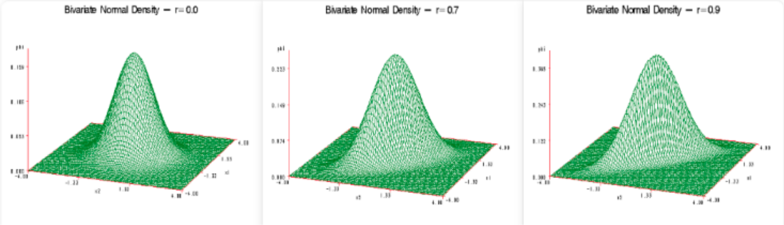
Covariance

$$\mathrm{cov}(X,Y) = \sigma_{xy} = E\{(X-\mu_x)(Y-\mu_y)\}$$

$$= \iint (x-\mu_x)(y-\mu_y)\, f(x,y)\, dx\, dy$$

With a bivariate (2 random variables) (X, Y) continuous distribution we have the density $f(x, y)$.

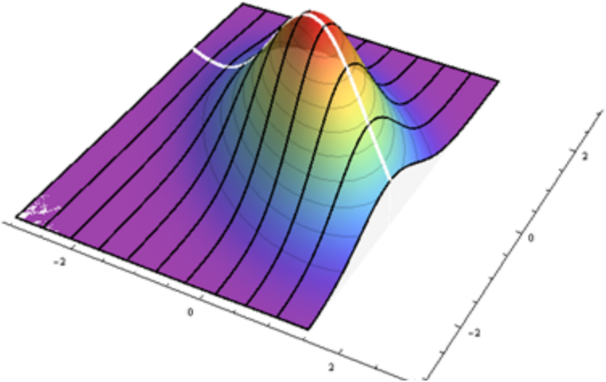
We can plot the density in 3D.



This gives us a sense of what *correlation* means with continuous random variables.

The conditional density is proportional to the joint $f(y|x) \propto f(x, y)$

$$f(y|x) \propto f(x, y)$$



11. Making Decisions under Uncertainty

In our drug discovery example we traced out the key things that could happen if we made the investment and assigned probabilities to the branches of the path.

This gave us a random variable representing the uncertain revenue we would get if we made the investment.

If we just go by the expected value, then our 1 million dollar investment gives us a random variable with expected revenue \$4,225,000 so it seems like the investment is a good idea.

However, we considered another scenario where the random variable representing the revenue had the same expected value even though the distributions were quite different.

We saw the standard deviations under the two scenarios were 9,134,721 and 2,836,141 which dramatically highlights the difference between the two distributions, even though they have the same mean.

In general, how do we make a decision when outcomes are uncertain !!!

If microeconomics, we often think of people as making decisions by maximizing utility.

We often assume that all the parameters of our decision (prices, budget....) are known.

This is often very unrealistic !!

Maximize Expected Utility

A general formulation for decision making under uncertainty is to again start with a utility function

$$U(Y, a)$$

representing your utility under the *uncertain outcome* represented by the random variable Y and the action $a \in A$, where A represents a (possibly contrained) set of actions.

We then maximize expected utility over our possible actions:

$$\max_{a \in A} E(U(Y, a))$$

This is a fundamental idea in both economics and statistics.

Minimize Expected Loss

In statistics/Machine Learning/Data Science/Predictive Analytics is of common to the think in terms of the *loss* rather than the utility.

We minimize our expected loss:

$$\min_{a \in A} E(L(Y, a))$$

Minimizing Expected Squared Loss

Often people will want you got come with a single number as your *prediction* for what Y will turn out to be.

Let a represent your prediction.

For a numeric Y, a commonly used loss function is squared error loss, giving,

$$\min_a E((Y-a)^2)$$

The optimal choice of a is $a^* = E(Y)$.

The expected value is the optimal prediction under squared error loss.