Section 1.2: Probability and Decisions

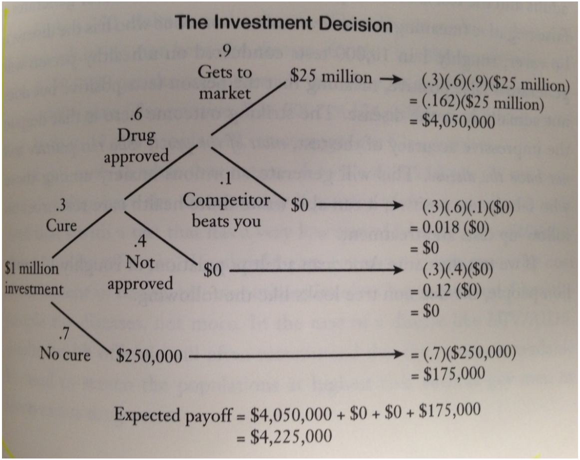
Jared S. Murray
The University of Texas at Austin
McCombs School of Business
OpenIntro Statistics, Chapters 2.4.1-3.
Decision Tree Primer Ch. 1 & 3 (on Canvas under Pages)

Probability and Decisions

- So you've tested positive for a disease. Now what?
- Let's say there's a treatment available. Do you take it?
- What additional information (if any) do you need?
- We need to understand the **probability distribution** of outcomes to assess **(expected) returns** and **risk**

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.



Probability and Decisions

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

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

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

So, should we invest or not?

Back to Targeted Marketing

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)

Back to Targeted Marketing

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

Revenue	$P(Revenue)$
$-0.8	0.95
$39.20	0.05

Should you do the promotion?

Probability and Decisions

Let's get back to the drug investment example...

What if you could choose this investment instead?

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

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

What is the difference?

Mean and Variance of a Random Variable

The Mean or Expected Value is defined as (for a discrete X with n possible outcomes):

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

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!

Example: Mean and Variance of a Binary Random Variable

Suppose

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

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

Another example: What is the $E(Revenue)$ for the targeted marketing problem?

Mean and Variance of a Random Variable

The Variance is defined as (for a discrete X with n possible outcomes):

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

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

Example: Mean and Variance of a Binary Random Variable

Suppose

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

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

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

What is the $Var(Revenue)$ in our example above?

How about the drug problem?

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, Variance, Standard Deviation: Summary

What to keep in mind about the mean, variance, and SD:

- The expected value/mean is usually our **best prediction** of an uncertain outcome. ("Best" meaning closest in distance to the realized outcome, for a particular measure of distance)
- The variance is often a reasonable summary of **how unpredictable** an uncertain outcome is (or how risky it is to predict)
- The standard deviation (square root of the variance) is another reasonable summary of risk that is on a meaningful scale.

Why expected values?

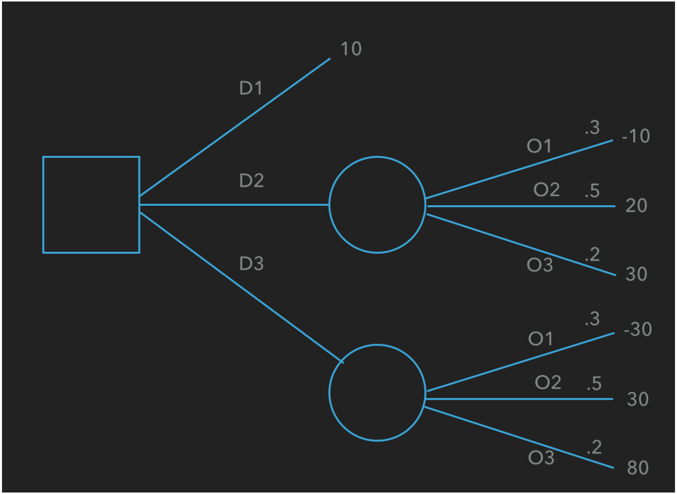
- When you have a repeated decision problem (or many decisions to make), make decisions to maximize your **expected utility**
- Utility functions provide a numeric value to outcomes; those with higher utilities are preferred
- Profit/payoff is one utility function. More realistic utilities allow for risk taking/aversion, but the concepts are the same.

Decision Trees

A convenient way to represent decision problems:

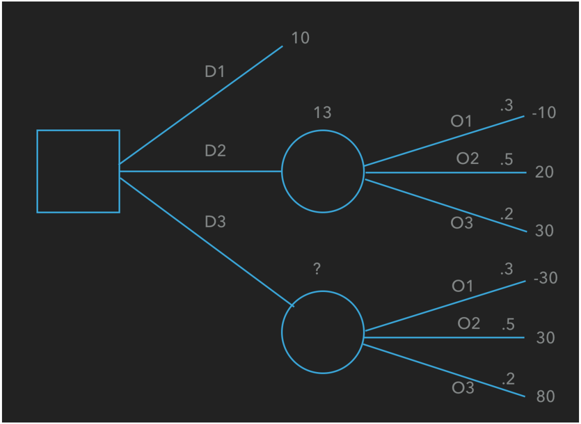
- Time proceeds from left to right.
- Branches leading out of a decision node (usually a square) represent the possible decisions.
- Probabilities are listed on probability branches, and are conditional on the events that have already been observed (i.e., they assume that everything to the left has already happened).
- Monetary values (utilities) are shown to the right of the end nodes.
- EVs are calculated through a "rolling-back" process.

Example



Rolling back: Step 1

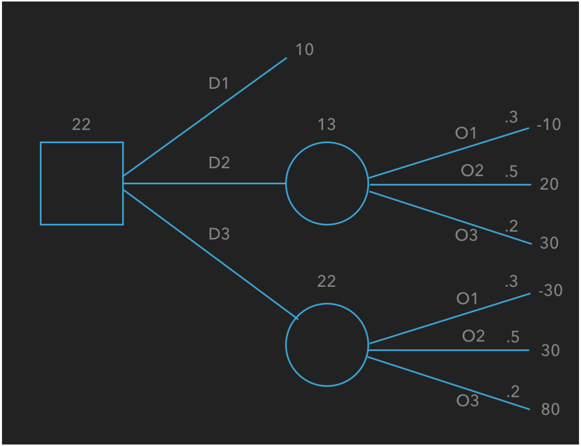
Calculate the expected value at each probability node:



$$E(Payoff \mid D2) = .3(-10) + .5(20) + .2(30) = 13$$

Rolling back: Step 2

Calculate the maximum at each decision node:



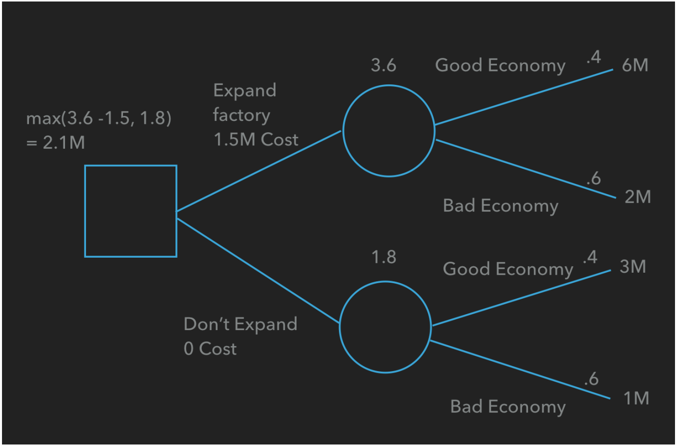
Take decision D3 since $22 = \max(10, 13, 22)$.

Sally Ann Soles' Shoe Factory

Sally Ann Soles manages a successful shoe factory. She is wondering whether to expand her factory this year.

- The cost of the expansion is $1.5M.
- If she does nothing and the economy stays good, she expects to earn $3M in revenue, but if the economy is bad, she expects only $1M.
- If she expands the factory, she expects to earn $6M if the economy is good and $2M if it is bad.
- She estimates that there is a 40 percent chance of a good economy and a 60 percent chance of a bad economy.

Should she expand?



$E(\text{expand}) = (.4(6) + .6(2)) - 1.5 = 2.1$

$E(\text{don't expand}) = (.4(3) + .6(1)) = 1.8$

Since $2.1 > 1.8$, she should expand, right? (Why might she choose not to expand?)

Sequential decisions

She later learns after she finishes the expansion, she can assess the state of the economy and opt to either:

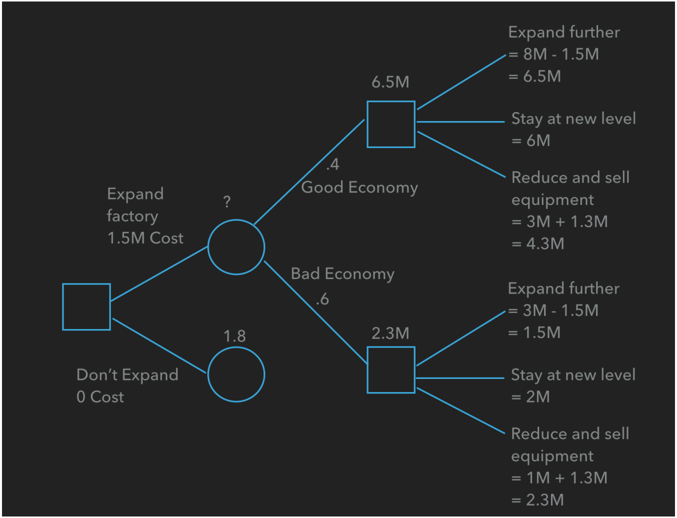
(a) expand the factory further, which costs \$1.5M and will yield an extra \$2M in profit if the economy is good, but \$1M if it is bad,

(b) abandon the project and sell the equipment she originally bought, for \$1.3M – obtaining \$1.3M, plus the payoff if she had never expanded, or

(c) do nothing.

How has the decision changed?

Sequential decisions



Expected value of the option

The EV of expanding is now

$$(.4(6.5) + .6(2.3)) - 1.5 = 2.48.$$

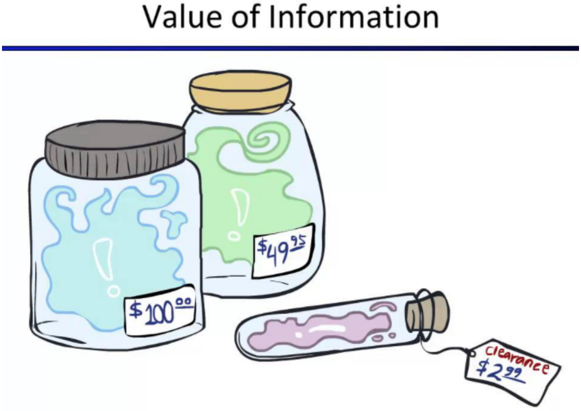
If the option were free, is there any reason not to expand?

What would you pay for the option? How about

$$E(\text{new}) - E(\text{old}) = 2.48 - 2.1 = 0.38,$$

or $380,000?

What Is It Worth to Know More About an Uncertain Event?



Value of information

- Sometimes information can lead to better decisions.
- How much is information worth, and if it costs a given amount, should you purchase it?
- The expected value of perfect information, or EVPI, is the most you would be willing to pay for perfect information.

Typical setup

- In a multistage decision problem, often the first-stage decision is whether to purchase information that will help make a better second stage decision
- In this case the information, if obtained, may change the probabilities of later outcomes
- In addition, you typically want to learn how much the information is worth
- Information usually comes at a price. You want to know whether the information is worth its price
- This leads to an investigation of the value of information

Example: Marketing Strategy for *Bevo: The Movie*

UT Productions has to decide on a marketing strategy for it's new movie, Bevo. Three major strategies are being considered:

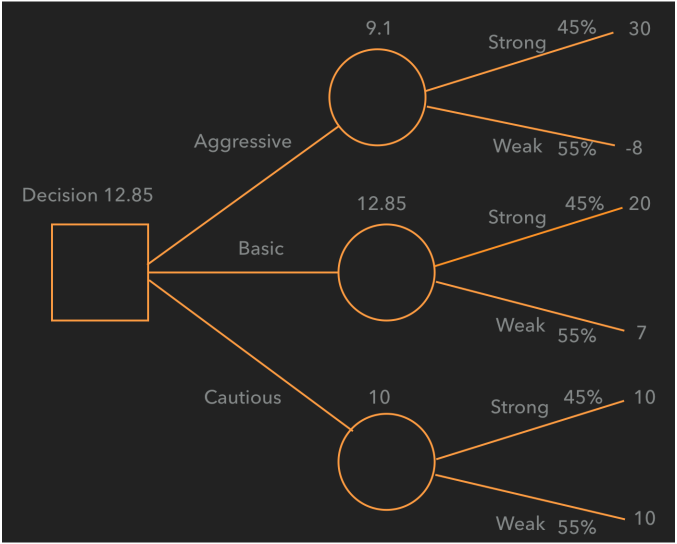
- (A) Aggressive: Large expenditures on television and print advertising.
- (B) Basic: More modest marketing campaign.
- (C) Cautious: Minimal marketing campaign.

Payoffs for *Bevo: The Movie*

The net payoffs depend on the market reaction to the film.

	Market Reaction	
Decisions	Strong	Weak
Aggressive	30	-8
Basic	20	7
Cautious	10	10
Probability	0.45	0.55

Decision Tree for *Bevo: The Movie*



Expected Value of Perfect Information (EVPI)

How valuable would it be to know what was going to happen?

- If a clairvoyant were available to tell you what is going to happen, how much would you pay her?
- Assume that you don't know what the clairvoyant will say and you have to pay her before she reveals the answer

EVPI = (EV with perfect information) - (EV with no information)

Finding EVPI with a payoff table

The payoffs depend on the market reaction to the film:

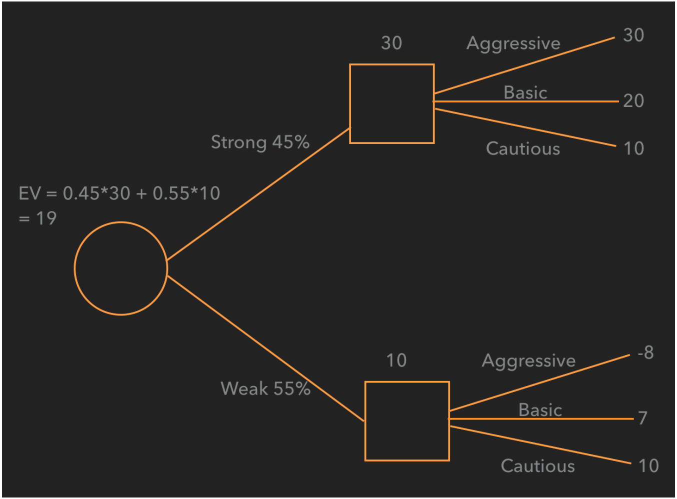
	Market Reaction	
Decisions	Strong	Weak
Aggressive	30	-8
Basic	20	7
Cautious	10	10
Probability	0.45	0.55

- With no information, the Basic strategy is best: EV $= 0.45(20) + 0.55(7) = 12.85$
- With perfect info, select the Agressive strategy for a Strong reaction and the Cautious strategy for a Weak reaction: EV $= 0.45(30) + 0.55(10) = 19$
- EVPI $= 19 - 12.85 = 6.15$

Finding EVPI with a decision tree

- Step 1: Set up tree without perfect information and calculate EV by rolling back
- Step 2: Rearrange the tree the reflect the receipt of the information and calculate the new EV
- Step 3: Compare the EV's with and without the information

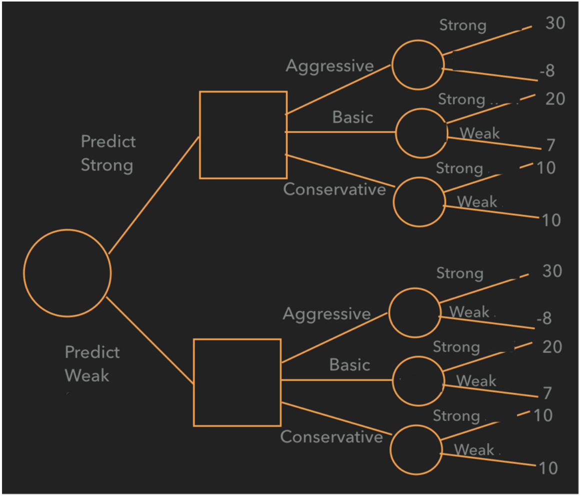
Finding EVPI with a decision tree



What about imperfect information?

Suppose that Myra the movie critic has a good record of picking winners, but she isn't clairvoyant. What is her information worth?

The decision tree with imperfect information



How does this compare with the perfect information tree?

We need to get the relevant conditional probabilities...

How good is the information?

Suppose that Myra the movie critic has a good record of picking winners.

- For movies where the audience reaction was strong, Myra has historically predicted that 70% of them would be strong.
- For movies where the audience reaction was weak, Myra has historically predicted that 80% of them would be weak.

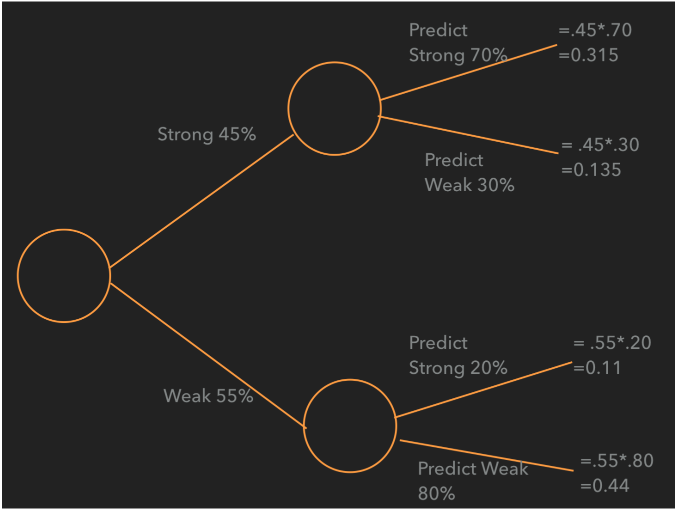
Remember that the chance of a strong reaction is 45% and of a weak reaction is 55%.

Suppose S and W means the audience reaction was strong or weak, respectively, and PS and PW means that Myra's prediction was strong or weak, respectively. Let's translate what we know:

- For movies where the audience reaction was strong, Myra has historically predicted that 70% of them would be strong.
  $P(PS|S) = .7, \qquad P(PW|S) = .3$
- For movies where the audience reaction was weak, Myra has historically predicted that 80% of them would be weak.
  $P(PW|W) = .8, \qquad P(PS|W) = .2$
- The chance of a strong reaction is 45% and of a weak reaction is 55%.
  $P(S) = .45, \qquad P(W) = .55$

Bayes rule to the rescue!

We have the wrong margin/conditionals, but we can get the correct ones. First compute the joint probabilties:



What distributions do we need?

The sequence is (Myra predicts) $\rightarrow$ (We decide)

First uncertain outcome in the new tree is Myra's prediction, so we need $P(PS)$ and $P(PW) = 1 - P(PS)$:

$$P(PS) = P(PS \mid S)P(S) + P(PS \mid W)P(W) = (0.315{+}0.11) = 0.425$$

What conditionals do we need?

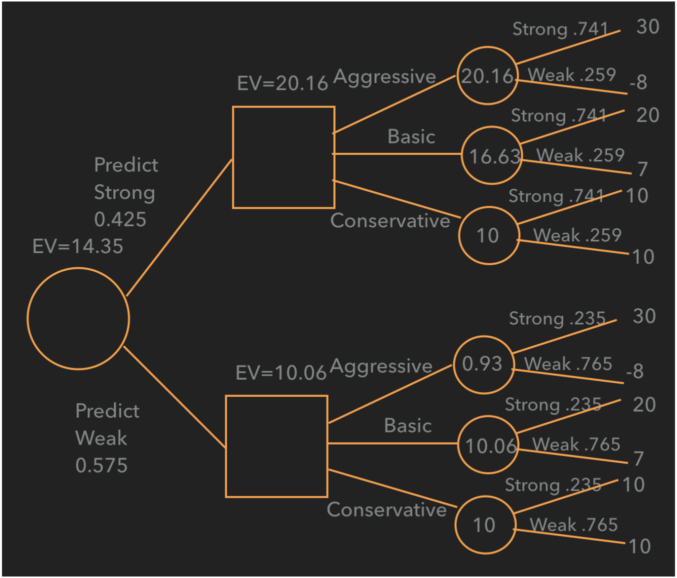
The sequence is (Myra predicts) $\rightarrow$ (We decide)

Next uncertain outcome is the true market response, so we need $P(S \mid PS)$ and $P(W \mid PW)$:

$$P(S \mid PS) = \frac{P(PS \mid S)P(S)}{P(PS)} = 0.315/0.425 = 0.741$$

$$P(W \mid PW) = \frac{P(PW \mid W)P(W)}{P(PW)} = 0.44/(1 - 0.425) = 0.765$$

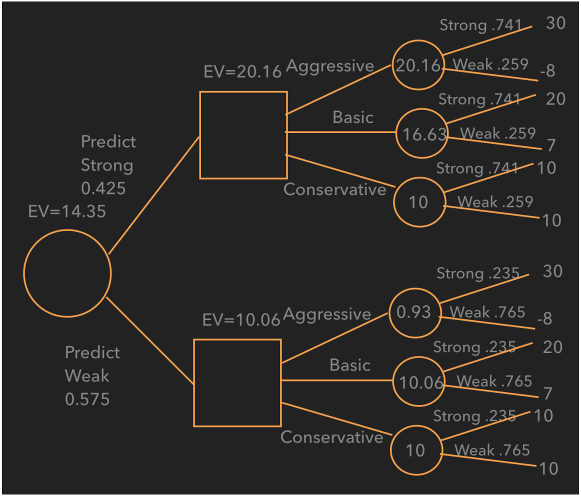
Tree with imperfect information



Myra's information is worth paying for

It changes the decision and adds $14.35 - 12.85 = 1.5$ in value.

(Compare this to the 6.15 the clairvoyant's prediction was worth.)



Things to remember about the value of information

- Perfect information is more valuable that any imperfect information
- Information cannot have negative value

Decision trees: Summary

- Useful framework for simplifying some probability & expectation calculations.
- "Under the hood" they are simply applications of conditional probability and expectation!
- Specialized software exists for complicated trees (e.g. Pallisade PrecisionTree in Excel or the free Radiant R package) but the concepts are the same.

Combining random variables

We've seen how the **expected value** (our best prediction) and **variance/standard deviation** (how risky our best prediction is) help us think about uncertainty and make decisions in simple scenarios

We need some more tools for thinking about multiple random variables (sources of uncertainty)

Covariance

- A measure of *dependence* between two random variables...
- It tells us how two unknown quantities tend to move together: Positive $\rightarrow$ One goes up (down), the other tends to go up (down). Negative $\rightarrow$ One goes down (up), the other tends to go up (down).
- If X and Y are independent, $Cov(X, Y) = 0$ **BUT** $Cov(X, Y) = 0$ does not mean X and Y are independent (more on this later).

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

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

Ford vs. Tesla

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

	t=-7%	t=0%	t=7%	Pr(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**
Pr(T=t)	**0.09**	**0.80**	**0.11**	**1**

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

Example: Ford vs. Tesla

	t=-7%	t=0%	t=7%	Pr(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**
Pr(T=t)	**0.09**	**0.80**	**0.11**	**1**

- $E(F) = 0.12$, $E(T) = 0.14$
- $Var(F) = 5.25$, $sd(F) = 2.29$, $Var(T) = 9.76$, $sd(T) = 3.12$
- What is the better stock?

Example: Ford vs. Tesla

	t=-7%	t=0%	t=7%	Pr(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**
Pr(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 Ford vs. Tesla example:

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

Linear Combination of Random Variables

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

To answer this question we need to understand the behavior of the weighted sum (linear combinations) of two random variables...

Let X and Y be two random variables:

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

Linear Combination of Random Variables

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

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

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

Risk Adjustment: Sharpe Ratio

The Sharpe ratio is a unitless quantity used to compare investments:

$$\frac{\text{(average return) - (return on a risk-free investment)}}{\text{standard deviation of returns}}$$

Idea: Standardize the average excess return by the amount of risk. ("Risk adjusted returns")

Ignoring the risk-free investment, what are the Sharpe ratios for Ford, Tesla, and the 50-50 portfolio?

Linear Combination of Random Variables

More generally...

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

- $Var(w_1X_1 + w_2X_2 + ...w_pX_p + c) = w_1^2Var(X_1) + w_2^2Var(X_2) +$
  $... + w_p^2Var(X_p) + 2w_1w_2 \times Cov(X_1, X_2) + 2w_1w_3Cov(X_1, X_3) +$
  $... = \sum_{i=1}^{p} w_i^2Var(X_i) + \sum_{i=1}^{p}\sum_{j\neq i} w_iw_jCov(X_i, X_j)$

where $w_1, w_2, \ldots, w_p$ and c are constants