AP Statistics – Lesson Notes - Chapter 14,15: Probability

Probability = relative frequency "in the long run"

Consider this situation: On your drive to work you encounter one signal light. When you arrive at the signal, sometimes it is red, sometimes it is green (or yellow). Because you don't leave at a specific time each day, the exact time you arrive at the signal is random.

But there is an underlying structure in the situation: the signal light is red for 2 minutes out of every 10 minutes and green/yellow for the rest of the time. For 20% of the arrival times, the signal will be red.

On any particular arrival time, anything can occur. But if you keep track of the signal at your arrival time over 100 days, this underlying structure will emerge: about 20% of the outcomes will be red and the rest green/yellow.

The **Law of Large Numbers (LLN)** says that the observed relative frequency of red lights will approach the **probability** that the light is red (20%).

Probability terms:

Outcome: One possible result that can occur in an experiment.

Examples of outcomes: A single roll of a die: 5

Tossing a coin twice: TH

Sample space: Set of all outcomes that can occur as the result of an experiment.

Examples of sample spaces: A single roll of a die: $S = \{1, 2, 3, 4, 5, 6\}$

Tossing a coin twice: $S = \{HH, HT, TH, TT\}$

Event: Any subset of the sample space.

Examples of events: Rolling an even number: $E = \{2, 4, 6\}$

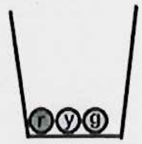
Obtaining at least one tail: $E = \{HT, TH, TT\}$

Probability of an Event: A numerical value between 0 and 1 representing the likelihood of the event occurring. Denoted $P(E)$.

a) $0 \le P(E) \le 1$
b) The sum of all the probabilities of the outcomes in $S = 1$.

Probability of a single event, its complement, and sum of all outcomes

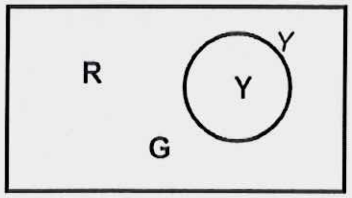
Drawing one ball from a jar...



$S = \{R, Y, G\}$

$Probabilities: \frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}$

Venn diagram:



Probability of the sample space:

$P(S) = \frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$

Probability of an event:

$P(Y) = \frac{1}{3}$

$P(Y) = \frac{\#\,desired\ outcomes}{\#\,all\ \ possible\ outcomes}$

Probability of the complement of an event:

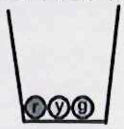
$P(Y^C) = P(\overline{Y}) = P(Y') = 1 - P(Y)$

$P(Y^C) = 1 - \frac{1}{3} = \frac{2}{3}$

Probability of a single event with equally likely outcomes

Drawing one ball from a jar...

S= {R RRR YYY G}

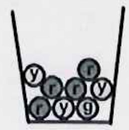


$S = \{R, Y, G\}$

$Probabilities: \frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}$

$P(Y) = \frac{1}{3}$

$P(Y) = \frac{\#\,desired\ outcomes}{\#\,all\ possible\ outcomes}$



$S = \{R, Y, G\}$

$Probabilities: \frac{4}{8}\ \frac{3}{8}\ \frac{1}{8}$

$P(Y) = \frac{3}{8}$

$P(Y) \neq \frac{1}{3} = \frac{\#\,desired\ outcomes}{\#\,all\ possible\ outcomes}$

Only true for **equally likely outcomes**

Conditional Probability

If we asked 60 girls and 40 boys at a local high school what their favorite entertainment activity was, the results might be something like this:

	girls	*boys*
read a book	18	4
video games	12	20
watch Netflix	30	16

Whenever we have a table of data, we should always add up the rows and columns to get the *marginal* values:

	girls	*boys*	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

Now we could ask various questions about the probability of certain events. Each probability would be the percentage which is the ratio of two numbers from this table.

We could ask: If I select a student at random, what is the probability of selecting a girl? Each student is equally likely to be selected, so this is an 'equally-likely cases' probability, and we use:

$$P(E) = \frac{number\ of\ outcomes\ in\ event\ E}{number\ of\ outcomes\ in\ the\ sample\ space\ S}$$

$$= \frac{number\ of\ 'desired'\ outcomes}{total\ number\ of\ outcomes}$$

If I select a student at random, what is the probability of selecting a girl?

Here, the total number of outcomes would be all the students, so the entire table. There are 100 students total.

The number which are girls is just the left column. There are 60 girls.

	girls	*boys*	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$P(girl) = \frac{60}{100} = .60$$

When we are selecting out of the entire sample space, this is referred to as a **Simple Probability**.

Or we could ask: If I select a student at random, what is the probability of selecting a student who likes video games?

Here, the total number of outcomes is still all the students, so the entire table. There are 100 students total.

The number which like video games is just the middle row of students. There are 32 students who like video games.

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$P(video\ games) = \frac{32}{100} = .32$$

But what if we ask this: What is the probability that a girl prefers video games?

Now the total number of outcomes isn't all the students, it is only the left column of girls. There are a total of 60 girls in this sample.

Staying within just the girls, the number of girls who prefer video games is 12.

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$= \frac{12}{60} = .20$$

This probability is 12 out of the 60 girls, which is 20%. But this situation is more complicated...it isn't what we call a *simple probability*. It is the probability of playing video games with the *condition* that we only look at the girls. This is known as a ***conditional probability.***

To indicate a conditional probability, in the probability statement we have to indicate both the condition (the denominator) and the event (the numerator). Here is how this is written:

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$P(video\ games \mid girl) = \frac{12}{60} = .20$$

event **condition**

We're still finding the probability of someone preferring video games, so that is listed first. But after a vertical line, we indicate any *condition.*

Some important things to notice...

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$P(video\ games \mid girl) = \frac{12}{60} = .20$$

event
The event is always contained within the conditional sample space.

condition
The condition is always just a portion of the sample space. This is called the *conditional sample space*.

The conditional sample space is a portion of the sample space.
The event is a portion of the conditional sample space.

The event (what comes before the line) goes in the numerator of the fraction.

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$P(video\ games \mid girl) = \frac{12}{60} = .20$$

The condition (what comes after the line) goes in the denominator of the fraction.

How about this one: What is the probability that a student who prefers watching Netflix will be a boy?

The most important (and sometimes trickiest) thing is to decide what is the event and what is the condition?

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

Let's dissect the question sentence:
What is the probability that a student who prefers watching Netflix will be a boy?

If we put the middle part in parentheses, the rest of the sentence is still asking about a probability:

What is the probability that a student (~~who prefers watching Netflix~~) will be a boy?

This question is really asking about the probability of being a boy. The stuff in the parentheses is the 'condition'. So the correct probability statement for this question is:

$$P(boy\,|\,watch\ Netflix)$$

Now we can start evaluating. First, let's identify the conditional sample space and get the number for the denominator. This would be all the students who watch Netflix, which is the row of 46.

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

Then, within the conditional sample space (only this row), how many students are boys? The numerator should be 16.

$$P(boy\,|\,watch\ Netflix) = \frac{16}{46} = .3478$$

One more from this scenario: What is the probability that a student is a girl if they prefer either playing video games or watching Netflix?

$$P(girl\,|\,video\ games\ or\ watch\ Netflix)$$

The conditional sample space would be all the students in the video games or Netflix cells, which is a total of 78 students.

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

The event would be only the girls within this conditional sample space, which is 42.

$$P(girl\,|\,video\ games\ or\ watch\ Netflix) = \frac{42}{78} = .5385$$

Note: it matters which part is the event and which is the condition:

$$P(boy\,|\,video\ games) = \frac{20}{32} = .6250$$

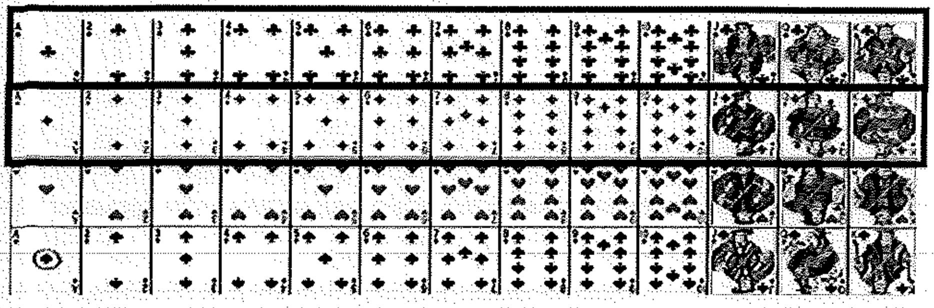
	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$P(video\ games\,|\,boy) = \frac{20}{40} = .5000$$

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

Compound Event Probability: one event OR another event happening

If we draw one card from a standard deck of 52 cards, what is the probability the card is a club or a diamond?



Each card is equally-likely to be drawn, so we could just list all the cards the event 'club or diamond' and count them, then divide by the total number of cards.

$$P(club \cup diamond) = \frac{26}{52} = \boxed{\frac{1}{2}}$$

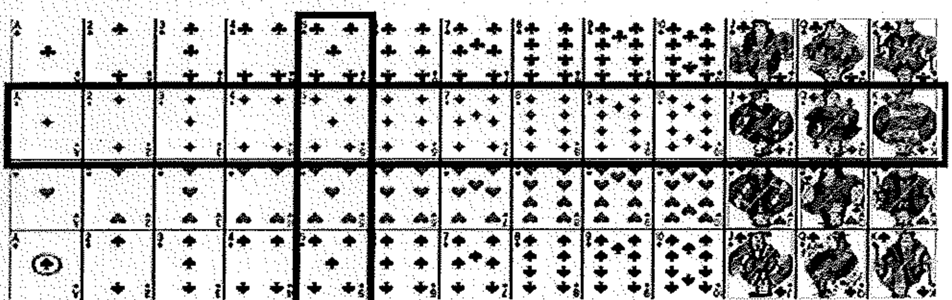
We could also add the separate probabilities for 'club' and 'diamond'

$$P(club) = \frac{13}{52} = \frac{1}{4}, \quad P(diamond) = \frac{13}{52} = \frac{1}{4}$$

OR = add (but there's a catch...)

$$P(club \cup diamond) = P(club) + P(diamond) = \frac{1}{4} + \frac{1}{4} = \boxed{\frac{1}{2}}$$

If we draw one card from a standard deck of 52 cards, what is the probability the card is a diamond or a 5?



Let's first solve by just counting up the individual cards in this event:

$$P(diamond \cup 5) = \boxed{\frac{16}{52}}$$

Now let's try adding the separate probabilities for 'diamond' and '5'

$$P(diamond) = \frac{13}{52}, \quad P(5) = \frac{4}{52}$$

$$P(diamond \cup 5) = P(diamond) + P(5) = \frac{13}{52} + \frac{4}{52} = \boxed{\frac{17}{52}}$$

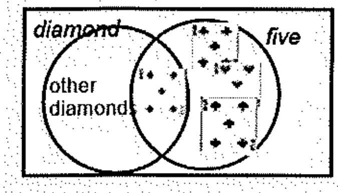
This is incorrect. We know the correct answer is 16/52.

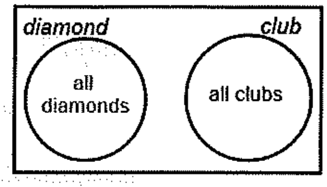
The reason adding produced a value too high is because the 5 of diamonds got included twice. **There is 'overlap' between the two events.** But this is easy to fix: we'll just subtract the overlap:

$$P(diamond \cup 5) = P(diamond) + P(5) - P(diamond \cap 5)$$

$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \boxed{\frac{16}{52}}$$

This wasn't a problem in our first example with diamonds and clubs. Compare its Venn diagram to the diamonds and five example:





We didn't have a problem just adding probabilities with the events 'diamonds' and 'clubs', because there was no 'overlap'...nothing to get counted twice.

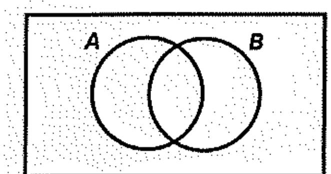
Some people memorize that for probabilities 'OR equals addition'. That idea is correct, but for it to work in all cases, it needs to be 'OR equals addition, but subtract off any overlap'.

Probability of 'OR' case compound events:

$$P(A\ OR\ B) = P(A) + P(B) - P(A\ AND\ B)$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

This is the true 'OR' formula which works in all cases.

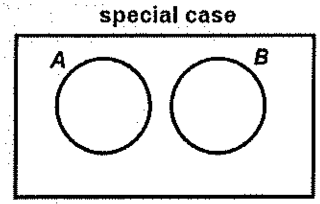
To help us visualize this, here are Venn diagrams for these two situations:



A and B are non mutually-exclusive
A and B are not disjoint events
A and B are joint events

$$P(A \cap B) \neq 0$$
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

special case



A and B are mutually-exclusive
A and B are disjoint events

$$P(A \cap B) = 0$$
$$P(A \cup B) = P(A) + P(B)$$

Test for disjoint (mutually-exclusive) events:

We can also define a 'test for disjoint events':

Two events are disjoint if $P(A \cap B) = 0$ (there is no overlap)

Compount Event Probability: one event AND another event happening

Let's look again at the student survey about entertainment choices:

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

The probability that the chosen student is a girl would be a simple probability because we are choosing from the entire sample space of all the students:

$$P(girl) = \frac{60}{100} = .6$$

Similarly, the probability that the chosen student likes to read a book would be:

$$P(book) = \frac{22}{100} = .22$$

To find the probability that the chosen student is a girl AND likes to read a book we are still choosing a student out of all 100 students, so this is still a simple (not conditional) probability:

$$P(girl\ AND\ book) = \frac{18}{100} = .18$$

The other way to write 'AND' is to use the intersection symbol $\cap$. Note that this is because the 18 students who are both girls AND like reading a book are at the intersection of the read a book row and girls column:

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

$$P(girl \cap book) = \frac{18}{100} = .18$$

Can we compute $P(girl \cap book)$ from $P(girl)$ and $P(book)$?

This number of students for 'girls' AND 'book', 18, is smaller than the number of girls (60) and the number who like to read a book (22) because, to be in this group, the student must meet *both* requirements so it makes sense that fewer students would meet this more stringent requirement.

For the OR case, we ended up 'adding' probabilities. That was because it is *easier* for success to occur if there are two ways to be successful, so the probabilities add to become larger than the individual event probabilities.

For the AND case, it is *harder* for success to occur, so the AND probability should be *lower* than the individual event probabilities. When fractions are multiplied, the result is a smaller fraction, which suggests we might try ***multiplying probabilities for AND cases.***

We already know what the AND probability should be from calculating it as a simple probability before:

$$P(girl \cap book) = \frac{18}{100} = .18$$

Let's try multiplying the individual event probabilities:

$$P(girl \cap book) \stackrel{?}{=} P(girl) \cdot P(book)$$

$$.18 \stackrel{?}{=} (.6) \cdot (.22)$$

$$.18 \neq .132$$

This doesn't work, so there must be something else going on (there is a 'catch')

To investigate, let's look at just the girls column:

	girls
read a book	18
video games	12
watch Netflix	30
	60

	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

Since we are choosing the girls column, we should use $P(girl)$ as the first probability:

The probabilities should be...

$$P(girl) = \frac{60}{100} = .6$$

$$P(book) = \frac{22}{100} = .22$$

$$P(girl \cap book) = \frac{18}{100} = .18$$

$$P(girl \cap book) = P(girl) \cdot \_\_\_\_$$

$$P(girl \cap book) = \frac{60}{100} \cdot \_\_\_\_$$

	girls
read a book	18
video games	12
watch Netflix	30
	60

$$P(book \mid girl) = \frac{18}{60}$$

But the 2nd probability can't be $P(book)$ because that includes some boys, and we are only considering girls.

Instead, it has to be the probability that the student likes books but from within only the girls. That would be the <u>conditional probability</u> $P(book \mid girl)$

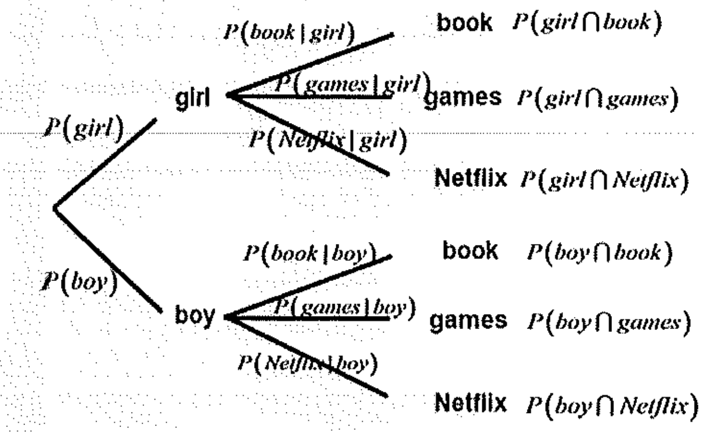
$$P(girl \cap book) = P(girl) \cdot P(book \mid girl)$$

$$P(girl \cap book) = \frac{60}{100} \cdot \frac{18}{60} = \frac{18}{100} \checkmark$$

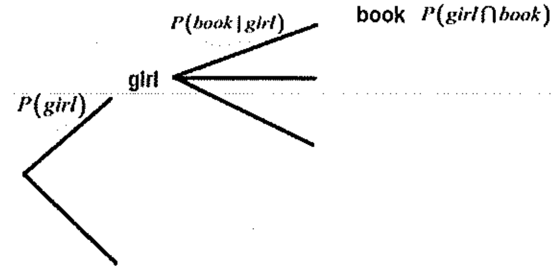
Now it is correct.
For AND cases we do multiply, but the 2nd probability is *<u>conditional</u>*.

Here's another way to picture what is going on. Even though we are just choosing a single student, we could imagine each event ('girl') ('book') as a 'choice' and imagine that these choices were made in sequence. We could then make a tree diagram showing possibilities:

If we label each branch with a probability, remember that the 2nd choices are all conditional probabilities, and the end nodes are all AND cases:



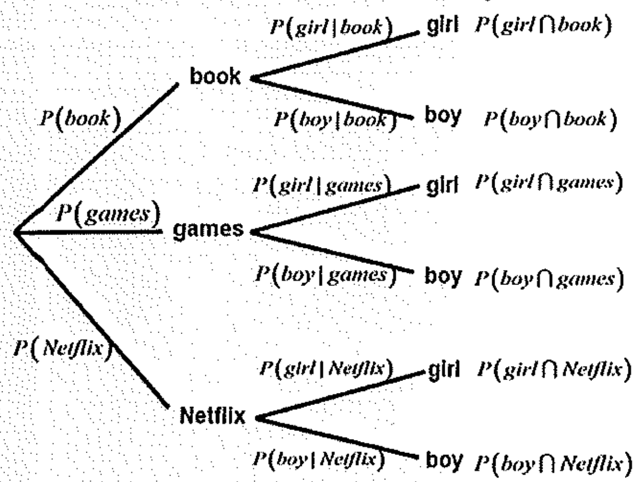
If we just look at the top set of branches...



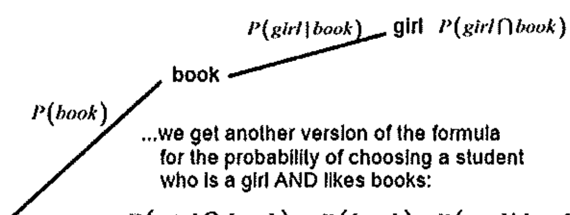
...it gives us an easy, visual way to remember the AND formula:

$$P(girl \cap book) = P(girl) \cdot P(book \mid girl)$$

Our decision to list the 'gender' choice first was arbitrary. We could have instead listed the 'activity' first:



If we again look at just the top branches...



...we get another version of the formula for the probability of choosing a student who is a girl AND likes books:

$$P(girl \cap book) = P(book) \cdot P(girl \mid book)$$

So it doesn't matter which event is listed first:

$$P(girl \cap book) = P(girl) \cdot P(book \mid girl)$$
$$P(girl \cap book) = P(book) \cdot P(girl \mid book)$$

Some people memorize that for probabilities 'AND equals multiplication'. That idea is correct, but for it to work in all cases, it needs to be 'AND equals multiplication, but one of the probabilities is a *conditional probability*'.

Probability of 'AND' case compound events:

$$P(A\ AND\ B) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$$
$$P(A \cap B) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)$$

This is the true 'AND' formula which works in all cases.

The special case for AND: Independent Events

In the 'OR' case we found that there was a 'special case' which was if the events were disjoint (mutually exclusive). This slightly altered (simplified) the 'OR' formula.

There is also a special case for the 'AND' situation: if the events are **independent**.

Events are considered **independent if the probability of one event happening does not change the probability that the other event will happen.**

The data we collected shows that 'favorite activity' is not independent from 'gender'.

To see why, consider what you would say if I asked you the question, "What is the probability that a student likes to play video games?"

	girls	*boys*	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

There are 3 probabilities you could find to try to answer this question:

$$P(games) = \frac{32}{100} \qquad P(games \mid girl) = \frac{12}{60} \qquad P(games \mid boy) = \frac{20}{40}$$

$$P(games) = .32 \qquad P(games \mid girl) = .20 \qquad P(games \mid boy) = .50$$

So your answer to the question would have to be, "It *depends.* Are we talking about boys, girls, or all students?"

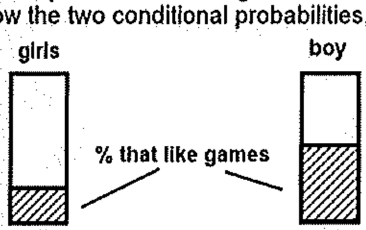
This means 'activity' and 'gender' are **not independent.**

How can we determine if two events are independent?

One way is to check **any two** of the three probabilities of one of the events, like we just did.

If you find that they don't match, then the events are not independent.

Sometimes, we are presented with segmented bar graphs that show the two conditional probabilities, like this:

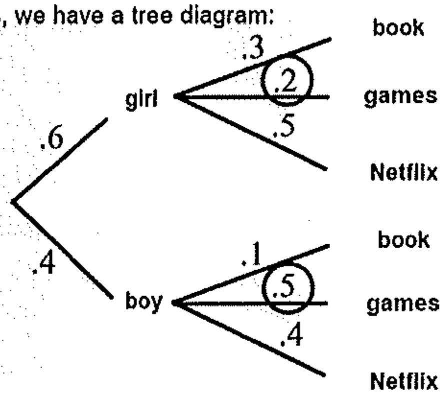


$$P(games \mid girl) = \frac{12}{60} \qquad P(games \mid boy) = \frac{20}{40}$$

$$P(games \mid girl) = .20 \qquad P(games \mid boy) = .50$$

If you find that the shaded regions don't match, then the events are not independent.

Sometimes, we have a tree diagram:



If you find that probabilities for the 2nd choice don't match, then the events are not independent.

Here's what it would look like if 'gender' and 'activity' ***were* independent:**

At first glance, this table doesn't look very different from the previous set of data.

Notice that the counts are the same between the boys and girls and still *not the same*.

	girls	*boys*	
read a book	12	8	20
video games	30	20	50
watch Netflix	18	12	30
	60	40	100

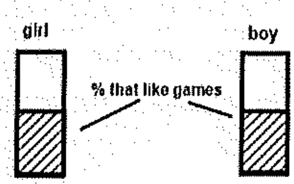
Yet, this data does show that 'gender' and 'activity' are independent.

We can't tell by comparing the counts between categories...we must compare the ***percentages.***

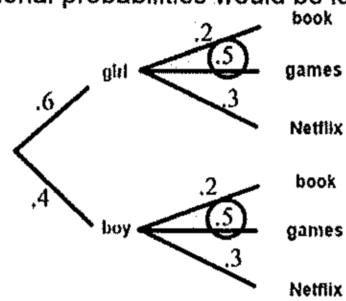
The ***percentages*** would now be the same:

$P(games)=\frac{50}{100}$ $\quad P(games \mid girl)=\frac{30}{60}$ $\quad P(games \mid boy)=\frac{20}{40}$

$P(games)=.50$ $\quad P(games \mid girl)=.50$ $\quad P(games \mid boy)=.50$



And on a tree diagram, corresponding conditional probabilities would be identical:



The special case 'disjoint events' modified the OR formula (it removed the overlap term).
Similarly, independent events are a special case for the AND formula and will modify (simplify) the formula:

<u>Special Case</u>: If the probability of an event does not change regardless of whether or not another event happens, then the events are <u>independent events.</u>

$$\text{If } P(B)=P(B \mid A)=P(B \mid \overline{A}), \text{ then } A \text{ and } B \text{ are independent.}$$

...and the 'AND' formula...

$$P(A \cap B)=P(A) \cdot P(B \mid A)$$

...simplifies to:

$$P(A \cap B)=P(A) \cdot P(B)$$

<u>Test for independent events</u>:

Two events are independent if $P(B)=P(B \mid A)=P(B \mid \overline{A})$

In other words: check any 2 of the 3 ways to find probability of B. If they don't match, then B is depending upon A and the events are not independent.

<u>Note</u>: Some books also use the simplified version of the AND formula as a 'test for independence'...

$$\text{If } P(A \cap B)=P(A) \cdot P(B), \text{ then } A \text{ and } B \text{ are independent.}$$

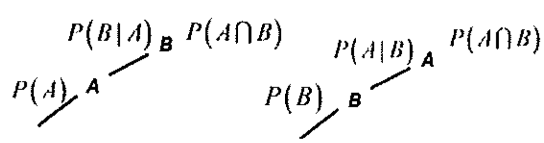
...but this is more a consequence of independence, not the reason the events are independent.

<u>Probability of 'AND' case compound events</u>:

$P(A \text{ AND } B)=P(A) \cdot P(B \mid A)=P(B) \cdot P(A \mid B)$

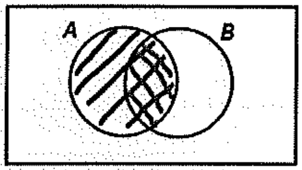
$P(A \cap B)=P(A) \cdot P(B \mid A)=P(B) \cdot P(A \mid B)$

Picture a part of a tree to remember the AND formula...



More about conditional probability

Now that we've defined how 'AND' works, we can say more about conditional probability. Conditional probability is a probability that is computed within only the part of the sample space that meets the condition:



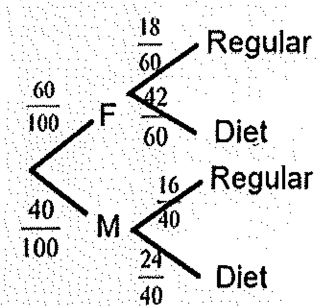
$$P(B \mid A) = \frac{\text{the part of } B \text{ that is within } A}{\text{all of } A}$$

But the red shading is the overlap. When we compute a probability fraction we can use either counts or probabilities, as long as we are consistent. So we can use probabilities for each shading:

Conditional Probability Formula: $P(B \mid A) = \dfrac{P(A \cap B)}{P(A)}$

Tables, Venn diagrams, Tree diagrams

	Regular	Diet	*total*
Female	18	42	*60*
Male	16	24	*40*
total	*34*	*66*	*100*



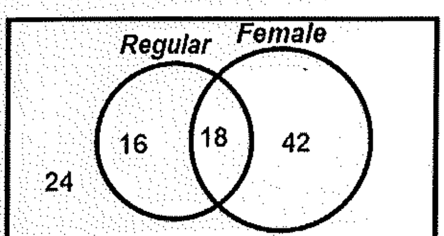
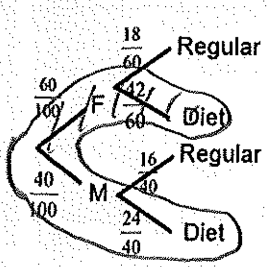


	Table	Tree	Venn
$P(M)=$	$\frac{40}{100}$	$\frac{40}{100}$ (1st branch)	$\frac{24+16}{24+16+18+42} = \frac{40}{100}$
$P(Diet \mid F)=$	$\frac{42}{60}$	(F part of tree) $\frac{42}{60}$	$\frac{42}{18+42} = \frac{42}{60}$
$P(Diet \mid M)=$	$\frac{24}{40}$	(M part of tree) $\frac{24}{40}$	$\frac{24}{24+16} = \frac{24}{40}$
$P(Diet \cap M)=$	(intersection) $\frac{24}{100}$	(end node) $\left(\frac{40}{100}\right)\left(\frac{24}{40}\right) = \frac{24}{100}$	$\frac{24}{24+16+18+42} = \frac{24}{100}$
$P(Diet)=$	$\frac{66}{100}$	(2 end nodes or'd) $\left(\frac{60}{100}\right)\left(\frac{42}{60}\right)+\left(\frac{40}{100}\right)\left(\frac{24}{40}\right)$; $\frac{42}{100}+\frac{24}{100}=\frac{66}{100}$	$\frac{24+42}{24+16+18+42} = \frac{66}{100}$
$P(F \mid Diet)=$	$\frac{42}{66}$	(Baye's formula or reverse conditional)	$\frac{42}{24+42} = \frac{42}{66}$

'Reversing the Condition' (Baye's Formula, a priori probabilities)



$$P(F \mid Diet) = \frac{\textit{probability of ways to get diet 'thru' } F}{\textit{total probability of ways to get diet}}$$

This determines the conditional sample space (the denominator)

$$= \frac{\left(\frac{60}{100}\right)\left(\frac{42}{60}\right)}{\left(\frac{60}{100}\right)\left(\frac{42}{60}\right)+\left(\frac{40}{100}\right)\left(\frac{24}{40}\right)}$$

$$= \frac{\frac{42}{100}}{\frac{42}{100}+\frac{24}{100}}$$

$$= \frac{42}{42+24}$$

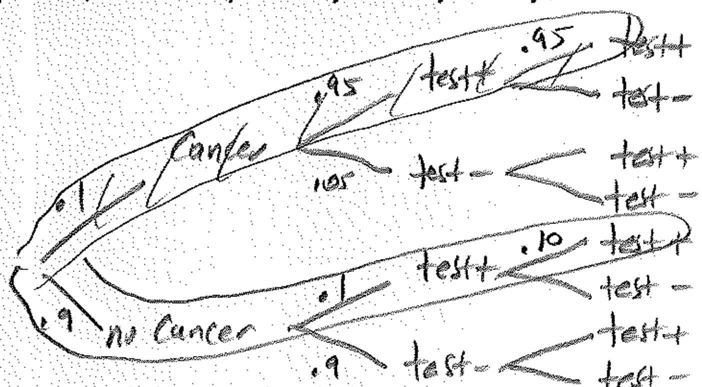
$$= \frac{42}{66}$$

Another example: A test exists to detect cancer. In people who have cancer, the test correctly detects cancer 95% of the time, and fails to detect the cancer 5% of the time. In people who do not have cancer, the test correctly reports no cancer 90% of the time, but falsely reports cancer 10% of the time. Long term data collection has shown that 10% of the population overall does have cancer.

If a person takes this test once and cancer is detected, what is the probability that they actually have cancer?

If a person who has tested positive undergoes a second test which is also positive, what is the probability that they actually have cancer?

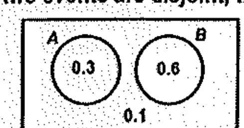
$$P(\text{cancer} \mid \text{test}+) = \frac{(.1)(.95)}{(.1)(.95)+(.9)(.1)} = .5135 \quad \boxed{51\%}$$

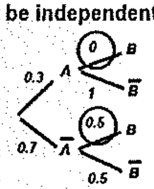


$$P(\text{cancer} \mid \text{test}++) = \frac{(.1)(.95)(.95)}{(.1)(.95)(.95)+(.9)(.1)(.1)} = .9093 \quad \boxed{91\%}$$

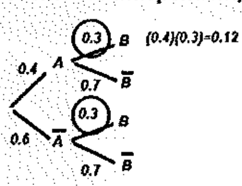
Independent and Disjoint

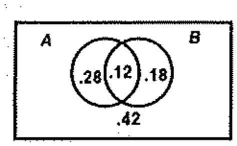
If two events are disjoint, they cannot be independent...





If two events are independent, they cannot be disjoint...





Summary of Probability

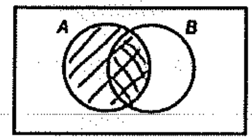
- Ways to list sample space: set, table, Venn diagram, tree diagram.
- $P(A) = \frac{\#\,outcomes\ in\ A}{\#\,total\ outcomes}$ but only for equally likely outcomes.

- <u>Conditional Probability:</u>

the event
...goes in the numerator of the fraction.

$$P(B \mid A) = \frac{\text{the part of } B \text{ that is within } A}{\text{all of } A}$$

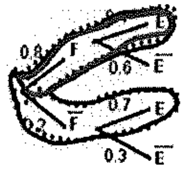
the condition
...goes in the denominator of the fraction.



	girls	boys	
read a book	18	4	22
video games	12	20	32
watch Netflix	30	16	46
	60	40	100

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

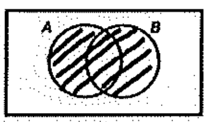
Reverse condition (Baye's) remember that it is a probability of paths through the tree:

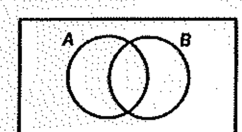


$$P(F \mid E) = \frac{(0.8)*(0.4)}{(0.8)*(0.4) + (0.2)*(0.7)}$$

- <u>OR case probabilities:</u>

  A OR B
  A union B
  $A \cup B$



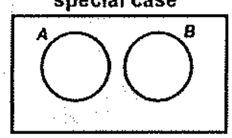


A and B are non mutually-exclusive
A and B are not disjoint events
A and B are joint events

$P(A \cap B) \neq 0$

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

special case



A and B are mutually-exclusive
A and B are disjoint events

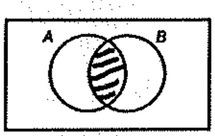
$P(A \cap B) = 0$

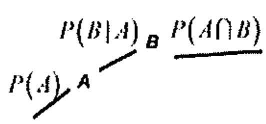
$$\boxed{P(A \cup B) = P(A) + P(B)}$$

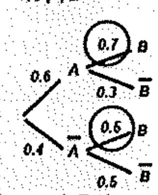
OR = addition, but subtract overlap (unless special case)

- <u>AND case probabilities:</u>

  A AND B
  A intersection B
  $A \cap B$







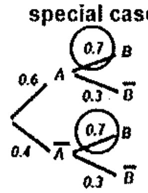
In general, B depends upon A...

$$P(B \mid A) \neq P(B \mid \overline{A}) \neq P(B)$$

...so 2nd probability must be conditional:

$$\boxed{P(A \cap B) = P(A) \bullet P(B \mid A)}$$

special case



A and B are independent if...

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

...so we can use the simpler AND formula:

$$\boxed{P(A \cap B) = P(A) \bullet P(B)}$$

AND = multiply, but 2nd probability is conditional (unless special case)