Unit 1 - Definitions of Probability and Basic Rules

Adam Petrie
Department of Business Analytics
University of Tennessee

January 30, 2021

1. What is Probability?
   - Think in terms of frequency
   - An investigation
   - Observations and Complications
2. Formal Discussion of Probability
   - Key Words
   - Frequentist Definition of Probability
   - Complications of the strict frequentist definition
3. Calculation when All Outcomes are Equally Likely
   - All outcomes equally likely
   - But what if infinite number of outcomes?
   - Problems with infinite sample spaces
   - Mind candy: Bertrand's paradox
4. Counting (Combinatorics)
   - Fundamental Theorem of Counting
   - Replacement vs. non-replacement
   - When order doesn't matter
   - Reordering a sequence
   - Summary
   - Using R
5. Basic Rules of Probability
   - Axioms
   - Complement rule
   - Addition Rule (A or B)
   - Multiplication Rule (A and B)

What is Probability?

What is Probability?

We often hear statements about the probability (or chance) that something occurs. Perhaps the weatherman just said there's a 70% of at least 2 inches of snow tomorrow.

What is Probability?

Perhaps the analytics team claims that there is a 1 in 4 chance that a customer who has called support 3 or more times will churn.

What is Probability?

Maybe a (bad) inventory control policy yields a stockout probability for a particular brand of soup is 35% during any given week.

What is Probability?

The Development Office might claim that the probability that an alumnus will donate more than $20,000 over his/her lifetime is 3%.

What do we mean when they say "the probability of an event is x" or that the chance of something happening is 1 in y?

Thinking in terms of frequency

In business analytics, we're often concerned with how *often* some event of interest occurs

- How often will a stockout occur if inventory is "reset" to 20 at the beginning of each week?
- How often will someone click on an ad when placed on the top-center of a webpage?
- How often will an Uber get a request for pickup from from Market Square?
- How often will someone who buys dry pasta also buy tomato sauce?

Frequentist view of probability

Sometimes, the event happens. Other times it does not.

Informally, the *frequentist view* of probability defines the probability of some event of interest as the long-run fraction (or proportion) of time when that event of interest occurs.

An investigation

Let's investigate what "long-run fraction of time when that event of interest occurs" means.

Classic example that we all "get": flipping a fair coin. Two options: heads or tails. "Obviously", the probability of the coin coming up heads is 50%. So, "in the long run, the fraction of coin flips that end up heads is 50%".

What does this mean and what are the implications?

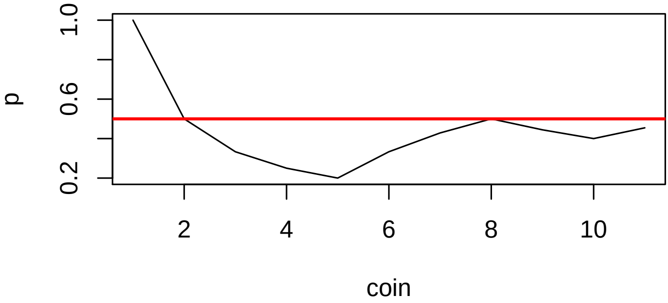
Investigation

Let's run a simulation to get a feeling for what 50% means. To simulate a coin, let's generate a 0 or 1 at random. If it is a 1, the coin comes up heads. If it is a 0, the coin comes up tails.

Flip	1	2	3	4	5	6	7	8	9	10	11
Random #	1	0	0	0	0	1	1	1	0	0	1
Heads	Yes	No	No	No	No	Yes	Yes	Yes	No	No	Yes
Fraction	1	1/2	1/3	1/4	1/5	2/6	3/7	4/8	4/9	4/10	5/11

Investigation

```
p <- c(1 , 1/2 , 1/3 , 1/4 , 1/5 , 2/6 , 3/7 , 4/8 , 4/9 , 4/10 , 5/11 )
coin <- 1:11
plot( p ~ coin , type= "l" )
abline(h=0.5,lwd=2,col="red")
```



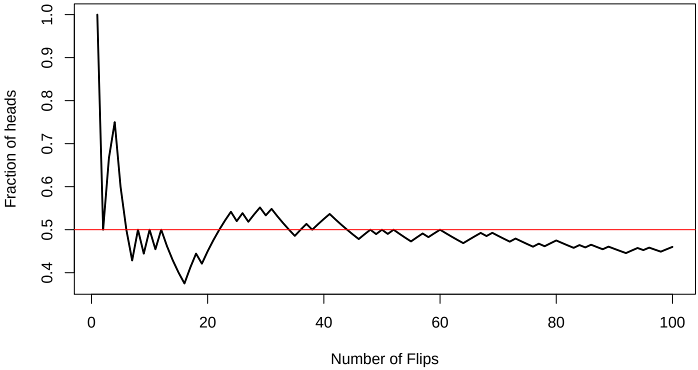
Investigation

In the example so far, 11 flips is "hardly" the long run. In this simulation, the fraction $5/11 \approx 45\%$ is indeed pretty close to 50%. In general, it's possible that, after 11 flips, the fraction of heads is much smaller or much larger than 50% (e.g., about a 1 in 10 chance of getting 3 or fewer heads out of 11 flips, for a fraction of 27% or smaller!).

Let's extend the simulation to a larger and larger number of trials.

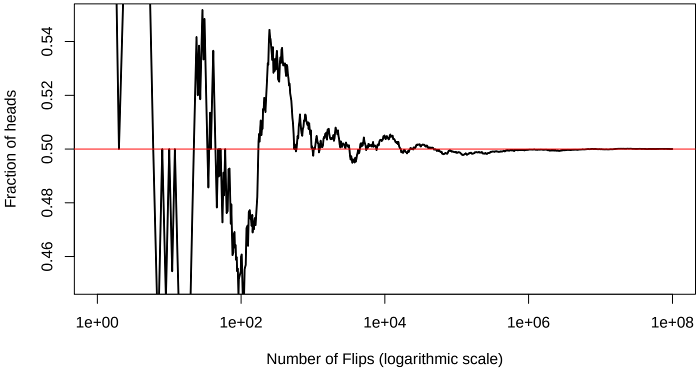
Extended investigation 1

After 100 flips, the fraction that have come up heads is still noticeably below 50%. 100 must not be "in the long run".



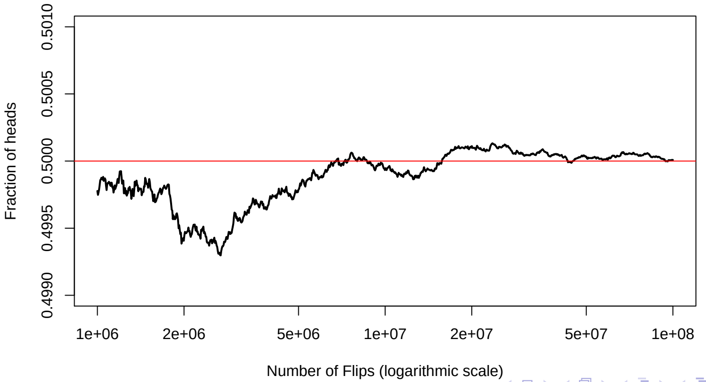
Extended investigation 2

After 100,000,000 $= 10^8$ flips, the fraction that have come up heads looks to be right at 50%. Is 10^8 what we mean by "in the long run"?



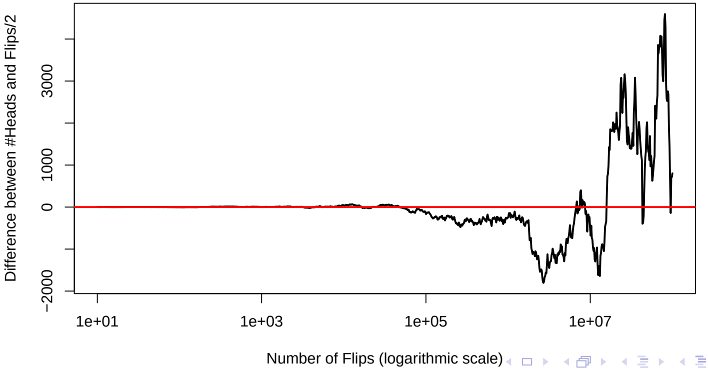
Extended investigation 3

If we zoom in a bit on the plot, the fraction that have come up heads isn't *right* at 50%, but it certainly is close. "In the long run" doesn't actually refer to a specific number of flips, rather the phrase is just a useful abstraction.



Extended investigation 4

After n flips, you might expect about $n/2$ flips to be heads. However, while the typical difference between the fraction of flips that have come up heads and 50% approaches 0, the typical difference between *number of heads* and $n/2$ does *not* approach 0.



What did we learn from the simulation?

- The proportion of flips that are heads seems compelled to get closer and closer to 50% as the number of flips grows, though the observed proportion might be a bit above or a bit below 50% after any particular number of flips.
- The deviation between the proportion of heads and 50% seems to get smaller and smaller (approaching 0) as the number of flips increases.

However . . .

- The *number* of flips that come up heads doesn't seem to be "driven" towards the number of flips divided by 2.
- Be absolutely clear on this: when discussing probability, it's about the *fraction* or proportion of time you observe the event, **not** about the number of times the event is observed.

Problematic Definition of Probability

Because the probability of an event refers to a number that emerges "in the long run" (technically, infinity!), its definition is somewhat problematic.

Except in special cases (where all possible outcomes an event are known and are equally likely to occur, like the faces of a die roll), the probability of an event is never directly observed and we can never truly *know* the numerical value of a probability.

Formal Discussion of Probability

Definitions

Trial - an experiment

- flipping a coin
- checking retention status after contract expires
- observing the amount of money someone has donated
- measuring time to go to work

Outcome - possible result of a trial

- heads/tails
- churn/stay
- 0, \$1, \$2, . . .
- 15 minutes and 12 seconds

Definitions

Sample space - list of all possible outcomes to a trial

- flipping a coin - {H,T}
- retention - {churn, stay}
- donation amount - any two decimal number that is 0 or more
- time - any positive number

Event - one *or more* outcomes of a trial (e.g., one or more items in the sample space)

- heads
- churn
- exactly \$300, between \$500-\$1000, over \$10,000, etc.
- between 13-19 minutes

Frequentist Definition

- We talk about the probability that an *event* occurs
- When we say that "probability is the fraction of time that an event occurs", we really mean P(event) = long run fraction of trials where the event occurs

$$P(event) = \lim_{\#trials \to \infty} \frac{\#\text{ trials in which event occurs}}{\#\text{ trials}}$$

- Special condition: each trial is *independent* and conducted under *identical* conditions (more on that later).

Frequentist Definition Example

P(heads) = 50%

- In the long run, 50% of flips turn up heads
- Remember that this does *not* mean that after 5000 flips, exactly 2500 will be heads

Probability vs. Chance vs. Odds

The terms *probability* and *chance* are used interchangeably. It's fine to say "the probability the coin comes up heads is 50%" or to say "the chances that the coin comes up heads is 50%".

The **odds** of an event is a completely different quantity, and we'll avoid using it in this class. If p is the probability that some event occurs, then the odds of that event occurring is:

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

Interpreting odds

To convert odds into probability.

- If the odds of an event equals $1/8$ (usually written 1:8), for every 1 time the event occurs it does not occur 8 times. In other words, the probability the event occurs is 1 out of 9, or $1/9$.
- If the odds of an event equals $8 = 8/1$ (usually written 8:1), for every 8 times the event occurs, it does not occur 1 time. In other words, the probability the event occurs is 8 out of 9, or $8/9$.
- Note: in sports betting, the ordering of the numbers is reversed so 8:1 odds instead means a 1/9 chance the team wins.

GEMS

My favorite slot machine of all time is GEMS-Wild Tiles

https://www.gametwist.com/en/games/slots/gems-wild-tiles/

GEMS

If 3 (or more) tiles match, the middle tile turns into a "wild" tile and the remainder disappear, with any tiles above the now-empty spaces falling down to take their place as necessary. Tiles match again, etc., and the process continues until no further matches are made. The amount you win depends on how many tiles are ultimately matched.

What's the probability that you win at least *something* during a game of GEMS? It turns out that it is about 45%. What does this mean?

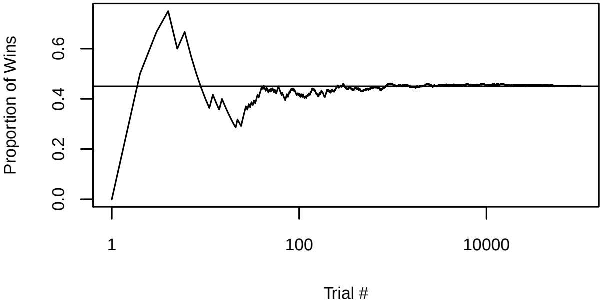
GEMS-Example

P(win) for game of GEMS is 45%. Run a lot of games and keep track of proportion of wins. Eventually, this fraction will get arbitrarily close to 0.45.

```
outcomes <- sample(c(0,1),size=1e5,prob=c(.55,.45),replace=TRUE)
head(outcomes,15)  #First 15 games
##  [1] 0 1 1 1 0 1 0 0 0 0 0 1 0 0 1
head( cumsum(outcomes), 15 ) #cumulative sum; gives you a running total of number wins so far
##  [1] 0 1 2 3 3 4 4 4 4 4 4 5 5 5 6
proportion <- cumsum(outcomes)/(1:1e5)  #Proportion of trials that result in wins
```

GEMS-Example

```
plot(1:1e5,proportion,log="x",type="l",xlab="Trial #",ylab="Proportion of Wins")
abline(h=0.45)  #True probability
```



Complications to the frequentist definition

The long-run fraction only converges to the probability if

- Trials are independent (the outcome of a trial does not depend on outcomes of any previous trials)
- Trials are repeated under *identical* conditions

Measured proportion after a large number of trials still won't *exactly* equal the true probability but it will be "close"

- Can derive a "confidence interval" for the true probability
- If $\hat{p}$ is the fraction of n trials where event occurs, you can be 95% confident that the true probability is between

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

- We'll "discover" this formula shortly so that it doesn't look like it comes out of thin air.

Complications regarding computation

Very difficult to ensure that trials are repeated under identical conditions (usually only possible in computer simulations)

- P(make a free throw) – fatigue, muscle memory, etc.
- P(Pat makes a purchase) – Pat's finances might change over time
- P(less than 15 min to work) – might be VERY dependent on what time you leave, day, weather; each trial "looks" different

Can be very time consuming to perform a trial

- P(a student succeeds after graduating from UT)
- P(get lung cancer after smoking 2 packs a day for 10 years)

Sometimes it's impossible to have repeated trials. P(rain tomorrow) $= 30\%$? But 2021-01-17 only happens once.

Further complications

The formal frequentist definition of probability is a bit more restrictive than what you'd like it to be.

- Take a coin and flip it, letting it fall to the floor.
- Without looking at it, cover it with your foot.
- What's the probability that the coin is heads?

Further complications

The answer (according to a strict interpretation of the frequentist definition) is NOT 50%!

- Since flip has already occurred, the coin either *is* or *is not* heads.
- If we consider a "trial" as checking the face of the coin, then either 100% of times we check it will be heads, or 100% of times we check it will be tails.

Technically, the probability the coin we are stepping on is heads is either 0 or 1, we just do not know which one it is.

Further complications

Under the frequentist definition, we do not use probability to quantify our ignorance about the state of the world.

If the outcome of some trial has been determine (you just don't know what it is), it doesn't make sense to use probability to describe the likelihood of different outcomes.

Further complications

In fact, the word probability really shouldn't be used to describe *specific* events at all.

Does it makes sense to talk about the probability that the *next* time we flip a coin there is a 50% chance of it coming up heads?

- It is true that, in the long run, 50% of flips in general result in heads.
- However, every time we check the result of a *specific* flip, it will either be heads 100% of the time, or be tails 100% of the time.

Under the strictest interpretation of the frequentist view of probability, it only makes sense to talk about what happens when we flip a coin *in general*, and not what happens on a particular flip.

Problematic Definition: Ken oversleeping

Another example. Take the statement: "Ken is late today; the probability he overslept is 0.80". We might "get" what this statement means intuitively, but mathematically it is problematic.

Ken either has or hasn't overslept, we're just unsure as to what happened. Under the strict frequentist definition, we don't use probability as a way to quantify our ignorance of reality. Thus, the event "Ken overslept" is not one we can use probability to describe the likelihood that it happened.

Problematic Definition: Ken oversleeping

The statement about Ken is problematic in other ways. There's no sense of repeated trials. "Today" happens just once and there's no way to repeat "today" under identical conditions.

It may be true that out the last 100 times Ken has been late, 80 times he overslept. But there's no reason to suspect this trend will continue into the future (these aren't independent trials repeated under identical conditions).

To say the probability that Ken overslept is 80% implies that the fraction of days where Ken is late is because he overslept creeps closer and closer to 80%.

Although we all intuitively *get* what "Ken is late today; 80% chance he overslept" means, it's not a statement that makes sense under the strict frequentist definition. Pedantic I know! But this is a class in probability!

Bayesian approach

Ultimately, the frequentist definition of probability is a bit of a letdown because it is so narrow.

The **Bayesian definition** compensates for these shortcomings. Under the Bayesian view:

- we *do* use probability to quantify our ignorance of events that have already happened
- we *do* use probability to talk about the outcome of a specific trial that has yet to be performed
- we can incorporate previous knowledge or beliefs about the probability of event occurring in probability calculations

Calculation when All Outcomes are Equally Likely

Shortcut when all outcomes in a sample space are equally likely

Occasionally, all outcomes of an experiment are equally likely (i.e., the outcome is picked *at random*).

- Pick a random # between 1-10
- Flip a fair coin
- Roll a balanced die
- Determine the winning Powerball number
- Generate an initial board for GEMS

When this is the case, the *probability of an event is the fraction of outcomes in the sample space that contain the event*.

$$P(event) = \frac{\text{\# outcomes that satisfy event}}{\text{total \# outcomes in sample space}}$$

Example 1 - Picking a random number

Question: A number if picked at random between 1-15. What is the probability it is a 5, a multiple of 5, an odd number, an odd number that is a multiple of 5?

Solution: Begin by enumerating the sample space.

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$$

- P(5) – There is only one outcome that satisfy this event: the number 5. Since one of the 15 outcomes in the sample space corresponds to the event of interest, P(pick 5) $= 1/15$.
- P(multiple of 5) – There are three outcomes that satisfy the event (5, 10, 15), so the probability is $3/15$.
- P(odd number) – There are eight outcomes that satisfy the event (1,3,...,15), so the probability is $8/15$.
- P(odd number that is a multiple of 5) – There are only two outcomes that satisfy this event (5,15), so the probability is $2/15$

Example 2 - Dice rolling

Question: Two dice are rolled and both faces are recorded. What is the probability that at least one 5 is rolled? The sum of both faces is 5? The maximum number on any face is 4?

Solution: Begin by enumerating the sample space, letting a roll be represented by a two-digit number.

11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

- P(at least one 5 is rolled) – There are 11 outcomes that satisfy this event (51, 52, 53, 54, 55, 56, 15, 25, 35, 45, 65), and there are 36 possible outcomes, the probability is 11/36.
- P(sum is 5) – There are four outcomes that satisfy the event (14, 23, 32, 41), so the probability is 4/36.
- P(maximum number at most 4) – There are 16 outcomes that satisfy the event (upper four rows/left four columns), so the probability is 16/36.

Caution: only apply this formula when outcomes are picked at random

One of the biggest fallacies people make in probability is assuming that because a trial has n outcomes that they are all equally likely. This is ONLY true if the outcomes are being picked at random.

- P(A in this class) = ?
  Seven outcomes (A, B+, B, C+, C, D, F), but the probability of getting an A is not 1/7 since I do not assign your grade at random.
- P(pregnant after intercourse) = ?
  Two outcomes (yes/no), but the probability of getting pregnant is not 50% since the outcome is not picked at random (many factors contribute like birth control use, age, ovulation status, etc.)
- P(intelligent life exists beyond Earth) = ?
  Two outcomes (yes/no), but the probability is not 50% since this outcome is not picked at random. In fact, this question is ill-posed because it can't be answered using our frequentist approach (it either does or does not, we just don't know it).

Infinite sample spaces

Sometimes the number of outcomes in a sample space is infinite.

- Time to get to work (assuming clock with arbitrary precision)
- Points on a dartboard

The formula

$$P(event) = \frac{\#\text{ outcomes that satisfy event}}{\text{total }\#\text{ outcomes in sample space}}$$

breaks down when the sample space is infinite because the anything divided by infinity is 0. We need a more general definition.

$$P(event) = \text{fraction of sample space that contains the event}$$

This formula works as long as all points in the sample space are equally likely and the outcomes are being picked at random. Mathematicians have a much more formal definition of probability in this case, but this definition is intuitive and easy to work with.

Example - Stoplights

Question: A stoplight is red for 30 seconds, green for 11 seconds, and yellow for 4 seconds. What is the probability that the light will be green when we show up?

Solution: The "sample space" looks like

In effect, our experiment is picking a random location on the line above (45 units in length) and noting its color. There are an infinite number of points to choose from so we need to be careful how we calculate the probability.

In this case, green points make up $11/45$ of the points in the sample space (the line is 45 units long, of which 11 are colored green), so $P(\text{green}) = 11/45$.

Example - Dartboard

Question: A dartboard with radius one has a bullseye with radius 0.05. What is the probability of hitting the bullseye, assuming our dart-throwing skills are consistent with hitting the dartboard at random?

Solution: P(bullseye) is the fraction of the sample space containing the event, i.e., the fraction of the dartboard that contains the bullseye.
The area of a circle with radius r is πr^2. Thus, the area of the dartboard is $\pi \times 1^2$ while the area of the bullseye is $\pi \times 0.05^2 = 0.0025\pi$.

$$P(bullseye) = \frac{\text{Area of bullseye}}{\text{Area of dartboard}} = \frac{0.0025\pi}{\pi} = 0.0025$$

Probability of any particular outcome for infinite sample spaces is zero

Question: A number between 0-1 is picked at random. What is the probability that it is greater than 0.75? Exactly 0.75?

Solution:

- P(number greater than 0.75) $= 25\%$. The sample space is like a line of length 1. The probability of getting a number greater than 0.75 is 25%, since 1/4th of this line has numbers greater than 0.75.
- P(number equals 0.75) $= 0$. Using the classic formula, there is exactly one outcome that satisfies our event (getting 0.75). Dividing 1 by ∞ gives 0.

Mindblowing paradox? If the experiment is performed and a number of picked at random, we *will* get a result (granted, the number will be infinitely long). However, the probability of getting that number is 0 according to the math, so something with 0 probability just occurred (zero probability is impossible).

Paradox resolution

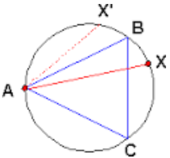
In the real world, there aren't infinite sample spaces

- continuous quantities are measured out only to a certain precision (e.g., time to the nearest second)
- some theories of physics posit that space and time itself is quantized, i.e., there is a finite smallest unit of length and chunk of time

Very often it is *easier* to treat a sample space as infinite because the math is more approachable. As it turns out, *infinity is more of a conceptual tool* that makes calculations easier, so let's use it.

Another mindblowing paradox: Bertrand

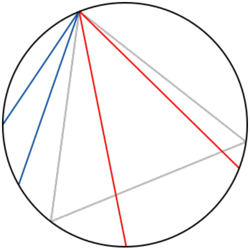
Consider an equilateral triangle inscribed in a circle. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle?



Note: A chord connects 2 points on the perimeter of a circle. One of the points doesn't have to coincide with a vertex of the triangle, it just does here to allow easy comparison.

Bertrand's Paradox: Solution 1

Pick two points on the circle at random and connect them to make the chord. For reference, put a vertex of the triangle at one of the end points to allow for easy illustration.

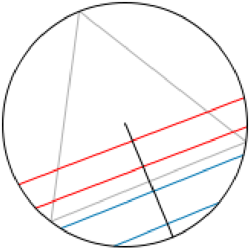


Red: longer than triangle edge. Blue: shorter than triangle edge.

$$P(chord\ longer) = 1/3$$

Solution 2

Pick a random point along some radius of the circle to serve as the midpoint of the chord (which is drawn perpendicular to the radius)

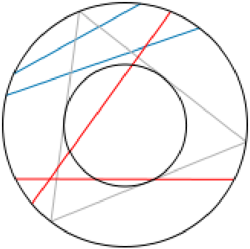


Chord is longer than a side of the triangle if the chosen point is nearer to the center of the circle than the point where the side of the triangle intersects the radius, so

$$P(chord\ \ longer) = 1/2$$

Solution 3

Pick a random point inside the circle to serve as the midpoint of the chord.



Chord is longer than a side of the triangle if the chosen point is inside the concentric circle with radius $1/2$, so

$$P(chord\ \ longer) = 1/4$$

Resolution

All solutions are correct. The issue is that "picking a chord at random" is an ill-defined procedure which (as we have seen) can be accomplished numerous ways.

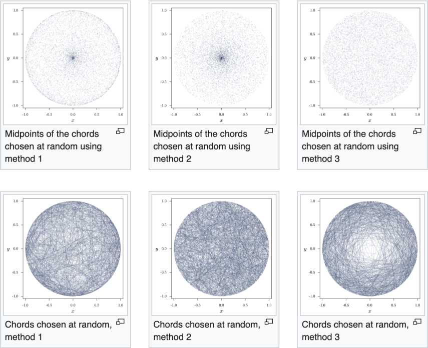
When there are an infinite number of possible outcomes, might there be numerous procedures for picking one "at random" that give rise to numerous valid answers?

Advice: use infinity as a tool to help you get answers. Like Santa Claus, the concept is useful and fun to think about, but (spoiler alert) I don't think it really exists. Besides, NASA has launched probes that have visited all the planets (and Pluto too) using the "wrong" theory of gravity, relying on the useful "tool" of Newtonian dynamics to get there.

Resolution

Which method of picking a random chord "feels" the most random to you?

Scatterplots Showing Simulated Bertrand Distributions,
Midpoints/Chords Chosen at Random Using 1 of 3 Methods.[citation needed]



Midpoints of the chords chosen at random using method 1

Midpoints of the chords chosen at random using method 2

Midpoints of the chords chosen at random using method 3

Chords chosen at random, method 1

Chords chosen at random, method 2

Chords chosen at random, method 3

Counting and Combinatorics

Importance of counting

Using the formula

$$P(event) = \frac{\text{\# outcomes that satisfy event}}{\text{total \# outcomes in sample space}}$$

hinges upon our ability to count the number of outcomes in the sample space and the number of outcomes that satisfy our event. Brute-force enumerating the sample space may not be possible.

- What is sample space if 100 coins are flipped (i.e., how many different sequences of 100 H's and T's are possible)?
- What is sample space of Powerball numbers (five unique white numbers 1-69 and one red number 1-26)?
- What is sample space if 25 people are randomly put in groups (could have five groups of five, or eight groups of three and one group of one, etc.)

Fundamental theorem of counting

Imagine a trial consists of k experiments to be performed in order. Experiment 1 has n_1 outcomes, experiment 2 has n_2 outcomes, etc., and experiment k has n_k outcomes. Then there are $n_1 \times n_2 \times \ldots \times n_k$ ways the sequence of experiments can be performed.

Fundamental theorem of counting

Fundamental theorem of counting

Example: A restaurant has three choices of appetizer, four choices of main course, and two desserts. How many different meals can be constructed?

Solution: from the fundamental theorem of counting the number of meals is $3 \times 4 \times 2 = 24$. If we wanted to be more clever and allow for the possibility of "none of the above" for each course, then there would be $4 \times 5 \times 3 = 60$ meals (well, perhaps only 59 since "none of the above" for all three wouldn't constitute a meal).

Match made in heaven

Question: You and your date are visiting the previously mentioned restaurant where there are three choices of appetizer, four choices of main course, and two desserts. Your date has picked their meal. If you pick yours at random, what is the probability that it is a match?

Match made in heaven

Solution: there is only one outcome that satisfies the event (your meal is one particular combination). So the probability is $1/n_{meals}$, where n_{meals} is the total number of meals that can be constructed. From the fundamental theorem of counting this is $3 \times 4 \times 2 = 24$. So there is a $1/24$ chance you'll match if you pick your meal at random.

Coin flips and Cards

Question: How many unique sequences of 100 (fair) coin flips are possible?

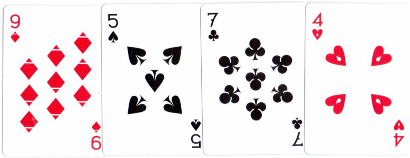
Coin flips and Cards

Question: How many unique sequences of 100 (fair) coin flips are possible?

Solution: Each flip has two outcomes, so the fundamental theorem of counting says there are $2 \times 2 \times \ldots \times 2 = 2^{100} = 1.27 \times 10^{30}$ possible sequences.

Coin flips and Cards

Question: How many different sequences can four cards be dealt from a well-shuffled deck?



Coin flips and Cards

Question: How many different sequences can four cards be dealt from a well-shuffled deck?

Solution: There are 52 cards in a deck. There are thus 52 possibilities for the first card dealt, 51 for the second, 50 for the third, and 49 for the fourth. From the fundamental theorem of counting, there are $52 \times 51 \times 50 \times 49 = 6497400$ sequences. Note: in a "hand" of poker, the order of the cards does not matter, e.g. 23456 is the same as 63425, so there are substantially fewer poker hands.

A note on replacement

Often, our experiment is a series of repeated trials, each picking an outcome from the same or related sample space.

- Flipping four coins is equivalent to performing four trials of flipping a single coin once
- Picking four cards out of a deck is equivalent to performing four trials of picking a single card out of a deck while allowing the deck composition to change between each trial.

A note on replacement

When the sample space of each repeated trial is the same, we say that the outcomes are being picked *with replacement*. In other words, it is possible to get the same outcome (e.g., heads) on multiple trials. In this case, the outcomes of each trial are said to be *independent* of each other.

When the sample space of each repeated trial changes (outcomes of previous trials are not allowed to occur again), we the outcomes are being picked *without replacement*. In this case the trials are not independent since the results of one depends on the results of the others.

Sequences from repeated trials using the same sample space

Imagine an experiment has n different outcomes. If a trial consists of k of these experiments then it is possible to count up the number of resulting sequences of results.

- If outcomes are picked with replacement there are n^k different sequences.
  - Flip 10 coins $\rightarrow 2^{10}$
  - Roll 5 dice $\rightarrow 6^5$
  - Picking 3 cards from a deck (and shuffling in one that is picked before picking again) $\rightarrow 52^3$
  - Tennessee Pick 4 $\rightarrow 10^4$ (numbers are 0-9)

Sequences from repeated trials using the same sample space

- If k outcomes are picked without replacement, the number of different sequences is

$$n \times (n-1) \times (n-2) \times \ldots \times (n-k+1) = \frac{n!}{(n-k)!}$$

  - Picking 3 cards from a deck (and NOT shuffling in any that have been picked) $\rightarrow 52 \times 51 \times 50$
  - PINs that cannot have duplicate numbers: $10 \times 9 \times 8 \times 7 = 5040$

Note: n! can be written in R as `factorial(n)`.

Ordering

The fundamental theorem of counting is not useful when the *ordering* of the experiments does not matter.

- Hand of poker consists of five cards drawn from a deck without replacement, but many hands are equivalent, e.g. the sequence 22773 and 37272 are the same hand.
- The white numbers in Powerball are five numbers drawn without replacement. While the sequence of numbers drawn may be 4 10 37 2 21, any ticket which has those five numbers is a winner.
- The order does not matter when choosing a *pair* of people from a group of four since the pair A and B is the same as the pair of B and A

Ordering

If a total of k objects are drawn from a list of n items, then

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

gives the number of unique *groups* (i.e. sequences where ordering does not matter). This symbol is pronounced "n choose k". In R, this can be written `choose(n,k)`.

Formula illustration

Imagine a lottery where four numbers 0-9 are picked without replacement. Thus, 4190 is a valid result while 4110 is not. Further, the ordering of the numbers does not matter when determining a winning ticket – 4190 is just as much as winner as 0941. How many possible sequences exist?

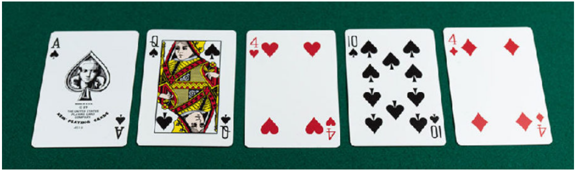
Formula illustration

If ordering *did* matter, then the fundamental theorem of counting tells us that there are $10 \times 9 \times 8 \times 7$ sequences possible (since numbers are picked *without* replacement). Since ordering does not matter, many of these sequences (e.g. 4190 and 0941) are redundant.

Given a sequence of four numbers like 4190, how many possible ways can we rearrange them? There are four choices for the first number, three choices for the second, etc., for a total of $4 \times 3 \times 2 \times 1 = 24$ ways. Thus, only 1/24th of the sequences results in a unique grouping.

$$\frac{1}{4 \times 3 \times 2 \times 1} \times (10 \times 9 \times 8 \times 7) = \frac{1}{4!} \times \frac{10!}{6!} = \frac{10!}{6!4!} = \binom{10}{4} = \binom{n}{k}$$

Examples



Question: How many possible poker hands are there?

Solution: since this is picking 5 cards without replacement from a deck of 52, the answer is "52 choose 5", or $(52!)/(5!47!) = 2598960$.

Examples

Question: How many Powerball tickets are there?

Solution: a Powerball ticket consists of 5 white numbers drawn without replacement from 1-69 and a red number from 1-26. There are "69 choose 5" = 11238513 groups of white numbers. Since a Powerball ticket consists of two experiments (picking the white numbers and picking the red), the fundamental theorem of counting says that there are $11238513 \times 26 =$ 292,201,338

Note: this agrees with
`http://www.powerball.com/powerball/pb_prizes.asp`

Full House

Question: What is the probability of a "full house" in poker (three cards with the same value and two cards of another).

Full House

Question: What is the probability of a "full house" in poker (three cards with the same value and two cards of another).

Solution: to find this probability we simply find

$$\frac{\#\text{ hands with a three-of-a-kind and a pair}}{\#\text{ poker hands}}$$

The denominator is 2598960 from the previous slide. What about the numerator?

Full House

Getting a full house can be cast as a sequence of four experiments

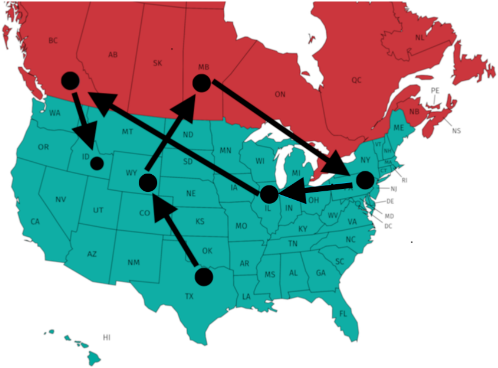
- Choosing a card value for the three-of-a-kind (there are "13 choose 1", or 13 different choices).
- Choosing three of the four cards with that card value to make the three-of-a-kind. Ordering doesn't matter, so this is "4 choose 3" or $4!/(3!1!) = 4$).
- Choosing a card value for the pair (there are now only "12 choose 1" or 12 choices).
- Choosing two of the four cards of that value to make the pair. Ordering doesn't matter so this is "4 choose 2" or $4!/(2!2!) = 6$).

The fundamental theorem of counting says that there are thus $13 \times 4 \times 12 \times 6 = 3744$ poker hands that are "full houses", so P(full house) $=$ $3744/2598960 = 0.00144$.

http://en.wikipedia.org/wiki/Poker_probability

Planning a trip

A couple is planning to visit 5 US states and 2 Canadian provinces.



Planning a trip

Question: A couple is planning to visit 5 US states and 2 Canadian provinces. How many trips can be planned if:

- Order of the visits does matter (i.e., we are counting unique itineraries)?
- Order of the visits does not matter.

Planning a trip

Solution when order doesn't matter.

Choose 5 of the 50 states, and 2 of the 10 provinces.

$$\binom{10}{2} \times \binom{50}{5} = 95,344,200$$

Planning a trip

Solution when order matters (unique itineraries)

- We just saw there were $\binom{10}{2} \times \binom{50}{5}$ itineraries if order doesn't matter.
- If order does matter, we can re-arrange the 7 stops in 7! different ways.

```
choose(10,2)*choose(50,5)*factorial(7)

## [1] 480534768000
```

Guessing Game

I've been to all 50 states! The last state I had to my list was Oklahoma.

Guessing Game

Question: I ask you to guess what the 48th and 49th states were that I added to my list. You decide to pick five states at random (not Oklahoma). If five states are picked at random, what is the probability that one (and only one) is on the list of not being visited?

Alabama	Indiana	Nebraska	Rhode Island
Alaska	Iowa	Nevada	South Carolina
Arizona	Kansas	New Hampshire	South Dakota
Arkansas	Kentucky	New Jersey	Tennessee
California	Louisiana	New Mexico	Texas
Colorado	Maine	New York	Utah
Connecticut	Maryland	North Carolina	Vermont
Delaware	Massachusetts	North Dakota	Virginia
Florida	Michigan	Ohio	Washington
Georgia	Minnesota	~~Oklahoma~~	West Virginia
Hawaii	Mississippi	Oregon	Wisconsin
Idaho	Missouri	Pennsylvania	Wyoming
Illinois	Montana		

Guessing Game

Solution:

$$P(1\ correct) = \frac{\#\text{ of groups of 5 states that contain exactly 1 unvisited state}}{\#\text{ groups of 5 states}}$$

Let's break it down.

The denominator is easy – "49 choose 5" or 1906884.

Guessing Game

Calculating the numerator requires realizing "picking 1 unvisited state out of 5" can be thought of as two sequential experiments:

- selecting 1 of the 2 states in the set *then*
- selecting 4 of the 47 states not in the set (you aren't allowing Oklahoma)

By the fundamental theorem of counting, this works out to be $\binom{2}{1} \times \binom{47}{4} = 356730$.

The probability is thus $356730/1906884 = 18.7\%$ of having exactly one correct guess.

Sequences when picking with replacement and order does not matter

What if a lottery is run where the same number can appear more than once and order doesn't matter, e.g., if 1174 and 4171 are both winning tickets. In this case, the number of combinations when choosing k objects from a list of n when order does not matter and picking is done with replacement is

$$\binom{n-1+k}{k} = \frac{(n-1+k)!}{k!(n-1)!}$$

The reasoning behind this formula is elegant and clever, but it takes some work to understand. Let's skip it and accept it as is!

Reordering a sequence

If a sequence contains n items, where there are n_1 items of type 1, n_2 items of type 2, …, n_r items of type r, then there are

$$\frac{n!}{n_1!n_2!\cdots n_r!}$$

ways of rearranging the sequence.

Reordering a sequence

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Example: how many ways are there of rearranging the letters in TENNESSEE? This is a sequence of 9 total letters with 4 Es, 2 Ns, 2Ss, and 1 T.

$$\frac{9!}{4!2!2!1!} = 3780$$

If an arrangement of the letters in TENNESSEE is selected at random, what is the probability it'll actually spell TENNESSEE?

Out of the 3780 ways of arranging the letters, exactly 1 results in the word we are looking for, so $1/3780$.

Binomial formula

Often a sequence is full of only two types of items

- Sequence of heads/tails
- Sequence of purchase/not purchase
- Sequence of "success" and "fail"

If k is the number of "successes", then $n-k$ is the number of failures and the number of ways to rearrange the items in the sequence while having the total number of successes be the same is

$$\frac{n!}{n_{success}!\,n_{fail}!} = \frac{n!}{k!(n-k)!}$$

This is the well known binomial formula.

Binomial formula

Example: I pass through 9 stoplights on the way to work. How many sequences have exactly 6 reds and 3 greens? One such sequence is RRRRRRGGG, but the total number of sequences we can get by rearranging these items is $9!/(6!3!) = 84$.

Summary

	With replacement	Without replacement
Order matters	n^k	$\frac{n!}{(n-k)!}$
Order does not matter	$\binom{n-1+k}{k}$	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Rearranging objects in a sequence (n_1 of type 1, etc.)

$$\frac{n!}{n_1!n_2!\cdots n_r!}$$

Number of sequences with k successes and $n-k$ failures (binomial formula)

$$\frac{n!}{n_{success}!n_{fail}!} = \frac{n!}{k!(n-k)!}$$

Using R for factorials and combinations

To get the factorial of something, like 5!, type `factorial(5)`. To get "n choose k", like 5 choose 2, type `choose(5,2)`

```
factorial(5)
## [1] 120
choose(5,2)
## [1] 10
```

To produce all "n choose k" combinations, use `combn(itemlist,k)`

```
itemlist <- c("A","B","C","D")
combn(itemlist,2)
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,] "A"  "A"  "A"  "B"  "B"  "C"
## [2,] "B"  "C"  "D"  "C"  "D"  "D"
```

expand.grid getting combinations of different items

Imagine you assign a letter A-C and a number 1-3 to someone. How many combinations exist?

```
letter <- c("A","B","C")
number <- 1:3
expand.grid(item1=letter,item2=number)
##   item1 item2
## 1     A     1
## 2     B     1
## 3     C     1
## 4     A     2
## 5     B     2
## 6     C     2
## 7     A     3
## 8     B     3
## 9     C     3
```

Basic Rules of Probability

Axioms

Let A be an event (one or more outcomes in a sample space S). For example, imagine picking a random integer between 1-100.

- S would be the integers 1-100 (set of all possible outcomes)
- A could be the event "the picked value is a 4".
- A could be the event "the picked value is an even number".
- A could be the event "the picked value is less than 57".

Axioms

- $0 \leq P(A) \leq 1$ If $P(A) = 0$, it is impossible for A to occur (e.g., P(obtain a sum of 14 on a roll of two dice)=0). If $P(A) = 1$ then A will occur with 100% certainty (e.g., P(sum of two dice will be between 0 and 20)=1). There are no negative probabilities.
- $P(S) = 1$ In other words, when an experiment is performed, an event from the sample space is guaranteed to occur.
- If A_1, A_2, ..., A_k are a collection of disjoint events (i.e., mutually exclusive, like A_1 is that a customer buys exactly 1 item, A_2 is that a customer buys 2 items, etc.) then

$$P(A_1 \text{ or } A_2 \text{ or } \ldots \text{ or } A_k) = P(A_1) + P(A_2) + \ldots + P(A_k)$$

Useful consequence: complement rule

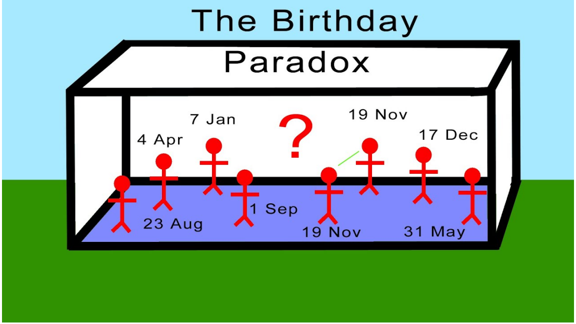
Let A be an event. Then A^c or A' is called the *complement* of A and means that "A does not occur".

$$P(A^c) = 1 - P(A)$$
$$P(A) = 1 - P(A^c)$$

- $P(green) = 0.20 \rightarrow P(not\ green) = 1 - 0.20 = 0.80$
- $P(pass) = 0.96 \rightarrow P(fail) = 1 - 0.96 = 0.04$

Complement rule illustration - Birthday Problem

A classic probability problem is the birthday problem, which asks the probability of the event "two (or more) people in the class share a birthday".



Complement rule illustration - Birthday Problem

It turns out that for this and for many problems it is easier to figure out the *complement* of the event and calculate accordingly.

If A is the event that at least two people in the room share a birthday, let us calculate $P(A^c)$, i.e., the probability that all birthdays in the class are unique. Let us ignore freak leap-year babies born Feb 29 and apply the fundamental theorem of counting to a class of 45 students.

$$P(\textit{all unique}) = \frac{\text{\# sequences where all birthdays unique}}{\text{\#number of possible sequences}}$$

Reminder: Fundamental theorem of counting (duplicate slide)

Imagine a trial consists of k experiments to be performed in order. Experiment 1 has n_1 outcomes, experiment 2 has n_2 outcomes, etc., and experiment k has n_k outcomes. Then there are $n_1 \times n_2 \times \ldots \times n_k$ ways the sequence of experiments can be performed.

Example: A restaurant has three choices of appetizer, four choices of main course, and two desserts. How many different meals can be constructed?

Solution: from the fundamental theorem of counting the number of meals is $3 \times 4 \times 2 = 24$. If we wanted to be more clever and allow for the possibility of "none of the above" for each course, then there would be $4 \times 5 \times 3 = 60$ meals (well, perhaps only 59 since "none of the above" for all three wouldn't constitute a meal).

Birthday Problem calculation

Denominator (#number of possible sequences)

- Each person has one of 365 birthdays (each equally likely).
- Total number of ways birthdays could turn out for a class of 45? This "trial" is like 45 sequential experiments: pick a birthday for the first person, pick a birthday for the second person, etc. Each has 365 possibilities, so it's $365 \times 365 \times 365 \times \ldots = 365^{45}$

Numerator (# sequences where all birthday unique)

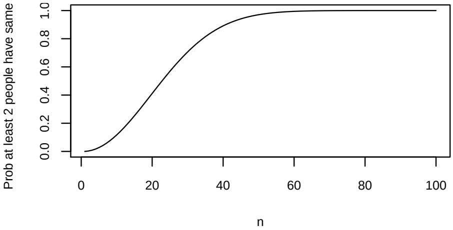
- This is a sequence of 45 experiments: Give 1st person a unique birthday (any of 365 choices), give 2nd person a unique birthday (any of 364 choices), etc.
- Fundamental theorem of counting says there are $365 \times 364 \times 363 \times \ldots \times 321$ ways for this to turn out.

Birthday Problem calculation II

$$P(all\ unique) = \frac{365 \times 364 \times 363 \times \ldots \times 321}{365^{45}} = 0.059$$

This means the probability that at not all birthdays will be unique in a class of 45 (at least two people will share a birthday) is $1 - 0.059$, or 94.1%!

The probability changes with n, the number of students in the class, as follows

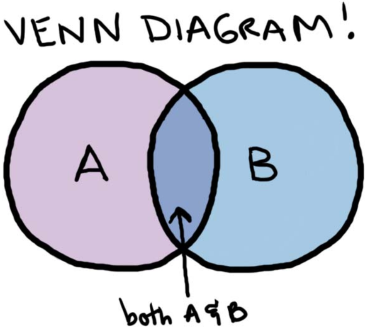


Addition Rule (A or B)

If A and B are two events and we want to know the probability of A or B occurring (i.e., at least one of the events A or B occur), then

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

The subtraction is so that events where both A and B occur are not double counted, as the Venn diagram below shows.



Addition Rule (A or B) example

A restaurant serves burgers, fries, and drinks. Some people buy food, some people just buy drinks. It is known that any particular order has a 70% chance of having a burger, 60% chance of having fries, and 45% chance of having both fries and a burger.

What is the probability that an order has a burger or fries (i.e., at least one of burger/fries)?

$$P(burger\ or\ fries) = P(burger) + P(fries) - P(both) = .7 + .6 - .45 = 0.85$$

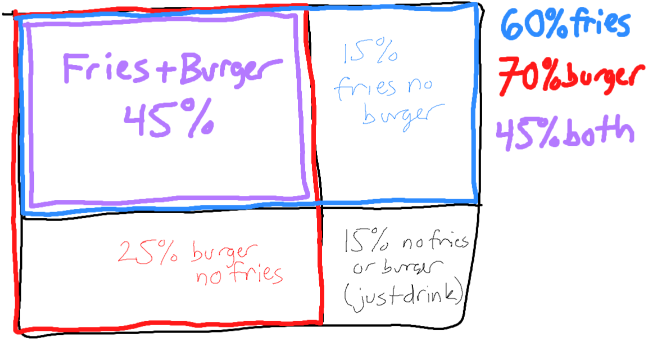
What is the probability that an order only has drinks?

"Just drinks" is the opposite of "at least one of burgers or fries", so this is $1 - 0.85 = 0.15$.

I think it's interesting that you can calculate the probability of only having drinks given information only on orders that have fries and/or burgers!

Addition Rule (A or B) example

Venn diagrams don't have to be circles. Blue rectangle = has fries. Red rectangle = has burger. Purple rectangle = has both.



Addition Rule (A or B) when A and B are mutually exclusive

If A and B are mutually exclusive (i.e., cannot occur simultaneously), then P(A and B) $= 0$ and $P(A \textit{ or } B) = P(A) + P(B)$.

Addition Rule (A or B) when A and B are mutually exclusive

Consider showing up at a stoplight. Only one outcome (green, red, yellow) is observed, so these events are mutually exclusive.

- If P(red) = 0.6, P(green)=0.35, P(yellow)=0.05, then what is the probability that we will get to pass through the intersection when we show up? A light can't be green and yellow at the same time so they are mutually exclusive.

$$P(green\ or\ yellow) = P(green) + P(yellow) = 0.35 + 0.05 = 0.40$$

- What is the probability of rolling an even number on a die? The numbers 1-6 occur with probability $1/6$, and only one number can occur so

$$P(even) = P(2\ or\ 4\ or\ 6) = P(2) + P(4) + P(6) = 1/6 + 1/6 + 1/6 = 50\%$$

.

Multiplication Rule (A and B)

If A and B are two events and we want to know the probability of both A and B occurring then

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

$P(A|B)$ is read as the probability of A given B, and is the probability that A will occur given that event B has occurred. Conditional probability is the subject of a future unit.

- P(brown hair) = 30%, but P(brown hair|brown eyes) = 70%.
- P(graduated HS with 4.0) = 0.05 but P(graduated HS with 4.0|admitted to Harvard) = 65%.

Multiplication Rule (A and B) for independent events

If A and B are independent, then the occurrence of one does not affect the probability that the other occurs, i.e. $\mathrm{P(A|B) = P(A)}$ and $\mathrm{P(B|A) = P(B)}$. In this case, the multiplication rule simplifies

$$P(A \textit{ and } B) = P(A) \times P(B) \qquad \text{A and B independent}$$

This extends to more than one event. If A_1, A_2, ..., A_k are independent events, then

$$P(A_1 \textit{ and } A_2 \textit{ and } \ldots \textit{ and } A_k) = P(A_1) \times P(A_2) \times \ldots \times P(A_k)$$

Example 1 (independent events)

What is the probability of rolling four 1's in a row on a die? The sequence 1, 2, 3, 4?

We could let A be the event "roll a one on the first roll", in which case P(A)=1/6. We could let B be the event "roll a one on the second roll", in which case P(B)=1/6. Since rolls are independent, if we follow this reasoning we get that P(four 1's in a row) $= (1/6)^4$. The same reasoning applies to the second sequence since each number individually has a 1/6 chance of being rolled.

Example 2 (independent events)

Imagine M & M colors have the following probabilities: red, green, blue = 20%; orange, yellow = 15%, brown 10%. Let the colors of sequential candies you draw out of a pack be **independent**.

- You pick 2 candies. What is the probability that both are orange? This is the probability of the event "orange first and orange second".
  $0.15 \times 0.15 = 0.0225$
- What is the probability of drawing 6 candies in rainbow order (brown last)? This the probability of the event "red first and orange second and yellow third and green fourth and blue fifth and brown sixth".
  $0.2 \times 0.15 \times 0.15 \times 0.2 \times 0.2 \times 0.1 = 0.000015$

Example 2 (independent events)

What is the probability of drawing *at least one* red out of 5 candies? Technically, this is the event "exactly 1 OR exactly 2 OR exactly 3 OR exactly 4 OR exactly 5", and that's a hard event to handle. Instead, let's calculate the *complement* of "at least one red", which would be the event "exactly 0 reds".

Exactly 0 reds is the event "not red first and not red second and . . .". The probability of a candy not being red is $1 - 0.2 = 0.8$. The probability of 5 not-reds in a row is thus $0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.8 = 0.8^5 = 0.32768$.

The probability of "at least one red", the complement of "exactly 0 reds" is thus $1 - 0.32768 = 0.67232$.

Example 2 (independent events)

What is the probability of drawing a red and a yellow?

There are two possible sequences that yield this event. 1st red and 2nd yellow or vice versa.

Red then yellow has probability $0.2 \times 0.15 = 0.03$; yellow then red is $0.15 \times 0.2 = 0.03$.

The probability of either of these sequences $P(RY \text{ or } YR)$, since they are mutually exclusive, is 0.06 by the addition rule.

Example 3 (non-independent events)

What is the probability of being dealt an ace and then a king from a deck of cards?

Let A be the event "1st card ace" and B be the event "second card king", then our task is to find P(A and B). A and B are not independent here (if an ace was dealt as the first card then the probability the second card is a king is higher than if the first card dealt was a king), so we use the general rule.

$$P(A \ and \ B) = P(B|A)P(A) = \frac{4}{51} \times \frac{4}{52}$$

This follows since if A has occurred then we know that the first card dealt was an ace. Of the 51 cards remaining, 4 are kinds so P(A|B)=4/51. The probability the first card is an ace is just 4/52 since there are 4 aces out of 52 cards.

Example 3 (birthday problem)

The multiplication rule extends to more than one event. For example:

$$P(A \text{ and } B \text{ and } C \text{ and } D) = P(D|A \text{ and } B \text{ and } C) \times P(A \text{ and } B \text{ and } C)$$

We can rewrite:

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

We can rewrite

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

So:

$$P(A \text{ and } B \text{ and } C \text{ and } D) = P(D|A \text{ and } B \text{ and } C) \times P(C|A \text{ and } B) \times P(B|A) \times P(A)$$

If we drop the *and* and just separate events that occur together with a comma, the formula becomes:

$$P(A, B, C, D) = P(D|A, B, C)P(C|A, B)P(B|A)P(A)$$

Example 3 (birthday problem)

Let

- A_1 be the event where person 1 does not share a birthday with anyone before him or her
- A_2 be the event where person 2 does not share a birthday with anyone before him or her
- A_k be the event where person k does not share a birthday with anyone before him or her

Then the probability that no one shares a birthday in a room of 45 people is

$$P(A_1 \textit{ and } A_2 \textit{ and } \ldots \textit{ and } A_k)$$

Example 3 (birthday problem) II

We can rewrite this as

$$P(A_k|A_1\,A_2\,\ldots\,A_{k_1}) \times P(A_{k-1}|A_1\,\ldots\,A_{k-2}) \times \cdots \times P(A_2|A_1) \times P(A_1)$$

It is easiest to work from the right side going back.

- $P(A_1) = 1$ since the first person doesn't share a birthday with anyone asked before
- $P(A_2|A_1) = 364/365$ since the second person can have 364 out of 365 days in order to have a unique birthday
- $P(A_3|A_2\,A_1)$ is the probability that the third person doesn't share a birthday with anyone asked before given that person 1 and 2 have unique birthdays. There are 363 available days out of 365, so this is 363/365

$$P(all\ unique) = \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \cdots \times \frac{321}{365} = 0.059$$