PROBABILITY

A BEGINNER'S GUIDE TO PERMUTATIONS & COMBINATIONS

SCOTT HARTSHORN

Probability – A Beginner's Guide To Permutations And Combinations

By Scott Hartshorn

Thank You!

Thank you for getting this book! This book gives examples of how to understand using permutations and combinations, which are a central part of many probability problems. That subject usually makes up a chapter in most statistics text books, but it is a chapter that doesn't do the subject its proper justice. Most chapters on this subject start and end with memorizing the permutation and combination equations, and miss the deeper understanding of them and also skip over the permutation and combination problems that can't be solved with those equations directly.

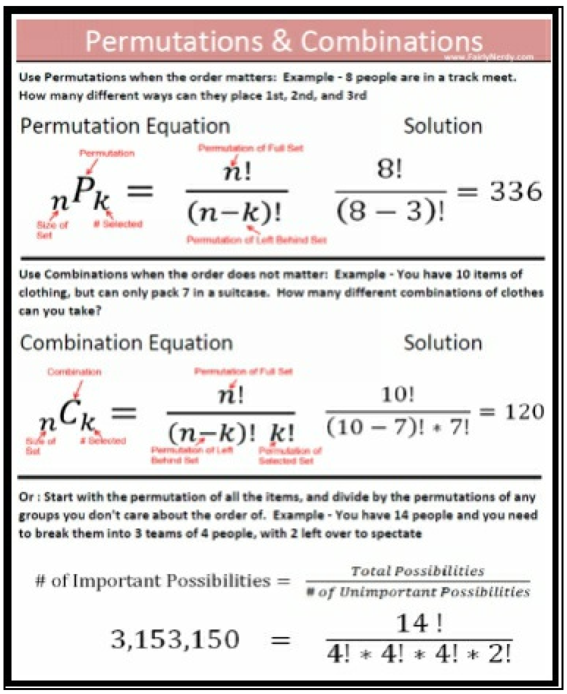
This book also devotes a large section to an example permutation problem of the kind that you might find in a programming challenge. Those problems are frequently in programming challenges because permutations are an easy way to ensure that naïve brute force solutions can't solve the problems in a reasonable amount of time, and that a more elegant understanding of the math is required.

If you want to help us produce more material like this, **then please leave a positive review on Amazon**. It really does make a difference!

Your Free Gift

As a way of saying thank you for your purchase, I'm offering this free permutations and combinations cheat sheet that's exclusive to my readers.

This FREE PDF cheat sheet has the permutation and combinations equations, an example for each of them, and an additional example to show how to expand those equations to handle situations that the basic equations don't cover.   This is a PDF document that I encourage you to print, save, and share. [You can download it by going here](http://www.fairlynerdy.com/permutations-and-combinations-cheat-sheet/)



http://www.fairlynerdy.com/permutations-and-combinations-cheat-sheet/

Permutations And Combinations Overview

Permutations and combinations are an essential part of statistics. They show up in a ton of different places when you are finding the probability of anything. But it can be hard to remember the exact formula for a permutation or a combination when you need it without looking it up.   This book will show you an easy, intuitive way to understand permutations and combinations so that you only need to remember one thing, and the rest you can just calculate when you need it.

This book goes through several permutation and combination problems that would be difficult to solve using just the permutation and combination equations without a deeper understanding. It also goes through one fairly challenging permutation problem of the kind that you might encounter during a programming challenge

Table of Contents

Combinations And Permutations – The Basics

The Basics

Permutations and combinations are a way of determining how many different possibilities of something there are.

Permutations are what you use when the order matters. For instance, if 8 people are racing in a track meet, and you want to find the different ways they could get 1^{st}, 2^{nd}, and 3^{rd} place, then the order matters. So you would use a permutation.

Combinations are what you use when the order doesn't matter. For instance, if you have 10 different pieces of clothing you want to take on a trip, but you can only fit 7 of them in your suitcase, it matters which 7 you pick, but it doesn't matter what order you put them in the suitcase, so you would use a combination

The Key Mathematical Symbol

There is one key thing to know with Permutations and Combinations, and that is the Factorial. Typically denoted with an exclamation point !

If you want to find how many different ways you can arrange 8 different items, it is 8 Factorial, which is 8 * 7 * 6 * 5 * 4 * 3 * 2 * 1. What this represents is that when you make your first choice of items to arrange, you have 8 to choose from. When you make your second choice, there is one less, so you then have 7 to choose from, then 6 and so on.

Everything with permutations and combinations are just different applications of the Factorial.

Example 1 – Straight Forward Permutation & Combination Problems

Permutations – Slightly Simpler Than Combinations

Let’s go back to the track example. Let’s say that you have 8 people racing on the track. The total different orders they could come in are 8 !

8 * 7 * 6 * 5 * 4 * 3 * 2 * 1 = 40320

Now let’s say you only care about the order of the first three people on the track. Clearly we would have fewer than the full 8! different permutations, because we only care about how the first 3 people finished, not all 8. So in that case you have 8 options for first place, 7 options for 2^{nd} place, and 6 options for 3^{rd} place, and that’s it.

8 * 7 * 6 = 336

This is the permutations of 8, choosing 3.

Now there isn’t a function that lets us just multiply 8 * 7 * 6 easily. If we wanted the order of everyone, then 8 Factorial lets us multiply 8, down through 1, but it doesn’t stop in the middle. The way we do this is by finding 8 Factorial, and then dividing by 5 factorial. We use 5! because there are 5 items left behind that we don’t care about (8-3 = 5) That ends up being

$$\frac{\overbrace{8*7*6*5*4*3*2*1}^{\text{Permutations of Full Set}}}{\underbrace{5*4*3*2*1}_{\text{Permutations of Left Behind Set}}}$$

The

(5 * 4 * 3 * 2 * 1)

Cancels out of both the numerator and denominator, and we are left with 8 * 7 * 6.

So to find the permutations of a subset of a group, what we have just done is

- Find the permutations of the entire group (8! in this case)
- Divide by the permutations of the part of the group left behind (5! In this case)

Why are we dividing by the permutations of the parts left behind ? Because we don't care what order they are in, so we need to cancel out all different orderings that they can be in.

This is a key point to remember

- If you want to find the permutations of something, use the factorial
- If you want to find the permutations of a subset, find the permutations of the entire group, and then divide by the permutations of the set left behind.

Combinations – Build on Permutations

Combinations simply start as permutations with a subset and add 1 more step.

- Permutations – We take the factorial of the entire set to find the number of possibilities
- Permutations with a subset – We take the factorial of the entire set, and then divide by the factorial of the left behind set

Well for combinations we still don't care about the order of the left behind set, but we also don't care about the order of the set that we have chosen. So we start with the permutations of the entire set, then divide by the permutations of the left behind set, then divide by the permutations of the chosen set.

So if we have 10 different items of clothing, and we can only choose 7 to pack, so there are 3 left behind, the number of possibilities are

$$\frac{\overbrace{10*9*8*7*6*5*4*3*2*1}^{\text{Permutations of All Items}}}{\underbrace{(3*2*1)}_{\text{Permutations of Left Behind}}*\underbrace{(7*6*5*4*3*2*1)}_{\text{Permutations of Selected Items}}}$$

Which is equal to 120

An important thing to know is that there will always be at least as many permutations of a set as combinations, and typically many more permutations than combinations.

The Traditional Permutations & Combinations Equations

At this point, it is worth showing the traditional permutations and combinations equations. These are the things that you might typically be expected to memorize for a class, but can be challenging to remember long term

Here is the permutation equation

$$\underset{\text{Size of Set}}{{}_{n}}\overset{\text{Permutation}}{P}_{\underset{\text{\# Selected}}{k}} = \frac{\overset{\text{Permutation of Full Set}}{n!}}{\underset{\text{Permutation of Left Behind Set}}{(n-k)!}}$$

And here is the combination equation

$$\underset{\text{Size of Set}}{{}_{n}}\overset{\text{Combination}}{C}_{\underset{\text{\# Selected}}{k}} = \frac{\overset{\text{Permutation of Full Set}}{n!}}{\underset{\text{Permutation of Left Behind Set}}{(n-k)!}\;\underset{\text{Permutation of Selected Set}}{k!}}$$

So while it is manageable to memorize those equations, it is easier to just intuitively understand

- To find the permutations of a full set, take the factorial
- To find the permutations of a partial set, find the permutations of the full set, then divide by the permutations of the items left behind

- To find the combinations of a partial set, find the permutations of the full set, divide by the permutations of the items left behind, and divide by the permutations of the selected items.

Example 2 - Taking It One Step Farther

The examples above were straight forward applications of the combination and permutation equations. However it turns out that there are a lot of problems where it is not easy to apply those equations directly. If you understand using the factorial, instead of memorizing the equations, you can apply it to cases that the equations don't cover. Take this example, say you have a group of 10 people that are traveling to a destination and you find yourself with 4 vehicles that can take them there. Those 4 vehicles can seat 1 person, 2 people, 3 people, and 4 people respectively, so all 10 people can go.

How many different ways can you organize the groups, assuming that people care about which vehicle they are going in, but they don't care about where they are sitting within a vehicle?

The first thing to recognize is that this seems like a combination problem. The fact that the people don't care where they are sitting means that order does not matter within a particular group. However the combination equation

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

is only made for 1 group. If our problem had been that we had 10 people, 4 of which were traveling somewhere and the other 6 were left behind, we could use the combination equation to get the answer. But with the multiple groups this would not work.

So let's think about what is happening. This is still pretty much a combination problem, but we need to do it slightly differently than the classic combination problem. For every one of these combination or permutation problems we start with the total number of possible permutations that there can be, and then divide out number of possibilities that we do not care about

$$ \#\ \text{of Important Possibilities} = \frac{Total\ Possibilities}{\#\ of\ Unimportant\ Possibilities} $$

To start with, what is the total number of different permutations of the group ? It is 10!

That means, if we cared about the order of everyone in the group, we would have 10! Different combinations. But we know we don't care about everyone's order, because we don't care about the order within a specific vehicle, so we know that we will have fewer than 10!

So how many different orders of people do we not care about ? To put it a different way, assume that we split the 10 people among the 4 vehicles, so now we know which car everyone will go in. How many different ways could they sit within the vehicles ?

- In the largest vehicle, the 4 people could sit 4 ! different ways
- In the next vehicle, the 3 people could sit 3 ! different ways
- In the next vehicle, the 2 people could sit 2 ! different ways
- And in the last vehicle, the 1 person could sit 1 ! different way

So for any given different groups of people traveling, after you determine what vehicle everyone is traveling in, there are 4! * 3! * 2! * 1! different ways that they can be seated

$$\text{\# of Groups of People} = \frac{\#\, Total\ Seating\ Arrangements}{\#\, Seating\ Arrangments\ Within\ The\ Cars}$$

We know that the total number of different orders is 10 !   and the different seating arrangements within the cars are 4!, 3!, 2!, 1!

This means the total different ways the people can be split between vehicles is

$$\text{\# of Groups of People} = \frac{10\,!}{4! * 3! * 2! * 1!} = 12600$$

Example 3: Mixing Permutations & Combinations

The last problem expanded the combination equation to use multiple groups. This one shows using both combination and permutation equations in a single problem.

For this problem let's keep the same 4 vehicles

- 1 Seat motorcycle
- 2 seat sidecar
- 3 seat golf cart
- 4 seat car

and we still have 10 people that need to go somewhere, however 5 of them are children and 5 of them are adults. Only an adult can be the driver of one of the vehicles.

How many different arrangements are there, assuming that for the passengers it only matters which vehicle they are in, but not where they sit in that vehicle (just like the previous problem) but that it does matter who is driving each vehicle?

To solve this problem, we will need to use both permutations and combinations. We have to split the problem into two parts because there are people who can't be in the driver's seat. So it isn't applicable to assume that anyone could be anywhere to start with

$$\text{Total Arrangements} = \#\ \text{Driver Arrangements} * \#\ \text{Passenger Arrangements}$$

The number of driver arrangements is a straight forward permutation problem. We have 5 people, and 4 seats. We care who is in what seat, and we will have one person left behind where we don't care about their order. The resulting number of driver arrangements is

$$\#\ \text{of Driver Arrangements} = \frac{5\,!\ \text{(total adults)}}{1!\ \text{(not a driver)}}$$

For the passengers we have 6 possible people, because we have the 5 children plus which ever adult isn't driving. This would mean that there are 6! different arrangements of passengers, however some of those arrangements aren't unique because they are just passengers switching seats within the same vehicle, which we don't care about.

To calculate the total number of unique passenger arrangements we recognize that the 4, 3, 2, & 1 seat vehicles will have 3, 2, 1, & 0 passengers in them respectively. So the

- 4 seat car could have 3! different arrangements of passengers in it
- 3 seat golf cart would have 2! different arrangements
- 2 seat sidecar would have 1! different arrangements
- 1 seat motorcycle would have 0! different arrangements

So the total number of unique passenger arrangements

$$\#\ \text{of Passenger Arrangements} = \frac{6\,!}{3! * 2! * 1! * 0\,!}$$

$$\#\ \text{of Passenger Arrangements} = \frac{6\,!}{6 * 2 * 1 * 1} = 60$$

When multiplied by the number of driver arrangements

$$\text{Total Arrangements} = \frac{5\,!}{1!} * \frac{6\,!}{3! * 2! * 1! * 0\,!} = 7200$$

A Side Note On Whole Numbers

One interesting thing to note in all these problems is that we always end up with a whole number. This is because we are asking about real life events where you can't have a fractional number of different arrangements. For instance, in the last problem you could physically make the people sit in each of the different arrangements and count them (assuming you could get 5 adults and 5 kids to put up with 7200 different seating arrangements) and there would be no way to do only part of an arrangement and have it still be legal.

More mathematically, what is occurring is that the factorials on the numerator and denominator of the permutation or combination equations are always canceling.

For instance if I had 13 things, and wanted a combination of 7 of them the equation would be

$$\text{Total Combinations} = \frac{13\,!}{6! * 7!} = 1716$$

The result of 1716 is a whole number, with all of the denominator being completely canceled even though this problem uses the prime numbers of 13 & 7.

The 13! & 7! will cancel out in a very straightforward manner

$$\text{Total Combinations} = \frac{13 * 12 * 11 * 10 * 9 * 8 \boxed{\cancel{* 7 * 6 * 5 * 4 * 3 * 2 * 1}}}{6! * \boxed{\cancel{7 * 6 * 5 * 4 * 3 * 2 * 1}}}$$

Which leaves you left with

$$\text{Total Combinations} = \frac{13 * 12 * 11 * 10 * 9 * 8}{6 * 5 * 4 * 3 * 2 * 1}$$

And the numbers in the numerator will cancel the denominator out. You could combine them different ways, but the 12 has a 6 factor, the 10 has a 5 factor, the 9 has a 3 factor, and 8 has a 4 & a 2 factor.

As it turns out, this always works and as a result the permutation or combination equations will always give a whole number

Example 4: The Lottery

The Powerball is one of the largest lotteries in the Unites States, and combinations are key to figuring out the odds of winning. This lottery is played as follows

- There are 69 balls in a drum, each with a unique number
- 5 of those balls are selected. The order of those balls do no matter
- There are 26 balls in a separate drum, each with a unique number
- 1 ball is selected from those 26. This is the Powerball.   It is possible, but not required, that the Powerball number could match 1 of the other 5 balls chosen because they are selected from different sources.

What are the odds of a single ticket matching all the drawn numbers and winning the lottery?

The solution to this is a simple application of the combination equation.

For the first drum, there are 69 options, chose 5. This means there are

- 69! total permutations of balls divided by
- 5! which are the permutations of the selected balls, the order of which doesn't matter. The result also divided by
- 64! which is the permutations of the left behind balls, the order of which doesn't matter

So the total number of possible combinations of balls from the first drum are

$$= \frac{69!}{5! * 64!} = 11{,}238{,}513$$

Additionally 1 ball is chosen out of 26 from the second drum. The odds of getting all of the balls correct are

$$= \frac{69!}{5! * 64!} * 26 = 292{,}201{,}338$$

So the odds of winning the jackpot are slightly worse than 1 in 292 million. Each ticket costs $2, so if you ignored all other factors a lottery with a jackpot of more than 584 million could be a statistically good value.

Interestingly there have been a few Jackpots larger than 584 million. In fact the largest Jackpot at the time of writing was ~1.6 billion. It is times like those you start to see articles about the profitability of buying every possible ticket. Leaving aside the impossible capital requirements and logistics for an individual or small group to buy 292 million pieces of paper (which, if the tickets are 3 x 5 x .01 inches would be ~25,000 cubic feet, or approximately 30 Uhaul's worth) the claimed profitability of buying all possible numbers ignores

- Shared Jackpots
- Taxes

- Knockdowns for taking a lump sum

To be fair though, I have ignored the lesser prizes that can be won by matching just some of the numbers on the tickets, which improves the odds a little bit.

The final conclusion? The lottery is organized intentionally with odds that make it won every so often in a nation of 300 million people. It is intentionally infrequent enough to yield large payouts, but the odds are small enough that it does get won by someone somewhere at some point. Probably not by you however, and certainly not by me

Just How Big Do Permutations Get?

One interesting thing about permutation problems is that as you increase the samples, the number of possible permutations gets so large, so fast, that it is not possible to analyze all of them. What this book is focused on is how to calculate the number of possible permutations, but what if you instead wanted to list or analyze all of the possibilities?

There are abundant examples where it would be useful to examine every permutation, if that were possible. One classic example is the traveling salesperson problem, which is as follows:

"You have a list of cities you need to visit, and a travel distance between every city pair. Find the route which visits every city exactly one time in the shortest possible distance."

This problem has widespread applications, not just for salesman, but for shipping companies, data routing companies, even machines which drill holes or which test circuit boards. At first glance it seems the problem seems simple, just make a list of all possible orders of cities and their distances. After all, computers are fast and getting faster all the time. You don't need a fancy algorithm, just analyze everything to find the best solution.

But it is here that the sheer magnitude of permutation problems becomes apparent. Let's say you wanted to calculate the shortest route between 48 points, for instance the 48 state capitals of the lower 48 states in the United States. You have a modern desk top computer, a C++ compiler, and can calculate the total distance of 100 million routes per second.

What happens?

You start your code with a simple data set, only 5 cities. There are 5! = 120 routes that you can take to visit each of those cities one time (although half of those routes are just the reverse direction of a different route) and your computer finds the shortest path of those 120 routes faster than you can measure, certainly less than one second.

Since the code worked for 5 cities, you bump it up to 10 cities. 10! = 3,628,800 different possible routes. So this calculation still takes less than one second.

At 12 cities however you start to notice a slowdown. 12! = 479,001,600

Since your machine can do 100 million routes per second, this takes approximately 5 seconds, or just enough to notice. 13! Is just over 6 billion, which takes your computer over a minute. By the time you get to 15 cities, it takes over 3 and a half hours to run the analysis on your computer.

Clearly your home computer doesn't have enough processing power, so it is you rent some time on the cloud. Even if you rent 1,000 times the computing power of your personal machine, it still takes almost a day to find the shortest route through 18 cities, and almost a whole year to find the shortest route through 20 cities.

At this point it is time to give up. There isn't anywhere close to enough computing power in the world to go through all the permutations of 48! That is 45 orders of magnitude greater than 18!, which took almost a day to run through on the expensive cloud server.

So what's the point of this? Simply to demonstrate how big permutations can get. If you have a problem to solve, anything larger than ~12!, it becomes challenging or impossible to iterate through all the permutations, so you would need a more efficient algorithm, or be willing to accept an approximation using a heuristic algorithm.

Example 5: Combinations Applied To Poker

Previously we saw the impracticality of iterating through every permutation of even smallish problems. Fortunately, there are problems of interest that don't require you to evaluate every possible outcome. Simply knowing how many outcomes are possible is useful information. Take for instance Texas Hold'em, which is a variant of poker.

In this game every player is dealt two cards, which are hidden and for their use alone. Then there are 3 cards dealt in the center of the table that all players share, then another single shared card is dealt in the center of the table, and then a final shared card is dealt to the center of the table, for a total of 5 shared cards over 3 separate rounds.

After you have been dealt your two cards, but before any of the shared cards are dealt, how many different variations of cards can be dealt?   For this assume that it doesn't matter what order the three cards are dealt in the first round of community cards (known as the flop), but it does matter which of the 3 rounds the card is dealt in. (round 4 is the turn, round 5 is the river)

To start the problem, you have two cards known to you in your hand. That means that there are 50 cards left to choose from.   For the flop, you need to choose 3 cards.   This is a combination problem, so the number of possibilities is 50 choose 3.   After you have chosen those 3, there are 47 cards left. You pick 1 out of 47 for the turn. Finally, you pick 1 out the remaining 46 for the river.

The final equation ends up being

$$= \frac{\overset{\text{Starting Deck}}{50!}}{\underset{\text{Flop}}{3!} * \underset{\text{Left behind}}{47!}} * \overset{\text{Turn}}{47} * \overset{\text{River}}{46} = 42{,}375{,}200$$

So the total number of different ways the hand could play out is just over 42 million.   That calculation might not actually be that useful. A different question is, "You have a starting hand with 2 hearts. What are the odds that you will get a flush by the end of the hand ?"   This is important because a flush is a powerful hand. You get a flush if there are at least 3 more hearts dealt on the board of the 5 community cards.

Here let's assume that you don't care what order the cards come out in. You only care about if you end up with a flush at the end or not.   That means that from your point of view, the total number of possible hands is 50 choose 5, which is different from the result above. The solution of the combination equation of 50 choose 5 is

$$= \frac{\overset{\text{Starting Deck}}{50!}}{\underset{\text{Community Cards}}{5!} * \underset{\text{Left Behind Cards}}{45!}} = 2,118,760$$

So there are 2,118,760 total hands which could be dealt. How many of those give you a flush?   To solve this, we need to separately solve for the cases where there are 3 hearts dealt on the board, and 4 hearts dealt on the board, and 5 hearts dealt on the board.

Each suit has 13 cards. You have 2 hearts in your hand, which means there are 11 hearts still in the deck. There are also 39 cards of other suits still in the deck.

- 3 Heart Scenario: There are 3 hearts on the board out of 11 left in the deck. There are 2 non-hearts on the board out of 39.   The total number of possible hands is 11 choose 3 * 39 choose 2
- 4 Heart Scenario: There are 4 hearts on the board out of 11 left in the deck. There is 1 non-heart on the board out of 39.   The total number of possible hands is 11 choose 4 * 39 choose 1
- 5 Heart Scenario: There are 5 hearts on the board out of 11 left in the deck. There are 0 non-hearts on the board out of 39.   The total number of possible hands is 11 choose 5 * 39 choose 0

The equations and results for all of those are

$$\text{3 Hearts} \quad = \frac{11!}{3! * 8!} * \frac{39!}{2! * 37!} = \quad 122{,}265$$

$$\text{4 Hearts} \quad + \frac{11!}{4! * 7!} * \frac{39!}{1! * 38!} = + \quad 12{,}870$$

$$\text{5 Hearts} \quad + \frac{11!}{5! * 6!} * \frac{39!}{0! * 39!} = + \quad 462$$

$$\text{Total Hands} \quad = \quad 136{,}597$$

So there are 135,597 total hands that could come that would give you a flush. That is out of possible 2,118,760 total hands. That means the odds of getting a flush with two suited starting cards are 135,597 / 2,118,760 = .064 or approximately 6.4 %

Example 6: Urn Problems

Urn problems are a classic permutation / combination type of problem. They involve drawing items out of an urn or bag. Those items are distinguished in different ways, often by color. A simple urn problem is:

You have an urn with 3 different colors of balls, red, green, and blue. There is an unlimited quantity of balls of each of the 3 colors. You draw out 5 balls. How many different arrangements of colors can you have? Assume that the order of the balls matters.

The answer to this problem is straight forwards, and doesn't use the permutation or combination equations that we have used before. The key differentiator that informs us not to use those equations was the unlimited quantity of balls. Essentially this is the same as using a random number generator to come up with 1 of 3 numbers for each of the 5 trials.

Since for each of the 5 trials there are 3 possible options, the number of different outcomes is

$$= \ 3 * 3 * 3 * 3 * 3 = 243$$

Urn problems can be set up to be more complicated. For instance, imagine you have an urn with balls of 3 different colors. In this urn you have 10 red balls, 9 green balls, and 8 blue balls. If you draw 7 balls randomly from the urn, what percentage of the time will you draw exactly 4 red balls?

These type of problems, where a percentage is being asked, almost always require two distinct steps. In the first step you have to calculate the total number of possible outcomes. In the second step you have to calculate the number of outcomes which satisfy the imposed criteria, which in this case is that there are exactly 4 red balls. Dividing the second step by the first step gives a percentage.

Solution: Step 1 Total Outcomes

As always, it is important to determine if the problem needs permutations or combinations. Since in this case it was requested a “total of 4 red balls” and no mention was made on what order they would be drawn in it is clear that combinations are important. So how many possible combinations of balls can be drawn?

There are 27 balls in total, and we are drawing 7 of them. So this is 27 choose 7

$$= \frac{\overset{\text{Total Balls}}{27!}}{\underset{\text{Chosen}}{7!} * \underset{\text{Left Behind}}{(27-7)!}} = 888{,}030$$

There ends up being 888,030 different ways to choose the 7 balls.

Step 2: Desired Outcomes

The second part of this is a little more complicated. How many outcomes have exactly 4 red balls?   The trick here is to break this into two smaller sub-problems. The first sub-problem is, of the 10 red balls, how many different ways can 4 balls be chosen.   That is the combination formula with 10 choose 4. For the second half of the problem, of the 7 choses balls 3 of them are not red. So this part is, of the 17 green and blue balls, how many ways can 3 balls be chosen. This is the combination formula with 17 choose 3.

Those two sub-problems are independent of each-other, and for any given outcome on one of the sub-problems, the other one can have its full range of outcomes. That means the two sub-problems need to be multiplied together to get the total number of possible outcomes

The resulting equations are

$$= \frac{\overset{\text{Red Balls}}{10!}}{\underset{\text{4 Red Balls Chosen}}{4!} * \underset{\text{6 Left Behind}}{6!}} * \frac{\overset{\text{Blue + Green Balls}}{17!}}{\underset{\text{3 Chosen}}{3!} * \underset{\text{14 Left Behind}}{14!}} = \quad 142{,}800$$

So there are 142,800 different ways that exactly 4 red balls can be chosen when 7 total balls are selected. This means that the odds of getting exactly 4 red balls are 142,800 / 888,030 = .1608 or approximately 16.1%

More Uses For Combination Equation – Binomial Theorem / Pascal's Triangle

One interesting use for the combination equation is to identify outcomes relating the binomial theorem. The binomial theorem deals with the likelihood of discrete events. For instance, if you wanted to answer the question: "If you flipped a coin 10 times, what is the probability that you would get heads 4 times" the way to solve that would be to use the binomial theorem.

The binomial theorem is a topic in its own right, and will only be briefly touched on here. It is mentioned in this book because the binomial theorem incorporates the combination equation into it.

The way to solve the problem of "What is the probability of getting 4 heads in 10 flips" is to

- First calculate how many possible outcomes there could be from flipping a coin 10 times
- Calculate how many of those outcomes result in 4 heads

The number of possible outcomes is simple. There are 2 discrete outcomes for each event, heads or tails. If you choose between those two outcomes 10 times, you end up with 2^10 possible chains of outcomes. 2 raised to the 10^{th} power is 1024

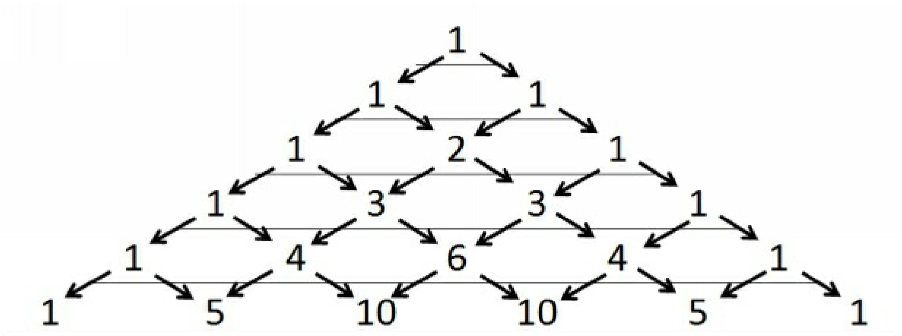
The number of outcomes that will be heads turns out to be the same as 10 choose 4

$$= \frac{10!}{4! * 6!} = 210$$

So there are 210 possible scenarios to get 4 heads, out of a total of 1024 possibilities. That means that you will get 4 heads 210 / 1024 = .2051 or 20.5% of the time.

Pascal’s Triangle

Pascal’s triangle is one way of expressing results for the binomial theorem. Every value in the triangle is the sum of the two numbers above it, i.e. the ones with arrows pointing from them in this picture.



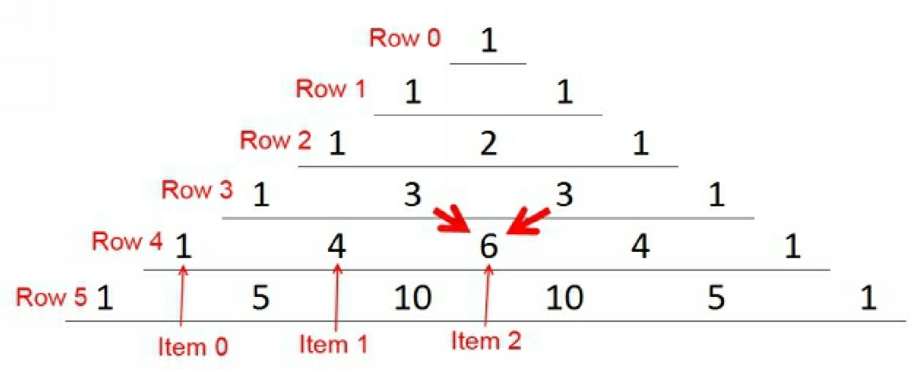
Every row of Pascal’s triangle represents 1 more discrete event relative to the row above it. For instance the top row represents flipping the coin zero times. The only possible outcome from flipping a coin zero times is to get zero heads.

The next row, 1,1 represents flipping a coin 1 time. The first 1 is the outcome of getting zero heads. The second 1 is the outcome from getting 1 head. Note that there are 2 possible outcomes from the 1 flip.

The 4th row with the numbers 1, 3, 3, 1 represents the possible outcomes from flipping a coin 3 times. There are 8 possible permutations of outcomes. In total if you flip the coin 3 times you will get zero heads 1 time, 1 head 3 times, 2 heads 3 times, and 3 heads 1 time.

How this relates to combinations and permutations is that any given value in Pascal’s triangle can be calculated using the combination formula. The row number is the number chosen from, and how many from the left a value is, is the number of items chosen. Remember to count both rows and items starting with zero.

As an example, this value of 6 is item number 2 in row number 4



That means it can be calculated using the combination formula of 4 choose 2.

$$= \frac{4!}{2! * 2!} = 6$$

Which in fact it can. The usefulness of using the combination formula to calculate results in the Binomial Theorem is that can save a lot of calculation compared to Pascal’s triangle. For instance, if you had to find the odds of getting 50 coins in 100 flips, it is a single equation using the combination equation, but would be 101 rows of numbers using Pascal’s triangle.

Example 7: A Programming Permutation Problem

The reason I wrote this book is that I was doing a programming challenge for a large tech company and there were a lot of permutation & combination type programming problems that highlighted the inadequacy of simply relying on the permutation & combination equations. There was a need for a deeper understanding.

This is an example of that kind of programming problem.

Say you have 5 people each with a different height. No two people have the same height. They all stand in a straight line and you stand directly in front of them looking at them.

You can see anyone that is taller than all of the people in front of them, but not anyone that has a taller person standing anywhere in front of them. So if they stood in ascending height order with the shortest person in front you could see everyone. If the tallest person stood in the front you could only see that one person regardless of the order of the people behind him.

There are 5 people in the line, and you can see 3 of them. How many different arrangements of those 5 people are possible that would have exactly 3 people visible to you ?

To begin with, this is a problem that is meant to be solved by writing some computer code. We can and will solve the 5 people with 3 visible manually, but getting too many more people than that would get tedious. Something like 10 people with 6 visible would be fairly hard to do manually due to requiring a fair bit of calculation. (Although that quantity is still within the range of solving in a spreadsheet like Excel with the proper equations set up. The Excel solution & Python solutions to this problem can be downloaded for free here http://www.fairlynerdy.com/permutations-example-problem/)

So why do it manually at all ? Why not just use the computer program ?

The reason is that it is worth understanding how the problem works. Without that understanding you can easily set up your computer code in a naïve way that makes it unable to solve the larger problems either. In fact, if you hit these kind of problems in a programming challenge you can pretty much guarantee that they have inputs that are large enough that the naïve brute force methods will be unable to solve them, and that you will need to have an understanding of the math to program it correctly.

If you were programming a computer to solve this problem of 5 people with 3 visible, the naïve brute force way would be to just check all 5! = 120 arrangements of people to see if 3 are visible. Checking 120 arrangements would be very tedious to do manually, and certainly you wouldn't want to read a bunch of pages of this book trying to do that, however it would be well within the capability of a computer.

However permutations grow so quickly that you usually can't try them all for anything but the smallest numbers. For instance, with 10 people, 6 visible there are 10! = ~3.6 million arrangements, which is still solve-able on a regular desktop computer but will probably take a measureable second or fraction of one. But solving 20 people 10 visible, there are 20! = $\sim 2.4 * 10^{18}$ arrangements, which would break any computer in the world to try to look through all the permutations.

Fortunately, we don't need to examine every permutation. We just need to calculate the number that satisfy our criteria for how many people are visible. We can do that more intelligently.

Back to the 5 people with 3 visible problem

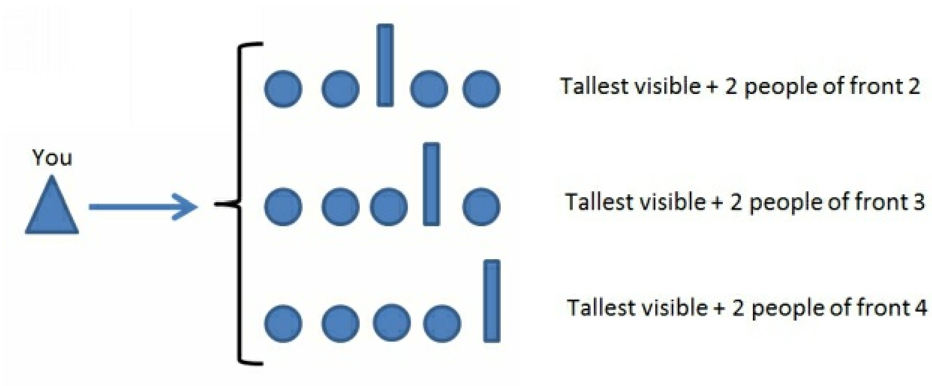
The key insight here is that you can't see over the head of the tallest person.

So you can solve this problem by finding where the tallest person is and then how many different arrangements are in front of that person multiplied by the number of arrangements behind them.

So to start with, where can the tallest person be ?

There are 5 people, and 3 visible, so the tallest person must be at least 3 back because if they were any closer to the front, 3 people couldn't be visible.

That means there are 3 spots the tallest person could be in, position 3, 4, or 5

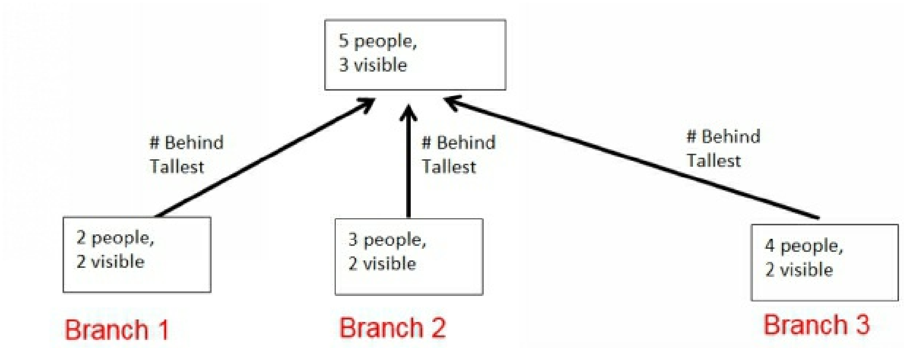


We are going to have to find the total possibilities for each of these cases and add them.

$$\begin{aligned}\text{Total Arrangements} &= \#\text{ when tallest is at position } 3\\ &+ \#\text{ when tallest is at position } 4\\ &+ \#\text{ when tallest is at position } 5\end{aligned}$$

In all of those cases, there are 2 people visible in front of the tallest person, and the only difference is how many other people are in front of the tallest, vs how many people are behind the tallest person.

One way to keep track of this is to make a tree. We want to solve for 5 people, 3 visible, and to do that we need to develop a function that includes 2 people - 2 visible, 3 people – 2 visible, and 4 people – 2 visible



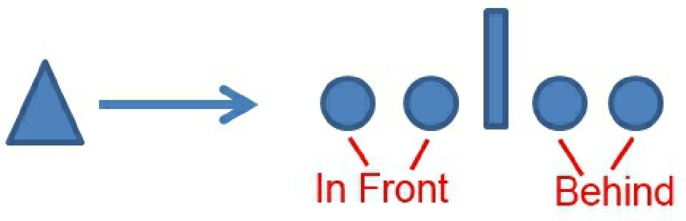
Another way of visualizing the problem, and one that is probably ultimately more useful, is to use a table

# of people					
5			Goal		
4		Branch 3			
3		Branch 2			
2		Branch 1			
1					
# visible	1	2	3	4	5

Solution For Main Branch 1

Let's start with solving for when the tallest person is at position 3. This turns out to be slightly easier than the other two positions as we will see in a minute

At position 3



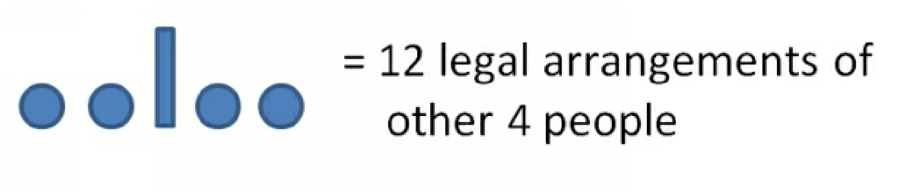
$$\text{Total Arrangements} = \#\ \text{Arrangements in Front} * \#\ \text{Arrangements Behind}$$

The possibilities behind is a simple permutation problem.   Using the classic permutation equation, 4 people, pick 2 (behind)

$$= \frac{4!}{(4-2)!} = 12$$

The number of possibilities for people in front is more complicated. It is not just a classic permutation or combination problem, because we need to take into account the fact that the exact amount of people need to be visible. However here is where this branch of the problem is the easiest. We have 2 people and we know that there have to be 2 people visible. So there is only 1 possible arrangement for them, the shorter person in the front and the taller person in the back.

That means that there are 12 possible arrangements behind the tallest person, and 1 possible arrangement in front.   12 * 1 = 12, so when the tallest person is in the 3rd position, there are 12 legal arrangements of the other people



Let’s step back from the problem to fill in some of the table and the tree. This will help us to keep sense of where we are in the problem, although it isn’t strictly necessary for a problem of the size of this 5 people, 3 visible.

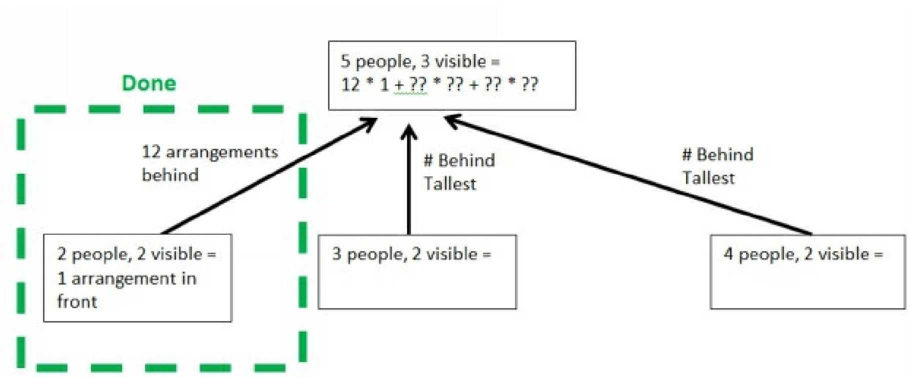
We just calculated that if there are 2 people, and 2 visible there is only 1 arrangement that works, so we can fill in that square

# of people					
5			Goal		
4		Branch 3			
3		Branch 2			
2		1			
1					
# visible	1	2	3	4	5

In fact, if there are the same number of people as are visible, there is always only 1 possible arrangement, so we can fill in the whole diagonal. Furthermore, there can never be more visible than there are people, so those squares can be grayed out

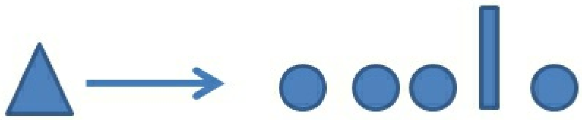
# of people					
5			Goal		1
4		Branch 3		1	
3		Branch 2	1		
2		1			
1	1				
# visible	1	2	3	4	5

For the tree, we have completed 1 out of the 3 branches



Solution For Main Branch 2

Now let’s do the part of the problem where the tallest is in the 4^{th} position.

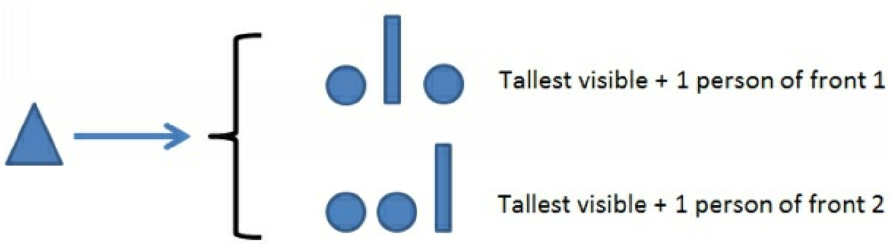


Once again, the total number of possibilities is the number of arrangements behind the tallest person, multiplied by the number in front of the tallest person.

$$\#\text{ of arrangements behind} = \frac{4\,!\ \ (people\ remaining)}{3\,!\ \ (people\ in\ front)} = 4$$

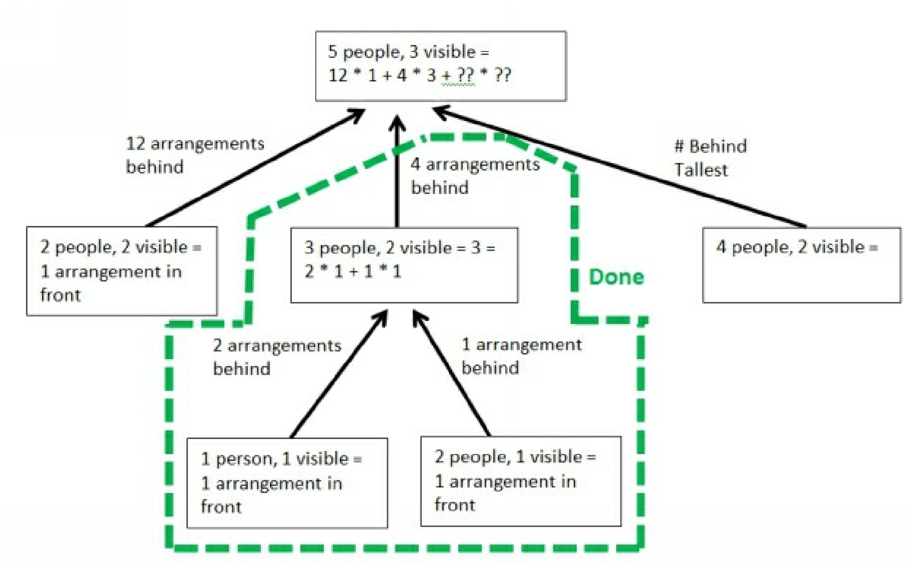
$$\#\text{ of arrangements in front} = \ \text{Solution with 3 people, 2 visible}$$

However this 3 people, 2 visible isn’t an obvious answer. So we need to break this down and solve it. Of these 3 people, the tallest can either be in position 2 or position 3. If the tallest was at the front, there couldn’t be 2 people visible



If the tallest is in position 2, there are 2 solutions because you can switch the people. If the tallest is in position 3, there is 1 solution because of the remaining 2 people, only 1 can be visible.

So 3 people, 2 visible has a total of 3 solutions. Visually, the updated tree is below.



We can also fill out some other parts of the table that we have solved

# of people					
5			Goal		1
4		Branch 3		1	
3		3	1		
2	1	1			
1	1				
# visible	1	2	3	4	5

Solution For Main Branch 3

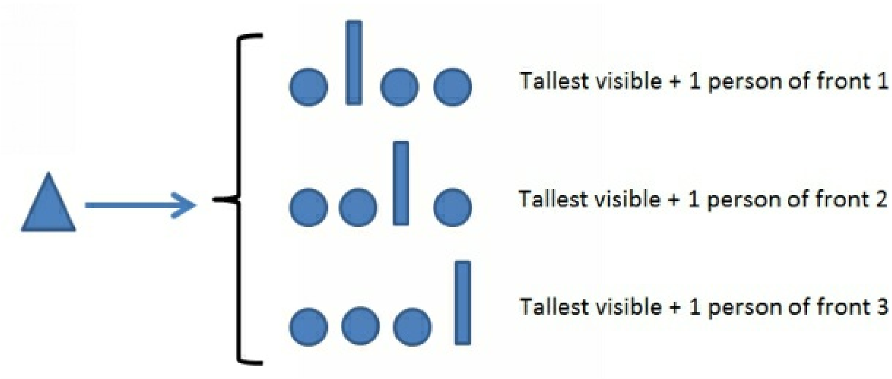
Now for the last step, we need to find the solution where the tallest person is at position 5.

For this there is only 1 possible arrangement behind, that of 0 people.   The permutation equation for this is 4 pick 0, so

$$\text{\# of arrangements behind} = \frac{4\,!\ \ (people\ remaining)}{(4-0)\,!\ \ (people\ in\ front)} = 1$$

For the number of arrangements in front we have to solve for 4 people, 2 visible.   Of these 4 people, there are 3 possible places that the tallest of this group could be. (Side note, you can see by now why this type of problem can get a bit tedious to do manually. Computers are good at this, and there is actually a fairly elegant way to do all of this with the table we have been building in Excel that will be shown at the end)

So of these 4 people, we need 2 visible. That means the tallest of this group could be at position 2, 3, or 4



Main Branch 3, Sub Branch 1

If the tallest is in position 2, then

- Behind = 3 pick 2 permutations =   3! / (3-2)! = 6
- In front is only 1 arrangement of the 1 person

Main Branch 3, Sub Branch 2

If the tallest is in position 3 then

- Behind = 3 pick 1 permutations = 3! / (3-1)! = 3
- In front is only 1 arrangement, since of the two people only the tallest can be visible, so they must be in front.

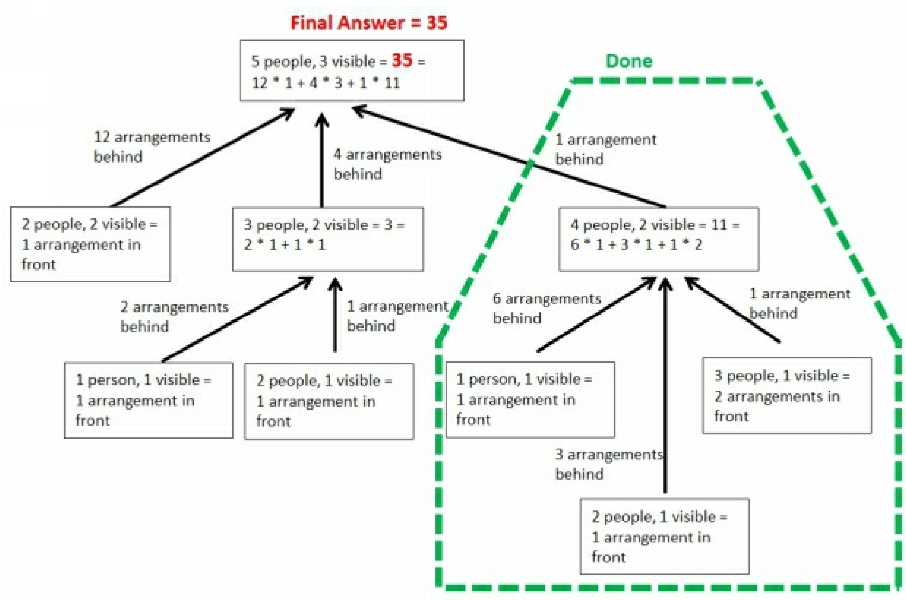
Main Branch 3, Sub Branch 3

If the tallest is in position 4 then

- Behind = 3 pick 0 permutations = 3! / (3-0)! = 1
- In front = 2 possibilities. The tallest of the three remaining must be in front since there is only 1 visible, and the other two can be swapped between themselves.

At this point, we have solved all of sub-problems, and the only steps left are to do some multiplication of arrangements in front or behind, and add them all up. **Note, do not be concerned if you lost track of a few of the steps.** This is a tedious problem that is hard to keep track of everything. Doing it as a table turns out to be easier, which we will get to after summing up this solution.

What we have is



*****Our final solution is that there are 35 possible arrangements with 5 people, 3 visible.** ***

This is are result of the sum of

- 12 arrangements when the tallest is in position 3
- 12 arrangements when the tallest is in position 4
- 11 arrangements when the tallest is in position 5

Drilling down another level, those had sub-sums of

- Tallest in position 3 – Total Sum = 12
  - 12 = 12 arrangements behind * 1 arrangement in front
- Tallest in position 4 – Total Sum = 12
  - 12 = 4 arrangements behind * 3 arrangements in front
- Tallest in position 5 – Total Sum = 11
  - 11 = 1 arrangements behind * 11 arrangements in front

Drilling down the final level

- Tallest in position 3 – Total Sum = 12
  - 12 = 12 arrangements behind * 1 arrangement in front
- Tallest in position 4 – Total Sum = 12
  - 12 = 4 arrangements behind * 3 arrangements in front
    - 3 arrangements =
      - 2 = 2 arrangements behind * 1 in front
      - 1 = 1 arrangement behind * 1 in front
- Tallest in position 5 – Total Sum = 11
  - 11 = 1 arrangements behind * 11 arrangements in front
    - 11 arrangements =
      - 6 = 6 arrangements behind * 1 in front
      - 3 = 3 arrangements behind * 1 in front
      - 2 = 1 arrangement behind * 2 in front

For the table, what we have calculated is

# of people					
5			35		1
4		11		1	
3	2	3	1		
2	1	1			
1	1				
# visible	1	2	3	4	5

The more elegant table.

We have built up a table for this problem although we haven't filled it in completely. Let's back up a step to where just the diagonals are filled in to see how this problem can be easier to solve using the table.

# of people					
5					1
4				1	
3			1		
2		1			
1	1				
# visible	1	2	3	4	5

The diagonals were fairly intuitive. If you have 5 people, and all 5 must be visible, there is only 1 possible arrangement.

Also fairly intuitive is the first column. If you have 5 people, and only 1 is visible, that must be the tallest person and they must be in front. So the other 4 people can be in any order. So the first column is just the factorial of one less than the number of people

# of people					
5	4! = 24				1
4	3! = 6			1	
3	2! = 2		1		
2	1! = 1	1			
1	0! = 1				
# visible	1	2	3	4	5

For the other numbers in the other columns, each value is a function of the numbers in the column immediately to the left, and also below it. For instance, 3 people, 2 visible is a function of every cell that has 1 person visible, and 2 or fewer people. In this case that happens to be 2 cells

# of people					
5	24				1
4	6			1	
3	2	3	1		
2	1	1			
1	1				
# visible	1	2	3	4	5

If we were looking at 5 people, 2 visible, that would be a function of 4 cells, because there are 4 cells in the column immediately to the left, that have fewer than 5 people. For 5 people, 3 visible, it would be a function of 3 cells, because there are 3 cells in the column immediately to the left that have fewer than 5 people.

# of people					
5	24	50	35	10	1
4	6	11	6	1	
3	2	3	1		
2	1	1			
1	1				
# visible	1	2	3	4	5

We haven't fully defined the function yet, just identified which cells are contributing. The function is the sum of all the arrangements of people standing in front of the tallest person multiplied by all the arrangements of people standing behind. The values in the cells are the values for the people standing in front, so we need to include the multiplication factor for the number of different ways people standing behind the tallest person can be arranged.

The equation for the how the people behind the tallest person are arranged is fairly intuitive when viewed as this table. It is simply the permutation equation with one less than the number of people in the cell, i.e. everyone except the tallest person as the total number of people, and then picking the number of people behind the tallest person.

Visually it is

# of people		
5	24	50
4	6 * 4p0	11
3	2 * 4p1	3
2	1 * 4p2	1
1	1 * 4p3	
# visible	1	2

In these charts, 4p2 stands for 4 people, selecting 2 using the permutation equation. So the value of 50 for 5 people 2 visible uses 4 people to select from, because it is one less than 5, and you increase the number of people selected for the permutation by 1 every cell you go down. The full equation is

$$50 = \ 6 * 4p0 \ + \ 2 * 4p1 \ + \ 1 * 4p2 \ + \ 1 * 4p3$$

$$50 = \ \frac{6 * 4!}{(4-0)!} + \frac{2 * 4!}{(4-1)!} + \frac{1 * 4!}{(4-2)!} + \frac{1 * 4!}{(4-3)!}$$

$$50 = \ 6 * 1 \ + \ 2 * 4 \ + \ 1 * 12 \ + \ 1 * 24$$

Another example for a different cell, the solution for 4 people, 3 visible is a function of 3 people 2 visible and 3 people 1 visible. The permutation equation for the number of people behind the tallest person uses 3 for the numerator because 3 is one less than the total number of 4 people.   The other two cells are then multiplied by 3 pick 0 and 3 pick 1.

# of people			
5	24	50	35
4	6	11	6
3	2	3 * 3p0	1
2	1	1 * 3p1	
1	1		
# visible	1	2	3

$$6 = \ \frac{3 * 3!}{(3-0)!} + \frac{1 * 3!}{(3-1)!}$$

Using these equations, it is relatively quick to set up a table for this problem in Excel, here is an 8x8 solution for this problem

# of people								
8	5040	13068	13132	6769	1960	322	28	1
7	720	1764	1624	735	175	21	1	
6	120	274	225	85	15	1		
5	24	50	35	10	1			
4	6	11	6	1				
3	2	3	1					
2	1	1						
1	1							
# visible	1	2	3	4	5	6	7	8

Which was set up using the PERMUT function, for permutations, in Excel

=E9+PERMUT(4,1)*E10+PERMUT(4,2)*E11+PERMUT(4,3)*E12

	D	E	F	G	H	I	J	K	L
4	# of people								
5	8	5040	13068	13132	6769	1960	322	28	1
6	7	720	1764	1624	735	175	21	1	
7	6	120	274	225	85	15	1		
8	5	24	=E9+PERMU	35	10	1			
9	4	6	11	6	1				
10	3	2	3	1					
11	2	1	1						
12	1	1							
13	# visible	1	2	3	4	5	6	7	8

The Excel file that has this table, and where all the Excel images in this book were taken from is available here http://www.fairlynerdy.com/permutations-example-problem/ for free. Since this was originally a programming problem, there is also a python program for free at the same location that solves this problem 3 different ways, first using a brute force try every permutation approach, then using the bottom up table like we have shown in Excel, and finally using a top down approach with recursion and memoization.

If You Find Bugs & Omissions:

We put some effort into trying to make this book as bug free as possible, and including what we thought was the most important information. However if you have found some errors or significant omissions that we should address please email us here

ImFairlyNerdy (at) gmail (dot) com

And let us know.   If you do, then let us know if you would like free copies of our future books.   Also, a big thank you!

More Books

If you liked this book, you may be interested in checking out some of my other books such as

- Hypothesis Testing: A Visual Introduction To Statistical Significance – This book demonstrates how to tell the difference between events that have occurred by random chance, and outcomes that are driven by an outside event. This book contains examples of all the major types of statistical significance tests, including the Z test and the 5 different variations of a T-test.
- Machine Learning With Random Forests And Decision Trees – If you like the book you just read on boosting, you will probably like this book on Random Forests. Random Forests are another type of Machine learning algorithm where you combine a bunch of decision trees that were generated in parallel, as opposed to in series like we did with boosting.
- Linear Regression and Correlation: Linear Regression is a way of simplifying a set of data into a single equation. For instance, we all know Moore's law: that the number of transistors on a computer chip doubles every two years. This law was derived by using regression analysis to simplify the progress of dozens of computer manufacturers over the course of decades into a single equation. This book walks through how to do regression analysis, including multiple regression when you have more than one independent variable. It also demonstrates how to find the correlation between two sets of numbers.

Thank You

Before you go, I'd like to say thank you for purchasing my eBook. I know you have a lot of options online to learn this kind of information. So a big thank you for reading all the way to the end.

If you like this book, then I need your help. **Please take a moment to leave a review on Amazon**. It really does make a difference, and will help me continue to write quality eBooks on Math, Statistics, and Computer Science.

P.S.

I would love to hear from you. It is easy for you to connect with us on Facebook here

https://www.facebook.com/FairlyNerdy

or on our webpage here

http://www.FairlyNerdy.com

But it's often better to have one-on-one conversations. So I encourage you to reach out over email with any questions you have or just to say hi!

Simply write here:

ImFairlyNerdy (at) gmail (dot) com

~ Scott Hartshorn