Contents

1 Week 1: Basic Counting

1.1 Starting to Count

- Why counting?
- Rule of Sum: If there are k objects of the first type and n objects of the second type, then there are $n + k$ objects of one of the two types
- Rule of Sum in Programming
  > Suppose there are 7 pizza places to eat and 5 burger places to eat. There are 12 places you can choose to eat.
- How Not to Use The Rule of Sum
  > Numbers from 1-10 that are divisible by 2 or 3. Naively you'd think $|\{2, 4, 6, 8, 10\}| + |\{3, 6, 9\}| = 5 + 3 = 8$. But there are actually 7.
  > Counting 6 twice is what gives the wrong result. The issue here is that the first and second types of objects are not disjoint.
- Convenient Language: Sets
  > Sets, Venn Diagrams, Intersection, Union, Cardinality
- Generalized Rule of Sum
  > If there is a set A with k elements, a set B with n elements, and these sets do not have common elements, then the set $A \cup B$ has $n + k$ elements
  > If they are not disjoint, then $|A \cup B| = |A| + |B| - |A \cap B|$.

1.2 Recursive Counting

- Number of Paths
  > How many paths on a graph from a source s to a sink t?
  > Describes BFS, where s starts with a path count of 1. Visiting a node means decorating that node with the sum of the path counts of the nodes that have an edge to it.
- Rule of Product
  > If there are k objects of the first type and n objects of the second type, then there are $k \times n$ pairs of objects, the first of the first type and the second of the second type.
  > If there is a finite set A and a finite set B, then there are $|A| \times |B|$ pairs of objects, the first from A and the second from B.
- Back to Recursive Counting

1.3 Tuples and Permutations

- Number of Tuples
  > How many passwords of exactly five symbols exist, if the symbols are chosen from the case-insensitive latin alphabet? 26^5
  > The number of tuples (sequences) of length k composed out of n symbols is n^k.
- License Plates
- Tuples with Restrictions

  - How many integers are there between 0 and 9999 that have exactly one 7 digits?
  - Model this as sequences of digits of length 4 / 4-tuples.
  - In the case that 7 is in position i then there are 9^3 possible tuples. So there are $4 * (9^3)$ possibilities.

- Permutations
  - Suppose we have a set of n symbols. How many different sequences of length k can we form out of these symbols if we aren't allowed to use the same symbol twice.
  - k-permutations: Tuples of length k without repetitions
  - Note that no k-permutations exist for $n < k$.
  - In total, there are $\frac{n!}{(n-k)!}$
  - n-permutations of n symbols are the simplest case of permutations, with the number of possible permutations being $\frac{n!}{(n-n)!} = n!$

2 Week 2: Binomial Coefficients

2.1 Combinations

- Previously on Combinatorics
- Generating Combinatorial Objects: Code
- Number of Games in a Tournament
  - In a tournament where each of n teams plays each other team once, how many games are played?
  - Each team plays a game against $n - 1$ other teams, so naively it seems that $n * (n - 1)$ games are played. But each game is a pairing of two teams, meaning that this technique counts each game twice (once as an (i, j) pairing and another time as a (j, i) pairing).
  - The number of games in a tournament with n teams (each pair of teams plays each other exactly once) is $\frac{n(n-1)}{2}$.
  - Another way of deriving this result: $T(n)$ is the number of games played. There are two types of games, the $n - 1$ games involving a given team, and the $T(n-1)$ games that do not involve that team. Hence, $T(n) = (n-1)+T(n-1)$. An arithmetic series.
  - A nice way to derive the value of an arithmetic series like $(n-1)+(n-2)+\ldots+2+1$ is to sum the series with itself, but reverse the order of the elements in one of them. So 2 times the series equals $((n-1)+1)+((n-2)+2)+\ldots+(2+(n-2))+(1+(n-1))$. In other words, a sum of $n - 1$ terms, each of which evaluates to n. So $T(n) = \frac{n(n-1)}{2}$
- Combinations
  - For a set S, its k-combination is a subset of S of size k.
  - The number of k-combinations of an n element set is denoted by $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, called "$n$ choose k".
  - One way of considering this is that if we take the k-permutations of n symbols, and remove those permutations where the items are reorderings of one another, then we have the k-combinations of that set. So the number of these is determined by the number of k-permutations of n symbols, divided by the number of k-permutations of k symbols, with $\frac{n!}{(n-k)!} \times \frac{(k-k)!}{k!} = \frac{n!}{k!(n-k)!}$.

2.2 Pascal's Triangle

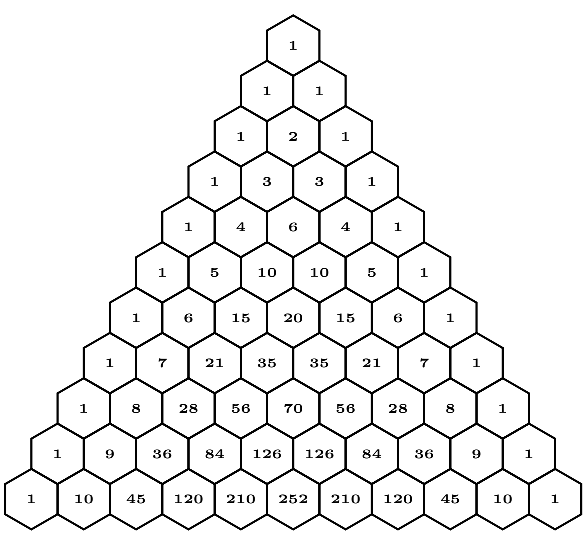
- Pascal's Triangle
  - There are n students. What is the number of ways of forming a team of k students? Well, it's $\binom{n}{k}$.
  - Let's derive it another way: For any given student, there are two types of teams: the $\binom{n-1}{k-1}$ teams with that student, and the $\binom{n-1}{k}$ teams without that student. So $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

- This brings us to Pascal's Triangle. On the i'th row the values are $\binom{i-1}{0} \dots \binom{i-1}{i-1}$.
- Boundary values are all of the form $\binom{n}{0} = 1$ or $\binom{n}{n} = 1$.
- For each interior value (a value at the i'th row and j'th column, for $i > 1$ and $1 < j < i$), it's the sum of its immediate 'parent' elements: $\binom{i-1}{j} = \binom{i-2}{j-1} + \binom{i-2}{j}$

- Symmetries
  - The left-right symmetry of Pascal's triangle expresses that $\binom{n}{k} = \binom{n}{n-k}$.
  - The intuition here is that to partition of students into a team of size k leaves $n-k$ students available to form a different team, and vice-versa. There are exactly as many ways of creating this partition in either case.

- Row Sums
  - The sum of all elements in the i'th row of Pascal's triangle is is 2^{i-1}.
  - Proving this by induction involves setting up a recurrence relation. Given that 'interior values' are the sum of two upstairs neighbors, the sum of any row is twice the sum of the previous row.
  - The combinatorial intuition behind a row's sum is each cell's value is $\binom{n}{k}$, the number of k-subsets of a set of size n for different values of k.
  - $\sum_{k=0}^{n} \binom{n}{k}$ is the number of all subsets of an n element sets.
  - This is 2^n by the product rule: for each of the n elements, each distinct subset may either contain it or not, so there are $\prod_{0}^{n} 2 = 2^n$ distinct subsets
  - Another interesting property is that if you take all elements of a row, alternate the plus/minus sign, and sum that row, you always get 0. So, for $n > 0$, $\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$.
  - Again, a convenient way of thinking about this is that each value from the row above is represented twice in the current row, and will be positive in one term of the current row, and negative in another term, thus producing a sum of 0.

- Binomial Theorem
  - The coefficients in the expansion of $(a+b)^i$ are taken from the i'th row of Pascal's triangle: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$.



2.3 Practice Counting

- How many five card hands out of a 52-card deck? $\binom{52}{5}$
- How many five card hands have 2 hearts and 3 spades? IOW, how many combinations of 2 hearts (out of the 13 in the deck), how many combinations of 3 spades (out of the 13), and how many combinations of these combinations? So $\binom{13}{2} \times \binom{13}{3}$.
- How many non-negative integers having at most four digits where the digits are strictly increasing?
  - Secretly, this is asking "how many combinations of 4 distinct digits exist, such that they could be rearranged into the single ordering which is strictly increasing?". Thus: $\binom{10}{4}$.

3 Week 3: Advanced Counting

3.1 Combinations with Repetitions

- Review
  - So far we've considered choosing k items out of n options, in 3 of 4 different ways: Ordered sequences, permitting and prohibiting repetition of items; and unordered sequences without repetitions.
  - Suppose we have an unlimited supply of tomatoes, bell peppers, and lettuce, and we want to make a salad out of 4 units of these ingredients (no need to use all ingredients). How many different salads can be made?
  - Since order doesn't matter in the final salad, we don't need to pick ingredients in any specific order. This means our representations.
  - Instead we can consider it recursively: If the salad contains $0 \leq j \leq 4$ tomatoes, then we count the number of $4-j$-ingredient salads that can be made using only bell peppers and lettuce.
- Combinations with Repetitions
  - Okay, now let's ask how many 7-ingredient salads can be made out of 4 distinct ingredients.
  - Since we can represent ingredients in any order, we really just need to choose an ordering for the ingredients, and the possible ingredient counts for each ingredient.
  - A simpler representation yet is: suppose each ingredient is a 1 in a string, and we signal the transition from one ingredient to another using a delimiter $\bigcirc$. For 4 distinct ingredients, we need to know where the 3 $\bigcirc$ symbols are inserted in a string of length 10. So in fact, there are $\binom{10}{3}$ distinct ways of making this salad.
  - The number of combinations of size k, chosen out of n objects with repetitions allowed, is equal to $\binom{k+n-1}{n-1}$.

In the end, this is what we have:

	Repetitions allowed	**No repetition allowed**
Ordered	Tuples; n^k	k-permutations; $\frac{n!}{(n-k)!}$
Unordered	combination with repetitions; $\binom{k+n-1}{n-1}$	Combinations; $\binom{n}{k}$

3.2 Problems in Combinatorics

- Distributing Assignments among People
  - Suppose there are 4 people and 9 different assignments, each person should receive one assignment.
  - We will count selections of assignments for 4 people. People are different so the selection is ordered. No assignment can be given two two people, so there are no repetitions.
  - It's a k-permutation, with $\frac{n!}{(n-k)!} = \frac{9!}{5!} = 9 \times 8 \times 7 \times 6$.

  - Suppose there are 4 people and 9 different assignments. We need to distribute all assignments among the people. No assignment should be assigned to multiple people. Each person can get any number of assignments (none or all 9).
  - This looks tricky. But if we think of giving people to assignments instead of giving assignments to people, it starts to look simple.
  - If we think of assigning people to assignments, then for each assignment we have 4 people to choose from. In this sense, we're looking at counting possible tuples, with 4^9 possible ways of assigning people to assignments.

- Distributing Candies Among Kids
  - There are 15 identical candies. How many ways are there to distribute them among 7 kids?
  - Repetitions are okay (necessary in fact, due to pigeonhole principle), and candies are identical so order doesn't matter. So in fact with $\binom{k+n-1}{n-1}$ evaluated with $k = 15$ and $n = 7$
  - There are 15 identical candies. How many ways are there to distribute them among 7 kids such that each kid receives at least 1 candy.
  - We can reduce this problem to the previous one.
  - So, first, we allocate one kid each candy. After that, the remaining candies can be assigned as in the previous problem. So there are $\binom{k+n-1}{n-1}$ ways of doing it, with $n = 7$ (still 7 kids), but now with $k = 15 - 7 = 8$ (only 8 candies are left after the equal allocation).
  - Something to note is how much these quantities differ: 3003 distributions where each kid gets a candy out of 54264 total distributions.

- Numbers with Fixed Sum of Digits
  - How many non-negative integers are there below 10000 such that the sum of their digits is equal to 9?
  - A useful approach: there are four positions that start at a value of 0. Consider distributing a weight of 9 among them, 1 at a time, and each time we pick a digit to increase. So this is a combination of 9 ones among 4 options, where order doesn't matter.
  - How about the same problem, except with a sum of digits of 10 instead of 9?
  - Applying the same approach doesn't work: the previous approach would have counted situations where position 1 was assigned all the ones.
  - One way to fix this is to subtract the illegal situations: for each of the 4 positions subtract 1 outcome where 10 was assigned to that position. This yields the correct result.

- Numbers with Non-Increasing Digits
  - How many 4-digit numbers are there such that their digits are not increasing.
  - Really we're picking any 4 digits, then re-ordering them in non-increasing order.
  - So it's a combination of size 4 from 10 options, with repetition allowed.

- Splitting into Working Groups
  - We have 12 students in the class. How many ways are there to split them into working groups of size 2?
  - We could go working-group by working-group: First group needs 2 people, second group needs 2 people, etc. So there should be $\binom{12}{2} \times \binom{10}{2} \times \binom{8}{2} \ldots \binom{2}{2}$.
  - This seems simple and straightforward to calculate. But it's wrong.
  - In this formulation, identical groups may be counted multiple times (e.g. groups 1 and 2 can be $\{\{3, 7\}, \{1, 5\}\}$ and $\{\{1, 5\}, \{3, 7\}\}$, respectively)
  - In fact, order within groups doesn't matter. But also order between groups doesn't matter.
  - What can we do? Fix it after the fact! We have counted each splitting-into-groups 6! times, once for each permutation of 6 groups.

4 Week 4: Probability

4.1 What Is Probability?

- The Paradox of Probability Theory
  - Attempts to predit the unpredictable
  - A canonical example: coin tossing
  - A single coin toss is very unpredictable
  - A small number of coin tosses should have close to 50% of each result
  - Many repeated coin tosses, however, should give very close to 50% of each result
- Galton Board
  - Galton Board = B Machine: a funnel sends beans down a single central channel, and there's an arrangement of pegs that they can hit on the way down. Below the pegs are a number of channels they can rest in.
  - The distribution of beans ends up looking a lot like Pascal's triangle, with all values on the i'th level divided by 2^i
- Natural Sciences and Mathematics
  - Natural sciences care about developing models (e.g of coins) and whether real coins behave according to a model
  - Mathematics cares about the implications of a model
  - What's the probability of getting 1 heads and 1 tails on a toss of 2 coins
  - Alice might say that there are 4 distinct outcomes, HH, HT, TH, TT, giving a $\frac{2}{4}$ probability
  - Bob might say that there are 3 distinct outcomes: 2H, 2T, and HT, giving a $\frac{1}{3}$ probability
  - a pure mathematician might say that both of these are correct given different assumptions (a particle physics application may favor Bob's answer)
  - Consider a Galton board where a bean hitting a peg (and going either left or right) will gain horizontal velocity, influencing the likelihood of its next move.
  - But we assumed *independence* between the movements
- Rolling Dice
  - Natural sciences may say that each face of six-sided die occurs equally often
  - Mathematics would then predict that even numbers occur 50% of the time, numbers divisible by 3 occur $\frac{1}{3}$ of the time, etc..
  - What if we roll two six-sided dice, a red one and a blue one?
  - There are 36 equiprobable outcomes if you consider them as an ordered pair (red observation, blue observation)
  - The *probability space*: the set of all *outcomes*
  - *event*: some outcome (which we typically assume to be favorable, important, or interesting)
  - *probability* of an event is the number of outcomes that match your event's description, divided by the set of available outcomes
  - For dice which are independent, this holds for rolling the dice simultaneous, or in any sequential order.
- More Probability Spaces
  - Let's consider a sequence of n coin tosses. The outcome is a sequence of n bits, there are 2^n of these.
  - The Galton Board can also be considered as a sequence of Left/Right moves of length n. So if we want to know the probability of a bean going Right 40%-60% of the time ($\#R \in [0.4n, 0.6n]$), then it is $\displaystyle\sum_{k \in [0.4n, 0.6n]} \frac{\binom{n}{k}}{2^n}$
  - "Why would we learn probability theory, it's just combinatorics/counting?"
  - We've made big assumptions, which limits the ability of these models to make useful predictions. Our assumptions have included: independence; uniform distributions; known distributions; discrete distributions

4.2 Probability: Do's and Don't's

- Not Equiprobable Outcomes
  - It's okay if outcomes are not equiprobable, as long as the frequencies of distinct outcomes stabilize around some specific number and the sum of probabilities of all outcomes is 1.
  - Finite probability space: a finite set X of outcomes. Each outcome i has some probability $0 \leq p_i \leq 1$, and $\sum p_i = 1$.
  - An event is any sum of outcomes.
  - Then the probability of the event is the sum of the outcome probabilities.
- More about Finite Spaces
  - The event must a set of outcomes in the same probability space.
  - Flipping a coin and rolling a die are not the same probability space, for example. Though you can construct a finite space out of their separate probability spaces to fit your problem.
  - If there are two mutually exclusive events, A and B, then $\Pr[AorB] = \Pr[A] + \Pr[B]$
  - $\Pr[notA] = 1 - \Pr[A]$
  - Sequential choice: Suppose we have 6 balls labelled 1..6, in two boxes, with $1, 2$ in box 1 and $3, .., 6$ in box 2, and pulling a ball from a box gives equal chances of pulling each ball in that box. We choose a random box and then a random ball: what are the probabilities of each ball being pulled? 1/4, 1/4, 1/8, 1/8, 1/8, 1/8
- Mathematics for Prisoners
  - King offers to a prisoner a means of being pardoned. There are 15 white balls and 15 black balls to be placed in two boxes. The prisoner may place the balls into the boxes however they'd like (though each box must have 1 ball). The king will select 1 box randomly, and then select a ball randomly out of that box. If the ball selected is white, the prisoner is pardoned. If the ball selected is black then the prisoner is killed.
  - Prisoner should put in one box a single white ball, and all remaining balls into the other box. Then the probability of being released is $(\frac{1}{2} \times 1) + (\frac{1}{2} \times \frac{14}{29}) = 0.74$.
  - Why is this optimal?
  - Can think of the prisoner as having two parameters: how many total balls to put into each box, then how many balls of each color in the first box.
  - A given ball's color matter more in a box which contains fewer balls: each ball in a box of i balls has a $\frac{1}{2i}$ chance of being selected.
  - So a high concentration of white balls is better in a box with few balls, and a box with a higher than 50% concentration of black balls is better the more balls there are in that box.
- Not All Questions Make Sense
  - The biggest danger in probability theory is computing something and making a wrong decision that means your computation is not based on probability theory.
  - "What is the probability that I have a dollar bill in my pocket?" can't be computed for example.
  - There is more to probability theory than is being covered here....
  - Infinite probability spaces (e.g., chances of getting a random natural number i, where their frequencies are defined by $\sum_{i=1}^{\infty} \frac{i}{2^i} = 1/2, 1/4, ...$, for example expressible as the odds of getting a specific number of heads before the first tail is observed)
  - "market interpretations", like when people say "the probability that X will be reelected is 80%" and from there that the obligation to pay \$1 if X is reelected trades around \$0.8.

4.3 Conditional Probability

- What Is Conditional Probability?
  - Conditional probabilities and independence: topics which distinguish probability theory from measure theory (combinatorics)
  - wikipedia says that about 14% of US is age 65 and over, also that the ratio of men:women in this age group is 3:4 (43% are males)
  - The fraction of total population that's males over 65 is given by $\Pr[male\& \geq 65] = \Pr[\geq 65] \times \Pr[male| \geq 65] = 0.14 \times 0.43 \approx 0.06$
  - Suppose we have a probability space of 6 equiprobable outcomes. We want to talk about outcomes A (the 4 numbers that are at least 3) and outcomes B (the 3 even numbers). The probability of A and B both being true is then the intersection of the two sets.
- How Reliable is the Test?
  - A disease that 1% of the population has. There's a test for it, the test gives a false negative (reports no disease for people who actually have it) 10% of the time, and a false positive (reports presence of the disease for people who don't have it) 10% of the time.
  - So what is the chance of having the disease if you test positive?
  - $\Pr[D] = 0.01$; $\Pr[T|D] = 0.9$; $\Pr[T|\neg D] = 0.1$. What's $\Pr[D|T]$?
  - 1% of people are ill, of which 90% test positive, so 0.9% people are positive and ill. 10% of people test positive among the 99% who are not ill, giving 9.9% of the population who are not ill but test positive.
  - $\Pr[T\&\neg D] = \Pr[\neg D] \times \Pr[T|\neg D] = 0.99 \times 0.1 = 0.099$
  - $\Pr[T] = \Pr[T\&D] + \Pr[T\&\neg D] = 0.009 + 0.099 = 0.108$
  - $\Pr[D \mid T] = \frac{\Pr[D\&T]}{\Pr[T]} = \frac{0.009}{0.108} \approx 0.833$
  - The Law of Total Probability: suppose there are a set of mutually exclusive (disjoint) and exhaustive cases ($X = \bigcup_1^n B_i$ and $\forall i \neq j : B_i \cap B_j = \emptyset$). Then any set A is also represented completely as a number of subsets, $A \cap B_1$ and so on. So $\Pr[A] = \Pr[A\&B_1] + \ldots + \Pr[A\&B_n] = (\Pr[B_1] \times \Pr[A \mid B_1]) + \ldots + (\Pr[B_n] \times \Pr[A \mid B_n])$
- Bayes' Theorem
  - Consider two events, H the hypothesis and E the evidence, which are in a single probability space. $\Pr[H|E] = \frac{\Pr[H\&E]}{\Pr[E]} = \frac{\Pr[E|H]}{\Pr[E]} \Pr[H]$.
  - explanation of what this means: condition E multiplies the probability of H by a factor, the likelihood to see evidence E when the hypothesis is true $(E|H)$ divided by the probability of that evidence being true in general.
  - One interpretation of Bayes Theorem is that *if condition B increases the probability of A by factor k, then condition A increases the probability of B by the same factor k*
- Conditional Probability: A Paradox
  - "Mary has two children. At least one of them is a girl. What is the probability that she has two daughters?", or cleaned up of biological realities: "Mary tossed a fair coin twice. At least one of the outcomes is a tail. What is the probability that she got two tails?"
  - faulty argument: either the first or the second bit was a tail. There are two outcomes, HT and TH. Assume without loss of generality it was the first, then of this subset of situations there are two options here: HT and TT. So the chances are 1/2 if the first was the tails, and 1/2 if not. So the probability is 1/2.
  - parse the problem:
  - "Mary tossed a fair coin twice" →probability space is 00, 01, 10, 11 (for heads=0 and tails=1)
  - "At least one of the outcomes is a tail." →prior evidence E, that she got at least one tail (contains 3 of the outcomes, probability 3/4)
  - "What is the probability that she got two tails?" →we want to find $\Pr[H|E]$.

  - two tails occurs in 1 of our 4 outcomes.
  - $\Pr[H|E] = \frac{\Pr[H\&E]}{\Pr[E]} = \frac{1/4}{3/4} = \frac{1}{3}$
  - The problem with the faulty argument is that we double-count the tails-tails case by taking the union of two non-disjoint cases. In some sense, we confused information about the prior/evidence/condition (one toss gave tails) with information about the underlying probability space.

- Past and Future
  - Imagine you have a die and you roll it twice. If we get a 6 in the first roll, does that influence our odds of getting a 6 the second roll?
  - According to our model, no.
  - But to borrow from "Rosencrantz and Guildenstern are Dead", if you've gotten 999 heads in a row, do you believe that the next toss will be a heads with 50% probability? No.
  - What's going on here is not a failure of the model, it's that we are using the wrong model for the situation.
  - There's a standing joke about this: an amateur statistician brings a bomb to an airplane because he thinks that two bombs on the same plane is quite improbable.

- Independence
  - A and B are *independent* if $\Pr[A|B] = \Pr[A]$.
  - Events with a probability of zero are always vacuously independent of any other event.
  - For independent events with nonzero probability, this provides a symmetric consequence: $\frac{\Pr[A\&B]}{\Pr[B]} = \Pr[A]$.
  - Thus, the product rule for probabilities is only a consequence of this definition.
  - If you assume non-independent events, then the product rule predicts that A makes B {more, less} probable if B makes A {more, less} probable, respectively.

4.4 Monty Hall Paradox

- Setting and Three Convincing Arguments
  - Player is asked to pick from any of 3 doors. The game host then opens either of the other two doors, revealing a goat. Then the host offers to the player whether they would like to keep their original door choice (in which case, the player wins a prize), or change to the door which was not opened. Only one door contains a prize behind it. What should a player do, keep their pick or change it?
  - Argument # 1: Why you should keep the guess
    - The opening of the other door does not prove anything since an empty door always exists.
    - Indeed, the door opening doesn't decisively show the location of the prize, but it does provide new information about the distribution of the prize.
  - Argument # 2: Make a random guess between the two non-opened doors
    - Now there are two doors where the prize can be; we don't know where it is, so we can only make a random guess
    - This is an enumeration of the probability space without taking into account the probabilities of the outcomes.
    - If you want to guess the result of a random process with known probability, you should *not* imitate this process. If you have a coin that gives 'head' 70% of the time, you should bet on heads every time (not 70% of the time).
  - Argument # 3: Why to change doors
    - The original probability that the goat was behind the door you already selected was $\frac{1}{3}$, with the remaining $\frac{2}{3}$ probability distributed evenly among the other two doors.
    - Once one of the other two doors is selected, that $\frac{2}{3}$ probability is concentrated in the door you hadn't selected which the host didn't open.

- 'Our Position'
  - General advice about paradoxes/unexpected results in probability theory: think about repeating the experiment many times

  - If you can't, then probably the experiment is not formulated rigorously
  - if it's not repeatable, probably the question is not well-formulated.
  - So the keep strategy should deliver a prize 1/3 of the time. But in the other 2/3 of the times when keep would lose, change would win every time.
  - 'Choose Random', by contrast, would win a prize half of the time.

5 Week 5: Random Variables

5.1 Random Variables and Expectations

- Random Variables
  - So far we've studied probability distributions; events (subsets of outcomes, characterized by whether a given outcome is or is not counted in this subset) and their probabilities
  - A random variable f is a variable whose value is determined bya random experiment.
  - We have a probability distribution on the finite set X of k outcomes
  - Outcomes have probabilities $p_1, \ldots, p_k$
  - To define f we assign a number a_i to each outcome, with probability p_i.
- Average
  - Problem: A student got scores 78, 72, and 87 on three tests. What's the average of these 3 tests?
  - So: $\frac{78+72+9=87}{3} = \frac{237}{3} = 79$
  - Another problem: suppose HR in some company decides to consistently fire everyone who performs below average
  - Note that unless everyone works equally well, someone will always be below average and will get fired. Once that person is fired, the new distribution will also have a least performer. And so on until all employees are fired but one.
  - Problem: suppose we throw a die many times. What is the average outcome?
  - The real average depends on the values observed: if you throw only 6's then the average is 6, if you throw only 1's then the average is 1.
  - But we can give an approximation that is good with high probability.
  - If we throw the dice n times for a very large n, then among the outcomes there are $\frac{n}{6}$ instances of each die face, giving $\frac{n(1+2+3+4+5+6)}{6} = \frac{21n}{6} = 3.5n$
  - This is the expected value, or the expectation, of the die throw.
- Expectation
  - If we have a random variable f with values $a_1, \ldots, a_k$ on a distribution with k outcomes, with probabilities on these outcomes $p_1, \ldots, p_k$ and we repeat this random experiment on this many times (n times). And we want to know the expectation of f.
  - A value a_i happens about $p_i n$ times; the expectation (denoted $\mathbb{E}f$) is $\mathbb{E}f \approx \frac{a_1 p_1 n + \ldots + a_k p_k n}{n} = a_1 p_1 + \ldots + a_k p_k$

5.2 Linearity of Expectation

- Linearity of Expectation
  - Suppose there are two random variables f and g over the same probability space, with outcomes $a_1, \ldots, a_k$ for f and $b_1, \ldots, b_k$ for g. Consider $f + g$; what can we say about its expectation?
  - It is also a random variable over the same probability distribution, with $\mathbb{E}(f + g) = \mathbb{E}f + \mathbb{E}g$.
  - Linearity is a very useful property for simplifying computations
  - Example 1: We throw two dice. What is the expected value of their sum? It's 7.

- Example 2: We toss a coin 5 times in a row. What is the expected number of tails?
  - We can compute it directly: there are 32 outcomes, we can count up the number of tails of each (or use combinatorics to compute the count)
  - Let f_i be an outcome of the i'th coin: it is 1 if the outcome is tails, 0 if heads. We are interested in $\mathbb{E}(f_1 + f_2 + f_3 + f_4 + f_5)$. The expectation for a single coin is trivially computed as $\mathbb{E}f_i = (0 \times \frac{1}{2}) + (1 \times \frac{1}{2})$, so actually we get to a value of 2.5.

- Birthday Problem
  - Consider 28 randomly chosen people. Consider the number of pairs (i, j) such that the i'th person has a birthday on the same day as the j'th person. Show that the expectation of this number is greater than 1.
  - If there are two people with the same birthday, they will contribute 1 to the number of pairs in the problem. If there are 3 people with the same birthday, they form 3 pairs.
  - Let's formalize the problem
  - Assume birthdays are distributed uniformly among 365 days of the year (in reality, nonuniform distributions only increase the expectation), and people are chosen independently and uniformly
  - Use the linearity of expectation. Denote the number of pairs of people with the same birthday by f. Enumerate people from 1 to 28; consider a random variable $g_{ij} = 1$ if persons i and j have same birthday, 0 otherwise. With this construction, $f = g_{ij}$ over all unordered pairs of i and j.
  - Since we know this, we need to compute $\mathbb{E}g_{ij}$ and to count the number of (i, j) pairs that exist.
  - $\mathbb{E}g_{ij} = (1 \times \frac{1}{365}) + (0 \times \frac{364}{365}) = \frac{1}{365}$. Another way of saying this is that no matter what one date is, the second date will match it in $\frac{1}{365}$ choices.
  - How many pairs of i and j do we have? With 28 people in total, there are $\binom{28}{2} = 378$ such pairs.
  - So $\mathbb{E}f = 378 \times \frac{1}{365} > 1$

5.3 Expectation is Not Everything

- Expectation Is Not Everything
  - Suppose Alice and Bob are playing a dice game, where the faces on Alice's die are 2,2,2,2,3,3 and the faces on Bob's are 1,1,1,1,6,6. They each throw their die, the one with the larger number showing wins. Who will win more often?
  - Let's see who has a better expected value? Alice has $(2 \times \frac{2}{3}) + (3 \times \frac{1}{3}) = \frac{7}{3}$, Bob has $(1 \times \frac{2}{3}) + (6 \times \frac{1}{3}) = \frac{8}{3}$. Bob has a better expected value.
  - Expected values on these dice are not relevant in this setting: the winner depends on the combination of throws, not on the real values.
  - Who wins actually depends only on what Bob throws. If Bob throws a 1, he loses (it is lower than every value on Alice's die), if he throws a 6 he wins (it is higher than every value on Alice's die).
  - Bob throws a 1 with probability $\frac{2}{3}$ so he'll lose with that probability.
  - In fact, expectation not only is irrelevant in this case, but it gives the opposite prediction of a better analysis. When Bob wins his die roll, it is by a large margin (6 versus either 3 or 2), and when he loses it, it is by a small margin (1 versus either 3 or 2). However, these margins aren't pertinent for the win versus lose decision.

5.4 Markov's Inequality

- From Expectation to Probability
  - Problem: Suppose there's a lottery, with each ticket costing \$10. 40% of the lottery budget goes to prizes. Show that chances to win \$500 or more are less than 1%.
  - Proof by contradiction: Assume the probability to win \$500 or more is at least 0.01. Denote the number of tickets sold by n.
  - Then the budget of the lottery is $10n$ dollars, with $10n \times 0.4 = 4n$ dollars spent on prizes

  > By our assumption at least $\frac{n}{100}$ win at least \$500 dollars, so these tickets in total win $\frac{n}{100} \times 500 = 5n$ dollars.
  > This exceeds the total price budget of $4n$.

- Markov's Inequality

  > Markov's Inequality: Suppose f is a non-negative random variable. Then for any $a > 0$ we have $\Pr[f \geq a] \leq \frac{\mathbb{E}f}{a}$.
  > This lets us use expected value to bound the probability of certain events.
  > How can we prove Markov's inequality?

    * First rewrite the inequality we wish to prove as $a \times \Pr[f \geq a] \leq \mathbb{E}f$
    * Construct a random variable g on the same probability space as f, and for each a_i in f, the corresponding value in g is defined as a if $a_i \geq a$ and is defined as 0 if $a_i < a$.
    * Note that by this construction, the values in g are less than or equal to the values in f for each outcome in their shared probability space: $\mathbb{E}g \leq \mathbb{E}f$
    * What's the expectation of g? It has only one nonzero value, a. So $\mathbb{E}g = (0 \times \Pr[f < a]) + (a \times \Pr[f \geq a])$
    * By now we have $\mathbb{E}g \leq \mathbb{E}f$ and $\mathbb{E}g = a \times \Pr[f \geq a]$.
    * Combining these, $\mathbb{E}f \geq \mathbb{E}g = a \times \Pr[f \geq a]$, showing Markov's inequality

- Application to Algorithms

  > Suppose there is a randomized algorithm that runs on average time a^2, where n is the size of the input. The algorithm outputs the correct answer. Construct another randomized algorithm that always stops in time cn^2 for some constant c and makes a mistake with probability at most 10^{-3}.
  > We'll apply Markov's inequality
  > Running time is a random variable f, with $\mathbb{E}f = n^2$.
  > New algorithm: let's run the original algorithm for $10^3 n^2$ steps; if it stops, we return the value it found; if it does not, then we return a wrong value or indicate failure
  > Claim: the probability that the original algorithm doesn't stop after $10^3 n^2$ steps is at most 10^{-3}.
  > $\Pr[f \geq 10^3 n^2] \leq \frac{\mathbb{E}f}{10^3 n^2} = \frac{n^2}{10^3 n^2} = 10^{-3}$

- Summary

  > Random variables let us study randomness quantitatively, via many analytical tools.
  > Expectation is a characteristic of a random variable which is useful, because it carries a lot of information about that random variable, and as a single number, is very convenient to use in computations.