CIS 2033 Lecture 6, Spring 2017

Instructor: David Dobor

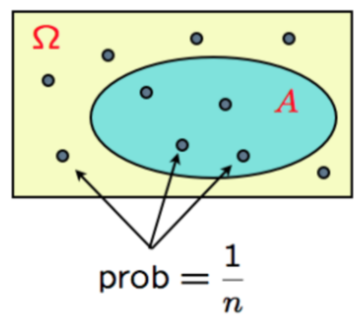
February 2, 2017

In this lecture, we introduce the basic principle of counting, use it to count subsets, permutations, combinations, and partitions, and apply it to some probability problems. As we mentioned at the beginning of this class, calculus is a prerequisite for this course. Counting, however, is not. So we begin by considering the very basics of counting.

Introduction

A basketball coach has 20 players available. Out of them, he needs to choose five for the starting lineup, and seven who would be sitting on the bench. In how many ways can the coach choose these 5 plus 7 players? It is certainly a huge number, but what exactly is it? In this lecture, we will learn how to answer questions of this kind.

More abstractly, we will develop methods for counting the number of elements of a given set which is described in some implicit way. Now, why do we care? The reason is that in many models, the calculation of probabilities reduces to counting. Counting the number of elements of various sets. Suppose that we have a probability model in which the sample space, Ω, is finite, and consists of n equally likely elements.



So each element has probability $1/n$. Suppose now that we're interested in the probability of a certain set, A, which has k elements. Since each one of the elements of A has probability $1/n$, and since A has k distinct elements, then by the additivity axiom, the probability of A is equal to $k \times 1/n$. Therefore to find the probability of A, all we have to do is to count the number of elements of Ω and the number of elements of A, and so determine the numbers k and n.

Discrete uniform law

– Assume Ω consists of n equally likely elements
– Assume A consists of k elements

Then : $$\mathbf{P}(A) = \frac{\text{number of elements of } A}{\text{number of elements of } \Omega} = \frac{k}{n}$$

Of course, if a set is described explicitly through a list of its elements, then counting is trivial. But when a set is given through some abstract description, as in our basketball team example, counting can be a challenge.

In this lecture, we will start with a powerful tool, the basic counting principle, which allows us to break a counting problem into a sequence of simpler counting problems. We will then count permutations, subsets, combinations, and partitions. We will see shortly what all of these terms mean.

In the process we will solve a number of example problems, and we will also derive the formula for the binomial probabilities, the probabilities that describe the number of heads in a sequence of independent coin tosses. So, let us get started.

Basic Counting Principle

In this segment we introduce a simple but powerful tool, the basic counting principle, which we will be using over and over to deal with counting problems. Let me describe the idea through a simple example.

YOU wake up in the morning and you find that you have in your closet

- 4 shirts
- 3 ties
- 2 jackets.

In how many different ways can you get dressed today?

To answer this question, let us think of the process of getting dressed as consisting of three steps, three stages. You first choose a shirt, any one of the shirts - you have 4 choices of shirts. But each shirt can be used together with 1 of the 3 available ties to make 3 different shirt-tie combinations. But since we had 4 choices for the shirt, this means that we have $4 \times 3 = 12$, shirt-tie combinations.

Finally, you choose a jacket. Each shirt-tie combination can go together with either jacket, and so the fact that you have 2 jackets available doubles the number of options that you have, leading to 24 different options overall.

So 24 is the answer to this simple problem. And how did the number 24 come about? Well, 24 is the same as the number of options you had in the first stage times the number of options you had in the second stage times the number of options you had in the third stage:

$$24 = 4 \times 3 \times 2$$

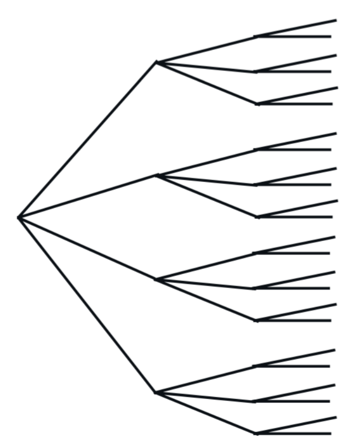


Figure 1: Picking attire in 3 stages: 4 different ways to pick a shirt, 3 to pick a tie, and 2 to pick a jacket. Thus this tree has $24 = 4 \times 3 \times 2$ leaves.

Let us generalize. Suppose we want to construct some kind of object, and we're going to construct it through a sequential process, through a sequence of r different stages (in the example that we just considered, the number of stages was equal to 3). At stage i of this selection process, suppose you have a number - call it n_i - of options that are available (in our example, at the first stage we had 4 options, at the second stage we had 3 options, and at the last stage we had 2 options).

What is important is that when you reach stage i, no matter what you chose, no matter what you did at the previous stages, the number of options that you will have available at stage i is going to be that fixed number, n_i.

So how many different objects can you construct this way? Well, just generalizing from what we did in our specific example, the answer is the product of the number of choices or options that you had at each stage.

$$\text{Number of objects we can construct } = n_1 \cdot n_2 \cdot \ldots \cdot n_r$$

This is the counting principle. It's a very simple idea, but it is powerful. It will allow us to solve fairly complicated counting problems. However, before we go into more complicated problems, let us first deal with a few relatively easy examples.

EXAMPLE 1. Consider license plates that consist of 2 letters followed by 3 digits. The question is, how many different license plates are there? We think of the process of constructing a license plate as a sequential process.

At the first stage we choose a letter, and we have 26 choices for the first letter. Then we need to choose the second letter, and we have 26 choices for that one. Then we choose the first digit. We have 10 choices for it. We choose the second digit, for which we have 10 choices. And finally, we choose the last digit, for which we also have 10 choices. So if you multiply these numbers, you can find the number of different license plates that you can make with 2 letters followed by 3 digits.

$$\text{Number of different license plates } = 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10$$

(REPETITIONS ALLOWED)

EXAMPLE 2. Now let us change the problem a little bit and require that no letter and no digit can be used more than once. So, let us think of a process by which we could construct license plates of this kind. In the first stage, we choose the first letter that goes to the license plate, and we have 26 choices. Now, in the second stage where we choose the second letter, because we used 1 letter in the first stage, there are only 25 available letters that can be used. We only have 25 choices at the second stage.

Now, let us start dealing with the digits. We choose the first digit, and we have 10 choices for it. However, when we go and choose the next digit we will only have 9 choices, because 1 of the digits has already been used. At this point, 2 digits have been used, which means that at the last stage we have only 8 digits to choose from. So by multiplying these numbers, we see that the number of license plates if repetition is prohibited is:

$$\text{Number of different license plates } = 26 \cdot 25 \cdot 10 \cdot 9 \cdot 8$$
$$(\text{ REPETITIONS NOT ALLOWED })$$

EXAMPLE 3. Now suppose that we start with a set that consists of n elements. What we want to do is to take these n elements and order them. A terminology that's often used here is that we want to form a *permutation* of these n elements. One way of visualizing permutations is to say that we're going to take these elements of the set, which are unordered, and we're going to place them in a sequence of slots.

So we create n slots. And we want to put each one of these elements into one of these slots. How do we go about it? We think of putting the elements into slots, one slot at a time.

We first consider the first slot. We pick one of the elements and put it there. How many choices do we have at this stage? We have n choices, because we can pick any of the available elements and place it in that slot.

Next, we pick another element and put it inside the second slot. How many choices do we have at this step? Well, we have already used one of the available elements, which means that there's $n-1$ elements to choose from at the next stage. At this point, we have used 2 of the elements. There are still $n-2$ elements left. We pick one of these $n-2$ elements and put it in the third slot this point.

We continue this way. At some point we have placed $n-1$ of the elements into slots. There's only one element left, and that element, necessarily, will get into the last slot. There are no choices to be made at this point.

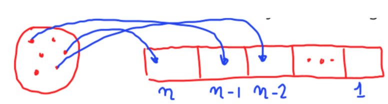


Figure 2: We can think of permutations of n elements as a number of different ways of placing unordered elements into a sequence, here represented by slots into which we place these elements.

So the overall number of ways that we can carry out this process, put the elements into the n slots, by the counting principle is going to be the product of the number of choices that we had at each one of the stages. So it's this product:

$$n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$$

.

And this product we denote as a shorthand this way,

$$n!$$

which we read as n factorial. $n!$ is the product of all integers from 1 all the way up to n. And in particular, the number of permutations of n elements is equal to $n!$.

EXAMPLE 4. Let us now consider another example. We start again with a general set, which consists of n elements. And we're interested in constructing a *subset* of that set. In how many different ways can we do that? How many different subsets are there?

Let us think of a sequential process through which we can choose the subset. The sequential process proceeds by considering each one of the elements of our set, one at a time. We first consider the first element, and here we have 2 choices: do we put it inside the set or not? So 2 choices for the first element.

Then we consider the second element. Again, we have 2 choices. Do we put it in the subset or not? We continue this way until we consider all the elements. There's n of them. And the overall number of choices that we have is the product of $2 \times 2 \times 2 \times \ldots \times 2$, this product taken n times (because we make that choice of either including or excluding an element for each of the n elements. This of course works out to be 2^n.

At this point, we can also do a sanity check to make sure that our answer is correct. Let us consider the simple and special case where n is equal to 1, which means we're starting with a set with 1 element and we want to find the number of subsets that it has. According to the answer that we derived, this should have 2^1, that is 2 subsets. Which ones are they?

One subset of this set is the set itself and the other subset is the empty set. So we do have, indeed, 2 subsets out of that set, which agrees with the answer that we found. Notice that when we count subsets of a given set, we count both the set itself, the whole set, and we also count the empty set.

Check Your Understanding: You are given the set of letters $\{A, B, C, D, E\}$.

1. How many three-letter strings (i.e., sequences of 3 letters) can be made out of these letters if each letter can be used only once?

2. How many subsets does the set $\{A, B, C, D, E\}$ have?

3. How many five-letter strings can be made if we require that each letter appears exactly once and the letters A and B are next to each other, as either "AB" or "BA"? (Hint: Think of a sequential way of producing such a string.)

1. There are 5 choices for the first letter, 4 choices for the second, and 3 for the last. Thus, the answer is $5 \cdot 4 \cdot 3 = 60$.

2. The number of subsets of a 5-element set is $2^5 = 32$.

3. We first choose whether the order will be "AB" or "BA" (2 choices). We then choose the position of the first letter in "AB" or "BA". There are 4 choices, namely positions 1, 2, 3, or 4. We are left with three positions in which the letters C, D, and E can be placed, in any order. The number of ways that this can be done is the number of permutations of these three letters, namely, $3! = 3 \cdot 2 \cdot 1 = 6$. Thus, the answer to this problem is $2 \cdot 4 \cdot 6 = 48$.

Another Example

We will now use counting to solve a simple probabilistic problem. We have in our hands an ordinary six-sided die which we are going to roll six times. We're *interested in the probability of the event that the six rolls result in different numbers*. So let us give a name to that event and call it event A:

We wish to calculate the $P(A)$. But before we can even get started answering this question, we need a probabilistic model. We need to make some assumptions, and the assumption that we're going to make is that all outcomes of this experiment are equally likely.

This is going to place us within a discrete uniform probabilistic model so that we can calculate probabilities by counting. In particular, as we discussed earlier, the probability of an event A is going to be the number of elements of the set A, the number of outcomes that make event A occur, divided by the total number of possible outcomes, which is the number of elements in our sample space

$$P(A) = \frac{\text{Number of Elements in } A}{\text{Total Number of Possible Outcomes}}$$

So let us start with the denominator, and let us look at the typical outcomes of this experiment. A typical outcome is something like this sequence, 2, 3, 4, 3, 6, 2. That's one possible outcome

How many outcomes of this kind are there? Well, we have 6 choices for the result of the first roll, 6 choices for the result of the second roll, and so on. And since we have a total of 6 rolls, this means that there is a total of 6 to the 6-th power possible outcomes, according to the *counting principle*. And since we have so many possible outcomes and we assume that they are equally likely, the probability of each one of them would be $1/6^6$

$$P(\text{ Sequence } \{2, 3, 4, 3, 6, 2\} \text{ occurs }) = \frac{1}{6^6}$$

(Incidentally, that's the same number you would get if you were to assume, instead of assuming directly that all outcomes are equally likely, to just assume that the different rolls are rolls of a fair six-sided die, so the probability of getting a 2, say, is 1/6, and also that the different rolls are independent of each other.

So in that case, the probability, let's say, of the particular sequence 2, 3, 4, 3, 6, 2 would be the probability of obtaining a 2, which is 1/6, times the probability that we get a 3 at the next roll, which is 1/6, times 1/6 times 1/6 and so on, and we get the same answer, $1/6^6$. So we see that this assumption of all outcomes being equally likely has an alternative interpretation in terms of having a fair die which is rolled independently 6 times.)

Now, let us look at the event of interest, A. What is a typical element of A? A typical element of A is a sequence of 6 rolls in which no number gets repeated. So, for example, it could be a sequence of results of this kind: 2, 3, 4, 1, 6, 5. So all the numbers appear exactly once in this sequence.

So to compute the number of outcomes that make event A happen, we basically need to count the number of permutations of the numbers 1 up to 6. These 6 numbers can appear in an arbitrary order. In how many ways can we order 6 elements? As discussed earlier, this is equal to 6!. So we have now the probability of event A:

$$P(A) = \frac{6!}{6^6}$$

Check Your Understanding: You are given the set of letters $\{A, B, C, D, E\}$. What is the probability that in a random five-letter string in which each letter appears exactly once, and with all such strings equally likely, the letters A and B are next to each other? The answer to a previous exercise may also be useful here.

From the previous exercise, the event of interest has 48 elements. The sample space has 5!=120 elements. Thus, the desired probability is 48/120=2/5=0.4.

Combinations

Let us now study a very important counting problem, the problem of counting combinations. What is a combination? We start with a set of n elements. We are also given a non-negative integer k. We want to construct or to choose a subset of the original set that has exactly k elements. In different language, we want to *pick a combination of* k *elements of the original set*. In how many ways can this be done?

Figure 3: We can think of combinations as number of different ways in which you we can choose a subset of size k from a set of size n.

Let us introduce some notation. We use the notation $\binom{n}{k}$, which we read as "n-choose-k," to denote exactly the quantity that we want to calculate, namely the number of subsets of a given n-element set, where we only count those subsets that have exactly k elements.

$\binom{n}{k}$: number of k-element subsets of a given n-element set

How are we going to calculate this quantity?

Instead of proceeding directly, we're going to consider a somewhat different counting problem which we're going to approach in two different ways, get two different answers, compare those answers, and by comparing them, get an equation which is going to give us the desired answer.

We start, as before, with our given set that consists of n elements. But instead of picking a subset, what we want to do is to construct a list, an ordered sequence, that consists of k distinct elements taken out of the original set. So we think of having k different slots, and we want to fill each one of those slots with one of the elements of the original set. In how many ways can this be done?

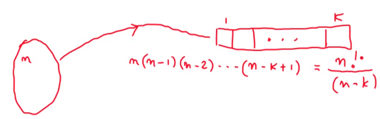


Figure 4: We fill the list of k slots with elements from a set of size n, one at time. Any of the n items can go into the first position in the list, then any of the $n-1$ remaining items can go into the second position, and so on, all the way down to the k-th position in the list, where any of the remaining $(n-k+1)$ remaining items can go. We then simply rewrite this product as $\frac{n!}{(n-k)!}$.

Well, we want to use the counting principle, so we want to decompose this problem into stages. We choose each one of the k items that go into this list one at a time. We first choose an item that goes to the first position, into the first slot. Having used one of the items in that set, we're left with $n-1$ choices for the item that can go into the second slot.

And we continue similarly. When we're ready to fill the last slot, we have already used $k-1$ of the items, which means that the number of choices that we're going to have at that stage is $n-k+1$. Thus, the number of ways to fill these k slots is

$$n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$$

At this point, it's also useful to simplify that expression a bit. We observe that this is the same as $\frac{n!}{(n-k)!}$. (Why is this the case? You can verify this is correct by moving the denominator to the other side. When you do that you realize that you have the product of all terms from n down to $n-k+1$. And then you have the product of $n-k$ going all the way down to 1. And that's exactly the product, which is the same as $n!$.)

$$\begin{aligned} \text{Number of different ways to fill the list of } k \text{ elements} &= \\ &= n \cdot (n-1) \cdot \ldots \cdot (n-k+1) \\ &= \frac{n!}{(n-k)!} \end{aligned}$$

So this was the first method of constructing the list that we wanted. How about the second method?

What we can do is to first choose k items out of the original set, and then take those k items and order them in a sequence to obtain an ordered list.

We construct our ordered list in two stages. In the first stage, how many choices do we have? That's the number of subsets with k elements out of the original set. We don't know what this number is - that's what we're trying to calculate. But we have a symbol for it: it's $\binom{n}{k}$.

How about the second stage? We have k elements, and we want to arrange them in a sequence. That is, we want to form a permutation of those k elements. This is a problem that we have already studied, and we know that the answer is $k!$. According to the counting principle, the number of ways that this two-stage construction can be made is equal to the product of the number of ways that we have in the first stage times the number of options that we have in the second stage:

$$\text{Number of different ways to fill the list of } k \text{ elements} = \binom{n}{k} \times k!$$

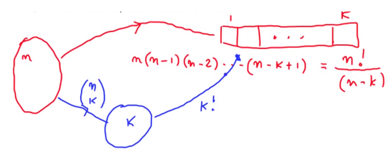


Figure 5: The second way of building the list is shown in blue. We first choose k items out of the original set - we don't yet know what that number is, but we have a symbol for it: $\binom{n}{k}$ - and then take those k items and order them in a sequence to obtain an ordered list.

So we have two different answers for the number of possible ordered sequences. Of course, both of them are correct. And therefore, they have to be equal.

$$\begin{aligned} \binom{n}{k} k! &= \frac{n!}{(n-k)!} \\ \Longrightarrow \binom{n}{k} &= \frac{n!}{(n-k)!\, k!} \end{aligned}$$

Now, this last formula is valid only for numbers that make sense: while n can be any non-negative integer, the only k's that make sense would be k's from 0, 1 up to n.

You may be wondering about some of the extreme cases of that formula. What does it mean for n to be 0 or for k to be equal to 0? So let us consider now some of these extreme cases and make a sanity check about this formula.

The first case to consider is the extreme case of $\binom{n}{n}$. What does that correspond to? Out of a set with n elements, we want to choose a subset that has n elements. There's not much of a choice here. We just have to take all of the elements of the original set and put them in the subset. The subset is the same as the set itself. So we only have one choice here - 1 should be the answer:

$$\binom{n}{n} = 1$$

Let's check it with the formula:

$$\binom{n}{n} = \frac{n!}{n!\,0!} = 1$$

Is this correct? Well, it becomes correct as long as we adopt the convention that $0! = 1$. We're going to adopt this convention and keep it throughout this course.

Let's look at another extreme case now, the coefficient $\binom{n}{0}$. This time let us start from the formula. Using the convention that we have, this is equal to 1:

$$\binom{n}{0} = \frac{n!}{0!\,n!} = 1$$

Is it the correct answer? How many subsets of a given set are there that have exactly zero elements? Well, there's only one subset that has exactly 0 elements, and this is the empty set, $\emptyset$.

Now, let us use our understanding of those coefficients to solve a somewhat harder problem. Suppose that for some reason, you want to calculate this sum:

$$\sum_{k=1}^{n} \binom{n}{k}$$

How do we compute this sum? One way would be to use the formula for these individual terms in the sum, do a lot of algebra.

And if you're really patient and careful, eventually you should be able to get the right answer.

But this would be very painful!

Let us think whether there's a clever way, a shortcut, of obtaining this answer. Let us try to think what this sum is all about. This sum includes the term, $\binom{n}{0}$, which is the number of zero-element subsets. The sum also includes the term $\binom{n}{1}$, which is the number of subsets that have one element. And we keep going all the way to the number of subsets that have exactly n elements. So just by rewriting the sum, we have:

$$\sum_{k=0}^{n} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n} =$$
$$= \text{Number of all subsets of a set of } n \text{ elements.}$$

So we're counting zero-element subsets, one-element subsets, all the way up to n element subsets. But this is really the number of *all* subsets of our given set, and this is a number that we already know! It is the number of subsets of a given set with n elements and is equal to 2^n:

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n$$

So by thinking carefully and interpreting the terms in this sum, we were able to solve this problem very fast, something that would be extremely tedious if we had tried to do it algebraically. For some practice with this idea, why don't you pause at this point and try to solve a problem of a similar nature?

Check Your Understanding: Counting committees.

We start with a pool of n people. A chaired committee consists of $k \geq 1$ members, out of whom one member is designated as the chairperson. The expression $k\binom{n}{k}$ can be interpreted as the number of possible chaired committees with k members. This is because we have $\binom{n}{k}$ choices for the k members, and once the members are chosen, there are then k choices for the chairperson. Thus,

$$c = \sum_{k=1}^{n} k \binom{n}{k}$$

is the total number of possible chaired committees of any size.

Find the value of c (as a function of n) by thinking about a different way of forming a chaired committee: first choose the chairperson, then choose the other members of the committee. The answer is of the form

$$c = (\alpha + n^{\beta}) 2^{\gamma n + \delta}.$$

What are the values of $\alpha, \beta, \gamma, \delta$?

The answer is $\alpha = 0, \beta = 1, \gamma = 1, \delta = -1$. We first choose the chairperson, for which there are n choices, and then choose an arbitrary subset of the remaining $n-1$ people, who will be the remaining commitee members. For example, this arbitrary subset could be the empty set, which would mean that the commitee is of size 1: only the chairperson. There are 2^{n-1} possible subsets of a set with $n - 1$ elements, and so there are 2^{n-1} ways of choosing the remaining commitee members. Thus, an alternative expression for the number of possible chaired commitees of any size is $n2^{n-1}$, from which we can extract the values of α, β, γ, and δ.

Binomial Probabilities

The coefficients $\binom{n}{k}$ that we calculated in the previous segment are known as the binomial coefficients. They are intimately related to certain probabilities associated with coin tossing models, the so-called binomial probabilities.

CONSIDER a coin which we toss n times in a row, independently. For each one of the tosses of this coin, we assume that there is a certain probability, p, that the result is heads, which of course implies that the probability of obtaining tails in any particular toss is going to be $1 - p$.

Assumptions:
- independence
- $\mathbf{P}(H) = p$

The question we want to address is the following. We want to calculate the probability that in those n independent coin tosses, we're going to observe exactly k heads.

$$P(\text{Exactly } k \text{ heads in } n \text{ tosses of a fair coin}) = ?$$

Let us start working our way towards the solution to this problem by looking first at a simple setting.

Let us answer this question first: what is the probability that we observe the particular sequence $HTTHHH$? What is:

$$P(HTTHHH) = ?$$

Of course here we take $n = 6$, and we wish to calculate the above probability. Now, because we have assumed that the coin tosses are independent, we can multiply probabilities. So the probability of this sequence is

$$\begin{aligned} P(HTTHHH) &= P(H)\ P(T)\ P(T)\ P(H)\ P(H)\ P(H) \\ &= p\ (1-p)\ (1-p)\ p\ p\ p \\ &= 4\ p\ (1-p)^2 \end{aligned}$$

More generally, if I give you a particular sequence of heads and tails, as in this example, and I ask you, what is the probability that this particular sequence is observed, then by generalizing from this answer you see that you're going to get p to the power number of heads and then there are factors associated with tails, each tail contributing a factor of $1 - p$, and so we're going to have here $1 - p$ to a power equal to the number of tails.

$$P(\text{particular sequence}) = p^{\text{number of heads}}\ (1-p)^{\text{number of tails}}$$

NOW, if I ask you about the probability of a particular sequence and that particular sequence has happened to have exactly k heads, what is the probability of that sequence?

Well, we already calculated what it is. It is the previous answer, except we use the symbol k instead of just writing out explicitly "number of heads." And the number of tails is the number of tosses minus how many tosses resulted in heads.

$$P(\text{particular } k\text{-head sequence}) = p^k\ (1-p)^{n-k}$$

Now, we're ready to consider the actual problem that we want to solve, which is calculate the probability of k heads.

The event of obtaining k heads can happen in many different ways. Any particular k-head sequence makes that event to occur. The overall probability of k heads is going to be the probability of any particular k-head sequence, times the number of k-head sequences that we have:

$$P(k \text{ heads}) = p^k\ (1-p)^{n-k} \times \text{total number of } k\text{-head sequences}$$

(As an aside, you may note that the reason why we can carry out this argument is the fact that any k-head sequence has the same probability. Otherwise, we wouldn't be able to write down an answer which is just the product of two terms. But because every k-head sequence has the same probability, to find the overall probability, we take the probability of each one of them and multiply it with the number of how many of these we have.)

To make further progress, now we need to calculate the number of possible k-head sequences. How many are there?

Well, specifying a k-head sequence is the same as the following. You think of having n time slots. These time slots corresponds to the different tosses of your coin. And to specify a k-head sequence, you need to tell me which ones of these slots happen to contain a head, and there should be k of them, of course.

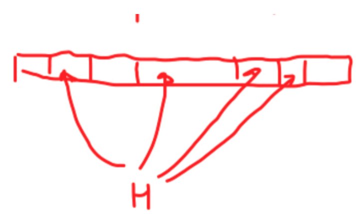


Figure 6: Think of having a sequence of n time slots, each corresponding to the different tosses of a coin. To specify a k-head sequence, you just identify which ones of these slots happen to contain a head.

In other words, what you're doing is you're specifying a subset of the set of these n slots, a subset that has k elements. You need to choose k of the slots out of the n and tell me that those k slots have heads. That's the way of specifying a particular k-head sequence.

So what's the number of k-head sequences? Well, it's the same as the number of ways that you can choose k slots out of the n slots, which is our binomial coefficient, $\binom{n}{k}$. Therefore, the answer to our problem is this expression:

$$\mathbf{P}(k \text{ heads}) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\begin{aligned} P(k \text{ heads}) &= \binom{n}{k} p^k\ (1-p)^{n-k} \\ &= \frac{n!}{k!\ (n-k)!}\ p^k\ (1-p)^{n-k} \end{aligned}$$

At this point, pause and consider a simple question to check your understanding of the binomial probabilities.

Check Your Understanding: Binomial probabilities.

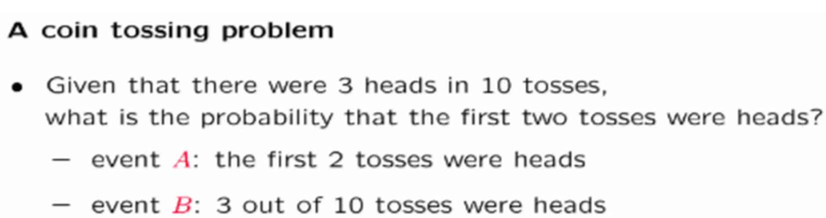
Recall that the probability of obtaining k k Heads in n independent coin tosses is $\binom{n}{k} p^k \ (1-p)^{n-k}$, where p is the probability of Heads for any given coin toss. Find the value of $\sum_{k=0}^{n} \binom{n}{k} p^k \ (1-p)^{n-k}$.

(Your answer should be a number.)

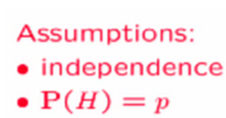
A coin tossing example

Let us now put to use our understanding of the coin-tossing model and the associated binomial probabilities. We will solve the following problem. We have a coin, which is tossed 10 times. And we're told that exactly 3 out of the 10 tosses resulted in heads. Given this information, we would like to calculate the probability that the first two tosses were heads:

This is a question of calculating a conditional probability of one event given another. The conditional probability of event A, namely that the first two tosses were heads, given that another event B has occurred, namely that we had exactly three heads out of the 10 tosses. However, before we can start working towards the solution to this problem, we need to specify a probability model that we will be working with.



We need to be explicit about our assumptions. To this effect, let us introduce the following assumptions. We will assume that the different coin tosses are independent. In addition, we will assume that each coin toss has a fixed probability, p, the same for each toss, that the particular toss results in heads. These are the exact same assumptions that we made earlier when we derived the binomial probabilities.



In the last section, we came up with the following formula that if we have n tosses, the probability that we obtain exactly k heads is given by this expression:

$$P(\text{exactly } k \text{ heads in } n \text{ tosses}) = \binom{n}{k} p^k \ (1-p)^{n-k}$$

So now, we have a model in place and also the tools that we can use to analyze this particular model. Let us start working towards a solution. Actually, we will develop two different solutions and compare them at the end.

The first approach, which is the safest one, is the following. Since we want to calculate a conditional probability, let us just start with the definition of conditional probabilities. The definition is shown again in Figure 7.

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

Figure 7: Recall the definition of the conditional probability of an event A given another event B.

Recall, in the numerator, we're talking about the probability that event A happens *and* event B happens. What does that mean? This means that event A happens – that is, the first two tosses resulted in heads, which I'm going to denote symbolically as H_1H_2 – *and* in addition to that event B happens, where event B requires that there is a total of three heads. This means that we had just one additional head in the remaining eight tosses; in other words, we have exactly one head in tosses 3, 4, all the way to 10.

So let's write this in the numerator of the following expression:

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{P(H_1H_2 \text{ and one } H \text{ in tosses } 3, \ \ldots \ ,10)}{P(B)} \end{aligned}$$

Now, here comes the independence assumption. Because the different tosses are independent, whatever happens in the first two tosses is independent from whatever happened in tosses 3 up to 10. So the probability of events A and B happening is the product of their individual probabilities. So we first have the probability that the first two tosses were heads, which is p^2. Now we multiply this p^2 with the probability that there was exactly one head in the tosses numbered from 3 up to 10. These are eight tosses. The probability of one head in eight tosses is given by the binomial formula with $k = 1$ and $n = 8$. So let's plug it in and rewrite the above expression:

$$
\begin{aligned}
P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(H_1 H_2 \text{ and one } H \text{ in tosses } 3, \ \dots \ ,10)}{P(B)} \\
&= \frac{p^2 \binom{8}{1} \ p^1 \ (1-p)^7}{P(B)}
\end{aligned}
$$

The denominator is easier to find. This is the probability that we had three heads in 10 tosses. So we just apply the binomial formula again.

$$P(B) = \binom{8}{1} p^3 \ (1-p)^7$$

Or, to recap what we have so far:

$$P(A \mid B) = \frac{p^2 \binom{8}{1} \ p^1 \ (1-p)^7}{\binom{10}{3} p^3 \ (1-p)^7}$$

And here we notice that terms in the numerator and denominator cancel out, and we obtain

$$
\begin{aligned}
P(A \mid B) &= \frac{p^2 \binom{8}{1} \ p^1 \ (1-p)^7}{\binom{10}{3} p^3 \ (1-p)^7} \\
&= \frac{\binom{8}{1}}{\binom{10}{3}} \\
&= \frac{8}{\binom{10}{3}}
\end{aligned}
$$

So this is the answer to the question.

Now let us work towards developing a *second approach* towards this particular answer. In our second approach, we start first by looking at the sample space and understanding what conditioning is all about. As usual, we denote the sample space by Ω. As usual, Ω contains a bunch of possible outcomes. A typical outcome is going to be a sequence of heads or tails that has length 10.

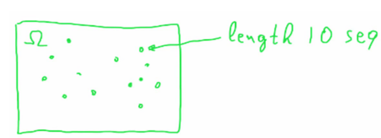


Figure 8: The outcomes in our Ω are sequences of heads or tails of length 10.

We want to calculate conditional probabilities. And this places us in a conditional universe. We have the conditioning event B, which is some set. A typical element of the set B is a sequence which is, again, of length 10, but has exactly three heads.

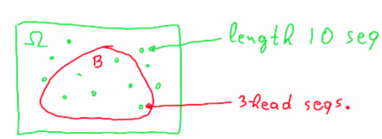


Figure 9: Within Ω we identify event B which consists of sequences of length 10 that have exactly three heads.

Now, since we're conditioning on event B, we can just work with conditional probabilities. Recall that any three-head sequence has the same probability of occurring in the original unconditional probability model, namely as we discussed earlier, any particular three-head sequence has a probability equal to

$$P(\text{ any particular three-head sequence }) = p^3\ (1-p)^7$$

So three-head sequences are all equally likely. This means that the *unconditional* probabilities of all the elements of B are the same. Moreover, when we construct *conditional* probabilities given an event B, what happens is that the ratio (or the relative proportions) of the probabilities of elements in B remain the same. So conditional probabilities are proportional to unconditional probabilities; the elements of B were equally likely in the original model, therefore, they remain equally likely in the conditional model as well.

In short, this means that given that B occurred, all the possible outcomes now have the same probability and thus we can now answer probability questions by just counting.

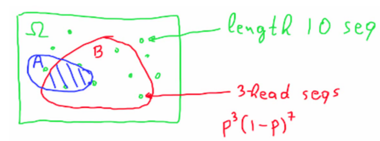


Figure 10: We now add event A to the picture where elements inside B are all equally likely.

Now, consider figure 10. We're interested in the probability of a certain event, A, given that B occurred. So we're interested in the probability of outcomes that belong in the blue shaded region as a proportion of those outcomes that belong within the set B. In short, we just need to count how many outcomes belong to the shaded region and divide them by the number of outcomes that belong to the set B.

How many elements are there in the intersection of A and B? These are the outcomes or sequences of length 10, in which the first two tosses were heads – no choice here - and there is one more head. That additional head can appear in one out of eight possible places. So there's eight possible sequences that have the desired property.

$$\text{Number of elements in the blue shaded region} = 8$$

How many elements are there in the set B? How many three-head sequences are there? Well, the number of three-head sequences is the same as the number of ways that we can choose three elements out of a set of cardinality 10. And this is $\binom{10}{3}$, as we discussed earlier.

So, final answer must be this:

$$P(A \mid B) = \frac{8}{\binom{10}{3}}$$

And this is the same answer as we derived before with our first approach. Both approaches, of course, give the same solution. This second approach is a little easier, because we never had to involve any p's in our calculation.

Check Your Understanding: Coin tossing
Use the second method in the preceding segment to find the probability that the 6-th toss out of a total of 10 tosses is Heads, given that there are exactly 2 Heads out of the 10 tosses. As in the preceding segment, continue to assume that all coin tosses are independent and that each coin toss has the same fixed probability of Heads.

The conditional universe consists of sequences of length 10 that contain exactly 2 Heads. There are

$$\binom{10}{2} = \frac{10!}{8!\,2!} = \frac{10 \cdot 9}{2} = 45$$

such sequences. Out of these 45 sequences, how many have the property that the 6-th toss was Heads? There are 9 sequences with this property: the 6-th toss is fixed to be Heads, and the other Heads can be any of the remaining 9 tosses. Therefore, the desired conditional probability is 9/45=1/5.

Partitions (optional, may skip on first reading)

We now come to our last major class of counting problems. We will count the number of ways that a given set can be partitioned into pieces of given sizes. We start with a set that consists of n different elements, and we have r persons.

We want to give n_1 items to the first person, give n_2 items to the second person, and so on. And finally, we want to give n_r items to the r-th person. These numbers, n_1, n_2, up to n_r are given to us. And these numbers must add up to n so that every item in the original set is given to some person.

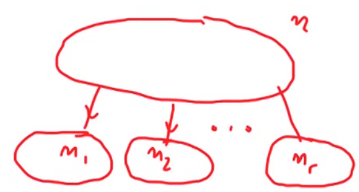


Figure 11: We would like to distribute n items among r persons.

We want to count to the number of ways that this can be done. This is the number of ways that we can partition a given set into subsets of prescribed sizes. Let's use c to denote the number of ways this can be done. We want to calculate c.

Instead of calculating directly, we're going to use the same trick that we employed when we counted combinations and derived the binomial coefficient. That is, we're going to consider, in a much simpler counting problem, the problem of ordering n items, taking the n items in our original set and putting them in an ordered list.

Of course, we know in how many ways this can be done. Ordering n items can be done in $n!$ ways. This is the count of the number of permutations of n items. But now let us think of a different way of ordering the n items, an indirect way.

It proceeds according to the following stages. We start with the n items. And we first distribute them to the different persons. Having done that, then we ask person one to take their items, order them, and put them in the first n_1 slots of our list.

Then person two takes their items and puts them into the next n_2 slots in our list. We continue this way. And finally, the last person takes the items that they possess and puts them in the last n_r slots in this list.

In how many ways can this process be carried out? We have c choices on how to partition the given set into subsets. Then person one has $n_1!$ choices on how to order the n_1 items that that person processes.

Person two has $n_2!$ factorial choices for how to order the n_2 items that it possesses, and so on until the last person, who has $n_r!$ factorial choices for ordering their elements. This multi-stage process results in an ordered list of the n items. So the number of ways this multi-stage process can be carried out is

$$c\ n_1!\ n_2! \ldots n_r!$$

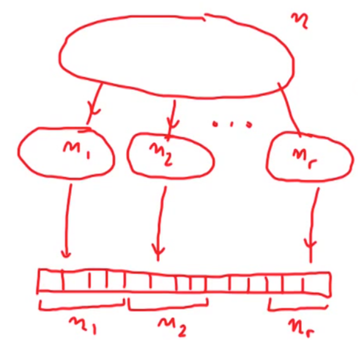


Figure 12: The i-th person takes her n_i items and orders them in the n_i slots alloted to her.

On the other hand, we know that the number of possible orderings of the items is $n!$, so we have this equality:

$$n! = c\ n_1!\ n_2! \ldots n_r!$$

Solving this for c gives

$$c = \frac{n!}{n_1!\ n_2! \ldots n_r!}$$

This particular expression is called the multinomial coefficient, and it generalizes the binomial coefficient. The binomial coefficient was referring to the case where we essentially split our set into one subset with k elements, and then the second subset gets the remaining elements. So the special case where $r = 2$, and $n_1 = k, n_2 = n - k$, this corresponds to a partition of a set into two subsets, or what is the same just selecting the first subset and putting everything else in the second subset. And you can check that in this particular case, the expression for the multinomial coefficient agrees with the expression that we had derived for the binomial coefficient:

$$c = \frac{n!}{k!\ (n-k)!} \quad \text{for the special case when } r = 2.$$

Check Your Understanding: Counting partitions.

We have 9 distinct items and three persons. Alice is to get 2 items, Bob is to get 3 items, and Charlie is to get 4 items.

1. As just discussed, this can be done in $\frac{a!}{b!\ 3!\ 4!}$ ways. Find a and b.

2. A different way of generating the desired partition is as follows. We first choose 2 items to give to Alice. This can be done in $\binom{c}{d}$ different ways. Find c and d. (There are 2 possible values of d that are correct. Find the smaller value.)

3. Having given 2 items to the Alice, we now give 3 items to Bob. This can be done in $\binom{e}{f}$ ways. Find e and f. (There are 2 possible values of f that are correct. Find the smaller value.)

Verify that the answer from part 1 agrees with the answer that you get by combining parts 2 and 3.

1. By the multinomial formula, $a = 9$ and $b = 2$.
2. We want the number of ways of choosing 2 items out of 9 items. This is the number of 2-element subsets of a 9-element set, so that $c = 9$ and $d = 2$.
3. We have 7 remaining items out of which we need to choose 3. Hence, $e = 7$ and $f = 3$.

From part 1, the number of ways of splitting up the 9 items between Alice, Bob, and Charlie in the specified manner is $\frac{9}{2!\,3!\,4!}$.
In parts 2 and 3, we calculate this answer in a different way. Let us now verify that the two methods produce the same answer.
From part 2, we can first give Alice her 2 items in $\binom{9}{2} = \frac{9!}{2!\,7!}$ ways. Then, from part 3, we can give Bob his 3 items from the remaining 7 items in $\binom{7}{3} = \frac{7!}{3!\,4!}$ ways. Finally, Charlie's 4 items are exactly the 4 items that remain, so there is only 1 way to give him his items. Combining these steps, we have a total of

$$\frac{9!}{2!\,7!} \cdot \binom{7}{3} = \frac{7!}{3!\,4!} \cdot 1 = \frac{9}{2!\,3!\,4!}$$

ways, which agrees with the answer from part 1.

Each person gets an ace (optional, may skip on first reading)

We will now apply our multinomial formula for counting the number of partitions to solve the following probability problem. We have a standard 52-card deck, which we deal to four persons. Each person gets 13 cards as, for example, in bridge. What is the probability that each person gets exactly one ace?

Well, before we start, as always, we will need a probability model. We deal the cards fairly, and this is going to be our model. But we still need to interpret our statement.

To give this interpretation, let us first think of the outcomes of the experiment. What are the possible outcomes? An outcome of this experiment is a partition of the 52 cards into the four persons so that each person gets exactly 13 cards. Our statement about dealing the cards fairly will be an *assumption* that all partitions are equally likely.

We make the assumption that all partitions of the 52 cards into 4 parts of 13 cards each are equally likely.

So since all partitions, all outcomes of the experiment, are equally likely, this means that we can solve a probability question by counting. We need to count the number of elements of our sample space, the number of possible outcomes, and then count the number of outcomes that make the event of interest to occur.

Let us start with the number of elements of the sample space. This is the problem that we just dealt with a little while ago – the number of outcomes, the number of partitions of 52 items into four persons, where we give 13 cards to person one, 13 cards to person two, 13 cards to person three, and 13 cards to person four. The number of possible ways of doing this is equal to this multinomial coefficient:

$$\frac{52!}{13!\ 13!\ 13!\ 13!}$$

So now let us count the number of outcomes that belong to the event of interest, namely the outcomes where each person gets an ace. We think of the process of constructing such an outcome as a multistage process. And we count the number of choices that we have at each stage.

The process is as follows. We first distribute the four aces. We take the ace of spades and give it to one person. In how many ways can we do it? We can do it in four ways.

Then we take the next ace. The next ace must be given to a different person. And so at that stage, we have three different choices about who to give that ace to.

$4 \cdot 3 \cdot 2 \cdot 1 = 24$ is the number of ways to distribute the 4 aces among the 4 players.

Then we consider the next ace. At this point, two persons already have aces. So we have two available choices for who can get the next ace. And finally for the last ace, we do not have any choice. We give it to the only remaining person who doesn't yet have an ace.

Having distributed the four aces, then we need to somehow distribute the remaining 48 cards to the four people. But we can do that in any way we want. So all we need to do is to just partition the 48 cards into four subsets of given cardinalities. And this can be done by a number of ways, which is the number of such partitions. We have already found what that number is. It is

$$\frac{48!}{12!\ 12!\ 12!\ 12!}$$

.

$48!/(12!\ 12!\ 12!\ 12!)$ is the number of ways to distribute the remaining 48 cards.

So the number of ways that we can distribute the cards so that each person gets an ace, according to the counting principle, is going to be the number of ways that we can distribute the aces times the number of ways that we can distribute the remaining cards. The product of this number gives us the count, gives us the cardinality, of the event of interest.

We also have the cardinality of the sample space. So the desired probability can be found by dividing these two numbers. And the final answer takes this form:

$$\frac{4 \cdot 3 \cdot 2 \cdot \frac{48!}{12!\ 12!\ 12!\ 12!}}{\frac{52!}{13!\ 13!\ 13!\ 13!}}$$

Let us now look at the same problem but in a different way. Probability problems can often be solved in multiple ways, and some can be faster than others. Is there a solution here that will get us to the desired answer faster?

We will use the following trick. We will think about a very specific way of dealing the cards which is the following. We take the 52 cards and stack it so that the four aces are at the top. So they are first.

And then we deal those cards to the players as follows. We think of each player having 13 slots of his own. And the cards will be placed randomly into the different slots.

So we can do this one card at a time, starting from the top. We take the first ace and send it to a random location. Then we will take the second ace, send it to a random location, and so on.

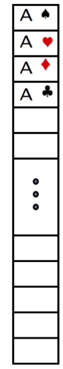


Figure 13: We stack the deck, aces on top.

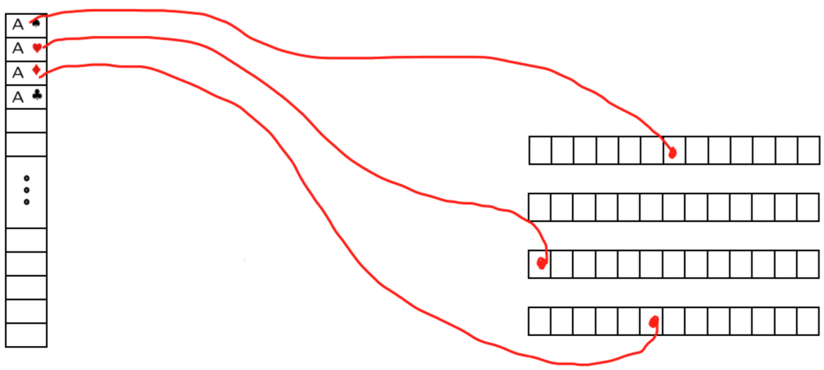


Figure 14: Each ace can go into any of the 13 slots for the player dealt the ace.

What we want to calculate is the probability that the four aces will end up in slots that are associated with different persons. So let us calculate this probability.

The first ace, the ace of spades, can go anywhere. It doesn't matter.

The second ace, the ace of hearts, has 51 slots to choose from. It's 51 because we started with 52, but one slot has already been taken by the ace of spades.

So for the ace of hearts, we have 51 slots that it can go to. And out of those 51, we have 39 that belong to people who do not yet have an ace. So the probability that the ace of hearts gets placed into a slot that belongs to a person who is different than the person who got the first ace is

$$\frac{39}{51}$$

Now let us consider the ace of diamonds. What is the probability that this ace will get into a slot which belongs to either of the remaining two persons? There are 26 slots out of the 50 available slots:

$$\frac{26}{50}$$

Finally, let us consider the ace of clubs. So having placed that ace and assuming that it got to a different person, what is the probability now that this ace is going to go to the fourth person who doesn't yet have an ace? The probability of this happening is the number of slots associated with that person, which is equal to 13 divided by the number of slots that this card can choose from. And the number of slots is 52 minus the 3 slots that have already been taken, that is, 49:

$$\frac{13}{49}$$

And so the answer to our problem is this:

$$\frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} = 0.105.$$

This expression looks very different from the expression that we derived a little earlier. But you do the arithmetic and simplify the answer, you will be able to verify that indeed it's exactly the same answer we got before. So there's about 10% chance that when you deal the cards in bridge, each one of theplayers is going to end up having exactly one ace.

The second way was a faster way of getting to the answer to our problem, compared to the first one. But it raises a legitimate question. Is the way that we dealt the cards by putting the aces on top and then dealing them a fair way of dealing the cards? Is it true that with this way of dealing the cards all partitions are equally likely?

It turns out that this is indeed the case. But it does require a bit of thinking. Maybe you can see it intuitively that this is the case. But if not, then it is something that one can prove.

It can be proved formally as follows. One first needs to check that all permutations, that is all possible allocations of cards into slots, are equally likely. And because of this, one can then argue that any possible partition into subsets of 13 is also equally likely.

This is an equivalent way of dealing the cards to the one that we considered earlier, which was that every partition is equally likely. Therefore, we did indeed solve the same problem, and so this is a legitimate alternative way of getting to the answer.