

An Exam P Study Guide

Actuary Helper

2020-01-12

Contents

1	An Introduction to Set Theory	1
1.1	Defining Sets	1
1.2	Set Equality and Subsets	2
1.3	Set Operations and Venn Diagrams	2
1.4	Special Sets and Identities	4
2	Counting	7
2.1	Size of a set	7
2.2	Principle of Inclusion-Exclusion	7
2.3	Multiplication Principle	8
2.4	Permutations	9
2.5	Combinations	10
2.6	Variations	11
3	Probability Basics	13
3.1	Equally Likely Events	13
3.2	Terminology	14
3.3	Probability Functions	15
3.4	Conditional Probability and Independence	17
3.5	Bayes Theorem	19

4	Random Variables	23
4.1	What is a Random Variable	23
4.2	Probability Distribution Functions	24
4.3	Cumulative Distribution Functions	25
4.4	Discrete vs. Continuous Random Variables	27
5	Functions of Discrete Random Variables	31
5.1	What is the Mean?	31
5.2	What is Variance?	35
6	Common Discrete Distributions	39
6.1	Bernoulli Distribution	39
6.2	Binomial Distribution	39
6.3	Hypergeometric Distribution	42
6.4	Poisson Distribution	44
6.5	Geometric Distribution	46
6.6	Negative Binomial Distribution	47

Chapter 1

An Introduction to Set Theory

1.1 Defining Sets

Sets are collections of objects. Usually objects are placed inside curly braces and separated by commas. Here is the set with the numbers 1 and 2. We often give sets a name. This set is named “A”.

$$A = \{1, 2\}$$

Objects inside the curly braces are called elements of the set. There is a special character that means “is an element of”. Using our notation we can say that 1 is an element of A.

$$1 \in A$$

We can also say that 3 is not an element of A by drawing a slash through our symbol.

$$3 \notin A$$

We can define sets in a something called set-builder notation. Set builder notation is useful for infinite sets or sets that are hard to enumerate. Below are some examples.

We can define the even numbers as $\{x | x \text{ is an even number}\}$

We can define hands of cards as $\{x | x \text{ is a 5 element subset of a deck of cards}\}$

To read set builder notation we translate the “|” as “where”. So the real numbers are “the set of x where x is a real number”.

1.2 Set Equality and Subsets

Sets are equal when they have the same elements. This means order doesn't matter in sets, $\{1, 2\} = \{2, 1\}$ because they have the same elements. Also, $\{1, 1, 2\} = \{1, 2\}$ and we say that sets don't have duplicate elements because duplicate elements have no purpose.

A set is a subset of another set when it fits inside of it. There is a symbol that looks like \leq that means "is a subset of". For example $\{1, 2\} \subseteq \{1, 2, 3\}$. It is also true that any set is a subset of itself, so $\{1, 2\} \subseteq \{1, 2\}$. In more precise terms $A \subseteq B$ if and only if every element of A is also an element of B.

A common technique to prove that two sets are equal is to show that $A \subseteq B$ and $B \subseteq A$. This means there is no element in either set that does not belong to the other and the sets are equal.

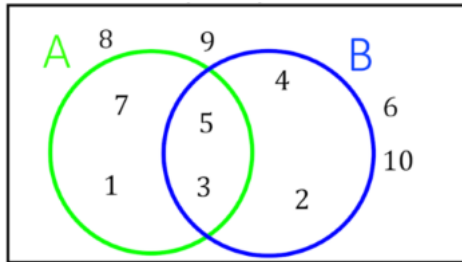
1.3 Set Operations and Venn Diagrams

Just like you can add numbers together to make a new number, you can combine sets and make a new set. Let's define some sets.

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} A = \{1, 3, 5, 7\} B = \{2, 3, 4, 5\}$$

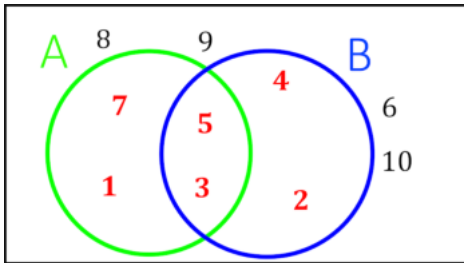
1.3.1 Venn Diagram

There is a visual representation of sets called a Venn Diagram. In a Venn Diagram each set is represented by a circle. The sample space is usually represented by a large box that the circles are inside of.



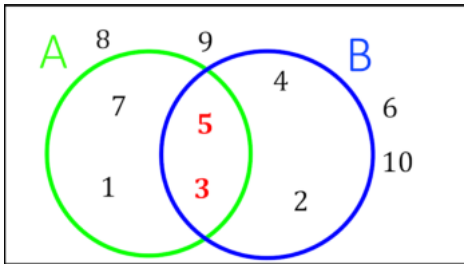
1.3.2 Union

An element is in the union of A and B if it is in A, B, or A and B. The notation for this operation is $A \cup B$ and it is pronounced "A union B". Here is the set that results from this union: $A \cup B = \{1, 2, 3, 4, 5, 7\}$



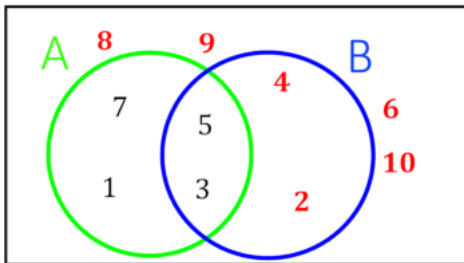
1.3.3 Intersection

An element is in the intersection of A and B if it is in A and B. The notation for this operation is $A \cap B$ and it is pronounced “A intersect B”. Here is the set that results from this intersection: $A \cap B = \{3, 5\}$



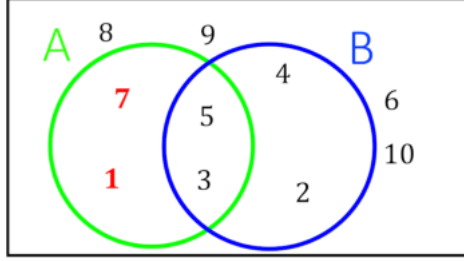
1.3.4 Complement

An element is in the complement of A if it is not in A, but is in the sample space. The notation for the complement of A is A^C and is pronounced “A complement” or “the complement of A”. Here is the result of taking the complement of A: $A^C = \{2, 4, 6, 8, 9, 10\}$. Note that in general $(A^C)^C = A$.



1.3.5 Set Difference

The set difference of A and B elements that are in A but not in B. The notation is either $A \setminus B$ or $A - B$ and is pronounced “A minus B”. Here is the result of this set difference: $A - B = \{7, 1\}$. It is worth noting that $A - B = A \cap B^C$ because the elements are in A and they are not in B.



1.4 Special Sets and Identities

1.4.1 Sample Space and Empty Set

The sample space is a type of special set because all other sets are subset of it. This leads to some commonsensical identities like $A \cup S = S$ and $A \cap S = A$. There is another special set called the empty set. This set has no elements. We could write it as $\{\}$ but the common way to write this is \emptyset . Some properties of the empty set are $A \cup \emptyset = A$ and $A \cap \emptyset = \emptyset$. It is also true that $\emptyset^C = S$ and $S^C = \emptyset$.

1.4.2 De Morgan's Laws

De Morgan's laws are formulas for the complement of a union or intersection of sets.

$$(A \cap B)^C = A^C \cup B^C \quad (A \cup B)^C = A^C \cap B^C$$

One way of memorizing these formulas is that you bring the complement inside the parentheses to both sets and flip the union or intersection upside down.

Let's see if these laws make any sense. Let T be the set of tennis players and H be the set of hockey players. $(T \cap H)^C$ is the set of people that don't play both tennis and hockey. $T^C \cup H^C$ is people that don't play tennis or they don't play hockey. If I don't play both sports then I either don't play tennis or I don't play hockey, so $(T \cap H)^C \subseteq T^C \cup H^C$. If I don't play tennis or I don't play hockey then it is true that I don't play both sports, so $T^C \cup H^C \subseteq (T \cap H)^C$. This means that the sets are equal, ponder this for some time. A similar argument can be made for the complement of the union. If this is not convincing spend some time with a Venn diagram and see if you can get it to make sense.

1.4.3 3 or More Sets

1.4.3.1 Associativity

Just like you can add more than two numbers together you can take the intersection or union of more than two sets. For example $\{1, 2\} \cap \{2, 3\} \cap \{3, 4\} = \emptyset$. There are no elements in the intersection of these three sets because no number appears in all three sets, so this is the empty set. Intersections are associative, meaning $(A \cap B) \cap C = A \cap (B \cap C)$, so it doesn't matter what order you take a group of intersections in. Unions are also associative, $(A \cup B) \cup C = A \cup (B \cup C)$.

1.4.3.2 Distributive Property

If you mix together unions and intersections in an expression it isn't associative. Can you think of an example?

$$A \cap (B \cup C) \neq (A \cap B) \cup C$$

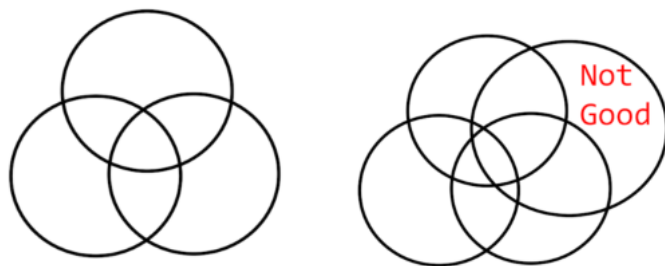
There are distributive formulas for situations like this.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Because of this it is said that intersection distributes over unions, and that unions distribute across intersections. It is easier to remember these distributive formulas by comparing them to the way multiplication distributes over addition. $a(b + c) = ab + ac$. Just pretend that the multiplication is an intersection and the addition is a union.

1.4.3.3 Messy Venn Diagrams

You can draw a Venn diagram for three sets with three circles. It gets a little complicated. I wouldn't bother drawing a Venn diagram for 4 sets.



Chapter 2

Counting

2.1 Size of a set

The set $A = \{5, 6\}$ has two elements. There is a notation to say the size of a set:

$$n(A) = 2$$

n is a function that takes in a set and outputs the number of elements in it:

$$n : \text{Sets} \rightarrow \text{Whole Numbers}$$

Let the sample space be $S = \{1, 2, 3, 4, 5, 6\}$. Since $A = \{5, 6\}$, $A^C = \{1, 2, 3, 4\}$ and $n(A^C) = 4$. There is a formula we can use for the size of the complement:

$$n(A^C) = n(S) - n(A) = 6 - 2 = 4$$

This is saying the size of the complement of a set is the size of the sample set minus the size of the set. I think this is pretty intuitive.

2.2 Principle of Inclusion-Exclusion

We define sets A and B :

$$A = \{3, 4, 5\} \quad B = \{4, 5, 6\}$$

We can see that $A \cup B = \{3, 4, 5, 6\}$ and $n(A \cup B) = 4$. Notice that $n(A \cup B) \neq n(A) + n(B)$.

There is a special formula for the number of elements in a union called the principle of inclusion-exclusion.

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

We subtract $n(A \cap B)$ from this expression because the elements of the intersection are counted twice - once in $n(A)$, and once in $n(B)$. Using the sets from earlier, $A \cap B = \{4, 5\}$. We can verify that the formula works by calculating

$$n(A \cup B) = n(A) + n(B) - n(A \cap B) = 3 + 3 - 2 = 4$$

If the intersection of two sets is the empty set then the size of the union is the sum of the size of the sets.

$$A \cap B = \emptyset \iff n(A \cap B) = 0 \iff n(A \cup B) = n(A) + n(B) - 0 = n(A) + n(B)$$

When two sets have no elements in common we call them **mutually exclusive**.

There is also a formula for the size of the union of three sets.

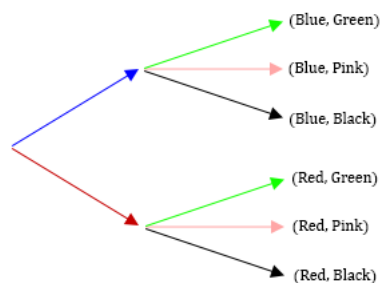
$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(B \cap C) - n(C \cap A) + n(A \cap B \cap C)$$

Consider how many times elements from each of the sections of a venn diagram with three sets are counted using the above formula. You should conclude that each element in $A \cup B \cup C$ is counted exactly once.

2.3 Multiplication Principle

I have a red shirt and a blue shirt. I have a green hat, a pink hat, and a black hat. How many different shirt/hat combinations can I wear?

Combinations of (shirt, hat)



The above diagram is an example of a **tree** diagram. The leaves (arrows with no arrows after them) of our tree represent the possible shirt/hat combinations. There are 6 possible shirt/hat combinations because there are 2 branches that each have 3 leaves which gives us $2 \times 3 = 6$ total combinations. In general if there are m shirts and n hats there will be $m \times n$ shirt/hat combinations because there will be m branches with n leaves. This generalizes to anything where you make two selections where the number of options in one of the selections does not depend on the other selection. For example the number of hats I can wear doesn't depend on which shirt I pick. If I decide that my pink hat is special

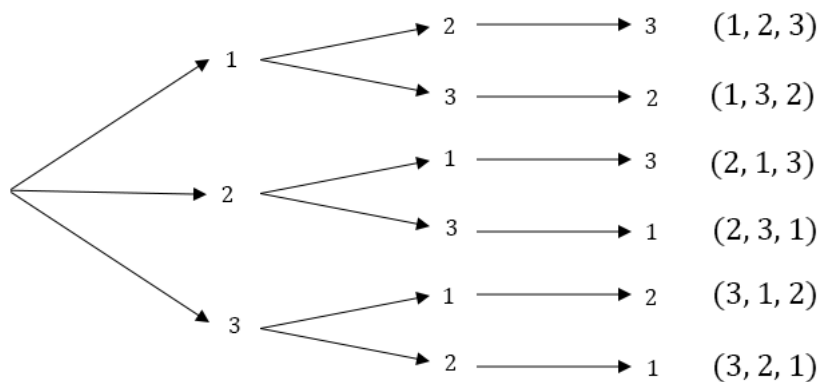
and not to be worn with my red shirt then there are 5 combinations and the multiplication principle doesn't apply.

If I have 4 pairs of pants to choose from then there will be $2 \times 3 \times 4$ shirt/hat/pant combinations. In general if you are making independent selections with a_1, a_2, \dots, a_n options, then the number of ways to make these selections is:

$$a_1 \times a_2 \times \dots \times a_n = \prod_{k=1}^n a_k$$

2.4 Permutations

Permutations are all about rearranging things into ordered outcomes. For example, how many ways can we order the numbers in the set $\{1, 2, 3\}$? Let's draw a tree diagram to figure this out.



There are three options for our first selection, two options for our second selection, and one option for our third selection. The multiplication principle gives us 6 permutations of the set $\{1, 2, 3\}$. Generally if $n(A) = m$ there are $m \times (m - 1) \times \dots \times 2 \times 1 = m!$ permutations of the set A .

Here are some natural examples of permutations. How many possible outcomes are there in a three person race? We could imagine a set containing the different people and the outcomes of the race are the possible orders of the set. Some things not obviously involving order are permutation problems. If I bought three different chocolate bars how many ways can I give each of my three family members a chocolate bar? I think the natural way to do this is to use the multiplication principle to see that there are three options for the first chocolate bar I give away, two for the second, and one for the third. If I gave away all of the chocolates at the same time and there is no ordering it doesn't matter because I could have given the gifts away one at a time for the same set of possible outcomes.

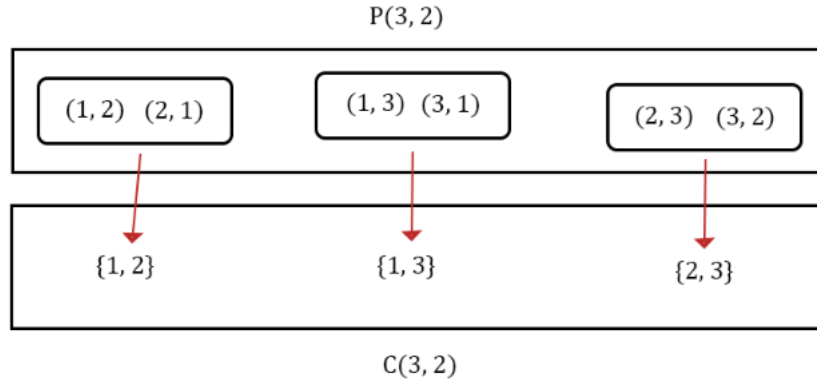
Let's consider something more general. In a 10 person race how many possible outcomes are there for the top 3? 10 people could have come in first place, there are 9 options for second once the first place is determined, and then 8 options for the third. The answer is then $10 \times 9 \times 8 = 720$ using the multiplication principle. The general formula for the top k racers in an n person race is:

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

The situation for the top k racers in an n person race has it's own notation, $P(n, k)$. $P(n, k)$ is pronounced " n permute k ".

2.5 Combinations

In permutations we consider how many possible orderings a subset of a certain size can have. What if we instead asked how many possible subsets there are of a certain size? For example, how many different ways can I pick 2 people from a group of 3 people? These problems are called combinations and the number of k element subsets of an n element set is $C(n, k)$. An different notation that is more common is $\binom{n}{k}$ and this is pronounced " n choose k ". There is a relationship between permutations and combinations that we illustrate below.



For every set counted in $C(3, 2)$ there are two ordered pairs in $P(3, 2)$. Order doesn't matter in a combination, but it does in a permutation. If we select two people from a group of three there will be two permutations for each combination. If we select three people, there are six permutations per combination because there are six possible permutations of three people. In general we see that $P(n, k) = C(n, k) \times k!$. Let's solve for $C(n, k)$.

$$C(n, k) = \frac{P(n, k)}{k!} = \frac{n!}{k!(n-k)!}$$

2.6 Variations

Often a problem will involve using the multiplication principle with permutations and combinations. Consider the following:

There is a set of 9 distinct pigs. Three go to the market, three go to Arby's, and three go home. How many ways can this happen?

There are $\binom{9}{3}$ ways to select the piggies for the market. After that selection is made there are $\binom{6}{3}$ ways to select the piggies for Arby's. The remaining three piggies go home. The answer is then:

$$\binom{9}{3} \times \binom{6}{3} \times \binom{3}{3} = \frac{9!}{3!6!} \times \frac{6!}{3!3!} \times \frac{3!}{3!0!} = \frac{9!}{3!3!3!}$$

There is actually a general principle at work here. If we divide a set into mutually exclusive subsets then the number on the top is the size of the set and the numbers on the bottom are the size of the subsets.

Is this true for dividing a set into two subsets? Why?

A rephrasing of this question is "How many ways can you rearrange the letters of 'aaabbbccc'?" To see that these questions are the same let each position of the letter be a piggy and assign the market, Arby's, and home to the letters a, b, and c. Select positions in groups of three to assign the letters to.

Here is an example that combines permutations and combinations. "How many ways could I rank my 10 favorite math problems from a list of 20 and then do 5 of them?" There are $P(20, 10)$ possible rankings of my top 10 problems. There are $\binom{10}{5}$ ways to choose from the selected problems. The answer is then:

$$P(20, 10) \times \binom{10}{5} = 168,951,528,345,600$$

Chapter 3

Probability Basics

3.1 Equally Likely Events

If we flip a coin it is said that heads and tails have the same probability because the percentage of both heads and tails tend toward 50% as the number of coin flips increases. Think of a **probability** as the percent of time an outcome occurs when many trials are performed. Here are the results of a simulation.

Number of Trials

Percent Heads

10

60%

1,000

48.1%

100,000

50.005%

When we make a “random selection” it means that every possible selection has equal probability. For example, a randomly selected card from a deck of 52 cards has probability $\frac{1}{52}$ of being a 2 of clubs.

When every possible outcome has equal probability, the formula for the probability of some subset of the outcomes is:

$$\frac{\text{Number of Selected Outcomes}}{\text{Number of Total Possible Outcomes}}$$

What is the probability of drawing 4 aces when we randomly select 5 cards from a deck? The number of ways to draw 4 aces is $\binom{4}{4} = 1$ because there are 4 aces

and we must select all of them. There are 48 cards to select that are not aces which leads to $\binom{48}{1} = 48$ outcomes.

The number of ways to draw 4 aces is:

$$\binom{4}{4}\binom{48}{1} = 48$$

The number of ways to draw 5 cards is:

$$\binom{52}{5}$$

The answer is then:

$$\frac{\binom{4}{4}\binom{48}{1}}{\binom{52}{5}} = \frac{48}{2598960} = \frac{1}{54145}$$

3.2 Terminology

3.2.1 Sample Spaces and Events

In probability we talk about experiments, sample spaces, and events.

When we flip a coin or roll a die that is an **experiment**. Experiments have a set of possible outcomes that happen with some probability.

The set of all possible outcomes of an experiment is the **sample space**. When we flip a coin once the sample space is $S = \{H, T\}$. When we flip a coin twice the sample space is $S = \{HH, HT, TH, TT\}$.

Events represent a set of possible outcomes from an experiment. When we flip two coins there is an event for flipping two heads, $E_{\text{both heads}} = \{HH\}$. There is also an event for not flipping two heads $E_{\text{not both heads}} = \{HT, TH, TT\}$. An event is a subset of the sample space. It may represent a single outcome of our experiment, or it may represent several of the possible outcomes:

$$E \subseteq S$$

We can rewrite our formula for probabilities when all outcomes of an experiment are equally likely using $n(E)$ for the number of selected outcomes.

$$\frac{\text{Number of Selected Outcomes}}{\text{Number of Total Possible Outcomes}} = \frac{n(E)}{n(S)}$$

Example: Rolling Dice

We roll two dice and want to calculate the probability that the dice sum to 4. Define the experiment, sample space, event, and calculate the probability of the event.

The experiment is rolling two dice.

The sample space has 36 elements, we can define it in set-builder notation:

$$\{(a, b) | a, b \in \{1, 2, 3, 4, 5, 6\}\}$$

The event contains the elements of the sample space summing to 4.

$$E = \{(1, 3), (3, 1), (2, 2)\}$$

The probability is

$$\frac{\text{Number of Selected Outcomes}}{\text{Number of Total Possible Outcomes}} = \frac{n(E)}{n(S)} = \frac{3}{36} = \frac{1}{12}$$

3.3 Probability Functions

There is a function called a probability function, denoted P , that calculates the probability of an event. When we flip a coin twice the probability of getting a head and a tail is:

$$P(\{HT, TH\}) = .5$$

For experiments where all outcomes are equally likely:

$$P(E) = \frac{n(E)}{n(S)}$$

A probability is between 0 and 1 because an event can't happen less than 0% of the time or more than 100% of the time. More formally, P is a function that takes an event as input and gives a number between 0 and 1 as output. In notation:

$$P : E \mapsto [0, 1]$$

If we are being more mathematically rigorous, this rule about probabilities being between 0 and 1 is derived from some more basic assumptions and not intuition about the percentage of times something occurs. Let's talk about these basic assumptions.

3.3.1 Probability Axioms

There are three fundamental assumptions (called axioms) about probability functions from which our other laws are derived.

First Axiom - For an event E , and probability function P :

$$P(E) \geq 0$$

Second Axiom - For a sample space S :

$$P(S) = 1$$

Third Axiom - If E_1, E_2, \dots, E_n are **mutually exclusive** events:

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = P(E_1 \cup E_2 \cup \dots \cup E_n \cup \dots) = P(E_1) + P(E_2) + \dots + P(E_n) + \dots$$

For **two mutually exclusive** events the third axiom is:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

There are several formulas that are useful for this exam that can be derived from these axioms. We can derive the formula for the probability of the complement of an event:

$$P(A) + P(A^C) = P(A \cup A^C) = P(S) = 1 \text{ using Axioms 2 and 3}$$

Since $P(A^C) + P(A) = 1 \implies P(A^C) = 1 - P(A)$.

Here are some useful formulas that can be derived from these axioms.

Complements : $P(A^C) = 1 - P(A)$ *Upper Bound* : $P(A) \leq 1$ *General Probability of 2 Unions* : $P(A \cup B)$

Notice that these formulas are the same as our formulas for the size of the set if we swap out P for n and 1 for $n(S)$.

Example: Coin Flipping and Axioms

If we flip a coin, heads and tails are mutually exclusive events.

$$P(\{H\} \cup \{T\}) = P(\{H\}) + P(\{T\}) = .5 + .5 = 1$$

We know that $P(\{H\}), P(\{T\}) = .5$ from our formula for equally likely events at the beginning of the chapter. Note that $S = \{H\} \cup \{T\}$ so this example also illustrates that $P(S) = 1$.

3.4 Conditional Probability and Independence

Fluffy

Not Fluffy

Total

Cats

21

18

39

Dogs

35

26

61

Total

56

44

100

We make estimates about the overall population of pets using the results of a survey about pets. To do this we pretend that the pets in the survey represent the entire population of pets.

We can calculate the probability of a pet being fluffy using our formula for equally likely events:

$$\frac{\text{Number of Fluffy Pets}}{\text{Total Number of Pets}} = \frac{n(\text{Fluffy})}{n(\text{Pets})} = \frac{56}{100}$$

We can calculate the probability of a pet being a fluffy cat as:

$$\frac{\text{Number of Fluffy Cats}}{\text{Total Number of Pets}} = \frac{n(\text{Fluffy} \cap \text{Cat})}{n(\text{Pets})} = \frac{21}{100}$$

We can also calculate the probability of a pet being fluffy, given that the pet is a cat:

$$\frac{\text{Number of Fluffy Cats}}{\text{Total Number of Cats}} = \frac{n(\text{Fluffy} \cap \text{Cat})}{n(\text{Cats})} = \frac{21}{39}$$

Calculating the probability of a pet being fluffy given that the pet is a cat is an example of conditional probability because in the denominator we do not

include all of the possible pets, but only the cats. There is a special notation for conditional probabilities:

Probability Pet is Fluffy given that Pet is a Cat = $P(\text{Pet is Fluffy} | \text{Pet is a Cat})$

The general formula for conditional probabilities uses probabilities instead of set sizes. If all outcomes are equally likely either the probability or set size formula will give the same result:

$$\frac{P(\text{Fluffy} \cap \text{Cat})}{P(\text{Cat})} = \frac{n(\text{Fluffy and Cat})/n(\text{Pets})}{n(\text{Cat})/n(\text{Pets})} = \frac{n(\text{Fluffy} \cap \text{Cat})}{n(\text{Cat})}$$

3.4.1 Conditional Probability Formulas

The probability of event A given that event B has occurred is denoted $P(A|B)$ and pronounced “the probability of A given B”. The general formula for this is

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

For experiments where all outcomes have equal probabilities $P(A|B)$ can be calculated with set sizes. The derivation for formula this was done in our example with fluffy cats.

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

We can calculate the probability of the intersection using the probability of the given event with the conditional probability. This formula is easiest to conceptualize if you imagine an event B happening with probability $P(B)$ and then an event A happening with probability $P(A|B)$.

$$P(A|B) \times P(B) = P(A \cap B)$$

3.4.2 Independence

Events are independent if they do not influence each other. The first flip of a coin will not impact the second flip so these events are independent.

Let the event that the first card pulled without replacement from a deck is a jack be J_1 and the event that the second card pulled is a jack is J_2 . If I pull a jack on my first try there will be less jacks in the deck for the second draw. The event J_1 has an effect on J_2 and these events are not independent.

Definition of Independence: Events A and B are said to be independent if $P(A|B) = P(A)$.

This definition is equivalent to $P(A \cap B) = P(A) \times P(B)$ after substitution with the conditional probability formulas. This identity has it's own name:

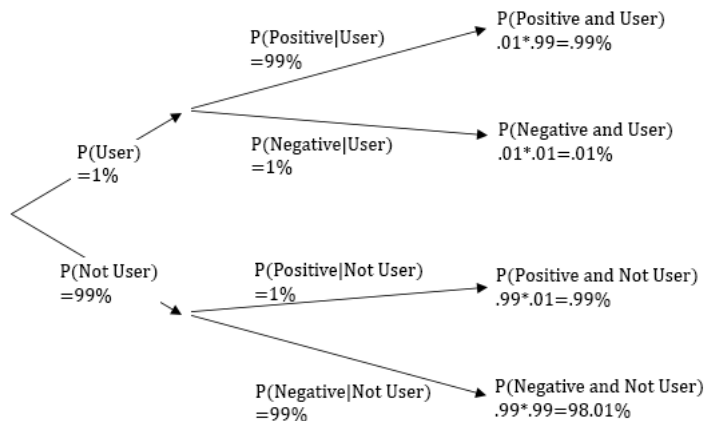
Multiplication Rule for Independent Events: If A and B are independent events then $P(A \cap B) = P(A) \times P(B)$.

Let H_n be the event that the n th flip of a coin is heads so that H_2 means the second coin flip is heads. The multiplication rule for independent events works for more than two events if all events are mutually independent so that $P(H_1 \cap H_2 \cap H_3 \cap H_4) = P(H_1) \times P(H_2) \times P(H_3) \times P(H_4) = (\frac{1}{2})^4$.

3.5 Bayes Theorem

3.5.1 Bayes Theorem Intuition

Suppose 1% of the population uses drugs. 98% of drug users test positive on a drug test and 2% of non-users test positive. What is the probability that a person testing positive for a drug test has used drugs?



Using the above tree diagram and the definition of conditional probability we calculate:

$$P(\text{User}|\text{Positive}) = \frac{P(\text{User} \cap \text{Positive})}{P(\text{Positive})} = \frac{P(\text{User} \cap \text{Positive})}{P(\text{Positive} \cap \text{Not User}) + (P(\text{Positive} \cap \text{User}))} = \frac{P(\text{Positive and User})}{P(\text{Positive and Not User}) + P(\text{Positive and User})}$$

Let's explain these steps in more detail.

3.5.2 Law of Total Probability

In our calculation when we go from the first to the second line we use the identity $P(\text{Positive}) = P(\text{Positive} \cap \text{User}) + P(\text{Positive} \cap \text{Not User})$, but how do we know this is true? $\text{Positive} \cap \text{User}$ only contains drug users and $\text{Positive} \cap \text{Not User}$ only contains non-users. These events are mutually exclusive so $P(\text{Positive} \cap \text{User}) + P(\text{Positive} \cap \text{Not User}) = P((\text{Positive} \cap \text{User}) \cup (\text{Positive} \cap \text{Not User}))$ using our third probability axiom. Using the distributive property for intersections with the identities

$$A^C \cup A = \text{Sample Space}, \quad A \cap (\text{Sample Space}) = A$$

we can see that:

$$(\text{Positive} \cap \text{User}) \cup (\text{Positive} \cap \text{Not User}) = \text{Positive} \cap (\text{User} \cup \text{Not User}) = \text{Positive} \cap (\text{Sample Space}) = \text{Positive}$$

This is how we know that $P(\text{Positive}) = P(\text{Positive} \cap \text{User}) + P(\text{Positive} \cap \text{Not User})$.

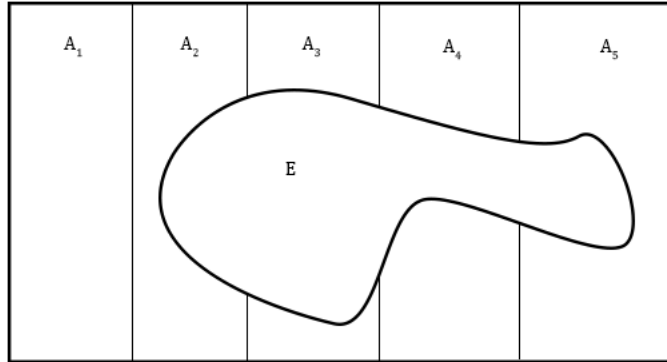
There is a more general formulation of this known as the **law of total probability**:

Events A_1, A_2, \dots, A_n are said to be a **partition** of the sample space S if $A_1 \cup A_2 \cup \dots \cup A_n = S$ and if for all i, j : $A_i \cap A_j = \emptyset$. This just means that the sets A_1, A_2, \dots, A_n cover the whole set and there is no overlap between the sets.

If sets A_1, A_2, \dots, A_n partition S then for any event E we have **the law of total probability**:

$$P(E) = P(E \cap (A_1 \cup A_2 \cup \dots \cup A_n)) = P(E \cap A_1) + P(E \cap A_2) + \dots + P(E \cap A_n) = P(E|A_1) \times P(A_1) + \dots + P(E|A_n) \times P(A_n)$$

Here is a visual explaining the intuition behind the equality in the first line of the equations above:



The proof of the law of total probability is outlined in our discussion of drug users.

3.5.3 Bayes Theorem Formula

3.5.3.1 Derivation Using Law of Total Probability

In our previous example every person is either a drug user or not so this partitions the sample space into two regions, *User* and *Not User*. We can use the definition of conditional probability to say:

$$P(\text{User}|\text{Positive}) = \frac{P(\text{User} \cap \text{Positive})}{P(\text{Positive})}$$

We use the law of total probability in the denominator to expand it and convert all expressions of the form $P(A \cap B)$ to expressions of the form $P(B|A) \times P(A)$. The result is:

$$\frac{P(\text{Positive}|\text{User}) \times P(\text{User})}{P(\text{Positive}|\text{User}) \times P(\text{User}) + P(\text{Positive}|\text{Not User}) \times P(\text{Not User})}$$

3.5.3.2 Bayes Theorem Statement

For a partition A_1, A_2, \dots, A_n and event E :

$$P(E|A_k) = \frac{P(E|A_k) \times P(A_k)}{P(E|A_1) \times P(A_1) + \dots + P(E|A_n) \times P(A_n)}$$

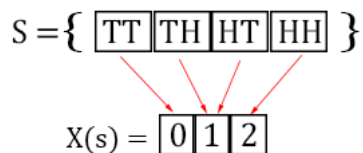
Chapter 4

Random Variables

4.1 What is a Random Variable

A random variable is a number determined by chance. What is the temperature in my city tomorrow? This is a number determined by chance, a random variable. OK sorry. I lied about a random variable being a number. Actually, **a random variable is a function** that takes an element of the sample space as input and produces a real number. What does that mean?

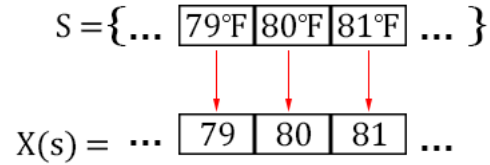
Example 4.1 (Counting Heads). If you flip a coin twice the sample space is $S = \{HH, HT, TH, TT\}$. We define a random variable X that maps the sequence of heads and tails to the number of heads. For example, $X(HT) = 1$ because there is exactly 1 head. See below that a random variable is a function that takes elements of the sample space and outputs a number.



Let's return to the question of tomorrow's temperature. I said this is a random variable because it is a random number. But I thought random variables were functions, not numbers? Why did I lie to you?

Example 4.2 (Tomorrow's Temperature). The set of possible temperatures measured in degrees fahrenheit is the sample space S . Our random variable is a function that sends a temperature in degrees fahrenheit to the real number that represents the temperature. Notice that there are no units for $X(s)$, it is a real number.

$$S = \{ \dots \boxed{79^\circ\text{F}} \boxed{80^\circ\text{F}} \boxed{81^\circ\text{F}} \dots \}$$

$$X(s) = \dots \boxed{79} \boxed{80} \boxed{81} \dots$$


See that a random variable is a function. It is just a very simple one.

Students can get confused about random variables because often the random variable is something simple like in the temperature example. The student then begins to overgeneralize and confuse random variables with sample spaces.

The important part of random variables isn't being able to describe them as being a function. The important part is being able to calculate the probability that tomorrow is a warm day. Let's get into that.

4.2 Probability Distribution Functions

4.2.1 Probability of an Outcome

We can calculate the probability of a random variable being a particular value. Let's work an example.

Example 4.3 (Probability of One Head). Again we are flipping a coin twice and $S = \{HH, HT, TH, TT\}$. Let $s \in S$, and X is the random variable for the number of heads. $X(s) = 1$ when $s \in \{HT, TH\}$. To calculate the probability of $X(s) = 1$:

$$P(X(s) = 1) = P(\{HT, TH\}) = \frac{1}{2}$$

Even though X is a function we usually don't write it as such. Instead of saying $X(s) = 1$ we usually just write $X = 1$ and it is understood that X is a function. In our previous example we could rewrite it as $P(X = 1) = P(\{HT, TH\}) = \frac{1}{2}$.

4.2.2 Probability Distribution Function Definition

It is common to calculate the probability of a random variable being a value x , $P(X = x)$. It is so common that we call it a “probability distribution function” (abbreviated **PDF**) and give it special notation.

Definition 4.1 (Probability Distribution Function). The probability distribution calculates $P(X = x)$. There is a special way to write this, instead of using a capital P we use a lowercase p and say that:

$$p(x) = P(X = x)$$

Some properties of the probability distribution function are that the probability of an outcome is between 0 and 1, and that the sum of the probability of all possible outcomes is 1:

For $x \in S$

- $0 \leq p(x) \leq 1$
 - $\sum p(x) = 1$
-

The table below shows the probabilities from our coin flipping example 4.1. The probabilities are between 0 and 1 and sum up to 1 as expected.

x
0
1
2
p(x)
.25
.5
.25

4.3 Cumulative Distribution Functions

Again we are flipping the coin twice and counting the heads. We can calculate the probability of 0 heads as .25, and the probability of 1 head as .5.

Now we can also calculate $P(X \leq 1) = p(0) + p(1) = .25 + .5 = .75$. Calculations of this sort are very common, we call it the cumulative distribution function (abbreviated **CDF**). There is a special notation for this.

Definition 4.2 (Cumulative Distribution Function). The cumulative distribution function calculates $P(X \leq x)$. There is a special way to write this:

$$F(x) = P(X \leq x)$$

We can calculate the cumulative distribution function by summing the probability density function:

$$F(x_0) = \sum_{x \leq x_0} p(x)$$

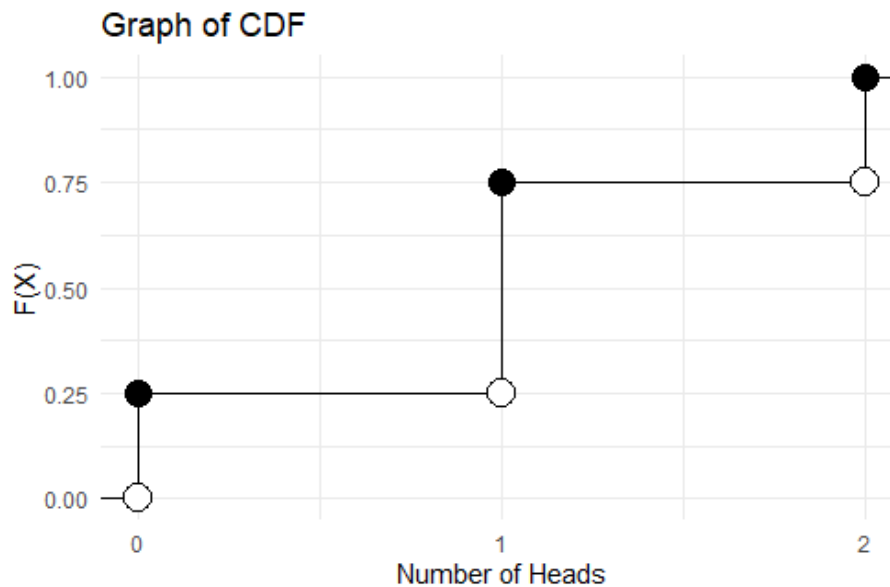
We calculate the cumulative distribution function for our coin flipping example, see that the CDF is the cumulative sum of the PDF.

x	
0	
1	
2	
p(x)	
.25	
.5	
.25	
F(x)	
.25	
.75	
1.00	

In the above table we only define the CDF at points 0, 1, 2 but we can also calculate the CDF at .5.

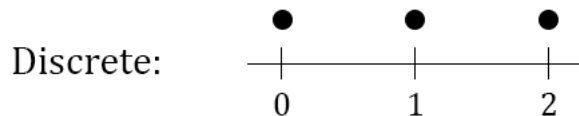
$$F(.5) = P(X \leq .5) = \sum_{x \leq .5} p(x) = p(0) = .25$$

The CDF only changes at values where there are outcomes that happen with some probability. Understand why the graph below looks the way it does.

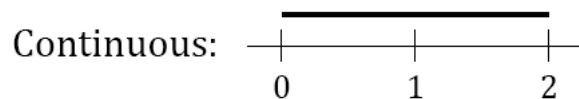


4.4 Discrete vs. Continuous Random Variables

So far we have only talked about flipping coins. When I flip a coin twice the only possible numbers for the heads are $\{0, 1, 2\}$. This is an example of a **discrete random variable**. Discrete means that only certain numbers are possible. Plotting the possible outcomes on a number line yields a series of points.



Some random variables are **continuous**. This means that the outcomes are intervals instead of points. The length of a cat is continuous, the length could truly be any number.



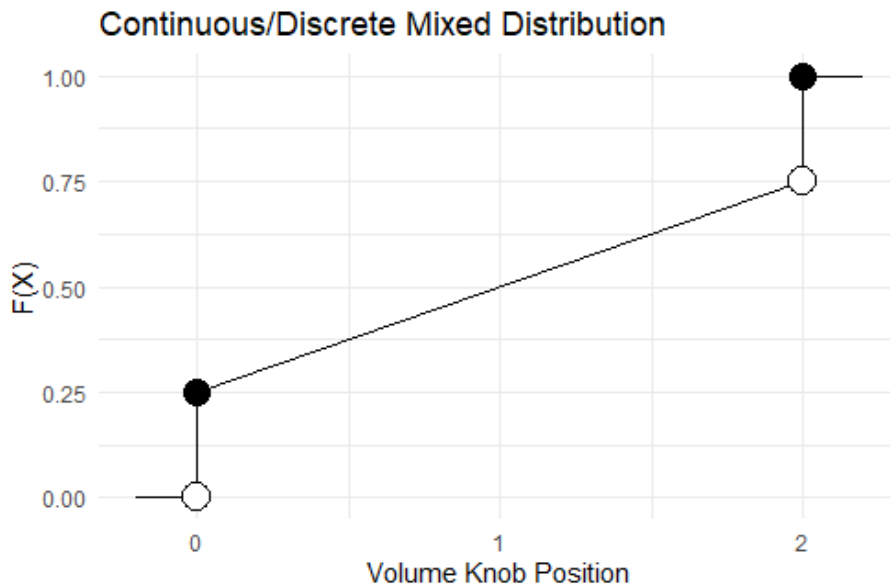
Some random variables are continuous and discrete **mixed** together. If my speakers volume knob must be between 0 and 2 then often my speakers will often be turned to exactly 0 (no sound) or 2 (max sound). If the volume knob is somewhere between 0 and 2 it could be turned to any number, like 1.5543..., so the volume knob is a mixture of discrete and continuous.

Mixed:

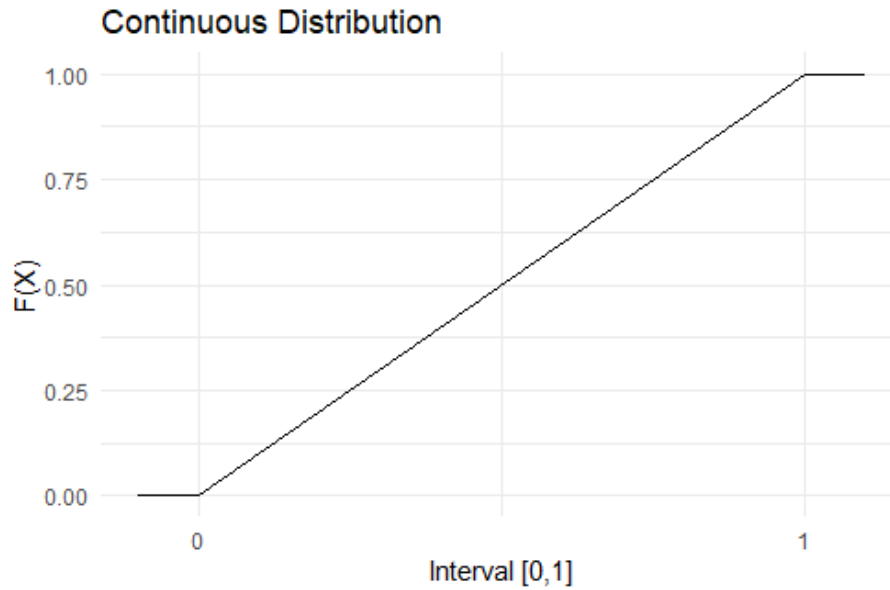


4.4.1 A Bit More Precise

In the volume knob example the possible outcomes come from the interval $[0, 2]$, this sounds continuous? What about it exactly is mixed? In the CDF there is a jump at 0 because $P(X \leq -0.01) = 0$ but $P(X \leq 0) =$ the probability that the volume is zero.



Generally discrete distributions are just going to bounce around without ever smoothly increasing over an interval. Consider the coin flipping CDF graph at the end of last section. Continuous distributions smoothly increase over intervals. Let our random variable be a randomly drawn number in $[0, 1]$. The CDF smoothly increases from 0 to 1 on $[0, 1]$.



The CDF jumps at a point x when $P(X = x)$ is nonzero, consider our volume knob example for evidence of this. A randomly selected number from $[0, 1]$ is **exactly** .5 with probability 0, so the continuous random variable has no jumps. There are infinitely many numbers in $[0, 1]$ and .5 is only one of these infinitely many numbers. From our formula for equally likely outcomes:

$$\frac{\text{Number of Selected Outcomes}}{\text{Number of Total Outcomes}} = \frac{1}{\infty} = 0$$

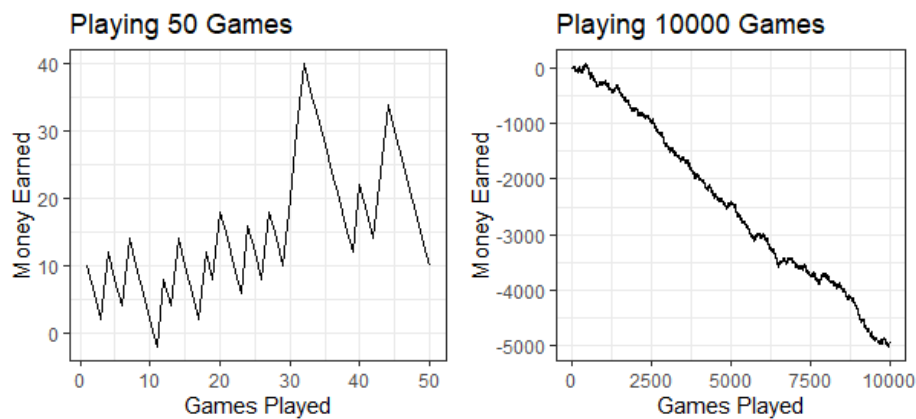
We are still not telling the whole story. Continuous distributions end up using a bunch of calculus and we will discuss them in a later chapter.

Chapter 5

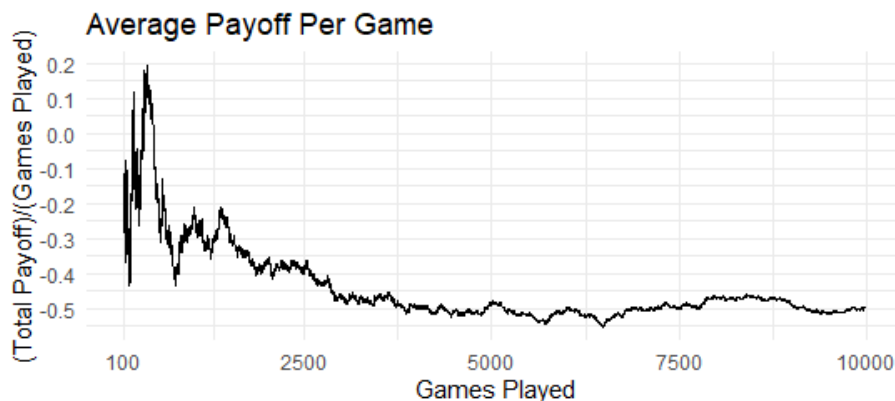
Functions of Discrete Random Variables

5.1 What is the Mean?

Your friend plays a game with you where he flips a coin twice. You win 10\$ if both flips are heads, and lose 4\$ otherwise. Let's simulate some results to see if we should play this game.



After 50 games we might think this is a fun game but in fact the odds are stacked against us. On average we lose .5 dollars per game. In our simulation notice that the average payoff approaches $-.5$.



We can calculate the average payoff over the 10,000 simulated games by adding up the payoffs and dividing by the number of games:

Games Played

Wins

Losses

10000

2501

7499

$$\frac{\sum \text{Payoff}}{\text{Number of Games}} = \frac{2501 \times 10 + 7499 \times -4}{10000} = -0.4986$$

We can do some rearranging:

$$\frac{2501 \times 10 + 7499 \times -4}{10000} = \frac{2501}{10000}(10) + \frac{7499}{10000}(-4) = \% \text{Wins} \times 10 + \% \text{Losses} \times -4$$

As the number of trials increases %Wins approaches $P(\{HH\})$ and %Losses approaches $P(\{HT, TH, TT\})$. We can define a random variable X to represent the payoff with a probability distribution function of:

$$p(10) = P(\{HH\}) = .25, p(-4) = P(\{HT, TH, TT\}) = .75$$

We can then write the long term average payoff as:

$$-4 \times p(-4) + 10 \times p(10) = .75 \times -4 + .25 \times 10 = -.5$$

So the long term payoff of our random variable is .5. When we have a random variable X that represents the payoff, we say that this is the **mean** of X .

Definition 5.1 (Expected Value). The mean of a random variable X is also called the **expected value** and is denoted by $E(X)$ or μ . To calculate the value of the mean we take the probability weighted sum of the possible values of the random variable.

$$\mu = E(X) = \sum p(x) \times x$$

Take a moment and see that the summation in our calculation is the same one performed in our calculation of the coin flipping game.

Example 5.1 (Average Heads). We flip a coin twice, let X be the number of heads. What is $E(X)$?

$$E(X) = p(0) \cdot 0 + p(1) \cdot 1 + p(2) \cdot 2 = .25 \cdot 0 + .5 \cdot 1 + .25 \cdot 2 = 1$$

5.1.1 $E(aX + b)$

Suppose we flip a coin twice. We play a game where we receive 2\$ for every head we flip. Let the random variable for the number of dollars we receive be Y and the random variable for the number of heads flipped be X . Notice that $Y = 2X$.

Probability

.25

.5

.25

Heads

0

1

2

Dollars

0

2

4

We calculate $E(Y)$:

$$E(Y) = .25 \cdot 0 + .5 \cdot 2 + .25 \cdot 4 = 2$$

$E(Y) = 2$ and from example 4.1, $E(X) = 1$. This makes sense because $Y = X - 1$. What if we got 4 dollars or 7 dollars for each head? In general if we get a dollars per head:

$$\begin{aligned} E(Y) &= .25 \cdot 0 + .5 \cdot a + .25 \cdot 2a \\ &= a(.25 \cdot 0 + .5 \cdot 1 + .25 \cdot 2) \\ &= aE(X) \end{aligned}$$

Let's make a new game. We still get 2 dollars for each head but this time we have to pay 1 dollar to play. Let the random variable representing the payout be Z . Note that $Z = Y - 1$. We calculate $E(Z)$:

$$E(Z) = .25 \cdot -1 + .5 \cdot 1 + .25 \cdot 3 = 1$$

Notice that $E(Z) = E(Y) - 1$ which makes sense since $Z = Y - 1$. What if we have to pay a different amount? Let's come up with a formula for $Z = Y - a$

$$\begin{aligned} E(Z) &= .25 \cdot (0 - a) + .5 \cdot (2 - a) + .25 \cdot (4 - a) \\ &= (.25 \cdot 0 + .5 \cdot 2 + .25 \cdot 4) + a(-.25 - .5 - .25) \\ &= E(Y) - a = 2E(X) - a \end{aligned}$$

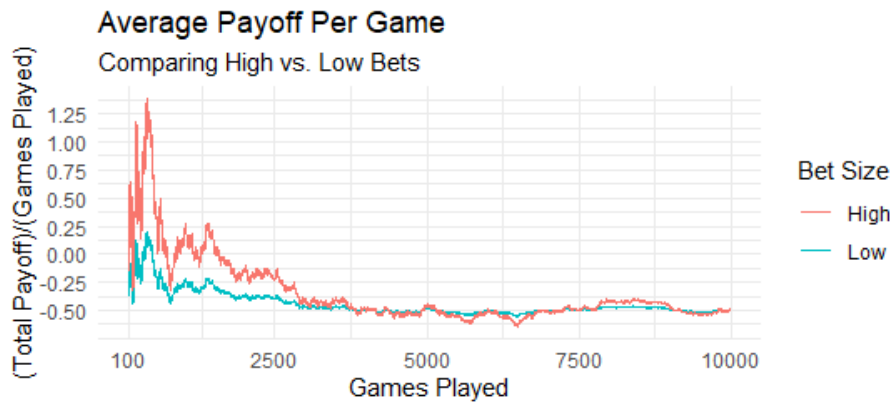
These tricks work for any random variable, not just coins. The proof of the following theorem is very similar to the calculations we have been doing.

Theorem 5.1 ($E(aX+b)$). *For a random variable X with constants a and b :*

$$E(aX + b) = aE(X) + b$$

5.2 What is Variance?

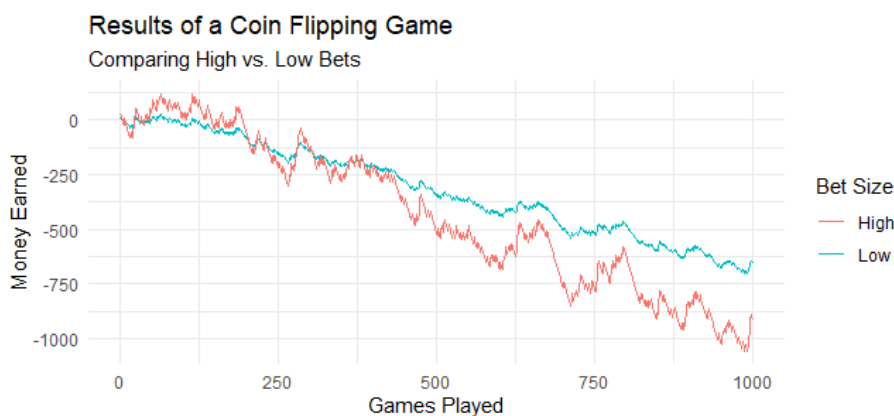
At the beginning of the chapter we played a game where we flip a coin twice. We win 10\$ if both flips are heads and lose 4\$ otherwise. Let's increase the bet size. In our new bet we win 28\$ if both flips are heads and lose 10\$ otherwise. Experimentally we can see that both of the bets have the same long term average loss of $-.5$.



We verify this using the formula for expected values (5.1):

$$.25 \cdot 28 + .75 \cdot -10 = -.5$$

A difference between the lower and the higher bet sizes is that with the higher bet is that there are bigger wins and bigger losses. Notice in the graph below how the increases and decreases in money earned are larger with the higher bet size. In this example the higher bet size has the same mean but a higher **variance**



5.2.1 Defining Variance

Definition 5.2 (Definition of Variance). The variance of a random variable X is denoted by $V(X)$. It is defined as:

$$V(X) = E((X - \mu)^2) = \sum (x - \mu)^2 \cdot p(x)$$

Let X be the payoff of the smaller bet sizes and Y be the payoff for the larger bet sizes. We calculate the variance of the payoff for both games.

Low Bet			High Bet		
p(x)	.25	.75	p(y)	.25	.75
x	10	-4	y	28	-10
μ	-.5	-.5	μ	-.5	-.5

$$V(X) = E((X - \mu)^2) = .25 \cdot (10 - (-.5))^2 + .75 \cdot (-4 - (-.5))^2 = 36.75$$

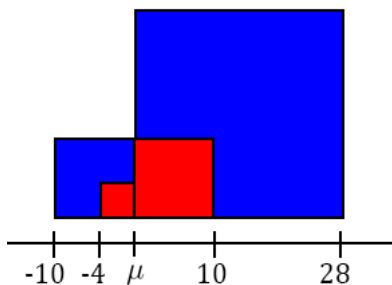
$$V(Y) = E((Y - \mu)^2) = .25 \cdot (28 - (-.5))^2 + .75 \cdot (-10 - (-.5))^2 = 270.75$$

The game with the larger bet sizes has a larger variance, agreeing with our experiments

Exercise 5.1. Let X be 1 if a coin flips heads and 0 if it is tails. Show that the $V(X) = .25$.

5.2.2 Variance Intuition

Consider that $V(X) = \sum (x - \mu)^2 \cdot p(x)$. $(x - \mu)^2$ is the area of a square with sides of length $x - \mu$. $V(X) = \sum (x - \mu)^2 \cdot p(x)$ is then the probability weighted sum of the area of squares. Do you see why the larger bet sizes cause a larger variance?



5.2.3 $V(aX+b)$

Suppose we flip a coin once. We receive 2\$ if the coin is heads. Let X be 1 if heads and 0 otherwise, and Y be the number of doillars received. Note that $Y = 2X$. $E(Y) = 1$ and $E(X) = .5$, remember that these numbers appear in our variance formula as μ . We calculate $V(X)$ and $V(Y)$:

$$V(X) = .5 \cdot (0 - .5)^2 + .5 \cdot (1 - .5)^2 = .25$$

$$V(Y) = .5 \cdot (0 - 1)^2 + .5 \cdot (2 - 1)^2 = 1$$

There is a good reason that $V(Y) = V(2X) = 2^2 \cdot V(X)$

$$\begin{aligned} V(Y) &= .5 \cdot (0 - 1)^2 + .5 \cdot (2 - 1)^2 \\ &= .5 \cdot (2 \cdot 0 - 2 \cdot .5)^2 + .5 \cdot (2 \cdot 1 - 2 \cdot .5)^2 \\ &= 2^2 \cdot (.5 \cdot (0 - .5)^2 + .5 \cdot (1 - .5)^2) \\ &= 2^2 \cdot V(X) \end{aligned} \tag{5.1}$$

In general it is true that $V(aX) = a^2V(X)$. The manipulations performed in the coin flipping example could be applied anywhere.

What if we make a new random variable Z where we make 2\$ for heads and 1\$ dollars for tails? In this case $Z = X + 1$. We calculate $V(Z)$, note that $E(Z) = E(X + 1) = E(X) + 1 = 1.5$:

$$V(Z) = .5 \cdot (1 - 1.5)^2 + .5 \cdot (2 - 1.5)^2 = .25$$

$V(Z) = V(X)$ because the values are spread out the same distance. In general $V(X + b) = V(X)$ because the translation does not change how far apart the values are. If you are feeling inspired try to write a proof of this.

Combining these observations we get the following theorem.

Theorem 5.2 ($V(aX+b)$). *For a random variable X with constants a and b :*

$$V(aX + b) = a^2V(X)$$

Exercise 5.2. Show that $V(-2X + 1) = 4V(X)$.

Chapter 6

Common Discrete Distributions

6.1 Bernoulli Distribution

The Bernoulli distribution is appropriate when there are two outcomes. Success happens with probability p and failure happens with probability $1 - p$.

$$p(x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \end{cases}$$

Exercise 6.1. Let X be a Bernoulli random variable. Show that $E(X) = p$ and that $V(X) = p \cdot (1 - p)$.

The number of heads from a single coin flip is an example of a Bernoulli random variable. The event that the coin lands on heads is success, tails is failure. If the coin is fair, $p = .5$.

6.2 Binomial Distribution

We have an unfair coin that is heads 60% of the time and tails 40% of the time. We flip the coin 5 times. We express the number of heads as the sum of Bernoulli random variables, $X = H_1 + H_2 + H_3 + H_4 + H_5$ where

$$H_i = \begin{cases} 1 & \text{if flip } i \text{ is heads} \\ 0 & \text{if flip } i \text{ is tails} \end{cases}$$

We wish to calculate the probability that all 5 flips are heads. Because coin flips are independent we can use the rule $P(A \cap B) = P(A) \cdot P(B)$

$$\begin{aligned} P(X = 5) &= P((H_1 = 1) \cap (H_2 = 1) \cap (H_3 = 1) \cap (H_4 = 1) \cap (H_5 = 1)) \\ &= P(H_1 = 1) \cdot P(H_2 = 1) \cdot P(H_3 = 1) \cdot P(H_4 = 1) \cdot P(H_5 = 1) \\ &= .6^5 \end{aligned}$$

Less formally, the only way to achieve all heads is to flip the sequence *HHHHH*. Each head happens with probability .6 and the answer is $.6^5$.

What about the probability of flipping exactly 4 heads? The sequences that can lead to this outcome are:

T H H H H
H T H H H
H H T H H
H H H T H
H H H H T

These sequences all represent distinct outcomes for X . Although some of the sequences are the same in some places, each represents a distinct outcome of X , disjoint from all other outcomes. So we can use the addition formula for the probability of disjoint events.

$$\begin{aligned} P(X = 4) &= \\ &P((T H H H H) \cup (H T H H H) \cup (H H T H H) \cup (H H H T H) \cup (H H H H T)) = \\ &P(T H H H H) + P(H T H H H) + P(H H T H H) + P(H H H T H) + P(H H H H T) = \\ &5 \cdot .6^4 \cdot .4 \end{aligned}$$

What about the probability of 3 heads? There are many sequences of flips that result in this, like *HTHTH*. Each of these sequences will have probability $.6^3 \cdot .4^2$ because there are three heads and two tails. To find out how many sequences there are we must answer the question “How many ways are there to assign three of the five coins to be heads?” The answer to this question is:

$$\binom{5}{3} = \frac{5!}{3!2!} = 10$$

There are $\binom{5}{3}$ ways to flip 3 heads and each way has probability $.6^3 \cdot .4^2$, so the probability of three heads is $\binom{5}{3} \cdot .6^3 \cdot .4^2$.

This technique works for the other calculations we have done. There are $\binom{5}{5} = \frac{5!}{5!0!} = 1$ ways to flip 5 heads and $\binom{5}{4} = \frac{5!}{4!1!} = 5$ ways to flip 4 heads.

Let's generalize from our coin flipping. We had 5 independent Bernoulli trials where each Bernoulli trial had a probability of success as .6. What if we had n Bernoulli trials each with a probability of success of p . This is called the **binomial distribution** with parameters n and p .

Theorem 6.1 (Binomial Distribution). *Let X have the binomial distribution with parameters n and p . We have:*

$$p(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Be sure that this makes perfect sense.

6.2.1 Mean and Variance of the Binomial Distribution

In exercise 6.1 the motivated student showed that if X is Bernoulli with parameter p then $E(X) = p$ and $V(X) = p(1 - p)$. These formulas for the binomial distribution are very similar.

Theorem 6.2 (Binomial Distribution). *Let X have the binomial distribution with parameters n and p . We have:*

$$\begin{aligned} E(X) &= np \\ V(X) &= np(1 - p) \end{aligned}$$

6.2.1.1 Intuition

Let X be the number of heads from a single coin flip. Let Y be the number of heads from 10 coin flips.

$$\begin{aligned} E(X) &= np = 1 \cdot .5 \\ E(Y) &= np = 10 \cdot .5 \end{aligned}$$

Think about how many heads you expect to get when you flip 10 coins. It seems like it should be 10 times the expected number of heads from 1 coin. Apparently the variance is also 10 times larger for 10 times the coins.

$$\begin{aligned} V(X) &= np(1 - p) = 1 \cdot .5^2 \\ V(Y) &= np(1 - p) = 10 \cdot .5^2 \end{aligned}$$

6.3 Hypergeometric Distribution

Example 6.1 (Hypergeometric Rocks). We have 15 rocks in a bag. 7 rocks are red and 8 are black. We select 5 rocks. What is the probability of selecting exactly 2 red rocks and 3 black rocks.

Any selection of rocks is equally likely. We can use our formula for the probability when all events are equally likely.

$$\frac{\text{Number of Selected Outcomes}}{\text{Total Possible Outcomes}}$$

For the denominator we need to find how many ways there are to draw 5 rocks from a set of 15. This is $\binom{15}{5}$.

For the numerator we need to select 2 red rocks from a set of 7. This can be done in $\binom{7}{2}$ ways. We can select 3 blue rocks from 8 in $\binom{8}{3}$ ways. There are then $\binom{7}{2}\binom{8}{3}$ ways to select the rocks.

The answer is then:

$$\frac{\binom{7}{2}\binom{8}{3}}{\binom{15}{5}}$$

Exercise 6.2 (Hypergeometric Cards). We have a deck of 40 cards. 30 cards are red and 10 are black. We draw a hand of 5 cards. Show that the probability of drawing 3 black cards is:

$$\frac{\binom{30}{2}\binom{10}{3}}{\binom{40}{5}}$$

I strongly recommend students know how to figure these things out without consulting any formula sheets. I think trying to memorize this formula would be far worse than using the basics to derive it. Here is the formula.

Definition 6.1 (Hypergeometric Distribution). We randomly select n items from a population of N items. Let r represent the number of items from the population classified as a success, and k be the number of items in the selection

classified as successes. Let X be the random variable representing the number of items in our selection considered successes.

$$P(X = k) = \frac{\binom{N-r}{n-k} \binom{r}{k}}{\binom{N}{n}}$$

Example 6.2 (Memorization Is Bad). We return to our example with 15 rocks in a bag. 7 rocks are red. We select 5 rocks. What is the probability of selecting exactly 2 red rocks. We abandon critical thought and use the formula.

$$P(X = k) = \frac{\binom{N-r}{n-k} \binom{r}{k}}{\binom{N}{n}} = \frac{\binom{15-7}{5-2} \binom{7}{2}}{\binom{15}{5}}$$

I think trying to memorize this formula is a mistake. There are also formulas for the variance and mean that are not discussed here because there is not much understanding to be gained.

6.3.1 Hypergeometric vs. Binomial

The hypergeometric distribution is closely related to the binomial distribution. We have a group of 600 actuaries and 400 astronauts. We select 4 people randomly from the 1000 to win a prize, what is probability that 3 people are actuaries and 1 is an astronaut?

$$\frac{\binom{600}{3} \binom{400}{1}}{\binom{1000}{4}} = 0.3459$$

The hypergeometric distribution is different than the binomial distribution because it samples **without replacement**. Let's change the problem and allow someone to win a prize multiple times. 4 names are drawn from a hat. Each time a name is drawn a prize is given and the name is put back in the hat (this is sampling **with replacement**). Since 600 of the 1000 people are actuaries, any time we make a selection to win a prize there is a probability of .6 that the person is an actuary. We select 4 people, so the distribution of prizes given to actuaries is binomial with parameters $n = 4$, $p = .6$. In this example the probability of 3 actuaries winning prizes and 1 astronaut winning a prize is:

$$\binom{4}{3} \cdot .6^3 \cdot .4 = .3456$$

The hypergeometric distribution (without replacement) gives .3459 and the binomial distribution (with replacement) gives .3456. In the hypergeometric distribution if an actuary's name is drawn from the hat, the name does not go back

in the hat. The probability that the first name drawn is an actuary is .6 just like our example with replacement. Once an actuary's name is drawn from the hat the probability of drawing another actuarial name is $\frac{599}{999} \approx .6$. This is why our hypergeometric and binomial distributions are so similar in this problem. This approximation works best when the overall population is large enough so that selecting without replacement changes the probabilities very little from selecting with replacement.

6.4 Poisson Distribution

A Poisson distribution is appropriate when we are counting the number of times something happens in an hour or some unit of time. How many texts do I get in an hour? How often does someone in the state of California say the word “dude”. How often does a person in the city of Minneapolis trip and fall? How often does a car accident happen on interstate 35? These are things that happen at some frequency.

If you ever see the phrase “**the average rate**” you should consider that the question might be related to the Poisson distribution.

Often they give you the rate per hour and you need to convert it to the rate per 2 hours, or the daily rate, or the rate per minute. λ

Example 6.3 (Poisson Scaling Rate). The average rate of people spilling their coffee in the office is 2 per hour. Generally this rate is called λ , so $\lambda = 2$. The rate of people spilling coffee per minute is $\frac{\lambda}{60} = \frac{1}{30}$. The rate of people spilling their coffee every day is $24\lambda = 48$.

In reality most of the coffee will be spilled in the morning so the time of day changes the rate of spilling. The Poisson model does not work this way though. In the Poisson model it is assumed that the rate is constant throughout the day. The Poisson model also assumes that one person spilling their coffee doesn't make another person spill their coffee, spills are .

Theorem 6.3 (Poisson Distribution). *Let X have a Poisson distribution with parameter λ . The probability of k occurrences of something happening is:*

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

Example 6.4 (Poisson Scaling Rate). Assume a Poisson model. The average rate of people spilling their coffee in the office throughout the work day is 2.5 spills per hour. What is the probability that 2 people spill their coffee between 8 A.M. and 9 A.M., and 6 people spill their coffee between 11 A.M. and 1 P.M.?

Because the number of spills across time intervals is independent:

$$P(2 \text{ spills from 8-9 A.M. and 6 spills from 11 A.M.-1 P.M.}) = P(2 \text{ spills from 8-9 A.M.}) \cdot P(6 \text{ spills from 11 A.M.-1 P.M.})$$

For the calculation from 8-9 A.M. we use $\lambda = 2.5$ since the time interval is 1 hour.

$$P(2 \text{ spills from 8-9}) = \frac{e^{-2.5} 2.5^2}{2!}$$

For the calculation from 11 A.M.- 1 P.M. we use $\lambda = \lambda_{1 \text{ hour}} \cdot t = 2.5 \cdot 2 = 5$ since the time interval is 2 hours.

$$P(6 \text{ spills from 11 A.M.-1 P.M.}) = \frac{e^{-5} 5^6}{6!}$$

We multiply these quantities to get the answer:

$$\frac{e^{-2.5} 2.5^2}{2!} \cdot \frac{e^{-5} 5^6}{6!} = .03751$$

This formula is sort of funny so we try to make some connections to things so it makes sense. People that payed attention in Calculus 2 will know that

$$e^x = \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Let's verify that the sum of all probabilities is 1 for a Poisson X .

$$\sum_{n=0}^{\infty} P(X = k) = \sum_{n=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

Using the expansion for e^x it can be proven that:

Theorem 6.4 (Poisson Distribution). *Let X have a Poisson distribution with parameter λ .*

$$E(X) = \lambda$$

$$V(X) = \lambda$$

Easy to remember right?

6.5 Geometric Distribution

Let X be the number of tries it takes to land a backflip if I land 10 of my backflips. This is an example of a geometric distribution. The geometric distribution is when you take independent bernoulli trials until the first success. Each backflip attempt is an independent bernoulli trial with the same probability of success (assuming I don't improve or get tired).

The probability of success on my first attempt is .1. The probability of success on my second attempt is $.9 \cdot .1$, first you fail, then you succeed. The probability of success on my third attempt is $.9 \cdot .9 \cdot .1$, first I fail twice, then I succeed. Do you see how this goes? There are failures until there is success. We call the probability of success p and the probability of failure is $1 - p$ but we often use $q = 1 - p$ to make our formulas shorter.

Definition 6.2 (Geometric Distribution). We perform independent Bernoulli trials, each having a probability of success p and probability of failure q . Let X represent the trial on which the first success occurs.

$$P(X = k) = q^{k-1}p$$

Definition 6.3 (Geometric Distribution). Let X be a geometric random variable with parameter p .

$$E(X) = \frac{1}{p}$$

$$V(X) = \frac{1-p}{p^2}$$

There are multiple forms of the geometric distribution. So far we have been counting the attempt on which we succeed. Another way to do this is to count the number of failures before success.

Suppose we land a backflip on our first attempt. $X = 1$ if X is the try on which we succeed. If we let Y be the number of failures until success, $Y = 0$. We see that there are two ways to do this.

I personally prefer the formulation where we use the try on which we succeed. This is because if I land $p = \frac{1}{10}$ of my backflips then the expected try on which I succeed is $\frac{1}{p} = 10$. This makes sense to me and it is pretty. If we use the other formulation where Y is the number of failures before success we can say that $Y = X - 1$ and

$$E(Y) = E(X - 1) = E(X) - 1 = \frac{1}{p} - 1 = \frac{1 - p}{p}$$

$V(Y)$ is unchanged because $V(Y) = V(X + b) = V(X)$.

$$V(Y) = \frac{1 - p}{p^2}$$

6.6 Negative Binomial Distribution

The geometric distribution has a parameter p and the value of the random variable is the first successful attempt. The negative binomial distribution has two parameters, p and r . The negative binomial distribution counts the attempt on which the r th success occurs.

We flip an unfair coin with a probability of .6 for heads and .4 for tails. Let X be the trial on which we get the 3rd head. We wish to know $P(X = 5)$. Let's first consider that the 5th flip must be a head. Of the first 4 flips, 2 of them are heads. We get 2 heads in the first four flips with probability:

$$\binom{4}{2} \cdot .6^2 \cdot .4^2$$

We need the fifth flip to be heads so we multiply this probability by .6 to get our answer of:

$$\binom{4}{2} \cdot .6^3 \cdot .4^2$$

In this case $P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$.

6.6.1 Other (more common) formulation

A more common way of representing this is to count the number of failures until the r th success.

Definition 6.4 (Negative Binomial Distribution). Let Y be the number of failures until the r th success. We calculate the p.d.f. of Y .

$$P(Y = y) = \binom{r + y - 1}{y} p^r (1 - p)^y$$

If X is the R.V. representing the attempt on which the r th success occurs and Y is the R.V. representing the number of failures before the r th success, then $Y = X - r$ because failures = attempts - successes. Regardless of if the question asks you to count failures or attempts, the prepared student should be able to deduce the probability using first principles.

We give formulas for the expectation and variance for this formulation of the negative binomial distribution.

Theorem 6.5 (Negative Binomial Distribution). Let Y be the number of failures until the r th success.

$$E(Y) = \frac{r(1 - p)}{p}$$

$$V(Y) = \frac{r(1 - p)}{p^2}$$
