STATS 230 Course Notes

Joshua Allum

April 2017

Contents

Chapter 1

Introduction

1.1 Defining Probability

The Classical Definition

The probability of an event is

$$\frac{\text{the number of ways the event may occur}}{\text{the total number of possible outcomes}}$$

provided all outcomes are equally likely.

Example 1.1.1

The probability of a fair dice landing on 3 is $1/6$ because there is one way in which the dice may land on 3 and 6 total possible outcomes of faces the dice may land on. The sample space of the experiment, $\mathbb{S}$, is $\{1, 2, 3, 4, 5, 6\}$ and the event occurs in only one of these six outcomes.

The main limitation of this definition is that it demands that the outcomes of a sample space are equally likely. This is a problem since a definition of "likelyhood" (probability) is needed to include this postulate in a definition of probability itself.

The Relative Frequency Definition

The probability of an event is the limiting proportion of times that an event occurs in a large number of repetitions of an experiment.

Example 1.1.2

The probability of a fair dice landing on 3 is $1/6$ because after a very large series of repetitions (ideally infinite) of rolling the dice, the fraction of times the face with 3 is rolled tends to $1/6$.

The main limitation of this definition is that we can never repeat a process indefinitely so we can never truly know the probability of an event from this definition. Additionally, in some cases we cannot even obtain a long series of repetitions of processes to produce an estimate due to restrictions on cost, time, etc.

The Subjective Definition

The probability of an event occurring is a measure of how sure the person making the statement is that the event will occur.

Example 1.1.3

The probability that a football team will win their next match can be predicted by experts who regard all the data of past matches and current situations to provide a subjective probability.

This definition is irrational and leads to many people having different probabilities for the same events, with no clear “right” answer. Thus, by this definition, probability is not an objective science.

Probability Model

To avoid many of the limitation of the definitions of probability, we can instead treat probability as a mathematical system defined by a set of axioms. Thus, we can ignore the numerical values of probabilities until we consider a specific application. The model is defined as follows

- A sample space of all possible outcomes of a random experiment is defined.
- A set of events, to which we may assign probabilities, is defined.
- A mechanism for assigning probabilities to events is specified.

Chapter 2

Mathematical Probability Models

2.1 Sample Spaces

A sample space, $\mathbb{S}$, is a set of distinct outcomes for an experiment or process, with the property that in a single trial, one and only one of these outcomes occurs. The outcomes that make up a sample space are called sample points or simply points.

Example 2.1.1

The sample space for a roll of a six-sided die is

$$\{a_1, a_2, a_3, a_4, a_5, a_6\} \quad \text{where } a_i \text{ is the event the top face is } i$$

More simply we could define the sample space as

$$\{1, 2, 3, 4, 5, 6\}$$

Note that a sample space of a probability model for a process is not necessarily unique. Often times, however, we try to chose sample points that are the smallest possible or "indivisible".

Example 2.1.2

If we define E to be the event that the top face of a six-sided die is even when rolled and O to be the event the top-face is odd, then the sample space, $\mathbb{S}$, can be defined as

$$\{E, O\}$$

This is the same process as Example 2.1.1 (rolling a six-sided die), so since the sample spaces differ, clearly, sample spaces are not unique. Moreover, if we are interested in the event that a 3 is rolled, this sample space is not suitable since it groups the event in question with other events.

A sample space can be either **discrete** or **non-discrete**. If a sample space is discrete, it consists of a finite or countably infinite number "simple events". A countably infinite set is one that can be put into a one-to-one correspondence with the set of real numbers. For example, $\left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right\}$ is countably infinite whereas $\{\, x \mid x \in \mathbb{R} \,\}$ is not.

Simple Events

An event in a discrete sample space is a subset of the sample space, i.e., $A \subset \mathbb{S}$. If the event is indivisible, so as to only contain one point, we call it a simple event, otherwise it is a compound event.

Example 2.1.3

A simple event for a roll of a six-sided die is $A = \{a_1\}$ where a_i is the event the top face is i. A compound event is $E = \{a_2, a_4, a_6\}$.

2.2 Assigning Probabilities

Let $\mathbb{S} = \{a_1, a_2, a_3, \ldots\}$ be a discrete sample space. We assign probabilities, $P(a_i)$, for $i = 1, 2, 3, \ldots$ to each sample point a_i such that the following two conditions hold

- $0 \leq P(a_i) \leq 1$
- $\sum_{\text{all } i} P(a_i) = 1$

The set of probabilities $\{\, P(a_i) \mid i = 1, 2, 3, \ldots \}$ is called a **probability distribution** on $\mathbb{S}$.

Note that P is a function with the sample space as its domain.

The second condition, that the sum of the probabilities of all sample points is 1, relates to the property that for a given experiment one simple event in the sample space must occur. Every experiment or process always has an outcome thus the probability of any outcome being achieved must be 1.

Compound Events

The probability of an event A is the sum of the probability of all the simple events that make up A.

$$P(A) = \sum_{a \in A} P(a)$$

Example 2.2.1

In the previous example we saw that $E = \{a_2, a_4, a_6\}$ is a compound event. Thus, the probability of the compound event E is

$$P(E) = P(a_2) + P(a_4) + P(a_6)$$

Note that the probability model that we defined does not specify what actual numbers to assign to the simple events of a process. It only defines the properties that guarantee mathematical consistency. Thus, if we assigned $P(a_2)$ to be 0.9, our model would still be mathematically consistent but would not be consistent with the frequencies we obtain in multiple repetitions of the experiment.

In actual practice we try to define probabilities that are approximately consistent with the frequencies of the events in multiple repetition of the process.

Complements

The complement of an event, A, is the set of all outcomes not included in A and is denoted by $\overline{A}$.

Example 2.2.2

If $E = \{ a_1, a_3, a_5 \}$ is a compound event on the sample space $\{ a_1, a_2, a_3, a_4, a_5, a_6 \}$, then the complement of E is

$$\overline{E} = \{ a_2, a_4, a_6 \}$$

Because of the nature of complementary events, two complementary events cannot both occur in one process. The events are **mutually exclusive**.

Chapter 3

Counting Techniques

3.1 Counting Arguments

If we have a sample space, $\mathbb{S}$, of some experiment that has a **uniform distribution** (all sample points are equally likely), then we can calculate the probability of a compound event A as the number of sample points in A divided by the total number of sample points.

$$P(A) = \frac{k}{n}$$

where k is the number of sample points in A and n is the total number of sample points in the sample space.

Addition Rule

Consider we can perform process 1 in p ways and process 2 in q ways. Suppose we want to do process 1 **or** process 2 **but not both**, then there are $p + q$ ways to do so.

Example 3.1.1

Suppose a keyboard only has 26 letters and 20 special characters (!%#$), there are 46 ways in which a typist may type a **single** character. (Process 1: typing a letter. Process 2: typing a special character).

Multiplication Rule

Again, consider we can perform process 1 in p ways and process 2 in q ways. Suppose we want to do process 1 **and** process 2, then there are $p \times q$ ways to do so. This is because **for each way** of doing process 1 we can do process 2 in q ways.

Example 3.1.2

Suppose the same typist with the same keyboard wants to type a single letter **and** a single special character. The typist can do so in 520 ways, since there are 26 ways to select the letter and **for each** possible letter selection there are 20 possible special character selections.

Try to associate **OR** and **AND** with **addition** and **multiplication** respectively in your mind.

Often times, **OR**'s and **AND**'s are not explicit or obvious so you must re-word your problem to identify implicit **OR**'s and **AND**'s.

Example 3.1.3

A young boy gets to pick 2 toys from a store for his birthday. How many ways can he pick 2 toys if the store contains 12 toys? He may pick the same toy multiple times and picks the toys at random.

We can re-word this problem as follows: A young boy selects one of 12 toys **and** again, selects one of 12 toys. Thus there are $12 \times 12 = 144$ ways in which he can select 2 toys. Furthermore, we have that since selections are random, each selection is equally likely. So the probability that the boy selects any pair of toys is 1/144.

In this case the boy was allowed to select the same toy more than once. This is often referred to as **with replacement**. The addition and multiplication rules are generally sufficient to find probability of processes with replacement but if processes occur without replacement solutions become more complex and other techniques are often used.

The phrase **at random** or **uniformly**, indicates that each point in the sample space is equally likely.

Example 3.1.4

Consider a farmer with 500 different seeds. How many ways can he select 3 seeds randomly to plant?

We can re-word this problem to become: A farmer selects one seed from 500 **and** then selects one seed of 499 **and** then one seed of 498. So there are $500 \times 499 \times 498$ ways to do so.

Now, how many ways can he select 5 and 50 seeds randomly?

He can select 5 seeds in $500 \times 499 \times 498 \times 497 \times 496$ ways and 50 seeds in $500 \times \cdots \times 451$ ways.

Generally, if there are n ways of doing a process and it is done k times **without replacement**, that is you can only do the process a specific way once, there are $n \times \cdots \times (n - k + 1)$ ways to do it.

3.2 Counting Arrangements

When the sample space of a process is a set of arrangements of elements, like $\{ abc, acb, bac, bca, cab, cba \}$, the sample points are called the **permutations**. Assuming all n elements we are arranging are unique, how many sample points are there?

Consider trying to fill n boxes: $\boxed{}\boxed{}\boxed{\cdots}\boxed{}\boxed{}$. We have n ways to fill the first box (each element can go in the first box), **and** we have $(n - 1)$ ways to fill the second box, **and** so on until we have 1 way to fill the n$^{\text{th}}$ box. So there are $n \times (n - 1) \times \cdots \times 1$ total permutations in the sample space.

Example 3.2.1

Consider the letters of the word "fiesta". A baby (who cannot spell) randomly rearranges the letters of the word. What is the probability that "fiesta" is the outcome?

There are six boxes to fill: $\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}$. We have 6 ways to fill the first position, 5 ways to fill the second and so on until we have 1 way to fill the 6$^{\text{th}}$ position. The number of points in the sample space is $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$. So the probability of each outcome in the sample space is $1/720$.

Example 3.2.2

Consider the letters of the word "snake". If arranged randomly what is the probability that the word formed begins with a vowel?

There are five boxes to fill: $\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}$. There are two ways to fill the first box:

$$\boxed{\text{a}}\boxed{}\boxed{}\boxed{}\boxed{} \text{ and } \boxed{\text{e}}\boxed{}\boxed{}\boxed{}\boxed{}$$

and for each of these ways there are four remaining boxes to fill. The number of ways to fill the 4 remaining boxes is $4 \times 3 \times 2 \times 1 = 24$ so the total number of outcomes in which the first letter is a vowel is $2 \times 24 = 48$. Therefore, the probability of the event occurring is $\frac{48}{\text{number of sample points}}$. The five boxes can be filled by any letter to obtain a point in the sample space, so there are $5 \times 4 \times 3 \times 2 \times 1 = 120$ sample points. So the probability of the event occurring is $48/120 = 4/15$.

Example 3.2.3

Suppose we have 7 meals to distribute randomly to 7 people (one each). Three of the meals are gluten free and the other four are not. Of the 7 people, two of them cannot eat gluten. How many ways are there to distribute the meals without giving gluten to someone who cannot eat it?

We can liken this to the boxes example with each person being a box. Let the first two boxes be the people who cannot eat gluten. We have

$$\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}$$

Since we cannot place a gluten meal in boxes 1 or 2, we have that we have 3 ways to fill box 1 then 2 ways to fill box 2. So there are 6 ways distribute meals to the gluten-free people. We have

$$\boxed{\text{G}}\boxed{\text{G}}\boxed{}\boxed{}\boxed{}\boxed{}\boxed{}$$

Now there are 5 boxes to be filled with any of 5 meals. So there are $5 \times 4 \times 3 \times 2 \times 1 = 120$ ways to distribute the meals to the other 5 people. This is an implicit **and** statement, thus there are $6 \times 120 = 720$ ways to distribute the meals.

3.3 Notations

Because some calculations occur very frequently in statistics we define a notation that helps us to deal with such problems.

Factorial

We define $n!$ for any natural number n to be

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 1$$

and in order to maintain mathematical consistency we define $0!$ to be 1. This is the number of arrangements of n possible unique elements, using each once.

n to k Factors

We define $n^{(k)}$ to be

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

This is the number of arrangements of length k using each element, of n possible unique elements, at most once.

Power of

As in ordinary mathematics $n^k = \underbrace{n \times n \times \cdots \times n}_{k}$. This represents the number of arrangements that can be made of length k using each element, of n possible unique elements, as often as we wish (with replacement).

For many problems it is simply impractical to try to count the number of cases by conventional means because of how big the numbers become. Notations such as $n!$ and n^k allow us to deal with these large numbers effectively.

Example 3.3.1

An evil advertising company randomly chooses 7-digit phone numbers to call to try to sell products. Find the probabilities of the following events:

- A: the number is your phone number
- B: the first three number are less than 5
- C: the first and last numbers match your phone number

Now assume that all 7-digits are unique (chosen without replacement):

- D: the number is 210-3869
- E: the first three number are less than 5
- F: the first and last numbers are 1 and 2 respectively

A: The initial sample space contains all the ways that one can select 7 numbers from the numbers 0 to 9 **with replacement**. There are 10 choices for each of the seven numbers, therefore the sample space contains 10^7 points. Thus, since all points are equally likely, $P(A) = 1/10^7$.

B: Now if the first three numbers are less than 5, there are 5 ways (0 to 4) to select each of the first three numbers and there are 10 ways to select each of the next four numbers. So there are $5^3 \times 10^4$ points in B. Therefore, $P(B) = \frac{5^3 \times 10^4}{10^7}$

C: There is only one way to select the first number such that it matches your number and the same is true for the last number. Thus, we must only consider the middle digits. There are 10 choices each for the middle five numbers, so there are 10^5 points in C. Therefore, $P(C) = 1/10^5$.

D: The new sample space contains all the ways that one can select 7 numbers from the numbers 0 to 9 **without replacement**. There are 10 choices for the first number, 9 for the second and so on until there are 4 choices for the last number. Thus, there are $10^{(7)}$ points in the sample space and since each is equally likely, $P(D) = 1/10^{(7)} = \frac{1}{10\times 9\times 8\times 7\times 6\times 5\times 4}$.

E: If the first three numbers are less than five, there are 5 ways to select the first number, 4 for the second and 3 for the third, so there are $5^{(3)}$ ways to select the first 3 numbers. The next 4 digits may be selected from any of the 7 digits that were not used as one of the first 3. So there are $7^{(4)}$ ways to select the final four digits. Therefore, there are $5^{(3)} \times 7^{(4)}$ points in E. So, $P(E) = \frac{5^{(3)} \times 7^{(4)}}{10^{(7)}}$.

F: There is only one way to select the first and last digits as 1 and 2 respectively, so we must only consider the middle 5 digits. The 5 digits are selected from 8 numbers without replacement, so there are $8^{(5)}$ ways to do this. Therefore, $P(F) = \frac{8^{(5)}}{10^{(7)}}$.

3.4 Counting Subsets

In many problems, you will encounter a sample space, $\mathbb{S}$, of some experiment that consists of fixed-length subsets of some set.

Combinations

We define $\binom{n}{k}$ to be the number of subsets of size k that can be selected from a set of n elements. We have

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

It is read "n choose k".

Derivation of Choose

Suppose we have a set of n unique elements and we wish to select a subset of size k, such that $k \leq n$, and the elements of the subset are unique (selected without replacement). If we use the boxes metaphor we have k empty boxes.

$$\underbrace{\boxed{}\boxed{}\boxed{\cdots}\boxed{}\boxed{}}_{k}$$

There are n ways to select the first element of the subset, $(n-1)$ ways to select the second and so on until there are $(n-k+1)$ ways to select the k$^{\text{th}}$ and last element.

So there are $n^{(k)}$ ways to fill the k boxes **but** note that some of the subsets will contain all the same elements as each other but in varying order. These subsets are not unique since we do not care for the arrangement of the items in a subset. Each unique subset can be arranged to form $k!$ permutations of its k elements. Thus, the number of unique subsets, $\binom{n}{k}$, multiplied by the number of arrangements of each subset, $k!$, is $n^{(k)}$. Therefore, we have

$$\binom{n}{k} \times k! = n^{(k)}$$

So it follows that

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

3.5 Properties of Combinations

Here are a few properties of $\binom{n}{k}$.

- $n^{(k)} = \frac{n!}{(n-k)!} = n(n-1)^{k-1}$ for $k \geq 1$
- $\binom{n}{k} = \frac{n!}{(n-k)!\,k!} = \frac{n^{(k)}}{k!}$
- $\binom{n}{k} = \binom{n}{n-k}$ for all $0 \leq k \leq n$
- $\binom{n}{0} = \binom{n}{n} = 1$
- $(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$ (Binomial Theorem)

3.6 Counting Arrangements of Set with Repeated Elements

Thus far we have only discussed counting arrangements of unique items. Now, we consider a case in which we want to count the number of unique arrangements of size k of a set of n elements that are not necessarily unique.

Consider we have a set of n elements with k of those elements being unique. Let n_i be the number of appearances of the i^{th} element of the k unique elements. Thus, $n_1 + \cdots + n_k = n$. The number of different ways of selecting an arrangement of size n that uses all symbols is:

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \binom{n-n_1-n_2}{n_3} \times \cdots \times \binom{n_k}{n_k} = \frac{n!}{n_1!\,n_2!\,n_3!\cdots n_k!}$$

Derivation

Suppose we have a set of n elements with k being unique. Let the k unique items be labelled u_1 to u_k and let n_i be the number of appearance of u_i in the set of n elements. We want to form an arrangement of length n so using the boxes metaphor we have

$$\underbrace{\boxed{}\boxed{}\boxed{\cdots}\boxed{}\boxed{}}_{n}$$

We must use each of the n elements once so we must select n_1 boxes to fill with u_1's. This can be done in $\binom{n}{n_1}$ ways. Next, we must select n_2 of the remaining $n - n_1$ boxes to fill with u_2's, n_3 of the remaining $n - n_1 - n_2$ boxes to fill with u_3's, and so on until we must select n_k of the $n - n_1 - n_2 - \cdots - n_{k+1} = n_k$ remaining boxes to fill with u_k's. Therefore, there are

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \cdots \times \binom{n_k}{n_k} = \frac{n!}{\cancel{(n-n_1)!}\,n_1!} \times \frac{\cancel{(n-n_1)!}}{(n-n_1-n_2)!\,n_2!} \times \cdots \times \frac{(n_{k-1}+n_k)!}{\cancel{n_k!}\,n_{k-1}!} \times \frac{\cancel{n_k!}}{0!\,n_k!}$$

ways, which simplifies to,

$$\frac{n!}{n_1!\,n_2!\,n_3!\cdots n_k!}$$

to do so.

Chapter 4

Probability Rules

4.1 General Rules

Here are a few basic rules of probabilities. They should be relatively straightforward.

Theorem 4.1.1

For a sample space, $\mathbb{S}$, the probability of a simple event in $\mathbb{S}$ occurring is 1. That is

$$P(\mathbb{S}) = 1$$

Proof 4.1.1:

$$P(\mathbb{S}) = \sum_{a \in \mathbb{S}} P(a) = \sum_{\text{all } a} P(a)$$

$\square$

Theorem 4.1.2

Any event A in a sample space has a probability between 0 and 1 inclusive. That is

$$0 \leq P(A) \leq 1 \text{ for all } A \subseteq \mathbb{S}$$

Proof 4.1.2:

Note that A is a subset of $\mathbb{S}$, so

$$P(A) = \sum_{a \in A} P(a) \leq \sum_{a \in \mathbb{S}} P(a) = 1$$

Now, recall that $P(a) \geq 0$ for any sample point a by our probability model. Thus, since $P(A)$ is the sum of non-negative real numbers, $P(A) \geq 0$. So we have

$$1 \leq P(A) \leq 1$$

$\square$

Theorem 4.1.3

If A and B are two events such that $A \subseteq B$, that is all the sample points in A are also in B, then

$$P(A) \leq P(B)$$

Proof 4.1.3:

$$P(A) = \sum_{a \in A} P(a) \leq \sum_{a \in B} P(a) = P(B)$$

$\square$

4.2 Venn Diagrams

As we have seen already, it is helpful to think of events in a sample space as subsets of the sample space. Consider a sample space, $\mathbb{S} = \{1, 2, 3, 4, 5, 6\}$. A number is picked at random, let E be the event that the number is even. We can think of E as the subsets of $\mathbb{S}$, $\{2, 4, 6\}$ and the probability of E is the probability of any sample points in A occurs, that is 2, 4, or 6 is selected. We can represent the relationships of events in the sample space using Venn diagrams.

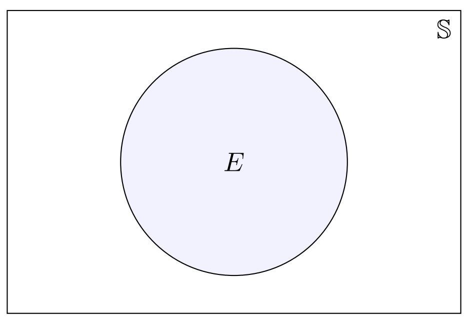


Figure 4.1: Single event E

Now, assuming the area of E is half the area of $\mathbb{S}$, we have that the probability of E is the probability that a randomly chosen point on the area of $\mathbb{S}$ will be within E.

Consider now we let $G = \{4, 5, 6\}$ be the event that that the number selected is greater than or equal to 4. We have

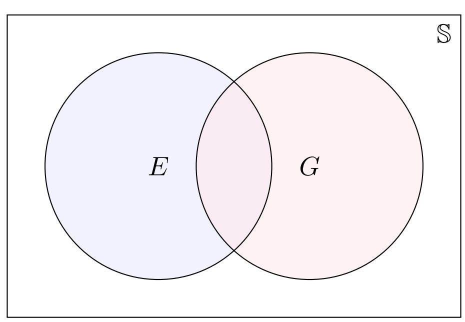


Figure 4.2: Events E and G

The total shaded region of the Venn diagram, $E \cup G$, contains all the sample points of E and G. It is the event that any outcome in either E or G, or both, occurs. Thus, $E \cup G$ is the event that E, G or both, occurs. Similarly, the union of three events is the event that at least one of the three events occur.

Consider now the intersection $E \cap G$. It is the set of all the points that are in both E and G, $\{4, 6\}$. Thus, it is the event that an outcome in both E and G occurs. So $E \cap G$ is the event that E and G both occur.

The sets $A \cap B$ and similarly $A \cap B \cap C$ are often denoted as AB and ABC respectively.

Finally, the unshaded space in Figure 4.1 is the set of all outcomes that are not in E. It is the complement of E and is denoted by $\overline{E}$. It is the event that E does not occur.

Note that the complement of $\mathbb{S}$ is the null set, that is $\overline{\mathbb{S}} = \emptyset$, and has a probability of 0.

4.3 De Morgan's Laws

Theorem 4.3.1

The following are De Morgan's Laws:

1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$

4.4 Rules for Unions of Events

Recall Figure 4.2, copied below.

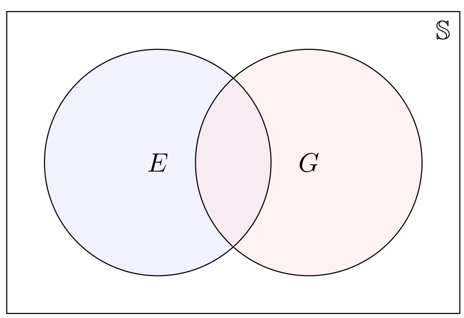


Figure 4.2: Events E and G (repeated from page 14)

We can see that the area of $E \cup G$ is not simply the sum of the areas of E and G. So we have that the probability of $E \cup G$ is not simply the sum of the probability of E and G. Rather, we must sum the probabilities and subtract the intersection (which gets included twice in the sum) to obtain $P(E \cup G)$.

Theorem 4.4.1

For any events, A and B, in a sample space, we have

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

Example 4.4.1

A number between 1 and 6 inclusive is chosen randomly. Let $E = \{2, 4, 6\}$ be the event the number is odd and let $G = \{4, 5, 6\}$ be the event that the number is greater than or equal to 4.

The probability of the number being even **or** greater than 4 is $P(E \cup G)$. Since both E and G contain 3 points of the six in the sample space, $P(E) = P(G) = 1/2$. Thus, we can see clearly that $P(E \cup G) \neq P(E) + P(G) = 1$ since $\{1\}$ is not in E or G and has a probability of $1/6$. Now, note $E \cap G = \{4, 6\}$, so $P(E \cap G) = 1/3$. We have

$$P(E \cup G) = P(E) + P(G) - P(E \cap G) = 1/2 + 1/2 - 1/3 = 2/3$$

Now consider the case of the union of three events.

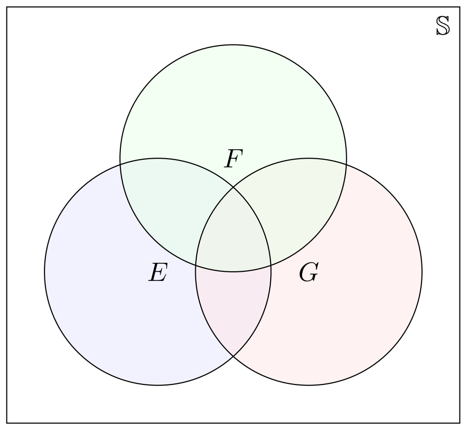


Figure 4.4: Three events

Let A_I be the area on the Venn diagram of the event I. The area of the union once again is not simply the sum of the areas $(A_E + A_G + A_F)$. Instead we can reason out that when we add the three areas we include $A_{E\cap G}$, $A_{G\cap F}$, and $A_{F\cap E}$ twice each and $A_{E\cap G\cap F}$ three times. The sum of these doubly counted areas $(A_{E\cap G} + A_{G\cap F} + A_{F\cap E})$ also includes $A_{E\cap G\cap F}$ three times. Thus, when we subtract the area of the doubly counted segments, $A_{E\cap G\cap F}$ is also subtracted three times leaving this area unaccounted for. Therefore we then add $A_{E\cap G\cap F}$ to find the complete area of $E \cup G \cup F$.

Theorem 4.4.2

For any events, A, B and C, in a sample space, we have

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

4.5 Mutually Exclusive Events

Events A and B are mutually exclusive if and only if $A \cap B = \emptyset$. More simply, the events A and B cannot both occur in one experiment because they share no points in common and only one sample point is achieved.

In general, events $A_1, A_2, A_3, \ldots, A_n$ are mutually exclusive if and only if $A_i \cap A_j = \emptyset$ for all $i \neq j$. This means that at most one of these events may occur in any one experiment.

Probability of the Unions of Mutually Exclusive Events

Consider the Venn diagram of two mutually exclusive events, E and G. Clearly the probability of the

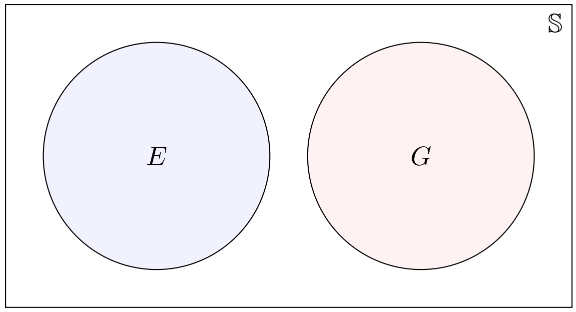


Figure 4.5: Two mutually exclusive events

intersection of two mutually exclusive events is 0, since it doesn't contains any sample points. So we have

$$P(E \cap G) = 0$$

Another intrinsic property of mutually exclusive events that we can see on a Venn diagram is that the area of $E \cup G$ is the sum of the areas of E and G. Therefore, unlike in previous examples, the probability of $E \cup G$ is the sum of the probabilities of E and G.

Theorem 4.5.1

For mutually exclusve events, A and B, in a sample space, we have

$$P(A \cup B) = P(A) + P(B)$$

Theorem 4.5.2

More generally for n mutually exclusive events, $A_1, A_2, \ldots, A_n$, in a sample space, we have

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = \sum_{i=1}^{n} P(A_i)$$

Probabilities of Complements

Theorem 4.5.3

For any event A, we have

$$P(A) = 1 - P(\overline{A})$$

Proof 4.5.1:

Recall the complement of an event consists of all the sample points not in the event. Thus, for any event A, its complement $\overline{A}$ contains no points in common with A. So $A \cap \overline{A} = \emptyset$ and A and $\overline{A}$ are mutually exclusive, by definition. Now, consider $A \cup \overline{A}$, it spans the whole of the sample space so we have $P(A \cup \overline{A}) = 1$ and since A and $\overline{A}$ are mutually exclusive, we have

$$P(A) + P(\overline{A}) = 1$$

and it follows that $P(A) = 1 - P(\overline{A})$, as required. □

4.6 Independence of Events

Events A and B are said to be independent if and only if $P(A \cap B) = P(A)P(B)$. Otherwise they are dependent events.

In general, events $A_1, A_2, A_3 \ldots, A_n$ are independent if and only if

$$P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \cdots \cap A_{i_n}) = P(A_{i_1}) + P(A_{i_2}) + P(A_{i_3}) + \cdots + P(A_{i_n})$$

for all sets $\{ i_1, i_2, i_3, \ldots, i_k \}$ of distinct subscripts chosen from $\{ 1, 2, 3, \ldots, n \}$.

Example 4.6.1

Consider an experiment in which a fair die is tossed twice. We define the following events:

- A: The first number rolled is a six
- B: The second number rolled is a six
- C: The sum of the numbers rolled is less than or equal to seven
- D: The sum of the numbers rolled is equal to seven

Suppose the event A occurs. Does this have any impact on the probability of B, C or D occurring?

It is quite clear to see that the events A and B are independent events since rolling a six on the first toss has no impact on the number that will be rolled on the second toss. Now, events B and C from the onset appear to be dependent since if you roll a six on the first toss you must roll a one to make your total less than or equal to seven. To confirm this consider the sample space

$$\left\{ \begin{array}{l} (1,1),\ (1,2),\ (1,3),\ (1,4),\ (1,5),\ (1,6) \\ (2,1),\ (2,2),\ (2,3),\ (2,4),\ (2,5),\ (2,6) \\ (3,1),\ (3,2),\ (3,3),\ (3,4),\ (3,5),\ (3,6) \\ (4,1),\ (4,2),\ (4,3),\ (4,4),\ (4,5),\ (4,6) \\ (5,1),\ (5,2),\ (5,3),\ (5,4),\ (5,5),\ (5,6) \\ (6,1),\ (6,2),\ (6,3),\ (6,4),\ (6,5),\ (6,6) \end{array} \right\}$$

We can count that 21 of the sample points have sums less than or equal to seven. So the probability of C occurring is $P(C) = 21/36 = 7/12$. We also have that $P(A) = 1/6$. So $P(A)P(C) = 7/72$ but we can count that $A \cap C$ contains only one sample point and hence has a probability of $1/36$. Thus, $P(A)P(C) \neq P(A \cap C)$ so A and C are dependent events.

At first glance, we see that upon rolling a six as the first number you must roll a 1 for the sum to equal seven. So at first glance, events A and D seem to be independent however it would be naïve to assume this. We can count from the sample space that event D contains 6 points and so has a probability $P(D) = 6/36 = 1/6$ and $P(A) = 1/6$. So $P(A)P(D) = 1/36$. Now, we can count that the event $A \cap D$ contains only one point, (1,6) and so has a probability $P(A \cap D) = 1/36$. Therefore, $P(A \cap D) = P(A)P(D)$ and the events A and D are independent.

Chapter 5

Conditional Probability

Often we need to calculate the probability of some event A occurring while knowing that some other event B has already occurred. We call this the conditional probability of A **given** B and denote it by $P(A|B)$.

The conditional probability of event A, given event B, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ for } P(B) > 0$$

5.1 Theorems and Rules for Conditional Probability

Theorem 5.1.1

For any two events A and B defined on the same sample space, with $P(A) > 0$ and $P(B) > 0$, events A and B are independent if and only if $P(A|B) = P(A)$ or $P(B|A) = P(B)$.

Proof 5.1.1:

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

and by definition of independence, A and B are independent if and only if $P(A \cap B) = P(A)P(B)$ which is true if and only if $P(A|B) = P(A)$. Without loss of generality we can swap events A and B and arrive at the conclusion. □

Product Rules

Theorem 5.1.2

Let A, B, C and D be events on a sample space, with $P(A), P(B), P(C), P(D) > 0$. We have

$$P(A \cap B) = P(A)P(B|A)$$
$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$
$$P(A \cap B \cap C \cap D) = P(A)P(B|A)P(C|A \cap B)P(D|A \cap B \cap C)$$

and so on...

Proof 5.1.2:

The first statement come directly from the definition of conditional probability

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

For the second we have

$$\begin{aligned} P(A)P(B|A)P(C|A \cap B) &= P(A \cap B)P(C|A \cap B) && \text{by the first statement} \\ &= P(A \cap B)\frac{P(A \cap B \cap C)}{P(A \cap B)} && \text{by definition of conditional probability} \\ &= P(A \cap B \cap C) \end{aligned}$$

and so on... □

Law of Total Probability

Theorem 5.1.3

Let $A_1, A_2, A_3, \ldots, A_k$ be mutually exclusive events on a sample space and let B be an event on the same sample space. We have

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3) + \cdots + P(B \cap A_k) = \sum_{i=1}^{k} P(A_i)P(B|A_i)$$

Proof 5.1.3:

Note that the events $A_i \cap B$ for $1 \leq i \leq k$ are all mutually exclusive events since A_i's are mutually exclusive. Thus, the union of the $A_i \cap B$'s is B, that is

$$(A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B) \cup \cdots \cup (A_k \cap B) = B$$

So by Theorem 4.5.1, we have

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + \cdots + P(A_k \cap B)$$

and by Theorem 5.1.2 (Product Rule), we have

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) + \cdots + P(A_k)P(B|A_k) = \sum_{i=1}^{k} P(A_i)P(B|A_i)$$

as required. □

Bayes' Theorem

Theorem 5.1.4

Let A and B be events on a sample space, with $P(B) > 0$. We have

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

Proof 5.1.4:

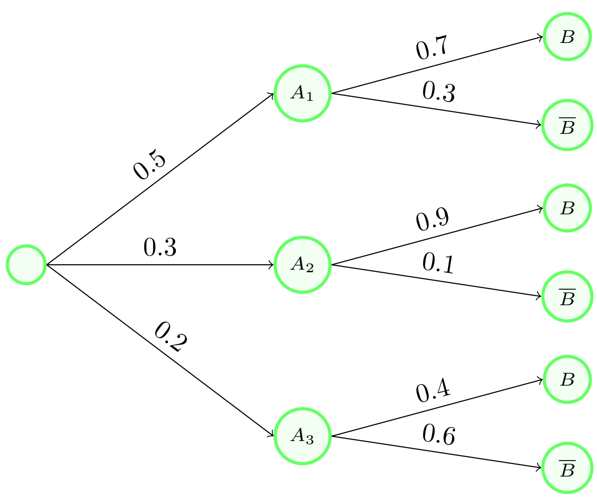
$$
\begin{aligned}
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} &\quad \text{by Theorem 4.7.2 (Product Rule)} \\
= \frac{P(B|A)P(A)}{P(A \cap B) + P(\overline{A} \cap B)} &\quad \text{by Theorem 4.7.3 (Law of Total Probability)} \\
= \frac{P(B|A)P(A)}{P(B|\overline{A})P(\overline{A}) + P(B|A)P(A)} &\quad \text{by Theorem 4.7.2 (Product Rule)}
\end{aligned}
$$

□

Bayes' Theorem allows us to find the conditional probability of some event A given B, in terms of the probability of B given A. It allows us calculate conditional probabilities using the reversed order of conditioning.

5.2 Tree Diagrams

Tree diagrams are a technique that we can use to keep track of conditional probabilities. We start from a single node and draw new branches to separate nodes for each event. Each node represents the event occurring. On each branch we write the probability of event it leads to occurring. To find the probability of any node we multiply the probabilities of the branches leading to the node.



$$P(A_1 \cap B) = P(B|A_1)P(A_1) = 0.5 \times 0.7$$
$$P(A_1 \cap \overline{B}) = P(\overline{B}|A_1)P(A_1) = 0.5 \times 0.3$$
$$P(A_2 \cap B) = P(B|A_2)P(A_2) = 0.3 \times 0.9$$
$$P(A_2 \cap \overline{B}) = P(\overline{B}|A_2)P(A_2) = 0.3 \times 0.1$$
$$P(A_3 \cap B) = P(B|A_3)P(A_3) = 0.2 \times 0.4$$
$$P(A_3 \cap \overline{B}) = P(\overline{B}|A_3)P(A_3) = 0.2 \times 0.6$$

Figure 5.1: Tree diagram

The probability of all the the branches leading outward from each node must sum to 1 since at least one outcome must occur.

Chapter 6

Useful Sums and Series

This chapter includes a few useful sums and series that show up in the following chapters.

6.1 Geometric Series

$$\sum_{i=0}^{n-1} r^i = 1 + r + r^2 + \cdots + r^{n-1} = \frac{1-r^n}{1-r}$$

For $|r| < 1$, we have

$$\sum_{i=0}^{\infty} r^i = 1 + r + r^2 + \cdots = \frac{1}{1-r}$$

Other identities can be obtained from this one by differentiation. For example we have

$$\frac{\mathrm{d}}{\mathrm{d}r} \sum_{i=0}^{\infty} r^i = \sum_{i=0}^{\infty} i r^{i-1} = \frac{\mathrm{d}}{\mathrm{d}r} \frac{1}{1-r} = \frac{1}{(1-r)^2}$$

6.2 Binomial Theorem

The binomial theorem describes the algebraic expansion of powers of a polynomial.

$$(1+t)^n = 1 + \binom{n}{1} t^1 + \binom{n}{2} t^2 + \cdots + \binom{n}{n} t^n = \sum_{x=0}^{n} \binom{n}{x} t^x$$

for any positive integer n and real number t.

A more general form of this theorem that holds even when n is not a positive integer is

$$(1+t)^n = \sum_{x=0}^{\infty} \binom{n}{x} t^x, \quad \text{for } |t| < 1$$

It is an important skill to be able to recognize if an infinite, or otherwise, polynomial with binomial coefficients can be reduced to a simple polynomial raised to a power.

6.3 Multinomial Theorem

The multinomial theorem is a generalization of the binomial theorem. It describes the algebraic expansion of powers of a sum in terms of powers of the terms in the sum.

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1+k_2+\cdots+k_m=n} \binom{n}{k_1, k_2, \ldots, k_m} \prod_{t=1}^{m} x_t^{k_t}$$

Another common form in which this theorem may be represented is

$$(x_1 + x_2 + \cdots + x_m)^n = \sum_{k_1+k_2+\cdots+k_m=n} \frac{1}{k_1!\,k_2!\cdots k_m!}(x_1^{k_1} x_2^{k_2} \cdots x_m^{k_m})$$

The summation is over all non-negative integers, $k_1, k_2, \ldots, k_m$ such that $k_1 + k_2 + \cdots + k_m = n$

6.4 Hypergeometric Identitiy

$$\sum_{x=0}^{\infty} \binom{a}{x}\binom{b}{n-x} = \binom{a+b}{n}$$

Proof 6.4.1:

We begin with the equality

$$(1+y)^{a+b} = (1+y)^a \times (1+y)^b$$

Now by Binomial Theorem we have

$$\sum_{k=0}^{a+b} \binom{a+b}{k} y^k = \sum_{i=0}^{a} \binom{a}{i} y^i \times \sum_{j=0}^{b} \binom{b}{j} y^j$$

Consider the coefficient of y^k on the right hand side. It is the sum of all the binomial terms such that $i + j = k$. Thus, the coefficient of y^k on the right hand side is

$$\sum_{i=0}^{\min\{a,k\}} \binom{a}{i}\binom{b}{k-i}$$

and since when $i > a$ or $i > k$ the term is 0 we can increase the sum to infinity. Thus, since the coefficient on the right hand side is equal to that on the left hand side we have

$$\binom{a+b}{k} = \sum_{i=0}^{\infty} \binom{a}{i}\binom{b}{k-i}$$

□

When x becomes significantly large, the terms of the summation become 0 since

$$\binom{n}{x} = \binom{n}{n-x} = 0, \quad \text{for } x > n$$

6.5 Exponential Series

This is an example of a Maclaurin series expansion.

$$e^t = \frac{t^0}{0!} + \frac{t^1}{1!} + \frac{t^2}{2!} + \cdots = \sum_{n=0}^{\infty} \frac{t^n}{n!}, \quad \text{for all } t \in \mathbb{R}$$

The following limit definition of the exponential function is also useful

$$e^t = \lim_{n\to\infty} \left(1 + \frac{1}{t}\right)^n, \quad \text{for all } t \in \mathbb{R}$$

6.6 Integer Series

The following are useful equalities involving sums of integers.

$$\begin{aligned} 1 + 2 + 3 + \cdots + n &= \frac{n(n+1)}{2} \\ 1^2 + 2^2 + 3^2 + \cdots + n^2 &= \frac{n(n+1)(n+2)}{6} \\ 1^3 + 2^3 + 3^3 + \cdots + n^3 &= \left[\frac{n(n+1)}{2}\right]^2 \end{aligned}$$