Chapter 1

Some Pre-Requisites on Calculus, Games of Chance, and Statistics

- Learning objectives:
  - review some important types of functions that will be used in the course
  - review differentiation, integration, and infinite series, with an emphasis on techniques/ideas that will be used later in the course
  - introduce and explain the terminology for certain games of chance
  - review some fundamental statistics that will be used throughout the course

1.1 Calculus Pre-Reqs

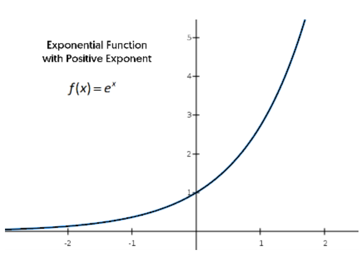
This section contains the important terms, ideas, and formulas necessary for success in this course. This is not a calculus book, so proofs are not included. The more unfamiliar this material seems to you, the more you should study it and do practice problems (refer to any calculus textbook).

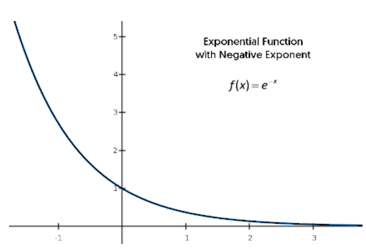
1.1.1 The Exponential Function and its Inverse (Natural Logarithm Function)

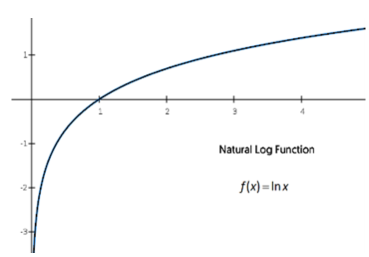
Two very important functions in many fields of mathematics are the exponential functions and the logarithmic functions. Probability and statistics are no exception to this fact. Knowledge of these two functions is essential to success in several chapters. It is assumed here that you have been exposed to these functions before and we only list some essential properties of these functions. We will exclusively be using the base e in this text. So the functions of interest to us are the (natural) exponential function and the natural logarithm function.

The exponential function is written $f(x) = e^x$, and the natural log function is written $f(x) = \ln x$. Note that this is the log function with base e and could also be written as $f(x) = \log_e x$. There are several important calculation rules for these two functions given below, along with the graphs of the functions.

The domain for the exponential function is all real numbers (i.e., $\mathbb{R}$). The range of the exponential function is all positive y-values (i.e., $y > 0$). The domain for the natural log function is all positive x-values (i.e., $x > 0$), and the range is all real numbers (i.e., $\mathbb{R}$).







The tables below collects some rules and properties for the exponential function and the natural log function:

$e^a e^b = e^{a+b}$	$\ln(ab) = \ln a + \ln b$
$(e^a)^b = e^{ab}$	$\ln a^b = b \ln a$
$\frac{e^a}{e^b} = e^{a-b}$	$\ln\left(\frac{a}{b}\right) = \ln a - \ln b$

$\lim_{x \to -\infty} e^x = 0$	The limit goes to zero very quickly.
$\lim_{x \to \infty} e^x = \infty$	The limit goes to infinity very quickly.
$\lim_{x \to \infty} \ln x = \infty$	The limit goes to infinity very slowly.
$\lim_{x \to 0^+} \ln x = -\infty$	The limit goes to negative infinity very quickly.

The exponential function and the natural log functions are inverses of each other. That is, $\ln e^x = x$ for all x and $e^{\ln x} = x$ for $x > 0$.

1.1.2 Derivatives

The derivative of a function measures the instantaneous rate of change of the y-variable (output) with respect to the rate of change of the x-variable (input). The derivative of a function tells us the slope of the tangent line to the graph of $y = f(x)$ for any x-value.

The definition of the derivative is

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

Fortunately, there are derivative formulas/rules which allow us to sidestep using the definition to determine the derivative of a function. If the function is differentiable at a given x-value, then it is "smooth" at that x-value. That is, the function does not jump (since differentiable implies continuous) or have a corner at that x-value. Most of the continuous functions used in this book are also differentiable.

There are many possible notations to denote the derivative of a function $y = f(x)$. Some notations include y', $f'(x)$, $\frac{dy}{dx}$, $D_x[f(x)]$, etc.

Below is a list of some simple and important functions. Remember, this is only derivative essentials. There are many other important functions that will not be needed for this course in probability, so we do not discuss them here. Let $y = f(x)$ be the function of interest and let c be a constant.

$f(x)$	$f'(x)$	Example
x^n	nx^{n-1}	$f(x) = x^7 \to f'(x) = 7x^6$
$\ln x$	$\frac{1}{x}$	
e^x	e^x	
$cf(x)$	$cf'(x)$	$f(x) = 4x^3 \to f'(x) = 12x^2$

It is important to know the list and also important to realize that if a function is not on the list, we are likely stuck. That is, it is important to know "the list is all we know" (until we expand the list or learn new rules).

Because the list is all we know, we need to convert some problems to look like something on the list. The table below includes some examples like this.

$f(x)$	Rewrite	$f'(x)$	Reason
$\sqrt{x}$	$x^{1/2}$	$\frac{1}{2}x^{-1/2}$	There is no square root in our list. Convert it to an exponent.
$\frac{1}{x^3}$	x^{-3}	$-3x^{-4} = \frac{-3}{x^4}$	There is no power in the denominator in our list. Convert it to an exponent.

The sum/difference, product, and quotient rules allow us to determine many more derivatives. These rules are needed throughout the text.

Theorem 1.1.1 The Sum/Difference Rule

The sum/difference rule tells us how to differentiate the sum or difference of two functions: if $y = f(x) \pm g(x)$, then

$$y' = f'(x) \pm g'(x).$$

Example 1.1.1

Differentiate $y = x^3 + 5x^4$.

Theorem 1.1.2 The Product Rule

The product rule tells us how to differentiate the product of two functions: if $y = f(x)g(x)$, then

$$y' = f'(x)g(x) + g'(x)f(x).$$

Example 1.1.2

Differentiate $y = x^3 \ln x$.

Theorem 1.1.3 The Quotient Rule

The quotient rule tells us how to differentiate the quotient of two functions: if $y = \frac{f(x)}{g(x)}$, then

$$y' = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}.$$

Example 1.1.3

Differentiate $y = \frac{x^3}{x^2+5}$.

The last rule we will consider allows us to differentiate a composition of functions (i.e., a function of a function).

Theorem 1.1.4 The Chain Rule

Let $f(x)$ and $u(x)$ be functions of x with derivatives $f'(x)$ and $u'(x)$. Then the derivative of $y = f(u(x))$ is

$$y' = f'(u(x)) \cdot u'(x).$$

Our derivative formula list expands to include:

$f(x)$	$f'(x)$	Example $f(x)$	Solution $f'(x)$
u^n	$nu^{n-1}u'(x)$	$(x^2+e^x)^4$	$4(x^2+e^x)^3(2x+e^x)$
$\ln[u(x)]$	$\frac{u'(x)}{u(x)}$	$\ln\left[x^4+x^2\right]$	$\frac{4x^3+2x}{x^4+x^2}$
$e^{u(x)}$	$e^{u(x)}u'(x)$	e^{6x^3}	$e^{6x^3} \cdot 18x^2$

1.1.3 Integrals

In this course, we will have a great need to determine areas under curves and volumes under surfaces. For this, we will need to use another fundamental calculus tool: the definite integral. We start by reviewing our basic integral formulas as indefinite integrals. As was the case with derivatives, we will have a set of simple known integral formulas and then progress into more general formulas. There are many integral formulas and techniques that we will not be reviewing, as they will not be needed in this course. The basic formulas we will need, along with some examples, include those in the table below.

$f(x)$	$\int f(x)\,dx$	Example
x^n $(n \neq -1)$	$\frac{x^{n+1}}{n+1} + C$	$f(x) = x^7 \to \int f(x)\,dx = \dfrac{x^8}{8} + C$
$\frac{1}{x}$	$\ln\lvert x\rvert + C$	
e^x	$e^x + C$	

Once again, it is important to know the list and also important to realize that if a function is not on the list, we are likely stuck. That is, it is important to know "the list is all we know" (until we expand the list or learn new rules or techniques). As we saw with derivatives, we need to convert some problems to look like something on the list.

$\int f(x)\,dx$	Rewrite	Answer	Reason
$\int \sqrt{x}\,dx$	$\int x^{1/2}\,dx$	$\frac{2}{3}x^{3/2} + C$	There is no square root on our list. Convert it to an exponent.
$\int \dfrac{1}{x^3}\,dx$	$\int x^{-3}\,dx$	$\frac{x^{-2}}{-2} + C = \frac{-1}{2x^2} + C$	There is no power in the denominator on our list. Convert it to an exponent.

As with the derivative, we have a more general form of the integral. Let $u = u(x)$ be some function of x and let $u' = u'(x)$ be its derivative with respect to x.

$f(x)$	$\int f(x)\,dx$	Example $f(x)$
$u^n u'$ $(n \neq -1)$	$\frac{u^{n+1}}{n+1} + C$	$\int (x^2+1)^7 \cdot 2x\,dx = \dfrac{(x^2+1)^8}{8} + C$
$\frac{u'}{u}$	$\ln\lvert u\rvert + C$	$\int \dfrac{2x}{x^2+6}\,dx = \ln\left(x^2+6\right) + C$
$e^u u'$	$e^u + C$	$\int e^{x^3+8} \cdot 3x^2\,dx = e^{x^3+8} + C$

A useful rule to know is that constants can come in or out of integrals (variables cannot):

$$\int c f(x)\,dx = c \int f(x)\,dx$$

for any function $f(x)$ and any constant c.

An example where this is helpful is the following problem.

Example 1.1.4

Determine $\int \left(x^2+1\right)^7 \cdot x\,dx$.

Our last formula and example is a type of integral that will come up a few times in this course. To solve this type of integral, we need a technique called integration by parts. An entire section would be devoted to this topic in a calculus book with a wide variety of examples. We will stick to the one that will be needed for our studies.

The general formula for integration by parts is $\int uv'\,dx = uv - \int vu'\,dx$. Note that this formula provides a technique to transform the integral you want, $\int uv'\,dx$ into a known quantity, uv, minus a new integral, $\int vu'\,dx$. The hope is that this new integral is one that is in our list.

Example 1.1.5

Determine $\int xe^x\,dx$.

Note that the above is just a review of the basics that will be needed, and this section does not take the place of a calculus textbook. If these simple examples of indefinite integrals are not clear to you, you will need to do some more calculus review.

It should be noted that there is a relationship between integration and differentiation. If you integrate a function and then differentiate your result, you end up where you started. In other words, if $\displaystyle\int f(x)\,dx = F(x) + C$, then $F'(x) = f(x)$. Do you recall the name of this theorem that is very fundamental to the theory of calculus?

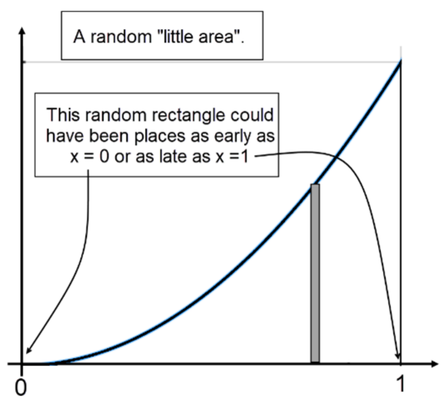
As stated earlier, we need to determine areas and volumes in this course. For areas, we need to compute a single integral that is then evaluated at the endpoints. This is known as a definite integral.

Example 1.1.6

Suppose that we want to determine the area under the graph of $y = x^2$ between $x = 0$ and $x = 1$.

Objectively, we want to create "little areas", or rectangles, under the curve. We then make the rectangles infinitesimally thin and add up their areas. The result of this process can be shown to converge to the area under the graph and is calculated via the definite integral. Isn't calculus great?!

Without calculus, we would have to compute and add the areas of thin rectangles that start at $x = 0$ and stop at $x = 1$. The rectangles all have height $y = x^2$ in this problem and $y = f(x)$ in a general problem. The width of all of the rectangles will become arbitrarily small (dx), and we determine the limit of this sum by the definite integral.



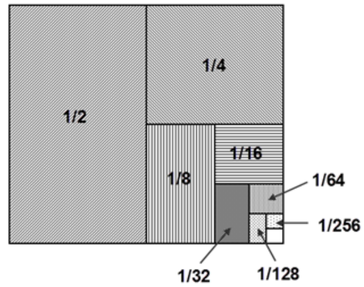
Recall that to evaluate a definite integral, we start by writing down any antiderivative of the integrand. We then substitute the upper limit into the evaluated integral and subtract the result when the lower limit is substituted into the evaluated integral. Below is how we would write the Riemann sum in this case, which converges to the definite integral as the width of the rectangles goes to 0.

$$A \approx \sum_{i=1}^{n} x_i^2 \Delta x_i \to \int_0^1 x^2\,dx = \frac{1}{3}x^3\Big|_0^1 = \frac{1}{3} - \frac{0}{3} = \frac{1}{3} = A$$

1.1.4 Infinite Series

We will be dealing with discrete sets that are not finite. Therefore, some knowledge of infinite series is required for those sets.

Recall that a sum of the form $\sum_{k=1}^{\infty} a_k$ is called an infinite series. Adding positive values indefinitely does not necessarily yield an infinite sum, as seen in the picture below:



Notice that the square is 1×1, so has an area of $A = 1$ square unit. If we start out by shading half of the square, then half of the remaining, then half of the remaining again, and so on, we will never completely fill the square. We can view the area of the square as the sum of the shaded areas, and since the total area of the square is 1, we have a sum of infinitely many positive terms that is not infinite. In fact, it can be shown that the infinite series in this case converges to 1, the area of the square.

The amount of area added each step of the way is half of the previous area. So the total area is increasing after each iteration, but will always be less than 1 (the total area of the square). The shaded area after n-iterations would be

$$S_n = \sum_{k=1}^{n} \left(\frac{1}{2}\right)^k = \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots + \frac{1}{2^n} = 1 - \frac{1}{2^n}.$$

Definition 1.1.1

We define the sum of an infinite series to be the limit of the n^{th} partial sum (if it exists):

$$\sum_{k=1}^{\infty} a_k = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k.$$

Thus, in the previous example, we have

$$A = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k = \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots = \lim_{n\to\infty} S_n = \lim_{n\to\infty} \left(1 - \frac{1}{2^n}\right) = 1.$$

Definition 1.1.2

Given an infinite series $\displaystyle\sum_{k=1}^{\infty} a_k$, if $\displaystyle\lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k = S$, then we say the series converges to the value S.

Definition 1.1.3

Given an infinite series $\displaystyle\sum_{k=1}^{\infty} a_k$, if $\displaystyle\lim_{n\to\infty} S_n = \lim_{n\to\infty} \sum_{k=1}^{n} a_k$ does not converge, then we say the series diverges.

In general, the ability to determine the actual sum of an infinite series is pretty rare. That is, just writing out some infinite series, there would be little hope of determining the actual sum (if one exists).

One type of series whose sum we can actually determine is called the geometric series. A geometric series is a series in which the ratio of consecutive terms is constant. Such a series can be written in the form

$$\sum_{n=0}^{\infty} ar^n = a + ar + ar^2 + ar^3 + \cdots.$$

The number r is called the common ratio. If $|r| < 1$, the series converges to $\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$. If $|r| \geq 1$, the series diverges.

Example 1.1.7

If possible, determine the sum of the series $\sum_{n=0}^{\infty} 6(0.8)^n$.

Example 1.1.8

If possible, determine the sum of the series $\sum_{n=2}^{\infty} \left(\frac{1}{3}\right)^{2n+1}$.

Another type of series whose sum can be computed is the telescoping (or collapsing) series. You studied this type of series in Calculus II, and we will not cover that here. If necessary, you can consult a calculus textbook for the details.

As noted above, there are very few series where the final sum can actually be calculated by hand. Because of this, we often focus on simply determining if the series converges or diverges. A special class of infinite series for which we can easily determine convergence/divergence are the p-series. These are series of the form $\sum_{n=1}^{\infty} \frac{1}{n^p}$. A p-series converges if $p > 1$ (although there is no formula for calculating the sum) and diverges if $p \leq 1$.

For instance, the so-called harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ diverges (since $p = 1$) and so does the series $\sum_{n=1}^{\infty} \frac{1}{n^{1/2}}$ (since $p = \frac{1}{2}$). But the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$ and $\sum_{n=1}^{\infty} \frac{1}{n^{4/3}}$ are examples of convergent p-series (since $p = 2$ and $p = \frac{4}{3}$, respectively).

Generally speaking, the terms in the series do not go to zero fast enough for convergence when $p \leq 1$, but do go to zero fast enough for convergence when $p > 1$. These types of series will allow us to look at some interesting probability distributions later.

We conclude this section with some final ideas and examples. When deciding if a series converges, all focus is on how fast the general term (the terms you are adding indefinitely) goes to 0. If the general term goes to 0 slowly, the series will diverge (no upper bound on the partial sums). If the general term goes to 0 quickly, the series will converge. The reason a geometric series converges when $|r| < 1$ is that r^n goes to 0 VERY quickly in this case. Other "fast" functions involve factorials (in the denominator). The p-series defined above has a general term that goes to 0 fast enough as long as $p > 1$. Some comparison rules (basic and limit) from calculus allow us to handle a variety of series that are not quite p-series, but resemble p-series.

Example 1.1.9

Determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{n}{n^2+1}$ converges or diverges by comparison with a p-series.

Example 1.1.10

Determine whether the series $\displaystyle\sum_{n=1}^{\infty} \frac{n^2}{n^4+1}$ converges or diverges by comparison with a p-series.

Example 1.1.11

Determine whether the series $\sum_{n=1}^{\infty} \frac{n^2}{2^n}$ converges or diverges.

Note that in order to show that the terms go to 0 in the previous example may require the use of L'Hôpital's rule from calculus.

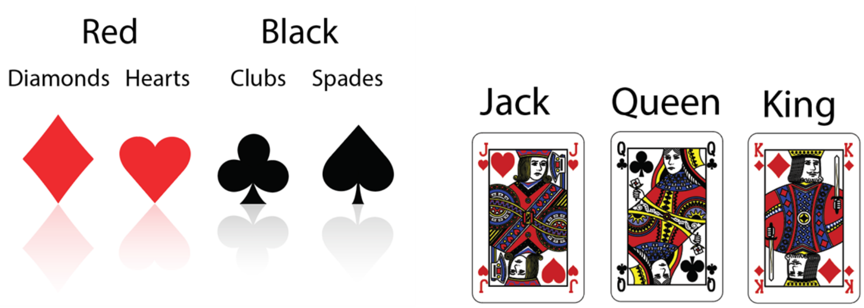
1.2 Games of Chance Pre-Reqs

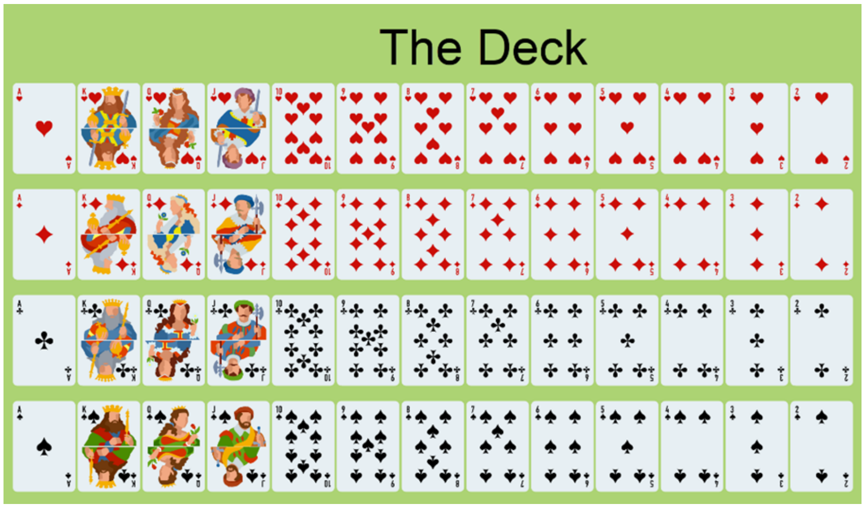
So much of the earliest development and study of probability was related to games of chance. We will restrict ourselves to modern games of chance and some made up games. Most of our examples will involve either a deck of playing cards, dice, or roulette. These three devices are chosen not just because they offer a look at real world examples of probability, but they are tangible and it is easy to visualize the collection of all things that can take place. The examples available with these three devices are plentiful and very helpful in our quest for an understanding of probability and the techniques used in probability.

1.2.1 A Deck of Playing Cards

There are many games that deal with a deck of cards. Here, we investigate the contents of a deck of playing cards. A deck of cards consists of 52 cards that can be decomposed in a variety of ways:

- there are two colors: red and black (26 cards of each color)
- there are four suits: diamonds, hearts, clubs, and spades (13 cards of each suit)
- within each suit, there are thirteen types of cards (often called the values): Ace=A, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack=J, Queen=Q, and King=K
- jacks, queens, and kings are referred to as face cards (because there is a face on the card)





1.2.2 Poker

Many nice examples in this book are related to the game of poker. We need to know the basics of each type of poker hand. We do not need to know which hand beats which in the game. We only need to know the cards that make up each type of hand. In a hand of poker, your final hand is the highest ranking hand of five cards. The order of the cards is irrelevant. Each type of hand will be illustrated through example(s).

- One Pair: consists of one pair of cards of the same value, and three other cards of values different from each other and different from the pair. Suits or color do not factor in.
  - Ace, Ace, 7, 3, 5 is a pair of aces
  - Ace, 6, 4, 6, 10 is a pair of sixes
  - 6, 7, 2, 7, 2 is not one pair since there are two pair
  - King, Jack, King, King, 8 is not one pair since there are three cards of the same value
- Two Pair: consists of two pair of different values and one card of value different from those in the two pair. Suits or color do not factor in.
  - King, 8, 8, 4, King is two pair
  - 6, 7, 2, 7, 2 is two pair
- Three of a Kind (slang is Trips): consists of three of one specific card value and two cards of values different from each other and different from the trips. Suits or color do not factor in.
  - 5, 6, 5, 5, 10 is three fives (trip fives)
  - 5, Jack, Jack, 5, Jack is not three of a kind since there are two fives
- A Straight: consists of five consecutive cards by rank (value) in any order. The Ace can be used as the card below a two (like as if it were a one) or as the card above a King. If all five of the straight cards are of the same suit, the hand is not considered a straight, but rather a straight flush (see below).
  - Ace, 2, 3, 4, 5 is a five-high straight
  - 2, 5, 3, Ace, 4 is the same five-high straight as above. The order they come in does not matter.
  - 10, Jack, Queen, King, Ace is an Ace-high straight
- A Flush: consists of any five cards of the same suit.
  - 2♥, 6♥, 4♥, 10♥, K♥ is a heart flush
- Full House: consists of three cards of one value and two cards of another value.
  - 5, Jack, Jack, 5, Jack is a full house, referred to as Jacks full of fives.

- Jack, Jack, 5, 5, 5 is a different full house, referred to as fives full of Jacks.

- Four of a Kind: consists of four cards of the same value and one card of some other value.
  - 7, 7, 7, 7, 3 is four of a kind
- Straight Flush: a straight in which all five cards are of the same suit.
  - 3♥, 4♥, 5♥, 6♥, 7♥ is a seven-high straight flush.
- Royal Flush: an Ace-high straight flush.
  - 10♥, J♥, Q♥, K♥, A♥ is a royal flush.

1.2.3 A Single Die and Set of Dice

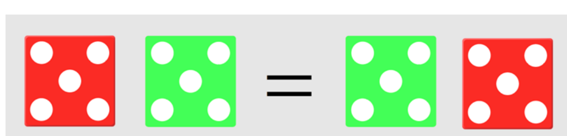
As with cards, dice offer a chance to develop our thinking and strategies as related to probability. Our dice examples will generally be devoted to six-sided dice. Each die has spots on it. The number of spots are one through six, each used exactly once.

When rolling two or more dice, it would be to our advantage to use dice of different colors. This will help ensure that we visualize all of the possibilities when rolling multiple dice.

For instance, consider rolling two dice. A simple mistake to make is thinking that a five and a three is the same as a three and a five. But if we use a green die and a red die, we easily see that the two are different.

Obviously, rolling more dice leads to a similar conclusion. That is, 3, 4, 1 is not the same as 4, 3, 1 or 1, 4, 3 or any other rearrangement.

Note that a red five and a green five is the same as a green five and a red five. In other words, there is only one 5, 5.



1.2.4 The Game of Roulette

A simple and mindless game in nearly every casino is the game of roulette. In this game, a wheel is subdivided into 2 green spaces, 18 red spaces, and 18 black spaces. A ball is then sent spiraling around the wheel and it eventually lands in one of the 38 spaces. Players wager on where the ball will land and are offered a variety of types of bets. The game takes zero skill and zero knowledge. It is a game of total chance and no amount of intelligence can alter your chances of winning.

1.3 The Mean of a Data Set

One of the things that we love to do with a data set is to summarize or describe it. Suppose that we have a data set that consists of $4, 4, 4, 4, 4, 4, 4, 4$.

Now somebody asks you to fully describe the data with a single number. This is a no-brainer. You would use the number 4. Using this number is "risk free", i.e., all of the numbers are equal to 4 so there is no risk in using this number to perfectly describe the data set. The number 4 accurately describes every number in the data set. So the number 4 is "spot on" in describing the data set.

Now suppose that we have a different data set: $4, 4, 4, 5, 5, 5$.

Again, somebody asks you to describe the data set with a single number. If you use the number 4 to describe the data, you would perfectly describe 3 of the numbers, but not the other 3 numbers. Using the number 5 to describe the data set has the same issue. Using either of these numbers comes with some risk. That is, they do not perfectly describe the data set.

Maybe we should use 4.5, which is the average of the data. That seems like a good idea, as we have seen the concept of average in other math classes and we use it all the time in everyday life. Of course in this case, if we use 4.5 to describe (or summarize) the data, then none of the data values is exactly described by our number. Would this be considered even more risky than choosing 4 or 5? To answer this, we need to decide how to quantify risk.

A reasonable way to quantify the risk of using a number to describe a data set would be to see how it differs from each number in the data set and add those differences. That is, see how much total loss of information there is in using a specific number to describe the data set.

Suppose that we decide to measure loss as the square of the distance between a data value and our description (summary) number. The risk can then be calculated as the sum of the losses. If we consider the number c to describe the data set above consisting of $n = 6$ values, then our risk would be $\sum_{i=1}^{6} (x_i - c)^2$, where x_i represents the i^{th} data point.

Example 1.3.1

Using the above definition of loss, calculate the risk for the summary numbers 3, 4, 5, and 4.5.

We see that, for instance, using $c = 4$ is less risky than using $c = 3$ to describe the data. A natural question to consider is what value of c results in the least risk? Since there are an infinite number of possibilities, checking them all is out of the question. Instead, we rely on calculus.

Think of the risk as a function of the summary number c, i.e., $R(c) = \sum_{i=1}^{6} (x_i - c)^2$. To find the value of c that minimizes risk, we differentiate the risk function and set it equal to 0.

We see that using the summary value $c = 4.5$ to describe the data will result in the least risk. We can generalize this to determine the value that describes an arbitrary data set of size n with minimal risk.

Example 1.3.2

Consider a data set $x_1, x_2, \ldots, x_n$. Using the same methods as above, determine the summary value c that results in minimum risk (using the same definition of loss as above).

We see that using the average of the data set to summarize it results in the minimum risk. Note that using the average in this way is not just what we're accustomed to doing, but it actually has great mathematical significance.

Definition 1.3.1

We define the **mean** of a data set $x_1, x_2, \ldots, x_n$ to be $\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$.

Note that there are other definitions of loss that could be considered, leading to different risk functions. For instance, we could consider the absolute value function as a measure of loss, leading to risk function

$$R(c) = \sum_{i=1}^{n} |x_i - c|.$$

Example 1.3.3

Determine the value that minimizes the risk if our measure of loss is the absolute value function and our data set is $4, 4, 5, 7, 8, 8, 9$.

Example 1.3.4

Determine the value that minimizes the risk if our measure of loss is the absolute value function and our data set is $x_1, x_2, \ldots, x_n$.

We have used two different loss functions so far, $L(x) = x^2$ and $L(x) = |x|$. We will not see any more loss functions, and it is not necessary to have a full understanding of loss functions to appreciate the last few examples. It is, however, worth a short discussion of loss functions to enhance the overall mathematical knowledge of the reader.

Definition 1.3.2

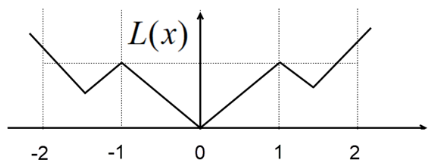
A **loss function** $L(x)$ is a function that satisfies the following:

- $L(0) = 0$ (i.e., no loss at 0)
- $L(x) \geq 0$ (i.e., the loss can't be negative)
- If $|a| < |b|$, then $L(a) \leq L(b)$ (i.e., loss cannot decrease as you move further away from 0)

Definition 1.3.3

A loss function is called **convex** if it has the additional property that for any two values a, b, we have $L(a + b) \leq L(a) + L(b)$.

Both of the loss functions that we considered above are convex loss functions. Below is an example of a function that is not a loss function (since, for instance, $L(1.5) < L(1)$).



The study of loss functions is not where we were headed, so we exit the topic now.

1.4 The Standard Deviation of a Data Set

Suppose that we have a loss function that satisfies a property stronger than the third property listed in the definition above: if $|a| < |b|$, then $L(a) < L(b)$. In other words, the loss function is increasing as we move away from 0, rather than simply non-decreasing. This is a very nice modification to be sure. This means that as the data set spreads out, the total loss increases. In statistics, one our favorite loss functions is the quadratic loss function that was used to arrive at the mean of a data set (as seen in the previous section).

Example 1.4.1

Suppose that we have a data set $x_1, x_2, \ldots, x_n$, and we calculate the mean $\overline{x}$ (recall that this is our risk minimizer using the quadratic loss function).

Suppose now that we add n data points, each with value equal to $\overline{x}$. Our new data set would have $2n$ pieces of data. The data set would be $x_1, x_2, \ldots, x_n, \overline{x}, \overline{x}, \ldots, \overline{x}$.

Since all of our new points are at the mean of the original data set, the new data set has the same mean.

When calculating the sum of losses using $\overline{x}$ to describe the data, it should be clear that the additional loss terms are all equal to 0. That is,

$$\sum_{i=1}^{2n} (x_i - \overline{x})^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 + \sum_{i=n+1}^{2n} (\overline{x} - \overline{x})^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

So the second data set has the same sum of losses as the first data set, but has twice as much data. In comparing the two data sets, the second data set seems more pleasant than the first, since it has a lot more data near $\overline{x}$. This suggests that we might be better off using the average loss as a measure of risk rather than the total loss. This example motivates the following definition.

Definition 1.4.1

The **variance** of a data set is denoted by s^2 and calculated using the following formula:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Notice that we have divided our total loss by $n-1$ instead of n. The reason for this will not be clear for quite some time. Suffice it to say, it can be approximately viewed as the average loss.

Note that the variance is measured in square units, which are not clearly interpretable. Instead, we often use another measure of risk that uses the same units as the original data.

Definition 1.4.2

The **standard deviation** of a data set is denoted by s and defined as the square root of the variance:

$$s = \sqrt{s^2}.$$

It will turn out that both the mean and standard deviation of a data set are very important in statistical analysis. We will see these quantities and terms often in this text.