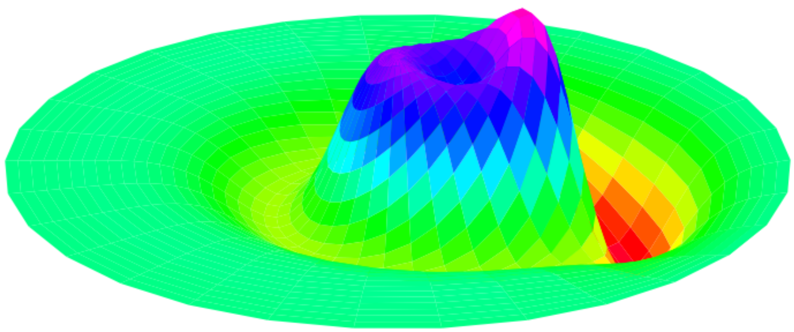
Machine Learning for Physicists



Lance J. Nelson

Department of Physics

Machine Learning for Physicists

Lance J. Nelson

Department of Physics

Brigham Young University–Idaho

Last Revised: April 27, 2020

Preface

This is a lab notebook intended to give you experience solving ordinary and partial differential equations numerically. The objectives of this course are

- to help you learn how to solve a differential equation numerically; in situations when a paper-and-pencil solution is impossible or impractical.
- to help you gain greater skills programming a computer and using loops, logic, functions, and classes.
- that your ability to produce a high-quality, professional scientific document will increase.

Text with a bold P designation (**P1.1** for example) indicate tasks that will be done together in class. Text with a bold H designation are homework problems and should be completed out of class. (working in groups is encouraged.)

Python is the programming language that we will be using You can obtain a free copy of Python here. Any computer code that you create should be uploaded to the Google Drive folder provided.

There is a companion book to this one entitled, "Introduction to Python". It is intended to help you learn to use Python to do the tasks contained herein.

Contents

Part I

Simple Probability

Chapter 1

Introduction to Probability

Simple Probability

For most people, the idea of simple probability is pretty intuitive. The probability of rolling a 5 on a 6-sided die is $\frac{1}{6}$ because of the 6 possible outcomes, only one of them produces a 5. The probability of drawing a red card in a standard, 52-card deck is $\frac{13}{52}$ because of the 52 total cards, 13 of them are red. Generalizing this idea yields the equation:

$$P(\text{event A}) = \frac{\text{\# of ways A can happen}}{\text{Total \# of things that could happen}} \tag{1.1}$$

This equation is the most general version for calculating probability and is a good one to fall back on when you get stuck. However, sometimes it's hard, or impossible, to count the number of possible outcomes. For example, consider rolling a fair, 6-sided dice, three times. What is the probability that the sum of these three rolls is less than 10? To answer the question using equation (1.1) you could write down all possible combinations of three dice rolls (there are 216 of them) and then identify how many of them summed to a number less than 10. Probably don't want to do that and if you feel tempted to tackle the task, then increase the number of rolls by 1 or 2 and reconsider. It isn't hard to get a situation where brute-force enumeration and counting is not a great idea.

What follows are some tricks and rules that you can use to calculate probability when brute-force counting is intractable.

Probability of multiple events

For example, a single dice roll has 6 possible outcomes, two dice rolls has 36 possible outcomes (illustrated in figure 1.1) and three dice rolls has 216 possible outcomes. Clearly you don't want to be enumerating these by hand. One scenario where outcome counting becomes challenging is when asking about the probability of multiple events occuring. This is because you have to consider whether the outcome of one event will affect the outcome of the next. For example, suppose you roll a 6-sided die twice and you want to know the probability that both rolls come up 5. Having a 5 come up on the first roll doesn't affect the probability that a 5 comes up on the second. These two events are independent of one another. The outcome of one event does not affect the outcome of the other.

In contrast, consider randomly selecting marbles out of a bag which initially contains 3 red marbles and 8 green marbles. If you reach into the bag and grab two marbles, what is the probability that they will both be red? Notice that

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

Figure 1.1 All 36 outcomes from rolling two fair dice. Out of 36 possible outcomes, only one yields a 5 on both rolls. Hence the probability of rolling two 5's is $\frac{1}{36}$.

selecting a red marble on the first reach into the bag modifies the number or red marbles available for selection on the second reach into the bag. Hence the two probabilities are dependent on each other. This is an example of dependent events.

Intersection of two events (event A and event B)

We'll start with the case where two events, A and B, both occur. (Also called the intersection of two events. See figure 1.2) The math for this sort of situation is:

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

but let's see some examples so it makes more sense.

Independent Events

Consider rolling a fair 6-sided dice. What is the probability that 5 comes up on the first two rolls? Your intuition is probably telling you that rolling two 5's is less likely than rolling just one, so you expect the probability to be less than $\frac{1}{6}$. In figure 1.1 you will see all 36 possible outcomes when rolling a fair dice two times. Of those 36 possible outcomes how many of occurrences of two 5's being rolled were there? Just one. Hence, the probability is $\frac{1}{36}$. You may have noticed that $\frac{1}{36} = \frac{1}{6} \times \frac{1}{6}$. In other words, the probability of event A happening **and** event B happening is just the product of the two individual probabilities:

$$P(A \text{ and } B) = P(A) \times P(B)$$

sometimes $P(A \text{ and } B)$ is written using the symbol for intersection ($\cap$):

$$P(A \cap B) = P(A) \times P(B) \tag{1.2}$$

Dependent Events

When the outcome of one event is affected by the outcome of a previous event, we say they events are dependent. For example, consider a bag containing 3 red marbles and 8 green marbles. If I reach into the bag and select two marbles, what is the probability that they are both red. The probability that my first selection is red is:

$$P(A) = \frac{3}{11}$$

because 3 out of the 11 choices are red. What about the second selection? After the first selection, I am left with 10 marbles, 2 of them being red. Hence, the probability that I draw a red marble on my second selection is:

$$P(B) = \frac{2}{10}$$

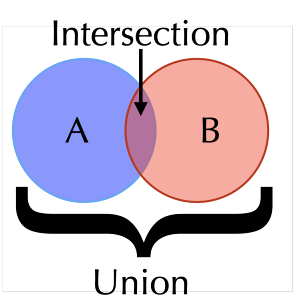


Figure 1.2 If the blue circle represents the probability of A occuring and the red circle represents the probability of B occuring, then the overlap of these two circles is the probability of both events occuring. By sliding one cirle left or right we can adjust the size of the overlap (aka $P(A|B)$)

The probability that my first selection will be a red marble (event A) **and** my second selection will also be a red marble (event B) is then

$$\begin{aligned} P(A \cap B) &= P(A) \times P(B|A) \\ &= \frac{3}{11}\frac{2}{10} \\ &= \frac{6}{110} \\ &= 0.055 \\ &= 5.5\% \end{aligned}$$

where $P(A|B)$ is called "conditional probability" and means the probability of B happening **given that** A already happened.

Notice that this is really the same thing we did for independent events (namely, multiply the two individual probabilities together) but we just had to be explicit state that when calculating the probability of event B we had to assume the event A had already happened.

If we want the probility of more than two events happening, we can continue to multiply conditional probabilities for each event, making sure to condition it on a positive result for all previous events.

Whether the events are dependent or independent doesn't change the general equation for multiple events, which is given by equation (1.2). However, to account for the possibility that event B's probability depends on event A, we'll write equation (1.2) as

$$P(A \cap B) = P(A) \times P(B|A) \tag{1.3}$$

where $P(A|B)$ is called "conditional probability" and means the probability of B happening **given that** A already happened.

We can do a little algebra on equation (1.3) to get:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{1.4}$$

Furthermore, if A and B are independent events, then $P(A \cap B) = P(B \cap A) = P(B) \times P(A|B)$. Hence, the above equation becomes:

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} && (1.5) \\ &= \boxed{\frac{P(A|B)P(B)}{P(A)}} && (1.6) \end{aligned}$$

This is known as Bayes' Theorem or Bayes' Rule and is of profound importance going forward.

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

Figure 1.3 All 36 outcomes from rolling two fair dice. <first roll> - <second roll>

Union of two events (event A or event B)

Things change a little if you are asking about the probability that either event A happens, or event B happens. Let's handle independent events first, then we'll do dependent events.

Independent Events

Let's take the dice-rolling example that we've been working with and ask ourselves what the probability is that either the first roll is a 5 or the second roll is a 5. Figure 1.3 will help us answer the question and then we'll try to generalize the result. The figure shows all possible results when rolling two fair die. Highlighted are the cases where one of the rolls was a 5. As you can see, there are 11 ways that one of the rolls can be a 5. Hence, the probability is $\frac{11}{36}$. That's a great way to do a really simple problem, but what if I roll the dice 10 times and ask what the probability is that one of them is a 5. Not so easy anymore right. Let's try to generalize this result. The general equation for or events is:

$$P(\text{A or B}) = P(A) + P(B) - P(A,B) \tag{1.7}$$

where $P(A)$ is the probability that event A occurs, $P(B)$ is the probability that event B occurs, and $P(A,B)$ is the probability that A and B both occur. Let's see what this works out to for this problem:

$$\begin{aligned} P(\text{A or B}) &= \frac{1}{6} + \frac{1}{6} - \frac{1}{36} && (1.8) \\ &= \frac{6}{36} + \frac{6}{36} - \frac{1}{36} && (1.9) \\ &= \frac{11}{36} && (1.10) \\ & && (1.11) \end{aligned}$$

which agrees with our result from the table.

Dependent Events

If event A and B are dependent, then we need to modify equation (1.7) like this:

$$P(\text{A or B}) = P(A) + P(B \cap \text{not A}) - P(A \cap B) \tag{1.12}$$

A tree diagram (see figure 1.4) will help to understand how this works. On the first selection, the probability of drawing a green marble is $P(A) = \frac{3}{5}$. $P(B|\text{not A})$ is the probability that a green marble is draw on the second selection assuming that a green marble was not drawn on the first. From the tree diagram we can see that this is equal to: We can use equation (1.3) to find $P(B \cap \text{not A})$:

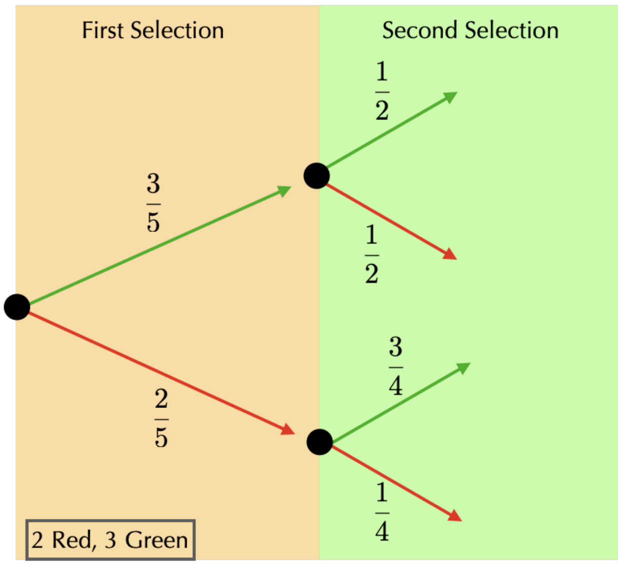


Figure 1.4 Testing

$$\begin{aligned}
P(B \cap \text{not A}) &= P(B|\text{not A}) \times P(A) \\
&= \frac{3}{4}\frac{2}{5}
\end{aligned}$$

$$P(A \cap B) = \frac{3}{5}\frac{1}{2}$$

Putting it all together we get:

$$\begin{aligned}
P(\text{A or B}) &= P(A) + P(B \cap \text{not A}) - P(A \cap B) \\
&= \frac{3}{5} + \frac{3}{4}\frac{2}{5} - \frac{3}{5}\frac{1}{2} \\
&= \frac{3}{5} + \frac{6}{20} - \frac{3}{10} \\
&= \frac{12}{20} + \frac{6}{20} - \frac{6}{20} \\
&= \frac{12}{20} \\
&= \frac{3}{5}
\end{aligned} \tag{1.13}$$

Another way to see this result is to enumerate all possible ways to draw these marbles from the basket: (There are only ten possible ways to do this.)

r, r, g, g, g
r, g, r, g, g

r, g, g, r, g
r, g, g, g, r
g, r, r, g, g
g, r, g, r, g
g, r, g, g, r
g, g, r, r, g
g, g, r, g, r
g, g, g, r, r

And notice that there are 6 ways where one (and only one) of the first two selections is green. Hence, the probability is $\frac{6}{10} = \frac{3}{5}$

Part II

Bayesian Statistics

Part III

Regression

Part IV

Optimization Algorithms

Commented out by Michael Ware. Code below inserts index

Index