1 Introduction

Why do we care about math?

If you ask this question to students of political science, they may give your very different opinions. To me, I think having some math background certainly helps you to help you along the way you read advanced statistic textbook or method paper. You can certainly not understand how different methods or formulas work but you can still pass the quant courses. However, I think a little bit of math background can help you build some confidence in learning new quant-related materials a little bit deeper. I'll give you my anecdotes to illustrate my point.

- quant-related method class in Berlin and econ class: don't set yourself limits because you are not good at math.
- classes beyond our department. In statistics, you are required to solve problems by hand. Before you start programming, in your mind, you should know how to do it manually. If you want to take some courses beyond our course load, like this screen shot from Coursera, which is an introduction to Bayesian analysis by University of California, you'll definitely need some math background to understand the basic.

My point is if you are not good at maths, it is totally okay. You really don't need maths to survive the first year. If you want to go a little bit deeper in quantitative methods, you will definitely need some basic calculus, such as limits, derivatives and integrals, some linear algebra, and some basic problem-solving skills. You may want to spend some time to have a good basic understanding of those. Taking some courses online or 400-level courses outside of this department helps a lot. Some of you may want to focus on political methodology in which case you may end up employing or more advanced methods in your research, or inventing new methods for other political scientists to use in the future. So don't set limits to yourself. And don't be intimidated by maths.

Why do we care about learning integral?

What is the learning objective in today's math camp?

After this morning's session, I hope you have an intuition about what integral stands for and link integral to a bigger picture context. Some of the concepts we talk about today will show up multiple times in the rest of the math camp and certainly in your quant class. If you ever have a chance to encounter integral in your future study, which I'm sure you will in some textbooks or papers, I hope you are confident that you have a little bit background in maths and do not afraid about reading and understanding those equations line-by-line.

How do you participate in the morning sessions?

Bring out your pen and pencils. Make some notes to yourself during the lectures. The slides are available to you after the session, but based on my own experiences, jotting down your ideas/key concepts during the lectures are important to help grasp and remember the basic idea. More importantly, when you write down them, you become familiarize yourself some mathematical symbols and the process of solving those problems you may not have seen before at all.

After illustrating some concepts, I will give you two or three problems at the end of each section to try out. I hope that you can try your best to solve them independently. Put some time and thoughts into it. Remember for some of you, this may be your first time seeing these concepts, so give yourself a little bit more time to reflect on those problems. Practice is very important in learning maths, which can help you to overcome an illusion of competence, meaning that you think you understand it, but in fact you don't, until you can solve the problem yourself. I'll also walk you through the solutions on the board, so I'll also encourage you to make some notes as well.

In short, you are going to spend three hours with me this morning. Please feel free to ask any questions or seek for clarification. In the first two hours, we will have three sections of lecturing followed by some practice problems printed in the handout for you to try out. We'll take 5-minute breaks at the end of first and second sections, and then 15 minute breaks after the third section. After the long break, you'll solve some homework problems yourself. I am here to offer you some help and address some questions to y'all if you need it.

2 Section 4: Integrals and Distributions

What is distribution, and why should we care?

Before introducing distributions, we want to know why we should bother studying distributions. The headline is: political scientist cares about generalization, and generalization involves the distributions of variables. Theories about political behaviour and institutions are generalization. It means they do not apply to a single specific individual; instead, it applies to most members of a class or populations at most moments in times. The relationship between a causal variable X and caused variable Y is not expected to hold in every case. Rather, theories are probabilistic. In fact, nearly all social science hypotheses are phrased as "If A, then B is more likely to occur." Social science hypotheses are rarely deterministic. The occurrence of A only makes B more likely. not inevitable.

Once we shift our attention away from specific individual members of a class or population toward most members of a class or population, we have already begun to think about distributions. And the mathematical study of distributions provides some powerful tools to aid one in both developing theory (which by definition involves generalization) and testing the hypotheses implied by theories.

Formally speaking, a distribution defines the set of values that a random variable may take. Random variables are those that have their value determined in part by chance. The value they take in any given circumstance can be described probablistically. The set of values the variable may take is the distribution of the random variable. The function that defines the probability that each value occurs is known as the probability mass function or the probability density function, depending on whether the distribution of values is discrete or continuous.

What is random variable?

In other words, we expect the value of each variable in our hypotheses (and many theories) to be a draw from some associated distribution. A simple example of this would be the roll of a fair die. The random variable corresponding to the die's value would have an equal probability (one-sixth) of having each of the integer values from 1 to 6. This is called a uniform distribution. However, more complicated distributions are allowed, particularly when we're discussing dependent random variables. For example, we could say that one's decision to vote is a random variable Y that can take the values of 1 (for yes) and 0 (for no). The probability that it takes the value 1 will depend on many other factors, such as past voting behavior, education, income, etc. In other words, the probability distribution of values for Y is a conditional probability distribution: one for which the likely outcomes vary as one or more other variables vary in value (more on this below). Increasing the value of an independent variable such as education or income might shift the distribution of 1s and 0s toward more 1s, implying that a more educated or higher-earning person will be more likely to vote, and more likely to vote more often, given multiple opportunities to do so.

Examples of discrete and continuous variables

Examples of discrete variables include number of heads when flipping three coins, numbers of civil wars in a year, etc. The values as such can be obtained by counting.

Examples of continuous variable are distance traveled between classes, time a country takes to recover from economic recession, how long a coalition government survives, countrys' GDP per capital etc. Continuous variables have an infinite number of possible values. The difference between the two types of variables can be thought of whether the variables are actually countable or not. If you can list each value the random variable can take and assign a probability to each, then you have a discrete random variable, represented by a discrete distribution. If you can't, you may have a continuous random variable, represented by a continuous distribution.

The probability that a random variable has any value is given by its probability distribution. Discrete random variables have discrete associated probability distributions; continuous random variables have continuous associated probability distributions. We will develop the concept of the probability function for each type of variable separately.

What is probability mass function (PMF)?

PMF is a function that specifies the probabilities of drawing discrete values from the sample. The probability that a specific case drawn at random from a sample has a specific value, i, is the relative frequency of the value i.

The sum of the empirical probabilities that each value is drawn at random from the sample equals to 1. Let's consider the simple case of tossing a coin. There are two outcomes for each time of tossing a coin. We can write the following PMF:

$$p(y_i = 0) = \pi$$

,

$$p(y_i = 1) = 1 - \pi$$

.

Here, π is the parameter of this PMF. Because probabilities range between 0 and 1, $\pi \in [0, 1]$.

If we know that the coin is unbiased, then the value of π equals 0.5. If the coin is biased and turns up heads twice as often as tails, then π is 0:67 and 1 - π is 0:33. What PMF tells us is that when we change the value of the parameter, the distribution of outcomes change.

Some of the theoretical PDF are Bernoulli, Binomial, Multinomial, Event count distribution, which you might be able to learn the basic in the third or fourth day of math camp.

What is probability density function (PDF)?

PDF of a continuous variable is related to the relative frequency distribution of that variable.

The density at any one point of X is vanishingly small, and we care not just about a single point but also values of the random variable in a neighborhood of points. Typically, then, we will deal with the cumulative or total probability in an range of values or up to a specific point.

A PDF differs from a PMF in that it does not describe the chance that any particular value of the random variable is drawn at random from the distribution. Rather, it describes the relative likelihood of drawing any specific value, and the exact probability of drawing a value within some range. This is why it is called a density function rather than a mass function: it describes the density of the probability within some range of values that the random variable may take, rather than the explicit mass of probability at a particular value.

An example on the graph can help us to consider why the difference between PDF and PMF exists. Let's consider x $\in$ $[0,1] forallvaluesxthatarandomvariableXmaytake.Sinceallrand$
$[0,1] is1.Nowconsidertherangex \in [0,0.5].Thechanceofbeinghereislessthan1, sowereducethisprobabilityf$
$[0,0.25].Again, wearelikelytoshrinktheprobabilityofbeinginthatregion.Ifwekeepdoingsooveranoveragai$
$Xisdefinedover[0;0:001];[0;0:00001], andsoon, forever.Theprobabilityatapointwouldbetheprobabilityat$
$defined.Sowedon'tdefineit, anddon'tingeneralspeakaboutprobabilitiesatspecificpoints, eventhoughtheP$
$zerovaluesatthesepoints.$

Since we usually deal with a range of values, rather than a specific point, this involves the integration of all values of the density function between two points, say a and b. Formally,

$$Pr(X \in [a, b]) = \int_a^b f(x)dx$$

. The probability of being in some interval [a, b] is the integral of the PDF over that region. If the probability of being in some interval [a, b] is the integral of the PDF over that region, then we are saying that the probability is computed by summing lots of f(x)dx. If we replace the dx with a delta x, we're back to the area under a rectangle, or, in other words, the area under the relative frequency histogram. A PDF in this sense is like a smoothed-out histogram.

Those of you who have had a course in statistics have already studied a number of these distributions, but even those of you who have never studied statistics have probably heard of the bell curve. The bell, normal, or Gaussian curve is a probability distribution that describes variables that take continuous values.

What is cumulative distribution function (CDF) for both discrete and continuous variables?

We mentioned that distribution function is critical for generating theory and conducting statistical hypothesis test. When conducting hypothesis tests, it is often useful to determine the probability that a value drawn at random from a sample is above or below a specific value. The cumulative distribution function (CDF) describes the function that covers a range of values below a specific value and is defined for both discrete and continuous random variables.

Intuitively, if we pick a number, and we want to know the probability that a random draw from a population that produces a value less than that specified number, we sum the individual probabilities of each value below the specified value.

We can write the CDF for discrete variables as

$$Pr(Y \leq y) = \sum_{i \leq y} p(i)$$

.

We sum the probabilities of each value for all values less than or equal to y.

For continuous variable, the idea is similar. Since we cannot write the CDF for a continuous variable as the sum of the probabilities of each discrete value below the specified value, we have to write it as the sum of all the value ranges below the specified value. We replace the sum with an integral to get the equation for the CDF:

$$Pr(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)dt$$

.

Sometimes you will see the notation f(x) for a probability distribution function (PDF or PMF) and F(x) for a CDF; using that notation makes clearer the connection between the two functions.

Also, note that

$$Pr(Y \leq y) + Pr(Y > y) = 1$$

, which means that

$$Pr(Y > y) = 1 - Pr(Y \leq y)$$

. It means the the probability that a random draw exceeds some quantity is equal to one minus the probability that it does not exceed that quantity. If y is the highest value that Y can take, then

$$Pr(Y \leq y) = 1$$

, since in this case we are adding the probability of all outcomes in the sample space. So all CDFs plateau at one.

We often want to know whether a given value is likely to have been drawn from a specific portion of a distribution (i.e., the chance that we would observe an outcome greater or less than some specific value, in this case, 1.96 in the normal distribution). The CDF helps us answer that question.

Example of normal distribution (back to slide).