Math for Data Analytics I

QBS 103: Foundations of Data Science

August 20, 2024

Lesson Objectives

At the end of this lecture you should be able to:

1. Read and interpret mathematical notation
2. Write and plot mathematical functions in R
3. Write logarithmic and trigonometric functions in R

Resources

LaTeX Cheat Sheet for Set Notation: https://www.overleaf.com/learn/latex/List_of_Greek_letters_and_math_symbols

Good Explanation of Mathematical Notation: https://www.youtube.com/watch?v=kaJuXx6uYR0

LaTeX in R Plots Using *latex2exp*: https://cran.r-project.org/web/packages/latex2exp/vignettes/using-latex2exp.html

Mathematical Notation

Mathematical equations and proofs often rely on set notation to efficiently state information. This is a quick overview (not a comprehensive list) of some mathematical notation you might encounter as you move forward in statistics.

For example, we can define

$$A = \{x, y, z\}$$

which means that A is a set that contains elements x, y, and z. We can refer to elements of the set using $\in$, such that we can say

$$x \in A$$

Similarly, we can use the element symbol dashed through like this $\notin$ to indicate something isn't an element of that set, such that

$$s \notin A$$

One special set that you might see is $\mathbb{R}$ which indicates the set of all real numbers.

We can also define a set as

$$\{x | x > 0\}$$

rather than listing out individual values which would indicate the set of x in which x is greater than 0.

We also use $\forall$ to indicate "for all" such that we can say

$$\forall\ x \in A, x > 0$$

Reading across this just says, "for all x elements of A, x is greater than zero" or, more simply, all the elements of A are greater than zero.

We can also use the $\exists$ symbol to indicate "exists" such that we can say $\forall \in A \ \exists$ such that $x > 0$ Reading across this means that there exists an element in A that is greater than zero. Similarly to $\notin$ we also have $\nexists$ which indicates "there does not exist".

Simple Mathematical Functions in R

In this course already, we have been using functions, simulating data, making plots, and even using some linear algebra - all of which are foundational to calculus!

For instance, supposed we are interested in the following mathematical function:

$$f(x) = mx + b$$

To create this function in R, we represent this function as:

And then if we were to solve this function for $m = 1$, $x = 2$, and $b = 3$ we could:

$$f(2) = 1(2) + 3 = 5$$

or, use our function:

It is then also super easy to evaluate this function over a range of x values. For instance, we may want to know the solution of $f(x)$ over the domain $\{x \in \mathbb{N} | 0 \leq x \leq 1000\}$. You could solve this by hand, or you could generate values of x over your domain and solve like so:

And you can see that we then have each solution for every value of x! Now we can easily plot these values.

Representing functions this way becomes increasingly useful with increasingly complex functions. For instance, suppose we have the following function:

$$f(x) = x\sin\left(\frac{1}{x}\right)$$

and we want to solve it over the domain $\{x \in \mathbb{R} | -1 \leq x \leq 1\}$. You could solve this by hand over many possibilities of x, or:

Sometimes, you might want to use LaTeX within your plots. To do so, we'll use the *latex2exp* library. Additional information and examples are linked at the start of the lecture under resources.

```
# Load package
#install.packages('latex2exp')
library(latex2exp)

# Increase margin size for function
# Syntax here is: par(mar = c(bottom, left, top, right)
par(mar = c(5, 5, 4, 2) + 0.1)
```

Exponents amd Logarithms in R

In R, *log(x)* calculates the natural log of x (i.e. base-e). If you type in ?log you'll see that you have the ability to customize the base within the log function as well.

Given that log_2 and log_10 are both commonly used logarithms, there are separate functions for these as well.

Given that log_2 and log_10 are both commonly used logarithms, there are separate functions for these as well. We've already gotten very used to using the ˆ symbol for creating exponents. When we want to raise something to the power of e we use the *exp()* function in R.

You'll recall that there are special rules for using math with logarithms. You'll want to remember these before you dive back into statistics and calculus. We can also confirm these in R pretty easily. They include:

$$log_{10}(ab) = log_{10}(a) + log_{10}(b)$$

$$ln(\frac{a}{b}) = ln(a) - ln(b)$$

$$ln(1/x) = -ln(x)$$

$$log_2(x^y) = ylog_2(x)$$

$$ln(e) = 1$$

$$ln(1) = 0$$

Similarly, there are a variety of exponent laws worth reviewing:

$$e^a e^b = e^{a+b}$$

$$e^{ln(xy)} = xy$$

$$2^{log_2(xy)} = xy$$

$$\frac{e^a}{e^b} = e^{a-b}$$

$$(e^a)^b = e^{ab}$$

$$(xy)^a = x^a y^a$$

$$x^{-a} = \frac{1}{x^a}$$

$$x^{\frac{a}{b}} = \sqrt[b]{x^a}$$

Trigonometry

You'll often need to use trigonometry in Statistics and Calculus, so naturally our trig functions are built into R as well. Above, we used the *sin()* function in a plot but there are also functions for *cos()* and *tan()* as well as arcsin, arcos, and arctan, all shown below.

You'll reall that there are also a variety of trig identities you can use to make solving functions with trigonometry easier! A complete cheat sheet can be found here: https://sciencenotes.org/trig-identities-study-sheet/ but below are some useful ones you'll frequently use:

$$\tan\theta = \frac{\sin\theta}{\cos\theta}$$

$$\sin^2\theta + \cos^2\theta = 1$$

$$\sin(2\theta) = 2\sin\theta\cos\theta$$

$$\cos(2\theta) = \cos^2\theta - sin^2\theta = 2\cos^2\theta - 1$$

$$sin(-\theta) = -\sin(\theta)$$

You may be wondering why we are reviewing so much of this in the context of R, and one very great reason is because R will be a useful tool in checking your work when you are required to use these laws to manually simplify functions. Consider for instance:

$$f(\theta) = \frac{1+\sin\theta}{\cos\theta} - \frac{\cos\gamma}{1-\sin\gamma}$$

And you simplify to:

$$f(\theta) = \frac{2\sin\theta - 2\sin\gamma}{\sin(\theta-\gamma) + \cos\theta - cos\gamma}$$

We can test this in R: