Mathematical Notation

Josef Fruehwald

8/20/22

Variables

The most common variables we're going to be seeing:

- x, y
  - Stand-ins for numbers, usually a list, or vector, of numbers
  - y is often some kind of "outcome" variable
  - x is often some kind of input, or predictor variables
  - e.g. "We want to predict how many ice cream cones, y, we'll sell if it's a specific temperature, x.
- X, Y
  - Stand ins for categorical variables
- w
  - A special variable for a word
- c
  - A "count" of something
- p
  - A probability
- N
  - Usually, the total number of something

- n
    - A contextual number. E.g. "In $n+1$ days (that is, tomorrow)…"
- n, m
    - When n and m are used together, it's often to describe the number of rows and columns of a matrix.
- A, B
    - These are almost always used for matrices
    - Capital roman letters are often a clue we're looking at a matrix (but not always)
- λ
    - "lambda"
    - Often used for an arbitrary value you multiply things by.
- k
    - Often used for an arbitrary value you add things to.
- δ
    - "delta"
    - Often used to describe a **d**ifference of some kind
- α, β, γ
    - "alpha", "beta", "gamma"
    - Used for model parameters
- θ
    - "theta"
    - Also used as a model parameter.
    - Or, to describe an angle (in radians)

"Decorated" variables

- $\hat{y}$
  - "y hat"
  - A predicted value for y
  - e.g. "I predicted $\hat{y}$ ice cream cones to be sold, but they actually sold y.
- $\bar{y}$
  - "y bar"
  - The average value of y
- y^*
  - "y star"
  - A modified value of y

Indices

One Dimensional

When we have a variable that contains a list of values, each individual value will be described with an "index". For example, if we had a variable X that contained the names of the week.

$$X = (\text{Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday})$$

Then the *first* value in the list would be X_1, and we could say $X_1 = \text{Monday}$. You'd usually pronounce X_1 as "X sub 1".

Sometimes we want to be able to refer to a generic value in the list X and for that we'd use an index variable like X_i (pronounced "X sub i"). The most common letters used to indicate generic indices are i, j, k.

We can do math on the indices also. So, if the name of today is X_i, then the name of tomorrow is X_{i+1}. The name of yesterday is X_{i-1}. The day after tomorrow would be X_{i+2}.

We can also include a range of numbers in the indices. So, the names of the first three days of the week are $X_{1:3} = (\text{Monday, Tuesday, Wednesday})$. The names of yesterday, today, and tomorrow are $X_{i-1:i+1}$.

Two Dimensional

We could imagine a Month as being a matrix M made up of 4 weeks, with each week being 7 days.

$$M = \begin{bmatrix} \text{Monday} & \text{Tuesday} & \text{Wednesday} & \text{Thursday} & \text{Friday} & \text{Saturday} & \text{Sunday} \\ \text{Monday} & & & \text{...} & & & \text{Sunday} \\ \vdots & & & \ddots & & & \\ \text{Monday} & & & \text{...} & & & \text{Sunday} \end{bmatrix}$$

If we wanted to get the name of the third day during the first week, it would be $M_{1,3} =$ Wednesday. When giving numeric indices of a matrix, the rows (the parts going across) come first, and the columns (the parts going up and down) come second. So, we'd describe the matrix M as being a 4×7 "four by seven" matrix.

To refer to a day of the month generically, we'd use $M_{i,j}$ (pronounced "M sub i, j"). We can also use ranges in these indices. So, to refer to the second week of the month, we'd use $M_{2,1:7}$. Or, to refer to all of the Saturdays in the month, we'd use $M_{1:4,6}$.

And, we can also use math in these indices. So, we could refer to "a week from today" with $M_{i+1,j}$.

Matrix Transposition

The one special notation related to matrices is "transposition", which basically takes a matrix and flips it.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$A^\top = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

The little $^\top$ indicates that we're transposing the matrix. You'd pronounce $A^\top$ as "A transpose".

The transpose operation can apply to single lists too. This will eventually be important.

$$x = [1, 2, 3, 4]$$

$$x^\top = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Again, $x^\top$ is pronounced "x transpose."

Functions

Generic

- $f(), g()$
  - These are the most common "functions" you'll see. Usually in a whole formula like $y = f(x)$ (pronounced "y equals f of x").
  - It means "we stick the value x in and y comes out.
  - *What* $f()$ or $g()$ (or whatever name we give the function) is needs to be defined. They don't have a fixed meaning.

Specialized

- $P()$
  - The function $P()$ refers to the probability of whatever we put in.
  - $P(X_i)$ returns the probability of a specific X_i value.
  - We'd refer to the specific value the $P()$ returns with the variable p, usually with the same index. So $p_i = P(X_i)$.
- $C()$
  - The function $C()$ returns the count of whatever we put in.
  - In the matrix above, $C(\text{Monday}) = 4$
  - We'd refer to the specific count value of something with the variable c, usually with the same index. So $c_i = C(X_i)$.

Summation and Product

Summation

$\sum$

This operator indicates that we are adding together numbers in a list. Let's look at the table of the height of actors, in cm who played Spider-Man in *Spider-Man: No Way Home*.

Actor	Height (cm)
Tobey Maguire	172
Andrew Garfield	179
Tom Holland	169

We would say that there are $N = 3$ actors who played Spider-Man in the movie. And we could represent their heights in a variable y, and say

$$y = (172, 179, 169)$$

The height of the first actor would be y_1, which equals 172, and the way to refer to any given height on the list would be y_i.

To get the total height of the actors (like if one stood on the head of the other), we would have to sum it up, which we could represent like this:

$$h = y_1 + y_2 + y_3$$

Or, we could use summation notation

$$h = \sum_{i=1}^{N} y_i$$

The way to read this out loud is "h equals the sum of y sub i from i equals 1 to N". The $i = 1$ part underneath the $\sum$ means "start getting values out of y starting with 1". The N at the top means "keep adding 1 to i until $i = N$."

What the whole notation is going to do is pull out every value of y and add them together.

If you know how to do some coding, sometimes it's easier to understand the mathematical notation to see it in code.

```{python}
y = [172, 179, 169]
N = 3 # = len(y)
h = 0

for i in range(N):
  h = h + y[i]

print(h)
```

520

Product

$\prod$

The product operator, $\prod$, works a lot like the the summation operator, except instead of adding numbers together, it multiplies them. For example, let's say we're keeping track of the day-to-day changes in the number of visitors to a website.

Day	Percent Change	Multiplier
1	up 1%	1.01
2	no change	1.00
3	down 5%	0.95

We can get the total proportional change over these three days by multiplying the proportions together. Writing it out the long way, it's

$$N = 3$$

$$y = (1.01, 1, 0.95)$$

$$t = y_1 \cdot y_2 \cdot y_3$$

In product notation, though, it looks like this.

$$t = \prod_{i=1}^{N} y_i$$

Again, if you're more comfortable with programming code, it's equivalent to this.

```
```{python}
y = [1.01, 1, 0.95]
N = 3
t = 1

for i in range(N):
  t = t * y[i]

print(f"{t:.4}")
```
```

```
0.9595
```

Probability

There's not space in this tutorial for an *entire* intro to probability theort, so here I'll just provide enough context to describe how we notate key concepts.

Conditional Probability

The "conditional probability" is the probability of some value or event Y, holding constant some other value or event X . For example, maybe

$$Y = \text{I am teaching}$$

and

$$X_1 = \text{It is a Monday}$$

and

$$X_6 = \text{It is a Saturday}$$

The probability that I am teaching a class is is a lot lower on a Saturday than on a Monday. We can express these like so.

$$p_1 = P(Y|X_1)$$

a.k.a "p sub 1 equals the probability I am teaching, given that it is a Monday".

$$p_6 = P(Y|X_6)$$

a.k.a. "p sub 6 equals the probability I am teaching given that it is a Saturday"

and

$$p_6 < p_1$$

The key piece of notation in the expressions above is the | (the "pipe") inside of the $P()$ function.

Joint Probability

The joint probability is the probability of two values or events happening together. It's **not** the same as a conditional probability of one event given the other, but explaining why requires more time and space.

If we stick with the same events as above ("I am teaching" and "It is a Monday") the joint probability of "I am teaching *and* it is a Monday" would be notated as

$$q_1 = P(Y, X_1)$$

The comma inside the $P()$ function means "and."

- pdf