Maths Day One: Calculus

Princeton Sociology Methods Camp

Varun Satish[1]

[1] These slides are the collective effort of everyone who has contributed to the Princeton Sociology Methods Camp.

Outline

1. What is a derivative and why we are learning about them
2. Special rules for logs and exponents
3. Higher-order derivative (e.g. the second derivative)
4. Partial derivatives

What is a derivative?

CHAPTER I.

TO DELIVER YOU FROM THE PRELIMINARY TERRORS.

THE preliminary terror, which chokes off most fifth-form boys from even attempting to learn how to calculate, can be abolished once for all by simply stating what is the meaning—in common-sense terms—of the two principal symbols that are used in calculating.

These dreadful symbols are:

(1) d which merely means "a little bit of."

Thus dx means a little bit of x; or du means a little bit of u. Ordinary mathematicians think it more polite to say "an element of," instead of "a little bit of." Just as you please. But you will find that these little bits (or elements) may be considered to be indefinitely small.

(2) $\int$ which is merely a long S, and may be called (if you like) "the sum of."

Thus $\int dx$ means the sum of all the little bits of x; or $\int dt$ means the sum of all the little bits of t. Ordinary mathematicians call this symbol "the

C.M.E. A

integral of." Now any fool can see that if x is considered as made up of a lot of little bits, each of which is called dx, if you add them all up together you get the sum of all the dx's, (which is the same thing as the whole of x). The word "integral" simply means "the whole." If you think of the duration of time for one hour, you may (if you like) think of it as cut up into 3600 little bits called seconds. The whole of the 3600 little bits added up together make one hour.

When you see an expression that begins with this terrifying symbol, you will henceforth know that it is put there merely to give you instructions that you are now to perform the operation (if you can) of totalling up all the little bits that are indicated by the symbols that follow.

That's all.

What is a derivative?

- A function is an expression that takes an input, x, and turns it into an outcome, y.
- A derivative is the answer to the question: what happens to the output of the function when I change the input by "a little bit".

Motivation

Why are we learning about derivatives?

- You will come across them in SOC500. Derivatives are a key feature in much of statistical theory
- **Example 1**: a probability density function (p.d.f) can be expressed as the first derivative of the cumulative density function (c.d.f)
- **Example 2**: Optimizing functions (i.e. finding where they are either maximized or minimized)
  - How? Set the first derivative equal to zero!

How to take a derivative

There are 4 main rules you need to remember:

1. Power rule (for most simple functions)
2. Rules for logs and exponents
3. Chain rule (for composite functions)
4. Product rule (for products of functions)

Quick note about notation

- To express the derivative of $f(x)$ you can use either $\frac{d}{dx}f(x)$ or $f'(x)$–they both mean the same thing!

Power rule

- Suppose we have a function $f(x)$ which is of the form x^n (where n is any real number).

  For example:
  - x^2
  - $x^4 + 1$
  - $x + x^2 + x^3$

- Then, $\frac{d}{dx} f(x) = n * x^{n-1}$.

Derivatives of logs and exponents

- $\frac{d}{dx} ln(f(x)) = \frac{f'(x)}{f(x)}$
- $\frac{d}{dx} e^{f(x)} = f'(x) * e^{f(x)}$

Review: derivatives of logs and exponents

Some examples to practice on your worksheet

1. $\frac{d}{dx} ln(x^5)$
2. $\frac{d}{dx} ln(x^2 + 3)$
3. $\frac{d}{dx} e^{5x+2}$

Chain rule

- Suppose we have have what's called a *composite function*- e.g., a function nested inside another function
- **Notation**:
  - We have two functions, $f(x)$ and $g(x)$.
  - We are interested in $f(g(x))$ which is the result of what happens if we apply the outside outside function to the inside function
  - **Example:** $f(x) = x^2$ and $g(x) = 3x + 2$. What is $f(g(x))$?

Chain rule

Chain rule: $\frac{d}{dx} f(g(x)) = f'(g(x)) * g'(x)$

1. Take the derivative of the outside function. Don't touch the inside function (treat it like a constant).
2. Take the derivative of the inside function, ignoring the outside function
3. Multiply (1) and (2).

Product rule

Product rule: $\frac{d}{dx}f(x)g(x) = f'(x)g(x) + f(x)g'(x)$

1. Common notation is to all $f(x)$ and $g(x)$ u and v, respectively.
2. Then, the derivative is equal to $u'v + uv'$ where

Exercise two

Find the answer to the following:

1. $\frac{d}{dx}3x^4 + 2$
2. $\frac{d}{dx}\sqrt{(3x^4 + 2)}$
3. $\frac{d}{dx}x^2e^x$

On your worksheet, work through these examples and then we'll review as a group...

Higher-order derivatives (i.e. second, third derivatives)

Motivation

When do we use higher-order derivatives?

- **Example 1:** Optimizing functions. Taking the first derivative and setting it equal to 0 gives us either a minimum or maximum. The second derivative allows us to discern between the two.
- **Example 2:** Later down the line, higher-order derivatives can be used to construct useful approximations (i.e. Taylor's approximation) to complex functions that are important in understanding the properties of various statistical estimators.
- These motivations aren't important to remember for now, but we want to give you some preview of *why* we are learning this concept

Higher-order derivatives

- Not much more to learn in addition to the *first derivatives* we've been finding.
  - **Second derivative**: $f''(x) = \frac{d^2 f(x)}{dx^2} = \frac{d}{dx}(\frac{d}{dx} f(x))$
  - **Third derivative**: $f'''(x) = \frac{d^3 f(x)}{dx^3}$
  - Etc. until there remains nothing left in the function of which to take the derivative (that is, all constants)
- **Example**:
  - $f(x) = 5x^3 + 3x^2$
  - $\frac{df(x)}{dx} = 15x^2 + 6x$
  - $\frac{d^2 f(x)}{dx^2} = 30x + 6$
  - $\frac{d^3 f(x)}{dx^3} = 30$
  - Notice that these are just derivatives of derivatives!

Practicing higher-order derivatives

Find the following on your worksheet:

1. $\frac{d^2}{dx^2} 3x^4 + 2$
2. $\frac{d^2}{dx^2} \ln(x^5)$

Partial derivatives

Motivation

When do we use partial derivatives?

- **Example:** The population coefficient in the OLS model can be expressed as the solution to a partial derivative.
- Again, this isn't important to remember for now, but we want to give you some preview of *why* we are learning this concept

Partial derivatives

- **What is a partial derivative**? Suppose we have a function that take two or more variables as inputs. A partial derivative tells us how the function changes if we vary one of the variables by a little bit *holding the other one constant*.
  - *What we've been doing*: $f(x) = x^2 \rightarrow f'(x) = 2x$
  - *When we'd use partial derivative*: $f(x, y) = x^2 + 5y^4$
- Notation:
  - $\frac{\partial f(x,y)}{\partial x}$: partial derivative of $f(x, y)$ with respect to (w.r.t.) x
  - $\frac{\partial f(x,y)}{\partial y}$: partial derivative of $f(x, y)$ with respect to (w.r.t.) y
- How do we find?
  - $\frac{\partial f(x,y)}{\partial x}$: take derivative of function but treat y as a constant
  - $\frac{\partial f(x,y)}{\partial y}$: take derivative of function but treat x as a constant

Tool two: partial derivatives, example of mechanics

$$f(x, y) = x^2 y^5 + e^y + ln(x)$$

Find the following on your worksheet:

1. $\frac{\partial f(x,y)}{\partial x}$
2. $\frac{\partial f(x,y)}{\partial y}$
3. $\frac{\partial f_x(x,y)}{\partial y}$ (first x, then y)
4. $\frac{\partial f_x(x,y)}{\partial x}$

Summing up

1. We learned about what a derivative is and previewed why they will come up in SOC500
2. We learned some important rules for taking derivatives
   - The power rule
   - The chain rule
   - The product rule
   - Rules for logs and exponents
3. We learned that we can use what we've learned to take higher-order and partial deriatives

Bonus Content: A simple case of univariate optimization

Example one: health navigators

Suppose a community health center is interested in hiring health navigators to help enroll patients into the ACA. They want to maximize the number of patients they enroll, and are aware that the following relationships hold, letting h = health navigator.

- Productivity of each health navigator $= 50h^{\frac{2}{3}}$
- Cost that detracts from other efforts at ACA enrollment $= 2h$
- Putting them together, and letting A = patients enrolled in ACA:

$$A(h) = 50h^{\frac{2}{3}} - 2h$$

Example one: health navigators

To help us get a sense of the function, we can graph it by using three R commands (also in the code file for this lecture):

1. Create a user-defined function that codes $A(h)$ that takes as its input different values of health navigators:

```
acafunc <- function(h){
  50*(h^(2/3)) - 2*h
}
```

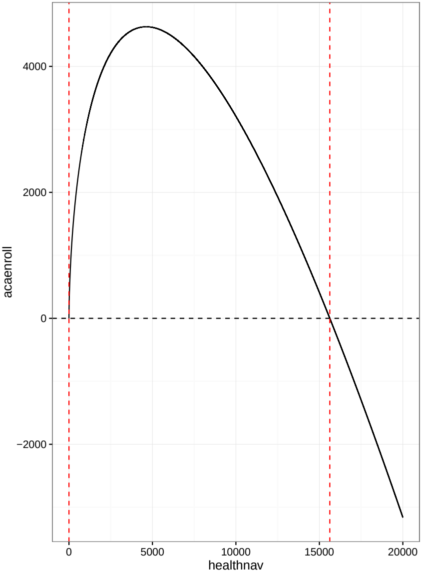
2. Create a vector of h (health navigators) at which to evaluate the function more efficiently using the sequence command seq

```
h <- seq(from = 0, to = 20000, by = 0.1)
```

3. Apply the function to that vector of health navigator values; store in a vector a that indicates the number of ACA enrollees associated with each health navigator value:

```
a <- sapply(h, acafunc)
```

Visualizing the relationship between health navigators and ACA enrollment



Example one: health navigators, motivation for optimization

We can see that...

- The curve flattens out a little before 5000 navigators
- That flat point appears to be a maximum
- How do we formalize the previous two observations?

Example one: health navigators, mechanics of optimization

1. **Define the domain of interest**: in this case, it is where $h \geq 0$ and $A(h) \geq 0$ (we don't want there to be fewer ACA enrollees as a result of the health navigators)
   - For this particular equation, $0 = 50h^{\frac{2}{3}} - 2h \implies h = 15625$
   - Domain of interest: $h = [0, 15625]$
2. **Find the critical point(s) of the function**: we find the critical point(s) by setting the derivative of the function equal to zero and solving for the number of health navigators:

Solve for critical point(s) on your worksheet

Example one: health navigators, mechanics of optimization

1. **Define the domain of interest**: in this case, it is where $h \geq 0$ and $A(h) \geq 0$ (we don't want there to be fewer ACA enrollees as a result of the health navigators)
   - For this particular equation, $0 = 50h^{\frac{2}{3}} - 2h \implies h = 15625$
   - Domain of interest: $h = [0, 15625]$
2. **Find the critical point(s) of the function**: we find the critical point(s) by setting the derivative of the function equal to zero and solving for the number of health navigators:

Solve for critical point(s) on your worksheet

$$A(h) = 50h^{\frac{2}{3}} - 2h$$

$$A'(h) = \frac{100}{3h^{\frac{1}{3}}} - 2 = 0$$

$$2 = \frac{100}{3h^{\frac{1}{3}}}$$

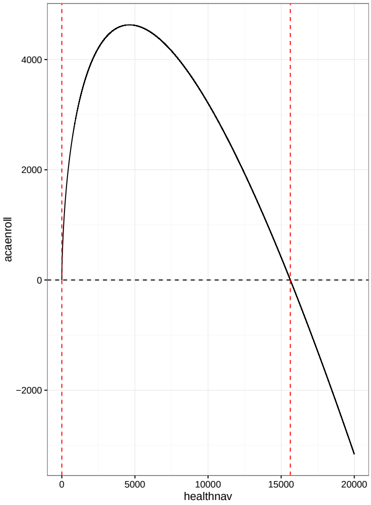
$$h^{\frac{1}{3}} = \frac{100}{6} = \frac{50}{3}$$

$$(h^{\frac{1}{3}})^3 = \frac{50^3}{3^3} = \frac{125000}{27} = 4629.23$$

Sidenote on terminology...

Critical points are also sometimes called stationary or stable points. For critical points, in addition to finding where $f'(x) = 0$, you should also find points at which the function is not differentiable (reviewed in summer assignment). The ACA function is differentiable at all points

Purpose of second derivative



- Now we've been able to formalize the intuition that the slope flattens a little before 4600 navigators, by solving for $\hat{h} \approx 4629$
- How do we formalize that this is indeed a local max and not a local min?

1. **Second derivative test**: We found the critical point(s), but how to make sure that it's a maximum rather than a minimum (if we didn't have the useful graph...)
   1.1 Take second derivative of original function:

$$A'(h) = \frac{100}{3h^{\frac{1}{3}}} - 2$$

$$A''(h) = -\frac{100}{9h^{\frac{4}{3}}}$$

   1.2 Evaluate the second derivative at each critical point (in this case, only one) to *check its sign*

$$A''(h) = -\frac{100}{9 * (4629)^{\frac{4}{3}}} = -0.00014$$

   1.3 Signs of second derivative

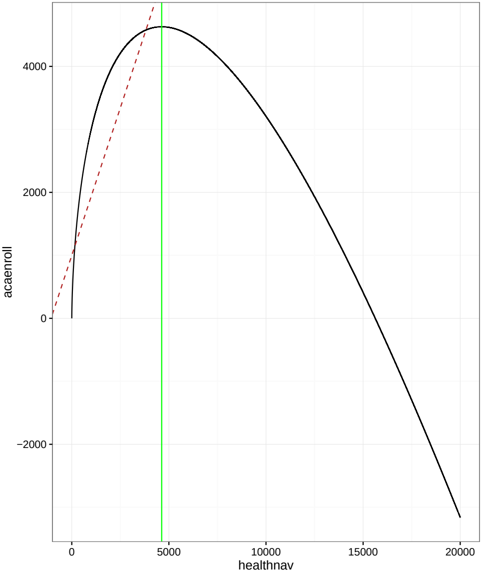
$f'(x)$	$f''(x)$	Classification
0	> 0	local min
0	< 0	local max
0	0	local min, local max, or inflection point (where f(x) changes concavity)

Graphical intuition for negative second derivative = local max; positive second derivative = local min

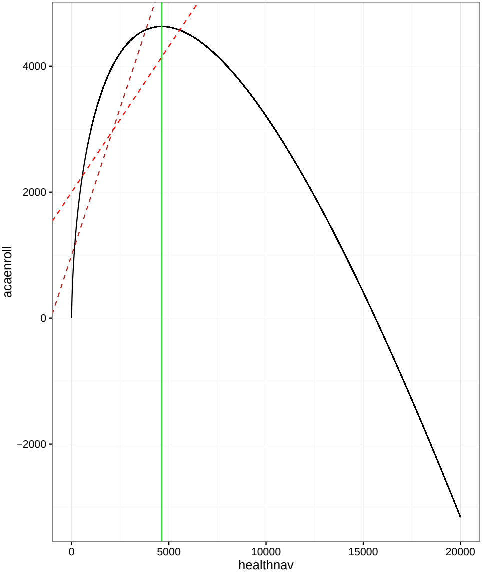
Basic idea: as we approach the maximum, the slope of the curve (first derivative) becomes less and less positive. Because the slope is going, for instance, from 0.937 around health navigators = 1000 to 0.463 around health navigators = 2000 to 0.199 around health navigators = 3000, the *slope of that slope* (the second derivative) is negative

To illustrate graphically (also in your code file)...

Graphical intuition for negative second derivative = local max; positive second derivative = local min



Graphical intuition for negative second derivative = local max; positive second derivative = local min



Graphical intuition for negative second derivative = local max; positive second derivative = local min