FAME Week Four Class Notes – Derivatives III

Andie Creel

11th November, 2025

1 Mean Value Theorem

Definition of Mean Value Theorem: For a smooth and continuous function, there is at least one point on a given interval where the derivative is equal to the average rate of change over the interval.

Intuition: The theorem says that if you draw the secant line (the straight line) connecting the point $(a, f(a))$ (where a is on the x-axis and $f(x)$ is on the y-axis) and the point $(b, f(b))$, then somewhere in the interval (a, b) there will be a point c where the tangent line (the line that touches the curve at exactly one point and has the same slope as the curve at that point) is parallel to the secant line. **In other words, at some point in the interval, the function's derivative matches the average rate of change over the interval,**

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{1}$$

2 Taylor Series: A Polynomial Approximation

We know that straight lines are good approximators because they minimize squared errors when fitting data—this is the foundation of linear regression. The "best fit" line through data points is the one that minimizes the sum of squared deviations.

Polynomial series (like Taylor series) extend this idea: they can approximate complex, nonlinear relationships by combining simple polynomial terms. Since any continuous function can be approximated arbitrarily well by polynomials, series give us a systematic way to build these approximations. This is very useful in science when we know our data was generated by some process, but we don't actually know what the function to describe that process/relationship is.

A **Taylor Series** is an extremely useful series for approximating the relationship of sequence of numbers (data). It's become the go-to way to approximate a function/relationship that you don't know. Taylor series underpins a lot of the math we do today, particularly for any relationship that isn't linear.

A Taylor Series expands a function around a specific x-value that is equal to a (called the center or point of expansion). The choice of a determines where your approximation will be most accurate:

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ... + \frac{f^k(a)}{k!}(x-a)^k + R \tag{2}$$

where:

- a is the **point of expansion** (you choose this!)
- $f(a)$ is the **value** of your function at point a
- $f'(a), f''(a), f'''(a)$... are the derivatives evaluated at point a
- R is the residual, or higher order terms

How to choose a:

- Pick a close to the x value you want to approximate $f(x)$ of. The closer a is to your target x, the better your approximation will be.
- The accuracy of your approximation does rely on how close the a you choose it to the value of x you're actually interested in.

How do you know/choose $f(x)$?

- Sometimes you know the function itself, but it's hard to calculate $f(x)$ at a specific point x that you're interested in, so you approximate it using a Taylor Series and a point of expansion a where it is easier to calculate $f(a)$ (see Section 3.2 for an example).
- Other times you do not know the function $y = f(x)$ that describes the relationship between x and y. You only know an unknown function could describe the relationship between these two variables. So you choose to use either an average, a line, a parabola, or a third order polynomial knowing that these simple, lower order polynomial models are a good approximation for how variables are plausibly related (Because the taylor series proves polynomials are good approximations of more complicated problems). You're not trying to be 100%, you're trying to get a useful approximation that gives you useful insight on important relationships.

2.1 0th order Taylor Series

Consider a 0th order taylor series: it would be a **mean** aka just a flat line equal to

$$f(x) \approx f(a)$$

If you're only looking at the means of a dataset, that means you're doing a 0th order taylor approximation.

2.2 First order Taylor Series

Consider a first order taylor series:

$$\begin{aligned} f(x) &\approx f(a) + \frac{f'(a)}{1!}(x-a)^1 \\ &= A + B(x-a) \\ &= A - aB + Bx \\ &= m + Bx \end{aligned}$$

It's a **line** with a slope. A first order Taylor series is a linear approximation. Linear regression is a first order Taylor approximation.

3 Mean Value Theorem and Taylor Series

Remember the Taylor series. A **Taylor Series** can approximate *any continuous function* $f(x)$. The modeler decides if they wants to approximate $f(x)$ with a 0th order taylor series (mean), a 1st order taylor series (line), a 2nd order taylor series (parabola), a 3rd order taylor series (at this point, you better have a LOT of data!!). We can approximate $f(x)$ with our taylor series,

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + ... + \frac{f^k(a)}{k!}(x-a)^k + R \tag{3}$$

where R is the residual.

Remember that what the **Mean Value Theorem** tells us: If you have two points, the slope between those two points will be parallel to the tangent line of the original function. So, the taylor series is a large application of the mean value theorem, where a taylor series approximates for the slope at a point many many times.

3.1 Discussion of models

Any Taylor Series approximation is a model. If someone says "I don't do models, let's only do means" they actually *are* suggesting a model. A mean is a model, it's a zero order Taylor Series approximation!

Higher order Taylor Series will approximate an underlying function better than a lower order. *However,* we do not always have enough data to fit a higher order Taylor series because it would require more estimating parameters and your data set may not have enough power (aka enough data) to do so.

3.2 Exercise: What is $\sqrt{10}$?

What is $\sqrt{10}$? The function we are considering is

$$f(x) = \sqrt{x} = x^{\frac{1}{2}}. \tag{4}$$

We're interested when $x = 10$. Let's do a **0th order taylor series approximation**:

$$f(x) = f(a) \tag{5}$$

what should we choose a to be? Let's choose $a = 9$ because 9 is close to 10.

$$\begin{aligned} f(10) &\approx f(a) && (6) \\ &\approx f(9) && (7) \\ &\approx \sqrt{9} && (8) \\ &\approx 3 && (9) \end{aligned}$$

So the 0th order approximation of $f(10) \approx f(9) = 3$. It's close!

Now, let's do a **first order taylor series approximation**

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x - a)^1. \tag{10}$$

We need to find $f'(a)$,

$$\begin{aligned} f(a) &= a^{1/2} && (11) \\ f'(a) &= \frac{1}{2}a^{-1/2} && (12) \end{aligned}$$

We can plug 12 back into 10 to get

$$f(x) \approx f(a) + \frac{1}{2}\frac{a^{-1/2}}{1}(x-a)^1 \tag{13}$$

Now, let's find $f(10)$ where $x = 10$ by using the taylor expansion around $f(9)$ where $a = 9$,

$$f(10) \approx f(9) + \frac{f'(9)}{1!}(10-9)^1 \tag{14}$$
$$\approx \sqrt{9} + \frac{1}{2}(9)^{-1/2} \times 1 \tag{15}$$
$$\approx 3 + \frac{1}{2}\frac{1}{(9)^{1/2}} \tag{16}$$
$$\approx 3\frac{1}{6} \tag{17}$$

So our first order approximation of $f(10) \approx 3\frac{1}{6}$. This is even closer than our original approximation, which was 3!

We can then do a **second order taylor series approximation** and get even closer.

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a)^1 + \frac{f''(a)}{2!}(x-a)^2 \tag{18}$$

we already found what $f'(a)$ is. What's $f''(a)$?

$$f(a) = a^{1/2} \tag{19}$$
$$f'(a) = \frac{1}{2}a^{-1/2} \text{ (power rule)} \tag{20}$$
$$f''(a) = -\frac{1}{4}a^{-3/2} \text{ (power rule again)} \tag{21}$$

We can plug in 21 into 18 (along with what we found in our first order approximation) to get

$$f(x) \approx f(a) + \frac{1}{2}\frac{a^{-1/2}}{1}(x-a)^1 + \frac{1}{2}(-\frac{1}{4})a^{-3/2}(x-a)^2 \tag{22}$$

we know from our first order that the first two terms simplify to $3\frac{1}{6}$. Lets simplify this expression

$$f(10) \approx 3 + \frac{1}{6} - \frac{1}{8}\frac{1}{9^{3/2}}(10-9)^2 \tag{23}$$
$$= 3 + \frac{1}{6} - \frac{1}{8}\frac{1}{27} \tag{24}$$
$$= 3.16203703704 \tag{25}$$

So what is $f(10) = \sqrt{10}$? $\sqrt{10} = 3.16227766017$. So we can see that our second order approximation got pretty close!

3.3 Example: Logistic growth

Consider the logistic growth function:

$$G(N) = rN(1 - \frac{N}{K}) \tag{26}$$

What would we get if we ”taylor expanded” this function? Let's work with the per capital growth rate

$$\frac{G(N)}{N} = r(1 - \frac{N}{K}) \tag{27}$$

and see if we can use a Taylor series that would approximate this right hand side equation

$$\frac{G(N)}{N} \approx G(a) + G'(a)(N - a) \tag{28}$$
$$\approx G(a) + G'(a)N - G'(a)a \tag{29}$$
$$\approx G(a) - G'(a)a + G'(a)N \tag{30}$$

Let $r = -G'(a)a$.

$$\approx r + G(a) + G'(a)N \tag{31}$$

Let $G(a) = 0$ and rename $G'(a) = \frac{-r}{K}$

$$\approx r + \frac{-r}{K}N \tag{32}$$
$$\approx r(1 - \frac{N}{K}) \tag{33}$$

And now we've shown that 33 is equivalent to 27. Which means we've shown that we can derive the logistic growth equation using a Taylor series approximation.

4 Partial derivatives

Consider a function that maps from $R^2 \rightarrow R^1$, *aka* it takes two variables as inputs (x, y) and you plug it into a function to get the value of Z,

$$Z = F(x, y). \tag{34}$$

Consider a situation where we can change x with policy, but we can't change y and we want to maximize outcome Z. In this case, we would take the derivative of $F(x,y)$ w.r.t x (treat y like a constant, assume it isn't changing).

We can take a **partial derivative** which notation that are all equivalent:

$$\frac{\partial F(x,y)}{\partial x} = F_x(x,y) = F_1(x,y) \tag{35}$$

This is how we find a **relative max/min**. Conditioned on y staying constant, we can find the relative max/min of the function $F(x,y)$ where we find the x that maximizes that function given some constant value of y.

4.1 Absolute max/min

If you want to find the absolute maximum or minimum you need to take *all* partial derivatives and set them all equal to zero and solve that system of equations. If your function mapped from $R^N \to R^1$ then you would take N partial derivatives and solve the system of N equations.

5 Total Derivatives

The total derivative of a function is equal to the sum of the partials multiplied by the change that we're interested. In other words, if x changed by Δx and y changed by Δy, how much did $f(x,y)$ change by?

Consider the function $F = f(x,y)$. We can get our partial derivatives. Recall that when we take a partial derivative wrt x, we're holding y constant.

$$\frac{\partial f}{\partial x} = f_x \tag{36}$$

$$\frac{\partial f}{\partial y} = f_y \tag{37}$$

Now consider how we would calculate the change in F if both x and y changed. We'd need to get a full derivative,

$$df = f_x dx + f_y dy \tag{38}$$

$$= \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy \tag{39}$$

$$\Delta f = = \frac{\partial f}{\partial x} \Delta x + \frac{\partial f}{\partial y} \Delta y \tag{40}$$

These are all equivalent ways of writing the full derivative.

5.1 Example of maximizing a multi-variate function

Consider the functions $F(x,y) = -x^2 - y^2$. Let's say that we want to maximize this function, and find the *global* maximum. To do so, we need to take the partial derivatives and set them equal to zero, then solve for x, y.

$$\frac{\partial F}{\partial x} = -2x = 0 \implies x = 0 \tag{41}$$
$$\frac{\partial F}{\partial y} = -2y = 0 \implies y = 0 \tag{42}$$

So the maximum point of $F(x,y) = -x^2 - y^2$ is $(0,0)$, where $x = 0$ and $y = 0$.

5.2 Implicit Function Theorem (envelope theorem in econ)

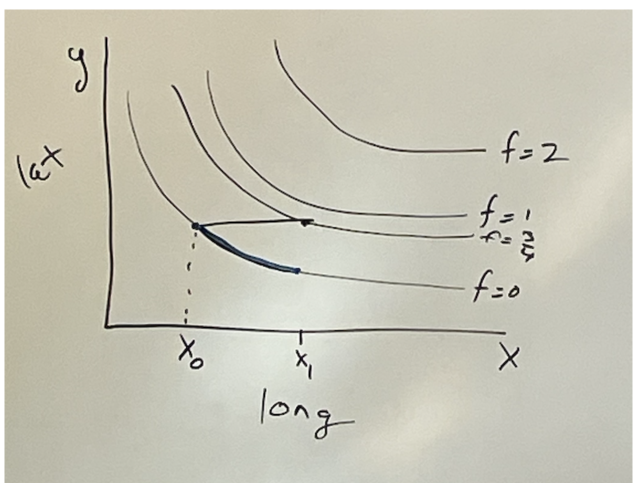

Figure 1: The f=0 represents an altitude

Now consider if you want to stay on one level set. For example, if you want to stay on a contour line or on a utility level. That would mean we want the value of $f(x,y)$ to stay the same, $df = \Delta f = 0$, even when we change both the x and y. We can set the full derivative to 0 in order to achieve this.

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy = 0 \implies \tag{43}$$

$$dy\frac{\partial f}{\partial y}dy = -dx\frac{\partial f}{\partial x} \implies \tag{44}$$

$$\frac{dy}{dx} = \frac{-\frac{\partial f}{\partial x}}{\frac{\partial f}{\partial y}} \tag{45}$$

Equation 45 is the implicit function theorem. What equation 45 tells us is "if I want to stay at the original value of $f(x, y)$, and I change x, how do I need to change y in order to stay at the original value of f". You could also think of this as you're hiking, and you're at a certain altitude. The same questions would be "if I change my latitude, how do I need to change my longitude in order to stay at the same altitude?"

The envelope theorem is a special case of the implicit function theorem when first derivatives are set to zero. This leads to math simplifying because we can say that, at the optimum, a bunch of derivatives will equal zero.