Derivatives

Mark Dean

Lecture Notes for Fall 2014 PhD Class - Brown University

I would guess that it would be almost impossible for you to get this far in your economics education without having a good intuitive (and probably quite good technical) understanding of how derivatives work, so this section will be very quick. First, a reminder of what a derivative is, and what we mean by a differentiable function on some open interval of $\mathbb{R}$

Definition 1 *Let $f:(a,b)\rightarrow\mathbb{R}$. Define the quotient function $\phi(t)$ as*

$$\phi(t)=\frac{f(t)-f(x)}{t-x}$$

We say that f is differentiable at $x\in(a,b)$, if $\lim_{t\rightarrow x}\ \phi(t)$ exists - in other words there exists some y such that, for every $\varepsilon>0$ there exists some $\delta>0$ such that $|t-x|<\delta$ implies that $|y-\phi(t)|<\varepsilon$. If this is the case, we define the derivative as $f'(x)=\lim_{t\rightarrow x}\ \phi(t)$

We say that a function is differentiable on (a,b) if it is differentiable at any $x\in(a,b)$. We say it is continuously differentiable if the function $f':(a,b)\rightarrow\mathbb{R}$ is continuous. Such functions are belong to the class $\mathcal{C}^1$. A function is twice (continuously) differentiable if $f':(a,b)\rightarrow\mathbb{R}$ is (continuously) differentiable. Such functions belong to the class $\mathcal{C}^2$

If $f:X\rightarrow\mathbb{R}$ where X is an open cube in $\mathbb{R}^n$, then, at any $x=(x_1,x_2,...x_n)$ we define the partial derivative with respect to x_i as

$$\frac{\partial f(x)}{\partial x_i}=\lim_{\varepsilon\rightarrow 0}\frac{f(x)-f(x+\varepsilon e_i)}{\varepsilon}$$

if such a limit exists, and define

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}$$

Obviously, the reason that we are here interested in derivatives is they tell us something about the slope of a function. In particular, we are going to be interested in how the derivative of a function can help us find local maxima and minima. For brevity we will deal with local maxima here, but local minima can be treated analagously.

Definition 2 *Let $f : X \to \mathbb{R}$ where $X \subset \mathbb{R}^n$. x^* is a local maximizer of f if there exists $\varepsilon > 0$ such that $f(x^*) \geq f(x)\ \forall\ x \in B(x^*, \varepsilon) \cap X$. It is a strict local maximizer if $f(x^*) > f(x)\ \forall$ $x \in B(x^*, \varepsilon) \cap X$ s.t. $x \neq x^*$*

Intuitively, we know that if x^* is a local maximizer in the interior of the domain of a function, then it must have a derivative of 0.

Lemma 1 *Let $f : (a, b) \to \mathbb{R}$ and $x^* \in (a, b)$ be a local maximizer, then $f'(x^*) = 0$*

Proof. *Note that, for any $\varepsilon > 0$,*

$$\begin{aligned} \phi(x+\varepsilon) &= \frac{f(x+\varepsilon) - f(x)}{\varepsilon} \\ &\Rightarrow f(x+\varepsilon) = \varepsilon\phi(x+\varepsilon) + f(x) \\ &= f(x) + \varepsilon f'(x) + \varepsilon\left(\phi(x+\varepsilon) - f'(x)\right) \end{aligned}$$

As $f(x^*)$ is a local maximizer, $f(x+\varepsilon) \leq f(x^*)$ for ε small enough and so

$$\begin{aligned} f(x^*) + \varepsilon f'(x^*) + \varepsilon\left(\phi(x^*+\varepsilon) - f'(x^*)\right) &\leq f(x^*) \\ &\Rightarrow f'(x^*) \leq -\left(\phi(x^*+\varepsilon) - f'(x^*)\right) \end{aligned}$$

but as $-(\phi(x^*+\varepsilon) - f'(x^*)) \to 0$ as $\varepsilon \to 0$, this implies $f'(x^*) \leq 0$

A similar argument gives

$$\begin{aligned} f(x-\varepsilon) &= -\varepsilon\phi(x-\varepsilon) + f(x) \\ &= f(x) - \varepsilon f'(x) + \varepsilon\left(f'(x) - \phi(x+\varepsilon)\right) \end{aligned}$$

and so

$$\begin{aligned} f(x^*) - \varepsilon f'(x^*) + \varepsilon \left(f'(x^*) - \phi(x^* - \varepsilon) \right) &\leq f(x^*) \\ &\Rightarrow f'(x^*) \geq \left(f'(x^*) - \phi(x^* - \varepsilon) \right) \end{aligned}$$

giving $f'(x^*) \geq 0$ *and so* $f'(x^*) = 0$ ■

We can use this result to derive Rolle's theorem.

Theorem 1 (Rolle) *Let* $f : [a, b] \to \mathbb{R}$ *be differentiable, and say* $f(a) = f(b)$*. Then* $f'(c) = 0$ *for some* $c \in (a, b)$

Proof. *As* f *is differentiable, it is continuous (see homework). This implies (by Weierstrass theorem) that there exists an* $a \leq x, y \leq b$ *such that* $f(x) \leq f(t) \leq f(y)\ \forall\ t \in [a, b]$*. If* $\{x, y\} = \{a, b\}$ *then* f *must be constant, and so* $f'(t) = 0\ \forall\ t \in [a, b]$*. Otherwise, either* $x \in (a, b)$ *or* $y \in (a, b)$*, and by lemma 1* $f'(x) = 0$ *or* $f'(y) = 0$ ■

Another useful thing we can do with derivatives is use them to approximate function: as the derivative gives us the slope of a function at a particular point x, then we can approximate $f(x+\varepsilon)$ by $f(x) + \varepsilon f(x)$. This is a Taylor series approximation. To make this precise, we are going to formally define the idea of an error being small:

Definition 3 $h : \mathbb{R}^n \to \mathbb{R}$ *is ''little oh' of order* k*, which we denote as* $h(x) = o(||x||^k)$ *if*

$$\lim_{x \to 0} \frac{h(x)}{||x||^k} = 0$$

Thus, if a function $h(x)$ is $o(||x||^k)$, then $h(x)$ gets small 'quickly', in the sense that it does so quicker than $\frac{1}{||x||^k}$ gets big.

Theorem 2 *Let* $f : [a, b] \to \mathbb{R}$ *be* $\mathcal{C}^2$*. Then for any* $x, x + \varepsilon \in (a, b)$

$$f(x + \varepsilon) = f(x) + f'(x)\varepsilon + o(\varepsilon)$$

Proof. *Note that this can be proved relatively easily from the definition of the derivative. Rearranging the above expression gives*

$$\frac{f(x + \varepsilon) - f(x)}{\varepsilon} - f'(x) = \frac{o(\varepsilon)}{\varepsilon}$$

The limit of the left hand side equals zero, and therefore so does the limit of the right hand side.

A more long winded proof, which is useful as it can be generalized to prove higher order approximations, is as follows:

Let

$$g(t) = f(t) - \left[f(x) + f'(x)(t-x)\right] - M(t-x)^2$$

Where

$$M = \frac{1}{\varepsilon^2}\left[f(x+\varepsilon) - f(x) - f'(x)\varepsilon\right]$$

Note that

$$\begin{aligned} g(x+\varepsilon) &= 0 \\ g(x) &= 0 \end{aligned}$$

Also note that

$$g'(t) = f'(t) - f'(x) - 2M(t-x)$$

and so $g'(x) = 0$

Applying Rolle's theorem tells us that there exists a $c_1 \in (x, x+\varepsilon)$ *such that* $g'(c_1) = 0$. *Applying Rolle's theorem again tells us that there is a* $c \in (x, c_1)$ *such that* $g''(c) = 0$,

As

$$g''(t) = f''(t) - 2M$$

This tells us that $f''(c) = 2M$ *and so*

$$\begin{aligned} f''(c) &= \frac{2}{\varepsilon^2}\left[f(x+\varepsilon) - f(x) - f'(x)\varepsilon\right] \\ &\Rightarrow f(x+\varepsilon) = f(x) + f'(x)\varepsilon + \frac{\varepsilon^2}{2}f''(c) \end{aligned}$$

Note here that c *here is really a function of the* ε *we initially chose. But, as* $f''(c)$ *is bounded (as it is continuous on* $[a,b]$*), then*

$$\frac{\frac{\varepsilon^2}{2}f''(c(\varepsilon))}{\varepsilon} = \frac{\varepsilon}{2}f''(c(\varepsilon))$$

tends to zero, giving the necessary result ■

An extension, which we will state but not prove, tells us that we can get an even better approximation if we also use a second derivative.

Theorem 3 *Let* $f : [a, b] \to \mathbb{R}$ *be* $\mathcal{C}^3$*. Then for any* $x, x + \varepsilon \in (a, b)$

$$f(x + \varepsilon) = f(x) + f'(x)\varepsilon + \frac{1}{2} f''(x)\varepsilon^2 + o(\varepsilon^2)$$