Econ 204
Taylor's Theorem

In this supplement, we give alternative versions of Taylor's Theorem. For univariate functions, we provide a different formulation of the error term using so-called "little oh" and "big Oh" notation. For multivariate functions, we provide the quadratic form of Taylor's Theorem (de la Fuente just provides the linear form, with quadratic error term) and analyze it as a quadratic form using the machinery in the Supplement to Section 3.6.

Definition 1 We say

$$y = o(x) \text{ as } x \to 0$$

if

$$\frac{|y|}{|x|} \to 0 \text{ as } x \to 0$$

and

$$y = O(x) \text{ as } x \to 0$$

if

$$\frac{|y|}{|x|} \text{ is bounded as } x \to 0$$

or more formally

$$\exists_M \; \exists_{\varepsilon > 0} \; |x| \leq \varepsilon \Rightarrow |y| \leq M|x|$$

The following theorem is a consequence of Theorem 1.9 on page 160 of de la Fuente. In my experience, knowing the exact form of the error term E_n as given in de la Fuente is not particularly useful, because one does not know in advance the location of $x + \lambda h$ at which E_n is evaluated. However, if f has an $(n+1)^{st}$ derivative which is continuous, one can obtain a $O\left(h^{n+1}\right)$ error term from the formula for E_n.

Theorem 2 (Taylor's Theorem for Univariate Functions) *Let* $f : I \to \mathbf{R}$ *be* n*-times differentiable, where* $I \subseteq \mathbf{R}$ *is an open interval. If* $x \in I$*, then*

$$f(x+h) = f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} + o\left(h^n\right) \; \text{ as } h \to 0$$

If f is $(n+1)$-times continuously differentiable, then

$$f(x+h) = f(x) + \sum_{k=1}^{n} \frac{f^{(k)}(x)h^k}{k!} + O\left(h^{n+1}\right) \text{ as } h \to 0$$

In the following theorem, Equation (1) is just a restatement of the definition of differentiability, while Equation (2) is a consequence of Theorem 4.4 on page 181 of de la Fuente. Note that the linear term $Df(x)(h)$ here and in de la Fuente is evaluated at the known point x. However, the quadratic term in de la Fuente is evaluated at the unknown point $x + \lambda h$; here, that term is incorporated into the "big Oh" error term. The version in de la Fuente is stated for functions from $\mathbf{R}^n$ to $\mathbf{R}^1$, while this version is stated for functions from $\mathbf{R}^n$ to $\mathbf{R}^m$; the restriction is needed in de la Fuente's formulation because the point $x + \lambda h$ will be different for different components in the range; the "big Oh" notation allows us to easily state Taylor's Theorem for functions taking values in $\mathbf{R}^m$.

Theorem 3 (Taylor's Theorem for Multivariate Functions–Linear Form) *Suppose $X \subseteq \mathbf{R}^n$ is open, $x \in X$, and $f : X \to \mathbf{R}^m$ is differentiable. Then*

$$f(x+h) = f(x) + Df(x)(h) + o\left(|h|\right) \text{ as } h \to 0 \tag{1}$$

If f is C^2, then

$$f(x+h) = f(x) + Df(x)(h) + O\left(|h|^2\right) \text{ as } h \to 0 \tag{2}$$

In understanding the geometry of preference relations and utility functions (including sufficient conditions for the differentiability of demand), it is very useful to have a quadratic version of the multivariate form of Taylor's Theorem. To keep notation simple, we restrict attention to the case of functions from $\mathbf{R}^n$ to $\mathbf{R}^1$; this suffices for the treatment of utility functions, and it is easy to generalize to functions from $\mathbf{R}^n$ to $\mathbf{R}^m$ by treating each component of the range separately.

Definition 4 Let $X \subseteq \mathbf{R}^n$ be open, $x \in \mathbf{R}$, and $f \in C^2(x)$. Let

$$D^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_1} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

denote the matrix of second partial derivatives of f, evaluated at x.

Recall that

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

so $D^2 f(x)$ is a symmetric matrix.

Theorem 5 (Taylor's Theorem for Multivariate Functions–Quadratic Form) *Suppose $X \subseteq \mathbf{R}^n$ is open, $x \in X$, and $f : X \to \mathbf{R}$ is C^2. Then*

$$f(x+h) = f(x) + Df(x)(h) + \frac{1}{2} h^T D^2 f(x) h + o\left(|h|^2\right) \text{ as } h \to 0$$

If f is C^3, then

$$f(x+h) = f(x) + Df(x)(h) + \frac{1}{2} h^T D^2 f(x) h + O\left(|h|^3\right) \text{ as } h \to 0$$

Remark 6 Theorem 5 is a stronger version of de la Fuente's Theorem 4.4. Note that we don't need to assume that X is convex. Since X is open, if $x \in X$, there exists $\delta > 0$ such that $B_\delta(x) \subseteq X$ and $B_\delta(x)$ *is* convex.

Because $D^2 f(x)$ is symmetric, we can apply the diagonalization results from the Supplement to Section 3.6, to obtain the following corollary:

Corollary 7 *Suppose $X \subseteq \mathbf{R}^n$ is open, $x \in X$, and $f : X \to \mathbf{R}$ is C^2. Then there is an orthonormal basis $\{v_1, \ldots, v_n\}$ of $\mathbf{R}^n$ such that*

$$\begin{aligned} & f(x + \gamma_1 v_1 + \ldots + \gamma_n v_n) \\ = \quad & f(x) + \sum_{i=1}^n \gamma_i \frac{\partial f}{\partial v_i}(x) + \frac{1}{2} \sum_{i=1}^n \lambda_i \gamma_i^2 + o\left(|\gamma|^2\right) \text{ as } \gamma \to 0 \end{aligned}$$

where

$$\frac{\partial f}{\partial v_i}(x) = Df(x) v_i$$

is the directional derivative of f in the direction v_i, evaluated at x. In addition,

1. *If f is C^3, then*

$$\begin{aligned} & f(x+\gamma_1 v_1+\ldots+\gamma_n v_n) \\ = \quad & f(x)+\sum_{i=1}^{n}\gamma_i\frac{\partial f}{\partial v_i}(x)+\frac{1}{2}\sum_{i=1}^{n}\lambda_i\gamma_i^2+O\left(|\gamma|^3\right) \text{ as } \gamma\to 0\end{aligned}$$

2. *If f has a local maximum or minimum at x, then $Df(x)=0$.*

3. *If $Df(x)=0$, then*

   (a) *If $\lambda_1,\ldots,\lambda_n>0$, then f has a local minimum at x.*

   (b) *If $\lambda_1,\ldots,\lambda_n<0$, then f has a local maximum at x.*

   (c) *If $\lambda_i<0$ for some i and $\lambda_j>0$ for some j, then f has a saddle at x (i.e. f has neither a local maximum nor a local minimum at x).*

   (d) *If $\lambda_1,\ldots,\lambda_n\geq 0$ and $\lambda_i>0$ for some i, then f has either a local minimum or a saddle at x.*

   (e) *If $\lambda_1,\ldots,\lambda_n\leq 0$ and $\lambda_i<0$ for some i, then f has either a local maximum or a saddle at x.*