Mean Value Theorem And Taylor Expansion

Mean Value Theorem:

Suppose $f(x)$ is continuous on $[a,b]$ and differentiable on (a,b). Then there exists a point $c \in (a,b)$ such that

$f(b) - f(a) = f'(c)(b-a)$

As a consequence:

(a) If $f'(x) \geq 0$ for all $x \in (a,b)$, then $f(x)$ is increasing.

(b) If $f'(x) = 0$ for all $x \in (a,b)$, then $f(x)$ is constant.

(c) If $f'(x) \leq 0$ for all $x \in (a,b)$, then $f(x)$ is decreasing.

Proof: Consider the function $g(x) = f(b) - f(x) + \frac{f(b)-f(a)}{b-a}(x-b)$.

$g(x)$ is continuous on $[a,b]$ and differentiable on (a,b).

$g(a) = g(b) = 0$.

It follows that $g(x)$ attains either a maximum or a minimum at some $c \in (a,b)$.

At this point, $g'(c) = 0$, which is equivalent to the mean value property.

Mean Value Theorem of Integrals:

If $f(x)$ is continuous on $[a,b]$, there exist a point $c \in [a,b]$, such that

$\int_a^b f(x)dx = f(c)(b-a)$

Proof: $f(x)$ is continuous on $[a,b]$, let the maximum and minimum value of $f(x)$ on $[a,b]$ be M and m, then $m \leq f(x) \leq M$.

Integral the last inequality, we have

$m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$, or $m \leq \int_a^b f(x)dx/(b-a) \leq M$.

Because $m \leq \int_a^b f(x)dx/(b-a) \leq M$, $m \leq f(x) \leq M$ and $f(x)$ is continuous, there exist a point c, such that

$\int_a^b f(x)dx/(b-a) = f(c)$, that's to say

$\int_a^b f(x)dx = f(c)(b-a)$.

Multivariate Mean Value Theorem:

Let function $f(x)$ map an open subset S of R^n to R^m. If $f(x)$ is differentiable on a neighborhood of $x \in S$, then

$f(y) = f(x) + \int_0^1 df[x + t(y-x)]dt(y-x)$

for y near x.

Proof: Integrating component by component, we only consider the case $m = 1$.

The real-valued function $g(t) = f[x + t(y-x)]$ of the scalar t has differential $dg(t) = df[x + t(y-x)](y-x)$.

Because differentials and derivatives coincide on the real line, the equality follows from the fundamental theorem of calculus applied to $g(t)$.

Taylor's Theorem:

Let

$L_2[0,1]$ be the set of all square integrable functions on the interval $[0,1]$.

$C^m[0,1]$ ={μ: $\mu^{(j)}$ is continuous, $j = 0, \cdots, m$}

$W_2^m[0,1]$ ={μ: $\mu^{(j)}$ is absolutely continuous, $j = 0, \cdots, m-1$, and $\mu^{(m)} \in L_2[0,1]$}

Then $W_2^m[0,1] \supset C^m[0,1]$ and $W_2^0[0,1] = L_2[0,1]$.

If $\mu \in W_2^m[0,1]$, then there exists coefficient $\theta_1, \cdots, \theta_m$ such that

$\mu(t) = \Sigma_{j=1}^m \theta_j t^{j-1} + \int_0^1 \frac{(t-u)_+^{m-1}}{(m-1)!} \mu^{(m)}(u)du$,

where $(x)_+^r = \begin{cases} x^r, & x \geq 0, \\ 0, & x < 0. \end{cases}$

Proof: Write $\mu(t) = \int_0^1 (t-u)_+^0 \mu'(u)du + \mu(0)$ and integrate by parts.

Taylor's Theorem suggests that if, for some positive integerλ, the remainder term

$Rem_\lambda(t) = [(\lambda-1)!]^{-1} \int_0^1 (t-u)_+^{\lambda-1} \mu^{(\lambda)}(u)du$

is uniformly small then we could write

$y_i \approx \Sigma_{j=1}^{\lambda} \theta_j t_i^{j-1} + \epsilon_i$, $i = 1, \cdots, n$.

In other words, the data would follow an approximate polynomial regression model.

If $f: R^d \to R$ and if $\ddot{f}(x)$ is continuous in the sphere $O_r(x_0)$, then for $|t| < r$,

$f(x_0 + t) = f(x_0) + \dot{f}(x_0)t + t^T \int_0^1 \int_0^1 v\ddot{f}(x_0 + uvt)dudvt.$

Proof: Since $\ddot{f}(x_0 + uvt)$

$$= \begin{pmatrix} \dot{f}_1 \\ \dot{f}_2 \\ \vdots \\ \dot{f}_d \end{pmatrix} = \begin{pmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f(x_0+uvt) & \frac{\partial^2}{\partial x_1 \partial x_2} f(x_0+uvt) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_d} f(x_0+uvt) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(x_0+uvt) & \frac{\partial^2}{\partial x_2 \partial x_2} f(x_0+uvt) & \cdots & \frac{\partial^2}{\partial x_2 \partial x_d} f(x_0+uvt) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2}{\partial x_d \partial x_1} f(x_0+uvt) & \frac{\partial^2}{\partial x_d \partial x_2} f(x_0+uvt) & \cdots & \frac{\partial^2}{\partial x_d \partial x_d} f(x_0+uvt) \end{pmatrix}$$

it follows that

$t^T \int_0^1 \int_0^1 v\ddot{f}(x_0 + uvt)dudvt$

$= \int_0^1 \int_0^1 vt^T \ddot{f}(x_0 + uvt)tdudv$

$= \int_0^1 \int_0^1 v\Sigma_{i=1}^d \Sigma_{j=1}^d t_i \frac{\partial^2}{\partial x_i \partial x_j} f(x_0 + uvt)t_j dudv$

$= \Sigma_{i=1}^d t_i \int_0^1 v \int_0^1 \Sigma_{j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(x_0 + uvt)t_j dudv$

$= \Sigma_{i=1}^d t_i \int_0^1 \{v \int_0^1 \Sigma_{j=1}^d \frac{\partial}{\partial x_j}[\frac{\partial}{\partial x_i} f(x_0 + uvt)]t_j\}dudv$

$= \Sigma_{i=1}^d t_i \int_0^1 \{\int_0^1 \Sigma_{j=1}^d \frac{\partial}{\partial x_j}[\frac{\partial}{\partial x_i} f(x_0 + uvt)]vt_j\}dudv$

$= \Sigma_{i=1}^d t_i \int_0^1 \{\int_0^1 (vt)^T \frac{\partial}{\partial x}[\frac{\partial}{\partial x_i} f(x_0 + uvt)]\}dudv$

$= \Sigma_{i=1}^d t_i \int_0^1 \{\int_0^1 \frac{\partial}{\partial u}[\frac{\partial}{\partial x_i} f(x_0 + uvt)]du\}dv$

$= \Sigma_{i=1}^d t_i \int_0^1 \{\frac{\partial}{\partial x_i} f(x_0 + vt) - \frac{\partial}{\partial x_i} f(x_0)\}dv$

$= \int_0^1 \Sigma_{i=1}^d t_i \frac{\partial}{\partial x_i} f(x_0 + vt)dv - \Sigma_{i=1}^d t_i \int_0^1 \frac{\partial}{\partial x_i} f(x_0)dv$

$= \int_0^1 \frac{\partial}{\partial v} f(x_0 + vt)dv - \dot{f}(x_0)t$

$= f(x_0 + t) - f(x_0) - \dot{f}(x_0)t$

References:

Optimization, Kenneth Lange

Nonparametric Regression and Spline Smoothing, Eubank R.
A Course in Large Sample Theory(Lecture notes), Xianyi Wu