What Are the Mean Value and Taylor Theorems Saying?

We have studied two propositions about the derivative of a function that sound vaguely alike.

(1) On the one hand, the **mean value theorem** (Week 13, Stewart 3.2) says that

$$f(x) = f(a) + f'(c)\,(x-a) \quad \text{(exactly!)}$$

for some c between a and x.

For example, $e^x = 1 + e^c x$.

(2) On the other hand, the **best linear approximation** (Week 16, Stewart 2.9) says that

$$f(x) = f(a) + f'(a)\,(x-a) + \text{something small}$$

if x is close to a.

For example, $e^x \approx 1 + x$ (that is, $e^x = 1 + x +$ a small error).

Quick exercise: What are the **two** crucial differences between (1) and (2)? Read closely.

An immediately interesting question about statement (2) is: **How** small? The word "small" by itself is too vague to be useful in technical work.

We can cast some light on the relation between (1) and (2) by stepping back a bit to the **definition of continuity**, which can be paraphrased this way:

(0) If f is continuous, then

$$f(x) = f(a) + \text{something small}$$

if x is close to a.

For example, $e^x \to e^0 = 1$ as $x \to 0$.

Again one should ask, "**How** small?" In this case the answer is given by the detailed definition of continuity (through the definition of a limit): For every ϵ there is a δ such that if $|x-a| < \delta$, then $|f(x) - f(a)| < \epsilon$. We will talk about statements like this in more depth in Week 26; for now the only thing to understand is that (0) merely says that as $x - a$ goes to 0,

$f(x) - f(a)$ also goes to 0 — it says nothing about **how fast** $f(x) - f(a)$ approaches 0.

But now look again at the mean value theorem (alias "Taylor's theorem with $N = 0$"): It says that **if f is differentiable**, then

$$f(x) = f(a) + f'(c)\,(x - a)$$

and so, if $|x - a| < \delta$, then*

$$|f(x) - f(a)| < \delta \max_{c \in (a,x)} |f'(c)|.$$

The important thing about this inequality is that it says that the difference between $f(x)$ and $f(a)$ is (at worst) **linear** in δ, the distance between x and a. Thus (1) is a sharpened form of (0), valid if f is differentiable, not just continuous.

But differentiability also gives us statement (2), and we want to know what "small" means in it. One type of answer is provided by Taylor's theorem with $N = 1$:

(3) If the **second** derivative of f exists, then

$$f(x) = f(a) + f'(a)\,(x - a) + \tfrac{1}{2} f''(c)\,(x - a)^2$$

for some c.

For example, $e^x = 1 + x + \frac{1}{2} e^c x^2$.

Therefore, if $|x - a| < \delta$, then

$$|f(x) - [f(a) + f'(a)\,(x - a)]| < \tfrac{1}{2}\delta^2 \max |f''(c)|.$$

That is, the error is **quadratically** small in its dependence on δ.

If $f'(a)$ exists but $f''(a)$ doesn't, then it turns out that the error in the linear approximation (2) typically goes to zero faster than δ but not as fast as δ^2. (Contemplate, for example, $f(x) = x^{4/3}$ around $x = 0$.) We shall not discuss this sort of situation further (leaving it to advanced courses in mathematics).

---

* $c \in (a, x)$ means $a < c < x$, which is the right formulation if $a < x$; if $x < a$, of course, we would have $c \in (x, a)$.

Of course, when f'' exists we expect to form a **best quadratic approximation**:

(4) If x is close to a, then

$$f(x) \approx f(a) + f'(a)\,(x-a) + \tfrac{1}{2} f''(a)\,(x-a)^2.$$

For example, $e^x \approx 1 + x + \frac{1}{2}x^2$.

We can ask the usual question: Exactly **how** small is the error in the approximation (4)? By now you can probably guess the answer:

(5) If the **third** derivative exists, then (Taylor's theorem with $N = 2$)

$$f(x) = f(a) + f'(a)\,(x-a) + \tfrac{1}{2} f''(a)\,(x-a)^2 + \tfrac{1}{6} f^{(3)}(c)\,(x-a)^3$$

for some c, and therefore the error in the quadratic approximation is **cubic** in δ.

Obviously we can continue this game forever; successfully completing each step invites us to attempt the next step. If $f^{(N)}(a)$ exists, then we can construct the Nth Taylor polynomial

$$T_N(x) = \sum_{j=0}^{N} \frac{f^{(j)}(a)}{j!}\,(x-a)^j$$

and expect it to be the **best approximation (near** a**) to** $f(x)$ **by a polynomial of degree** N. (This is proposition number $(2N)$.) We ask how good this approximation really is. Then Taylor's theorem says that if $f^{(N+1)}(a)$ exists, then

$$f(x) = T_N(x) + R_N(x),$$

with

$$R_N(x) = \frac{f^{(N+1)}(c)}{(N+1)!}\,(x-a)^{N+1},$$

and therefore the error in the Nth-degree approximation vanishes as fast as δ^{N+1} as $\delta = |x-a|$ approaches 0. (This is proposition number $(2N+1)$.) This formula tempts us to consider the even better approximation T_{N+1}, but we have to stop somewhere.