Stat 8112 Lecture Notes

The Finite-Dimensional Delta Method

Charles J. Geyer

January 23, 2013

Theorem 7 in Ferguson (1996) is not as general as it can be in two respects. The conditions are overly strong in two ways: the limit distribution need not be normal and the conditions on the function g are overly strong.

We say a function $g : U \to W$, where U is an open set in a finite-dimensional vector space V and W is another finite-dimensional vector space is *differentiable* at a point $x \in U$ (Browder, 1996, Definition 8.9) if

$$g(x + h) = g(x) + Th + o(h) \tag{1}$$

for some linear transformation $T : V \to W$, in which case it can be shown that T is unique, and we write $g'(x)$ or $\nabla g(x)$ in place of T (the $o(h)$ here means as $h \to 0$).

Thus the elegant way to think about differentiation is that derivatives are linear transformations. Of course, in finite-dimensional vector spaces, linear transformations can be represented by matrices, and we sometimes confuse the linear transformation and the matrix that represents it by saying the derivative is a matrix.

When all partial derivatives exist and are continuous at x, then the function is differentiable at x (Browder, 1996, Theorem 8.23) and the matrix representing the derivative is the matrix of partial derivatives. But this condition is only sufficient not necessary. Thus the condition used by Ferguson (continuous partial derivatives) is overly strong.

Theorem 1 (Delta Method)**.** *Suppose X_1, X_2, ..., is a sequence of random vectors, Y is another random vector, ξ is a constant vector, a_n is a sequence of positive scalars converging to infinity,*

$$a_n(X_n - \xi) \xrightarrow{w} Y, \tag{2}$$

g is a function from an open subset of the finite-dimensional vector space where X_n, Y, and ξ live to another finite-dimensional vector space, and g is differentiable at ξ. Then

$$a_n\bigl[g(X_n) - g(\xi)\bigr] \xrightarrow{w} g'(\xi)Y, \tag{3}$$

the right-hand side denoting the linear transformation $g(\xi)$ applied to the random vector Y.

Proof. We can rewrite (1) as

$$g(y) = g(x) + g'(x)(y - x) + \|y - x\| w(y - x)$$

where $w(z)$ is $o(1)$ as $z \to 0$. Another way to say this is that w is continuous at zero (perhaps nowhere else) with $w(0) = 0$. Hence

$$a_n \bigl[g(X_n) - g(\xi) \bigr] = g'(\xi) \cdot a_n (X_n - \xi) + \| a_n (X_n - \xi) \| \cdot w(X_n - \xi)$$

By (2) and the continuous mapping theorem, the first term on the right-hand side converges to the right-hand side of (3). By (2) and the converse part of Prohorov's theorem

$$a_n (X_n - \xi) = O_p(1),$$

hence

$$X_n - \xi = O_p(a_n^{-1}) = o_p(1). \tag{4}$$

By the continuous mapping theorem, (4), and the continuity of w at zero,

$$w(X_n - \xi) = o_p(1).$$

The theorem now follows from $O_p(1) o_p(1) = o_p(1)$, which will be a homework problem, and Slutsky's theorem. □

References

Browder, A. (1996). *Mathematical Analysis: An Introduction.* New York: Springer.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* London: Chapman & Hall.