Calculating higher order derivatives via differentials

Mwiza Kunda

January 26, 2021

Content summarised from:

Magnus, J. R. and H. Neudecker (2019). Matrix Differential Calculus with Applications in Statistics and Econometrics, Third Edition, John Wiley, New York.

1 First differentials

special case of the first identification theorem: $\mathrm{d}\phi = \alpha(x)\mathrm{d}x \Longleftrightarrow \frac{d\phi(x)}{dx} = \alpha(x)$

Jacobian in **numerator** layout:

$$\frac{\partial f(x)}{\partial x'} = \begin{pmatrix} \partial f_1(x)/\partial x_1 & \partial f_1(x)/\partial x_2 & \dots & \partial f_1(x)/\partial x_n \\ \partial f_2(x)/\partial x_1 & \partial f_2(x)/\partial x_2 & \dots & \partial f_2(x)/\partial x_n \\ \vdots & \vdots & & \vdots \\ \partial f_m(x)/\partial x_1 & \partial f_m(x)/\partial x_2 & \dots & \partial f_m(x)/\partial x_n \end{pmatrix}$$

First identification theorem:

$$\mathrm{d}f = A(x)\mathrm{d}x \Longleftrightarrow \frac{\partial f(x)}{\partial x'} = A(x)$$

2 Second differentials

Defintion:

$$\mathrm{d}^2 f = \mathrm{d}(\mathrm{d}f)$$

For a scalar function ϕ, the second differential is a quadratic form in :

$$\mathrm{d}^2\phi = (\mathrm{d}a)'\mathrm{d}x = (\mathrm{d}x)'(\mathrm{H}\phi)\mathrm{d}x$$

Second identification theorem (symmetric hessian matrix)

$$\mathrm{d}^2\phi = (\mathrm{d}x)'B(x)\mathrm{d}x \Longleftrightarrow \mathrm{H}\phi = \frac{B(x)+B(x)'}{2}$$

3 Vec and Kronecker product

3.1 Vec

First, the vec operator. Consider an $m \times n$ matrix A. This matrix has n columns, say $a_1, \dots, a_n$. Now define the $mn \times 1$ vector vec A as the vector which stacks these columns one underneath the other:

$$\operatorname{vec} A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

3.1.1 Properties

- $\mathrm{d} \operatorname{vec} X = \operatorname{vec} \mathrm{d} X$
- $\operatorname{tr} A'B = (\operatorname{vec} A)'(\operatorname{vec} B)$

3.2 Kronecker product

Let A be an $m \times n$ matrix and B a $p \times q$ matrix. The $mp \times nq$ matrix defined by

$$\begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix}$$

is called the Kronecker product of A and B and is written as $A \otimes B$.

3.2.1 Properties

- left and right associativity (=) $A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C)$
- Distributivity: $(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D$,
- if A and B have the same order and C and D have the same order (not necessarily equal to the order of A and B), then $(A \otimes B)(C \otimes D) = AC \otimes BD$
- Transpose: $(A \otimes B)' = A' \otimes B'$
- Trace: If A and B are square matrices (not necessarily of the same order), then $\operatorname{tr}(A \otimes B) = (\operatorname{tr} A)(\operatorname{tr} B)$
- Inverses: if A and B are nonsingular, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

3.3 Other tools

3.3.1 Commutation Matrix

Let A be an $m \times n$ matrix. The vectors vec A and vec A' contain the same mn elements, but in a different order. Hence, there exists a unique $mn \times mn$ matrix, which transforms vec A into vec

A'. This matrix contains mn ones and $mn(mn-1)$ zeros and is called the commutation matrix, denoted by K_{mn}. (If $m=n$, we write K_n instead of K_{nn}.) Thus,

$$K_{mn} \operatorname{vec} A = \operatorname{vec} A'$$

It can be shown that K_{mn} is orthogonal, i.e. $K'_{mn} = K^{-1}_{mn}$. Also, premultiplying (20) by K_{nm} gives $K_{nm}K_{mn}$ vec $A = \operatorname{vec} A$, which shows that $K_{nm}K_{mn} = I_{mn}$. Hence,

$$K'_{mn} = K^{-1}_{mn} = K_{nm}$$

The key property of the commutation matrix enables us to interchange (commute) the two matrices of a Kronecker product:

$$K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$$

Key use is in manipulating expressions to be in line with identification theorems:

$$\mathrm{d} \operatorname{vec} F = \operatorname{vec} \mathrm{d}X' = K_{nq} \operatorname{vec} \mathrm{d}X$$

3.3.2 Duplication Matrix

When differentiating with respect to symmetric matrices, symmetry must be taken into account. In comes the Duplication Matrix. Let A be a square $n \times n$ matrix. Then vech (A) will denote the $\frac{1}{2}n(n+1) \times 1$ vector that is obtained from vec A by eliminating all elements of A above the diagonal. In this way, vech (A) contains the unique elements of the matrix A.

There exists a unique matrix of order $n^2 \times \frac{1}{2}(n)(n+1)$ which transforms for a symmetric matrix A, vech A into vec A. This matrix is called the duplication matrix and is denoted by D_n:

$$D_n \operatorname{vech}(A) = \operatorname{vec} A \quad (A = A')$$

and

The matrix D_n has full column rank $\frac{1}{2}n(n+1)$, so that D'_nD_n is nonsingular. This implies that vech(A) can be uniquely solved from (23), and we have

$$\operatorname{vech}(A) = (D'_nD_n)^{-1} D'_n \operatorname{vec} A \quad (A = A')$$

The duplication matrix is linked to the commutation matrix:

$$K_nD_n = D_n, \quad D_n \left(D'_nD_n\right)^{-1} D'_n = \frac{1}{2}\left(I_{n^2} + K_n\right)$$

Much of the interest in the duplication matrix is due to the importance of the matrix $D'_n(A \otimes A)D_n$, where A is an $n \times n$ matrix. This matrix is important, because the scalar function $\phi(X) = \operatorname{tr} AX'AX$ occurs frequently in statistics and econometrics, for example in the next section on maximum likelihood. When A and X are known to be symmetric we have

$$\begin{aligned} \mathrm{d}^2\phi &= 2 \operatorname{tr} A(\mathrm{~d}X)'A \mathrm{~d}X = 2(\mathrm{~d} \operatorname{vec} X)'(A \otimes A)\mathrm{d} \operatorname{vec} X \\ &= 2(\mathrm{~d} \operatorname{vech}(X))'D'_n(A \otimes A)D_n \mathrm{~d} \operatorname{vech}(X) \end{aligned}$$

and hence, $\mathrm{H}\phi = 2D'_n(A \otimes A)D_n$.

3.4 Relation between Vec and Kronecker product

Vec and the kronecker product are related via:

$$\operatorname{vec} ab' = b \otimes a$$

And more generally:

Theorem 18.5: For any matrices A, B, and C for which the product ABC is defined, we have

$$\operatorname{vec} ABC = (C' \otimes A) \operatorname{vec} B$$

3.5 Identification using vectorization

First identification, $A(X)$ is the derivative

$$\mathrm{d} \operatorname{vec} F = A(X) \mathrm{d} \operatorname{vec} X \Longleftrightarrow \frac{\partial \operatorname{vec} F(X)}{\partial (\operatorname{vec} X)'} = A(X)$$

Second identification, $H\phi$ is the hessian matrix (ϕ is a scalar function)

$$\mathrm{d}^2 \phi = (\mathrm{d} \operatorname{vec} X)' B(X) \mathrm{d} \operatorname{vec} X \Longleftrightarrow \mathrm{H}\phi = \frac{B(X) + B(X)'}{2}$$