Mathematics and Statistics Primer
QBUS6830

1. Review of Some Basic Statistics

See Tsay Chapter 2

2. Rudiments of Matrix Algebra

See Tsay Chapter 8, Appendix A (and other notes provided)

3. Some Useful Algebraic Notation

a) The "summation" or "sigma" notation is a way of writing sums of numbers in shorthand: $x_1 + x_2 + \ldots + x_n = \sum_{i=1}^{n} x_i$

b) The product notation is similar to the sigma notation but for multiplication:

$x_1 x_2 \ldots x_n = \prod_{i=1}^{n} x_i$ is shorthand for multiplying together all the numbers from x_1 up to x_n

c) When we have two products, with both of them having their index going <u>over the same range</u>, ie. in the case below both go from 1 to n, then we can write:

$$\left(\prod_{i=1}^{n} x_i\right)\left(\prod_{j=1}^{n} y_i\right) = (x_1 x_2 \ldots x_n)(y_1 y_2 \ldots y_n) = \prod_{i=1}^{n} (x_i y_i)$$

d) A summation over two indexes can be written as:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j} = (x_{1,1} + x_{1,2} + \ldots + x_{1,n}) + (x_{2,1} + x_{2,2} + \ldots + x_{2,n}) +$$
$$\ldots\ldots + \ldots + (x_{m,1} + x_{m,2} + \ldots + x_{m,n})$$

In all, the above formula adds together $m \times n$ numbers.

e) In the case when $m = n$ and $x_{i,j} = x_{j,i}$, there is a symmetry in the set of numbers and we can write the above formula as:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{i,j} = \sum_{i=1}^{n} x_{i,i} + 2\sum_{i=1}^{n}\sum_{j<i} x_{i,j}$$

Think of the $x_{i,j}$ as being the elements of a matrix, and you will see why there is a symmetry when $m = n$ and $x_{i,j} = x_{j,i}$.

4. Exponentials and Logarithms

a) The exponential of a value *x* is written as e^x or exp(*x*)

b) The (natural) logarithm of a value *x* is written as ln(*x*). Since logarithms to the base 10 are rarely used in econometrics, some texts and computer programs also use log(*x*) for the natural logarithm

c) The exponential is the inverse function of the natural logarithm, so that:
$\ln(e^x) = \exp(\ln(x)) = x$

d) One important property of logarithms is that:

$$\ln\left(\prod_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} \ln(x_i) \, .$$

See Algebra Rules 4a) and 4b) for the notations $\sum_{i=1}^{n} x_i$ and $\prod_{i=1}^{n} x_i$.

e) One important property of exponentials is that:

$$\exp\left(\sum_{i=1}^{n} x_i\right) = \prod_{i=1}^{n} \exp(x_i)$$

Some index rules for exponentials and logarithms

f) $e^a e^b = e^{(a+b)}$ (from (e) above with x_1 = a and x_2 = b)

g) $\left(e^a\right)^b = e^{ab}$

h) $e^a / e^a = e^a e^{-b} = e^{a-b}$

i) $\ln(x^a) = a \ln(x)$

j) $\ln(1/x) = \ln(x^{-1}) = -\ln(x)$

k) $\ln\left(x/y\right) = \ln(x) + \ln\left(1/y\right) = \ln(x) - \ln(y)$

Some differentials and integrals for exponentials and logarithms

l) $\dfrac{d\left(e^{ax}\right)}{dx} = ae^{ax}$.

When $a = 1$ we get $\dfrac{d\left(e^{x}\right)}{dx} = e^{x}$

m) $\dfrac{d\left(\ln(ax)\right)}{dx} = \dfrac{d\left(\ln(a)+\ln(x)\right)}{dx} = \dfrac{d\left(\ln(a)\right)}{dx} + \dfrac{d\left(\ln(x)\right)}{dx} = 0 + 1/x = 1/x$.

When $a = 1$ we get $\dfrac{d\left(\ln(x)\right)}{dx} = 1/x$

n) $\int\left(e^{ax}\right)dx = c + \dfrac{1}{a}e^{ax}$ where c is the constant of integration.

When $a = 1$ we get $\int e^{x}dx = c + e^{x}$

o) $\int\left(\dfrac{1}{a+x}\right)dx = c + \ln(a+x)$ where c is the constant of integration.

When $a = 0$ we get $\int\left(\dfrac{1}{x}\right)dx = c + \ln(x)$

5. Partial Differentiation

a) When we differentiate the function $f(x)$ with respect to x we write the differential, or derivative, as $\dfrac{df(x)}{dx}$, or sometimes simply $\dfrac{df}{dx}$.
If f is a function of two variables, say of x and y, then we can differentiate with respect to either x (treating y like it's a constant), or with respect to y (treating x like it's a constant). This is called partial differentiation and the two partial derivatives are written as: $\dfrac{\partial f(x,y)}{\partial x}$ and $\dfrac{\partial f(x,y)}{\partial y}$, or in shorthand, $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$.
It's the use of the curly ∂ instead of d that indicates partial differentiation.

b) As an example, consider $f(x, y) = x^2 + 3xy - \frac{y^2}{2}$.

Looking at it as a function of *x* and treating *y* as an unknown constant, we get:
$\frac{\partial f(x, y)}{\partial x} = 2x + 3y$.

Looking at it as a function of *y* and treating *x* as unknown constant, we get:
$\frac{\partial f(x, y)}{\partial y} = 3x - y$.

6. Some Useful Rules for Expectations, Variances and Covariances

a) If $X_1, X_2, \ldots, X_n$ is a sequence of random variables, then the weighted sum, $\sum_{i=1}^{n} b_i X_i = b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$, is called a linear combination of the random variables.

If we write $X_1, X_2, \ldots, X_n$ as a column vector $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$, and the constants $b_1, \ldots, b_n$ as a row vector $b' = (b_1, b_2, \ldots b_n)$, then we can also write this linear combination as: $\sum_{i=1}^{n} b_i X_i = b' X$.

b) Again, write the sequence of random variables $X_1, X_2, \ldots, X_n$ as a column vector *X*. This is sometimes called a random vector. Also, let *B* be an $m \times n$ matrix of constants (i.e. not random variables). Then we get:

- $E(BX) = BE(X)$ which has dimension $m \times 1$
- $\text{var}(BX) = B\text{var}(X)B'$ which has dimension $m \times m$.

In many cases, instead of a matrix *B* we have a row vector *b'* as in 6(a), which can be thought of as one row of the matrix *B*¸ or as a $1 \times n$ matrix. Then we get:

- $E(b'X) = b'E(X)$ which has dimension 1×1
- $\text{var}(b'X) = b'\text{var}(X)b$ which has dimension 1×1.

The above results give us the expectation (ie. the mean) and the variance of the linear combination $\sum_{i=1}^{n} b_i X_i = b'X$.

c) We can use Rule 4e) and the expectation rules, to show that if we have a set of random variables X_i, then the variance of the sum of the X_i can be written as:

$$\operatorname{var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}\left(X_{i,j}\right) = \sum_{i=1}^{n}\operatorname{cov}\left(X_{i,i}\right) + 2\sum_{i=1}^{n}\sum_{j<i}\operatorname{cov}\left(X_{i,j}\right)$$

$$= \sum_{i=1}^{n}\operatorname{var}\left(X_i\right) + 2\sum_{i=1}^{n}\sum_{j<i}\operatorname{cov}\left(X_{i,j}\right)$$

d) The Law of Conditional Expectation states that if X and Y are two random variables (not necessarily independent of each other), then:

$$E(Y) = E_X\left[E(Y/X)\right]$$

where $E(Y/X)$ is the expectation of Y conditional on X (ie. while treating X as if it's a constant), and $E_X[.]$ means that we take the expectation of the expression in the square brackets only with respect to X.

e) A similar result for the variance gives us the Decomposition of Variance rule:

$$\operatorname{var}(Y) = \operatorname{var}_X\left[E(Y/X)\right] + E_X\left[\operatorname{var}(Y/X)\right]$$

where $\operatorname{var}_X[.]$ means we're taking the variance of the expression in the square brackets only over the distribution of X.

f) If X and Y are two random variables, then the correlation coefficient between X and Y is: $\rho = \dfrac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)\operatorname{var}(\mathrm{Y})}}$ which satisfies the restriction $-1 \leq \rho \leq +1$.

7. Matrix Differentiation

Rule 1

$$\frac{\partial(\mathrm{a}'\mathrm{x})}{\partial \mathrm{x}} = \frac{\partial(\mathrm{x}'\mathrm{a})}{\partial \mathrm{x}} = \mathrm{a}$$

where a= $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ and x= $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

***Proof*:**

$$\mathrm{a}'\mathrm{x} \;=\; \mathrm{x}'\mathrm{a} = a_1x_1 + a_2x_2 + \;....+a_nx_n$$

$$\therefore \frac{\partial(\mathrm{a}'\mathrm{x})}{\partial x_1} = a_1$$

$$\frac{\partial(\mathrm{a}'\mathrm{x})}{\partial x_2} = a_2$$

etc

Rule 1*

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

where A= $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ and x= $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

***Proof*:**

Write A in terms of its column vectors

$\mathrm{A}=(\mathrm{a}_1,\ \mathrm{a}_2,\ ...,\ \mathrm{a}_\mathrm{n})$.

Then apply Rule 1 to each of $\mathrm{x}'\mathrm{a}_1,\ \mathrm{x}'\mathrm{a}_2$ etc.

Rule 2

$$\frac{\partial(\mathrm{x'Ax})}{\partial \mathrm{x}} = 2\mathrm{Ax}$$

where A= $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ and x= $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

and A is symmetric; ie. $a_{12} = a_{21}$ etc. (missing in Gujurati)

***Proof*:**

$\mathrm{x'Ax} = a_{11}x_1^2 + a_{22}x_2^2 + \;....+ a_{nn}x_n^2 +$ terms like

$a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + a_{21}x_2x_1$ etc. (all cross-products)

$$\therefore \frac{\partial(\mathrm{x'Ax})}{\partial x_1} = 2a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + a_{21}x_2 \text{ etc.}$$

Since $a_{12} = a_{21}$, and collecting pairs of terms, we get

$$\frac{\partial(\mathrm{x'Ax})}{\partial x_1} = 2a_{11}x_1 + 2a_{12}x_2 + \ldots + 2a_{1n}x_n = 2\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

= 2 $\times$ (first row of A) $\times$ x

Similarly, $\frac{\partial(\mathrm{x'Ax})}{\partial x_2}$ = 2 $\times$ (second row of A) $\times$ x ,

etc for the other *x*'s.

Since $\frac{\partial(\mathrm{x'Ax})}{\partial \mathrm{x}} = \begin{pmatrix} \frac{\partial(\mathrm{x'Ax})}{\partial x_1} \\ \frac{\partial(\mathrm{x'Ax})}{\partial x_2} \\ \vdots \\ \frac{\partial(\mathrm{x'Ax})}{\partial x_n} \end{pmatrix}$

$$\frac{\partial(\mathrm{x'Ax})}{\partial \mathrm{x}} = 2\times\begin{pmatrix} \text{first row of A} \\ \text{second row of A} \\ \vdots \\ \mathrm{n}^{\mathrm{th}} \text{ row of A} \end{pmatrix}\times \mathrm{x} = 2\mathrm{Ax}$$

The symmetry of A is important. For a NOT symmetric matrix A,

$$\frac{\partial(\mathrm{x'Ax})}{\partial \mathrm{x}} = (\mathrm{A+A'})\mathrm{x}$$

Rule 3

$$\frac{\partial^2(\mathrm{x'Ax})}{\partial \mathrm{x}\partial \mathrm{x'}} = 2\mathrm{A}$$

where A is symmetric.

Pr oof:

For Rule 2 we can write the transpose as $\frac{\partial(\mathrm{x'Ax})}{\partial \mathrm{x'}} = 2\mathrm{x'A}$.

Hence,

$$\frac{\partial^2(\mathrm{x'Ax})}{\partial \mathrm{x}\partial \mathrm{x'}} = \frac{\partial}{\partial x}\left[\frac{\partial(\mathrm{x'Ax})}{\partial \mathrm{x'}}\right] = \frac{\partial}{\partial x}\left[2\mathrm{x'A}\right] = 2\frac{\partial}{\partial x}\left[\mathrm{x'A}\right] = 2\mathrm{A}.$$

where Rule 1* is used to obtain the last step.