Summary of Mathematical and Statistical Concepts for Advanced Macro II

Andrea Chiavari and Mario Giarda

April 13, 2020

1 Some Statistical Concepts

This section devotes some space to remind some statistical concepts. Our aim is not to make an extensive analysis but a reminder of some key concepts we will use throughout the course.

1. **Expected value.** We denote an expected value by $\mathbb{E}[X]$. Formally, it is the *probability weighted average* of a random variable X. Therefore, an expected value is an average but considering the probability distribution of the variable X. An expected value is different to the sample average as the latter considers that the weight of each observation is the same. Expectations are linear operators:

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

It is worth to mention that as a linear operator, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ is not necessarily true. Also, since, for instance, ln is a nonlinear function, $\mathbb{E}[\ln X] \neq \ln \mathbb{E}[X]$.

Finally, recall that there is the law of iterated expectations, that implies that

$$E[E[X]] = E[X]$$

which is important when we take expectations in time, like

$$E_t[E_{t+1}[X_{t+2}]] = E_t[X_{t+2}]$$

we must take the expectation with the least information we have, in that case, with the expectation up to period t.

2. **Variance.** The variance is a measure of dispersion. It represents the *squared average deviation* from the expected value:

$$var(X) = \mathbb{E}[X - \mathbb{E}[X]]^2.$$

Working with this equation it can be seen that

$$var(X) = \mathbb{E}[X^2] + \mathbb{E}[\mathbb{E}[X]]^2 - \mathbb{E}[2X\mathbb{E}[X]],$$

$$var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

The variance is a nonlinear operator, two properties are:

$$var(aX) = a^2 var(X),$$

and

$$var(X + a) = var(X),$$

We can finally define the **standard deviation**:

$$sd(X) = \sqrt{var(X)}$$

3. **Covariance and Correlation.** These concepts are useful to analyze the relation between two variables. The covariance is defined as

$$cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Distributing this expression:

$$cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y]$$

hence,

$$cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Notice that if $X = Y$ the covariance is equal to the variance. An additional property of the variance and covariance is:

$$var(aX \pm bY) = a^2 var(X) + b^2 var(Y) \pm 2cov(X, Y)$$

The covariance of two sums is given by:

$$cov\left(\sum_i a_i X_i, \sum_i b_i Y_i\right) = \sum_i \sum_j a_i b_j cov(X_i, Y_j)$$

The **correlation** between two variables is given by

$$corr(X,Y) = \frac{cov(X,Y)}{sd(X)sd(Y)}$$

4. Sample estimators. Let's remind the sample analog of these concepts. Given two random variables $\{X_i\}_{i=1}^T$ and $\{Y\}_{i=1}^T$ the sample moments are given by

   - Mean: $\mu_x = \frac{\sum_{i=1}^T X_i}{T}$
   - Variance: $\sigma_x^2 = \frac{\sum_{i=1}^T (X_i - \mu_x)^2}{T}$. Standard deviation $\sigma_x = \sqrt{\sigma_x^2}$
   - Covariance: $c\hat{o}v(X,Y) = \frac{\sum_{i=1}^T (X_i - \mu_x)(Y_i - \mu_y)}{T}$
   - Correlation: $\rho(X,Y) = \frac{c\hat{o}v(X,Y)}{\sigma_x \sigma_y}$

2 An Application of the Statistical Concepts to an AR(1) Process

A class of processes we will use a lot in the course is AR(1), which is an Autoregressive process of order one. In this section we study the properties and moments of an autoregressive process of order one.

Let's y_t follow an AR(1) process:

$$y_t = (1-\rho)\bar{y} + \rho y_{t-1} + \sigma_\varepsilon \varepsilon_t, \qquad \varepsilon_t \sim iid(0,1), \qquad \rho \in (0,1).$$

1. Compute the *unconditional* moments: mean $\mathbb{E}[y_t]$, variance $\mathbb{V}[y_t]$, and auto-covariance function $\gamma(\tau) \equiv Cov[y_t, y_{t-\tau}]$. Sketch the shape of $\gamma(\tau)$ as a function of τ. Is this process Markovian? Is it 2^{nd} order stationary? *(Hint: $\gamma(\tau)$ can be computed recursively. Also note that $\gamma(0) = \mathbb{V}[y_t]$).*

   Answer: The AR(1) prcess can be re-written in a more convenient way

$$\begin{aligned} y_t &= (1-\rho)\bar{y} + \rho y_{t-1} + \varepsilon_t \\ &= [(1-\rho)\bar{y} + \varepsilon_t] + \rho[(1-\rho)\bar{y} + \varepsilon_{t-1}] + \rho^2[(1-\rho)\bar{y} + \varepsilon_{t-2}] + \ldots \qquad (1) \\ &= \bar{y} + \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \ldots \end{aligned}$$

Taking expectations of (1) we get

$$\mathbb{E}[y_t] = \bar{y} + 0 + 0 + 0 + \dots$$

so that the mean of the AR(1) process is

$$\mu = \bar{y}.$$

The variance is

$$\begin{aligned}
\mathbb{V}[y_t] &= \mathbb{E}\big[(y_t - \mu)^2\big] \\
&= \mathbb{E}\big[(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \dots)^2\big] \\
&= (1 + \rho^2 + \rho^4 + ...)\sigma_\varepsilon^2 \\
&= \sigma_\varepsilon^2/(1 - \rho^2),
\end{aligned}$$

while the τ*th autocovariance is*

$$\begin{aligned}
\gamma(\tau) &= \mathbb{E}[(y_t - \mu)(y_{t-\tau} - \mu)] \\
&= \mathbb{E}[(\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \dots \\
&\qquad + \rho^\tau\varepsilon_{t-\tau} + \rho^{\tau+1}\varepsilon_{t-\tau-1} + \rho^{\tau+2}\varepsilon_{t-\tau-2})(\varepsilon_{t-\tau} + \rho\varepsilon_{t-\tau-1} + \rho^2\varepsilon_{t-\tau-2} + \dots)] \\
&= (\rho^\tau + \rho^{\tau+2} + \rho^{\tau+4} + \dots)\sigma_\varepsilon^2 \\
&= \rho^\tau(1 + \rho^2 + \rho^4 + \dots)\sigma_\varepsilon^2 \\
&= [\rho^\tau/(1 - \rho^2)]\sigma_\varepsilon^2 \\
&= \rho^\tau\gamma(0) \\
&= \rho^\tau\mathbb{V}[y_t].
\end{aligned}$$

The AR(1) process is Markovian, i.e.

$$Pr[y_t|y_{t-1}, y_{t-2}, \cdots, y_1, y_0] = Pr[y_t|y_{t-1}].$$

The AR(1) process is 2^{nd} *order stationary when* $|\rho| < 1$*. Note that the variance and the autocovariance have been derived under this assumption!*

2. Compute the *conditional* moments (**not needed for the course**): mean $\mathbb{E}_t[y_t]$, variance $\mathbb{V}_t[y_t]$, and auto-covariance function $Cov_t[y_t, y_{t-\tau}]$, where the notation means

$\mathbb{E}_t[\cdot] \equiv \mathbb{E}[\cdot|y^{t-1}]$ with $y^{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_1, y_0\}$, i.e., the information set at time t are all the observations up to that point. Analogously for $\mathbb{V}_t[\cdot] \equiv \mathbb{V}[\cdot|y^{t-1}]$ and $\gamma_t(\tau) \equiv Cov_t[\cdot] \equiv Cov[\cdot|y^{t-1}]$.

- Are the unconditional and conditional moments of the AR(1) process different? Why is it so?
- How do the unconditional and conditional moments change with the persistence parameter ρ?
- To compute the moments above, did you make any distributional assumptions on the ε process?

Answer: The conditional expectation for the AR(1) process is given by

$$\begin{aligned}\mathbb{E}_t[y_t] &= \mathbb{E}[y_t|y_{t-1}, y_{t-2}, \ldots, y_1, y_0] \\ &= \mathbb{E}[(1-\rho)\bar{y} + \rho y_{t-1} + \varepsilon_t|y_{t-1}, y_{t-2}, \ldots, y_1, y_0] \\ &= (1-\rho)\bar{y} + \rho y_{t-1}.\end{aligned}$$

The conditional variance is given by

$$\begin{aligned}\mathbb{V}_t[y_t] &= \mathbb{V}[y_t|y_{t-1}, y_{t-2}, \ldots, y_1, y_0] \\ &= \mathbb{V}[(1-\rho)\bar{y} + \rho y_{t-1} + \varepsilon_t|y_{t-1}, y_{t-2}, \ldots, y_1, y_0] \\ &= \sigma_\varepsilon^2,\end{aligned}$$

while the τth conditional autocovariance is

$$\begin{aligned}\gamma_t(\tau) &= Cov_t[(1-\rho)\bar{y} + y_{t-1} + \varepsilon_t, (1-\rho)\bar{y} + y_{t-\tau-1} + \varepsilon_{t-\tau}] \\ &= 0.\end{aligned}$$

The only moment that is affected by ρ is conditional expectation. Particularly, as $\rho \to 0$ the conditional expectation approaches the unconditional mean $\bar{y}$, whereas, if $\rho \to 1$ the conditional expectation goes to the previous observation y_{t-1}.

3 Properties of logarithms and log-linearization

We will, by using logarithms, linearize our models that are essentially non-linear. This is, we convert our systems of equations from non-linear to linear in logs. First, remind some properties of the logarithms:

$$\begin{aligned}
\ln(exp(X)) &= X \\
\ln(1) &= 0 \\
\ln(0) &\rightarrow -\infty \\
exp(0) &= 1 \\
exp(-\infty) &\rightarrow 0 \\
\ln(XY) &= \ln(X) + \ln(Y) \\
\ln(X/Y) &= \ln(X) - \ln(Y) \\
\ln(X+Y) &\neq \ln(X) + \ln(Y) \\
\ln(X^{\alpha}) &= \alpha \ln(X)
\end{aligned}$$

The last example is a good explanation of what the logarithm will do to our equations. In fact, what logs does is to transform a variable which is nonlinear x^{α} to a variable which is linear in logs $\alpha \log(x)$. As you can see, now the variable is not still an exponential (which is nonlinear) but it is linear, growing at a constant rate given by α.

As we said above, logarithms convert our systems from fully nonlinear to equations that are linear-in-logs. We do that by using a powerful device, the Taylor approximation.

A **Taylor approximation** of the function $f(x)$ at the n-th order around a point x_0 is given by:

$$\hat{f}(x)|_{x_0} \approx f(x_0) + \frac{f^1(x_0)}{1!}(x_t - x_0) + \frac{f^2(x_0)}{2!}(x_t - x_0)^2 + ... + \frac{f^n(x_0)}{n!}(x_t - x_0)^n \qquad (2)$$

where $f^i(x_0)$ is the i-th order derivative f the function $f(.)$ evaluated at x_0. The Taylor Theorem shows that if $n \rightarrow \infty$, $\tilde{f}(x_t) \rightarrow f(x_t)$.

A Taylor expansion will be useful for several reasons in the course. The main application we utilize is the approximation of our model around the steady state. In concrete, $f(x_t)$ will be a system of equations and we will approximate it around a point we call the steady state, which will be the x_0 in that case.[1] A second (a very practical) application of the

[1] We will see later in the course that a steady state is the solution of the system in which variables don't move anymore. We will come back to this later.

Taylor theorem is to prove that we can approximate a rate of growth with the difference of the log of the variable with its lagged value. This is,

$$\frac{x_t - x_{t-1}}{x_{t-1}} \approx \ln(x_t) - \ln(x_{t-1})$$

To prove this just take a first order approximation of the $\ln(x_t)$ around the point x_{t-1}:

$$\ln(x_t) \approx \ln(x_{t-1}) + \frac{1}{x_{t-1}}(x_t - x_{t-1}),$$

which rearranging we get the previous equation.[2]

In other applications of these concepts we can define variables like $\hat{x}_t = \ln(x_t) - \ln(x)$, that we read as the percent deviation of the variable x_t with respect to the value x, which typically will be a steady state.

Also, we use the same concept of "percent deviation with respect to a point x", to any variable we work with. In particular, we will express our variables as log-deviations from the steady state (which we will denote as $\hat{x}_t$), where the steady state is the point in which the variables don't move. A deviation with respect to the steady state is given by

$$\hat{x}_t = \frac{x_t - x}{x} \approx \ln(x_t) - \ln(x). \tag{3}$$

Finally, we will use logs plus the taylor expansions to approximate other several class of equations. This is, we want to loglinearize the system (or a equation) around a point (typically the steady state). To do that, throughout the course we will use a first-order approximation around the steady state.[3] Again, before doing that let's remind how a first-order Taylor approximation with more than one argument looks. Lets assume now $X_t = [x_{1t},\ x_{2t},\ ..., x_{nt}]'$, which is a $n \times 1$ vector of variables. A first order Taylor approximation of $f(X_t)$ around X_0 is:

$$\hat{f}(X_t) \approx f(X_0) + \sum_{i=1}^{n} f^1(X_0)(x_{it} - x_{i0}) \tag{4}$$

[2] Remind that $\frac{\partial \ln(x)}{\partial x} = \frac{1}{x}$.

[3] During the course we use first order approximations, but this is not restricted in general. You can use any approximation. Remind that the larger the order, the closer we are to the actual function. We must be smart in the order we choose because if your model is not sufficiently non-linear you do not need to use higher orders. Packages like Dynare allows us to use up to a third order approximation.

which is the sum of the approximations of the function $f(.)$ with respect to all its dimensions.

We will typically use log-linearization of any kind of equations with several variables. Let's see some examples:

1. A labor supply. An equation we will work a lot with is the labor supply

$$N_t^{\varphi} C_t^{\sigma} = W_t.$$

   This equation is easy to log-linearize as it is composed by multiples of different functions of the three variables. Let's take logs:

$$\varphi \ln(N_t) + \sigma \ln(C_t) = \ln(W_t) \tag{5}$$

   where we used the rule for exponentials and multiplications. The same equation can also be expressed and it holds in a point where we want to make the approximation like the steady state:

$$\varphi \ln(N) + \sigma \ln(C) = \ln(W) \tag{6}$$

   Substracting (6) to (5), we get the equation with variables "in percent differences with respect to the steady state":

$$\varphi(\ln(N_t) - \ln(N)) + \sigma(\ln(C_t) - \ln(C)) = \ln(W_t) - \ln(W) \tag{7}$$

   Using our notation (equation (3, this equation becomes

$$\varphi \hat{n}_t + \sigma \hat{c}_t = \hat{w}_t$$

2. A macroeconomic identity. We will also make use of the national accounts identity where we just consider consumption and government spending:

$$Y_t = C_t + G_t.$$

   This equation is a bit harder to log-linearize as $\ln(X + Y) \neq \ln(X) + \ln(Y)$. To implement the log-linearization, we first apply logs:

$$\ln(Y_t) = \ln(C_t + G_t).$$

Now let's take the right-hand-side of this equation and apply a first-order approximation around the steady state:

$$\ln(C_t + G_t) = \ln(C + G) + \frac{\partial \ln(C+G)}{\partial C}(C_t - C) + \frac{\partial \ln(C+G)}{\partial G}(G_t - G)$$

using $\frac{\partial \ln(C+G)}{\partial G} = \frac{1}{C+G}$:

$$\log(C_t + G_t) = \log(C + G) + \frac{1}{C+G}(C_t - C) + \frac{1}{C+G}(G_t - G)$$

multiplying by one in both terms of the Taylor approximation (C/C and G/G) and noticing that $Y = C + G$:

$$\ln(C_t + G_t) = \ln(Y) + \frac{C}{Y}\frac{(C_t - C)}{C} + \frac{G}{Y}\frac{(G_t - G)}{G}$$

Plugging this back into the original equation:

$$\ln(Y_t) = \ln(Y) + \frac{C}{Y}\frac{(C_t - C)}{C} + \frac{G}{Y}\frac{(G_t - G)}{G}$$

denoting $g_y = G/Y$ and using our definitions of $\hat{x}_t$ (equation(3)), we get

$$\hat{y}_t = (1 - g_y)\hat{c}_t + g_y\hat{g}_t$$

3. Growth rates (or interest rates). We can also show with a Taylor expansion that $\ln(1 + r_t) \approx r_t$. Take a first-order approximation of $\ln(1 + r_t)$ around $r_t = 0$ (usually these rates are close to zero):

$$\ln(1 + r_t) \approx \ln(1 + 0) + \frac{1}{1+0}(r_t - 0)$$
$$\ln(1 + r_t) \approx r_t,$$

as $\ln(1) = 0$ we get the expected result.

4 Summations

Throughout the course, we will use expected values of forward looking variables and also of the sum of a sequence of variables in the future. Let's remind what is a sum by taking a sequence of length T of a variable $\{X_{t+s}\}_{s=0}^{T}$. This sequence starts at the date t and lasts until $T+t$. We compute the sum of this sequence as

$$S' = \sum_{s=0}^{T} X_{t+s}$$

Some properties of summations:

- $\sum_{s=0}^{T} \alpha X_{t+s} = \alpha \sum_{s=0}^{T} X_{t+s}$ if α is a constant with respect to the index s.
- $\sum_{s=0}^{T} (X_{t+s} + Y_{t+s}) = \sum_{s=0}^{T} X_{t+s} + \sum_{s=0}^{T} Y_{t+s}$

We will frequently use a the sum of a geometric series that evolves like this (assuming $0 < \beta < 1$):

$$X_{t+1} = \beta X_t,$$

$$X_{t+2} = \beta^2 X_t,$$

$$...$$

$$X_{t+j} = \beta^j X_t.$$

If we sum this series:

$$S' = \sum_{s=0}^{T} X_{t+s} = \sum_{s=0}^{T} \beta^s X_t = X_t \sum_{s=0}^{T} \beta^s,$$

where the last equalyty follos from the fact that X_t does not depend on s. We solve for the sum S' just by solving for the sume of the β's, as follows. Take the summation S

$$S = \beta^0 + \beta^1 + \beta^2 + ...\beta^T$$

and multiply by β:

$$\beta S = \beta^1 + \beta^2 + \beta^3 + ...\beta^{T+1}$$

Substracting these two summations:

$$S(1-\beta) = 1 - \beta^{T+1}$$

Thus,

$$S = \frac{1-\beta^{T+1}}{1-\beta}$$

As $0 < \beta < 1$ if $T \to \infty$:

$$S = \frac{1}{1-\beta}.$$

We will use this last expression many times.

Sum of discounted sums. Another useful tool we will use during the course is the sum of expected values. Take the equation in differences given by

$$y_t = x_t + \beta \mathbb{E}_t y_{t+1}$$

In macroeconomics, equations like that appear constantly. In the course the Dynamic IS equation and the New-Keynesian Phillips curve are examples of that. We can solve those equations very easily by iterating forward. Notice that for sequences of y_t and x_t, the following holds

$$\begin{aligned} y_t &= x_t + \beta \mathbb{E}_t y_{t+1} \\ y_{t+1} &= x_{t+1} + \beta \mathbb{E}_{t+1} y_{t+2} \\ &... \\ y_{t+s} &= x_{t+s} + \beta \mathbb{E}_{t+s} y_{t+s+1} \\ &... \end{aligned}$$

Then, notice also that we can replace sequencially to get:

$$y_t = \mathbb{E}_t \sum_{s=0}^{\infty} \beta^s x_{t+s} \tag{8}$$

Which is a *discounted expected sum.*

Sometimes we will know the stochastic process of x_t. In particular, we will use AR(1). So let's assume x_t follows an AR(1):

$$x_{t+1} = \rho x_t + \epsilon_{t+1}$$

We know that as we know that, we can write *any future* value of x_{t+s} as:

$$x_{t+s} = \rho^s x_t + \sum_{k=0}^{s} \epsilon_{t+k}$$

However, the expectation of this process writes:

$$\mathbb{E}_t x_{t+s} = \rho^s x_t$$

as $\epsilon \sim IID(0, \sigma^2)$ which replacing back into (8):

$$y_t = \sum_{s=0}^{\infty} \beta^s \rho^s x_t = x_t \sum_{s=0}^{\infty} \beta^s \rho^s.$$

Next, by solving for the discounted sum, we get:

$$y_t = \frac{x_t}{1 - \beta\rho}.$$