Law of Iterated Expectations

Anmol Somanchi

The objective of this very brief note is to get an intuition of the Law of Iterated Expectations and importantly to not get confused or lost in the notations.

Case 1: Expectation of a random variable

Let x and y be two random variables and assume for now that x is continuous with support over $(-\infty, +\infty)$ and y is discrete with support over a finite set $Y = \{Y_1, Y_2, \ldots, Y_K\}$. This means y can only take on values $Y_1, \ldots, Y_K$ with some finite probability for each of these outcomes. The Law of Iterated Expectations (LIE) says:

$$\begin{aligned} \text{LIE}: \ \mathbb{E}(x) &= \mathbb{E}_y[\mathbb{E}(x \mid y)] \\ &= \sum_y \mathbb{E}(x \mid y)\, \mathbb{P}(y = Y) \end{aligned} \tag{1}$$

At the outset, useful to clarify notation. Note that the expectation with subscript y in $\mathbb{E}_y(\mathbb{E}(x|y))$ means we are taking an expectation over y of the conditional expectation inside that bracket. This means evaluating the conditional expectation at different values of y and then taking their average accounting for the probabilities of observing the different values of y. This notation of the LIE is sometimes also written as $\mathbb{E}(x) = \mathbb{E}_y[\mathbb{E}_x(x \mid y)]$ to be even more explicit that the outer expectation is over y and the inner expectation over x.

Example 1. (Analytical)

Let h be a continuous random variable denoting the height of M1 APE students. We are interested in estimating the mean height of M1 APE students. Suppose there is a discrete random $g \in \{M, F\}$ denoting gender. ***The LIE says we can estimate the unconditional expectation of height as the weighted average of the conditional expectations****, with the weights being given by the probability of the respective conditions holding (i.e. the probability of being male and female in this example).*

$$\mathbb{E}(h) = \mathbb{E}_g[\mathbb{E}(h \mid g)] = \sum_g \mathbb{E}(h \mid g = G)\, \mathbb{P}(g = G) = \mathbb{E}(h \mid g = F)\, \mathbb{P}(g = F) + \mathbb{E}(h \mid g = M)\, \mathbb{P}(g = M)$$

Note that between the first equality and the last, we expressed an unconditional expectation in terms of its conditional expectation. Also note that $\mathbb{P}(g = M)$ and $\mathbb{P}(g = F)$ are the probabilities of being male and female respectively in population of M1 APE students which represent the "weights" which need to used when averaging the conditional mean heights across males and females.

Example 2. (Numerical)

Consider the following data on heights (h) and gender (g) for some hypothetical population made up of only 3 individuals:

$$\begin{bmatrix} i & h & g \\ 1 & 10 & F \\ 2 & 15 & M \\ 3 & 20 & F \end{bmatrix}$$

As in the earlier example, we are interested in the expected height in the population, $\mathbb{E}(h)$*. We can first compute this expectation in the standard manner:*

$$\mathbb{E}(h) = \frac{10+15+20}{3} = 15$$

Now to convince ourselves that LIE works, we compute this using the LIE formula.

$$\begin{aligned}
\mathbb{E}(h) &= \mathbb{E}_g[\mathbb{E}(h \mid g)] \\
&= \sum_g \mathbb{E}(h \mid g = G)\,\mathbb{P}(G = g) \\
&= \mathbb{E}(h \mid g = M)\,\mathbb{P}(g = M) + \mathbb{E}(h \mid g = F)\,\mathbb{P}(g = F) \\
&= \frac{10+20}{2} \times \frac{2}{3} + 15 \times \frac{1}{3} \\
&= 15
\end{aligned}$$

Case 2: Expectation of product of random variables

The LIE formula in equation (1) focuses on the expectation of a single variable (x). But the LIE extends to cases where we are interested in the expectation of a product:

$$\mathbb{E}(yx) = \mathbb{E}_y[y\mathbb{E}(x \mid y)] = \sum_y y\ \mathbb{E}(x \mid y = Y)\,\mathbb{P}(y = Y) \tag{2}$$

Example 3. (Numerical)

Following up on the height (h) - gender (g) example, let us re-code the gender variable as a 0-1 dummy taking value 1 if the individual is female. The same data can now be expressed as:

$$\begin{bmatrix} i & h & g & h \times g \\ 1 & 10 & 1 & 10 \\ 2 & 15 & 0 & 0 \\ 3 & 20 & 1 & 20 \end{bmatrix}$$

We are now interested in estimating $\mathbb{E}(hg)$*. Let us first compute this in the usual manner as we compute a mean.*

$$\mathbb{E}(hg) = \frac{10+0+20}{3} = 10$$

Let us now compute it using the LIE formula.

$$\begin{aligned}
\mathbb{E}(hg) &= \mathbb{E}_g[g\mathbb{E}(h \mid g)] \\
&= \sum_g g\,\mathbb{E}(h \mid g = G)\,\mathbb{P}(g = G) \\
&= \Big(1 \times \mathbb{E}(h \mid g = 1) \times \mathbb{P}(g = 1)\Big) + \Big(0 \times \mathbb{E}(h \mid g = 0) \times \mathbb{P}(g = 0)\Big) \\
&= \left(1 \times \frac{10+20}{2} \times \frac{2}{3}\right) + \left(0 \times \frac{15}{1} \times \frac{1}{3}\right) \\
&= 10
\end{aligned}$$

Hopefully you are now convinced that the LIE works and understand the way the notation works.

Case 3+: LIE for more general/complex cases

The LIE notation and examples that we have dealt with so far were such that the conditioning variable has been discrete. But the LIE very much applies to the case of both variables being continuous. Let x and y both be continuous variables with support over $(-\infty, +\infty)$:

$$\mathbb{E}(x) = \mathbb{E}_y[\mathbb{E}(x \mid y)] = \int_{-\infty}^{+\infty} \mathbb{E}(x \mid y)\, f(y)\, dy$$

where $f(y)$ is the probability density function of y (as opposed to the probability mass function $\mathbb{P}(y = Y)$ in the case of discrete y).

In similar vein, the LIE can be extended to cases beyond what is covered in this note. For instance, let x, y, and z be 3 random variables:

$$\mathbb{E}(x \mid y) = \mathbb{E}_z[\mathbb{E}(x \mid y,\ z)]$$

That is, we can apply the LIE to a conditional expectation by additionally conditioning it on another random variable. One can likely think of other useful flavours of the LIE but the ones covered here are more than likely to suffice for this course.

We now proceed to prove that OLS is unbiased and then consistent using two different flavours of LIE covered in this note.

Proof OLS is unbiased using LIE

Consider a linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$. The OLS estimator is given by $\hat{\beta}_{OLS} = (\mathbf{X'X})^{-1}\mathbf{X'y}$. We assume that $A^3_{OLS} : \mathbb{E}(\mathbf{u} \mid \mathbf{X}) = 0$ is satisfied. To prove OLS is unbiased, we need to show that:

$$\mathbb{E}[\hat{\beta}_{OLS}] = \beta \implies \mathbb{E}[(\mathbf{X'X})^{-1}\mathbf{X'y}] = \beta$$

This is complicated by the fact that for our purposes in econometrics, the $\mathbf{X}$ matrix is *stochastic* (i.e. it is a random variable) given that if we take different samples, we will get different values of the explanatory/independent variables that make up the $\mathbf{X}$ matrix. Given this stochastic nature of $\mathbf{X}$, we cannot simply take it out of the expectation operator. This poses challenges. To proceed, we begin by evaluating the conditional expectation of the OLS estimator, conditioning on $\mathbf{X}$ (i.e. evaluating the expectation of the OLS estimator for a given $\mathbf{X}$):

$$\begin{aligned}
\mathbb{E}[\hat{\beta}_{OLS} \mid \mathbf{X}] &= \mathbb{E}[(\mathbf{X'X})^{-1}\mathbf{X'y} \mid \mathbf{X}] \\
&= \mathbb{E}[(\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X}\beta + \mathbf{u}) \mid \mathbf{X}] \qquad [\because \mathbf{y} = \mathbf{X}\beta + \mathbf{u}] \\
&= \mathbb{E}[(\mathbf{X'X})^{-1}\mathbf{X'X}\beta + (\mathbf{X'X})^{-1}\mathbf{X'u} \mid \mathbf{X}] \\
&= \mathbb{E}[\beta + (\mathbf{X'X})^{-1}\mathbf{X'u} \mid \mathbf{X}] \\
&= \beta + (\mathbf{X'X})^{-1}\mathbf{X'}\ \underbrace{\mathbb{E}[\mathbf{u} \mid \mathbf{X}]}_{=0\ \because\ A^3{OLS}} \\
&= \beta
\end{aligned}$$

So we have that $\mathbb{E}(\hat{\beta}_{OLS} \mid \mathbf{X}) = \beta$. This result tells us that for any possible values of $\mathbf{X}$ that we hold fixed, the conditional expectation of $\hat{\beta}_{OLS}$ equals β. By conditioning on $\mathbf{X}$, we are circumventing the troubles that its stochastic nature causes for the expectation operator. At this point we can call upon the LIE to prove our final result in the following manner:

$$\mathbb{E}(\hat{\beta}_{OLS}) = \underbrace{\mathbb{E}_{\mathbf{X}}[\mathbb{E}(\hat{\beta}_{OLS} \mid \mathbf{X})]}_{\because LIE} = \mathbb{E}_{\mathbf{X}}[\beta] = \beta$$

And so we have proved that if A^3_{OLS} holds, then $\mathbb{E}(\hat{\beta}_{OLS}) = \beta$, i.e. OLS is unbiased, relying on LIE. To be clear of what we did here, we first proved that the conditional expectation of the OLS estimator equals β. We then used LIE to express the unconditional expectation of $\hat{\beta}_{OLS}$ as an average over this conditional expectation we deduced. Finally, given that β is a constant, a weighted average of this constant over all possible values of $\mathbf{X}$ (weighted by the probabilities of observing the various values of $\mathbf{X}$) will equal the constant β itself.

Proof OLS is consistent using LIE

We will now apply the LIE applied to a product of random variable to prove consistency of OLS. We once again assume A^3_{OLS}. To prove consistency of OLS, we need to show that:

$$plim\,\hat{\beta}_{OLS} = \beta \implies plim\,(\mathbf{X'X})^{\mathbf{-1}}\mathbf{X'y} = \beta$$

To see this is indeed the case, we apply the LIE in the following manner:

$$\begin{aligned}
plim\,\hat{\beta}_{OLS} &= plim\,[(\mathbf{X'X})^{\mathbf{-1}}\mathbf{X'y}] \\
&= plim\,[(\mathbf{X'X})^{\mathbf{-1}}\mathbf{X'}(\mathbf{X}\beta + \mathbf{u})] \\
&= plim\,[(\mathbf{X'X})^{\mathbf{-1}}\mathbf{X'X}\beta + (\mathbf{X'X})^{\mathbf{-1}}\mathbf{X'u}] \\
&= plim\,[\beta + (\mathbf{X'X})^{\mathbf{-1}}\mathbf{X'u}] \\
&= \beta + plim\,(\mathbf{X'X})^{\mathbf{-1}}\,plim\,\mathbf{X'u} \\
&= \beta + plim\left(\frac{\mathbf{X'X}}{N}\right)^{-1} plim\left(\frac{\mathbf{X'u}}{N}\right) \\
&= \beta + plim\left(\frac{\mathbf{X'X}}{N}\right)^{-1} \mathbb{E}(\mathbf{X'u}) && [\because LLN] \\
&= \beta + plim\left(\frac{\mathbf{X'X}}{N}\right)^{-1} \mathbb{E}_{\mathbf{X}}[\mathbf{X'}\underbrace{\mathbb{E}(\mathbf{u}\,|\,\mathbf{X})}_{=0\,\because A^3_{OLS}}] && [\because LIE] \\
&= \beta
\end{aligned}$$

where LLN stands for the Law of Large Numbers, $plim\,\frac{1}{N}\sum_{i=1}^{N} x_i = \mathbb{E}(x_i)$, which says that the sample mean of a random variable x converges in probability to its population expectation. And so we have proved that $plim\,\hat{\beta}_{OLS} = \beta$, i.e. OLS is consistent. Note that we have done so applying a different flavour of LIE compared to the unbiasedness proof - in this case applying it to a product of random variables ($\mathbf{X'u}$).