Law of Iterated Expectations

Samiran Dutta

This note is meant to guide you through the derivation of the Law of Iterated Expectations. This result will prove to be extremely useful when deriving properties of the OLS estimators.

Let Y and X be two discrete random variables with some joint distribution $P(Y, X)$. The Law of Iterated Expectations (LIE) states:

$$E(Y) = E_X\Big\{E(Y|X)\Big\} \tag{1}$$

Where E_X denotes the expectation over the random variable X. Let the observations for Y be denoted by: Y_j where $j = \{1, 2, .., n\}$. Similarly, denote the observations of X by: X_i where $i = \{1, 2, .., m\}$. The number of observations for both RV's can be different ($m \neq n$) or the same ($m = n$). The LIE works in both cases.

The proof of equation (1) proceeds as follows. Using the formula for the conditional expectation $E(Y|X)$:

$$E_X\Big\{E(Y|X)\Big\} = E_X\Big\{\sum_{j=1}^{n} Y_j P(Y = Y_j|X)\Big\} \tag{2}$$

Note that the conditional expectation is done over all observations Y_j, conditional on a given X. Solving the outer expectation E_X is equivalent to taking a sum over all observations X_i, weighted by the probability $P(X = X_i)$. This gives:

$$E_X\Big\{E(Y|X)\Big\} = \sum_{i=1}^{m}\Big\{\sum_{j=1}^{n} Y_j P(Y = Y_j|X = X_i)\Big\}P(X = X_i) \tag{3}$$

Simply removing the brackets:

$$E_X\Big\{E(Y|X)\Big\} = \sum_{i=1}^{m}\sum_{j=1}^{n} Y_j P(Y = Y_j|X = X_i)P(X = X_i) \tag{4}$$

Using the formula for conditional probability, we have:

$$P(Y = Y_j|X = X_i) = \frac{P(Y = Y_j, X = X_i)}{P(X = X_i)} \implies P(Y = Y_j|X = X_i)P(X = X_i) = P(Y = Y_j, X = X_i)$$

Using this relation in equation (4) gives:

$$E_X\Big\{E(Y|X)\Big\} = \sum_{i=1}^{m}\sum_{j=1}^{n} Y_j P(Y = Y_j, X = X_i) \tag{5}$$

Since we are working with finite sums, the order of the summation can be changed:

$$E_X\Big\{E(Y|X)\Big\} = \sum_{j=1}^{n}\sum_{i=1}^{m} Y_j P(Y = Y_j, X = X_i) \tag{6}$$

To convince yourself that equations (5) and (6) are equivalent, go back to the simple example of a joint distribution (seen before in PBS2 $\mathcal{Q}$.2). Note that $m = n = 2$ in this simple example.

	Rain ($X = 0$)	No Rain ($X = 1$)	$P(Y)$
Long commute ($Y = 0$)	0.15	0.07	0.22
Short commute ($Y = 1$)	0.15	0.63	0.78
$P(X)$	0.3	0.7	1.00

Working through equation (5), we have two cases for X:

1. When $X = 0$

$$\sum_{j=1}^{2} Y_j P(Y = Y_j, X = 0) = 0 * 0.15 + 1 * 0.15 = 0.15$$

2. When $X = 1$

$$\sum_{j=1}^{2} Y_j P(Y = Y_j, X = 1) = 0 * 0.07 + 1 * 0.63 = 0.63$$

Summing over both cases of X:

$$\sum_{i=1}^{2}\sum_{j=1}^{2} Y_j P(Y = Y_j, X = X_i) = 0.15 + 0.63 = 0.78 \tag{7}$$

Working through equation (6) instead, we have two cases for Y:

1. When $Y = 0$

$$\sum_{i=1}^{2} 0 * P(Y = 0, X = X_i) = 0 * 07 + 0 * 0.63 = 0$$

2. When $Y = 1$

$$\sum_{i=1}^{2} 1 * P(Y = 1, X = X_i) = 1 * 07 + 1 * 0.63 = 0.78$$

Now summing over both cases of Y:

$$\sum_{j=1}^{2}\sum_{i=1}^{2} Y_j P(Y = Y_j, X = X_i) = 0 + 0.78 = 0.78 \tag{8}$$

Thus, equations (7) and (8) are equivalent. This implies that by generalization, equations (5) and (6) are also equivalent. Rewriting equation (6):

$$E_X\Big\{E(Y|X)\Big\} = \sum_{j=1}^{n}\sum_{i=1}^{m} Y_j P(Y = Y_j, X = X_i) \tag{9}$$

$$E_X\Big\{E(Y|X)\Big\} = \sum_{j=1}^{n} Y_j \sum_{i=1}^{m} P(Y = Y_j, X = X_i) \tag{10}$$

Where $\sum_{i=1}^{m} P(Y = Y_j, X = X_i)$ is the sum over all possibilites of X, given a value of Y_j. This is equivalent to the marginal probability of Y *i.e.*, $P(Y)$.

$$E_X\Big\{E(Y|X)\Big\} = \sum_{j=1}^{n} Y_j P(Y_j) \tag{11}$$

$$E_X\Big\{E(Y|X)\Big\} = E(Y) \tag{12}$$