More on Conditional Expectation

Francis J. DiTraglia

Econ 103

This handout contains some extra information on conditional expectation. I will not test you on anything that appears in this note but not in the lecture slides. However, if you want a deeper understanding of conditional expectation, reading through this material will be very helpful. Mathematically, the material that appears below is straightforward, using only tools and techniques that should be familiar from class. Conceptually, however, some of the discussion below goes beyond what we'll cover in class.

Evaluating a Prediction of Y Suppose that we are asked to submit a prediction m of a random variable Y. What value of m might we choose? A good prediction should be "close" to Y on average, but we have to define what we mean by "close." One possibility is to choose m to minimize *expected squared error*, namely

$$\text{choose } m \text{ to minimize } E\left[(Y-m)^2\right]$$

Expanding, and using the linearity of expectation and the shortcut formula,

$$
\begin{aligned}
E\left[(Y-m)^2\right] &= E\left[Y^2 - 2mY + m^2\right] \\
&= E[Y^2] - 2mE[Y] + m^2 \\
&= \left(\sigma_Y^2 + \mu_Y^2\right) - 2m\mu_Y + m^2
\end{aligned}
$$

Now, we need to minimize over m. Remember, σ_Y^2 and μ_Y are *constants* with respect to this minimization problem. The first order condition is:

$$2m - 2\mu_Y = 0$$

and hence our solution is $m = \mu_Y$, noting that the second order condition for a minimum is satsified. We have shown that the best forecast of Y is μ_Y, if our criterion of "best" is minimizing expected squared error.

Using X to Predict Y   Above we were asked to submit a *fixed* prediction of Y. Suppose now that we observe another random variable X and are allowed to make our forecast for Y depend on X. In this case we will choose a *function* m such that our prediction of Y is $m(X)$. Again, the question is what to choose for m? Suppose that we once again decide to minimize expected squared error:

$$\text{choose } m \text{ to minimize } E\left[\{Y - m(X)\}^2\right]$$

So what function m should we choose? To answer this, we re-write $\{Y - m(X)\}^2$ by adding and subtracting $E[Y|X]$ and expanding:

$$
\begin{aligned}
\{Y - m(X)\}^2 &= \{Y - E[Y|X] + E[Y|X] - m(X)\}^2 \\
&= \left[\{Y - E[Y|X]\} - \{m(X) - E[Y|X]\}\right]^2 \\
&= \{Y - E[Y|X]\}^2 + \{m(X) - E[Y|X]\}^2 - 2\{Y - E[Y|X]\}\{m(X) - E[Y|X]\}
\end{aligned}
$$

Now, we want to choose m to minimize the expected value of the preceding expression, namely

$$E\left[\{Y - E[Y|X]\}^2\right] + E\left[\{m(X) - E[Y|X]\}^2\right] - 2E\left[\{Y - E[Y|X]\}\{m(X) - E[Y|X]\}\right]$$

Notice that the first of the three terms doesn't involve m, so we can treat it as a constant in our minimization problem and just ignore it. Thus it suffices to find the function m that minimizes

$$E\left[\{m(X) - E[Y|X]\}^2\right] - 2E\left[\{Y - E[Y|X]\}\{m(X) - E[Y|X]\}\right]$$

The second term in this expression is quite complicated, but it turns out that one can show (using the law of iterated expectations) that it is always equal to zero. Thus, it suffices to find the function m that minimizes $E\left[\{m(X) - E[Y|X]\}^2\right]$ and the answer is clearly $m(X) = E[Y|X]$. Hence, we have shown that $E[Y|X]$ is the forecast of Y that minimizes expected squared error.

Deriving the Law of Iterated Expectations: Discrete RV

$$
\begin{aligned}
E_X\left[E_{Y|X}\left[Y|X\right]\right] &= E_X\left[\sum_y y\, p_{Y|X}(y|x)\right] = \sum_x \left(\sum_y y\, p_{Y|X}(y|x)\right) p_X(x) \\
&= \sum_x \sum_y y\, p_X(x)\, p_{Y|X}(y|x) = \sum_x \sum_y y\, p_{XY}(x,y) \\
&= \sum_y y \sum_x p_{XY}(x,y) = \sum_y y\, p_Y(y) = E[Y]
\end{aligned}
$$

The second step is the tricky one. Here what we're doing is calculating the expected value of the function $g(X) = \sum_y p_{Y|X}(y|X)$ with respect to the *marginal* pmf of X.

Deriving the Law of Iterated Expectations: Continuous RV   Same as above, but replace sums with itegrals.