What's this "E"-thing?

1 Introduction

The E-thing is a fundamental building block of econometrics. It is called the expectation operator and it is used to summarize the distribution of a random variable (or the joint distributions of multiple random variables). It is very flexible on one hand and very boring on the other. One thing to understand here is that the expectation operator ALWAYS works the same way. Therefore, we will start out with a rather non-formal exposition:

Let's assume that "stuff" is some random thing that has N realizations and $p(\text{stuff})$ gives the probabilities with which these realizations occur. Then the expectation of "stuff" equals

$$E(\text{stuff}) = \text{stuff}_1 \cdot p(\text{stuff}_1) + \text{stuff}_2 \cdot p(\text{stuff}_2) + ... + \text{stuff}_N \cdot p(\text{stuff}_N)$$

+

We see that an expectation has two ingredients

1. The thing we called "stuff" which is a function of one (ore more) random variable(s). For example, X, X^2, $(X - \mu_x)^2$, $X \cdot Y$, $(X - \mu_x)(Y - \mu_y)$
2. The probability distribution of "stuff". For example, p(X), P(X,Y), p(X|Y) etc.

With those two ingredients in hand expectations are always constructed in the same way. We calculate the function (1) for all possible realizations of the random variable(s) (i.e. x_1^2, x_2^2....) and weights (multiply) each one with the probability of its occurrence (2) (i.e. $x_1^2 \cdot p(x_1)$, $x_2^2 \cdot p(x_2)$....) and finally sums it all up. Done! Combining these two steps, for stuff $= X^2$ we get:

$$\begin{aligned} E(X^2) &= x_1^2 \cdot p(x_1) + x_2^2 \cdot p(x_2) + ... + x_N^2 \cdot p(x_N) \\ &= \sum_{i=1}^{N} x_i^2 \cdot p(x_i) \end{aligned}$$

Now the big secret lies in realizing that the expectations operator (E-thing) always works in the same way but has different name that depends on which function we use.

1.1 stuff $= X$: The Population Mean

E(X) merely uses the random variable X itself and is called the population mean.

$$\begin{aligned} E(X) &= x_1 \cdot p(x_1) + x_2 \cdot p(x_2) + ... + x_N \cdot p(x_N) \\ &= \sum_{i=1}^{N} x_i \cdot p(x_i) \\ &= \mu_x \end{aligned}$$

Note here that we can interchangeably refer to the population mean as E(X) or μ_x.

1.2 stuff $= (X - \mu_X)^2$: The Population Variance

Instead of merely using the random variable X itself the population variance uses the squared deviation from the population mean $(X - \mu_x)^2$ but the "mechanics" stay the same.

$$\begin{aligned} E\left[(X - \mu_X)^2\right] &= (x_1 - \mu_x)^2 \cdot p(x_1) + (x_2 - \mu_x)^2 \cdot p(x_2) + ... + (x_N - \mu_x)^2 \cdot p(x_N) \\ &= \sum_{i=1}^{N} (x_i - \mu_x)^2 \cdot p(x_i) \\ &= \sigma_x^2 \end{aligned}$$

1.3 stuff $= (X - \mu_X)(Y - \mu_Y)$: The Population Covariance

Nothing changes when we use more than one random variable. The function that we are weighting and summing is now a function of two (or more) random variables. For example, we can take the function $(X - \mu_x) \cdot (Y - \mu_y)$ and use the expectations operator. Naturally, since we have two random variables, we need to multiply with the joint probability of X and Y. Here we go.

$$\begin{aligned}
E[(X-\mu_x)(Y-mu_y)] &= (x_1-\mu_x)(y_1-\mu_y)\cdot p(x_1,y_1)+....+(x_1-\mu_x)(y_M-\mu_y)\cdot p(x_1,y_M) \\
&+ (x_2-\mu_x)(y_1-\mu_y)\cdot p(x_2,y_1)+....+(x_2-\mu_x)(y_M-\mu_y)\cdot p(x_2,y_M) \\
&+ ...... \\
&+ (x_N-\mu_x)(y_1-\mu_y)\cdot p(x_N,y_1)+....+(x_N-\mu_x)(y_M-\mu_y)\cdot p(x_N,y_M) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{M}(x_i-\mu_x)(y_j-\mu_y)\cdot p(x_i,y_j)=\sigma_{xy}
\end{aligned}$$

Note here that there are suddenly a lot of terms to add up. This comes from the fact that we have 2 random variables. Think of a table where each row contains a realization of X and each column a realization of Y. If there are N realizations of X and M realizations of Y there will be $N \cdot M$ terms to add up. It should now become clear that using the summation sign simplifies things a LOT!

1.4 The Conditional Population Mean

So far we have focused "stuff", the first part of the expectation (X, $X \cdot Y$, X^2 etc.) Now let's just take expectation of random variable X but change the second part, the probability distribution we are using. Let's use the conditional probability distribution of X given Y=5. That way we can calculate, for example, the expectation of random variable X if random variable Y=5. Again nothing really changes.

$$\begin{aligned}
E(X|Y=5) &= x_1\cdot p(x_1|y=5)+x_2\cdot p(x_2|y=5)+....+x_N\cdot p(x_N|y=5) \\
&= \sum_{i=1}^{N} x_i\cdot p(x_i|y=5) \\
&= \mu_{x|y=5}
\end{aligned}$$