Expectation and Vector Random Variables

Guy Lebanon

January 6, 2006

Conditional expectation comes in two flavors. The first a number $E(Y|X = x) \in \mathbb{R}$ and the second $E(Y|X)$ is a random variable itself (a function from Ω to $\mathbb{R}$). We cover these cases and then proceed to discuss covariance and correlation which are the analogue of variance for vector random variables.

Definition 1. *The conditional expectation of the RV $Y|X = x$ is*

$$E(Y|X=x) = \begin{cases} \int_{-\infty}^{\infty} y f_{Y|X=x}(y)\,dy & Y|X=x \text{ is a discrete RV} \\ \sum_{y\in\mathbb{R}} y p_{Y|X=x}(y) & Y|X=x \text{ is a continuous RV} \end{cases}$$

Intuitively, it represents the mean or average value of Y if we know that $X = x$. The above definition extends naturally to conditioning on multiple RVs e.g. $E(X_i|\{X_j = x_j : j \neq i\})$ (just use the appropriate conditional pdf or pmf in the definition above).

The conditional expectation $E(Y|X = x)$ is a real number, assuming that x is fixed ahead of time. If we look at it as a function of x i.e., $g(x) = E(Y|X = x)$, we obtain a function that assigns a real number $g(x)$ for every value $x \in \mathbb{R}$. This leads to the following definition. It is an elusive concept which require careful thinking.

Definition 2. *The conditional expectation $E(Y|X)$ is a RV $E(Y|X) : \Omega \to \mathbb{R}$ defined as follows:*

$$E(Y|X)(\omega) = E(Y|X = X(\omega)).$$

In other words, for every value $\omega \in \Omega$ we obtain a value $X(\omega) \in \mathbb{R}$ which we may denote as x and this in turn leads to the real number $E(Y|X = x)$. Note that $E(Y|X)$ is a RV that is a function of the RV X.

Since $E(Y|X)$ is a random variable, we can compute its expectation. The following theorem is sometimes useful

Theorem 1. *For any two RVs X, Y we have $E(E(Y|X)) = E(Y)$.*

Proof. We prove the result for the continuous case. The discrete case can be proven using an analogous proof.

$$\begin{aligned} E(E(Y|X)) &= \int_{-\infty}^{\infty} E(Y|X=x) f_X(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{Y|X=x}(y)\,dy f_X(x)\,dx \\ &= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dxdy = \int_{-\infty}^{\infty} y f_Y(y)\,dy = E(Y) \end{aligned}$$

where the first equality holds by the formula for expectation of a function of a random variable $E(g(X)) = \int g(x) f_X(x)dx$. □

Example: Suppose that X is uniform on $[0, 1]$ and that $Y|X = x$ is uniform on $[x, 1]$. What is $E(Y)$?

For each given value of x between 0 and 1, $E(Y|X = x)$ will equal the midpoint $(x + 1)/2$ of the interval $[x, 1]$. Therefore $E(Y|X) = (X + 1)/2$ and by the linearity of the expectation,

$$E(Y) = E(E(Y|X)) = (E(X) + 1)/2 = \left(\frac{1}{2} + 1\right)/2 = 3/4.$$

As for RVs, the expectation of a function of a vector RV $Y = g(\vec{\boldsymbol{X}})$ (Y is a one dimensional RV here) is

$$E(Y) = \begin{cases} \int_{\mathbb{R}^n} g(\vec{\boldsymbol{x}}) f_{\vec{\boldsymbol{X}}}(\vec{\boldsymbol{x}})\, d\vec{\boldsymbol{x}} & \vec{\boldsymbol{X}} \text{ is continuous} \\ \sum_{\vec{\boldsymbol{x}}\in\mathbb{R}^n} g(\vec{\boldsymbol{x}}) p_{\vec{\boldsymbol{X}}}(\vec{\boldsymbol{x}}) & \vec{\boldsymbol{X}} \text{ is discrete} \end{cases} \tag{1}$$

Important note: When you see expectation over several RVs, for example $E(X+Y)$, it is assumed that the expectation is taken with respect to (the integral, or sum) the joint distribution or all variables that appear in the argument.

We have that (if X, Y are discrete replace integrals with sum and pdf with pmf)

$$\begin{aligned} E(X+Y) &= \iint_{-\infty}^{\infty} (x+y) f_{X,Y}(x,y)\, dxdy = \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dydx + \int_{-\infty}^{\infty} y \int f_{X,Y}(x,y)\, dxdy \\ &= E(X) + E(Y). \end{aligned}$$

By induction we obtain the linearity property of expectation for a finite sum of RVs (not necessarily independent): $E(X_1 + \ldots + X_n) = \sum_{i=1}^n E(X_i)$.

If X, Y are independent, we have (again, for discrete RV, replace integrals with sums and pdf with pmf) for some functions g_1, g_2 (that could be the identity)

$$\begin{aligned} E(g_1(X)g_2(Y)) &= \iint_{-\infty}^{\infty} g_1(x)g_2(y) f_{X,Y}(x,y)\, dxdy = \iint_{-\infty}^{\infty} g_1(x)g_2(y) f_X(x) f_Y(y)\, dxdy \\ &= \left(\int_{-\infty}^{\infty} g_1(x) f_X(x) dx\right)\left(\int_{-\infty}^{\infty} g_2(y) f_Y(y) dy\right) = E(g_1(X))E(g_2(Y)). \end{aligned}$$

In particular, we have that for independent X, Y, $E(XY) = E(X)E(Y)$. Again the above result may be generalized by induction to a finite product of functions of RVs.

The covariance of X, Y is the generalization of the variance $E((X - E(X))^2)$.

Definition 3. *The covariance of two RV X, Y is $\mathsf{Cov}(X,Y) = E((X - E(X))(Y - E(Y)))$.*

An alternative expression that is sometimes more convenient is

$$\mathsf{Cov}(X,Y) = E(XY - XE(Y) - YE(X) + E(X)E(Y)) = E(XY) - 2E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y).$$

Recall that for independent X, Y $E(XY) = E(X)E(Y)$ and so $\mathsf{Cov}(X,Y) = 0$. However, the converse statement is false as there exists random variables that have covariance 0 but are dependent. Intuitively, the covariance measures the extent to which there exists a linear relationship between X, Y e.g. $X = \alpha Y + \beta$. If there is no linear relationship, the covariance is zero but the variables may still be dependent.

Definition 4. *For two random variables X, Y the correlation coefficient $\rho_{X,Y}$ is defined as*

$$\rho_{X,Y} = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}(X)}\sqrt{\mathsf{Var}(Y)}}.$$

Its virtue is in the fact that it is a normalized version of the covariance. While $\mathsf{Cov}(X,Y)$ can take on any real value, $-1 \leq \rho_{X,Y} \leq 1$ always with $|\rho_{X,Y}| = 1$ if there is a linear relationship between X, Y e.g. $X = \alpha Y + \beta$ and 0 if X, Y are independent.

To see that $-1 \leq \rho_{X,Y} \leq 1$ observe that since the expectation of a non-negative RV is non-negative,

$$0 \leq E\left(\left(\frac{X - E(X)}{\sqrt{\mathsf{Var}(X)}} \pm \frac{Y - E(Y)}{\sqrt{\mathsf{Var}(Y)}}\right)^2\right) = \frac{E((X - E(X))^2)}{\mathsf{Var}\, X} + \frac{E((Y - E(Y))^2)}{\mathsf{Var}\, Y} \pm 2\rho_{X,Y} = 2(1 \pm \rho_{X,Y})$$

which implies that $0 \leq 1 \pm \rho$ that is equivalent to $-1 \leq \rho_{X,Y} \leq 1$.