Statistics 3657 : Conditional Expectation

See file ch3-3-eg1.tex for some conditional distributions and conditional expectations for a specific bivariate pdf.

The conditional expectation of Y given X is defined to be a function of X, that is

$$\mathrm{E}(Y|X) = h(X) \tag{1}$$

where the function h is given by

$$\begin{aligned} h(x) &= \sum_y y P(Y = y|X = x) \text{ (discrete case)} \\ h(x) &= \int_{-\infty}^{\infty} y f_{Y|X=x}(y) dy \text{ (continuous case).} \end{aligned}$$

The function $h(x)$ is the usual formula for expectation, but using the conditional distribution of Y given $X = x$. Notice this can also be extended to the conditional expectation of $g(Y)$ given X by using the function

$$\mathrm{E}(g(Y)|X) = h(X) \tag{2}$$

where now we take

$$h(x) = \sum_y g(y) P(Y = y|X = x) \text{ (discrete case)} \tag{3}$$

$$h(x) = \int_{-\infty}^{\infty} g(y) f_{Y|X=x}(y) dy \text{ (continuous case).} \tag{4}$$

Again notice this is the usual formula for the expectation of $g(Y)$, a function of Y, but using the conditional distribution of Y given $X = x$.

This formulation above explicitly defines a function h, which maps numbers x (from an appropriate domain, but a domain that includes at least the support of the distribution of the r.v. X) into a value $h(x)$, that is a mapping or a function $x \mapsto h(x)$. Using this function h defined above we now consider the function of the random variable X, that is the r.v. $h(X)$. As before we can ask and answer various questions about the transformed r.v. $h(X)$. For example what is its distribution? What is its expectation? What is its variance, or any other moment? etc.

Since the function h used in (1) is obtained from the joint distribution of X, Y, the random variable $h(X)$ has some special properties and is called the conditional expectation of Y given X. For this we use the notation $\mathrm{E}(Y|X)$. Sometimes we write $\mathrm{E}(Y|X = x) = h(x)$. $\mathrm{E}(Y|X = x)$ is not the conditional expectation random variable.

We could ask about the conditional expectation of Y^2 given X. Since the joint distribution of X, Y^2 is different from the joint distribution of X, Y it not surprising that $\mathrm{E}(Y^2|X)$ is a different function of X than is the random variable $\mathrm{E}(Y|X)$. To find the random variable $\mathrm{E}(Y^2|X)$ we use the definition or equation (2) with the function $g(x) = x^2$.

Remark : The function h given by equations (3) or (4) is used only to specify the function to define $\mathrm{E}(g(Y)|X) = h(X)$. It allows us to deal with conditional expectation as a function of a random variable X. The notation $\mathrm{E}(g(Y)|X)$ is just a more convenient notation for this concept. This notion is useful in practical applications since as we saw earlier some specific models in actuarial science, biostatistics or other statistical models were written in the form of a two stage game, first a random variable X and then Y defined or constructed in terms of X. Some examples are given below.

We now study some special properties of $\mathrm{E}(Y|X)$.

Some models of a two stage game, or ones that can be written in this way, give rv's X for stage one and Y for stage 2, but do not give an explicit form for the distribution of (X, Y). However some of these models give a simple way of obtaining a formula for $\mathrm{E}(Y|X)$ and $\mathrm{E}(Y^2|X)$ (see some later examples). In connection with Theorem 1 below this allows us to calculate $E(Y)$ and $\mathrm{Var}(Y)$ without ever having to explicitly calculate either the marginal distribution of Y or the joint distribution of (X, Y).

See Theorem A, p137

Theorem 1 *$E(Y) = E(E(Y|X))$.*

This says the two random variables, Y and $E(Y|X)$ which is a random variable that is a function of X, both have the same expectation.

Theorem 1 is perhaps the most important property and use of conditional expectation.

Proof (in the discrete case); do in class.

Theorem 2 *Suppose that $a(X)$ and $b(X)$ are random variables that are functions of X but not of Y. Suppose also that $E(Y)$ is finite. Then*

$$E\left(a(X)Y + b(X)|X\right) = a(X)E(Y|X) + b(X) \ .$$

Theorem 2 says that conditional expectation has a linearity property similar to ordinary expectation. In particular the random coefficients $a(X)$ and $b(X)$ behave in the conditional expectation as if they are constants. This is sensible as conditional upon X, $a(X)$ and $b(X)$ are in fact fixed with respect to X.

Theorem 3 *Suppose that X and Y are independent random variables and that $E(Y)$ is a finite number. Then*

$$E\left(Y|X\right) = E(Y) \ .$$

Proof (in the discrete case); do in class.

Example [Random Sums of Random Variables]

N is a non negative integer valued random variable (eg Poisson or Binomial).

Suppose that X_i, $i = 1, 2, 3, \ldots$ are iid and independent of N. Consider the random sum

$$\begin{aligned} S &= \sum_{i=1}^{N} X_i \\ &= \begin{cases} \sum_{i=1}^{n} X_i & \text{if} \quad N = n \geq 1 \\ 0 & \text{if} \quad N = 0 \end{cases} \end{aligned}$$

Suppose that $\mu_N = \mathrm{E}(N)$ and $\mu_X = \mathrm{E}(X)$. We now calculate $\mathrm{E}(S|N)$. To do this we first find the corresponding function h used in the definition.

Given $N = n = 0$ we obtain $h(0) = 0$. Given $N = n \geq 1$ we obtain

$$\begin{aligned} h(n) &= \mathrm{E}\left(\sum_{i=1}^{N} X_i | N = n\right) \\ &= \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) \quad (*) \\ &= \sum_{i=1}^{n} \mathrm{E}(X_i) \\ &= \sum_{i=1}^{n} \mu_X \\ &= n\mu_X \end{aligned}$$

Notice the function $h(n) = n\mu_X$ also holds for $n = 0$. Thus we have

$$\mathrm{E}(S|N) = h(N) = N\mu_X \ .$$

We should investigate further why line (*) above holds. For now consider

$$\begin{aligned} h(1) &= \mathrm{E}\left(\sum_{i=1}^{N} X_i | N = 1\right) \\ &= \mathrm{E}\left(X_1 \mid N = 1\right) \end{aligned}$$

In this model the X_i's are independent of N. However S is not independent of N. By Theorem 3, since X_1 is independent of N we thus have

$$\mathrm{E}\left(X_1 \mid N = 1\right) = \mathrm{E}(X_1) = \mu_X \ .$$

Similarly, since $X_1 + X_2$ is independent of N (aside : the student should look at the material in the Section on independence and find the Theorem that tells us why this is so)

$$\begin{aligned} h(2) &= \mathrm{E}\left(\sum_{i=1}^{N} X_i | N = 2\right) \\ &= \mathrm{E}\left(X_1 + X_2 \mid N = 2\right) \\ &= \mathrm{E}\left(X_1 + X_2\right) \\ &= 2\mu_X \ . \end{aligned}$$

The student should now consider why this holds for $n > 2$.

Aside : This example is interesting since in general S is neither continuous nor discrete and hence we do not have any tools in this course to obtain the marginal distribution of S. The notion of conditional expectation however allows us to obtain $\mathrm{E}(Y)$ despite this problem. Below we see that we can also calculate the variance of S, that is $\mathrm{Var}(S)$, again despite the fact that we cannot explicitly calculate the marginal distribution of S.

To see that S might be neither continuous or discrete consider N Poisson and X_i normal. Then

$$P(S = 0) \geq P(N = 0) > 0 \ .$$

Continuous random variables cannot have this property.

For any interval (a, b) we also have

$$\begin{aligned} P(a < S < b) &= \sum_{i=1}^{\infty} P(a < S < b | N = i) P(N = i) \\ &\geq P(a < S < b | N = 1) P(N = 1) \\ &= P(a < X_1 < b) P(N = 1) \\ &> 0 \end{aligned}$$

This last line is a strict inequality. The student should this about why this is so.

However for discrete integer valued rv-s M, the student should check that there are some intervals (a, b) for which $P(a < M < b) = 0$. For example consider $(a, b) = \left(\frac{1}{4}, \frac{1}{2}\right)$. In general for any discrete r.v. there are intervals for which the probability of falling into this interval must be equal to 0.

Thus we see S does not satisfy necessary properties that must hold for all discrete r.v.'s and does not satisfy properties that must hold for all continuous r.v.s.

If however the X_i are all integer valued r.v.s then it will be the case that S is discrete.

End of Aside

Now suppose that $\sigma_N^2 = \mathrm{Var}(N)$ and $\sigma_X^2 = \mathrm{Var}(X)$. We now calculate $\mathrm{E}(S^2|N)$. To do this we first find the corresponding function h used in the definition.

$$h(n) = \mathrm{E}\left(\left(\sum_{i=1}^{N} X_i\right)^2 \mid N = n\right) \ . \tag{5}$$

Notice this will be a different function h than in the previous calculation. For the student you may wish to use a different name, for example h_1, to help distinguish this fact.

We can take advantage of variance, specifically conditional variance, which is the variance formula, but using conditional distributions. For $n > 0$

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i | N = n\right) = \mathrm{E}\left(\left(\sum_{i=1}^{N} X_i - \mathrm{E}(S|N = n)\right)^2 \mid N = n\right)$$

$$
\begin{aligned}
&= \mathrm{E}\left(\left(\sum_{i=1}^{n} X_i - n\mu\right)^2\right) \\
&= n\sigma_X^2
\end{aligned}
$$

Thus using linearity we find

$$
\begin{aligned}
h(n) &= \mathrm{E}\left(\left(\sum_{i=1}^{N} X_i\right)^2 |N = n\right) \\
&= n\sigma_X^2 + (n\mu)^2 \\
&= n\sigma_X^2 + n^2\mu_X^2 \ .
\end{aligned}
$$

This gives

$$
\mathrm{E}(S^2|N) = N\sigma_X^2 + N^2\mu_X^2 \ .
$$

End of Example

Remark 1 *We could also work with equation (5) is a direct manner. For $n \geq 1$ we have*

$$
\begin{aligned}
h(n) &= E\left(\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i X_j) \\
&= \sum_{i=1}^{n} E(X_i^2) + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} E(X_i)E(X_j) \\
&= n(\sigma_X^2 + \mu_X^2) + n(n-1)\mu_X^2 \\
&= n\sigma_X^2 + n^2\mu_X^2 \ .
\end{aligned}
$$

Remark 2 *In the example above notice that we did not have to find the marginal distribution of S.*

If the r.v.'s X are continuous, the random variable S is not continuous. It has $P(S = 0) = P(N = 0)$, which is greater than 0 for many interesting N, for example $N \sim$ Poisson. In this case S is neither continuous or discrete, but we can still obtain the first two moments above.

The random variable S above is called a compound random variable, and is very useful in actuarial science. A typical model has N as the random number of claims or accidents, with a payout of X_i for claim i. Such compound random variables are used to describe or model the payout for a given period of time.

Example Suppose that $X \sim f_X$, and $Y|X = x \sim f_{Y|X=x}$. This information allows us to determine the joint distribution of X, Y. This in turn will allow us to obtain formula such as the mean and variance of Y. However these could be obtained directly using conditional expectation. We now suppose that Y has a finite mean and variance (so that the conditional expectations are well defined).

Let

$$
h_1(x) = \int_R y f_{Y|X=x}(y) dy
$$

$$h_2(x) = \int_R y^2 f_{Y|X=x}(y)dy$$

Then

$$\begin{aligned} \mathrm{E}(Y|X) &= h_1(X) \\ \mathrm{E}(Y^2|X) &= h_2(X) \ . \end{aligned}$$

As a specific example, suppose that $X \sim \mathrm{Unif}(0,1)$ and $Y|X = x \sim \mathrm{Unif}(0,x)$.

Then

$$\begin{aligned} h_1(x) &= \int_0^x y\frac{1}{x}dy \\ &= \frac{x}{2} \\ h_2(x) &= \int_0^x y^2\frac{1}{x}dy \\ &= \frac{x^2}{3} \end{aligned}$$

Thus

$$\begin{aligned} \mathrm{E}(Y|X) &= \frac{X}{2} \\ \mathrm{E}(Y^2|X) &= \frac{X^2}{3} \ . \end{aligned}$$

Therefore

$$\begin{aligned} \mathrm{E}(Y) &= \frac{1}{2}\mathrm{E}(X) \\ &= \frac{1}{4} \\ \mathrm{E}(Y^2) &= \frac{1}{3}\mathrm{E}(X^2) \\ &= \frac{1}{9} \ . \end{aligned}$$

Thus $\mathrm{Var}(Y) = \frac{7}{144}$.

In this example the student could find the marginal distribution of Y, and then use this to find the first two moments of Y, that is $\mathrm{E}(Y^k)$, $k = 1, 2$.

Let us next calculate the covariance of X with Y, that is

$$\mathrm{Cov}(X,Y) = \mathrm{E}\left((X - \mu_X)(Y - \mu_Y)\right) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \ .$$

Notice that

$$\mathrm{E}(XY|X) = X\mathrm{E}(Y|X) = \frac{1}{2}X^2 \ .$$

Thus

$$\mathrm{E}(XY) = \mathrm{E}\left(X\mathrm{E}(Y|X)\right) = \frac{1}{2}\mathrm{E}(X^2) = \frac{1}{6}$$

and therefore

$$\mathrm{Cov}(X,Y) = \frac{1}{6} - \frac{1}{2}\frac{1}{4} = \frac{1}{24} \ .$$

The student can now calculate the correlation of X with Y.

End of Example

Remark 3 *Notice in the above example that we only have to calculate conditional means, so that either X could be discrete and Y continuous, or X continuous and Y discrete. This is something that we could not work with using other results and methods in this course.*

Example

Another type of example that occurs in actuarial science, reliability and other types of applications are mixture distributions. Here consider one type.

Suppose that Λ is a positive random variable with density f. Suppose that X given $\Lambda = \lambda$ is a random variable on the non negative integers (eg Poisson). What is the marginal distribution of X?

We consider a particular example.

Suppose that $\Lambda \sim \text{exponential}, \alpha$, and that $X|\Lambda = \lambda \sim \text{Poisson}, \lambda$. We could find the expectation of X by the method of conditional expectation. The student should show $\mathrm{E}(X|\Lambda = \lambda) = \lambda$ and then find that $\mathrm{E}(X|\Lambda) = \Lambda$.

In our problem we are trying to obtain something more than just $\mathrm{E}(X)$, in particular to find the pmf of X. Thus we need to find $P(X = k)$ for all non negative integers k, since that is the support of the distribution of X.

To put our calculation in the framework of conditional probability consider, for a fixed non negative integer k,

$$W = \begin{cases} 1 & \text{if } X = k \\ 0 & \text{otherwise} \end{cases}$$

Then since W is a Bernoulli random variable we obtain $\mathrm{E}(W) = Pr(W = 1) = P(X = k)$. We find the conditional expectation of W given Λ which then happens to give us $P(X = k)$. Using the definition we have $\mathrm{E}(W|\Lambda) = h(\Lambda)$ where

$$h(\lambda) = \frac{\lambda^k}{k!}e^{-\lambda} \ .$$

Thus

$$\begin{aligned}
P(X = k) &= \mathrm{E}(W) \\
&= \mathrm{E}(\mathrm{E}(W|\Lambda)) \\
&= \mathrm{E}(h(\Lambda)) \\
&= \int_0^\infty \frac{\lambda^k}{k!}e^{-\lambda}\alpha e^{-\lambda\alpha}d\lambda \\
&= \frac{\alpha}{k!}\int_0^\infty \lambda^k e^{-\lambda(1+\alpha)}d\lambda \\
&= \frac{\alpha}{k!}\frac{1}{(1+\alpha)^{k+1}}\Gamma(k+1) \\
&= \frac{\alpha}{1+\alpha}\left(1 - \frac{\alpha}{1+\alpha}\right)^k
\end{aligned}$$

Notice this is the geometric distribution, in the form of counting the number of failures before the first success, where the chance of success is $p = \frac{\alpha}{1+\alpha}$.

End of Example

Example

Consider X, Y with the bivariate distribution and parameters given in class. We found that

$$Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right) .$$

Therefore

$$\mathrm{E}(Y|X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) .$$

Aside : Notice that in this case we can easily calculate the distribution of $\mathrm{E}(Y|X)$ since this is a linear function of X. The methods in Chapter 2 will apply, and we can determine that $\mathrm{E}(Y|X)$ has a normal distribution. The student should verify this random variable has mean μ_Y and variance $\rho^2\sigma_Y^2$. Thus

$$\mathrm{E}(Y|X) \sim N(\mu_Y, \rho^2\sigma_Y^2)$$

Notice this is a normal distribution but with smaller variance, unless $\rho^2 = 1$, than the distribution of Y.

End of Aside

We can then obtain

$$\begin{aligned}
\mathrm{E}(XY) &= \mathrm{E}\left(X\mathrm{E}(Y|X)\right) \\
&= \mathrm{E}(X)\mu_Y + \mathrm{E}\left(X\rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X)\right) \\
&= \mu_X\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}\mathrm{Var}(X) \\
&= \mu_X\mu_Y + \rho\sigma_Y\sigma_X .
\end{aligned}$$

We then obtain $\mathrm{Cov}(X, Y) = \rho\sigma_Y\sigma_X$, which is consistent with our previous knowledge about bivariate normal distributions.

End of Example

Example

See example in the handout ch3-3-eg1.pdf

There we found

$$f_{Y|X=x}(y) = \begin{cases} \frac{1}{1-x^2} & \text{if } 0 < y < 1 - x^2 \\ 0 & \text{otherwise} \end{cases}$$

Thus for $-1 < x < 1$, $Y|X = x \sim \mathrm{Unif}(0, 1 - x^2)$. Consider the function

$$h(x) = \int_{-\infty}^{\infty} y f_{Y|X=x}(y)dy = \int_0^{1-x^2} y\frac{1}{1-x^2}dy = \frac{1-x^2}{2}$$

Thus

$$\mathrm{E}(Y|X) = \frac{1-X^2}{2}$$

Recall in the original handout we also found $\mathrm{E}(Y) = \frac{2}{7}$. We now calculate the expected value of $\mathrm{E}(Y|X)$ and verify that this expectation is in fact $\frac{2}{7}$.

$$\begin{aligned}
\mathrm{E}\left(\mathrm{E}(Y|X)\right) &= \mathrm{E}(h(X)) \text{ with } h \text{ being the function above} \\
&= \mathrm{E}\left(\frac{1-X^2}{2}\right) \\
&= \int_{-1}^{1} \frac{1-x^2}{2}\frac{15}{4}x^2(1-x^2)dx \\
&= \frac{15}{8}\int_{-1}^{1}(1-x^2)x^2(1-x^2)dx \\
&= \frac{15}{4}\int_{0}^{1}(1-x^2)x^2(1-x^2)dx \\
&= \frac{15}{4}\int_{0}^{1}x^2(1-2x^2+x^4)dx \\
&= \frac{15}{4}\int_{0}^{1}(x^2-2x^4+x^6)dx \\
&= \frac{15}{4}\left(\frac{1}{3}-\frac{2}{5}+\frac{1}{7}\right) \\
&= \frac{2}{7}
\end{aligned}$$

End of Example