Chapter 5, Section 11

Multivariate Distributions

Conditional Expectations

We can talk about the expectation of any distribution, including a conditional one.

Definition. If Y_1 and Y_2 are any two random variables, the **conditional expectation** of Y_1 given $Y_2 = y_2$ is:

$$E(Y_1 | Y_2 = y_2) = \int_{-\infty}^{\infty} y_1 f(y_1 | y_2) dy_1 \qquad \text{if } Y_1 \text{ is continuous,}$$

or

$$E(Y_1 | Y_2 = y_2) = \sum_{y_1} y_1 p(y_1 | y_2) \qquad \text{if } Y_1 \text{ is discrete.}$$

$E(Y_2 | Y_1 = y_1)$ is defined in a similar manner.

Notes:

1. $E(Y_1 \mid Y_2 = y_2)$ is only defined when y_2 is in the support of Y_2; that is, only when $f_2(y_2) \neq 0$ or $p_2(y_2) \neq 0$.

2. $E(Y_1 \mid Y_2 = y_2)$ is a function of y_2 alone, since the y_1 has been "integrated out" or "summed out" in its computation.

3. If we let $h(y_2) = E(Y_1 \mid Y_2 = y_2)$, its graph is the locus of the means of the conditional distributions of $(Y_1 \mid y_2)$. This is the idea behind regression, one of next semester's topics.

4. If $g(Y_1)$ is a function of the random variable, Y_1, then, just as in the univariate case, we can calculate $E(g(Y_1) \mid Y_2 = y_2)$ as

$$E\big(g(Y_1) \mid Y_2 = y_2\big) = \int_{-\infty}^{\infty} g(y_1) f(y_1 \mid y_2)\,dy_1 \quad \text{if } Y_1 \text{ is continuous,}$$

or

$$E\big(g(Y_1) \mid Y_2 = y_2\big) = \sum_{y_1} g(y_1) p(y_1 \mid y_2) \quad \text{if } Y_1 \text{ is discrete.}$$

3

Example. Consider again $(Y_1, Y_2) \sim f(y_1, y_2) = \begin{cases} 6(1-y_2), & 0 \le y_1 \le y_2 \le 1 \\ 0, & \text{elsewhere} \end{cases}$

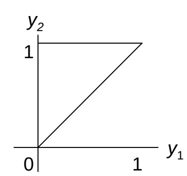
We have already calculated:

$$f_1(y_1) = 3(1-y_1)^2, \quad 0 \le y_1 \le 1; \quad = 0, \text{ elsewhere}$$

$$f_2(y_2) = 6y_2(1-y_2), \quad 0 \le y_2 \le 1; \quad = 0, \text{ elsewhere}$$

$$f(y_1 \mid y_2) = \frac{1}{y_2}, \quad 0 \le y_1 \le y_2; \quad = 0, \text{ elsewhere}$$

$$f(y_2 \mid y_1) = \frac{2(1-y_2)}{(1-y_1)^2}, \quad y_1 \le y_2 \le 1; \quad = 0, \text{ elsewhere}$$



$$E(Y_1) = \tfrac{1}{4}; \quad E(Y_2) = \tfrac{1}{2}; \quad V(Y_1) = \tfrac{3}{80}; \quad V(Y_2) = \tfrac{1}{20}; \quad \text{Cov}(Y_1, Y_2) = \tfrac{1}{40}.$$

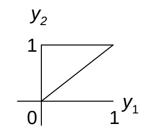
1. Find $E(Y_1 \mid Y_2 = y_2)$ and $V(Y_1 \mid Y_2 = y_2)$ for $0 < y_2 < 1$.

2. Find $E(Y_2 \mid Y_1 = y_1)$ and $V(Y_2 \mid Y_1 = y_1)$ for $0 \le y_1 < 1$.

4

1. Find $E(Y_1|Y_2=y_2)$ and $V(Y_1|Y_2=y_2)$ for $0<y_2<1$.

$f(y_1|y_2)=\dfrac{1}{y_2}$, $0\le y_1\le y_2$; $f(y_1|y_2)=0$, elsewhere.



$$E(Y_1|y_2) = \int_{-\infty}^{\infty} y_1 f(y_1|y_2)dy_1 = \int_0^{y_2} y_1\left(\frac{1}{y_2}\right)dy_1$$

$$= \frac{y_1^2}{2y_2}\Bigg|_{y_1=0}^{y_1=y_2} = \frac{y_2}{2}, \text{ for } 0<y_2<1.$$

$$E\left(Y_1^2|y_2\right) = \int_{-\infty}^{\infty} y_1^2 f(y_1|y_2)dy_1 = \int_0^{y_2} y_1^2\left(\frac{1}{y_2}\right)dy_1$$

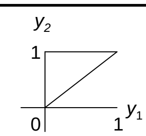
$$= \frac{y_1^3}{3y_2}\Bigg|_{y_1=0}^{y_1=y_2} = \frac{y_2^2}{3}, \text{ for } 0<y_2<1.$$

$$V(Y_1|y_2)=E\left(Y_1^2|y_2\right)-\left[E(Y_1|y_2)\right]^2 = \frac{y_2^2}{3}-\frac{y_2^2}{4} = \frac{y_2^2}{12}.$$

5

1. Find $E(Y_1|Y_2=y_2)$ and $V(Y_1|Y_2=y_2)$ for $0<y_2<1$.

$f(y_1|y_2)=\dfrac{1}{y_2}$, $0\le y_1\le y_2$; $f(y_1|y_2)=0$, elsewhere.



Since the conditional distribution of Y_1 given $Y_2 = y_2$ is uniform on the interval $[0, y_2]$, we can also obtain these values using what we know about the mean and variance of a continuous uniform distribution:

$$E(Y_1|Y_2=y_2) = \frac{0+y_2}{2} = \frac{y_2}{2} \quad \text{for } 0<y_2<1,$$

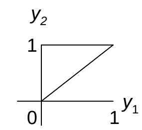
and

$$V(Y_1|Y_2=y_2) = \frac{(y_2-0)^2}{12} = \frac{y_2^2}{12} \quad \text{for } 0<y_2<1.$$

6

2. Find $E(Y_2 | Y_1 = y_1)$ and $V(Y_2 | Y_1 = y_1)$ for $0 \le y_1 < 1$.

$$f(y_2 | y_1) = \frac{2(1-y_2)}{(1-y_1)^2}, \quad y_1 \le y_2 \le 1; \quad f(y_2 | y_1) = 0 \text{, elsewhere.}$$



$$E(Y_2 | y_1) = \int_{-\infty}^{\infty} y_2 f(y_2 | y_1) dy_2 = \int_{y_1}^{1} \frac{y_2 \cdot 2(1-y_2)}{(1-y_1)^2} dy_2 = \int_{y_1}^{1} \frac{2y_2 - 2y_2^2}{(1-y_1)^2} dy_2$$

$$= \frac{y_2^2 - \frac{2}{3}y_2^3}{(1-y_1)^2} \Bigg|_{y_2 = y_1}^{y_2 = 1} = \frac{\left(1 - \frac{2}{3}\right) - \left(y_1^2 - \frac{2}{3}y_1^3\right)}{(1-y_1)^2} = \frac{1 - 3y_1^2 + 2y_1^3}{3(1-y_1)^2}$$

$$= \tfrac{1}{3}(2y_1 + 1) \quad \text{for} \quad 0 \le y_1 < 1.$$

Exercise: Show that $E\left(Y_2^2 | y_1\right) = \frac{1}{6}\left(3y_1^2 + 2y_1 + 1\right)$ and $V\left(Y_2 | y_1\right) = \frac{1}{18}\left(y_1^2 - 2y_1 + 1\right)$ for $0 \le y_1 < 1$.

Notice that $E(Y_1 | y_2) = \dfrac{y_2}{2}$ is a function of y_2 alone wherever it is defined (the interval $0 < y_2 < 1$). Thus we can calculate its expectation with respect to the distribution of Y_2.

$$E\left[E(Y_1 | Y_2)\right] = E\left(\tfrac{1}{2}Y_2\right) = \tfrac{1}{2}E(Y_2)$$

Since $E(Y_2) = \frac{1}{2}$, $E\left[E(Y_1 | Y_2)\right] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$.

Notice that this is $E(Y_1)$. It is not a coincidence.

Theorem. Let Y_1 and Y_2 be jointly distributed random variables. Then:

1. $E(Y_1) = E[\,E(Y_1 \mid Y_2)\,]$.

Expectation wrt the marginal distribution of Y_1

Expectation wrt the conditional distribution of Y_1 given Y_2

Expectation wrt the marginal distribution of Y_2

2. $V(Y_1) = E[V(Y_1 \mid Y_2)] + V[E(Y_1 \mid Y_2)]$.

Notes. (1) is called the **Double Expectation Theorem.**

(2) is called the **Conditional Variance Formula** or the **Double Expectation Theorem for Variance.**

Example. In the previous example, $E(Y_1 \mid Y_2) = \frac{1}{2} Y_2$,

$V(Y_1 \mid Y_2) = \frac{1}{12} Y_2^2$, $E(Y_2) = \frac{1}{2}$, and $V(Y_2) = \frac{1}{20}$, so

$$E[V(Y_1 \mid Y_2)] = E\left[\frac{1}{12} Y_2^2\right] = \frac{1}{12} E\left[Y_2^2\right] = \frac{1}{12}\left(V(Y_2) + [E(Y_2)]^2\right)$$

$$= \frac{1}{12}\left(\frac{1}{20} + \left(\frac{1}{2}\right)^2\right) = \frac{1}{12}\left(\frac{6}{20}\right) = \frac{1}{40}$$

and

$$V[E(Y_1 \mid Y_2)] = V\left(\frac{1}{2} Y_2\right) = \frac{1}{4} V(Y_2) = \frac{1}{4}\left(\frac{1}{20}\right) = \frac{1}{80}.$$

Therefore,

$$E[V(Y_1 \mid Y_2)] + V[E(Y_1 \mid Y_2)] = \frac{1}{40} + \frac{1}{80} = \frac{3}{80} = V(Y_1).$$

Example. A quality control program involves sampling $n = 10$ items per day and counting the number, Y, of defective items. If p denotes the probability of observing a defective item, then $Y \sim \text{bin}(10, p)$, assuming that a large number of items are produced each day. However, p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to 1/4.

1. Find the expected value of Y.

2. Find the variance of Y.

11

Solution. $Y \sim \text{bin}(10, p)$ and $p \sim U(0, 1/4)$,

1. $E(Y) = E[E(Y|p)]$ by Theorem 1. For any value of p, Y has a binomial distribution with parameters $n = 10$ and p, so

$$E(Y|p) = np = 10p.$$

Therefore,

$$E(Y) = E[E(Y|p)] = E(10p) = 10E(p).$$

Since $p \sim U(0, 1/4)$,

$$E(p) = \frac{\left(0 + \frac{1}{4}\right)}{2} = \frac{1}{8},$$

so

$$E(Y) = 10E(p) = 10\left(\tfrac{1}{8}\right) = \tfrac{5}{4}.$$

12

2. $V(Y) = E[V(Y|p)] + V[E(Y|p)]$. Since $Y \sim \text{bin}(10, p)$,

$V(Y|p) = 10p(1-p) = 10p - 10p^2$ and $E(Y|p) = 10p$.

Thus, $E[V(Y|p)] = E(10p - 10p^2) = 10E(p) - 10E(p^2)$.

Since $p \sim U(0, 1/4)$, $E(p) = \frac{1}{8}$, $V(p) = \frac{[(1/4)-0]^2}{12} = \frac{1}{192}$, and

$E(p^2) = V(p) + [E(p)]^2 = \frac{1}{192} + \frac{1}{64} = \frac{1}{48}$.

Thus, $E[V(Y|p)] = 10E(p) - 10E(p^2) = \frac{10}{8} - \frac{10}{48} = \frac{50}{48}$.

Also, $V[E(Y|p)] = V(10p) = 10^2 V(p) = \frac{100}{192} = \frac{25}{48}$,

So $V(Y) = E[V(Y|p)] + V[E(Y|p)] = \frac{50}{48} + \frac{25}{48} = \frac{75}{48} = \frac{25}{16}$.

Example. Expectation of a Random Number of Random Variables

Suppose that we can regard insurance claim amounts as independent random variables $X_1, X_2, \ldots$, having a common distribution, X. Let N denote the number of claims that will be filed next year. Assume that N is independent of the X_i; that is, assume that the number of claims filed next year is independent of their amounts. Find the expected total value of next year's claims.

Solution. Let S_N denote the total value of next year's claims, so that

$$S_N = \sum_{i=1}^{N} X_i.$$

Notice that in the summation, both the terms, X_i, and the number of terms, N, are (independent) random variables.

So, $S_N = \sum_{i=1}^{N} X_i$ and $E(S_N) = E[E(S_N | N)]$.

independence of N and the X_i

But, for each value, n, that N can assume,

$$E(S_N | N = n) = E\left(\sum_{i=1}^{N} X_i \,\middle|\, N = n\right) = E\left(\sum_{i=1}^{n} X_i \,\middle|\, N = n\right) = E\left(\sum_{i=1}^{n} X_i\right)$$

Then, since the X_i have common distribution X,

$$E(S_N | N = n) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} E(X) = n\,E(X).$$

Thus, $E(S_N \,|\, \mathrm{N}) = N\,E(X)$. Then, since $E(X)$ is a constant,

$$E(S_N) = E[E(S_N \,|\, \mathrm{N})]$$
$$= E[N\,E(X)] = E(X)\,E(N).$$

15

Thus, the Expected Total Value of the Claims is

$$E(S_N) = E(N)\,E(X),$$

the product of the Expected Number of Claims and the Expected Value of a Single Claim.

What is the variance of S_N?

We can use the Conditional Variance Formula

$$V(S_N) = E[V(S_N | N)] + V[E(S_N | N)]$$

after evaluating $V(S_N | N)$ using a procedure similar to the one used to evaluate $E(S_N | N)$.

16

$$V(S_N \mid N=n) = V\left(\sum_{i=1}^{N} X_i \,\middle|\, N=n\right) = V\left(\sum_{i=1}^{n} X_i \,\middle|\, N=n\right) = V\left(\sum_{i=1}^{n} X_i\right)$$

$$= \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} V(X) = n\,V(X).$$

independence of N and the X_i

Thus, $V(S_N \mid N) = N\,V(X)$. We have shown $E(S_N \mid N) = N\,E(X)$,

so
$$\begin{aligned} V(S_N) &= E[V(S_N \mid N)] + V[E(S_N \mid N)] \\ &= E[N\,V(X)] + V[N\,E(X)] \\ &= E(N)V(X) + V(N)[E(X)]^2 \end{aligned}$$

In the last step, we used the fact that $E(X)$ and $V(X)$ are constants.

17

Example. The number of defects per square yard, X, in a certain fabric is known to have a Poisson distribution with parameter λ, which itself is a random variable with density function $f(\lambda) = \begin{cases} e^{-\lambda}, & \lambda \ge 0 \\ 0, & \text{elsewhere} \end{cases}$

Thus, $\lambda \sim \text{Exp}(1)$.

1. Find $E(X)$ and $V(X)$.

2. Is it likely that X exceeds 9?

18

Solution.

1. From the given information, $(X \mid \lambda) \sim$ Poisson (λ), so

$$E(X \mid \lambda) = \lambda \quad \text{and} \quad V(X \mid \lambda) = \lambda .$$

Also, $\lambda \sim$ Exp(1), so

$$E(\lambda) = 1 \quad \text{and} \quad V(\lambda) = 1^2 = 1.$$

Thus,

$$E(X) = E\left[E(X \mid \lambda)\right] = E(\lambda) = 1$$

and

$$\begin{aligned} V(X) &= E\left[V(X \mid \lambda)\right] + V\left[E(X \mid \lambda)\right] \\ &= E(\lambda) + V(\lambda) \\ &= 1 + 1 \\ &= 2 \end{aligned}$$

19

2. Since $E(X) = 1$ and $\sigma_X = \sqrt{2}$,

$$9 = 1 + 8 = E(X) + 4\sqrt{2}\,\sigma_X$$

is $4\sqrt{2} = 5.657$ standard deviations from the mean of X,

so it is unlikely that $X > 9$.

20