

CHAPTER SIX

EXPECTATION

- 6.1 Expectation of a random variable*
 - 6.2 Expectation of a function of a random variable*
 - 6.3 Properties of expectation*
 - 6.4 Variance of a random variable and its Properties*
 - 6.5 Moments and moment generating function*
 - 6.6 Chebyshev's inequality*
 - 6.7 Covariance and correlation coefficient*
-

Introduction

Most probability distributions are characterized by their mean and variance. The mean of a random variable is referred to as expectation, provided the expected value converges to a constant. In addition to characterizing a given probability distribution, expected values are used in further statistical analyses. For instance, in the theory of estimation verifying whether a given estimator is an unbiased estimator or not we need to have the concept of expectation. For that matter, calculation of variance of a given random variable itself calls for expectation. Thus, we are not exaggerating if we say that almost all statistical analyses are, directly or indirectly, related to expectation.

The concept of expected value is used extensively in disciplines such as management where expected values play a key role in decision trees. Moreover, the expected monetary value (EMV) and other related variables required in decision making all call for the notion of expectation.

If we want to summarize a random variable by a single number, then this number should undoubtedly be its expected value. The expected value, also called the expectation or mean, gives the center in the sense of average value of the distribution of the random variable. If we allow a second number to describe the random variable, then we look at its variance, which is a measure of spread of the distribution of the random variable.

6.1 Expectation of a random variable

Definition (Expectation of a discrete random Variable) : Let X be a discrete random variable with possible values $x_1, x_2, x_3, \dots, x_n$. Let $P(x_i) = P(X = x_i) \ i = 1, 2, 3 \dots n$. Then the expected

value of X , denoted by $E(X)$, is defined as:

$$E(X) = \sum_{i=1}^{\infty} x_i P(x_i)$$

Note: If the series converges absolutely, this number is also called the mean of X .

Example 1: A manufacturer produces items such that 10% are defective and 90% are non-defective. If a defective item is produced the manufacturer loses \$1 and if a non-defective item is produced the manufacturer gets \$5 profit. Find the expected net profit of the manufacturer per item.

Solution

Let X be net profit per item. The values of X are -1 (loss) and 5 with corresponding probabilities 0.1 and 0.9 . Thus the expected profit $E(X) = -1(0.1) + 5(0.9) = \$4.40$

Example 2: A lot consists of 10 motors. It must either be totally rejected or is sold depending on the outcome of the following procedure. Two motors are chosen at random and inspected. If one or more are defective, the lot is rejected. Otherwise, it is accepted. Suppose that each motor costs \$75 and is sold for \$100. If the lot contains 1 defective motor, what is the manufacturer's expected profit?

Solution:

In the lot there are 9 non-defective and 1 defective motors.

$$\text{Hence } P(\text{one or more defective}) = 1 - P(\text{no defective}) = 1 - \frac{\binom{9}{2} \binom{1}{0}}{\binom{10}{2}} = 0.2$$

Thus if X is net profit, Expected net profit is $E(X) = -75(0.2) + 25(0.8) = \$5$

Exercise: A coin is tossed three times. Let X be the number of heads. Find the mean of X .

Solution:

The possible outcomes are $S = \{TTT, TTH, THT, HTT, HHT, THT, THH, HHH\}$

x	0	1	2	3	Total
$P(X = x)$	1/8	3/8	3/8	1/8	1
$xP(X = x)$	0	3/8	6/8	3/8	12/8

$$\text{Mean of } X = E(X) = \sum_{i=0}^3 x_i P(X = x_i) = 12/8$$

Definition (Expectation of a Continuous Random Variable): Let X be a continuous random variable with probability density function f . Then the expected value of X , denoted by $E(X)$, is

defined as:

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx$$

Example: Let X have the P.d.f $f(x) = \begin{cases} 2-2x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$. Find the expected value of X.

Solution:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x(2-2x)dx = \left[x^2 - \frac{2x^3}{3} \right]_0^1 = 1 - \frac{2}{3} = \frac{1}{3}$$

Exercise: Let $f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$, find expected value of X

$$\text{Solution: } E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 2xdx = \left[x^2 \right]_0^1 = 1$$

6.2 Expectation of a function of a random variable

Definition: Let X be a random variable, and g(x) be a function of the random variable X whose domain and range is the real line.

The expected value of g(x) is defined by

$$E(g(x)) = \begin{cases} \sum g(x_i)P(X = x_i) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x)f(x)dx, & \text{if } X \text{ is continuous} \end{cases}$$

Example 1: Let V be wind velocity and suppose that P.d.f of V is given as follows

$$f(v) = \begin{cases} 1/10, & 0 < v < 10 \\ 0 & \text{elsewhere} \end{cases}$$

The pressure W on the surface of an air plane wing is given by the relationship $W=0.003V^2$. Find E (W).

Solution:

$$E(W) = E(0.003V^2) = \int_0^{10} 0.003v^2 f(v)dv = \int_0^{10} 0.003v^2 \cdot 1/10 dv = 0.1$$

Example 2: Suppose that X has P.d.f $f(x) = \begin{cases} \frac{8}{x^3}, & x > 2 \\ 0, & \text{otherwise} \end{cases}$

Let $W=1/3 X$ find E (W)

Solutions:

$$E(W) = E(1/3X) = \int_2^{\infty} \frac{1}{3}x \frac{8}{x^3} dx = \frac{4}{3}$$

Exercise: Let X have the P.d.f $f(x) = \begin{cases} 2-2x, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$. Find $E(X^2)$.

Solution: $E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^1 x^2 (2-2x) dx = \frac{1}{6}$

6.2.1 Expectation of two-dimensional random variable

Definition: Let (X, Y) be a two-dimensional random variable and $Z = H(x, y)$ be real-valued function of (X, Y) . The expected value of the random variable Z is defined as follows:

a. If Z is discrete random variable with possible values $z_1, z_2, \dots$ and $P(z_i) = P(Z = z_i)$ then

$$E(Z) = \sum_{i=1}^{\infty} z_i P(z_i)$$

b. If Z is a continuous random variable with P.d.f f we have $E(Z) = \int_{-\infty}^{\infty} z f(z) dz$

Theorem: Let (X, Y) be a two-dimensional random variable and let $Z = H(x, y)$

$$E(Z) = \begin{cases} \sum_i \sum_j H(x_i, y_j) P(x_i, y_j), & \text{for discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) f(x, y) dx dy, & \text{for continuous} \end{cases}$$

Example 1: Assume that the current I and the resistance R are independent continuous random variables with the following P.d.f's

$$g(i) = \begin{cases} 2i, & 0 < i < 1 \\ 0, & \text{otherwise} \end{cases} \quad h(r) = \begin{cases} \frac{r^2}{9}, & 0 < r < 3 \\ 0, & \text{elsewhere} \end{cases}$$

The voltage $Z = IR$ with joint P.d.f of I and R

$$f_{IR}(i, r) = \begin{cases} \frac{2ir^2}{9}, & 0 < i < 1 \quad 0 < r < 3 \\ 0, & \text{elsewhere} \end{cases}$$

Find $E(Z)$

Solution: $E(Z) = \int_0^3 \int_0^1 ir f(i, r) di dr = \int_0^3 \int_0^1 ir \left(\frac{2ir^2}{9}\right) di dr = 3/2$

Example 2: Let (X, Y) be a two-dimensional random variable with P.d.f

$$f(x, y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Let $Z = XY$, evaluate $E(Z)$

Solution:

$$E(Z) = \int_0^1 \int_0^y 2xy \, dx \, dy = 1/4$$

Exercise: Let X have the P.d.f $f(x) = \begin{cases} \frac{x+2}{18}, & -2 < x < 4 \\ 0 & \text{Otherwise} \end{cases}$ Find a) $E(X)$ b) $E((X+2)^2)$

6.3 Properties of expectation

1. If C is any constant value, then $E(C) = C$
2. $E(CX) = C * E(X)$
3. If X and Y are any two random variables, then $E(X + Y) = E(X) + E(Y)$
4. If X and Y be two independent random variables, then $E(XY) = E(X) * E(Y)$
5. If a and b are any constant numbers, then $E(a + bX) = a + bE(X)$
6. $E(X) > 0$ if $X > 0$
7. $|E(X)| \leq E(|X|)$
8. $E(g_1(x)) \leq E(g_2(x))$ if $g_1(x) \leq g_2(x)$ for all x

6.4 Variance of a random variable and its Properties

Definition: Let X be a random variable. The variance of X , denoted by $V(X)$ is defined as:

$$V(X) = E[(X - E(X))^2]$$

That is, the mean value of the square of the deviations of X from its mean is called the variance of X or the variance of the distribution. The positive square root of $V(X)$ is called the standard deviation of X and it is denoted by $s.d(X)$.

It is worthwhile to observe that

$$\begin{aligned} V(X) &= E((X - E(X))^2) = E(X^2 - 2XE(X) - [E(X)]^2) \\ &= E(X^2) - 2E(X)E(X) - [E(X)]^2 \\ &= \underline{\underline{E(X^2) - [E(X)]^2}} \end{aligned}$$

or

$$\hookrightarrow \text{Var}(X) = \sum_i (x_i - \mu_X)^2 P(x_i) \text{ if } X \text{ is a discrete random variable}$$

$$\hookrightarrow \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \text{ if } X \text{ is a continuous random variable}$$

This expression affords an easier way of computing the variance of X .

The standard deviation of X may be interpreted as a measure of the dispersion of the points of the space relative to the mean value $E(X)$. We note that if the space contains only one point x for which $f(x) > 0$, then $s.d=0$.

Example1: Let X have the P.d.f $f(x) = \begin{cases} \frac{x+1}{2}, & -1 < x < 1 \\ 0 & \text{Otherwise} \end{cases}$ Find the variance of X .

Solution:

In order to find the variance of X we have to first find the expected value of X .

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-1}^1 x \frac{x+1}{2} dx = \int_{-1}^1 \left(\frac{x^2}{2} + \frac{x}{2}\right) dx = \frac{1}{3}$$

$$\text{Moreover, } E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{-1}^1 x^2 \frac{x+1}{2} dx = \int_{-1}^1 \left(\frac{x^3}{2} + \frac{x^2}{2}\right) dx = \left[\frac{x^4}{8} + \frac{x^3}{6}\right]_{-1}^1 = \frac{1}{3}$$

$$\text{Thus, } V(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \left(\frac{1}{3}\right)^2 = \frac{2}{9}$$

Example 2: Find the variance of the number of points, which come up on a single toss of a fair die.

Solution:

Let X : The number that come up.

$$\text{Thus } P(X = x) = \begin{cases} \frac{1}{6}, & x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{Otherwise} \end{cases}$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = \sum_{x=1}^6 x^2 P(X = x) = \frac{1}{6} \sum_{x=1}^6 x^2 = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}$$

$$E(X) = \sum_{x=1}^6 x P(X = x) = \frac{1}{6} \sum_{x=1}^6 x = \frac{1}{6} \left(\frac{6(6+1)}{2}\right) = \frac{21}{12}$$

$$\Leftrightarrow Var(X) = \frac{91}{6} - \left(\frac{21}{12}\right)^2 = 12.104$$

$$\Leftrightarrow s.d(X) = \sqrt{Var(X)} = \sqrt{12.104} = 3.48$$

Properties of Variance

Let X be a random variable:

☞ For any real number k , $Var(kX) = k^2 Var(X)$

☞ The expression $E[(X - k)^2]$ assumes its minimum value when $k = E(X)$

☞ If Y is another variable that is independent from X , then

$$\Rightarrow \text{Var}(X + Y) = \text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$$

☞ If X and Y are not independent, i.e., if they are dependent then

$$\Rightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) - 2(E(XY) - E(X)E(Y));$$

☞ The expression $E(XY) - E(X)E(Y)$ is called the Covariance of X and Y .

$$\Rightarrow \text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

☞ Note that: If X and Y are independent then $\text{Cov}(X, Y) = 0$

Exercise: A continuous random variable X has P.d.f $P(X = x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0 & \text{Otherwise} \end{cases}$. Find the

variance of X .

Solution:

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$\text{But, } E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{+\infty} x(2e^{-2x})dx = 2 \int_0^{+\infty} e^{-2x} dx = 2 \left[x \left(\frac{e^{-2x}}{-2} \right) - \left(\frac{e^{-2x}}{4} \right) \right]_0^{+\infty} = \frac{1}{2}$$

$$\text{Likewise, } E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_0^{+\infty} x^2 (2e^{-2x})dx = \frac{1}{2}$$

$$\Leftrightarrow \text{Var}(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

Now let's state a theorem, called Chebyshev's Inequality, which enables us to find upper (or lower) bounds for certain probabilities.

6.5 Chebyshev's Inequality

This inequality gives us a means of understanding how precisely the variance measures variability about the expected value of a random variable.

If we know the probability distribution of a R.V. X (either a pdf in the continuous case or the point probabilities in the discrete case), we may then compute $E(X)$ and $\text{Var}(X)$ if these exist. However, the converse is not true. That is, for a knowledge of $E(X)$ and $\text{Var}(X)$ we cannot reconstruct the probability distribution of X and hence cannot compute quantities such as $P(|X - E(X)| \leq C)$. Nonetheless, it turns out that although we cannot evaluate such probabilities (a knowledge of $E(X)$ and $\text{Var}(X)$), we can give a very useful upper (or lower) bound to such probabilities. This result is contained in what is known as Chebyshev's inequality.

Chebyshev's inequality: Let X be a random variable with $E(X) = \mu$ and let C be any real number. The, if $E(X - C)^2$ is finite and e is any positive number, we have

$$P[|X - C| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} E(X - C)^2$$

The following forms are equivalent to the above equation are immediate

a) By considering the complementary event we obtain

$$P[|X - C| < \varepsilon] > 1 - \frac{1}{\varepsilon^2} E(X - C)^2$$

b) Choosing $C = \mu$ we obtain

$$\begin{aligned} P[|X - \mu| \geq \varepsilon] &\leq \frac{1}{\varepsilon^2} E(X - \mu)^2 \\ &= P[|X - \mu| \geq \varepsilon] \leq \frac{Var(X)}{\varepsilon^2} \end{aligned}$$

c) Choosing $C = \mu$ and $\varepsilon = k\sigma$, where $\sigma^2 = Var(X) > 0$ we obtain

$$P[|X - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

6.6 Covariance and Correlation Coefficient

In the previous section of this unit we have stated that the covariance of two random variables is given by:

$$\begin{aligned} \sigma_{XY}^2 &= Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Now, if we divide the covariance to the product of the standard deviations of X and Y we get dimensionless (unit less) quantity called the correlation coefficient of X and Y.

Definition (Correlation Coefficient): Let X and Y be two random variables. The correlation coefficient of X and Y is denoted by ρ and is defined as:

$$\rho = \frac{Cov(X, Y)}{S.d(X) \times S.d(Y)} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - (E(X))^2} \sqrt{E(Y^2) - (E(Y))^2}}$$

Remarks:

- ✓ $-1 \leq \rho \leq 1$
- ✓ If $\rho = -1$ there is perfect inverse (negative) relationship or dependence between the two variables X and Y.
- ✓ If $-1 < \rho < 0$, then there is inverse relationship (correlation) between X and Y. The strength of the relationship will be determined based on the magnitude of ρ . If $-0.5 < \rho < 0$, then the relationship is inverse-weak. On the other hand, if $-1 < \rho < -0.5$, then the relationship is inverse-strong.

- ✓ If $\rho=0$, then the two variable X and Y are uncorrelated, no linear correlation or relationship. It has to be stressed that absence of correlation is not a guarantee for independence. But, the converse is true.
- ✓ If $0 < \rho < 1$, then there is direct (positive) relationship between X and Y. The strength of the relationship will be determined from the magnitude of ρ . If $0 < \rho < 0.5$, the relationship is direct-weak. If, on the other hand, $0.5 < \rho < 1$, the relationship is direct-strong.
- ✓ Finally if $\rho=1$ there is perfect direct relationship between the two random variables X and Y.

Example: Let the random variables X and Y have the joint P.d.f,

$$f(x, y) = \begin{cases} x + y & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases} \quad \text{Find the correlation coefficient of X and Y.}$$

Solution:

$$E(X) = \int_0^1 \int_0^1 x(x+y) dx dy = \int_0^1 \left(\frac{x^3}{3} + \frac{x^2 y}{2} \right)_0^1 dy = \int_0^1 \left(\frac{1}{3} + \frac{y}{2} \right) dy = \left[\frac{y}{3} + \frac{y^2}{4} \right]_0^1 = \frac{1}{3} + \frac{1}{4} = \frac{7}{12}$$

$$E(X^2) = \int_0^1 \int_0^1 x^2(x+y) dx dy = \int_0^1 \left(\frac{x^4}{4} + \frac{x^3 y}{3} \right)_0^1 dy = \int_0^1 \left(\frac{1}{4} + \frac{y}{3} \right) dy = \left[\frac{y}{4} + \frac{y^2}{6} \right]_0^1 = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

$$\text{Similarly } E(Y) = \int_0^1 \int_0^1 y(x+y) dx dy = \int_0^1 \left(\frac{yx^2}{2} + y^2 x \right)_0^1 dy = \int_0^1 \left(\frac{y}{2} + y^2 \right) dy = \left[\frac{y^2}{4} + \frac{y^3}{3} \right]_0^1 = \frac{1}{4} + \frac{1}{3} = \frac{7}{12}$$

$$E(Y^2) = \int_0^1 \int_0^1 y^2(x+y) dx dy = \int_0^1 \left(\frac{y^2 x^2}{2} + y^3 x \right)_0^1 dy = \int_0^1 \left(\frac{y^2}{2} + y^3 \right) dy = \left[\frac{y^3}{6} + \frac{y^4}{4} \right]_0^1 = \frac{1}{6} + \frac{1}{4} = \frac{5}{12}$$

$$E(XY) = \int_0^1 \int_0^1 xy(x+y) dx dy = \int_0^1 \left(\frac{yx^3}{3} + \frac{x^2 y^2}{2} \right)_0^1 dy = \int_0^1 \left(\frac{y}{3} + \frac{y^2}{2} \right) dy = \left[\frac{y^2}{6} + \frac{y^3}{6} \right]_0^1 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{5}{12} - \left(\frac{7}{12} \right)^2 = \frac{11}{144}$$

$$\Leftrightarrow S.d(X) = \sqrt{\text{Var}(X)} = \sqrt{\frac{11}{144}} = \frac{\sqrt{11}}{12}$$

$$\text{Var}(Y) = E(Y^2) - (E(Y))^2 = \frac{5}{12} - \left(\frac{7}{12} \right)^2 = \frac{11}{144}$$

$$\Leftrightarrow S.d(Y) = \sqrt{\text{Var}(Y)} = \sqrt{\frac{11}{144}} = \frac{\sqrt{11}}{12}$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{3} - \left(\frac{7}{12} \right) \left(\frac{7}{12} \right) = \frac{-1}{144}$$

$$\rho = \frac{\text{Cov}(X, Y)}{S.d(X) * S.d(Y)} = \frac{-1/144}{\sqrt{\frac{11}{12} * \frac{11}{12}}} = -\frac{1}{11}$$

The result seems to suggest that, we have a weak inverse relationship or correlation between X and Y.

Exercise: Let X and Y be two random variables with P.d.f

$$f(x, y) = \begin{cases} \frac{x(2x + y)}{210} & \text{if } 2 < x < 6 \text{ and } 0 < y < 5 \\ 0 & \text{elsewhere} \end{cases}$$

Find the correlation coefficient of X and Y.

Solution:

$$E(X) = \int_2^6 \int_0^5 \frac{x(2x + y)}{210} dx dy = \frac{268}{63}$$

$$E(X^2) = \int_2^6 \int_0^5 \frac{x^2(2x + y)}{210} dx dy = \frac{1220}{63}$$

$$\text{Similarly } E(Y) = \int_2^6 \int_0^5 \frac{y(2x + y)}{210} dx dy = \frac{170}{63}$$

$$E(Y^2) = \int_2^6 \int_0^5 \frac{y^2(2x + y)}{210} dx dy = \frac{1175}{126}$$

$$E(XY) = \int_2^6 \int_0^5 \frac{xy(2x + y)}{210} dx dy = \frac{80}{7}$$

$$\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{1220}{63} - \left(\frac{268}{63}\right)^2 = \frac{5036}{3969}$$

$$\Leftrightarrow S.d(X) = \sqrt{\text{Var}(X)} = \sqrt{\frac{5036}{3969}}$$

$$\text{Var}(Y) = E(Y^2) - (E(Y))^2 = \frac{1175}{126} - \left(\frac{170}{63}\right)^2 = \frac{16225}{7938}$$

$$\Leftrightarrow S.d(Y) = \sqrt{\text{Var}(Y)} = \sqrt{\frac{16225}{7938}}$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{80}{7} - \left(\frac{268}{63}\right)\left(\frac{170}{63}\right) = \frac{-200}{3969}$$

$$\rho = \frac{\text{Cov}(X, Y)}{S.d(X) \times S.d(Y)} = \frac{\frac{-200}{3969}}{\sqrt{\frac{5036}{3969}} \sqrt{\frac{16225}{7938}}} = -0.03129$$

Theorem: Suppose that $\rho^2=1$. The $Y = AX + B$, where A and B are constants. In words: If the correlation coefficient is ± 1 , then Y is the linear function of X (with probability 1).

Theorem: Suppose that X and Y are two random variables for which $Y = AX + B$, where A and B are constants. Then $\rho^2 = 1$. If $A > 0, \rho = +1$; If $A < 0, \rho = -1$

Proof:

Theorem: If ρ_{XY} is the correlation coefficient between X and Y , and if $V = AX + B$ and $W = CY + D$, where A, B, C , and D are constants, then:

$$\rho_{VW} = \left(\frac{AC}{|AC|} \right) \rho_{XY}; \quad A \neq 0 \text{ and } C \neq 0$$

Proof:

6.7 Conditional Expectation

Definition (Conditional Expectation):

a) If (X, Y) is a two-dimensional discrete random variable, we define the conditional expectation of X for given $Y = y_j$ as

$$E(X/Y = y_j) = \sum_{i=1}^{\infty} x_i P(x_i / y_j)$$

b) If (X, Y) is a two-dimensional continuous random variable, we define the conditional expectation of X for given $Y = y$ as

$$E(X/Y) = \int_{-\infty}^{\infty} x f_X(x/y) dx$$

Conditional expectation of Y for given X is defined analogously.

Theorem:

$$E(E(X/Y)) = E(X); \quad E(E(Y/X)) = E(Y)$$

Proof:

$$E(X/Y) = \int_{-\infty}^{\infty} x f_X(x/y) dx = \int_{-\infty}^{\infty} x \frac{f_{XY}(x, y)}{f_Y(y)} dx$$

Hence

$$\begin{aligned} E(E(X/Y)) &= \int_{-\infty}^{\infty} E(X/Y) f_Y(y) dy = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \frac{f_{XY}(x, y)}{f_Y(y)} dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{XY}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{XY}(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= E(X) \end{aligned}$$

Theorem: suppose that X and Y are independent random variable. Then

$$E(X/Y) = E(X); \quad E(Y/X) = E(Y)$$

Example 1: Let $f_{XY}(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & 0 < x < 1, 0 < y < 1 \\ 0 & \text{Otherwise} \end{cases}$

Find the conditional expectation of X given $Y = 1/2$

Solution:

$$f_X(x/y) = \begin{cases} \frac{2x + 4y}{1 + 4y} & 0 < x < 1 \\ 0 & \text{Otherwise} \end{cases}$$

$$f_X\left(x/Y = \frac{1}{2}\right) = \begin{cases} \frac{2}{3}(x + 1) & 0 < x < 1 \\ 0 & \text{Otherwise} \end{cases}$$

Hence

$$E\left(\frac{X}{Y} = \frac{1}{2}\right) = \int_0^1 x \frac{2}{3}(x + 1) dx = \frac{5}{9}$$

Example 2: Consider the following joint probability distribution.

Y \ X	0	1	2	Marginal
0	0	1/8	1/8	1/4
1	1/8	1/4	1/8	1/2
2	1/8	1/8	0	1/4
Marginal	1/4	1/2	1/4	1

Use the table to find the conditional probability functions and conditional expectations of (i) Y given $X=0$; (ii) Y given $X=1$; and (iii) Y given $X=2$.

Solution

The table has the marginal probability functions of X and Y added to it. You can see that once the joint probability mass function has been given in a table it is easy to write down the marginal (we just sum the rows and columns of the table). The conditional probability functions are obtained by dividing each column by the column total.

$$p_{Y|X}(y | x = 0) = \begin{cases} 0 & y = 0 \\ 1/2 & y = 1; \\ 1/2 & y = 2 \end{cases}$$

$$p_{Y|X}(y | x = 1) = \begin{cases} 1/4 & y = 0 \\ 1/2 & y = 1. \\ 1/4 & y = 2 \end{cases}$$

$$p_{Y|X}(y | x = 2) = \begin{cases} 1/2 & y = 0 \\ 1/2 & y = 1. \\ 0 & y = 2 \end{cases}$$

We can now use the conditional probability functions to find the conditional expectations:

$$E(Y|X = 0) = 0 * 0 + \left(\frac{1}{2}\right) * 1 + \left(\frac{1}{2}\right) * 2 = \frac{3}{2}$$

$$E(Y|X = 1) = \left(\frac{1}{4}\right) * 0 + \left(\frac{1}{2}\right) * 1 + \left(\frac{1}{4}\right) * 2 = 1$$

$$E(Y|X = 2) = \left(\frac{1}{2}\right) * 0 + \left(\frac{1}{2}\right) * 1 + 0 * 2 = \frac{1}{2}$$

Example 3:

I throw a fair die two times. Let

X be the sum of the scores on my two throws

Y be the number of sixes I obtain

Find the joint probability mass function of (X, Y). Find (i) the marginal probability function of Y; (ii) the conditional probability function of X given Y=1; (iii) the conditional probability function of X given Y=2; (iv) the conditional probability function of Y given X=7; (v) the conditional expectation of X given Y=1; (vi) F(8, 1); (vii) F(7.3, 1.4).

Solution

X \ Y	2	3	4	5	6	7	8	9	10	11	12
0	1/36	2/36	3/36	4/36	5/36	4/36	3/36	2/36	1/36	0	0
1	0	0	0	0	0	2/36	2/36	2/36	2/36	2/36	0
2	0	0	0	0	0	0	0	0	0	0	1/36

We now give the answers to the question with some hints on how to obtain them.

y	0	1	2
q(y)	25/36	10/36	1/36

....

6.8 Moments and moment generating function

It is difficult to overemphasize the importance of a moment generating function when it exists. This importance stems from the fact that the moment generating function of a given random variable is unique and completely determines the distribution of the random variable; thus if two random variables have the same moment generating function, they have the same distribution. In fact, it is not every distribution that has a moment generating function.

Definition (The r^{th} Moment about the origin): If X is a random variable, the r^{th} moment of the random variable X, usually denoted by μ'_r , is defined as:

$$\mu'_r = E(X^r) = \begin{cases} \sum_i x_i^r P(x_i) & \text{If } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} x^r f_X(x) dx & \text{If } X \text{ is a Continuous random variable} \end{cases}$$

Verbally, a moment is the expectations of the powers of the random variable that has the given distribution. Moments are defined for a random variable or a distribution.

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \mu'_2 - \mu_1'^2$$

$$\text{where } \mu_1' = E[X^1] = E(X) \text{ and } \mu_2' = E(X^2)$$

Definition (Central-Moments): The r^{th} moment about the mean of the random variable X , denoted by μ_r , is the expected value of $(X - \mu_X)^r$; symbolically,

$$\mu_r = E(X - \mu_X)^r = \begin{cases} \sum_i (x_i - \mu_X)^r P(x_i) & \text{If } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} (x - \mu_X)^r f_X(x) dx & \text{If } X \text{ is a Continuous random variable} \end{cases}$$

Remarks:

✚ The 1st central moment about is zero.

✚ The second central moment is the variance of the random variable.

✚ All odd moments of X about μ_X are 0 if the density function of X is symmetrical about μ_X , provided that such moment exist.

Example: Let X have the P.d.f $f(x) = \begin{cases} \frac{x+1}{2} & \text{if } -1 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$

Find a) the first moment about the origin b) the second central moment

Solution:

$$a) \mu_1' = E(X) = \int_{-1}^1 x f(x) dx = \int_{-1}^1 x \left(\frac{x+1}{2} \right) dx = \left[\frac{x^3}{6} + \frac{x^2}{4} \right]_{-1}^1 = \frac{5}{12} - \frac{1}{12} = \frac{4}{12} = \frac{1}{3}$$

$$b) \mu_2 = E((X - E(X)))^2 = E(X^2 - 2XE(X) + (E(X))^2) = E(X^2) - (E(X))^2$$

$$\text{But } E(X^2) = \int_{-1}^1 x^2 \left(\frac{x+1}{2} \right) dx = \left[\frac{x^4}{8} + \frac{x^3}{6} \right]_{-1}^1 = \frac{7}{24} - \frac{-1}{24} = \frac{1}{3}$$

$$\text{From part "a" above we have } E(X) = \frac{1}{3}$$

$$\Leftrightarrow \mu_2 = \frac{1}{3} - \left(\frac{1}{3} \right)^2 = \frac{3-1}{9} = \frac{2}{9}$$

Moment Generating Functions

Definition (Moment Generating Function): Let X be a random variable with P.d.f f . The expected value of e^{tx} is defined to be the moment generating function of X if in some interval $-h < t < h$; $h > 0$. The moment generating function, denoted by $M_X^{(t)}$, is

$$M_X^{(t)} = E(e^{tx}) = \begin{cases} \sum_i e^{tx_i} P(x_i) & \text{if } X \text{ is a discrete random variable.} \\ \int_{-\infty}^{\infty} e^{tx} f_X(x) dx & \text{if } X \text{ is a continuous random variable, and} \end{cases}$$

To explain why we refer to this function as “moment-generating” function, let us substitute for e^{tx} its Taylor series expansion, namely,

$$e^{tx} = 1 + \frac{t^1 x^1}{1!} + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \dots + \frac{t^r x^r}{r!} + \dots$$

For the discrete case, we thus get

$$\begin{aligned} M_X^{(t)} &= \sum_x e^{tx} P(x) = \sum_x \left(1 + \frac{t^1 x^1}{1!} + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \dots + \frac{t^r x^r}{r!} + \dots \right) P(x) \\ &= \sum_x P(x) + \sum_x x P(x) + \frac{t^2}{2!} \sum_x x^2 P(x) + \dots + \frac{t^r}{r!} \sum_x x^r P(x) + \dots \\ &= 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \dots + \mu'_r \frac{t^r}{r!} + \dots \end{aligned}$$

and it can be seen that in the Taylor series of moment generating function of X the coefficient of $\frac{t^r}{r!}$ is μ'_r , the r^{th} moment about the origin of the random variable X . In the continuous random variable the argument is the same.

Notes:

1. From the above expression one can show that;

$$\mu'_r = \frac{\partial^r}{\partial t^r} M_X^{(t)} \Big|_{t=0}$$

i.e., μ'_r is the r^{th} derivative of $M_X^{(t)}$ with respect to t evaluated at $t = 0$.

2. If a and b are constants, then

$$\begin{aligned} \text{I. } M_{X+a}^{(t)} &= e^{at} E(e^{tx}) \\ \text{II. } M_{bX}^{(t)} &= E(e^{btX}) \\ \text{III. } M_{\frac{X+a}{b}}^{(t)} &= e^{\frac{a}{b}t} E\left(e^{\frac{t}{b}X}\right) \end{aligned}$$

3. If X and Y are independent r.vs having moments generating functions $M_X^{(t)}$ and $M_Y^{(t)}$ respectively, then

$$M_{X+Y}^{(t)} = M_X^{(t)} * M_Y^{(t)}$$

Proof:

4. In general, If $X_1, X_2, X_3, \dots, X_n$ are independent random variables with moment generating function's $M_{X_i}^{(t)}; i = 1, 2, 3, \dots, n$ then the moment generating function of their sum, $M_Z^{(t)}$, where $Z = X_1 + X_2 + X_3 + \dots + X_n$, is given by

$$M_Z^{(t)} = M_{X_1}^{(t)} * M_{X_2}^{(t)} * \dots * M_{X_n}^{(t)} = \prod_{i=1}^n M_{X_i}^{(t)}$$

Proof: (Exercise)

5. **Uniqueness theorem:** Suppose that X and Y are random variables having moment generating function $M_X^{(t)}$ and $M_Y^{(t)}$ respectively. Then X and Y have the same probability distribution iff $M_X^{(t)} = M_Y^{(t)}$ identically.

Example: Find the moment generating function of the random variable X having P.d.f

$$f(x) = \begin{cases} \frac{1}{3} & \text{if } -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Solution:

$$M_X^{(t)} = E(e^{tx}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \int_{-1}^2 \frac{e^{tx}}{3} dx = \left[\frac{e^{tx}}{3t} \right]_{-1}^2 = \frac{e^{2t} - e^{-t}}{3t} \text{ provided } t \neq 0$$

If $t = 0$ we have $E(e^{0x}) = E(1) = 1$

Therefore,

$$M_X^{(t)} = \frac{e^{2t} - e^{-t}}{3t}$$

Remark: In more advanced courses, we would not work with the moment-generating functions because so many distributions do not have moment-generating functions.

Example: Consider the random variable X with P.d.f $f(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$

Find a) The second moment about origin b) the moment-generating function

$$a) \mu_2' = E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_0^1 x^2 (2x) dx = \left[\frac{x^4}{2} \right]_0^1 = \frac{1}{2}$$

$$b) M_X^{(t)} = E(e^{tx}) = \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \int_0^1 e^{tx} (2x) dx = \frac{2(te^t - e^t + 1)}{t^2} \dots \dots (\text{using integration by parts})$$