MATHEMATICAL EXPECTATION: DISCRETE RANDOM VARIABLES

TABLE OF CONTENTS

Content Developer

Chandra Goswami, Associate Professor, Department of Economics

Dyal Singh College, University of Delhi

Reference

Jay L. Devore: *Probability and Statistics for Engineering and the Sciences*, Cengage Learning, 8th edition [Chapter 3]

.

MATHEMATICAL EXPECTATION: DISCRETE RANDOM VARIABLES

***Learning objectives*:**

In this chapter you will learn how to obtain two main characteristics of the probability distribution of a discrete random variable. The mean of the distribution is the point on the number line where the distribution is centered and the variance is a measure of the spread of the distribution. You will learn how to derive these characteristics of distributions of discrete random variables. You will also learn how to apply the rules of mathematical expectation to functions of random variables as well as to sums of random variables.

Chapter Outline

1. Expected value of a discrete random variable
2. Expectation of a function of a discrete random variable
3. Rules of mathematical expectation
4. Variance of a discrete random variable
5. Variance of a function of a discrete random variable
6. Covariance and variance of sums of random variables
7. Parameters of the probability mass function

1 EXPECTED VALUE OF A DISCRETE RANDOM VARIABLE

Mathematical expectation of a random variable (rv) is a very important concept in probability theory. Graphical presentation of the probability distribution of a rv is valuable in reaching conclusions about the form of the distribution. However, mathematical expectation helps us to obtain summary measures of the characteristics of the probability distribution. Mathematical expectation of a rv is referred to simply as its expected value.

If X is a discrete rv with a set of possible values D and pmf p(x), then we can define the ***expected value of X***, denoted by E(X) or μ_x, as follows

***Definition* 1**

E(X) = $\sum_{x \in D} x.p(x)$

If D = [x_1, x_2, x_3,……x_n], then E(X) = μ_x, = x_1.p(x_1) + x_2. p(x_2) + x_3. p(x_3) +…….. + x_n. p(x_n)

If it is clear to which *X* the expected value refers, μ_x may be used instead of μ_x,

The expected value of a rv is called its mean value. We can interpret the expected value as the long-run average value that the rv takes over a large number of repeated trials of an experiment performed in identical and independent fashion. When the trials are conducted in this fashion then the outcome of any trial is independent of outcomes of the other trials.

Suppose a random experiment is repeated *N* times and the outcome X = x is observed in N_x number of these trials. For each possible value of X= x_1, x_2, x_3,…… in the set *D*, we can obtain $x\frac{N_x}{N}$. The sum of $x\frac{N_x}{N}$ over all possible values of *x* is then the average of values taken by the rv over all *N* trials, ie, $\sum_{x \in D} \frac{xN_x}{N}$. As the number of trials increases and *N* becomes infinitely large, the ratio $\frac{N_x}{N}$ tends to the probability of occurrence of *x*. In other words, $\frac{N_x}{N}$ → P(x) as $N \to \infty$. Thus, E(X) is the mean of the probability distribution of the random variable *X*.

To compute the population average value of *X* we need only the possible values of *X* along with their respective probabilities. The size of the population is immaterial as long as the pmf is given. The mean value of *X* is a weighted average of the possible values of *X,* where the weights are the probabilities of these values. The expected value μ may not coincide with any of the possible values of *X*. Note that the mean will coincide with the median if the distribution is symmetric.

The expected value of a rv *X* is also referred to as the first moment of *X* about the origin or simply the first moment. The quantity $E(X^n)$ is similarly the n^{th} moment of *X* where $n \geq 1$.

Example 1.1

If *X* is a Bernoulli rv with pmf

$$p(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & otherwise \end{cases}$$

Then, E(X) = 0.p(0) + 1.p(1) = 0(1- p) + 1(p) = p

Hence, expected value of the Bernoulli distribution is the probability that *X* takes the value 1. If a population consists of only 0's and 1's in the proportions (1-p) and p respectively, then the population mean is μ = p.

Example 1.2

If *X* has a pmf as follows

x	1	2	3	4
p(x)	0.002	0.146	0.588	0.264

Then μ = E(X) = 1(0.002) + 2(0.146) + 3(0.588) + 4(0.264)

= 0.002 + 0.292 + 1.764 + 1.056

= 3.114

Note that 3.114 is not one of the possible values of X since x = 1, 2, 3, 4. Also population size is not given nor is it required.

Example 1.3

Let *X* = number of trials till the first success is observed, and *p* = the probability of success. The pmf of *X* is

$$p(x) = \begin{cases} p(1-p)^{x-1} & x = 1,2,3...... \\ 0 & otherwise \end{cases}$$

The mean of *X* is E(X) obtained as follows

E(X) = $\sum_{x \in D} x.p(x)$ = $\sum_{x=1}^{\infty} x(1-p)^{x-1}$ = $p\sum_{x=1}^{\infty} x(1-p)^{x-1}$

Now $\frac{d}{dp}(1-p)^x = -x(1-p)^{x-1}$

Substituting we get

E(X) = $p\sum_{x=1}^{\infty}\left[-\frac{d}{dp}(1-p)^x\right]$

= $p\frac{d}{dp}\left[(-1)\sum_{x=1}^{\infty}(1-p)^x\right]$

Since $\left[\sum_{x=0}^{\infty}(1-p)^x\right] = \frac{1}{1-(1-p)}$,

therefore $\sum_{x=1}^{\infty}(1-p)^x = \frac{1}{p} \; -1$,

and $(-)\sum_{x=1}^{\infty}(1-p)^x = 1-\frac{1}{p}$

[This is a convergent geometric series as p < 1 and (1-p) < 1]

$\therefore$ E(X) = $p\frac{d}{dp}\left(1-\frac{1}{p}\right) = p\left[0+\frac{1}{p^2}\right] = \frac{1}{p}$

= $p\left[0+\frac{1}{p^2}\right] = \frac{1}{p}$

Alternately

$\sum_{x=1}^{\infty} xp(1-x)^{x-1} = \sum_{x=1}^{\infty} xpq^{x-1}$ = $pq^0 + 2pq^1 + 3pq^2 + 4pq^3 + 5pq^4 + 6pq^5 +$…………..

= $p(1 + 2q + 3q^2 + 4q^3 + 5q^4 + 6q^5 +$ …………)

Using series expansion $\frac{1}{(1-x)^2}$ = $1 + 2x + 3x^2 + 4x^3 + 5x^4 +$ ……… , where x = q

we get E(X) = $p\frac{1}{(1-q)^2} = p\frac{1}{p^2} = \frac{1}{p}$ since (1-q) = p

If p = 0.5 then E(X) = $\frac{1}{0.5} = 2$. ie, a success will be observed after 2 trials on the average.

If p = 0.2 then E(X) = 5, ie, on the average a success will be observed after 5 trials.

As *p* approaches 1, there will be few failures before a success is observed. As *p* approaches 0, we expect many failures before a success is observed.

It is possible to have a probability distribution where larger values of the rv *X* have higher probabilities. Such distributions with "heavy tails" may result in a mean value that is not finite.

Example 1.4

$$p(x) = \begin{cases} \frac{k}{x^2} & x = 1,2,3,.. \\ 0 & otherwise \end{cases}$$ where k is chosen so that $\sum_{x=1}^{\infty} \frac{k}{x^2} = 1$

E(X) = μ = $\sum_{x=1}^{\infty} x \frac{k}{x^2} = k \sum_{x=1}^{\infty} \frac{1}{x}$

E(X) is not finite as the harmonic series $\sum_{x=1}^{\infty} \frac{1}{x}$ is equal to infinity and p(x) does not decrease sufficiently fast as *x* increases.

Exercise 1

Find the expected value of the rv *X* having the pmf

$$p(x) = \begin{cases} \frac{|x-2|}{7} & x = -1,0,1,3 \\ 0 & otherwise \end{cases}$$

Solution

E(X) = $\sum x \frac{|x-2|}{7} = (-1)\left(\frac{3}{7}\right) + 0 + (1)\left(\frac{1}{7}\right) + (3)\left(\frac{1}{7}\right) = \frac{1}{7}$

$$= (-1)\left(\frac{3}{7}\right) + 0 + (1)\left(\frac{1}{7}\right) + (3)\left(\frac{1}{7}\right)$$

$$= \frac{1}{7}$$

Exercise 2

Find E(X) where *X* is the outcome when we roll a fair die.

Solution

For a fair die each face of the die is an equally likely outcome.

$\therefore$ p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = $\frac{1}{6}$

where, x = 1, 2, 3, 4, 5, 6 denote the number of dots on the six faces of the die or outcome when we roll the die.

$$E(X) = (1)\frac{1}{6} + (2)\frac{1}{6} + (3)\frac{1}{6} + (4)\frac{1}{6} + (5)\frac{1}{6} + (6)\frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3.5$$

Note that we can never observe the outcome to be 3.5 as *X* can only take the integer values 1, 2, 3, 4, 5, 6.

E(X) is simply the average value of *X* if we roll a fair die a large number of times.

Exercise 3

An investor is considering three strategies for a $1,000 investment. The estimated probable returns are:

Strategy 1: A profit of $10,000 with probability 0.15 and a loss of $1,000 with probability 0.85

Strategy 2: A profit of $1,000 with probability 0.50, a profit of $500 with probability 0.30 and a loss of $500 with probability 0.20

Strategy 3: A certain profit of $400.

Which strategy has the highest expected profit?

Solution

Let X_j = returns from investment in j^{th} strategy, where j = 1, 2, 3

Strategy 1: $E(X_1) = (0.15)(10000) + (0.85)(-1000) = 1500 – 850 = \$650$

Strategy 2: $E(X_2) = (0.50)(1000) + (0.30)(500) + (0.20)(-500) = 500 + 150 – 100 = \$550$

Strategy 3: $E(X_3) = (1)(400) = \$400$

Since $E(X_1) > E(X_2) > E(X_3)$ therefore strategy 1 is most profitable.

Exercise 4

A group of 500 persons participate in a lottery with a first prize of ₹1000, two second prizes of ₹500 each and five third prizes of ₹100 each. If the lottery is equitable, so that each player's expectation is zero, then what is the fair price of the lottery ticket?

Solution

Since there are 500 players, the probability of winning the first prize is $\frac{1}{500}$, the probability of winning the second prize is $\frac{2}{500}$, and that of winning the third prize is $\frac{5}{500}$. Then,

$$E(X) = (1000)\frac{1}{500} + (500)\frac{2}{500} + (100)\frac{5}{500} + (0)\frac{492}{500} = 2 + 2 + 1 = 5$$

The lottery will be equitable if ticket price is ₹5, in which case expected earnings are ₹5 and the cost to player is ₹5 so that each player's expectation is E(X) – 5 = 0

2 EXPECTATION OF A FUNCTION OF A DISCRETE RANDOM VARIABLE

Let h(X) be a function of a discrete rv *X*. Since *X* is a rv, h(X) is also a rv. The pmf of h(X) is the same as the pmf of *X*. Let *Y* denote the rv h(X). Let *D** denote all possible values of *Y* and *D* denote all possible values of *X*.

Proposition 1

$$E(Y) = E[h(X)] = \sum_{y \in D^*} y.p(y) = \sum_{x \in D} h(x).p(x) \text{ provided that } \sum_{x \in D} |h(x)|.p(x) < \infty$$

The expected value of *Y* or $\mu_{h(X)}$ is thus a weighted average of possible values of h(x) and the weights are the corresponding probabilities. E[h(X)] is computed in the same way as E(X), except that we substitute h(X) in place of X. Examples of h(X) are aX+b, e^X, lnX, etc.

Example 2.1

Let *X* be the damage incurred (in $) in a certain type of accident during a given year. Possible *X* values are 0, 1000, 5000 and 10000, with probabilities 0.8, 0.1, 0.08 and 0.02 respectively. A particular company offers a $500 deductible policy. The company wishes to fix the premium amount to be charged so that its expected profit is $100.

Since the company offers \$500 deductible policies, the amount to be paid in case of accident claim will be 0, 500, 4500 and 9500 with respective probabilities 0.8, 0.1, 0.08 and 0.02.

Expected payment for accident claim = (0)(0.8) + (500)(0.1) + (4500)(0.08) + (9500)(0.02)

= 50 + 360 + 190 = \$600

Expected profit is the difference between the premium charged and the expected expenditure on accident insurance claims.

Since expected profit is \$100, the company should charge a premium of \$700 so that
100 = 700 – 600.

Example 2.2

The pmf for the rv *X* is as follows

x	4	6	8
p(x)	0.5	0.3	0.2

and

$Y = h(X) = 20 + 3X + 0.5X^2$

The possible *Y* values are 40, 56 and 76, obtained by substituting x = 4, 6, 8 in h(x)

E(Y) = E[h(X)] = (40)(0.5) + (56)(0.3) + (76)(0.2) = 20 + 16.8 + 15.2 = 52

3 RULES OF MATHEMATICAL EXPECTATION

Let h(X) be a linear function of *X* such that h(X) = aX + b. Then ,

***Proposition* 2**

E[h(X)] = E(aX + b) = a.E(X) + b

ie, $\mu_{aX+b} = a\mu_X + b$

Proof

$$E(aX + b) = \sum_{x \in D} (ax + b).p(x)$$

$$= a\Sigma x.p(x) + b\,\Sigma p(x) = aE(X) + b, \text{ since } \Sigma p(x) = 1$$

This proposition yields three rules of expected values:

***Rule* 1**

If $b = 0$, then for any constant a, E(aX) = aE(X)

Multiplication of *X* by the constant *a* changes the unit of measurement. The rule says that expected value in the new units equals the expected value in the old units multiplied by the factor *a*.

***Rule* 2**

If $a = 1$, then for any constant b, E(X + b) = E(X) + b

If a constant is added to each possible value of *X*, there is a change in origin. Then the expected value will be shifted by the same amount *b*.

***Rule* 3**

If $a = 0$, then for any constant b, E(b) = Σb.p(x) = b Σp(x) = b

That is, the expected value of a constant is just its value. This is only logical. As a constant value is a certainty, there is no probability associated with it. The expected value is the value of the constant itself.

***Proposition* 3**

If n $\geq$ 1, $E(X^n) = \sum_{x \in D} x^n . p(x)$

This follows from *proposition* 2. Thus, the second moment (about the origin) is

$E(X^2) = \sum_{x \in D} x^2 . p(x)$

Proposition 2 can be extended to more than one rv. Let *X* and *Y* be two discrete random variables. If *X* has a set of possible values *D* with pmf *p(x),* and *Y* has a set of possible values *D** with pmf *p(y),* then for the function of the two random variables g(X,Y) = X + Y we have

***Proposition* 4**

E(X + Y) = E(X) + E(Y)

Proof

$$E(X + Y) = E[g(X,Y)] = \sum_{y \in D^*} \sum_{x \in D} g(x, y).p(x, y) = \sum_{D} x.p(x) + \sum_{D^*} y.p(y) = E(X) + E(Y)$$

We can similarly show that the expected value of the sum of any number of random variables equals the sum of their individual expectations.

Example 3.1

$E(X + Y + Z) = E[(X + Y) + Z] = E(X + Y) + E(Z) = E(X) + E(Y) + E(Z)$

Exercise 5

An individual who has automobile insurance from a certain company is randomly selected. Let *Y* be the number of traffic rule violations for which the individual was booked during the last three years. The pmf of *Y* is

y	0	1	2	3
p(y)	0.60	0.25	0.10	0.05

(a) Compute E(X)

(b) Suppose an individual with *Y* violations incurs a surcharge of ₹$200Y^2$. Calculate the expected amount of the surcharge.

Solution

(a) $E(Y) = \Sigma y.p(y) = (0)(0.60) + (1)(0.25) + (2)(0.10) + (3)(0.05)$

$= 0 + 0.25 + 0.20 + 0.15 = 0.60$

(b) Let $Z = 200Y^2$

$E(Z) = 200E(Y^2) = 200\ \Sigma y^2.p(y)$

$= 200[(0)(0.60) + (1)(0.25) + (4)(0.10) + (9)(0.05)$

$= 200[0 + 0.25 + 0.40 + 0.45] = 200(1.10) =$ ₹220

Exercise 6

An appliance dealer sells three different models of upright freezers having 13.5, 15.9, and 19.1 cubic feet of storage space, respectively. Let X = the amount of storage space of the freezer purchased by the next customer.

Suppose *X* has the following pmf

x	13.5	15.9	19.1
p(x)	0.2	0.5	0.3

(a) Compute (i) E(X), and (ii) $E(X^2)$

(b) If price of a freezer having capacity *X* cu. ft. is 25X – 8.5, what is the expected price paid by the next customer to buy a freezer?

(c) Suppose that although the rated capacity of a freezer is *X*, the actual capacity is $h(X) = X – 0.01X^2$. What is the expected actual capacity of the freezer purchased by the next customer?

Solution

(a) (i) E(X) = (13.5)(0.2) + (15.9)(0.5) + (19.1)(0.3) = 2.7 + 7.95 + 5.73 = 16.38 cu ft

(ii) $E(X^2) = (13.5)^2(0.2) + (15.9)^2(0.5) + (19.1)^2(0.3)$

= 36.45 + 126.405 + 109.443 = 272.298

(b) Let *Y* = price of freezer, where Y = 25X – 8.5

E(Y) = 25E(X) – 8.5 = (25)(16.38) – 8.5 = $401

(c) Actual capacity = $h(X) = X – 0.01X^2$

$E[h(X)] = E(X) – (0.01)E(X^2)$ = 16.38 – (0.01)(272.298) = 16.38 – 2.72 = 13.66 cu ft

Exercise 7

Let X = the outcome when a fair die is rolled once. If before the die is rolled you are offered either $\frac{1}{3.5}$ dollars or h(X) = $\frac{1}{X}$ dollars, would you accept the guaranteed amount or would you gamble?

Solution

Guaranteed amount = \$ $\frac{1}{3.5}$ = 0.2857 = \$0.29

Otherwise, h(X) = $\frac{1}{X}$, where x = 1, 2, 3, 4, 5, 6

Since this is a fair die, P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = $\frac{1}{6}$

$$E[h(X)] = \left(\frac{1}{1}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{4}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{5}\right)\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{6}\right)$$

$$= \frac{1}{6}\left[1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}\right] = \frac{1}{6}\left[\frac{60+30+20+15+12+10}{60}\right] = \frac{147}{360} = 0.4083$$

Therefore E[h(X)] = \$0.41 > guaranteed amount \$0.29. It is a better option to gamble.

One useful application of *Proposition* 1 is to obtain $E(X^2)$, ie, the second moment of *X* about the origin. Here $h(X) = X^2$. Let *D* is the set of all possible values that *X* can take. Since the pmf of X^2 is same as the pmf of X, therefore $E(X^2) = \sum_D x^2 p(x)$

4 VARIANCE OF A DISCRETE RANDOM VARIABLE

Expected value of the rv *X* is the mean of the probability distribution or pmf of *X*. It tells us what will be the value of *X* on the average when the experiment is repeated a very large number of times in an identical and independent fashion (ie, replicated a very large number of times). We need to also obtain the variance of *X* to examine the amount of variability in the probability distribution of *X*. The mean and variance are useful measures for summarizing the essential properties of the pmf.

Example 4.1

Let the rv *X* have the pmf *p(x)* = ½ for x = -1, 1 and let the rv *Y* have the pmf *p(y)* = ½ for y = -100, 100.

E(X) = E(Y) = 0 but the pmf for *Y* is more spread out than that for *X*

***Definition* 2**

Let *X* have pmf p(x) and expected value μ. Then the **variance of *X*** is

$$V(X) = \sigma_x^2 = \sum_{x \in D} (x - \mu)^2 p(x) = E(X - \mu)^2$$

and the standard deviation of *X* is $\sigma_X = \sqrt{E(X - \mu)^2}$

Variance is thus the expected value of the function $h(X) = (X - \mu)^2$, ie, the squared deviation of X from its mean. Hence, variance is the expected squared deviation. Variance is thus the weighted average of squared deviations, where the weights are the probabilities.

An alternative formula for Var(X) can be derived as follows, where $x \in D$ and D is the set of all possible values of the rv X:

$$
\begin{aligned}
Var(X) &= E[(X - \mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= E(X^2) - E(2\mu X) + E(\mu^2) \\
&= E(X^2) - \Sigma 2\mu x.p(x) + \Sigma\mu^2.p(x) && \text{where summation is over all } x \in D \\
&= E(X^2) - 2\mu\Sigma x.p(x) + \mu^2\Sigma p(x) && \text{since } \mu \text{ is a parameter of the distribution it is a constant} \\
&= E(X^2) - 2\mu E(X) + \mu E(X) \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}
$$

Variance of *X* is, therefore, equal to the expected value of the square of *X* minus the square of the expected value of *X*. It is often easier to compute V(X) using this formula than the definitional formula $E[(X - \mu)^2]$.

Example 4.2

Given the pmf of *X* in *example* 2.2, we can compute the mean, variance and standard deviation of the probability distribution of the random variable *X*.

x	p(x)	x.p(x)	x^2	$x^2.p(x)$
4	0.5	2	16	8
6	0.3	1.8	36	10.8
8	0.2	1.6	64	12.8

$E(X) = \Sigma x.p(x) = 2 + 1.8 + 1.6 = 5.4$

$V(X) = E(X^2) - [E(X)]^2$ where $E(X^2) = \Sigma x^2.p(x) = 8 + 10.8 + 12.8 = 31.6$

$\therefore V(X) = 31.6 - (5.4)^2 = 31.6 - 29.16 = 2.44$ and $\sigma_X = \sqrt{2.44} = 1.562$

5 VARIANCE OF A FUNCTION OF A DISCRETE RANDOM VARIABLE

Let *h(x)* denote a function of the random variable *X*. By *definition* 2,

$$V[h(x)] = E[h(x) - E(h(x))]^2 = \sum_D [h(x) - E(h(x))]^2.p(x)$$

If we have a linear function h(x) = aX + b, then

V[h(x)] $= \sum_D [(ax+b) - E(ax+b)]^2 . p(x)$

Now, E(aX + b) = a.E(X) + b

If we denote E(X) = μ, then E(aX + b) = aμ + b, so that

V[h(x)] $= \sum_D [(ax+b) - (a\mu+b)]^2 . p(x)$

$$= \sum_D [a(x-\mu)]^2 . p(x)$$

$$= a^2 \sum_D [x-\mu]^2 . p(x)$$

$$= a^2 E[X-\mu]^2$$

$$= a^2 V(X)$$

Thus we get a simple relationship between V[h(x)] and V(X) for the linear function h(x) = aX + b

***Proposition* 5**

$V(aX+b) = \sigma^2_{aX+b} = a^2\sigma^2_X$ and $\sigma_{aX+b} = |a|.\sigma_X$

We need to take the absolute value |a| since $\sqrt{a^2}$ = ± a, and standard deviation can never be negative.

Proposition 5 yields the following two rules of variance and standard deviation for the function of a random variable.

***Rule* 4**

$V(aX) = \sigma^2_{aX} = a^2 V(X)$ and $\sigma_{aX} = |a|.\sigma_X$ when b= 0 in h(x) = aX + b

***Rule* 5**

$V(X+b) = \sigma^2_{X+b} = V(X)$ and $\sigma_{X+b} = \sigma_X$ when a= 1 in h(x) = aX + b

Thus change in origin by adding or subtracting a constant *b* does not affect the variability of the distribution. It just shifts the distribution to the left for $b < 0$, and to the right for $b > 0$. However, change in the unit of measurement by multiplication or division by a constant *a* impacts the variability. The new standard deviation is a product of the old standard deviation and the absolute value of the conversion factor *a*. If $0 < |a| < 1$ the distribution becomes narrower. If $|a| > 1$, the new distribution is more spread out than before.

Exercise 8

The total cost for the production process is equal to $1000 plus two times the number of units produced. The mean and variance for the number of units produced are 500 and 900 respectively. Find the mean and standard deviation of the total cost.

Solution

Let *X* denote the number of units produced where *X* is a rv. Then the cost function is

h(x) = 1000 +2X.

Given that E(X) = 500 and V(X) = 900,

E[h(x)] = E[1000 + 2X] = 1000 + 2E(X) = 1000 + (2)(500) = $2000

and

V[h(x)] = V(1000 + 2X) = 2V(X) = (4)(900) = 3600

so that $\sigma_{h(x)} = \sqrt{3600}$ = $60

6 COVARIANCE AND VARIANCE OF SUMS OF RANDOM VARIABLES

We have seen in *example* 4.1 that, using *proposition* 4, the expected value of a sum of random variables is the sum of their expected values, ie,

E(X + Y + Z) = E(X) + E(Y) + E(Z).

However, the variance of a sum of random variables is *generally* not equal to the sum of the individual variances.

Example 6.1

V(X + X) = V(2X) = 4 V(X) whereas V(X) + V(X) = 2 V(X)

so that [V(X + X)] ≠ [V(X) + V(X)]

If the random variables are, however, *independent* then variance of a sum of the rv's will equal the sum of the respective variances. Recall that if *X* and *Y* are two independent rv's then the probability of occurrence of one variable is not affected by the probability of occurrence of the other. To prove that V(X+Y) = V(X) + V(Y) when *X* and *Y* are independent we need to first define the concept of *covariance* of two random variables.

***Definition* 3**

The ***covariance*** of two random variables *X* and *Y* is defined by

Cov(X,Y) = E[(X - μ_X)(Y - μ_Y)], where μ_X and μ_Y are the means of X and Y respectively.

From the definition of covariance we can derive an alternative formula.

Cov(X,Y) = E[(X - μ_X)(Y - μ_Y)]

= E[XY – μ_XY - μ_YX + μ_X μ_Y]

= E(XY) – μ_XE(Y) - μ_YE(X) + μ_X μ_Y

Since E(X) =μ_X and E(Y) = μ_Y are constants of the probability distributions of *X* and *Y* respectively,

Cov(X,Y) = E(XY) - μ_X μ_Y - μ_X μ_Y + μ_X μ_Y

= E(EY) - μ_X μ_Y

= E(X,Y) – E(X).E(Y)

If *X* and *Y* are independent rv's then Cov(X,Y) = 0. This is because variability in *X* is unrelated to variability in *Y*. To show that Cov(X,Y) = 0 for independent rv's *X* and *Y*, we need to prove that E(XY) = E(X).E(Y)

Proof

Let the rv *X* have a pmf p(x). Let the domain of *X* be denoted by D = { x_1, x_2,……., x_m}

Let the rv *Y* have a pmf p(y). Let the domain of *Y* be denoted by D* = { y_1, y_2,……., y_n}

$$E(XY) = \sum_{x_i \in D} \sum_{y_j \in D^*} x_i y_j \, p(x_i y_j)$$

Since *X* and *Y* are independent, p(xy) = p(x).p(y), therefore

$$E(XY) = \sum_{x_i \in D} \sum_{y_j \in D^*} x_i y_j \, p(x_i) p(y_j) = \sum_{x_i \in D} \sum_{y_j \in D^*} x_i \, p(x_i).y_j \, p(y_j)$$

$$= \sum_{x_i \in D} x_i \, p(x_i).E(Y) = E(X).E(Y)$$

Thus, Cov(X,Y) = E(XY) – E(X).E(Y)

= E(X).E(Y) – E(X).E(Y) = 0, if *X* and *Y* are independent.

Now let X and Y be two independent random variables. Then,

$V(X + Y) = E[(X + Y) – E(X + Y)]^2$

$= E[(X + Y) –\{E(X) + E(Y)\}]^2$

$= E[\{X – E(X)\} + \{Y – E(Y)\}]^2$

$= E[X – E(X)]^2 + E[Y – E(Y)]^2 + 2 E[\{X – E(X)\}\{ Y – E(Y)\}]$

= V(X) + V(Y) + 2 Cov(X,Y)

Since Cov(X,Y) = 0 when *X* and *Y* are independent random variables, therefore,

V(X + Y) = V(X) + V(Y)

From this it follows that if X_1, X_2,…….. are independent random variables, then

$V(X_1 + X_2 +, \ldots\ldots) = \Sigma V(X_i)$ if each pair of X_i, X_j are mutually independent ($i \neq j$).

Exercise 9

A company produces and sells security devices in two countries which do not permit international trade in this item. Let *X* and *Y* denote the number of devices sold weekly in the first country and the second country respectively. The profit function in the two countries are h(x) = 200X - 100 and h(y) = 500Y- 250

Compute the mean and standard deviation of weekly total profits (measured in $) of the company if the pmf's of *X* and *Y* are as follows:

x	3	4	5	6
p(x)	0.1	0.2	0.3	0.4

and

y	1	2	3	4
p(y)	0.2	0.4	0.3	0.1

Solution

E(X) = 3(0.1) + 4(0.2) + 5(0.3) + 6(0.4) = 0.3 + 0.8 + 1.5 + 2.4 = 5

$E(X^2)$ = 9(0.1) + 16(0.2) + 25(0.3) + 36 (0.4) = 0.9 + 3.2 + 7.5 + 14.4 = 26

V(X) = 26 – 25 = 1

E(Y) = 1(0.2) + 2(0.4) + 3(0.3) + 4(0.1) = 0.2 + 0.8 + 0.9 + 0.4 = 2.3

$E(Y^2) = 1(0.2) + 4(0.4) + 9(0.3) + 16(0.1) = 0.2 + 1.6 + 2.7 + 1.6 = 6.1$

$V(Y) = 6.1 – 5.29 = 0.81$

Expected profits in the first country = E[h(x)] = E[200X - 100]

= 200.E(X) -100

= 200(5) - 100 = $900

Variance of profits in first country = V[200X -100] = 40000 V(X) = 40000(1) = 40,000

Expected profits in the second country = E[h(y)] = E[500Y – 250]

= 500.E(Y) – 250

= 500(2.3) – 250 = $900

Variance of profits in second country = V[500Y – 250] = 250000 V(Y)

= 250000(0.81) = 202,500

Expected total profits = E[h(x)] + E[h(y)] = 900 + 900 = $1,800

Variance of total profits = V[h(x)] + V[h(y)] since profits in the two countries are independent as there is no trade in this device in either country so that Cov(X,Y) = 0

V[h(x)] + V[h(y)] = 40000 + 202500 = 242500

Therefore standard deviation of total profits = $\sqrt{242500}$ = $492.44

Exercise 10

Show that Cov(aX+b,cY+d) = acCov(X,Y)

Solution

Cov(aX+b,cY+d) = E[{aX + b – E(aX + b)}{cY + d – E(cY + d)}]

= E[{aX + b – aE(X) +b}{cY + d – cE(Y) + d}]

= E[(a{X – E(X)})(c{Y- E(Y)})]

= acE[{X – E(X)}{Y – E(Y)}] = acCov(X,Y)

7 PARAMETERS OF THE PROBABILITY MASS FUNCTION

When the pmf specifies a mathematical model for the distribution of population values, the expected value or mean μ measures the value of the rv at which the distribution is centered. Both σ^2 and σ measure the spread of the population distribution where σ^2 is the population variance and σ is the population standard deviation.

If most of the population values are close to μ, the spread of the distribution is small and σ^2 is relatively small. If, however, there are x values that are far from μ that have large p(x), then σ^2 will be quite large.

Example 7.1

In *example* 5.1, p(x) = ½ for x = -1, 1 and p(y) = ½ for y = -100, 100

$E(X) = \mu_X = ½(-1) + ½(1) = 0$,

and $E(Y) = \mu_Y = ½(-100) + ½(100) = 0$,

so that $\mu_X = \mu_Y = 0$

$V(X) = E(X^2) – [E(X)]^2 = [½ (1) + ½(1)] – 0 = 1$, and

$V(Y) = E(Y^2) – [E(Y)]^2 = [½ (10000) + ½(10000)] – 0 = 10,000$

Therefore, V(Y) > V(X)

The characteristics of the distribution can now be specified. We can obtain the calculated values of the mean and variance. The histrogram of the distribution will show whether the distribution is symmetric or asymmetric. It will also show whether the distribution is unimodal, bimodal or multomodal.

Practice Questions

1 If *X* takes on the values 0, 1, 2, and 3 with probabilities $\frac{1}{125}, \frac{12}{125}, \frac{48}{125}, \ and\ \frac{64}{125}$ respectively, find E(X) and V(X). Use these results to find the mean and variance of Y = 3X + 2

2. The pmf of the amount of memory *X*(GB) in a flash drive is given as follows:

x	1	2	4	8	16
p(x)	0.05	0.10	0.35	0.40	0.10

Compute the following

(a) E(X)

(b) V(X) using the definitional formula

(c) Standard deviation of X

(d) V(X) using the shortcut formula

3. Use the proposition involving V(aX +b) to establish a general relationship between V(X) and V(-X).

4. If X denotes a temperature recorded in degrees Fahrenheit, then $\frac{5}{9}(X-32)$ is the corresponding temperature in degrees Celsius. If the standard deviation for a set of temperatures is 15.7°F, what is the standard deviation of the equivalent Celsius temperatures?

5. If E(W) = μ and V(W) = σ^2, show that

$$E\left(\frac{W-\mu}{\sigma}\right)=0 \quad and \quad V\left(\frac{W-\mu}{\sigma}\right)=1$$

6. Suppose that X_i is a rv for which $E(X_i) = \mu$, I = 1, 2,……,n. Under what conditions will the following be true?

$$E\left(\sum_{i=1}^{n} a_i X_i\right)=\mu$$

7. A stationery shop orders copies of a certain magazines each week. Let X = demand for the magazine, with pmf

x	1	2	3	4	5	6
p(x)	1/15	2/15	3/15	4/15	3/15	2/15

Suppose the shop actually pays ₹5 for each copy of the magazine and the price to customers is ₹10. If magazines left at the end of the week have no salvage value, is it better to order three or four copies of the magazine?

8. An industrialist has a choice of two alternative proposals to start a new project. *Proposal A* will yield a profit of ₹10 lakhs with a probability of 0.4 or a loss of ₹ 2 lakhs with a probability of 0.6. *Proposal B* will yield a profit of ₹ 4.5 lakhs with probability 0.8 or a loss of ₹ 50,000 with probability 0.2. Which proposal should he prefer?

9. Given that variables *X* and *Y* are independent and Z = aX – bY, prove that $Var(Z) = a^2Var(X) + b^2Var(Y)$

10. Arun and Barun play a game in which they toss a fair coin three times. The one obtaining heads first wins the game. If Arun tosses the coin first and if the total value of the stakes is ₹20, how much should be contributed by each in order that the game be considered fair?

11. A bakery sells bread for Rs. 15 each. Daily sales *X* is a random variable and has a distribution with mean 530 and standard deviation 69
    (i) Find the mean daily total revenues from the sale of bread
    (ii) Find the standard deviation of total revenues from the sale of bread
    (iii) If daily costs (in Rs) for making bread are given by C=1000+0.95X, find the mean and variance of daily profits from sales of bread

12. A chemical supply company currently has in stock 100 kg of a certain compound, which it sells to customers in 5-kg batches. Let *X* = the number of batches ordered by a randomly chosen customer, and suppose that *X* has pmf

x	1	2	3	4
p(x)	0.2	0.4	0.3	0.1

Compute E(X) and V(X). Then compute the expected number of kgs left after the customer's order is shipped and the variance of the number of kgs left.

13. A sports promoter is contemplating buying a rain insurance for an event he is sponsoring. If it does not rain he expects to earn ₹ 10,000 but only ₹ 2,000 if it does. If the probability of rain is 3/7, what is his expected earnings? If the insurance policy costs him ₹ 3,000 and assures him ₹ 7,000 if it rains, is it profitable to purchase the insurance?

14. Use the proposition involving V(aX +b) to establish a general relationship between V(X) and V(-X)