MTH302 Unit 1 (iii)
Mathematical expectation(Mean) , Variance , Covariance and Chebyshev's theorem

Mean, or Expected value of r. v. X

Definition 1:

Let X be a random variable with probability distribution $f(x)$. The **mean**, or **expected value**, of X is

$$\mu = E(X) = \sum_x x f(x)$$

if X is discrete, and

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

if X is continuous.

Theorem 1

Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x) f(x)$$

if X is discrete, and

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

if X is continuous.

Definition 2

Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

$$\mu_{g(X,Y)} = E[g(X,Y)] = \sum_x \sum_y g(x, y) f(x, y)$$

if X and Y are discrete, and

$$\mu_{g(X,Y)} = E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy$$

if X and Y are continuous.

NOTE: In the definition 2 above

(i)

if $g(X,Y) = X$

Then

$$E(X) = \begin{cases} \sum_x \sum_y x f(x,y) = \sum_x x g(x) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y)\, dy\, dx = \int_{-\infty}^{\infty} x g(x)\, dx & \text{(continuous case)}, \end{cases}$$

where $g(x)$ is the marginal distribution of X

(ii)

if $g(X,Y) = Y$

Then

$$E(Y) = \begin{cases} \sum_y \sum_x y f(x,y) = \sum_y y h(y) & \text{(discrete case)}, \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y)\, dx dy = \int_{-\infty}^{\infty} y h(y)\, dy & \text{(continuous case)} \end{cases}$$

where $h(y)$ is the marginal distribution of the random variable Y

Variance of Random Variables X

Definition 3

Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

Theorem 2

The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

Theorem 3

Let X be a random variable with probability distribution $f(x)$. The variance of the random variable $g(X)$ is

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$

if X is discrete, and

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)\, dx$$

if X is continuous.

Covariance of Random Variables

Definition 4

Let X and Y be random variables with joint probability distribution $f(x, y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_y) f(x, y)$$

if X and Y are discrete, and

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_y) f(x, y)\, dx\, dy$$

if X and Y are continuous.

Theorem 4

The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y.$$

Correlation Coefficient

Although the covariance between two random variables does provide information regarding the nature of the relationship, the magnitude of **σ**XY does not indicate anything regarding the strength of the relationship , since **σ**XY is not scale-free.Its magnitude will depend on the units used to measure both **X** and **Y** . There is a scale-free version of the covariance called the correlation coefficient that is used widely in statistics.

Definition 5

Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Note:

- ρ_{XY} is free of the units of X and Y
- $-1 \le \rho_{XY} \le 1$
- It assumes a value of zero when **σ**XY = 0.
- Where there is an exact linear dependency say Y ≡ a + bX ,

  $\rho_{XY} = 1$ if $b > 0$ and

  $\rho_{XY} = -1$ if $b < 0$.

Solved Examples

1. (Example for discrete r.v.)

   Let X be a random variable with the following probability distribution

x	-2	3	5
$f(x)$	0.3	0.2	0.5

   Find the mean, variance and standard deviation of the random variable X.

   Solution:

   $\mu = (-2)(0.3) + (3)(0.2) + (5)(0.5) = 2.5$ and
   $E(X^2) = (-2)^2(0.3) + (3)^2(0.2) + (5)^2(0.5) = 15.5$.
   So, $\sigma^2 = E(X^2) - \mu^2 = 9.25$ and $\sigma = 3.041$.

2. (Example for cont. r.v.)

For a laboratory assignment, if the equipment is working, the density function of the observed outcome X is

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

Find the mean, variance and standard deviation of the random variable X.

Solution:

$E(X) = 2\int_0^1 x(1-x)\,dx = 2\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1 = \frac{1}{3}$ and

$E(X^2) = 2\int_0^1 x^2(1-x)\,dx = 2\left(\frac{x^3}{3} - \frac{x^4}{4}\right)\Big|_0^1 = \frac{1}{6}$. Hence,

$Var(X) = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$, and $\sigma = \sqrt{1/18} = 0.2357$.

3. (Example for discrete r.v.)

Let X be a random variable with the following

Let X be a random variable with the following probability distribution

x	-3	6	9
$f(x)$	$1/6$	$1/2$	$1/3$

Find mean variance and the standard deviation of the random variable **g(X)** where $g(X) = (2X+1)^2$.

Solution:

The probability density function is,

x	-3	6	9
$f(x)$	$1/6$	$1/2$	$1/3$
$g(x)$	25	169	361

$$\mu_{g(X)} = E[(2X+1)^2] = (25)(1/6) + (169)(1/2) + (361)(1/3) = 209.$$

$$\sigma^2_{g(X)} = \sum_x [(2X+1)^2 - 209]^2 g(x)$$
$$= (25-209)^2(1/6) + (169-209)^2(1/2) + (361-209)^2(1/3) = 14,144.$$

So, $\sigma_{g(X)} = \sqrt{14,144} = 118.9$.

4. (Example for cont. r.v.)

The length of time, in minutes, for an airplane to obtain clearance for takeoff at a certain airport is a random variable Y = 3X −2, where X has the density function

$$f(x) = \begin{cases} \frac{1}{4}e^{-x/4}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the mean and variance of the random variable Y .

Solution:

$$\mu_Y = E(3X - 2) = \frac{1}{4}\int_0^\infty (3x-2)e^{-x/4}\, dx = 10. \text{ So}$$
$$\sigma_Y^2 = E\{[(3X-2) - 10]^2\} = \frac{9}{4}\int_0^\infty (x-4)^2 e^{-x/4}\, dx = 144.$$

5. (Example for two rvs with discrete data)

The two random variables X and Y with the following joint probability distribution f(x,y).

	$f(x,y)$	$x=0$	$x=1$	$x=2$	$h(y)$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
	$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

(i)Find the covariance of X and Y .

(ii) Find the correlation coefficient between X and Y

Solution:

(i) Formula to find Cov(X,Y) is

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y.$$

Where

$$\begin{aligned} E(XY) &= \sum_{x=0}^{2}\sum_{y=0}^{2} xyf(x,y) \\ &= (0)(0)f(0,0) + (0)(1)f(0,1) \\ &\quad + (1)(0)f(1,0) + (1)(1)f(1,1) + (2)(0)f(2,0) \\ &= f(1,1) = \frac{3}{14}. \end{aligned}$$

$$\mu_X = \sum_{x=0}^{2} xg(x) = (0)\left(\frac{5}{14}\right) + (1)\left(\frac{15}{28}\right) + (2)\left(\frac{3}{28}\right) = \frac{3}{4},$$

and

$$\mu_Y = \sum_{y=0}^{2} yh(y) = (0)\left(\frac{15}{28}\right) + (1)\left(\frac{3}{7}\right) + (2)\left(\frac{1}{28}\right) = \frac{1}{2}.$$

Therefore,

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

(ii) Formula for correlation coefficient

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$E(X^2) = (0^2)\left(\frac{5}{14}\right) + (1^2)\left(\frac{15}{28}\right) + (2^2)\left(\frac{3}{28}\right) = \frac{27}{28}$$

and

$$E(Y^2) = (0^2)\left(\frac{15}{28}\right) + (1^2)\left(\frac{3}{7}\right) + (2^2)\left(\frac{1}{28}\right) = \frac{4}{7},$$

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112}$$

and

$$\sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}.$$

Therefore, the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-9/56}{\sqrt{(45/112)(9/28)}} = -\frac{1}{\sqrt{5}}.$$

6.(Example for two cont. rvs)

A fast-food restaurant operates both a drive through facility and a walk-in facility. On a randomly selected day, let X and Y , respectively, be the proportions of the time that the drive-through and walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y), & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y.

Solution:

$g(x) = \frac{2}{3}\int_0^1 (x+2y)\ dy = \frac{2}{3}(x+1)$, for $0 < x < 1$, so $\mu_X = \frac{2}{3}\int_0^1 x(x+1)\ dx = \frac{5}{9}$;
$h(y) = \frac{2}{3}\int_0^1 (x+2y)\ dx = \frac{2}{3}\left(\frac{1}{2}+2y\right)$, so $\mu_Y = \frac{2}{3}\int_0^1 y\left(\frac{1}{2}+2y\right)\ dy = \frac{11}{18}$; and
$E(XY) = \frac{2}{3}\int_0^1\int_0^1 xy(x+2y)\ dy\ dx = \frac{1}{3}$.
So, $\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{1}{3} - \left(\frac{5}{9}\right)\left(\frac{11}{18}\right) = -0.0062$.

(Here μ_x and μ_Y can also be calculated directly with double integral)

7. The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x,y) = \begin{cases} 8xy, & 0 \le y \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(i) Find the covariance of X and Y .
(ii) Find the correlation coefficient between X and Y

Solution:

(i)

We first compute the marginal density functions. They are

$$g(x) = \begin{cases} 4x^3, & 0 \le x \le 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and

$$h(y) = \begin{cases} 4y(1-y^2), & 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

From these marginal density functions, we compute

$$\mu_X = E(X) = \int_0^1 4x^4 \, dx = \frac{4}{5} \text{ and } \mu_Y = \int_0^1 4y^2(1-y^2) \, dy = \frac{8}{15}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2y^2 \, dx \, dy = \frac{4}{9}.$$

Then

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{4}{9} - \left(\frac{4}{5}\right)\left(\frac{8}{15}\right) = \frac{4}{225}.$$

(ii)

$$E(X^2) = \int_0^1 4x^5 \, dx = \frac{2}{3} \text{ and } \quad \mu_X = E(X) = \int_0^1 4x^4 \, dx = \frac{4}{5}$$

$$\sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112}$$

$$E(Y^2) = \int_0^1 4y^3(1-y^2) \, dy = 1 - \frac{2}{3} = \frac{1}{3}, \text{ and } \mu_Y = \int_0^1 4y^2(1-y^2) \, dy = \frac{8}{15}.$$

$$\sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}.$$

Therefore, the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{-9/56}{\sqrt{(45/112)(9/28)}} = -\frac{1}{\sqrt{5}}.$$

(Chebyshev's Theorem)

(Chebyshev's Theorem) The probability that any random variable X will assume a value within k standard deviations of the mean is at least $1 - 1/k^2$. That is,

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

Also,

$$\mathbf{P(|X - \mu| \geq k\sigma)} < \frac{1}{k^2}$$

Example

A random variable X has a mean $\mu = 8$, a variance $\sigma^2 = 9$, and an unknown probability distribution. Find

(a) $P(-4 < X < 20)$,

(b) $P(|X - 8| \geq 6)$.

Solution: (a) $P(-4 < X < 20) = P[8 - (4)(3) < X < 8 + (4)(3)] \geq \frac{15}{16}$.

(b) $P(|X - 8| \geq 6) = 1 - P(|X - 8| < 6) = 1 - P(-6 < X - 8 < 6)$

$$= 1 - P[8 - (2)(3) < X < 8 + (2)(3)] \leq \frac{1}{4}.$$

[(b) can also be done using alternative inequality as discussed in the class]

Ex.

A random variable X has a mean $\mu = 10$ and a variance $\sigma^2 = 4$. Using Chebyshev's theorem, Find

(a) $P(|X - 10| \geq 3)$;

(b) $P(|X - 10| < 3)$;

(c) $P(5 < X < 15)$;

(d) the value of the constant c such that

$$P(|X - 10| \geq c) \leq 0.04.$$