Random Variables and probability Distributions-II

Mr. Anup Singh
Department of Mathematics
Mahatma Gandhi Central university
Motihari-845401, Bihar, India
E-mail: anup.singh254@gmail.com

Mathematical Expectation: Let X be a random variable having a probability distribution function f(x). Expected value of the random variable is the arithmetic mean of the random variable. The mean or expected value of the random variable u(X), Then

If X is discrete type of random variable

$$\mu_{u(X)} = E[u(x)] = \sum_{x} u(x) f(x)$$

If X is continuous type of random variable

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

Special Mathematical Expectation: Let $u(X) = X$, where X is a random variable of the discrete type having a p.d.f. $f(x)$. Then

1. Mean of the random variable is $E[X] = \mu_X$
2. If a is constant, $E[a] = a$
3. If a and b are constants, $E[aX \pm b] = a\,E[X] \pm b$
4. $E[f(X) \pm g(X)] = E[f(X)] \pm E[g(X)]$
5. Variance of the random variable

   $\sigma_X^2 = Var[X] = E[(X - \mu_X)^2] = E[X^2] - E[X]^2$
6. If a is constant, $Var[a] = 0$
7. $Var[aX \pm b] = a^2\,Var[X]$

Example (i) In a gambling game a man is paid Rs. 5 if he gets all heads or all tails when three coins are tossed, and he will pay out Rs. 3 if either one or two heads show,. What is his expected gain?

Solution: The sample space for the possible outcomes when three coins are tossed simultaneously, or equivalently if 1 coin is tossed three times, is

S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}.

The amount the gambler can:

win is Rs.5; if event

E_1 = {HHH, TTT} occurs and

lose Rs.3; if event

E_2 = {HHT, HTH, THH, HTT, THT, TTH} occurs.

Since E_1 and E_2 occur with probabilities 1/4 and 3/4 respectively, it follows that

$$\mu = E[X] = \sum_x x\, p(x) = (5)\left(\frac{1}{4}\right) + (-3)\left(\frac{3}{4}\right) = -1$$

In this game, the gambler will on an average lose Rs. 1 per toss of the three coins.

Example (ii) Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \dfrac{20{,}000}{x^3}, & x > 100 \\ 0, & \text{elsewhere} \end{cases}$$

Find the expected life of this type of device.

Solution: We have,

$$\mu = E[X] = \int_{100}^{\infty} x \frac{20000}{x^3} dx = \int_{100}^{\infty} \frac{20000}{x^2} dx = 200$$

Therefore, we can expect this type of device to last, on average 200 hours.

Example (iii) Suppose that the number of cars X that pass through a car wash between 4.00p.m and 9.00p.m. on any day has the following distribution:

x	4	5	6	7	8	9
P(X = x)	1/12	1/12	1/4	1/4	1/6	1/6

Let g(X) = 2X – 1 represent the amount of money in rupees, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution: The attendant can expect to receive

$$E[g(X)] = E[2X - 1]$$

$$= \sum_{x=4}^{9} (2x-1)\,p(x)$$

$$= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) + (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right)$$

$$= 12.67$$

Example (iv) Let X be a random variable with density function

$$f(x) = \begin{cases} \dfrac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the expected value of g(X) = 4X + 3.

Solution: we have

$$E[g(X)] = E[4X + 3]$$

$$= \int_{-1}^{2} \frac{(4x+3)x^2}{3} dx = \frac{1}{3} \int_{-1}^{2} (4x^3 + 3x^2) dx = 8$$

Example (v) The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X *having the* probability density

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

Find the mean and variance of X.

Solution: Calculating *E(X) and E(X^2), we have*

$$\mu = E[X] = 2\int_1^2 x(x-1)\,dx = \frac{5}{3}$$

$$E[X^2] = 2\int_1^2 x^2(x-1)\,dx = \frac{17}{6}$$

$$\therefore Var[X] = \frac{17}{6} - \left[\frac{5}{3}\right]^2 = \frac{1}{18}$$

Moment Generating Function: Moment generating function (MGF) of a r.v. X (discrete or continuous) is defined as $E[e^{tX}]$, where t is a real variable and denoted as M(t).

i.e., $M(t) = E[e^{tX}]$

If X is discrete type of random variable

$$M(t) = \sum_x e^{tx} f(x)$$

If X is continuous type of random variable

$$M(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Also, we have,

$$M(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n) \qquad E(X^n) = \left[\frac{d^n}{dt^n} M(t) \right]_{t=0}$$

Example(i) If X represents the outcome, when a fair die is tossed, find the MGF of X and hence find E(X) and Var(X).

Solution: The probability distribution of X is given by

$$P(X = i) = 1/6, \; i = 1,2,3,4,5,6$$

$$M(t) = \sum_x e^{tx} p_x = \frac{1}{6}(e^t + e^{2t} + e^{3t} + e^{4t} + e^{5t} + e^{6t})$$

$$E(X) = [M'(t)]_{t=0}$$

$$= \left[\frac{1}{6}(e^t + 2e^{2t} + 3e^{3t} + 4e^{4t} + 5e^{5t} + 6e^{6t})\right]_{t=0} = \frac{7}{2}$$

$$E(X^2) = [M''(t)]_{t=0}$$

$$= \frac{1}{6}(e^t + 4e^{2t} + 9e^{3t} + 16e^{4t} + 25e^{5t} + 36e^{6t})_{t=0} = \frac{91}{6}$$

$$Var(X) = E(X^2) - [E(X)]^2 == \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

Chebyshev's Inequality If X is a random variable with mean μ and variance σ^2, then for any positive number K, we have

$$P\{|X-\mu|\geq K\sigma\}\leq\frac{1}{K^2} \quad \text{or} \quad P\{|X-\mu|<K\sigma\}\geq 1-\frac{1}{K^2}$$

Proof: We know that

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

$$= \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$$

$$= \int_{-\infty}^{\mu-K\sigma}(x-\mu)^2 f(x)dx + \int_{\mu-K\sigma}^{\mu+K\sigma}(x-\mu)^2 f(x)dx + \int_{\mu+K\sigma}^{\infty}(x-\mu)^2 f(x)dx$$

$$\geq \int_{-\infty}^{\mu-K\sigma}(x-\mu)^2 f(x)dx + \int_{\mu+K\sigma}^{\infty}(x-\mu)^2 f(x)dx$$

From the first Integral	From the second Integral
$x \le \mu - K\sigma$ $-(x-\mu) \ge K\sigma$ $(x-\mu)^2 \ge K^2\sigma^2$	$x \ge \mu + K\sigma$ $(x-\mu) \ge K\sigma$ $(x-\mu)^2 \ge K^2\sigma^2$

$$\sigma^2 \ge \int_{-\infty}^{\mu-K\sigma} K^2\sigma^2 f(x)dx + \int_{\mu+K\sigma}^{\infty} K^2\sigma^2 f(x)dx$$

$$1 \ge \int_{-\infty}^{\mu-K\sigma} K^2 f(x)dx + \int_{\mu+K\sigma}^{\infty} K^2 f(x)dx$$

$$1 \ge K^2\{P[X \le \mu - K\sigma] + P[X \ge \mu + K\sigma]$$

$$1 \ge K^2\{P[X - \mu \le -K\sigma] + P[X - \mu \ge K\sigma]$$

$$1 \ge K^2\{P[|X - \mu| \ge K\sigma]$$

Hence

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

Also we know,

$$P[|X-\mu|\geq K\sigma]+P[|X-\mu|<K\sigma]=1, \quad \because P(A)+P(A^C)=1$$

$$P[|X-\mu|<K\sigma]=1-P[|X-\mu|\geq K\sigma]$$

$$\therefore P[|X-\mu|<K\sigma]\geq 1-\frac{1}{K^2}$$

Example (i) A discrete RV X takes the values $-1,0,1$ with probabilities 1/8, 3/4, 1/8 respectively. Evaluate $P[|X-\mu| \geq 2\sigma]$ and compare it with the upper bound given by chebyshev's inequality.

Solution:

We have,

$$E[X] = -1 \times \frac{1}{8} + 0 \times \frac{3}{4} + 1 \times \frac{1}{8} = 0$$

$$E[X^2] = 1 \times \frac{1}{8} + 0 \times \frac{3}{4} + 1 \times \frac{1}{8} = \frac{1}{4}$$

$$Var[X] = E[X^2] - E[X]^2 = \frac{1}{4}$$

$$\sigma = \sqrt{Var[X]} = \sqrt{\frac{1}{4}} = \frac{1}{2}$$

$$\therefore P[|X - \mu| \geq 2\sigma] = P[|X| \geq 1] = P[X = -1] + P[X = 1] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

Using Chebyshev's inequality,

$$P[|X - \mu| \geq 2\sigma] \leq \frac{1}{2^2} = \frac{1}{4}, here K = 2$$

Example (ii) Let X be a continuous random variable whose probability density function given by $f(x)=e^{-x}$, $0 \le x \le \infty$. Using Chebyshev's inequality verify

$$P[|X-\mu|>2] \le \frac{1}{4}$$

and show that actual probability is e^{-3}.

Solution: We have $E[X]=\int_0^\infty xe^{-x}dx$

$$=\left[-xe^{-x}-e^{-x}\right]_0^\infty=1$$

$$E[X^2]=\int_0^\infty x^2e^{-x}dx$$

$$=\left[-x^2e^{-x}-2xe^{-x}-2e^{-x}\right]_0^\infty=2$$

$$Var[X]=E[X^2]-E[X]^2=2-1=1 \Rightarrow \sigma=1$$

Therefore, $P[|X-1|>2] = P(-\infty \le X-1 \le -2) + P(2 \le X-1 \le \infty)$

$$= P(-\infty \le X \le -1) + P(3 \le X \le \infty)$$

$$= 0 + \int_3^{\infty} f(x)dx$$

$$= e^{-3}$$

Using the Chebyshev's inequality, $P\{|X-\mu| \ge K\sigma\} \le \dfrac{1}{K^2}$

$$K\sigma = 2;\ K = 2,\quad \because \sigma = 1$$

$$P[|X-\mu|>2] \le \frac{1}{4}$$

Two Dimensional Random Variables: Let S be the sample space. Let X= X(S), Y= Y(S) be the two functions each assigning a real number to each outcome s $\in$ S then (X, Y) is a two dimensional random variable.

If the possible values of (X, Y) are finite or countably infinite, then (X, Y) is called a two dimensional discrete random variable.

If (X, Y) can assume all values in a specified region R in the (X, Y) plane then (X, Y) is called a two dimensional continuous random variables.

Joint Probability Function:

1-Discrete Random Variables: If (X,Y) is a 2-dimensional discrete random variable such that

$P(X = x_i, Y = y_j) = p_{ij}$

Then p_{ij} is called the joint probability mass function provided the following conditions satisfied

1. $p_{ij} \geq 0$
2. $\sum_j \sum_i p_{ij} = 1$

2-Continuous Random Variable: If X and Y are continuous random variables with $f(x,y)$ is called the joint probability density function provided the following conditions are satisfied

1. $f(x, y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$

Marginal Probability Function: If the joint probability distribution of two random variables X and Y are given

1- Discrete Random Variables

Marginal probability function of X is given by

$$P(X = x_i) = p_{i*} = \sum_j p_{ij}$$

Marginal probability function of X is given by

$$P(Y = y_j) = p_{*j} = \sum_i p_{ij}$$

2- Continuous Random Variables

Marginal probability function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

Marginal probability function of X is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Marginal Distribution Function

1-Discrete Random Variable: The collection of pairs $\{x_i, p_{i*}\}$ where i takes the value 1, 2, 3, … etc. is called the Marginal distribution of X. Similarly the collection of pairs $\{y_j, p_{*j}\}$ where j takes the value 1, 2, 3, … etc. is called the Marginal distribution of Y.

2-Continuous Random Variable: If the joint distribution of the random variables X, Y is F[x, y], then

the marginal distribution of X is $F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x, y)dydx$

the marginal distribution of Y is $F_Y(y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(x, y)dxdy$

Conditional Probability Density Function

Discrete Random Variable

Conditional Probability of X given Y = y_j is given by

$$P\left(X = x_i \big/ Y = y_j \right) = \frac{p(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{*j}}$$

Conditional Probability of Y given X = x_i is given by

$$P\left(Y = y_j \big/ X = x_i \right) = \frac{p(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i*}}$$

Continuous Random Variable

The conditional probability function of X given Y

$$f\left(x/y\right) = \frac{f(x,y)}{f(y)}$$

The conditional probability function of Y given X

$$f\left(y/x\right) = \frac{f(x,y)}{f(x)}$$

Conditional Expectation: Let f(x,y) be the joint probability density function of X,Y and u(Y) is a function of y. Then the conditional expectation of u(Y), given X = x is

$$E[u(y)/x] = \int_{-\infty}^{\infty} u(y) f(y/x) dy$$

In particular,

Conditional Mean,

$$E[Y/x] = \int_{-\infty}^{\infty} y f(y/x) dy$$

Conditional Variance,

$$E\left[[Y - E(Y/x)]^2 / x\right] = E[Y^2/x] - E[Y/x]^2$$

Example (i) Find the marginal distributions given that the joint distribution of X and Y is,

$$f(x, y) = \frac{x+y}{21}$$

where X = 1, 2, 3; *Y* = 1, 2. Also determine the conditional mean and variance of X, given Y=1.

Solution:

X \ Y	1	2	
1	2/21	3/21	5/21
2	3/21	4/21	7/21
3	4/21	5/21	9/21
	9/21	12/21	1

- Marginal Distribution of X

$$P(X = x_i) = p_{i*} = \sum_j p_{ij}$$

X	1	2	3
p_{i*}	5/21	7/21	9/21

- Marginal Distribution of Y

$$P(Y = y_j) = p_{*j} = \sum_i p_{ij}$$

Y	1	2
p_{*j}	9/21	12/21

X	$f\left(\frac{y}{x}\right)=\frac{f(x,y)}{f(x)}$
1	$\frac{2/21}{5/21}=\frac{2}{5}$
2	$\frac{3/21}{5/21}=\frac{3}{5}$

- Conditional mean

$$E[Y/x]=\sum_{y=1}^{2} yf(y/x)=1\times f(1/1)+2\times f(2/1)=7/5$$

$$E[Y^2/x]=\sum_{y=1}^{2} y^2 f(y/x)=1^2\times f(1/1)+2^2\times f(2/1)=13/5$$

- Conditional Variance

$$E[Y^2/x]-E[y/x]^2=\frac{13}{5}-\frac{49}{25}=\frac{16}{25}$$

Example (ii) Joint probability mass function of X , Y is given by P(x, y) = k (2x + 3y) ; x = 0, 1, 2; y = 1, 2, 3. Find all the marginal and conditional probability and also find probability distribution of X + Y.

X \ Y	1	2	3	
0	3 k	6 k	9 k	18k
1	5 k	8 k	11 k	24k
2	7 k	10 k	13 k	30k
	15k	24k	33k	72k

Solution: Since $\sum_j \sum_i p_{ij} = 1$, k = 1/72

- Marginal Distribution of X $P(X = x_i) = p_{i*} = \sum_j p_{ij}$

X	0	1	2
p_{i*}	18/72	24/72	30/72

- Marginal Distribution of Y $P(Y = y_j) = p_{*j} = \sum_i p_{ij}$

Y	1	2	3
p_{*j}	15/72	24/72	33/72

Conditional probability distribution of X given Y = y_j

$$P\left(X = x_i \big/ Y = y_j\right) = \frac{p(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{*j}}$$

X	$P(X \le x_i/Y = 1)$	$P(X \le x_i/Y = 2)$	$P(X \le x_i/Y = 3)$
0	$\frac{3/72}{15/72} = \frac{3}{15}$	$\frac{6/72}{24/72} = \frac{6}{24}$	$\frac{9/72}{33/72} = \frac{9}{33}$
1	$\frac{5/72}{15/72} = \frac{5}{15}$	$\frac{8/72}{24/72} = \frac{8}{24}$	$\frac{11/72}{33/72} = \frac{11}{33}$
2	$\frac{7/72}{15/72} = \frac{7}{15}$	$\frac{10/72}{24/72} = \frac{10}{24}$	$\frac{13/72}{33/72} = \frac{13}{33}$

Conditional probability distribution of Y given X = x_i

$$P\left(Y = y_j / X = x_i \right) = \frac{p(X = x_i, Y = y_j)}{P(X = x_i)} = \frac{p_{ij}}{p_{i*}}$$

Y	P(Y ≤ x_i/X = 0)	P(Y ≤ x_i/X = 1)	P(Y ≤ x_i/X = 2)
1	$\frac{3/72}{18/72} = \frac{3}{18}$	$\frac{5/72}{24/72} = \frac{5}{24}$	$\frac{7/72}{30/72} = \frac{7}{30}$
2	$\frac{6/72}{18/72} = \frac{6}{18}$	$\frac{8/72}{24/72} = \frac{8}{24}$	$\frac{10/72}{30/72} = \frac{10}{30}$
3	$\frac{9/72}{18/72} = \frac{9}{18}$	$\frac{11/72}{24/72} = \frac{11}{24}$	$\frac{13/72}{30/72} = \frac{13}{30}$

- Probability distribution of X + Y

X + Y	Probabilities
1	3/72
2	11/72
3	24/72
4	21/72
5	13/72

Example(iii) Given joint probability density function

$$f(x, y) = \begin{cases} e^{-y}, & x > 0, y > x \\ 0, & otherwise \end{cases}$$

1. Find $P\left(X > 1 \middle/ Y < 5\right)$
2. Find the marginal density function of X and Y.

Solution: 1.

$$P\left(X>1 \big/ Y<5\right)=\frac{P(X>1,Y<5)}{P(Y<5)}$$

$P\left(X>1,Y<5\right)=\int_1^5\int_1^y e^{-y}dxdy$	$P(Y<5)=\int_0^5\int_0^y e^{-y}dxdy$
$=\int_1^5 e^{-y}\left[x\right]_1^y dy$	$=\int_0^5\left[e^{-y}x\right]_0^y dy$
$=\int_1^5\left[1-y\right]e^{-y}\,dy$	$=\int_0^5 e^{-y}y\,dy$
$0<1<x<y<5$ $\quad=-5e^{-5}+e^{-1}$	$0<x<y<5$ $\quad=-6e^{-5}+1$

$$\therefore\ P\left(X > 1 \big/ Y < 5\right) = \frac{P(X > 1, Y < 5)}{P(Y < 5)} = \frac{e^4 - 5}{e^5 - 6}$$

2. Marginal density function of X

$$f_X(x) = \int_x^{\infty} e^{-y}\,dy = e^{-x}$$

Marginal density function of Y

$$f_Y(y) = \int_0^{y} e^{-y}\,dx = ye^{-y}$$

Covariance: If X, Y is a two dimensional random variable then the co-variance of X and Y is denoted by C_{xy} and defined as

$$C_{xy} = E[(X - E(X))(Y - E(Y))] = E[XY] - E[X]\,E[Y]$$

Note: $C_{xy} \leq \sigma_x \sigma_y$

Correlation: The coefficient of correlation between X and Y is denoted by

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$$

where

$$\sigma_x^2 = E[X^2] - [E(X)]^2 = E[X - E(X)]^2, \qquad \sigma_y^2 = E[Y^2] - [E(Y)]^2 = E[Y - E(Y)]^2$$

Note: If X and Y are independent random variables, then

$$E(XY) = E(X)E(Y)$$

$$\therefore C_{xy} = 0 \qquad \therefore \rho_{xy} = 0$$

In a bivariate distribution if the change in one variable effects the change in other variable, the variable are said to be correlated. If the increase (or decrease) in one variable results in the corresponding increase (or decrease) then the correlation said to be positive correlation of direct correlation. If the increase (or decrease) in one variable results in the corresponding decrease (or increase) then the correlation said to be negative correlation of inverse correlation.

Example: Let the random variables X and Y have the joint probability function

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

Find the correlation coefficient.

Solution:

Correlation coefficient: $\rho_{xy} = \dfrac{C_{xy}}{\sigma_x \sigma_y}$

$$E[X] = \int_0^1 \int_0^1 x(x+y)dxdy = \frac{7}{12}, \qquad E[Y] = \int_0^1 \int_0^1 y(x+y)dxdy = \frac{7}{12}$$

$$E[X^2] = \int_0^1 \int_0^1 x^2(x+y)dxdy = \frac{5}{12}, \qquad E[Y^2] = \int_0^1 \int_0^1 y^2(x+y)dxdy = \frac{5}{12}$$

$$E[XY] = \int_0^1 \int_0^1 xy(x+y)dxdy = \frac{1}{3}$$

$$\sigma_X^2 = E[X^2] - E[X]^2 = \frac{5}{12} - \left[\frac{7}{12}\right]^2 = \frac{11}{144}$$

$$\sigma_Y^2 = E[X^2] - E[X]^2 = \frac{5}{12} - \left[\frac{7}{12}\right]^2 = \frac{11}{144}$$

- Covariance

$$Cov[X,Y] = E[XY] - E[X]E[Y] = \frac{1}{3} - \frac{49}{144} = -\frac{1}{144}$$

- Correlation Coefficient

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} = \frac{-1/144}{\sqrt{(11/144) \times (11/144)}} = -\frac{1}{11}$$

Reference Books:

1. Erwin Kreyszig, Advance Engineering Mathematics, 9th Edition, John Wiley & Sons, 2006.

2. Sheldon Ross, A first course in Probability, 8th Edition, Pearson Education India.

3. W. Feller An Introduction to Probability Theory and its Applications, Vol. 1, 3rd Edition, Wiley, 1968.

4. S. C. Gupta and V. . Kapoor, Fundamentals of Mathematical Statistics, Sultan Chand & Sons.

5. P. G. Hoel, S. C. Port and C. J. Stone, Introduction to Probability Theory, Universal Book Stall, 2003.

6. A. M. Mood, F. A. Graybill and D. C. Bose, Introduction to Theory of Statistics. 3rd Edition, Tata McGraw-Hill Publication.

Thank You