Expectation and Variance

Given a random variable, we often compute the expectation and variance, two important summary statistics. The expectation describes the average value and the variance describes the spread (amount of variability) around the expectation.

Let X be a random variable whose possible values $x_1, x_2, x_3, \ldots, x_n$ occur with probabilities $p_1, p_2, p_3, \ldots, p_n$, respectively. The mean of X, denoted by μ, is the number $\sum x_i p_i$ i.e. the **mean of X is the weighted average of the possible values of X**, each value being weighted by its probability with which it occurs. The mean of a random variable X is also called the expectation of X.
Thus, $\boldsymbol{\mu = E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n}$.

Definition: Let X be a **continuous** random variable with p.d.f. $f_X(x)$. The expected value of X is

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

Definition: Let X be a **discrete** random variable with probability function $f_X(x)$. The expected value of X is

$$\mathbb{E}(X) = \sum_x x f_X(x) = \sum_x x \mathbb{P}(X = x).$$

Properties of Expectation:

1. E(c) = c
   i.e E(c) = $\sum cP(X = c) = c.1 = c$
2. E(aX+b)= $\sum(aX + b)P(X = x) = a\sum xP(X = x) + b\sum P(X = x)$
   = aE(x)+b.1= aE(X)+b
3. E(E(X))= E(X)

Example Let a pair of dice be thrown and the random variable X be the sum of the numbers that appear on the two dice. Find the mean or expectation of X.
Solution The sample space of the experiment consists of 36 elementary events in the form of ordered pairs (x_i , y_i), where x_i = 1, 2, 3, 4, 5, 6 and y_i = 1, 2, 3, 4, 5, 6. The random variable X i.e. the sum of the numbers on the two dice takes the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12.

X or x_i	2	3	4	5	6	7	8	9	10	11	12
P(X) or p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Now $\mu = 7$ Thus, the mean of the sum of the numbers that appear on throwing two fair dice is 7.

Variance

The mean of the random variable does not give us information about the variability in the values of the random variable. In fact, if the variance is small, then the values of the random variable are close to the mean. Also random variables with different probability distributions can have equal means.

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$\sigma_x^{\,2} = \operatorname{Var}(X) = \sum_{i=1}^{n}(x_i - \mu)^2\, p(x_i)$$

$$= \sum_{i=1}^{n} x_i^{\,2}\, p(x_i) + \sum_{i=1}^{n} \mu^2\, p(x_i) - \sum_{i=1}^{n} 2\mu x_i\, p(x_i)$$

$$= \sum_{i=1}^{n} x_i^{\,2}\, p(x_i) + \mu^2 \sum_{i=1}^{n} p(x_i) - 2\mu \sum_{i=1}^{n} x_i\, p(x_i)$$

$$= \sum_{i=1}^{n} x_i^{\,2}\, p(x_i) + \mu^2 - 2\mu^2 \left[\text{since} \sum_{i=1}^{n} p\,(x_i) = 1 \text{ and } \mu = \sum_{i=1}^{n} x_i\, p(x_i)\right]$$

$$= \sum_{i=1}^{n} x_i^{\,2}\, p(x_i) - \mu^2$$

$$\operatorname{Var}(X) = \sum_{i=1}^{n} x_i^{\,2}\, p(x_i) - \left(\sum_{i=1}^{n} x_i\, p(x_i)\right)^2$$

$$\operatorname{Var}(X) = E(X^2) - [E(X)]^2, \text{ where } E(X^2) = \sum_{i=1}^{n} x_i^{\,2}\, p(x_i)$$

Example Find the variance of the number obtained on a throw of an unbiased die.

Solution The sample space of the experiment is S = {1, 2, 3, 4, 5, 6}.
Let X denote the number obtained on the throw. Then X is a random variable which can take values 1, 2, 3, 4, 5, or 6.

$$E(X^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} = \frac{91}{6}$$

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

$$= \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{91}{6} - \frac{441}{36} \quad \frac{35}{12}$$

Properties of Variance:

1. **V(c) =0**
   i.e V(c) = E(c^2) –[E(c)]^2= $c^2 - [c]^2 = 0$
2. **V(aX+b) = $a^2 V(X)$**
   WKT, V(X)= E(X^2) –[E(X)]^2
   V(aX+b)= E({aX + b}^{2}) –[E(aX + b)]^2
   = E($a^2x^2 + b^2 + 2abX$) –{aE(X) + b}^{2}
   = $a^2E(X^2) + b^2 + 2abE(X) - [a^2\{E(X)\}^2 + b^2 + 2abE(X)]$
   = a^2\{E(X^2) –[E(X)]^2\} – 0
   = $a^2V(X)$

Example: Let X be a continuous random variable with p.d.f.

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

Find $\mathbb{E}(X)$ and $\text{Var}(X)$.

$$\begin{aligned} E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_1^2 x \times 2x^{-2}\, dx \; &= \; \int_1^2 2x^{-1}\, dx \\ &= \Big[2\log(x)\Big]_1^2 \\ &= 2\log(2) - 2\log(1) \\ &= 2\log(2). \end{aligned}$$

For $\text{Var}(X)$, *we use*

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

Now

$$\begin{aligned} E(X^2) = \int_{-\infty}^{\infty} x^2\, f_X(x)\, dx = \int_1^2 x^2 \times 2x^{-2}\, dx \; &= \; \int_1^2 2\, dx \\ &= \Big[2x\Big]_1^2 \\ &= 2 \times 2 - 2 \times 1 \\ &= 2. \end{aligned}$$

Thus

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X^2) - \{\mathbb{E}(X)\}^2 \\ &= 2 - \{2\log(2)\}^2 \\ &= 0.0782. \end{aligned}$$

Mean, Median and MODE

If X is A CRV then median M, $\int_{-\infty}^{M} f(x)dx = \int_{M}^{\infty} f(x)\,dx = \frac{1}{2}$
And MODE of X for which f(x) is maximum, $f^{|} = 0$ and $f^{||} < 0$.

Problems:

1. Find mean, Median and mode also variance of a random variable X having pdf f(x)= $\begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & otherwise \end{cases}$.
   Ans: E(x)= ½ , E(x^2)= 3/10, V(x)= 1/20, M=½, MODE= ½

2. Find pdf, mean, Median and mode also variance of a random variable X having cdf F(x)= $\begin{cases} 1 - e^{-x} - xe^{-x} & x \ge 0 \\ 0 & otherwise \end{cases}$.
   Ans: E(x)= 2 , E(x^2)= 6, V(x)= 2, M= 5/3, MODE=1

   Formulas: MODE= 3M-2E(X) AND $\int UV$ PARTS: $\int UV = U \int V - \int U' \int V$

3. If $F(x) = \begin{cases} -e^{-\frac{x^2}{2}} & x > 0 \\ 0 & x \le 0 \end{cases}$ then find V(X).

   Solution:

   The pdf of X is $f(x) = \begin{cases} xe^{-\frac{x^2}{2}} & x > 0 \\ 0 & x \le 0 \end{cases}$,

   E(X) = $\int_0^\infty x^2\ e^{-\frac{x^2}{2}}\,dx = \sqrt{2}\int_0^\infty \sqrt{t}\, e^{-t}\,dt$ using $\frac{x^2}{2} = t,\ x\,dx = dt$

   WKT, $\Gamma(z) = \int_0^\infty t^{z-1}\, e^{-t}\, dt$ and $\Gamma(n+1) = n\,\Gamma(n)$, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

   E(X)= $\sqrt{2}\,\Gamma(3/2) = \sqrt{2}\,.\,\frac{1}{2}.\,\sqrt{\pi} = \frac{\sqrt{\pi}}{\sqrt{2}}$

   E(X^2)= $\int_0^\infty x^3\ e^{-\frac{x^2}{2}}\,dx = 2\int_0^\infty te^{-t}\,dt$ using $\frac{x^2}{2}$ =t

   = 2 $\Gamma(2)$ =2

   V(X) = 2 - $\left(\sqrt{\pi}\frac{1}{\sqrt{2}}\right)^2 = 4 - \frac{\pi}{2}$

Uniform distribution

Let X be a continuous random variable assuming all values in the interval [a, b] where a and b are finite. If the pdf of X is given by

$$f(x) = \begin{cases} \frac{1}{(b-a)} & a \leq x \leq b \\ 0 & else\ where \end{cases}$$

then we say that X has uniform distribution defined over [a, b].

Note that, for any sub interval [c, d],

$$\mathrm{P}\ (\mathrm{c} < \mathrm{X} < \mathrm{d}) = \int_c^d f(x)dx = \int_c^d \frac{1}{(b-a)} dx = \frac{(d-c)}{(b-a)}$$

$$\textbf{Cdf} = F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{(b-a)} & a < x < b \\ 1 & x \ge b \end{cases}$$

$$\textbf{Mean } E(X) = \int_a^b x f(x)\,dx = \frac{1}{(b-a)}\left\{\frac{x^2}{2}\right\}_a^b = \frac{(b-a)(a+b)}{2(b-a)} = \frac{(a+b)}{2}$$

$$E(X^2) = \int_a^b x^2 f(x)\,dx = \frac{1}{(b-a)}\left\{\frac{x^3}{3}\right\}_a^b = \frac{(b^3 - a^3)}{3(b-a)} = \frac{(a^2 + b^2 + ab)}{3}$$

$$\textbf{Variance } V(X) = E(X^2) - [E(X)]^2 = \frac{(a^2 + b^2 + ab)}{3} - \left\{\frac{(a+b)}{2}\right\}^2 = \frac{(b-a)^2}{12}$$

Problems:

1. If X is uniformly distributed over (-2, 2) then find i) P(X<1) ii)P $\left(|X-1| \geq \frac{1}{2}\right)$.

Solution: Given that $X \in U(-2,2)$.

Therefore, $f(x) = \begin{cases} \frac{1}{4} & -2 \leq x \leq 2 \\ 0 & else\ where \end{cases}$

i) P(X<1) = $\int_{-\infty}^{1} f(x)dx = \int_{-2}^{1} \frac{1}{4}\, dx = \frac{3}{4}$

ii) P $\left(|X-1| \geq \frac{1}{2}\right)$ = 1 - P $\left(|X-1| < \frac{1}{2}\right)$

= 1- P($-\frac{1}{2}$ < X – 1< $\frac{1}{2}$)

= 1 – P($\frac{1}{2}$ < X < $\frac{3}{2}$)

= 1 - $\int_{\frac{1}{2}}^{\frac{3}{2}} \frac{1}{4}\, dx = 1 - \frac{1}{4} = \frac{3}{4}$

2. If K is uniformly distributed over (0, 5) then what is the probability that the roots of the equation $4x^2 + 4xK + K + 2 = 0$ are real ?

Solution: Given that K$\in U(0,5)$.

Therefore, $f(k) = \begin{cases} \frac{1}{5} & 0 \le k \le 5 \\ 0 & else\,where \end{cases}$

P { Roots are real} = P{ $(4\text{K})^2 - 4.4(k+2) \ge 0$ }

= P{ $K^2 - K - 2 \ge 0$ } = P{ $(K+1)(K-2) \ge 0$ }

= P{ $(K+1) \ge 0, (K-2) \ge 0$ or $(K+1) \le 0, (K-2) \le 0$}

= P{ $K \ge -1, K \ge 2$ or $K \le -1, K \le 2$ }

= P{ $K \ge 2$ or $K \le -1$ }

= P{ $K \ge 2$ } + P{ $K \le -1$ }

$= \int_2^5 \frac{1}{5} dx + \int_{-\infty}^{-1} 0\, dx = 3/5$

Problems on Variance and Expectation:

1. A student takes a multiple choice test consisting of 2 problems. The first one has 3 possible answers and the second one has 5. The student chosen at random, one answer as the right answer for each of the 2 problems. Let X denote the number of right answers of student. Find V(X).

Solution: Let X : the number of right answers

X:	0	1	2
P(X=x):	8/15	6/15	1/15

{P(X =0) = 2/3 . 4/5 and P(X =1) = 1/3 . 4/5 + 2/3.1/5 = 6/15 and P(X =2) = 1/3 . 1/5 =1/15}

E(X) = 0 + 1.(6/15) + 2. (1/15) = 8/15, E(X^2) = 0 + 6/15 + 4.(1/15) =10/15 = 2/3 and

V(X) = E(X^2) – $\{E(X)\}^2$ = 2/3 – $\{8/15\}^2$ = 86/225=0.3822

2. Three balls are randomly selected from an urn containing 3 white, 3 red, 5 black balls. The person who selects the ball wins \$ 1.00 for each white ball selected and lose \$1.00 for each red ball selected. Let X be the total winnings from the experiment. Find the Probability distribution of X and V(X).

Solution: Let X : total winnings

X:	-3	-2	-1	0	1	2	3
P(X = x):	1/165	15/165	39/165	55/165	39/165	15/165	1/165

{P(X =0) = P{Selection of 3B or 1W, 1B, 1R balls} = $\frac{5C_3}{11C_3} + \frac{3C_1 \times 3C_1 \times 5C_1}{11C_3}$ = 55/165

P(X =1) = P(X = -1) = P{Selection of 2B, 1W or 2W, 1R balls} = $\frac{3C_1 \times 3C_2}{11C_3} + \frac{3C_1 \times 5C_2}{11C_3}$ = 39/165

P(X =2) = P(X = - 2) = P{Selection of 1B, 2W balls} = $\frac{5C_1 \times 3C_2}{11C_3}$ = 15/165

and P(X =3) =P(X = -3) = P{Selection of 3W balls} = $\frac{3C_3}{11C_3}$ = 1/165}

E(X) = 0 , E(X^2) = 2{9/165 + 15X4/165 + 39/165 }=216/165 and V(X)= 216/165

3. A coin is tossed till head appears then find the probability distribution on number of tosses. Let X denote the number of tosses. Find E(X).

Solution:

X denote the number of tosses

P(H)= p

P(T)= q

X	**1**	**2**	**3**	**…**
P(X)	p	qp	q^2p	…

Therefore, P(X=k)= pq^{k-1}

Hence, E(X) = $\sum_1^\infty k\ pq^{k-1}$ = p $\sum_1^\infty k\ q^{k-1}$= $\frac{p}{(1-q)^2} = \frac{1}{p}$

4. Suppose that an electronic device has a life length X (in units' of 1000 hours) which is considered as a continuous random variable with the following pdf: f(x) = e^{-x}, x > 0, Suppose that the cost of manufacturing one such item is 2 rupees. The manufacturer sells the item for 5 rupees but guarantees a total refund if $x \le 0.9$. What is the manufacturer's expected profit per item?

Solution:

Y- expected profit

Profit, p= $\begin{cases} 3 & x > 0.9 \\ -2 & x \le 0.9 \end{cases}$

E(Y)= 3 (probability of x>0.9) +(-2) (probability of $x \le 0.9$)

E(Y)= 3 ($\int_{0.9}^{\infty} e^{-x}\, dx$) +(-2) ($\int_{-\infty}^{0.9} e^{-x}\, dx$)

E(Y)= 3 ($\int_{0.9}^{\infty} e^{-x}\, dx$) +(-2) (1-$\int_{0.9}^{\infty} e^{-x}\, dx$)

E(Y)= 3 ($\int_{0.9}^{\infty} e^{-x}\, dx$) +(-2) (1-$\int_{0.9}^{\infty} e^{-x}\, dx$)

E(Y)= 3 (0.4066) +(-2) (1-0.466)

E(Y)=0.033

5. Let X be a random variable with probability function

$$P(X = k) = p\,(1-p)^{k-1}\ , k = \ 1,2,3 \ldots n. \text{ Find V(X).}$$

Solution: Given $P(X = k) = p\,(1-p)^{k-1}\ , k = 1,2,3 \ldots n$

E(X) = $\sum_1^n x\, p(x) = \sum_1^n k\, p(1-p)^{k-1}$

= p{ 1 + 2 (1-p) + 3 $(1-p)^2$+….}

= p . $\frac{1}{(1-(1-p))^2}$ = 1/p

P(H)is not equal to P(T). P(H) =p, P(T) = 1-p

1+ x + $(x)^2$ +……..=1/1-x

1+2x+3$(x)^2$+ ….= $1/(1-x)^2$

X	1	2	3	…	N
P(X=k)	p	P(1-p)	(1-p)(1-p)p	…	(1-p) (1-p)….. (1-p) (1-p)p

$\sum_1^n p(x)$ =p [1+ (1-p)+$(1-p)^2$+….] =p/1-(1-p)=1

$E(X^2)= \sum_1^n x^2\, p(x) \;= \sum_1^n k^2\, p(1-p)^{k-1}$ = p{ 1 + 4 (1-p) + 9 $(1-p)^2$+….} =pS

Where S = { 1 + 4 (1-p) + 9 $(1-p)^2$+….} ……………..(1)

Multi ply (1) by 1-p we get,

S (1-p) = { 1-p + 4 $(1-p)^2$+ 9 $(1-p)^3$+….} …………(2)

Eq(1)- Eq (2) simplifies to Sp = {1 +3(1-p) + 5 $(1-p)^2$ +………..} …………………..(3)

Multi ply (3) by 1-p we get,

Sp (1-p) = {1-p +3$(1-p)^2$ + 5 $(1-p)^3$ +………..} …………………..(4)

Eq(3)- Eq (4) simplifies to

Sp^2 = {1 +2(1-p) + 2 $(1-p)^2$ +………..} …………………..(3)

Therefore, Sp^2 = {1 +2(1-p) { 1+ (1-p) +$(1-p)^2$ +………..} }

S = $\frac{1}{p^2}$ [1+ 2(1-p) $\frac{1}{1-(1-p)}$] =$\frac{2}{p^3} - \frac{1}{p^2}$

$E(X^2)$= p($\frac{2}{p^3} - \frac{1}{p^2}$) = $\frac{2}{p^2} - \frac{1}{p}$

V(X) = $\frac{1-p}{p^2}$

Two-Dimensional Random Variable

Let S be the sample space associated with a random experiment E. Let X=X(S) and Y=Y(S) be two functions each assigning a real number to each s∈S. Then (X,Y) is called a two dimensional random variable.

Joint Probability distribution function

Definition. (a) Let (X, Y) be a two-dimensional discrete random variable. With each possible outcome (x_i, y_j) we associate a number $p(x_i, y_j)$ representing $P(X = x_i, Y = y_j)$ and satisfying the following conditions:

$$(1) \quad p(x_i, y_j) \geq 0 \quad \text{for all} \quad (x, y),$$

$$(2) \quad \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} p(x_i, y_j) = 1. \tag{6.1}$$

Joint Probability density function

(b) Let (X, Y) be a continuous random variable assuming all values in some region R of the Euclidean plane. The *joint probability density function* f is a function satisfying the following conditions:

$$(3) \quad f(x, y) \geq 0 \quad \text{for all} \quad (x, y) \in R,$$

$$(4) \quad \iint_R f(x, y)\, dx\, dy = 1. \tag{6.2}$$

Joint Cumulative distribution function

For two dimentional random variable (X, Y) the CDF F(x,y) is defined as

$$\mathbf{F(x, y)= P(X\leq x, Y \leq y)}$$

For Discrete RV: $F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(x_i, y_j)$

For Continuous RV: $F(x, y)= \int_{-\infty}^{x} \int_{-\infty}^{y} f(x, y)\,dy\,dx$

Note:

If F is the cdf of a two-dimensional random variable with joint pdf f, then

$$\partial^2 F(x, y)/\partial x\, \partial y = f(x, y)$$

The Marginal probability functions.

For Discrete RV:

In the *discrete* case we proceed as follows: Since $X = x_i$ must occur with $Y = y_j$ for some j and can occur with $Y = y_j$ for only one j, we have

$$p(x_i) = P(X = x_i) = P(X = x_i,\ Y = y_1 \text{ or } X = x_i,\ Y = y_2 \text{ or } \cdots)$$
$$= \sum_{j=1}^{\infty} p(x_i, y_j).$$

$$q(y_j) = P(Y = Y_j) = \sum_{i=1}^{\infty} p(x_i, y_j)$$

Y \ X	0	1	2	3	4	5	Sum
0	0	0.01	0.03	0.05	0.07	0.09	0.25
1	0.01	0.02	0.04	0.05	0.06	0.08	0.26
2	0.01	0.03	0.05	0.05	0.05	0.06	0.25
3	0.01	0.02	0.04	0.06	0.06	0.05	0.24
Sum	0.03	0.08	0.16	0.21	0.24	0.28	1.00

For Continuous RV:

In the *continuous* case we proceed as follows: Let f be the joint pdf of the continuous two-dimensional random variable (X, Y). We define g and h, the *marginal probability density functions* of X and Y, respectively, as follows:

$$g(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy; \qquad h(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx.$$

These pdf's correspond to the basic pdf's of the one-dimensional random variables X and Y, respectively. For example

$$P(c \le X \le d) = P[c \le X \le d, -\infty < Y < \infty]$$
$$= \int_c^d \int_{-\infty}^{+\infty} f(x, y)\, dy\, dx$$
$$= \int_c^d g(x)\, dx.$$

EXAMPLE 6.1. Two production lines manufacture a certain type of item. Suppose that the capacity (on any given day) is 5 items for line I and 3 items for line II. Assume that the number of items actually produced by either production line is a random variable. Let (X, Y) represent the two-dimensional random variable yielding the number of items produced by line I and line II, respectively. Table 6.1 gives the joint probability distribution of (X, Y). Each entry represents

$$p(x_i, y_j) = P(X = x_i, Y = y_j).$$

TABLE 6.1

$Y \backslash X$	0	1	2	3	4	5
0	0	0.01	0.03	0.05	0.07	0.09
1	0.01	0.02	0.04	0.05	0.06	0.08
2	0.01	0.03	0.05	0.05	0.05	0.06
3	0.01	0.02	0.04	0.06	0.06	0.05

Thus $p(2, 3) = P(X = 2, Y = 3) = 0.04$, etc. Hence if B is defined as

$$B = \{\text{More items are produced by line I than by line II}\}$$

we find that

$$\begin{aligned} P(B) &= 0.01 + 0.03 + 0.05 + 0.07 + 0.09 + 0.04 + 0.05 + 0.06 \\ &\qquad + 0.08 + 0.05 + 0.05 + 0.06 + 0.06 + 0.05 \\ &= 0.75. \end{aligned}$$

EXAMPLE 6.5. Two characteristics of a rocket engine's performance are thrust X and mixture ratio Y. Suppose that (X, Y) is a two-dimensional continuous random variable with joint pdf:

$$\begin{aligned} f(x, y) &= 2(x + y - 2xy), \qquad 0 \le x \le 1, \quad 0 \le y \le 1, \\ &= 0, \qquad \text{elsewhere.} \end{aligned}$$

(The units have been adjusted in order to use values between 0 and 1.) The marginal pdf of X is given by

$$\begin{aligned} g(x) = \int_0^1 2(x + y - 2xy)\, dy &= 2(xy + y^2/2 - xy^2)|_0^1 \\ &= 1, \qquad 0 \le x \le 1. \end{aligned}$$

That is, X is uniformly distributed over $[0, 1]$.

The marginal pdf of Y is given by

$$\begin{aligned} h(y) = \int_0^1 2(x + y - 2xy)\, dx &= 2(x^2/2 + xy - x^2y)|_0^1 \\ &= 1, \qquad 0 \le y \le 1. \end{aligned}$$

Hence Y is also uniformly distributed over $[0, 1]$.

Definition. We say that the two-dimensional continuous random variable is *uniformly distributed* over a region R in the Euclidean plane if

$$f(x, y) = \text{const} \quad \text{for } (x, y) \in R,$$
$$= 0, \quad \text{elsewhere.}$$

Because of the requirement $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1$, the above implies that the constant equals $1/\text{area}(R)$. We are assuming that R is a region with finite, nonzero area.

Note: This definition represents the two-dimensional analog to the one-dimensional uniformly distributed random variable.

EXAMPLE 6.6. Suppose that the two-dimensional random variable (X, Y) is *uniformly distributed* over the shaded region R indicated in Fig. 6.5. Hence

$$f(x, y) = \frac{1}{\text{area}(R)}, \qquad (x, y) \in R.$$

We find that

$$\text{area}(R) = \int_0^1 (x - x^2)\, dx = \tfrac{1}{6}.$$

Therefore the pdf is given by

$$f(x, y) = 6, \qquad (x, y) \in R$$
$$= 0, \qquad (x, y) \notin R.$$

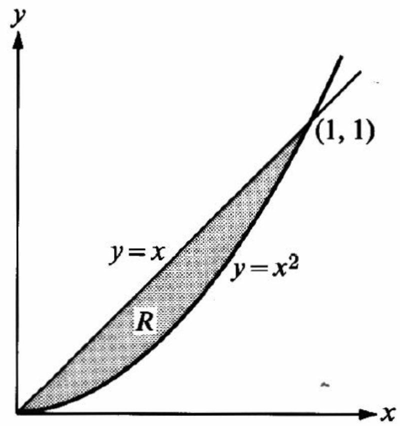

FIGURE 6.5

In the following equations we find the marginal pdf's of X and Y.

$$g(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy = \int_{x^2}^{x} 6\, dy$$
$$= 6(x - x^2), \qquad 0 \le x \le 1;$$
$$h(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx = \int_{y}^{\sqrt{y}} 6\, dx$$
$$= 6(\sqrt{y} - y), \qquad 0 \le y \le 1.$$

Conditional probability function:

$$p(x_i \mid y_j) = P(X = x_i \mid Y = y_j) = \frac{p(x_i, y_j)}{q(y_j)} \quad \text{if} \quad q(y_j) > 0,$$

$$q(y_j \mid x_i) = P(Y = y_j \mid X = x_i) = \frac{p(x_i, y_j)}{p(x_i)} \quad \text{if} \quad p(x_i) > 0.$$

Y \ X	0	1	2	3	4	5	Sum
0	0	0.01	0.03	0.05	0.07	0.09	0.25
1	0.01	0.02	0.04	0.05	0.06	0.08	0.26
2	0.01	0.03	0.05	0.05	0.05	0.06	0.25
3	0.01	0.02	0.04	0.06	0.06	0.05	0.24
Sum	0.03	0.08	0.16	0.21	0.24	0.28	1.00

evaluate the conditional probability $P(X = 2 \mid Y = 2)$. According to the definition of conditional probability we have

$$P(X = 2 \mid Y = 2) = \frac{P(X = 2, Y = 2)}{P(Y = 2)} = \frac{0.05}{0.25} = 0.20.$$

Definition. Let (X, Y) be a continuous two-dimensional random variable with joint pdf f. Let g and h be the marginal pdf's of X and Y, respectively.

The *conditional* pdf of X for given $Y = y$ is defined by

$$g(x \mid y) = \frac{f(x, y)}{h(y)}, \qquad h(y) > 0. \tag{6.7}$$

The *conditional* pdf of Y for given $X = x$ is defined by

$$h(y \mid x) = \frac{f(x, y)}{g(x)}, \qquad g(x) > 0. \tag{6.8}$$

EXAMPLE 6.3. Suppose that the two-dimensional continuous random variable (X, Y) has joint pdf given by

$$f(x, y) = x^2 + \frac{xy}{3}, \qquad 0 \le x \le 1, \quad 0 \le y \le 2,$$
$$= 0, \qquad \text{elsewhere.}$$

$$g(x) = \int_0^2 \left(x^2 + \frac{xy}{3}\right) dy = 2x^2 + \frac{2}{3}x,$$

$$h(y) = \int_0^1 \left(x^2 + \frac{xy}{3}\right) dx = \frac{y}{6} + \frac{1}{3}.$$

Hence,

$$g(x \mid y) = \frac{x^2 + xy/3}{1/3 + y/6} = \frac{6x^2 + 2xy}{2 + y}, \qquad 0 \le x \le 1, \quad 0 \le y \le 2;$$

$$h(y \mid x) = \frac{x^2 + xy/3}{2x^2 + 2/3(x)} = \frac{3x^2 + xy}{6x^2 + 2x} = \frac{3x + y}{6x + 2},$$

$$0 \le y \le 2, \quad 0 \le x \le 1.$$

To check that $g(x \mid y)$ is a pdf, we have

$$\int_0^1 \frac{6x^2 + 2xy}{2 + y}\, dx = \frac{2 + y}{2 + y} = 1 \qquad \text{for all } y.$$

A similar computation can be carried out for $h(y \mid x)$.

Independent Random Variable

Just as we defined the concept of independence between two events A and B, we shall now define *independent random variables*. Intuitively, we intend to say that X and Y are independent random variables if the outcome of X, say, in no way influences the outcome of Y. This is an extremely important notion and there are many situations in which such an assumption is justified.

Definition. (a) Let (X, Y) be a two-dimensional discrete random variable. We say that X and Y are independent random variables if and only if $p(x_i, y_j) = p(x_i)q(y_j)$ for all i and j. That is, $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$, for all i and j.

(b) Let (X, Y) be a two-dimensional continuous random variable. We say that X and Y are independent random variables if and only if $f(x, y) = g(x)h(y)$ for all (x, y), where f is the joint pdf, and g and h are the marginal pdf's of X and Y, respectively.

EXAMPLE 6.10. Suppose that a machine is used for a particular task in the morning and for a different task in the afternoon. Let X and Y represent the number of times the machine breaks down in the morning and in the afternoon, respectively. Table 6.3 gives the joint probability distribution of (X, Y).

An easy computation reveals that for *all* the entries in Table 6.3 we have

$$p(x_i, y_j) = p(x_i)q(y_j).$$

TABLE 6.3

$Y \backslash X$	0	1	2	$q(y_j)$
0	0.1	0.2	0.2	0.5
1	0.04	0.08	0.08	0.2
2	0.06	0.12	0.12	0.3
$p(x_i)$	0.2	0.4	0.4	1.0

EXAMPLE 6.11. Let X and Y be the life lengths of two electronic devices. Suppose that their joint pdf is given by

$$f(x, y) = e^{-(x+y)}, \qquad x \geq 0, \quad y \geq 0.$$

Since we can factor $f(x, y) = e^{-x}e^{-y}$, the independence of X and Y is established.

PROBLEMS

6.1. Suppose that the following table represents the joint probability distribution of the discrete random variable (X, Y). Evaluate all the marginal and conditional distributions.

$Y \backslash X$	1	2	3
1	$\frac{1}{12}$	$\frac{1}{6}$	0
2	0	$\frac{1}{9}$	$\frac{1}{5}$
3	$\frac{1}{18}$	$\frac{1}{4}$	$\frac{2}{15}$

6.2. Suppose that the two-dimensional random variable (X, Y) has joint pdf

$$\begin{aligned} f(x, y) &= kx(x - y), \quad 0 < x < 2, \quad -x < y < x, \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

(a) Evaluate the constant k.
(b) Find the marginal pdf of X.
(c) Find the marginal pdf of Y.

6.3. Suppose that the joint pdf of the two-dimensional random variable (X, Y) is given by

$$\begin{aligned} f(x, y) &= x^2 + \frac{xy}{3}, \quad 0 < x < 1, \quad 0 < y < 2, \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

Compute the following.

(a) $P(X > \frac{1}{2})$; (b) $P(Y < X)$; (c) $P(Y < \frac{1}{2} \mid X < \frac{1}{2})$.

1. Check f(x,y) is a pdf. 2. Marginal pdfs 3. P(X+Y >=1) 4. g(X/Y) 5. h(y/x)
Ans: 1. yes 2. 2x^2+(2/3)x and (1/3)+(y/6) 3. 65/72 4. 6x^2+2xy/(2+y) 5. 3x^2+xy/(6x^2+2x)

6.5. For what value of k is $f(x, y) = ke^{-(x+y)}$ a joint pdf of (X, Y) over the region $0 < x < 1,\ 0 < y < 1$?

6.6. Suppose that the continuous two-dimensional random variable (X, Y) is uniformly distributed over the square whose vertices are $(1, 0)$, $(0, 1)$, $(-1, 0)$, and $(0, -1)$. Find the marginal pdf's of X and of Y.

6.14. Suppose that the joint pdf of (X, Y) is given by

$$\begin{aligned} f(x, y) &= e^{-y}, \quad \text{for } x > 0, \quad y > x, \\ &= 0, \quad \text{elsewhere.} \end{aligned}$$

(a) Find the marginal pdf of X.
(b) Find the marginal pdf of Y.
(c) Evaluate $P(X > 2 \mid Y < 4)$.

Answers:

6.2. (a) $k = \frac{1}{8}$ (b) $h(x) = x^3/4,\ 0 < x < 2$

(c) $g(y) = \begin{cases} \frac{1}{3} - y/4 + y^3/48, & 0 \le y \le 2 \\ \frac{1}{3} - y/4 + (5/48)y^3, & -2 \le y \le 0 \end{cases}$

6.3. (a) $\frac{5}{6}$ (b) $\frac{7}{24}$ (c) $\frac{5}{32}$ **6.5.** $k = 1/(1 - e^{-1})^2$

6.6. (a) $k = \frac{1}{2}$ (b) $h(x) = 1 - |x|,\ -1 < x < 1$

(c) $g(y) = 1 - |y|,\ -1 < y < 1$

6.14. (a) $g(x) = e^{-x},\ x > 0$ (b) $h(y) = ye^{-y},\ y > 0$

Problems:

1. Find C for which f(x,y)= Cx+Cy^2. $Ans: C = \frac{1}{37}$
2. If f(x,y)= $\begin{cases} \frac{2}{a^2} & 0 \le x \le y \le a \\ & elsewhere \end{cases}$. Find f(y/x) and f(x/y).

   Ans: $\frac{1}{a-x}$ and $\frac{1}{y}$
3. If f(x,y)= $\begin{cases} 8xy & 0 < x < y < 1. \\ 0 & elsewhere \end{cases}$ Find the marginal pdf of X and Y. Check whether they are independent.

   Ans: 4x- $4x^3, 4\ y^3$ and are not independent.

EXAMPLE 6.2. Suppose that a manufacturer of light bulbs is concerned about the number of bulbs ordered from him during the months of January and February. Let X and Y denote the number of bulbs ordered during these two months, respectively. We shall assume that (X, Y) is a two-dimensional continuous random

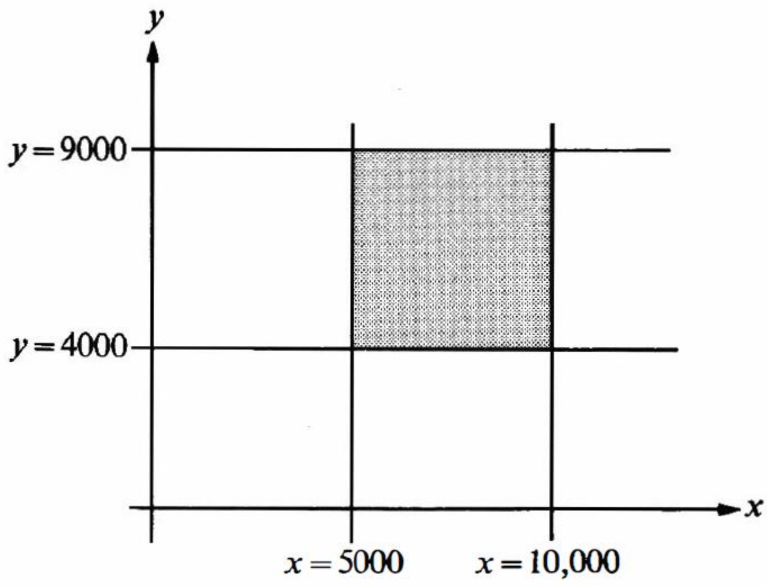

FIGURE 6.3

variable with the following joint pdf (see Fig. 6.3):

$$f(x, y) = c \quad \text{if} \quad 5000 \le x \le 10{,}000 \quad \text{and} \quad 4000 \le y \le 9000,$$
$$= 0, \quad \text{elsewhere.}$$

To determine c we use the fact that $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\,dx\,dy = 1$. Therefore

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\,dx\,dy = \int_{4000}^{9000} \int_{5000}^{10{,}000} f(x, y)\,dx\,dy = c[5000]^2.$$

Thus $c = (5000)^{-2}$. Hence if $B = \{X \ge Y\}$, we have

$$P(B) = 1 - \frac{1}{(5000)^2} \int_{5000}^{9000} \int_{5000}^{y} dx\,dy$$
$$= 1 - \frac{1}{(5000)^2} \int_{5000}^{9000} [y - 5000]\,dy = \frac{17}{25}.$$

EXAMPLE 6.3. Suppose that the two-dimensional continuous random variable (X, Y) has joint pdf given by

$$f(x, y) = x^2 + \frac{xy}{3}, \qquad 0 \le x \le 1, \quad 0 \le y \le 2,$$
$$= 0, \qquad \text{elsewhere.}$$

To check that $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1$:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, dx\, dy = \int_0^2 \int_0^1 \left(x^2 + \frac{xy}{3}\right) dx\, dy$$
$$= \int_0^2 \frac{x^3}{3} + \frac{x^2 y}{6}\bigg|_{x=0}^{x=1} dy$$
$$= \int_0^2 \left(\frac{1}{3} + \frac{y}{6}\right) dy = \frac{1}{3}y + \frac{y^2}{12}\bigg|_0^2$$
$$= \tfrac{2}{3} + \tfrac{4}{12} = 1.$$

Let $B = \{X + Y \ge 1\}$. (See Fig. 6.4.) We shall compute $P(B)$ by evaluating $1 - P(\overline{B})$, where $\overline{B} = \{X + Y < 1\}$. Hence

Let $B = \{X + Y \ge 1\}$. (See Fig. 6.4.) We shall compute $P(B)$ by evaluating $1 - P(\overline{B})$, where $\overline{B} = \{X + Y < 1\}$. Hence

$$P(B) = 1 - \int_0^1 \int_0^{1-x} \left(x^2 + \frac{xy}{3}\right) dy\, dx$$
$$= 1 - \int_0^1 \left[x^2(1 - x) + \frac{x(1 - x)^2}{6}\right] dx$$
$$= 1 - \tfrac{7}{72} = \tfrac{65}{72}.$$

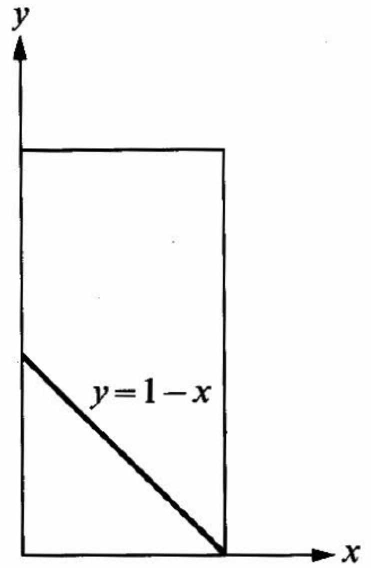

FIGURE 6.4

Expectation of 2D RV

For discrete RV:

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

$= \sum_{i=1}^{\infty} x_i\ \{\sum_{j=1}^{\infty}\ P\ (x_i, y_i)\}$

$= \sum_{i=1}^{\infty} x_i\ p(x_i)\}$ where $p(x_i)$ is marginal pmf of x.

E(X) $= \sum_{i=1}^{\infty} x_i\ p(x_i)\}$ and E(Y) $= \sum_{i=1}^{\infty} y_j\ q(y_i)\}$

For continuous RV:

E(X) $= \int_{-\infty}^{\infty} x$ g(x) dx and E(Y)= $\int_{-\infty}^{\infty} y$ h(y) dy

Properties:

1. E(c)= c
2. E(cX)= cE(x)
3. E(X+Y)= E(X)+E(Y)

Property 7.6. Let (X, Y) be a two-dimensional random variable and suppose that X and Y are *independent*. Then $E(XY) = E(X)E(Y)$.

Proof

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyg(x)h(y)\, dx\, dy$$

$$= \int_{-\infty}^{+\infty} xg(x)\, dx \int_{-\infty}^{+\infty} yh(y)\, dy = E(X)E(Y).$$

Property 7.9. If (X, Y) is a two-dimensional random variable, and if X and Y are *independent* then

$$V(X + Y) = V(X) + V(Y). \tag{7.15}$$

Proof

$$\begin{aligned} V(X+Y) &= E(X+Y)^2 - (E(X+Y))^2 \\ &= E(X^2 + 2XY + Y^2) - (E(X))^2 - 2E(X)E(Y) - (E(Y))^2 \\ &= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 = V(X) + V(Y). \end{aligned}$$

Problem:

If f(x,y)= $\begin{cases} x^2 + \frac{xy}{3} & 0 \le x \le 1, 0 \le y \le 2 \\ 0 & elsewhere \end{cases}$. Find E(X), E(Y) and V(Y).

Solution:

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

$= \int_{-\infty}^{\infty} x \left\{2x^2 + \frac{2}{3}x\right\} dx = \frac{13}{18}$

E(Y) = $\int_{-\infty}^{\infty} y$ h(y) dy= $\frac{10}{19}$

E(Y^2) = $\int_{-\infty}^{\infty} y^2$ h(y) dy= $\frac{14}{9}$

V(Y) = E(Y^2) – $\{E(Y)\}^2$= 0.3209

Conditional Expectation:

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

$$E(X \mid y) = \int_{-\infty}^{+\infty} xg(x \mid y)\, dx. \tag{7.23}$$

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

$$E(X \mid y_j) = \sum_{i=1}^{\infty} x_i p(x_i \mid y_j). \tag{7.24}$$

Chebyshev’s inequality

There is a well-known inequality due to the Russian mathematician Chebyshev which will play an important role in our subsequent work. In addition, it will give us a means of understanding precisely how the variance measures variability about the, expected value of a random variable.

If we know the probability distribution of a random variable X (either the pdf in the continuous case or the point probabilities in the discrete case), we may then compute E(X) and V(X), if these exist. However, the converse is not true. That is, from a knowledge of E(X) and V(X) we cannot reconstruct the probability distribution of X.

Nonetheless, it turns out that although we cannot evaluate such probabilities [from a knowledge of E(X) and V(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. Then, if $E(X - c)^2$ is finite and ϵ is any positive number, we have

$$P[|X - c| \geq \epsilon] \leq \frac{1}{\epsilon^2} E(X - c)^2. \tag{7.20}$$

NOTE:

(a) By considering the complementary event we obtain

$$P[|X - c| < \epsilon] \geq 1 - \frac{1}{\epsilon^2} E(X - c)^2. \tag{7.20a}$$

(b) Choosing $c = \mu$ we obtain

$$P[|X - \mu| \geq \epsilon] \leq \frac{\operatorname{Var} X}{\epsilon^2}. \tag{7.20b}$$

(c) Choosing $c = \mu$ and $\epsilon = k\sigma$, where $\sigma^2 = \operatorname{Var} X > 0$, we obtain

$$P[|X - u| \geq k\sigma] \leq k^{-2}. \tag{7.21}$$

NOTE:

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

Problems:

1. Apply Chebyshev's inequality to find (with $\mu = 10$ and $\sigma^2 = 4$)

i) P(5<X<15) Ans: $\geq \frac{21}{25}$ where, k=5/2

ii) P(|X-10|≤ 3) Ans: $\geq \frac{5}{9}$ where, k= 3/2

iii) P(|X-10|<3) Ans: $\geq \frac{4}{9}$

2. A random variable has mean 3 and variance 2. Find an upper bound for

   i) P(|X-3|≥ 2) Ans: $\leq \frac{1}{2}$

   ii) P(|X-3|≥ 1) Ans: ≤ 2

3. Find a smallest value of k in Chebyshev's inequality for which the probability is at most 0.95.

   Solution:

   P(|x-μ|$\leq ka) \geq 1 - 1/k^2$

   $0.95 \geq \; 1 - 1/k^2$

   K= $\sqrt{20}$

Correlation Coefficient

Parameter which measures "degree of association" between X and Y.

Definition. Let (X, Y) be a two-dimensional random variable. We define ρ_{xy}, the *correlation coefficient*, between X and Y, as follows:

$$\rho_{xy} = \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{V(X)V(Y)}}. \tag{7.22}$$

The numerator is called as covariance of X and Y and is denoted by σ_{xy} or COV(X,Y).

Theorem 7.9

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}.$$

Proof: Consider

$$\begin{aligned} E\{[X - E(X)][Y - E(Y)]\} &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned}$$

Theorem 7.10. If X and Y are independent, then $\rho = 0$.

Proof: This follows immediately from Theorem 7.9, since

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

if X and Y are independent.

NOTE:
The converse of Theorem 7.10 is in general not true. That is, we may have ρ= 0, and yet X and Y need not be independent. If p = 0, we say that X and Y are uncorrelated. Thus, being uncorrelated and being independent are, in general, not equivalent.
Example: Consider the RV Y=X^2 where f(x)= ½ , -1≤ $x \leq$ 1.
E(XY)- E(X)E(Y)
= E(X^3)- E(X)E(X^2)
=0
$\rho = 0$ but X and Y are not independent.

Theorem:

$$-1 \leq \rho \leq 1$$

Proof:
We have E$(X) \geq 0$
Since $V(X) \geq 0$
E(X^2)- $[E(X)]^2 \geq 0$

$$E(X^2) \geq 0$$

$$E\left\{\frac{X - E(X)}{\sqrt{V(X)}} \pm \frac{Y - E(Y)}{\sqrt{V(Y)}}\right\}^2 \geq 0$$

Simplification

$$2 \pm 2\rho \geq 0$$

$$1 \pm \rho \geq 0$$

$1 + \rho \geq 0$ and $1 - \rho \geq 0$
$\rho \geq -1$ and $\rho \geq 1$
Hence, $-1 \leq \rho \leq 1.$

Theorem:
If X and Y are linearly related then $\rho = \pm 1.$
Proof:
Let X and Y are linearly related.
Y = a+bX

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}$$

Here: $E(XY) = E(X(a + bX)) = E(Xa + BX^2) = aE(X) + bE(X^2)$
Therefore,

$$\rho = \frac{aE(X) + bE(X^2) - E(X)E(a + bX)}{\sqrt{V(X)V(Y)}}$$

$$\rho = \frac{aE(X) + bE(X^2) - E(X)\big((\,a + bE(X)\big)}{\sqrt{V(X)V(Y)}}$$

$$\rho = \frac{aE(X) + bE(X^2) - aE(X) - bE(X)^2}{\pm b\,V(X)}$$

$$\rho = \frac{E(X^2) - E(X)^2}{\pm V(X)} = \pm 1$$

Problems:

1. With usual notation, prove that $\rho_{uv} = \pm \rho_{xy}$ where u=ax+b and v= c+dx.
2. The random variable (X, Y) has a joint pdf given by
   f(x, y)= x+y, $0 \le x \le 1\,, 0 \le y \le 1$compute correlation between X &Y.
   Solution:
   $$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}}$$
   E(X)= $\int_0^1 \int_0^1 x(x+y)dx\,dy = \frac{7}{12}$
   E(Y)= $\int_0^1 \int_0^1 y(x+y)dx\,dy = \frac{7}{12}$
   E(XY)= $\int_0^1 \int_0^1 xy(x+y)dx\,dy = \frac{1}{3}$
   V(X)= 11/144
   V(Y)= 11/144
   $\rho = \frac{-1}{11}$
3. If (X, Y) has the joint density function f(x,y)= 2-x-y, , $0 \le x \le 1\,, 0 \le y \le 1$ compute the correlation between x and y. Ans: -1/11
4. Prove that V(aX+bY)=$a^2V(X) + b^2V(Y) + 2ab\;Cov\;(X,Y)$.
5. And when X and Y are independent V(aX+bY)=$a^2V(X) + b^2V(Y)$.
6. Two independent random variable X_1 *and* X_2has mean 5 and 10 and variance 4 and 9 respectively. Find the covariance between u = $3x_1 + 4x_2$ v = $3x_1 - x_2$.
7. If X_1, X_2, X_3 be uncorrelated random variables having same standard deviation. Find the correlation coefficient between $X_1 + X_2$ and $X_3 + X_2$ Ans:1/2
8. Suppose that 2 dimensional random variable is uniformly distributed over the triangular region R={(x,y)/ 0<x<y <1}
   i) Find pdf Ans: 2
   ii) Marginal pdf of X and Y Ans: 2(1-x) and 2y
   iii) Find ρ. Ans: E(x)= 1/3, E(Y)= 2/3,

   E(XY)= 1/4, E(X^2)=1/6, E(Y^2)= ½ , V(X)= 1/18, V(Y)= 1/8, $\rho = \frac{1}{2}$

9. The random Variable (X, Y) has a joint pdf by f(x,y)= x+y, $0 \le x \le 1, 0 \le y \le 1$ compute the correlation coefficient between X and Y.
   Ans: g(x)= x+1/2, h(y)= y+1/2 , E(X)= 7/12, E(Y)= 7/12, E(X^2)= 5/12, E(Y^2)= 5/12, V(X) = 11/144, V(Y) = 11/144, E(XY)=1/3 and $\rho = -1/11$
10. Given E(XY)=43, P(X=x_i)= 1/5 and P(Y=y_i)= 1/5. Find ρ.

X	1	3	4	6	8
Y	1	2	24	12	5

   Ans: E(X)=22/5, E(Y)= 44/5, E(X^2)= 126/5, E(Y^2)= 150, V(X)= 5.84, V(Y)= 7256 and $\rho = 0.20791$.

11. If X,Y and Z are uncorrelated random variable with standard deviation 5, 12, 9 respectively. If U=X+Y and V= Y+Z. Evaluate ρ between U and V.
   Ans: COV (X, Y) = COV (Y, Z) = COV (X, Z)= 0
   V(U) = 169, V(V) = 225 and $\rho = 0.73$.

Example 7.21. Suppose that the two-dimensional random variable (X, Y) is uniformly distributed over the triangular region

$$R = \{(x, y) \mid 0 < x < y < 1\}.$$

(See Fig. 7.9.) Hence the pdf is given as

$$f(x, y) = 2, \quad (x, y) \in R,$$
$$= 0, \quad \text{elsewhere.}$$

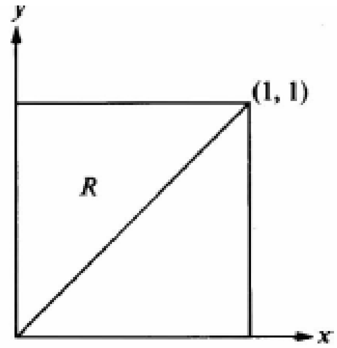

Figure 7.9

Thus the marginal pdf's of X and of Y are

$$g(x) = \int_x^1 (2)\,dy = 2(1-x), \quad 0 \le x \le 1;$$
$$h(y) = \int_0^y (2)\,dx = 2y, \quad 0 \le y \le 1.$$

Therefore

$$E(X) = \int_0^1 x2(1-x)\,dx = \tfrac{1}{3}, \qquad E(Y) = \int_0^1 y2y\,dy = \tfrac{2}{3};$$
$$E(X^2) = \int_0^1 x^2 2(1-x)\,dx = \tfrac{1}{6}, \qquad E(Y^2) = \int_0^1 y^2 2y\,dy = \tfrac{1}{2};$$
$$V(X) = E(X^2) - [E(X)]^2 = \tfrac{1}{18}, \qquad V(Y) = E(Y^2) - [E(Y)]^2 = \tfrac{1}{18};$$
$$E(XY) = \int_0^1 \int_0^y xy2\,dx\,dy = \tfrac{1}{4}.$$

Hence

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{V(X)V(Y)}} = \frac{1}{2}$$